NLPTEA 2016

The 3rd Workshop on Natural Language Processing Techniques for Educational Applications

Proceedings of the NLPTEA 2016 Workshop

December 12, 2016
Osaka, Japan
Copyright of each paper stays with the respective authors (or their employers).

ISBN978-4-87974-717-4
Preface

Welcome to the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA 2016), with a Shared Task on Chinese Grammatical Error Diagnosis (CGED).

The development of Natural Language Processing (NLP) has advanced to a level that affects the research landscape of many academic domains and has practical applications in many industrial sectors. On the other hand, educational environment has also been improved to impact the world society, such as the emergence of MOOCs (Massive Open Online Courses). With these trends, this workshop focuses on the NLP techniques applied to the educational environment. Research issues in this direction have gained more and more attention, examples including the activities like the workshops on Innovative Use of NLP for Building Educational Applications since 2005 and educational data mining conferences since 2008.

This is the third workshop held in the Asian area, with the first one NLPTEA 2014 workshop being held in conjunction with the 22nd International Conference on Computer in Education (ICCE 2014) from Nov. 30 to Dec. 4, 2014 in Japan. The second edition NLPTEA 2015 workshop was held in conjunction with the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015) from July 26-31 in Beijing, China. This year, we continue to promote this research line by holding the workshop in conjunction with the COLING 2016 conference and also holding the third shared task on Chinese Grammatical Error Diagnosis. We receive 14 valid submissions for research issues, each of which was reviewed by three experts, and have 15 teams participating in the shared task, with 9 of them submitting their testing results. In total, there are 10 oral papers and 10 posters accepted. We also organize a keynote speech in this workshop. The invited speaker Professor Glenn Stockwell is expected to deliver a great talk entitled as "Technology and the Changing Face of Language Education".

We would like to thank the program committee members for their hard work in completing the review tasks. Their collective efforts achieved quality reviews of the submissions within a few weeks. Great thanks should also go to the speaker, authors, and participants for the tremendous supports in making the workshop a success.

Welcome you to the Osaka city, and wish you enjoy the city as well as the workshop.

Workshop Chairs
Hsin-Hsi Chen, National Taiwan University
Yuen-Hsien Tseng, National Taiwan Normal University
Vincent Ng, The University of Texas at Dallas
Xiaofei Lu, The Pennsylvania State University
Organization

Workshop Organizers

Hsin-Hsi Chen, National Taiwan University
Yuen-Hsien Tseng, National Taiwan Normal University
Vincent Ng, The University of Texas at Dallas
Xiaofei Lu, The Pennsylvania State University

Shared Task Organizers

Lung-Hao Lee, National Taiwan Normal University
Gaoqi Rao, Beijing Language and Culture University
Liang-Chih Yu, Yuan Ze University
Endong Xun, Beijing Language and Culture University
Baolin Zhang, Beijing Language and Culture University
Li-Ping Chang, National Taiwan Normal University

Program Committee

Yuki Arase, Osaka University
Rafael E. Banchs, Institute of Infocomm Research
Chris Brockett, Microsoft Research
Tao Chen, National University of Singapore
Barbara Di Eugenio, University of Illinois at Chicago
Vidas Daudaravicius, VTex Solutions for Science Publishing
Mariano Felice, Cambridge University
Cyril Goutte, National Research Council Canada
Na-Rae Han, University of Pittsburgh
Trude Heift, Simon Fraser University
Mamoru Komachi, Tokyo Metropolitan University
John Lee, City University of Hong Kong
Chen Li, Microsoft
Chuan-Jie Lin, National Taiwan Ocean University
Shervin Malmasi, Harvard University
Tomoya Mizumoto, Tohoku University
Courtney Napoles, John Hopkins University
Arti Ramesh, University of Maryland
Alla Rozovskaya, Virginia Tech
Mathias Schulze, University of Waterloo
Yukio Tono, Tokyo University of Foreign Studies
Elena Volodina, University of Gothenburg
Shih-Hung Wu, Chaoyang University of Technology
Jiajin Xu, Beijing Foreign Studies University
Jui-Feng Yeh, National Chiayi University
Marcos Zampieri, Saarland University
Trosten Zesch, University of Duisburg-Essen
# Table of Contents

*Simplification of Example Sentences for Learners of Japanese Functional Expressions*
Jun Liu and Yuji Matsumoto ................................................................. 1

*Effectiveness of Linguistic and Learner Features to Listenability Measurement Using a Decision Tree Classifier*
Katsunori Kotani and Takehiko Yoshimi ................................................. 6

*A Two-Phase Approach Towards Identifying Argument Structure in Natural Language*
Arkanath Pathak, Pawan Goyal and Plaban Bhowmick ................................ 11

*Distributed Vector Representations for Unsupervised Automatic Short Answer Grading*
Oliver Adams, Shourya Roy and Raghuram Krishnapuram .......................... 20

*A Comparison of Word Embeddings for English and Cross-Lingual Chinese Word Sense Disambiguation*
Hong Jin Kang, Tao Chen, Muthu Kumar Chandrasekaran and Min-Yen Kan ................................................................. 30

*Overview of NLP-TEA 2016 Shared Task for Chinese Grammatical Error Diagnosis*
Lung-Hao Lee, Gaoqi RAO, Liang-Chih Yu, Endong XUN, Baolin Zhang and Li-Ping Chang . 40

*Chinese Grammatical Error Diagnosis with Long Short-Term Memory Networks*
Bo Zheng, Wanxiang Che, Jiang Guo and Ting Liu .................................. 49

*Automatic Grammatical Error Detection for Chinese based on Conditional Random Field*
Yajun Liu, Yingjie Han, Liyan Zhuo and Hongying Zan ......................... 57

*CYUT-III System at Chinese Grammatical Error Diagnosis Task*
CHEN PO-LIN, Shih-Hung Wu, Liang-Pu Chen and ping-che yang ........................ 63

*Word Order Sensitive Embedding Features/Conditional Random Field-based Chinese Grammatical Error Detection*
Wei-Chieh Chou, Chin-Kui Lin, Yuan-Fu Liao and Yih-Ru Wang ................... 73

*A Fluctuation Smoothing Approach for Unsupervised Automatic Short Answer Grading*
Shourya Roy, Sandipan Dandapat and Y. Narahari ................................ 82

*Japanese Lexical Simplification for Non-Native Speakers*
Muhaimin Hading, Yuji Matsumoto and Maki Sakamoto .............................. 92

*A Corpus-based Approach for Spanish-Chinese Language Learning*
Shuyuan Cao, Iria da Cunha and Mikel Iruskieta ...................................... 97

*Syntactic Well-Formedness Diagnosis and Error-Based Coaching in Computer Assisted Language Learning using Machine Translation*
Luís Morgado da Costa, Francis Bond and Xiaoling He .......................... 107

*An Aligned French-Chinese corpus of 10K segments from university educational material*
Ruslan Kalitvianski, Lingxiao Wang, Valérie Bellynck and Christian Boitet ................................. 117

*Analysis of Foreign Language Teaching Methods: An Automatic Readability Approach*
Nasser Zalmout, Hind Saddiki and Nizar Habash ..................................... 122

*Generating and Scoring Correction Candidates in Chinese Grammatical Error Diagnosis*
Shao-Heng Chen, Yu-Lin Tsai and Chuan-Jie Lin .................................. 131
Grammatical Error Detection Based on Machine Learning for Mandarin as Second Language Learning
   Jui-Feng Yeh, Tsung-Wei Hsu and Chan-Kun Yeh ......................................................... 140

Bi-LSTM Neural Networks for Chinese Grammatical Error Diagnosis
   Shen Huang and Houfeng WANG ................................................................. 148

Chinese Grammatical Error Diagnosis Using Single Word Embedding
   Jinnan Yang, Bo Peng, Jin Wang, Jixian Zhang and Xuejie Zhang ......................... 155
Workshop Program

December 12, 2016

09:00–09:10  Opening Ceremony

09:10–10:00  Keynote Speech

10:00–10:30  Coffee Break

10:30–12:00  Regular Paper Session
10:30–10:45  Simplification of Example Sentences for Learners of Japanese Functional Expressions
             Jun Liu and Yuji Matsumoto

10:45–11:00  Effectiveness of Linguistic and Learner Features to Listenability Measurement Using a Decision Tree Classifier
             Katsunori Kotani and Takehiko Yoshimi

11:00–11:20  A Two-Phase Approach Towards Identifying Argument Structure in Natural Language
             Arkanath Pathak, Pawan Goyal and Plaban Bhowmick

11:20–11:40  Distributed Vector Representations for Unsupervised Automatic Short Answer Grading
             Oliver Adams, Shourya Roy and Raghuram Krishnapuram

11:40–12:00  A Comparison of Word Embeddings for English and Cross-Lingual Chinese Word Sense Disambiguation
             Hong Jin Kang, Tao Chen, Muthu Kumar Chandrasekaran and Min-Yen Kan
December 12, 2016 (continued)

12:00–14:00  Luch

14:00–15:15  Shared Task Session

14:00–14:15  Overview of NLP-TEA 2016 Shared Task for Chinese Grammatical Error Diagnosis
Lung-Hao Lee, Gaoqi RAO, Liang-Chih Yu, Endong XUN, Baolin Zhang and Li-Ping Chang

14:15–14:30  Chinese Grammatical Error Diagnosis with Long Short-Term Memory Networks
Bo Zheng, Wanxiang Che, Jiang Guo and Ting Liu

14:30–14:45  Automatic Grammatical Error Detection for Chinese based on Conditional Random Field
Yajun Liu, Yingjie Han, Liyan Zhuo and Hongying Zan

14:45–15:00  CYUT-III System at Chinese Grammatical Error Diagnosis Task
CHEN PO-LIN, Shih-Hung Wu, Liang-Pu Chen and ping-che yang

15:00–15:15  Word Order Sensitive Embedding Features/Conditional Random Field-based Chinese Grammatical Error Detection
Wei-Chieh Chou, Chin-Kui Lin, Yuan-Fu Liao and Yih-Ru Wang

15:15–15:50  Coffee Break

15:50–16:50  Poster Session

A Fluctuation Smoothing Approach for Unsupervised Automatic Short Answer Grading
Shourya Roy, Sandipan Dandapat and Y. Narahari

Japanese Lexical Simplification for Non-Native Speakers
Muhaimin Hading, Yuji Matsumoto and Maki Sakamoto

A Corpus-based Approach for Spanish-Chinese Language Learning
Shuyuan Cao, Iria da Cunha and Mikel Iruskieta

Syntactic Well-Formedness Diagnosis and Error-Based Coaching in Computer Assisted Language Learning using Machine Translation
Luís Morgado da Costa, Francis Bond and Xiaoling He
December 12, 2016 (continued)

An Aligned French-Chinese corpus of 10K segments from university educational material
Ruslan Kalitvianski, Lingxiao Wang, Valérie Bellynck and Christian Boitet

Analysis of Foreign Language Teaching Methods: An Automatic Readability Approach
Nasser Zalmout, Hind Saddiki and Nizar Habash

Generating and Scoring Correction Candidates in Chinese Grammatical Error Diagnosis
Shao-Heng Chen, Yu-Lin Tsai and Chuan-Jie Lin

Grammatical Error Detection Based on Machine Learning for Mandarin as Second Language Learning
Jui-Feng Yeh, Tsung-Wei Hsu and Chan-Kun Yeh

Bi-LSTM Neural Networks for Chinese Grammatical Error Diagnosis
Shen Huang and Houfeng WANG

Chinese Grammatical Error Diagnosis Using Single Word Embedding
Jinnan Yang, Bo Peng, Jin Wang, Jixian Zhang and Xuejie Zhang

16:50–17:00 Closing Remarks