Identifying Comparative Customer Requirements from Product Online Reviews for Competitor Analysis

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.

Keywords: customer requirement; review analysis; competitor analysis; product comparison; representative yet comparative sentences; product design

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, Loh and Liu, 2009; Li, Hitt and Zhang, 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 NPD (new product design), the strengths and weakness are often analyzed exhaustively for probable opportunities to succeed in the fierce market competition. Competitor analysis is also an indispensable step in QFD (Quality Function Deployment), 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 Jim Blythe's definition (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 is 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. proposed a review summarization framework at the sentence level (Zhuang et al., 2006). 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 took tokens, POS tags, short dependency paths, distances between opinion words and others as features of online reviews (Jakob and Gurevych, 2010). 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 SVM (Support Vector Machine) 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 PLSA (Probabilistic Latent Semantic Analysis), 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. proposed two probabilistic graph models to cluster reader comments in news articles (Ma et al., 2012). 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 review recommendation or review sampling, which aims to obtain a small number of reviews with critical customer viewpoints.

A matrix factorization model and a tensor factorization model were proposed to recommend reviews for different raters (Moghaddam et al., 2011). The matrix factorization model examines the information about the raters and reviews. The tensor factorization model reckons the information about raters, reviewers and products. They found that the matrix factorization model is more effective in the recommendation of online reviews. Based on the tensor factorization model, an extended tensor factorization method is reported in which the overall rating of a product was incorporated as one constraint on the tensor factorization model (Moghaddam et al., 2012). In addition, in this study, another unbiased extended tensor factorization model is introduced to capture the biases associated with the tendencies that some reviewers give higher ratings than others and the tendencies that some reviews receive higher ratings than others.

Three aspects are utilized to select a small set of comprehensive reviews, which include the discussed attributes, the sentiment polarities and the quality of reviews (Tsaparas et al., 2011). Then, different coverage functions are defined for the selection of reviews, and various greedy algorithms are therefore proposed to coordinate coverage functions. However, review samples should be proportionate to the sentiment polarities (Lappas et al., 2012). With this purpose, a greedy algorithm, an integer-regression algorithm and an iterative-random algorithm were developed to sample a characteristic set of reviews.

There are also some studies regarding review sampling for opinion mining because manually labeled data are usually expensive to obtain. To avoid random sampling, the selection of informative samples for opinion mining was discussed (Ju et al., 2012). In this study, the informativeness about a word or a document is evaluated. The informativeness of words is defined as the product of the proportion between a certain POS and its occurring frequency. The informativeness of sentences is defined as the sum of informativeness of words that are normalized by the logarithm of the document length. A new sampling strategy was presented to select reviews for imbalanced opinion mining (Li et al., 2012). Two classifiers are trained with a disjoint feature subspace and a labeled dataset. One classifier is to select the top k positive and k negative samples with the highest probabilities. The other classifier is to select one positive sample and one negative sample with the lowest probabilities. Finally, the two approaches are applied in a pool-based active learning algorithm for imbalanced sentiment classification.

All of these models for review recommendation or review sampling help designers to screen a big volume of customer online reviews and focus on helpful opinions only. However, few of them are capable to be applied directly for competitor analysis.

2.3 Contrastive Viewpoints Extraction

The extraction of contrastive viewpoints from textual data benefits to make efficient summarizations and comparisons of competitors. It motivates some researchers to conduct a comparative text summarization task. Also, with comparative sentences, some researchers tried to recognize the exact relationship within the comparative sentences.

A two-stage method was proposed to summarize multiple contrastive viewpoints from opinionated text by Paul et al. (2010). In the first stage, an extended LDA (Latent Dirichlet Analysis) model is utilized to extract topics and viewpoints from texts with different types of features. In the second stage, a modified PageRank method is employed to summarize comparative sentences. Mukherjee and Liu first utilized a topic model to extract topics and expressions indicating contention and agreement topics (Mukherjee and Liu, 2012). However, this model was argued to neglect the topics through the reply-to relation and the interaction between authors. This model was then improved by considering these two characteristics. A topic model with cross perspective was also employed for mining contrastive opinions in political documents (Fang et al., 2012).

A framework for contrastive opinion summarization was proposed (Kim and Zhai, 2009). In this framework, two aspects are considered in contrastive opinion summarization: the content similarity with the same polarity and the contrastive similarity with opposite polarities. Accordingly, an optimization problem is developed to generate comparative summaries of contradictory opinions. Another unsupervised learning method was developed to identify two groups of opposing opinions in forums

(Lu et al., 2012). The sentiments of threads are first determined by SentiWordnet. Then, the agree-or-disagree relations in forums are inferred utilizing the reply-to and user relation consistency.

To identify comparative patterns, an algorithm for sequential pattern mining with multiple minimum supports was applied on POS (Part of Speech) tags of review sentences and sentences with a small number of keywords (Jindal and Liu, 2006). Then, a naive Bayesian classifier was utilized to handle the case that a single review sentence matches several rules. Finally, the prediction from the classifier was utilized to decide whether a sentence is comparative. However, Jindal and Liu's approach fails to cover all cases of comparative sentences, and a two-level CRFs (Conditional Random Fields) model was built to identify comparative sentences in online reviews (Xu et al., 2011). The first level is to model the relationship between product relations with entities and words. The second level is to model the relationship between relations of products. According to the word formation, non-equal gradable comparatives and superlative comparatives are summarized (Ganapathibhotla and Liu, 2008). Opinionated comparatives and comparatives with context-dependent opinions are considered for both types. Then, a rule-based approach is suggested to identify which entity in the comparative sentence is preferable.

In these studies, most of them models the contrastive viewpoints extraction as a text summarization problem, which make large details of online customer concerns are abandoned. In addition, only a small number of comparative sentences are typically found in product online reviews. Hence, it leads designers potentially overlook a significant proportions of online customer concerns if a summarization task is conducted on comparative sentences only.

2.4 Product Online Reviews for Engineering Design

The utilization of product online reviews for engineering design is relatively new, and only a few relevant studies were reported to cover a limited number of subtopics.

Netzer et al. utilized a CRFs model to extract product names from online reviews (Netzer et al., 2012). Then, with the identified product features, two applications regarding product comparisons are demonstrated. In addition to online reviews, community-based question answering was also utilized by a graph propagation approach to compare products (Li et al., 2011a). A product comparison network was built by analyzing comparative sentences in online reviews (Zhang et al., 2013). In consideration of the number of the overall sentiments about customers, three types of graphs are built, which include single-link graphs, dichotomic-link graphs and multi-link graphs. Moreover, regression models are employed to analyze how various factors influence the product rank. In this approach, comparative sentences in online reviews are also utilized. Specially, the information regarding the number of preferences between comparative products is considered. Then, weights in product pairs, which are calculated as the number of preferences between products, are utilized for information propagation.

Some studies investigate the prediction of product ranks for the near future. For instance, Li et al. extracted affinity rank history, average ratings, and affinity evolution distance from product reviews (Li et al., 2010b). Then, an Autoregressive model with exogenous inputs was presented to predict product sales rank. Tucker and Kim employed online reviews to forecast product preference trends (Tucker and Kim, 2011). Sentiment polarities in the product feature level are extracted from online reviews, and the Holt-Winters exponential smoothing method is employed to predict the preference trends. Another customer opinions monitoring system was developed based on from a large volume of textual data (Goorha and Ungar, 2010). Frequent phrases and phrases that are near the terms of interest are extracted. Then, three metrics are utilized to judge which one is a dramatically appearing interesting phrase. These metrics include how frequently they are referred to, how frequently they are referred to compared with before, and how specific they refer to a topic.

Some research has also begun to analyze the usability of online reviews in product design. For instance, how to identify helpful online reviews from the perspective of designers was discussed (Liu et al., 2013). Four categories of features are extracted from product reviews, and a regression approach is utilized to infer the helpfulness of online reviews. In addition, with three domain-independent features only, it is found that there is no significant loss of the helpfulness prediction. An SVM-based method was reported to classify the information in online reviews into usability information and user experience information (Hedegaard and Simonsen, 2013). To build training samples, review sentences are manually labeled according to several categories of dimensions that relate to usability and user experience.

How to utilize online reviews directly in engineering design has also been explored. Wang et al. utilized a three-step method for customer-driven product design selection by analyzing online reviews (Wang et al., 2011). In the first step, product attributes were extracted. In the second step, a hierarchical customer preference model was developed by using a Bayesian linear regression method in which product ratings, category ratings, attribute ratings and product specifications were considered. An optimization problem was formulated in the last step to maximize the potential profit by considering constraints of ECs (engineering characteristics). Recently, based on product online reviews, an ordinal classification approach was advised to prioritize ECs for QFD (Jin et al., 2014). It is a pairwise approach in which customer online opinions are deemed features and the overall customer satisfaction is the target value. In addition, an integer linear programming model is suggested to convert the results from pairwise approaches into the original customer satisfaction ratings.

Designers' considerations are involved in these studies. However, as pointed out in the previous sections, few of them step forward to mining the value of customer online opinions for competitor analysis, which is an indispensable task in NPD.

2.5 A Brief Summary

Different models were reported to extract valuable customer information intelligently, such as contrastive viewpoints extraction, high-quality review recommendation, review summarization, etc. However, few studies take a step further toward how to use these findings to help designers' work in NPD, with the designers being the actual promoters to improve products and meet potential consumers. A few pioneering research studies have been conducted regarding the utilization of big opinion data in engineering design. But only a limited number of aspects are concerned in these studies and some practical problems in NPD are not extensively explored.

A comprehensive study regarding the analysis of online reviews toward product comparisons should be conducted, and more abundant customer preference messages for competitive products are highly desired. Accordingly, in this study, how to obtain pairs of comparative yet representative sentimental sentences with specific product features from big customer opinion data efficiently is investigated, which is one significant problem for designers' work in NPD.

3. A Framework for Mining Comparative Viewpoints from Product Online Reviews

3.1 Framework

To identify representative yet comparative sentimental sentences with specific product features from product online reviews, a framework is presented in Figure 1. As observed from this figure, POS tagging is conducted first, which is utilized for the analysis of sentiment polarities and the identification of product features. Particularly, Standford Part-Of-Speech Tagger was utilized for POS tagging. It is a famous tool and widely utilized for POS tagging in the field of natural language processing. In this research, two simple but effective supervised learning approaches are utilized for these tasks. The details about the two models will be elaborated upon in Section 3.2. Given online reviews in the same product domain, the two approaches aid designers in extracting product features with the corresponding sentiment polarities efficiently.

[Insert Figure 1]

These opinion data in different product features are then categorized to positive ones, negative ones and objective ones. Note that, in this research, if one sentence expresses either positive or negative opinions, it is deemed as a "subjective" one and it is referred as an "opinionated sentence". Otherwise, it is assumed as an objective one. This assumption is also applied in many related studies for sentiment analysis (Kim and Hovy, 2006; Ding et al., 2008; Lin and He, 2009). Also, an objective sentence might also contain valuable information about customer concerns. However, in this research, the focus is stressed on the contrastive viewpoint on similar topics for competitor analysis in the perspective of product designers. Hence, as presented in Figure 1, in this research, only positive sentences and negative sentences are taken into considerations for selecting a small number of pairs of opinionated sentences.

As discussed, selected sentences are expected to be descriptive and representative of general customer requirements. It illustrates that selected sentences are required to cover as many topics as possible. Hence, a critical subtask is to understand which topics are referred to in different product reviews. Accordingly, topic analysis is conducted on categorized opinion data, which helps to distinguish topic distributions regarding customers concerns. In addition, for the second aspect of the review sentence selection, selected sentences are expected to be comparative. It means that for the selected review sentences of different products, similar topics are referred to. Therefore, categorized opinionated sentences with the same product features are clustered in which similar customer topics are discussed. Actually, these cluster results help to obtain groups of opinionated sentences with similar topics of different products. For instance, strengths of the screen of mobile phone 1 are expected to be compared with the weakness of that of mobile phone 2. It requires that topics discussed in the selected sentences of two products need to be similar. More specifically, the selected sentences of two products must come from the same cluster of review sentences. Conversely, to gain the same clusters of review sentences of these two products, the cluster set of positive sentences referring to the screen of mobile phone 1 is intersected with that of negative ones of mobile phone 2.

Now, representative yet comparative sentences are extracted from each cluster in the intersection, which indexes groups of opinionated sentences with similar topics of different products. The details about how to select representative yet comparative sentences in each cluster will be explained in Section 4. Eventually, all of the selected sentences from each cluster are sorted according to an overall score, which evaluates a combined value of information representativeness, information comparativeness and information diversity about a group of selected sentences.

3.2 Product Feature Extraction and Sentiment Analysis

Two major tasks in the sentiment analysis on product online reviews include how to extract product features and how to judge the sentiment polarities in different levels. Many publications have reported on these tasks in the area of opinion mining (Ding et al., 2008; Lin and He, 2009; Zhai et al., 2010). However, some models are quite complex to implement for product designers, especially for those who do not have a solid background in computer science and statistics. In this research, a simple but effective approach is employed with the help of pros and cons reviews, which smoothens the difficulty on the comprehension and implementation of these tasks. Similar approaches for product feature identification and sentiment analysis were also reported in (Kim and Hovy, 2006; Yu et al., 2011).

Many review sites invite customers to post both compliments and criticisms of products they have purchased. For instance, a review of the Samsung Galaxy S III GT-I9300 is presented in Epinions.com. In this review, the pros and cons of the I9300 are highlighted clearly in which the pros are described as "Great battery life, 4.8 HD Super Amoled Display, and S Beam sharing 1.4GHz Quad-Core Processor" and the cons include that "Images tend to get overexposed, Hangs with heavy usage, and Screen dim for outdoor use". Note that the most frequently referred to nouns or noun phrases in this pros and cons list are product features. Accordingly, POS tagging is conducted, and frequently referred to nouns or noun phrases are regarded as product features. These results help to extract product features from customer online reviews in a general format, such as reviews in Amazon.com. In addition, customers may utilize different words to describe the same product feature. For example, customers use "memory" or "storage" to refer to the same feature. To cluster synonyms that refer to the same product feature, WordNet distance is utilized. Moreover, abbreviations also frequently appear in product online reviews. For instance, "apps" and "applications" are utilized interchangeably by mobile customers. Many abbreviations are occasionally defined in WordNet or other web thesauruses. Hence, a small group of manually defined synonyms are provided to improve the WordNet based clustering. Finally, with the extracted candidates from pros and cons lists, product features are identified from online reviews.

In addition, Pang and Lee (2004) developed a publicly available subjective dataset, which includes 5,000 subjective and 5,000 objective sentences. This dataset helps to build a binary classifier to discern subjective sentences from online reviews. Accordingly, the bag of words representation (BOW) is utilized to denote each review sentence with a specific product feature, and a binary Naive Bayes classifier is employed to judge whether a subjective or objective opinion is expressed. Furthermore, another subtask is to identify whether customers hold a positive or negative sentiment regarding the product feature. The good news is that sentimental information is listed clearly in pros and cons reviews, which provide a large number of non-manually labeled training samples to analyze the sentiment polarities. By employing such sentimental information in pros and cons reviews, rather than the BOW representation, sentimental terms in MPQA project (Wilson et al., 2005) are employed in a binary Naive Bayes classifier. This classifier is utilized to analyze the sentiment polarity of review sentences.

4. Comparative Viewpoints Identification

4.1 Problem Definition

Take two competitive products a and b, for instance. Suppose that designers expect to analyze the strengths and weakness of a and b associated with the product feature f. Initially, two review sentence sets, A_f and B_f, are prepared, which contain sentences referring to f. However, it is generally time-consuming to understand all sentences in A_f and B_f, whose sizes are $|A_f|= S_a$ and $|B_f|= S_b$, respectively. To help designers make a sound comparison with a and b regarding customer concerns about f efficiently, two small subsets of opinionated sentences with f, P_f and Q_f, are selected from A_f and B_f. Obviously, P_f and Q_f satisfy that $P_f \subseteq A_f$ and $Q_f \subseteq B_f$. In addition, the sizes of the two selected small subsets are expected to be equal. It is denoted as $|P_f|= K$ and $|Q_f|= K$, in which $K \leqslant S_a$ and $K \leqslant S_b$.

Generally, review sentences in P_f and Q_f from A_f and B_f are expected to:

(a) be descriptive and representative about general customer requirements. It requires that the similarity between P_f and A_f, *similarity*(P_f, A_f) or *similarity*(P_f, A_f - P_f) should be as high as possible, where A_f - P_f denotes the difference set between P_f and A_f. Likewise, *similarity*(Q_f, B_f - Q_f), should be as high as possible.

(b) be comparative, which means that they discuss similar topics of different products. It implies that the similarity between P_f and Q_f, *similarity*(P_f, Q_f), should be as high as possible or, specifically, similar topics regarding customer concerns are discussed in P_f and Q_f.

(c) be diversified to reflect various customer requirements. It suggests that the similarity between each pair of sentences within P_f, *similarity*(P_f), needs to be as low as possible, which demonstrates that the selected sentences within P_f are expected to describe multiple aspects regarding *f*. Likewise, the similarity, *similarity*(Q_f), should be as low as possible.

In particular, in this research, the review selection problem can be described as follows: how to select two small sets of review sentences, P_f and Q_f, and their size *K*, from two big sets, A_f and B_f, with the above three principles.

4.2 An Optimization Perspective

Generally, three principles can be expressed mathematically as,

(a) $P_f = \arg\max_K similarity(P_f, A_f - P_f)$ and $Q_f = \arg\max_K similarity(Q_f, B_f - Q_f)$

(b) $P_f, Q_f = \arg\max_K similarity(P_f, Q_f)$

(c) $P_f = \arg\min_K similarity(P_f)$ and $Q_f = \arg\min_K similarity(Q_f)$

The third principle, which requires that the similarity *similarity*(P_f) and *similarity*(Q_f) are a minimization problem, can be equally rewritten as a maximization problem as $P_f = \arg\min_K similarity(P_f) = -\arg\max_K similarity(P_f)$. Accordingly, P_f and Q_f should intuitively satisfy all three principles. It is denoted as,

$$\begin{aligned} P_f, Q_f = \arg\max_K \{ & \lambda_1 (similarity(P_f, A_f - P_f) + similarity(Q_f, B_f - Q_f)) \\ & + \lambda_2 similarity(P_f, Q_f) \\ & - (1 - \lambda_1 - \lambda_2)(similarity(P_f) + similarity(Q_f)) \} \end{aligned} \tag{1}$$

λ_1 and λ_2 are two parameters that control the relative importance of the three principles. They are confined as,

$$\begin{gathered} 0 \le \lambda_1 \le 1 \\ 0 \le \lambda_2 \le 1 \\ 0 \le 1 - \lambda_1 - \lambda_2 \le 1 \end{gathered} \tag{2}$$

A further step can be performed on Equation (1) in the sentence level. Mathematically, it is equivalent to,

$$\begin{aligned} P_f, Q_f = &\arg\max_K \{\lambda_1 (similarity(P_f, A_f - P_f) + similarity(Q_f, B_f - Q_f)) \\ &+ \lambda_2 similarity(P_f, Q_f) \\ &- (1 - \lambda_1 - \lambda_2)(similarity(P_f) + similarity(Q_f))\} \\ = &\arg\max_K \{\lambda_1 (\sum_{k=1}^{K} \sum_{t_a=1}^{S_a - K} similarity(p_f^k, u_f^{t_a}) + \sum_{k=1}^{K} \sum_{kt_b=1}^{S_b - K} similarity(q_f^k, v_f^{t_b})) \\ &+ \lambda_2 \sum_{k=1}^{K} similarity(p_f^k, q_f^k) \\ &- (1 - \lambda_1 - \lambda_2)(\sum_{i=1}^{K} \sum_{j=1, j \ne i}^{K} similarity(p_f^i, p_f^j) + \sum_{i=1}^{K} \sum_{j=1, j \ne i}^{K} similarity(q_f^i, q_f^j))\} \end{aligned} \tag{3}$$

p_f^k is the k^{th} sentence in the review sentence set P_f, and $u_f^{t_a}$ is the t_a^{th} sentence in the review sentence set A_f - P_f. They are denoted as $p_f^k \in P_f$ and $u_f^{t_a} \in A_f - P_f$. Similarly, q_f^k belongs to the set Q_f, where $q_f^k \in Q_f$, and $v_f^{t_b}$ is one sentence in the set B_f - Q_f, where $v_f^{t_b} \in B_f - Q_f$.

With such an optimization perspective, to find an optimal set of representative yet contrasting review sentences of comparative products, the similarity between sentences needs to be defined and an efficient approach is expected to analyze the maximization problem.

4.3 Similarity Functions

According to the optimization perspective, one of the central tasks is how to define the similarity between review sentences.

As discussed in Section 3.1, topics are identified from online reviews of competitive products. These topics help to discern customer concerns of products. In addition, in this research, it is required that selected sentences are descriptive about the general topics of customer concerns. Correspondingly, the similarity between review sentences is evaluated by the distance between different topics that are referred to in each sentence.

On the basis of the referred topics in review sentences, in this research, two variants of similarity metrics are testified. Let Ω_f^k and Ψ_f^k be referred topics in sentence p_f^k and q_f^k, respectively. Take $similarity(p_f^k, q_f^k)$, for instance; the two similarity functions are defined as follows with different nominators.

$$similarity(p_f^k, q_f^k) = \frac{|\Omega_f^k \cap \Psi_f^k|}{|\Omega_f^k \cup \Psi_f^k|} \tag{4}$$

$$similarity(p_f^k, q_f^k) = \frac{|\Omega_f^k \cap \Psi_f^k|}{|\Omega_f^k| + |\Psi_f^k|} \tag{5}$$

Note that in this research, these two variants are presented. However, other sophisticated functions that evaluate the similarities between review sentences are also applicable for this optimization problem, such as the similarity functions proposed by Kim and Zhai (2009).

4.4 Greedy Algorithms

The objective of the optimization problem is to choose K pairs of review sentences to build P_f and Q_f from A_f and B_f, whose sizes are S_a and S_b, respectively. It is a nonlinear integer programming problem. A brute force approach is not computationally applicable because it involves $\binom{S_a}{K} \times \binom{S_b}{K} \times K! \times K!$ comparisons. To find suboptimal solutions, in this research, greedy algorithms are employed.

In Equation (3), the three principles are required to be followed at the same time. Now, in the proposed greedy algorithms, this constraint is relaxed. In particular, if only one principle is followed, it will lead to a much simpler computation to gain a suboptimal pair of sentence sets from A_f and B_f. Accordingly, three greedy algorithms are developed according to each principle.

For the first principle, $similarity(P_f, A_f - P_f)$ and $similarity(Q_f, B_f - Q_f)$ are considered. In this research, it is called information representativeness first or "R-First". Mathematically, to obtain a suboptimal pairs of sentence sets from A_f and B_f, it can be denoted as

$$\begin{aligned}\tilde{P}_f, \tilde{Q}_f &= \arg\max_k \{similarity(P_f, A_f - P_f) + similarity(Q_f, B_f - Q_f)\} \\ &= \arg\max_k \{\sum_{t_a=1}^{S_a-K} \sum_{k=1}^{K} similarity(p_f^k, u_f^{t_a}) + \sum_{t_b=1}^{S_b-K} \sum_{k=1}^{K} similarity(q_f^k, v_f^{t_b})\}\end{aligned} \quad (6)$$

This equation leads to $\binom{S_a}{K} \times K + \binom{S_b}{K} \times K$ comparisons because the top K pairs of sentences are only built from top K sentences from A_f and the top K sentences from B_f. It is more computationally economical than the primal problem that involves $\binom{S_a}{K} \times \binom{S_b}{K} \times K! \times K!$ comparisons.

For the second principle, *similarity*(P_f, Q_f) is considered. It is referred to as comparativeness first or "C-First". A suboptimal solution, by employing this approach, is written as

$$\begin{aligned}\tilde{P}_f, \tilde{Q}_f &= \arg\max_k \{similarity(P_f, Q_f)\} \\ &= \arg\max_k \sum_k similarity(p_f^k, q_f^k)\end{aligned} \quad (7)$$

The computation cost is only $S_a \times S_b$ because the top K pairs of sentences are selected from an $S_a \times S_b$ similarity matrix.

For the third principle, *similarity*(P_f) and *similarity*(Q_f) are considered. Correspondingly, it is named as diversity first or "D-First". A suboptimal solution with this approach can be denoted as

$$\begin{aligned}\tilde{P}_f, \tilde{Q}_f &= \arg\max_k \{-(similarity(P_f) + similarity(Q_f))\} \\ &= \arg\min_k \{(\sum_{i=1}^{k} \sum_{j=1, j\neq i}^{k} similarity(p_f^i, p_f^j) + \sum_{i=1}^{k} \sum_{j=1, j\neq i}^{k} similarity(q_f^i, q_f^j))\}\end{aligned} \quad (8)$$

The comparison cost is $S_a \times S_a + S_b \times S_b$ because sentences are selected from the top K sentences in a $S_a \times S_a$ matrix and the top K sentences in $S_b \times S_b$ matrix.

5. Experimental Study and Discussion

5.1 Experimental Setup

In this section, a case study is presented to clarify how the proposed approach can be utilized by product designers to identify representative yet comparative sentimental sentences with specific product features from product online reviews efficiently.

21,952 pros and cons reviews of 583 intelligent mobile phones were collected from Cnet.com. They are utilized as the training corpus for the product feature extraction and sentiment polarity identification. To verify the availability of the proposed approach, in particular, 4,055 reviews of four popular mobile phones of different brands were obtained from Amazon.com. In consideration of data privacy, the names of the four products are represented as P1, P2, P3 and P4. The number of reviews of the four mobile phones is 905, 1,108, 1,088 and 954, respectively. In Figure 2, some statistics of these reviews are presented.

[Insert Figure 2]

As observed from this figure, in general, most reviews contain less than 10 sentences and are within 100 words, and only a few of them are found to have more than 60 sentences with more than 600 words. In particular, in this dataset, on average, there are 6.162 sentences in each review, but they are not distributed evenly, with the maximum of 80 sentences in a single review. A similar phenomenon is also found in terms of the word number per review, with an average 115.413 and a maximum of 2120. All 4,055 mobile reviews of the four products are employed in this case study to demonstrate how the proposed approach is applied to identify pairs of representative yet comparative sentimental sentences.

5.2 Evaluation Metrics

To evaluate the performance on a typical information retrieval problem, conventional metrics such as precision, recall and F_1 value are often utilized. However, in this research, it is generally difficult to obtain some training samples for evaluation from a big volume of product online reviews manually. Even for a small exemplary dataset that is less than 100 review sentences of two comparative products, it is still a tricky task to select some comparative sentence pairs by hand. This dilemma makes some metrics fail to be applied here to evaluate the performance of the proposed approach. Therefore, three evaluation metrics are borrowed, including information comparativeness, information representativeness and information diversity.

(a) Information comparativeness

The information comparativeness denotes to what extent the selected pairs of sentences cover similar topics. It is defined as the similarity between P_f and Q_f.

$$C(P_f, Q_f) = \frac{1}{K}\sum_{k=1}^{K} similarity(p_f^k, q_f^k) \tag{9}$$

(b) Information representativeness

The information representativeness denotes to what extent the selected pairs of sentences are capable to cover the information that is mentioned in the source review set. It is evaluated by the percentage of topics that are covered by the selected pairs of sentences. Let T_f^A and T_f^B be the topic set discovered from A_f and B_f, and let T_f^P and T_f^Q be the topic set from P_f and Q_f, then the information representativeness is

$$R(P_f, Q_f) = \frac{|T_f^P \cup T_f^Q|}{|T_f^A \cup T_f^B|} \tag{10}$$

(c) Information diversity

The information diversity denotes to what extent the selected sentences cover different topics. It is evaluated by the similarity within P_f and Q_f,

$$\begin{aligned} D(P_f, Q_f) &= 1 - \frac{1}{2}(similarity(P_f) + similarity(Q_f)) \\ &= 1 - \frac{1}{2}(avg(\sum_{i=1}^{K}\sum_{j=1, j\neq i}^{K} similarity(p_f^i, p_f^j)) + avg(\sum_{i=1}^{K}\sum_{j=1, j\neq i}^{K} similarity(q_f^i, q_f^j))) \end{aligned} \tag{11}$$

5.3 Results and Discussion

To show how competitive products can be compared by selecting a small number of pairs of sentences, 905 reviews of P1 and 1,108 reviews of P2 are analyzed as an illustrative example. Now, suppose designers care about opinions regarding the battery in these reviews. According to the approach introduced in Section 3.2, product features and related opinions are extracted from online reviews, and the number of sentences that refer to the battery of the two products is listed in Table 1.

[Insert Table 1.]

As observed from Table 1, it is time consuming to read all of battery-related sentences one by one for competitor analysis. Now, suppose two pairs of representative yet comparative sentences regarding the battery are expected to be selected. With the help of the proposed approach in Section 4, two pairs of representative yet comparative sentimental sentences regarding the battery can be extracted. Also, in Section 4, three strategies, namely "R-First", "C-First" and "D-First", are developed to select pairs of sentences, which potentially induces that different pairs of sentences are selected. To demonstrate the variance of selected sentences in different optimization strategies, selected sentences with different strategies are listed in Table 2, Table 3 and Table 4.

[Insert Table 2.]

[Insert Table 3.]

[Insert Table 4.]

As seen from these tables, with three optimization approaches, P1 and P2 are compared in terms of four sentimental sentence groups. In each group, two pairs of sentences are listed. Take the "Positive vs. Positive" group for instance. Two positive sentences in P1 reviews referring to the battery and comparative two positive sentences in P2 are considered. But these pairs of sentences are not ordered according to the information coverage degree since three strategies stress on different optimization targets, such as information representativeness, information comparativeness and information divergence. Also, as compared with results in three tables, the variance in concentrations of these approaches makes that different pairs of sentences are selected. Specifically, in Table 2, the information representativeness is the focus and selected pairs of sentences are expected to denote customer topics as many as possible. In Table 3, sentences are selected according to the information comparativeness and these sentences are observed to present similar topics. However, in Table 4, diversified information is expected to be drawn from online reviews, which makes selected sentences of the same product are dissimilar with each others. Additionally, as presented in these tables, these selected sentences seem to discuss more about the battery life. But designers might care also about comparative customer concerns in a fine grain level such as different aspects of product features. For instance, some consumers might discuss about the battery life, while others might complaint about the loose battery case. This question requires to identify customer concerns in a fine grain level and make further comparisons. This research topic is quite interesting and left for our future work.

To examine the performance of three greedy approaches, categories of experiments are conducted by analyzing review sentences referring to the battery, the application as well the screens of P1 and P2. In these experiments, three pairs of sentences are selected from the corresponding opinionated review sentence set, and the results are presented in Figure 3.

[Insert Figure 3.]

In these experiments, four groups of sentimental sentences are analyzed. In this figure, "P vs. P" indicates that positive sentences of P1 and positive sentences of P2 are compared, while "N vs. P" illustrates that negative sentences of P1 and positive sentences of P2 are analyzed. As observed from this figure, pairs of sentences that are selected by the "R-first" and the "C-First" approach present a high information comparativeness value. It can be claimed that pairs of sentences that are selected by both two approaches are highly similar to each other. However, if the information diversity is a major concern, the "D-First" approach is capable of selecting pairs of sentences that give different topics, which perform significantly better than the other two approaches.

Nevertheless, it can also be found that moderately low representative information values are obtained by all three approaches. The reason perhaps is that in these experiments, only three pairs of opinionated sentences are selected from each opinionated set. They account for a minor proportion of sentences. Hence, it is reasonable to cover only a few topics from the reviews. Another interesting phenomenon found is that somewhat higher information representativeness is gained by the "D-First" approach. Note that in each selection of the "D-First" approach, candidates that are more dissimilar with selected ones are prone to be chosen. This causes more different topics to be selected, which must lead to a higher information representativeness value.

In Figure 4, the categories of experiments regarding different numbers of pairs are conducted by analyzing review sentences referring to the batteries of P1 and P2. In this figure, "C", "R" and "D" denote the information comparativeness, information representativeness and information diversity.

[Insert Figure 4.]

As observed from this figure, the information representativeness begins to climb higher with an increasing number of selected pairs of sentences. It confirms the conjecture that a higher information representativeness will be gained if more sentences are selected. For the information diversity, it also increases as more pairs are involved. Nevertheless, it reaches a relatively stable peak and does not fluctuate much after about four pairs are chosen. Another interesting observation is that information comparativeness declines gradually with more selected pairs of sentences. A higher similarity within each pair of sentences is easy to achieve if only a few are selected. However, it is generally difficult to choose many pairs of comparative sentences from the review set of the different products, which leads to relatively lower information comparativeness values.

Note that in all of the above experiments, the similarity function that is denoted in Equation (4) is utilized. To check the influence induced by the difference of similarity functions, similar categories of experiments are conducted by employing the similarity function of Equation (5). All of these results are shown in Figure 5. Compared with Figure 4, similar trends are observed in terms of all three evaluation metrics, including that relatively higher information representativeness and relatively lower comparativeness are gained if more pair of sentences are selected and that the information diversity climbs to a plateau quickly once a few pairs are chosen.

[Insert Figure 5]

To demonstrate how the proposed approach benefit product designers, an example is presented in Figure 6. In this example, at first, customer reviews of two competitive products are obtained from a database of product review corpus. With the proposed approach, several pairs of comparative review sentences in the product feature level are obtained. For instance, exemplary review sentence pairs are selected associating with each product feature. Next, to analyze the advantages and disadvantages in a QFD house for customer requirement analysis, review sentence pairs are clustered. These sentence clusters can be viewed as voices of the customers, which are the input of QFD house. Also, different categories of product features are consolidated and translated into a list of engineering characteristics. Finally, with comparative pairs of review sentences and consolidated engineering characteristics, a QFD house can be formed, from which the advantages and disadvantages of products can be compared exactly.

[Insert Figure 6]

6. Conclusion

Big opinionated product review data provide valuable customer feedback, which indicates the strength and weakness of products for both potential consumers and product designers. In the past decade, many researchers in computer science and information management have paid much attention to how to extract and analyze customer requirements efficiently from big opinion data. It is extremely critical for product designers to understand customer requirements, give effective response and improve their products in the fierce market competition.

In this research, how to select a small number of opinionated sentences from product online reviews for competitor analysis is investigated. Its core is the selection of a small number of representative yet comparative sentences from reviews of competitive products. In particular, an optimization problem is formulated in which the information representativeness, the information comparativeness and the information diversity are considered. Different similarity functions that evaluate the similarity between sentences are analyzed, and three greedy algorithms are proposed to gain suboptimal solutions for the optimization problem. Moreover, categories of comparative experiments and profound analysis are conducted on a large number of real reviews. The sampled results demonstrate the effectiveness of the proposed approach. Potential research work can be extended in many directions, such as how to visualize these results in an interactive graphical user interface and how to compare products with the help of big opinionated product reviews in QFD, etc.

Acknowledgement

This work was supported by XXX.

References

Abbasi, A., Chen, H., and Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems*, 26(3):12:1-12:34.

Blythe, J. (2008). Essentials of Marketing (4th edition). Pearson Education. ISBN 978-0-273-71736-2.

Chen, L., Qi, L., and Wang, F. (2012). Comparison of feature-level learning methods for mining online consumer reviews. *Expert Systems with Applications*, 39(10):9588-9601.

Das, A. and Bandyopadhyay, S. (2010). Topic-based bengali opinion summarization. In *COLING*'10, pages 232-240.

Ding, X. and Liu, B. (2007). The utility of linguistic rules in opinion mining. In *SIGIR*'07.

Ding, X., Liu, B., and Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In *WSDM*'08.

Dou, X., Walden, J. A., Lee, S., and Lee, J. Y. (2012). Does source matter? Examining source effects in online product reviews. *Computers in Human Behavior*, 28(5):1555-1563.

Fang, Y., Si, L., Somasundaram, N., and Yu, Z. (2012). Mining contrastive opinions on political texts using cross-perspective topic model. In *WSDM*'12, pages 63-72.

Ganapathibhotla, M. and Liu, B. (2008). Mining opinions in comparative sentences.

In *COLING*'08, pages 241-248.

Goorha, S. and Ungar, L. (2010). Discovery of significant emerging trends. In *KDD*'10, pages 57-64.

Hedegaard, S. and Simonsen, J. G. (2013). Extracting usability and user experience information from online user reviews. In *CHI*'13, pages 2089-2098.

Jakob, N. and Gurevych, I. (2010). Extracting opinion targets in a single- and cross-domain setting with conditional random fields. In *EMNLP*'10, pages 1035-1045.

Jin, J., Ji, P. and Liu, Y. (2014) Prioritising engineering characteristics based on customer online reviews for quality function deployment. *Journal of Engineering Design*. 25:303-324

Jindal, N. and Liu, B. (2006). Identifying comparative sentences in text documents. In *SIGIR*'06, pages 244-251.

Ju, S., Li, S., Su, Y., Zhou, G., Hong, Y., and Li, X. (2012). Dual word and document seed selection for semi-supervised sentiment classification. In *CIKM*'12, pages 2295-2298.

Kim, H. D. and Zhai, C. (2009). Generating comparative summaries of contradictory opinions in text. In *CIKM*'09, pages 385-394.

Kim, S.-M. and Hovy, E. (2006). Automatic identification of pro and con reasons in online reviews. In *COLING'06*, pages 483-490.

Korenek, P. and Simko, M. (2013). Sentiment analysis on microblog utilizing appraisal theory. *World Wide Web*, August:1-21.

Lappas, T., Crovella, M., and Terzi, E. (2012). Selecting a characteristic set of reviews. In *KDD*'12, pages 832-840.

Li, F., Han, C., Huang, M., Zhu, X., Xia, Y.-J., Zhang, S., and Yu, H. (2010). Structure-aware review mining and summarization. In *COLING*'10, pages 653-661.

Li, H., Bhowmick, S. S., and Sun, A. (2010b). Affnity-driven prediction and ranking of products in online product review sites. In *CIKM*'10, pages 1745-1748.

Li, S., Ju, S., Zhou, G., and Li, X. (2012). Active learning for imbalanced sentiment classification. In *EMNLP-CoNLL*'12, pages 139-148.

Li, S., Zha, Z.-J., Ming, Z., Wang, M., Chua, T.-S., Guo, J., and Xu, W. (2011a). Product comparison using comparative relations. In *SIGIR*’11, pages 1151-1152.

Li, X., Hitt, L. M., and Zhang, Z. J. (2011b). Product reviews and competition in markets for repeat purchase products. *Journal of Management Information Systems*, 27(4):9-42.

Lin, C. and He, Y. (2009). Joint sentiment/topic model for sentiment analysis. In *CIKM'09*, pages 375-384.

Liu, Y., Jin, J., Ji, P., Harding, J. A., and Fung, R. Y. K. (2013). Identifying helpful online reviews: A product designers perspective. *Computer-Aided Design*, 45:180-194.

Lu, Y., Wang, H., Zhai, C., and Roth, D. (2012). Unsupervised discovery of opposing opinion networks from forum discussions. In *CIKM*’12, pages 1642-1646.

Lu, Y., Zhai, C., and Sundaresan, N. (2009). Rated aspect summarization of short comments. In *WWW*’09, pages 131-140.

Ly, D. K., Sugiyama, K., Lin, Z., and Kan, M.-Y. (2011). Product review summarization from a deeper perspective. In *JCDL*’11, pages 311-314.

Ma, Z., Sun, A., Yuan, Q., and Cong, G. (2012). Topic-driven reader comments summarization. In *CIKM*’12, pages 265-274.

Moghaddam, S., Jamali, M., and Ester, M. (2011). Review recommendation: Personalized prediction of the quality of online reviews. In *CIKM*’11, pages 2249-2252.

Moghaddam, S., Jamali, M., and Ester, M. (2012). ETF: Extended tensor factorization model for personalizing prediction of review helpfulness. In *WSDM*’12, pages 163-172.

Mukherjee, A. and Liu, B. (2012). Mining contentions from discussions and debates. In *KDD*’12, pages 841-849.

Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3):521-543.

Pang, B. and Lee, L. (2004). A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In *ACL'04*.

Paul, M. J., Zhai, C., and Girju, R. (2010). Summarizing contrastive viewpoints in opinionated text. In *EMNLP*’10, pages 66-76.

Tsaparas, P., Ntoulas, A., and Terzi, E. (2011). Selecting a comprehensive set of reviews. In *KDD*’11, pages 168-176.

Tucker, C. and Kim, H. M. (2011). Predicting emerging product design trend by mining publicity available customer review data. In *ICED*’11, pages 43-52.

Wang, L., Youn, B. D., Azarm, S., and Kannan, P. K. (2011). Customer-driven product design selection using web based user-generated content. In *IDETC*’11, pages 405-419.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *EMNLP'05*, pages 347-354.

Xu, K., Liao, S. S., Li, J., and Song, Y. (2011). Mining comparative opinions from customer reviews for competitive intelligence. *Decision Support Systems*, 50(4):743-754.

Yu, J., Zha, Z.-J., Wang, M., and Chua, T.-S. (2011). Aspect ranking: identifying important product aspects from online consumer reviews. In *ACL'11*, pages 1496-1505.

Zhai, Z., Liu, B., Xu, H., and Jia, P. (2010). Grouping product features using semi-supervised learning with soft-constraints. In *COLING'10*, pages 1272-1280.

Zhan, J., Loh, H. T., and Liu, Y. (2009). Gather customer concerns from online product reviews - a text summarization approach. *Expert Systems with Applications*, 36(2 Part 1):2107-2115.

Zhang, Z., Guo, C., and Goes, P. (2013). Product comparison networks for competitive analysis of online word-of-mouth. *ACM Transactions on Management Information Systems*, 3(4):1-20.

Zhuang, L., Jing, F., and Zhu, X. Y. (2006). Movie review mining and summarization. In *CIKM*’06, pages 43-50.

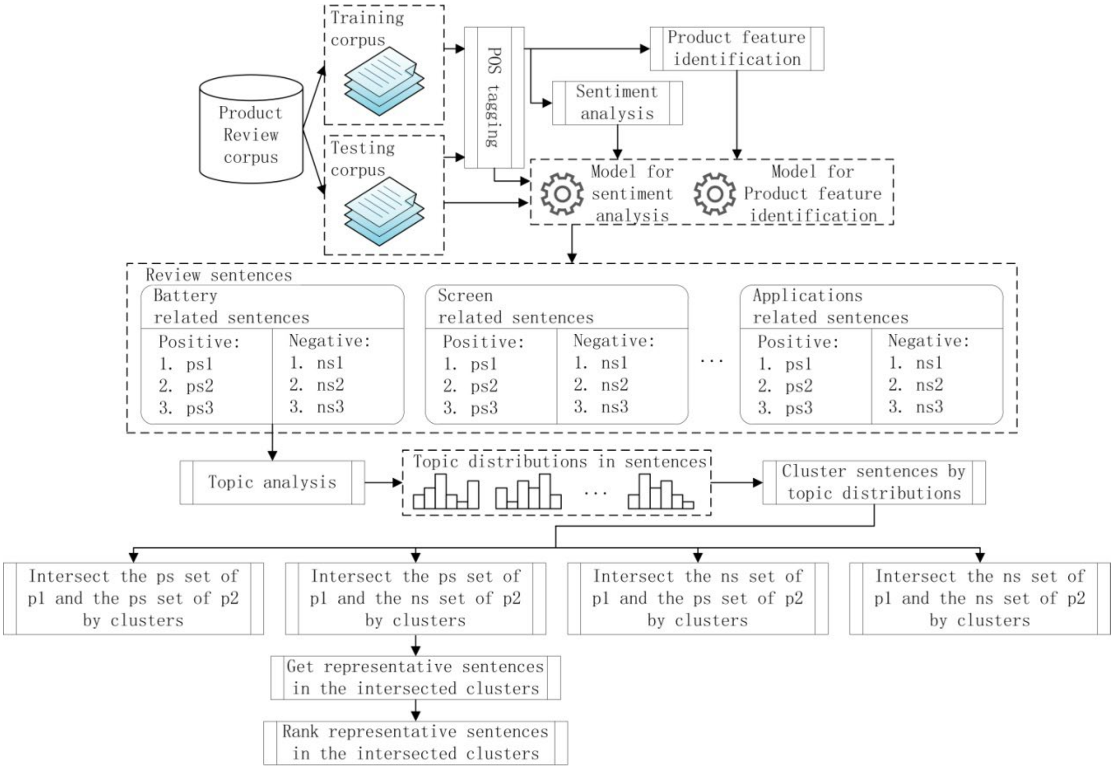
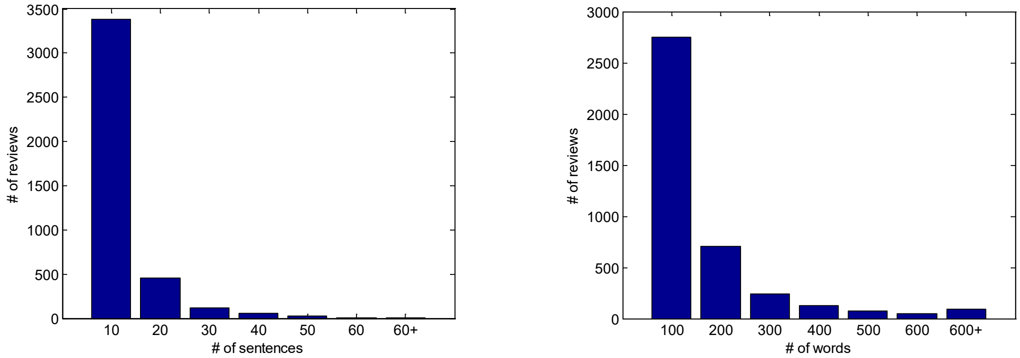
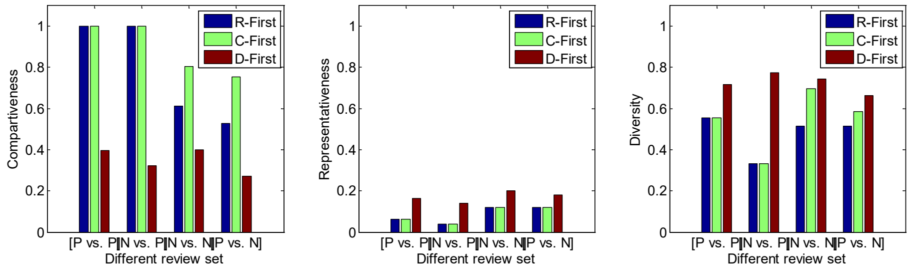


Figure 1. A Framework for Mining Comparative Viewpoints from Product Online Reviews

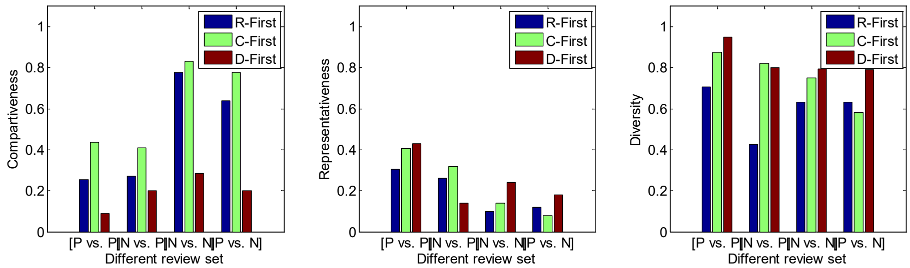


(a) # of reviews vs. # of sentences (b) # of reviews vs. # of words

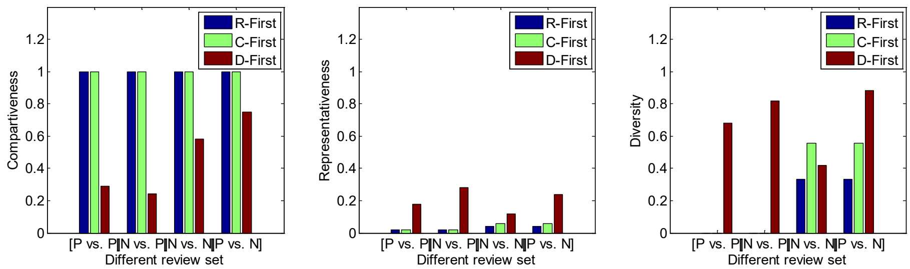
Figure 2. Some statistics of 4,055 reviews of four popular mobiles



(a)Battery Reviews



(b)Application Reviews



(c)Screen Reviews

Figure 3. Selecting three pairs of opinionate review sentences of P1 and P2.

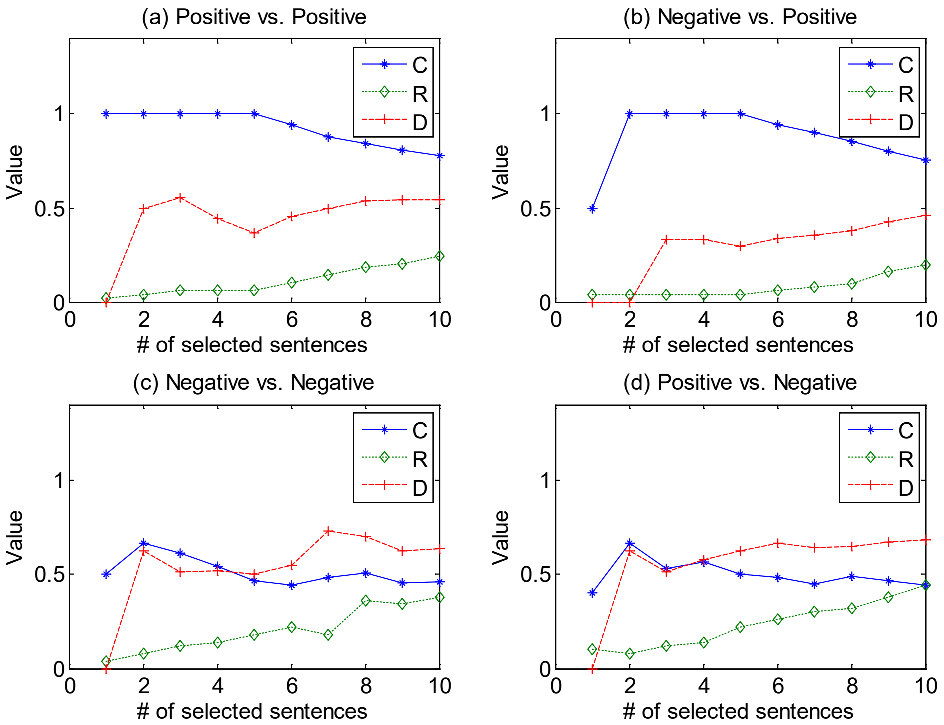


Figure 4. Sentences referring to the battery in reviews of P1 and P2 with Equation (4)

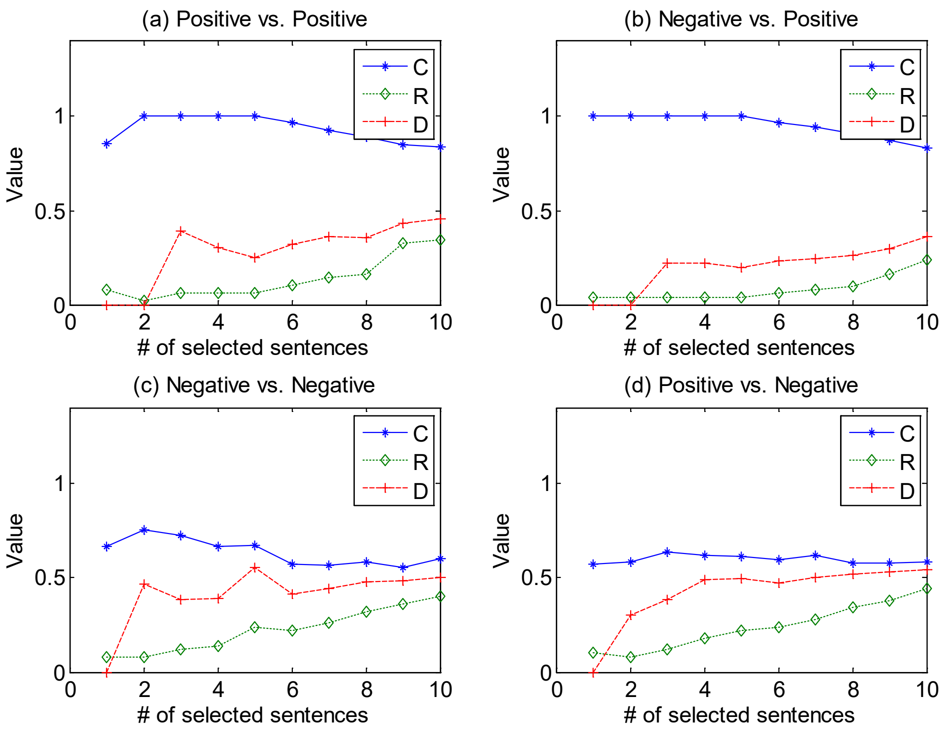


Figure 5. Sentences referring to the battery in reviews of P1 and P2 with Equation (5)

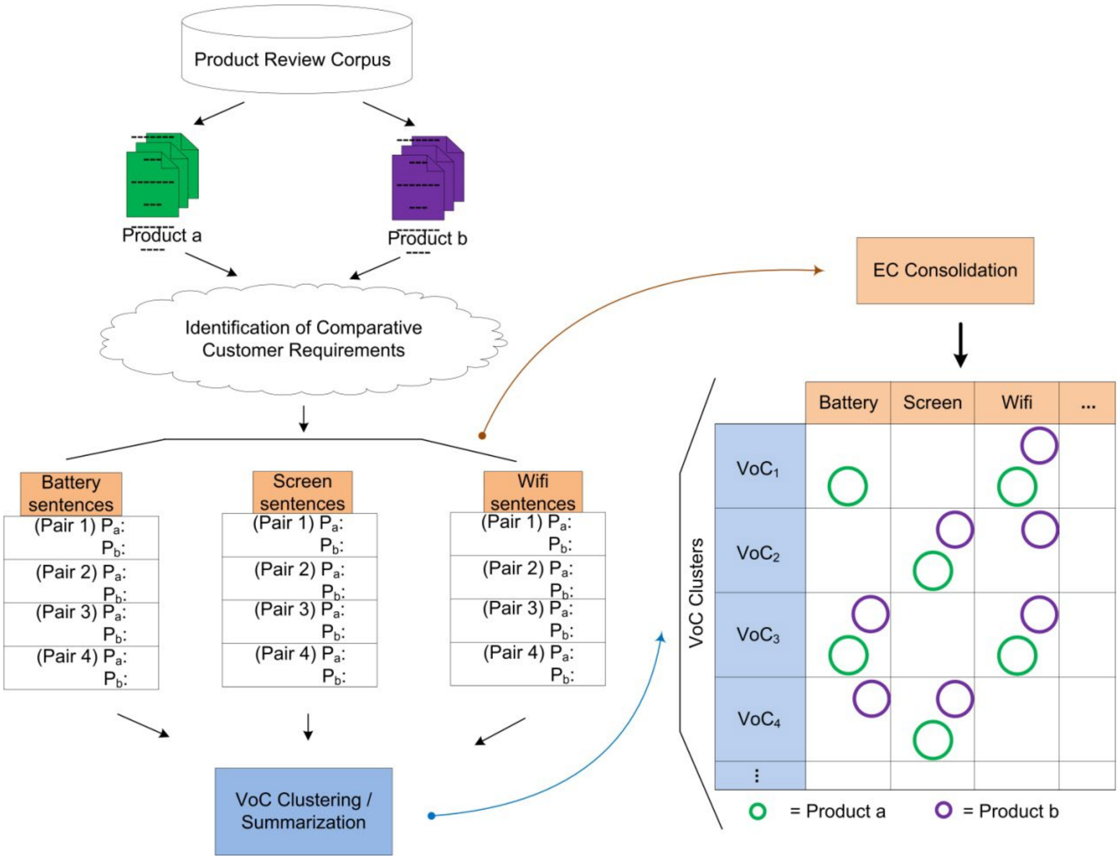


Figure 6. An example to utilize the proposed approach for competitor analysis.

	# of positive	# of negative	# of neutral	Total
P1	45	78	31	154
P2	28	53	24	105

Table 1. # of sentences referring to the battery in reviews of P1 and P2

Sentiment	Group #	Pair of sentences
Positive vs. Positive	1	The battery life is really good . THe battery life is pretty good .
	2	Very good battery life too . Good battery life .
Negative vs. Negative	1	Battery life got worse as it was used . Battery life is not much longer than I expected .
	2	Battery life so far average . I 've accidently drained the battery .
Negative vs. Positive	1	Battery life got worse as it was used . Battery life is excellent as well .
	2	Battery life so far average . Battery life is good .
Positive vs. Negative	1	Battery life is excellent as well . Battery life is not much longer than I expected .
	2	Battery life is good enough for the amount of processing the phone does . I 've accidently drained the battery .

Table 2. Two pairs of sentences that are selected from reviews of P1 and P2 with the "R-First" approach

Sentiment	Group #	Pair of sentences
Positive vs. Positive	1	does have very good battery life Works great , good reception and battery life .
	2	The battery life is really good . The battery life is pretty good .
Negative vs. Negative	1	Battery life got worse as it was used . Battery life is not much longer than I expected .
	2	Battery life so far average . I 've accidently drained the battery .
Negative vs. Positive	1	Battery life got worse as it was used . Battery life is excellent as well .
	2	Battery life so far average . Battery life is good .
Positive vs. Negative	1	Battery life is excellent as well . Battery life is not much longer than I expected .
	2	Battery life is good enough for the amount of processing the phone does . I 've accidently drained the battery .

Table 3. Two pairs of sentences that are selected from reviews of P1 and P2 with the "C-First" approach

Sentiment	Group #	Pair of sentences
Positive vs. Positive	1	Battery life is more than 24 hours with moderate use . Battery life is excellent as well .
	2	Battery life is great and camera takes good pictures . Battery life is good , as expected with a GSM phone .
Negative vs. Negative	1	but not much better battery life than my old phone. The battery life was relavively poor
	2	When I first bought the old Lumia 521 it had about the same battery life . sometimes I would even have to remove the battery and put it back in before I could get the phone to turn back on .
Negative vs. Positive	1	The battery life is not better than my old phone . Battery life is excellent as well .
	2	Battery life , call quality , apps , all of the things are not better. Battery life is good , as expected with a GSM phone .
Positive vs. Negative	1	Overall , a phone with good battery life . On occasion the phone will not turn on and you have to take the battery out and put it back in to get it to respond , battery life is horrible as well .
	2	does have very good battery life The battery had n't a longer life than I expected .

Table 3. Two pairs of sentences that are selected from reviews of P1 and P2 with the "D-First" approach