

## **An Approach to Effective Co- Referential Aspect and Entity Based Sentiment Classification of Tourist Reviews using Neural Networks and Machine Learning**

*Dr Kamlesh A. Waghmare, Sheetal K. Bhala\**  
*Department of Computer Science and Engineering,*  
*Government College of Engineering, Amravati, India.*  
*\*Corresponding Author*  
*E-mail Id:-sheetalbhala1995@gmail.com*

### **ABSTRACT**

*Tourism is a widely increasing industry and tourist reviews plays the vital role for the people to select the place , hotel , restaurants and others. Sometimes there is collection of irrelevant data which leads to the unwanted errors. So as to minimize this noisy data we make use of aspect based sentiment classification. In this paper we are trying to identify the implicit explicit and co referential aspects and perform the sentiment classification with the higher efficiency. Here in this paper we would be using the Machine learning Algorithms and the neural network algorithms for the sentiment classification.*

**Keywords:-***Aspect based sentiment analysis, Machine learning, Neural network, Tourist reviews.*

### **INTRODUCTION**

Tourism is a largely expanding industry with increasing importance of countries internationally. With the increasing fad of social media platforms, people are more intended to sharing traveling photos and their sentiments about the place, food, restaurants and hotels ambience. People all over the world are giving out their emotions and opinions about the travel left for others to judge and get opinionated and reviewed virtually on anything. Actually this dataset of the reviews adds values for various other applications.

Sentiment analysis, we use natural language processing domain for focusing on the expression implied by the text review. In sentiment analysis, we often come across opinion mining which is data mining and information retrieval system. Subjective Analysis is used for both sentiment analysis and opinion mining. These all the terms study the emotion, opinion, sentiment, evaluation, appraisal and attitude. There are in all three steps:

1. Identifying of sentiment-target pairs in the text.
2. Classifying of the sentiment-target pairs, according to a predefined set of sentiment values, for instance positive and negative. At times the target is classified according to a predefined set of aspects as well.
3. The sentiment values are aggregated for each aspect depending on the specific needs and requirements of the application.

For example we refer to a sentence Ambience is awesome, but taste is not so, so here we imply ambience is the atmosphere and taste is about the food. Here the first part is a positive sentiment and second is about the lesser satisfaction i.e. negative sentiment. Here the sentiments are classified for the two different aspects, those are atmosphere and food. So the task here was to identify the aspects and its polarity. Here the aspect ambience is co referential aspect and taste is an implicit aspect.

The field of sentiment analysis is basically gathering, aggregating and analysing from the text reviews. However, the number of reviews are ever increasing and hence classifying them and identifying the aspects also becomes tough. Hence we require to have certain process models which help us with the working and processing. This paper illustrates the help from the machine learning algorithms and neural network algorithms to identify aspects and classification of the sentiments. These algorithms help us to understand the polarity of the sentence and identifying the aspects for the score calculation.

## **RELATED WORK**

### **Aspect Extraction**

Three main categories are rule based, seed method, and topic method. Rule based method derived from importance, frequency, and appearance. Marrese Taylor *et.al* [8, 9] proposes that firstly it will be using Part of Speech Tagger on sentences and then followed by the extraction of noun as aspects. Score Calculation is given by Hai *et al.* [10] using the Intrinsic Domain Relevance (IDR) and the Extrinsic Domain Relevance(EDR). FURIA machine learning algorithm was proposed by Afzaal *et al.* [11] as an improvement in the rule based methods. Identifying the aspects by deriving a grammatical connection in between the sentiment words and seed words is called as Seed Based method. Bootstrapping is applied to determine the value of overlapping and dependency between the sentiment words and seed words. Zhu *et al.* [15] proposed that identifying the important aspects becomes better by using bootstrapping. They used two techniques for aspect determining 1)N grams and 2)POS Tagger. The inefficiency of rule based method and seed based

methods is disability of identifying the lesser frequent aspects.

Topic model based technique supposes that every sentiment is the combination of various topics and every idea is the probability distribution over different words. Wu and Ester [16] used the added substance generative techniques to identify the aspects. Xianghua *et al.* [17] proposed the sliding window technique for identifying the aspects from the tourist reviews, but at the initial stage of some of the aspects was a little lesser accurate.

Xueke *et al.* [18] proposed that topic modeling is done by applying Latent Dirichlet Allocation.

### **Aspect-Based Sentiment Classification**

This is being used to find the orientation of the reviews and the sentiments in the distinguished classes those are like good, bad and many others. Wang *et al.* [14] applied the Latent Rating Regression (LRR) to manage the specific words to be associated to different classes on the rating scale. Xueke *et al.* [18] suggest the usage of SVM algorithm for identifying the specific aspects by preprocessing the NLP Toolkit for the sentence segmentation and the POS Tagger. Pontiki *et al.* [19, 20] proposed the method for binary classification which increased robustness and stability. SVM with linear kernel was used. De Albornoz *et al* [7] represented survey rating as VFI and have made us of lodging audits to verify the method. Afzaal *et al.* [11] worked on classifying sentiments into negative and positive using the three stage fuzzy aspect based classification method.

## **PROPOSED WORK**

### **Data Collection**

It is done from the social media platforms and is done by collecting the reviews related to various aspects.

### **Data Pre-processing**

It is by removing the unwanted words or grammar which adds no value to the algorithm or computations. The noisy data is cleared. Removing of punctuation marks like exclamation mark, question mark and others.

### **Aspect Identification**

These are identified from the sentences. POS Tagger is used to retrieve each lexicon from the sentences. Nouns and noun Phrases are extracted from the POS tags.

### **Aspect Based Opinion Classification**

The following three steps are followed:

#### *Discard opinion sentences*

This is done to remove the unwanted and irrelevant sentences. Sentiments word and aspect has to be checked if any dependency exists.

#### *Feature Extraction*

Extracting the tokens such as verbs, adverbs, adjectives and analyzing them with the help of POS Tagger. The bigram, unigram and trigram is also used for this method.

#### *Classifier Training*

The relation between the sentiment word and aspects is classified as a positive or negative sentence.

#### *Data Cleaning, tokenization*

Removal of spelling errors, punctuations, exclamation, grammatical errors, abbreviations, and others. Removal of any redundant or irrelevant sentences that add no value to the computations or classifications. Here tokens are formed by unigram, bigram or trigram helping to

extract the sentiment words and their relation with the aspect.

### **METHODOLOGY**

#### **SVM**

SVM works by plotting data points in n-dimensional space. The number of dimensions depends on the number of features. Then classification is performed by finding the hyperplane which divides the categories very well. It works on multi-classification problem too. It has advantages over imbalanced dataset. SVM algorithm is used for detecting aspect categories from aspect terms. Aspect terms are plotted using word embedding in hyper space and then categorized into particular categories.

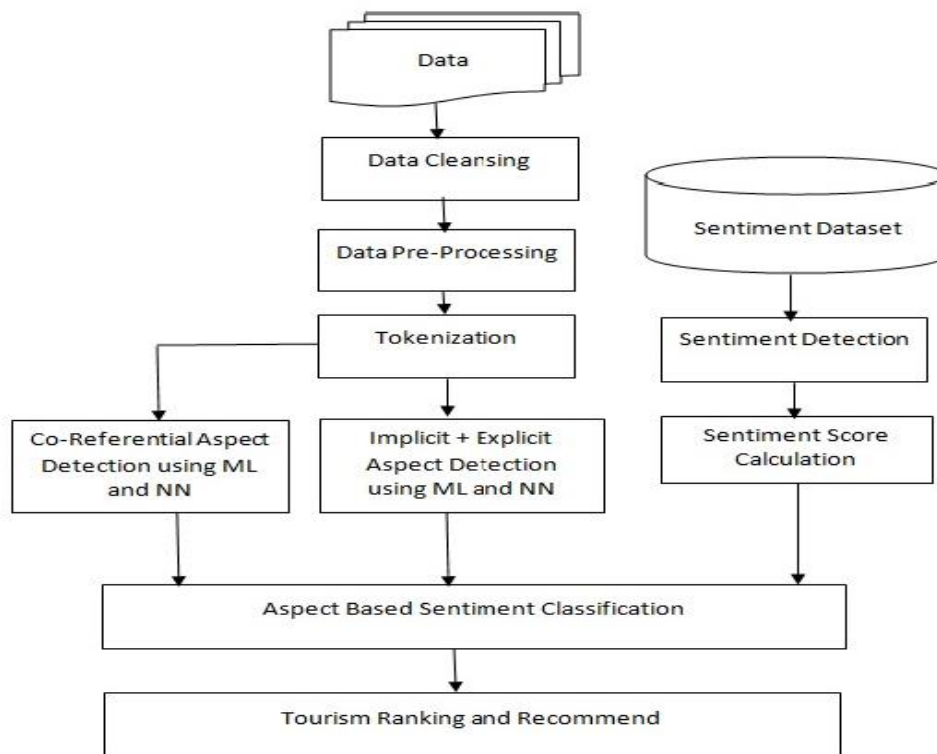
#### **NBM**

NBM is part of naïve Bayes classifiers which is based on applying Bayes' theorem with strong independence assumptions between the features. It is also used for detecting aspect categories.

#### **RNN and LSTM**

Detecting aspect categories is complicated because there are three types of aspect categories viz. explicit, implicit and co-referential. Hence, we proposed to use different algorithms for detecting different categories. For detecting explicit and implicit categories, RNN algorithm is used. RNN is feedforward neural network that has an internal memory.

For detecting co-referential aspect categories, LSTM is modified version of recurrent neural networks, which makes it easier to remember past data in memory. To detect different types of aspects, different algorithms are applied and their result is then combined to get better accuracy and precision.



*Fig.1: Proposed System*

## CONCLUSION

This paper presented the sentiment based classification of the co-referential aspects and implicit and explicit aspect. Here we have used the Machine learning algorithms such as NBM and SVM along with the neural network algorithms RNN and LSTM. The way co-referential, explicit and implicit extracts features using Noun (NN) and Noun Phrases from the opinion text, reviews text, the removal of irrelevant and redundant text. The deep neural network gives a better performance in accuracy and gives high efficiency in classifying and extracting sentiments.

## REFERENCES

1. Kim Schouten and Flavius Frasincar. *Survey on Aspect-Level Sentiment Analysis. IEEE transactions on knowledge and data engineering.*2018.28(3).
2. Muhammad Afzaal, Muhammad Usman, Alvis Fong. *Tourism Mobile App with Aspect-Based Sentiment*

*Classification Framework for Tourist Reviews. IEEE Transactions on Consumer Electronics.*DOI 10.1109/TCE.2019.2908944,May 2019.10.

3. C. A. Martin ,J. M. Torres,R. M. Aguilar,and S. Dias. *Using Deep Learning to Predict Sentiments: Case Study in Tourism.* 2018. Article ID 7408431.9p.
4. C. S. Khoo and S. B. Johnkhan. *Lexicon-based sentiment analysis:Comparative evaluation of six sentiment lexicons.* Jour. Inform.Scien.2018.44(4):491-511p. DOI:10.1177/0165551517703514
5. A. Moreo, M. Romero, J. Castro, and J. M. Zurita. *Lexicon-based comments-oriented news sentiment analyzer system.* Expert Sys.with Appli.2012.39.(10):9166-9180p. DOI:10.1016/j.eswa.2012.02.057
6. Qurat Tul Ain , Mubashir Ali , Amna Riaz, Amna Noureen , Muhammad Kamran , Babar Hayat and A. Rehman,. *Sentiment analysis using*

- deep learning techniques :A Review.* International Journal of Advanced Computer Science and Applications.2017.8.(6).
7. M. Pontiki et al., SemEval-2016 task 5: *Aspect based sentiment analysis.* Proc. 10th Int. Workshop Semantic Evaluation. 2016:19-30p.
  8. G. Wang, J. Sun, J. Ma, K. Xu, and J. Gu. *Sentiment classification: The contribution of ensemble learning,* Decision Supp.2014.5:77-93p. Jan. 2014, DOI: 10.1016/j.dss.2013.08.002
  9. J. C. De Albornoz, L. Plaza, P. Gervás, and A. Díaz. *A joint model of feature mining and sentiment analysis for product review rating.* Eur. Conf. Inform. Retrieval. 2011:55-66p. DOI: 10.1007/978-3-642-20161-5\_8
  10. E. Marrese-Taylor, J. D. Velásquez, and F. Bravo-Marquez. *Opinion zoom: A modular tool to explore tourism opinions on the web.* Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. And Intell. Agent Technol. 2013: 261-264p. DOI: 10.1109/WIIAT. 2013.193
  11. E. Marrese-Taylor, J. D. Velásquez, and F. Bravo-Marquez. *A novel deterministic approach for aspect-based opinion mining in tourism products reviews.* Expert Sys. With Appli.2014.41(17):7764-7775p. Dec. 2014, DOI: 10.1016/j.eswa.2014.05.045
  12. Z. Hai, K. Chang, J.-J. Kim, and C. C. Yang. *Identifying features in opinion mining via intrinsic and extrinsic domain relevance.* IEEETrans. Know. and Data Engi.2014.26(3):623-634p. Mar. 2014, DOI: 10.1109/TKDE.2013.26
  13. 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. Oct. 2016, DOI:10.1155/2016/6965725
  14. X. Xu, X. Cheng, S. Tan, Y. Liu, and H. Shen. *Aspect-level opinion mining of online customer reviews.* China Commun. 2013.10(3):25-41p.Mar. 2013, DOI: 10.1109/CC.2013.6488828
  15. A. Mukherjee and B. Liu. *Aspect extraction through semisupervised modeling.* Proc. 50th Annu. Meeting Assoc. for Computational Linguistics. 2012:339-348p.
  16. H. Wang, Y. Lu, and C. Zhai. *Latent aspect rating analysis on review text data: a rating regression approach.* Proc. 16th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining. 2010,pp. 783-792, DOI: 10.1145/1835804.1835903
  17. J. Zhu, H. Wang, M. Zhu, B. K. Tsou, and M. Ma. *Aspect-based opinion polling from customer reviews.* IEEE Trans. Affective Compu. 2011.2(1):37-49p. Jan. 2011, DOI: 10.1109/TAFFC.2011.2
  18. 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. 1987.2:740–741p. A [Digests 9th Annual Conf. Magnetism Japan.1982:301p].
  19. F. Xianghua, L. Guo, G. Yanyan, and W. Zhiqiang. *Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon.* Knowledge-Based Sys.2013.37:186-195p. DOI: 10.1016/j.knosys.2012.08.003