



# Identifying comparative customer requirements from product online reviews for competitor analysis



Jian Jin<sup>a,b,d,\*</sup>, Ping Ji<sup>b,c</sup>, Rui Gu<sup>e</sup>

<sup>a</sup> Department of Information Management, Beijing Normal University, Beijing 100875, China

<sup>b</sup> The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, China

<sup>c</sup> Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

<sup>d</sup> School of Management, Xian Jiaotong University, Xian 710049, China

<sup>e</sup> School of Information Technology and Management, University of International Business and Economics, Beijing 100029, China

## ARTICLE INFO

### Article history:

Received 26 April 2015

Received in revised form

4 November 2015

Accepted 6 December 2015

Available online 22 December 2015

### Keywords:

Customer requirement

Review analysis

Competitor analysis

Product comparison

Representative yet comparative sentences

Product design

## ABSTRACT

A large volume of product online reviews are generated from time to time, which contain rich information regarding customer requirements. These reviews help designers to make exhaustive analyses of competitors, which is one indispensable step in market-driven product design. How to extract critical opinionated sentences associated with some specific features from product online reviews has been investigated by some researchers. However, few of them examined how to employ these valuable resources for competitor analysis. Hence, in this research, a framework is illustrated to select pairs of opinionated representative yet comparative sentences with specific product features from reviews of competitive products. With the help of the techniques on sentiment analysis, opinionated sentences referring to a specific feature are first identified from product online reviews. Then, information representativeness, information comparativeness and information diversity are investigated for the selection of a small number of representative yet comparative opinionated sentences. Accordingly, an optimization problem is formulated, and three greedy algorithms are proposed to analyze this problem for suboptimal solutions. Finally, with a large amount of real data from Amazon.com, categories of extensive experiments are conducted and the final encouraging results are realized, which prove the effectiveness of the proposed approach.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Nowadays, millions of customers gain opportunities to compare similar products and pick their favorites in digital retailers, such as Amazon.com and Taobao.com. Customers, especially novices, often make comparisons, find the pros and cons among the competitors, and choose the most suitable ones. On the other hand, product designers are required to understand customer choices on alternatives regarding their compliments and complaints. Perhaps one simple approach to understand the pros and cons among competitors is to read online reviews of different products. Product online reviews provide rich information about customers' concerns and they allow designers to get a general idea regarding competitors which may assist to improve products.

However, it is generally difficult to understand all reviews in different websites for competitive products and obtain insightful

suggestions manually. In the past decade, some researchers, especially in computer science, paid much attention to how to analyze such big customer data intelligently and efficiently (Ding and Liu, 2007; Abbasi et al., 2008; Chen et al., 2012; Korenek and Simko, 2013). For instance, many studies about opinion mining for online reviews were reported to infer sentiment polarities from online reviews in different levels. Nonetheless, most researchers in this field ignore how to make their findings be seamlessly utilized by designers. Recently, a limited number of studies were noted to utilize the latest development in artificial intelligence and data mining in the design community (Zhan et al., 2009; Li et al., 2011b; Dou et al., 2012). These studies help designers to understand a large amount of customer requirements in online reviews for product improvements. But these discussions are far from sufficient, and some potential problems have not been fully investigated such as, with product online reviews, how to conduct a thorough competitor analysis. Actually, in a typical scenario of a customer-driven new product design (NPD), the strengths and weakness are often analyzed exhaustively for probable opportunities to succeed in the fierce market competition. Competitor

\* Corresponding author at: Department of Information Management, Beijing Normal University, Beijing 100875, China. Tel.: +86 10 5880 5387.

E-mail address: [jinjian.jay@bnu.edu.cn](mailto:jinjian.jay@bnu.edu.cn) (J. Jin).

analysis is also an indispensable step in Quality Function Deployment (QFD), which is a famous tool for customer-driven NPD.

Hence, in this research, the ultimate goal is to identify several pairs of representative yet comparative opinionated sentences with specific product features from product online reviews. The essence of this problem is that these review sentences should be descriptive about general customer concerns and, at the same time, they are expected to be comparative to reflect similar customer feedback of different products. Specifically, opinionated sentences referring to specific product features are initially extracted from product reviews by supervised learning approaches. Next, three aspects of review sentences that characterize online customer requirements are considered and an optimization problem is formulated for the selection of review sentences. Additionally, different functions that evaluate the similarity between sentences are utilized, and greedy algorithms are proposed to analyze the optimization problem for suboptimal solutions.

To avoid unnecessary misunderstanding, the difference between "customers" and "consumers" is emphasized in consideration of their exact roles. According to Blythe (2008), "customers are the people who buy the product; consumers are those who consume it." Therefore, strictly speaking, online reviews are often posted by customers and customer requirements that are revealed in these reviews. Furthermore, there exist different types of online reviews, such as restaurant reviews, product reviews, political reviews, etc. In this research, the focus is limited on product online reviews only.

Accordingly, selected pairs of sentences are expected to discuss similar consumer concerns with specific product features. The identification of these sentences is assumed to be an effective approach for product comparisons as well as the reduction of the information to a set of representative sentences. These results will not only benefit designers to obtain pros and cons of comparative products from a big volume of online customer feedback efficiently, but also make pairwise comparisons with competitors beyond the product feature level in different sentiment polarities. Besides pairwise comparisons within specific product features, other types of comparisons might also be interesting. For instance, customer concerns within a specific feature for one product do not have corresponding similar feedback for competitor products. It is absolutely valuable information to interests product designers for competitor analysis, which is not concerned in this research. These types of comparisons are left for further careful considerations.

The contributions of this research are at least threefold. First, a framework for mining comparative viewpoints from product online reviews is presented. This problem is fundamentally different from many existing studies. In the competitive analysis of customer-driven NPD, it is a crucial question for designers, which has not been answered but is urgently needed. Second, how to identify pairs of representative yet comparative review sentences is formulated as an optimization problem. Three critical aspects of comparative sentences are considered in this optimization problem. Finally, evaluation metrics are proposed for this new problem, and categories of experiments are conducted to illustrate the effectiveness of these approaches.

The rest of this research is structured as follows. In Section 2, relevant studies are briefly reviewed. Section 3 outlines a framework for mining comparative viewpoints from product online reviews. Accordingly, in Section 4, an optimization perspective is proposed to identify representative review sentences for competitive products. Greedy algorithms are described for the optimization problem. In Section 5, comprehensive details of the experimental study are presented by utilizing a large number of mobile phone reviews from Amazon.com. Section 6 concludes this research.

## 2. Related work

### 2.1. Review summarization

To gain a general sketch outline regarding customer online reviews, various summarization approaches were innovated at the sentence level, the sentiment level or the product feature level.

Zhuang et al. (2006) proposed a review summarization framework at the sentence level. Dependency relation templates are derived from a dependency grammar graph. Then, these templates are employed to detect feature-opinion pairs. Finally, a group of organized sentences are regarded as the review summary. A review summarization system was illustrated to cluster review sentences with similar facets in the same sentiment polarity (Ly et al., 2011). In this system, product facets are identified from online reviews with an association mining approach. Then, the similarity between sentences is calculated. Finally, both hierarchical groupwise-average clustering and the non-hierarchical exchange method are applied to cluster review sentences. Those sentences with maximal information coverage are selected as representative sentences.

Jakob and Gurevych (2010) took tokens, POS tags, short dependency paths, distances between opinion words and others as features of online reviews. Then, a CRFs algorithm was utilized to detect opinion target in online reviews. Four types of CRFs models were compared to identify product features and related opinion words (Li et al., 2010a). The linear CRFs was utilized to model the sequential dependency between continuous words. The skip-chain CRFs was to model the long distance dependency with conjunctions. The tree CRFs was to model the syntactic tree structure. The skip-tree CRFs was to combine both the skip-chain CRFs and the tree CRFs. Some researchers also proposed an opinion summarization for Bengali news articles (Das and Bandyopadhyay, 2010). First, an Support Vector Machine (SVM) classifier was utilized to identify subjective sentences, and a CRFs model was utilized to recognize the theme word. Then, sentences are clustered by *K*-means according to the cosine similarity. In addition, a semantic graph is constructed to denote the connection between documents. Finally, a PageRank-like approach is applied to select sentences for each cluster.

A summarization approach was proposed for rated aspects in short reviews (Lu et al., 2009). Rated aspects are identified from short reviews by unstructured Probabilistic Latent Semantic Analysis (PLSA), structured PLSA and structured PLSA with aspect priors. Next, aspect ratings are predicted utilizing two methods: a local prediction method and a global prediction method. Finally, the top three phrases with the highest frequency are selected to represent rated aspects. Ma et al. (2012) proposed two probabilistic graph models to cluster reader comments in news articles. In the first model, topics in reader comments are connected with topics in news articles. In the second model, topics in reader comments are derived from topics in news articles and all comments themselves. Finally, representative sentences are selected by the approach of Maximal Marginal Relevance and the approach of Ranting & Length.

However, these text summarization models neglect some raw messages, which is not expected by product designers, especially for designers who want to conduct fine-grained level analysis on customer concerns for product improvements.

### 2.2. Review recommendation and review sampling

A huge number of online reviews are widely available. However, only some of them contain valuable information. This dilemma interests researchers to develop different approaches for

Download English Version:

<https://daneshyari.com/en/article/380244>

Download Persian Version:

<https://daneshyari.com/article/380244>

[Daneshyari.com](https://daneshyari.com)