

28th International Conference on Computational Linguistics (COLING 2020)

Industry Track

Held online due to COVID-19

Barcelona, Spain
12 December 2020

ISBN: 978-1-7138-2522-7

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© (2020) by the Association for Computational Linguistics
All rights reserved.

Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, (<https://creativecommons.org/licenses/by-nd/4.0/>)

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

<i>Evaluating Cross-Lingual Transfer Learning Approaches in Multilingual Conversational Agent Models</i> Lizhen Tan and Olga Golovneva	1
<i>Data-Efficient Paraphrase Generation to Bootstrap Intent Classification and Slot Labeling for New Features in Task-Oriented Dialog Systems</i> Shailza Jolly, Tobias Falke, Caglar Tirkaz and Daniil Sorokin	10
<i>Leveraging User Paraphrasing Behavior In Dialog Systems To Automatically Collect Annotations For Long-Tail Utterances</i> Tobias Falke, Markus Boese, Daniil Sorokin, Caglar Tirkaz and Patrick Lehnen	21
<i>Query Distillation: BERT-based Distillation for Ensemble Ranking</i> Wangshu Zhang, Junhong Liu, Zujie Wen, Yafang Wang and Gerard de Melo	33
<i>Semantic Diversity for Natural Language Understanding Evaluation in Dialog Systems</i> Enrico Palumbo, Andrea Mezzalana, Cristina Marco, Alessandro Manzotti and Daniele Amberti	44
<i>An Empirical Study on Multi-Task Learning for Text Style Transfer and Paraphrase Generation</i> Pawel Bujnowski, Kseniia Ryzhova, Hyungtak Choi, Katarzyna Witkowska, Jaroslaw Piersa, Tymoteusz Krumholz and Katarzyna Beksa	50
<i>Best Practices for Data-Efficient Modeling in NLG: How to Train Production-Ready Neural Models with Less Data</i> Ankit Arun, Soumya Batra, Vikas Bhardwaj, Ashwini Challa, Pinar Donmez, Peyman Heidari, Hakan Inan, Shashank Jain, Anuj Kumar, Shawn Mei, Karthik Mohan and Michael White	64
<i>Interactive Question Clarification in Dialogue via Reinforcement Learning</i> Xiang Hu, Zujie Wen, Yafang Wang, Xiaolong Li and Gerard de Melo	78
<i>Towards building a Robust Industry-scale Question Answering System</i> Rishav Chakravarti, Anthony Ferritto, Bhavani Iyer, Lin Pan, Radu Florian, Salim Roukos and Avi Sil	90
<i>Delexicalized Paraphrase Generation</i> Boya Yu, Konstantine Arkoudas and Wael Hamza	102
<i>Multi-task Learning of Spoken Language Understanding by Integrating N-Best Hypotheses with Hierarchical Attention</i> Mingda Li, Xinyue Liu, Weitong Ruan, Luca Soldaini, Wael Hamza and Chengwei Su	113
<i>Misspelling Detection from Noisy Product Images</i> Varun Nagaraj Rao and Mingwei Shen	124
<i>hinglishNorm - A Corpus of Hindi-English Code Mixed Sentences for Text Normalization</i> Piyush Makhija, Ankit Kumar and Anuj Gupta	136
<i>Assessing Social License to Operate from the Public Discourse on Social Media</i> Chang Xu, Cecile Paris, Ross Sparks, Surya Nepal and Keith VanderLinden	146
<i>Extreme Model Compression for On-device Natural Language Understanding</i> Kanthashree Mysore Sathyendra, Samridhi Choudhary and Leah Nicolich-Henkin	160

<i>Scalable Cross-lingual Treebank Synthesis for Improved Production Dependency Parsers</i> Yousef El-Kurdi, Hiroshi Kanayama, Efsun Sarioglu Kayi, Vittorio Castelli, Todd Ward and Radu Florian	172
<i>An Industry Evaluation of Embedding-based Entity Alignment</i> Ziheng Zhang, Hualuo Liu, Jiaoyan Chen, Xi Chen, Bo Liu, YueJia Xiang and Yefeng Zheng .	179
<i>Learning Domain Terms - Empirical Methods to Enhance Enterprise Text Analytics Performance</i> Gargi Roy, Lipika Dey, Mohammad Shakir and Tirthankar Dasgupta.....	190
<i>Model-agnostic Methods for Text Classification with Inherent Noise</i> Kshitij Tayal, Rahul Ghosh and Vipin Kumar	202
<i>ScopeIt: Scoping Task Relevant Sentences in Documents</i> Barun Patra, Vishwas Suryanarayanan, Chala Fufa, Pamela Bhattacharya and Charles Lee	214
<i>Uncertainty Modeling for Machine Comprehension Systems using Efficient Bayesian Neural Networks</i> Zhengyuan Liu, Pavitra Krishnaswamy, Ai Ti Aw and Nancy Chen	228
<i>Regularized Graph Convolutional Networks for Short Text Classification</i> Kshitij Tayal, Nikhil Rao, Saurabh Agarwal, Xiaowei Jia, Karthik Subbian and Vipin Kumar ..	236