



2019

人工智能发展报告

Report of Artificial Intelligence Development



清华大学-中国工程院知识智能联合研究中心
中国人工智能学会吴文俊人工智能科学技术奖评选基地

2019年12月

編寫委員會 (按姓氏拼音排序)

主 編：李涓子 唐 傑

編 委：曹 楠 程 健 賈 珈 李國良 劉華平

宋德雄 喻 純 余有成 朱 軍

責任編輯：景 晨 劉 佳

編 輯：畢小俊 程時偉 韓 騰 侯 磊 劉德兵

劉 越 駱昱宇 麻曉娟 仇 瑜 王若琳

徐 菁

技術支援：北京智譜華章科技有限公司

CONTESTS

1. 中国人工智能发展报告编制概要

1.1 编制背景	1
1.2 编制目标与方法	1

2. 机器学习

2.1 机器学习概念	4
2.2 机器学习发展历史	6
2.3 机器学习经典算法	7
2.4 深度学习	21
2.5 人才概况	37
2.6 代表性学者简介	39
2.7 论文解读	60

3. 计算机视觉

3.1 计算机视觉概念	85
3.2 计算机视觉发展历史	87
3.3 人才概况	89
3.4 论文解读	91
3.5 计算机视觉进展	105

4. 知识工程

4.1 知识工程概念	107
4.2 知识工程发展历史	108
4.3 人才概况	111
4.4 论文解读	113
4.5 知识工程最新进展	129

5. 自然语言处理

5.1 自然语言处理概念	131
5.2 自然语言的理解发展历史	132
5.3 人才概况	133
5.4 论文解读	136
5.5 自然语言处理最新进展	153

6. 语音识别

6.1 语音识别概念	155
6.2 语音识别发展历史	156
6.3 人才概况	158
6.4 论文解读	160
6.5 语音识别进展	173

7. 计算机图形学

7.1 计算机图形学概念	175
7.2 计算机图形学发展历史	175
7.3 人才概况	178
7.4 论文解读	181
7.5 计算机图形学进展	194

8. 多媒体技术

8.1 多媒体概念	197
8.2 多媒体技术发展历史	198
8.3 人才概况	200
8.4 论文解读	203
8.5 多媒体技术进展	215

9. 人机交互技术

9.1 人机交互概念	217
9.2 人机交互发展历史	218
9.3 人才概况	222
9.4 论文解读	225
9.5 人机交互进展	239

10. 机器人

10.1 机器人概念	241
10.2 机器人发展历史	242
10.3 人才概况	245
10.4 论文解读	247
10.5 机器人进展	260

11.	数据库技术	
	11.1 数据库概念	263
	11.2 数据库技术历史	264
	11.3 人才概况	266
	11.4 论文解读	269
	11.5 数据库技术重要进展	287
12.	可视化技术	
	12.1 可视化技术概念	289
	12.2 可视化技术发展历史	290
	12.3 人才概况	294
	12.4 论文解读	296
	12.5 可视化进展	312
	12.6 可视化应用	314
13.	数据挖掘	
	13.1 数据挖掘概念	321
	13.2 数据挖掘的发展历史	323
	13.3 人才概况	324
	13.4 论文解读	326
	13.5 数据挖掘进展	337
14.	信息检索与推荐	
	14.1 信息检索与推荐概念	339
	14.2 信息检索和推荐技术发展历史	341
	14.3 人才概况	345
	14.4 论文解读	347
	14.5 信息检索与推荐进展	362
15.	结束语	365
	参考文献	366
	附录	372

1 編制概要

1.1 編制背景

21 世紀前兩個十年，在大規模 GPU 伺服器平行計算、大資料、深度學習演算法和類腦晶片等技術的推動下，人類社會相繼進入互聯網時代、大資料時代和人工智慧時代。當前，隨著移動互聯網發展紅利逐步消失，後移動時代已經來臨。當新一輪產業變革席捲全球，人工智慧成為產業變革的核心方向：科技巨頭紛紛把人工智慧作為後移動時代的戰略支點，努力在雲端建立人工智慧服務的生態系統；傳統製造業在新舊動能轉換，將人工智慧作為發展新動力，不斷創造出新的發展機遇。

現今，人工智慧的發展對國民經濟具有重要意義，人工智慧通過綜合各生產要素作用於國民經濟活動，有利於提高生產力水準，助力實體經濟發展，主要表現在以下四個方面：一是人工智慧可以依託大資料，對龐大的資訊資源進行處理，分析得到有效資料，避免了錯誤的經濟決策，推進經濟持續穩定的發展。二是人工智慧可以通過智慧化的精準控制來達到減少資源浪費、提高生產水準和生產效率的目的。三是人工智慧可以賦能於商業生態，以電能為動力源的人工智慧可以做到減少碳排放，達到節能環保的效果。四是在人工智慧的驅動下，產業經濟與資訊經濟相互整合，改變了傳統的“需求-設計-製造-銷售-服務”的生產模式。由於互聯網等資訊技術的應用，使得不同產業間的關聯關係不斷改變，新的產業不斷湧現，跨界和融合發展成為產業生態的重要特徵，提高了經濟增長的品質，推動了經濟整體結構的調整。

人工智慧處於第四次科技革命的核心地位，在該領域的競爭意味著一個國家未來綜合國力的較量。我國在人工智慧領域的發展上有其獨特優勢，如穩定的發展環境、充足的人才儲備、豐富的應用場景等；同時，需要注意的是，我國人工智慧發展起步較晚，與以美國為主的發達國家相比還有一定差距。人工智慧對於任何國家來說既是機遇又是挑戰，世界格局極有可能因此而重新洗牌，對於錯過前三次科技革命的我國來說，此次機遇尤為重要。近年來，我國政府高度重視人工智慧的發展，相繼出臺多項戰略規劃，鼓勵指引人工智慧的發展。2015 年，《國務院關於積極推進“互聯網+”行動的指導意見》頒佈，提出“人工智慧作為重點佈局的 11 個領域之一”；2016 年，在《國民經濟和社會發展第十三個五年規劃綱要（草案）》中提出“重點突破新興領域人工智慧技術”；2017 年，人工智慧寫入十九大報告，提出推動互聯網、大資料、人工智慧和實體經濟深度融合；2018 年，李克強總理在政府工作報告中再次談及人工智慧，提出“加強新一代人工智慧研發應用”；2019 年，習近平主席主持召開中央全面深化改革委員會第七次會議並發表重要講話，會議審議通過了《關於促進人工智慧和實體經濟深度融合的指導意見》。目前，在多層次戰略規劃的指導下，無論是學術界還是產業界，我國在人工智慧國際同行中均有不錯的表現，在世界人工智慧舞臺上扮演了重要的角色，我國人工智慧的發展已駛入快車道。

我國人工智慧的發展也離不開人工智慧團體組織與先進平臺的參與和協助。中國人工智慧學會（Chinese Association for Artificial Intelligence，CAAI）成立於 1981 年，是經國家民政部正式註冊的我國智慧科學技術領域唯一的國家級學會，目前擁有 48 個分支機構，包括 40 個專業委員會和 8 個工作委員會，覆蓋了智慧科學與技術領域，基本任務是團結全國智慧科學技術工作者和積極分子通過學術

研究、國內外學術交流、科學普及、學術教育、科技會展、學術出版、人才推薦、學術評價、學術諮詢、技術評審與獎勵等活動促進我國智慧科學技術的發展，為國家的經濟發展、社會進步、文明提升、安全保障提供智慧化的科學技術服務。

科技情報大資料採擷與服務平臺 (AMiner) 2006 年上線，經過十多年的建設發展，已收錄 2.3 億篇論文與 1.3 億位學者，吸引了全球 220 個國家/地區、800 多萬獨立 IP 的訪問，年度訪問量 1100 萬次。AMiner 平臺曾獲得 2017 年北京市科學技術獎一等獎，2013 年中國人工智慧學會科學技術進步一等獎。AMiner 平臺已經服務於科技部、中國科協、自然科學基金委、北京科委等政府機構，以及騰訊、華為、阿裡巴巴、搜狗等企業機構。人工智慧團體組織與先進平臺的成立和發展已經成為團結優勢資源共同促進人工智慧發展的重要力量，見證並融入到了我國人工智慧偉業的發展。

1.2 編制目標與方法

本報告由清華大學知識智慧聯合研究中心團隊負責編寫。依託于 AMiner 平臺的資料資源及技術挖掘成果生成相關資料包告及圖表，邀請清華大學、同濟大學等高校專家解讀核心技術及提出觀點建議。報告遴選 13 個人工智慧的重點領域進行重點介紹，包括：機器學習、知識工程、電腦視覺、自然語言處理、語音辨識、電腦圖形學、多媒體技術、人機交互、機器人、資料庫技術、視覺化、資料採擷、資訊檢索與推薦等。在述說各領域概念及發展情況等內容的基礎上，報告著重介紹了各領域人才情況以及對代表性文章的解讀。

AMiner 平臺推薦了各領域代表性的期刊/會議，並由專家進行補充，挖掘這些期刊/會議近 10 年論文，確定了 h-index 排名前 2000 的學者，構建各領域學者

庫。我們將這些學者供職機構的位置資訊繪製於地圖上得到了學者分佈地圖，研究各領域學者在世界及我國的分佈規律；同時，我們進一步統計分析了各領域學者性別比例、h-index 分佈等情況。對於中國在各領域的合作情況也進行了挖掘分析，通過統計中文合作論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的情況。

報告還選取這些期刊/會議上發表的高水準論文作為代表，對近年來的熱點及前沿技術進行深度解讀，既包括高引論文、最佳論文，又有專家推薦的代表性工作。解讀前沿熱點研究問題，深入探討研究方法，展現最新研究成果。為讀者瞭解近期人工智慧相關領域的發展動向、基礎及應用研究的代表性成果提供了資訊視窗。

當前，人工智慧正處在爆發期。我國在人工智慧領域的科學技術研究和產業發展起步稍晚，但在最近十餘年的時間裡抓住了機遇，進入了快速發展階段。在這個過程中，技術突破和創造性高端人才對人工智慧的發展有著至關重要的作用。本報告對人工智慧 13 個領域的人才情況及技術發展等內容進行了挖掘分析，希望能對我國人工智慧的發展起到借鑒參考作用。以下各章將對各人工智慧領域的基本概念、發展歷史、人才情況、代表性論文解讀以及近期重要進展進行詳細介紹。

2 機器學習

2.1 機器學習概念

機器學習已經成為了當今的熱門話題，但是從機器學習這個概念誕生到機器學習技術的普遍應用經過了漫長的過程。在機器學習發展的歷史長河中，眾多優秀的學者為推動機器學習的發展做出了巨大的貢獻。

從 1642 年 Pascal 發明的手搖式電腦，到 1949 年 Donald Hebb 提出的赫布理論——解釋學習過程中大腦神經元所發生的變化，都蘊含著機器學習思想的萌芽。

事實上，1950 年圖靈在關於圖靈測試的文章中就已提及機器學習的概念。到了 1952 年，IBM 的亞瑟·撒母耳(Arthur Samuel, 被譽為“機器學習之父”)設計了一款可以學習的西洋跳棋程式。它能夠通過觀察棋子的走位來構建新的模型，用來提高自己的下棋技巧。撒母耳和這個程式進行多場對弈後發現，隨著時間的推移，程式的棋藝變得越來越好^[1]。撒母耳用這個程式推翻了以往“機器無法超越人類，不能像人一樣寫代碼和學習”這一傳統認識。並在 1956 年正式提出了“機器學習”這一概念。他認為“機器學習是在不直接針對問題進程式設計的情況下，賦予電腦學習能力的一個研究領域”。

對機器學習的認識可以從多個方面進行，有著“全球機器學習教父”之稱的 Tom Mitchell 則將機器學習定義為：對於某類任務 T 和性能度量 P ，如果電腦程式在 T 上以 P 衡量的性能隨著經驗 E 而自我完善，就稱這個電腦程式從經驗 E 學習。這些定義都比較簡單抽象，但是隨著對機器學習瞭解的深入，我們會發現隨著時間的變遷，機器學習的內涵和外延在不斷的變化。因為涉及到的領域和應

用很廣，發展和變化也相當迅速，簡單明瞭地給出“機器學習”這一概念的定義並不是那麼容易。

普遍認為，機器學習 (Machine Learning , 常簡稱為 ML) 的處理系統和演算法是主要通過找出資料裡隱藏的模式進而做出預測的識別模式，它是人工智慧 (Artificial Intelligence , 常簡稱為 AI) 的一個重要子領域，而人工智慧又與更廣泛的資料採擷 (Data Mining , 常簡稱為 DM) 和知識發現 (Knowledge Discovery in Database , 常簡稱為 KDD) 領域相交叉。为了更好的理解和區分人工智慧 (Artificial Intelligence)、機器學習(Machine Learning)、資料採擷(Data Mining)、模式識別(Pattern Recognition)、統計(Statistics)、神經計算(Neuro Computing)、資料庫(Databases)、知識發現(KDD)等概念，特繪製其交叉關係如下圖所示：

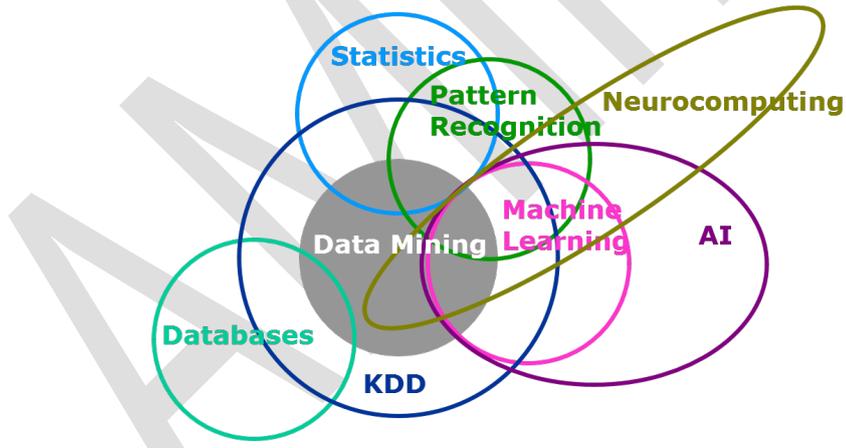


圖 2-1 機器學習相關概念的辨識

機器學習是一門多領域交叉學科，涉及概率論、統計學、逼近論、凸分析、演算法複雜度理論等多門學科。專門研究電腦怎樣類比或實現人類的學習行為，以獲取新的知識或技能，重新組織已有的知識結構使之不斷改善自身的性能。其過程可以用下圖簡單表示：

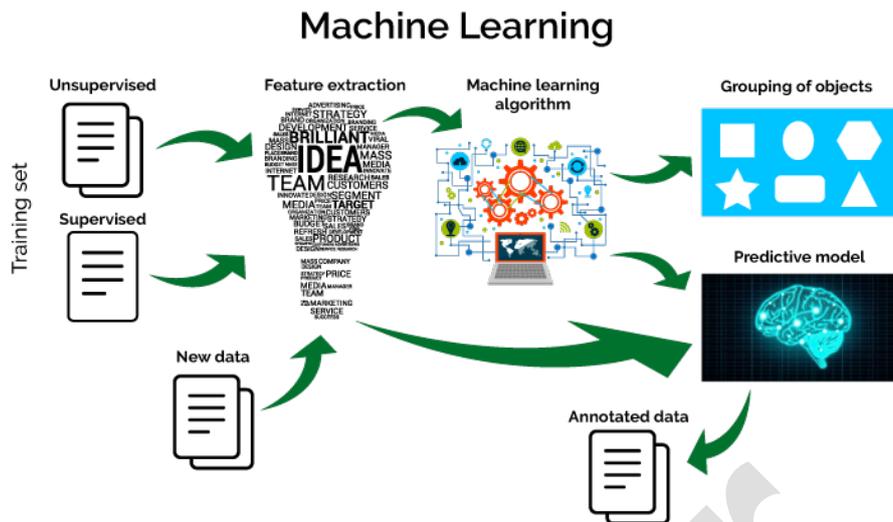


圖 2-2 機器學習基本過程

2.2 機器學習發展歷史

● 奠基時期

1950 年，阿蘭·圖靈創造了圖靈測試來判定電腦是否智慧。圖靈測試認為，如果一台機器能夠與人類展開對話(通過電傳設備)而不能被辨別出其機器身份，那麼稱這台機器具有智慧。這一簡化使得圖靈能夠令人信服地說明“思考的機器”是可能的。

1952，IBM 科學家亞瑟·撒母耳開發的跳棋程式。駁倒了普羅維登斯提出的機器無法超越人類的論斷，像人類一樣寫代碼和學習的模式，他創造了“機器學習”這一術語，並將它定義為：“可以提供電腦能力而無需顯式程式設計的研究領域”。

● 瓶頸時期

從 60 年代中到 70 年代末，機器學習的發展步伐幾乎處於停滯狀態。無論是理論研究還是電腦硬體限制，使得整個人工智慧領域的發展都遇到了很大的瓶頸。

雖然這個時期溫斯頓 (Winston) 的結構學習系統和海斯·羅思 (Hayes Roth) 等的基於邏輯的歸納學習系統取得較大的進展，但只能學習單一概念，而且未能投入實際應用。而神經網路學習機因理論缺陷也未能達到預期效果而轉入低潮。

- 重振時期

偉博斯在 1981 年的神經網路反向傳播 (BP) 演算法中具體提出多層感知機模型。雖然 BP 演算法早在 1970 年就已經以“自動微分的反向模型 (reverse mode of automatic differentiation)”為名提出來了，但直到此時才真正發揮效用，並且直到今天 BP 演算法仍然是神經網路架構的關鍵因素。有了這些新思想，神經網路的研究又加快了。在 1985-1986 年，神經網路研究人員相繼提出了使用 BP 演算法訓練的多參數線性規劃 (MLP) 的理念，成為後來深度學習的基石。在另一個譜系中，昆蘭在 1986 年提出了一種非常出名的機器學習演算法，我們稱之為“決策樹”，更具體的說是 ID3 演算法。在 ID3 演算法提出來以後，研究社區已經探索了許多改進 (如 ID4、回歸樹、CART 演算法等)，這些演算法至今仍然活躍在機器學習領域中。

- 成型時期

支持向量機 (SVM) 的出現是機器學習領域的另一大重要突破，演算法具有非常強大的理論地位和實證結果。那一段時間機器學習研究也分為神經網路 (Neural Network, NN) 和 SVM 兩派。然而，在 2000 年左右提出了帶核函數的支援向量機後，SVM 在許多以前由 NN 占優的任務中獲得了更好的效果。此外，

SVM 相對於 NN 還能利用所有關於凸優化、泛化邊際理論和核函數的深厚知識，因此 SVM 可以從不同的學科中大力推動理論和實踐的改進。

- 爆發時期

神經網路研究領域領軍者 Hinton 在 2006 年提出了神經網路 Deep Learning 演算法，使神經網路的能力大大提高，向支援向量機發出挑戰。2006 年，Hinton 和他的學生 Salakhutdinov 在頂尖學術刊物《Science》上發表了一篇文章，開啟了深度學習在學術界和工業界的浪潮。2015 年，為紀念人工智慧概念提出 60 周年，LeCun、Bengio 和 Hinton 推出了深度學習的聯合綜述。深度學習可以讓那些擁有多個處理層的計算模型來學習具有多層次抽象的資料的表示，這些方法在許多方面都帶來了顯著的改善。深度學習的出現，讓圖像、語音等感知類問題取得了真正意義上的突破，離實際應用已如此之近^[2]，將人工智慧推進到一個新時代。

2.3 機器學習經典演算法

機器學習演算法可以按照不同的標準來進行分類。比如按函數 $f(x, \theta)$ 的不同，機器學習演算法可以分為線性模型和非線性模型；按照學習準則的不同，機器學習演算法也可以分為統計方法和非統計方法。

但一般來說，我們會按照訓練樣本提供的資訊以及回饋方式的不同，將機器學習演算法分為以下幾類：

- 監督學習 (Supervised Learning)

監督學習中的資料集是有標籤的，就是說對於給出的樣本我們是知道答案的。如果機器學習的目標是通過建模樣本的特徵 x 和標籤 y 之間的關係： $f(x, \theta)$ 或

$p(y|x, \theta)$ ，並且訓練集中每個樣本都有標籤，那麼這類機器學習稱為監督學習。根據標籤類型的不同，又可以將其分為分類問題和回歸問題兩類。前者是預測某一樣東西所屬的類別（離散的），比如給定一個人的身高、年齡、體重等資訊，然後判斷性別、是否健康等；後者則是預測某一樣本所對應的實數輸出（連續的），比如預測某一地區人的平均身高。我們大部分學到的模型都是屬於監督學習，包括線性分類器、支援向量機等。常見的監督學習演算法有： k -近鄰演算法（ k -Nearest Neighbors, kNN）、決策樹（Decision Trees）、樸素貝葉斯（Naive Bayesian）等。監督學習的基本流程如下圖所示：

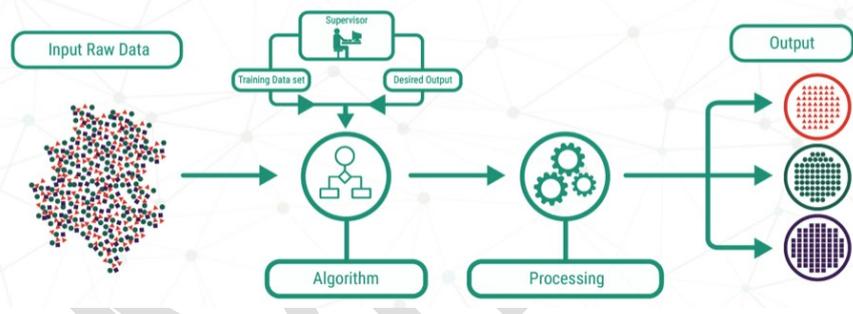


圖 2-3 監督學習的基本流程

- 無監督學習（Unsupervised Learning, UL）

跟監督學習相反，無監督學習中資料集是完全沒有標籤的，依據相似樣本在資料空間中一般距離較近這一假設，將樣本分類。

常見的無監督學習演算法包括：稀疏自編碼（sparse auto-encoder）、主成分分析（Principal Component Analysis, PCA）、K-Means 演算法（K 均值演算法）、DBSCAN 演算法（Density-Based Spatial Clustering of Applications with Noise）、最大期望演算法（Expectation-Maximization algorithm, EM）等。

利用無監督學習可以解決的問題可以分為關聯分析、聚類問題和維度約減。

關聯分析是指發現不同事物之間同時出現的概率。在購物籃分析中被廣泛地應用。如果發現買麵包的客戶有百分之八十的概率買雞蛋，那麼商家就會把雞蛋和麵包放在相鄰的貨架上。

聚類問題是指將相似的樣本劃分為一個簇 (cluster)。與分類問題不同，聚類問題預先並不知道類別，自然訓練資料也沒有類別的標籤。

維度約減：顧名思義，是指減少資料維度的同時保證不丟失有意義的資訊。利用特徵提取方法和特徵選擇方法，可以達到維度約減的效果。特徵選擇是指選擇原始變數的子集。特徵提取是將資料從高維度轉換到低維度。廣為熟知的主成分分析演算法就是特徵提取的方法。

非監督學習的基本處理流程如下圖所示：

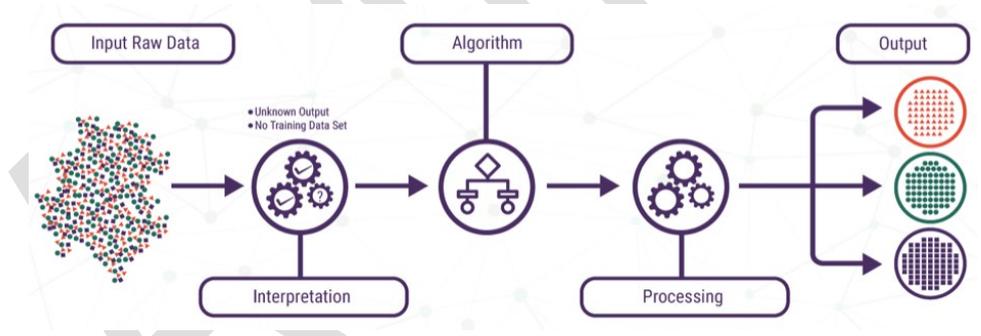


圖 2-4 非監督學習的基本流程

可以很清楚的看到相對於監督學習，非監督學習的過程中沒有監督者 (Supervisor) 的干預。下圖是一個典型的監督學習和非監督學習的對比，左圖是對一群有標籤資料的分類，而右圖是對一群無標籤資料的聚類。

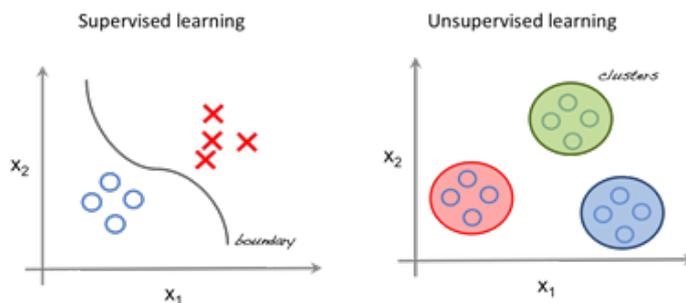


圖 2-5 一個典型的監督學習和非監督學習對比

- 半監督學習 (Semi-Supervised Learning)

半監督學習是監督學習與無監督學習相結合的一種學習方法。半監督學習一般針對的問題是資料量大，但是有標籤資料少或者說標籤資料的獲取很難很貴的情況，訓練的時候有一部分是有標籤的，而有一部分是沒有的。與使用所有標籤資料的模型相比，使用訓練集的訓練模型在訓練時可以更為準確，而且訓練成本更低。常見的兩種半監督的學習方式是直推學習 (Transductive learning) 和歸納學習 (Inductive learning)。

直推學習 (Transductive learning)：沒有標記的資料是測試資料，這個時候可以用測試的資料進行訓練。這裡需要注意，這裡只是用了測試資料中的特徵 (feature) 而沒有用標籤 (label)，所以並不是一種欺騙的方法。

歸納學習 (Inductive learning)：沒有標籤的資料不是測試集。

半監督學習的基本流程如圖 2-6 所示：

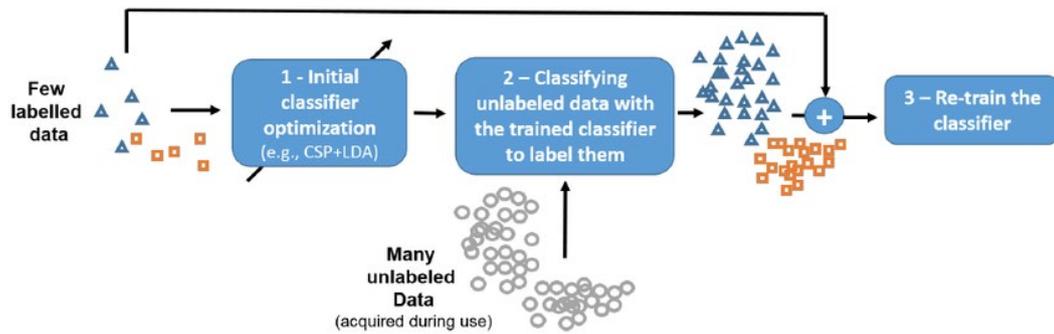


圖 2-6 半監督學習的基本流程

監督學習、半監督學習和非監督學習之間的區別可以用圖 2-7 表示：

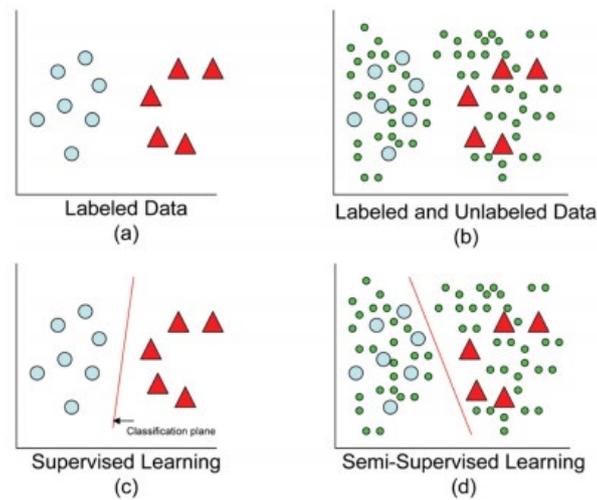


圖 2-7 監督學習、半監督學習和非監督學習的簡單對比

可以看到，圖 2-7 (a) 中，紅色三角形數據和藍色圓點數據為標注數據；圖 2-7 (b) 中，綠色的小圓點為非標注數據。圖 2-7 (c) 顯示監督學習將有標籤的資料進行分類；而半監督學習如圖 2-7 (d) 中部分是有標籤的，部分是沒有標籤的，一般而言，半監督學習側重於在有監督的分類演算法中加入無標記樣本來實現半監督分類。

- 強化學習 (Reinforcement Learning, RL)

強化學習從動物學習、參數擾動自我調整控制等理論發展而來，基本原理是：如果 Agent 的某個行為策略導致環境正的獎賞(強化信號)，那麼 Agent 以後產生這個行為策略的趨勢便會加強。Agent 的目標是在每個離散狀態發現最優策略以使期望的折扣獎賞和最大。

強化學習在機器人學科中被廣泛應用。在與障礙物碰撞後，機器人通過感測器收到負面的回饋從而學會去避免衝突。在視頻遊戲中，可以通過反復試驗採用一定的動作，獲得更高的分數。Agent 能利用回報去理解玩家最優的狀態和當前應該採取的動作。

下圖採用一隻老鼠來模擬強化學習中的 Agent，其任務是走出迷宮，每走一步都有一個方法來衡量其走的好與壞，基本學習過程是當其走得好的時候就給其一定的獎勵（如一塊蛋糕）。通過這種方式，Agent 在行動評價的環境中獲得知識，改進行動方案以適應環境。

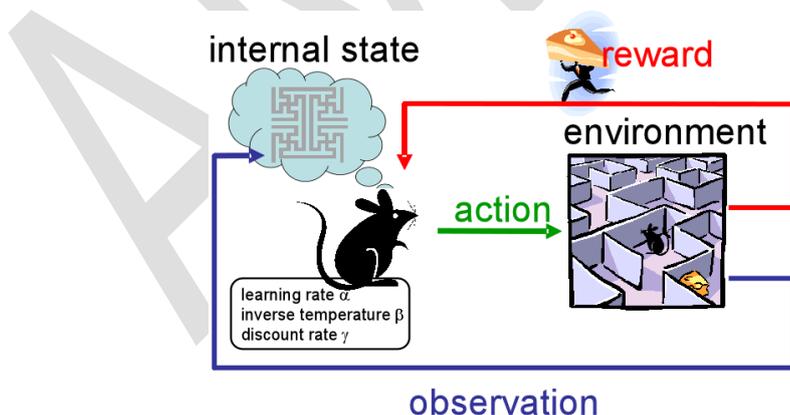


圖 2-8 強化學習的基本學習流程

下面內容對部分機器學習代表演算法進行了介紹。

- 線性回歸

在機器學習中，我們有一組輸入變數 (x) 用於確定輸出變數 (y)。輸入變數和輸出變數之間存在某種關係，機器學習的目標是量化這種關係。

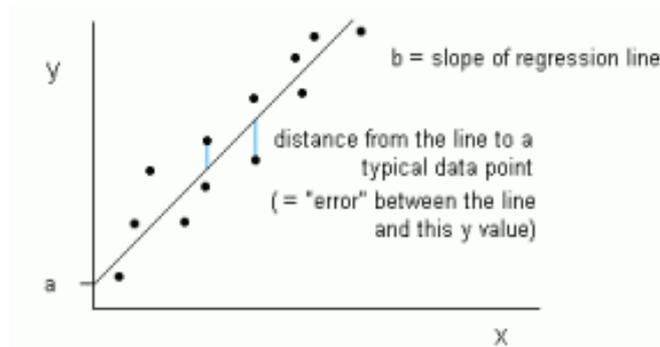


圖 2-9 資料集的繪製 x 和 y 值

在線性回歸中，輸入變數 (x) 和輸出變數 (y) 之間的關係表示為 $y = ax + b$ 的方程。因此，線性回歸的目標是找出係數 a 和 b 的值。這裡， a 是直線的斜率， b 是直線的截距。上圖顯示了資料集的 x 和 y 值，線性回歸的目標是擬合最接近大部分點的線。

- 分類與回歸樹 (CART)

CART 是決策樹的一個實現方式，由 ID3，C4.5 演化而來，是許多基於樹的 bagging、boosting 模型的基礎。CART 可用於分類與回歸。

CART 是在給定輸入隨機變數 x 條件下輸出隨機變數 y 的條件概率分佈，與 ID3 和 C4.5 的決策樹所不同的是，ID3 和 C4.5 生成的決策樹可以是多叉的，每個節點下的叉數由該節點特徵的取值種類而定，比如特徵年齡分為(青年，中年，老年)，那麼該節點下可分為 3 叉。而 CART 為假設決策樹為二叉樹，內部結點特徵取值為“是”和“否”。左分支取值為“是”，右分支取值為“否”。這樣

的決策樹等價於遞迴地二分每一個特徵，將輸入空間劃分為有限個單元，並在這些單元上預測概率分佈，也就是在輸入給定的條件下輸出條件概率分佈。

- 隨機森林 (Random Forest)

隨機森林指的是利用多棵決策樹對樣本進行訓練並預測的一種分類器。它包含多個決策樹的分類器，並且其輸出的類別是由個別樹輸出的類別的眾數而定。

隨機森林是一種靈活且易於使用的機器學習演算法，即便沒有超參數調優，也可以在大多數情況下得到很好的結果。隨機森林也是最常用的演算法之一，因為它很簡易，既可用於分類也能用於回歸。

其基本的構建演算法過程如下：

1. 用 N 來表示訓練用例 (樣本) 的個數， M 表示特徵數目。
2. 輸入特徵數目 m ，用於確定決策樹上一個節點的決策結果；其中 m 應遠小於 M 。
3. 從 N 個訓練用例 (樣本) 中以有放回抽樣的方式，取樣 N 次，形成一個訓練集 (即 bootstrap 取樣)，並用未抽到的用例 (樣本) 作預測，評估其誤差。
4. 對於每一個節點，隨機選擇 m 個特徵，決策樹上每個節點的決定都是基於這些特徵確定的。根據這 m 個特徵，計算其最佳的分裂方式。
5. 每棵樹都會完整成長而不會剪枝，這有可能在建完一棵正常樹狀分類器後被採用)。

一個簡單的隨機森林演算法示意如下：

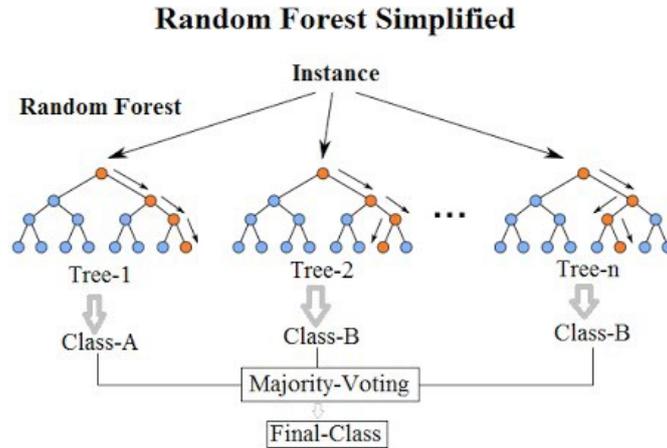


圖 2-10 一個簡單的隨機森林演算法示意

隨機森林集成了所有的分類投票結果，將投票次數最多的類別指定為最終的輸出，這就是一種最簡單的 Bagging 思想。

- 邏輯回歸

邏輯回歸最適合二進位分類 ($y = 0$ 或 1 的資料集，其中 1 表示默認類) 例如：在預測事件是否發生時，發生的事件被分類為 1 。在預測人會生病或不生病，生病的實例記為 1)。它是以其中使用的變換函數命名的，稱為邏輯函數 $h(x) = 1 / (1 + e^{-x})$ ，它是一個 S 形曲線。

在邏輯回歸中，輸出是以缺省類別的概率形式出現的。因為這是一個概率，所以輸出在 $0-1$ 的範圍內。輸出 (y 值) 通過對數轉換 x 值，使用對數函數 $h(x) = 1 / (1 + e^{-x})$ 來生成。然後應用一個閾值來強制這個概率進入二元分類。

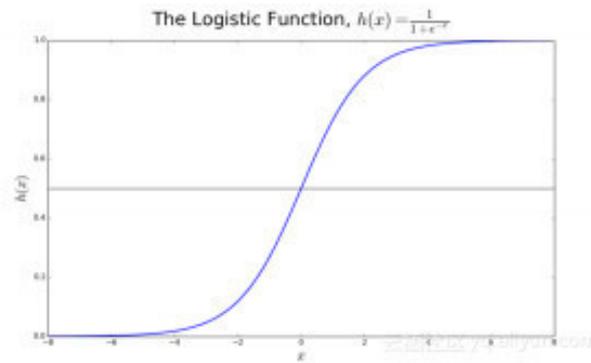


圖 2-11 邏輯函數曲線圖

上圖判斷了腫瘤是惡性還是良性。預設變數是 $y = 1$ (腫瘤=惡性) ; x 變數可以是腫瘤的資訊，例如腫瘤的尺寸。如圖所示，邏輯函數將資料集的各种實例的 x 值轉換成 0 到 1 的範圍。如果概率超過閾值 0.5 (由水平線示出) ，則將腫瘤分類為惡性。

$$p(x) = \frac{e^{b_0 + b_1 x}}{1 + e^{b_0 + b_1 x}}$$

$$\log\left(\frac{p(x)}{1-p(x)}\right) = b_0 + b_1 x$$

邏輯回歸的目標是使用訓練資料來找到係數 b_0 和 b_1 的值，以使預測結果與實際結果之間的誤差最小化。這些係數是使用最大似然估計來計算的。

- 樸素貝葉斯 (Naive Bayesian)

樸素貝葉斯法是基於貝葉斯定理與特徵條件獨立假設的分類方法。樸素貝葉斯分類器基於一個簡單的假定：給定目標值時屬性之間相互條件獨立。

通過以上定理和“樸素”的假定，我們知道：

$$P(\text{Category} | \text{Document}) = P(\text{Document} | \text{Category}) * P(\text{Category}) / P(\text{Document})$$

樸素貝葉斯的基本方法：在統計資料的基礎上，依據條件概率公式，計算當前特徵的樣本屬於某個分類的概率，選擇最大的概率分類。

對於給出的待分類項，求解在此項出現的條件下各個類別出現的概率，哪個最大，就認為此待分類項屬於哪個類別。其計算流程表述如下：

(1) $x = \{a_1, a_2, \dots, a_m\}$ 為待分類項，每個 a_i 為 x 的一個特徵屬性

(2) 有類別集合 $C = \{y_1, y_2, \dots, y_n\}$

(3) 計算 $P(y_1|x), P(y_2|x), \dots, P(y_n|x)$

(4) 如果 $P(y_k|x) = \max\{P(y_1|x)\}$

- k 最近鄰 (kNN)

kNN (k -Nearest Neighbor) 的核心思想是如果一個樣本在特徵空間中的 k 個最相鄰的樣本中的大多數屬於某一個類別，則該樣本也屬於這個類別，並具有這個類別上樣本的特性。該方法在確定分類決策上只依據最鄰近的一個或者幾個樣本的類別來決定待分樣本所屬的類別。kNN 方法在做類別決策時，只與極少量的相鄰樣本有關。由於 kNN 方法主要靠周圍有限的鄰近的樣本，而不是靠判別類域的方法來確定所屬類別的，因此對於類域的交叉或重疊較多的待分樣本集來說，kNN 方法較其他方法更為適合。

kNN 演算法不僅可以用於分類，還可以用於回歸。通過找出一個樣本的 k 個最近鄰居，將這些鄰居的屬性的平均值賦給該樣本，就可以得到該樣本的屬性。

如下圖是 kNN 演算法中， k 等於不同值時的演算法分類結果：

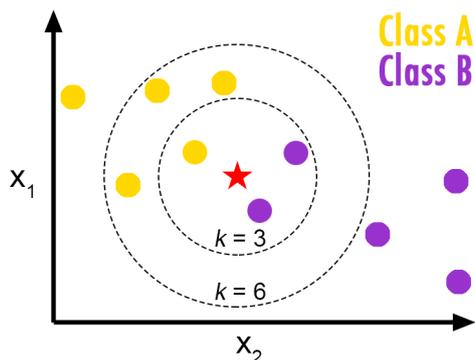


圖 2-12 kNN 演算法簡單示例

簡單來說，kNN 可以看成：有那麼一堆你已經知道分類的資料，然後當一個新資料進入的時候，就開始跟訓練資料裡的每個點求距離，然後挑離這個訓練資料最近的 k 個點，看看這幾個點屬於什麼類型，然後用少數服從多數的原則，給新資料歸類。

- AdaBoost

Adaptive Boosting 或稱為 AdaBoost，是多種學習演算法的融合。它是一種反覆運算演算法，其核心思想是針對同一個訓練集訓練不同的分類器（弱分類器），然後把這些弱分類器集合起來，構成一個更強的最終分類器（強分類器）。其演算法本身是通過改變資料分佈來實現的，它根據每次訓練集之中每個樣本的分類是否正確，以及上次的總體分類的準確率，來確定每個樣本的權值。將修改過權值的新資料集送給下層分類器進行訓練，然後將每次訓練得到的分類器融合起來，作為最終的決策分類器。

AdaBoost 是最常用的演算法。它可用於回歸或者分類演算法。相比其他機器學習演算法，它克服了過擬合的問題，通常對異常值和雜訊資料敏感。為了創建一個強大的複合學習器，AdaBoost 使用了多次反覆運算。因此，它又被稱為“Adaptive Boosting”。通過反覆運算添加弱學習器，AdaBoost 創建了一個強學

習器。一個新的弱學習器加到實體上，並且調整加權向量，作為對前一輪中錯誤分類的樣例的回應。得到的結果，是一個比弱分類器有更高準確性的分類器。

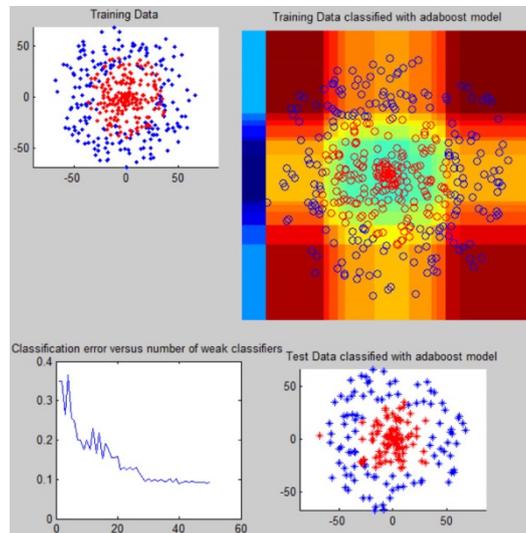


圖 2-13 AdaBoost 執行

AdaBoost 有助於將弱閾值的分類器提升為強分類器。上面的圖像描述了 AdaBoost 的執行，只用了簡單易於理解的代碼在一個檔中就實現了。這個函數包含一個弱分類器和 boosting 元件。弱分類器在一維的資料中嘗試去尋找最理想的閾值來將資料分離為兩類。boosting 組件反覆運算調用分類器，經過每一步分類，它改變了錯誤分類示例的權重。因此，創建了一個級聯的弱分類器，它的行為就像一個強分類器。

目前，對 Adaboost 演算法的研究以及應用大多集中於分類問題，同時近年也出現了一些在回歸問題上的應用。Adaboost 系列主要解決了：兩類問題、多類單標籤問題、多類多標籤問題、大類單標籤問題和回歸問題。它用全部的訓練樣本進行學習。

- K-均值演算法 (K-Means)

K-均值是著名聚類演算法，它找出代表聚類結構的 k 個質心。如果有一個點到某一質心的距離比到其他質心都近，這個點則指派到這個最近的質心所代表的簇。依次，利用當前已聚類的資料點找出一個新質心，再利用質心給新的資料指派一個簇。

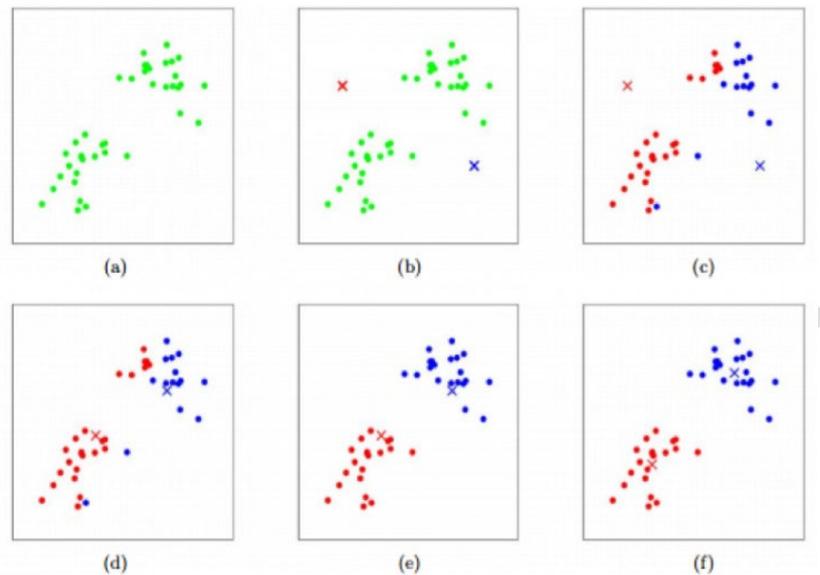


圖 2-14 K-均值演算法圖示

K-均值演算法——在上圖中用“x”表示聚類質心，用點表示訓練樣本：

- a) 原始資料集
- b) 隨機初始化聚類質心
- c) (c-f)k-均值反覆運算 2 次的示意圖

在每次反覆運算中每個訓練樣例都被指派到一個最近的聚類質心，每個聚類質心被移動到分配給它的點的平均值的位置。

- 支持向量機 (SVM)

支持向量機 (Support Vector Machine, SVM) 是一類按監督學習 (supervised learning) 方式對資料進行二元分類 (binary classification) 的廣義線性分類器 (generalized linear classifier) , 其決策邊界是對學習樣本求解的最大邊距超平面 (maximum-margin hyperplane) 。基本思想是 : 找到集合邊緣上的若干資料 (稱為支援向量 (Support Vector)) , 用這些點找出一個平面 (稱為決策面) , 使得支援向量到該平面的距離最大。由簡至繁的 SVM 模型包括 :

- ◆ 當訓練樣本線性可分時 , 通過硬間隔最大化 , 學習一個線性可分支援向量機 ;
- ◆ 當訓練樣本近似線性可分時 , 通過軟間隔最大化 , 學習一個線性支援向量機 ;
- ◆ 當訓練樣本線性不可分時 , 通過核技巧和軟間隔最大化 , 學習一個非線性支援向量機 ;

在分類問題中 , 很多時候有多個解 , 如下圖左邊所示 , 在理想的線性可分的情況下其決策平面會有多個。而 SVM 的基本模型是在特徵空間上找到最佳的分離超平面使得訓練集上正負樣本間隔最大 , SVM 演算法計算出來的分界會保留對類別最大的間距 , 即有足夠的餘量 , 如下圖右邊所示。

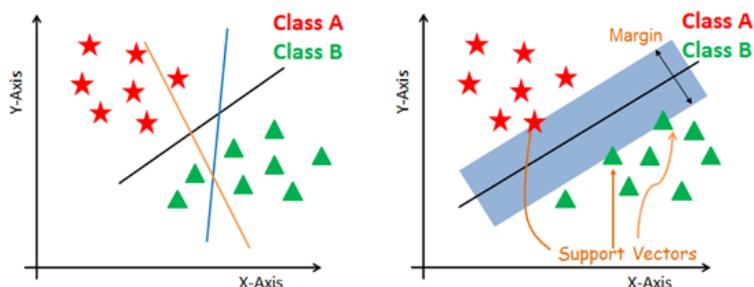


圖 2-15 SVM 的決策平面

在解決線性不可分問題時，它可以通過引入核函數，巧妙地解決了在高維空間中的內積運算，從而很好地解決了非線性分類問題。如下圖所示，通過核函數的引入，將線性不可分的資料映射到一個高維的特徵空間內，使得資料在特徵空間內是可分的。如下圖所示：

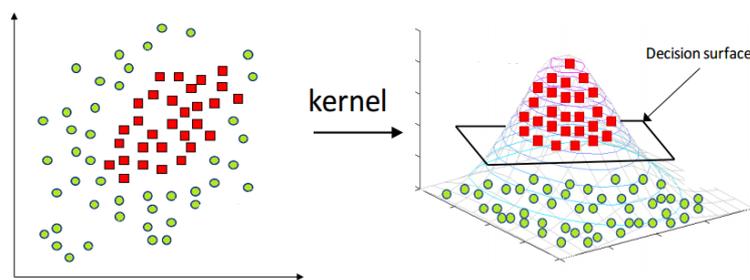


圖 2-16 SVM 的核函數

- 人工神經網路 ANN (Artificial Neural Network)

人工神經網路 ANN (Artificial Neural Network) 是由大量處理單元互聯組成的非線性、自我調整資訊處理系統。它是一種模仿動物神經網路行為特徵，進行分散式並行資訊處理的演算法數學模型。其基本過程可以概述如下：外部刺激通過神經末梢，轉化為電信號，傳導到神經細胞 (又叫神經元) ；無數神經元構成神經中樞；神經中樞綜合各種信號，做出判斷；人體根據神經中樞的指令，對外部刺激做出反應。其過程表述如下圖所示：

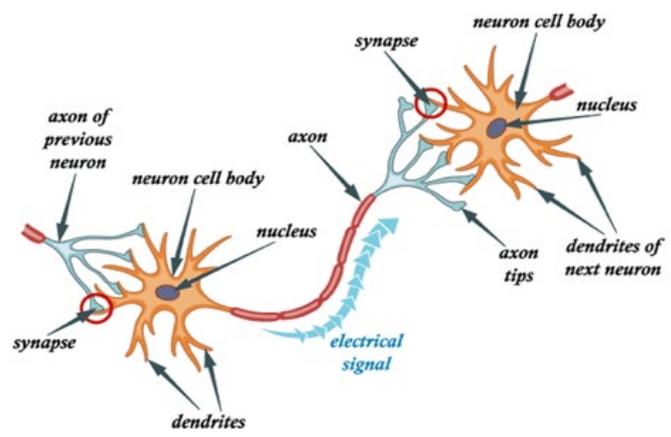


圖 2-17 神經網路的傳導過程

人工神經網路經歷了漫長的發展階段。最早是上個世紀六十年代提出的“人造神經元”模型，叫做“感知器”(perceptron)。感知機模型是機器學習二分類問題中的一個非常簡單的模型。它的基本結構如下圖所示：

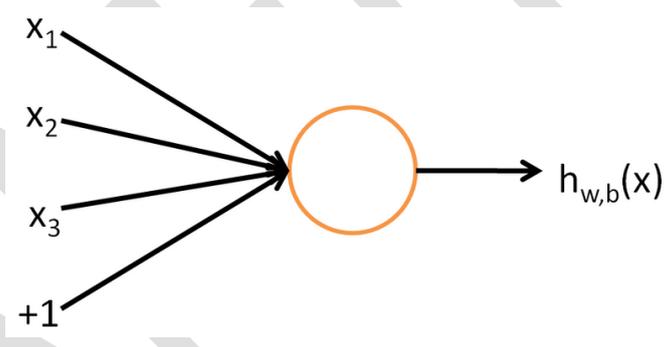


圖 2-18 感知機

隨著反向傳播演算法、最大池化 (max-pooling) 等技術的發明，神經網路進入了飛速發展的階段。神經網路就是將許多個單一“神經元”聯結在一起，這樣，一個“神經元”的輸出就可以是另一個“神經元”的輸入。典型的人工神經網路具有以下三個部分：

結構 (Architecture) 結構指定了網路中的變數和它們的拓撲關係。

激勵函數 (Activity Rule) 大部分神經網路模型具有一個短時間尺度的動力學規則，來定義神經元如何根據其他神經元的活動來改變自己的激勵值。

學習規則 (Learning Rule) 指定了網路中的權重如何隨著時間推進而調整。

一個典型的人工神經網路結構如下圖所示：

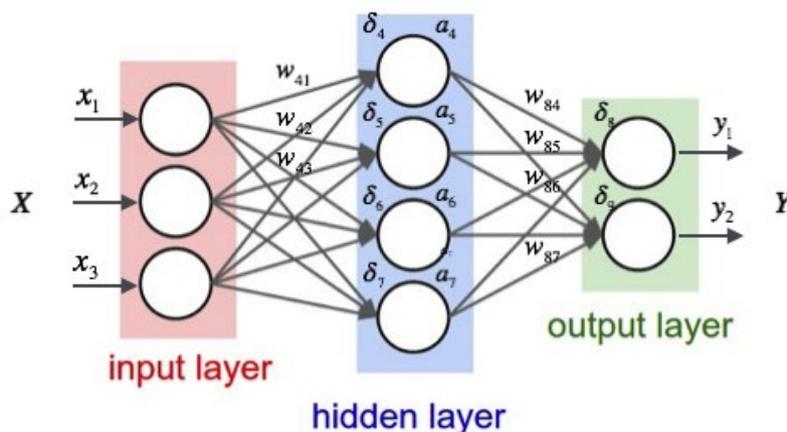


圖 2-19 典型的人工神經網路結構

人工神經網路具有四個基本特徵：非線性、非局限性、非常定性和非凸性。

人工神經網路的特點和優越性，主要表現在三個方面：具有自學習功能、具有聯想存儲功能和具有高速尋找最優解的能力。

2.4 深度學習

深度學習是近 10 年機器學習領域發展最快的一個分支，由於其重要性，三位教授 (Geoffrey Hinton、Yann Lecun、Yoshua Bengio) 因此同獲圖靈獎。深度學習模型的發展可以追溯到 1958 年的感知機 (Perceptron)。1943 年神經網路就已經出現雛形 (源自 Neuroscience)，1958 年研究認知的心理學家 Frank 發明了感知機，當時掀起一股熱潮。後來 Marvin Minsky (人工智慧大師) 和 Seymour

Papert 發現感知機的缺陷：不能處理異或回路等非線性問題，以及當時存在計算能力不足以處理大型神經網路的問題。於是整個神經網路的研究進入停滯期。

最近 30 年來取得快速發展。總體來說，主要有 4 條發展脈絡。

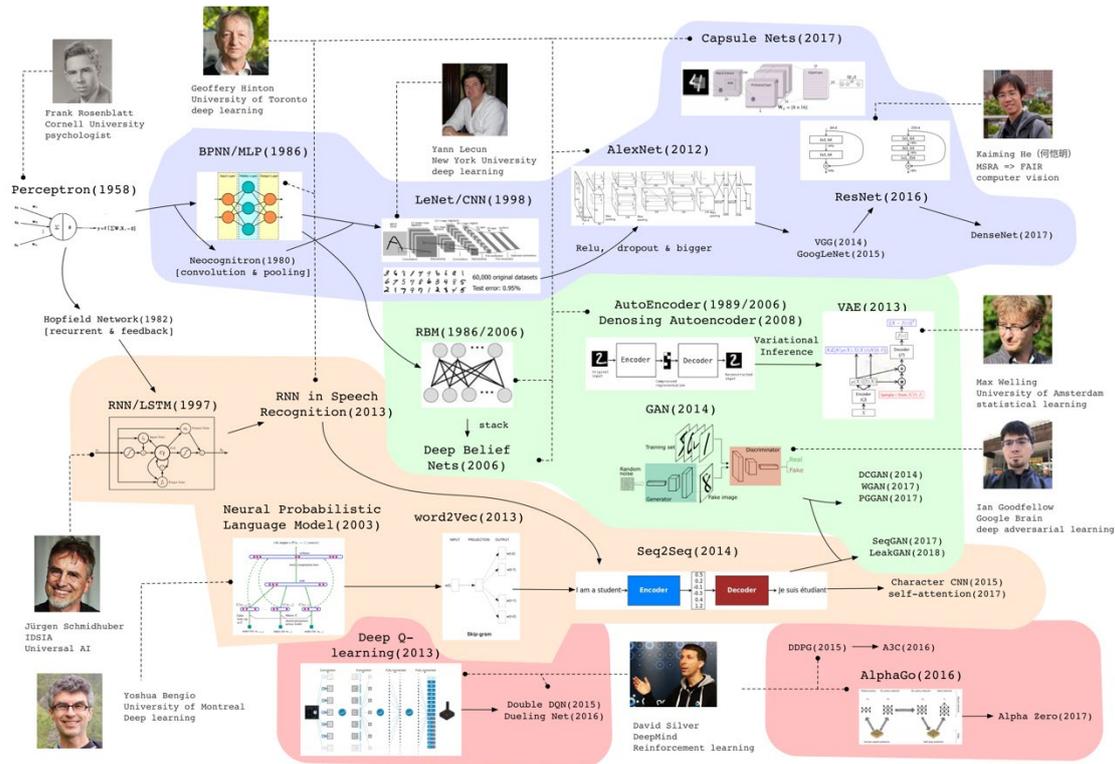


圖 2-20 深度學習模型最近若干年的重要進展

第一個發展脈絡 (上圖淺紫色區域) 以電腦視覺和卷積網路為主。這個脈絡的進展可以追溯到 1979 年，Fukushima 提出的 Neocognitron。該研究給出了卷積和池化的思想。1986 年 Hinton 提出的反向傳播訓練 MLP (之前也有幾個類似的研究)，該研究解決了感知機不能處理非線性學習的問題。1998 年，以 Yann LeCun 為首的研究人員實現了一個七層的卷積神經網路 LeNet-5 以識別手寫數位。現在普遍把 Yann LeCun 的這個研究作為卷積網路的源頭，但其實在當時由於 SVM 的迅速崛起，這些神經網路的方法還沒有引起廣泛關注。真正使得卷積神經網路榮耀登上大雅之堂的事件是，2012 年 Hinton 組的 AlexNet (一個設計

精巧的 CNN 在 ImageNet 上以巨大優勢奪冠，這引發了深度學習的熱潮。AlexNet 在傳統 CNN 的基礎上加上了 ReLU、Dropout 等技巧，並且網路規模更大。這些技巧後來被證明非常有用，成為卷積神經網路的標配，被廣泛發展，於是後來出現了 VGG、GoogLeNet 等新模型。2016 年，青年電腦視覺科學家何愷明在層次之間加入跳躍連接，提出殘差網路 ResNet。ResNet 極大增加了網路深度，效果有很大提升。一個將這個思路繼續發展下去的是近年的 CVPR Best Paper 中黃高提出的 DenseNet。在電腦視覺領域的特定任務出現了各種各樣的模型（Mask-RCNN 等），這裡不一一介紹。2017 年，Hinton 認為反向傳播和傳統神經網路還存在一定缺陷，因此提出 Capsule Net，該模型增強了可解釋性，但目前在 CIFAR 等資料集上效果一般，這個思路還需要繼續驗證和發展。

第二個發展脈絡（上圖淺綠色區域）以生成模型為主。傳統的生成模型是要預測聯合概率分佈 $P(x, y)$ 。機器學習方法中生成模型一直佔據著一個非常重要的地位，但基於神經網路的生成模型一直沒有引起廣泛關注。Hinton 在 2006 年的時候基於受限玻爾茲曼機（RBM，一個 19 世紀 80 年代左右提出的基於無向圖模型的物理模型）設計了一個機器學習的生成模型，並且將其堆疊成為 Deep Belief Network，使用逐層貪婪或者 wake-sleep 的方法訓練，當時模型的效果其實並沒有那麼好。但值得關注的是，正是式基於 RBM 模型 Hinton 等人開始設計深度框架，因此這也可以看做深度學習的一個開端。Auto-Encoder 也是上個世紀 80 年代 Hinton 就提出的模型，後來隨著計算能力的進步也重新登上舞臺。Bengio 等人又提出了 Denoise Auto-Encoder，主要針對資料中可能存在的噪音問題。Max Welling（這也是變分和概率圖模型的高手）等人後來使用神經網路訓練一個有一層隱變數的圖模型，由於使用了變分推斷，並且最後長得跟 Auto-Encoder 有點

像，被稱為 Variational Auto-Encoder。此模型中可以通過隱變數的分佈採樣，經過後面的 Decoder 網路直接生成樣本。生成對抗模型 GAN(Generative Adversarial Network)是 2014 年提出的非常火的模型，它是一個通過判別器和生成器進行對抗訓練的生成模型，這個思路很有特色，模型直接使用神經網路 G 隱式建模樣本整體的概率分佈，每次運行相當於從分佈中採樣。後來引起大量跟隨的研究，包括：DCGAN 是一個相當好的卷積神經網路實現，WGAN 是通過維爾斯特拉斯距離替換原來的 JS 散度來度量分佈之間的相似性的工作，使得訓練穩定。PGGAN 逐層增大網路，生成逼真的人臉。

第三個發展脈絡 (上圖橙黃色區域) 是序列模型。序列模型不是因為深度學習才有的，而是很早以前就有相關研究，例如有向圖模型中的隱瑪律科夫 HMM 以及無向圖模型中的條件隨機場模型 CRF 都是非常成功的序列模型。即使在神經網路模型中，1982 年就提出了 Hopfield Network，即在神經網路中加入了遞迴網路的思想。1997 年 Jürgen Schmidhuber 發明了長短期記憶模型 LSTM (Long-Short Term Memory)，這是一個里程碑式的工作。當然，真正讓序列神經網路模型得到廣泛關注的還是 2013 年 Hinton 組使用 RNN 做語音辨識的工作，比傳統方法高出一大截。在文本分析方面，另一個圖靈獎獲得者 Yoshua Bengio 在 SVM 很火的時期提出了一種基於神經網路的語言模型 (當然當時機器學習還是 SVM 和 CRF 的天下)，後來 Google 提出的 word2vec (2013) 也有一些反向傳播的思想，最重要的是給出了一個非常高效的實現，從而引發這方面研究的熱潮。後來，在機器翻譯等任務上逐漸出現了以 RNN 為基礎的 seq2seq 模型，通過一個 Encoder 把一句話的語義資訊壓成向量再通過 Decoder 轉換輸出得到這句話的翻譯結果，後來該方法被擴展到和注意力機制 (Attention) 相結合，也大大擴展了

模型的表示能力和實際效果。再後來，大家發現使用以字元為單位的 CNN 模型在很多語言任務也有不俗的表現，而且時空消耗更少。Self-attention 實際上就是採取一種結構去同時考慮同一序列局部和全域的資訊，Google 有一篇很有名的文章“attention is all you need”把基於 Attention 的序列神經模型推向高潮。當然 2019 年 ACL 上同樣有另一篇文章給這一研究也稍微降了降溫。

第四個發展脈絡(上圖粉色區域)是增強學習。這個領域最出名的當屬 Deep Mind，圖中標出的 David Silver 博士是一直研究 RL 的高管。Q-learning 是很有名的傳統 RL 演算法，Deep Q-learning 將原來的 Q 值表用神經網路代替，做了一個打磚塊的任務。後來又應用在許多遊戲場景中，並將其成果發表在 Nature 上。Double Dueling 對這個思路進行了一些擴展，主要是 Q-Learning 的權重更新時序上。DeepMind 的其他工作如 DDPG、A3C 也非常有名，它們是基於 Policy Gradient 和神經網路結合的變種。大家都熟知的 AlphaGo，裡面其實既用了 RL 的方法也有傳統的蒙特卡洛搜索技巧。Deep Mind 後來提出了一個用 AlphaGo 框架，但通過主學習來玩不同（棋類）遊戲的新演算法 Alpha Zero。

下面對深度學習的不同方面進行分別解讀。有些地方解讀可能稍微會簡單一些，不完整的地方還請見諒。

2.4.1 卷積神經網路

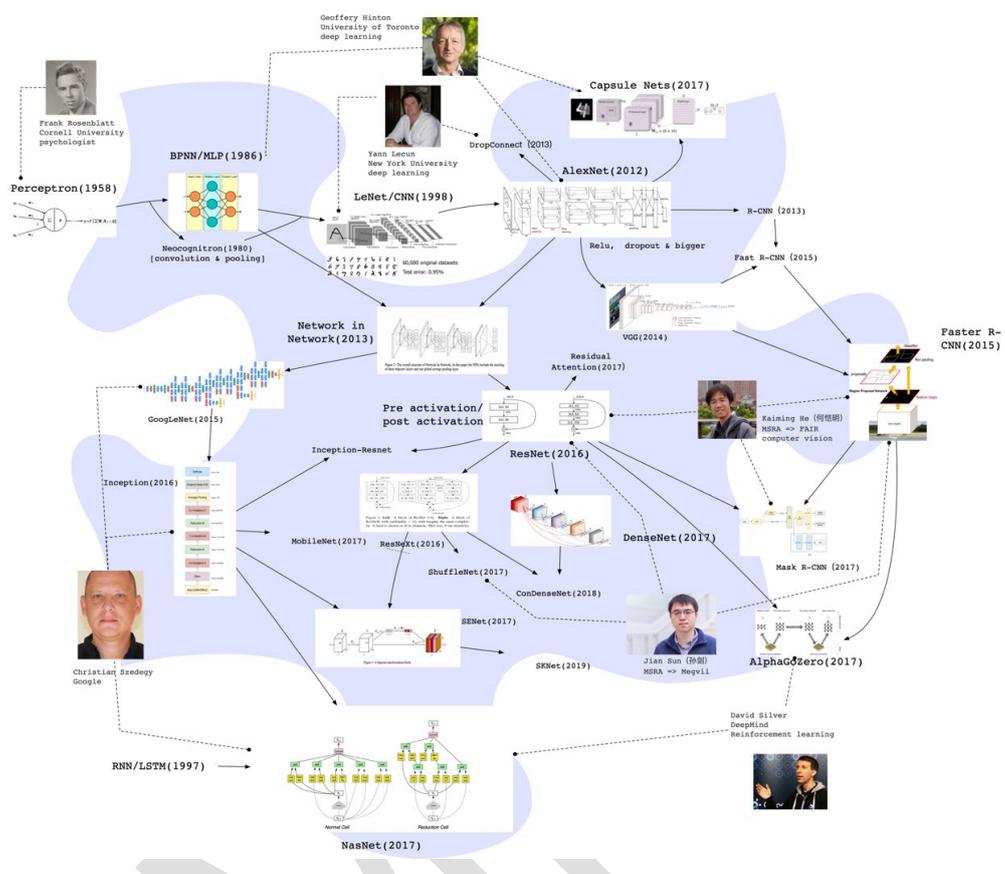


圖 2-21 卷積神經網路的重要進展

卷積神經網路的發展，最早可以追溯到 1962 年，Hubel 和 Wiesel 對貓大腦中的視覺系統的研究。1980 年，一個日本科學家福島邦彥(Kunihiko Fukushima) 提出了一個包含卷積層、池化層的神經網路結構。在這個基礎上，Yann Lecun 將 BP 演算法應用到這個神經網路結構的訓練上，就形成了當代卷積神經網路的雛形。

其實最初的 CNN 效果並不算好，而且訓練也非常困難。雖然也在閱讀支票、識別數字之類的任務上有一定的效果，但由於在一般的實際任務中表現不如 SVM、Boosting 等演算法好，因此一直處於學術界的邊緣地位。直到 2012 年，ImageNet 圖像識別大賽中，Hinton 組的 AlexNet 引入了全新的深層結構和 Dropout

方法，一下子把 error rate 從 25%降低到了 15%，這顛覆了圖像識別領域。AlexNet 有很多創新，儘管都不是很難的方法。其最主要的結果是讓人們意識到原來那個福島邦彥提出的、Yann LeCun 優化的 LeNet 結構原來是有很大改進空間的：只要通過一些方法能夠加深這個網路到 8 層左右，讓網路表達能力提升，就能得到出人意料的好結果。

順著 AlexNet 的思想，LeCun 組 2013 年提出一個 DropConnect，把 error rate 降低到了 11%。而 NUS 的顏水成組則提出了一個重要的 Network in Network (NIN) 方法，NIN 的思想是在原來的 CNN 結構中加入了一個 1×1 conv 層，NIN 的應用也得到了 2014 年 ImageNet 另一個挑戰——圖像檢測的冠軍。Network in Network 更加引發了大家對 CNN 結構改變的大膽創新。因此，兩個新的架構 Inception 和 VGG 在 2014 年把網路加深到了 20 層左右，圖像識別的 error rate (越小越好) 也大幅降低到 6.7%，接近人類錯誤率的 5.1%。2015 年，MSRA 的任少卿、何愷明、孫劍等人，嘗試把 identity 加入到卷積神經網路中提出 ResNet。最簡單的 Identity 卻出人意料的有效，直接使 CNN 能夠深化到 152 層、1202 層等，error rate 也降到了 3.6%。後來，ResNeXt, Residual-Attention, DenseNet, SENet 等也各有貢獻，各自引入了 Group convolution, Attention, Dense connection, channelwise-attention 等，最終 ImageNet 上 error rate 降到了 2.2%，大大超過人類的錯誤率。現在，即使手機上的神經網路，也能達到超過人類的水準。而另一個挑戰——圖像檢測中，也是任少卿、何愷明、孫劍等優化了原先的 R-CNN, fast R-CNN 等通過其他方法提出 region proposal，然後用 CNN 去判斷是否是 object 的方法，提出了 faster R-CNN。Faster R-CNN 的主要貢獻是使用和圖像識別相同的方法，提出了 faster R-CNN。Faster R-CNN 的主要貢獻是使用和圖像識別相同的 CNN feature，發現 feature 不僅可以識別圖片內容，還可以用來識別圖片的位

置。也就是說，CNN 的 feature 非常有用，包含了大量的資訊，可以同時用來做不同的任務。這個創新一下子把圖像檢測的 MAP 也翻倍了。在短短的 4 年中，ImageNet 圖像檢測的 MAP (越大越好) 從最初的 0.22 達到了最終的 0.73。何愷明後來還提出了 Mask R-CNN，即給 faster R-CNN 又加了一個 Mask Head，發現即使只在訓練中使用 Mask Head，其資訊可以傳遞回原先的 CNN feature 中，獲得了更精細的資訊。由此，Mask R-CNN 得到了更好的結果。何愷明在 2009 年時候就以一個簡單有效的去霧演算法得到了 CVPR Best Paper，在電腦視覺領域聲名鵲起。後來更是提出了 ResNet 和 Faster R-CNN 兩大創新，直接顛覆了整個電腦視覺/機器學習領域。

當然，CNN 結構變得越來越複雜，很多結構都很難直覺的來解釋和設計。於是穀歌提出了自動架構學習方法 NasNet (Neural Architecture Search Network) 來自動用 Reinforcement Learning 去搜索一個最優的神經網路結構。Nas 是目前 CV 界一個主流的方向，可以自動尋找出最好的結構，以及給定參數數量/運算量下最好的結構(這樣就可以應用於手機)，是目前圖像識別的一個重要發展方向。今年何愷明 (2019 年 4 月) 又發表了一篇論文，表示即使 Random 生成的網路連接結構 (只要按某些比較好的 Random 方法)，都會取得非常好的效果，甚至比標準的好很多。Random 和 Nas 哪個是真的正確的道路，這有待進一步的研究了。

卷積神經網路 CNN 的發展引發了其他領域的很多變革。比如：利用 CNN，AlphaGo 戰勝了李世石，攻破了圍棋 (基礎版本的 AlphaGo 其實和人類高手比起來是有勝有負的)。後來利用了 ResNet 和 Faster-RCNN 的思想，一年後的 Master 則完全戰勝了所有人類圍棋高手。後來又有很多複現的開源圍棋 AI，每一個都

能用不大的計算量超過所有的人類高手。以至於現在人們講棋的時候，都是按著 AI 的勝率來講了。

2.4.2 AutoEncoder

AutoEncoder 的基本思想是利用神經網路來做無監督學習，就是把樣本的輸入同時作為神經網路的輸入和輸出。本質上是希望學習到輸入樣本的表示 (encoding)。早期 AutoEncoder 的研究主要是資料過於稀疏、資料高維導致計算複雜度高。比較早用神經網路做 AutoEncoder 的可以追溯到 80 年代的 BPNN 和 MLP 以及當時 Hinton 推崇的 RBM。後來到了 2000 年以後還堅持在做的只剩下 Hinton 的 RBM 了。從 2000 年以後，隨著神經網路的快速興起，AutoEncoder 也得到快速發展，基本上有幾條線：稀疏 AutoEncoder、噪音容忍 AutoEncoder、卷積 AutoEncoder、變分 AutoEncoder。最新的進展是結合對抗思想的對抗 AutoEncoder。

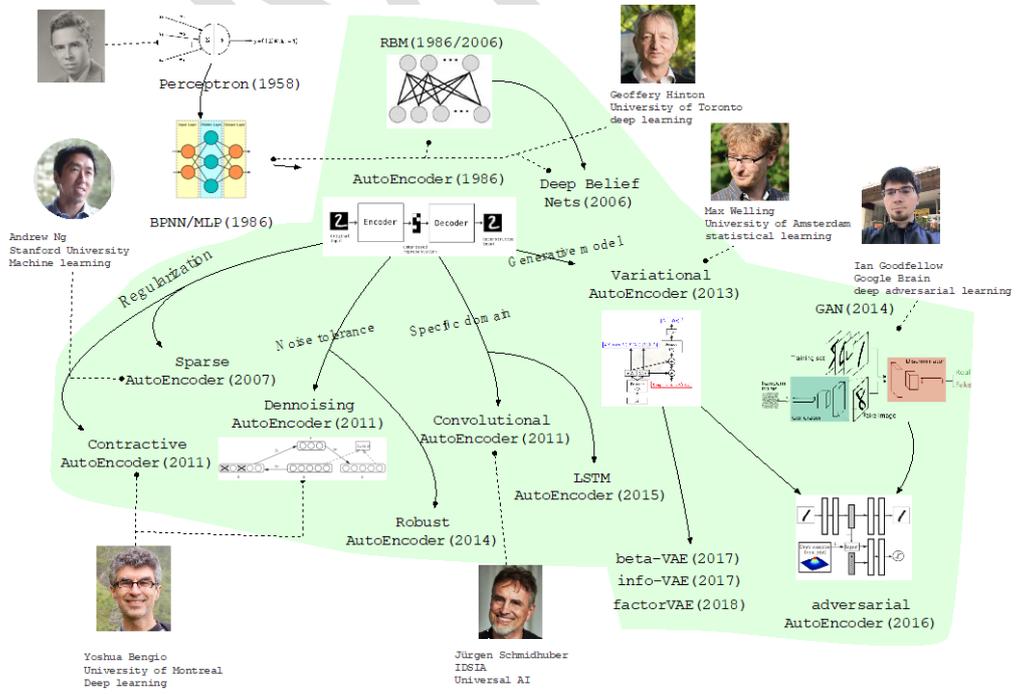


圖 2-22 Auto-Encoder 的重要進展

稀疏 AutoEncoder 在學習輸入樣本表示的時候可以學習到相對比較稀疏的表示結果，這在 Overcomplete AutoEncoder (就是學習得到高維表示) 方法中尤為重要。代表性人物包括斯坦福大學的 Andrew Ng 和蒙特利爾的 Yoshua Bengio 教授。具體方法就是在原來的損失函數中加一個控制稀疏化的正則化項，通過控制優化過程來實現。

Denoising AutoEncoder 的核心思想就是提高 Encoder 的魯棒性，本質上就是避免可能的 overfitting。一個辦法是在輸入中加入隨機噪音 (比如隨機置 0 一些輸入，或者隨機把部分輸入變為 marked)，這些思想後來在 BERT 等模型中也有廣泛使用；另一個辦法就是結合正則化的思想，比如在目標函數中加上 eEncoder 的 Jacobian 範數。Jacobian 範數可以讓學習到的特徵表示更具有差異性。

著名研究者 Jurgen Schmidhuber 提出了基於卷積網路的 AutoEncoder 以及後來的 LSTM AutoEncoder。Max Welling 基於變分思想提出變分 AutoEncoder 方法 VAE，這也是一個里程碑式的研究成果。後面很多研究者在這個工作上進行了擴展，包括 info-VAE、beta-VAE 和 factorVAE 等。最近還有人借鑒 Ian Goodfellow 等人提出的對抗建模思想提出 Adversarial AutoEncoder，也取得了很好的效果。這和之前的噪音容忍的 AE 學習也有一定呼應。除了上面的思想，就是可以把上面的各種方法 stacking 起來。

2.4.3 迴圈神經網路 RNN

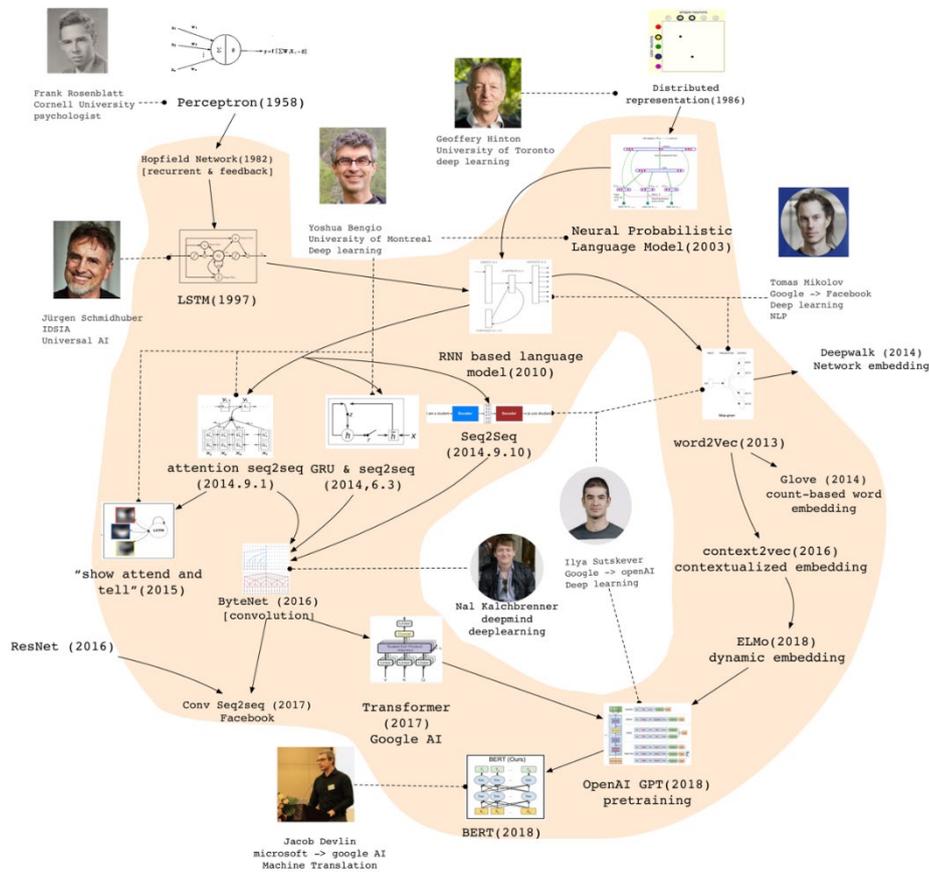


圖 2-23 迴圈神經網路 RNN 的重要進展

1982 年，美國加州理工學院物理學家 John Hopfield 發明了一種單層回饋神經網路 Hopfield Network，用來解決組合優化問題。這是最早的 RNN 的雛形。86 年，另一位機器學習的泰斗 Michael I. Jordan 定義了 Recurrent 的概念，提出 Jordan Network。1990 年，美國認知科學家 Jeffrey L. Elman 對 Jordan Network 進行了簡化，並採用 BP 演算法進行訓練，便有了如今最簡單的包含單個自連接節點的 RNN 模型。但此時 RNN 由於梯度消失 (Gradient Vanishing) 及梯度爆炸 (Gradient Exploding) 的問題，訓練非常困難，應用非常受限。直到 1997 年，瑞士人工智慧研究所的主任 Jurgen Schmidhuber 提出長短期記憶 (LSTM)，LSTM 使用門控單元及記憶機制大大緩解了早期 RNN 訓練的問題。同樣在 1997 年，Mike

Schuster 提出雙向 RNN 模型 (Bidirectional RNN)。這兩種模型大大改進了早期 RNN 結構，拓寬了 RNN 的應用範圍，為後續序列建模的發展奠定了基礎。此時 RNN 雖然在一些序列建模任務上取得了不錯的效果，但由於計算資源消耗大，後續幾年一直沒有太大的進展。

2010 年，Tomas Mikolov 對 Bengio 等人提出的 feedforward Neural network language model (NNLM)進行了改進，提出了基於 RNN 的語言模型(RNN LM)，並將其用在語音辨識任務中，大幅提升了識別精度。在此基礎上 Tomas Mikolov 於 2013 年提出了大名鼎鼎的 word2vec，與 NNLM 及 RNNLM 不同，word2vec 的目標不再專注於建模語言模型，而是如何利用語言模型學習每個單詞的語義化向量 (distributed representation)，當然 distributed representation 概念最早要來源於 Hinton 1986 年的工作。word2vec 引發了深度學習在自然語言處理領域的浪潮，除此之外還啟發了 knowledge representation，network representation 等新的領域。

另一方面，2014 年 Bengio 團隊與 Google 幾乎同時提出了 seq2seq 架構，將 RNN 用於機器翻譯。沒過多久，Bengio 團隊又提出注意力 Attention 機制，對 seq2seq 架構進行改進。自此機器翻譯全面進入到神經機器翻譯(NMT)的時代，NMT 不僅過程簡單，而且效果要遠超統計機器翻譯的效果。目前主流的機器翻譯系統幾乎都採用了神經機器翻譯的技術。除此之外，Attention 機制也被廣泛用於基於深度學習的各種任務中。

近兩年，相關領域仍有一些突破性進展，2017 年，Facebook 人工智慧實驗室提出基於卷積神經網路的 seq2seq 架構，將 RNN 替換為帶有門控單元的 CNN，提升效果的同時大幅加快了模型訓練速度。此後不久，Google 提出 Transformer

架構，使用 Self-Attention 代替原有的 RNN 及 CNN，更進一步降低了模型複雜度。在詞表示學習方面，Allen 人工智慧研究所 2018 年提出上下文相關的表示學習方法 ELMo，利用雙向 LSTM 語言模型對不同語境下的單詞，學習不同的向量表示，在 6 個 NLP 任務上取得了提升。OpenAI 團隊在此基礎上提出預訓練模型 GPT，把 LSTM 替換為 Transformer 來訓練語言模型，在應用到具體任務時，與之前學習詞向量當作特徵的方式不同，GPT 直接在預訓練得到的語言模型最後一層接上 Softmax 作為任務輸出層，然後再對模型進行微調，在多項任務上 GPT 取得了更好的效果。

不久之後，Google 提出 BERT 模型，將 GPT 中的單向語言模型拓展為雙向語言模型 (Masked Language Model)，並在預訓練中引入了 sentence prediction 任務。BERT 模型在 11 個任務中取得了最好的效果，是深度學習在 NLP 領域又一個里程碑式的工作。BERT 自從在 arXiv 上發表以來獲得了研究界和工業界的極大關注，感覺像是打開了深度學習在 NLP 應用的潘朵拉魔盒。隨後湧現了一大批類似於“BERT”的預訓練 (pre-trained) 模型，有引入 BERT 中雙向上下文資訊的廣義自回歸模型 XLNet，也有改進 BERT 訓練方式和目標的 RoBERTa 和 SpanBERT，還有結合多工以及知識蒸餾 (Knowledge Distillation) 強化 BERT 的 MT-DNN 等。這些種種，還被大家稱為 BERTology。

2.4.4 網路表示學習與圖神經網路 (GNN)

這個方面的研究可以追溯到 Hinton 當年 1986 的 Distributed Representation，後來 Stanford 的 Andrew Ng 實驗室做了個 Neural Tensor Network，本質就是把知識之間的關係和表示學習一起放到 tensor 裡面來做，算是一個 smart 的擴展。後

來 Facebook 的 Antonie Bordes 提出了 TransE 是一個 milestone 的工作，把知識網路的三元組融合到了表示學習中，這是 NLP 和知識圖譜中的一個非常重要的研究，後面延續了一系列的工作，包括 TransH、TransR、TransA、TransG。

從表示學習本身來看，Neural Language Model 是對於單詞和文本的表示，是對原來向量模型的一個自然擴展，其實本質上類似一個隱含語義分析，只是這裡用的是神經網路來做學習。RNN based language model 是利用 RNN 進行表示學習，更好的保持了語言模型的連續性。但這個階段的研究當時大部分都沒有火起來，一是當時深度學習還沒火起來，二是這些演算法都還比較慢。2013 年 Tomas Mikolov 和 Jeff Dean 等人做 word2vec，可以說佔據了“天時、地利、人和”：深度學習開始發熱、演算法簡單有效、大神作品。現在 word2vec 已經輕鬆超過 1 萬多引用了。後面的擴展也很多，如 paragraph2vec、doc2vec，context2vec。以至於後面有一段時間，“2vec”成了流行取名字的方法。最近的進展是 ELMo、OpenAI 的 GPT 和穀歌的 BERT。

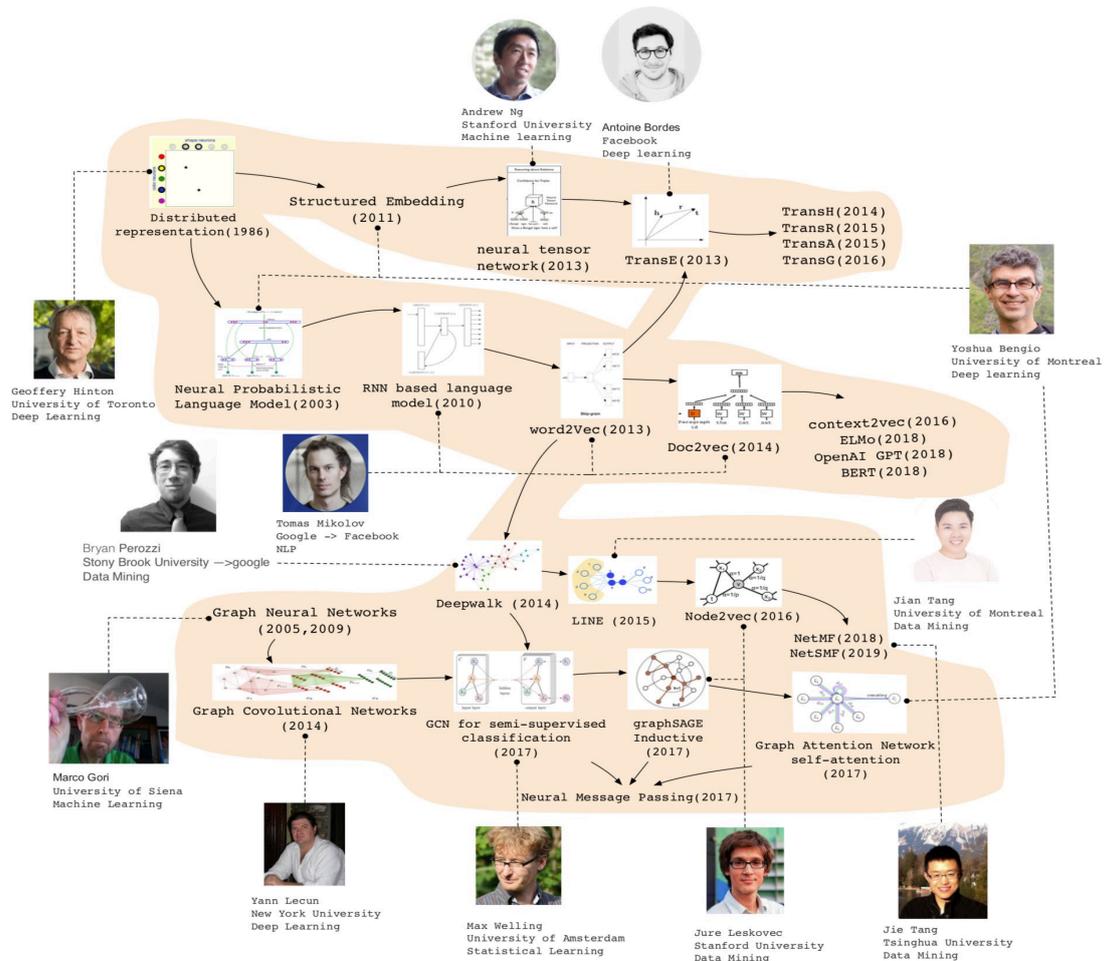


圖 2-24 網路表示學習與圖神經網路的重要進展

表示學習的另一個脈絡就是擴展到網路資料上，在 NLP 領域的 Structured Embedding、TransE 等模型更多的是語言中的局部結構資訊，而網路中還有更加複雜的拓撲結果。Bryan(原 Stony Brook 大學的，現在去了穀歌)提出 DeepWalk，這個算最早把 word2vec 稍微擴展了一下，應用於網路資料，這篇文章獲得了當年 KDD 的最佳論文和後來 KDD 的最佳博士論文。很快這個工作吸引了大量關注，Jian Tang(原北大、微軟，現在去了 Bengio 那邊)等人做了兩階擴展，斯坦福的 Jure Leskovec 做了面向社交網路的“三階”擴展 node2vec，後來清華也給出了一個理論證明，證明這些不同方法本質上都在做一個矩陣分解，並基於此提出了一個 NetMF 的演算法以及其適用於大規模網路的實現 NetSMF。ProNE 是

另一個清華作品，其主要特點是高效和高精度。該方法非常簡單，本質上是在原來的表示學習上引入了一個類似卷積但又不是卷積的操作，大大提高了精度。

最近的網路表示學習更多的是用卷積網路直接對圖做，大方向是 Graph Neural Network，最早是 Siena 大學的 Marco 等人在 2005 和 2009 年提出的，但當時沒引起太大關注。後來 Yann Lecun 提出的 Graph Convolutional Networks，還有 Kipf & Welling 等人提出的 semi-supervised 的 GCN。這一系列的研究本質上就是 Neural Message Passing，在最近引起大量關注。斯坦福的 Jure 也提出了 GraphSage，利用 NMP 簡化了卷積，提高了速度，並且支援 inductive learning，再後來 Yoshua 他們團隊又提出了 Graph Attention Network，進一步提高了圖卷積精度。最近網路表示學習非常熱，前前後後都能看到三大巨頭 Hinton、Yoshua 和 Yann 的影子。在未來若干年還會繼續是個研究熱點。

2.4.5 增強學習

Deep Mind 的是一家英國人工智慧公司，這是一個對增強學習影響最大的公司。創立於 2010 年，2014 年被 Google 收購。創始人哈薩比斯出身于倫敦，母親為新加坡華裔，13 歲便已經獲得國際象棋大師的頭銜，19 歲開始學習圍棋，當前是圍棋業餘初段。DeepMind 于 2014 年開始開發 AlphaGO。來看看 AlphaGO 的戰績吧。2015 年 10 月，AlphaGO 5:1 樊麾；2016 年 3 月，AlphaGo 4:1 李世石；2017 年 5 月，AlphaGO 3:0 柯潔；2017 年 10 月 19 日，AlphaGo Zero 發表在 Nature，其思路是從零開始，自我對弈，40 天超過所有版本。2018 年 12 月 7 日，AlphaZero 再次發表於 Science，AlphaZero 使用與 AlphaGo Zero 類似但更一般性的演算法，在不作太多改變的前提下，並將演算法從圍棋延伸到將棋與國際

象棋上。2018 年 12 月，Deep Mind 公司推出 AlphaFold，可以根據基因序列預測蛋白質結構。2019 年 1 月 25 日，Deep Mind 公司 AlphaStar，在《星海爭霸 II》以 10：1 戰勝人類職業玩家。另一條在美國的戰線，可能最著名的是 Open AI 公司，這是 Hinton 的高徒 Ilya Sutskever (AlexNet 發明人) 創立的。2019 年 4 月，Open AI 推出 five dota2，2-0 戰勝 Dota2 的 TI8 冠軍戰隊 OG。

在研究方法上 Deep Q-Network(DQN)利用神經網路對 Q 值進行函數近似，並利用了 experience replay 和 fixed target network 的策略讓 DQN 可以收斂，在 Atari 的不少遊戲上都超過了人類水準。Double DQN 是深度學習版本的 double Q-learning，它通過微小的修改就成功減小了 DQN 中 max 操作帶來的 bias。再後來，Dueling DQN 將 Q-network 分成了 action-dependent 和 action-independent 兩個部分，從而提高了 DQN。DQN 是為 Value 的期望建模，greedy 的時候也是最大化期望的形式，Categorical DQN 的想法是直接為 Value 的分佈進行建模。Noise DQN 在網路中添加了雜訊，從而達到 exploration 的效果。DQN 還有非常多的提升版本，rainbow 整合了多種 DQN 版本。Ape-X 從 Rainbow 的工作中發現 Replay 的優先順序對於性能影響是最大的，故擴大 Prioritised Replay Buffer，並使用 360 個 actor 做分散式的訓練，比 rainbow 更快，也更好。

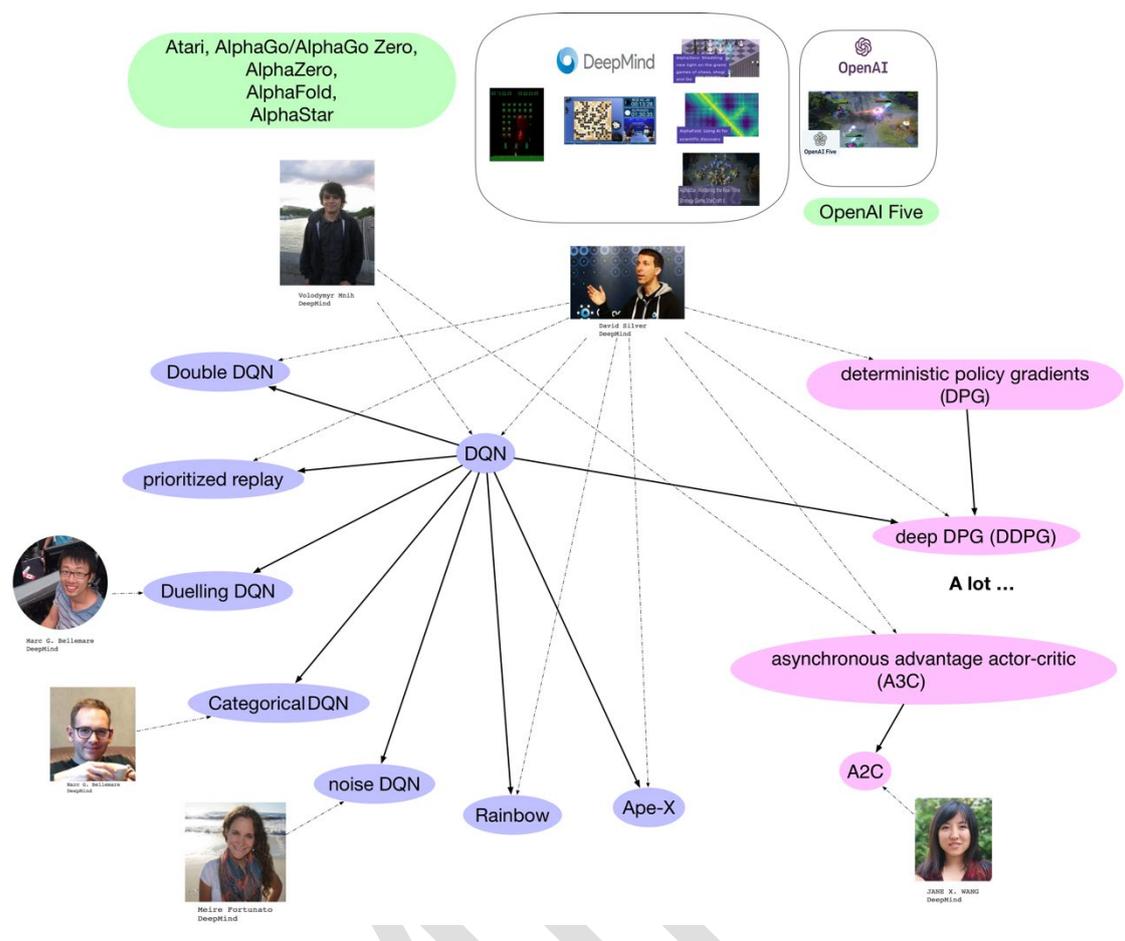


圖 2-25 增強學習的重要進展

Deterministic policy gradients (DPG) 將 policy gradients 方法中隨機的 policy 推廣為確定性 policy。Deep DPG 使用了神經網路表示高維 state, 是結合了 DQN 和 DPG 的 actor critic 演算法。A3C 是經典的 policy gradient 方法, 可以並行 multiple agent 的訓練, 並非同步更新參數。A2C 是 A3C 的同步、確定性 policy 版本, 同步的梯度更新, 可以讓並行訓練更快收斂。

2.4.6 生成對抗網路

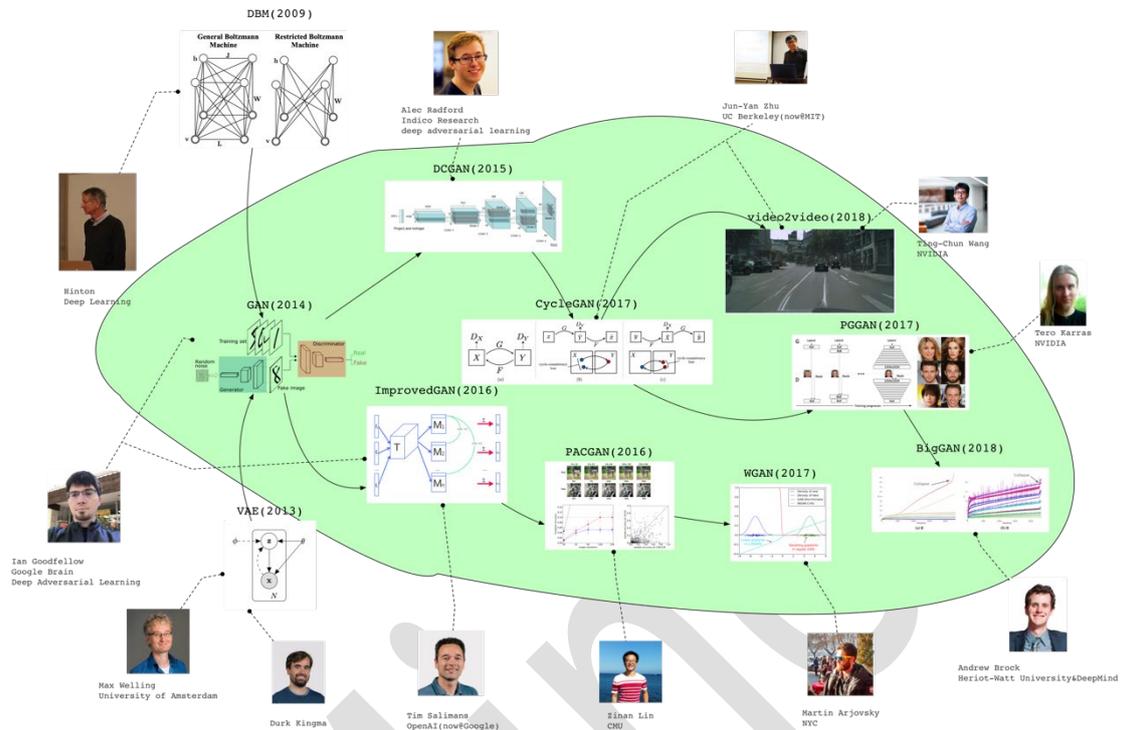


圖 2-26 生成對抗網路的重要進展

GAN 最近幾年發展非常快，這也是 Yoshua Bengio 獲得圖靈獎的貢獻之一。傳統的生成模型是要預測聯合概率分佈 $P(x, y)$ 。首先玻爾茲曼機 (RBM) 這個模型其實是一個基於能量的模型，1986 年的時候就有，Hinton 在 2006 年的時候重新拿出來作為一個生成模型，並且將其堆疊成為 Deep Belief Network，使用逐層貪婪或者 wake-sleep 的方法訓練。

AutoEncoder 也是上個世紀 80 年代 Hinton 就提出的模型，此時由於計算能力的進步也重新登上舞臺。Bengio 等人又搞了 Denoise AutoEncoder。Max welling 等人使用神經網路訓練一個有一層隱變數的圖模型，由於使用了變分推斷，並且最後長得跟 AutoEncoder 有點像，被稱為 Variational AutoEncoder。此模型中可以通過隱變數的分佈採樣，經過後面的 decoder 網路直接生成樣本。

在生成模型方面，最近一個最重要的進展就是對抗生成網路 (GAN) ，可以說是現在最火的生成模型。2014 年 Ian Goodfellow 在 NIPS 上發表了最初的 GAN 文章，到現在已經有近九千引用。為什麼這個模型引起如此大的關注呢？一個原因是這個模型理論上非常優雅，大家理解起來簡單方便；二就是效果確實好。看上面這一排，是基於 GAN 的一些應用文章，下面這些是改進 GAN 的訓練的一些文章。這些文章都引起了廣泛關注。可以看出，GAN 的發明人 Ian Goodfellow 是少年得志的典範。他本科在斯坦福，碩士在 Andrew Ng 手下，博士就跑到蒙特利爾 Yoshua Bengio 手下了。他另外還有一個導師 Aaron Courville。大家現在經常用的教科書《Deep Learning》，作者就是 Ian Goodfellow 和他兩個博士導師。他是 85 年人，發表 GAN 在 2014 年，29，還差一年才 30。GAN 這個工作也給 Goodfellow 帶來了很多榮譽，比如 17 年就被 MIT under 35 選中了。Goodfellow 博士畢業後去了 Google Brain，後來又跳到 Open AI，又跳回 Google，現在在蘋果做特別項目機器學習項目負責人，實際上現在他也才 34 歲。另外，GAN 是 Ian Goodfellow 在蒙特利爾的時候的工作。大家知道今年圖靈獎給了深度學習三巨頭，其中的 Bengio，在圖靈獎官網上給獲獎理由，選的三個貢獻之一就是 GAN。另外兩個貢獻分別是 90 年代的序列概率模型和 00 年代的語言模型。GAN 可以說是 Bengio 的代表作之一了，甚至可以說幫助他拿圖靈獎。

另外還有幾個有名的 GAN 的擴展，包括：cycleGAN 和 vid2vid。去年 NIPS 企業展示會場，英偉達把 vid2vid 配合方向盤，做了個實物 demo，非常引人關注。

2.4.7 老虎機

老虎機也是機器學習的一個重要分支，和深度學習有著或多或少的聯繫。老虎機實際上是個賭博機器。走進拉斯維加斯賭場，你就能看到一排排閃亮的機器。老虎機模型這個數學模型，現在追本溯源基本認為是一個病理學家 Thompson 在 1933 年提出的。他當時覺得驗證新藥的醫學的隨機雙盲實驗有些殘酷的地方，對於被分到藥效較差的新藥的那一組病人並不公平。他想知道能否在實驗中途就驗證藥物藥效，從而避免給病人帶來痛苦，因此他提出了一個序列決策模型。但是，實際使用還是有很多問題，比如中途效果不好評價。所以直到現在，美國 FDA 對在醫學隨機雙盲實驗中使用這種自我調整調整的多臂老虎機方法，仍然只是建議使用。就現在而言，老虎機模型實際是在搜索和推薦方面的應用很多。

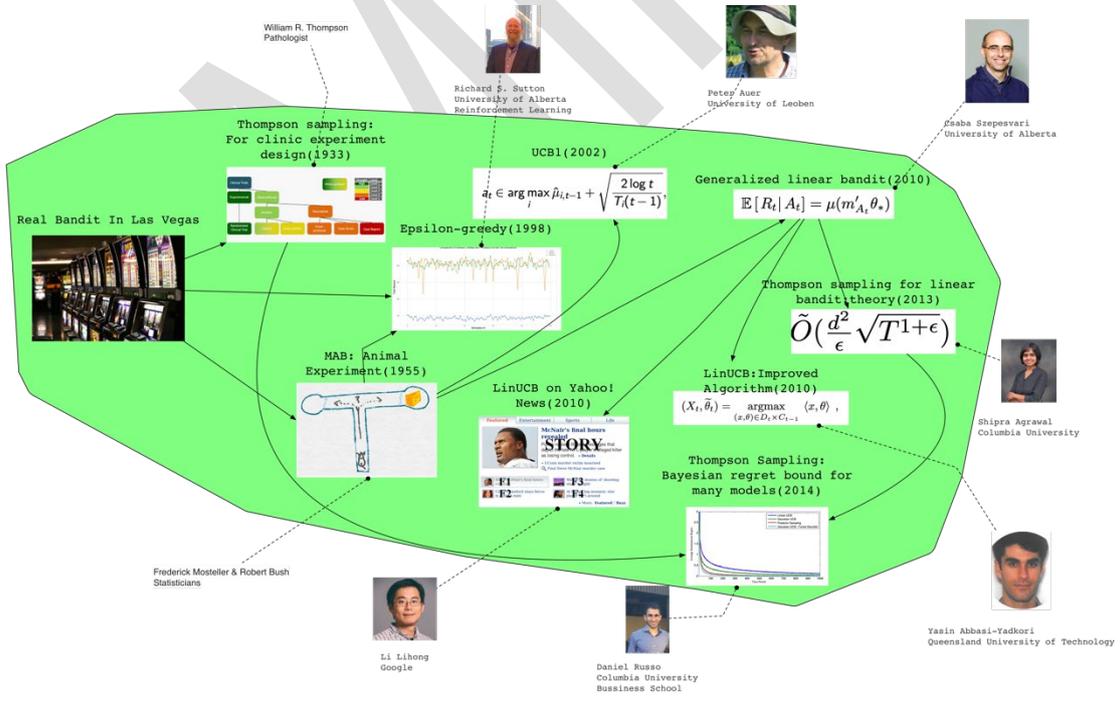


圖 2-27 老虎機的重要進展

Epsilon-greedy 是種預留一點點機會去嘗試的思想。這種想法很自然，學術界也不清楚最初的 credit 該給誰。現在就放在 Sutton 名下。他是強化學習方面的

大佬，寫的那本教材 *Reinforcement Learning* 引用五位數，裡邊講解了這個演算法。Peter Auer 這個工作不僅分析了 UCB 演算法的理論性質，還順道分析了 Epsilon-greedy 的理論性質。這篇文章用到的技術，是此後很多更複雜技術的基礎，很值得一看。這篇純理論文章的引用量也達到了兩千多。Frederick Mosteller 是哈佛統計系奠基人，20 世紀統計學界的超級牛人。他們當時做老虎機模型，主要是想給真實的動物或者人的序列決策建模，想抽象一個框架出來。所以他們作了一個老鼠找蛋糕的實驗。當然，也做了關於人玩賭博用的老虎機的實驗。Li Lihong 是清華 02 級校友。他在 Yahoo! news 上的 LinUCB 的工作發表在 WWW 上，這篇應用文章獲得了大量關注，引用上千。他後來又翻出來 Thompson sampling 這個很古舊的方法，作了一些系統性的實驗，從實驗結果的角度說明 Thompson sampling 效果很好。這篇文章發在 NIPS2011 上，也獲得了大量關注。後來大批做理論的人就跟進，就把 Thompson sampling 在線性模型上的理論基礎建立起來了。比如 Russo 這篇文章。那可以看到，從 Thompson 1933 年用 Thompson sampling，到 2010 年後這個方法的理論基礎才建立起來，這個時間跨度是很大的。當然，因為線性情況下都還比較簡單，所以 2011 年後收到廣泛關注沒幾年，理論就建立。這個現象和神經網路的理論建立基本是一個樣子，都是線性的容易又基礎，就先做著。研究老虎機模型確實比較偏理論，但老虎機應用也很廣。上圖裡邊除了有做醫學的、做統計的、做電腦科學的，還有在商學院任教的，就是這個 Russo。

2.5 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為機器學習領域全球學者分佈情況：



圖 2-28 機器學習領域全球學者分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量遙遙領先且主要分佈在其東西海岸；歐洲中西部也有較多的人才分佈；亞洲的人才主要分佈於我國東部及日韓地區；其他諸如非洲、南美洲等地區的學者非常稀少；機器學習領域的人才分佈與各地區的科技、經濟實力情況大體一致。此外，在性別比例方面，機器學習領域中男性學者占比 89.8%，女性學者占比 10.2%，男性學者占比遠高於女性學者。

- h-index 分佈

機器學習學者的 h-index 分佈如下圖所示，大部分學者的 h-index 都在 20 以上，其中 h-index 在 20-30 區間的人數最多，有 584 人，占比 28.8%，小於 20 區間的人數最少，共 7 人。

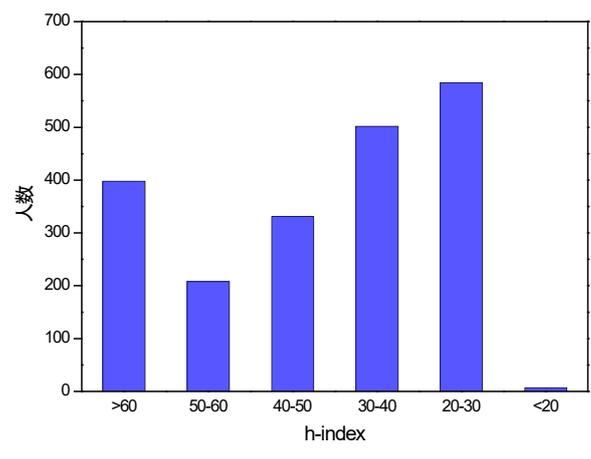


圖 2-29 機器學習領域學者 h-index 分佈

- 中國人才分佈

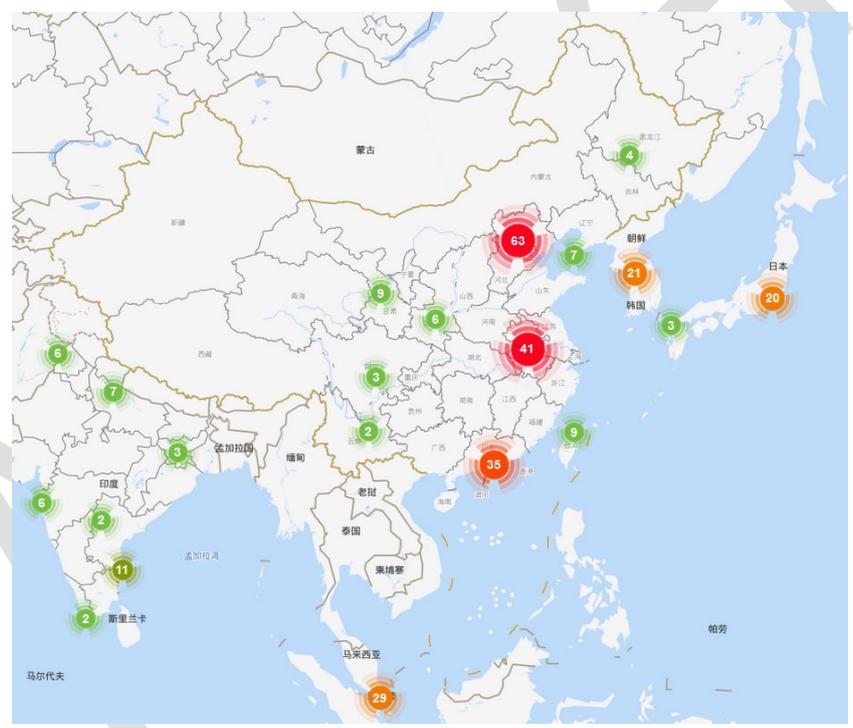


圖 2-30 機器學習領域中國學者分佈

我國專家學者在機器學習領域的分佈如上圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是長三角和珠三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，

通過觀察中國周邊國家的學者數量情況，特別是與日韓、東南亞等亞洲國家相比，中國在機器學習領域學者數量較多。

中國與其他國家在機器學習的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 2-1 機器學習領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	總的學者數
中國-美國	511	26694	52	819
中國-英國	44	1398	32	73
中國-新加坡	36	1189	33	56
中國-澳大利亞	31	744	24	42
中國-印度	22	1123	51	19
中國-德國	17	419	25	39
中國-瑞士	11	233	21	22
中國-荷蘭	6	93	16	10
中國-巴基斯坦	4	82	21	3
中國-以色列	3	23	8	6

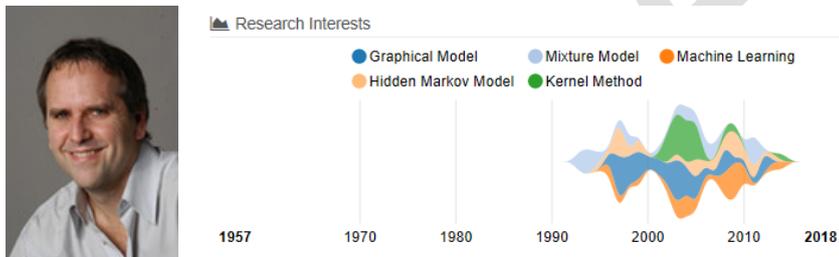
從上表資料可以看出，中美合作的論文數、引用數、平均引用數、學者數遙遙領先，表明中美間在機器學習領域合作之密切；從地域角度看，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 4 席；中國與印度合作的論文數雖然不是最多，但是擁有平均引用數依然位列第二，說明在合作品質上中印合作也達到了較高的水準。

2.6 代表性學者簡介

綜合 h-index 以及領域知名度與活躍度，下面我們將對國內外機器學習領域代表性學者進行簡要介紹，排名不分先後。此外，限於報告篇幅，我們對所有學者不能逐一羅列，如有疏漏，還請與 AMiner 編者聯繫，或者登錄 <https://www.aminer.cn/> 獲取更多資料。

2.6.1 國際頂級學者

- Michael I. Jordan



Michael I. Jordan，美國三院（美國國家科學院、美國國家工程院、美國藝術與科學院）院士，機器學習泰斗，被譽為人工智慧領域的“根目錄”之一，伯克利大學機器學習實驗室 AMP Lab 聯合主任，IEEE Fellow，ACM Fellow。

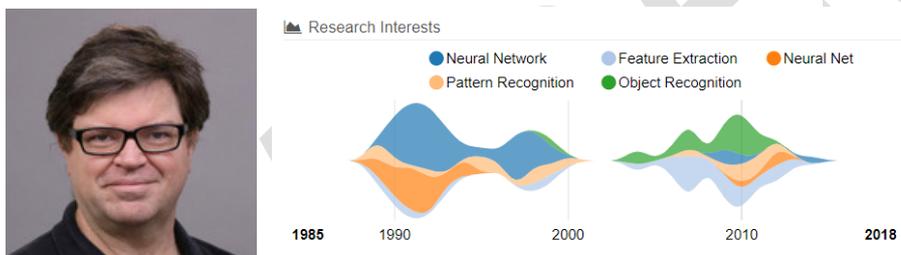
Michael I. Jordan 是美國科學家，加州大學伯克利分校電子工程系、電腦科學和統計系傑出教授，機器學習、統計學和人工智慧研究員。他是機器學習領域的領軍人物之一，並且在 2016 年被 *Semantic Scholar*（科學報）稱為世界上最有影響力的電腦科學家。同年也被 AMiner 評為機器學習最有影響力學者。

他于 1985 年獲得加利福尼亞大學聖地牙哥分校博士學位。自 1988 年至 1998 年，Michael I. Jordan 任麻省理工學院教授，他的研究方向包括了計算學、統計學、認知科學以及生物科學。近年來，他的研究興趣集中在貝葉斯非參數分析、

概率圖模型、譜方法、核方法、分散式運算系統、自然語言處理、信號處理和統計遺傳學等問題的應用上。深度學習領域的權威 Yoshua Bengio，貝葉斯學習領域權威 Zoubin Ghahramani 及前百度首席科學家吳恩達等人都是其門下學生。

他曾獲得眾多獎項，在 2016 年獲得 IJCAI 研究卓越獎 (IJCAI Research Excellence Award)，2015 年獲得了 David E. Rumelhart 獎，並在 2009 年獲得了 ACM / AAAI Allen Newell 獎，2004 年獲得 ICML 最佳學生論文獎。同時，他是 AAAI、ACM、ASA、CSS、IEEE、IMS、ISBA 和 SIAM 成員。

● Yann LeCun



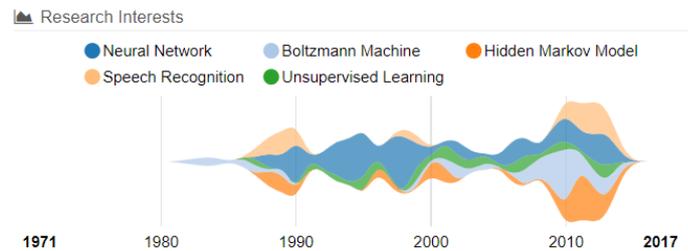
Yann LeCun，人工智慧領域三大奠基人之一，被稱為“卷積網路之父”。

Yann LeCun 是美國工程院院士，Facebook 人工智慧研究院院長，紐約大學 Sliver 教授，同時還兼職於科學資料中心，數學科學交流學院，神經科學中心，以及電子工程電腦系。他以使用卷積神經網路 (CNN) 進行光學字元辨識和電腦視覺方面的工作而聞名，並且是卷積網路的創始人。

他獲得巴黎第六大學 (Pierre et Marie Curie) 大學的電腦科學博士學位，1987 年至 1988 年，是多倫多大學 Geoffrey Hinton 實驗室的博士後研究員。他于 2003 年加入紐約大學，之後還在普林斯頓的 NEC 研究院短暫任職。在 2012 年，他創建了紐約大學資料科學中心，並擔任主任。2013 年底，他被任命為 Facebook 人

工智慧研究總監，並繼續在紐約大學做兼職教授。2015-2016 年，他在巴黎法蘭西工學院做客座教授。

● Geoffrey Hinton



Geoffrey Hinton，人工智慧領域三大奠基人之一，被稱為“神經網路之父”，“深度學習鼻祖”。

Geoffrey Hinton 是英國電腦科學家，擔任多倫多大學電腦科學系教授，多倫多大學向量學院(Vector Institute)首席科學顧問。人工智慧三大奠基人之一 Yann LeCun 以及穀歌大腦研究科學家 Hugo LaRochelle 都是其博士後。

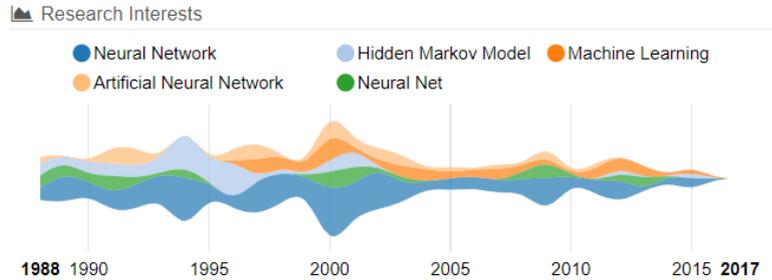
他于 1975 年獲得愛丁堡大學人工智慧方向博士學位，神經網路是他的研究重點。

2013 年，他加入穀歌並帶領 AI 團隊，將神經網路帶入到研究與應用的熱潮，將“深度學習”從邊緣課題變成了穀歌等互聯網公司的依賴的核心技術，並將 Backpropagation (反向傳播) 演算法應用到神經網路與深度學習。

Geoffrey Hinton 獲得諸多獎項。2016 年獲得 NEC C&C Award，IEEE/RSE James Clerk Maxwell Medal；2014 年獲得 IEEE Frank Rosenblatt Medal；2013 年獲得 Doctorat honorifique, University of Sherbrooke；2012 年，獲得了加拿大基廉

獎 (Killam Prizes , 有 “ 加拿大諾貝爾獎 ” 之稱的國家最高科學獎) 。 2011 年獲得蘇賽克斯大學理學博士榮譽學位 ; 2005 年獲得 JICAI 卓越研究獎項。

● Yoshua Bengio



Yoshua Bengio , 加拿大電腦科學家 , 與 Geoffrey Hinton、Yann LeCun 一起 , 被稱為人工智慧三大奠基人。根據 MILA 的資料 , 在 h 指數至少為 100 的電腦科學家中 , Yoshua Bengio 是每天都有被引用的一個。他在人工神經網路和深度學習方面做出了突出貢獻。

Yoshua Bengio 于 1991 年獲得加拿大麥吉爾大學電腦科學博學位元 , 並是麻省理工學院和貝爾實驗室的博士後。他自 1993 年以來擔任蒙特利爾大學教授 , 擔任電腦科學與運籌學系主任。他撰寫了三本書 , 超過 500 種出版物 (h -index 為 125 , 超過 135000 次引用) , 最常被引用在深度學習 , 複現神經網路 , 概率學習演算法 , 自然語言處理和多元學習領域 , 其中 , *Deep Learning* 是他于 GAN 之父 Ian Goodfellow 等人合著的入門深度學習必讀經典教程。

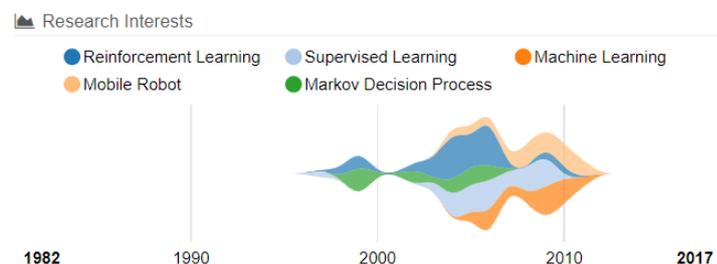
他是加拿大最受歡迎的電腦科學家之一 , 也是 (或曾經是) 機器學習和神經網路中頂尖期刊的副主編。自 2000 年起 , 他在統計學習演算法中擔任加拿大研究主席 , 自 2006 年成為 NSERC 工業主席 , 自 2005 年以來 , 他是加拿大高級研究所高級研究員 , 自 2014 年以來 , 他一直致力於深入學習。他是 NEURIPS 基金

會的董事會成員，也是 NEURIPS 的課程主席和總裁。他共同組織了 14 年的學習研討會，還共同組織了新的國際學習代表會議。他目前的興趣集中於通過機器學習對 AI 的追求，並且包括關於深度學習和表徵學習的基本問題，高維空間中的泛化幾何，多元學習，生物學啟發式學習演算法以及統計機器學習的具有挑戰性的應用。

2016 年 10 月，Yoshua Bengio 聯合創立了 Element AI，這是一家位於蒙特利爾的企業孵化器，致力於將人工智慧 (AI) 研究轉化為實際的商業應用。2017 年 5 月，Bengio 宣佈他將加入蒙特利爾的法律創業公司 Botler AI，擔任戰略顧問。他是 2017 年瑪麗維克多魁北克獎獲得者，加拿大皇家學會會員，他是 CIFAR 高級研究員並共同指導其在機器和大腦學習計畫。此外，他還是 MILA (蒙特利爾大學學習演算法學院) 的創始人們兼科學主任。

Yoshua Bengio 的論文“*A neural probabilistic language model*”開創了神經網路 language model (語言模型) 的先河。該論文的思路影響、啟發了之後的很多基於神經網路做 NLP (自然語言處理) 的文章。

- Andrew Y. Ng (吳恩達)



Andrew Y. Ng 最知名的事情是他所開發的人工神經網路通過觀看一周 YouTube 視頻，自主學會識別哪些是關於貓的視頻。這個案例為人工智慧領域翻開嶄新一頁。

Andrew Y. Ng 於 2002 年獲得了加州大學伯克利分校的博士學位，並從這年開始在斯坦福大學工作。他是前文介紹的 Michael I Jordan 的弟子。他的主要興趣領域在機器學習、深度學習、機器人、人工智慧和電腦視覺等方面。2010 年，時任斯坦福大學教授的 Andrew Y. Ng 加入谷歌開發團隊 XLab——這個團隊已先後為谷歌開發無人駕駛汽車和穀歌眼鏡兩個知名項目，Andrew Y. Ng 加入後開始“穀歌大腦”專案。2014 年 5 月，吳恩達加入百度，擔任百度公司首席科學家，負責百度研究院的領導工作，尤其是 Baidu Brain 計畫。2017 年 10 月，吳恩達出任 Woebot 公司新任董事長，該公司擁有一款同名聊天機器人

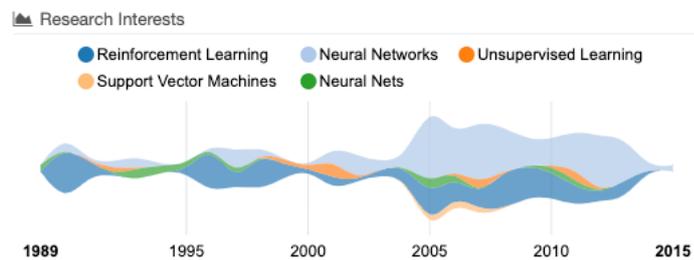
Andrew Y. Ng 最知名的事情是，他所開發的人工神經網路通過觀看一周 YouTube 視頻，自主學會識別哪些是關於貓的視頻。這個案例為人工智慧領域翻開嶄新一頁。

他 2007 年獲得了斯隆獎(Sloan Fellowship),2008 年入選“the MIT Technology Review TR35”，即《麻省理工科技創業》雜誌評選出的科技創新 35 俊傑，以及電腦思維獎(Computers and Thought Award)，並在 2013 年入選《Time》雜誌年度全球最有影響力的 100 人之一，其中共 16 位科技界人物。他也是“電腦和思想獎”的獲得者。

他現在的興趣主要是深度學習。他在 2013 年前共有 128 項學術著作，如 *Deep Learning with COTS HPC Systems* (Adam Coates, Brody Huval, Tao Wang, David

J.Wu, Bryan Catanzaro and Andrew Y. Ng 等人在 ICML 2013 上發表)、*Parsing with Compositional Vector Grammars* 等, 限於篇幅, 本報告不一一列舉。他所著 *Machine Learning Yearning* 於 2018 年出版, 該書面向的用戶群體為機器學習從業者, 主要介紹機器學習實際使用時的一些策略和技巧, 以便為開發指明方向, 提升開發效率。

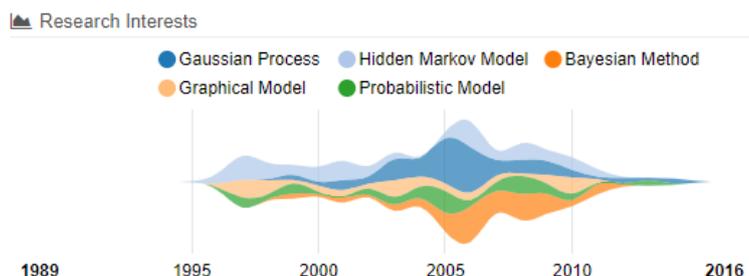
- Jürgen Schmidhuber



Jürgen Schmidhuber 出生於德國, 是瑞士人工智慧實驗室 (IDSIA) 的研發主任, 是 LSTM 的發明人、深度學習元老, 被稱為遞迴神經網路之父。Schmidhuber 本人創立的公司 Nnaisense 正專注於人工智慧技術研發。此前, 他開發的演算法讓人類能夠與電腦對話, 還能讓智慧手機將普通話翻譯成英語。

德國電腦科學家 Jürgen Schmidhuber 在接受英國《衛報》採訪時表示, 宇宙史上重大事件的發生間隔似乎在幾何式地縮短——前後兩個里程碑事件的間隔約為前一個間隔的四分之一。按照這一規律, 人工智慧可能在 2050 年超過人類智商。人工智慧將造就一種新型的生命, 像是生物大爆炸。

- Zoubin Ghahramani



Zoubin Ghahramani，是劍橋大學資訊工程教授，他領導了由大約 30 名研究人員組成的機器學習小組，並領導了 Uber-AI 實驗室的首席科學家。他曾擔任英國國家資料科學研究所阿蘭圖靈研究所(Alan Turing Institute)的創始劍橋主任，勒沃胡姆未來情報中心 (Leverhulme Centre for the Future of Intelligence) 副學術主任，劍橋聖約翰學院 (St John's College Cambridge) 院士。

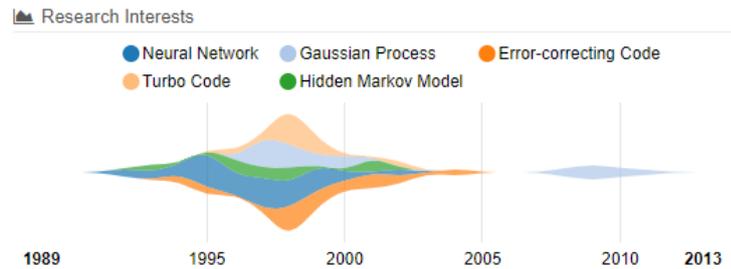
他在賓夕法尼亞大學學習電腦科學和認知科學，1995 年從麻省理工學院獲得博士學位，並在多倫多大學做博士後。他的學術生涯包括同時被任命為倫敦蓋茨比計算神經科學部門的創始成員之一，以及 CMU 機器學習部門的教員超過 10 年。

他目前的研究興趣包括統計機器學習、貝葉斯非參數、可伸縮推理、概率規劃等。他發表了 250 多篇論文，獲得 38000 多條引文 (h 指數 84)。他的工作得到了 EPSRC、DARPA、微軟、穀歌、Infosys、Facebook、亞馬遜、FX Concepts、NTT 和其他一些工業合作夥伴的資助和捐贈。

2013 年，他獲得了 75 萬美元的穀歌獎，用於研究如何建立自動統計師。他曾擔任微軟劍橋研究院(Microsoft Research Cambridge)、VocalIQ(被蘋果收購)、劍橋資本管理公司(Cambridge Capital Management)、Echobox、Informetis、Opera Solutions 和其他幾家公司的顧問。他還擔任過一些領導職務，擔任機器學習領域

主要國際會議的項目和總主席：AISTATS(2005 年)、ICML(2007 年、2011 年) 和 NIPS (2013 年、2014 年)。2015 年，他被選為皇家學會會員，2016 年，他被評為機器學習領域十大最具影響力的學者之一。

- David J.C. MacKay



David J.C. MacKay，曾任劍橋大學卡文迪什實驗室物理系自然哲學教授，現為劍橋大學工程系教授，能源和氣候變化部首席科學顧問。

1967 年 4 月 22 日出生於英國特倫特的斯托克。在紐卡斯爾接受萊姆學校和劍橋三一學院的教育後，他於 1991 年在加州理工學院完成了計算和神經系統博士學位。

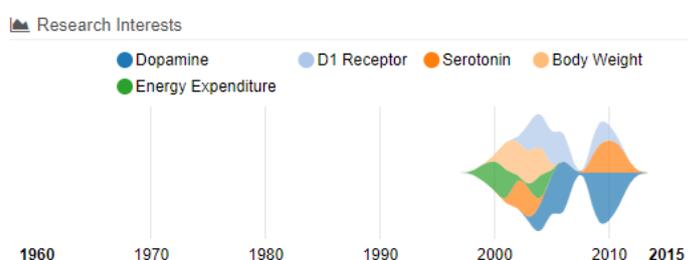
他的興趣包括構建和實現發現資料模式的分層貝葉斯模型，開發神經網路的概率方法，以及改錯碼的設計和解碼。

他在機器學習、資訊理論和通信系統方面的研究在國際上享有盛名，其中包括 Dasher 的發明，Dasher 是一種軟體介面，可以用任何肌肉在任何語言中進行有效的通信。他從 1995 年開始在劍橋教物理。自 2005 年以來，他將越來越多的時間用於能源方面的公共教學。他是世界經濟論壇全球氣候變化議程理事會成員。

1985 年南斯拉夫國際物理奧林匹克運動會：銀牌；一等獎，1999 年通信學會 Leonard G.Abraham 獎論文獎（與 R.J.McEliece 一起以及 J.-F.Cheng），2001

年、1999 年 IBM 合作夥伴獎，2009 年當選物理研究所院士、皇家學會會員，2010 年當選土木工程師學會會員，2013 年獲梅爾切特獎。

● Christopher Bishop



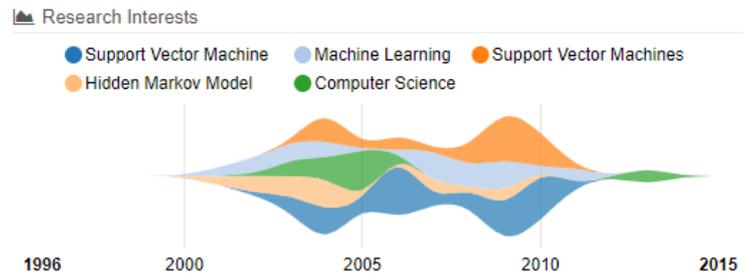
Christopher Bishop，微軟劍橋研究院技術研究員兼實驗室主任，愛丁堡大學電腦科學教授，劍橋達爾文學院院士。

Chris 在牛津大學獲得物理學學士學位，在愛丁堡大學獲得理論物理學博士學位，並發表了一篇關於量子場論的論文。從那時起，他對模式識別產生了興趣，並成為 AEA 技術應用神經計算中心的負責人。隨後，他被選為阿斯頓大學電腦科學和應用數學系的主席，並在那裡成立和領導了神經計算研究小組。

克裡斯是兩本被廣泛引用的機器學習教科書的作者：《神經網路模式識別》（1995）和《模式識別與機器學習》（2006）。他還致力於機器學習在從電腦視覺到醫療保健等領域的廣泛應用。克裡斯是公眾參與科學的積極宣導者，2008 年，他發表了著名的皇家學會聖誕講座，1825 年由邁克爾法拉第創立，並在國家電視臺播出。

他於 2004 年當選皇家工程院院士，2007 年當選愛丁堡皇家學會院士，2017 年當選皇家學會院士。

- Tony Jebara



哥倫比亞大學電腦科學系副教授，哥倫比亞大學機器學習實驗室負責人。

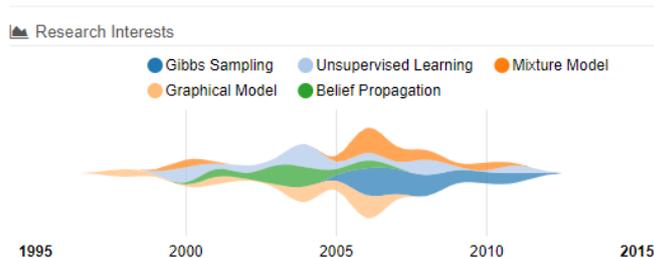
主要研究方向為電腦科學和統計學的交叉融合，在視覺、學習和時空建模等方面成就很高。

他于 2002 獲得麻省理工學院博士學位。他指導哥倫比亞機器學習實驗室 (Columbia Machine Learning Laboratory)，該實驗室的研究與電腦科學和統計學交叉，開發新的資料學習框架，並將其應用於視覺、網路、時空資料和文本。Tony Jebara 已經創立了包括 Sense Networks、Agolo、Ninoh 和 Bookt 在內的幾家初創公司，並為其提供諮詢服務。他在會議、研討會和期刊上發表了 100 多篇同行評議論文，包括 NIPS、ICML、UAI、COLT、JMLR、CVPR、ICCV 和 AISTAT。他是《機器學習：辨別與生成》一書的作者，也是視覺、學習和時空建模領域多項專利的共同發明人。

2004 年，Tony Jebara 獲得了國家科學基金會的職業獎。他的作品在第 26 屆機器學習國際會議上獲得最佳論文獎，在第 20 屆機器學習國際會議上獲得最佳學生論文獎，並在 2001 年獲得模式識別學會的傑出貢獻獎。Jebara 的研究已經在電視上 (ABC, BBC, New York One, TechTV 等) 和大眾媒體 (紐約時報, 斜線點, 有線電視, 商業週刊, IEEE 頻譜等) 上出現。《紳士》雜誌將他評為 2008 年最優秀、最聰明的人物之一。Jebara 還是《機器學習研究》雜誌和《機器

學習》編輯委員會的副主編。2007 年至 2011 年，Jebara 任機器學習副主編，2010 年至 2012 年任 IEEE 模式分析和機器智慧事務副主編。2006 年，他與人共同創立了 NYAS 機器學習研討會，並從那時起一直擔任該研討會的指導委員會成員。Tony Jebara 還擔任了 2014 年第 31 屆機器學習國際會議 (ICML) 的項目主席。

● Max Welling



Max Welling，阿姆斯特丹大學的“研究主席”，加州大學歐文分校 (UCI) 電腦科學與統計學教授，加拿大高級研究所 (CIFAR) 副研究員，Scyfer BV 聯合創始人。

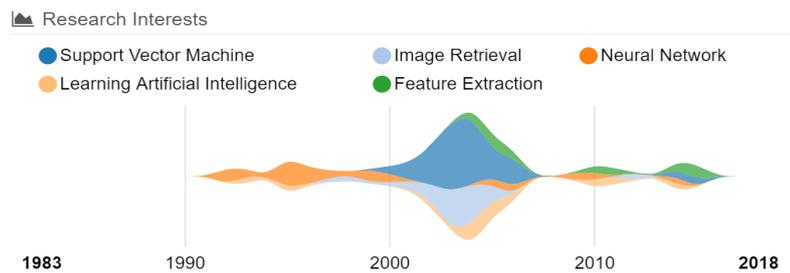
在過去，他曾在加州理工學院 (Caltech) (1998-2000)、加州大學洛杉磯分校 (UCL) (2000-2001) 和多倫多大學 (U.Toronto) (2001-2003) 擔任博士後。1998 年，他在諾貝爾獎獲得者霍夫特教授的指導下獲得了博士學位。

Max Welling 從 2011 年至 2015 年擔任 IEEE TPAMI 的副主編。他自 2015 以來擔任 NIPS 基金會的董事會成員 (在機器學習方面規模最大的會議)，分別擔任 2013 和 2014 年度 NIPS 的計畫主席和總主席。2009 年，他還是 AISTATS 和 2016 年 ECCV 的專案主席，2018 年 MIDL 的總主席。他曾在 JMLR 和 JML 的編輯委員會任職，並擔任神經計算、JCGS 和 TPAMI 的副主編。他從穀歌、Facebook、雅虎、NSF、NIH、NWO 和 ONR-MURI 獲得了多項資助，其中一項是 2005 年的 NSF 職業資助。他是 2010 年 ECCV Koenderink 獎的獲得者。Welling

是阿姆斯特丹資料科學研究中心的董事會成員，他領導阿姆斯特丹機器學習實驗室 (AMLAB)，並共同領導高通公司的 UvA 深度學習實驗室 (QUVA) 和博世公司的 UvA 深度學習實驗室 (DELTA)。

2.6.2 國內知名學者

● 張鉞



張鉞，中國科學院院士，清華大學電腦科學與技術系教授，清華大學人工智慧研究院院長。

張鉞于 1958 年畢業於清華大學自動控制系，是國家第一批自動控制專業的畢業生。1995 年他當選為中國科學院院士。

他早期從事自動控制理論與系統研究，1979 年開始電腦科學與技術研究。從事人工智慧理論、人工神經網路、遺傳演算法、分形和小波等理論研究；以及把上述理論應用於模式識別、知識工程、智慧型機器人與智慧控制等領域的應用技術研究。

他針對人工智慧問題求解計算複雜性、指數爆炸的主要困難，提出了問題分層求解的商空間理論，解決了不同細微性空間的描述、它們之間相互轉換、複雜性分析等理論問題。在此基礎上提出統計啟發式搜索演算法，基於拓撲的空間規

劃方法和關係矩陣的規劃演算法，對克服計算量的指數爆炸很有成效。還提出了研究不確定性處理、定性推理、模糊分析、證據合成等新原理。指導並參加建成了陸地自主車、圖像與視頻檢索等實驗平臺。

張鈹和同期同事成了國內最早接觸到人工智慧的研究者，並成為我國在這方面的首批專家。

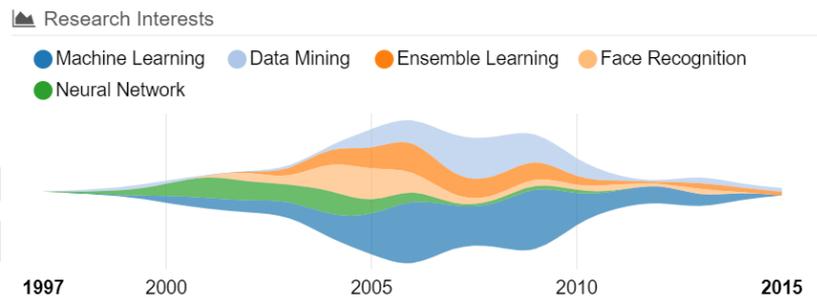
在學術研究上的主要貢獻是提出問題分層求解的商空間理論，通過代數的方法，系統地解決了不同層次求解空間的問題表達、複雜性分析、不同層次空間之間資訊、運算元及推理機制等的相互轉換關係。在上述理論基礎上，他進一步提出了統計啟發式搜索演算法，基於拓撲的空間規劃方法以及基於關係矩陣的時間規劃演算法等，極大降低了計算複雜性，具有重要的應用價值。其專著《問題求解理論及應用》全面總結了他在人工智慧理論研究中的成果，其英文版於 1992 年由 Elsevier Science Publishers B.V.(Nortn-Holland)出版，中文版獲國家教委頒發的高校出版社優秀學術專著特等獎。澳大利亞專家 Ronald Walts 在電腦雜誌 *The Australian Computer Journal* (1995) 對《問題求解理論及應用》(英文版) 的評論為“這是一部在重要研究領域的優秀著作”。美國學者 Harold S.Stone 認為，張鈹等在統計啟發式搜索等方面的工作，是“最近幾年中國學者作出的很有意義的貢獻”，“將新一代計算技術的前沿向前推進了”。

他在國內外共發表論文 100 多篇，中英文專著有《問題求解理論及應用》(中英版) 以及《人工神經網路理論及應用》等。

他於 1994 年當選為俄羅斯自然科學院外籍院士；1995 年當選為中國科學院院士；2011 年德國漢堡大學授予自然科學名譽博士；2015 年 1 月 31 日，張鈞獲得 2014 CCF 終身成就獎。

他的社會任職有：智慧技術與系統國家重點實驗室主任、校學位委員會副主任、資訊科學與技術學院學術委員會主任；中國自動化學會機器人專業委員會副主任及智慧控制專業委員會副主任；《電腦學報》副主編；國家高技術“863”計畫智慧型機器人主題專家組成員；河南科技大學兼職院士；電腦學學術委員會主任。

- 周志華



周志華，南京大學教授，博士生導師；教育部長江學者特聘教授，國家傑出青年基金獲得者；南京大學電腦科學與技術系副主任、軟體新技術國家重點實驗室常務副主任，機器學習與資料採擷研究所（LAMDA）所長，校學術委員會委員、南京大學人工智慧學院院長（兼）。

周志華於 2000 年獲得南京大學電腦科學與技術系博士學位，2001 年 1 月起留校任教，2002 年 3 月被破格聘任為副教授，2003 年，在他 29 歲時獲得國家傑出青年科學基金，隨後被聘為教授。

他于 2006 年入選教育部長江學者特聘教授，2012 年當選 IEEE Fellow 和 IAPR Fellow (國際模式識別學會會士)，2013 年當選 ACM Distinguished Scientist (ACM 傑出科學家)和中國電腦學會(CCF)會士，成為大陸高校首位當選 ACM 傑出科學家的學者。2007 年創建南京大學機器學習與資料採擷研究所(LAMDA)，2010 年 11 月任軟體新技術國家重點實驗室常務副主任，2013 年 5 月任電腦系副主任。

2016 年，他當選 AAAI Fellow(國際人工智慧學會)，成為我國大陸第一位，也是此次入選的唯一來自美歐之外的學者，並且是唯一在中國大陸取得博士學位的 AAAI Fellow。2016 年 11 月，當選美國科學促進會會士 (AAAS Fellow)。2016 年 12 月，當選 ACM Fellow，成為第一位在中國大陸取得全部學位的 ACM Fellow。

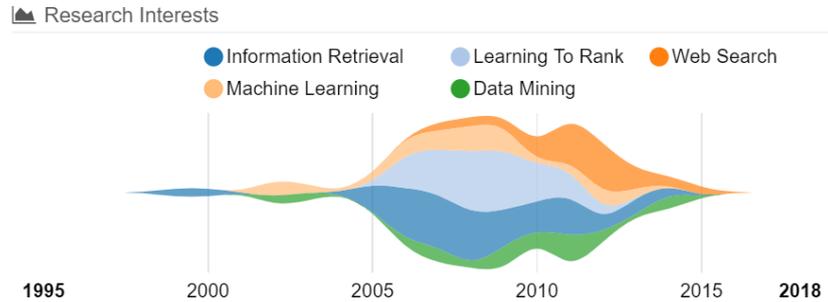
2017 年 2 月，當選人工智慧領域頂級學術會議 AAAI 2019 程式委員會主席，是該會議自 1980 年成立以來首位華人主席、也是首次由美歐之外國家的學者出任主席。

兼任 AAAI Fellow, IEEE Fellow, IAPR Fellow, ACM Fellow 和 AAAS Fellow，周志華成為國際上與人工智慧相關的重要學會“大滿貫”Fellow 華人第一人。

此外，他還擔任 IJCAI 程式委員會主席，是中國內地首位任此職位學者。

周志華主要從事人工智慧、機器學習、資料採擷等領域的研究工作。他著有機器學習入門書籍《機器學習》。

- 李航



李航，北京大學、南京大學兼職教授。

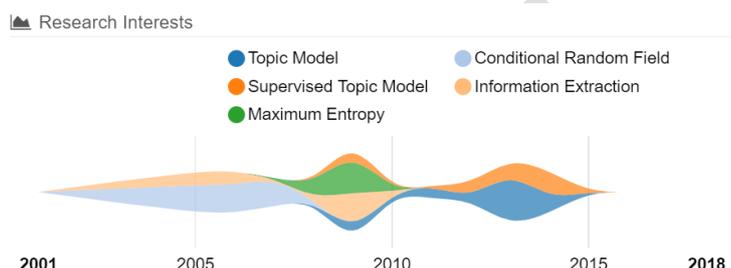
李航畢業于日本京都大學電氣電子工程系，於 1998 年獲得日本東京大學電腦科學博士學位。曾任日本 NEC 公司中央研究所研究員，微軟亞洲研究院高級研究員與主任研究員、華為技術有限公司諾亞方舟實驗室主任。現任今日頭條人工智慧實驗室主任。

他的研究方向包括資訊檢索、自然語言處理、統計機器學習及資料採擷。他一直活躍在相關學術領域，曾出版過三部學術專著，並在頂級國際學術會議和國際學術期刊上發表過上百篇學術論文，擁有 42 項授權美國專利。

他出版的三本技術書籍其中最廣為人知的是 2012 年出版的《統計學習方法》，他發表超過 120 項技術論文，包括 SIGIR、WWW、WSDM、ACL、EMNLP、ICML、NeurIPS、SIGKDD、AAAI、IJCAI 等頂級國際會議以及包括 CL、NLE、JMLR、TOIS、IRJ、IPM、TKDE、TWEB、TIST。他和他同事的論文收到了 SIGKDD'08 最佳應用論文獎，SIGIR'08 最佳學生論文獎，ACL'12 最佳學生論文獎。

他是 ACM 傑出科學家。他的社會任職包括 AIRS-2008 程式委員會主席，SIGIR-2008 Poster & Demo 委員會主席，KDD-2009 宣傳主席，EMNLP-2009 領域主席，ACM Transaction on Asian Language Information Processing 副主編，Journal of Computer Science and Technology 編委等。

● 朱軍



朱軍，清華大學電腦科學系教授，智慧技術與系統國家重點實驗室副主任，卡內基梅隆大學兼職教授。2013 年，入選 IEEE Intelligent Systems 的“人工智慧 10 大新星” (AI's 10 to Watch)。

朱軍於 2009 年獲得清華大學電腦博士學位，主要從事機器學習研究。

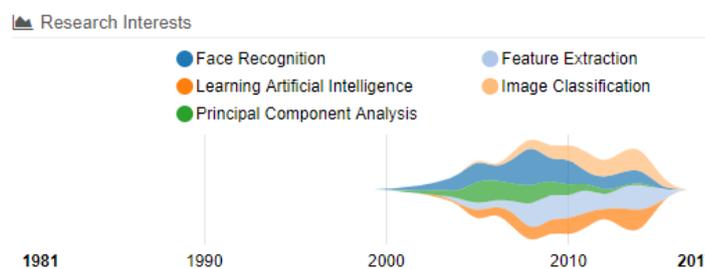
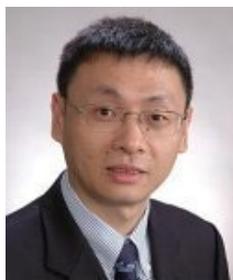
他在國際重要期刊與會議發表學術論文 80 餘篇，他開源了 ZhuSuan，一個用於貝葉斯深度學習（貝葉斯方法和深度學習的結合）的 GPU 庫，可以在 TensorFlow 上使用

他是 AAAI 2019，NeurIPS 2018，ICML 2018，UAI 2018，IJCAI 2018 的區域主席，擔任國際期刊 IEEE TPAMI 和 Artificial Intelligence 的編委、國際會議 ICML 2014 地區聯合主席、以及 ICML、NEURIPS 等國際會議的領域主席。

他於 2006 年被評為微軟學者，2009 年入選卡內基梅隆大學 Innovation Fellow；中國電腦學會優秀博士論文獎獲得者（2009）；清華大學 221 基礎研究計畫入選

者 (2012) ; 中國電腦學會青年科學家 (2013) ; IEEE Intelligent Systems 雜誌評選的“AI’s 10 to Watch” (2013) ; 國家優秀青年科學基金獲得者 (2013) , 同年獲得、中國電腦聯合會 (CCF) 頒發的“CCF 青年科學家”獎 ; 2014 年 , 他獲得清華-MSRA 聯合研究實驗室頒發的最佳協作獎 ; 2015 年獲得全國青年頂尖人才支持計畫的支持 , 同年收到了“CVIC SE 人才”獎 ; 2017 年 , 他被麻省理工學院 TR35 中國選為“先驅者”之一。

- 顏水成



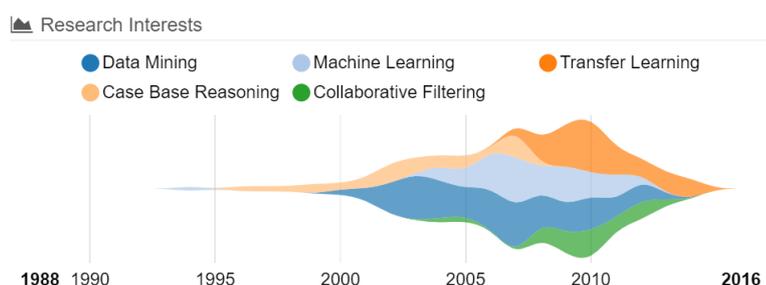
顏水成，新加坡國立大學副教授、360 集團副總裁、人工智慧研究院院長、第十三批國家“千人計畫”專家。

主要研究興趣是為電腦視覺、多媒體和資訊檢索應用開發機器學習理論。

他在眾多研究課題上撰寫/合著了數百篇技術論文，其中谷歌學者引文 2 萬餘次。他是 2014 年、2015 年和 2016 年 ISI 被高度引用的研究員。顏水成博士率領的團隊共獲得了 10 次電腦視覺領域兩大核心競賽 Pascal VOC 和 ImageNet 大規模視覺識別 (ILSVRC) 冠軍和榮譽獎，10 餘次最佳 (學生) 論文獎。他的團隊還曾獲得多媒體領域頂會 ACM MM 最佳論文獎、最佳學生論文獎和最佳技術演示獎的大滿貫。

顏水成博士團隊提出的“Network in Network” (NIN) 網路結構的核心 1x1 卷積是近年來幾乎所有電腦視覺深度學習模型的標準模組，在學術界和工業界影響深遠，其思想也被後期的 GoogleNet、殘差網路 (ResNet) 等模型所採用。

● 楊強



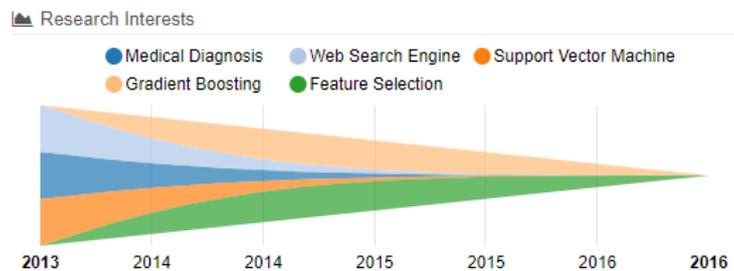
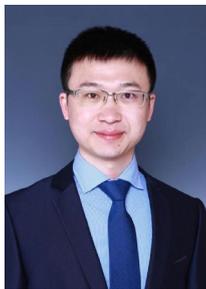
楊強，香港科技大學新明工程學講席教授，電腦科學和工程學系主任，大資料研究所所長。IEEE Fellow，IAPR Fellow，AAAS Fellow，ACM 傑出科學家，KDD 中國主席。

楊強于 1989 年獲得馬里蘭大學電腦科學博士學位，之後直到 1995 年，于加拿大滑鐵盧大學電腦系任助理教授及副教授。其主要研究領域為機器學習、資料採擷和自動規劃。他是人工智慧研究的國際專家和領軍人物，在學術界和工業界做出了傑出的服務和貢獻，尤其近些年為中國人工智慧(AI)和資料採擷(KDD) 的發展起了重要引導和推動作用。迄今為止，楊強已發表逾 400 篇關於人工智慧和資料採擷方面的論文，引用超過 20000 次。

2009 年，他創建了 ACM 刊物 *Transactions on Intelligent Systems and Technology (TIST)* 並任首屆主編。2012 年至 2015 年，出任華為諾亞方舟實驗室創始主任；2015 年，任香港科技大學電腦與工程學系主任；2016 年，在香港科技大學創建大資料研究所。

他的社會兼職有：2013 年 7 月當選為國際人工智慧協會 (AAAI) 院士，是第一位獲此殊榮的華人，之後又於 2016 年 5 月當選為 AAAI 執行委員會委員，是首位也是至今為止唯一的 AAAI 華人執委，同年，任 ACM 資料採擷中國分會 (KDD China) 主席。2017 年 8 月他當選為國際人工智慧聯合會 (IJCAI，國際人工智慧領域創立最早的頂級國際會議) 理事會主席，是第一位擔任 IJCAI 理事會主席的華人科學家。

- 黃高



黃高，清華大學助理教授，博士生導師。

主要研究領域為深度神經網路的結構設計與優化演算法，以及深度學習在電腦視覺中的應用。

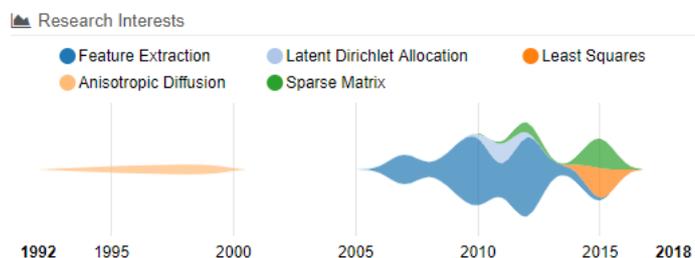
2009 年本科畢業於北京航空航太大學，2015 年獲得清華大學控制科學與工程博士學位，2015 年至 2018 年為美國康奈爾大學電腦系博士後。其博士論文獲選中國自動化學會優秀博士學位論文以及清華大學優秀博士論文一等獎。該獲獎論文的主要貢獻是提出了一種全新的卷積神經網路架構“密集連結卷積網路”(DenseNet)，顯著地提升了模型在圖片識別任務上的準確率。

目前在 NIPS，ICML，CVPR 等國際頂級會議及 IEEE 多個彙刊共計發表學術論文 30 餘篇。2016 年曾獲得全國百篇最具國際影響學術論文、2017 年國際電

腦視覺頂級會議 CVPR 最佳論文獎、2018 年世界人工智慧創新大賽 SAIL 先鋒獎和吳文俊人工智慧自然科學一等獎等獎勵和榮譽。

他是 AAAI2018 高級程式委員，擔任 NeurIPS、ICML、CVPR、ICCV、ECCV、ICLR、AAAI 等國際學術會議和 JMLR、TPAMI、TIP、TNNLS 等國際期刊審稿人。

● 林宙辰



林宙辰，北京大學資訊科學技術學院機器感知與智慧教育部重點實驗室教授。

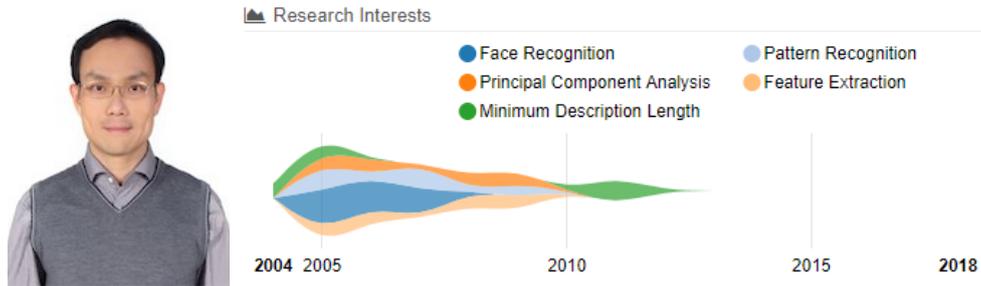
主要研究領域為機器學習、模式識別、電腦視覺、影像處理、數值優化。

1993 年于南開大學獲理學學士學位，1996 年于北京大學獲理學碩士學位，1998 年於香港理工大學獲得哲學碩士學位，2000 年于北京大學獲得理學博士學位。

2007 年榮獲 Microsoft SPOT Award，2015 年 ImageNet 大規模視覺識別競賽 (ILSVRC) 場景分類專案冠軍，2016 年獲國家自然科學傑出青年基金資助。

他是 CVPR 2014/2016、ICCV 2015、NIPS 2015 的領域主席和 AAAI 2016/2017、IJCAI 2016 的高級程式委員。他也是 IEEE Transactions on Pattern Analysis And Machine Intelligence 和 International Journal of Computer Vision 的編委。

- 王立威



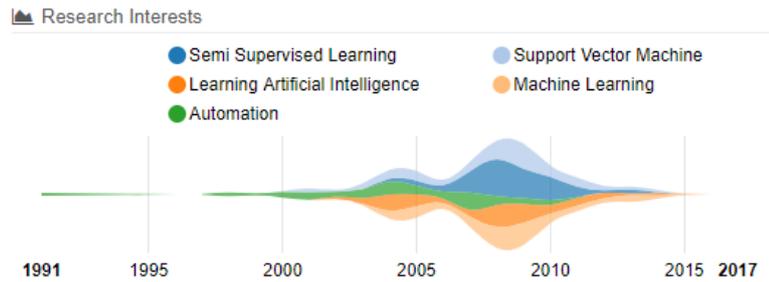
王立威，1999 年獲清華大學學士學位，2002 年獲清華大學碩士學位，2005 年獲北京大學博士學位。現為北京大學資訊科學技術學院教授。

長期從事機器學習相關研究，目前主要致力於機器學習基礎理論，即泛化理論的研究，差分隱私演算法的設計與分析以及醫療影像診斷演算法與系統的開發。

自 2002 年以來，在 PAMI、CVPR、ICML 等國際頂級期刊和會議上發表論文 60 餘篇，並參與編寫《機器學習及其應用》2009 版“關於 Boosting 演算法的 Margin 解釋”及 2015 版“差分隱私保護的機器學習”相關章節。

曾獲得第 11 屆 Meeting on Image Recognition and Understanding 會議最佳論文獎，2010 年獲得 *Pattern Recognition Letters* 期刊最高引用論文獎(2005-2010)，2010 年入選 AI's 10 to Watch，是首位獲得該獎項的亞洲學者。2012 年獲得首屆國家自然科學基金優秀青年基金，新世紀優秀人才。任 NIPS 等權威會議 Area Chair，以及多家學術期刊編委。

- 張長水



張長水，男，1965 年生，河北人。智慧技術與系統國家重點實驗室學術委員會委員，清華大學自動化系教授、博士生導師，智慧技術與系統國家重點實驗室副主任，自動化系主任。

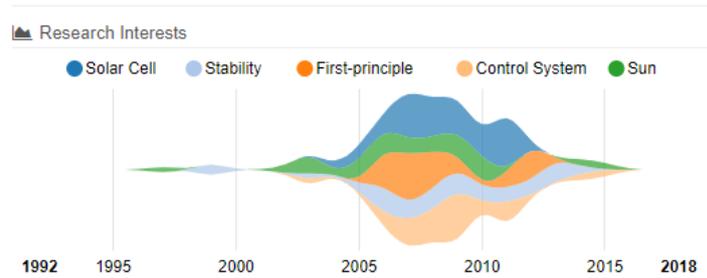
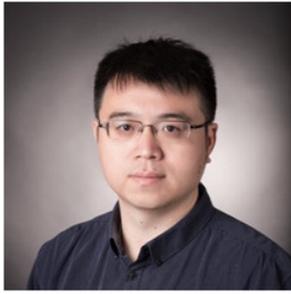
主要從事影像處理、信號處理、模式識別與人工智慧、進化計算等研究領域以及和工業界的合作。

1986 年 7 月畢業於北京大學數學系，獲得理學學士學位。1992 年 7 月畢業於清華大學自動化系，獲得博士學位。

1992 年 7 月—1994 年 12 月，在清華大學自動化系任講師；1995 年 1 月—2000 年 8 月，在清華大學自動化系任副教授；2000 年 9 月起，在清華大學自動化系任教授；2001 年起，任清華大學博士生導師。

近幾年在國際期刊和會議上發表學術論文超過 100 篇，其中包括國際權威期刊 Pattern Recognition、TNN、TKDE、IEEE Transaction on Multimedia 以及國際頂級會議 IJCAI、AAAI、NIPS、ICML、ECML、SIGIR、CVPR 等。他還是國際權威期刊 Pattern Recognition 編委。

- 孫劍



孫劍，男，前微軟亞研院首席研究員，現就職於北京曠視科技有限公司，任曠視首席科學家、曠視研究院院長。

其主要研究方向是電腦視覺和深度學習。

1993 年—1997 年就讀於西安交通大學自動控制專業，獲工學學士學位。2000 年、2003 年在西安交通大學模式識別與智慧控制專業研究生畢業，分別獲得工學碩士學位和工學博士學位。2003 年孫劍加入微軟亞洲研究院。

孫劍自 2002 年以來在 CVPR、ICCV、SIGGRAPH、PAMI 等頂級學術會議和期刊上發表學術論文 100 余篇，孫劍博士擁有超過 40 項美國或國際專利。

2009 年孫劍帶領團隊發表的論文 *Single Image Haze Removal Using Dark Channel Prior* 贏得了國際電腦視覺與模式識別會議 (CVPR) 的最佳論文獎 (CVPR Best Paper)，這是亞洲人第一次獲得該獎；2010 年，孫劍被美國科技評論期刊《麻省理工科技評論》(MIT Technology Review) 評選為“全球 35 歲以下傑出青年創新者”。2012 年至 2014 年，孫劍加入法國國家資訊與自動化研究院(INRIA)/巴黎高等師範學院 Willow 組。2016 年，孫劍帶領的團隊憑藉 *Deep Residual Learning for Image Recognition* 再次獲得了國際電腦視覺與模式識別會議 (CVPR) 的最佳論文獎 (CVPR Best Paper)。2016 年 7 月，孫劍正式加入曠視任首席科學家、曠視研究院院長。2017 年 8 月，孫劍擔任中國自動化學會 (Chinese

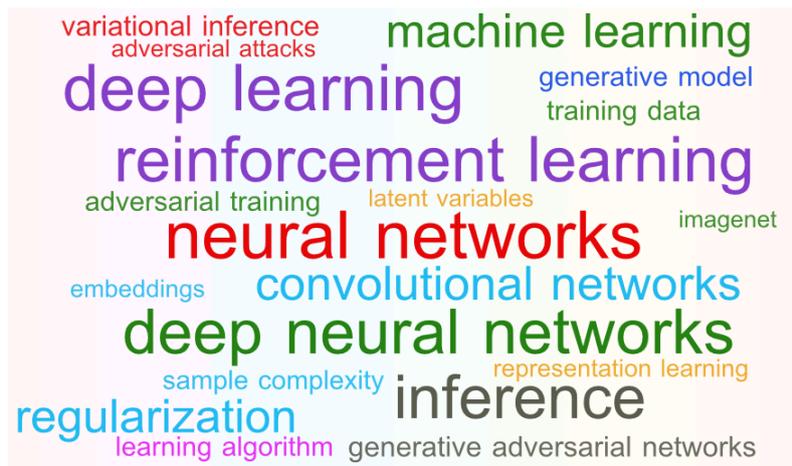
Association of Automation, CAA) 混合智慧專委會副主任。2018 年 5 月，2018 年第一批國家重點研發計畫公示孫劍博士擔任變革性技術關鍵科學問題專項項目負責人。2019 年 1 月，孫劍出任西安交通大學人工智慧學院首任院長。

2.7 論文解讀

本節對本領域的高水準學術會議論文進行挖掘，解讀這些會議在近年的部分代表性工作，會議具體包括：

International Conference on Machine Learning

Conference and Workshop on Neural Information Processing Systems



我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如上圖所示。其中，出神經網路 (neural networks)、深度學習 (deep learning)、強化學習 (reinforcement learning) 是本領域中最熱的關鍵字。ICML 和 NeurIPS 是機器學習領域非常具有代表性的會議，限於報告篇幅，我們選取 ICML 和 NeurIPS 近十年若干最佳論文進行解讀。

表 2-2 ICML 近 10 年 best paper

ICML (International Conference on Machine Learning)		
年份	論文標題	作者
2019	Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations	Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, Olivier Bachem
	Rates of Convergence for Sparse Variational Gaussian Process Regression	David R. Burt, Carl E. Rasmussen, Mark van der Wilk
2018	Delayed Impact of Fair Machine Learning	Lydia T. Liu, University of California Berkeley; et al.
	Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples	Anish Athalye, Massachusetts Institute of Technology; et al.
2017	Understanding Black-box Predictions via Influence Functions	Pang Wei Koh & Percy Liang, Stanford University
2016	Ensuring Rapid Mixing and Low Bias for Asynchronous Gibbs Sampling	Christopher De Sa, Stanford University; et al.
	Pixel Recurrent Neural Networks	Aaron Van den Oord, Google; et al.
	Dueling Network Architectures for Deep Reinforcement Learning	Ziyu Wang, Google; et al.
2015	A Nearly-Linear Time Framework for Graph-Structured Sparsity	Chinmay Hegde, Massachusetts Institute of Technology; et al.
	Optimal and Adaptive Algorithms for Online Boosting	Alina Beygelzimer, Yahoo! Research; et al.
2014	Understanding the Limiting Factors of Topic Modeling via Posterior Contraction Analysis	Jian Tang, Peking University; et al.
2013	Vanishing Component Analysis	Roi Livni, The Hebrew University of Jerusalem; et al.
	Fast Semidifferential-based Submodular Function Optimization	Rishabh Iyer, University of Washington; et al.
2012	Bayesian Posterior Sampling via Stochastic Gradient Fisher Scoring	Sungjin Ahn, University of California Irvine; et al.
2011	Computational Rationalization: The Inverse Equilibrium Problem	Kevin Waugh, Carnegie Mellon University; et al.
2010	Hilbert Space Embeddings of Hidden Markov Models	Le Song, Carnegie Mellon University; et al.
2009	Structure preserving embedding	Blake Shaw, Tony Jebara, Columbia University

表 2-3 NeurIPS 近 10 年 best paper

NeurIPS (Conference and Workshop on Neural Information Processing Systems)		
年份	論文標題	作者
2018	Non-delusional Q-learning and Value-iteration	Tyler Lu, Dale Schuurmans, Craig Boutilier
	Optimal Algorithms for Non-Smooth Distributed Optimization in Networks	Kevin Scaman, Francis Bach, Sebastien Bubeck, Laurent Massoulié, Yin Tat Lee
	Nearly Tight Sample Complexity Bounds for Learning Mixtures of Gaussians via Sample Compression Schemes	Hassan Ashtiani, Shai Ben-David, Ick Harvey , Christopher Liaw, Abbas Mehrabian, Yaniv Plan
	Neural Ordinary Differential Equations	Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt , David Duvenaud

NeurIPS (Conference and Workshop on Neural Information Processing Systems)		
2017	Safe and Nested Subgame Solving for Imperfect-Information Games	Noam Brown, Tuomas Sandholm
	Variance-based Regularization with Convex Objectives	Hongseok Namkoong, John Duchi
	A Linear-Time Kernel Goodness-of-Fit Test	Wittawat Jitkrittum, Wenkai Xu, Zoltan Szabo , Kenji Fukumizu, Arthur Gretton
2016	Value Iteration Networks	Aviv Tamar, Yi Wu, Garrett Thomas, Sergey Levine, Pieter Abbeel
	Matrix Completion has No Spurious Local Minimum	Rong Ge, Jason Lee, Tengyu Ma
	Interactive musical improvisation with Magenta	Adam Roberts, Jesse Engel, Curtis Hawthorne, Ian Simon, Elliot Waite, Sageev Oore, Natasha Jaques, Cinjon Resnick, Douglas Eck
2015	Competitive Distribution Estimation: Why is Good-Turing Good	Alon Orlitsky, Ananda Theertha Suresh
	Fast Convergence of Regularized Learning in Games	Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo , Robert Schapire
2014	Asymmetric LSH (ALSH) for sublinear time Maximum Inner Product Search (MIPS)	Anshumali Shrivastava, Ping Li
	A* Sampling	Chris J. Maddison, Daniel Tarlow, Tom Minka
2013	A Memory Frontier for Complex Synapses	Subhaneil Lahiri, Surya Ganguli
	Submodular Optimization with Submodular Cover and Submodular Knapsack Constraints	Rishabh Iyer, Jeff Bilmes
	Scalable Influence Estimation in Continuous-Time Diffusion Networks	Nan Du, Le Song, Manuel Gomez-Rodriguez , Hongyuan Zha
2012	No voodoo here! Learning discrete graphical models via inverse covariance estimation	Po-Ling Loh, Martin Wainwright
	Discriminative Learning of Sum-Product Networks	Robert Gens, Pedro Domingos
2011	Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials	Philipp Krähenbühl, Vladlen Koltun
	Priors Over Recurrent Continuous Time Processes	Ardavan Saeedi, Alexandre Bouchard-Côte
	Fast and Accurate K-means for Large Datasets	Michael Shindler, Alex Wong, Adam Meyerson
2010	Construction of dependent dirichlet Processes based on Poisson Processes	Dahua Lin, Eric Grimson, John Fisher
	A Theory of Multiclass Boosting	Indraneel Mukherje , Robert E Schapire
2009	An LP View of the M-Best MAP Problem	Menachem Fromer , Amir Globerson
	Fast Subtree Kernels on Graphs	Nino Shervashidze , Karsten Borgwardt

2.7.1 ICML 歷年最佳論文解讀

- 2019 年最佳論文

論文題目：*Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations*

中文題目：挑戰無監督分離式表徵的常見假設

論文作者：Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, Olivier Bachem

論文地址：<https://aminer.cn/pub/5c04967517c44a2c74709162/challenging-common-assumptions-in-the-unsupervised-learning-of-disentangled-representations>

論文解讀：文章主要從理論和實踐兩方面對這一領域中的一些基本假設提出了挑戰。文章從理論上證明，如果沒有對所考慮的學習方法和資料集產生歸納偏置，那麼解耦表示的無監督學習基本上是不可能的。文章還採用了完善的無監督解耦學習實驗方案，進行了一個超級大規模的實驗研究。最後還發佈了 `disentanglement_lib`，這是一個用於訓練和評估解耦表示的新庫。由於複製這個結果需要大量的計算工作論文還發佈了超過 10000 個預訓練的模型，可以作為未來研究的基線方法。

論文題目：*Rates of Convergence for Sparse Variational Gaussian Process Regression*

中文題目：稀疏變分高斯過程回歸的收斂速度

論文作者：David R. Burt, Carl E. Rasmussen, Mark van der Wilk

論文地址：<https://www.aminer.cn/pub/5ced106da562983788e64b9/rates-of->

convergence-for-sparse-variational-gaussian-process-regression

論文解讀：這篇文章來自英國劍橋大學。自從許多研究人提出了對高斯過程後驗的變分近似法後，避免了資料集大小為 N 時 $O(N^3)$ 的縮放。它們將計算成本降低到 $O(NM^2)$ ，其中 $M \leq N$ 是誘導變數的數量。雖然 N 的計算成本似乎是線性的，但演算法的真正複雜性取決於 M 如何增加以確保一定的近似品質。論文證明了稀疏 GP 回歸變分近似到後驗變分近似的 KL 散度的界限，該界限僅依賴於先驗核的協方差運算元的特徵值的衰減。這些邊界證明了直觀的結果，平滑的核、訓練資料集中在一個小區域，允許高品質、非常稀疏的近似。這些邊界證明了用 $M \leq N$ 進行真正稀疏的非參數推理仍然可以提供可靠的邊際似然估計和點後驗估計。對非共軛概率模型的擴展，是未來研究的一個有前景的方向。

- 2018 年最佳論文

論文題目：*Delayed Impact of Fair Machine Learning*

中文題目：公正機器學習的滯後影響

論文作者：Lydia T.Liu, Sarah Dean, Esther Rolf, Max Simchowitz, Moritz Hardt

論文地址：<https://arxiv.org/abs/1803.04383>

論文解讀：機器學習的公平性主要在靜態分類設置中得到研究，但卻沒有關注這些決策如何隨時間改變潛在的群體。傳統的觀點認為公平性准能提升他們想保護的群體的長期利益。本文研究了靜態公平性標準如何與暫時的利益指標相互作用，例如利益變數的長期提升、停滯和下降。本文證實了即使在一步回饋模型中，常見的公平性準則沒有隨時間帶來改善，實際上可能給特定案例帶來了傷害。

論文題目：*Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples*

中文題目：混淆梯度的虛假安全感：對抗樣本防禦

論文作者：Anish Athalye, Nicholas Carlini, David Wagner

論文地址：<https://arxiv.org/abs/1802.00420v1>

論文解讀：如果在一張圖片添加干擾，可能就可以騙過分類器。為了抵禦對抗樣本的攻擊，使得神經網路在受到反覆運算攻擊時不受對抗樣本干擾，研究人員在尋找強大的對抗樣本防禦器，使其在面對基於優化的攻擊之下，可以實現對對抗樣本的魯棒性防禦。

- 2017 年最佳論文

論文題目：*Understanding Black-box Predictions via Influence Functions*

中文題目：利用影響函數理解黑箱預測

論文作者：Pang Wei Koh, Percy Liang

論文地址：<https://arxiv.org/abs/1703.04730>

論文解讀：這篇論文利用影響函數（穩健統計學中的經典技術），通過學習演算法跟蹤模型的預測並追溯到訓練資料，從而確定對給定預測影響最大訓練點來解釋黑箱模型的預測。為了將影響函數擴展到現代機器學習中，論文中設計了一個簡單高效的實驗，僅需梯度 oracle 訪問和 Hessian 向量積。而且即使在非凸和非微分模型上，影響函數的近似值演算法仍然可以提供有價值的資訊。在線性

模型和卷積神經網路中，論文中也證明，影響函數可用于理解模型行為，調試模型，檢測資料集錯誤，甚至是生成視覺上無法區分的訓練集攻擊。

- 2016 年最佳論文

論文題目：*Ensuring Rapid Mixing and Low Bias for Asynchronous Gibbs Sampling*

中文題目：確保非同步吉布斯採樣的快速混合和低偏差

論文作者：Christopher De Sa, Kunle Olukotun, Christopher Ré

論文地址：<https://arxiv.org/abs/1602.07415>

論文解讀：吉布斯採樣 (Gibbs Sampling) 是一種常被用於估計邊緣分佈 (marginal distribution) 的瑪律可夫鏈蒙特卡羅技術 (Markov chain Monte Carlo technique)。為了加速吉布斯採樣，人們最近產生了通過非同步執行並行處理它的興趣。儘管一些經驗結果表明許多模型都可以有效地進行非同步採樣，但傳統的瑪律克夫鏈分析卻無法應用於非同步的情況，因此對非同步吉布斯採樣只有很少的瞭解。在這篇論文中，我們設法更好地瞭解了非同步吉布斯的兩個主要挑戰：偏差 (bias) 和混合時間 (mixing time)。我們通過實驗證明了我們的理論結果是符合實際結果的。

論文題目：*Pixel Recurrent Neural Networks*

中文題目：圖元迴圈神經網路

論文作者：Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu

論文地址：<https://arxiv.org/abs/1601.06759>

論文解讀：在無監督學習中，給自然圖像分佈建模是一個里程碑式的問題。這項任務要求得到可以同時表現圖像、易於處理並且具備可擴展性的圖像模型。我們展示了一個可以沿二維空間維度依次預測圖像中圖元的深度神經網路。我們的方法建立了原始圖元值的離散概率模型，並且編碼了圖像中完整的依賴關係集合。該架構的不同之處在於它包括快速二維迴圈層 (recurrent layers) 和對深度迴圈網路中殘差連接 (residual connections) 的有效利用。我們完成了自然圖像上的對數似然分數，其比之前最先進的還要好很多。我們主要的成果還包括提供多樣化的 ImageNet 資料集基準。從模型中生成了新鮮多樣且全域同一的樣本。此論文提出了一系列生成模型，可直接對圖元的統計依賴關係進行建模。這些模型包括兩個 PixelRNN：Row LSTM 和 Diagonal BiLSTM (區別主要在於它們進行預測使用到的條件資訊所在的領域)；一個 PixelCNN，以及一個多尺度 PixelRNN。

論文題目： *Dueling Network Architectures for Deep Reinforcement Learning*

中文題目：深度強化學習中的競爭網路架構

論文作者：Ziyu Wang, Tom Schaul, Matteo Hessel, Hado van Hasselt, Marc Lanctot, Nando de Freitas

論文地址：<https://arxiv.org/abs/1511.06581>

論文解讀：近幾年，已經有很多在強化學習中使用深度表徵獲得成功的例子。然而，這些應用中的很多例子仍然使用傳統的架構，比如卷積網路、LSTMs，或者是自動編碼器。在此論文中，我們提出了一個新的用於無模型 (model free) 強化學習的神經網路架構。我們的競爭網路 (dueling network) 表示了兩種獨立的評估量：一個用於狀態價值函數 (state value function)，一個用於狀態依存動作優勢

函數 (state-dependent action advantage function) 。這一分解的主要好處是在沒有將任何變化強加於低層的強化學習演算法的情況下，在動作 (action) 間歸納學習。我們的結果顯示，這一架構在多種價值相似的動作面前能引發更好的政策評估。此外，這一競爭架構使得我們的強化學習代理勝過 Atari 2600 領域最前沿的研究。在這篇論文中，作者基於分開建模狀態值和動作優勢的想法，提出了一款可供選擇的用於深度 Q 網路 (DQN) 的架構和相關的學習方案。當被應用於 Atari 學習環境 (Atari Learning Environment) 基準時，這項技術顯著推進了當前最先進的研究成果。

- 2015 年最佳論文

論文題目： *A Nearly-Linear Time Framework for Graph-Structured Sparsity*

中文題目：圖結構稀疏性的近似線性時間框架

論文作者：Hegde, Chinmay, Indyk, Piotr, Schmidt, Ludwig

論文地址：<http://proceedings.mlr.press/v37/hegde15.pdf>

論文解讀：本文引入了一個通過圖定義的稀疏結構框架。其方法較靈活，並且推廣到了以前研究過的幾個稀疏模型。此外，本文還為該稀疏度模型提供了有效的投影演算法，該模型幾乎在線性時間內運行。在稀疏恢復的背景下，本文證明了該框架在理論上實現了廣泛參數下的資訊最優樣本複雜性。本文用實驗來補充該理論分析，證明該演算法在實踐中也改進了先前的工作。

論文題目： *Optimal and Adaptive Algorithms for Online Boosting*

中文題目：Online Boosting 的優化和自我調整演算法

論文作者：Alina Beygelzimer, Satyen Kale, Haipeng Luo

論文地址：<https://arxiv.org/abs/1502.02651>

論文解讀：我們學習線上促進，這是一項將任何一個弱的線上學習者轉變為強的線上學習者的任務。基於對網路學習能力弱的一個新的自然定義，我們開發了兩種線上增強演算法。第一種演算法是線上版本的 Boost by Majority。通過證明一個匹配下界，我們證明了該演算法對於弱學習者的數量和達到指定精度所需的樣本複雜度是本質上最優的。然而，這種優化演算法並不具有自我調整性。利用線上損失最小化的工具，推導了一種無參數但非最優的自我調整線上增強演算法。這兩種演算法都與基礎學習者一起工作，基礎學習者可以直接處理示例重要性權重，也可以使用升遷者定義的概率拒絕抽樣示例。結果與廣泛的實驗研究相輔相成。

- 2014 年最佳論文

論文題目：*Understanding the Limiting Factors of Topic Modeling via Posterior Contraction Analysis*

中文題目：經由過程後驗收縮分析理解主題建模的限制因素

論文作者：Ming Zhang, Jian Tang, Zhaoshi Meng

論文地址：<http://proceedings.mlr.press/v32/tang14.pdf>

論文解讀：潛在狄利克雷分佈 (LDA) 已經成為了機器學習建模工具箱中的一個標準工具。它們已被應用於各種不同程度的資料集、背景和任務，但是迄今為止，幾乎沒有正式的理论來解釋 LDA 的行為，並且儘管對其很熟悉，但是對

影響模型推理性能的資料的性質幾乎沒有系統性的分析和指導。本文試圖通過對影響 LDA 的性能的因素進行系統分析來解決此問題。本文提出的定理闡明了隨著資料量的增加後驗概率，並使用綜合和真實的資料集進行了全面的支持性實證研究。基於這些結果，本文對如何為主題模型識別合適的資料集以及如何制定特定的模型參數提供了實際指導。

- 2013 年最佳論文

論文題目：*Vanishing Component Analysis*

中文題目：經由過程後驗收縮分析理解主題建模的限制因素

論文作者：Roi Livni, Shai Shalevshwartz, Amir Globerson

論文地址：<http://proceedings.mlr.press/v28/livni13.pdf>

論文解讀：傳統的特徵選擇方法通常是在採樣中選擇顯著的特徵，作者研究的是，在特徵選擇時，是否能夠選擇一些不變的特徵。文章描述並分析了構造一組消失理想生成器的有效過程。該過程是數值穩定的，並且可以用於近似消失多項式。這由此得到的多項式捕捉資料中的非線性結構，例如可用於監督學習。與核方法的實證比較表明，文章提出的方法構造了更緊湊的分類器，具有相當的精度。

論文題目：*Fast Semidifferential-based Submodular Function Optimization*

中文題目：基於半微分的快速子模組函數優化

論文作者：Iyer, Rishabh, Jegelka, Stefanie, Bilmes, Jeff

論文地址：<http://export.arxiv.org/pdf/1308.1006>

論文解讀：本文提出了一種實用而強大的基於離散半微分(子微分和超微分)的無約束和約束子模函數優化新框架。所得到的演算法反復計算並有效地優化了子模半梯度，為子模優化提供了新的、通用的方法。此外，本文的方法還採取步驟，提供適用於次模最小化和最大化的統一範式，這些問題在歷史上得到了相當明顯的處理。本文的演算法的實用性很重要，因為子模性由於其自然和廣泛的適用性，最近在機器學習中佔據了優勢。分析了本文的極小化和最大化演算法的理論性質，表明許多最先進的最大化演算法都是特殊情況。最後，本文將理論分析與實證實驗相補充。

- 2012 年最佳論文

論文題目：*Bayesian Posterior Sampling via Stochastic Gradient Fisher Scoring*

中文題目：通過隨機梯度 Fisher 得分進行貝葉斯後驗採樣

論文作者：S.Ahn, A.Korattikara, M.Welling

論文地址：<https://arxiv.org/ftp/arxiv/papers/1206/1206.6380.pdf>

論文解讀：在本文中，討論了以下問題：如果我們只允許對生成的每個樣本接觸一小批資料項目，那麼我們是否可以近似地從貝葉斯後驗分佈中提取樣本？本文提出了一種基於隨機梯度朗格文方程 (SGLD) 的混合演算法，但是其混合速率慢。通過利用貝葉斯中心極限定理，我們擴展了 SGLD 演算法，使其在高混合速率下從後驗函數的正態近似中採樣，而在慢混合速率下，它將使用預調節矩

陣模擬 SGLD 的行為。作為一個額外的好處，該演算法使人想起費希爾評分 (隨機梯度) ，因此在老化過程中是一個有效的優化器。

- 2011 年最佳論文

論文題目：*Computational Rationalization: The Inverse Equilibrium Problem*

中文題目：計算合理化：逆向平衡問題

論文作者：Kevin Waugh, Brian D.Ziebart, J. Andrew (Drew) Bagnell

論文地址：https://www.ri.cmu.edu/pub_files/2011/6/paper.pdf

論文解讀：從少量的觀察結果中模擬不完美因素的有目的行為是一項具有挑戰性的任務。當限制在單智慧體決策理論設置下時，逆最優控制技術假定觀測行為是未知決策問題的近似最優解。這些技術學習解釋示例行為的實用函數，然後可用於準確預測或模擬類似觀察到或未觀察到的情況下的未來行為。在這項工作中，我們考慮了競爭和合作多代理領域中的類似任務。在這裡，不同于單一代理設置，玩家不能近似地最大化其回報-它必須推測其他代理如何行動，以影響遊戲的結果。利用遺憾博弈論的概念和最大熵原理，提出了一種預測和概括行為的方法，並在此領域中恢復了獎勵函數。

- 2010 年最佳論文

論文題目：*Hilbert Space Embeddings of Hidden Markov Models*

中文題目：隱瑪律科夫模型的希爾伯特空間嵌入

論文作者：Le Song, Byron Boots, Sajid M. Siddiqi, Geoffrey J. Gordon

論文地址：<https://storage.googleapis.com/pub-tools-public-publication-data/pdf/36408.pdf>

論文解讀：隱瑪律可夫模型是序列資料建模的重要工具。然而，它們僅限於離散的潛在狀態，並且主要限於高斯和離散觀測。而且，HMM 的學習演算法主要依賴於局部搜索啟發式演算法，除了下面描述的譜方法。本文提出了一個非參數 HMM，它將傳統的 HMM 擴展到結構化和非高斯連續分佈。此外，本文還匯出了一個學習這些 HMM 的局部最小自由核譜演算法。本文將該方法應用於機器人視覺資料、槽車慣性感測器資料和音訊事件分類資料，結果表明，在這些應用中，嵌入式 HMM 的性能超過了以往的先進水準。

- 2009 年最佳論文

論文題目：*Structure preserving embedding*

中文題目：結構保留嵌入

論文作者：Blake Shaw, Tony Jebara

論文地址：<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.149.7500&rep=rep1&type=pdf>

論文解讀：結構保留嵌入 (SPE) 是一種用於在歐幾裡得空間中嵌入圖形的演算法，該嵌入是低維的，並保留了輸入圖的全域拓撲屬性。如果諸如 k 最近鄰之類的連通性演算法可以在嵌入後僅從節點的座標輕鬆恢復輸入圖的邊緣，則可以保留拓撲。SPE 公式化為半定程式，該程式學習受一組線性不等式約束的低階內核矩陣，該不等式捕獲輸入圖的連通性結構。SPE 在圖形的視覺化和無失真壓縮方面提供了顯著的改進，勝過了諸如光譜嵌入和 Laplacian 特徵

圖之類的流行方法。我們發現，僅需使用幾個維度就可以正確嵌入許多經典圖和網路。

2.7.2 NeurIPS 歷年最佳論文解讀

- 2018 年最佳論文

論文題目：*Non-delusional Q-learning and Value-iteration*

中文題目：非妄想 Q 學習和價值反覆運算

論文作者：Tyler Lu, Dale Schuurmans, Craig Boutilier

論文地址：<http://120.52.51.18/papers.NeurIPS.cc/paper/8200-non-delusional-q-learning-and-value-iteration.pdf>

論文解讀：本文用函數逼近法確定了 Q-學習和其他形式的動態規劃中誤差的根本來源。當近似結構限制了可表達的貪婪策略的類別時，就會產生偏差。由於標準 Q-updates 對可表達的策略類做出了全域不協調的動作選擇，因此可能導致不一致甚至衝突的 Q 值估計，從而導致病態行為，例如過度/低估、不穩定甚至發散。為了解決這個問題，本文引入了策略一致性的新概念，並定義了一個本地備份流程，通過使用資訊集，也就是記錄與備份 Q 值一致的策略約束集，來確保全域一致性。本文證明了使用這種備份的基於模型和無模型的演算法都能消除妄想偏差，從而產生第一種已知演算法，保證在一般條件下的最優結果。此外，這些演算法只需要多項式的多個資訊集（從潛在的指數支援）。最後，本文建議使用其他實用的啟發式價值反覆運算和 Q 學習方法去嘗試減少妄想偏差。

論文題目： *Optimal Algorithms for Non-Smooth Distributed Optimization in Networks*

中文題目：非光滑凸函數的分散式優化演算法

論文作者：Kevin Scaman, Francis Bach, Sebastien Bubeck, Laurent Massoulié, Yin Tat Lee

論文地址：<https://arxiv.org/pdf/1806.00291.pdf>

論文解讀：在本文中，我們考慮使用計算單元網路的非光滑凸函數的分散式優化。我們在兩個正則性假設下研究這個問題：(1) 全域目標函數的 Lipschitz 連續性；(2) 局部個體函數的 Lipschitz 連續性。在局部正則性假設下，本文給出了稱為多步原對偶(MSPD)的一階最優分散演算法及其相應的最優收斂速度。這個結果的一個顯著特點是，對於非光滑函數，當誤差的主要項在 $O(1/t)$ 中時，通信網路的結構僅影響 $O(1/t)$ 中的二階項，其中 t 是時間。換言之，即使在非強凸目標函數的情況下，由於通信資源的限制而導致的誤差也以快速率減小。在全域正則性假設下，基於目標函數的局部平滑，給出了一種簡單而有效的分散式隨機平滑(DRS)演算法，並證明了DRS在最優收斂速度的 $d^{1/4}$ 乘因數內，其中 d 為底層。

論文題目： *Nearly Tight Sample Complexity Bounds for Learning Mixtures of Gaussians via Sample Compression Schemes*

中文題目：通過樣本壓縮方案學習混合高斯模型的近乎緊密的樣本複雜性邊界

論文作者：Hassan Ashtiani, Shai Ben-David, Nicholas J. A. Harvey, Christopher Liaw, Abbas Mehrabian, Yaniv Plan

論文地址：<https://papers.NeurIPS.cc/paper/7601-nearly-tight-sample-complexity-bounds-for-learning-mixtures-of-gaussians-via-sample-compression-schemes.pdf>

論文解讀：本文證明了 $O(k d^2/\epsilon^2)$ 樣本對於學習 R^d 中 k 個高斯的混合，直至總變差距離中的誤差 ϵ 來說，是充分必要條件。這改善了已知的上界和下界這一問題。對於軸對準高斯混合，本文證明了 $O(k d/\epsilon^2)$ 樣本匹配一個已知的下界是足夠的。上限是基於樣本壓縮概念的分佈學習新技術。任何允許這種樣本壓縮方案的分佈類都可以用很少的樣本來學習。此外，如果一類分佈具有這樣的壓縮方案，那麼這些產品和混合物的類也是如此。本文主要結果的核心是證明了 R^d 中的高斯類能有效的進行樣本壓縮。

論文題目：*Neural Ordinary Differential Equations*

中文題目：神經常微分方程

論文作者：Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, David Duvenaud

論文地址：<https://arxiv.org/pdf/1806.07366.pdf>

論文解讀：本文介紹了一系列新的深度神經網路模型。本文使用神經網路參數化隱藏狀態的導數，而不是指定隱藏層的離散序列。使用黑盒微分方程求解器計算網路的輸出。這些連續深度模型具有恒定的記憶體成本，使其評估策略適應每個輸入，並且可以明確地交換數值精度以獲得速度。本文在連續深度殘差網路和連續時間潛變數模型中證明了這些性質。本文還構建了連續歸一化流，這是一種可以通過最大似然進行訓練的生成模型，無需對資料維度進行分區或排序。為了訓練，本文展示了如何通過任何 ODE 求解器進行可擴展反向傳播，而無需訪問其內部操作。這允許在較大模型中對 ODE 進行端到端訓練。

- 2017 年最佳論文

論文題目： *Safe and Nested Subgame Solving for Imperfect-Information Games*

中文題目：不完全資訊博弈的安全嵌套子博弈求解

論文作者：Noam Brown, Tuomas Sandholm

論文地址：<https://arxiv.org/pdf/1705.02955.pdf>

論文解讀：和完美資訊博弈不同，不完美資訊博弈不能通過將博弈分解為可獨立求解的子博弈而求得占優策略。因此本文越來越多地使用計算密集的均衡判定技術，並且所有的決策必須將博弈的策略當作一個整體。本文提出了一種無論在理論上還是在實踐上都超越了之前方法的子博弈求解技術。本文還展示了如何對它們和以前的子博弈求解技術進行調整，以對超出初始行動提取 (original action abstraction) 的對手的行動做出應答；這遠遠超越了之前的頂尖方法，即行動轉化 (action translation)。最後，本文展示了當博弈沿著博弈樹向下進行時，子博弈求解可能會重複進行，從而大大降低可利用性。

論文題目： *Variance-based Regularization with Convex Objectives*

中文題目：帶有凸物件的基於方差的正則化方法

論文作者：Hongseok Namkoong, John Duchi

論文地址：<https://arxiv.org/pdf/1610.02581.pdf>

論文解讀：本文研究了一種風險最小化和隨機優化的方法，該方法可以為方差提供一個凸屬性的替代項，並允許在逼近和估計誤差間實現近似最優與高效計

算間的權衡。本文的方法建立在分佈魯棒性優化和 Owen 經驗性似然度的基礎上，並提供了一些有限樣本 (finite-sample) 和漸進結果以展示估計器的理論性能。具體來說，本文證明了該過程具有最優性保證 (certificates of optimality)，並通過逼近和最優估計誤差間良好的權衡在更一般的設定下比經驗風險最小化方法有更快的收斂率。本文還給出了確鑿的經驗性證據，表明估計器在實踐中會在訓練樣本的方差和絕對性能之間進行權衡。此外，估計器也會提升標準經驗風險最小化方法在許多分類問題上的測試性能。

論文題目： *A Linear-Time Kernel Goodness-of-Fit Test*

中文題目：一種線性時間核的擬合優度測試方法

論文作者：Wittawat Jitkrittum, Wenkai Xu, Zoltan Szabo, Kenji Fukumizu, Arthur Gretton.

論文地址：<https://arxiv.org/pdf/1705.07673.pdf>

論文解讀：本文提出了一個全新的擬合優度 (goodness-of-fit) 的適應性測試法，其中計算資源的消耗與樣本數呈線性關係。本文通過最小化假負類率來學習最能展示觀察樣本和參考模型之間差異的測試特徵。這些特徵是通過 Stein 法構造的——這意味著沒有必要計算模型的歸一化常數。本文分析了新測試的 Bahadur 漸進效率，並證明了在均值偏移 (mean-shift) 的情況下，無論選擇哪個測試參數，本文的測試總是比先前的線性時間核測試具有更高的相對效率。在高維和模型結構可用的情況下，本文的擬合優度測試在模型中抽取樣本，表現遠遠超越基於最大平均差異 (Maximum Mean Discrepancy) 的二次時序雙樣本測試。

- 2016 年最佳論文

論文題目： *Value Iteration Networks*

中文題目：價值反覆運算網路

論文作者：Aviv Tamar, Yi Wu, Garrett Thomas, Sergey Levine, Pieter Abbeel

論文地址：<https://arxiv.org/pdf/1602.02867.pdf>

論文解讀：本文介紹了一個價值反覆運算網路 (VIN)：一種完全可微分的神經網路，內置“規劃模組”。VIN 可以學習計畫，並且適用於預測涉及基於計畫的推理的結果，例如加強學習的政策。我們的方法的關鍵是一種新的可微近似值反覆運算演算法，它可以表示為卷積神經網路，並使用標準的反向傳播訓練端到端。本文基於離散和連續路徑規劃域以及基於自然語言的搜索任務評估基於 VIN 的策略。本文表明，通過學習一個明確的規劃計算，VIN 策略可以更好地推廣到新的、未發現的領域。

論文題目： *Matrix Completion has No Spurious Local Minimum*

中文題目：矩陣填充沒有假的局部最小值

論文作者：Rong Ge, Jason Lee, Tengyu Ma

論文地址：<https://arxiv.org/pdf/1605.07272.pdf>

論文題目：矩陣填充是一個基本的機器學習問題，具有廣泛的應用，尤其是在協作過濾和推薦系統中。簡單的非凸優化演算法在實踐中很流行且有效。我們證明了用於矩陣填充的常用非凸目標函數沒有假的局部最小值——所有局部最小值也必須是全域的。因此，許多流行的優化演算法（例如隨機梯度下降）可以通過多項式時間內的任意初始化可證明地解決矩陣填充問題。當觀察到的條目包

含雜訊時，結果可以推廣到該設置。我們認為，我們的主要證明策略對於理解其他涉及部分或嘈雜觀測值的統計問題的幾何性質很有用。

論文題目： *Interactive musical improvisation with Magenta*

文題目：基於 Magenta 的即興音樂交互體驗

論文作者：Adam Roberts, Jesse Engel, Curtis Hawthorne, Ian Simon, Elliot Waite, Sageev Oore, Natasha Jaques, Cinjon Resnick, Douglas Eck

論文地址：<https://nips.cc/Conferences/2016/ScheduleMultitrack?event=6307>

論文解讀：作者結合了基於 LSTM 的迴圈神經網路和 Deep Q-learning 建立了即時生成音樂序列。LSTM 的任務是學習音樂評分（編碼為 MIDI，而不是音訊檔）的一般結構。Deep Q-learning 用來改進基於獎勵的序列，如期望的類型，組成正確性和預測人類協作者演奏的內容。基於 RNN 模型的生成與強化學習的結合是一種生成音樂的全新方式。這種方式比單獨使用 LSTM 更為穩定，生成的音樂更加好聽。該方法有兩個任務：生成對短旋律輸入的回應，以及即時生成對旋律輸入的伴奏，持續對未來輸出進行預測。本方法在 TensorFlow 中加入了一個全新的 MIDI 介面產生即興的音樂體驗，讓使用者可以與神經網路即時交互。

● 2015 年最佳論文

論文題目： *Competitive Distribution Estimation: Why is Good-Turing Good*

中文題目：競爭分佈估計：為什麼 Good-Turing 好

論文作者：Alon Orlitsky, Ananda Theertha Suresh

論文地址：<http://120.52.51.17/papers.NeurIPS.cc/paper/5762-competitive-distribution-estimation-why-is-good-turing-good.pdf>

論文解讀：該論文屬於統計學習的理論研究範疇，它對估計離散變數的分佈律這一普遍問題，提出了基於 Good-Turing 估計量的兩種改進方法，借助對先驗的最優估計量，給出了針對任意分佈律的近似最優的高效估計。論文不僅指出這兩種方法可以快速收斂，同時還給出相應的理論分析。

論文題目：*Fast Convergence of Regularized Learning in Games*

中文題目：博弈中正則化學習的快速收斂

論文作者：Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, Robert Schapire

論文地址：<http://120.52.51.17/papers.NeurIPS.cc/paper/5763-fast-convergence-of-regularized-learning-in-games.pdf>

論文解讀：我們證明了具有新近偏置形式的自然類正則化學習演算法，可以在多人正常形式博弈中達到更快的收斂速度，從而有效地近似並達到粗略的相關均衡。當博弈中的每個玩家使用我們類中的演算法時，它們會在 $O(T^{-3/4})$ 處衰減，而效用的總和會在 $O(T^{-1})$ 處收斂至最佳值——在最差的情況 $O(T^{-1/2})$ 比率下有所改善情況。我們展示了該類中任何演算法的黑盒衰減，以針對對手達到 $\tilde{O}(T^{-1/2})$ 的速率，同時保持該類中演算法的較快速率。我們的結果擴展了 Rakhlin 和 Shridharan 和 Daskalakis 等人的結果，它們只針對特定演算法分析了兩人零和博弈。

- 2014 年最佳論文

論文題目：*Asymmetric LSH (ALSH) for sublinear time Maximum Inner Product Search (MIPS)*

中文題目：次線性時間的不對稱 LSH (ALSH) 最大內積檢索 (MIPS)

論文作者：Anshumali Shrivastava, Ping Li

論文地址：<https://arxiv.org/pdf/1405.5869.pdf>

論文解讀：我們提出了第一個可證明的次線性時間雜湊演算法，用於近似最大內積檢索 (MIPS)。使用 (未歸一化的) 內積作為基礎相似性度量進行檢索是一個已知的難題，並且為 MIPS 查找雜湊方案是很困難。雖然現有的本地敏感雜湊 (LSH) 框架不足以解決 MIPS，但在本文中，我們將 LSH 框架擴展為允許非對稱雜湊方案。我們的方法基於一個關鍵的觀察，即在獨立的不對稱變換之後，找到最大內積的問題可以轉化為經典設置中的近似近鄰搜索問題。這個關鍵的發現使針對 MIPS 的高效亞線性雜湊方案成為可能。我們提出的演算法簡單易實現。所提出的散列方案與協作過濾中的兩種流行的常規 LSH 方案相比，顯著節省了計算量：(i) 符號隨機投影 (SRP) 和 (ii) 基於 L-2 範數的 p 穩定分佈 (L2LSH)，在 Netflix 和 Movielens (10M) 資料集上的專案推薦任務。

論文題目：*A* Sampling*

中文題目：A*採樣

論文作者：Chris J. Maddison, Daniel Tarlow, Tom Minka

論文地址：<https://arxiv.org/pdf/1411.0030.pdf>

論文解讀：從離散分佈中提取樣本的問題可以轉化為離散優化問題。在這項工作中，本文展示了如何將連續分佈的採樣轉化為連續空間上的優化問題。該方法的核心是最近在數學統計學中描述的一個隨機過程，本文稱之為 Gumbel 過程。本文提出了一種新的 Gumbel 過程和 A*採樣結構，這是一種實用的通用採樣演算法，它使用 A*搜索來搜索 Gumbel 過程的最大值。本文分析了 A*抽樣的正確性和收斂時間，並從經驗上證明了它比最相關的自我調整拒絕抽樣演算法更有效地利用了邊界和似然估計。

- 2013 年最佳論文

論文題目：*A Memory Frontier for Complex Synapses*

中文題目：複雜突觸的記憶邊界

論文作者：Subhaneil Lahiri, Surya Ganguli

論文地址：<http://120.52.51.14/papers.NeurIPS.cc/paper/4872-a-memory-frontier-for-complex-synapses.pdf>

論文解讀：一個令人難以置信的鴻溝將突觸的理論模型分開，通常僅由表示突觸後電位大小的單個標量值描述，來自真實突觸下的分子信號傳導途徑的巨大複雜性。為了理解這種分子複雜性對學習和記憶的功能貢獻，必須將突觸的理論概念從單個標量擴展到具有許多內部分子功能狀態的整個動力系統。這裡產生了一個基本問題，突觸複雜性如何產生記憶？為了解決這個問題，本文開發了新的數學定理，闡明了複雜突觸的結構組織和記憶特性之間的關係，這些突觸本身就是分子網路。此外，在證明這些定理時，本文發現了一個基於第一次通過時間理

論的框架，對複雜突觸模型的內部狀態施加順序，從而簡化了突觸結構和功能之間的關係。

論文題目： *Submodular Optimization with Submodular Cover and Submodular Knapsack Constraints*

中文題目：具有子模組覆蓋和子模組背包約束的子模組優化

論文作者：Rishabh Iyer, Jeff Bilmes

論文地址：<https://static.aminer.org/pdf/20160902/web-conf/NEURIPS/NEURI PS-2013-2874.pdf>

論文解讀：我們研究了兩個新的優化問題——最小化受子模組化下界約束（子模組化覆蓋）的子模函數和最大化子模函數受下模組化上限約束（子模組化背包）的約束。我們受到機器學習中許多實際應用的啟發，這些應用包括感測器放置和資料子集選擇，這些應用要求最大化某個子模組功能（例如覆蓋範圍或分集），同時最小化另一個子模組功能（例如合作成本）。我們發現通過將這些問題表述為約束優化（對於許多應用程式而言更自然），可以實現許多有界逼近的保證。我們還表明，這兩個問題都是密切相關的，可以使用求解一個問題的近似演算法來獲得對另一個問題的近似保證。我們提供了兩個問題的結果，從而表明我們的逼近因數嚴格到對數因數。最後，我們通過實驗證明了演算法的性能和良好的可伸縮性。

論文題目： *Scalable Influence Estimation in Continuous-Time Diffusion Networks*

中文題目：連續時間擴散網路中的可擴展影響估計

論文作者：Nan Du, Le Song, Manuel Gomez-Rodriguez, Hongyuan Zha

論文地址：<https://static.aminer.org/pdf/20160902/web-conf/NEURIPS/NEURIPS-2013-2951.pdf>

論文解讀：如果從媒體網站發佈一條資訊，我們能否預測它是否可以在一個月內傳播到一百萬個網頁？由於需要同時處理任務的時間敏感性和可伸縮性要求，因此影響估計問題非常具有挑戰性。在本文中，我們提出了一種用於連續時間擴散網路中影響估計的隨機演算法。我們的演算法可以用 $|V|$ 估計網路中每個節點的影響。節點和 $|\varepsilon|$ 使用 $n = O(1/\varepsilon^2)$ 隨機化並以對數因數 $O(n|\varepsilon| + n|V|)$ 計算邊緣到 ε 的精度。當在貪婪影響最大化方法中用作子常式時，我們提出的演算法可確保找到一個至少受 $(1 - 1/|\varepsilon|) \text{OPT} - 2C\varepsilon$ 影響的 C 節點集，其中 OPT 是最佳值。對合成資料和實際資料進行的實驗均表明，該演算法可以輕鬆擴展至數百萬個節點的網路，同時在估計影響力的準確性和品質方面都大大優於以前的最新技術，選擇節點以最大程度地發揮影響力。

- 2012 年最佳論文

論文題目：*No voodoo here! Learning discrete graphical models via inverse covariance estimation*

中文題目：通過逆協方差估計學習離散圖模型

論文作者：Po-Ling Loh, Martin Wainwright

論文地址：<https://pdfs.semanticscholar.org/85fb/8ff18d1a588b7b314faaa6a17f6b68818683.pdf>

論文解讀：本文研究了廣義協方差矩陣的逆的支援與離散圖形模型的結構之間的關係。本文證明了對於某些圖結構，指標變數的逆協方差矩陣對圖的頂點的支持反映了圖的條件獨立結構。本文的工作擴展了以前僅針對多元高斯分佈建立的結果，並且部分地回答了關於非高斯分佈的逆協方差矩陣含義的開放問題。本文提出了基於可能損壞的觀測值的具有有界度的一般離散圖形模型的圖選擇方法，並通過模擬驗證本文的理論結果。在此過程中，本文還在基於損壞和缺失觀測的稀疏高維線性回歸設置中建立支援恢復的新結果。

論文題目： *Discriminative Learning of Sum-Product Networks*

中文題目：和積網路的判別學習

論文作者：Robert Gens, Pedro Domingos

論文地址：<http://papers.nips.cc/paper/4516-discriminative-learning-of-sum-product-networks.pdf>

論文解讀：Sum-product 網路是一種新的深度架構，可以對高樹寬模型進行快速、準確的推斷。迄今為止，僅提出了用於生成 SPN 的生成方法。在本文中，我們提出了第一種針對 SPN 的判別式訓練演算法，將前者的高精度與後者的表示能力和易處理性相結合。我們表明，可分辨的判別式 SPN 的類別比可處理的可區分性 SPN 的類別更廣泛，並提出了一種有效的反向傳播演算法來計算條件對數似然度的梯度。我們在標準圖像分類任務上測試判別式 SPN。我們使用迄今在 CIFAR-10 資料集上獲得最佳結果的方法，其性能比具有 SPN 架構的方法(具有判別性學習本地圖像結構)的性能要少。即使僅使用資料集的標記部分，我們也報告了 STL-10 上公佈的最高測試準確性。

- 2011 年最佳論文

論文題目：*Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials*

中文題目：具有高斯邊緣電位的完全連接 CRF 中的有效推理

論文作者：Philipp Krähenbühl, Vladlen Koltun

論文地址：<https://static.aminer.org/pdf/20160902/web-conf/NEURIPS/NEURIPS-2011-1998.pdf>

論文解讀：用於多類圖像分割和標記的大多數最新技術使用在圖元或圖像區域上定義的條件隨機欄位。儘管區域級模型通常具有密集的成對連通性，但圖元級模型卻要大得多，並且只允許稀疏圖結構。在本文中，我們考慮在圖像的完整圖元集上定義的完全連接的 CRF 模型。生成的圖具有數十億條邊，這使得傳統的推理演算法不切實際。我們的主要貢獻是針對全連接 CRF 模型的高效近似推理演算法，其中成對邊緣勢能由高斯核的線性組合定義。我們的實驗表明，圖元級的密集連線性可顯著改善分割和標記的準確性。

論文題目：*Priors Over Recurrent Continuous Time Processes*

中文題目：連續時間過程的優先順序

論文作者：Ardavan Saeedi, Alexandre Bouchard-Côte

論文地址：<https://static.aminer.org/pdf/20160902/web-conf/NEURIPS/NEURIPS-2011-2195.pdf>

論文解讀：本文引入 Gamma 指數過程 (GEP) ，這是一個大型連續時間過程系列的先驗。該先驗的分層版本 (HGEP; Hierarchical GEP) 產生用於分析複雜時間序列的有用模型。基於 HGEP 的模型顯示出許多有吸引力的特性：等待時間的共軛性, 可交換性和封閉形式預測分佈, 以及時間尺度參數的精確 Gibbs 更新。在建立這些屬性之後, 本文展示了如何使用粒子 MCMC 方法有效地進行後驗推理。本文將本文的模型應用於估計多發性硬化症的疾病進展和 RNA 進化建模的問題。在這兩個領域, 本文發現本文的模型優於標準的速率矩陣估計方法。

論文題目：*Fast and Accurate K-means for Large Datasets*

中文題目：大型資料集的快速準確 K-means

論文作者：Michael Shindler, Alex Wong, Adam Meyerson

論文地址：<http://120.52.51.15/papers.NeurIPS.cc/paper/4362-fast-and-accurate-k-means-for-large-datasets.pdf>

論文解讀：群集是許多應用程式中的一個普遍問題。在資料太大而無法存儲在主記憶體中、並且必須循序存取 (例如從磁片)、必須使用盡可能少的記憶體的情況下, 我們考慮 k 均值問題。我們的演算法基於最新的理論結果, 並進行了重大改進以使其實用。然後, 我們合併近似最近鄰搜索以計算 $o(nk)$ 中的 k 均值 (其中 n 是資料點的數量; 請注意, 在給定解的情況下, 計算成本需要 $\Theta(nk)$ 時間)。我們證明了我們的演算法在理論上和實驗上都優於現有演算法, 從而在理論和實踐上均提供了最先進的性能。

- 2010 年最佳論文

論文題目：*Construction of dependent dirichlet Processes based on Poisson Processes*

中文題目：基於泊松過程的 DDP 構建

論文作者：Dahua Lin, Eric Grimson, John Fisher

論文地址：<https://static.aminer.org/pdf/20160902/web-conf/NEURIPS/NEURIPS-2010-3901.pdf>

論文解讀：本文提出了一種構造依賴 Dirichlet 過程的方法。新的方法揭示了 Dirichlet 和泊松過程之間的內在關係，以便創建一個適合用作先前演化混合模型的 Dirichlet 過程的瑪律可夫鏈。該方法允許元件模型隨時間的創建、移除和位置變化，同時保持隨機測量略微 DP 分佈的屬性。此外，本文推導出用於模型推理的 Gibbs 採樣演算法，並在合成和實際資料上進行測試。實證結果表明該方法可有效地估算動態變化的混合模型。

論文題目：*A Theory of Multiclass Boosting*

中文題目：多類別 Boosting 演算法的理論

論文作者：Indraneel Mukherje, Robert E Schapire

論文地址：<https://static.aminer.org/pdf/20160902/web-conf/NEURIPS/NEURIPS-2010-3934.pdf>

論文解讀：Boosting 將弱分類器組合在一起，以形成高度準確的預測器。儘管二進位分類的情況已廣為人知，但在多類設置中，缺少對弱分類器的“正確”要求或最有效的增強演算法的概念。在本文中，我們創建了一個廣泛而通用的框架，

在此框架內，我們可以對弱分類器進行精確確定並確定最佳要求，並在某種意義上設計最有效的 Boosting 演算法來滿足此類要求。

- 2009 年最佳論文

論文題目：*An LP View of the M-Best MAP Problem*

論文作者：Menachem Fromer, Amir Globerson

論文地址：<https://static.aminer.org/pdf/20160902/web-conf/NEURIPS/NEURIPS-2009-4089.pdf>

論文解讀：本文考慮在概率圖形模型中以最大概率找到 M 指派的問題。本文展示了如何將這個問題表述為特定多面體上的線性程式 (LP)。本文證明，對於樹狀圖 (和一般的交叉樹)，這個多面體具有特別簡單的形式，並且與單個不等式約束中的邊際多面體不同。本文使用這種表徵來為非樹圖提供近似方案，通過使用這些圖上的生成樹集。本文提出的方法在 LP 鬆弛的背景下提出了 M-最佳推理問題，LP 鬆弛最近得到了相當多的關注，並且已經證明在解決困難的推理問題方面是有用的。本文憑經驗證明，本文的方法經常為高樹寬度的問題找到可證明的精確 M 最佳配置。

論文題目：*Fast Subtree Kernels on Graphs*

中文題目：圖的快速子樹核

論文作者：Nino Shervashidze, Karsten Borgwardt

論文地址：<https://static.aminer.org/pdf/20160902/web-conf/NEURIPS/NEURIPS-2009-4199.pdf>

論文解讀：在本文中，我們提出了圖的快速子樹核。在具有 n 個節點和 m 個邊，且最大度為 d 的圖上，這些高度為 h 的比較子樹核可以用 $O(mh)$ 計算，而 Ramon & Gartner 經典子樹核的縮放比例為 $O(n^2 4^{dh})$ 。效率性的關鍵是觀察到，根據圖論進行的 Weisfeiler-Lehman 同構檢驗很好地計算了作為副產品的子樹核。我們的快速子樹核可以處理帶標籤的圖形，可以輕鬆擴展到大型圖形，並且可以在準確性和執行時間方面在多個分類基準資料集上勝過最新的圖形核。

AMiner

3 電腦視覺

3.1 電腦視覺概念

電腦視覺 (computer vision) ，顧名思義，是分析、研究讓計算機智能化的達到類似人類的雙眼“看”的一門研究科學^[3]。即對於客觀存在的三維立體化的世界的理解以及識別依靠智慧化的電腦去實現。確切地說，電腦視覺技術就是利用了攝像機以及電腦替代人眼使得電腦擁有人類的雙眼所具有的分割、分類、識別、跟蹤、判別決策等功能。總之，電腦視覺系統就是創建了能夠在 2D 的平面圖像或者 3D 的三維立體圖像的資料中，以獲取所需要的“資訊”的一個完整的人工智慧系統。

電腦視覺技術是一門包括了電腦科學與工程、神經生理學、物理學、信號處理、認知科學、應用數學與統計等多門科學學科的綜合性科學技術。由於電腦視覺技術系統在基於高性能的電腦的基礎上，其能夠快速的獲取大量的資料資訊並且基於智慧演算法能夠快速的進行處理資訊，也易於同設計資訊和加工控制資訊集成。

電腦視覺本身包括了諸多不同的研究方向，比較基礎和熱門的方向包括：物體識別和檢測 (Object Detection) ，語義分割 (Semantic Segmentation) ，運動和跟蹤 (Motion & Tracking) ，視覺問答 (Visual Question & Answering) 等^[4]。

- 物體識別和檢測

物體檢測一直是電腦視覺中非常基礎且重要的一個研究方向，大多數新的演算法或深度學習網路結構都首先在物體檢測中得以應用如 VGG-net, GoogLeNet,

ResNet 等等，每年在 imagenet 資料集上面都不斷有新的演算法湧現，一次次突破歷史，創下新的記錄，而這些新的演算法或網路結構很快就會成為這一年的熱點，並被改進應用到電腦視覺中的其它應用中去。

物體識別和檢測，顧名思義，即給定一張輸入圖片，演算法能夠自動找出圖片中的常見物體，並將其所屬類別及位置輸出出來。當然也就衍生出了諸如人臉檢測 (Face Detection)，車輛檢測 (Vehicle Detection) 等細分類的檢測演算法。

- 語義分割

語義分割是近年來非常熱門的方向，簡單來說，它其實可以看作一種特殊的分類——將輸入圖像的每一個圖元點進行歸類，用一張圖就可以很清晰地描述出來。很清楚地就可以看出，物體檢測和識別通常是將物體在原圖像上框出，可以說是“宏觀”上的物體，而語義分割是從每一個圖元上進行分類，圖像中的每一個圖元都有屬於自己的類別。

- 運動和跟蹤

跟蹤也屬於電腦視覺領域內的基礎問題之一，在近年來也得到了非常充足的發展，方法也由過去的非深度演算法跨越向了深度學習演算法，精度也越來越高，不過即時的深度學習跟蹤演算法精度一直難以提升，而精度非常高的跟蹤演算法的速度又十分之慢，因此在實際應用中也很難派上用場。

學術界對待跟蹤的評判標準主要是在一段給定的視頻中，在第一幀給出被跟蹤物體的位置及尺度大小，在後續的視頻當中，跟蹤演算法需要從視頻中去尋找到被跟蹤物體的位置，並適應各類光照變換，運動模糊以及表觀的變化等。但實際上跟蹤是一個不適定問題 (ill posed problem)，比如跟蹤一輛車，如果從車的

尾部開始跟蹤，若是車輛在行進過程中表觀發生了非常大的變化，如旋轉了 180 度變成了側面，那麼現有的跟蹤演算法很大的可能性是跟蹤不到的，因為它們的模型大多基於第一幀的學習，雖然在隨後的跟蹤過程中也會更新，但受限於訓練樣本過少，所以難以得到一個良好的跟蹤模型，在被跟蹤物體的表觀發生巨大變化時，就難以適應了。所以，就目前而言，跟蹤算不上是電腦視覺內特別熱門的一個研究方向，很多演算法都改進自檢測或識別演算法。

- 視覺問答

視覺問答也簡稱 VQA (Visual Question Answering)，是近年來非常熱門的一個方向，其研究目的旨在根據輸入圖像，由使用者進行提問，而演算法自動根據提問內容進行回答。除了問答以外，還有一種演算法被稱為標題生成演算法 (Caption Generation)，即電腦根據圖像自動生成一段描述該圖像的文本，而不進行問答。對於這類跨越兩種資料形態 (如文本和圖像) 的演算法，有時候也可以稱之為多模態，或跨模態問題。

3.2 電腦視覺發展歷史

儘管人們對電腦視覺這門學科的起始時間和發展歷史有不同的看法，但應該說，1982 年瑪律 (David Marr) 《視覺》 (Marr , 1982) 一書的問世，標誌著電腦視覺成為了一門獨立學科。電腦視覺的研究內容，大體可以分為物體視覺 (object vision) 和空間視覺 (spatial vision) 二大部分。物體視覺在於對物體進行精細分類和鑒別，而空間視覺在於確定物體的位置和形狀，為“動作 (action)”服務。正像著名的認知心理學家 J.J.Gibson 所言，視覺的主要功能在於“適應外界

環境，控制自身運動”。適應外界環境和控制自身運動，是生物生存的需求，這些功能的實現需要靠物體視覺和空間視覺協調完成。

電腦視覺 40 多年的發展中，儘管人們提出了大量的理論和方法，但總體上說，電腦視覺經歷了三個主要歷程。即：瑪律計算視覺、多視幾何與分層三維重建和基於學習的視覺。下面將對這三項主要內容進行簡要介紹^[5]。

- 瑪律計算視覺 (Computational Vision)

現在很多電腦視覺的研究人員，恐怕對“瑪律計算視覺”根本不瞭解，這不能不說是一件非常遺憾的事。目前，在電腦上調“深度網路”來提高物體識別的精度似乎就等於從事“視覺研究”。事實上，瑪律的計算視覺的提出，不論在理論上還是研究視覺的方法論上，均具有劃時代的意義。

瑪律的計算視覺分為三個層次：計算理論、表達和演算法以及演算法實現。由於瑪律認為演算法實現並不影響演算法的功能和效果，所以，瑪律計算視覺理論主要討論“計算理論”和“表達與演算法”二部分內容。瑪律認為，大腦的神經計算和電腦的數值計算沒有本質區別，所以瑪律沒有對“演算法實現”進行任何探討。從現在神經科學的進展看，“神經計算”與數值計算在有些情況下會產生本質區別，如目前興起的神經形態計算 (Neuromorphological computing)，但總體上說，“數值計算”可以“模擬神經計算”。至少從現在看，“演算法的不同實現途徑”，並不影響瑪律計算視覺理論的本質屬性。

- 多視幾何與分層三維重建

上世紀 90 年代初電腦視覺從“蕭條”走向進一步“繁榮”，主要得益於以下二方面的因素：首先，瞄準的應用領域從精度和魯棒性要求太高的“工業應用”轉到要

求不太高，特別是僅僅需要“視覺效果”的應用領域，如遠端視訊會議（teleconference），考古，虛擬實境，視頻監控等。另一方面，人們發現，多視幾何理論下的分層三維重建能有效提高三維重建的魯棒性和精度。

多視幾何的代表性人物首數法國 INRIA 的 O.Faugeras，美國 GE 研究院的 R.Hartely 和英國牛津大學的 A.Zisserman。應該說，多視幾何的理論於 2000 年已基本完善。2000 年 Hartley 和 Zisserman 合著的書（Hartley & Zisserman 2000）對這方面的內容給出了比較系統的總結，而後這方面的工作主要集中在如何提高“大資料下魯棒性重建的計算效率”。大資料需要全自動重建，而全自動重建需要反復優化，而反復優化需要花費大量計算資源。所以，如何在保證魯棒性的前提下快速進行大場景的三維重建是後期研究的重點。舉一個簡單例子，假如要三維重建北京中關村地區，為了保證重建的完整性，需要獲取大量的地面和無人機圖像。假如獲取了 1 萬幅地面高解析度圖像（ 4000×3000 ），5 千幅高解析度無人機圖像（ 8000×7000 ）（這樣的圖像規模是當前的典型規模），三維重建要匹配這些圖像，從中選取合適的圖像集，然後對相機位置資訊進行標定並重建出場景的三維結構，如此大的資料量，人工干預是不可能的，所以整個三維重建流程必須全自動進行。這樣需要重建演算法和系統具有非常高的魯棒性，否則根本無法全自動三維重建。在魯棒性保證的情況下，三維重建效率也是一個巨大的挑戰。所以，目前在這方面的研究重點是如何快速、魯棒地重建大場景。

- 基於學習的視覺

基於學習的視覺，是指以機器學習為主要技術手段的電腦視覺研究。基於學習的視覺研究，文獻中大體上分為二個階段：本世紀初的以流形學習為代表的子空間法和目前以深度學習為代表的視覺方法。

物體表達是物體識別的核心問題，給定圖像物體，如人臉圖像，不同的表達，物體的分類和識別率不同。另外，直接將圖像圖元作為表達是一種“過表達”，也不是一種好的表達。流形學習理論認為，一種圖像物體存在其“內在流形”(intrinsic manifold)，這種內在流形是該物體的一種優質表達。所以，流形學習就是從圖像表達學習其內在流形表達的過程，這種內在流形的學習過程一般是一種非線性優化過程。深度學習的成功，主要得益於資料積累和計算能力的提高。深度網路的概念上世紀 80 年代就已提出來了，只是因為當時發現“深度網路”性能還不如“淺層網路”，所以沒有得到大的發展。目前似乎有點電腦視覺就是深度學習的應用之勢，這可以從電腦視覺的三大國際會議：國際電腦視覺會議 (ICCV)，歐洲電腦視覺會議 (ECCV) 和電腦視覺和模式識別會議 (CVPR) 上近年來發表的論文可見一般。目前的基本狀況是，人們都在利用深度學習來“取代”電腦視覺中的傳統方法。“研究人員”成了“調程式的機器”，這實在是一種不正常的“群眾式運動”。牛頓的萬有引力定律，麥克斯韋的電磁方程，愛因斯坦的質能方程，量子力學中的薛定諤方程，似乎還是人們應該追求的目標。

3.3 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為電腦視覺領域全球學者分佈情況：



圖 3-1 電腦視覺領域全球學者分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量優勢明顯且主要分佈在其東西海岸；亞洲也有較多的人才分佈，主要集中在我國東部及日韓地區；歐洲的人才主要分佈在歐洲中西部；其他諸如非洲、南美洲等地區的學者非常稀少；電腦視覺領域的人才分佈與各地區的科技、經濟實力情況大體一致。

此外，在性別比例方面，電腦視覺中男性學者占比 91.0%，女性學者占比 9.0%，男性學者占比遠高於女性學者。

電腦視覺學者的 h-index 分佈如下圖所示，大部分學者的 h-index 分佈在中間區域，其中 h-index 在 20-30 區間的人數最多，有 706 人，占比 34.7%，小於 20 的區間人數最少，有 81 人。

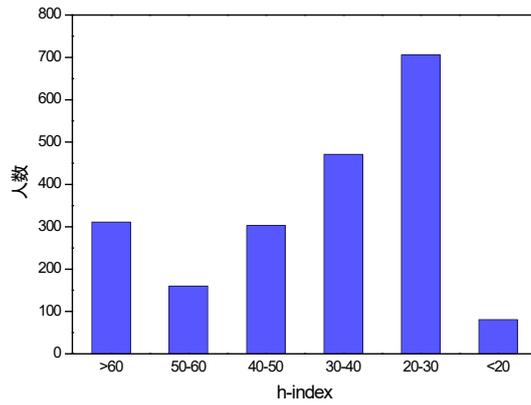


圖 3-2 電腦視覺領域學者 h-index 分佈

● 中國人才分佈



圖 3-3 電腦視覺領域中國學者分佈

我國專家學者在電腦視覺領域的分佈如下圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是珠三角和長三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，通過觀察中國周邊國家的學者數量情況，特別是與日韓、東南亞等亞洲國家相比，中國在電腦視覺領域學者數量相對較多。

中國與其他國家在電腦視覺的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 3-1 電腦視覺領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	1034	88585	86	1459
中國-新加坡	210	20194	96	283
中國-澳大利亞	110	6815	62	147
中國-英國	101	7769	77	148
中國-加拿大	70	7070	101	109
中國-日本	36	2093	58	69
中國-巴基斯坦	26	1933	74	35
中國-瑞士	25	2071	83	46
中國-德國	23	655	28	42
中國-韓國	22	1325	60	51

從上表資料可以看出，中美合作的論文數、引用數、學者數遙遙領先，表明中美間在電腦視覺領域合作之密切；同時，中國與世界各地區之間的合作非常廣泛，前 10 名合作關係裡包含了歐洲、亞洲、北美洲以及大洋洲等；中國與加拿大合作的論文數雖然不是最多，但是擁有最高的平均引用數說明在合作品質上中加合作達到了較高的水準。

3.4 論文解讀

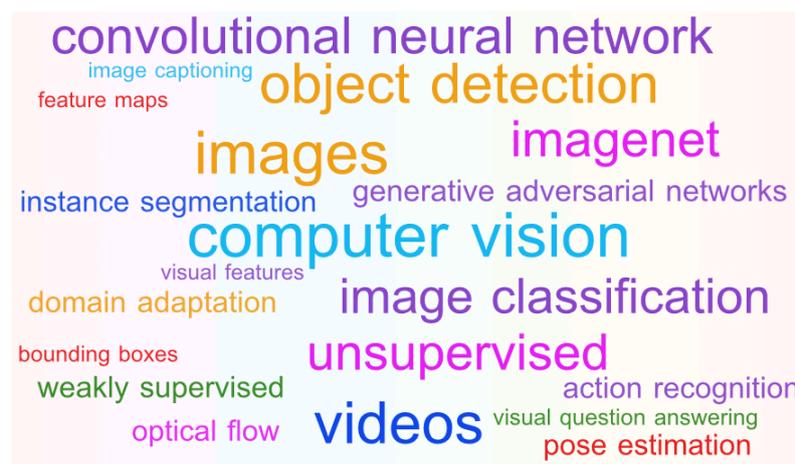
本節對本領域的高水準學術會議論文進行挖掘，解讀這些會議在 2018-2019 年的部分代表性工作。會議具體包括：

IEEE Conference on Computer Vision and Pattern Recognition

IEEE International Conference on Computer Vision

European Conference on Computer Vision

我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖。其中，電腦視覺 (computer vision)、圖像 (images)、視頻 (videos) 是本領域中最熱的關鍵字。



論文題目：*Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*

中文題目：具有空洞分離卷積的編碼-解碼器用於語義圖像分割

論文作者：Liang-Chieh Chen , Yukun Zhu , George Papandreou , Florian Schroff , Hartwig Adam

論文出處 : Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.

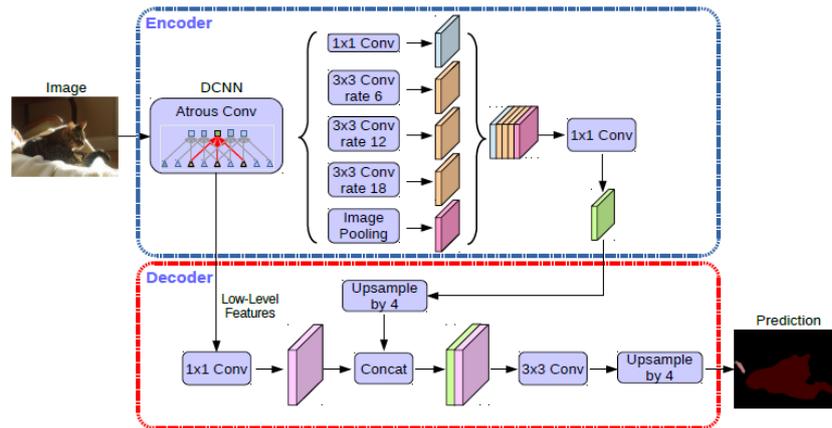
論文地址 : https://link.springer.com/chapter/10.1007%2F978-3-030-01234-2_49

研究問題 :

語義分割是電腦視覺中一項基本且重要的研究內容,它是為圖像中的每個圖元分配語義標籤。在深度學習語義分割任務中經常會使用空間金字塔池化和編碼-解碼器結構。空間金字塔池化可以通過不同解析度的池化特徵捕捉豐富的上下文資訊,但網路中具有步進操作的池化或卷積會導致與物件邊界有關的詳細資訊丟失。這可以通過空洞卷積提取更密集的特徵圖來緩解,但大大增加了計算資源的消耗。而編碼-解碼器結構則可以通過逐漸恢復空間資訊來捕獲更清晰的物件邊界。通過組合兩種方法的優點,提出新的模型—DeepLabv3+。

研究方法 :

如下圖是 DeepLabv3+ 的網路結構,通過添加一個簡單但有效的解碼器模組來優化分割結果,尤其是對象邊界的分割結果,擴展了 DeepLabv3。編碼器模組 (DeepLabv3) 通過在多個尺度上應用空洞卷積,編碼多尺度上下文資訊。空洞卷積可以明確控制由深度卷積神經網路所提特徵的解析度,並調整濾波器的感受野以捕獲多尺度資訊。而簡單而有效的解碼器模組則沿物件邊界調整分割結果。為了進一步提高模型的性能和速度,將深度分離卷積應用於 ASPP (空洞空間金字塔池化) 和解碼器模組。深度分離卷積將傳統的卷積分解為一個深度卷積和一個 1×1 的逐點卷積,在深度卷積操作時應用膨脹率不同的空洞卷積,以獲取不同的尺度資訊。



研究結果：

以用 ImageNet-1k 預訓練的 ResNet-101 和修改的對齊 Xception(更多的層、步進深度分離卷積替代最大池化、額外的 BN 和 ReLU) 為骨架網路，通過空洞卷積提取稠密特徵。在 PASCAL VOC 2012 和 Cityscapes 資料集上證明了 DeepLabv3+ 的有效性和先進性，無需任何後處理即可實現 89% 和 82.1% 的測試集性能。但是對非常相近的物體 (例如椅子和沙發)、嚴重遮擋的物體和視野極小的物體較難進行分割。

論文題目： *MobileNetV2: Inverted Residuals and Linear Bottlenecks*

中文題目：MobileNetV2: 反向殘差和線性瓶頸

論文作者：Sandler Mark , Howard Andrew , Zhu Menglong , Zhmoginov Andrey , Chen Liang-Chieh

論文出處：2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018

論文地址：<https://ieeexplore.ieee.org/document/8578572>

研究問題：

在眾多電腦視覺領域中，深度神經網路正扮演越來越重要的角色。但是優秀性能的獲得通常是以高昂計算資源為代價的，從而大大限制了在計算資源嚴重受限的移動端或嵌入式設備中使用。因此輕量化網路的研究在近期收到了大量關注，本文提出了一種新的移動端輕量化模型—MobileNetV2，在保持相同精度的同時顯著減少了所需的操作和記憶體需求，關鍵是設計了具有線性瓶頸的反向殘差模組。將上述模型應用於移動端目標檢測，介紹了一種有效的方法—SSDLite。此外，通過簡化的 DeepLabv3 構建移動端語義分割模型—Mobile DeepLabv3。

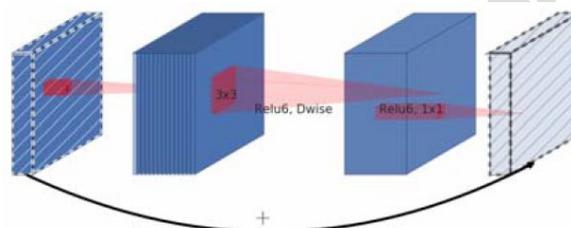
研究方法：

MobileNetV2 的關鍵是具有線性瓶頸的反向殘差模組，該模組以低維壓縮表示作為輸入，首先將其擴張到高維，然後使用羽量級的深度卷積進行過濾，最後使用線性卷積將特徵投影回低維表示。其包含兩個主要的技術：深度分離卷積和殘差模組。

深度分離卷積是很多有效的神經網路結構中關鍵的組成部分，其基本思想是將傳統卷積分解為兩部分：第一層稱為深度卷積，它通過對每個輸入通道應用單個卷積濾波器來執行輕量化濾波；第二層是 1×1 卷積，稱為逐點卷積，它通過計算輸入通道的線性組合來構建新特徵。深度分離卷積的計算量相對於傳統卷積減少了大約 k^2 (k 是卷積核大小)，但是性能只有極小的降低。

我們可以認為深度神經網路中任意層的啟動組成一個“感興趣流形”，它可以嵌入到低維子空間中。也就是說，深度卷積層中所有單個通道的圖元，其中編碼的資訊實際上位於某種流形中，而該流形可以嵌入到低維子空間中。通過分析作者得到兩個屬性：(1) 如果感興趣流形在 ReLU 變換後仍保持非零值，則對

應於線性變換；(2) ReLU 能夠保留輸入流形的完整資訊，但前提是輸入流形位於輸入空間的一個低維子空間中。基於以上兩個觀點，幫助作者優化現有的神經網路結構：假設感興趣流形是低維的，可以通過向卷積塊插入線性瓶頸獲得該流形，即本文核心具有線性瓶頸的反向殘差模組，其結構如下圖所示。先使用逐點卷積擴大通道數+ReLU 啟動，然後使用逐深度卷積提取特徵+ReLU 啟動，最後使用逐點卷積降低通道數+線性啟動，並且使用了 shortcut 連接。



研究結果：

研究者首先通過實驗驗證了反向殘差連接和線性瓶頸的有效性，然後在圖像分類、目標檢測和語義分割三個任務上證明了本文網路結構的先進性。ImageNet 圖像分類任務上 MobileNetV2 的 Top1 最好可達 74.7，優於 MobileNetV1、ShuffleNet 和 NASNet-A。在目標檢測任務上，MNetV2+SSDLite 與 MNetV1+SSDLite 的 mAP 很接近，但參數量和計算時間都明顯減少。在語義分割任務上保持較好性能的同時減少了參數量和計算資源的消耗。

論文題目：*The Unreasonable Effectiveness of Deep Features as a Perceptual Metric*

中文題目：深度特徵作為感知度量的有效性

論文作者：Zhang Richard，Isola Phillip，Efros Alexei A.，Shechtman Eli，Wang Oliver

論文出處：2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018

論文地址：<https://ieeexplore.ieee.org/document/8578166>

研究方法：

對於人類來說，評估兩幅圖像之間的感知相似度幾乎是毫不費力且快速的，但其潛在過程卻被認為是相當複雜的。視覺模式是高維且高度相關的，視覺相似性的概念十分主觀。例如在圖像壓縮領域，壓縮圖像是為了人類看來與原始圖像沒有很大區別，而不關注它們在圖元值上可能有很大的差別。當今最廣泛使用的、傳統的基於圖元值的度量方法（例如 L2 歐式距離、PSNR）或感知距離度量（如 SSIM、MSSIM 等）是簡單的淺層函數，無法解決人類感知的許多細微差別，一個最典型的例子就是模糊會造成圖像在感知上的很大不同，但是在 L2 範數上卻差別不大。如下圖所示，傳統的評價指標與人類的感知判斷是完全相反的。近期深度學習社區發現，將在 ImageNet 分類中訓練的 VGG 網路模型所提取的深度特徵，用作圖像合成的訓練損失是非常有用，一般將這種損失稱為“感知損失”（perceptual losses）。但是這些感知損失的作用有多大？哪些要素對其成功至關重要？本文研究者們嘗試探討了這些問題。



研究方法：

為了研究將深度神經網路提取的深度特徵作為感知損失的有效性，本文研究者們構造了一個人類感知相似性判斷的新資料集——Berkeley-Adobe Perceptual Patch Similarity Dataset (BAPPS 資料集)。該資料集包括 484K 個人類判斷，具有大量傳統失真，如對比度、飽和度和雜訊等；還有基於 CNN 模型的失真，例如自編碼、降噪等造成的失真；以及一些真實演算法的失真，如超解析度重建、去模糊等真實應用。

論文用如下公式計算在給到一個網路 F 時，參考和失真圖像塊的距離。首先提取特徵，然後將通道維度的啟動歸一化，用向量 w 縮放每個通道，並採用 2 距離。最後對空間維度的所有層取平均。

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \square (\hat{y}'_{hw} - \hat{y}'_{0hw}) \right\|_2^2$$

研究結果：

作者進行了大量的實驗，系統地評估了不同網路結構和任務中的深度特徵，並將它們與經典指標進行比較，發現深度特徵是一種非常好的感知度量指標。更令人驚訝的是，該結果不僅限於 ImageNet 訓練的 VGG 提取的深度特徵，而且還適用於不同的深度網路結構和不同的訓練方式（監督，自監督，甚至無監督）。

論文題目： *Residual Dense Network for Image Super-Resolution*

中文題目：基於殘差密集網路的圖像超解析度重建

論文作者：Yulun Zhang，Yapeng Tian，Yu Kong，Bineng Zhong，Yun Fu

論文出處：2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018

論文地址：<https://ieeexplore.ieee.org/document/8578360>

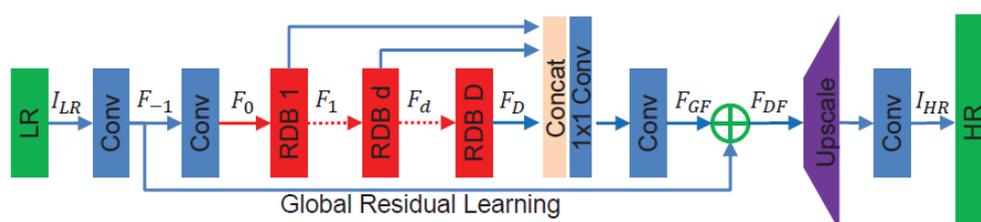
研究內容：

單幅圖像超解析度 (SISR) 旨在通過其退化的低解析度 (LR) 觀測結果生成視覺上令人愉悅的高解析度 (HR) 圖像。最近，深度卷積神經網路在圖像超解析度重建方面取得了巨大的成功，網路的不斷加深使模型能提供豐富分層特徵，圖像中的目標具有不同的比例、視角和寬高比，來自非常深的網路的分層特徵能為重建提供更多線索。但是，大多數基於卷積神經網路的深層圖像超解析度模型都沒有充分利用原始低解析度 (LR) 圖像中的分層特徵，以致獲得了相對較低的性能。在本文中，研究者提出了一種新穎的殘差密集網路 (RDN) 來解決圖像超解析度中的上述問題，使模型能充分利用所有卷積層提取的分層特徵。

研究方法：

如下圖是殘差密集網路 RDN，主要包含四部分：淺層特徵提取網路 (SFEnet)、殘差密集塊 (RDBs)、密集特徵融合 (DFF) 和上採樣網路 (UPNet)。一個非常深的網路直接提取 LR 空間中每個卷積層的輸出是困難且不切實際的，所以使用殘差密集塊 (RDB) 作為 RDN 的構建模組。RDB 由密集連接層和具有局部殘差學習能力的局部特徵融合 (LFF) 組成。RDB 還支援 RDB 之間的連續存儲，一個 RDB 的輸出可以直接訪問下一個 RDB 中每一層，形成連續的狀態傳遞。RDB 中的每個卷積層都可以訪問所有後續層，並傳遞需要保留的資訊。局部特徵融合將先前的 RDB 和當前 RDB 中所有先前層的狀態連接在一起，通過自我調整保留資訊來提取局部密集特徵。LFF 通過更高的增長率來穩定更寬網路的訓練。在提取多層局部密集特徵後，進一步進行全域特徵融合 (GFF)，以全域方

式自我調整地保留分層特徵。在 RDN 中每個卷積層卷積核大小為 3×3 ，局部和全域特徵融合卷積核大小為 1×1 。在上採樣部分使用 ESPCNN 提升圖像的解析度。



研究結果：

使用 DIV2K 資料集中全部的 800 幅訓練圖像訓練模型，測試選用 5 個標準基準資料集：Set5、Set14、B100、Urban 和 Manga109。為了全面地說明所提方法的有效性，類比了三種圖像退化過程：（1）雙三次下採樣（BI）；（2）高斯核模糊 HR 圖像，再下採樣（BD）；（3）先雙三次下採樣，再加入高斯雜訊（DN）。作者進行了大量的實驗發現：（1）RDB 數量或 RDB 中卷積層數量越多，模型性能越好；增長率越大也會獲得更好的性能。當上述模組使用數量較少時 RDN 依然比 SRCNN 性能好。（2）進行了消融實驗，驗證了所提模型中連續存儲、局部殘差學習和全域特徵融合的有效性。（3）在三種退化模型上與六種先進的模型進行了對比：SRCNN、LapSRN、DRNN、SRDenseNet、MemNet 和 MDSR。在不同比例因數、退化模型和資料集中，RDN 都表現出了相近甚至更好的性能。

論文題目：*ShuffleNet V2: Practical guidelines for efficient cnn architecture design*

中文題目：ShuffleNet V2: 高效 CNN 網路結構設計實用指南

論文作者：Ma Ningning，Zhang Xiangyu，Zheng Hai-Tao，Sun Jian

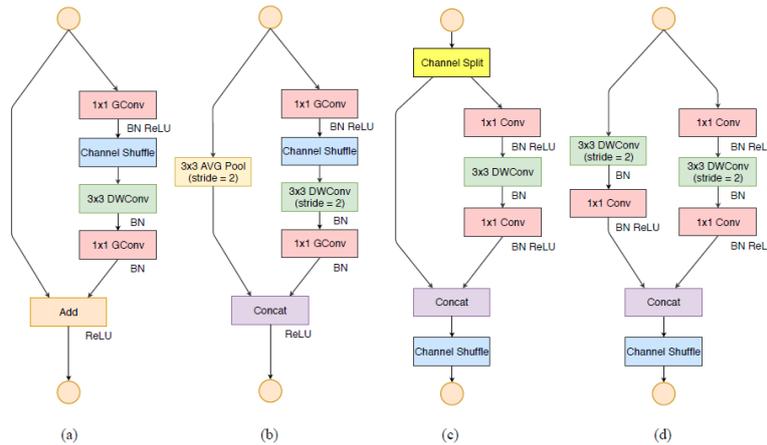
論文出處：Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) , v 11218 LNCS, p 122-138, 2018, Computer Vision - ECCV 2018 - 15th European Conference, 2018, Proceedings

論文連結：https://link.springer.com/chapter/10.1007%2F978-3-030-01264-9_8

研究內容：

自 AlexNet 之後，ImageNet 圖像分類準確率被很多新的網路結構如 ResNet 和 DenseNet 等不斷提高，但是除準確率外，計算複雜度也是 CNN 網路需要考慮的重要指標。實際任務通常是要在有限的計算資源下獲得最佳的精度，過複雜的網路由於速度原因難以在移動端等設備中應用。為此，研究者們提出了很多輕量化的 CNN 網路如 MobileNet 和 ShuffleNet 等，在速度和準確度之間做了較好地平衡。以往的移動端 CNN 網路結構設計在考慮計算複雜度時，直接致力於優化整體網路計算所需的 FLOPs，並沒有考慮真正關心的速度和延遲，且具有相似 FLOPs 的網路速度也是不同的。像記憶體訪問開銷 (MAC)、計算平臺等也是需要考慮的方面。為了實際需求，本文研究者不局限于追求理論 FLOPs 的減少，從更直接的角度為輕量化網路設計提供指導意見。

研究方法：



作者建議有效的網路結構設計應考慮兩個原則。首先，應使用直接指標（例如速度）代替間接指標（例如 FLOP）。其次，應在目標平臺上評估此類指標。通過對兩個代表性最新網路的分析，作者得出了關於有效網路設計的四項準則：

(1) 卷積層的輸入和輸出特徵通道數相等時 MAC 最小；(2) 過多的組卷積會增大 MAC；(3) 網路碎片化會降低並行度；(4) 元素級的操作 (element-wise) 會增加時間消耗。遵循以上準則提出了一個更有效的網路結構——ShuffleNet V2。

下圖是 ShuffleNet V1 (圖中 a 和 b) 和 ShuffleNet V2 (圖中 c 和 d) 組成模組的對比。對比 (a) 和 (b)，ShuffleNet V2 首先用 Channel Split 操作將輸入按通道分成兩部分，一部分直接向下傳遞，另外一部分則用於計算；然後棄用了 1x1 的組卷積，將通道混洗操作 (Channel Shuffle) 移到了最後，並將前面的 Add 操作用 Concat 代替。

研究結果：

論文進行了大量的實驗，與 MobileNet V1/V2、ShuffleNet V1、DenseNet、Xception、IGCV3-D、NASNet-A 等模型在速度、精度、FLOPs 上進行了詳細的對比。實驗中不少結果都和前面幾點發現吻合，ShuffleNet V2 在準確率和速度方面達到了很好的平衡。

論文題目：*A Theory of Fermat Paths for Non-Line-of-Sight Shape Reconstruction*

中文題目：非視距形狀重建的費馬路徑理論

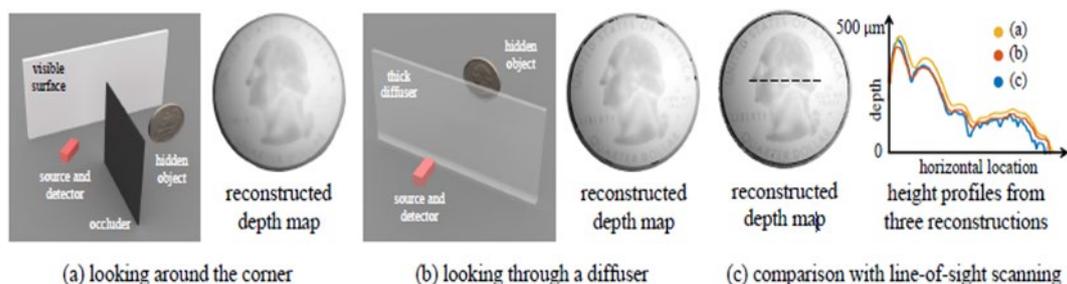
論文作者：Shumian Xin, Sotiris Nousias, Kiriakos N. Kutulakos, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan, and Ioannis Gkioulekas.

論文出處：CVPR 2019: IEEE Conference on Computer Vision and Pattern Recognition.

論文地址：<https://www.ri.cmu.edu/wp-content/uploads/2019/05/cvpr2019.pdf>

研究問題：

很多時候攝像頭可能無法拍攝全部場景或物體，例如，面對攝像機的物體背面，拐角處的物體或通過漫射器觀察到的物體。非視距(non-line-of-sight, NLOS) 成像對於許多安全保護應用至關重要。一些傳統方法通過分析隱藏場景投射陰影的細微本影和半影，以估計粗糙的運動和結構，或使用光的相干特性來定位隱藏的物件，但很難重建任意隱藏場景的 3D 形狀。基於主動照明的瞬態 NLOS 成像大多採用快速調製光源和時間分辨感測器，但現有的 SPAD 強度估計不理想，而且重建 NLOS 物件的朗伯反射率假設。作者使用 NLOS 瞬態測量得出幾何約束而非強度約束的方法來克服上述限制。



上圖為非視距成像示例：被遮光板遮擋 (a) 和被漫射板遮擋 (b) 的物體表面重建結果與視距掃描結果 (c) 對比。

研究方法：

作者提出了一個新的光費馬路徑 (Fermat path) 理論，即光在已知的可見場景和不處於瞬態相機視線範圍內的未知物體之間，這些光要麼從鏡面反射，要麼被物體的邊界反射，從而編碼了隱藏物體的形狀。作者證明，費馬路徑對應於瞬態測量中的不連續性，間斷點的位置僅與 NLOS 物件的形狀有關，與其反射率 (BRDF) 無關。並推導出一個新的約束條件，它將這些不連續處的路徑長度的空間導數與曲面的曲率相關聯。基於此理論，作者提出了一種稱為費馬流 (Fermat Flow) 的演算法，用於估計非視距物體的形狀。其關鍵在於，費馬路徑長度的空間導數可唯一確定隱藏場景點的深度和法線，再擬合和估算平滑路徑長度函數，進一步結合深度和法線獲得光滑的網格，從而精確恢復了對複雜物件 (從漫反射到鏡面反射) 形狀，範圍從隱藏在拐角處以及隱藏在漫射器後面的漫反射到鏡面反射。最後，該方法與用於瞬態成像的特定技術無關。

研究結果：

作者使用了一些不同 BRDF 的凹凸幾何形狀的日常物品，包括半透明 (塑膠壺)，光滑 (碗，花瓶)，粗糙鏡面 (水壺) 和光滑鏡面 (球形) 等。分別開展了使用 SPAD 和超快鐳射從皮秒級瞬態中恢復毫米級形狀，以及使用干涉法實現從飛秒級瞬態中恢復毫米級形狀的兩種實驗，實驗結果顯示重建細節與 groundtruth 形狀非常吻合。

論文題目：*Implicit 3D Orientation Learning for 6D Object Detection from RGB Images*

中文題目：從 RGB 圖像檢測 6 維位元姿的隱式三維朝向學習

論文作者：Martin Sundermeyer, Zoltan-Csaba Marton, Maximilian Durner, Rudolph Triebel

論文出處：ECCV 2018: European Conference on Computer Vision.

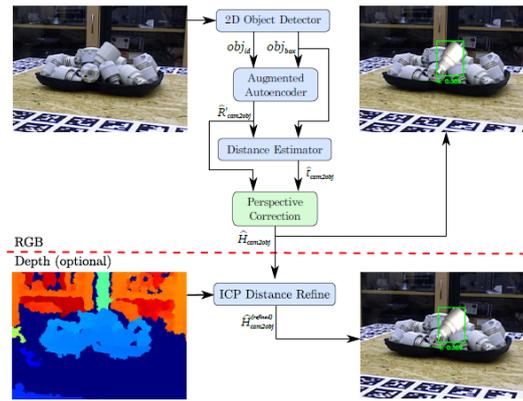
論文地址：

http://openaccess.thecvf.com/content_ECCV_2018/papers/Martin_Sundermeyer_Implicit_3D_Orientation_ECCV_2018_paper.pdf

研究問題：

對於諸如移動機器人控制和增強現實之類的應用而言，現代電腦視覺系統中最重要元件之一就是可靠且快速的 6D 目標檢測模組。至今尚無通用，易於應用，強大且快速的解決方案。原因是多方面的：首先，當前的解決方案通常不足以有效處理典型的挑戰；其次，現有方法通常需要某些目標屬性。而且，當前的方法在執行時間以及所需帶標注的訓練資料的數量和種類方面效率都不高。作者提出對單個 RGB 圖像進行操作，可在很大程度上不需要深度資訊，顯著增加可用性。

研究方法：



上圖為 6D 目標檢測管道具有齊次座標變換 H_{cam2obj} (右上) 和深度細化結果 $H_{\text{cam2obj}}^{(\text{refined})}$ (右下)。作者提出了一種基於 RGB 的即時目標檢測和 6D 姿態估計流程。首先使用 SSD (Single Shot Multibox Detector) 來提供目標邊界框和識別字。其次，在此基礎上，採用新穎的 3D 方向估計演算法，該演算法基於之前的降噪自動編碼器 (Denoising Autoencoder) 的通用版本，增強型自動編碼器 (AAE)。AAE 使用一種新穎的域隨機化策略，模型學到的並不是從輸入圖像到物體位元姿的顯式映射，而是會根據圖像樣本在隱含空間內建立一個隱式的物體位元姿表徵。因而，訓練獨立於目標方向的具體表示 (例如四元數)，避免從圖像到方向的一對多映射，由此 AAE 可處理由對稱視圖引起的模糊姿態。另外學習專門編碼 3D 方向的表徵，同時實現對遮擋，雜亂背景的魯棒性，並可推廣到對不同環境和測試感測器。而且，AAE 不需要任何真實的姿勢標注訓練資料。相反，它被訓練為以自我監督的方式編碼 3D 模型視圖，克服了對大型姿勢標注資料集的需要。下圖為 AAE 訓練過程。

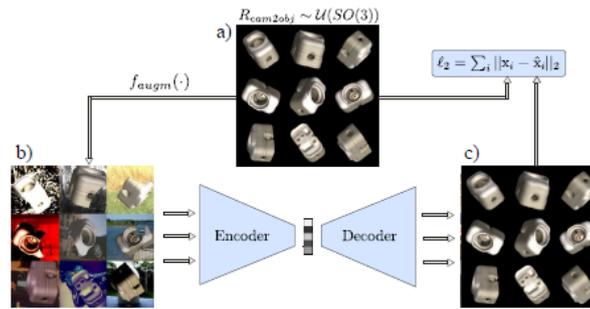


Fig. 4: Training process for the AAE; a) reconstruction target batch \mathbf{x} of uniformly sampled $SO(3)$ object views; b) geometric and color augmented input; c) reconstruction $\hat{\mathbf{x}}$ after 30000 iterations

研究結果：

作者在 T-LESS 和 LineMOD 資料集上評估了 AAE 和整個 6D 檢測管道，僅包括 2D 檢測，3D 方向估計和投影距離估計。與最先進的深度學習方法相比，AAE 準確性更好，同時效率更高。另外，作者也分析了一些失敗案例，主要源於檢測失敗或強遮擋。

論文題目：*SinGAN: Learning a Generative Model from a Single Natural Image*

中文題目：SinGAN:從單張圖像學習生成模型

論文作者：Tamar Rott Shaham ,Technion Tali Dekel ,Google Research ,Tomer Michaeli ,Technion

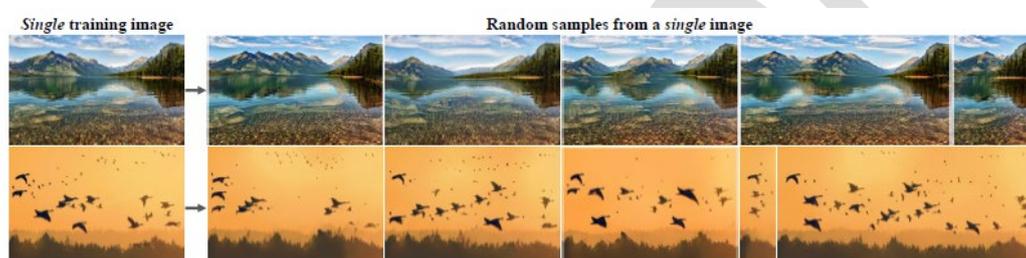
論文出處：ICCV 2019 : IEEE International Conference on Computer Vision.

論文地址：<https://arxiv.org/pdf/1905.01164.pdf>

研究問題：

生成對抗網路 (Generative Adversarial Nets, GAN) 在類比視覺資料的高維分佈方面取得了巨大飛躍。特別是用特定類別的資料集 (如人臉、臥室) 進行訓練時，非條件 GAN 在生成逼真的、高品質的樣本方面取得了顯著成功。但對高度

多樣化、多種類別的資料集 (如 ImageNet) 的模擬仍然是一項重大挑戰，而且通常需要根據另一輸入信號來調整生成或為特定任務訓練模型。對單個自然圖像中各種圖像塊的內部分佈進行建模已被公認為是許多電腦視覺任務的有用先驗。作者將 GAN 帶入到一個新領域—從單個自然圖像中學習非條件生成模型。單個自然圖像通常具有足夠的內部統計資訊，可學習到強大的生成模型，而不必依賴某個相同類別的資料集。為此，作者提出了一個新的單圖像生成模型 SinGAN，能夠處理包含複雜結構和紋理的普通自然圖像的神經網路。



相對於左邊的原始圖像，SinGAN 生成新的逼真的圖像樣本，該樣本在創建新的物件配置和結構的同時保留原始圖像塊分佈。

研究方法：

作者的目標是學習一個非條件生成模型，該模型可捕獲單個訓練圖像的內部統計資料。此任務在概念上與常規 GAN 設置相似，不同之處在於，訓練樣本是單個圖像的多尺度的圖像塊，而非整個圖像樣本。為此，SinGAN 生成框架由具有層級結構的 patch-GANs (瑪律可夫判別器) 組成，其中每個判別器負責捕獲不同尺度的分佈，這是第一個為從單個圖像進行內部學習而探索的網路結構。圖像樣本從最粗尺度開始，然後依次通過所有的生成器直到最細尺度，且每個尺度都注入雜訊。所有生成器和判別器具有相同的感受野，因此，隨著生成過程推進可

以捕獲更細尺寸的結構。在訓練時，對抗損失採用 WGAN-GP 損失，以增加訓練穩定性。並設計了一種重建損失來確保可以生成原始圖像的特定雜訊圖譜集合。

研究結果：

作者在圖像場景跨度很大的資料集上進行了測試。直觀上，SinGAN 很好地保留目標的全域結構和紋理資訊，很真實地合成了反射和陰影效果。再使用 AMT 真假用戶調研和 FID 的單幅圖像版本進行量化。AMT 測試結果表明可以生成很真實的樣本，對於細節保留的也更多，人類判別的混淆率較高。FID 結果與 AMT 一致。

3.5 電腦視覺進展

近年來，巨量資料的不斷湧現與計算能力的快速提升，給以非結構化視覺資料為研究物件的電腦視覺帶來了巨大的發展機遇與挑戰性難題，電腦視覺也因此成為學術界和工業界公認的前瞻性研究領域，部分研究成果已實際應用，催生出人臉識別、智慧視頻監控等多個極具顯示度的商業化應用。

電腦視覺的研究目標是使電腦具備人類的視覺能力，能看懂圖像內容、理解動態場景，期望電腦能自動提取圖像、視頻等視覺資料中蘊含的層次化語義概念及多語義概念間的時空關聯等。電腦視覺領域不斷湧現出很多激動人心的研究成果，例如，人臉識別、物體識別與分類等方面的性能已接近甚至超過人類視覺系統。本文根據近兩年電腦視覺領域頂級會議最佳論文及高引論文，對該領域中的技術現狀和研究前沿進行了綜合分析。

近兩年大多數研究都集中在深度學習、檢測和分類以及面部/手勢/姿勢、3D傳感技術等方面。隨著電腦視覺研究的不斷推進，研究人員開始挑戰更加困難的電腦視覺問題，例如，圖像描述、事件推理、場景理解等。單純從圖像或視頻出發很難解決更加複雜的圖像理解任務，一個重要的趨勢是多學科的融合，例如，融合自然語言處理領域的技術來完成圖像描述的任務。圖像描述是一個融合電腦視覺、自然語言處理和機器學習的綜合問題，其目標是翻譯一幅圖片為一段描述文字。目前主流框架為基於遞迴神經網路的編碼器解碼器結構其核心思想類似於自然語言機器翻譯。但是，由於遞迴網路不易提取輸入圖像和文本的空間以及層次化約束關係，層次化的卷積神經網路以及啟發自認知模型的注意力機制受到關注。如何進一步從認知等多學科汲取知識，構建多模態多層次的描述模型是當前圖像描述問題研究的重點。

事件推理目標是識別複雜視頻中的事件類別並對其因果關係進行合理的推理和預測。與一般視頻分析相比，其難點在於事件視頻更加複雜，更加多樣化，而最終目標也更具挑戰性。不同於大規模圖像識別任務，事件推理任務受限於訓練資料的規模，還無法構建端到端的事件推理系統。目前主要使用圖像深度網路作為視頻的特徵提取器，利用多模態特徵融合模型，並利用記憶網路的推理能力，實現對事件的識別和推理認知。當前研究起源於視頻的識別和檢測，其方法並未充分考慮事件資料的複雜和多樣性。如何利用視頻資料豐富的時空關係以及事件之間的語義相關性，應是今後的關注重點。

場景理解的目的是電腦視覺系統通過分析處理自身所配置的感測器採集的環境感知資料，獲得周圍場景的幾何/拓撲結構、組成要素（人、車及物體等）及其時空變化，並進行語義推理，形成行為決策與運動控制的時間、空間約束。近

年來，場景理解已經從一個初期難以實現的目標成為目前幾乎所有先進電腦視覺系統正在不斷尋求新突破的重要研究方向。利用社會-長短記憶網路 (Social-LSTM) 實現多個行人之間的狀態聯繫建模，結合各自運動歷史狀態，決策出未來時間內的運動走向。此外神經網路壓縮方向也是目前深度學習研究的一個熱門的方向，其主要的研究技術有壓縮，蒸餾，網路架構搜索，量化等。

綜上所述，視覺的發展需要設計新的模型，它們需要能考慮到空間和時間資訊；弱監督訓練如果能做出好的結果，下一步就是自監督學習；需要高品質的人類檢測和視頻物件檢測資料集；結合文本和聲音的跨模態集成；在與世界的交互中學習。

4 知識工程

4.1 知識工程概念

1994 年圖靈獎獲得者、知識工程的建立者費根鮑姆給出知識工程定義——將知識集成到電腦系統從而完成只有特定領域專家才能完成的複雜任務。在大資料時代，知識工程是從大資料中自動或半自動獲取知識，建立基於知識的系統，以提供互聯網智慧知識服務。大資料對智慧服務的需求，已經從單純的搜集獲取資訊，轉變為自動化的知識服務。我們需要利用知識工程為大資料添加語義/知識，使資料產生智慧（Smart Data），完成從資料到資訊到知識，最終到智慧應用的轉變過程，從而實現對大資料的洞察、提供使用者關心問題的答案、為決策提供支援、改進使用者體驗等目標。知識圖譜在以下應用中已經凸顯出越來越重要的應用價值：

- 知識融合：當前互聯網大資料具有分佈異構的特點，通過知識圖譜可以對這些資料資源進行語義標注和連結，建立以知識為中心的資源語義集成服務；
- 語義搜索和推薦：知識圖譜可以將使用者搜索輸入的關鍵字，映射為知識圖譜中客觀世界的概念和實體，搜索結果直接顯示出滿足使用者需求的結構化資訊內容，而不是互聯網網頁；
- 問答和對話系統：基於知識的問答系統將知識圖譜看成一個大規模知識庫，通過理解將用戶的問題轉化為對知識圖譜的查詢，直接得到用戶關心問題的答案；

-
- 大資料分析與決策：知識圖譜通過語義連結可以說明理解大資料，獲得對大資料的洞察，提供決策支援。

我們根據知識工程生命週期各個階段的關鍵技術，利用 AMiner 中近年來知識圖譜領域的高水準學術論文，挖掘出了包括知識表示 (knowledge representation) 知識獲取(knowledge acquisition) 知識推理(knowledge reasoning) 知識集成 (knowledge integration) 和知識存儲 (knowledge storage) 等相關關鍵字近年來全球活躍的學術研究。此外，結合知識圖譜技術，本報告將以上研究領域表示為三級圖譜結構，具體分析和處理的方法如下：

1. 使用自然語言處理技術，提取每篇論文文獻的關鍵字，據此，結合學科領域知識圖譜，將文章分配到相應領域；
2. 依據學科領域對論文進行聚類，並統計論文數量作為領域的研究熱度；
3. 領域專家按照領域層級對學科領域劃分等級，設計了三級圖譜結構，最後根據概念熱度定義當前研究熱點。

知識工程三級知識圖譜的詳細資料可以參見本報告附錄，或到 <https://www.aminer.cn/data> 中直接下載原始資料。鑒於自動分析技術和論文採集的限制性，圖譜還可以進一步完善，歡迎讀者批評指正，我們會根據讀者的回饋定期更新。

4.2 知識工程發展歷史

回顧知識工程四十年多來發展歷程，總結知識工程的演進過程和技術進展，可以將知識工程分成五個標誌性的階段，前知識工程時期、專家系統時期、萬維網 1.0 時期，群體智慧時期以及知識圖譜時期，如下圖所示。

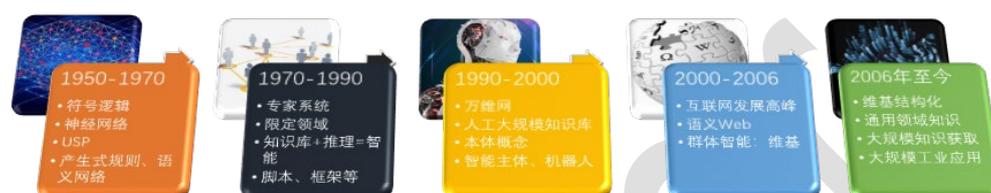


圖 4-1 知識工程發展歷程

● 1950-1970 時期：圖靈測試——知識工程誕生前期

人工智慧旨在讓機器能夠像人一樣解決複雜問題，圖靈測試是評測智慧的手段。這一階段主要有兩個方法：符號主義和連結主義。符號主義認為物理符號系統是智慧行為的充要條件，連結主義則認為大腦（神經元及其連接機制）是一切智慧活動的基礎。這一階段具有代表性的工作是通用問題求解程式（GPS）：將問題進行形式化表達，通過搜索，從問題初始狀態，結合規則或表示得到目標狀態。其中最成功應用是博弈論和機器定理證明等。這一時期的知識表示方法主要有邏輯知識表示、產生式規則、語義網路等。這一時代人工智慧和知識工程的先驅 Minsky, McCarthy 和 Newell 以 Simon 四位學者因為他們在感知機、人工智慧語言和通用問題求解和形式化語言方面的傑出工作分別獲得了 1969 年、1971 年、1975 年的圖靈獎。

● 1970-1990 時期：專家系統——知識工程蓬勃發展期

通用問題求解強調利用人的求解問題的能力建立智慧系統，而忽略了知識對智慧的支援，使人工智慧難以在實際應用中發揮作用。70 年開始，人工智慧開始轉向建立基於知識的系統，通過“知識庫+推理機”實現機器智慧，這一時期湧現出很多成功的限定領域專家系統，如 MYCIN 醫療診斷專家系統、識別分子結構的 DENRAL 專家系統以及電腦故障診斷 XCON 專家系統等。斯坦福人工智慧實驗室的奠基人 Feigenbaum 教授在 1980 年的一個專案報告《Knowledge Engineering: The Applied Side of Artificial Intelligence》中提出知識工程的概念，從此確立了知識工程在人工智慧中的核心地位。這一時期知識表示方法有新的演進，包括框架和腳本等。80 年代後期出現了很多專家系統的開發平臺，可以幫助將專家的領域知識轉變成電腦可以處理的知識。

- 1990-2000 時期：萬維網

在 1990 年到 2000 年，出現了很多人工構建大規模知識庫，包括廣泛應用的英文 WordNet，採用一階謂詞邏輯知識表示的 Cyc 常識知識庫，以及中文的 HowNet。Web1.0 萬維網的產生為人們提供了一個開放平臺，使用 HTML 定義文本的內容，通過超連結把文本連接起來，使得大眾可以共用資訊。W3C 提出的可延伸標記語言 XML，實現對互聯網文檔內容的結構通過定義標籤進行標記，為互聯網環境下大規模知識表示和共用奠定了基礎。這一時期在知識表示研究中還提出了本體的知識表示方法。

- 2000-2006 時期：群體智慧

在 2001 年，萬維網發明人、2016 年圖靈獎獲得者 Tim Berners-Lee 在科學美國人雜誌中發表的論文《The Semantic Web》正式提出語義 Web 的概念，旨在對

互聯網內容進行結構化語義表示，利用本體描述互聯網內容的語義結構，通過對網頁進行語義標識得到網頁語義資訊，從而獲得網頁內容的語義資訊，使人和機器能夠更好地協同工作。W3C 進一步提出萬維網上語義標識語言 RDF (資源描述架構) 和 OWL (萬維網本體表述語言) 等描述萬維網內容語義的知識描述規範。

萬維網的出現使得知識從封閉知識走向開放知識，從集中構建知識成為分佈群體智慧知識。原來專家系統是系統內部定義的知識，現在可以實現知識源之間相互連結，可以通過關聯來產生更多的知識而非完全由固定人生產。這個過程中出現了群體智慧，最典型的代表就是維琪百科，實際上是用戶去建立知識，體現了互聯網大眾用戶對知識的貢獻，成為今天大規模結構化知識圖譜的重要基礎。

- 2006 年至今：知識圖譜——知識工程新發展時期

從 2006 年開始，大規模維琪百科類富結構知識資源的出現和網路規模資訊提取方法的進步，使得大規模知識獲取方法取得了巨大進展。與 Cyc、WordNet 和 HowNet 等手工研製的知識庫和本體的開創性項目不同，這一時期知識獲取是自動化的，並且在網路規模下運行。當前自動構建的知識庫已成為語義搜索、大資料分析、智慧推薦和資料集成的強大資產，在大型行業和領域中正在得到廣泛使用。典型的例子是穀歌收購 Freebase 後在 2012 年推出的知識圖譜(Knowledge Graph)，Facebook 的圖譜搜索，Microsoft Satori 以及商業、金融、生命科學等領域特定的知識庫。最具代表性大規模網路知識獲取的工作包括 DBpedia、Freebase、KnowItAll、WikiTaxonomy 和 YAGO，以及 BabelNet、ConceptNet、DeepDive、NELL、Probase、Wikidata、XLORE、Zhishi.me、CNDBpedia 等。這

些知識圖譜遵循 RDF 資料模型，包含數以千萬級或者億級規模的實體，以及數十億或百億事實（即屬性值和與其他實體的關係），並且這些實體被組織在成千上萬的由語義體現的客觀世界的概念結構中。

目前知識圖譜的發展和應用狀況，除了通用的大規模知識圖譜，各行業也在建立行業和領域的知識圖譜，當前知識圖譜的應用包括語義搜索、問答系統與聊天、大資料語義分析以及智慧知識服務等，在智慧客服、商業智慧等真實場景體現出廣泛的應用價值，而更多知識圖譜的創新應用還有待開發。

在我國知識工程領域研究中，中科院系統所陸汝鈞院士、計算所史忠植研究員等老一代知識工程研究學者為中國的知識工程研究和人才培養做出了突出貢獻，例如，陸汝鈞院士因在知識工程和基於知識的軟體工程方面作出的系統和創造性工作，以及在大知識領域的開創性貢獻，榮獲首屆“吳文俊人工智慧最高成就獎”^[6]。

4.3 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為知識工程領域全球學者分佈情況：



圖 4-2 知識工程領域全球學者分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量優勢明顯且主要分佈在其東西海岸；歐洲及亞洲東部也有較多的人才分佈；其他諸如非洲、南美洲等地區的學者非常稀少；知識工程領域的人才分佈與各地區的科技、經濟實力情況大體一致。

此外，在性別比例方面，知識工程領域中男性學者占比 89.7%，女性學者占比 10.6%，男性學者占比遠高於女性學者。

知識工程領域學者的 h-index 分佈如下圖所示，大部分學者的 h-index 分佈在中低區域，其中 h-index 在 20-30 區間的人數最多，有 783 人，占比 38.9%，小於 20 區間的人數最少，有 90 人。

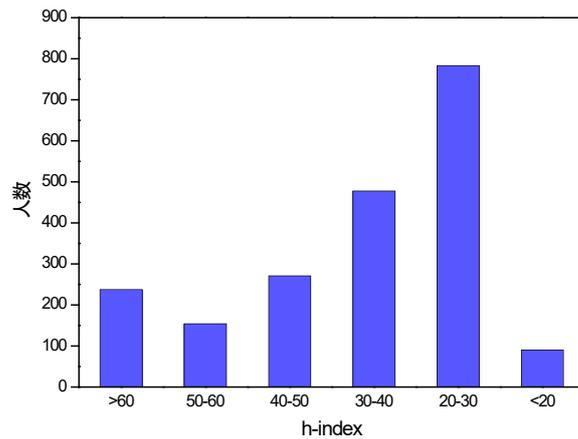


圖 4-3 知識工程領域學者 h-index 分佈

- 中國人才分佈

我國專家學者在知識工程領域的分佈如下圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是珠三角和長三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，通過觀察中國周邊國家的學者數量情況，特別是與日韓、東南亞等亞洲國家相比，中國在知識工程領域學者數量較多。



圖 4-4 知識工程領域中國學者分佈

中國與其他國家在知識工程領域的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 4-1 知識工程領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	541	17306	32	1092
中國-新加坡	116	4107	35	244
中國-澳大利亞	111	3634	33	237
中國-英國	27	352	13	52
中國-加拿大	24	632	26	58

中國-日本	21	572	27	56
中國-丹麥	14	328	23	23
中國-德國	10	344	34	20
中國-印度	10	76	8	22
中國-希臘	10	197	20	20

從上表資料可以看出，中美合作的論文數、引用數、學者數遙遙領先，表明中美間在知識工程領域合作之密切；此外，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 4 席；中國與新加坡合作的論文數雖然不是最多，但是擁有最高的平均引用數說明在合作品質上中國與新加坡合作達到了較高的水準。

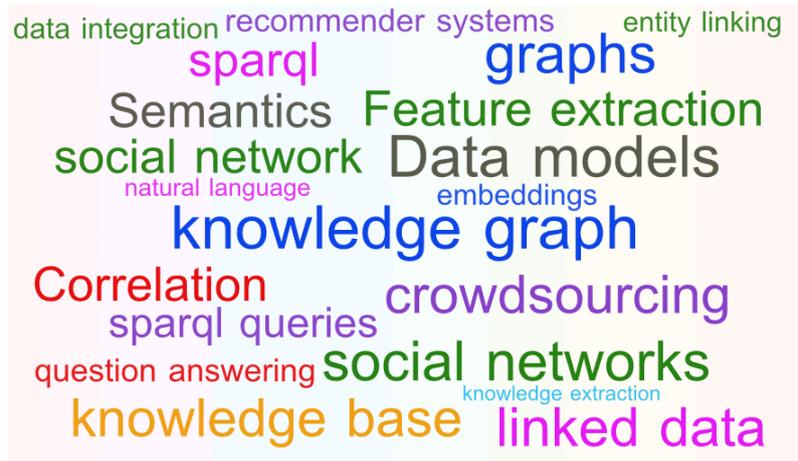
4.4 論文解讀

本節對本領域的高水準學術會議及期刊論文進行挖掘，解讀這些會議和期刊在 2018-2019 年的部分代表性工作。這些會議和期刊包括：

IEEE Transactions on Knowledge and Data Engineering

International Conference on Information and Knowledge Management

我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如下圖所示。其中，知識圖譜 (knowledge graph)、資料模型 (Data models)、社交網路 (social networks) 是本領域中最熱的關鍵字。



論文題目： *Convolutional 2D Knowledge Graph Embeddings*

中文題目： 基於二維卷積的知識圖譜嵌入表示學習

論文作者： Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, Sebastian Riedel

論文出處： The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI 2018)

論文地址：

<https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/download/17366/15884>

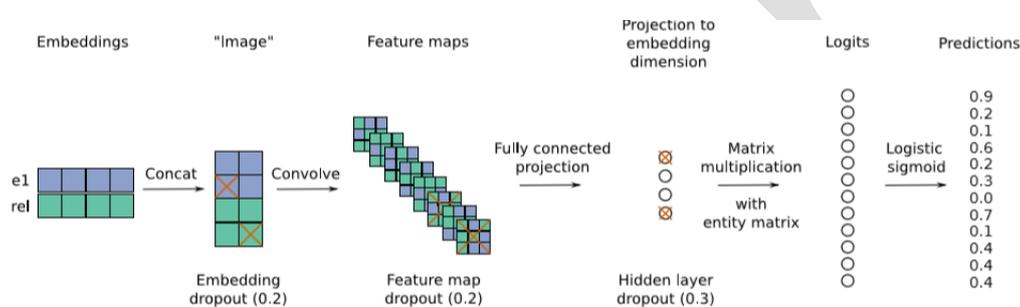
研究問題：

知識圖譜的連結預測任務是預測節點之間潛在的關係。傳統的連結預測方法專注於淺的、快速的模型，因為這樣可以擴展到大規模的 KG 中。但是淺層模型學習到的特徵比深沉模型少很多，大大限制了模型的性能。解決該問題的方法之一是增加 embedding 的維度，但是會增加模型參數量，不方便擴展到大規模 KG 中。此外，部分現有資料集中有測試集洩露問題：訓練集中的三元組稍微翻轉一下就可以得到測試集三元組，然後使用基於規則的模型就能達到最佳性能。文章

通過構造一個簡單的翻轉來衡量這個問題嚴重性，並清洗了部分資料來解決該問題。

研究方法：

文章提出一種多層卷積神經網路模型用於知識圖譜的連結預測任務。與自然語言處理中常用的一維卷積不同，文章通過把多個向量堆疊成矩陣，就可以像圖形一樣用二維卷積核來抽取 embedding 之間的關係。



如上圖所示，評分函數如下公式

$$\psi_r(\mathbf{e}_s, \mathbf{e}_o) = f(\text{vec}(f([\bar{\mathbf{e}}_s; \bar{\mathbf{r}}_r] * \omega)))\mathbf{W})\mathbf{e}_o$$

模型的流程總結為：

- 經過 look-up embedding 得到實體和關係的向量表示，然後通過變形和堆疊轉化為 2D 版本。
- 用多個卷積核對堆疊後的矩陣進行卷積操作，得到一個特徵圖 γ 。
- 把 γ 向量化，然後通過一個全連接層映射到 k 維空間中。
- 最後與目標實體的 embedding 相乘獲得相應的得分。
- 將得分進行 sigmoid 操作得到概率 p ，從而最小化交叉熵來訓練模型。

值得一提的是，與傳統模型對三元組關係 (s, r, o) 打分的 1-1 scoring 模式不同，ConvE 以實體關係對 (s, r) 作為輸入，同時對所有實體 o 進行打分，即 1-N scoring。這種方式極大加快了計算速度。實驗結果表明，即使實體個數擴大 10 倍，計算時間也只是增加了 25%。

研究結果：

作者在 4 個資料集 WN18、FB15K、YAGO3-10、Countries 上進行實驗，與 DisMult、R-GCN 等模型進行了對比。實驗結果表明：0.23M 個參數的 ConvE 就與 1.89M 個參數的 DistMult 有相近的性能表現，總的來說 ConvE 的參數效率是 R-GCN 的 17 倍以上，是 DistMult 的 8 倍以上。此外，作者還發現 ConvE 在 YAGO3-10 和 FB15k-237 上的表現比在 WN18RR 上好，因為前兩者包含入度很大的結點，比如結點 United States 在“was born in”上的入度超過 10000，這種複雜的 KG 需要 deeper 模型，而淺層模型比如 DistMult 則在較簡單的 KG 上有優勢。

Model	Param. count	Emb. size	MRR	Hits		
				@10	@3	@1
DistMult	1.89M	128	.23	.41	.25	.15
DistMult	0.95M	64	.22	.39	.25	.14
DistMult	0.23M	16	.16	.31	.17	.09
ConvE	5.05M	200	.32	.49	.35	.23
ConvE	1.89M	96	.32	.49	.35	.23
ConvE	0.95M	54	.30	.46	.33	.22
ConvE	0.46M	28	.28	.43	.30	.20
ConvE	0.23M	14	.26	.40	.28	.19

論文題目：*Explainable Reasoning over Knowledge Graphs for Recommendation*

中文題目：基於知識圖譜路徑推理的可解釋推薦

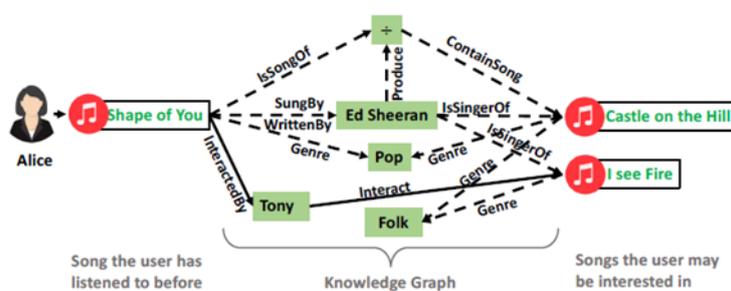
論文作者：Xiang Wang, DingxianWang, Canran Xu, Xiangnan He, Yixin Cao, Tat-Seng Chua

論文出處：Proceedings of the AAAI Conference on Artificial Intelligence. 2019 (AAAI'19).

論文地址：https://www.aaai.org/ojs/index.php/AAAI/article/view/4470/4348

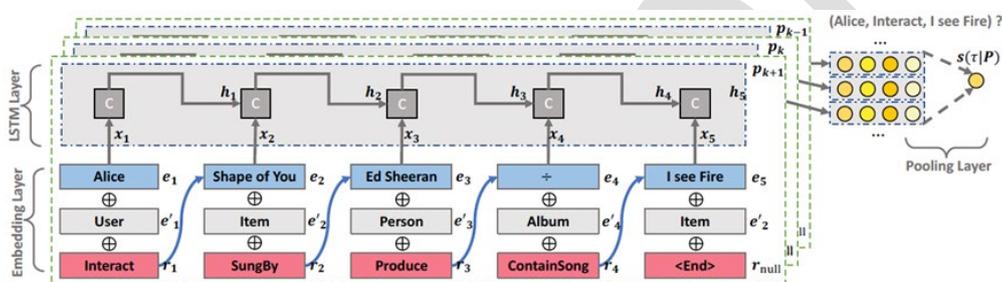
研究問題：

近年來如何將知識圖譜融入推薦系統得到越來越多的關注，通過探索知識圖譜中的使用者到商品的路徑，可以為使用者與商品的交互行為提供豐富的補充資訊。這些路徑不僅揭示了實體和關係的語義，還能幫助理解用戶的興趣。然而現有的模型沒能充分利用路徑來推斷使用者偏好，尤其是在建模的順序依賴關係和路徑的整體語義方面。文章構建了知識感知路徑遞迴網路模型(Knowledge aware Path Recurrent Network , KPRN)，通過組合實體和關係的語義來生成路徑表示。利用路徑中的順序依賴關係，可以基於路徑進行有效推理，從而推斷出用戶-項目交互場景中的基本原理。此外，文章設計了一種新的權重池化操作，以區分用戶與專案連接的不同路徑的優勢，賦予我們的模型一定的可解釋性。下圖為基於知識圖譜的音樂推薦場景實例，虛線為關係，實線為使用者-商品交互路徑。



研究方法：

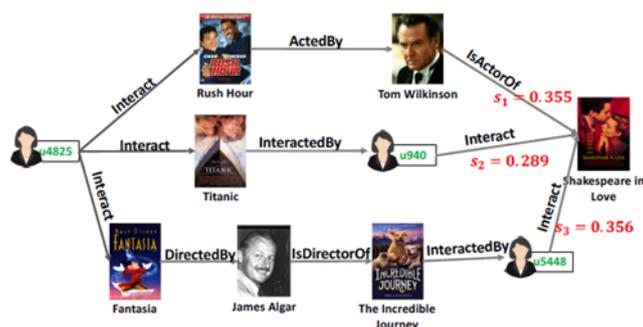
知識圖譜和路徑：知識圖譜由一組三元組 (h, r, t) 組成，代表實體 h 和實體 t 構成關係 r 。文章中知識圖譜還融合了使用者-商品的交互資訊，即包含了三元組 (用戶，對話模式，商品)，其中對話模式為預定義的關係。路徑的定義是一個由實體或關係組成的序列，並且以用戶作為起始點，以商品作為終點。給定一個用戶，商品，以及連接該使用者和商品的路徑的集合，我們希望模型能夠計算該使用者和商品之間存在交互的可能性，即是否存在三元組 (用戶，對話模式，商品)。



模型一共分為三層：在 Embedding 層對路徑的每一個行為做 embedding 的計算。對於給定三元組，分別計算實體名稱、實體類型和關係 (或對話模式) 的 embedding 後再拼接得到最終特徵表示。LSTM 層將路徑上的每個單元的特徵表示按照時間順序輸入並且將最後時刻的隱藏層狀態作為該路徑的特徵表示。在 pooling 層，將所有路徑的特徵表示集合輸入兩層前饋神經網路，再對輸出做帶權重的池化操作得到最後的預測結果。

研究結果：

文章在公開電影資料集 MI 和音樂資料集 KKBox 上進行了實驗，驗證了所提出的模型的有效性，並且相對於僅將實體映射為一個向量表示的方法，KPRN 還能夠從路徑中挖掘使用者和商品之間的交互關係，這提高了模型的可解釋性。



如上圖所示，在 MovieLens-1M 中隨機選擇的一個用戶 u_{4825} ，並從她的交互記錄中選擇電影“戀愛中的莎士比亞”。然後，我們提取連接用戶-項對的所有有限定路徑，得到每個路徑的分數 $s_1 = 0.355$ ， $s_2 = 0.289$ ， $s_3 = 0.356$ ，即模型更傾向於認為使用者 u_{4825} 是通過路徑 3 和電影“戀愛中的莎士比亞”產生交互關係。

論文題目： *Knowledge Graph Embedding with Iterative Guidance from Soft Rules*

中文題目：基於規則反覆運算引導的增強知識圖譜表示學習

論文作者：Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, Tat-Seng Chua

論文出處：Proceedings of the AAAI Conference on Artificial Intelligence. 2018 (AAAI'18).

論文地址：<https://arxiv.org/abs/1711.11231v1>

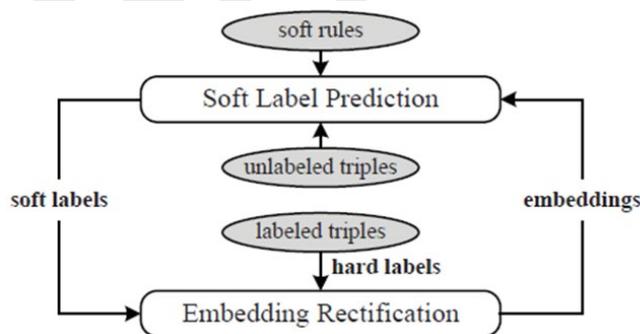
研究問題：

知識圖譜表示學習旨在將實體和關係嵌入到向量空間，同時保留知識圖譜的內在結構。傳統方法主要基於關係三元組學習知識圖譜的嵌入表示。本文認為，知識圖譜中的邏輯規則對於表示學習也很有幫助，於是提出了一個新的表示學習

模型 RUGE (Rule-Guided Embedding)。RUGE 的主要特點是，利用從知識圖譜自動抽取的規則反覆運算地預測未標記三元組，並將其加入訓練來增強表示學習。

研究方法：

為了在知識圖譜進行分散式表示學習的過程中加入邏輯規則進行引導，RUGE 同時利用標注三元組(labeled Triples) 未標注三元組(Unlabeled Triples) 自動抽取出的軟規則 (soft rules) 這三種資源以反覆運算的方式進行知識圖譜表示學習。軟規則指的是不總是成立，帶置信度的規則。每一輪反覆運算在軟標籤預測和 embedding 修正這兩個步驟間交替進行。前者利用當前學到的 embedding 和軟規則為未標注三元組預測軟標籤；後者進一步利用標注三元組 (硬標籤) 和未標注三元組(軟標籤)對當前的 embedding 進行修正。通過這個反覆運算過程，RUGE 可以成功建模分散式知識表示學習和邏輯推理二者間的交互性，邏輯規則中蘊含的豐富知識也能被更好地傳遞到所學習的分散式表示中。



上圖為 RUGE 框架圖，可以看出，RUGE 使嵌入模型能夠以反覆運算方式同時從標記的三元組、未標記的三元組和軟規則中學習。在每次反覆運算中，模型交替在軟標籤預測階段和 embeddings 校正階段之間。

學習資源構建：文章假設在知識圖譜中觀測到的三元組集合為正三元組，使用隨機替換頭尾實體的方式構建負三元組，此外還考慮那些能被軟規則 (soft rules) 編碼的未標記三元組，其中規則為不同置信度水準的 FOL 規則。

三元組和規則建模：對於三元組建模，文章採用現有模型 ComplEx，三元組的真值可以直接計算得到。對於規則建模，也就是建模規則 groundings 的真值，文章採用 T-norm fuzzy logics，規則的真值等於其三元組真值的邏輯組合。

軟標籤預測：可以基於 embedding 表示來計算標記和未標記三元組的“真值”；也可以基於規則 groundings 的真值來計算三元組的實際真值，即軟標籤 (soft label)。文章希望這兩個真值應該是接近的，且應當使得規則 groundings 為真為此設置了相應的優化目標。

Embedding 校正：得到了未標記三元組的軟標籤之後，文章結合已標記三元組，使用交叉熵進行統一優化。

研究結果：

文章的主要實驗任務是傳統的關係預測。資料集採用了 FB15K 和 YAGO37。實驗結果如下表所示，可以看見，RUGE 相比基線方法取得了較好的結果。文章創新性在於提出了軟規則，並可以成功建模分散式知識表示學習和邏輯推理二者間的交互性，邏輯規則中蘊含的豐富知識也能被更好地傳遞到所學習的分散式表示中。

Method	FB15K						YAGO37					
	MRR	MED	HITS@N				MRR	MED	HITS@N			
			1	3	5	10			1	3	5	10
TransE	0.400	4.0	0.246	0.495	0.576	0.662	0.303	13.0	0.218	0.336	0.387	0.475
DistMult	0.644	1.0	0.532	0.730	0.769	0.812	0.365	6.0	0.262	0.411	0.493	0.575
HolE	0.600	2.0	0.485	0.673	0.722	0.779	0.380	7.0	0.288	0.420	0.479	0.551
ComplEx	0.690	1.0	0.598	0.756	0.793	0.837	0.417	4.0	0.320	0.471	0.533	0.603
PTransE	0.679	1.0	0.565	0.768	0.810	0.855	0.403	9.0	0.339	0.444	0.473	0.506
KALE	0.523	2.0	0.383	0.616	0.683	0.762	0.321	9.0	0.215	0.372	0.438	0.522
RUGE	0.768	1.0	0.703	0.815	0.836	0.865	0.431	4.0	0.340	0.482	0.541	0.603

論文題目：*Variational Reasoning for Question Answering with Knowledge Graph*

中文題目：基於知識圖譜的問答變分推理

論文作者：Yuyu Zhang, Hanjun Dai, Zornitsa Kozareva, Alexander J. Smola, and Le Song

論文出處：Proceedings of the AAAI Conference on Artificial Intelligence, 2018 (AAAI'18).

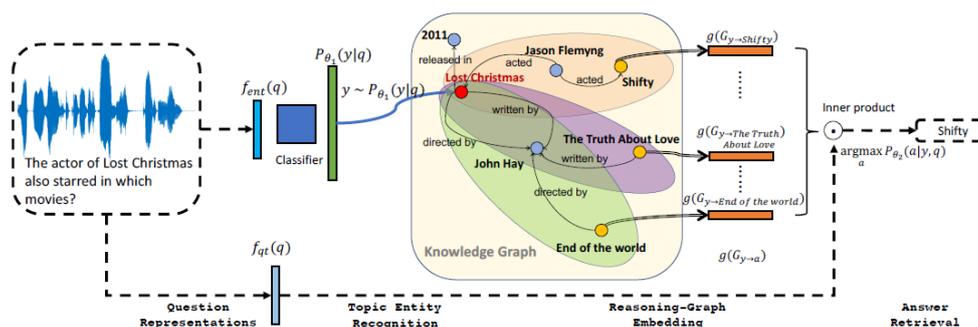
論文地址：<https://arxiv.org/abs/1709.04071v1>

研究問題：

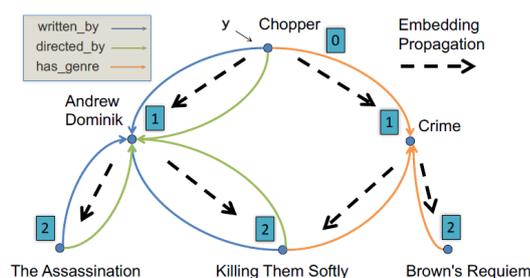
傳統的知識圖譜問答主要是基於語義解析的方法，這種方法通常是將問題映射到一個形式化的邏輯運算式，然後將這個邏輯表達轉化為知識圖譜的查詢。問題答案可以從知識圖譜中通過轉化後的查詢得到。然而傳統的基於語義解析的知識庫問答會存在一些挑戰，如基於查詢的方法只能獲取一些明確的資訊，對於知識庫中需要多跳才能獲取的答案則無法回答；此外實際的場景中用戶的輸入可能是通過語音辨識轉換而來或者是用戶通過打字輸入而來，很難確保不存在一定的雜訊。在具有雜訊的場景下，問句中的實體很難直接準確的匹配到知識庫上。因此文章提出了一個端到端的知識庫問答模型來解決以上兩個問題。

研究方法：

論文提出了變分推理網路 (VRN)，模型分為兩部分：通過概率模型來識別問句中的實體 (得到圖譜中每個實體是問句中實體的概率)，這避免了語義解析帶來的誤差。具體而言，將問句 (基於語音或者文本) 映射成向量，然後對其做 softmax 多分類，以計算問句中的實體的概率。



問答時在知識圖譜上做邏輯推理，且推理規則將被學習出來。給定 query 和問題實體 y ，希望模型從知識圖譜中找到答案 a 。文章首先將 query 通過另一個網路編碼成向量 q ，然後從實體 y 沿著知識圖譜向相鄰實體擴展搜索答案，形成推理路徑 G_y 。下圖是兩跳推理路徑的示意圖：



研究結果：

實驗結果顯示在 Vanilla、NTM 和 Audio 資料集下，演算法的效果都超過傳統的 QA 系統，同時在需要推理的問題中性能更為顯著。其中，在文章新發佈的 KBQA 資料集 MetaQA 上相比對照模型提升較為明顯，尤其是要求多跳推理的

問題。另外，在問題的形式是語音，機器翻譯後的結果，以及訓練時不給定標注好的 topicentity 的情況下，都有較大的提升。

	NTM-EU 1-hop	NTM-EU 2-hop	NTM-EU 3-hop	Audio-EU 1-hop	Audio-EU 2-hop	Audio-EU 3-hop
VRN	81.3	69.7	38.0	37.0	24.6	21.1
Bordes et al. [22]'s QA system	32.5	32.3	25.3	18.5	19.3	15.3
KV-MemNN	33.9	8.7	10.2	4.3	7.0	15.3
Supervised embedding	16.1	22.8	24.2	4.1	6.1	12.1

	Vanilla 1-hop	Vanilla 2-hop	Vanilla 3-hop	Vanilla-EU 1-hop	Vanilla-EU 2-hop	Vanilla-EU 3-hop
VRN	97.5	89.9	62.5	82.0	75.6	38.3
Bordes et al. [22]'s QA system	95.7	81.8	28.4	39.5	38.3	26.9
KV-MemNN	95.8	25.1	10.1	35.8	10.3	10.5
Supervised embedding	54.4	29.1	28.9	18.1	23.2	25.3

論文題目： *TorusE: Knowledge Graph Embedding on a Lie Group*

中文題目：TorusE：一種基於李氏群的知識圖譜嵌入表示學習

論文作者：Yuyu Zhang, Hanjun Dai, Zornitsa Kozareva, Alexander J. Smola, and Le Song

論文出處：Proceedings of the AAAI Conference on Artificial Intelligence. 2018
(AAAI'18)

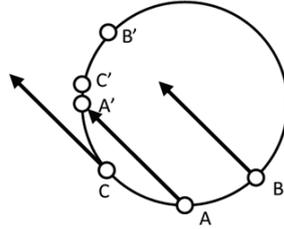
論文地址：

<https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/viewPDFInterstitial/16227/15885>

研究問題：

知識圖譜在許多人工智慧任務都扮演重要的角色。知識圖譜通常是用一個三元組 (h, r, t) 來表示一條知識。為了獲得它們的低維稠密表示，通常採用分散式

向量表示，例如 TransE 模型。然而 TransE 的正則約束會迫使實體的向量表示在一個球面上，而這與之前的優化條件又是相互矛盾的。



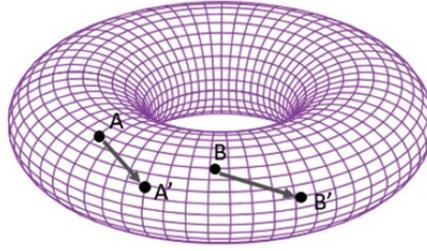
這種矛盾還會影響實體間連接預測的準確性。以下圖為例，箭頭方向表示關係 r ， A, B, C 及 A', B', C' 表示實體，對於 (C, r, C') 和 (B, r, B') 的正則約束和優化目標就是相互矛盾的。

研究方法：

文章提出的 TorusE 擁有類似 TransE 遵循的優化目標和正則項。為了避免上述的正則項帶來的矛盾，TorusE 不再將特徵學習到一個開流形 (open manifold) 的歐式空間，而是在緊空間 (compact space) 上學習知識圖譜的嵌入表示。可以證明，緊李群可以滿足 TransE 遵循的優化目標和正則項條件，即嵌入空間為可微的流形空間，滿足阿貝爾群性質，群運算可微且能夠定義距離函數。證明了任意一個阿貝爾李群符合嵌入空間的需求後，文章構建了一個緊李群的圓環空間 T^n 和圓環空間上的不同範式的距離函數 d_{L_1} ， d_{L_2} ， d_{eL_2} 。具體定義如下左圖所示。

Definition 2 An n -dimensional torus T^n is a quotient space, $\mathbb{R}^n / \sim = \{[x] | x \in \mathbb{R}^n\} = \{[y \in \mathbb{R}^n | y \sim x] | x \in \mathbb{R}^n\}$, where \sim is an equivalence relation and $y \sim x$ if and only if $y - x \in \mathbb{Z}^n$.

- d_{L_1} : A distance function d_{L_1} on T^n is derived from the L_1 norm of the original vector space by defining $d_{L_1}([x], [y]) = \min_{(x', y') \in [x] \times [y]} \|x' - y'\|_1$.
- d_{L_2} : A distance function d_{L_2} on T^n is derived from the L_2 norm of the original vector space by defining $d_{L_2}([x], [y]) = \min_{(x', y') \in [x] \times [y]} \|x' - y'\|_2$.
- d_{eL_2} : T^n can be embedded on \mathbb{C}^n by g . A distance function d_{eL_2} on T^n is derived from the L_2 norm of the \mathbb{C}^n by defining $d_{eL_2}([x], [y]) = \|g([x]) - g([y])\|_2$.



類似於 TransE 在 \mathbb{R}^n 上的優化目標 $h + r = t$, TorusE 在 T^n 上構建 $[h] + [r] = [t]$, 並根據距離函數的不同定義三個對應的評分函數：以 2 維的 TorusE 模型為例 (上右圖) , 箭頭方向表示關係 r , 對於三元組 (A, r, A') 和 (B, r, B') , 映射至嵌入空間後 , 我們仍可以得到 $[A'] - [A]$ 與 $[B'] - [B]$ 在圓環空間上是相似的。

研究結果：

文章在知識表示嵌入的可擴展性和連接預測上將 TorusE 和 TransE 進行對比。實驗結果表明 , TorusE 具有比 TransE 更快的計算表現 , 證明了 TorusE 具有更低的計算複雜度。在連接預測的任務上 , TorusE 比當今最好的模型仍要表現出色 (見下表) 。

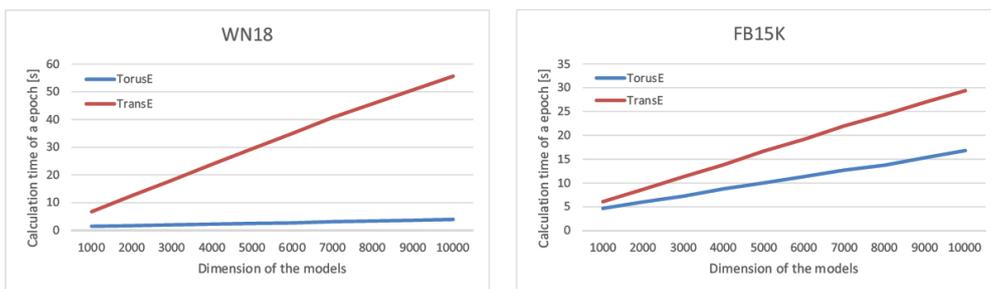


Figure 4: Calculation time of TorusE and TransE on WN18 and FB15K

Model	WN18					FB15K				
	MRR		HITS@			MRR		HITS@		
	Filtered	Raw	1	3	10	Filtered	Raw	1	3	10
TransE	0.397	0.306	0.040	0.745	0.923	0.414	0.235	0.247	0.534	0.688
TransR	0.605	0.427	0.335	0.876	0.940	0.346	0.198	0.218	0.404	0.582
RESCAL	0.890	0.603	0.842	0.904	0.928	0.354	0.189	0.235	0.409	0.587
DistMult	0.822	0.532	0.728	0.914	0.936	0.654	0.242	0.546	0.733	0.824
ComplEx	0.941	0.587	0.936	0.945	0.947	0.692	0.242	0.599	0.759	0.840
TorusE	0.947	0.619	0.943	0.950	0.954	0.733	0.256	0.674	0.771	0.832

論文題目： *Commonsense Knowledge Aware Conversation Generation with Graph Attention*

中文題目：基於圖注意力機制的常識感知對話生成

論文作者：Hao Zhou, Tom Young, Minlie Huang, Haizhou Zhao, Jingfang Xu, Xiaoyan Zhu

論文出處：*Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*

論文地址：<https://www.ijcai.org/proceedings/2018/0643.pdf>

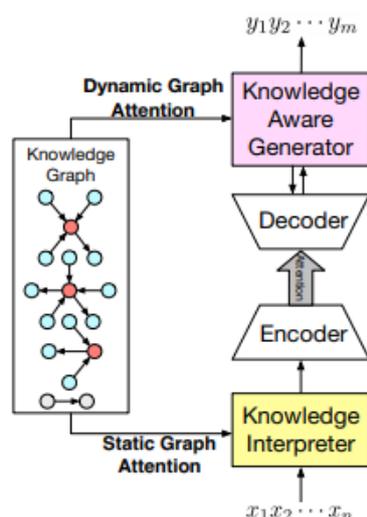
研究問題：

文章著重研究了基於常識知識庫的開放域對話生成問題。具體來說，該問題的輸入為一個常識知識庫和開放域的問題，模型需要生成一個符合上下文語境的回答作為輸出。

研究方法：

整體來說，對於一個輸入的問題，文章提出的模型會從常識知識庫中檢索相應的知識圖，然後基於靜態圖注意力機制對其進行編碼，圖注意力機制有助於提升語義資訊，從而說明系統更好地理解問題。接下來，在語句的生成過程中，模

型會逐個讀取檢索到的知識圖以及每個圖中的知識三元組，並通過動態的圖注意力機制來優化語句的生成。



具體的，模型可以分為 3 個主要的部分，分別如下：（1）知識解析器，這部分模型旨在優化問題理解這一環節。它通過引入每個單詞對應的圖向量，來增強單詞的語義。知識解析器會把問題中的每個單詞作為關鍵實體，從整個常識知識庫中檢索圖；（2）靜態圖注意力，靜態圖注意力機制可以看做是知識解析器的一部分，它可以為檢索到的知識圖提供一個表現形式；（3）知識感知生成器與動態圖注意力，這部分模型會讀取所有的知識圖和每個圖中的所有三元組，用來生成最終的回復。

研究結果：

該論文使用 ConceptNet 作為常識知識庫，並基於 reddit 上一問一答形式的對話資料，構建了常識對話資料集。由於論文的目標是用常識知識優化語言理解和生成，所以資料集僅保留了帶有知識三元組的原始語料資料。

論文使用自動評估和人工評估兩種方法同時對模型的表現進行評價，並選取了 Seq2Seq、MemNet 和 CopyNet 三種基線模型進行對比。最終研究結果表明，本篇論文提出的模型在兩種評價方法下，都比幾個基線模型表現突出。此外，該論文還做了案例研究，對於一個具體的問題，本篇論文的模型可以生成更為合理，資訊也更豐富的回復。

論文題目： *That's Interesting, Tell Me More! Finding Descriptive Support Passages for Knowledge Graph Relationships*

中文題目：知識圖譜關係的描述性段落查找

論文作者：Sumit Bhatia, Purusharth Dwivedi, Avneet Kaur.

論文出處：*Proceedings of the International Semantic Web Conference 2018*

論文地址：https://link.springer.com/chapter/10.1007/978-3-030-00671-6_15

研究問題：

近年來，知識圖譜的應用日益增多，其可靠性成為一個待解決的關鍵問題。文章研究如何從文本語料中為知識圖譜的三元組尋找支持段落，從而增強用戶對知識圖譜的信任並幫助其做出更正確的決策。

研究方法：

文章提出了一種概率方法來表示一個段落描述了特定三元組的概率。通過使用貝葉斯公式，並假設三元組的頭實體、謂語、尾實體三個元素條件獨立，可以將目標概率拆解為段落描述頭實體、段落描述謂語、段落描述尾實體三個概率的乘積。進一步，由於頭實體、謂語、尾實體均由一個或多個單詞組成，可以將這些單詞提取出來，將目標概率繼續拆解為給定段落描述了某個單詞的概率的乘積。

為了表示段落 p 描述單詞 w 的概率，文章基於語言模型進行改進，提出考慮三個層次的因素，分別是段落因素、文檔因素和集合因素。其中段落因素考慮 w 直接在 p 中出現的概率；文檔因素考慮 w 在 p 所在文檔中出現的概率，從而解決段落中使用代詞來指代實體時直接概率過低的問題；集合因素考慮整個文本語料，相當於背景語言模型，與資訊檢索領域常用的 IDF 作用相當。文章使用統計頻率來計算以上三個因素的概率，計算方法如下。

$$\text{Passage Evidence: } P(w|\theta_p) = \frac{\text{count}(w, p) + 1}{|p| + |V|}$$

$$\text{Document Evidence: } P(w|\theta_d) = \frac{\text{count}(w, d) + 1}{|d| + |V|}$$

$$\text{Collection Evidence: } P(w|\theta_c) = \frac{\text{count}(w, c)}{|C|}$$

將三個概率加權相加，得到段落描述單詞的概率，再進行累乘從而得到段落描述三元組的概率。最終為每個三元組對所有候選段落根據該概率進行排序。

研究結果：

文章從 WikiData 中構造了 50 個三元組作為目標三元組，以 Wikipedia 作為文本語料，對每一個三元組選出了語料中排名最高的 5 個段落作為結果，並依靠

3 名標注人員對結果進行評價。評價分為三個等級，分別是該段落與三元組相關部分相關和不相關。

	Evaluator 1	Evaluator 2	Evaluator 3	Final
non-relevant	406	438	444	449
partially-relevant	41	11	43	12
relevant	248	246	208	234

對於基線方法，使用相同的三元組進行抽取和評價。最終結果表明，文章提出的方法相比於基線方法有顯著改進，排名第一的段落的準確率從 0.251 提高到了 0.860。

	P@1	Precision	MRR
Inf. N/w	0.251	0.156	0.272
Inf. N/w + Rel. Exp.	0.165	0.088	0.144
Proposed Approach	0.860	0.727	0.805

論文題目：*HighLife: Higher-arity Fact Harvesting*

中文題目：HighLife：高精度事實獲取

論文作者：Patrick Ernst, Amy Siu, Gerhard Weikum

論文出處：Proceedings of the 2018 World Wide Web Conference

論文地址：<https://dl.acm.org/citation.cfm?id=3186000>

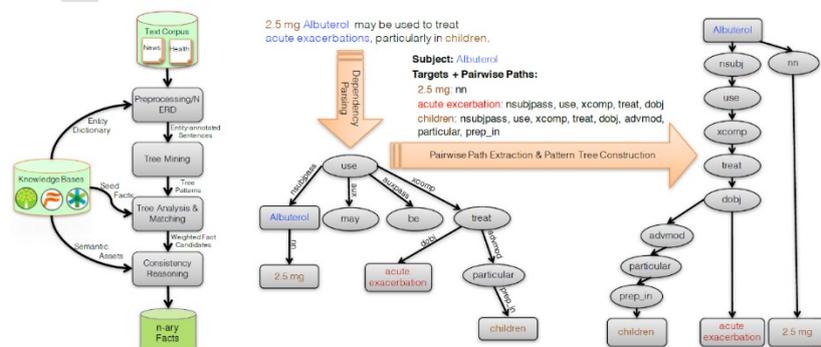
研究問題：

現有的知識抽取方法主要關注二元實事，即兩個實體之間的關係。然而在實際中，我們往往需要考慮三元關係甚至是更加高階的關係，這種高階的關係能夠更加精確地表達一個事實。比如在醫療領域，我們需要知道一種藥品治療哪一種疾病，這種藥品的使用劑量是什麼以及適合哪種類型的患者（例如，兒童或者是

成人)。這篇文章提出了一種從文本中獲取高階實事的方法，基於遠端監督的假設，從一些種子集合出發。一方面為了提高召回率，利用二元的事實範本去發現盡可能多的事實；另一方面為了提高準確率，設計了一種基於約束的推理方法來去除錯誤的候選。主要的創新點是解決了高階事實在文本中的表達不集中，分佈不規律的問題。例如，一句話可以指一種藥物、一種疾病和一組患者，而另一句話則是指藥物、其劑量和目標人群，而沒有提到疾病。這篇文章的方法在模式學習和約束推理階段都能很好地處理這些部分觀察到的事實。對健康相關文檔和新聞文章的實驗證明了這種方法的可行性。

研究方法：

這篇文章的框架名叫 HighLife，由兩部分組成。首先使用種子事實作為遠端監控來學習範本，使用這些範本來提取候選事實，並反覆運算這些步驟，保證召回率。進而將事實從二元擴展到更高階的情況。在實現高召回率的同時，這種方法容易受到雜訊和目標漂移的影響。因此，作者使用約束推理來消除虛假的候選事實，將基於 MaxSat 的推理器擴展到更高階情況。例如，可以應用類型約束來確定何時贏得普利策獎的事實是電影或歌曲（而不是書籍）。



研究結果：

為了證明方法的有效性和通用性，文章在新聞和生物醫學領域上分別進行了實驗。將 HighLife 模型及其變種和語義角色標注的基線方法進行對比，結果表明 Highlife 模型顯著優於已有的 SRL 方法，進一步分析了 Highlife 模型在兩個資料集上的表現，發現在醫學領域的效果更好。

4.5 知識工程最新進展

近兩年知識獲取、推理和應用研究取得了顯著的進展，主要表現在如下幾個方面：

- 資源匱乏情況下的知識獲取

知識圖譜的構建始終是知識圖譜領域的核心問題之一，近年來除了傳統的有監督的實體、關係、事件知識獲取的研究外，也湧現了一批在弱資源情況下的知識獲取方法。例如：在集合擴展（實體集擴展）研究中，Learning to Bootstrap for Entity Set Expansion 使用蒙特卡洛樹搜索策略的 bootstrap 方法有效地提升了實體集擴展方法的穩定性，尤其是在與分類體系相關任務的同時優化上。HiExpan: Task-Guided Taxonomy Construction by Hierarchical Tree Expansion 提出一個知識分類體系的擴展框架，模型利用弱監督關係抽取模型，從一個小型的上下位關係樹出發，抽取擴展的節點並擴展成一個更加豐富的上下位體系。FewRel 2.0: Towards More Challenging Few-Shot Relation Classification 提出了少次學習任務，通過設計少次學習機制，能夠利用從過往資料中學到的泛化知識，結合新類型資料的少量訓練樣本，實現快速遷移學習。COMET: Commonsense Transformers for Automatic Knowledge Graph Construction 提出常識 Transformer 架構，將 GPT-2

等語言模型與種子知識圖譜相結合，學習其結構和關係，根據圖表徵形成語言模型，從而生成新的知識並將它們添加到種子圖中。

- 知識圖譜的知識補全和可解釋推理

傳統的表示學習缺乏可解釋性，知識圖譜推理越來越受到關注，其中既有使用強化學習方法尋找路徑的方法，也有使用實體鄰居和注意力權重做可解釋性推理方法。Multi-Hop Knowledge Graph Reasoning with Reward Shaping 是基於多跳推理的知識庫問答方法，基於強化學習擴展在知識圖譜的推理路徑，以獲得問題的正確答案。Learning Attention-based Embeddings for Relation Prediction in Knowledge Graphs 提出一種基於注意力機制的特徵嵌入方法，獲取實體鄰近範圍內的實體和關係特徵，引入關係聚類和多跳關係，有效提升了基於多跳推理的知識圖譜補全的效果。Iteratively Learning Embeddings and Rules for Knowledge Graph Reasoning 研究如何反覆運算地進行知識表示學習和規則學習，提出的 IterE 模型可以利用學習的規則改進稀疏實體的表示學習，進而提升規則學習和連結預測效果。

- 基於知識圖譜的推薦和對話問答

將知識圖譜作為輔助資訊引入到推薦系統中可以有效地解決傳統推薦系統存在的稀疏性和冷開機問題，近幾年吸引大量研究人員在相關工作。隨著圖卷積神經網路，圖注意力機制等技術的逐漸興起，基於圖表示學習的推薦模型達到了更高的表現效果，並為推薦系統的可解釋性提供了幫助。KGAT: Knowledge Graph Attention Network for Recommendation 利用知識圖譜中商品之間的關係，訓練了一個端到端的含注意力機制的模型，用於提高推薦系統的能力。AKUPM:

Attention-Enhanced Knowledge-Aware User Preference Model for Recommendation 使用注意力模型，利用知識圖譜對用戶進行建模，顯著提升了推薦系統的效果。

Reinforcement Knowledge Graph Reasoning for Explainable Recommendation 結合強化學習的框架和知識圖譜推理來提供對推薦結果的解釋。在對話問答方面，以前對話生成的資訊源是文本與對話記錄，但如果遇到詞表之外的 (Out-of-Vocabulary) 的詞，模型往往難以生成合適的、有信息量的回復，而會產生一些低品質的、模稜兩可的回復。Commonsense Knowledge Aware Conversation Generation with Graph 提出一種基於常識知識圖譜的對話模型 CCM 來理解對話，產生資訊豐富且合適的回復。

5 自然語言處理

5.1 自然語言處理概念

自然語言是指漢語、英語、法語等人們日常使用的語言，是人類社會發展演變而來的語言，而不是人造的語言，它是人類學習生活的重要工具。概括說來，自然語言是指人類社會約定俗成的，區別於如程式設計的語言的人工語言。在整個人類歷史上以語言文字形式記載和流傳的知識占到知識總量的 80%以上。就電腦應用而言，據統計，用於數學計算的僅占 10%，用於程序控制的不到 5%，其餘 85%左右都是用於語言文字的資訊處理。

處理包含理解、轉化、生成等過程。自然語言處理，是指用電腦對自然語言的形、音、義等資訊進行處理，即對字、詞、句、篇章的輸入、輸出、識別、分析、理解、生成等的操作和加工。實現人機間的資訊交流，是人工智慧、電腦科學和語言學所共同關注的重要問題。自然語言處理的具體表現形式包括機器翻譯、文本摘要、文本分類、文本校對、資訊抽取、語音合成、語音辨識等。可以說，自然語言處理就是要電腦理解自然語言，自然語言處理機制涉及兩個流程，包括自然語言理解和自然語言生成。自然語言理解是指電腦能夠理解自然語言文本的意義，自然語言生成則是指能以自然語言文本來表達給定的意圖。



圖 5-1 自然語言理解層次

自然語言的理解和分析是一個層次化的過程，許多語言學家把這一過程分為五個層次，可以更好地體現語言本身的構成，五個層次分別是語音分析、詞法分析、句法分析、語義分析和語用分析。

- 語音分析是要根據音位規則，從語音流中區分出一個個獨立的音素，再根據音位形態規則找出音節及其對應的詞素或詞。
- 詞法分析是找出詞彙的各個詞素，從中獲得語言學的資訊。
- 句法分析是對句子和短語的結構進行分析，目的是要找出詞、短語等的相互關係以及各自在句中的作用。
- 語義分析是找出詞義、結構意義及其結合意義，從而確定語言所表達的真正含義或概念。
- 語用分析是研究語言所存在的外界環境對語言使用者所產生的影響。

在人工智慧領域或者是語音資訊處理領域中，學者們普遍認為採用圖靈試驗可以判斷電腦是否理解了某種自然語言，具體的判別標準有以下幾條：

- 第一，問答，機器人能正確回答輸入文本中的有關問題；
- 第二，文摘生成，機器有能力生成輸入文本的摘要；
- 第三，釋義，機器能用不同的詞語和句型來複述其輸入的文本；
- 第四，翻譯，機器具有把一種語言翻譯成另一種語言的能力。

5.2 自然語言的理解發展歷史

自然語言處理是包括了電腦科學、語言學心理認知學等一系列學科的一門交叉學科，這些學科性質不同但又彼此相互交叉。因此，梳理自然語言處理的發展歷程對於我們更好地瞭解自然語言處理這一學科有著重要的意義。

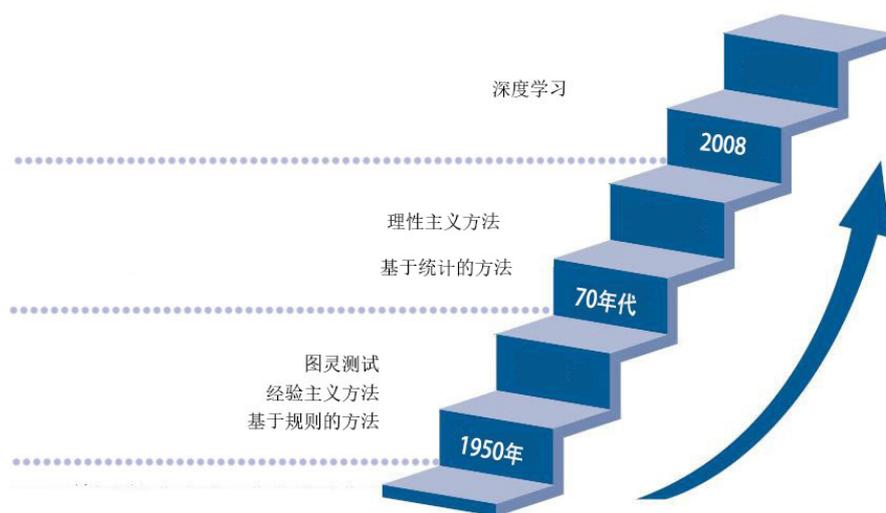


圖 5-2 自然語言發展歷程

1950 年圖靈提出了著名的“圖靈測試”，這一般被認為是自然語言處理思想的開端，20 世紀 50 年代到 70 年代自然語言處理主要採用基於規則的方法，研究人員們認為自然語言處理的過程和人類學習認知一門語言的過程是類似的，所以大量的研究員基於這個觀點來進行研究，這時的自然語言處理停留在理性主義思潮階段，以基於規則的方法為代表。但是基於規則的方法具有不可避免的缺點，首先規則不可能覆蓋所有語句，其次這種方法對開發者的要求極高，開發者不僅要精通電腦還要精通語言學，因此，這一階段雖然解決了一些簡單的問題，但是無法從根本上將自然語言理解實用化。

70 年代以後隨著互聯網的高速發展，豐富的語料庫成為現實以及硬體不斷更新完善，自然語言處理思潮由經驗主義向理性主義過渡，基於統計的方法逐漸

代替了基於規則的方法。賈裡尼克和他領導的 IBM 華生實驗室是推動這一轉變的關鍵，他們採用基於統計的方法，將當時的語音辨識率從 70% 提升到 90%。在這一階段，自然語言處理基於數學模型和統計的方法取得了實質性的突破，從實驗室走向實際應用。

從 2008 年到現在，在圖像識別和語音辨識領域的成果激勵下，人們也逐漸開始引入深度學習來做自然語言處理研究，由最初的詞向量到 2013 年的 word2vec，將深度學習與自然語言處理的結合推向了高潮，並在機器翻譯、問答系統、閱讀理解等領域取得了一定成功。深度學習是一個多層的神經網路，從輸入層開始經過逐層非線性的變化得到輸出。從輸入到輸出做端到端的訓練。把輸入到輸出對的資料準備好，設計並訓練一個神經網路，即可執行預想的任務。RNN 已經是自然語言處理最常用的方法之一，GRU、LSTM 等模型相繼引發了一輪又一輪的熱潮。

近年自然語言處理在詞向量 (word embedding) 表示、文本的 (編碼) encoder 和 decoder (反編碼) 技術以及大規模預訓練模型 (pre-trained) 上的方法極大地促進了自然語言處理的研究^[7]。

5.3 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為自然語言處理領域全球學者分佈情況：



圖 5-3 自然語言處理領域全球人才分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量優勢明顯且主要分佈在其東西海岸；歐洲也有較多的人才分佈，主要集中在歐洲中西部；亞洲的人才主要分佈在我國東部及日韓地區；其他諸如非洲、南美洲等地區的學者非常稀少；自然語言處理領域的人才分佈與各地區的科技、經濟實力情況大體一致。此外，在性別比例方面，自然語言處理領域中男性學者占比 89.3%，女性學者占比 10.7%，男性學者占比遠高於女性學者。

自然語言處理領域學者的 h-index 分佈如下圖所示，分佈情況大體呈階梯狀，大部分學者的 h-index 分佈在中低區域，其中 h-index 在小於 20 區間的人數最多，有 929 人，占比 45.3%，50-60 區間的人數最少，有 98 人。

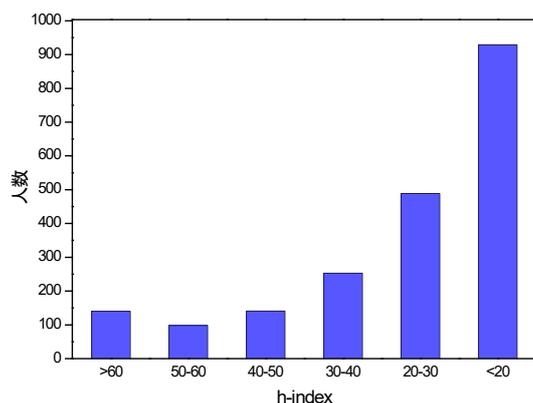


圖 5-4 自然語言處理領域學者 h-index 分佈

- 中國人才分佈

我國專家學者在自然語言處理領域的分佈如下圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是長三角和珠三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，通過觀察中國周邊國家的學者數量情況，特別是與日韓、東南亞等亞洲國家相比，中國在自然語言處理領域學者數量較多。



圖 5-5 自然語言處理領域中國學者分佈

中國與其他國家在自然語言處理領域的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 5-1 自然語言處理領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	250	7589	30	472
中國-愛爾蘭	41	826	20	34
中國-新加坡	37	1537	42	77
中國-英國	34	1223	36	61
中國-日本	24	513	21	41

中國-印度	23	1368	59	32
中國-加拿大	19	307	16	35
中國-德國	15	124	8	30
中國-澳大利亞	8	120	15	20
中國-柬埔寨	8	101	13	9

從上表資料可以看出，中美合作的論文數、引用數、學者數遙遙領先，表明中美間在自然語言處理領域合作之密切；此外，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 3 席；中國與印度合作的論文數雖然不是最多，但是擁有最高的平均引用數說明在合作品質上中印合作達到了較高的水準。

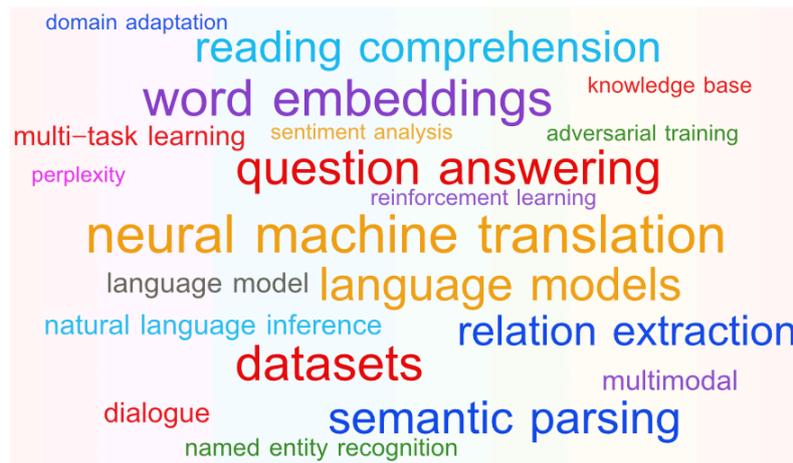
5.4 論文解讀

本節對本領域的高水準學術會議論文進行挖掘，解讀這些會議在 2018-2019 年的部分代表性工作。會議具體包括：

Annual Meeting of the Association for Computational Linguistics

Conference on Empirical Methods in Natural Language Processing

我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如下圖所示。其中，神經機器翻譯 (neural machine translation)、詞嵌入 (word embeddings)、資料集 (datasets) 是本領域中最熱的關鍵字。



論文題目： *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*

中文題目： BERT: 語言理解的深層雙向轉換器的預訓練

論文作者： Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova

論文出處： In Proceedings of the 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics.

論文地址： <https://arxiv.org/abs/1810.04805>

研究問題：

文章介紹一種新的語言表示模型 BERT (Bidirectional Encoder Representations from Transformers)，通過聯合上下文資訊從未標記文本中預訓練深層雙向表示形式，只需一個額外的輸出層，就可以對預訓練模型進行調整，在不需要對特定任務的體系結構進行大量修改的前提下，在多種語言相關任務上獲得。

研究內容：

模型包含預訓練和微調兩個步驟：在預訓練階段，對不同訓練任務的未標記資料進行訓練。在微調階段，首先用預訓練參數初始化 BERT 模型，然後，使用來自下游任務的標記資料對預訓練的參數進行微調。

BERT 是一個多層的雙向 Transformer，輸入包括三個部分，分別為詞向量、單詞所屬句子向量和單詞的位置向量，形象的表示如下圖所示，其中[CLS]和[SEP]是放在每個輸入最前和使用者分隔句子的特殊符號。

文章提出兩種無監督任務來預訓練 BERT，分別是遮罩語言模型 (Masked Language Model, MLM) 和下旬預測模型 (Next Sentence Prediction, NSP)：MLM 通過遮罩一句話中部分詞然後讓模型來預測遮罩詞來訓練模型。在實驗設置中，大約 15%的詞被隨機遮罩。但是這樣的訓練方法也有缺陷，遮罩詞相當於從資料集中抹去，且可能預訓練階段與微調階段不一致。因此，對於遮罩詞有如下三種處理方式：80%用[MASK]替換，10%用隨機的詞語替換，另外 10%不做改變。NSP 任務是為了增強模型對句子間關係的理解能力，訓練時選擇的句對 A、B 中，B 有 50%的概率真的是 A 的下一句，50%的概率不是 A 的下一句。預訓練語料使用 BooksCorpus 和英語維琪百科的文本段落。

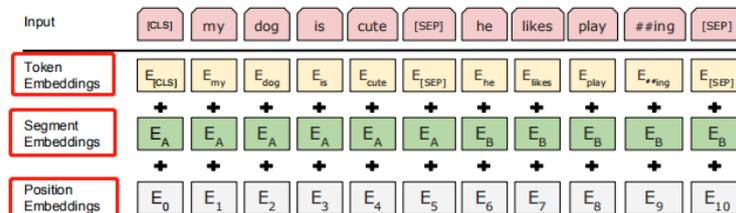


Figure 2: BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

模型微調測試了 11 個自然語言處理任務上的效果，包括 General Language Understanding Evaluation(GLUE)基準測試集中的 8 項評測 SQuAD 1.1 和 SQuAD

2.0 兩個閱讀理解資料集和 Situations With Adversarial Generations (SWAG)資料集。BERT 均穩定優於基線方法，下表展示了 GLUE 上的對比結果。

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	Average
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
BERT _{BASE}	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
BERT _{LARGE}	86.7/85.9	72.1	92.7	94.9	60.5	86.5	89.3	70.1	82.1

Table 1: **GLUE Test results**, scored by the evaluation server (<https://gluebenchmark.com/leaderboard>). The number below each task denotes the number of training examples. The “Average” column is slightly different than the official GLUE score, since we exclude the problematic WNLI set.⁸ BERT and OpenAI GPT are single-model, single task. F1 scores are reported for QQP and MRPC, Spearman correlations are reported for STS-B, and accuracy scores are reported for the other tasks. We exclude entries that use BERT as one of their components.

研究結論：

文章提出的 BERT 模型在 11 項自然語言處理任務上取得了最先進的效果。由語言模型轉移學習帶來的模型效果改進表明，豐富的、無監督的預訓練是許多語言理解系統的組成部分。特別地，即使是資源匱乏的任務也可以從深層的單向架構中獲益。文章主要貢獻是進一步將這些發現推廣到深層的雙向架構，允許相同的預訓練模型成功地應用於廣泛的 NLP 任務。

論文題目：*Semi-Supervised Learning for Neural Machine Translation*

中文題目：神經機器翻譯的半監督學習機制

論文作者：Yong Cheng, Wei Xu, Zhongjun He, Wei He, Hua Wu, Maosong Sun and Yang Liu

論文出處：Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics

論文地址：https://link.springer.com/chapter/10.1007/978-981-32-9748-7_3

研究問題：

近年來，端到端神經機器翻譯(neural machine translation, NMT)取得了顯著的進展，但 NMT 系統僅依靠並行語料庫進行參數估計。由於平行語料庫在數量、品質和覆蓋範圍等方面都存在一定的局限性，尤其是對資源相對較少的語言而言。所以利用單語語料庫來提高網路機器翻譯的性能就變得很有吸引力了。文章就提出了一種半監督的方法來訓練 NMT 模型。其核心思想是使用一個自編碼器重建單語語料庫，其中源-目標和目標-源轉換模型分別充當編碼器和解碼器。該方法不僅可以利用目標語的單語語料庫，而且還可以利用源語的單語語料庫。

研究內容：

首先，將觀察到的目標句編碼為潛在的源句(圖中藍色箭頭的過程)。然後，使用源到源的翻譯模型，對源句進行解碼(圖中黃色箭頭的過程)，利用源到目標的模型重構所觀察到的目標句。

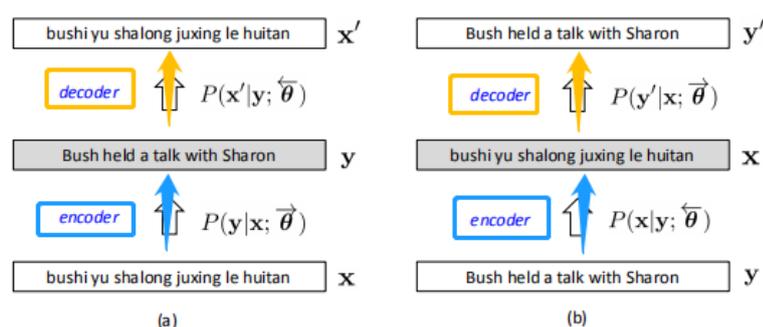


Figure 1: Examples of (a) source autoencoder and (b) target autoencoder on monolingual corpora. Our idea is to leverage autoencoders to exploit monolingual corpora for NMT. In a source autoencoder, the source-to-target model $P(y|x; \vec{\theta})$ serves as an encoder to transform the observed source sentence x into a latent target sentence y (highlighted in grey), from which the target-to-source model $P(x'|y; \vec{\theta})$ reconstructs a copy of the observed source sentence x' from the latent target sentence. As a result, monolingual corpora can be combined with parallel corpora to train bidirectional NMT models in a semi-supervised setting.

目標編碼器公式如下：

$$\begin{aligned}
& P(y'|y; \vec{\theta}, \vec{\theta}) \\
&= \sum_x P(y', x|y; \vec{\theta}, \vec{\theta}) \\
&= \sum_x \underbrace{P(x|y; \vec{\theta})}_{\text{encoder}} \underbrace{P(y'|x; \vec{\theta})}_{\text{decoder}},
\end{aligned}$$

x 是潛在的原句, y 是目標句, y' 是要重構的目標句的副本。同樣的, 源句的編碼器公式如下:

$$\begin{aligned}
& P(x'|x; \vec{\theta}, \vec{\theta}) \\
&= \sum_y P(x', y|x; \vec{\theta}) \\
&= \sum_y \underbrace{P(y|x; \vec{\theta})}_{\text{encoder}} \underbrace{P(x'|y; \vec{\theta})}_{\text{decoder}}.
\end{aligned}$$

由於自動編碼器同時涉及到源到目標和目標到源的模型, 所以很自然地要將並行語料庫和單語語料庫結合起來, 在半監督的環境中學習雙向 NMT 翻譯模型。

平行語料庫: $D = \{(x^n, y^n)\}_{n=1}^N$

目的語言的單語料庫: $T = \{y^t\}_{t=1}^T$

來源語言的單語料庫: $S = \{x^s\}_{s=1}^S$

訓練的目標函數為:

$$\begin{aligned}
& J(\vec{\theta}, \vec{\theta}) \\
&= \underbrace{\sum_{n=1}^N \log P(y^{(n)}|x^{(n)}; \vec{\theta})}_{\text{source-to-target likelihood}} \\
&+ \underbrace{\sum_{n=1}^N \log P(x^{(n)}|y^{(n)}; \vec{\theta})}_{\text{target-to-source likelihood}} \\
&+ \lambda_1 \underbrace{\sum_{t=1}^T \log P(y^t|y^t; \vec{\theta}, \vec{\theta})}_{\text{target autoencoder}} \\
&+ \lambda_2 \underbrace{\sum_{s=1}^S \log P(x^s|x^s; \vec{\theta}, \vec{\theta})}_{\text{source autoencoder}},
\end{aligned}$$

其中， λ_1 和 λ_2 是超參數。很明顯，源到目標和目標到源的模型是通過自動編碼器連接起來的，並且有望在聯合訓練中相互受益。

模型用小批量隨機梯度下降演算法進行優化。 $\vec{\theta}$ 的導數如下所示：

$$\begin{aligned} & \frac{\partial J(\vec{\theta}, \vec{\theta})}{\partial \vec{\theta}} \\ &= \sum_{n=1}^N \frac{\partial \log P(\mathbf{y}^{(n)}|\mathbf{x}^{(n)}; \vec{\theta})}{\partial \vec{\theta}} \\ & \quad + \lambda_1 \sum_{t=1}^T \frac{\partial \log P(\mathbf{y}'|y^{(t)}; \vec{\theta}, \vec{\theta})}{\partial \vec{\theta}} \\ & \quad + \lambda_2 \sum_{s=1}^S \frac{\partial \log P(\mathbf{x}'|\mathbf{x}^{(s)}; \vec{\theta}, \vec{\theta})}{\partial \vec{\theta}}. \end{aligned}$$

由於第二部分和第三部分依賴於搜索空間難以計算，文章又提出了一種近似計算的方法，縮小 y 的搜索空間。用 $\tilde{x}(y)$ 近似替代 $x(y)$ ，即用 $x(y)$ 候選翻譯的前 top-k 個作為 $\tilde{x}(y)$ 。

用文章的方法和最先進的 SMT 和 NMT 方法進行比較，實驗結果圖如下：

System	Training Data			Direction	NIST06	NIST02	NIST03	NIST04	NIST05
	CE	C	E						
MOSES	✓	×	×	C → E	32.48	32.69	32.39	33.62	30.23
				E → C	14.27	18.28	15.36	13.96	14.11
	✓	×	✓	C → E	34.59	35.21	35.71	35.56	33.74
	✓	✓	×	E → C	20.69	25.85	19.76	18.77	19.74
RNNSEARCH	✓	×	×	C → E	30.74	35.16	33.75	34.63	31.74
				E → C	15.71	20.76	16.56	16.85	15.14
	✓	×	✓	C → E	35.61****	38.78****	38.32****	38.49****	36.45****
				E → C	17.59++	23.99++	18.95++	18.85++	17.91++
	✓	✓	×	C → E	35.01**	38.20****	37.99****	38.16****	36.07****
				E → C	21.12***	29.52****	20.49****	21.59****	19.97**

Table 2: Comparison with MOSES and RNNSEARCH. MOSES is a phrase-based statistical machine translation system (Koehn et al., 2007). RNNSEARCH is an attention-based neural machine translation system (Bahdanau et al., 2015). “CE” donates Chinese-English parallel corpus, “C” donates Chinese monolingual corpus, and “E” donates English monolingual corpus. “✓” means the corpus is included in the training data and × means not included. “NIST06” is the validation set and “NIST02-05” are test sets. The BLEU scores are case-insensitive. “*”: significantly better than MOSES ($p < 0.05$); “***”: significantly better than MOSES ($p < 0.01$); “+”: significantly better than RNNSEARCH ($p < 0.05$); “++”: significantly better than RNNSEARCH ($p < 0.01$).

Method	Training Data			Direction	NIST06	NIST02	NIST03	NIST04	NIST05
	CE	C	E						
Sennrich et al. (2015)	✓	×	✓	C → E	34.10	36.95	36.80	37.99	35.33
	✓	✓	×	E → C	19.85	28.83	20.61	20.54	19.17
this work	✓	×	✓	C → E	35.61**	38.78**	38.32**	38.49*	36.45**
				E → C	17.59	23.99	18.95	18.85	17.91
	✓	✓	×	C → E	35.01**	38.20**	37.99**	38.16	36.07**
				E → C	21.12**	29.52**	20.49	21.59**	19.97**

Table 3: Comparison with Sennrich et al. (2015). Both Sennrich et al. (2015) and our approach build on top of RNNSEARCH to exploit monolingual corpora. The BLEU scores are case-insensitive. “*”: significantly better than Sennrich et al. (2015) ($p < 0.05$); “***”: significantly better than Sennrich et al. (2015) ($p < 0.01$).

研究結果：

文章提出了一種訓練神經機器翻譯模型的半監督方法。其核心思想是在單語語料庫上引入自動編碼器，採用源對目標和目標對源的翻譯模型作為編碼器和解碼器。在漢英 NIST 資料集上的實驗表明，與最先進的 SMT 和 NMT 方法進行，該方法帶來了顯著的改善。

論文題目： *Know What You Don't Know: Unanswerable Questions for SquAD*

中文題目：知道你所不知道的：針對 SquAD 中不可回答的問題

論文作者：Pranav Rajpurkar, Robin Jia, Percy Liang

論文出處：Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018

論文地址：<https://arxiv.org/abs/1806.03822>

研究問題：

閱讀理解系統(模型)通常可以在上下文文檔中找到問題的正確答案，但對於沒有在上下文中說明正確答案的問題，它們給出的答案也不那麼可靠。現有的資料集要麼只關注可回答的問題，要麼使用容易識別的自動生成的不可回答的問題作為資料集。為了彌補這些不足，文章介紹了斯坦福問答資料集(SQuAD)的最新版本——SQuAD 2.0，它整合了現有的 SQuAD 中可回答的問題和 50000 多個由大眾工作者編寫的難以回答的問題，其中那些難以回答的問題與可回答的問題題目設置相似。為了在 SQuAD 2.0 中表現的更好，系統不僅要在可能的情況下

回答問題，還要確定什麼時候段落的上下文不支持回答，並且避免回答問題。

SQuAD 2.0 資料集是自然語言理解任務中對現有模型的一個挑戰。

研究方法：

資料集：在 Daemo 平臺上雇傭了眾包工作人員來編寫無法回答的問題。每個任務由來自 SQuAD 1.1 的一整篇文章組成。對於文章中的每個段落，工作人員最多可提出 5 個僅憑段落是無法回答的問題，同時還要參考段落中出現的實體並給出一個合理的答案。同時給工作人員展示 SQuAD 1.1 中的問題作為參考，儘量使難以回答的那些問題與可回答的問題相似。

文章評估了三種現有的模型架構在兩個資料集上的表現，讓這些模型不僅去學習答案的分佈，而且也去預測一個問題是不可回答問題的機率。當模型預測某個問題無法回答的機率超過某個閾值時，模型就放棄學習答案分佈。下表展示了三個模型在兩個資料集 (SQuAD 1.1 和 SQuAD 2.0) 上的表現，結果顯示：

表現最好的模型 (DocQA + ELMo) 在 SQuAD 2.0 上與人類仍有 23.2 的差距，意味著模型有很大的改進空間；

在兩個資料集上運用相同模型架構，相比於 SQuAD1.1，最優模型和人的 F1 值差距在 SQuAD 2.0 上更大，說明對現有模型來說 SQuAD 2.0 是一個更難學習的資料。

System	SQuAD 1.1 test		SQuAD 2.0 dev		SQuAD 2.0 test	
	EM	F1	EM	F1	EM	F1
BNA	68.0	77.3	59.8	62.6	59.2	62.1
DocQA	72.1	81.0	61.9	64.8	59.3	62.3
DocQA + ELMo	78.6	85.8	65.1	67.6	63.4	66.3
Human	82.3	91.2	86.3	89.0	86.9	89.5
Human-Machine Gap	3.7	5.4	21.2	21.4	23.5	23.2

文章在 SQuAD 1.1 資料集上利用 TFIDF 和規則隨機生成了一些難以回答的問題，仍採用相同的模型進行對比。結果顯示(如下表)最好的模型還是在 SQuAD 2.0 資料集上表現最低，再次證明了 SQuAD 2.0 對現有的語言理解模型來說是一個有難度的挑戰。

System	SQuAD 1.1 + TFIDF		SQuAD 1.1 + RULEBASED		SQuAD 2.0 dev	
	EM	F1	EM	F1	EM	F1
BNA	72.7	76.6	80.1	84.8	59.8	62.6
DocQA	75.6	79.2	80.8	84.8	61.9	64.8
DocQA + ELMo	79.4	83.0	85.7	89.6	65.1	67.6

研究結果：

文章證明了 SQuAD 2.0 是一個具有挑戰性的、多樣化的、大規模的資料集，它迫使模型去學習什麼情況下一個問題在給定的環境中是無法回答的。我們有理由相信，SQuAD 2.0 將會促進新的閱讀理解模型的發展，這些模型能夠知道他們不知道的東西是什麼，從而能在更深層次上理解語言文字。

論文題目： *GLUE: A MULTI-TASK BENCHMARK AND ANALYSIS PLATFORM FOR NATURAL LANGUAGE UNDERSTANDING*

中文題目：GLUE: 一個用於自然語言理解的多工基準測試和分析平臺

論文作者：Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy & Samuel R. Bowman

論文出處：Proceedings of the International Conference on Learning Representations (ICLR). 2019

論文地址：<https://arxiv.org/abs/1804.07461>

研究問題：

人類理解語言的能力是靈活的、強健的。相比之下，單詞級以上的大多數自然語言理解 (Natural Language Understanding, NLU) 模型都是為特定的任務設計的。我們期望開發一個能夠學習在不同領域執行一系列不同語言任務的更統一的模型，它必須能夠以一種不局限於單個任務、類型或資料集的方式來理解人類的語言。為了實現這一目標，文章設計了一個通用語言理解評估基準 (General Language Understanding Evaluation, GLUE) 用於評估模型在不同的現有 NLU 任務集上的性能。

研究方法：

文章設計了一個通用語言理解評估基準 (General Language Understanding Evaluation, GLUE)，它包含一組 NLU 任務，包括問答系統、情感分析和文本蘊涵，以及一個用於模型評估、比較和分析的線上平臺。GLUE 傾向於讓模型在任務之間共用一般的語言知識。GLUE 還提供了一個人工設計的測試集(診斷集)，可以對模型進行詳細的分析。

Corpus	Train	Test	Task	Metrics	Domain
Single-Sentence Tasks					
CoLA	8.5k	1k	acceptability	Matthews corr.	misc.
SST-2	67k	1.8k	sentiment	acc.	movie reviews
Similarity and Paraphrase Tasks					
MRPC	3.7k	1.7k	paraphrase	acc./F1	news
STS-B	7k	1.4k	sentence similarity	Pearson/Spearman corr.	misc.
QQP	364k	391k	paraphrase	acc./F1	social QA questions
Inference Tasks					
MNLI	393k	20k	NLI	matched acc./mismatched acc.	misc.
QNLI	105k	5.4k	QA/NLI	acc.	Wikipedia
RTE	2.5k	3k	NLI	acc.	news, Wikipedia
WNLI	634	146	coreference/NLI	acc.	fiction books

為了評測這個 GLUE 評估基準，文章在公共資料集上評估了句子表示的 baseline 和最優模型，下表展示了資料集的任務表述和相關的統計資訊。資料集的任務表述和統計資訊如上表所示。

結果顯示如下表所示，針對所有任務的多工訓練比針對每個任務單獨訓練模型的效果更好。然而，最佳模型的低性能表明模型仍存在改進空間。

Model	Single Sentence			Similarity and Paraphrase			Natural Language Inference			
	Avg	CoLA	SST-2	MRPC	QQP	STS-B	MNLI	QNLI	RTE	WNLI
Single-Task Training										
BiLSTM	63.9	15.7	85.9	69.3/79.4	81.7/61.4	66.0/62.8	70.3/70.8	75.7	52.8	<u>65.1</u>
+ELMo	66.4	35.0	<u>90.2</u>	69.0/80.8	85.7/65.6	64.0/60.2	72.9/73.4	71.7	50.1	<u>65.1</u>
+CoVe	64.0	14.5	88.5	<u>73.4/81.4</u>	83.3/59.4	<u>67.2/64.1</u>	64.5/64.8	75.4	<u>53.5</u>	<u>65.1</u>
+Attn	63.9	15.7	85.9	68.5/80.3	83.5/62.9	59.3/55.8	74.2/73.8	<u>77.2</u>	51.9	<u>65.1</u>
+Attn, ELMo	<u>66.5</u>	35.0	<u>90.2</u>	68.8/80.2	86.5/66.1	55.5/52.5	76.9/76.7	76.7	50.4	<u>65.1</u>
+Attn, CoVe	63.2	14.5	88.5	68.6/79.7	84.1/60.1	57.2/53.6	71.6/71.5	74.5	52.7	<u>65.1</u>
Multi-Task Training										
BiLSTM	64.2	11.6	82.8	74.3/81.8	84.2/62.5	70.3/67.8	65.4/66.1	74.6	57.4	<u>65.1</u>
+ELMo	67.7	32.1	89.3	78.0/84.7	82.6/61.1	67.2/67.9	70.3/67.8	75.5	57.4	<u>65.1</u>
+CoVe	62.9	18.5	81.9	71.5/78.7	84.9/60.6	64.4/62.7	65.4/65.7	70.8	52.7	<u>65.1</u>
+Attn	65.6	18.6	83.0	76.2/83.9	82.4/60.1	72.8/70.5	67.6/68.3	74.3	58.4	<u>65.1</u>
+Attn, ELMo	70.0	<u>33.6</u>	90.4	78.0/84.4	84.3/63.1	<u>74.2/72.3</u>	<u>74.1/74.5</u>	79.8	<u>58.9</u>	<u>65.1</u>
+Attn, CoVe	63.1	8.3	80.7	71.8/80.0	83.4/60.5	69.8/68.4	68.1/68.6	72.9	56.0	<u>65.1</u>
Pre-Trained Sentence Representation Models										
CBoW	58.9	0.0	80.0	73.4/81.5	79.1/51.4	61.2/58.7	56.0/56.4	72.1	54.1	<u>65.1</u>
Skip-Thought	61.3	0.0	81.8	71.7/80.8	82.2/56.4	71.8/69.7	62.9/62.8	72.9	53.1	<u>65.1</u>
InferSent	63.9	4.5	<u>85.1</u>	74.1/81.2	81.7/59.1	75.9/75.3	66.1/65.7	72.7	58.0	<u>65.1</u>
DisSent	62.0	4.9	83.7	74.1/81.7	82.6/59.5	66.1/64.8	58.7/59.1	73.9	56.4	<u>65.1</u>
GenSen	<u>66.2</u>	<u>7.7</u>	83.1	<u>76.6/83.0</u>	<u>82.9/59.8</u>	79.3/79.2	<u>71.4/71.3</u>	<u>78.6</u>	59.2	<u>65.1</u>

研究結果：

首先，文章實現了一個通用語言理解評估基準（GLUE 基準），包含 9 個句子或句對的 NLU 任務。所有任務建立在帶標注的資料集上，資料集覆蓋了各種文本類型、不同資料規模和不同難度係數。其次，建立了一個主要基於私有評測資料的線上模型評估平臺。該平臺與模型無關，並且可以評估任何能夠在所有 9 個任務上產生結果的模型。然後，文章還構建了專門的診斷評價資料集，以用作誤差分析、模型的定性比較以及對抗性資料的補充。最後，文章實驗了句子表示學習的幾種主要現有方法的結果。

論文題目：*Linguistically-Informed Self-Attention for Semantic Role Labeling*

中文題目：用於語義角色標注的基於語言學資訊的自我注意力方法

論文作者：Emma Strubell, Patrick Verga, Daniel Andor, David Weiss and Andrew McCallum

論文出處：Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing

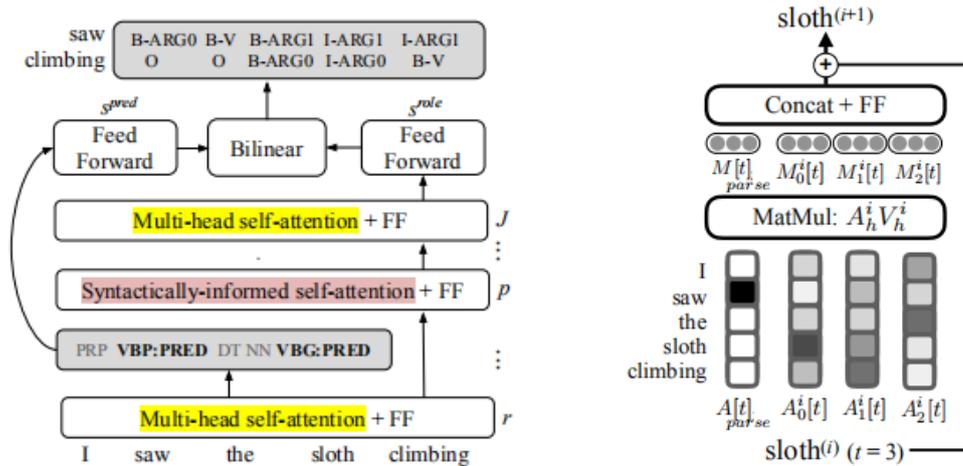
論文地址：<https://www.aclweb.org/anthology/D18-1548/>

研究問題：

語義角色標注 (Semantic Role Labeling, SRL) 是一種提取文本高級表示的技術。目前最先進的基於深度神經網路的語義角色標注模型並沒有顯式使用文本的語言特徵。然而，有研究已經指出，語法樹可以有效改善 SRL 任務。鑒於此，文章提出了一種用於語義角色標注的基於語言學的自注意力方法 (linguistically-informed self-attention , LISA) 。該模型將多頭自注意力機制 (multi-head self-attention) 與多工學習相結合，包括句法依賴解析、詞性標注、謂詞檢測和語義角色標記。與先前需要大量預處理來準備語言特徵的模型不同，LISA 可以僅使用原始的 token 對序列進行一次編碼，來同時執行多個預測任務。

研究方法：

文章設計了一個高效的利用語言資訊有效執行端到端語義角色標注任務的神經網路模型。該模型融合了神經網路的注意力機制預測句法依賴關係，並在 4 個相關任務上進行了多工學習。下圖 (左) 是模型結構：詞向量輸入到具有多頭注意力機制的 J 層，在 p 層訓練自注意力機制去關注父節點的語義資訊。P 層的詳細操作見下圖 (右) 。



具體地，模型的基礎是一個多頭自注意力機制的 token 編碼，基於 ELMo 預訓練得到的 token 表示結合一個由正弦函數確定的位置編碼向量學習 H 個注意力頭，拼接後組成每個 token 的最終自注意力表示。句法資訊的自注意力機制將注意力中的 key、value 和 query 均增加語義解析資訊，其中 key、value 分別對應父節點和依賴關係。最後，共用模型中較低層的參數來預測 POS 詞性標記和謂詞。

文章把 LISA 模型和四個先進的基線方法比較，下表的結果顯示，未加入語義資訊的模型已經實現了現有的最優模型性能（如紅框所示），當 LISA 加入自己的語義解析時模型性能並沒有很大的提升（如綠框所示），但當再加入了目前最優的 D&M 語義解析以後，模型性能有了較大的提升。Gold 表示加入了最優的語義解析，以期模型能有更好的表現。

GloVe	Dev			WSJ Test			Brown Test		
	P	R	F1	P	R	F1	P	R	F1
He et al. (2017) PoE	81.8	81.2	81.5	82.0	83.4	82.7	69.7	70.5	70.1
He et al. (2018)	81.3	81.9	81.6	81.2	83.9	82.5	69.7	71.9	70.8
SA	83.52	81.28	82.39	84.17	83.28	83.72	72.98	70.1	71.51
LISA	83.1	81.39	82.24	84.07	83.16	83.61	73.32	70.56	71.91
+D&M	84.59	82.59	83.58	85.53	84.45	84.99	75.8	73.54	74.66
+Gold	87.91	85.73	86.81	—	—	—	—	—	—
ELMo									
He et al. (2018)	84.9	85.7	85.3	84.8	87.2	86.0	73.9	78.4	76.1
SA	85.78	84.74	85.26	86.21	85.98	86.09	77.1	75.61	76.35
LISA	86.07	84.64	85.35	86.69	86.42	86.55	78.95	77.17	78.05
+D&M	85.83	84.51	85.17	87.13	86.67	86.90	79.02	77.49	78.25
+Gold	88.51	86.77	87.63	—	—	—	—	—	—

Table 1: Precision, recall and F1 on the CoNLL-2005 development and test sets.

研究結果：

文章提出了一種多工神經網路模型，該模型有效地融合了豐富的語言資訊用於語義角色標注。通過一系列實驗證明了 LISA 的性能優於最先進的現有模型。具體實驗結果：在 CoNLL-2005SRL 資料集上，LISA 模型在謂詞預測、詞嵌入任務上比當前最好的演算法在 F1 值上高出了 2.5 (新聞專線資料) 和 3.5 以上 (其他領域資料)，減少了約 10% 的錯誤。在 ConLL-2012 英文角色標記任務上，該方法也獲得了 2.5F1 值的提升。LISA 同時也比當前最好的基於上下文的詞表示學習方法 (ELMo) 高出了 1.0 的 F1 (新聞專線資料) 和多於 2.0 的 F1 值 (其他領域資料)。

論文題目： *OpenKiwi: An Open Source Framework for Quality Estimation*

中文題目: OpenKiwi: 一個用於品質評估的開源框架

論文作者：Fabio Kepler、Jonay Trenous、Marcos Treviso、Miguel Vera、Andre F. T. Martins

論文出處：Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019

論文地址：<https://arxiv.org/abs/1902.08646>

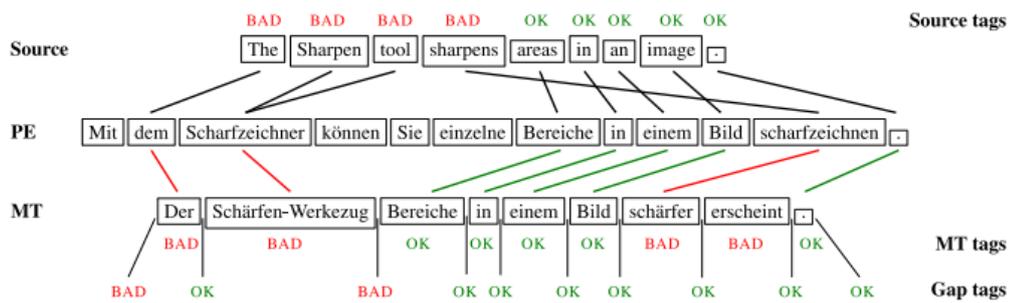
研究問題：

文章介紹了一個基於 PyTorch 的用於翻譯品質評估的開源框架——OpenKiwi。該框架支援單詞級和句子級的品質評估系統的訓練和測試，實現和集成了 WMT 2015-18 品質評估比賽的獲獎系統。文章在 WMT2018 的兩個資料集 (English-German SMT and NMT) 上對 OpenKiwi 進行基準測試。實驗結果證

明了，該框架在單詞級任務上實現了最先進的性能，在句子級任務上實現了幾乎最先進的性能。

研究方法：

品質評估(Quality Estimation, QE)提供了機器翻譯和人工翻譯之間缺失的一環，目標是在沒有參考譯文的情況下評估翻譯系統的品質。句子級的品質評估旨在預測整個翻譯句子的品質，如基於人後期編輯所花費的時間，或者修改它需要的編輯運算元。單詞級別品質評估的目標是給機器翻譯的每個詞、單詞之間的間隙（根據上下文需要插入的）和來源語言單詞(原句中被錯譯或省略的單詞)分配品質標籤。下圖是一個單詞級標注示例。



文章研發的 OpenKiwi(<https://github.com/Unbabel/OpenKiwi>)實現和集成了 WMT 2015-18 品質評估比賽的獲獎系統，且允許輕鬆地添加和運行新模型，而不需要過多地關注輸入資料處理、輸出生成和評估。OpenKiwi 基於 PyTorch 深度學習框架實現，可以單獨運行或通過 API 集成到其他項目。此外，OpenKiwi 提供了根據 WMT2018 資料的預訓練模型，並支援根據新資料訓練新 QE 模型。

文章對該框架進行了基準測試，使用了 WMT 2018 品質評估比賽的資料集，結果顯示，這些系統的集成版本表現最好，堆疊的架構在預測單詞級標籤方面非

常有效。文章還比較了另一個現有的開源工具 deepQuest，在單詞級和句子級均獲得更優的表現。

研究結果：

文章介紹了一個新的機器翻譯品質評估 (QE) 開源框架—OpenKiwi。OpenKiwi 是在 PyTorch 中實現的，並支援在新資料上訓練單詞級和句子級的 QE 系統。它在單詞級和句子級上都優於其他開源工具包，並產生了新的最先進的單詞級 QE 結果。OpenKiwi 一經發佈就作為 WMT 2019 QE 的基線系統。此外，所有 WMT 2019 QE 的單詞、句子和文檔級任務的獲獎系統都使用 OpenKiwi 作為其構建基礎。

論文題目：*Bridging the Gap between Training and Inference for Neural Machine Translation*

中文題目：架起一座在基於神經元的機器翻譯訓練和推理之間的橋樑

論文作者：Wen Zhang, Yang Feng, Fandong Meng, Di You, Qun Liu

論文出處：Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. (2019).

論文地址：<https://arxiv.org/abs/1906.02448>

研究問題：

神經機器翻譯(Neural Machine Translation, NMT)是根據上下文預測下一個詞，從而依次生成目標詞。訓練時用真值詞作為上下文進行預測，而推理時必須從頭開始生成整個序列，這會導致誤差的積累。此外，單詞級訓練要求生成的序列與

真值序列嚴格匹配會導致對不同但合理的翻譯的過度矯正。針對這一問題，文章提出了一種過矯正恢復的方法。該方法不僅從真值序列中提取上下文，而且通過訓練模型從預測序列中提取上下文，即翻譯過程中模型不需要再逐詞對比標準來確定損失函數。在中文→英文和英語→德語的翻譯任務的實驗結果表明，該方法可以在多個資料集上實現顯著的改進。

研究內容：

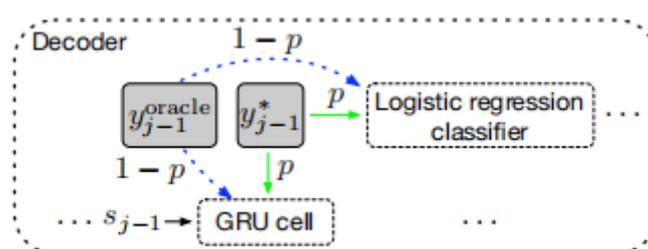


Figure 1: The architecture of our method.

模型主要結構如上圖所示，核心思想是：不僅使用真值序列進行約束，在訓練過程中，也利用訓練模型預測出的上一個詞語作為其中的備選詞語來約束模型。對於 oracle 詞的選擇有兩種方法，一種是用貪心搜索演算法在詞級上選擇，另一種是在句級上選擇最優的 oracle 序列。

在詞語級的選擇上，在時間步為 j 時，獲取 $j-1$ 時間步模型預測出的每個詞語的預測分數。為了提高模型的魯棒性，在預測分數的基礎上加上 Gumbel noise，取分數最高的詞語作為此時的 Oracle Word。

在句子級的選擇上，使用集束搜索，選擇集束寬為 k 的句子，然後計算每個句子的 BLEU 分數，選擇分數最高的句子。對於生成的實際句子超出或短于這一長度的情況，文章採用強制解碼的方式進行干預。

最終選擇的 Oracle Word 也會和真值序列的詞語混合，然後使用衰減式採樣 (Decay Sampling) 的方法從中挑選出作為約束模型訓練的詞。

Systems	Architecture	MT03	MT04	MT05	MT06	Average
<i>Existing end-to-end NMT systems</i>						
Tu et al. (2016)	Coverage	33.69	38.05	35.01	34.83	35.40
Shen et al. (2016)	MRT	37.41	39.87	37.45	36.80	37.88
Zhang et al. (2017)	Distortion	37.93	40.40	36.81	35.77	37.73
<i>Our end-to-end NMT systems</i>						
this work	RNNsearch	37.93	40.53	36.65	35.80	37.73
	+ SS-NMT	38.82	41.68	37.28	37.98	38.94
	+ MIXER	38.70	40.81	37.59	38.38	38.87
	+ OR-NMT	40.40^{†‡*}	42.63^{†‡*}	38.87^{†‡*}	38.44[‡]	40.09
	Transformer	46.89	47.88	47.40	46.66	47.21
	+ word oracle	47.42	48.34	47.89	47.34	47.75
	+ sentence oracle	48.31*	49.40*	48.72*	48.45*	48.72

Table 1: Case-insensitive BLEU scores (%) on Zh→En translation task. “†”, “‡”, “*” and “*” indicate statistically significant difference ($p < 0.01$) from RNNsearch, SS-NMT, MIXER and Transformer, respectively.

文章對 NIST 中文→英文 (Zh→En) 和 WMT14 英語→德語 (En→De) 的翻譯任務進行了實驗。結果表明，文章提出的方法可以在多個資料集上實現提升。同時在 RNNsearch 模型和 Transformer 模型上也驗證了該方法。結果表明，新方法可以顯著提高兩種模型的性能。

研究結果：

端到端的 NMT 模型在訓練時逐字逐句地生成翻譯，將真實單詞作為上下文，而不是將模型生成的前一個單詞作為上下文進行推理。為了減少訓練和推理之間的差異，在預測一個單詞時，文章使用抽樣方法將真實單詞或先前預測的單詞作為上下文輸入。被預測的單詞 (稱為 oracle 單詞) 可以通過單詞級或句子級優化生成。與詞級 oracle 相比，句子級 oracle 進一步賦予了該模型過度矯正恢復的能力。通過兩個基線模型和實際翻譯任務的相關工作驗證了該方法的有效性，並對所有資料集進行了顯著的改進。文章還指出，句子級的 oracle 優於單詞級別的 oracle。

論文題目： *Do you know that Florence is packed with visitors? Evaluating state-of-the-art models of speaker commitment*

中文題目：你知道佛羅倫斯到處都是遊客嗎?評估說話者結論確定性的最新模型

論文作者：Nanjiang Jiang, Marie-Catherine de Marneffe

論文出處：Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019.

論文地址：<https://www.aclweb.org/anthology/P19-1412/>

研究問題：

當一個演講者問“你知道佛羅倫斯擠滿了遊客嗎”時，我們可能會相信，但如果她問的是“你認為佛羅倫斯擠滿了遊客嗎”，我們可能就不會相信了。推斷說話者承諾（或稱事件真實性）對於資訊抽取和問答均至關重要。對說話者承諾的預測，是判斷說話者在句子中對某一事件承諾到何種程度，是實際的、非實際的還是不確定的。本篇文章通過在資料集上分析模型誤差的語言相關性，來探討語言缺陷會導致說話者承諾模型出現錯誤模式的假設。

研究內容：

評估資料集選取的 CommitmentBank 包含 1200 條資料，涉及四種包含時態嵌入動詞的取消蘊涵環境(否定句、情態動詞、疑問句、條件句)。對於每條資料，從 Mechanical Turk 尋找至少 8 個母語為英語的人標注其說話者承諾。

(1)	Context	The answer is no, no no. Not now, not ever.
	Target	<i>I never believed</i> <i>I could wish anyone dead</i> but last night changed all that. <small>Gold: 1.56, Rule-based: 3.0, Hybrid: 0.50</small>
(2)	Context	Revenue is estimated at \$18.6 million. The maker of document image processing equipment said the state procurement division had declared FileNet in default on its contract with the secretary of state uniform commercial code division.
	Target	FileNet said it <i>doesn't believe</i> the state has a valid basis of default and is reviewing its legal rights under the contract, but said it can't predict the outcome of the dispute. <small>Gold: 0.47, Rule-based: 3.0, Hybrid: 1.08</small>
(3)	Context	A: Yeah, that's crazy. B: and then you come here in the Dallas area, um.
	Target	<i>I don't believe</i> that <i>people should be allowed to carry guns in their vehicles</i> . <small>Gold: 2.64, Rule-based: 3.0, Hybrid: 1.40</small>

Table 1: Examples from the CommitmentBank, with gold scores and predictions from rule-based and hybrid models. **Embedding verbs in bold**, entailment-canceling environments italicized. The gold score is the mean annotators' speaker commitment judgments towards the content of the complement.

文章評估了兩種最先進的說話者承諾模型：Stanovsky 等人提出的基於規則的方法和 Rudinger 等人提出的神經網路方法，結果顯示基於規則的模型表現得更好，但整體表現均不是很好，因為 CommitmentBank 與其他任何資料集相比，其相關性更低，絕對錯誤率更高。

為了更好地解釋模型的輸出，文章在分類實驗中對它們進行了評估。使用高斯混合模型來獲得三個平均得分的聚類，用兩個模型進行預測。將均值最高的聚類標記為真 (+)，均值最低的聚類標記為假 (-)，剩餘真值不確定 (o)。結果顯示兩類模型對 o 類都沒有預測能力。

	Precision		Recall		F1		Count
	Rule	Hybr.	Rule	Hybr.	Rule	Hybr.	
+	0.58	0.64	0.91	0.51	0.71	0.56	251
-	0.99	0.67	0.55	0.20	0.70	0.31	268
o	0.00	0.06	0.00	0.46	0	0.11	37
Total	0.74	0.61	0.67	0.35	0.66	0.41	556

Table 4: Classification performance of the models.

研究結果：

文章在 CommitmentBank 上評估了兩種最先進的說話者承諾模型。研究發現，帶有語言學知識的模型比基於 LSTM 的模型表現更好，這表明如果要在這樣的有挑戰性的自然語言資料中捕捉說話者承諾資訊的話，語言學知識是不可避免的。根據語言特徵對標注資料進行分類可以發現不對稱的錯誤模式。雖然模型在某些情況下（如否定）取得了良好的性能，但很難泛化到其他更豐富的自然語言的語言學結構（如條件句），從而指出了研究的改進方向。

5.5 自然語言處理最新進展

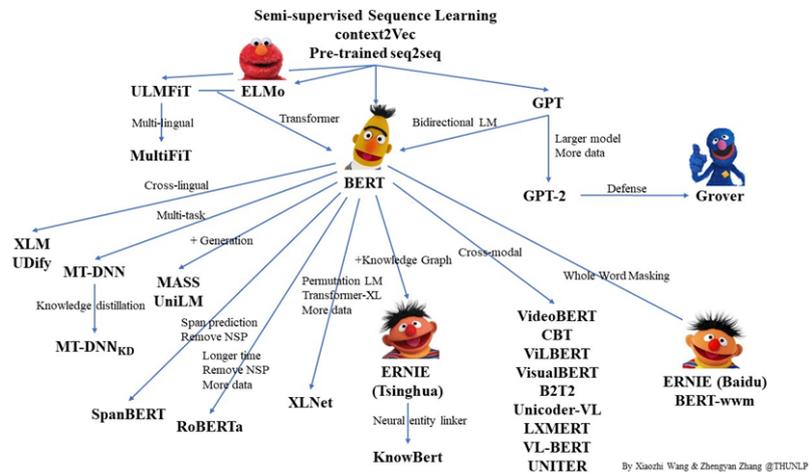
近年來，預訓練語言模型在自然語言處理領域有了重要進展。預訓練模型指的是首先在大規模無監督的語料上進行長時間的無監督或者是自監督的預先訓練（pre-training），獲得通用的語言建模和表示能力。之後在應用到實際任務上時對模型不需要做大的改動，只需要在原有語言表示模型上增加針對特定任務獲得輸出結果的輸出層，並使用任務語料對模型進行少許訓練即可，這一步驟被稱作微調（fine tuning）。

自 ELMo、GPT、BERT 等一系列預訓練語言表示模型（Pre-trained Language Representation Model）出現以來，預訓練模型在絕大多數自然語言處理任務上都展現出了遠遠超過傳統模型的效果，受到越來越多的關注，是 NLP 領域近年來最大的突破之一，是自然語言處理領域的最重要進展。

BERT（Bidirectional Encoder Representation from Transformer）是 Google AI 於 NAACL2019 提出的一個預訓練語言模型。BERT 的創新點是提出了有效的無

監督預訓練任務，從而使得模型能夠從無標注語料中獲得通用的語言建模能力。

模型的部分細節在前文的論文解讀中已經給出，不再贅述。



BERT 之後湧現了許多對其進行擴展的模型（如上圖所示），包括：跨語言預訓練的 XLM 和 UDify，跨模態預訓練的模型，融合知識圖譜的 ERNIE，將 seq2seq 等語言生成任務整合入 BERT 類模型的 MASS, UniLM 等。其中幾個重要的進展包括：

(1) XLNet 使用 Transformer-XL 替代了 Transformer 作為基礎模型，擁有編碼超長序列的能力。XLNet 提出了一個新的預訓練語言任務：Permutation Language Modeling（排列語言模型），模型將句子內的詞語打亂順序，從而使得預測當前詞語時可以利用雙向資訊。XLNet 相對 BERT 也使用了更多的語料。

(2) RoBERTa 採用了與 BERT 具有相同的模型結構，同樣採用了遮罩語言模型任務進行預訓練，但捨棄了 BERT 中下句預測模型。此外，RoBERTa 採用了更大規模的資料和更魯棒的優化方法，從而取得了更好的表現。

(3) ALBERT 模型針對 BERT 參數量過大難以訓練的問題做了優化，一是對詞向量矩陣做分解，二是在層與層之間共用參數。此外，ALBERT 將下句預測模型替換為句序預測任務，即給定一些句子預測它們的排列順序。

AMiner

6 語音辨識

6.1 語音辨識概念

語音辨識是讓機器識別和理解說話人語音信號內容的新興學科，目的是將語音信號轉變為文本字元或者命令的智慧技術，利用電腦理解講話人的語義內容，使其聽懂人類的語音，從而判斷說話人的意圖，是一種非常自然和有效的人機交流方式。它是一門綜合學科，與很多學科緊密相連，比如語言學、信號處理、電腦科學、心理和生理學等^[8]。

語音辨識首先要對採集的語音信號進行預處理，然後利用相關的語音信號處理方法計算語音的聲學參數，提取相應的特徵參數，最後根據提取的特徵參數進行語音辨識。總體上，語音辨識包含兩個階段：第一個階段是學習和訓練，即提取語音庫中語音樣本的特徵參數作為訓練資料，合理設置模型參數的初始值，對模型各個參數進行重估，使識別系統具有最佳的識別效果；第二個階段就是識別，將待識別語音信號的特徵根據一定的準則與訓練好的範本庫進行比較，最後通過一定的識別演算法得出識別結果。顯然識別結果的好壞與範本庫是否準確、模型參數的好壞以及特徵參數的選擇都有直接的關係。

實際上，語音辨識也是一種模式識別，其基本結構如下圖所示。和一般模式識別過程相同，語音辨識包括如圖所示 3 個基本部分。實際上，由於語音資訊的複雜性以及語音內容的豐富性，語音辨識系統要比模式識別系統複雜的多。

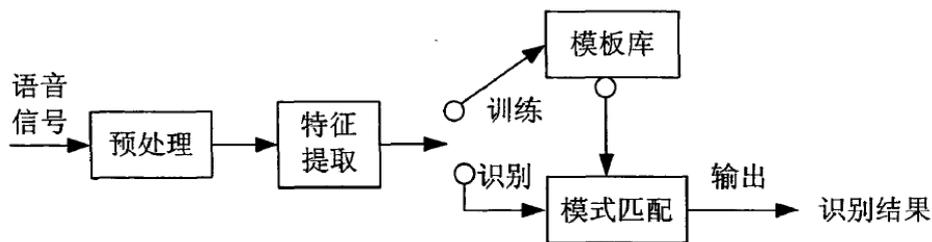


圖 6-1 語音辨識系統框架

其中，預處理主要是對輸入語音信號進行預加重和分段加窗等處理，並濾除其中的不重要資訊及背景雜訊等，然後進行端點檢測，以確定有效的語音段。特徵參數提取是將反映信號特徵的關鍵資訊提取出來，以此降低維數減小計算量，用於後續處理，這相當於一種資訊壓縮。之後進行特徵參數提取，用於語音訓練和識別。常用的特徵參數有基於時域的幅度、過零率、能量以及基於頻域的線性預測倒譜系數、Mel 倒譜系數等。

6.2 語音辨識發展歷史

語音辨識的研究工作可以追溯到 20 世紀 50 年代。在 1952 年，AT & T 貝爾研究所的 Davis, Biddulph 和 Balashek 研究成功了世界上第一個語音辨識系統 Audry 系統，可以識別 10 個英文數位發音。這個系統識別的是一個人說出的孤立數位，並且很大程度上依賴於每個數字中的母音的共振峰的測量。1956 年，在 RCA 實驗室，Olson 和 Belar 研製了可以識別一個說話人的 10 個單音節的系統，它同樣依賴於母音帶的譜的測量。1959 年，英國的 Fry 和 Denes 研製了一個能夠識別 4 個母音和 9 個輔音的識別器，他們採用了譜分析儀和模式匹配器。所不同的是他們對音素的序列做了限制（相當於現在的語法規則），以此來增加字識別的準確率。但當時存在的問題是的理論水準不夠，都沒有取得非常明顯的成功。

60 年代，電腦的應用推動了語音辨識技術的發展，使用了電子電腦進行語音辨識，提出了一系列語音辨識技術的新理論——動態規劃線性預測分析技術，較好的解決了語音信號產生的模型問題。該理論主要有三項研究成果。首先是美國新澤西州普林斯頓 RCA 實驗室的 Martin 和他的同事提出一種基本的時間歸一化方法，這種方法有效的解決了語音事件時間尺度的非均勻性，能可靠的檢測到語音的起始點和終止點，有效地解決了識別結果的可變性。其次，蘇聯的 Vintsyuk 提出了用動態規劃的方法將兩段語音的時間對齊的方法，這實際上是動態時間規整 (Dynamic Time Warping) 方法的最早版本，儘管到了 80 年代才為外界知曉。第三個是卡耐基梅隆大學的 Reddy 採用的是音素的動態跟蹤的方法，開始了連續語音辨識的研究工作，為後來的獲得巨大成功的連續語音辨識奠定了基礎。

70 年代，語音辨識研究取得了重大的具有里程碑意義的成果，伴隨著自然語言理解的研究以及微電子技術的發展，語音辨識領域取得了突破性進展。這一時期的語音辨識方法基本上是採用傳統的模式識別策略。其中蘇聯的 Velichko 和 Zagoruyko 的研究為模式識別應用於語音辨識這一領域奠定了基礎；日本的迫江和千葉的研究則展示了如何利用動態規劃技術在待識語音模式與標準語音模式之間進行非線性時間匹配的方法；日本的板倉的研究則提出了如何將線性預測分析技術加以擴展，使之用於語音信號的特徵抽取的方法。同時，這個時期還提出了向量量化和隱瑪律可夫模型理論。

80 年代，語音辨識研究進一步走向深入。這一時期所取得的重大進展有：

- (1) 隱瑪律科夫模型 (HMM) 技術的成熟和不斷完善，並最終成為語音辨識的主流方法。
- (2) 以知識為基礎的語音辨識的研究日益受到重視。在進行連續語音辨識的時候，除了識別聲學資訊外，更多地利用各種語言知識，諸如構詞、句

法、語義、對話背景等方面的知識來幫助進一步對語音辨識和理解。同時在語音辨識研究領域，還產生了基於統計概率的語言模型。(3)人工神經網路(ANN)在語音辨識中的應用研究的興起。ANN 具有較好的區分複雜分類邊界的能力，顯然它十分有助於模式識別。在這些研究中，大部分採用基於反向傳播演算法(BP 演算法)的多層感知網路^[9]。

20 世紀 90 年代，語音辨識技術逐漸走向實用化，在建立模型、提取和優化特徵參數方面取得了突破性的進展，使系統具有更好的自我調整性。許多發達國家和著名公司都投入大量資金用以開發和研究實用化的語音辨識產品，從而許多具有代表性的產品問世。比如 IBM 公司研發的漢語 ViaVoice 系統，以及 Dragon 公司研發的 DragonDictate 系統，都具有說話人自我調整能力，能在用戶使用過程中不斷提高識別率。

21 世紀之後，深度學習技術極大的促進了語音辨識技術的進步，識別精度大大提高，應用得到廣泛發展。2009 年，Hinton 將深度神經網路(DNN)應用於語音的聲學建模，在 TIMIT 上獲得了當時最好的結果。2011 年底，微軟研究院的俞棟、鄧力又把 DNN 技術應用在了大詞彙量連續語音辨識任務上，大大降低了語音辨識錯誤率。從此語音辨識進入 DNN-HMM 時代。DNN 帶來的好處是不再需要對語音資料分佈進行假設，將相鄰的語音幀拼接又包含了語音的時序結構資訊，使得對於狀態的分類概率有了明顯提升。同時 DNN 還具有強大環境學習能力，可以提升對雜訊和口音的魯棒性。

目前，語音辨識技術已逐漸被應用於工業、通信、商務、家電、醫療、汽車電子以及家庭服務等各個領域。例如，現今流行的手機語音助手，就是將語音辨

識技術應用到智慧手機中，能夠實現人與手機的智慧對話功能。其中包括美國蘋果公司的 Siri 語音助手，智慧 360 語音助手，百度語音助手等^[10]。

6.3 人才概況

● 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為語音辨識領域全球學者分佈情況：



圖 6-2 語音辨識領域全球學者分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量優勢明顯且主要分佈在其東西海岸；亞洲也有較多的人才分佈，主要在我國東部及日韓地區；歐洲的人才主要集中在歐洲中西部；其他諸如非洲、南美洲等地區的學者非常稀少；語音辨識領域的人才分佈與各地區的科技、經濟實力情況大體一致。

此外，在性別比例方面，語音辨識領域中男性學者占比 87.3%，女性學者占比 12.7%，男性學者占比遠高於女性學者。

語音辨識領域學者的 h-index 分佈如下圖所示，大部分學者的 h-index 分佈在中間區域，其中 h-index 在 30-40 區間的人數最多，有 752 人，占比 37.3%，小於 20 區間的人數最少，只有 6 人。

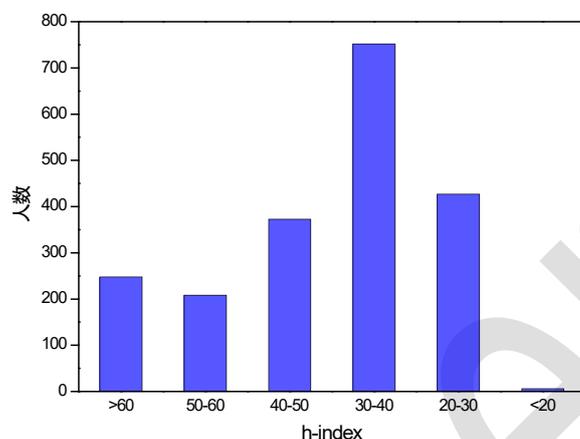


圖 6-3 語音辨識領域學者 h-index 分佈

- 中國人才分佈

我國專家學者在語音辨識領域的分佈如下圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是長三角和珠三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，通過觀察中國周邊國家的學者數量情況，特別是與日韓、東南亞等亞洲國家相比，中國在語音辨識領域學者數量較多且有一定的優勢。



圖 6-4 語音辨識領域中國學者分佈

中國與其他國家在語音辨識領域的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 6-1 語音辨識領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	922	14529	16	1548
中國-英國	207	3088	15	358
中國-新加坡	131	1788	14	221
中國-澳大利亞	92	577	6	194
中國-加拿大	84	921	11	165

中國-法國	76	1318	17	132
中國-日本	75	921	12	151
中國-德國	68	1099	16	110
中國-丹麥	32	501	16	31
中國-巴基斯坦	25	772	31	40

從上表資料可以看出，中美合作的論文數、引用數、學者數遙遙領先，表明中美間在語音辨識領域合作之密切；此外，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 4 席；中國與巴基斯坦合作的論文數雖然不是最多，但是擁有最高的平均引用數說明在合作品質上中巴合作達到了較高的水準。

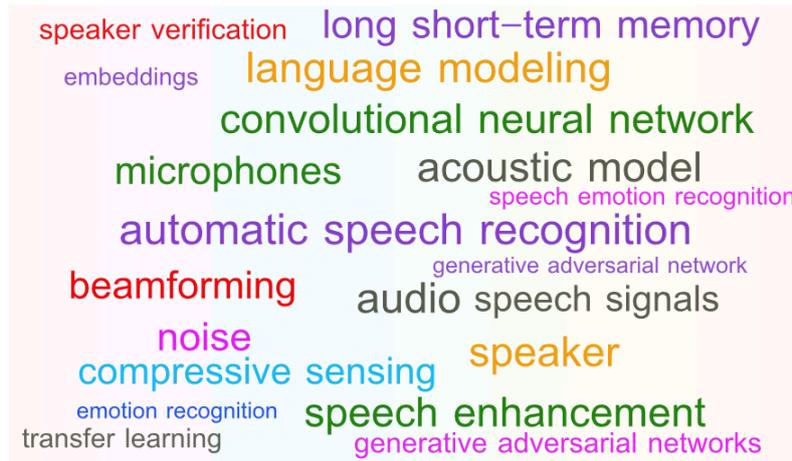
6.4 論文解讀

本節對本領域的高水準學術會議及期刊論文進行挖掘，解讀這些會議和期刊在 2018-2019 年的部分代表性工作。這些會議和期刊包括：

IEEE International Conference on Acoustics, Speech and Signal Processing

IEEE Transactions on Audio, Speech, and Language Processing

我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如下圖所示。其中，雜訊 (noise)、語言模型 (language modeling)、音訊 (audio) 是本領域中最熱的關鍵字。



論文題目： *X-Vectors: Robust DNN Embeddings for Speaker Recognition*

中文題目：X 向量：用於說話人識別的魯棒 DNN 嵌入

論文作者：David Snyder, Daniel Garcia-Romero, Gregory Sell, Daniel Povey and Sanjeev Khudanpur. X-Vectors: Robust DNN Embeddings for Speaker Recognition.

論文出處：2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

論文地址：<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8461375>

研究問題：

捕捉說話者特徵是語音辨識領域具有重大意義的研究內容。大多數說話人識別系統都是基於 i-vectors 來實現的。標準的基於 i-vectors 的方法由通用背景模型 (UBM) 和大型投影矩陣 T 組成，該模型以無監督方式來學習。在早期的系統中，神經網路經訓練後，被用來分離說話者，從網路中提取幀級表示，並將其用作高斯說話者模型的特徵。近年來，使用深度神經網路 (DNN) 捕獲說話者特徵是當前非常活躍的研究領域。DNN 嵌入性能也隨著訓練資料量的增加而高度擴展。

研究方法：

在本文中，研究者們使用資料增強來提高用於說話人識別的深度神經網路 (DNN) 嵌入的性能。經過訓練後，用於區分說話者的 DNN 將可變長度話語映射到固定維嵌入，並將其稱為 x 向量。本文使用一系列資料增強的方法，包括增加雜訊、增加混響等，用以增加訓練資料的數量並提高其魯棒性。訓練後，DNN 可對訓練資料中的 N 個說話者進行分類。一個訓練示例包括大量語音特徵 (平均約 3 秒) 和相應的說話者標籤。該模型從圖層 segment6 的仿射成分中提取嵌入。不包括 softmax 輸出層和 segment7 (因為訓練後不需要它們)，總共有 420 萬個參數。基於 i -vector 和 x -vector 的系統均使用了 PLDA 分類器。 x -vector 和 i -vector 需先中心化，其後再使用 LDA 投影。在 SITW 開發中將 LDA 尺寸調整為 i -vector 為 200， x -vector 為 150。降維後，再使用自我調整 s 范數進行長度歸一化和建模。

研究結果：

沒有資料擴充的系統分別在 SWBD 和 SRE 資料集上對提取器進行了訓練。不使用增強，SITW 上的最佳結果是通過 i -vector (BNF) 獲得的結果，比 DCF10-2 處的 x -vector 系統好 12%。與 SITW 上的 x -vector 系統相比，聲學 i -vector 系統還實現了稍低的錯誤率。但是，即使不進行擴展，也可以通過 x -vector 獲得 SRE16 粵語的最佳結果。就 DCF10-2 而言，這些嵌入比任意 i -vector 系統效果好約 14%。使用了資料增強的對比實驗結果表明，PLDA 增強對所有系統都有明顯的改進。 x -vector 可以從 PLDA 增強中獲得比 baseline 系統更高的改進效果。在 SITW 上， x -vector 系統的誤碼率略低於 i -vector (聲學)，但在大多數工作點上

仍落後於 i-vector (BNF)。在 SRE16 上，在 DCF10-2 中，x-vector 比 i-vector 保持約 14% 的優勢。

論文題目： *Boosting Noise Robustness of Acoustic Model via Deep Adversarial Training*

中文題目：通過深度對抗訓練提高聲學模型的雜訊魯棒性

論文作者：Bin Liu, Shuai Nie, Yaping Zhang, Dengfeng Ke, Shan Liang, Wenju Liu
Boosting Noise Robustness of Acoustic Model via Deep Adversarial Training

論文出處：2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

論文地址：<https://ieeexplore.ieee.org/document/8462093>

研究問題：

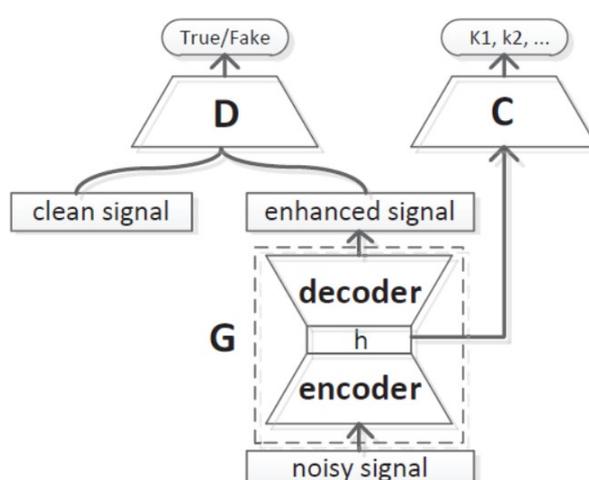
在真實環境中，語音信號很容易受到雜訊和混響的干擾，尤其在遠場條件下，由於聲波在傳播過程中其能量隨傳播距離呈指數衰減，語音信號受到雜訊和混響的干擾更加嚴重，因此自動語音辨識系統性能很難得到有效提升。針對語音辨識系統在雜訊環境下識別性能下降的問題，作者提出了深度對抗聲學模型訓練框架，減小了雜訊環境語音資料和真實訓練資料的分佈差異，從而提升聲學模型的魯棒性。

研究方法：

語音辨識模型的雜訊魯棒性問題主要來源於純淨訓練資料和帶噪測試資料的分佈差異。生成式對抗網路 (Generative Adversarial Networks, GAN) 可以通過對抗訓練的方式，連續逼近指定的資料分佈。GAN 由生成器和判別器組成，生

成器用來生成樣本，判別器用來判斷樣本是否來自真實訓練集。二者進行對抗訓練，使得生成器生成的樣本盡可能逼近真實訓練資料。

針對語音辨識系統在雜訊環境下識別性能下降的問題，作者提出深度對抗和聲學模型聯合訓練的框架，如下圖所示，框架由生成器 (G)、判別器 (D) 以及分類器 (C) 組成。生成器用來把帶噪語音資料分佈變成純淨語音；判別器用來判定語音信號是否來自真實純淨訓練集；聲學模型作為分類器，指導生成器提取區分性特徵。生成器、判別器和聲學模型進行聯合對抗訓練，三者相互配合相互促進。



通過深度對抗的聯合訓練策略，可以減小雜訊環境語音資料和真實訓練資料的分佈差異，提升聲學模型的魯棒性。相對於語音增強方法，該框架沒有增加計算的流程和複雜度，而且不需要一一對應的帶噪資料和純淨資料，可作為通用訓練框架提升已有聲學模型的雜訊魯棒性。

研究結果：

作者使用了 CHiME-4 資料及對提出的方法進行測試，結果表明該方法能夠有效提升聲學模型的魯棒性，在詞錯誤率 (WER) 上相比於基線系統有較大的提升。

論文題目： *Modality Attention for End-to-end Audio-visual Speech Recognition*

中文題目：基於模態注意力的端到端音視覺語音辨識

論文作者：Pan Zhou, Wenwen Yang, Wei Chen, Yanfeng Wang, Jia Jia.

論文出處：2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

論文地址：<https://ieeexplore.ieee.org/document/8683733>

研究問題：

隨著語音辨識的快速發展，純粹靠聲音的識別技術越來越成熟，識別準確率達到 95% 以上，但是在嘈雜環境中語音辨識的準確率會明顯下降。而人在嘈雜環境中不僅靠聲音資訊，還結合講話者的嘴唇和面部視覺資訊來理解講話者的意思。解決嘈雜環境下的語音辨識問題可以通過在語音基礎上加入視覺資訊，通過視聽模態資訊的融合來增強語音辨識的效果 (Automatic Visual Speech Recognition)。

利用視覺資訊來增強語音辨識的效果需要解決兩個難題：一是兩者幀率不同，如何將兩種模態資訊融合在一起，二是如何選擇音訊和視頻的權重。

研究方法：

作者提出一種基於模態重要程度的注意力機制，可以根據模態的資訊含量自我調整調整模態的權重來融合音視覺特徵。

具體方法是分別使用兩個神經網路編碼器對輸入的不同模態的序列進行逐層特徵抽取，得到高層特徵表達。然後，由解碼器分別對不同模態的特徵表達進行注意力計算，得到聲音和視覺模態資訊中對應於當前解碼時刻的上下文向量（context vector）。不同模態的上下文向量經由模態間的注意力自動對不同模態的資訊進行融合，最後輸入到輸出層得到解碼輸出預測。

在編解碼框架下，由於融合的是不同模態的上下文向量，而不是對原始特徵或者編碼器輸出的高層特徵進行直接融合，解決了不同模態的特徵長度不同的問題。

同時，這種模態注意力（Modality Attention）依據不同模態各自的重要程度計算出相應的融合係數，反應了不同模態在當前解碼時刻的不同貢獻度，可以隨著不同解碼時刻的不同信噪比等得到不同的模態融合權重，得到更加魯棒的融合資訊。

研究結果：

作者在 150 小時的電視新聞類音視覺資料上進行了測試，在信噪比為 0dB（信號與雜訊大小相當）時，多模態識別將準確率有很大程度的提高。而且模型在不同雜訊下，體現出了對語音和視頻兩種不同模態間的依賴。隨著雜訊的提升，模型在融合音視覺時，對視覺資訊的依賴比例在逐漸提升。

論文題目：*State-of-the-Art Speech Recognition with Sequence-to-Sequence Models*

中文題目：先進的序列識別語音辨識模型

論文作者：Chung-Cheng Chiu, Tara N. Sainath, Yonghui Wu, Rohit Prabhavalkar, Patrick Nguyen, Zhifeng Chen, Anjuli Kannan, Ron J. Weiss, Kanishka Rao, Ekaterina Gonina, Navdeep Jaitly, Bo Li, Jan Chorowski, Michiel Bacchiani.

論文出處：2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

論文地址：<https://ieeexplore.ieee.org/abstract/document/8462105>

研究問題：

序列到序列模型已經在自動語音辨識 (ASR) 社區中獲得了普及，這是一種將常規 ASR 系統的分離的聲學，發音和語言模型 (AM , PM , LM) 折疊到單個神經網路中的方法。但到目前為止，我們還不清楚這樣的方法是否可以取代當前基於 HMM 的最新技術的神經網路聲學模型。儘管序列到序列模型是完全神經網路化的，無需有限的狀態轉換器、詞典或文本規範化模組。訓練這種模型比傳統的 ASR 系統更簡單：它們不需要決策樹進行引導，也不需要從單獨的系統生成的時間對齊。但是，迄今為止，這些模型都無法在大型詞彙連續語音辨識(LVCSR) 任務上勝過最先進的 ASR 系統。

研究方法：

本文的目的是探索各種結構和優化方面的改進，以允許序列到序列模型在語音搜索任務上明顯優於傳統的 ASR 系統。在此工作中我們將重點放在對 LAS 模型的改進上。LAS 模型是一個單一的神經網路，其中包括類似於常規聲學模型的編碼器。我們既考慮對模型結構的修改，也考慮優化過程。在結構方面，首先，我們探索單詞模型 (WPM)，我們比較了 LAS 的字素和 WPM，並發現 WPM 有

適度的改進。接下來，我們探索合併多頭注意力，它使模型能夠學習到編碼特徵的多個位置。

研究結果：

實驗結果顯示，結構改進 (WPM , MHA) 後，在 WER 方面提高了 11% ，而優化改進 (MWER , SS , LS 和同步訓練) 後又提高了 27.5% ，而語言模型記錄的改進又提高了 3.4% 。應用於 Google 語音搜索任務後，我們的 WER 為 5.6% ，而混合 HMM-LSTM 系統的 WER 為 6.7% 。在命令任務中測試了相同的模型，在 WER 指標方面，我們的模型達到 4.1% ，混合系統達到 5% 。

論文題目： *Deep Audio-visual Speech Recognition*

中文題目：深度視聽語音辨識

論文作者：Triantafyllos Afouras; Joon Son Chung; Andrew Senior; Oriol Vinyals; Andrew Senior.

論文出處：IEEE Transactions on Pattern Analysis and Machine Intelligence

論文地址：<https://ieeexplore.ieee.org/abstract/document/8585066>

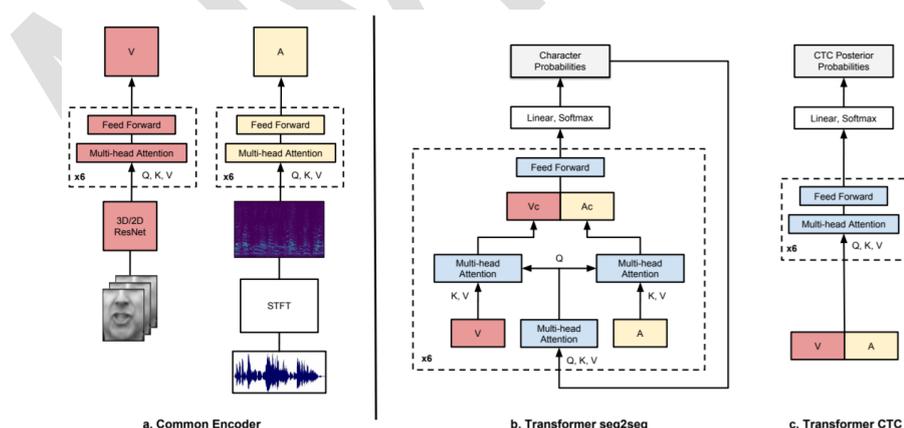
研究問題：

唇讀，作為一種僅憑視覺資訊就能識別所說內容的能力，是一項令人印象深刻的技能。由於同音字的存在，它在字面上本質上是模稜兩可的-不同的字元會產生完全相同的口音序列(例如“p”和“b”)。合理的使用句子中相鄰單詞的上下文和/或語言模型在一定程度上解決此類歧義。唇讀技術可以應用於許多場景：例如，在嘈雜的環境中“命令”向手機發送指令或消息；轉錄和重新複製檔案無聲

電影；解決多人同時語音並且總體上改善了自動語音辨識的性能。由於在電腦視覺任務中眾所周知的兩個發展，使得上述這些應用成為可能。基於為語音辨識和機器翻譯而開發的最新編碼器-解碼器體系結構——唇讀模型變得尤為重要。

研究方法：

與以前的工作著重於識別有限數量的單詞或短語不同，我們將唇讀作為一個開放世界的問題來解決-無限制的自然語言句子和野外視頻。首先，我們比較了兩種唇讀模型，一種使用 CTC 損失，另一種使用序列間損失。兩種模型都建立在變壓器自我關注架構的基礎上。其次，我們研究了唇讀在多大程度上與音訊語音辨識相輔相成，特別是當音訊信號有雜訊時；再次，我們引入並公開發佈了兩個用於視聽語音辨識的新資料集：LRS2-BBC，由英國電視臺的數千個自然句子組成；和 LRS3-TED，其中包括從 YouTube 獲得的數百小時的 TED 和 TEDx 演講。我們訓練的模型在唇讀基準資料集上大大超越了所有先前的工作。



研究結果：

實驗結果顯示，效果最佳的網路是 TM-seq2seq，使用語言模型進行解碼時，LRS-BBC 的 WER 達到 50%，與之前的 70.4% 的最新水準相比，提高了 20% 以

上。在 LRS2-BBC 上進行評估時，TM-seq2seq 模型展示出增加波束寬度的效果。對比實驗表明，當音訊信號有雜訊時，嘴巴的運動為語音辨識提供了重要線索。甚至在音訊信號乾淨的情況下也可以提高性能。例如，使用視聽 TM-CTC 模型時，單詞錯誤率從僅音訊的 10.1% 降低到 LRS2-BBC 的 8.2%，從 LRS3-TED 的 6.0% 降低到 5.0%。與僅音訊模型相比，使用視聽 TM-seq2seq 時獲得的收益相似。當在原始話語中添加雜訊來合成的嘈雜音訊與兩個資料集的純音訊情況相比時，性能下降了 60% 以上。這表明在僅限於音訊模型的性能上，該模型對單詞錯誤率的評分與僅使用嘴唇獲得的錯誤率相似。但是，將這兩種方式組合起來可帶來顯著的改進，所有模型和資料集的字錯誤率均下降 20%-30%。因此，在存在較大背景雜音的情況下，視聽模型的性能要比僅視頻或僅音訊的模型好得多。

論文題目： *Parameter Uncertainty for End-To-End Speech Recognition*

中文題目：端到端語音辨識中的參數不確定性研究

論文作者：Stefan Braun and Shih-Chii Liu.

論文出處：2019 IEEE International Conference on Acoustics, Speech and Signal Processing.

論文地址：<https://ieeexplore.ieee.org/abstract/document/8683066>

研究問題：

近期端到端 (End-to-End) 的自動語音辨識 (Automatic Speech Recognition, ASR) 研究相比於 DNN-HMM 混合系統，在模型結構和訓練過程方面有了明顯簡化。傳統的端到端模型通常使用確定性參數 (Deterministic Parameters)，即每個參數對應一個確定的實數值。然而在對相關任務進行訓練過程中，該類方法僅

對參數大小進行編碼，沒有直接對參數的不確定性 (Uncertainty) 或重要性 (Importance) 進行直接編碼，但是這些內容也包含重要的資訊。因此學者們開始研究將參數以概率的形式進行編碼，來探索神經網路中的參數不確定性。相關研究工作表明，在自動語音辨識之外的其他多個任務的剪枝實驗中，參數信噪比 (Signal-to-Noise Ratio, SNR) 與參數重要性展現出很高的相關性。目前在語音辨識領域相關的研究還很少，已知僅有的一項研究從貝葉斯模型角度使用變分推斷框架 (variational inference framework) 匯出概率網路圖。本文從參數角度提出另一種概率網路，避免了對貝葉斯模型解釋的需求。

研究方法：

本文研究使用不確定性參數的端到端方法將自動語音辨識任務的領域適用性，包括純淨語音和帶噪語音；提出基於信噪比 (SNR-based) 的正則化方案來控制參數根據其重要性來更新；使用不同的信噪比水準的概率網路來評估；對比了在領域適用過程中不同信噪比水準的網路如何容忍 (tolerate) 參數剪枝及災難性遺忘 (catastrophic forgetting) 程度在網路中是如何變化的。

基礎端到端的模型包括 5 層的雙向 LSTM 網路(每個方向包含 320 個單元) 和最終 640×59 的網路映射至輸出標籤。

確定性模型 (deterministic models) 使用預設 LSTM 單元，參數集合 θ_D 包含 LSTM 權重 w^{LSTM} ，偏差 b^{LSTM} 及映射權重 w^{PROJ}

$$\theta_D = \{w^{LSTM}, b^{LSTM}, w^{PROJ}\}$$

概率模型 (probabilistic models) 使用具有高斯權重的 LSTM 模型 , 參數集合 θ_p 包含 LSTM 權重的均值 μ^{LSTM} , 參數化權重標準差 β^{LSTM} , 偏差 b^{LSTM} 及映射權重 w^{PROJ}

$$\theta_p = \{\mu^{LSTM}, \beta^{LSTM}, b^{LSTM}, w^{PROJ}\}$$

使用 Xavier uniform initialization (下式) 初始化 w^{LSTM} , μ^{LSTM} 和 w^{PROJ} 。

$$A_{ij} \sim \mathcal{U}\left(-\frac{\sqrt{6}}{i+j}, \frac{\sqrt{6}}{i+j}\right), A \in \mathbb{R}^{i \times j}$$

參數化標準差 β^{LSTM} 使用下式進行初始化

$$B_{ij} = \log\left(\exp\left(\frac{1}{2} \frac{\sqrt{6}}{i+j}\right) - 1\right), B \in \mathbb{R}^{i \times j}$$

在概率模型中 , 對參數化標準差 β^{LSTM} 通過使用權值衰減 (衰減項 $\mathcal{L}_\beta = \|\beta\|_2^2$) 來增強低信噪比參數。

研究結果 :

本文對使用具有概率權重參數的 LSTM 對端到端的語音辨識模型進行了評估。測試集使用 Wall Street Journal (在純淨條件下的資料) 和 CHiME-4 的語音辨識任務 (含有噪音資料)。實驗結果表明在參數剪枝和領域適用性方面概率模型獲得了比確定性模型更好的結果。概率模型的關鍵優勢是對特定參數信噪比的可用性 , 在訓練時與參數的重要程度相關性較高。

論文題目 : *Stochastic Adaptive Neural Architecture Search for Keyword Spotting*

中文題目 : 面向關鍵字檢出的隨機自我調整神經網路結構搜索

論文作者 : Tom Véniat, Olivier Schwander and Ludovic Denoyer

論文出處：2019 IEEE International Conference on Acoustics, Speech and Signal Processing.

論文地址：<https://ieeexplore.ieee.org/document/8683305>

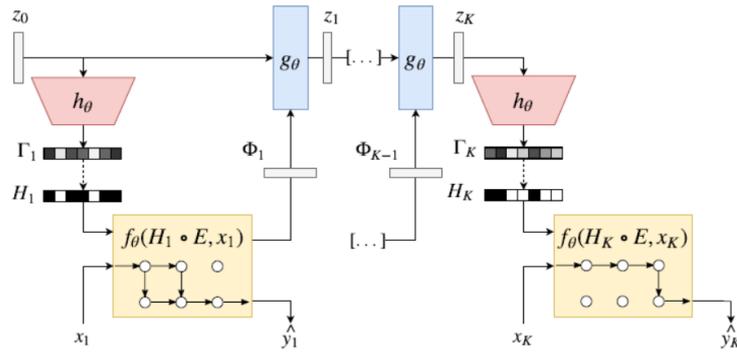
研究問題：

目前關鍵字定位 (Keyword Spotting) 問題 (如在即時音訊流中確定關鍵字) 的主要方法是在連續的滑動視窗中使用神經網路模型進行識別。在目前神經網路搜索 (Neural Architecture Search) 的研究中發現的網路結構都是靜態的 (相同的神經網路結構在預測時重用)。由於任務的複雜性，目前基準模型通常很大，導致預測階段計算資源及能源消耗水準較高。

研究方法：

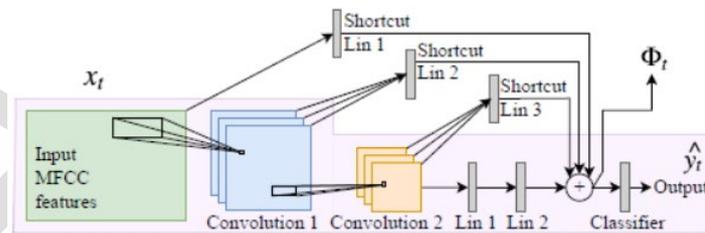
文章提出了隨機自我調整神經網路結構搜索 (Stochastic Adaptive Neural Architecture Search, SANAS) 模型，能夠在模型推斷階段自我調整地線上調整神經網路的結構 (當任務簡單時使用較小的結構，當任務複雜時使用較大的結構)。關鍵字定位 (Keyword Spotting) 可以抽象為一個音訊流序列標注問題，在每個時間步長 (timestep)，系統接收一個數據點 x_t ，生成一個輸出標籤 y_t (在音訊流中 x_t 通常為一個時頻特徵圖， y_t 為給定關鍵字是否存在的判斷)。

文章定義了一種可以根據上下文的隱含表示預測在每個時步 (timestep) 進行網路結構變化的設置。



在時步 t 中，從前一隱藏狀態生成的結構分佈 Γ_t ， $\Gamma_t = (z_t, \theta)$ ，其中 z_t 為上文 $x_1, y_1, \dots, x_{t-1}, y_{t-1}$ 的編碼隱含表示，在每一步 z_t 的更新根據神經網路結構 \mathcal{A} ，參數 θ 確定 $z_{t+1} = g(z_t, x_t, \theta, \mathcal{A})$ 。然後從 Γ_t 中抽取出離散結構 H_t ，並通過輸入 x_t 進行評估。評估過程給出特徵向量 $\Phi_t(x_t, \theta, E \circ H_t)$ 來計算下一個隱含狀態，並根據 $f(z_t, x_t, \theta, E \circ H_t)$ 來預測模型 \hat{y}_t 。虛線代表 sampling 操作，在推理階段，每個 timestep 中選出具有最高概率的結構。

下圖為基於卷積神經網路 (cnn-trad-fpool3) 的 SANAS 結構：



網路層之間的連接通過上述模型採樣生成，高亮的網路結構是增加了快捷連接 (shortcut connections) 的基準模型。

研究結果：

實驗評估資料使用了 Speech Commands 資料集。實驗對比了傳統的靜態模型和本文提出的方法，結果表明 SANAS 方法能夠很大程度上降低 FLOPs (每秒浮點運算次數)，同時相對於基準方法識別出了更多的關鍵字，準確率也更高。

論文題目：Adversarially Trained End-to-end Korean Singing Voice Synthesis System

中文題目：基於對抗訓練的端到端韓語歌聲語音合成系統

論文論文作者：Juheon Lee, Hyeong-Seok Choi, Chang-Bin Jeon, Junghyun Koo, Kyogu Lee.

論文出處：20th Annual Conference of the International Speech Communication Association INTERSPEECH 2019.

論文地址：https://www.isca-speech.org/archive/Interspeech_2019/pdfs/1722.pdf

研究問題：

深度學習技術的發展催生了一批基於神經網路的歌聲語音合成系統，系統基於歌聲、歌詞及樂譜平行資料來訓練聲學模型 (acoustic model)，用來預測聲碼器 (vocoder) 的輸入特徵。儘管這些基於神經網路的歌聲合成系統能夠有較好的表現，但是預測聲碼器特徵的網路仍然具有局限：不能超過聲碼器的性能上限。因此，本文研究能夠直接生成線性頻譜圖來代替聲碼器特徵的端到端模型，其挑戰在於端到端模型增加了模型的複雜性，而且需要大量的訓練資料。

研究方法：

本文提出的歌聲合成系統能夠基於適量的訓練資料使用端到端的方式進行訓練。系統結構如下圖：

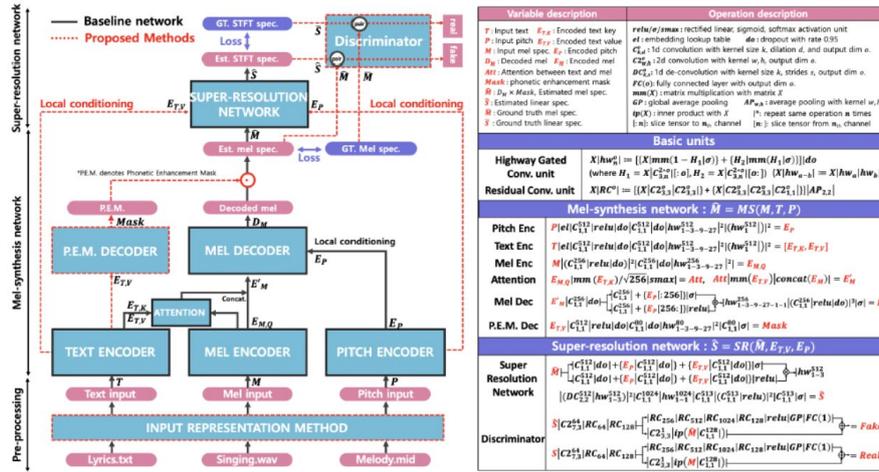


Figure 1: Proposed system overview (left), detailed structure of each sub-module (right). $X[F]$ denotes $F(X)$.

系統主要由兩個模組構成：旋律合成網路 mel-synthesis 和超解析度網路 super-resolution。mel-synthesis 網路根據前面的旋律輸入 $M_{0:L-1}$ ，時序對齊的文本 $T_{1:L}$ 及音調輸入 $P_{1:L}$ 訓練生成旋律譜圖；super-resolution 網路根據文本和音調資訊作為條件輸入，將生成的旋律譜圖 M 進行上採樣 (upsample)；最後判別器 (discriminator) 將上採樣結果和生成的旋律譜圖以對抗的方式訓練網路。

在測試階段，從給定文本及音調輸入中以自回歸的方式生成旋律譜圖的幀序列，然後通過 super-resolution 網路上採樣為線性譜圖，最後通過 Griffin-Lim 演算法轉換為聲波形式 (waveform)。

研究結果：

實驗使用手工收集整理的歌聲資料集，包含了 60 首流行歌曲。實驗表明使用文本資訊對 phonetic enhancement mask 進行建模是有效的，能夠生成更為準確的發音。同時在 super-resolution 階段使用條件對抗 (conditional adversarial) 訓練方法能夠獲得更高的聲音品質。

6.5 語音辨識進展

隨著人工智慧的迅速發展，語音辨識的技術越來越成為國內外研究機構的焦點。人們致力於使機器能夠聽懂人類的話語指令，並希望通過語音實現對機器的控制。作為一項人機交互的關鍵技術，語音辨識在過去的幾十年裡取得了飛速的發展，在研究和探索過程中針對語音辨識的各部流程進行了各種各樣的嘗試和改造，以期發現更好的方法來完成語音辨識流程中的各個步驟，以此來促進在不同環境下語音辨識的效率和準確率。研究人員從最簡單的非常小詞彙量的閱讀式的語音辨識問題開始，逐漸轉向越來越複雜的問題。

近年來智慧語音進入了快速增長期，語音辨識作為語音領域的重要分支獲得了廣泛的關注，如何提高聲學建模能力和如何進行端到端的聯合優化是語音辨識領域中的重要課題。

語音辨識經歷了從 2012 年最開始的 DNN 的引入時的 Hybrid HMM 結構，再到 2015 年開始吸引大家研究興趣的 CTC 演算法，而後到 2018 年的 Attention 相關結構的研究熱點。Attention 相關演算法在語音辨識或者說話人識別研究的文章中出現頻率極高。從最開始 Attention，到 Listen-Attend-Spell，再到 Self-Attention（或者 Transformer），在不同的文章被作者多次介紹和分析，頻繁出現在了相關文章的 Introduction 環節中。在 Attention 結構下，依然還有很多內容需要研究者們進一步地探索：例如在一些情況下 Hybrid 結構依然能夠得到 State-of-the-art 的結果，以及語音資料庫規模和 Attention 模型性能之間的關係。

在近兩年的研究中，端到端語音辨識仍然是 ASR（Automatic Speech Recognition）研究的一大熱點，正如上文提到的，基於 Attention 機制的識別系統

已經成為了語音技術研究主流。同時，隨著端到端語音辨識框架日益完善，研究者們對端到端模型的訓練和設計更加的關注。遠場語音辨識 (far-field ASR) ，模型結構 (ASR network architecture) ，模型訓練 (model training for ASR) ，跨語種或者多語種語音辨識 (cross-lingual and multi-lingual ASR) 以及一些端到端語音辨識 (end-to-end ASR) 成為研究熱點。

在語音合成方面，高音質語音生成演算法及 Voice conversion 是近兩年研究者關注的兩大熱點，Voice Conversion 方向的研究重點主要集中在基於 GAN 的方法上。在語言模型方面 (Language Model) 的研究熱點主要包括 NLP 模型的遷移，低頻單詞的表示，以及深層 Transformer 等。

在說話人識別方面，說話人資訊，特別是說話人識別及切分，正被越來越多的研究者所重視。目前 Attention 在說話人方面更類似一種 Time Pooling，比 Average Pooling 及 Stats Pooling 更能捕捉對說話人資訊更重要的資訊，從而帶來性能提升。說話人識別技術經歷深度學習帶來的性能飛躍後，在模型結構、損失函數等方面的探討已經較為成熟，以 TDNN、ResNet 加上 LMCL、ArcFace 的主流模型開始不斷刷新各資料集的性能上限。模型以外的因素逐漸成為制約說話人系統的瓶頸。說話人技術目前也逐漸暴露出與人臉識別同樣的易受攻擊的問題。因此，ASVspoof 這樣的 Challenge 從 2015 年起就開始關注聲紋反作弊問題。相信隨著此類研究的不斷深入，結合聲紋系統的性能提升，聲紋將有望變成我們的“聲音身份證”。

7 電腦圖形學

7.1 電腦圖形學概念

國際標準組織 ISO 將電腦圖形學定義為：電腦圖形學是一門研究通過電腦將資料轉換成圖形，並在專門顯示裝置上顯示的原理方法和技術的學科。它是建立在傳統的圖形學理論、應用數學及電腦科學基礎上的一門邊緣學科。這裡的圖形是指三維圖形的處理。簡單來講，它的主要研究內容是研究如何在電腦中表示圖形，以及利用電腦進行圖形的計算處理和顯示的相關原理和演算法。

在電腦圖形學的開創之初，他主要解決的問題是在電腦中表示三維結合圖形以及如何利用電腦進行圖形的生成處理和顯示的相關原理和演算法，目的是產生令人賞心悅目的真實感圖像，這僅僅是狹義的電腦圖形學。隨著近些年的發展，電腦圖形學的內容已經遠遠不止這些，廣義的電腦圖形學研究內容非常廣泛，包括圖形硬體、圖形標準、圖形交互技術、柵格圖形生成演算法、曲線曲面造型、實體造型、真實版圖形的計算、顯示演算法、科學計算視覺化、電腦動畫、虛擬實境、自然景物模擬等等。

電腦圖形學的總體框架可以包括以下幾個部分：數學和演算法基礎、建模、渲染以及人機交互技術。電腦圖形學需要一些基本的數學演算法，例如向量和幾何的變化、幾何建模式的三維空間變化、三維到二維的圖形變換等等。建模是進行圖形描述和計算，由於在多維空間中有各種組合模型，有一些是解析式表達的簡單形體，也有一些隱函數表達的複雜曲線，因此需要進行複雜的建模工作。渲染也叫繪製，指的是模型的視覺實現過程，例如對光照紋理等理論和演算法進行

處理，其中也需要大量的計算。交互技術可以說是圖形學交互的重要工具，是電腦圖形學的重要應用。

7.2 電腦圖形學發展歷史

20 世紀 50 年代：1950 年，美國 MIT 的旋風一號 (whirlwind I) 電腦配備了世界上第一台顯示器——陰極射線管 (CRT) 來顯示一些簡單的圖形，使得電腦擺脫了純數值計算的單一用途，能夠進行簡單的圖形顯示，從此電腦具有了圖像顯示功能，但是還不能對圖形進行交互操作，這時的電腦圖形學處於準備和醞釀時期，並稱之為“被動式”圖形學。

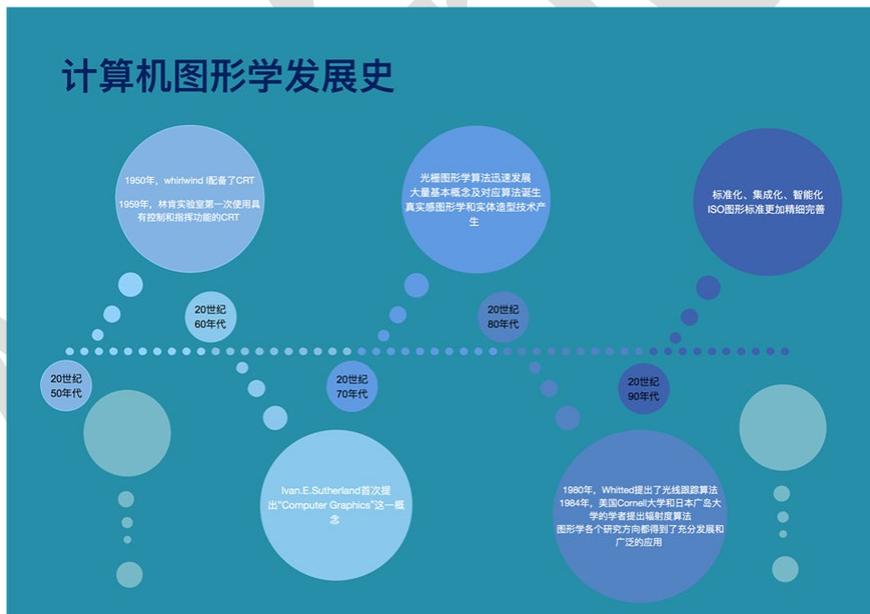


圖 7-1 電腦圖形學發展歷史

50 年代末期，MIT 的林肯實驗室在“旋風”電腦上開發 SAGE(Semi-Automatic Ground Environment System) 空中防禦體系。SAGE 於 1957 年投入試運行，已經能夠將雷達信號轉換為顯示器上的圖形並具有簡單的人機交互功能，操作者使用光筆點擊螢幕上的目標即可獲得敵機的飛行資訊，這是人類第一次使用光筆在螢

幕上選取圖形。1959 年，麻省理工學院林肯實驗室第一次使用了具有指揮和控制功能的 CRT，“被動式”圖形學開始邁向互動式電腦圖形學。

20 世紀 60 年代：1962 年美國 MIT 林肯實驗室的 IvanE.Sutherland 發表了一篇題為“sketchpad：一個人機交互通信的圖形系統”的博士論文，首次使用了“Computer Graphics”這一概念，證明了互動式電腦圖形學是一個可行的、有應用價值的研究領域，從而確立了電腦圖形學正式成為一個獨立學科的分支。1968 年 IvanE.Sutherland 又發表了《頭戴式三維顯示器》的論文，在頭盔的封閉環境下，利用電腦成像的左右視圖匹配，生成立體場景，使人置身於虛擬實境中。IvanE.Sutherland 為電腦圖形學技術做出了巨大的貢獻，被稱作電腦圖形學的開山鼻祖，1988 年 IvanE.Sutherland 被授予 A.M 圖靈獎。並且這一時期，點陣圖形學演算法開始萌芽。

20 世紀 70 年代：圖形學在這一時期進入了興盛期，點陣圖形學演算法迅速發展，區域填滿、裁剪、消隱等概念及其相應演算法紛紛被提出，實用的 CAD 圖形系統也開始出現。除此之外，真實感圖形學和實體造型技術的產生也是 70 年代電腦圖形學的兩個重要進展。1970 年 J.Bouknight 在 ACM 上發表論文，提出了第一個光反射模型，指出物體表面朝向是確定物體表面上一點光強的主要因素，並用 Lambert 漫反射定律計算物體表面上各多邊形的光強，對光照射不到的地方用環境光代替。1971 年 Henri Gouraud 在 IEEE Trans.Computer 上提出被稱為 Gouraud 明暗處理的“漫反射模型 + 插值”思想，對多面體模型，用漫反射模型計算多邊形頂點的光亮度，再用增量法插值計算多邊形的其他內部點。1975 年 Phong 在 ACM 上發表論文提出了著名的簡單光照模型“Phong 模型”，Phone 模型雖然只是一個經驗模型，但其真實度已經達到了較好的顯示效果。這些都是真實

感圖形學最早的開創性工作。從 1973 年開始，相繼出現了英國劍橋大學 CAD 小組的 Build 系統、美國羅徹斯特大學的 PADL-1 系統等實體造型系統，這些都為 CAD 領域的發展做出了重要貢獻。

70 年代圖形軟體標準化程度提高，1974 年，ACMSIGGRAPH“與機器無關的圖形技術”的工作會議的召開，提出了圖形軟體標準化問題，ACM 成立圖形標準化委員會，制定“核心圖形系統”（core graphics system），ISO 發佈 CGI、CGM、GKS、PHIGS 一系列的圖形標準，其中 1977 年的 GKS 是 ISO 批准的第一個圖形軟體標準軟體，是一個二維圖形軟體標準，1986 年，ISO 公佈了程式師級的分層結構交互圖形系統 PHIGS，這是一些非官方的圖形軟體，廣泛應用於工業界並成為事實上的標準，PHIGS 是對 GKS 的擴充，增加的功能有物件建模、彩色設定、表面繪製和圖形管理等。隨後 PHIGS 的擴充成為 PHIGS+；1988 年的 GKS3D，是 ISO 批准的第二個圖形軟體標準軟體，是一個三維圖形軟體標準。

20 世紀 80 年代以後：出現了帶有光柵掃描顯示器的微型電腦和圖形工作站，極大的推動了電腦圖形學的發展，如 Machintosh、IBM 公司的 PC 及其兼容機，Apollp、Sun 工作站等。隨著奔騰 III 和奔騰 IV 系列 CPU 的出現，電腦圖形軟體功能開始部分地由硬體實現。高性能顯卡和液晶顯示幕的使用，高傳輸率大容量硬碟的出現，特別是 Internet 的普及使得微型電腦和圖形工作站在運算速度、圖形顯示細節上的差距越來越小，這些都為圖形學的飛速發展奠定了物質基礎。

1980 年 Turner Whitted 提出了光透視模型，並第一次給出光線跟蹤演算法的範例，實現了 Whitted 模型；1984 年美國 Cornell 大學和日本廣島大學的學者分別將熱輻射工程中的輻射度演算法引入到電腦圖形學中，用輻射度的方法成功地模擬了理想漫反射表面間的多重漫反射效果。以上二者的提出，標誌著真實感圖形

的顯示演算法已逐漸成熟。80 年代中期以後，超大型積體電路的發展，電腦運算能力的提高，圖形處理速度的加快，促使了圖形學各個研究方向都得到了充分發展和廣泛的應用。

20 世紀 90 年代以後：微機和軟體系統的普及使得圖形學的應用領域日益廣泛，電腦圖形學朝著標準化、集成化和智慧化的方向發展，多媒體、人工智慧、電腦視覺化、虛擬實境等分支蓬勃發展，三維造型也獲得了長足發展。ISO 公佈的圖形標準越來越精細，更加成熟。這時存在著一些事實上的標準，如 SGI 公司開發的 OpenGL 開放式三維圖形標準，微軟公司為 PC 遊戲開發的應用程式介面標準 DirectX 等，Adobe 公司 Postscript 等，均朝著開放、高效的方向發展^[11]。

7.3 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為電腦圖形學全球學者分佈情況：



圖 7-2 電腦圖形學領域全球學者分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量優勢明顯；歐洲也有較多的人才分佈，主要在歐洲中西部；亞洲的人才主要集中在我國東部及日韓地區；其他諸如非洲、南美洲等地區的學者非常稀少；電腦圖形學的人才分佈與各地區的科技、經濟實力情況大體一致。

此外，在性別比例方面，電腦圖形學領域中男性學者占比 93.7%，女性學者占比 6.3%，男性學者占比遠高於女性學者。

電腦圖形學領域學者的 h-index 分佈如下圖所示，分佈情況大體呈階梯狀，大部分學者的 h-index 分佈在中低區域，其中 h-index 在小於 20 區間的人數最多，有 1240 人，占比 60.1%，50-60 區間的人數最少，有 50 人。

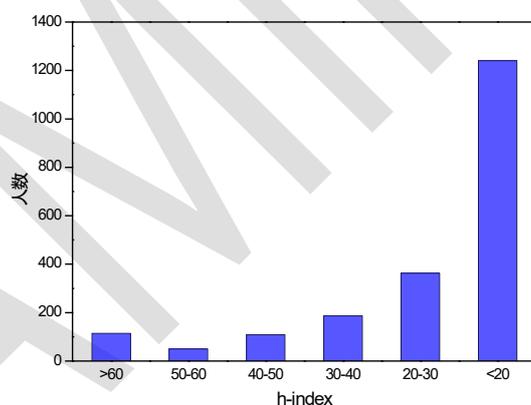


圖 7-3 電腦圖形學領域學者 h-index 分佈

● 中國人才分佈

我國專家學者在電腦圖形領域的分佈如上圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是長三角和珠三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，

通過觀察中國周邊國家的學者數量情況，特別是與日韓等地相比，中國在電腦圖形領域學者數量略多但差距較小。



圖 7-4 電腦圖形學領域中國學者分佈

中國與其他國家在電腦圖形學領域的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 7-1 電腦圖形學領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	237	8729	37	407
中國-加拿大	69	3550	51	85
中國-以色列	59	3203	54	58
中國-英國	34	2299	68	72
中國-新加坡	28	1080	39	37

中國-德國	24	625	26	42
中國-瑞士	21	779	37	37
中國-印度	16	784	49	28
中國-沙烏地阿拉伯	16	468	29	22
中國-法國	15	485	32	36

從上表資料可以看出，中美合作的論文數、引用數、學者數遙遙領先，表明中美間在電腦圖形領域合作之密切；此外，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 4 席；中國與英國合作的論文數雖然不是最多，但是擁有最高的平均引用數說明在合作品質上中英合作達到了較高的水準。

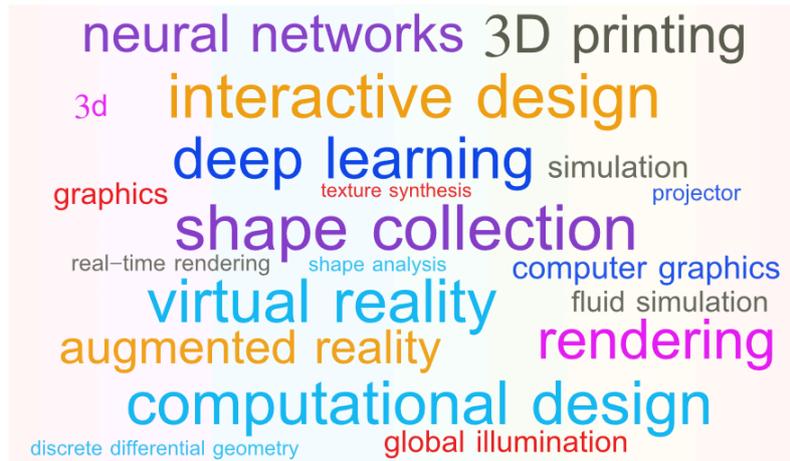
7.4 論文解讀

本節對本領域的高水準學術會議及期刊論文進行挖掘，解讀這些會議和期刊在 2018-2019 年的部分代表性工作。這些會議和期刊包括：

ACM SIGGRAPH Conference

ACM Transactions on Graphics

我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如下圖所示。其中，形狀集合 (shape collection)、交互設計 (interactive design)、計算設計 (computational design) 是本領域中最熱的關鍵字。



論文題目： *A Style-based Generator Architecture for Generative Adversarial Networks*

中文題目：基於樣式的生成式對抗網路生成器架構

論文作者：Tero Karras, Samuli Laine, Timo Aila.

論文出處：The IEEE Conference on Computer Vision and Pattern Recognition- CVPR 2019

論文地址：

http://openaccess.thecvf.com/content_CVPR_2019/papers/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.pdf

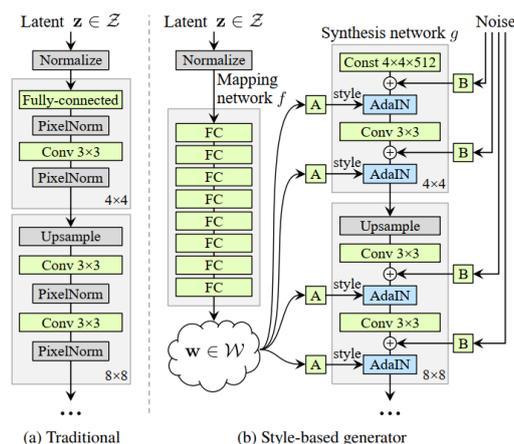
研究問題：

本文針對自動的無監督的習得圖像的高層屬性(譬如人臉對應的身份資訊以及拍攝姿態) 以及對於生成的每幅圖像產生一些特定的隨機化的變換(譬如臉部瑕疵以及頭髮的細節) , 生成較為直觀且可控的合成結果進行了研究。通過借鑒風格遷移的思想, 提出了一種新的對抗網路中的生成器架構。該架構不僅在傳統

的分佈距離的度量上優勢明顯，並且較好地將控制圖像變化的隱變數分離出來進行獨立建模。

研究方法：

下圖為提出的方案與傳統的生成器的對比：



傳統的潛在編碼表示由前向神經網路的第一層生成，而作者提出不再使用網路編碼，而是如圖(b)的右側子圖所示，由一個可通過反向傳播學習的常數作為網路輸入。在這一新的架構中，由映射網路 f 將採樣自隱空間 \mathcal{Z} 的隱變數 z ，經由非線性映射網路 f ，投射到潛在的中間表達空間 \mathcal{W} 中，該空間控制著不同的視覺特徵的表達。通過可學習的仿射變換，又將潛在空間中的編碼 w 解碼為樣式資訊 y_0 。而 y 通過對自我調整樣本歸一化 (adaptive instance normalization, AdaIN) 後的特徵分佈進行放縮和偏置操作：

$$\text{AdaIN}(x_i, y) = y_{s,i} \frac{x_i - \mu(x_i)}{\sigma(x_i)} + y_{b,i},$$

影響生成網路 g 的不同解析度的卷積層輸出通道的重要性，進而決定了輸出圖片的視覺特徵。相較於風格遷移中使用圖片作為遷移目標，這裡使用解碼自向

量 w 的具有空間無關特性的樣式編碼 y ，結合高斯雜訊來輸入生成網路 g 創建新的圖像。

研究結果：

作者提出的生成器設計將不同的樣式資訊引入到生成網路 g 的不同層，可以實現在不同解析度上對 w 施加影響，進而控制高級屬性以及細小特徵上的改變。如圖 (a-c) 所示在不同解析度上將源分佈生成的樣式編碼引入靶心圖表像，產生不同的由高層 (a) 到細節 (c) 的改變。作者展示在經典的名人臉部資料集以及新提出的更加多樣化的人臉資料集上的當前最佳性能。並且，作者將其應用於車輛以及臥室場景資料集上，也取得了較好的視覺效果。



論文題目：*TempoGAN: A Temporally Coherent, Volumetric GAN for Super-resolution Fluid Flow*

中文題目：用於超分流體的時間一致性生成對抗網路

論文作者：You Xie, Erik Franz, Mengyu Chu, Nils Thuerey

論文出處：ACM Transaction on Graphics – SIGGRAPH-2018

論文地址：<https://arxiv.org/pdf/1801.09710.pdf>

研究問題：

生成對抗網路 (GAN) 過去幾年在表示和生成複雜的自然圖像方面取得了很大的進展。這些工作表明，深度卷積神經網路 (CNNs) 能夠有效提取輸入圖片的特徵分佈，並用於生成全新的，不同於原始輸入的圖片。同時，相關工作如 PG-GAN 在從粗糙輸入生成高解析度的自然圖像方面也很成功。但是，這些生成模型在其推導和改進過程中並沒有將資料的時間資訊考慮進來，而這些時間資訊在現實的物理系統中是十分重要的。因此本文考慮資料的時間資訊，擴展現有的 GAN 方法，以生成高解析度資料集，並確保其時間連續性。

研究方法：

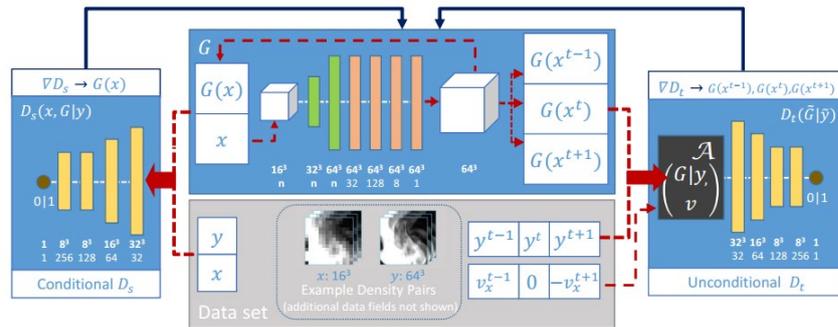
本文在之前特徵空間損失相關工作的基礎上，進一步改善生成自然圖像的逼真度和解決模式崩潰問題。為了實現上述目標，本文在生成器中引入神經網路的部分特徵空間的 L2 損失。具體來講，生成網路的中間結果將被神經網路的間接參考資料所約束，為此，本文引入新的損失函數形式：

$$\mathcal{L}_f = \mathbb{E}_{n,j} \lambda_f^j \|F^j(G(x)) - F^j(y)\|_2^2$$

其中 j 是判別網路中的某一層， F^j 表示相應層的啟動回應，係數 λ_f^j 是權重項，用於平衡各個層之間的不平衡性。進一步地，為了使生成的資料具有時間一致性，文章還設計了下面的 L2 損失函數：

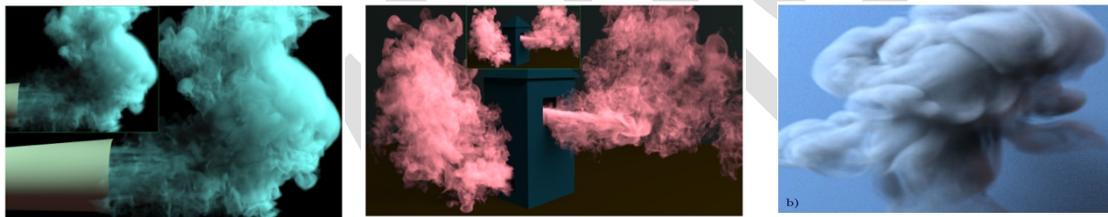
$$\mathcal{L}_{2,t} = \|G(x^t) - \mathcal{A}(G(x^{t-1}), v_x^{t-1})\|_2^2$$

下圖是文章設計的網路結構，使用全部由卷積層構成的網路作為生成器，使得模型能夠應用于生成任意尺寸的資料。網闢結構借鑒了 U-Net 和 ResNet，通過 skip 連接綜合 low-level 資訊和 high-level 資訊。實驗證明這種結構生成的自然圖像品質最高。



研究結果：

文章設計的生成網路能夠學會生成高精度的，逼真的且時間上具有一致性的流體圖形。所設計的方法僅使用單個時間步長的低解析度流體資料即可完成。在二維和三維資料上都取得了不錯的效果。



論文題目：*Temporal Segment Networks for Action Recognition in Videos*

中文題目：用於視頻行為識別的時間分段網路

論文作者：Wang, Limin and Xiong, Yuanjun and Wang, Zhe and Qiao, Yu and Lin, Dahua and Tang, Xiaoou and Van Gool, Luc.

論文出處：IEEE transactions on pattern analysis and machine intelligence (2018)

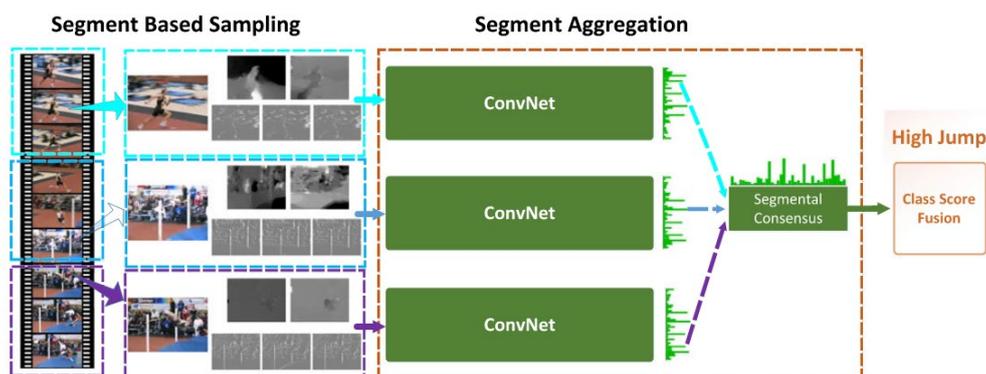
論文地址：<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8454294>

研究問題：

行為識別是電腦影像處理的重要應用領域，基於視頻的行為識別是行為識別問題中的重要方法。傳統的方法往往對單幀圖像進行分析，這種方法缺乏前後幀

的資訊建模。有一種前後幀建模的方式是提取前後幀的光流資訊，利用光流資訊來完成運動的建模。然而光流只能描述短暫的順序關係，無法對較長的時序進行建模。

研究方法：



這篇文章對整個視頻進行分段，每一段作為一個時序模組。具體來講每個時序模組中都進行兩種建模，一個是單幀圖像獨立的卷積神經網路 (CNN) 模型，而另一個則是採用光流資訊進行輸入的卷積神經網路 (CNN) 模型。與之前方法只進行較短時間的建模不同的是，這篇文章將視頻分成三個較長的片段，對三個長片段進行分析。從而完成較長時序上的建模。

研究結果：

這篇文章所提出的 TSN 方法在 HMDB51, UCF101, THUMOS14, ActivityNet v1.2 以及 Kinetics400 資料庫上進行了廣泛的實驗，取得了很高的精度。當採用 RGB 的前後幀差分圖來代替光流圖時，這篇文章的方法能夠取得很快的實測速度。

論文題目：*Non-local Neural Networks*

中文題目：非局部神經網路

論文作者：Xiaolong Wang, Ross Girshick, Abhinav Gupta, Kaiming He

論文出處：The IEEE Conference on Computer Vision and Pattern Recognition- CVPR
2018

文章連接：

http://openaccess.thecvf.com/content_cvpr_2018/papers/Wang_Non-Local_Neural_Networks_CVPR_2018_paper.pdf

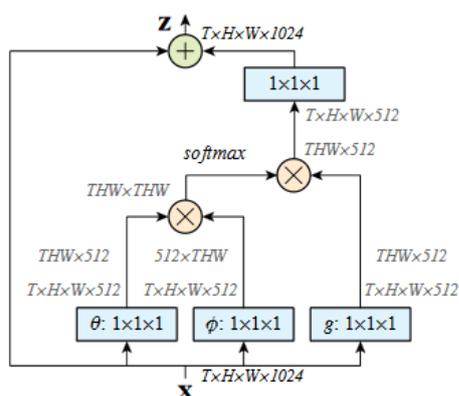
研究問題：

在圖像目標檢測、視頻分類等任務中，捕捉長距離、大範圍內的資料依賴關係是一個很重要的問題。在傳統卷積神經網路中，通常使用較大的卷積核來達到捕捉遠距離圖元之間關係的目的。在序列化資料如視頻中，通常使用迴圈神經網路建立時序上距離遠的圖元之間的聯繫。但是卷積和迴圈操作本身都是局部操作，只能在有限的空間範圍內捕捉，難以獲得更遠甚至全域的圖元間依賴關係。

研究方法：

受到 non-local mean 方法的啟發，作者提出非局部模組，它可以很好地捕捉時間或空間維度上較遠距離的圖元間的依賴關係。非局部模組的計算過程如上圖所示。

以視頻分類為例，輸入到非局部模組的特徵圖是一個形狀為(T, H, W, C)的四維張量，首先將這個經過一個卷積的特徵圖與另一個經過卷積的自身的轉置在 C 維度上做矩陣乘，之後經過 softmax 得到一個形狀為(THW, THW)的張量，該張量與另一個經過卷積層的特徵圖做矩陣乘，結果再經過一個卷積層後與輸入的特徵圖相加得到輸出。



研究結果：

作者在 Kinetics、Charades 和 COCO 資料集上針對視頻分類、目標檢測等任務進行了實驗。視頻分類任務中，在 Kinetics 資料集上，帶有非局部模組的網路的 top1 精度會比其基準網路提升大約 1.6 個百分點，top5 精度提升約 1 個百分點；在 Charades 資料集上測試集精度大約有 2.3 個百分點的提升。目標檢測任務中，在 COCO 資料集上，作者提出的網路相比其基準網路提升約 1.4 點 mAP。實驗結果證明了非局部模組對於建立圖像中或圖像間長距離依賴關係的有效性。

論文題目： *Squeeze-and-Excitation Networks*

中文題目：壓縮與激勵網路

論文作者：Jie Hu, Li Shen, Gang Sun

論文出處：The IEEE Conference on Computer Vision and Pattern Recognition- CVPR 2018

文章連接：

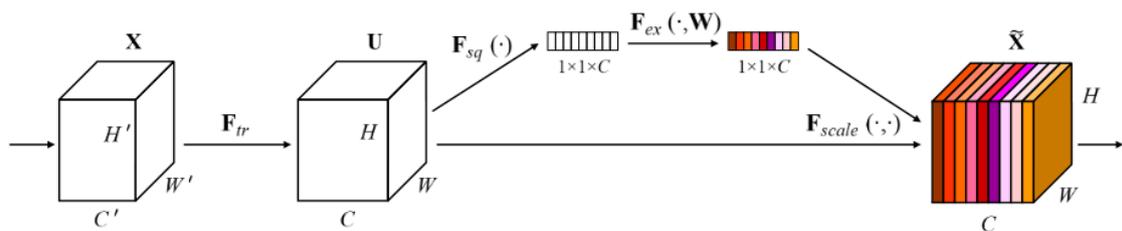
http://openaccess.thecvf.com/content_cvpr_2018/papers/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.pdf

研究問題：

如何通過建模特征通道之間的依賴關係提升網路性能，通過學習，使得有效特征圖權重變大，而削弱無效圖權重，從而使網路關注到重要部分。

研究方法：

提出的 SE 模組 (Squeeze-and-Excitation Block) 包括壓縮和激勵兩個關鍵操作，利用它對特征通道間的資訊進行融合。融合時不引入新的空間維度，而是採用了一種全新的特征重標定策略。通過學習的方式來自動獲取到每個特征通道的重要程度，然後依照這個重要程度去提升有用的特征並抑制對當前任務用處不大的特征。如下圖所示，對於一張形狀為 (H, W, C) 的特征圖，SE 模組首先進行壓縮操作，在空間維度將特征圖進行全域平均得到一個表示全域資訊的特征向量。其次是激勵操作，作者設計這一步的思想類似於 RNN 中的門控機制，通過一個全連接和啟動層為每個特征通道生成權重，全連接層中的參數用來顯式地建模各個通道間的相關性。最後進行重加權操作，通過乘法逐通道地將特征權重乘到輸入的特征圖上，完成特征重標定操作。



研究結果：

作者分別在 ImageNet 和 COCO 資料集上進行了圖像分類和目標檢測任務的實驗。分類任務中，分別以 ResNet50、ResNet101、ResNet152 為基準，加入了 SE 模組的網路與基準相比，top1 分類誤差分別減小了 1.51%、0.79%、0.85%，而相應的 GFLOPs 只增加了 0.01、0.02、0.02。在目標檢測任務上，使用 Faster

RCNN 模型，在骨幹絡上加入 SE 模組的模型所得 AP 值比基準網路高出大約 2.2~2.4 個百分點。從實驗結果可以看出 SE 模組對於說明網路建立特徵通道間的依賴關係有很大幫助。

論文題目： *Neural Ordinary Differential Equations*

中文題目：神經網路常微分方程

論文作者：Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, David Duvenaud .

論文出處：The 32nd Conference on Neural Information Processing Systems - NeurIPS 2018

論文地址：<https://arxiv.org/abs/1806.07366>

研究問題：

已有的神經網路模型由大量隱層堆疊而成，是一種離散的序列化的表示形式。作者考慮將神經網路的隱狀態參數化為連續時間下的隱變數模型，實現消除離散的層的概念的連續的新型神經網路模型。該種方案將神經網路視為帶求解的隱狀態序列，通過已有的常微分方程求解方案(ODE-solver)來求解網路的輸出。為避免常規的數值解法的較高計算代價及累積誤差，作者利用伴隨敏感度方法(adjoint sensitivity method)將 ODE-solver 視為黑箱，利用兩次對於增廣 ODE 反向傳播演算法求解對於參數的梯度，而無需真正關注內部的具體操作。此種求解方案廣泛的適用於各類 ODE-solver，可以在求解精度與速度間進行平衡。

研究方法：

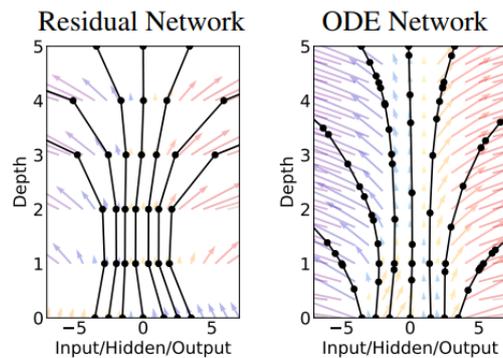
下式可以視為 ResNet 中的殘差連接，表示某卷積層的輸出與該層的輸入的加和：

$$\mathbf{h}_{t+1} = \mathbf{h}_t + f(\mathbf{h}_t, \theta_t)$$

而這種殘差連接又可以被視為 ODE-solver 中歐拉法的對連續系統的一種離散表示。在傳統的神經網路中，離散的時間步體現為層的概念。通過不斷增加層，將得到盡可能小的時間步，進而得到參數化的神經網路隱狀態的連續表示，也即上式的微分形式：

$$\frac{d\mathbf{h}(t)}{dt} = f(\mathbf{h}(t), t, \theta)$$

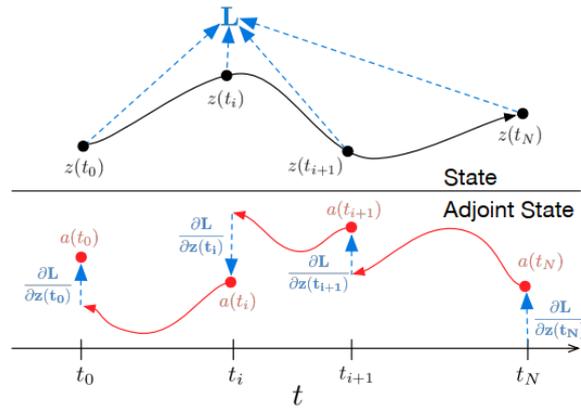
從初始層 $\mathbf{h}(0)$ 觸發，輸出層 $\mathbf{h}(T)$ 可以被定義為關於該初始值的在時間步 T 時的上述 ODE 的解。下圖給出了殘差網路所對應的隱變數的離散化表示形式，以及神經微分方程網路的連續化狀態表示：



若通過反向傳播演算法來求解連續狀態的網路，因其前向過程往往對應於密集的數值積分方法，對積分求解微分將會帶來巨大的記憶體開銷和數值誤差。通過將 ODE-solver 視為黑箱模型，採用伴隨敏感度方法將使得計算隨問題規模線性增加，同時伴隨著較低的記憶體開銷。考慮如下損失：

$$L(\mathbf{z}(t_1)) = L\left(\mathbf{z}(t_0) + \int_{t_0}^{t_1} f(\mathbf{z}(t), t, \theta) dt\right) = L(\text{ODESolve}(\mathbf{z}(t_0), f, t_0, t_1, \theta))$$

伴隨敏感度方法求解方案如下圖：



考慮如圖所示的增廣的連續時間系統，其包含原始狀態和損失函數關於狀態的敏感性兩部分。如果損失直接依賴於在多個觀測點的狀態，那麼伴隨狀態也應相應的在損失於觀測點處的偏導數方向進行更新。對於該初值問題的參數 θ 的反向傳播演算法如下：

Algorithm 1 Reverse-mode derivative of an ODE initial value problem

Input: dynamics parameters θ , start time t_0 , stop time t_1 , final state $\mathbf{z}(t_1)$, loss gradient $\frac{\partial L}{\partial \mathbf{z}(t_1)}$
 $s_0 = [\mathbf{z}(t_1), \frac{\partial L}{\partial \mathbf{z}(t_1)}, \mathbf{0}_{|\theta|}]$ ▷ Define initial augmented state
def aug_dynamics($[\mathbf{z}(t), \mathbf{a}(t), \cdot], t, \theta$): ▷ Define dynamics on augmented state
 return $[f(\mathbf{z}(t), t, \theta), -\mathbf{a}(t)^\top \frac{\partial f}{\partial \mathbf{z}}, -\mathbf{a}(t)^\top \frac{\partial f}{\partial \theta}]$ ▷ Compute vector-Jacobian products
 $[\mathbf{z}(t_0), \frac{\partial L}{\partial \mathbf{z}(t_0)}, \frac{\partial L}{\partial \theta}] = \text{ODESolve}(s_0, \text{aug_dynamics}, t_1, t_0, \theta)$ ▷ Solve reverse-time ODE
return $\frac{\partial L}{\partial \mathbf{z}(t_0)}, \frac{\partial L}{\partial \theta}$ ▷ Return gradients

研究結果：

作者在圖像相關的監督學習任務，MNIST 書寫數位分類上驗證了 ODE 網路的性能。實驗結果顯示，使用相較於對應離散表示的殘差網路，連續化狀態表示的 ODE 網路具有與之相當的性能，但參數量減小為原來的 1/3，並且具有恒定的記憶體複雜度以及可以隨求解精度變化的計算複雜度。

論文題目： *The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks*

中文題目：彩票假設：尋找稀疏可訓練的神經網路

論文出處：The 7th International Conference on Learning Representations – ICLR2019

論文地址：<https://arxiv.org/pdf/1803.03635.pdf>

研究問題：

網路稀疏是神經網路加速壓縮的常用策略之一。目前主流的稀疏方法由三個階段組成，即預訓練、裁剪和重訓練。在重訓練過程中已有的方法都是將預訓練得到的權重用於初始化，這裡隱含的假設是裁剪後的子網路難以從頭訓練，因此需要借助預訓練的結果輔助優化。本文則指出在預訓練前的原始大網路中，存在特定的子網路結構和其初始化權重的組合，作者稱其為“中獎彩票”，可以在與原大網路相同反覆運算次數的訓練設置下獲得更好的測試精度和收斂表現，也就是本文提出的“彩票假設”。以此為動機，作者一方面給出通過非結構化稀疏找到“中獎彩票”的方法，另一方面通過詳盡的對比實驗驗證了“中獎彩票”的存在性以及方法的有效性。

研究方法：

為了找到給定初始化權重下原始網路中的“中獎彩票”，本文提出基於非結構化稀疏的搜索和訓練策略，具體兩種策略如下所示。

Strategy 1: Iterative pruning with resetting.

1. Randomly initialize a neural network $f(x; m \odot \theta)$ where $\theta = \theta_0$ and $m = 1^{|\theta|}$ is a mask.
2. Train the network for j iterations, reaching parameters $m \odot \theta_j$.
3. Prune $s\%$ of the parameters, creating an updated mask m' where $P_{m'} = (P_m - s)\%$.
4. Reset the weights of the remaining portion of the network to their values in θ_0 . That is, let $\theta = \theta_0$.
5. Let $m = m'$ and repeat steps 2 through 4 until a sufficiently pruned network has been obtained.

Strategy 2: Iterative pruning with continued training.

1. Randomly initialize a neural network $f(x; m \odot \theta)$ where $\theta = \theta_0$ and $m = 1^{|\theta|}$ is a mask.
2. Train the network for j iterations.
3. Prune $s\%$ of the parameters, creating an updated mask m' where $P_{m'} = (P_m - s)\%$.
4. Let $m = m'$ and repeat steps 2 and 3 until a sufficiently pruned network has been obtained.
5. Reset the weights of the remaining portion of the network to their values in θ_0 . That is, let $\theta = \theta_0$.

對比這兩種策略，不同之處在於策略 1 在每次反覆運算裁剪後仍用原初始權重開始訓練，策略 2 則是用上一次反覆運算得到的權重繼續訓練。而在確定子網路結構後，兩種策略都是用原初始權重進行訓練。實驗驗證策略 1 下子網路最終收斂的性能會明顯好於策略 2，因此文中之後的對比實驗都是基於前者進行。

研究結果：

文章先在 MNIST 和 CIFAR10 上用淺層的全連接和卷積神經網路進行對比實驗，結果表明採用上述策略得到的稀疏子網路用原網路的初始權重初始化相比于重新初始化在相同訓練設置下最終模型收斂性能有顯著性提升，並且在一定稀疏率條件下甚至超過原大網路。更進一步地，同樣在 CIFAR10 上，當採用更深的網路如 VGG19 和 ResNet18 時，需要對子網路採用更小的學習速率並結合

Warmup 等訓練策略才能找到“中獎彩票”。從這些實驗結果出發，作者希望能啟發更多溝通權重初始化、網路結構與歸納偏置、神經網路泛化性及優化的研究和
工作。

論文題目： *Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations*

中文題目：挑戰無監督表徵學習中的解耦合假設

論文出處：The 36th International Conference on Machine Learning – ICML2019

論文地址：<https://arxiv.org/pdf/1811.12359.pdf>

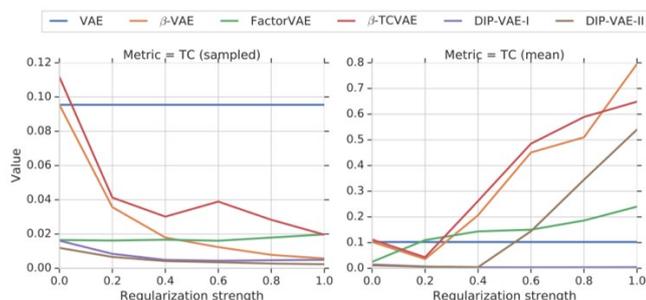
研究問題：

在使用生成模型進行表徵學習時通常假設資料的生成過程由式 $P(x) = P(z)P(x|z)$ 表示，其中 x 表示原始資料如圖片， z 則對應於語義明確的可解釋變數並假設不同分量間相互獨立。解耦合表徵學習的目的即學習到某種特徵變換 $r(x)$ 能與這些解釋變數間構成對應關係，使得學到的特徵更加高效魯棒便於下游任務提取任務相關的資訊。本文通過搭配已有解耦合表徵學習方法和不同資料集及評價指標構建不同實驗設置，批判性指出現有工作的不足和基本假設上的缺陷。

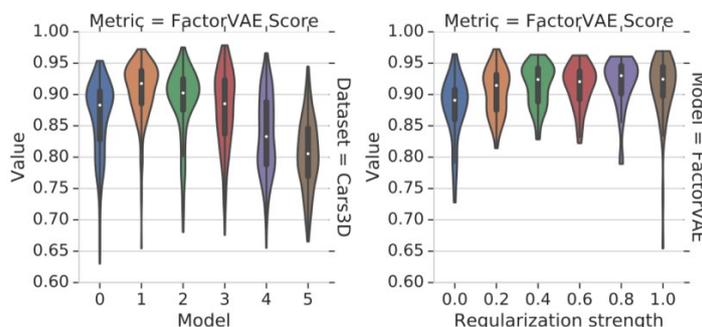
研究方法：

首先文章從理論上給出一個直覺性的結論，即任何有效的解耦合表徵學習必須基於某種對模型和資料分佈的歸納偏置。接著為了進一步驗證其他基本假設，文章考慮了在 6 種不同已有方法與評價指標，以及 7 個不同資料集下進行對比實驗。這裡我們簡要說明幾個主要實驗及其結果：

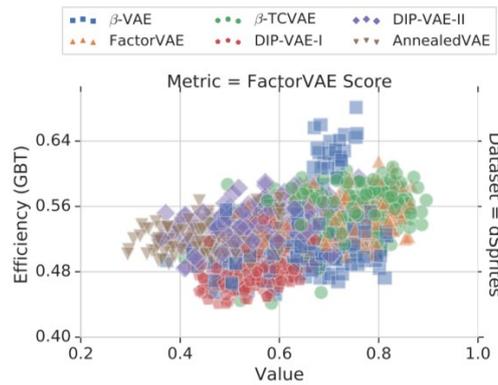
- 1) 已有工作中為了使學得的滿足正態分佈先驗的隱層特徵 z 解相關，大都在訓練過程中對編碼器的輸出加入相關性的正則項。然而如下右圖所示，這種方式下訓練最後得到的表徵不同維度間的相關性反而不斷上升。



- 2) 相比於不同方法的選擇，訓練超參甚至亂數種子對模型最終的性能表現更為重要，這從下圖每個方法在不同超參和亂數種子下性能表現的方差上可以看出。



- 3) 直覺上我們認為解耦合的表徵能更好輔助下游任務的訓練，下圖衡量了不同評分表徵學習模型的樣本利用效率，在實驗給定評價指標下並未明顯觀察到表徵解耦合帶來的樣本利用效率的提升。



研究結果：

總的來說，本文首先強調了歸納偏置在無監督解耦合表徵學習中的重要性並給出理論上的證明。然後實驗部分，文章主要歸納出以下幾點結論：1、對隱層特徵 z 後驗分佈加入的可分解約束並不能保證學得解耦合的特徵表示；2、訓練超參和亂數種子對不同方法性能的影響甚至超過方法本身帶來的性能差異；3、在當前解耦合表徵性能的衡量指標下，更好的解耦合表示並未使得下游任務樣本利用效率上的明顯提升。

7.5 電腦圖形學進展

隨著數位化技術和互聯網的發展，電腦圖形學在許多領域都已經得到了廣泛的應用，如遙感圖像分析、多媒體通信、醫療診斷、機器人視覺等。當前電腦圖形學的研究逐漸向多學科交叉融合方向發展，即有與認知計算、計算器學習、人機交互的融合，也有與大資料分析、視覺化的融合；不僅針對三維數位模型，而且涵蓋了圖像視頻，與電腦視覺深度交叉。電腦圖形學的快速發展，一個潛在的趨勢是不再有明確清晰的主題，更多的體現出方法和技術的創新。

針對近兩年電腦圖形學重要期刊會議的相關論文，對該領域內容熱點研究內容和前沿技術方法進行了綜合分析。目前，熱點研究內容主要集中在自監督學習 (Self-Supervised Learning)、全景分割 (Panoptic Segmentation)、網路結構搜索 (Neural Architecture Search) 和生成式對抗網路 (Generative Adversarial Networks) 等方面。

自監督學習研究早期主要集中在代理任務的設計和選取上，怎樣的代理任務才能最好地提取出有益於下游任務的特徵以及為何這些代理任務能夠有效，這些是理論層面上自監督學習仍需要解決的問題。隨著大量圍繞著實例判別代理任務的相關工作的提出，有一些工作將其中的核心思想進行展開提出了所謂對比學習的概念。通過將原來兩個圖片實例特徵間的對比延伸到任意兩個模態間特徵的對比，使得模型學習不同模態間一致的特徵表達並用最大化互資訊作為新的衡量準則。

在已有的工作中，比較典型的代理任務有將圖片分塊然後預測不同分塊間的相對位置或者將分塊打亂後重排得到原圖，以及基於圖片的上下文資訊進行補全和圖片不同顏色通道間的相互預測等。目前在圖像與圖形學領域，取得性能突破的方法主要仍局限在監督學習的框架之下，隨著無標記資料的不斷爆增和模型性能進一步提升的需求，無監督學習將會越來越受到學術界和工業界的重視。而作為目前無監督學習中的一支，自監督學習因其良好的特徵判別能力和對大規模資料擴展能力，也將受到更廣泛的關注^[12-23]。

全景分割作為一個統一的任務在 2018 年被提出，它的目標是為圖像中的所有圖元點都分配一個語義類別和一個實例編號，從另一個角度來說，全景分割演

算法需要預測出圖像中每一個圖元點的所屬類別和所屬實例。在全景分割任務的基礎上，近期的進展主要體現在三個方面：（1）從圖像整體的角度考慮全景分割，共用網路主幹（backbone）形成設計整體網路結構；（2）考慮圖像中不同元素之間的交互，建模物體與語義概念之間的關係；（3）提出可學習模組，解決預測結果層面的衝突。接下來，我們將分別介紹有代表性的工作。全景分割作為一個最近被提出的視覺任務，受到了很大的關注，目前方法也在探討的過程中，具有很大的發展潛力^{[24-38][24]}。

目前深度學習的方法在各類圖像與圖形分析任務中取得了非常大的成功，伴隨這一成功而來的是對網路結構設計要求的不斷提高。自動化網路設計自然而然地成為了自動化機器學習的下一個目標。早期的相關工作證明了使用強化學習演算法可以發現好的網路架構，但是這些方法在計算過程中需要消耗大量計算資源，因此後續的工作都集中在如何減少計算負擔上。搜索空間的設計也是一項重要研究熱點，同時，研究人員又拓寬了神經結構搜索的視野，將多種優化目標考慮在內，而不僅僅是減少搜索時間和提高網路精度。具有代表性的工作如嘗試限制模型參數的數量或類似的方法，以有效地部署在移動設備上。在此基礎上，還有一些工作將網路結構搜索技術擴展到搜索深度網路相關元件上^[39-48]。

在圖像合成方面，近期最引人關注的工作就是生成對抗網路，生成對抗網路由一個生成網路 G 和一個判別網路 D 組成。生成網路 G 和判別網路 D 在訓練階段使用對抗的方式進行學習，生成網路 G 的目標是生成盡可能真實的圖片使得判別網路認為這是一張真實的圖片；而判別網路 D 的任務則是判別合成的圖像是真實的還是生成的。在這種兩者對抗的學習過程中，生成 G 學會如何生成真

實的图片。目前在生成對抗網路研究中，條件生成對抗網路、損失函數的改進、模型結構的改進及訓練方法的改進是主要研究方向^[49-55]。

AMiner

8 多媒體技術

8.1 多媒體概念

“多媒體”一詞譯自英文“Multimedia”，而該詞又是由 multiple 和 media 複合而成，核心詞是媒體。媒體 (medium) 在電腦領域有兩種含義：一是指存儲資訊的實體，如磁片、光碟、磁帶、半導體記憶體等，中文常譯為媒質；二是指傳遞資訊的載體，如數位、文字、聲音、圖形和圖像等，中文譯作媒介，多媒體技術中的媒體是指後者。其實，“媒體”的概念範圍是相當廣泛的。“媒體”有下列五大類：

(1) 感覺媒體 (Perception medium) 指的是能使人產生直接感覺的媒體。如聲音、動畫、文本等；(2) 表示媒體 (Representation medium) 指的是為了傳送感覺媒體而人為研究出來的媒體。諸如語言編碼、電報碼、條碼等等；(3) 顯示媒體 (Presentation medium) 指的是用於通信中使電信號和感覺媒體之間產生轉換用的媒體。如鍵盤、滑鼠、印表機等；(4) 存儲媒體 (Storage medium) 指的是於存放某種媒體的媒體。如紙張、磁帶、磁片、光碟等；(5) 傳輸媒體 (Transmission medium) 指的是用於傳輸某些媒體的媒體。常用的有如電話線、電纜、光纖等^[56]。

多媒體技術就是融電腦、聲音、文本、圖像、動畫、視頻和通信等多種功能於一體的技術，它借助日益普及的高速資訊網，可實現電腦的全球聯網和資訊資源分享，並且它給傳統的電腦系統、音訊和視頻設備帶來了方向性的變革，將對大眾傳媒產生深遠的影響。因此多媒體將加速電腦進入家庭和社會各個方面的進程，給人們的工作、生活和娛樂帶來深刻的革命。多媒體技術涉及的內容包括：

- 多媒體資料壓縮：多模態轉換、壓縮編碼；

-
- 多媒體處理：音訊資訊處理，如音樂合成、語音辨識、文字與語音相互轉換；影像處理，虛擬實境；
 - 多媒體資料存儲：多媒體資料庫；
 - 多媒體資料檢索：基於內容的圖像檢索，視頻檢索；
 - 多媒體著作工具：多媒體同步、超媒體和超文字；
 - 多媒體通信與分散式多媒體：CSCW、會議系統、VOD 和系統設計；
 - 多媒體專用設備技術：多媒體專用晶片技術，多媒體專用輸入輸出技術；
 - 多媒體應用技術：CAI 與遠端教學，GIS 與數位地球、多媒體遠端監控等^[57]。

8.2 多媒體技術發展歷史

- 啟蒙發展階段

多媒體技術的一些概念和方法，起源於 20 世紀 60 年代。1965 年，納爾遜 (Ted Nelson) 為電腦處理文字檔提出了一種把文本中遇到的相關文本組織在一起的方法，並為這種方法杜撰了一個詞，稱為“hypertext (超文字)”。與傳統的方式不同，超文字以非線性方式組織文本，使電腦能夠回應人的思維以及能夠方便地獲取所需要的資訊。萬維網 (WWW) 上的多媒體資訊正是採用了超文字思想與技術，組成了全球範圍的超媒體空間。

多媒體技術實現於 20 世紀 80 年代中期。1984 年美國 Apple 公司在研製 Macintosh 電腦時，為了增加圖形處理功能，改善人機交互介面，創造性地使用了位元映射 (bitmap)、視窗 (window)、圖符 (icon) 等技術。這一系列改進

所帶來的圖形化使用者介面 (GUI) 深受用戶的歡迎，加上引入滑鼠作為交互設備，配合 GUI 使用，大大方便了用戶的操作。Apple 公司在 1987 年又引入了“超級卡”(Hypercard)，使 Macintosh 機成為更容易使用、易學習並且能處理多媒體資訊的機器，受到電腦使用者的一致讚譽。

- 標準化階段

自 20 世紀 90 年代以來，多媒體技術逐漸成熟。多媒體技術從以研究開發為重心轉移到以應用為重心。

由於多媒體技術是一種綜合性技術，它的實用化涉及到電腦、電子、通信、影視等多個行業技術協作，其產品的應用目標，既涉及研究人員也面向普通消費者，涉及各個使用者層次，因此標準化問題是多媒體技術實用化的關鍵。在標準化階段，研究部門和開發部門首先各自提出自己的方案，然後經分析、測試、比較、綜合，總結出最優、最便於應用推廣的標準，指導多媒體產品的研製。

靜態圖像的一個標準，是國際電信聯盟 (ITU) 的 T.81。靜態圖像的主要標準稱為 JPEG 標準 (ISO/IEC10918)。它是 ISO 和 IEC 聯合成立的專家組 JPEG (Joint Photographic Experts Group) 建立的適用於單色和彩色、多灰度連續色調靜態圖像國際標準。該標準在 1991 年通過，成為 ISO/IEC10918 標準，全稱為“多灰度靜態圖像的數位壓縮編碼”。

視頻/運動圖像的主要標準是國際標準組織 (ISO) 下屬的一個專家組 MPEG (Moving Picture Experts Group) 制定的 MPEG-1 (ISO/IEC11172)、MPEG-2 (ISO/IEC13818) 和 MPEG-4 (ISO/IEC14496) 三個標準。與 MPEG-1、4 等效

的國際電信聯盟(ITU)標準,在運動圖像方面有用於視訊會議的 H.261(Px64),用於可視電話的 H.263。

在多媒體數位通信方面(包括電視會議等)制定了一系列國際標準(表 01-03-2),稱為 H 系列標準。這個系列標準分為兩代。H.320、H.321 和 H.322 是第一代標準,都以 1990 年通過的 ISDN 網路上的 H.320 為基礎。H.323、H.324 和 H.310 是第二代,使用新的 H.245 控制協定並且支援一系列改進的多媒體編、解碼器。

更深層次的多媒體技術標準也開始推出或列入開發中。一個典型的標準是稱作“多媒體內容描述介面”的 MPEG-7 標準(ISO/IEC15938)。與已經推出的幾個 MPEG 標準不同,MPEG-7 是一個關於表示音/視資訊的標準。它的七個組成部件中,系統、描述定義語言(DDL)、視頻、音訊和多媒體描述方案等已經成為正式標準,參考軟體和一致性測試則計畫在 2002 年 9 月成為標準。

- 蓬勃發展時期

隨著多媒體各種標準的制定和應用,極大地推動了多媒體產業的發展。很多多媒體標準和實現方法(如 JPEG、MPEG 等)已被做到晶片級,並作為成熟的商品投入市場。與此同時,涉及到多媒體領域的各種軟體系統及工具,也如雨後春筍,層出不窮。這些既解決了多媒體發展過程必須解決的難題,又對多媒體的普及和應用提供了可靠的技術保障,並促使多媒體成為一個產業而迅猛發展。

代表之一是進一步發展多媒體晶片和處理器。1997 年 1 月美國 Intel 公司推出了具有 MMX 技術的奔騰處理器(Pentium processor with MMX),使它成為多媒體電腦的一個標準。奔騰處理器在體系結構上有三個主要的特點:(1)增加

了新的指令，使電腦硬體本身就具有多媒體的處理功能（新添 57 個多媒體指令集），能更有效地處理視頻、音訊和圖形資料。（2）單條指令多資料處理（SIMD, Single Instruction Multiple Dataprocess）減少了視頻、音訊、圖形和動畫處理中常有的耗時的多迴圈。（3）更大的片內快取記憶體，減少了處理器不得不訪問片外低速記憶體的次數。奔騰處理器使多媒體的運行速度成倍增加，並已開始取代一些普通的功能卡板。

隨著網路電腦（Internet PC、NC）及新一代消費性電子產品，如電視機上盒（Set-Top Box）、DVD、視頻電話（Video Phone）、視訊會議（Video Conference）等觀念的崛起，強調應用於影像及通訊處理上最佳的數位訊號處理器（DSP），經過另一番的結構包裝，可由軟體驅動組態的方式，進入諮詢及消費性的多媒體處理器市場。

現在多媒體技術及應用正在向更深層次發展。下一代使用者介面，基於內容的多媒體資訊檢索，保證服務品質的多媒體全光通信網，基於高速互聯網的新一代分散式多媒體資訊系統等等，多媒體技術和它的應用正在迅速發展，新的技術、新的應用、新的系統不斷湧現^[58]。

8.3 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為多媒體領域全球學者分佈情況。

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量優勢明顯且主要分佈在其東西海岸；亞洲東部也有較多的人才分佈；歐洲的人才主要集中在歐洲中西部；其他諸如非洲、南美洲等地區的學者非常稀少；多媒體領域的人才分佈與各地區的科技、經濟實力情況大體一致。



圖 8-1 多媒體領域全球學者分佈

此外，在性別比例方面，多媒體領域中男性學者占比 91.7%，女性學者占比 8.3%，男性學者占比遠高於女性學者。

多媒體領域學者的 h-index 分佈如下圖所示，大部分學者的 h-index 分佈在中低區域，其中 h-index 在 20-30 區間的人數最多，有 691 人，占比 34.2%，50-56 區間的人數最少，有 124 人。

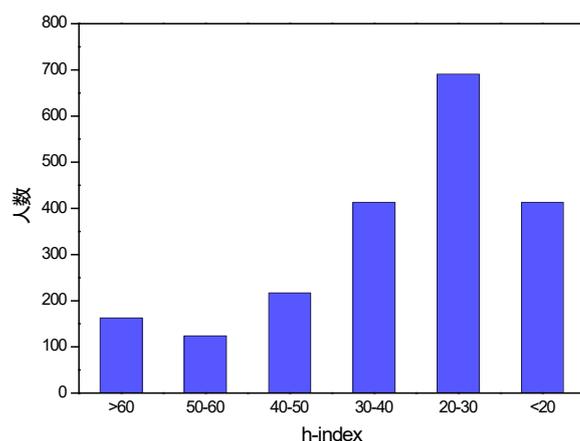


圖 8-2 多媒體領域學者 h-index 分佈

- 中國人才分佈

我國專家學者在多媒體領域的分佈如下圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是長三角和珠三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，通過觀察中國周邊國家的學者數量情況，特別是與日韓、東南亞等亞洲國家相比，中國在多媒體領域學者數量較多且有一定的優勢。

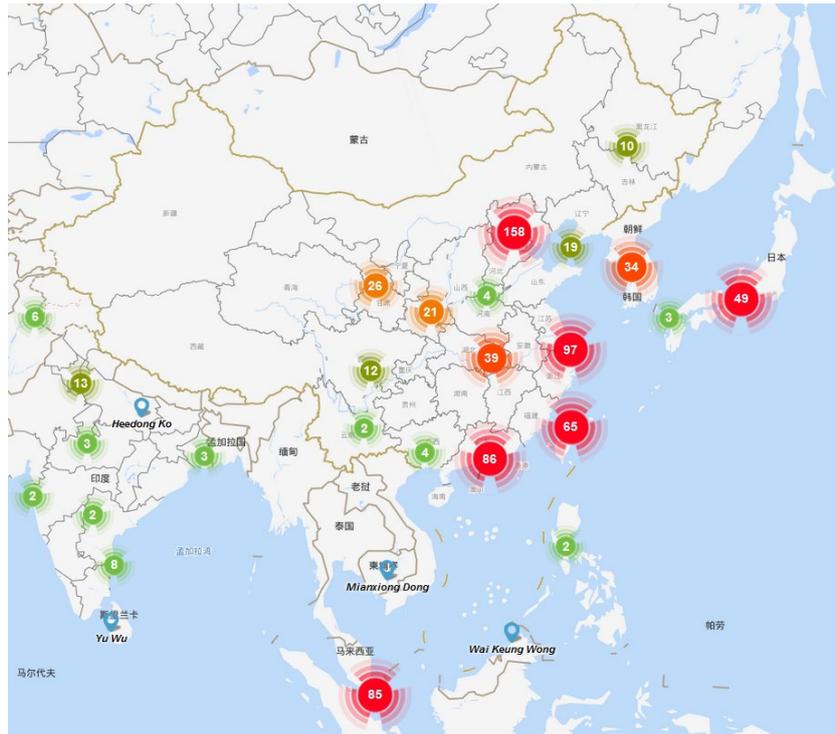


圖 8-3 多媒體領域中國學者分佈

中國與其他國家在多媒體領域的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 8-1 多媒體領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	676	18348	27	1107
中國-新加坡	231	6827	30	364
中國-澳大利亞	101	2919	29	166
中國-英國	71	1315	19	143
中國-加拿大	63	1261	20	117

中國-義大利	24	636	27	34
中國-巴基斯坦	17	412	24	23
中國-荷蘭	17	252	15	34
中國-德國	16	817	51	39
中國-法國	12	271	23	31

從上表資料可以看出，中美合作的論文數、引用數、學者數遙遙領先，表明中美間在多媒體領域合作之密切；同時，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 5 席；中國與德國合作的論文數雖然不是最多，但是擁有最高的平均引用數說明在合作品質上中德合作達到了較高的水準。

8.4 論文解讀

本節對本領域的高水準學術會議及期刊論文進行挖掘，解讀這些會議和期刊在 2018-2019 年的部分代表性工作。這些會議和期刊包括：

ACM International Conference on Multimedia

IEEE Transactions on Multimedia

我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如下圖所示。其中，多媒體(multimedia)、視頻(videos)、音訊 (audio) 是本領域中最熱的關鍵字。



論文題目： *Beyond Narrative Description: Generating Poetry from Images by Multi-Adversarial Training*

中文題目： 超越敘事描述:通過多重對抗訓練，從意象中生成詩歌

論文作者： Bei Liu , Jianlong Fu , Makoto P. Kato , Masatoshi Yoshikawa

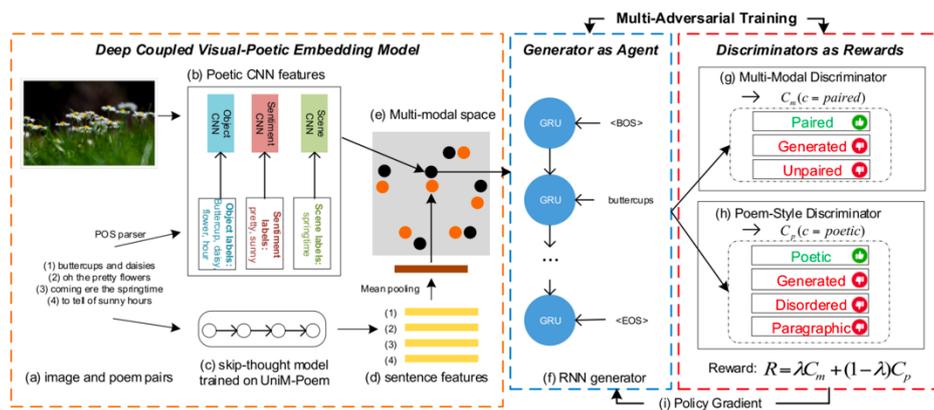
論文出處： 26th ACM International Conference on Multimedia – ACMMM’18

論文地址： <https://arxiv.org/pdf/1804.08473v4.pdf>

研究問題：

本文主要研究了從圖像自動生成詩歌的方法。這項任務涉及多個挑戰，包括從圖像中發現詩意線索（例如，從綠色中獲得希望），以及生成滿足圖像相關性和語言水準的詩意的詩歌。

研究方法：



本文通過使用策略梯度的多對抗訓練將詩歌生成的任務劃分為兩個相關的子任務，從而確保跨模式的關聯性和詩歌語言風格。為了從圖像中提取詩意線索，提出深度耦合的視覺詩意嵌入，可以同時學習圖像中的物件，情感和場景的詩歌表達。進一步引入兩個判別網路來指導詩歌的產生，包括多模式判別器和詩歌風格判別器。

如圖，為了更好地從圖像中學習詩歌線索以產生詩歌，首先從多模式詩歌資料集中學習具有圖像 CNN 特徵和具有 skip-thought 向量特徵的深度耦合視覺詩詞嵌入模型。然後，使用該嵌入模型從詩歌語料庫 UniM-Poem 中檢索相關且多樣的詩歌。這些檢索到的詩歌和 MultiM-Poem 的圖像一起構成了一個放大的圖像-詩對資料集 MultiM-Poem (Ex)。

進一步利用最新的順序學習技術來訓練端到端的圖像到詩歌模型。為避免因為長序列的長度過長 (所有詩行在一起) 而導致的曝光偏差問題，以及沒有可用于評分生成的詩的特定損失的問題，使用遞迴神經網路 (RNN) 來生成多首詩，並採用對抗訓練，通過政策梯度進一步優化。

研究結果：

本文發佈了人類注釋的具有兩個不同屬性的兩個詩歌資料集：1) 第一個人類注釋的圖像-詩歌對資料集 (共有 8 對 292 對) ; 2) 迄今為止最大的公共英語詩歌語料庫資料集 (共包含 92,265 種不同的詩歌) 。在 MultiM-Poem , UniM-Poem 和 MultiM-Poem (Ex) 上進行實驗 , 以生成圖像詩歌。生成的詩歌以客觀和主觀兩種方式進行評估。其中定義了有關相關性、新穎性和翻譯一致性的自動評估指標 , 並就生成的詩歌的相關性、連貫性和想像力進行了用戶研究 , 本文提出的模型在客觀評估和主觀評估中都顯示出優於從圖像生成詩歌的最新方法的優異性能。對 500 多個人類受試者進行的圖靈測試 , 其中 30 名評估者是詩歌專家 , 證明了方法的有效性。

論文題目 : *Audiovisual Zooming: What You See Is What You Hear*

中文題目 : 視聽縮放 : 你所看到的就是你所聽到的

論文作者 : Arun Asokan Nair , Austin Reiter , Changxi Zheng , Shree Nayar

論文出處 : 27th ACM International Conference on Multimedia – ACM-MM'19

論文地址 : <https://dl.acm.org/citation.cfm?id=3351010>

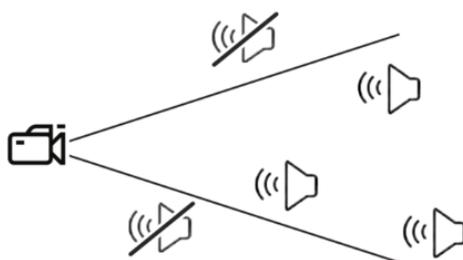
研究問題 :

在移動平臺上捕獲視頻時 , 感興趣的目標通常會被周圍的環境污染。為了消除視覺上的不相關性 , 攝像機平移和縮放提供了隔離所需視場的方法。但是 , 捕獲的音訊仍然被 FOV 外部的信號污染。這種效果是不自然的。本文提出了視聽縮放的概念 , 從而形成聽覺場以匹配視覺。框架建立在經典的波束成形概念之上 , 波束成形是一種使用麥克風陣列從單個方向增強聲音的計算方法。但是 , 波束成

形本身不能包含聽覺 FOV，因為 FOV 可能包括任意數量的定向源。本文將視聽縮放公式化為廣義特徵值問題，並提出了一種在移動平臺上進行有效計算的演算法。

研究方法：

在這項工作中，增加了麥克風陣列和波束成形方法，以實現視聽縮放，而無需從訓練資料中學習。受麥克風陣列波束成形的激勵，可將單個麥克風採樣的信號視為某些潛在隨機過程的隨機變數。從這個角度出發，本文在頻域中估計了兩個複值矩陣，稱為頻譜矩陣：一個描述來自 FOV 內部的麥克風信號的自相關和互相關，另一個描述來自 FOV 外部的信號。使用這兩個矩陣，可以將增強向 FOV 的問題公式化為可以在移動設備上輕鬆解決的廣義特徵值問題。為了分析本方法，本文得出了頻譜矩陣估計的理論誤差範圍，並揭示了誤差殘差與經典最小方差無失真回應 (MVDR) 波束形成器性能的聯繫。根據經驗，通過模擬，瞭解各種設計參數如何影響麥克風陣列。這些推論為實驗提供了資訊。最終演算法很簡單，可以輕鬆地部署在移動設備上。



研究結果：

本文提供了演算法元件的理論分析以及用於理解麥克風陣列各種設計選擇的數值研究。最後，在兩個不同的移動平臺上演示了視聽縮放：移動智慧手機和用於視訊會議設置的 360°球形成像系統。

論文題目： *Emotion Recognition in Speech using Cross-Modal Transfer in the Wild*

中文題目：在語音辨識中應用跨模態傳輸進行情感識別

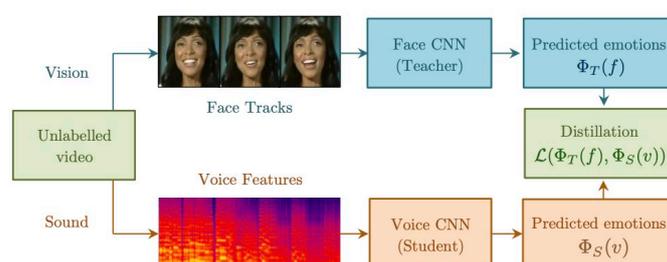
論文作者：Samuel Albanie, Arsha Nagrani, Andrea Vedaldi, Andrew Zisserman

論文出處: MM2018

論文地址：<https://arxiv.org/pdf/1808.05561.pdf>

研究問題：

傳統的語音情感識別往往需要大量的標注資料，與此同時，資料的標注往往是模糊不清的，有大量雜訊。本文基於這樣一個簡單的假設：人在說話時臉像的情感和語音情感是一致的，利用這種一致性，本文提出了一種 teacher-student 模型，將視覺的情感資訊遷移到語音，取得了很好的實驗結果。



研究方法：

本文提出了一種基於 teacher-student 的 cross-modal transfer 模型，將視覺上的臉像情感資訊遷移到語音。對於一段無標注的視頻，視覺上臉像的預測結果為

音訊的預測結果提供了 soft-label , 相比於人手工標注的 hard label , teacher-student 的監督模式極大地提升了語音情感識別的性能。

研究結果：

本文已經證明了使用大資料集的情緒未標記的視頻的價值，情緒的跨模態轉移從臉到語言。這種方法的好處是顯而易見的，語音情感模型在標準的基準上獲得了合理的分類性能，結果遠遠高於隨機。作者還在 FERPlus 基準(supervised) 上實現了面部情緒識別的更新性能，並在兩個標準資料集(RML 和 eNTERFACE) 上為語音情緒識別的交叉模式蒸餾方法設置了基準。這種方法的優點是視頻資料幾乎是無限的，可以從 YouTube 和其他來源免費獲得。

論文題目： *Multi-View Image Generation from a Single-View*

中文題目：從單視圖生成多視圖圖像

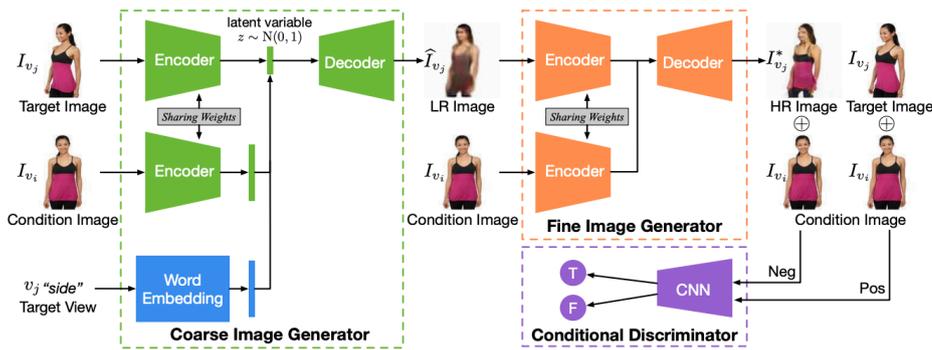
論文作者：Bo Zhao, Xiao Wu, Zhi-Qi Cheng, Hao Liu, Zequn Jie, Jiashi Feng

論文出處：MM2018

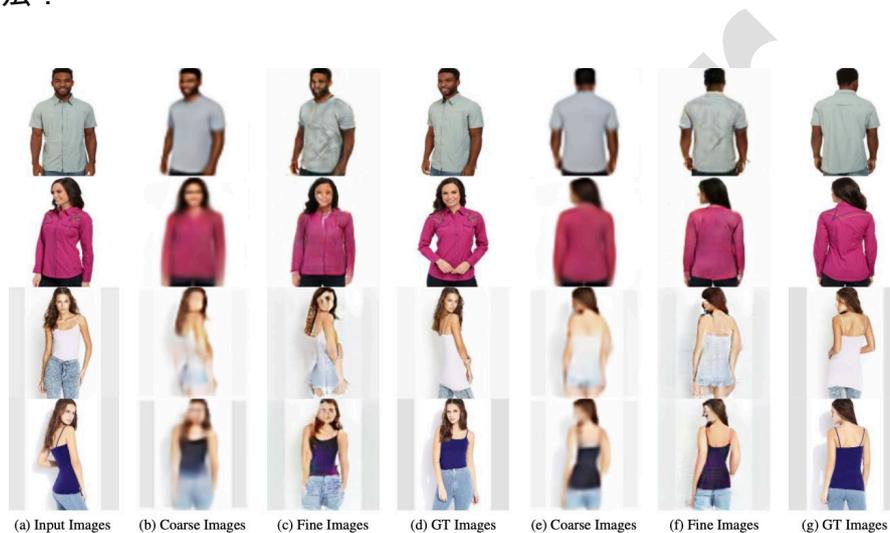
論文地址：<https://arxiv.org/pdf/1704.04886.pdf>

研究問題：

本文旨在解決利用一個視角的圖像合成多視角圖像的問題，這一問題極具挑戰性。作者提出了將變分推斷和對抗生成網路結合的 VariGAN，使用了一種從粗到細的方式，首先生成低解析度的對應視角圖片，再利用對抗訓練的方法提升圖像解析度。該方法在多個資料集上取得了很好的結果。



研究方法：



為了解決單視角圖片合成多視角圖片的問題，作者提出了 VariGAN，具體如下：（1）粗細微性生成器首先用變分推斷的方式生成對應視角的圖像，變分自編碼器以原始圖像，目標視角為輸入，輸出低解析度圖像，並用變分自編碼器對應的損失監督訓練。（2）細細微性的模型將低解析度的圖像精細化，用對抗訓練的損失作為監督，生成了更加細細微性的圖像。在具體實現中，作者設計了雙路徑 U-Net 來實現細細微性生成器，最終在 MVC 和 Deepfashion 資料集上取得了較好的結果。

研究結果：

本文提出了一種變分生成對抗網路 (VariGANs) 來合成以不同視角為輸入

圖像的現實服裝圖像。該方法利用變分推理的方法，實現了由粗到精的圖像生成。具體地說，提供具有一定視圖的輸入圖像，粗圖像生成器首先生成具有目標視圖的物件的基本形狀。然後用精細圖像生成器將細節資訊填充到粗圖像中，並對缺陷進行修正。通過大量的實驗，該模型可以得到比現有方法更合理的結果。消融研究也證實了所提出的靜脈曲張各組成部分的重要性。

論文題目： *Understanding the Teaching Styles by an Attention based Multi-task Cross-media Dimensional modelling*

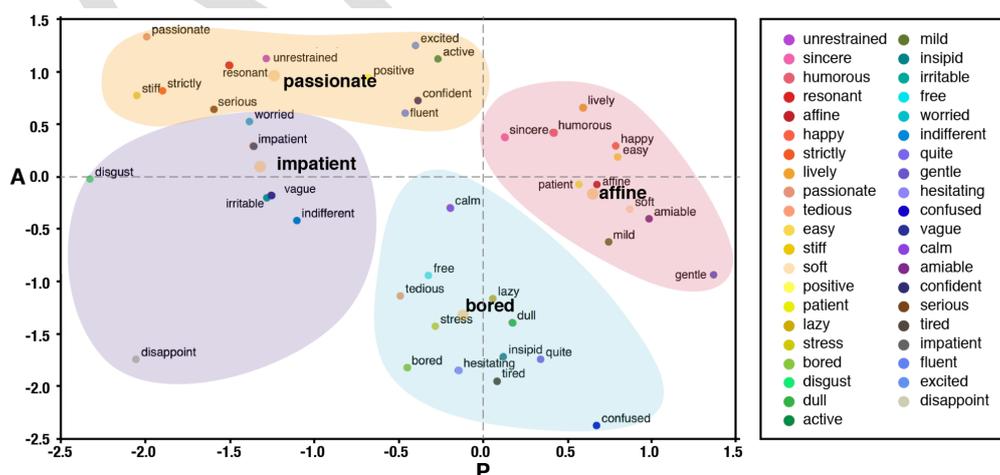
中文題目：通過一個基於注意力的多工跨媒體維度建模來理解教學風格

論文作者：Suping Zhou, Jia Jia, Yufeng Yin, Xiang Li, Yang Yao, Ying Zhang, Zeyang Ye, Kehua Lei, Yan Huang, Jialie Shen

論文出處：27th ACM International Conference on Multimedia – ACMMM'19

論文地址：<https://dl.acm.org/citation.cfm?id=3351059>

研究問題：



教師的授課風格在幫助學生更好學習中起著重要的作用。文章探討了如何有效的分析理解教師的授課風格。具體來說，文章研究了 1) 如何定量描述不同教

師的教學風格；2) 如何對教師的跨模態教學資料(言語、面部表情和身體動作、內容等) 與教學風格之間的關係進行建模。

研究方法：

首先，基於愉悅度啟動度維度理論，文章構建了一個二維教學風格語義空間 (TSSS)，對教學風格進行定量、全面的描述。從好未來教育集團提供的 10000 多份家長對教師評價的回饋問卷中，篩選出了最常用的 41 個授課風格形容詞，並在 TSSS 上手工標注授課風格形容詞的座標，即愉悅度啟動度值。

同時，文章提出了一種基於注意力機制的多路徑多工深度神經網路 (AMMDNN)，該神經網路能夠準確、可靠地捕捉跨模態特徵與 TSSS 之間的內在聯繫。文章利用從好未來教育集團收集的 4541 句跨模態教學資料集，設計了大量的測試實驗來評估 AMMDNN 對跨模態特徵和 TSSS 上座標值之間的映射效果。

研究結果：

結果表明，所提出的 AMMDNN 模型性能均優於基線模型 (對於 CCC 一致性相關係數指標，平均提升 0.0842)。同時，文章將二維座標與教學風格形容詞聯繫起來，可以更合理、更生動地描述教學風格。最後，文章還進行了一些有趣的案例研究，包括不同教師和課程之間的教學風格比較，以及利用所提出的方法進行教學品質分析。

論文題目：*Dance with Melody: An LSTM-autoencoder Approach to Music-oriented Dance Synthesis*

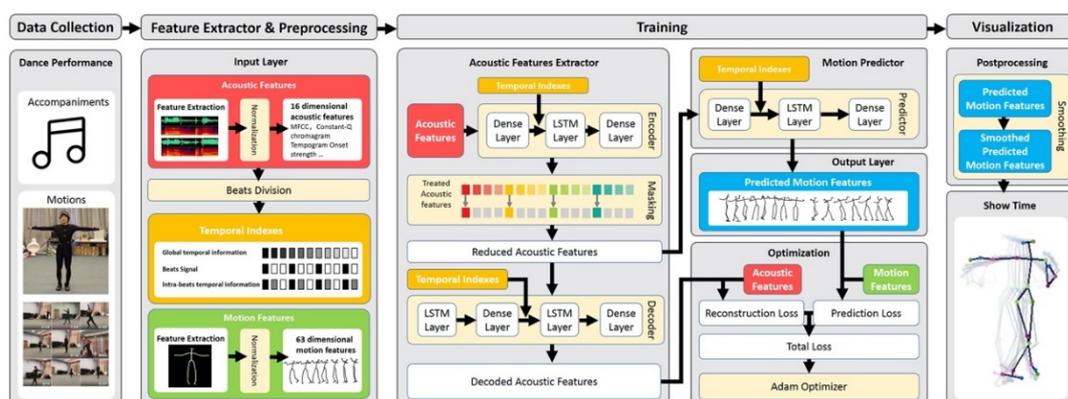
中文題目：與旋律共舞：一種面向音樂的舞蹈合成的 LSTM-autoencoder 方法

論文作者： Taoran Tang , Jia Jia , Hanyang Mao

論文出處： 26th ACM International Conference on Multimedia – ACMMM'18

論文地址： <https://dl.acm.org/citation.cfm?id=3240526>

研究問題：



舞蹈深受音樂的影響。對音樂舞蹈編舞的綜合研究可以促進舞蹈教學和人類行為研究等多個領域的研究。儘管在研究音樂和舞蹈之間的關係方面已經付出了相當大的努力，但是基於音樂的舞蹈編排的綜合仍然是一個開放的問題。目前該問題主要有兩個挑戰：1) 如何根據音樂選擇合適的舞姿，即技術舞蹈手冊中命名和指定的舞步組；2) 如何根據音樂藝術地加強舞蹈編排。

研究方法：

為了解決這些問題，文章提出了一種基於長短時記憶網路 (LSTM)-自動編碼器的模型來解決面向音樂的舞蹈自動編排方法。具體地說，該模型旨在提取聲音和運動特徵之間的映射，以便音樂的情感將被合成的舞蹈所反映。因此，該模型可以學習舞者如何調整自己的局部關節姿勢和動作節奏，以表達音樂情感的變化，以及舞蹈編排中的動作選擇規則。

由於缺乏模型訓練所需的資料，文章構建了一個音樂舞蹈資料集，包含 40 個四種舞蹈類型的完整舞蹈編排，共 907200 幀，用光學運動捕捉設備 (Vicon) 採集。該資料集還記錄了舞蹈編排所用的音樂，使收集到的資料特別有助於面向音樂的舞蹈合成。提取了包括 63 維運動特徵、16 維聲學特徵和 3 維時間指標在內的綜合特徵。從而為神經網路的訓練提供了有針對性、精確性和完整性的特徵。目前這是最大的音樂舞蹈資料集，且該文章公開了資料集。

研究結果：

文章進行了一些定性和定量實驗來量化模型性能。由於舞蹈是一種藝術創作，除了考慮常用的歐幾裡德損失函數外，還考慮了使用者評價來評價模型的表現。實驗結果表明，與多個基線模型相比，文章所提出的模型成功地提取了與舞姿選擇相關的聲學特徵，在選擇與一段動畫片的長度、節奏和情感相匹配的舞姿時表現良好。

論文題目：*MMGCN : Multi-modal Graph Convolution Network for Personalized Recommendation of Micro-video*

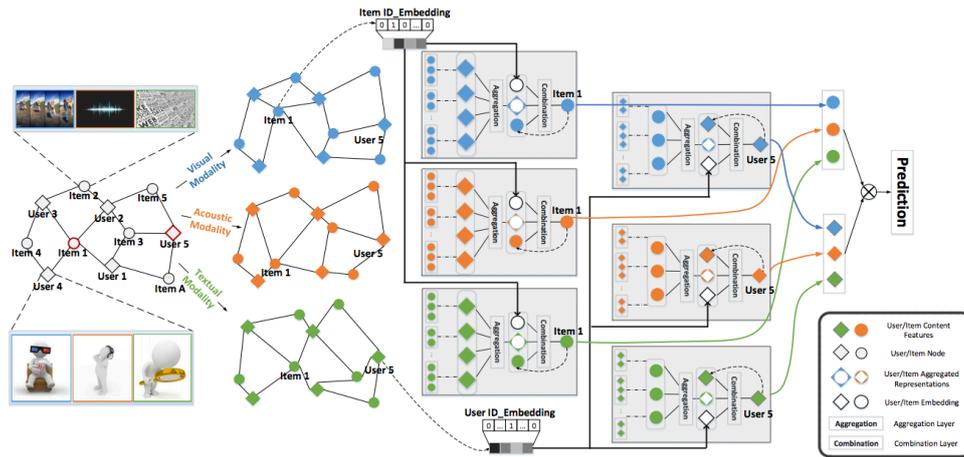
中文題目：MMGCN：用於微視頻個性化推薦的多模態圖卷積網路

論文作者：Yinwei Wei , Xiang Wang , Liqiang Nie , Xiangnan He , Richang Hong , Tat-Seng Chua

論文出處：Proceedings of the 27th ACM International Conference on Multimedia

論文地址：<https://dl.acm.org/citation.cfm?id=3351034>

研究問題：



個性化推薦在許多線上內容共用平臺中起著核心作用。為了提供優質的微視頻推薦服務，重要的是考慮用戶與專案（即短視頻）之間的交互以及來自各種模態（例如視覺，聽覺和文本）的專案內容。現有的多媒體推薦作品在很大程度上利用多模態內容來豐富專案表示，而為利用使用者和專案之間的資訊交換來增強使用者表示並進一步捕獲使用者對不同模式的細細微性偏好所做的工作卻較少。本文中利用用戶-專案交互來指導每種模式中的表示學習，並進一步個性化微視頻推薦。基於圖神經網路的消息傳遞思想設計了一個多模態圖卷積網路（MMGCN）框架，該框架可以生成使用者和微視頻的特定模態表示，以更好地捕獲用戶的偏好，如上圖所示。

研究方法：

在每個模態中構造一個 user-item 二部圖，並用其鄰居的拓撲結構和特徵豐富每個節點的表示。使用者的歷史記錄反應了個人的興趣，用戶群組也可以對項目進行建模。多模態圖卷積網路（MMGCN）包括三個部分：聚合層、組合層、預測層。聚合層對於每個模態（如視頻），將交互過專案的相應內容（如幀）通過彙總函式來衡量鄰居的影響，將結構資訊和鄰居分佈編碼得到使用者的表示；組合層融合了結構資訊、啟發式資訊、模態關聯，通過用戶群組來增強專案的表

示。通過遞迴地進行這種聚合和組合，使用者和專案的表示可以體現多跳鄰居的資訊，使得使用者關於特定模態的偏好可以被很好的表示。最後，對於未知交互的預測可以通過使用者和微視頻表示的相似度來得到。

研究結果：

在三個公開可用的資料集 Tiktok, Kwai 和 MovieLens 上進行的大量實驗，證明了論文提出的模型能夠明顯優於目前最新的多模態推薦方法。在本文中，作者明確地建模特定于模式的使用者偏好來增強微視頻推薦。作者設計了一個新的基於 gcn 的框架，稱為 MMGCN，以利用使用者和微視頻之間在多種模式下的資訊交換，細化它們的特定模式表示，並進一步類比使用者對微視頻的細微性偏好。在三個公開的微視頻資料集上的實驗結果很好地驗證了該的模型。此外，作者還視覺化了一些示例來演示特定於模式的使用者首選項

論文題目： *Routing Micro-videos via A Temporal Graph-guided Recommendation System*

作文題目：通過臨時圖形引導推薦系統路由微視頻

論文作者：Yongqi Li, Meng Liu, Jianhua Yin, Chaoran Cui, Xin-Shun Xu, Liqiang Nie

論文出處：Proceedings of the 27th ACM International Conference on Multimedia

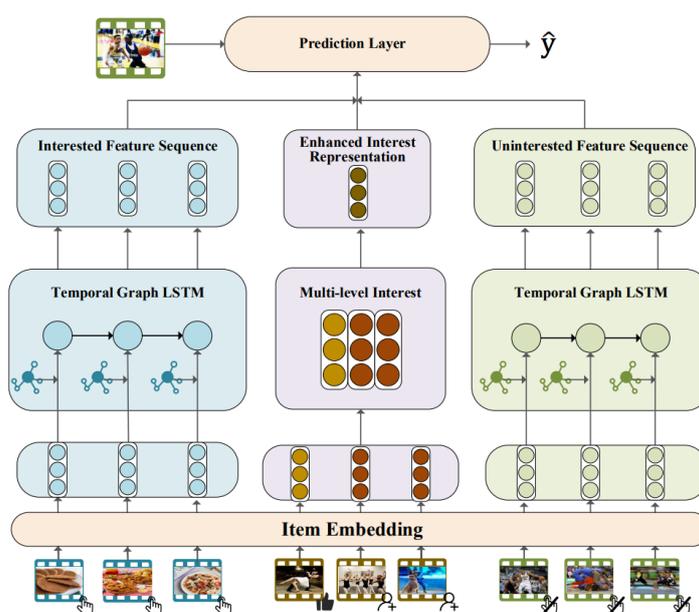
論文地址：<https://dl.acm.org/citation.cfm?id=3350950>

研究問題：

在過去的幾年中，短視頻已成為社交媒體時代的主流趨勢。同時，隨著短視頻數量的增加，使用者經常被他們不感興趣的視頻所淹沒。儘管現有的針對各種

社區的推薦系統已經取得了成功，但由於短視頻平臺中的使用者具有其獨特的特徵：多樣化的動態興趣，多層次的興趣以及負樣本，因此它們無法以一種好的方式應用於短視頻。為了解決這些問題，論文提出了一個時間圖指導的推薦系統（ALPINE）。首先設計了一個新穎的基於圖的順序網路，能夠同時對使用者的動態興趣和多樣化興趣進行建模。同樣，可以從使用者的真實負樣本中捕獲不感興趣的資訊。除此之外，通過使用者矩陣將使用者的多層次興趣引入推薦模型，該矩陣能夠學習用戶興趣的增強表示。最後，系統可以通過考慮上述特徵做出準確的推薦。

研究方法：



如上圖所示，ALPINE 模型由三個部分組成：基於圖的順序網路、多層次興趣建模層、預測層。基於圖的順序網路將興趣圖與長短時記憶網路結合，利用使用者的點擊和未點擊歷史行為，分別提取動態的、多樣的感興趣和不感興趣的特徵序列。多層次興趣建模層利用用戶的關注、點贊等行為，對用戶興趣的表示進行增強。最後，基於使用者的感興趣特徵序列、不感興趣特徵序列、增強的興趣

表示，分別將未知短視頻的表示與相應特徵通過 Vanilla Attention 得到增強的表示，再通過多層感知機得到相應的評分，將三個評分加權得到最終的預測輸出。

研究結果：

在兩個公共資料集上進行的實驗，證明了提出的 ALPINE 模型較其他推薦演算法取得了更好的性能，也證明了模型各組成部分的有效性。在這項工作中，作者提出了一個基於時間圖的 LSTM 模型來智慧地將微視頻路由到目標使用者。為了捕獲使用者動態的、多樣的興趣，作者將使用者的歷史交互序列編碼到時序圖中，設計了一種基於時序圖的 LSTM 模型。由於不同的交互反映了不同程度的興趣，作者構建了多層次的興趣建模層來增強用戶的興趣表示。此外，該模型從真實的陰性樣本中提取不感興趣的資訊，以提高推薦性能。為了驗證該方案，作者在兩個公共資料集上進行了大量的實驗，實驗結果證明了該模型的有效性。

8.5 多媒體技術進展

近年來，隨著數位化技術的發展，多媒體技術突飛猛進，音視頻技術是當前最活躍、發展最迅速的高新技術領域之一。多媒體分析以文本、圖像、聲音、視頻等多種不同類型媒體的資料為研究物件，主要的研究目的一方面是使電腦具備人類的多媒體（如視、聽）理解能力，另一方面是從多媒體資料中挖掘資訊和知識、說明人類更好地理解世界。

多媒體技術研究領域包括多媒體資訊處理、多媒體資料壓縮編碼、多媒體內容分析與檢索技術、多媒體交互與集成、多媒體通信與網路、多媒體內容安全、

多媒體系統與虛擬實境等。在近幾年的研究中，多媒體技術呈現出與電腦體系結構、電腦網路、人機交互、資訊安全、社會網路等多學科交叉融合的發展趨勢。

近兩年多媒體領域研究熱點主要集中在大規模圖像視頻分析、社會媒體研究、多模態人機交互、計算視覺、計算圖像、即時視頻流化等方面。

由於多媒體資料往往是多種資訊的傳遞媒介(例如一段視頻中往往會同時使得文字資訊、視覺資訊和聽覺資訊得到傳播)，多模態學習已逐漸發展為多媒體內容分析與理解的主要手段。

在計算圖像方面，大規模資料集的構建仍是一個熱點研究方向，尤其語義物件的圖元級標注需求越來越強烈，能夠人機交互標注的過程中不斷學習的協同標注方法得到了廣泛關注。

無監督學習是多媒體資料分析的長遠目標。目前很多領域擁有大量的資料，但是這些資料都是沒有經過標記的。因此除了基本的資料勘探和異常檢測場景，這些資料基本無法使用。近期在使用未標記的資料來改進(標記資料)監督學習過程方面已經取得了許多進展。

此外自動機器學習(AutoML)和元學習(Meta Learning)的最新研究成果及其在多媒體上的應用也逐漸增多。

在圖像壓縮處理方面，也有一些研究工作將深度學習用於圖像或視訊壓縮後處理，並得到了一定的效果。然而，現有工作的一個主要問題是用於後處理的深度網路較為複雜，計算速度慢，不滿足實際應用的需求。如何在處理效果和處理速度之間取得一個折中，是壓縮後處理的一個主要挑戰。

圖神經網路 (Graph Neural Network, GNN) 在多媒體領域的應用是近兩年的熱點研究方向, 應用場景包括: 個性化推薦, 如基於多模態圖卷積網路(MMGCN) 的多模態推薦方法; 短視頻推薦, 如使用基於圖的順序網路進行建模; 多視頻摘要, 如採用圖卷積網路衡量每個視頻的重要性和相關性; 基於文本的行人搜索, 如使用深度對抗圖注意力卷積網路 (A-GANet) 利用文本和視覺場景圖學習聯合特徵空間; 視頻關係監測, 如使用轉移圖神經網路 (DoT-GNN) 解決圖像外觀變化的問題。

隨著 Mask-RCNN 與 RetinaNet 的發展, 物體檢測研究日趨成熟, 但即便如此, 就應用而言, 當前的技術依然存在諸多缺陷, 為此, 針對現代目標檢測的基本框架 (backbone、head、scale、batchsize 與 post-processing), 神經網路架構搜索 (NAS) 以及細細微性圖像分析 (FGIA) 等 3 個方面的潛在難題成為主要研究內容, 尤其是後兩者, 將成為未來視覺物體檢測的兩個重要研究維度。

9 人機交互技術

9.1 人機交互概念

人機交互 (Human-Computer Interaction, HCI) , 是人與電腦之間為完成某項任務所進行的資訊交換過程, 是一門研究系統與使用者之間的交互關係的學問。系統可以是各種各樣的機器, 也可以是電腦化的系統和軟體。人機交互介面通常是指使用者的可見部分, 使用者通過人機交互介面與系統交流, 並進行操作。人機交互技術是電腦使用者介面設計中的重要內容之一, 與認知學、人機工程學、心理學等學科領域有密切的聯繫。

目前關於人機交互的定義主要有三種: 一是 ACM(Association for Computing Machinery) 的觀點, 它將人機交互定義為: 有關交互電腦系統設計、評估、實現以及與之相關現象的學科; 二是伯明罕大學教授 AlanDix 的觀點: 他認為人機交互是研究人、電腦以及他們之間相互作用方式的學科, 學習人機交互的目的是使電腦技術更好地為人類服務; 三是賓夕法尼亞州立大學 JohnM.Carroll 的觀點: 他認為人機交互指的是有關可用性的學習和實踐, 是關於理解和構建使用者樂於使用的軟體和技術, 並能在使用時發現產品有效性的學科。無論是哪一種定義方式, 人機交互所關注的首要問題都是人與電腦之間的關係問題。

人機交互技術的發展與國民經濟發展有著直接的聯繫, 它是使信息技術融入社會, 深入群體, 達到廣泛應用的技術門檻。任何一種新交互技術的誕生, 都會帶來其新的應用人群, 新的應用領域, 帶來巨大的社會經濟效益, 從企業的角度, 改善人機交互能夠提高員工的生產效率; 學習人機交互能夠降低產品的後續支援

成本。從個人的角度，可以說明使用者有效地降低錯誤發生的概率，避免由於錯誤引發的損失。在現代和未來的社會裡，只要有人利用通信、電腦等資訊處理技術進行社會活動時，人機交互都是永恆的主題，鑒於它對科技發展的重要性，人機交互是現代資訊技術、人工智慧技術研究的熱門方向^[59]。

9.2 人機交互發展歷史

人機交互的發展歷史，是從人適應電腦到電腦不斷地適應人的發展史。交互的資訊也由精確的輸入輸出資訊變成非精確的輸入輸出資訊。

9.2.1 簡單人機交互

由於受到製造技術和成本等原因限制，早期的人機交互在設計上較少考慮人的因素，強調輸入輸出資訊的精確性，使用不夠自然和高效^[60]。

- 早期的手工作業

這個時期交互的特點是由設計者（或本部門同事）來使用電腦，他們採用手工操作和依賴機器（二進位機器代碼）的方法去適應電腦。

- 作業控制語言及交互命令語言

這一階段特點是電腦的主要使用者——程式師可採用批次處理作業語言或交互命令語言的方式和電腦打交道，雖然要記憶許多命令和熟練地敲鍵盤，但已可用較方便的手段來偵錯工具、瞭解電腦執行情況。

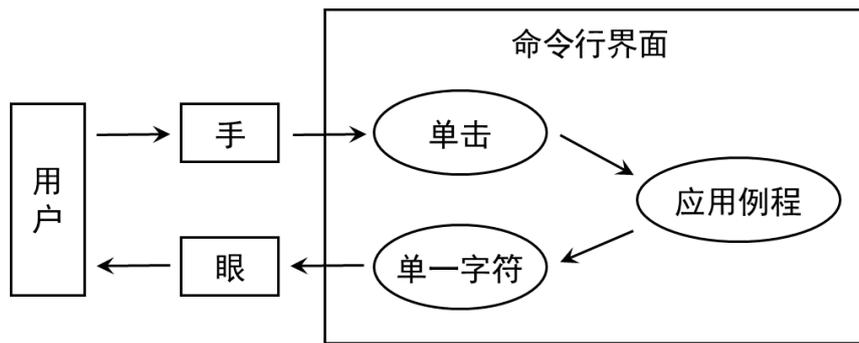


圖 9-1 命令列介面概念模型

- 圖形化使用者介面 (GUI)

圖形化使用者介面的主要特點是桌面隱喻、WIMP (Window / Icon / Menu / Pointing Device) 技術、直接操縱和“所見即所得 (WYSIWYG)”。由於 GUI 簡明易學、減少了敲鍵盤、實現了“事實上的標準化”。因而使不懂電腦的普通使用者也可以熟練地使用，開拓了用戶人群。它的出現使資訊產業得到空前的發展。

- 網路使用者介面

以超文字標記語言 HTML 及超文字傳輸協定 HTTP 為主要基礎的網路流覽器是網路使用者介面的代表。由它形成的萬維網 (World Wide Web, WWW) 已經成為當今 Internet 的支柱。這類人機交互技術的特點是發展快，新的技術不斷出現，如搜尋引擎、網路加速、多媒體動畫、聊天工具等。

9.2.2 自然人機交互

隨著網路的普及性發展和無線通訊技術的發展，人機交互領域面臨著巨大的挑戰和機遇，傳統的圖形介面交互已經產生了本質的變化，人們的需求不再局限於介面美學形式的創新，用戶更多的希望在使用多媒體終端時，有著更便捷、更符合他們使用習慣，同時又有著比較美觀的操作介面。利用人的多種感覺通道和

動作通道 (如語音、手寫、姿勢、視線、表情等輸入) ，以並行、非精確的方式與 (可見或不可見的) 電腦環境進行交互，使人們從傳統的對話模式的束縛解脫出來，使人們進入自然和諧的人機交互時期。這一時期的主要研究內容包括：多通道交互、情感計算、自然語言理解、虛擬實境、智慧使用者介面等方面。

● 多通道交互

多通道交互 (Multi Modal Interaction , MMI) 是近年來迅速發展的一種人機交互技術，它既適應了“以人為中心”的自然交互準則，也推動了互聯網時代資訊產業 (包括移動計算、移動通信、網路服務器等) 的快速發展^[61]。

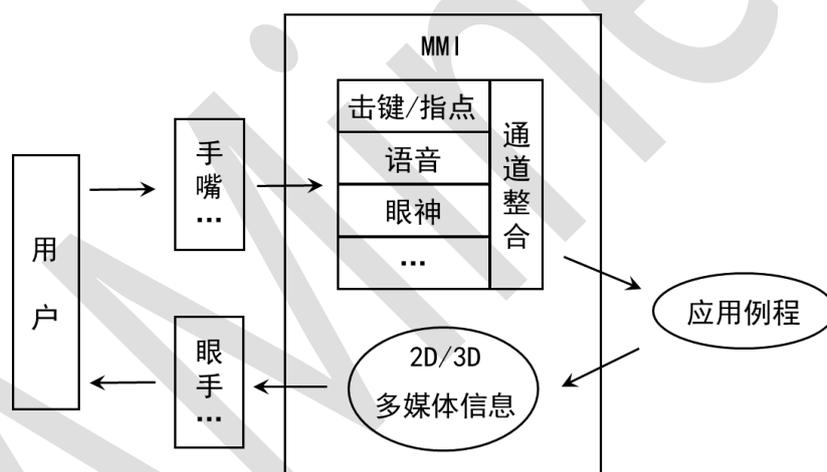


圖 9-2 命令列介面概念模型

MMI 是指“使用多種通道與電腦通信的人機對話模式。通道 (modality) 涵蓋了用戶表達意圖、執行動作或感知回饋資訊的各種通信方法，如言語、眼神、臉部表情、唇動、手動、手勢、頭動、肢體姿勢、觸覺、嗅覺或味覺等”。採用這種方式的電腦使用者介面稱為“多通道使用者介面”。目前，人類最長使用的多通道交互技術包括手寫辨識、筆式交互、語音辨識、語音合成、數位墨水、視線跟蹤技術、觸覺通道的力回饋裝置、生物特徵識別技術和人臉表情識別技術等方面。

- 情感人機交互

讓電腦具有情感能力首先是由美國 MIT 大學 Minsky 教授(人工智慧創始人之一) 提出的。他在 1985 年的專著“*The Society of Mind*”中指出，問題不在於智慧型機器能否有任何情感，而在於機器實現智慧時怎麼能夠沒有情感^[62]。從此，賦予電腦情感能力並讓電腦能夠理解和表達情感的研究、探討引起了電腦界許多人士的興趣。這方面的工作首推美國 MIT 媒體實驗室 Picard 教授領導研究小組的工作。情感計算一詞也首先由 Picard 教授于 1997 年出版的專著“*Affective Computing (情感計算)*”中提出並給出了定義，即情感計算是關於情感、情感產生以及影響情感方面的計算。

MIT 對情感計算進行全方位研究，正在開發研究情感機器人，最終有可能人機融合。其媒體實驗室與 HP 公司合作進行情感計算的研究。IBM 公司的“藍眼計畫”，可使電腦知道人想幹什麼，如當人的眼瞄向電視時，它竟然知道人想打開電視機，它便發出指令打開電視機。此外該公司還研究了情感滑鼠，可根據手部的血壓及溫度等感測器感知使用者的情感。CMU 主要研究可穿戴電腦。日本在對感性資訊處理的研究中，有眾多研究單位參與，主要集中在研究所和高校。特別值得一提的是，日本歐姆龍公司研製生產的機器玩具曾風行一時，最高價達 4000 美元。隨後其它公司也進行機器狗等玩具的生產。情感計算的研究不僅具有重要的科學和學術價值，也存在著巨大的商機，有很好的經濟效益。

- 虛擬實境

虛擬實境 (*Virtual Reality, VR*) 是以電腦技術為核心，結合相關科學技術，生成與一定範圍真實環境在視、聽、觸感等方面高度近似的數位化環境，用戶借

助必要的裝備與數位化環境中的物件進行交互作用、相互影響，可以產生親臨對應真實環境的感受和體驗。虛擬實境是人類在探索自然、認識自然過程中創造產生，逐步形成的一種用於認識自然、類比自然，進而更好地適應和利用自然的科學方法和科學技術。

虛擬實境技術具有很強的應用性。軍事方面，將 VR 技術應用於軍事演練，帶來軍事演練觀念和方式的變革，推動了軍事演練的發展。如美國的 SIMNET、ACTDSTOW、WARSIM2000 和虛擬之旗 2006 等一系列分散式虛擬戰場環境。醫學方面，VR 技術已初步應用於虛擬手術訓練、遠端會診、手術規劃及導航、遠端協作手術等方面，某些應用已成為醫療過程不可替代的重要手段和環節。工業領域方面，VR 技術多用於產品論證、設計、裝配、人機工效和性能評價等。代表性應用，如模擬訓練、虛擬樣機技術等已受到許多工業部門的重視。教育文化領域方面，VR 已經成為數字博物館/科學館、大型活動開閉幕式彩排模擬、沉浸式互動遊戲等應用系統的核心支撐技術。紐約大都會博物館、大英博物館、俄羅斯冬宮博物館和法國盧浮宮等都建立了自己的數字博物館。



圖 9-3 VR 參觀盧浮宮概念圖

- 智慧使用者介面

智慧使用者介面 (Intelligent User Interface , IUI) 是致力於改善人機交互的高效率、有效性和自然性的人機界面。它通過表達、推理，按照使用者模型、領域模型、任務模型、談話模型和媒體模型來實現人機交互。智慧使用者介面主要使用人工智慧技術去實現人機通信，提高了人機交互的可用性：如知識表示技術支援基於模型的使用者介面生成，規劃識別和生成支援使用者介面的對話管理，而語言、手勢和圖像理解支援多通道輸入的分析，用戶建模則實現了對自我調整交互的支持等。當然，智慧使用者介面也離不開認知心理學、人機工程學的支持。

智慧體、代理 (Agents) 在智慧技術中的重要性已“不言而喻”了。Agent 是一個能夠感知外界環境並具有自主行為能力的以實現其設計目標的自治系統。智慧的 Agent 系統可以根據使用者的喜好和需要配置具有個性化特點的應用程式。基於此技術，我們可以實現自我調整使用者系統、使用者建模和自我調整腦介面。自我調整系統方面，如說明使用者獲得資訊，推薦產品，介面自我調整，支援協同，接管例行工作，為使用者裁剪資訊，提供說明，支援學習和管理引導對話等。用戶建模方面，目前機器學習是主要的用戶建模方法，如神經網路、Bayesian 學習以及在推薦系統中常使用協同過濾演算法實現對個體使用者的推薦。自我調整腦介面方面，如神經分類器通過分析用戶的腦電波識別出用戶想要執行什麼任務 (該任務既可以是運動相關的任務如移動手臂，也可以是認知活動如做算術題)

- 自然語言人界交互

在“電腦文化”到來的社會裡，語言已不僅是人與人之間的交際工具，而且是人機對話的基礎，是最自然的一種人機對話模式。自然語言處理 (Natural Language Processing, NLP) 是使用自然語言同電腦進行通訊的技術，因為處理自

然語言的關鍵是要讓電腦“理解”自然語言，所以自然語言處理又叫做自然語言理解 (Natural Language Understanding, NLU)。

近年來自然語言理解技術在搜索技術方面得到了廣泛的應用，現在，已經有越來越多的搜尋引擎宣佈支援自然語言搜尋特性，自然語言人機交互介面在智慧短信服務、情報檢索、人機對話等方面也具有廣闊的發展前景和極高的應用價值，並有一些階段性成果出現在商業運用中。

9.3 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為人機交互領域全球學者分佈情況：



圖 9-4 人機交互領域全球學者分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量優勢明顯且主要分佈在其東西海岸；歐

洲也有較多的人才分佈；亞洲的人才主要集中在日韓地區；其他諸如非洲、南美洲等地區的學者非常稀少；人機交互領域的人才分佈與各地區的科技、經濟實力情況大體一致。

此外，在性別比例方面，人機交互領域中男性學者占比 84.6%，女性學者占比 15.4%，男性學者占比遠高於女性學者。

人機交互領域學者的 h-index 分佈如下圖所示，大部分學者的 h-index 分佈在中低區域，其中 h-index 在 20-30 區間的人數最多，有 842 人，占比 42.1%，50-60 區間的人數最少，有 136 人。

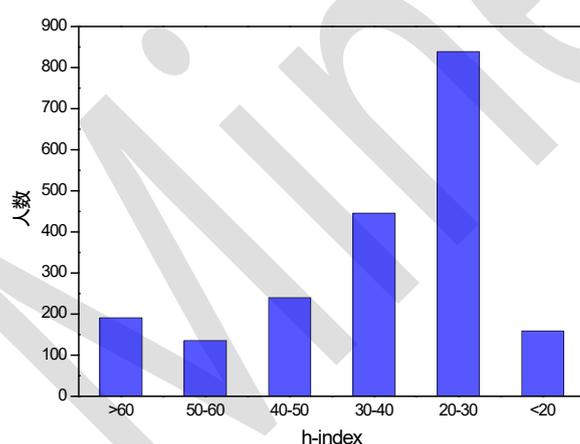


圖 9-5 人機交互領域學者 h-index 分佈

● 中國人才分佈

我國專家學者在人機交互領域的分佈如下圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是長三角和珠三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，通過觀察中國周邊國家的學者數量情況，特別是與日韓等地相比，中國在人機交互領域學者數量較少。

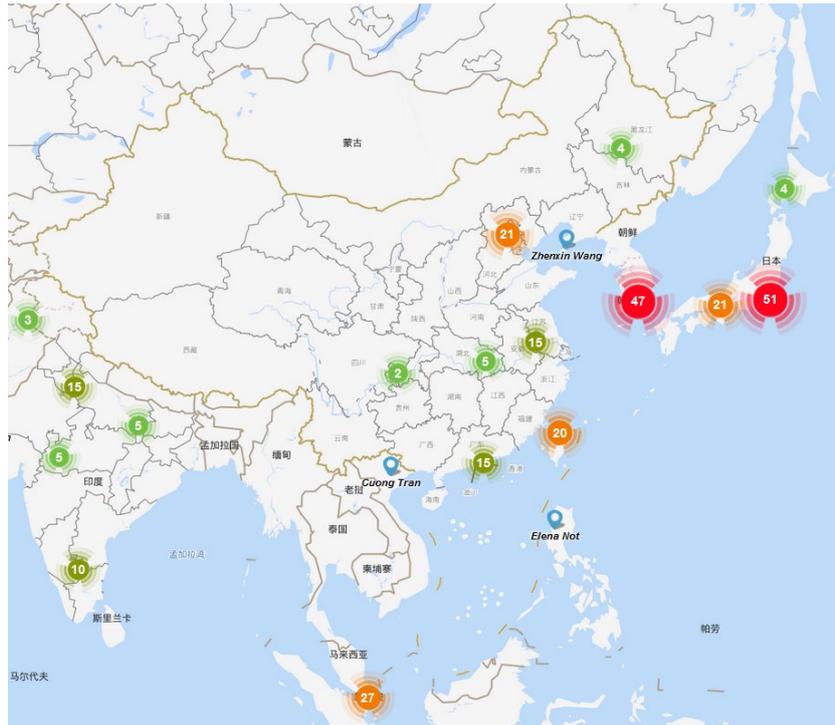


圖 9-6 人機交互領域中國學者分佈

中國與其他國家在人機交互領域的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 9-1 人機交互領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	142	3505	25	375
中國-日本	22	371	17	45
中國-英國	20	207	10	59
中國-新加坡	16	279	17	45
中國-德國	11	224	20	39
中國-加拿大	10	35	4	21
中國-芬蘭	9	26	3	22
中國-澳大利亞	8	193	24	20
中國-法國	7	174	25	20
中國-荷蘭	6	140	23	13

從上表資料可以看出，中美合作的論文數、引用數、學者數遙遙領先，表明中美間在人機交互領域合作之密切；此外，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 5 席；中國與美國，中國與法國合作的論文平均引用數都達到了最高，說明在合作品質上中美、中法合作達到了較高的水準。

9.4 論文解讀

本節對本領域的高水準學術會議及期刊論文進行挖掘，解讀這些會議和期刊在 2018-2019 年的部分代表性工作。這些會議和期刊包括：

ACM CHI Conference on Human Factors in Computing Systems

ACM Symposium on User Interface Software and Technology

ACM International Conference on Ubiquitous Computing

International Journal of Human Computer Studies

ACM Transactions on Computer-Human Interaction



我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如上圖所示。其中，用戶 (users)、虛擬實境 (virtual reality)、增強現實 (augmented reality) 是本領域中最熱的關鍵字。

論文題目： *Guidelines for human-AI interaction*

中文題目：人工智慧交互指南

論文作者：Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz

論文出處：ACM CHI Conference on Human Factors in Computing Systems 2019 (CHI 2019)

論文地址：<https://doi.org/10.1145/3290605.3300233>

研究問題：

人工智慧(AI)領域的快速發展給使用者介面和交互設計帶來了新的機遇和挑戰。雖然人機交互屆對人和 AI 交互的原則原理已經進行了 20 多年的探討，我們仍需要更多的研究和創新來解決人工智慧新技術及其面向人類的應用不斷湧現而帶來的新科學及社會問題。作者提出了 18 條具有通用性的、可適用於多種應用場景的、針對人和 AI 交互的設計指導，指出現有知識的空缺及未來的探索方向。這份指南不僅為 AI 設計師提供了具體、可操作的建議，還旨在推動使用者體驗和工程開發從業者就設計決策的相關問題展開討論，推動這一領域研究的

研究方法：

文章提出了 18 條人工智慧交互設計指導，並進行多輪的實例評估來驗證其有效性，包括通過一個用戶實驗，邀請 48 位設計師以這些設計指導為工具來測試 20 項廣泛使用的有 AI 技術支援的使用者產品。

指南內容包括：G1.說明使用者清晰的理解 AI 系統的能力所及；G2.幫使用者樹立對 AI 系統表現（如出錯率）的正確期待。；G3.根據上下文設計好服務回應和中斷的時機；G4.顯示與上下文（使用者當前任務及環境）相關的資訊；G5.確保使用者體驗與其社會及文化規範相一致；G6.切實減少 AI 系統的語音及行為中可能存在的刻板印象和偏見。G7.保證需要時能容易且迅速的啟動或請求 AI 系統的服務；G8.保證能夠方便的取消或忽略不適當的 AI 系統服務；G9.保證 AI 系統出錯時使用者可以對其進行編輯、修改、或者恢復設置；G10.AI 系統在不確定使用者需求或目的時可進行適當問詢或者以得體的方式降級服務；G11.向使用者適當解釋 AI 系統的行為及背後依據。G12.記住使用者最近的交互模式，通過短期記憶提升使用效率；G13.學習使用者長期的行為模式，提供個性化的體驗；G14.降低 AI 系統升級或進行自我調整調整時可能給用戶帶來的干擾；G15.引導使用者在日常使用過程中向 AI 系統回饋其交互體驗及偏好；G16.及時向使用者傳遞其行為可能對 AI 系統將來的服務帶來的影響；G17.讓使用者能夠在全域層面定制 AI 系統對資料的監測及其行為模式；G18.將 AI 系統的變更適時通知給用戶。

本文著重於在使用者介面審查過程中可以很容易評估的設計指導意見。因此，沒有包含諸如“建立信任”這樣的抽象原則，而是側重可觀測的、可能可以幫助建立信任的具體措施。

過去的工作也提出了不少影響人和 AI 交互系統可用性的設計方針，但主要適用於 AI 模型建立階段。未來可以進一步加強設計師和 AI 工程師的合作，在實際應用中進一步理解如何更好的評定不同指導意見的優先順序並平衡各方面的用戶體驗。

研究結果：

本文通過三輪的評估，從 150 多條 AI 相關的設計建議中提取並驗證了 18 條針對人與 AI 交互領域的通用設計指導意見。隨著越來越多的科技應用以 AI 作為驅動力，本文希望能通過推廣這些設計指導實現真正以人為本的 AI 交互系統。

本文研究整合了人機交互屆對人工智慧交互設計 20 多年的探索、思考、經驗和教訓，是一篇非常及時的綜述性文章。不同於時下對於如何讓 AI 模型更可用的針對 AI 從業人員研究，作者從設計師的角度出發，側重討論了在交互設計過程中需要注意的問題。另外，在提出設計指導意見時很好的權衡了適用範圍以及可操作性，為其結果真正在實際應用中推廣奠定了良好的基礎，有著現實的指導意義。

論文題目： *Voice Interfaces in Everyday Life*

中文題目：日常生活中的語音交互介面

論文作者：ACM CHI Conference on Human Factors in Computing Systems 2018(CHI 2018)

論文出處：CHI 2018, April 21–26, 2018, Montreal, QC, Canada

論文地址：<https://doi.org/10.1145/3173574.3174214>

研究問題：

隨著越來越多智慧產品內嵌語音交互介面（後簡稱 VUI），工業界和學界產出了不少計算語言學或自然語言理解方面的研究成果，但卻鮮有語音交互設備在生活場景中的實證使用調研。作者以此空白為契機，著重調研了語音交互和其他

日常活動並行時的使用者回饋，以及多方對話場景下的話論連續處理。文章也涉及更為寬泛的三個領域：VUI 發展進程，VUI 調研中的會話分析，設計、部署和研究 VUI 時需要的方法論。

研究方法：

該研究沿用 HCI 文獻中普遍的常人方法學和會話分析方法，記錄了五戶家庭一個月內使用 Echo 音箱的資料，收集了使用者與機器六小時的對話語料進行解讀。喚醒語音助手後，祈使指令和提問是常見的兩種觸發形式。對語音助手的指令會開啟和機器間的話論轉換；未收到音時的靜默能推動話論前進；用戶自發糾正說法也能推動話論延續。語音設備回復後，使用者會有後續回饋，本文暫且討論三種形式：靜默也是使用者的回復方式；用戶會明確指出問題；用戶也會糾正會話。

作者不認為語音交互的介面本質是對話的，用“對話介面”來命名基於語音的人機對話模式也不妥當。日常對話中的話論轉換和毗鄰對是自然產生的，而語音交互介面中的問法其實會刻意迎合會有的答案。因此，作者主張把任務型的 VUI 設計稱為請求/回復設計，而非對話設計。

研究結果：

本文通過解讀真實場景中和音箱的交互語料，展現了語音設備是如何並行參與到家庭對話中的；也發現除卻設備自身的功能可用性，用戶在社交場景中的行為也要為最終的 VUI 效果負責；同時本文還探討了用戶使用 VUI 時的觸發和接應方式；最後提出三點概念性問題為後續 VUI 的設計和研究提供方向。

本文研究可謂填補了語音設備在真實場景中的調研空白。除此之外，很多易於被忽視的事實也被作者拎了出來。技術人員常常關注語言計算，業界設計師常常掉入邏輯或話術細節，鮮有人高屋建瓴地研究 VUI 使用情況。諸如靜默也能推動話論前進、“對話設計”的命名糾正、關照用戶發問的易用性等觀點都是“跳出了盒子”的嶄新思考。

論文題目： *TipText: Eyes-free Text Entry on a Fingertip Keyboard*

中文題目：指尖上的鍵盤：可盲打的指尖鍵盤輸入法

論文作者：Zheer Xu, Pui Chung Wong, Jun Gong, Te-yen Wu, Aditya Nittala, Xiaojun Bi, Jürgen Steimle, Hongbo Fu, Kening Zhu, Xing-Dong Yang

論文出處：ACM Symposium on User Interface Software and Technology 2019 (UIST 2019)

論文地址：<https://doi.org/10.1145/3332165.3347865>

研究問題：

隨著可穿戴設備以及普適計算的普及，人們越來越需要一種易於攜帶且跨平臺通用的輸入裝置來進行交互。與此同時，基於拇指和食指指尖的微手勢作為一種新型的跨平臺對話模式有著得天獨厚的優勢：快速，簡單易學（把食指指尖當作觸控板，用大拇指進行點擊）而又隱蔽（有利於保護隱私）。這使得微手勢特別適合輸入空間十分有限的可穿戴設備。本文的主要目標就是為基於微手勢的文字輸入設計一種合理而高效的盲打鍵盤來論證可行性，並通過用戶實驗證明了這種輸入方法可以達到理想的輸入效率（在完成 40 個短句後速度可以達到每分鐘最高約 13.9 個單詞）。

研究方法：

本研究使用用戶調研 (user study) 的資料獲取方法和基於使用者盲打資料的電腦模擬 (computer-based simulation) 來為微手勢盲打鍵盤挑選出最好的鍵盤配置，並使用統計解碼 (statistical decoding) 方法來實現文本輸入，最後通過使用者評估實驗 (user evaluation) 來衡量整個輸入法系統的性能。

本文將字母聚合的概念針對性地引入到指尖鍵盤上(即每個鍵對應多個字母，類似於傳統的 9 鍵鍵盤)，並和觸控屏上文字輸入的統計解碼器創新性結合在一起，克服了高混淆度對文字輸入效率的影響。

本文選擇了基於電腦類比的計算設計 (computational design) 方法：通過先採集少量使用者資料，再使用電腦進行大規模類比使用者輸入的方式來量化評價每個候選佈局。這樣的流程提高了對使用者資料的利用效率，從而在一個極大的設計空間中科學而高效地尋找到了最符合人類行為模型的最優解；

在採集使用者資料的時候，本文開創性地採用了“在虛擬世界中重構現實世界”的方法：通過將動作追蹤系統和 3D 遊戲引擎結合的方式來精確追蹤拇指指尖和食指指尖的運動軌跡，從而計算出兩者接觸時的碰撞點。這種方法將感測器對用戶行為的潛在影響降低到最小，最大程度保證了實驗結論的嚴謹性。

研究結果：

本文通過開創性地將新興的微手勢交互技術運用到生活中常見的文字輸入任務上，證明了在指尖鍵盤上進行文字盲打的可行性，不僅進一步打開了應用微手勢交互技術的思路，也給進入了可穿戴設備時代的文字輸入系統設計帶來了更多的可能性。本文遵循計算設計的思想，根據使用者的行為資料最大程度為指尖

鍵盤優化了鍵盤配置，並在此基礎上實現了一個概念論證的原型系統，在使用者評估中表現出了令人滿意的系統性能，從而進一步證明了盲打指尖鍵盤可以高效地完成文字輸入任務。

本文首創了在極小的鍵盤上進行盲打文字輸入任務，為“在可穿戴設備上缺乏足夠的文字輸入空間”這一問題提供了全新的解決思路。與此同時，本文遵循著計算設計的理念，在一個極其複雜的設計空間中抽絲剝繭，最終找到了一個充分考慮指尖盲打鍵盤的特點的最優鍵盤配置。在使用者評估的過程中，整個系統展現出了很高的魯棒性和有效性，以及良好的性能。值得一提的是，本文在人機交互頂級會議 UIST 2019 上獲得了最佳論文的榮譽，究其原因在於為一個迫切的現實問題提供了嶄新而又極具啟發性的想法，並用科學嚴謹的實驗流程向讀者展示了如何將一個看似不可行的想法一步步實現的全過程，其中許多實驗及設計方法上的貢獻已經超越了文字輸入這一特定問題，給整個人機交互領域帶來更多的啟發和思考。

論文題目：*ElectroDermis: Fully Untethered, Stretchable, and Highly-Customizable Electronic Bandages*

中文題目：電子皮膚：完全不受束縛，可拉伸且高度可定制的電子創可貼

論文作者：Eric Markvicka, Guanyun Wang, Yi-Chin Lee, Gierad Laput, Carmel Majidi, and Lining Yao

論文出處：ACM CHI Conference on Human Factors in Computing Systems 2019(CHI 2019)

論文地址：<https://dl.acm.org/citation.cfm?id=3300862>

研究問題：

近些年柔性電子材料飛速發展，並在電子產品、醫療、健康檢測、柔性交互介面等領域進行著顛覆性的創新。然而，相比于傳統的電子設備或者電子器件，柔性電子顯示出來的功能和應用還並不強大。很大程度上這是由於缺少對柔性電子工藝的集成和製作流程的設計，使得其從技術到應用還存在不小的差距。本文以此為動機，研究可用於皮膚表面的柔性電子交互介面的完整的設計到製造的流程，並更好地支援柔性電子可拉伸及可定制的特性。

研究方法：

本文採用 HCI 領域常用的原型設計、製作、展示及評估的方法。文章首先對所研究的問題進行了闡述，提出了設計上需考慮的因素，並針對每一個研究因素提出了解決方案。然後文章通過原型及示例應用的展示提出了解決方案。最後文章進行了必要的評測及對未來工作的討論。

設計工具：文章首先開發了一套設計軟體，支援使用者通過簡單的操作在 3D 掃描的人體模型上標出想要應用的區域，如膝蓋、手肘等部位。軟體自動完成所選幾何平面的鋪平，以準備後續的實物切割和製作。軟體預存了電路設計圖並自動在鋪平的表面完成佈置。同時，軟體還嵌入了圖形樣式生成器，在不影響功能的同時提高了設備的美觀。

多層製造：本文採用了多層製造的工藝步驟。本文首先設計並製備了小型化的電子元器件，包括電池、控制器、若干感測器等，同時採用光刻機制成波浪形的連接銅導線，具有很好的伸縮性和可彎曲。本文考慮市面上常見的、容易買到的、可直接進行鐳射切割或模切成型、且具有很好的粘附性能的材料：選用氨綸

作為基質，可提高設備的彈性和實用性；選用醫用薄膜來粘接設備和皮膚，可支援快速的貼合和去掉。在進行多層拼接時，首先將銅包板層壓到 PDMS 基底上並進行切割，然後通過熱塑性的聚氨酯熱敏膜與之進行鍵合，形成的薄膜背面貼上之前準備的電子元器件，並將整體通過熱處理與氨綸基質進行粘合。

功能展示：本文作者通過該工藝展示了多種柔性交互介面的應用場景，包括可以檢測體溫，心跳的隨身貼，檢測飲食活動的智慧項鍊，檢測傷口癒合情況的創可貼，檢測環境中出現的顏色的指示燈，捕捉肢體運動的感測器等。每一個設備從設計到製作完成，都不超過 1 個小時。

性能評測：本文通過拉伸對設備進行了應變測試。通過觀察發現該工藝製作的連接線具有很好的拉伸能力，在 171%應變的情況下其導電能力才受到明顯的影響。本文也進行了旋扭拉伸測試，併發現在旋轉兩圈和 70%應變的情況下，設備的性能會收到明顯的影響。總體而言，本文展示的設計方案所呈現的柔性介面表現出了優秀的柔韌度。

研究結果：

本文展示了一套完整的柔性交互介面的設計和製作過程，通過多層結構的設計和材料的選取，使得設備不僅能夠適用於不同的交互場景，同時具有優秀的拉伸特性。這一製造工藝的設計，有利於新手設計師更便捷的進入到柔性電子設備的開發中。

本文的研究背景是柔性電子產業的飛速發展，帶來了電子產品新形態的產生，也推動著可穿戴設備的發展。在關注柔性電子技術的同時，本文作者發現從技術的發展到應用的落地還存在不小的差距，而彌補這一差距的途徑之一就是通過設

計和製造，使得柔性電子的優秀技術能夠快速的得到應用。本文的優勢是借薦了如柔性導線結構、柔性多層製造工藝等先進材料領域的成果，並帶入了如使用氮綸作為基質，醫用薄膜作為粘附層等創新的設計思路，完整地考慮了從設計到製造到交互應用的各個步驟。本文展示了近幾年人機交互與材料學科的交叉融合、設計及創新，已成為人機交互技術發展的一個重要領域。

論文題目： *ReconViguration: Reconfiguring Physical Keyboards in Virtual Reality*

中文題目：ReconViguration: 在虛擬實境中重新配置物理鍵盤

論文作者：Daniel Schneider, Alexander Otte, Travis Gesslein, Philipp Gagel, Bastian Kuth, Mohamad Shahm Damlakhi, Oliver Dietz, Eyal Ofek, Michel Pahud, Per Ola Kristensson, Jörg Müller, Jens Grubert

論文出處：IEEE International Symposium on Mixed and Augmented Reality 2019 (ISMAR 2019)

論文地址：https://ieeexplore_ieee.xilesou.top/abstract/document/8794572

研究問題：

迄今為止，鍵盤的物理佈局通常被移植到虛擬實境(VR)中，以便在虛擬的標準辦公環境中複製打字體驗。本文探討了如何充分利用 VR 的沉浸性，改變 VR 環境下物理鍵盤交互的輸入輸出特性。作者探索了一組輸入和輸出映射，用於重新配置物理鍵盤的虛擬模型，並通過具體設計、實施和評估 9 個與 VR 相關的應用程式來探索最終的設計空間：表情符號、語言和特殊字元、應用程式快捷方式、虛擬文本處理巨集、視窗管理器、照片流覽器、打地鼠遊戲、安全密碼輸入和虛擬觸控條。作者在 20 名用戶參與的研究中評估了這些應用程式的可行性，發現

它們在 VR 中是可用的。在實證研究和分析的基礎上，討論了 VR 中物理鍵盤輸入輸出特性重新映射的局限性和可能性，並指出了該領域未來的研究方向。

研究方法：

本研究在 VR 中重新配置物理鍵盤的按鍵，並招募了 20 名用戶來評估這種做法的可行性。首先，讓使用者使用 9 個 VR 應用程式來評估在 VR 環境中重新配置物理鍵盤的基本使用者體驗。其次，通過安全密碼輸入應用深入探討了客觀安全感和感知安全感之間的關係以及感知安全感和文本輸入性能之間的權衡。最後，通過虛擬觸控條應用分析了改變物理鍵盤的視覺表現形式對使用者體驗和性能的影響。應用程式細節如下：

語言和特殊字元：對於多語言輸入或特殊字元輸入的場景，可以把傳統鍵盤映射成為相應語言或特殊字元的鍵盤（類似于智慧手機中的多語種鍵盤）。

應用程式快捷方式：將傳統鍵盤中的部分按鍵映射成瀏覽器的後退、前進、刷新、主頁等快捷按鍵。

虛擬文本處理巨集：將按鍵映射成為 Microsoft Word 中的巨集命令（插入簽名/寄件者位址/圖片）。

視窗管理器：將鍵盤重新配置為視窗管理器，按下按鍵可以切換到對應的視窗。

照片流覽器：在鍵盤的一個按鍵或幾個按鍵上方顯示相應的照片縮略圖，按下相應按鍵可以流覽對應的照片。

打地鼠遊戲：將物理鍵盤重新配置為打地鼠遊戲，按下地鼠相應位置的按鍵可以打地鼠。

安全密碼輸入：將物理鍵盤原始的按鍵順序打亂，實現虛擬環境中安全輸入。

虛擬觸控條：將鍵盤上的一行 10 個按鍵虛擬為控制視頻播放進度的觸控條（類似於 MacBook Pro 上的 Touch Bar），通過按下相應按鍵控制播放進度。

研究結果：

本文通過重新配置單個按鍵或整個鍵盤的輸入輸出，設計了 9 個與 VR 相關的應用程式：表情符號、語言和特殊字元、應用程式快捷方式、虛擬文本處理巨集、視窗管理器、照片流覽器、打地鼠遊戲、安全密碼入口和虛擬觸控條，並通過招募 20 名參與者進行用戶研究，評估了在 VR 中重新配置物理鍵盤的可行性，發現這些應用程式在 VR 中是可用的。研究結果表明物理鍵盤可以在 VR 中以多種靈活的方式集成，作為 VR 中多種不同任務的輸入裝置，並且可以基於當前任務即時進行重新配置，在未來的 VR 應用中有著光明的前景。

本文利用 VR 的沉浸性特點，通過在 VR 環境中重新配置鍵盤的視覺顯示和功能，評估了重新配置鍵盤的可行性，提出了一種在 VR 環境中集成物理鍵盤的新思路。本文設計了 9 個 VR 應用，覆蓋了 VR 中辦公和遊戲的大部分應用場景，展現了傳統鍵盤輸入在 VR 中的應用潛力。值得一提的是，本文在混合現實和增強現實頂級會議 ISMAR 2019 上獲得了最佳論文的榮譽，究其原因在於將最傳統的輸入裝置以一種嶄新的應用形式帶入了 VR 中，並通過合理的用戶研究分析了其局限性和可行性，給 VR 中的人機交互帶來了更多的啟發和思考。

論文題目： *VIPBoard: Improving Screen-Reader Keyboard for Visually Impaired People with Character-Level Auto Correction*

中文題目：VIPBoard: 通過字元級別的自動糾錯為視障使用者優化讀屏鍵盤

論文作者：Weinan Shi, Chun Yu, Shuyi Fan, Feng Wang, Tong Wang, Xin Yi, Xiaojun Bi, Yuanchun Shi

論文出處：ACM CHI Conference on Human Factors in Computing Systems 2019(CHI 2019)

論文地址：<https://dl.acm.org/citation.cfm?doid=3290605.3300747>

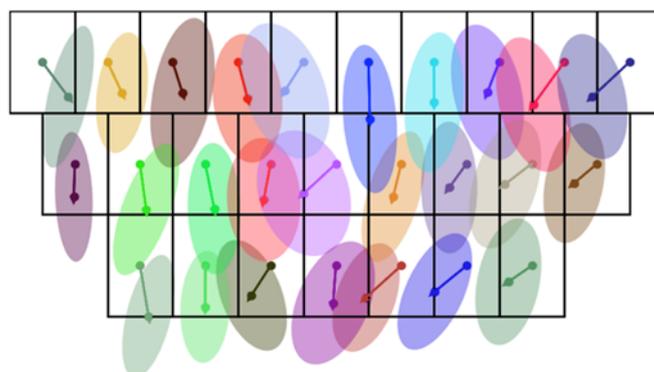
研究問題：

在使用軟鍵盤進行文本輸入時，視障使用者會根據讀屏軟體讀出的字元來確認輸入。因此視障使用者只能逐字元地、確保正確地輸入目標文本，輸入效率較低且無法使用當代先進的自動糾錯功能。本文提出了 VIPBoard，一種在不改變原有使用方式的前提下，專為視障使用者設計的智慧鍵盤。該鍵盤可以使用字元級別的自動糾錯演算法預測出使用者最有可能輸入的字元，並提供該預測字元的語音回饋，從而減少使用者輸入錯誤時所需的調整時間，提升輸入效率。使用者實驗證明 VIPBoard 可以顯著減少使用者的輸入錯誤 (63.0%)，提升文本輸入速度 (12.6%)。

研究方法：

本研究通過用戶調研 (user study) 的資料獲取方法得到了視障使用者在文本輸入時點擊的位置，擬合得出反映使用者行為的觸摸模型。演算法上使用基於貝葉斯原理的概率計算的方法預測用戶的輸入，通過多輪反覆運算的方法設計出了

最小化用戶學習成本的交互設計。最後通過用戶評估實驗 (user evaluation) 來衡量整個輸入法的性能。視障使用者進行文本輸入時的觸摸模型如下圖：



本文首次將自動糾錯應用到視障人士使用的讀屏軟鍵盤演算法中。由於通常地的詞級別糾錯演算法並不符合視障使用者逐字正確輸入的特徵，本文創新性地使用了字元級的自動糾錯演算法，提高了使用者輸入的準確性，從而減少了輸入過程中的調整次數，提高輸入速度。

本文結合視障使用者的使用行為特徵，設計了一套佈局自我調整策略。該策略首先保證了鍵盤使用時的魯棒性，即使用者可以在鍵盤上輸入任意想輸入的字元，無論預測結果是否正確。同時，策略允許用戶在不改變原有讀屏鍵盤使用習慣的前提下使用，從而大大減小了用戶的學習成本。

本文的用戶實驗評估進行地較為完備。文中實現了中文和英文兩種語言以及 VIPBoard 和傳統鍵盤兩種演算法下的系統原型，並讓使用者分別在兩種語言和兩種鍵盤下進行文本輸入。實驗結果不僅表明了 VIPBoard 相對於傳統鍵盤的巨大優勢，也說明了文中演算法的普適性和實際應用價值。

研究結果：

本文設計了一款面向視障使用者的智慧鍵盤，通過使用字元級的自動糾錯演算法減少了使用者輸入錯誤時的調整時間，提高了文本輸入的效率。同時，本文精心設計了與糾錯演算法相對應的佈局調整策略與對話模式，使得用戶可以在幾乎沒有學習成本的前提下使用該鍵盤。該鍵盤將現代智慧技術帶到了視障人士的生活中，為他們提供便利。

本文提出了世界上首款面向視障使用者的智慧鍵盤，開創性地將字元級的自動糾錯思想引入到了視障用戶的日常使用中，提高了輸入效率。在用戶評估中，整個技術體現出了相比于傳統顯著的性能優勢和使用者積極的主觀偏好。值得一提的是，本文在人機交互頂級會議 CHI2019 上獲得了最佳論文提名的榮譽，究其原因在於從特殊群體的特殊使用方式出發將已有的解決問題的思路遷移到該特定的場景下，提出了一種優雅的解決方案。同時，針對用戶的學習成本進行的交互設計優化和詳實的實驗設計，讓該工作實際應用價值體現地更為明顯，從技術角度給視障群體帶來一絲光明。

論文題目：*EarTouch: Facilitating Smartphone Use for Visually Impaired People in Mobile and Public Scenarios*

中文題目：耳勢交互：提升視力障礙用戶在移動和公眾場景下使用手機的體驗

論文作者：Ruolin Wang, Chun Yu, Xing-Dong Yang, Weijie He, Yuanchun Shi

論文出處：ACM CHI Conference on Human Factors in Computing Systems 2019(CHI 2019)

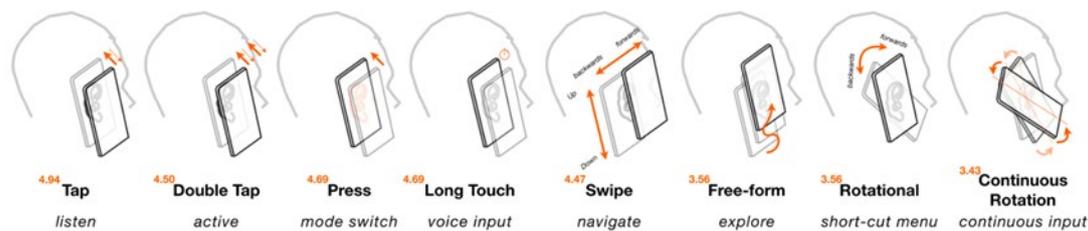
論文地址：<https://dl.acm.org/citation.cfm?id=3300254>

研究問題：

智慧觸屏手機已經成為視力障礙人群同世界溝通的重要工具。視力障礙使用者使用手機時主要採用的姿勢需要佔用雙手：一隻手握持手機，用另一隻手的手指在螢幕上觸摸，讀屏軟體能夠把交互內容轉化為語音讀出來。在公眾場景下，通過手機揚聲器聽取語音回饋會受到噪音的影響，引起洩漏隱私等擔憂，使用者往往需要將手機揚聲器舉近耳邊進行交互。而在移動出行的場景下，用戶一隻手被盲杖等物品佔用時，完成交互需要騰出雙手，轉換姿態，費時費力。此前針對視障使用者單手使用移動設備的研究工作集中於盲文的閱讀和輸入等有限的應用場景，而缺少一套面向手機上日常交互的無障礙解決方案。本文設計和開發了 EarTouch，一種基於耳勢的單手交互技術，能夠支援視力障礙用戶在移動和公眾場景下以一種相對穩定的使用姿態便捷地完成常用交互任務並輕鬆有效地獲取語音回饋。

研究方法：

本研究開展了有 30 位視障用戶參與的用戶訪談和 23 位視障用戶參與的設計工作坊，提出了一套包含八類耳勢動作的交互設計；基於電容感測器資料和慣性感測器資料識別耳勢動作，並在 16 位視力障礙用戶的操作資料集上實現了高準確度的評估效果；最後通過有 22 位視力障礙用戶參與的用戶實驗對可用性進行評估。結果表明，EarTouch 易於學習，在多數任務上執行效率更高，富有趣味性，能夠保護用戶的隱私並被用戶所接受，用戶期待將來能夠在自己的手機上使用到這項技術。EarTouch 選用的八類耳勢交互動作、功能，及視障用戶對於各類動作易於操作的評分的平均值(5 點李克特量表，5=非常同意該動作易於操作)。



EarTouch 拓展了智慧手機的輸入能力，使得耳朵能夠代替手作為觸摸輸入，因而用戶只需要佔用一隻手來握持手機。特殊地，耳朵靠近螢幕進行操作時也能夠同時用來從聽筒更加隱秘地接收語音回饋。另外，使用 EarTouch 時的握持姿態（耳朵位於聽筒附近、嘴位於話筒附近）也支援用戶在同一姿態下完成語音輸入。

研究結果：

基於分析電容感測器資料和慣性感測器資料的智慧演算法，EarTouch 支援手機識別耳朵在螢幕上的接觸、相對移動等動作，從而支持一系列交互操作如按一下、按兩下、旋轉等，使得視力障礙用戶在公眾和移動場景下能夠便捷地單手完成高頻、緊急的交互任務如接打電話、發送語音訊息、快捷導航等。EarTouch 支援語音回饋從手機上方的聽筒播放，無論是觸摸輸入、語音輸入或是獲取回饋，用戶都可以在一個相對穩定輕鬆的類似打電話的單手姿態下完成。在不便使用耳機時，從聽筒聽取回饋也能解決公眾場合下洩漏隱私和引起他人關注等問題。懷著基於用戶能力水準進行設計的理念，EarTouch 可以使更多用戶更容易地與智慧手機進行交互。雖然本文為視力障礙智慧手機用戶設計了 EarTouch，但該技術也可能使運動障礙的用戶（如單臂殘疾人等）和具有情境障礙的非殘疾用戶（如父母用一隻手抱著嬰兒時）受益。從更長遠的層面上來講，EarTouch 也為智慧手機面向不同交互能力的所有用戶的包容性設計邁出了重要的一步。

9.5 人機交互進展

最近的十年，是人機交互向自然交互蓬勃發展的十年。毋庸置疑，電腦是世紀最偉大的發明，其作用從科學計算工具迅速發展為資訊處理和資訊交互工具，起引領作用的則是人機交互技術的變革，即以滑鼠發明為標誌的圖形化使用者介面 (Graphical User Interface, GUI) 的產生，一改規範命令與電腦交互的命令列介面模式 (Command Line Interface , CLI) ， GUI 提供了普通人與電腦便捷交互的工具和方法，讓電腦從實驗室走進辦公室、走入家庭，十多年前，觸屏技術成為產品技術，GUI 中的滑鼠被人的天然指點 (pointing) 工具——手指所取代，電腦又變身出手機，成為更多人方便使用的隨身掌上工具。更少依賴操控工具，發展學習和使用成本更小的自然交互技術，一直是人機交互研究的價值追求，最近十年，隨著感知和計算技術的進步，自然交互技術創新層出不窮，並能迅速成為新型產品技術，《麻省理工科技評論》總結和評論人機交互領域的突破技術 (breakthroughs) ，為人機交互技術、未來終端技術的發展建立了一個高端的技術論壇，影響深遠。我把這些突破技術分為 3 大類：支援自然動作的感知技術，面向穿戴的新型終端和基於語音辨識的對話交互。

人體動作蘊含豐富的語義，動作交互技術一方面需要感知技術的進步，另一方面需要發現或設計有明確交互語義的動作 (gesture , 姿態 , 由於人手的靈巧性，手勢成為主要的交互動作，通常叫做手勢) ，如今，二維表面上，多指觸摸動作在觸屏上已普遍可用，三維空間中，嵌入了深度攝像頭的手持和固定設備，能比較準確識別人的姿態和動作，做出回應。不同於人臉識別等目標明確的視覺識別任務，動作交互不僅要求視覺識別的準確度，更需要研究基於交互任務的動作表

達的自然性與一致性，難以發現和突破，所以，除了動作語義很直白的動作遊戲（body game），三維動作交互尚缺少普遍認知和接受的交互動作語義。而無論二維還是三維，手勢的不可見性，是動作交互的主要難題。

穿戴（wearable）取代手持（handheld）曾是前幾年的一個革命口號，目前看，市場上的確出現了一定規模的新產品，但穿戴仍是補充的地位。穿戴設備中，手環設備基本只有健康和活動檢測功能，智慧手錶可以算做創新終端，但作為縮小版的手機，由於交互介面的縮小和操作方式的限制（通常是小介面上雙手參與操作），其承載功能也較手機縮減很多。VR / AR（虛擬實境 / 增強現實）的一個理想載體是頭戴式設備，最近幾年，多款智慧眼鏡產品面世，較之前笨重的頭盔輕便了許多，逼真的虛擬場景和準確的現實物件識別資訊都可以清晰呈現在眼前，並在特定領域開拓著增強體驗的應用；然而，智慧眼鏡尚缺少與其三維真實顯示匹配的準確的自然輸入技術，以及從眼手綁定在手機上轉變到眼手分離的眼鏡設備上時，尚未建立起相應的交互模式。

自然語言對話式交互得益於大資料和智慧技術的進步，多語言的自然語音辨識技術在使用者終端上都達到了很高的可用水準，並且，語音辨識超越文本輸入方式，成為智慧軟體助理的使能技術，近兩年，更是有基於語音介面的家居產品如雨後春筍般出現，VUI（Voice User Interface，語音使用者介面）已經成為交互術語。然而，VUI的局限也是顯而易見的，相對並行模式的視覺通道，串列模式的語音通道的頻寬顯然窄的多，出聲的使用方式在很多場合是不合適的，但作為一種可用的自然交互技術，有效提升了用戶體驗。

人機交互作為終端產品的引領技術的作用已經是產業界的普遍認識，欣喜看到很多種自然交互技術和新型交互終端面世，但 GUI 仍是交互的主導模式。計算無所不在，交互自然高效是發展趨勢，人機交互的研究和開發空間很大，需要綜合地探索自然交互技術的科學原理，建立明確的優化目標，結合智慧技術，發展高可用的自然交互技術。

AMiner

10 機器人

10.1 機器人概念

機器人廣義上包括一切模擬人類行為或思想以及模擬其他生物的機械(如機器狗, 機器貓等)。狹義上對機器人的定義還有很多分類法及爭議, 有些電腦程式甚至也被稱為機器人(例如爬蟲機器人)。聯合國標準化組織採納了美國機器人協會給機器人下的定義: “一種可程式設計和多功能的操作機; 或是為了執行不同的任務而具有可用電腦改變和可程式設計動作的專門系統。一般由執行機構、驅動裝置、檢測裝置和控制系統和複雜機械等組成”。機器人是綜合了機械、電子、電腦、感測器、控制技術、人工智慧、仿生學等多種學科的複雜智慧機械。

目前, 智慧型機器人已成為世界各國的研究熱點之一, 成為衡量一國工業化水準的重要標誌。機器人是自動執行工作的機器裝置, 因此, 它既可以接受人類指揮, 又可以運行預先編排的程式, 也可以根據以人工智慧技術制定的原則綱領行動。在當代工業中, 機器人指能自動執行任務的人造機器裝置, 用以取代或協助人類工作, 一般會是機電裝置, 由電腦程式或電子電路控制。機器人的範圍很廣, 可以是自主或是半自主的, 從本田技研工業的 ASIMO 或是 TOSY 的 TOPIO 等擬人機器人到工業機器人, 也包括多台一起動作的群機器人, 甚至是納米機器人。借由模仿逼真的外觀及自動化的動作, 理想中的高模擬機器人是高級整合控制論、機械電子、電腦與人工智慧、材料學和仿生學的產物。機器人可以作一些重複性高或是危險, 人類不願意從事的工作, 也可以做一些因為尺寸限制, 人類

無法作的工作，甚至是像外太空或是深海中，不適人類生存的環境。機器人在越來越多方面可以取代人類，或是在外貌、行為或認知，甚至情感上取代人類。

機器人技術最早應用於工業領域，但隨著機器人技術的發展和各行業需求的提升，在電腦技術、網路技術、MEMS 技術等新技術發展的推動下，近年來，機器人技術正從傳統的工業製造領域向醫療服務、教育娛樂、勘探勘測、生物工程、救災救援等領域迅速擴展，適應不同領域需求的機器人系統被深入研究和開發。過去幾十年，機器人技術的研究與應用，大大推動了人類的工業化和現代化進程，並逐步形成了機器人的產業鏈，使機器人的應用範圍也日趨廣泛^[63]。

10.2 機器人發展歷史

“機器人”一詞最早出現在 1920 年捷克斯洛伐克劇作家 Karel Capek 的科幻情節劇《羅薩姆的萬能機器人》中。

機器人從幻想世界真正走向現實世界是從自動化生產和科學研究的發展需要出發的。1939 年，紐約世博會上首次展出了由西屋電氣公司製造的家用機器人 Elektro，但它只是掌握了簡單的語言，能行走、抽煙，並不能代替人類做家務。

現代機器人的起源則始於二十世紀 40-50 年代，美國許多國家實驗室進行了機器人方面的初步探索。二次世界大戰期間，在放射性材料的生產和處理過程中應用了一種簡單的遙控操縱器，使得機械抓手就能複現人手的動作位置和姿態，代替了操作人員的直接操作。在這之後，橡樹嶺和阿爾貢國家實驗室開始研製遙控式機械手作為搬運放射性材料的工具。1948 年，主從式的遙控機械手正式誕生於此，開現代機器人製造之先河。美國麻省理工學院輻射實驗室(MIT Radiation

Laboratory) 1953 年研製成功數控銑床，把複雜伺服系統的技術與最新發展的數位電腦技術結合起來，切削模型以數位形式通過穿孔紙帶輸入機器，然後控制銑床的伺服軸按照模型的軌跡作切削動作。

上世紀 50 年代以後，機器人進入了實用化階段。1954 年，美國的 GeorgeC. Devol 設計並製作了世界上第一台機器人實驗裝置，發表了《適用於重複作業的通用性工業機器人》一文，並獲得了專利。GeorgeC. Devol 巧妙地把遙控操作器的關節型連杆機構與數控機床的伺服軸連接在一起，預定的機械手動作一經程式設計輸入後，機械手就可以離開人的輔助而獨立運行。這種機器人也可以接受示教而能完成各種簡單任務。示教過程中操作者用手帶動機械手依次通過工作任務的各個位置，這些位置序列記錄在數位記憶體內，任務執行過程中，機器人的各個關節在伺服驅動下再現出那些位置序列。因此，這種機器人的主要技術功能就是“可程式設計”以及“示教再現”。

上世紀 60 年代，機器人產品正式問世，機器人技術開始形成。1960 年，美國的 Consolidated Control 公司根據 GeorgeC. Devol 的專利研製出第一台機器人樣機，並成立 Unimation 公司，定型生產了 Unimate (意為“萬能自動”) 機器人。同時，美國“機床與鑄造公司” (AMF) 設計製造了另一種可程式設計的機器人 Versatran (意為“多才多藝”)。這兩種型號的機器人以“示教再現”的方式在汽車生產線上成功地代替工人進行傳送、焊接、噴漆等作業，它們在工作中表現出來的經濟效益、可靠性、靈活性，使其它發達工業國家為之傾倒。於是 Unimate 和 Versatran 作為商品開始在世界市場上銷售，日本、西歐也紛紛從美國引進機器人技術。這一時期，可實用機械的機器人被稱為工業機器人。

在機器人嶄露頭角于工業生產的同時，機器人技術研究不斷深入。1961年，美國麻省理工學院 Lincoln 實驗室把一個配有接觸感測器的遙控操縱器的從動部分與一台電腦連結在一起，這樣形成的機器人可以憑觸覺決定物體的狀態。隨後，用電視攝像頭作為輸入的電腦影像處理、物體辯識的研究工作也陸續取得成果。1968年，美國斯坦福人工智慧實驗室 (SAIL) 的 J. McCarthy 等人研究了新穎的課題——研製帶有手、眼、耳的電腦系統。於是，智慧型機器人的研究形象逐漸豐滿起來。

上世紀 70 年代以來，機器人產業蓬勃興起，機器人技術發展為專門的學科。1970 年，第一次國際工業機器人會議在美國舉行。工業機器人各種卓有成效的實用範例促成了機器人應用領域的進一步擴展；同時，又由於不同應用場合的特點，導致了各種坐標系統、各種結構的機器人相繼出現。而隨後的大型積體電路技術的飛躍發展及微型電腦的普遍應用，則使機器人的控制性能大幅度地得到提高、成本不斷降低。於是，導致了數百種類的不同結構、不同控制方法、不同用途的機器人終於在 80 年代以來真正進入了實用化的普及階段。進入 80 年代後，隨著電腦、感測器技術的發展，機器人技術已經具備了初步的感知、回饋能力，在工業生產中開始逐步應用。工業機器人首先在汽車製造業的流水線生產中開始大規模應用，隨後，諸如日本、德國、美國這樣的製造業發達國家開始在其他工業生產中也大量採用機器人作業。

上世紀 80 年代以後，機器人朝著越來越智慧的方向發展，這種機器人帶有多種感測器，能夠將多種感測器得到的資訊進行融合，能夠有效的適應變化的環境，具有很強的自我調整能力、學習能力和自治功能。智慧型機器人的發展主要經歷了三個階段，分別是可程式設計試教、再現型機器人，有感知能力和自我調

整能力的機器人，智慧型機器人。其中所涉及到的關鍵技術有多感測器資訊融合、導航與定位、路徑規劃、機器人視覺智慧控制和人機介面技術等。

進入 21 世紀，隨著勞動力成本的不斷提高、技術的不斷進步，各國陸續進行製造業的轉型與升級，出現了機器人替代人的熱潮。同時，人工智慧發展日新月異，服務機器人也開始走進普通家庭的生活。世界上許多機器人科技公司都在大力發展機器人技術，機器人的特質與有機生命越來越接近。

最近，波士頓動力公司在機器人領域的成就已經成為人們的焦點，其產品機器狗 Spot 和雙足人形機器人 Atlas 都讓人大為驚歎。Spot 的功能十分先進，可以前往你告訴它要去的目的地，避開障礙，並在極端情況下保持平衡。Spot 還可以背負多達四個硬體模組，為公司提供其他多款機器人完成特定工作所需的任何技能；Atlas 已經掌握了倒立、360 度翻轉、旋轉等多項技能，繼表演跑酷、後空翻等絕技之後，Atlas 又掌握了一項新技能—體操，再次讓我們大開眼界。



圖 10-1 波士頓動力機器人 Spot 與 Atlas

經過幾十年的發展，機器人技術終於形成了一門綜合性學科——機器人學 (Robotics)。一般地說，機器人學的研究目標是以智慧電腦為基礎的機器人的基本組織和操作，它包括基礎研究和應用研究兩方面內容，研究課題包括機械手

設計、機器人動力和控制、軌跡設計與規劃、感測器、機器人視覺、機器人控制語言、裝置與系統結構和機械智慧等。由於機器人學綜合了力學、機械學、電子學、生物學、控制論、電腦、人工智慧、系統工程等多種學科領域的知識，因此，也有人認為機器人學實際上是一個可分為若干學科的學科門類。同時，由於機器人是一門不斷發展的科學，對機器人的定義也隨著其發展而變化，目前國際上對於機器人的定義紛繁複雜，RIA、JIRA、NBS、ISO 等組織都有各自的定義，迄今為止，尚沒有一個統一的機器人定義。

10.3 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為機器人領域全球學者分佈情況：



圖 10-2 機器人領域全球學者分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量優勢明顯；歐洲也有較多的人才分佈；

亞洲的人才主要集中在我國東部及日韓地區；其他諸如非洲、南美洲等地區的學者非常稀少；機器人領域的人才分佈與各地區的科技、經濟實力情況大體一致。

此外，在性別比例方面，機器人領域中男性學者占比 90.7%，女性學者占比 9.3%，男性學者占比遠高於女性學者。

機器人領域學者的 h-index 分佈如下圖所示：

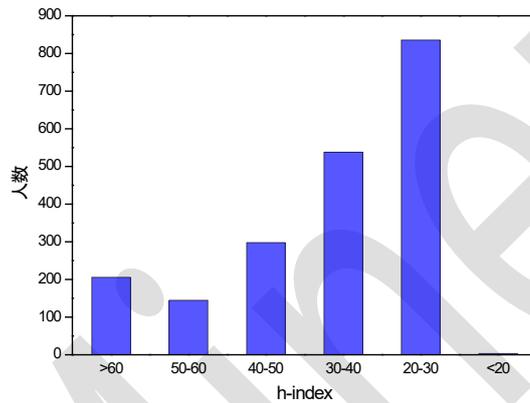


圖 10-3 機器人領域學者 h-index 分佈

從上圖可以看出，大部分學者的 h-index 分佈在中低區域，其中 h-index 在 20-30 區間的人數最多，有 836 人，占比 41.3%，小於 20 區間的人數最少，只有 3 人。

● 中國人才分佈

我國專家學者在機器人領域的分佈如下圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是珠三角和長三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，通過觀察中國周邊國家的學者數量情況，特別是與日韓等地相比，中國在機器人領域學者數量較少。

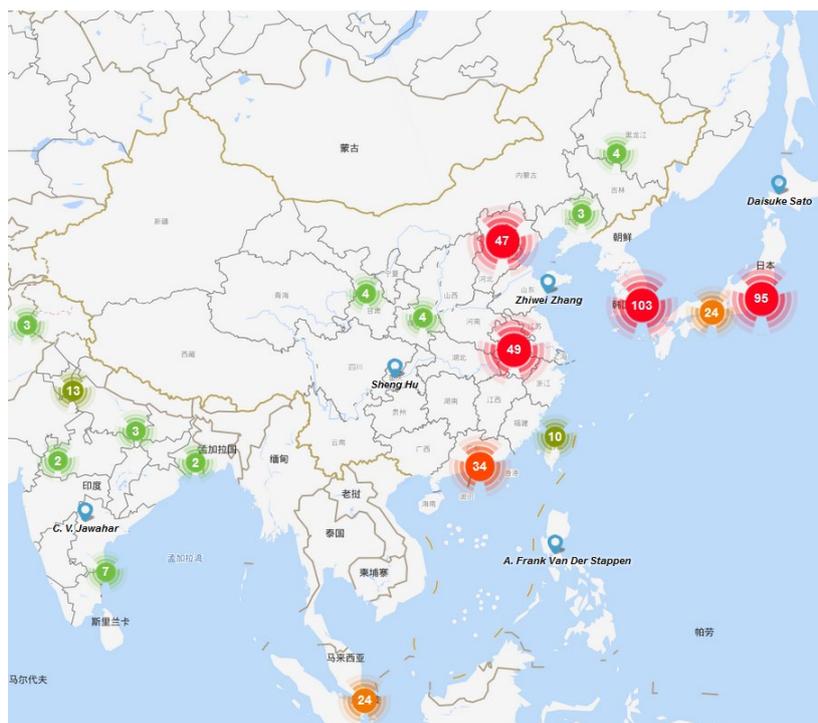


圖 10-4 機器人領域中國學者分佈

中國與其他國家在機器人領域的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 10-1 機器人領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	445	6606	15	756
中國-日本	90	998	11	196
中國-加拿大	69	761	11	128
中國-新加坡	62	586	9	123
中國-德國	57	780	14	98
中國-英國	50	1270	25	94

中國-法國	41	860	21	63
中國-澳大利亞	27	351	13	45
中國-瑞典	20	246	12	27
中國-義大利	18	318	18	33

從上表資料可以看出，中美合作的論文數、引用數、學者數遙遙領先，表明中美間在機器人領域合作之密切；此外，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 4 席；中國與英國合作的論文數雖然不是最多，但是擁有最高的平均引用數說明在合作品質上中英合作達到了較高的水準。

10.4 論文解讀

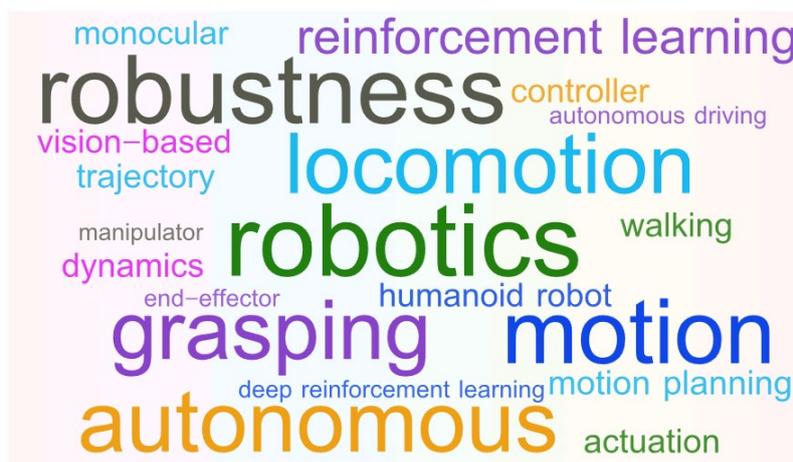
本節對本領域的高水準學術會議及期刊論文進行挖掘，解讀這些會議和期刊在 2018-2019 年的部分代表性工作。這些會議和期刊包括：

IEEE International Conference on Robotics and Automation

IEEE/RSJ International Conference on Intelligent Robots and Systems

Robotics: Science and Systems·A Robotics Conference

IEEE Transactions on Robotics



我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如上圖所示。其中，機器人(robotics) 魯棒性(robustness) 動作 (motion) 是本領域中最熱的關鍵字。

論文題目 : *Robotic Pick-and-Place of Novel Objects in Clutter with Multi-Affordance Grasping and Cross-Domain Image Matching*

中文題目：通過多 affordance 抓取和跨域圖像匹配完成雜亂環境下對新物體的檢放操作

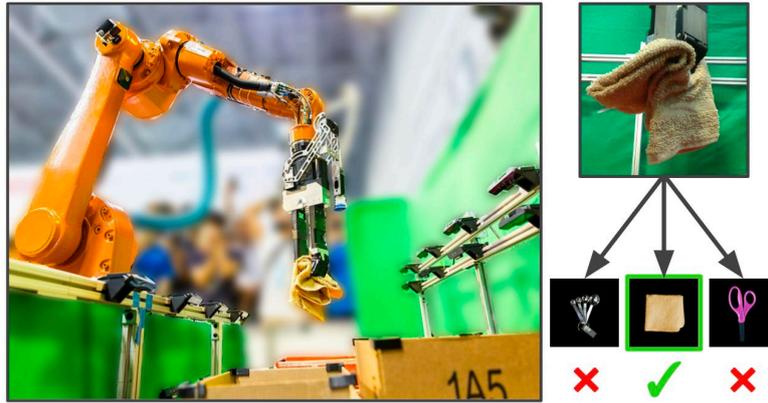
論文作者：Andy Zeng, Shuran Song, Kuan-Ting Yu, Elliott Donlon, Francois R. Hogan, Maria Bauza, Daolin Ma, Orion Taylor, Melody Liu, Eudald Romo, Nima Fazeli, Ferran Alet, Nikhil Chavan Dafle, Rachel Holladay, Isabella Morona, Prem Qu Nair, Druck Green, Ian Taylor, Weber Liu, Thomas Funkhouser, Alberto Rodriguez

論文出處：IEEE International Conference on Robotics and Automation, 2018

論文地址：<https://ieeexplore.ieee.org/abstract/document/8461044>

研究問題：

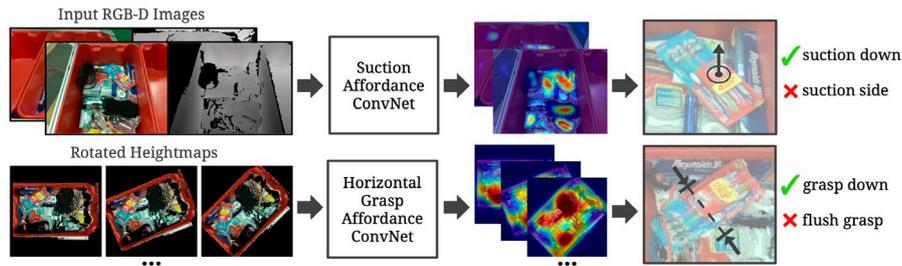
人類可以在僅掌握少量先驗知識的前提下識別和抓取陌生目標物，這一能力一直是機器人研究的靈感來源，也是很多實際應用的核心。為此，提出一種能在雜亂環境下對新目標物進行識別和檢放操作的機器人系統，整個系統可直接用於新目標物（在測試過程中首次出現），而無需額外的資料收集或重新訓練，如下圖所示。



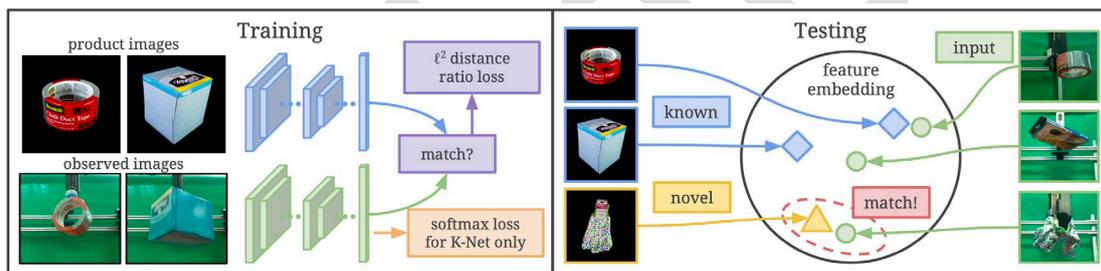
研究方法：

專門設計該機器人識別撿放系統，該系統由兩部分組成：1) 具有四個基礎行為的多模式抓取框架，該框架使用深度卷積神經網路 (ConvNets) 來預測場景 affordance，而無需事先對目標物進行分割和分類。2) 跨域圖像匹配框架，用於通過將抓取的物件與產品圖像進行匹配來識別抓取的物件，該框架使用了 ConvNet 架構，該架構可直接用於新目標物而無需重新進行訓練。這兩部分互相配合，可以在雜亂的環境中完成對新目標物的抓取操作。

機器人吸取和抓取的 affordance 預測如下圖所示，考慮各個視角的 RGBD 圖像，可通過一個全卷積殘差網路估算出每張圖片的吸取 affordance。然後，將預測匯總在 3D 點雲上，並基於表面法線生成向下吸取或側向吸取的建議。並行地，我們將 RGB-D 圖像合併為 RGB-D 高度圖，將其旋轉 16 個不同的角度，並估計每個高度圖的水準抓取。這有效地生成了針對 16 個不同抓取角度的 affordance 圖，從中可得到向下抓取和其他抓取的建議。



新物體的識別框架如下圖所示。訓練一個雙流的卷積神經網路，其中一個流計算得到產品圖像的 2048 維特徵向量，而另一個流計算得到觀察圖像的 2048 維特徵向量，並對兩個流進行優化，以使相同圖像的特徵更加相似，反之則不同。在測試期間，已知物件和新物件的圖像都映射到公共特徵空間上。通過在相同的特徵空間找到與其最近的特徵來匹配來識別它們。



研究結果：

提出一種系統，該系統能夠以很少的先驗資訊 (少數產品圖片) 來拾取和識別新對象。該系統首先使用與類別無關的 affordance 預測演算法在四種不同的抓取動作元之間進行選擇，然後通過將抓取的物件與它們的產品圖像進行匹配來識別抓取的對象。通過評估證明，該機器人系統可以拾取並在雜亂無章的環境中識別出新物體。

論文題目： *Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping*

中文題目：使用模擬和領域適應來提高深度機器人抓取的效率

論文作者：Konstantinos Bousmalis, Alex Irpan, Paul Wohlhart, Yunfei Bai, Matthew Kelcey, Mrinal Kalakrishnan, Laura Downs, Julian Ibarz, Peter Pastor, Kurt Konolige, Sergey Levine, Vincent Vanhoucke

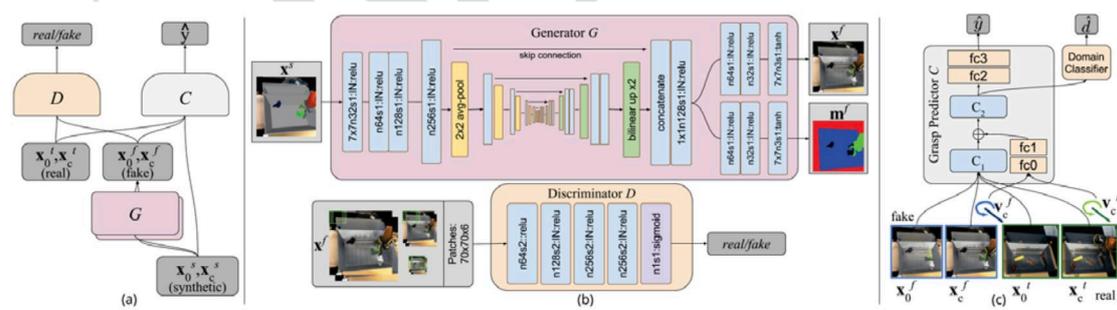
論文出處：IEEE International Conference on Robotics and Automation, 2018

論文地址：<https://ieeexplore.ieee.org/document/8460875>

研究問題：

收集帶注釋的視覺抓取資料集以訓練現代機器學習演算法可能是非常耗時的。一個替代方法是使用現成的模擬器來合成資料，這樣就可以自動產生這些資料的真實標注。不幸的是，僅基於類比資料訓練的模型通常無法泛化到現實世界。研究如何將隨機模擬環境和域適應方法應用到真實場景，訓練一種抓取系統，可以通過原始 RGB 圖像中進行規劃，抓取新的目標物。

研究方法：



研究類比環境中的 3D 目標模型、類比的場景和模擬的動力學如何影響機器人最終的抓取性能，以及將類比與真實場景集成以實現最大程度的遷移。具體方法如上圖所示。

(a) 為圖元級的域適應模型 GraspGAN 的概述。從模擬器中得到的圖像元組 x^s 並輸入到生成器 G 中，生成真實版本的圖像 x^f ，鑒別器 D 獲得未標注的真

實世界圖像 x^t 和 x^f ，並經過訓練以區分它們。真實的和經過適應的圖像也被送到抓取預測網路 C 中，一同進行並行訓練。因此，生成器 G 從鑒別器 D 和預測器 C 中獲得回饋，以使適應的圖像看起來更加真實並保持其語義資訊；(b) 為生成器 G 和鑒別器 D 的體系結構；(c) 為 DANN 模型，其中 C_1 包含 7 個卷積層， C_2 包含 9 個卷積層。

研究結果：

研究將類比資料合併到基於學習的抓取系統中的方法，以提高抓取性能並減少資料需求。通過使用合成資料和域適應，僅使用少量隨機生成的類比資料，就可以達到給定性指標的 50 倍。還表明，僅使用未標注的真實資料和 GraspGAN 的方法，就可以在沒有任何真實資料標注的情況下獲得與真實世界相同的抓取性能。

論文題目： *Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics*

中文題目：Dex-Net 2.0：利用合成點雲進行魯棒抓取和分析抓取指標的深度學習

論文作者：Bohg Jeffrey Mahler, Jacky Liang, Sherdil Niyaz, Michael Laskey, Richard Doan, Xinyu Liu, Juan Aparicio Ojea, and Ken Goldberg

論文出處：Robotics: Science and Systems, 2017

論文地址：<https://arxiv.org/pdf/1703.09312.pdf>

研究問題：

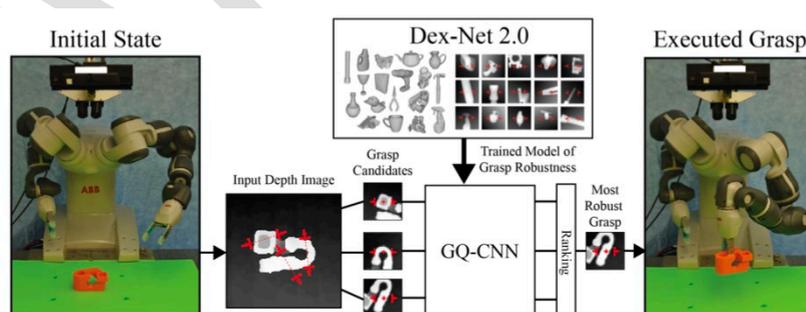
為了減少採用深度學習進行魯棒機器人抓取策略學習所需的資料收集時間，探索了從 670 萬點雲，抓取姿態和抓取指標的合成資料集進行訓練，這些資料是

從 Dex Net 1.0 的數千個三維模型中以隨機姿勢在桌子上生成的。利用得到的資料集 Dex-Net 2.0 訓練抓取品質卷積神經網路 (GQ-CNN) 模型，該模型可快速從深度圖像預測抓取成功的概率，其中抓取被指定為相對於 RGB-D 感測器的夾持器的平面位置、角度和深度。

研究方法：

研究基於深度點雲的處於桌面上的單剛體的平行爪抓取規劃問題。學習一個函數，它以一個候選抓取和一個深度圖像作為輸入，並輸出一個魯棒性的估計值，或者在傳感和控制的不確定性下的成功概率。

Dex Net 2.0 的架構如下圖所示。GQ-CNN 是抓取品質卷積神經網路，它是離線訓練的，使用由 670 萬個合成點雲、相關魯棒抓取指標的資料集 Dex-Net 1.0 資料集，可從深度圖像預測魯棒候選抓取。當一個物體呈現給機器人時，深度相機返回一個三維點雲，識別出一組幾百個候選抓取點。GQ-CNN 迅速確定了最穩健的抓取位姿，並由 ABB YuMi 機器人執行操作。



研究結果：

開發了一個抓取品質卷積神經網路 (GQ-CNN) 體系結構，它可以預測基於點雲模型抓取的穩定性，並在 Dex-2.0 資料集上對其進行訓練，它是一個包含 670 萬點雲、平行抓取和穩定性抓指標的資料集。在 1000 多個物理評估中，發現 Dex-

Net 2.0 抓取規劃器是一種可靠的、速度比基於點雲配准方法快 3 倍的，並且在 40 個新目標的測試集上具有 99%的精度的抓取規劃器。

論文題目： *Deep Predictive Policy Training using Reinforcement Learning*

中文題目：深度預測策略的強化學習訓練方法

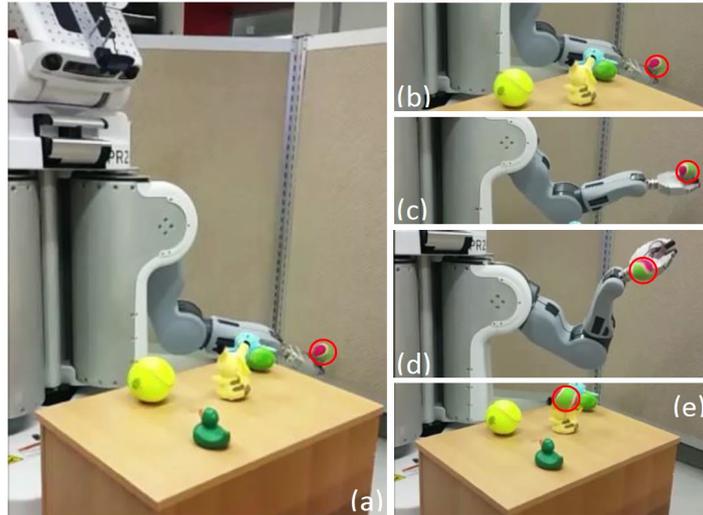
論文作者：Ali Ghadirzadeh, Atsuto Maki, Danica Kragic and Marten Bjorkman.

論文出處：Robotics: Science and Systems , 2019

論文地址：<https://arxiv.org/pdf/1903.11239.pdf>

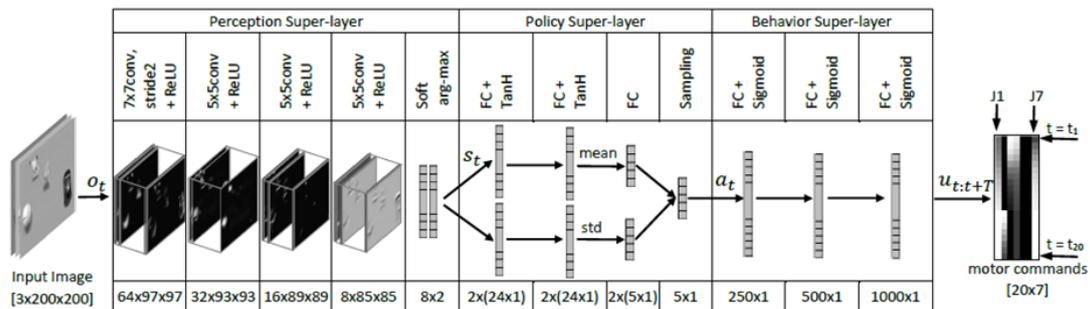
研究問題：

由於感知運動過程的固有延遲，機器人任務學習最好通過預測動作策略來實現。然而，訓練這樣的預測策略是具有挑戰性的，因為它涉及到在整個動作過程中找到運動啟動的軌跡。本文中，提出一個基於深度神經網路的資料高效深度預測策略訓練 (DPPT) 框架，將圖像觀測映射到一系列的運動啟動。該體系結構由三個子網路組成，分別稱為感知層、策略層和行為層。感知層和行為層迫使我們對視覺和行為進行抽象分別用合成訓練樣本和類比訓練樣本訓練資料。策略層是一個較小的子網路，具有較少的參數來映射抽象流形之間的資料。使用策略搜索強化學習的方法對每個任務進行訓練。通過在 PR2 機器人上訓練熟練抓取和投球的預測策略。下圖表示在機器人拋擲 ball 過程的瞬間圖。



研究方法：

由感知層、策略層和行為層組成的深度預測策略體系結構如下圖所示。作為網路輸入，給出了一個中心 RGB 圖像。感知層將圖像資料抽象為與任務相關的物件對應的若干空間位置。策略層將抽象狀態隨機映射到操作流形中的一個點。最後，針對給定的採樣動作，行為層生成一長軌跡的電機指令，並應用於機器人連續 T 個時間步長。



研究結果：

文章證明了所提出的結構和學習框架的適用性。該方法的有效性通過以下事實得到了證明:這些任務僅使用 180 次真正的機器人進行訓練，並提供定性的最終獎勵。

論文題目：*Learning Agile and Dynamic Motor Skills for Legged Robots*

中文題目：面向腿式機器人的敏捷動態特性的技能學習

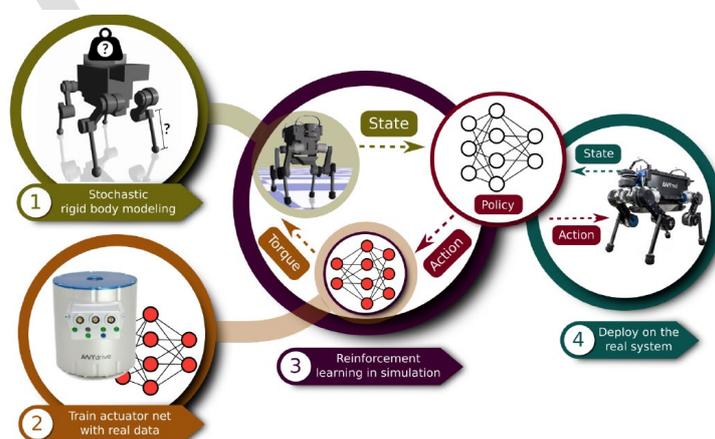
論文作者：Jemin Hwangbo, Joonho Lee, Alexey Dosovitskiy, Dario Bellicoso, Hoonho Lee, Vassilios Tsounis, Vladlen Koltun and Marco Hutter.

論文出處：Science Robotics, 2019

論文地址：<https://arxiv.org/pdf/1901.08652.pdf>

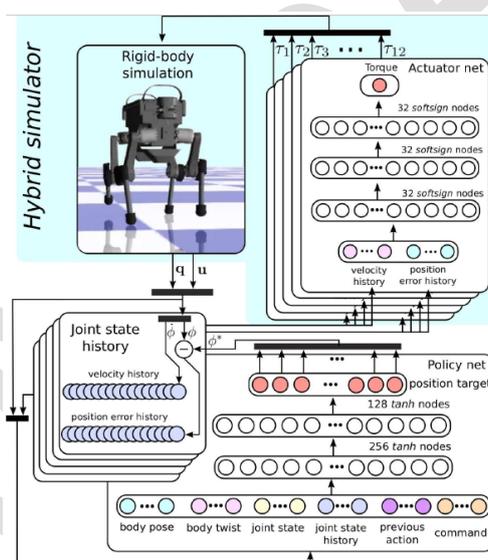
研究問題：

近年來，腿式機器人是機器人技術中最大的挑戰之一。動物的動態和敏捷的動作是現有的方法無法模仿的，而這些方法是由人類精心設計的。一個令人信服的替代方案是強化學習，它需要最少的技能並促使控制策略的自然演化更新。然而，到目前為止，對腿式機器人的強化學習研究主要局限於模擬，在實際系統中應用比較簡單的例子較少。主要原因是，使用真正的機器人進行訓練，尤其是使用動態平衡系統，既複雜又昂貴。在本論文中，我們提供了一種新的方法，在類比中訓練一個神經網路策略，並將其遷移到一個最先進的腿系統，因此我們利用快速、自動化和經濟有效的資料生成方案。



研究方法：

對於腿式機器人的敏捷動態性技能學習的過程，首先是系統建模，針對於四足機器人的物理參數的辨識以及確定機器人動態參數的不確定性指標，這個過程可能需要環境參數估計，物理動態性能估計等；其次是訓練驅動神經網路，這個過程一般通過構建機器人狀態到機器人電機控制的映射函數實現，隨著深度神經網路的廣泛認可，這樣的非線性映射函數現大多採用深度神經網路擬合；然後在模擬中完成基於強化學習的驅動神經網路的學習過程，最後將訓練好的驅動神經網路擬合的控制策略應用在實際的系統中。



整個系統的控制網路由三部分構成，首先是策略網路，用於將當前的觀測量和之前的關節狀態量映射到目標關節量（下一時刻關節控制量），然後是驅動網路，用於在剛體關節控制中將歷史關節狀態映射到關節力矩控制量上，機器人狀態量有各關節的位置資訊 q 與速度資訊 u 。

研究結果：

應用於一個複雜的中型犬大小的四足系統 ANYmal 機器人，使得在模擬中訓練的四足機器人的運動策略超越了以前的方法，ANYmal 能夠精確和高效地遵循高水準的身體速度指令，比以前跑得更快，甚至在複雜的配置中也能從跌倒中恢復過來。

論文題目： *Making Sense of Vision and Touch: Self-Supervised Learning of Multimodal Representations for Contact-Rich Tasks*

中文題目：理解視覺和觸覺：接觸任務多模態表達的自監督學習

論文作者：Michelle A. Lee, Yuke Zhu, Krishnan Srinivasan, Parth Shah, Silvio Savarese, Li Fei-Fei, Animesh Garg, and Jeannette Bohg

論文出處：IEEE International Conference on Robotics and Automation, 2019

論文地址：<https://ieeexplore.ieee.org/abstract/document/8793485>

研究問題：

非結構化環境中需要接觸的操作任務通常需要觸覺和視覺回饋。但是，人工設計融合各個不同模態的機器人控制器並非易事。儘管深度強化學習已經成功地應用於針對高維輸入的控制策略學習，但由於樣本複雜性，這些演算法通常難以部署在實際的機器人上。提出使用自監督來學習感官輸入的緊湊和多模態表示，以用來提高策略學習的樣本效率。

研究方法：

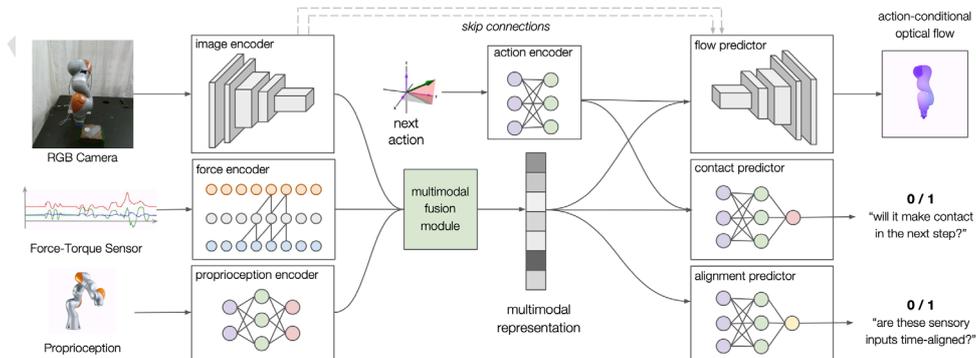
以學習機器人執行需要接觸操作任務的策略為目標。希望評估融合多感測器資訊的價值以及多模態表示在跨任務傳遞的能力。為了提高樣本效率，首先學習

了基於神經網路的多感測器資料特徵表示。得到的壓縮特徵向量用作通過強化學習學習到的策略的輸入。

將操作任務建模為一個有限時間的離散瑪律科夫決策過程 M ，狀態空間 S ，動作空間 A ，狀態轉移動力學 $T: S \times A \rightarrow S$ ，初始狀態分佈 ρ_0 ，回報函數 $r: S \times A \rightarrow R$ ，時間 T ，折扣係數 $\gamma \in (0, 1]$ ，為了確定最優隨機策略 $\pi: S \rightarrow P(A)$ ，我們希望最大化期望折扣獎勵

$$J(\pi) = \mathbb{E}_{\pi} \left[\sum_{t=0}^{T-1} \gamma^t r(s_t, \mathbf{a}_t) \right]$$

自監督多模態表示學習的神經網路結構如下圖所示，該網路將來自三個不同感測器的資料作為輸入：RGB 圖像、32ms 視窗上的 F/T 讀數以及末端執行器的位置和速度。它將這些資料編碼並融合到一個多模態表示中，基於此，可以學習包含接觸操作的控制器。這種表示學習網路是通過自監督形式進行端到端訓練的。



我們將具有接觸的操作作為一個無模型強化學習問題，研究它在依賴多模態回饋以及在幾何、間隙和構型不確定的情況下的性能。由於選擇無模型，還消除了對精確動力學模型的需要，這是存在接觸的操作中的典型困難。

研究結果：

提出了一種新穎的模型，將異構感官輸入編碼為多模態表示。一旦經過訓練，當用作用於強化學習的淺層神經網路策略的輸入時，該表示就保持固定。通過自我監督來訓練表示模型，從而無需手動標注。實驗表明，需要接觸的任務需要視覺和觸覺的多模式回饋，此外，還進一步證明了多模態表示可以很好地遷移到其他新任務中。

論文題目：A Magnetically-Actuated Untethered Jellyfish-Inspired Soft Milliswimmer

中文題目：一個受水母啟發的磁力驅動軟體游泳機器人

論文作者：Ziyu Ren , Tianlu Wang , Wenqi Hu , and Metin Sitti

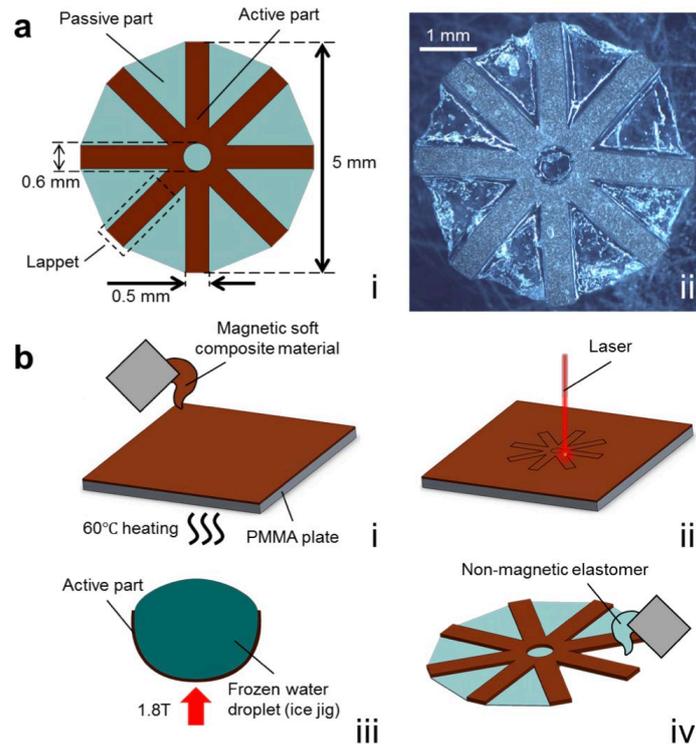
論文出處：Robotics: Science and Systems, 2019

論文地址：<http://www.roboticsproceedings.org/rss15/p13.pdf>

研究問題：

不受限制的小型軟機器人可以用於醫療和生物醫學應用。他們可以進入狹小空間並以可程式設計方式改變形狀，以適應非結構化環境並具有多種動態行為。但是，當前的微型軟機器人的功能有限，從而限制了它們在醫療中的應用。利用磁性軟複合材料形狀可程式設計的優勢，提出一種不受束縛的軟體機器人，它可以像水母一樣在時間和軌跡上不對稱地上下跳動，可以通過調節外部振盪磁場的大小，頻率和方向來控制其游泳速度和方向。

研究方法：



該機器人的設計如上圖所示，機器人主體由兩部分組成：主動部分像肌肉一樣工作以實現劃槳運動，而被動部分則填充了主動部分的間隙，使身體成為連續的流體動力表面。身體的主動部分由軟磁性材料製成，可在外部 B 場下變形。通過將釹鐵硼 (NdFeB) 磁性微粒 (MQP-15-7, Magnequench; 平均直徑: $5\mu\text{m}$) 與聚合物 (Ecoflex 00-10, Smooth-On Inc.) 混合來製備材料，品質比為 1:1。將該混合物澆鑄到塗覆有聚對二甲苯-C 的聚合物 (甲基丙烯酸甲酯) (PMMA) 板上。聚合物在 60°C 固化形成厚度約為 $96\mu\text{m}$ 的薄膜 (下圖 b-i)。使用鐳射切割機從該薄膜上切出主動部分的幾何形狀 (下圖 b-ii)。從平板上移開主動部分後，使用移液器將水滴滴在其上。活性部分可以立即包裹水滴並在表面張力作用下形成橢圓形 (下圖 b-iii)。然後將有效成分放入冰箱進行冷凍，以保持橢圓形的形狀。橢圓形主動部分在振動樣品磁力計 (VSM, EZ7, Microsense) 內部被 1.8T 均勻磁場磁化。磁化後，用非磁性彈性體 (Ecoflex 00-10) 填充主動部分的間隙以形成厚度約為 $20\mu\text{m}$ 的薄層薄膜 (下圖 b-iv)。最終機器人如下圖 a-ii 所示。

研究結果：

提出了一種使用磁性軟複合材料製作的軟體游泳機器人。只需調節外部磁場的波形，頻率和振盪方向即可實現對其控制。已經進行了初步研究以發現其推進速度與輸入控制信號之間的關係。當驅動頻率增加時，由於流體動力阻尼力，機器人的跳動幅度單調減小。實驗資料和模型預測都顯示了對於特定控制波形的最佳驅動頻率的的存在。實驗表明，該機器人可用于多種潛在醫療功能。

論文題目：*Robust Visual-Inertial State Estimation with Multiple Odometries and Efficient Mapping on an MAV with Ultra-Wide FOV Stereo Vision*

中文題目：魯棒多測度視覺慣性狀態估計及其在具有超廣角立體視覺的微型飛行器上的高效映射

論文作者：M. G. Müller, F. Steidle, M. J. Schuster, P. Lutz, M. Maier, S. Stoneman, T. Tomic, and W. Sturzl

論文出處：IEEE International Conference on Intelligent Robots and Systems , 2018

論文地址：<https://ieeexplore.ieee.org/document/8594117>

研究問題：

近年來，微型飛行器 (MAV) 已用於各種各樣的應用中。他們能夠快速到達感興趣的點或獲得以前難以或不可能到達的視角，這使它們對於諸如勘探，檢查，搜索和救援之類的任務變得非常有用。提出了一種配備兩對廣角立體相機和一個慣性測量單元 (IMU) 的多旋翼系統，以實現強大的視覺慣性導航和省時的全向 3D 映射，如下圖所示。



研究方法：

四個攝像頭垂直覆蓋了 240 度的立體視野 (FOV) ，這使得該系統也適用於狹窄和狹窄的環境，例如洞穴。在所提出的方法中，從四個廣角攝像頭合成了八個虛擬針孔攝像頭。所得的四個合成針孔立體系統中的每一個都為獨立的視覺測距法 (VO) 提供輸入。隨後，基於它們與狀態估計的一致性，將四個單獨的運動估計與來自 IMU 的資料融合。

研究結果：

提出了配備有四個廣角攝像機的 MAV。多達 240° 的垂直立體視野使 MAV 能夠感知其下方，上方和前方的物件，這與避障，路徑規劃和有效的映射等任務有關。實驗表明，由四個具有獨立關鍵幀的立體測距儀提供的魯棒運動估計，也可以從較大的視野中受益，從而可以進行狀態估計。

10.5 機器人進展

- 機器人學習

在 AI 興起的時代，機器人擁有了一種新型的學習方式：深度強化學習。這一新方式借助通用化的神經網路表示，處理複雜的感測器輸入，來讓機器人從自

己的經驗活動中直接學習行為。相比傳統方式，它解放了工程設計人員們的雙手，不再需要程式師們手動設計機器人每一個動作的每一項精確參數。但是，現有的強化學習演算法都還不能夠適用於有複雜系統的機器人，不足以支撐機器人在短時間內就學習到行為，另外在安全性上也難以保障。針對這種困境，2019 年初，穀歌 AI 與 UC 伯克利大學合作研發了一種新的強化學習演算法：SAC (Soft Actor-Critic)。SAC 非常適應真實世界中的機器人技能學習，可以在幾個小時內學會解決真實世界的機器人問題，而且它的一套超參數能夠在多種不同的環境中工作，效率十分之高。SAC 的開發基於最大熵強化學習這個框架。此框架嘗試讓預期回報最大化，同時讓策略的熵最大化。一般而言，熵更高的策略具有更高的隨機性。從直覺上看，這意味著，最大熵強化學習能取得高回報策略中具有最高隨機性的那個策略。SAC 學習一個隨機策略，這個策略會把狀態映射到動作，也映射到一個能夠估計當前策略目標價值的 Q 函數，這個 Q 函數還能通過逼近動態程式設計來優化它們。SAC 通過這樣的方式，來讓經過熵強化的回報最大化。此過程中，目標會被看作一個絕對真的方法，來匯出更好的強化學習演算法，它們有足夠高的樣本效率，且表現穩定，完全可以應用到真實世界的機器人學習中去。

- 機器人應用

2019 年 6 月，亞馬遜在 MARS 人工智慧大會上最新發佈的倉庫機器人 Pegasus，該機器人已正式加入亞馬遜 Kiva 機器人行列。Pegasus 是一種新型包裹分揀機器人，外觀上看，Pegasus 機器人十分類似亞馬遜既有的 Kiva 機器人，外觀還是橙色不變，2 英尺高，3 英尺寬，約相當於一個手提包的大小。Pegasus 機器人更像是對原有 Kiva 機器人的改良版，在原有機器人底座上增加了一個載貨平臺+皮帶傳送帶對各個包裹進行分類和移動，有助於最大限度地減少包裹損壞

並縮短交貨時間。Pegasus 機器人可以自主將右側盒子放在正確的位置。倉庫作業人員將包裹掃描完放到 Pegasus 機器人上，Pegasus 機器人載著包裹到指定地點。機器人配備的攝像機可以感知任何意外障礙。到了指定地點，機器人載貨平臺上的傳送帶將包裝從機器人上移開，然後包裹沿著滑槽向下移動，準備送出。機器人在大約 2 分鐘內完成整個包裹運送過程。據亞馬遜介紹，Pegasus 機器人具有與 Kiva 機器人驅動器相同的容量。Pegasus 機器人目前已經在丹佛分揀中心上線的六個多月，行駛約 200 萬英里，經測試，它能將當前系統的包裹分揀錯誤率大幅降低 50%。本次 MARS 人工智慧大會上，除了推出 Pegasus 機器人，亞馬遜還發佈了一種大型模組化運輸機器人 Xanthus。依據上方安裝的模組，執行多種不同的任務 Xanthus 擁有透過改變上方配備，勝任不同任務的能力。相較過去使用的系統，Xanthus 不僅用途更為廣泛，體積也只有前輩的 1/3，成本甚至直接砍半。

- 機器人平臺

如何將機器人技術落地、實踐商業化一直是備受關注的問題。波士頓動力的策略是要希望其成為平臺公司，通過授權或開源方式，使其技術能被廣為被使用。2018 年這個傳言似乎得到了證實，在《連線》雜誌舉辦的峰會上，波士頓動力創始人暨首席執行官 Marc Raibert 指出，他們的定位是成為平臺公司，讓生態圈包括協力廠商夥伴、客戶，一起來找到技術真正適合使用的地方。Marc Raibert 表示波士頓動力在開發機器人時是以“平臺”的概念來出發，客戶可以增加硬體，例如手臂及其他元件，“當然，我們也可以針對單一領域打造一個有特殊應用的機器人方案，但我們不知道哪一個領域合適，所以我們從平臺的角度出發，希望生態圈幫我們一起來找到技術真正可落地之處”、“我們要打造的是‘通用用途的平臺’

(general purpose platform) , 讓協力廠商夥伴、客戶、波士頓動力自己的應用開發團隊，可以一同來設計產品以符合定制化需求。

AMiner

11 資料庫技術

11.1 資料庫概念

資料庫是按一定的結構和規則組織起來的相關資料的集合，是綜合各使用者資料形成的資料集合，是存放資料的倉庫(我國資料庫的發展現狀與趨勢--陳黎)。隨著電腦技術與網路通信技術的快速發展，資料庫技術已經成為當今資訊社會中對大量資料進行組織與管理的重要技術手段，是網路資訊化管理系統的基礎。目前，新一代資料庫系統不僅保持和繼承了傳統資料庫系統的各項功能，支援知識管理、資料管理和物件管理，而且還對其它應用系統開放，在網路上支援標準網路通訊協定，具有良好的可連線性、可攜性、可互通性和可擴展性。

資料庫技術與網路通信技術、人工智慧技術、物件導向程式設計技術、平行計算技術等互相滲透和結合，是當前資料庫技術應用的主要特徵，當前具有此類特徵的新型資料庫系統包括如分散式資料庫系統、知識庫系統和主動資料庫系統等^[64]。

- 分散式資料庫系統

分散式資料庫系統由一組分佈在網路中的不同電腦上的資料組成。系統中每台伺服器有自己的資料庫系統及若干台客戶機，3 台伺服器之間通過網路相連。網路中的每個節點具有獨立處理的能力，可以執行局部應用，同時每個節點也能通過網路子系統執行全域應用。使用者通過客戶機可以對本機伺服器中的資料庫執行某些應用，也可以對 2 個或 2 個以上節點中的資料庫執行某些應用。

- 知識庫系統

知識庫系統是資料庫和人工智慧兩種技術相結合的產物,簡單來說就是在資料庫技術中引入人工智慧技術,把資料庫看作一個人工智慧系統,利用人工智慧技術來提高 DBMS 的表達、推理和查詢能力。其功能主要體現在資料庫系統推理能力的擴充、語義知識的引入、知識的獲取、知識和資料的有效組織及管理等方面；而效率則體現在資料庫對使用者查詢的快速回應和查詢優化上。

- 主動資料庫

在實際應用中，如電腦集成製造系統、執行資訊系統、辦公自動化系統等通常希望資料庫系統在緊急情況下能根據資料庫的當前狀態主動做出反應，並執行相應的操作，向使用者提供特定資訊。主動資料庫是相對於傳統資料庫的被動性而言的，其主要任務是提供對緊急情況的及時反應能力，同時提高資料庫管理系統的模組化程度。主動資料庫通常採用的方法是在傳統資料庫系統中嵌入事件、條件、動作規則,在某一事件發生時引發資料庫管理系統去檢測資料庫的當前狀態；判斷是否滿足設定的條件,如果條件滿足便觸發規定動作執行。

11.2 資料庫技術歷史

在處理器、電腦記憶體、電腦存儲和電腦網路等領域的技術進步之後，資料庫的大小、性能和性能以及它們各自的 DBMS 都以數量級增長。資料庫技術的發展可以根據資料模型或結構劃分為三個時代：導航資料庫時代、SQL/關係時代、後關係時代。其中後關係包含了物件導向型資料庫及新型資料庫技術的發展。下面按照時間順序介紹每個時期的主流資料庫技術^[65]。

資料庫的起源

60 多年前，資料以一種原始與粗糙的方式進行管理。通過大量的分類、比較和表格繪製的機器運行數百萬穿孔卡片來進行資料的處理，其運行結果在紙上列印出來或者製成新的穿孔卡片。而資料管理就是對所有這些穿孔卡片進行物理的儲存和處理。但是這樣並不能滿足對日益增長資料管理控制的需求。資料以某種抽象的方式組織起來，使存儲和檢索更加有效，這是不可避免的。

20 世紀 60 年代，基於導航的資料庫管理系統

20 世紀 60 年代電腦開始廣泛地應用於資料管理，但對資料的共用提出了越來越高的要求。傳統的檔案系統已經不能滿足人們的需要。能夠統一管理和共用資料的資料庫管理系統 (DBMS) 應運而生。資料模型是資料庫系統的核心和基礎，各種 DBMS 軟體都是基於某種資料模型的。所以通常也按照資料模型的特點將傳統資料庫系統分成網狀資料庫 (Network database)、層次資料庫 (Hierarchical database) 和關聯式資料庫 (Relational database) 三類。

20 世紀 70 年代初，關聯式資料庫的發展

1970 年，Codd 撰寫了大量論文，概述了資料庫建設的新方法，最終實現了大型共用資料庫的開創性關聯資料模型。Codd 描述了一個用於存儲和處理大型資料庫的新系統。關係模型通過將資料分解為一系列規範化表，可選元素從主表移出到只有在需要時才佔用空間的位置。可以在這些表中自由插入，刪除和編輯資料，DBMS 可以進行任何維護，以向應用程式/使用者呈現表視圖。關係模型還允許資料庫的內容發展，而不必不斷重寫連結和指標。關係部分來自引用其他實體的實體，稱為一對多關聯性，如傳統的層次模型，以及像導航 (網路) 模型這樣的多對多關係。因此，關係模型既可以表達層次結構模型，也可以表示導航

模型，也可以表示其原生表格模型，從而根據應用程式的需要對這三種模型進行純粹的或組合的建模。

20 世紀 70 年代後期，SQL 的發展

在 20 世紀 70 年代早期，IBM 開始基於 Codd 的概念鬆散地開發原型系統 SystemR。第一個版本於 1974 年 5 月準備就緒，然後在多表系統上開始工作，在該系統中可以拆分數據，以便記錄的所有資料不必存儲在單個大塊“塊”。隨後的多用戶版本在 1978 年和 1979 年由客戶進行了測試，此時添加了一種標準化的查詢語言——SQL (Structured Query Language)。結構化查詢語言是高級的非過程化程式設計語言，允許使用者在高層資料結構上工作。它不要求使用者指定對資料的存放方法，也不需要使用者瞭解具體的資料存放方式，所以具有完全不同底層結構的不同資料庫系統，可以使用相同的結構化查詢語言作為資料登錄與管理的介面。結構化查詢語言語句可以嵌套，這使它具有極大的靈活性和強大的功能。

20 世紀 90 年代，物件導向資料庫

資料庫面臨的首要問題之一就是非字元 (字母數位) 資料的崛起。越來越多的圖像、音效檔、地圖和視頻需要存儲、操作和檢索。這導致了大量的創新，但也支離破碎的標準。這時物件導向資料庫應運而生。物件導向資料庫系統(OODBS) 支援定義和操作 OODB，應滿足兩個標準：首先它是資料庫系統，其次它也是物件導向系統。第一個標準即作為資料庫系統應具備的能力 (持久性、事務管理、併發控制、恢復、查詢、版本管理、完整性、安全性)。第二個標準就是要求物件導向資料庫充分支援完整的物件導向 (OO) 概念和控制機制。綜上所述，我們將物件導向資料庫簡寫為：物件導向資料庫=物件導向系統+資料庫能力。

21 世紀初，NoSQL 與 NewSQL 技術的發展

NoSQL，泛指非關聯式的資料庫。NoSQL 資料庫的產生是為了解決大規模資料集合多重資料種類帶來的挑戰，尤其是大資料應用難題。NoSQL 早期就有人提出，發展至 2009 年趨勢越發高漲。NoSQL 的擁護者們提倡運用非關聯式的資料存儲，相對於鋪天蓋地的關係型數據庫運用，這一概念無疑是一種全新的思維的注入。NewSQL 一詞是由 451Group 的分析師 Matthew 在研究論文中提出的，是對各種新的可擴展/高性能資料庫的簡稱，這類資料庫不僅具有 NoSQL 對海量資料的存儲管理能力，還保持了傳統資料庫支援 ACID 和 SQL 等特性。

圖資料庫也是非關聯式的資料庫的一種，是一個使用圖結構進行語義查詢的資料庫，它使用節點、邊和屬性來表示和存儲資料。該系統的關鍵概念是圖，它直接將存儲中的資料項目，與資料節點和節點間表示關係的邊的集合相關聯。這些關係允許直接將存儲區中的資料鏈接在一起，並且在許多情況下，可以通過一個操作進行檢索。圖資料庫將資料之間的關係作為優先順序。查詢圖資料庫中的關係很快，因為它們永久存儲在資料庫本身中。可以使用圖資料庫直觀地顯示關係，使其對於高度互連的資料非常有用。

NewSQL 屬於分散式資料庫，分散式資料庫是指利用高速電腦網路，將物理上分散的多個資料存儲單元，連接起來組成一個邏輯上統一的資料庫。分散式資料庫的基本思想是，將原來集中式資料庫中的資料，分散存儲到多個通過網路連接的資料存儲節點上，以獲取更大的存儲容量和更高的併發訪問量。

11.3 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為資料庫領域全球學者分佈情況：



圖 11-1 資料庫領域全球學者分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量優勢明顯且主要分佈在其東西海岸；歐洲也有較多的人才分佈；亞洲的人才主要集中在我國東部；其他諸如非洲、南美洲等地區的學者非常稀少；資料庫領域的人才分佈與各地區的科技、經濟實力情況大體一致。

此外，在性別比例方面，資料庫領域中男性學者占比 91.7%，女性學者占比 8.3%，男性學者占比遠高於女性學者。

資料庫領域學者的 h-index 分佈下圖所示，分佈情況大體呈階梯狀，其中 h-index 小於 20 區間的人數最多，有 921 人，占比 45.4%，50-60 區間的人數最少，有 87 人。

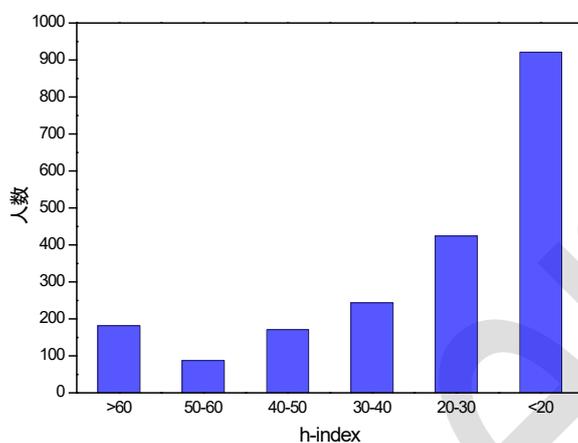


圖 11-2 資料庫領域學者 h-index 分佈

- 中國人才分佈

我國專家學者在資料庫領域的分佈如下圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是珠三角和長三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，通

過觀察中國周邊國家的學者數量情況，特別是與日韓、東南亞等地相比，中國在資料庫領域學者數量較多但差距不大。



圖 11-3 資料庫領域中國學者分佈

中國與其他國家在資料庫領域的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 11-1 資料庫領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	259	9679	37	584
中國-新加坡	89	3968	45	139
中國-澳大利亞	47	1565	33	81

中國-丹麥	25	318	13	45
中國-加拿大	17	290	17	41
中國-法國	15	181	12	14
中國-希臘	11	911	83	17
中國-卡達	10	206	21	17
中國-韓國	7	227	32	16
中國-瑞士	6	66	11	21

從上表資料可以看出，中美合作的論文數、引用數、學者數遙遙領先，表明中美間在資料庫領域合作之密切；此外，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 4 席；中國與希臘合作的論文數雖然不是最多，但是擁有最高的平均引用數說明在合作品質上中希合作達到了較高的水準。

11.4 論文解讀

本節對本領域的高水準學術會議及期刊論文進行挖掘，解讀這些會議和期刊在 2018-2019 年的部分代表性工作。這些會議和期刊包括：

ACM SIGMOD International Conference on Management of Data

International Conference on Very Large Data Bases

我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如下圖所示。其中，大資料 (big data)、資料庫系統 (database systems)、解析法 (analytics) 是本領域中最熱的關鍵字。



論文題目： Self-Driving Database Management Systems

中文題目：自動駕駛的資料庫管理系統

論文作者：Andrew Pavlo, Gustavo Angulo, Joy Arulraj and, Haibin Lin, Jiexi Lin, Lin Ma, et al.

論文出處：7th Biennial Conference on Innovative Data Systems Research (CIDR) – CIDR 2017

論文地址：<https://www.pdl.cmu.edu/PDL-FTP/Database/p42-pavlo-cidr17.pdf>

研究問題：

在過去的二十年中，研究人員和資料庫系統供應商都嘗試開發了各式輔助工具以在資料庫系統的調優和物理設計等各個方面協助資料庫管理員（Database Administrator，DBA）。但是，大多數的工作還是不足夠完善的，因為它們仍然需要 DBA 對資料庫的任何更改做出最終決定，並且是在問題發生後解決問題的反應性措施。尤其是隨著雲資料庫的發展，不需要人工干預的 DBMS 就成為了一個迫切的需求，於是能“自動駕駛”的資料庫管理系統（Database Management System，DBMS）便成為了必然的選擇。真正地能“自動駕駛”的資料庫管理系統

所需要的是一種為自治操作而設計的新體系結構。與早期的各種 DBMS 不同的是，該類系統的所有方面都由集成的計畫元件控制，該元件不僅可以針對當前工作負載 (Workload) 優化系統，而且還能預測未來的工作負載的變化趨勢，以便系統可以相應地進行準備。這樣，DBMS 可以支援所有以前的調優技術，而無需人工確定正確的方式和適當的時間來部署它們。

研究方法：

該論文指出“自動駕駛” (Self-Driving) 的 DBMS 會面臨以下三點問題：

1) 需要理解業務應用的工作負載 (Workload)：簡單的來說，我們可以將應用分為連線交易處理過程 (Online Transaction Processing , OLTP)、連線分析處理過程 (Online Analytical Processing , OLAP) 和混合交易-分析處理過程 (Hybrid Transaction-Analytical Processing , HTAP) 三種類型。不同類型的應用需要不同的設置，OLTP 以交易型場景為主，資料採用行存儲，便於快速的對一條記錄進行增刪改查；HTAP 則是混合型場景，不僅有頻繁的條目增刪改查，也有大量的單列分析應用。

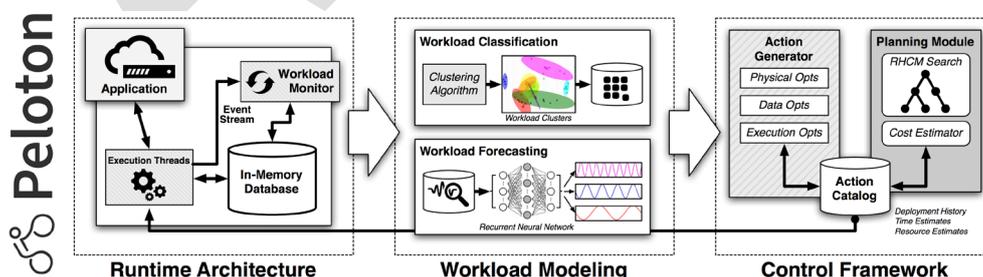
2) 需要能夠預測資源的使用趨勢：有助於系統根據未來可能的情況和資源需求進行動態資源調配以確保對性能的影響保持最低。由於很多應用的使用模式與人的活動模式密切相關，資料庫管理員通常會在業務低峰期進行一些優化操作，避免影響應用的服務品質。不可否認，現實中難免會存在一些工作負載的異常狀況是 DBMS 無法避免的。但是，這些預測模型是可以儘早地提供一些預警使得 DBMS 盡可能完成相應操作。有了預測模型之後，DBMS 可以通過一些調優操作使資料庫在預期工作負載上工作的更好。自動駕駛的 DBMS 可以支援的操作主

要有如下幾種：(1) 資料庫的物理設計 (Database's Physical Design) ； (2) 數據組織形式的變更 (Data Organization) ； (3) DBMS 運行時的行為 (DBMS's Runtime Behavior) 。這三類操作的詳細內容如下表所示，這裡不一一贅述。

	Types	Actions
PHYSICAL	Indexes	AddIndex, DropIndex, Rebuild, Convert
	Materialized Views	AddMatView, DropMatView
	Storage Layout	Row→Columnar, Columnar→Row, Compress
DATA	Location	MoveUpTier, MoveDownTier, Migrate
	Partitioning	RepartitionTable, ReplicateTable
RUNTIME	Resources	AddNode, RemoveNode
	Configuration Tuning	IncrementKnob, DecrementKnob, SetKnob
	Query Optimizations	CostModelTune, Compilation, Prefetch

3) 要一個靈活的，記憶體級的架構，便於快速應用優化操作：如果 DBMS 不夠靈活，那麼模型推薦的一些優化措施不能及時地實施，也就失去了優化的意義。同時，也不能算得上是真正意義的“自動駕駛”資料庫。

現有的 DBMS 對於自動化操作的支持都不太友好，經常需要通過重啟來使得配置生效。針對這些，Peloton 採用了多版本併發控制的模式，可以在不阻塞 OLAP 的情況下，提供 OLTP 的服務，並採用了無鎖資料結構的記憶體存儲管理器，能支援快速地執行 HTAP 操作。

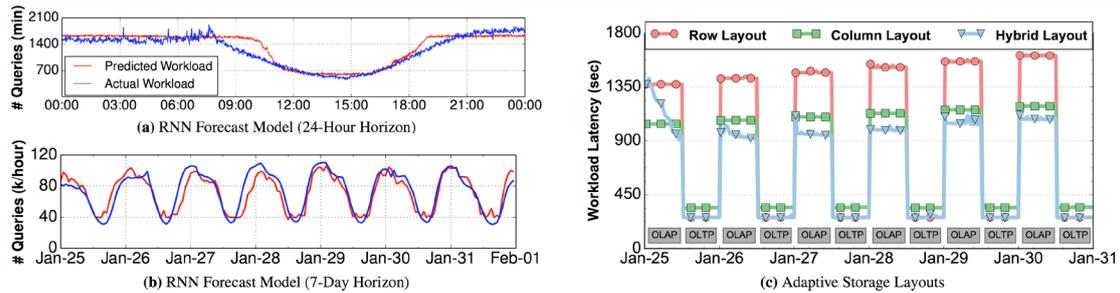


上圖是 Peloton 系統的“自動駕駛”工作流程的概述。除了環境設置 (例如，記憶體閾值，目錄路徑) 以外，Peloton 的目標在沒有任何人工提供指導資訊的情況下能有效的運行。系統會自動學習如何改善應用程式查詢和事務的延遲

(Latency)。由於延遲是反映 DBMS 性能的最重要指標，Peloton 系統主要將延遲作為優化的指標。除此之外，該系統還可以優化在分散式環境中其它的重要指標，例如服務成本和能源。Peloton 還包含一個嵌入式監視器，該監視器遵循已執行查詢的系統內部事件流。在 Workload 建模部分，該系統首先對工作負載進行聚類，這樣子可以減少 DBMS 管理的預測模型的數量，降低了預測應用行為的複雜度，同時採用了 DBSCAN 聚類演算法，使得效果進一步提升。在工作負載預測部分，該系統可以預測週期性的工作負載和資料變化趨勢，能更好的提供服務性能。例如，每當 DBMS 執行完一個查詢操作之後，對系統該查詢的聚類中心進行表示，並按照預定義的統計區間去記錄這些查詢的請求次數。基於此，系統可以使用這些資料來預測“未來”的查詢數量，並可能提前執行相關的優化。通過系統的控制模組，可以提供對系統的持續監控並選擇合適的優化措施去提升應用的服務性能。

研究結果：

Peloton 系統首先對工作負載的不同細微性和時間跨度進行預測。如下圖(a) 和 (b) 所示，Peloton 系統對於短期內的查詢預測是比較準確的，但是長期的預測效果不佳。從下圖(c)中可以看出，該實驗主要測試了 OLAP、OLTA 和 HTAP 場景下的優化效果。可以看出，系統可以看出 Peloton 會隨時間的推移而收斂。例如，在第一個 OLAP 操作之後，系統會將存儲的元組改為列存儲的佈局，這是 OLTP 查詢的理想選擇，因此其延遲下降。本篇論文開創性地提出了 Self-Driving DBMS 的概念，並提出了一個初步的系統驗證了該概念的可行性。



論文題目：*Neo: A Learned Query Optimizer*

中文題目：Neo：學習型查詢最佳化工具

論文作者：Ryan Marcus, Parimarjan Negi, Hongzi Mao, Chi Zhang, Mohammad Alizadeh, Tim Kraska, Olga Papaemmanouil, Nesime Tatbul

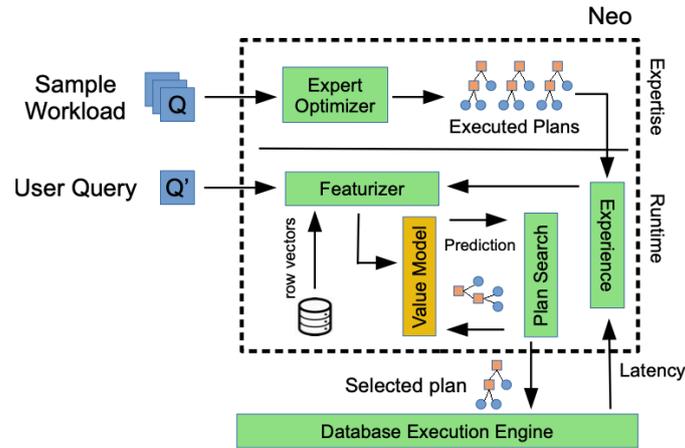
論文出處：45th International Conference on Very Large Data Bases – VLDB 2019

論文地址：<http://www.vldb.org/pvldb/vol12/p1705-marcus.pdf>

研究問題：

查詢優化是資料庫系統中最具挑戰性的問題之一。儘管在過去幾十年中取得了進步，但是查詢最佳化工具仍然是極其複雜的元件，需要針對特定工作負載和資料集進行大量手動調整。受此缺點的激勵，並受到將機器學習應用於資料管理挑戰的最新進展的啟發，我們引入了 Neo (神經優化器)，這是一種新型的基於學習的查詢最佳化工具，它依賴於深度神經網路來生成查詢執行計畫。Neo 從現有的優化器中引導其查詢優化模型，並繼續從傳入的查詢中學習，以成功為基礎，從失敗中學習。此外，Neo 會自然地適應基礎資料模式，並且對估計錯誤具有魯棒性。實驗結果表明，即使從 PostgreSQL 之類的簡單優化器啟動後，Neo 仍可以學習一種模型，該模型可提供與最新的商業優化器相似的性能，甚至在某些情況下甚至可以超越它們。

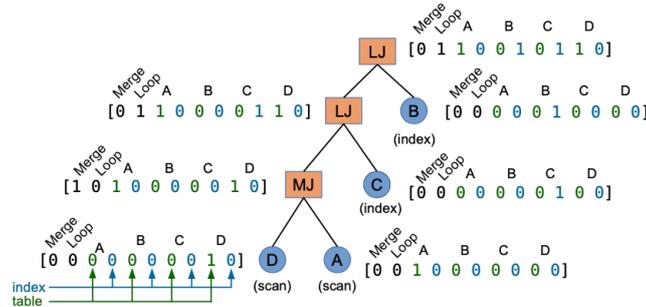
研究方法：



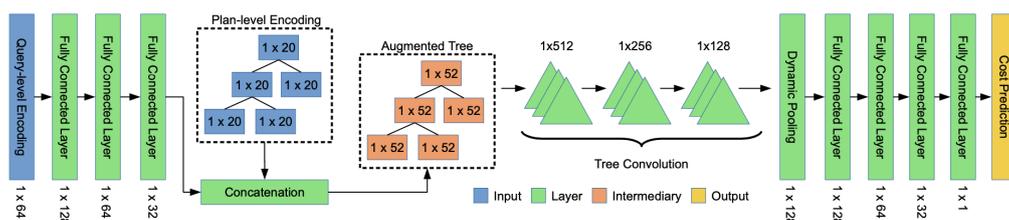
Neo 的設計模糊了傳統查詢最佳化工具主要元件 (基數估計，成本模型和計畫搜索演算法) 之間的界限。Neo 將基數估計和成本模型兩個功能組合在一個價值網路中，在價值網路的指導下，Neo 對查詢計畫空間進行了簡單的搜索以制定決策。隨著 Neo 發現更好的查詢計畫，Neo 的價值網路會不斷完善，將搜索重點放在更好的計畫上，這是一個強化學習的過程。強化學習過程將會持續進行，直到 Neo 的決策政策收斂為止。

Neo 的運行分為兩個階段：初始階段和運行時階段，如上圖所示。在初始階段，Neo 使用傳統的查詢最佳化工具為樣本工作負載中的每個查詢創建查詢執行計畫。這些執行計畫及其延遲會添加到 Neo 的經驗中。這些經驗被用作價值模型訓練的起點。建立初始價值模型可預測給定部分或完整執行計畫的最終執行時間。在運行時階段，Neo 使用價值模型在查詢執行計畫的空間中進行搜索，並以最短的預測執行時間發現計畫。隨著 Neo 優化更多查詢，價值模型將得到改進並針對使用者資料庫進行量身定制。

為了對查詢進行合理的表示，Neo 對查詢本身和查詢計畫分別進行了編碼。查詢編碼由兩個部分組成，首先將查詢的連接圖編碼為鄰接矩陣，由於該矩陣是對稱的，因此僅對上三角部分進行編碼；為了對列謂詞向量進行編碼，Neo 基於自然語言處理模型 word2vec，將列謂詞向量中的每個條目編碼為包含與謂詞相關的包含語義資訊的向量。這種編碼需要在資料庫中的資料上建立模型進行訓練，行中頻繁出現的值被映射成相似的向量。查詢計畫編碼將執行計畫的每個節點編碼為一個向量，並按照節點在查詢計畫中的位置將這些向量組織成向量樹，如下圖所示。

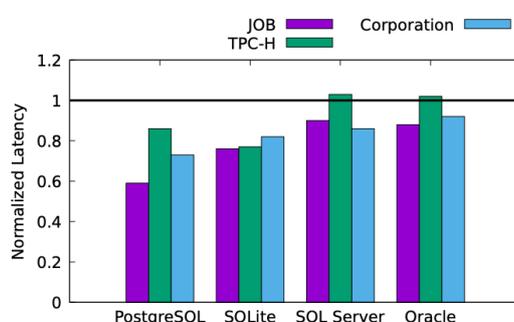


Neo 的價值網路結構如下圖所示，Neo 通過稱為樹卷積的技術從樹狀查詢計畫表示中進行學習。查詢編碼首先通過多個全連接層的，每個層的大小都減小。第三個全連接層輸出的向量與樹狀查詢計畫編碼的每一個向量連接在一起。與查詢編碼合併以後的樹狀查詢計畫編碼接著就通過幾個樹卷積層，之後，進入動態池化層，將樹結構展平為單個向量。最後，幾個附加的全連接層用於將此向量，並將此向量映射成單個值，用作輸入計畫的代價估計值。



研究結果：

如下圖所示，顯示了 Neo 在測試資料集上進行 100 次訓練反覆運算後 Neo 的相對性能（越低越好）。例如，利用 PostgreSQL 和 JOB 工作負載，Neo 生成的查詢僅比原始 PostgreSQL 優化器創建的查詢佔用平均執行時間的 60%。此外，對於 SQL Server 以及 JOB 和 Corp 工作負載，Neo 生成的查詢計畫也比 SQL Server 商業優化程式創建的計畫快 10%。總體而言，該實驗表明 Neo 的表現，與開源優化器以及同類的商業產品一樣好，有時甚至好於這些產品。



論文題目：*SageDB: A learned database system*

中文題目：SageDB：學習型資料庫系統

論文作者：T Kraska, M Alizadeh, A Beutel, EH Chi, J Ding, et al.

論文出處：9th Biennial Conference on Innovative Data Systems Research (CIDR) – CIDR 2019

論文地址：<https://ai.google/research/pubs/pub47669>

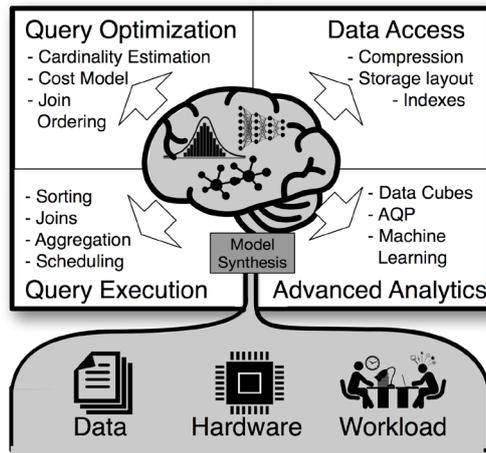
研究問題：

現代資料處理系統通常被設計為通用的，因為它們可以處理各種不同的資料模式、資料類型和資料分佈，並旨在通過使用優化器 (Optimizer) 和代價模型 (Cost

Model) 來提供對資料的高效訪問。這種通用資料庫參數設定自然導致資料庫系統無法利用特定應用程式和使用者資料的特徵來進行特定的優化。該論文的研究者提出了 SageDB , 一種對新型資料處理系統。該系統高度專注于通過代碼合成和機器學習來開發應用程式。通過對資料分佈 , 工作負載和硬體進行建模 , SageDB 可以學習資料的結構以及最佳存取方法和查詢計畫。這些學習的模型通過代碼合成被深深地嵌入到資料庫的每個元件中。

研究方法 :

SageDB 的核心思想是建立一個或多個資料和工作負載分佈有關的模型 , 並基於這些模型自動地為資料庫系統的所有元件構建最佳的資料結構和演算法。這種稱為“合成資料庫”的方法將使每個資料庫元件的實現都專用於特定的資料庫 , 查詢工作負載和執行環境 , 從而使資料庫系統的性能大大提高。如下圖所示 , SageDB 從總體上構建了一個能對資料負載分佈感知的模型 , 並自動地為資料庫每個元件選擇合適的演算法和資料特徵。儘管這套系統目前還停留在理論論證層面 , 還有很多突出的問題沒有徹底解決 , 但是它為未來的資料庫架構和組織管理方法提供了非常大膽的方向。當每個元件乃至於決策系統都具有自主學習和優化能力 , 資料庫將能更好的為使用者服務。



接下來，我們以學習排序模型為例，詳細介紹 SageDB 的工作原理。加快排序的基本思想是使用 CDF 模型將記錄大致按排序順序放置，再對該結果進行更正成幾乎完美排序的結果。根據模型的執行代價及其精度，這種排序技術可以在性能方面優於其他分類技術。例如，假定以下查詢：“SELECT * FROM customer c ORDER BY c.name”，並假定表的大小為 N ，在 $c.name$ 這一系列上有一個學習索引 (Learned Index)，且該列的資料是以 $customer\ id$ 而不是名稱的順序存儲。為了快速按 $customer\ name$ 的資料進行排序，我們可以通過 CDF 模型將每條記錄的鍵 k 放置到合適的位置，這樣子可以得到一個大致的排序結果。至此，我們將每個記錄映射到一個在輸出陣列中的位置，如果發生衝突，我們將其存儲一些溢出陣列 (演算法 1 中的第 5-10 行)。在為了減少碰撞次數，我們可以改為分配一個比 N 大 m 倍的輸出陣列 (例如 $m = 1.37$)，然後刪除此映射中所有為空的位置。注意，在這種情況下，預測的位置的計算方式為 $pos = F(k) * m * N$ (演算法 1 中的第 6 行)。假設如果我們將存儲桶設為緩存行的倍數，那麼我們可以在桶中進行快排，同時減少衝突的可能性。一個類似的想法用作 Cuckoo Hashing 的一部分。無論哪種情況，如果我們模型是非單調的，我們必須使用高效的 localsort 演算法 (即插入排序) 以糾正所有排序錯誤 (演算法 1 中的第 12 行)。最後，

我們對溢出陣列進行排序（演算法 1 中的第 13 行）並且合併排好序的陣列和溢出陣列，同時刪除空元素（演算法 1 中的第 15 行）。

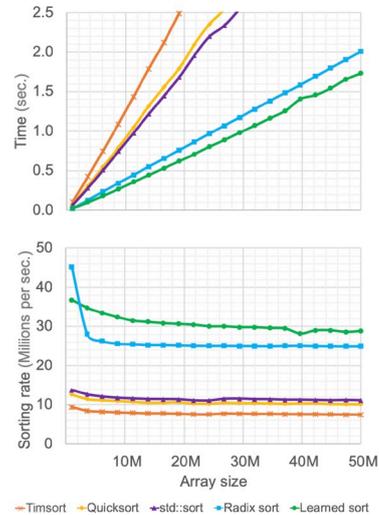
Algorithm 1 Learned sorting algorithm

Input a - the array to be sorted
Input F - the CDF model for the distribution of a
Input m - the over-allocation parameter
Output o - the sorted version of array a

```

1: procedure LEARNED-SORT( $a, F, m$ )
2:    $o \leftarrow [\infty] * (a.length * m)$ 
3:    $s \leftarrow \{\}$ 
4:   // STEP 1: Approximate ordering
5:   for  $i$  in  $a$  do
6:      $pos \leftarrow F(i) * a.length * m$ 
7:     if  $o[pos] = \infty$  then
8:        $o[pos] \leftarrow i$ 
9:     else
10:       $s \leftarrow s \cup \{i\}$ 
11:   // STEP 2: Touch-up
12:   INSERTION-SORT( $o$ )
13:   QUICKSORT( $s$ )
14:   // STEP 3: Merging
15:   return MERGE-AND-REMOVE-EMPTY( $o, s$ )

```



研究結果：

以學習排序為例，該論文以陣列大小為變數，對比了多種排序方法。從上圖的結果可以看出，在排序時間（Time）和排序比率（Sorting Rate）兩個指標上，SageDB 提出的 Learned Sort 方法均比其他方法好，驗證了所提方法的有效性。SageDB 為資料庫的設計提供了一種新的思路，並在初步的實驗上驗證了這種思路的可行性。

論文題目： *QTune: A Query-Aware Database Tuning System with Deep Reinforcement Learning*

中文題目：QTune：一種基於深度強化學習的對查詢感知的資料庫調優系統

論文作者：Guoliang Li, Xuanhe Zhou, Shifu Li, Bo Gao.

論文出處：45th International Conference on Very Large Data Bases – VLDB 2019

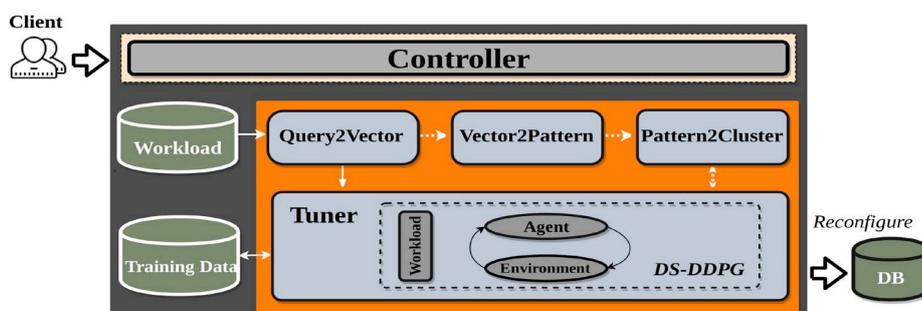
論文地址：<http://www.vldb.org/pvldb/vol12/p2118-li.pdf>

研究問題：

傳統的資料庫調優通常依靠雇傭專家(DBA)來完成:專家針對指定的負載,線上下反復進行瓶頸檢測、參數調整和性能對比,直至達到滿意效果,這是一項非常耗時的工作,而且嚴重依賴於 DBA 自身的經驗和知識。此外,在雲資料庫環境下,要面臨更加繁多的資料庫狀態和負載類型,極大地增加了這項工作的難度。雖然目前有一些自動配置工具可以供使用,但是這些調優工具多是由特定供應商創建,很難支持其他 DBMS,如 MySQLTuner。此外,這些調優工具都是基於有限的規則,很容易重複推薦錯誤的方案,無法自動的從以前的調優工作中獲得知識來優化規則。因此該論文的研究者提出了 QTune,一種基於學習的資料庫自動調優系統。該系統高度專注于利用歷史資料學習負載和資料庫環境到最優配置的映射策略,從查詢編碼、參數學習、多細微性調參三個方面提供智慧的資料庫調優服務。

研究方法：

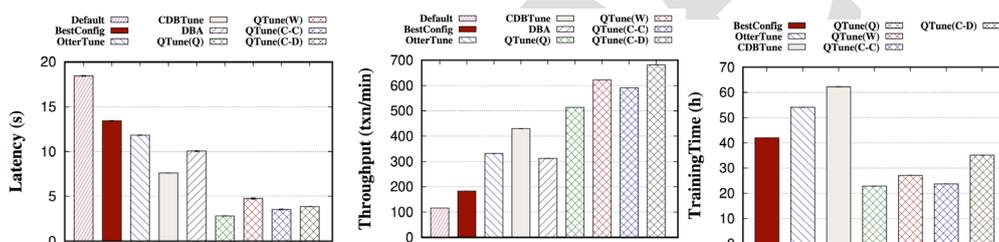
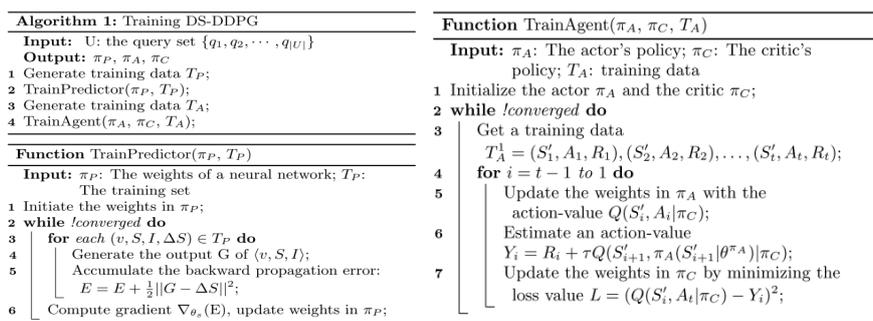
QTune 是一套對查詢感知的資料庫調優系統,其核心思想是基於深度強化學習演算法學習在不同的環境條件下推薦參數的策略。如下圖所示,QTune 主要從三個方面提供智慧的資料庫調優服務。



其一，為了提高對不同負載的適應能力，QTune 在查詢計畫級別對作業進行編碼，對執行開銷作預評估。其二，為了提高調優表現，QTune 的調優模型基於一種面向調優問題的深度強化學習方法——DS-DDPG，不僅基於 Actor-Critic 演算法，大大提高學習效率，而且能夠根據當前負載、資料庫狀態特徵綜合決策，提高調優表現的同時增強了對不同負載、資料庫實例等的適應能力。其三，為了更好地滿足不同使用者對資料庫性能的需求，QTune 增加 Query2Cluster 模組，先對作業進行配置類別劃分、批量聚類，再根據結果進行查詢簇級別的批量調優，能夠有效地平衡輸送量和延遲的優化程度。此外，QTune 還支援查詢級別、負載級別的調優模式，分別滿足延遲敏感型、輸送量敏感性用戶的需求。

接下來，我們詳細介紹 QTune 是如何訓練調優模型的。整個調優模型的訓練分為兩個部分。首先，預測模型 (Predictor) 是一個五層的神經網路，負責將負載特徵(v)轉換成資料庫的狀態變化量 ΔS 。因此，QTune 用批量梯度下降(Batch Gradient Descent) 的方法訓練預測模型：批量梯度演算法每一輪累積所有樣本的誤差值。對於負載樣本 W ，QTune 將 W 解析成特徵向量 v ；然後預測模型輸入特徵向量 v 、資料庫外部狀態 s 、內部配置 I ，輸出估計的狀態變化量 ΔS (函數 TrainPredictor 中的第 4 行)；然後計算 ΔS 相對於真實值得平方差值，並累加到總誤差 E 上 (函數 TrainPredictor 中的第 5 行)。預測完每個樣本後，QTune 用整體的平均誤差 E 更新網路權重 (函數 TrainPredictor 中的第 6 行)。其次，代理模型 (Agent) 用於決定如何調整資料庫參數，其訓練採用 Actor-Critic 演算法。在代理模型中，行動方 (Actor) 負責實際調參，而評判方 (Critic) 負責給每個調參決策打分。訓練中，對於每個樣本 (S'_i, A_i, R_i)， A_i 是一個推薦的參數集合， S'_i 是查詢在 A_i 下執行後的資料庫狀態， R_i 是查詢執行的表現增益。對於每個($S'_i,$

A_i), 評判方先輸出一個 Q 值 (函數 TrainAgent 中的第 5 行), 行動方根據 Q 值更新自己的網路權重; 然後評判方根據蒙特卡洛理論用 R_i 計算長期受益 Y_i (函數 TrainAgent 中的第 6 行), 用 Y_i 計算誤差值 L; 最後 QTune 用誤差值 L 更新評判方的網路權重 (函數 TrainAgent 中的第 7 行)。



研究結果：

如上圖所示，經過在 TPC-H、JOB、Sysbench 等標準測試集的測試，QTune 在延遲、輸送量、訓練時間等方面都有較大的提升，如同比業界最優的 CDBTune 輸送量平均有 23.4% 的提高，延遲平均有 29.3% 的降低。目前 QTune 支持 GaussDB PostgreSQL、MySQL 三種關係型數據庫和 MongoDB 一種非關係型數據庫，主要在單節點上提供自調優服務。

論文題目：*An End-to-End Learning-based Cost Estimator*

中文題目：一個端到端的基於學習的代價估計器

論文作者：Ji Sun, Guoliang Li.

論文出處：46th International Conference on Very Large Data Bases (VLDB 2020)

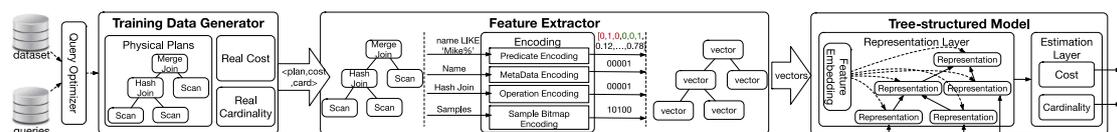
論文地址：<http://www.vldb.org/pvldb/vol13/p307-sun.pdf>

研究問題：

代價和基數估計是資料庫查詢優化的核心問題，他用於指導優化器選擇最高效的執行計畫。其中，基數指的是子查詢計畫得到的結果的行數，而代價指的是子查詢計畫所消耗的 CPU 和 IO 時間。由於傳統依賴於經驗的基數和代價估計方法無法識別多個資料表之間的相關關係，所以他們無法提供高品質的多表連接基數 (代價) 估計。最近資料庫社區越來越多的研究顯示基於學習的基數估計和基於經驗的方法相比，性能有顯著提升。但是，已有的基於學習的方法依然存在一些局限性。比如，他們主要關注估計基數，而忽略了對於代價的估計，他們使用的模型要麼無法表示複雜的查詢計畫，要麼模型過於複雜而無法高效完成代價估計的任務。

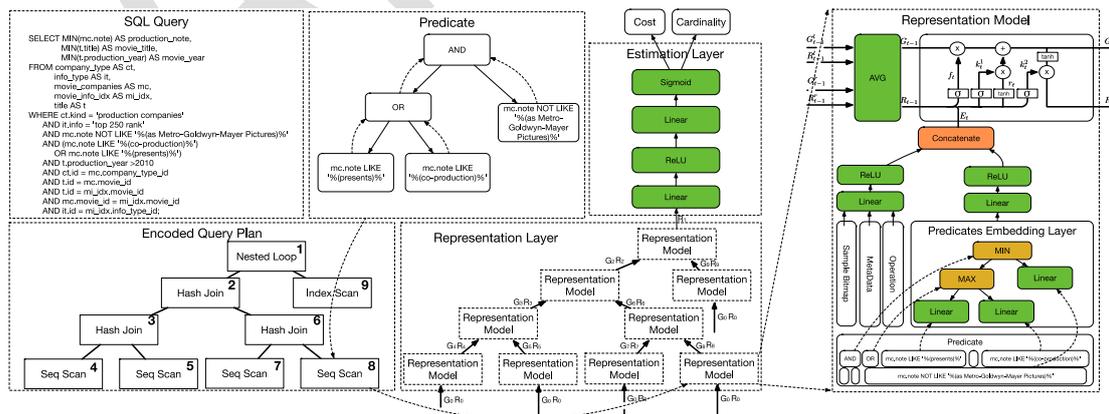
研究方法：

文章提出一個高效的基於樹形遞迴神經網路的代價估計框架，這個框架支持同時估計查詢基數和代價。對於估計演算法，文章提出了包含查詢以及具體物理執行操作的高效的特徵抽取和編碼技術。對於包含字串模式匹配的查詢準則，文章提出一個高效的基於詞向量的編碼方法以及對應高效詞典抽取方法用於從資料庫文本中抽取訓練詞向量所需要的字元子串。



如上圖所示，文章提出的框架包含四個部分，1) 訓練資料生成器用於根據資料表連接圖隨機生成一系列的查詢以及對應的真實基數和代價，輸入模型進行訓練。2) 特徵抽取模組從查詢計畫中抽取重要的特徵並且編碼成浮點向量作為輸入。3) 樹形結構的模型是一個隨著執行計畫變化而變化的動態遞迴神經網路，他的每個節點用子節點的輸出和對應執行計畫節點的特徵作為輸入，輸出作為上層節點的輸入，依次迴圈直到根節點。4) 表示記憶池用來記錄已經計算過的子查詢的表示，這樣對於執行優化器選擇計畫的過程中對於同樣的子查詢不需要重複計算。

如下圖所示，神經網路模型也分為三個層次，1) 嵌入層將操作符，中繼資料，謂詞和採樣從係數矩陣編碼成稠密矩陣。特別的，對於複合查詢謂詞，使用與或池化進行嵌入編碼。2) 展示層，這一層使用 LSTM 神經元組，接受孩子節點的輸出表示，和本節點的特徵進行線性非線性融合變換，輸入當前的子查詢表示。3) 輸出層，將查詢根節點的輸出表示作為輸入，通過兩層全連接神經網路輸出要估計的基數和代價。



Synthetic	median	90th	95th	99th	max	mean	Synthetic	median	90th	95th	99th	max	mean
PostgreSQL	1.69	9.57	23.9	465	373901	154	PostgreSQL	15.1	65.1	173	1200	8040	62.7
MySQL	2.07	22.6	50.6	625	458835	353	MySQL	4.51	39.7	94.7	449	7203	32.4
Oracle	1.97	12.4	40.1	473	545912	378	Oracle	6.72	41.1	124	796	6674	56.1
MSCN-NoSamp	2.14	6.72	11.5	114	1870	23.6	MSCN-NoSamp	10.3	24.7	234	569	2110	31.6
TLSTM-NoSamp	1.97	5.53	9.13	81.5	988	10.3	TLSTM-NoSamp	5.34	21.2	153	328	1345	19.8
MSCN	1.19	3.32	6.84	30.51	1322	2.89	MSCN	3.14	7.43	18.1	65.8	739	10.3
TNN	1.40	5.51	10.7	43.1	441	3.57	TNN	1.49	4.50	10.6	61.5	718	4.35
TLSTM	1.20	3.21	6.12	25.2	357	2.87	TLSTM	1.56	4.47	10.7	57.7	689	4.45
TPool	1.18	3.19	6.05	24.5	323	2.81	TLSTM-Multi	1.49	4.33	10.2	55.8	624	4.16
							TPool	1.48	4.12	10.1	47.6	532	3.99
Scale	median	90th	95th	99th	max	mean	Scale	median	90th	95th	99th	max	mean
PostgreSQL	2.59	200	540	1816	233863	568	PostgreSQL	13.3	38.9	81.1	718	1473	35.7
MySQL	3.08	90.1	329	7534	54527	426	MySQL	4.25	37.4	131	577	5157	40.7
Oracle	2.43	114	482	3412	102833	397	Oracle	6.49	27.7	61.4	623	3612	31.5
MSCN-NoSamp	2.33	96.1	257	1110	4013	131	MSCN-NoSamp	3.32	20.9	30.5	274	1173	21.2
TLSTM-NoSamp	2.06	69	176	931	3295	78.2	TLSTM-NoSamp	2.19	13.4	21.7	228	1162	14.9
MSCN	1.42	37.4	140	793	3666	35.1	MSCN	1.79	10.6	27.1	88.8	1027	8.22
TNN	1.59	58.7	141	573	2238	31.3	TNN	1.61	5.37	13.5	72.7	714	5.53
TLSTM	1.43	38.8	139	469	1892	28.1	TLSTM	1.58	5.51	14.4	70.1	611	5.21
TPool	1.42	37.3	125	345	1813	26.3	TLSTM-Multi	1.56	5.56	12.2	68.6	254	4.41
							TPool	1.54	5.29	11.9	67.6	254	4.39
JOB-light	median	90th	95th	99th	max	mean	JOB-light	median	90th	95th	99th	max	mean
PostgreSQL	7.93	164	1104	2912	3477	174	PostgreSQL	26.8	332	696	2740	3020	173
MySQL	9.55	303	685	2256	2578	149	MySQL	9.47	102	342	1293	2228	84.5
Oracle	8.32	374	976	2761	3331	157	Oracle	12.3	157	278	1366	1825	102.1
MSCN-NoSamp	5.43	126	978	1310	2020	100	MSCN-NoSamp	12.4	152	231	1071	1553	62.7
TLSTM-NoSamp	5.18	97.3	613	864	1541	72.3	TLSTM-NoSamp	10.4	105	217	986	1271	38.3
MSCN	3.82	78.4	362	927	1110	57.9	MSCN	4.75	11.3	40.1	563	987	27.4
TNN	2.95	76.8	275	799	902	49.8	TNN	2.06	25.5	134	293	401	19.1
TLSTM	3.73	50.8	157	256	289	24.9	TLSTM	3.66	32.1	80.3	445	583	17
TPool	3.51	48.6	139	244	272	24.3	TLSTM-Multi	1.85	13.2	22.9	95	123	5.81
							TPool	1.85	11.1	20.3	101	125	5.76

Cardinality	median	90th	95th	99th	max	mean	Cost	median	90th	95th	99th	max	mean
PostgreSQL	184	8303	34204	106000	670000	10416	PostgreSQL	4.90	80.8	104	3577	4920	105
MySQL	104	28157	213471	1630689	2487611	60229	MySQL	7.94	691	1014	1568	1943	173
Oracle	119	55446	179106	697790	927648	34493	Oracle	6.63	149	246	630	1274	55.3
TLSTM-Hash	11.1	207	359	824	1371	83.3	TLSTM-Hash	4.47	53.6	149	239	478	24.1
TLSTM-Emb	11.6	181	339	777	1142	70.2	TLSTM-Emb	4.12	18.1	44.1	105	166	10.3
TLSTM-EmbRule	10.9	136	227	682	904	55.0	TLSTM-EmbRule	4.28	13.3	22.5	104	126	8.6
TPool	10.1	74.7	193	679	798	47.5	TPool	4.07	11.6	17.5	63.1	67.3	7.06

研究結果：

文章提出了完整的對於查詢計畫的基數和代價估計方法，如上圖實驗結果顯示，對於 IMDB 資料集，無論是在 Synthetic, Scale, JOB-light 還是 JOB-full 查詢上，該方法對於基數和代價的估計都要比傳統方法和目前最新的基於深度學習的方法準確高效。

論文題目：*Software Engineering for Machine Learning: a Case Study*

中文題目：面向機器學習的軟體工程：一個案例研究

論文作者：Amershi, Saleema, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann

論文出處：41st International Conference on Software Engineering (ICSE) ---- ICSE 2019

論文地址：<https://dl.acm.org/citation.cfm?id=3339967>

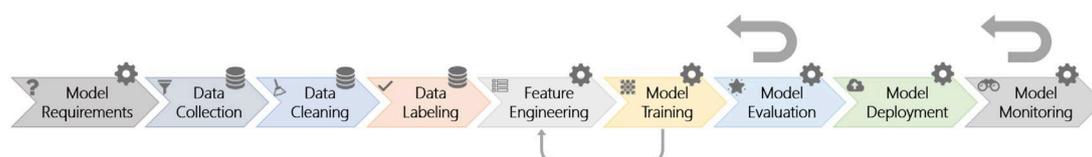
研究問題：

從個人計算、互聯網、移動計算到雲計算，每個轉變都為軟體工程帶來了新的目標，這些目標促使研究新的開發模式，以解決軟體工程在這些新領域遇到的問題。該論文首先概述了衝擊軟體行業的最新趨勢，即在軟體發展週期中集成人工智慧(AI)的功能。該論文調研了一些使用機器學習來開發應用程式(例如 Bing Search 或 Cortana 虛擬助手) 的 Microsoft 產品團隊，和使用即時翻譯文本，語音和視頻的平臺 (例如 Microsoft Translator) ，以及為客戶構建自己的機器學習應用程式 Azure AI 的平臺。

研究方法：

在該論文中，通過案例分析，瞭解到了一些 Microsoft 軟體團隊是如何構建以客戶為中心的 AI 功能的軟體應用程式。為此，Microsoft 將現有的敏捷開發軟體工程流程與 AI 專用 workflow 集成在一起，這些 workflow 得益於開發早期 AI 和資料科學應用程式的先前經驗。在這項研究中，他們調研了 Microsoft 員工是如何應對 AI 軟體發展以及大規模 AI 基礎架構和應用程式開發所帶來的日益增長的挑戰。研究指出，由於整個公司的團隊在 AI 方面擁有不同的工作經驗，隨著團隊的成熟，新團隊報告中出現許多問題的在這些成熟的團隊中的重要性急劇下降，而某些問題對於大規模 AI 的實踐仍然至關重要。這項研究嘗試創建新的 workflow 度量標準，以說明開發團隊在開發 AI 應用程式週期中做到進度管理。如下圖所示，該論文調研了面向機器學習的軟體工程的開發週期，這個開發週期通常包含了九個階段，有些階段是面向資料的 (例如，收集，清潔和標記) ，而另一些階段是面向模型的 (例如，模型需求，功能工程，培訓，評估，部署和監視) 。

工作流程中有許多回饋迴圈。較大的回饋箭頭表示模型評估和監視可能會迴圈回到之前的任何階段，例如在模型評估階段，如果發現當前的模型效果不理想，有可能會回到資料清洗的階段來進一步清洗資料以提高模型的品質。較小的回饋箭頭說明模型訓練可以迴圈回到特徵工程（例如，在表示學習中）。



該論文在進行案例分析研究的過程中，主要有兩個階段收集資料：1、通過一組訪談來收集與研究問題相關的主要主題；2、以及對已收集的有意義的主題進行大規模調查。

Id	Role	Product Area	Manager?
I1	Applied Scientist	Search	Yes
I2	Applied Scientist	Search	Yes
I3	Architect	Conversation	Yes
I4	Engineering Manager	Vision	Yes
I5	General Manager	ML Tools	Yes
I6	Program Manager	ML Tools	Yes
I7	Program Manager	Productivity Tools	Yes
I8	Researcher	ML Tools	Yes
I9	Software Engineer	Speech	Yes
I10	Program Manager	AI Platform	No
I11	Program Manager	Community	No
I12	Scientist	Ads	No
I13	Software Engineer	Vision	No
I14	Software Engineer	Vision	No

1. Part 1
 - 1.1. Background and demographics:
 - 1.1.1. years of AI experience
 - 1.1.2. primary AI use case*
 - 1.1.3. team effectiveness rating
 - 1.1.4. source of AI components
 - 1.2. Challenges*
 - 1.3. Time spent on each of the nine workflow activities
 - 1.4. Time spent on cross-cutting activities
2. Part 2 (repeated for two activities where most time spent)
 - 2.1. Tools used*
 - 2.2. Effectiveness rating
 - 2.3. Maturity ratings
3. Part 3
 - 3.1. Dream tools*
 - 3.2. Best practices*
 - 3.3. General comments*

受訪者的基本資訊如上圖左圖所示，可以看出，受訪者有從事一線軟體發展工作的工程師、應用科學家、專案經理、架構師以及研究員，從事的領域包括了搜索、語音和視頻等。採訪的內容如上圖右圖所示，主要詢問受訪者的 AI 工作背景、目前工作中遇到的難題、開發週期中主要的工具和一些開發訴求和感想等。該論文研究者將調查問卷以內部郵件的形式發送給有關 AI 和 ML 相關的 4195 位員工。551 名員工填寫了調查問卷，這項問卷的回應率為 13.6%。受訪者分佈在公司的所有部門中，並且來自各種職位：資料和應用科學（42%），軟體工程

(32%) , 程式管理 (17%) , 研究 (7%) 和其他 (1%) 。 21% 的受訪者是經理 , 這有助於該論文在訪談中平衡大多數經理的觀點。

研究結果 :

如下表所示 , 可以看出對於開發經驗較少(Low) 中(Medium)和高(High) 的受訪者來說 , 最困難的就是資料的收集、清洗和管理工作。然而 , 對於 AI 領域的新知識的教育和訓練 , 則認為沒有那麼重要。通過這項案例研究 , 我們可以看到資料管理在面向機器學習的軟體工程週期中是十分重要的 , 也是十分具有挑戰性的。

Challenge	Frequency			Rank		
	Medium vs. Low	High vs. Low	Trend	Low	Experience Medium	High
Data Availability, Collection, Cleaning, and Management	-2%	60%		1	1	1
Education and Training	-69%	-78%		1	5	9
Hardware Resources	-32%	13%		3	8	6
End-to-end pipeline support	65%	41%		4	2	4
Collaboration and working culture	19%	69%		5	6	6
Specification	2%	50%		5	8	8
Integrating AI into larger systems	-49%	-62%		5	16	13
Education: Guidance and Mentoring	-83%	-81%		5	21	18
AI Tools	144%	193%		9	3	2
Scale	154%	210%		10	4	3
Model Evolution, Evaluation, and Deployment	137%	276%		15	6	4

論文題目 : *Declarative Recursive Computation on an RDBMS, or, Why You Should Use a Database for Distributed Machine Learning*

中文題目 : RDBMS 上的聲明式遞迴計算 , 或者為什麼要使用資料庫進行分散式機器學習

論文作者 : Jankov, Dimitrije, Shangyu Luo, Binhang Yuan, Zhuhua Cai, Jia Zou, Chris Jermaine, and Zekai J. Gao

論文出處 : 45th International Conference on VeryLarge Data Bases – VLDB 2019

論文地址：<http://www.vldb.org/pvldb/vol12/p822-jankov.pdf>

研究問題：

諸如 TensorFlow 之類的現代機器學習 (ML) 平臺主要是為了支援資料並行性而設計的，其中在一組計算單元上並存執行一組幾乎相同的計算 (例如，梯度計算)。計算之間的唯一區別是，每個計算都針對不同的訓練資料 (稱為“批次”) 進行操作。對於非同步資料並行性而言，在每次計算完成後，都將本地結果載入到參數伺服器；對於同步資料並行性而言，在每次計算完成後，都將本地結果全域聚合並用於更新模型。但問題是，資料並行性有其局限性。例如，資料並行性隱含地假設正在學習的模型 (以及使用批次處理更新模型時生成的中間資料) 可以適合計算單元 (可以是伺服器或 GPU) 的 RAM。但是，該假設有時候不成立。例如，目前最好的 NVIDIA Tesla V100 Tensor Core GPU 具有 32GB 的 RAM，然而這 32GB 的 RAM 無法存儲全連接層的 200,000 個類別的條目的向量編碼。處理這樣規模的模型需要並行化，在這種模型中，要學習的統計模型不是簡單地在不同的計算單元上進行複製，而是要進行分區和平行作業，並由一系列批量同步操作執行。然而，現有的分散式機器學習系統為模型並行性提供的支援十分有限。該論文研究如何對現代關聯式資料庫管理系統 (RDBMS) 進行少量更改，以使其適合於分散式學習計算。

研究方法：

該論文認為可以使用關係型數據庫的相關技術來實現模型並行化。更加具體的，模型的不同部分可以存儲在一組表中，部分模型上的計算通常可以通過一些 SQL 查詢來表示。實際上，對於開發人員來說，學習演算法的模型並行 SQL 與

資料並行實現很相似。關聯式資料庫管理系統 (RDBMS) 提供了聲明性的程式設計介面，這意味著程式師或自動演算法生成器，如果通過自動微分自動生成了 ML 演算法，則只需指定他/她/它想要的內容，而無需指定需要寫出如何計算它。計算將由系統自動生成，然後進行優化和執行以匹配資料大小，佈局和計算硬體。無論計算是在本地電腦上還是在分散式環境中運行，代碼都是相同的。相反，諸如 TensorFlow 之類的系統提供的聲明性相對較弱，因為必須在某個物理計算單元 (例如 GPU) 上指定並執行計算圖中的每個邏輯運算 (例如矩陣乘法)。使用關係技術的另一個好處是，RDBMS 中的分散式運算已經研究了三十多年，並且速度快，功能強大。RDBMS 附帶的查詢最佳化工具對於優化分散式運算非常有效。但是，有兩個原因導致 RDBMS 無法作為大多數大型 ML 演算法的平臺直接使用。首先，RDBMS 缺乏足夠的遞迴支持。在深度學習中，“遍歷”深度神經網路的各個層，然後向後“遍歷”整個網路以傳播錯誤是必要的操作。雖然在 RDBMS 中，這種“迴圈”可以通過表之間的遞迴依賴關係來聲明性地表示，但是 RDBMS 對遞迴的支援通常僅限於通過諸如傳遞閉包之類的集合來計算 fixed-points。另一個挑戰是，典型的深度學習計算的查詢計畫可能會運行到成千上萬的運算子，而現有的 RDBMS 優化器將無法處理該運算子。簡單來說，這篇文章向資料庫表引入了多維的，類似陣列的索引。當一組表共用相似的計算模式時，可以將它們壓縮並替換為具有多個版本 (由其索引指示) 的表。此外，該論文還修改了資料庫的查詢最佳化工具，以使其能夠應對非常大的查詢圖 (Query Graphs)。一個查詢圖被劃分為一組可運行的 *frames*，優化了操作的代價，並將圖劃分問題 (Graph-cutting Problem) 形式化為廣義二次賦值 (Generalized Quadratic Assignment) 問題。

FFNN				Word2Vec		
Hidden Layer Size	RDBMS (CPU)	RDBMS (GPU)	TensorFlow (GPU)	Embedding Dimensions	RDBMS	TensorFlow
\$3 per hour budget				100	00:16:43 (00:01:59)	00:08:03
10000	04:50	06:25	00:24	1000	00:17:05 (00:01:53)	01:14:58
20000	07:07	07:12	Fail	10000	00:29:18 (00:01:53)	Fail
40000	11:52	11:48	Fail			
80000	16:30	Fail	Fail			
160000	Fail	Fail	Fail			
\$7 per hour budget				Collapsed LDA		
10000	04:53	04:58	00:15	Number of Topics	RDBMS	TensorFlow
20000	05:54	06:08	Fail	1000	00:06:25	00:05:06
40000	09:32	08:26	Fail	5000	00:06:54	00:25:22
80000	12:03	17:50	Fail	10000	00:07:05	00:52:35
160000	Fail	Fail	Fail	50000	00:08:32	04:51:51
\$15 per hour budget				100000	00:09:58	Fail
10000	05:12	5:00	00:12			01:42:35
20000	05:36	06:30	Fail			
40000	09:08	08:39	Fail			
80000	12:24	12:20	Fail			
160000	39:40	Fail	Fail			

研究結果：

該論文在 SimSQL 的基礎上進行了實驗，主要測試了三種機器學習演算法的分散式實現：(1) 多層前饋神經網路 (FFNN)；(2) Word2Vec 演算法；(3) 分散式的 LDA 話題模型。如上圖所示，在 FFNN 上，當隱層的大小較小的時候 (10000)，該論文提出的方法並不能超過 TensorFlow，但是當隱層的大小增大的時候，TensorFlow 並不能完成計算，而該論文提出的方法能進行有效地計算，同樣的結論也可以在 Word2Vec 演算法和 LDA 話題模型上得出。研究結果表明，該論文提出的方法能夠處理需要在機器或計算單元之間分佈的大型複雜模型。與 TensorFlow 相比，基於 RDBMS 的模型並行機器學習計算框架具有很好的伸縮性，對於 Word2Vec 和 LDA，基於 RDBMS 的計算可以比 TensorFlow 更快。但是，對於基於 GPU 的神經網路實現，RDBMS 的速度比 TensorFlow 慢。

11.5 資料庫技術重要進展

步入大資料時代，面對 PB 乃至 EB 級海量資料、複雜多變的應用場景、異構的硬體架構和層次不齊的使用者使用水準，傳統的資料管理技術難以滿足新時代的需求。例如，一個雲資料庫系統通常具有百萬級別的資料庫實例，每一個資料庫實例通常都有各自的應用場景，不同用戶的使用水準往往也有著比較大的差

別，資料庫中傳統的啟發式演算法在這些場景中難以取得較好的效果，而有經驗的資料庫管理員也難以直接干預和優化數量如此之多的資料庫實例。

近年來，以機器學習為代表的人工智慧技術因其強大的學習和適應能力，在多個領域都大放異彩。同樣的，在資料管理領域，傳統機器學習和深度學習等技術也有著巨大的潛力和廣闊的應用前景。例如，資料庫系統所積累的海量歷史查詢記錄可以為基於學習的資料庫智慧優化技術提供資料支撐。一方面，我們可以構建包含查詢、視圖或資料庫狀態的有標籤資料，比如，在視圖選擇問題中，這個標籤是指每個候選視圖是否被選中。另一方面，在缺乏標籤資料的時候，我們可以利用（深度）強化學習技術探索性地（從選擇結果的回饋中學習）選擇最優的候選視圖。此外，人工智慧技術讓自治資料庫的自動決策管理、自動調優和自動組裝等需求成為可能。在以深度學習為代表的人工智慧技術的加持下，讓資料庫朝著更加智慧的方向發展，資料管理技術也隨之智慧化。近些年湧現的自治資料庫和人工智慧原生資料庫（如 SageDB，XuanyuanDB），通過融合人工智慧技術到資料庫系統的各個模組（優化器、執行器和存儲引擎等）和資料管理的生命週期，可以大幅度提升資料庫各方面的性能，為下一代資料庫和人工智慧技術的發展指明了一個方向。

在另外一方面，資料管理技術也能以基礎設施的身份來支援人工智慧的發展。目前的人工智慧在落地過程中還面臨著一些挑戰性。例如，人工智慧演算法訓練效率較低，現有人工智慧系統缺少執行優化技術（如大規模緩存、資料分塊分區、索引等），不僅會導致大量的計算、存儲資源浪費，而且會提高程式異常的發生率（如記憶體溢出、進程阻塞等），嚴重影響單個任務的執行效率。其次，人工智慧技術往往依賴高品質的訓練資料，現實中的訓練資料往往是包含很多缺失值、

異常值和別名等類型的錯誤，這些錯誤通常會影響訓練效率，對模型的品質造成干擾。面向人工智慧的資料管理技術可以為解決上述挑戰做出貢獻。

AMiner

12 視覺化技術

12.1 視覺化技術概念

視覺化技術是把各種不同類型的資料轉化為可視的表示形式，並獲得對資料更深層次認識的過程。視覺化將複雜的資訊以圖像的形式呈現出來，讓這些資訊更容易、快速地被理解，因此，它也是一種放大大人類感知的圖形化表示方法。

視覺化技術充分利用電腦圖形學、影像處理、使用者介面、人機交互等技術，以人們慣於接受的表格、圖形、圖像等形式，並輔以資訊處理技術（例如：資料採擷、機器學習等）將複雜的客觀事物進行圖形化展現，使其便於人們的記憶和理解。視覺化為人類與電腦這兩個資訊處理系統之間提供了一個介面，對於資訊的處理和表達方式有其獨有的優勢，其特點可總結為可視性、交互性和多維性。

目前，資料視覺化針對不同的資料類型及研究方向，可以進一步劃分為科學資料視覺化、資訊視覺化，以及可視分析學三個子領域。這三個領域既緊密相關又分別專注於不同類型的資料及視覺化問題。具體而言，科學視覺化是針對科學資料的視覺化展現技術。科學資料，例如，醫療過程中由 CT 掃描生成的影像資料、風洞實驗而產生的流體資料、以及分子的化學結構等，是對物理世界的客觀描述，往往是通過科學儀器而測量得到的資料。這類資料的視覺化主要關注於如何以清晰直觀的方式展現資料所刻畫的真實物理狀態。因此，科學視覺化往往呈現的是三維場景下的時空資訊。資訊視覺化注重於如何以圖形的方式直觀展現抽象資料，它涉及到了對人類圖形認知系統的研究。在這裡，抽象資料（例如：圖形資料、多維度數據、文本資料等）往往是對各應用領域所產生資料的高層次概

括，記錄的是抽象化的資訊。針對這樣的資料，資訊視覺化著眼於多維度資訊的可視編碼技術，即如何以低維度(2D)的圖形符號來直觀展現並揭示抽象資料中所隱藏的潛在規律與模式；可視分析學是多領域技術結合的產物，旨在結合並利用資訊視覺化、人機交互、以及資料採擷領域的相關技術，將人的判斷與回饋作為資料分析中重要的一環，從而達到精準資料分析、推理及判斷的目的。

視覺化技術的重要性在於，通過提供對資料和知識的展現，建立使用者與資料系統交互的良好溝通管道，利用人類對圖形資訊與生俱來的模式識別能力，通過以直觀的圖像化方式展現資料，從而說明使用者快速發覺資料中的潛在規律，並借助分析人員的領域知識與經驗，對模式進行精準分析、判斷、推理，從而達到輔助決策的目的。

目前視覺化技術在各行各業中均得到了廣泛的應用。其中，視覺化技術在資訊安全、智慧醫療、電子商務、機器學習、智慧城市、文化體育、數字新聞、氣象預報、地質勘測等諸多領域產生了非常廣泛的應用，並逐漸成為這些領域當中越來越重要的組成部分。

當下視覺化存在的挑戰是：如何進一步深入挖掘人類對於圖形、動畫、以及交互的感知及認知模式，從而進一步完善視覺化的相關理論；如何打破“手工作坊”式的針對每一個問題，單獨定制資料視覺化設計方案的傳統模式，大規模批量創造生成風格化的可視展現；以及，如何根據使用者的資料分析任務與需求自動推薦合適的視覺化展現方式。

主要的研究趨勢：海量、異構、時變、多維資料的視覺化展示方案；視覺化在可解釋性深度學習領域的應用；自動視覺化生成技術的研究；基於形式概念分

析理論的知識視覺化方法；視覺化模式識別；整體可視與局部詳細可視相結合的新方法研究等。

12.2 視覺化技術發展歷史



圖 12-1 視覺化發展歷程

19 世紀-19 世紀中葉：現代圖形學設計雛形

十九世紀前葉，因為受視覺表達方法創新的影響，統計圖形及專題繪圖領域應用得到快速發展。目前，我們看到的絕大多數統計圖形都是在這一時間段被發明的。同期，因政府開始著重關注人口、教育、犯罪、疾病等領域，資料的收集整理範圍明顯擴大，超乎以往的社會管理方面的資料被收集起來用於社會分析。

1801 年英國地質學家 William Smith 繪製了第一幅地質圖，引領了一場在地圖上表現量化資訊的潮流。這一時期，資料的收集整理從科學技術和經濟領域擴展到社會管理領域，對社會公共領域資料的收集標誌著人們開始以科學手段進行社會研究。與此同時科學研究對資料的需求也變得更加精確，研究資料的範圍也有明顯擴大，人們開始有意識地使用視覺化方式來嘗試研究、解決更廣泛領域的問題。

19 世紀中葉-末期：資料圖形繪製發展加速

在十九世紀中葉，統計圖形、概念圖等概念迅猛發展，此時的人們已經掌握了整套統計資料視覺化工具，資料視覺化領域發展進入了加速期，隨著數位資訊對社會、工業、商業直至交通規劃的影響不斷增大，歐洲開始著力發展資料分析技術。一群學者發起的統計理論給出了多種資料的意義，資料視覺化迎來了它歷史上的第一個發展加速期。統計學理論的建立是視覺化發展的重要一步，此時資料由政府機構進行收集，資料的來源變得更加規範化。隨著社會統計學的影響力越來越大，在 1857 年維也納的統計學國際會議上，學者就已經開始對視覺化圖形的分類和標準化進行討論。不同資料圖形開始出現在書籍、報刊、研究報告和政府報告等正式場合之中。這一時期法國工程師 Charles Joseph Minard 繪製了多幅有意義的視覺化作品，被稱為“法國的 Playfair”，他最著名的作品是用二維的表達方式，展現六種類型的資料，用於描述拿破崙戰爭時期軍隊損失的統計圖，如下圖所示。並且在這一時期出現了三維的資料表達方式，這種創造性的成果對後來的研究有十分突出的作用。

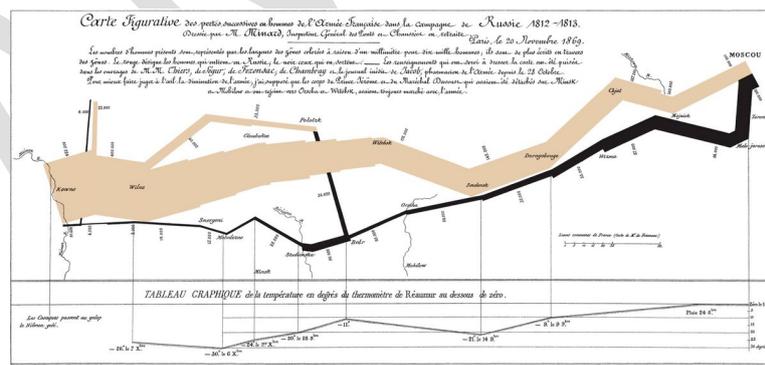


圖 12-2 拿破崙進軍莫斯科的歷史事件

20 世紀-20 世紀中葉：停滯期

20 世紀的上半葉，隨著數理統計這一新數學分支的誕生，追求數理統計嚴格的數學基礎並擴展統計的疆域成為這個時期統計學家們的核心任務。資料視覺

化成果在這一時期得到了推廣和普及，並開始被用於嘗試著解決天文學、物理學、生物學的理论新成果，Hertzsprung-Russell 繪製的溫度與恒星亮度圖成為了近代天體物理學的奠基之一；倫敦地鐵線路圖的繪製形式如今依舊在沿用（如下圖所示）；E.W.Maunder 的“蝴蝶圖”用於研究太陽黑子隨時間的變化。然而，這一時期人類收集、展現資料的方式並沒有得到根本上的創新，統計學在這一時期也沒有大的發展，所以整個上半葉都是休眠期。但這一時期的蟄伏與統計學者潛心的研究才讓資料視覺化在本世紀後期迎來了復蘇與更快速的發展。



圖 12-3 HenryBeck 設計的倫敦地鐵圖

1950-1974：復蘇

從 20 世紀上半葉末到 1974 年這一時期被稱為資料視覺化領域的復蘇期，在這一時期引起變革的最重要的因素就是電腦的發明，電腦的出現讓人類處理資料的能力有了跨越式的提升。在現代統計學與電腦計算能力的共同推動下，資料視覺化開始復蘇。隨著電腦的普及，上世紀六十年代末，各研究機構逐漸開始使用電腦程式取代手繪的圖形。由於電腦的資料處理精度和速度具有強大的優勢，高精度分析圖形已不能用手繪製。在這一時期，數據縮減圖、多維標度法 MDS、聚類圖、樹狀圖等更為新穎複雜的資料視覺化形式開始出現。人們嘗試著在一張

圖上表達多種類型資料，或用新的形式表現資料之間的複雜關聯，這也成為這一時期資料處理應用的主流方向。資料和電腦的結合讓資料視覺化迎來了新的發展階段。

1975-2011：科學計算視覺化、資訊視覺化相繼誕生

這段時期，電腦成為資料處理的一個重要工具，資料視覺化進入了新的黃金時代，隨著應用領域的增加和資料規模的擴大，更多新的資料視覺化需求逐漸出現。二十世紀七十年代到八十年代，人們主要嘗試使用多維定量資料的靜態圖來表現靜態資料，八十年代中期出現了動態統計圖，最終在上世紀末兩種方式開始合併，致力於實現動態、可交互的資料視覺化，動態互動式的資料視覺化方式成為新的發展主題。資料視覺化的這一時期的最大潛力來自動態圖形方法的發展，允許對圖形物件和相關統計特性的即時和直接的操縱。這一時段初期就已經出現交互系統，通過調整控制來選擇參考分佈的形狀參數和功率變換。這可以看作動態互動式視覺化發展的起源，並推動了這一時期資料視覺化的發展。

2012-至今：基於大資料的視覺化

步入 21 世紀互聯網資料量猛增，人們逐漸開始對大資料的處理進行了重點關注。之後全球每天的新增資料量就已經開始以指數倍膨脹，使用者對於資料的使用效率也在日益提升，資料的服務商開始需要從多個維度向使用者提供服務，大資料時代就此正式開啟。2012 年，我們進入資料驅動的時代。人們對資料視覺化技術的依賴程度也不斷加深。大資料時代的到來對資料視覺化的發展有著衝擊性的影響，繼續以傳統展現形式來表達龐大的資料量中的資訊是不可能的，大規模的動態化資料要依靠更有效的處理演算法和表達形式才能夠傳達出有價值的

資訊，因此大資料視覺化的研究成為新的時代命題。我們在應對大資料時，不但要考慮快速增加的資料量，還需要考慮到資料類型的變化，這種資料擴展性的問題需要更深入的研究才能解決；互聯網的加入增加了資料更新的頻率和獲取的管道，而即時資料的巨大價值只有通過有效的視覺化處理才可以體現，於是在上一歷史時期就受到關注的動態交互的技術已經向互動式即時資料視覺化發展。綜上，如何建立一種有效的、可互動式的大資料視覺化方案來表達大規模、不同類型的即時資料，成為了資料視覺化這一學科的主要的研究方向^[66-67]。

12.3 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為視覺化領域全球學者分佈情況：



圖 12-4 視覺化領域全球學者分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量優勢明顯；歐洲也有較多的人才分佈；

亞洲的人才主要集中在我國東部地區；其他諸如非洲、南美洲等地區的學者非常稀少；視覺化領域的人才分佈與各地區的科技、經濟實力情況大體一致。此外，在性別比例方面，視覺化中男性學者占比 91.7%，女性學者占比 8.3%，男性學者占比遠高於女性學者。

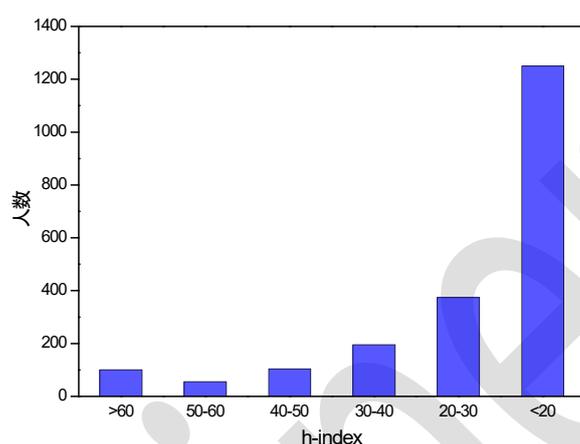


圖 12-5 視覺化領域學者 h-index 分佈

視覺化學者的 h-index 分佈如上圖所示，分佈情況大體呈階梯狀，其中 h-index 小於 20 區間的人數最多，有 1250 人，占比 60.1%，50-60 的區間人數最少，有 56 人。

● 中國人才分佈

我國專家學者在視覺化領域的分佈如下圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是長三角和珠三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，通過觀察中國周邊國家的學者數量情況，特別是與日韓等地相比，中國在視覺化領域學者數量較多但差距較小。



圖 12-6 視覺化領域中國學者分佈

中國與其他國家在視覺化的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。從表中資料可以看出，中美合作的論文數、引用數、學者數遙遙領先，表明中美間在視覺化領域合作之密切；此外，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 5 席；中國與瑞士、中國與瑞典的平均引用數都達到了最高，說明在合作品質上中瑞合作達到了較高的水準。

表 12-1 視覺化領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	152	4996	33	311
中國-新加坡	34	684	20	64
中國-英國	32	1393	44	78

中國-瑞士	31	2298	74	40
中國-瑞典	31	2298	74	40
中國-德國	23	422	18	49
中國-加拿大	19	659	35	37
中國-法國	7	83	12	17
中國-澳大利亞	2	71	36	7
中國-日本	2	58	29	7

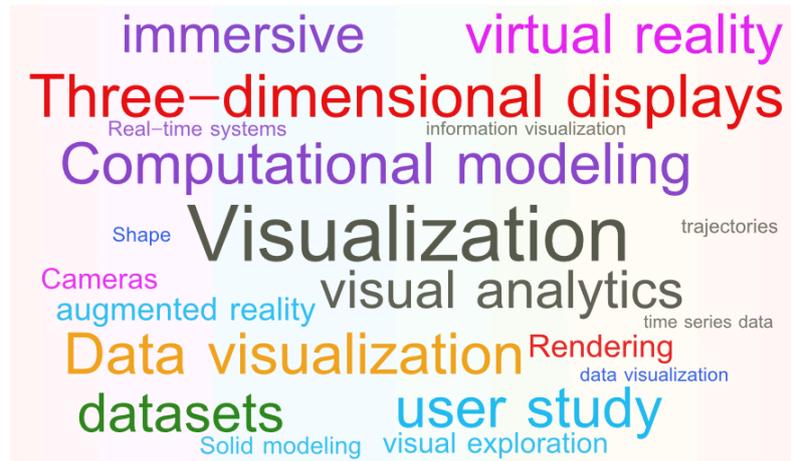
12.4 論文解讀

本節對本領域的高水準學術會議及期刊論文進行挖掘，解讀這些會議和期刊在 2018-2019 年的部分代表性工作。這些會議和期刊包括：

IEEE Transactions on Visualization & Computer Graphics

IEEE Visualization Conference

我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如下圖所示。其中，視覺化 (Visualization)、三維顯示 (Three-dimensional displays)、計算建模 (Computational modeling) 是本領域中最熱的關鍵字。



論文題目： *Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips*

中文題目： 城市大時空資料的視覺化研究:紐約市計程車出行研究

論文作者： Nivan Ferreira, Jorge Poco, Huy T. Vo, Juliana Freire, Cláudio T. Silva

論文出處： IEEE Transactions on Visualization and Computer Graphics, 2013

論文地址：

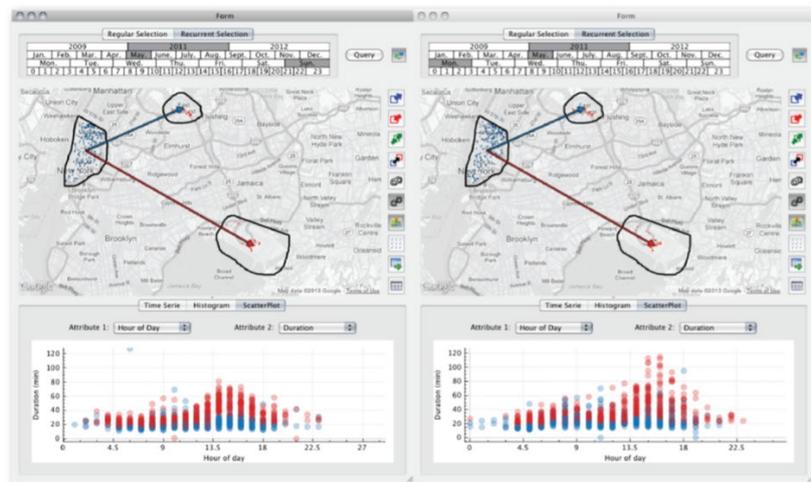
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6634127&tag=1>

研究問題：

計程車資料是城市中極具價值的資訊，收集並利用好計程車的資料可以有效的說明決策者和社會學家理解城市的狀況並做出正確的決策。但高效的探索計程車資料其實是一個充滿挑戰的事情。計程車資料十分複雜且龐大，包含了時間和空間上的資訊，很難快速查詢並進行比較。在採訪城市規劃和交通專家後，該文作者瞭解到，他們目前沒有合適的工具來完成分析。一些簡單的工具和語言只能分析一些小規模的資料，能分析的資料比較片面；而複雜一些的工具，雖然可以對大資料進行分析，則需要掌握高級的資料查詢語言，對分析人員而言很困難。

所以該文提出了一種支持在起點-終點 (OD) 資料上進行複雜時空視覺化查詢的模型。

研究方法：

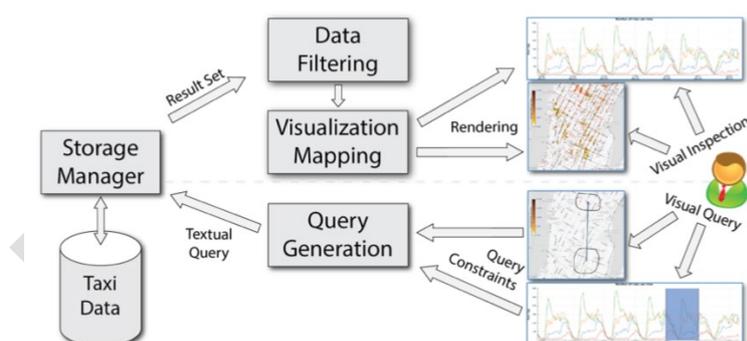


該文提出了一個視覺化系統 TaxiVis，對紐約市的計程車資料進行視覺化，讓使用者更便捷的進行可視查詢。系統的主要介面如上圖所示，主要包括三個部分，最上面的是時間選擇器，中間的是可以用來查詢和展示結果的地圖，位於下方的是資料資訊總結的一個視圖，可以將查詢到的結果通過不同的形式進行展示，例如，時間序列，長條圖，不同屬性的散點圖，支持視覺化的過濾。

系統的主要流程是，使用者通過和地圖和其他的一些視覺元素進行交互，直觀地進行查詢。系統接收到視覺化的查詢後，會在內部自動生成文字查詢。關於資料查詢的設計，作者對城市規劃和交通專家進行了採訪，將查詢需求歸納成下面三點：時間、地點和對象。對應的支持三種查詢：時間+地點->物件，時間+物件->地點，地點+對象->時間。作者提出了一個可視模型，`SELECT * FROM trips WHERE <constraints>`。Constraints 就是上述三種類別，時間，空

間，以及物件屬性上的限制。時間限制通過時間選擇器實現；空間限制可以在地圖上進行區域圈選；而屬性限制可以通過在屬性長條圖上篩選完成。

之後，在存儲管理器查詢獲得結果，為支援時空資料的互動式查詢，作者採用一種高效的資料存儲模式，構建了一個基於 k-d 樹的特殊索引。有了查詢結果後，系統會立刻將結果顯示在地圖上。用戶可以依照這個查詢結果不斷的來優化查詢。由於查詢到的結果可能會很大，因此作者還利用了自我調整的細節層次 (LOD) 和密度熱力圖來視覺化地展示結果。具體的系統結構如下所示：



研究結果：

該文構建了一個視覺化系統 TaxiVis，系統支援多種對話模式，說明使用者對資料的所有維度進行查詢，更加便捷的探索與計程車行程相關的屬性。作者通過一系列案例研究驗證了系統的有效性，如比較了不同區域上下客流量隨時間的變化，交通樞紐地區的交通流量變化，以及颶風 Sandy 和 Irene 對紐約市計程車系統的影響。案例表示了作者提出的系統確實可以說明行業專家完成之間無法實現的研究任務，通過作者設計的視覺化系統，專家們能更高效的挖掘出更多的有效資訊。

論文題目： *UpSet: Visualization of Intersecting Sets*

中文題目：UpSet: 相交集的視覺化

論文作者：Alexander Lex, Nils Gehlenborg, Hendrik Strobel, Romain Vuillemot, and Hanspeter Pfister

論文出處：IEEE Transactions on Visualization and Computer Graphics (Volume: 20, Issue: 12, Dec. 31 2014)

論文地址：http://sci.utah.edu/~vdl/papers/2014_infovis_upset.pdf

研究問題：

瞭解集合之間的關係是一項重要的資料分析任務。如果集合的數量超過一定的閾值，集合交叉點的數量就會爆炸性增長。所以對於數量多的集合資料來說，對其進行分析和視覺化是具有挑戰性的。為了解決這個問題，該文提出了UpSet——一種新穎的視覺化技術，用於對集合及其交集進行定量分析。

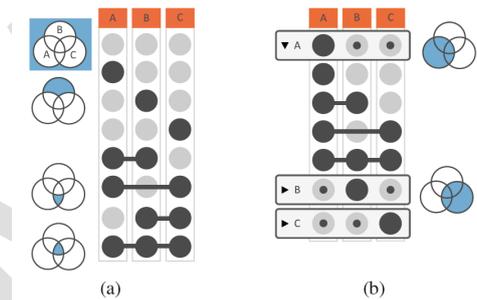
研究方法：

UpSet 專注於創建任務驅動的聚合，研究集合之間在大小，所包含元素及其相關屬性方面的相互關係。UpSet 引入矩陣佈局，可有效表示關聯資料，例如交集中元素的數量，以及從子集或元素屬性派生的其他摘要統計資訊。根據各種度量進行分類，可以對相關的交叉集合進行任務驅動的分析。集合中表示的元素及其關聯的屬性可以在單獨的視圖中視覺化。

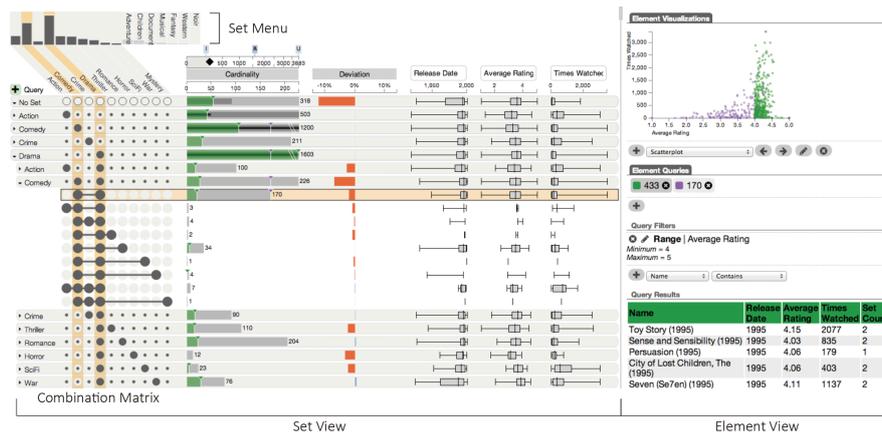
UpSet 採用“分而治之”的概念方法，將 k 個集合的資料集劃分為所有可能的 2^k 個交叉點。這些交點對應於卡氏圖表的原子區域。將這些基本構建塊被稱為互

斥交集。使用專有交集作為基本構建模組，使分析人員能夠創建聚合。聚合是使用任務驅動方法定義的互斥交集的集合。交叉集可以根據某種聚合語義進行聚合，也可以通過查詢進行聚合。

UpSet 將集合的交集繪製為矩陣，如下圖所示。每一列對應於一組，並且每一行對應於卡氏圖表中的一個分段。若儲存格為空 (淺灰色圓圈)，表示此集合不是該交集的一部分；若儲存格為已填充 (黑色圓圈)，表示該集合是交集的一部分。帶有黑點的淺灰色圓圈表示這些集合可能是互斥交集的一部分。如下圖(a) 中的第一行完全為空，說明它不是交集的一部分，第二行對應于僅在集合 A 中的元素 (且不在 B 或 C 中)。



下圖為使用 UpSet 對電影資料集進行關於集合分析的示例，該圖清楚的顯示了電影流派的關係。設置視圖顯示了集合間的相交關係及其聚合，元素數量和屬性統計資訊。元素視圖顯示了過濾後的元素和一個散點圖，比較了兩組過濾後的元素。



研究結果：

該文提出一種關於集合的視覺化技術 UpSet 並進行了用例研究，通過採訪來自各個領域(宏觀經濟學、遺傳學、藥理學和社會網路分析)的多位元研究人員，總結他們在研究中遇到的各類與集合相關的分析任務，並且讓每位研究人員都利用 UpSet 都對他們提供的資料集進行了分析。最終實驗結果表明，UpSet 可以有效解決幾位研究人員所描述的 26 個與集合相關的任務中的 23 個。由研究結果以及專家評論可以得出，UpSet 這一技術對於分析集合資料來說是一種很有用的工具，同時在不同領域的多個用例研究結果展示了其通用性。

論文題目： *Towards Better Analysis of Deep Convolutional Neural Networks*

中文題目：針對深度卷積神經網路的進一步分析

論文作者：Mengchen Liu, Jiaxin Shi, Zhen Li, Chongxuan Li, Jun Zhu, Shixia Liu

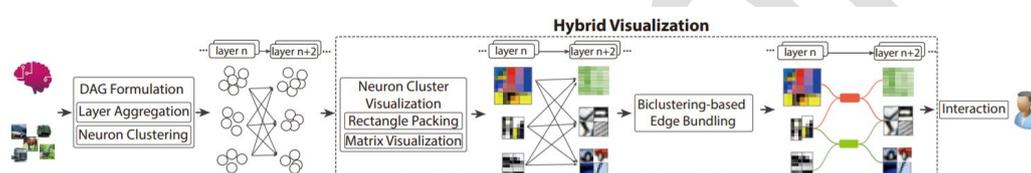
論文出處：IEEE transactions on visualization and computer graphics, 2016

論文地址：<https://arxiv.org/pdf/1604.07043.pdf>

研究問題：

深卷積神經網路 (CNNs) 在圖像分類等模式識別任務中取得了突破性的性能，然而開發高品質的深度學習模型通常依賴于大量的嘗試和錯誤。因為對於深度學習模型如何學習以及為什麼有效仍然沒有明確的解釋。為了幫助專家們更好地探索和深入理解 CNN，並且分析每個神經元的作用和神經元之間的聯繫，該文提出了一個新的互動式視覺化系統 CNNVis，說明專家更好地理解、診斷和提煉深度 CNN。

研究方法：



CNNVis 是一個互動式的視覺化分析系統，旨在說明機器學習專家更好地理解、診斷和完善 CNN 模型。基於深度 CNN 的特點，CNNVis 將其表示為有向無環圖 (DAG)，其中每個節點表示了一個神經元，每條邊代表一對神經元之間的連接。若需要視覺化大型 CNN，CNNVis 首先對網路中的層進行聚類，然後從每個層聚類中選擇一個具有代表性的層，然後 CNNVis 在每個代表層中對神經元進行聚類，並從每個神經元聚類中選擇幾個代表性的神經元。在 DAG 表示的基礎上，CNNVis 開發了一種混合視覺化方法，通過顯示不同類型圖像中神經元的作用來揭示神經元與每個神經元之間的多種相互作用關係。需要特別指出的是，CNNVis 提出了一種分層矩形排列演算法來顯示神經元簇的派生特徵。在 hold-Karp 演算法 (state compression dynamic programming，狀態壓縮動態規劃) 的基礎上，CNNVis 設計了一種矩陣重排序演算法，演示了每個神經元素啟動時的簇模式。這裡，啟動值是神經元的輸出值。這個值由將神經元輸入值轉換為神經元

輸出值的啟動函數決定。此外，CNNVis 還提出了一種雙聚類邊緣捆綁方法來減少神經元之間大量連線所造成的視覺混亂。

研究結果：

通過三個案例研究，該文評估了 CNNVis 的高效性與有用性。三個案例研究包括：利用 CNNVis 分析模型結構對於模型表現的影響，利用 CNNVis 診斷一個失敗的模型訓練過程和利用 CNNVis 提升模型的表現。研究結果證實 CNNVis 可以說明專家診斷模型結構的潛在問題，並對 CNN 進行優化，從而使模型構建過程的反覆運算速度更快，收斂速度更快。

論文題目： *Visualizing Dataflow Graphs of Deep Learning Models in TensorFlow*

中文題目：在 TensorFlow 中視覺化深度學習模型的資料流程圖

論文作者：Kanit Wongsuphasawat, Daniel Smilkov, James Wexler, Jimbo Wilson, Dandelion Mané, Doug Fritz, Dilip Krishnan, Fernanda B. Viégas, and Martin Wattenberg

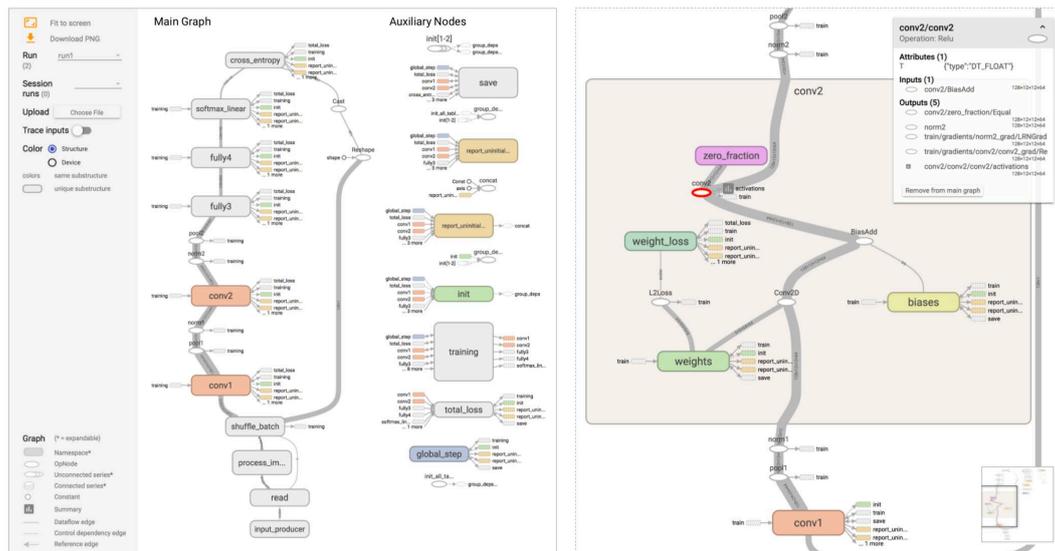
論文出處：IEEE Transactions on Visualization and Computer Graphics (Volume: 24, Issue: 1, Jan. 2018)

論文地址：<https://idl.cs.washington.edu/files/2018-TensorFlowGraph-VAST.pdf>

研究問題：

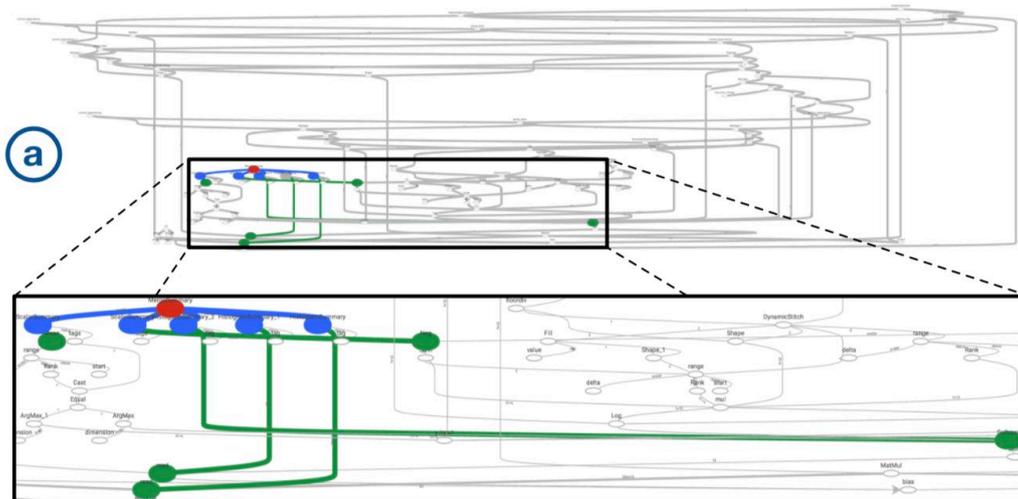
近年來，機器學習取得了一系列突破，其中，深度學習最為突出。深度學習的最大特點是其多層次的計算網路，而這些網路的複雜性導致其實現上的困難，所以 TensorFlow、Theano 和 PyTorch 等深度學習平臺提供了高級的 API 來降低這些困難，開發人員可以利用這些 API 生成資料流程圖，以支援各種演算法模型

和分散式運算。然而，深度學習的網路結構通常都很複雜，開發人員憑藉著自己的記憶或代碼本身，很難對演算法各個部分進行調試或者互相溝通。因此，該文提出了一種基於 TensorFlow 的視覺化工具 TensorFlow Graph Visualizer，旨在幫助開發者在 TensorFlow 中進行演算法的分析與開發。



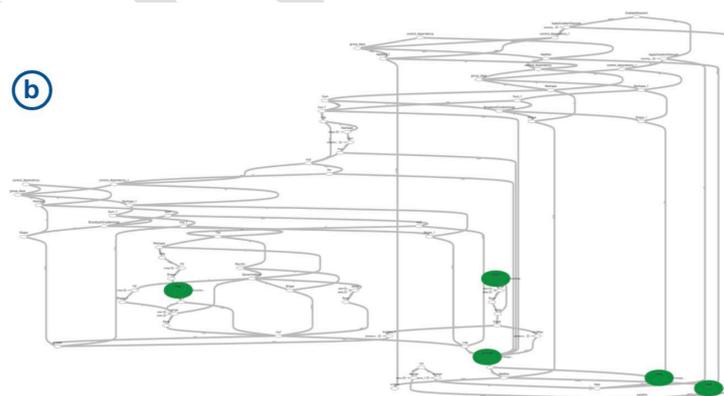
研究方法：

如下圖所示，該文首先直接利用 Sugiyama 流佈局對 TensorFlow 的資料流程圖進行佈局。由於資料流程圖裡的結構複雜，這種佈局容易出現交錯，並且缺乏層次分佈。



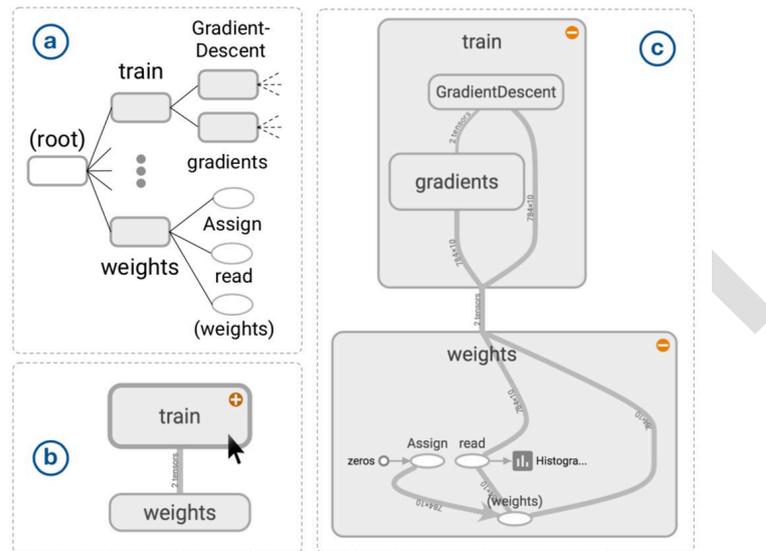
根據這種佈局方式的缺點，文章提出了三點優化：

提取非重要節點。該文作者將常量節點和統計節點提取出來作為非重要節點。其原因在於，常量節點通常是某個操作的輸入，只有出度，沒有入度；而統計節點需要利用到負責日誌記錄的“仲介節點”，對實際的資料流程操作沒有影響且容易破壞佈局。將這些節點提取出來，使用小的圖示進行展示，對佈局效果有很大改善（見下圖）。

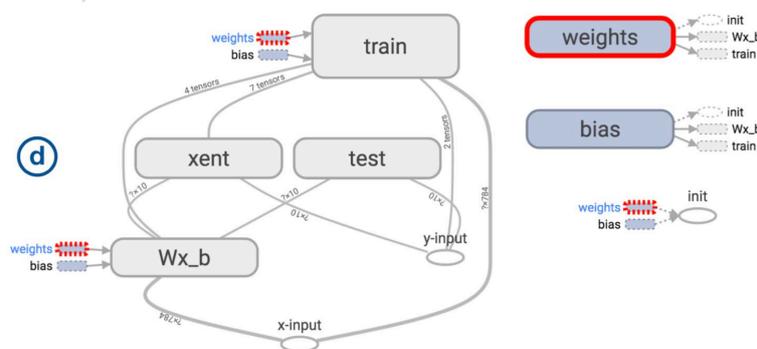


建立層次關係。資料流程本身是存在層次結構的，多種簡單操作可能組合為一個複雜操作。因此，該文作者利用各個操作的命名空間這種層級關係，對佈局進行優化：只顯示命名空間，將底層的操作隱藏，並進行邊綁定，直到用戶對負

節點進行點擊操作才展開底層的資料操作。圓角矩形代表一組操作，節點的高度代表其包含的運算元目。對於內部結構相同的節點使用相同的顏色著色。節點間的連邊根據關係有所不同：實線代表資料傳輸，虛線代表存在演算法依賴關係，而連邊的粗細則代表資料流程的大小。



提取輔助節點。對於當前佈局，圖中仍存在許多度數高的節點，造成邊纏繞，降低了圖的可讀性。同時，深度學習專家回饋目前圖上存在著一些分析過程中不被重視的節點，比如常量或日記記錄。因此，作者將這些點提取出來作為輔助節點，將其畫在資料流程圖之外，然後在資料流程圖中利用小的代理圖示來表示輔助節點與其他節點的相連關係，簡化圖的結構（見下圖）。



隨後，該文作者通過對資料流程圖結構進行分析，從組節點中提取範本和檢查子圖間的相似性，並且加入額外的定量資料，來說明演算法開發人員理解和分析演算法。

研究結果：

該文利用經過優化後的 TensorFlow Graph Visualizer 分別展示卷積神經網路和 Inception 網路架構，展示了 TensorFlow Graph Visualizer 的應用場景，並通過用戶回饋證明了 TensorFlow Graph Visualizer 對於理解、調試和共用模型的結構的有效性。

論文題目： *ACTIVIS: Visual Exploration of Industry-Scale Deep Neural Network Models*

中文題目：ACTIVIS: 工業規模深層神經網路模型的視覺化探索

論文作者：Minsuk Kahng, Pierre Y. Andrews, Aditya Kalro, and Duen Horng (Polo) Chau.

論文出處：IEEE Transactions on Visualization and Computer Graphics (Volume: 24, Issue: 1, Jan. 2018)

論文地址：<https://arxiv.org/pdf/1704.01942.pdf>

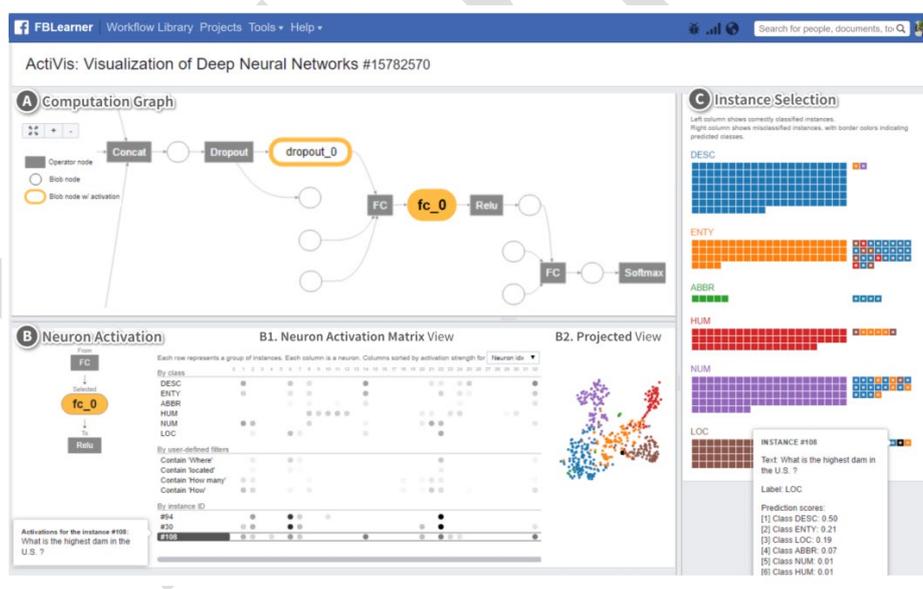
研究問題：

深度學習是機器學習的子域，它應用非監督式或監督式的特徵學習和分層特徵提取的方法來代替傳統的手工特徵提取，是一種十分高效的演算法。雖然深度學習模型已經在許多預測任務中實現了一定的精準度，但是理解這些模型仍然是一個挑戰。近年來，研究人員嘗試著開發視覺化工具來說明人們理解深度學習模

型。然而，由於工業界模型的複雜性和多樣性，以及他們所使用的大規模資料集，導致現有的工作不能充分解決這些問題。因此提出了一個用於解釋大規模深度學習模型和結果的互動式視覺化系統 (ACTIVIS)。它集成了包括模型架構視圖、神經元啟動矩陣視圖以及實例選擇視圖在內的多個子視圖。

研究方法：

通過與深度學習領域從業者 (Facebook 員工) 進行訪談，凝練出三個主要的設計目標：比較不同實例下神經元啟動情況；對模型架構和神經元活動情況進行低級別檢查；支援工業界模型和資料集的導入分析。根據這些設計目標，對系統做出了如下設計和開發：



如上圖所示，ACTIVIS 系統主要由三個部分組成，包括模型架構視圖、神經元啟動矩陣視圖和投影視圖。模型架構視圖：對整個深度學習網路結構的概覽，矩形節點代表操作函數，圓形節點代表一個張量；神經元啟動矩陣視圖：左圖中的每一列代表一個實例下神經元的啟動情況，用圓形的顏色深淺來編碼其啟動情況，顏色越深表示啟動值越大；右圖為投影視圖，展示實例的分類情況；實例選

擇視圖：每一行代表一個實例的分類情況，左半部分代表被正確分類的實例，右半部分代表被錯誤分類的實例，右半部分矩形顏色代表其被錯誤分類的類別。

研究結果：

使用者在使用該系統進行神經網路模型探索時，通過概覽和探索整個深度神經網路的結構，啟動矩陣視圖和投影視圖觀察分類情況，以及實例選擇圖中的具體情況對模型進行調整，可以在較大範圍內提高開發效率。

論文題目： *FlowSense: A Natural Language Interface for Visual Data Exploration within a Dataflow System*

中文題目：FlowSense：用於可視資料的自然語言介面資料流程系統內的探索

論文作者：Bowen Yu, Claudio Silva.

論文出處：EEE Transactions on Visualization and Computer Graphics

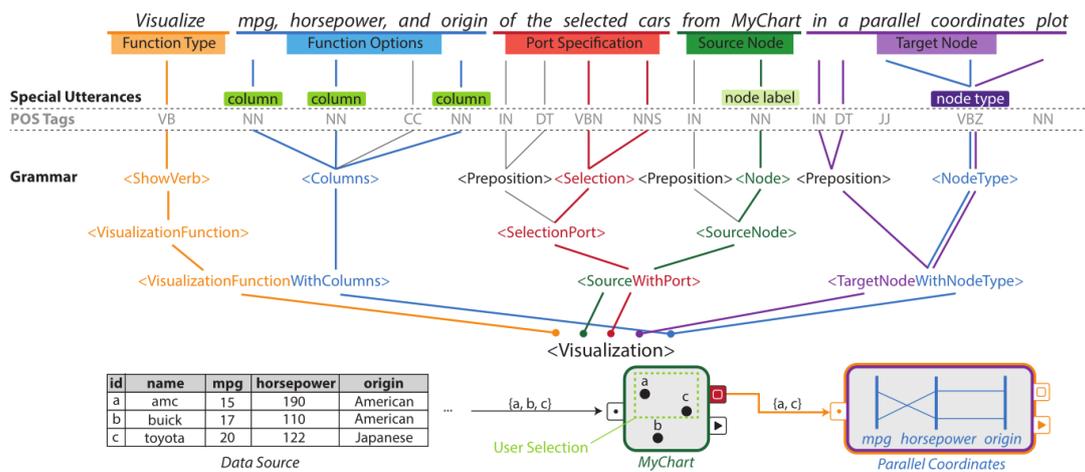
論文地址：<https://arxiv.org/pdf/1908.00681.pdf>

研究問題：

資料流程圖 (DFD) 是描述系統中資料流動過程的分析工具，它以圖形元素和流動指向來構建資料變化的內部邏輯。與傳統的流程圖或框圖不同，它是從資料的角度描述一個系統。這也就意味著它更抽象，不便於理解和學習。另一方面，自然語言作為人類行為學認知的的基本單元，往往是最便於理解的。因此，該文提出了使用一個帶有特殊文本標記和預留位置的語義分析器 (FlowSense) 來概括不同的資料集和資料流程圖。FlowSense 試圖將自然語言作為系統的切入口，經過自然語言處理轉化為電腦語言，輔助使用者完成資料流程圖的構建。

研究方法：

首先通過研究示例資料流程圖集構造 VisFlow 函數，該函數支援資料流程圖的自動生成。FlowSense 從使用者輸入的自然語言中提取一組特殊的標記，該標記是指表示為資料集或資料流程圖中的實體的單詞（包括表列名、節點標籤、節點類型和資料集名稱等），它們可作為 VisFlow 函數的參數和運算元。於是，FlowSense 再通過執行 VisFlow 函數生成資料流程圖。具體實現流程如下圖所示：



使用者輸入一段文字。系統會捕獲到其中的特殊文本，然後利用自然語言處理技術將其轉化為 VisFlow 函數可以識別的語法，即該函數的參數或運算元，最後繪製資料流程圖。

當用戶展開資料流程關係圖時，總是存在一個編輯焦點，儘管該焦點通常是隱式的。FlowSense 根據使用者的交互行為判斷焦點，通過以下方法計算每個節點 X 的焦點得分：

$$score(X) = activeness(X, t) + \alpha \left(1 - \frac{1}{1 + e^{-(distanceToMouse(X)/\gamma - \beta)}} \right)$$

X 的活躍度在每次使用者在系統中點擊時重新反覆運算：

$$activeness(X,t) = activeness(X,t-1)/2 + click(X,t)$$

其中, $click(X,t) = 1$ 代表用戶在 t 時刻對 X 節點進行了點擊操作。FlowSense 選擇焦點得分最高的節點作為圖表編輯焦點。如果需要多個源節點 (例如, 在合併查詢中) , 那麼 FlowSense 將按節點的焦點得分遞減順序選擇節點。

研究結果 :

該文結合 VisFlow 框架進行了使用者研究來評估 FlowSense 系統的有效性。結果表明, 用戶通過查詢得到有效資料流程圖的概率為 68.455%。後續研究通過研究人員對系統的改進, 將有效率提高到 76.911%。這說明系統的整體運行情況良好。

論文題目 : *Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco*

中文題目 : 以形式化視覺化設計知識為約束 : Draco 系統中可操作和擴展的模型

論文作者 : Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Adam M. Smith, Bill Howe, Jeffrey Heer

論文出處 : IEEE Transactions on Visualization and Computer Graphics (Volume: 25, Issue: 1, Jan. 2019)

論文地址 : <https://adamsmith.as/papers/08440847.pdf>

研究問題 :

在視覺化設計中, 程式師往往根據經驗對視覺化圖表進行設計、編寫, 他們無法正確判斷自己編寫的圖表是否符合設計標準, 導致最後的效果不盡如人意。

雖然經驗研究可以提供設計指導，但人們缺少一個系統性的判斷體系和框架來表示設計標準、整合研究結果，從而促進有效的視覺化編碼和視覺化探索。該文將視覺化設計標準作為約束集，並根據實驗對約束集中各個子項進行權重分配，生成符合設計標準的視覺化。此外，該文還實現了一個基於答案集程式設計(ASP)的約束系統，旨在實用設計規範說明使用者製作好的視覺化設計。

研究方法：

該文將視覺化建模為事實集合，並解釋了該模型的設計空間以及如何使用約束查詢模型。約束求解器可以有效地在定義的空間內搜索最優的視覺化規範。系統採用列舉生成和測試的方法，將設計標準和演算法結合起來，生成最有效的設計。

在求解最優視覺化的過程中，使用了硬性視覺化約束規範和軟性視覺化規範。

硬性視覺化規範包括一些視覺化圖表內部的邏輯(比方說折線圖無法表示種類等等)，還有許多用戶自己定義的偏好，這些規則必須被滿足。比如 $Mark \in \{\text{bar, line, area, point}\}$ 、 $Channel \in \{\text{x, y, color, text, shape}\}$ 、 $Field \in \{\text{site, year, age}\}$ 、 $Type \in \{\text{categorical, continuous}\}$ 、 $Aggregate \in \{\text{sum, mean, count, median}\}$ 以及 $Zero \in \{\text{yes, no}\}$ 等等。

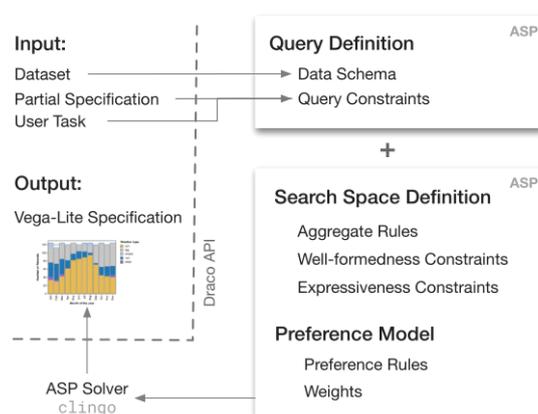
假設有 m 個軟性約束， p_i 為第 i 個約束，每個約束的懲罰為 w_i ，令 $S = \{(p_1, w_1) \dots (p_m, w_m)\}$ ， $n_{p_i}(v)$ 是 v 這種視圖違反軟約束 p_i 的次數，那麼 $Cost$ 可以定義為：

$$Cost(v) = \sum_{i=1..k} w_i \cdot n_{p_i}(v)$$

這樣求解器就可以通過這些權值得到不同視覺化圖表的偏好。軟性視覺化規範是視覺化的結構特徵，它捕捉了視覺化屬性之間隱藏的關係。

使用以上約束實現了最優的編碼搜索過程。Draco 將使用者查詢（包括資料集、部分規範和任務）編譯成一組規則，並將它們與現有的知識庫組合起來形成 ASP（回答集程式設計）程式。然後，Draco 調用 Clingo 來解決程式，以獲得最佳答案集。最後，Draco 將答案集轉換為 Vega-Lite 規範（Vega-Lite 是一種資料視覺化的高級語法，能夠快速定義一些基本的互動式資料視覺化）。

如下圖所示：



研究結果：

該文將 Draco 應用到三個不同的場景以測試它的表達性、可拓展性和可用性。結果表明使用約束程式設計，使得自動化視覺化設計工具的開發和維護更加容易。它還可以結合來自不同研究領域的資料學習權重，進一步加速建模工作。

論文題目：*InSituNet: Deep Image Synthesis for Parameter Space Exploration of Ensemble Simulations*

中文題目：InSituNet:用於集成模擬的參數空間探索的深度圖像合成

論文作者：Wenbin He, Junpeng Wang, Hanqi Guo, Ko-Chih Wang, Han-Wei Shen, Mukund Raj, Youssef S. G. Nashed, Tom Peterka

論文出處：IEEE Transactions on Visualization and Computer Graphics (SciVis 2019)

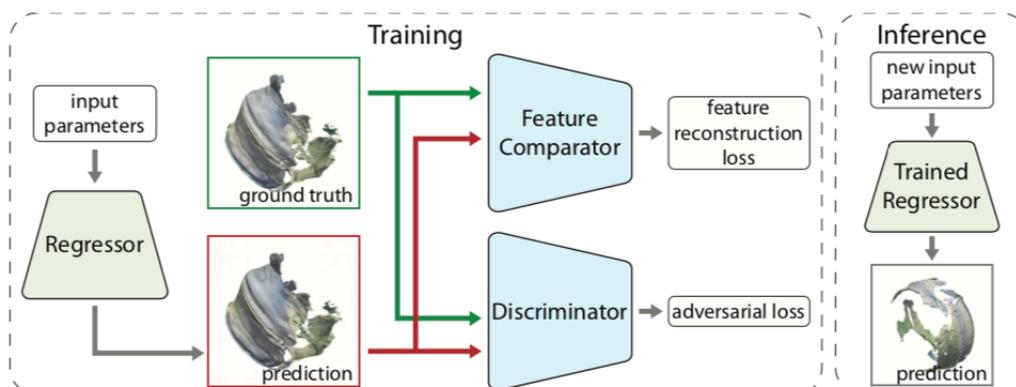
論文地址：<https://arxiv.org/pdf/1908.00407.pdf>

研究問題：

集成模擬在各種科學和工程學科中發揮著重要作用，如計算流體動力學、宇宙學和天氣研究。由於 I/O 和存儲方面的限制，在處理大規模模擬時，通過模擬生成原位視覺化效果的做法越來越普遍。然而，由於無法獲得原始的類比資料，原位視覺化方法限制了事後分析的靈活性。雖然現有研究已經提出了多種基於圖像的方法來解決這一問題，但這些方法仍然缺乏探索模擬參數的能力。因此，該文提出了一種基於深度學習的方法——InSituNet，來實現在大規模集成模擬中靈活地探索參數空間。

研究方法：

通過現有的集成模擬，利用不同的模擬參數、視覺映射參數和視圖參數生成出對應的視覺化圖像。這些參數和相應的視覺化圖像構成一組組資料對，這些資料對將被存儲並用於訓練 InSituNet。



InSituNet 由三部分組成，分別為回歸器、特徵比較器和鑒別器。

回歸器 R_o 是一個深度神經網路模型，將輸入參數映射到視覺化圖像。首先，三種參數分別輸入到三組全連接層中，然後將輸出連接到另一組全連接層中，將它們編碼成一個隱向量。隨後，潛向量被重新構造為低解析度圖像，通過 Residual blocks 進行 2D 卷積和上採樣，將其映射為高解析度輸出圖像。

隨後，利用特徵比較器和鑒別器對原始圖像和預測圖像進行對比。特徵比較器是一個預訓練的 VGG-19 模型，用於提取和比較預測圖像和真實圖像之間的圖像特徵（如邊緣、形狀），從而獲得特徵重構損失。而鑒別器是一個深度神經網路，其權值在訓練過程中不斷更新，來估計預測的分佈與真實圖像之間的差異，稱為對抗損失。二者一起構成了 InSituNet 的損失函數。

利用訓練後的 InSituNet，用戶可以在不同的模擬與視覺化參數設置下生成不同模擬圖像，而不需通過真正渲染得到模擬圖像，從而能夠快速對集成模擬和參數設置進行深入分析。

研究結果：

通過定量和定性的評估，該文作者證明了 InSituNet 對燃燒、宇宙學和海洋三種模擬模擬的有效性，且生成結果優於現有的演算法模型。

12.5 視覺化進展

- 視覺化語法及工具

隨著大資料時代的到來，視覺化已經成為一個必不可少的工具。現有的視覺化軟體及工具，可用于設計視覺化應用程式和構建視覺化分析系統，有助於視覺

化的廣泛使用。為了減少製作視覺化的技術負擔，一些視覺化工具提供了聲明性語法，其中包括了 Vega-Lite 和 P5。聲明性語法可以將視覺化設計與執行細節分離，這使分析人員可以專注于特定于應用程式的設計決策。同時 Vega-Lite 和 P5 都提供了易於使用的程式設計介面。Vega-Lite 是一套能夠快速構建互動式視覺化的高階語法，它是基於 Vega 和 D3 等底層視覺化語法的上層封裝。相比於其它比較底層視覺化語法，Vega-Lite 可以通過幾行 JSON 配置代碼即可完成一些通用的圖表創建，而相反地，想要用 D3 等去構建一個基礎的統計圖表則可能需要編寫多行代碼，如果涉及到交互的話代碼量更是會大大增加。P5 是一個基於 Web 的視覺化工具包，它能集成了 GPU 計算與漸進式處理，並且提供了帶有聲明性語法的 API，可用於指定漸進式資料轉換和視覺化操作，從而幫助分析人員構建融合了高性能計算和漸進式分析工作流的視覺化系統^[68-69]。

- 視覺化與故事敘述

故事敘述是視覺化研究的一個重要且新興的方向。與傳統的、強調資料分析的視覺化思路不同，故事敘述強調資料的傳達與溝通，強調資料與人（且通常是普通人）的連結。在此思路下，研究者們致力於探究：何種資料呈現與講述技巧，可以使資料具備吸引力、記憶度；資料故事的創作流程是怎樣的，存在哪些需求和痛點；以及如何自動生成資料故事等等。用講故事的方式來呈現視覺化，本質上是體現了一種人本導向，即以人的需求出發，提取和分析資料，並以對人友好的方式，將資料中的資訊傳達出去。隨著我們的社會越來越依賴資料賦能，更好地構建資料與人的關係，將成為一條必經之路。敘述視覺化的應用，不僅在於那些以“敘述”為主業的領域，如新聞媒體、廣告宣傳，更在於需要用資料來影響人、

說服人、打動人的各行各業。對於研究者來說，相關的研究方向則包括視覺化設計、人機交互、認知與感知、智慧生成與推薦等等^[70-71]。

- 視覺化的自動生成

資料視覺化領域中大多數的視覺化生成系統往往是基於資料的互動式探索，也包括商業領域的知名的視覺化工具 Tableau 和 PowerBI。而近些年來，為了避免繁雜的資料分析步驟並提升使用者效率，視覺化的自動生成逐漸成為行業領域中的研究熱點。一系列基於規則和機器學習的推薦方法層出不窮，在自動生成視覺化的最新研究中，研究者希望在保證準確表現資料的同時，也能將視覺設計的因素考慮在內，確保視覺化的美觀性和資料的表現力。例如，DataShot 和 Text-to-Viz，分別從資料和自然語言兩個角度去自動生成富有設計感的資料視覺化，前者直接從表格資料生成資訊簡報，後者根據使用者的自然語言輸入生成對應的資訊圖。製作一個有效且美觀的資料視覺化往往需要跨專業領域的技能，尤其是需要同時具備資料分析能力和平面設計能力，而這對於一個沒有專業訓練的普通用戶來說是比較困難的。DataShot 和 Text-to-Viz 等前沿的技術研究均通過自動化的方法從資料洞察和設計美學兩個方面幫助用戶生成視覺化，降低用戶製作視覺化的門檻，並有效提高生產效率^[72-73]。

可解釋性深度學習

LSTMVis 是一個遞迴神經網路的視覺化分析工具，它著重於對 RNNs 中的隱藏特徵進行視覺化分析。LSTMVis 結合了一個基於時間序列的選擇介面和一個互動式的匹配工具來搜索大型資料集中相似的隱藏狀態模式。系統的主要功能是理解模型中動態變化的隱藏狀態。該系統允許使用者選擇一個假設的輸入範圍

來關注局部的改變，將這些狀態改變與大型資料集中類似的模式進行匹配，並將這些選擇出來的模式進行對齊分析。RNNs 在序列建模方面有著重要的作用，但是模型中的隱藏層含義很難被解釋清楚。對於一個完成訓練的 RNN 模型，分析人員並不清楚這個模型是如何理解序列中不同節點之間的關係的。LSTMVis 能夠說明使用者互動式地探索 RNN 模型複雜的網路結構，並將模型中抽象表示的隱藏層資訊與人類可理解的原始輸入進行關聯^[74]。

12.6 視覺化應用

隨著 21 世紀大資料的興起和發展，大資料視覺化廣泛應用於各個領域，本節重點介紹其中的社交媒體視覺化、醫療資訊視覺化和體育資料視覺化。

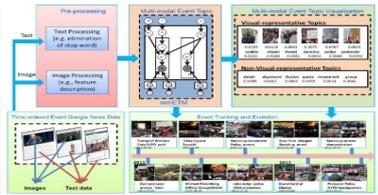
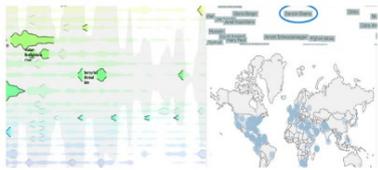
12.6.1 社交媒體視覺化

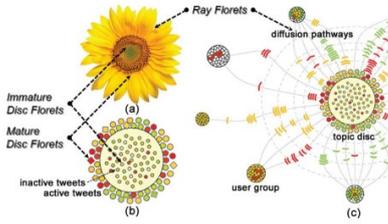
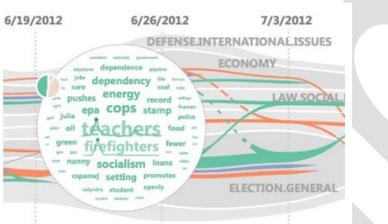
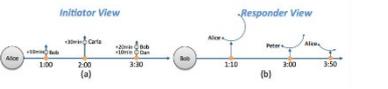
社交媒體，比如最近幾年非常流行的 Twitter、Facebook、微博。它們可以作為強大的線上交流平臺，允許數百萬用戶在任何時間、任何地點產生、傳播、共用或交換資訊。這些資訊通常包括多種多媒體內容，如文本、圖像和視頻。在社交媒體上傳播的大量多媒體資料，涵蓋了全球範圍內大規模和即時發生的社會動態資訊，這種現象為社交媒體視覺化提供了很多機會。

社交媒體技術層面上的視覺化，主要包括：基於關鍵字方法的視覺化，基於主題方法的視覺化和多元方法的視覺化。現有的研究大多集中于集體行為的視覺化，這類研究的主題包括：資訊擴散的視覺化，社會競爭與合作的視覺化，人的流動性的視覺化。

社交媒體資料的視覺化分析正在迅速發展，每年都有大量的新方法出現。然而，該領域仍處於起步階段，面臨許多挑戰和懸而未決的問題。許多挑戰不能僅使用來自一個規程的技術來解決。但是，將視覺化、多媒體、NLP 和人機交互相結合的多學科研究，將帶來處理和理解社交媒體資料會有更強大、更可行的方法和技術。具體社交媒體視覺化如下表所示：

表 12-2 社交媒體視覺化介紹

視覺化方法	圖例	特點
基於關鍵字方法的視覺化		<p>這種方法，用於跟蹤和探索社交媒體上大型活動的線上對話。提供了從時間、主題、社交和圖像方面對會話的視覺化總結。可以靈活使用時間和空間篩檢程式，從社交媒體上海量的資訊中獲取自己想要的資訊。</p>
基於主題方法的視覺化		<p>主題方法使用基於三個標準(視覺相關性、視覺連貫性和獨特性)確定的代表性圖像進行視覺化。一個多模態框架，用於從多模態資訊中檢測主題，跟蹤主題的演變，並隨時間的變化使用文本和圖像視覺化主題。</p>
多元方法的視覺化		<p>多元方法幫助用戶從多個角度獲取關於社交媒體事件的資訊。與其他主要側重於理解社交媒體消息文本內容的方法相比，多元方法通過結合高級資料採擷方法來保持態勢</p>

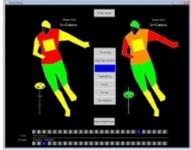
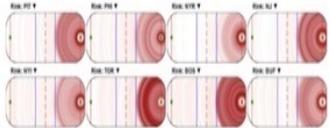
		感知，並提供全面的概述。
<p>資訊擴散的視 化</p>		<p>在資訊擴散的視覺化方法上，有不同的模型來描述資訊在社交媒體上的傳播，比如綜合的多源驅動的非同步擴散模型。往往直觀地呈現出社會化媒體傳播過程的三個主要特徵，即時間趨勢、社會空間範圍和社區對某一興趣話題的反應。</p>
<p>社會競爭與合作 的視覺化</p>		<p>為了便於視覺化分析，引入流式視覺化，用以說明每個主題的競爭力隨時間的變化。並展示出促進或降低主題傳播行為被視覺化為執行緒，這表明他們的角色隨著時間的推移而變化。從不同的角度展示社交媒體用戶群體之間的合作與競爭行為。</p>
<p>人的流動性的視 覺化</p>		<p>用戶可以在此基礎上探索人流動性的語義，如流動方式、頻繁訪問序列、關鍵字描述等。採用啟發式模型來減少資料的不確定性，從而增加對可靠資料的適當選擇。</p>

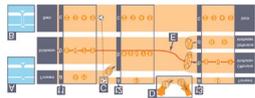
12.6.2 體育資料視覺化

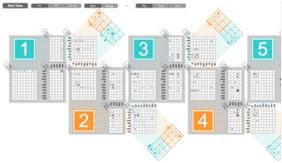
視覺化技術在體育行業的運用越來越普遍。例如觀眾通過電視觀看體育賽事時，經常能看到關於比賽情況的解讀。以前使用技術資料統計與空間資料的方法來演示比賽實況，但使用者並不能從視覺化結果中看出比賽的發展趨勢。現在，

將技術資料統計和時間資料相結合，就能很好地反映比賽過程。基於對整場比賽或者比賽中某個時間段的賽況的視覺化，大大說明了使用者分析比賽進程的動態變化。體育視覺化使用的時間概念已經對時間做了一定程度的簡化，它將客觀事件發生的時間作為規律的時間單元進行處理，這種視覺化方法能夠反映比賽的趨勢，很大程度上增加了用戶的體驗效果。從資料角度出發，可以將體育資料視覺化分為以下幾類，如下表所示：

表 12-3 體育資料視覺化

視覺化方法	圖例	特點
<p>技術統計資料的視覺化方法</p>		<p>根據統計資料的不同，具體的視覺化方法千差萬別。通常會根據資料的屬性選擇較好的呈現方式，並輔之以良好的交互手段。</p>
<p>技術統計資料和空間資料結合的視覺化方法</p>		<p>大部分體育比賽都是在一定規模的場地中進行，球員位置和事件產生位置等都是至關重要的資料。因此僅對技術統計視覺化可能會遺漏掉重要的資訊。技術統計資料和空間資料結合的視覺化方法能夠對比賽場地進行等比例的簡化繪製，並且將技術統計資料繪製到對應的空間位置上。</p>

<p>技術統計資料和時間資料 結合的視覺化方法</p>		<p>如果採用技術統計資料與空間資料的方法展示比賽,用戶並不能從視覺化結果中看出比賽的發展趨勢。然而技術統計資料和時間資料相結合的方法就能很好地反映比賽過程。基於對整場比賽或者比賽中某個時間段的賽況的視覺化,有利於說明使用者分析比賽進程的動態變化。</p>
<p>技術統計資料和時空資料 視覺化方法</p>		<p>足球比賽中陣型是一個時空資料,隨比賽的進行在時間和空間維度上都會有很大的變化,球隊分析師很難通過觀看視頻和查看統計資料看出球隊整場比賽下來陣型的一個變化。使用流視圖對陣型時空資料進行視覺化能給出陣型在整場比賽的一個直觀的變化概覽。分析師通過查看概覽發現異常的地方後,還可以對流進行選定,進一步查看這一段時間內球隊的技術統計資料以及視頻重播,進行細緻的分析。</p>

<p>技術統計資料的關聯視 化方法</p>		<p>單靠技術統計資料的整體時序變化趨勢或者兩兩對比結果,很難找到比賽中運用的高層次的技戰術策略,而不同技術統計資料之間的關聯分析則可以說明使用者發現許多技術資料之間的關係,從而揭示出比賽中涉及的一些技戰術策略。尤其是對於像乒乓球這樣包含豐富戰術變化的運動來說,不同種類技術之間的關聯分析可以幫助分析運動員的打法風格和技術上的優勢與劣勢。</p>
<p>軌跡視覺化</p>		<p>對於體育資料分析而言,比賽的軌跡繪製不容置疑是非常重要的。在球類比賽中,球的軌跡或者運動員的軌跡都能夠反映出動態規律,通過對重要軌跡的分析可以掌握重要的比賽和球員線索。</p> <p>再現比賽:通常低精度的重現可以作為一種更生動的圖文直播;高精度的再現則可以做到對體育比賽的完全掌握和分析,它具有強大的分析能力。在一場冰球比賽中,這個系統可以即時地顯示球場上發生的事件,新的事件通過特定標誌表示,顯示在球場上的對應位置以</p>

		及下方的時間線上。
--	--	-----------

12.6.3 醫療資料視覺化

醫療健康領域是與每一個直接密切相關的重要科學領域。一直以來，科學家在探索生命奧秘，及疾病產生機理的過程中，一直重視對跨學科技術的運用。從基於虛擬實境技術的模擬手術到手術機器人，從醫學成像技術到醫學影像處理，從大資料分析到人工智慧，越來越多的新興科技正在被應用到醫療領域，每一次技術上的革新與成功應用，都給醫療領域帶來了全新診療技術，及科研手段，也提高了就診治療過程中患者的安全。

2010 年，醫療 2.0 伴隨著互聯網 2.0 的熱潮應運而生，旨在利用一切先進科學技術幫助提升診療手段，攻克醫學難題。2011 年，美國國家醫學院 (MOI) 發佈的該年度報告中特別指出，相對於其他技術在醫療領域的引用，資訊視覺化技術在醫療領域的應用顯得尤為滯後，現有技術無法滿足資訊展現、使用者交互、資料分析等眾多需求。近些年來，伴隨著大資料及人工智慧技術在醫療領域的應用與普及，資訊視覺化及相應的可視分析技術在醫療領域的應用也得到了長足的發展。

視覺化技術，尤其是科學視覺化技術在醫療領域長久以來一直扮演者重要角色。例如，無論是平面 X 光掃描，還是三維 CT 影像，都應用到了科學視覺化的相關技術。然而這些技術仍然局限於對於具象資料 (例如，人體的骨骼、器官組織結構等) 的展現。隨著互聯網的普及以及可穿戴設備的廣泛應用，越來越多的與醫療相關的抽象資料被採集了上來，因此對資訊視覺化技術提出新的需求，這

也正是 MOI 報告中所指出來的相關問題。針對這些新的資料與需求，在醫療 2.0 概念的範疇下，新的資訊視覺化技術被主要開發用來：

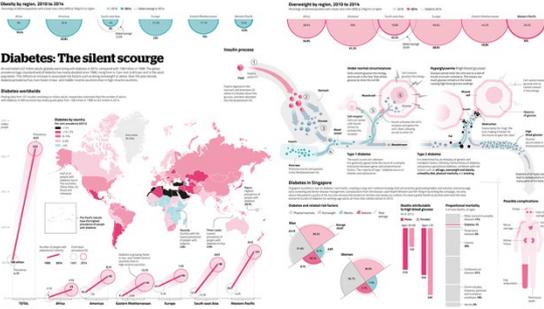
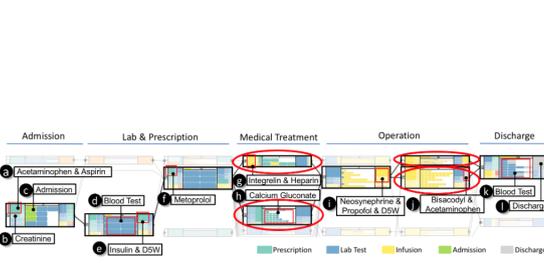
1) 展現使用者的個人健康資訊。例如，用於展現使用者心跳、血壓等狀態的視覺化使用者介面；

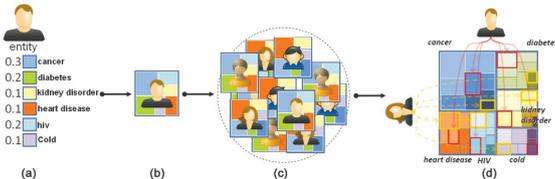
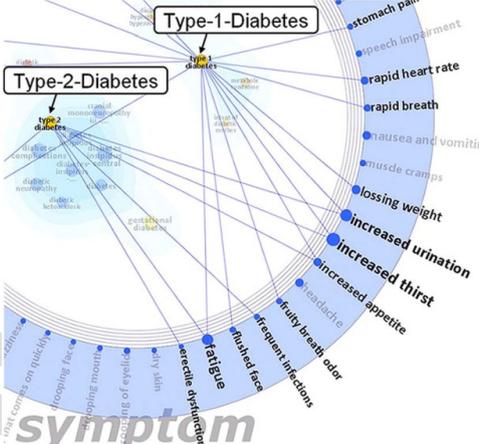
2) 匯總並展現公眾健康資訊。例如，用於匯總展現禽流感在擴散趨勢，或不同地區的人民健康狀況等統計資訊的視覺化大屏；

3) 分析並展現臨床電子病歷記錄中的規律與模式。例如，疾病的演變過程，以及診療方案的療效等。前兩類視覺化面向的一般是不具備醫療知識的普通用戶，因此，往往採用傳統直觀的資訊視覺化形式，如，柱狀統計圖，折線圖等，以便於用戶的理解與閱讀。第三類技術，主要面向的是醫生等具有專業背景，需要對資料進行深入調查，並做出職業判斷的用戶，因此這類視覺化更具有針對性，視覺化及相關分析技術的設計也更具挑戰性。

表 12-4 醫療資料視覺化

視覺化方法	案例	特點
醫療圖像資料的視覺化		採用針對標量場的視覺化技術，通過顏色映射、等值線和高度圖三類等方式展現資料在空間的分佈，常見於呈現 X 光及 CT 成像。

<p>針對個人統計健康資訊的視覺化方法</p>	 <p>The screenshot shows the PatientsLikeMe website for 'Bipolar Type II'. It features a navigation bar with 'Home', 'Patients', 'Treatments', 'Conditions & Symptoms', and 'Research'. The main content includes a description of the condition, a 'Who has this Condition?' section with a bar chart showing 11,879 patients and 485 new patients this month, and a 'Top Treatments' section with a horizontal bar chart listing drugs like Seroquel, Aripiprazole, and Topiramate. There is also a 'Patient Spotlight' for Rachel, a 57-year-old female from Lancaster, PA, and a 'Links' section for research updates.</p>	<p>採用簡單的以統計圖、視覺化圖示等簡單形式為主的視覺化技術，針對無技術背景的終端使用者，展現其個人健康資料。這類視覺化的設計旨在採用最直觀易懂的方式表達並展現如心跳、血壓、體溫等資訊。</p>
<p>針對公眾資訊的視覺化</p>	 <p>The infographic is titled 'Diabetes: The silent scourge'. It features a world map with color-coded regions indicating the prevalence of diabetes. To the right, there are detailed diagrams of the human body showing the effects of diabetes on various organs, including the heart, kidneys, and eyes. The text explains that diabetes is a chronic condition that can lead to serious complications if not managed properly.</p>	<p>採用資訊圖或視覺化大屏的直觀表達方式展現疾病的分佈、傳播等公眾健康資料。</p>
<p>針對電子病例中時間序列資料的視覺化方法</p>	 <p>The screenshot shows a medical dashboard with several time-series plots. On the left, there is a circular gauge chart. The main area contains multiple line graphs showing various physiological data points over time, such as heart rate, blood pressure, and brain waves. The plots are color-coded and labeled with patient IDs and specific data series.</p>	<p>採用時間序列視覺化的展現形式表達心電、血壓、腦電波等隨時間連續變化的資料。</p>
<p>針對電子病歷中事件序列資料的視覺化方法</p>	 <p>The visualization is a horizontal timeline showing a patient's medical history from admission to discharge. The timeline is divided into stages: Admission, Lab & Prescription, Medical Treatment, Operation, and Discharge. Various events are marked along the timeline, such as 'Admission', 'Blood Test', 'Metoprolol', 'Calcium Gluconate', 'Integrilin & Heparin', 'Nesynphrine 2', 'Propofol & D5W', 'Bisoprolol & Acetaminophen', 'Blood Test', and 'Discharge'. A legend at the bottom identifies the colors used for Prescription, Lab Test, Infusion, Admission, and Discharge.</p>	<p>匯總並展現電子病歷中所記錄的醫療事件發生的階段性過程，從而揭示疾病在不同人群中發展的規律，以及不同診</p>

		<p>療方案所帶來的不同療效。</p>
<p>針對電子病歷中病人個體及群體特徵的視覺化</p>	 <p>(a) Single patient icon. (b) Patient icon with associated conditions. (c) Group of patient icons. (d) Heatmap visualization of multiple patients with various conditions.</p>	<p>採用多維度數據視覺化技術，展現電子病歷資料中病人的個體及群體多維度特徵，主要用於群體分析 (Cohort Analysis) 以區分不同類型的病人。</p>
<p>醫療知識圖譜的視覺化</p>	 <p>The diagram shows 'Type-1-Diabetes' and 'Type-2-Diabetes' as central nodes connected to numerous symptoms including stomach pain, speech impairment, rapid heart rate, rapid breath, nausea and vomitir, muscle cramps, lossing weight, increased urination, increased thirst, headache, increased appetite, frequent infections, flushed face, fatigue, dry mouth, frequent urination, blurred vision, and dry skin.</p>	<p>採用圖的視覺化技術，展現大規模異構醫療知識圖譜，用於方便知識的檢索與查詢。</p>

13 資料採擷

13.1 資料採擷概念

資料採擷 (Data Mining) , 是指從大量的資料中自動搜索隱藏於其中的有著特殊關係性的資料和資訊, 並將其轉化為電腦可處理的結構化表示, 是知識發現的一個關鍵步驟。資料採擷的廣義觀點: 從資料庫中抽取隱含的、以前未知的、具有潛在應用價值的模型或規則等有用知識的複雜過程, 是一類深層次的資料分析方法。資料採擷是一門綜合的技術, 涉及統計學、資料庫技術和人工智慧技術的綜合, 它的最重要的價值在於用資料採擷技術改善預測模型。資料採擷涉及的常見的任務有^[75] :

資料表征: 是對目標類資料的一般特徵或特徵的總結。對應於使用者指定類的資料通常通過資料庫查詢收集。例如, 要研究上一年銷售額增長 10% 的軟體產品的特徵, 可以通過執行 SQL 查詢來收集與此類產品相關的資料。

異常檢測: 資料庫可能包含不符合資料一般行為或模型的資料物件, 這些資料物件即被成為異常值。大多數資料採擷方法將異常值視為雜訊或異常。但是, 在諸如欺詐檢測等應用中, 罕見事件可能比更常見的事件更有價值。異常值資料的分析被稱為異常值挖掘。

關聯規則學習: 搜索變數之間的關係。例如, 一個超市可能會收集顧客購買習慣的資料。運用關聯規則學習, 超市可以確定哪些產品經常一起買, 並利用這些資訊說明行銷。這有時被稱為市場購物籃分析。

聚類：是在未知數據的結構下，發現資料的類別與結構。聚類演算法基於最大化類內相似性和最小化類間相似性的原則對物件進行聚類或分組。也就是說，形成物件集群，使得集群內的物件彼此之間具有較高的相似性，但與其他集群中的物件非常不相似。每個形成的集群都可以被視為一類物件，從中可以派生出規則。聚類還可以促進分類的形成，也就是將觀察組織成一個將類似事件歸類在一起的類的層次結構。

分類：分類是查找描述和區分資料類別或概念的模型（或函數）的過程，目的是為了能夠使用模型來預測類別標籤未知的物件的類別。例如，一個電子郵件程式可能試圖將一個電子郵件分類為“合法的”或“垃圾郵件”。

回歸：試圖找到能夠以最小誤差對該資料建模的函數。回歸分析是最常用於數字預測的統計方法，但也存在其他方法。預測還可以根據現有資料確定趨勢。

資料演化分析：描述並建模其行為隨時間變化的物件的規則或趨勢。雖然這可能包括時間相關資料的表徵，區分，關聯和相關分析，分類，預測或聚類，但這種分析的明顯特徵包括時間序列資料分析，序列或週期性模式匹配以及基於相似性的資料分析。

本報告分析了近年來資料採擷領域的高水準學術論文，挖掘出了包括社交網路、大資料、情報分析、聚類分析、文本挖掘、使用者行為、推薦系統、離群檢測、專家系統等相關關鍵字近年來全球活躍的學術研究。此外，結合知識圖譜技術，本報告將以上研究領域表示為如下圖譜結構

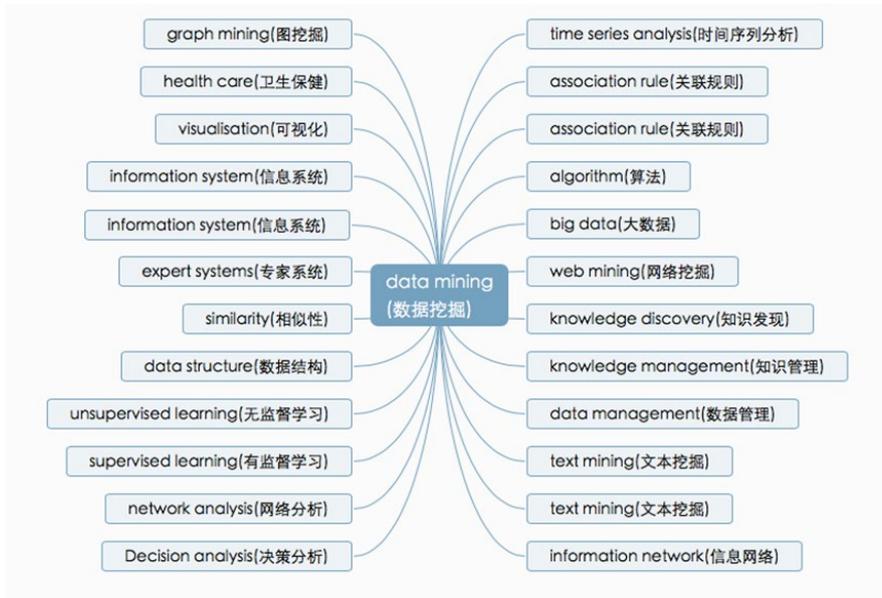


圖 13-1 Data Mining 知識圖譜

具體分析和處理的方法如下：

1. 使用自然語言處理技術，提取每篇論文文獻的關鍵字，據此，結合學科領域知識圖譜，將文章分配到相應領域；
2. 依據學科領域對論文文獻進行聚類，並統計論文數量作為領域的研究熱度；
3. 領域專家按照領域層級對學科領域劃分等級，設計了三級圖譜結構，最後根據概念熱度定義當前研究熱點。三級詳細資料可以參見本報告附錄，或到 <https://www.aminer.cn/data> 中直接下載原始資料。

13.2 資料採擷的發展歷史

隨著資料體量的快速增加，人們希望有一種方法可以幫助處理這些紛繁複雜的資料，從中發現有價值的資訊或知識為決策服務，資料採擷在此背景下應運而生，下文將介紹資料採擷的起源和發展歷程。

20 世紀 60 年代，資料搜集階段。在這個階段受到資料存儲能力的限制，特別是當時還處在磁片存儲的階段，因此主要解決的是資料搜集的問題，而且更多是針對靜態資料的搜集與展現，所解決的商業問題，也是基於歷史結果的統計資料上的。

20 世紀 70 年代，資料存儲階段。隨著資料庫管理系統趨於成熟，存儲和查詢百萬百萬位元組甚千萬億位元組成為可能。而且，資料倉庫允許用戶從面向事物處理的思維方式向更注重資料分析的方式進行轉變。然而，從這些多維模型的資料倉庫中提取複雜深度資訊的能力是非常有限的。

20 世紀 80 年代，資料分析訪問階段。關係性資料庫與結構性查詢語言的出現，使得動態的資料查詢與展現成為可能，人們可以用資料來解決一些更為聚焦的商業問題。在這個階段，KDD 出現了，資料採擷走進了歷史舞臺。也正是在這個時期，出現了些成熟的演算法，能夠“學習”資料間關係，相關領域的專家能夠從中推測出各種資料關係的實際意義。

20 世紀 90 年代，資料倉庫決策與支持階段。OLAP 與資料倉庫技術的突飛猛進使得多層次的資料回溯與動態處理成為現實，人們可以用資料來獲取知識，對經營進行決策，零售公司和金融團體使用資料採擷分析資料和觀察趨勢以擴大客源，預測利率的波動，股票價格以及顧客需求。

21 世紀至今，真正的資料採擷的時代。現在是大資料的時代，因為資訊化的發展非常快，資料也在不斷更新，相應的資料研究也越來越難。我們需要對這些大資料進行處理，從中提取出有價值的資訊。隨著資訊技術的發展，資料採擷已經越來越成熟，成為一門交叉學科。一般來說，資料採擷結合了資料庫，人工智

慧，模式識別，神經網路，機器學習，統計，高性能計算，資料視覺化，空間資料分析和資訊檢索等很多方面的知識^[76]。

13.3 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為資料採擷領域全球學者分佈情況：



圖 13-2 資料採擷領域全球人才分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量遙遙領先且主要分佈於其東西海岸；歐洲、亞洲也有較多的人才分佈；其他諸如非洲、南美洲等地區的學者非常稀少；視覺化領域的人才分佈與各地區的科技、經濟實力情況大體一致。

此外，在性別比例方面，資料採擷領域中男性學者占比 89.4%，女性學者占比 10.6%，男性學者占比遠高於女性學者。

資料採擷領域學者的 h-index 分佈如下所示，大部分學者的 h-index 分佈在中間區域，其中 h-index 在 20-30 區間的人數最多，有 683 人，占比 33.9%，小於 20 區間的人數最少，共 138 人。

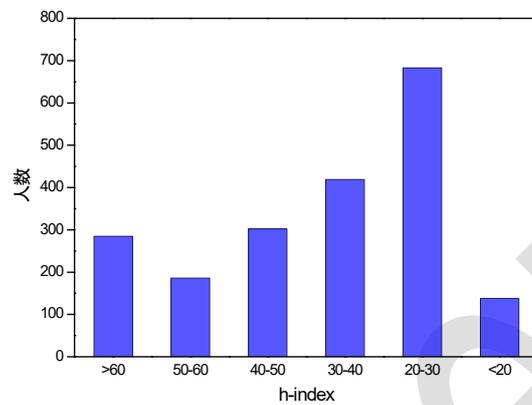


圖 13-3 資料採擷領域學者 h-index 分佈

● 中國人才分佈



圖 13-4 資料採擷領域中國學者分佈

我國專家學者在資料採擷領域的分佈如上圖所示。通過下圖我們可以發現，京津地區在本領域的人才數量最多，其次是珠三角和長三角地區，相比之下，內

陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，通過觀察中國周邊國家的學者數量情況，特別是與日本、東南亞等亞洲國家相比，中國在資料採擷領域學者數量較多且有一定的優勢。

中國與其他國家在資料採擷領域的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 13-1 資料採擷領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	475	35338	74	986
中國-新加坡	49	3841	78	106
中國-澳大利亞	34	1266	37	65
中國-加拿大	32	2938	92	65
中國-英國	23	515	22	41
中國-德國	12	177	15	21
中國-印度	9	638	71	25
中國-瑞士	7	528	75	20
中國-荷蘭	6	101	17	10
中國-沙烏地阿拉伯	5	183	37	8

從上表資料可以看出，中美合作的論文數、引用數、學者數遙遙領先，表明中美間在資料採擷領域合作之密切；此外，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 4 席；中國與加拿大合作的論文數雖然不是最多，但是擁有最高的平均引用數說明在合作品質上中加合作達到了較高的水準。

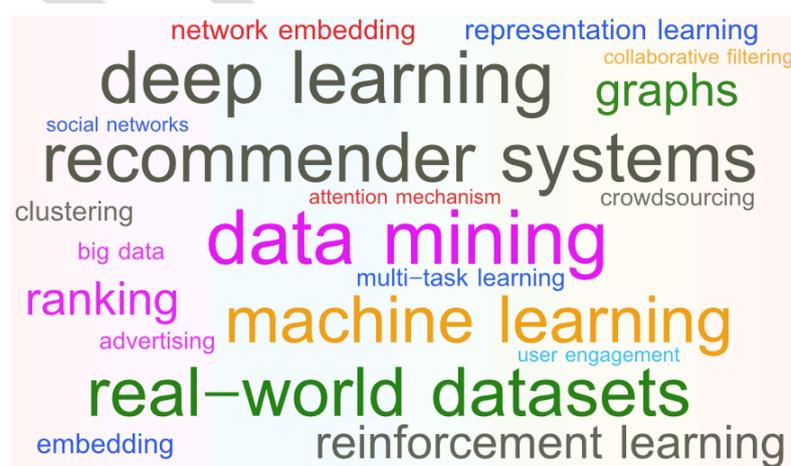
13.4 論文解讀

本節對本領域的高水準學術會議論文進行挖掘，解讀這些會議在 2018-2019 年的部分代表性工作。會議具體包括：

ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

ACM International Conference on Web Search and Data Mining

我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如下圖所示。其中，資料採擷 (data mining)、深度學習 (deep learning)、推薦系統 (recommender systems) 是本領域中最熱的關鍵字。



論文題目：*Graph Convolutional Neural Networks for Web-Scale Recommender Systems*

中文題目：圖卷積神經網路應用於網路規模推薦系統

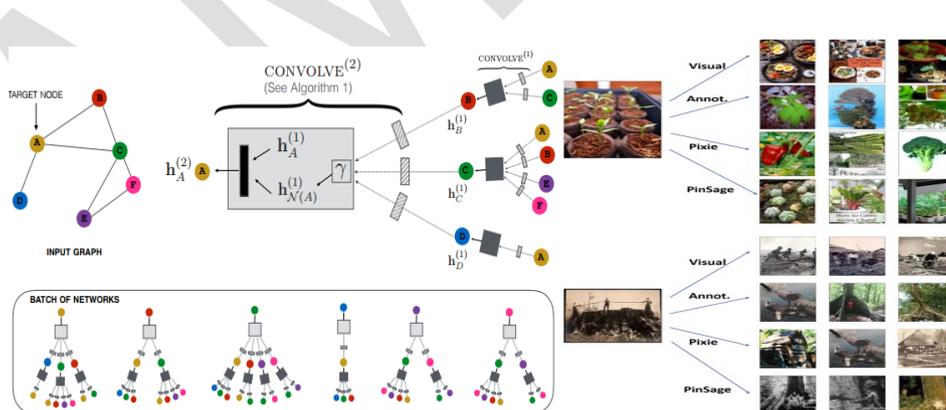
論文作者：Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L. Hamilton, and Jure Leskovec.

論文出處：In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18).

論文地址：<https://dl.acm.org/citation.cfm?doi=3219819.3219890>

研究問題：

圖結構資料作為深層神經網路最新研究成果應用於推薦系統基準的最新的指標，傳統的深度學習網路主要針對圖片，語音等歐氏空間內規則型資料，但是現實中存在很多不是歐氏資料，它們的結構不規則，難以用卷積神經網路對其進行結構資訊的聚合，故將其擴展到有數十億使用者的網路級推薦系統是一個巨大的挑戰。



研究方法：

本研究提出了一個資料有效格拉夫卷積網路 (GDN) 演算法 PinSage ，並結合高效的隨機遊動和局部圖卷積來生成有兩個圖結構和節點 (專案) 特徵的嵌入模型。該方法的關鍵計算工作是局部圖卷積的概念。為了產生用於一個節點 (一

個項目)的嵌入,我們應用其聚集來自所述節點的本地圖形附近的特徵的資訊(例如,視覺,文本特徵)(上圖所示)的多個卷積模組。每個模組學習如何聚合資訊從一個小圖附近,並通過堆疊多個這樣的模組,我們的方法可以得到有關本地網路拓撲資訊。重要的是,這些局部卷積模組的參數在所有節點共用,使得獨立於輸入圖形大小的我們的方法的參數的複雜性。與之前 GCN 方法相比,該方法提高了模型的魯棒性和收斂性;用 2 個深度卷積來概述提出的方法,左邊將目標節點輸入圖形,經過底部計算目標節點的兩層神經網路,每個目標的神經網路不同,但它們都有相同的共用參數集(即右邊的函數和演算法,矩形框和細框表示緊密連接的多層神經網路),輸出與之關聯的目標節點。

研究結果：

本文評估了 PinSage 在兩個任務中生成的嵌入:推薦相關的 pin 和在用戶的 home/news feed 中推薦 pin。為了推薦相關的引腳,我們在嵌入空間中選擇查詢引腳的 K 個近鄰。我們使用離線排名和受控用戶研究來評估這個相關 pin 推薦任務的性能。對於 homefeed 推薦任務,我們選擇嵌入空間中距離用戶最近的固定項 pin。我們使用 A / B 測試來評估完全部署的生產系統在此任務上的性能,以度量對使用者參與的總體影響。在這種測試情況下 PinSage 是表現最好的,60% 的情況下 PinSage 的推薦結果是優於 Pixie 的。

論文題目：*Unbiased Learning-to-Rank with Biased Feedback*

中文題目：基於有偏差的回饋資料進行無偏差的 LTR

論文作者：Joachims, Thorsten and Swaminathan, Adith and Schnabel, Tobias.

論文出處：In Proceeding WSDM '17 Proceedings of the Tenth ACM International Conference on Web Search and Data Mining.

論文地址：<https://dl.acm.org/citation.cfm?id=3018699>

研究問題：

隱藏的回饋資料(如點擊,停留時間等)是人類交互系統的資料的豐富來源。雖然隱藏的回饋資料具有很多優點(如採集費用低,以用戶為中心且即時性強),但如何從有偏差的回饋資料中,訓練無偏差的排序模型,是這些資料有效利用的一個主要障礙。例如,搜索排名位置偏差會強烈影響檢索結果的點擊次數,往往只對排名較前的資料進行點擊,後面的即使與檢索內容相關,也不會被點擊,用戶的回饋就受到現在系統的影響,導致檢索無法提升到全域最優的情況。

研究方法：

本文使用一種可證明的無偏估計器,用來評估有偏差的回饋資料評估排序性能,並且提出一種針對LTR的傾向加權經驗風險最小化的方法,得到了一個傾向加權排序支援向量機(傾向加權是用來糾正選擇偏差時,丟棄查詢沒有點擊在學習排序,此工作的關鍵是認識到逆傾向得分可以使用的更有力,排名位置偏差,相信偏差,語境效果,文檔等使用較多的點擊模型來估計每個點擊的傾向而不是查詢獲得點擊的傾向),將其用於隱式回饋中進行判別學習,其中點擊模型扮演傾向估計器的角色。與應用點擊模型來消除點擊資料偏差的大多數傳統方法相比,這個方法允許對排序函數進行訓練。

$$\begin{aligned}
\hat{w} &= \operatorname{argmin}_{w, \xi} \frac{1}{2} w \cdot w + \frac{C}{n} \sum_{j=1}^n \frac{1}{q_j} \sum_{y \in Y_j} \xi_{jy} & \operatorname{rank}(y_i | y) - 1 &= \sum_{y \neq y_i} \mathbb{1}_{w \cdot [\phi(x_i, y) - \phi(x_i, y_i)] > 0} \\
s.t. \quad & \forall y \in Y_1 \setminus \{y_1\} : w \cdot [\phi(x_1, y_1) - \phi(x_1, y)] \geq 1 - \xi_{1y} & &\leq \sum_{y \neq y_i} \max(1 - w \cdot [\phi(x_i, y_i) - \phi(x_i, y)], 0) \\
& \vdots & &\leq \sum_{y \neq y_i} \xi_{iy}. \\
& \forall y \in Y_n \setminus \{y_n\} : w \cdot [\phi(x_n, y_n) - \phi(x_n, y)] \geq 1 - \xi_{ny} \\
& \forall j \forall y : \xi_{jy} \geq 0.
\end{aligned}$$

本文提出的模型通過一個鬆弛公式有效地求解這類二次規劃問題，並且我們使用支援向量機 rank1 適當修改基於逆傾向得分 (IPS) 局部資訊特徵的權值 $1/q_j$ 生成的代碼將應用於線上推理模型。通過 Position-based Click Model with Click Noise (PCMCN) 進行排序文檔的建模，用以方便預測點擊概率 (前系統選擇各個文檔的概率)。

研究結果：

通過提供了廣泛的經驗證據 [https://www.joachims.org/svm light/svm rank.html](https://www.joachims.org/svm%20light/svm%20rank.html)，證明了使用傾向加權排序支援向量機產生實體經驗風險最小化方法，並通過廣泛的經驗證據表明，本文提出的學習方法對選擇偏差、雜訊和模型錯誤規範具有魯棒性。此外，我們在一個即時搜尋引擎上的實驗表明，該方法在學習過程中沒有任何啟發或手動干預的情況下，可以顯著地提高檢索效果。

題目： *Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding*

中文題目：基於卷積序列嵌入的個性化 Top-N 序列推薦

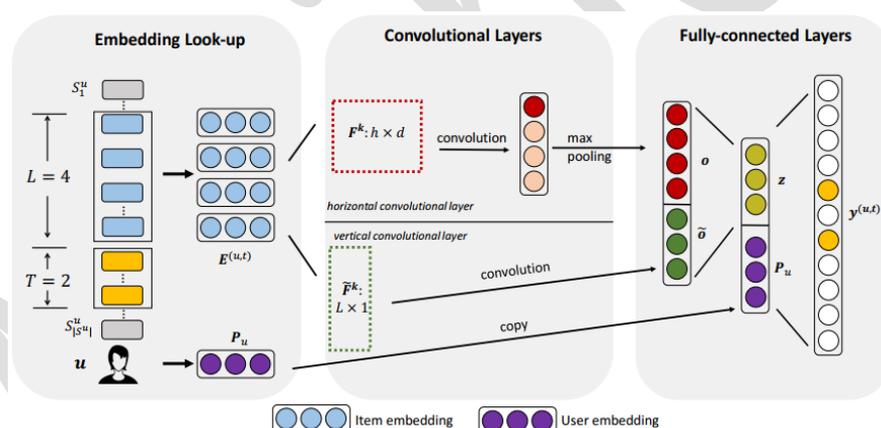
論文作者：Tang, Jiayi, Wang, Ke.

論文出處：In Proceeding WSDM '18 Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining.

論文地址：<https://arxiv.org/abs/1809.07426>

研究問題：

推薦系統漸漸成為許多應用軟體的核心技術，目前往往只考慮擁有一般偏好的推薦系統在銷售 iPhone 後將錯過推薦手機配件的機會，因為購買手機配件並不是一個長期的用戶行為，現有的 Top-N 序列模型不能對聯合級別的順序模式進行建模並且沒有考慮先後互換的跨越行為，Top -N 順序推薦將每個使用者建模為過去交互的項目序列，並旨在預測用戶在“不久的將來”可能交互的排名前 n 的項目。交互順序意味著順序模式扮演了一個重要的角色，其中序列中最近的專案對下一個專案有更大的影響。鑒於此，本文提出一種卷積序列嵌入推薦模型(簡稱 Caser) 來解決 top-N 序列推薦問題。



研究方法：

本文提出一個卷積序列嵌入推薦模型(Caser)作為一個解決方案來解決這一需求。我們的想法是在時間和潛在空間中，把一系列最近的項目嵌入到一個“圖像”中，將順序模式作為圖像的局部特徵並使用卷積濾波器學習這些特徵，其使用水準和垂直卷積篩檢程式來捕獲點級、聯合級和跳過行為的順序模式，再對使用者的一般偏好和順序模式進行建模，並在一個統一的框架中概括了幾種現有的先進方法。故此方法可以提供一個統一的、可構建(可性)的網路結構，用於獲

取一般偏好和順序模式。Caser 網路結構圖如上圖所示，矩形框表示用戶序列中的條目 $S_1^u, \dots, S_{|S_u|}^u$ ，內嵌有圓圈的矩形框表示某個向量，如用戶嵌入 P_u 。虛線矩形框是具有不同尺寸的卷積濾波器。卷積層中的紅色圓圈表示每個卷積結果的最大值。我們使用之前的 4 個行為 ($L=4$) 來預測這個用戶將在接下來的 2 個步驟 ($T=2$) 中與哪些項目交互。

研究結果：

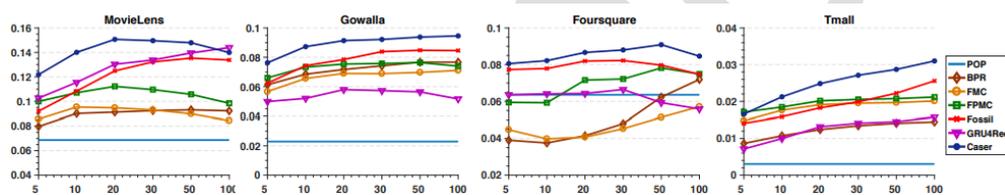
通過公共資料集上的實驗表明，Caser 優於最先進的序列推薦方法，作為一種新穎的 top-N 序列推薦解決方案，Caser 將最近的動作建模為時間和潛在維度之間的“圖像”，並應用卷積濾波器學習序列模式。Caser 提供了一個統一和靈活的網路結構捕捉序列推薦的許多重要的功能（實現點級和聯合級順序模式、跳過行為和用戶一般偏好）。

Table 1: Statistics of the datasets

Datasets	Sequential Intensity	#users	#items	avg. actions per user	Sparsity
MovieLens	0.3265	6.0k	3.4k	165.50	95.16%
Gowalla	0.0748	13.1k	14.0k	40.74	99.71%
Foursquare	0.0378	10.1k	23.4k	30.16	99.87%
Tmall	0.0104	23.8k	12.2k	13.93	99.89%

Table 2: Performance comparison on the four data sets.

Dataset	Metric	POP	BPR	FMC	FPMC	Fossil	GRU4Rec	Caser	Improv.
MovieLens	Prec@1	0.1280	0.1478	0.1748	0.2022	0.2306	0.2515	0.2502	-0.5%
	Prec@5	0.1113	0.1288	0.1505	0.1659	0.2000	0.2146	0.2175	1.4%
	Prec@10	0.1011	0.1193	0.1317	0.1460	0.1806	0.1916	0.1991	4.0%
	Recall@1	0.0050	0.0070	0.0104	0.0118	0.0144	0.0153	0.0148	-3.3%
	Recall@5	0.0213	0.0312	0.0432	0.0468	0.0602	0.0629	0.0632	0.5%
	Recall@10	0.0375	0.0560	0.0722	0.0777	0.1061	0.1093	0.1121	2.6%
	MAP	0.0687	0.0913	0.0949	0.1053	0.1354	0.1440	0.1507	4.7%
Gowalla	Prec@1	0.0517	0.1640	0.1532	0.1555	0.1736	0.1050	0.1961	13.0%
	Prec@5	0.0362	0.0983	0.0876	0.0936	0.1045	0.0721	0.1129	8.0%
	Prec@10	0.0281	0.0726	0.0657	0.0698	0.0782	0.0571	0.0833	6.5%
	Recall@1	0.0064	0.0250	0.0234	0.0256	0.0277	0.0155	0.0310	11.9%
	Recall@5	0.0257	0.0743	0.0648	0.0722	0.0793	0.0529	0.0845	6.6%
	Recall@10	0.0402	0.1077	0.0950	0.1059	0.1166	0.0826	0.1223	4.9%
	MAP	0.0229	0.0767	0.0711	0.0764	0.0848	0.0580	0.0928	9.4%
Foursquare	Prec@1	0.1090	0.1233	0.0875	0.1081	0.1191	0.1018	0.1351	13.4%
	Prec@5	0.0477	0.0543	0.0445	0.0555	0.0580	0.0475	0.0619	6.7%
	Prec@10	0.0304	0.0348	0.0309	0.0385	0.0399	0.0331	0.0425	6.5%
	Recall@1	0.0376	0.0445	0.0305	0.0440	0.0497	0.0369	0.0565	13.7%
	Recall@5	0.0800	0.0888	0.0689	0.0959	0.0948	0.0770	0.1035	7.9%
	Recall@10	0.0954	0.1061	0.0911	0.1200	0.1187	0.1011	0.1291	7.6%
	MAP	0.0636	0.0719	0.0571	0.0782	0.0823	0.0643	0.0909	10.4%
Tmall	Prec@1	0.0010	0.0111	0.0197	0.0210	0.0280	0.0139	0.0312	11.4%
	Prec@5	0.0009	0.0081	0.0114	0.0120	0.0149	0.0090	0.0179	20.1%
	Prec@10	0.0007	0.0063	0.0084	0.0090	0.0104	0.0070	0.0132	26.9%
	Recall@1	0.0004	0.0046	0.0079	0.0082	0.0117	0.0056	0.0130	11.1%
	Recall@5	0.0019	0.0169	0.0226	0.0245	0.0306	0.0180	0.0366	19.6%
	Recall@10	0.0026	0.0260	0.0333	0.0364	0.0425	0.0278	0.0534	25.6%
	MAP	0.0030	0.0145	0.0197	0.0212	0.0256	0.0164	0.0310	21.1%



論文題目：*Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution*

中文題目：機器學習的理論局限性與因果革命的七個火花

論文作者：Judea Pearl.

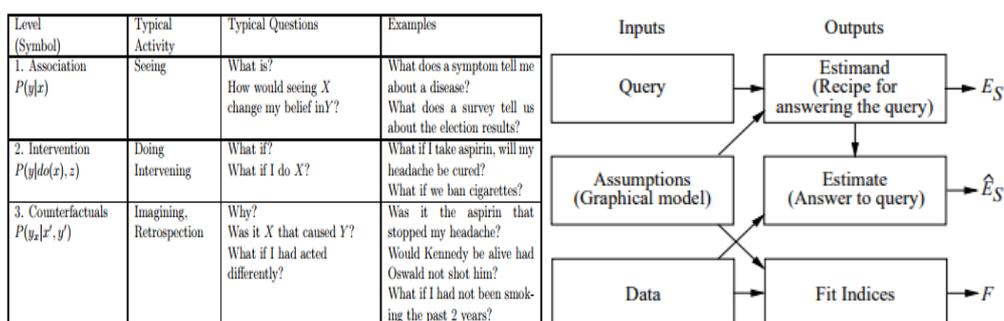
論文出處：Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining

論文地址：<https://dl.acm.org/citation.cfm?id=3176182>

研究問題：

當前的機器學習系統幾乎完全以一種統計的或無模型的模式運行，即通過輸入進行參數優化以提高系統的性能，這對機器學習系統的能力和性能造成了嚴重

的理論限制。這樣的系統不能對干預和回顧進行推理，因此這種系統不能作為強人工智慧的基礎。為了實現人類級別的智慧，學習系統應該需要真實的模型的指導，就如同在因果推理任務中的模型一樣。為證明這些模型的重要作用，文章概述了七個當前機器學習系統無法實現、但可以使用因果建模工具完成的任務。



研究方法：

圖形和結構模型的進步使得假設在計算上易於管理，從而使模型驅動的推理成為一個更有前途的方向，可以在此基礎上建立強人工智慧，為描述機器學習系統所面臨的障礙，即左圖所示，使用一個從關聯（看正在發生什麼）到干涉（做出干預措施）、最後到反事實（想像，回顧）的三層結構來體現因果推理中的推論，並且在因果關係的層次結構中，只有在 i 級或更高級別的資訊可用時，才能回答 i 級的問題。哈佛大學教授 Garry King 認為，“在過去的幾十年裡，人們已經學到了比以往所有歷史記錄中所學到的所有內容的總和還要多的關於因果推斷的知識”，文章稱其為“因果革命”，相應的數學框架稱為結構因果模型 (Structural Causal Models, SCM)。右圖為以推理機的形式表述結構因果模型，其包括詢問，假設和資料 3 個輸入和預估計（回答問題的秘密方法），估計（問題的答案）及擬合指數（用來測量資料與模型所傳達的假設之間的相容性。）三個輸出。接下來，文章介紹了 SCM 框架的 7 項最重要的特性，即需要結構因果模型來完成的

七個重點任務：1.編碼因果假設的透明度和可測試性；2.do-calculus 和混雜控制；3.反事實演算法；4.調節分析和直接、間接影響評估；5.外部效度和樣本選擇偏差；6.缺失數據；7.因果關係發現。

研究結果：

為了達到人類的智慧水準，機器學習需要一個現實模型的指導，類似於在因果推理任務中使用的模型，本文已經描述了其中的一些認知任務，並展示了它們是如何在結構因果模型的框架內運行且體現了基於模型的方法對認知任務是至關重要的。本文提出了結構因果模型的七個任務，這七個任務是目前的機器學習系統無法達到的，使用因果建模工具和科學的資料可以達到說明實現人類水準的強人工智慧。

論文題目：*Network Density of States*

中文題目：狀態網路密度

論文作者：Kun Dong, Austin R. Benson, and David Bindel.

論文出處：In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19).

論文地址：<https://dl.acm.org/citation.cfm?id=3330891>

研究問題：

本論文旨在解決真實世界圖形光譜內部存在的難以計算和解釋的問題。與幾何學不同，通過頻譜密度的整體分佈來研究圖形在很大程度上限於簡單的隨機圖形模型。真實世界圖譜的內部仍未開發，難以計算和解釋。本文旨在深入研究了

現實世界圖譜密度的核心，並展示頻譜密度的估計如何促進許多常見中心度度量的計算，使用頻譜密度來估計關於圖結構的有意義的資訊。

研究方法：

態密度在理解凝聚態物理學中的電子能帶結構中起著重要作用，以往已經有文獻提出了幾種方法來估計光譜密度。本文採用了其中兩種方法：核多項式方法（KPM），其中涉及 DOS/LDOS 的多項式展開，以及通過 Lanczos 反覆運算實現的高斯正交（GQL）。這些方法由 Cohen-Steiner 等人在很早以前提出，但尚未應用在網路環境中。本文在此基礎上，提出了一種用於 LDOS 的新的直接嵌套剖析方法（ND），以及針對圖形的特定修改方法以改進 KPM 和 GQL 方法的收斂性。之後，本文由提出了主題濾波方法（Motif Filtering），來分離頻譜中尖峰成分，從而使用更少的切比雪夫矩來獲得更準確的近似值。如圖所示

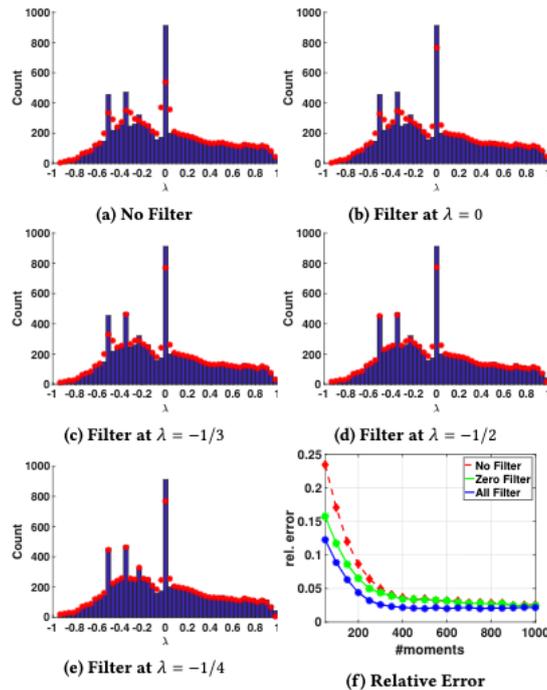


Figure 3: The improvement in accuracy of the spectral histogram approximation on the normalized adjacency matrix for the High Energy Physics Theory (HepTh) Collaboration Network, as we sweep through spectrum and filter out motifs. The graph has 8638 nodes and 24816 edges. Blue bars are the real spectrum, and red points are the approximated heights. (3a-3e) use 100 moments and 20 probe vectors. (3f) shows the relative L_1 error of the spectral histogram when using no filter, filter at $\lambda = 0$, and all filters.

研究結果：

本文使頻譜密度的計算成為分析大型實際網路的實用工具，具體方法借鑒了凝聚態物理學中的方法，但是通過對圖形圖案的特殊處理保留了獨特的光譜指紋，這些調整可改善網路分析設置中的性能。本文證明了光譜密度對於圖中的小變化都是穩定的，並提供了對方法中近似誤差的分析。實驗通過僅使用一個計算節點來處理具有數千萬個節點和數十億條邊的圖形來驗證了方法的效率。

論文題目：*Real-time Personalization using Embeddings for Search Ranking at Airbnb*

中文題目：利用嵌入式表示的 Airbnb 即時個性化搜索排序

論文作者：Mihajlo Grbovic, Haibin Cheng

論文出處：In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18).

論文地址：<https://dl.acm.org/citation.cfm?id=3219885>

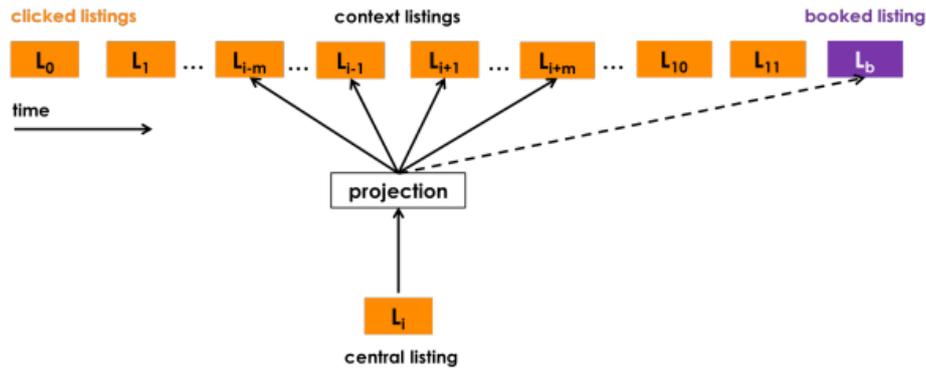
研究問題：

Airbnb 是全世界最大的房屋短租網站，提供了一個連接房主和租客的仲介平臺，因此也需要一個高水準的搜索與推薦服務來支援其使用。系統中房主和租客的對話模式包括：租客點擊某個房屋商品、租客預訂某個房屋商品、房主(可能)拒絕某個預訂請求。基於上述幾種交互資料，本文提出了一個即時的排序模型來捕捉使用者的短時興趣以及長期興趣，從而達到一個更好的搜索效果。

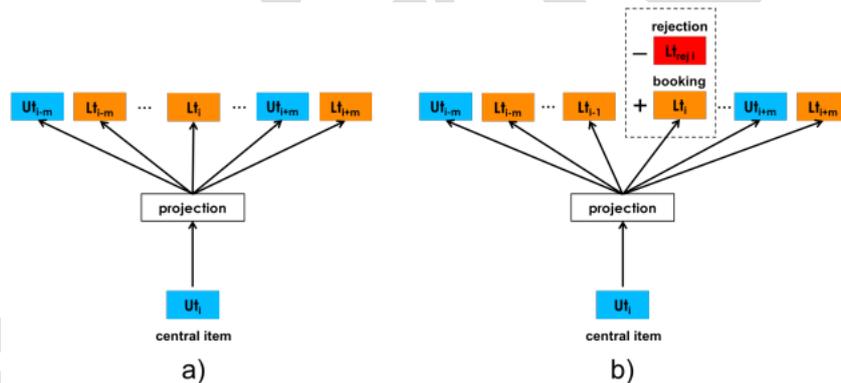
研究方法：

文章認為，使用者的搜索目標分為短時興趣和長時期興趣兩部分，其中短時興趣代表用戶因為當前的需求所導向的商品傾向，而長時期興趣則表示用戶選擇商品的習慣，為了更好地捕獲這兩部分資訊，本文通過兩種方式生成了不同的嵌入式表示來進行分別建模。

通過使用者點擊資料生成 listing 的嵌入表示，旨在進行 listing 的相似推薦，以及對用戶會話內的即時個性化推薦，主要亮點是在訓練時增加了一個確定的全域上下文視窗。



通過預定資訊生成 user-type 和 listing-type 的嵌入表示,目的是捕捉不同 user-type 的長期喜好。由於預定資訊過於稀疏, Airbnb 對同屬性的 user 和 listing 進行了聚合,另一個亮點是提出利用用戶的負反饋(即房主的拒絕行為)來修正嵌入表示的訓練。



研究結果：

本文通過兩部分的實驗表明了其有效性。

其一是提出了一種離線的推薦系統測試方法,用本文訓練出的嵌入表示來對商品進行排序,用這個排序結果與實際用戶的點擊和預訂進行對比來實現離線評測。從這個角度看,本文提出的嵌入表示方法顯著優於已有的方法。

其二是给出了一些應用了本文嵌入表示方法得出的相似商品及用戶習慣的案例分折,並提供了工具,通過案例觀察可以發現,本文給出的相似商品不僅體

現在價格等已知條目中，也體現了商品的風格資訊，這表明即使不利用圖片，本文給出的方法也可以挖掘出相似風格的商品。

總的來說，本文是一篇實用性很強的文章，不僅從實際角度出發，對用戶的短期興趣和長期偏好進行了建模，而且針對推薦系統嵌入表示訓練中的冷開機問題等都進行了很有效的處理，並提出了一些針對作業系統的離線測試手段，是很有效的將理論應用到工程中的工作。

13.5 資料採擷進展

近幾年，我們已經迎來了大資料時代，各大互聯網企業每天都在產生數以億計的資料。各類資料往往都隱含著一些有價值的資訊，如果人們手動地進行資料分析，往往需要耗費大量的時間。同時，大量未經處理的資料可能會被人們所忽視。資料採擷就是想自動地從大規模的資料中挖掘出有意義的知識或者模式。這裡，我們將資料採擷領域近期的主要發展歸為兩大類：複雜資料採擷、分散式資料採擷。

複雜資料包括序列資料、圖資料等。在序列資料採擷中，基於注意力（Attention）機制的 Transformer 模型表現出了巨大的潛力，在機器翻譯等任務上取得了非常好的效果。隨後，BERT 模型使用雙向 Transformer 通過預訓練方式在各種自然語言處理的任務上都達到了當時最好的結果。在圖資料採擷研究中，網路表示學習仍然是近年來非常熱門的話題。從 DeepWalk 演算法開始，基於隨機遊走的演算法在無監督的表示學習任務中表現良好。NetMF 演算法將幾種基於隨機遊走的演算法統一寫成了矩陣分解的形式，給網路表示學習演算法提供了

理論基礎。圖卷積神經網路是另一種處理圖資料的有效方法，借鑒了圖譜論中的圖卷積並使用圖的拉普拉斯矩陣，在半監督的節點分類任務和圖分類任務中都表現出很好的效果。除此之外，異構網路的表示與挖掘也逐漸被大家所關注。

分散式資料採擷已成為資料採擷領域非常有前途的方向。隨著資料採擷計算成本的增加和資料隱私保護的問題，分散式資料採擷開始備受關注。分散式資料採擷利用分散式網站的資源來降低計算成本並增強資料保密性。由於分散式資料採擷採用了不同的計算方式，傳統的資料採擷技術很難直接應用於分散式資料採擷。目前，資料安全與資料隱私開始被大家所關注。2018年5月，通用資料保護條例（GDPR）在歐盟正式生效，這也使得基於隱私保護的分散式資料採擷方法逐漸被研究者所重視。

資料採擷已經被廣泛地應用於各類實際問題，包括金融資料分析、推薦系統等。資料採擷相關研究需要結合實際問題，注重與機器學習、統計學科等的交叉，從大資料中挖掘出有價值的資訊。

14 資訊檢索與推薦

14.1 資訊檢索與推薦概念

- 資訊檢索

R.Baeza-Yates 教授在其著作《現代資訊檢索中》中指出，資訊檢索 (Information Retrieval , IR) 是電腦科學的一大領域，主要研究如何為使用者訪問他們感興趣的資訊提供各種便利的手段，即：資訊檢索涉及對文檔、網頁、連線目錄、結構化和半結構化記錄及多媒體物件等資訊的表示、存儲、組織和訪問，資訊的表示和組織必須便於使用者訪問他們感興趣的資訊^[77]。

在範圍上，資訊檢索的發展已經遠超出了其早期目標，即對文檔進行索引並從中尋找有用的文檔。如今，資訊檢索的研究包括使用者建模、Web 搜索、文本分析、系統構架、使用者介面、資料視覺化、過濾和語言處理等技術。

資訊檢索的主要環節包括資訊內容分析與編碼、組成有序的資訊集合以及使用者提問處理和檢索輸出。其中資訊提問與資訊集合的匹配、選擇是整個環節中的重要部分。當使用者向系統輸入查詢時，資訊檢索過程開始，接著使用者查詢與資料庫資訊進行匹配。返回的結果可能是匹配或不匹配查詢，而且結果通常被排名。大多數資訊檢索系統對資料庫中的每個物件與查詢匹配的程度計算數值分數，並根據此值進行排名，然後向使用者顯示排名靠前的物件，資訊檢索框架如下圖所示。

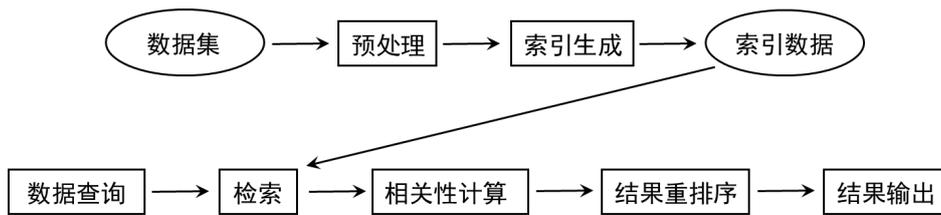


圖 14-1 資訊檢索系統流程

● 推薦系統

推薦系統 (Recommendation System , RS) 是指資訊過濾技術，從海量專案 (專案是推薦系統所推薦內容的統稱，包括商品、新聞、微博、音樂等產品及服務) 中找到使用者感興趣的部分並將其推薦給用戶，這在用戶沒有明確需求或者項目數量過於巨大、凌亂時，能很好地為使用者服務，解決資訊超載問題^[78]。

如下圖所示，一般推薦系統模型流程通常由 3 個重要的模組組成：使用者特徵收集模組，使用者行為建模與分析模組，推薦與排序模組。推薦系統通過使用者特徵收集模組收集使用者的歷史行為，並使用用戶行為建模和分析模組構建合適的數學模型分析使用者偏好，計算專案相似度等，最後通過推薦與排序模組計算使用者感興趣的專案，並將專案排序後推薦給用戶^[79]。

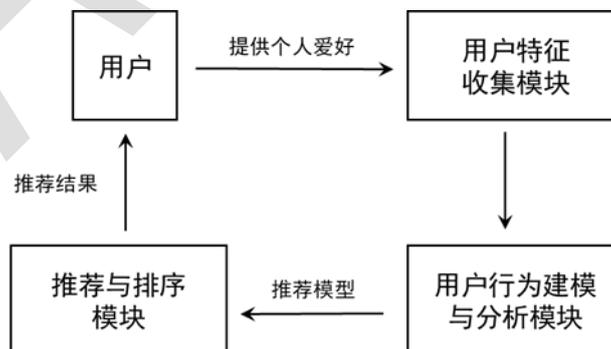


圖 14-2 推薦系統模型流程

● 聯繫與區別

資訊的檢索與推薦都是使用者獲取資訊的手段，無論是在互聯網上，還是線上的生活場景裡，這兩種方式都大量並存，兩者之間的關係是互補的：搜尋引擎需要使用者主動提供準確的關鍵字來尋找資訊，因此不能解決使用者的很多其他需求，比如當用戶無法找到準確描述自己需求的關鍵字時，搜尋引擎就無能為力了。和搜尋引擎一樣，推薦系統也是一種說明使用者快速發現有用資訊的工具。與搜尋引擎不同的是，推薦系統不需要使用者提供明確的需求，而是通過分析用戶的歷史行為給用戶的興趣建模，從而主動給用戶推薦能夠滿足他們興趣和需求的資訊。因此，從某種意義上說，推薦系統和搜尋引擎對於使用者來說是兩個互補的工具。搜尋引擎滿足了使用者有明確目的時的主動查找需求，而推薦系統能夠在使用者沒有明確目的的時候說明他們發現感興趣的新內容。在實際生活中也有很多運用。

同時，資訊的檢索與推薦也有著一定的區別，可以分為以下幾個方面：

首先是主動與被動的不同。搜索是一個非常主動的行動，用戶的需求也十分明確，在搜尋引擎提供的結果裡，用戶也能通過流覽和點擊來明確的判斷是否滿足了用戶需求。然而，推薦系統接受資訊是被動的，需求也都是模糊而不明確的。

其次是個性化程度的高低。搜尋引擎雖然也可以有一定程度的個性化，但是整體上個性化運作的空間是比較小的，因為當需求非常明確時，找到結果的好壞通常沒有太多個性化的差異。但是推薦系統在個性化方面的運作空間要大很多，雖然推薦的種類有很多，但是個性化對於推薦系統是非常重要的，以至於在很多時候大家索性就把推薦系統稱為“個性化推薦”甚至“智慧推薦”。

再次就是需求時間不同。在設計搜索排序演算法裡，需要想盡辦法讓最好的結果排在最前面，往往搜尋引擎的前三條結果聚集了絕大多數的用戶點擊。簡單來說，“好”的搜索演算法是需要讓使用者獲取資訊的效率更高、停留時間更短。但是推薦恰恰相反，推薦演算法和被推薦的內容往往是緊密結合在一起的，用戶獲取推薦結果的過程可以是持續的、長期的，衡量推薦系統是否足夠好，往往要依據是否能讓用戶停留更多的時間，對用戶興趣的挖掘越深入，越“懂”用戶，那麼推薦的成功率越高，用戶也越願意留在產品裡。

最後是評價方法不同。搜尋引擎通常基於 Cranfield 評價體系，整體上是將優質結果盡可能排到搜索結果的最前面，讓用戶以最少的點擊次數、最快的速度找到內容是評價的核心。而推薦系統的評價要寬泛很多，既可以用諸如 MAP (Mean Average Precision) 的常見量化方法評價，也可以從業務角度進行側面評價^[80]。

14.2 資訊檢索和推薦技術發展歷史

● 資訊檢索

資訊檢索的目的是獲取所需資訊，而這要基於比較完善的檢索技術，使用者需求的變化和資訊技術的進步對資訊檢索的發展有著重要的影響。根據技術的演化，我們將資訊檢索發展歷程分為三個階段：

(1) 數位圖書館 / 文檔電子化時代

1954 年，Vannevar Bush (範內瓦·布希) 在“Atlantic Monthly”7 月號發表了一篇名為“*As We May Think*”的文章，這篇文章影響了幾代的電腦科學家。文章提

到：“未來人們能夠實現對海量圖書資源（1M）進行快速的訪問”。概括出了資訊檢索在數位圖書館時代的特徵，即對文檔全文內容的快速檢索。

範內瓦·布希在擔任美國科學研究與發展辦公室主任期間推進了美國軍隊研究機構與高校研究機構的合作，正是當時在這種合作關係中發揮最重要影響的三所大學（哈佛大學、麻省理工學院、加州大學伯克利分校）與後來成立的美國國防部高等研究計畫署（ARPA）合作開發出了互聯網的雛形：ARPANET。



圖 14-3 範內瓦·布希（1890-1974）

1957 年，Luhn 在論文“*A Statistical Approach to Mechanized Encoding and Searching of Literary Information*”裡提到“...a writer chooses that level of subject specificity and that combination of words which he feels will convey the most meaning.”，這是一種以單詞作為索引單元的文檔檢索方法。

20 世紀 60 年代，Gerard Salton 創造了資訊檢索系統 SMART（Salton’s Magic Automatic Retrieval of Text），推進了資訊檢索相關研究的水準提升。SMART 系統並非搜尋引擎，但它具備搜尋引擎所具有的文本索引、查詢處理、結果排序等功能。

20 世紀 60 年代後期另外兩個研究領域需要提及。第一個是 Julie Beth Lovins 於 1968 年在麻省理工學院開發的詞幹演算法 (Stemming Algorithm) ; 另一個研究涉及評估指標, 例如 William Cooper 在 1968 年提出的“Cooper”, 這個度量標準目前已在多個應用程式中大量使用。

在數位圖書館時代, 資訊檢索技術主要應用於封閉資料集合、單機模式或專網內的主機-終點模式, 在商業應用方面, 則是提供軟體/解決方案, 專網內的查詢服務。

(2) 早期互聯網時代

隨著資訊技術的爆炸式發展, 資訊檢索的發展發生了質的飛躍。Tim Berners-Lee (蒂姆·伯納斯·李) 基於尚未被商用的互聯網提出了萬維網 (Web) 的原型建議。1991 年 8 月, 蒂姆·伯納斯·李在一台 NeXT 電腦上建立了第一個網站 <http://nxoc01.cern.ch/>。他一直堅持將公開和開放作為萬維網的靈魂。



圖 14-4 蒂姆·伯納斯·李和他的 NeXT 電腦

從事檢索業務的公司隨著互聯網的發展而快速崛起, 如雅虎、百度等。在眾多公司中, 穀歌被公認為全球最大的搜尋引擎公司, 其業務包括互聯網搜索、雲計算、廣告技術, 開發並提供大量基於互聯網的產品與服務。隨著互聯網的發展,

面對眾多雜亂無章的資訊，如何對數以億計的相關網頁進行排序成為搜尋引擎演算法的核心問題，為此谷歌開發出了著名的 PageRank 演算法。

PageRank 的主要原理是用連結數量作為搜索排序的一個因數。在互聯網上，如果一個網頁被很多其他網頁所連結，說明它受到普遍的承認和信賴，那麼它的排名就高，這就是 PageRank 的核心思想。PageRank 演算法將互聯網中大多數的網頁通過基於連結來計算網頁品質的方式進行排名，為搜尋引擎使用者提供較好的基於連結查詢的搜索結果，同時該演算法能夠進行離線分析處理，大大縮短了搜尋引擎使用者的服務回應時間，實屬電腦科學史上一項偉大成就，它以及其簡明的邏輯，發明了迄今為止在搜尋引擎領域還相當有代表性的演算法，解決了數以億計的網頁品質評估問題，拋開它難以估量的商業價值不談，就說其學術方面，這種依靠資料自身結構的學習方法，也依然還在當前很多資訊檢索領域啟發著我們。

業界主要表現為第一代搜尋引擎和第二代搜尋引擎的出現，國外有 AltaVista、Excite、WebCrawler 和 Yahoo！，國內有應用于國防和安全領域的“天羅”和面向公眾提供服務的天網。第二代搜尋引擎的代表是 1998 年成立的 Google 和 2000 年 1 月創建的中文搜尋引擎——百度。在百度之後，多家中文搜尋引擎相繼出現，例如中搜、搜狗、搜搜和有道。

這個時期資訊檢索的應用形態的特徵是開放的、大規模的、即時的、多媒體的，尤其巨型搜尋引擎採集到的公開資料和使用者訪問日誌等非公開資料深刻地影響著這一時期資訊檢索領域的創新模式。

(3) Web 2.0 時代

在 Web 2.0 時代，用戶對 Web 有更深入的參與需求，這就對資訊檢索提出了更高的要求。資訊搜索的發展開始更加關注使用者需求，以實現內容與行為的精確 Web 搜索。

這個時期的資訊檢索實現了內容資料與社會各側面的電子化資料（萬維網、社交網、物聯網、地理資訊等）的全面融合；尤其是對社交網路資料的採集和大資料處理技術出現了社會化趨勢。

- 資訊推薦

上個世紀最後二十年以來，互聯網的發展和普及為人們提供了一個全新的資訊存儲、加工、傳遞和使用的載體，網路資訊也迅速成為了社會成員獲取知識和資訊的主要管道之一。

一般認為推薦系統的研究始於 1994 年明尼蘇達大學，Group Lens 研究組推出了 Group Lens 系統，該工作不僅首次提出了協同過濾的思想，並且為推薦問題建立了一個形式化的模型，為隨後幾十年推薦系統的發展帶來了巨大影響。

之後，推薦系統的相關技術得到了進一步發展和重視。1995 年 3 月，卡耐基·梅隆大學的 Robert Armstrong 等人在美國人工智慧協會提出了個性化導航系統 Web Watcher；斯坦福大學的 Marko Balabanovic 等人在同一會議上推出了個性化推薦系統 LIRA；1997 年，AT&T 實驗室提出了基於協同過濾的個性化推薦系統 PHOAKS 和 Referral Web；2000 年，NEC 研究院的 Kurt 等人為搜尋引擎 Cite Seer 增加了個性化推薦功能；2003 年，Google 開創了 AdWords 盈利模式，通過使用者搜索的關鍵字來提供相關的廣告。2007 年開始，Google 為 AdWords 添加了個性化元素，不僅僅關注單詞搜索的關鍵字，而且對使用者一段時間內的

推薦歷史進行記錄和分析，據此瞭解用戶的喜好和需求，更為精確地呈現相關的廣告內容。

資訊推薦系統的演變始終伴隨著網路的發展，第一代資訊推薦系統使用傳統網站從以下三個來源收集資訊：來自購買或使用過的產品的基礎內容資料；使用者記錄中收集的人口統計資料；以及從用戶的項目偏好中收集的基於記憶的資料。第二代推薦系統通過收集社交資訊，例如朋友、關注者、粉絲等。第三代推薦系統使用網上集成設備提供的資訊。

14.3 人才概況

- 全球人才分佈

學者地圖用於描述特定領域學者的分佈情況，對於進行學者調查、分析各地區競爭力現況尤為重要，下圖為資訊檢索與推薦領域全球學者分佈情況：



圖 14-5 資訊檢索與推薦領域全球人才分佈

地圖根據學者當前就職機構地理位置進行繪製，其中顏色越深表示學者越集中。從該地圖可以看出，美國的人才數量優勢明顯且主要分佈於其東西海岸；歐

洲、亞洲也有較多的人才分佈；其他諸如非洲、南美洲等地區的學者非常稀少；資訊檢索與推薦領域的人才分佈與各地區的科技、經濟實力情況大體一致。此外，在性別比例方面，資訊檢索與推薦領域中男性學者占比 90.6%，女性學者占比 9.4%，男性學者占比遠高於女性學者。

資訊檢索與推薦領域學者的 h-index 分佈如下圖所示，大部分學者的 h-index 分佈在中低區域，其中 h-index 小於 20 區間的人數最多，有 870 人，占比 42.8%，50-60 區間的人數最少，有 82 人。

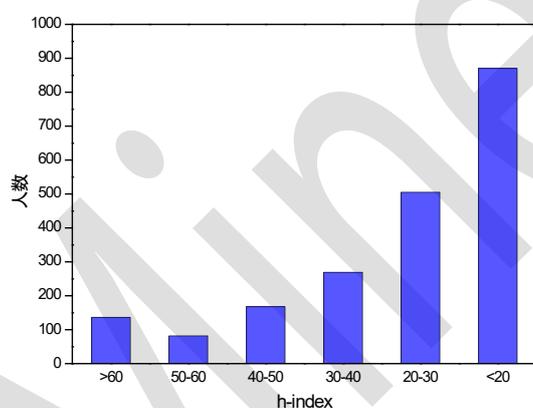


圖 14-6 資訊檢索與推薦領域學者 h-index 分佈

- 中國人才分佈

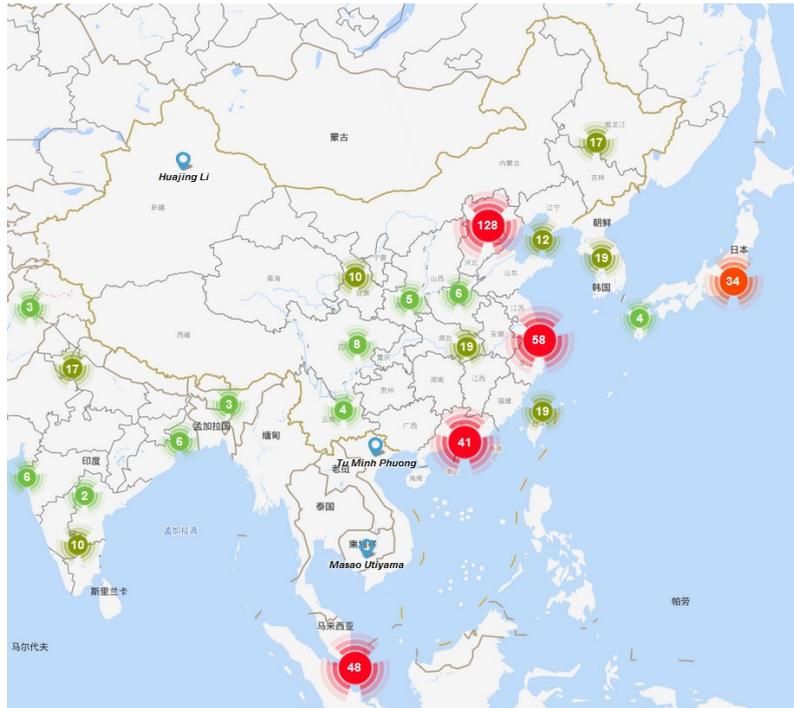


圖 14-7 資訊檢索與推薦領域中國學者分佈

我國專家學者在資訊檢索與推薦領域的分佈如上圖所示，從中可以發現京津地區在本領域的人才數量最多，其次是長三角和珠三角地區，相比之下，內陸地區的人才較為匱乏，這種分佈與區位因素和經濟水準情況不無關係。同時，通過觀察中國周邊國家的學者數量情況，特別是與日韓、東南亞等地相比，中國在資訊檢索與推薦領域學者數量較多且優勢較大。

中國與其他國家在資訊檢索與推薦領域的合作情況可以根據 AMiner 資料平臺分析得到，通過統計論文中作者的單位資訊，將作者映射到各個國家中，進而統計中國與各國之間合作論文的數量，並按照合作論文發表數量從高到低進行了排序，如下表所示。

表 14-1 資訊檢索與推薦領域中國與各國合作論文情況

合作國家	論文數	引用數	平均引用數	學者數
中國-美國	204	6858	34	443

中國-新加坡	49	1067	22	84
中國-英國	33	1057	32	44
中國-澳大利亞	32	746	23	63
中國-荷蘭	22	502	23	20
中國-加拿大	19	514	27	42
中國-日本	7	63	9	17
中國-印度	4	58	15	10
中國-希臘	4	47	12	7
中國-德國	3	34	11	11

從上表資料可以看出，中美合作的論文數、引用數、平均引用數以及學者數遙遙領先，表明中美間在資訊檢索與推薦領域合作之密切；此外，中國與歐洲的合作非常廣泛，前 10 名合作關係裡中歐合作共占 4 席。

14.4 論文解讀

本節對本領域的高水準學術會議及期刊論文進行挖掘，解讀這些會議和期刊在 2018-2019 年的部分代表性工作。這些會議和期刊包括：

International ACM SIGIR Conference on Research and Development in Information Retrieval

ACM Transactions on Information Systems

ACM Recommender Systems

我們對本領域論文的關鍵字進行分析，統計出詞頻 Top20 的關鍵字，生成本領域研究熱點的詞雲圖，如下圖所示。其中，推薦 (recommendation)、檢索 (retrieval)、排序學習 (learning to rank) 是本領域中最熱的關鍵字。



論文題目： *Adversarial Personalized Ranking for Recommendation*

中文題目： 對抗式個性化推薦排名

論文作者： Xiangnan He, Zhankui He, Xiaoyu Du and Tat-Seng Chua.

論文出處： The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '18

論文地址： <https://arxiv.org/pdf/1808.03908.pdf>

研究問題：

貝葉斯個性化排名 (Bayesian Personalized Ranking, BPR) 是一種成對學習的排序方法，用於優化個性化排序的推薦模型。它以內隱回饋學習為目標，假定觀察到的交互比未觀察到的交互排在更高的位置。矩陣因數分解 (Matrix Factorization, MF) 是最基本也是最有效的推薦模型。MF 將每個用戶和項表示為

嵌入向量，通過嵌入向量之間的內積來估計用戶對某一項的偏好程度。在資訊檢
 索領域，貝葉斯個性化排名訓練的矩陣分解模型 (MF-BPR) 學習一個與訓練資
 料相適應的複雜函數，不能很好地泛化，且其魯棒性較差，易受參數的對抗性擾
 動。因此本文提出了一種新的個性化排名訓練方法-對抗的個性化排名
 (Adversarial Personalized Ranking, APR) 。

研究方法：

以 BPR 為基礎，APR 中引入一個額外的目標函數，對其進行優化來量化推
 薦模型在參數擾動下的損失，使推薦模型既適合於個性化排序，又具有對抗性擾
 動的魯棒性。

$$L_{BPR}(\mathcal{D}|\Theta) = \sum_{(u,i,j) \in \mathcal{D}} -\ln \sigma(\hat{y}_{ui}(\Theta) - \hat{y}_{uj}(\Theta)) + \lambda_{\Theta} \|\Theta\|^2,$$

$$L_{APR}(\mathcal{D}|\Theta) = L_{BPR}(\mathcal{D}|\Theta) + \lambda L_{BPR}(\mathcal{D}|\Theta + \Delta_{adv}),$$

$$\text{where } \Delta_{adv} = \arg \max_{\Delta, \|\Delta\| \leq \epsilon} L_{BPR}(\mathcal{D}|\hat{\Theta} + \Delta),$$

上式和下式分別是 BPR 和 APR 的目標函數， Δ_{adv} 是對抗性擾動，旨在最大化
 BPR 目標函數的擾動。APR 可以看作是在玩一個極小極大的遊戲，在這個遊戲
 中，優化擾動使 BPR 損失最大化，並且在對抗擾動的情況下訓練模型使 BPR 損
 失和附加損失最小化。APR 指定了一個與模型無關的通用學習框架，只要底層模
 型是可微的，就可以在 APR 框架下使用反向傳播或者基於梯度的優化演算法訓
 練模型。

具體地，由於 APR 的目標函數中含有非線性函數，且訓練實例數目龐大，
 故使用隨機梯度下降法 (Stochastic Gradient Descent, SGD) 對 APR 進行優化。
 SGD 的思想是隨機選擇一個訓練實例，並只針對單個實例更新模型參數，因此

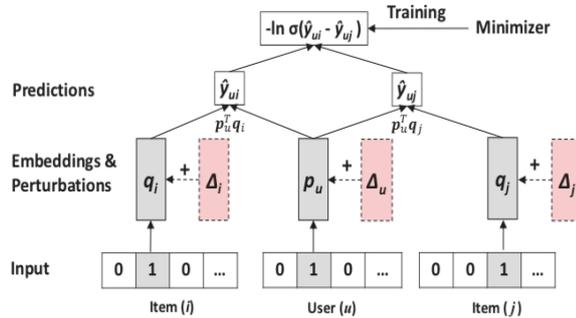
如何根據一個隨機採樣實例 (u, i, j) 優化模型參數是關鍵。本文提出的求解框架包括對抗性擾動構建和模型參數學習兩步，具體步驟詳見演算法 1。

Algorithm 1: SGD learning algorithm for APR.

Input: Training data \mathcal{D} , adversarial noise level ϵ , adversarial regularizer λ , L_2 regularizer λ_Θ , learning rate η ;

Output: Model parameters Θ ;

- 1 Initialize Θ from BPR;
- 2 **while** Stopping criteria is not met **do**
- 3 Randomly draw (u, i, j) from \mathcal{D} ;
- 4 // Constructing adversarial perturbations
- 4 $\Delta_{adv} \leftarrow$ Equation (8);
- 4 // Updating model parameters
- 5 $\Theta \leftarrow$ Equation (11);
- 6 **end**
- 7 **return** Θ



為了說明 APR 是如何工作的，本文提出了一個基於 MF 的推薦解決方案。首先用 BPR 訓練 MF，然後在 APR 框架下進一步優化它，因此把這種方法稱為對抗性矩陣分解 (AMF)。AMF 如上圖所示。由於 MF 的參數是用戶和項的嵌入向量，故對嵌入向量加以對抗性擾動，再將演算法 1 應用到 AMF 中，這需要對 AMF 進行小批量訓練，直到 AMF 達到收斂狀態或性能開始下降。

研究結果：

本文在 Yelp、Pinterest 和 Gowalla 三個公共資料集上進行了大量的實驗，這三個資料集分別代表不同的應用場景。定量分析和定性分析都證明了對個性化排名進行對抗性訓練的有效性和合理性。AMF 優於 MF-BPR，歸一化折現累積增益 (NDCG) 和命中率 (HR) 平均提高了 11%，它也優於最近提出的推薦模型，成為最前沿的推薦模型。

論文題目： *Neural Compatibility Modeling with Attentive Knowledge Distillation*

中文題目：基於注意力知識蒸餾的神經相容性建模

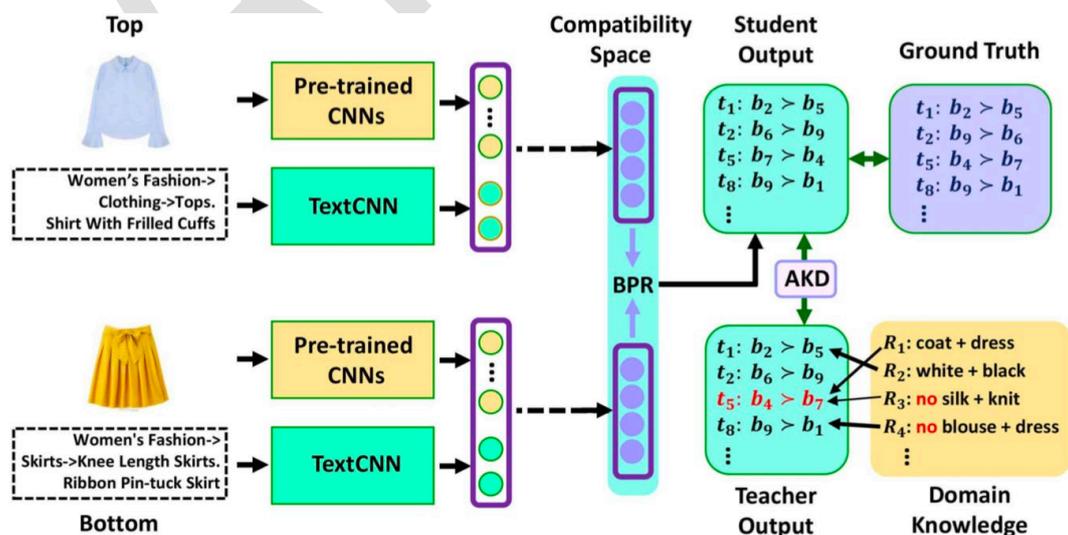
論文作者：Xuemeng Song, Fuli Feng, Xianjing Han, Xin Yang, Wei Liu and Liqiang Nie.

論文出處：The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '18

論文地址：<https://arxiv.org/pdf/1805.00313.pdf>

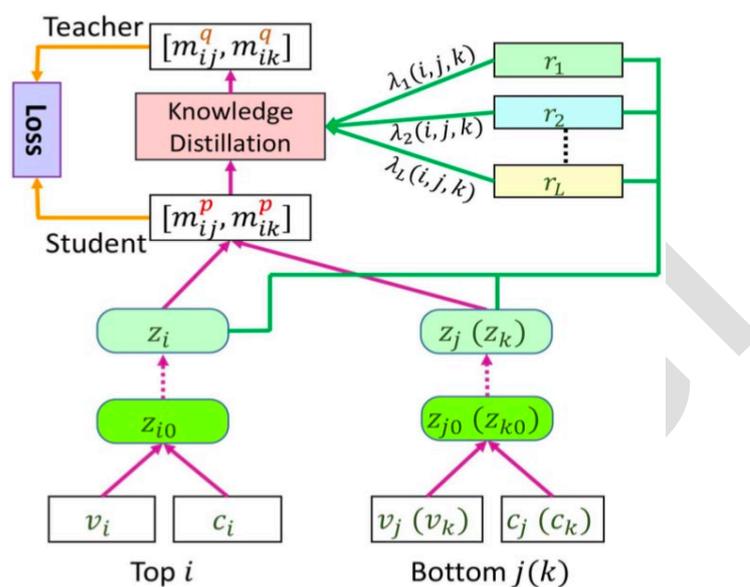
研究問題：

服裝搭配與人們的日常生活息息相關，現有研究大多依賴深度神經網路來提取時尚單品的有效表徵來解決服裝搭配問題。但作為純資料驅動方法的神經網路不僅具有較差的解釋性，而且也忽視了搭配領域知識。即使從資料驅動和知識驅動的角度對服裝單品之間的相容性進行全面建模也面臨許多挑戰，如搭配領域知識是無結構且模糊的，如何將搭配規則無縫地編碼到純資料驅動的學習框架中以及對於不同的樣本知識規則可能表現出不同的置信度，從而提供不同的搭配指導。為此，本文提出了一種基於注意力知識蒸餾的神經相容性建模方法（AKD-DBPR）。



研究方法：

AKD-DBPR 能夠從特定資料樣本和一般領域知識中學習，採用教師-學生模式來整合領域知識 (教師) 並提高神經網路 (學生) 的性能。其基本思想類似於人類教育，教師知道幾個專業規則，因此教師可以用自己對特定問題的解決方案指導學生。



學生網路 p 作為一種純資料驅動的神經網路，其目標是學習一個隱含的相容空間，用雙路徑神經網路將來自異構空間的時尚單品統一起來。為了對不同模式之間的相容性和語義關係全面建模，學生網路通過在視覺和語境表示的連接向量上添加隱含層，無縫地集成時尚單品的視覺和語境模式。此外，為了更好地描述時尚單品之間的相對相容性，構建基於貝葉斯個性化排名 (BPR) 框架的學生網路來研究互補單品之間的配對偏好。同時，用一組靈活的結構化邏輯規則對領域知識進行編碼，並利用正則化器將這些知識規則編碼到教師網路 q 中。但不同的規則對於不同的樣本可能有不同的置信水準，因此引入注意力機制來分配規則置信度，進一步用於指導學生網路的訓練。上圖是注意力知識蒸餾的流程。 v 和 c 分別表示單品的視覺和上下文語境， $m_{i,j}$ 表示上裝 i 和下裝 j 的相容性， r 表示規

則。最後，鼓勵學生網路達到良好的相容性建模性能，而且能很好地類比規則正則化的教師網路。

研究結果：

在真實資料集上進行的大量實驗證明了 AKD-DBPR 在服裝搭配領域具有良好性能且能應用到互補時尚單品檢索的實踐中，除此之外也證明了引入注意機制有助於克服人為定義的模糊規則的局限性。

論文題目： *Improving Sequential Recommendation with Knowledge-Enhanced Memory Networks*

中文題目：基於知識增強記憶網路的序列推薦

論文作者：Jin Huang, Wayne Xin Zhao, Hong-Jian Dou, Ji-Rong Wen and Edward Y. Chang.

論文出處：The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '18

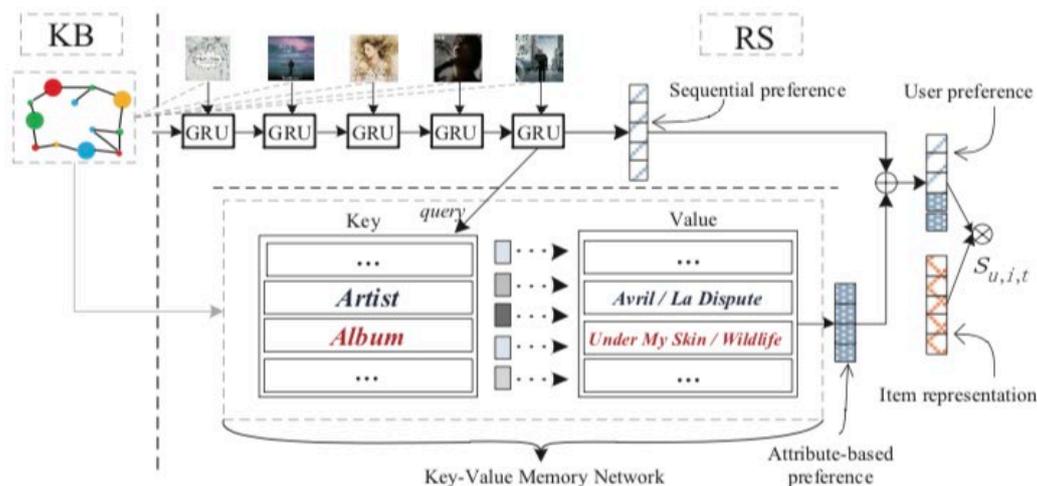
論文地址：<https://sci-hub.tw/10.1145/3209978.3210017>

研究問題：

推薦系統可以為使用者推薦其感興趣的內容並給出個性化的建議。基於 RNN 的網路可以將歷史交互記錄編碼為隱藏狀態向量，但是它很難從交互序列中捕獲細微性的用戶偏好，而且隱向量表示的可解釋性也較差。為了以一種可解釋的方式增強細微性使用者偏好建模的能力，本文提出一種知識增強的序列推薦 (KSR) 模型。

研究方法：

KSR 模型將基於 RNN 的網路 (GRU) 與鍵值記憶網路 (KV-MNs) 相結合來增強推薦系統的特徵捕獲能力與解釋性。GRU 部分用於捕捉使用者序列偏好特徵，而 KV-MNs 用於捕捉基於屬性的使用者偏好特徵。該模型如下圖所示。

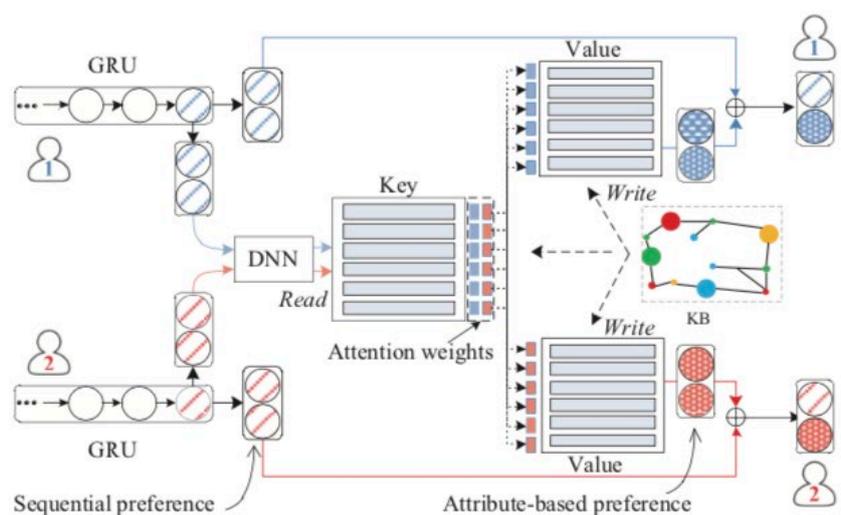


給定一組用戶 u 的交互序列 $\{i_1, \dots, i_t\}^1$ ，在表示好 GRU 模型的隱藏層和使用者 u 的序列偏好向量之後，對每個物件使用基於貝葉斯後驗優化的個性化排序演算法 (BPR) 進行預訓練，再通過下式計算排名得分對候選物件排名，將得分最高的物件推薦給目標使用者。

$$s_{u,i,t} = g(u, i, t) = \mathbf{h}_t^{u\top} \cdot \mathbf{q}_i,$$

KSR 模型的整體工作機制如上圖所示。基於 RNN 的知識增強序列推薦模型的短期記憶能力有限，不適合長期存儲知識庫資訊，因此通過 KV-MNs 來整合知識庫知識。KV-MNs 將鍵向量設置為從知識資料中學習的嵌入關係，對應於實體屬性。此外，對於給定的鍵向量設置一個特定於使用者的值向量來存儲相應屬性的使用者偏好特徵。通過這種方式，外部知識庫知識被有效地整合到 KV-MNs 中。準備好 KV-MNs 後，將其與基於 RNN 的序列推薦模型集成。在每次推薦時

使用來自 RNNs 的序列偏好作為查詢來讀取特定於用戶的 KV-MNs 值向量相關內容。值向量被組合到具有注意力權值的基於屬性的偏好表示中，基於屬性的偏好表示與序列偏好表示相結合形成用戶偏好的最終表示。



研究結果：

本文在四個推薦系統資料集上進行實驗，結果表明 KSR 模型在每個資料集的推薦性能都比表現最好的基準推薦模型有所提升，能夠生成更高品質的序列推薦。此外，還定量分析了 KSR 模型的可解釋性，結果表明該模型具有高度的可解釋性。

論文題目：*Equity of Attention: Amortizing Individual Fairness in Rankings*

中文題目：注意力的公平性：在排名中平攤個體公平

論文作者：Asia J. Biega, Krishna P. Gummadi and Gerhard Weikum

論文出處：The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '18

論文地址：<https://arxiv.org/pdf/1805.01788.pdf>

研究問題：

從招聘網站到共用經濟平臺，人員和物件的排名是選拔、配對和推薦系統的核心。由於排名位置會影響被排名物件受到的關注度，排名中的偏差會導致機會和資源的不公平分配。因此，本文提出了新的措施和機制來量化和減輕所有排名的不公平性。一個物件得到的關注應該與其相關性成正比並且注意力在排名中需公平分配。所提出的方法關注的是個體物件層面的公平，並將群體公平作為一個特例納入其中，還設計了一種新的機制來確保排名中注意力平攤的公平性。

研究方法：

物件的關注受位置偏差的影響較大，即多個具有相似相關性的物件並沒有得到相同的排名位置和相近的關注度。本文定義的考慮位置偏差的注意力公平 (equity of attention) 認為一個序列排名中每個物件獲得的累積注意力 (A) 與其累積相關性 (R) 成比例。注意力公平是個體層次的，當個體的注意力達到公平時也會在群體層次上達到公平。注意力公平定義如下：

$$\frac{\sum_{l=1}^m a_{i1}^l}{\sum_{l=1}^m r_{i1}^l} = \frac{\sum_{l=1}^m a_{i2}^l}{\sum_{l=1}^m r_{i2}^l}, \forall u_{i1}, u_{i2}$$

下式將不公平度量為累積注意力和累積相關性之間的距離(ρ^1, \dots, ρ^m 是一個序列排名)：

$$unfairness(\rho^1, \dots, \rho^m) = \sum_{i=1}^n |A_i - R_i| = \sum_{i=1}^n \left| \sum_{j=1}^m a_i^j - \sum_{j=1}^m r_i^j \right|$$

當相關程度較低的物件排名高於相關程度較高的物件排名時，為了滿足公平標準而對排名進行過濾可能會導致品質損失。因此使用 IR 評估的度量指標來量化排名品質。解決公平性的方法是在原有排名的基礎上重新排名，可將原始排名

ρ 作為參考來評估重新排名 ρ^* 的品質。即通過量化新排名 ρ^* 與原始排名 ρ 的差異來量化排名品質，如下式所示：

$$NDCG\text{-quality}@k(\rho, \rho^*) = \frac{DCG@k(\rho^*)}{DCG@k(\rho)}$$

為了提高公平性，對基於相關性的排名加以擾動，這會導致排名品質的下降，為解決此問題，可通過在品質約束下最小化不公平（即對可接受的排名品質設置下限）來對二者進行權衡。

將此方法轉化為線上優化問題，排名攤銷需要以線上方式完成，在不瞭解未來查詢負載的情況下，就要對當前排名進行重新排序，從而在當前排名品質受到約束的情況下，將排名中累積注意力和相關性分佈的不公平性降到最低。此線上優化問題可以通過整型線性規劃（ILP）來解決。假設要在一系列排名中重新排列第 1 位，引入 n^2 個決策變數 $X_{i,j}$ ，如果物件分配到排名位置 j ，設置決策變數的值為 1，否則為 0。對每個物件 u_i ，累積注意力和相關性分別初始化為 $A_i^0 = 0$ ， $R_i^0 = 0$ ，ILP 定義如下：

$$\begin{aligned} & \text{minimize} && \sum_{i=1}^n \sum_{j=1}^n |A_i^{l-1} + w_j - (R_i^{l-1} + r_j^l)| \cdot X_{i,j} \\ & \text{subject to} && \sum_{j=1}^k \sum_{i=1}^n \frac{2^{r_i^l} - 1}{\log_2(j+1)} X_{i,j} \geq \theta \cdot IDCG@k \\ & && X_{i,j} \in \{0, 1\}, \forall i, j \\ & && \sum_i X_{i,j} = 1, \forall j \\ & && \sum_j X_{i,j} = 1, \forall i \end{aligned}$$

第一項約束限制了排名品質損失，其他約束確保解是物件到排名位置的雙向映射。當且僅當物件 u_i 映射到位置 j 時， $A_i^{l-1} + w_j$ 和 $R_i^{l-1} + r_j^l$ 這兩項分別更新累

積注意力和相關性。還需注意的是 ILP 運行在一個巨大的組合空間中，所以又設計了篩檢程式來精簡 ILP 的組合空間。

研究結果：

實驗驗證在兩類資料集上進行，一是通過三個人工合成資料集來分析模型在不同相關性分佈下（均勻分佈、線性分佈和指數分佈）的性能，二是分析模型在真實場景中的性能，資料集採用世界不同地域三個城市的 Airbnb 公寓排名和基於 StackExchange 查詢日誌和文檔集合構建的資料集。實驗結果表明了提高注意力的公平性的重要性，在共用經濟或排名影響人們經濟生活的市場平臺中尤為重要，而且這是在不犧牲排名品質的情況下做到的。

論文題目：*Impact of Item Consumption on Assessment of Recommendations in User Studies*

中文題目：關於使用者是否消費過物品對推薦系統的評價影響研究

論文作者：Benedikt Loepp, Tim Donkers, Timm Kleemann, Jürgen Ziegler

論文出處：ACM Recommender Systems

論文地址：<https://doi.org/10.1145/3240323.3240375>

研究問題：

在對推薦系統的使用者研究中，參與者通常不能消費推薦的商品。儘管如此，他們還是被要求通過問卷的方式來評估推薦品質和與用戶體驗相關的其他方面。然而，如果沒有聽過推薦的歌曲或看過推薦的電影，這可能是一個很容易出錯的

任務，可能會限制這些研究結果的有效性。在這篇文章中，作者調查了實際消費的推薦專案的影響。

研究方法：

作者提出了兩項在不同領域進行的用戶研究，研究表明在某些情況下，推薦評估和問卷調查結果存在差異。顯然，在不允許使用者消費專案的情況下，並不總能充分衡量出用戶體驗。另一方面，根據領域和提供的資訊，參與者有時似乎很好地估計了推薦的實際價值。

一項是在音樂領域，用聽歌做測試(記為 S)，一項是在電影領域(記為 M)。通過對兩組 40 名平均年齡為二十幾歲的男女大致均衡的參與者，分別進行了音樂領域和電影領域的對照實驗。控制的變數條件是消費前(前)和消費後(後)進行問卷調查，以及僅在消費後(後)進行問卷調查。

研究結果：

實驗結果認為，我們有必要對推薦系統的問卷調查結果有所保留(即保持懷疑態度)。在某些情況下，如果不使用推薦的物品，參與者無法充分評估推薦系統的各個方面，特別是與用戶體驗相關的方面。

舉例來說，比如音樂領域的參與者通常傾向於低估歌曲，並且在只能從包含描述性資訊做判斷，而不能夠聽歌時，對被要求在推薦列表中進行選擇時滿意度較低。聽歌對滿意度會產生正向影響，導致相關問卷項目的得分明顯更高。比如在對選擇滿意度的研究中表明人們會傾向於高估過去事件的影響，並且在某些情況下，幾周後滿意度會下降，因此有必要調查這些結果的穩定性。主觀系統方面，如感知的召回品質的評價與是否消費過無關。對於電影而言，這在更多方面

似乎是正確的，尤其是在可以使用高品質的文字說明的情況下，與具有抽象情感內容的音樂相比，電影的推薦更生動具體。確實一些為人熟知的歌曲也有電影的這種效果，這會帶來一些推薦的偏差。還有就是，短電影在這方面是完全未知的，被人耳熟能詳的歌曲數量也不多，所以也沒有產生出相關性。但對物品的熟悉程度及其對評估推薦的影響還是應該進行更詳細的研究。另外，需要更多地考慮廣泛使用問卷的一般影響，因為有意識地思考決策並不總是有益的（對於決策結果來說，結果越直觀、簡單越好）。

總的來說，定量結果表明，實際經驗是否足以被替代，在很大程度上取決於推薦的領域，以及推薦系統提供的資訊的類型和數量。本文通過研究定性評論反映了這一點。因此，作者建議避免不同設置之間的比較（注意控制變數），並在要求參與者在不消費的情況下對推薦進行評級時，要對用戶實驗更加注意才行（注重實踐和對結果的把控）還有，現實世界中提供的物品評分不會抑制評分，例如在看電影或烹飪食譜之前的評分，應該與看過後以及吃過後的評分謹慎的有機結合。只有當提供了足夠資訊的時候，參與者的反應才可能是可靠的，在這種情況下，可能就不需要（使用者實際）消費過這個物品了。因為他們已經在心裡形成了一個模型，考慮了自己的喜好，使這個建議像他已經真的買了或用了一樣可靠（也就是自己說服自己）。

下圖是作者針對條件（S1, S2/M1, M2）和時間點（事前，事後）交互作用的使用者研究的混合模型結果。正差異結果表明用戶消費後的結果會更好。

Table 1: Results of our mixed models for both user studies for interaction of condition (S1, S2/M1, M2) and point in time (Pre, Post). Positive differences indicate better results after consumption (*Choice Diff.*, *Effort* and *Doubts* are reversed accordingly).

Study 1	Interaction			S1-Pre vs. S1-Post			S1-Pre vs. S2-Post			Study 2	Interaction			M1-Pre vs. M1-Post			M1-Pre vs. M2-Post		
	Sig.	Est. Diff.	Std. Err.	Sig.	Est. Diff.	Std. Err.	Sig.	Est. Diff.	Std. Err.		Sig.	Est. Diff.	Std. Err.	Sig.	Est. Diff.	Std. Err.	Sig.	Est. Diff.	Std. Err.
Perceived Rec. Quality [11]	.390	0.38	0.28	.183	0.15	0.29	.611			.467	-0.14	0.17	.411	-0.27	0.27	.328			
Mean Recommendation Rating	.009*	0.59	0.18	.004*	0.30	0.24	.226			.771	-0.08	0.14	.578	-0.11	0.21	.574			
Choice Satisfaction [11]	.000*	0.71	0.21	.003*	1.29	0.28	.000*			.020*	-0.19	0.25	.450	0.03	0.35	.937			
Choice Difficulty [11]	.001*	1.14	0.29	.001*	0.55	0.38	.156			.968	0.05	0.31	.877	-0.05	0.37	.905			
Effort [11]	.415	0.21	0.16	.196	0.10	0.23	.664			.012*	-0.07	0.08	.383	-0.47	0.15	.003*			
Effectiveness [11]	.000*	0.81	0.19	.000*	1.08	0.33	.002*			.479	-0.14	0.22	.520	-0.41	0.34	.229			
Diversity [11]	.056	-0.38	0.26	.151	0.42	0.31	.184			.117	0.24	0.19	.224	-0.37	0.34	.288			
Novelty [15]	.288	-0.19	0.13	.144	0.11	0.30	.731			.218	0.14	0.09	.106	0.14	0.20	.472			
Information Sufficiency [15]	.000*	1.48	0.38	.000*	1.67	0.38	.000*			.041*	-0.33	0.23	.149	-0.37	0.32	.250			
Transparency [15]	.104	0.48	0.22	.051	0.61	0.38	.113			.763	-0.14	0.21	.499	-0.16	0.36	.658			
Confidence and Trust [15]	.017*	0.54	0.20	.014*	0.64	0.26	.020*			.787	0.04	0.16	.826	-0.18	0.28	.527			
Doubts	.000*	2.19	0.33	.000*	1.71	0.38	.000*			.680	-0.14	0.27	.605	-0.29	0.35	.407			
Overall Satisfaction [15]	.005*	0.62	0.20	.005*	0.89	0.31	.007*			.442	-0.14	0.22	.525	-0.36	0.30	.235			

論文題目： *Should I Follow the Crowd? A Probabilistic Analysis of the Effectiveness of Popularity in Recommender Systems*

中文題目：我應該跟隨潮流（人群）嗎？推薦系統中流行度有效性的概率分析

論文作者：Rocío Cañamares, Pablo Castells

論文出處：International ACM SIGIR Conference on Research and Development in Information Retrieval

論文地址：<https://dl.acm.org/citation.cfm?id=3210014>

研究問題：

近年來，在推薦系統的評價中使用 IR 方法已經成為一種普遍的做法。然而，人們發現 IR 指標對推薦熱門物品的獎勵演算法存在強烈的偏差 (bias)，這與目前最先進的推薦演算法存在的偏差是相同的。最近也有一些研究證實並測量了這種偏差，並提出了避免這種偏差的方法。

可是，最根本的問題仍懸而未決——一個物品的流行度 (popularity) 的偏差是否是我們應當避免的；無論這個偏差是一個對於推薦來說的好信號，還是可能因為受到實驗偏差而導致的壞信號。

研究方法：

論文方法是根據關鍵隨機變數之間的依賴關係，識別和建模可以確定答案的條件，包括物品評級、發現和相關性等。作者發現了保證流行有效或完全相反的條件，以及度量值反映真實有效性或在品質上偏離真實有效性的條件。

作者構建了一個眾包資料集，其中沒有公共可用資料顯示的常見偏差，在這個資料集中，作者說明了在一個常見的有偏差的離線實驗設置中所能測量的準確性與通過無偏差的觀察所能測量的實際準確性之間的矛盾。

研究結論：

本文作者通過研究證實了普遍流行的有效性趨勢，並用公式證明與解釋了原因。同時作者還發現，在許多情況下表觀的準確度可能會產生誤導（即與真實準確度不匹配）。這大多是因為推薦的物品出現與使用者興趣相差較大的情況。作者發現，通常的實驗（即觀察到的準確度為度量值）可能對平均評級（average rating）及其個性化衍生品（personalized derivatives）相當不公平。與目前文獻中觀察到的結果相反，平均評級在大多數情況下，相對於正評級（positive ratings），在真實的準確率上可能更好、更安全，魯棒性更強。

作者進一步發現，當涉及到流行度的有效性或平均評分及其度量時，喜歡的物品被評分的次數多少並不重要，重要的是評級是否依賴相關性，依賴是完全的還是部分的；當發現主要依賴於相關性，或者幾乎不依賴於相關性時，流行度和平均評分才是真正準確的理想條件。平均評級似乎比流行度對相關性獨立性的影響更大，因此在有高偏差的情況下（如行銷驅動），它比流行度更可取。

論文題目：*Variance Reduction in Gradient Exploration for Online Learning to Rank*

中文題目：一種基於方差縮減梯度搜索的線上學習排序方法

論文作者：Huazheng Wang, Sonwoo Kim, Eric McCord-Snook, Qingyun Wu, Hongning Wang

論文出處：International ACM SIGIR Conference on Research and Development in Information Retrieval

論文地址：<https://dl.acm.org/citation.cfm?id=3331264>

研究問題：

線上學習排名 (OL2R) 演算法從使用者的即時回饋中學習，演算法的關鍵是對梯度的無偏估計，通常是通過從整個參數空間均勻採樣來實現的。然而，這導致了梯度估計的高方差，會使模型更新時的效果不佳，特別是在參數空間的維數較大的時候。本文旨在降低 OL2R 演算法中梯度估計的方差。

研究方法：

在交叉測試之後，作者將選擇的更新方向 (如目標方向 winning direction) 投影到當前查詢下被檢索 (examined) 文檔的特徵向量所跨越的空間 (簡稱為“文檔空間 document space”)。作者的主要觀點是，交叉測試的結果完全由用戶對所檢索文檔的相關性評估控制。因此，該測試引入的真實梯度只反映在構建的文檔空間中，為了減少方差，我們可以安全地刪除與文檔空間正交的梯度分量。作者證明了這個投影梯度仍然是一個真實梯度的無偏估計，並且證明了這種低方差梯度估計能夠顯著減少 regret。

在本文中，作者提出並發展了文檔空間投影 (Document Space Projection, DSP) 方法來減少梯度估計中的方差，提高線上學習的排序性能。DSP 的核心思想是認識到交叉測試只揭示了真實梯度在被檢文檔的跨空間上的投影。包含任何

超出此空間的模型更新只會引入雜訊。因此，作者將選擇的模型更新方向投射回文檔空間以減少方差。作者同時證明了 DSP 保持著一個無偏的梯度估計，並且通過減少方差可以顯著提高 DBGD 類型演算法的 regret 界。通過大量的實驗，發現 DSP 能夠在方差減少和整體性能方面，特別是在排序特徵數量較大的情況下，對幾種最先進的 OL2R 模型提供統計上顯著的改進。下表是線上和離線 NDCG@10，經過一萬次查詢，每種演算法的文檔空間投影的標準差和相對改進。

Table 2: Online NDCG@10, standard deviation and relative improvement of document space projection of each algorithm after 10,000 queries.

Click Model	Algorithm	MQ2007	MQ2008	MSLR-WEB10K	NP2003	Yahoo
Perfect	DBGD	679.3 _(21.4)	847.1 _(38.4)	532.2 _(15.3)	1130.2 _(43.3)	1165.5 _(22.4)
	DBGD-DSP	689.1 _(19.9) (+1.44%)	858.0 _(39.2) (+1.29%)	553.6 _(13.1) (+4.02%)	1198.8 _(40.0) (+6.07%)	1198.8 _(31.5) (+2.86%)
	MGD	689.1 _(34.4)	859.4 _(38.1)	558.3 _(7.9)	1192.9 _(44.6)	1201.9 _(36.9)
	MGD-DSP	757.3 _(16.2) (+9.90%)	919.5 _(42.2) (+6.99%)	626.4 _(9.6) (+12.20%)	1335.3 _(39.1) (+11.94%)	1309.4 _(10.6) (+8.94%)
	NSGD	684.4 _(20.9)	867.5 _(40.3)	589.5 _(14.2)	1274.9 _(47.4)	1162.3 _(32.9)
	NSGD-DSP	732.5 _(20.0) (+7.03%)	904.3 _(38.0) (+4.24%)	635.6 _(12.8) (+7.82%)	1368.5 _(41.1) (+7.34%)	1270.1 _(25.5) (+9.27%)
Navigational	DBGD	646.1 _(23.4)	817.9 _(45.5)	517.5 _(20.9)	1062.3 _(55.4)	1133.3 _(40.0)
	DBGD-DSP	664.9 _(26.9) (+2.91%)	830.3 _(44.1) (+1.52%)	543.1 _(14.4) (+4.95%)	1140.1 _(32.5) (+7.32%)	1199.4 _(34.6) (+5.83%)
	MGD	632.7 _(15.5)	827.5 _(35.5)	538.2 _(7.2)	1115.4 _(44.6)	1171.3 _(20.4)
	MGD-DSP	694.5 _(15.7) (+9.77%)	882.3 _(40.0) (+6.62%)	586.9 _(9.5) (+9.05%)	1300.9 _(39.6) (+16.63%)	1290.2 _(15.3) (+10.15%)
	NSGD	660.1 _(24.5)	849.1 _(16.6)	562.1 _(13.8)	1211.1 _(66.5)	1186.2 _(36.0)
	NSGD-DSP	724.6 _(24.5) (+9.77%)	895.8 _(34.2) (+5.50%)	608.3 _(12.1) (+8.22%)	1296.2 _(24.1) (+7.03%)	1283.4 _(7.2) (+8.19%)
Informational	DBGD	583.4 _(46.0)	763.9 _(55.1)	472.4 _(14.6)	849.8 _(34.5)	1107.3 _(46.4)
	DBGD-DSP	620.1 _(40.6) (+6.29%)	782.4 _(11.6) (+2.42%)	522.1 _(13.6) (+10.52%)	992.5 _(81.1) (+16.79%)	1158.5 _(22.0) (+4.62%)
	MGD	621.2 _(18.2)	817.5 _(45.3)	538.3 _(10.8)	1107.9 _(46.2)	1146.6 _(37.5)
	MGD-DSP	671.4 _(18.9) (+8.08%)	865.9 _(37.7) (+5.92%)	580.5 _(10.4) (+7.84%)	1274.5 _(42.9) (+15.04%)	1268.1 _(16.4) (+10.60%)
	NSGD	629.7 _(25.9)	814.9 _(37.1)	532.9 _(15.2)	1123.5 _(59.8)	1110.5 _(30.9)
	NSGD-DSP	703.6 _(29.2) (+11.74%)	871.3 _(48.1) (+6.92%)	597.9 _(14.1) (+12.20%)	1222.8 _(41.1) (+9.03%)	1204.7 _(5.6) (+8.48%)

Table 3: Offline NDCG@10, standard deviation and relative improvement of document space projection of each algorithm after 10,000 queries.

Click Model	Algorithm	MQ2007	MQ2008	MSLR-WEB10K	NP2003	Yahoo
Perfect	DBGD	0.484 _(0.023)	0.683 _(0.020)	0.331 _(0.009)	0.737 _(0.056)	0.688 _(0.011)
	DBGD-DSP	0.480 _(0.020) (-0.83%)	0.685 _(0.024) (+0.29%)	0.333 _(0.011) (+0.6%)	0.738 _(0.039) (+0.14%)	0.681 _(0.011) (-1.02%)
	MGD	0.495 _(0.022)	0.691 _(0.020)	0.334 _(0.003)	0.746 _(0.048)	0.715 _(0.002)
	MGD-DSP	0.501 _(0.021) (+1.21%)	0.695 _(0.021) (+0.58%)	0.409 _(0.006) (+22.46%)	0.748 _(0.051) (+0.27%)	0.725 _(0.003) (+1.40%)
	NSGD	0.488 _(0.019)	0.689 _(0.024)	0.397 _(0.012)	0.743 _(0.030)	0.691 _(0.005)
	NSGD-DSP	0.491 _(0.022) (+0.61%)	0.691 _(0.025) (+0.29%)	0.398 _(0.008) (+0.25%)	0.750 _(0.042) (+0.94%)	0.717 _(0.004) (+3.76%)
Navigational	DBGD	0.463 _(0.020)	0.667 _(0.021)	0.320 _(0.012)	0.728 _(0.054)	0.663 _(0.020)
	DBGD-DSP	0.465 _(0.021) (+0.43%)	0.668 _(0.025) (+0.15%)	0.327 _(0.011) (+2.19%)	0.734 _(0.052) (+0.82%)	0.656 _(0.013) (-1.06%)
	MGD	0.426 _(0.019)	0.664 _(0.016)	0.321 _(0.003)	0.740 _(0.048)	0.703 _(0.010)
	MGD-DSP	0.467 _(0.021) (+9.62%)	0.684 _(0.017) (+3.01%)	0.331 _(0.005) (+3.12%)	0.744 _(0.051) (+0.54%)	0.714 _(0.006) (+1.56%)
	NSGD	0.473 _(0.022)	0.676 _(0.024)	0.389 _(0.013)	0.732 _(0.051)	0.686 _(0.008)
	NSGD-DSP	0.478 _(0.020) (+1.06%)	0.683 _(0.020) (+1.04%)	0.376 _(0.014) (-3.34%)	0.788 _(0.006) (+7.65%)	0.711 _(0.001) (+3.64%)
Informational	DBGD	0.410 _(0.034)	0.641 _(0.013)	0.294 _(0.023)	0.699 _(0.061)	0.623 _(0.017)
	DBGD-DSP	0.427 _(0.027) (+4.15%)	0.632 _(0.011) (-1.4%)	0.309 _(0.011) (+32.65%)	0.692 _(0.062) (-1.00%)	0.63 _(0.030) (1.12%)
	MGD	0.406 _(0.030)	0.651 _(0.020)	0.317 _(0.003)	0.726 _(0.030)	0.668 _(0.014)
	MGD-DSP	0.444 _(0.025) (+4.44%)	0.669 _(0.018) (+0.67%)	0.325 _(0.004) (+0.33%)	0.738 _(0.054) (+0.74%)	0.701 _(0.005) (+4.94%)
	NSGD	0.469 _(0.018)	0.674 _(0.023)	0.360 _(0.013)	0.733 _(0.056)	0.663 _(0.015)
	NSGD-DSP	0.466 _(0.019) (-0.64%)	0.668 _(0.026) (-0.89%)	0.340 _(0.010) (-5.56%)	0.789 _(0.013) (+7.64%)	0.685 _(0.004) (+3.32%)

研究結果：

本文通過與幾種最先進的 OL2R 演算法的大量實驗比較，驗證了作者提出的方法在減少方差和提高整體排名性能方面的有效性（如表 2 和 3 所示）。但是觀察到在模擬用戶點擊回饋時，DSP 在不同點擊模式下的性能有所不同。未來，作者計畫合併不同的點擊模式解決方案，以更精確地構建文檔空間。在交叉測試之前，研究如何進行基於文檔空間的探索性方向生成也是有意義的。探索性方向預選有望進一步加速梯度探索，提高線上學習過程中的用戶滿意度，不過同時也必須確保它是無偏差的。

14.5 資訊檢索與推薦進展

隨著互聯網中數位資訊數量的增長，商品、書籍、新文章、歌、電影、研究檔等日常基礎性事物，其數量和種類填滿了多個資料倉庫和資料庫。蘊含著智慧推薦系統和強大的搜尋引擎的線上商店、線上音樂、線上視頻和圖片庫等已成為人們快速尋找資訊的主要方式。此類系統的流程度和有用性在於它們能夠便捷地顯示幾乎無限的物品資訊。比如，Amazon、Netflix 等推薦系統嘗試瞭解使用者興趣，並向用戶推薦他們感興趣的商品。儘管這些系統由於使用場景而各不相同，但其尋找用戶感興趣商品的核心機制都是用戶興趣與商品匹配的機制。

為了提高資訊檢索與推薦系統中演算法模型的準確性和可解釋性，研究人員近年來主要關注無偏的線上排序學習模型，以及利用知識資訊增強推薦系統的表現和可解釋性等方面的研究。其中，無偏的線上排序學習模型是指自動利用大規模使用者點擊資料訓練搜索結果的排序模型。使用者點擊資料是現代搜尋引擎的重要資料來源，具有成本低廉，並且對以用戶為中心的檢索應用程式（如搜索排名）特別有用等優點。為了充分利用使用者點擊資料開發一個無偏的學習排名系

統，研究人員試圖消除使用者偏見對排名模型訓練的影響。近年來，一種基於反事實學習和圖形模型的無偏學習排名框架引起了人們的廣泛關注。該框架側重於使用反事實學習直接訓練帶有偏倚點擊資料的排名模型。這個無偏的學習排名框架對待點擊偏差作為一個反事實的影響和去偏用戶回饋加權每點擊與他們的反向傾向加權。它使用傾向性模型來量化點擊的偏差，並沒有明確地估計查詢文檔與培訓資料的相關性。研究人員從理論上證明，在正確的偏差估計下，在該框架下使用點擊資料訓練的排序模型將收斂於使用真實相關信號訓練的排序模型。

資訊檢索與推薦系統可以為使用者推薦其感興趣的內容並給出個性化的建議。而現在的推薦系統大都著眼於被推薦物件的序列建模，而忽略了它們細細微性的特徵。為了解決以上問題，研究人員提出了多工可解釋推薦模型(Multi-Task Explainable Recommendation, MTER) 和知識增強的序列推薦模型 (Knowledge-enhanced Sequential Recommender, KSR)。其中，MTER 模型是一個用於可解釋推薦任務的多工學習方法，通過聯合張量分解將使用者、產品、特徵和觀點短語映射到同一向量空間，來從使用者評論中提取產品細細微性的個性化特徵。KSR 模型提出了利用結合知識庫的記憶網路來增強推薦系統的特徵捕獲能力與解釋性，解決序列化推薦系統不具有解釋性，且無法獲取使用者細細微性特徵的不足。MTER 和 KSR 模型通過對推薦結果的解釋，分析被推薦物件的特徵，可以讓使用者可以對使用哪些推薦結果做出更明智，更準確的決策，從而提高他們的滿意度。

近年來，資訊檢索與推薦領域比較流行的開源平臺主要包括基於深度學習的檢索模型 (MatchZoo)、基於 tensorflow 的 learning to rank 模型 (TF-Ranking) 和 microsoft recommenders。其中，MatchZoo 是由中國科學院計算技術研究所網

路資料科學與技術重點實驗室近期發佈的深度文本匹配開源專案。MatchZoo 是一個 Python 環境下基於 TensorFlow 開發的開源文本匹配工具，使用了 Keras 中的神經網路層，並有資料預處理，模型構建，訓練與評測三大模組組成，旨在讓大家更加直觀地瞭解深度文本匹配模型的設計、更加便利地比較不同模型的性能差異、更加快捷地開發新型的深度匹配模型。MatchZoo 提供了基準資料集(TREC MQ 系列資料、WiKiQA 資料等) 進行開發與測試，整合了當前最流行的深度文本匹配的方法 (包括 DRMM, MatchPyramid, DUET, MVLSTM, aNMM, ARC-I, ARC-II, DSSM, CDSSM 等演算法的統一實現)，旨在為資訊檢索、資料採擷、自然語言處理、機器學習等領域內的研究與開發人員提供便利，可以應用到的任務場景包括文本檢索，自動問答，複述問題，對話系統等等。

TF-Ranking 是一個可擴展的基於 tensorflow 的用於排序的庫，由 google 於 2018 年提出。TF-Ranking 提供了一個統一的框架，其中包括一套最先進的學習排序演算法，並支援成對或清單損失函數、多項評分、排序度量優化和無偏學習排序。TF-Ranking 速度很快並且易於使用，可以創建高品質的排序模型。統一的框架使機器學習的研究人員、實踐者和愛好者能夠在一個庫中評估和選擇一系列不同的排序模型。此外，這個開源庫不僅提供了合理的預設模型，還可以讓使用者能夠開發自己的定制模型，且提供了靈活的 API，用戶可以在其中定義和插入自己定制的損失函數、評分函數和指標。Microsoft Recommenders 是微軟雲計算和人工智慧開發團隊與微軟亞洲研究院團隊深度合作，基於多年來各類大型企業級客戶的專案經驗以及最新學術研究成果，搭建的完整推薦系統的最新實操技巧開源項目。該項目有效解決了定制和搭建企業級推薦系統中的幾個難點，包括如何將學術研究成果或開源社區提供的範例適用於企業級應用、如何集成資訊檢

索與推薦領域的學習指導資源協助從業人員深入理解並實際搭建完整推薦系統、如何選擇最優演算法以應對具體應用場景等。

為了協助資訊檢索與推薦領域的演算法模型的訓練和優化，微軟公司提供了一個大規模支持機器閱讀理解和問答系統等多種領域研究的資料集，簡稱 MS MACRO。該資料集從必應 (bing) 的搜索查詢記錄中取樣，每個問題都有人工生成的答案和完全人工重寫的答案。此外，資料集包含從通過 bing 檢索的 web 文檔中提取的百萬個密碼，這些密碼提供了管理自然語言答案所需的資訊。使用這個資料集，本文提出三個不同層次的難度不同的任務：(i) 根據一組上下文段落預測一個問題是否可以回答，然後像人類一樣提取和合成答案 (ii) 基於根據問題和段落語境資訊可以被理解的上下文段落，來生成格式良好的答案 (如果可能) ，最後 (iii) 根據給定的一個問題，對檢索得到的段落進行排序。資料集的大小和問題來自真實用戶搜索查詢的事實，該資料集的規模和真實世界的性質使它對基準測試機器閱讀理解和問答模型具有吸引力。

15 結束語

本報告依託於清華大學科技情報大資料採擷與服務平臺 AMiner 完成編制，對人工智慧各領域進行了詳細介紹。在介紹各領域概念及發展情況等內容的基礎上，報告著重介紹了各領域人才分佈情況，並對代表性論文進行了深度解讀。

在人才情況方面，總的來看，美國的人才數量遙遙領先且主要分佈在其東西海岸，獨成第一梯隊，凸顯了其在人工智慧領域的人才優勢；歐洲及亞洲也有較多的人才分佈，歐洲的人才主要集中在中西部，亞洲的人才主要集中於我國東部及日韓地區；其他諸如非洲、南美洲等地區的學者非常稀少。人才分佈與各地區的科技、經濟實力情況大體一致。對於我國而言，人才數量在大部分領域領跑第二梯隊，但與位居首位的美國相比，中國高影響力學者數量明顯不足，頂尖學者相對匱乏，中美之間還存在較大的趕超空間。我國高影響力學者主要集中在京津冀、東部、南部沿海地區及港臺地區，相較而言，中西部地區人才分佈較少，這種分佈與國內區位因素和經濟水準情況不無關係。從國際合作關係可以看出，在各領域中，中美合作論文數量均居首位，引用數及學者數也有明顯優勢，中美在人工智慧領域的積極合作，有利於我國在人工智慧前沿領域的學習發展；同時，我國也重視與歐洲、新加坡、日本等國的合作，博採眾長。

在論文解讀方面，報告對各領域的高水準學術會議及期刊論文進行挖掘，這些論文既包括近年高引論文、會議最佳論文，又有專家推薦的代表性工作。報告對各領域近年的代表性論文進行了詳盡解讀，解讀面向前沿熱點研究問題，深入探討研究方法，展現最新研究成果。同時，報告對各領域論文的關鍵字進行了挖掘分析，統計出高頻關鍵字，通過詞雲圖反映研究焦點。

當前，人工智慧已經成為引領新一輪科技革命和產業變革的戰略性技術，我國在人工智慧領域的科學技術研究和產業發展，起步稍晚於以美國為代表的發達國家，但是，在最近十餘年的人工智慧爆發發展期我國抓住了機遇，進入了快速發展階段。在這個階段，能夠推動技術突破和創造性應用的高端人才對人工智慧的發展有著至關重要的作用。本報告對人工智慧各領域技術與人才情況等內容的挖掘分析，希望能對推動我國實施人工智慧發展起到借鑒參考作用。

AMiner

參考文獻

- [1] Samuel A L. Some studies in machine learning using the game of checkers[J]. IBM Journal of research and development, 1959, 3(3): 210-229.
- [2] 機器學習簡介 (一) 機器學習簡史; <http://www.studyai.com/article/ad75a319>.
- [3] 黃佳. 基於 OPENCV 的電腦視覺技術研究[D]. 華東理工大學, 2013.
- [4] 電腦視覺入門系列 (一) 綜;
<https://blog.csdn.net/wangss9566/article/details/54618507>.
- [5] 電腦視覺簡介; https://blog.csdn.net/xiangz_csdn/article/details/78628521.
- [6] 人工智慧之知識圖譜;
https://www.aminer.cn/research_report/5c3d5a8709e961951592a49d?download=true&pathname=knowledgegraph.pdf.
- [7] 自然語言處理研究報告;
https://www.aminer.cn/research_report/5c35cdc55a237876dd7f127e?download=true&pathname=nlp.pdf.
- [8] 高維深. 基於 HMM/ANN 混合模型的非特定人語音辨識研究[D]. 電子科技大學, 2013.
- [9] 陳碩. 深度學習神經網路在語音辨識中的應用研究[D]. 華南理工大學, 2013.
- [10] 白琳. 基於語音辨識的機器人控制技術研究[D]. 西南石油大學, 2014.
- [11] 電腦圖形學研究報告;
https://www.aminer.cn/research_report/5c2edcae81ecb9818a800700?download=true&pathname=cg.pdf.
- [12] Doersch C, Gupta A, Efros A A. Unsupervised Visual Representation Learning by Context Prediction[C] // 2015 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2015.
- [13] Noroozi M, Favaro P. Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles[J]. 2016.

- [14] Pathak D, Krahenbuhl P, Donahue J, et al. Context Encoders: Feature Learning by Inpainting[J]. 2016.
- [15] Richard Zhang, Phillip Isola, Alexei A Efros. Colorful Image Colorization[C] // European Conference on Computer Vision. Springer, Cham, 2016.
- [16] Dosovitskiy A, Fischer P, Springenberg J, et al. Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015:1-1.
- [17] Wu Z, Xiong Y, Yu S, et al. Unsupervised Feature Learning via Non-Parametric Instance-level Discrimination[J]. 2018.
- [18] Wu Z, Xiong Y, Yu S, et al. Unsupervised Feature Learning via Non-Parametric Instance-level Discrimination[J]. 2018.
- [19] Zhuang C, Zhai A L, Yamins D. Local aggregation for unsupervised learning of visual embeddings[C] // Proceedings of the IEEE International Conference on Computer Vision. 2019: 6002-6012.
- [20] He K, Fan H, Wu Y, et al. Momentum Contrast for Unsupervised Visual Representation Learning[J]. arXiv preprint arXiv:1911.05722, 2019.
- [21] Bachman P, Hjelm R D, Buchwalter W. Learning representations by maximizing mutual information across views[J]. arXiv preprint arXiv:1906.00910, 2019.
- [22] Oord A, Li Y, Vinyals O. Representation learning with contrastive predictive coding[J]. arXiv preprint arXiv:1807.03748, 2018.
- [23] Tian Y, Krishnan D, Isola P. Contrastive Multiview Coding[J]. arXiv preprint arXiv:1906.05849, 2019.
- [24] Kirillov A, He K, Girshick R, et al. Panoptic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 9404-9413.
- [25] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C] // European conference on computer vision. Springer, Cham, 2014: 740-755.
- [26] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 3213-3223.

- [27] Zhou B, Zhao H, Puig X, et al. Scene parsing through ade20k dataset[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 633-641.
- [28] Neuhold G, Ollmann T, Rota Bulo S, et al. The mapillary vistas dataset for semantic understanding of street scenes[C] // Proceedings of the IEEE International Conference on Computer Vision. 2017: 4990-4999.
- [29] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C] // Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
- [30] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.
- [31] de Geus D, Meletis P, Dubbelman G. Panoptic segmentation with a joint semantic and instance segmentation network[J]. arXiv preprint arXiv:1809.02110, 2018.
- [32] Kirillov A, Girshick R, He K, et al. Panoptic feature pyramid networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 6399-6408.
- [33] Li J, Raventos A, Bhargava A, et al. Learning to fuse things and stuff[J]. arXiv preprint arXiv:1812.01192, 2018.
- [34] Li Y, Chen X, Zhu Z, et al. Attention-guided unified network for panoptic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 7026-7035.
- [35] Liu H, Peng C, Yu C, et al. An end-to-end network for panoptic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 6172-6181.
- [36] Xiong Y, Liao R, Zhao H, et al. Upsnet: A unified panoptic segmentation network[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 8818-8826.
- [37] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. arXiv preprint arXiv:1207.0580, 2012.
- [38] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C] // Advances in neural information processing systems. 2012: 1097-1105.

- [39] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, and Quoc V. Le. Mnasnet: Platform-aware neural architecture search for mobile. CoRR, abs/1807.11626, 2018.
- [40] Bowen Baker, Otkrist Gupta, Nikhil Naik, and Ramesh Raskar. Designing neural network architectures using reinforcement learning. In 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings, 2017.
- [41] Ye-Hoon Kim, Bhargava Reddy, Sojung Yun, and Chanwon Seo. Nemo : Neuro-evolution with multi objective optimization of deep neural network for speed and accuracy. In AutoML Workshop at ICML 2017, 2017.
- [42] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions. In 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Workshop Track Proceedings, 2018.
- [43] Ekin Dogus Cubuk, Barret Zoph, Dandelion Mané, Vijay Vasudevan, and Quoc V. Le. Autoaugment: Learning augmentation policies from data. CoRR, abs/1805.09501, 2018a.
- [44] Zoph B, Vasudevan V, Shlens J, et al. Learning transferable architectures for scalable image recognition[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8697-8710.
- [45] Zoph B, Le Q V. Neural architecture search with reinforcement learning[J]. arXiv preprint arXiv:1611.01578, 2016.
- [46] Zhong Z, Yan J, Wu W, et al. Practical block-wise neural network architecture generation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2423-2432.
- [47] Real E, Moore S, Selle A, et al. Large-scale evolution of image classifiers[C] // Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017: 2902-2911.
- [48] Hanxiao Liu, Karen Simonyan, Yiming Yang. DARTS: Differentiable Architecture Search. ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings, 2018.
- [49] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C] // Advances in neural information processing systems. 2014: 2672-2680.

- [50]Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv: 1411.1784, 2014.
- [51]Qi G J. Loss-sensitive generative adversarial networks on Lipschitz densities. arXiv preprint arXiv: 1701.06264, 2017.
- [52]Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. arXiv preprint arXiv: 1802.05957, 2018
- [53]Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv: 1511.06434, 2015.
- [54]Zhang H, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks[J]. arXiv preprint arXiv:1805.08318, 2018.
- [55]Karras T, Aila T, Laine S, Lehtinen J. Progressive Growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv: 1710.10196, 2017.
- [56]王昕. 關於媒體和多媒體的概念[J]. 現代電信科技, 1998(10): 44.
- [57]多媒體技術;
<https://wiki.mbalib.com/wiki/%E5%A4%9A%E5%AA%92%E4%BD%93%E6%8A%80%E6%9C%AF>.
- [58]多媒體技術的發展簡史; http://blog.sina.com.cn/s/blog_4cf2499101000a19.html.
- [59]袁保宗,阮秋琦,王延江,劉汝傑,唐曉芳.新一代(第四代)人機交互的概念框架特徵及關鍵技術[J].電子學報,2003 (S1) : 1945-1954.
- [60]徐菁.一種CAD/CAM平臺虛擬交互模組的研究與開發[D].瀋陽航空工業學院, 2009.
- [61]李兆堃. 基於 Kinect 體感技術的人機交互環境[J]. 數位技術與應用,2013 (09) : 65-66.
- [62]Minsky M.Society of mind[M]. Simon and Schuster, 1988.
- [63]任福繼, 孫曉. 智慧型機器人的現狀及發展[J]. 科技導報,2015,33 (21) : 32-38.

- [64]張俊, 吳紹輝. 資料庫技術的研究現狀及發展趨勢[J]. 工礦自動化, 2011, 37 (07) : 34-36.
- [65]資料庫的發展歷史和當前主流技術和產品;
http://www.omegaxyz.com/2018/03/26/brief_history_of_database/.
- [66]陳為, 沈則潛, 陶煜波等. 數據視覺化[M]. 北京:電子工業出版社, 2013:2-124.
- [67]Liu S, Cui W, Wu Y, et al. A survey on information visualization: recent advances and challenges[J]. The Visual Computer, 2014, 30(12):1373-1393.
- [68]Satyanarayan A, Moritz D, Wongsuphasawat K, et al. Vega-lite: A grammar of interactive graphics[J]. IEEE transactions on visualization and computer graphics, 2016, 23(1): 341-350.
- [69]Li J K, Ma K L. P5: Portable Progressive Parallel Processing Pipelines for Interactive Data Analysis and Visualization[J]. IEEE transactions on visualization and computer graphics, 2019.
- [70]Segel E, Heer J. Narrative visualization: Telling stories with data[J]. IEEE transactions on visualization and computer graphics, 2010, 16(6): 1139-1148.
- [71]Chen S, Li J, Andrienko G, et al. Supporting Story Synthesis: Bridging the Gap between Visual Analytics and Storytelling[J]. IEEE transactions on visualization and computer graphics, 2018.
- [72]Wang Y, Sun Z, Zhang H, et al. DataShot: Automatic Generation of Fact Sheets from Tabular Data[J]. IEEE transactions on visualization and computer graphics, 2019.
- [73]Cui W, Zhang X, Wang Y, et al. Text-to-Viz: Automatic Generation of Infographics from Proportion-Related Natural Language Statements[J]. IEEE transactions on visualization and computer graphics, 2019.
- [74]Strobelt H, Gehrmann S, Pfister H, et al. Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks[J]. IEEE transactions on visualization and computer graphics, 2017, 24(1): 667-676.
- [75]Han J. Data Mining: Concepts and Techniques[M]. Morgan Kaufmann Publishers Inc. 2005.

[76] 琚會婧. 資料採擷技術在客戶關係管理 (CRM) 中的應用研究[D]. 華北理工大學, 2019.

[77] Ricardo Baeza-yates, Berthier Ribeiro-Neto 著. 黃宣菁, 張奇, 邱錫鵬譯. 《現代資訊檢索》. 機械工業出版社. 2012 年 10 月第 1 版.

[78] 劉士琛. 面向推薦系統的關鍵問題研究及應用[D]. 中國科學技術大學, 2014.

[79] 李川. 即時個性化推薦系統的設計與實現[D]. 北京郵電大學, 2015.

[80] 詳細分析推薦系統和搜尋引擎的差異.

<https://blog.csdn.net/cserchen/article/details/50422553>

附錄

附錄 1 知識圖譜/知識工程知識樹

<p>知識圖譜/知識工程的知識樹共包括 10 個二級分類和 212 個三級分類。</p> <p>圖中帶“<>”的節點表示關係，沒有標“<>”的標明的節點關係是上下位關係。</p>		
一級分類	二級分類	三級分類
Knowledge Engineering	<is_kind_of>	knowledge technology 知識技術
		semantic (web) technology 語義技術
		web science 萬維科學
		information science 情報科學
	<multidiscipline_of>	cognitive science 認知科學
		semantic web 語義網
		artificial intelligence 人工智慧
		computer science 電腦科學
		natural language processing 自然語言處理
		information processing 資訊處理
		social machine 社交機器
	<using_techniques>	unified modeling language 統一模組化語言
		pattern recognition 模式識別
		information processing 資訊處理
		clustering 聚類

		clustering algorithms 聚類演算法
		data visualization 數據視覺化
		data mining 資料採擷
		quality management 品質管制
		design methodology 設計方法論
		feature extraction 特徵提取
		feature space 特徵空間
		feature selection 特徵選擇
		human centered computing 人機交互技術
		support vector machine 支持向量機
		statistical model 統計模型
		service oriented architecture 面向服務的體系結構
		markov chains 瑪律可夫鏈
		social network analysis 社會網路分析
		decision models 決策模型
		data management 資料管理
		human interaction 人機交互
		decision tree 決策樹
		genetic algorithm 遺傳演算法
		machine learning 機器學習
		information retrieval 資訊檢索
		semantic similarity 語義相似度

		semantic relatedness 語義相關性
		semantic computing 語義計算
		semantic analysis 語義分析
		graph theory 圖論
		explanation based learning 解釋學習
		data integrity 資料完整性
		text analysis 文本分析
		text mining 文本挖掘
		bootstrapping method 拔靴法
		reinforcement learning 強化學習
		human computer interaction 人機交互
		transfer learning 遷移學習
		domain experts 領域專家
		situation aware 情境感知
		graphical user interface 圖形化使用者介面
		predictive model 預測模型
	<aims_at>	better understanding 內涵理解
		computational intelligence 智慧計算
		knowledge based system 知識系統
		knowledge base 知識庫
		RDF repository 資源描述架構存儲庫
		knowledge management 知識管理

	knowledge management systems 知識管理系統
	decision support system 決策支援系統
	decision models 決策模型
	decision maker 決策者
	adaptive systems 自我調整系統
	recommender systems 推薦系統
	multiagent systems 多智慧體系統
	multi agent systems 多智慧體系統
	autonomous systems 自動系統
	autonomous agent 自動代理
	semantic search 語義檢索
	question answering system 問答系統
	human robot interaction 人機交互
	intelligent assistant 智慧輔助
	knowledge reuse 知識再利用
	knowledge sharing 知識共用
	expert system 專家系統
	intelligent systems 智慧系統
<using_information_sources>	social networks 社會網路
	web resource 網路資源
	world wide web 萬維網
	distributed databases 分散式資料庫

		big data 大數據
		information sources 信息源
		xml database 可擴展標誌語言資料庫
		heterogeneous database 異構資料庫
		heterogeneous data source 異構資料來源
		multimedia 多媒體
		wireless networks 無線網路
		relational database 關聯式資料庫
	<applications_in>	electric commerce 電子商務
		disaster management 災害管理
		computational biology 計算生物學
		biomedical domain 生物醫學領域
		health care 衛生保健
		scientific domain 科學域
		education 教育
		open government data 政府公開資料
		life science 生命科學
		gene expression data 基因表達資料
	knowledge representation	data model 資料模型
		concept modelling 概念模型
		concept model 概念模型
		conceptual model 概念模型

		semantic model 語義模型
		knowledge model 知識模型
		structured data 結構化資料
		formal specification 形式描述
		formal meaning prepresentation 形式意義表示
		formal semantics 形式語義
		commonsense knowledge 常識
		world knowledge 世界知識
		web of data 數據網
		background knowledge 背景知識
		domain knowledge 領域知識
		semantic network 語義網路
		ontology 本體論
		rough set 粗糙集
		rough set theory 粗糙集理論
		concept map 概念圖
		fuzzy sets 模糊集合
		rule based 基於規則
		rule based system 基於規則系統
		heuristic rule 啟發式規則
		object oriented 物件導向
		semantic workflow 語義 workflow

		first order logic 一階邏輯
		logic programming 邏輯程式設計
		frame based system 框為本的系統
		fuzzy logic 模糊邏輯
		fuzzy systems 模糊系統
		formal logic 形式邏輯
		decision rule 決策規則
		temporal logic 時態邏輯
		dynamic logic 動態邏輯
		domain specific language 領域專用語言
		resource description framework 資源描述架構
		ontology language 本體語言
		web ontology language 網路本體語言
		semantic web rule language 語義網規則語言
		owl 2
		collabrative ontology engineering 聯合本體工程
		ontology engineering 本體工程
		ontology development 本體開發
		collabrative ontology development 聯合本體開發
		ontology extraction 本體抽取
		ontology evolution 本體演化
		ontology versioning 本體版本

		knowledge extraction 知識提取
		knowledge capture 知識獲取
		knowledge construction 知識建構
		knowledge building 知識建構
		information extraction 資訊提取
		entity resolution 實體解析
		entity recognition 實體識別
		entity disambiguation 實體消歧
		semantic annotation 語義標注
		taxonomy induction 感應規範
	knowledge acquisition	concept clustering 概念聚類
		concept formation 概念形成
		concept learning 概念學習
		attribute value taxonomy 屬性分類規範
		event detection 事件檢測
		event identification 事件識別
		event extraction 事件抽取
		relation extraction 關係抽取
		semantic relation learning 語義關係學習
		relational learning 關係學習
		inference rule 推理規則
		rule learning 規則學習

	knowledge reasoning	case based reasoning 實例推理
		logical implication 邏輯蘊涵
		inference mechanisms 推理機制
		knowledge verification 知識驗證
		semantic interpretation 語義解釋
		uncertainty reasoning 不精確推理
		causal models 因果模型
		nonmonotonic reasoning 非單調推理
		spatial reasoning 空間推理
		temporal reasoning 時序推理
		abductive reasoning 溯因推理
	default reasoning 默認推理	
	knowledge integration	knowledge fusion 知識融合
		semantic integration 語義集成
		data fusion 資料融合
		inconsistent ontology 本體不一致
		heterogenous ontology 異構本體
		ontology interoperability 互用性本體
		ontology mapping 本體映射
		ontology alignment 本體映射
ontology matching 本體匹配		
schema mapping 模式映射		

		schema matching 模式匹配
		matching function 匹配函數
		instance matching 實例匹配
		date linking 日期連結
		date interlinking 日期互聯
		record linkage 記錄連結
		thesaurus alignment 同義對齊
	knowledge storage	triple store 三元組存儲
		RDF database 資源描述架構資料庫
		RDF storage 資源描述架構存儲
		graph database 圖資料庫
		exhaustive indexing 完整索引
		query language 查詢語言
		conjunctive queries 合取查詢
		RDF query 資源描述架構查詢
		graph query 圖查詢
		query rewrite 查詢重寫
		distributed query 分散式查詢
		subgraph matching 子圖匹配
		graph partitioning 圖劃分
		data partitioning 資料劃分

附錄 2 Data Mining 知識圖譜 (共包含二級節點 15 個 , 三級節點 93 個)

領域	二級分類	三級分類
data mining (資料採擷)	time series analysis(時間序列分析)	data streams(資料流程)
		time series data(時間序列資料)
		real time(即時)
		time series(時間序列)
		complex dynamical networks(複雜動態網路)
		dynamic system(動態系統)
		nonlinear dynamics(非線性動力學)
		system dynamics(系統動力學)
		time frequency analysis(時頻分析)
	association rule(關聯規則)	rule induction (規則歸納)
		rule learning (規則學習)
		sequential pattern(序列模式)
		frequent itemsets(頻繁項目集)
		pattern mining(模式挖掘)
		pattern matching(模式匹配)
		pattern classification(模式分類)
		frequent pattern(頻繁模式)
	algorithm(演算法)	algorithm design and analysis(演算法設計與分析)
		upper bound(上界)

		prediction algorithms(預測演算法)
		efficient algorithm(有效演算法)
		computational modeling(計算模型)
		predictive models(預測模型)
		reinforcement learning(強化學習)
		neural networks(神經網路)
		computational complexity(計算複雜性)
		probabilistic logic(概率邏輯)
		structural risk minimization (結構風險最小化)
		constrained least squares (約束最小二乘)
		incremental learning(增量學習)
		pruning technique(修剪技術)
		matrix decomposition(矩陣分解)
		generative model(生成模型)
		hidden markov models(隱瑪律可夫模型)
	big data(大數據)	dynamic databases(動態資料庫)
	big data(大數據)	heterogeneous data(異構數據)
	big data(大數據)	text data(文本資料)
	big data(大數據)	data models(資料模型)
	big data(大數據)	sensor data(感測器資料)
	big data(大數據)	data warehouses(資料倉庫)
	big data(大數據)	query processing(查詢處理)

		data structure(資料結構)
		data analysis(資料分析)
		data privacy(數據隱私)
		personal data(個人資料)
		cloud computing(雲計算)
		user behavior(用戶行為)
		parallel processing(並行處理)
		graph data(圖形資料)
		data intensive computing(資料密集型計算)
		data stream(資料流程)
		distributed databases(分散式資料庫)
		data handling(資料處理)
		data center(資料中心)
		data management(資料管理)
		data warehouse(資料倉庫)
		data security(資料安全)
		data warehousing(資料倉庫)
		privacy preservation(隱私保護)
		database management systems(資料庫管理系統)
		data generation(資料生成)
	web mining(網路挖掘)	web search (網路檢索)
		information retrieval(資訊檢索)

		link analysis (連結分析)
		image retrieval (圖像檢索)
		utility mining(效用挖掘)
		relevance feedback (相關回饋)
		recommender systems(推薦系統)
		mobile computing(移動計算)
		location based services(基於位置的服務)
		web pages(web 頁面)
		collaborative filtering(協同過濾)
		social network(社交網路)
		social interaction(社交互動)
		social media(社交媒體)
		information filtering(資訊過濾)
		social network analysis(社交網路分析)
		graph theory(圖論)
		sentiment analysis(情感分析)
		opinion mining(意見挖掘)
		semantic web(語義網)
		social web(社交網頁)
		online social network(線上社交網路)
		world wide web(萬維網)
		web 2.0(網路 2.0)

		linked data(關聯資料)
		social tagging system(社交標籤系統)
		user generated content(使用者生成內容)
		social tagging(社交標籤)
		tag recommendation(標籤推薦)
		link prediction(連結預測)
		web usage mining(web 使用挖掘)
		online community(網路社區)
		interaction network(交互網路)
		web forum(web 論壇)
	knowledge discovery(知識發現)	
	knowledge management(知識管理)	project management(專案管理)
		information technology(資訊技術)
		information system(資訊系統)
		database management(資料庫管理)
		customer relationship management(客戶關係管理)
		management system(管理系統)
	data management(資料管理)	data integration(數據整合)
		data compression(資料壓縮)
		data point(資料點)
		spatial database(空間資料庫)

		time series data(時間序列資料)
		range query(範圍查詢)
	text mining(文本挖掘)	text analysis(文本分析)
		text classification (文本分類)
		information retrieval(資訊檢索)
		natural language processing(自然語言處理)
		language model(語言模型)
		retrieval models(檢索模型)
		feature selection(特徵選擇)
		text mining technique(文本挖掘技術)
		information retrieval models(資訊檢索模型)
		text data(文本資料)
		topic model(主題模型)
		recommender system(推薦系統)
		opinion mining(意見挖掘)
		feature extraction(特徵提取)
		event detection(事件檢測)
		information filtering(資訊過濾)
		opinion analysis(輿情分析)
		sentiment analysis(情感分析)
		social media(社交媒體)
		disastrous event(災難性事件)

		text summarization(文本摘要)
		query language(查詢語言)
		query expansion(查詢擴展)
		language modeling approach(語言模型方法)
		machine translation(機器翻譯)
		biomedical text(生物醫學文本)
	image mining(圖像挖掘)	image reconstruction(圖像重建)
		image segmentation(圖像分割)
		image classification(圖像分類)
		object recognition(目標識別)
	information network(資訊網路)	information network mining(資訊網路挖掘)
		heterogeneous information network(異構資訊網路)
		graph theory(圖論)
		online social networks(線上社交網路)
		recommender system(推薦系統)
		graph mining(圖挖掘)
		location based service(基於位置的服務)
		network analysis(網路分析)
		link prediction(連結預測)
		graph data(圖數據)
		factor graph(因數圖)
		complex network(複雜網路)

		network topology(網路拓撲)
		homogeneous network(同構網路)
		information network analysis(資訊網路分析)
		graph classification(圖分類)
		graph clustering(圖聚類)
		graph structure(圖結構)
		random walk(隨機遊走)
		biological network(生物網路)
		computer networks(電腦網路)
		information integration(資訊集成)
		graph database(圖資料庫)
		large graph(大圖)
		heterogeneous network(異構網路)
		entity recognition(實體識別)
	graph mining(圖挖掘)	large graph(大圖)
	graph mining(圖挖掘)	graph classification(圖分類)
	graph mining(圖挖掘)	random graph(隨機圖)
	graph mining(圖挖掘)	directed graph(有向圖)
	graph mining(圖挖掘)	undirected graph(無向圖)
	health care(衛生保健)	electronic health records(電子健康檔案)
	health care(衛生保健)	gene expression(基因表達)
	health care(衛生保健)	biomedical research(生物醫學研究)

		adverse drugs reactions(藥物不良反應)
		genome wide association study(全基因組關聯分析)
		patient care(病人醫療護理)
		computational biology(計算生物學)
		biological sciences(生物科學)
		medical research(醫學研究)
	visualisation(視覺化)	information visualization(信息視覺化)
		data visualization(數據視覺化)
		visual analytics(視覺化分析)
		data visualisation(數據視覺化)
		data analysis(資料分析)
		network visualization(網路視覺化)
		visualization technique(視覺化技術)
		visual content(視覺內容)
		visualization tool(視覺化工具)
		interactive visualization(互動式視覺化)
		graph visualization(圖形視覺化)
		graphical user interfaces(圖形化使用者介面)
		computer animation(電腦動畫)
		visual representation(視覺表徵)
	information system(資訊系統)	

	fuzzy data mining(模糊資料採擷)	fuzzy set theory(模糊集合論)
		fuzzy set(模糊集)
		fuzzy clustering (模糊聚類)
	expert systems(專家系統)	knowledge management(知識管理)
		knowledge representation(知識表達)
		knowledge discovery(知識發現)
	similarity(相似性)	kernel operator (核算子)
		similarity relationship (相似關係)
		nearest neighbor (近鄰)
		dissimilarity (相異性)
		citation matching (引文匹配)
		similarity search(相似搜索)
		similar kernel function(相似核函數)
		earth mover's distance(EMD 距離)
		kernel function(核函數)
search problems(搜索問題)		
string matching(串匹配)		
similarity measure(相似性度量)		
keyword search(關鍵字檢索)		
semantic similarity(語義相似度)		
data structure(資料結構)	data hierarchy (資料階層)	
	complex data(複雜資料)	

unsupervised learning(無監督學習)	clustering (聚類)
	document clustering (文檔聚類)
	hierarchical clustering (層次聚類)
	image clustering (圖像聚類)
	data clustering (數據聚類)
	fuzzy clustering (模糊聚類)
	collaborative filtering (協同過濾)
	nonnegative matrix factorization (非負矩陣分解)
	cluster-based retrieval (聚類檢索)
	fuzzy clustering (模糊聚類)
	clustering algorithms(聚類演算法)
	outlier detection(孤立點檢測)
	topic modeling(主題模型)
	subspace clustering(子空間聚類)
	pattern recognition(模式識別)
	mixture of gaussians(混合高斯模型)
	gaussian processes(高斯過程)
	density estimation(密度估計)
	dimensionality reduction(降維)
	dimension reduction(降維)
maximum likelihood estimation(最大似然估計)	
matrix decomposition(矩陣分解)	

		nonnegative matrix factorization(非負矩陣分解)
		sparse representation(稀疏表示)
		sparse matrices(疏鬆陣列)
		probability distribution(概率分佈)
		probabilistic model(概率模型)
		hidden markov model(隱瑪律可夫模型)
	supervised learning(有監督學習)	classification (分類)
		feature selection (特徵選擇)
		neural networks (神經網路)
		inductive learning (歸納學習)
		markov processes(瑪律可夫過程)
		belief propagation(置信傳播)
		decision tree(決策樹)
		support vector machines(支持向量機)
		semi supervised learning(半監督學習)
		action recognition(行為識別)
		pattern recognition(模式識別)
		statistical analysis(統計分析)
		sparse coding(稀疏編碼)
		object detection(目標檢測)
		object recognition(目標識別)
	probabilistic logic(概率邏輯)	

		regression(回歸)
		manifold learning(流形學習)
		linear programming(線性規劃)
		convex programming(凸規劃)
		active learning(主動學習)
		random forest(隨機森林)
		inference mechanisms(推理機制)
		bayes methods(貝葉斯方法)
		neural network(神經網路)
		classification algorithms(分類演算法)
		bayesian methods(bayes 方法)
		random processes(隨機過程)
		deep learning(深度學習)
		feature extraction(特徵提取)
		recurrent neural network(遞迴神經網路)
		restricted boltzmann machines(受限玻爾茲曼機)
		hidden markov model(隱瑪律可夫模型)
		boltzmann machine(玻爾茲曼機)
		bayesian inference(貝葉斯推斷)
		convolutional neural networks(卷積神經網路)
		conditional random field(條件隨機場模型)
		generative model(生成模型)

		probability distribution(概率分佈)
		probabilistic model(概率模型)
		deep belief network(深度信念網路)
		logistic regression(logistic 回歸)
	network analysis(網路分析)	social network(社交網路)
		social media(社交媒體)
		graph theory(圖論)
		sensor networks(感測器網路)
		network analysis(網路分析)
		information diffusion(資訊擴散)
		community detection(社區發現)
		network structure(網路結構)
		link prediction(連結預測)
		dynamic network(動態網路)
		network formation(組網)
		social learning(社會學習)
		social science(社會科學)
		information cascades(信息追隨)
		communication networks(通訊網路)
		social influence(社會影響)
		complex network(複雜網路)
	network theory(網路理論)	

		social interaction(社交互動)
		shortest path(最短路徑)
		social behavior(社交行為)
		social life networks(社交生活網路)
	Decision analysis(決策 分析)	decision support systems (決策支援系統)
		decision making (決策)
		data envelopment analysis (資料包絡分析)
		information resource management (資訊資源管理)

AMiner



学术头条



学术君

关注“学术头条”并回复“caai”下载电子版