International Journal of Electronics Engineering (ISSN: 0973-7383)
Volume 11 • Issue 1 pp. 882-888 Jan 2019-June 2019 www.csjournals.com

Classification of Reviews using Artificial Neural Network

[1]Pooja Rani, [2]Jaswinder Singh
[1]M.Tech Scholar, [2] Assistant Professor, Department of Computer Science & Engineering,
Guru Jambheshwar University of Science & Technology, Hisar, Haryana, India
poojabhatia46@gmail.com, jaswinder_singh_2k@rediffmail.com

Abstract: Sentiment classification is an important area of research. The concept of sentiment classification or sentiment analysis is considered as the computational analysis of people's opinion. People frequently express their sentiments through internet by giving ratings and reviews like positive, negative and neutral.This paper highlights certain discussions regarding the review-based classification and performance analysis of collected data through the use of Artificial Neural Network classifier.
Keywords:.Sentiment analysis,reviews, Artificial Neural Network

I. INTRODUCTION

The concept of sentiment analysis is understood by combining the terms "Senitiment" and "Analysis". The word sentiment represent feeling that can be joyful, confusing, irritating, distracting. The sentiments are the feelings based on certain attitudes and opinions rather than facts due to which sentiments are of subjective nature [16]. The sentiment implies an emotion usually motivated by opinion or perception of a person. The psychologists attempts to present multitude of emotions classified into six distinct classes: joy, love, fear, sadness, surprise and anger.The emotions based on sadness and joy are experienced on daily basis at different levels. The main concern about sentiment analysis is detecting a positive, negative and neutral response. The major significance of sentiment analysis is that every emotion is linked to human perception forming an ingrained part of humans which means that every human has the potential to generate different opinions acting as a tool for sentiment analaysis. It deals with the analysis automation of a known text regulating the distinct types of feelings conveyed. The term sentiment analysis and opinion mining can be used substitutably [15][17]. Sentiment analysis as defined as an information extraction and natural language processing task with an aim to gain the feelings of writer expressed positively or negatively based on requests, comments or questions analysing large datasets or documents. It basically intends to define writer's feeling concerning a specific area established on writer's own viewpoint [14]. It models a branch that can help in providing a judgement over distinct fields. The measurement of sentiments is a biased technique with it is really complex to achieve high accuracy of automated systems.

Approaches of Sentiment Analysis:

Sentiment analysis is an approach to Natural Language Processing that determines the polarity of emotions of users given in the form of ratings and reviews. Existing approaches of Sentiment Analysis is mainly divided into two distinct approaches [12], [7], [18]. One is the machine learning and another one is the lexicon-based approach. There is one another technique which is union of both machine learning and lexicon- based is known as the hybrid approach.

Machine Learning Approach: It relies over sentiment analysis or sentiment classification treatment as a problem of text classification. The classification of text is basically used in business that needs proper text processing. It uses records in order to train a design or model which is used in prediction of fresh records. Each and every record gets assigned to a specified class [9]. If a new form of unlabelled record is known, then the model helps in predicting its class label. Such classes may be positive, neutral, and negative. In this kind of approach, two types of its sub-approaches are distinguished based on the methods of learning. First is the supervised learning and second is the un-supervised learning.

Supervised Learning: Classification is a case of supervised learning because in supervised learning class label of each training tuple is provided. It uses classifier i.e. specifically supervised that learns from the assigned trained type of documents. The trained labelled documents contain words related to topic as key features. The review-based words express a positive or a negative review. Classification technique is used in many fields like marketing, medical diagnosis, performance prediction and credit analysis.

Unsupervised Learning: Clustering is the most common task of unsupervised learning. It is also used for representation learning, and density estimation. Unsupervised learning includes some common algorithms such as k-means Clustering and DB scan. This technique is used for document-based clustering analysis as it does not

depend over pre-defined labeled training documents. The summary to be noted is that the supervised approach learns by examples whereas the unsupervised approach learns by the observation method [4, 8].

Lexicon-Based Approach: Such type of approach depends upon sentiment lexicon, a collection of familiar and precompiled terms. This method represents another type of sentiment analysis, but in such cases, a dictionary with synonyms and antonyms of opinionated phrases and words could be used along with their respective orientation of sentiments. There are further two main sub-classes of this approach. First one is called the dictionary-based approach and other is known as the corpus-based approach.

Hybrid Approach: With the help of other machine learning methods which also combine machine learning techniques researchers have initiated to investigate hybrid approaches.

II. ARTIFICIAL NEURAL NETWORK

Connectionist systems or ANNs represents the system of computing which has structure and functions similar to the biological neurons. Information flow thorough the network affects the network structure because a neural network learns on the basis of input and output. ANN is nonlinear statistical data modeling tool in which input output relationship are modeled. The basic processing unit of ANN is artificial neurons which is simply called nodes. Neurons perform the summing and mapping function. These neurons operate in parallel and setup in regular architecture. Neurons are organized in layers and have feedback connection within the layer and for next layer. These connections have weight associated with it. Fig. 1 presents the basic block diagram of neuron. In the given figure X1, X2, ……, Xn represents the input variables. B_k is used for biasing. Wk1, Wk2, ……, Wkn shows the weights. The weight may be positive as well as negative. Σ is known as adder. It sums the input signal which has some weight associated with it. Activation function converts the value obtained from summing function into some real number output. Yk represents the output variable.

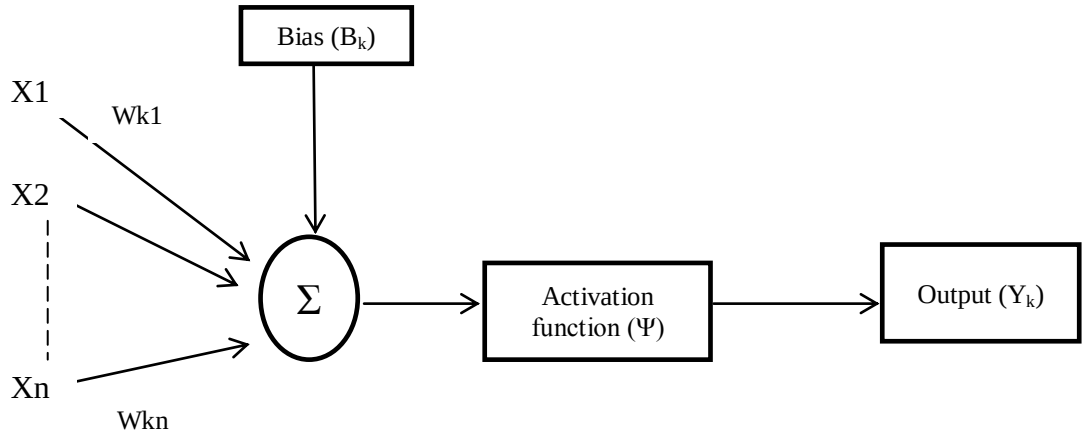

Fig.1 Block Diagram of Neuron [20]

ANN consists of three layers in which first layer is known as the input layer, one or more intermediate (hidden) layer and in the end of structure there is one output layer. The layers are connected by the link which has weight associated with it. ANN learns by updating these weights. The following Fig. 2 shows simple structure of Artificial Neural Network.

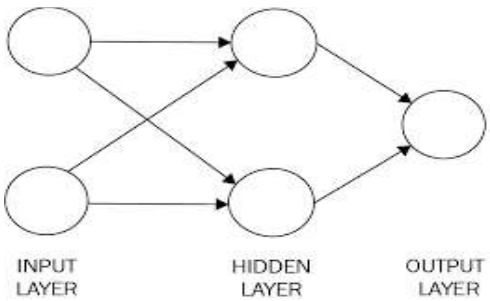

Fig. 2 ANN Structure [19]

ANN Topology

ANN topology is the settlement of network with its nodes and connection lines. On the basis of these connections it can be classified into two categories.

Feed-Forward network: These networks are non-recurrent network i.e. there is no cycle from next layer to previous layer. Information can flow only in one direction only. The non-recurrent structure is also known as Associative and can be further classified into single layer, multilayer perceptron, and Radial basis functions.

Feed-Back: These networks are recurrent network means there is feedback path between layers. The recurrent structure is also known as Auto associative. Information can flow in both directions. This makes it nonlinear dynamic system which changes constantly until network reaches as state of equilibrium.

Characteristics of Neural Networks

Artificial neural network is a non-parametric approach. Performance and accuracy depends upon the structure of network and number of inputs. It has various features [17] like:

- It includes high tolerance to noisy data
- It learns from examples.
- Neural networks are inherently parallel, so it speeds up the computation time.
- There are various techniques which are used to extract rule from trained neural network so it can also be used for classification and prediction.
- It provides collective solution which means result is summation of all the outputs which comes after processing various inputs.

III. RELATED WORK

Yelmen, *et.al* [1] presented their work by applying three distinct classification algorithms i.e. Support Vector Machine, Artificial Neural Network and Centroid Based Algorithms have been applied together with feature selection methods i.e. GI, IG, and GA on the preprocessed tweet data of followers of GSM operators for sentiment analysis. Ensemble of feature selection method and classification algorithm is used to find the succession rate of analysis. It is specially used when classifier Support Vector Machines are used with the Genetic Algorithms as a hybrid. 96.8% favorable result has been gained for the analysis of the reviews as +ve or -ve. Anna Pomeranets, *et.al* [2] proposed a study to find and observe that whether the Deep Learning process can be adapted for improving the sentiment analysis outcome in context of StockTwits. The researchers have applied various neural network models like Convolution Neural Networks, doc2vec, and long short-term memory to stock the opinions of the market posted in StockTwits dataset. The results have shown that the process of DL can be employed efficiently for classification. The model of CNN is considered to be the best model for prediction of author's sentiment in StockTwits dataset. Shahin Amiriparian, *et.al* [3] proposed a review on some of the papers in terms of research on sentiment analysis on Twitter-based data that describes certain framework and methodologies adopted and various techniques were applied, along with description of a generalized approach deployed on Python. This kind of research was evolved during last decade along with the models that reached an efficiency of about 85% to 90%. Gerasimos Spanakis, *et.al* [6] proposed and evaluated another procedure for prediction of reviews. These reviews are related to the customers posts on social pages of firms such as supermarket chains. Posts were collected from the Facebook pages of large supermarket chains. After collection of reviews there will be construction of a dataset which can be accessible for another investigation. In regulation to estimate the circulation of reviews of new posts, neural network architectures were tested using pre-trained conversation inserting. Willian Becker, *et.al* [7] proposed a translation-free language-agnostic method for the analysis of Twitter sentiment which used the methodology of Deep Convolution with embeddings of character-level for targeting the exact tweet confliction that might be written in different languages. The method proposed was more correct than various other architectures of deep neural network and substantially required limited learned specifications. The resulting model was capable to learn all the inherent features in a straight-forward manner. P. Borele, *et.al* [10] explained the sentiment analysis in detail and highlighted the particular approach, methodology and model to ascertain the sentiment orientation and its analysis. The author focused on the different approaches of machine learning in the research. The proposed work was based on ANN. The investigation suggested that the executions by using artificial neural network would result in improved analysis, after grouping the perfect of supervised and unsupervised machine learning techniques. R. M. Chandrasekaran, *et.al* [11] compared neural network build on sentiment analysis methods i.e. Back Propagation Neural Network, Probabilistic Neural Network and combination of both by applying different levels of word granularity as features for sentiment analysis. The procedure was classified by adopting distinct quality measures and the research concluded that the homogeneous grouping of the neural network technique gives finer achievement. The predicted analysis can still be improved by increasing and combining the number of different classifiers. D. A. Borikar, *et.al* [13] focused on the different approaches of machine learning and also the usage of Artificial Neural Networks in classification of sentiments. The salient features of techniques used in sentiment analysis

were also discussed. Machine learning approach-based sentiment classification includes different well-known supervised learning techniques i.e. NB, ME, SVM, and KNN K-nearest. For enhancing the achievements regarding correlation and dependencies between variables, an approach combining ANN and fuzzy logic is often used.

The related work concludes that in machine learning method, Artificial Neural Network is one of those classification techniques that is used by many authors for performance enhancement. Next section of paper describes the implementation details of classification of reviews using artificial Neural Network.

IV. IMPLEMENTATION

The information is gathered for test from reviews and put away in database for pre-handling. The steps of implementation are described in this section of paper.

Step1: Collection of the data

The facts given as input to the recommended model has been assembled from the movie reviews of IMDb (Internet Movie Database). IMDb is the most famous and authentic source for movie reviews and ratings. The information recovered from the web-based social networking is in unstructured form because of concentrated data. Some pre-handling steps are required to separate the valuable data from the dataset.

Step 2: Fetching of data

The recovered comments are put away as .csv arrangement records, and after that these documents are brought in the PyCharm instrument of Python. A large number of reviews are stored to prepare and test the datasets. Information mining calculation is utilized for preparation and experiment of the reviews.

Step 3: Pre-Handling/Processing of data

In the progression cleaning of information is finished. Data pre-handling evacuates the boisterous, excess information from the crude information and afterward makes the prepared dataset for other work. Various advances pursue to immaculate the information depicted underneath.

a) Change all the capitalized in to lowercase.
b) Discard all web informal speech from the information.
c) Eliminate all extra void areas.
d) Squeeze the duplicate words.
e) All the hash tags are evacuated yet the hash label writings are saved.

Step 4: Feature Extraction and Classifier Modeling

Feature extraction is the procedure of taking useful facts from the dataset. Extracted features are presented in the form of matrix of finite dimensions. When feature extraction and reduction is completed, these reduced features are fed to artificial neural network is known as learning. Neural network has the capability to classify the reviews. Classifier modeling is the next process that performs the classification of reviews in positive and negative.

Step 5: Optimization of result

Result optimization is the last step that is used to check the model that features are learned by training data set and no mistake rate is figured. As per the informational collection cross approval is utilized to get the exact outcome. To construct the model in Python, Artificial Neural Network (ANN) is used for the training and testing of the data.

V. RESULTS

This section deals with the result and analysis of the results obtained. The proposed approach represents the comparative analysis of classifier ANN. Experimental analysis shows the overall Accuracy, Precision, Recall and F-measure. Cross validation is used for splitting the dataset. K-fold cross validation is applied because it is easy to use and simple. Cross validation method first partitioned the dataset into k mutually exclusive subset, almost equal size. During ith iteration each subset is used as a test set and remaining subset is used as a training set. This technique makes efficient use of data and avoids the overlapping of test set.

Table 1 Accuracy of ANN

No. of Validation	ANN (Accuracy)
5-Fold	52
10-Fold	54.34

Table 1 shows the accuracy of used classifier i.e. ANN in different forms of validations. Accuracy shown by 5-Fold cross validation process is 52 and 10-Fold shows 54.34. Accuracy increases with the increase of number of validations. The overall accuracy is 62.73 obtained by the averaging all accuracy obtained from each fold. The accuracy of classifier is defined as the percentage of test tuple that the classifier correctly classifies. There can be misclassification rate which is 1-accuracy. Accuracy is also known as overall recognition rate of a classifier.

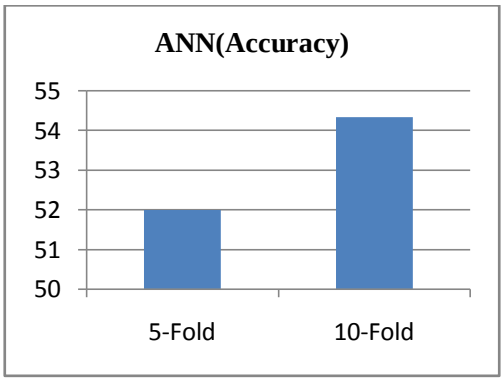

Fig. 3 Accuracy of ANN

Fig. 3 depicts the accuracy of the classifier that is ANN. The X-axis on graph shows the validation fold and Y-axis presents the values of accuracy. The algorithm represents the accuracy of used classifier in 5-Fold and 10-Fold. 52 is the minimum accuracy represented by ANN in 5-fold cross validation process.

Table 2 Precision after using ANN

No. of Validation	ANN (Precision)
5-Fold	62
10-Fold	63.23

Table 2 represents the precision after using ANN in distinct forms of cross validation folds i.e. 5-Fold and 10-Fold. Precision shown in 5-Fold validation is 62 and in 10-Fold is 63.23.

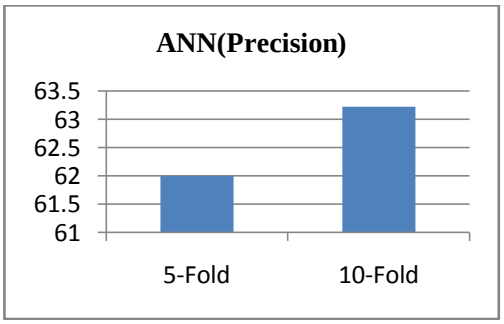

Fig. 4 Precision on different Folds

Fig. 4 depicts the Precision of the classifier that is ANN. The X-axis on graph stands for the validation fold and Y-axis stands for the values of Precision. The algorithm represents the precision in folds i.e. 5-Fold and 10-Fold of cross validation. 62 is the minimum precision represented by ANN in 5-fold cross validation process. There are various parameters which can be used for accuracy measure. Table 3 describes the different parameters and their formulae.

Table 3 Different Parameters

Accuracy	TP+TN/ALL
TP rate or Recall	TP/TP+FN
FP rate	FP/FP+TN

Specificity or TN rate	TN/TN+FP
Precision	TP/TP+FP
F-Measure or F-Score	$2.\frac{Precision\,.Recall}{Precision+Recall}$

Table 4 Parametric analysis of ANN

Classifier	Overall Accuracy	Precision	Recall	F-Measure
ANN	62.73	72.13	70.23	73.13

After experiment analysis Table 4 represents the parametric values of neural network. Value of overall accuracy is 62.73, precision is 72.13, recall is 70.23, and F-measure is 73.13.

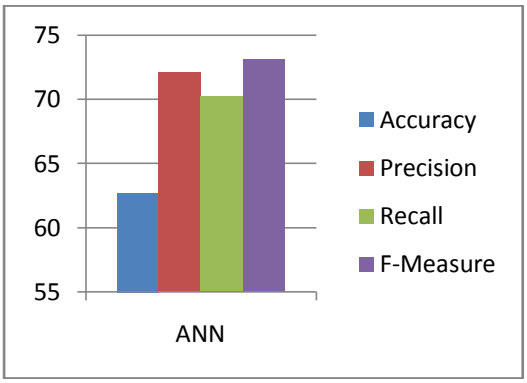

Fig. 5 Results on all Parameters

Fig. 5 shows the results on all parameters. The X-axis on graph correlates to the classifiers and Y-axis corresponds to the values of Accuracy, Precision, Recall and F-measure.

VI. CONCLUSION

Classification of reviews or comments is one of the most important filed for the research in data mining. ANN is one of classification technique used for analysis of sentiments. The experimental analysis consists of a specified or particular form of dataset and the process of learning will be done to improve the accuracy. The features obtained will be used for the classification and to improve the results of the model. The performance of ANN can be further improved by combining it with some other classifier.

VII. REFERENCES

[1] I. Yelmen, M. Zontul, O. Kaynar, and F. Sonmez, "A Novel Hybrid Approach for Sentiment Classification of Turkish Tweets for GSM Operators," *International Journal of Circuits, Systems and Signal Processing*, vol. 12, pp. 637-645, 2018.
[2] Sohangir Sahar, Dingding Wang, Anna Pomeranets, and Taghi M. Khoshgoftaar, "Big Data: Deep learning for financial sentiment analysis," *Journal of Big Data*, vol. 5, no. 1, pp. 1-25, 2018.
[3] Cummins Nicholas, Shahin Amiriparian, Sandra Ottl, Maurice Gerczuk, Maximilian Schmitt, and Björn Schuller, "Multimodal Bag-of-Words for Cross Domains Sentiment Analysis," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Calgary, Canada*, pp. 4954-4958, 2018.
[4] Brinda, H. S. Nagashree, and Madhura Prakash, "Sentiment analysis of Twitter data: A machine learning approach to analyse demonetization tweets," 2018.
[5] Desai Radhi, "Sentiment Analysis of Twitter Data: A Survey," International *Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, vol. 3, no. 1, pp. 464-470, 2018.
[6] Krebs Florian, Bruno Lubascher, Tobias Moers, Pieter Schaap, and GerasimosSpanakis, "Social Emotion Mining Techniques for Facebook Posts Reaction Prediction," *arXiv preprint arXiv: 1712.03249,* 2017.

[7] Wehrmann Joonatas, Willian Becker, Henry EL Cagnini, and Rodrigo C. Barros, "A character-based convolutional neural network for language-agnostic Twitter sentiment analysis," In *Neural Networks (IJCNN), 2017 International Joint Conference on*, pp. 2384-2391, IEEE, 2017.
[8] Roy Kaustav, Disha Kohli, Rakesh kumarKathirvel Senthil Kumar, Rupaksh Sahgal, and Wen-Bin Yu, "Sentiment Analysis of Twitter Data for Demonetization in India–A Text Mining Approach," *Issues in Information Systems*, vol. 18, no. 4, pp. 9-15, 2017.
[9] Shoeb Md, and Jawed Ahmed, "Sentiment Analysis and Classification of Tweets Using Data Mining," in *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 12, pp. 1471-1474, 2017.
[10] P. Borele and D. A. Borikar, "A Survey on Evaluating Sentiments by Using Artificial Neural Network," *International Research Journal of Engineering and Technology (IRJET)*, vol. 03, no. 02, pp. 1402-1406, 2016.
[11] G. Vinodhini and R. M. Chandrasekaran, "A comparative performance evaluation of neural network-based approach for sentiment classification of online reviews," *Journal of King Saud University – Computer and Information Science*, vol. 28, no. 1, pp. 2-12, 2016.
[12] Nakov Preslav, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov, "SemEval-2016 task 4: Sentiment analysis in Twitter," in *Proceedings of the 10th international workshop on semantic evaluation (semeval-2016)*, pp. 1-18, 2016.
[13] P. Borele and D. A. Borikar, "An Approach to Sentiment Analysis using Artificial Neural Network with Comparative Analysis of Different Techniques," *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 18, no. 2, pp. 64-69, 2016.
[14] Sahayak Varsha, Vijaya Shete, and Apashabi Pathan, "Sentiment analysis on twitter data," in *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, vol. 2, no. 1, pp. 178-183, 2015.
[15] Kouloumpis Efthymios, Theresa Wilson, and Johanna Moore, "Twitter Sentiment Analysis: The good the bad and the omg!" in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, (*Icwsm*-2011), pp. 538-541, 2011.
[16] Agarwal Apoorv, BoyiXie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau, "Sentiment Analysis of Twitter Data," in *Proceedings of the Workshop on Language in Social Media*, pp. 30-38, Association for Computational Linguistics, 2011.
[17] Suman, Jaswinder Singh, "Sentiment Analysis: A Survey," *International Journal for Research in Applied Science & Engineering Technology*, vol. 5, issue 8, pp 1957-1963, 2017.
[18] Suman Rani, Jaswinder Singh, "Sentiment Analysis of Tweets using Support Vector Machine", *International Journal of Computer Science and Mobile Applications*, vol.5, issue 10, 2017.
[19] https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_neural_networks.
[20] S.S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, 1999.