


RESEARCH

Open Access



Helping consumers to overcome information overload with a diversified online review subset

Zhang Jin¹ , Weng Zhangwen¹ and Ni Naichen^{2*}

* Correspondence: ncni@umich.edu

²Department of Statistics, University of Michigan, Ann Arbor MI48109, USA

Full list of author information is available at the end of the article

Abstract

Redundant online reviews often have a negative impact on the efficiency of consumers' decision-making in their online shopping. A feasible solution for business analytics is to select a review subset from the original review corpus for consumers, which is called review selection. This study aims to address the diversified review selection problem, and proposes an effective review selection approach called Simulated Annealing-Diversified Review Selection (SA-DRS) that considers the semantic relationship of review features and the content diversity of selected reviews simultaneously. SA-DRS first constructs a feature taxonomy by utilizing the Latent Dirichlet Allocation (LDA) topic model and the Word2vec model to measure the topic relation and word context relation. Based on the established feature taxonomy, the similarity between each pair of reviews is defined and the review quality is estimated as well. Finally, diversified, high-quality reviews are selected heuristically by SA-DRS in the spirit of the simulated annealing method, forming the selected review subset. Extensive experiments are conducted on real-world e-commerce platforms to demonstrate the effectiveness of SA-DRS compared to other extant review selection approaches.

Keywords: Business analytics, Online reviews, Feature taxonomy, Diversified subset, Review selection, Simulated annealing-diversified review selection (SA-DRS), E-commerce

Introduction

With the rapid development of the Internet, more and more people are writing textual reviews on e-commerce platforms, leading to a proliferation of online reviews. Prior to purchasing products, consumers usually read online reviews from previous buyers, demonstrating their trust in such kinds of user-generated content (Archak et al. 2011; Chen and Xie 2008). Online reviews provide detailed information about products and help customers better understand product quality and functions (Sun 2012). Thus, the valuable information provided by online reviews plays an important role in the purchase decision-making of potential consumers (Dhar and Chang 2009).

Resource-matching theory explains that the balance between the cognitive resources available to process the information and the mental resources required for the task is of huge importance in decision-making. When current cognitive resources and mental

resources required are matched, the decision-making processing is more efficient and effective; otherwise decision performance is low (Anand et al. 1989; Mantel and Kellaris 2003). Nevertheless, on a popular e-commerce website, the online reviews of a product can easily exceed hundreds or even thousands (Hu and Liu 2004; Park and Lee 2009), leading to information overload (Bawden and Robinson 2009). Due to the limitations in customers' browsing time and the small screens of smart phones widely used in consumers' online shopping (the mental resources), it is impossible for consumers to read all related reviews about a particular product (the cognitive resources). This motivated firms to design an IT artifact to balance mental and cognitive resources, by providing consumers with a subset of reviews (e.g., the top 5 or 10 results) that reflect various aspects of the original large-scale review collection.

In recent years, review selection has become an attractive business intelligence and analytics research direction that aims to address the information overload problem of online reviews. In particular, some researchers have attempted to select a comprehensive subset of reviews, which would fully cover all opinions appearing in a review corpus (Tsaparas et al. 2011). Moreover, Lappas et al. (2012) examined the consistency of opinion proportions between the selected review subset and the original review corpus. E-commerce platforms like Amazon rank reviews according to their helpfulness votes or product ratings to provide consumers with several high-ranked helpful reviews. However, most research and e-commerce platforms rarely consider the diversity of selected subsets. As such, diversified review subset selection has not been well addressed. Furthermore, previous research simply assumes that the extracted features are of the same semantic level, which cannot reflect the multi-level features of certain products.

To overcome the limitations of previous studies, this research focuses on addressing the diversified review subset selection problem with the constraint of high-quality for selected reviews, where the semantic hierarchy of features has been properly considered. In particular, we apply the LDA topic model and the Word2vec model to measure the topic relation and context relation between each pair of features and then construct a feature taxonomy. Based on the established feature taxonomy, we calculate the similarity between each pair of reviews and then propose a diversified review selection approach similar to the simulated annealing method. Through extensive experiments on real data, we demonstrate that the proposed approach can select a diversified review subset, and provide concise, high-quality information for consumers in their decision-making.

From the perspective of design science research, this work can be positioned in the "improvement" quadrant of the design science knowledge contribution framework (Gregor and Hevner 2013). The review selection problem is a known research field which has been recognized by existing studies in e-commerce. To further improve the solutions to the problem, this study develops a new formulation of the online review selection problem taking into consideration both the diversity and quality of selected review subsets, and proposes a heuristic approach to effectively solve the problem. In the proposed approach, the multi-level features of certain products are well reflected by the feature taxonomy constructed based on the topic relation and content relation between features. To achieve a rigorous design evaluation, experimental analyses on real-world data collected

from Tmall and Amazon have been conducted to demonstrate the performance advantages of the proposed approach over other approaches. In this way, the contributions of this study can be summarized as follows:

- (1) A diversified review selection problem considering both diversity and quality of selected review subsets is formulated in this study. In order to address this problem, a heuristic diversified review subset selection approach that can achieve high diversity and high quality in the selected results is proposed.
- (2) To reflect the multi-level features of certain products, a feature taxonomy construction method taking into consideration both topic relation and content relation is explored, based on which the diversity of reviews is measured.

The rest of the paper is organized as follows. Section 2 discusses related work on review selection, review helpfulness and quality prediction, and ontology learning. Section 3 defines the proposed diversified review subset selection problem and provides an example to illustrate it. Section 4 proposes the feature taxonomy construction method and the diversified review subset selection approach, i.e., SA-DRS. Section 5 presents the experimental results on real data. Section 6 concludes the entire work and highlights some future research directions.

Literature review

This study has ties with existing research in the domains of online review selection, review helpfulness and quality prediction, and ontology learning. In this section, we will provide a comprehensive review of studies in these three fields.

Review selection

The early work on review selection proposed by Hu and Liu (2004) tried to address the review selection problem by creating a statistical summary of useful information such as features and opinions from the original review corpus. They managed to mine the product features and corresponding opinions expressed by consumers. Tsaparas et al. (2011) proposed a greedy method to select a comprehensive subset of high-quality reviews that covered many different aspects of the reviewed item. Yu et al. (2013) clustered reviews based upon their opinions and selected reviews proportionally from different clusters.

Some other researchers focused on opinion distribution when selecting review subsets. Lappas et al. (2012) proposed an approach to select no more than k reviews that can emulate the opinion distribution of the original review collection as accurately as possible. Thus, the selected reviews could reflect the proportion of both positive and negative opinions for each feature in the original review corpus. Tian et al. (2015) proposed a greedy review selection approach, which could not only choose the individual review with high comprehensiveness, but also ensure the extracted reviews as a whole to reflect the opinion distribution of the review corpus. Jian et al. (2017) used a clustering algorithm to identify review sentences regarding the same features for a list of products. Then, they aimed to select representative review sentences from each product for multiple-to-multiple comparisons.

In summary, the problem of diversified review subset selection has not been well addressed. Moreover, the assumption in previous studies that extracted features of the same semantic level is not appropriate in solving the problem of review selection. Thus, it is desirable to completely address the diversified review selection problem and meanwhile take the semantic relations of reviews into consideration, which highly motivates this study.

Review helpfulness and quality prediction

The early work on review helpfulness prediction was done by Kim et al. (2006). They evaluated the impacts of many types of variables on the helpfulness of reviews, and found that review length, product rating and the unigram TF-IDF scores of each word were the most useful attributes for predicting helpfulness. Moreover, Liu et al. (2007) found that the number of product features mentioned in a review was one of the most important indicators of the review quality. Liu et al. (2008) showed that the helpfulness of a review depended on three important factors: reviewer expertise, writing style, and review timeliness. Then they proposed a regression model to predict the quality of a review. Ghose and Ipeirotis (2011) examined the impact of different factors, such as subjectivity levels, readability, and spelling errors on perceived usefulness of reviews. Chen and Tseng (2011) extracted four review characteristics—reputation of authors, review completeness, ease of understanding, and concise representation—and proved that those four characteristics could assist users in evaluating the quality of information in product reviews. Recently, Paul et al. (2017) used review word embedding and word position information as inputs to a dynamic convolutional neural network to estimate review quality.

Ontology learning

Maedche and Staab (2001) proposed an ontology learning framework for ontology extraction, merging, and management. They employed data mining approaches including a hierarchical clustering method and some background knowledge to learn concepts, hierarchical relations, and associative relations from texts. Tang et al. (2009) used a three-stage probabilistic model to define four divergence measures to evaluate the relations between tags, based on which the ontological hierarchy of social tags can be learned. Djuana et al. (2012) also contributed to this research area by proposing a tag ontology from the folksonomy based on WordNet. They performed user clustering based on user profiles and then generated personalized tag ontologies for each user cluster. Recently, Tian et al. (2014) presented a method that identified useful association rules between features, based on which a feature taxonomy structure could be constructed by applying the LDA model. The feature relations captured by the feature taxonomy provided more detailed information about products, which was deemed as one step further towards profiling products from a multi-level perspective.

However, to the best of our knowledge, none of the previous research considered the context relation between extracted features in constructing the feature taxonomy, leading to the deviation between the established semantic relation and human cognition.

Thus, our paper aims to explore the semantic relations of features according to the topic relation and word context relation simultaneously.

Problem definition

This section formally defines the diversified review subset selection problem and takes an example to illustrate the proposed review selection problem.

The diversified review selection problem

Let $R = (r_1, r_2, \dots, r_n)$ denote a set of n reviews with respect to a certain product, and r_i represent the i th review in R . Let $F = (f_1, f_2, \dots, f_m)$ denote a set of m qualified features commented by reviews in R .

Given the review set R and the feature set F , we use matrix $A = (a_{ij})_{n \times m}$ to record the features contained in each review. The element in the i th row and j th column in matrix A is defined as Equation (1):

$$a_{ij} = \begin{cases} 1, & \text{review } r_i \text{ covers feature } f_j; \\ 0, & \text{review } r_i \text{ not covers feature } f_j. \end{cases} \tag{1}$$

a_{ij} represents the (i, j) th element in matrix A . When $a_{ij} = 1$, it implies the j th feature appears in the i th review. As shown in Equation (2), we use x_i to indicate whether review r_i is included in the selected review set:

$$x_i = \begin{cases} 1, & \text{if review } r_i \text{ is selected;} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

When $x_i = 1$, it implies the i th review is included in the selected review subset. We then use $X = (x_1, x_2, \dots, x_n)$ to represent the selected subset for the review selection problem. The typical constraint of review selection is that the review subset S has k reviews, which means $\|X\| = k$. In this study, we aim to select a high-quality review subset with diversified content. In other words, the similarity between each pair of selected reviews is supposed to be relatively low and the quality of selected reviews should be relatively high. We use $SR = (sr_1, sr_2, \dots, sr_k)$ to represent the selected review subset with k reviews. Let $q = (q_1, q_2, \dots, q_n)$ denote the quality of each review in R , so the quality of the selected review subset can be calculated as $\sum q_i \quad i = 1 \dots k$, which is simplified as qX . Meanwhile, let $\text{sim}(sr_i, sr_j)$ denote the similarity between selected review sr_i and sr_j . There are C_k^2 pairs of $\text{sim}(sr_i, sr_j)$ values in the selected review subset. Thus, the objective function of the diversified review selection problem is defined in Equation (3):

$$\begin{aligned} \max \quad & qX \times \frac{C_k^2}{\sum_{i=1}^{k-1} \sum_{j=i+1}^k \text{sim}(sr_i, sr_j)}, \\ \text{s.t.} \quad & \|X\| = k, \\ & x_i \text{ is binary.} \end{aligned} \tag{3}$$

In Equation (3), qX represents the quality of the selected review subset. When the value of qX becomes larger, the objective function value becomes larger. Moreover, $C_k^2 / \sum_{i=1}^{k-1} \sum_{j=i+1}^k \text{sim}(sr_i, sr_j)$ represents the reciprocal of the average similarity

within the selected review subset. When the average similarity within the review subset becomes smaller, the objective function value accordingly becomes larger.

An illustrative example

Example 1

Given a review set R with six reviews, $R = (r_1, r_2, r_3, r_4, r_5, r_6)$. The similarity matrix of the six reviews is shown as follows:

$$\begin{bmatrix} 1 & 0.95 & 0.03 & 0.05 & 0.12 & 0.21 \\ 0.95 & 1 & 0.13 & 0.08 & 0.15 & 0.01 \\ 0.03 & 0.13 & 1 & 0.87 & 0.92 & 0.78 \\ 0.05 & 0.08 & 0.87 & 1 & 0.85 & 0.95 \\ 0.12 & 0.15 & 0.92 & 0.85 & 1 & 0.77 \\ 0.21 & 0.01 & 0.78 & 0.95 & 0.77 & 1 \end{bmatrix}.$$

Without loss of generality, we assume that the quality vector $q = (0.7, 0.8, 0.8, 0.6, 0.5, 0.4)$. If the review subset is supposed to contain 3 reviews, we can calculate the objective function value shown as Equation (3) for each possible selection result, which is shown in Table 1.

In Table 1, the value of the review subset $\{r_1, r_3, r_4\}$ is the largest. Reviews r_3, r_4 increase the quality of the selected review subset. Moreover, $sim(r_1, r_4)$ and $sim(r_1, r_3)$ are pretty small, and improve the diversity of the review subset. Thus, according to Equation (3), the review subset $\{r_1, r_3, r_4\}$ with diversified content and high quality is the desirable subset which we aim to select in this study.

The proposed approach

This section introduces the proposed review selection approach. A feature taxonomy was constructed by applying the LDA topic model and the Word2vec model. Based on the feature taxonomy, similarities between features, opinions, and reviews were calculated and the overall quality of individual online reviews was

Table 1 Objective function values of review subsets

S	Value	S	Value
$\{r_1, r_2, r_3\}$	6.2162	$\{r_2, r_3, r_4\}$	6.1111
$\{r_1, r_2, r_4\}$	5.8333	$\{r_2, r_3, r_5\}$	5.2500
$\{r_1, r_2, r_5\}$	4.9180	$\{r_2, r_3, r_6\}$	6.5217
$\{r_1, r_2, r_6\}$	4.8718	$\{r_2, r_4, r_5\}$	5.2778
$\{r_1, r_3, r_4\}$	6.6316	$\{r_2, r_4, r_6\}$	5.1923
$\{r_1, r_3, r_5\}$	5.6075	$\{r_2, r_5, r_6\}$	5.4839
$\{r_1, r_3, r_6\}$	5.5882	$\{r_3, r_4, r_5\}$	2.1591
$\{r_1, r_4, r_5\}$	5.2941	$\{r_3, r_4, r_6\}$	2.0769
$\{r_1, r_4, r_6\}$	4.2149	$\{r_3, r_5, r_6\}$	2.0648
$\{r_1, r_5, r_6\}$	4.3636	$\{r_4, r_5, r_6\}$	1.7510

estimated. Thereafter, based on the calculated similarities and review quality, reviews can be heuristically selected by the proposed approach, whose framework is shown in Fig. 1.

Feature taxonomy construction

When commenting on a product, some consumers may discuss various features (e.g., price performance and appearance, each of which represents a key feature of a mobile phone), while some others prefer to pay close attention to one feature from different detailed angles (e.g., duration, charging time and battery life, which focus on the feature “battery” of a phone). Thus, a hierarchical semantic structure is required to reveal and differentiate the relations among all the features of one product. Without considering the semantic structure among the features of certain products, the extracted features at the same semantic level may result in semantic overlap and informative duplication within the selected subset, which affects the cognition of consumers. On the other hand, the feature taxonomy construction can provide the semantic information of features at different levels, which is in line with the target of this study. Although Tian et al. (2014) proposed a feature taxonomy construction method based on association rules, they neglected the word relations in various contexts which contains more semantic information. LDA uses global documents and words to calculate the topic distributions of a document and word distributions on each topic correspondingly, while Word2vec relies on local words (context window) to capture the context relations. The integration of these two methods could effectively improve the quality of the feature similarity measurement for constructing a feature taxonomy that is fit for consumer cognition. Therefore, meanwhile, we propose a method considering the global level and local level semantic information based upon the topic relations and word context relations.

As for topic relations, we employ the LDA topic model to calculate the relations between features in different topics. Blei et al. (2003) proposed the LDA topic model, which is a three-level hierarchical Bayesian model, including document level, topic level and word level. We assume that the extracted features are attributed to word level. Let

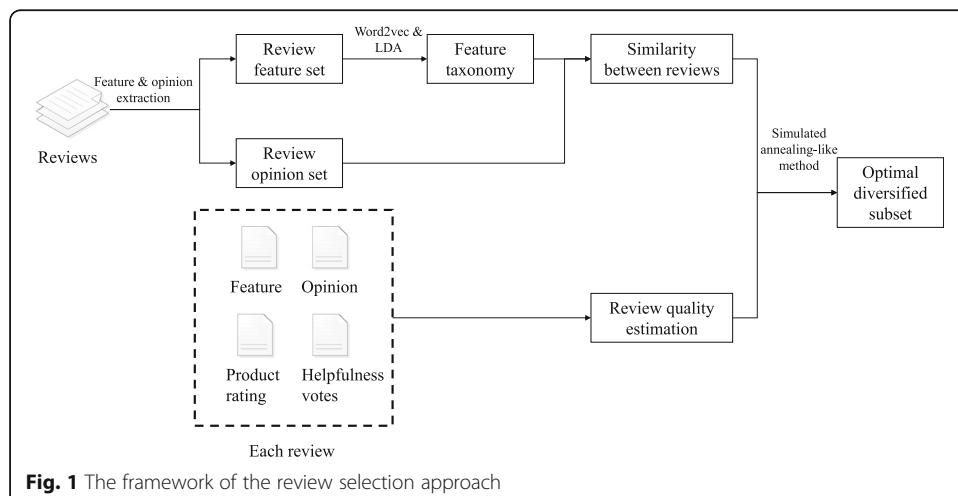


Fig. 1 The framework of the review selection approach

$F = (f_1, f_2, \dots, f_m)$ be a set of extracted features appearing in a review corpus, and $Z = (Z_1, Z_2, \dots, Z_h)$ be a set of hidden topics. LDA is applied to generate topic models to represent the review collection as a whole. At a review collection level, each topic Z_t is represented by a probability distribution over features, $P(f_j|Z_t)$ is the probability of feature f_j appearing in reviews on topic Z_t . Based on the probability $P(f_j|Z_t)$, we can choose the top features to represent the specific topic Z_t .

Definition 1 (Topic features): Let $\phi_k = (P(f_1|Z_t), P(f_2|Z_t), \dots, P(f_m|Z_t))$ be the topic representation for topic Z_t calculated by LDA and $0 \leq \sigma \leq 1$ be a threshold. The topic features for Z_t , denoted as $TF(Z_t)$ is defined as $TF(Z_t) = \{f_j | P(f_j|Z_t) > \sigma, f_j \in F\}$.

Topic features for specific topic Z_t are the features whose appearance probabilities are larger than the threshold σ . Then we define single topic relation to measure the topic relation of two features.

Definition 2 (Single topic relation): For $f_i, f_j \in F$, the topic relation between two features with respect to a certain topic Z_t is defined as Equation (4):

$$STR_{Z_t}(f_i, f_j) = \begin{cases} 1 - |P(f_i|Z_t) - P(f_j|Z_t)|, & f_i, f_j \in TF(Z_t); \\ 0, & \text{otherwise;} \end{cases} \tag{4}$$

where $|P(f_i|Z_t) - P(f_j|Z_t)|$ is the absolute value of the difference between appearing probabilities of feature f_i and feature f_j . A single topic relation reflects the semantic relation of two features on one particular topic. When feature f_i and feature f_j are both topic features of hidden topic Z_t and $P(f_i|Z_t)$ is near to $P(f_j|Z_t)$, the value of $STR_{Z_t}(f_i, f_j)$ is near to 1, so the two features share similar semantic meaning on this single topic. Then we can calculate the topic relation between two features on all topics.

Definition 3 (Topic relation): Let $f_i, f_j \in F$ be two features appearing in review corpora, and $Z(f_i, f_j)$ be a set of topics that contain both features. The topic relation between two features with respect to all topics is defined in Equation (5):

$$TR(f_i, f_j) = \frac{\sum_{Z_k \in Z(f_i, f_j)} STR_{Z_k}(f_i, f_j)}{|Z(f_i, f_j)|}, \tag{5}$$

where $|Z(f_i, f_j)|$ represents the number of topics containing both feature f_i and feature f_j . If the topic relation between two features is large, two features share similar semantic meaning across all topics.

Moreover, as for word context relations, we utilize the Word2vec model to train word vectors, then calculate word distances. Mikolov et al. (2013) proposed the Word2vec model, which could train high-quality word vectors with a much lower computational complexity. After the full training of Word2vec, we can acquire the vector representation of feature f_j as $v_j = (v_{j1}, v_{j2}, \dots, v_{jd})$. So, we can calculate the cosine distance of features, which is deemed as the word context relation.

Definition 4 (Word context relation): Let $f_i, f_j \in F$ be two features appearing in review corpus, v_i and v_j are the vector representations of feature f_i and feature f_j . The word context relation is defined as Equation (6):

$$WCR(f_i, f_j) = \frac{v_i \cdot v_j}{\|v_i\| \times \|v_j\|}, \tag{6}$$

where $v^i \cdot v^j$ equals to the inner product of vector representations of feature f^i and feature f^j . $\|v^i\|$ represents the Euclidean norm of v^i .

Based on the topic relations and word context relations, the feature taxonomy can be constructed. First, we set some constraints for the feature taxonomy:

- (1) The root of the taxonomy represents the feature with the largest appearance probability in reviews. The root appears most frequently in review corpora.
- (2) The taxonomy is structured as a tree, which means that each feature has only one parent feature except for the root.
- (3) A parent feature is more general than its sub features. The proportion of a parent feature is larger than its sub features.

The input of the proposed feature taxonomy construction method is the review matrix A , feature set $F = (f_1, f_2, \dots, f_m)$, topic relation $TR(f_i, f_j)$, $i, j = 1 \dots m, i \neq j$, word context relation $WCR(f_i, f_j)$, $i, j = 1 \dots m, i \neq j$ and proportion vector $\theta = (\theta_1, \theta_2, \dots, \theta_m)$ whose element θ_j equals to the number of appearances of feature f_j , $j = 1 \dots m$ divided by the number of reviews. First, the feature taxonomy is an empty tree. Then we sort the proportion vector at a descending sequence $\theta' = (\theta'_1, \theta'_2, \dots, \theta'_m)$, whose corresponding feature is $f' = (f'_1, f'_2, \dots, f'_m)$. We add the feature f'_1 to the root of the feature taxonomy, which is named as the function **addroot**. Second, the feature f'_2 is directly added to the feature taxonomy as the sub feature of f'_1 , which is named as the function **addsubfeature**. Third, we process features according to the descending sequence in proportion vector θ , and add feature f'_i as the sub feature of f'_j only if f'_i satisfies that $TR(f'_i, f'_j) + WCR(f'_i, f'_j)$ is the largest. The method proceeds until there is no feature in θ . The pseudo code of the feature taxonomy construction method is shown in Table 5 Appendix.

Review quality estimation

In order to estimate the quality of a review, according to Chen and Tseng (2011) and Jindal and Liu (2008), we adopt a quality measurement capturing four dimensions of a review: completeness, objectivity, believability and deviation, which can be assessed from its content, sentiment and feedback information. For each review r_i , we extract the number of features to reflect its completeness. The objectivity is measured by the number of opinion words o_i . In light of previous literature (Ghose and Ipeirotis 2011; Korfiatis et al. 2012; Lee 2018; Tian et al. 2015), the helpfulness vote means that people find the review helpful, which is usually used to fully or partly represent the believability of a review as proxy. Thus, the believability of a review is estimated by its helpfulness vote s_i . The deviation of a review is calculated as the difference between the product rating of the review and the overall average rating, denoted as d_i . Each dimension value is normalized respectively, which is shown in Equation (7):

$$scl(x_i) = \frac{x_i - \min(x)}{\max(x) - \min(x)}. \tag{7}$$

Afterwards, we use the average of all the normalized values to measure the overall quality of a review, as shown in Equation (8). It is worth noting that other quality measurements can also be adopted according to actual requirements:

$$q_i = \frac{scl(f_i) + scl(o_i) + scl(s_i) + scl(d_i)}{4}. \tag{8}$$

Review selection approach

The similarity between two extracted features can be defined based on the established feature taxonomy. Let $f_i, f_j \in F$ be two features appearing in the review corpus, H be the height of the taxonomy and $d(f_i, f_j)$ be the shortest path length in the taxonomy. The similarity between features is defined as Equation (9):

$$sim(f_i, f_j) = \begin{cases} -\log \frac{d(f_i, f_j)}{2H}, & i \neq j; \\ 1, & i = j. \end{cases} \tag{9}$$

In the review selection problem, two typical sentiment polarities, positive and negative, are considered for each feature (Hu and Liu 2004; Tsaparas et al. 2011). Therefore, there are at most $2m$ opinions collectively. Let $O = (o_1, o_2, \dots, o_p)$, $p = 2m$, be the complete set of opinions for all the features. Considering the semantic polarity, we define the similarity of two opinions in the review corpus. Let $o_i, o_j \in O$ be two corresponding opinions of feature f_i and f_j , the similarity between two opinions is defined in Equation (10):

$$sim(o_i, o_j) = sim(f_i, f_j) \times \begin{cases} 1, & o_i, o_j \text{ have the same polarity;} \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Let $r_i, r_j \in R$ be two reviews in the review corpus. Review r_i has m opinions, whose opinion vector is $\vec{o}_i = (o_{i1}, o_{i2}, \dots, o_{im})$ and r_j has n opinions with the opinion vector $\vec{o}_j = (o_{j1}, o_{j2}, \dots, o_{jn})$. The similarity between two reviews is defined as Equation (11):

$$sim(r_i, r_j) = \frac{\sum_{t=1}^m \sum_{p=1}^n sim(o_{it}, o_{jp})}{\|\vec{o}_i\| \times \|\vec{o}_j\|}, \tag{11}$$

where the numerator represents the sum of similarities between opinions appearing in review r_i and r_j . The denominator equals to the product of the number of opinions in review r_i and r_j . Thus, the similarity matrix of the review corpus $S = (sim_{ij})_{n \times n}$, which records the similarity between each pair of reviews, can be calculated. The similarity matrix is a symmetric matrix, whose diagonal elements are 1.

The proposed review selection approach aims to select a review subset whose reviews are high-quality and meanwhile diversified in term of low similarities. In light of previous literature on review selection, we assume that the selected subset has k reviews. The aggregated similarity of reviews in the selected subset is $\sum_{i=1}^{k-1} \sum_{j=i+1}^k sim(sr_i, sr_j) / C_k^2$, where sr_i, sr_j are selected reviews in subset. Therefore, the proposed review selection problem can also be formulated as Equation (3).

Because the proposed review selection problem is a NP-complete, a simulated annealing-like method is proposed to solve this single objective optimization problem. Simulated annealing is a compact and robust algorithm that provides excellent solutions to single and multiple objective optimization problems with a great reduction in computation time (Suman and Kumar 2006). It is a kind of stochastic search algorithm

based on Monte-Carlo iterations, which is inspired by heating and controlled cooling of a material. The algorithm normally begins from a very high initial cooling temperature, changes the initial value of variables and then gets a new solution. If the objective function value of the new solution becomes better, it will be kept unconditionally. If the objective function value of the new solution becomes worse, it will be kept with a probability as shown in Equation (12):

$$p = e^{\frac{\delta}{T}}, \quad (12)$$

where δ is the difference of the objective function value on two consecutive iterations and T is the current temperature. Finally, this algorithm avoids local convergence and reaches global optimization by making use of the stochastic search strategy as the temperature drops.

The input of the proposed approach SA-DRS is similarity matrix $S = (sim_{ij})_{n \times m}$, quality vector $q = (q_1, q_2, \dots, q_n)$, initial vector X_0 , the lowest temperature T_{min} , initial temperature T_0 , and cooling parameter ε . In the initialization stage, the temperature is pretty high and we only calculate the aggregated similarities for the combination of non-zero elements in the X by the function of *combination* and *nonzero*. By doing so, we calculate the initial objective function value $V(X_0)$. Then, we can generate a new vector X_{N+1} from the current state X_N . The function *randomint* is to generate an integer randomly. After that, the difference of objective values between the new solution and the current state is calculated. If the difference $\delta \geq 0$, the new solution can achieve a larger objective function value and could be accepted as the next state. Otherwise, rather than reject the new solution directly, it is accepted with a certain probability (*acceptProb*). The acceptance probability is determined by the difference δ and the current temperature T . As the iteration continues, the temperature drops quickly and thus *acceptProb* becomes smaller, which means it is impossible to accept the new solution.

At the end, to avoid the best solution, the best solution with the largest objective function value is obtained as the final output, rather than the result of the last iteration. Thus, the final output is the optimal review vector X . The pseudo code of SA-DRS approach is shown in Table 6 Appendix.

To further illustrate the performance of SA-DRS, it is leveraged to find the optimal subset for Example 1. We set the initial temperature $T_0 = 10,000$, the cooling parameter $\varepsilon = 0.999$, the initial vector $X_0 = [1, 1, 1, 0, 0, 0]$, and the lowest temperature $T_{min} = 0.001$. After running more than 30,000 iterations, the approach ends and we acquire the largest objective value 6.6316. And the 1st, 3rd, 4th review subsets are selected, which are the same as the results in Section 3. It shows that SA-DRS could find the subset with the largest diversity. The same parameter setting is also used in the following experiments with real-world data.

Experiments

In this section, extensive comparative experiments are conducted to evaluate the effectiveness of SA-DRS. The experiments are carried out using real online review data collected from two typical e-commerce platform Tmall (<https://www.tmall.com>) with Chinese reviews and Amazon (<https://www.amazon.com>) with English reviews. With two experiments based on different languages from two of the largest e-commerce platforms, it can

demonstrate that the proposed approach is robust and can be extended across different languages. The experiments on Tmall are conducted by using a collection of six different product categories including Books, Digital Cameras, Laptops, Mobile Phones, Diapers, and Shirts, and the experiments on Amazon are conducted by randomly selecting 89 products which covers most categories of commodities on Amazon. The details about the experiment datasets are given in Table 2.

Feature taxonomy results

Online reviews used in the experiments are first structured by feature extraction and sentiment analysis. In this paper, feature extraction is implemented by the method proposed by Hu and Liu (2004), and sentiment analysis is implemented by the technique proposed by Ding et al. (2008). By applying these methods, we extract the structured features and opinions from the review corpus. The average number of features in Tmall reviews is 12.3, and 18.2 in Amazon reviews. Moreover, we apply the proposed Feature Taxonomy Construction Method to explore the hierarchical relationship of extracted features and establish different feature taxonomies for different categories. For example, the mobile phone feature taxonomy is shown in Fig. 2.

From the feature taxonomy above, we can find that the root feature is quality which is pretty general in online reviews. It has several second-level features, including speed, price, etc., which reveals more detailed information than quality. Moreover, as for the third level of the feature taxonomy, the features battery, sound and camera describe pretty specific information about performance, so they become the second-level features of performance. The feature taxonomy is consistent with our semantic cognition. Based on the feature taxonomy, we can select a diversified and high-quality review subset which is demonstrated in the following sections.

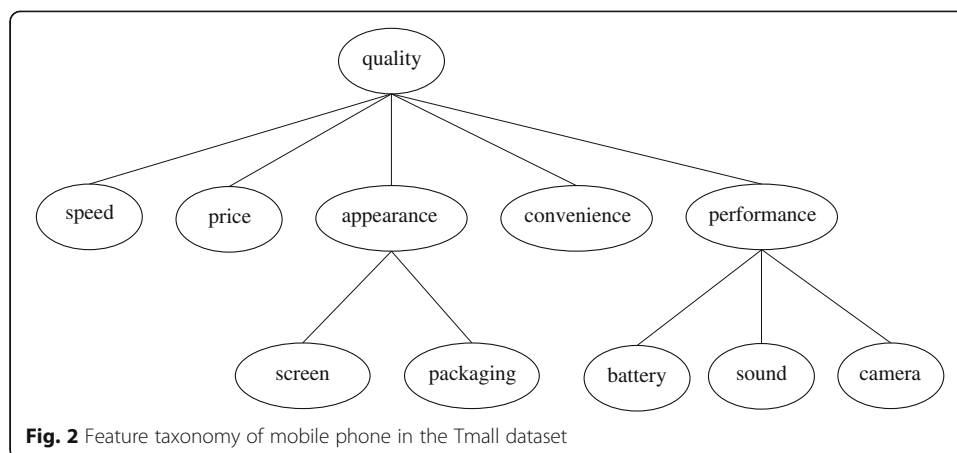
Effectiveness experiment on Tmall and Amazon datasets

In this section, we compare SA-DRS with several state-of-the-art benchmark approaches of review selection listed below:

- (1) Greedy-U & Greedy-Q: Tsaparas et al. (2011) proposed several greedy approaches for review selection. Greedy-U and Greedy-Q are selected for comparison. Greedy-U aims to cover as many features as possible, while Greedy-Q aims to select reviews with high quality.

Table 2 Experiment datasets

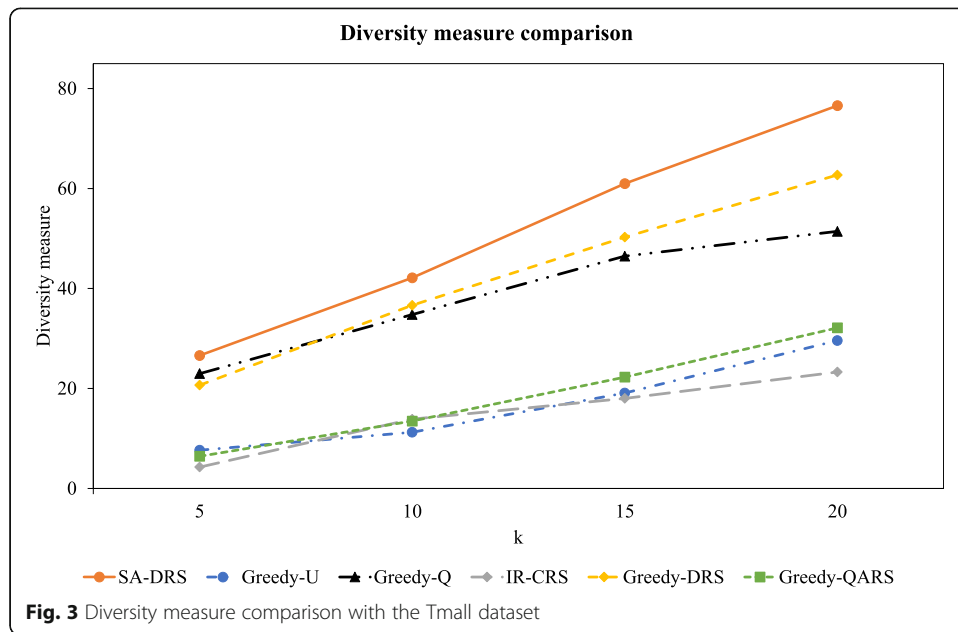
Platform	Category	Product	# of Average reviews
Tmall	Books	61	672.0
	Digital cameras	53	180.3
	Laptops	60	1110.3
	Mobile phones	61	2507.6
	Diapers	59	4745.2
	Shirts	70	596.1
Amazon	Random	89	306.5



- (2) IR-CRS: Lappas et al. (2012) proposed an integer-regression-based approach to select a consistent review subset with k reviews, denoted as IR-CRS.
- (3) Greedy-DRS: Yu et al. (2013) employed a greedy approach to cover important features, denoted as Greedy-DRS.
- (4) Greedy-QARS: Tian et al. (2015) proposed a greedy approach to select a consistent review subset based on a feature taxonomy, denoted as Greedy-QARS.

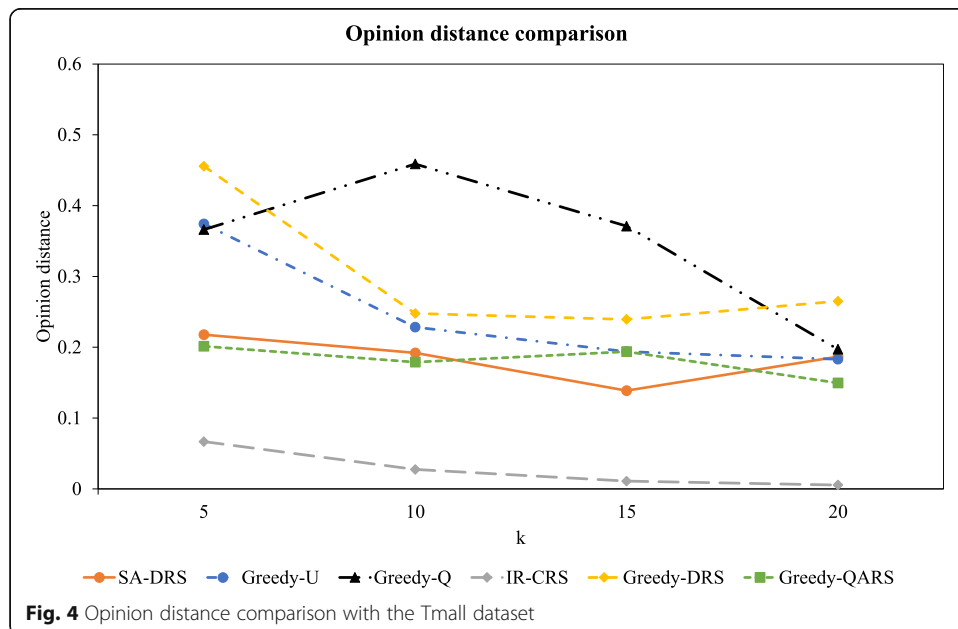
As for evaluation metrics, we focus on the three important aspects of the performance of review subsets: diversity measure formulated as Equation (3), coverage measure proposed by Tsaparas et al. (2011) and opinion distance measure proposed by Lappas et al. (2012), which reveal three traits of review subsets: diversity, coverage and consistency. We calculate the average value of these measures respectively. In the first experiment part, we evaluate the effectiveness of six approaches with the Tmall dataset by the metrics of diversity which is shown in Fig. 3.

In Fig. 3, we see that SA-DRS always performs best on this metric and it achieves the largest diversity values all the time. The results demonstrate that SA-DRS can select a high-quality and diversified subset, which will assist potential consumers to make purchase decisions. In contrast, Greedy-DRS and Greedy-Q perform worse than SA-DRS, because neither of them considers the quality and diversity in selecting review subsets. Although, Greedy-QARS considers the quality of reviews to some extent, it is mainly designed to optimize the consistent object similar to the opinion distance measure proposed by Lappas et al. (2012). Moreover, Greedy-QARS is a two-stage optimization approach which first selects a review subset with high consistent scores and then broadens the restriction of the candidate reviews to choose the review with high comprehension value. The two-stage optimization process limits the optimization effect of Greedy-QARS, while the proposed SA-DRS can optimize the object in terms of high quality and high diversity simultaneously. As for the other two approaches, they have a comparatively bad performance because they only aim to select a comprehensive or consistent review subset.



Then we compare the effectiveness on the metric of consistency, which is also called the opinion distance measure. The consistency measures the distance between the opinion proportion in the review corpus and that in the review subset. The corresponding results are shown in Fig. 4.

From the results shown in Fig. 4, we find that IR-CRS and Greedy-QARS rank first and second on the opinion distance measure, since they are designed just to optimize the consistency between opinions of the review corpus and the review subset. Moreover, SA-DRS also performs pretty well on this measure because it is proposed to select diversified reviews, which leads to the consistent opinions



between the review corpus and the review subset to a large extent. However, SA-DRS is not directly designed to optimize the object of consistency and it indirectly reaches a satisfactory consistent value by selecting a diversified subset of reviews, which could be seen as a limitation of the proposed approach when users only want to read a subset with consistent opinions. Moreover, Greedy-DRS, Greedy-U and Greedy-Q have fairly poor performance because none of these approaches take consistency into consideration.

Furthermore, we compare the coverage measure of review subsets and the results are shown in Fig. 5.

According to Fig. 5, Greedy-U performs better than the other approaches in terms of coverage, since it directly aims to optimize the coverage object. In a general way, a high-quality review often covers the great majority of important features of the product. Thus, the proposed approach SA-DRS and Greedy-Q also perform pretty well (but not the best) on the measure of coverage, since they are designed to select a subset of high-quality reviews. As for IR-CRS and Greedy-QARS, they perform the worst because they neglect the object of coverage. Although SA-DRS can select a subset with a satisfactory coverage value, it definitely does not perform the best in terms of coverage, which could also be viewed as a limitation under the circumstances that users just want to read a comprehensive subset of reviews.

In order to clearly illustrate the results, we rank their performance on the three measures above. The ranking comparison of the six approaches is shown in Table 3.

According to the results shown in Table 3, it can be seen that the proposed approach ranks first in the diversity measure, third in the opinion distance measure and third in the coverage measure. On average, SA-DRS has the highest rank among six approaches, demonstrating that SA-DRS can not only select a high-quality, diversified review subset, but also is effective on the coverage measure and consistency measure.

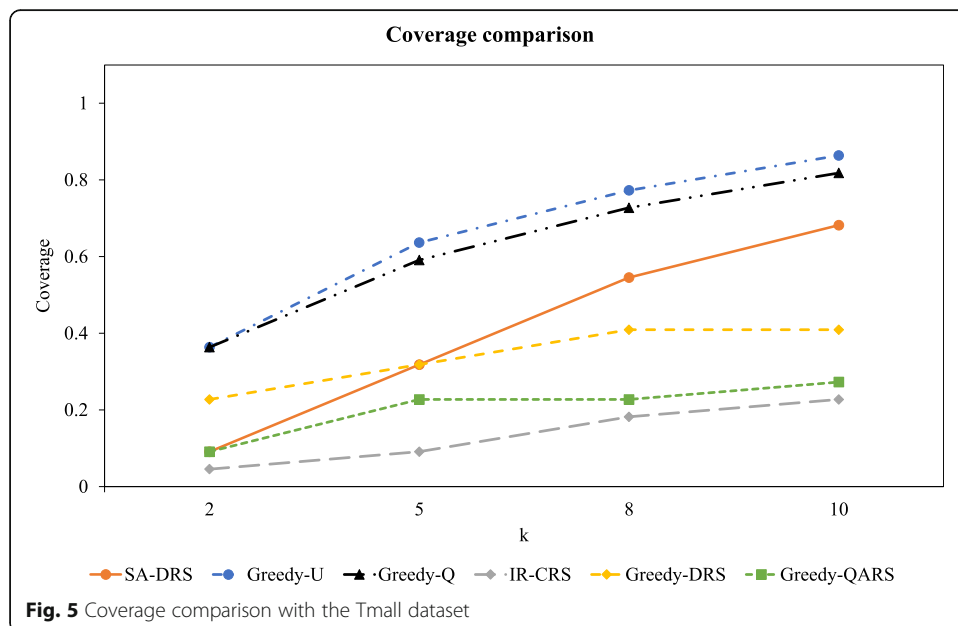


Table 3 Ranking comparison of the six approaches with the Tmall dataset

Approach	Diversity rank	Opinion distance rank	Coverage rank	Average rank
SA-DRS	1	3	3	2.3
Greedy-U	6	4	1	3.7
Greedy-Q	3	6	2	3.7
IR-CRS	4	1	6	3.7
Greedy-DRS	2	5	4	3.7
Greedy-QARS	5	2	5	4.0

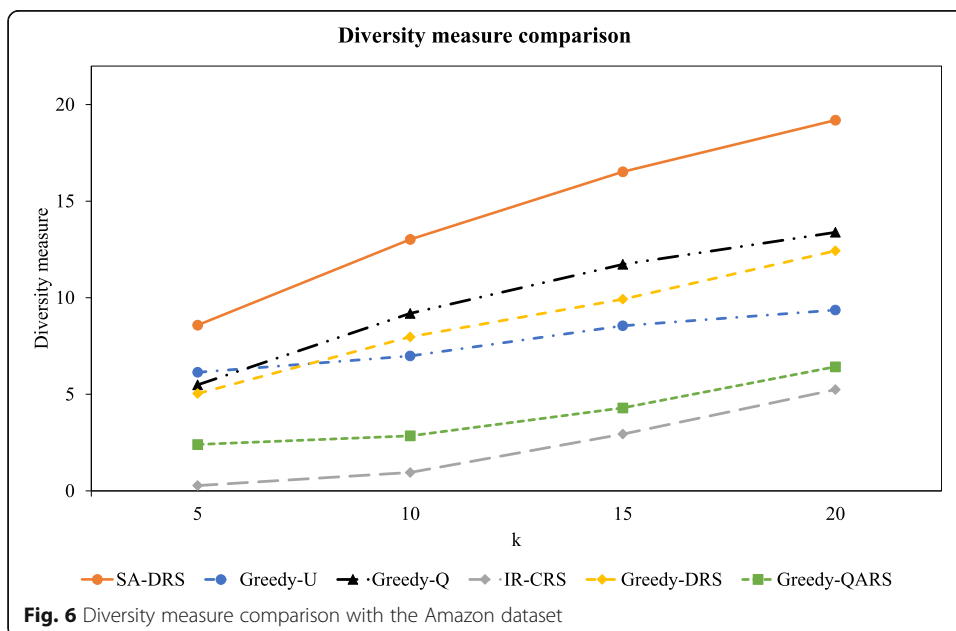
To further examine whether different structures and expressions of languages make a difference, we also utilize online English reviews on the e-commerce platform of Amazon to conduct experiments comparing SA-DRS with the other five approaches on those three measures.

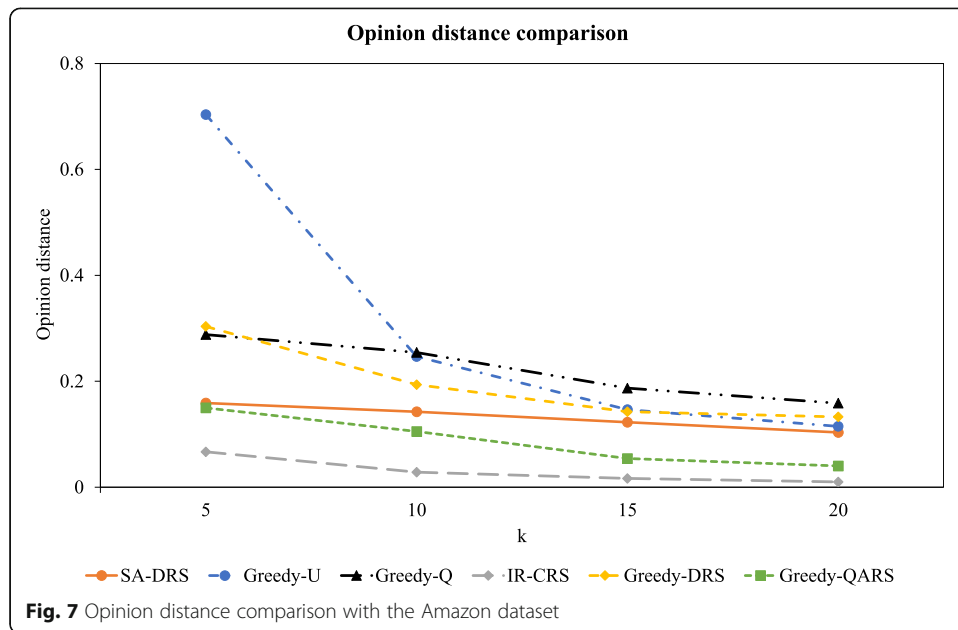
As for the results of diversity which is shown in Fig. 6, we find that SA-DRS ranks first on this measure among the six approaches, which confirms that the proposed approach could select a high-quality and diversified review subset no matter which language the platform uses.

From the results shown in Fig. 7, IR-CRS and Greedy-QARS are also the best two approaches on this measure, which is consistent with the intuition. SA-DRS also ranks third on this measure, which shows that the selected review subsets of SA-DRS can also reflect the original distribution of opinions in the review corpus to some extent.

According to Fig. 8, Greedy-U and Greedy-Q can select a review subset with high coverage. SA-DRS is also effective on coverage for English reviews.

In order to compare these six approaches more clearly, we also rank their performance on the three measures above. The ranking comparison of the six approaches with the Amazon dataset is shown in Table 4.





According to Table 4, we find that SA-DRS has the highest rank among six approaches, which is the same as the result for the Tmall dataset. It also demonstrates that SA-DRS performs well for both English and Chinese reviews. In addition to the numerical experiments above, we also provide a more concrete example with respect to a set of mobile phone reviews to further illustrate the effectiveness of the proposed approach. The example shows the same conclusion. Due to space limitations, please refer to the appendix of this paper.

In summary, we come to three conclusions as follows:

- (1) SA-DRS has the best performance on the diversity metric.

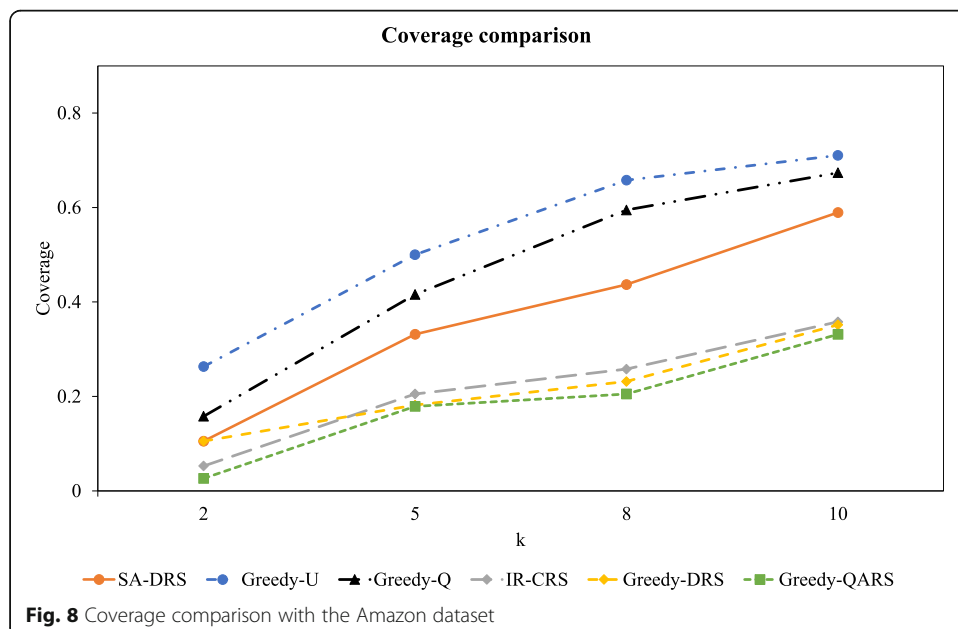


Table 4 Ranking comparison of the six approaches with the Amazon dataset

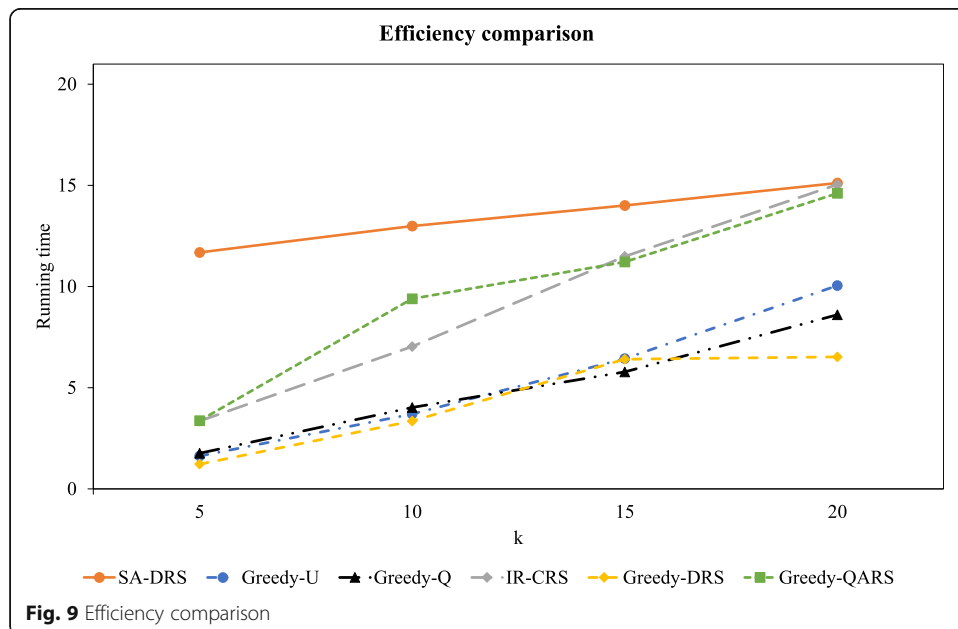
Approach	Diversity rank	Opinion distance rank	Coverage rank	Average rank
SA-DRS	1	3	3	2.3
Greedy-U	4	4	1	3.0
Greedy-Q	2	6	2	3.3
Greedy-DRS	3	5	4	4.0
Greedy-QARS	5	2	5	4.0
IR-CRS	6	1	6	4.3

- (2) In terms of coverage and opinion distance measure, SA-DRS also performs well, which shows that the selected subset is comprehensive and consistent on opinions.
- (3) No matter which language the platforms use, SA-DRS is always the best when considering measures of diversity, coverage and consistency simultaneously.

Efficiency experiment

In this section, we compare the online running time of SA-DRS with the five approaches to validate its efficiency. For one product randomly chosen from the Tmall dataset, the review subset process is conducted 100 times and the average running time of the six approaches is recorded as the final running time to reduce deviation in the efficiency comparison. Figure 9 illustrates the average running time of each approach.

From Fig. 9, we can see all the approaches are efficient in review selection. Although SA-DRS is not the best in terms of efficiency, its running time is still comparable parallel with the other benchmark approaches. The reason why SA-DRS needs to take additional time is that it takes into consideration more factors such as the similarity calculation between two reviews which is time-consuming. However, from the effectiveness experiment results, SA-DRS can not only select a high-



quality, diversified review subset, but also be effective in terms of coverage and consistent opinions, demonstrating that the extra time cost of SA-DRS is helpful in selecting a review subset that consumers need in their decision-making process. Moreover, it is also worth noting that the efficiency experiment platform used in this study is just a local laptop computer with limited computing power, which has a huge gap with the real computing platforms maintained by mainstream e-commerce sites like Tmall and Amazon. The efficiency gap between SA-DRS and the other benchmark approaches can be definitely narrowed by real powerful computing platforms such as parallel computing and high-performance computing. Thus, it is believed that the proposed approach can provide actual users with a desirable review subset in an effective way by being adopted by Tmall or Amazon.

Conclusions

To help consumers overcome the information overload problem in e-commerce, this study focuses on the research on review selection and proposes an approach which can select a diversified and high-quality review subset. The proposed approach first establishes a feature taxonomy based on the topic relation and word context relation. Then, the diversified review selection problem is transformed into a constraint optimization problem according to the constructed feature taxonomy and is solved by a simulated annealing-like method. Extensive real data experiments conducted in this paper demonstrate the effectiveness of the proposed approach compared to other review selection approaches. According to the experiment results, the proposed approach can not only select a diversified review subset, but also have a good performance in terms of coverage and consistency as well.

Although there exist prior studies focused on coverage and/or other measures to find a representative subset, this paper selects a review subset considering diversity and quality based on the constructed feature taxonomy which contributes to the literature of ontology learning, review selection, and consumer behavior studies. First, this paper designs a feature taxonomy method based on the topic relation and word context relation to reveal the hierarchical structure of key features in user generated content. Second, the subset considering diversity based on the feature taxonomy also contributes to the literature with respect to opinion analysis for user generated content. The experimental results reveal that the proposed approach has superiority in selecting a high coverage and consistent subset. Last, the new IT artifact of the proposed review selection approach could have different influences on the decision-making process of consumers when online shopping, which further motivates behavioral science in terms of online consumer behavior studies.

The results of this study can provide technical support for e-commerce platforms. It is very common that consumers cannot read the whole body of reviews of certain products. However, when they are keen to master all details of the candidate products and read high-quality reviews in a short time simultaneously, e-commerce platforms can show a diversified review subset selected by the proposed approach to help consumers improve their decision-making efficiency. When consumers read the selected subsets, they will quickly grasp the features

of the product and the useful feedback of other consumers. Therefore, the implications of this work are twofold. For one thing, the application of the proposed approach can tailor the browsing time of consumers, enhance their satisfaction and help them make purchase decisions. For the other, the proposed approach can simplify the review corpus on e-commerce platforms and raise the readability of large-scale online reviews which are very common in the era of big data.

Future work can be focused on two aspects. One is to incorporate more factors into consideration when estimating the quality of reviews, such as the reputation of authors, the temporal information of reviews and the length of review. The other is to add other factors like the semantic words shared by features into the proposed approach when exploring the semantic relations of extracted features, which will help establish a more feasible feature taxonomy.

Appendix

An illustrative example

In this example, 30 stratified sampling reviews of a mobile phone are selected according to the ratings of the original reviews: 20 reviews with 5-stars, 3 reviews with 4-stars, 4 reviews with 3-stars, and 3 reviews with 1-star. After feature extraction and sentiment analysis, the features of sampling reviews include camera, packaging, appearance, screen, quality, speed, battery, and price, and the corresponding opinion distribution is [11/30, 2/30; 2/30, 0/30; 1/30, 0/30; 8/30, 3/30; 4/30, 0/30; 9/30, 1/30; 9/30, 1/30; 12/30, 0/30]. As well, the number of features, the number of opinion words, the helpfulness vote, the product rating of the review and the overall average rating are used to estimate review quality. All brands mentioned in the sampling reviews are anonymized and

Table 5 Feature taxonomy construction method

Algorithm 1 Feature taxonomy construction

Input: The review matrix $A = (a_{ij})_{n \times p}$; the feature set $F = (f_1, f_2, \dots, f_m)$; the proportion vector $\theta = (\theta_1, \theta_2, \dots, \theta_m)$; the topic relation $TR(f_i, f_j) \ i, j = 1, \dots, m, \ i \neq j$; the word context relation $WCR(f_i, f_j)$;

Output: The feature taxonomy FT ;

- 1: $FT = ()$, $\theta' = \text{sort}(\theta, \text{descending} = \text{TRUE})$
- 2: $FT = \text{addroot}(FT, f_1')$ add root feature to feature taxonomy
- 3: $FT = \text{addsubfeature}(FT, f_1', f_2')$ add sub feature to existing feature in taxonomy
- 4: **for** $i = 3$ **to** m **do**
- 5: $j = \text{argmax}_{j=1 \dots i-1}(TR(f'_i, f'_j) + WCR(f'_i, f'_j))$
- 6: $FT = \text{addsubfeature}(FT, f'_j, f'_i)$
- 7: **end for**
- 8: **return** FT

Table 6 Simulated annealing-diversified review selection approach**Algorithm 2** Simulated annealing-diversified review selection approach

Input: The similarity matrix $S = (sim_{ij})_{n \times n}$; the quality vector $q = (q_1, q_2, \dots, q_n)$; the initial vector X_0 ; the lowest temperature T_{min} ; the initial temperature T_0 ; the cooling parameter ε ;

Output: The review vector X ;

```

1:  $index(X_0) = nonzero(X_0)$ 
2:  $com\_list(X_0) = combination(index(X_0), 2)$ 
3:  $sim(X_0) = sum(S[com\_list(X_0)])$ 
4:  $qua(X_0) = q \times X_0$ 
5:  $V(X_0) = qua(X_0) / (\frac{sim(X_0)}{c_k^2})$ 
6:  $T = T_0, N = 1$ 
7: while  $T > T_{min}$  do
8:    $i = randomint(0, n - 1)$ 
9:    $j = randomint(0, n - 1)$ 
10:  while  $i = j$  or  $X_{N+1}[j] \leq 0$  or  $X_{N+1}[i] \geq 1$  do
11:     $j = randomint(0, n - 1)$ 
12:     $i = randomint(0, n - 1)$ 
13:  end while
14:   $X_{N+1}[i] \leftarrow X_N[i] + 1$ 
15:   $X_{N+1}[j] \leftarrow X_N[j] - 1$ 
16:   $\delta = V(X_{N+1}) - V(X_N)$ 
17:  if  $\delta \geq 0$  do
18:     $V(X_{N+1}) = V(X_N)$ 
19:     $X_{N+1} = X_N$ 
20:  else do
21:     $acceptProb = e^{\delta/T}$ 
22:     $RandomProb = Random(0, 1)$ 
23:    if  $acceptProb \geq RandomProb$  do
24:       $X_{N+1}[i] \leftarrow X_N[i]$ 
25:       $X_{N+1}[j] \leftarrow X_N[j]$ 
26:    end if
27:  end else
28:  end if
29:   $T \leftarrow T * \varepsilon$ 
30:   $N \leftarrow N + 1$ 
31: end while
32: return  $X \leftarrow argmax V(X)$ 

```

Table 7 Example dataset

#	Review
1	This one just seems reasonable and you get more value for your money, compared to similar phones that want to mark up the price ridiculously. If you are on the fence wondering if this is a good option for you, I would say it is for what you pay. MP just needs to focus on expanding into the U.S. market better and give more support for U.S. cell bands. But other than that, this phone is superb. I will definitely consider purchasing more phones like this one from them in the future. I am increasingly surprised by what MP can produce. No major problems here, keep it up.
2	I sold cell phones for over 18 years and the quality of this phone is top notch!
3	Great phone for the price . A lot of people here are negative and I had a lot of second thoughts of buying this phone, I just said I did and I love it and cannot complain.
4	As one coming from the MP2, I do miss the Ultra Power Saving mode, but I am pleased with the phone/ screen size, memory space and quality of the camera .
5	For this price point, it is an amazing deal. MP really has done it again. Really fast speed , paired with MP providing an almost MP1 level of smoothness. Battery life is amazing. With a monster 4000 mAh battery, this phone easily lasts for two days of charge. Call quality is great. Tons of memory, and a great screen makes this phone an amazing deal.
6	The phone was much higher quality than I expected! Very impressive battery life, good camera , and sharp design. Granted, I have only had the phone for a couple days. I am sure that some problems will reveal themselves in the future, but at this point I have no complaints.
7	This is the first time I have purchased a 100% Chinese handset and I must say I am very impressed. The packaging came very nicely packed and in a timely manner, a usual forte of Amazon.
8	Running speed is smooth. Excellent camera . Full day battery and charges very quickly. Screen does great job of being scratch free. The price \$200 unlocked an excellent buy.
9	The battery life is really awesome, this phone has everything for a good price . It does not have the quick charge. But with the great battery life, it is not necessary.
10	I was nervous about buying this phone since I had never heard of the brand but I could not pass it up. Those specs at such a low price ? Count me in! I am super glad I took the chance too. This is one of the best phones I have ever had, hands down. The camera is incredible, the battery just lasts forever, and the speed is really fast. Seriously, take the chance, you will not be disappointed!
11	I have only had the phone a few days but so far so good. Its speed is very fast for the price . You cannot go wrong. The only bad thing is that it does not have wifi calling nor does it have a face recognition software but otherwise great phone for the price.
12	I have had this phone for about a month now. The cameras are great. The screen is beautiful and huge and speed has been fast enough for anything I can throw at it. I am shocked with how amazing this phone is for this price . MP is pretty good and the themes are clean.
13	I ordered this phone about four months ago switching from an MP1 to this. The price is incredible for what it is. The screen is amazing, has faster download speeds , camera is astonishing, the battery is amazing, and it is easy to navigate through MP. Recommend it all the way.
14	The product came early more than I had expected. Packaging with safety box. The phone was in original box with all tools.
15	I am loving MP. It has some pretty cool features. Battery life is great and camera is also no slouch. Definitely recommend.
16	A great value! I love this phone, it does everything I need and more. The battery life is outstanding and the camera produces high-quality images. I highly recommend this phone!
17	Phone is absolutely fabulous. Touch screen and screen size is amazing. As mentioned in specification RAM and processors, the speed is fast. Only thing you are not used to use bigger screen phone then you may find it difficult initially to operate this phone.
18	Great phone with good price .
19	So far I have no complaints... The camera is great and its speed is really quick to respond even with a lot of apps open...
20	Overall appearance of the phone is great. Camera quality is awesome... A quality product in less price .
21	Gets the job done, pretty big battery .
22	Bigger screen that I thought, like the MP1's. Overall, good value.
23	Great phone for the price , for those of you that have issues as to how long the screen takes to pop after it is locked...
24	From the surface this phone is a steal of a value, and it is. The screen is gorgeous, the UI is quick and snappy (but definitely lacks customization options), and overall very satisfying phone to use.

Table 7 Example dataset (Continued)

#	Review
25	Really like the phone. Everything is good— speed, camera , and so on. Did not like ads in system apps but there is a way how to turn them off.
26	This is a beautiful phone but the battery in my device seems to be defective because the battery goes from 100 to 0 within 30 mins. This seller and Amazon refused to help fix this problem and wants me to spend more money than I paid for the phone to do a return, so I have lost all my money on this defective product.
27	The problem is sometimes when I touch the screen , the screen is senseless it happens every day since I bought this phone. So is this really brand-new?
28	The camera quality is piss poor, having four cameras does not make for better pictures, especially when two of them are dormant and useless outside the camera app. Selfies are blurry, and colors are off balance. Unless the phone is completely still on a clear, sunny day, the picture quality is just garbage. They turn out blurry and unclear. And color reproduction is far off key. I cannot express just how bad the camera is on this phone, among the worst I have used.
29	I found the big screen and camera disappointing!
30	Screen scratches easily like phones from 2008. Speed becomes slow after two weeks of use.

replaced by keywords such as MP, MP1, and MP2. The example dataset is given in Table 7.

The constructed feature taxonomy (Fig. 2) is used to calculate feature similarities and review similarities. Then, SA-DRS and other 5 benchmark approaches are carried out to select 5 reviews. The results are shown as follows.

The results in Table 8 illustrate that SA-DRS can select a diversified subset, which covers 6 features including both positive and negative comments, while Greedy-U, IR-CRS, Greedy-DRS and Greedy-QARS only cover positive comments, and Greedy-Q only covers one negative comment. For example, #28 review is a negative but high-quality comment, which can offer consumers more comprehensive details of the product. However, that review is excluded from the selection results of Greedy-U, IR-CRS, Greedy-DRS and Greedy-QARS. As for Greedy-Q, its selection results only have diversified comments on the feature of camera. In contrast, SA-DRS can provide both positive and negative comments on the features of camera, speed, and screen, which is highly desirable in the process of online consumers’ decision-making.

Table 8 Result comparison of the six approaches with the example dataset

Approach	Review subset	Diversity	Opinion distance	Coverage
SA-DRS	#7, #10, #24, #28, #30	35.94	0.186	0.75
Greedy-U	#1, #2, #7, #8, #20	3.28	0.199	1.00
Greedy-Q	#5, #7, #8, #20, #28	12.76	0.226	1.00
IR-CRS	#1, #3, #6, #15, #17	3.01	0.052	0.75
Greedy-DRS	#5, #8, #10, #12, #13	5.33	1.599	0.75
Greedy-QARS	#1, #2, #3, #4, #5	2.72	0.199	0.75

Acknowledgements

All the authors would like to thank those two anonymous reviewers of 2018 FBR International Conference for their valuable comments on the previous version of this study.

Authors’ contributions

All the authors have made an equal contribution in this study.

Funding

This work has been supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (19XNB021).

Availability of data and materials

The authors declare that all the supporting data of this study is from public sources, which are available and feasible for everyone.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Business, Renmin University of China, Beijing 100872, China. ²Department of Statistics, University of Michigan, Ann Arbor MI48109, USA.

Received: 17 December 2018 Accepted: 17 July 2019

Published online: 24 September 2019

References

- Anand, P., & Sternthal, B. (1989). Strategies for designing persuasive messages: Deductions from the resource matching hypothesis. In P. Cafferata & A. M. Tybout (Eds.), *Cognitive and affective responses to advertising* (pp. 135–159). Lexington: Lexington Books/D. C. Heath and Com.
- Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Social Science Electronic Publishing*, 57(8), 1485–1509.
- Bawden, D., & Robinson, L. (2009). The dark side of information: Overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2), 180–191.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(4–5), 993–1022.
- Chen, C. C., & Tseng, Y. D. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4), 755–768.
- Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54(3), 477–491.
- Dhar, V., & Chang, E. A. (2009). Does chatter matter? The impact of user-generated content on music sales. *Journal of Interactive Marketing*, 23(4), 300–307.
- Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 international conference on web search and data mining* (pp. 231–240).
- Djuana, E., Xu, Y., Li, Y., & Cox, C. (2012). Personalization in tag ontology learning for recommendation making. In *Proceedings of the 14th international conference on information integration and web-based applications and services* (pp. 368–377).
- Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10), 1498–1512.
- Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(3), 337–355.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 168–177).
- Jin, J., Ji, P., & Yan, S. X. (2017). Comparison of series products from customer online concerns for competitive intelligence. *Journal of Ambient Intelligence & Humanized Computing*, 10(3), 937–952.
- Jindal, N., & Liu, B. (2008). Opinion spam and analysis. In *Proceedings of the 2008 international conference on web search and data mining* (pp. 219–230).
- Kim, S. M., Pantel, P., Chklovski, T., & Pennacchiotti, M. (2006). Automatically assessing review helpfulness. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (pp. 423–430).
- Korfatis, N., García-Bariocanal, E., & Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11(3), 205–217.
- Lappas, T., Crovella, M., & Terzi, E. (2012). Selecting a characteristic set of reviews. In *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 832–840).
- Lee, I. (2018). Usefulness, funniness, and coolness votes of viewers: An analysis of social shoppers' online reviews. *Industrial Management & Data Systems*, 118(4), 700–713.
- Liu, J., Cao, Y., Lin, C. Y., Huang, Y., & Zhou, M. (2007). Low-quality product review detection in opinion summarization. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning* (pp. 334–342).
- Liu, Y., Huang, X., An, A., & Yu, X. (2008). Modeling and predicting the helpfulness of online reviews. In *Proceedings of the 8th international conference on data mining* (pp. 443–452).
- Maedche, A., & Staab, S. (2001). Ontology learning for the semantic web. *Intelligent Systems IEEE*, 16(2), 72–79.
- Mantel, S. P., & Kellaris, J. J. (2003). Cognitive determinants of consumers' time perceptions: The impact of resources required and available. *Journal of Consumer Research*, 29(4), 531–538.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 26, pp. 3111–3119). USA: Curran Associates, Inc.
- Park, D. H., & Lee, J. (2009). Ewom overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research & Applications*, 7(4), 386–398.
- Paul, D., Sarkar, S., Chelliah, M., Kalyan, C., & Sinai Nadkarni, P. P. (2017). Recommendation of high quality representative reviews in e-commerce. In *Proceedings of the 11th ACM Conference on Recommender Systems* (pp. 311–315).

- Suman, B., & Kumar, P. (2006). A survey of simulated annealing as a tool for single and multiobjective optimization. *Journal of the Operational Research Society*, 57(10), 1143–1160.
- Sun, M. (2012). How does the variance of product ratings matter? *Management Science*, 58(4), 696–707.
- Tang, J., Leung, H. F., Luo, Q., Chen, D., & Gong, J. (2009). Towards ontology learning from folksonomies. In *Proceedings of the 21st international joint conference on artificial intelligence* (pp. 2089–2094).
- Tian, N., Xu, Y., Li, Y., Abdel-Hafez, A., & Josang, A. (2014). Product feature taxonomy learning based on user reviews. In *WEBIST 2014 10th International Conference on Web Information Systems and Technologies* (pp. 184–192).
- Tian, N., Xu, Y., Li, Y., & Gabriella, P. (2015). Quality-aware review selection based on product feature taxonomy. In *Proceedings of the 11th Asia information retrieval societies conference* (pp. 68–80).
- Tsaparas, P., Ntoulas, A., & Terzi, E. (2011). Selecting a comprehensive set of reviews. In *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 168–176).
- Yu, W., Zhang, R., He, X., & Sha, C. (2013). Selecting a diversified set of reviews. In Ishikawa, Y., Li, J., Wang, W., Zhang, R., & Zhang W. (Eds.), *Web technologies and applications. APWeb 2013. Lecture notes in computer science* (Vol. 7808). Berlin: Springer.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
