ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Sentiment Analysis in E-Commerce and Information Security

Aruna Sathish

M.Tech, School of Computing Science and Engineering, Vellore Institute of Technology, Vellore, India

ABSTRACT: This paper focuses on the process of sentiment analysis or the automated mining of the opinions and emotions from texts. With the help of Natural Language Processing(NLP) the words in the text are classified as positive or negative. This provides the most useful and fundamental way of extracting the user's preferences for various websites with respect to the products, topics, services, etc. Based on the sentiment analysis of existing data it is possible to predict the ratings of websites, analyze the brand reputation, customer satisfaction, etc. The aim of this work is to analyze the application of sentiment analysis in e-commerce as well as information security. A brief overview of the techniques and processes is given. As a further important application of opinion analysis, twitter data of users as well as hackers is analyzed to study the sentiment content and the rates of security attacks are also predicted. The application of Recurrent Neural Network in data mining and prediction is also explored with the help of Long Short Term Memory(LSTM) architecture and the results are compared with that of Naïve Bayes Classifier.

KEYWORDS: Hacking, LSTM, NLP, NLTK, Rating prediction, RNN, SentiWordNet, Sentiment Analysis, Tensorflow, Twitter

I. INTRODUCTION

Sentiment analysis also known as opinion mining refers to the use of natural language processing and computational linguistics to extract subjective information from the given data and classify opinions. It is a broader concept and many tasks are involved in it. The most important are as follows:

- **Sentiment Classification-** The process is also known as sentiment orientation or sentiment polarity. The opinion is classified into any one of the following categories- positive, negative or neutral. There are three levels of classification.
    - Document level- It considers the entire document as an opinion and classifies accordingly. With this classification, it is possible to predict whether the review expresses a positive or negative opinion.
    - Sentence level- If the given sentence is subjective, this level of classification classifies it as positive or negative. Hence, the process is more related to the task of subjectivity classification that distinguishes sentences that express factual information from sentences that express subjective opinions.
    - Aspect level- The sentiment is classified with respect to specific aspects of the entities. Therefore, this is the finest-grained and the most complex level as it is essential to extract many features and their relationships with high precision.

- **Subjectivity Classification-** It consists of evaluating whether a given sentence is subjective or not. It can be considered as the step that precedes sentiment classification. The accuracy of sentiment classification can be improved by employing a better subjectivity classification.
- **Opinion summarization-** It is focused on extracting the main features of an entity shared with one or several documents and the sentiments regarding them. [Wang et, al 2013]. The task involves single document as well as multi-document summarization. Single document summarization analyzes the changes in the sentiment orientation throughout the document. In multi-document summarization, once the features have been detected, the system must group the different sentences which express sentiments related to those features.
- **Opinion retrieval-** The documents that express opinions or views are retrieved based on the given query. In this task, the documents are ranked as per the relevance score and the opinion score with respect to the query.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Several approaches are there to perform sentiment analysis. A few of them are mentioned below.

1. **Knowledge-based approach-** Here, the sentiment is evaluated as a function of keywords and involves the construction of lexicons which indicate a positive or negative class. The sentiment values of the words in the lexicon are determined prior to the sentiment analysis. Lexicons can be created by starting with seed words and then using some linguistic heuristics to add more words to them. SentiWordNet is a lexical resource for sentiment analysis and opinion mining.
2. **Relationship-based approach-** Different relationships may exist between features and components. These are analyzed to perform sentiment analysis. The relationships may be relationships between different product features.
3. **Language-based approach-** Classification is done by building n-gram language models. A gram is token considered for training and classification. In computational linguistics and conditional probability, n-gram represents a set or sequence of chosen tokens or lexicons. The frequency of n-grams is converted to TF-IDF (term frequency- inverse document frequency). Researches have indicated that unigrams give better results for sentiment analysis.
4. **Discourse structures and semantics approach-** This approach is used in applications where the prior classification is not possible (positive and negative classes). Text is classified when it is encountered into the best category it fits. Based on the similarity of the semantics of words in it, they are tagged into classes. This is useful for documents where the overall review is expressed in the last paragraph. The sentiment of the whole review is obtained as a function of the sentiment of the different discourse relations that exist between them.

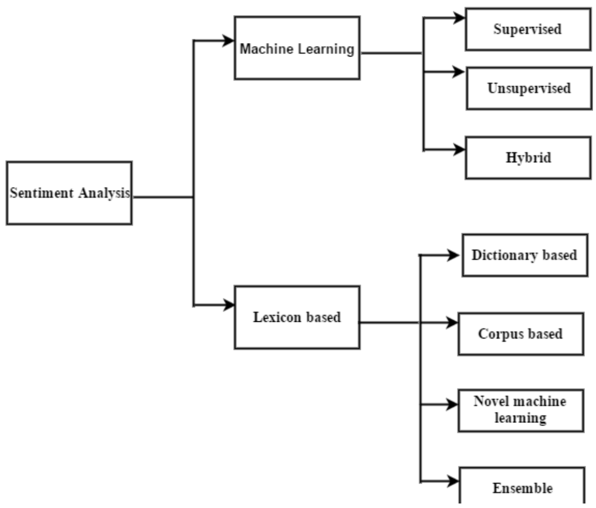

Fig 1: Sentiment analysis techniques

Techniques: Sentiment analysis can be performed using machine learning approach or lexicon based approach or a combination of both.

- **Machine learning approach-** It involves both supervised and unsupervised techniques. Supervised techniques are implemented by building a classifier. The commonly used algorithms are Naïve Bayes, Support Vector Machines(SVM) and Maximum Entropy classifier. Cui et.al have stated that SVM are more appropriate for sentiment classification as they perform better when the reviews contain both positive and negative words. However, it requires a large set of training data. Naïve Bayes works quite efficiently with a small set of training data. Appropriate set of features must be selected for sentiment classification. The

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

commonly used features are: -

- Term presence and frequency- This includes unigrams or n-grams and their frequency. For sentiment analysis unigrams, have proved to be more effective.
- POS information- Part of speech is used to disambiguate sense to guide feature selection. Text indexing and retrieval uses POS information. POS tags could be for identifying and treating differently the different meanings of polysemous words.
- Negations- These are important as they have the capability to reverse the sentiment of a review. E.g.: - "good", "not good".

In unsupervised technique, the classification is done by comparing the features of a given text against word lexicons whose sentiment values are determined prior to their use. The process starts with positive and negative word lexicons and the document is analyzed. If the document has more positive lexicons, it is positive. If not, it is considered as negative. The machine learning approach can adapt and create trained models which are suitable for specific contexts and purposes. However, the availability of training data is an issue. Moreover, it is difficult to integrate new features into the classifier which are not acquired from the training data. The learnt models often depend on the domain specific features of the training data. Therefore, the models have low adaptability.

- **Lexicon-based approach-** The lexicon-based approaches depend on a collection of known and precompiled sentiment terms and phrases, which is called a sentiment lexicon. The main methods involved in it are dictionary based approach, corpus based approach, novel machine learning approach and ensemble approach. The dictionary based approach uses an initial set of terms that are collected and annotated in a manual way. The set will be further extended by looking up the synonyms and antonyms. Examples of this approach are Wordnet and SentiWordNet. However, the method cannot deal with domain and context specific orientations. The corpus based method provides dictionaries related to a domain. The dictionaries are created from a seed of opinion terms that is built by searching related words using either semantic or statistical way. This method can produce opinion words with high accuracy. The lexicon-based approach has wider term coverage. The major issue is the availability of finite number of words in the lexicons. This can create problems in dynamic environments. The lexicons assign a fixed sentiment score to the words, irrespective of the way they are used in the input text.

- **Hybrid approach-** The sentiment classification is improved using both machine learning as well as lexicon based approaches. The sentiments are detected and measured at the concept level. However, noise removal is a cause of concern in this method.

In the current era of information retrieval and processing, it has become quite essential to focus not only on the facts and representations but also on the textual elements in the information such as opinions, sentiments, emotions and attitudes which have subjective characteristics. The detection as well as the analysis of the subjective information has become important in e-commerce, politics and security due to the generation of massive amounts of public opinion propelled by Web 2.0. Personal comments and reviews in the texts can help in the process of decision making particularly in online shopping websites. For every website or portal user ratings might not be available. Hence, we can make use of the user reviews on items or products to predict the rating of unrated items.

The explosion of web services has allowed the users to interact with each other over many platforms. One of the most common platforms are the social networking websites. These websites handle sensitive information as they make use of the micro-blog format where the contents involve plain texts, emoticons, files and live interactions such as instant messaging services. The users tend to express their opinions regarding current affairs on such platforms. Therefore, the mining of the contents from the social networking sites can help predict the outcome of major events, rate the current governance and even predict the current market trends. Since such sensitive data are handled in social networks, security of the user contents is a major issue. The user accounts are often compromised by hackers and terrorist outfits. With the help of sentiment analysis, it is possible to predict the future attacks based on the mining of the user contents as well as the response of hackers to those contents.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

- **Recurrent Neural Networks and LSTM:** Traditional neural networks cannot use its reasoning about previous events to classify or predict later ones. Recurrent neural networks address this issue. They are networks with loops in them, allowing information to persist.

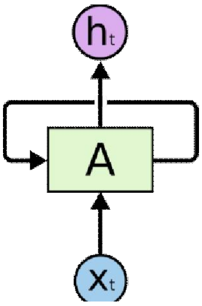

Fig 2: RNN[Deep learning documentation]

Long Short Term Memory networks, also known as LSTMs are a special category of RNN that are capable of learning long-term dependencies. They were introduced by Hochreiter and Schmidhuber in 1997. These are explicitly designed to avoid the long term dependency problem as they can remember information for long period of time. The LSTM model introduces a new structure called the memory cell. A memory cell is composed of four main elements: an input gate, a neuron with a self-recurrent connection (a connection to itself), a forget gate and an output gate. The self-recurrent connection has a weight of 1.0 and ensures that, barring any outside interference, the state of a memory cell can remain constant from one timestep to another. The gates serve to modulate the interactions between the memory cell itself and its environment. The input gate can allow incoming signal to alter the state of the memory cell or block it. On the other hand, the output gate can allow the state of the memory cell to have an effect on other neurons or prevent it. Finally, the forget gate can modulate the memory cell's self-recurrent connection, allowing the cell to remember or forget its previous state, as needed.

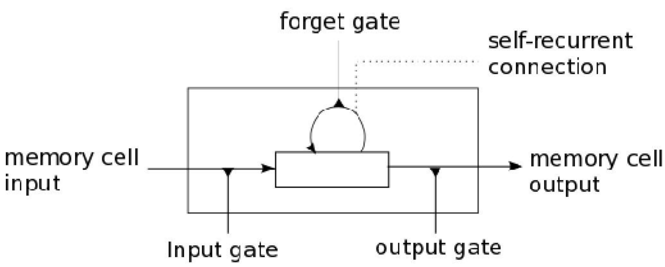

Fig 3: LSTM Cell

Recursive Neural Tensor Networks (RNTNs) were conceived by Richard Socher of MetaMind as part of his PhD thesis at Stanford. The purpose is to analyze data that has a hierarchical structure. This is extremely helpful for sentiment analysis, where the sentiment of a sentence depends not just on its component words, but on the order in which they are syntactically grouped. Each group of nodes in RNTN is a collection of neurons, where the number of neurons depends on the complexity of the input data. RNTNs are trained with backpropagation by comparing the predicted sentence structure with the proper sentence structure obtained from a set of labelled training data. Once trained, the net will give a higher score to structures that are more like the parse tress seen during the training.

In this paper, datasets from online e-commerce websites and social networking platforms like Twitter are collected. The user tweets and the hacker activities on those tweets are subjected to sentiment analysis and future attacks are predicted. Both the machine learning approach as well as the lexicon based approach will be implemented. The results are compared with the implementation of sentiment analysis through recurrent neural network with the help of Tensorflow.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

II. RELATED WORK

Li et.al have proposed a Recursive Neural Deep Model for the sentiment analysis of social data. Chinese SentimentTreebank is used on movie reviews from social websites. Andthen RNDM is ued to predict sentiment label of moviereviews on sentence level. Combination with ChineseTreebank, even baselines have showed a good performance.However, RNDM achieves the highest accuracy in predictingbinary sentiment label on sentence level.

Alessia et.al [2] have conducted a survey on the major tools and approaches used for implementing sentiment analysis. The sentiment classification approaches such as machine learning, hybrid and lexicon approach are discussed in this paper. The different tools and approaches discussed in this study has applications in several domains such as politics, finance, business and public actions.

In the study conducted by Vohra et.al [3], several applications as well as the challenges involved in the process of sentiment analysis are discussed. Knowledge-based, relationship-based, language model based approaches for sentiment analysis are highlighted. The supervised and unsupervised techniques for sentiment classification are also explained. Challenges such as the detection of spam, sarcastic sentences are also mentioned.

In a major study conducted by Olivas et.al [4] a collection of web services that provide the functionalities related to sentiment analysis are explained in elaborate detail. Different web services such as the Alchemy, Semantira, etc. have been assesses based on several criteria. Moreover, the different concepts that encompass sentiment analysis such as sentiment classification, subjectivity classification, opinion summarization, spam detection, etc. are discussed.

Lei et.al [5] have proposed a sentiment based rating prediction method to improve the accuracy in website recommender systems. The user’s sentiments on products are considered and based on the interpersonal similarity and item’s reputation, the rating is predicted. A unified matrix factorization framework has been used to implement the rating prediction task.

Hernandez et. al [6] have used Twitter, one of the major social networking platforms for sentiment analysis. By using a linear regression model, sentiment analysis is done on the tweets of users. Based on the responses of hacking activists on the tweets, future security attacks are predicted.

Apoorva et.al have proposed a novel model for data collection from Twitter [7]. The model also takes into consideration the tree like hierarchical retweeting structure to confirm that the general public's opinion matches with the experts' opinion. Lexicon based approaches have a wider coverage of sentiment classification. The use of lexicons like Wordnet and SentiWordNet can help increase the accuracy of score generation. Rahim et.al have use SentiWordNet in their work to analyze product reviews from Amazon [8]. Using the data mining and text processing tool known as GATE (General Architecture for Text Engineering) positive and negative scores of the review comments were calculated. The preprocessed and POS tagged tokens were given as input to the SentiWordNet and polarity was evaluated.

A major work involving the sentiment analysis of Twitter was contributed by A.P Jain et.al. The sentiments of users are analyzed using data mining classifiers [9]. The performance of single classifiers for sentiments analysis is compared with an ensemble of classifiers. Experimental results obtained demonstrated that k-nearest neighbour classifier has very high predictive accuracy. Result also demonstrate that single classifiers outperform ensemble of classifiers approach.

Chen et.al conducted a research on document-level sentiment classification. The study highlights the use of temporal relations of reviews for learning user and product embedding and proposes a sequence model to embed these temporal relations into user and product representations so as to improve the performance of document-level sentiment analysis. A distributed representation of each review is learnt by a one-dimensional convolutional neural network. Then, taking these representations as pretrained vectors, a recurrent neural network with gated recurrent units is used to learn distributed representations of users and products [10]. Finally, the user, product and review representations are

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

given to a machine learning classifier for sentiment classification.

An interactive visualization system was proposed by Wang et.al to analyze the public sentiments on popular topics. The system, known as SentiView [11], combines uncertainty modeling and model-driven adjustment. By searching and correlating frequent words in text data, it mines and models the changes of the sentiment on public topics. In addition, using a time-varying helix together with an attribute astrolabe to represent sentiments, it can visualize the changes of multiple attributes and relationships among demographics of interest and the sentiments of participants on popular topics. The proposed system was tested on several forums and blogs including Twitter.

In 2002, Turney proposed a significant algorithm for performing unsupervised sentiment classification. The classification of a review is predicted by the average semantic orientation of the phrases in the review that contain adjectives or adverbs. A phrase has a positive semantic orientation when it has good associations (e.g., "subtle nuances") and a negative semantic orientation when it has bad associations (e.g., "very cavalier"). In this paper, the semantic orientation of a phrase is calculated as the mutual information between the given phrase and the word "excellent" minus the mutual information between the given phrase and the word "poor" [12].

III. METHODOLOGY

The first step involved in any text processing system is the collection of datasets. Several datasets are available over the websites for research purposes. The product review datasets as well as the tweets are collected and stored in a file. Before evaluating the sentiments, the data must be preprocessed to prepare it for mining.

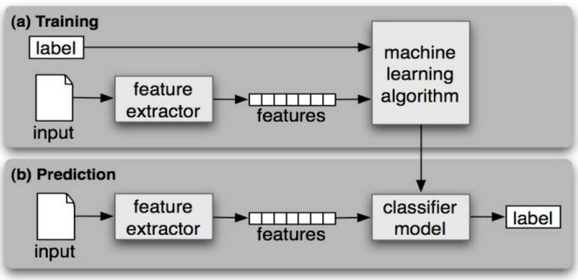

Fig 4: Supervised classification[www.nltk.org]

Data preprocessing-Preprocessing helps to reduce the size of the dataset. All the words are converted to lower case. With the help of regular expressions, the symbols “@”, “#” are matched and removed. Additional punctuations, white spaces and URLs are also removed. In a normal text document, stop words account 20-30% of the total word count. Hence, the removal of stop words will increase the speed of the sentiment analyzer. After removing stop words, tokenization is done in order to divide a large string into sub strings. This helps to break down sentences and statements in the tweets and treat them as a collection of simple words. The process is done by using regular expressions. The tokens are further subjected to stemming, where the variant forms of a specific word are replaced with its root. This can be used to match similar words in a text document and thereby improve the efficiency of information retrieval systems. The modified data will be then subjected to POS tagging which is the process of assigning each word in a sentence the part of speech that it assumes in that sentence. The Brown Corpus is used as reference for this. It is available as a package in nltk library.

The preprocessed data is fed to the sentiment analysis module. The basic algorithm is as follows: -

1. The preprocessed text having the words or the tokens is filtered to generate a feature vector. It involves the following steps:
    a. Removal of stop words. Stop words do not convey any meaning. Therefore, they can be removed.
    b. Repeating letters (usually used to stress the statements) can be removed. Eg: - hunnnggrrryyy..

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

   c. Removal of punctuations.
   d. The words that do not start with an alphabet are removed.
2. The filtered words are used to generate a feature vector. The feature vector is used to build a model which the classifier learns, from the training data and can be used to predict the supplied data. Each word from the tweets can be added to the feature vector. This is called the unigrams approach.
3. For a given tweet or text, if a feature word is present, we mark as 1 otherwise 0.
4. Based on the model of 0s and 1s, the classifier learns and predicts the sentiment.

In this work, Naïve Bayes, Maximum Entropy and SVM classifiers are implemented.The Naive Bayes classifier uses the prior probability of each label which is the frequency of each label in the training set, and the contribution from each feature. The major advantage of Naïve Bayes is that it is easy to implement and works with large datasets. When the assumption of independence holds, Naïve Bayes works better than other models such as Logistic Regression and requires less training data.

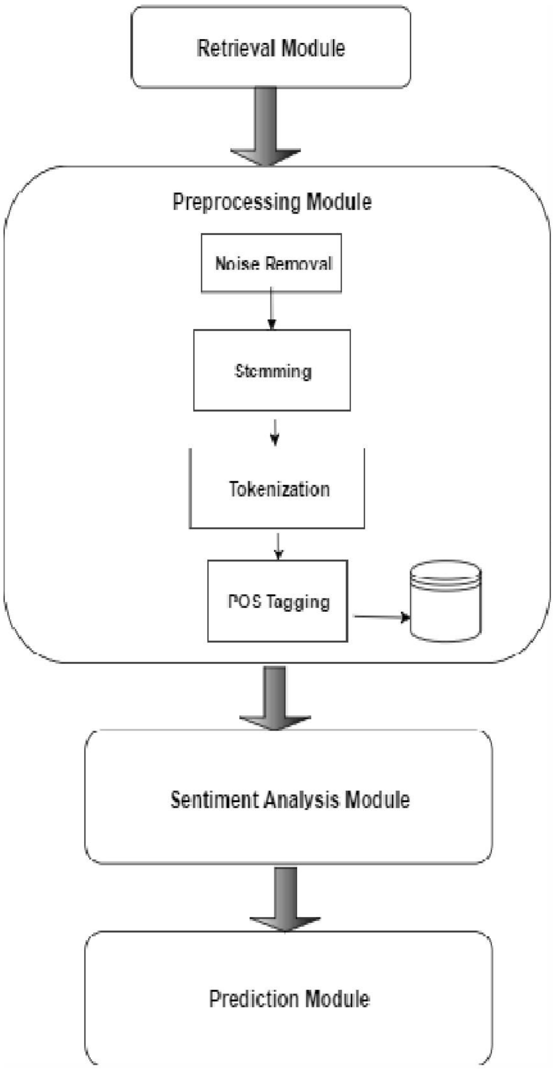

Fig 5: System architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

As an application of deep learning, Recurrent Neural Tensor Networks are also used to perform sentiment analysis. In a feed forward neural network, signals flow in only one direction from input to output, one layer at a time. In a recurrent net, the output of a layer is added to the next input and fed back into the same layer. Hence, unlike feedforward nets, a recurrent net can receive a sequence of values as input, and it can also produce a sequence of values as output. RNNs run into the Vanishing Gradient problem. Gating is used to address this problem. Gating helps the net decide when to forget the current input and when to remember it for future time steps. The gating used in this work is LSTM (Long Short Term Memory). Recursive neural tensor networks (RNTNs) are neural nets useful for natural-language processing. They have a tree structure with a neural net at each node. Recursive neural tensor networks can be used for boundary segmentation, to determine which word groups are positive and whichare negative. The same applies to the sentences. RNNs are able to process structured inputs by repeatedly applying the same neural network at each node of a Directed Acyclic Graph (DAG) [1]. The inputs to all these replicated feed forward networks are either given by using the children's labels to look up the associated representation or by their previously computed representation. Word vectors are used as features and serve as the basis of sequential classification. They are then grouped into subphrases, and the subphrases are combined into a sentence that can be classified by sentiment and other metrics. All the algorithms and techniques including RNN with Tensorflow have been implemented using Python and NLTK.

IV. SIMULATION RESULTS

The use of Naïve Bayes yields good results for sentiment analysis and classification. We follow a bag of words approach where each document is scanned and sets of positive and negative word lists are maintained. With the Maximum Entropy classifier, the probability distribution is estimated from the data rather assumption. Both perform well with comparatively smaller datasets. With large datasets, SVM has proved to be more accurate. However, with SVM, only binary classification is possible where as in Naïve Bayes and Maximum Entropy, positive, negative and neutral classifications are possible. The analysis was don with IMDB movie review datasets.

The use of RNTN with logistic regression has yielded much better accuracy compared to Naïve Bayes. Live tweets were taken and classification was done initially to identify tweets which are vulnerable to hackers. The hackers' responses to such feeds were analyzed and it was found that a significant number of accounts with negative tweets were hacked, based on the social relevance of the topic of tweets. For example, in March 2017, McDonald's account was hacked after the U.S President posted a picture of himself have food at the outlet. The current negative comments related to the presidential election have contributed a lot to hactivism.

In the sentiment analysis of Twitter data, RNTNs were giving much better accuracy compared to the machine learning approach of Naïve Bayes. The RNTN model outperforms Naïve Bayes with accurately classifying and predicting negative sentiments. The scope is captured at various levels of the tree because of the tensor-based composition function. The above graph shows the accuracy when RNTN is implemented with tensorflow.

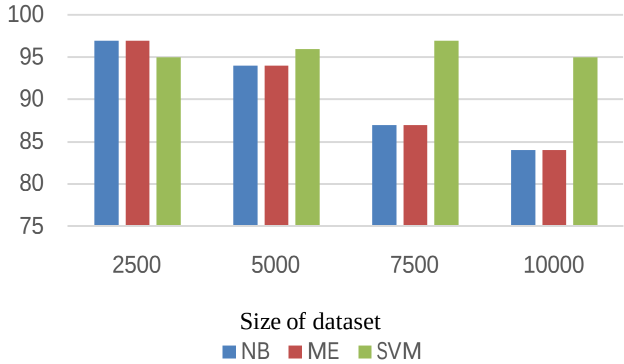

Fig 6: Comparison of accuracies of different algorithms

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

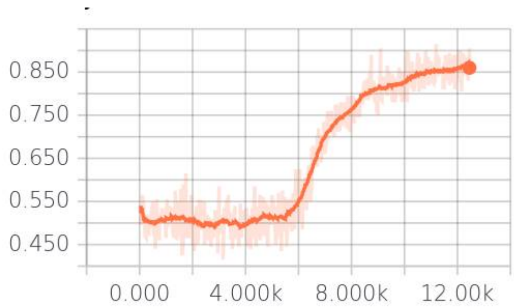

Fig 7: Accuracy of RNTN with increase in data size

V. Conclusion and Future Work

This paper focuses on the main applications of sentiment analysis in e-commerce as well as information security. By tracking the user comments a clear idea of brand perception and customer satisfaction is obtained. Moreover, the analysis of social media conversations and activities, the vulnerabilities experienced by the users based on the negativity in their comments are also highlighted. The use of RNTNs have proved to be a better option compared to the traditional machine learning algorithms.

Performing sentiment analysis on huge chunks of data is not an easy task. Several challenges such as Named Entity Recognition, Anaphora Resolution, Sarcasm are to be overcome by the application of a combination of right choice of algorithms and techniques. Along with the automated machine learning, human knowledge should also be applied to churn out the best results. The future scope of the work would be the resolution of these challenges and also achieving better accuracy in prediction as well as classification.

References

1. Vohra, M. S., &Teraiya, J. (2013, February). Applications and challenges for sentiment analysis: A survey. In *International Journal of Engineering Research and Technology* (Vol. 2, No. 2 (February-2013)). ESRSA Publications.
2. Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: a review and comparative analysis of web services. *Information Sciences*, *311*, 18-38.
3. Lei, X., Qian, X., & Zhao, G. (2013). Rating Prediction based on Social Sentiment from Textual Reviews.IEEE
4. Hern, A., Sanchez, V., Toscano, K., Nakano, M., & Martinez, V. (2016, March). Security attack prediction based on user sentiment analysis of Twitter data. In *2016 IEEE International Conference on Industrial Technology (ICIT)* (pp. 610-617). IEEE.
5. Apoorva, G., Vaishnav, N. R., Chowdary, E. D., &Uddagiri, C. (2016, January). An approach to sentiment analysis in Twitter using expert tweets and retweeting hierarchy. In *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)* (pp. 1-8). IEEE.
6. Hamouda, A., &Rohaim, M. (2011, January). Reviews classification using sentiwordnet lexicon. In *World Congress on Computer Science and Information Technology*.
7. Jain, A. P., &Katkar, V. D. (2015, December). Sentiments analysis of Twitter data using data mining. In *2015 International Conference on Information Processing (ICIP)* (pp. 807-810). IEEE.
8. Chen, T., Xu, R., He, Y., Xia, Y., & Wang, X. (2016). Learning User and Product Distributed Representations Using a Sequence Model for Sentiment Analysis. *IEEE Computational Intelligence Magazine*, *11*(3), 34-44.
9. Zhang, L., Jia, Y., Zhu, X., Zhou, B., & Han, Y. (2014). User-level sentiment evolution analysis in microblog. *China Communications*, *11*(12), 152-163.
10. Turney, P. D. (2002, July). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417-424). Association for Computational Linguistics.
11. Wang, C., Xiao, Z., Liu, Y., Xu, Y., Zhou, A., & Zhang, K. (2013). SentiView: Sentiment analysis and visualization for internet popular topics. *IEEE transactions on human-machine systems*, *43*(6), 620-630.
12. Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, *21*(4), 315-346.
13. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093-1113.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

14. Chaovalit, P., & Zhou, L. (2005, January). Movie review mining: A comparison between supervised and unsupervised classification approaches. In *Proceedings of the 38th annual Hawaii international conference on system sciences* (pp. 112c-112c). IEEE.
15. Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. N. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *Acm Sigkdd Explorations Newsletter*, *1*(2), 12-23.
16. Khan, F. H., Qamar, U., & Bashir, S. (2016). A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet. *Knowledge and Information Systems*, 1-22.
17. Boudia, M. A., Hamou, R. M., & Amine, A. (2016). A New Approach Based on the Detection of Opinion by SentiWordNet for Automatic Text Summaries by Extraction. *International Journal of Information Retrieval Research (IJIRR)*, *6*(3), 19-36.
18. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *Nature* 521.7553 (2015): 436-444.
19. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, 85-117.
20. Srivastava, N., & Salakhutdinov, R. R. (2012). Multimodal learning with deep boltzmann machines. In *Advances in neural information processing systems* (pp. 2222-2230).
21. Ravuri, S., & Stolcke, A. (2016, March). A comparative study of recurrent neural network models for lexical domain classification. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on* (pp. 6075-6079). IEEE
22. Wu, D., & Chi, M. (2017). Long Short-Term Memory with Quadratic Connections in Recursive Neural Networks for Representing Compositional Semantics. *IEEE Access*.
23. Ravuri, S. V., & Stolcke, A. (2015, September). Recurrent neural network and LSTM models for lexical utterance classification. In *INTERSPEECH* (pp. 135-139).
24. Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning* (pp. 137-142). Springer Berlin Heidelberg.
25. Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, *34*(1), 1-47.

BIOGRAPHY

Aruna Sathish is a final year M. Tech student at Vellore Institute of Technology, Vellore, India. Her research areas include Web Services, Distributed Systems, Machine Learning and Data Mining.