PAK Publishing Group
Growing Knowledge for Future

Proceedings Book of ICEFMO, 2013, Malaysia
Handbook on the Economic, Finance and Management Outlooks
ISBN: 978-969-9347-14-6

Business Website Interface Design for Feature Based Review Posting and Handle Spam: Concise Model

Atika Qazi
University of Malaya Kuala Lumpur, Malaysia
Email: atika@siswa.um.edu.my

Ram Gopal Raj
University of Malaya Kuala Lumpur, Malaysia
Email: ram_prime@fsktm.um.edu.my

Muhammad Tahir
COMSATS Institute of Information Technology Islamabad, Pakistan
Email: Muhammad_tahir@comsats.edu.pk

Syed Karim Bux
University of Malaya Kuala Lumpur, Malaysia
Email: kbsayed@siswa.um.edu.my

Zia ud din Sheikh
COMSATS Institute of Information Technology Islamabad, Pakistan

Abstract

Opinion mining is a novel filed of research. Though a lot of work has been carried out on sentiment analysis, little work has done on feature analysis. In order to strengthen businesses in competitive market, there is a need for particular and precise feature analysis. There is a lot of online information available on web sites that offer the visibility of reviews to use. A minimum work is being done on developing the business websites from that companies may retrieve the data for analysis. Companies can apply opinion mining techniques identify users' reviews in order to improve their businesses. We present a model to analyze customers' opinion by providing an improved and convenient interface regarding hotels in Malaysia named featured based opinion posting. Such results can be used to find public perception in order to better understand business demands and full fill customer's need and strengthen economic values. The feature based review posting can minimize the work of feature extraction that is required for regular opinion posting and can be helpful in reducing spam. The feature based opinion posting may give related and concise data that is used for potential users and owners to improve the business strategies according to users' feedback.

Key words: Opinion Mining, Product Features, Spam Handling, Customers Feed Back

1. Introduction

Opinion mining (OM) is a novel field of information retrieval and computational linguistics. OM is concerned with the opinion that is expressed by opinion holder. Basic components of an opinion mining are (1) opinion holder, that holds a specific opinion on a particular object, (2) the object on which an opinion is expressed and (3) the opinion a view, attitude, or appraisal on an object from an opinion holder. The fact and opinions are two main categories of text[1].The facts are objective speech about entities and opinions are subjective speech that shows the sentiment about related entities .

The excessive research on text information processing is focusing on factual information e.g web search, information retrieval, natural language processing and text mining [1, 2]. The two major types are (1) Regular opinion and (2) comparative[3].

The reviews like simple review is about a single product to show its pros and corns i.e " The battery life of App cell phone is long" and other type is " I'm sad for football match performance by team A" other one is " The camera A picture quality is much better than camera B picture Quality" . In all such kind of reviews that are simple, emotional and comparative sentences there is entity and its attributes, objects and its features that need to explore with the polarity.

The main task of OM is doing this job to identify the features and sentiments in different kinds of sentences and make them useful. In different situations different type of mining is required to perform necessary action.

The opinions have great significance and commonly used by users and potential buyers to make any decision about a relevant entity. The organizations are more and more focusing on user generated text (opinions) to make future and present decisions about products or any related domain. The decision making process required others opinions on different products and features in order to improve them for getting better competitive advantages.

2. Problem Statement

Opinion mining have multiple challenging issues like, document level sentiment classification, sentence level sentiment classification, opinion lexicon generation, feature-based opinion mining, opinion mining of comparative sentences, opinion spam detection [4].

We aim to focus on website design level and provide an idea of such an interactive interface that accepts the reviews based on features. The proposed model includes website design at interface level that is offering users to post reviews. The system need to have global standard those companies should follow for web 2 development. The Web 2.0 development is the standard for review based sites. As everybody is free to read and write the reviews in any way on supposed sites. If we focused on the platform on that reviews are posted and user may post any review under specific format under related feature, it can make feature extraction process easier and also by applying authenticity check the spam users can be avoided up to some extent.

3. Literature Survey

As e-commerce is becoming more and more popular, the number of reviews also increased . To ask from family or friends about their opinion on product before buying was very common trend before the evolution of Web 2.0. The surveys and focused group studies were also conducted at organizational level to find expert feedback. The way of talking opinion has changed with the rapid growth of user generated content on Web 2.0[4]. The people from multiple sites and domains used to share information via web, where spam users are also very common. In the related studies of opinion mining, spam detection is a new research problem in the field. The spam related early study was reported by[5]. In the study of Alisa Kongthon, discusses that the agoda also worked on opinion mining on gathered data and then gives the positive and negative summaries[6].

Different users post different type of information through web 2.0 to share experience. It is very difficult for a human reader to find relevant sources and true opinions from the huge amount of opinions that are provided on web sources. In context of opinion mining, the fake opinion information

is known as spam and the one who participate in this job is known as spammers[7].Actually spam is deliberately doing false activity against competitors to degrade its position or post fake post in favor of product to improve company own position. To handle spam reviews, there are different approaches proposed by different authors such as collaborative filtrating [8] , duplicate detection [9], group spam detection [10] and Suspicious Reviews finding [11] etc. Also in the opinion mining systems we have examined, product features are extracted and then sentiment of each feature is assigned. Next the systems are summarized and presented in various forms. Most of the systems that currently exist extract product features largely based on a statistical approach. On the contrary, various methods are used for assigning sentiment to the extracted features: PMI method, unsupervised classification method, and syntactic analysis [12, 13]. Some of the OM systems use linguistic resources which contain sentiment lexicons and the others, instead, use star rating or thumbs up/down[14] .

There is no particular study to avoid spam users at neither interface level nor posting reviews based on features. Therefore in this proposed work, we provide a model that will take the reviews based on features to overcome the efforts for feature detection and also handle spam by spam detection at interface level of a Website.

Here we propose an approach to post the reviews under defined structure through feature based posting and handle spam by applying user authenticity check before posting review. The propose study focus on feature based review posting to minimize the effort of feature analysis.

A. Definition (opinion passage on a feature)

The opinion is passed on a feature ‘f’ of the object ‘O’ evaluated in ‘d’ that expresses a positive or negative opinion on f. This means that it is possible that a sequence of sentences together expresses an opinion on an object or a feature of the object. The opinion on feature of an hotel can be like this “The room service of this hotel is excellent”.

Feature-Based Opinion Mining: consist of An object O is represented with a finite set of features, F ={f1, f2, …, fn}, which includes the object itself. Each feature fi ∈ F can be expressed with a finite set of words or phrases Wi, which are synonyms. That is, there is a set of corresponding synonym sets W ={W1, W2, …,Wn} for the n features.

The opinion mining task discovers all these hidden pieces of information from a given evaluative document d. Mining output is to evaluate the sentiment on opinions that either the review is positive, negative or neutral.

4. Proposed Model

Customer’s feedback is necessary to improve business plans and it also makes decision making easy for future plans by management. As here we proposed how sites could be developed and how conveniently and particular user can post the reviews, share feedback on different features. The flow chart described below the described the feature based review posting model.

Fig-1. Feature based classification

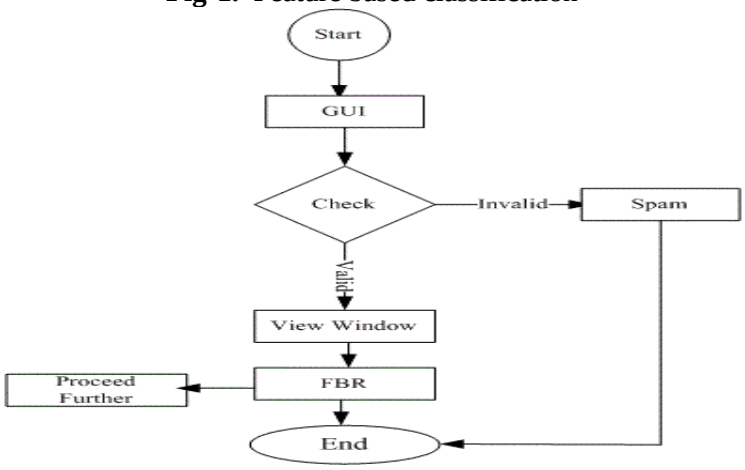

Table-1. View Classification for Different Class of Users

Hotel Features	**Available for Level 1**
Facilities	• Restaurant
	• Family room
	• Coffee shop
	• Salon
	• Business centre
	• Executive floor
	• Safety deposit boxes
	• Smoking area
	• Casino
	• Tours
	• Dry clean service
Sports & Recreation	• Garden
	• Indoor pool
	• Private beach
	• Gym
	• Water sports
	• Squash court
	• Tennis court
	• Disabled facilities
Internet	• Internet Access
	LAN/WAN
	complimentary
Car Park	• Car Park
	• Car Wash
Facilities	• 24 hours room service
	• Airport transfer
	• Pets allowed
	• Baby sitting
	• Restaurant
	• Family room
	• Coffee shop
	• Salon
	• Laundry service
Sports & Recreation	• Garden
	• Outdoor pool
	• Gym
	• Tennis court
	• Spa
	• Kids pool
Internet	• Internet Access WAN

If we limit the feature for users to pass reviews it can be intelligent approach, as there are multiple features that are not concerned with the user to pass comments. If we provide the feature based choices to comment it can help users to write more precisely in particular domain. As spam can also be reduced in this way that is immense problem in the novel era of opinion mining. Here a proposed architecture is described to define the proposed solution for FBRP and Spam handling. As its human nature when we are free to use something there are many chances of miss use in it, as here is the case we are tackling natural language to gain advantages so there are same number of chances of opinion spam.

Table-2. For Class-Level 1

Class-Level 1	**Class-level 2**
Feature view	Feature view
window level	window level

One way to reduce spam on particular areas like hotel lines may limit the user up to some extent by applying user authentication check. Many spammers are working for earring money they might be in the form of group or individual also as they are free to pass negative reviews to destroy the reputation of competitors organization. The given architecture that is providing the feature based view and feature based review posting system for registered users. The opinion mining techniques can effectively work over classified data also it will reduce the chance of spammers to get in easily to damage the related competitor position; it can be time and cost effective. As opinion mining approach may work more efficiently on classified set of data. In place of combination data the classified data can be easy to process. The data that is coming from different websites if once classified according to proposed approach it can make job of opinion mining easier in comparison that are presently required on feature extraction also data spam issues can be minimize to some extent.

5. Conclusion & Future Work

This paper describes an idea of feature based reviews posting to minimize the efforts of feature analysis and handle spam in order to increase system reliability and meet customer/manufacture need. The customer reviews are posted in the form of natural language and it needs natural langue tools to process. The reviews will be analyzed, once reviews are gathered free from spammers and according to feature based review posting model. The main objective is to provide a model that allows the valid users to post the review based on features on and block the spam users to interact with the online review system. In minutes, the positional customer and manufacturer can view the reviews on particular features that are posted in fairer manner. Furthermore, one can post the reviews time to time about pros and cons of product features in more reliable way. In future, we intend to make such system and compare with current approaches and noted down the results.

References

1. Pang, B. and L. Lee, Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2008. 2(1-2): p. 1-135.
2. Liu, B. and L. Zhang, A survey of opinion mining and sentiment analysis, in Mining Text Data. 2012, Springer. p. 415-463.
3. Jindal, N. and B. Liu. Identifying comparative sentences in text documents. in Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. 2006. ACM.
4. Liu, B., Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 2012. 5(1): p. 1-167.

5. Lim, E.-P., et al. Detecting product review spammers using rating behaviors. in Proceedings of the 19th ACM international conference on Information and knowledge management. 2010. ACM.
6. Kongthon, A., et al. Using an opinion mining approach to exploit Web content in order to improve customer relationship management. in Technology Management for Global Economic Growth (PICMET), 2010 Proceedings of PICMET'10:. 2010. IEEE.
7. Jindal, N. and B. Liu. Opinion spam and analysis. in Proceedings of the international conference on Web search and web data mining. 2008. ACM.
8. Mehta, B., T. Hofmann, and P. Fankhauser. Lies and propaganda: detecting spam users in collaborative filtering. in Proceedings of the 12th international conference on Intelligent user interfaces. 2007. ACM.
9. Jindal, N., B. Liu, and E.-P. Lim. Finding unusual review patterns using unexpected rules. in Proceedings of the 19th ACM international conference on Information and knowledge management. 2010. ACM.
10. Mukherjee, A., et al. Detecting group review spam. in Proceedings of the 20th international conference companion on World wide web. 2011. ACM.
11. Wu, G., et al. Distortion as a validation criterion in the identification of suspicious reviews. in Proceedings of the First Workshop on Social Media Analytics. 2010. ACM.
12. Lee, D., O.-R. Jeong, and S.-g. Lee. Opinion mining of customer feedback data on the web. in Proceedings of the 2nd international conference on Ubiquitous information management and communication. 2008. ACM.
13. Esuli, A. and F. Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. in Proceedings of LREC. 2006.
14. Turney, P.D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. in Proceedings of the 40th annual meeting on association for computational linguistics. 2002. Association for Computational Linguistics.