

Juanzi Li · Heng Ji · Dongyan Zhao  
Yansong Feng (Eds.)

# Natural Language Processing and Chinese Computing

4th CCF Conference, NLPCC 2015  
Nanchang, China, October 9–13, 2015  
Proceedings

# Contents

## Fundamentals on Language Computing

A Maximum Entropy Approach to Discourse Coherence Modeling . . . . .	3
<i>Rui Lin, Muyun Yang, Shujie Liu, Sheng Li, and Tiejun Zhao</i>	
Transition-Based Dependency Parsing with Long Distance Collocations. . . . .	12
<i>Chenxi Zhu, Xipeng Qiu, and Xuanjing Huang</i>	
Recurrent Neural Networks with External Memory for Spoken Language Understanding . . . . .	25
<i>Baolin Peng, Kaisheng Yao, Li Jing, and Kam-Fai Wong</i>	
Improving Chinese Dependency Parsing with Lexical Semantic Features . . . . .	36
<i>Lvexing Zheng, Houfeng Wang, and Xueqiang Lv</i>	

## Machine Translation and Multi-Lingual Information Access

Entity Translation with Collective Inference in Knowledge Graph . . . . .	49
<i>Qinglin Li, Shujie Liu, Rui Lin, Mu Li, and Ming Zhou</i>	
Stochastic Language Generation Using Situated PCFGs . . . . .	64
<i>Caixia Yuan, Xiaojie Wang, and Ziming Zhong</i>	

## Machine Learning for NLP

Clustering Sentiment Phrases in Product Reviews by Constrained Co-clustering . . . . .	79
<i>Yujie Cao, Minlie Huang, and Xiaoyan Zhu</i>	
A Cross-Domain Sentiment Classification Method Based on Extraction of Key Sentiment Sentence. . . . .	90
<i>Shaowu Zhang, Huali Liu, Liang Yang, and Hongfei Lin</i>	
Convolutional Neural Networks for Correcting English Article Errors . . . . .	102
<i>Chengjie Sun, Xiaoqiang Jin, Lei Lin, Yuming Zhao, and Xiaolong Wang</i>	

## NLP for Social Media

Automatic Detection of Rumor on Social Network . . . . .	113
<i>Qiao Zhang, Shuiyuan Zhang, Jian Dong, Jinhua Xiong, and Xueqi Cheng</i>	

Multimodal Learning Based Approaches for Link Prediction in Social Networks . . . . .	123
<i>Feng Liu, Bingquan Liu, Chengjie Sun, Ming Liu, and Xiaolong Wang</i>	
Sentiment Analysis Based on User Tags for Traditional Chinese Medicine in Weibo . . . . .	134
<i>Junhui Shen, Peiyan Zhu, Rui Fan, Wei Tan, and Xueyan Zhan</i>	
Predicting User Mention Behavior in Social Networks . . . . .	146
<i>Bo Jiang, Ying Sha, and Lihong Wang</i>	
Convolutional Neural Networks for Multimedia Sentiment Analysis . . . . .	159
<i>Guoyong Cai and Binbin Xia</i>	
<b>Applications on Language Computing</b>	
An Adaptive Approach to Extract Characters from Digital Ink Text in Chinese Based on Extracted Errors . . . . .	171
<i>Hao Bai</i>	
Context-Dependent Metaphor Interpretation Based on Semantic Relatedness . . . . .	182
<i>Chang Su, Shuman Huang, and Yijiang Chen</i>	
Context Vector Model for Document Representation: A Computational Study . . . . .	194
<i>Yang Wei, Jinmao Wei, and Hengpeng Xu</i>	
<b>NLP for Search Technology and Ads</b>	
Refine Search Results Based on Desktop Context . . . . .	209
<i>Xiaoyun Li, Ying Yu, and Chunping Ouyang</i>	
Incorporating Semantic Knowledge with MRF Term Dependency Model in Medical Document Retrieval . . . . .	219
<i>Zhongda Xie, Yunqing Xia, and Qiang Zhou</i>	
A Full-Text Retrieval Algorithm for Encrypted Data in Cloud Storage Applications . . . . .	229
<i>Wei Song, Yihui Cui, and Zhiyong Peng</i>	
How Different Features Contribute to the Session Search? . . . . .	242
<i>Jingfei Li, Dawei Song, Peng Zhang, and Yuexian Hou</i>	

**Web Mining**

- Beyond Your Interests: Exploring the Information Behind User Tags . . . . . 257  
*Weizhi Ma, Min Zhang, Yiqun Liu, Shaoping Ma, and Lingfeng Chen*
- Nonparametric Symmetric Correspondence Topic Models for Multilingual  
 Text Analysis . . . . . 270  
*Rui Cai, Miaohong Chen, and Houfeng Wang*

**Knowledge Acquisition and Information Extraction**

- Mining RDF from Tables in Chinese Encyclopedias . . . . . 285  
*Weiming Lu, Zhenyu Zhang, Renjie Lou, Hao Dai, Shansong Yang,  
 and Baogang Wei*
- Taxonomy Induction from Chinese Encyclopedias by Combinatorial  
 Optimization. . . . . 299  
*Weiming Lu, Renjie Lou, Hao Dai, Zhenyu Zhang, Shansong Yang,  
 and Baogang Wei*
- Recognition of Person Relation Indicated by Predicates . . . . . 313  
*Zhongping Liang, Caixia Yuan, Bing Leng, and Xiaojie Wang*
- Target Detection and Knowledge Learning for Domain Restricted  
 Question Answering . . . . . 325  
*Mengdi Zhang, Tao Huang, Yixin Cao, and Lei Hou*

**Short Papers**

- An Improved Algorithm of Logical Structure Reconstruction  
 for Re-flowable Document Understanding . . . . . 339  
*Lin Zhao, Ning Li, Xin Peng, and Qi Liang*
- Mongolian Inflection Suffix Processing in NLP: A Case Study . . . . . 347  
*Xiangdong Su, Guanglai Gao, Yupeng Jiang, Jing Wu, and Feilong Bao*
- Resolving Coordinate Structures for Chinese Constituent Parsing . . . . . 353  
*Yichu Zhou, Shujian Huang, Xinyu Dai, and Jiajun Chen*
- P-Trie Tree: A Novel Tree Structure for Storing Polysemantic Data . . . . . 362  
*Xin Zhou*
- Research on the Extraction of Wikipedia-Based Chinese-Khmer Named  
 Entity Equivalents . . . . . 372  
*Qing Xia, Xin Yan, Zhengtao Yu, and Shengxiang Gao*

Bilingual Lexicon Extraction with Temporal Distributed Word Representation from Comparable Corpora . . . . .	380
<i>Chunyue Zhang and Tiejun Zhao</i>	
Bilingually-Constrained Recursive Neural Networks with Syntactic Constraints for Hierarchical Translation Model . . . . .	388
<i>Wei Chen and Bo Xu</i>	
Document-Level Machine Translation Evaluation Metrics Enhanced with Simplified Lexical Chain . . . . .	396
<i>Zhengxian Gong and Guodong Zhou</i>	
Cross-Lingual Tense Tagging Based on Markov Tree Tagging Model . . . . .	404
<i>Yijiang Chen, Tingting Zhu, Chang Su, and Xiaodong Shi</i>	
Building a Large-Scale Cross-Lingual Knowledge Base from Heterogeneous Online Wikis . . . . .	413
<i>Mingyang Li, Yao Shi, Zhigang Wang, and Yongbin Liu</i>	
Refining Kazakh Word Alignment Using Simulation Modeling Methods for Statistical Machine Translation . . . . .	421
<i>Amandyk Kartbayev</i>	
A Local Method for Canonical Correlation Analysis . . . . .	428
<i>Tengju Ye, Zhipeng Xie, and Ang Li</i>	
Learning to Rank Microblog Posts for Real-Time Ad-Hoc Search . . . . .	436
<i>Jing Li, Zhongyu Wei, Hao Wei, Kangfei Zhao, Junwen Chen, and Kam-Fai Wong</i>	
Fuzzy-Rough Set Based Multi-labeled Emotion Intensity Analysis for Sentence, Paragraph and Document . . . . .	444
<i>Chu Wang, Shi Feng, Daling Wang, and Yifei Zhang</i>	
What Causes Different Emotion Distributions of a Hot Event? A Deep Event-Emotion Analysis System on Microblogs . . . . .	453
<i>Yanyan Zhao, Bing Qin, Zhenjiang Dong, Hong Chen, and Ting Liu</i>	
Deceptive Opinion Spam Detection Using Deep Level Linguistic Features . . . . .	465
<i>Changge Chen, Hai Zhao, and Yang Yang</i>	
Multi-sentence Question Segmentation and Compression for Question Answering . . . . .	475
<i>Yixiu Wang, Yunfang Wu, and Xueqiang Lv</i>	
A User-Oriented Special Topic Generation System for Digital Newspaper . . . . .	484
<i>Xi Xu, Mao Ye, Zhi Tang, Jian-Bo Xu, and Liang-Cai Gao</i>	

**Shared Task (Long Papers)**

Exploiting Heterogeneous Annotations for Weibo Word Segmentation and POS Tagging . . . . .	495
<i>Jiayuan Chao, Zhenghua Li, Wenliang Chen, and Min Zhang</i>	
Entity Recognition and Linking in Chinese Search Queries . . . . .	507
<i>Jinwei Yuan, Yan Yang, Zheng Jia, Hongfeng Yin, Junfu Huang, and Jie Zhu</i>	
BosonNLP: An Ensemble Approach for Word Segmentation and POS Tagging . . . . .	520
<i>Kerui Min, Chenggang Ma, Tianmei Zhao, and Haiyan Li</i>	
Research on Open Domain Question Answering System . . . . .	527
<i>Zhonglin Ye, Zheng Jia, Yan Yang, Junfu Huang, and Hongfeng Yin</i>	
Overview of the NLPCC 2015 Shared Task: Chinese Word Segmentation and POS Tagging for Micro-blog Texts . . . . .	541
<i>Xipeng Qiu, Peng Qian, Liusong Yin, Shiyu Wu, and Xuanjing Huang</i>	
Overview of the NLPCC 2015 Shared Task: Entity Recognition and Linking in Search Queries . . . . .	550
<i>Yansong Feng, Zhe Han, and Kun Zhang</i>	
Overview of the NLPCC 2015 Shared Task: Weibo-Oriented Chinese News Summarization. . . . .	557
<i>Xiaojun Wan, Jianmin Zhang, Shiyang Wen, and Jiwei Tan</i>	
Overview of the NLPCC 2015 Shared Task: Open Domain QA . . . . .	562
<i>Nan Duan</i>	

**Short Task (Short Papers)**

Word Segmentation of Micro Blogs with Bagging . . . . .	573
<i>Zhenting Yu, Xin-Yu Dai, Si Shen, Shujian Huang, and Jiajun Chen</i>	
Weibo-Oriented Chinese News Summarization via Multi-feature Combination. . . . .	581
<i>Maofu Liu, Limin Wang, and Liqiang Nie</i>	
Linking Entities in Chinese Queries to Knowledge Graph . . . . .	590
<i>Jun Li, Jinxian Pan, Chen Ye, Yong Huang, Danlu Wen, and Zhichun Wang</i>	

A Hybrid Re-ranking Method for Entity Recognition and Linking in Search Queries . . . . .	598
<i>Gongbo Tang, Yuting Guo, Dong Yu, and Endong Xun</i>	
<b>Author Index</b> . . . . .	607