# The 19th Chinese National Conference on Computational Linguistic

## Proceedings of the 19th Chinese National Conference on Computational Linguistics

October 30 - November 1, 2020 Haikou, China ©The 19th Chinese National Conference on Computational Linguistic

Order copies of this and other CCL proceedings from:

Chinese National Conference on Computational Linguistic (CCL) Courtyard 4, South Fourth Street, Zhongguancun , Haidian District, Beijing 100190 China

Tel: +010-62562916 Fax: +010-62661046 cips@iscas.ac.cn

#### Introduction

Welcome to the proceedings of the 19th China National Conference on Computational Linguistics (19th CCL). The conference and symposium were hosted online and co-organized by Hainan University, China. CCL is an annual conference (bi-annual before 2013) that started in 1991. It is the flagship conference of the Chinese Information Processing Society of China (CIPS), which is the largest NLP scholar and expert community in China. CCL is a premier nation-wide forum for disseminating new scholarly and technological work in computational linguistics, with a major emphasis on computer processing of the languages in China such as Mandarin, Tibetan, Mongolian, and Uyghur.

The Program Committee selected 109 papers (74 Chinese papers and 35 English papers) out of 303 submissions for publication. The acceptance rate is 35.97 – Machine Translation and Multilingual Information Processing (7)

- Fundamental Theory and Methods of Computational Linguistics (10)
- Minority Language Information Processing (9)
- Social Computing and Sentiment Analysis (12)
- Text Generation and Summarization (5)
- Information Retrieval, Dialogue and Question Answering (11)
- Linguistics and Cognitive Science (15)
- Language Resource and Evaluation (15)
- Knowledge Graph and Information Extraction (11)
- NLP Applications (14)

The final program for the 19th CCL was the result of intense work by many dedicated colleagues. We want to thank, first of all, the authors who submitted their papers, contributing to the creation of the high-quality program. We are deeply indebted to all the Program Committee members for providing high-quality and insightful reviews under a tight schedule, and extremely grateful to the sponsors of the conference. Finally, we extend a special word of thanks to all the colleagues of the Organizing Committee and secretariat for their hard work in organizing the conference, and to Springer for their assistance in publishing the proceedings in due time. We thank the Program and Organizing Committees for helping to make the conference successful, and we hope all the participants enjoyed the first online CCL conference.

September 2020

Maosong Sun Sujian Li Yue Zhang Yang Liu

#### **Organizers**

#### 19th CCL Program Chair:

Sujian Li, Peking University (China) Yue Zhang, Westlake University (China)

Yang Liu, Tsinghua University (China)

#### 19h CCL Area Co-Chairs:

Gaoqi Rao, Beijing Language and Culture University (China)

Timothy O'Donnell, McGill University (Canada)

Wanxiang Che, Harbin Institute of Technology (China)

Fei Xia, University of Washington (USA)

Xin Zhao, Renmin University of China (China)

Hongzhi Yin, University of Queensland (Australia)

Xiaojun Wan, Peking University (China)

Jinge Yao, Microsoft Research Asia

Kang Liu, Institute of Automation, CAS (China)

Ruihong Huang, Texas A&M University (USA)

Yang Feng, Institute of Computing Technology, CAS (China)

Haitao Mi, Ant Financial

Hongxu Hou, Mongolian university (China)

Fanglin Wang, Leading Intelligence Corporation

Weiguang Qu, Nanjing Normal University (China)

Nianwen Xue, University of Bratis (USA)

Meishan Zhang, Tianjin University (China)

Jiang Guo, Massachusetts Institute of Technology (USA)

Qi Zhang, Fudan University (China)

Jun Lan, Alibaba

#### 19th CCL Local Arrangement Chairs:

Chunjie Cao, Hainan University (China)

Ting Jin, Hainan University (China)

#### 19th CCL Evaluation Chairs:

Ting Liu, Harbin Institute of Technology (China)

Wei Song, Capital Normal University (China)

#### 19th CCL Publications Chairs:

Shizhu He, Institute of Automation, CAS (China)

Gaoqi Rao, Beijing Language and Culture University (China)

#### 19th CCL Workshop Chairs:

Jun Xu, Renmin University of China (China)

Xipeng Qiu, Fudan University (China)

#### 19th CCL Sponsorship Chairs:

Zhongyu Wei, Fudan University (China) Binyang Li, University of International Relations (China)

#### 19th CCL Publicity Chairs:

Zhiyuan Liu, Tsinghua University (China)

#### 19th CCL System Demonstration Chairs:

Shujian Huang, Nanjing University (China) Zhaopeng Tu, Tencent (China)

#### 19th CCL Student Counseling Chairs:

Pengyuan Liu, Institute of Computing Technology, CAS (China)

#### 19th CCL Student Seminar Chairs:

Jinsong Su, Xiamen University (China) Hongyu Lin , Institute of Computing Technology, CAS (China )

#### 19th CCL Student Finance Chairs:

Yuxing Wang, Tsinghua University (China)

### **Table of Content**

<i>基于规则的双重否定识别——以"不v1 不v2"为例</i> 王昱·······1
基于语料库的武侠与仙侠网络小说文体、词汇及主题对比分析 张三乐, 刘鹏远, 张虎·······10
<i>基于计量的百年中国人名用字性别特征研究</i> 杜冰洁, 刘鹏远, 田永胜····································
<i>伟大的男人和倔强的女人:基于语料库的形容词性别偏度历时研究</i> 朱述承,刘鹏远····································
用计量风格学方法考察《水浒传》的作者争议问题——以罗贯中《平妖传》为参照 宋丽, 刘颖·······43
<i>多轮对话的篇章级抽象语义表示标注体系研究</i> 黄彤, 李斌, 闫培艺, 计婷婷, 曲维光······54
<i>发音属性优化建模及其在偏误检测的应用</i>
基于抽象语义表示的汉语疑问句的标注与分析 闫培艺,李斌,黄彤,霍凯蕊,陈瑾,曲维光······77
<i>语用视角下复述句生成方式的类型考察</i> 马天欢······88
<i>面向汉语作为第二语言学习的个性化语法纠错</i> 张生盛,庞桂娜,杨麟儿,王辰成,杜永萍,杨尔弘,黄雅平····································
<i>中文问句的形式分类和资源建设</i> 黎江涛,饶高琦·······107
<i>基于组块分析的汉语块依存语法</i> 钱青青,王诚文·······117
<i>新支话题的句法成分和语义角色研究</i> 卢达威······128
<i>眼动记录与主旨结构标注的关联性分析研究</i> 单昊聪,周强······141
汉语竞争类多人游戏语言中疑问句的形式与功能

张文贤, 苏琪153
<u>融合目标端句法的 AMR-to-Text 生成</u> 朱杰, 李军辉·······162
基于神经网络的连动句识别
孙超, 曲维光, 魏庭新, 顾彦慧, 李斌, 周俊生172
<b>融合全局和局部信息的汉语宏观篇章结构识别</b> 范亚鑫, 蒋峰, 褚晓敏, 李培峰, 朱巧明·······183
基于图神经网络的汉语依存分析和语义组合计算联合模型 汪凯,刘明童,陈圆梦,张玉洁,徐金安,陈钰枫······195
<i>基于强负采样的词嵌入优化算法</i> 王雨晨, 林淼哲, 詹杰凡····································
联合依存分析的汉语语义组合模型           陈圆梦, 张玉洁, 徐金安, 陈钰枫·······215
基于对话约束的回复生成研究           管梦雨, 王中卿, 李寿山, 周国栋····································
多模块联合的阅读理解候选句抽取           吉宇, 王笑月, 李茹, 郭少茹, 关勇
<i>基于层次化语义框架的知识库属性映射方法</i> 李豫,周光有····································
<i>面向垂直领域的阅读理解数据增强方法</i> 吕政伟,杨雷,石智中,梁霄,雷涛,刘多星·······256
<i>融入对话上文整体信息的层次匹配回应选择</i> 司博文, 孔芳····································
一种结合话语伪标签注意力的人机对话意图分类方法 丁健德,黄沛杰,许嘉宝,彭佑铭·······277
基于BERTCA 的新闻实体与正文语义相关度计算模型 向军毅, 胡慧君, 毛瑞彬, 刘茂福·······288
<i>基于多任务学习的生成式阅读理解</i>
基于多头注意力和 BiLSTM 改进 DAM 模型的中文问答匹配方法 秦汉忠,于重重,姜伟杰,赵霞····································
基于 Graph Transformer 的知识库问题生成

胡月,周光有3	24
<i>基于 BERT 与柱搜索的中文释义生成</i>	36
基于深度学习的实体关系抽取研究综述 夏振涛,曲维光,顾彦慧,周俊生,李斌····································	49
<i>小样本关系分类研究综述</i>	163
基于阅读理解框架的中文事件论元抽取 陈敏,吴凡,王中卿,李培峰,朱巧明·······37	76
<i>基于BERT 的端到端中文篇章事件抽取</i> 张洪宽, 宋晖, 王舒怡, 徐波····································	90
面向微博文本的融合字词信息的轻量级命名实体识别 陈淳,李明扬,孔芳····································	02
引入源端信息的机器译文自动评价方法研究 罗琪,李茂西····································	14
细粒度英汉机器翻译错误分析语料库"的构建与思考 裘白莲,王明文,李茂西,陈聪,徐凡····································	24
陈林卿,李军辉,贡正仙····································	34
满志博,毛存礼,余正涛,李训宇,高盛祥,朱俊国·············44  基于跨语言双语预训练及 Bi-LSTM 的汉-越平行句对抽取方法	46
刘畅, 高盛祥, 余正涛, 黄于欣, 尤丛丛···································	.57
	l67
闫晓东,解晓庆,邹煜,李维····································	79
白双成,呼斯勒····································	
沙九,周鹭琴,冯冲,李洪政,张天夫,慧慧···································	199

加羊东周,才智杰,才让卓玛,三毛措50
<i>《动词句法语义信息词典》知识内容说明书</i> 袁毓林,曹宏·······51
<i>面向中文AMR 标注体系的兼语语料库构建及识别研究</i> 侯文惠, 曲维光, 魏庭新, 李斌, 顾彦慧, 周俊生·······52
<i>面向人工智能伦理计算的中文道德词典构建方法研究</i> 王弘睿, 刘畅, 于东····································
<i>汉语否定焦点识别研究:数据集与基线系统</i> 盛佳璇,邹博伟,沈龙骧,叶静,洪宇·······55
<i>面向医学文本处理的医学实体标注规范</i> 张欢, 宗源, 常宝宝, 穗志方, 昝红英, 张坤丽······56
<i>汉语块依存语法与树库构建</i>
<i>汉语学习者依存句法树库构建</i> 师佳璐,罗昕宇,杨麟儿,肖丹,胡正声,王一君,袁佳欣,余婧思,杨尔弘······58
CDCPP: 跨领域中文标点符号预测         刘鹏远, 王伟康, 邱立坤, 杜冰洁····································
<i>多目标情感分类中文数据集构建及分析研究</i> 刘鹏远,田永胜,杜成玉,邱立坤·······60
基于 Self-Attention 的句法感知汉语框架语义角色标注 王晓晖,李茹,王智强,柴清华,韩孝奇······61
基于词语聚类的汉语口语教材自动推送素材研究 杨冰冰,赵慧周,王治敏·······624
基于半监督学习的中文社交文本事件聚类方法 郭恒睿, 王中卿, 李培峰, 朱巧明·······634
基于多粒度语义交互理解网络的幽默等级识别 张瑾晖,张绍武,樊小超,杨亮,林鸿飞·······64
<i>文本情感分析中的重叠现象研究</i> 娜仁图雅, 徐晓音····································
基于 BiLSTM-CRF 的社会突发事件研判方法 胡慧君, 王聪, 代建华, 刘茂福·······66
结合金融领域情感词典和注意力机制的细粒度情感分析

祝清麟, 梁斌, 刘宇瀚, 陈奕, 徐睿峰, 毛瑞彬676
基于层次注意力机制和门机制的属性级别情感分析 冯超,黎海辉,赵洪雅,薛云,唐靖尧····································
<i>基于循环交互注意力网络的问答立场分析</i>
<i>新型冠状病毒肺炎相关的推特主题与情感研究</i> 梁帅龙,黄辉,张岳·······707
<i>融入多尺度特征注意力的胶囊神经网络及其在文本分类中的应用</i> 王超凡, 琚生根, 孙界平, 陈润····································
<i>结合深度学习和语言难度特征的句子可读性计算方法</i> 唐玉玲, 于东········731
基于预训练语言模型的案件要素识别方法 刘海顺,王雷,陈彦光,张书晨,孙媛媛,林鸿飞·······743
基于拼音约束联合学习的汉语语音识别 梁仁凤,余正涛,高盛祥,黄于欣,郭军军,许树理·······754
基于数据增强和多任务特征学习的中文语法错误检测方法 谢海华, 陈志优, 程静, 吕肖庆, 汤帜···········761
基于有向异构图的发票明细税收分类方法 赵珮瑶,郑庆华,董博,阮建飞,罗敏楠······771
半监督跨领域语义依存分析技术研究         毛达展,李华勇,邵艳秋·······783
<i>汉英篇章衔接对齐语料构建研究</i> 李艳翠, 冯继克, 来纯晓, 冯洪玉·······795
Cross-Lingual Dependency Parsing via Self-Training  Meishan Zhang and Yue Zhang 807
A Joint Model for Graph-based Chinese Dependency Parsing  Xingchen Li, Mingtong Liu, Yujie Zhang, Jinan Xu and Yufeng Chen820
Semantic-aware Chinese Zero Pronoun Resolution with Pre-trained Semantic Dependency Parser  Lanqiu Zhang, Zizhuo Shen and Yanqiu Shao······831
Improving Sentence Classification by Multilingual Data Augmentation and Consensus Learning  Yanfei Wang, Yangdong Chen and Yuejie Zhang·····842
Attention-Based Graph Neural Network with Global Context Awareness for Document Understanding  Yuan Hua, Zheng Huang, Jie Guo and Weidong Oiu

Combining Impression Feature Representation for Multi-turn Conversational Question Answering
Shaoling Jing, Shibo Hong, Dongyan Zhao, Haihua Xie and Zhi Tang······863
Chinese Long and Short Form Choice Exploiting Neural Network Language Modeling Approaches  Lin Li, Kees van Deemter and Denis Paperno 874
Refining Data for Text Generation
Qianying Liu, Tianyi Li, Wenyu Guan and Sujian Li881
Plan-CVAE: A Planning-based Conditional Variational Autoencoder for Story Generation
Lin Wang, Juntao Li,Dongyan Zhao and Rui Yan·····892
Towards Causal Explanation Detection with Pyramid Salient-Aware Network
Xinyu Zuo, Yubo Chen, Kang Liu and Jun Zhao903
Named Entity Recognition with Context-Aware Dictionary Knowledge
Chuhan Wu, Fangzhao Wu, Tao Qi and Yongfeng Huang·····915
Chinese Named Entity Recognition via Adaptive Multi-pass Memory Network with Hierarchical Tagging Mechanism
Pengfei Cao, Yubo Chen, Kang Liu and Jun Zhao·····927
A Practice of Tourism Knowledge Graph Construction based on Heterogeneous Information  Dinghe Xiao, Nannan Wang, Jiangang Yu, Chunhong Zhang and Jiaqi Wu939
A Novel Joint Framework for Multiple Chinese Events Extraction
Nuo Xu, Haihua Xie and Dongyan Zhao·····950
Entity Relative Position Representation based Multi-head Selection for Joint Entity and Relation Extraction
Tianyang Zhao, Zhao Yan, Yunbo Cao and Zhoujun Li·····962
A Mixed Learning Objective for Neural Machine Translation
Wenjie Lu, Leiying Zhou, Gongshen Liu and Quanhai Zhang······974
Multi-Reward based Reinforcement Learning for Neural Machine Translation
Shuo Sun, Hongxu Hou, Nier Wu, Ziyue Guo and Chaowei Zhang·····984
Low-Resource Text Classification via Cross-lingual Language Model Fine-tuning
Xiuhong Li, Zhe Li, Jiabao Sheng and Wushour Slamu·····994
Constructing Uyghur Name Entity Recognition System using Neural Machine Translation Tag Projection  Azmat Anwar, Xiao Li, Yating Yang, Rui Dong and Turghun Osman······1006
Recognition Method of Important Words in Korean Text based on Reinforcement Learning
Feiyang Yang, Yahui Zhao and Rongyi Cui·····1017
Mongolian Quartions Classification Pased on Mulit Hoad Attention

Guangyi Wang, Feilong Bao and Weihua Wang······10	026
The Annotation Scheme of English-Chinese Clause Alignment Corpus  Shili Ge, Xiaopin Lin and Rou Song·····	035
Categorizing Offensive Language in Social Networks: A Chinese Corpus, Systems and an Explainal Tool  Xiangru Tang, Xianjun Shen, Yujie Wang, Yujuan Yang	
LiveQA: A Question Answering Dataset over Sports Live  Qianying Liu, Sicong Jiang, Yizhong Wang and Sujian Li	
Chinese and English Elementary Discourse Units Segmentation based on Bi-LSTM-CRF Model  Yancui Li, Chunxiao Lai, Jike Feng, Hongyu Feng10	)68
Better Queries for Aspect-Category Sentiment Classification  Yuncong Li, Cunxiang Yin, Sheng-hua Zhong, Huiqiang Zhong, Jinchang Luo, Siqi Xu and Xiao  Wu·····  10	
Multimodal Sentiment Analysis with Multi-perspective Fusion Network Focusing on Sense Attenta  Language  Xia Li and Minping Chen 10	
CAN-GRU: a Hierarchical Model for Emotion Recognition in Dialogue  Ting Jiang, Bing Xu, Tiejun Zhao and Sheng Li······11	101
A Joint Model for Aspect-Category Sentiment Analysis with Shared Sentiment Prediction Layer  Yuncong Li, Zhe Yang, Cunxiang Yin, Xu Pan, Lunan Cui, Qiang Huang and Ting Wei11	12
Compress Polyphone Pronunciation Prediction Model with Shared Labels  Pengfei Chen, Lina Wang, Hui Di, Kazushige Ouchi and Lvhong Wang······	122
Multi-task Legal Judgement Prediction Combining a Subtask of Seriousness of Charge Zhuopeng Xu, Xia Li, Yinlin Li, Zihan Wang, Yujie Fanxu and Xiaoyan Lai······11	32
Clickbait Detection with Style-aware Title Modeling and Co-attention  Chuhan Wu, Fangzhao Wu, Tao Qi and Yongfeng Huang········  11	143
Konwledge-Enabled Diagnosis Assistant Based on Obstetric EMRs and Knowledge Graph  Kunli Zhang, Xu Zhao, Xiabing Zhou, Qi Xie and Hongying Zan······	155
Reusable Phrase Extraction Based on Syntactic Parsing  Xuemin Duan, Hongying Zan, Xiaojing Bai and Christoph Zahner······11	166
WAE_RN: Integrating Wasserstein Autoencoder and Relational Network for Text Sequence Xinxin Zhang, Xiaoming Liu, Guan Yang and Fangfang Li	172