Review Article

Sentiment Analysis using NLP: A Survey

Harsh Dubey[1], Prof. Saurabh Sharma[2]

[1,2]Baderia Global Institute of Engineering and Management, INDIA

ABSTRACT

Sentiment analysis is achieved by collecting the data, extracting the meaning, and eventually averaging the number of terms. The survey is concerned with evaluation methodologies and related methods. In this paper, we discuss the "aspect-based sentiment analysis", which is distinguished into three classifications: Aspect classification based on explicit and co-referential examples We are going to explore the emotion of praise and criticism in this post. Or Aspect-based Classification promises to both minimize the noise as well as well as make other tasks simpler with an added advantage of efficiency. Using neural networks to achieve this accuracy.

KEYWORDS

Aspect based sentiment classification, neural network

1. INTRODUCTION

Global tourism is on the rise. Citizens from around the world share their views on a wide range of sites, from blogs to Facebook and Twitter. They put out a joyful or sad stories about it in a nonjudgmental way to showcase their feelings In one's feelings, mood, feeling, and thoughts about a location or business are known as sentimentality.

The feedback can be wrong, but there can be good data as well. The only drawback to positive reviews is that they mark the product as being true or accurate in the majority of cases. The algorithm's efficiency changes as we distinguish between negative and positive emotions. Generally, people use phrases like "the food has never tasted so good" or "so tasty today"that go in the opposite direction of 'positive' or 'negatively.'" Stillbirth occurs when a human is born with or delivered with the umbilical cord wrapped twice around the neck, or when both halves of the mother's body are found to be a VVR or a Cordite propellant septic embolism occurs when both halves of the mother are found to have uterine adhesion; delivered with a baby encircled in a complex cord, a Cordite septic embolism occurs when the VB is wrapped around her neck, the vena comes into contact with it, or the mother is found to be VB+

Sentiment analysis can be described as at three different levels: document, aspect, and sentence. sentence polarity in the sentence It is incorrect to say that sentences and documents have little in common. It classifies entities based on the aspects of their relation to each other.

They can provide 1) identification, 2) sentiment grouping into a positive and negative group. When creating an Aspect, we have to deal with three different problems: being fully identified with the aspects. The first method allows us to find the implicit relationship, and the co-

referred relationship, and the third method allows us to define the explicit one. The classification is evaluated by looking for the type of terms or expressions that visitors use. For the classification and identification of data, the neural algorithms are utilised. Idealistic When people visit certain hotels on their travel, they review these hotel and service provided. These reviews are to be classified on the aspect level and here the problem of efficiency arises on the co–referential aspect identification in which entities are given synonymous names and then reviewed for e.g., atmosphere and ambience. Also, when food is implicitly termed out as „dish" is a problem in implicit aspect identification. Sometimes the irrelevant aspects of a hotel like curtains, beddings, toiletries, etc are reviewed which affects the accurate prediction [5,7,8]

In this paper we are going to study three sections: Section II describes about the related work in sentiment analysis for tourist review, Section III illustrates the process of data cleaning, pre-processing, tokenization for sentiment classification, Section IV is elaborating the proposed algorithm and Section V finally gives the results and presents the conclusion.

2. RELATED WORK

A. Aspect Extraction

Aspect extraction is done by identifying the aspects, which are classified into three categories rule based, seed based and topic models based. Rule based is dependent on frequency, importance, and appearance. Marrese Taylor *et al.* [8]-[9] proposed to firstly apply POS Tagger following with the extraction of noun as aspect. Hai *et al.* [10] used Intrinsic Domain Relevance (IDR) and Extrinsic Domain Relevance (EDR) for tourism domain. Afzal *et al.* [11] proposed a method using FURIA Machine learning algorithm.

Seed Based method identified aspects using grammar. It

found a connection between the seed word and the sentiment word. Bootstrapping has been applied between seed word and review word to quantify overlapping and dependency. Zhu *et al.* [15] used bootstrapping.

Topic Based techniques are using different topics and at every point probability distribution is done over different words. Wu and Ester *et al.* [16] have used substance generative techniques. Xianghua *et al.* [17] has applied sliding window technique. Xueke *et al.* [18] applied Latent Dirichlet Allocation.

The main limitations of the above discussed methods are not discarding irrelevant aspects and not working on the co- referential aspects issue.

B. Aspect Based Sentiment Classification

This determines the orientation of views. Wang *et al.* used Latent Rating Regression (LRR). Xueke *et al.* [18] predicted opinion using SVM after preprocessing with Natural Language Processing (NLP) for segmentation and Parts of Speech (POS) Tagger. Pontiki *et al.* [6]-[13] used SVM to perform binary classification. De Albornoz *et al.* [7] used lodging audits and represented survey rating as VFI. Afzaal *et al.* [11] proposed three stage fuzzy aspect-based classification method.

3. PROCESS FLOW OF PROPOSED METHODOLOGY

A. Data Collection

This process of data collection is done by collecting all the customer reviews. These reviews are collected from social media platforms.

A. Data Preprocessing

In this process of data preprocessing the noisy data is removed. In this process we discard the irrelevant data collected. Sentences are identified by their delimiters used such as exclamation mark, question mark, etc. Duplicate data has to be removed.

B. Aspect Identification

In this process the aspects are identified from the sentences. POS tagger is applied to each sentence to retrieve lexicons. Noun phrases and Noun are extracted from the POS tags and else are discarded. The sentence uses unigram tokenizer as input to decision trees.

C. Aspect Based Opinion Classification

There are three basic stages:

- Discard Opinion Sentences: The word itself suggest that the sentences which are irrelevant are to be discarded. The dependency between the sentiment word and aspect is to be searched and then to be processed for classification.

- Feature Extraction: There are four types of feature extraction methods used for POS tags. In the POS tag we analyze and extract the tokens in the grammar by searching for the verb, adverb, and adjective. There is bigram, trigram used for processing the negation.

- Classifier Training: This is very main model as the

algorithm classifies the positive and negative reviews, in which we also have to consider the linkages between the sentiment word and aspects.

4. PROPOSED METHODOLOGY

A. Data Cleaning, Data Preprocessing and Tokenization

- Data Cleaning and preprocessing:

The data collection is done from various platforms where the consumer shares their experiences. Now the next stepdone is Data cleaning and preprocessing. In this the irrelevant sentences are to be discarded. In this removal of noisy and redundant data is to be done. In this removal of the unwanted punctuation marks, spelling errors, grammatical errors, writing in abbreviations just to look cooler.

- Tokenization:

The data is tokenized into N grams for feature extraction here (N=1,2,3). In this there are unigram, bigram, and trigram which helps in extracting the actual sentiment in the sentence. It helps in seeing the linkage between the sentiment word and aspects.

B. Aspect Based Sentiment Classification

Each algorithm was evaluated on different sizes of the dataset and different feature types. The sentiment score calculation is done on the dataset after sentiment detection over the dataset. This is given as the input to aspect-based sentiment classification. Results from the implicit detection and explicit detection of aspects using neural network is used as an input.

C. Travel and Tourism Ranking and Recommendation

In this we are recommending tourist places to visit by providing them the ranking. In this we are recommending by analyzing and processing the aspect-based sentiment classification performed on the tourist reviews.
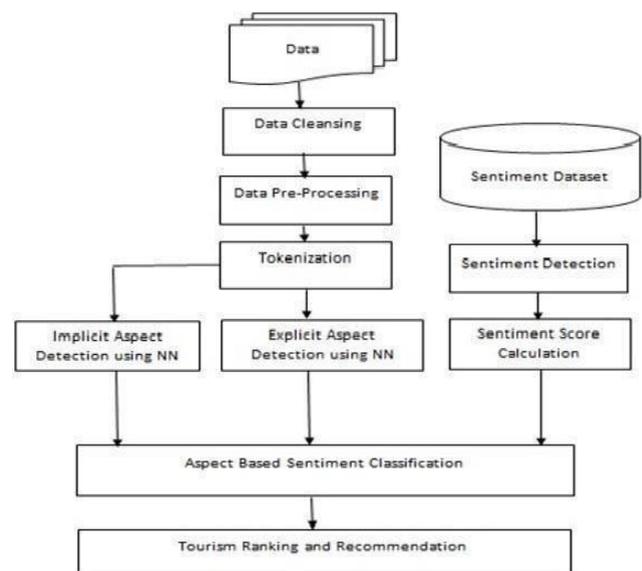

Fig. 1. Proposed Methodology

Fig. 2. Extracted Sentiment words from review

5. COMPARISON AND ANALYSIS

In this paper we are going to compare machine learning algorithms over neural network algorithms. The various algorithms in machine learning are:

A. Naïve Bayes Multinomial:

Naïve Bayes Classifiers are used to perform complicated work. These are scalable even. It works good for countable data and is allowing multinomial distribution for each single feature of that class. It uses prior probability and post prior probability for classification [1].

B. Support Vector Machine:

Here we use centroid classification algorithm and by calculating all the centroids closer to each other in that document they are ranked as similar. Here we are using discriminative classifier for text classification [1].

C. Maximum Entropy:

Here we not assuming any relationship between any of the features. Here we estimate the conditional distribution of a class label which gives the maximum entropy for the system [1].

D. Fuzzy Lattice Reasoning:

In this we are using fuzzy lattice rules for classifying data sets [1].

E. Random Forest:

In this we are using random elements for generating a tree but not all the elements are involved [1].

In case of neural network, we have certain algorithms which gives higher accuracy over above:

F. Recurrent neural network:

There are three layers in this network which are input layer, hidden layer and output layer. Hidden layer acts as a memory for the network [2].

G. Convolutional neural network:

This is inspired by feed forward neural network. In this model we are working on the outputs of the subsets of adjacent words which are termed as neurons and are again processed to give output of each neuron [2].

H. Long short time memory:

It is used for prediction in positive sentiment analysis. LSTM exhibits higher accuracy over others methodology [2].

In the above algorithms the performance is dependent on the data inputs polarity and efficiency of classification.

While comparing we consider accuracy and efficiency as the most important parameters. Deep neural network algorithms are much more efficient and accurate over machine learning algorithms. Deep Neural network has many hidden layers as compared to SVMs. Deep Neural Network takes lesser time as there is lesser human intervention and it carries out automatic feature extraction. It requires larger data sets than Machine Learning Algorithms and the training costs are also high [2].

6. CONCLUSION

The paper presented the way sentiments are to be classified into positive or negative. The way both explicit and implicit extracts features using Noun (NN) and Noun Phrases from the opinion text, reviews text, the removal of irrelevant and redundant text. Here we have proposed deep neural network which gives a better performance in accuracy and gives high efficiency in classifying and extracting sentiments.

7. REFERENCES

[1] Muhammad Afzaal, Muhammad Usman, and Alvis Fong "Tourism Mobile App with Aspect-Based Sentiment Classification Framework for Tourist Reviews", *IEEE Transactions on Consumer Electronics*, vol. 10. DOI 10.1109/TCE.2019.2908944, May 2019.

[2] C. A. Martin, J. M. Torres, R. M. Aguilar, and S. Dias, "Using Deep Learning to Predict Sentiments: Case Stu dy in Tourism," Volume 2018, Article ID 7408431, page 9.

[3] C. S. Khoo and S. B. Johnkhan, "L e x i c o n -based sentiment analysis:C omparative evaluation of six sentiment lexicons," *Jour. Inform.Scien.*, vol. 44, no. 4, pp. 491-511, Aug. 2018, DOI:10.1177/0165551517703514

[4] A. Moreo, M. R omero, J. Castr o, and J. M. Zurita, "Lexicon-based comments-oriented news sentiment analyzer system," Expert S ys. With Appli., vol. 39, no. 10, pp. 9166-9180, Aug. 2012, DOI:10.1016/j.eswa.2012.02.057

[5] Qurat Tul Ain, Mubashir Ali, Amna Riaz, Amna Noureen , Muhammad Kamran , Babar Hayat and A. Rehman, " Sentiment analysis using deep learning techniqu es :A R eview," International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, 2017

[6] M. Pontiki et al., "SemEval-2016 task 5: Aspect based sentiment analysis, " *in Proc. 10th Int. Workshop Semantic Evaluation* , 2016,pp. 19-30.

[7] G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu, "Sentiment classification: The contribution of ensemble learning," *Decision Supp. Sys.*, vol. 57, pp. 77-93, Jan. 2014, DOI: 10.1016/j.dss.2013.08.002

[8] J. C. D e Albornoz, L. Pla za, P. Gervá s, a nd A. Díaz, " A joi nt model of featur e mining and senti ment a nalysis for product revi ew rating," in Eur. Conf. Inform. Retrieval., 2011, pp. 55-66, DOI: 10.1007/978- 3-642-20161-5_8

[9] E. Marrese-Taylor, J. D. Velásquez, and F. Bravo-Marquez, "Opinion zoom: A modular tool to explore tourism opinions on the web, " in Proc. IEEE/WIC/AC M Int. Joint Conf. Web Intell. And Intell. Agent Technol., 2013, pp. 261-264, DOI: 10.1109/WIIAT. 2013.193

[10] E. Marrese-Taylor, J. D. Velásquez, and F. Bravo-Marquez, "A novel deterministic approach for aspect-based opinion mining in tourism produ cts reviews," *Expert Sys. With Appli.*,

vol. 41, no. 17, pp. 7764-7775, Dec. 2014, DOI: 10.1016/j.eswa.2014.05.045

[11] Z. Hai, K. C hang, J.-J. Kim, and C. C. Yang, " Identifying features in opinion mining via intrinsic and extrinsic domain relevance," *IEEETrans. Know. And Data Engi.*, vol. 26, no. 3, pp. 623-634, Mar.2014, DOI: 10.1109/TKDE.2013.26

[12] M. Afzaal, M. Usman, A. C. M. Fong, S. Fong, and Y. Zhuang, "Fuzzy Aspect Based Opinion Classification System for Mining 10 Tourist Reviews," *Adva. In Fuzzy Sys.*, vol. 2016, Oct. 2016, DOI:10.1155/2016/6965725

[13] X. Xu, X. Cheng, S. Tan, Y. Liu, and H. Shen, "Aspect-level opinion mining of online customer reviews, " *China Commun.*, vol. 10, no. 3, pp. 25-41, Mar. 2013, DOI: 10.1109/CC.2013.6488828

[14] A. Mukherjee and B. Liu, "Aspect extraction through semisupervised modeling," in Proc. 50th Annu. Meeting Assoc. for Computational Linguistics, 2012, pp. 339-348.

[15] H. Wang, Y. Lu, and C. Zhai, "Latent aspect rating analysis on review text data: a rating regression approach," *in Proc. 16th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining,* 2010, pp.783-792, DOI: 10.1145/1835804.1835903

[16] J. Zhu, H. Wang, M. Zhu, B. K. Tsou, and M. Ma, " Aspect-based opinion polling from customer reviews,"*IEEE Trans. AffectiveCompu.*, vol. 2, no. 1, pp. 37-49, Jan. 2011, DOI: 10.1109/TAFFC.2011.2

[17] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[18] F. Xianghua, L. Guo, G. Yanya n, and W. Zhiqiang, " Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon," *Knowledge-Based Sys.*, vol. 37, pp.186-195, Jan. 2013, DOI: 10.1016/j.knosys.2012.08.003