A FUZZY APPROACH FOR SENTIMENT ANALYSIS BASED ON WEB CUSTOMER REVIEWS

1N.S.AMBEKAR, 2N.L.BHALE

1Department of Computer Engineering, Matoshri College of Engineering and Research Center, Nashik, India.
2Department of Information Technology, Matoshri College of Engineering and Research Center, Nashik, India.
E-mail: ambekarneelima@gmail.com, nbhale@gmail.com

Abstract - Sentiment analysis has become an important research area in today’s competitive business world. Customers now no longer ask friends or relatives for their suggestions to buy a product; everything is available just at a click on the World Wide Web. Opinions expressed in the user comments on the have become an important source for analysis e.g., customer reviews of products, ratings, posts, and blogs. In this paper, we focus on customer reviews of products. Specifically, we study the sentiments/opinions expressed (positive, negative or neutral) in the customer reviews. Hereby, we propose a new method i.e. fuzzy ontology tree for giving a more clear and precise analysis of sentiments to give best quality poll to customers. This approach allows the system to handle opinion words that are context dependent, which cause major difficulties for existing algorithms. In the reviews, a single sentence may exhibit multiple aspects for opinions. Bootstrapping algorithm is used to handle this problem. The proposed method doesn’t require labelled data and hence easy to implement.

Keywords: Fuzzy ontology tree; Opinion polling; Opinion mining; Sentiment analysis.

I. INTRODUCTION

Internet is an important platform for expressing ideas or views in the form of reviews, posts, blogs etc. Opinion mining (Sentiment Analysis) is a computational study of opinions, sentiments, subjectivity, evaluations, attitudes, appraisal, affects, views, emotions, etc., expressed in text. Sentiment analysis is more widely used in industry as well as in academia. “Opinions” are key influencers of our behaviors as whenever we need to make a decision, we often find out the opinions of others. In the past, Individuals seek opinions from friends and family while Organizations use surveys, focus groups, opinion polls, consultants. Organizations spend a huge amount of money to find consumer opinions using consultants, surveys and focus groups, etc. Many websites such as Amazon, IMDb etc provoke customer to post reviews on products, features, services etc. In order to enhance business growth, customer satisfaction in the form of these reviews is discovered which was either done traditionally by designing questions that are to be answered or now-a-days as free-form, on the spot customer feedback. Such structured survey method was useful for only small data but with the explosion of social media structured survey technique suffered with the drawbacks like expenses in designing questions, lack of participation of customers. Hence, in this paper we focus on sentiment analysis based on unstructured spontaneous free-form customer reviews. The uneven and dynamic nature of the web has lead to the research in the field of opinion mining or sentiment analysis. Many recent studies have put forward sentiment-based or opinion-oriented summarization [5], [6], [13], [8]. The basic approach of sentiment analysis is to collect, analyze and extract sentiments or opinions from the customer reviews and produce the summarized results in the form of an opinion poll.

Supervised document classification algorithms were used in some previous studies [12], [19], [16] to analyze customer reviews in text form. However, this approach suffered with the drawbacks (1) It required labelled training data which was costly and time consuming,(2) Opinion poll generated may not be in proper meaningful format. For ex, to analyze a customer review on a mobile” Mobile battery life is not so good”. Here the review expresses a negative opinion on the battery aspect. Therefore, aspect-based opinion analysis techniques [18] are needed. Previous work feature based sentiment summarization [5], [6] that addressed aspect-based opinion polling is implemented on the sentence level instead of the document level, and does not address the polarity conflicting issue of document-level aspect-based opinion analysis. Implicit aspect expression problem, multiword terms problem couldn’t be solved by this technique. Customers may express different opinions on multiple aspects in the single review statement. For example, for a mobile review “Mobile has very attractive color and body but it has less memory storage” contains two aspects i.e. positive body aspect as well as negative memory aspect. To deal with such multiple aspect sentences and segment them into single multiple aspects, aspect-based opinion polling method is used [22]. This approach worked at sub-sentence level but failed at word level and seeds were manually selected which is troublesome.

To overcome these limitations, this paper presents an automatic sentiment classification problem using fuzzy ontology tree method. This technique is used to
generate a opinion poll from unlabeled free-form textual customer mobile reviews.

II. RELATED WORK

Most of the research work has been carried out for opinion mining or sentiment analysis for summarizing the sentiments from customer reviews [13], [18], [9], [14], [5], [6]. Recent research work focused upon product reviews for aspect/feature based opinion analysis. Segmentation of text is an issue in retrieval of information. Previous studies had focused on text segmentation at document level instead of the sentence level. Feature-based classification and sentence summarization [5], [6] revealed the results in form of sentiment summary as per the feature of the product. B. Pang, Lee et.al [18] proposed a technique to analyse textual reviews to predict polarities using supervised document classification algorithms. M. Hu and B. Liu [7] proposed a set of techniques for mining and summarizing product reviews based on data mining and natural language processing methods. They focused only on explicit features but couldn’t deal with implicit aspect identification. Stemming and fuzzy matching techniques used by them failed to deal the multiple aspect sentences and document-level polarity conflict problems. B. Pang and L. Lee, [10] focused on supervised or semi-supervised learning techniques for sentiment analysis which need labelled data for training.

Titov and McDonald [18] proposed a statistical model for sentiment analysis based on the labelled customer reviews; customers manually entered the ratings for different aspects. J. Zhu, H. Wang, et.al, [21] in their work titled Multi-Aspect Opinion Polling from Textual Reviews, presented an unsupervised approach to aspect-based opinion polling from raw textual reviews without explicit ratings. Jingbo Zhu, et.al, [22] proposed opinion polling method that does not require labelled training data. They proposed an MAS model to segment a multi-aspect sentence into multiple single-aspect units for aspect-based opinion polling. To overcome the challenges we propose a new approach for sentiment analysis using fuzzy ontology tree method to give more precise opinion poll.

III. DESIGN OF OPINION POLL GENERATION MODEL BASED ON FOT

In the proposed model, initially data is collected in form of the Customer reviews from the World Wide Web. These reviews are collected from mixed domains hence there is a need for separating the data using clustering technique. Separation is followed by conversion of Web Customer textual reviews to extract meaningful multiword terms from unlabeled reviews. This is followed by pre-processing of documents and Conceptual Analysis of the documents using Multi-Aspect Bootstrapping (MAB). Lastly the FOT technique used for segmentation and tree generation to generate a precise opinion poll.

3.1 Problem Definition:

This paper explores the problem of sentiment-based opinion polling from unlabeled free-form textual customer reviews and dynamically analyze them using fuzzy ontology method. To improve and implement a quality Sentiment analysis in customer reviews and to automatically construct fuzzy ontology tree (FOT) based on the product reviews, including the extraction of sentiment words, product features and the relations among features. Ontology consists of a hierarchical description of important classes (or concepts) in a particular domain, along with the description of the properties (of the instances) of each concept. The fuzzy ontology tree technique is based on fuzzy sets and semantic relations which can deal with the dynamic nature of web for sentiment analysis. The main goal is to construct a tree model to reveal the semantic relation between product features and sentiment words. The product features (aspects) mean product components for ex. battery, camera, memory etc, and attributes such as color, screen size, keyboard etc. Sentiments or opinions may be positive or negative such as attractive, costly, and good and so on.

Let Y= y1, y2, ...... yn be the input set of customer review, our goal is to find the final opinion poll by calculating score value S(FxW(Y)). For every pair (f, c) ∈ F × C, we can find score as follows,

\[ S(f,c|Y) = \frac{\sum_{i=1}^{n} y_i \times s_{yi}}{n} \]

Where S(f, c|yi) is the voting score for the pair (f, c) stated by the review yi, that can be defined as follows

\[ S(f, c|yi) = \begin{cases} 1 & \text{if } yi \text{ expresses polarity on feature } f \\ 0 & \text{otherwise} \end{cases} \]

For example, for a review statement y = “Mobile has a good appearance, battery life is not so good but is cheap in cost”. This review statement expresses...
negative sentiment for the aspect battery so, S(battery,negative)=1. S(battery,positive) = 0 and S(battery,neutral) = 0. Thus its important to calculate S(F,C) for individual review y. But a review yi may have multiple features / aspects such as keyboard, we have to analyze every review statement yi for different aspects and determine the polarity for each one. Finally determine the total polarity of review yi for feature f.

3.2 Sentiment Analysis Model Description:
Previous studies on sentiment analysis [1][2],used features or aspects in their study. In our proposed work, we focus on aspect related terms for aspect identification such as noun, verbs etc.

3.2.1. Feature related terms identification:
We firstly extract useful multi word terms from unlabelled data by using the Cvalue method [3] which takes as input a review data set and produces list of multi-word terms. 

Cvalue score of a multi-word term t can be calculated by:
If t is not contained by any other terms, Cvalue(t) = \[\log(|t|) \times freq(t)\], Otherwise, 
\[\text{Cvalue}(t) = \log(|t|) \times \frac{1}{n(L)} \times \sum_{i=1}^{n} \text{freq}(t_i)\]
where |t| denotes the number of words contained by t, \(freq(t)\) indicates the frequency of occurrence of t in the corpus, L is the set of multi-word terms containing t, and \(n(L)\) denotes the number of terms in AS.

3.2.2. Multi-feature extraction using bootstrapping:
A bootstrapping method is used for learning using some seed aspect related terms for each aspect with the assistance of free form unlabeled data. Bootstrapping is same as that of iterative clustering where in each learning cycle the most valuable candidate is chosen to be added to the present seed set, and the learning procedure continues until the predefined stopping criterion is satisfied. We use the RlogF metric [4] to find the value of each candidate aspect related terms t by

\[\text{RlogF}(t) = \log(freq(t,T)) \times R(t,T)\]  
(1)

where T is the current seed set, \(freq(t,T)\) is the frequency of occurrence of t and T within a limited context (i.e., k words to left or right of t), \(freq(t)\) is the frequency of co-occurrence of t in the corpus, and

\[R(t,T) = freq(t,T)/freq(t)\]  
(2)

We assign a score to each learned FRT that is important for the aspect identification. This score represent the degree of its ability of reflecting the corresponding aspect. The score \(BS_i(t)\) [22] of an FRT t for the \(i^{th}\) aspect can be measured by means of a rank function as,

\[BS_i(t) = \frac{R(t,T)}{\text{amb}(t)}\]  
(3)

Where \(AS=[t_{i1}, t_{i2}... t_{ik}]\) is the feature related term set of the \(i^{th}\) aspect produced by BS. Notice that \(t_{ik}\) is learned in the \(i_{th}\) iteration, \(|AS_i|\) indicates the number of aspect related term in \(AS_i\), and \(rf(t)\) represents the rank of t in ASi indicating in which iteration it was learned. A higher \(BS_i(t)\) value indicates that t is a more important term for the \(i_{th}\) aspect. Generally we apply the bootstrapping method to learn an aspect related term set for each aspect separately.

The ambiguity degree \(\text{amb}(t)\) [22] of a multi-aspect aspect related term t can be calculated. When aspect related terms appear for more than two aspects then we use multi-aspect bootstrapping method.

\[BM_i(t) = BS_i \times (1-\text{amb}(t))\]  
(4)

Algorithm 1: Bootstrapping Method to learn different features

Input: initial feature seed sets \(AS=[S1, S2, ..., Sm]\) for m aspects, and collected free form reviews I

Step 1: Extraction of feature related terms
Extract nouns, verbs, adjectives, adverbs and top-n multi-word terms using C-value method from I to form a candidate Feature Related Term set \(\mu\) for bootstrapping.

Step 2: Bootstrapping Learning
Proceed learning with the seed set ASi for the \(i_{th}\) aspect;

Repeat
1. Use Equation (1) to calculate RlogF score of each candidate in \(\mu\) ;
2. Select the candidate with the highest RlogF score to be added to ASi, and remove it from \(\mu\) ;
Until the predefined stopping criteria is met.

Step 3: Recalculate the score
For each FRT set ASi produced by Single Feature Bootstrapping
1. Use Equation (4) to calculate MB score of every aspect related term in ASi;
2. Arrange the feature related terms in the descending order of MB score to produce final ASi.
Output: Final created Feature Related Term sets AS* for x aspects.

3.2.3. Sentence Segmentation based on Fuzzy Score:
Till now we were working on single - feature/aspect but when more than one aspect/feature in single sentence or combination of two or more such sub sentences, we use fuzzy score to produce multiple single-feature segments. Let \(R=r_1 r_2 r_3... r_n\) be a sentence R that contains n sub-sentences, and let \(V=v_1 v_2 v_3... v_k\) be a segmentation of R consisting of k single-aspect segments. The MAFS model can be formed by adding a new function \(Z(R,V)\) which helps
to justify each candidate to be selected for segmentation and assign a fuzzy score to it.

\[ V^*_{\text{def}} = \max_{v \in V} \sum_{(f, p)} \text{fuzzy score} (f) \]  

(5)

In this working model, we identify feature/aspect of a segment by using FRT sets AS = \{S1, S2, ..., Sm\} for m aspects, produced by Single feature or Multi-feature Bootstrapping method. The \( n^\text{th} \) aspect score of a segment \( v_i \) in \( V \) can be calculated by taking the sum of important scores of all FRTs of the \( n^\text{th} \) aspect in the segment, that is,

\[ \Phi_{(v_i)} = \sum_{(f, p)} \text{fuzzy score} (f) \]  

(6)

The familiar aspect/feature \( n^* \) of the given segment \( v_i \) is found by,

\[ \max_{v \in V} \Phi(n(v_i)) \]  

(7)

To find whether two adjacent segments in \( V \) have different features, we use a function \( \mathcal{L}(v_i, v_j) \) to show value 1 if segments \( v_i, v_j \) are used as two different aspects and 0 otherwise. We assume value of \( \mathcal{L}(v_0, v_1) \) to be 1. The value for the selection function can be calculated as,

\[ \mathcal{L}(v_i, v_j) = \sum_{(f, p)} \mathcal{L}_{(f, p)} (v_{i+1}, v_i) \times \mathcal{L}_{(f, p)} (v_i) \]  

(8)

For the final segmentation \( V^* \), we apply following rule to adjust an aspect: If the feature score of an aspect of segment \( v_i \) is less than previously defined frequency, the feature is considered as Null.

3.3. Fuzzy ontology tree for sentiment analysis:
Fuzzy ontology tree is a hierarchy of concepts in which the product becomes the root node, the non-leaf nodes consist of the features of that product and lastly the end nodes represent the sentiment in the form of positive or negative related to their respective parent. Automated analysis of sentiments has been carried out in previous research works using lexicon-based and machine learning based systems. Here in our paper, we use the FOT technique to represent product features and sentiments in each review sentence in form of a tree. There are huge number of reviews for the product from which we can find out sentiments that strongly connect to a given product feature by matching the candidate feature/aspect and sentiment/opinion word with the same aspect, polarity pairs in our FOT. The final score of the pair is found by matching the results obtained from opinion words in customer review with the opinion words from fuzzy method. For the implementation of the proposed work we are using Java platform and Oracle as database.

IV. EXPERIMENTAL RESULTS AND DISCUSSION:

4.1 Data Set used:
We have used customer reviews for different brands of mobile phones in the market. The dataset used for experiments to be carried out are from following links: https://github.com/ ucdspatial/ Datasets, https://github.com/hackreduce/Hackathon/wiki/Amaz on-review-dataset etc. To carry out the evaluations, we build an evaluation set in advance. The data set consist of only raw customer review which is in unlabelled form. This is followed by pre-processing step in which word segmentation and tagging is done. In the preprocessing step, we utilized the SentiWordnet tool to implement English word segmentation and POS tagging. Random selection of 1,000 sentences is done to find whether they belong to single aspect, multiple aspects or does not contain any aspect. Manual annotation of each review is done. To start the evaluation of the proposed method, we use the precision, recall, and F-measure to measure how effectively aspect-opinion pairs and their sentiment polarity are identified.

4.2 Experimental outcomes:
In this section, we carry out the selection of some feature seeds and then construct a FOT tree. In the pre-processing we carry out the segmentation and tagging of the parts of speech. Further we select only relevant data that will be used as input for the construction of the FOT.

The following diagram shows a sample tree created and we can find the precise results for opinion poll as we assign an ontology weight factor from child to parent i.e. Positive with higher one and negative with lower one.
A Fuzzy Approach For Sentiment Analysis Based on Web Customer Reviews

To perform the evaluation, we determine the performance of our system in terms of accuracy. Accuracy means number of (feature, opinion) pairs correctly identified automatically. We have given partly results in form of tree and table but we predict that our method can achieve higher precision and F-measure score due to automatic learning of sentiments/opinions knowledge for the opinion pairs approximately 72% Precision value and 65% F-measure value which are better than previous Aspect based segmentation method.

The bold number denotes the best performance.

CONCLUSION

Sentiment Analysis is a very important issue to give accurate analysis of opinions from customer reviews. Hereby, a new approach is presented in this research work i.e. fuzzy ontology tree method is proposed for the segmentation of multi-aspect sentence. Fuzzy ontology technique is one of the solutions to give proper semantic meaning for data in uncertain and inconsistent web world. Fuzzy ontology tree can be automatically constructed that helps for faster sentiment analysis. Basically, unlabelled data is taken as an input and unsupervised learning technique is be used for the analysis. We deal with both the problems of feature/aspect and sentiment/polarity change occurring in input customer review sentence for sentiment analysis.

ACKNOWLEDGEMENT

The author wish to express her sincere gratitude to Dr.Varsha.Patil madam, Vice-Principal, and Prof.N.L.Bhale Sir, HOD, IT Department, Matoshri College of engineering, Nashik for their constant encouragement, important comments and helpful suggestions. The author is also thankful to all friends and family for their support. This paper is submitted in subject to University of Pune for postgraduate degree.

REFERENCES

A Fuzzy Approach For Sentiment Analysis Based on Web Customer Reviews


