

Text Analysis of Users Feedback In Data Mining

#¹S.Selvaraj, #²B.Jegatheeskumar, #³D.Babukannan,

¹selvaraj10@gmail.com
²jegatheesbe@gmail.com
³babutec@gmail.com



¹Assistant prof Department of CSE,RVSETGI,Dindigul

²Assistant prof Department of ECE,RVSETGI,Dindigul

³Assistant prof Department of ECE,RVSETGI,Dindigul

ABSTRACT

Unlike large concern, start-up concern usually does not have the available resources to afford traditional mass marketing campaigns such as TV commercials or magazine advertisements. However social networking services (e.g., Facebook, Twitter etc). Provide a more economically more viable chance for these new companies to directly communicate with their potential users. Social networks are significant marketing communication device for established, and in particular, start-up concerns. Some FB Pages contain hundreds of responses and receive good opinions from the users whereas some Pages do not. By explain these Pages, this is possible to make the key factors that attract customers and for the young business executive to react to new postings. Text mining the Pages helps to greater understands and run their Pages and builds a closer connection with the target audience. This is the logical process to interpret the dialogue between young business executive and their audience of Facebook Pages. First, collect customer feedback from social networks, like FB. The explanation of the dialogues into significant statistics, especially when attempting to design, cluster, and analyze the critical elements of posted Internet satisfy, requires new text analysis techniques and methodologies. Then clustering is used to make the critical points that customers care about and then to explore key factors that attracts customers and resolves their needs. Therefore, business executive better understand how to apply an interesting topic to strengthen their marketing communications and increase their market share.

Keywords- customer feedback, text mining, Facebook, social networks,

ARTICLE INFO

Article History

Received: 27th December 2015

Received in revised form :29th December 2015

Accepted : 30th December , 2015

Published online :

31st December 2015

I. INTRODUCTION

With the rapid expansion of social network, many new companies create a Page on facebook and post news or topics that they want to share with their customers. They hope to build a closer relationship with their customer, just as facebook promotes on their Create a Page website. As time passes though, many start-ups find they haven't actually generated the results or sales desired and the facebook relations neither satisfy their customers or the company owners. Many companies set up facebook URLs or Pages for marketing communications, but most pages are a one-way form of communication, with limited operations and functions facebook pages can only send out the information or messages to the audience who press the like button of the Page and in return receive no information the marketplace, the product users, or the resent status or stories

about the product benefits unless the audience response. Therefore, the concern cannot study what their users want or care about unless their customers or audience post their opinions on the Page wall. In order to invoke the audience's interest to respond and gain more information about customers' attitudes and opinions, Page editor should know what the reasons that make the audience response are. Therefore, how to analyze the responses become research issue and a new approach is needed. This approach will change one-way marketing communications into a two-way communications dialogue critical for inventors and designers because they know how to communicate with the audience.

In general condition, users will also provide direct feedback and comments regarding the product or service they have, including both positive and negative aspect. Such

information will be very important to start-up companies, since the information is directly from the user side. When compared to traditional questionnaire distribution, such information will not need to set up boundaries or limitations for answering questions. User may give any suggestions or comments regarding the service or product. Utilizing such information will provide the company a customer-orientated overview of its competitors and also concerned factors. Besides, those responses are the critical comments the customers care most. Therefore, the critical incident technique can be applied to analyze that information.

II. EXISTING SYSTEM

Clustering is widely used for text mining, pattern recognition, webpage analysis, and marketing analysis. The purpose of clustering is to select elements that are as similar as possible within groups but as different as possible between groups. Groups are clustered depends on entities' similarity according to specified variables and the meanings of clusters depend on the context of the analysis. Clustering algorithms are usually used for exploring data structure based on defined data format. Whenever clustering algorithm evaluation analysis was utilized only for preprocessing, it is not necessary to evaluate clustering result. Existing approaches are based on different direct methods and indirect methods using opinion words and phrases and the grammar information. One key issue is to identify opinion words and phrases like good, bad, poor, or great, which are instrumental to sentiment analysis. However, there are seemingly an unlimited number of expressions. That people use to express opinions, and in different domains, they can be significantly different. Even in the same domain, the same word might indicate different opinions in different contexts.

DISADVANTAGES OF EXISTING SYSTEM

It depends on defined data format. One key issue is to identify opinion words and phrases such as good, bad, poor, or great

PROPOSED SYSTEM

In Proposed system, analyze the facebook post, comments, opinions and related phrases, then extracted key phrases and transformed user comments into key phrase matrix. Based on key phrase matrix, clustering algorithm was utilized to grouping similar key phrases together for further analysis. In order to decide the number of cluster within each test implemented an iteration process to calculate clustering quality that is an index composed by cohesion and separation. This perspective more accurately represents the diversity of the constituency groups participating in the FB page and closely aligns the analysis.

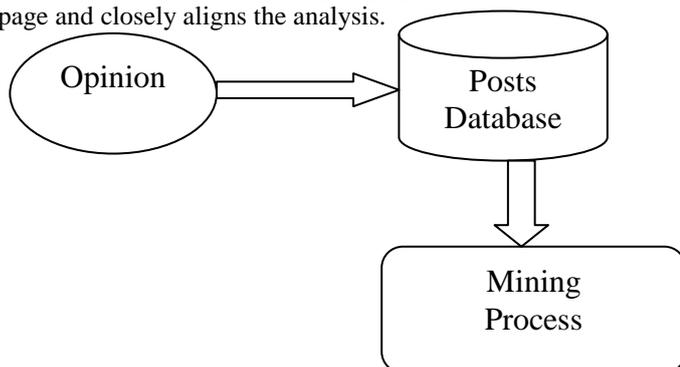


Fig-1

III. ADVANTAGES OF PROPOSED SYSTEM

Have to determine the type of opinion whether it is direct opinion or comparative opinion.

A. Key phrases and extraction

The data mining is the language processing application; machine learning and retrieval have to understand so words in a text so that the text can be processed. In order to extraction and ordering, each FB comments were transformed into a phrase matrix as shown in Table 1.

TABLE 1. PHRASE MATRIX

Facebook input	Key phrase			
FB comment 1	f11	f1 2	...	F1m1
FB comment 2	f21	f2 2	...	F2m2
...
FB comment p	fp1	fp 2	...	Fpmp

B. Clustering algorithm: K-means

In order to know the main theme that customers concern about, clustering is implementing. Based on the phrases frequency matrix and their similarity, we use K-means clustering analysis with the key phrase dimensions. The k-means algorithm is one of clustering algorithm implemented within WEKA, which is a method of cluster analysis aims to separate n samples into k clusters in which each sample belongs to the cluster with the nearest mean. Although clustering analysis is most often used for sample partition, it can also be used for key phrase features partition if we transpose the data matrix. Given a set of key phrases (k1, k2, ... , kn), where each key phrase is a d-dimensional real vector, k-means clustering aims to separate the n key phrases into k sets (k n) $S = \{S1, S2, \dots , Sk\}$ so as to minimize the within-cluster sum of squares (WCSS):

$$\operatorname{argmin}_S \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \quad (1)$$

where μ_i is the mean of points in S_i .

C. Clustering quality evaluation: cohesion and separation

In order to decide the number of clusters, we executed a clustering quality criteria composed of cohesion and separation. These two values would help us review the clustering result and decide the final structure. Cohesion and separation calculations help us understand the variations between clusters within cluster structure whereas cohesion evaluates variations within clusters and separation evaluates variations between clusters. Where C_i denotes each cluster within clustering result and similarity was defined by Euclidean distance. x and y represent key phrase pairs found within the cluster. The algorithm reviews each cluster and aggregates the cluster cohesions values as a synthetic clustering quality criterion. If there is only one key phrase within a cluster,

the cohesion of the cluster will be 1. Finally the average cohesion of clustering result is calculated as followresulting structure. A better clustering result will have a larger separation value. Each pair of clusters within the result will have a separation value. After summation of every cluster pair separation and dividing to the number of cluster pairs, the average separation of the clustering result is computed:

$$\text{Separation} = \frac{\sum_{i=1}^K \text{separation}(C_i)}{K} \quad (2)$$

For comparing cluster results, we use Cartesian coordinates to display both values. To combine cohesion and separation as an integrate index, we combine them with the following formula:

$$\text{Cohesion-separation} = \frac{\text{cohesion}}{\sqrt{\text{separation}}} \quad (3)$$

IV. IMPLEMENTATION AND RESULTS

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

MODULES

- Posting opinion
- Object identification
- Features extraction
- Opinion-orientation determination
- Integration

1) Posting opinion

In this module, we get the opinion from various people about business, e-commerce and products through online or facebook. The opinions may be of two types, direct and relative opinion. Direct opinion is to post an opinion about the element and attributes of products directly. Comparative opinion is to post an opinion based on comparison of two or more products. The opinions may be positive or negative

2) Object identification

In general, people can express opinions on any target entity like outcomes, services, creature, organizations, or events. In this module, the term object is used to denote the target entity that has been commended on. For each comment, we have to identify an object. Based on object, we have to integrate and generate ratings for opinions.

3) Feature extraction

An object can have a set of components or parts and a set of attributes or properties which we have collect all the report of the object. For example, a mobile phone is an object. It has a set of element such as battery and screen and a set of attributes such as voice quality and size, which are all called characteristic or aspects. An opinion can be indicating any feature of the object and also on the object itself. With these ideas in mind, we can define an object design, a model of an opinionated text, and the mining

objective, which are collectively called the feature-based sentiment analysis model.

4) Opinion-orientation determination

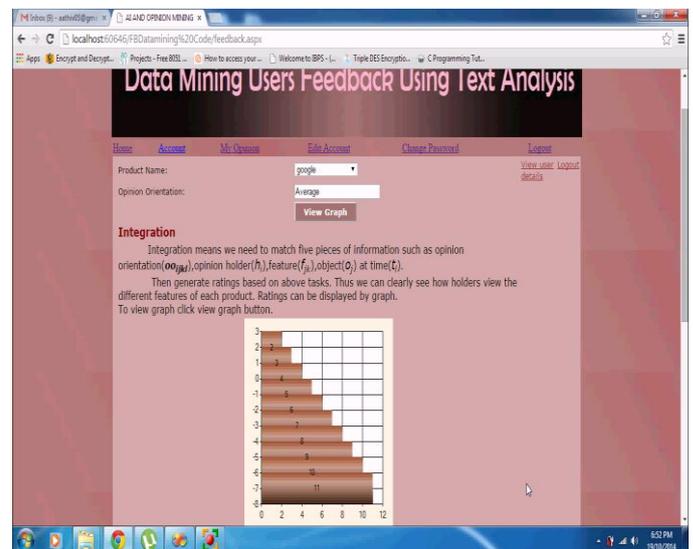
The opinion holder is the person or organization that expressed the opinion. In the case of product reviews and opinion holders are usually the authors of the posts. An opinion on a characteristic f or object o is a positive or negative view or evaluation on f or o from an opinion holder. Positive and negative comments are called opinion orientations. From this opinion direction we have to determine the type of opinion whether it is direct opinion. The opinion aim can be positive, negative, these are the comparative opinion. A comparative opinion conveys a preference relation of two or more objects based their shared features. A comparative opinion is usually fetched using the comparative or superlative form of an adjective or adverb, such as coke tastes better than Pepsi.

5) Integration

Integrating these tasks is also complicated because we need to match the five pieces of information in the quintuple. That is, the opinion must be given by opinion holder on feature of object at time. To make matters worse, a sentence might not explicitly mention some pieces of information, but they are suggested using pronouns, language conversion, and context. Then generate ratings based on above tasks. Thus we can clearly see how holders view the different features of each product.

V. RESULT

Finally all the comments are compared through the integration module, what we received the comments from customer through the social network. Then analyze the comments after that we conclude what the customer wants from the concern. This result also needed for company development by using the rating or integration modules.



VI. CONCLUSION

Thus set of techniques for mining and summarizing Facebook posts and comments based on data mining and natural language processing methods are proposed. Other than distribute questionnaires to end users, it is a method to provide a more direct and open-minded way for accessing customer's feedbacks, so start-up vendors can know who is the existing competitors on the market and also their succeed factors and features during decision making

process. The practical case study indicate that the proposed analysis process is very promising for analyzing data extracted from social network community site. Since more and more people using social network to communicate and to express their feedback and opinions on the Internet and proposed analysis process will become increasingly important to start-up companies who have few financial resources to do large scale market area. The key phrase clustering is to generate the key factors so the start-up companies know what the main concern that customers care about. In future, utilize the user information for further and deep learning analysis. Incorporating with user statistics, our analysis process would also connect user opinion to specific type of customers. Moreover the implementation of this analysis process which is used to start-ups of other type of industry.

REFERENCES

- [1] J.C. Flanagan, "The Critical Incident Technique," *Psychological Bulletin*, vol. 51, pp. 327-358, 1954. J.C. Flanagan, "The Critical Incident Technique," *Psychological Bulletin*, vol. 51, pp. 327-358, 1954.
- [2] R.A. Feinberg, K. DeRuyter, C.V. Trappey, and T.Z. Lee, "Consumer-Defined Service Quality in International Retailing," *Total Quality Management*, vol. 6, no. 1, pp. 61-67, 1995.
- [3] C.V. Trappey, T.T. Hsio, T.C. Chang, M.H. Che, W.J. Chiu, "Consumer Driven Game Design," *Proceedings of DiGRA 2005 Conference: Changing Views – Worlds in Play*, 2005.
- [4] Berkhin, P. (2002), "Survey of clustering and data mining techniques," *Technical Report*, Accrue Software, Inc.
- [5] Chen, B., Tai, P.C., Harrison, R., and Yi, P. (2005), "Novel hybrid hierarchical-K-means clustering method (H-K-means) for microarray analysis," *Computational Systems Bioinformatics Conference*, Aug. 8-11, StanfordCA, USA.
- [6] Berry, M. J. A. and Linoff, G. (1997), *Data Mining Techniques: For Marketing, Sale, and Customer Support*, John Wiley & Sons Inc, New York, NY, pp. 55.
- [7] Tan, P.-N., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining*. Addison Wesley, 1 edition.
- [8] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H., "The WEKA Data Mining Software: An Update," *ACM SIGKDD Explorations Newsletter*, vol. 11, No. 1, pp. 10-18, 2009.