ACL 2014

BioNLP 2014
Workshop on Biomedical Natural Language Processing

Proceedings of the Workshop

June 27-28, 2014
Baltimore, Maryland, USA

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 978-1-941643-18-1

Introduction

The first day of the BioNLP 2014 workshop continues following the course set by the first ACL workshop on Natural Language Processing in the Biomedical Domain that was held in 2002: BioNLP 2014 provides a venue for exploring challenges and techniques in processing biomedical language and brings together researchers from computational linguistics and biomedical informatics. The submissions to the first day of 2014 workshop organized by SIGBioMed were traditionally very strong and continued demonstrating the considerable breadth of research in biomedical language processing. The 2014 workshop has accepted 12 full and short papers for oral presentations and 7 posters. The first day of the workshop features a keynote that expands the scope of BioNLP beyond its already remarkable breadth

Keynote BioNLP as the Pioneering field of linking text, knowledge and data

Professor Jun'ichi Tsujii, Principal Researcher at Microsoft Research Asia (MSRA), Chair of Text Mining and Scientific Director of the National Centre for Text Mining (NaCTeM) at the University of Manchester, UK

The second day of the workshop features a paper submitted to the special track on NLP approaches for assessment of clinical conditions. Kathleen C. Fraser presents the featured talk on using statistical parsing to detect agrammatic aphasia. The track organizers, Thamar Solorio and Yang Liu, serve as discussants.

The second day further features an exciting panel that brings together organizers of several shared tasks in biomedical information retrieval and natural language processing. The panel introduces the workshop participants to the long-standing and relatively new community-wide challenges in biomedical and clinical language processing. It also provides an opportunity to discuss the future of the shared tasks in this domain.

Panel Life cycles of BioCreative, BioNLP-ST, i2b2, TREC Medical tracks, and ShARe /CLEF/ SemEval

Lynette Hirschman & John Wilbur, Sophia Ananiadou, Ellen Voorhees, Ozlem Uzuner, Danielle Mowery & Sumithra Velupillai & Sameer Pradhan

The second day of the BioNLP 2014 workshop concludes with two tutorials on the fundamental resources widely used in the biomedical domain.

Tutorial 1 UMLS in biomedical text processing

Olivier Bodenreider, Branch Chief, Cognitive Science Branch, LHNCBC, NLM, NIH

Tutorial 2 Using MetaMap Alan R. Aronson, Senior Researcher, Cognitive Science Branch, LHNCBC, NLM, NIH

Acknowledgments

As always, we are profoundly grateful to the authors who chose BioNLP as venue for presenting their innovative research. The authors' willingness to share their work through BioNLP consistently makes the workshop noteworthy among the increasing numbers of available venues. We are equally indebted to the program committee members (listed elsewhere in this volume) who produced at least three thorough reviews per paper on a tight review schedule and with an admirable level of insight.

Organizers:

Kevin Bretonnel Cohen, University of Colorado School of Medicine
Dina Demner-Fushman, US National Library of Medicine
Sophia Ananiadou, University of Manchester and National Centre for Text Mining, UK
John Pestian, University of Cincinnati, Cincinnati Children's Hospital Medical Center
Jun'ichi Tsujii, Microsoft Research Asia and National Centre for Text Mining, UK

Program Committee:

Emilia Apostolova, DePaul University, USA
Eiji Aramaki, University of Tokyo, Japan
Alan Aronson, US National Library of Medicine
Sabine Bergler, Concordia University, Canada
Olivier Bodenreider, US National Library of Medicine
Kevin Cohen, University of Colorado, USA
Nigel Collier, National Institute of Informatics, Japan
Dina Demner-Fushman, US National Library of Medicine
Marcelo Fiszman, US National Library of Medicine
Filip Ginter, University of Turku, Finland
Graciela Gonzalez, Arizona State University, USA
Antonio Jimeno Yepes, NICTA, Australia
Halil Kilicoglu, US National Library of Medicine
Jin-Dong Kim, University of Tokyo, Japan
Robert Leaman, US National Library of Medicine
Yang Liu, The University of Texas at Dallas, USA
Zhiyong Lu, US National Library of Medicine
Makoto Miwa, National Centre for Text Mining, UK
Aurelie Neveol, LIMSI, France
Naoaki Okazaki, Tohoku University, Japan
Jong Park, KAIST, South Korea
Rashmi Prasad, University of Wisconsin-Milwaukee, USA
Sampo Pyysalo, National Centre for Text Mining, UK
Bastien Rance, Georges Pompidou European Hospital, France
Thomas Rindflesch, US National Library of Medicine
Kirk Roberts, US National Library of Medicine
Andrey Rzhetsky, University of Chicago, USA
Matthew Simpson, US National Library of Medicine
Thamar Solorio, The University of Alabama at Birmingham, USA
Yoshimasa Tsuruoka, University of Tokyo, Japan
Karin Verspoor, NICTA, Australia
W. John Wilbur, US National Library of Medicine

Invited Speaker:

Jun'ichi Tsujii, Microsoft Research Asia and National Centre for Text Mining, UK

Panelists

Sophia Ananiadou, University of Manchester and National Centre for Text Mining, UK
Lynette Hirschman, The MITRE Corporation , USA

Danielle Mowery, University of Pittsburgh, USA
Sameer Pradhan, Harvard Medical School, USA
Ozlem Uzuner, State University of New York, Albany, USA
Sumithra Velupillai, Stockholm University, Sweden
Ellen Voorhees, National Institute of Standards and Technology, USA
W. John Wilbur, US National Library of Medicine

Table of Contents

Conference Program

Thursday, June 26, 2014

9:00–9:10 Opening remarks

Session 1: Processing biomedical publications

9:10–9:30 *Natural Language Processing Methods for Enhancing Geographic Metadata for Phylogeography of Zoonotic Viruses*
Tasnia Tahsin, Robert Rivera, Rachel Beard, Rob Lauder, Davy Weissenbacher, Matthew Scotch, Garrick Wallstrom and Graciela Gonzalez

9:30–9:50 *Temporal Expression Recognition for Cell Cycle Phase Concepts in Biomedical Literature*
Negacy Hailu, Natalya Panteleyeva and Kevin Cohen

9:50–10:10 *Classifying Negative Findings in Biomedical Publications*
Bei Yu

10:10–10:30 *Automated Disease Normalization with Low Rank Approximations*
Robert Leaman and Zhiyong Lu

10:30–11:00 Coffee Break

Keynote by Junichi Tsujii

11:00–11:50 BioNLP as the Pioneering field of linking text, knowledge and data

Session 2: Processing consumer language

11:50–12:10 *Decomposing Consumer Health Questions*
Kirk Roberts, Halil Kilicoglu, Marcelo Fiszman and Dina Demner-Fushman

12:10–12:30 *Detecting Health Related Discussions in Everyday Telephone Conversations for Studying Medical Events in the Lives of Older Adults*
Golnar Sheikhshab, Izhak Shafran and Jeffrey Kaye

12:30–14:00 Lunch

Thursday, June 26, 2014 (continued)

Session 3: Processing clinical text and gray literature

14:00–14:20 *Coreference Resolution for Structured Drug Product Labels*
Halil Kilicoglu and Dina Demner-Fushman

14:20–14:40 *Generating Patient Problem Lists from the ShARe Corpus using SNOMED CT/SNOMED CT CORE Problem List*
Danielle Mowery, Mindy Ross, Sumithra Velupillai, Stephane Meystre, Janyce Wiebe and Wendy Chapman

14:40–15:00 *A System for Predicting ICD-10-PCS Codes from Electronic Health Records*
Michael Subotin and Anthony Davis

15:00–15:20 *Structuring Operative Notes using Active Learning*
Kirk Roberts, Sanda Harabagiu and Michael Skinner

15:30–16:00 Afternoon Break

16:00–16:20 *Chunking Clinical Text Containing Non-Canonical Language*
Aleksandar Savkov, John Carroll and Jackie Cassell

16:20–16:40 *Decision Style in a Clinical Reasoning Corpus*
Limor Hochberg, Cecilia Ovesdotter Alm, Esa M. Rantanen, Caroline M. DeLong and Anne Haake

(16:40–17:30) Poster session

Temporal Expressions in Swedish Medical Text – A Pilot Study
Sumithra Velupillai

A repository of semantic types in the MIMIC II database clinical notes
Richard Osborne, Alan Aronson and Kevin Cohen

Extracting drug indications and adverse drug reactions from Spanish health social media
Isabel Segura-Bedmar, Santiago de la Peña González and Paloma Martínez

Symptom extraction issue
Laure Martin, Delphine Battistelli and Thierry Charnois

Thursday, June 26, 2014 (continued)

Seeking Informativeness in Literature Based Discovery
Judita Preiss

Towards Gene Recognition from Rare and Ambiguous Abbreviations using a Filtering Approach
Matthias Hartung, Roman Klinger, Matthias Zwick and Philipp Cimiano

FFTM: A Fuzzy Feature Transformation Method for Medical Documents
Amir Karami and Aryya Gangopadhyay

Friday, June 27, 2014

Session 1: NLP approaches for assessment of clinical conditions

9:00–9:40 *Using statistical parsing to detect agrammatic aphasia*
Kathleen C. Fraser, Graeme Hirst, Jed A. Meltzer, Jennifer E. Mack and Cynthia K. Thompson

Panel: Life cycles of BioCreative, BioNLP-ST, i2b2, TREC Medical tracks, and ShARe /CLEF/ SemEval

9:40–10:05 BioCreative by Lynette Hirschman and John Wilbur

10:05–10:30 BioNLP-ST by Sophia Ananiadou and Junichi Tsujii

10:30–11:00 Coffee Break

11:00–11:25 TREC Medical tracks by Ellen Voorhees

11:25–11:50 i2b2 by Ozlem Uzuner

11:50–12:10 ShARe/CLEF/SemEval by Danielle Mowery, Sumithra Velupillai and Sameer Pradhan

12:10–12:30 Discussion

12:30–14:00 Lunch

Friday, June 27, 2014 (continued)

Tutorials

14:00–15:30 UMLS in biomedical text processing by Olivier Bodenreider

15:30–16:00 Afternoon Break

16:00–17:30 Using MetaMap by Alan R. Aronson