19th Workshop of the Australasian Language Technology Association (ALTA 2021)

Online
8-10 December 2021

ISBN: 978-1-7138-9160-4

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© (2021) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

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

Long Papers

Findings on Conversation Disentanglement 1
Rongxin Zhu, Jey Han Lau and Jianzhong Qi

An Approach to the Frugal Use of Human Annotators to Scale up Auto-coding for Text Classification Tasks . 12
Li'An Chen and Hanna Suominen

Curriculum Learning Effectively Improves Low Data VQA 22
Narjes Askarian, Ehsan Abbasnejad, Ingrid Zukerman, Wray Buntine and Gholamreza Haffari

Using Word Embeddings to Quantify Ethnic Stereotypes in 12 years of Spanish News 34
Danielly Sorato, Diana Zavala-Rojas and Maria del Carme Colominas Ventura

Multi-modal Intent Classification for Assistive Robots with Large-scale Naturalistic Datasets 47
Karun Varghese Mathew, Venkata S Aditya Tarigoppula and Lea Frermann

Harnessing Privileged Information for Hyperbole Detection 58
Rhys Biddle, Maciek Rybinski, Qian Li, Cecile Paris and Guandong Xu

Combining Shallow and Deep Representations for Text-Pair Classification 68
Vincent Nguyen, Sarvnaz Karimi and Zhenchang Xing

Phone Based Keyword Spotting for Transcribing Very Low Resource Languages 79
Eric Le Ferrand, Steven Bird and Laurent Besacier

Evaluation of Review Summaries via Question-Answering 87
Nannan Huang and Xiuzhen Zhang

Exploring Story Generation with Multi-task Objectives in Variational Autoencoders 97
Zhuohan Xie, Jey Han Lau and Trevor Cohn

Principled Analysis of Energy Discourse across Domains with Thesaurus-based Automatic Topic Labeling . . 107
Thomas Scelsi, Alfonso Martinez Arranz and Lea Frermann

Robustness Analysis of Grover for Machine-Generated News Detection 119
Rinaldo Gagiano, Maria Myung-Hee Kim, Xiuzhen Zhang and Jennifer Biggs

Document Level Hierarchical Transformer 128
Najam Zaidi, Trevor Cohn and Gholamreza Haffari

Exploring the Vulnerability of Natural Language Processing Models via Universal Adversarial Texts 138
Xinzhe Li, Ming Liu, Xingjun Ma and Longxiang Gao

Generating and Modifying Natural Language Explanations 149
Abdus Salam, Rolf Schwitter and Mehmet Orgun

Short Papers

Does QA-based intermediate training help fine-tuning language models for text classification? 158
Shiwei Zhang and Xiuzhen Zhang

Retrodiction as Delayed Recurrence: the Case of Adjectives in Italian and English 163
Raquel G. Alhama, Francesca Zermiani and Atiqah Khaliq

Automatic Post-Editing for Vietnamese 169
Thanh Vu and Dai Quoc Nguyen

Using Discourse Structure to Differentiate Focus Entities from Background Entities in Scientific Literature . . 174
Antonio Jimeno Yepes, Ameer Albahem and Karin Verspoor

Evaluating Hierarchical Document Categorisation 179
Qian Sun, Aili Shen, Hiyori Yoshikawa, Chunpeng Ma, Daniel Beck, Tomoya Iwakura and Timothy Baldwin

BERT's The Word : Sarcasm Target Detection using BERT 185
Pradeesh Parameswaran, Andrew Trotman, Veronica Liesaputra and David Eyers

Cross-Domain Language Modeling: An Empirical Investigation 192
Vincent Nguyen, Sarvnaz Karimi, Maciej Rybinski and Zhenchang Xing

Shared Task Papers

Overview of the 2021 ALTA Shared Task: Automatic Grading of Evidence, 10 years later 201
Diego Mollá

Quick, get me a Dr. BERT: Automatic Grading of Evidence using Transfer Learning 205
Pradeesh Parameswaran, Andrew Trotman, Veronica Liesaputra and David Eyers

An Ensemble Model for Automatic Grading of Evidence 213
Yuting Guo, Yao Ge, Ruqi Liao and Abeed Sarker

Handling Variance of Pretrained Language Models in Grading Evidence in the Medical Literature 218
Fajri Koto and Biaoyan Fang