EMNLP 2020

The 3rd Clinical Natural Language Processing Workshop

Proceedings of the Workshop

November 19, 2020

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-952148-74-3

Introduction

This volume contains papers from the 3rd Workshop on Clinical Natural Language Processing (ClinicalNLP), held at EMNLP 2020.

Clinical text offers unique challenges that differentiate it not only from open-domain data, but from other types of text in the biomedical domain as well. Notably, clinical text contains a significant number of abbreviations, medical terms, and other clinical jargon. Clinical narratives are characterized by nonstandard document structures that are often critical to overall understanding. Narrative provider notes are designed to communicate with other experts while at the same time serving as a legal record. Finally, clinical notes contain sensitive patient-specific information that raise privacy and security concerns that present special challenges for natural language systems. This workshop focuses on the work that develops methods to address the above challenges, with the goal of advancing state-of-the-art in clinical NLP.

This year, we received the total of 48 submissions, out of which 14 were accepted as oral presentations and 19 as posters.

Organizers:

Anna Rumshisky, UMass Lowell
Kirk Roberts, University of Texas Health Science Center at Houston
Steven Bethard, University of Arizona
Tristan Naumann, Microsoft Research

Program Committee:

John Aberdeen, The MITRE Corporation
Emily Alsentzer, Massachusetts Institute of Technology
Sabine Bergler, Concordia University
Parminder Bhatia, Amazon
William Boag, Massachusetts Institute of Technology
Cheryl Clark, The MITRE Corporation
Surabhi Datta, University of Texas Health Science Center at Houston
Murthy Devarakonda, Arizona State University
Dmitriy Dligach, Loyola University
Jungwei Fan, Mayo Clinic
Jason Fries, Stanford University
Travis Goodwin, National Library of Medicine
Yoan Gutiérrez, University of Alicante
Sadid Hasan, CVS Health
Saeed Hassanpour, Dartmouth University
Djoerd Hiemstra, Radboud University
Eben Holderness, Brandeis University
Baotian Hu, Harbin Institute of Technology
Di Jin, Massachusetts Institute of Technology
Alistair Johnson, Massachusetts Institute of Technology
Yoshinobu Kano, Shizuoka University
Egoitz Laparra, University of Arizona
Chen Lin, Boston Children's Hospital
Sijia Liu, Mayo Clinic
Matthew McDermott, Massachusetts Institute of Technology
Bridget McInnes, Virginia Commonwealth University
Timothy Miller, Boston Children's Hospital
Yifan Peng, Weill Cornell Medicine
Hoifung Poon, Microsoft Research
Preethi Raghavan, IBM Research
Frank Rudzicz, University of Toronto
Allen Schmaltz, Harvard University
Thomas Searle, King's College London
Chaitanya Shivade, Amazon
Yuqi Si, University of Texas Health Science Center at Houston
Sarvesh Soni, University of Texas Health Science Center at Houston
Karin Verspoor, The University of Melbourne
Byron Wallace, Northeastern University
Yanshan Wang, Mayo Clinic
Ben Wellner, The MITRE Corporation
Stephen Wu, University of Texas Health Science Center at Houston
Dongfang Xu, University of Arizona
Rui Zhang, University of Minnesota

Invited Speaker:

Hong Yu, UMass Lowell

Talk Title: Deep Learning is Conquering Human Tasks. How Far Can it Go in Medicine? Advances, Challenges, and Future Directions

Biography: Hong Yu is faculty at the Department of Computer Science and Director of Biomedical and Health Research in Data Sciences, University of Massachusetts Lowell. She is also a Research Health Scientist at the Edith Nourse Rogers Memorial Veterans Hospital. She holds adjunct faculty positions at UMass Amherst and UMass Medical School. She is an elected fellow of the American College of Medical Informatics. She has published over 200 peer-reviewed articles in natural language processing and applied machine learning.

Table of Contents

Conference Program

19 Nov 2020

09:00–10:00 Opening Session

09:00–09:45 *Keynote Talk: Deep Learning is Conquering Human Tasks. How Far Can it Go in Medicine? Advances, Challenges, and Future Directions*
Hong Yu

09:45–10:00 *Keynote Q&A*
Hong Yu

10:00–10:40 Oral Session 1

10:00–10:05 *Various Approaches for Predicting Stroke Prognosis using Magnetic Resonance Imaging Text Records*
Tak-Sung Heo, Chulho Kim, Jeong-Myeong Choi, Yeong-Seok Jeong and Yu-Seop Kim

10:05–10:15 *Multiple Sclerosis Severity Classification From Clinical Text*
Alister D'Costa, Stefan Denkovski, Michal Malyska, Sae Young Moon, Brandon Rufino, Zhen Yang, Taylor Killian and Marzyeh Ghassemi

10:15–10:25 *BERT-XML: Large Scale Automated ICD Coding Using BERT Pretraining*
Zachariah Zhang, Jingshu Liu and Narges Razavian

10:25–10:40 *Oral Session 1: Q&A*

19 Nov 2020 (continued)

10:40–10:50 Coffee Break

10:50–11:40 Oral Session 2

10:50–10:55 *Incorporating Risk Factor Embeddings in Pre-trained Transformers Improves Sentiment Prediction in Psychiatric Discharge Summaries*
Xiyu Ding, Mei-Hua Hall and Timothy Miller

10:55–11:05 *Information Extraction from Swedish Medical Prescriptions with Sig-Transformer Encoder*
John Pougué Biyong, Bo Wang, Terry Lyons and Alejo Nevado-Holgado

11:05–11:15 *Evaluation of Transfer Learning for Adverse Drug Event (ADE) and Medication Entity Extraction*
Sankaran Narayanan, Kaivalya Mannam, Sreeranga P Rajan and P Venkat Rangan

11:15–11:25 *BioBERTpt - A Portuguese Neural Language Model for Clinical Named Entity Recognition*
Elisa Terumi Rubel Schneider, João Vitor Andrioli de Souza, Julien Knafou, Lucas Emanuel Silva e Oliveira, Jenny Copara, Yohan Bonescki Gumiel, Lucas Ferro Antunes de Oliveira, Emerson Cabrera Paraiso, Douglas Teodoro and Cláudia Maria Cabral Moro Barra

11:25–11:40 ***Oral Session 2: Q&A***

11:40–12:00 Poster Teasers

11:40–11:41 *Dilated Convolutional Attention Network for Medical Code Assignment from Clinical Text*
Shaoxiong Ji, Erik Cambria and Pekka Marttinen

11:41–11:42 *Classification of Syncope Cases in Norwegian Medical Records*
Ildiko Pilan, Pål H. Brekke, Fredrik A. Dahl, Tore Gundersen, Haldor Husby, Øystein Nytrø and Lilja Øvrelid

11:42–11:43 *Comparison of Machine Learning Methods for Multi-label Classification of Nursing Education and Licensure Exam Questions*
John Langton, Krishna Srihasam and Junlin Jiang

11:43–11:44 *Clinical XLNet: Modeling Sequential Clinical Notes and Predicting Prolonged Mechanical Ventilation*
Kexin Huang, Abhishek Singh, Sitong Chen, Edward Moseley, Chih-Ying Deng, Naomi George and Charolotta Lindvall

19 Nov 2020 (continued)

11:44–11:45 *Automatic recognition of abdominal lymph nodes from clinical text*
Yifan Peng, Sungwon Lee, Daniel C. Elton, Thomas Shen, Yu-xing Tang, Qingyu Chen, Shuai Wang, Yingying Zhu, Ronald Summers and Zhiyong Lu

11:45–11:46 *How You Ask Matters: The Effect of Paraphrastic Questions to BERT Performance on a Clinical SQuAD Dataset*
Sungrim (Riea) Moon and Jungwei Fan

11:46–11:47 *Relative and Incomplete Time Expression Anchoring for Clinical Text*
Louise Dupuis, Nicol Bergou, Hegler Tissot and Sumithra Velupillai

11:47–11:48 *MeDAL: Medical Abbreviation Disambiguation Dataset for Natural Language Understanding Pretraining*
Zhi Wen, Xing Han Lu and Siva Reddy

11:48–11:49 *Knowledge Grounded Conversational Symptom Detection with Graph Memory Networks*
Hongyin Luo, Shang-Wen Li and James Glass

11:49–11:50 *Pretrained Language Models for Biomedical and Clinical Tasks: Understanding and Extending the State-of-the-Art*
Patrick Lewis, Myle Ott, Jingfei Du and Veselin Stoyanov

11:50–11:51 *Assessment of DistilBERT performance on Named Entity Recognition task for the detection of Protected Health Information and medical concepts*
Macarious Abadeer

11:51–11:52 *Distinguishing between Dementia with Lewy bodies (DLB) and Alzheimer's Disease (AD) using Mental Health Records: a Classification Approach*
Zixu Wang, Julia Ive, Sinead Moylett, Christoph Mueller, Rudolf Cardinal, Sumithra Velupillai, John O'Brien and Robert Stewart

11:52–11:53 *Weakly Supervised Medication Regimen Extraction from Medical Conversations*
Dhruvesh Patel, Sandeep Konam and Sai Prabhakar

11:53–11:54 *Extracting Relations between Radiotherapy Treatment Details*
Danielle Bitterman, Timothy Miller, David Harris, Chen Lin, Sean Finan, Jeremy Warner, Raymond Mak and Guergana Savova

11:54–11:55 *Cancer Registry Information Extraction via Transfer Learning*
Yan-Jie Lin, Hong-Jie Dai, You-Chen Zhang, Chung-Yang Wu, Yu-Cheng Chang, Pin-Jou Lu, Chih-Jen Huang, Yu-Tsang Wang, Hui-Min Hsieh, Kun-San Chao, Tsang-Wu Liu, I-Shou Chang, Yi-Hsin Connie Yang, Ti-Hao Wang, Ko-Jiunn Liu, Li-Tzong Chen and Sheau-Fang Yang

11:55–11:56 *PHICON: Improving Generalization of Clinical Text De-identification Models via Data Augmentation*
Xiang Yue and Shuang Zhou

19 Nov 2020 (continued)

11:56–11:57 *Where's the Question? A Multi-channel Deep Convolutional Neural Network for Question Identification in Textual Data*
George Michalopoulos, Helen Chen and Alexander Wong

11:57–11:58 *Learning from Unlabelled Data for Clinical Semantic Textual Similarity*
Yuxia Wang, Karin Verspoor and Timothy Baldwin

11:58–11:59 *Joint Learning with Pre-trained Transformer on Named Entity Recognition and Relation Extraction Tasks for Clinical Analytics*
Miao Chen, Ganhui Lan, Fang Du and Victor Lobanov

12:00–13:20 Poster Session & Lunch

13:20–14:10 Oral Session 3

13:20–13:25 *Extracting Semantic Aspects for Structured Representation of Clinical Trial Eligibility Criteria*
Tirthankar Dasgupta, Ishani Mondal, abir naskar and Lipika Dey

13:25–13:35 *An Ensemble Approach to Automatic Structuring of Radiology Reports*
Morteza Pourreza Shahri, Amir Tahmasebi, Bingyang Ye, Henghui Zhu, Javed Aslam and Timothy Ferris

13:35–13:45 *Utilizing Multimodal Feature Consistency to Detect Adversarial Examples on Clinical Summaries*
Wenjie Wang, Youngja Park, Taesung Lee, Ian Molloy, Pengfei Tang and Li Xiong

13:45–13:55 *Advancing Seq2seq with Joint Paraphrase Learning*
So Yeon Min, Preethi Raghavan and Peter Szolovits

13:55–14:10 ***Oral Session 3: Q&A***

19 Nov 2020 (continued)

14:10–14:55 EMNLP Findings Session 1

14:10–14:15 *Learning to Generate Clinically Coherent Chest X-Ray Reports*
Justin Lovelace, Bobak Mortazavi

14:15–14:20 *Learning Visual-Semantic Embeddings for Reporting Abnormal Findings on Chest X-rays*
Jianmo Ni, Chun-Nan Hsu, Amilcare Gentili, Julian McAuley

14:20–14:30 *Characterizing the Value of Information in Medical Notes*
Chao-Chun Hsu, Shantanu Karnwal, Sendhil Mullainathan, Ziad Obermeyer, Chenhao Tan

14:30–14:40 *PharmMT: A Neural Machine Translation Approach to Simplify Prescription Directions*
Jiazhao Li, Corey Lester, Xinyan Zhao, Yuting Ding, Yun Jiang, V.G.Vinod Vydiswaran

14:40–14:55 *EMNLP Findings Session 1: Q&A*

14:55–15:05 Coffee Break

15:05–15:50 EMNLP Findings Session 2

15:05–15:15 *A Dual-Attention Network for Joint Named Entity Recognition and Sentence Classification of Adverse Drug Events*
Susmitha Wunnava, Xiao Qin, Tabassum Kakar, Xiangnan Kong, Elke A. Rundensteiner

15:15–15:25 *Dr. Summarize: Global Summarization of Medical Dialogue by Exploiting Local Structures*
Anirudh Joshi, Namit Katariya, Xavier Amatriain, Anitha Kannan

15:25–15:35 *Generating Accurate Electronic Health Assessment from Medical Graph*
Zhichao Yang, Hong Yu

15:35–15:50 *EMNLP Findings Session 2: Q&A*

19 Nov 2020 (continued)

15:50–16:35 Best Paper Session

15:50–16:00 *On the diminishing return of labeling clinical reports*
Jean-Baptiste Lamare, Oloruntobiloba Olatunji and Li Yao

16:00–16:10 *The Chilean Waiting List Corpus: a new resource for clinical Named Entity Recognition in Spanish*
Pablo Báez, Fabián Villena, Matías Rojas, Manuel Durán and Jocelyn Dunstan

16:10–16:20 *Analyzing Text Specific vs Blackbox Fairness Algorithms in Multimodal Clinical NLP*
John Chen, Ian Berlot-Attwell, Xindi Wang, Safwan Hossain and Frank Rudzicz

16:20–16:35 ***Best Paper Session: Q&A***

16:35–16:50 Concluding Session