Research Article

Aspect Based Opinion Mining and Ranking: Survey

Aarati Mahadik†* and Asha Bharambe†

†Department of Information Technology, V.E.S. Institute of Technology, Mumbai University, India

Abstract

Aspect based opinion mining plays important role for both academic and business perspective. Whenever people wants to take decision they always go for others opinion, now a days it has always been an important that what others thinks, from where they collect important piece of information.(e.g. which car I should buy, which school should I apply to etc.). Now a day's e-commerce sites playing vital role in our life, they are providing products/services online for which they allow us to leave our reviews/opinion on which success of product/services is depends. Reviews usually have an overall rating about product/ service, but most of the times there are sub-texts in the review body that describe certain features of the product/service. These features are also known as aspects. In this paper various techniques and specifications for aspect identifications are provided.

1. Introduction

As user generated content is increasing day by day on web, applying mining on sentiment and opinion has grown rapidly and become crucial. Work previously done on sentiment analysis focus on over all polarity of content provided by user. Users are concerned not only in the overall quality of a product/services, but also its quality in certain specific aspect areas. For example, someone looking at reviews of a new car will probably be interested in individual aspects such as the power of engine, working of AC, mileage, and the perceived value with respect to price. This suggests that rating to specific aspect is more useful than overall rating on product/services and plays vital role in gathering information as per user requirements. These sentiments are expressed in the form of text review on different blogs, forums and social networking sites. There are many e-commerce sites which provides online shopping like amazon, eBay, flipkart, mantra, snapdeal etc. most of the sites allow user to post their review about products so that these reviews will be used by the other users in future and also firms can get feedback regarding their products, whether peoples liking or disliking it. Accordingly improvement can done as per user requirements which exceed sale of their products. Aspect based opinion mining include aspect mining, sentiment prediction, and sentiment classification [Hu, M., and Liu, B. *et al*, 2004]. As the number of users for ecommerce sites are getting

increase number of reviews written by users is also increasing. Reviews written by users are unstructured, long in text size so it become very difficult for users to read all the review and find exact opinion about product/service, it is also time consuming task for firm to keep record about their manufactured goods. To overcome this problem we need some automatic approach. So here opinion mining comes in picture. In proposed work firstly important aspect will be identify. And then will calculate the rank of aspects with ranking algorithm.

2. Literature survey

In this section we go over related work previously done. In General, opinions are found from user's text (from Blogs, Review sites, Micro blogging). Text can be a word, phrase, sentence or document. According to Bing Liu, Opinion mining tasks are generally classified into three major levels: Document level, Sentence level, and Phrase level (Aspect based).

Document Level

Document level is to classify whether a whole opinion document expresses a positive or negative sentiment. For example, given a product review, the system determines whether the review expresses an overall positive or negative opinion about the product.

Sentence Level

The task at sentence level goes to the sentences and determines whether each sentence expressed a

*Corresponding author: **Aarati Mahadik; Asha Bharambe** is working as Assistant Professor

positive, negative, or neutral opinion. Neutral usually means no opinion. This level of analysis is closely related to *subjectivity classification*, which distinguishes sentences (called *objective sentences*) that express factual information from sentences (called *subjective sentences*) that express subjective views and opinions. Both the document level and the sentence level analyses do not discover what exactly people liked and did not like.

Aspect Level

Aspect level performs finer-grained analysis. Aspect level was earlier called *feature level* (*feature-based opinion mining and summarization*). Instead of looking at language constructs (documents, paragraphs, sentences, clauses or phrases), aspect level directly looks at the opinion itself. A drawback in document or sentence level is that they cannot provide complete information of a product, For example, a positive or negative review of a particular product doesn't mean that the reviewer likes or dislikes all aspects of that product.

Previously done work in aspect based opinion mining are mentioned below: By using frequently found noun method

Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang [Zheng-Jun Zha *et al*, 2014] extract the apsects. In order to obtain more precise identification of aspects, they propose to exploit the *Pros* and *Cons* reviews as auxiliary knowledge to assist identify aspects in the free text reviews. In particular, they first split the free text reviews into sentences, and parse each sentence using Stanford parser.

Part of speech tagging (POS) is performed to generate feature list, which is used by Azra Shamim, Vimala Balakrishnan, Muhammad Tahir, andMuhammad Shiraz [Azra Shamim *et al*, 2014]. Part of speech (POS) tagging: Parts of speech or POS tagging is a linguistic technique used since 1960 and has recently got particular attention of NLP researchers for product feature extraction as product aspects are generally nouns or noun phrases. POS tagging is an important phase of opinion mining, it is necessary to determine the features and opinion words from the reviews.POS tagging can be done manually or with the help of POS tagger. Manual POS tagging of the reviews take lots of time. Here, POS tagger is used to tag all the words of reviews POS tagging assigns a tag to each word in a text and classifies a word to a specific morphological category such as norm, verb, adjective, etc. POS taggers are efficient for explicit feature extraction in terms of accuracy they achieved, however problem arises when review contains implicit features.

Mita K. Dalal and Mukesh A. Zaveri [Mita K. Dalal *et al*, 2013] use frequently occurring nouns (N) and noun phrases (NP) as possible opinion features and associated adjectives describing them as indicators of their opinion orientation perform parts-of-speech (POS) tagging on the review sentences using the Link Grammar Parser.

POS tagging and Seed listing is used by the Richa Sharma, Shweta Nigam and Rekha Jain[Richa Sharma *et al*, 2014] for extracting opinion on reviews. With the help of seed list, the polarity of the sentences is determined for each feature. Polarity is determined on the basis of majority of opinion words. They also handle negation in reviews .

2.1 Aspect Identification Methods

Previously used opinion mining approaches for Aspect identification are frequency based, relation based, and hybrid method.

2.1.1. Frequency based method

Frequency-based methods usually apply a set of constraints on high-frequency noun phrases to identify aspects. An aspect can be expressed by a noun, adjective, verb or adverb. However, recent research shows that 60-70% of the aspects are explicit nouns. In reviews people are more likely to talk about relevant aspects which suggests that aspects should be frequent nouns. However, not all of the frequent nouns are aspects. Therefore, different filtering techniques are applied on frequent nouns to filter out non-aspects.

Strength: Although these methods are very simple, they are actually quite effective. Many companies are using these techniques for analyzing their user feedback.

Limitations: These methods tend to produce too many non-aspects and miss low-frequency aspects. In addition, they require the manual tuning of various parameters (thresholds) which makes them hard to port to another dataset.

2.1.2. Relation-based Methods

Relation-based methods exploit aspect-sentiment relationships to extract new aspects and sentiments. The intuition behind this approach is that each sentiment expresses an opinion on an aspect and sentiments are often known or easy-to-find. As a result, their relationship can be used for identifying new aspects (and sentiments).

Strength: Compared to frequency-based methods, relation-based approaches can find low frequency aspects.

Limitation: The main limitation of these methods is that they produce many non-aspects matching with the relation patterns.

2.1.3. Hybrid Methods

Hybrid approaches combine both frequency- and relation-based techniques to extract aspects. These methods normally use aspect-sentiment relations for filtering frequent noun phrases. The intuition behind the hybrid approach is that aspects are mostly frequent

nouns and are normally described by some sentiments. So, the relationship between aspects and sentiments can be used for filtering non-aspects.

Strength: Compared to frequency- and relation-based approaches, the number of non-aspects is more limited in the result of hybrid methods since they apply more constraints (frequency threshold and relation pattern).

Limitation: While hybrid methods can limit the number of non-aspects, they still miss low frequency ones. They also require the manual tuning of various parameters.

2.2. Aspect Sentiment Classification Techniques

2.2.1. Lexicon Based Approach

Three approaches used to collect this opinion word list are manual, dictionary based and corpus based approaches.

Dictionary-based approach

In this approach a small set of opinion words is collected manually with known orientations. Then, synonyms and antonyms of these words is added to this set which is increase by searching the words in the well known corpora Word Net or thesaurus. The newly found words are added to the seed list then the next iteration starts. This iterative process stops when no new words are found. After completion of process list is checked manually to remove or correct errors.

Corpus-based approach:

The drawback of dictionary based approach is overcome in Corpus- based approach which helps to solve the problem of finding opinion words with context specific orientations. Its methods depend on syntactic patterns.

2.2.2. Holistic Lexicon Based Approach

This method does not look at the current sentence alone rather it uses the external information and evidences in other sentence and other reviews. Some linguistic conventions in natural language expression are used to find the orientation of opinion word. This method required previous domain knowledge or user inputs are required. This approach is highly effective when sentence contain multiple contradictory opinion words.

2.2.3 Supervised Learning Techniques

Naive Bayes Classifier (NB):

Naive Bayesian networks are composed of acyclic graph with only one parent and several children. There is a very strong assumption of independence with child nodes in the context of their parents. When these two probabilities are compared, larger probability is more likely to be the actual class label. Advantage of naive bayes classifier is its short computational time for learning the dataset. Bayes classifiers are usually less accurate than that of other learning algorithms.

Maximum Entropy Classifier (ME):

Another classifier is Maximum Entropy classifier. The name Maximum Entropy comes from the fact that the classifier finds the probabilistic model which is the simplest and least constrained. Yet it has some specific constraints.

The idea behind maximum entropy is that one should prefer the most uniform models that also satisfy any given constraints.

Support Vector Machine

SVM is a useful technique for data classification. A classification task usually involves with training and testing data which consist of some data instances. Each instance in the training set contains one target values and several attributes. The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes. Classification in SVM is an example of Supervised Learning. Known labels help indicate whether the system is performing in a right way or not. This information points to a desired response, validating the accuracy of the system, or be used to help the system learn to act correctly. A step in SVM classification involves identification as which are intimately connected to the known classes. This is called feature selection or feature extraction. Feature selection and SVM classification together have a use even when prediction of unknown samples is not necessary. They can be used to identify key sets which are involved in whatever processes distinguish the classes.

3. Proposed Work

The goal of proposed system is to develop system will extract the aspects from reviews given by the users and mine opinions on extracted aspects which will be recommended to users.

3.1. Data Pre-processing

Reviews submitted by users are not in structured form, so before we proceed to extract aspect need to do pre processing for review data set, because if opinion mining will apply on unstructured data will not get accurate result as per requirements. There are many methods include in pre processing.

Stemming and Lemmatization

Stemming and Lemmatization are two essential morphological processes of pre-processing module during feature extraction. The stemming process

converts all the inflected words present in the text into a root form called a stem. For example, 'automatic,' 'automate,' and 'automation' are each converted into the stem 'automat.' Stemming gives faster performance in applications where accuracy is not major issue. The lemma of a word includes its base form plus inflected forms. For example the words plays, played and playing have play as their lemma. Lemmatization groups together various inflected forms of word into a single one. Stemming removes word inflections only whereas; Lemmatization replaces words with their base form. For example, the words caring and cars are reduced to car in a stemming process whereas lemmatization reduces it to care and car respectively, hence lemmatization is considered to be more accurate.
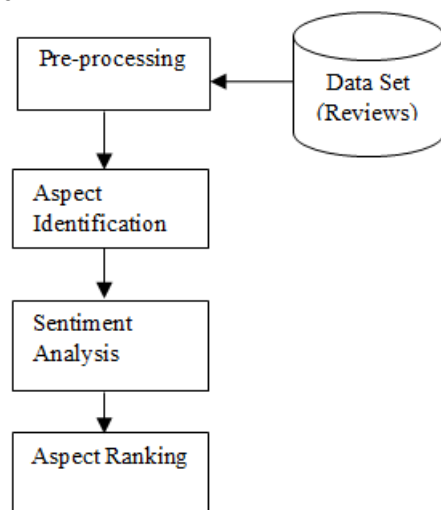
Fig: 1 System Overview

3.2. Aspect Identification

We classify aspects from consumers reviews which are available from different forum websites in Aspect identification step. Reviews found in forum websites are either positive, negative or neutral. Previously done work shows that aspect identification can be done on basis of nouns and noun phrases which are frequently present in reviews. Hu and Liu proposed most notable approach for aspect identification. In this approach it first identifies the noun and noun phrases in the document. Phrase dependency parser used to extract the noun phrase by Wu *et al*.

3.3. Sentiment Analysis

Sentiment analysis is used to extract sentiment (Opinion) from reviews posted by users on products/ services on web either they are liking it or not. Users normally express their opinions as positive , negative or neutral. So in sentiment analysis reviews are separated on positive , negative or neutral basis. For example Moto G mobile have good display. Here the opinion they have used is good, which is positive sentiment , here the users are expressing their emotions about mobile display which is positive.

3.4. Aspect Ranking

Aspect based Ranking algorithm is used to identify important aspect of product from reviews given by the users. The important aspects are commented again and again in consumer review and the consumers opinions on the important aspects are greatly influence their overall opinions on the product. The overall opinion in a review is an aggregation of the opinions given to specific aspects in the review, and various aspects have different contributions in the aggregation.

Conclusion

Aspect based opinion mining for reviews is approach for full fill user requirement as per their requirements in searching product/services and choosing product/services. This is achieved by a fine-grained aspect based review ranking, review recommendations. With the massive advancement in Social networking and ecommerce the amount of reviews or opinions of individuals has grown leaps and limits. For the users to make informed decision on the products/services, ranking these reviews based on their preferences and friend circle becomes usable. Further a finer-grained aspect based opinion mining of reviews helps users to filter or rank based to the features they are most interested in.

References

Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang (May 2014), Product Aspect Ranking and Its Applications , *IEEE* Vol. 26, No. 5, , pp 1211-122

Azra Shamim, Vimala Balakrishnan, Muhammad Tahir and Muhammad Shiraz (2014), Critical Product Features Identification Using an Opinion Analyzer, *Hindawi Publishing Corporation Scientific World Journal* Volume, pp 1-9

Mita K. Dalal and Mukesh A. Zaveri (2013), Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews, *Hindawi Publishing Corporation Applied Computational Intelligence and Soft Computing* Volume, pp 1-8

Hu, Minqing, and Bing Liu. (2004), Mining and Summarizing Customer Reviews. *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining KDD 04* 04: 168.

Richa Sharma, Shweta Nigam and Rekha Jain (May 2014) , Mining Of Product Reviews At Aspect Level, *IJFCST,* Vol.4, No.3, pp 87-95

Akshi Kumar and Teeja Mary Sebastian (July 2012), Sentiment Analysis on Twitter, *IJCSI,* Vol. 9, Issue 4, No 3, pp 372-378

K. Vivekanandan, J. Soonu Aravindan (November 2014), Aspect-based Opinion Mining: A Survey , *International Journal of Computer Applications Volume 106 – No.3*, pp 21-26

Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou (May 2012), Movie Rating and Review Summarization in Mobile Environment, *IEEE Transaction on systems*, Vol. 42, No. 3, pp 397-407

A. Nisha Jebaseeli, E. Kirubakaran (June 2012), A Survey on Sentiment Analysis of (Product) Reviews, *International Journal of Computer Applications* Volume 47– No.11, pp 36-39

Hu, Minqing, and Bing Liu. (2004) Mining Opinion Features in Customer Reviews, AAAI'04 *Proceedings of the 19th national conference on Artifical intelligence,* pp 755-760