21st Workshop on Biomedical Language Processing (BioNLP 2022)

Dublin, Ireland
26 May 2022

ISBN: 978-1-7138-6728-9

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© (2022) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2023)

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

Explainable Assessment of Healthcare Articles with QA
Alodie Boissonnet, Marzieh Saeidi, Vassilis Plachouras and Andreas Vlachos 1

A sequence-to-sequence approach for document-level relation extraction
John Giorgi, Gary Bader and Bo Wang 10

Position-based Prompting for Health Outcome Generation
Micheal Abaho, Danushka Bollegala, Paula Williamson and Susanna Dodd 26

How You Say It Matters: Measuring the Impact of Verbal Disfluency Tags on Automated Dementia Detection
Shahla Farzana, Ashwin Deshpande and Natalie Parde 37

Zero-Shot Aspect-Based Scientific Document Summarization using Self-Supervised Pre-training
Amir Soleimani, Vassilina Nikoulina, Benoit Favre and Salah Ait Mokhtar 49

Data Augmentation for Biomedical Factoid Question Answering
Dimitris Pappas, Prodromos Malakasiotis and Ion Androutsopoulos 63

Slot Filling for Biomedical Information Extraction
Yannis Papanikolaou, Marlene Staib, Justin Joshua Grace and Francine Bennett 82

Automatic Biomedical Term Clustering by Learning Fine-grained Term Representations
Sihang Zeng, Zheng Yuan and Sheng Yu 91

BioBART: Pretraining and Evaluation of A Biomedical Generative Language Model
Hongyi Yuan, Zheng Yuan, Ruyi Gan, Jiaxing Zhang, Yutao Xie and Sheng Yu 97

Incorporating Medical Knowledge to Transformer-based Language Models for Medical Dialogue Generation
Usman Naseem, Ajay Bandi, Shaina Raza, Junaid Rashid and Bharathi Raja Chakravarthi 110

Memory-aligned Knowledge Graph for Clinically Accurate Radiology Image Report Generation
Sixing Yan 116

Simple Semantic-based Data Augmentation for Named Entity Recognition in Biomedical Texts
Uyen Phan and Nhung Nguyen 123

Auxiliary Learning for Named Entity Recognition with Multiple Auxiliary Biomedical Training Data
Taiki Watanabe, Tomoya Ichikawa, Akihiro Tamura, Tomoya Iwakura, Chunpeng Ma and Tsuneo Kato 130

SNP2Vec: Scalable Self-Supervised Pre-Training for Genome-Wide Association Study
Samuel Cahyawijaya, Tiezheng Yu, Zihan Liu, Xiaopu ZHOU, Tze Wing Tiffany MAK, Yuk Yu Nancy IP and Pascale Fung 140

Biomedical NER using Novel Schema and Distant Supervision
Anshita Khandelwal, Alok Kar, Veera Raghavendra Chikka and Kamalakar Karlapalem 155

Improving Supervised Drug-Protein Relation Extraction with Distantly Supervised Models
Naoki Iinuma, Makoto Miwa and Yutaka Sasaki 161

Named Entity Recognition for Cancer Immunology Research Using Distant Supervision
Hai-Long Trieu, Makoto Miwa and Sophia Ananiadou 171

Intra-Template Entity Compatibility based Slot-Filling for Clinical Trial Information Extraction
Christian Witte and Philipp Cimiano 178

Pretrained Biomedical Language Models for Clinical NLP in Spanish
Casimiro Pio Carrino, Joan Llop, Marc Pàmies, Asier Gutiérrez-Fandiño, Jordi Armengol-Estapé, Joaquín Silveira-Ocampo, Alfonso Valencia, Aitor Gonzalez-Agirre and Marta Villegas 193

Few-Shot Cross-lingual Transfer for Coarse-grained De-identification of Code-Mixed Clinical Texts
Saadullah Amin, Noon Pokaratsiri Goldstein, Morgan Wixted, Alejandro Garcia-Rudolph, Catalina Martínez-Costa and Guenter Neumann 200

VPAI_Lab at MedVidQA 2022: A Two-Stage Cross-modal Fusion Method for Medical Instructional Video Classification
Bin Li, Yixuan Weng, Fei Xia, Bin Sun and Shutao Li 212

GenCompareSum: a hybrid unsupervised summarization method using salience
Jennifer Bishop, Qianqian Xie and Sophia Ananiadou 220

BioCite: A Deep Learning-based Citation Linkage Framework for Biomedical Research Articles
Sudipta Singha Roy and Robert E. Mercer 241

Low Resource Causal Event Detection from Biomedical Literature
Zhengzhong Liang, Enrique Noriega-Atala, Clayton Morrison and Mihai Surdeanu 252

Overview of the MedVidQA 2022 Shared Task on Medical Video Question-Answering
Deepak Gupta and Dina Demner-Fushman 264

Inter-annotator agreement is not the ceiling of machine learning performance: Evidence from a comprehensive set of simulations
Russell Richie, Sachin Grover and Fuchiang (Rich) Tsui 275

Conversational Bots for Psychotherapy: A Study of Generative Transformer Models Using Domain-specific Dialogues
Avisha Das, Salih Selek, Alia R. Warner, Xu Zuo, Yan Hu, Vipina Kuttichi Keloth, Jianfu Li, W. Jim Zheng and Hua Xu 285

BEEDS: Large-Scale Biomedical Event Extraction using Distant Supervision and Question Answering
Xing David Wang, Ulf Leser and Leon Weber 298

Data Augmentation for Rare Symptoms in Vaccine Side-Effect Detection
Bosung Kim and Ndapa Nakashole 310

Improving Romanian BioNER Using a Biologically Inspired System
Maria Mitrofan and Vasile Pais 316

BanglaBioMed: A Biomedical Named-Entity Annotated Corpus for Bangla (Bengali)
Salim Sazzed 323

ICDBigBird: A Contextual Embedding Model for ICD Code Classification
George Michalopoulos, Michal Malyska, Nicola Sahar, Alexander Wong and Helen Chen 330

Doctor XAvIer: Explainable Diagnosis on Physician-Patient Dialogues and XAI Evaluation
Hillary Ngai and Frank Rudzicz 337

DISTANT-CTO: A Zero Cost, Distantly Supervised Approach to Improve Low-Resource Entity Extraction Using Clinical Trials Literature
Anjani Dhrangadhariya and Henning Müller 345

EchoGen: Generating Conclusions from Echocardiogram Notes
Liyan Tang, Shravan Kooragayalu, Yanshan Wang, Ying Ding, Greg Durrett, Justin F. Rousseau and Yifan Peng 359

Quantifying Clinical Outcome Measures in Patients with Epilepsy Using the Electronic Health Record
Kevin Xie, Brian Litt, Dan Roth and Colin A. Ellis 369

Comparing Encoder-Only and Encoder-Decoder Transformers for Relation Extraction from Biomedical Texts: An Empirical Study on Ten Benchmark Datasets
Mourad Sarrouti, Carson Tao and Yoann Mamy Randriamihaja 376

Utility Preservation of Clinical Text After De-Identification
Thomas Vakili and Hercules Dalianis 383

Horses to Zebras: Ontology-Guided Data Augmentation and Synthesis for ICD-9 Coding
Matúš Falis, Hang Dong, Alexandra Birch and Beatrice Alex 389

Towards Automatic Curation of Antibiotic Resistance Genes via Statement Extraction from Scientific Papers: A Benchmark Dataset and Models
Sidhant Chandak, Liqing Zhang, Connor Brown and Lifu Huang 402

Model Distillation for Faithful Explanations of Medical Code Predictions
Zach Wood-Doughty, Isabel Cachola and Mark Dredze 412

Towards Generalizable Methods for Automating Risk Score Calculation
Jennifer J Liang, Eric Lehman, Ananya Iyengar, Diwakar Mahajan, Preethi Raghavan, Cindy Y. Chang and Peter Szolovits 426

DoSSIER at MedVidQA 2022: Text-based Approaches to Medical Video Answer Localization Problem
Wojciech Kusa, Georgios Peikos, Óscar Espitia, Allan Hanbury and Gabriella Pasi 432