59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021)

Student Research Workshop

Online
5-6 August 2021

ISBN: 978-1-7138-3311-6

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
and the Asian Federation of Natural Language Processing
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

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

Investigation on Data Adaptation Techniques for Neural Named Entity Recognition
Evgeniia Tokarchuk, David Thulke, Weiyue Wang, Christian Dugast and Hermann Ney 1

Stage-wise Fine-tuning for Graph-to-Text Generation
Qingyun Wang, Semih Yavuz, Xi Victoria Lin, Heng Ji and Nazneen Rajani 16

Transformer-Based Direct Hidden Markov Model for Machine Translation
Weiyue Wang, Zijian Yang, Yingbo Gao and Hermann Ney 23

AutoRC: Improving BERT Based Relation Classification Models via Architecture Search
Wei Zhu 33

How Low is Too Low? A Computational Perspective on Extremely Low-Resource Languages
Rachit Bansal, Himanshu Choudhary, Ravneet Punia, Niko Schenk, Émilie Pagé-Perron and Jacob Dahl 44

On the Relationship between Zipf's Law of Abbreviation and Interfering Noise in Emergent Languages
Ryo Ueda and Koki Washio 60

Long Document Summarization in a Low Resource Setting using Pretrained Language Models
Ahsaas Bajaj, Pavitra Dangati, Kalpesh Krishna, Pradhiksha Ashok Kumar, Rheeya Uppaal, Bradford Windsor, Eliot Brenner, Dominic Dotterrer, Rajarshi Das and Andrew McCallum 71

Attending Self-Attention: A Case Study of Visually Grounded Supervision in Vision-and-Language Transformers
Jules Samaran, Noa Garcia, Mayu Otani, Chenhui Chu and Yuta Nakashima 81

Video-guided Machine Translation with Spatial Hierarchical Attention Network
Weiqi Gu, Haiyue Song, Chenhui Chu and Sadao Kurohashi 87

Stylistic approaches to predicting Reddit popularity in diglossia
Huikai Chua 93

"I've Seen Things You People Wouldn't Believe": Hallucinating Entities in GuessWhat?!
Alberto Testoni and Raffaella Bernardi 101

How do different factors Impact the Inter-language Similarity? A Case Study on Indian languages
Sourav Kumar, Salil Aggarwal, Dipti Misra Sharma and Radhika Mamidi 112

COVID-19 and Misinformation: A Large-Scale Lexical Analysis on Twitter
Dimosthenis Antypas, Jose Camacho-Collados, Alun Preece and David Rogers 119

Situation-Based Multiparticipant Chat Summarization: a Concept, an Exploration-Annotation Tool and an Example Collection
Anna Smirnova, Evgeniy Slobodkin and George Chernishev 127

Modeling Text using the Continuous Space Topic Model with Pre-Trained Word Embeddings
Seiichi Inoue, Taichi Aida, Mamoru Komachi and Manabu Asai 138

Semantics of the Unwritten: The Effect of End of Paragraph and Sequence Tokens on Text Generation with GPT2
He Bai, Peng Shi, Jimmy Lin, Luchen Tan, Kun Xiong, Wen Gao, Jie Liu and Ming Li 148

Data Augmentation with Unsupervised Machine Translation Improves the Structural Similarity of Cross-lingual Word Embeddings
Sosuke Nishikawa, Ryokan Ri and Yoshimasa Tsuruoka 163

Joint Detection and Coreference Resolution of Entities and Events with Document-level Context Aggregation
Samuel Kriman and Heng Ji 174

"Hold on honey, men at work": A semi-supervised approach to detecting sexism in sitcoms
Smriti Singh, Tanvi Anand, Arijit Ghosh Chowdhury and Zeerak Waseem 180

Observing the Learning Curve of NMT Systems With Regard to Linguistic Phenomena
Patrick Stadler, Vivien Macketanz and Eleftherios Avramidis 186

Improving the Robustness of QA Models to Challenge Sets with Variational Question-Answer Pair Generation
Kazutoshi Shinoda, Saku Sugawara and Akiko Aizawa 197

Tools Impact on the Quality of Annotations for Chat Untangling
Jhonny Cerezo, Felipe Bravo-Marquez and Alexandre Henri Bergel 215

How Many Layers and Why? An Analysis of the Model Depth in Transformers
Antoine Simoulin and Benoit Crabbé 221

Edit Distance Based Curriculum Learning for Paraphrase Generation
Sora Kadotani, Tomoyuki Kajiwara, Yuki Arase and Makoto Onizuka 229

Changing the Basis of Contextual Representations with Explicit Semantics
Tamás Ficsor and Gábor Berend 235

Personal Bias in Prediction of Emotions Elicited by Textual Opinions
Piotr Milkowski, Marcin Gruza, Kamil Kanclerz, Przemyslaw Kazienko, Damian Grimling and Jan Kocon 248

MVP-BERT: Multi-Vocab Pre-training for Chinese BERT
Wei Zhu 260

CMTA: COVID-19 Misinformation Multilingual Analysis on Twitter
Raj Pranesh, Mehrdad Farokhenajd, Ambesh Shekhar and Genoveva Vargas-Solar 270

Predicting pragmatic discourse features in the language of adults with autism spectrum disorder
Christine Yang, Duanchen Liu, Qingyun Yang, Zoey Liu and Emily Prud'hommeaux 284

SumPubMed: Summarization Dataset of PubMed Scientific Articles
Vivek Gupta, Prerna Bharti, Pegah Nokhiz and Harish Karnick 292

A Case Study of Analysis of Construals in Language on Social Media Surrounding a Crisis Event
Lolo Aboufoul, Khyati Mahajan, Tiffany Gallicano, Sara Levens and Samira Shaikh 304

Cross-lingual Evidence Improves Monolingual Fake News Detection
Daryna Dementieva and Alexander Panchenko 310

Neural Machine Translation with Synchronous Latent Phrase Structure
Shintaro Harada and Taro Watanabe 321

Zero Pronouns Identification based on Span prediction
Sei Iwata, Taro Watanabe and Masaaki Nagata 331

On the differences between BERT and MT encoder spaces and how to address them in translation tasks
Raúl Vázquez, Hande Celikkanat, Mathias Creutz and Jörg Tiedemann 337

Synchronous Syntactic Attention for Transformer Neural Machine Translation
Hiroyuki Deguchi, Akihiro Tamura and Takashi Ninomiya 348