Electronic Commerce Research (2021) 21:423–444
https://doi.org/10.1007/s10660-019-09372-5

The user preference identification for product improvement based on online comment patch

Shugang Li[1] · Yuqi Zhang[1] · Yueming Li[1,2] · Zhaoxu Yu[3]

Published online: 14 August 2019

Abstract

Online comments have become a valuable source for designers for the purpose of product improvement. However, the implicitly expressed users' preferences on multiple product attributes in incomplete online comments make it difficult to extract useful information to improve the product from the online comments. In order to identify the users' preference from the perspective of product improvement, the comment extension mining model is proposed to patch up the online reviews based on the semantic similarity and emotional resemblance. First of all, for the sake of mining the full context semantic information of the keywords in comments, they are transferred into vectors by the word2vec method. Next, smart semantic distance measurement models are developed to match the online comments with the standard comment templates that express users' sentiment on product attributes based on the semantic similarity. Moreover, the fine-grained matching neural networks are designed to further match the reviews of each product attribute to its standard sentiment templates according to the user's overall emotions towards the product. Finally, the KANO model is introduced to depict users' preferences and develop product improvement strategies. The experiment on laptop product confirms that our method is effective.

Keywords Online comments · Comment extension mining model · Fine-grained matching neural network · Product improvement · KANO model

✉ Shugang Li
westside_li@163.com

1 School of Management, Shanghai University, Shanghai 200444, People's Republic of China

2 China United Network Communications Group Co., Ltd, Anhui 230000, People's Republic of China

3 Department of Automation, East China University of Science and Technology, Shanghai 200237, People's Republic of China

1 Introduction

With the rapid proliferation of e-commerce, people increasingly use online comment mechanism to depict detailed aspects of products and services and reflect their perceptions, sentiments, and preferences. User-generated reviews on the Web has become an important information source that remarkably influences consumer decisions and product sales performance. Gartner reports that consumers trust online reviews as much as a personal recommendation, 67% say they are influenced by online reviews [1]. More and more companies are paying close attention to the comments in the e-commerce environment, for instance, retailers (e.g., Amazon.com, Shopping.com, Walmart.com) and product manufacturers (e.g., Hewlett-Packard, Nike, Levi's) have established their own opinion sharing communities, where customers can express their opinions on products they have purchased, or they are interested in. From the perspective of enterprises, it is of significant importance to learn from Big Data of online user-generated reviews for product enhancement and/or new product design.

To exploit Big Data of online comments, i.e., the volume, velocity, and variety of raw user-generated data from the perspective of product improvement, identifying customer preference is the first and the most critical step. Measuring customer preference can help not only to product improvement and product design but also to planning and decision-making pertaining to pricing, positioning, market segmentation and advertising [2, 3]. Traditionally, companies and business researchers collect consumers' preferences through offline surveys or virtual experiments, such as paper-and-pencil surveys and conjoint analysis [4], which can easily become expensive in terms of time and money. In recent years, some research seeks to exploit online reviews source for the purpose of exploring consumer preference measures (i.e., a preference measurement model) that capture the effects of consumers' sentiment towards product features [5].

"Listen carefully to what your customers want and then respond with new products that meet or exceed their needs", that undoubtedly led to great products [6]. Compared to information obtained from other sources, online comments, which are voluntarily generated by consumers, provide a more reliable, more accessible, and richer information source to understand customer requirements and consumer preferences [7, 8], especially for those consumer-oriented businesses. In addition, online comments are excellent sources of some innovative ideas which may bring subversive innovations.

However, despite the huge volume of comments could contribute to consumer preference identification and product improvement, it is difficult to mine valuable information from implicit and incomplete comments. Moreover, users are more inclined to express their sentiment polarity towards the product attributes, rather than explicitly reveal their preference on the attributes. In other words, whether a customer cares about one attribute above others is recessive. For example, when a user commented that "the appearance of the notebook is far from pleasant", we can only draw the conclusion that the user gave a negative evaluation on appearance. But we cannot assert that negative emotion to the appearance property necessarily leads to dissatisfaction, because the user may do not care about the appearance at all, and even cannot measure the degree of the user's preference for this feature. However, most of the extant studies which explore consumer preference measures utilizing online reviews ignore the patch of incomplete information. In this context, we develop the following research questions:

(1) How to extract consumers' sentiment towards product features from implicit and incomplete reviews?
(2) How to mine consumers' preferences for the purpose of developing product improvement strategies?

To answer these questions, we propose a comment extension mining model (CEMM) to supplement the information about the product attributes which user cares about in a review based on the semantic similarity and emotional resemblance. Firstly, a set of standard templates that expresses user sentiment towards product attributes is designed to describe all possible comment patterns. Each standard sentiment template not only contains the information on the user's sentiment polarity for the product attributes but also contains the new extended information on the impact of the attribute on the satisfaction level of the user. In order to improve the precision of the match, each keyword in the comment is transformed into a word vector by word2vec technology to fully disclose its context semantic information in a specific comment. Secondly, smart semantic distance measurement models (SSDMs) are adopted to match online reviews and standard templates to make up the incomplete comment information. Thirdly, fine-grained matching neural networks (FGMNNs) are designed to further match the reviews of attributes and their standard comment templates by considering the joint effect of the semantic similarity between comments and standard templates and user's overall emotions towards the product.

To uncover users' preferences from the perspective of developing appropriate product improvement strategies, based on accurate extended information, the KANO model is applied. KANO model can map the product attributes into five categories, on each of which consumers show different degree of preferences. Finally, we use laptop product data from JD.com, one of China's largest online marketplaces, to test our approach. The experiment results prove the validity of our method. Furthermore, we provide suggestions for the enterprises' product improvement decisions.

The rest of this paper is arranged as follows. Section 2 provides a brief introduction to related work and techniques. Section 3 presents our research framework. Section 4 presents the CEMM and preference mining method utilizing KANO model from the perspective of product improvement. Section 5 is the experimental part. In Sect. 6, the implications and limitations are discussed.

2 Literature review

2.1 Sentiment analysis

In recent years, sentiment analysis has been widely applied to understand consumers' emotions embedded in various user-generated content. Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [9]. The sentiment can be classified by various approaches, such as the rating of the comment, the emotion (i.e. anger, happiness), the polarity (i.e. positive, negative, or neutral), and some more sophisticated approaches (such as a "5-star" scale), etc. A common method of sentiment measurement is based on polarity [10].

Hu and Liu [11] mined the features of the product and customers' positive or negative opinions. They manually set the initial adjectives and then expanded the set using synonym and antonym relations in WordNet and predicted the orientations of user-generated texts. Conjunctions have a substantial impact on the overall sentiment of a sentence, thus Meena and Prabhakar [12] used word dependencies and dependency trees to analyze the sentence constructs and presented how atomic sentiments of individual phrases combined in the presence of conjuncts to decide the overall sentiment of a sentence. Pang et al. [13] first applied three machine learning approaches (naive Bayes, maximum entropy, and support vector machines), with unigram (bag-of-words) as features for binary sentiment classification of movie reviews. Then, other features, specifically the frequency of terms, POS, sentiment words, syntactic dependency, were considered. Ye et al. [14] compared three supervised machine learning algorithms of Naive Bayes, SVM, and the character-based N-gram model for sentiment classification of the reviews on travel blogs for seven popular travel destinations in the US and Europe. In order to understand unstructured online reviews, Tripathy et al. [15] utilized four different machine learning algorithms, such as Naive Bayes (NB), Maximum Entropy (ME), Stochastic Gradient Descent (SGD), and Support Vector Machine (SVM) for classification of human sentiments.

In reality, user-generated comments are usually incomplete; and preference expressions are implicit. However, existing sentiment analysis methods mostly neglect the patch of implicit and incomplete information in online comments. Therefore, by defining the standard sentiment templates of online comments, this paper puts forward the CEMM to match the reviews of each attribute to standard templates by considering the joint effect of the semantic similarity between comments and standard templates and user's overall emotions towards the product.

2.2 Preference mining

In decision theory, the user's preferences are value judgments and tendentious choices when comparing the alternatives available for choice. Although sentiment analysis is adequate when mining collective sentiments [16], however, preference mining and representation is called for, as soon as we wish to depict users' value judgments on every kind of product features according to sentiment polarity and consumers' overall satisfaction.

Substantial research has proposed many methods to extract consumer preferences, such as survey-, behavior-, and online review-based approaches. Online review-based measure leverages the huge amount of existing, publicly available online user-generated reviews and thus avoids time-consuming and costly data collection process [5]. Additionally, it has been demonstrated that the review-based method is more favorable compared with traditional conjoint analysis techniques [17]. Over viewing extant research, sentiment analysis is the prerequisite for user's preferences elicitation from the online user-generated content source.

Zhao et al. [18] proposed a unified probabilistic model to capture topical-region preference and category aware topical-aspect preference simultaneously. Their model considered several different factors including topical aspect, sentiment, and spatial information and captured the interaction of these factors. Decker and Trusov [17] proposed a preference measurement based on online consumer review data using a negative binomial regression approach, which allows inferences on the relative effect of product attributes and brand names on the overall evaluation of the products; specifically, opinion heterogeneity was taken into account. Thus, the available plentitude of individual consumer emotional polarity was turned into aggregate consumer preferences. To learn consumers' relative preferences for different product features, Archak et al. [19] presented an econometric modeling framework to mine online review text information. After identifying users' opinions about the features that are embedded in the reviews, an econometric model was proposed to quantitatively report the pragmatic and economic value of these evaluations. Xiao et al. [5] proposed a novel econometric preference measurement model, the modified ordered choice model (MOCM), to extract aggregate consumer preferences from online product reviews.

Despite ample research on preference mining has provided several managerial implications, such as marketing decision supporting [20], online shop optimizations [19], and user recommendation [18]. To the best of our knowledge, users' preferences can contribute to product improvement or product design; yet only a very few studies have done some relevant explorations, and most of them stop at measuring feature preference (or feature importance). For example, Li et al. [20] proposed a product portfolio construction module which can derive insights (product feature specification and feature importance) and help enterprises to design next-generation products portfolio, through extracting and consolidating the reviews expressed via social media. More relevant to our research, Qin et al. [21] provided an innovative way to use the KANO model on product attributes classification from the perspective of enterprises, revealing the non-linear relations between product features performance and consumer's overall satisfaction. Thus, KANO model is introduced to extract consumers' preference from the perspective of product improvement and design.

2.3 KANO model for preference mining and product improvement

The Kano model is a theory for product development and customer satisfaction. KANO model can classify product attributes and solve the positioning problem of product attributes. The KANO model classifies customer preferences into five categories [22], as is shown in Fig. 1: (1) Basic attribute (B) is the must-be requirement of customers for product or service attributes provided by enterprises. When it is met, customers are just neutral, but when it performs poorly, customers are very dissatisfied; (2) Expected attribute (E) results in satisfaction when fulfilled and dissatisfaction when not fulfilled. In other words, the satisfaction degree of the customer is proportional to the performance of this kind of attribute. (3) Attractive attribute (A) refers to that customer satisfaction increases sharply with the increase of meeting customer expectation. This kind of attribute provides satisfaction when satisfied fully, but will not cause dissatisfaction when not fulfilled. Once achieved, the customer's satisfaction increases even if the performance is imperfect. (4) Indifference attribute (I), no matter this attribute is provided or not, there is no effect on the user experience and it will not result in customer satisfaction or dissatisfaction. (5) Reverse attribute (R), is an attribute that leads to strong dissatisfaction when fulfilled. For example, some customers prefer for high-tech products while others prefer ordinary products.

To integrate the voice of the customer into the subsequent processes of product development, the KANO model was adopted by some research. Sharif et al.

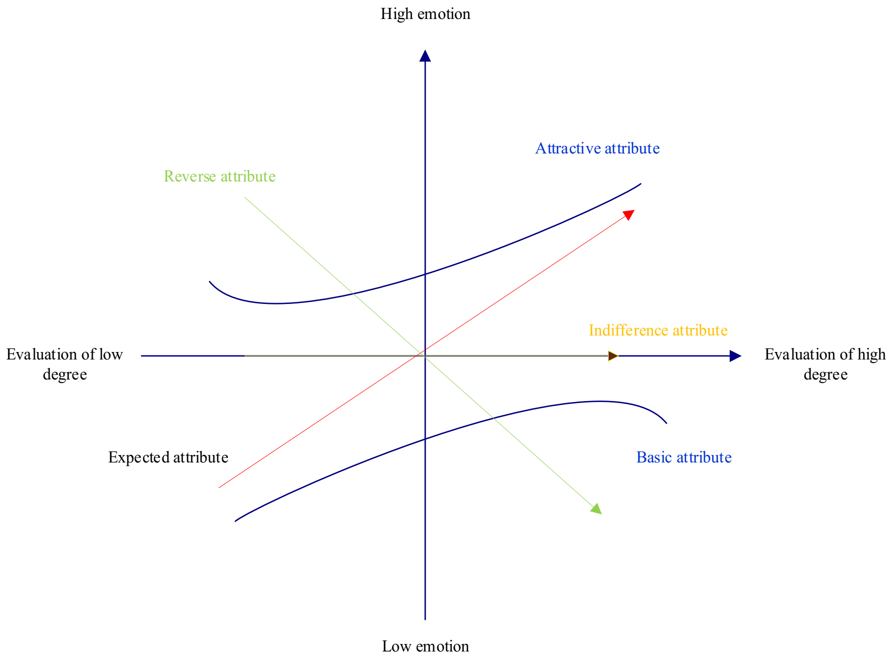

Fig. 1 KANO model

[23] presented a method for analyzing customers' preferences obtained by using the KANO model. Focusing on the early phases of product development, Tan et al. [24] suggested the combined use of quality function deployment (QFD) and KANO model. Senthil et al. [25] implemented Customer KANO Model for turning the design perception into a precise knowledge as more creative activities from the hidden knowledge of designing using statistical science.

In this study, after extracting emotional polarity of customers, we map product attributes into five categories according to the KANO model for the purpose of preference identification. Through analyzing consumers' preference on five categories of product features, some product improvement strategies can be proposed.

2.4 Critical techniques: word2vec and Convolutional Neural Network (CNN)

As the basic algorithm in the field of natural language processing, the word vector algorithm plays an important role in sequence annotation, question answering system, text classification, and many other tasks [26, 27]. Word2vec extracts the semantic information of words and represents them with a low-dimensional real vector [28]. Its advantage is that similar words are closer in distance, reflecting the correlation and dependence between different words. Since word vectors can be acquired by pre-training corpus, their introduction effectively reduces the network depth and makes deep learning an efficient method for text classification. In our proposed CEMM, word2vector is employed to generate word vectors for semantic analysis.

In recent years, deep learning has shown great advantages in the fields of speech, image and text processing, and made breakthroughs in core issues. In the field of Natural Language Processing (NLP) and deep learning, Convolutional Neural Network (CNN) is commonly used for text classification. For example, Kalchbrenner et al. [29] proposed that a convolutional neural network should be applied to sentences and applied to semantic analysis and classification problems. Zhang and Wallace [30] used deep learning to conduct an emotional analysis of texts. Yin et al. [31] adopted a general Attention Based Convolutional Neural Network (ABCNN) for modeling a pair of sentences. These classification problems using deep learning have achieved better classification results than before. He et al. [32] proposed a Multi-Perspective Sentence Similarity Modeling with Convolutional Neural Networks. This method uses the word vector to represent the sentence as a matrix as input, and uses a variety of pooling operations after convolution to obtain multi-angle sentence features, and represents the sentence as a vector for similarity comparison. Kim et al. [33] proposed a sentence classification method based on single-hidden CNN based on word2vec. This method uses the idea of multi-channel, sets up a number of different convolution Windows, flexibly obtains a variety of context features in the sentence, and then classifies the text after it is expressed as a vector. As a typical deep learning model, CNN is adopted as a test benchmark model for the evaluation of proposed CEMM in this study.

3 The product improvement model based on online comments

This paper mainly studies how to use online comments to mine users' preferences from the product improvement perspective. Our research framework is illustrated in Fig. 2. First, we take laptops as our research object and use Chinese online reviews as the source of data. Thus, we must translate the unstructured short Chinese text into structured data for further analysis, which is the basic portion of our research. This process includes data investigation, data crawling, lexicon building (attribute lexicon, emotional lexicon, and evaluation lexicon) and data set building. Second, we build a CEMM to supplement the users' information about their emotions towards attributes in online reviews, which includes the SSDMs and the FGMNNs. After creating standard sentiment templates of online comments and calculating word vector, the SSDMs are applied to do preliminary sentiment analysis of online reviews by getting the semantic distance between the comments and the standard sentiment templates and identifying the standard sentiment templates that are closest to the semantics of the comments. Based on the identified standard sentiment templates, the FGMNNs are designed for product attributes to further match comments

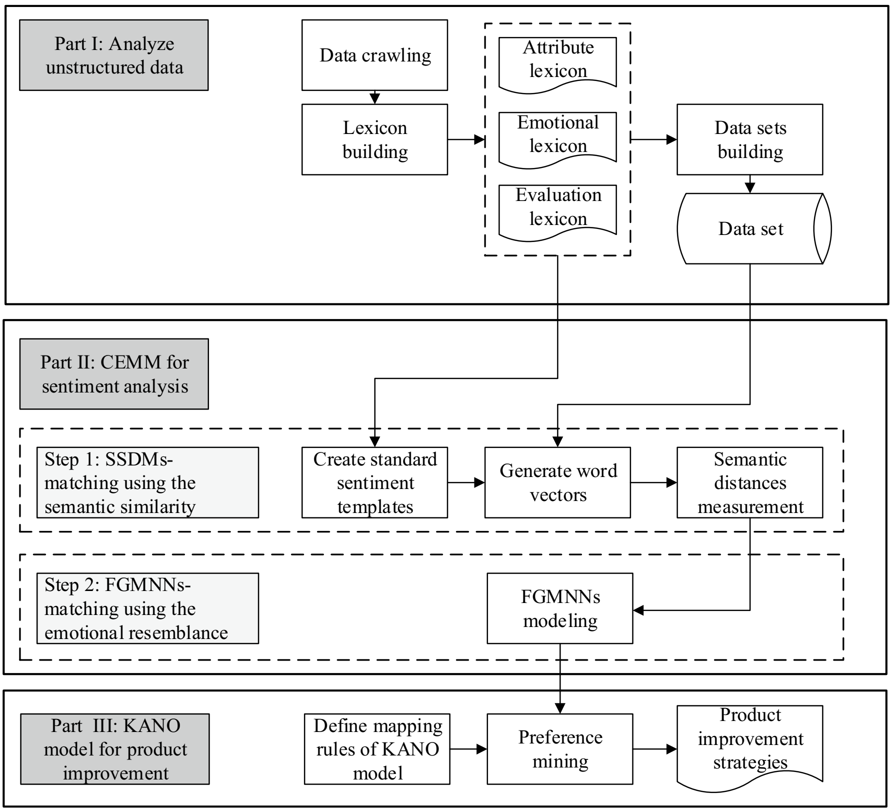

Fig. 2 Research framework

and standard sentiment templates in accordance with the user's overall satisfaction on the product. Third, we build a set of attribute classification rules to map the attributes according to KANO model. The results of laptop attributes preferences are obtained and several product improvement strategies are proposed on the basis of the results.

4 Comment extension mining model (CEMM) and KANO model for product improvement

In this study, we aim to discover customer preference when the customer has a positive evaluation or a negative evaluation of the product attributes. Therefore, we need to find out all of the product attributes, positive evaluation terms, negative evaluation terms, and emotional words towards the product. Accordingly, corresponding standard sentiment templates of online comments that express users' emotions towards product attributes are formulated. Utilizing KANO model according to the results of CEMM, product attributes can be mapped into five categories, on each of which consumers show different degree of preferences. Based on users' preferences, corresponding product improvement strategies can be proposed.

4.1 Data crawling and lexicon building

We selected JD.com as online comment resource. And we developed a data crawler to collect all online comments of a laptop on JD.com. More than 20,000 online comments were eventually captured from JD.com. By removing some repetitive comments and comments in which features are not obvious, we finally selected 4021 comments for this study. Next, we analyzed unstructured data to translate the short Chinese text comments into structured data for further analysis.

The first step is to identify the product attributes. We applied POS (part-of-Speech) tagging [34, 35] and removed stemming and stop-words. Then, according to the frequency and semantic relationship of terms, the Latent Dirichlet Allocation (LDA) and Page Rank are used to rank terms. A total of 3404 terms were extracted as candidate attributes. Then we divided the filtered attributes into 13 categories by consulting professional notebook designers. The attributes include Boot speed, Screen, System, Price, Peripheral, Service, Memory, Appearance, Function, Temperature, Logistics, CPU and Battery. Then, according to the existing HowNet resources, we built the emotional words library and evaluation words library of laptop product. Among them, the evaluation words were divided into two categories: positive and negative evaluation words, which are the same as HowNet, as is shown in the second column of Table 1. To map online comments to the KANO model, we divided emotional words into three levels according to their emotions, including Satisfied, Neutral and Unsatisfied, as is shown in the third column of Table 1.

According to attribute words, evaluation words and emotional words, comments on each attribute can be divided into six categories. For example, if a customer gave a positive evaluation of the battery, there are three possibilities for his satisfaction:

Table 1 Three lexicons and six categories of comments related to attributes

Attribute	Evaluation words	Sentiment words	Categories of comments related to attributes
Battery	Positive evaluation	Satisfied	P1
		Neutral	P2
		Unsatisfied	P3
	Negative evaluation	Satisfied	N1
		Neutral	N2
		Unsatisfied	N3

satisfied, neutral, or dissatisfied. Similarly, when a customer had a negative evaluation of batteries, there are also three possibilities for their satisfaction. Therefore, comments on each attribute can be divided into six categories: P1, P2, P3, N1, N2, and N3, as is shown in the fourth column of Table 1. So we hired five experts in this field to divide the collected online comments into 6 categories for each attribute, consequently, the data set used to train and test the model can be obtained.

4.2 Standard sentiment templates of online comments definition

In the later model application and comment classification process, we need a standard comment statement to make a baseline statement to categorize other comments. Therefore, we need to define standard sentiment templates for six categories of comments for each attribute, as is shown in Table 1 (P1, P2, P3, N1, N2, and N3).

Usually, customer comments are mostly compound sentences, which contain attribute words, evaluation words and (or) emotional words, so we apply the compound sentences structure as the standard comment sentence pattern of this paper. For example, for the attribute of price, we set the standard sentiment template of the P1 level as "Computer price is affordable, and I am very satisfied". This is a compound sentence, where the "computer price" is the attribute word, the "affordable" is the evaluation word, and "satisfied" is the emotional word.

Three-Plane Theory summarizes sentence type, sentence pattern and sentence category from three planes: syntax, semantics, and pragmatics. Sentence type refers to the grammatical structure of a sentence [36]. Sentence pattern refers to the semantic structure pattern of sentences. Sentence category refers to the pragmatic value category of a sentence. Neither of the three can represent the whole of a sentence alone; thus a sentence should be a structure composed of sentence type, sentence pattern, and sentence category. The standard sentiment templates of online comments for this paper contain these three planes, so our templates of comments are valid, as is shown in Fig. 3.

4.3 Semantic analysis of the comments using word2vec

Consumer online reviews represent consumer satisfaction on the attributes of the purchased product. To obtain the overall consumer preference, we try to acquire

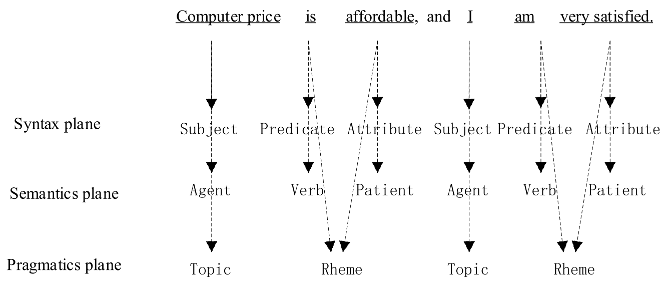

Fig. 3 The template contains three planes of syntax, semantics, and pragmatics

the customers' emotion patterns on each attribute by classifying the comments based on the evaluation polarity and emotional intensity. This comment classification process is accomplished by measuring the semantic distances between a comment and standard comment templates. Based on word2vec and Euclidean distance, SSDMs are proposed to calculate the sentence semantic distances, according to which the online comments are matched to the closest sentiment template.

To train the corpus, this paper used the word2vec learning method, which uses the neural network language model [37]. This language model contains two word-vector learning structures, namely CBOW and Skip-Gram. The former uses the context to predict the current word, while the latter uses the current word to predict its context. Consequently, the CBOW word vector learning structure is adopted in this study.

The CBOW model is mainly adopted to predict target words by using words in a given sliding window. Given a word sequence $W = w_1, \ldots, w_m$, the goal of CBOW model is to maximize the average logarithmic probability:

$$l(W) = \frac{\sum_{i=L}^{M-L} \log p(w_i | w_{i-L}, \ldots, W_{i+L})}{M} \tag{1}$$

where L is the context window size of the target word. The SoftMax function is adopted to formulize

$$p\left(w_i | w_{i-L}, \ldots, W_{i+L}\right) = \frac{\exp\left(\boldsymbol{x}_{\boldsymbol{o}}^{\boldsymbol{T}} \cdot \boldsymbol{x}_{\boldsymbol{i}}\right)}{\sum_{x_i \in D} \exp\left(\boldsymbol{x}_{\boldsymbol{o}}^{\boldsymbol{T}} \cdot \boldsymbol{x}'_{\boldsymbol{i}}\right)} \tag{2}$$

$$\boldsymbol{x}_{\boldsymbol{o}} = \frac{\sum_{j=i-L,\ldots,i+L, j \neq i} \boldsymbol{x}_{\boldsymbol{j}}}{2L} \tag{3}$$

where D is the dictionary, $\boldsymbol{x}_{\boldsymbol{i}}$ is the vector representation of the target word w_i, and $\boldsymbol{x}_{\boldsymbol{o}}$ is the average of all the contextual word vectors. In this paper, each word is converted to a 200-dimensional word vector (D = 200).

4.4 The smart semantic distance measurement model (SSDM)

Transformed into word vectors, the semantic distance between two sentences is calculated by the semantic distance between words, as is shown in Sect. 4.3. In this study, we consider not only the weight of words but also the weight of word alignment between different sentences. Furthermore, the minimum semantic distance between two sentences is used to measure the similarity between them. Specifically, S and S' are two text sentences respectively represented by CBOW model. Then, a SSDM model is proposed to measure the semantic similarity between a comment and a standard sentiment template according to the semantic distance distribution of words. The minimum semantic distance between S to S' can be solved by the following linear programming:

$$SD\left(S, S'\right) = \min_{T>0} \sum_{i=1}^{m} \sum_{j=1}^{h} T_{ij} D(i,j) \tag{4}$$

s.t.

$$\sum_{j=1}^{h} T_{ij} = d_i, \quad \forall i \in \{1, \dots, m\} \tag{5}$$

$$\sum_{i=1}^{m} T_{ij} = d_j', \quad \forall j \in \{1, \dots, h\} \tag{6}$$

where $\boldsymbol{T} \in R^{m\times h}$ is the matrix of word alignment weight when a word moving to another [38]. m and h represent the number of words in the sentence S and S', respectively. $T_{ij} \geq 0$ denotes the alignment weight between word w_i in S and word w_j in S'. d_i and d_j' represent the weight of word w_i and word w_j respectively, that is, if word w_i appears c_i times in sentence S, then its weight is $d_i = \frac{c_i}{\sum_{j=1}^{|S|} c_j}$, and d_j' is obtained in a similar way. Moreover, $\sum_j T_{ij} = d_i$, $\sum_i T_{ij} = d_j'$.

To calculate the semantic distance between a comment and a standard comment template, we need to calculate the semantic distance $D(i,j)$ between words. The traditional word vector representation is mainly achieved through the dictionary, that is, the dimension of the vector is the size of the dictionary, and the word vector is the index position of the word. Such a word vector is very sparse. But the distributed representation of words overcomes the problem of too large dimensions and embeds the contextual semantic information of the word into the word vector representation. So, the semantic distance measurement of these words is computed with Euclidean distance in the word2vec embedding space. Here, the European distance between the word w_i and w_j is calculated by the formula (7):

$$D(i,j) = \| \boldsymbol{x}_i - \boldsymbol{x}_j \|_2 \tag{7}$$

where $D(i,j)$ represents the semantic distance between word w_i and w_j, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ represent their word vectors, respectively. If the semantics of the two words are very close, then in the semantic space, the value of $D(i,j)$ is smaller; otherwise, the value of $D(i,j)$ is larger.

Through the above optimization problem, using the simplex tableau of the Simplex algorithm, we can solve T_{ij}. Finally, the constraint optimization problem in formula (4) is a special case of the optimization of Earth Mover's distance metric (EMD) [39], which has been widely used for the distance measurement between images and text [37, 40]. Therefore, the proposed sentence semantic distance model also uses the EMD's optimization method to accomplish the semantic distance calculation. For example, the sentence 1 is "Bought this computer because of cheap, light and beautiful. Fingerprint unlock is convenient. Just look at the good reviews". And the sentence 2 is "The overall feeling is ok, I'm very satisfied". The semantic distance between the two sentences is 0.999984741979.

4.5 The fine-grained matching neural network (FGMNN)

We use the Earth Mover's distance metric and word2vec methods in CEMM to match the comments and standard templates preliminarily according to the context semantic information. Through preliminary matching, we can get the semantic distance between a comment and all the standard sentiment templates of online comments. Next, we design the FGMNNs to further match the reviews and their standard sentiment templates of online comments based on both the content of comments on this attribute and user's overall emotions towards the product.

Neural networks can learn past cases and suggest the final matching template. Using the neural network, the final matched template is estimated by the learned relationships between the determinants (semantic distance between templates and comments) and the past review cases. As a result, product designers can reduce biases from their limited individual memory and experience, as they can learn from others' experiences by estimating the weights of neural networks to solve problems.

In this study, we respectively build an FGMNN for each attribute, namely, 13 FGMNNs are built. The FGMNN model for one attribute, as is shown in Fig. 4, has three layers. The input layer is the distance between all the attributes contained in the reviews and the individual templates. For the sake of investigating the influence of commentator's overall emotions towards the product on his interest for a specific attribute, in the input layer, besides the attribute to be analyzed, the other attributes are also input. Specifically, the input $SD_n^{(s)}(1 \le n \le 13; 1 \le s \le 6)$ represents the semantic distance between the nth attribute and the sth template corresponding to the nth attribute. The hidden layer is the matching probabilities between one of the attributes studied and its templates. The output layer outputs the preference comment template with the maximum matching probability.

Among them, the matching probability of the hidden layer is calculated by the following formula (8). The probability values between attribute a and the matching template are obtained using the logical function, which is as follows:

$$h_a^s\left(SD^{(i)}\right) = \frac{1}{1 + e^{-\theta^T SD^{(i)}}} \tag{8}$$

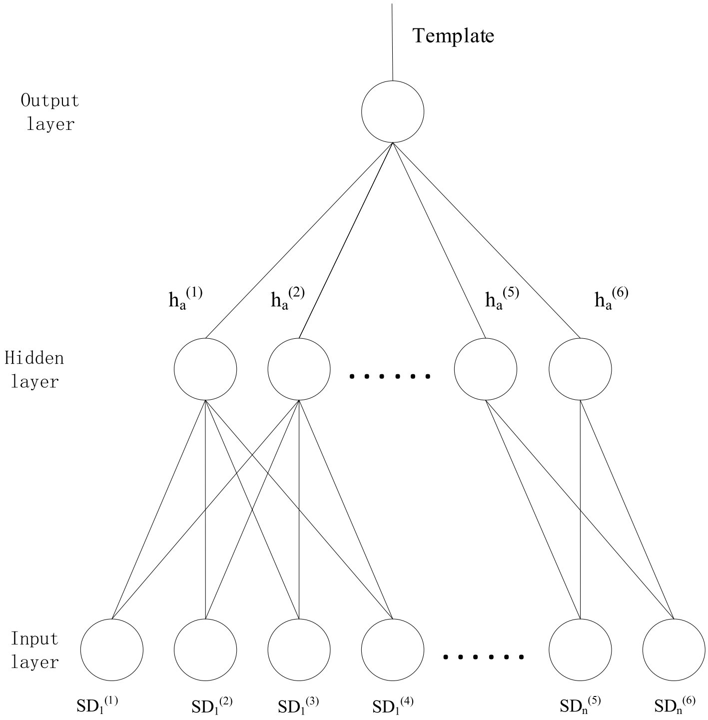

Fig. 4 FGMNN model for one attribute

Here, $\theta^T \mathrm{SD}^{(i)}$ is the sum of the weighted inputs and $h_a^s\left(SD^{(i)}\right)$ is the matching probability between the ath ($a = 1, 2, \ldots, 13$) attribute studied and its sth ($s = 1, 2, \ldots, 6$) template. $\mathrm{SD}^{(i)}$ is input vector of sample i, namely the semantic distance between the attributes and their templates, and θ is a parameter vector, which can be estimated using the gradient descent algorithm with the aim to minimize the following cost function:

$$J(\theta) = -\frac{1}{r}\left[\sum_{i=1}^{r} y^{(i)} \log h_a^s\left(SD^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_a^s\left(SD^{(i)}\right)\right)\right] \tag{9}$$

where r is the number of samples, and $y^{(i)}$ is 1 if the output of sample i is a sth template, else $y^{(i)}$ is 0.

4.6 Users' preferences mining from the product improvement perspective using KANO model

After mapping online comments of each attribute into the specific standard sentiment template using the proposed CEMM model, then we can classify laptop attributes into different categories according to the KANO model, and the classification rules are shown in Table 2.

Table 2 Attribute classification rules

	N1	N2	N3
P1	Suspicious results	A	E
P2	I	I	B
P3	R	I	Suspicious results

Table 3 Confusion matrix

	The true labels of the test set	
	Positive	Negative
Predicted labels of the test set		
Positive	True positives (TP)	False positives (FP)
Negative	False negatives (FN)	True negatives (TN)

For example, when customers have a positive evaluation of the battery, their emotion is satisfied. But when customers have a negative evaluation of the battery, their emotion is unsatisfied. So, the battery is an expected attribute. In other words, the emotion degree of the customer is proportional to the performance of the battery. When customers have a negative evaluation on battery, yet their emotion is neutral; however, positive evaluation on the battery is accompanied by satisfying emotion. That means that battery is a kind of attractive attributes. Indifference attributes are related to three kinds of categories, as is shown in Table 2.

5 Experiment results

5.1 Data set and model evaluation criteria

The data set of this paper is the consumers' comments on the laptop on JD.com. 4021 pieces of comments were used in this study. They were divided into two parts: the training set and the test set. In the experimental evaluation, the method used is a confusion matrix. A confusion matrix is a tool to evaluate the accuracy of the classification algorithm. By comparing the predicted results of the models with the test samples, the following confusion matrix was introduced, where the columns are the true label of the test set. The rows are predicted results, which are classified according to the predicted probability. The true labels of the test set and predicted labels were divided into Positive and Negative categories.

Based on the confusion matrix in Table 3, Precision, Recall, and F1 were used to evaluate the classification effect of the models. They are defined as follows:

$$Precision = \frac{TP}{TP + FP} = \frac{the\ correct\ number\ of\ similar\ sentences\ detected}{the\ number\ of\ similar\ sentences\ detected} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} = \frac{the\ correct\ number\ of\ similar\ sentences\ detected}{the\ number\ of\ similar\ sentences\ in\ the\ corpus} \tag{11}$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{12}$$

The denominator in formula (10) refers to the number of sentences like the template, which are detected by the test method. Numerator refers to the correct number of sentences detected. The denominator in formula (11) refers to the actual number of sentences like the template in the data set.

5.2 Contrast experiment and results

We used other methods based on lexical, convolutional neural networks (CNN), support vector machine (SVM), Bayes classifier, logistic regression (LR), WordNet, LSTM, and TF-IDF to compare the experimental results of our method. The methods based on lexical and WordNet are unsupervised, while the LR, CNN method, SVM, and Bayes are the supervised methods. All the comparison methods and CEMM method in this paper were implemented through Python software. As described in Sect. 4, we need to match online comments with standard sentiment templates for each attribute. The cross-validation method we used is the Hold Out Method. And we used 80% of online comments of each attribute as the training set, and the rest of online comments as the test set. The input of each benchmark model was a standard sentiment template for a certain attribute and the comment related to this attribute. The output was the semantic distance between each comment and the standard sentiment template. Then the template with the minimum semantic distance was selected as the output.

Table 4 shows the value of TP, FP, FN, TN, Precision, Recall and F1 of test results. As can be seen from Table 4 and the line diagram in Fig. 5, the value of the Precision, Recall and F1 of CEMM are the highest among all the contrast methods. So, it can be safely drawn the conclusion that the performance of CEMM is better

Table 4 Experimental results of different methods

	TP	FN	FP	TN	Precision (%)	Recall (%)	F1 (%)
LR	1418	1309	764	1334	65	52	58
CNN	1473	1255	1415	683	51	54	52
SVM	1636	1091	1572	526	51	60	55
Bayes	1527	1200	936	1162	62	56	59
TF-IDF	2236	491	1313	784	63	82	71
LSTM	2264	464	566	1532	80	83	81
Lexical	1336	1391	2090	8	39	49	43
WordNet	1527	1200	1793	305	46	56	51
SSDM	1800	927	969	1129	65	66	65
CEMM	2291	436	227	1871	91	84	87

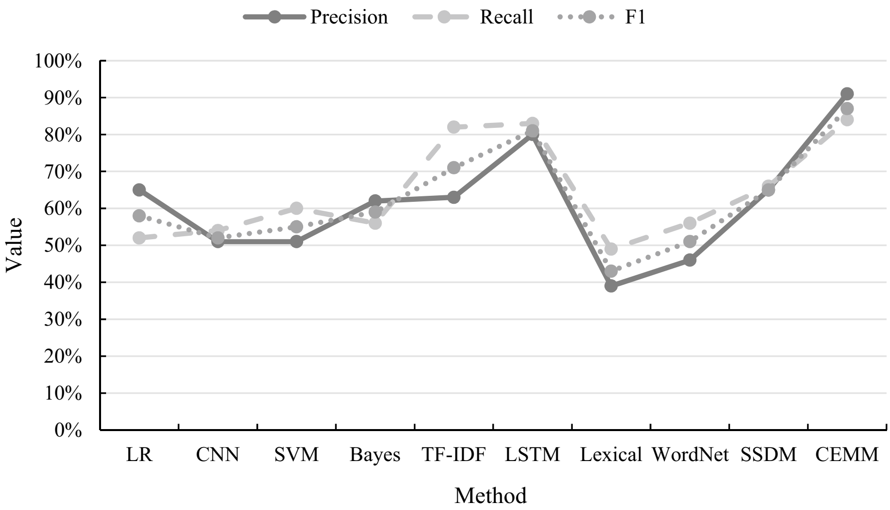

Fig. 5 Experimental comparison of different methods

than that of other methods. In addition, we can find that the performance of CEMM is much better than that of SSDM, so the FGMNN can improve the precision of the model. Therefore, we can use the CEMM model to classify comments reliably according to semantic similarity and emotional resemblance.

At the same time, we also compared the precision of our model when the number of attributes contained in the same comment is different. By observing each comment sample, we found that in a comment the user describes 6 attributes at most, so we conducted experiments on 6 attributes for the upper bound. The results of the experiment are shown in Fig. 6. Obviously, in Fig. 6, when the number of attributes contained in the comments is more, the accuracy of the model is higher. This is because the more attributes we consider, the better we can fully portray the user's emotion towards the product.

Next, we compared the precision of our model when different numbers of the most similar templates are selected to calculate the semantic distance in the input layer of FGMNNs. The experimental results based on all samples, as is shown in Fig. 7, show that when the input layer selects the first 6 most similar templates, the accuracy is the highest. So, in the experiments we did above, the first 6 most similar templates were selected to calculate the semantic distance in the input layer.

In addition, we tested whether the model performed equally well in both positive and negative reviews. As is shown in the following Fig. 8, the model performed better in positive reviews than in negative reviews.

Finally, we chose a famous text mining software in the market to compare with our model, which was developed by the leading large data semantic analysis application service provider in the Chinese field. In order to protect the developer's interests, the name of the text mining software is not provided here. We selected 10% comments to do comparative experiments. The experimental results show that the prediction accuracy of the text software is 0.846, and the prediction accuracy of our

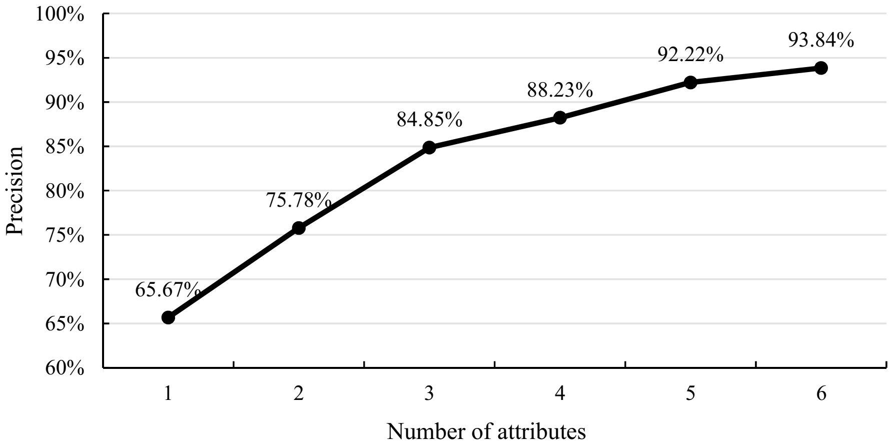

Fig. 6 Precision of the model when adopting different number of attributes

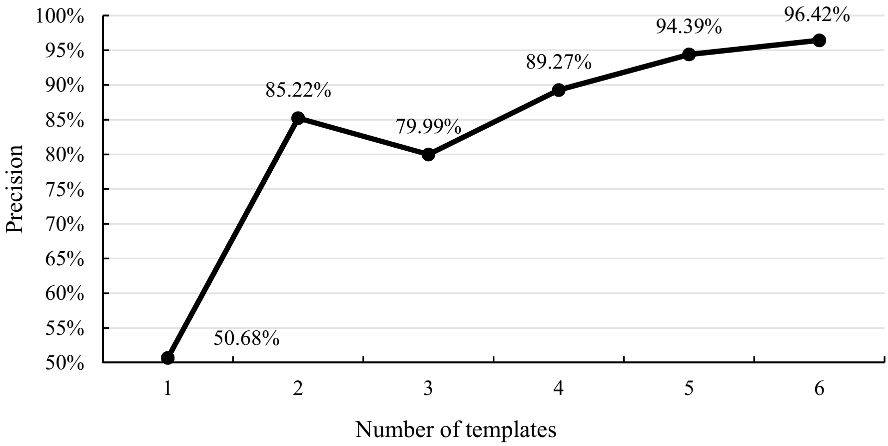

Fig. 7 Accuracy of the model when adopting different number of the most similar templates

model is 0.964. Therefore, the prediction accuracy of our model is higher than that of the software, and our model has market value.

5.3 Users' preferences results and product improvement strategies

The attribute classification results are shown in Table 5. Attributes that are identified as basic attributes include logistics, peripheral, service, memory, and CPU. These attributes can cause great dissatisfaction when they are bad, so they should be given a high priority in product improvement. The battery is an expected attribute whose variation has a positive linear relationship with satisfaction. Attractive attributes include boot speed, screen, function, temperature, price, appearance, and

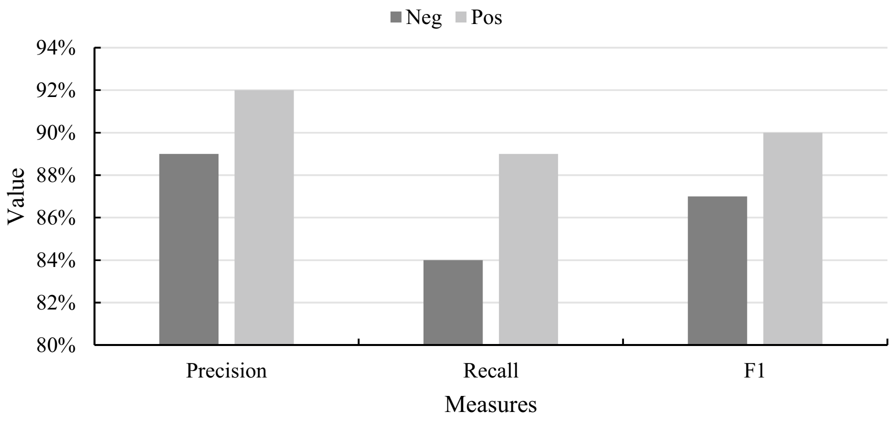

Fig. 8 Model performance on positive and negative reviews

Table 5 The table of classification results

Attribute	Classification	Attribute	Classification
Boot speed	A	Logistics	B
Screen	A	Peripheral	B
System	A	Service	B
Price	A	Memory	B
Appearance	A	CPU	B
Function	A	Battery	E
Temperature	A		

system. These attributes can greatly increase satisfaction when they are fulfilled. But they will not cause a serious decline in satisfaction when they are not. As such, enterprises should first meet the needs of the basic attributes and expected attributes. When adequate funding and technical support are available, enterprises can meet the demands of attractive attributes furtherly.

6 Conclusions and discussion

In this study, our goal is to mine consumers' preferences from incomplete online reviews source for the purpose of product improvement. To this end, CEMM is proposed to realize the patch of comment information to find valuable customer sentiment information, and based on which, the KANO model is developed to support the designers to acquire preference information from the perspective of products enhancement and design. The practical implications of this study are provided as follows.

First, our study shows that online user-generated reviews can be a valuable and exploitable source for business insights discovering. In the era of big data commerce, data-driven approach can provide insights into the market and consumers for enterprises. With regard to the rich content and high reliability of big data of online comments, many scholars have made attempts to exploit them for managerial inspirations and innovations. Unlike previous studies which focus on revealing online reviews' influence on consumers' purchase decisions, we focus on using online reviews to develop strategies for improving product design.

Second, our study proposes a methodology to mend the incomplete and implicit information in the online review. Mining valuable information from the unstructured, incomplete, and great volume of online reviews has great practical significance. We creatively design a set of standard sentiment templates of online comments for product attributes to portray all possible users' comment patterns. The standard sentiment templates we designed are applicable to all reviews of consumer electronic products. The structure of the template is the same for different products, but the attribute words, emotion words and evaluation words need to be changed according to the research object. Therefore, our template is both flexible and fixed. And the SSDMs are proposed to map the online comments into the specific template initially based on their semantic similarity. Subsequently, the FGMNNs are designed to further match the comments of attributes and their standard comment templates according to the user's overall emotions towards the product. Thus, customer sentiment on product attributes is obtained.

Third, to support product improvement decisions, our research provides a practical procedure to mine users' preferences from online comments. We combine the CEMM model and traditional KANO model to explore customer preference information. Based on users' sentiment polarity on product attributes obtained by CEMM, we introduce the KANO model to extract the attributes that users value most and the ones that users pay little attention to, and accordingly provide smarter, more profitable suggestions for product improvement. Although we choose laptops as research object, our models are generic and can easily be applied to other industries, such as other digital products.

Utilizing comments of laptop product crawled from JD.com, the effectiveness of our proposed model was validated and some practical suggestions can be given. Logistics, peripheral, service, memory, and CPU are must-be and can cause great dissatisfaction when not well-performed, so they should be given the first priority in product improvement. However, excessive resources input is completely unnecessary when their performances have met the expectations of consumers. As for the expected attribute, the quality of the battery should be guaranteed to satisfy consumers. And if resources permitted, improving battery quality should always be valued. Attractive attributes, such as boot speed, screen, function, temperature, price, appearance, and system, are likely to become competitive factors that differentiate products and catch target consumers' eyes. If permitted, the enterprise should actively invest resources to improve some of these attributes according to its market positioning to attract target customers.

Our study has some limitations that warrant additional research attention. First, our study is based on the assumption that all consumers are homogeneous. But as the customer segmentation theory shows, customers in different market segments may have various types of preferences. Therefore, conducting a consumers clustering process before making a preference analysis may help to improve our model and is worthy to study in the future. Second, in this study, we assume that all online reviews which are available on e-commerce platforms are generated by real consumers and express their real feelings and points. In other words, so-called user-generated reviews are not manipulated or untruthful. However, spam reviews can easily be found on e-commerce platforms and identifying fake reviews is a growing research issue [5]. A spam review detection process may contribute to identify truthful user-generated reviews and, thus, help to mine the genuine preference of consumers.

Acknowledgements This work was supported by the Chinese National Natural Science Foundation (No. 71871135).

References

1. Gartner Research (2018). Market trends: Social analytics show consumers want more health features from their Wearables. https://www.gartner.com/en/documents/3871126
2. Lee, T., & Bradlow, E. T. (2007). Automatic construction of conjoint attributes and levels from online customer reviews. *University Of Pennsylvania, The Wharton School Working Paper*.
3. Ghose, A., Ipeirotis, P., & Sundararajan, A. (2007). Opinion mining using econometrics: A case study on reputation systems. In *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 416–423).
4. Chamorro-Mera, A., & Palacios-González, M. M. (2019). Socially responsible investment: An analysis of the structure of preferences of savers. In *Corporate social responsibility and environmental management*.
5. Xiao, S., Wei, C. P., & Dong, M. (2016). Crowd intelligence: Analyzing online product reviews for preference measurement. *Information & Management, 53*(2), 169–182.
6. Thomke, S., & Von Hippel, E. (2002). Customers as innovators: A new way to create value. *Harvard Business Review, 80*(4), 74–85.
7. Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box-office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science, 29*(5), 944–957.
8. Dellarocas, C., Zhang, X., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing, 21*(4), 23–45.
9. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1–167.
10. Amplayo, R. K., Lee, S., & Song, M. (2018). Incorporating product description to sentiment topic models for improved aspect-based sentiment analysis. *Information Sciences, 454–454,* 200–215.
11. 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). ACM.
12. Meena, A., & Prabhakar, T. V. (2007). Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis. In *European conference on information retrieval* (pp. 573–580). Berlin: Springer
13. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on empirical methods in natural language processing-volume 10* (pp. 79–86). Association for Computational Linguistics.
14. Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. *Expert Systems with Applications, 36*(3), 6527–6535.
15. Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications, 57,* 117–126.
16. Grandi, U., Loreggia, A., Rossi, F., & Saraswat, V. A. (2014). From sentiment analysis to preference aggregation. In *ISAIM*.
17. Decker, R., & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing, 27*(4), 293–307.

18. Zhao, K., Cong, G., Yuan, Q., & Zhu, K. Q. (2015). SAR: A sentiment-aspect-region model for user preference analysis in geo-tagged reviews. In *2015 IEEE 31st international conference on data engineering* (pp. 675–686). IEEE.
19. Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science, 57*(8), 1485–1509.
20. Li, Y. M., Chen, H. M., Liou, J. H., & Lin, L. F. (2014). Creating social intelligence for product portfolio design. *Decision Support Systems, 66,* 123–134.
21. Qi, J., Zhang, Z., Jeon, S., & Zhou, Y. (2016). Mining customer requirements from online reviews: A product improvement perspective. *Information & Management, 53*(8), 951–963.
22. Löfgren, M., & Witell, L. (2005). Kano's theory of attractive quality and packaging. *Quality Management Journal, 12*(3), 7–20.
23. Sharif Ullah, A. M. M., & Tamaki, J. I. (2011). Analysis of Kano-model-based customer needs for product development. *Systems Engineering, 14*(2), 154–172.
24. Tan, K. C., Xie, M., & Shen, X. X. (1999). Development of innovative products using Kano's model and quality function deployment. *International Journal of Innovation Management, 3*(03), 271–286.
25. Senthil Kumar, P. S., Balasubramanian, S., & Suresh, R. K. (2012). Pairing of intelligent design concept method and Kano model for product development. *Australian Journal of Mechanical Engineering, 10*(2), 91–99.
26. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics, 5,* 135–146.
27. Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543).
28. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint* arXiv:1301.3781.
29. Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint* arXiv:1404.2188.
30. Zhang, Y., & Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint* arXiv:1510.03820.
31. Yin, W., Schütze, H., Xiang, B., & Zhou, B. (2016). Abcnn: Attention-based convolutional neural network for modeling sentence pairs. *Transactions of the Association for Computational Linguistics, 4,* 259–272.
32. He, H., Gimpel, K., & Lin, J. (2015). Multi-perspective sentence similarity modeling with convolutional neural networks. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1576–1586).
33. Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint* arXiv:1408.5882.
34. Nguyen, D. Q., & Verspoor, K. (2019). From POS tagging to dependency parsing for biomedical event extraction. *BMC Bioinformatics, 20*(1), 72.
35. Bach, N. X., Linh, N. D., & Phuong, T. M. (2018). An empirical study on POS tagging for Vietnamese social media text. *Computer Speech & Language, 50,* 1–15.
36. Fan, X. (1996). *Grammar view of three planes*. Beijing: Beijing Language Institute Press.
37. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).
38. Kusner, M. J., Sun, Y., Kolkin, N. I., Weinberger, K. Q. (2015). From word embeddings to document distances. In *International conference on machine learning*. JMLR.org, pp. 957–966.
39. Pele, O., & Werman, M. (2009). Fast and robust earth mover's distances. In *2009 IEEE 12th international conference on computer vision* (pp. 460–467). IEEE.
40. Ren, Z., Yuan, J., & Zhang, Z. (2011). Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In *Proceedings of the 19th ACM international conference on multimedia* (pp. 1093–1096). ACM.