

Chin-Yew Lin · Nianwen Xue
Dongyan Zhao · Xuanjing Huang
Yansong Feng (Eds.)

Natural Language Understanding and Intelligent Applications

5th CCF Conference on Natural Language Processing
and Chinese Computing, NLPCC 2016 and
24th International Conference on Computer Processing
of Oriental Languages, ICCPOL 2016
Kunming, China, December 2–6, 2016
Proceedings

Contents

Fundamentals on Language Computing

Integrating Structural Context with Local Context for Disambiguating Word Senses	3
<i>Qianlong Du, Chengqing Zong, and Keh-Yih Su</i>	
Tibetan Multi-word Expressions Identification Framework Based on News Corpora	16
<i>Minghua Nuo, Congjun Lun, and Huidan Liu</i>	
Building Powerful Dependency Parsers for Resource-Poor Languages	27
<i>Junjie Yu, Wenliang Chen, Zhenghua Li, and Min Zhang</i>	
Bidirectional Long Short-Term Memory with Gated Relevance Network for Paraphrase Identification	39
<i>Yatian Shen, Jifan Chen, and Xuanjing Huang</i>	
Syntactic Categorization and Semantic Interpretation of Chinese Nominal Compounds	51
<i>Taizhong Wu, Jian Liu, Xuri Tang, Min Gu, Yanhui Gu, Junsheng Zhou, and Weiguang Qu</i>	
TDSS: A New Word Sense Representation Framework for Information Retrieval	63
<i>Liwei Chen, Yansong Feng, and Dongyan Zhao</i>	
A Word Vector Representation Based Method for New Words Discovery in Massive Text	76
<i>Yang Du, Hua Yuan, and Yu Qian</i>	

Machine Translation and Multi-lingual Information Access

Better Addressing Word Deletion for Statistical Machine Translation	91
<i>Qiang Li, Dongdong Zhang, Mu Li, Tong Xiao, and Jingbo Zhu</i>	
A Simple, Straightforward and Effective Model for Joint Bilingual Terms Detection and Word Alignment in SMT	103
<i>Guoping Huang, Jiajun Zhang, Yu Zhou, and Chengqing Zong</i>	
Bilingual Parallel Active Learning Between Chinese and English	116
<i>Longhua Qian, JiaXin Liu, Guodong Zhou, and Qiaoming Zhu</i>	

Study on the English Corresponding Unit of Chinese Clause 129
Wenhe Feng, Yi Yang, Yancui Li, Xia Li, and Han Ren

Research for Uyghur-Chinese Neural Machine Translation 141
Jinying Kong, Yating Yang, Xi Zhou, Lei Wang, and Xiao Li

MaxSD: A Neural Machine Translation Evaluation Metric Optimized
 by Maximizing Similarity Distance 153
*Qingsong Ma, Fandong Meng, Daqi Zheng, Mingxuan Wang,
 Yvette Graham, Wenbin Jiang, and Qun Liu*

Automatic Long Sentence Segmentation for Neural Machine Translation 162
Shaohui Kuang and Deyi Xiong

Machine Learning for NLP

Topic Segmentation of Web Documents with Automatic Cue Phrase
 Identification and BLSTM-CNN 177
Liang Wang, Sujian Li, Xinyan Xiao, and Yajuan Lyu

Multi-task Learning for Gender and Age Prediction on Chinese Microblog. . . 189
Liang Wang, Qi Li, Xuan Chen, and Sujian Li

Dropout Non-negative Matrix Factorization for Independent Feature
 Learning 201
*Zhicheng He, Jie Liu, Caihua Liu, Yuan Wang, Airu Yin,
 and Yalou Huang*

Analysing the Semantic Change Based on Word Embedding 213
Xuanyi Liao and Guang Cheng

Learning Word Sense Embeddings from Word Sense Definitions 224
Qi Li, Tianshi Li, and Baobao Chang

Information Extraction, Question Answering and Knowledge Acquisition

Character-Based LSTM-CRF with Radical-Level Features for Chinese
 Named Entity Recognition 239
*Chuanhai Dong, Jiajun Zhang, Chengqing Zong, Masanori Hattori,
 and Hui Di*

Improving First Order Temporal Fact Extraction with Unreliable Data 251
Bingfeng Luo, Yansong Feng, Zheng Wang, and Dongyan Zhao

Reducing Human Effort in Named Entity Corpus Construction Based
 on Ensemble Learning and Annotation Categorization 263
Tingming Lu, Man Zhu, and Zhiqiang Gao

A Convolution BiLSTM Neural Network Model for Chinese Event Extraction	275
<i>Ying Zeng, Honghui Yang, Yansong Feng, Zheng Wang, and Dongyan Zhao</i>	
Detection of Entity Mixture in Knowledge Bases Using Hierarchical Clustering	288
<i>Haihua Xie, Xiaoqing Lu, Zhi Tang, and Xiaojun Huang</i>	
Knowledge Base Question Answering Based on Deep Learning Models	300
<i>Zhiwen Xie, Zhao Zeng, Guangyou Zhou, and Tingting He</i>	
An Open Domain Topic Prediction Model for Answer Selection	312
<i>Zhao Yan, Nan Duan, Ming Zhou, Zhoujun Li, and Jianshe Zhou</i>	
Joint Event Extraction Based on Skip-Window Convolutional Neural Networks	324
<i>Zhengkuan Zhang, Weiran Xu, and Qianqian Chen</i>	
Improving Collaborative Filtering with Long-Short Interest Model	335
<i>Chao Lv, Lili Yao, Yansong Feng, and Dongyan Zhao</i>	
Discourse Analysis	
Leveraging Hierarchical Deep Semantics to Classify Implicit Discourse Relations via Mutual Learning Method	349
<i>Xiaohan She, Ping Jian, Pengcheng Zhang, and Heyan Huang</i>	
Transition-Based Discourse Parsing with Multilayer Stack Long Short Term Memory	360
<i>Yanyan Jia, Yansong Feng, Bingfeng Luo, Yuan Ye, Tianyang Liu, and Dongyan Zhao</i>	
Predicting Implicit Discourse Relation with Multi-view Modeling and Effective Representation Learning	374
<i>Haoran Li, Jiajun Zhang, Yu Zhou, and Chengqing Zong</i>	
A CDT-Styled End-to-End Chinese Discourse Parser	387
<i>Fang Kong, Hongling Wang, and Guodong Zhou</i>	
NLP for Social Media	
Events Detection and Temporal Analysis in Social Media	401
<i>Yawei Jia, Jing Xu, Zhonghu Xu, and Kai Xing</i>	
Discovering Concept-Level Event Associations from a Text Stream	413
<i>Tao Ge, Lei Cui, Heng Ji, Baobao Chang, and Zhifang Sui</i>	

A User Adaptive Model for Followee Recommendation on Twitter 425
Yang Liu, Xuan Chen, Sujian Li, and Liang Wang

Who Will Tweet More? Finding Information Feeders in Twitter 437
Beibei Gu, Zhunchen Luo, and Xin Wang

Short Papers

Discrete and Neural Models for Chinese POS Tagging: Comparison and Combination. 451
Meishan Zhang, Nan Yu, and Guohong Fu

Improving Word Vector with Prior Knowledge in Semantic Dictionary 461
Wei Li, Yunfang Wu, and Xueqiang Lv

Adapting Attention-Based Neural Network to Low-Resource Mongolian-Chinese Machine Translation 470
Jing Wu, Hongxu Hou, Zhipeng Shen, Jian Du, and Jinting Li

Sentence Similarity on Structural Representations 481
Meng Yang, Peifeng Li, and Qiaoming Zhu

Word Sense Disambiguation Using Context Translation. 489
Zhizhuo Yang, Hu Zhang, Qian Chen, and Hongye Tan

Cyrillic Mongolian Named Entity Recognition with Rich Features 497
Weihua Wang, Feilong Bao, and Guanglai Gao

Purchase Prediction via Machine Learning in Mobile Commerce. 506
Chao Lv, Yansong Feng, and Dongyan Zhao

Exploring Long Tail Data in Distantly Supervised Relation Extraction. 514
Yaocheng Gui, Qian Liu, Man Zhu, and Zhiqiang Gao

Detecting Potential Adverse Drug Reactions from Health-Related Social Networks 523
Bo Xu, Hongfei Lin, Mingzhen Zhao, Zhihao Yang, Jian Wang, and Shaowu Zhang

Iterative Integration of Unsupervised Features for Chinese Dependency Parsing 531
Te Luo, Yujie Zhang, Jinan Xu, and Yufeng Chen

Can We Neglect Function Words in Word Embedding?. 541
Gongbo Tang, Gaoqi Rao, Dong Yu, and Endong Xun

A Similarity Algorithm Based on the Generality and Individuality of Words	549
<i>Yinfeng Zou, Chunping Ouyang, Yongbin Liu, Xiaohua Yang, and Ying Yu</i>	
An Improved Information Gain Algorithm Based on Relative Document Frequency Distribution.	559
<i>Jian Peng, Xiao-Hua Yang, Chun-Ping Ouyang, and Yong-Bin Liu</i>	
Finding the True Crowds: User Filtering in Microblogs	568
<i>Bin Hao, Min Zhang, Weizhi Ma, Jiashen Sun, Yiqun Liu, Shaoping Ma, Xuan Zhu, and Hengliang Luo</i>	
Learning to Recognize Protected Health Information in Electronic Health Records with Recurrent Neural Network	575
<i>Kun Li, Yumei Chai, Hongling Zhao, Xiaofei Nan, and Yueshu Zhao</i>	
Sentiment Classification of Social Media Text Considering User Attributes	583
<i>Junjie Li, Haitong Yang, and Chengqing Zong</i>	
Learning from User Feedback for Machine Translation in Real-Time.	595
<i>Guoping Huang, Jiajun Zhang, Yu Zhou, and Chengqing Zong</i>	
GuideRank: A Guided Ranking Graph Model for Multilingual Multi-document Summarization.	608
<i>Haoran Li, Jiajun Zhang, Yu Zhou, and Chengqing Zong</i>	
Fast-Syntax-Matching-Based Japanese-Chinese Limited Machine Translation	621
<i>Wuying Liu and Lin Wang</i>	
Value at Risk for Risk Evaluation in Information Retrieval	631
<i>Meijia Wang, Peng Zhang, Dawei Song, and Jun Wang</i>	
Chinese Paraphrases Acquisition Based on Random Walk N Step	639
<i>Jun Ma, Yujie Zhang, Jinan Xu, and Yufeng Chen</i>	
A Micro-topic Model for Coreference Resolution Based on Theme-Rheme Structure	648
<i>Xue-feng Xi and Guodong Zhou</i>	
Learning from LDA Using Deep Neural Networks	657
<i>Dongxu Zhang, Tianyi Luo, and Dong Wang</i>	
Relation Classification: CNN or RNN?	665
<i>Dongxu Zhang and Dong Wang</i>	

Shared Tasks

Ensemble of Feature Sets and Classification Methods for Stance Detection. . .	679
<i>Jiaming Xu, Suncong Zheng, Jing Shi, Yiqun Yao, and Bo Xu</i>	
Exploiting External Knowledge and Entity Relationship for Entity Search . . .	689
<i>Le Li, Junyi Xu, Weidong Xiao, Shengze Hu, and Haiming Tong</i>	
A Flexible and Sentiment-Aware Framework for Entity Search.	701
<i>Kerui Min, Chenghao Dong, Shiyuan Cai, and Jianhao Chen</i>	
Word Segmentation on Micro-Blog Texts with External Lexicon and Heterogeneous Data	711
<i>Qingrong Xia, Zhenghua Li, Jiayuan Chao, and Min Zhang</i>	
Open Domain Question Answering System Based on Knowledge Base	722
<i>Yuxuan Lai, Yang Lin, Jiahao Chen, Yansong Feng, and Dongyan Zhao</i>	
Recurrent Neural Word Segmentation with Tag Inference	734
<i>Qianrong Zhou, Long Ma, Zhenyu Zheng, Yue Wang, and Xiaojie Wang</i>	
Chinese Word Similarity Computing Based on Combination Strategy	744
<i>Shaoru Guo, Yong Guan, Ru Li, and Qi Zhang</i>	
An Empirical Study on Chinese Microblog Stance Detection Using Supervised and Semi-supervised Machine Learning Methods.	753
<i>Liran Liu, Shi Feng, Daling Wang, and Yifei Zhang</i>	
Combining Word Embedding and Semantic Lexicon for Chinese Word Similarity Computation	766
<i>Jiahuan Pei, Cong Zhang, Degen Huang, and Jianjun Ma</i>	
Football News Generation from Chinese Live Webcast Script.	778
<i>Tang Renjun, Zhang Ke, Na Shenruoyang, Yang Minghao, Zhou Hui, Zhu Qingjie, Zhan Yongsong, and Tao Jianhua</i>	
Convolutional Deep Neural Networks for Document-Based Question Answering	790
<i>Jian Fu, Xipeng Qiu, and Xuanjing Huang</i>	
Research on Summary Sentences Extraction Oriented to Live Sports Text . . .	798
<i>Liya Zhu, Wenchao Wang, Yujing Chen, Xueqiang Lv, and Jianshe Zhou</i>	

Short Papers

Statistical Entity Ranking with Domain Knowledge.	811
<i>Xiao-Bo Jin, Guang-Gang Geng, Kaizhu Huang, and Zhi-Wei Yan</i>	

Study on the Method of Precise Entity Search Based on Baidu’s Query 819
Teng Wang, Xueqiang Lv, Xun Ma, Pengyan Sun, Zhian Dong, and Jianshe Zhou

Overview of the NLPCC-ICCPOL 2016 Shared Task: Chinese Word Similarity Measurement 828
Yunfang Wu and Wei Li

Exploring Various Linguistic Features for Stance Detection 840
Qingying Sun, Zhongqing Wang, Qiaoming Zhu, and Guodong Zhou

Overview of Baidu Cup 2016: Challenge on Entity Search. 848
Ke Sun, Tingting Li, Shiqi Zhao, Yajuan Lv, Yansong Feng, Xiaojun Wan, and Dongyan Zhao

A Feature-Rich CRF Segmenter for Chinese Micro-Blog 854
Yabin Leng, Weiwei Liu, Sheng Wang, and Xiaojie Wang

NLPCC 2016 Shared Task Chinese Words Similarity Measure via Ensemble Learning Based on Multiple Resources 862
Shutian Ma, Xiaoyong Zhang, and Chengzhi Zhang

Overview of the NLPCC-ICCPOL 2016 Shared Task: Sports News Generation from Live Webcast Scripts 870
Xiaojun Wan, Jianmin Zhang, Jin-ge Yao, and Tianming Wang

Sports News Generation from Live Webcast Scripts Based on Rules and Templates 876
Maofu Liu, Qiaosong Qi, Huijun Hu, and Han Ren

A Deep Learning Approach for Question Answering Over Knowledge Base 885
Linjie Wang, Yu Zhang, and Ting Liu

Stance Detection in Chinese MicroBlogs with Neural Networks 893
Nan Yu, Da Pan, Meishan Zhang, and Guohong Fu

Overview of the NLPCC-ICCPOL 2016 Shared Task: Chinese Word Segmentation for Micro-Blog Texts. 901
Xipeng Qiu, Peng Qian, and Zhan Shi

Overview of NLPCC Shared Task 4: Stance Detection in Chinese Microblogs. 907
Ruifeng Xu, Yu Zhou, Dongyin Wu, Lin Gui, Jiachen Du, and Yun Xue

Combining Deep Learning with Information Retrieval for Question Answering	917
<i>Fengyu Yang, Liang Gan, Aiping Li, Dongchuan Huang, Xiaohui Chou, and Hongmei Liu</i>	
A Hybrid Approach to DBQA	926
<i>Fangying Wu, Muyun Yang, Tiejun Zhao, Zhongyuan Han, Dequan Zheng, and Shanshan Zhao</i>	
A Chinese Question Answering Approach Integrating Count-Based and Embedding-Based Features.	934
<i>Benyou Wang, Jiabin Niu, Liqun Ma, Yuhua Zhang, Lipeng Zhang, Jingfei Li, Peng Zhang, and Dawei Song</i>	
Overview of the NLPCC-ICCPOL 2016 Shared Task: Open Domain Chinese Question Answering	942
<i>Nan Duan</i>	
Author Index	949