The 11th CCF International Conference on Natural Language Processing and Chinese Computing

NLPCC 2022 Handbook

22 September – 25 September, 2022
Guilin, China

Preface by the Program Committee co-Chairs

Welcome to NLPCC 2022, the eleventh CCF International Conference on Natural Language Processing and Chinese Computing. Following the success of previous conferences held in Beijing (2012), Chongqing (2013), Shenzhen (2014), Nanchang (2015), Kunming (2016), Dalian (2017), Hohhot (2018), Dunhuang (2019), Zhengzhou (2020), and Qingdao (2021), this year's NLPCC will be held in Guilin. As a premier international conference on natural language processing and Chinese computing, organized by the CCF-NLP (Technical Committee of Natural Language Processing, China Computer Federation, formerly known as Technical Committee of Chinese Information, China Computer Federation), NLPCC serves as an important forum for researchers and practitioners from academia, industry, and government to share their ideas, research results, and experiences, and to promote their research and technical innovations.

The fields of natural language processing (NLP) and Chinese computing (CC) have boomed in recent years. Following NLPCC's tradition, we welcomed submissions in ten areas for the main conference: Fundamentals of NLP; Machine Translation and Multilinguality; Machine Learning for NLP; Information Extraction and Knowledge Graph; Summarization and Generation; Question Answering; Dialogue Systems; Social Media and Sentiment Analysis; NLP Applications and Text Mining; Multimodality and Explainability. This year, despite the non-negligible influence of COVID-19, we still received 327 valid submissions to the main conference on the submission deadline.

After a thorough reviewing process, including meta reviewing, out of 327 submissions (some of which were desk-rejected due to policy violations), 83 papers were finally accepted as regular papers to appear in the main conference, where 73 were written in English and 10 in Chinese, resulting in an acceptance rate of 25.4%. Among them, 62 submissions will be presented as oral papers and 21 as poster papers at the conference. Six papers were nominated by our area chairs for the best paper award. An independent best paper award committee was formed to select the best papers from the shortlist. This proceeding includes only the accepted English papers; the Chinese papers will appear in the ACTA Scientiarum Naturalium Universitatis Pekinensis. In addition to the main proceedings, three papers were accepted to the Student workshop, 21 papers were accepted to the Evaluation workshop.

We are honored to have four internationally renowned keynote speakers, Jason Eisner (Johns Hopkins University), Ray Mooney (The University of Texas at Austin), Alexander Rush (Cornell University), and Luke Zettlemoyer (University of Washington), sharing their findings on recent research progress and achievements in natural language processing.

We would like to thank all the people who have contributed to NLPCC 2022. First of all, we would like to thank our 20 area chairs for their hard work recruiting reviewers, monitoring the review and discussion processes, and carefully rating and recommending submissions. We would like to thank all 330 reviewers for their time and efforts to review the submissions. We are very grateful to Preslav Nakov (chair), Min-Yen Kan, Pascale Fung, Minlie Huang, William Wang, Juanzi Li and Xuanjing Huang for their participation in the best paper committee. We are also grateful for the help and support from the general chairs, Bonnie Webber and Ya Zhou, and from the organization committee chairs, Guimin Huang and Xiaojun Wan. Special thanks go to Yu Hong and Xiabing Zhou, the publication chairs. We greatly appreciate all your help!

Finally, we would like to thank all the authors who submitted their work to NLPCC 2022, and thank our sponsors for their contributions to the conference. Without your support, we could not have such a strong conference program.

We are looking forward to seeing you at NLPCC 2022 in Guilin and hope you enjoy the conference!

Wei Lu, Singapore University of Technology and Design
Shujian Huang, Nanjing University
NLPCC 2022, Program Committee co-Chairs

Organizing Committee

Organizer
China Computer Federation (CCF)

Host
Guilin University of Electronic Technology

Publishers
Lecture Notes on Artificial Intelligence (LNAI), Springer Verlag
ACTA Scientiarum Naturalium Universitatis Pekinensis

General Chairs

Bonnie Webber	University of Edinburgh
Ya Zhou	Guilin University of Electronic Technology

Program Committee Chairs

Wei Lu	Singapore University of Technology and Design
Shujian Huang	Nanjing University

Student Workshop Chairs

Zhongqing Wang	Soochow University
Piji Li	Nanjing University of Aeronautics and Astronautics

Evaluation Chairs

Yunbo Cao	Tencent
Youzheng Wu	JD AI Research

Tutorial Chairs

Hai Zhao	Shanghai Jiao Tong University
Yang Feng	Institute of Computing Technology, Chinese Academy of Sciences

Publication Chairs

Yu Hong	Soochow University
Xiabing Zhou	Soochow University

Journal Coordinator

Yunfang Wu	Peking University

Conference Handbook Chair

Jun Li	Guilin University of Electronic Technology

Sponsorship Chairs

Haofen Wang	Tongji University
Ruifeng Xu	Harbin Institute of Technology (Shenzhen)
Guoyong Cai	Guilin University of Electronic Technology

Publicity Chairs

Jianxing Yu	Sun Yat-Sen University

Siyou Liu	Macao Polytechnic University

Organization Committee Chairs

Guimin Huang	Guilin University of Electronic Technology
Xiaojun Wan	Peking University

Area Chairs

Fundamentals of NLP

Nanyun Peng	University of California, Los Angeles
Zhenghua Li	Soochow University

Machine Translation and Multilinguality

Lei Li	University of California, Santa Barbara
Jiajun Zhang	Chinese Academy of Sciences

Machine Learning for NLP

Zhiting Hu	University of California, San Diego
Piji Li	Nanjing University of Aeronautics and Astronautics

Dialogue Systems

Lili Mou	University of Alberta
Weinan Zhang	Harbin Institute of Technology

Question Answering

Rui Zhang	Penn State University
Nan Yang	MSRA

Summarization and Generation

Chenghua Lin	The University of Sheffield
Deyi Xiong	Tianjin University

Information Extraction and Knowledge Graph

Ruihong Huang	Texas A&M University
Kang Liu	Chinese Academy of Sciences

Social Media and Sentiment Analysis

Anh Tuan Luu	Nanyang Technological University, Singapore
Rui Xia	Nanjing University of Science and Technology

NLP Applications and Text Mining

Jey Han Lau	The University of Melbourne
Lemao Liu	Tencent

Multimodality and Explainability

Lizi Liao	Singapore Management University
Mingxuan Wang	ByteDance

Treasurer

Yajing Zhang	Soochow University
Xueying Zhang	Peking University

Webmaster

Hui Liu	Peking University

Program Committee

Program Committee Chairs

Wei Lu	Singapore University of Technology and Design
Shujian Huang	Nanjing University

Program Committee Members

Bo An	The Institute of Ethnology and Anthropology, Chinese Academy of Social Sciences
Xiang Ao	Institute of Computing Technology, Chinese Academy of Sciences
Guirong Bai	Institute of Automation, Chinese Academy of Sciences
Yu Bao	ByteDance AI Lab
Qiming Bao	The University of Auckland
Junwei Bao	JD AI Research
Xiangrui Cai	Nankai University
Yi Cai	South China University of Technology
Yuan Cao	Google Brain
Yixin Cao	Singapore Management University
Pengfei Cao	Institute of Automation, Chinese Academy of Sciences
Jun Cao	ByteDance AI Lab
Zhangming Chan	Alibaba Group
Shuaichen Chang	The Ohio State University
Xiuying Chen	Peking University
Yufeng Chen	Beijing Jiaotong University
Xinchi Chen	Amazon AWS
Kehai Chen	School of Computer Science and Technology, Harbin Institute of Technology
Bo Chen	Institute of Software, Chinese Academy of Sciences
Yubo Chen	Institute of Automation, Chinese Academy of Sciences
Yi Chen	Harbin Institute of Technology, Shenzhen
Shuang Chen	Harbin Institute of Technology
Wei Chen	School of Data Science, Fudan University
Hanjie Chen	University of Virginia
Pei Chen	Texas A&M University
Yulong Chen	Zhejiang University, Westlake University
Jiangjie Chen	Fudan University
Chen Chen	Nankai University
liang chen	Peking University
Guanyi Chen	Utrecht University
Shanbo Cheng	Bytedance AI Lab
Chenhui Chu	Kyoto University
Yiming Cui	Joint Laboratory of HIT and iFLYTEK Research
mao cunli	kunming university of science and technology
Xiang Deng	The Ohio State University
Chenchen Ding	NICT
Qianqian Dong	ByteDance AI Lab
Zi-Yi Dou	UCLA

Longxu Dou	Harbin Institute of Technology
Rotem Dror	University of Pennsylvania
Xinya Du	UIUC
Jinhua Du	Investments AI, AIG
Chaoqun Duan	Harbin Institute of Technology
Junwen Duan	Central South University
Xiangyu Duan	Soochow University
Alex Fabbri	Salesforce AI Research
Zhihao Fan	Fudan University
Biaoyan Fang	The University of Melbourne
Zhiyuan Fang	Amazon
Xiachong Feng	Harbin Institute of Technology
Jiazhan Feng	Peking University
Yansong Feng	Peking University
Shi Feng	Northeastern University
Yang Feng	Institute of Computing Technology, Chinese Academy of Sciences
Mauajama Firdaus	University of Alberta
Lea Frermann	Melbourne University
Qiankun Fu	Zhejiang University
Xingyu Fu	UPenn
Guohong Fu	Soochow University
Jun Gao	Harbin Institute of Technology, Shenzhen
Shen Gao	Peking University
Ruiying Geng	Alibaba Group
Yu Gu	The Ohio State University
Jia-Chen Gu	University of Science and Technology of China
Yi Guan	School of Computer Science and Technology, Harbin Institute of Technology
Tao Gui	Fudan University
Shaoru Guo	Institute of Automation, Chinese Academy of Sciences
Jiale Han	Beijing University of Posts and Telecommunications
Lifeng Han	The University of Manchester
Xudong Han	The university of Melbourne
Tianyong Hao	School of Computer Science; South China Normal University
Yongchang Hao	University of Alberta
Hongkun Hao	Shanghai Jiao Tong University
Ziwei He	Shanghai Jiao Tong University
Ruifang He	Tianjin University
Yanqing He	Institute of Scientific and Technical Information of China
I-Hung Hsu	USC Information Sciences Institute
Jingwen Hu	Harbin Institute of Technology
Minghao Hu	Information Research Center of Military Science
Chenyang Huang	Dept. of Computing Science, University of Alberta
Ruihong Huang	Texas A&M University
Jiangping Huang	Chongqing University of Posts and Telecommunications
Qingbao Huang	Guangxi University, South China University of Technology
Fei Huang	Tsinghua University
Changzhen Ji	Harbin Institute of Technology
Tong Jia	Northeastern University
Hao Jia	School of Computer Science and Technology, Soochow University
Zhongtao Jiang	Institute of Automation, Chinese Academy of Sciences
Xuhui Jiang	Institute of Computing Technology, Chinese Academy of Sciences

Jingchi Jiang	Harbin Institute of Technology
Yong Jiang	Alibaba DAMO Academy
Wenbin Jiang	Baidu Inc.
Yiping Jin	FreeDa Language Space
Peng Jin	Leshan Normal University
Zhu Junguo	Kunming University of Science and Technology
Fajri Koto	The University of Melbourne
Tuan Lai	University of Illinois at Urbana-Champaign
Yuxuan Lai	Peking University
Yuanyuan Lei	Texas A&M University
Miao Li	The University of Melbourne
Irene Li	Yale University
Mingzhe Li	Peking University
Fei Li	Wuhan University
Fenghuan Li	Guangdong University of Technology
mingda Li	Harbin Institute of Technology
Jiajun Li	University of Chinese Academy of Sciences, Beijing, China; Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
Zhenghua Li	Soochow University
Maoxi Li	School of Computer Information Engineering, Jiangxi Normal University
Jing Li	Department of Computing, The Hong Kong Polytechnic University
Mingda Li	University of California, Los Angeles
Jiaqi Li	Harbin Institute of Technology
Yanran Li	The Hong Kong Polytechnic University
Shasha Li	College of Computer, National University of Defense Technology
Lishuang Li	Dalian University of Technology
Yanyang Li	The Chinese University of Hong Kong
Qintong Li	The University of Hong Kong
Bin Li	Nanjing Normal University
Xinyi Li	National University of Defense Technology
Zuchao Li	Shanghai Jiao Tong University
Hongzheng Li	School of Foreign Languages, Beijing Institute of Technology
zheng Li	Stockton University
Xin Li	Alibaba Group
Haonan Li	The University of Melbourne
Zhixu Li	Fudan University
Yucheng Li	University of Surrey
Chenliang Li	Wuhan University
Dongfang Li	Harbin Institute of Technology, Shenzhen
Zujie Liang	Sun Yat-Sen University
Lizi Liao	Singapore Management University
Ying Lin	Apple
Yuchen Liu	National Laboratory of Pattern Recognition, CASIA
Kang Liu	Institute of Automation, Chinese Academy of Sciences
Xiao Liu	Microsoft Research Asia
pengyuan liu	Beijing Language and Culture University
Xuebo Liu	Harbin Institute of Technology, Shenzhen
Qian Liu	Beihang University

Yongbin Liu	School of Computer Science, University of South China
Qingbin Liu	Tencent
Chunhua Liu	The University of Melbourne
Chuang Liu	Tianjin University
Yan Liu	Tianjin University
Xianggen Liu	Sichuan University
Yuanxing Liu	Harbin Institute of Technology
Jian Liu	Beijing Jiaotong University
zhicheng liu	ByteDance
Lemao Liu	Tencent AI Lab
Qun Liu	Chongqing University of Posts and Telecommunications
Shujie Liu	Microsoft Research Asia, Beijing, China
Puyuan Liu	University of Alberta
Yaojie Lu	Institute of Software, Chinese Academy of Sciences
Xin Lu	Harbin Institute of Technology
Hengtong Lu	Beijing University of Posts and Telecommunications
Yinglong Ma	North China Electric Power University
Longxuan Ma	Research Center for Social Computing and Information Retrieval, Harbin Institute of Technology
Mingyu Derek Ma	UCLA
Xuezhe Ma	Information Sciences Institute
Yunshan Ma	National University of Singapore
Zhao Meng	ETH Zurich
Tao Mingxu	Peking University
Xiangyang Mou	Meta
Yulia Otmakhova	University of Melbourne
Jiaxin Pei	University of Michigan
Xutan Peng	The University of Sheffield
Jianzhong Qi	The University of Melbourne
Jun Qi	Georgia Institute of Technology
Xian Qian	ByteDance AI LAB
Lihua Qian	ByteDance
Tao Qian	Hubei University of Science and Technology
Yanxia Qin	School of Computing, National University of Singapore
Libo Qin	Harbin Institute of Technology
Liang Qiu	Amazon Alexa AI
Yuqi Ren	Tianjin University
Shuo Ren	Microsoft Research Asia
Yubin Ruan	Harbin Institute of Technology
Lei Sha	University of Oxford
Wei Shao	City University of Hong Kong
Lingfeng Shen	Johns Hopkins University
Haoran Shi	Amazon Inc.
Xing Shi	Bytedance Inc.
Lei Shu	Google Research
Jyotika Singh	Placemakr
Linfeng Song	Tencent AI Lab
Kaiqiang Song	Tencent AI Lab
Zhenqiao Song	ByteDance AI Lab
Haoyu Song	Harbin Institute of Technology

Jinsong Su	Xiamen university
Dianbo Sui	Institute of Automation, Chinese Academy of Sciences
Kai Sun	Cornell University
Zewei Sun	ByteDance
Chengjie Sun	Harbin Institute of Technology
Simon Suster	University of Melbourne
Zhixing Tan	Tsinghua University
Minghuan Tan	Singapore Management University
Ping Tan	Universiti Malaysia Sarawak
Yun Tang	Facebook
Buzhou Tang	Harbin Institute of Technology (Shenzhen)
Duyu Tang	Tencent
Jialong Tang	Institute of Software, Chinese Academy of Sciences
Xunzhu Tang	University of Luxembourg
Zhiyang Teng	Westlake University
Lin Tian	RMIT University
Thinh Hung Truong	The University of Melbourne
Zhaopeng Tu	Tencent AI Lab
Gisela Vallejo	The University of Melbourne
Lijie Wang	Baidu
Le Wang	Chinese Academy of Military Science
Tao Wang	Department of Biostatistics & Health Informatics, King's College London
Xing Wang	Tencent
Shaolei Wang	Harbin Institute of Technology
Yunli Wang	Beihang University
Wei Wang	Shenzhen International Graduate School, Tsinghua University
Ruize Wang	Academy for Engineering and Technology, Fudan University
Hongling Wang	Soochow University
Danqing Wang	Bytedance AI Lab
Liang Wang	Microsoft Research Asia
Yuxuan Wang	Zhejiang Lab
Lingzhi Wang	The Chinese University of Hong Kong
Siyuan Wang	Fudan University
Zijian Wang	AWS AI Labs
Jun Wang	University of Melbourne
Sijia Wang	Virginia Tech.
Bo Wang	Tianjin University
Yaqiang Wang	Chengdu University of Information Technology
Yufei Wang	Macquaire University
Hongwei Wang	Tencent AI Lab
Qingyun Wang	University of Illinois at Urbana-Champaign
Tao Wang	ByteDance AI Lab
Zhen Wang	The Ohio State University
Jiaqiang Wang	ByteDance
Xuesong Wang	Harbin Institute of Technology
Xuepeng Wang	Tencent AI Lab
Xiao Wang	Fudan University
Wei Wei	Huazhong University of Science and Technology
Yang Wei	Bytedance AI Lab
Haoyang Wen	Carnegie Mellon University

Yuqiao Wen	University of Alberta
Sixing Wu	School of Computer Science, Peking University
Yuxia Wu	Xi'an Jiaotong University
shun wu	Institute of Automation，Chinese Academy of Sciences
Liwei Wu	Bytedance AI Lab
Lijun Wu	Microsoft Research
Changxing Wu	East China Jiaotong University
Qingrong Xia	Soochow University
Congying Xia	Salesforce Research
Yang Xiang	Peng Cheng Laboratory
Tong Xiao	Northeastern University
ruiyu Xiao	Harbin Institute of Technology
Jun Xie	Alibaba DAMO Academy
Yuqiang Xie	Institute of Information Engineering, Chinese Academy of Sciences
Xin Xin	Beijing Institute of Technology
kang xu	Nanjing University of Posts and Telecommunications
Yan Xu	Hong Kong University of Science and Technology
Wenda Xu	University of California at Santa Barbara
Xiao Xu	Harbin Institute of Technology
Zhixing Xu	Nanjing Normal University
Jinan Xu	Beijing Jiaotong University
Peng Xu	Google
Ying Xu	Oracle Australia
jiahao xu	Nanyang Technological University
Yiheng Xu	Microsoft Research Asia
Wang Xu	Harbin Institute of Technology
Lingyong Yan	Baidu Inc.
Yuanmeng Yan	Beijing University of Posts and Telecommunications
Hang Yang	Institute of Automation, Chinese Academy of Sciences
Ziqing Yang	iFLYTEK Research
Liner Yang	Beijing Language and Culture University
Shiquan Yang	The University of Melbourne
Ziqing Yang	Peking University
Muyun Yang	Harbin Institute of Technology
Jun Yang	Marcpoint Co.,Ltd.
Qiang Yang	KAUST
Baosong Yang	Alibaba Damo Academy, Alibaba Inc.
Songlin Yang	ShanghaiTech University
Zhiwei Yang	College of Computer Science and Technology, Jilin University
Liang Yang	Dalian University of Technology
Haoran Yang	The Chinese University of Hong Kong
Wenlin Yao	Tencent AI Lab
Jianmin Yao	Soochow University
Rong Ye	ByteDance AI Lab
Tiezheng Yu	The Hong Kong University of Science and Technology
Dong Yu	Beijing Language and Culture University
Junjie Yu	Soochow University
Zhe Yu	Sun Yat-sen University
Heng Yu	Alibaba

Pengfei Yu	Department of Computer Science, University of Illinois at Urbana-Champaign
Guoxin Yu	Institute of Computing Technology, Chinese Academy of Sciences
Chunyuan Yuan	Institute of Information Engineering，Chinese Academy of Sciences
Xiang Yue	The Ohio State University
Xiangrong Zeng	Kuaishou
Qi Zeng	University of Illinois at Urbana-Champaign
Daojian Zeng	Hunan Normal University
Shuang (Sophie) Zhai	University of Oklahoma
Weidong Zhan	Dept. of Chinese, Peking University
Wenxuan Zhang	DAMO Academy, Alibaba Group
Zhirui Zhang	Tencent AI Lab
Zhihao Zhang	Beihang University
Shuaicheng Zhang	Virginia Polytechnic Institute and State University
Peng Zhang	Tianjin University
Xiuzhen Zhang	RMIT University
Xingxing Zhang	Microsoft Research Asia
Dakun Zhang	SYSTRAN
Yuanzhe Zhang	Institute of Automation, Chinese Academy of Sciences
Zhuosheng Zhang	Shanghai Jiao Tong University
Biao Zhang	University of Edinburgh
Yazhou Zhang	Zhengzhou University of Light Industry
Tongtao Zhang	Siemens Corporate Technology
Kaiyan Zhang	Harbin Institute of Technology
Mengjie Zhao	Center for Information and Language Processing, LMU Munich
Zhenjie Zhao	Nanjing University of Information Science and Technology
Xiang Zhao	National University of Defense Technology
Jun Zhao	Fudan University
YUFAN ZHAO	Microsoft
Zhedong Zheng	National University of Singapore
Wanjun Zhong	Sun Yat-Sen University
Bo Zhou	Institute of Automation, Chinese Academy of Sciences
Guangyou Zhou	School of Computer Science, Central China Normal University
Wenxuan Zhou	University of Southern California
Chunting Zhou	Meta AI
Wangchunshu Zhou	Beihang Univeristy
Xin Zhou	Fudan University
Peng Zhou	Kuaishou
Junnan Zhu	Institute of Automation, Chinese Academy of Sciences
Qingfu Zhu	Harbin Institute of Technology
Tong Zhu	Soochow University
Rongxin Zhu	The University of Melbourne
Yaoming ZHU	ByteDance AI lab
Jie Zhu	Soochow University
Conghui Zhu	Harbin Institute of Technology
Muhua Zhu	Meituan Group
Zhou Qian	ByteDance AI Lab

Organizers and Supports


Diamond Sponsors
Alibaba Group
阿里巴巴集团
Platinum Sponsors
HUAWEI
Tencent AI Lab
Baidu 百度
GTCOM
ByteDance
字节跳动
Golden Sponsors
乐言科技
Leyan Technologies
mi
vivo
PARATERA 并行
中业科技
zoneyet

Contents

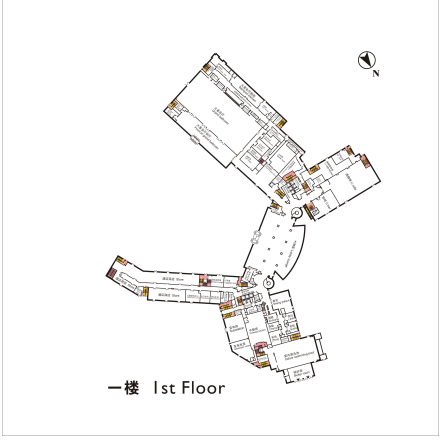
1 Organization

NLPCC (CCF International Conference on Natural Language Processing & Chinese Computing) is the annual conference of CCF TCNLP (Technical Committee of Natural Language Processing, China Computer Federation). As a leading conference on the field of NLP & Chinese Computing of CCF, NLPCC is the premier forum for the NLP researchers and practitioners from academia, industry, and government in China and Pacific Asia area to share their ideas, research results and experiences, which will highly promote the research and technical innovation in these fields domestically and internationally.

2 Conference Venues

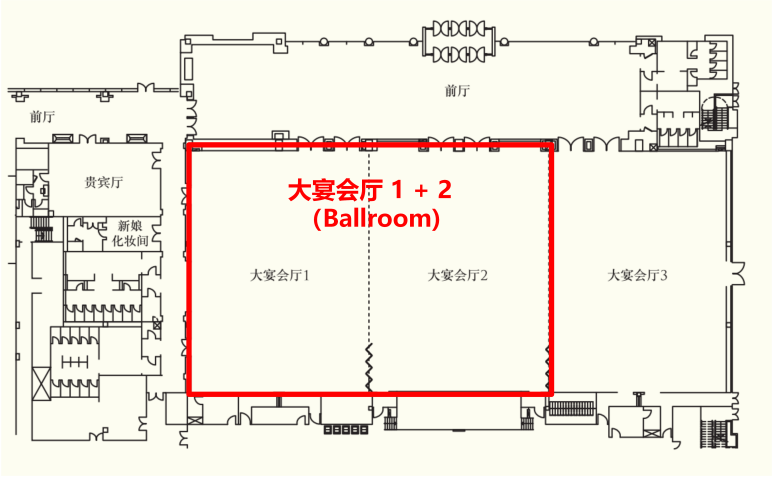
Venue

Shangri-La Hotel Guilin, 桂林香格里拉大酒店
1st Floor Plan

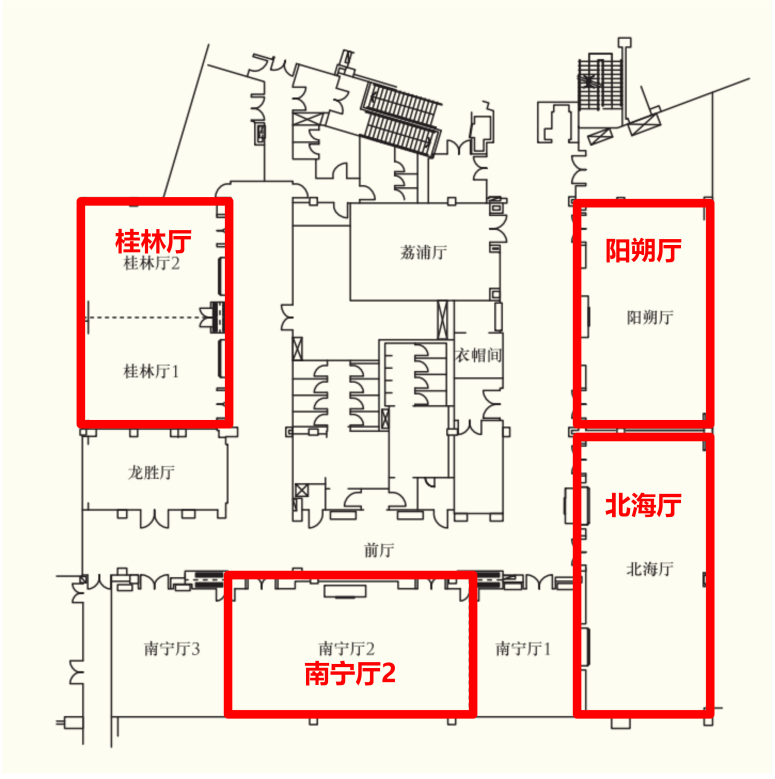


Main Conference & Workshops

Shangri-La Hotel, Guilin
1st Floor Ballroom 1 & 2 Diagram



Shangri-La Hotel, Guilin
2nd Floor Boardroom Diagram



3 Conference Agenda

Tutorials (Online)

Sep. 22		
08:55-09:00	Opening Remarks	Tutorial Chairs
09:00-10:30	Lecture 1: Language Understanding in Task-oriented Dialogue Systems	Wanxiang Che
10:45-12:15	Lecture 2: Cross-modal Knowledge Acquisition and Fusion	Peng Li
14:00-15:30	Lecture 3: Affective Computing for Recommender Systems	Chenliang Li
15:45-17:15	Lecture 4: Quantum Uncertainty in Natural language	Peng Zhang
18:45-20:15	Lecture 5: Conversational Information Seeking and Recommendation	Zhaochun Ren
20:30-22:00	Lecture 6: Automatic Solving Math Problem in Neuro-symbolic Computing	Yichun Yin
Sep. 23		
09:00-10:30	Lecture 7: Towards Automated Fact Checking	Yansong Feng
10:45-12:15	Lecture 8: K-nearest-neighbor Machine Translation	Shujian Huang

Main Conference & Workshops

Sep. 24 Main Conference & Workshops

08:30-09:00	NLPCC 2022 Opening Ceremony	Ballroom 1&2, 1st floor
09:00-10:00	Keynote Talk 1, by Raymond J. Mooney (Online)	Ballroom 1&2, 1st floor
10:00-10:15	Industrial Talk, by Jun Xie	Ballroom 1&2, 1st floor
10:15-10:45	Break & Group Photo	Outdoor plaza
10:45-11:45	Keynote Talk 2, by Jason Eisner (Online)	Ballroom 1&2, 1st floor
11:30-13:30	Lunch	Li Cafe (Buffet), 1st floor
13:30-15:30	Posters Session 1	2nd floor public area
	Oral 1: NLP Fundementals	Yangshuo room (阳朔厅), 2nd floor
	Oral 2: NLP Applications	Beihai room (北海厅), 2nd floor
	Student Workshop	Guilin room (桂林厅), 2nd floor
	Evaluation Workshop 1	Nanning room 2 (南宁厅 2), 2nd floor
15:30-16:00	Break	
16:00-18:00	Posters Session 2	2nd floor public area
	Oral 3: NLP Fundamentals	Yangshuo room (阳朔厅), 2nd floor
	Oral 4: Question Answering	Beihai room (北海厅), 2nd floor
	Pretrained Models and Applications Workshop	Guilin room (桂林厅), 2nd floor
	Evaluation Workshop 2	Nanning room 2 (南宁厅 2), 2nd floor
18:30-21:00	Dinner	Ballroom 1&2 (Banquet), 1st floor

Sep. 25 Main conference & Workshop

09:00-10:00	Keynote Talk 3, by Luke Zettlemoyer (Online)	Ballroom 1&2, 1st floor
10:00-10:30	Break	
10:30-11:30	Keynote Talk 4, by Alexander Rush (Online)	Ballroom 1&2, 1st floor
11:30-13:30	Lunch	Li Cafe (Buffet), 1st floor
13:30-15:30	Oral 5: Sentiment Analysis	Yangshuo room (阳朔厅), 2nd floor
	Oral 6: Information Extraction	Beihai room (北海厅), 2nd floor
	Oral 7: Machine Learning	Guilin room (桂林厅), 2nd floor
	Low-Resource NLP Workshop 1	Nanning room 2 (南宁厅 2), 2nd floor
15:30-15:45	Break	
15:45-17:00	Oral 8: Machine Translation	Yangshuo room (阳朔厅), 2nd floor
	Oral 9: Diaglog	Beihai room (北海厅), 2nd floor
	Oral 10: Generation	Guilin room (桂林厅), 2nd floor
	Low-Resource NLP Workshop 2	Nanning room 2 (南宁厅 2), 2nd floor
17:15-17:45	NLPCC Awards & Closing Ceremony	Ballroom 1&2, 1st floor
18:00-21:00	Dinner	Li Cafe (Buffet), 1st floor

4

Tutorials : Sep, 22 – 23 (Online)

Schedule

Thursday, 22 Sep (Online)

08:55-09:00	Opening Remarks	Tutorial Chairs
09:00-10:30	Lecture 1: Language Understanding in Task-oriented Dialogue Systems	Wanxiang Che
10:45-12:15	Lecture 2: Cross-modal Knowledge Acquisition and Fusion	Peng Li
14:00-15:30	Lecture 3: Affective Computing for Recommender Systems	Chenliang Li
15:45-17:15	Lecture 4: Quantum Uncertainty in Natural language	Peng Zhang
18:45-20:15	Lecture 5: Conversational Information Seeking and Recommendation	Zhaochun Ren
20:30-22:00	Lecture 6: Automatic Solving Math Problem in Neuro-symbolic Computing	Yichun Yin

Friday, 23 Sep (Online)

09:00-10:30	Lecture 7: Towards Automated Fact Checking	Yansong Feng
10:45-12:15	Lecture 8: K-nearest-neighbor Machine Translation	Shujian Huang

Tutorial 1

Language Understanding in Task-oriented Dialogue Systems

Wanxiang Che

Thursday, 22 Sep, 09:00 – 10:30

Abstract: Task-based dialogue systems have received extensive attention from industry and academia. Among them, dialogue language understanding (DLU) is the core component, which has developed rapidly in recent years. This report first reviews and summarizes the development of the DLU field in recent years, especially the methods in fewer-labeled scenarios, and then gives the future development trend of the DLU field.

Lecturer: Wanxiang Che is a professor and doctoral supervisor of Computing Faculty at Harbin Institute of Technology, Deputy Dean of the Institute of Artificial Intelligence, and Deputy Director of the Research Center for Social Computing and Information Retrieval. Heilongjiang Province "Longjiang Scholar" young scholar, visiting scholar at Stanford University. He is currently a deputy director and secretary-general of the Professional Committee of Computational Linguistics of the Chinese Information Society of China; executive committee and secretary-general of the Asia-Pacific Chapter of the Association for Computational Linguistics (AACL); a senior member of the China Computer Federation. He has published more than 200 academic papers in high-level journals and conferences at home and abroad, such as ACL, EMNLP, AAAI, and IJCAI, among which AAAI's 2013 paper won the Outstanding Paper Honorable Mention Award. He has published 4 textbooks and 2 translated books. At present, he is undertaking a number of scientific research projects such as the 2030 "New Generation Artificial Intelligence" major project and the National Natural Science Foundation of China. He has won many awards such as the 2020 Heilongjiang Youth Science and Technology Award.

Tutorial 2

Cross-modal Knowledge Acquisition and Fusion

Peng Li

Thursday, 22 Sep, 10:45 – 12:15

Abstract: Text, image, and audio signals carry rich knowledge of the world and human experience. Due to their inherent differences in various aspects, knowledge does not distribute uniformly across these signals. Acquiring and fusing knowledge from multi-modal signals to improve both the quantity and quality of knowledge accessible by models has attracted extensive attention in recent years. In this talk, we will summarize the progress of cross-modal knowledge acquisition and fusion, hoping to highlight the trends in this field. And we will also discuss the future directions worth pursuing in this field.

Lecturer: **Peng Li** is a Research Associate Professor at Institute for AI Industry Research (AIR), Tsinghua University. His main research interests include natural language processing, pre-trained models, cross-modal information processing, question answering, information extraction, machine translation, and dialogue system. He has published more than 60 papers in international conferences and journals on artificial intelligence and won the first place on a few highly influential leaderboards. He served as the Area Chair of COLING 2022 and the Senior Area Chair of AACL 2022. He was the recipient of First Prize of Qian Weichang Chinese Information Processing Science and Technology Award and Second Prize of Chinese Institute of Electronics Science and Technology Progress Award.

Tutorial 3

Affective Computing for Recommender Systems

Chenliang Li

Thursday, 22 Sep, 14:00 – 15:30

Abstract: Product reviews contain a wealth of semantic information that provides a basis for us to enhance personalized recommendation. The powerful semantic extraction and comprehension capacity of deep neural network can overcome the defects of the bag-of-word model in traditional text semantic analysis, help us better portray the user's preferences and product characteristics, and also provide material for the interpretability of recommendations through affective computing. This tutorial will introduce the review-based recommendation techniques in recent years, and will mainly focuse the design choices and the interpretability of recommendation, and finally look discuss the potential trend of review-based recommendation systems.

Lecturer: Chenliang Li is a full professor at School of Cyber Science and Engineering, Wuhan University. His research interests inlcude information retrieval, natural language processing and social media analysis. He has published over 70 research papers on leading academic conferences and journals such as SIGIR, ACL, WWW, IJCAI, AAAI, TKDE and TOIS. He has served as Associate Editor / Editorial Board Member for ACM TOIS, ACM TALLIP, IPM and JASIST. His research won the SIGIR 2016 Best Student Paper Honorable Mention and TKDE Featured Spotlight Paper. He is a recipient of ACM Wuhan Rising Star Award.

Tutorial 4

Quantum Uncertainty in Natural language

Peng Zhang

Thursday, 22 Sep, 15:45 – 17:15

Abstract: In recent years, the cross research of quantum theory and artificial intelligence has attracted more and more attention. This lecture will explain the fundamental connection between natural language and quantum mechanics, and introduce the frontier progress of quantum natural language processing and future challenges for the key scientific problems of quantum uncertainty in natural language.

Lecturer: **Peng Zhang** is a professor and doctoral supervisor, vice Dean of the School of Computer Science of Tianjin University. He has devoted himself to the research work of quantum information retrieval and quantum language modeling for more than ten years, and actively promoted the application of the quantum language model. He published top conference papers including NeurIPS, SIGIR, ICLR, ACL, IJCAI, AAAI, WWW, CIKM, EMNLP and leading journal papers such as TNNLS, TKDE, TIST, IP&M. He has received SIGIR 2017 Best Paper Award Honorable Mention and ECIR 2011 Best Poster Award.

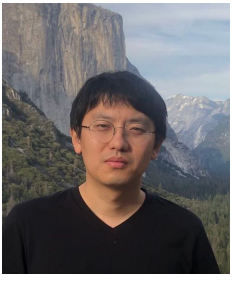
Tutorial 5

Conversational Information Seeking and Recommendation

Zhaochun Ren

Thursday, 22 Sep, 18:45 – 20:15

Abstract: In contrast to traditional information retrieval and recommendation systems, conversational information seeking and conversational recommender systems aim to solve search and recommendation tasks through conversational interactions between the system and the user. As a new paradigm for interacting with search engines, conversational information seeking can better capture user intent through conversational interactions, and users can get answers directly without having to search results, in response to the inflexibility of traditional retrieval for complex intent queries. Conversational recommender systems can reveal users' current preferences more quickly through real-time, multi-round online conversational interactions and better understand the reasons behind consumer behaviour. In recent years, research on conversational information seeking and conversational recommender systems has received increasing attention and has been developing rapidly. This tutorial will summarize recent studies on conversational information seeking techniques and conversational recommendation from different perspectives, and present the progress of research on conversational information seeking and conversational recommender systems in terms of methods, data, evaluation metrics, and the challenges.



Lectuer: Zhaochun Ren is a professor at Shandong University. His research interests lie in information retrieval and natural language processing, with emphasis on conversational information retrieval, recommender systems, social media analysis, dialogue systems, opinion mining, and summarization. I aim to develop intelligent agents that can address complex user requests and solve core challenges in IR and NLP towards that goal. His research has appeared at various prestigious conferences and journals, including SIGIR, WWW, ACL, KDD, WSDM, CIKM, AAAI, IJCAI, EMNLP, ACM TOIS, IEEE TKDE, IPM, etc.. He has received the Best Full Paper Runner Up award in CIKM 2017, and the Best Student Paper award in WSDM 2018.

Tutorial 6

Automatic Solving Math Problem in Neuro-symbolic Computing

Yichun Yin

Thursday, 22 Sep, 20:30 – 22:00

Abstract: Automatically solving mathematical problems is a fundamental challenge in neural symbolic computing, which is divided into: (1) Formal mathematical problems rely on pre-built formal systems, requiring models to perform multi-step reasoning and interact with the formal system to complete problem solving or theorem proving. (2) Informal math problems do not depend on formal systems. Given a mathematical problem described in natural language, the model needs to complete the reasoning process and return the final answer by itself. Recently pre-trained language models have been successfully applied to formal/informal math problems, but SOTA results are still below expert level. There are still many unanswered questions on this topic. In this tutorial, we will first introduce the definition and background of mathematical problems, then take two classic problems, automatic theorem proving and math word problems as examples, review the technical development, and finally discuss the future direction of the topic.

Lecturer: Yichun Yin received his Ph.D. from Peking University in 2018. Currently working as a senior researcher in Speech and Semantics Lab of Huawei's Noah's Ark, he is mainly engaged in the research of efficient pre-trained language models and neuro-symbolic computing. He published many papers at natural language processing conferences such as ACL and EMNLP, one of which is the most cited paper in EMNLP2020.

Tutorial 7

Towards Automated Fact Checking

Yansong Feng

Friday, 23 Sep, 09:00 – 10:30

Abstract: Automated Fact Checking aims to verify a given claim according to retrieved evidence, and has received increasing attention in recent years. The key components to solve this task is to collect sufficient relevant evidence according to the given claim from multiple sources, digest the evidence, and finally make the veracity prediction. The community has made promising progress by addressing multiple evidence formats, the fusion of evidence pieces, model explainability and robustness. In this tutorial, we will provide a gentle introduction to automated fact checking, including the task setups, datasets, models, and applications. We will focus on the most recent advance, including effective evidence selection and organization, fusion of structured and unstructured evidence pieces, and how to improve model robustness in real-world scenarios.

Lecturer: **Yansong Feng** is an associate professor in the Wangxuan Institute of Computer Technology at Peking University. Before that, he obtained his PhD from ICCS (now ILCC) at the University of Edinburgh. His current research interests include using probabilistic methods to distill knowledge from large volumes of natural language texts, and supporting intelligent human-computer interfaces, such as question answering and dialogue. He has served as Action Editor and Area Chair for ARR and ACL conferences. Yansong received the IBM Faculty Award in 2014 and 2015, and the IBM Global Shared University Research Award in 2016.

Tutorial 8

K-nearest-neighbor Machine Translation

Shujian Huang

Friday, 23 Sep, 10:45 – 12:15

Abstract: One important change for machine translation in the deep learning era is that the translation knowledge are no longer represented in a symbolic way but embedded in the parameters of the neural networks. However, even large scale neural networks cannot learn all the knowledge in the training data, especially for the low frequency events. k-nearest-neighbor machine translation is a retrieval based technique. kNNMT employs a translation datastore with symbolic translation knowledge to assist neural machine translation models, showing great potential in modeling low-frequency events, fast adaptation, etc. This talk covers both the basis of kNN-MT and its recent advances, including dynamically integrate the symbolic knowledge into the neural system, how to control the size of the symbolic knowledge base and the interpretability of this framework.

Lecturer: Shujian Huang is currently an associate professor and phd advisor in Nanjing University. His research interests includes machine translation, text analysis and understanding, etc. He is now the dupty director of the Techinical Committee on Machine Translation of CIPSC, and senior member of CCF. He currently serves as AC/AE/SPC for major NLP/AI conferences such as ACL, EMNLP, NAACL, AAAI, IJCAI, etc.

5

Main Conference & Workshops : Sep, 24

	Session		Venue
08:30-09:00	NLPCC 2022 Opening Ceremony		Ballroom 1&2, 1st floor
09:00-10:00	Keynote Talk 1, by Raymond J. Mooney (Online)		
10:00-10:15	Industrial Talk, by Jun Xie		
10:15-10:45	Break & Group Photo		Outdoor plaza
10:45-11:45	Keynote Talk 2, by Jason Eisner (Online)		Ballroom 1&2, 1st floor
11:30-13:30	Lunch		Li Cafe (Buffet), 1st floor
13:30-15:30	Sessions & Workshop	Chair	Venue
	Posters Session 1	-	2nd floor public area
	Oral 1: NLP Fundemental	Wenliang Chen	Yangshuo room (阳朔厅)
	Oral 2: NLP Applications	Xianling Mao	Beihai room (北海厅)
	Student Workshop	Zhongqing Wang Piji Li	Guilin room (桂林厅)
	Evalution Workshop 1	Yunbo Cao	Nanning room 2 (南宁厅 2)
15:30-16:00	Break		
16:00-18:00	Sessions & Workshop	Chair	Venue
	Posters Session 2	-	2nd floor public area
	Oral 3: NLP Fundamentals	Yue Zhang	Yangshuo room (阳朔厅)
	Oral 4: Question Answering	Xiaojun Quan	Beihai room (北海厅)
	Pretrained Models and Applications Workshop	Xiaojun Wan	Guilin room (桂林厅)
	Evaluation Workshop 2	Youzheng Wu	Nanning room 2 (南宁厅 2)
18:00-21:00	Dinner		Ballroom 1&2 (Banquet), 1st floor

Keynote Talk 1 (Online)

Sep. 24 , Chair: *Wei Lu*
09:00-10:00, Ballroom 1&2, 1st floor

Answering Why Questions about Narrative Text

Raymond J. Mooney

Answering questions about why people perform particular actions is central to understanding and reasoning about narrative text. Despite recent progress in QA, it is not clear if existing models have the ability to answer "why" questions, which generally require using commonsense knowledge external to the narrative and inferring people's plans and goals. We have been developing novel data and methods for such why-question answering. TellMeWhy is a new crowd-sourced corpus consisting of more than 30k questions and free-form answers concerning why characters in short narratives perform the actions described. Given the limitations of automated evaluation for this task, we have also designed a systematized human evaluation interface for this dataset. Our evaluation of state-of-the-art models show that they are significantly below human performance on answering such questions. We have also explored what aspects of the knowledge required to answer why questions are accessible in current large language models and what aspects can be made accessible via external commonsense knowledge resources. Not surprisingly, larger models perform better, but all of the variable-sized models we explored benefited from the injection of question-specific knowledge from the COMET knowledge base. We also developed a simple ontology of knowledge types and analyzed the relative coverage of different models across these categories.

Raymond J. Mooney is a Professor in the Department of Computer Science at the University of Texas at Austin. He received his Ph.D. in 1988 from the University of Illinois at Urbana/Champaign. He is an author of over 180 published research papers, primarily in the areas of machine learning and natural language processing. He was the President of the International Machine Learning Society from 2008-2011, program co-chair for AAAI 2006, general chair for HLT-EMNLP 2005, and co-chair for ICML 1990. He is a Fellow of AAAI, ACM, and ACL and the recipient of the Classic Paper award from AAAI-19 and best paper awards from AAAI-96, KDD-04, ICML-05 and ACL-07.

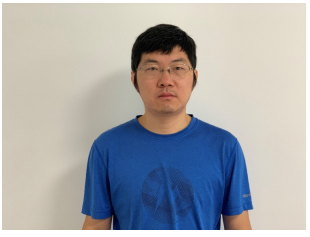
Industrial Talk

Sep. 24, Chair: *Wei Lu*
10:00-10:15, Ballroom 1&2, 1st floor

Enhanced Model, Unified Quality Evaluation, and Light-Weight Domain Adaptation for Nerual Machine Translation

Jun Xie
Alibaba Damo Academy

High quality, trustability and transferability is always fundemental and crucial for a practical machine translation system to meet different demands on different domains. In this talk, I would like to introduce our efforts on these three aspects, CSANMT (continous semantic augmented neural machine translation), UniTE (Unified Translation Evaluation) , Adapter-NMT and revised KNN-MT, respectively. CSANMT equips Transformer with a continuous semantic space to simulate the real distribution of semantic equivalence between source and target languages, which is capable to make more unseen instances under generalization with very limited training data. UniTE is a universial framework with better human correlation, which can evaluate the translation quality of hypothesis with source (quality estimation setting), reference (metrics setting), and both source and reference. With Adapter-NMT and revised KNN-MT, we can transfer a base NMT model to new domains with few new trainable parameters or even without training procedure at all.



Jun Xie is a principal algorithm engineer at Alibaba Damo Academy. He recieved his Ph.D. in 2012 from ICT The Institute of Computing Technology of the Chinese Academy of Sciences. His research focuses on machine translation, text generation and natural language processing. And he has published 20+ research papers at top-tier conferences including ACL, EMNLP, AAAI. He is now working on a high quality practical neural machine translation system which supports multiple langages and is flexible to transfer to multiple domains.

Keynote Talk 2 (Online)

Sep. 24 , Chair: *Shujian Huang*
10:45-11:45, Ballroom 1&2,1st floor

Imposing Structure on Sequence Models

Jason Eisner

LSTMs and Transformers do very well at predicting the future from the past. They develop predictive internal representations based on large training sets. Sometimes, however, we may like to use domain knowledge to enforce hard constraints on the output. The same handwritten grammar rules that enforce constraints can also be used to guide representation learning. By structuring how the model attends to the past, this strategy may improve generalization and interpretability.After giving some introductory examples, I will present two interesting architectures that make use of rules.A finite-state transducer (FST) is a classical architecture that can be built from a regular expression. It describes the legal outputs that can be aligned to each input, in terms of a path through a labeled graph. I'll describe how to featurize those paths and score them neurally, breaking the usual Markov assumption in a way that requires approximate inference algorithms such as beam search or particle smoothing. The neural architecture can be made to follow the structure of the regular expression.Datalog is a classical language that uses logic rules to derive facts from other facts within a database. We extend Datalog to handle temporal databases, so that its rules model how the facts can change over time. At any time, the database determines the set of possible next events, and when an event happens, it modifies some facts, causing the derived facts and possible next events to change. This is a kind of automaton that can be used to model either language (e.g., the events are words) or irregularly spaced event sequences such as transactions, clicks, posts, news events, or medical events. I'll explain how to turn it into an autoregressive probability model, whose neural architecture embeds the database facts and events by following the structure of the Datalog proofs. This is joint work with Chu-Cheng Lin, Hongyuan Mei, and other collaborators.

Jason Eisner is Professor of Computer Science at Johns Hopkins University, as well as Director of Research at Microsoft Semantic Machines. He is a Fellow of the Association for Computational Linguistics. At Johns Hopkins, he is also affiliated with the Center for Language and Speech Processing, the Mathematical Institute for Data Science, and the Cognitive Science Department. His goal is to develop the probabilistic modeling, inference, and learning techniques needed for a unified model of all kinds of linguistic structure. His 150+ papers have presented various algorithms for parsing, machine translation, and weighted finite-state machines; formalizations, algorithms, theorems, and empirical results in computational phonology; and unsupervised or semi-supervised learning methods for syntax, morphology, and word-sense disambiguation. He is also the lead designer of Dyna, a new declarative programming language that provides an infrastructure for AI research. He has received two school-wide awards for excellence in teaching, as well as recent Best Paper Awards at ACL 2017, EMNLP 2019, and NAACL 2021 and an Outstanding Paper Award at ACL 2022.

Oral Paper Sessions

Oral Paper Session 1: NLP Fundementals (Yangshuo room (阳朔厅), 2nd floor)
Session Chair: *Wenliang Chen*

13:30-13:45	Exploiting Word Semantics to Enrich Character Representations of Chinese Pre-trained Models *Wenbiao Li, Rui Sun and Yunfang Wu*
13:45-14:00	多模态与文本预训练模型的文本嵌入差异研究 *Yuchong Sun, Xiwei Cheng, Ruihua Song, Wanxiang Che, Zhiwu Lu and Jirong Wen*
14:00-14:15	PromptAttack: Prompt-based Attack for Language Models via Gradient Search *Yundi Shi, Piji Li, Changchun Yin, Zhaoyang Han, Lu Zhou and Zhe Liu*
14:15-14:30	MCER: A Multi-domain Dataset for Sentence-level Chinese Ellipsis Resolution *Jialu Qi, Yanqiu Shao, Wei Li and Zizhuo Shen*
14:30-14:45	Break
14:45-15:00	Two-layer Context-enhanced Representation for Better Chinese Discourse Parsing *Qiang Zhu, Kedong Wang and Fang Kong*
15:00-15:15	PGBERT: Phonology and Glyph Enhanced Pre-training for Chinese Spelling Correction *Lujia Bao, Xiaoshuai Chen, Junwen Ren, Yujia Liu and Chao Qi*
15:15-15:30	Online Self-boost Learning for Chinese Grammatical Error Correction *Jiaying Xie, Kai Dang and Jie Liu*

Oral Paper Session 2: NLP Applications (Beihai room (北海厅), 2nd floor)
Session Chair: *Xianling Mao*

13:30-13:45	Teaching Text Classification Models Some Common Sense via Q&A Statistics: A Light and Transplantable Approach *Hanqing Tao, Guanqi Zhu, Tong Xu, Qi Liu and Enhong Chen*
13:45-14:00	KGAT: An Enhanced Graph-based Model for Text Classification *Xin Wang, Chao Wang, Haiyang Yang, Xingpeng Zhang, Qi Shen, Kan Ji, Yuhong Wu and Huayi Zhan*
14:00-14:15	Multi-view document clustering with joint contrastive learning *Ruina Bai, Ruizhang Huang, Yongbin Qin and Yanping Chen*
14:15-14:30	Joint Optimization of Multi-vector Representation with Product Quantization *Yan Fang, Jingtao Zhan, Yiqun Liu, Jiaxin Mao, Min Zhang and Shaoping Ma*
14:30-14:45	Break
14:45-15:00	Generative Text Steganography via Multiple Social Network Channels Based on Transformers *Long Yu, Yuliang Lu, Xuehu Yan and Xianhui Wang*
15:00-15:15	Automatic Academic Paper Rating Based on Modularized Hierarchical Attention Network *Kai Kang, Huaping Zhang, Yugang Li, Xi Luo and Silamu Wushour*
15:15-15:30	A Joint Label-Enhanced Representation Based on Pre-trained Model for Charge Prediction *Jing Pei Dan, Xiaoshuang Liao, Lanlin Xu, Weixuan Hu and Tianyuan Zhang*

Oral Paper Session 3: NLP Fundamentals (Yangshuo room (阳朔厅), 2nd floor)
Session Chair: *Yue Zhang*

16:00-16:15	How Effective and Robust is Sentence-level Data Augmentation for Named Entity Recognition? *Runmin Jiang, Xin Zhang, Jiyue Jiang, Wei Li and Yuhao Wang*
16:15-16:30	Rethinking the Value of Gazetteer in Chinese Named Entity Recognition *Qianglong Chen, Xiangji Zeng, Jiangang Zhu, Yin Zhang, Bojia Lin, Yang Yang and Daxin Jiang*
16:30-16:45	Adversarial Transfer for Classical Chinese NER with Translation Word Segmentation *Yongjie Qi, Hongchao Ma, Lulu Shi, Hongying Zan and Qinglei Zhou*
16:45-17:00	BG-EFRL: Chinese Named Entity Recognition Method and Application Based on Enhanced Feature Representation *Xiankun Zhang and Siyuan Peng*
17:00-17:15	Break
17:15-17:30	MGCN: A Novel Multi-Graph Collaborative Network for Chinese NER *Yingqi Zhang, Wenjun Ma and Yuncheng Jiang*
17:30-17:45	Distill-AER: fine-grained Address Entity Recognition from Spoken Dialogue via Knowledge Distillation *Yitong Wang, Xue Han, Feng Zhou, Yiting Wang, Junlan Feng and Chao Deng*
17:45-18:00	Adversarial Transfer Learning for Named Entity Recognition Based on Multi-Head Attention Mechanism and Feature Fusion *Dandan Zhao, Pan Zhang, Jiana Meng and Yue Wu*

Oral Paper Session 4: Question Answering (Beihai room (北海厅), 2nd floor)
Session Chair: *Xiaojun Quan*

16:00-16:15	Coarse-to-fine Retriever for Better Open-domain Question Answering *Xue Liu and Fang Kong*
16:15-16:30	LoCSGN: Logic-Contrast Semantic Graph Network for Machine Reading Comprehension *Xi Zhao, Tingrui Zhang, Yuxiao Lu and Guiquan Liu*
16:30-16:45	Modeling Temporal-sensitive Information for Complex Question Answering over Knowledge Graphs *Yao Xiao, Guangyou Zhou and Jin Liu*
16:45-17:00	Knowledge-enhanced Iterative Instruction Generation and Reasoning for Knowledge Base Question Answering *Haowei Du, Quzhe Huang, Chen Zhang and Dongyan Zhao*
17:00-17:15	Break
17:15-17:30	基于多信息感知的多方对话问答方法 *Xiaoqian Gao, Xiabing Zhou and Min Zhang*
17:30-17:45	QuatSE: Spherical Linear Interpolation of Quaternion for Knowledge Graph Embeddings *Jiang Li, Xiangdong Su, Xinlan Ma and Guanglai Gao*
17:45-18:00	Entity Difference Modeling based Entity Linking for Question Answering over Knowledge Graphs *Meiling Wang, Min Li, Kewei Sun and Zhirong Hou*

Posters Session

Posters Session 1 (2nd floor public area)

P1-1	Emotion-Cause Pair Extraction via Transformer-based Interaction Model with Text Capsule Network *Cheng Yang and Jie Ding*
P1-2	Deep structure-aware Approach for QA over Incomplete Knowledge Bases *Qian Chen, Xiaoying Gao, Xin Guo and Suge Wang*
P1-3	FuDFEND: Fuzzy-domain for Multi-domain Fake News Detection *Chaoqi Liang, Yu Zhang, Xinyuan Li, Jinyu Zhang and Yongqi Yu*
P1-4	基于角色信息引导的多轮事件论元抽取 *Yuanfang Yu, Yong Zhang, Haoyang Zuo, Lianfa Zhang and Tingting Wang*
P1-5	Employing Internal and External Knowledge to Factuality-oriented Abstractive Summarization *Zhiguang Gao, Feng Jiang, Xiaomin Chu and Peifeng Li*
P1-6	Continuous Prompt Enhanced for Biomedical Entity Normalization *Zhaohong ZH L. Lai, Biao Fu, Shangfei Wei and Xiaodong Shi*
P1-7	Abstractive Summarization Model with Adaptive Sparsemax *Shiqi Guo, Yumeng Si and Jing Zhao*
P1-8	基于预训练模型和混合神经网络的医疗实体关系抽取 *Dandan Zhao, Junpeng Zhang, Jiana Meng, Zhihao Zhang and Wen Su*
P1-9	Bidirectional Multi-channel Semantic Interaction Model of Labels and Texts for Text Classification *Yuan Wang, Yubo Zhou, Peng Hu, Maoling Xu, Tingting Zhao and Yarui Chen*
P1-10	An On-Device Machine Reading Comprehension Model with Adaptive Fast Inference *Fulai Nan, Jin Wang and Xuejie Zhang*
P1-11	Exploiting Dynamic and Fine-grained Semantic Scope for Extreme Multi-label Text Classification *Yuan Wang, Huiling Song, Peng Huo, Tao Xu, Jucheng Yang, Yarui Chen and Tingting Zhao*
P1-12	依据《中图法》的英文图书分类探索——结合图情与一般领域的文本增强方法 *Yanting Jiang*
P1-13	KBRTE: A Deep Learning Model for Chinese Textual Entailment Recognition based on Synonym Expansion and Sememe Enhancement *Yalei Liu, Lingling Mu and Hongying Zan*
P1-14	融合语境语义差异特征的短文本匹配模型 *Wenhui Zhang, Meiling Wang and Zhirong Hou*

Posters Session 2 (2nd floor public area)

P2-1	知识驱动的事件双曲嵌入时序关系抽取方法研究 *Jianyong Duan, Shiwei Dai, Hao Wang, Li He and Xin Li*
P2-2	基于联合学习的少样本多类别情感分类方法 *Zicheng Li, Xiaoqin Chang, Yameng Li, Shoushan Li and Guodong Zhou*
P2-3	MGEDR: a Molecular Graph Encoder for Drug Recommendation *Kaiyuan Shi, Shaowu Zhang, Haifeng Liu, Yijia Zhang and Hongfei Lin*
P2-4	Hierarchical Planning of Topic-Comment Structure for Paper Abstract Writing *Mingyue Han, Ruifang He and Huijie Wang*
P2-5	Faster and Better Grammar-based Text-to-SQL Parsing via Clause-level Parallel Decoding and Alignment Loss *Kun Wu, Lijie Wang, Zhenghua Li and Xinyan Xiao*
P2-6	Two-stage Query Graph Selection for Knowledge Base Question Answering *Yonghui Jia, Chuanyuan Tan, Yuehe Chen, Muhua Zhu, Pingfu Chao and Wenliang Chen*
P2-7	Plug-and-Play Module for Commonsense Reasoning in Machine Reading Comprehension *Damai Dai, Hua Zheng, Zhifang Sui and Baobao Chang*
EW-3	Multi-task Hierarchical Cross-Attention Network for Multi-label Text Classification *Junyu Lu, Hao Zhang, Zhexu Shen, Kaiyuan Shi, Liang Yang, Bo Xu, Shaowu Zhang and Hongfei Lin*
EW-4	An Interactive Fusion Model for Hierarchical Multi-Label Text Classification *Xiuhao Zhao, Zhao Li, Xianming Zhang, Jibin Wang, Tong Chen, Zhengyu Ju, Canjun Wang, Chao Zhang and Yiming Zhan*
EW-7	CDAIL-BIAS MEASURER: A Model Ensemble Approach for Dialogue Social Bias Measurement *Jishun Zhao, Shucheng Zhu, Ying Liu and Pengyuan Liu*
EW-18	A Multi-Task Learning Model for Fine-Grain Dialogue Social Bias Measurement *Hanjie Mai , Xiaobing Zhou and Liqing Wang*

Student Workshop

Sep.24, Guilin room (桂林厅), 2nd floor
Chairs:
Zhongqing Wang, Soochow University
Piji Li, Nanjing University of Aeronautics and Astronautics

13:30-14:10	How to write a NLP paper?	Wayne Xin Zhao (Online)
14:10-14:40	How to Fail a Reseach Project?	Deng Cai (Online)
14:40-15:00	Manage Experiments in NLP Research	Ning Ding (Online)
15:00-15:30	Paper Presentation	

Student Workshop: Invitated Talk

Chair: Zhongqing Wang, Soochow University
Guilin room (桂林厅), 2nd floor

How to write a NLP paper?

Wayne Xin Zhao (Renmin University of China) (Online)
13:30-14:10

Abstract: Scientific writing is very important for gradudate students. However, the speaker does observe that many students don't know how to write a good research paper. In this talk, the speaker will sysmetically talk about the writing skills for NLP papers, and introduce the key points for writing each part of a research paper (e.g., introduction, model and experiments). He will also given some general guidelines or experiences to practice the writing skills for research purpose.

Wayne Xin Zhao is currently a professor at Gaoling School of Artificial Intelligence, Renmin University of China. He received the PhD degree from Peking University, in 2014. His research interests are recommender systems and natural language processing, and he has published more than 100 referred papers in international conferences and journals. His ECIR'11 paper "Comparing Twitter and Traditional Media Using Topic Models" has won "Test-of-time Award" at ECIR'21.

How to Fail a Reseach Project?

Deng Cai (Tencent AI Lab) (Online)
14:10-14:40

Abstract: "There are a thousand no's for every yes." Failure is common in our research life. Sadly, some research projects might not work out eventually. This talk will discuss this topic from a student's perspective. By analyzing the reasons for failure, we hope to speed up the next research project and have greater success.

Deng Cai recieved his PhD from The Chinese University of Hong Kong. His research interests include semantic parsing, machine translation, dialogue systems, and text generation. In particular, his research is concerned with improving the interpretability and generalization of deep learning models with symbolic semantics and reasoning, and external and explicit memory.He has published research papers at prestigious NLP/AI conferences and journals, such as ACL, EMNLP, NAACL, and AAAI. His work won an Outstanding Paper Award at ACL (2021).

Manage Experiments in NLP Research

Ning Ding (Tsinghua University) (Online)

14:40-15:00

Abstract: In the research of natural language processing and machine learning, experimentation is often one of the most important parts. Excellent and rigorous experimental design can accurately and comprehensively verify the effectiveness of the proposed method. This report will discuss the design experiments in NLP research, focusing on the following aspects. (1) Basic experimental guidelines: elaborate those rigorous experimental guidelines that are easily overlooked in machine learning but need to be followed; (2) Design of main experiments and auxiliary experiments: Explain how experiments should be designed in outstanding natural language processing research; (3) Experiment process: describe the methods and details of the experimental process such as fairness comparison, hyper-parameter tuning, and model selection; (4) Open source work: explain how to open source your own code to the natural language processing community in a standardized, clear, and elegant manner . (5) Tool construction: Explain the transition from experiment to tool construction, as well as the design concept and development process of NLP tools.

Ning Ding is a Ph.D student at Tsinghua Univeristy, studying natural language processing and machine learning. His research has been published at ICLR, ACL and EMNLP, etc. He is a recipient of ACL Best Demo Paper Award, Baidu Ph.D Fellowship and China National Scholarship.

Student Workshop: Paper Presentation

Chair: Piji Li, Nanjing University of Aeronautics and Astronautics
Guilin room (桂林厅), 2nd floor

15:00-15:10 Semi-Supervised Protein-Protein Interactions Extraction Method Based on Label Propagation and Sentence Embedding
Zhan Tang, Xuchao Guo, Lei Diao, Zhao Bai, Longhe Wang and Lin Li

15:10-15:20 Construction and Application of a Large-Scale Chinese Abstractness Lexicon Based on Word Similarity
Huidan Xu and Lijiao Yang

15:20-15:30 Stepwise Masking: A Masking Strategy Based on Stepwise Regression for Pre-Training
Jie Pan, Shuxia Ren, Dongzhang Rao, Zongxian Zhao and Wenshi Xue

Pretrained Models and Applications Workshop

Sep.24, Guilin room (桂林厅), 2nd floor
Chair: Xiaojun Wan

16:00-16:20	A Cross-lingual Knowledge-centric Pretrained Framework and Application	Ziyan Chen
16:20-16:40	Leveraging Multilingual Pretrained Models for Machine Translation	Wenxiang Jiao
16:40-17:00	Overview of Recent Progress on Foundation Models at Huawei Noah's Ark Lab	Fei Mi
17:00-17:10	Break	
17:10-17:30	ERNIE-ViLG 2.0: a Vision-Language Generation Model in Baidu Wenxin	Shikun Feng
17:30-17:50	Leveraging Causal Inference for Legal Text Analysis	Yansong Feng
17:50-18:00	Closing remarks	

Pretrained Models and Applications Workshop

Sep. 24, Chair: *Xiaojun Wan*
16:00-16:20, Guilin room (桂林厅), 2nd floor

A Cross-lingual Knowledge-centric Pretrained Framework and Application

Ziyan Chen
GTCOM

With the globalization of open source data, multilingual nlp technology and cross-lingual alignment have become the core issues to be solved. In recent years, the rapid development of multilingual pre-trained models has brought profound changes to the global linguistic intelligence technology. GTCOM has built a multilingual pre-trained model, based on its large-scale cross-lingual industry knowledge graph and a global multilingual big data system. Meanwhile, based on its big models, GTCOM has released a natural language processing algorithm library covering 61 kinds and more than 60 languages, enabling vertical industry applications such as military, finance, and technology.

Ziyan Chen currently the director of NLP field of GTCOM 2030 Artificial Intelligence Research Institute, graduated from the Institute of Electronics, Chinese Academy of Sciences with a Ph.D. His main research interests include information extraction, natural language generation, and cross-lingual knowledge graph construction. He has published more than 10 papers in international academic journals and conferences, and participated in a number of major national projects, such as pre-research projects and National Natural Science Foundation.

Pretrained Models and Applications Workshop

Sep. 24, Chair: *Xiaojun Wan*
16:20-16:40, Guilin room (桂林厅), 2nd floor

Leveraging Multilingual Pretrained Models for Machine Translation

Wenxiang Jiao
Tencent AI Lab

Multilingual pretraining is an effective approach to boost the performance of NLP tasks across languages by learning representations from large-scale unlabeled multilingual corpora. It is well suited for machine translation (MT) tasks, which usually involve two or more languages. While previous multilingual pretraining for MT generally focuses only on Transformer encoder, recent studies like mBART pretrain a complete autoregressive sequence-to-sequence (Seq2Seq) model, which remedy the architecture gap between pretraining and finetuning.In this talk, we present a substantial step in better understanding such multilingual Seq2Seq pretrained models through three questions: (1) How much does the jointly pretrained decoder matter? (2) How do the discrepancies between pretraining and finetuning affect the downstream performance? (3) How does multilingual Seq2Seq pretraining perform for multilingual MT? We find that multilingual Seq2Seq pretraining is a double-edged sword: On one hand, it helps MT models to produce more diverse translations and reduce adequacy-related translation errors. On the other hand, the discrepancies between multilingual Seq2Seq pretraining and MT finetuning limit the translation quality (i.e., domain discrepancy) and induce the over-estimation issue (i.e., objective discrepancy). As for multilingual MT, multilingual Seq2Seq pretrained models consistently improve the performance of supervised translation but harms that of zero-shot translation by introducing more off-target issues. Based on these findings, we propose simple yet effective approaches to better leverage the multilingual Seq2Seq pretrained models and achieve significant improvements on various translation tasks.

Wenxiang Jiao is now a senior researcher at Tencent AI Lab. He received his Ph.D degree from the Chinese University of Hong Kong in 2021, under the supervision by Prof. Irwin King and Prof. Michael R. Lyu. Before that, he received his Bachelor degree and Mphil degree at Nanjing University in 2015 and 2017, respectively. Wenxiang is interested in research directions like conversational emotion recognition, machine translation, and multilingual pretraining, and has published papers in top conferences and journals such as ACL, EMNLP, NAACL, AAAI, TASLP, etc.

Pretrained Models and Applications Workshop

Sep. 24, Chair: *Xiaojun Wan*
16:40-17:00, Guilin room (桂林厅), 2nd floor

Overview of Recent Progress on Foundation Models at Huawei Noah's Ark Lab

Fei Mi
Huawei Noah's Ark Lab

Foundation Models (a.k.a., Pre-trained Language Models, Large Models) is a new paradigm of enabling NLP technologies to prosper different NLP applications. Foundation Models are equipped with increasingly large model parameters and are trained on broad data at scale through self-supervision to cope with a wide range of downstream tasks. A variety of foundation models have been recognized as major transformation in how powerful AI systems can be built in different scenarios, such as language understanding (BERT), language generation (GPT), dialogue applications (LaMDA), multi-modality (DALL-E, Flamingo, Florence), decision making (GATO), and etc. In this talk, the speaker is going to briefly introduce recent leading researches w.r.t. the topic of foundation models at Huawei Noah's Ark Lab, including (1) large-scale (200 billion) language model pre-training fully based on Huawei technology stack (PanGu-Alpha); (2) model compression and acceleration (TinyBert, QuantGPT, ⋯); (3) open-domain dialogue model pre-training (PanGu-Bot); (4) Code intelligence pre-training (PanGu-Coder); (5) multi-modality pre-training (SPIRAL, FILIP). Besides foundation models, the speaker will also briefly overview some other research interests and roadmaps at Huawei Noah's Ark Lab –Speech & NLP Lab.

Fei Mi is a research scientist at Huawei Noah's Ark Lab specialized in foundation NLP models and dialog systems. He obtained his Ph.D. degree in Computer Science from The Swiss Federal Institute of Technology Lausanne (EPFL) in 2021, supervised by Prof. Boi Faltings. Prior to that, he obtained his MPhill degree from Hong Kong University of Science and Technology (HKUST) under the supervision of Prof. Dit-Yan Yeung and Bachelor degree from a joint program between Sun-Yat Sun University (SYSU) and HKUST. His research interests mainly lies on Dialog Systems, Foundation Models, Few-shot Learning, Domain Adaptation, Recommendation, and he has published more than 10 papers at top AI conferences such as ACL, EMNLP, NAACK, AAAI, IJCAI.

Pretrained Models and Applications Workshop

Sep. 24, Chair: *Xiaojun Wan*
17:10-17:30, Guilin room (桂林厅), 2nd floor

ERNIE-ViLG 2.0: a Vision-Language Generation Model in Baidu Wenxin

Shikun Feng
Baidu

Recently the field of text-to-image synthesis has attracted more and more attention, and its goal is to generate art or realistic images given an input prompt (text). The techniques behind this field have been evolving very quickly lately, from GAN to Seq2Seq models, and then to the Diffusion Model. In this talk, I will first give a brief introduction to the text-to-image synthesis techniques, and then share the two generations of the image-text synthesis models in Baidu Wenxin, which are ERNIE-ViLG 1.0, the Unified Generative Pre-training for Bidirectional Vision-Language Generation, and ERNIE-ViLG 2.0, a new state-of-the-art text-to-image model that generates images from Chinese text. In addition, I will share serveral useful skills (Prompt Books) for playing with ERNIE-ViLG 2.0 and show some interesting generated cases. The audiences are welcome to click this link (https://wenxin.baidu.com/moduleApi/ernieVilg) to try out the ERNIE-ViLG 2.0 API service.

Shikun Feng, currently the principal architect of natural language processing department at Baidu Inc., graduated from the Institute of Automation, Chinese Academy of Sciences. He is responsible for the semantic representation, graph learning, intelligent document understanding, and other directions. His research and development results are widely used in search engines, information flow, smart speakers, maps, and other products in Baidu, and significantly improve the user experience of hundreds of millions of netizens. He has won more than ten AI competition world championships, including KDD CUP, GLUE, SuperGLUE, SemEval, DocVQA, and so on. He has published several high-level papers in top international conferences on artificial intelligence, such as CVPR, AAAI, IJCAI, KDD, ACM MM, and CIKM, and one of the papers was rated as one of the most influential academic papers in AAAI 2020 by Paper Digest. He has more than 40 domestic and foreign technology patents and won the China Excellent Patent Award. He won the highest award of the World Artificial Intelligence Conference SAIL Award, the Outstanding Science and Technology Achievement Award of the Chinese Association for Artificial Intelligence, and two times Baidu Highest Award.

Pretrained Models and Applications Workshop

Sep. 24, Chair: *Xiaojun Wan*
17:30-17:50, Guilin room (桂林厅), 2nd floor

Leveraging Causal Inference for Legal Text Analysis

Yansong Feng
Peking University

Causal inference is the process of capturing cause-effect relationship among factors, which are often carefully prepared by domain experts. However, mining causal relationship among factors from unstructured data, like plain text, has been less examined. In this talk, I will present our recent work on enabling causal inference over plain texts without much human involvement. We design a novel Graph-based Causal Inference (GCI) framework to build causal graphs from fact descriptions to facilitate legal practitioners to make proper decisions. In our experiments, we find that GCI can capture the nuance from fact descriptions among multiple confusing charges and provide explainable discrimination, especially in few-shot settings.

Dr. Yansong Feng is an associate professor in the Wangxuan Institute of Computer Technology at Peking University. Before that, he obtained his PhD from ICCS (now ILCC) at the University of Edinburgh. His current research interests include using probabilistic methods to distill knowledge from large volumes of natural language texts, and supporting intelligent human-computer interfaces, such as question answering. He has served as Action Editor and Area Chair for ARR and *ACL conferences. Yansong received the IBM Faculty Award in 2014 and 2015, and the IBM Global Shared University Research Award in 2016.

Evaluation Workshop

Evaluation Workshop Session 1: Question Answering with Knowledge Models & Speech Entity Linking (Nanning room 2 (南宁厅 2), 2nd floor)

13:30-13:40	A Pre-trained Language Model for Medical Question Answering based on Domain Adaption *Lang Liu, Junxiang Ren, Yuejiao Wu, Ruilin Song, Zhen Cheng and Sibo Wang*
13:40-13:50	Overview of the NLPCC2022 Shared Task on Speech Entity Linking *Ruoyu Song, Sijia Zhang, Xiaoyu Tian and Yuhang Guo*
13:50-14:00	DAMO-NLP at NLPCC-2022 Task 2: Knowledge Enhanced Robust NER for Speech Entity Linking *Shen Huang, Yuchen Zhai, Xinwei Long, Yong Jiang, Xiaobin Wang, Yin Zhang and Pengjun Xie*
14:00-14:10	Enhancing Entity Linking with Contextualized Entity Embeddings *Zhenran Xu, YuLin Chen, Senbao Shi and Baotian Hu*
14:10-14:20	Break

Evaluation Workshop Session 2: Multimodal Product Summarization (Nanning room 2 (南宁厅 2), 2nd floor)

14:20-14:30	Overview of the NLPCC 2022 Shared Task on Multimodal Product Summarization *Haoran Li, Peng Yuan, Haoning Zhang, Weikang Li, Song Xu, Youzheng Wu and Xiaodong He*
14:30-14:40	Knowledge Enhanced Pre-trained Language Model for Product Summarization *Wenbo Yin, Junxiang Ren, Yuejiao Wu, Ruilin Song, Lang Liu, Zhen Cheng and Sibo Wang*
14:40-14:50	Break

Evaluation Workshop Session 3: Multimodal Dialogue Understanding and Generation (Nanning room 2 (南宁厅 2), 2nd floor)

14:50-15:00	Overview of the NLPCC 2022 Shared Task: Multi-modal Dialogue Understanding and Generation *Yuxuan Wang, Xueliang Zhao and Dongyan Zhao*
15:00-15:10	Scene-Aware Prompt for Multi-modal Dialogue Understanding and Generation *Bin Li, Yixuan Weng, Ziyu Ma, Bin Sun and Shutao Li*
15:10-16:00	Break

Evaluation Workshop Session 4: Multi-label Classification, NER, Content Extraction for Scientific Literature (Nanning room 2 (南宁厅 2), 2nd floor)

16:00-16:10	Overview of NLPCC2022 Shared Task 5 Track 1: Multi-label Classification for Scientific Literature *Ming Liu, He Zhang, Yangjie Tian, Tianrui Zong, Borui Cai, Ruohua Xu and Yunfeng Li*
16:10-16:20	NLPCC-2022 Task5 Track1: Hierarchical Multi-label Classification via Label-aware Graph Convolutional Network *Bo Wang, Yifan Lu, Xiaochi Wei, Xiao Liu, Ge Shi, Changsen Yuan, Heyan Huang, Chong Feng and Xianling Mao*
16:20-16:30	Overview of NLPCC2022 Shared Task 5 Track 2: Named Entity Recognition *Borui Cai, He Zhang, Fenghong Liu, Ming Liu, Tianrui Zong, Zhe Chen and Yunfeng Li*
16:30-16:40	Context Enhanced and Data Augmented W2NER System for Named Entity Recognition *Chunping Ma, Zijun Xu, Minwei Feng, Jingcheng Yin, Liang Ruan and Hejian Su*
16:40-16:50	Break

Evaluation Workshop Session 5: Dialogue Text Analysis, Topic Extraction and Dialogue Summary (Nanning room 2 (南宁厅 2), 2nd floor)

16:50-17:00	Overview of the NLPCC 2022 Shared Task on Multimodal Product Summarization *Qingliang Miao, Tao Guan, Yifan Yang, Yifan Zhang, Hua Xu and Fujiang Ge*
17:00-17:10	Dialogue Topic Extraction as Sentence Sequence Labeling *Dinghao Pan, Zhihao Yang, Haixin Tan, Jiangming Wu and Hongfei Lin*
17:10-17:20	Augmented Topic-Specific Summarization for Domain Dialogue Text *Zhiqiang Rao, Daimeng Wei, Zongyao Li, Hengchao Shang, Jinlong Yang, Zhengzhe Yu, Shaojun Li, Zhanglin Wu, Lizhi Lei, Hao Yang and Ying Qin*
17:20-17:30	Break

Evaluation Workshop Session 6: Fine-Grain Dialogue Social Bias Measurement (Nanning room 2 (南宁厅 2), 2nd floor)

17:30-17:40	Overview of NLPCC 2022 Shared Task 7: Fine-Grain Dialogue Social Bias Measurement *Jingyan Zhou, Fei Mi, Helen Meng and Jiawen Deng*
17:40-17:50	A Fine-Grained Social Bias Measurement Framework for open-domain dialogue systems *Aimin Yang, Qifeng Bai, Jigang Wang, Nankai Lin, Xiaotian Lin, Guanqiu Qin and Junheng He*

6

Main Conference & Workshop : Sep, 25

	Session		Venue
09:00-10:00	Keynote Talk 3, by Luke Zettlemoyer (Online)		Ballroom 1&2, 1st floor
10:00-10:30	Break		
10:30-11:30	Keynote Talk 4, by Alexander Rush (Online)		Ballroom 1&2, 1st floor
11:30-13:30	Lunch		Li Cafe (Buffet), 1st floor
13:30-15:30	Oral Paper Sessions 5-7 & Workshop	Chair	Venue
	Oral 5: Sentiment Analysis	Liang Yang	Yangshuo room (阳朔厅)
	Oral 6: Information Extraction	Buzhou Tang	Beihai room (北海厅)
	Oral 7: Machine Learning	Zhaochun Ren	Guilin room (桂林厅)
	Low-Resource NLP Workshop 1	Deyi Xiong Degen Huang	Nanning room 2 (南宁厅 2)
15:30-15:45	Break		
15:45-17:00	Oral Paper Sessions 8-10 & Workshop	Chair	Venue
	Oral 8: Machine Translation	Xiangyu Duan	Yangshuo room (阳朔厅)
	Oral 9: Diaglog	Wei Wei	Beihai room (北海厅)
	Oral 10: Generation	Kehai Chen	Guilin room (桂林厅)
	Low-Resource NLP Workshop 2	Guogen Cheng Zhengshan Xue	Nanning room 2 (南宁厅 2)
17:15-17:45	NLPCC Awards & Closing Ceremony		Ballroom 1&2, 1st floor
18:00-21:00	Dinner		Li Cafe (Buffet), 1st floor

Keynote Talk 3 (Online)

Sep. 25 , Chair: *Wei Lu*
09:00-10:00, Ballroom 1&2, 1st floor

Large Language Models: Will they keep getting bigger? And, how will we use them if they do?

Luke Zettlemoyer

He trend of building ever larger language models has dominated much research in NLP over the last few years. In this talk, I will discuss our recent efforts to (at least partially) answer two key questions in this area: Will we be able to keep scaling? And, how will we actually use the models, if we do? I will cover our recent efforts on learning new types of sparse mixtures of experts (MoEs) models. Unlike model-parallel algorithms for learning dense models, which are very difficult to further scale with existing hardware, our sparse approaches have significantly reduced cross-node communication costs and could possibly provide the next big leap in performance, although finding a version that scales well in practice remains an open challenge. I will also present our recent work on prompting language models that better controls for surface form variation, to improve performance of models that are so big we can only afford to do inference, with little to no task-specific fine tuning. Finally, time permitting, I will discuss work on new forms of supervision for language model training, including learning from the hypertext and multi-modal structure of web pages to provide new signals for both learning and prompting the model. Together, these methods present our best guesses for how to keep the scaling trend alive as we move forward to the next generation of NLP models.

Luke Zettlemoyer is a Professor in the Paul G. Allen School of Computer Science & Engineering at the University of Washington, and a Research Scientist at Meta. His research focuses on empirical methods for natural language semantics, and involves designing machine learning algorithms, introducing new tasks and datasets, and, most recently, studying how to best develop self-supervision signals for pre-training. His honors include being named an ACL Fellow as well as winning a PECASE award, an Allen Distinguished Investigator award, and multiple best paper awards. Luke received his PhD from MIT and was a post-doc at the University of Edinburgh.

Keynote Talk 4 (Online)

Sep. 25, Chair: *Shujian Huang*
10:30-11:30, Ballroom 1&2, 1st floor

Prompting, Metadatasets, and Zero-Shot NLP

Alexander Rush

The paradigm of NLP tasks is changing, expanding from mostly single-dataset supervised learning in structured form to multi-dataset semi-supervised learning expressed in natural language. This talk focuses on T0, a large-scale language model trained on multitask prompted data (Sanh et al 2022). Despite being an order of magnitude smaller than GPT-3 class models, T0 exhibits similar zero-shot accuracy on unseen task categories. In addition to the modeling elements, this talk highlights the community processes of collecting data, dataset, and prompts for models of this scale. The work was done as part of BigScience, an international, collaborative effort to study large language models.

Alexander Rush is a Professor at Cornell Tech and researcher at Hugging Face. His work is at the intersection of natural language processing and probabilistic deep learning with applications in text generation and efficient inference. He has written several popular open-source software projects supporting NLP research and data science, as well as pedagogical implementations of popular libraries. He is the secretary of ICLR and developed the MiniConf software used to run ML/NLP virtual conferences during COVID. His work has received paper and demo awards at major NLP, visualization, and hardware conferences, an NSF Career Award, and a Sloan Fellowship.

Low-Resource NLP Workshop

Sep.25, Nanning room 2 (南宁厅 2), 2nd floor
Organizers: Deyi Xiong, Degen Huang, Guogen Cheng, Zhengshan Xue

13:30-14:10	Solving Low-resource & Multilingual NLP Problems with Data Augmentation and Regularization	Lidong Bing (Online)
14:10-14:50	Language • Intelligence • Future	Tsering Gyal
14:50-15:00	Break	
15:00-15:40	Unified Schema Prompt for NLP Task Generalization	Nan Duan
15:40-16:20	Understanding and Improving Cross-Lingual Transfer in Multilingual Translation	Biao Zhang (Online)
16:20-16:30	Break	
16:30-17:00	Panel: Challenges and Opportunities in NLP4LRL	Deyi Xiong, Tsering Gyal, Degen Huang, Nan Duan, Biao Zhang
17:00-17:05	Closing remarks	

Low-Resource NLP Workshop

Sep. 25, Organizers: *Deyi Xiong, Degen Huang, Guogen Cheng, Zhengshan Xue*
13:30-14:10, Nanning room 2 (南宁厅 2), 2nd floor

Solving Low-resource & Multilingual NLP Problems with Data Augmentation and Regularization

Lidong Bing (Online)

A unique aspect of natural language is that there are many different languages used in different parts of the world. Linguistic studies have revealed that there are more than 6,500 human languages in the world. In addition, new domains are always encountered in real applications. Therefore, low-resource & multilingual NLP problems are everywhere. In this talk, the presenter will discuss two major techniques for solving these problems, namely, Data Augmentation (DA) and Regularization. Broadly speaking, all solvers for low-resource problems are of DA-based. DA for fine-grained tasks such as NER is difficult because a small context change will mislead the model training. The presenter will introduce several methods for fine-grained DA, including pseudo-training data generation, translate train and code-switch methods, etc. The presenter will also introduce methods that regularize the model behavior of the main task with limited training data, such as auxiliary task-based methods, warm-up methods, and the consistency training method. Finally, the presenter will introduce some works on unsupervised representation learning which is an important backbone for many downstream tasks.

Lidong Bing Ph.D. from the Chinese University of Hong Kong, and post-doctoral fellow at Carnegie Mellon University. He is currently working in the Language Technology Laboratory of Alibaba DAMO Academy and is the head of the multilingual NLP team. Dr. Bing has been engaged in research and development in NLP for more than 10 years. His current research interests focus on low-resource NLP, sentiment analysis, text generation, etc. In recent years, he has published more than 100 papers in top artificial intelligence conferences and journals. He has participated in the organization and review of top NLP and machine learning journals and conferences as an associate editor, area chair, and senior program committee member.

Low-Resource NLP Workshop

Sep. 25, Organizers: *Deyi Xiong, Degen Huang, Guogen Cheng, Zhengshan Xue*
14:10-14:50, Nanning room 2 (南宁厅 2), 2nd floor

Language ·Intelligence ·Future

Tsering Gyal

This report focuses on the function, value and transmission of language. It also introduces the relationship between natural language processing (in the case of Tibetan natural language processing) and artificial intelligence and natural language understanding (in the case of Tibetan natural language understanding) and artificial intelligence. Some superficial thoughts on machine intelligence and human intelligence and the problem of language gaming and language disappearance faced by human beings are presented.

Tsering Gyal Ph.D. supervisor, is the deputy director of the State Key Laboratory of Tibetan Language Intelligence Information Processing and Application, the deputy director of the Key Laboratory of Tibetan Language Information Processing of the Ministry of Education, the director of Qinghai Provincial Tibetan Information Processing Engineering Technology Research Center, the governmental counselor of Qinghai Province, the editorial board member of the Journal of Chinese Information.

Low-Resource NLP Workshop

Sep. 25, Organizers: *Deyi Xiong, Degen Huang, Guogen Cheng, Zhengshan Xue*
15:00-15:40, Nanning room 2 (南宁厅 2), 2nd floor

Unified Schema Prompt for NLP Task Generalization

Nan Duan

Recent research attempts to improve the task generalization ability of pre-trained language models using human-readable prompts. However, these approaches require laborious and inflexible manual collection of prompts. We propose Unified Schema Prompt to automatically customize the learnable prompts for each task according to the task input schema. It models the shared knowledge between tasks, while keeping the characteristics of different task schema, and thus enhances task generalization ability. We conduct schema prompt-based multitask pre-training on a wide variety of general NLP tasks. The resulting model achieves strong zero-shot and few-shot generalization performance on 16 unseen downstream tasks from 8 task types.

Nan Duan is a senior principal researcher & research manager at Microsoft Research Asia. He is an adjunct Ph.D. supervisor at University of Science and Technology of China and an adjunct professor at Tianjin University. He is a Distinguished Member of CCF. His research interests include natural language processing, code intelligence, multimodal intelligence, and machine reasoning.

Low-Resource NLP Workshop

Sep. 25, Organizers: *Deyi Xiong, Degen Huang, Guogen Cheng, Zhengshan Xue*
15:40-16:20, Nanning room 2 (南宁厅 2), 2nd floor

Understanding and Improving Cross-Lingual Transfer in Multilingual Translation

Biao Zhang (Online)

Neural machine translation has achieved impressive performance on many language pairs, but these successes often assume the availability of large-scale training data. Unfortunately, most language pairs in our world are short of parallel resources, regardless of sentences or documents. What if we have limited and even no parallel corpora? In this talk, I will present our efforts on (massively) multilingual neural machine translation (MNMT) with a particular focus on cross-lingual transfer for low-resource languages. I will first show the challenges when scaling the number of languages up in MNMT and discuss the off-target problem in zero-shot translation. Then, I will further explore document-level translation and present whether and how contextual modeling capability is achievable for languages without document data via cross-lingual transfer.

Biao Zhang is a postdoc at the University of Edinburgh. He got his Ph.D. at the ILCC, University of Edinburgh, advised by Prof. Rico Sennrich and Prof. Ivan Titov. His research focuses on improving neural machine translation (NMT), particularly its efficiency and universality, including developing lightweight (fast and effective) architectures for NMT, low-resource NMT, massively multilingual NMT, speech-to-text translation, context-aware NMT, and their intersections. He has published several papers at top-tier NLP/ML conferences, such as ACL, EMNLP, NeurIPS, ICLR, ICML, AAAI, and IJCAI.

Low-Resource NLP Workshop

Sep. 25, Organizers: *Deyi Xiong, Degen Huang, Guogen Cheng, Zhengshan Xue*
16:30-17:00, Nanning room 2 (南宁厅 2), 2nd floor

Panel: Challenges and Opportunities in NLP4LRL

Deyi Xiong, Tsering Gyal, Degen Huang, Nan Duan, Biao Zhang

In the past several decades, research works presented in the prestigious natural language processing (NLP) conferences (e.g., ACL, COLING) have been usually devoted to a very few high-resource languages while hundreds of languages have been left behind. With the fast development of (multilingual) large language models and the availability of public data, recent years have witnessed a growing interest in natural language processing for low-resource languages. The goals of this workshop are to identify key challenges in natural language processing for low-resource languages, to discuss how resources and tools can be built to facilitate NLP for low-resource languages, and to find new opportunities emerging from massive multilingual language models for low-resource language processing. Our key interest is to provide insights into resources, tools and methodologies for NLP4LRL, especially for ethnic minority language processing. Specifically, the motivations of this workshop are: To bring researchers from both academia and industry together and to cultivate new ideas for resources, tools, cutting-edge models and algorithms for low-resource language processing. To explore new research horizons for low-resource language processing and its applications.

Deyi Xiong is a Professor of Computer Science at Tianjin University (TJU), Director of both the Natural Language Processing Laboratory at the College of Intelligence and Computing, TJU and the International Joint Research Center of Language Intelligence and Technology at TJU. Prior to joining TJU, he was a professor at Soochow University (2013-2018) and a research scientist at the Institute for Infocomm Research, Singapore (2007-2013). His research focuses on natural language processing, specifically machine translation, dialogue, natural language generation and commonsense reasoning. He has published over 100 papers in prestigious journals and conferences, including Computational Linguistics, IEEE TPAMI, IEEE TASLP, Artificial Intelligence, AAAI, IJCAI ACL, and EMNLP. He was the program co-chair of IALP 2021 and CWMT 2017. He has also served as an area chair of conferences including ACL, EMNLP, NAACL and COLING. He was the founder and co-organizer of multiple ACL/EMNLP/NAACL-affiliated workshops such as S2MT 2015, SedMT 2016 and DiscoMT 2019. He is a member of the standing committee of reviewers of CL, action editor of both TACL and ARR, and an editorial board member of International Journal of Asian Language Processing. He has been active in developing a variety of resources for natural language processing. He has recently led teams to create BiPaR that is a bilingual parallel novel-style machine reading comprehension dataset developed to support multilingual and cross-lingual reading comprehension, RiSAWOZ that is to date the largest fully annotated human-to-human task-oriented dialogue dataset, TED-CDB that is the largest PDTB-style Chinese discourse corpus built on a large set of TED talks, Chinese WPLC that is a Chinese dataset for evaluating pretrained language models on word prediction given long-range context.

Degen Huang is full professor of Nature Language Processing in the School of Computer Science and Technology at Dalian University of Technology. He holds MSc degrees in Computer Science and Engineering(1988，Dalian University of Technology), and a PhD in Computer Software and Theory(2003， Dalian University of Technology). He joined Dalian University of Technology in 1988, and was appointed full professor in 2006. He is a senior member of ACM, CAAI, ACL, CCF. He leads the Nature Language Processing and Machine Translation group. His research focus is on Machine Translation, with projects on machine translation in Cross-Language Information Retrieval (CLIR), on Multi-language Neural Machine Translation (MNMT). His research interests include developing nature language processing models and machine translation algorithms for solving Chinese-Japanese translation. He has published over 200 papers appeared at venues such as ACL, ACM, IEEE, COLING, EMNLP and a book about English shallow parsing for machine translation.

Guogen Cheng is the Board Member, Vice President and Chief Technology Officer of GTCOM Technology Co., Ltd., Council Member of Chinese Information Society of China, Standing Committee Member of Multilingual Intelligent Information Processing Technology Committee of China Artificial Intelligence Society. He graduated from China University of Geosciences (Wuhan) majoring in computer science in 2000. He served as the technical director of Peking University Founder with a senior technical title (information system project manager, system architect), and one of the top ten digital talents of China Publishing Group. In the past 8 years, he has undertaken more than 10 national-level scientific research and industrial development projects from the Central Propaganda Department, the Ministry of Science and Technology, and the Ministry of Industry and Information Technology, and obtained more than 10 invention patents. He is one of principle investigators in the 2020 Science and Technology Innovation 2030 New Generation Artificial Intelligence Major Project - "Research on Multilingual Automatic Translation with Chinese as the Core", the project leader of the new generation artificial intelligence industry innovation task project of the Ministry of Industry and Information Technology.

Zhengshan Xue has been devoted to the research on machine translation & natural language processing and their industrial applications. He is leading the machine translation team in OPPO. His team has developed a variety of machine translation products embedded in OPPO platforms, benefiting over ten million users. He has published more than 10 papers, and has won the first place in multiple domestic and international machine translation evaluation campaigns (e.g., WMT, IWSLT, CCMT).

Oral Paper Sessions

Oral Paper Session 5: Sentiment Analysis (Yangshuo room (阳朔厅), 2nd floor)
Session Chair: *Liang Yang*

13:30-13:45	A Multibias-mitigated and Sentiment Knowledge Enriched Transformer for Debiasing in Multimodal Conversational Emotion Recognition *Jinglin Wang, Fang Ma, Yazhou Zhang and Dawei Song*
13:45-14:00	Aspect-specific Context Modeling for Aspect-based Sentiment Analysis *Fang Ma, Chen Zhang, Bo Zhang and Dawei Song*
14:00-14:15	Memeplate: A Chinese Multimodal Dataset for Humor Understanding in Meme Templates *Zefeng Li, Hongfei Lin, Liang Yang, Bo Xu and Shaowu Zhang*
14:15-14:30	FuncSA: Function Words-guided Sentiment-aware Attention for Chinese Sentiment Analysis *Jiajia Wang, Hongying Zan, Yingjie Han and Juan Cao*
14:30-14:45	Break
14:45-15:00	Prompt-Based Generative Multi-label Emotion Prediction with Label Contrastive Learning *Yuyang Chai, Chong Teng, Hao Fei, Shengqiong Wu, Jingye Li, Ming Cheng, Donghong Ji and Fei Li*
15:00-15:15	Learning Emotion-Aware Contextual Representations for Emotion-Cause Pair Extraction *Baopu Qiu and Lin Shang*
15:15-15:30	A Multi-step Attention and Multi-level Structure Network for Multimodal Sentiment Analysis *Chuanlei Zhang, Hongwei Zhao, Bo Wang, Wei Wang, Ting Ke and Jianrong Li*
15:30-15:45	Unimodal and Multimodal Integrated Representation Learning via Improved Information bottleneck for Multimodal Sentiment Analysis *Tonghui Zhang, Changfei Dong, Jinsong Su, Haiying Zhang and Yvzheng Li*

Oral Paper Session 6: Information Extraction (Beihai room (北海厅), 2nd floor)
Session Chair: *Buzhou Tang*

13:30-13:45	Temporal Relation Extraction on Time Anchoring and Negative Denoising *Liang Wang, Peifeng Li and Qiaoming Zhu*
13:45-14:00	TEMPLATE: TempRel Classification Model Trained with Embedded Temporal Relation Knowledge *Tiesen Sun and Lishuang Li*
14:00-14:15	Dual Interactive Attention Network for Joint Entity and Relation Extraction *Lishuang Li, Zehao Wang, Xueyang Qin and Hongbin Lu*
14:15-14:30	BART-Reader: Predicting Relations between Entities via Reading Their Document-Level Context Information *Hang Yan, Yu Sun, Junqi Dai, Xiangkun Hu, Qipeng Guo, Xipeng Qiu and Xuanjing Huang*
14:30-14:45	Break
14:45-15:00	DuEE-Fin: A Large-scale Dataset for Document-level Event Extraction *Cuiyun Han, Jinchuan Zhang, Xinyu Li, Guojin Xu, Weihua Peng and Zengfeng Zeng*
15:00-15:15	Label Semantic Extension for Chinese Event Extraction *Zuohua Chen, Guohua Wu, Qiuhua Wang, Zhen Zhang, Qisen Xi, Yizhi Ren and Lifeng Yuan*
15:15-15:30	ArgumentPrompt: Activating Multi-category of Information for Event Argument Extraction with Automatically Generated Prompts *Shenpo Dong, Wei Yu, Hongkui Tu, Xiaodong Wang, Yunyan Zhou, Haili Li, Jie Zhou and Tao Chang*

Oral Paper Session 7: Machine Learning (Guilin room (桂林厅), 2nd floor)
Session Chair: *Zhaochun Ren*

13:30-13:45	Multi-task Learning with Auxiliary Cross-attention Transformer for Low-resource Multi-dialect Speech Recognition *Zhengjia Dan, Yue Zhao, Xiaojun Bi, Licheng Wu and Qiang Ji*
13:45-14:00	Regularized Contrastive Learning of Semantic Search *Mingxi Tan, Alexis Rolland and Andong Tian*
14:00-14:15	Kformer: Knowledge Injection in Transformer Feed-Forward Layers *Yunzhi Yao, Shaohan Huang, Li Dong, Furu Wei, Huajun Chen and Ningyu Zhang*
14:15-14:30	Doge Tickets: Uncovering Domain-general Language Models by Playing Lottery Tickets *Yi Yang, Chen Zhang, Benyou Wang and Dawei Song*

Oral Paper Session 8: Machine Translation (Yangshuo room (阳朔厅), 2nd floor)
Session Chair: *Xiangyu Duan*

15:45-16:00	Random Concatenation: A Simple Data Augmentation Method for Neural Machine Translation *Nini Xiao, Huaao Zhang, Chang Jin and Xiangyu Duan*
16:00-16:15	基于单词领域特征敏感的多领域神经机器翻译 *Zengcheng Huang, Zhibo Man, Yujie Zhang, Jinan Xu and Yufeng Chen*
16:15-16:30	篇章约束的译文质量评估模型 *Qin Feng, Zhengxian Gong, Guodong Zhou and Heng Ye*
16:30-16:45	Contrastive Learning for Robust Neural Machine Translation with ASR Errors *Dongyang Hu and Junhui Li*
16:45-17:00	An Enhanced New Word Identification Approach Using Bilingual Alignment *Ziyan Yang, Huaping Zhang, Jianyun Shang and Silamu Wushour*

Oral Paper Session 9: Diaglog (Beihai room (北海厅), 2nd floor)
Session Chair: *Wei Wei*

15:45-16:00	MedDG: An Entity-Centric Medical Consultation Dataset for Entity-Aware Medical Dialogue Generation *Wenge Liu, Yi Cheng, Jianheng Tang, Wenjie Li, Yefeng Zheng and Xiaodan Liang*
16:00-16:15	DialogueTRGAT: Temporal and Relational Graph Attention Network for Emotion Recognition in Conversations *Junjun Kang and Fang Kong*
16:15-16:30	Training Two-Stage Knowledge-grounded Dialogues with Attention Feedback *Zhen Li, Jiazhan Feng, Chongyang Tao and Dongyan Zhao*
16:30-16:45	Generating Emotional Responses with DialoGPT-based Multi-task Learning *Shuai Cao, Yuxiang Jia, Changyong Niu, Hongying Zan, Yutuan Ma and Shuo Xu*
16:45-17:00	MCIC: Multimodal Conversational Intent Classification for E-commerce Customer Service *Shaozu Yuan, Xin Shen, Yuming Zhao, Hang Liu, Zhiling Yan, Ruixue Liu and Meng Chen*

Oral Paper Session 10: Generation (Guilin room (桂林厅), 2nd floor)
Session Chair: *Kehai Chen*

15:45-16:00	Topic-Features for Dialogue Summarization *Zhen Zhang and Junhui Li*
16:00-16:15	Adversarial Fine-grained Fact Graph for Factuality-oriented Abstractive Summarization *Zhiguang Gao, Feng Jiang, Xiaomin Chu and Peifeng Li*
16:15-16:30	Retrieval, Selection and Writing: A Three-Stage Knowledge Grounded Storytelling Model *Wentao Qin and Dongyan Zhao*
16:30-16:45	An Adversarial Approach for Unsupervised Syntax-Guided Paraphrase Generation *Tang Xue, Yuran Zhao, Gongshen Liu and Xiaoyong Li*
16:45-17:00	ADS-Cap: A Framework for Accurate and Diverse Stylized Captioning with Unpaired Stylistic Corpora *Kanzhi Cheng, Zheng Ma, Shi Zong, Jianbing Zhang, Xinyu Dai and Jiajun Chen*

Index