International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 1 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT183812

Aspect Ranking Technique for Efficient Opinion Mining using Sentiment Analysis : Review

Prof. Sonali D. Borase#[1], Prof. Prasad P. Mahale*[2]

#Assistant Professor, Department of Computer Engineering, NMIMS Mukesh patel school of technology, Shirpur, Maharashtra, India

*Assistant Professor, Department of Computer Engineering, SES's R.C Patel institute of technology, Shirpur, Maharashtra, India

ABSTRACT

Opinion mining, also called sentiment analysis, is the field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. Even though facts still play a very important role when information is sought on a topic, opinions have become increasingly important as well. Opinions expressed in blogs and social networks are playing an important role influencing everything from the products people buy to the presidential candidate they support. Thus, there is a need for a new type of search engine which will not only retrieve facts, but will also enable the retrieval of opinions. Such a search engine can be used in a number of diverse applications like product reviews to aggregating opinions on a political candidate or issue. This paper consist review works have been designed for opinion mining by using classification and ranking techniques.

Keywords : Sentiment Analysis, Opinion Mining, POS, Ranking Algorithm, Feature Selection Method, Semantic Orientation.

I. INTRODUCTION

Sentiment analysis, also called opinion mining, is the process of extracting subjective information, such as opinions, sentiments and attitudes in the source materials towards an entity. It is an interdisciplinary research field that combines tools and techniques from natural language processing, text mining and computational linguistics [1]. Opinions are important factors and influencers of the decision-making process. The determination of people's opinions toward a particular event can be extremely important in several fields, such as management sciences, political science, economics and other disciplines of social sciences [2]. The Web provides a rich and progressively expanding source of information to reach opinions/sentiments regarding a particular topic, product, event or individual. Sentiment analysis process can be modelled as a classification problem. Sentiment analysis can be conducted at different levels of detail. Based on the levels, sentiment analysis can be broadly divided into three main levels: document level, sentence-level and aspect-level sentiment analysis [3]. Document-level sentiment classification aims to determine the overall sentiment orientation of an entire document, such as a review text, assuming that each document contains information regarding a single entity. Sentence-level sentiment analysis aims to identify subjective and objective sentences. In sentence-level sentiment

analysis, the sentiment orientation of subjective sentences is also identified.[6] The classification of review documents at document or sentence level of granularity does not fully reveal the opinions regarding different features of a particular entity. Hence, aspect-level sentiment analysis concerns the classification of sentiments by focusing on the particular features/aspects of entities.

II. LITERATURE SURVEY

Blenty proposed algorithm works on word level. POS tagged words are extracted from pre-processing steps are used. Each word acts as feature. Algorithm is executed for total number of features provided in input dataset. According to this weight words are classified into two categories. Classified words are having POS tag, polarity tag, and weight value. From this list, common and uncommon words are picked and used for bipartite graph clustering.

Probability distribution of class c is calculated based on term frequency. Classification process predicts the polarity of the target domains lexicon from source domain. The clustering algorithm is applied on classified word lists and documents until it reaches convergence. All extracted words from source domain are tagged, and weight is calculated for each word using mutual information available for words. Target words are extracted and compared with the source. If they match then they will be categorized as domain independent, otherwise domain-specific. Domain-independent words are from both source and target domains; whereas, domain-specific are from target domain only. A graph is constructed between domain dependent and domain-independent words. Co-occurrence relationship between these words represents edge. Occurrence of domain-specific word along with domain-independent word means that both a related to each other and assign edge. Using domain-independent words weight is assigned to domain-specific words and classified accordingly. Each file form target domain is assigned score on which basis it is classified as positive or negative. Each word has weight assigned to it. Summation of weights of words in each sentence gives score to sentence. Then addition of all sentence score is nothing but score of file. On the basis of this, file is classified. Clustering helps in reducing mismatch between domain specific words of source and target domains. Two sets of lexicons are extracted as an output with polarity which is compared with SentiWordNet..[1]

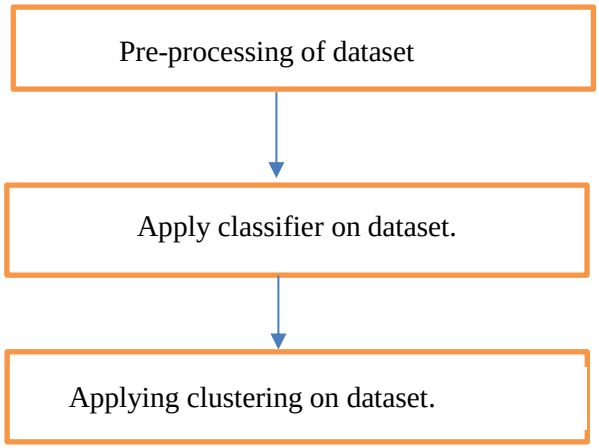

Fig 1 : major phases in entropy based classifier

Nyaung proposed system in which review crawler crawl the related web pages and retrieves review comments only. The filtered review comments will be proceeding for more processing steps. Subjective sentences express the reviewer's sentiment about the product and objective sentences do not have any direct or support of that sentiment. Subjectivity/objectivity determination works in two phases – training and classification. For training phase, manually labelled sentences are used as trained data, which is later used to identify subjective unigrams for new dataset. In the second phase, the classification is centred on the probability of unigrams from test dataset using the training data. In the pre-processing, only subjective sentences are submitted to a pipeline for Parts-Of-Speech (POS) tags. POS tagging is used for sentence splitting and to

assign lexical categories to the words in text. The feature and opinion pairs, the parser from previous step are analysed and generate all possible information components from them. The extracted pairs from dependency relation can contain which are relevant or not. Therefore, maximum entropy model is used to predict opinion-relevant product feature or not by choosing the class with the highest conditional probability p according to this model. Domain ontology to get the domain related features and to define the synonym set for features. Product feature candidates are identified by POS tags and only the features which are stored in the domain ontology are valid. Maximum Entropy modelling for prediction of feature and opinion pairs in a different way it has been used so far, using weights for both to emphasize the importance of each one of them in the relation of feature and opinion pairing task. Maximum entropy model is used to predict which feature word should be related with the opinion word with maximum probability.[3]

Table 1: Summery

Year	Author	Method	Limitation
2018	Blenty et al	Modified Naïve Byes Algorithm	It is not support unstructured keyword.
2018	Deshmukh et al	Extracts and classifies opinion words from one domain called source domain and predicts opinion words of another domain called target domain using a semi-supervised approach, Which combines modified maximum entropy and bipartite graph clustering.	It is not consider non-word feature
2017	Aytug et al	The presented framework obtains individual feature rankings with information gain, chi-square, gain ratio, symmetrical uncertainty, Pearson correlation coefficient, ReliefF algorithm and probabilistic significance measure-based feature selection methods. Then, these individual lists are amalgamated into a single ranking list via a genetic algorithm.	
2015	Dim En Nyaung et al	Polarity of extracted opinion word calculated with sentiword net.	Comparative sentence not included
2014	Zhang et al	Probabilistic aspect ranking algorithm.	It used only document level sentiment analysis.

Deshmukh et al proposed method collected data from site that provides with reviews on electronic gadgets and gives the respective user reviews. Pre-processing phase on data through Natural Language processing is done and then feature/aspect list is displayed. Extracting nouns as features and adjectives as opinions. Next step is Enhanced Aspect Identification and Opinion Word Extraction In this technique High frequency aspects are ranked and the corresponding opinion words for the aspects are extracted for identifying polarity of each aspect n finally the product itself. Modify Naivè Bayes model to create an aspect ranking algorithm which eventually will rank the products where this step calculates the weight of features from reviews. [2]

Zhang et al proposed system in which given the consumer reviews of a particular product, we first identify aspects in the reviews by a shallow dependency parser and then analyse consumer

opinions on these aspects via a sentiment classifier. System develop a probabilistic aspect ranking algorithm, which effectively exploits the aspect frequency as well as the influence of consumers' opinions given to each aspect over their overall opinions on the product in a unified probabilistic model. The overall opinion in a review is generated based on a weighted aggregation of the opinions on specific aspects, where the weights essentially measure the degree of importance of these aspects. A probabilistic regression algorithm is developed to infer the importance weights by incorporating aspect frequency and the associations between the overall opinion and the opinions on specific aspects. In order to evaluate the proposed product aspect ranking framework. A product aspect ranking framework to automatically identify the important aspects of products from numerous consumer reviews. A probabilistic aspect ranking algorithm to infer the importance of various aspects by simultaneously exploiting aspect frequency and the influence of consumers' opinions given to each aspect over their overall opinions on the product.[4].

Aytug at al evaluated Feature selection is the process of obtaining an appropriate feature subset from the data set so that the classification

Algorithms can deal efficiently with high-dimensional feature spaces. Feature selection methods aim to eliminate irrelevant or redundant features and to reduce the training time required to build a classification model Feature selection method can be broadly divided into two groups: filter-based and Wrapper-based feature selection methods.

Filter-based methods evaluate the usefulness of features based on heuristics/evaluation metrics, whereas wrapper-based methods select the features based on the performance of a machine learning algorithm to optimize the predictive performance. System evaluated the different individual filter-based measures like Information gain-based feature ranking is a filter-based feature selection method, which is widely utilized in text mining domain. The second is Chi-squared feature ranking evaluates the merit of each feature individually with the chi-squared statistical measure. Next is Information gain is a biased measure towards features with high values. Similar to gain ratio, the symmetrical uncertainty coefficient aims to eliminate the bias of information gain towards the attributes with higher values. Pearson correlation coefficient is used to measure the correlation between two attributes. Relief algorithm is a filter-based method that uses a feature relevance criterion to rank the features. Proposed genetic rank aggregation-based feature selection model uses Filter-based feature selection methods can obtain different rankings for the same data set. The aggregation of several different feature rankings may be beneficial to obtain an enhanced ranking of the features. The proposed method models the feature selection as a rank aggregation problem and the feature rankings obtained by different filter-based methods are combined. Given a set of rankings of the same candidate sets, the rank aggregation seeks to find a single better ranking from these multiple rankings. The rank aggregation problem can be modelled as an optimization problem where the objective is to obtain a final ranking that is the closest to the all individual ranking lists. [5]

III. CONCLUSION

This paper give brief information on Opinion Mining and the recent literature review putting light on the steps involved in Opinion Mining. Opinion Mining is the field of study that tracks the mood or attitude of people towards a particular product or topic. It helps to find the product with best quality, which is predicted based on the user reviews. It is also used for recommendation system and business intelligence. There are many challenges and future developments possible in Opinion Mining approach like short

length and irregular structure of the content such as named entity recognition, parsing. Opinion mining is to find the opinion of a person from sentences and then classify them on the basis of polarity. Opinion mining helps people to know the semantic orientation for a product by classifying it as positive, negative or neutral.

IV. REFERENCES

[1]. J. Deshmukh and A. Tripathy, "Entropy based classifier for cross-domain opinion mining", Applied Computing and Informatics, vol. 14, no. 1, pp. 55-64, 2018.

[2]. B. Alengadan and S. Khan, "Modified Aspect/Feature Based Opinion Mining for a Product Ranking System", 2018 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC), pp. 1-5, 2018.

[3]. D. Nyaung and T. Lai Thein, "Feature-Based Summarizing and Ranking from Customer Reviews", International Journal of Computer and Information Engineering, vol. 9, no. 3, pp. 734-739, 2015.

[4]. Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang and Tat-Seng Chua, "Product Aspect Ranking and Its Applications", IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 5, pp. 1211-1224, 2014.

[5]. A. Onan and S. Korukoglu, "A feature selection model based on genetic rank aggregation for text sentiment classification", Journal of Information Science, vol. 43, no. 1, pp. 25-38, 2016.

[6]. B. Liu, "Sentiment Analysis and Opinion Mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167, May 2012

Cite this article as :

Prof. Sonali D. Borase, Prof. Prasad P. Mahale, "Aspect Ranking Technique for Efficient Opinion Mining using Sentiment Analysis : Review ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 1, pp. 45-49, January-February 2019.
Available at doi :
https://doi.org/10.32628/CSEIT183812
Journal URL : http://ijsrcseit.com/CSEIT183812