Machine Learning Applications for Sentiment Analysis Aspect based Opinion Mining

Rajani Devi

Student, Department of Computer Science and Engineering

Baba Institute of Technology and Sciences, Visakhapatnam

A V S Pavan

Asst.Prof, Department of Computer Science and Engineering

Baba Institute of Technology and Sciences, Visakhapatnam

Abstract- Presently multi day's online consumer review is most incredible asset for decision making. This term fill in as electronic verbal (EWM) which turn into an inexorably famous. A large number of individuals are currently purchase items and services by means of online. Web services are given element to clients straightforwardly. The web can gives a broad wellspring of consumer reviews. The client can peruse every one of the reviews and assess reasonable view about item or administration. This can material just for predetermined number of reviews displayed on web. The web contain more than many reviews then issue arrived and tedious too. A content preparing system is alluring which abridge every one of the reviews. This system would be discover general viewpoint classification tended to in all review sentences. The technique introduced in this system which applies affiliation rule mining on co-event recurrence information to discover these angle classifications. From this outcome, produce extremity score for every perspective class. This extremity score assesses reasonable decision making for customer and additionally organization.

Keywords— sentiment analysis, polarity, aspect, opinion mining, Feature extraction, machine learning technique,

I. INTRODUCTION

Information mining is the way toward finding designs in colossal informational collections including strategies at the crossing point of machine learning, insights, and database frameworks. Information mining is an interdisciplinary subfield of software engineering with a general question mine data (with insightful strategies) from an informational index and change the data into an understandable structure for further use. Information mining is the investigation venture of the "learning revelation in databases" process, or KDD. Beside the crude analysis step, it likewise includes database and information the executive's viewpoint, information preprocessing, model and induction generosity, intriguing quality measurements, intricacy thought, post-preparing of found structures, representation, and online refreshing. The expression "information mining" is in reality a misnomer, in light of the fact that the point is the extraction of examples and learning from a lot of information, not simply the extraction (mining) of information. It likewise is a popular expression and is regularly connected to any type of huge scale information or data preparing (gathering, extraction, warehousing, analysis, and insights) and additionally any case of PC decision emotionally supportive network, and also man-made consciousness (e.g., machine learning) and business knowledge. The genuine information mining errand is the self-loader or routine analysis of huge amounts

of information to extricate beforehand obscure, intriguing examples, for example, gatherings of information records (group analysis), unordinary records (peculiarity location), and reliance (affiliation rule mining, consecutive example mining). These typically include utilizing database system, for example, spatial lists. These examples would then be able to be viewed as a sort of rundown of the info information, and might be utilized in further analysis or, for instance, in machine learning and prescient investigation. For instance, the information mining step may characterize different gatherings in the information, which would then be able to be utilized to get progressively precise expectation results by a decision emotionally supportive network. Neither the information gathering, information planning, nor result comprehension and announcing is a piece of the information mining step, yet do have a place with the general KDD process as extra advances. While the dictionary based methodology does not require any earlier preparing so as to mine the information. It utilizes a predefined rundown of words, where each word is related with an explicit sentiment. At long last in the cross breed approach, the mix of both the machine learning and the vocabulary based methodologies can possibly enhance the sentiment order execution. On considering the apparatuses utilized for sentiments analysis, the most utilized devices for recognizing the sentiments extremity are Emoticons, LIWC, SentiStrengh, Senti WordNet, SenticNet, Happiness Index, AFINN, PANAS-t, Sentiment140, NRC, EWGA and FRN. Sentiment analysis is utilized essentially in various fields, for example, advertising, political and sociological. In showcasing field organizations use it to build up their methodologies, to comprehend customers' sentiments towards items or brand how individuals react to their battles or item dispatches and why consumers don't get a few items. In political field, it is utilized to track of political view, to recognize consistency and irregularity among proclamations and activities at the administration level; it very well may be utilized to anticipate race results. Sentiment analysis additionally is utilized to screen and investigate social wonders, for the spotting of conceivably hazardous circumstances and deciding the general state of mind of the blogosphere. The sentiment analysis at that point speaks to an imperative component for any subject (strategy producers, partners, organizations and so on.) to perform various types of exercises, for example, foresee budgetary execution [5], comprehend consumers' discernment [6] give early admonitions [7; 8], characterize decision results and so forth. In these precedents, the sentiment input is whether a given consumer supposition has negative, positive or unbiased extremity with respect to the distinctive focus of intrigue [9]. The expansive measure of these substance required the utilization of mechanized methods for breaking down since physically it is preposterous. As indicated by [10], analysts have discovered approaches to keep away from the utilization of manual explanation by using existing online literary substance produced from destinations, for example, Epinion, Amazon, Rotten Tomatoes, Twitter, and Facebook.

II. RELATED WORK

In [1], Bing Liu et al. depicted a model where the errand of highlight based supposition synopsis is performed by first mining the item includes that have been remarked on by customers utilizing affiliation mining method, at that point recognizing sentiment sentences in each review and choosing whether every conclusion sentence is certain or negative utilizing a lot of seed modifiers alongside their introductions that becomes later utilizing WorldNet lastly condensing the outcomes. In [2] T. Ahmad built up a supposition mining framework where the highlights and

conclusions are separated utilizing semantic and etymological analysis of content records; the extremity of the assessment sentences is found utilizing extremity scores given by SentiWordNet and the created synopsis is exhibited utilizing a perception module comprehendibly. In [4], W. Zhang et al. built up a framework considered Weakness Finder that encourages the makers to discover their item shortcoming from Chinese reviews by utilizing angle based sentiment analysis. The framework concentrates and gathering express highlights by utilizing Morpheme based technique and How net based comparability measure. Next it recognizes and bunches verifiable highlights with collocation choice strategy for every viewpoint. At last the extremity is controlled by sentence based sentiment analysis strategy. in [5], A. Dengel exhibited an extractive perspective based sentiment synopsis framework which comprise of a viewpoint identifier for highlight extraction that happens as often as possible, a bunching module to bunch every one of the records that have the event of same angle word inside them in one gathering, a crossover extremity recognition framework alongside their produced list of capabilities for deciding assessment introduction and a literary and graphical outline generator module which utilizes an unsupervised extremity discovery and positioning calculation created by them for rundown age. In [7], R. Kumar gave a technique to mine distinctive item highlights and sentiment words dependent on customer feeling communicated in the review utilizing a semantic put together methodology based with respect to composed reliance relations. In [9], M. Dalal displayed a semi-directed methodology for mining online client reviews to produce relative component based factual synopses. It incorporates stages like preprocessing, highlight extraction, trailed by sentiment grouping and rundown. They performed essential cleaning errands like sentence limit identification and spell-blunder rectification in the preprocessing stage. At that point in the wake of performing POS labeling utilizing Link Grammar Parser, often happening things (N) and thing phrases (NP) are considered as the conceivable supposition highlights dependent on multiword approach which are separated alongside the related descriptors portraying them, as indictors of their feeling introduction. When highlights and assessment words are extricated, the sentiment extremity of the suppositions is resolved utilizing SentiWordNet. In [10], D. Wang et al. created Sum View, a Web-based review outline framework, to consequently separate the most delegate articulations and customer feelings in the reviews on different item includes. The framework centers around conveying the greater part of data contained in the review records by choosing the most delegate review sentences for each removed item highlight.

III. DOCUMENT SENTIMENT CLASSIFICATION

This approach we consider the whole document as a single entity and classify the sentiment as positive or negative or neutral. Document-level sentiment classification assumes that the opinionated document expresses opinions on a single target and the opinions belong to a single person or product or service. It is clear that this assumption is true for customer reviews of products documents which usually focus on one product and single reviewer writes it.

There are two type of classification techniques that have been used in document-level sentiment classification such as supervised method and unsupervised method.

1) Supervised Method

Sentiment classification task can be spontaneous as a supervised learning problem with two classes, positive and negative. Product reviews are used as training and testing data. Another method such as online reviews have rating scores assigned by their reviewers, e.g., 1-5 stars, the positive and negative classes are determined using the ratings. A review with 4 or 5 stars is considered a positive review, and a review with 1 to 2 stars is considered a negative review.

A training set is used by an automatic classifier to learn the differentiating characteristics of documents, and a test set is used to check how well the classifier performs. The most commonly used features in sentiment classification are introduced below

1.1. Terms and their frequency:

The features can be considered as individual words termed as unigram and their n-grams with associated frequency counts. It can be calculated as weightage form.

1.2. Part of speech (POS):

POS information is a very important indicator of sentiment expression.

1.3. Sentiment words and phrases:

Sentiment words and phrases that is express in positive and negative sentiments. For example good, fantastic, amazing words are positive sentiment and bad, boring, slow, worst and poor are negative sentiment. For example rubbish (noun), hate and like (verb) can indicate opinion in some documents.

1.4. Negations:

Negation words are very important to evaluate the polarity of a sentence because they can transform the sentiment orientation in a sentence. For instance, the sentence “I don’t like this phone” has negative orientation.

1.5. Syntactic dependency:

It is represented as a parse tree and it contains word dependency based features.

SUPERVISED LEARNING TECHNIQUES:

Supervised learning techniques can be used to sentiment classification, such as

- Naive Bayes Classification (NB)
- Maximum Entropy Rule (ME)
- Support Vector Machine (SVM)

In Fig we can analyses the supervised learning techniques. In research works which is used as supervised method to solve sentiment classification problem. In these papers, we used three machine learning techniques - Naïve Bayes, Maximum Entropy classification (ME), and SVM to classify sentiment of product review documents. They test

several features to find Unigrams, Bigrams, Parts of Speech (POS) and position of words were used as features in these techniques. The results show that the best performance is achieved is used in SVM classifier.

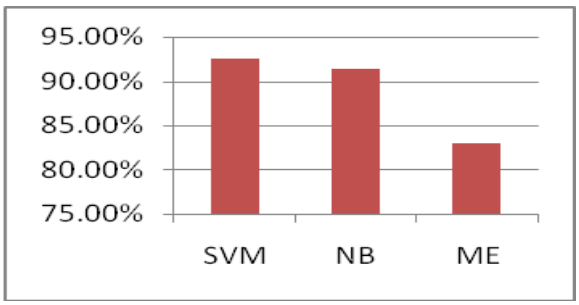

FIG SUMMARY OF OPINION MINING TECHNIQUES

2) Unsupervised Method:

The Unsupervised learning method has a set of inputs, like clustering, labels are not known during training. Classification is performed using some fixed viti patterns which are used to express opinions. The partof-speech (POS) tags are used to compose viti patterns. The unsupervised learning can be done by extracting the opinion words inside a document .The point-wise mutual information can be made use of to find the semantics of the extracted words

E. Sentence-Level Sentiment Classification:

In the sentence level sentiment analysis, the polarity of each sentence is calculated. The same document level classification methods can be applied to the sentence level classification problem. Objective and subjective sentences must be found out. The subjective sentences contain opinion words which help in determining the sentiment about the entity. After which the polarity classification is done into positive and negative classes.

Knowing that a sentence is positive or negative is of lesser use than knowing the polarity of a particular feature of a product. The advantage of sentence level analysis lies in the subjectivity/ objectivity classification.

F. Sentiment Lexicon Construction:

Sentiment lexicon words are identified by opinion words are also known as opinion-bearing words. Sentiment words are always divided into two categories such as: positive or negative. Positive opinion words are used to express some desired states while negative opinion words are used to express some undesired states.

There are three methods to construct a sentiment lexicon: manually construction, corpus-based methods

and dictionary-based methods.

1) Manually Construction:

The manual construction of sentiment lexicon is a very hard and time-consuming task and always cannot be used alone but it can be combined with other methods to improve the accuracy of these methods.

IV. SENTIMENT CLASSIFICATION TECHNIQUES

Sentiment classification techniques can be roughly divided into machine learning approach, lexicon based approach and hybrid approach [5]. The *Machine Learning Approach (ML)* applies the famous ML algorithms and uses linguistic features. The *Lexicon-based Approach* relies on a sentiment lexicon, a collection of known and precompiled sentiment terms. It is divided into dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity. The *Hybrid Approach* combines both approaches and is very common with sentiment lexicons playing a key role in the majority of methods.

The text classification methods using *Machine learning* approach can be divided into supervised and unsupervised learning methods. The supervised methods use a large number of labelled training documents. The unsupervised methods are used when it is difficult to find these labelled training documents.

The lexicon-based approach depends on finding the opinion lexicon which is used to analyze the text. The dictionary based approach which depends on finding opinion seed words, and then searches the dictionary of their synonyms and antonyms. The corpus-based approach starts with a seed list of opinion words, and then finds other opinion words in a large corpus to help in finding opinion words with context specific orientations. This could be done by using statistical or semantic methods.

A. *Machine Learning*

Machine learning approach relies on the famous Machine Learning algorithms to solve the sentiment analysis as a regular text classification problem that makes use of syntactic and/or linguistic features.

Text Classification Problem Definition: We have a set of training records = { , ,…. ., } where each record is labeled to a class. The hard classification problem is when one label is assigned to an instance. The soft classification problem is when a probabilistic value of labels is assigned to an instance.

I.A.1 *Supervised Learning*

The supervised learning methods depend on the existence of labelled training documents. There are many supervised classifiers in literature. Next subsection, we present it with brief details about some of the most frequently used classifier in Sentiment Analysis.

I.A.1.1 *Probabilistic Classifiers*

Probabilistic classifiers use mixture models for classification. The mixture model assumes that each class is a component of the mixture. These kind of classifiers are also called generative classifiers because each mixture component is a generative model that provides the probability of sampling a particular term for that component. Three of the most famous probabilistic classifiers are discussed in next subsections.

I.A.1.1.1 *Naive Bayes Classifier(NB)*

Naïve Byes classifier is the simplest and most commonly used classifier. Naïve Bayes classification model computes the posterior probability of a class based on the distribution of the words in the document. It uses Bayes Theorem to predict the probability that a given feature set belongs to a particular label.

$$(l) = \frac{(\quad) * (\quad | \quad)}{(\quad)} \qquad (1)$$

P(label) is the prior probability of a label or the likelihood that a random feature set label. *P(features|label)* is the prior probability that a given feature set is being classified as a label. *P(feature)* is the prior probability that a given feature set is occurred. Given the Naïve assumption which states that all features are independent, the equation could be rewritten as follows:

$$(l) = \frac{(\quad) * (\quad | \quad) * \ldots\ldots * (\quad | \quad)}{(\quad)} \qquad (2)$$

An improved NB classifier was proposed by Kang and Yoo [2] to solve the problem of tendency for the positive classification accuracy to appear up to approximately 10% higher than the negative classification accuracy. They showed that using this algorithm with restaurant reviews narrowed the gap between the positive accuracy and the negative accuracy compared to NB and SVM. The accuracy is improved in recall and precision compared to both NB and SVM.

I.A.1.1.2 *Bayesian Network(NB)*

The main assumption of the NB classifier is the independence of features. The other extreme assumption is to assume that all the features are fully dependent. This leads to the Bayesian Network model which is a directed acyclic graph whose nodes represent random variables, and edges represent conditional dependencies. BN is considered a complete model for the variables and their relationships. In text mining, the computation complexity of BN is very expensive; that is why, it is not frequently used [1].

BN was used by Hernandez and Rodriguez [3] to consider a real-world problem in which the attitude of author is characterized by three different but related target variables. They propose the use of multi-dimensional Bayesian network classifiers. It joined the different target variables in the same classification task in order to exploit the potential relationships between them. They showed that their semi-supervised multi-dimensional approach out performs the most common SA approaches, and that their classifier is the best solution in a semi-supervised framework because it matches the actual underlying domain structure.

I.A.1.2 *Linear Classifiers*

Given = { …. . } is the normalized document word frequency, vector $\bar{}$= { …. . } is a vector of linear coefficients with the same dimensionality as the feature space, and b is a scalar; the output of the linear predictor is defined as =$\bar{}$. + , which is the output of the *linear classifier*. The prediction p is a separating hyper plane between different classes. There are many kinds of linear classifiers; among them is *Support Vector Machines (SVM)* [4,6] which is form of

classifiers that attempt to determine good linear separators between classes. Two of most famous linear classifiers are discussed in the following subsections.

I.A.1.2.1 *Support Vector Machines Classifiers (SVM).*

The main principle of SVM is to determine linear separators in the search space which can best separate the different classes. In Fig. there are 2 classes *x*, *o* and there are 3 hyperplanes A, B and C. Hyperplane is provide the best separation between the classes, because the normal distance of any of the data points is the largest, so it represents the maximum margin of separation.

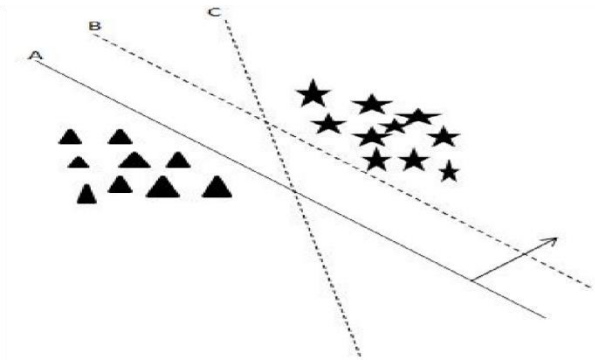

Fig: Using support vector machine on a classification problem.

I.A.2 *Weakly, Semi and Unsupervised Learning*

The main goal of the classification is to classify documents into number of categories. In large number of labeled training document, it is difficult to create labeled training document, but easy to collect the unlabeled documents. To solve this problem unsupervised learning methods are used. Koa and Seo present the research work in this field and they propose a method that divides document into sentences, and categorized each sentence using keyword lists of each category and sentence similarity measure [1].

B. *Lexicon Approach*

Opinion words are divided in many categories. Positive opinion words are used to express some necessary things, and negative opinion words are used to describe unnecessary things. Opinion phrases and idioms are also there which together are called *opinion lexicon.* To collect the opinion word list there are three main methods. One of them is Manual method not used alone and which is very time consuming.

The basic steps of the lexicon based techniques are outlined below [2]:

1. Preprocess each text (i.e. remove noisy characters and HTML tags)
2. Initialize the total text sentiment score: s<-0.
3. Tokenize text. For each token, check if it is present in a sentiment dictionary.

(a). If token is present in dictionary,

I. If token is positive , then s<-s+w. II. If token is negative, then s<-s-w.

4. Look at total text sentiment score s,

(a). If s> threshold, then classify the text as positive.

(b). If s< threshold, then classify the text as negative.

I.B.1 *Dictionary Based Approach*

Manually collect a small set of opinion words and the main strategy of dictionary-based method is presented in [11, 14]. Then, this set is grown by finding their synonyms and antonyms in the WordNet [13] and thesaurus [15]. After found new words these words are added to the seed list and the next process starts. This process stops when no new words are found. To remove or correct the errors manual inspection process will be done.

Disadvantage: This method cannot find the opinion words with domain and context specification orientations are the major disadvantage.

I.B.2 *Corpus Based Approach*

It is hard to prepare a huge corpus to cover all English words so it is not effective as the dictionary-based method when it used alone. But it can help to find domain and context specific opinion words using a domain corpus is the huge advantage of this method. The corpus-based approach is performed in statistical approach or semantic approach.

I.B.2.1 Statistical Approach

Statistical techniques find the co-occurrence patterns or seed opinion words. It can be done by obtaining posterior polarities in corpus, as proposed by Fahrni and Klenner [7]. By using the entire set of indexed document it is possible to solve the problem of the unavailability of some words [6].

The word has positive polarity if it occurs more frequently in positive texts, or its polarity is negative if it occurs more and more time in negative texts. If it has same occurrence, then it is neutral word. So, the polarity of word can be identified by analyzing the occurrence frequency [9].

I.B.2.2 Semantic Approach

For computing the similarity between words this method gives sentiment values directly and depends on different principles. For semantically close words this principle gives similar semantic values. By using relative count of positive and negative synonyms of this word determine the sentiment polarity of an unknown word [8].

To build a lexicon model for the description of verbs, nouns and adjectives to be used in SA and also in many other applications. The detailed subjectivity relations among the actors in a sentence expressing separate attitudes for each

actor described by Maks and Vossen model [3]. Their results showed that the sometimes actor's subjectivity and speaker's subjectivity can be constantly distinguished.

To perform sentiment analysis task semantic methods can be mixed with the statistical methods. As per the Zhang and Xu [20] to find the product weakness from online reviews they used both methods. To find the frequent and infrequent denotative features they used Hownet-based similarity measure. By applying semantic methods they have grouped products feature words appropriate aspects. They took the impact of adverbs of degree and they have utilized sentencebased sentiment analysis method.

V. PROPOSED METHODOLOGY

The proposed research is in the area of sentiment analysis to determine whether a document or a sentence express positive or negative sentiments [1]. With the development of the internet, we find that massive reviews are practically sentence level, appearing in wechat, blogs and ecommerce. Opinion mining that is based on sentence level can get the sentiment expressed in review contents. A new method which takes into the consideration of not only opinion words but also context information [7].

Earlier works that focus on implicit feature extraction are where implicit features are found using semantic association analysis on point wise mutual information which uses co-occurrence association rule mining to link opinion words as antecedents to implicit features as consequents[2].

At sentence level aspect based sentiment analysis, generally a review consists of several sentences and one single sentence may contains mixed opinion tuples to words the differ OTE (Opinion Target Expression) and Entity # (null) Attribute (A) [E#A]. Therefore, in order to extract features from the relevant fragment, we proposed a two-step method to acquire potential words related to given aspect as pending words for future feature extraction.

1. Segmentation step:

- It is to split each sentence into several fragments.

2. Selection step:

- It selects out one more fragments from sentence for each aspects.

By Blair-Goldensohnet in 2008, assumed that the product class is known in advance. Their algorithm detects whether a noun or noun phrase is a product feature by computing the point wise mutual information between the noun phrase and products class or service class. User feedback to LDA as a response variable related to each document with proposed semi-supervised model [4]. The propose work is to provide abstract of more number of customer review of online business based n aspects summarizing the review not only helpful for buyer, but also important for manufacturer and sellers. Online review system is becoming very useful and serving as vital source of information for people [6].

LDA means Latent Dirichlet Allocation.LDA is a technique that automatically discover topics that respected Documents contain. Keyword Extracted using following Parameters:

- Frequency
- Location within document
- Co-occurrence with other words.

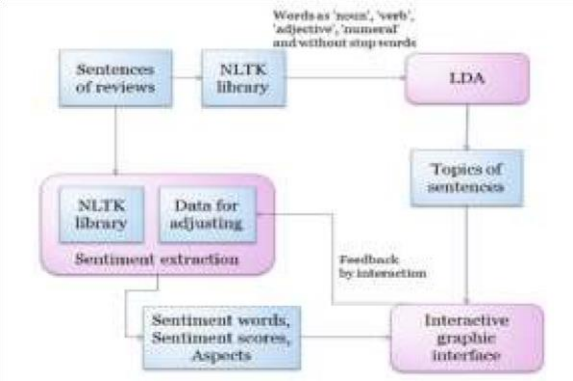

Fig. Proposed Architecture Diagram

VI. CONCLUSION

Most existing techniques for document level sentiment classification are based on supervised learning, for Example, n-gram features, Naïve Bayes, Maximum Entropy classification, and SVM can be applied to achieve the high performance. The results produced via machine learning techniques are quite better in comparison to the baselines Naïve Bayes tends to do the worst and SVMs tend to do the best, although the differences are not very large. It is clear that although we may be able to build comprehensive lexicons of sentiment-annotated words, it is still a challenge to accurately locate it in text. Few studies have been done outside the realm of short documents like product reviews, and especially in difficult domains like political commentaries. This is true partially because there is little annotated data available for realms outside reviews. The Method called syntax tree pruning and tree kernel based approach to sentiment classification to sentence-level sentiment classification can be applied. If the opinion words are included in dictionary then it must contain all words. It is important for lexicon based approach. Because it will reduce the performance if there are fewer words present in dictionary. Another significant challenge to this approach is that the polarity of many words is domain and context dependent. For example, ‘funny movie’ is positive in movie domain and ‘funny taste’ is negative in food domain. Such words are associated with sentiment in a particular domain.

REFERENCES

[1] M. Hu and B. Liu, "Mining Opinion Features in Customer Reviews," in Proceedings of the 19th National Conference on Artifical Intelligence (AAAI 2004). AAAI, 2004, pp. 755–760. "Mining and Summarizing Customer Reviews," in Proceedings of 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004). ACM, 2004, pp. 168–177.

[2] C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, H. Ng, and C. Jin, "Red Opal: Product-Feature Scoring from Reviews," in Proceedings of the 8th ACM Conference on Electronic Commerce (EC 2007). ACM, 2007, pp. 182–191.

[3] Z. Li, M. Zhang, S. Ma, B. Zhou, and Y. Sun, "Automatic Extraction for Product Feature Words from Comments on the Web," in Proceedings of the 5th Asia Information Retrieval Symposium on Information Retrieval Technology (AIRS 2009). Springer, 2009, pp. 112–123.

[4] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "Semeval- 2014 task 4: Aspect based sentiment analysis," in Proceedings of the 8thInternational Workshop on Semantic Evaluation (SemEval 2014). Association for Computational Linguistics and Dublin City University, 2014, pp. 27–35.

[5] Y. Zhao, B. Qin, S. Hu, and T. Liu, "Generalizing Syntactic Structures for Product Attribute Candidate Extraction," in Proceedings of the Conference of the North American Chapter ofthe Association of Computational Linguistics: Human Language Technologies 2010 (HLT-NAACL 2010). ACL, 2010, pp. 377– 380.

[6] G. Qiu, B. Liu, J. Bu, and C. Chen, "Expanding Domain Sentiment Lexicon Through Double Propagation," in Proceedings of the 21st International Jont Conference on Artifical Intelligence (IJCAI 2009). Morgan Kaufmann Publishers Inc., 2009, pp.1199–1204.

[7] L. Zhang, B. Liu, S. H. Lim, and E. O"Brien-Strain, "Extracting and Ranking Product Features in Opinion Documents," in Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010). ACL, 2010, pp. 1462–1470.

[8] G. Qiu, B. Liu, J. Bu, and C. Chen, "Opinion Word Expansion and Target Extraction through Double Propagation," Computational Linguistics, vol. 37, no. 1, pp. 9–27, 2011.

[9] N. Jakob and I. Gurevych, "Extracting Opinion Targets in a Single- and Cross-Domain Setting with Conditional Random Fields," in Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2010). ACL, 2010, pp. 1035–1045.

[10] L. Zhuang, F. Jing, and X.-Y. Zhu, "Movie Review Mining and Summarization," in Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM 2006), vol. 6, no. 11. ACM, 2006, pp. 43–50.