

Obstacles and Challenges of Opinion Mining of Customers Reviews

Mubarak Himmat¹, Naomie Salim²

Faculty of Computing, Universiti Teknologi Malaysia, Malaysia.

barakamub@yahoo.com

Abstract: The customer's reviews and opinion mining become one of the wealthy areas in data mining. Nowadays, as the vast developments of using web applications and spreading of sites that provides a good platform for the customers to express their opinions directly on online shopping and companies web sites like Amazon.com, Cnet.com, customers opinion becomes helpful tools to manufactories to evaluation and measure satisfying ratio and weakness of the products [3]. Recently, there are many researcher works on this area of opinion mining, using different techniques, the developed techniques are good but also still there are many challenges and obstacles are still found. In this paper, we try to review some of the feature base opinion mining base concepts and techniques that are used now, and the challenges that face the opinion mining, and what the future works need to solve is.

Keywords: Data mining, sentiment analysis, opinion mining, Natural language processing.

1. Introduction

We can say that this is the era of the social media and wide spreading of its applications for that more people to share information and easily post reviews about their opinion and comments on the products, politic issues, books, and so forth. The posted opinions of people are important and helpful to companies and customers in past when we need to buy product or book or any other object, we need to hear and ask colleagues, friends, family members and others about this product before making decision of buying but now we need not to do that we just could visit a forums, blog, tweeter and company web sites to find users opinions are they satisfied it, what are it's good and weak features of products what are their recommendations and this is also an important issue to the companies to hear opinion about products and measure the customer satisfaction and also measure the strengths and weaknesses of the product by focusing on customers reviews. Because a new product development cannot only be pursuant to the business of the design and manufacturing capability one also has to consider the customer's needs and preferences and translate then into the design stage. When companies developing new products, they must be fully aware of the needs of customers, and really the model of product development driven by sales has been gradually replaced by the customers and market orientation. If an enterprise can exactly understand what the customer wants, preferences, and buying, for that most care about to study the people's opinions, attitudes appraisals, and emotions toward entities. For all that opinion mining opened its doors and challenges in this area of data mining. Sentiment or opinion mining concerned of that issues and there are many research done in this area and now the opinion of customers become major feedback tools to the future trend of product improvement. But really there are still many challenges faced the opinion mining.

In this paper we will discuss some of opinion mining techniques that used and then demonstrate the challenges that face opinion mining and try to share some idea of sentiment analysis and how it is difficult and also try to predict future challenges.

2. People opinion

While much work has recently focused on sentiment analysis or opinion mining and attitude of reviewers on social media in order to get dive in what people thinking about products and what are the features that they preferred or unsatisfied with, using NLP techniques, still many challenges are faced opinion mining, for opinion mining are opinionated and wrote as text and the available text mining systems are originally designed for regular kinds of texts of opinion, and some of the opinionated text are not regular. News methods may need to be adapted to deal with this type of text. The Natural Language Processing and it applications are represented some useful tools for sentiment analysis and it also faced some difficulties in some aspects of documents, because of different style of opinion holder thinking and their way of writing. During this paper we will try to identify some of these aspects. Most of the previous works dealing with Opinion perspectives and it determined the polarity and orientation of the opinion they tried to predict the polarity by analyzing and extracting the opinion documents sets, and some others works take the adjectives and subjective of opinion and then tried to find the polarity of the opinion and we will mention some basic of opinion classification as introduction of this reviews paper.



2.1 Customer Opinion classifications

Customers express their opinion on their own viewpoints, skill of writing, and thought and these Regular opinions or Sentiment expressions on some target entities can be divided into the following:

➤ **Direct opinions:**

This type of opining we can called explicit if a feature or any of its synonyms appears in a sentence, the feature could be identified as explicit or direct opinion [3]. The explicit features are features which appearing directly in a reviews. Such as: "The speed of the phone is slow."

➤ **Indirect opinions:**

This type of opining we can called implicit If a feature or any of its synonyms dose not appears in a sentence, The feature could be could be identified as explicit or in direct opinion [3]. The implicit features are that features which not appearing directly in reviews. Such as: "my friend said that you lost your money by purchasing this phone".

➤ **Comparative opinions:** Comparisons of more than one entity.

This kind of opinion the customers or reviewers make a comparison of similar products. E.g.: "I Samsung is better than Nokia."

2.2 Opinion orientation and extraction

When customers comments or reviews bout some products, their reviews will be taken as positive or negative or natural, this is the process of determine the opinion polarity.

The extraction process is to catch and find the words of pinion in opinions documents, the opinion holders, and the contextual information should be considered as clues during extracting opinion sentences and determining their polarity. Therefore, the extraction algorithm are to detect sentiment words firstly, then identifying the opinion polarities of sentences and then finally detect overall documents.

2.3 Sentiment analysis and classifications

Sentiment analysis is the procedure by which information is extracted from the opinions, appraisals and emotions of people in regards to entities, events and their attributes. In decision making, the opinions of others have a significant effect on customers, Facilitate the purchase process and making choices regards to online shopping, and choosing events, products, entities, etc. [1]. And these sentiment or opinion can be classified into the following categories.

2.3.1 Document level

Document level classify a whole opinion document (e.g., a review) based on the overall sentiment of the opinion holder [2] to find check the polarity of the opinion Positive, negative or neutral.

2.3.2 Sentence level

Classifying the Document to sentence and determine the polarity of each sentence and detect the overall opinion polarity [2] to C Positive, negative or neutral.

2.3.2 Corpus passed approach

Corpus-based approaches this method is determine and find co-occurrence of the patterns of words or phrases [8].

2.3.4 Dictionary base approach

Dictionary-based approaches use synonyms and antonyms in WordNet to determine word sentiments based on a set of seed opinion words [3].

2.4 Feature based opinion mining

Classifying evaluative texts at the overall opinion or as document level, and also evaluating it in sentence level is not accurate way of evaluating customers opinion, polarity likes and dislikes. A positive overall document on product or object does not mean that the opinion holder has positive opinions on all aspects or features of the object. Likewise, a negative document does not mean that the opinion holder dislikes everything about the object. In an evaluative document such as a customer reviews of a product, because every opinion holders express their both positive and negative feature of the product or the object.



Although the general sentiment on the object may be positive or negative, to obtain such detailed aspects, Feature-based opinion mining has been proposed [3] to summarize the overall opinion. And word or feature level sentiment analysis becomes importance by applying the natural language processing and its different methods. And there are many researchers worked on extraction of features and opinion words using a predefined seed word list for extracting and applying semantic orientation and opinion classification, and here, the individual words or phrase must be extracted as from reviews and used Natural Language Processor Linguistic Parser to parse each review, split text into sentences and to produce (POS) Part Of Speech tags for each word like noun, verb, adverb, adjective, etc. They use WorldNet to find different senses of the same term.

3. Experimental issues of Feature based opinion mining

To have a deep look in reviewer's opinions and discover every aspects of what opinion contains, some researchers have tried to mine and extract opinions at the feature level [7]. The task of opinion mining at the feature level needs to identify the individual elements of the opinion text. This means to define each individual words or phrases and extracted them as elements from reviews sentences. This task was regarded as an information extraction (IE) problem, which inherently handles the extraction of needed information from unstructured text: relation extraction, terminology extraction and named entity recognition [6]. The purpose of relation extraction is to find out the relationship between the elements. The terminology extraction is to extract the relevant terms from a given corpus [7], while the named entity recognition aims to classify the extracted elements into a specific entity category such as location, person, etc. These three tasks can be in nature related to feature level opinion mining for product reviews. For example, the step of discovering the specific entities (e.g., features, functions from a digital camera's reviews) can be related to the named entity recognition. The mining of opinions that are associated with these entities can be then taken as a variation of the relation extraction problem [4].

3.1 Features based techniques

There are many techniques use feature based opinion mining here we will discuss some of these techniques.

3.1.1 A Method for Opinion Mining of Product Reviews using Association Rules

This technique proposed opinion mining method for product reviews. In this approach, firstly they used POS tagging techniques on each review sentence, and then extract features and opinion words in form of transaction data. Then discover association rules of needed type from the transaction data, and provide information that is summarized advantages and disadvantages using The Pointwise Mutual Information (PMI)-IR algorithm. Four types of association rule are extracted from product reviews. The type rule one indicates overall opinion of customers for product, this rule is useful for customers wanting to know about satisfaction of a product, the second rule indicates features that appear frequently in product reviews. This frequently appeared feature in product reviews is important information about product. The third rule indicates opinion by each feature. This rule is useful to customers who wanting to know about each feature of a product, the fourth rule is a combination of type second rule and third rule. This rule is used to determine whether the feature is an advantage or a disadvantage using APRIORI algorithm [9].

3.1.2 Mining Feature opinion technique.

These methods proposed for mining product feature and opinion based on the consideration of syntactic information and semantic information. By applying dependency relations and ontological knowledge with probabilistic based model, the result of this methods experiments shows that this approach is more flexible and effective.

3.1.3 Sentiment Classification from Online Customer Reviews Using Lexical Contextual Sentence Structure.

The proposed method classifies subjective and objective sentences from reviews and blog comments. The semantic score of subjective sentences is extracted from SentiWordNet to calculate their polarity as positive, negative or neutral based on the contextual sentence structure. For classifying and analyzing sentiments from online reviews and blog comments, they use lexical contextual information at the sentence level to check whether sentences are objective or subjective, and to classify subjective sentences into positive, negative or neutral opinions. They used a machine learning algorithm for classifying sentences into objective and subjective and for finding their polarity. In their work, they proposed a rule based lexicon method to determine subjectivity from objectivity sentences. From subjective sentences, they extract the opinion expression and check their semantic scores using the SentiWordNet directory. The final weight of each individual sentence is calculated after considering the whole sentence structure, contextual information and word sense disambiguation [1].

3.1.4 Feature-based opinion mining and ranking

This ranking method proposed an opinion mining algorithm called Ask Us [5]. This method proposed a search engine that can assist potential customers to make a wise decision based on the features of the products being offered by each company. It can also help companies understand which feature of the product is well received by the customers and which is not. The method classifies a review as positive, negative or neutral and also identifies the most representative features and assigns overall weights to each one of them, and classifies each feature as positive, negative, or neutral in various levels of importance and presents the most important ones to the end user. At the end the algorithm summarizes and ranks the opinions about these product features by giving those scores



4. Evaluation

All the mentioned techniques are following similar steps of featured based approach when trying to find automatically the product features and polarity (negative, Positive, neutral) of opinions that are expressed by customers. Using all the mentioned techniques are good in some manners and could achieve a good results but really opinion mining is still facing difficulties for opinion is subjective for there are no standard of expressing of writing, every opinion holder expresses his opinion by using his own words and they are not care about the syntax of language, when writing opinion and they wrote opinions are not clear and concise enough to understand by using the recently available methods of natural language processing NLP methods of sentences and words detections (POS). and also using slang words, sarcasm words, negation, abbreviations and spelling mistakes are difficult recognize and identified when it compared with WorldNet, also the in direct sentence and the comparison sentence causes ambiguous to understand by NLP, Reliability in the opinion holder, because some time there are many opinion spam in product reviews, those spam is written to draw a bad impression of product from Competitors companies or to embed a good comment to attract people to their products and these need more research to solve these problems in future.

5. References

- [1]. Aurangzeb Khan, Baharum Baharudin, Khairullah Khan, "Sentiment Classification from Online Customer Reviews Using Lexical Contextual Sentence Structure", J. M. Zain et al. (Eds.): ICSECS 2011, Part I, CCIS 179, pp. 317–331, 2011 © Springer-Verlag Berlin Heidelberg 2011.
- [2]. Pang, B., Lee, A. L., "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts", In: Proceedings of the 42nd ACL, November 16-24.
- [3]. Hu, M., & Liu, B. (2004), "Mining opinion features in customer reviews, in AAAI 2004 (pp. 755–760).
- [4]. Li Chen, Luole Qi, Feng Wang, Department of Computer Science, Hong Kong Baptist University, Hong Kong, China, "Comparison of feature-level learning methods for mining online consumer reviews", L. Chen et al. / Expert Systems with Applications 39 (2012) 9588–9601.
- [5]. Magdalini Eirinaki, Shamita Pisal I, Japinder Singh ,M. Eirinaki , Feature-based opinion mining and ranking/ Journal of Computer and System Sciences 78 (2012) 1175–1184.
- [6]. Cowie, J., & Lehnert, W. (1996), "Information extraction. Communications of the ACM, 39, 80–91.
- [7]. Bourigault D., & Jacquemin, C. (2009), "Term extraction + term clustering: An integrated platform for computer-aided terminology", In Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics (pp. 15–22), Association for Computational Linguistics.
- [8]. Yu, H., Hatzivassiloglou, V., "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences, In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2003), USA, pp. 129–136 (2003), doi:10.3115/1119355.1119372.
- [9]. Agrawal, R. and Srikant, R. 1994, "Fast algorithm for mining association rules" VLDB 94, 1994.

