Sentiment Analysis in Deep Learning: A Study on Bias Evaluation

Murali Tanneru
Illinois Institute of Technology
Chicago, Illinois
mtanneru@hawk.iit.edu

Subhash Prabhakar Kamath
Illinois Institute of Technology
Chicago, Illinois
skamath3@hawk.iit.edu

Somu Medaka
Illinois Institute of Technology
Chicago, Illinois
smedaka@hawk.iit.edu

October 1, 2023

Abstract

Sentiment Analysis in social media has gotten a lot of attention in recent years, with Twitter being a popular venue for evaluating public mood. While correct sentiment categorization on Twitter data is critical, it remains a difficult task due to the elusiveness of establishing consistent results. The purpose of this project is to identify a robust deep learning model for Twitter sentiment analysis that not only accurate in classification but also emphasizes the critical issue of bias in sentiment evaluation.

1 Project Description

In recent times, there has been a noticeable shift in human behavior due to our increased engagement with innovation, especially in the realm of social media. People are now using the internet and online platforms more frequently, and this shift is gradually reshaping our behavior. Among these platforms, Twitter stands out as a trustworthy source of information and is frequently used authoritatively by many individuals. To enhance the quality of discourse and reduce objectionable language, our project aims to develop an effective deep learning model for sentiment analysis of tweets, while minimizing bias in the process.

For our project, we intend to utilize the Sentiment140 [1]dataset, which comprises 1,600,000 tweets obtained through the Twitter API and manually annotated with sentiment labels (0 for negative and 4 for positive). We will explore various deep learning models, including Convolutional Neural Networks (CNNs), Bidirectional Long Short-Term Memory networks (LSTMs), Bidirectional LSTMs combined with CNNs, and RoBERTa, to train and evaluate their performance in sentiment analysis.

Furthermore, we are committed to assessing and addressing potential biases in our models. To achieve this, we will employ the Equity Evaluation Corpus dataset obtained from paper [2], which will enable us to identify and mitigate any sources of bias that may exist within our models. By integrating these measures into our project, our objective is to identify the most accurate, reliable, and fair sentiment analysis model for Twitter data.

1.1 Existing Work

Numerous research studies have been conducted to analyze sentiment in tweets on various topics, utilizing a range of machine learning models. However,

it's important to note that most of these studies have primarily focused on just individual system, without giving significant consideration to the potential bias within their models.

"Comparative study of Twitter Sentiment On COVID-19 Tweets" [3] in this study researchers compare the effectiveness of Logistic Regression sentiment analysis, VADER sentiment analysis, and BERT sentiment analysis in determining sentiment within COVID-19 tweets.

"Twitter Sentiment Classification: Positive, Negative, or Neutral?" [5] In this paper, sentiment analysis is performed on Twitter datasets manually annotated as positive, negative, or neutral. Naive Bayes, RandomForest, and the J48 Decision Tree Multiclass classifier are used alongside an iterative classifier optimizer to improve accuracy.

"Analysis of Sentiments on Russia-Ukraine Conflict using Machine Learning Techniques" [4]This study explores sentiment in tweets related to the Russia-Ukraine conflict, employing machine learning techniques such as Random Forest, Xtreme Gradient Boosting, Support Vector Machine, and Logistic Regression. Comparative performance analysis is conducted using various metrics.

While these studies offer valuable insights into sentiment trends, they underscore the need for future research to address potential bias within sentiment analysis models, particularly in the context of social media data.

References

[1] A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision, 01 2009. Dataset URL: http://www.sentiment140.com/.

[2] S. Kiritchenko and S. Mohammad. Examining gender and race bias in two hundred sentiment analysis systems, 01 2018. Dataset URL: http://saifmohammad.com/WebPages/Biases-SA.html.

[3] A. J. Nair, V. G, and A. Vinayak. Comparative study of twitter sentiment on covid - 19 tweets. pages 1773–1778, 2021.

[4] S. Negi, I. Dawar, R. Prakash, P. Tiwari, M. Bhushan, and A. Kumar. Analysis of sentiments on russia-ukraine conflict using machine learning techniques. 1:1–6, 2023.

[5] M. F. Çeliktuğ. Twitter sentiment analysis, 3-way classification: Positive, negative or neutral? pages 2098–2103, 2018.

1.2 MileStones

MileStone	Task	Date	Team Member
Milestone 1	CNN Training and Evaluation	10-Oct	Murali
Milestone 2	Bidirectional LSTM Training and Evaluation	10-Oct	Prabhakar
Milestone 3	Bidirectional LSTM with CNN Training and Evaluation	15-Oct	Somu
Milestone 4	Roberta Training and Evaluation	15-Oct	Murali
Milestone 5	Intermediate Results	30-Oct	Murali Prabhakar Somu
Milestone 6	CNN Bias Evaluation	5-Nov	Prabhakar
Milestone 7	Bidirectional LSTM Bias Evaluation	5-Nov	Somu
Milestone 8	Bidirectional LSTM with CNN Bias Evaluation	15-Nov	Murali
Milestone 9	Roberta Bias Evaluation	15-Nov	Murali
Milestone 10	Project Presentation and Report	29-Nov	Murali Prabhakar Somu