Survey Paper on Sentiment Analysis Using Machine Learning Techniques

Afroz Abdulhamid Bagwan
Dept. Of Computer Engineering
Modern Education Society's College Of Engineering
Pune – 01,India

Arti Vitthal Ganage
Dept. Of Computer Engineering
Modern Education Society's College Of Engineering
Pune – 01,India

Ashwini Baban Yedge
Dept. Of Computer Engineering
Modern Education Society's College Of Engineering
Pune – 01,India

Yamini Pandharinath Shewale
Dept. Of Computer Engineering
Modern Education Society's College Of Engineering
Pune – 01,India

Rubeena A. Khan
Dept. Of Computer Engineering
Modern Education Society's College Of Engineering
Pune – 01,India

Mahesh Shinde
Dept. Of Computer Engineering
Modern Education Society's College Of Engineering
Pune – 01,India

***

***Abstract*— In the recent year Sentiment analysis has gain much attention. Sentiment analysis system classifies text data into their corresponding sentiments of negative polarity, positive polarity or neutral. The aim of the paper is to enhance the precision of the model by utilizing Ensembling numerous Artificial Intelligence algorithms. The purpose of this paper is to inspect the various machine learning methods to recognise its significance and also to increase interest in this research field. This paper shows the survey of main approaches used for sentiment classification.**

Keywords— Sentiment analysis, Ensemble Method, Machine Learning, Deep Learning

I. INTRODUCTION

Sentiment analysis has become extensively spread in many areas and is helpful to obtain useful information from otherwise context-less circumstances. With the rise of personal assistants such as Google Assistant and Amazon Alexa, it has become more valuable to give relevant feedback to queries based on the user's mood. Sentiment analysis is a Artificial intelligence and natural language processing strategy used to decide if information is negative, positive or neutral. Sentiment analysis is performed on text data to understand customer needs, to help businesses monitor brand and product sentiment in customer feedback. Machine Learning Classifiers are used to performed sentiment analysis. Sentiment analysis models aim on polarity (positive, negative, neutral) yet additionally on sentiments and feelings (sad, angry, happy, etc).

Sentiment analysis is emerged as a feasible tool for any business, when it comes to recognize customer feedback. For example, to make sense of client's feedback in a client feedback survey with open-ended questions and reactions, sentiment analysis algorithms are being utilized.

Ensemble methods are learning algorithms that builds a group of classifiers and afterward classify new information focuses by taking a dominant part of their prediction. To train number of base learners as ensemble members and integrate their output into a single output that should have better execution is the basic concept of ensemble learning. The advantageous part of an ensemble learning is that the ability of boosting the weak learners to enhance the overall accuracy of the learning algorithm on training data.

II. LITRATURE REVIEW

YUAN LIN et al. [1] proposed Comparison Enhanced Bi-LSTM with Multi-Head Attention for classification task in sentiment analysis. The experiment was performed on three datasets, which are Large Movie Review Dataset, Semeval2017-task4-A English and Stanford Sentiment Treebank. In this model, bidirectional LSTM is utilized for inceptive component extraction, and Multi-Head Attention is utilized for valuable data from various angles and representation of subspaces. The aim of correlation component is to acquire the element vectors by dissecting with the marked vectors. The results show that Comparison Enhanced Bi-LSTM with Multi-Head Attention has improved execution than many existing models on three opinion investigation datasets.

Nora Al-Twairesh and Hadeel Al-Negheimish [2] proposed a feature ensemble model of deep features and surface. The model was estimated on three different datasets the SemEval 2017 Arabic tweet dataset, AraSenTi-Tweet dataset and ASTD dataset. The deep features are sentiment specific word embedding's and generic word embedding's and surface features are physically extracted features. The huge experiments was executed to test the advantages of the deep features ensemble and surface feature ensemble, cross-dataset models, embedding's size and pooling functions. The experimental outcome signify that the highest performing model is the ensemble of deep and surface features and the methodology accomplished best in class results on multiple standard datasets.

Azwa Abdul and Andrew Starkey [3] gave a Contextual Analysis (CA) strategy, a cycle that construct a connection among words and sources that is builded in a tree structure known as Hierarchical Knowledge Tree (HKT). To discover comparability and changes among train and real dataset they produce Tree Similarity Index (TSI) and Tree Differences Index (TDI) equation from tree structure. The made expectation model showed assessment mistake inside 2.75 to 3.94 and 2.30 for 3.51 for normal absolute contrasts and furthermore this technique can group the opinion words into positive and negative without having any semantic assets utilized and simultaneously changing of opinion words when another dataset is tried.

Tong Gu et al. [4] planned MBGCV model to lighten over-fitting, vanishing gradient, and to enhance the precision, MBGCV utilizes a multichannel model and combine Bidirectional Gated Recurrent Unit, Convolutional Neural Network and Variational Information Bottleneck. Chinese hotel reviews and the Chinese Product Review datasets were utilized in this experiment. The model integrate the benefits of Variational Information Bottleneck and Maxout actuation capacity to reduce the limitation , for example, over-fitting, vanishing gradient in existent sentiment analysis models. Through real survey datasets, he author implement large experiments to indicate that the proposed system can accomplish better precision and enhance the performance of text sentiment analysis.

Cem Rifki Aydin and Tunga Gungor [5] proposed a novel neural network structure for viewpoint based sentiment analysis that integrate recursive and recurrent neural models. In this study, the author perform experiment on the two datasets of the SemEval-2014 Task 4 competition which is composed of two corpora laptop and restaurant reviews. Initially the author partition every review into sub reviews that contains the sentiment data applicable to the respective viewpoint label by using electorate and reliance parsers. After training and generating, recursive neural trees constructed from the parses of the sub reviews, then their result is provided to the recurrent model. It was observed that integrating the recurrent and recursive neural networks give more extensive and a powerful model.

Yuling Chen [6] proposed a SVM integrated with Convolutional Neural Networks (CNNs) model text opinion analysis assessment task data set was used in this experiment .The author integrate the benefit of SVM and CNN, and implement a text sentiment analysis model dependent on SVM and CNN. For input pre-prepared word vector is utilized, and for automatic feature learner CNNs is used, and SVM is the final text classifier. The test outcome show that the intended model enhanced the precision of text sentiment classification adequately differentiated with conventional CNN and strengthens the efficiency of sentiment analysis build on CNNs and SVM.

Jie Zhou et al. [7] proposed a ADeCNN model for viewpoint level sentiment analysis, by integrating the attention mechanism with the deformable CNN model. The experiment was carried out on the Laptop and Restaurant datasets in SemEval 2014 Task4 and SemEval 2017 Task4. In ADeCNN, the deformable convolutional layers and bi-directional long short-term memory network, integrated with sentence-level consideration, to extract sentiment features were used by the author. And then a gated end-to-end memory network is used to join the objective into the conclusion include extraction method, so as to acquire sentiment features. Comprehensive test results conveys that ADeCNN exceed its rivals, building an impressive rise of the classification precision on all the three datasets.

Thien Khai Tran and Tuoi Thi Phan [8] proposed an effectual ensemble learning model for the problem of sentiment classification. In this experiment, three review datasets of HOTEL-Reviews, UIT-VSFC and FOODY-Reviews were used. The author integrate the different state-of-the-art deep learning models and rule-based methods to capture the contextual information from text. The integration of the attention mechanism and word embedding representation, along with specific-domain sentiment dictionaries and pre-defined rules are useful in managing with large valence-shifting cases. The model provides numerous distinctive features and gives effective outcomes than other approaches.

Andreea Salinca [9] utilizes numerous methodologies sentiment classification, using two feature extraction techniques such as one way is to create a custom word reference from the training dataset and the other way is performing lexical analysis of text reviews and four AI models Naive Bayes, Linear Support Vector Classification, Logistic Regression and Stochastic Gradient Descent. The author had given a sentiment analysis approaches to deal with business reviews classification using a huge reviews dataset given by Yelp: Yelp Challenge dataset. This paper analyse the use of multiple feature extraction methods and classifiers for classifying business text reviews. The Linear SVC and SGD classifiers results more accuracy than other algorithm.

III. Conclusion

This paper deals with a fundamental problem of sentiment analysis, sentiment polarity categorization. We studied and analyse text sentiment Analysis of comparison Enhanced Bi-LSTM with Multi-Head Attention is improved, compared with CNN and RNN classification models. RNN, CNN with attention were used as baselines. We studied ensemble learning and its perspective such as, Boosting that construct a strong classifier from the number of base learners. The availability of standard databases, different classification algorithm and the comparison of different methods are discussed. CNN and ANN are the most often used Deep Learning algorithms for solving Sentiment Classification problem and considered a reference model where many proposed algorithms are compared.

IV. References

[1] Hongfei Lin, Liang Yang, Yuan Lin, Kan Xu and Jiaping Li, "Sentiment Analysis with Comparison Enhanced Deep Neural Network".

[2] Hadeel Al-Negheimish and Nora Al-Twairesh, "Surface and Deep Features Ensemble for Sentiment Analysis of Arabic Tweets", I*EEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 496_509, June 2019.

[3] Andrew Starkey and Azwa Abdul Aziz, "Predicting supervise machine learning performance for sentiment analysis using contextual-based approaches", IEEE Access, vol. 7, pp. 22983–27992, 2019, sept 25, 2019.

[4] Guoliang Xu, Jiangtao Luo, and Tong Gu, "Sentiment Analysis via Deep Multichannel Neural Networks with Variational Information Bottleneck", IEEE Access, vol. 7, Pp-216-237, 29 July, 2020.

[5] Tunga Gungor and Cem Rifki Aydin, "Combination of Recursive and Recurrent Neural Networks for Aspect-Based Sentiment Analysis Using Inter-Aspect Relations", pp.342-221, April 17, 2020.

[6] Yuling Chen and Yuling Chen,"Research on text sentiment analysis based on CNNs and SVM", pp.232-221, 2018.

[7] Xinli Huang, Siqi Jin and Jie Zhou, "ADeCNN: An Improved Model for Aspect-Level Sentiment Analysis Based on Deformable CNN and Attention*",* International Conference on Inventive Communication and Computational Technologies, pp.216-221, 20, July 21, 2020.

[8] Tuoi Thi Phan and Thien Khai Tran, "Capturing Contextual Factors in Sentiment Classification: An Ensemble Approach", *IEEE Access*, vol. 7, pp. 114795_114807, June 2020, doi: *10.1109/ACCESS.2020.3004180.*

[9] Andreea Salinca, "Business reviews classification using sentiment analysis", 978-1-5090-0461-4/16 $31.00 © 2016 IEEE