

Deep Learning Techniques for Sentiment Analysis of Users on Social Media.

Wiam Khalid Alharbi.

Table Contents

ABSTRACT	3
INTRODUCTION.....	4
BACKGROUND	5
PROBLEM STATEMENT	6
RESEARCH OBJECTIVES	8
RESEARCH SCOPE	9
ANALYSIS OF CURRENT LIMITATIONS IN SENTIMENT ANALYSIS:	9
EVALUATION OF EXISTING MODELS AND TECHNIQUES:	9
DEVELOPMENT OF A HYBRID DEEP LEARNING FRAMEWORK:	9
VALIDATION ACROSS SOCIAL MEDIA DOMAINS	9
LITERATURE REVIEW	10
RESEARCH METHODOLOGY	11
EXPECTED OUTCOMES.....	12
GANTT CHART	13
REFERENCES.....	14

Abstract

Sentiment analysis has become an essential part in shaping how we understand public opinion, particularly within the ever-shifting discourse of the internet. However, machine learning models based on traditional methods are rarely able to track this phenomenon as it runs counter to the weak assumptions of language and causality that language has and transport them to modern #mediacorpology, especially in social media vectors on the Internet. In this paper, a hybrid deep learning framework combining CNN, LSTM and SVM architectures with cutting-edge embedding techniques including BERT and Word2Vec is proposed. It aims at improved accuracy of sentiment classification, solving adaptability problems for a certain domain, emotion detection and real-time applications. This research aims to provide more generalizable and scalable approaches to sentiment analysis in multiple settings by applying the suggested models to a wide variety of datasets from Twitter, Yelp, and Amazon.

Introduction

The meteoric rise of social media has forever changed how we express ourselves: opinions, emotions, reactions. Platforms such as Twitter, Facebook, and Yelp are potential treasure troves of unstructured data that can provide insights into public sentiment on matters ranging from product preferences to political opinions. For businesses, researchers, and policymakers aiming to respond to public opinion, understanding these sentiments has been critical.

The opinion mining or sentiment analysis has built on high standard with a integration of deep learning technique. Yet the relatively fluid nature of online communication — full of slang, sarcasm, emerging phrases, and multilingual dialogue — poses profound challenges to existing models. Traditional machine-learning algorithms typically do not have the robustness to handle these nuances of language, while various deep learning methods, although accurate, still do not transfer well and generalize well over new domains.

Many of these challenges are addressed in this research study investigating a hybrid deep learning model that combines the strengths of several architectures. When CNN extract spatial features, LSTM is set to learn the temporal sequence of data and an SVM based classifier is built to classify those extracted features Hence a more robuste combination is used. Also, you can use advanced Word2Vec, BERT words embedding methods for your models to understand context and semantic relationships.

This study aims not only to enhance the quality of sentiment classification but also guarantee that such models are scalable and relevant in practical settings. The results enhance the conventional unintended compliance with emotions, thus this paper supports the evolution of more complex and cognitive psychology forms for the corpus processing.

Background

The emergence of digital communication, in particular via social media channels, has made public sentiment analysis an important focus of NLP research. In this project, we will carry out a sentiment analysis, also referred to as opinion mining, which aims to extract subjective information from text data and identify the overall emotional tone behind the texts, for example, positivity, negativity or neutrality. Again, this analysis is not only helpful for businesses looking to capture consumer sentiment, but also for political, social, and psychological studies to examine population attitudes and behaviors.

Early sentiment analysis approaches were based on lexicons and were using shallow machine learning classifiers like Naïve Bayes, Decision Trees, and Support Vector Machines (SVM). However, these methods were fundamentally limited in their capacity to reflect the richness of human language, particularly in informal and fluid environments such as social media. These models were poor at handling context, sarcasm, slang, and domain-specific language. Enter deep learning, which cut up some big chunks of the cake. Recurrent Neural Networks (RNN) specifically Long Short Term (LSTM) networks became widely popular because they can deal with successive data and contextual relations between words. Convolutional Neural Networks (CNNs) Despite the fact that CNNs were originally created for image data processing, the exploration of CNNs applied to the text data was also the consequence of the properties of extraction of space interaction and patterns. Recently, techniques like transformers, BERT and such have propelled the field further, allowing bidirectional context understanding and transfer learning from large pre-trained language models. However, there still exist a few challenges.

Deep learning models are often data-hungry, computationally and resource-intensive, and it can lead to overfitting the particular dataset as well. In addition, models trained on one platform or domain can behave completely differently when trained on another platform due to differences in language use, tone, and structure. As a result, researchers have begun looking toward hybrid methods which effectively combine the strengths of various architectures and techniques. Having realized the merits of CNNs local feature detection, LSTMs for sequential and temporal modeling, and SVMs (or other dense classifiers) for classification, some researchers have begun to combine them into hybrid models. We start from there and implement embedding techniques Word2Vec and BERT, that converts the text into numerical formats by capturing its semantics and syntaxes. The above models are augmented by introducing user context (U), emotion analysis module and ABSA applied at sentence level to enhance detection and provide holistic view of the sentiment context.

The constant transformation of online conversations — driven by new slang, multilingual phrases and whatever else is trending — requires increasingly flexible and clever models. Propelled by preceding works in hybrid deep learning, this study aims at creating a sentiment analysis framework with a high degree of accuracy across a variety of domains and real-time settings.

Problem Statement

Social media has become a powerful place for expressing opinions, reactions, and emotions. Understanding these sentiments is crucial for businesses, policymakers, and researchers looking to analyze the public's opinion and trends. Sentiment analysis, powered by deep learning, has significantly improved in recent years, yet challenges remain—specifically in adapting to new trends, slang, and delving into topics like the COVID-19 pandemic (Alsayat, 2022). While deep learning models are highly effective, they often struggle to maintain accuracy across different datasets and real-world situations (Seo et al., 2020).

One of the most promising approaches to sentiment analysis is hybrid deep learning models. Studies show that combining Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and Support Vector Machines (SVMs) leads to better sentiment classification than using any single method alone (Li & Li, 2023). Advanced word embedding techniques, such as BERT and Word2Vec, also play a key role in improving accuracy by capturing deeper contextual meaning (Zhang et al., 2024). However, there is no single method that works perfectly in every scenario, and each approach has its limitations. Some models work well on short texts like tweets but would struggle with long reviews, while others may perform well in general sentiment classification but fail to detect sarcasm or complex emotions (Dang et al., 2020).

Another area of interest in recent research is the impact of user-specific data on sentiment prediction. Some studies suggest that analyzing a user's past comments can improve the accuracy of sentiment analysis by considering their typical tone (Li & Li, 2023). For example, a user who frequently uses sarcastic language may be better understood by a model that takes their history into account. However, other research argues that sentiment analysis models can still achieve high accuracy without user-specific data, making this an ongoing debate (Alsayat, 2022).

Recent experiments have shown that hybrid deep learning models outperform traditional methods, especially when trained on diverse datasets from social media platforms like Twitter, Yelp, and Amazon (Seo et al., 2020). For example, CNN-BiLSTM models have proven effective at detecting emotions such as joy, anger, and sadness, while aspect-based sentiment analysis techniques help identify the sentiment toward specific topics like customer service or product quality (Zhang et al., 2024). However, there are still major challenges, including handling multilingual data, identifying sarcasm—as previously mentioned—and making models efficient enough for real-time applications (Dang et al., 2020).

Despite these obstacles, sentiment analysis continues to develop, driven by the need for more adaptable and accurate models. The key to future improvements lies in developing more flexibility and understanding the bigger picture in deep learning techniques.

By refining hybrid models, improving word embeddings, and striking a balance between complexity and efficiency, researchers can create more powerful sentiment analysis tools that work across different domains and real-world applications (Li & Li, 2023).

Research Objectives

1) **Improve Sentiment Analysis**

This research aims to enhance sentiment analysis by improving its accuracy, adaptability, and efficiency. It focuses on better handling the diverse and complex language patterns found in social media, which is essential for understanding public opinion and detecting emotions in online conversations.

2) **Incorporate Hybrid Deep Learning Models**

The goal is to improve sentiment classification by combining word embedding techniques with LSTM networks. This approach focuses on improving how sentiment analysis handles context, especially for trending topics like the COVID-19 pandemic. The research will test this hybrid model on popular datasets like Twitter, Amazon, and Yelp to boost accuracy.

3) **Develop Hybrid Frameworks for Sentiment Analysis**

By integrating models like SVM, CNN, and LSTM, this research aims to create more reliable sentiment analysis tools. The focus is on evaluating how well these hybrid models perform across different datasets, ensuring they can handle various domains and data structures effectively.

4) **Enhance Emotion Recognition**

Another objective is to improve emotion detection within sentiment analysis, focusing on emotions such as joy, anger, and sadness. By using CNN-BiLSTM architectures and advanced feature extraction techniques like LDA, this research will enhance the ability to identify specific emotional expressions in social media content.

5) **Refine Aspect-Based Sentiment Analysis (ABSA)**

This research will explore how incorporating user behavior data can improve sentiment analysis, allowing for more detailed and accurate sentiment classification related to specific aspects like product reviews or customer service feedback.

6) **Develop Real-Time, Efficient Sentiment Analysis Models**

The final objective is to develop sentiment analysis models that are both lightweight and efficient, making it possible to handle large-scale social media data in real-time. This involves optimizing hybrid models to balance accuracy with computational efficiency, ensuring practical use in fast-paced online environments.

Research Scope

Analysis of Current Limitations in Sentiment Analysis:

Focusing on the key challenges that hinder accuracy and adaptability in sentiment analysis, especially within the context of fast-evolving social media language. This includes issues such as sarcasm detection, domain transferability, emotion recognition, and the handling of multilingual content.

Evaluation of Existing Models and Techniques:

Investigating current sentiment analysis approaches, particularly deep learning-based models such as LSTM, CNN, SVM, and embedding methods like BERT and Word2Vec. The evaluation will highlight the performance, strengths, and weaknesses of these models across various datasets and social media platforms.

Development of a Hybrid Deep Learning Framework:

Designing a hybrid sentiment analysis model that integrates multiple deep learning architectures and advanced feature extraction techniques. The goal is to improve classification accuracy, enhance emotion and aspect detection, and maintain efficiency in real-time processing scenarios.

Validation Across social media Domains:

Testing the proposed framework using diverse datasets from platforms like Twitter, Yelp, and Amazon. This phase also includes comparing results across short and long text formats, different topics, and user-specific sentiment trends to evaluate generalizability and real-world performance.

Literature Review

Sentiment analysis has progressed from rule-based and traditional machine learning methods to advanced deep learning approaches. Early models like Naïve Bayes and SVM struggled with the informal and complex language of social media. Deep learning models such as LSTM and CNN improved performance by capturing sequential and local patterns, though each has limitations.

The use of word embeddings like Word2Vec and GloVe enhanced semantic representation, while BERT introduced contextual understanding through bidirectional encoding. Hybrid and ensemble models—such as CNN-BiLSTM (Seo et al., 2020) and user-aware models (Li & Li, 2023)—have shown improved accuracy and robustness.

Alsayat (2022) proposed an ensemble model tailored for social media sentiment analysis, while Zhang et al. (2024) and Dang et al. (2020) highlighted the effectiveness of deep learning in aspect-based and comparative sentiment tasks. Despite progress, challenges remain in handling sarcasm, multilingual data, and domain adaptation, motivating continued innovation in hybrid frameworks.

Research Methodology

The methodology is divided into the following phases:

1. Data Collection

Datasets will be sourced from widely used social media and review platforms, including:

- Twitter (short informal texts)
- Yelp (review-based emotional expression)
- Amazon (aspect-based product reviews)

Each dataset will be preprocessed for noise removal, tokenization, and normalization.

2. Model Architecture

A hybrid architecture will be developed, consisting of:

- CNN layers for local feature extraction
- BiLSTM layers for sequential pattern learning
- Fully connected layers followed by an SVM classifier

Word embeddings will be generated using both Word2Vec and BERT to evaluate the effect of contextual representations.

3. Training and Validation

The model will be trained using a stratified split of the datasets (e.g., 80% training, 20% testing). Performance metrics such as accuracy, precision, recall, and F1-score will be used to evaluate the models. Cross-validation will be applied to ensure robustness.

4. Emotion and Aspect Analysis

Additional modules will be included to classify emotional states (e.g., joy, anger, sadness) and perform aspect-based sentiment analysis using topic modeling techniques like Latent Dirichlet Allocation (LDA).

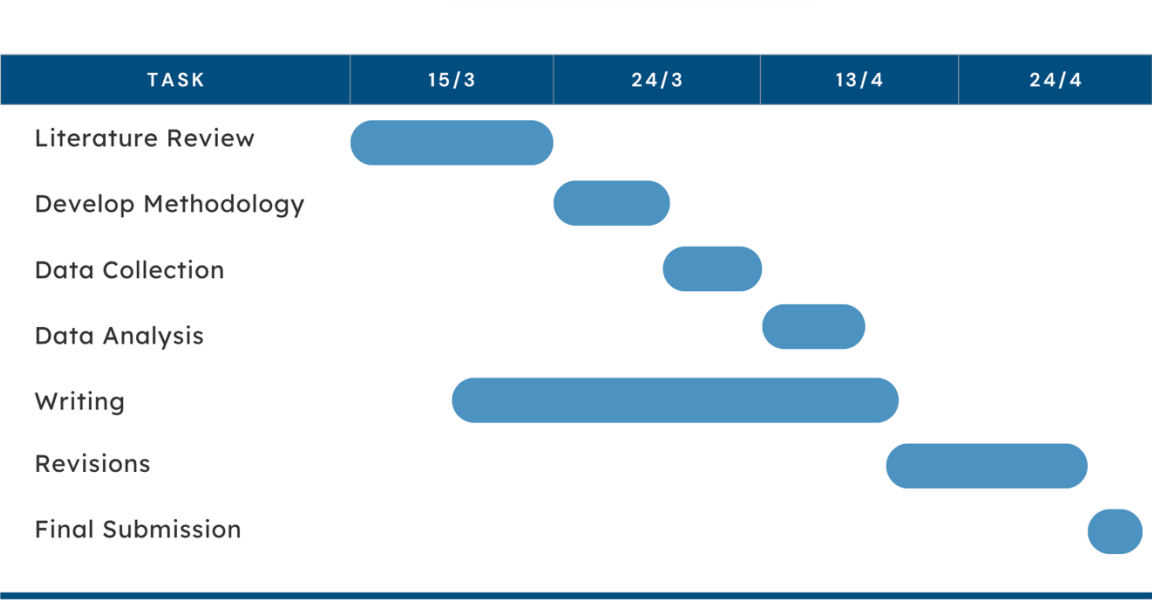
5. Performance Benchmarking

The proposed model will be compared against existing baselines including standalone LSTM, CNN, and SVM models, as well as transformer-based models like BERT fine-tuned for sentiment classification.

Expected Outcomes

- A robust hybrid model that combines CNN, BiLSTM, and SVM will outperform individual models in sentiment classification tasks.
- Integrate contextual embeddings (BERT) will improve sentiment detection in complex texts and trending topics such as COVID-19.
- The hybrid model will demonstrate improved performance in identifying emotions and aspect-based sentiments across diverse datasets.
- The resulting framework will be computationally efficient enough for near real-time analysis on large-scale social media streams.
- This research will provide deep insights into the trade-offs between model complexity, accuracy, and generalizability across domains.

Gantt Chart



References

- Alsayat, Ahmed. "Improving Sentiment Analysis for Social Media Applications Using an Ensemble Deep Learning Language Model." *Arabian Journal for Science and Engineering*, vol. 47, no. 6, 2022, pp. 2499-2511. <https://doi.org/10.1007/s13369-021-06227-w>.
- Seo, Seungwan, et al. "Comparative Study of Deep Learning-Based Sentiment Classification." *IEEE Access*, vol. 8, 2020, pp. 6861-6875. <https://doi.org/10.1109/ACCESS.2019.2963426>.
- Li, C., & Li, F. (2023). Emotion recognition of social media users based on deep learning. *PeerJ Computer Science*, 9, e1414. [DOI: 10.7717/peerj-cs.1414](https://doi.org/10.7717/peerj-cs.1414)
- Zhang, M., Chai, J., Cao, J., Ji, J., & Yi, T. (2024). Aspect-Level Sentiment Analysis Based on Deep Learning. *CMC*, 78(3), 3744-3762. <https://doi.org/10.32604/cmc.2024.048486>.
- Dang, Cach N., María N. Moreno-García, and Fernando De la Prieta. "Sentiment Analysis Based on Deep Learning: A Comparative Study." *Electronics*, vol. 9, no. 3, 2020, p. 483. MDPI, <https://doi.org/10.3390/electronics9030483>.