IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

TO PERFORM SUPERVISED CLASSIFICAION OF PRODUCT REVIEWS FOR SENTIMENT ANALSIS USING EMERGING TECHNOLOGIES

[1]Ms.J.DeviGowsalyaRenuga, MCA., B.Ed**., [2]Mrs.P.Nithya**, M.Sc., M.Phil., **[3]Mrs.M.Meena**, MCA., M.Phil.,

[1]Assistant Professor, [2]Assistant Professor, [3]Assistant Professor,

Department of Computer Science, Nadar Saraswathi College of Arts and Science, Theni, India.

Abstract: Product reviews are valuable for upcoming buyers it helping them to make decisions. Sentiment analysis recognized user views, emotions, opinions and arguments from different online products, brands. The big challenge of sentiment analysis is lack of sufficient classified data in the field of Natural Language Processing (NLP)[1]. Recently deep learning has emerged as an effective means for solving sentiment classification problems. A neural network intrinsically learns a useful representation automatically without human efforts. However, the success of deep learning highly relies on the availability of large-scale training data. It propose a novel framework, named Net Spam[2], which utilizes spam features for modeling review datasets as heterogeneous information networks to map spam detection procedure into a classification problem in such networks. As the result of this weighting step, it can use fewer features with more weighs to obtain better accuracy with less time complexity. In addition, categorizing features in four major categories helps us to understand how much each category of features is contributed to spam detection. A weighting algorithm is employed to calculate each feature importance. These weights are utilized to calculate the final labels for reviews using both unsupervised and supervised approaches.

Index Terms - Sentiment analysis, NLP, deep learning, neural networks, Net spam.

I. INRODUCTION

A big part of people rely on available content in social media in their decisions. The possibility that anybody can leave a review about products a golden opportunity for spammers to write spam reviews about products and services for different interests. Identifying these spammers and the spam content is a hot topic of research and although a considerable number of studies have been done recently toward this end, but so far the methodologies put forth still barely detect spam reviews and none of them show the importance of each extracted feature type. In this project propose a novel framework, named Net Spam[2] which utilizes spam features for modeling review data sets as heterogeneous information networks to map spam detection procedure into a classification problem in such networks. Using the importance of spam features help to obtain better results in terms of different metrics experimented on real world review datasets from Yelp and Amazon websites. The results show that Net Spam outperforms the existing methods and among four categories of features including review behavioral, user behavioral, review linguistic, user linguistic, the first type of features performs better than the other categories.

II. SENTIMENT ANALYSIS

Sentiment analysis, also known as opinion mining is a Natural Language Processing (NLP)[1] technique that involves determining the sentiment or emotional tone expressed in a piece of text. It aims to understand whether the opinion conveyed is positive, negative or neutral. This analysis is often applied to various sources, such as customer reviews, social media posts or any textual data to identify public or individual opinions about particular products.

2.1 Features of sentiment analysis

Sentiment analysis[3] involves evaluating and interpreting subjective information in text data to determine the sentiment or emotional tone expressed. Key features include:

1. **Text Preprocessing:** Cleaning and preparing text data by removing noise, stop words, and irrelevant information.
2. **Tokenization:** Breaking down sentences or phrases into individual words or tokens for analysis.
3. **Lexicon-based Analysis:** Using predefined dictionaries or lexicons to assign sentiment scores to words or phrases.
4. **Machine Learning Models:** Employing algorithms like Naive Bayes, Support Vector Machines, or deep learning models to learn and predict sentiment from labelled training data.
5. **NLP Techniques:** Leveraging Natural Language Processing techniques to understand context, relationships, and nuances in language.

6. **Feature Extraction:** Identifying relevant features from text, such as keywords, n-grams, or word embeddings, to improve model accuracy.
7. **Sentiment Polarity:** Classifying sentiments into positive, negative, or neutral categories.
8. **Aspect-based Sentiment Analysis:** Analyzing sentiments towards specific aspects or topics within a piece of text.
9. **Emotion Analysis:** Going beyond simple positive/negative classification to identify specific emotions like joy, anger, or sadness.
10. **Contextual Analysis:** Considering the context and tone of the entire document or conversation rather than individual sentences.
11. **Domain Adaptation:** Customizing sentiment analysis models to perform well in specific industries or domains.
12. **Real-time Analysis:** Providing quick and timely insights into changing sentiments in dynamic environments.
13. **Multilingual Support:** Extending analysis capabilities to different languages and accommodating diverse linguistic expressions.
14. **Evaluation Metrics:** Using metrics like accuracy, precision, recall, and F1 score to assess the performance of sentiment analysis models.
15. **Sentiment Visualization:** Presenting results in a visually interpretable format, such as charts or graphs, to aid understanding.
16. **Continuous Learning:** Adapting the model over time to evolving language trends and user feedback for improved accuracy.

These features collectively enable sentiment analysis systems to effectively analyze and interpret opinion in diverse textual data using emerging technologies.

III. DEEP LEARNING FOR SENTIMENT CLASSIFICATION

Deep learning is a subset of machine learning that involves neural networks with multiple layers. It's particularly effective for tasks like image and speech recognition. Neural networks are influenced by human brain and it contains several neurons that make an impressive network. Deep learning [4] networks are capable for providing training to supervised and unsupervised data labels.

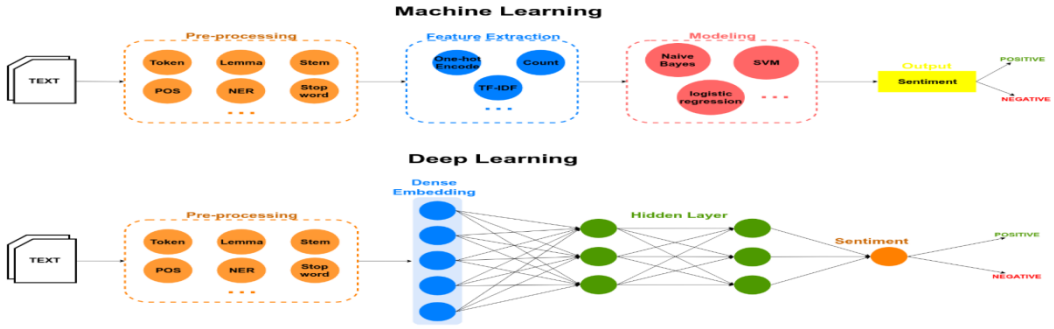

Figure 1: Machine Learning and Deep Learning for Sentiment Classification

Deep learning includes many networks they are Convolutional Neural Networks (CNN), Recurrent Neural Networks or Recursive Neural Network (RNN), Deep Belief Networks (DBN) [5]and many more.

3.1 DNN

Deep Neural Network (DNN) for sentiment analysis involves training a model to understand and classify the sentiment expressed in text data, such as reviews or social media posts. Here's a high-level overview:

1. **Data Collection:** Gather a dataset with labeled examples of text and their corresponding sentiment (positive, negative, or neutral).
2. **Data Preprocessing:** Clean and preprocess the text data, which may involve tasks like removing stop words, stemming, and tokenization.
3. **Word Embeddings:** Represent words as vectors using techniques like Word2Vec, Glove, or embeddings layers in the neural network.
4. **Model Architecture:** Design a DNN architecture suitable for sentiment analysis. Common architectures include recurrent neural networks (RNNs), long short-term memory networks (LSTMs), or more advanced models like transformers.
5. **Training:** Train the model using the labeled dataset. Adjust parameters, such as learning rate and batch size, to optimize performance.
6. **Evaluation:** Assess the model's performance on a separate test dataset to ensure it generalizes well to new, unseen data.
7. **Fine-Tuning:** Iterate on the model, adjusting hyper parameters or adding regularization techniques to improve accuracy.
8. **Deployment:** Once satisfied with the model's performance, deploy it for sentiment analysis on new text inputs.

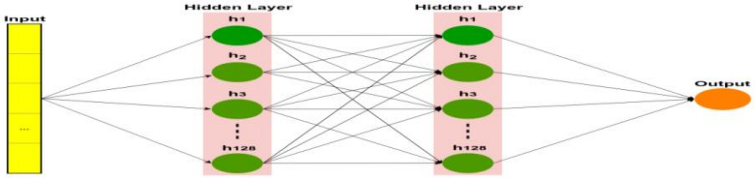

Figure 2: DNN for Sentiment Analysis

3.2 CNN

Convolutional Neural Network (CNN) in sentiment analysis is following these general steps:

1. **Data Collection:** Gather a dataset with labeled examples of text and their corresponding sentiment (positive, negative, or neutral).
2. **Data Preprocessing:** Clean and preprocess the text data, which may involve tasks like removing stop words, stemming, and tokenization.
3. **Word Embeddings:** Represent words as vectors using techniques like Word2Vec, GloVe, or embeddings layers in the neural network.
4. **Model Architecture:** Design a DNN architecture suitable for sentiment analysis. Common architectures include recurrent neural networks (RNNs), long short-term memory networks (LSTMs), or more advanced models like transformers.

5. **Training:** Train the model using the labeled dataset. Adjust parameters, such as learning rate and batch size, to optimize performance.
6. **Evaluation:** Assess the model's performance on a separate test dataset to ensure it generalizes well to new, unseen data.
7. **Fine-Tuning:** Iterate on the model, adjusting hyper parameters or adding regularization techniques to improve accuracy.
8. **Deployment:** Once satisfied with the model's performance, deploy it for sentiment analysis on new text inputs.

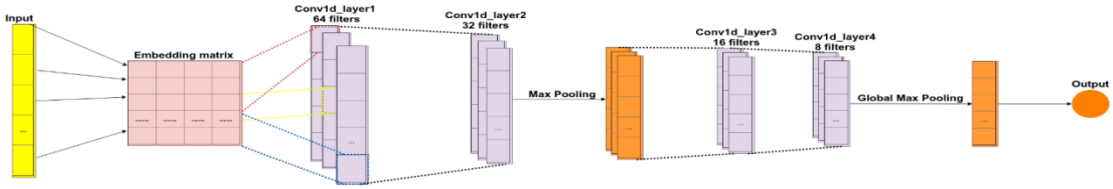

Figure 3: CNN for Sentiment Analysis

3.3 RNN

Recurrent Neural Networks (RNNs) [5] are commonly used for sentiment analysis tasks.

1. **Data Preparation:** Collect a dataset with labeled sentiment data (positive, negative, or neutral).Preprocess the text data by tokenizing and converting words to numerical representations.
2. **Embedding Layer:** Use an embedding layer to convert the numerical representations of words into dense vectors. This helps the model understand the semantic meaning of words.
3. **RNN Layers:** Stack one or more recurrent layers (such as LSTM or GRU) to capture sequential dependencies in the text.
4. **Output Layer:** Add a dense layer with soft max activation for multi-class sentiment classification or a sigmoid activation for binary sentiment classification.
5. **Training:** Split the dataset into training and testing sets. Train the model using back propagation and optimize it with an appropriate optimizer (e.g., Adam) and a suitable loss function (e.g., categorical cross entropy or binary cross entropy).
6. **Hyper parameter Tuning:** Experiment with hyper parameters like learning rate, batch size, and the number of hidden units to optimize the model's performance.
7. **Evaluation:** Evaluate the model on the testing set to assess its performance. Common metrics include accuracy, precision, recall, and F1 score.
8. **Fine-tuning:** Adjust the model architecture or hyper parameters based on evaluation results to improve performance.

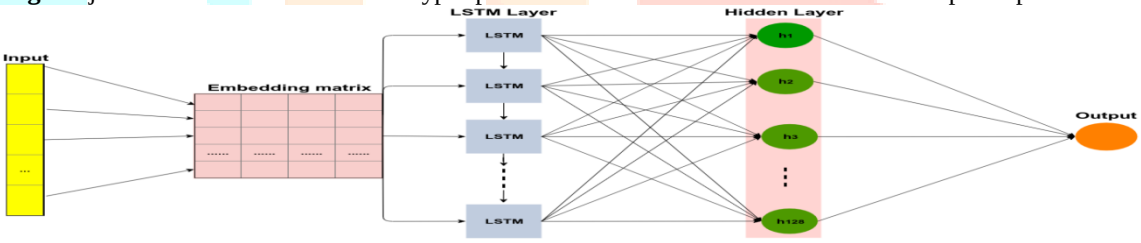

Figure 4: RNN for Sentiment Analysis

IV. CONCLUSION

In this paper described the supervised classification of product reviews for Sentiment analysis using Deep learning technologies. Deep learning is used to identify the human opinion based on the online products or brands reviews. The sentiment analysis is being implementing through deep learning techniques. Numerous problems have been resolved both fields of sentiment analysis and deep learning.

V. REFERENCES

[1] https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP
[2] https://ieeexplore.ieee.org/document/7865975
[3] https://ieeexplore.ieee.org/document/8876896/
[4] https://www.sciencedirect.com/science/article/abs/pii/S2214785322011439
[5] https://www.mdpi.com/2079-9292/9/3/483