E-Commerce Reviews Management System Based on Online Customer Reviews Mining

Deng Bin
School of mgmt. and Econ.
Univ. of Elec. Sci. & Tech. of China
Chengdu, P.R.China
dengbin@uestc.edu.cn

Shao Peiji
School of mgmt. and Econ.
Univ. of Elec. Sci. & Tech. of China
Chengdu, P.R.China
shaopieji@163.com

Zhao Dan
School of mgmt. and Econ.
Southwest University for Nationalities
Chengdu, P.R.China
zd324@126.com

***Abstract*—E-commerce website accumulates a large number of customer reviews for merchandise and online shopping services. E-commerce enterprises and manufacturers could get customer opinion to improve service and merchandise through mining customer reviews. The paper presents a prototype system could be used to track and manage customer reviews, through mining topics and sentiment orientation from online customer reviews. The paper studied Chinese reviews mining processing steps and methods. Finally, this study applied text mining to costumer reviews. Through case study, the method can efficiently identify topics to enterprises.**

Keywords- E-commerce; online customer reviews; text mining; opinion mining

I. INTRODUCTION

Online customer reviews, which provided by consumers who have purchased a product previously. These have become a major information source for consumers and marketers know the quality of products. Customer reviews on products or services that companies publish on their web sites are widely read now. Customers are putting less trust in corporate marketing messages and becoming more influenced by recommendations from other people reviews. E-commerce companies need to have a long-term client relationship strategy to keep customer satisfied and improve services. Manufacturers need discover customer opinions about company products, features, services.

Text mining is an interdisciplinary field which draws on information retrieval, data mining, machine learning, statistics, and computational linguistics [1]. With the advent of automatic techniques for text mining free-form customer reviews can be processed efficiently and distilled down to essential topics and recurring patterns of content. Through the text mining analysis of reviews, the enterprise can effective manage reviews. It is believed that satisfied customers not only continue to use the product or service but also help to recruit more customers through word of mouth.

The paper organized as follows. Section 2 reviews the previous studies that related to review mining. Section 3 presents a framework for E-commerce reviews management system (EC-RMS). Section 4, we described reviews mining on the steps and the algorithms in detail. Section 5, we apply the real data of website to the test text mining in order to achieve our objectives. The conclusion and direction for further research are given in Section 6.

II. RELATED WORKS

Review mining is a complicated text understanding technology involving review extraction and sentiment analysis. The ideal review-mining tool would process a set of search results for a given item, generating a list of product attributes and aggregating reviews about each of them [2]. Hong Yu separated facts from reviews and identified the polarity of review sentences [3]. Bethard proposed a semantic parser based approach to identify opinion propositions and their holders [4]. Kim and Hovy presented a series of articles about review analysis, including automatic detecting review bearing words and sentences [5], identifying and analyzing judgment reviews [6]. Review analysis in Chinese is still in an elementary stage [7-9]. Existing works on review mining and summarization mainly focused on product reviews. Existing research emphasis a step approach research did not a systematic study and discussion the reviews of management.

III. FRAMEWORK FOR E-COMMERCE REVIEWS MANAGEMENT SYSTEM

A. Online Customer Reviews

E-Commerce websites have customer reviews module, which allows customers to submit reviews for the products had bought and collects positive and negative reviews. Negative reviews will not be deleted, because research has shown that negative reviews increase the credibility of the products. E-Commerce websites can export reviews to a delimited text file which facilitate the analysis.

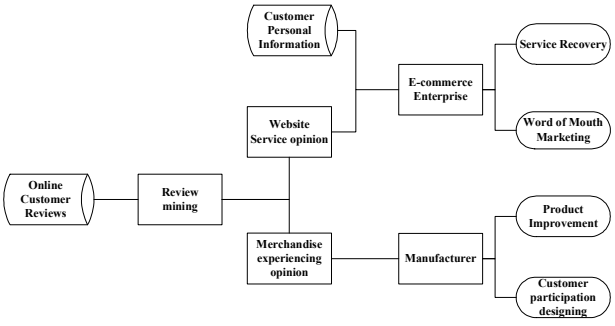

Figure 1. Framework for E-commerce reviews management system.

B. Review Mining

This module achieves preprocessing for reviews text files. And then use of text mining for classification the

978-0-7695-3942-3/10 $26.00 © 2010 IEEE
DOI 10.1109/CICC-ITOE.2010.100

reviews, one part of reviews for the website service, another for the goods. The main steps include word segmentation, part-of-speech tagging, feature representation, semantic classification/categorization, and sentiment analysis/classification.

Word segmentation is the process of dividing written text into meaningful units, such as words, sentences, or topics. Part-of-speech tagging is the process of marking up the words in a text as corresponding to a particular part of speech, based on both its definition. Feature representation turns the content of the text into the format that the computer can deal with. Semantic classification/categorization identifying what topics customers are talking about, it would be useful to characterize the reviews that they express through those topics. Sentiment analysis/classification aims at mining reviews of people for a certain event's topic or product by automatic classifying the reviews into positive or negative reviews.

C. *Website Service Opinion and Management*

Through analysis the reviews of E-commerce sites, the paper summarized a website service opinion dimensions (Tab 1). EC-RMS watches which products are most highly reviews rated by customers, and then displays those products prominently. EC-RMS was design to analyze customer reviews and automate mining customer complaints. The system can flag reviews left by dissatisfaction customers. That information should feedback to customer service representatives, who may attempt to solve the customer's problem or address their complaint, and implement complaint handling and service recovery.

Complaint handling and service recovery are keys to an effective customer loyalty program or customer retention program. A highly effective service recovery means a service or product failure offers a chance to achieve higher satisfaction ratings from customers than if the failure had never happened. In fact it can create better will than if things had gone smoothly in the first place. Nevertheless not all service recovery efforts will lead to increased satisfaction ratings as several studies have already shown. The key is to understand that there is certain situation when it is highly likely that a service recovery will lead to increased customer satisfaction. Services recoveries that are likely to be efficient are obviously those where the service failure is perceived to be not systematic or that the company had little control over it. But in cases when it was a systematic failure and the company had control over the failure there is benefit for when service recovery activities are put into action to ensure that one can win-back customer's and eliminate the failure source.

TABLE I. WEBSITE SERVICE OPINION DIMENSIONS

Object	Dimensions	Explain
Pre-sales	Purchasing consultation	Q & A solution is to provide online customer queries
	Comprehensive information	Commodity information meet customer needs
Merchandise	Price	Price fluctuations impact to customers
	Quality	User experience
Sale of services	Payment	Payment convenience, variety, security
	Internal Distribution	Internal order processing speed
	Communication	Timely inform the customers transaction information
Logistics	Timeliness	Time is reasonable and logistics
	Attitude of service	Courier services, door-to-door approach
	Quality	Outer wear, safety
After-sales service	Returns& Replacements	To deal with the issue of goods

D. *Merchandise Experiencing Opinion and Application*

Through analysis the reviews of E-commerce sites, the paper concluded a merchandise experiencing opinion dimensions (Tab 2). Customer reviews include various aspects of goods, which can be used to character design and improve products. The insight that online reviews provide to retailers extends to merchandising and even product development. Manufacturer used reviews to correct problems with packaging that customers' complain, and fix a manufacturing problem when supplier wasn't quite hitting the mark. While many companies often market house-brand products primarily on the basis of their lower price than name brand items, performance has positioned its products based on a combination of features, performance and price. Direct feedback on these items is helping performance merchandisers better understand how to improve house-brand products to cater to customers' requirement.

The automatic analysis of sentiments on customer review is useful for any company or institution who caring about quality control. At the same time, getting user feedback means bothering him or her for surveys on every aspect of the company is interested in. The problems with this approach are making a survey for each product or feature; the format, distribution and the time of survey. Customer makes commentaries about a certain subject or talks about his or her personal experiences, inviting readers to provide their own reviews.

TABLE II. MERCHANDISE EXPERIENCING OPINION DIMENSIONS

Object	Features	Performance	Individual Requirements
Dimensions	Color	Perceived Usefulness	Function
	Texture	Perceived Ease of Use	Appearances
	Shape	Perceived Durability	Humanity Design
	Humanity Design		

IV. REVIEW MINING METHODOLOGIES

A. *Word segmentation*

The subject of natural language processing is an artificial process implemented in computers, different from the term applies both to mental processes used by humans when reading text. Word segmentation is the problem of dividing a string of written language into its component

words. In English and many other languages using some forms of the Latin alphabet, the space is a good approximation of a word delimiter. Languages which do not have a trivial word segmentation process include Chinese, where sentences but not words are delimited. Commonly used algorithms are maximum matching method; reverse maximum matching method, optimum matching method.

B. Part-of-Speech Tagging (POS)

POS is the act of eliminating syntax ambiguity with appropriate methods, base on syntax relations in context. Result of part-of-speech tagging direct affects research in many fields such as paring, semantic analysis, speech recognition, machine translation, information search, and information filtration and so on. Much research has been done to tag Chinese part-of-speech tagging using several different models and methods, including: statistical model, rule-based methods, genetic algorithms and artificial neural network etc. Product features are usually nouns or noun phrases in review sentences. EC-RMS parses each review to split text into sentences and produce the part-of-speech tag for each word. The process also identifies simple noun and verb groups.

C. Feature Representation

Dimension Reduction in the Vector Space Mode (VSM) is generally adopted to accomplish the text processing. The basic idea of VSM is to express text using vector like (w1,w2,.....wn). W is the weight of given term in the text. TF-IDF is a typical method for word weight computing. The following formula is used to compute weight of words: Wi is the importance of word and n is the number of occurrence of the word in sentence. M means the total number of sentences in the text and m is the number of the other sentences contained this word.

D. Semantic Classification/Categorization

In order to train a classifier for semantic recognition in text, classic supervised learning techniques can be used [10]. The most competitive classification methods include SVM, NBM, and ME.

---Support Vector Machines. The SVM method draws the widest possible frontier between the different classes of samples in the vector space representing the corpus [11]. SVM operate by constructing a hyper plane with maximal Euclidean distance to the closest training examples.

---Naive Bayes Multinomial. The Bayes Multinomial method combines the use of the probability well known Bayes law and the multinomial law. A naive Bayes classifier uses Bayes rule as its main equation, under the naive assumption of conditional independence: each individual feature is assumed to be an indication of the assigned class, independent of each other.

---Maximum Entropy. The approach tries to pre-serve as much uncertainty as possible. A number of models are computed, where each feature corresponds to a constraint on the model. The model with the maximum entropy over all models that satisfy these constraints is selected for classification. This way no assumptions are made that are not justified by the empirical evidence available.

E. Sentiment Analysis/Classification

The task of sentiment classification is to determine the semantic orientations of words, sentences or documents. Text sentiment orientation is decided by nouns or adjectives with the strong commendatory or derogatory semantic orientation. The evaluative character of a word is called its sentiment orientation. Positive sentiment orientation indicates praise and negative sentiment orientation indicates criticism. Sentiment analysis/classification applied methods in similar to semantic classification/categorization.

V. Review Mining Methodologies

The paper selected the number of more than 100 reviews on 8 kinds products, download 1000 reviews of customers from Amazon.cn and DangDang which are B2C sites. The paper used the algorithm to classify topics of reviews. Word segmentation used CSW carve the word component. Part-of-speech tagging used a semantic tagging corpus which was built by Institute of Computational Linguistics, Peking University [12]. Feature representation used Vector Space Mode (VSM). Semantic classification used SVM algorithm in MATLAB. A review maybe involves several topics, the result shows in Table 3. In the sample, consumers are most concerned about the reviews of three topics are: timeliness, quality and attitude of service.

TABLE III. Result Website Service Opinion

Object	Dimensions	Results
Pre-sales	Purchasing consultation	145
	Comprehensive information	264
Merchandise	Price	443
	Quality	760
Sale of services	Payment	45
	Internal Distribution	48
	Communication	169
Logistics	Timeliness	764
	Attitude of service	684
	Quality	246
After-sales service	Returns& Replacements	63

VI. Conclusion

The paper study a prototype system can be used to track and manage customer reviews. With the rapid development of e-commerce, customer reviews will be-come more and more important for e-commerce enterprises and manufacturers. The prototype system model can be a reference for e-commerce enterprises, which is a cost-effective solution available to manage and analyze online reviews. Further study, we will build the system and use it to commercial applications.

References

[1] S. Dan, et al., "Opinion Mining in e-Learning System," Network and Parallel Computing Workshops NPC Workshops IFIP International Conference, 2007, pp. 788-792.

[2] Kushal Dave, Steve Lawrence, David M. Pennock. "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews." In Proceedings of WWW, 2003, pp.519–528.

[3] Yu, H., V. Hatzivassiloglou. "Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences". Proceedings of the EMNLP-03,

2003, pp.341-349.
[4] S. Bethard, H. Yu, A. Thornton. “Automatic extraction of opinion propositions and their holders “. Proceedings of AAAI-04 Spring Symposium on Exploring Attitude and Affect in Text, 2004, pp.121-128.
[5] S. Kim, E. Hovy. “Automatic Detection of Opinion Bearing Words and Sentences “. Second International Joint Conference on Natural Language Processing, 2005, pp. 61-66.
[6] G. Wei, L. Hongfei. “Automatic Opinion Analysis Based on SVM and Distance-Weighted Computing”. Advanced Language Processing and Web Information Technology International Conference, 2008, pp. 93-98.
[7] X.-J. Tong, et al., “Research on Chinese Text Automatic Categorization Based on VSM”. Wireless Communications, Networking and Mobile Computing International Conference, 2007, pp.3863-3866.
[8] X. Ruifeng, et al. “Learning Knowledge from Relevant Webpage for Opinion Analysis”. Web Intelligence and Intelligent Agent Technology IEEE/WIC/ACM International Conference, 2008, pp.307-313.
[9] Erik Boiy, Pieter Hens. Moens. “Automatic Sentiment Analysis in On-line Text “. Proceedings ELPUB2007 Conference on Electronic, 2007(6) , pp.349-360.
[10] A. Balahur, A. Montoyo. “A feature dependent method for opinion mining and classification “. Natural Language Processing and Knowledge Engineering International Conference, 2008, pp. 1-7.
[11] A. A. Michael Gamon, Simon Corston-Oliver. “Pulse: Mining Customer Opinions from Free Text”, Springer-Verlag Berlin Heidelberg, 2005, pp. 121-132.
[12] Y. Wu, P. Jin, Y. Zhang and S. Yu. 2006. “A Chinese Corpus with Word Sense Annotation”. In: Proceeding of ICCPOL 06. 2006, pp.414-421.
[13] MENG Xiangping GAO Yan. “Electric systems analysis”. Beijing: Higher Education Press, 2004, pp.3−21