GeBNLP 2019

Gender Bias in Natural Language Processing

Proceedings of the First Workshop

2 August 2019
Florence, Italy

©2019 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 978-1-950737-40-6

Introduction

Natural Language Processing (NLP) is pervasive in the technologies people use to curate, consume and create information on a daily basis. Moreover, it is increasingly found in decision-support systems in finance and healthcare, where it can have a powerful impact on peoples' lives and society as a whole. Unfortunately, far from the objective and calculating algorithms of popular perception, machine learned models for NLP can include subtle biases drawn from the data used to build them and the builders themselves. Gender-based stereotypes and discrimination are problems that human societies have long struggled with, and it is the community's responsibility to ensure fair and ethical progress. The increasing volume of work in the last few years is a strong signal that researchers and engineers in academia and industry do care about fairer NLP.

This volume contains the proceedings of the First Workshop on Gender Bias in Natural Language Processing held in conjunction with the 57th Annual Meeting of the Association for Computational Linguistics in Florence. The workshop received 19 submissions of technical papers (8 long papers, 11 short papers), of which 13 were accepted (5 long, 8 short), for an acceptance rate of 68%. We have to thank the high-quality selection of research works thanks to the Program Committee members which provided extremely valuable reviews. The accepted papers cover a diverse range of topics related to the analysis, measurement and mitigation of gender bias in NLP. Many of the papers investigate how automatically learned vector space representations of words are affected by gender bias, but the programme also features papers on NLP applications such as machine translation, sentiment analysis and text classification. In addition to the technical papers, the workshop also included a very popular shared task on gender-fair coreference resolution, which attracted submissions from 263 participants. Many of them achieved excellent performance. Of the 11 submitted system description papers, 10 were accepted for publication. We are very grateful to Google for providing a generous prize pool of 25,000 USD for the shared task, and to the Kaggle team for their great help with the organisation of the shared task.

Finally, the workshop counts on two impressive keynote speakers: Melvin Johnson and Pascale Fung, who will provide insides in gender-specific translations in the Google system and the gender roles in the artificial intelligence world, respectively.

We are very excited about the interest that this workshop has generated and we look forward to a lively discussion about how to tackle bias problems in NLP applications when we meet in Florence!

June 2019

Marta R. Costa-jussà
Christian Hardmeier
Will Radford
Kellie Webster

Organizers:

Marta R. Costa-jussà, Universitat Politècnica de Catalunya (Spain)
Christian Hardmeier, Uppsala University (Sweden)
Will Radford, Canva (Australia)
Kellie Webster, Google AI Language (USA)

Program Committee:

Kai-Wei Chang, University of California, Los Angeles (USA)
Ryan Cotterell, University of Cambridge (UK)
Lucie Flekova, Amazon Alexa AI (Germany)
Mercedes García-Martínez, Pangeanic (Spain)
Zhengxian Gong, Soochow University (China)
Ben Hachey, Ergo AI (Australia)
Svetlana Kiritchenko, National Research Council (Canada)
Sharid Loáiciga, University of Gothenburg (Sweden)
Kaiji Lu, Carnegie Mellon University (USA)
Saif Mohammad, National Research Council (Canada)
Marta Recasens, Google (USA)
Rachel Rudinger, Johns Hopkins University (USA)
Bonnie Webber, University of Edinburgh (UK)

Invited Speakers:

Melvin Johnson, Google Translate (USA)
Pascale Fung, Hong Kong University of Science and Technology (China)

Table of Contents

Gendered Ambiguous Pronoun (GAP) Shared Task at the Gender Bias in NLP Workshop 2019
Kellie Webster, Marta R. Costa-jussà, Christian Hardmeier and Will Radford 1

Proposed Taxonomy for Gender Bias in Text; A Filtering Methodology for the Gender Generalization Subtype
Yasmeen Hitti, Eunbee Jang, Ines Moreno and Carolyne Pelletier 8

Relating Word Embedding Gender Biases to Gender Gaps: A Cross-Cultural Analysis
Scott Friedman, Sonja Schmer-Galunder, Anthony Chen and Jeffrey Rye 18

Measuring Gender Bias in Word Embeddings across Domains and Discovering New Gender Bias Word Categories
Kaytlin Chaloner and Alfredo Maldonado 25

Evaluating the Underlying Gender Bias in Contextualized Word Embeddings
Christine Basta, Marta R. Costa-jussà and Noe Casas 33

Conceptor Debiasing of Word Representations Evaluated on WEAT
Saket Karve, Lyle Ungar and João Sedoc 40

Filling Gender & Number Gaps in Neural Machine Translation with Black-box Context Injection
Amit Moryossef, Roee Aharoni and Yoav Goldberg 50

The Role of Protected Class Word Lists in Bias Identification of Contextualized Word Representations
João Sedoc and Lyle Ungar 56

Good Secretaries, Bad Truck Drivers? Occupational Gender Stereotypes in Sentiment Analysis
Jayadev Bhaskaran and Isha Bhallamudi 65

Debiasing Embeddings for Reduced Gender Bias in Text Classification
Flavien Prost, Nithum Thain and Tolga Bolukbasi 72

BERT Masked Language Modeling for Co-reference Resolution
Felipe Alfaro, Marta R. Costa-jussà and José A. R. Fonollosa 79

Transfer Learning from Pre-trained BERT for Pronoun Resolution
Xingce Bao and Qianqian Qiao 85

MSnet: A BERT-based Network for Gendered Pronoun Resolution
Zili Wang 92

Look Again at the Syntax: Relational Graph Convolutional Network for Gendered Ambiguous Pronoun Resolution
Yinchuan Xu and Junlin Yang 99

Fill the GAP: Exploiting BERT for Pronoun Resolution
Kai-Chou Yang, Timothy Niven, Tzu Hsuan Chou and Hung-Yu Kao 105

On GAP Coreference Resolution Shared Task: Insights from the 3rd Place Solution
Artem Abzaliev 110

Resolving Gendered Ambiguous Pronouns with BERT
Matei Ionita, Yury Kashnitsky, Ken Krige, Vladimir Larin, Atanas Atanasov and Dennis Logvinenko 116

Anonymized BERT: An Augmentation Approach to the Gendered Pronoun Resolution Challenge
Bo Liu 123

Gendered Pronoun Resolution using BERT and an Extractive Question Answering Formulation
Rakesh Chada 129

Gendered Ambiguous Pronouns Shared Task: Boosting Model Confidence by Evidence Pooling
Sandeep Attree 137

Equalizing Gender Bias in Neural Machine Translation with Word Embeddings Techniques
Joel Escudé Font and Marta R. Costa-jussà 150

Automatic Gender Identification and Reinflection in Arabic
Nizar Habash, Houda Bouamor and Christine Chung 158

Quantifying Social Biases in Contextual Word Representations
Keita Kurita, Nidhi Vyas, Ayush Pareek, Alan W Black and Yulia Tsvetkov 169

On Measuring Gender Bias in Translation of Gender-neutral Pronouns
Won Ik Cho, Ji Won Kim, Seok Min Kim and Nam Soo Kim 176

Workshop Program

Friday, August 2, 2019

09:00–09:10 Opening

09:10–10:00 Keynote 1

Providing Gender-Specific Translations in Google Translate and beyond
Melvin Johnson

10:00–10:30 Shared Task Overview

Gendered Ambiguous Pronoun (GAP) Shared Task at the Gender Bias in NLP Workshop 2019
Kellie Webster, Marta R. Costa-jussà, Christian Hardmeier and Will Radford

10:30–11:00 Refreshments

11:00–12:30 Oral presentations 1: Representations

11:00–11:20 *Proposed Taxonomy for Gender Bias in Text; A Filtering Methodology for the Gender Generalization Subtype*
Yasmeen Hitti, Eunbee Jang, Ines Moreno and Carolyne Pelletier

11:20–11:35 *Relating Word Embedding Gender Biases to Gender Gaps: A Cross-Cultural Analysis*
Scott Friedman, Sonja Schmer-Galunder, Anthony Chen and Jeffrey Rye

11:35–11:50 *Measuring Gender Bias in Word Embeddings across Domains and Discovering New Gender Bias Word Categories*
Kaytlin Chaloner and Alfredo Maldonado

11:50–12:05 *Evaluating the Underlying Gender Bias in Contextualized Word Embeddings*
Christine Basta, Marta R. Costa-jussà and Noe Casas

12:05–12:25 *Conceptor Debiasing of Word Representations Evaluated on WEAT*
Saket Karve, Lyle Ungar and João Sedoc

Friday, August 2, 2019 (continued)

12:30–14:00 Lunch Break

14:00–14:50 Keynote 2

Gender Roles in the AI World
Pascale Fung

14:50–16:00 Poster session

Filling Gender & Number Gaps in Neural Machine Translation with Black-box Context Injection
Amit Moryossef, Roee Aharoni and Yoav Goldberg

The Role of Protected Class Word Lists in Bias Identification of Contextualized Word Representations
João Sedoc and Lyle Ungar

Good Secretaries, Bad Truck Drivers? Occupational Gender Stereotypes in Sentiment Analysis
Jayadev Bhaskaran and Isha Bhallamudi

Debiasing Embeddings for Reduced Gender Bias in Text Classification
Flavien Prost, Nithum Thain and Tolga Bolukbasi

BERT Masked Language Modeling for Co-reference Resolution
Felipe Alfaro, Marta R. Costa-jussà and José A. R. Fonollosa

Transfer Learning from Pre-trained BERT for Pronoun Resolution
Xingce Bao and Qianqian Qiao

MSnet: A BERT-based Network for Gendered Pronoun Resolution
Zili Wang

Look Again at the Syntax: Relational Graph Convolutional Network for Gendered Ambiguous Pronoun Resolution
Yinchuan Xu and Junlin Yang

Friday, August 2, 2019 (continued)

Fill the GAP: Exploiting BERT for Pronoun Resolution
Kai-Chou Yang, Timothy Niven, Tzu Hsuan Chou and Hung-Yu Kao

On GAP Coreference Resolution Shared Task: Insights from the 3rd Place Solution
Artem Abzaliev

Resolving Gendered Ambiguous Pronouns with BERT
Matei Ionita, Yury Kashnitsky, Ken Krige, Vladimir Larin, Atanas Atanasov and Dennis Logvinenko

Anonymized BERT: An Augmentation Approach to the Gendered Pronoun Resolution Challenge
Bo Liu

Gendered Pronoun Resolution using BERT and an Extractive Question Answering Formulation
Rakesh Chada

Gendered Ambiguous Pronouns Shared Task: Boosting Model Confidence by Evidence Pooling
Sandeep Attree

15:30–16:00 Mid-afternoon Snacks (Poster session continues)

16:00–17:15 Oral Presentations 2: Applications

16:00–16:20 *Equalizing Gender Bias in Neural Machine Translation with Word Embeddings Techniques*
Joel Escudé Font and Marta R. Costa-jussà

16:20–16:40 *Automatic Gender Identification and Reinflection in Arabic*
Nizar Habash, Houda Bouamor and Christine Chung

16:40–16:55 *Quantifying Social Biases in Contextual Word Representations*
Keita Kurita, Nidhi Vyas, Ayush Pareek, Alan W Black and Yulia Tsvetkov

16:55–17:15 *On Measuring Gender Bias in Translation of Gender-neutral Pronouns*
Won Ik Cho, Ji Won Kim, Seok Min Kim and Nam Soo Kim

17:15–17:30 Closing Remarks