NLPCSS 2022

The Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS)

Held at the 2022 Conference on Empirical Methods in Natural Language Processing

November 7, 2022

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

Gold

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 978-1-959429-20-3

Introduction

Welcome to the Fifth Workshop on Natural Language Processing (NLP) and Computational Social Science (CSS)! This workshop series builds on a successful string of iterations, with many interdisciplinary contributions to make NLP techniques and insights standard practice in CSS research—as well as improve NLP through insights from the social sciences.

We received a record 63 submissions and after a rigorous review process by our committee, we accepted 23 archival entries and 2 non-archival papers. We are also pleased to include 8 Findings of EMNLP papers for presentation at the workshop. This year we have organized a hybrid event with both virtual and in-person components with an evening soirée in Gather Town to allow mingling between folks in different time zones. We continue to be excited to see so many submissions from outside of NLP, and hope to continue the tradition to foster a dialogue between researchers in NLP and these other fields.

We would like to thank the Program Committee members who reviewed the papers this year. They did a heroic job proving some top-notch reviews in a time when reviewing requests are abundant. We would also like to thank the workshop participants both in-person and virtual for the opportunities to connect (or reconnect) and learn from each other. Last, a word of thanks also goes to our sponsor Google, who enabled us to support valuable participation and activities.

David Bamman, Dirk Hovy, Katherine Keith, David Jurgens, Brendan O'Connor, and Svitlana Volkova (Co-Organizers)

Organizing Committee

Organizers

David Bamman, University of California, Berkeley
Dirk Hovy, Bocconi University
David Jurgens, University of Michigan
Katherine Keith, Williams College
Brendan O'Connor, University of Massachusetts Amherst
Svitlana Volkova, Pacific Northwest National Laboratory

Program Committee

Chairs

David Bamman, University of California, Berkeley
Dirk Hovy, Bocconi University
David Jurgens, University of Michigan
Katherine Keith, University of Massachusetts, Amherst
Brendan O'connor, University of Massachusetts Amherst
Svitlana Volkova, Pacific Northwest National Laboratory

Program Committee

Natalie Ahn, UC Berkeley
Kenan Alkiek, University of Michigan
Kristen M. Altenburger, Facebook
Timothy Baldwin, The University of Melbourne
Alon Bartal, Ben-Gurion University
Eric Bell, Self
Chris Biemann, Universität Hamburg
Laura Biester, University of Michigan
Michael Bloodgood, The College of New Jersey
Amanda Buddemeyer, University of Pittsburgh, School of Computing and Information
Aoife Cahill, Dataminr
Marine Carpuat, University of Maryland
Samuel Carton, University of Chicago
Sky Ch-wang, Columbia University
Mohit Chandra, Georgia Institute of Technology
Kaiping Chen, University of Wisconsin-Madison
Zhiyu Chen, Meta
Minje Choi, University of Michigan
Aron Culotta, Tulane University
Walter Daelemans, University of Antwerp, CLiPS
Debarati Das, University of Minnesota Twin Cities
Leon Derczynski, IT University of Copenhagen
Jonathan Dunn, University of Canterbury
Valery Dzutsati, University of Kansas
Micha Elsner, The Ohio State University
Samuel Fraiberger, World Bank
Kathleen C. Fraser, National Research Council Canada
Dan Garrette, Google Research
Nabeel Gillani, MIT
Ann-sophie Gnehm, University of Zurich
Xiaobo Guo, Dartmouth College
Shirley Anugrah Hayati, University of Minnesota
Marti A. Hearst, UC Berkeley
Joseph Hoover, University of Southern California
Tiancheng Hu, ETH Zurich
Rebecca Hwa, University of Pittsburgh
Loring Ingraham, George Washington University

Mali Jin, University of Sheffield
Kristen Johnson, Michigan State University
Kenneth Joseph, University at Buffalo
Shima Khanehzar, University of Melbourne
Sopan Khosla, Amazon Web Services, Amazon Inc
Roman Klinger, University of Stuttgart
Ekaterina Kochmar, University of Bath
Jiazhao Li, University of Michigan
Jinfen Li, Syracuse University
Mingyang Li, University of Pennsylvania
Ruibo Liu, Dartmouth College
Nikola Ljubešić, Jožef Stefan Institute
Julia Mendelsohn, University of Michigan
Rada Mihalcea, University of Michigan
David Mimno, Cornell University
Shubhanshu Mishra, Twitter Inc.
Kunihiro Miyazaki, The University of Tokyo
Jason Naradowsky, University of Tokyo
John Nay, Stanford University - Center for Legal Informatics
Dong Nguyen, Utrecht University
Pierre Nugues, Lund University
Alice Oh, KAIST
Silviu Oprea, University of Edinburgh
Sebastian Padó, Stuttgart University
Shriphani Palakodety, Onai
Elizabeth Palmieri, University of Virginia
Jiaxin Pei, University of Michigan
Massimo Poesio, Queen Mary University of London
Thierry Poibeau, LATTICE (CNRS & ENS/PSL)
Afshin Rahimi, The University of Queensland
Ange Richard, Laboratoire PACTE, Laboratoire Informatique de Grenoble (LIG)
Matīss Rikters, National Institute of Advanced Industrial Science and Technology
Ellen Riloff, University of Utah
Anthony Rios, University of Texas at San Antonio
Molly Roberts, University of California, San Diego
Frank Rudzicz, St Michael's Hospital; University of Toronto, Department of Computer Science
Asad Sayeed, University of Gothenburg
David Schlangen, University of Potsdam
Hope Schroeder, MIT
Djamé Seddah, Inria
Felix Soldner, GESIS - Leibniz Institute for the Social Science
Nikita Soni, Stony Brook University
Mark Steedman, University of Edinburgh
Sara Tonelli, FBK
Oren Tsur, Ben Gurion University
Zijian Wang, AWS AI Labs
Maximilian Weber, Goethe University
Charles Welch, University of Marburg
Swede White, Louisiana State University Dept. of Sociology
Gregor Wiedemann, Leibniz Institute for Media Research | Hans-Bredow-Institute
Steven Wilson, Oakland University

Winston Wu, University of Michigan
Wei Xu, Georgia Institute of Technology
Xiao Xu, NIDI-KNAW / University of Groningen
Ivan Yamshchikov, Max Planck Institute for Mathematics in the Sciences; ISEG & CEMAPRE, University of Lisbon
Michael Yeomans, Imperial College London
Sourabh Zanwar, RWTH Aachen University
Shuang (sophie) Zhai, University of Oklahoma
Yongjun Zhang, Stony Brook University
Tongtong Zhang, Stanford University
Xixuan Zhang, Freie Universität Berlin
Mian Zhong, ETH Zürich
Naitian Zhou, University of Michigan

Keynote Talk: Tackling social challenges with data science and AI

Jisun An
Singapore Management University

Abstract: Artificial Intelligence (AI) and technology have become increasingly embedded in our daily lives. Billions of people interact with AI systems on various online platforms every day. Those tremendous interactions enable us to understand individual or collective human behavior: what people think, what people care about, how people feel, where people go, what people buy, what people do, and with whom people associate. Tools that can extract hidden patterns and insights from large-scale data allow researchers to understand complex social phenomena better. Also, the recent advancement of AI has empowered researchers and practitioners to investigate such massive data in an innovative way. My main research theme is developing AI-based methods and tools to 1) understand, predict, and nudge online human behavior and 2) tackle a wide range of social problems. In this talk, I will introduce two of my work on developing natural language processing methods and models to tackle social challenges.
First, I will introduce our novel technique, 'SemAxis,' to measure semantic changes in words across communities. Using transfer learning of word embeddings, SemAxis offers a framework to examine and interpret words on diverse semantic axes (732 systematically created semantic axes that capture common antonyms, such as safe vs. dangerous). Second, I will present my recent work on tackling anti-Asian hate during COVID-19. We used natural language processing techniques to characterize social media users who began to post anti-Asian hate messages during COVID-19 and build a prediction model to investigate the predictors of those users.

Bio: Jisun An is an Assistant Professor at the School of Computing and Information Systems, Singapore Management University (SMU-SCIS). She is a member of SODA (Social Data and AI) Lab (https://soda-labo.github.io), where she develops AI and NLP methods to understand, predict, and nudge online human behavior and to tackle various social problems, from media bias and framing, polarization, online hate, to healthy lifestyle and urban changes. Before joining SMU-SCIS, she was a scientist at Qatar Computing Research Institute, HBKU, and she received her Ph.D. in Computer Science from the University of Cambridge, UK.

Keynote Talk:

Elliot Ash
ETH Zurich

Abstract:

Bio: Elliott Ash is Assistant Professor of Law, Economics, and Data Science at ETH Zurich's Center for Law & Economics, Switzerland. Elliott's research and teaching focus on empirical analysis of the law and legal system using techniques from applied micro-econometrics, natural language processing, and machine learning. Prior to joining ETH, Elliott was Assistant Professor of Economics at University of Warwick, and before that a Postdoctoral Research Associate at Princeton University's Center for the study of Democratic Politics. He received a Ph.D. in economics and J.D. from Columbia University, a B.A. in economics, government, and philosophy from University of Texas at Austin, and an LL.M. in international criminal law from University of Amsterdam.

Keynote Talk: Challenges in NLP for Analyzing Social Media during Emerging Events

Anjalie Field
Stanford University

Abstract: Abstract: Social media has become a driving force in both online and offline events. Given huge volumes of organizations and people who generate text in short time periods, NLP should be a valuable tool in analyzing new data. However, developing NLP approaches that are robust to emerging domains and useable for research questions of interest remains difficult.
In this talk I will review some challenges we have encountered in using NLP to analyze social media during unfolding conflicts and social movements. First, I will present an analysis of emotions in tweets about the Black Lives Matter movement and our findings on how emotion data can shed interesting light on on-the-ground activism. Second, I will present an analysis of Twitter and VK posts in the ongoing Ukraine-Russia war in an attempt to identify propaganda strategies employed by state media. In both parts, I will highlight what worked and what didn't in using state-of-the-art NLP approaches and will suggest some directions for future research.

Bio: Anjalie Field is currently a postdoctoral researcher in the Stanford NLP Group and Stanford Data Science Institute working with Dan Jurafsky and Jennifer Eberhardt and an incoming Assistant Professor in Computer Science at Johns Hopkins University in the Fall of 2023. She completed her PhD at the Language Technologies Institute at Carnegie Mellon University, where she was advised by Yulia Tsvetkov and a member of TsvetShop. She was also a visiting student at the University of Washington in 2021-2022. Her primary interests involve using Natural Language Processing (NLP) to model social science concepts. Her current work is focused on identifying social biases in various domains, including Wikipedia, social media, and social workers' notes.

Table of Contents

Program

Wednesday, December 7, 2022

09:00 - 09:10 *Opening Remarks*

09:10 - 10:00 *Invited Speaker: Anjalie Field*

10:00 - 10:30 *Synchronous Talk Session 1: Influence due to Linguistic Variation*

Professional Presentation and Projected Power: A Case Study of Implicit Gender Information in English CVs
Jinrui Yang, Sheilla Njoto, Marc Cheong, Leah Ruppanner and Lea Frermann

Conditional Language Models for Community-Level Linguistic Variation
Bill Noble and Jean-philippe Bernardy

Predicting Long-Term Citations from Short-Term Linguistic Influence
Jacob Eisenstein, David Bamman and Sandeep Soni

10:30 - 11:00 *Coffee Break*

11:00 - 11:30 *Synchronous Talk Session 2: Political Frames and Stances*

Quotatives Indicate Decline in Objectivity in U.S. Political News
Tiancheng Hu, Manoel Horta Ribeiro, Robert West and Andreas Spitz

Capturing Topic Framing via Masked Language Modeling
Soroush Vosoughi, Weicheng Ma and Xiaobo Guo

Examining Political Rhetoric with Epistemic Stance Detection
Ankita Gupta, Su Lin Blodgett, Justin Gross and Brendan O'connor

11:30 - 12:30 *In-person Poster Session*

Improving the Generalizability of Text-Based Emotion Detection by Leveraging Transformers with Psycholinguistic Features
Sourabh Zanwar, Daniel Wiechmann, Yu Qiao and Elma Kerz

Professional Presentation and Projected Power: A Case Study of Implicit Gender Information in English CVs
Jinrui Yang, Sheilla Njoto, Marc Cheong, Leah Ruppanner and Lea Frermann

Quotatives Indicate Decline in Objectivity in U.S. Political News
Tiancheng Hu, Manoel Horta Ribeiro, Robert West and Andreas Spitz

Measuring Harmful Representations in Scandinavian Language Models
Samia Touileb and Debora Nozza

Fine-Grained Extraction and Classification of Skill Requirements in German-Speaking Job Ads
Ann-sophie Gnehm, Eva Bühlmann, Helen Buchs and Simon Clematide

Extracting Associations of Intersectional Identities with Discourse about Institution from Nigeria
Pavan Kantharaju and Sonja Schmer-galunder

No Word Embedding Model Is Perfect: Evaluating the Representation Accuracy for Social Bias in the Media
Henning Wachsmuth, Maximilian Keiff and Maximilian Spliethöver

You Are What You Talk About: Inducing Evaluative Topics for Personality Analysis
Jan Snajder, Iva Vukojević and Josip Jukić

Capturing Topic Framing via Masked Language Modeling
Soroush Vosoughi, Weicheng Ma and Xiaobo Guo

Opening up Minds with Argumentative Dialogues
Andreas Vlachos, Svetlana Stoyanchev, Tom Stafford, Paul Piwek, Jacopo Amidei, Charlotte Brand and Youmna Farag

Are Neural Topic Models Broken?
Philip Resnik, Pranav Goel, Rupak Sarkar and Alexander Miserlis Hoyle

Logical Fallacy Detection
Bernhard Schoelkopf, Rada Mihalcea, Mrinmaya Sachan, Zhiheng Lyu, Yiwen Ding, Xiaoyu Shen, Tejas Vaidhya, Abhinav Lalwani and Zhijing Jin

Wednesday, December 7, 2022 (continued)

12:30 - 14:00 *Lunch Break*

14:00 - 15:00 *Invited Speaker: Jisun An*

15:00 - 15:30 *Synchronous Talk Session 2: Heterogenous Real-World Social Data*

Fine-Grained Extraction and Classification of Skill Requirements in German-Speaking Job Ads
Ann-sophie Gnehm, Eva Bühlmann, Helen Buchs and Simon Clematide

Analyzing Norm Violations in Real-Time Live-Streaming Chat
Jihyung Moon, Dong-ho Lee, Hyundong Cho, Woojeong Jin, Chan Young Park, Minwoo Kim, Jay Pujara and Sungjoon Park

Status Biases in Deliberation Online: Evidence from a Randomized Experiment on ChangeMyView
Alan Montgomery, Yohan Jo and Emaad Manzoor

15:30 - 16:00 *Coffee Break*

16:00 - 17:00 *Invited Speaker: Elliot Ash*

17:00 - 20:00 *Break*

20:00 - 21:00 *Virtual Poster Session*

Fine-Grained Extraction and Classification of Skill Requirements in German-Speaking Job Ads
Ann-sophie Gnehm, Eva Bühlmann, Helen Buchs and Simon Clematide

Experiencer-Specific Emotion and Appraisal Prediction
Maximilian Wegge, Enrica Troiano, Laura Ana Maria Oberlaender and Roman Klinger

Understanding Narratives from Demographic Survey Data: a Comparative Study with Multiple Neural Topic Models
Xiao Xu, Gert Stulp, Antal Van Den Bosch and Anne Gauthier

An Analysis of Acknowledgments in NLP Conference Proceedings
Winston Wu

Extracting Associations of Intersectional Identities with Discourse about Institution from Nigeria
Pavan Kantharaju and Sonja Schmer-galunder

OLALA: Object-Level Active Learning for Efficient Document Layout Annotation
Zejiang Shen, Weining Li, Jian Zhao, Yaoliang Yu and Melissa Dell

Towards Few-Shot Identification of Morality Frames using In-Context Learning
Shamik Roy, Nishanth Sridhar Nakshatri and Dan Goldwasser

Analyzing Norm Violations in Real-Time Live-Streaming Chat
Jihyung Moon, Dong-ho Lee, Hyundong Cho, Woojeong Jin, Chan Young Park, Minwoo Kim, Jay Pujara and Sungjoon Park

Utilizing Weak Supervision to Create S3D: A Sarcasm Annotated Dataset
Jordan Painter, Helen Treharne and Diptesh Kanojia

A Robust Bias Mitigation Procedure Based on the Stereotype Content Model
Eddie Ungless, Amy Rafferty, Hrichika Nag and Björn Ross

Who is GPT-3? An exploration of personality, values and demographics
Marilù Miotto, Nicola Rossberg and Bennett Kleinberg

No Word Embedding Model Is Perfect: Evaluating the Representation Accuracy for Social Bias in the Media
Henning Wachsmuth, Maximilian Keiff and Maximilian Spliethöver

You Are What You Talk About: Inducing Evaluative Topics for Personality Analysis
Jan Snajder, Iva Vukojević and Josip Jukić

Status Biases in Deliberation Online: Evidence from a Randomized Experiment on ChangeMyView
Alan Montgomery, Yohan Jo and Emaad Manzoor

Predicting Long-Term Citations from Short-Term Linguistic Influence
Jacob Eisenstein, David Bamman and Sandeep Soni

Capturing Topic Framing via Masked Language Modeling
Soroush Vosoughi, Weicheng Ma and Xiaobo Guo

MBTI Personality Prediction for Fictional Characters Using Movie Scripts
Jeffrey Stanton, Jing Li, Dakuo Wang, Mo Yu, Xiangyang Mou and Yisi Sang

Are Neural Topic Models Broken?
Philip Resnik, Pranav Goel, Rupak Sarkar and Alexander Miserlis Hoyle

Logical Fallacy Detection
Bernhard Schoelkopf, Rada Mihalcea, Mrinmaya Sachan, Zhiheng Lyu, Yiwen Ding, Xiaoyu Shen, Tejas Vaidhya, Abhinav Lalwani and Zhijing Jin

A Critical Reflection and Forward Perspective on Empathy and Natural Language Processing
Lucie Flek, David Jurgens, Charles Welch and Allison Lahnala