Twelfth Workshop on Innovative Use of NLP for Building Educational Applications 2017

Copenhagen, Denmark
8 September 2017

ISBN: 978-1-5108-4756-9

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571

Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2017) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2017)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006
Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Question Difficulty – How to Estimate Without Norming, How to Use for Automated Grading
Ulrike Pado 1

Combining CNNs and Pattern Matching for Question Interpretation in a Virtual Patient Dialogue System
Lifeng Jin, Michael White, Evan Jaffe, Laura Zimmerman and Douglas Danforth 11

Continuous fluency tracking and the challenges of varying text complexity
Beata Beigman Klebanov, Anastassia Loukina, John Sabatini and Tenaha O'Reilly 22

Auxiliary Objectives for Neural Error Detection Models
Marek Rei and Helen Yannakoudakis 33

Linked Data for Language-Learning Applications
Robyn Loughnane, Kate McCurdy, Peter Kolb and Stefan Selent 44

Predicting Specificity in Classroom Discussion
Luca Lugini and Diane Litman 52

A Report on the 2017 Native Language Identification Shared Task
Shervin Malmasi, Keelan Evanini, Aoife Cahill, Joel Tetreault, Robert Pugh, Christopher Hamill, Diane Napolitano and Yao Qian 62

Evaluation of Automatically Generated Pronoun Reference Questions
Arief Yudha Satria and Takenobu Tokunaga 76

Predicting Audience's Laughter During Presentations Using Convolutional Neural Network
Lei Chen and Chong Min Lee 86

Collecting fluency corrections for spoken learner English
Andrew Caines, Emma Flint and Paula Buttery 91

Exploring Relationships Between Writing & Broader Outcomes With Automated Writing Evaluation
Jill Burstein, Dan McCaffrey, Beata Beigman Klebanov and Guangming Ling 101

An Investigation into the Pedagogical Features of Documents
Emily Sheng, Prem Natarajan, Jonathan Gordon and Gully Burns 109

Combining Multiple Corpora for Readability Assessment for People with Cognitive Disabilities
Victoria Yaneva, Constantin Orasan, Richard Evans and Omid Rohanian 121

Automatic Extraction of High-Quality Example Sentences for Word Learning Using a Determinantal Point Process
Arseny Tolmachev and Sadao Kurohashi 133

Distractor Generation for Chinese Fill-in-the-blank Items
Shu Jiang and John Lee 143

An Error-Oriented Approach to Word Embedding Pre-Training
Youmna Farag, Marek Rei and Ted Briscoe 149

Investigating neural architectures for short answer scoring
Brian Riordan, Andrea Horbach, Aoife Cahill, Torsten Zesch and Chong Min Lee 159

Human and Automated CEFR-based Grading of Short Answers
Anaïs Tack, Thomas François, Sophie Roekhaut and Cédrick Fairon 169

GEC into the future: Where are we going and how do we get there?
Keisuke Sakaguchi, Courtney Napoles and Joel Tetreault 180

Detecting Off-topic Responses to Visual Prompts
Marek Rei 188

Combining Textual and Speech Features in the NLI Task Using State-of-the-Art Machine Learning Techniques
Pavel Ircing, Jan Svec, Zbynek Zajic, Barbora Hladka and Martin Holub 198

Native Language Identification Using a Mixture of Character and Word N-grams
Elham Mohammadi, Hadi Veisi and Hessam Amini 210

Ensemble Methods for Native Language Identification
Sophia Chan, Maryam Honari Jahromi, Benjamin Benetti, Aazim Lakhani and Alona Fyshe . . 217

Can string kernels pass the test of time in Native Language Identification?
Radu Tudor Ionescu and Marius Popescu 224

Neural Networks and Spelling Features for Native Language Identification
Johannes Bjerva, Gintare Grigonyte, Robert Östling and Barbara Plank 235

A study of N-gram and Embedding Representations for Native Language Identification
Sowmya Vajjala and Sagnik Banerjee 240

A Shallow Neural Network for Native Language Identification with Character N-grams
Yunita Sari, Muhammad Rifqi Fatchurrahman and Meisyarah Dwiastuti 249

Fewer features perform well at Native Language Identification task
Taraka Rama and Çağrı Çöltekin 255

Structured Generation of Technical Reading Lists
Jonathan Gordon, Stephen Aguilar, Emily Sheng and Gully Burns 261

Effects of Lexical Properties on Viewing Time per Word in Autistic and Neurotypical Readers
Sanja Štajner, Victoria Yaneva, Ruslan Mitkov and Simone Paolo Ponzetto 271

Transparent text quality assessment with convolutional neural networks
Robert Östling and Gintare Grigonyte 282

Artificial Error Generation with Machine Translation and Syntactic Patterns
Marek Rei, Mariano Felice, Zheng Yuan and Ted Briscoe 287

Modelling semantic acquisition in second language learning
Ekaterina Kochmar and Ekaterina Shutova 293

Multiple Choice Question Generation Utilizing An Ontology
Katherine Stasaski and Marti A. Hearst 303

Simplifying metaphorical language for young readers: A corpus study on news text
Magdalena Wolska and Yulia Clausen 313

Language Based Mapping of Science Assessment Items to Skills
Farah Nadeem and Mari Ostendorf 319

Connecting the Dots: Towards Human-Level Grammatical Error Correction
Shamil Chollampatt and Hwee Tou Ng 327

Question Generation for Language Learning: From ensuring texts are read to supporting learning
Maria Chinkina and Detmar Meurers 334

Systematically Adapting Machine Translation for Grammatical Error Correction
Courtney Napoles and Chris Callison-Burch 345

Fine-grained essay scoring of a complex writing task for native speakers
Andrea Horbach, Dirk Scholten-Akoun, Yuning Ding and Torsten Zesch 357

Exploring Optimal Voting in Native Language Identification
Cyril Goutte and Serge Léger 367

CIC-FBK Approach to Native Language Identification
Ilia Markov, Lingzhen Chen, Carlo Strapparava and Grigori Sidorov 374

The Power of Character N-grams in Native Language Identification
Artur Kulmizev, Bo Blankers, Johannes Bjerva, Malvina Nissim, Gertjan van Noord, Barbara Plank and Martijn Wieling 382

Classifier Stacking for Native Language Identification
Wen Li and Liang Zou 390

Native Language Identification on Text and Speech
Marcos Zampieri, Alina Maria Ciobanu and Liviu P. Dinu 398

Native Language Identification using Phonetic Algorithms
Charese Smiley and Sandra Kübler 405

A deep-learning based native-language classification by using a latent semantic analysis for the NLI Shared Task 2017
Yoo Rhee Oh, Hyung-Bae Jeon, Hwa Jeon Song, Yun-Kyung Lee, Jeon-Gue Park and Yun-Keun Lee 413

Fusion of Simple Models for Native Language Identification
Fabio Kepler, Ramón Astudillo and Alberto Abad 423

Stacked Sentence-Document Classifier Approach for Improving Native Language Identification
Andrea Cimino and Felice Dell'Orletta 430

Using Gaze to Predict Text Readability
Ana Valeria Gonzalez-Garduño and Anders Søgaard 438

Annotating Orthographic Target Hypotheses in a German L1 Learner Corpus
Ronja Laarmann-Quante, Katrin Ortmann, Anna Ehlert, Maurice Vogel and Stefanie Dipper . . 444

A Large Scale Quantitative Exploration of Modeling Strategies for Content Scoring
Nitin Madnani, Anastassia Loukina and Aoife Cahill 457