Opinion Mining and Sentiment Analysis - Its Tools and Challenges

Surabhi Thorat, MCA Department, Marathwada Institute of Technology(E), Aurangabad(M.S), India

ABSTRACT

The explosive growth of social media also sparks today’s competitive scenario of market that always concern about the sales growth chart of their products. The businesses always want to find public or consumer opinions about their products and services. Potential customers also want to know the opinions of existing users before they use a service or purchase a product. The customers opinion can be taken online by means of customer feedback, internet forums, customer reviews, comments, newsgroups, post, discussion groups or blogs which is collectively called user generated contents. Users express their views and opinions regarding products and services that lead to overcome the requirements of marketing intelligence and product benchmarking in the production industry. These opinions are subjective information which represents user’s sentiments, feelings or appraisal related to the same. Gaining information effectively and easily for businesses and individuals is a technically a very challenging task and at the same time practically beneficial too. It requires lot of natural language processing tasks. In this paper, I tried to analyze the basics of opinion mining, key players of opinion mining, different text evaluation task involved in opinion mining, tools and challenges of opinion mining.

Keywords

Opinion Mining, data source, opinion mining model, major players, text evaluation task, tools, challenges.

I. INTRODUCTION

In the last few years, web documents arise as a matter of great interest that describes individual experiences and opinions. This situation focuses on the techniques for automatically extracting or analyzing personal opinions from web documents such as posts on message board and weblogs. Such technologies can be an alternative to traditional questionnaire-based social or customer research and would also benefit Web users who seek reviews on certain consumer products of their interest.

Opinion mining is a combination of information retrieval and computational linguistic techniques deals with the opinions expressed in a document [1]. Opinion Mining or Sentiment Analysis is the field to extract the opinionated text datasets and summarize in understandable form for end user [2]. The opinion is the subjective expression which describes people’s opinions, emotions and sentiments towards entities and their properties, particular topics, product or services. Opinion mining is to extract the positive, negative or neutral opinion summary from unstructured data.

II. DATA SOURCE

User opinion is one of the major criterions for the improvement of the quality of services. Blogs, review sites, data and micro blogs provide a good understanding for the deliverable level of the products and services provided to customers.

A. Blogs

As internet usage is increasing day by day, blogging and blog pages are growing rapidly. Blog pages contain the expression of one's personal opinions. Many of these blogs contain reviews on many products, issues, etc. Blogs act as one of the sources of expressing opinion in many of the studies related to sentiment analysis[3]. The name associated to universe of all the blog sites is called blogosphere. People write about the topics they want to share with others on a blog. Blog pages [4] have become the popular means to express one’s personal opinions about any product or topic.

B. Review sites

Previous buyer's comments on any particular product plays a vital role in decision making related to purchase of the product. The user generated reviews for products and
services are largely available on internet. The data given by reviewers are collected from the e-commerce websites like www.amazon.com (product reviews). A large number of user-generated reviews are available on the Internet. The reviewers data used in most of the sentiment classification studies are collected from the e-commerce websites [5] like www.amazon.com (product reviews).

C. Data Set
The work in the field uses movie reviews data for classification. The dataset contains different types of product reviews extracted from Amazon.com including Books, DVDs, Electronics and Kitchen appliances.

D. Micro-blogging
A very popular communication tool among Internet users is micro-blogging. Millions of messages appear daily in popular web-sites for micro-blogging such as Twitter, Facebook. Twitter messages sometimes express opinions which are used as data source for classifying sentiment. [6]

III. OPINION MINING MODEL

In general, opinions can be expressed on anything, e.g., a product, a service, a topic, an individual, an organization, or an event. Opinion from different sources form a text database that raises the need of web mining. That involves extraction and computational linguistics techniques by means of opinion mining. A last opinion facts database is created that provide informational data about opinion facts and trends.

IV. MAJOR PLAYERS IN OPINION MINING

A. Object:
An object $O$ is an entity which can be a product, a service, a topic, an individual, an organization, or an event. It is associated with a pair, $O: (T, A)$, where $T$ is a hierarchy or taxonomy of components (or parts) and sub-components of $O$, and $A$ is a set of attributes of $O$. Each component has its own set of sub-components and attributes. In this hierarchy or tree, the root is the object itself. Each non-root node is a component or subcomponent of the object. Each link is a part-of relationship. Each node is associated with a set of attributes. An opinion can be expressed on any node and any attribute of the node. However, for an ordinary user, it is probably too complex to use a hierarchical representation. To simplify it, the tree is flattened. The word “features” is used to represent both components and attributes. Using features for objects (especially products) is quite common in practice. Note that in this definition the object itself is also a feature, which is the root of the tree.

Let an evaluative document be $d$, which can be a product review, a forum post or a blog that evaluates a particular object $O$. In the most general case, $d$ consists of a sequence of sentences $d= \langle s_1, s_2, \ldots, s_m \rangle$.

B. Opinion passage on a feature:
The opinion passage on a feature $f$ of the object $O$ evaluated in $d$ is a group of consecutive sentences in $d$ that expresses a positive or negative opinion on $f$. This means that it is possible that a sequence of sentences (at least one) together expresses an opinion on an object or a feature of the object. It is also possible that a single sentence expresses opinions on more than one feature, e.g., “The picture quality of this camera is good, but the battery life is short”.

C. Opinion holder:
The holder of a particular opinion is a person or an organization that holds the opinion. In the case of product reviews, forum postings and blogs, opinion holders are usually the authors of the posts. Opinion holders are important in news articles because they often explicitly state the person or organization that holds a particular opinion [7].

V. TEXT EVALUATION TASK

The text evaluation task which is the documents that expresses the opinion is broadly divided into 3 categories:

A. Sentiment classification:
In this task opinion mining is treated like text classification problem. It classifies an evaluative text as being positive or negative. The classification is usually done at the document-level. No details are discovered about what people liked or didn’t like. Sentiment classification is being defined as a given set of evaluative documents $D$, it determines whether each document $d \in D$ expresses a positive or negative opinion (or sentiment) on an object. For example, given a set of product reviews, the system classifies them into positive reviews and negative reviews.

B. Featured-based opinion mining and summarization:
This task works on the sentence level to discover details. It focuses on extracting the features of the commented object and after that determine the opinion of the object i.e. positive or negative and then group the feature synonyms and produce the summary report. Liu [8] used supervised pattern learning method to extract the object features for identification of opinion orientation. To identify the orientation of opinion he used lexicon based approach.

For example, in a product review, this task identifies product features that have been commented on by reviewers and determines whether the comments are positive or negative. In the sentence, “the processing speed of this processor is very slow,” the comment is on the “processing speed” and the opinion is negative. A structured summary will also be produced from the mining results. This approach basically uses opinion words and phrase in a sentence to determine the opinion. The working of lexicon based approach [9] is described in following steps.

- Identification of opinion words
- Role of Negation words
- But-clauses

C. Comparative sentence and relation mining:
Comparative sentence and relation mining directly compares one object against one or more other similar objects. For example, the following sentence compares two cameras: “the picture quality of camera A is much poor than that of camera B.” We want to identify such sentences and extract comparative relations expressed in them. The comparison in a comparative sentence is usually expressed using the comparative or the superlative form of an adjective or adverb [10].

VI. OPINION MINING TOOLS

The tools which are used to track the opinion or polarity from the user generated contents are:

A. Review Seer tool:
This tool is used to automate the work done by aggregation sites. The Naive Bayes classifier approach is used to collect positive and negative opinions for assigning a score to the extracted feature terms. The results are shown as simple opinion sentence [11].

B. Web Fountain:
It uses the beginning definite Base Noun Phrase (bBNP) heuristic approach for extracting the product features. It is possible to develop a simple web interface.

C. Red Opal:
It is a tool that enables the users to determine the opinion orientations of products based on their features. It assigns the scores to each product based on features extracted from the customer reviews. The results to be shown with a web based interface [12].

D. Opinion observer:
This is an opinion mining system for analyzing and comparing opinions [13] on the Internet using user generated contents. This system shows the results in a graph format showing opinion of the product feature by feature. It uses WordNet Exploring method to assign prior polarity.

VII. CHALLENGES AND ISSUES

Opinion mining is a relatively new field where lot many work still need to done, thus there are several challenges to be faced like.

A. Object identification:
In opinion mining, firstly you have to identify the objects in a review on which opinion have been given. This problem is important because without knowing the object on which an opinion has been expressed, the opinion is of little use. However, there is a difference. In for opinion mining, only those objects in the review are to be considered which are in competition to each other. The system thus needs to separate relevant objects and irrelevant objects.

B. Grouping synonyms:
Different words or phrases can be used to refer to the same feature of the object. So, such words (synonyms) should be identified and grouped together. It is a difficult task to identify these words. A lot of research is required to be done on this issue as it has not been much addressed in the past. To produce a summary similar to the one in Figure 9, it is needed to group synonym features, as people often use different words or phrases to describe the same feature. In this example, “voice” and “sound” both refer to the same feature.

C. Identify comparison words:
Identify comparison words and whether they are giving positive or negative feedback totally depends on their context. So, it’s not an easy job as sometimes good are bad and bad are good.

D. Opinions Change with Time:
Another challenge lies in the issue of being able to monitor opinions changing with the passage of time. This helps us to observe if a certain product gets improved with time, or people change their opinion about a product and get convinced for it with time. A research work is done to identify how the people's mood changes over time in Reference [14]. The work done observes blogs where the mood is explicitly specified either by selecting from a predefined list of moods or by entering it as free text.
E. Misleading Opinions due to sarcastic and ironic statements:
Sarcastic and ironic sentences exist in text. In such a scenario, positive words can have negative sense of usage in a metaphorical manner. So, text in a statement can be hard to identify as sarcastic or ironic which can lead to erroneous orientation and misleading opinion mining. Reference [15] discusses this issue.

F. Strength of Opinions:
Identification of the strength of an opinion is another challenge faced in opinion mining. The strength of an opinion can change as the discussion continues in a forum i.e. arguments used during discussion are strong enough to change the strength of opinions. To identify strength of opinions SentiWordNet application has been used [16]. Reference [17, 18] addresses this issue.

VIII. CONCLUSION
Opinions play vital role in decision making. Opinion mining is an emerging field of data mining used to extract the precious knowledge from huge volume of customer comments, feedback and reviews on any product or topic etc.

Opinion mining is helpful for individuals when they want to buy a product and they can decide which product to buy, by studying the summarized opinions generated after opinion mining.

Based on these summarized opinions companies can modify their products according to customers opinions in a better and faster way. Thus, companies can establish better customer relationship by giving them exactly what they need. The companies can find, attract and retain customers; they can save on production costs by utilizing the acquired insight of customer requirements.

In this paper I tried to high lighten major concepts related to opinion mining which, opinion model, its tools and challenges in field of opinion mining.

REFERENCES
[6] Alexander Pak and Patrick Paroubek, Twitter as a Corpus for Sentiment Analysis and Opinion Mining
[10]Bing Liu, Web Data Mining Exploring Hyperlinks, Contents, and Usage Data