Extraction of Comparative Opinionate Sentences from Product Online Reviews

Ping JI[1,2]

Jian JIN[3,1]

[1] The Hong Kong Polytechnic University Shenzhen Research Institute
Shenzhen 518057, China

[2] Department of Industrial and Systems Engineering,
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

[3] Department of Information Management
Beijing Normal University
Beijing 100875, China

Abstract—**Big 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 opinionate sentences associated with some specific features from product online reviews has been investigated by some researchers. However, few of them examined how to select a small number of representative yet comparative sentences for competitor analysis. In this research, a framework is illustrated to select pairs of opinionate sentences referring to a specific feature from reviews of competitive products. With the help of the techniques on sentiment analysis, opinionate sentences referring to a specific feature are first identified from product online reviews. Then, for the selection of a small number of representative yet comparative opinionate sentences, information representativeness, information comparativeness and information diversity are investigated. 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; product design; text mining;

I. INTRODUCTION

Nowadays, consumers are offered more options to select and differentiate products online. They gain opportunities to compare similar products and pick their favorites. With hundreds or even thousands of products easily found by product search engines, a new dilemma is that too many similar ones are recommended in e-commerce websites. Consumers must make comparisons, find the pros and cons among the competitors, and choose the most suitable ones. Perhaps one simple approach to understand the pros and cons among competitors is to read online reviews of products. Online reviews provide rich information about consumers' concerns. They allow potential consumers to get a general idea regarding different products. In addition, these reviews provide suggestions from consumers' feedback to product designers, which may assist to improve their products.

However, it is generally difficult to understand all reviews in different websites for competitive products manually. If only a limited number of reviews are covered, critical consumer comments might be neglected. In the past decade, some researchers paid much attention to how to analyze such big consumer data intelligently [1-3]. For instance, many publications about opinion mining for online reviews were reported to have discussed how to infer sentiment polarities 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 [4-5]. These studies help designers to understand a large amount of customer requirements in online reviews and strive for product improvements to achieve a higher level of customer satisfaction. However, 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 product improvements to seek any 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. The essence of this problem is how to digest big consumer data to offer designers some representative review sentences of different products. Specially, these review sentences should be descriptive about general consumer concerns and, at the same time, they are expected to be comparative to reflect contrasting consumer feedback of different products. Thus, in this research, the ultimate goal is to identify several pairs of representative yet comparative sentimental sentences with specific product features from product online reviews. Accordingly, selected review sentences of different products are required to be opinionate ones, and they necessitate referring to specific product features. In addition, selected review sentences should characterize online consumer requirements in several aspects:

a) selected review sentences should be descriptive and representative about general consumer requirements;

b) selected review sentences need to be comparative, which means that they discuss similar topics of products;

c) selected review sentences are expected to be diversified to reflect various consumer requirements.

Hence, in this research, opinionate sentences referring to specific product features are initially extracted from product reviews by employing a supervised learning approach. Next, to take all three aspects into consideration, an optimization problem is formulated for review sentence selection. In addition, different functions that evaluate the similarity between sentences are utilized, and greedy algorithms are proposed to analyze the optimization problem for suboptimal solutions. These approaches aid designers in obtaining a small number of pairs efficiently for competitor analysis.

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. According to this framework, in Section 4, an optimization perspective is proposed to identify representative review sentences for competitive products. Different similarity functions between sentences are designed, and greedy algorithms are described for the optimization problem. Section 5 presents comprehensive details of the experimental study utilizing a large number of reviews from Amazon.com and discusses the results. Section 6 concludes this research.

II. Related Work

A. Contrastive Viewpoints Extraction

A two-stage method was proposed to summarize multiple contrastive viewpoints from opinionated text [6]. In the first stage, an extended LDA (Latent Dirichlet Analysis) model was utilized to extract topics and viewpoints from texts with different types of features. In the second stage, a modified PageRank method was employed to summarize comparative sentences. Mukherjee and Liu first utilized a topic model to extract topics and expressions indicating contention and agreement topics [7]. 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 [8].

A framework for contrastive opinion summarization was proposed [9]. 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 [10]. 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 [11]. Then, a naive Bayesian classifier was utilized to handle the case that a single review sentence matches several rules. Finally, the prediction from the naive Bayesian classifier was utilized to decide whether a sentence is comparative. However, Xu et al. argued that 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 [12]. 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 formation of English words, non-equal gradable comparatives and superlative comparatives are summarized [13]. 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.

B. Review Sampling

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 [14]. 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 [15]. 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 [16]. 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 [17]. 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, two approaches are applied in an active learning algorithm for imbalanced sentiment classification.

C. Product Online Reviews for Engineering Design

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 [18]. 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 [19]. 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 [20]. 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 [21]. 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 [22]. 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 [23]. First, product attributes were extracted. Next, 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 customer reviews, an ordinal classification approach was advised to prioritize ECs for QFD [24]. It is a pairwise approach in which customer opinions in online reviews are deemed features and the overall customer satisfaction is the target value.

III. A FRAMEWORK FOR MINING COMPARATIVE VIEWPOINTS FROM PRODUCT ONLINE REVIEWS

A. Framework

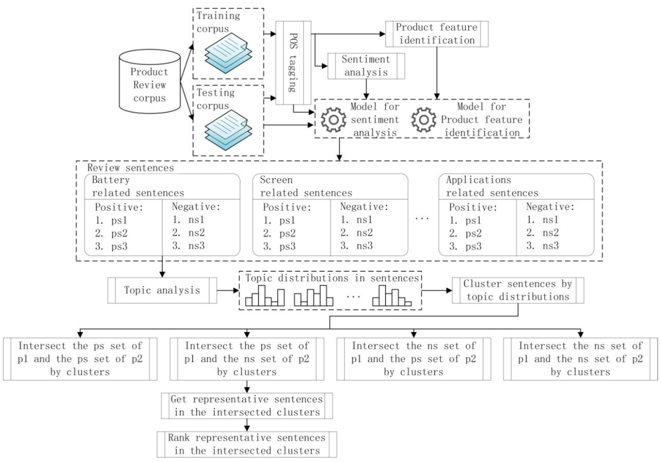


Figure 1. A Framework for Mining Comparative Viewpoints

To identify representative yet comparative sentimental sentences with specific product features from product online reviews, a framework is presented in Figure 1. POS tagging is conducted first, which is utilized for the analysis of sentiment polarities and the identification of product features. In this research, two simple but effective supervised learning models are utilized for these tasks. Given customer online reviews in the same product domain, the two models aid designers in extracting product features with the corresponding sentiment polarities efficiently.

Opinion consumer data of different products can be categorized by product features with contrasting sentiment polarities. Note that one central objective of this research is to identify representative yet contrasting sentences from a big volume of customer reviews. As discussed in Section 1, one of the three aspects of the review sentence selection is that selected sentences need to be descriptive and representative of general consumer 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 by consumers in different product reviews. Accordingly, topic analysis is conducted on categorized opinion data, which helps to distinguish topic distributions regarding consumers 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 opinionate sentences with the same product features are clustered in which similar customer topics are discussed. Actually, these cluster results help to obtain groups of opinionate 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 opinionate 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.

B. 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 [25, 26]. 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 [27, 28].

Many review sites invite consumers to post both compliments and criticisms of products they have purchased. For instance, a typical 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 on the pros and cons lists, and frequently referred to nouns or noun phrases are regarded as product features. In addition, consumers may utilize different words to describe the same product feature. For example, consumers 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 customer online reviews. For instance, "apps" and "applications" are utilized interchangeably by mobile consumers. 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 customer online reviews.

In addition, Pang and Lee developed a publicly available subjective dataset, which includes 5,000 subjective and 5,000 objective sentences [29]. 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 consumers 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 [30] are employed in a binary Naive Bayes classifier. This classifier is utilized to analyze the sentiment polarity of review sentences.

IV. COMPARATIVE VIEWPOINTS IDENTIFICATION

A. 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 consumer concerns about f efficiently, two small subsets of opinionate 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, as discussed in Section 1, review sentences in P_f and Q_f from A_f and B_f should follow three principles:

(a) 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. It illustrates that selected sentences in P_f are representative ones that describe a general idea about those that are described in A_f. Likewise, *similarity*(Q_f, B_f - Q_f), should be as high as possible.

(b) The similarity between P_f and Q_f, *similarity*(P_f, Q_f), should be as high as possible. It means that selected sentences in P_f and selected sentences in Q_f are expected to discuss similar topics regarding customer concerns.

(c) The similarity between each pair of sentences within P_f, *similarity*(P_f), should be as low as possible. It 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.

B. An Optimization Perspective

Generally, three principles are pressed 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,

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

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

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

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.

C. *Similarity Functions*

As discussed, topics are identified from online reviews of competitive products. These topics help to discern consumers concerns of products. In addition, in this research, it is required that selected sentences are descriptive about the general topics of consumers 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; two similarity functions are defined as.

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

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

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 [9].

D. *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 since 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

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

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

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

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

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

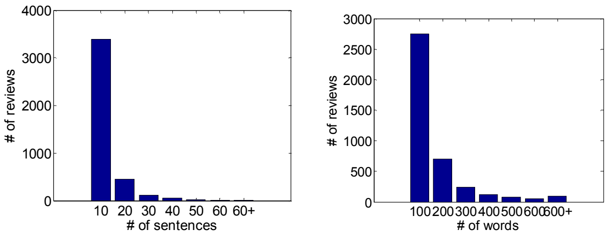
The comparison cost is $S_a \times S_a$ + $S_b \times S_b$ because sentences are selected from the top K sentences in a $S_a \times S_a$ matrix and the top K sentences in $S_b \times S_b$ matrix.

V. Experimental Study and Discussion

A. Experimental Setup

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



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

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

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.

B. Evaluation Metrics

In Section 4, an optimization problem is defined to extract pairs of representative yet comparative sentimental sentences, and different greedy algorithms are presented to obtain suboptimal solutions. In addition, two similarity functions are nominated in the optimization problem. To evaluate the proposed greedy algorithms, three statistics are utilized, which are also employed for the performance comparison regarding different approaches.

(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,

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

C. 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.

TABLE I. # OF SENTENCES REFERRING TO THE BATTERY IN REVIEWS OF P1 AND P2

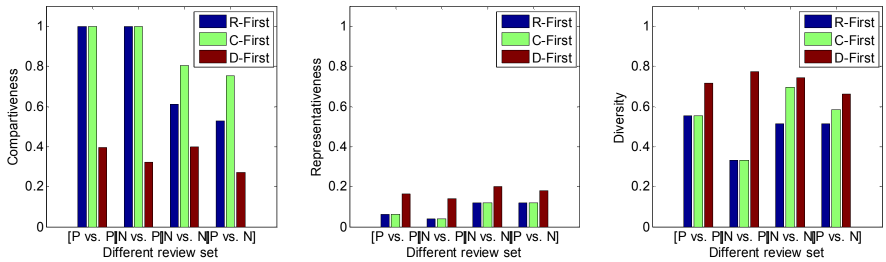
	# of positive	# of negative	# of neutral	Total
P1	45	78	31	154
P2	28	53	24	105

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 are shown in Table 2.

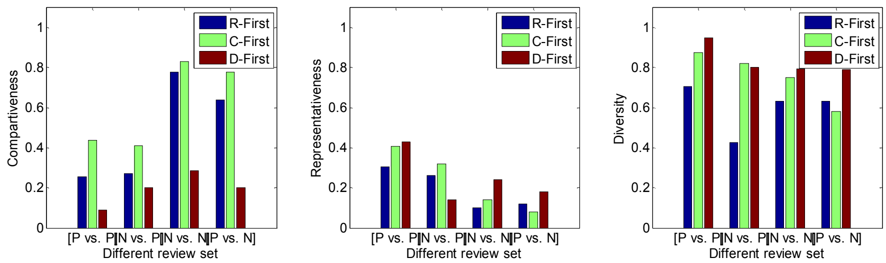
In this table, P1 and P2 are compared in terms of four sentimental sentence groups. In each sentence group, two pairs of sentences are listed. Take the "Positive vs. Positive" group for instance. Positive sentences in P1 reviews referring to the battery and that of positive sentences in P2 are considered. To facilitate designers to obtain a general idea of these sentences and make comparisons with the two products, two pairs of sentences are selected. In the first pair, the consumer of P1 and the consumer of P2 describe the good "battery life", while in the second pair, both selected sentences of P1 and P2 praised for the fact that the battery lasts long enough.

TABLE II. TWO PAIRS OF REVIEW SENTENCES OF P1 AND P2 IN EACH SENTIMENT GROUP

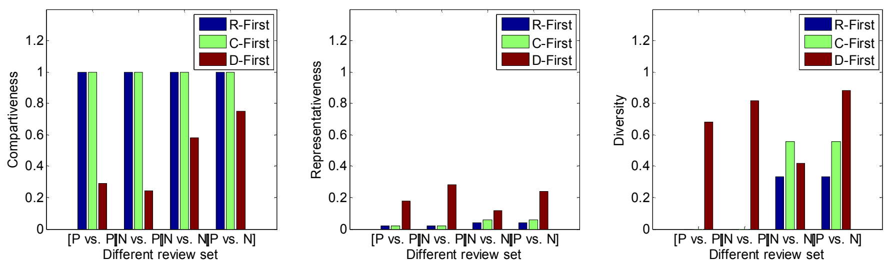
Sentiment	P #	Pair of sentences
Positive vs. Positive	1	P1: Fast response time , great battery life , easy set up , fun apps . P2: Battery life is good , as expected with a GSM phone .
	2	P1: Ihad no issues , battery lasts long enough . P2: The battery lasts a long time for me since I do n't leave the phone on all the time .
Negative vs. Negative	1	P1: The back cover popped off and the battery flew out . P2: I have another problem because I can't open the damn back cover to get to the damn battery .
	2	P1: I am not a fan of the back cover, that is having to press in on it and pry it off with my nails to access the battery , SIM and memory card . P2: Because I have bought a rubber protective cover so now I just kept the back cover off to make it ease to access the battery as I expect removing the battery may happen more often than not
Positive vs. Negative	1	P1: Great battery , it is very efficient with its battery usage , so although it is a small battery , I have easily made it through the day still with 50 % battery left . P2: And even though I do n't use my phone much , after a year of light use , the battery completely drains within a day or so even when it is mostly on standby , so I have to recharge it a lot .
	2	P1: battery life great - approx day and a half without using battery saver . P2: Sometimes when I do n't use it for a day , it takes up to a minute to turn on , but it 'll be frozen (to reboot , you take the battery out for a few minutes) .
Negative vs. Positive	1	P1: If the battery is designed to only work well for one month , then I 'd say send me at least a year 's supply of batteries rather than replacing the entire phone (Which I love when it 's working , by the way !) P2: The battery lasts a long time for me since I do n't leave the phone on all the time .
	2	P1: Battery lasts for couple days with the wifi on . P2: As an added bonus , the battery charge appears to last significantly longer than with my last LG phone .



(a)Battery Reviews



(b)Application Reviews



(c)Screen Reviews

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

Three heuristic greedy approaches are proposed in Section IV to obtain suboptimal solutions. To examine the performance of the different 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 opinionate review sentence set, and the results are presented in 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 opinionate sentences are selected from each opinionate 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.

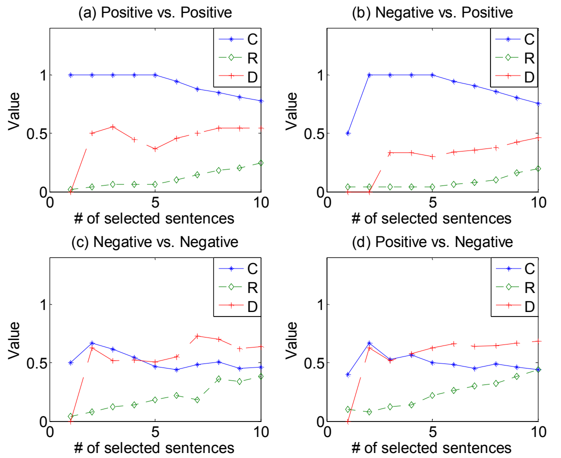


Figure 4. Sentences referring to the battery in reviews of P1 and P2 with Equation (4)

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.

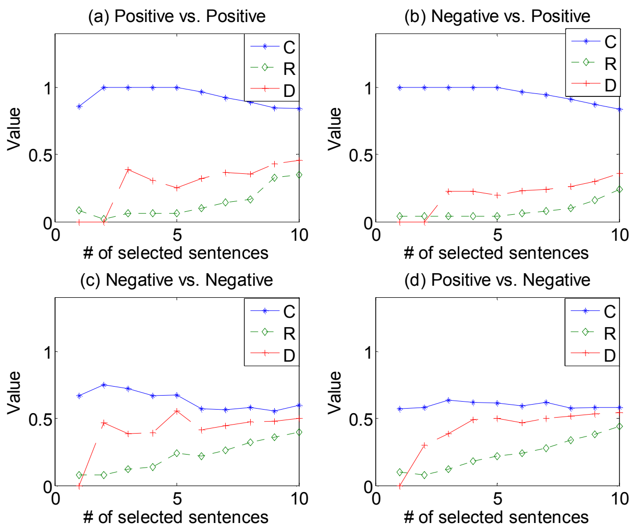


Figure 5. Sentences referring to the battery in reviews of P1 and P2 with Equation (5)

VI. CONCLUSION

Competitor analysis in customer-driven NPD is one essential task. It requires designers to comprehend the strengths and weakness of competitors. Online reviews enable designers to obtain sufficient messages regarding customer concerns of different products. However, the sheer volume of online reviews surpasses the ability of designers to grasp critical information for competitor analysis. Hence, an efficient approach is imperative to extract informative, representative and comparative customer concerns.

In this research, how to select a small number of opinionate 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, how to compare products with the help of big opinionate product reviews in QFD, etc.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. NSFC 71271185/G011201), the Fundamental Research Funds for the Central Universities (No. SKZZX2013091) and the Research Project of The Hong Kong Polytechnic University (No. G-YBE3).

REFERENCES

[1] 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.

[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.

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

[4] 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. 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.

[5] 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.

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

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

[8] 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.

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

[10] 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.

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

[12] 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.

[13] Ganapathibhotla, M. and Liu, B. (2008). Mining opinions in comparative sentences. In *COLING*'08, pages 241-248.

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

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

[16] 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.

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

[18] 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.

[19] 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.

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

[21] 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.

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

[23] 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.

[24] 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

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

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

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

[28] 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.

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

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