DOI: 10.21917/ijsc.2025.0538

SARCASTIC NEWS AND POSTS DETECTION USING HIDDEN MARKOV MODEL AND BERT

Saberi Goswami, Supratim Bhattacharya and Jayanta Poray
Department of Computer Science and Engineering, Techno India University, India

Abstract

Sentiment analysis has been quite a critical task in natural language processing or NLP which is used in the classification of opinions or sentiments in texts as positive, neutral or negative. But there are few challenges in understanding sarcastic, ambiguous and implicit texts, which usually limit the performance of traditional models like Naïve Bayes, SVM, Machine Learning, LSTM. Here we have taken social media and news into consideration, to show how different news or posts on twitter on Facebook can be analysed to extract information about the sentiments or opinion of the user even if it is sarcastic or ambiguous in nature, which can be a very challenging task. This paper proposes a novel framework using Hidden Markov Model with Contextual Embedding Layers using BERT (HMM-BERT), to address these said challenges. The unique combination of deep contextual understanding provided by pre-trained contextual embeddings like BERT with the sequential pattern learning of Hidden Markov Models enhances sentiment prediction by integrating sarcasm detection and word sense disambiguation techniques to manage complex contextual relationships within texts, and achieves an accuracy of 89-95% which is almost 10-20% better than the traditional and base models Thus, this method marks a considerable improvement in the sentiment analysis of ambiguous and sarcastic texts.

Keywords:

Hidden Markov Model, BERT, Contextual Embeddings, Sentiment Analysis, Sarcasm Detection

1. INTRODUCTION

In this paper we define ways to analyse social media content which is usually dynamic. Our main aim is to encounter and solve the challenges related to these social media contents, which include micro posts which are typically ambiguous, containing unusual capitalization, grammar and language variations, and include emoticons, abbreviations and hashtags [1]. Some posts also show sarcasm as well as irony, which are not only challenging for machines to understand, but also for humans. A good understanding of the contexts surrounding the situation, subject or individuals involved in any sarcastic argument is essential [2].

In our proposed model illustrated in Fig.1, Transformer models such as DistilBERT or BERT (Bidirectional Encoder Representations from Transformers), have been used to obtain with contextual embeddings especially in order to improve NLP tasks by considering both the meaning and context of texts.

But to increase the accuracy of the analysis, consideration of sequential dependencies in sentiment transitions are also required, especially for longer text sequences, which has been achieved by using HMM.

Therefore, this paper proposes a novel hybrid model, the Hidden Markov Model with Contextual Embedding Layers using BERT (HMM-BERT), which fuses sequential learning ability of the HMM model. with the powerful contextual learning of the BERT model, enhanced with sarcasm detection and word sense disambiguation mechanisms, to ensure enhanced sentiment analysis to solve the above discussed challenges.

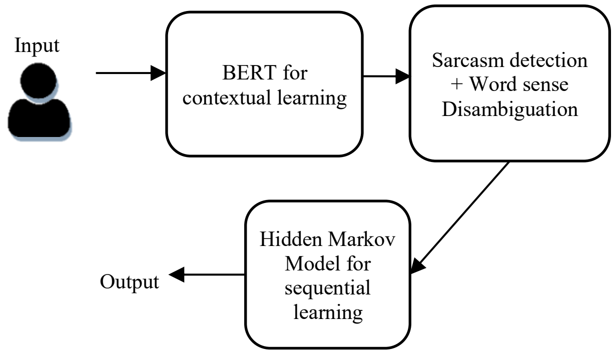

Fig.1. Proposed HMM-BERT Model Flowchart

Our contribution in this research work are as follows:

- Propose a novel hybrid HMM-BERT Model, that integrates Hidden Markov Models (HMMs) with Contextual Embedding Layers of BERT, to combine the advantages of pre-trained embeddings like BERT, for the enhancement of sentiment analysis performance.
- Integrate sequential learning, to benefit from the sequential pattern learning capabilities of HMMs while utilizing the deep contextual understanding provided by contextual embeddings.
- Address the challenges involved in Sentiment Analysis, like Sarcasm and Ambiguity, using the proposed HMM-BERT model, for improving overall accuracy.
- Therefore, a novel HMM-BERT model has been proposed to classify sarcastic and ambiguous news, comments, and posts in social media, which also aids to detect vague and harmful news and posts.
- Performance of the proposed model have been tested using three datasets from Reddit, News Headlines, Twitter and have been found to outperform other models with accuracy levels of 89-94%.

2. LITERATURE REVIEW

Diana et al. [1], [2] discusses the challenges in detecting sarcastic texts and finds the polarity of the sentiment being expressed. It also takes Hashtags into consideration to extract useful words which describes the actual meaning of the text.

Detection of sarcastic phrases, questions, swear words, conditional statements and the entity-centric approaches by Diana Maynard, et al. [4] aims to analyse opinions on definite issues and uses linguistic relations.

Diana Maynard et al. [5] look at opinions about individual entities. Here, the opinion extraction and classification is inspired by the work of Taboada et al. [6], except that it has to deal with linguistic issues.

Social Media posts and comments have been used for the prediction of social trends [7], and also for the product sales [8]. Adam et al. [9] developed a supervised machine-learning system that uses language data to classify text by rating it as either good or bad, or on a scale from 1 to 5 stars.

In another work, by Thelwall, [10], it was observed that prevalent events are related to increased number of negative opinions, by utilizing Twitter records. Kramer et al, in a research work [11], discovered that the Facebook status updates depend on the changing events.

Balog et al. utilize LiveJournal mood labels for the detection of events based on temporal data [12]. Likewise, Nguyen et al. also identified the positive and negative phases over a given time period, known as macro events, as well as short term fluctuations or local bursts or micro events [13].

Akcora et al. [14], also aimed to detect shifts in user opinions by monitoring the frequency of specific words used by the users.

Leonard DAvolio Dina Demner–Fushman Wendy W. Chapman [15] finds ways to detect Nosocomial Infections, and identifies symptoms of lung cancer, by finding all related phrases in ED Report and mapping the phrases to the standard features.

Diana Maynard et al. [4] [1] [5] [2] have studied the challenges that is ususally associalted with analyzing Social Media posts and comments. They have used NLP on Social Media texts for opinion mining and sentiment analysis, including sarcastic texts, swear words, emoticons and hashtags. Based on their work, we analyze such texts and posts and try to separate them into positive, negative and neutral emotions, and categorized them into different emotions like excited, happy, sad, depressed, frustrated, angry and pleasant.

Hidden Markov Model has been used earlier for sentiment analysis by Perikos et al. [15], where the effectiveness of Hidden Markov Models (HMM) in analyzing public sentiment, emotions, and opinions in social media text, has been evaluated, to demonstrate how HMMs outperform traditional machine learning methods by leveraging the sequential nature of textual data.

Table.1. Literature Study

Approach	Contributions	Results
Diana et al. (2014)	Detection of swear words, sarcasm, questions.	Emphasized the use of linguistic relations to focus opinions on specific topics.
Thelwall et al.(2011)	Social media posts used for predicting social trends.	Identified correlations between popular events and increased negative opinions.
Irsalinda et al. (2021)	Hidden Markov Models (HMM) with Viterbi algorithm to analyze sentiment in social media text	Forecasts candidate sentiment accurately, aiding political parties in refining campaign strategies.
Najafabadi (2024)	Hybrid sentiment analysis technique combining CNN and HMM.	Enhance text classification and sentiment detection
Zhao and Chen (2024)	HMMs to transform investor comments into sentiment scores.	Compares sentiment factors with technical analysis based on SSE 50 index data HMM's strong predictive power for stock returns

Irsalinda et al. [16] applies Hidden Markov Models (HMM) with the Viterbi algorithm to analyze sentiment in social media text, focusing on predicting sentiment during the 2015 Surabaya elections based on Twitter data. The model forecasts candidate sentiment accurately, aiding political parties in refining campaign strategies. Najafabadi [17] introduces a hybrid sentiment analysis technique combining Convolutional Neural Networks (CNN) and Hidden Markov Models (HMM) to enhance text classification and sentiment detection, demonstrating superior performance on benchmark datasets. Zhao and Chen [18] analyzes investor sentiment on Chinese financial social media by applying Hidden Markov Models (HMM) to transform investor comments into sentiment scores, comparing the resulting sentiment factor with technical analysis using SSE 50 index data, showing HMM's predictive strength for stock returns. A summary of a few relevant literatures study is given in Table.1.

3. PROPOSED HMM-BERT MODEL

Initial sentiment analysis models used machine learning algorithms, whereas in recent times, deep learning models have been employed. However, these models cannot properly understand context-dependent sentiment shifts, like in sarcastic or ambiguous language. Therefore, we propose a model to eradicate these limitations. The flow diagram of the model is explained using Fig.2. The HMM-BERT model addresses the challenges by integrating three key components:

3.1 CONTEXTUAL EMBEDDINGS LAYER (BERT)

This layer is used to produce contextualized word embeddings, in order to capture the semantic meaning and context of each word or phrase. It provides word embeddings which varies with the context. However, these embeddings are not enough to capture the sequential nature of sentiment transitions across longer texts.

3.2 HIDDEN MARKOV MODEL LAYER

The contextual embeddings are passed to the HMM layer for sequential sentiment transitions, to predict the most probable sentiment from a text. HMMs can be successfully used for sequential data analysis, like part-of-speech tagging, speech recognition, and time-series forecasting. However, HMMs normally use simple features like token frequencies which are not sufficient to capture the semantic meaning of the texts and contextual data required in robust sentiment classification.

3.3 SARCASM DETECTION AND WORD SENSE DISAMBIGUATION MODULES

Sarcasm is one of the biggest tasks in sentiment analysis, where the literal meaning given in a phrase is different from the intended sentiment. Word sense disambiguation (WSD) is also required for the correct analysis of words with multiple meanings, depending on the context. These modules help to identify sarcastic expressions and resolve ambiguities in meanings.

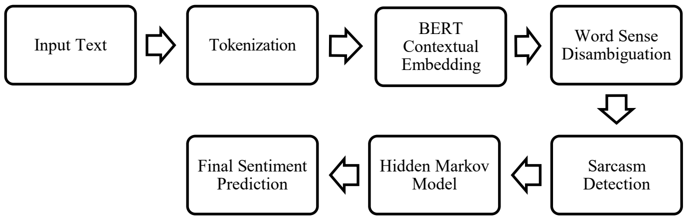

Fig.2. Proposed HMM-BERT Model Flow Diagram

4. METHODOLOGY

The methodology of the proposed model, shown in Fig.5, has been explained step by step in more detail with some examples, especially how the model handles sarcasm and ambiguous phrases and word meanings.

4.1 MODEL ARCHITECTURE

4.1.1 Input Processing:

The input text is tokenized into individual words using a tokenizer (BERT tokenizer). These tokens are delivered to the BERT layer to produce contextualized embeddings. Each of the tokens will be processed by the BERT embedding layer to generate contextualized embeddings, which are mainly based on the surrounding phrases or texts. The steps are illustrated in Fig.3.

4.1.2 Contextual Embedding Layer:

BERT-based embedding layer converts the tokenized words into contextualized embeddings, which reflect the semantics and context of the words. It identifies that a sentence may convey sarcasm, even though it uses positive words.

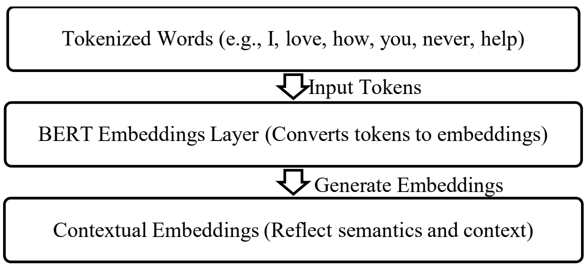

Fig.3. Tokenization and Contextual Embedding Layer

4.1.3 Word Sense Disambiguation (WSD) Module:

A WSD module is used to interpret the contextual embeddings correctly to resolve ambiguities in meaning. The WSD module identifies the word's meaning based on the surroundings as "cold" in the emotional sense.

4.1.4 Sarcasm Detection Module:

A sarcasm detection module uses auxiliary classifiers to detect sarcastic expressions. The incongruity between the positive word "great" and a negative expression implies sarcasm, which indicates accurate sentiment detection. The flow of the above modules is given in Fig.4.

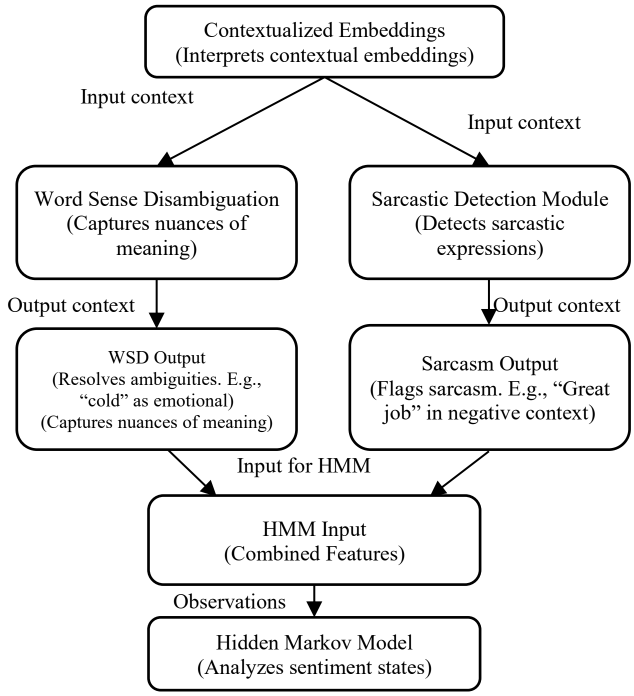

Fig.4. WSD and Sarcasm Detection modules

4.1.5 Hidden Markov Model (HMM) Layer:

- The contextualized embeddings obtained in the above layer are the inputs (observations) to the HMM layer. The hidden states relate to sentiment classes: positive, negative, and neutral, as shown in Fig.5.
- The transition probabilities are used to analyse the chances of moving from one sentiment state to another, while the emission probabilities compute the possibility of creating a particular contextual embedding.
- The forward-backward algorithm (used for training) finds the probability of the sequence of sentiment states, and the Viterbi algorithm (used for prediction) helps to find the most likely sequence of sentiment states from the observations.

4.1.6 Output Layer:

The output layer generates the final sentiment prediction (positive, negative, or neutral) for each sentence or document, along with the probability distribution for all sentiment classes.

The HMM layer detects a positive sentiment in the first part *"I had a great day"* but it then changes to negative sentiment in the second part *"but the end was bad."*. Here the model uses transition probabilities to capture this sentiment change, whereas the emission probabilities help in mapping of the contextual embeddings from the BERT embedding layer to the accurate sentiment states.

- Transition probabilities A_{ij} will be defined as:

$$A_{ij}=P(s_{t+1}=j|s_t=\mathrm{i})$$

where s_t and s_{t+1} denote the sentiment states at times t and t+1, respectively.

- Emission probabilities $B_j(o_t)$ is defined as:

$$B_j(o_t)=\mathrm{P}(\mathrm{o_t}|s_t=j)$$

where o_t is the contextualized embedding at time t, and s_t is the sentiment state. The forward-backward algorithm is employed to calculate the probabilities of several sentiment transitions during training. The Viterbi algorithm is used to determine the most likely sequence of sentiment states by analysing the first part *"I had a great day"* as positive and the second part *"but the end was terrible"* as negative.

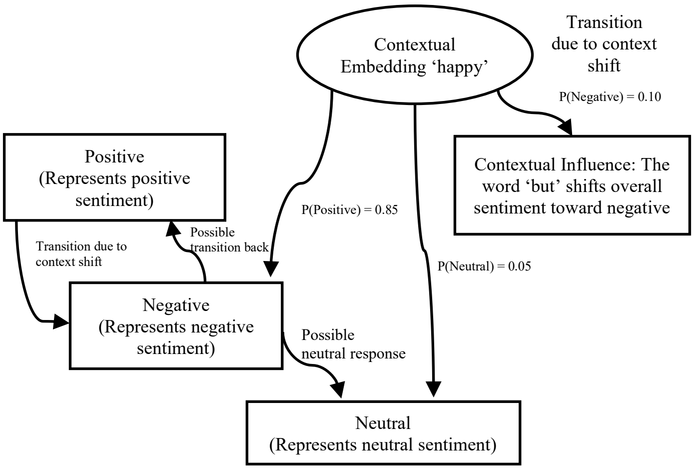

Fig.5. HMM and Output Layer

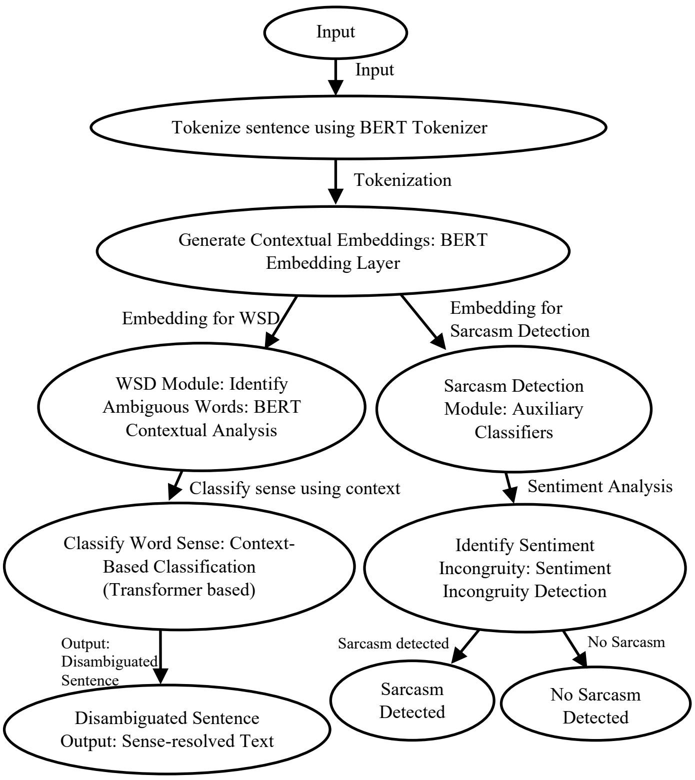

Fig.6. Proposed HMM-BERT model architecture

4.2 TRAINING AND INFERENCE

Training is performed using a Sarcastic Tweets dataset. The likelihood of sentiment sequences is calculated using the forward-backward algorithm and the transition and emission probabilities are assessed using maximum likelihood estimation (MLE). The Viterbi algorithm determines the most probable sequence of sentiment states.

4.2.1 Example:

In the sentence: *"I love how you never help.",* after all steps are processed through all the layers (tokenization, embedding, WSD, sarcasm detection, HMM), the model predicts the final sentiment as **negative** with a probability distribution as shown below:

4.2.2 Sentiment Distribution:

- Negative: 0.85
- Neutral: 0.10
- Positive: 0.05

The distribution obtained from the model shows that the sentence is negative in sentiment, despite using the positive word *"love"*.

5. RESULTS AND DISCUSSION

Experiments were conducted on three different datasets, specifically suitable for detecting and analysing sarcasm and ambiguous meanings in sentiment analysis, for the evaluation of the performance of the recommended and proposed HMM-BERT model or Hidden Markov Model with Contextual Embedding.

5.1 REDDIT SARCASM DETECTION DATASET

The Sarcasm Detection dataset is one of the most comprehensive datasets that can be used for sarcasm detection, containing 1.3 million sarcastic comments, which are extracted from various online platforms like social media web-sites, online forums, or reviews. The dataset is focused on sarcastic expressions and context-sensitive sentiment and has been used to train and test the accuracy of the HMM-BERT model.

Table.2. Section from the Reddit Dataset

Label	Comment
0	NC and NH.
0	You do know west teams play against west teams more than east teams right?
0	They were underdogs earlier today, but since Gronk's announcement this afternoon, the Vegas line has moved to patriots -1
0	This meme isn't funny none of the "new york nigga" ones are.
0	I could use one of those tools.

Fig.7. Distribution of data in the Reddit dataset

A section from the dataset is shown in Table.2. The data in this dataset contains sentences or comments column as well a label column which indicates whether the comment is non-sarcastic or sarcastic. The sarcastic and non-sarcastic comments in this dataset have been distributed equally which ensures that the training is done without any bias towards one particular sentiment over another. Equal distribution helps to make the dataset more balanced and effective in the evaluation of the capability of the model in the differentiation within sarcasm and non-sarcasm, which is fundamental for context-aware sentiment analysis.

The Fig.7 demonstrates how the data is distributed, where 50% texts are labelled as sarcastic comments, while the remaining 50% are labelled as non-sarcastic comments, whereas Fig.8 illustrates the accuracy graphs of the model on this dataset.

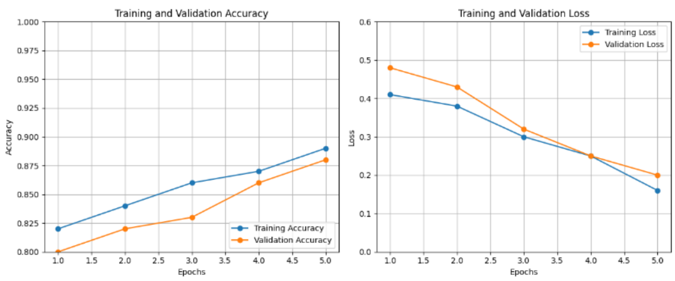

Fig.8. Training and Validation accuracy and loss graphs using proposed Model on Reddit dataset

The HMMCE model has shown a significant increase in accuracy with better performance as illustrated by the accuracy and loss graphs in Fig.8.

5.2 NEWS HEADLINES DATASET

Created for sarcasm detection, where the headlines have been gathered from two different sources [19] [20]: The Onion, for its sarcastic nature of presenting news and HuffPost which is non-sarcastic, to ensure a balanced illustration of news, comprising of around 28,000 texts.

Table.3. Section from the Sarcastic News Dataset

label	headline	article_link
1	thirtysomething scientists unveil doomsday clock of hair loss	https://www.theonion.com/thirtysomething-scientists-unveil-doomsday-clock-of-hai-1819586205
0	dem rep. totally nails why congress is falling short on gender, racial equality	https://www.huffingtonpost.com/entry/donna-edwards-inequality_us_57455f7fe4b055bb1170b207
0	eat your veggies: 9 deliciously different recipes	https://www.huffingtonpost.com/entry/eat-your-veggies-9-delici_b_8899742.html
1	inclement weather prevents liar from getting to work	https://local.theonion.com/inclement-weather-prevents-liar-from-getting-to-work-1819576031
1	mother comes pretty close to using word ‘streaming’ correctly	https://www.theonion.com/mother-comes-pretty-close-to-using-word-streaming-cor-1819575546

The Fig.9 demonstrates how the data is distributed, with 52.36% non-sarcastic and 47.64% sarcastic headlines. The proposed model gives an accuracy of around 94%, as shown by the graphs in Fig.10.

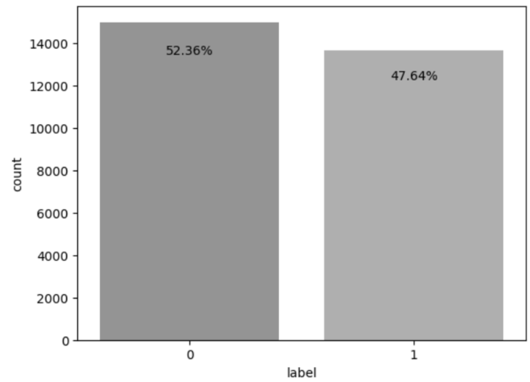

Fig.9. Distribution of data in the Sarcastic News dataset

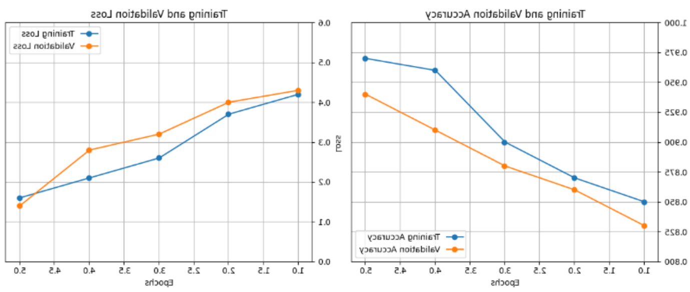

Fig.10. Training and Validation accuracy and loss graphs using proposed Model in Sarcastic News dataset

5.3 TWEETS WITH SARCASM AND IRONY:

It consists of 8000 tweets as shown in Table.4, classified as irony, sarcasm or regular.

Table.4. Section from the Sarcastic Tweets Dataset

tweets	class
Be aware dirty step to get money #staylight #staywhite #sarcastic #moralneeded @â€¦ https://t.co/Oj6BdyX3WG	0
#sarcasm for #people who don’t understand #diy #artattack http://t.co/rtyYmuDVUS	0
@IminworkJeremy @medsingle #DailyMail readers being sensible as always #shocker #sarcastic #dailyfail #inHuntspocket #theyhatethenhs	0
@wilw Why do I get the feeling you like games? #sarcasm	0
-@TeacherArthurG @rweingarten You probably just missed the text. #sarcastic	0

The Fig.11 demonstrates how the data is distributed in training set and test set, whereas the graphs illustrated in Fig.11 show that the proposed model gives an accuracy of around 94%.

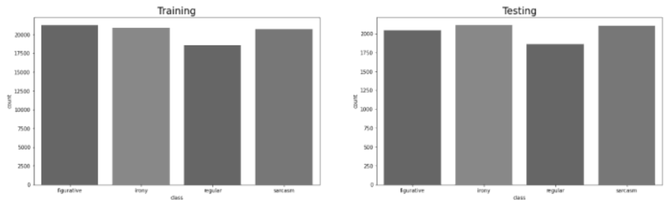

Fig.11.Distribution of data in the Sarcastic Tweets dataset

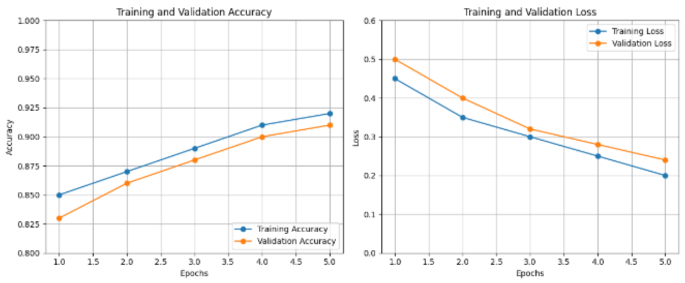

Fig.12. Training and Validation accuracy and loss graphs using proposed Model in Sarcastic Tweets dataset

The HMM-BERT model’s performance has been compared to those of some traditional sentiment analysis models, like the probabilistic model “NaiveBayes”, supervised learning model “SVM”, and a recurrent neural network “LSTM”. These models represent basic probabilistic methods (Naive Bayes) as well as sophisticated deep learning techniques (LSTM).

Table.5. Comparison Chart of SA Models with different datasets

Model	Reddit Dataset	News Dataset	Twitter Dataset
Naive Bayes	75.4	70.1	69.0
SVM	82.3	77.0	88.2
LSTM	70.8	75.5	87.1
CNN	74.0	79.0	90.0
BERT	77.3	92.0	82.6
HMM-BERT(Proposed)	**89.0**	**94.2**	**91.1**

The HMM-BERT model has shown better performance compared to these traditional models, especially because of its most focal capability to leverage contextual embeddings from BERT, combined with the Hidden Markov Model (HMM) framework, and achieved improved accuracy on sarcasm and context-sensitive sentiment detection, as presented in the Table.5.

From the above table, it is evident that the HMM-BERT model has outperformed the other models by achieving an accuracy of 89%, 94% and 91% on the above datasets respectively, especially in handling sarcasm. The Fig.12 illustrates the performance of the proposed model on the three datasets used in the experiments.

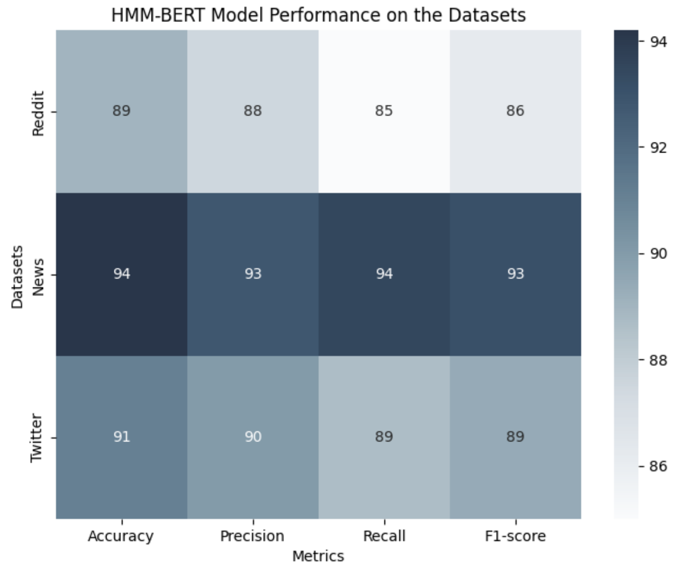

Fig.13. Heat map to show performance of proposed Model on the three datasets

The contextual embeddings were quite crucial to understand the subtleties of language more efficiently, especially for the detection of sarcastic texts, where literal meanings are not same as the intended sentiment. The HMM-BERT was also able to detect sentiment changes in a comment or phrase, which is mainly useful in texts with phrases or words having contradictory meanings within the same sentence.

6. CONCLUSION AND FUTURE WORK

The HMM-BERT model proposed in this research paper presents a novel solution for sentiment analysis by mixing the sequential learning ability of Hidden Markov Model with the deep contextual understanding using BERT embeddings. This combination allows the model to:

- Understand the semantic context of texts more efficiently, which is vital for the detection of sarcasm.
- Analyse the changes of different sentiment states (positive, negative, neutral) over time in context-sensitive sentiment detection., which result in the improvement of accuracy

The model resolves the discussed challenges of sarcasm and ambiguity. In future we can enhance the efficiency of the suggested model by refining the sarcasm analysis module and investigating more contextual embedding models. The model could be enhanced for fake news detection, which would help social media platforms, news agencies, and regulatory bodies to filter out false content, to ensure reliable information dissemination and mitigating harm caused by misinformation, or personalized sentiment analysis, in which user-specific patterns in emotional expression may be considered, which could be useful mainly in mental health monitoring systems, where emotional dynamics play a crucial role.

REFERENCES

[1] D.G. Maynard and K. Bontcheva, "Challenges of Evaluating Sentiment Analysis Tools on Social Media", *Proceedings of the International Conference on Language Resources and Evaluation*, pp. 1142-1148, 2016.

[2] D.G. Maynard and M.A. Greenwood, "Who Cares About Sarcastic Tweets? Investigating the Impact of Sarcasm on Sentiment Analysis", *Proceedings of the International Conference on Language Resources and Evaluation*, pp. 1-6, 2014.

[3] A. Funk, Y. Li, H. Saggion, K. Bontcheva and C. Leibold, "Opinion Analysis for Business Intelligence Applications", *Proceedings of the International Workshop on Ontology-Supported Business Intelligence*, pp. 1-9, 2008.

[4] D. Maynard, K. Bontcheva and D. Rout, "Challenges in Developing Opinion Mining Tools for Social Media", *Proceedings of the International Conference on Language Resources and Evaluation*, pp. 15-22, 2012.

[5] D. Maynard, G. Gossen, A. Funk and M. Fisichella, "Should I Care About Your Opinion? Detection of Opinion Interestingness and Dynamics in Social Media", *Future Internet*, Vol. 6, No. 3, pp. 457-481, 2014.

[6] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, "Lexicon-based Methods for Sentiment Analysis", *Computational Linguistics*, Vol. 37, No. 2, pp. 267-307, 2011.

[7] B. O'Connor, R. Balasubramanyan, B. Routledge and N. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series", *Proceedings of the International Conference on Web and Social Media*, Vol. 4, No. 1, pp. 122-129, 2010.

[8] Y. Liu, X. Huang, A. An and X. Yu, "ARSA: A Sentiment-Aware Model for Predicting Sales Performance using Blogs", *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 607-614, 2007.

[9] M. Thelwall, K. Buckley and G. Paltoglou, "Sentiment in Twitter Events", *Journal of the American Society for Information Science and Technology*, Vol. 62, No. 2, pp. 406-418, 2011.

[10] A.D. Kramer, "An Unobtrusive Behavioral Model of Gross National Happiness", *Proceedings of the International Conference on Human Factors in Computing Systems*, pp. 287-290, 2010.

[11] K. Balog, G. Mishne and M. De Rijke, "Why are they Excited? Identifying and Explaining Spikes in Blog Mood Levels", *Demonstrations*, Vol. 54, pp. 207-210, 2006.

[12] T. Nguyen, D. Phung, B. Adams and S. Venkatesh, "Event Extraction using Behaviors of Sentiment Signals and Burst Structure in Social Media", *Knowledge and Information Systems*, Vol. 37, pp. 279-304, 2013.

[13] C.G. Akcora, M.A. Bayir, M. Demirbas and H. Ferhatosmanoglu, "Identifying Breakpoints in Public Opinion", *Proceedings of the First Workshop on Social Media Analytics*, pp. 62-66, 2010.

[14] I. Perikos, S. Kardakis and I. Hatzilygeroudis, "Sentiment Analysis using Novel and Interpretable Architectures of Hidden Markov Models", *Knowledge-Based Systems*, Vol. 229, pp. 1-7, 2021.

[15] I. Perikos, S. Kardakis, M. Paraskevas and I. Hatzilygeroudis, "Hidden Markov Models for Sentiment Analysis in Social Media", *Proceedings of the International Conference on Big Data, Cloud Computing, Data Science and Engineering*, pp. 130-135, 2019.

[16] Y. Zhao and D. Chen, "HMM Quantifying Investor Sentiment: Research on Sentiment Factors based on Text Analysis and Hidden Markov Model", *Proceedings of the International Conference on Digital Economy and Artificial Intelligence*, pp. 635-640, 2024.

[17] M.K. Najafabadi, "Sentiment Analysis Incorporating Convolutional Neural Network into Hidden Markov Model", *Computational Intelligence*, Vol. 40, No. 2, pp. 1-9, 2024.

[18] N. Irsalinda, H. Haswat, S. Sugiyarto and M. Fitrianawati, "Hidden Markov Model for Sentiment Analysis using Viterbi Algorithm", *EKSAKTA: Journal of Sciences and Data Analysis*, pp. 18-23, 2021.

[19] Misra Rishabh and Prahal Arora, "Sarcasm Detection using News Headlines Dataset", *AI Open*, Vol. 4, No. 1, pp. 13-18, 2023.

[20] Misra Rishabh and Jigyasa Grover, "*Sculpting Data for ML: The first act of Machine Learning*", Springer, 2021.