Chapter 15
Gender-Based Tweet Analysis (GTA)

Dipti P. Rana
https://orcid.org/0000-0002-5058-1355
Sardar Vallabhbhai National Institute of Technology, Surat, India

Navodita Saini
Sardar Vallabhbhai National Institute of Technology, Surat, India

ABSTRACT

Each gender is having special personality and behavior characteristics that can be naturally reflected in the language used on social media to review, spread information, make relationships, etc. This information is used by different agencies for their profits. The magnified study of this information can reflect the implicit biases of their creators' gender. The ratio of gender is imbalanced across the global world, social media, discussion, etc. Twitter is used to discuss the issues caused by COVID-19 disease like its symptoms, mental health, advice, etc. This fascinating information motivated this research to propose the methodology gender-based tweet analysis (GTA) to study and magnify gender's impact on emotions of tweet data. The analysis of the experiment discovered the biases of gender on emotions of tweet data and highlighted the future real-world applications which may become more productive if gender biases are considered for the safety and benefit of society.

INTRODUCTION

The modern digital era has generated big data which are used by many researchers to discover the pattern based on content, emotions, behavior, network surfing, etc. (Gaind, Syal, Padgalwar, 2019; Kaushal & Patwardhan, 2018). These all patterns if analyzed in detail, one can observe that these patterns have an impact on gender. Each gender is having its special personality and behavior characteristics, which they are showing with the help of the language of expression when providing a review, information spreading, make friends, follow the topics/other users, etc. All this information without considering the biases of gender used by different agencies for their profits to recommend the products in terms of the

DOI: 10.4018/978-1-7998-7371-6.ch015

Copyright © 2021, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

recommendation of items, links, movies, songs, etc. The syntactic and semantic linguistic features used by the user can reproduce the implicit biases of their creators' gender.

The ratio of both genders male and female is imbalanced across the global world. In addition, this ratio is imbalanced on social media as registered users, users' status, users per discussion, etc. The author (Top Twitter Demographics That Matter to Social Media Marketers, 2021) said that non-technical social media networks like Snapchat, Pinterest, Instagram, Facebook, etc. have more females than males, whereas the technical and scholarly social media networks like Twitter, LinkedIn, etc. have fewer females than the males. One of the most popular social network applications "twitter" shows gender discrimination in its users. The global audience of Twitter shows skewness towards the male gender with 62% over 38% female. From which only in the US, the gender division is 50 - 50 without the age information. However, among U.S. adults, the most active users are women with 65% and having more private Twitter accounts than men are. While in India, the most active users are men with 85%, which is the third country in the highest male-to-female ratio ranking. Other countries like the Brunei, Indonesia, Laos, Malaysia, Philippines, and Thailand have the highest female Twitter users.

The purpose of the usage of social media for females and males is completely different. Generally, social networking sites used by female gender is to make contacts and stay in connection with friends and family, while social networking sites used by male gender, to assimilate the data required to form influence. Social media aids in research, gather relevant contacts, and ultimately increase their status. This gender imbalance is having its impact and researchers are discussing this.

In the medical science field especially in psychology, the researchers are studying gender-based information for a long time. They identified gender-linked features using the comparison of text samples from known males and females and compared the calculated frequencies of interesting features from text like usage of the first-person singular, a positive self-presentation or other, and then interpreted the results in terms of psychological theory (Leaper & Ayres, 2007; Newman, Groom, Handelman, & Pennebaker, 2008; Fast, & Funder, 2010).

However, in the machine learning research area, gender-based research is still in its fantasy. The imbalance ratio of gender attracted the researchers to look at its impact on the emotions of text data. During the tough time like the pandemic of COVID-19, different social media platforms used by the users during the pandemic year 2019-20 to discuss the issues caused by the novel virus disease like its symptoms, mental health, health services, advice, recovery rate, hospitalization/death ratio, containment area, supply chain, etc. These period tweets involved a variety of discussions with different emotions from different categories of people like gender, profession, age, culture, etc. From all these categories of people, the author identified "gender-biased issue" as an eye-catching issue because of its applicability in a wide variety of domains.

To yield the promising knowledge outcome from the shared information by assessing the gender's impact requires gender identification. Traditionally the only way to identify the gender is through the explicit gender feature or separate gender-based questionnaire. However, in this digital world, the challenges posed to the researchers for the identification of gender due to unavailability of users' personal information hidden behind the fantasy or ambiguous name and due to the policy, the data provider is not availing the gender information.

To analyze the impact on the emotion of tweet text from the perspective of gender, this research proposed the methodology Gender based Tweet Analysis (GTA). The preprocessing of the tweet text performed to use the data for the further step. To consider the data partition based on gender, the username utilized here to identify the type of user is male, female, or organization followed by the gender

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/gender-based-tweet-analysis-gta/280922

Related Content

An Experimental Evaluation of Debayering Algorithms on GPUs for Recording Panoramic Video in Real-Time
Ragnar Langseth, Vamsidhar Reddy Gaddam, Håkon Kvale Stensland, Carsten Griwodz, Pål Halvorsenand Dag Johansen (2015). *International Journal of Multimedia Data Engineering and Management (pp. 1-16).*
www.irma-international.org/article/an-experimental-evaluation-of-debayering-algorithms-on-gpus-for-recording-panoramic-video-in-real-time/132684

Blockchanging Money: Reengineering the Free World Incentive System
Dario de Oliveira Rodrigues (2021). *Political and Economic Implications of Blockchain Technology in Business and Healthcare (pp. 69-117).*
www.irma-international.org/chapter/blockchanging-money/282336

The Real Blockchain Game Changer: Protocols and DAOs for Coordinating Work to Provide Goods and Services
Mitchell Loureiro, Ana Pêgoand Inês Graça Raposo (2021). *Political and Economic Implications of Blockchain Technology in Business and Healthcare (pp. 160-172).*
www.irma-international.org/chapter/the-real-blockchain-game-changer/282338

Comparison of Light Field and Conventional Near-Eye AR Displays in Virtual-Real Integration Efficiency
Wei-An Teng, Su-Ling Yehand Homer H. Chen (2023). *International Journal of Multimedia Data Engineering and Management (pp. 1-17).*
www.irma-international.org/article/comparison-of-light-field-and-conventional-near-eye-ar-displays-in-virtual-real-integration-efficiency/333609

Fetal ECG Extraction: Principal Component Analysis Method for Extraction of Fetal ECG
Vidya Sujit Kurtadikarand Himangi Milind Pande (2022). *Designing User Interfaces With a Data Science Approach (pp. 275-294).*
www.irma-international.org/chapter/fetal-ecg-extraction/299757