ACL 2017

The 55th Annual Meeting of the Association for Computational Linguistics

Proceedings of the Student Research Workshop

July 30 - August 4, 2017
Vancouver, Canada

©2017 The Association for Computational Linguistics

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 978-1-945626-56-2

Introduction

Welcome to the ACL 2017 Student Research Workshop!

The ACL 2017 Student Research Workshop (SRW) is a forum for student researchers in computational linguistics and natural language processing. The workshop provides a unique opportunity for student participants to present their work and receive valuable feedback from the international research community as well as from selected mentors.

Following the tradition of the previous years' student research workshops, we have two tracks: research papers and research proposals. The research paper track is a venue for Ph.D. students, Masters students, and advanced undergraduates to describe completed work or work-in-progress along with preliminary results. The research proposal track is offered for advanced Masters and Ph.D. students who have decided on a thesis topic and are interested in feedback on their proposal and ideas about future directions for their work.

We received in total 58 submissions: 16 research proposals and 42 research papers. We accepted 6 research proposals and 17 research papers, giving an acceptance rate of 40% overall. This year, all of the accepted papers will be presented as posters in an evening poster session during the main conference.

Mentoring is at the heart of the SRW. In keeping with previous years, students had the opportunity to participate in a pre-submission mentoring program prior to the submission deadline. This program offered students a chance to receive comments from an experienced researcher, in order to improve the quality of the writing and presentation before making their submission. Twenty-two authors participated in the pre-submission mentoring. In addition, authors of accepted SRW papers are matched with mentors who will meet with the students in person during the workshop. Each mentor prepares in-depth comments and questions prior to the student's presentation, and provides discussion and feedback during the workshop.

We are grateful for the support of the Don and Betty Walker Scholarship Fund and the National Science Foundation under award No. 1714855, which help support the travel expenses of workshop participants.

We would also like to thank our program committee members for their careful reviews of each paper, and all of our mentors for donating their time to provide feedback to our student authors. Thank you to our faculty advisors Cecilia Ovesdotter Alm, Marine Carpuat, and Mark Dredze for their essential advice and guidance, and to members of the ACL 2017 organizing committee, including Chris Callison-Burch, Min-Yen Kan, Regina Barzilay, Charley Chan, Christian Federmann and Priscilla Rasmussen, for their helpful support. Finally, thank you to our student participants!

Organizers:

Allyson Ettinger, University of Maryland
Spandana Gella, University of Edinburgh
Matthieu Labeau, University Paris-Saclay

Faculty Advisors:

Cecilia Ovesdotter Alm, Rochester Institute of Technology
Marine Carpuat, University of Maryland
Mark Dredze, Johns Hopkins University

Program Committee:

Alexandre Allauzen, University Paris-Saclay
Hadi Amiri, University of Maryland
Antonios Anastasopoulos, University of Notre Dame
Bharat Ram Ambati, Apple Inc.
Lauriane Aufrant, University Paris-Saclay
Timothy Baldwin, The University of Melbourne
Kalika Bali, Microsoft Research, India
Rachel Bawden, University Paris-Saclay
Julian Brooke, The University of Melbourne
John Carroll, University of Sussex
Danqi Chen, Stanford University
Wanxiang Che, Harbin Institute of Technology
David Chiang, University of Notre Dame
Kyunghyun Cho, New York University
Tejaswini Deoskar, University of Amsterdam
Long Duong, Voicebox Technologies
Lea Frermann, University of Edinburgh
Kevin Gimpel, Toyota Technological Institute at Chicago
Ralph Grishman, New York University
Alvin Grissom, University of Colorado
Jiang Guo, Harbin Institute of Technology
Xiaodong He, Microsoft Research, USA
Eduard Hovy, Carnegie Mellon University
Mohit Iyyer, University of Maryland
Kristiina Jokinen, University of Cambridge
Daisuke Kawahara, Kyoto University
Eric Laporte, University Paris-Est Marne-la-Vallée
Haizhou Li, A*STAR Institute for Infocomm Research
Junyi Jessy Li, University of Pennsylvania
Zhenghua Li, Soochow University
Annie Louis, University of Essex
Julie Medero, Harvey Mudd College
Mitchell Marcus, University of Pennsylvania
Prashant Mathur, Ebay Research
Diana McCarthy, University of Cambridge
Saif Mohammad, National Research Council Canada

Taesun Moon, IBM Research
Maria Nadejde, University of Edinburgh
Karthik Narasimhan, Massachusetts Institute of Technology
Shashi Narayanan, University of Edinburgh
Graham Neubig, Carnegie Mellon University
Yannick Parmentier, University of Orléans
Gerald Penn, University of Toronto
Emily Pitler, Google Research, USA
Massimo Poesio, University of Essex
Emily Prud'hommeaux, Rochester Institute of Technology
Avinesh PVS, Technishe Universität Darmstadt
Siva Reddy, University of Edinburgh
Brian Roark, Google Research, USA
Michael Roth, University of Illinois Urbana-Champaign
Markus Saers, Hong Kong University of Science and Technology
David Schlangen, Bielefeld University
Carina Silberer, Saarland University
Ekaterina Shutova, University of Cambridge
Sunayana Sitaram, Microsoft Research, India
Kevin Small, Amazon
Richard Sproat, Google Research, USA
Christoph Teichmann, University of Potsdam
Sowmya Vajjala, Iowa State University
Marlies van der Wees, University of Amsterdam
William Yang Wang, Carnegie Mellon University
Bonnie Webber, University of Edinburgh
Alessandra Zarcone, Saarland University
Meishan Zhang, Singapore University of Technology and Design
Yuan Zhang, Massachusetts Institute of Technology
Yue Zhang, Singapore University of Technology and Design

Table of Contents

Computational Characterization of Mental States: A Natural Language Processing Approach
Facundo Carrillo 1

Improving Distributed Representations of Tweets - Present and Future
Ganesh Jawahar 4

Bilingual Word Embeddings with Bucketed CNN for Parallel Sentence Extraction
Jeenu Grover and Pabitra Mitra 11

nQuery - A Natural Language Statement to SQL Query Generator
Nandan Sukthankar, Sanket Maharnawar, Pranay Deshmukh, Yashodhara Haribhakta and Vibhavari Kamble 17

V for Vocab: An Intelligent Flashcard Application
Nihal V. Nayak, Tanmay Chinchore, Aishwarya Hanumanth Rao, Shane Michael Martin, Sagar Nagaraj Simha, G. M. Lingaraju and H. S. Jamadagni 24

Are You Asking the Right Questions? Teaching Machines to Ask Clarification Questions
Sudha Rao 30

Building a Non-Trivial Paraphrase Corpus Using Multiple Machine Translation Systems
Yui Suzuki, Tomoyuki Kajiwara and Mamoru Komachi 36

Segmentation Guided Attention Networks for Visual Question Answering
Vasu Sharma, Ankita Bishnu and Labhesh Patel 43

Text-based Speaker Identification on Multiparty Dialogues Using Multi-document Convolutional Neural Networks
Kaixin Ma, Catherine Xiao and Jinho D. Choi 49

Variation Autoencoder Based Network Representation Learning for Classification
Hang Li, Haozheng Wang, Zhenglu Yang and Masato Odagaki 56

Blind Phoneme Segmentation With Temporal Prediction Errors
Paul Michel, Okko Räsänen, Roland Thiolliere and Emmanuel Dupoux 62

Automatic Generation of Jokes in Hindi
Srishti Aggarwal and Radhika Mamidi 69

Word Embedding for Response-To-Text Assessment of Evidence
Haoran Zhang and Diane Litman 75

Domain Specific Automatic Question Generation from Text
Katira Soleymanzadeh 82

SoccEval: An Annotation Schema for Rating Soccer Players
Jose Ramirez, Matthew Garber and Xinhao Wang 89

Accent Adaptation for the Air Traffic Control Domain
Matthew Garber, Meital Singer and Christopher Ward 95

Generating Steganographic Text with LSTMs
Tina Fang, Martin Jaggi and Katerina Argyraki 100

Predicting Depression for Japanese Blog Text
Misato Hiraga 107

Fast Forward Through Opportunistic Incremental Meaning Representation Construction
Petr Babkin and Sergei Nirenburg 114

Modeling Situations in Neural Chat Bots
Shoetsu Sato, Naoki Yoshinaga, Masashi Toyoda and Masaru Kitsuregawa 120

An Empirical Study on End-to-End Sentence Modelling
Kurt Junshean Espinosa 128

Varying Linguistic Purposes of Emoji in (Twitter) Context
Noa Naaman, Hannah Provenza and Orion Montoya 136

Negotiation of Antibiotic Treatment in Medical Consultations: A Corpus Based Study
Nan Wang 142

Conference Program

Monday July 31 - Evening poster session

Computational Characterization of Mental States: A Natural Language Processing Approach
Facundo Carrillo

Improving Distributed Representations of Tweets - Present and Future
Ganesh Jawahar

Bilingual Word Embeddings with Bucketed CNN for Parallel Sentence Extraction
Jeenu Grover and Pabitra Mitra

nQuery - A Natural Language Statement to SQL Query Generator
Nandan Sukthankar, Sanket Maharnawar, Pranay Deshmukh, Yashodhara Haribhakta and Vibhavari Kamble

V for Vocab: An Intelligent Flashcard Application
Nihal V. Nayak, Tanmay Chinchore, Aishwarya Hanumanth Rao, Shane Michael Martin, Sagar Nagaraj Simha, G. M. Lingaraju and H. S. Jamadagni

Are You Asking the Right Questions? Teaching Machines to Ask Clarification Questions
Sudha Rao

Building a Non-Trivial Paraphrase Corpus Using Multiple Machine Translation Systems
Yui Suzuki, Tomoyuki Kajiwara and Mamoru Komachi

Segmentation Guided Attention Networks for Visual Question Answering
Vasu Sharma, Ankita Bishnu and Labhesh Patel

Text-based Speaker Identification on Multiparty Dialogues Using Multi-document Convolutional Neural Networks
Kaixin Ma, Catherine Xiao and Jinho D. Choi

Variation Autoencoder Based Network Representation Learning for Classification
Hang Li, Haozheng Wang, Zhenglu Yang and Masato Odagaki

Blind Phoneme Segmentation With Temporal Prediction Errors
Paul Michel, Okko Räsänen, Roland Thiolliere and Emmanuel Dupoux

Monday July 31 - Evening poster session (continued)

Automatic Generation of Jokes in Hindi
Srishti Aggarwal and Radhika Mamidi

Word Embedding for Response-To-Text Assessment of Evidence
Haoran Zhang and Diane Litman

Domain Specific Automatic Question Generation from Text
Katira Soleymanzadeh

SoccEval: An Annotation Schema for Rating Soccer Players
Jose Ramirez, Matthew Garber and Xinhao Wang

Accent Adaptation for the Air Traffic Control Domain
Matthew Garber, Meital Singer and Christopher Ward

Generating Steganographic Text with LSTMs
Tina Fang, Martin Jaggi and Katerina Argyraki

Predicting Depression for Japanese Blog Text
Misato Hiraga

Fast Forward Through Opportunistic Incremental Meaning Representation Construction
Petr Babkin and Sergei Nirenburg

Modeling Situations in Neural Chat Bots
Shoetsu Sato, Naoki Yoshinaga, Masashi Toyoda and Masaru Kitsuregawa

An Empirical Study on End-to-End Sentence Modelling
Kurt Junshean Espinosa

Varying Linguistic Purposes of Emoji in (Twitter) Context
Noa Naaman, Hannah Provenza and Orion Montoya

Negotiation of Antibiotic Treatment in Medical Consultations: A Corpus Based Study
Nan Wang