ClinicalNLP 2016

Clinical Natural Language Processing Workshop

Proceedings of the Workshop

December 11, 2016
Osaka, Japan

Copyright of each paper stays with the respective authors (or their employers).

ISBN978-4-87974-710-5

Preface

This volume contains papers from the Workshop on Clinical Natural Language Processing (ClinicalNLP).

Clinical text offers unique challenges that differentiate it both from open-domain data and from other types of text in the biomedical domain. Narrative clinical notes are written by experts for other experts, while also intended to serve as a legal record. Clinical jargon, non-standard document strucuture, privacy and security concerns all present special challenges for natural language systems. The workshop's goal is to attract more quality NLP researchers to the clinical domain, and thereby to allow us as a community to contribute more to the advancement of critical clinical research.

Organizers

Anna Rumshisky
Kirk Roberts
Steven Bethard
Tristan Naumann

Program Committee

Sabine Bergler
Kevin Cohen
Dina Demner-Fushman
Dmitriy Dligach
Oladimeji Farri
Michele Filannino
Nazli Goharian
Cyril Grouin
Sanda Harabagiu
Sadid Hasan
Lynette Hirschman
Guergana Savova
Hongfang Liu
Stephane Meystre
Timothy Miller
Marie-Francine Moens
Danielle Mowery
Ted Pedersen
Amber Stubbs
Sumithra Velupillai
Ben Wellner
Jenna Wiens
Stephen Wu
Hua Xu
Pierre Zweigenbaum

Table of Contents

The impact of simple feature engineering in multilingual medical NER
Rebecka Weegar, Arantza Casillas, Arantza Diaz de Ilarraza, Maite Oronoz, Alicia Pérez and Koldo Gojenola 1

Bidirectional LSTM-CRF for Clinical Concept Extraction
Raghavendra Chalapathy, Ehsan Zare Borzeshi and Massimo Piccardi 7

MedNLPDoc: Japanese Shared Task for Clinical NLP
Eiji Aramaki, Yoshinobu Kano, Tomoko Ohkuma and Mizuki Morita 13

Feature-Augmented Neural Networks for Patient Note De-identification
Ji Young Lee, Franck Dernoncourt, Ozlem Uzuner and Peter Szolovits 17

Semi-supervised Clustering of Medical Text
Pracheta Sahoo, Asif Ekbal, Sriparna Saha, Diego Molla and Kaushik Nandan 23

Deep Learning Architecture for Patient Data De-identification in Clinical Records
Shweta Yadav, Asif Ekbal, Sriparna Saha and Pushpak Bhattacharyya 32

Neural Clinical Paraphrase Generation with Attention
Sadid A. Hasan, Bo Liu, Joey Liu, Ashequl Qadir, Kathy Lee, Vivek Datla, Aaditya Prakash and Oladimeji Farri 42

Assessing the Corpus Size vs. Similarity Trade-off for Word Embeddings in Clinical NLP
Kirk Roberts 54

Inference of ICD Codes from Japanese Medical Records by Searching Disease Names
Masahito Sakishita and Yoshinobu Kano 64

A fine-grained corpus annotation schema of German nephrology records
Roland Roller, Hans Uszkoreit, Feiyu Xu, Laura Seiffe, Michael Mikhailov, Oliver Staeck, Klemens Budde, Fabian Halleck and Danilo Schmidt 69

Detecting Japanese Patients with Alzheimer's Disease based on Word Category Frequencies
Daisaku Shibata, Shoko Wakamiya, Ayae Kinoshita and Eiji Aramaki 78

Prediction of Key Patient Outcome from Sentence and Word of Medical Text Records
Takanori Yamashita, Yoshifumi Wakata, Hidehisa Soejima, Naoki Nakashima and Sachio Hirokawa 86

Unsupervised Abbreviation Detection in Clinical Narratives
Markus Kreuzthaler, Michel Oleynik, Alexander Avian and Stefan Schulz 91

Automated Anonymization as Spelling Variant Detection
Steven Kester Yuwono, Hwee Tou Ng and Kee Yuan Ngiam 99

Conference Program

Sunday, December 11, 2016

09:00–09:30 ***Introductory Remarks***

09:30–10:30 *Invited Talk*
Tim Baldwin

10:30–10:45 Coffee Break

10:45–11:30 Short Talks

10:45–11:00 *The impact of simple feature engineering in multilingual medical NER*
Rebecka Weegar, Arantza Casillas, Arantza Diaz de Ilarraza, Maite Oronoz, Alicia Pérez and Koldo Gojenola

11:00–11:15 *Bidirectional LSTM-CRF for Clinical Concept Extraction*
Raghavendra Chalapathy, Ehsan Zare Borzeshi and Massimo Piccardi

11:15–11:30 *MedNLPDoc: Japanese Shared Task for Clinical NLP*
Eiji Aramaki, Yoshinobu Kano, Tomoko Ohkuma and Mizuki Morita

11:30–12:15 Best Student Paper Award Session

11:30–11:45 *Feature-Augmented Neural Networks for Patient Note De-identification*
Ji Young Lee, Franck Dernoncourt, Ozlem Uzuner and Peter Szolovits

11:45–12:05 *Semi-supervised Clustering of Medical Text*
Pracheta Sahoo, Asif Ekbal, Sriparna Saha, Diego Molla and Kaushik Nandan

Sunday, December 11, 2016 (continued)

12:15–14:00 Lunch

14:00–15:00 Long Talks

14:00–14:20 *Deep Learning Architecture for Patient Data De-identification in Clinical Records*
Shweta Yadav, Asif Ekbal, Sriparna Saha and Pushpak Bhattacharyya

14:20–14:40 *Neural Clinical Paraphrase Generation with Attention*
Sadid A. Hasan, Bo Liu, Joey Liu, Ashequl Qadir, Kathy Lee, Vivek Datla, Aaditya Prakash and Oladimeji Farri

14:40–15:00 *Assessing the Corpus Size vs. Similarity Trade-off for Word Embeddings in Clinical NLP*
Kirk Roberts

15:00–15:45 Poster Session

15:00–15:45 *Inference of ICD Codes from Japanese Medical Records by Searching Disease Names*
Masahito Sakishita and Yoshinobu Kano

15:00–15:45 *A fine-grained corpus annotation schema of German nephrology records*
Roland Roller, Hans Uszkoreit, Feiyu Xu, Laura Seiffe, Michael Mikhailov, Oliver Staeck, Klemens Budde, Fabian Halleck and Danilo Schmidt

15:00–15:45 *Detecting Japanese Patients with Alzheimer's Disease based on Word Category Frequencies*
Daisaku Shibata, Shoko Wakamiya, Ayae Kinoshita and Eiji Aramaki

15:00–15:45 *Prediction of Key Patient Outcome from Sentence and Word of Medical Text Records*
Takanori Yamashita, Yoshifumi Wakata, Hidehisa Soejima, Naoki Nakashima and Sachio Hirokawa

15:00–15:45 *Unsupervised Abbreviation Detection in Clinical Narratives*
Markus Kreuzthaler, Michel Oleynik, Alexander Avian and Stefan Schulz

15:00–15:45 *Automated Anonymization as Spelling Variant Detection*
Steven Kester Yuwono, Hwee Tou Ng and Kee Yuan Ngiam

Sunday, December 11, 2016 (continued)

15:45–16:00 Coffee Break

16:00–17:00 Panel on Facilitating Access to Clinical Data for the NLP Community