

# **Lecture Notes in Artificial Intelligence**

**11221**

Subseries of Lecture Notes in Computer Science

LNAI Series Editors

Randy Goebel

*University of Alberta, Edmonton, Canada*

Yuzuru Tanaka

*Hokkaido University, Sapporo, Japan*

Wolfgang Wahlster

*DFKI and Saarland University, Saarbrücken, Germany*

LNAI Founding Series Editor

Joerg Siekmann

*DFKI and Saarland University, Saarbrücken, Germany*

More information about this series at <http://www.springer.com/series/1244>

Maosong Sun · Ting Liu  
Xiaojie Wang · Zhiyuan Liu  
Yang Liu (Eds.)

# Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data

17th China National Conference, CCL 2018  
and 6th International Symposium, NLP-NABD 2018  
Changsha, China, October 19–21, 2018  
Proceedings

*Editors*

Maosong Sun  
Tsinghua University  
Beijing, China

Ting Liu  
Harbin Institute of Technology  
Harbin, China

Xiaojie Wang  
Beijing University of Posts  
and Telecommunications  
Beijing, China

Zhiyuan Liu  
Tsinghua University  
Beijing, China

Yang Liu  
Tsinghua University  
Beijing, China

ISSN 0302-9743 ISSN 1611-3349 (electronic)  
Lecture Notes in Artificial Intelligence  
ISBN 978-3-030-01715-6 ISBN 978-3-030-01716-3 (eBook)  
<https://doi.org/10.1007/978-3-030-01716-3>

Library of Congress Control Number: 2018956472

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer Nature Switzerland AG 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

## Preface

Welcome to the proceedings of the 17th China National Conference on Computational Linguistics (17th CCL) and the 6th International Symposium on Natural Language Processing Based on Naturally Annotated Big Data (6th NLP-NABD). The conference and symposium were hosted by Changsha University of Science and Technology located in Changsha City, Hunan Province, China.

CCL is an annual conference (bi-annual before 2013) that started in 1991. It is the flagship conference of the Chinese Information Processing Society of China (CIPS), which is the largest NLP scholar and expert community in China. CCL is a premier nation-wide forum for disseminating new scholarly and technological work in computational linguistics, with a major emphasis on computer processing of the languages in China such as Mandarin, Tibetan, Mongolian, and Uyghur.

Affiliated with the 17th CCL, the 6th International Symposium on Natural Language Processing Based on Naturally Annotated Big Data (NLP-NABD) covered all the topics of NLP, with particular focus on methodologies and techniques relating to naturally annotated big data. In contrast to manually annotated data such as treebanks that are constructed for specific NLP tasks, naturally annotated data come into existence through users' normal activities, such as writing, conversation, and interactions on the Web. Although the original purposes of these data typically were unrelated to NLP, they can nonetheless be purposefully exploited by computational linguists to acquire linguistic knowledge. For example, punctuation marks in Chinese text can help word boundaries identification, social tags in social media can provide signals for keyword extraction, and categories listed in Wikipedia can benefit text classification. The natural annotation can be explicit, as in the aforementioned examples, or implicit, as in Hearst patterns (e.g., "Beijing and other cities" implies "Beijing is a city"). This symposium focuses on numerous research challenges ranging from very-large-scale unsupervised/semi-supervised machine learning (deep learning, for instance) of naturally annotated big data to integration of the learned resources and models with existing handcrafted "core" resources and "core" language computing models. NLP-NABD 2018 was supported by the National Key Basic Research Program of China (i.e., "973" Program) "Theory and Methods for Cyber-Physical-Human Space Oriented Web Chinese Information Processing" under grant no.2014CB340500 and the Major Project of the National Social Science Foundation of China under grant no. 13&ZD190.

The Program Committee selected 102 papers (69 Chinese papers and 33 English papers) out of 277 submissions from China, Hong Kong (region), Singapore, and USA for publication. The acceptance rate is 36.82%. The 33 English papers cover the following topics:

- Semantics (3)
- Machine translation (6)
- Knowledge graph and information extraction (7)
- Linguistic resource annotation and evaluation (2)

- Information retrieval and question answering (4)
- Text classification and summarization (5)
- Social computing and sentiment analysis (2)
- NLP applications (4)

The final program for the 17th CCL and the sixth NLP-NABD was the result of a great deal of work by many dedicated colleagues. We want to thank, first of all, the authors who submitted their papers, and thus contributed to the creation of the high-quality program that allowed us to look forward to an exciting joint conference. We are deeply indebted to all the Program Committee members for providing high-quality and insightful reviews under a tight schedule. We are extremely grateful to the sponsors of the conference. Finally, we extend a special word of thanks to all the colleagues of the Organizing Committee and secretariat for their hard work in organizing the conference, and to Springer for their assistance in publishing the proceedings in due time.

We thank the Program and Organizing Committees for helping to make the conference successful, and we hope all the participants enjoyed a remarkable visit to Changsha, a historical and beautiful city in South China.

August 2018

Maosong Sun  
Ting Liu  
Xiaojie Wang  
Randy Goebel  
Heng Ji

# Organization

## General Chairs

Sheng Li	Harbin Institute of Technology, China
Changning Huang	Tsinghua University, China
Kaiying Liu	Shanxi University, China

## Program Committee

### 17th CCL Program Chairs

Maosong Sun	Tsinghua University, China
Ting Liu	Harbin Institute of Technology, China
Xiaojie Wang	Beijing University of Posts and Telecommunications, China

### 17th CCL Area Co-chairs

#### Linguistics and Cognitive Science

Weiguang Qu	Nanjing Normal University, China
Yulin Yuan	Peking University, China

#### Fundamental Theory and Methods of Computational Linguistics

Wanxiang Che	Harbin Institute of Technology, China
Yue Zhang	Singapore University of Technology and Design, Singapore

#### Information Retrieval and Question Answering

Jun Xu	Institute of Computing Technology, CAS, China
Aixin Sun	Nanyang Technological University, Singapore

#### Text Classification and Summarization

Xipeng Qiu	Fudan University, China
Xiaodan Zhu	Queen's University, Canada

#### Knowledge Graph and Information Extraction

Xianpei Han	Institute of Software, CAS, China
Ni Lao	Google Inc., USA

#### Machine Translation

Shujian Huang	Nanjing University, China
Haitao Mi	Ant Financial, USA

#### Minority Language Information Processing

Xiaobing Zhao	Minzu University of China, China
Celimuge Wu	The University of Electro-Communications, Japan

### **Language Resource and Evaluation**

Muyun Yang	Harbin Institute of Technology, China
Rui Wang	National Institute of Information and Communications Technology, NICT, Japan

### **Social Computing and Sentiment Analysis**

Jia Jia	Tsinghua University, China
Yanyan Zhao	Harbin Institute of Technology, China

### **NLP Applications**

Qingcai Chen	Harbin Institute of Technology, Shenzhen, China
Cui Tao	The University of Texas Health Science Center at Houston, USA

### **6th NLP-NABD Program Chairs**

Maosong Sun	Tsinghua University, China
Randy Goebel	University of Alberta, Canada
Heng Ji	Rensselaer Polytechnic Institute, USA

### **17th CCL and 6th NLP-NABD Local Arrangements Chairs**

Feng Li	Changsha University of Science and Technology, China
Jin Wang	Changsha University of Science and Technology, China
Yang Liu	Tsinghua University, China

### **17th CCL and 6th NLP-NABD Evaluation Chairs**

Ting Liu	Harbin Institute of Technology, China
Wei Song	Capital Normal University, China

### **17th CCL and 6th NLP-NABD Publication Chairs**

Zhiyuan Liu	Tsinghua University, China
Shizhu He	Institute of Automation, CAS, China

### **17th CCL and 6th NLP-NABD Publicity Chairs**

Ruifeng Xu	Harbin Institute of Technology, Shenzhen, China
William Wang	University of California, Santa Barbara, USA

### **17th CCL and 6th NLP-NABD Advance Lectures Chairs**

Xiaojun Wan	Peking University, China
Jiafeng Guo	Institute of Computing Technology, CAS, China

### 17th CCL and 6th NLP-NABD Sponsorship Chairs

Qi Zhang	Fudan University, China
Jiajun Zhang	Institute of Automation, CAS, China

### 17th CCL and 6th NLP-NABD System Demonstration Chairs

Tong Xiao	Northeastern University, China
Binyang Li	University of International Relations, China

### 17th CCL and 6th NLP-NABD Student Seminar Chairs

Yang Feng	Institute of Computing Technology, CAS, China
Dong Yu	Beijing Language and Culture University, China

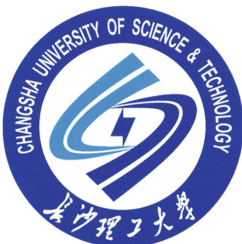
### 17th CCL and 6th NLP-NABD Organizers



Chinese Information Processing Society of China



Tsinghua University



Changsha University of Science & Technology, China

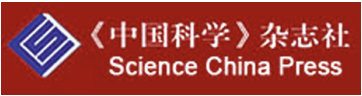
**Publishers**



Lecture Notes in Artificial Intelligence,  
Springer



Journal of Chinese Information  
Processing



Science China



Journal of Tsinghua University  
(Science and Technology)

**Sponsoring Institutions**

**Platinum**



犀语科技  
AlphaInsight



**Gold**



**Silver**



**Bronze**



# Contents

## Semantics

Radical Enhanced Chinese Word Embedding . . . . .	3
<i>Zheng Chen and Keqi Hu</i>	
Syntax Enhanced Research Method of Stylistic Features . . . . .	12
<i>Haiyan Wu and Ying Liu</i>	
Addressing Domain Adaptation for Chinese Word Segmentation with Instances-Based Transfer Learning . . . . .	24
<i>Yanna Zhang, Jinan Xu, Guoyi Miao, Yufeng Chen, and Yujie Zhang</i>	

## Machine Translation

Collaborative Matching for Sentence Alignment . . . . .	39
<i>Xiaojun Quan, Chunyu Kit, and Wuya Chen</i>	
Finding Better Subword Segmentation for Neural Machine Translation . . . . .	53
<i>Yingting Wu and Hai Zhao</i>	
Improving Low-Resource Neural Machine Translation with Weight Sharing . . . . .	65
<i>Tao Feng, Miao Li, Xiaojun Liu, and Yichao Cao</i>	
Identifying Word Translations in Scientific Literature Based on Labeled Bilingual Topic Model and Co-occurrence Features . . . . .	76
<i>Mingjie Tian, Yahui Zhao, and Rongyi Cui</i>	
Term Translation Extraction from Historical Classics Using Modern Chinese Explanation . . . . .	88
<i>Xiaoting Wu, Hanyu Zhao, and Chao Che</i>	
Research on Chinese-Tibetan Neural Machine Translation . . . . .	99
<i>Wen Lai, Xiaobing Zhao, and Xiaqing Li</i>	

## Knowledge Graph and Information Extraction

Metadata Extraction for Scientific Papers . . . . .	111
<i>Binjie Meng, Lei Hou, Erhong Yang, and Juanzi Li</i>	
Knowledge Graph Embedding with Logical Consistency . . . . .	123
<i>Jianfeng Du, Kunxun Qi, and Yuming Shen</i>	

An End-to-End Entity and Relation Extraction Network with Multi-head Attention . . . . .	136
<i>Lishuang Li, Yuankai Guo, Shuang Qian, and Anqiao Zhou</i>	
Attention-Based Convolutional Neural Networks for Chinese Relation Extraction . . . . .	147
<i>Wenya Wu, Yufeng Chen, Jinan Xu, and Yujie Zhang</i>	
A Study on Improving End-to-End Neural Coreference Resolution . . . . .	159
<i>Jia-Chen Gu, Zhen-Hua Ling, and Nitin Indurkha</i>	
Type Hierarchy Enhanced Heterogeneous Network Embedding for Fine-Grained Entity Typing in Knowledge Bases . . . . .	170
<i>Hailong Jin, Lei Hou, and Juanzi Li</i>	
Scientific Keyphrase Extraction: Extracting Candidates with Semi-supervised Data Augmentation. . . . .	183
<i>Qianying Liu, Daisuke Kawahara, and Sujian Li</i>	
<b>Linguistic Resource Annotation and Evaluation</b>	
Using a Chinese Lexicon to Learn Sense Embeddings and Measure Semantic Similarity . . . . .	197
<i>Zhuo Zhen and Yuquan Chen</i>	
Revisiting Correlations between Intrinsic and Extrinsic Evaluations of Word Embeddings. . . . .	209
<i>Yuanqian Qiu, Hongzheng Li, Shen Li, Yingdi Jiang, Renfen Hu, and Lijiao Yang</i>	
<b>Information Retrieval and Question Answering</b>	
Question-Answering Aspect Classification with Hierarchical Attention Network. . . . .	225
<i>Hanqian Wu, Mumu Liu, Jingjing Wang, Jue Xie, and Chenlin Shen</i>	
End-to-End Task-Oriented Dialogue System with Distantly Supervised Knowledge Base Retriever . . . . .	238
<i>Libo Qin, Yijia Liu, Wanxiang Che, Haoyang Wen, and Ting Liu</i>	
Attention-Based CNN-BLSTM Networks for Joint Intent Detection and Slot Filling. . . . .	250
<i>Yufan Wang, Li Tang, and Tingting He</i>	
Multi-Perspective Fusion Network for Commonsense Reading Comprehension . . . . .	262
<i>Chunhua Liu, Yan Zhao, Qingyi Si, Haiou Zhang, Bohan Li, and Dong Yu</i>	

**Text Classification and Summarization**

A Hierarchical Hybrid Neural Network Architecture for Chinese Text Summarization . . . . .	277
<i>Yunheng Zhang, Leihan Zhang, Ke Xu, and Le Zhang</i>	
TSABCNN: Two-Stage Attention-Based Convolutional Neural Network for Frame Identification . . . . .	289
<i>Hongyan Zhao, Ru Li, Fei Duan, Zepeng Wu, and Shaoru Guo</i>	
Linked Document Classification by Network Representation Learning . . . . .	302
<i>Yue Zhang, Liying Zhang, and Yao Liu</i>	
A Word Embedding Transfer Model for Robust Text Categorization . . . . .	314
<i>Yiming Zhang, Jing Wang, Weijian Deng, and Yaojie Lu</i>	
Review Headline Generation with User Embedding. . . . .	324
<i>Tianshang Liu, Haoran Li, Junnan Zhu, Jiajun Zhang, and Chengqing Zong</i>	

**Social Computing and Sentiment Analysis**

A Joint Model for Sentiment Classification and Opinion Words Extraction. . .	337
<i>Dawei Cong, Jianhua Yuan, Yanyan Zhao, and Bing Qin</i>	
Network Representation Learning Based on Community and Text Features . . .	348
<i>Yu Zhu, Zhonglin Ye, Haixing Zhao, and Ke Zhang</i>	

**NLP Applications**

Learning to Detect Verbose Expressions in Spoken Texts. . . . .	363
<i>Qingbin Liu, Shizhu He, Kang Liu, Shengping Liu, and Jun Zhao</i>	
Medical Knowledge Attention Enhanced Neural Model for Named Entity Recognition in Chinese EMR . . . . .	376
<i>Zhichang Zhang, Yu Zhang, and Tong Zhou</i>	
Coherence-Based Automated Essay Scoring Using Self-attention. . . . .	386
<i>Xia Li, Minping Chen, Jianyun Nie, Zhenxing Liu, Ziheng Feng, and Yingdan Cai</i>	
Trigger Words Detection by Integrating Attention Mechanism into Bi-LSTM Neural Network—A Case Study in PubMed-Wide Trigger Words Detection for Pancreatic Cancer. . . . .	398
<i>Kaiyin Zhou, Xinzhi Yao, Shuguang Wang, Jin-Dong Kim, Kevin Bretonnel Cohen, Ruiying Chen, Yuxing Wang, and Jingbo Xia</i>	

<b>Author Index . . . . .</b>	<b>411</b>
-------------------------------	------------