

SocialNLP 2022

**The Tenth International Workshop on Natural Language
Processing for Social Media**

Proceedings of the Workshop

July 14-15, 2022

©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 978-1-955917-88-9

SocialNLP 2022@NAACL Chairs' Welcome

Happy 10th Anniversary to SocialNLP!

It is our great pleasure to welcome you to the Tenth Workshop on Natural Language Processing for Social Media-SocialNLP 2022, associated with NAACL 2022. SocialNLP is an inter-disciplinary area of natural language processing (NLP) and social computing. We hold SocialNLP twice a year: one in the NLP venue, the other in the associated venue such as those for web technology or artificial intelligence. This year the other version has been successfully held in conjunction with TheWebConf 2022 (formerly WWW), and we are very happily looking forward to the NLP version in NAACL 2022. We are very glad that the number of submissions to this year's workshop keeps increasing, and the submissions themselves were still of high quality with the accepted threshold of 3.33 (maximum 5), which again leads to a competitive selection process. We received submissions from Asia, Europe, and the United States. Considering the review process is rigorous and we want to encourage authors to participate in the workshop, we accepted 8 oral papers. These exciting papers include novel and practical topics for researchers working on NLP for social media, such as bias mitigation, domain transfer, and dataset constructed for the newly emerged research problems. We believe they will benefit our research community.

On this 10th anniversary, we would like to share our happiness and celebrate the success together with our community members with the special speaker reunion event. Most of our previous invited speakers, including Prof. Saif Muhammad from National Research Council Canada, Prof. Yohei Seki from University of Tsukuba, Prof. Tim Wenginger and Mr. Nicholas Botzer from University of Notre Dame, Prof. Sonjanya Poria from Singapore University of Technology and Design, Prof. Cristian Danescu-Niculescu-Mizil from Cornell University, Prof. Dan Goldwasser from Purdue University, Dr. Ian Stewart from University of Michigan, and Prof. Thamar Solorio from University of Houston will participate in this event and give an insightful talk. We deeply appreciate their support. Their talk should reveal the past and the future of related research topics to participants in the hope of encouraging more researchers to join us. We couple each invited talk with one oral paper presentation to encourage attendees to (virtually) attend both sessions to have more discussions with outstanding researchers.

Putting together SocialNLP 2022 was a team effort. We first thank the authors for providing the quality content of the program. We are grateful to the program committee members, who worked very hard in reviewing papers and providing feedback to authors. Finally, we especially thank the NAACL Workshop chairs Prof. Dan Goldwasser, Dr. Yunyao Li, and Dr. Ashish Sabharwal for helping us with all the complicated logistics for this year's online version.

We hope you enjoy the workshop and keep supporting us!

Organizers

Lun-Wei Ku, Academia Sincia, Taiwan

Cheng-Te Li, National Cheng Kung University, Taiwan

Yu-Che Tsai, National Taiwan University, Taiwan

Wei-Yao Wang, National Yang Ming Chiao Tung University, Taiwan

Organizing Committee

Organizers

Lun-Wei Ku, Academia Sinica

Cheng-Te Li, National Cheng Kung University

Yu-Che Tsai, National Taiwan University

Wei-Yao Wang, National Yang Ming Chiao Tung University

Program Committee

Program Committee

Silvio Amir, CCIS, Northeastern University
Yung-Chun Chang, Graduate Institute of Data Science, Taipei Medical University
Zhiyu Chen, University of California, Santa Barbara
Monojit Choudhury, Microsoft Research
Min-Yuh Day, Tamkang University
David Jurgens, University of Michigan
Tsung-Ting Kuo, University of California San Diego
Chuan-Jie Lin, National Taiwan Ocean University
Manuel Montes, INAOE
Derek Ruths, McGill University
Saurav Sahay, Intel Labs
Yohei Seki, University of Tsukuba
Steven Wilson, Oakland University
Shih-Hung Wu, Chaoyang University of Technology
Liang-Chih Yu, Yuan Ze University
Zhe Zhang, Meta AI
Huaping Zhang, Beijing Institute of Technology

Keynote Talk: Facilitating Citizens' Voices on Social Media to Address Regional Issues

Yohei Seki

University of Tsukuba

Abstract: I will talk about our research on extracting issues and evaluations for the administrative aspects of local governments from the analysis of citizens' voices on social media. Specifically, I will explain our method of collecting and analyzing citizens' voices from Twitter on a city-by-city basis and analyze how the actions of local governments in Japan changed citizens' voices. As an applied case study, I will also present our work on understanding the mood of citizens' interest for the region and our analysis of citizens' opinions on childcare and restaurant take-out services in the COVID-19 disaster.

Bio: Dr. Yohei Seki is currently an associate professor, Faculty of Library, Information and Media Science, University of Tsukuba, Japan. He received his Ph.D. degree in Informatics from the Graduate University for Advanced Studies (SOKENDAI) in 2005. He was visiting scholars at Columbia University in 2008 and at National University of Singapore in 2018, respectively. His main research interests are natural language processing, sentiment analysis, and information access. He leads sentiment analysis work as one of co-organizers in NTCIR multilingual opinion analysis task from 2006 to 2010. He received best paper award at CEA 2014 and nominated best paper award runner-up at ICADL 2020. He recently published international standards for smart cities ISO/IEC 30146 in 2019 and ISO/IEC 30145-3 in 2020 as co-editor.

Keynote Talk: Conversational Flow and Moral Judgment on Reddit

Tim Weninger and Nicholas Botzer
University of Notre Dame

Abstract: The focus of this talk will revolve around two works that both leverage the large amount of conversational data that is available on Reddit. In the first work we will look at how conversations flow across entities in discussions threads on Reddit. In the study of social networks the typical perspective is to view users as nodes and concepts as flowing through user-nodes within the social network. In this work we take the opposite perspective: we extract and organize group discussion into a concept space we call an entity graph where concepts and entities are static and human communicators move about the concept space via their conversations. In the second work, we will look at how users cast moral judgements of each other by extracting and analyzing self-contained labels from the subreddit /r/AmITheAsshole. These labels allow us to train a BERT classifier to determine when someone is casting a moral judgment on another.

Bio: Tim Weninger is the Frank M. Freimann Collegiate Associate Professor of Engineering in the Department of Computer Science and Engineering in the College of Engineering at the University of Notre Dame. His research is at the intersection of machine learning, network science and social media. Generally speaking, his work is to uncover how humans consume and curate information.

Nicholas Botzer is a 5th year PhD student at the University of Notre Dame advised by Tim Weninger. His interests are focused in computational social science, natural language processing, and machine learning. His current research focuses on how conversations form online and how people make moral judgements of others.

Keynote Talk: Ethics Sheets for Social NLP: Task-Specific Guides to Responsible Research

Saif Mohammad

National Research Council Canada

Abstract: Several high-profile events, such as the mass testing of emotion recognition systems on vulnerable sub-populations and using question answering systems to make moral judgments, have highlighted how technology will often lead to more adverse outcomes for those that are already marginalized. At issue here are not just individual systems and datasets, but also the AI tasks themselves. In this talk, I make a case for thinking about ethical considerations not just at the level of individual models and datasets, but also at the level of AI tasks. I will present a new form of such an effort, Ethics Sheets for AI Tasks, dedicated to fleshing out the assumptions and ethical considerations hidden in how a task is commonly framed and in the choices we make regarding the data, method, and evaluation. I will also present a template for ethics sheets with 50 ethical considerations, using the task of emotion recognition as a running example. Ethics sheets are a mechanism to engage with and document ethical considerations before building datasets and systems. Similar to survey articles, a small number of ethics sheets can serve numerous researchers and developers. I will wrap things up with concrete steps for students and early researchers.

Bio: Dr. Saif M. Mohammad is a Senior Research Scientist at the National Research Council Canada (NRC). He received his Ph.D. in Computer Science from the University of Toronto. Before joining NRC, Saif was a Research Associate at the Institute of Advanced Computer Studies at the University of Maryland, College Park. His research interests are in Computational Linguistics and Natural Language Processing (NLP), especially Lexical Semantics, Emotions in Language, Sentiment Analysis, Computational Creativity, Fairness in NLP, Psycholinguistics, and Information Visualization. He has published over 100 scientific articles (journal articles, book chapters, and conference papers). He has served in various capacities at prominent journals and conferences, including: action editor for Computational Linguistics, senior action editor for ACL Rolling review, chair of the Canada–UK symposium on Ethics in AI, co-chair of SemEval 2017-19 (the largest platform for semantic evaluations), workshops co-chair for ACL 2020, co-organizer of WASSA 2017 and 2018 (a sentiment analysis workshop), and area chair for ACL, NAACL, and EMNLP (in the areas of sentiment analysis, lexical semantics, and fairness in NLP). His team developed a sentiment analysis system which ranked first in shared task competitions. His word–emotion resources, such as the NRC Emotion Lexicon, are used for analyzing affect in text. His work has garnered media attention, including articles in Time, SlashDot, LiveScience, io9, The Physics arXiv Blog, PC World, and Popular Science.

Keynote Talk: Linguistic Bias as a Window into Social Attitudes

Ian Stewart
University of Michigan

Abstract: When considering the presence of bias in language models, NLP researchers have generally treated it as a problem to be solved, e.g. removing the association between nurse and woman in word embeddings. Alternately, researchers can use the bias in language models as a window into social attitudes expressed through user-generated text. While promising, such bias as a window work should be careful to define the appropriate task and construct for the research question at hand. I present work from three studies that address linguistic bias in user-generated text written in everyday settings, including a controlled experiment, online blog posts, and college course reviews. Using supervised classification, word embeddings, and domain-specific lexicons, we identify specific stereotypes (e.g. associating Black people with basketball) and general attitudes (describing female professors as generally nicer) that are directed toward minority groups. Such analyses of social attitudes demonstrate the wealth of information that exists in “biased” language models, as well as a reminder of the social inequality that is constructed through everyday communication.

Bio: Ian is interested in building natural language processing models that incorporate social information to improve the writing experience for system users. Ian is also interested in computational social science to better understand the benefits and limitations of discussions on social media platforms such as Twitter and Reddit.

Keynote Talk: New Avenues in Dialogue Systems

Soujanya Poria

Singapore University of Technology and Design

Abstract: Lately, the topic of dialogue systems has witnessed a significant surge in research interest due to the vast availability of conversational data on the Web and elsewhere. Dialogue systems also have wide applications in healthcare, e-commerce, and many other sectors. However, the progress in this research area has mostly been limited to the tasks of — (a) dialogue generation using seq2seq frameworks; (b) dialogue act classification, (c) slot filling, and (d) dialogue state tracking. While these tasks are of prime importance when creating a dialogue system, one should not overlook the other aspects of natural language such as emotions, and causal and commonsense reasoning as they are equally vital to attain a superior dialogue understanding. To this end, in this talk, I will explain some of the emerging and challenging dialogue-level tasks related to the above-mentioned aspects — (a) (multimodal) emotion recognition in conversations, (b) empathetic dialogue generation, and (c) recognizing emotion cause in conversations, and (d) commonsense inference in dialogues. Further, I will present strong baselines to address these tasks and shed light on the numerous associated challenges when you attempt to solve these tasks using today's Language Models.

Bio: Soujanya Poria is an assistant professor of Computer Science at the Singapore University of Technology and Design (SUTD), Singapore. He holds a Ph.D. degree in Computer Science from the University of Stirling, UK. Soujanya was a recipient of the prestigious early career research award called 'NTU Presidential Postdoctoral Fellowship' in 2018. Before taking up the presidential fellowship position at the NTU, he was a scientist at the A*STAR and the Temasek Laboratory, NTU. He is also PI of multiple academic and industrial grants with the amount totaling US2,500,000. *He has co-authored more than 100 papers, published in top-tier conferences and journals such as ACL, EMNLP, AAAI, NAACL, STAR, Singapore as a senior scientist. He was an area co-chair at several ACL, NAACL, and EMNLP conferences. SEM 2019. He served as a senior PC member at several AAAI, and IJCAI conferences, and often serves as a PC member. His index is 58. Soujanya is a recipient of several academic awards such as the IEEE CIM outstanding paper award, and*

Keynote Talk: TBD

Cristian Danescu-Niculescu-Mizil

Cornell University

Bio: Online communication is gaining a central role in our society and the opportunities for more and more people to interact online in potentially fruitful ways continue to grow. Sadly, online interactions have also acquired a reputation for not going well: this extends all the way from unproductive and inefficient online collaboration to outright antagonism and harassment in online discussions. During his fellowship, Cristian Danescu-Niculescu-Mizil will take a mixed-methods approach to explore the benefits and potential risks of using novel computational tools to increase the quality of online discussions.

Danescu-Niculescu-Mizil is an associate professor in the information science department at Cornell University. His research aims at developing computational methods that can lead to a better understanding of our conversational practices, supporting tools that can improve the way we communicate with each other. He is the recipient of several awards – including an NSF CAREER Award, the WWW 2013 Best Paper Award, a CSCW 2017 Best Paper Award, and two Google Faculty Research Awards – and his work has been featured in popular media outlets such as The Wall Street Journal, NBC’s The Today Show, NPR and The New York Times.

Keynote Talk: TBD

Dan Goldwasser

Purdue University

Bio: I am an associate professor at the department of computer science at Purdue university. I am broadly interested in connecting natural language with real world scenarios, and using them to guide natural language understanding. Before starting at Purdue I was a postdoctoral researcher at the University of Maryland in College Park. I completed my Ph.D. studies at the University of Illinois at Urbana-Champaign in the department of Computer Science.

Keynote Talk: Code-switching and social media data: an overview of common challenges and recent developments

Thamar Solorio
University of Houston

Abstract: In this talk I will aim to expose the audience to major developments in the brief history of research in NLP for code-switching data. Starting from the first non-empirical paper in a *CL venue on the topic, to the more recent, transformer based papers. I will then discuss how addressing challenges in code-switching data can help advance NLP for social media and vice versa. To conclude, I will point to the outstanding questions in processing this type of data.

Bio: Thamar Solorio is a Professor of Computer Science at the University of Houston (UH) and she is also a visiting scientist at Bloomberg LP. She holds graduate degrees in Computer Science from the Instituto Nacional de Astrofísica, Óptica y Electrónica, in Puebla, Mexico. Her research interests include information extraction from social media data, enabling technology for code-switched data, stylistic modeling of text, and more recently multimodal approaches for online content understanding. She is the director and founder of the Research in Text Understanding and Language Analysis Lab at UH. She is the recipient of an NSF CAREER award for her work on authorship attribution, and recipient of the 2014 Emerging Leader ABIE Award in Honor of Denice Denton. She is currently serving a second term as an elected board member of the North American Chapter of the Association of Computational Linguistics.

Table of Contents

<i>Mask and Regenerate: A Classifier-based Approach for Unpaired Sentiment Transformation of Reviews for Electronic Commerce Websites.</i>	
Shuo Yang	1
<i>Exploiting Social Media Content for Self-Supervised Style Transfer</i>	
Dana Ruitter, Thomas Kleinbauer, Cristina España-Bonet, Josef van Genabith and Dietrich Klakow	11
<i>Detecting Rumor Veracity with Only Textual Information by Double-Channel Structure</i>	
Alex Gunwoo Kim and Sangwon Yoon	24
<i>Leveraging Dependency Grammar for Fine-Grained Offensive Language Detection using Graph Convolutional Networks</i>	
Divyam Goel and Raksha Sharma	34
<i>A Comparative Study on Word Embeddings and Social NLP Tasks</i>	
Fatma Elsafoury, Steven R. Wilson and Naeem Ramzan	44
<i>Identifying Human Needs through Social Media: A study on Indian cities during COVID-19</i>	
Sunny Rai, Rohan Joseph, Prakruti Singh Thakur and Mohammed Abdul Khaliq	54
<i>Towards Toxic Positivity Detection</i>	
Ishan Sanjeev Upadhyay, KV Aditya Srivatsa and Radhika Mamidi	64
<i>OK Boomer: Probing the socio-demographic Divide in Echo Chambers</i>	
Henri-Jacques Geiss, Flora Sakketou and Lucie Flek	72

Program

Thursday, July 14, 2022

- 08:30 - 08:40 *Opening Remarks*
- 08:40 - 09:40 *Day-1 Keynote Speech 1 by Prof. Saif Muhammad (National Research Council Canada)*
- 09:40 - 10:00 *Towards Toxic Positivity Detection*

Towards Toxic Positivity Detection
Ishan Sanjeev Upadhyay, KV Aditya Srivatsa and Radhika Mamidi
- 10:00 - 10:30 *Day-1 Coffee Break*
- 10:30 - 11:30 *Day-1 Keynote Speech 2 by Prof. Yohei Seki (University of Tsukuba)*
- 11:30 - 11:50 *Mask and Regenerate: A Classifier-based Approach for Unpaired Sentiment Transformation of Reviews for Electronic Commerce Websites*

Mask and Regenerate: A Classifier-based Approach for Unpaired Sentiment Transformation of Reviews for Electronic Commerce Websites.
Shuo Yang
- 11:50 - 13:30 *Day-1 Lunch Break*
- 13:30 - 14:30 *Day-1 Keynote Speech 3 by Prof. Tim Wenerger (University of Notre Dame)*
- 14:30 - 14:50 *Detecting Rumor Veracity with Only Textual Information by Double-Channel Structure*

Detecting Rumor Veracity with Only Textual Information by Double-Channel Structure
Alex Gunwoo Kim and Sangwon Yoon
- 14:50 - 15:30 *Day-1 Coffee Break*
- 15:30 - 16:30 *Day-1 Keynote Speech 4 by Prof. Sonjanya Poria (Singapore University of Technology and Design)*
- 16:30 - 16:50 *OK Boomer: Probing the socio-demographic Divide in Echo Chambers*

Thursday, July 14, 2022 (continued)

OK Boomer: Probing the socio-demographic Divide in Echo Chambers

Henri-Jacques Geiss, Flora Sakketou and Lucie Flek

Friday, July 15, 2022

08:40 - 09:40 *Day-2 Keynote Speech 1 by Prof. Cristian Danescu-Niculescu-Mizil (Cornell University)*

09:40 - 10:00 *Exploiting Social Media Content for Self-Supervised Style Transfer*

Exploiting Social Media Content for Self-Supervised Style Transfer

Dana Ruitter, Thomas Kleinbauer, Cristina España-Bonet, Josef van Genabith and Dietrich Klakow

10:00 - 10:30 *Day-2 Coffee Break*

10:30 - 11:30 *Day-2 Keynote Speech 2 by Prof. Dan Goldwasser (Purdue University)*

11:30 - 11:50 *Identifying Human Needs through Social Media: A study on Indian cities during COVID-19*

Identifying Human Needs through Social Media: A study on Indian cities during COVID-19

Sunny Rai, Rohan Joseph, Prakruti Singh Thakur and Mohammed Abdul Khaliq

11:50 - 13:30 *Day-2 Lunch Break*

13:30 - 14:30 *Day-2 Keynote Speech 3 by Dr. Ian Stewart (University of Michigan)*

14:30 - 14:50 *A Comparative Study on Word Embeddings and Social NLP Tasks*

A Comparative Study on Word Embeddings and Social NLP Tasks

Fatma Elsafoury, Steven R. Wilson and Naeem Ramzan

14:50 - 15:30 *Day-2 Coffee Break*

15:30 - 16:30 *Day-2 Keynote Speech 4 by Prof. Thamar Solorio (University of Houston)*

16:30 - 16:50 *Leveraging Dependency Grammar for Fine-Grained Offensive Language Detection using Graph Convolutional Networks*

Leveraging Dependency Grammar for Fine-Grained Offensive Language Detection using Graph Convolutional Networks

Divyam Goel and Raksha Sharma

Friday, July 15, 2022 (continued)

16:50 - 17:00 *Day-2 Closing Remarks*