NAACL 2022

The 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies

Proceedings of the Student Research Workshop

July 10-15, 2022

The NAACL organizers gratefully acknowledge the support from the following sponsors.

SRW Sponsors

Google Research

Microsoft

©2022 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 978-1-955917-73-5

Introduction

Welcome to the NAACL 2022 Student Research Workshop.

The NAACL 2022 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 faculty mentors.

Following the tradition of the previous student research workshops, we have archival and non-archival tracks for research papers and thesis 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 thesis 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.

This year, we received 96 submissions in total: 8 thesis proposals and 88 research papers. We accepted 5 thesis proposal and 41 research papers, resulting in an acceptance rate of 63% for thesis proposals and 47% for research papers. Out of the 41 research papers, 9 were non-archival and 33 are presented in these proceedings. Out of the 5 thesis proposals, 1 was non-archival and 4 are presented in these proceedings.

Mentoring is at the heart of the SRW. In line with previous years, we had a pre-submission mentoring program before the submission deadline. A total of 28 papers participated in the pre-submission mentoring program. This program offered students the opportunity to receive comments from an experienced researcher to improve the writing style and presentation of their submissions.

We are deeply grateful to our sponsors, the National Science Foundation, Microsoft and Google. We thank the program committee members for their careful reviews of each paper and all of the mentors for donating their time to provide feedback to the student authors. We thank our faculty advisors, Danqi Chen and Nianwen Xue, for their essential advice and guidance, and the NAACL 2022 organizing committee for their support. Finally, we thank all the student authors for submitting their work and participating in the NAACL 2022 edition of the SRW.

Program Committee

Student Research Workshop Student Chairs

Daphne Ippolito, University of Pennsylvania
Liunian Harold Li, University of California, Los Angeles
Maria Leonor Pacheco, Purdue University

Faculty Advisors

Danqi Chen, Princeton University
Nianwen Xue, Brandeis University

Pre-submission Mentors

Archita Pathak, Amazon
Arkaitz Zubiaga, Queen Mary University of London
Atul Kumar Ojha, National University of Ireland, Galway
Dan Goldwasser, Purdue University
Eduardo Blanco, Arizona State University
Greg Durrett, University of Texas at Austin
Hao Tan, Adobe
Kasturi Bhattacharjee, AWS AI
Kristen Johnson, Michigan State University
Mascha Kurpicz-Briki, Bern University of Applied Sciences
Melissa Roemmele, RWS Language Weaver
Mihir Kale, Google
Muhao Chen, University of Southern California
Nedjma Ousidhoum, The University of Cambridge
Rajaswa Patil, TCS Research
Sunipa Dev, University of California, Los Angelesv
Shomir Wilson, Pennsylvania State University
Valerio Basile, University of Turin
Vincent Ng, University of Texas at Dallas
Vivek Srivastava, TCS Research
Yonatan Bisk, Carnegie Mellon University

Program Committee

Abeer Aldayel, King Saud University
Abhinav Joshi, Indian Institute of Technology, Kanpur
Adithya Pratapa, Carnegie Mellon University
Adithya Shah, Virginia Polytechnic Institute and State University
Agnieszka Falenska, University of Stuttgart
Alessandra Teresa Cignarella, University of Turin
Alexander Panchenko, Skoltech
Alina Karakanta, Fondazione Bruno Kessler
Andrea Varga, Theta Lake Ltd
Arjun Subramonian, University of California, Los Angeles

Arkaitz Zubiaga, Queen Mary University London
Arya McCarthy, Johns Hopkins University
Ashima Suvarna, University of California, Los Angeles
Ayushi Pandey, University of Dublin, Trinity College
Beatrice Savoldi, Fondazione Bruno Kessler
Bonnie Webber, University of Edinburgh
Brian Davis, Dublin City University
Bruno Martins, InstituInstituto Superior Técnico
Bryan Li, University of Pennsylvania
Chenglei Si, University of Maryland, College Park
Dage Särg, University of Tartu
Dan Goldwasser, Purdue University
Eduardo Blanco, Arizona State University
Esma Balkir, National Research Council Canada
Evgeniia Tokarchuk, University of Amsterdam
Fajri Koto, The University of Melbourne
Fatemehsadat Mireshghallah, University of California, San Diego
Fernando Alva-Manchego, Cardiff University
Francielle Vargas, Universidade de São Paulo
Gabriel Bernier-Colborne, National Research Council Canada
Gemma Boleda, ICREA
Hai Pham, Carnegie Mellon University
Haoyu Wang, University of Pennsylvania
Hossein Rouhizadeh, Shahid Beheshti University
Hritik Bansal, University of California, Los Angeles
J. Alberto Conejero, Universidad Politécnica de Valencia
Jakob Prange, Georgetown University
Jakob Prange, Georgetown University
Jasy Suet Yan Liew, Universiti Sains Malaysia
Joseph Marvin Imperial, University of Bath
Julia Kiseleva, Microsoft Research
Koichi Takeda, Nagoya University
Kristen Johnson, Michigan State University
Labiba Jahan, Augustana College
Laurie Burchell, University of Edinburgh
Lei Li, Peking University
Lina Maria Rojas-Barahona, Orange Labs
Luca Soldaini, Allen Institute for Artificial Intelligence
Maaz Amjad, Instituto Politecnico Nacional
Maria Antoniak, Cornell University
Maria Kunilovskaya, University of Wolverhampton
Marija Stanojevic, Temple University
Masaaki Nagata, NTT Corporation
Mascha Kurpicz-Briki, Bern University of Applied Sciences
Max Müller-Eberstein, IT University of Copenhagen
Meishan Zhang, Harbin Institute of Technology, Shenzhen
Michael Miller Yoder, Carnegie Mellon University
Michael Sejr Schlichtkrull, University of Cambridge
Mike Zhang, IT University of Copenhagen
Mingda Chen, Toyota Technological Institute at Chicago
Mona Jalal, Boston University

Neville Ryant, Linguistic Data Consortium
Nikhil Mehta, Purdue University
Parsa Bagherzadeh, Concordia University
Peiyuan Liao, Praxis Pioneering
Prashant Kodali, International Institute of Information Technology Hyderabad
Qing Lyu, University of Pennsylvania
Qing Lyu, University of Pennsylvania
Rajkumar Pujari, Purdue University
Ritam Dutt, Carnegie Mellon University
Ryohei Sasano, Nagoya University
Ryohei Sasano, Nagoya University
Sandeep Mathias, Presidency University, Bangalore
Sashank Santhanam, Apple
Satwik Dutta, University of Texas at Dallas
Shabnam Behzad, Georgetown University
Shamik Roy, Purdue University
Shubham Mittal, Indian Institute of Technology Delhi
Shubhanshu Mishra, Twitter
Sowmya Vajjala, National Research Council Canada
Stan Szpakowicz, University of Ottawa
Sunny Rai, Mahindra University
Surangika Ranathunga, University of Moratuwa
Tanmay Parekh, University of California, Los Angeles
Tatiana Litvinova, Voronezh State Pedagogical University
Tom Sherborne, University of Edinburgh
Tsvetomila Mihaylova, Instituto de Telecomunicações, Portugal
Tunazzina Islam, Purdue University
Vincent Nguyen, Australian National University
Vivek Srivastava, Tata Consultancy Services Limited, India
Xabier Soto, University of the Basque Country UPV/EHU
Xiang Dai, CSIRO
Yoshinari Fujinuma, AWS AI Labs
Younghun Lee, Purdue University
Yuan Zhang, Google
Zhong Zhou, Carnegie Mellon University

Table of Contents

Systematicity Emerges in Transformers when Abstract Grammatical Roles Guide Attention
Ayush K Chakravarthy, Jacob Labe Russin and Randall O'Reilly 1

Grounding in social media: An approach to building a chit-chat dialogue model
Ritvik Choudhary and Daisuke Kawahara 9

ExtraPhrase: Efficient Data Augmentation for Abstractive Summarization
Mengsay Loem, Sho Takase, Masahiro Kaneko and Naoaki Okazaki 16

Regularized Training of Nearest Neighbor Language Models
Jean-Francois Ton, Walter Talbott, Shuangfei Zhai and Joshua M. Susskind 25

Again, Dozens of Refugees Drowned: A Computational Study of Political Framing Evoked by Presuppositions
Qi Yu 31

Methods for Estimating and Improving Robustness of Language Models
Michal Stefanik 44

Retrieval-augmented Generation across Heterogeneous Knowledge
Wenhao Yu 52

Neural Retriever and Go Beyond: A Thesis Proposal
Man Luo 59

Improving Classification of Infrequent Cognitive Distortions: Domain-Specific Model vs. Data Augmentation
Xiruo Ding, Kevin Lybarger, Justin Tauscher and Trevor Cohen 68

Generate, Evaluate, and Select: A Dialogue System with a Response Evaluator for Diversity-Aware Response Generation
Ryoma Sakaeda and Daisuke Kawahara 76

Impact of Training Instance Selection on Domain-Specific Entity Extraction using BERT
Eileen Salhofer, Xing Lan Liu and Roman Kern 83

Analysing the Correlation between Lexical Ambiguity and Translation Quality in a Multimodal Setting using WordNet
Ali Hatami, Paul Buitelaar and Mihael Arcan 89

Building a Personalized Dialogue System with Prompt-Tuning
Tomohito Kasahara, Daisuke Kawahara, Nguyen Tung, Shengzhe Li, Kenta Shinzato and Toshinori Sato 96

MM-GATBT: Enriching Multimodal Representation Using Graph Attention Network
Seung Byum Seo, Hyoungwook Nam and Payam Delgosha 106

Simulating Feature Structures with Simple Types
Valentin D. Richard 113

Dr. Livingstone, I presume? Polishing of foreign character identification in literary texts
Aleksandra Konovalova, Antonio Toral and Kristiina Taivalkoski-Shilov 123

Zuo Zhuan Ancient Chinese Dataset for Word Sense Disambiguation
Xiaomeng Pan, Hongfei Wang, Teruaki Oka and Mamoru Komachi 129

ViT5: Pretrained Text-to-Text Transformer for Vietnamese Language Generation
Long Phan, Hieu Tran, Hieu Nguyen and Trieu H. Trinh 136

Compositional Generalization in Grounded Language Learning via Induced Model Sparsity
Sam Spilsbury and Alexander Ilin 143

How do people talk about images? A study on open-domain conversations with images.
Yi-Pei Chen, Nobuyuki Shimizu, Takashi Miyazaki and Hideki Nakayama 156

Text Style Transfer for Bias Mitigation using Masked Language Modeling
Ewoenam Kwaku Tokpo and Toon Calders 163

Differentially Private Instance Encoding against Privacy Attacks
Shangyu Xie and Yuan Hong 172

A Simple Approach to Jointly Rank Passages and Select Relevant Sentences in the OBQA Context
Man Luo, Shuguang Chen and Chitta Baral 181

Multimodal Modeling of Task-Mediated Confusion
Camille Mince, Skye Rhomberg, Cecilia Alm, Reynold Bailey and Alex Ororbia 188

Probe-Less Probing of BERT's Layer-Wise Linguistic Knowledge with Masked Word Prediction
Tatsuya Aoyama and Nathan Schneider 195

Multimodal large language models for inclusive collaboration learning tasks
Armanda Lewis 202

Neural Networks in a Product of Hyperbolic Spaces
Jun Takeuchi, Noriki Nishida and Hideki Nakayama 211

Explicit Use of Topicality in Dialogue Response Generation
Takumi Yoshikoshi, Hayato Atarashi, Takashi Kodama and Sadao Kurohashi 222

Automating Human Evaluation of Dialogue Systems
Sujan Reddy A 229

Strong Heuristics for Named Entity Linking
Marko Čuljak, Andreas Spitz, Robert West and Akhil Arora 235

Static and Dynamic Speaker Modeling based on Graph Neural Network for Emotion Recognition in Conversation
Prakhar Saxena, Yin Jou Huang and Sadao Kurohashi 247

Few-shot fine-tuning SOTA summarization models for medical dialogues
David Fraile Navarro, Mark Dras and Shlomo Berkovsky 254

Unifying Parsing and Tree-Structured Models for Generating Sentence Semantic Representations
Antoine Simoulin and Benoit Crabbé 267

Multiformer: A Head-Configurable Transformer-Based Model for Direct Speech Translation
Gerard Sant, Gerard I. Gállego, Belen Alastruey and Marta Ruiz Costa-jussà 277

Defending Compositionality in Emergent Languages
Michal Auersperger and Pavel Pecina 285

Exploring the Effect of Dialect Mismatched Language Models in Telugu Automatic Speech Recognition
Aditya Yadavalli, Ganesh Sai Mirishkar and Anil Vuppala 292

Program

Tuesday, July 12, 2022

08:00 - 09:00 *Virtual Poster Q&A Session 1*

Explicit Use of Topicality in Dialogue Response Generation
Takumi Yoshikoshi, Hayato Atarashi, Takashi Kodama and Sadao Kurohashi

Automating Human Evaluation of Dialogue Systems
Sujan Reddy A

Probe-Less Probing of BERT's Layer-Wise Linguistic Knowledge with Masked Word Prediction
Tatsuya Aoyama and Nathan Schneider

Regularized Training of Nearest Neighbor Language Models
Jean-Francois Ton, Walter Talbott, Shuangfei Zhai and Joshua M. Susskind

Analysing the Correlation between Lexical Ambiguity and Translation Quality in a Multimodal Setting using WordNet
Ali Hatami, Paul Buitelaar and Mihael Arcan

12:15 - 10:45 *Panel Discussion for Starting Researchers*

14:15 - 15:45 *In-Person Poster Session*

Systematicity Emerges in Transformers when Abstract Grammatical Roles Guide Attention
Ayush K Chakravarthy, Jacob Labe Russin and Randall O'Reilly

Grounding in social media: An approach to building a chit-chat dialogue model
Ritvik Choudhary and Daisuke Kawahara

ExtraPhrase: Efficient Data Augmentation for Abstractive Summarization
Mengsay Loem, Sho Takase, Masahiro Kaneko and Naoaki Okazaki

Neural Retriever and Go Beyond: A Thesis Proposal
Man Luo

Improving Classification of Infrequent Cognitive Distortions: Domain-Specific Model vs. Data Augmentation
Xiruo Ding, Kevin Lybarger, Justin Tauscher and Trevor Cohen

Generate, Evaluate, and Select: A Dialogue System with a Response Evaluator for Diversity-Aware Response Generation
Ryoma Sakaeda and Daisuke Kawahara

Building a Personalized Dialogue System with Prompt-Tuning
Tomohito Kasahara, Daisuke Kawahara, Nguyen Tung, Shengzhe Li, Kenta Shinzato and Toshinori Sato

MM-GATBT: Enriching Multimodal Representation Using Graph Attention Network
Seung Byum Seo, Hyoungwook Nam and Payam Delgosha

ViT5: Pretrained Text-to-Text Transformer for Vietnamese Language Generation
Long Phan, Hieu Tran, Hieu Nguyen and Trieu H. Trinh

Compositional Generalization in Grounded Language Learning via Induced Model Sparsity
Sam Spilsbury and Alexander Ilin

How do people talk about images? A study on open-domain conversations with images.
Yi-Pei Chen, Nobuyuki Shimizu, Takashi Miyazaki and Hideki Nakayama

A Simple Approach to Jointly Rank Passages and Select Relevant Sentences in the OBQA Context
Man Luo, Shuguang Chen and Chitta Baral

Multimodal Modeling of Task-Mediated Confusion
Camille Mince, Skye Rhomberg, Cecilia Alm, Reynold Bailey and Alex Ororbia

Neural Networks in a Product of Hyperbolic Spaces
Jun Takeuchi, Noriki Nishida and Hideki Nakayama

Strong Heuristics for Named Entity Linking
Marko Čuljak, Andreas Spitz, Robert West and Akhil Arora

Unifying Parsing and Tree-Structured Models for Generating Sentence Semantic Representations
Antoine Simoulin and Benoit Crabbé

Defending Compositionality in Emergent Languages
Michal Auersperger and Pavel Pecina

Exploring the Effect of Dialect Mismatched Language Models in Telugu Automatic Speech Recognition
Aditya Yadavalli, Ganesh Sai Mirishkar and Anil Vuppala

Multimodal large language models for inclusive collaboration learning tasks
Armanda Lewis

16:15 - 17:45 *Thesis Proposals*

Methods for Estimating and Improving Robustness of Language Models
Michal Stefanik

Retrieval-augmented Generation across Heterogeneous Knowledge
Wenhao Yu

Neural Retriever and Go Beyond: A Thesis Proposal
Man Luo

Multimodal large language models for inclusive collaboration learning tasks
Armanda Lewis

Wednesday, July 13, 2022

08:00 - 09:00 *Virtual Poster Q&A Session 3*

Again, Dozens of Refugees Drowned: A Computational Study of Political Framing Evoked by Presuppositions
Qi Yu

Impact of Training Instance Selection on Domain-Specific Entity Extraction using BERT
Eileen Salhofer, Xing Lan Liu and Roman Kern

Text Style Transfer for Bias Mitigation using Masked Language Modeling
Ewoenam Kwaku Tokpo and Toon Calders

Differentially Private Instance Encoding against Privacy Attacks
Shangyu Xie and Yuan Hong

Methods for Estimating and Improving Robustness of Language Models
Michal Stefanik

Static and Dynamic Speaker Modeling based on Graph Neural Network for Emotion Recognition in Conversation
Prakhar Saxena, Yin Jou Huang and Sadao Kurohashi

Simulating Feature Structures with Simple Types
Valentin D. Richard

09:15 - 10:15 *Virtual Poster Q&A Session 4*

Dr. Livingstone, I presume? Polishing of foreign character identification in literary texts
Aleksandra Konovalova, Antonio Toral and Kristiina Taivalkoski-Shilov

Zuo Zhuan Ancient Chinese Dataset for Word Sense Disambiguation
Xiaomeng Pan, Hongfei Wang, Teruaki Oka and Mamoru Komachi

Few-shot fine-tuning SOTA summarization models for medical dialogues
David Fraile Navarro, Mark Dras and Shlomo Berkovsky