Findings

Findings of the Association for Computational Linguistics

Findings of ACL: EMNLP 2020

16 – 20 November, 2020
Preface by the Program Committee Co-Chairs

Welcome to the very first volume of *Findings of ACL*, a new publication designed to operate as a “companion” to conferences in the ACL stable. This first volume, *Findings of ACL: EMNLP 2020* comprises papers selected from those submitted to EMNLP 2020 which were not selected to appear at the main conference. Despite this, the Programme Committee recognised the value of these submissions, and deemed them to be strong work, worthy of publication.

Papers published in *Findings* count as full publications, however they are not assigned a presentation slot in the conference, but rather are published purely online in a separate volume in the ACL Anthology. There are a number of motivations for this new publication, from allowing timely work to be published quickly, to being more accepting of solid work, and helping to manage the increasing reviewing burden on the community.

This new publication option allowed us to accept many more worthy papers that would have otherwise been rejected from EMNLP. From the 3,359 submissions reviewed for EMNLP, a total of 752 were accepted to the main conference, and a further 520 papers were accepted to *Findings*. The papers accepted to *Findings* equates to 15.5% of the reviewed papers, giving an aggregate acceptance rate for EMNLP and *Findings* of 37.9%. We requested that authors withdraw their *Findings* papers if they preferred to resubmit their work to subsequent conferences. After withdrawals, 447 papers remained, of which 332 were long papers, and 115 short.

Papers were accepted to *Findings* based on their reviews and AC assessments, particularly whether reviewers found merit in the ideas and approaches, and found no fundamental issues in terms of the work’s motivation, theory, experimentation, analysis and ethical considerations. Where there was broad agreement of the robustness of the work, but the papers would otherwise have been rejected from EMNLP, instead these papers were accepted into *Findings*. Judgements as to the suitability of a paper to *Findings* was decided by the Senior Areas Chairs (SACs) and Programme Chairs (PCs), initially through SACs proposing a large suite of borderline papers and a cutoff for acceptance to the main conference. Final decisions were made for the main conference before considering papers for *Findings*. We are confident that this procedure did not disadvantage particular papers, or classes of papers. Instead it gave authors a publication option for papers that would otherwise have been rejected, and based on the fact that 86% of papers accepted to *Findings* were not withdrawn, there is clear community support for the endeavour.

We appreciate that *Findings* may have an effect on the EMNLP workshops through removing from circulation some papers that would have been submitted to these venues. For this reason we have provided a mechanism for authors of *Findings* papers to present their work at a workshop as a non-archival presentation. This was done as a match-making service between authors of *Findings* papers and workshops, resulting in 139 *Findings* papers being assigned workshop presentations.

The reviewing process for *Findings* is largely the same as for the main conference, and accordingly we wish to thank all involved in EMNLP 2020 for their efforts, as detailed in the *Preface to the Proceedings of EMNLP 2020*. We would like to specifically thank:

- Tim Baldwin, the architect of *Findings*, and all the members of the ACL Reviewing Committee who helped to refine the proposal;
- SACs for making paper recommendation decisions for *Findings*;
- Jing Li and Lemao Liu, the Publication Chairs responsible for *Findings* papers;
- The workshop chairs, Jackie Cheung and Lonneke van der Plas, and the coordinators of the many workshops, who helped in providing a venue for *Findings* authors to present their work;
• The many participants in the ACL 2020 Reviewing meeting, and others who provided their constructive feedback on this publication.

We hope that *Findings* will continue to serve as a companion to future conferences, and become an important venue for excellent, widely-read, and highly cited work in NLP.

EMNLP 2020 Program Co-Chairs

*Trevor Cohn*, University of Melbourne, Australia
*Yulan He*, University of Warwick, UK
*Yang Liu*, Amazon – Alexa AI, USA
# Table of Contents

**Fully Quantized Transformer for Machine Translation**  
Gabriele Prato, Ella Charlaix and Mehdi Rezagholizadeh ................................................................. 1

**Summarizing Chinese Medical Answer with Graph Convolution Networks and Question-focused Dual Attention**  
Ningyu Zhang, Shumin Deng, Juan Li, Xi Chen, Wei Zhang and Huajun Chen ............................... 15

**Stay Hungry, Stay Focused: Generating Informative and Specific Questions in Information-Seeking Conversations**  
Peng Qi, Yu Hao Zhang and Christopher D. Manning ................................................................. 25

**Adapting BERT for Word Sense Disambiguation with Gloss Selection Objective and Example Sentences**  
Boon Peng Yap, Andrew Koh and Eng Siong Chng ................................................................. 41

**Adversarial Text Generation via Sequence Contrast Discrimination**  
Ke Wang and Xiaojun Wan ................................................................. 47

**GRACE: Gradient Harmonized and Cascaded Labeling for Aspect-based Sentiment Analysis**  
Huaishao Luo, Lei Ji, Tianrui Li, Daxin Jiang and Nan Duan ................................................................. 54

**Reducing Sentiment Bias in Language Models via Counterfactual Evaluation**  
Po-Sen Huang, Huan Zhang, Ray Jiang, Robert Stanforth, Johannes Welbl, Jack Rae, Vishal Maini, Dani Yogatama and Pushmeet Kohli ................................................................. 65

**Improving Text Understanding via Deep Syntax-Semantics Communication**  
Hao Fei, Yafeng Ren and Donghong Ji ................................................................. 84

**GRUEN for Evaluating Linguistic Quality of Generated Text**  
Wanzheng Zhu and Suma Bhat ................................................................. 94

**A Greedy Bit-flip Training Algorithm for Binarized Knowledge Graph Embeddings**  
Katsuhiko Hayashi, Koki Kishimoto and Masashi Shimbo ................................................................. 109

**Difference-aware Knowledge Selection for Knowledge-grounded Conversation Generation**  
Chujie Zheng, Yunbo Cao, Daxin Jiang and Minlie Huang ................................................................. 115

**An Attentive Recurrent Model for Incremental Prediction of Sentence-final Verbs**  
Wenyan Li, Alvin Grissom II and Jordan Boyd-Graber ................................................................. 126

**Transformer-GCRF: Recovering Chinese Dropped Pronouns with General Conditional Random Fields**  
Jingxuan Yang, Kerui Xu, Jun Xu, Si Li, Sheng GAO, Jun Guo, Ji-Rong Wen and Nianwen Xue ................................................................. 137

**Neural Speed Reading Audited**  
Anders Søgaard ................................................................. 148

**Converting the Point of View of Messages Spoken to Virtual Assistants**  
Gunhee Lee, Vera Zu, Sai Srujana Buddi, Dennis Liang, Purva Kulkarni and Jack FitzGerald ................................. 154

**Robustness to Modification with Shared Words in Paraphrase Identification**  
Zhouxing Shi and Minlie Huang ................................................................. 164
Few-shot Natural Language Generation for Task-Oriented Dialog
Baolin Peng, Chenguang Zhu, Chunyuan Li, Xiujun Li, Jinchao Li, Michael Zeng and Jianfeng Gao

Mimic and Conquer: Heterogeneous Tree Structure Distillation for Syntactic NLP
Hao Fei, Yafeng Ren and Donghong Ji

A Hierarchical Network for Abstractive Meeting Summarization with Cross-Domain Pretraining
Chenguang Zhu, Ruochen Xu, Michael Zeng and Xuedong Huang

Active Testing: An Unbiased Evaluation Method for Distantly Supervised Relation Extraction
Pengshuai Li, Xinsong Zhang, Weijia Jia and Wei Zhao

Semantic Matching for Sequence-to-Sequence Learning
Ruiyi Zhang, Changyou Chen, Xinyuan Zhang, Ke Bai and Lawrence Carin

How Decoding Strategies Affect the Verifiability of Generated Text
Luca Massarelli, Fabio Petroni, Aleksandra Piktus, Myle Ott, Tim Rocktäschel, Vassilis Plachouras, Fabrizio Silvestri and Sebastian Riedel

Minimize Exposure Bias of Seq2Seq Models in Joint Entity and Relation Extraction
Ranran Haoran Zhang, Qianying Liu, Aysa Xuemo Fan, Heng Ji, Daojian Zeng, Fei Cheng, Daisuke Kawahara and Sadao Kurohashi

Gradient-based Analysis of NLP Models is Manipulable
Junlin Wang, Jens Tuyls, Eric Wallace and Sameer Singh

Pretrain-KGE: Learning Knowledge Representation from Pretrained Language Models
Zhiyuan Zhang, Xiaojian Liu, Yi Zhang, Qi Su, Xu SUN and Bin He

A Self-Refinement Strategy for Noise Reduction in Grammatical Error Correction
Masato Mita, Shun Kiyono, Masahiro Kaneko, Jun Suzuki and Kentaro Inui

Understanding tables with intermediate pre-training
Julian Eisenschlos, Syrine Krichene and Thomas Müller

Enhance Robustness of Sequence Labelling with Masked Adversarial Training
Luoxin Chen, Xinyue Liu, Weitong Ruan and Jianhua Lu

Multilingual Argument Mining: Datasets and Analysis
Orith Toledo-Ronen, Matan Orbach, Yonatan Bilu, Artem Spector and Noam Slonim

Improving Grammatical Error Correction with Machine Translation Pairs
Wangchunshu Zhou, Tao Ge, Chang Mu, Ke Xu, Furu Wei and Ming Zhou

Machines Getting with the Program: Understanding Intent Arguments of Non-Canonical Directives
Won Ik Cho, Youngki Moon, Sangwhan Moon, Seok Min Kim and Nam Soo Kim

The RELX Dataset and Matching the Multilingual Blanks for Cross-Lingual Relation Classification
Abdullatif Köksal and Arzucan Özgür

Control, Generate, Augment: A Scalable Framework for Multi-Attribute Text Generation
Giuseppe Russo, Nora Hollenstein, Claudiu Cristian Musat and Ce Zhang

Open-Ended Visual Question Answering by Multi-Modal Domain Adaptation
Yiming Xu, Lin Chen, Zhongwei Cheng, Lixin Duan and Jiebo Luo
Interpretable Entity Representations through Large-Scale Typing
Yasumasa Onoe and Greg Durrett .......................................................... 612

Empirical Studies of Institutional Federated Learning For Natural Language Processing
Xinghua Zhu, Jianzong Wang, Zhenhou Hong and Jing Xiao ........................ 625

NeuReduce: Reducing Mixed Boolean-Arithmetic Expressions by Recurrent Neural Network
Weijie Feng, Binbin Liu, Dongpeng Xu, Qilong Zheng and Yun Xu .......................... 635

From Language to Language-ish: How Brain-Like is an LSTM’s Representation of Nonsensical Language Stimuli?
Maryam Hashemzadeh, Greta Kaufeld, Martha White, Andrea E. Martin and Alona Fyshe .... 645

Revisiting Pre-Trained Models for Chinese Natural Language Processing
Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Shijin Wang and Guoping Hu ............... 657

Cascaded Semantic and Positional Self-Attention Network for Document Classification
Juyong Jiang, Jie Zhang and Kai Zhang ........................................................... 669

Toward Recognizing More Entity Types in NER: An Efficient Implementation using Only Entity Lexicons
Minlong Peng, Ruotian Ma, Qi Zhang, Lujun Zhao, Mengxi Wei, Changlong Sun and Xuanjing Huang .............................................................................. 678

From Disjoint Sets to Parallel Data to Train Seq2Seq Models for Sentiment Transfer
Paulo Cavalin, Marisa Vasconcelos, Marcelo Grave, Claudio Pinhanez and Victor Henrique Alves Ribeiro ........................................................................... 689

Jiannan Xiang, Xin Wang and William Yang Wang ............................................. 699

Document Ranking with a Pretrained Sequence-to-Sequence Model
Rodrigo Nogueira, Zhiying Jiang, Ronak Pradeep and Jimmy Lin ............................ 708

Pruning Redundant Mappings in Transformer Models via Spectral-Normalized Identity Prior
Zi Lin, Jeremiah Liu, Zi Yang, Nan Hua and Dan Roth ........................................ 719

Rethinking Self-Attention: Towards Interpretability in Neural Parsing
Khalil Mrini, Franck Dernoncourt, Quan Hung Tran, Trung Bui, Walter Chang and Ndapa Nakashole ................................................................................. 731

PolicyQA: A Reading Comprehension Dataset for Privacy Policies
Wasi Ahmad, Jianfeng Chi, Yuan Tian and Kai-Wei Chang .................................. 743

A Linguistic Analysis of Visually Grounded Dialogues Based on Spatial Expressions
Takuma Udagawa, Takato Yamazaki and Akiko Aizawa ........................................ 750

Efficient Context and Schema Fusion Networks for Multi-Domain Dialogue State Tracking
Su Zhu, Jieyu Li, Lu Chen and Kai Yu .................................................................. 766

Syntactic and Semantic-driven Learning for Open Information Extraction
Jialong Tang, Yaojie Lu, Hongyu Lin, Xianpei Han, Le Sun, Xinyan Xiao and Hua Wu .... 782

Group-wise Contrastive Learning for Neural Dialogue Generation
Hengyi Cai, Hongshen Chen, Yonghao Song, zhuoye Ding, Yongjun Bao, Weipeng Yan and Xiaofang Zhao ................................................................. 793
E-BERT: Efficient-Yet-Effective Entity Embeddings for BERT
Nina Poerner, Ulli Waltinger and Hinrich Schütze .................................................. 803

A Multi-task Learning Framework for Opinion Triplet Extraction
Chen Zhang, Qiuchi Li, Dawei Song and benyou wang ........................................... 819

Event Extraction as Multi-turn Question Answering
Fayuan Li, Weihua Peng, Yuguang Chen, Quan Wang, Lu Pan, Yajuan Lyu and Yong Zhu .... 829

Improving QA Generalization by Concurrent Modeling of Multiple Biases
Mingzhu Wu, Nafise Sadat Moosavi, Andreas Rücklé and Iryna Gurevych ................. 839

Actor-Double-Critic: Incorporating Model-Based Critic for Task-Oriented Dialogue Systems
Yen-chen Wu, Bo-Hsiang Tseng and Milica Gasic ..................................................... 854

Controlled Hallucinations: Learning to Generate Faithfully from Noisy Data
Katja Filippova ........................................................................................................... 864

Sequential Span Classification with Neural Semi-Markov CRFs for Biomedical Abstracts
Kosuke Yamada, Tsutomu Hirao, Ryohei Sasano, Koichi Takeda and Masaaki Nagata .... 871

Where to Submit? Helping Researchers to Choose the Right Venue
Konstantin Kobs, Tobias Koopmann, Albin Zehe, David Fernes, Philipp Krop and Andreas Hotho

AirConcierge: Generating Task-Oriented Dialogue via Efficient Large-Scale Knowledge Retrieval
Chieh-Yang Chen, Pei-Hsin Wang, Shih-Chieh Chang, Da-Cheng Juan, Wei Wei and Jia-Yu Pan

DocStruct: A Multimodal Method to Extract Hierarchy Structure in Document for General Form Understanding
Zilong Wang, Mingjie Zhan, xuebo liu and ding liang ................................................. 898

Pretrained Language Models for Dialogue Generation with Multiple Input Sources
Yu Cao, Wei Bi, Meng Fang and Dacheng Tao ........................................................ 909

A Study in Improving BLEU Reference Coverage with Diverse Automatic Paraphrasing
Rachel Bawden, Biao Zhang, Lisa Yankovskaya, Andre Tätter and Matt Post ............. 918

Cross-lingual Alignment Methods for Multilingual BERT: A Comparative Study
Saurabh Kulshreshtha, Jose Luis Redondo Garcia and Ching-Yun Chang ...................... 933

Hybrid Emoji-Based Masked Language Models for Zero-Shot Abusive Language Detection
Michele Corazza, Stefano Menini, Elena Cabrio, Sara Tonelli and Serena Villata ........ 943

SeNsER: Learning Cross-Building Sensor Metadata Tagger
Yang Jiao, Jiacheng Li, Jianman Wu, Dezhi Hong, Rajesh Gupta and Jingbo Shang .... 950

Persian Ezafe Recognition Using Transformers and Its Role in Part-Of-Speech Tagging
Ehsan Doostmohammadi, Minoo Nassajian and Adel Rahimi .................................. 961

Scene Graph Modification Based on Natural Language Commands
Xuanli He, Quan Hung Tran, Gholamreza Haffari, Walter Chang, Zhe Lin, Trung Bui, Franck Dernoncourt and Nhan Dam ................................................................. 972
LiMiT: The Literal Motion in Text Dataset
Irene Manotas, Ngoc Phuoc An Vo and Vadim Sheinin ........................................ 991

Transition-based Parsing with Stack-Transformers
Ramón Fernandez Astudillo, Miguel Ballesteros, Tahira Naseem, Austin Blodgett and Radu Florian
1001

Generative Data Augmentation for Commonsense Reasoning
Yiben Yang, Chaitanya Malaviya, Jared Fernandez, Swabha Swayamdipta, Ronan Le Bras, Ji-Ping
Wang, Chandra Bhagavatula, Yejin Choi and Doug Downey ........................................ 1008

HybridQA: A Dataset of Multi-Hop Question Answering over Tabular and Textual Data
Wenhu Chen, Hanwen Zha, Zhiyu Chen, Wenhan Xiong, Hong Wang and William Yang Wang
1026

PhoBERT: Pre-trained language models for Vietnamese
Dat Quoc Nguyen and Anh Tuan Nguyen ................................................................. 1037

ESTeR: Combining Word Co-occurrences and Word Associations for Unsupervised Emotion Detection
Sujatha Das Gollapalli, Polina Rozenshtein and See-Kiong Ng .................................... 1043

Make Templates Smarter: A Template Based Data2Text System Powered by Text Stitch Model
Bingfeng Luo, Zuo Bai, Kunfeng Lai and Jianping Shen ............................................. 1057

GCDST: A Graph-based and Copy-augmented Multi-domain Dialogue State Tracking
Peng Wu, Bowei Zou, Ridong Jiang and AiTi Aw ....................................................... 1063

Incorporating Stylistic Lexical Preferences in Generative Language Models
Hrituraj Singh, Gaurav Verma and Balaji Vasan Srinivasan ........................................ 1074

Why do you think that? Exploring Faithful Sentence-Level Rationales Without Supervision
Max Glockner, Ivan Habernal and Iryna Gurevych ..................................................... 1080

Semi-Supervised Learning for Video Captioning
Ke Lin, Zhuoxin Gan and Liwei WANG ................................................................. 1096

Multi2OIE: Multilingual Open Information Extraction Based on Multi-Head Attention with BERT
Youngbin Ro, Yukyung Lee and Pilsung Kang ......................................................... 1107

LGPSolver - Solving Logic Grid Puzzles Automatically
Elgun Jabrayilzade and Selma Tekir ................................................................. 1118

Using the Past Knowledge to Improve Sentiment Classification
Qi Qin, Wenpeng Hu and Bing Liu ................................................................. 1124

High-order Semantic Role Labeling
Zuchao Li, Hai Zhao, Rui Wang and Kevin Parnow .................................................. 1134

Undersensitivity in Neural Reading Comprehension
Johannes Welbl, Pasquale Minervini, Max Bartolo, Pontus Stenetorp and Sebastian Riedel... 1152

HyperText: Endowing FastText with Hyperbolic Geometry
Yudong Zhu, Di Zhou, Jinghui Xiao, Xin Jiang, Xiao Chen and Qun Liu .................... 1166

AutoETER: Automated Entity Type Representation for Knowledge Graph Embedding
Guanglin Niu, Bo Li, Yongfei Zhang, Shiliang Pu and Jingyang Li ............................. 1172
Learning Robust and Multilingual Speech Representations
Kazuya Kawakami, Luyu Wang, Chris Dyer, Phil Blunsom and Aaron van den Oord . . . . . . . . 1182

FQuAD: French Question Answering Dataset
Martin D’Hoffschmidt, Wacim Belblidia, Quentin Heinrich, Tom Brendlé and Maxime Vidal 1193

Dynamic Semantic Matching and Aggregation Network for Few-shot Intent Detection
Hoang Nguyen, Chenwei Zhang, Congying Xia and Philip Yu ................................. 1209

Quantifying the Contextualization of Word Representations with Semantic Class Probing
Mengjie Zhao, Philipp Dufter, Yadollah Yaghoobzadeh and Hinrich Schütze ........................ 1219

Learning to Generate Clinically Coherent Chest X-Ray Reports
Justin Lovelace and Bobak Mortazavi ................................................................. 1235

FELIX: Flexible Text Editing Through Tagging and Insertion
Jonathan Mallinson, Aliaksei Severyn, Eric Malmi and Guillermo Garrido ...................... 1244

What Can We Do to Improve Peer Review in NLP?
Anna Rogers and Isabelle Augenstein ................................................................. 1256

Unsupervised Relation Extraction from Language Models using Constrained Cloze Completion
Ankur Goswami, Akshata Bhat, Hadar Ohana and Theodoros Rekatsinas ..................... 1263

Biomedical Event Extraction with Hierarchical Knowledge Graphs
Kung-Hsiang Huang, Mu Yang and Nanyun Peng .................................................. 1277

Language Generation via Combinatorial Constraint Satisfaction: A Tree Search Enhanced Monte-Carlo Approach
Maosen Zhang, Nan Jiang, Lei Li and Yexiang Xue .............................................. 1286

Examining the Ordering of Rhetorical Strategies in Persuasive Requests
Omar Shaikh, Jiaao Chen, Jon Saad-Falcon, Polo Chau and Diyi Yang ......................... 1299

Evaluating Models’ Local Decision Boundaries via Contrast Sets

Parsing with Multilingual BERT, a Small Corpus, and a Small Treebank
Ethan C. Chau, Lucy H. Lin and Noah A. Smith ..................................................... 1324

OptSLA: an Optimization-Based Approach for Sequential Label Aggregation
Nasim Sabetpour, Adithya Kulkarni and Qi Li ....................................................... 1335

Optimizing Word Segmentation for Downstream Task
Tatsuya Hiraoka, Sho Takase, Kei Uchiumi, Atsushi Keyaki and Naoaki Okazaki ........... 1341

Dynamically Updating Event Representations for Temporal Relation Classification with Multi-category Learning
Fei Cheng, Masayuki Asahara, Ichiro Kobayashi and Sadao Kurohashi ...................... 1352

A Compare Aggregate Transformer for Understanding Document-grounded Dialogue
Longxuan Ma, Wei-Nan Zhang, runxin sun and Ting Liu ........................................ 1358
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TextHide: Tackling Data Privacy in Language Understanding Tasks</td>
<td>Yangsibo Huang, Zhao Song, Danqi Chen, Kai Li and Sanjeev Arora</td>
<td>1368</td>
</tr>
<tr>
<td>Modeling Intra and Inter-modality Incongruity for Multi-Modal Sarcasm Detection</td>
<td>Hongliang Pan, Zheng Lin, Peng Fu, Yatao Qi and Weiping Wang</td>
<td>1383</td>
</tr>
<tr>
<td>Investigating Transferability in Pretrained Language Models</td>
<td>Alex Tamkin, Trisha Singh, Davide Giovanardi and Noah Goodman</td>
<td>1393</td>
</tr>
<tr>
<td>Improving Knowledge-Aware Dialogue Response Generation by Using Human-Written Prototype Dialogues</td>
<td>Sixing Wu, Ying Li, Dawei Zhang and Zhonghai Wu</td>
<td>1402</td>
</tr>
<tr>
<td>Filtering before Iteratively Referring for Knowledge-Grounded Response Selection in Retrieval-Based Chatbots</td>
<td>Jia-Chen Gu, Zhenhua Ling, Quan Liu, Zhigang Chen and Xiaodan Zhu</td>
<td>1412</td>
</tr>
<tr>
<td>Privacy-Preserving News Recommendation Model Learning</td>
<td>Tao Qi, Fangzhao Wu, Chuhan Wu, Yongfeng Huang and Xing Xie</td>
<td>1423</td>
</tr>
<tr>
<td>exBERT: Extending Pre-trained Models with Domain-specific Vocabulary Under Constrained Training Resources</td>
<td>Wen Tai, H. T. Kung, Xin Dong, Marcus Comiter and Chang-Fu Kuo</td>
<td>1433</td>
</tr>
<tr>
<td>Balancing via Generation for Multi-Class Text Classification Improvement</td>
<td>Naama Tepper, Esther Goldbraich, Naama Zwerdling, George Kour, Ateret Anaby Tavor and Boaz Carmeli</td>
<td>1440</td>
</tr>
<tr>
<td>Conditional Neural Generation using Sub-Aspect Functions for Extractive News Summarization</td>
<td>Zhengyuan Liu, Ke Shi and Nancy Chen</td>
<td>1453</td>
</tr>
<tr>
<td>Research Replication Prediction Using Weakly Supervised Learning</td>
<td>Tianyi Luo, xingyu li, Hainan Wang and Yang Liu</td>
<td>1464</td>
</tr>
<tr>
<td>Open Domain Question Answering based on Text Enhanced Knowledge Graph with Hyperedge Infusion</td>
<td>Jiale Han, Bo Cheng and Xu Wang</td>
<td>1475</td>
</tr>
<tr>
<td>Inexpensive Domain Adaptation of Pretrained Language Models: Case Studies on Biomedical NER and Covid-19 QA</td>
<td>Nina Poerner, Ulli Waltinger and Hinrich Schütze</td>
<td>1482</td>
</tr>
<tr>
<td>Semantically Driven Sentence Fusion: Modeling and Evaluation</td>
<td>Eyal Ben-David, Orgad Keller, Eric Malmi, Idan Szpektor and Roi Reichart</td>
<td>1491</td>
</tr>
<tr>
<td>Pseudo-Bidirectional Decoding for Local Sequence Transduction</td>
<td>Wangchunshu Zhou, Tao Ge and Ke Xu</td>
<td>1506</td>
</tr>
<tr>
<td>Predicting Responses to Psychological Questionnaires from Participants' Social Media Posts and Question Text Embeddings</td>
<td>Huy Vu, Suhai Abdurahman, Sudeep Bhatia and Lyle Ungar</td>
<td>1512</td>
</tr>
<tr>
<td>Will it Unblend?</td>
<td>Yuval Pinter, Cassandra L. Jacobs and Jacob Eisenstein</td>
<td>1525</td>
</tr>
</tbody>
</table>
CodeBERT: A Pre-Trained Model for Programming and Natural Languages
Zhangyin Feng, Daya Guo, Duyu Tang, Nan Duan, Xiaocheng Feng, Ming Gong, Linjun Shou, Bing Qin, Ting Liu, Daxin Jiang and Ming Zhou ................................................................. 1536

StyleDGPT: Stylized Response Generation with Pre-trained Language Models
Ze Yang, wei wu, Can Xu, Xinnian Liang, jiaqi bai, Liran Wang, Wei Wang and Zhoujun Li . 1548

Enhancing Automated Essay Scoring Performance via Fine-tuning Pre-trained Language Models with Combination of Regression and Ranking
Ruosong Yang, Jiannong Cao, Zhiyuan Wen, Youzheng Wu and Xiaodong He ................. 1560

Neural Dialogue State Tracking with Temporally Expressive Networks
Junfan Chen, Richong Zhang, Yongyi Mao and Jie Xu ............................................. 1570

Inferring about fraudulent collusion risk on Brazilian public works contracts in official texts using a Bi-LSTM approach
MARCOS LIMA, Roberta Silva, Felipe Lopes de Souza Mendes, Leonardo R. de Carvalho, Aleteia Araujo and Flavio de Barros Vidal ............................................................. 1580

Data-to-Text Generation with Style Imitation
Shuai Lin, Wentao Wang, Zichao Yang, Xiaodan Liang, Frank F. Xu, Eric Xing and Zhiting Hu 1589

Teaching Machine Comprehension with Compositional Explanations
Qinyuan Ye, Xiao Huang, Elizabeth Boschee and Xiang Ren ................................... 1599

A Knowledge-Driven Approach to Classifying Object and Attribute Coreferences in Opinion Mining
Jiahua Chen, Shuai Wang, Sahisnu Mazumder and Bing Liu ..................................... 1616

SimAlign: High Quality Word Alignments Without Parallel Training Data Using Static and Contextual-ized Embeddings
Masoud Jalili Sabet, Philipp Dufter, François Yvon and Hinrich Schütze ...................... 1627

TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification
Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke and Leonardo Neves .... 1644

Octa: Omissions and Conflicts in Target-Aspect Sentiment Analysis
Zhe Zhang, Chung-Wei Hang and Munindar Singh ............................................. 1651

On the Language Neutrality of Pre-trained Multilingual Representations
Jindrich Libovicky, Rudolf Rosa and Alexander Fraser ......................................... 1663

Cost-effective Selection of Pretraining Data: A Case Study of Pretraining BERT on Social Media
Xiang Dai, Sarvnaz Karimi, Ben Hachey and Cecile Paris ..................................... 1675

TopicBERT for Energy Efficient Document Classification
Yatin Chaudhary, Pankaj Gupta, Khushbu Saxena, Vivek Kulkarni, Thomas Runkler and Hinrich Schütze ................................................................. 1682

Improving Constituency Parsing with Span Attention
Yuanhe Tian, Yan Song, Fei Xia and Tong Zhang ................................................ 1691

RecoBERT: A Catalog Language Model for Text-Based Recommendations
Itzik Malkiel, Oren Barkan, Avi Caciularu, Noam Razin, Ori Katz and Noam Koenigstein . 1704
Multi-Agent Mutual Learning at Sentence-Level and Token-Level for Neural Machine Translation
Baohao Liao, Yingbo Gao and Hermann Ney .............................................................. 1715

DomBERT: Domain-oriented Language Model for Aspect-based Sentiment Analysis
Hu Xu, Bing Liu, Lei Shu and Philip Yu ................................................................. 1725

RMM: A Recursive Mental Model for Dialogue Navigation
Homero Roman Roman, Yonatan Bisk, Jesse Thomason, Asli Celikyilmaz and Jianfeng Gao . 1732

Will This Idea Spread Beyond Academia? Understanding Knowledge Transfer of Scientific Concepts across Text Corpora
Hancheng Cao, Mengjie Cheng, Zhepeng Cen, Daniel McFarland and Xiang Ren ............. 1746

Recurrent Inference in Text Editing
Ning Shi, Ziheng Zeng, Haotian Zhang and Yichen Gong ........................................ 1758

An Empirical Exploration of Local Ordering Pre-training for Structured Prediction
Zhishong Zhang, Xiang Kong, Lori Levin and Eduard Hovy ..................................... 1770

Unsupervised Extractive Summarization by Pre-training Hierarchical Transformers
Shusheng Xu, Xingxing Zhang, Yi Wu, Furu Wei and Ming Zhou ............................... 1784

Active Learning Approaches to Enhancing Neural Machine Translation
Yuekai Zhao, Haoran Zhang, Shuchang Zhou and Zhihua Zhang ............................... 1796

AGIF: An Adaptive Graph-Interactive Framework for Joint Multiple Intent Detection and Slot Filling
Libo Qin, Xiao Xu, Wanxiang Che and Ting Liu ....................................................... 1807

Continual Learning Long Short Term Memory
Xin Guo, Yu Tian, Qinghan Xue, Panos Lampropoulos, Steven Eliuk, Kenneth Barner and Xiaolong Wang ......................................................... 1817

CommonGen: A Constrained Text Generation Challenge for Generative Commonsense Reasoning
Bill Yuchen Lin, Wangchunshu Zhou, Ming Shen, Pei Zhou, Chandra Bhagavatula, Yejin Choi and Xiang Ren ................................................................. 1823

Constrained Decoding for Computationally Efficient Named Entity Recognition Taggers
Brian Lester, Daniel Pressel, Amy Hemmeter, Sagnik Ray Choudhury and Srinivas Bangalore . 1841

On the Potential of Lexico-logical Alignments for Semantic Parsing to SQL Queries
Tianze Shi, Chen Zhao, Jordan Boyd-Graber, Hal Daumé III and Lillian Lee .................. 1849

TED: A Pretrained Unsupervised Summarization Model with Theme Modeling and Denoising
Ziyi Yang, Chenguang Zhu, Robert Gmyr, Michael Zeng, Xuedong Huang and Eric Darve . 1865

Improving End-to-End Bangla Speech Recognition with Semi-supervised Training
Nafis Sadeq, Nafs Tahmid Chowdhury, Farhan Tanvir Utshaw, Shafayat Ahmed and Muhammad Abdullah Adnan ........................................ 1875

No Gestures Left Behind: Learning Relationships between Spoken Language and Freeform Gestures
Chaitanya Ahuja, Dong Won Lee, Ryo Ishii and Louis-Philippe Morency ........................ 1884

UNIFIEDQA: Crossing Format Boundaries with a Single QA System
Daniel Khashabi, Sewon Min, Tushar Khot, Ashish Sabharwal, Oyvind Tafjord, Peter Clark and Hannaneh Hajishirzi ......................................................... 1896
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robust and Interpretable Grounding of Spatial References with Relation Networks</td>
<td>Tsung-Yen Yang, Andrew Lan and Karthik Narasimhan</td>
<td>1908</td>
</tr>
<tr>
<td>Pragmatic Issue-Sensitive Image Captioning</td>
<td>Allen Nie, Reuben Cohn-Gordon and Christopher Potts</td>
<td>1924</td>
</tr>
<tr>
<td>PTUM: Pre-training User Model from Unlabeled User Behaviors via Self-supervision</td>
<td>Chuhan Wu, Fangzhao Wu, Tao Qi, Jianxun Lian, Yongfeng Huang and Xing Xie</td>
<td>1939</td>
</tr>
<tr>
<td>Adversarial Subword Regularization for Robust Neural Machine Translation</td>
<td>Jungsoo Park, Mujeen Sung, Jinhuyuk Lee and Jaewoo Kang</td>
<td>1945</td>
</tr>
<tr>
<td>SynET: Synonym Expansion using Transitivity</td>
<td>Jiale Yu, Yongliang Shen, Xinyin Ma, Chenghao Jia, Chen Chen and Weiming Lu</td>
<td>1961</td>
</tr>
<tr>
<td>Scheduled DropHead: A Regularization Method for Transformer Models</td>
<td>Wangchunshu Zhou, Tao Ge, Furu Wei, Ming Zhou and Ke Xu</td>
<td>1971</td>
</tr>
<tr>
<td>Multi-Turn Dialogue Generation in E-Commerce Platform with the Context of Historical Dialogue</td>
<td>Weisheng Zhang, Kaisong Song, Yangyang Kang, Zhongqing Wang, Changlong Sun, Xiaozhong Liu, Shoushan Li, Min Zhang and Luo Si</td>
<td>1981</td>
</tr>
<tr>
<td>Ruler: Data Programming by Demonstration for Document Labeling</td>
<td>Sara Evensen, Chang Ge and Cagatay Demiralp</td>
<td>1996</td>
</tr>
<tr>
<td>Dual Reconstruction: a Unifying Objective for Semi-Supervised Neural Machine Translation</td>
<td>Weijia Xu, Xing Niu and Marine Carpuat</td>
<td>2006</td>
</tr>
<tr>
<td>Focus-Constrained Attention Mechanism for CVAE-based Response Generation</td>
<td>Zhi Cui, Yanran Li, Jiayi ZHANG, Jianwei Cui, Chen Wei and Bin Wang</td>
<td>2021</td>
</tr>
<tr>
<td>Chunk-based Chinese Spelling Check with Global Optimization</td>
<td>Zuyi Bao, Chen Li and Rui Wang</td>
<td>2031</td>
</tr>
<tr>
<td>Multi-pretraining for Large-scale Text Classification</td>
<td>Kang-Min Kim, Bumsu Hyeon, Yeachan Kim, Jun-Hyung Park and SangKeun Lee</td>
<td>2041</td>
</tr>
<tr>
<td>End-to-End Speech Recognition and Disfluency Removal</td>
<td>Paria Jamshid Lou and Mark Johnson</td>
<td>2051</td>
</tr>
<tr>
<td>Characterizing the Value of Information in Medical Notes</td>
<td>Chao-Chun Hsu, Shantanu Karnwal, Sendhil Mullainathan, Ziad Obermeyer and Chenhao Tan</td>
<td>2062</td>
</tr>
<tr>
<td>KLearn: Background Knowledge Inference from Summarization Data</td>
<td>Maxime Peyrard and Robert West</td>
<td>2073</td>
</tr>
<tr>
<td>Extracting Chemical-Protein Interactions via Calibrated Deep Neural Network and Self-training</td>
<td>Dongha Choi and Hyunju Lee</td>
<td>2086</td>
</tr>
</tbody>
</table>
Rethinking Topic Modelling: From Document-Space to Term-Space  
Magnus Sahlgren ................................................................. 2250

Sparse and Decorrelated Representations for Stable Zero-shot NMT  
Bokyung Son and Sungwon Lyu ............................................. 2260

A Semi-supervised Approach to Generate the Code-Mixed Text using Pre-trained Encoder and Transfer Learning  
Deepak Gupta, Asif Ekbal and Pushpak Bhattacharyya .................... 2267

BERT-MK: Integrating Graph Contextualized Knowledge into Pre-trained Language Models  
Bin He, Di Zhou, Jinghui Xiao, Xin Jiang, Qun Liu, Nicholas Jing Yuan and Tong Xu .......... 2281

Recursive Top-Down Production for Sentence Generation with Latent Trees  
Shawn Tan, Yikang Shen, Alessandro Sordoni, Aaron Courville and Timothy J. O’Donnell .. 2291

Guided Dialogue Policy Learning without Adversarial Learning in the Loop  
Ziming Li, Sungjin Lee, Baolin Peng, Jinchao Li, Julia Kiseleva, Maarten de Rijke, Shahin Shayan-deh and Jianfeng Gao ........................................................... 2308

MultiDM-GCN: Aspect-guided Response Generation in Multi-domain Multi-modal Dialogue System using Graph Convolutional Network  
Mauajama Firdaus, NIDHI THAKUR and Asif Ekbal .......................... 2318

Edge-Enhanced Graph Convolution Networks for Event Detection with Syntactic Relation  
Shiyao Cui, Bowen Yu, Tingwen Liu, Zhenyu Zhang, Xuebin Wang and Jinqiao Shi ......... 2329

Semi-supervised Formality Style Transfer using Language Model Discriminator and Mutual Information Maximization  
Kunal Chawla and Diyi Yang .................................................. 2340

Differentially Private Representation for NLP: Formal Guarantee and An Empirical Study on Privacy and Fairness  
Lingjuan Lyu, Xuanli He and Yitong Li ...................................... 2355

Helpful or Hierarchical? Predicting the Communicative Strategies of Chat Participants, and their Impact on Success  
Farzana Rashid, Tommaso Fornaciari, Dirk Hovy, Eduardo Blanco and Fernando Vega-Redondo 2366

Learning Knowledge Bases with Parameters for Task-Oriented Dialogue Systems  
Andrea Madotto, Samuel Cahyawijaya, Genta Indra Winata, Yan Xu, Zihan Liu, Zhaojiang Lin and Pascale Fung .......................................................... 2372

Adapting Open Domain Fact Extraction and Verification to COVID-FACT through In-Domain Language Modeling  
Zhenghao Liu, Chenyan Xiong, Zhuyun Dai, Si Sun, Maosong Sun and Zhiyuan Liu .......... 2395

ProphetNet: Predicting Future N-gram for Sequence-to-Sequence Pre-training  
Weizhen Qi, Yu Yan, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang and Ming Zhou ................................................................. 2401

DivGAN: Towards Diverse Paraphrase Generation via Diversified Generative Adversarial Network  
Yue Cao and Xiaojun Wan .......................................................... 2411
Plug-and-Play Conversational Models
Andrea Madotto, Etsuko Ishii, Zhaojiang Lin, Sumanth Dathathri and Pascale Fung 2422

Event-Driven Learning of Systematic Behaviours in Stock Markets
Xianchao Wu 2434

Learning Improvised Chatbots from Adversarial Modifications of Natural Language Feedback
Makens Narimohan Sreedhar, Kun Ni and Siva Reddy 2445

Adapting Coreference Resolution to Twitter Conversations
Berfin Aktaş, Veronika Solopova, Annalena Kohnert and Manfred Stede 2454

On Romanization for Model Transfer Between Scripts in Neural Machine Translation
Chantal Amrhein and Rico Sennrich 2461

COSMIC: CommonSense knowledge for eMotion Identification in Conversations
Deepanway Ghosal, Navonil Majumder, Alexander Gelbukh, Rada Mihalcea and Soujanya Poria 2470

Improving Compositional Generalization in Semantic Parsing
Inbar Oren, Jonathan Herzig, Nitish Gupta, Matt Gardner and Jonathan Berant 2482

Answer Span Correction in Machine Reading Comprehension
Revanth Gangi Reddy, Md Arafat Sultan, Efsun Sarioglu Kayi, Rong Zhang, Vittorio Castelli and Avi Sil 2496

On the Interplay Between Fine-tuning and Sentence-level Probing for Linguistic Knowledge in Pre-trained Transformers
Marius Mosbach, Anna Khokhlova, Michael A. Hedderich and Dietrich Klakow 2502

Zero-shot Entity Linking with Efficient Long Range Sequence Modeling
Zonghai Yao, Liangliang Cao and Huapu Pan 2517

Jian Liu, Yubo Chen, Kang Liu, Yantao Jia and Zhicheng Sheng 2523

Adaptive Feature Selection for End-to-End Speech Translation
Biao Zhang, Ivan Titov, Barry Haddow and Rico Sennrich 2533

Abstractive Multi-Document Summarization via Joint Learning with Single-Document Summarization
Hanqi Jin and Xiaojun Wan 2545

Blockwise Self-Attention for Long Document Understanding
Jiezhong Qiu, Hao Ma, Omer Levy, Wen-tau Yih, Sinong Wang and Jie Tang 2555

Unsupervised Few-Bits Semantic Hashing with Implicit Topics Modeling
Fanghua Ye, Jarana Manotumruksa and Emine Yilmaz 2566

Grid Tagging Scheme for Aspect-oriented Fine-grained Opinion Extraction
Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai and Rui Xia 2576

Learning Numerical Embedding
Chengyue Jiang, Zhonglin Nian, Kaihao Guo, Shanbo Chu, Yinggong Zhao, Libin Shen and Kewei Tu 2586
An Investigation of Potential Function Designs for Neural CRF
Zechuan Hu, Yong Jiang, Nguyen Bach, Tao Wang, Zhongqiang Huang, Fei Huang and Kewei Tu
2600

Fast End-to-end Coreference Resolution for Korean
Cheoneum Park, Jamin Shin, Sungjoon Park, Joonho Lim and Changki Lee ................. 2610

Toward Stance-based Personas for Opinionated Dialogues
Thomas Scialom, Serra Sinem Tekiroğlu, Jacopo Staiano and Marco Guerini .............. 2625

Hierarchical Pre-training for Sequence Labelling in Spoken Dialog
Emile Chapuis, Pierre Colombo, Matteo Manica, Matthieu Labeau and Chloé Clavel ........ 2636

Extending Multilingual BERT to Low-Resource Languages
Zihan Wang, Karthikeyan K, Stephen Mayhew and Dan Roth ............................... 2649

Out-of-Sample Representation Learning for Knowledge Graphs
Marjan Albooyeh, Rishab Goel and Seyed Mehran Kazemi .................................. 2657

Fine-Grained Grounding for Multimodal Speech Recognition
Tejas Srinivasan, Ramon Sanabria, Florian Metze and Desmond Elliott .................. 2667

Unsupervised Expressive Rules Provide Explainability and Assist Human Experts Grasping New Domains
Eyal Shnarch, Leshem Choshen, Guy Moshkovich, Ranit Aharonov and Noam Slonim .... 2678

Textual Supervision for Visually Grounded Spoken Language Understanding
Bertrand Higy, Desmond Elliott and Grzegorz Chrupała ....................................... 2698

Universal Dependencies According to BERT: Both More Specific and More General
Tomasz Limisiewicz, David Mareček and Rudolf Rosa ............................................ 2710

Visual Objects As Context: Exploiting Visual Objects for Lexical Entailment
Masayasu Muraoka, Tetsuya Nasukawa and Bishwaranjana Bhattacharjee .............. 2723

Learning to Plan and Realize Separately for Open-Ended Dialogue Systems
Sashank Santhanam, Zhuo Cheng, Brodie Mather, Bonnie Dorr, Archana Bhatia, Bryanna Hebenstreit, Alan Zemel, Adam Dalton, Tomek Strzalkowski and Samira Shaikh ...................... 2736

Be Different to Be Better! A Benchmark to Leverage the Complementarity of Language and Vision
Sandro Pezzelle, Claudio Greco, Greta Gandolfi, Eleonora Gualdoni and Raffaella Bernardi . 2751

Cross-Lingual Training of Neural Models for Document Ranking
Peng Shi, He Bai and Jimmy Lin ................................................................. 2768

Improving Word Embedding Factorization for Compression Using Distilled Nonlinear Neural Decomposition
Vasileios Lioutas, Ahmad Rashid, Krin Kumar, Md. Akmal Haidar and Mehdi Rezagholizadeh 2774

PharmMT: A Neural Machine Translation Approach to Simplify Prescription Directions
Jiazhao Li, Corey Lester, Xinyan Zhao, Yuting Ding, Yun Jiang and V.G. Vinod Vydisharan . 2785

LSTMs Compose—and Learn—Bottom-Up
Naomi Saphra and Adam Lopez ............................................................... 2797
Natural Language Rationales with Full-Stack Visual Reasoning: From Pixels to Semantic Frames to Commonsense Graphs
Ana Marasović, Chandra Bhagavatula, Jae sung Park, Ronan Le Bras, Noah A. Smith and Yejin Choi .......................................................... 2810

Corpora Evaluation and System Bias Detection in Multi-document Summarization
Alvin Dey, Tanya Chowdhury, Yash Kumar and Tammoj Chakraborty .................. 2830

Graph-to-Tree Neural Networks for Learning Structured Input-Output Translation with Applications to Semantic Parsing and Math Word Problem
Shucheng Li, Lingfei Wu, Shiwei Feng, Fangli Xu, Fengyuan Xu and Sheng Zhong ........ 2841

Target Conditioning for One-to-Many Generation
marie-anne lachaux, Armand Joulain and Guillaume Lample .................................. 2853

Can Pre-training help VQA with Lexical Variations?
Shailza Jolly and Shubham Kapoor ................................................................. 2863

FENAS: Flexible and Expressive Neural Architecture Search
Ramakanth Pasunuru and Mohit Bansal ........................................................... 2869

Inferring symmetry in natural language
Chelsea Tanchip, Lei Yu, Aotao Xu and Yang Xu ............................................... 2877

A Concise Model for Multi-Criteria Chinese Word Segmentation with Transformer Encoder
Xipeng Qiu, Hengzhi Pei, Hang Yan and Xuanjing Huang .................................. 2887

LEGAL-BERT: The Muppets straight out of Law School
Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras and Ion Androutsopoulos ........................................ 2898

Enhancing Content Planning for Table-to-Text Generation with Data Understanding and Verification
Heng Gong, Wei Bi, Xiaocheng Feng, Bing Qin, Xiaojiang Liu and Ting Liu ............. 2905

Contextual Text Style Transfer
Yu Cheng, Zhe Gan, Yizhe Zhang, Oussama Elachqar, Dianqi Li and Jingjing Liu .... 2915

DiPair: Fast and Accurate Distillation for Trillion-Scale Text Matching and Pair Modeling
Jiecao Chen, Liu Yang, Karthik Raman, Michael Bendersky, Jung-Jung Yeh, Yun Zhou, Marc Najork, Danyang Cai and Ehsan Emadzadeh ........................................ 2925

Cross-Lingual Dependency Parsing by POS-Guided Word Reordering
Lu Liu, Yi Zhou, Jianhan Xu, Xiaoqing Zheng, Kai-Wei Chang and Xuanjing Huang .... 2938

Assessing Robustness of Text Classification through Maximal Safe Radius Computation
Emanuele La Malfa, Min Wu, Luca Laurenti, Benjie Wang, Anthony Hartshorn and Marta Kwiatkowska 2949

Social Commonsense Reasoning with Multi-Head Knowledge Attention
Debjit Paul and Anette Frank ................................................................. 2969

TurnGPT: a Transformer-based Language Model for Predicting Turn-taking in Spoken Dialog
Erik Ekstedt and Gabriel Skantze ................................................................. 2981

A little goes a long way: Improving toxic language classification despite data scarcity
Mika Juuti, Tommi Gröndahl, Adrian Flanagan and N. Asokan .......................... 2991

xx
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>KoBE: Knowledge-Based Machine Translation Evaluation</td>
<td>Zorik Gekhman, Roee Aharoni, Genady Beryozkin, Markus Freitag and Wolfgang Macherey</td>
<td>3200</td>
</tr>
<tr>
<td>Pushing the Limits of AMR Parsing with Self-Learning</td>
<td>Young-Suk Lee, Ramón Fernandez Astudillo, Tahira Naseem, Revanth Gangi Reddy, Radu Florian and Salim Roukos</td>
<td>3208</td>
</tr>
<tr>
<td>Towards Zero-Shot Conditional Summarization with Adaptive Multi-Task Fine-Tuning</td>
<td>Travis Goodwin, Max Savery and Dina Demner-Fushman</td>
<td>3215</td>
</tr>
<tr>
<td>Multilingual Knowledge Graph Completion via Ensemble Knowledge Transfer</td>
<td>Xuelu Chen, Muhao Chen, Changjun Fan, Ankith Uppunda, Yizhou Sun and Carlo Zaniolo</td>
<td>3227</td>
</tr>
<tr>
<td>Towards Controllable Biases in Language Generation</td>
<td>Emily Sheng, Kai-Wei Chang, Prem Natarajan and Nanyun Peng</td>
<td>3239</td>
</tr>
<tr>
<td>RobBERT: a Dutch RoBERTa-based Language Model</td>
<td>Pieter Delobelle, Thomas Winters and Bettina Berendt</td>
<td>3255</td>
</tr>
<tr>
<td>Regularization of Distinct Strategies for Unsupervised Question Generation</td>
<td>Junmo Kang, Giwon Hong, Haritz Puerto San Roman and Sung-Hyon Myaeng</td>
<td>3266</td>
</tr>
<tr>
<td>Graph-to-Graph Transformer for Transition-based Dependency Parsing</td>
<td>Alireza Mohammadshahi and James Henderson</td>
<td>3278</td>
</tr>
<tr>
<td>WER we are and WER we think we are</td>
<td>Piotr Szymański, Piotr Żelasko, Mikolaj Morzy, Adrian Szymczak, Marzena Żyła-Hoppe, Joanna Banaszczuk, Łukasz Augustyniak, Jan Mizgajski and Yishay Carmiel</td>
<td>3290</td>
</tr>
<tr>
<td>DeSMOG: Detecting Stance in Media On Global Warming</td>
<td>Yiwei Luo, Dallas Card and Dan Jurafsky</td>
<td>3296</td>
</tr>
<tr>
<td>Improve Transformer Models with Better Relative Position Embeddings</td>
<td>zhiheng huang, Davis Liang, Peng Xu and Bing Xiang</td>
<td>3297</td>
</tr>
<tr>
<td>A Sentiment-Controllable Topic-to-Essay Generator with Topic Knowledge Graph</td>
<td>Lin Qiao, Jianhao Yan, Fandong Meng, Zhendong Yang and Jie Zhou</td>
<td>3336</td>
</tr>
<tr>
<td>What-if I ask you to explain: Explaining the effects of perturbations in procedural text</td>
<td>Dheeraj Rajagopal, Niket Tandon, Peter Clark, Bhavana Dalvi and Eduard Hovy</td>
<td>3345</td>
</tr>
<tr>
<td>RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models</td>
<td>Samuel Gehman, Suchin Gururangan, Maarten Sap, Yejin Choi and Noah A. Smith</td>
<td>3356</td>
</tr>
<tr>
<td>Improving Event Duration Prediction via Time-aware Pre-training</td>
<td>Zonglin Yang, Xinya Du, Alexander Rush and Claire Cardie</td>
<td>3370</td>
</tr>
<tr>
<td>Composed Variational Natural Language Generation for Few-shot Intents</td>
<td>Congying Xia, Caiming Xiong, Philip Yu and Richard Socher</td>
<td>3379</td>
</tr>
<tr>
<td>Literature Retrieval for Precision Medicine with Neural Matching and Faceted Summarization</td>
<td>Jiho Noh and Ramakanth Kavuluru</td>
<td>3389</td>
</tr>
</tbody>
</table>
Cross-Lingual Text Classification with Minimal Resources by Transferring a Sparse Teacher
Giannis Karamanolakis, Daniel Hsu and Luis Gravano .................................................. 3604

A Multi-Persona Chatbot for Hotline Counselor Training
Orianna Demasi, Yu Li and Zhou Yu ................................................................. 3623

Narrative Text Generation with a Latent Discrete Plan
Harsh Jhamtani and Taylor Berg-Kirkpatrick ......................................................... 3637

Graph Transformer Networks with Syntactic and Semantic Structures for Event Argument Extraction
Amir Pouran Ben Veyseh, Tuan Ngo Nguyen and Thien Huy Nguyen ...................... 3651

The Box is in the Pen: Evaluating Commonsense Reasoning in Neural Machine Translation
Jie He, Tao Wang, Deyi Xiong and Qun Liu .......................................................... 3662

Using Visual Feature Space as a Pivot Across Languages
Ziyan Yang, Leticia Pinto-Alva, Franck Dernoncourt and Vicente Ordonez .................. 3673

CDEvalSumm: An Empirical Study of Cross-Dataset Evaluation for Neural Summarization Systems
Yiran Chen, Pengfei Liu, Ming Zhong, Zi-Yi Dou, Danqing Wang, Xipeng Qiu and Xuanjing Huang 3679

Attending to Long-Distance Document Context for Sequence Labeling
Matthew Jörke, Jon Gillick, Matthew Sims and David Bamman ................................ 3692

Global Bootstrapping Neural Network for Entity Set Expansion
Lingyong Yan, Xianpei Han, Ben He and Le Sun .................................................... 3705

Document Classification for COVID-19 Literature
Bernal Jimenez Gutierrez, Jucheng Zeng, Dongdong Zhang, Ping Zhang and Yu Su ........ 3715

Adversarial Augmentation Policy Search for Domain and Cross-Lingual Generalization in Reading Comprehension
Adyasha Maharana and Mohit Bansal ............................................................... 3723

Denoising Multi-Source Weak Supervision for Neural Text Classification
Wendi Ren, Yinghao Li, Hanxing Su, David Kartchner, Cassie Mitchell and Chao Zhang 3739

Dr. Summarize: Global Summarization of Medical Dialogue by Exploiting Local Structures.
Anirudh Joshi, Namit Katariya, Xavier Amatriain and Anitha Kannan ................. 3755

Generating Accurate Electronic Health Assessment from Medical Graph
Zhichao Yang and hong yu .............................................................................. 3764

Do Models of Mental Health Based on Social Media Data Generalize?
Keith Harrigian, Carlos Aguirre and Mark Dredze .............................................. 3774

Context Analysis for Pre-trained Masked Language Models
Yi-An Lai, Garima Lalwani and Yi Zhang ......................................................... 3789

Controllable Text Generation with Focused Variation
Lei Shu, Alexandros Papangelis, Yi-Chia Wang, Gokhan Tur, Hu Xu, Zhaleh Feizollahi, Bing Liu and Piero Molino ............................................................... 3805
Modeling Preconditions in Text with a Crowd-sourced Dataset
Heeyoung Kwon, Mahnaz Koupae, Pratyush Singh, Gargi Sawhney, Anmol Shukla, Keerthi Kumar Kallur, Nathanael Chambers and Niranjan Balasubramanian .......................... 3818

Reevaluating Adversarial Examples in Natural Language
John Morris, Eli Lifland, Jack Lanchantin, Yangfeng Ji and Yanjun Qi ......................... 3829

Question Answering with Long Multiple-Span Answers
Ming Zhu, Aman Ahuja, Da-Cheng Juan, Wei Wei and Chandan K Reddy .................. 3840

Inserting Information Bottlenecks for Attribution in Transformers
Zhiying Jiang, Raphael Tang, Ji Xin and Jimmy Lin ........................................... 3850

Event-Related Bias Removal for Real-time Disaster Events
SALVADOR MEDINA MAZA, Evangelia Spiliopoulou, Eduard Hovy and Alexander Hauptmann 3858

It’s not a Non-Issue: Negation as a Source of Error in Machine Translation
Md Mosharaf Hossain, Antonios Anastasopoulos, Eduardo Blanco and Alexis Palmer .... 3869

Incremental Text-to-Speech Synthesis with Prefix-to-Prefix Framework
Mingbo Ma, Baigong Zheng, Kaibo Liu, Renjie Zheng, Hairong Liu, Kainan Peng, Kenneth Church and Liang Huang ................................................................. 3886

Joint Turn and Dialogue level User Satisfaction Estimation on Multi-Domain Conversations
Praveen Kumar Bodigutla, Aditya Tiwari, Spyros Matsoukas, Josep Valls-Vargas and Lazaros Poly-menakos ................................................................. 3897

ArraMon: A Joint Navigation-Assembly Instruction Interpretation Task in Dynamic Environments
Hyounghun Kim, Abhaysinh Zala, Graham Burri, Hao Tan and Mohit Bansal ............... 3910

Fluent and Low-latency Simultaneous Speech-to-Speech Translation with Self-adaptive Training
Renjie Zheng, Mingbo Ma, Baigong Zheng, Kaibo Liu, Jiahong Yuan, Kenneth Church and Liang Huang ................................................................. 3928

Towards Context-Aware Code Comment Generation
Xiaohan Yu, Quzhe Huang, Zheng Wang, Yansong Feng and Dongyan Zhao .............. 3938

MCMH: Learning Multi-Chain Multi-Hop Rules for Knowledge Graph Reasoning
Lu Zhang, Mo Yu, Tian Gao and yue yu .......................................................... 3948

Finding the Optimal Vocabulary Size for Neural Machine Translation
Thamme Gowda and Jonathan May ................................................................. 3955

Weakly- and Semi-supervised Evidence Extraction
Danish Pruthi, Bhuwan Dhingra, Graham Neubig and Zachary C. Lipton .................... 3965

Making Information Seeking Easier: An Improved Pipeline for Conversational Search
Vaibhav Kumar and Jamie Callan ................................................................. 3971

Generalizable and Explainable Dialogue Generation via Explicit Action Learning
Xinting Huang, Jianzhong Qi, Yu Sun and Rui Zhang ........................................ 3981

More Embeddings, Better Sequence Labelers?
Xinyu Wang, Yong Jiang, Nguyen Bach, Tao Wang, Zhongqiang Huang, Fei Huang and Kewei Tu 3992
NLP Service APIs and Models for Efficient Registration of New Clients
Sahil Shah, Vihari Piratla, Soumen Chakrabarti and Sunita Sarawagi ......................... 4007

Effects of Naturalistic Variation in Goal-Oriented Dialog
Jatin Ganhotra, Robert Moore, Sachindra Joshi and Kahini Wadhawan ......................... 4013

Determining Event Outcomes: The Case of #fail
Srikala Murugan, Dhivya Chinnappa and Eduardo Blanco ........................................ 4021

WikiLingua: A New Benchmark Dataset for Cross-Lingual Abstractive Summarization
Faisal Ladhak, Esin Durmus, Claire Cardie and Kathleen McKeown ............................. 4034

Adversarial Training for Code Retrieval with Question-Description Relevance Regularization
Jie Zhao and Huan Sun ............................................................................................... 4049

Large Product Key Memory for Pretrained Language Models
Gyuwan Kim and Tae Hwan Jung .................................................................................. 4060

Temporal Reasoning in Natural Language Inference
Siddharth Vashishtha, Adam Poliak, Yash Kumar Lal, Benjamin Van Durme and Aaron Steven White .......................................................... 4070

A Pilot Study of Text-to-SQL Semantic Parsing for Vietnamese
Anh Tuan Nguyen, Mai Hoang Dao and Dat Quoc Nguyen ........................................ 4079

STANDER: An Expert-Annotated Dataset for News Stance Detection and Evidence Retrieval
Costanza Conforti, Jakob Berndt, Mohammad Taher Pilehvar, Chryssi Giannitsarou, Flavio Tossa-vaerd and Nigel Collier ............................................................ 4086

An Empirical Methodology for Detecting and Prioritizing Needs during Crisis Events
M. Janina Sarol, Ly Dinh, Rezvaneh Rezapour, Chieh-Li Chin, Pingjing Yang and Jana Diesner 4102

SupMMD: A Sentence Importance Model for Extractive Summarization using Maximum Mean Discrep-ancy
Umanga Bista, Alexander Mathews, Aditya Menon and Lexing Xie ................................. 4108

Towards Low-Resource Semi-Supervised Dialogue Generation with Meta-Learning
Yi Huang, Junlan Feng, Shuo Ma, Xiaoyu Du and Xiaoting Wu .................................. 4123

Connecting the Dots: A Knowledgeable Path Generator for Commonsense Question Answering
Peifeng Wang, Nanyun Peng, Filip Ilievski, Pedro Szekely and Xiang Ren ..................... 4129

No Answer is Better Than Wrong Answer: A Reflection Model for Document Level Machine Reading Comprehension
Xuguang Wang, Linjun Shou, Ming Gong, Nan Duan and Daxin Jiang .......................... 4141

Reference Language based Unsupervised Neural Machine Translation
Zuchao Li, Hai Zhao, Rui Wang, Masao Utiyama and Eiichiro Sumita ............................ 4151

TinyBERT: Distilling BERT for Natural Language Understanding
Xiaoqi Jiao, Yichun Yin, Lifeng Shang, Xin Jiang, Xiao Chen, Linlin Li, Fang Wang and Qun Liu 4163

Poison Attacks against Text Datasets with Conditional Adversarially Regularized Autoencoder
Alvin Chan, Yi Tay, Yew-Soon Ong and Aston Zhang .................................................. 4175
A Pointer Network Architecture for Joint Morphological Segmentation and Tagging
Amit Seker and Reut Tsarfaty ................................................................. 4368

Beyond Language: Learning Commonsense from Images for Reasoning
Wanqing Cui, Yanyan Lan, Liang Pang, Jiafeng Guo and Xueqi Cheng ................. 4379

A BERT-based Distractor Generation Scheme with Multi-tasking and Negative Answer Training Strategies.
Ho-Lam Chung, Ying-Hong Chan and Yao-Chung Fan .................................. 4390

How Effective is Task-Agnostic Data Augmentation for Pretrained Transformers?
Shayne Longpre, Yu Wang and Chris DuBois ............................................ 4401

Visually-Grounded Planning without Vision: Language Models Infer Detailed Plans from High-level Instructions
Peter Jansen ............................................................................................... 4412

Consistent Response Generation with Controlled Specificity
Junya Takayama and Yuki Arase ................................................................. 4418

Internal and external pressures on language emergence: least effort, object constancy and frequency
Diana Rodríguez Luna, Edoardo Maria Ponti, Dieuwke Hupkes and Elia Bruni ........ 4428

Parsing All: Syntax and Semantics, Dependencies and Spans
Junru Zhou, Zuchao Li and Hai Zhao ............................................................ 4438

LIMIT-BERT: Linguistics Informed Multi-Task BERT
Junru Zhou, Zhuosheng Zhang, Hai Zhao and Shuailiang Zhang ......................... 4450

Improving Limited Labeled Dialogue State Tracking with Self-Supervision
Chien-Sheng Wu, Steven C.H. Hoi and Caiming Xiong .................................... 4462

On the Branching Bias of Syntax Extracted from Pre-trained Language Models
Huayang Li, Lemao Liu, Guoping Huang and Shuming Shi ............................... 4473

The Pragmatics behind Politics: Modelling Metaphor, Framing and Emotion in Political Discourse
Pere-Lluis Huguet Cabot, Verna Dankers, David Abadi, Agneta Fischer and Ekaterina Shutova 4479

SMRT Chatbots: Improving Non-Task-Oriented Dialog with Simulated Multiple Reference Training
Huda Khayrallah and João Sedoc ..................................................................... 4489

PrivNet: Safeguarding Private Attributes in Transfer Learning for Recommendation
Guangneng Hu and Qiang Yang ..................................................................... 4506

Learning to Learn to Disambiguate: Meta-Learning for Few-Shot Word Sense Disambiguation
Nithin Holla, Pushkar Mishra, Helen Yannakoudakis and Ekaterina Shutova ........ 4517

An Empirical Investigation of Beam-Aware Training in Supertagging
Renato Negrinho, Matthew R. Gormley and Geoff Gordon ............................... 4534

Improving Aspect-based Sentiment Analysis with Gated Graph Convolutional Networks and Syntax-based Regulation
Amir Pouran Ben Veyseh, Nasim Nouri, Franck Dernoncourt, Quan Hung Tran, Dejing Dou and Thien Huu Nguyen ................................................................. 4543
<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decoding Language Spatial Relations to 2D Spatial Arrangements</td>
<td>Gorjan Radevski, Guillem Collell, Marie-Francine Moens and Tinne Tuytelaars</td>
<td>4549</td>
</tr>
<tr>
<td>The Dots Have Their Values: Exploiting the Node-Edge Connections in Graph-based Neural Models for Document-level Relation Extraction</td>
<td>Hieu Minh Tran, Minh Trung Nguyen and Thien Huu Nguyen</td>
<td>4561</td>
</tr>
<tr>
<td>Why and when should you pool? Analyzing Pooling in Recurrent Architectures</td>
<td>Pratyush Maini, Keshav Kolluru, Danish Pruthi and</td>
<td>4568</td>
</tr>
<tr>
<td>Structural and Functional Decomposition for Personality Image Captioning in a Communication Game</td>
<td>Minh Thu Nguyen, Duy Phung, Minh Hoai and Thien Huu Nguyen</td>
<td>4587</td>
</tr>
<tr>
<td>Long Document Ranking with Query-Directed Sparse Transformer</td>
<td>Jyun-Yu Jiang, Chenyan Xiong, Chia-Jung Lee and Wei Wang</td>
<td>4594</td>
</tr>
<tr>
<td>Visuo-Linguistic Question Answering (VLQA) Challenge</td>
<td>Shailaja Keyur Sampat, Yezhou Yang and Chitta Baral</td>
<td>4606</td>
</tr>
<tr>
<td>Byte Pair Encoding is Suboptimal for Language Model Pretraining</td>
<td>Kaj Bostrom and Greg Durrett</td>
<td>4617</td>
</tr>
<tr>
<td>Exploring BERT’s Sensitivity to Lexical Cues using Tests from Semantic Priming</td>
<td>Kanishka Misra, Allyson Ettinger and Julia Rayz</td>
<td>4625</td>
</tr>
<tr>
<td>Multi-hop Question Generation with Graph Convolutional Network</td>
<td>Dan Su, Yan Xu, Weniang Dai, Ziwei Ji, Tiezheng Yu and Pascale Fung</td>
<td>4636</td>
</tr>
<tr>
<td>Thinking Like a Skeptic: Defeasible Inference in Natural Language</td>
<td>Rachel Rudinger, Vered Shwartz, Jena D. Hwang, Chandra Bhagavatula, Maxwell Forbes, Ronan Le Bras, Noah A. Smith and Yejin Choi</td>
<td>4661</td>
</tr>
<tr>
<td>Guiding Attention for Self-Supervised Learning with Transformers</td>
<td>Ameet Deshpande and Karthik Narasimhan</td>
<td>4676</td>
</tr>
<tr>
<td>Language-Conditioned Feature Pyramids for Visual Selection Tasks</td>
<td>Taichi Iki and Akiko Aizawa</td>
<td>4687</td>
</tr>
<tr>
<td>Learning to Classify Events from Human Needs Category Descriptions</td>
<td>Haibo Ding and Zhe Feng</td>
<td>4698</td>
</tr>
<tr>
<td>Automatic Term Name Generation for Gene Ontology: Task and Dataset</td>
<td>Yanjian Zhang, Qin Chen, Yiteng Zhang, Zhongyu Wei, Yixu Gao, Jiakie Peng, Zengfeng Huang, Weijian Sun and Xuanjing Huang</td>
<td>4705</td>
</tr>
<tr>
<td>BERT-QE: Contextualized Query Expansion for Document Re-ranking</td>
<td>Zhi Zheng, Kai Hui, Ben He, Xianpei Han, Le Sun and Andrew Yates</td>
<td>4718</td>
</tr>
</tbody>
</table>
ZEN: Pre-training Chinese Text Encoder Enhanced by N-gram Representations
Shizhe Diao, Jiaxin Bai, Yan Song, Tong Zhang and Yonggang Wang .......................... 4729

Finding Friends and Flipping Frenemies: Automatic Paraphrase Dataset Augmentation Using Graph Theory
Hannah Chen, Yangfeng Ji and David Evans ............................................................... 4741

Probabilistic Case-based Reasoning for Open-World Knowledge Graph Completion
Rajarshi Das, Ameya Godbole, Nicholas Monath, Manzil Zaheer and Andrew McCallum . . 4752

TLDR: Extreme Summarization of Scientific Documents
Isabel Cachola, Kyle Lo, Arman Cohan and Daniel Weld ........................................... 4766

Tri-Train: Automatic Pre-Fine Tuning between Pre-Training and Fine-Tuning for SciNER
Qingkai Zeng, Wenhao Yu, Mengxia Yu, Tianwen Jiang, Tim Weninger and Meng Jiang . . 4778

Hierarchical Region Learning for Nested Named Entity Recognition
Xinwei Long, Shuzi Niu and Yucheng Li ................................................................. 4788

Understanding User Resistance Strategies in Persuasive Conversations
Youzhi Tian, Weiyan Shi, Chen Li and Zhou Yu ......................................................... 4794

On the Sub-layer Functionalities of Transformer Decoder
Yilin Yang, Longyue Wang, Shuming Shi, Prasad Tadepalli, Stefan Lee and Zhaopeng Tu . 4799

Extremely Low Bit Transformer Quantization for On-Device Neural Machine Translation
Insoo Chung, Byeongwook Kim, Yoonjung Choi, Se Jung Kwon, Yongkweon Jeon, Baeseong Park, Sangha Kim and Dongsoo Lee ................................................................. 4812

Robust Backed-off Estimation of Out-of-Vocabulary Embeddings
Nobukazu Fukuda, Naoki Yoshinaga and Masaru Kitsuregawa .................................. 4827

Exploiting Unsupervised Data for Emotion Recognition in Conversations
Wenxiang Jiao, Michael Lyu and Irwin King .............................................................. 4839

Tensorized Embedding Layers
Oleksii Hrinchuk, Valentin Khrulkov, Leyla Mirvakhabova, Elena Orlova and Ivan Oseledets 4847

Speaker or Listener? The Role of a Dialog Agent
Yafei Liu, Hongjin Qian, Hengpeng Xu and Jinmao Wei ............................................. 4861

Bridging Textual and Tabular Data for Cross-Domain Text-to-SQL Semantic Parsing
Xi Victoria Lin, Richard Socher and Caiming Xiong ................................................ 4870

Do Language Embeddings capture Scales?
Xikun Zhang, Deepak Ramachandran, Ian Tenney, Yanai Elazar and Dan Roth ............. 4889

Paraphrasing vs Coreferring: Two Sides of the Same Coin
Yehudit Meged, Avi Caciularu, Vered Shwartz and Ido Dagan .................................... 4897

Active Sentence Learning by Adversarial Uncertainty Sampling in Discrete Space
Dongyu Ru, Jiangaotao Feng, Lin Qiu, Hao Zhou, Mingxuan Wang, Weinan Zhang, Yong Yu and Lei Li ................................................................. 4908

xxx
Coming to Terms: Automatic Formation of Neologisms in Hebrew
Moran Mizrahi, Stav Yardeni Seelig and Dafna Shahaf ........................................... 4918

Dual Inference for Improving Language Understanding and Generation
Shang-Yu Su, Yung-Sung Chuang and Yun-Nung Chen ............................................ 4930

Joint Intent Detection and Entity Linking on Spatial Domain Queries
lei zhang, Runze Wang, Jingbo Zhou, Jingsong Yu, Zhenhua Ling and Hui Xiong ........ 4937

IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages
Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra and Pratyush Kumar ................................................................. 4948

Weakly-Supervised Modeling of Contextualized Event Embedding for Discourse Relations
I-Ta Lee, Maria Leonor Pacheco and Dan Goldwasser ................................................. 4962

Enhancing Generalization in Natural Language Inference by Syntax
Qi He, Han Wang and Yue Zhang ................................................................. 4973