International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 4 712 - 716

Opinion Mining and Review Quality Research

Amruta D. Sawant, Ashutosh R. Bhumkar, Anudnya Pujari
Department of Computer Engineering
Vishwatmak Om Gurudev College of Engineering
Aghai, India.
amzs1995sawant@gmail.com, ashutosh.bhumkar@outlook.com, anudnyap@gmail.com

Abstract-One of the most important measures for reviewing any product these days is getting views about the product and analyzing them so as to see where exactly the product stands in the market. The best source is to collect views from the people who are indulged in buying these products or entities. Almost every item or entity these days is subjected to classification.These can be rated asgood, bad, worse, better, or best. Opinions form tobe the real strong parameters that help in categorization of the items available in the market.The rating becomes so important for comparison that the entire market value is dependent on the public's say.With the growing competition against the products, reviews of the people have become a vital domain for their characterization. Hence a concept accentuating the opinions collected from the peoplehas been defined and is used to make things smooth for analysis of these items. This concept is called concept of Opinion Mining.

Keywords:*opinion mining,sentiment analysis, entities, services*

*****

I. INTRODUCTION

Opinion mining is a natural language processing technique,which is used in business process to gather various opinions from the customers in order to review a particular product. It is extremely essential for both, the customers or the businessmen to get a clear idea about their product in order to knowits market value and the amount of usage of that specific item. Business processes regularly use it for marketing, developing different items, identifying the needs of the customers and creating products which satiate it, comparing existing products and then developing more efficient ones also for customer service and many more.

Opinion mining is also known as sentiment analysis. It constitutes the usage of natural language processing and alsoanalysis of various sentences or texts which basically are reviews, so as to classify these sentences or opinions into different categories. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service.It can be thought of as a concept which works on the sentiments or emotions of the people. Each person associatesa certain feeling with respect to some entity and inspection of such a feeling is sentiment analysis.Opinion mining basically includes collecting various perceptions or views about a particular service or product.These perceptions are then collected and used for analysis of any product.

The core concept of opinion mining includes using a scale to rate various parameters that form to be the characteristics of a certain entity and then using it for analysis. As we see the world of web is fast growing and also rightly gives justice to interconnecting many people at a time. The discussions and interactions that happen amongst the people are done at a global level. These interactions include discussions about almost everything that forms a part of our social surroundings. Giving views and commenting about all the newly arrived products or services, comparison of these with the old ones etc, forms to be a major part of these discussions. Blogs and other social media prove to be the platforms for letting people express their viewsabout different commodities. However as per Technorati, which is a publisher advertising platform, 75000 new blogs are created daily. Also millions of people put up posts. These blogs include a vast number and a variety of views about any service or product. This makes it difficult for the analyzers to read every opinion and then draw a conclusion from these opinions. Hence software or method which takes opinions as input and does analysis on them are developed inorder to ameliorate the review scrutiny, thus helping all thebusinesses to get a big picture of their products and assisting them to get a fair market review regarding the same. In general, any opinion mining method, mines the opinions from various sources and then these mined opinions are used to perform analysis, so as to get the actual knowledge about the service or the commodity.

A diagrammatic representation of opinion is given below.

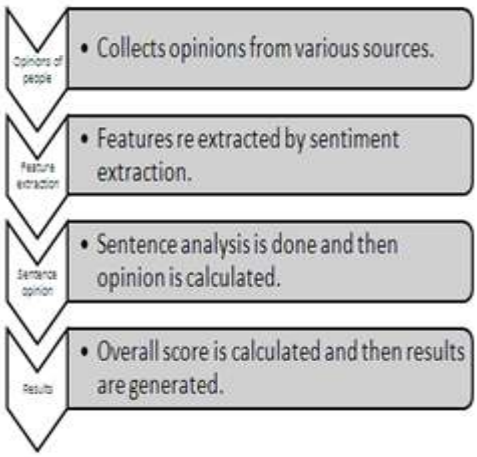

Fig.1 Opinion mining basic flow

Opinion mining has three basic terminologies on which it can be processed. These are,

a. Object: that is the entity or commodity regarding which opinion has to be given.
b. Opinion holder: the one who gives the opinion, also known as the author.

c. Opinion: the particular statement or the corresponding opinion which is expressed by the opinion holder or the author.

Opinion mathematically can be expressed as follows:

Opinion is aquintuple (O_j,f_{jk},h_i,t_l,SO_{ijkl}) where,

O_j is the object which is to be opinionated.

F_{jk} is the k^{th} feature of j^{th} object which is opinionated.

h_i is the i^{th} opinion holder.

SO_{ijkl} is the sentiment for the j^{th} object's k^{th} feature, byi^{th}opinion holder, at time l.

$t_l$is the time at which the opinion is created.

Example: a certain person has to comment on a book. Then in this case the person is the opinion holder, the book the object and "this is a good book" is the opinion.

II. BASICS OF OPINIONS AND SENTIMENT ANALYSIS

Opinions can be of varied types consistingof direct simple sentences as well as compound sentences. These sentences can include views upfront or via comparisons. Example

"Audi A6 is better than Maruti Suzuki Alto". Here two entities are compared. Also there are reviewers who like to talk in sarcasm and hence make it extremely complicated to analyze the hidden or the true meaning of their views. English is a large language and includes a variety of ways in which one can put up his or her ideas. This can include encompassing two or more thoughts in the same sentence or making up a sentence which includes only one idea. The extensive use of adjectives and nouns also makes it difficult to interpret exact meanings of the sentences as the number of synonyms found is huge.Constructing opinions out of such myriad types of sentences isthus indeed a skill.

In general the task of opinion mining includes following concepts.

a. Sentiment classification:Classify sentence/document/feature based on sentiments expressed by author as positive, negative or neutral.
b. Comparative mining: Mining of the comparative sentences and extraction of the comparison that is stated.
c. Opinion Integration: Collection of different opinions from different sources into one single source.
d. Opinion Spam/Trustworthiness: Detection of how true is the opinion as many of it can be faked in order to give false views about entities.
e. Opinion Retrieval: Getting the opinions regarding a particular query or a product. After collection of opinions, they are retrieved and arranged as ranked documents.
f. Opinion Question Answering: Analogous to opinion retrieval but it includes answering questions about the particular topic. Answers need to be in natural language format.

In our paper we will be focusing more on Sentiment classificationand Quality of opinions.For any opinion mining task, there are three levels at which it can be performed. These levels include opinion mining at sentence level, document level and feature level. Let us see each of it in brief.

a. Opinion mining at sentence level: In this level of opinion mining we consider the sentence for analysis. Here the statement is taken into account and is first checked to be a subjective or an objective sentence. An objective sentence is that sentence which includes true parameters or characteristics of any entity. We can say that an objective sentence is basically a sentence which is factual. The truth is encompassed in an objective sentence and gives all the actual information about it. A subjective sentence on the other hand is the one which includes biased opinions of people associated to any entity.A subjective sentence basically depicts the views or emotions of the opinion holder attached to an object.Hence the difference between the subjective and objective sentences can be thought of as a difference between facts and opinions.Example: "The car has airbag system" is an objective sentence and "this is a wonderful car" is a subjective sentence. When opinion mining is done at a sentence level we first identify if the given statement is a subjective or objective sentence.Then is the job of polarizing the sentence. Polarizing the sentence includes fining out the polarity of that particular statement. We have three polarities that include positive, negative and neutral. Any sentence can either be positive, negative or neutral. Example: "The car which has airbag systems is a wonderful car." This is a subjective and positive polarity sentence. Hence opinion mining at sentence level includes identification of subjective/objective sentence and then polarizing it.

b. Opinion mining at document level: The document level classification of opinions relates to analysis of an entire document which gives idea or opinions regarding only a single product.This document includes the opinions of a single author or opinion holder. The opinions are documented and sent for analysis. After analysis the document is checked for its polarity. And then it is classified as positive, negative or neutral. A neutral document does not convey much of an opinion about that product or service as its polarity is unable to suggest anything strongly. A wide range of algorithms like Naïve Bayesian algorithm, Maximum entropy algorithm, support vector machine(S.V.M) etc, are given. The works of document level opinion mining is done largely by Pang, Lee,Vaithyanathan and Turney in 2002.

c. Opinion mining at feature level:Features ofany object can be thought of as parameters of that object. Feature level opinion mining involves the rating given to these features and then calculating the overall polarity for that object. Objects can have a lot of features. Each feature can be scaled independently and then be considered to get the overall gain of the service or product.Example: if we are to compare cars then we will look out for the aspects like the average given by the car, the looks of the car, the seating of the car, the credit rating of the car etc. Based on all these characteristics we will see if we want to buy the car or choose other option.Hence when services or commodities are compared based on its features we do feature level opinion mining. Feature level mining is now-a –days also called as aspect based opinion mining.

Opinion mining also known as sentimentanalysisis one of the most burgeoning topics in the field of research. The domains of soft computing, machine learning and data mining etc have the evidence of their roots fixed in this field. Rise of various commercial operations and pervasive real life problems or obstacles are one of the important factors affecting the growth of sentiment analysis. Hence we can find a whole lot of applications which are effected by sentiment analysis. As the name itself suggests, carrying out operations on emotions or feelingscan be thought be sentiment analysis. The emotions can be depicted in a number of ways ranging fromfacial expressions, way of actions, through art forms like dance and songs, through written work such and poems, blogs orarticles and many such vivid sources.

However opinion mining basically focuses on sentiments which are shown via writing. The sources for opinion extraction majorly focus on social networking sites and review websites. Also discussion forums and blogs are useful repositories for mining.

III. APPROACHES OF SENTIMENT ANALYSIS.

Sentiment analysis consists of various elements out of which lexicons are an integral part. These lexicons are nothing but the words which make up a sentence. These lexicons can also be known as sentiment lexicon or opinion lexicons. It is important that we classify the lexicons we have into positives or negatives in order to achieve review analysis. One of the methods forsentiment analysis and classification is done on the basis of these lexicons.There are two broad approaches for sentiment analysis. These are lexicon based approaches and machine learning based approaches.

These approaches are further divided into different methods as shown in the figure given below.

Fig. 2 Sentiment analysis flow

Let us consider each approach in more detail.

A) Machine Learning Approaches

Machine learning methods include the concepts from soft computing field. These methods are classified as supervised and unsupervised methods. The supervised methods can either be probabilistic classifiers else it can be linear classifiers. Probabilistic classifiers include algorithms which use probability to find the classification and then separate the lexicons based on probabilistic calculations.In this approach we use data driven methods, which include the following algorithms.

a. Naïve Baye's Classifier: This algorithm includes the use of Baye'stheorem for determining the probabilistic value of the sentiments.
b. Maximum entropy method: It focuses on maximizing the entropy calculated on the probability distribution.

Hence are the probabilistic classifiers. The linear classifiers are as shown.

a. Support vector machine: This method includes linearly separating the inputs. A training set is present and based on it mathematical equations are found out. Weights and bias are the major terms involved in it. After classification is done a line is obtained which separates positive sentimentsfrom the negative ones.
b. Neural networks: These are those processing systems which are inspired by the structure of a human brain. These contain cells called the neurons and have weights associated with them. These have the capability to learn on their own.

Hence are the supervised learning algorithms.The unsupervised methods are those types of methods which include inspection of lexicons on a large amount. The entire input sentence is read first and then based on the patterns of parts of speechoccurring after each word the decision of accepting them is made. A certain set of pattern is predefined and if it matches with the sentence that we have, then it is accepted.The concept of "parts of speech" used inEnglish language isdeployed for preparation of these different predefined patterns. After the pattern is extracted, the sentiment orientation is calculated by using a statistical method known as point wise mutual information(P.M.I). The last step includes the segregation of the phrase as positive or negative.

B) Lexicon Based Approach

Lexical analysis checks the syntax of the statement or the opinion. If the syntax is appropriate then the sentence is split in different types. These sentences are analyzed on the basis of document level, sentence level and aspect or entity level. Lexical analysis includes corpus based analysis.

Corpus based analysis:

Corpus when all is said in done, implies any composed body around a specific content. Corpus is an arrangement of machine intelligible content of a specific dialect. It additionally comprises of the gathering of short structures utilized for some words that are utilized day by day illustration lyk, wid, and so forth. The corpus examination chips away at the example of the words and linguistic use in the content that adds to the significance of the content. There are numerous sorts of sentences like positive, negative, snide, similar sentences and so on to perform corpus examination on them we have to assess these on the premise of verbs and descriptive words utilized as a part of the sentences. Positive and negative sentences are anything but difficult to break down utilizing corpus. Be that as it may, it is hard to break down relative and wry articulations. Different semantic standards are mulled over, and a profound investigation of different things, modifiers, and the examples in which they happen is likewise thought about. Corpus construct examination significantly centers with respect to the route in which English dialect is talked or composed. Corpus based examination can be further thought as investigation done on semantic level and measurement level.

Semantic Analysis:

In semantic analysis sentences are converted into probability so as to calculate the average.These sentences or opinions are collected from different sources (Probably Internet sources) this is called as data streaming. This analysis is done by taking the word count of different required adjectives and verbs and their synonyms from the corpus. Thus this analysis is called as Corpus based analysis. From the collected data some adjectives are collected, the system contains synonyms of those adjectives against which we calculate respective occurrences of words.The semantic analysis is done including document, sentences or entities/aspects sentences. In general, in semantic analysis, the type of sentence is determined and then the words which can be synonyms are calculated. After deciding the types of sentences and counting the number of similar words or the synonyms of the given word then a table is created with the words and its counts. Then the statistic analysis comes in picture.

Statistic analysis

After getting the data from the semantic analysis,this data is converted into the histograms,bar charts,pie charts,line graph etc. Various statistical computation techniques are then used to enable the corpus based approach, one of the techniques being P.M.I that is point wise mutual information formula.

$$PMI(term_1, term_2) = \log_2\left(\frac{\Pr(term_1 \wedge term_2)}{\Pr(term_1)\Pr(term_2)}\right).$$

Challenges faced by Corpus analysis

a. Opinions are subjective but even objective based sentences can be opinions. Hence it is bit difficult to analyze the type.
b. Sarcastic statements are not computed properly.

C) Other Approaches

The above seen method or approaches are traditional approaches. They were developed along with the commencement of sentiment analysis. However, a few hitches were identified. These include the below stated ideas.

a. Cross domain sentiment analysis: The emergence of different domains has led to an increase in the amountof the training sets. Different domains need different opinions. An approach hence was taken into consideration which included, using the training set that was adapted from other domains and utilized in domains excluding the ones,to which the training sets belonged.Sentiment classification, being very domain sensitive from which the data is extracted for training, a trainer usually performs poorly on the test data of any other domain. The main reason behind thisis, either the words or the construction of sentences used for expressing the opinion. These words vary from domain to domain. Additionally the meanings of these words also vary. The only solution here is transfer learning or adaption of domain. Hence various techniques which solved the cross domain problems are developed.

b. Cross language sentiment analysis: It has been seen that researchers are found comfortable using their own native language. Hence, they are working on building sentiment analysis software in their own mother tongue. However English is a language, which is used commonly at a global level hence, a lot of research has been found in English language only. Thus at this point of time there are not many ways which can be used or implemented to build good sentiment classification in native languages. Hence there is a good amount of scope in area that is the one which will include native natural language processing.Thus above are a few approaches of sentiment analysis.

IV.QUALITY OF REVIEWS

As the world of web has grown out toalmost each and every part of our planet and has made impossible communication so simple that even interconnecting with every soul these days is possible. Also it has a huge impact on the way people discuss about various things or entities that they encounter in their day to day lives. The way things get opinionated has drastically changed. Whenever we need any advice regarding a certain commodity or service, internet is the first thing that strikes our mind. Also review sites and forums or groups created on the internet. Various social networking media and applications are developed so as to increase the interaction amongst the people. These sources also form to be great repositories for the collection of opinions and are subjected to opinion mining. However, one of the greatest doubts arises about the authenticity of the opinion. Many people know that websites and forums are the biggest sources for getting reviews and some of them can miss use it for their own good or for harming others.Quality of reviews in general, takes into considerationthe spam opinions and also considers if the all the parameters of that particular opinion are given justice. Quality of reviews is useful for rating the reviews and also for checking the helpfulness of the reviews. We often come across statements like “was this review helpful?” on the sites. This is done in order to, check how much has that opinion helped the visitor or user of the site. In many cases we also have “thumbs up or thumbs down” symbol, which also does the same task.Determining the quality of reviews is usually formulated as a regressionproblem. The mining model assigns a score to the quality of each review. The opinions are then ranked and recommended as per their score. The user feedback hold a very important place in quality of reviews as, only that forms the database or training set for the determining the authenticity of the reviews. An SVM regression is used to calculate the quality of the opinion. The features of the review are used inthe algorithm. The features are of different types. Features related to structure of the review like its length, number of sentences, percentage of punctuations, bold tags and breaks used, number of adjectives or nouns, adverbs,verbs or any other parts of speech used. Also how famous is the product in the market and what is its current review rating etc. are considered.

Other parameters which can be used to analyze any review is who is the author of that review, what is the author’s qualification, on what basis has he characterized the review as good, bad or anything. How much knowledge does the author has in product’s field, does his opinion have any reasoning or any experience associated with it, has the opinion been generated on a purely emotional basis or are there and logical bolsters to it. All these aspects are taken into consideration before analysis of the opinion quality. Also now-a-days, use of emoticons has increased to a very huge level. We can also segregate the reviews based on these emoticons as they are direct depictions of what exactly a person feels. Another aspect which can be used to see the opinions are how correctly the grammar has been used in the sentence and if there are any spelling mistakes. This can show that people usually having a correct way of putting up sentences and have correct spellings usually are more capable of giving good reviews. However, this cannot completely be true as giving opinions is basically analyzing a particular product or service and merging it with one’s own experience hold more grounds. Hence this component of opinion mining is fast growingand has a vivid sector to study.The year of 2009marks the advent of a new approach to classify reviews by considering some extra features like reputation, content, social and sentiment features.

V. CONCLUSION

This paper consists of understanding of core concepts of opinion mining or sentiment analysis. It also includes various approaches that are used to define or carry out the classification amongst various opinions. A study of all the approaches is shown in brief. Further down a perception is given about spam reviews and quality of the opinions, so as to focus on one of the major aspects involved in opinion mining. Finally we conclude the paper by giving a brief method for quality of reviews.

VI. ACKNOWLEDGEMENTS

We would like to firstly thank our parents for being so patient with us and for motivating us to take extra steps and run errands to share something we know. We also thank Prof. AmrutaMhatre, as she has given us her valuable time and the tremendous push that was so necessary for the completion of this paper.We also thank our friends who were very co-operative during the entire process. Last but not the least, we are obliged of our entire team of teachers who guided us throughthick and thin and most importantly the patrons of ICMtest for giving us this opportunity to put up our knowledge before the people.

REFERENCES

[1] B. Liu 2007. Web Data Mining,Exploring Hyperlinks, Contents and Usage data.
[2] B. Liu 2011. Opinion Mining and Sentiment Analysis, AAAI, San Francisco, USA.

[3] B.Liu 2010. Opinion Mining and Sentiment Analysis: NLP Meets Social Sciences", STSC, Hawaii.
[4] B. Liu 2008. OpinionMining and Summarization, World Wide Web Conference, Beijing, China.
[5] B. Liu. 2010. Sentiment Analysis: A Multifaceted Problem., Invited paper, IEEE Intelligent Systems.
[6] B. Liu. 2010 Sentiment Analysis and Subjectivity Second Edition, The Handbook of Natural Language Processing.
[7] B. Pang, L. Lee, and S. Vaithyanathan, 2002. Sentiment classification using machine learning techniques," Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 79–86.
[8] T.Khushboo "Mining of Sentence Level Opinion Using Supervised Term Weighted Approach of Naïve Bayesian Algorithm" ,Int. Journal. Computer Technology & Applications, Vol 3 IJCTA | MAY-JUNE 2012.
[9] N, Answer and A, Rashid "Feature Based Opinion Mining of Online Free Format Customer Reviews Using Frequency Distribution and Bayesian Statistics" Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference on 16-18 Aug.2010.
[10] 278. Qiu, Guang, Bing Liu, Jiajun Bu, and Chun Chen. Opinion Word Expansion and Target Extraction through Double Propagation.Computational Linguistics, Vol. 37,No. 1: 9.27, 2011.
[11] Qiu, Likun, Weish Zhang, Changjian Hu, and Kai Zhao.Selc: a selfsupervised model for sentiment classification. In Proceeding of the 18th ACM conference on Information and knowledge management (CIKM-2009).
[12] Raaijmakers, Stephan and Wessel Kraaij. A shallow approach to subjectivity classification, in Proceedings of ICWSM-2008, 2008. p. 216-217.
[13] Classification of Opinion Mining Techniques,by Nidhi Mishra and C.KJha.
[14] B.Liu 2010. Opinion Mining and Sentiment Analysis: NLP Meets Social Sciences", STSC, Hawaii.
[15] P.Turney 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proceeding of Association for Computational Linguistics, pp. 417--424.
[16] B. Liu, and J. Cheng, 2005. Opinion observer: Analyzing and comparing opinions on the web, Proceedings of WWW.
[17] X. Ding, B. Liu, and P. S. Yu, 2008. A holistic lexicon-based approach to opinion mining, Proceedings of the Conference on Web Search and Web Data Mining (WSDM).
[18] Alexander Pak and Patrick Paroubek, Twitter as a Corpus for Sentiment Analysis and Opinion Mining.
[19] (2001). LIBSVM: A library for support vector machines [Online]. Available: http://www.csie.ntu.edu.tw/cjlin/libsvm.