Yijie Bai, Tabib Wasit Rahman, Zhisheng Jin

CSC447: Natural Language Processing

November 15, 2023

Course Project Proposal

Social Media Study and Opinion Analysis of Public Opinion on the Palestinian-Israeli Conflict in 2023

The Israeli-Palestinian conflict has deep historical roots, tracing back to the early 20th century. It primarily revolves around two national movements laying claim to the same territory: the Zionist movement (Jewish nationalism) and Palestinian nationalism. The establishment of Israel in 1948 and the subsequent wars, particularly the 1948 Arab-Israeli War and the 1967 Six-Day War, led to profound territorial changes and a complex refugee crisis. This longstanding conflict has been marked by periods of intense violence, failed peace initiatives, and enduring political and territorial disputes.

The conflict has garnered immense international attention, with opinions ranging from strong support for one side to calls for peaceful coexistence. Media portrayal has had a significant impact on shaping global perspectives, but with the rise of social media, there's been a shift. Platforms like Reddit have democratized information dissemination, allowing individuals worldwide to voice their opinions, share experiences, and influence public sentiment. This shift has introduced a more diverse array of narratives, making the understanding of public opinion more complex yet more necessary. [1]

Social media has evolved into a battleground for narratives and ideologies. In the context of the Israeli-Palestinian conflict, it's a tool for advocacy, activism, and sometimes misinformation. For many, it's a primary source of news and a platform for solidarity or protest. [2] The diverse user base of Reddit, with its myriad of perspectives, provides a unique and rich data source to analyze public opinion. Understanding these sentiments can offer insights into how the conflict is perceived globally and how it affects and is affected by public opinion.

Analyzing social media data to gauge public opinion on the Israeli-Palestinian conflict holds significant value. It can reveal how narratives and sentiments evolve, differ, or converge over time and across different demographic groups. This analysis can contribute to a more nuanced understanding of the conflict's perception, aiding policymakers, conflict resolution practitioners, and scholars in their efforts to address the complex dynamics of this enduring conflict.

Current Outline:

Problem Statement:

The Israeli-Palestinian conflict has been a source of global concern for decades. With the advent of social media, people worldwide express their views on this issue, influencing public opinion and policy-making. However, the vast amount of data on social platforms like Reddit makes it challenging to gauge the general sentiment accurately. This project seeks to analyze these expressions to categorize public opinion into positive, negative, and neutral stances, and to identify the focal themes within these perspectives, such as civilian casualties or longstanding segregation policies.

Data Collection:

We will employ custom-built web scraping tools to collect relevant posts from Reddit. This data will encompass various threads and comments that discuss the Israeli-Palestinian conflict. The data will be timestamped to allow for temporal analysis and understand how opinions may have shifted over time.

Methodologies and algorithms:

- Data Preprocessing: We will clean and preprocess the data, including handling missing values, removing irrelevant information, and standardizing the format for analysis.
- Exploratory Data Analysis (EDA): This involves statistically analyzing the data, identifying patterns, and visualizing these insights using graphs and charts.
- Sentiment Analysis: We will utilize a XLNet model [3] to categorize sentiments and employ machine learning algorithms to enhance accuracy.
- Feature Importance Analysis: By employing various analytical techniques, using LDA model, we'll identify the most influential factors shaping public opinion.
- Predictive Modeling: Different machine learning algorithms, such as Random Forest, Support Vector Machines, and Neural Networks, will be experimented with to find the best fit. Parameter tuning will be conducted to optimize the models.
- Other possible algorithms: we may use other algorithms in this project.

Performance Evaluation:

- Accuracy of Sentiment Analysis: This will be evaluated using precision, recall, and F1-score metrics, ensuring that our model correctly classifies sentiments.
- Relevance of Feature Importance Analysis: We'll measure the impact of identified features on public opinion through correlation coefficients and significance testing.
- Predictive Model Accuracy: The performance of our predictive models will be evaluated using metrics like accuracy, area under the ROC curve, and confusion matrices. We will also conduct cross-validation to ensure the model's robustness. [4]

By meticulously following this proposed approach, we aim to provide a comprehensive analysis of public opinion on the Israeli-Palestinian conflict through social media data, offering valuable insights for researchers, policymakers, and the general public.

Reference:

[1] Imtiaz, Arsal, Danish Khan, Hanjia Lyu and Jiebo Luo, Taking sides: Public Opinion over the Israel-Palestine Conflict in 2021, ArXiv abs/2201.05961 (2022): n. pag.

[2] Hanjia Lyu, Junda Wang, Wei Wu, Viet Duong, Xiyang Zhang, Timothy D. Dye, Jiebo Luo, Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination, Intelligent Medicine, Volume 2, Issue 1, 2022, Pages 1-12, ISSN 2667-1026

[3] Enting Zhou, Yurong Liu, Hanjia Lyu, et al. A Fine-Grained Analysis of Public Opinion toward Chinese Technology Companies on Reddit. arXiv:2201.05538, 2023 GB/T 7714

[4] Singh, J., Singh, G. & Singh, R. Optimization of sentiment analysis using machine learning classifiers. Hum. Cent. Comput. Inf. Sci. 7, 32 (2017).