COLING 2020

Proceedings of the 3rd NLP4IF Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda

Co-located with the 28th International Conference on Computational Linguistics COLING'2020

NLP4IF'2020

December 12, 2020
Barcelona, Spain (Online)

ISBN 978-1-952148-36-1

Preface

We are living in unprecedented times. Humanity is working together to fight the pandemic. The COVID-19 pandemic has altered our way of life and how we go about our daily routines. For many, kitchen and dining room tables have become the new office, classroom, or social meeting place. Virtual is the new reality, and while it allows many of us to continue to work, learn and socialize safely, it comes with its own unique set of challenges and dangers. As we transition to using more online platforms, scammers, hackers and other cybercriminals are finding new ways to target people. With social distancing as the new norm, socialization has turned even more digital than ever before. Unfortunately, cyber bullying is rampant and has been linked to a variety of negative effects, including anxiety, depression, and substance abuse. In addition, according to Freedom House[1], disinformation campaigns have been on the rise since March 2020. Cooordinated and covert attempts by China-linked actors to manipulate information, particularly regarding COVID-19, have been detected in many countries, including the United States, Argentina, Serbia, Italy, and Taiwan, with the relevant content often delivered in local languages. Our workshop tries to address these challenges.

Vijjali et al., for example, develop a two stage automated pipiline for COVID-19 fake news detection: the first model retrieves the most relevant facts concerning user claims about specific COVID-19-related claims; the second model verifies the level of truth in the claim by computing the textual entailment between the claim and the true facts retrieved from a manually curated COVID-19 dataset. The authors demonstrate the effectiveness of pre-trained Transformer-based language models in retrieving and classifying fake news and suggest that the model performance can be further boosted by adding domain-specific knowledge.

Shin et al. describe the communication patterns on the coronvirus pandemic from three Asian countries using a time-topic cohesive model, which can detect contextualized events based on topical and temporal information via contrastive learning. It then can be applied to multiple languages, enabling a comparison of risk communication across cultures. They present a case study and discuss future implications of the proposed model.

Information warfare has recently shifted its focus to social media, but traditional media continue to play an important role. In their work with civial society organizations and journalists around the world, Niven & Kau have identified the usefulness of NLP tools to detect and to track media that are aligned with the Chinese Communist Party (CPP). Niven & Kau operationalize alignment in terms of sociological definitions of media bias. They take as a case study the alignment of four Taiwanese media outlets to the CPP state media. They present the results of an initial investigation using the frequency of words in psychologically meaningful categories. Their findings suggest that the chosen word categories correlate with framing choices.

Prakash & Madabushi propose a novel architecture for the task of stance detection, which integrates count-based features into pre-trained models such as BERT and RoBERTa, with specific regard to the identification of disinformation and propaganda on social media, where handcrafted and count-based features can be particularly helpful.

Lily et al. focus on the task of satire detection, noticing that satire is often misinterpreted by readers as legitimate news. They propose a multi-modal approach based on the state-of-the-art visiolinguistic model ViLBERT. They create a new dataset consisting of images and headlines of regular and satirical news, fine-tune ViLBERT on the dataset and train a convolutional neural network that uses an image forensics technique. Evaluation on the dataset shows that their proposed multi-modal approach outperforms image-only, text-only, and simple fusion baselines.

[1] `http://freedomhouse.org/`

We are also thrilled to be able to bring two invited speakers: Stephan Lewandowsky from the University of Bristol with a talk on social media coverage and human decision making and Aylin Caliskan from George Washington University with a talk on the impact of machine intelligence on society, especially threats to fairness, privacy, and democracy.

We further feature a panel including Roya Ensafi and Veronica Perez-Rosas from the University of Michigan, Andreas Vlachos from the University of Cambridge, and Stephan Lewandowsky from the University of Bristol.

Last but not least, we would like to thank the program committee for their help with reviewing the papers, and with advertising the workshop.

The NLP4IF 2020 Organizers:

Giovanni Da San Martino, Qatar Computing Research Institute, Chris Brew, Giovanni Luca Ciampaglia, Anna Feldman, Chris Leberknight, Preslav Nakov

`http://www.netcopia.net/nlp4if/`

Organizers:

Giovanni Da San Martino, Qatar Computing Research Institute, HBKU
Chris Brew, LivePerson
Giovanni Luca Ciampaglia, University of South Florida
Anna Feldman, Montclair State University
Chris Leberknight, Montclair State University
Preslav Nakov, Qatar Computing Research Institute, HBKU

Program Committee:

Tariq Alhindi, Columbia University
Alberto Barrón-Cedeño, Università di Bologna
Jedidiah Crandall, Arizona State University
Anjalie Field, Carnegie Mellon University
Yiqing Hua, Cornell Tech
Jeffrey Knockel, The University of New Mexico
Henrique Lopes Cardoso, University of Porto
Hannah Rashkin. University of Washington

Invited Speakers:

Aylin Caliskan, George Washington University
Stephan Lewandowsky, University of Bristol

Panelists:

Roya Ensafi, University of Michigan
Stephan Lewandowsky, University of Bristol
Veronica Perez-Rosas, University of Michigan
Joshua Tucker, New York University
Andreas Vlachos, University of Cambridge

Invited Talk: Implications of Biased AI on Democracy, Equity, and Justice

Aylin Caliskan
George Washington University

Abstract: Billions of people on the internet are exposed to the outputs of downstream natural language processing (NLP) applications on a daily basis. Many of these NLP applications use word embeddings as general purpose language representations. In this talk, Aylin Caliskan will introduce the Word Embedding Association Test (WEAT) to demonstrate that word embeddings trained on language corpora embed the biases and associations documented by the Implicit Association Test in social psychology. In particular, WEAT measures how statistical regularities of language capture biases and stereotypes, such as racism, sexism, and attitudes toward social groups. Word embeddings are trained on text collected from the internet, that includes society's organic natural language data in addition to text from information influence operations. The adaptation of WEAT to the information influence domain automatically characterizes overall attitudes and biases associated with emerging information influence operations. Accurate analysis of these emerging topics usually requires laborious, manual analysis by experts to annotate a large set of data points to identify biases in new topics. We validate our practical and non-parametric method using known information operation-related tweets from Twitter's Transparency Report. We perform a case study on the COVID-19 pandemic to evaluate our method's performance on non-labeled Twitter data, demonstrating its usability in emerging domains.

Bio: Aylin Caliskan is an Assistant Professor of Computer Science at George Washington University. Her research interests lie in AI ethics, bias in AI, machine learning, and the implications of machine intelligence on fairness and privacy. She investigates the reasoning behind biased AI representations and decisions by developing explainability methods that uncover and quantify biases of machines. Building these transparency enhancing algorithms involves the heavy use of machine learning, natural language processing, and computer vision in novel ways to interpret AI and gain insights about bias in machines as well as society. In her recent publication in Science, she demonstrated how semantics derived from language corpora contain human-like biases. Prior to that, she developed novel privacy attacks to de-anonymize programmers using code stylometry. Her presentations on both de-anonymization and bias in machine learning are the recipients of best talk awards. Her work on semi-automated anonymization of writing style furthermore received the Privacy Enhancing Technologies Symposium Best Paper Award. Aylin holds a PhD in Computer Science from Drexel University and a Master of Science in Robotics from the University of Pennsylvania. Before joining the faculty at George Washington University, she was a Postdoctoral Researcher and a Fellow at Princeton University's Center for Information Technology Policy.

Invited Talk: Diversionary Agenda Setting and Micro-Targeted Persuasion

Stephan Lewandowsky
University of Bristol

Abstract: We are said to live in a "post-truth" era in which "fake news" has replaced real information, denial has compromised science, and the ontology of knowledge and truth has taken on a relativist element. I argue that to defend evidence-based reasoning and knowledge against those attacks, we must understand the strategies by which the post-truth world is driven forward. I depart from the premise that the post-truth era did not arise spontaneously but is the result of a highly effective political movement that deploys a large number of rhetorical strategies. I focus on two strategies: Diversionary agenda-setting by social media and the use of "micro-targeting" of political messages online. I present evidence for the existence of each strategy and its impact, and how it might be countered.

Bio: Professor Stephan Lewandowsky is a cognitive scientist at the University of Bristol. He was an Australian Professorial Fellow from 2007 to 2012, and was awarded a Discovery Outstanding Researcher Award from the Australian Research Council in 2011. His research examines people's memory, decision making, and knowledge structures, with a particular emphasis on how people update information in memory. His most recent research interests examine the potential conflict between human cognition and the physics of the global climate, which has led him into research in climate science and climate modeling. He has published more than 200 scholarly articles, chapters, and books, including numerous papers on how people respond to corrections of misinformation and what variables determine people's acceptance of scientific findings. Professor Lewandowsky is an award-winning teacher and was Associate Editor of the Journal of Experimental Psychology: Learning, Memory, and Cognition from 2006-2008. He has also contributed around 50 opinion pieces to the global media on issues related to climate change "skepticism" and the coverage of science in the media. He is currently serving as Digital Content Editor for the Psychonomic Society and blogs routinely on cognitive research.

Panelists

Roya Ensafi is an assistant professor in computer science and engineering at the University of Michigan, where her research focuses on computer networking, security, and privacy. Her notable projects with real-world impact include founding Censored Planet, a global censorship observatory, and researching the Kazakhstan HTTPS MitM interception, the Great Cannon of China, and large-scale study of server-side geoblocking. She has received the NSF CISE Research Initiation Initiative award, the Google Faculty Research Award, and Consumer Report Digital Lab fellowship. Roya's work has appeared in the popular press publications such as the NY Times, Wired, Business Insider, and ArsTechnica.

Stephan Lewandowsky (see above for his biography)

Veronica Perez-Rosas is an assistant research scientist at the University of Michigan. Her research interests include machine learning, natural language processing, computational linguistics, affect recognition, and multimodal analysis of human behavior. Her research focuses on developing computational methods to analyze, recognize, and predict human affective responses during social interactions. She has authored papers in leading conferences and journals in Natural Language Processing and Computational Linguistics, and she served as a program committee member for multiple international journals and conferences in the same fields.

Joshua Tucker is Professor of Politics, affiliated Professor of Russian and Slavic Studies, and affiliated Professor of Data Science at New York University. He is the Director of NYU's Jordan Center for Advanced Study of Russia, a co-Director of the NYU Center for Social Media and Politics, and a co-author/editor of the award-winning politics and policy blog The Monkey Cage at The Washington Post. He serves on the advisory board of the American National Election Study, the Comparative Study of Electoral Systems, and numerous academic journals, and was the co-founder and co-editor of the Journal of Experimental Political Science. His original research was on mass political behavior in post-communist countries, including voting and elections, partisanship, public opinion formation, and protest participation. More recently, he has been at the forefront of the newly emerging field of study of the relationship between social media and politics. His research in this area has included studies on the effects of network diversity on tolerance, partisan echo chambers, online hate speech, the effects of exposure to social media on political knowledge, online networks and protest, disinformation and fake news, how authoritarian regimes respond to online opposition, and Russian bots and trolls, and he is currently the co-Chair of the independent academic advisory board of the 2020 Facebook Election Research Study. An internationally recognized scholar, he has served as a keynote speaker for conferences in Sweden, Denmark, Italy, Brazil, the Netherlands, Russia, and the United States, and has given more than 100 invited research presentations at top domestic and international universities and research centers over the past decade. His research has appeared in over two-dozen scholarly journals and has been supported by a wide range of philanthropic foundations, as well as multiple grants from the National Science Foundation. His most recent books are the co-authored Communism's Shadow: Historical Legacies and Contemporary Political Attitudes (Princeton University Press, 2017), and the co-edited Social Media and Democracy: The State of the Field (Cambridge University Press, 2020).

Andreas Vlachos is a senior lecturer at the Natural Language and Information Processing group at the Department of Computer Science and Technology at the University of Cambridge. Current projects include dialogue modelling, automated fact checking and imitation learning. He has also worked on semantic parsing, natural language generation and summarization, language modelling, information extraction, active learning, clustering, and biomedical text mining. His research is supported by ERC, EPSRC, ESRC, Facebook, Amazon, Google and Huawei.

Table of Contents

Conference Program

Two Stage Transformer Model for COVID-19 Fake News Detection and Fact Checking
Rutvik Vijjali, Prathyush Potluri, Siddharth Kumar and Sundeep Teki

Measuring Alignment to Authoritarian State Media as Framing Bias
Timothy Niven and Hung-Yu Kao

Incorporating Count-Based Features into Pre-Trained Models for Improved Stance Detection
Anushka Prakash and Harish Tayyar Madabushi

A Multi-Modal Method for Satire Detection using Textual and Visual Cues
Lily Li, Or Levi, Pedram Hosseini and David Broniatowski

A Risk Communication Event Detection Model via Contrastive Learning
Mingi Shin, Sungwon Han, Sungkyu Park and Meeyoung Cha