

A Survey on Opinion Mining and its Challenges

Rohit kumar¹ Mrs.R.Subhashini² Dr. G. Rosline Nesa Kumari³

¹U.G. Scholar, ²Assistant Professor, ³Associate Professor

^{1,2,3}Department of Computer Science & Engineering

Saveetha School of Engineering, Saveetha University, Chennai(Tamil Nadu), India

Abstract— Today opinion mining has become one of the latest emerging fields of technology, where people are becoming keen observers of the opinions. Opinion mining is a process of extracting the opinions of the users given when they buy some product or they have the knowledge related to that domain. Thus in this paper we have shown the various aspect elements of the opinion mining as a survey. There are various way in which the opinion can be analysed and retrieved, here we have surveyed on the technique called sentiment analysis which has various classifications in it. As the number of Internet users increases the reviews also keep on increasing, thus we have some major challenges which prevail in the opinion mining hence we have tried to classify the various problems related to it.

Key words: Internet, opinion mining, sentiment analysis, surveys, extraction

I. INTRODUCTION

This paper is based on the opinion mining and sentiment analysis, in today's world every human may be considered as an Internet user, and out ten at least six people use internet daily for satisfying all need, the virtual digital market is growing day-by-day and the users are also increasing rapidly. Thus users now a day tends to buy their entire product through Internet in order to save their time and resources, it also ensures that they are satisfied with the products what they buy. As the number of online shopping websites are increasing, the user generated reviews based on these product also increases with time. When a person buys a product he/she first checks for the reviews upon that product. Based on the reviews he decides to buy the product. Due to rapid increase of the reviews of a single product it is becoming more difficult for a user to get the correct opinion of the product. The basic opinion mining was done through the search engines which are not a good technique, the search engine will produce too many links of a single product given by the user and thus it is a waste of time visiting each link one by one. The second problem associated with the opinion mining using search engines is that it produces too many irrelevant information which are of no use. In some cases people get through some reviews but these reviews are very long due to which user uses his interest on going through those reviews. There are various other challenges which are faced during opinion mining and create hindrance for the users to get the desired information.

II. FUNDAMENTALS OF OPINION MINING

A. Model of Opinion Mining

The model of opinion mining includes sharing of reviews and comments by the user on any type of product, these opinions can be used by the other internet users while siting for reviews about the same product. The product here is termed as the object on which the opinions are given. Thus for an object X, y and z can be given as its constituents and the sub constituents. In this model each y that is the constituent has its own sub constituents present. These characteristics can also be termed as features of the

product. the features plays an important role in determining the opinions about a product.. thus features can be used as a great source of opinion mining. Collecting all the features of the model and defining them as an model is also termed as feature based opinion model. Thus the opinion.

B. Mining Opinion Based on Comparative And Superlative Sentences

Basically there are two types of sentences in which opinions are expressed by the reviewers, first is the sentence or the phrase will describe the feature in a subjective, which means there will be no comparison words used in the phrase, for example "The sound quality of Samsung S4 phone is so good", here the feature that is the sound quality is described for the object phone. Such types of opinions are called as direct opinion sentences. Secondly we have some reviews which are given by people in a comparative way, for example "the camera of S5 phone is better than S4" thus here better word represent a comparative word and hence these types of opinions are called as in direct opinions or comparative statements.

III. DATA SOURCES IN OPINION MINING

There are various kinds of data sources used in opinion mining, these data sources are used to extract the relevant opinions based on the reviews given by the users. The data sources are used to identify various orientations in which the opinion has been declared.

A. Blogs

Blogs are the oldest approach towards opinion mining, the blogs are written by many people, these bloggers have the daily works of uploading events based on their real life time experiences. They also pay attention and write on different aspect and issues related to the product or any kind of substances. They tend to give the relevant information by expressing their emotions and feelings on various issues. But today as the number of internet users are increasing the number of bloggers and their sites are also multiplied, hence to find the relevant information such as about a product it is now a very difficult task, because in blogs the user has to go through the whole content to understand about the various features of the product. model is deduced to get the basic understanding of the working of the opinion mining.

Some blogs may contain the content which is of no use hence it is a mere waste of time reading all the blogs.

B. Review Sites

There are various kinds of review sites emerging day by day which gives the whole collection of reviews based on any kind of product. Every user wants that the product that is bought is free from all kinds of defects so that their money is not wasted, thus to make a decision to buy the products online is one of the difficult decisions, when the user is unaware about the product. Therefore in this type of circumstances the user gets dependent on various review sites. The reviews given by these review sites are a major constituent for studying of various sentiment analysis classifications. The e-commercial websites such as

“amazon.com”, “justdial.com” are some websites which provides the reviews on the product which they sell. Thus there are some specialized websites present for the reviews they are “znet.com” and “dpreview.com”.

C. Data sets

There are some types of reviews which are classified on the basis of features of the object. The data sets are also known as multi domain datasets which can be accessed online. Thus features play an important role in classifying of reviews in data sets. As a general the data sets referred to as a set of collected reviews sorted on the basis of object features, which can be used by the user to get an idea of the product features.

D. Micro-Blogging

Micro blogging generally refers to the mining of opinions through social media; these social sites such as Facebook, Twitter have become one of the prominent ways for opinion mining. The social networking sites provide a tool or facility of micro-blogging; it is a free format of text writing space provided by these sites for the reviewers to express their thoughts and comments on the related aspect. Due to this large amount of internet users have shifted towards the micro blogging technique rather than using conventional blogging techniques. This helps the user to write their reviews in short and precise which only contains the relevant content and saves the writer's time from writing the length reviews.

IV. SENTIMENT ANALYSIS

Sentiment analysis is one of the common terms associated with the opinion mining. Sentiment analysis deal mainly with the extraction of opinions, as we saw above there are numerous data sources available for the opinions, thus the work of sentiment analyser is to extract these sentences and phrases to get the correct meaning in which the opinion has been defined. The user generated comments signifies three kinds of meanings or sentiments that are positive, negative, or neutral. There are three levels in which the sentiment analysis can be done, document level, sentence level and entity /aspect level.

A. Document Level

In this level of classification the whole document is considered as an opinion, and classified based on the sentiments it expresses as positive or negative. For an example the review of the product is considered which whole document of opinion has either expressed in positive or negative sentiments. This type of classification is called as document level classification.

B. Sentence Level

In this type of technique sentiments are analysed based on the sentence or phrase level, each sentence in the document describes the positivity or negativity of the sentences these sentences then can be referred as a positive opinion or a negative one given on the various product.

C. Entity/Aspect Level

The entity here is referred to as targets, instead of looking for sentences, words, or phrases the targets are analysed in an opinion. By the help of these targets we are able to analyse the various statements and which helps us to understand the problem of sentiment analysis better thus the above mentioned two level are not convenient for analysing

the sentiment problems and hence we shift to the entity level sentiment analysis. These targets can also be called as features of the given product.

problems and hence we shift to the entity level sentiment analysis. These targets can also be called as features of the given product.

V. CHALLENGES IN OPINION MINING

Opinion mining is one of the emerging fields and thus have various types of difficulties present in this concept. As the opinion mining just cuts data from any source and tries to find the opinions on it, but these data are present in unstructured form which has many kinds of defects in it. The various types of challenges are described below.

A. Object Identification

There are various users over internet who gives their reviews based on the objects and their features and hence it is difficult to determine the exact features in the context of objects. The correct objects have to be identified and the irrelevant objects should be removed in order to get an effective opinion sets. Thus object identification is one of the major challenge prevailing in opinion mining.

B. Feature Extraction

Features are the characteristics which define the object, these are generally defined using various nouns and noun phrases, these nouns are frequently used and need to be identified. The pictures of Nikkon camera are very clear here the object feature is picture clarity and hence identifying the features is a very difficult task. There are many other types of methods which are used for extracting the information from the opinions. Such models are conditional random fields (CRF), hidden markov model (HMM). Features are said to be the basic identifiable elements in the opinion mining but they can be retrieved only if the comments present on the features in an structured way that is the reviews which is made on a particular feature should explain about the features in a clear manner which make the work of the extraction easy, due to some kind of.

C. Grouping Synonyms

The features of the object plays a major role in determining the opinions, a single opinion can be expressed by different types of words or phrases. these words which express the features one or more time should be grouped together. This becomes a difficult task, to identify these types of words and group it together. For example a sound quality of can be expressed by either sound or voice which becomes two similar words for the same feature, thus it creates a confusion in sorting these synonyms and grouping it together.

D. Opinion Oriented Classification

In this type we need to identify the orientation of the opinions, which means whether the opinion features are determined as positive, negative, or neutral. For example “the voice of my phone is not clear”. In this opinion the voice is the feature of the object phone which is given in negative. Thus these types of phrases determine various orientations present in an opinion. Usually lexicon based approach is used to determine the opinion orientations; this method basically uses the phrases and sentences to extract the orientation of an opinion on an feature. A word could be considered positive in one situation and in another case it may give a negative meaning, hence it makes difficult

classifying an opinion into positive, negative, or neutral in an opinion mining. Thus the task of extracting opinions from different ways leads to the following challenges as mentioned above.

VI. CONCLUSION AND FUTURE WORK

This paper gives a quick survey on the various aspects of opinion mining, it helps us to understand the various ways in which opinions can be detected and extracted from different types of comments and reviews, next we have the sentiment analysis which shows us how the opinions are classified and can be extracted, it shows the different levels of sentiment analysis, which helps us to get the sentiments, word and phrases the feature extraction gets difficult and becomes a hindrance in opinion mining.

concept of sentiment analysis more clear. Lastly we have discussed about the various types of challenges in opinion mining, these challenges are the difficult tasks which have to be performed for the opinion extraction but the user is not sure whether the desired outputs will be produced. The task such as orientation analysis, feature extraction is one of the major challenges faced in the field of opinion mining. Our future work suggest that instead of collecting reviews from different sites, and extracting it using various methods, we can generate a form survey, this form survey contains a form with different sets of questions which have to answered by the reviewer these review forms could be saved in an database and a web host can be generated on which these forms can be used as data for the opinion analysis and on the basis of weightage given to the sentiments such as positive, negative, or neutral the correct results about the product can be determined.

VII. ACKNOWLEDGEMENT

This paper work was done under the guidance of my mentor Dr. G. Rosline Nesa Kumari and my guide Mrs.R.Subhashini. I express my gratitude in thanking the two personalities as mentioned above, for guiding and supporting me for completion of this paper successfully.

REFERENCES

- [1] Dr. M S Vijaya, V Prema Sudha " Research Directions in Social networking mining with Empirical Study on Opinion mining", CSI Communication Dec2013 pp 23-26
- [2] Liu, B.: Sentiment analysis and opinion mining. Morgan & Claypool, San Rafael (2012)
- [3] Dave K., Lawrence, S. & Pennock, D.M. (2003), "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews". In Proceedings of the 12th International Conference on World Wide Web, p. 519-528.
- [4] Hu, M. & Lui, B. (2004), "Mining and Summarizing Customer Reviews". In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2004 (KDD-2004), p. 168-177.
- [5] Pang, B. & Lee, L. (2004), "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts". In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics, p. 271-278.
- [6] Gamon, M. (2004), "Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors, and the Role of Linguistic Analysis". In Proceedings of the International Conference on Computational Linguistics (COLING 2004), p. 841-847.
- [7] Soujanya Poria, Alexander Gelbukh, Amir Hussain, Dipankar Das, Sivaji Bandyopadhyay, "Enhanced SenticNet with Affective Labels for Concept-based Opinion Mining", IEEE 2013
- [8] N.Anwar, A. Rashid, S.Hassan, "Feature Based Opinion Mining of Online Free Format Customer Reviews Using Frequency Distribution and Bayesian Statistics", IEEE.
- [9] Cui, H., Mittal, V., Datar, M.: Comparative Experiments on Sentiment Classification for Online Product Reviews. In: Proceedings of AAAI-2006, pp. 1265-1270 (2006)
- [10] Chaovalit, P., Zhou, L.: Movie Review Mining: A Comparison Between Supervised and Unsupervised Classification Approaches. In: Proceedings of the 38th Annual Hawaii International Conference on System Sciences, pp. 112-121 (2005)
- [11] Moghaddam, S., Ester, M.: On the Design of LDA Models for Aspect-based Opinion Mining. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 803-812 (2012)
- [12] Mishne, G., Glance, N.S.: Predicting Movie Sales from Blogger Sentiment. In: Proceedings of the 21st National Conference on Artificial Intelligence, pp. 11-14. AAAI Press, Boston (2006)
- [13] Sik Kim, Y., Lee, K., Ryu, J.-H.: Algorithm for Extrapolating Blogger's Interests through Library Classification Systems. In: Proceedings of the IEEE International Conference on Web Services, Beijing, China, September 23-26, pp. 481-488. IEEE (2008)
- [14] Liu, Y., Huang, X., An, A., et al.: ARSA: A Sentiment-Aware Model for Predicting Sales Performance Using Blogs. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 607-614. ACM Press, New York (2007)
- [15] Sadikov, E., Parameswaran, A., Venetis, P.: Blogs as Predictors of Movie Success. In: Proceedings of the Third International Conference on Weblogs and Social Media, pp. 304-307 (2009)
- [16] Liu, F., Wang, D., Li, B., et al.: Improving Blog Polarity Classification via Topic Analysis and Adaptive Methods. In: Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL, pp. 309-312 (2010)
- [17] Liu, F., Li, B., Liu, Y.: Finding Opinionated Blogs Using Statistical Classifiers and Lexical Features. In: Proceedings of the Third International ICWSM Conference, pp. 254-257 (2009)
- [18] Chmiel, A., Sobkowicz, P., Sienkiewicz, J., et al.: Negative emotions boost user activity at BBC forum. *Physica A* 390(16), 2936-2944 (2011)
- [19] Softic, S., Hausenblas, M.: Towards opinion mining through tracing discussions on the web. In: Social Data on the Web Workshop at the 7th International Semantic Web Conference (2008)
- [20] Go, A., Bhayani, R., Huang, L.: Twitter Sentiment Classification using Distant Supervision. CS224N Project Report, Stanford (2009)
- [21] Pak, A., Paroubek, P.: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In: Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC), Valletta, Malta, pp. 1320-1326 (2010)
- [22] Barbosa, L., Feng, J.: Robust Sentiment Detection on Twitter from Biased and Noisy Data. In: Proceedings of the 23rd International Conference on Computational Linguistics. Posters, pp. 36-44 (2010)