



论道机器翻译——离我们还有多远？

AI Time 第六期

主办：北京智源人工智能研究院BAAI
AI Time

AI Time rallies researchers specializing in the field of artificial intelligence. The goal of AI Time is to explore the essence of AI foundations, theories, algorithms, and applications, with an emphasis to encourage debate on AI-related topics.



■ 嘉宾



宗成庆
中科院自动化所研究员
国科大岗位教授



李长栓
北京外国语大学高级翻
译学院教授，副院长



张民
苏州大学特聘教授
计算机学院副院长



朱靖波
小牛翻译创始人
东北大学教授



刘洋
清华大学计算机系长
聘副教授

■ 主持人



张阔
清华大学工学博士
教授级高级工程师
小鹿医馆联合创始人

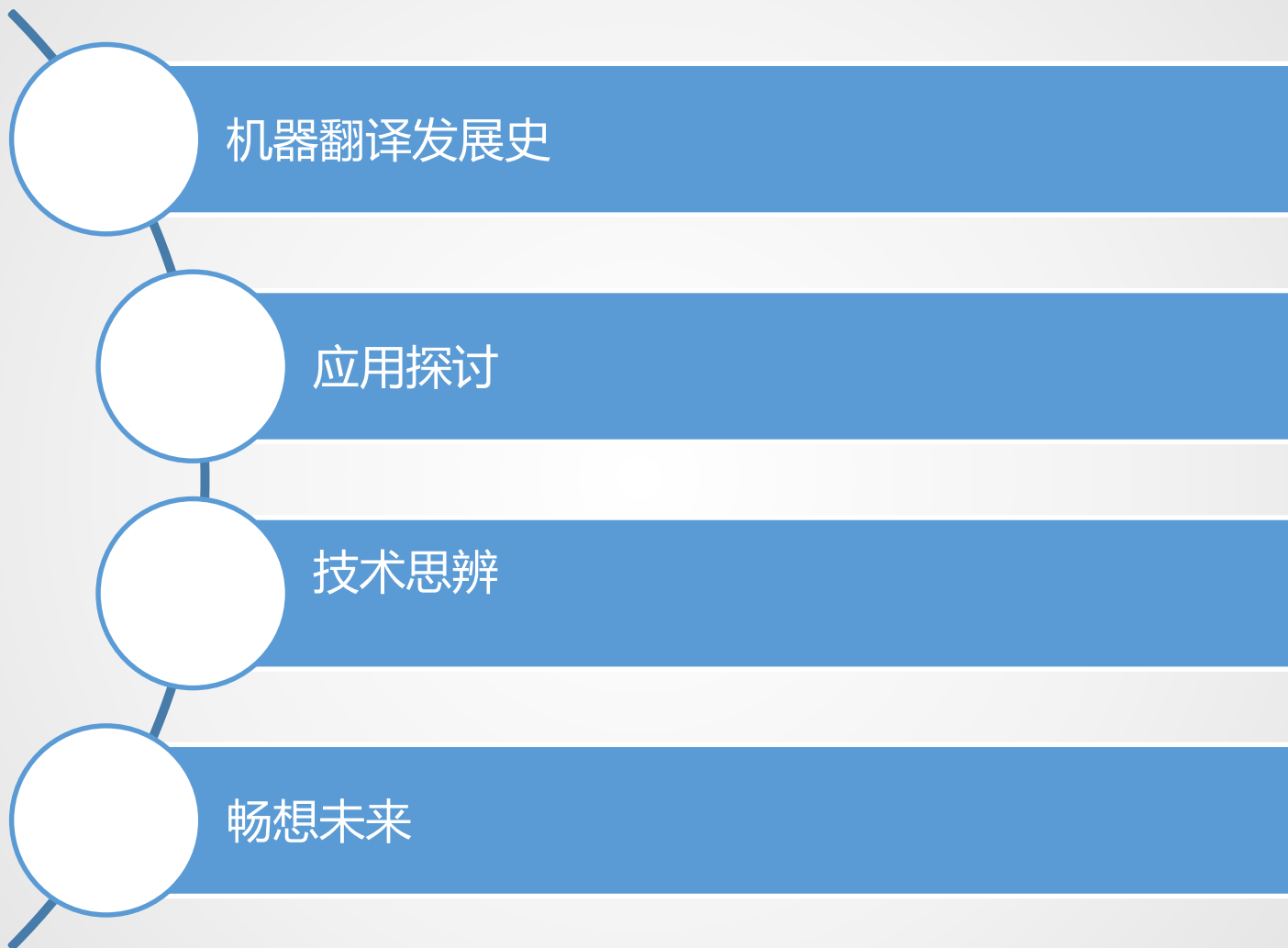


李文珏
中科创星投资总监
CCF YOCSEF学术秘书

现场观众微信群二维码



发送“AI6”进群



■ 什么是机器翻译?

I love you

我爱你 당신을 사랑합니다

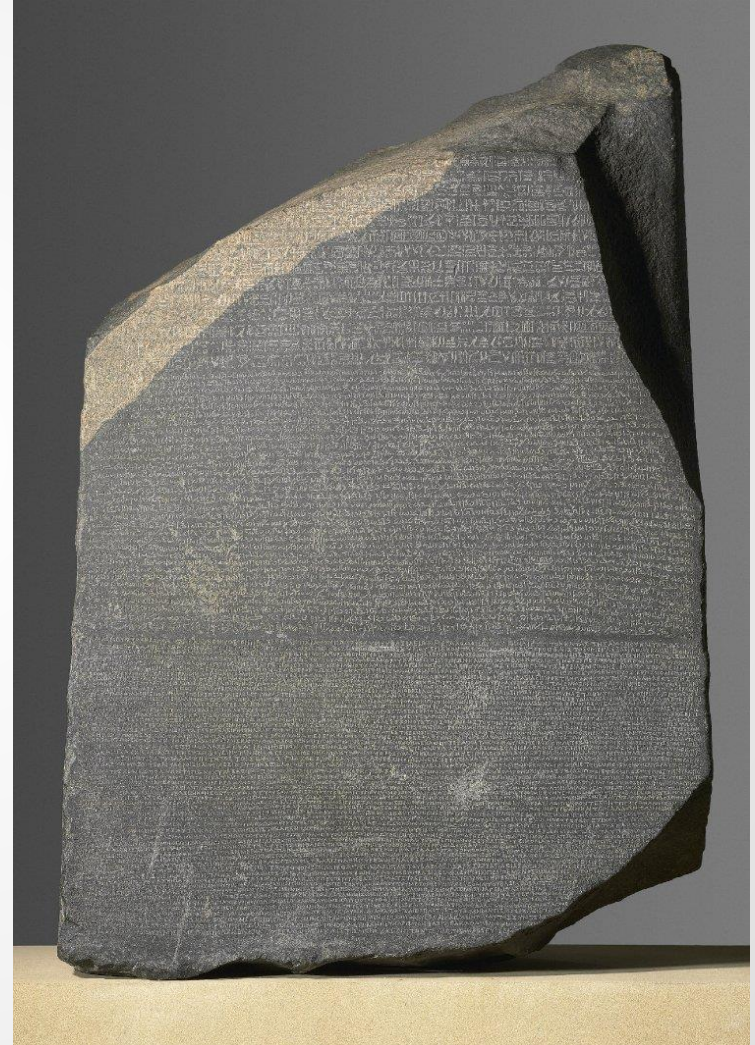
Je t'aime אני אוהב אותך

Ich liebe dich چن شما را دوست دارم

Te quiero Miluji tě

わたしは、あなたを愛しています Ik hou van je

Jag älskar dig



■ 什么是机器翻译?

- 目标: 利用计算机实现自然语言的**全自动、高质量**自动翻译

布什

与

沙龙

举行

了

会谈



Bush

held

a

talk

with

Sharon



■ 机器翻译发展史



规则驱动
机器翻译
1967-1992

统计学习驱动
机器翻译
1993-2014

表示学习驱动
机器翻译
2014-至今



■ 机器翻译发展史

- 基于语言学、字典和语法分析

基于字典的机器翻译系统

基于语法分析的机器翻译系统

基于语义的机器翻译系统



Apertium



GRAMTRANS

规则驱动
机器翻译
1967-1992

统计学习驱动
机器翻译
1993-2014

表示学习驱动
机器翻译
2014-至今



机器翻译发展史

The Mathematics of Statistical Machine Translation: Parameter Estimation

Peter F. Brown*
IBM T.J. Watson Research Center

Vincent J. Della Pietra*
IBM T.J. Watson Research Center

Stephen A. Della Pietra*
IBM T.J. Watson Research Center

Robert L. Mercer*
IBM T.J. Watson Research Center



Google Translate

Baidu 翻译

搜狗翻译

规则驱动
机器翻译
1967-1992

统计学习驱动
机器翻译
1993-2014

表示学习驱动
机器翻译
2014-至今

■ 机器翻译发展史

- 隐变量对数线性模型：在隐式语言结构上设计特征

x 布什 与 沙龙 举行 了 会谈



$$P(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta}) = \sum_{\mathbf{z}} \frac{\exp(\boldsymbol{\theta} \cdot \phi(\mathbf{x}, \mathbf{y}, \mathbf{z}))}{\sum_{\mathbf{y}'} \sum_{\mathbf{z}'} \exp(\boldsymbol{\theta} \cdot \phi(\mathbf{x}, \mathbf{y}', \mathbf{z}'))}$$



y Bush held a talk with Sharon

■ 机器翻译发展史

- 长距离调序问题

Chinese English Spanish

美国总统布什昨天在白宫与以色列总理沙龙就中东局势
举行了一个小时的会谈。

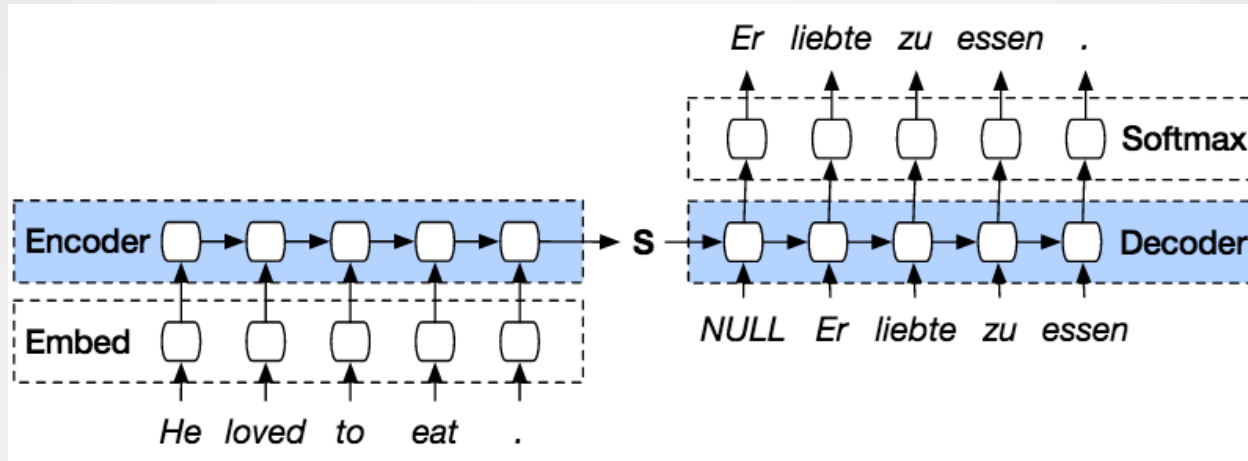
☐ Allow phonetic typing

🔊 🗨

English Chinese (Simplified) Spanish

Yesterday, U.S. President George W. Bush at the White
House with Israeli Prime Minister Ariel Sharon on the
situation in the Middle East held a one-hour talks.

■ 机器翻译发展史



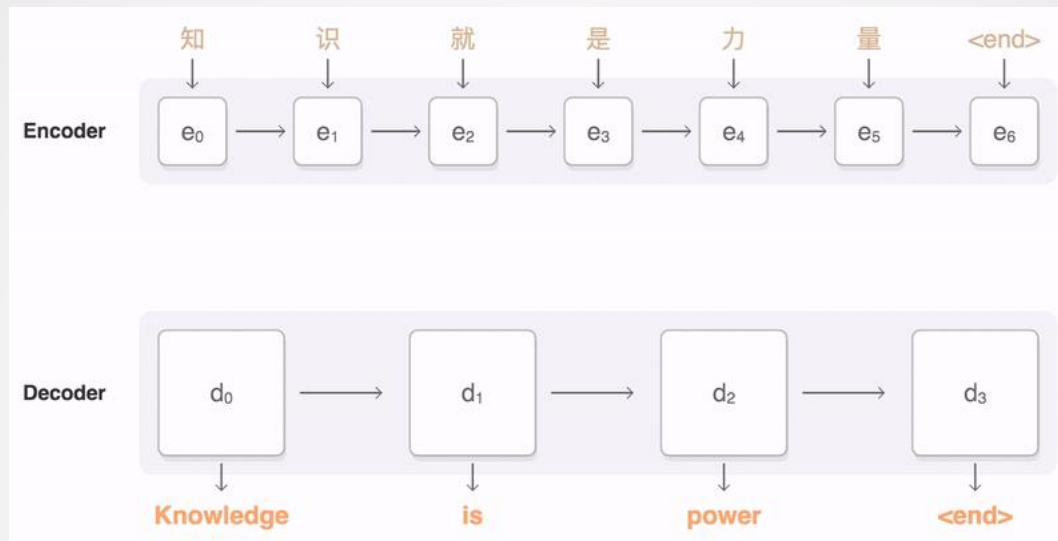
end-to-end machine translation
(seq2seq, transformer...)

规则驱动
机器翻译
1967-1992

统计学习驱动
机器翻译
1993-2014

表示学习驱动
机器翻译
2014-至今

■ 机器翻译发展史



Seq2Seq (Sutskever, I. *NIPS* 2014)

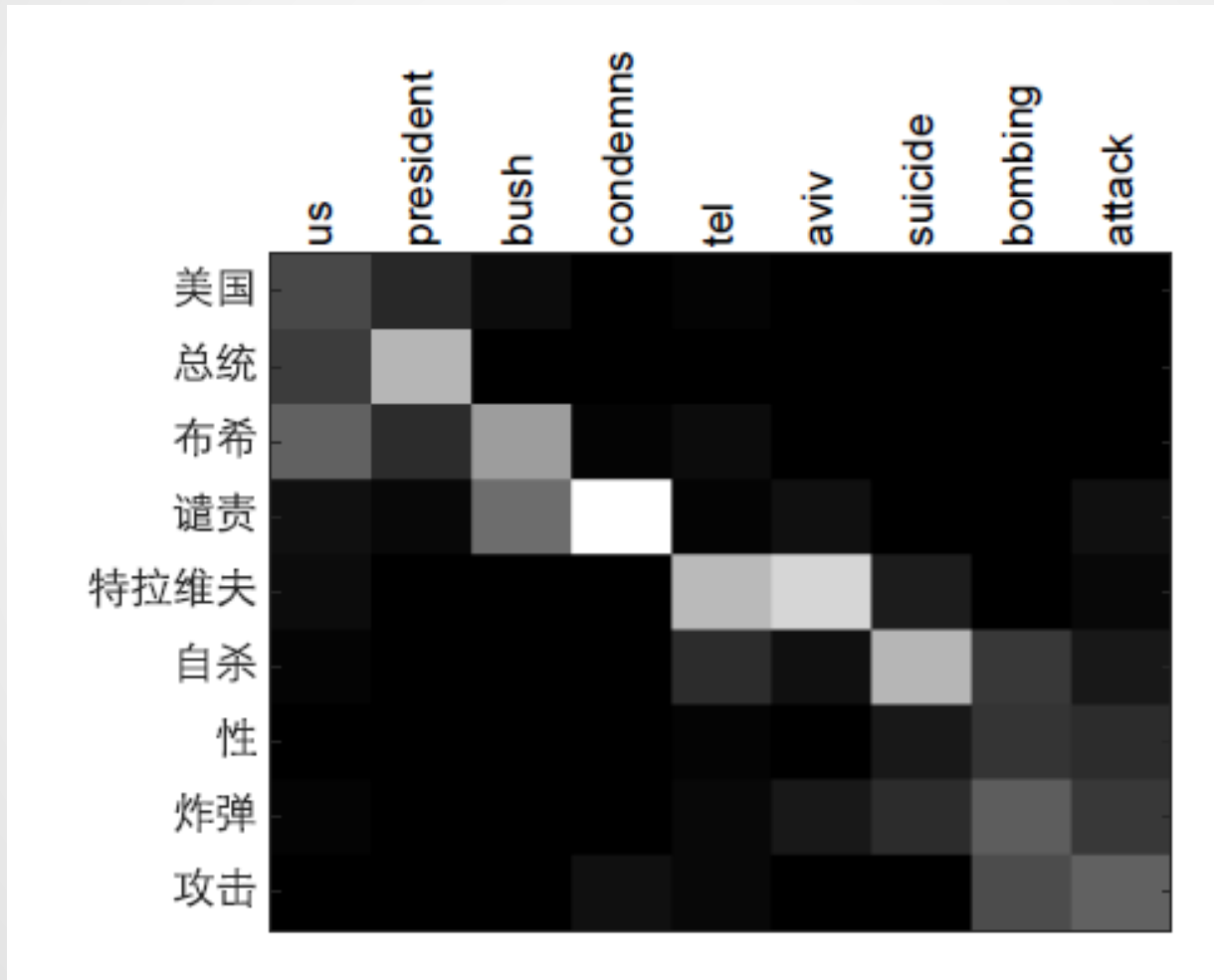
规则驱动
机器翻译
1967-1992

统计学习驱动
机器翻译
1993-2014

表示学习驱动
机器翻译
2014-至今

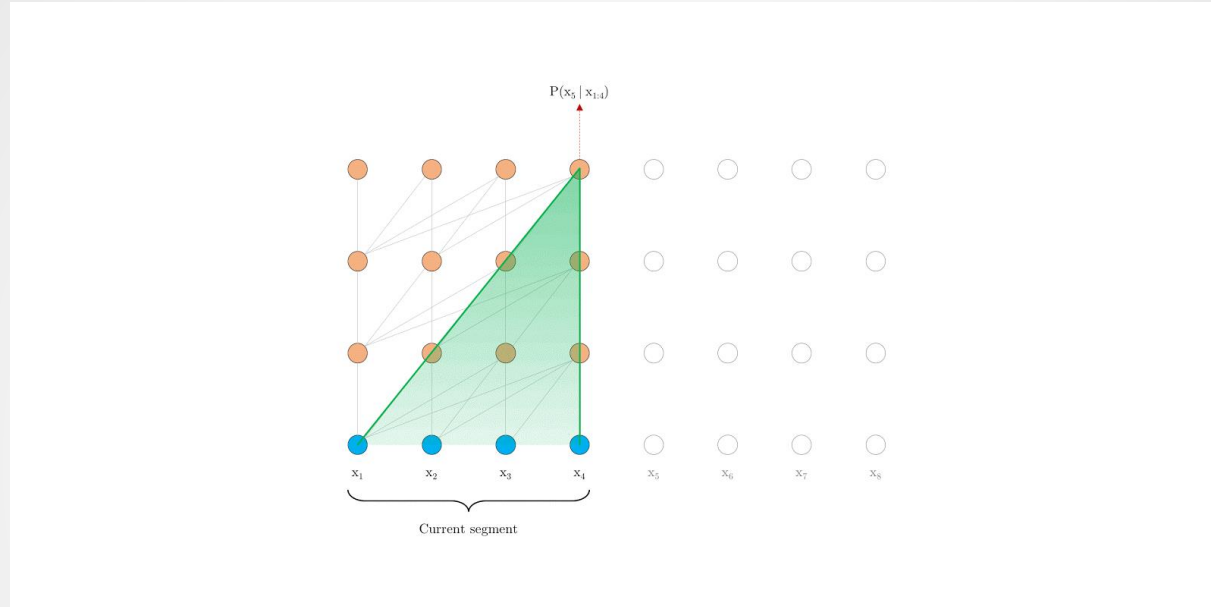
■ 机器翻译发展史

- 注意力机制：集中关注影响当前词的上下文



(Cheng et al., 2016a)

机器翻译发展史



Transformer (Vaswani, A. *NIPS* 2017)

规则驱动
机器翻译
1967-1992

统计学习驱动
机器翻译
1993-2014

表示学习驱动
机器翻译
2014-至今

应用篇

猎豹移动翻译机

小轻薄 1键翻译

猎豹微软双引擎

180天待机 双向翻译

¥299

天猫 猎豹翻译机小豹智能AI中英日韩翻译棒 出国旅行神器随身翻译机

快递: 0.00 月销427 北京

优惠 购物津贴 店铺 领津贴每满200减20 领券 >

有券 您已有3C数码购物券, 满500减5...

活动 天猫APP有「退货运费险」退货省运费! >

保障 21:00前付款, 预计9月6日(明天)送达 >

天猫直送 天猫保障的配送服务

假一赔四·极速退款·七天无理由退换

加入购物车 立即购买

7.58 搜狗官方旗舰店 99划算节 3期免息

54种语言互译

同声传译/对话翻译

能翻译的录音笔

钜惠价: 339元

¥398 品牌钜惠

价格 ¥399

99划算价 ¥359

津贴立减20元

99划算节 距开始还有4天

领如下优惠, 99划算价预计再减20

购物津贴 减20

更多优惠 3C数码券 满500减50 返积分

该商品最高可享3期分期免息

天猫【录音翻译】搜狗翻译笔翻译神器 随身 出国旅游智能翻译机 同声传译 翻译器...

快递: 0.00 月销57 北京

99划算节 此商品9月9日开卖, 请提前加购!

加入购物车 立即购买

PHILIPS 出国旅游

28种语言互译

一键翻译

拍照翻译

分期免息 视频 英语 日语 韩语 德语

¥399 品牌促销

价格 ¥699

99划算价 ¥358

99划算节 距开始还有4天

飞利浦VTR5080翻译机出国旅游中英译汉文越南泰多国语言互译翻译棒笔实时语音同声智能随身拍照即时翻译神器

快递: 0.00 月销量 214件 北京

99划算节 此商品9月9日开卖, 请提前加购!

优惠 购物津贴 店铺 领津贴每满200减20 领取 >

促销 积分 购买可得39积分 >

权益 芝麻分≥650有机会享, 0元下单满意付款 >

信用购

加入购物车 立即购买 可体验后付

应用篇

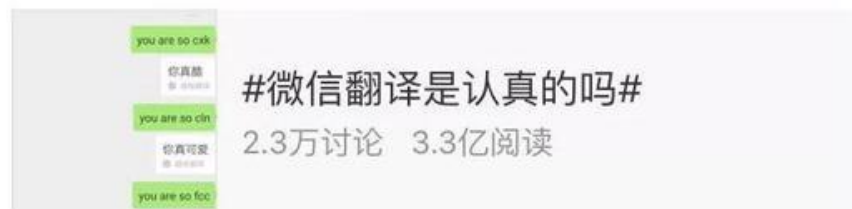


腾讯微信团队

3-3 来自iPhone客户端

+关注

#微信翻译是认真的吗#很抱歉，由于我们的翻译引擎在翻译一些没有进行过训练的非正式英文词汇时出现误翻，导致部分语句翻译出现问题，目前正在紧急修复中。



140人发表了态度 😊

■ 应用篇

- 机器翻译在哪些场景可以取代人类，哪些还有距离？
- 机器和人如何配合，能更好的进行翻译？



同声传译



专业文本
(法律/合同/技术文本)



日常/旅游

遇到的问题：

- 机器翻译以句子为基础，缺乏宏观思维
无法考虑上下文逻辑
需要补充主语时，并不考虑上一句的内容
不考虑词语翻译上下文是否统一
- 机器翻译进行语言形式转换，不关注意思传递
原文形式和意义统一时才会传递意思
不能发现和明示隐含的逻辑关系(如“条件”“因果”)。
只转换文字，不去补充任何隐含的背景信息
不会发现和编辑任何的冗余信息
不能识别和纠正原文的瑕疵（除了拼写错误）



遇到的问题

- 机器翻译依照概率而非逻辑确定词义和修饰关系
一词多义
结构歧义(并列关系还是条件、因果、方式)
- 机器翻译基于人类智慧现有成果，无查证创新精神
不常见的词、短语
成语、俗语、口号
- 机器翻译无法顾及意思的细微差别



■ 技术思辨

- ◆ 在翻译界有人认为，要成为一个好的翻译，除了翻译技能外，背景知识更重要。机器如何快速补充背景知识？
- ◆ 机器翻译研究中，是发现语言学规律更重要，还是从数据中学习表示更重要？
- ◆ 目前的机器翻译采用概率的方法，一犯错误经常是大错。如何从技术上提高机器翻译的可信性和可控性？
- ◆ 许多语言（如少数民族语言）的语言材料不足，如何在这种情况下构建机器翻译？



■ 畅想未来

■ 离实现**全自动、高质量**的机器翻译还要多久？

■ 未来的突破口可能在哪里？

■ It Is Your Turn



▶▶▶
AI Time 6 | 论道机器翻译

离我们还有多远

宗成庆 李长栓 张民 朱靖波 刘洋

时间：9月6日15:00—17:00 地点：搜狐大厦二楼1911



谢谢大家



AI Time 长期招募志愿者，欢迎志同道合的小伙伴们与AI Time 共同成长！