BioTxtM 2016

Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining

Proceedings of the Workshop

December 11-16, 2016
Osaka, Japan

ISBN978-4-87974-719-8

Preface

Biomedical natural language processing has grown from its roots in clinical language processing and bioinformatics into a thriving research field of its own. The search *("natural language processing") OR ("text mining")* performed in PubMed today returns 5,056 hits, versus 1,903 at the turn of the century and 3,485 in 2010. The papers appearing in this volume reflect the diversity of trends in biomedical natural language processing today—movement from English-language texts to clinical texts in other languages; exploration of social media in addition to clinical documents and traditional scientific publications; and processing of full-text articles, versus abstracts only. In addition to reflecting the diversity of the field, the papers in this volume also reflect the homogenisation of approaches that has characterised some recent approaches, with 4 out of 15 papers involving some combination of neural networks and/or distributional semantics. The organisers thank the authors for sharing their science with this community, and the programme committee (listed elsewhere in this volume) for their contribution to maintaining the high standards of the BioTxtM series of meetings.

Keynote Talk by Dr. Makoto Miwa

Learning for Information Extraction in Biomedical and General Domains

Information extraction (IE) has been widely studied in various domains since IE is a key to bridge the gap between knowledge and texts. IE includes several core sub-problems, such as named entity recognition, relation extraction, and event extraction, and these sub-problems have been tackled using machine learning techniques. In this talk, I will give an overview of learning approaches for IE in biomedical and general domain, especially on corpus-based classification and structured learning approaches. I will then introduce recent deep learning approaches including our recent recurrent neural network (RNN)-based approach, and discuss the limitations and future directions.

Speaker Biography
Makoto Miwa is an associate professor of Toyota Technological Institute (TTI). He received his Ph.D. from the University of Tokyo in 2008. His research mainly focuses on information extraction from texts, deep learning, and representation learning. His projects include AkaneRE, EventMine, PathText and LSTM-ER.

Organisers

Sophia Ananiadou, National Centre for Text Mining, University of Manchester UK

Riza Batista-Navarro, National Centre for Text Mining, University of Manchester UK

Kevin Bretonnel Cohen, Computational Bioscience Program, University of Colorado School of Medicine, USA

Dina Demner-Fushman, National Library of Medicine, USA

Paul Thompson, National Centre for Text Mining, University of Manchester, UK

Programme Committee

Eiji Aramaki, Nara Institute of Science and Technology (NAIST), Japan

Hercules Dalianis, Stockholm University, Sweden

Graciela Gonzalez, Arizona State University, USA

Wen-Lian Hsu, Academia Sinica, Taipei, Taiwan

Rezarta Islamaj, NCBI/NLM/NIH, USA

Roman Klinger, University of Stuttgart, Germany

Robert Leaman NCBI/NLM/NIH, USA

Shervin Malmasi, Harvard Medical School, USA

Makoto Miwa, Toyota Technological Institute, Japan

Sung-Hyon Myaeng, Korea Advanced Institute of Science and Technology (KAIST), Korea

Claire Nedellec, French National Institute of Agronomy (INRA), France

Naoaki Okazaki, Tohoku University, Japan

Arzucan Özgür, Bogazici University, Turkey

Martha Palmer, University of Colorado at Boulder, USA

Stelios Piperidis, Institute for Language and Speech Processing, Greece

Guergana Savova, Boston Children's Hospital and Harvard Medical School, USA

Hagit Shatkay, University of Delaware, USA

Mark Stevenson, University of Sheffield, UK

Yoshimasa Tsuruoka, University of Tokyo, Japan

Lucy Vanderwende, Microsoft, USA

Karin Verspoor, University of Melbourne, Australia

Stephen Wu, Oregon Health & Science University, USA

Yan Xu, Microsoft Research Asia, China

Pierre Zweigenbaum, Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI), France

Table of Contents

Conference Program

12th December 2016

9:00–9:10 Welcome remarks

9:10–10:20 Session 1

9:10–9:30 *Cancer Hallmark Text Classification Using Convolutional Neural Networks*
Simon Baker, Anna Korhonen and Sampo Pyysalo

9:30–9:50 *Learning Orthographic Features in Bi-directional LSTM for Biomedical Named Entity Recognition*
Nut Limsopatham and Nigel Collier

9:50–10:00 *Building Content-driven Entity Networks for Scarce Scientific Literature using Content Information*
Reinald Kim Amplayo and Min Song

10:00–10:10 *Named Entity Recognition in Swedish Health Records with Character-Based Deep Bidirectional LSTMs*
Simon Almgren, Sean Pavlov and Olof Mogren

10:10–10:20 *Entity-Supported Summarization of Biomedical Abstracts*
Frederik Schulze and Mariana Neves

10:20–10:50 Coffee break and Poster Session 1

12th December 2016 (continued)

10:50–12:00 Session 2

10:50–11:10 *Fully unsupervised low-dimensional representation of adverse drug reaction events through distributional semantics*
Alicia Pérez, Arantza Casillas and Koldo Gojenola

11:10–12:00 *Keynote Talk: Learning for Information Extraction in Biomedical and General Domains*
Dr. Makoto Miwa

12:00–14:00 Lunch break

14:00–15:20 Session 3

14:00–14:20 *A Dataset for ICD-10 Coding of Death Certificates: Creation and Usage*
Thomas Lavergne, Aurelie Neveol, Aude Robert, Cyril Grouin, Grégoire Rey and Pierre Zweigenbaum

14:20–14:40 *A Corpus of Tables in Full-Text Biomedical Research Publications*
Tatyana Shmanina, Ingrid Zukerman, Ai Lee Cheam, Thomas Bochynek and Lawrence Cavedon

14:40–14:50 *Supervised classification of end-of-lines in clinical text with no manual annotation*
Pierre Zweigenbaum, Cyril Grouin and Thomas Lavergne

14:50–15:00 *BioDCA Identifier: A System for Automatic Identification of Discourse Connective and Arguments from Biomedical Text*
Sindhuja Gopalan and Sobha Lalitha Devi

15:00–15:10 *Data, tools and resources for mining social media drug chatter*
Abeed Sarker and Graciela Gonzalez

15:10–15:20 *Detection of Text Reuse in French Medical Corpora*
Eva D'hondt, Cyril Grouin, Aurelie Neveol, Efstathios Stamatatos and Pierre Zweigenbaum

12th December 2016 (continued)

15:20–15:50 Coffee break and Poster Session 2

15:50–16:50 Session 4

15:50–16:10 *Negation Detection in Clinical Reports Written in German*
Viviana Cotik, Roland Roller, Feiyu Xu, Hans Uszkoreit, Klemens Budde and Danilo Schmidt

16:10–16:30 *Scoring Disease-Medication Associations using Advanced NLP, Machine Learning, and Multiple Content Sources*
Bharath Dandala, Murthy Devarakonda, Mihaela Bornea and Christopher Nielson

16:30–16:50 *Author Name Disambiguation in MEDLINE Based on Journal Descriptors and Semantic Types*
Dina Vishnyakova, Raul Rodriguez-Esteban, Khan Ozol and Fabio Rinaldi

16:50–17:00 Closing remarks