

NLPTEA 2017

**The 4th Workshop on Natural Language Processing
Techniques for Educational Applications**

Proceedings of the NLPTEA 2017 Workshop

December 1, 2017
Taipei, Taiwan

©2017 Asian Federation of Natural Language Processing

ISBN 978-1-948087-08-7

Preface

Welcome to the 4th Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA 2017), with a Shared Task on Chinese Spelling Check.

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 fourth 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. The third version NLPTEA 2016 workshop was held in conjunction with the 26th International Conference on Computational Linguistics (COLING 2016) from December 11- 16 in Osaka, Japan. This year, we continue to promote this research line by holding the workshop in conjunction with the IJCNLP 2017 conference and also holding the fourth shared task on Chinese Spelling Check. We receive 19 valid submissions for research issues, each of which was reviewed by at least two experts, and have 7 teams participating in the shared task, with 2 of them submitting their testing results. In total, there are 7 oral papers and 5 posters accepted. We also organize a keynote speech in this workshop. The invited speaker Professor Vincent Ng is expected to deliver a great talk entitled as "Towards Content-Based Essay Scoring".

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 Taipei city, and wish you enjoy the city as well as the workshop.

NLPTEA 2017 Workshop Chairs
Yuen-Hsien Tseng, National Taiwan Normal University
Hsin-Hsi Chen, National Taiwan University
Lung-Hao Lee, National Taiwan Normal University
Liang-Chih Yu, Yuan Ze University

Organization

Workshop Organizers:

Yuen-Hsien Tseng, National Taiwan Normal University
Hsin-Hsi Chen, National Taiwan University
Lung-Hao Lee, National Taiwan Normal University
Liang-Chih Yu, Yuan Ze University

Shared Task Organizers:

Gabriel Pui Cheong Fung, The Chinese University of Hong Kong
Jia Zhu, South China Normal University

Program Committee:

Yuki Arase, Osaka University
Chris Brockett, Microsoft Research
Christopher Bryant, Cambridge University
Dominique Brunato, National Council of Research, Italy
Lei Chen, Educational Testing Service
Mei-Hua Chen, Tunghai University
Tao Chen, Johns Hopkins University
Vidas Daudaravicius, VTeX Solutions for Science Publishing
Mariano Felice, Cambridge University
Cyril Goutte, National Research Council Canada
Homa B. Hashemi, University of Pittsburgh
Trude Heift, Simon Fraser University, Canada
Tomoyuki Kajiwara, Tokyo Metropolitan 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
Simon Ostermann, Saarland University
Arti Ramesh, SUNY - Binghamton University
Gaoqi Rao, Beijing Language and Culture University
Livy Real, IBM Research, Brazil
Alla Rozovskaya, The City University of New York
Elizabeth Salesky, Massachusetts Institute of Technology
Yukio Tono, Tokyo University of Foreign Studies
Elena Volodina, University of Gothenburg
Thuy Vu, University of California, Los Angeles
Shih-Hung Wu, Chaoyang University of Technology
Huichao Xue, Google
Jui-Feng Yeh, National Chiayi University
Zheng Yuan, Cambridge University
Marcos Zampieri, University of Wolverhampton
Torsten Zesch, University of Duisburg-Essen

Invited Talk: Towards Content-Based Essay Scoring

Vincent Ng

Professor of Computer Science
Human Language Technology Research Institute, University of Texas at Dallas

Abstract

State-of-the-art automated essay scoring engines such as E-rater do not grade essay content, focusing instead on providing diagnostic trait feedback on categories such as grammar, usage, mechanics, style and organization. Content-based essay scoring is very challenging: it requires an understanding of essay content and is beyond the reach of today's automated essay scoring technologies. As a result, content-dependent dimensions of essay quality are largely ignored in existing automated essay scoring research. In this talk, we describe our recent and ongoing efforts on content-based essay scoring, sharing the lessons we learned from automatically scoring two of the arguably most important content-dependent dimensions of persuasive essay quality, thesis clarity and argument persuasiveness.

Biography

Vincent Ng (Ph.D., Cornell) is a Professor in the Computer Science Department at the University of Texas at Dallas. He is also the director of the Machine Learning and Language Processing Laboratory in the Human Language Technology Research Institute at UT Dallas. He is currently an associate editor of the ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP) and an information officer of the ACL Special Interest Group on Chinese Language Processing (SIGHAN). Since 2009, he has become increasingly interested in an NLP application that brings about a number of under-studied but fascinating discourse-level problems, automated essay grading, where he has been focusing on modeling those facets of persuasive student essays that require an understanding of essay content, such as thesis clarity and argument persuasiveness.

Table of Contents

<i>NTUCLE: Developing a Corpus of Learner English to Provide Writing Support for Engineering Students</i> Roger Vivek Placidus Winder, Joseph MacKinnon, Shu Yun Li, Benedict Christopher Tzer Liang Lin, Carmel Lee Hah Heah, Luís Morgado da Costa, Takayuki Kuribayashi and Francis Bond	1
<i>Understanding Non-Native Writings: Can a Parser Help?</i> Jirka Hana and Barbora Hladka	12
<i>Carrier Sentence Selection for Fill-in-the-blank Items</i> Shu Jiang and John Lee	17
<i>Hindi Shabdmitra: A Wordnet based E-Learning Tool for Language Learning and Teaching</i> Hanumant Redkar, Sandhya Singh, Meenakshi Somasundaram, Dhara Gorasia, Malhar Kulkarni and Pushpak Bhattacharyya	23
<i>NLPTEA 2017 Shared Task – Chinese Spelling Check</i> Gabriel Fung, Maxime Debosschere, Dingmin Wang, Bo Li, Jia Zhu and Kam-Fai Wong	29
<i>Chinese Spelling Check based on N-gram and String Matching Algorithm</i> Jui-Feng Yeh, Li-Ting Chang, Chan-Yi Liu and Tsung-Wei Hsu	35
<i>N-gram Model for Chinese Grammatical Error Diagnosis</i> Jianbo Zhao, Hao Liu, Zuyi Bao, Xiaopeng Bai, Si Li and Zhiqing Lin	39
<i>The Influence of Spelling Errors on Content Scoring Performance</i> Andrea Horbach, Yuning Ding and Torsten Zesch	45
<i>Analyzing the Impact of Spelling Errors on POS-Tagging and Chunking in Learner English</i> Tomoya Mizumoto and Ryo Nagata	54
<i>Complex Word Identification: Challenges in Data Annotation and System Performance</i> Marcos Zampieri, Shervin Malmasi, Gustavo Paetzold and Lucia Specia	59
<i>Suggesting Sentences for ESL using Kernel Embeddings</i> Kent Shioda, Mamoru Komachi, Rue Ikeya and Daichi Mochihashi	64
<i>Event Timeline Generation from History Textbooks</i> Harsimran Bedi, Sangameshwar Patil, Swapnil Hingmire and Girish Palshikar	69

Workshop Program

Friday, December 1, 2017

09:30–09:40 **Opening Remarks**

09:40–10:30 *Towards Content-Based Essay Scoring*
Vincent Ng

10:30–11:00 **Coffee Break**

11:00–12:20 **Regular Paper Session**

11:00–11:20 *NTUCLE: Developing a Corpus of Learner English to Provide Writing Support for Engineering Students*

Roger Vivek Placidus Winder, Joseph MacKinnon, Shu Yun Li, Benedict Christopher Tzer Liang Lin, Carmel Lee Hah Heah, Luís Morgado da Costa, Takayuki Kuribayashi and Francis Bond

11:20–11:40 *Understanding Non-Native Writings: Can a Parser Help?*

Jirka Hana and Barbora Hladka

11:40–12:00 *Carrier Sentence Selection for Fill-in-the-blank Items*

Shu Jiang and John Lee

12:00–12:20 *Hindi Shabdmitra: A Wordnet based E-Learning Tool for Language Learning and Teaching*

Hanumant Redkar, Sandhya Singh, Meenakshi Somasundaram, Dhara Gorasia, Malhar Kulkarni and Pushpak Bhattacharyya

Friday, December 1, 2017 (continued)

12:20–14:00 Lunch

14:00–15:00 Shared Task Session

14:00–14:20 *NLPTEA 2017 Shared Task – Chinese Spelling Check*
Gabriel Fung, Maxime Debosschere, Dingmin Wang, Bo Li, Jia Zhu and Kam-Fai Wong

14:20–14:40 *Chinese Spelling Check based on N-gram and String Matching Algorithm*
Jui-Feng Yeh, Li-Ting Chang, Chan-Yi Liu and Tsung-Wei Hsu

14:40–15:00 *N-gram Model for Chinese Grammatical Error Diagnosis*
Jianbo Zhao, Hao Liu, Zuyi Bao, Xiaopeng Bai, Si Li and Zhiqing Lin

15:00–15:20 Coffee Break

15:20–15:50 Poster Session

The Influence of Spelling Errors on Content Scoring Performance
Andrea Horbach, Yuning Ding and Torsten Zesch

Analyzing the Impact of Spelling Errors on POS-Tagging and Chunking in Learner English
Tomoya Mizumoto and Ryo Nagata

Complex Word Identification: Challenges in Data Annotation and System Performance
Marcos Zampieri, Shervin Malmasi, Gustavo Paetzold and Lucia Specia

Suggesting Sentences for ESL using Kernel Embeddings
Kent Shioda, Mamoru Komachi, Rue Ikeya and Daichi Mochihashi

Event Timeline Generation from History Textbooks
Harsimran Bedi, Sangameshwar Patil, Swapnil Hingmire and Girish Palshikar

Friday, December 1, 2017 (continued)

15:50–16:00 Closing Remarks

