IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Sentiment Analysis on Customer Reviews using Deep Learning

M. Lal[1*], A. Jain[2], M. Avatade [3]

[1,2,3] Computer Dept, DYPIEMR, SPPU, Akurdi, Pune, India

[*]*Corresponding Author: milylike@gmail.com, Tel.: +91-9096942668*

Available online at: www.ijcseonline.org

Accepted: 13/Jul/2018, Published: 31/Jul/2018

***Abstract*—** The rapid growth of Web and Social Media Website brought about the need for sentiment analysis and opinion mining. Sentiment analysis and Opinion mining aims to explore the opinions or sentiments of customer reviews found in different social media platforms through deep learning technique. Deep learning is found to be more efficient to overcome the challenges faced by sentiment analysis and can handle the multiplicities involved. Deep Learning can perform sentiment analysis on any unstructured data with minimal restrictions and with no specific manual feature engineering. This paper proposes a sentiment analysis algorithm for the analysis of customer reviews by applying deep learning algorithm like Autoencoder Neural Network. Sentiment classification using deep learning promises to perform much better than the traditional supervised algorithms like Naive Bayes and SVM, with minimal constraints on the task or data for sentiment analysis.

***Keywords*—**Opinion Mining, Sentiment Analysis, Sentiment Classification, Deep Learning

I. INTRODUCTION

Previous research was mostly based on traditional machine learning tools like LSA, Naïve Bayes and SVM. Deep learning has recently proved to show better performance as a sentiment analysis tool in recent years [2]. Deep learning uses nonlinear processing units of multiple layers for aspect extraction and analysis. The lower layers learn simple features, while higher layers learn more complex features derived from lower layer features. This paper proposes to replace sentiment classification methodologies like Latent Semantic Analysis (LSA) to the use of deep learning algorithm like Auto Encoder Neural Networks for finding the sentiments in customer reviews. [3].

The rest of paper is organized as follows. Section 2 provides the related work on sentiment analysis. Section 3 gives the methodologies used. Section 4 explains about the expected results. Section 5 gives will be conclusion. Section 6 will be references.

II. RELATED WORK

This work is mainly related in replacing the traditional machine learning techniques with deep leaning.

Mostly deep learning models in NLP need word embedding results as input features. It is a technique for language modelling and feature learning, which transforms words in a vocabulary to vectors of continuous real numbers. The learning of word embedding can be done using neural networks or matrix factorization. [3] Autoencoder Neural Network is a three-layer neural network, which sets the target values to be equal to the input values. It is able to learn nonlinear representations, which gives better performance results than its linear counterparts, such as Principal Component Analysis (PCA) or Latent Semantic Analysis (LSA).[3] Another deep learning method called Convolutional Neural Network (CNN) can be applied to numerous Natural Language Processing tasks like Part-Of-Speech Tagging, Parsing, Chunking, Semantic Role Labelling and Named Entity Recognition. It takes concatenated word vectors of the text as input and involves convolutional and max-pooling layers prior to the general neural network framework [1]. Recurrent Neural Network (RNN) is a class of neural networks whose connections between neurons form a directed cycle. RNN is capable of conditioning the network on all the previously seen inputs (words in case of a sentence). In addition to dependency on the current input, the value of each hidden layer unit also depends on its previous state, thereby propagating the effects of words over the sentence [3].A recursive neural tensor network (RNTN) is a kind of deep learning model in which the same set of weights is applied recursively over a structure (e.g. tree), to produce a structured or a scalar prediction over variable length input, by traversing the given structure in topological order [4]. Long Short Term Memory network (LSTM) is a special type of RNN, which is capable of learning long-term dependencies. This model allows retention of information over a much longer period through

the use of the memory cell and hence produces appreciable results when applied to NLP tasks [3].

III. METHODOLOGY

Aspects which are mostly nouns/noun phrases and Opinion which are mostly adjective modifiers are to be extracted by Stanford POS Tagger. This proposed design is domain independent and unsupervised, avoiding labelling data for supervised learning methods which is tedious and consumes time.[1] Scoring of opinions is proposed to be done SentiWordNet which is a lexical resource of opinion words for terms in the English language which help in the extraction of opinion words [4]. Aspect reduction is to be carried out by Autoencoder Neural network.

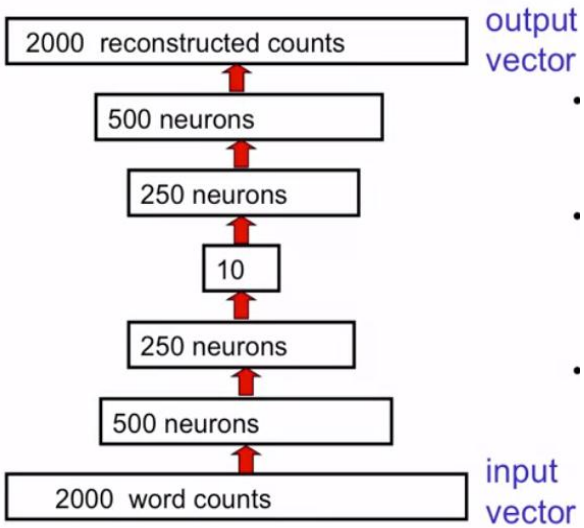

Figure 1. How to compress the count vector.

The neural vector will be trained to reproduce input vector as output vector. This forces it to compress as much information as possible. These compresses numbers are then good enough to compare aspects in customer reviews. We have to divide the aspects extracted by N, where N is the total non-stop words in the customer review. The resulting vector gives the probability of getting a particular word if any aspect is picked from the review. At the output of the autoencoder we can use softmax. To train the network, the word can be treated as possibilities. The visible to Hidden weights is to be made N times bigger than the hidden to visible. [8]

$$X_{i,j}=\log(1+C_{i,j})/\max_j \log(1+C_{i,j}) \quad (1)$$

Where ci,j denotes the number of occurrences of the jth word in the ith document, xi,j denotes the normalized count.[9]

IV. RESULTS AND DISCUSSION

We have to train on set of aspects extracted in the initial step for other set of training cases of customer reviews. First train the neural network then fine tune with backdrop. After this step it has to be tested on another set of separate customer reviews. One review will be picked query. Other reviews will be ranked using consine of angles between codes. This step has to be repeated for other test reviews as query as well. For performance measurement, the number of reviews retrieved has to be plotted against the proportion that is in the same hand labelled class as the query document.

V. CONCLUSION and Future Scope

The proposed algorithm is expected to perform better than the traditional techniques such as LSA and PCA [3]. The algorithm has to be implemented, trained and tested against various customer reviews. Performance has to be recorded and evaluated against our expectations.

REFERENCES

[1] Aspect Extraction & Segmentation In Opinion Mining , Mily Lal , Kavita Asnani, International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 3 Issue 5 May, 2014 Page No. 5873-5878

[2] Sentiment Analysis and Deep Learning: A Survey - CFILT , Prerana Singhal and Pushpak Bhattacharyya, 2016.

[3] Deep Learning for Sentiment Analysis: A Survey, Lei Zhang, Shuai Wang, Bing Liu, 2018

[4] Esuli, A., and F. Sebastiani. "Determining term subjectivity and term orientation for opinion mining." In Proceedings of Annual Conference of the European Chapter of the Association of Computational Linguistics (EACL-2006)

[5] Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. International Conference on Knowledge Discovery and Data Mining (ICDM).

[6] Liu, B., "Sentiment Analysis and Subjectivity" Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau), 2010

[7] Pang B and Lee L. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2008. 2(1–2): pp. 1–135.

[8] Vincent P, Larochelle H, Bengio Y, and Manzagol P-A. Extracting and composing robust features with denoising autoencoders. In Proceedings of the International Conference on Machine Learning (ICML 2008), 2008.

[9] Zhai S, Zhongfei (Mark) Zhang. Semisupervised autoencoder for sentiment analysis. In Proceedings of AAAI Conference on Artificial Intelligence (AAAI 2016), 2016.

[10] AR. PonPeriasamy, G. Vijayasree, "Data Mining Techniques for Customer Relationship Management", International Journal Of Computer Sciences and Engineering, Vol.5, Issue.4, pp.120-126 , 2017

Authors Profile

Mrs Mily Lal received B.Tech-Computers degree from Calicut University and M.E-I.T degree from Goa University. She is Assistant Professor of Computer department in DYPIEMR,Pune.

Mrs Abha Jain received B.E-Computers degree from and M.E degree from Rajasthan University. She is Assistant Professor of Computer department in DYPIEMR,Pune.

Mrs Mohini Avatade received B.E degree from and M.E degree from Pune University. She is Assistant Professor of Computer department in DYPIEMR,Pune.