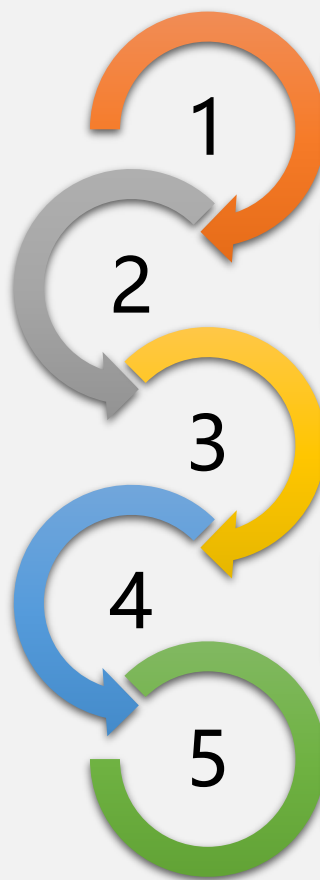




2021年6月6日

内容提纲



团队介绍

研究方向

承担的重要项目

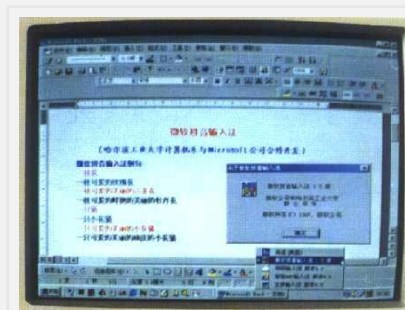
主要研究成果

代表论文及专著



团队介绍

- 教师队伍
 - 教授5人、副教授7人、讲师2人
 - 博导7人，硕导7人
- 学生队伍
 - 博士生：30+，硕士生：50+
- 成立于2003年
 - “教育部-微软”语言语音联合重点实验室依托单位
- 标志性成果
 - 语句级汉字输入系统—微软拼音
 - 汉英机器翻译系统
- 杰出校友



微软拼音-- 基于语句级的拼音输入系统



汉英机器翻译系统多次获奖



哈尔滨工业大学
HARBIN INSTITUTE OF TECHNOLOGY

团队介绍



刘铭 单丽莉 孙承杰 刘远超 林磊 刘秉权 赵铁军 李生 关毅 徐志明 杨沐昀 朱聪慧 曹海龙 徐冰



团队介绍

• 学科带头人: 李生教授

- 中国中文信息学会理事长 (2011~2016)
- 国际计算语言学会终身成就奖
 - 首位华人获奖者、亚洲第2位获奖者
- 中文信息学会终身成就奖
- 汉英机器翻译开拓者
 - 1985年国内最早汉英机器翻译研究之一
 - 1989年国内第一个通过鉴定的汉英翻译系统



团队介绍

- 学科带头人: 王晓龙教授
 - 微软拼音输入法发明者
 - 获2011年黑龙江省技术发明一等奖
 - 中国中文信息学会常务理事
 - 深圳市鹏城学者特聘教授



团队介绍

- 学科带头人: 赵铁军教授
 - “教育部-微软” 语言语音联合重点实验室主任
 - 语言技术研究中心党支部书记、主任
 - 人民网与哈尔滨工业大学共建的“社会科学
与计算协同创新中心” 常务副主任
 - 中国中文信息学会常务理事
 - CCF中文信息技术专业委员会副主任
 - 近 5 年来先后在国内高水平学术期刊和会议上发表论文 50 余篇。先后出版专著 3 部, 获得部级科技进步奖 6 项。



研究方向

行业应用

智能医疗

智慧司法

智能检务

社交网络
分析

智能金融

智能教育

结构分析技术

语言模型

词法分析

组块分析

实体识别

短语结构

依存分析

关系抽取

指代消解

语言服务技术

智能问答

检索与推荐

智能助理

智能客服

智能输入

翻译服务

语言生成技术

机器翻译

对话生成

自动摘要

图像到文本

篇章生成

数据到文本

语义理解技术

主题分析

表示学习

推理技术

知识图谱

情感分析

语义计算

机器学习技术

线性模型

随机过程

主题模型

无监督学习

深度学习

迁移学习

强化学习

对抗学习



近5年承担的重要项目

- **国家重点研发计划项目：**“基于案件集中管理的办案多维评估及检务公开技术研究”（2018YFC0830700，2018~2021，项目牵头）
- **国家重点研发计划项目：**“全媒体内容质量智能分析与推荐分发关键技术及核心系统”（2020YFB1406902，2020-2023）
- **国家重点研发计划项目：**“面向自然口语交互的情境化语义理解和多轮对话交互管理技术”（2017YFB1002102，2017~2021，参与课题研究）
- **国家自然科学基金重点项目：**“面向汉语文本理解的语义计算方法”，2016~2017
- **国家高技术发展计划项目：**“面向基础教育知识关联与推理类问题求解关键技术及系统”，2015~2017
- **多项国家自然科学基金项目、国际合作项目、企业合作项目等**

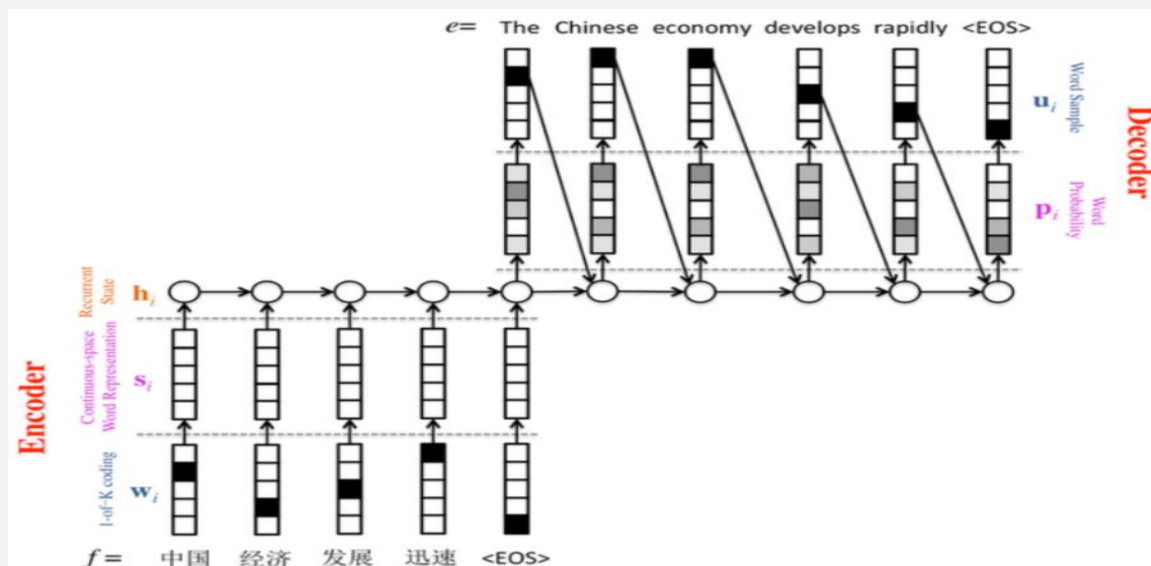


主要研究成果

机器翻译技术

• 技术基础及水平

- “奥运数据”多语翻译性能第一
 - 2012国际口语机器翻译评测 (IWSLT)
- 机器翻译自动评价性能第一
 - 2010 欧洲(WMT)和美国NIST评测(Metrics MATR)
- 2016年构建了全新的神经网络机器翻译模型
 - 能够进行中-俄, 中-英, 中-日, 中-韩, 中-越等多语言翻译



主要研究成果

机器翻译技术

• 市场前景

- 随着经济全球化及“一带一路”战略的推进，机器翻译技术在促进政治、经济、文化交流等方面将起到越来越重要的作用。



Company Logo

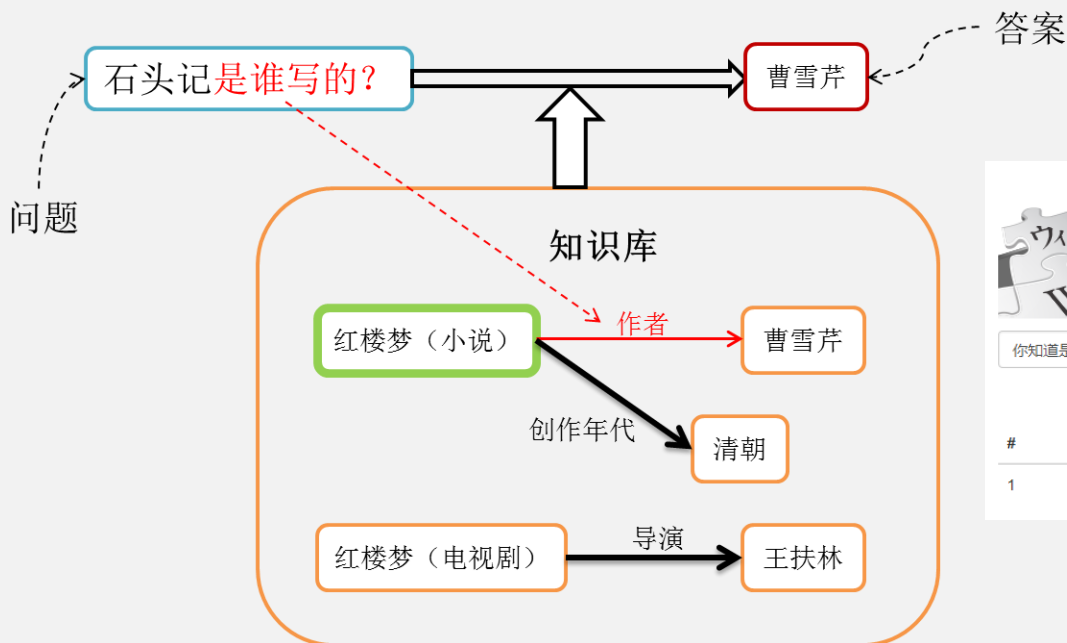
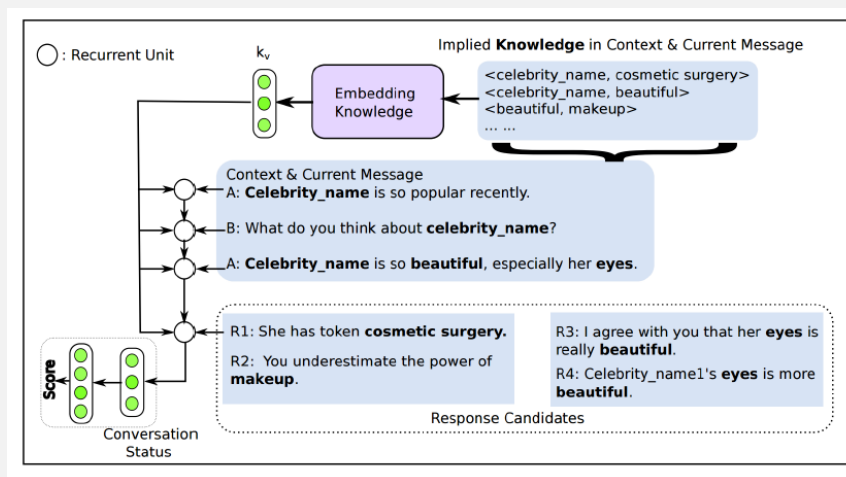


哈尔滨工业大学
HARBIN INSTITUTE OF TECHNOLOGY

主要研究成果

智能问答技术

- 基于语义理解的智能问答技术
- 基于检索的智能对话技术
- 基于知识库的智能问答技术
- 问答对资源挖掘技术



INSUN KBQA
Knowledge is Power
— Francis Bacon

6,502,738 ENTITIES
587,875 PROPERTIES
43,063,796 FACTS

你知不知道是谁写了石头记吗?

Ask!

#	Entity Name	Property	Value
1	红楼梦	原著	曹雪芹



主要研究成果

智能问答技术

- 市场前景
 - 智能客服

传统呼叫中心面临困境!

2015年呼叫中心座席
增长21.3%

2015年劳动力人口
减少487万人

呼叫中心年平均离职率
40%~100%

工作单调重复
环境拥挤嘈杂
薪酬低晋升难



2020年, 智能机器人座席能满足**40%**的客服市场需求!

中国将达到**1000万**的智能机器人座席。



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY

主要研究成果

智能问答技术

• 市场前景



全媒体呼叫中心
Internet of Things

+



AI人工智能
AI of Things

=

新一代呼叫中心：
智能交互中心

人工客服+机器人客服成为新一代呼叫中心的核心，成本更低、服务更智能！

	传统呼叫中心	新一代呼叫中心：智能交互中心
• 服务模式	人工坐席	智能机器人坐席+人工坐席
• 商业模式	成本中心	利润中心
• 组织模式	中心化	去中心化
• 渠道模式	电话+PC+手机	电话+PC+手机+APP+微信+机器人+IoT
• 沟通模式	基于服务支持的联络中心	服务即营销的客户终生交互中心
• 主要功能	呼叫中心+在线客服（IM）+工单	智能机器人+呼叫中心+在线客服+工单
• 技术架构	PSTN+SIP的硬件与软件系统集成架构	全数字化、软交换为主的云计算技术架构体系（SaaS化）
• 质检模式	抽样质检	100%智能质检+行为分析
• 成本模式	高投入建设、高运营成本	成本为原来的（10%~1%）



主要研究成果

智能输入技术

• 技术基础及水平

- 领先的语句级输入技术
- PC版、智能手机版、智能终端版
 - 点滑输入、支持九键输入、支持手写
 - 支持Android、iPhone、iPad
 - 创新浮动输入

- 韩国三星公司采用了WI输入法的技术。WI输入法荣获2010中国互联网最佳技术创新产品提名奖（全国仅有五款产品获此殊荣）。

哈尔滨是个美丽的城市

0 shi

1 是 2 诗歌 3 师哥 4 时 5 师 6 市 7 实 8 使



Wi 输入法
for iOS



Wi 输入法
for Android



哈尔滨工业大学
HARBIN INSTITUTE OF TECHNOLOGY

主要研究成果

对话系统

- 技术基础及水平
 - 检索与生成相结合的对话系统
 - 融合知识的对话生成技术
 - 多轮对话的历史建模
 - 基于深度学习的意图理解与对话管理

2019-03-09 16:44:31 You

帮我查一下北欧两项的比赛时间安排

Chatbot 2019-03-09 16:44:41

为您查询到多条北欧两项比赛安排, 可选时间为: 2018/2/22 16:30, 2018/2/20 19:00, 2018/2/14 15:00

2019-03-09 16:45:06 You

我还想知道花样滑冰的比赛时间

Chatbot 2019-03-09 16:45:07

为您查询到多条花样滑冰比赛安排, 可选时间为: 2018/2/20 10:00, 2018/2/12 10:00, 2018/2/16 10:00, 2018/2/15 10:00, 2018/2/17 10:00, 2018/2/21 10:00, 2018/2/19 10:00, 2018/2/23 10:00, 2018/2/11 10:00, 2018/2/9 10:00, 2018/2/14 10:00, 2018/2/25 9:30

冬奥会赛事时间查询

2019-03-09 19:43:27 You

我想买2月17号花样滑冰的比赛门票

Chatbot 2019-03-09 19:43:37

我们有A, B, C三类票可供选择, 您要选择哪一种?

2019-03-09 19:43:51 You

A类票吧

Chatbot 2019-03-09 19:43:53

请问您需要看什么时候的花样滑冰比赛?

2019-03-09 19:44:02 You

10:00

Chatbot 2019-03-09 19:44:03

您几个人同行?

2019-03-09 19:44:13 You

我和妈妈2个人

Chatbot 2019-03-09 19:44:14

为您预定: 2018/2/9 13:25, 花样滑冰-团体 男子单人滑 短节目 团体 男子双人滑 短节目, 价格4600的门票2张, 共计9200.0元!

冬奥会赛事门票预订

2019-03-09 16:56:54 You

请问跳台滑雪的门票价格是多少?

Chatbot 2019-03-09 16:57:04

为您查询到多条跳台滑雪比赛安排, 可选票价为: 1000.0, 900.0, 1700.0, 500.0

2019-03-09 16:57:20 You

还有速度滑冰的门票价格是多少呀

Chatbot 2019-03-09 16:57:22

为您查询到多条速度滑冰比赛安排, 可选票价为: 2100.0, 1300.0

冬奥会赛事门票价格查询

主要研究成果

智能医疗信息处理系统

• 技术基础及水平

- 中文电子病历的医疗语言多级加工系统
- 中文电子病历的医疗实体与实体关系抽取系统
- 中文电子病历语言处理规范及语料库
- 中文电子病历实体与实体关系标注规范及语料库
- 心血管疾病风险因子标注规范及语料库
- 系统涉及疾病、症状、检查和治疗四大类医疗实体，还包括这些实体之间的15种医疗关系。
 - 各个模块的模型评价 (F值) 分别是：分词，98.2%；词性标注，96.7%；组块分析，95.1%；命名实体识别，92.3%；实体关系抽取，75.2%。
- <https://github.com/WILAB-HIT/Resources>


基于中文电子病历文本的在线演示

请输入一个中文电子病历文本中的句子

1. 患者既往有高血压病史，最高达180/100mmHg，口服“北京降压灵、利血平”治疗，有脑梗死病史，遗留语言及左侧肢体无力。
2. 门诊行头CT检查，显示左侧侧脑室旁低密度病灶，以“脑梗死”收入我科。

执行

分词

1. 患者既往有高血压病史，最高达180/100mmHg，口服“北京降压灵、利血平”治疗，有脑梗死病史，遗留语言及左侧肢体无力。
2. 门诊行头CT检查，显示左侧侧脑室旁低密度病灶，以“脑梗死”收入我科。

词性

1#CD. #PU 患者#NN 既往#NT 有#VE 高血压#NN 病史#NN, #PU 最高#AD 高#AD 达#VV 180#CD /#PU 100#CD mmHg, #PU 口服#VV “#PU 北京#NR 降压灵#NR、#PU 利血平#NN”#PU 治疗#VV, #PU 有#VE 脑梗死#NN 病史#NN, #PU 遗留#VV 语言#NN 及#CC 左侧#JJ 肢体#NN 无力#VA. #PU
2#CD. #PU 门诊#NN 行#VV 头#NN CT#NN 检查#NN, #PU 显示#VV 左侧#JJ 侧脑室#NN 旁#LC 低#JJ 密度#NN 病灶#NN, #PU 以#P “#PU 脑梗死#NN”#PU 收入#VV 我科#NN. #PU

块级句法分析

{ LST (CD 1) } (PU) { NP (NN 患者) } (NP (NT 既往)) { VE } (NP (NN 高血压) (NN 病史)) (PU ,) { ADVP (AD 最) } { ADVP (AD 高) } { VV 达 } (CD 180) (PU /) (CD 100) (CLP (M mmHg)) (PU ,) { VV 口服 } (PU “) (NP (NR 北京) (NR 降压灵) (PU .) (NN 利血平) (PU “) { VV 治疗 } (PU ,) { VE 有 } { NP (NN 脑梗死) (NN 病史) } (PU ,) { VV 遗留 } (NP (NN 语言)) { VP (VA 及) } (CC 及) { ADJP (JJ 左侧) } (NP (NN 肢体)) { VP (VA 无力) } (PU .)
{ LST (CD 2) } (PU) { NP (NN 门诊) } { VV 行 } (NP (NN 头) (NN CT)) { NP (NN 检查) } (PU ,) { VV 显示 } (ADJP (JJ 左侧)) (NP (NN 侧脑室)) (LC 旁) { ADJP (JJ 低) } (NP (NN 密度)) (NP (NN 病灶)) (PU ,) { P 以 } (NP (PU “) (NN 脑梗死) (PU “) { VV 收入 } (NP (NN 我科)) (PU .)

医疗实体：疾病 疾病诊断分类 临床症状 异常检查结果 检查 治疗

1. 患者既往有高血压病史，最高达180/100mmHg，口服“北京降压灵、利血平”治疗，有脑梗死病史，遗留语言及左侧肢体无力。
2. 门诊行头CT检查，显示左侧侧脑室旁低密度病灶，以“脑梗死”收入我科。

实体关系

<北京降压灵, R=TIAD, 高血压病史>
<利血平, R=TIAD, 高血压病史>
<脑梗死病史, R=DCS, 左侧肢体无力>
<头CT检查, R=TeRS, 左侧侧脑室旁低密度病灶>
<头CT检查, R=TeRD, 脑梗死>
<左侧侧脑室旁低密度病灶, R=SID, 脑梗死>

注：实体关系类型参见 备注规范

主要研究成果

微博的传播分析

• 技术基础及水平

- 大规模社会舆情信息采集提取与分类技术
- 社会舆情热点发现技术
- 社会舆情热点话题多网络关联和结构分析技术
- 舆情热点真实性评估技术

• 市场前景

- 微博舆情分析
- 目标群体行为分析



意见领袖排名

排名	用户名	影响力	粉丝数	微博数	中心性	中介性
1	南方都市报	51.0	4052842	21544	0.07	0.52
2	月光博客	39.0	274038	6630	0.06	0.34
3	俞敏洪	36.0	8959607	555	0.07	0.38
4	王利芬	31.0	1690317	13635	0.15	0.33
5	洪晃look	31.0	6530317	12288	0.07	0.34
6	微天下	31.0	1984413	65896	0.00	0.32
7	创业邦杂志	26.0	279946	19734	0.21	0.22
8	OnlySwan	24.0	22021	16570	0.13	0.11
9	tanhu	16.0	5400	1814	0.04	0.12
10	官鑫Sam	13.0	57543	2405	0.00	0.06

个人基本信息


所在地: 广东 广州
关注: 385
粉丝: 4052842
微博: 21544
认证: 已认证
简介: 欢迎在南方都市报网上订阅中心
(http://db.oeeee.com) 订阅《南方都市报》! 咨询QQ:
800088030、190290531
标签: 南部 新闻传媒

结构划分

- 1 80后(30人)
- 2 电影(21人)
- 3 自由(18人)
- 4 音乐(18人)
- 5 IT(15人)
- 6 旅行(15人)
- 7 美食(13人)
- 8 互联网(13人)
- 9 旅游(13人)
- 10 听歌(11人)

标签划分


Grid of user avatars and names, including: 赵有宜, 上海 闵行区; A_Kris, 上海 浦东新区; 胡国亮, 广东 广州; 张子凯, 天津 河东区; 袁老远, 北京 朝阳区; Jasonhua198x, 重庆 沙坪坝区; doublesixwings, 四川 成都; 张鑫-19860624, 北京 大兴区; tanhu, 北京 海淀区; ShenghuiWei, 北京 海淀区; 一袋大米么么哒, 四川 成都; Lenovo刁然, 北京 海淀区; 胡国亮, 浙江 绍兴; 爱宠米米的加速联盟, 广东 珠海; Mr. C. 浮生未歇, 湖南 岳阳; 李银虹, 山东 日照; 逍遥, 河北 廊坊; 潘子 行云, 上海 杨浦区; 黄江江, 江西 萍乡; 魔术师的小鱼, 北京 东城区; Z09ap, 广东 广州; SHHH, 广东 广州; Jeslin_zheng, 上海 闸北区; yun1129, 台湾 桃园县



哈尔滨工业大学
HARBIN INSTITUTE OF TECHNOLOGY

主要研究成果

近五年获授权专利

- 一种基于机器翻译的中文语义知识库的自动构建方法, 201610111365.0
- 一种面向医疗领域基于深度学习的命名实体识别方法, 201610590151.6
- 基于多语平行语料的语义向量的机器翻译方法, 201610590241.5
- 一种问答系统中检索子模块的自动评价方法, 2017104699019
- 一种上下文敏感的神经网络机器翻译中未登录词翻译装置和方法, 2017105149355

研究中心杨沐昀副教授被中国知识产权报 (2019.12.7) 评为 “中国人工智能核心发明人” 之一。



代表论文及专著 ——2020

1. Kehai Chen, Rui Wang, Masao Utiyama, Eiichiro Sumita, Tiejun Zhao, Muyun Yang, Hai Zhao. Towards More Diverse Input Representation for Neural Machine Translation. IEEE ACM Trans. Audio Speech Lang. Process. 28: 1586-1597 (2020)
2. Mingming Yang, Rui Wang, Kehai Chen, Xing Wang, Tiejun Zhao, Min Zhang. A Novel Sentence-Level Agreement Architecture for Neural Machine Translation. IEEE ACM Trans. Audio Speech Lang. Process. 28: 2585-2597 (2020)
3. Muyun Yang, Shujie Liu, Kehai Chen, Hongyang Zhang, Enbo Zhao, Tiejun Zhao. A Hierarchical Clustering Approach to Fuzzy Semantic Representation of Rare Words in Neural Machine Translation. IEEE Trans. Fuzzy Syst. 28(5): 992-1002 (2020)
4. Haipeng Sun, Rui Wang, Kehai Chen, Masao Utiyama, Eiichiro Sumita, Tiejun Zhao. Knowledge Distillation for Multilingual Unsupervised Neural Machine Translation. ACL 2020: 3525-3535
5. Guanhua Zhang, Bing Bai, Junqi Zhang, Kun Bai, Conghui Zhu, Tiejun Zhao. Demographics Should Not Be the Reason of Toxicity: Mitigating Discrimination in Text Classifications with Instance Weighting. ACL 2020: 4134-4145
6. Changzhen Ji, Xin Zhou, Yating Zhang, Xiaozhong Liu, Changlong Sun, Conghui Zhu, Tiejun Zhao. Cross Copy Network for Dialogue Generation. EMNLP (1) 2020: 1900-1910
7. Chang Gao, Chengjie Sun, Lili Shan, Lei Lin, Mingjiang Wang. Rotate3D: Representing Relations as Rotations in Three-Dimensional Space for Knowledge Graph Embedding. CIKM 2020: 385-394



代表论文及专著 ——2019

1. Junwei Bao, Duyu Tang, Nan Duan, Zhao Yan, Ming Zhou, Tiejun Zhao: Text Generation From Tables. IEEE ACM Trans. Audio Speech Lang. Process. 27(2): 311-320 (2019)
2. Xuefeng Bai, Hailong Cao, Kehai Chen, Tiejun Zhao: A Bilingual Adversarial Autoencoder for Unsupervised Bilingual Lexicon Induction. IEEE ACM Trans. Audio Speech Lang. Process. 27(10): 1639-1648 (2019)
3. Kehai Chen, Rui Wang, Masao Utiyama, Eiichiro Sumita, Tiejun Zhao: Neural Machine Translation With Sentence-Level Topic Context. IEEE ACM Trans. Audio Speech Lang. Process. 27(12): 1970-1984 (2019)
4. Zhen Xu, Chengjie Sun, Yinong Long, Bingquan Liu, Baoxun Wang, Mingjiang Wang, Min Zhang, Xiaolong Wang: Dynamic Working Memory for Context-Aware Response Generation. IEEE ACM Trans. Audio Speech Lang. Process. 27(9): 1419-1431 (2019)
5. Haipeng Sun, Rui Wang, Kehai Chen, Masao Utiyama, Eiichiro Sumita, Tiejun Zhao: Unsupervised Bilingual Word Embedding Agreement for Unsupervised Neural Machine Translation. ACL (1) 2019: 1235-1245 (顶级会议)
6. Mingming Yang, Rui Wang, Kehai Chen, Masao Utiyama, Eiichiro Sumita, Min Zhang, Tiejun Zhao: Sentence-Level Agreement for Neural Machine Translation. ACL (1) 2019: 3076-3082 (顶级会议)
7. Guanhua Zhang, Bing Bai, Jian Liang, Kun Bai, Shiyu Chang, Mo Yu, Conghui Zhu, Tiejun Zhao: Selection Bias Explorations and Debias Methods for Natural Language Sentence Matching Datasets. ACL (1) 2019: 4418-4429 (顶级会议)
8. Guanlin Li, Lemao Liu, Guoping Huang, Conghui Zhu, Tiejun Zhao: Understanding Data Augmentation in Neural Machine Translation: Two Perspectives towards Generalization. EMNLP/IJCNLP (1) 2019: 5688-5694 (顶级会议)
9. Guanlin Li, Lemao Liu, Xintong Li, Conghui Zhu, Tiejun Zhao, Shuming Shi: Understanding and Improving Hidden Representations for Neural Machine Translation. NAACL-HLT (1) 2019: 466-477
10. Chunpeng Ma, Akihiro Tamura, Masao Utiyama, Eiichiro Sumita, Tiejun Zhao: Improving Neural Machine Translation with Neural Syntactic Distance. NAACL-HLT (1) 2019: 2032-2037

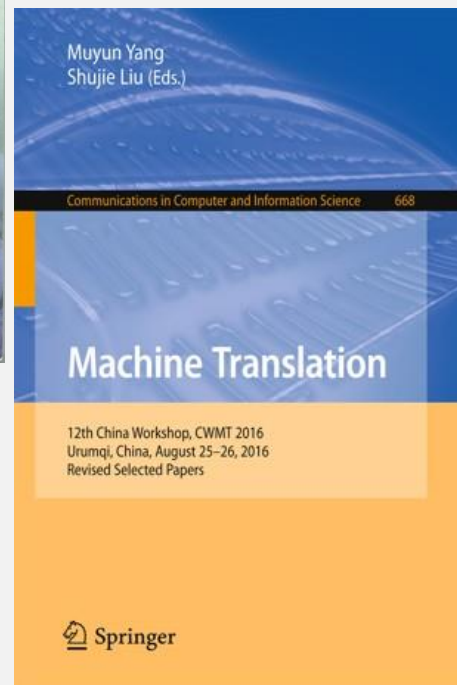
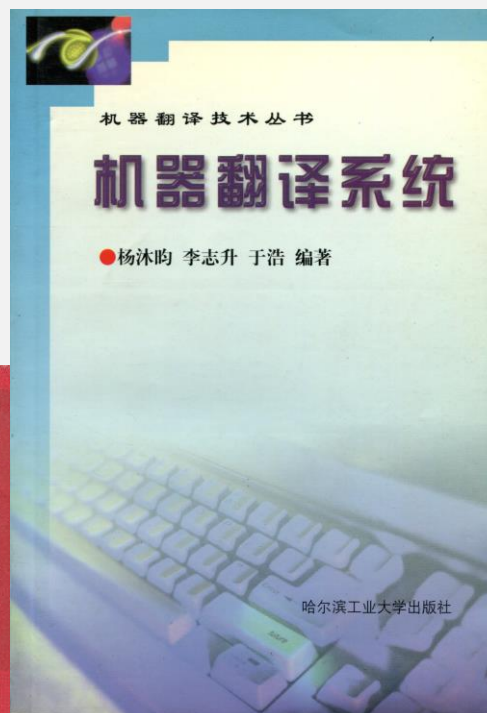
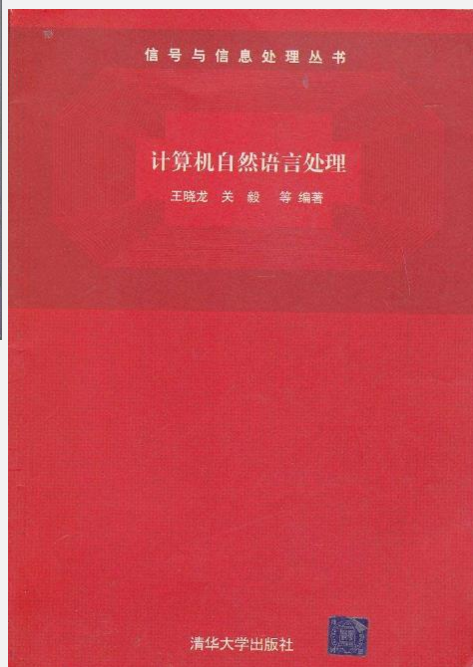
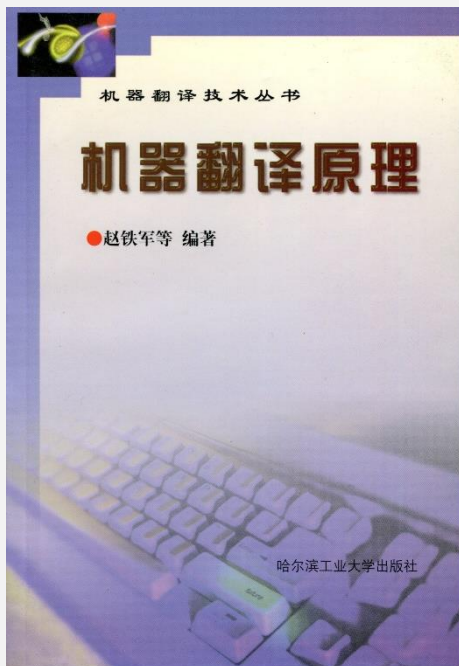


代表论文及专著——2017、2018

1. Point Set Registration For Unsupervised Bilingual Lexicon Induction, Hailong Cao and Tiejun Zhao, IJCAI2018 (顶级会议)
2. attention fused deep matching network for natural language inference, Chaoqun Duan, Lei Cui, Xinchu Chen, Furu Wei, Conghui Zhu, Tiejun Zhao, IJCAI2018 (顶级会议)
3. Forest-Based Neural Machine Translation, Chunpeng Ma, Akihiro Tamura, Masao Utiyama, Tiejun Zhao, Eiichiro Sumita, ACL-2018 (顶级会议)
4. A Neural Approach to Source Dependence Based Context Model for Statistical Machine Translation, Kehai Chen, Tiejun Zhao, Muyun Yang, Lemao Liu, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2018
5. Translation Prediction with Source Dependency-Based Context Representation, Kehai Chen, Tiejun Zhao, Muyun Yang, Lemao Liu, AAAI-2017 (顶级会议)
6. Deterministic Attention for Sequence-to-Sequence Constituent Parsing, Chunpeng Ma, Akihiro Tamura, Lemao Liu, Tiejun Zhao, Eiichiro Sumita, AAAI-2017 (顶级会议)



代表论文及专著



联系方式

- 机器智能与翻译 (MIT) 研究室
 - 徐冰 hitxb@hit.edu.cn
- 智能技术与自然语言处理 (ITNLP) 研究室
 - 刘秉权 liubq@hit.edu.cn
- 网络智能 (WI) 研究室
 - 关毅 guanyi@hit.edu.cn
- 社会计算 (SC) 研究室
 - 徐志明 xuzm@hit.edu.cn



谢谢各位同学



哈爾濱工業大學
HARBIN INSTITUTE OF TECHNOLOGY