Comparison of Series Products from Customer Online Concerns for Competitive Intelligence

Journal of Engineering Design

Jian Jin[a], Ping Ji[b] and Sixing Yan[a]

[a] Department of Information Management, School of Government,
Beijing Normal University,
Beijing 100875, China

[b] Department of Industrial and Systems Engineering
The Hong Kong Polytechnic University,
Hong Kong SAR, China

* Corresponding author:

Tel: +86-10-5880 5387

Email: jinjian.jay@bnu.edu.cn

Acknowledgement

The work described in this paper was supported by a grants from the National Nature Science Foundation of China (project no. NSFC 71701019/G0114) and two Research Projects of The Hong Kong Polytechnic University (Project No. G-YBLR and G-YBFE).

ABSTRACT Online reviews provide valuable information for product designers and the integration of online concerns into new product design has been investigated by different researchers. However, few of them exploit the value of online concerns on the comparison of series products. Analyzing online concerns of series products facilitates designers to obtain shared customer preferences regarding products in a series and recognize the strength and weakness of products in competitive series. Accordingly, a framework is designed to discover shared pros and cons of series

products by exploring online customer concerns, in which representative opinionated sentences are sampled from reviews of series products. In particular, opinionated sentences of specific features are initially identified from product reviews. Then, opinionated sentences regarding the same series products are clustered, which helps to extract similar customer concerns. Finally, an optimization problem is formulated for the sampling of a few opinionated representative sentences. With a large number of real data from Amazon.com, categories of experiments were conducted to evaluate the effectiveness of the proposed approach. This study explores to integrate big consumer data for competitive intelligence in the market driven new product design, which helps the theoretical development on customer requirement management in the fierce market.

Keywords: product comparison; series product; customer concerns; online reviews; competitive intelligence

Comparison of Series Products from Customer Online Concerns for Competitive Intelligence

1. INTRODUCTION

With the prevailing of digital retailers like Amazon.com and Taobao.com, product online reviews became an important type of information for both potential customers and product designers since valuable opinionated text about customer concerns are expressed. Online reviews help potential consumers to grasp the pros and cons of products and benefit product designers in understanding customer concerns as well. Generally, a big volume of online reviews are posted from time to time, which induces that it is time-consuming and tedious to digest the entire set of customer opinions. This problem interests researchers in different fields and various methods are reported to analyze such big customer data intelligently (Korenek and Šimko, 2013; Chen et al., 2015; Tao et al, 2010; Jin et al., 2016; Shen and Wang, 2016).

With different intelligent algorithms for processing big customer data available, however, there is still lack of operative intelligent approaches for designers to analyze a large volume of customer online concerns about products in the same series. A limited number of studies are noted to utilize customer online reviews for product comparison (Chen, Qi and Wang, 2012; Jin, Ji and Gu, 2016; H. D. Kim and Zhai, 2009). These studies select opinionated sentences from customer reviews to construct comparative summaries, which help to detect the contradictory opinions from a big volume of online reviews. But product serializations usually imply that a series of products in one brand should concerned at the same time and, then, product comparison becomes the comparison of multiple similar products in the same series. Nonetheless, many existing methods report that the extraction of contrastive viewpoint from customer online concerns mainly focuses on one-to-one product comparison, which is not aligned with the analysis about customer concerns or product defects in series products well. For instance, exemplary online customer review sentences of three smart phones in one series are shown in Table 1.

Table 1. Sentences extracted from reviews of three smart phones in one series

Series product	Review sentence
Product 1	The camera takes great photos and video is amazing, it has so many features and it's simple and fun to use the phone with family
Product 2	I'd also like to add that I have successfully changed my entire family over for school and professional photography development reasons
Product 3	And with the simplify music app I can share music and photos on my phone with my family and friends over 3g

As presented, in all of these three sentences, customers discuss the experience about sharing photos with their families and present positive opinions on this particular feature. An effective approach to identify shared comparative customer concerns over series products are required, which help to enlighten designers in understanding the strength and weakness about the brand. Practically, analyzing series products enable designers to identify the shared pros and cons of different products and facilitate to understand customer final purchase decisions towards various products of different series in competitive brands as well. Also, theoretically, comparison of series products from online concerns assists to make up research studies towards customer requirement management from the perspective of competitor analysis and promote the development of comparative customer requirement identification for new product design.

Hence, in this research, a framework that representative shared customer feedbacks are extracted from reviews of series products is proposed, which aims to highlight and compare similar customer concerns of series products, as well as conduct multiple-to-multiple comparisons on comparative products. In particular, product features and corresponding customer sentiment polarities are initially extracted. Next, sentences regarding the same product feature aspect are identified via a clustering algorithm. Finally, both information coverage and information diversity are reckoned to formulate an optimization problem for representative sentences sampling, in which different functions regarding sentence similarity are tested in the experiments.

The contributions of this research are at least two-fold.

(1) A framework of representative sentence sampling from online reviews of series products is outlined. Specifically, extracted sentences regarding strengths and weaknesses of multiple similar but competitive products can be utilized for intelligent marketing and product comparisons. It also highlights the significance of research studies associated with identifying comparative customer requirements of series products via mining online concerns for new product design

(2) To make effective comparisons on series products, representative sentences sampling from online customer concerns is formulated as an optimization problem and a greedy algorithm is designed to obtain sampling results effectively. The sampling results contain details of product features and cover different products of the series, in which relations of series products are illustrated. It facilitates product designers to make effective comparisons on series products of brand competitors in specific features.

The rest of this research is structured as follows. In Section 2, relevant studies are briefly reviewed. Section 3 outlines the problem statement of this research. In Section 4, technical details about the proposed method are carefully described. In Section 5, comparative experiment studies are presented and analyzed with a large number of online reviews. Finally, this research is concluded in Section 6.

2. RELATED WORK

In this study, to identify shared comparative customer concerns over series products, a representative sentence subset is sampled from a given review corpus, which relates to extant research studies on review summarization and review selection. Additionally, comparisons on similar products are often formulated as how to detect compared opinionated segments. Hence, research studies on contrastive viewpoint extraction are also briefly reviewed.

2.1 Review summarization

Studies of review summarization mainly focus on how to extract topics of customer reviews and present the sketch outline of a given reviews corpus.

An optimization model was built for summarization via salient sentences (Alguliev, Aliguliyev and Isazade, 2013). The salient sentence was defined as whether it covers the main content of the entire documents. Accordingly, sentences were weighted according to the term frequency–inverse sentence frequency and three types of relations between were considered for sentence selection, including sentence-to-document, summary-to-document, sentence-to-sentence. Another approach for informative sentence selection was introduced in (Zhu et al., 2013). The review summarization was formulated as a community-leader detection problem. The community was defined as a cluster of sentences regarding the same aspect of an entity while the leader was the most informative sentences among that cluster. Specifically, first, a graph was constructed where the nodes were the sentence and links was denoted as the similarity between two sentences. Second, the informative score of each sentence was estimated according to the helpfulness of the related original review. Finally, communities were detected in this graph and sentences that were considered as leaders were selected as results. An extractive summarization was generated via sentences removing (Bonzanini, Martinez-Alvarez and Roelleke, 2013). First, topics of each sentence were estimated using a Bayesian modeling approach. Then sentences that were relevant to the given topics were defined as representative one. Accordingly, sentences with low representativeness were removed until the total representativeness of selected subset was the highest.

A phrase summarization for rated aspects of short comments was generated (Lu, Zhai and Sundaresan, 2009). Rated aspects were taken as the various topics of customer opinions and phrases were taken as the detailed sentence pieces regarding different topics. First, aspects were identified by a clustering algorithm. Then the rating of each aspect was predicted and representative phrases of each aspect were extracted according to the predicted ratings of that aspect. Two models, namely, Master-Slave Topic Model and Extended Master-Slave Topic Model, were applied to summarize readers' comments, in which the maximal marginal relevance, the rating and the length of reviews were considered in the comment selection (Ma, Sun, Yuan and Cong, 2012). Topics from news articles and related comments were extracted using the above two topic models, respectively. Then comments were selected considering relations between articles and comments.

These studies aim to provide a concise report to extract critical topics from online reviews. However, what has been neglected is that customer concentrations of different products are not always the same. Given summarized online concerns of each product, it still needs designers to highlight the strength and weakness and make clear comparisons with competitors. In this research, online concerns of series products are analyzed at the same time, which helps to pinpoint shared pros and cons.

2.2 Review selection and recommendation

A number of recent studies focus on how to select a small number of representative reviews, which aim to cover various aspects about customer viewpoints.

Given a collection reviews and related words about a topic, an efficient review subset was selected (Nguyen, Lauw and Tsaparas, 2013). First, the coverage of a review was defined as the number of topic words or names of entities it covered and the efficiency of a sentence was defined similarly in the sentence level. Next, subsets of reviews that consider different levels of efficiency of review sentences were selected. Lappas et al. (Lappas, Crovella and Terzi, 2012) sampled a characteristic subset of reviews, in which sampled subset are expected to follow the sentiment distribution about the original review set. Besides the sentiment distribution, reviews' quality was also reckon for review selection by analyzing a taxonomy tree of product features in (Tian, Xu, Li and Pasi, 2015), in which the quality of each review was estimated by how many sub-features about a product feature covers. Similarly, a few reviews in the fine-grained product aspect level were selected by a probabilistic graphical model (Hai, Cong, Chang, Liu and Cheng, 2014), in which the helpfulness of sentences regarding aspects was estimated according to the online voting of reviews and review sentences with high helpfulness score were selected as the final summary. Other factors were also considered for review selection in (Tsaparas, Ntoulas and Terzi, 2011), which include coverage ratio of attributes, quality of selected sentences, subsets of the entire review corpus, different partitions of original review corpus, multiple levels of similarity and coverage, etc.

To determine which reviews should be recommended, different ranking algorithms are also introduced. A weighted and directed graph was constructed to rank sentences regarding product features (K. Zhang, Narayanan and Choudhary, 2010). First, product features were identified and their occurrence frequencies and relative usages were analyzed. Then, subjective and comparative sentences regarding features were assigned. Accordingly, relations among products were modeled and relative quality of products was mined. A framework was proposed to rank product reviews according to different ranking strategies (Krestel and Dokoohaki, 2011). Different topics in reviews were detected using the Latent Dirichlet Allocation. Also, customer-assigned rating scores were introduced as an indicator about sentimental polarity in the product level. Finally, three ranking strategies were designed to rank reviews with different requirements.

These studies mainly focus on sampling a small helpful opinionated review subset. Still, few of them are applicable to extract critical shared customer concerns from online opinions of multiple products.

2.3 Contrastive viewpoints extraction

The generation of contrastive summarization from large customer data benefits the detection of comparative requirements, which can be utilized to analyze competitors of products or bands. Different relevant studies were reported, and most of which modeled this problem as the extraction of sentences from customer online opinions.

A pattern discovery approach was proposed to discover comparative sentences from textual data (Jindal and Liu, 2006a), in which categories of sequential rules were studied. In their later work, label sequential rules were also utilized to further extract relations of comparative sentences (Jindal and Liu, 2006b). According to these techniques, opinionated entities of customer opinions were extracted from comparative sentences (Ganapathibhotla and Liu, 2008).

To uncover contrastive customer concerns in details, opinionated sentences are initially extracted for the detection of comparative viewpoints in many studies. A two-stage approach was proposed to summarize contrastive opinions (Paul, Zhai and Girju, 2010). Customer viewpoints were initially modeled and extracted accordingly to lexical and syntactic features. Then, pairs of sentences were scored by a comparative LexRank considering the representativeness and the contractiveness. Similarly, a framework was proposed to select pairs of representative yet comparative sentences about a specific feature from competitive products (Jin, Ji and Gu, 2016). The selected subsets of sentences from online opinions of competitive products were compared in different sentimental polarities. The comparison of review pieces was then utilized to help designers analyze competitive products.

Comparative sentences were identified from customer reviews using a two-level Conditional Random Fields (CRFs) model (Xu, Liao, Li and Song, 2011). A novel graphical model was proposed to extract comparative relations between products, in which interdependency relations of products were taken into consideration. An undirected network for relation analysis was constructed (Netzer, Feldman, Goldenberg and Fresko, 2012). A product graph was built according to the co-occurrence of product names. Then, product relations were analyzed and the market structure in the related domain was explored. A product comparison network was investigated to capture comparative opinions (Z. Zhang, Guo and Goes, 2013). Comparative sentiments in social media were exploited. Customer sentiments are represented as a network, in which products were denoted as nodes and directed link indicated a comparative relation between products. According to the network, comparative opinions were investigated regarding the impact on product sales.

Kim et al. (H. D. Kim and Zhai, 2010) proposed a novel approach for contrastive opinion summarization (COS), in which the content similarity with the same sentiment polarity and the contrastive similarity with the opposite polarity were considered. Two greedy algorithms that select representative and contrastive sentences were designed were utilized to select sentence pairs. Then sentence pairs with opposed sentiment polarities were sampled as the final comparative summaries. In this research, these two greedy algorithms, denoted as COS-1 and COS-2, were utilized for the benchmark, which is elaborated in Section 5.

Different approaches for contrastive viewpoints extraction are reported to initialize one-to-one comparisons about competitive products. However, it is arguably to be applied for the comparison of series product groups, which focuses on multiple-to-multiple comparisons of comparative products.

2.4 A brief summary

To sum up, different models were reported to extract valuable information from online reviews. But few studies focus on extracting shared pros and cons of series products and providing a summarization for the comparison on series products. It potentially provides valuable information for product designers to hold a competitive position in the scenario of new product design. Accordingly, in this research, how to obtain shared strength and weakness of different products in a series by analyzing online customer concerns is investigated. It is believed to be beneficial to provide operative actions for competitive intelligence in fierce market and theoretical development of customer requirement management by mining online concerns from the perspective of competitor analysis.

3. Problem Statement

One of the central task of this research is to facilitate designers to digest customer concerns from a large number of online reviews of series product efficiently. This concern information was used to mirror the major shared customer thoughts about series products. It is expected to facilitate designers to understand customer decisions towards products of different series and to make up research studies towards the identification of comparative customer requirement for new product design. To explain this problem clearly, some definitions of terms are clarified with running examples.

Definition 1 Product Feature: A feature of certain product refers to an attribute or component of that product which has been discussed in reviews. For example, the “camera” of smart phone in the following four smart phone review sentences is referred as a product feature.

Example 1: “The pixel of front camera could be higher.”

Example 2: “The details of selfie camera are not clear.”

Example 3: “I can hardly use the front camera.”

Example 4: “The front camera is too easy to hit by accident.”

Definition 2 Feature Aspect: An aspect of a certain feature refers to a sub-attribute or component of that feature which has been detected in the feature related review segments. In another word, these aspects of a given feature can be denoted as subtopics about the feature, which contain detailed information about customer concerns. For example, “camera picture” and “camera operating” are two different aspects of “camera”, which are shown in the Example 1, 2 and Example 3, 4, respectively.

Suppose that a series of N products in the same brand, $P=\{p_1, p_2..., p_N\}$, and M shared features of these products, $F=\{f_1, f_2..., f_M\}$, are reckoned. Specially, for a specific product feature, various aspects might be concerned by customers. Then, sentences that talk are a particular aspect k of one feature f_m can be defined as a_m^k and $A_m=\{a_m^1, a_m^2..., a_m^K\}$ can be regarded as the sentence set that refers all K different aspects of f_m. Additionally, it can be also observed that some feature aspects are discussed across reviews of each product in P and, in this research, such aspects are referred as serial aspects

$C_m = \{c_1^m, c_2^m, ..., c_{|C_m|}^m\}$ and $|C_m| \le K$. Then, the corresponding sentence set of a particular serial aspect $c_\mu^m \in C_m$ can be denoted as $S_{c_\mu^m} = \{s_1^{c_\mu^m}, s_2^{c_\mu^m} ..., s_N^{c_\mu^m}\}$.

Arguably, time factors should be reckoned regarding customer concerns on shared features of these products. Indeed, some studies investigated customer preferences trend by mining online reviews and time series analysis based approaches were utilized (Conrad and Harrison, 2011). In particular, seasonally or monthly averaged sentiment polarities on product features of a particular product are extracted and temporal sentimental changes in a relative long period, say five years, were analyzed. However, note that the objective of this research is to provide a concise summarization about customer concerns for a series of products, which often involve only a small number of products, for instance five products in a series. Hence, time series analysis might not be a proper approach for this problem.

Typically, a review sentence set of series products regarding a feature is taken as input and a representative sentence subset T is sampled. This list of sentences regarding a serial aspect c_μ^m is denoted as $T_{c_\mu^m} = \{t_1^{c_\mu^m}, t_2^{c_\mu^m} ..., t_N^{c_\mu^m}\}$, where $t_n^{c_\mu^m} \subseteq s_n^{c_\mu^m}$ indicates a group of sentences that are sampled from reviews of product p_n. T is expected to highlight and compare customer concerns about series products in new product design. To clarify what characteristics about sampled sentences would be, fundamental concepts are introduced.

Information Coverage, reveals the information that is covered by a subset of sentences T from a given sentence set S, which is denoted as *Coverage* (T, S). The value of information coverage is high when the subset of sentences covers a major part of content of S.

Information Diversity reveals the non-overlap information that is covered by a subset of sentences T, which is denoted as *Diversity* (T). The value of information diversity is high when different messages are covered in T.

Information Representativeness, indicates the representatives about a subset of sentences T that is compared with a given sentence set S, which is denoted as $Representative(T, S)$. A high information representative subset of sentences should intuitively have both high information coverage and information diversity. Thus, given a sentence set S and its subset T, the problem of sentence sampling representative sentences T^* from S can be modeled as an optimization problem,

$$T^* = \arg\max_T \big(Representativeness(T, S)\big)$$

$$= \arg\max_T \big((1-\alpha)Coverage(T, S) + \alpha Diversity(T)\big) \quad (1)$$

α is a coupling parameter to trade-off *Coverage* and *Diversity* of sampled sentences. Thus, the major concern of this study becomes how to sample a subset of sentences T from reviews of series products S that maximizes the information representativeness.

Note that, in this study, strength of opinions are neglected (Wilson, Wiebe and Hwa, 2004), i.e., whether the sentiment polarity is strongly (or weakly) negative (or positive). In addition, the helpfulness level of each review is not considered, which means all customer reviews are regarded to be equally important. Admittedly, these subtle details might be valuable for product designers. However, the focus of this research is on comparisons of multiple comparable products. Actually, in our previous study (Liu et al., 2013), the helpfulness of online reviews was initially defined from the perspective of product designers. In that study, four categories of features were extracted from online reviews for the prediction about the helpfulness, such as the information divergence of sentiment sentences, the information strength of sentiment sentences, etc. In the future, models on the helpfulness prediction of online reviews will be considered to be embedded with the problem about series product comparisons, which might help to polish the study in this research.

4. Methodology

In this section, technical details about the proposed approach will be explained. For the sake of reference, symbols are listed in Table 2.

Table 2. Notations and definitions

Symbol	Definition
P	The series product set, $P=\{p_i\}_N$
F	The product feature set of the series products, $F=\{f_j\}_M$
R_m	The reviews of all series products regarding feature f_m, $R_m=\{r_\phi\}_{\lvert R_m\rvert}$
A_m	The aspect set of a product feature f_m, $A_m=\{a_k^m\}_{k=1}^{\lvert A_m\rvert}$
C_m	The serial aspect set of a product feature f_m, $C_m=\{c_\mu^m\}_{\mu=1}^{\lvert C_m\rvert}(C_m \subseteq A_m)$
$S_{c_\mu^m}$	The related sentence set regarding a serial aspect c_μ^m, $S_{c_\mu^m}=\{s_i^{c_\mu^m}\}_{i=1}^N$
$T_{c_\mu^m}$	The subset of sampled sentences extracted from $S_{c_\mu^m}$, $T_{c_\mu^m}=\{t_i^{c_\mu^m}\}_{i=1}^N(T_{c_\mu^m} \subseteq S_{c_\mu^m})$
H	The limited count of sampling sentences of set $T_{c_\mu^m}$
h	The minimum number of sampling sentences of subset $t_i^{c_\mu^m} \in T_{c_\mu^m}$, where $N*h \le H$

4.1 Framework overview

In this research, a framework named **S**ampling **R**epresentative **S**entences of **S**erial products (**SRSS**) is proposed. This framework is to sample a subset of representative sentences for analyzing online customer concerns of series products in the same brand and competitors. It consists three phases: (1) **SRSS-I:** Extract Product Features; (2) **SRSS-II**: Identify Feature Aspects; (3) **SRSS-III**: Sample Representative Sentences. The overview of this framework is presented in Figure 1.

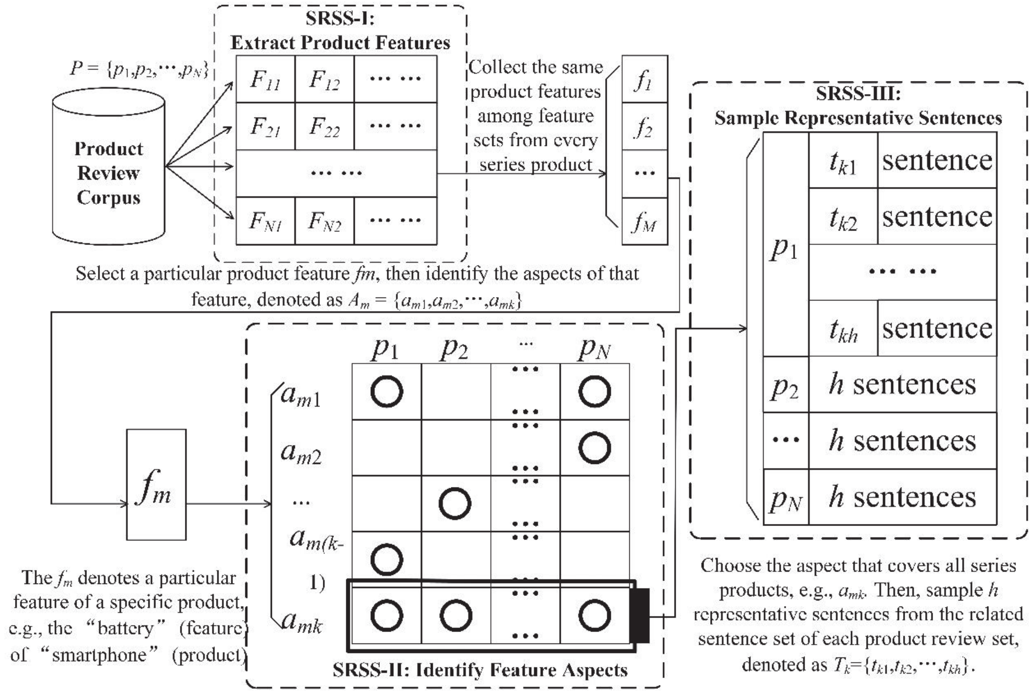


Figure 1 Overview of the framework SRSS

4.2 SRSS-I: Extract product features

Given online reviews in one domain, the first task is to extract product features and analyze the corresponding sentimental polarities. In this research, with the help of Pros and Cons reviews, such as reviews in Cnet.com and Epinions.com, a simple yet effective method is employed. Similar approaches for product feature identification and sentiment analysis are also reported in (Jin, Ji and Kwong, 2016; S. M. Kim and Hovy, 2006; Yu, Zha, Wang and Chua, 2011).

First, a typical review of the Samsung Galaxy S III GT-I9300 is presented in Epinions.com. As seen from this review, the strength and the weakness are listed clearly in the corresponding parts. Some other similar examples can be easily found in Pros and Cons reviews. But note that, in these reviews, most frequently referred nouns or noun phrases are product features. Hence, accordingly, in the beginning, POS tagging is conducted and frequently referred nouns or noun phrases are regarded as product feature. The assumption that frequently referred nouns or noun phrases are assumed to be product features are often utilized in relevant research fields on mining product online reviews, such as studies in (Liu, 2010, Conrad and Harrison, 2011). Then, analyzing results from Pros and Cons reviews help to extract product features from customer online reviews in a general format, such as reviews in Amazon.com. Next, for example, "battery", "camera", "screen", "media" and "application" can be extracted from reviews of smart phones. Then, sentences regarding product features can be identified from customer reviews.

Accordingly, with the help of review corpus in (Pang and Lee, 2004), a binary classifier can be built to classify whether a sentence is subjective or objective. Next, with the opinionate information in Pros and Cons reviews, another binary classification approach is utilized to classify whether a subjective sentence is positive or negative. In this approach, each sentence is represented as a bag of sentimental terms, which is defined in MPQA (Wilson, Wiebe and Hoffmann, 2005), and review sentences in Pros and Cons reviews are utilized as training corpus.

Note that an objective sentence might also contain valuable information about customer concerns. Similar to assumptions in other relevant studies (H. D. Kim and Zhai, 2010), in this study, only positive and negative sentences are taken as the major concerns for sampling representative sentences. Accordingly, sentences in the negative and positive polarities of specific features are collected into different sets.

4.3 SRSS-II: Identify feature aspects

The second task in SRSS is to identify different aspects of product features from online reviews.

Intuitively, since that the same feature aspect may be discussed in sentence pieces with similar details, they can be identified by grouping sentences with similar topics. In this research, one straightforward approach is utilized to cluster review sentences according to the content similarity. Specifically, first, given review sentences R_m of all series products regarding a feature f_m, the similarity matrix of sentences U is constructed, where $u_{ij} = Similarity(r_i, r_j)(r_i, r_j \in R)$. The measure function *Similarity* will be elaborated in Section 4.4. Next, feature-related sentences are clustered according to U. In this research, a self-adaptation clustering method in (Frey and Dueck, 2007) is employed and the clustering results are taken as different groups of relevant sentences regarding different aspects A.

Note that some aspects of features might be observed across different products of that series. In this research, these aspects are defined as serial aspects, denoted as $\boldsymbol{C}$. Serial aspects echo that similar topics of a feature which are frequently discussed across different products in this series, and these topics might become consistent customer concerns about the whole series products. Take two features of a smart phone for example, "battery overheating" of "battery" and "system-hogging app" of "application" are continually observed in customer reviews of each product in that series. Accordingly, "overheating" is a serial aspect of "battery", as well as "system-hogging" of "application". Thus, review sentences regarding a serial aspect $c \in C$ can be grouped into the same set S. Next, a few representative sentences are expected to be sampled from S.

4.4 SRSS-III: Sample representative sentences

In this section, how to sample representative sentences from the sentence set regarding a serial aspect will be discussed. Recall that sampled sentences are expected to balance two criterions, i.e., *information coverage* and *information diversity*. Hence, according to Eq. (1), how to model the *information coverage* and *information diversity* of sampling results, as well as how to sample a subset of representative sentences will be discussed.

Information Coverage: As aforementioned, sentence samplings should cover general details of aspects, which means that the sampled sentence set T is expected to cover the main content of the original sentence set S. It can be denoted as,

$$Coverage(T,S) = \frac{1}{|T| \times |S|} \sum_{t \in T, s \in S} Similarity(t,s) \tag{2}$$

Information Diversity: Sentence sampling should cover different topics of aspects, which means each sentence in T is expected to be dissimilar to others. It can be denoted as,

$$Diversity(T) = \frac{1}{|T| \times |T|} \sum_{t,t' \in T} Distance(t,t') \tag{3}$$

Now, the problem becomes to the evaluation about the similarity and the distance between two sentences.

4.5 Similarity models

According to Section 4.4, one of the central tasks is how to define the similarity between review sentences. Different approaches are reported to model the sentence similarity in the research field. However, some models are quite complex for product designers. To smooth the difficulty in the implementation and focus on the application of the proposed framework, simple but effective models that estimate the similarity between two sentences X and Y are employed, which is a semantic matching similarity model $Similarity(X, Y)_{semantic}$ and a cosine similarity model $Similarity(X, Y)_{CS}$. Note that the semantic matching model was initially utilized in (Kim and Zhai, 2009) to evaluate the similarity between two sentences, while the cosine similarity model is a widely applied in the research area of information retrieval for document similarity (Alguliev, Aliguliyev and Isazade, 2013).

4.5.1 Semantic matching model

In Eq. (4), the sentence similarity between two sentences X and Y is defined, which considers semantic matching between terms,

$$Similarity(X,Y)_{semantic} = \frac{\sum_{x\in X}\max_{y'\in y}\phi(x,y') + \sum_{y\in Y}\max_{x'\in X}\phi(x',y)}{|X|+|Y|} \tag{4}$$

$\phi(u, v)\in[0, 1]$ is a term similarity function and $|X|$ and $|Y|$ are the total counts of words in sentences X and Y. A similar definition about the sentences similarity can be found in (H. D. Kim and Zhai, 2010). Actually, many other similarity models, e.g., vector space model (VSM), can also be used, they will be tested in the future. Depending on how ϕ is defined, different variations can be obtained. In this research, two natural variants are investigated, where related similarity models are denoted as $Similarity(X, Y)_{WO}$ and $Similarity(X, Y)_{WN}$.

Word Overlap $Similarity(X, Y)_{WO}$: $\phi_{\mathrm{WO}}(x, y) = 1$ if $x = y$, and $\phi_{\mathrm{WO}}(x, y) = 0$ otherwise. It is naturally the *Jaccard* similarity function that considers word overlap.

Semantic Word Matching $Similarity(X, Y)_{WN}$: $\phi_{\mathrm{SWM}}(x, y) = 1$ if $x = y$, and otherwise $\phi_{\mathrm{SWM}}(x, y) = sim(x, y)$. $sim(x, y)$ refers to the semantic term similarity, which estimates the similarity of semantic content between word x and word y. In this research, the normalized value $WordNet(x, y)\in[0, 1]$ is employed. $WordNet(x, y)$ evaluates shortest path distance about the conceptual relations of two terms that is defined in WordNet (Fellbaum and Miller, 1998). Then, $sim(x, y)$ is set to be the 1-$WordNet(x, y)$, which measures how similar the two terms are.

4.5.2 Cosine similarity model

In Eq. (5), the sentence similarity between two sentences X and Y is measured, which considers the co-occurrence between word frequency vectors of X and Y over the related entire sentence set c_η,

$$Similarity(X,Y)_{CS} = \frac{\sum_{e\in X\cap Y, X\in S, Y\in S}(W(e,S))^2}{\sqrt{\sum_{x\in X}(W(x,S))^2}\sqrt{\sum_{y\in Y}(W(y,S))^2} \tag{5}$$

$W(e, S)$ indicates the frequency of the word e in sentence set S.

In short, three similarity models, namely, $Similarity(X, Y)_{WO}$, $Similarity(X, Y)_{WN}$ and $Similarity(X, Y)_{CS}$, are defined and they will be employed to evaluate the similarity between two sentences. Accordingly, the distance between two sentences X and Y can be defined as,

$$Distance(X,Y) = 1 - Similarity(X,Y) \tag{6}$$

4.6 An optimization problem

Generally, sampled review sentences are expected to have high information coverage and high information diversity. Thus, the problem of representative sentence sampling in Eq. (1) can be denoted as,

$$T^* = \arg\max_T ((1-\alpha)\frac{\sum_{t\in T, s\in S} Similarity(t,s)}{|T|\times|S|} + \alpha\frac{\sum_{t,t'\in T} Distance(t,t')}{|T|\times|T|}) \tag{7}$$

$|T|$ and $|S|$ were the numbers of T and S, respectively.

Actually, different methods are available to analyze such optimization problem, such as the greedy-based approach (McDonald, 2007), the clustering approach like fast approximate spectral clustering (Wei et al., 2016), particle swarm optimization (Tao et al., 2008), etc. In this research, the method in (McDonald, 2007) is adopted to analyze this optimization problem.

5. Experiment study and discussion

5.1 Experimental setup

In this section, a case study is presented to demonstrate how the proposed approach can be utilized to sample the representative sentences from series product reviews by product designers. 21,952 pros and cons reviews of intelligent mobile phones were collected from Cnet.com. They were utilized as a training corpus for product feature extraction and sentiment polarity identification. 10,815 reviews of popular series mobile phones of two brands were obtained from Amazon.com to verify the availability of the proposed approach. The numbers of reviews about these products are presented in Table 3. In this case study, online reviews of three series products in each brand are analyzed. For data privacy, the names of these two brands are represented as Brand1 and Brand2.

Table 3. # of reviews about series three products in Brand 1 and Brand 2

# of reviews	P1	P2	P3	Total
Brand 1	1,086	1,872	2,147	5,108
Brand 2	429	1,275	3,993	5,697

In this experiment, to evaluate the effectiveness of the proposed approach, review sentences that refer to battery and camera in Brand 1 and Brand 2 are explored. With the help of pros and cons reviews in Cnet.com, product features are extracted and sentiment polarities are analyzed by using approaches in SRSS-I and SRSS-II. Then, opinionated sentences are filtered from original customer reviews. Some statistics regarding the number of sentences are show in Table 4.

Table 4. # of sentences in positive and negative polarities

Sentiment polarity	Brand 1		Brand 2	
	Battery	Camera	Battery	Camera
# of positive sentences	211	332	347	618
# of negative sentences	592	263	749	257

5.2 Evaluation metrics

Conventionally, it is difficult to acquire training samples for this problem that is built from a big volume of product online reviews. Even for a small exemplary data set, it is still difficult to select some representative sentences manually. It induces that some widely utilized metrics, such as precision and recall, potentially fail to be applied to evaluate the performance of the proposed approach.

Many relevant studies employed the information coverage only to evaluate the performance about the task of summarization (Lappas, Crovella and Terzi, 2012; Nguyen, Lauw and Tsaparas, 2013). Note that in (Wei et al., 2016), information redundancy and information diversity were reckoned in an integrated framework for the summarization of categorized community answers. Motivated by the study in (Wei et al., 2016), hence, in this research, three evaluation metrics are employed, i.e., information redundancy, information coverage and information centralization.

a) Information redundancy (IRD), evaluates to what extent sampled sentences are similar to each other. To avoid the content overlap of sampled sentences, the result is expected to have a low IRD value. IRD can be estimated by evaluating the similarity among different sentences in *T*, and given a subset *T*, it is formulated as,

$$Redundacy(T) = \frac{1}{|T| \times |T|} \sum_{t,t' \in T} Simialrity(t,t') \quad (8)$$

b) Information coverage (ICR), evaluates to what extent sampled sentences are similar to sentences in the original review set. To obtain representative sentences, the sampled sentence set is expected to have a high ICR value. ICR can be estimated by evaluating the similarity between *T* and *S*, and given a subset *T* and its original set *S*, it can be formulated as,

$$Coverage(T,S) = \frac{1}{|T| \times |S|} \sum_{t \in T, s \in S} Similarity(t,s) \quad (9)$$

c) Information centralization (ICT), evaluates to what extent sampled sentences cover the content of original reviews. This metric aims to figure out how sampled sentences are affected by high-frequent words. ICT will increase if sampled sentences cover lots of high-frequent words. Given a frequent content word set *V* of the original reviews, the ICT of a given sentence subset *T* can be formulated as,

$$Centralization(T,V) = \frac{1}{|D(V,T)|} \sum_{v \in V} W(v,V) \times D(v,T) \quad (10)$$

$W(v, V)$ is the frequency of word v, $D(v, T)$ is an indicator variable that $D(v, T) = 1$ if T contains word v and $D(v, T) = 0$ otherwise, as well as $|D(V, T)|$ is the count of frequent words that T covers. To cover content words with different frequency, the sampled sentence set is hence expected to have a low ICT value.

In the following experiments, to evaluate the performance of the proposed approach regarding different metrics, $\text{Similarity}(X, Y)_{WN}$ is employed to evaluate the word similarity in different pairs of sentences.

5.3 Benchmark approaches

For the benchmark with similar algorithms, two simple approaches of review sentence selection, the RANDOM Sampling and TOPLEN Sampling, are introduced as basic benchmark approaches, which were also utilized as in (Tsaparas, Ntoulas and Terzi, 2011). Additionally, two greedy algorithms for the one-to-one product comparison analysis regarding the contrastive opinion summarization (COS) was proposed in (H. D. Kim and Zhai, 2010) and they are also employed to benchmark the proposed approach. Notice that two different selection strategies will be tested in this study and they are denoted as COS-1 and COS-2. Accordingly, given a customer review corpus of N products in the same series, the proposed method is compared with all these four benchmark approaches.

a) RANDOM Sampling (RD): h sentences are selected randomly. In this experiment, 3000 runs will be performed for the sampling and average values of different evaluation metrics are reported.

b) TOPLEN Sampling (TL): Sort all sentences according to the length and select h longest ones as the sample subset. This is meant to serve as a basic baseline since that longer review sentences are usually expected to contain valuable information.

c) COS-1 Sampling (COS-1): Given a review sentence set R_m regarding a particular feature f_m of a product, q sentences in negative polarity as well as q sentences in positive polarity are selected from each set of product reviews, denoted as Q_1 and Q_2, respectively. Q_1 and Q_2 are expected to have high similarity to the original set R_m. Then one sentence from Q_1 and one sentence from Q_2 are selected and combined as a pair, respectively. Sentences in each pair are expected to have high similarity regardless the sentiment polarity. To sample sentence pairs across N series products, $2*q*N$ sentences are selected as the result.

d) COS-2 Sampling (COS-2): Given a review sentence set R_m regarding a particular feature f_m of a product, one negative sentence and one positive sentence are collected and combined, respectively, which is denoted as Q. Sentences in each pair of Q are expected to have high similarity regardless the sentiment polarity. Then q pairs in Q are selected as the sampling results, which are expected to have the highest similarity to the original set R_m. Similar to the approach of COS-1, to sample sentence pairs across N series products, $2*q*N$ sentences are selected as the result.

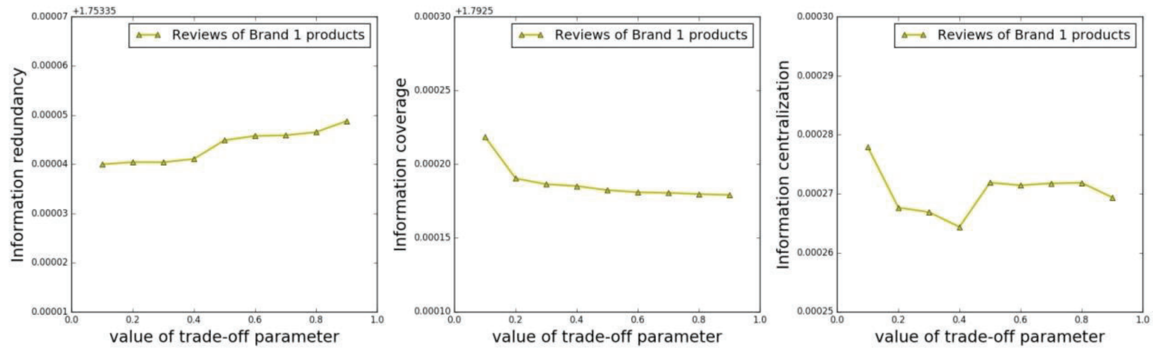
Note that, three products in each series are selected and it makes N equal to 3. To evaluate the performance of the proposed approaches and different approaches, the number of sentences that are sampled from the original review set is equal. Then, q, in COS-1 and COS-2, is set to be 2 to 8, which means that 12, 18, 24, 30, 36, 42, 48 sentences are sampled, respectively, since that these two approaches totally select $2*q*N$ sentences. Then, for different categories of experiments in Section 5.4, an averaged result will be reported according to all these seven categories of parameter settings.

5.4 Evaluation results and analysis

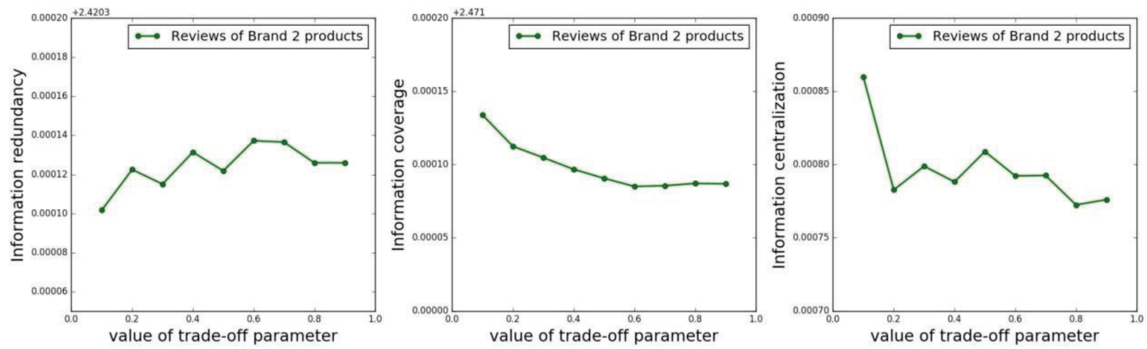
5.4.1 On the sensitivity of parameter α

Different experiments that evaluate the impacts of α in Eq. (6) are conducted. α is a coupling parameter that balances the *coverage* and the *diversity* of sampled sentences.

Different values of α are tested and $Similarity(X, Y)_{WO}$ is utilized. In Figure 2, the approach of SRSS is compared by using reviews of both Brand 1 and Brand 2.



(a) Reviews in Brand 1



(b) Reviews in Brand 2

Figure 2 Performance comparisons regarding different α

As observed form Figure 2(a) and Figure 2(b), higher IRD and lower ICR are observed if α is larger. However, a lower value of ICT is observed if α is moderate. One of potential reason is that with an increasing weight of *coverage* in Eq. (1), of which weight is 1 - α, more sentences that are similar with the original sentence set tends to be sampled, while the diversity of sampled sentences becomes smaller. As a result, the value of ICR is achieved higher, while sampled sentences are semantically similar, which leads the value of ICT get higher.

α can be set as various values according to which metrics matter the most. In this experiment, α is chosen to be 0.6. As seen from this figure, it is generally reported as an optima parameter for sampling representative sentences, where a lower value of IRD and ICT and a higher value of ICR are balanced. In the following experiments, α is set to 0.6, in which better performance is observed in terms of all three metrics.

5.4.2 Performance benchmarking with different approaches

Different benchmark approaches are compared in terms of the three evaluation metrics and the final results are reported in Table 5, in which the bold number indicates that it is the best desired performance among other benchmarks except RD and TL sampling approaches. Both positive and negative sentences are investigated in different experiments and results are the summation of the average value considering all sampling sentences.

Table 5. Performance benchmarking with different approaches

	Brand1						Brand2					
	Battery			Camera			Battery			Camera		
	IRD	ICR	ICT	IRD	ICR	ICT	IRD	ICR	ICT	IRD	ICR	ICT
RD	2.253	2.316	7.59E-04	2.255	2.320	9.76E-04	2.255	2.320	9.76E-04	2.255	2.318	1.13E-03
TL	2.247	2.315	8.66E-04	2.254	2.319	1.08E-03	2.254	2.319	1.08E-03	2.250	2.319	4.44E-03
Word Overlap(WO)												
SRSS+*WO*	**1.689**	1.730	**2.30E-04**	**1.817**	1.856	**3.26E-04**	**2.359**	2.403	7.01E-04	2.482	2.540	**8.99E-04**
COS1+*WO*	2.478	2.547	3.90E-04	2.478	2.547	4.62E-04	2.481	2.551	**6.86E-04**	**2.481**	2.550	1.19E-03
COS2+*WO*	2.480	**2.550**	5.85E-04	2.482	**2.550**	4.61E-04	2.482	**2.552**	8.02E-04	2.483	**2.552**	2.40E-03
Sematic Word Matching with WordNet(WN)												
SRSS+*WN*	2.481	**2.558**	**3.08E-04**	2.482	2.525	4.69E-04	**2.479**	**2.552**	**6.03E-04**	**2.481**	**2.554**	**8.53E-04**
COS1+*WN*	2.479	2.549	5.32E-04	**2.481**	2.547	4.59E-04	2.481	2.551	9.90E-04	2.482	2.551	1.21E-03
COS2+*WN*	**2.470**	2.547	5.37E-04	2.483	**2.548**	**4.32E-04**	2.481	2.550	7.93E-04	2.483	2.552	1.29E-03
Cosine(CS)												
SRSS+*CS*	**1.866**	1.906	**2.75E-04**	**1.712**	1.734	**2.95E-04**	**2.480**	2.527	7.89E-04	**2.479**	2.528	**9.75E-04**
COS1 +*CS*	2.480	2.548	3.75E-04	2.481	**2.547**	7.37E-04	2.481	**2.550**	**7.21E-04**	2.480	2.550	1.68E-03
COS2 +*CS*	2.478	**2.549**	5.56E-04	2.483	2.550	3.54E-04	2.481	2.551	7.63E-04	2.483	**2.552**	1.18E-03

Note that, IRD and ICT are the lower the better, while the value of ICR is the higher the better. As seen from Table 5, a lower IRD value and a lower ICT value are obtained by the approach of **SRSS**. It can be claimed that review sentences containing various details are sampled and they cover words in different frequency instead of high-frequent words only. These results indicated that, comparing to other benchmark approaches, for purpose of sampling review sentences across several series products, **SRSS** is capable to sample the richly detailed sentences that reveal the relations of customer concerns about different series products. In addition, it can also be found that, compared with benchmark approaches, moderate ICR values were obtained by the approach of **SRSS**. It reveals that sampling sentences are similar to the rest set of non-sampled sentences in a certain degree. Perhaps the reason behind is that, for a particular feature, a large proportion of customers may focus on similar topics and the word-overlap is pretty high. If these similar customer concerns are sampled, an obviously high value of coverage while low value of diversity will be gained. To reveal different topics of customer concerns and avoid sentences that reveal the same topics are sampled in SRSS, sentences that contains similar words are seldom sampled, which leads to a moderate level of ICR.

On the contrary, the approach of **COS-1** mainly considers the main content of the original sentence set, which leads the ICR is high. Also, the approach of **COS-2** focuses on the similarity between sentences in opposite sentiment polarities and samples the most similar sentence pairs as results. They both aim to reveal main content of entire sentence set, which leads to the values of ICR by both approaches are high. In short, **COS-1** and **COS-2** have strong ability to sample sentence subsets with a higher ICR by utilizing $Similarity(X, Y)_{WO}$ and $Similarity(X, Y)_{CS}$. Even in experiments utilizing $Similarity(X, Y)_{WN}$, **COS-1** and **COS-2** just obtained moderate lower ICR than **SRSS**. However, the proposed approach of **SRSS** performs better in sampling sentences that not only cover the main content of the original sentence set but also contain various detailed information, as well as avoid containing too much high-frequent words.

5.4.3 Performance comparisons on different sentiment polarities

To estimate the impact of sentiment polarity affecting the sampling results, two categories of experiments that consider positive sentences only and negative sentences only are conducted. The final results are showed in Table 6, in which the bold number indicates that it is the best desired performance among other benchmarks except the approach of random sampling.

Table 6. Performance comparisons on sentiment polarities

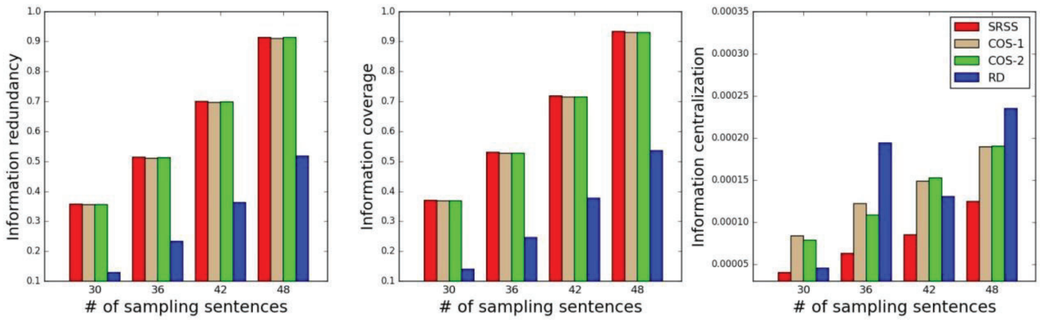
	Brand1						Brand2					
	Battery			Camera			Battery			Camera		
	IRD	ICR	ICT	IRD	ICR	ICT	IRD	ICR	ICT	IRD	ICR	ICT
Word Overlap(WO)												
WO+Pos	**2.481**	2.542	8.44E-04	**2.319**	2.356	**5.05E-04**	**2.482**	2.532	5.04E-04	2.483	**2.553**	3.26E-04
WO+Neg	2.482	**2.546**	**8.04E-04**	2.370	**2.407**	5.83E-04	2.483	**2.540**	**1.57E-03**	**1.767**	1.799	**3.23E-04**
Sematic Word Matching with WordNet(WN)												
WN+Pos	**2.474**	2.511	1.05E-03	2.481	**2.530**	5.85E-04	**2.481**	2.532	5.62E-04	**2.388**	2.335	4.42E-04
WN+Neg	2.482	**2.537**	**5.30E-04**	**2.447**	2.482	**4.82E-04**	2.482	**2.538**	**2.45E-03**	2.408	**2.663**	**4.00E-04**
Cosine(CS)												
CS+Pos	**1.524**	1.531	7.18E-04	2.428	**2.456**	**5.20E-04**	**1.725**	1.740	4.84E-04	**1.799**	1.819	1.13E-03
CS+Neg	2.480	**2.526**	**6.01E-04**	**2.415**	2.445	6.26E-04	2.481	**2.587**	**2.83E-03**	2.479	**2.528**	**6.28E-04**

As seen from Table 6, SRSS achieves better in ICR and ICT using negative reviews, while better in IRD with positive reviews. One of the possible reasons might be that, in these reviews, customers might discuss various topics about what they favor about these products, while flaws of products or something they dislike tend to be more focused. In addition, customers prefer to use simple words to indicate why they like a particular feature, while they might use affluent words to describe problems or flaws in details, which leads to that the word overlaps in negative opinionated texts tend to be low.

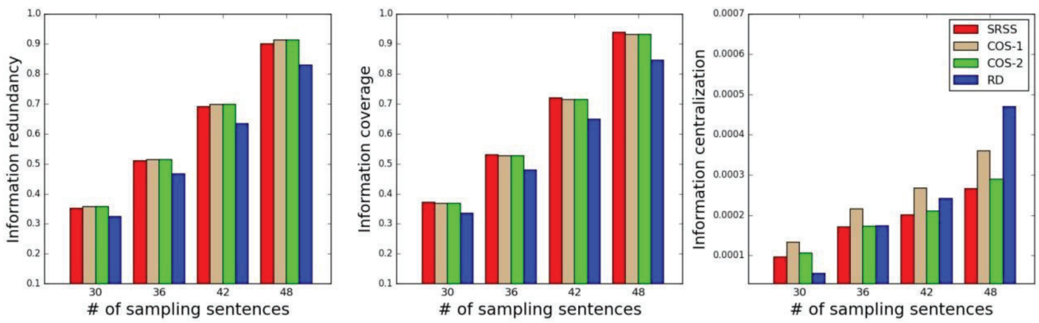
5.4.4 Performance comparisons on different number of sample sentences

To evaluate the performance regarding the number of sample sentences, experiments were conducted using reviews of "battery" and similarity measure ϕ_{WO} and the number of sample sentences is set to 30, 36, 42 and 48. The final evaluation results are presented in Figure 3.

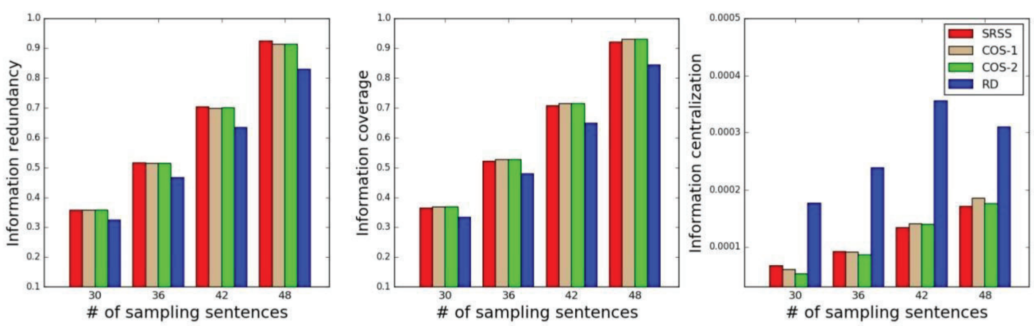
As seen from Figure 3, with the raising about the number of sampled sentences, increasing ICR, IRD and ICT are observed. It can be inferred that SRSS achieves a little lower IRD than COS-1 and COS-2, while ICR is a moderate higher than them. Also, a much lower ICT is reported than other three approaches. It is also observed that, by using the approach of Random Sampling, a much lower IRD value is reported than other approaches. One of the probable reason might be that, compared with the approach of Random, sentences sampled by other approaches have a certain level of word-overlap. As discussed in the previous experiments, for a particular feature, if central topics of consumers are sampled, a higher degree of IRD will be obtained due to the word overlapping problem. However, to balance the coverage and the diversity, sentences that discuss similar topics with various subtopics are sampled, which meet the requirement about the high degree of both information coverage and information diversity. It induces that sampled sentences have a certain level of information redundancy.



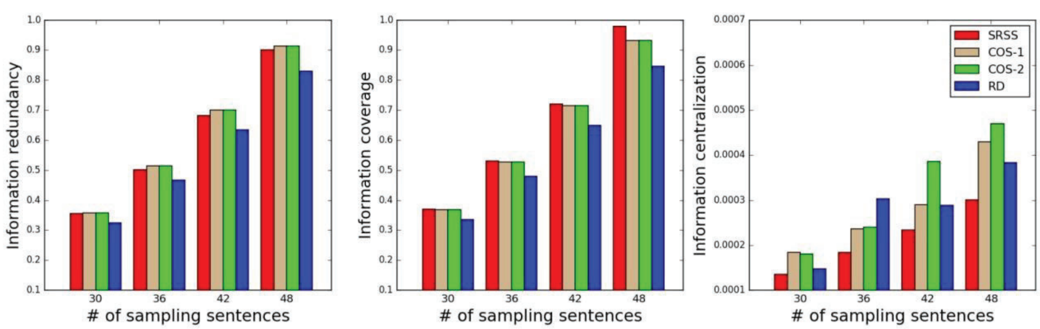
(a) The battery of series mobile phones in Brand 1



(b) The battery of series mobile phones in Brand 2



(c) The camera of series mobile phones in Brand 1



(d) The camera of series mobile phones in Brand 2

Figure 3. Performance comparison regarding battery and camera in two brands

5.5 Case study

To show how the proposed approach can be utilized by product designers, 1,078 reviews of the battery in Brand 1 and Brand 2 are utilized as an illustrative example.

Now, suppose phone designers care about battery opinions only. According to the proposed framework in Section 4, product features are extracted and serial aspects of features are identified with pros and cons reviews. To evaluate the effectiveness of the sampling results, in this experiment, opinionated sentences in negative polarity are taken into consideration only. Accordingly, 592 negative sentences related battery are obtained. Then, a few representative sentences are expected be sampled. The exemplary result is presented in Figure 4. As seen from Figure 4, a few representative sentences are sampled from each series products in Brand 1, in which customer concerns regarding different products in this series are briefly presented. Specifically, the battery "charger" of different products is frequently discussed by their customers.

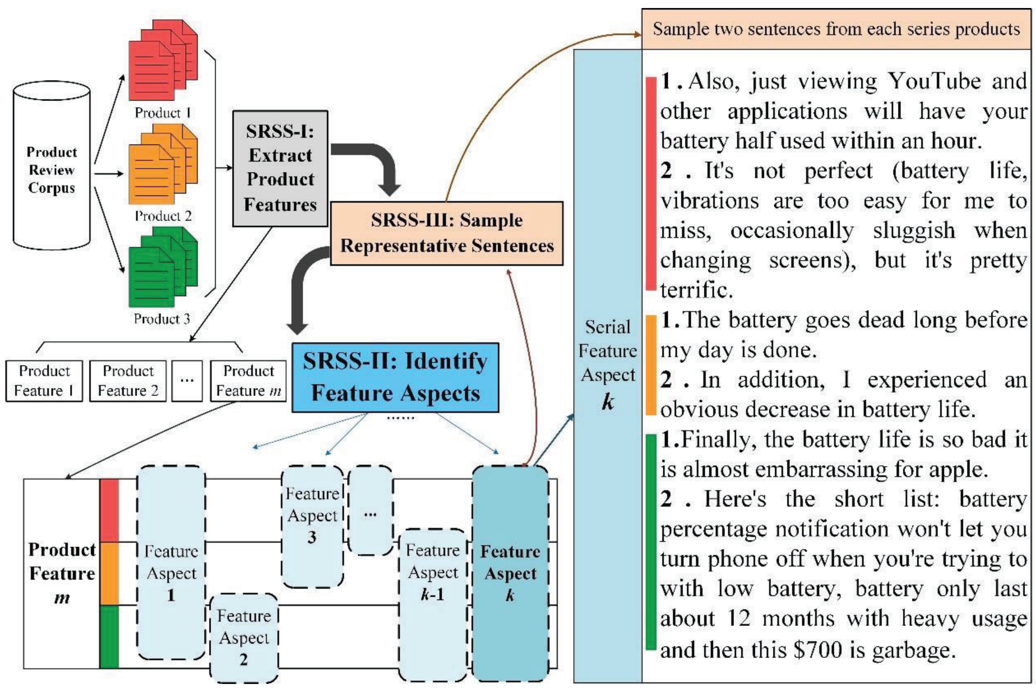


Figure 4. An example of sentence sampling regarding battery in series products of Brand 1

To demonstrate the variance of sampled sentences in **SRSS**, sampled sentences regarding two aspects of "Battery" of Brand 1 are listed in Table 7. As seen from this table, six sentences are sampled from two serial aspects, respectively. It can be observed that c_1 is related to the battery life, while c_2 is related to the charge function of battery. In general, sampled sentences contain detailed information about customer concerns in different conditions, e.g., battery life may run out quickly when using some applications, such as YouTube.

In Figure 5, an example is illustrated to show how the proposed approach benefits product designers for the comparison of series products. In this example, customer online reviews of series products in two brands are analyzed. To make cross comparisons with series products, sentences of customer concerns in two brands are consolidated together and the final representative sentences help to enlighten product designers about strengths and weaknesses of series products in competitor brands, respectively. Compared with the example in Figure 4, series product comparisons in competitor brands are conducted effectively by analyzing the voices of online customers.

Table 7. Sampled sentences regarding two aspects of “Battery” in Brand 1

Aspect	Series product	Sentence #	Sampling sentences
c_1	P1	1	Also, just viewing YouTube and other applications will have your battery half used within an hour.
		2	It’s not perfect (battery life, vibrations are too easy for me to miss, occasionally sluggish when changing screens), but it’s pretty terrific
	P2	1	The battery goes dead long before my day is done.
		2	In addition, I experienced an obvious decrease in battery life.
	P3	1	Finally, the battery life is so bad it is almost embarrassing for apple.
		2	Here's the short list: battery percentage notification won't let you turn phone off when you're trying to with low battery, battery only last about 12 months with heavy usage and then this $700 is garbage.
c_2	P1	1	It did not come with a charger; thus, the 4 star instead of 5.
		2	Charger is too easy to break down.
	P2	1	The phone is ok but the battery needs to be charged daily.
		2	Will have it plugged in for an hour or more and it will still show the battery in the red.
	P3	1	Works well except for the battery which doesn't hold a charge well.
		2	Will have it plugged in for an hour or more and it will still show the battery in the red.

6. Conclusions

In this research, how to identify shared comparative customer concerns regarding different products in a series and recognize the strength and weakness of products in competitive series is studied. Particularly, to sample a subset of representative sentences from online reviews of series products in the same brand and competitors is conducted, which helps to enlighten designers in understanding the strength and weakness about the series for competitive intelligence. A three-phase framework is proposed and an optimization problem is formulated, in which the information coverage and the information diversity are expected to be maximized. Moreover, categories of comparative experiments were conducted on a large volume of real reviews of Amazon.com and how the proposed approach facilitates product designers is presented as an illustrative case study. It demonstrates the effectiveness of the proposed approach.

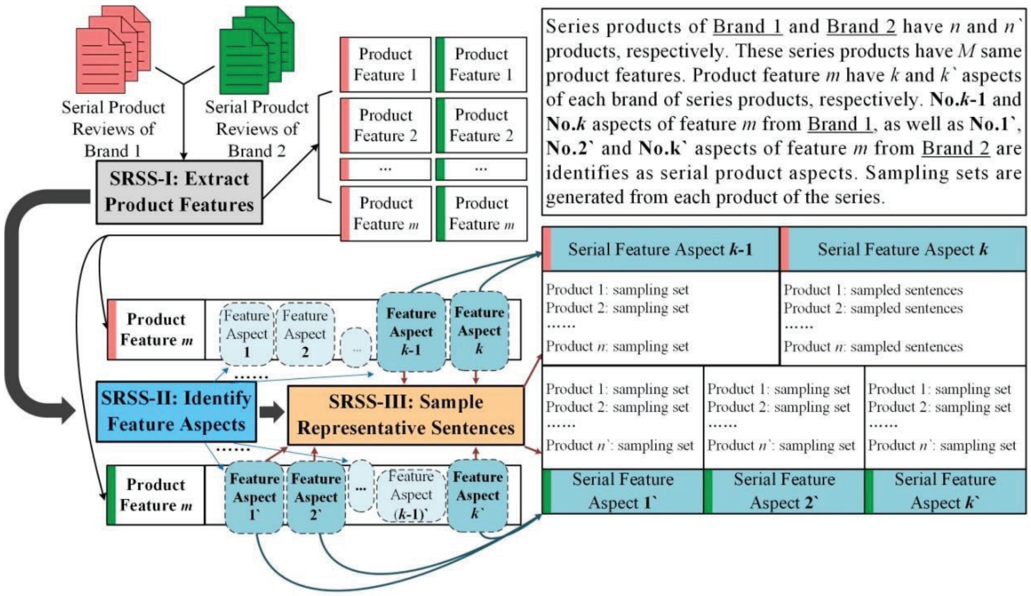


Figure 5. An example of series product analysis of two brand competitors

A visual research gap about multiple-to-multiple comparisons of series products is neglected by many existing studies. Theoretically, this study spotlights the significance of comparative comparisons on series products in customer requirement management via the identification of informative and critical opinions from a large volume of textual online opinions. On the other hand, in practice, provided a large volume of online opinions of series products, this study assists designers to identify the shared strength and weakness about series products for competitive intelligence when they are conceiving new products.

As explained in the previous section, one of the limitation of this research is that the utility of each review for product designers is neglected. In the future, the helpfulness or the utility from the perspective of product designers will be considered for representative sentence sampling and comparisons. Also, note that, in this research, a greedy algorithm is employed to analyze the proposed optimization model. However, a convex optimization model might be more persuading which helps to obtain a global optimization value. Hence, in the future, whether a reasonable convex optimization problem can be built will be examined. Besides, some other potential valuable research studies, such as how to make comparisons of the proposed approach with different similarity functions according to the sentence alignment, can be extended.

Acknowledgement
The work described in this paper was supported by a grants from the National Nature Science Foundation of China (project no. NSFC 71701019/G0114) and two Research Projects of The Hong Kong Polytechnic University (Project No. G-YBLR and G-YBFE).

References

Alguliev RM, Aliguliyev RM, Isazade NR (2013). Multiple documents summarization based on evolutionary optimization algorithm. Expert Systems with Applications, 40(5), 1675-1689.

Bonzanini M, Martinez-Alvarez M, Roelleke T (2013). Extractive summarisation via sentence removal: Condensing relevant sentences into a short summary. In SIGIR '13, pp. 893-896. ACM.

Chen J, Huang DP, Hu S, Liu Y, Cai Y.(2015) An opinion mining framework for Cantonese reviews. Journal of Ambient Intelligent Human Computing, 6(5), 541-547.

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

Fellbaum C, Miller G. (1998). WordNet:An Electronic Lexical Database: MIT Press.

Frey BJ, Dueck D. (2007). Clustering by passing messages between data points. science, 315(5814), 972-976.

Ganapathibhotla M, Liu, B. (2008). Mining opinions in comparative sentences. In Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1 (pp. 241-248). ACL.

Hai Z, Cong G, Chang K, Liu W, Cheng P. (2014). Coarse-to-fine review selection via supervised joint aspect and sentiment model. In SIGIR '14, pp. 617-626. ACM.

Jin J, Ji P, Gu R. (2016). Identifying comparative customer requirements from product online reviews for competitor analysis. Engineering Applications of Artificial Intelligence, 49, 61-73.

Jin J, Ji P, Kwong CK. (2016). What makes consumers unsatisfied with your products: Review analysis at a fine-grained level. Engineering Applications of Artificial Intelligence, 47, 38-48.

Jin J, Liu Y, Ji P, Liu HG. (2016). Understanding Big Consumer Opinion Data for Market-driven Product Design, International Journal of Production Research, Volume 54, Issue 10, 3019-3041.

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

Jindal N, Liu B. (2006). Mining comparative sentences and relations. In AAAI '06 pp. 1331-1336

Kim HD, Zhai C. (2009). Generating comparative summaries of contradictory opinions in text. In CIKM '09, pp. 385-394. ACM.

Kim SM, Hovy E. (2006). Automatic identification of pro and con reasons in online reviews. In Proceedings of the COLING/ACL on Main conference poster sessions (pp. 483-490). ACL.

Korenek P, Šimko M. (2014). Sentiment analysis on microblog utilizing appraisal theory. World Wide Web, 17(4), 847-867.

Krestel R, Dokoohaki N. (2011). Diversifying product review rankings: Getting the full picture. In WI-IAT '11, 138-145. IEEE.

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

Liu B. (2010). Sentiment Analysis and Subjectivity. In N. I. a. F. J. Damerau (Ed.), *Handbook of Natural Language Processing, Second Edition* (Vol. 30, pp. 152-153). New York, NY, USA: Marcel Dekker.

Liu Y, Jin J, Ji P, Harding JA, Fung RYK, 2013. Identifying Helpful online reviews: a product designer's perspective. Computer Aided Design. 45(2), 180–194.

Lu Y, Zhai C, Sundaresan N. (2009). Rated aspect summarization of short comments. In WWW '09, 131-140. ACM.

Ma Z, Sun A, Yuan Q, Cong G. (2012). Topic-driven reader comments summarization. In CIKM '12, pp. 265-274. ACM.

McDonald R. (2007). A study of global inference algorithms in multi-document summarization. In ECIR'07, pp. 557-564

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

Nguyen TS, Lauw HW, Tsaparas P. (2013). Using micro-reviews to select an efficient set of reviews. In CIKM'13, pp. 1067-1076. ACM.

Pang B, Lee L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In ACL'04, pp. 271, ACL.

Paul MJ, Zhai C, Girju R. (2010). Summarizing contrastive viewpoints in opinionated text. In EMNLP '10, pp. 66-76

Shen HC, Wang KC. (2016). Affective product form design using fuzzy Kansei engineering and creativity. Journal of Ambient Intelligent Human Computing, 7(6):875-888

Tao F, Zhao D, Yefa H, Zhou Z. (2010) Correlation-aware resource service composition and optimal-selection in manufacturing grid. European Journal of Operational Research, 201(1):129-143

Tao F, Zhao Y, Hu H, Zhou Z. (2008) Resource service composition and its optimal-selection based on particle swarm optimization in manufacturing grid system. IEEE Transactions on industrial informatics, 4(4):315-327

Tian N, Xu Y, Li Y, Pasi G. (2015). Quality-Aware Review Selection Based on Product Feature Taxonomy. In AIRS '15, pp. 68-80.

Tsaparas P, Ntoulas A, Terzi E. (2011). Selecting a comprehensive set of reviews. In KDD '11, pp. 168-176. ACM.

Conrad T, Harrison MK (2011). Predicting emerging product design trend by mining publicity available customer review data. In ICED'11, pp. 43-52, Copenhagen, Denmark

Wei W, Ming Z, Nie L, Li G, Li J, Zhu F, Shang T, Luo C. (2016). Exploring heterogeneous features for query-focused summarization of categorized community answers. Information Sciences, 330, 403-423.

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

Wilson T, Wiebe J, Hwa R. (2004). Just how mad are you? Finding strong and weak opinion clauses. In AAAI '04, Vol. 4, pp. 761-769.

Xu K, Liao SS, Li J, Song Y. (2011). Mining comparative opinions from customer reviews for Competitive Intelligence. Decision support systems, 50(4), 743-754.

Yu J, Zha, ZJ, Wang M, Chua TS. (2011). Aspect ranking: identifying important product aspects from online consumer reviews. In HLT '11, pp. 1496-1505. ACL

Zhang K, Narayanan R, Choudhary AN. (2010). Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking. In WOSN '10, 11-11.

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

Zhu L, Gao S, Pan SJ, Li H, Deng D, Shahabi C. (2013). Graph-based informative-sentence selection for opinion summarization. In International Conference on Social Networks Analysis and Mining, pp. 408-412.