NLP Based Review Categorization: A Survey

Amogh Karhadkar
Student of CSE Department
BMS College of Engineering
Bengaluru, India

Akanksha Laddha
Student of CSE Department
BMS College of Engineering
Bengaluru, India

Revanth N M
Student of CSE Department
BMS College of Engineering
Bengaluru, India

Antara Roy Choudhury
Assistant Professor BMS College of Engineering
Bengaluru, India

Dr. K Pa nimozhi
Assistant Professort
BMS College of Engineering
Bengaluru, India

Abstract—User reviews on social media were quickly gaining interest in the use of emotional analysis which serves as a response to government, public and private companies. Text Mining has various applications like emotional analysis, spam detection, humor detection and news sharing. The rapid growth of unorganized mountains of text data accompanied by an increase in analytics tools opens up great opportunities and challenges for mining research. Automatic labeling of text data is difficult because people tend to express ideas in complex ways. In addition, Emotional data sets are often very sensitive to the domain and difficult to perform analysis on it because emotions like feelings, attitudes and ideas are often full of sayings, onomatopoeia, synonyms, phonemes, symbols and abbreviations. In this paper, we survey the major contributions by previous researchers in this area by using Classical machine learning, LSTM (Long Short-Term Memory), CNN (Convolutional Neural Network), GRU (Gated Recurrent Unit), RNN (Recurrent Neural Network), Ontology Learning, Summarization, Hybrid Models etc. Then we propose a deep neural network model using Long-Short-Term Memory (LSTM) with word embedding features. The design includes sub-systems such as pre-processing, train-test splitting of the dataset, feature extraction using Word2Vec model and Deep Neural Network. The proposed model will be evaluated using several different metrics such as accuracy, precision, recall and F1-Score.

Keywords- **LSTM, Sentiment Analysis, Word Embeddings**

1. INTRODUCTION

In today’s digital world, people are keener in expressing themselves on different platforms and put forward their opinions, reviews, feedback and sentiments on diverse topics. Reviews of users on social media and other platforms have gained rapid interest in the usage of sentiment analysis which serves as feedback to the government, public and private companies. Review classification and categorization using sentiment analysis is an important and collaborative task for many organizations to learn about user trends and moods.

These reviews are highly influential in customers’ decision-making as well as come in handy in analyzing huge amounts of big data that is exponentially growing. With the availability of reviews on the internet in large numbers, it becomes cumbersome for customers to manually go through all the reviews and categorize them accordingly. Another challenge is the qualitative nature of the data and the number of reviews. Therefore it requires an efficient sentiment analysis system to quickly categorize the reviews and process large amounts of text without human intervention.

Existing techniques for classification, clustering, summarization and categorization include opinion mining, multi- nominal Naive Bayes classifier, classification using KNN, keyword matching, ontology-based learning, and extractive text summarization. In recent years, text classification is usually studied with machine learning models and handcrafted features that are not able to give promising results on short text classification.

Hence we propose a solution by categorizing reviews into multiple categories using a deep neural network-based model that includes LSTMs (long short term memory) and Word Embeddings. This approach overcomes the shortcoming of the above-listed techniques and approaches and is proved to be more efficient and accurate.

This paper has been organized into sections where Section 2 tackles the different approaches for sentiment analysis and their related articles. A comparison of existing methodologies is explained in Section 3. The proposed system is in section 4 and finally, we conclude this paper in Section 5.

2. RELATED WORK

Researchers have used many techniques/methodologies for user review categorization namely Classical machine learning, LSTM (Long Short-Term Memory), CNN (Convolutional Neural Network), GRU (Gated Recurrent Unit), RNN (Recurrent Neural Network), Ontology Learning, Summarization, Hybrid Models etc. Here, we explain the major work that has been done on each area with respect to review classification.

A. Classical Machine learning techniques

Several machine learning techniques such as K-means clustering, SVM (support vector machines), naive Bayes classifier, etc. have been discussed in [4], [14], [15], [16], [17], [19], [18], [23], [28], [29], [30] which provided either binary classification or grouping of similar data.

Multinomial Naive Bayes classification is extremely popular due to their simplicity and quickness. Arif Abdurrahman Farisi et al. [4] proposed Feature Selection - based on frequency or eliminating features with the lowest frequency of word. The characteristics with a tiny difference in the potential and negative values are deleted. The Multinomial Naive Bayes model is then employed and K-folds cross verification is performed to validate it. It achieves an F1-Score of 91.4 percent with preprocessing.

Shilpi Chawla et al.[14] used smartphone reviews to analyze sentiment and a machine learning approach to classify reviews. Reviews are categorized into three categories like Positive, Negative and Neutral. Document-level sentiment classification, Sentence Level sentiment classification and Feature Level sentiment classification are commonly known as Polarity Classification. The classification is done using a Naive Bayes Classifier. A total of 1000 reviews, both favorable and negative, have been gathered. When the model is tested, it turned out that the accuracy is just 40%, which is far from sufficient. As a result, the authors tried to use the Support Vector Machines (SVM) method, which produces an excellent 90% outcome.

M. Ikonomakis et al.[15] investigated various machine learning algorithms for text classification. After feature selection and transformation stage, the text is represented as a vector space document. Feature Transformation seeks to reduce the feature set size. The vocabulary is reduced using a method based on feature concurrencies. For text categorization, machine learning methods such as decision trees, naive Bayes, rule induction, neural networks, closest neighbors, and, more recently, support vector machines are employed and their performances are assessed and compared.

Harshali P. Patil[16] presented a survey of pros and cons of several methodologies for sentiment analysis. The authors discussed approaches such as Matrix Factorization, Dual Sentiment Analysis, Naive-Bayes multi-label classification algorithm, Latent Dirichlet Allocation based model, Weakly supervised joint sentiment-topic, SVM classifier, Hybrid approach(lexicon-based +M/c learning), Sentiment PLSA and other NLP techniques in the context of social data.

Shaik Daniya1 et al. [17] focused on the task of automated summarization. It makes use of an identification derived from the original text. It goes through methods like Lex Rank, LSA (Latent Semantic Analysis) and K-means Clustering. LexRank is an unsupervised graph-based method. It is used to determine the relevance of a text in a text sequence. Latent Semantic Analysis is a technique for discovering hidden semantic patterns in words and phrases. The K-means technique assigns words and terms to a grouping based entirely on the categorization algorithm.

Nirag T. Bhattet al. [18] provide a standard and generic framework for sentiment analysis techniques, which must be adhered to if success is to be achieved. The suggested system includes a few phrases that are only a rough framework and may vary depending on techniques and datasets. Data collection, data preparation, feature extraction, classifier, and result analysis are the steps involved.

The immediate goal of this research article by Hassan Raza at al. [19]is to assist and guide other researchers in applying sentiment analysis to discover high-quality research publications. Lot of preprocessing processes are employed to clean up the data. The Naive-Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbor (KNN), and Random Forest machine learning techniques are used. SVM outperforms all other models among all algorithms.

Kudakwashe Zvarevashe et al. [23] proposed emotion polarity in this paper which automatically modifies a set of sensory data to extract neutral hotel service perspectives from reviews. A comparison was made using Naive Bayes multinomial and Composite hypercubes for finding a suitable reading machine for the framework's partition method.

Tushar Ghorpade et al.[28] proposed the use of The jolly and pleasant exercise (JAPE). JAPE is a mathematical technique that preprocesses all words from given sentences and uses a Bayesian algorithm to classify them into positive or negative categories. At first, data is collected manually or with the help of a crawler. The Stanford parser is then used to determine the sentence's grammatical structure. Each word in a sentence is given a POS tag by a POS tagger. The verbs, adverbs, adjectives and nouns are then identified using domain ontology. After that, the word list is classified using the Bayesian classification method. The accuracy was found to be 96.09 percent.

Khushbu Khama et al.[29] mentions Short text classification as the process of categorizing diverse input short texts into different groups based on their content. The short text is nothing more than a short document with a limited number of words. Initially, other algorithms such as Naive Bayes, KNN, Support Vector Machine etc. were examined and it was discovered that KNN produced the best results for short texts. This document contains ways for reducing processing time and improving testing accuracy. The distance or similarity function, such as the Euclidean distance function, is used by KNN. The separation time is very long with this method, and finding the exact value of K is challenging. High values of k also lessen the effect of noise on isolation but also make the boundaries between classes slightly different. A good 'k' can be chosen using a variety of heuristic strategies. The following settings can be used to improve KNN. (1) The standard Euclidean distance is used to measure the difference or resemblance between two instances using the distance/similarity function. (2) The input parameter 'k' which stands for neighborhood size, which is arbitrarily assigned. (3) The simple voting-based class probability assumption is employed. The Euclidean range function has the flaw of being able to override other features if one input element has a particularly big distance.

The Senti-lexicon algorithm is used by Deebha Mumtaza et al.[30] to analyze the emotional content of movie reviews found on Twitter. The Senti-lexicon algorithm was invented in this study, and it was used to perform a sensory analysis of the data. In comparison to other machine learning algorithms the Senti-lexicon approach appears to be simple, adaptable and practicable. The breadth, levels, and forms of opinion mining are also discussed in the article. The following are some examples of opinion mining techniques: (1)Naive Bayes, KNN, Support Vector Machines, and other machine learning approaches (2)Lexicon-Based Method. In a lexicon-based approach, pre-built sentiment dictionaries or lexicons are employed in place of training data. It is based on the premise that the sum of the individual polarity of words determines the final orientation of a sentence. Machine learning approaches are more accurate than lexicon-based methods, but they take longer and are highly dependent on the training database's performance. A lexicon-based strategy, on the other hand, is quick, simple, and simple to set up. They operate effectively with simple data sets where the positive and negative statements are clearly separated.

B. LSTM (Long Short-Term Memory)

LSTM network is a type of recurrent neural network that can learn order dependence in sequence prediction challenges. Most of the time word embeddings are used with deep neural networks. The word embedding approach is to generate word vector representations that also reflect the semantic meaning of the words.

Abid Ishaq et al. examine opinion surveys and aggregate feedback, reviews, and comments received from various stakeholders on the university's and institution's operations. The data is processed initially, then ontology research is performed. The weight is assigned to each term and combined for each response after the synonyms are combined to produce tokens. The results show that the proposed LSTM network-based model exceeds modern techniques by providing 97% accuracy.
The goal of the work by Fie Long[12] is to use BiLSTM with Multi-Head Attention Mechanism to assess feelings in Chinese social media material. The Taobao e-commerce platform provided the data for this study. Customers rated the reviews on a scale of 1 to 5, with 1-2 indicating a negative review, 3 indicating a neutral review, and 4-5 indicating a positive review. There are approximately 19,465 comments in the dataset, with 7,854 being positive, 4,124 being neutral and 7,487 being negative. The dataset is separated into training and testing sets with an 8:2 ratio. When compared to the Embedding layer, the word2vec model improved precision, recall, f1-score, and accuracy by 0.9 percent, 0.97 percent, 0.95 percent, and 0.91 percent, respectively. When compared to Glove, the word2vec model improved precision, recall, f1-score, and accuracy by 0.82 percent, 0.79 percent, 0.78 percent, and 0.80 percent, respectively.
Baek Cheol Jang et al. [13] propose a hybrid model that combines Bi-directional LSTM, CNN and attention mechanisms, merging the benefits of each into a single model for improved accuracy. Using their pooling layers, CNN models are good at extracting high-level information, but LSTM is good at memorizing word relationships between word sequences. As a result, combining the two could improve accuracy, but the challenge was the number of parameters that needed to be taught. The BiLSTM and CNN hybrid model generated a large number of parameters. As a result, an attention mechanism was implemented at the end to reduce the number of parameters. For training and testing, the IMDB movie review dataset was used. The hybrid attention BiLSTM + CNN model beat models like Multi-Layer Perceptron (MLP), CNN, LSTM, and any other hybrid models of that type, according to the test findings. CNN had an accuracy of 0.8874, LSTM had an accuracy of 0.8940, MLP had an accuracy of 0.7129, the hybrid model had an accuracy of 0.8906, and the proposed model had an accuracy of 0.9141. The CNN model received an f1-score of 0.8875, 0.8816 for LSTM, and 0.7708 for MLP, the hybrid model received an f1-score of 0.8887, and the suggested model surpassed all of the other models with an f1-score of 0.9018
Long-short term memory (LSTM) with GloVe characteristics is investigated by Winda Kurnia Sari et al.[21]. The AGNews dataset, which contains 400,000 data samples, was employed in this study. The data is preprocessed by eliminating punctuation and tokenizing it. Then LSTM is used to classify it. The accuracy obtained is 95.17% with an average precision, recall, and F1-score of 95.
Reza Amalia Priyantina et al. [22] propose a method for determining the tone of a review based on hotel features. The names of hotel guests are pre-processed from the reviews. The Latent Dirichlet Allocation (LDA) discovers the hidden themes of a list of names first, and then Semantic Parallel divides a list of titles into five hotel sections depending on a subject caused by the LDA. The list of terms is then enlarged using the Term Frequency-Inverse Cluster Frequency (TF-ICF) approach, which counts matches. Finally, a combination of Voice embedding and temporary memory is used to separate the client's feelings (pleasure or unhappiness) (LSM). The findings indicate that the proposed method might divide the review into five hotel aspects. High phase performance is achieved by using LDA + TF-ICF 100 percent + Semantic comparison access 85 percent; emotional separation of high-quality emotional analysis performance is achieved by using Word + LSTM embedding up to 93 percent, and the luxury feature produces less intense emotions than the other features. And the findings indicate that emotions have a role.

C. CNN (Convolutional Neural Network)

The architecture of a CNN is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex

Normally, Convolutional Neural Networks (CNNs) are used for image classification tasks, however, the scientists took a risk and employed CNNs for text categorization, with promising results. Hannah Kim et al. [9] tested three datasets and found that utilizing consecutive convolutional layers is efficient for long texts and so performs exceptionally well. For the three datasets, the CNN model achieved weighted F1 scores of 80.96 percent, 81.4 percent, and 70.2 percent, respectively.

The word embedding approach generates word vector representations that also reflect the semantic meaning of the words. Xi Ouyang et al. [10] suggest a hybrid model, Word2Vec + CNN, in this study. The authors then create a sentiment analysis CNN architecture that incorporates convolutional and pooling layers. The experiment is conducted using a publicly available dataset of movie reviews with five labels: positive, slightly positive, neutral, negative, and somewhat negative. In this dataset, this model gets a testing accuracy of 45.4 percent which is not sufficient enough.

Yuling Chen et al. [11] mix Convolutional Neural Networks (CNN) with SVM for text sentiment analysis in this paper. When compared to just plain CNN models, the findings reveal that the suggested strategy actually improves text sentiment classification accuracy. When there are many anomalies in the data and a large dataset is used, typical feature extraction approaches fail. As a result, the authors emphasize the need of adopting deep learning methodologies to handle such difficulties. Although CNN may extract relevant and informative information from word vector representations, the fully connected dense layer's classification abilities for non-linear separable data are limited. SVMs may map data into high-dimensional space and then transform it into a linear separable issue using their kernel functions. As a result, these two elements are blended to produce better and more ideal results. The NLPCC2014 emotional analysis evaluation task data set was used in this study. The training data is separated into 5000 positive and 5000 negative samples and the test data is split into 1250 negative and 1250 positive samples. For positive samples, the model obtains 85.6 percent precision, 86.6 percent recall, and 86.1 percent f1-score, respectively, while for negative samples, 86.4 percent, 85.5 percent, and 86.0 percent accuracy, recall, and f1-score, respectively

D. GRU (Gated Recurrent Unit)

Gated recurrent units (GRUs) are a way to access recurrent neural networks. GRU is similar to long-term memory (LSTM) with a memory gateway, but has fewer limits than LSTM, as it has no exit gate. The GRU's performance in specific polyphonic music modeling functions, speech signal model and native language processing was found to be similar to that of LSTM. The GRUs show better performance on certain smaller and less frequent databases.

For sentiment analysis of Twitter data from US airlines, Yixin Tang et al. [8] employs a bi-directional Gated Recurrent Unit network technique. Each tweet must be classified into one of three categories: positive, neutral, or negative. The dataset used contains roughly 14,605 genuine tweets from United, Virgin America, and American Carriers, among other US airlines. The authors also used the skip-gram model to train their word vectors, starting with the GloVe model. The 3-Layer GRU has high accuracy but a poor F1-score with unbalanced data, but it has great accuracy and a high F1-score with balanced data.

Seenaiah Pedipina et al.[25] discuss the approach of employing an Artificial Neural Network with GRU for sentiment analysis. The information is made up of skewed political tweets. The proposed algorithm trained and predicted positive, negative, and neutral sentiments into three categories using 4500 tweets. The use of neural networks in sentiment analysis has been demonstrated to be more accurate at predicting sentiment. In this paper, a GRU plus sigmoid activation function artificial neural network technique was applied. The GRU layer aids in remembering the context of a series of statements that are interconnected. Web scraping is the first step, followed by pre-processing, web embedding, and categorization. GRU surpasses the competition in terms of time and memory use. The proposed model has an accuracy rate of 85%.

Ru Ni1 et al. [26] discuss the approach of employing an Artificial Neural Network with GRU for sentiment analysis. The information is made up of skewed political tweets. The proposed algorithm trained and predicted positive, negative, and neutral sentiments into three categories using 4500 tweets. The use of neural networks in sentiment analysis has been demonstrated to be more accurate at predicting sentiment. In this paper, a GRU plus sigmoid activation function artificial neural network technique was applied. The GRU layer aids in remembering the context of a series of statements that are interconnected. Web scraping is the first step, followed by pre-processing, web embedding, and categorization. GRU surpasses the competition in terms of time and memory use. The proposed model has an accuracy rate of 85%.

On an Amazon review dataset, this paper by Sharat Sachin et al. [27] implemented the baseline models for LSTM, GRU, Bi-LSTM, and Bi-GRU. A balanced dataset containing 120,000 reviews total in total, 90,000 reviews for testing, and training (30,000 reviews) have been used. The gated recurrent units perform better than the long short-term memory units if the text is short, and they are faster to train. The bidirectional gated recurrent showed an accuracy of 71.19%.

E. RNN (Recurrent Neural Network)

RNN is a kind of in-depth supervised learning algorithm. Here, neurons are connected to them over time. The idea of RNN is to recall what information was present in previous neurons so that these neurons could transmit information to them in the future for further analysis. The information holding capability of RNN is small, hence is in small text analysis.

Lilis Kurniasari et al. [6] used a deep learning approach with Recurrent Neural Networks coupled with Word Embeddings from the pre-trained Word2Vec model. They had only two classes namely positive and negative. Data was obtained by scraping the Traveloka website. The model reaches an accuracy of 91.1%. The dataset was broken down into training dataset (75%) and testing dataset (25%). The model is then trained and its accuracy approaches 92% and loss approaches 0%.

Alpna Patel et al. [7] give a thorough introduction to a variety of sentiment analysis methodologies and approaches. Positive, neutral, and negative sentiments are separated into three groups. The author used the RNN model, the LSTM model, and the CNN model to compare deep learning methodologies. The dataset came from the IMDB movie reviews. On the IMDB movie review dataset, the model obtains an accuracy of 87.42 percent, recall of 87.17 percent, precision of 87.53 percent, and an f-measure of 87.34 percent.

F. Ontology Learning

The automatic or semi-automatic creation of ontologies, which includes extracting the corresponding domain's terms and the relationships between the concepts that these terms represent and encoding them with an ontology language for easy retrieval, is referred to as ontology learning.

Deepshikha Chaturvedi et al.[1] focus on extracting features from bank reviews submitted by reviewers on the sites youthshut.com and myBankTracker.com. The Ontology-based feature of bank reviews has been released and it analyses the sentiments of financial institution reviews.
Jai Prakash Verma et al.[2] examines opinion surveys and aggregates feedback, reviews and comments received from various stakeholders on the university's and institution's operations. The data is processed initially, then ontology research is performed. The weight is assigned to each term and combined for each response after the synonyms are combined to produce tokens. This technique makes use of the Sequential Pattern Mining Framework (SPMF) and Elki's integration tool.

G. Summarization

The extractive text summarization technique is an unsupervised technique which involves pulling keyphrases from the source document and combining them to make a summary.

For evaluating and summarizing vast amount of Big Data, an extractive text summarization-based recommendation system model is used by Jai Prakash Verma et al.[3]. Cluster and Summary analysis are performed using three approaches: K-means, MiniBatch K-Means, and Graph-Based Text Summarization. An extractive text summarization system based on unsupervised learning is constructed and assessed using several techniques. MiniBatchKMeans outperforms K-Means in terms of accuracy. With recall, precision, and f-measure, Graph-Based Text Summarization improves the outcomes.

H. Hybrid Models

Jin Wang et al.[24] propose a regional CNN-LSTM model that forecasts VA ratings for documents using two components: CNN regional and LSTM. CNN regional was then used to generate text vectors for provided texts based on the vectors' names. When CNN regional and LSTM are combined, the prediction process can take into account both local (regional) information inside sentences and large distances in each phrase. Each region's proposed CNN-LSTM will release relevant pieces and learn the linguistic linkages between them to help with VA predictions.

I. Other approaches

R is a language and environment for statistical computing and graphics. R is highly extendable and offers a wide range of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, etc.) and graphical tools.

This research paper[20] by Dipak Kawade et al. is a result of studies on finding emotions and polarity of tweets over the Uri attack that was condemned by the world. Approximately around 5000 tweets were taken and were filtered out for duplicates and then the final dataset consisted of only 1788 tweets of all the 5000. Data pre-preprocessing is used in order to enhance the performance of machine learning algorithms. R was used as the language of choice to mine emotions and polarity determination.The results of the sentiment analysis were that the algorithm outputted six different emotions with their counts and polarities. The six different emotions were: anger, disgust, fear, joy, sadness, and surprise. The results show that 94.3% of people were disgusted by such an act of terrorism. The highest among all was fear with 55.59% followed by disgust and anger with 19.63% and 18.46%.

3. COMPARISON OF EXISTING SYSTEM

This section summarizes each of the techniques discussed so far along with the limitations imposed by them.

Method	Description	Characteristics	Advantages	Limitations
LSTM	Deep neural network for long texts	LSTMs can understand the context for long and short texts	Performs exceptionally well for long and short texts	They take time to train and are easily overfit
GRU	Evolved LSTM for efficient computation	GRUs are better in sense of computations	GRUs are good for mostly short texts	GRUs do not perform well for long texts
ML techniques	machine learning techniques and algorithms like k-means clustering, naive Bayes classifier, SVM etc. have been used	fast and simple to implement and scalable with large datasets	Can perform large scale analysis and be able to get more objective and accurate analysis.	techniques provide either binary classification or clustering of similar data
CNN	Neural Networks that extract local information well	Convolutional Neural Networks usually work with matrices and kernels	CNNs are good at extracting local information	CNNs to take too long to train and are overfit easily

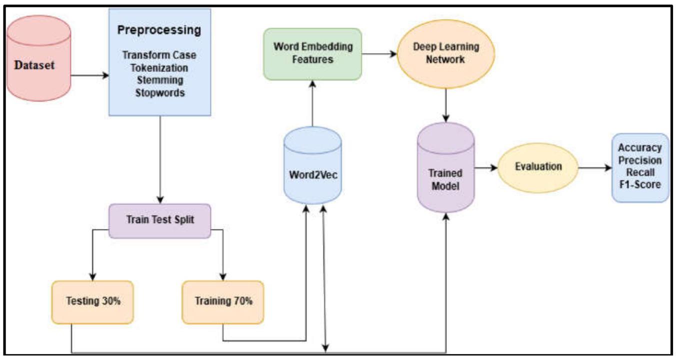
Ontology	Domain specific features text tree structure.	domain-dependent feature tree trained using large test samples.	Easy to use. Define a semantic model of the data combined with the related domain knowledge	Domain dependent trees make it hard to use the same if the domain changes.
Summarization	extracts representative text from large texts of data	domain-independent	better than supervised techniques which rely on human-generated features	execution time required for large scale review and feedback is too much

Table 3.1 Comparison of different methods

4. Proposed System

Deep neural networks (DNN) have gained a lot of popularity in NLP tasks because of the strong expressive power they possess and the lesser need to engineer features than the traditional models. . After a detailed survey of various techniques used and their drawbacks, we propose to use deep neural network-based model involving Long Short Term Memory (LSTM) along with word embedding features in our proposed model.

The RNN model could also be used but the model wouldn't perform any closer to the one proposed due to the fact that LSTMs have both long-term memory and short-term memory that they use to classify text. The LSTM can have numerous hidden layers, and as it goes through each layer, it keeps the important information while discarding the irrelevant data from each cell. It boosts performance by memorizing essential data and identifying patterns. In the traditional non-neural network classification techniques, they use multiple words as separate inputs and predict the output according to statistics and not according to meaning, which entails every single word is classified into one of the categories. While in LSTM, multiple word strings are used to find out the class to which they belong. This results in improved performance as the meaning of the complete sentence is preserved. Coupled with the Word Embedding which also maintains the semantic meaning of each and every word and the features are fed into LSTM and then the results are obtained in terms of classes.



[5] Fig 4.1. Workflow of the model

The data flow in the model is depicted in the diagram above. The high-level design that is being proposed is as follows. If the data is not already available, web scraping is used to construct the dataset. In order to optimize performance, several pre-processing techniques are used. The data is then divided into training (70 percent) and testing (30 percent) (30 percent). The features are extracted and fed to the deep neural network, which in this case is the LSTM layers, using the Word2Vec pre-trained word embedding model. The trained model is then evaluated, and if necessary, hyper parameter tweaks are made to improve the accuracy and f1-score.

5. Conclusion

In this paper, we survey the major contributions by previous researchers in the area of review categorization by using Classical machine learning, LSTM (Long Short-Term Memory), CNN (Convolutional Neural Network), GRU (Gated Recurrent Unit), RNN (Recurrent Neural Network), Ontology Learning, Summarization, Hybrid Models etc.
Then we discuss the pros and cons of each of the mentioned methods. Then we propose a deep neural network model using Long-Short-Term Memory (LSTM) with word embedding features. The design includes sub-systems such as pre-processing, train-test splitting of the dataset, feature extraction using Word2Vec model and Deep Neural Network. The proposed model will be evaluated using several different metrics such as accuracy, precision, recall and F1-Score.

References

1. [1] Deepshikha Chaturvedi, Shalu Chopra, “Customers Sentiment on Banks”, International Journal of Computer Applications (0975 – 8887) Volume 98– No.13, July 2014

2.

3. [2] Jai Prakash Verma, Bankim Patel, Atul Patel, “Web Mining: Opinion and Feedback Analysis for Educational Institutions”, International Journal of Computer Applications (0975 – 8887) Volume 84 – No 6, December 2013

4.

5. [3] Jai Prakash Verma and Atul Patel, “Evaluation of Unsupervised Learning-based Extractive Text Summarization Technique for Large Scale Review and Feedback Data”, Indian Journal of Science and Technology, Vol 10(17), DOI: 10.17485/ijst/2017/v10i17/106493, May 2017

6.

7. [4] Arif Abdurrahman Farisi , Yuliant Sibaroni, and Said Al Faraby, “Sentiment analysis on hotel reviews using Multinomial

Naïve Bayes classifier", IOP Conf. Series: Journal of Physics: Conf. Series 1192 (2019) 012024

8.

9. [5] Abid Ishaq, Muhammad Umer, Muhammad Faheem Mushtaq, Carlo Medaglia, Hafeez Ur Rehman Siddiqui, Arif Mehmood, Gyu Sang Choi, "Extensive hotel reviews classification using long short term memory", Journal of Ambient Intelligence and Humanized Computing · October 2021

10.

11. [6] Lilis Kurniasari1 and Arif Setyanto, "Sentiment Analysis using Recurrent Neural Network", IOP Conf. Series: Journal of Physics: Conf. Series 1471 (2020) 012018

12.

13. [7] Alpna Patel and Arvind Kumar Tiwari, "Sentiment Analysis by using Recurrent Neural Network", 2nd INTERNATIONAL CONFERENCE ON ADVANCED COMPUTING AND SOFTWARE ENGINEERING (ICACSE-2019)

14.

15. [8] Yixin Tang and Jiada Liu, "Gated Recurrent Units for Airline Sentiment Analysis of Twitter Data"

16.

17. [9] Hannah Kim and Young-Seob Jeong, "Sentiment Classification Using Convolutional Neural Networks"

18.

19. [10] Xi Ouyang, Pan Zhou, Cheng Hua Li, and Lijun Liu, "Sentiment Analysis Using Convolutional Neural Network", 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing

20. [11] Yuling Chen and Zhi Zhang, "Research on text sentiment analysis based on CNNs and SVM", 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)

21.

22. [12] FEI LONG 1,2, KAI ZHOU, AND WEIHUA OU, "Sentiment Analysis of Text Based on Bidirectional LSTM With Multi-Head Attention"

23.

24. [13] Baek Cheol Jang, Myeonghwi Kim, Gaspard Harerimana, Sang-ug Kang and Jong Wook Kim, "Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism"

25.

26. [14] Shilpi Chawla, Gaurav Dubey, Ajay Rana, "Product Opinion Mining Using Sentiment Analysis on Smartphone Reviews", 2017 6th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Sep. 20-22, 2017, AIIT, Amity University Uttar Pradesh, Noida, India

27.

28. [15] M. IKONOMAKIS, S. KOTSIANTIS, V. TAMPAKAS, "Text Classification Using Machine Learning Techniques", WSEAS TRANSACTIONS on COMPUTERS, Issue 8, Volume 4, August 2005, pp. 966-974

29.

30. [16] Harshali P. Patil and Dr. Mohammad Atique, "Sentiment Analysis for Social Media: A Survey"

31. [17] Shaik Daniya1, M N Chandan2, "Text Summarization and Classification using NLP", International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 08 | Aug 2021

32.

33. [18] Nirag T. Bhatt, Asst. Prof. Saket J. Swarndeep, "Sentiment Analysis using Machine Learning Technique: A Literature Survey", International Research Journal of Engineering and Technology (IRJET) Dec 2020

34.

35. [19] Hassan Raza, M. Faizan, Ahsan Hamza, Ahmed Mushtaq, Naeem Akhtar, "Scientific Text Sentiment Analysis using Machine Learning Techniques", (IJACSA) International Journal of Advanced Computer Science and Applications 2019

36.

37. [20] Dipak Kawade and Kavita Oza, "Sentiment Analysis: Machine Learning Approach", International Journal of Engineering and Technology · June 2017

38. [21] Winda Kurnia Sari, Dian Palupi Rini, Reza Firsandaya Malik, "Text Classification Using Long Short-Term Memory with GloVe Features", Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI) Vol. 5, No. 1, June 2019, pp. 85~100

39.

40. [22] Reza Amalia Priyantina, Riyanarto Sarno, "Sentiment Analysis of Hotel Reviews Using Latent Dirichlet Allocation, Semantic Similarity and LSTM", International Journal of Intelligent Engineering & systems, March 18, 2019.

41.

42. [23] Kudakwashe Zvarevashe, Oludayo O Olugbara, "A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews", 2018 Conference on Information Communications Technology and Society (ICTAS)

43.

44. [24] Jin Wang, Liang-Chih Yu, Member, IEEE, K. Robert Lai, and Xuejie Zhang, "Tree-Structured Regional CNN-LSTM Model for Dimensional Sentiment Analysis"

45.

46. [25] Seenaiah Pedipina, Sankar S and R Dhanalakshmi, "Sentiment Analysis of Twitter Political Data using GRU Neural Network", International Journal of Advanced Science and Technology Vol. 29, No. 6, (2020), pp.5307-5320

47.

48. [26] Ru Ni1, Huan Cao, "Sentiment Analysis based on GloVe and LSTM-GRU", Proceedings of the 39th Chinese Control Conference July 27-29, 2020, Shenyang, China

49.

50.

51. [27] Sharat Sachin1, Abha Tripathi, Navya Mahajan, Shivani Aggarwal, Preeti Nagrath, "Sentiment Analysis Using Gated Recurrent Neural Networks"

52.

53. [28] Tushar Ghorpade, Lata Ragha, "Featured Based Sentiment Classification for Hotel Reviews using NLP and Bayesian Classification", 2012 International Conference on Communication, Information & Computing Technology (ICCICT), Oct. 19-20

54.

55. [29] Khushbu Khamar, "Short Text Classification Using kNN Based on Distance Function", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 4, April 2013

56.

57. [30] Deebha Mumtaza, Bindiya Ahuja, "Sentiment Analysis of Movie Review Data Using Senti-Lexicon Algorithm"

58.