Extracting Customer Concerns from Online Reviews of Series Products for Competitor Analysis

Sixing Yan[1], Jian Jin[1,2], Ping Ji[2,3] and Zihao Geng[4]
[1] Department of Information Management, Beijing Normal University, Beijing 100875, China
[2] The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, China
[3] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China
[4] H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

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 researchers exploit how to apply online concerns in the competitor analysis about the merits and drawbacks of series products. Accordingly, in this research, a framework is presented to sample representative sentences from online reviews, aiming to highlight similar customer concerns of series products. First, opinionated sentences of specific features are identified. Then, opinionated sentences in the same series products are clustered to extract similar customer concerns. Finally, an optimization problem is formulated for sampling a few representative sentences. With real data from Amazon.com, categories of experiments were conducted to evaluate the effectiveness of the proposed approach. This study explores the possibility about integrating big consumer data into competitor analysis in the market driven product design, which is essentially critical in fierce market competition.
Keywords: review analysis; online reviews; competitor analysis; series product comparison; customer concerns
Citation: Yan, S., Jin, J., Ji, P., & Geng, Z. (2017). Extracting Customer Concerns from Online Reviews of Series Products for Competitor Analysis. In *iConference 2017 Proceedings* (pp. 142–156). https://doi.org/10.9776/17017

Acknowledgements: The work described in this paper was supported by two grants from the National Nature Science Foundation of China (project no. NSFC 71272177/G020902 and project no. NSFC 71271185/G011201)
Contact: jinjian.jay@bnu.edu.cn

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 because valuable opinionated text about customer concerns are expressed. Online reviews help potential consumers to grasp the merits and drawbacks of products and benefit 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 (Abbasi, Chen, & Salem, 2008; Ding & Liu, 2007; Korenek & Šimko, 2013). However, with all different intelligent algorithms available, one practical and significant problem is that, designers are still embarrassed with the comparisons about the merits and drawbacks of series products. Specifically, for product designers, there is still lack of intelligent approaches to make efficient comparisons on a list of similar products in one brand and competitors. For example, in Table 1, similar online opinions of three smart phones in one series are compared. As presented, in these three sentences, customers discuss their similar experiences about sharing photos with their families and present positive opinions on this feature. To analyze such customer concerns, an effective approach to identify comparative customer concerns over series products are required, with a purpose of enlightening designers in understanding the strength and weakness about a series of products in a specific brand.

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.

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

Recently, a limited number of studies are reported to utilize customer online reviews for competitor analysis (Chen, Qi, & Wang, 2012; Jin, Ji, & Gu, 2016; H. D. Kim & Zhai, 2010). These studies select opinionated sentences from customer reviews to construct comparative summaries, aiming to detect the contradictory opinions from a big volume of online reviews. But product serializations usually imply that every product of one series in one brand should concerned at the same time, which induces that the competitor analysis on the brand level becomes the comparison of multiple similar products. Nonetheless, many existing methods report that the contrastive viewpoint extraction for competitor analysis focuses on one-to-one comparison. These solutions are not aligned with the analysis about ongoing customer concerns or product defects in a series of products efficiently. Therefore, in this research, a framework for competitor analysis is outlined which extracts customer concerns from reviews of a series of products. It aims to highlight and compare similar customer concerns of series products. Specifically, an optimization problem is formulated to sample representative sentences from reviews of competitive series products, in which the information coverage and the information diversity of the sampled sentence subset is concerned.
The contributions of this research are at least three-fold:

a) A framework of representative sentence sampling from reviews of series products is outlined to analyze the merits and drawbacks of competitive products of ongoing customer concerns. The crucial information about multiple similar but competitive products is presented, aiming to intelligent marketing and product benchmarking.
b) To highlight and compare customer concerns of series products, the sampling of representative sentences is formulated as an optimization problem. The sampling results contain details of product features which cover all products of that series, as well as illustrate relationships of different series products. This framework helps product designers to compare series products of brand competitors in specific product features.
c) With a real large dataset from Amazon.com, categories of experiments were conducted for performance evaluation and comparisons with different approaches. The experiment results illustrates the effectiveness of the proposed approach in the task of representative sentence sampling.

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

One of the major objectives about this research is to sample a sentence subset from a given review corpus, which relates to some existing research studies on review summarization and review selection. Additionally, comparisons on products are often formulated as how to detect comparative opinionate segments. Therefore, studies on contrastive viewpoint extraction are also briefly reviewed.

2.1 Review Summarization and Review Selection

Studies of review summarization focus on the topics extraction of customer reviews to present the sketch outline of a given reviews corpus.

An optimization-based model was built for summarization via salient sentences (Alguliev, Aliguliyev, & Isazade, 2013). The relationships between sentences, summaries and reviews were utilized in the sentence selection. A phrases summarization for rated aspect of short comments was generated (Lu, Zhai, & Sundaresan, 2009). The rated aspects were taken as the various customer-opinion topics, as well as the phrases were taken as the detailed sentence pieces regarding those topics. An informative sentence selection approach was introduced in (Zhu et al., 2013), in which the review summarization was formulated as a problem of community-leader detection.

Many recent studies select a small number of representative reviews which covers various aspects about customer viewpoints. For instance, product reviews were selected according to review's quality, which was estimated by opinion distribution with a taxonomy tree of product features (Tian, Xu, Li, & Pasi, 2015). A few reviews in the fine-grained product aspect level were selected by a novel probabilistic graphical model (Hai, Cong, Chang, Liu, & Cheng, 2014). Lappas et al. (Lappas, Crovella, & Terzi, 2012) sampled a characteristic subset of reviews based to the opinion distribution of original review corpus. Multiple levels of similarity and coverage were also considered in review selection. Given a collection reviews and related words about a topic, an efficient review subset was selected (Nguyen, Lauw, & Tsaparas, 2013). The efficiency of a review was defined as the topic words it contained and review sets in various efficient levels were identified using greedy algorithms.

These approaches focus on sampling a helpful opinionated review subset. However, few of them are applicable for analyzing customer concerns of series products from valuable customer online opinions.

2.2 Contrastive Viewpoints Extraction

The generation of contrastive summarization benefits comparative opinion detection from large customer data. It is utilized to analyze competitors of products or bands. Different relevant studies were reported to investigate this problem as sentences extraction of customer online reviews.

A pattern discovery approach was proposed to discover comparative sentences in texts (Jindal & Liu, 2006), in which several categorized of class sequential rules were studied. Opinionated entities of customer opinions were extracted from comparative sentences (Ganapathibhotla & Liu, 2008). Comparative sentences were identified and concerned entities were then extracted. A two-stage approach was proposed to summarize contrastive opinions in text(Paul, Zhai, & Girju, 2010). Customer viewpoints were initially modeled and extracted according to lexical and syntactic features. Then pairs of sentences were scored considering the representativeness and the contractiveness. Similarly, consider a specific feature, a framework was proposed to select pairs of representative yet comparative sentences from competitive products (Jin, Ji, & Gu, 2016). Kim et al. (H. D. Kim & Zhai, 2010) proposed a contrastive opinion summarization, in which the content similarity with the same sentiment polarity and the opposite polarity were considered. Comparative sentences were identified from customer reviews using a two-level Conditional Random Fields (CRFs) model (Xu, Liao, Li, & Song, 2011).

Many approaches for contrastive viewpoints extraction initializes one-to-one comparisons about competitive products, which is arguably to be applied for the comparison of series product groups at the same time.

2.3 A Brief Summary

To sum up, different models were reported to extract valuable information from online reviews. However, few studies focused on providing a contrastive summarization for competitor analysis on series products level. Accordingly, in this research, how to obtain samples of representative sentences from reviews of series products efficiently is investigated, which is believed to be beneficial for the product development and competitor analysis of series products.

3 Problem Statement

The ultimate goal of this research is to facilitate designers to digest customer concerns from a large number of online reviews of series products efficiently. One of the effective approaches is to sample a limited number of representative review sentences that mirror critical customer concerns of a series of products. To futher clarify this problem, some definitions with examples are present in the following:

Definition 1 Product Feature, a feature of a particular product is an attribute or component of the product that has been discussed in reviews. For example, the "camera" of one smart phone in the following two opinionated sentences is referred as a feature:

- Example 1: "The front camera could be better."
- Example 2: "Only the selfie camera worked."

Definition 2 Feature Aspect, an aspect of a particular product feature is a component of the feature that has been detected in the feature-related review segments. In another word, these aspects can be denoted as subtopics about that feature, containing detailed information about customer concerns. For example, "battery life" and "battery charger" are two different aspects of "battery", which are showed in the following opinionate sentences.

- Example 3: "I was using the phone for about 1 and a half hours almost straight on the internet starting at around 95 battery life and ended up with 5 at the end."
- Example 4: "Only problem is not appropriate charger cable plug."

Suppose that a series of N products in the same brand, $P = \{p_1, p_2 \dots, p_N\}$, and M shared features of these products, $F = \{f_1, f_2 \dots, f_M\}$, are reckoned. Specially, for a specific product feature, various aspects might be concerned by customers. Then, sentences that talk about a particular aspect k of one feature f_m can be defined as a_{mk}, and $A_m = \{a_{m1}, a_{m2} \dots, a_{mK}\}$ can be regarded as the sentence set that refers all K different aspects of f_m. Additionally, it can be also observed that some special feature aspects are discussed across each product in P and, in this research, such aspects are referred as serial aspects $C_m = \{c_{m1}, c_{m2} \dots, c_{m|C_m|}\}$ and $|C_m| \leq K$. Then, the corresponding sentence sets of a particular serial aspects $c_{m\mu}$ of C_m can be denoted as $S_{m\mu} = \{s_1^{m\mu}, s_2^{m\mu} \dots, s_N^{m\mu}\}$ where $s_n^{m\mu}$ of $S_{m\mu}$ is denoted as a collected sentence set regarding product p_n.

Note that the objective of this research is to provide a concise summarization about customer concerns for a series of products. Typically, consider a particular feature, a review sentence set of series products is taken as input, and a representative sentence subset T is sampled from S. This list of sentences $T_{m\mu}$ regarding a serial aspect $c_{m\mu}$ is denoted as $T_{m\mu} = \{t_1^{m\mu}, t_2^{m\mu} \dots, t_N^{m\mu}\}$, and $t_n^{m\mu}$ indicates a group of sentences sampled from reviews of product p_n. T is expected to highlight and compare customer concerns about series products, in support of competitor analysis in the design process for the next model in the series. To clarify what characteristics about sampled sentences would be, some other fundamental concepts are introduced.

Definition 3 Information Coverage, reveals the information that is covered by a sampling 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 sentences subset covers a major part of S.

Definition 4 Information Diversity, reveals the non-overlap information that is covered a sampling subset T, which is denoted as $Diversity(T)$. The value of information diversity is high when different messages are covered in T.

Definition 5 Information Representativeness, indicates the representatives about a sampling subset of sentences T that is compared with a given sentence set S, which is denoted as $Representativeness(T,S)$. A high information representative subset of sampled sentences should intuitively have both high information coverage and information diversity. Thus, the problem of sentence sampling is 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 for trade-off the *Coverage* and *Diversity* of sampling sentences. Therefore, the major concern of this research becomes how to sample a subset of sentences from reviews of series products that maximize the information representativeness.

Note that, in this study, strength levels of opinions are neglected (Wilson, Wiebe, & 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 also valuable for product designers. However, the focus of this research is on comparisons of multiple comparable products and these problems will be investigated in the future work.

4 Methodology

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

Symbol	Definition
P	The series product set, $P = \{p_i\}_N$
F	The shared feature set of the series products, $F = \{f_j\}_M$
R_m	The review sentences of all series products regarding feature f_m, $R_m = \{r_{m\varphi}\}_{\|R_m\|}$
U_m	The similarity matrix of sentences from R_m, $U_m = u^m_{\varphi\varphi'}$, $(\varphi \le \|R_m\|)$
A_m	The aspect set of a particular product feature f_m, $A_m = \{a_{mk}\}_{k=1}^{K}$
C_m	The serial aspect set of a particular feature f_m, $C_m = \{c_{m\mu}\}_{\mu=1}^{\|C_m\|}$, $(C_m \subseteq A_m)$
$S_{m\mu}$	The set of related sentence regarding a serial aspect $c_{m\mu}$, $S_{m\mu} = \{s_i^{m\mu}\}_{i=1}^{N}$
$T_{m\mu}$	The set of sampled sentence regarding a particular serial aspect $S_{m\mu}$, $T_{m\mu} = \{t_i^{m\mu}\}_{i=1}^{N}$, $(T_{m\mu} \subseteq S_{m\mu})$
H	The limited count of sampling sentences of set $T_{m\mu}$
h	The minimum number of sampling sentences of subset $t_i^{m\mu} \in T_{m\mu}$, where $N \times h \le H$

Table 2. Notations and definitions

4.1 Framework Overview

In this research, a framework is proposed to sample a subset of representative sentences for the analysis of series products of same brand and competitors. It is named **S**erial **R**epresentative **S**entence **S**ampling (**SRSS**). 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 workflow is presented in Figure 1.

4.2 SRSS-I: Extract Product Features

Given online reviews of one particular product domain, the first task is to extract product features and analyze their corresponding sentimental polarities. In this research, with the help of merits and drawbacks reviews, such as reviews in Cnet.com and Epinions.com, a simple yet effective method is employed.

In the beginning, product features are extracted from merits and drawbacks reviews, in which features are defined as the nouns or noun phrases with a high frequency (Liu, 2010). For example, "battery", "camera", "screen", "media" and "application" are extracted from merits and drawbacks reviews of smartphones. Then, sentences regarding product features are identified from customer reviews.

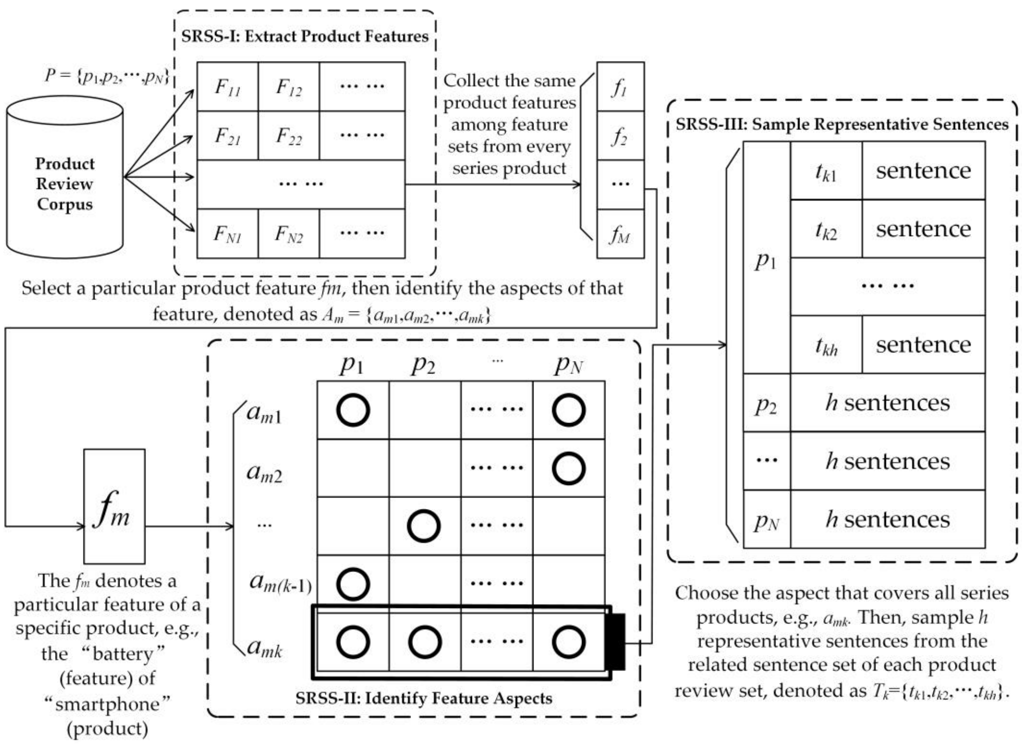


Figure 1. Overview of the steps in SRSS

Accordingly, with the help of review corpus in (Pang & Lee, 2004), a binary classifier is built to classify the attitude of a sentence in subjective or objective. Next, with the opinionate information in merits and drawbacks reviews, another binary classification approach is utilized to classify the sentiment of a subjective sentence in positive or negative. In this approach, each sentence is represented as a bag of sentimental terms defined in MPQA (Wilson, Wiebe, & Hoffmann, 2005), and review sentences in merits and drawbacks reviews are utilized as training corpus. Similar approaches for product feature identification and sentiment analysis were also reported in (Jin, Ji, & Gu, 2016; Jin, Ji, & Kwong, 2016; S. M. Kim & Hovy, 2006; Yu, Zha, Wang, & Chua, 2011).

Note that an objective sentence might also contain valuable information about customer concerns. Similar to the assumptions made in other relevant studies (Jin, Ji, & Gu, 2016; H. D. Kim & Zhai, 2010), in this study, only positive and negative sentences are taken as the major concerns for sampling representative sentences. Accordingly, these sentences of specific features are collected into different sets.

4.3 SRSS-II: Identify Feature Aspects

The second task in the SRSS is to identify different aspects of product features and classify customer concerns from online reviews.

Since the same feature aspect might 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 set R_m regarding a feature f_m, the similarity matrix of sentences U_m is constructed, where $u^m_{\varphi\varphi'} = Similarity(r_{m\varphi}, r_{m\varphi'}), (r_{m\varphi}, r_{m\varphi'}) \in R_m$. Next, feature-related sentences are clustered according to U_m. In this research, a self-adaptation clustering method in (Frey & Dueck, 2007) is employed and the clustering results are taken as different groups of relevant sentences regarding different aspects A_m.

Notice that some feature aspects might be observed across different products of that series. In this research, these aspects are defined as serial aspects C. Serial aspects echo 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 a feature of a smart phone for example. "battery overheating" of "battery" is continually observed in customer reviews of each product in that series. Accordingly, in this research, "overheating" is taken as a serial aspect of "battery". Thus, review sentences regarding a serial aspect c of 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, given a serial aspect, how to sample representative sentences from the sentence set regarding will be discussed. Recall that the representative sentences in sampling results are expected to balance two criterions, i.e., *information coverage* and *information diversity*. Therefore, with the objective function of sentence sampling in Eq. (1), how to model these two criterions and how to sample a subset of representative sentences are studied.

Information Coverage. As aforementioned, the sentence sampling is expected to cover general details of aspects, which means the sampled sentence set T are expected to cover the main content of the entire sentence set S. This expectation can be denoted as,

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

Information Diversity. The sentence sampling is expected to cover different subtopics of aspects, which means each sampled sentence of T are expected to be dissimilar to the rest of sampled sentences. This expectation 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 evaluate the similarity and the distance between two sentences. In this research field, different approaches are reported to model the sentence similarity. However, some models are quite complex for product designers. To lessened the difficult implementation of these tasks and focus on the application of the proposed framework, in this research, a simple but effective similarity model is employed. In Eq. (4), the sentence similarity between two sentences X and Y is defined, considering semantic matching between terms,

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

where $\Phi(x, y)$ is a term similarity function and $|X|$ and $|Y|$ are the total counts of words in X and Y. A similar definition about the sentence similarity can be also found in (H. D. Kim & Zhai, 2010). Depending on how Φ is defined, different variations can be obtained. In this research, two natural variants are experimented:

a) **Word Overlap**: $Similarity(X,Y)_{WO}$: $\Phi_{WO}(x,y)=1$ if $x=y$, and $\Phi_{WO}(x,y)=0$ otherwise. It is naturally the *Jaccard* similarity function that considers word overlap.
b) **Semantic Word Matching**: $Similarity(X,Y)_{WN}$: $\Phi_{SWM}(x,y)=1$ if $x=y$, and $\Phi_{SWM}(x,y)=sim(x,y)$ otherwise. $sim(x,y)$ refers to the semantic term similarity. In this research, the normalized value $WordNet(x,y)\in[0,1]$ is employed. $WordNet(x,y)$ evaluates the shortest path distance about the conceptual relations of two terms that is defined in WordNet (Fellbaum & Miller, 1998). Then, $sim(x,y)$ is set to be $1-WordNet(x,y)$ which measures how similar the two terms are.

Similar to $Similarity(X,Y)$, the dissimilarity between two sentences X and Y can be defined as,

$$Distance(X,Y)=1-Similarity(X,Y) \tag{5}$$

where $Distance(X,Y)_{WO}$ and $Distance(X,Y)_{SWM}$ are conducted with $\Phi_{WO}(x,y)$ and $\Phi_{SWM}(x,y)$.

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

$$T=\arg\max_{T}\left((1-\alpha)\frac{1}{|T|\times|S|}\sum_{t\in T,s\in S}Similarity(t,s)+\alpha\frac{1}{|T|\times|T|}\sum_{t,t'\in T}Distance(t,t')\right) \tag{6}$$

where $|T|\le H$ and $|t|\ge h(t\in T)$. Actually, different methods are available to analyze such optimization problem, such as the *greedy-based approach* (McDonald, 2007) and the *fast approximate spectral clustering* (Wei et al., 2016). In this research, the method in (McDonald, 2007) is adopted to solve this optimization problem.

5 Experiment Study and Discussion

5.1 Experimental Setup

In this section, a case study is presented to demonstrate how to utilize **SRSS** to sample the representative sentences from series product reviews. 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 number of reviews about these products is presented in Table 3. In this case study, customer reviews of three series products in two given brands are analyzed. For data privacy, the names of these two brands are represented as Brand 1, 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

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

5.2 Evaluation Metrics

Conventionally, for review analysis, it is burdensome to acquire training samples 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 by hand, which thus makes it hard to apply some widely-utilized metrics to evaluate the performance of the proposed approach. Therefore, in this research, three simply evaluation metrics are employed, i.e., information redundancy, information coverage and information centralization.

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

$$Redundancy(T) = \frac{1}{|T|\times|T|}\sum_{t,t'\in T} Similarity(t,t') \tag{7}$$

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

$$Coverage(T,S) = \frac{1}{|T|\times|S|}\sum_{t\in T,s\in S} Similarity(t,s) \tag{8}$$

c) **Information centralization (ICT)**, evaluates to what the extent sampled sentences cover the content of original reviews. This metric aims to figure out how sampled sentences are affected by high-frequent words. The sampled sentence set is expected to have a low ICT value to cover content words with different frequency. ICT can be estimated by evaluating the count and frequency of words that T covers. It will increase if T covers lots of high-frequent words. Given a frequent content word set V of the original reviews, ICT can be formulated as:

$$Centralization(T,V) = \frac{1}{|D(V,T)|}\sum_{v\in V} W(v,V)\times D(v,T) \tag{9}$$

$W(v,V)$ 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.

In the following experiments, to evaluate the performance of **SRSS**, $Similarity(X,Y)_{WN}$ is employed to evaluate the word similarity in different pairs of sentences. Furthermore, for benchmarking with similar algorithms, a simple approach of review sentence selection, the RANDOM Sampling, is introduced as basic benchmark approaches. It was also utilized in (Tsaparas, Ntoulas, & Terzi, 2011). Additionally, two greedy-based approaches for the one-to-one product comparison analysis (COS), which proposed in (H. D. Kim & Zhai, 2010), are also employed to benchmark. They are denoted as **COS-1** and **COS-2**. Accordingly, given a customer review corpus of N products in the same series, **SRSS** is compared with all these four benchmark approaches.:

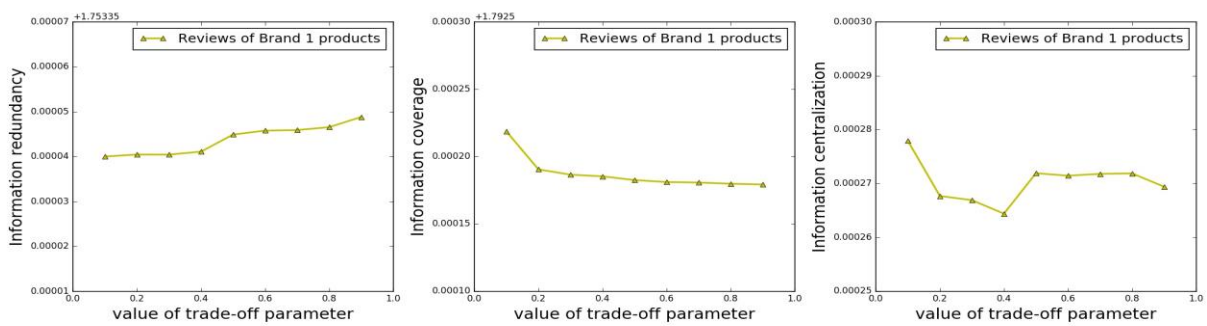
a) **RANDOM Sampling (RD)**: h sentences are selected randomly as the sample subset by the RANDOM algorithm, which is meant to serve as a sanity check that calibrates the results of the other algorithms. In this experiment, 2000 runs will be performed for the RANDOM sampling and the average value of performance are taken;
b) **COS-1 Sampling (COS-1)**: Given a review sentence set R_m regarding f_m of a particular product p_i, q sentences in negative polarity and q sentences in positive polarity are selected. Accordingly, two sets of $q\times N$ sentences regarding all products in P are collected, which are 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, and q pairs are selected as the result. Selected sentences in each pair are expected to have high similarity regardless the sentiment polarity. To sample sentence pairs across N series products, $2\times q\times N$ sentences are selected as the result eventually;
c) **COS-2 Sampling (COS-2)**: Given a review sentence set R_m regarding f_m of product set P, negative sentences and positive sentences are collected and combined as pairs, respectively. Collected

sentences in each pair are expected to have high similarity regardless the sentiment polarity. Then q pairs of the whole set are selected as the sampling results which are expected to have the highest similarity to the original set R_m. Similar to **COS-1**, to sample sentence pairs across N series products, $2{\times}q{\times}N$ sentences are selected as the final result.

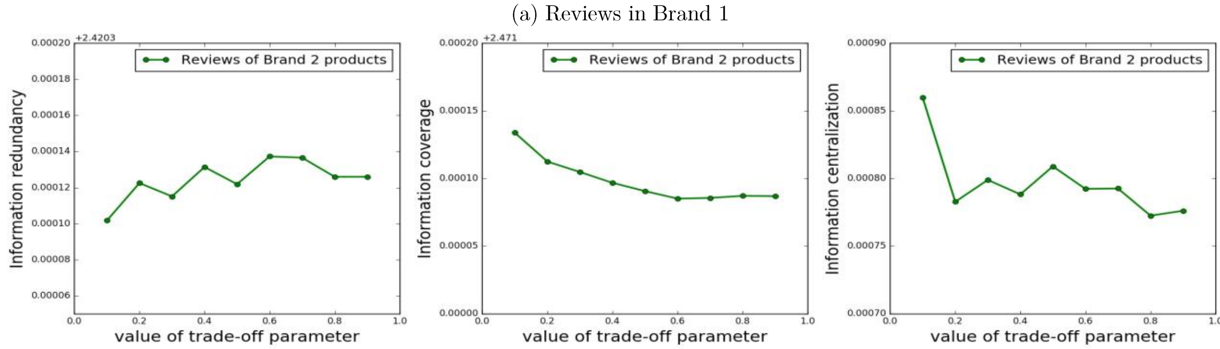
To evaluate the performance of each approach, the number of samplings is expected to be equal. Note that, in this experiment, three products in each series are selected and thus N equals to three. q is set to be 2 to 8, which means that 12, 18, 24, 30, 36, 42, 48 sentences are sampled, respectively, since that **COS-1** and **COS-2** totally select $2{\times}q{\times}N$ sentences.

5.3 On the Sensitivity of Parameter

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 also observed if α is moderate. One of potential reason is that with an increasing weight of $Coverage$ in Eq. (1), of which weight is $1-\alpha$, more sentences that are similar to the original sentence set tends to be sampled, which causes the $Diversity$ of sampled sentences becomes smaller. Thus, the value of ICR is achieved higher, while sampled sentences are semantically similar, which then leads the value of ICT get higher.

α can be set as various values according to which metrics matter higher. 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 lower values of IRD and ICT and a higher value of ICR are balanced. In the following, α is set to 0.6, in which better performance is observed in terms of all three metrics.

5.4 Evaluation Results and Analysis

For the effectiveness of the proposed approach, categories of experiments were conducted to analyze review sentences that refer to the battery and the camera of Brand 1 and Brand 2. In these experiments, sentences are selected from the corresponding opinionated review sentences. Then, sentences are sampled from corresponding opinionated review sentences and evaluated in terms of the three metrics. The results are reported in Table 4, in which the bold number indicates that it is the best desired performance among other benchmarks except **RD** 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.

Note that, IRD and ICT are the lower the better while ICR is the higher the better. As seen from Table 4, a lower IRD value and a lower ICT value are obtained by **SRSS**. It can be claimed that review sentences with various details are sampled and they cover words in different frequency instead of high-frequent words only. These results indicate that, compared to other benchmark approaches, **SRSS** is capable to sample the richly detailed sentences across several series products. They reveal the relationships of customer concerns about different series products. In addition, it can also be found that moderate ICR values are obtained by **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, in these experiments, for purpose of obtaining a high $Diversity$, sentences with the same words are seldom sampled which thus causes the $Coverage$ low. Note that, for a particular feature, large parts of customers concern about some similar topics are sampled. Accordingly, sampling results, which cover large number of these same topics and have a high level of word-overlap, obtain an obviously high value of $Coverage$ while low value of diversity. **SRSS** avoids sampling too many sentences that reveal the same topics, thus it obtains a moderate ICR.

	Brand 1						Brand 2					
	Battery			Camera			Battery			Camera		
	IRD	ICR	ICT	IRD	ICR	ICT	IRD	ICR	ICT	IRD	ICR	ICT
RD	2.253	2.316	7.59E-04	2.253	2.314	1.12E-03	2.255	2.320	9.76E-04	2.255	2.318	1.13E-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**
COS-1+*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
COS-2+*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(SWM)												
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**
COS-1+*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
COS-2+*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

Table 4. Evaluate the proposed approach with other benchmarks

On the contrary, the approach of **COS-1** considers more on covering the main content of the original sentence set, leading the ICR be high; also, the approach of **COS-2** focuses on obtaining high 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 their ICR are high. In short, **COS-1** and **COS-2** have strong ability to sample sentence subsets with a higher ICR by utilizing $Similarity(X,Y)_{WO}$. Even in experiments utilizing $Similarity(X,Y)_{WN}$, **COS-1** and **COS-2** just obtained moderate lower ICR than **SRSS**. However, **SRSS** performs better in sampling sentences. Its sampling results 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.

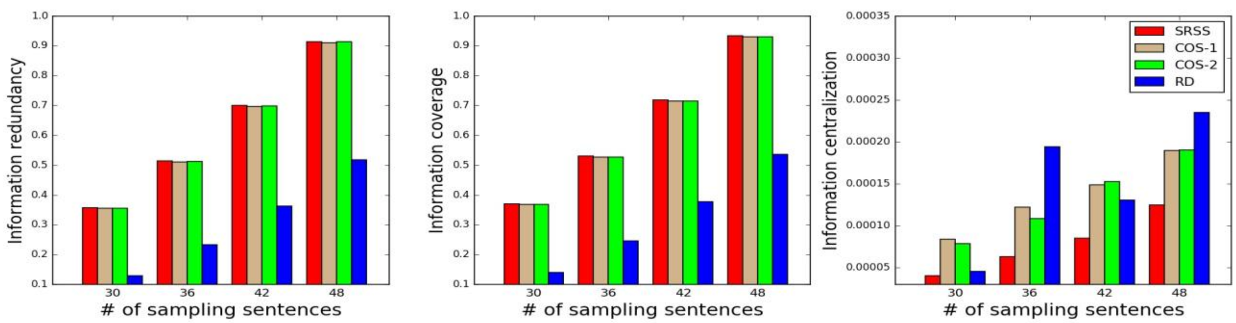
To evaluate the performance regarding the number of sample sentences, experiments were conducted using reviews of "battery" and similarity measure Φ_{WO}. The number of sample sentences is set to 30, 36, 42, 48. The final evaluation results are presented in Figure 3.

5.5 Case Study

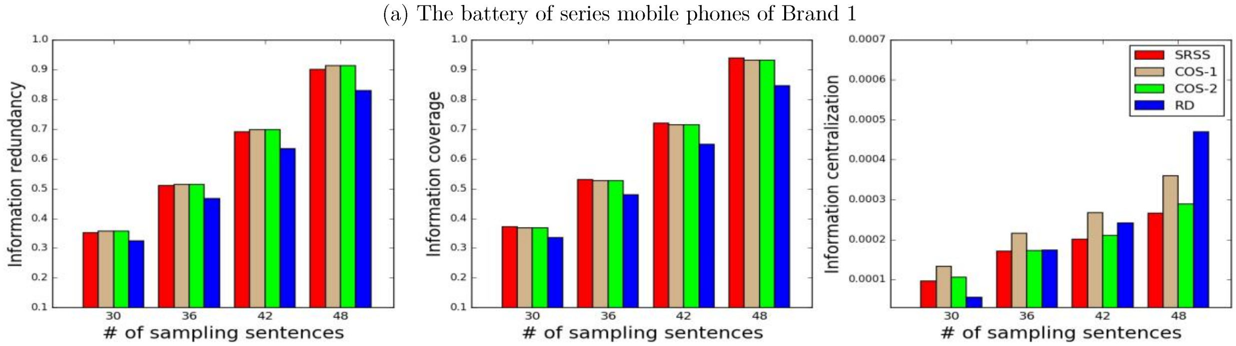
To show how to utilize **SRSS** by product designers, 1,078 reviews of the battery in Brand 1 and Brand 2 are utilized as an illustrative example.

Now, suppose designers of the mobile phone care about opinions on the battery only. As presented in the proposed framework in Section 4, product features are extracted and serial aspects of features are identified with merits and drawbacks reviews. To evaluate the effectiveness of the sampling results, in this experiment, opinionated sentences in negative polarity are taken into consideration only. Accordingly, 706 negative sentences referring to the 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 which customer concerns regarding different products of this series are briefly presented.

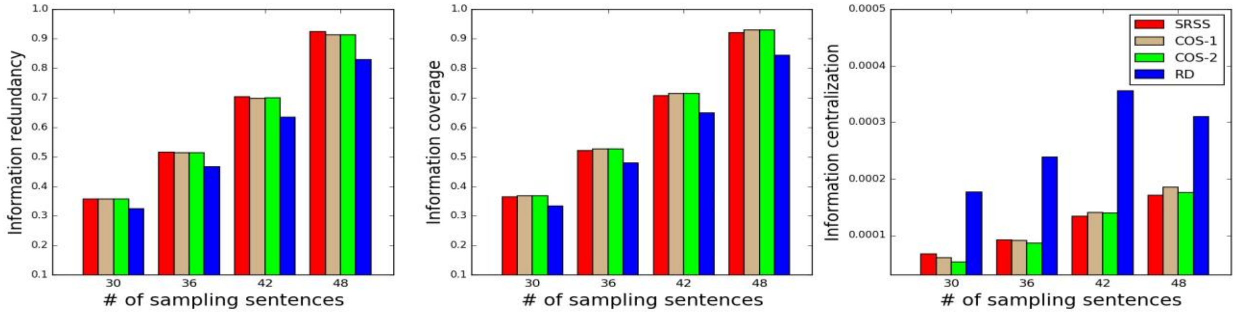
In Figure 5, an example is illustrated to show how **SRSS** benefits product designers for the competitor analysis at the level of series products. In this example, customer online reviews for series products of two brands are analyzed by **SRSS**. The proposed approach is utilized to sample representative sentences that reveal customer concerns about products of these two series. To make cross-comparison with these series products, review sentences of two brands are consolidated where comparable customer concerns are compared clearly. The representative sentence sets help to enlighten product designers with strengths and weaknesses of series products in multiple brands, respectively. Comparing with the example in Figure 4, series product comparison of multiple brands is conducted efficiently by analyzing the voices of online customers.



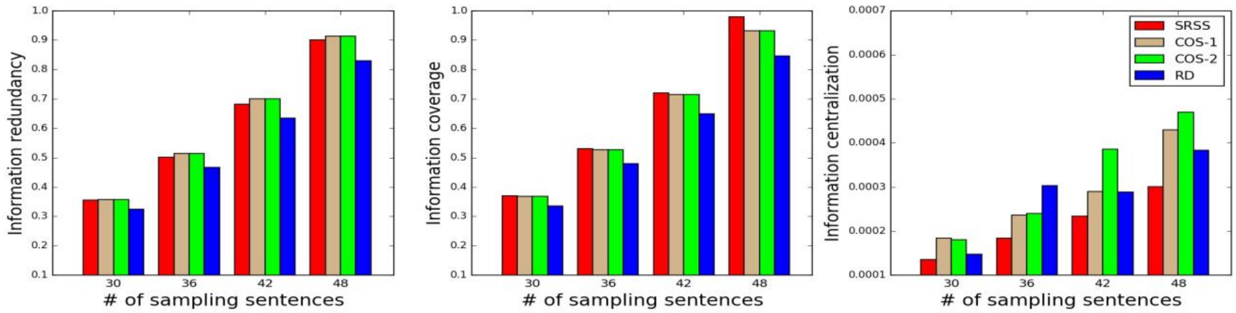
(a) The battery of series mobile phones of Brand 1



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



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



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

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

6 Conclusions

In this research, the problem of how to extract customer online concerns of series products for competitor analysis is investigated. Its core is to sample a subset of representative sentences from reviews of series products of the same brand and competitors. Specifically, a three-phase framework is proposed for this problem 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 wide range of real reviews of Amazon.com. These results demonstrate the effectiveness of the proposed approach.

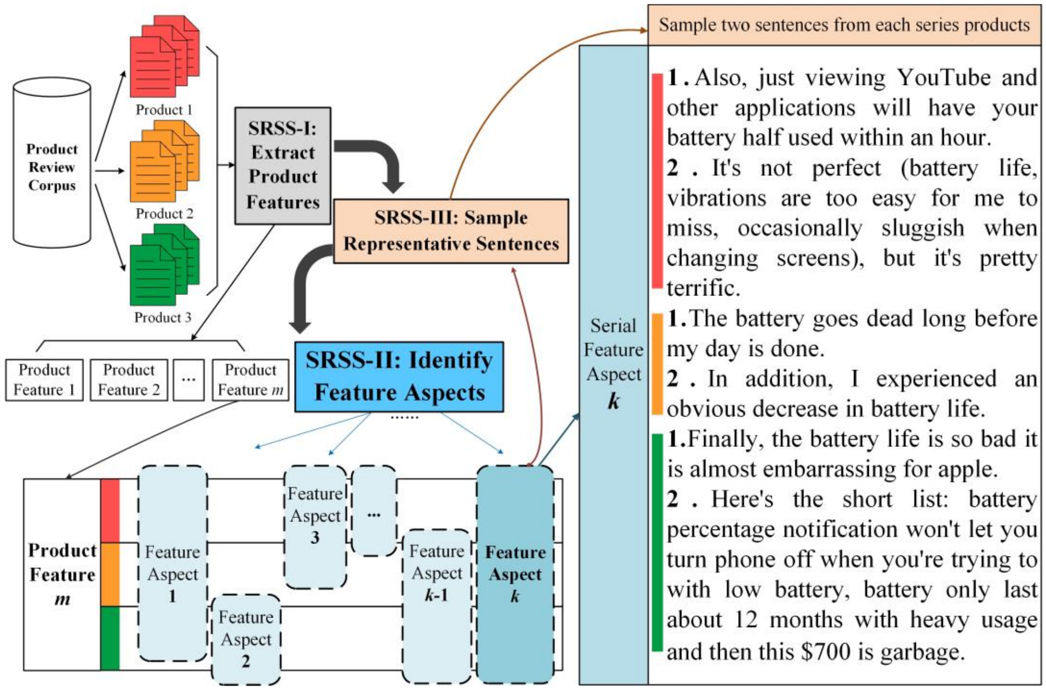


Figure 4. An example of sentence sampling regarding battery in series products of Brand 1

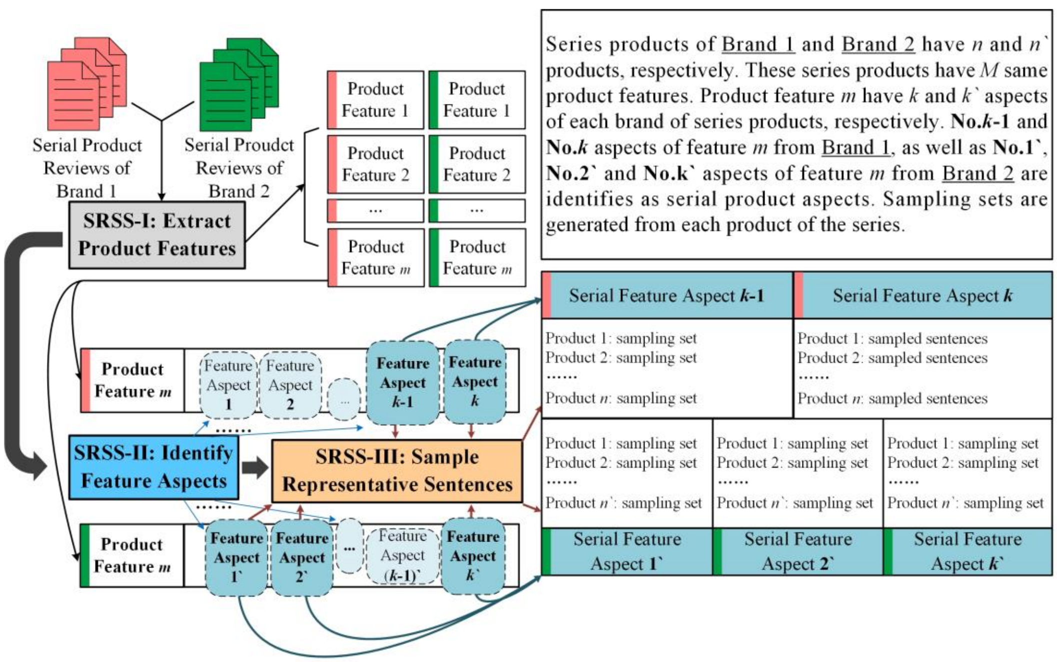


Figure 5. An example of series product analysis of two brand competitors

Some potential valuable research studies can be extended, such as how to make comparisons of the proposed approach with different similarity functions according to the sentence alignment, how to develop sophisticated algorithms to achieve better approximation solution for the optimization problem, etc.

7 References

Abbasi, A., Chen, H., & Salem, A. (2008). Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: feature selection for opinion classification in web forums. *Acm Transactions on Information Systems*, 26(3), 55-59.

Alguliev, R. M., Aliguliyev, R. M., & Isazade, N. R. (2013). Multiple documents summarization based on evolutionary optimization algorithm. *Expert Systems with Applications*, 40(5), 1675-1689.

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.

Ding, X., & Liu, B. (2007). The utility of linguistic rules in opinion mining. *Proceedings of ACM SIGIR'07*, 811-812.

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

Frey, B. J., & 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. *Proceedings of ICCL'08*, 241-248.

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

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, C. K. (2016). What makes consumers unsatisfied with your products: Review analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence, 47*, 38-48.

Jindal, N., & Liu, B. (2006). Identifying comparative sentences in text documents. *Proceedings of ACM SIGIR'06*, 244-251.

Kim, H. D., & Zhai, C. X. (2010). Generating comparative summaries of contradictory opinions in text. *Proceedings of ACM CIKM'10*, 385-394.

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

Korenek, P., & Šimko, M. (2013). Sentiment analysis on microblog utilizing appraisal theory. *World Wide Web*, 17(4), 847-867.

Lappas, T., Crovella, M., & Terzi, E. (2012). Selecting a characteristic set of reviews. *Proceedings of ACM SIGKDD'12*, 832-840.

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.

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

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

Nguyen, T. S., Lauw, H. W., & Tsaparas, P. (2013). Using micro-reviews to select an efficient set of reviews. *Proceedings of ACM CIKM'13*, 1067-1076.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of ACL'04*, 271-278.

Paul, M. J., Zhai, C. X., & Girju, R. (2010). Summarizing contrastive viewpoints in opinionated text. *Proceedings of EMNLP'10*, 66-76.

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

Tsaparas, P., Ntoulas, A., & Terzi, E. (2011). Selecting a comprehensive set of reviews. *Proceedings of ACM SIGKDD'11*, 168-176.

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. *Proceedings of EMNLP-CoNLL'05*, 347-354.

Wilson, T., Wiebe, J., & Hwa, R. (2004). Just how mad are you? Finding strong and weak opinion clauses. *Proceedings of AAAI'04*, 761-769.

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

Yu, J., Zha, Z. J., Wang, M., & Chua, T. S. (2011). Aspect ranking: identifying important product aspects from online consumer reviews. *Proceedings of NAACL HLT'11*, 1496-1505.

Zhu, L., Gao, S., Pan, S. J., Li, H., Deng, D., & Shahabi, C. (2013). Graph-based informative-sentence selection for opinion summarization. *Proceedings of IEEE/ACM ASONAM'13*, 408-412.