PRODUCT COMPETITIVENESS AND SENTIMENT ANALYSIS BASED ON DSM CLUSTERING OF CHINESE ONLINE COMPARATIVE COMMENTS

Fan YANG[1,2], Luokai HU[1,2,*], Yu CHEN[1,2]

Mining online customer reviews is crucial for analyzing product competitiveness. The effective identification of comparative comments is one of the important prerequisites for influencing marketing decision-making. Sometimes, customers publish a number of comparative reviews after they purchased products, especially if they bought similar products before. Mining these comparative reviews requires a sentimental mining method. A comparative sentence in a review can be transformed into a sentiment score of a particular feature of both compared products. The purpose of this paper is to calculate such emotional scores of product features using comparative sentences and to design a clustering method for analyzing the hierarchical relationship between different brands. To ensure better accuracy of the unsupervised algorithm, an improved computing model based on the sentiment dictionary has been used to obtain weighted sentiment scores. Then, these scores were entered into a design structure matrix, which was used for clustering different brands of similar products. Experiments with a sample of Taobao customer reviews showed that the proposed method can analyze comparative relationships more accurately than conventional methods. This developed design structure matrix is not only suitable for engineering applications, but also for product clustering.

Keywords: comparative opinion mining, DSM clustering, sentiment analysis, Chinese natural language processing, customer preferences

1. Introduction

For humans, making comparisons is a basic method that facilitates understanding. By comparing various things, people can identify similarities and differences in different attributes and can often identify the reasons for these. Especially in the commercial field, comparison is an important method for testing the competitiveness of products [1-3]. Consumers often learn about a product by comparing it with similar products, because a deeper understanding of the quality and performance of a product is often gained by comparison. Sellers also identify the competitive advantage and deficiency of a product by comparing, which provides the basis for a marketing strategy and product improvement.

[1] College of Computer, Hubei University of Education, Wuhan, 430205, China
[2] Hubei Education Cloud Service Engineering Technology Research Center, China
[*]Corresponding author' e-mail: huluokai@hue.edu.cn

Among product purchase reviews on the Internet, comparison comments include comparative information on a product feature, compared between two different brands. In contrast to other comments, comments that include comparative sentences are most important for the mining of both advantages and disadvantages of products. They are useful to analyze the competitiveness of products in the market. Therefore, data mining of comparative comments increasingly attracts the attention scholars. Many scholars have conducted sentiment analysis and obtained comparative data of products at an attribute level.

Based on Chinese text sentiment analysis and comparative text mining, the present study proposes a dictionary-based method for sentiment analysis of Chinese comparative comments. Furthermore, a design structure matrix (DSM) of comparative products is designed to realize DSM clustering. Finally, the competitive relationship among different brands was analyzed with a division for similar products obtained from DSM clustering.

2. Related Work

Research of comparative comment mining is derived from relationship extraction technology in text mining. Relationship extraction is an important branch of information extraction, and a method that can identify related entities from text [4]. This method applies four types of achievement processes: supervised learning, semi-supervised learning, unsupervised learning, and remote supervised learning. Supervised learning is a procedure with strong manual participation, while unsupervised learning is an automatic processing procedure without the involvement of human behavior. Semi-supervised learning is a procedure between supervised and unsupervised learning. Remote supervision represents supervised learning with the help of remote data.

The sentence structure consists of a comparative subject, comparative objective, comparative word, comparative point, and comparative result. In relationship extraction technology, sentiment analysis of comparative comments requires the following steps: comparative sentence recognition, the extraction of comparative factors (i.e., comparative subject, comparative objective, comparative word, and comparative point), and the calculation of comparative emotional scores (i.e., the comparative result).

DSM clustering is a method that combines clustering with symmetric matrix division. DSM is mainly applied to the production process of products, such as clustering of dependent product parts, auxiliary engineering management, production process model formation, and task assignment [5]. Its main function is the formation of a symmetric matrix with relationships between elements and the formation of a division of all elements after several steps.

2.1. Identification of Comparative Comments

Related research started with English comparative sentence recognition. For example, Jindal et al. proposed a method that combines the classifier and class sequence rules (CSR), and obtained a suitable F value [6]. In China, however, early research was conducted by Xiaojiang. He discussed the scope, the extension, and characteristics of Chinese comparative sentences, defined the task of Chinese comparative sentence recognition, and proposed a method for classifying Chinese sentences into two categories by using a support vector machine (SVM) classifier: comparative sentences and non-comparative sentences [7]. Huang significantly improved the accuracy and recall rate, and he developed a new model by using SVM as a classifier, that can discover features by feature words and CSRs, and can extract the entries by the conditional random field (CRF) algorithm [8].

The main task of comparative sentence recognition is to identify comparative sentences among various types of sentences using the recognition algorithm. Because of the flexibility of Chinese sentence patterns, the recognition algorithm for Chinese comparative sentences needs to be more complex than that for English comparative sentences. However, a number of scholars have proposed a simple algorithm to identify Chinese comparative sentences, especially for disparity sentences. The method identifies a sentence as a comparative sentence if the statement contains two comparative items and the word "bi" [9]. The present paper applies this simple and effective method to recognize comparative sentences.

2.2. Extraction of the Comparative Point

Comparative point extraction differs from the extraction of other factors. A comparative point in a sentence represents the feature that is compared between two items, such as “speed” in the following example: “This phone has a lower speed than an iPhone”. However, this feature is sometimes less obvious in a sentence, and is often replaced by other words, e.g., “fast” or “slow”, which indicate “speed”. Therefore, to conduct comparative point extraction, an appropriate algorithm is required.

At present, there are two main methods. The first method manually summarizes the feature dictionary, i.e., collects product features from product specifications, relevant experts, or test texts. Then, string matching is used to identify the matching feature words as well as the comparative points in a sentence [10]. The second method automatically mines for feature words. Because of the absence of a feature dictionary, this mining algorithm can identify and summarize feature words from a test data set, dynamically form a feature dictionary, and also apply the string-matching method to extract comparative points [11].

However, extraction of comparative points also needs to consider the expression of implicit features, i.e., features that are presented without a feature word in a comparison sentence, where the comparative point needs to be identified according to other relevant words. For example, “speed” can be derived from “fast” or “slow”.

Therefore, the method proposed in this paper incorporates such implicit feature words into the feature dictionary, maps them to corresponding feature words, and more accurately identifies comparative points from comparative sentences.

2.3. Sentiment Score Calculation

The calculation sentiment score is the third research direction of comparative comment mining. However, most approaches are merely a direct application of the common emotional scoring method to comparative sentences [12]. A simpler method is the determination of the emotional tendency of a sentence using the sentiment dictionary, the degree adverb dictionary, and the negative word dictionary. Currently, two methods are commonly used for emotional scoring. The first method only assesses the emotional polarity. A positive emotion is classified as +1, and a negative emotion is classified as -1 [12]. Another approach is to determine the degree of polarity after assessing the polarity itself. The influencing degree is determined by the emotional intensity of the emotional word itself as well as the degree adverb before and following the emotional word [13]. The present paper focuses on the intensity of the portrayed emotion; therefore, an improved version of second method has was used for scoring.

2.4. Design Structure Matrix Clustering

The design structure matrix (DSM) was originally proposed by Steward, as a matrix tool for planning and analyzing the product development process [14]. A DSM is a square matrix that consists of the same row and column elements (as shown in Table 1, where A, B, C, D, and E are row and column elements). Row and column elements represent the parts or production activities of the product. The white cells represent the relationship between the corresponding row and column elements. Cell i represents the connection of D to B, and cell ii represents the connection of B to D. The purpose of DSM clustering is the formation of modules that consist of closely connected product parts (or production activities). Strong convergence exists within the module and the connectivity between modules is as low as possible. This facilitates the modular production of products or the unitization management of production activities, thus simplifying the production process [15].

Table 1

Example of a Design Structure Matrix

	A	B	C	D	E	F
A	■					
B		■		i		
C			■			
D		ii		■		
E					■	
F						■

This paper adopts DSM clustering, which is suitable for analyzing product competitiveness. The row and column elements in the DSM represent products of the same type from different brands. A white cell in the matrix represents the sum of emotional scores (or close value) between corresponding products, e.g., cell i represents the evaluation value of product D to product B, and cell ii represents the evaluation value of product B to product D. The evaluation value is the result of the sum of the sentiment scores of all comparative reviews of both compared products. Then, using the DSM clustering model, closely related products are classified as one cluster, and a better partitioning solution can thus be obtained. Products within a particular cluster have a strong competitive relationship. Therefore, the results of clustering can be used to analyze the competition between similar products by different brands.

3. Methodology

3.1. Research Design

The model applied the following implementation steps: firstly, the evaluation data set from the network is crawled; then, after obtaining the statistical data, the DSM clustering implements preprocessing, word segmentation, and emotion dictionary training; finally, the competitiveness of the products is analyzed. Figure 1 shows the framework of this model.

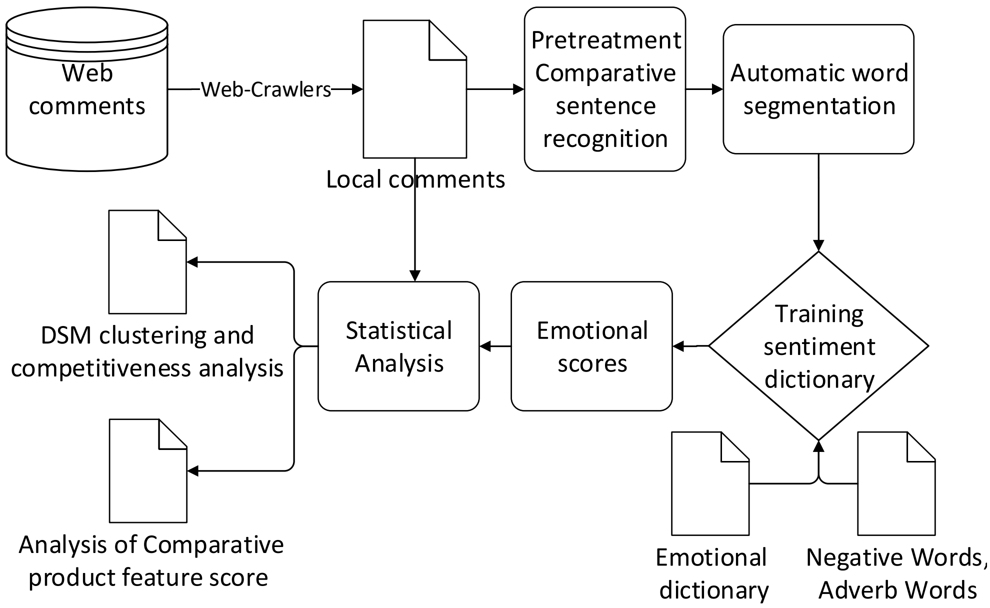

Fig. 1. Framework for mining comparative opinions based on online customer reviews

3.2. Data Collection

Taobao provides many customer reviews of the offered products. Based on the method of URL link recognition, the reviews of phones within the same price range are obtained for five brands from 15 online stores in Taobao, yielding a total of 26,481 reviews.

Table 2

Statistics for different online stores for phones within the same price range (partial data)

Type	Price	Store name	Number of evaluations obtained by the web crawler	Cumulative number of evaluations
Huawei honor/honour 9	2300-3000	Honor official flagship store 1	1979	58392
Huawei honor/honour 9	2300-3000	Honor official flagship store 2	1979	2859
Huawei honor/honour 9	2300-3000	Zhongkuan Communication Store	1242	1260
Huawei honor/honour 9	2300-3000	Red Japan Communications official authorized store	4999	5955

A batch-type web crawler was adopted, which limits the properties of the crawled content, including the scope of the crawl, the specific target, the crawl time, the amount of data, and the crawl page.

3.3. Comparative Sentence Recognition

After the online reviews were obtained, the second step was to identify the comparative sentences. A simple approach was adopted. When a sentence contains two product names and comparative words such as “bi” (“bi” is a Chinese word that can be roughly translated to “than”), it is classified as a comparative sentence [16,17]. Fig. 2 presents a sample of the obtained comparison statements.

biMeizu-Pro5-manlebushao	slower speed than Meizu PRO5
xinhaobiXiaomi-4s-qiangduole	a stronger signal than Xiaomi 4s
ganjuebipingguoyouxiu	better than iPhone

Fig. 2. Recognition results of comparative sentences (three representative examples)

3.4. Comparative Feature Extraction

Comparative features, or, more precisely, product attributes, are mainly used to identify the comparative point between two products using comparative sentences. For example, the comparative feature in “the performance of Huawei is much better than that of Samsung” is “performance”. The implementation of this extraction mainly includes manual acquisition and machine identification, or a combination of both. Since attributes of a product are limited, this method is often adopted to manually collect features. That is, product attributes that appear in the product manual or in the comments with high frequency are collected and manually classified to form a feature classification dictionary. Then, comparative points are extracted from a sentence using a method of string matching based on the dictionary.

Features of comparison objects are collected manually from the test comments and the specifications of phones. A total of 13 categories of features were designed. As shown in Table 3, each feature has its corresponding vocabulary. The 14th feature is a general evaluation, which represents the lack of obvious characteristic words, but the availability of an overall evaluation in the sentence. Then, the comment is classified as a general evaluation. The feature classification and its corresponding vocabulary are designed as a feature dictionary that can be used to extract comparative points from comments.

Table 3

Feature classification

Number	Classification	Corresponding words
1	Sudu (speed)	Sudu, kuai, man, liuchang (speed, fast, slow, smooth)
2	Xinhao (signal)	Xinhao, xinhaoqiangdu, shoujixinhao (signal, signal strength, cell phone signal)
3	Paizhao (photograph)	Pai, paizhao, shexiangtou, ganguang, zhaoxiang, xiangsu (take a photo, photograph, camera, photosensitive, photograph, pixel)
4	Xingneng (performance)	Xingneng, gongneng, xingjiabi (performance, function, cost performance ratio)
5	Xianshi (display)	Xianshi, fenbianlv, pingmu (display, resolution, screen)
6	Waiguan (appearance)	Waiguan, yanzhi, kuanxing, zuogong, gongyi (appearance, style, workmanship)
7	Jiage (price)	Gui, huasuan, pianyi, jiage (expensive, a good deal, cheap, price)
8	Wendu (temperature)	Fare, fatang, wendu (hot, phone heat dissipation temperature)
9	Haodian (power consumption)	Dianliang, haodian, haodianliang (power, power consumption, power consumption)
10	Jiesuo (unlock)	Jiesuo, zhiwen, zhiwenshibie (unlock, fingerprinting, fingerprint identification)
11	Jianpan (keyboard)	Jianpan (keyboard)
12	Yinzhi (sound)	Yinzhi (sound)
13	Taidu (service attitude)	Taidu (service attitude)
14	Zongping (general evaluation)	no corresponding words

3.5. Sentiment Score Calculation

The sentiment analysis module uses a dictionary-based analysis and sets both the polarity and weight for each emotional word. Considering the adverb and the negative dictionary, the sentiment score of the comparative sentence is computed with weighting. Therefore, the emotional dictionary is the key factor of the developed model. Therefore, a number of existing dictionaries were considered (e.g., the Knowledge Online Emotion Dictionary, the Taiwan University NTUSD, and the Tsinghua University Li Jun's derogatory Dictionary), and the industry characteristics, expressed by the trained text materials, were designed. In addition, to extract features, obtain comparison pairs, and calculate emotional weight, words included in the product names, attributes, feature classifications, negations, degree adverbs, and synonyms were also used. The degree adverb dictionary is divided into a prefix and a suffix dictionary. A number of examples of the dictionary are shown in Table 4.

Table 4

Examples of sentiment dictionary

Emotional word	Score	Emotional word	Score
Hao (good)	1	Qiangxie (strong)	0.5
Kuai (fast)	1	Ka (stuck)	-1
Man (slow)	1	Liangyidian (bright)	0.2
Gui (expensive)	-1	Geili (powerful)	1

The equations for calculating the sentiment score are presented in the following:

$$\text{Score} = (-1) * n * w1 * w2 \tag{1}$$

Or:

$$\text{Sco} = w3 \tag{2}$$

The value of *n* represents the number of negative words before the emotional word. The emotional word weight (w_1) was obtained from the sentimental dictionary, w_2 represents the weight of two adverbs before and after the emotional word, and w_3 represents the negative word weight. The first word or second before the emotional word may be an adverb, such as “hen\bucuo” (i.e., “very\not bad”), or “feichang\de\hao” (i.e., “very\\good”). An adverb can also appear after an emotional word, such as “hao\taiduo”, “hao\le\bushao” (i.e., “good\too much”, or “good\\not a few”). There was also the special case where a sentence only includes negative words without emotional words, such as “meifa” (i.e., “cannot”) in “meifa\bi” (i.e., “cannot\compare to”). Therefore, Eq. (2) is needed. Sentiment score calculation is based on the following assumptions:

Hypothesis 1: The weight of each sentimental word is not identical. The weighting is based on common knowledge and existing sentiment dictionaries (as mentioned above). For example, the weight of “hen” (i.e., “hate”) is higher than that of “taoyan” (i.e., “dislike”).

Hypothesis 2: The weights are linear. This is true in most cases.

Hypothesis 3: The weight of each adverb is different. For the influence of emotional intensity by negative words and adverbs, simple inversion and weighting are chosen. For example, the emotional intensity of “feichangxihuan” (i.e., “like very much”) is stronger than that of “tingxihuan” (i.e., “quite like”).

Based on the dictionaries of emotional words and related words, the sentiment score can be obtained by summing the scores from all sentences, and the total score of each group of two compared products for each feature can be computed as shown in Table 5. Here, the score of Huawei to Xiaomi on the camera function was 1.33, while in contrast, that of Xiaomi to Huawei was 0.2, which implies that Huawei is better than Xiaomi in total (1.33 - 0.2 = 1.13 > 0).

Table 5

Feature scores of a group of comparative brands

Product 1	Product 2	Feature	Score
Huawei	Xiaomi	Camera function	1.333333
Xiaomi	Huawei	Camera function	0.2
Huawei	Xiaomi	Phone performance	2
Xiaomi	Huawei	Phone performance	1
Huawei	Xiaomi	phone heat dissipation	1
Xiaomi	Huawei	phone heat dissipation	1

A number of other methods exist for calculating the sentiment score of compared text, e.g., support vector machine (SVM), SVM+N (which considers the name of the entity), pure keyword strategy (T), keyword pair co-occurrence strategy + sequential rule pattern (TP+CSR), and keyword pair co-occurrence strategy + sequential rule pattern + comparison entity secondary recognition (TP+CSR+N). The method used in this paper is superior to previously developed methods. Firstly, the first adverb before the emotional word is identified, and the second is also used for the calculation of sentiment scores. In addition, negative words are not simply identified from the whole sentence, but from the content before the emotional word. Moreover, the first or second word before the emotional word may adjust the emotional polarity. Therefore, the accuracy of the emotional recognition of a sentence could be improved compared to prior methods, as shown in Table 6.

Table 6

Performance metrics of methods

Methods	Accuracy	Precision	Recall	F-score
SVM	70.04%	59.59%	72.62%	63.61%
SVM+N	82.83%	72.73%	83.81%	77.88%
T	94.85%	63.55%	99.03%	77.42%
TP+CSR	97.12%	76.71%	97.17%	85.74%
TP+CSR+N	98.37%	85.99%	97.61%	91.44%
Scoring with one adverb	92.0%	93.7%	95.4%	94.5%
Scoring with two adverbs	92.7%	94.6%	95.5%	95.0%

3.6. Design Structure Matrix Clustering Implementation

i. Statistical analysis of sentiment scores

According to the contents of Tables 3 and 5, the 14th feature (i.e., general evaluation) is taken as example. Partial of the results are presented in Table 7.

Table 7

General evaluation scores of comparative brands (partial representation)

Brand 1	Brand 2	General evaluation score
Huawei	Vivo	1.5
Huawei	Xiaomi	1.409091
Huawei	Samsung	1.5
Huawei	Meizu	1.4
Xiaomi	Huawei	0.853846
Xiaomi	Samsung	2
Xiaomi	Oppo	1
Xiaomi	Nubiya	1.333333

Table 6 shows that Huawei is 1.409091 higher than Xiaomi, while Xiaomi is 0.853846 higher than Huawei. Overall, the mobile phone glory 9 by Huawei is perceived as substantially better than the mobile phone 6 of Xiaomi (1.409091 - 0.853846 > 0). This score is consistent with the mobile phone index ranking in Taobao, where Huawei ranked 7th and Xiaomi ranked 9th.

ii. DSM arrangement

Feature scores of each group identified above can be transformed to a matrix. Taking the 14th feature (i.e., general evaluation, as shown in Table 3) as example, eight group scores of comparison brands form a DSM, which is shown in Table 8.

Table 8

DSM for eight groups of comparison products

		Huawei	Vivo	Xiaomi	Samsung	Oppo	Meizu	Nubiya	Apple
		a1	a2	a3	a4	a5	a6	a7	a8
Huawei	a1		0.4	0.84	0.6	0.64	0.64	0.56	1
Vivo	a2	0.64			0.8			0.8	0.73
Xiaomi	a3	0.67	0.66		0.8		0.8	0.73	0.9
Samsung	a4	0.64	0.72	0.48		0.74		0.16	0.8
Oppo	a5		0.72	0.8					0.85
Meizu	a6	0.67	0.48					0.8	
Nubiya	a7			0.69		0.8			0.81
Apple	a8	0.72	0.65	0.75	0.79	0.65	0.69		

iii. Selection of clustering solution

Next, products can be clustered by the division method of row and column transformation. However, prior to the first step, row and column elements need to be divided into three categories: independent, bus, and common elements. An independent element implies that this element has no relationship with others. A bus element implies that it has a relationship with most other elements. The rest is the common element.

The specific steps of DSM clustering are described in the following [2]:

Step 1: Weak contact tear. In this step, specific elements are excluded from the matrix, which are weakly associated with other elements and have less impact on the final clustering results.

Step 2: Identification and separation of independent elements. In this step, individual elements are removed from the matrix.

Step 3: Row and column transformation. Row and column homogenous exchanges move the non-zero cells in the model as close as possible to the diagonal position of the matrix, while the relationship between elements remains unchanged.

Step 4: Identification of the bus element. Identify all bus elements and move them to the end of the queue of both row and column elements via homogeneous row and column transformation. Elements that are moved to the last position are grouped into a bus cluster.

Step 5: Cluster division. According to the intensity of non-zero cells in the DSM model obtained by the preceding four steps, non-bus elements are divided into several clusters, so that non-zero cells are included in the cluster as much as possible, while the number of clusters remains reasonable. All bus elements form a cluster. Each individual element forms a cluster. Weak elements that have been excluded also form a cluster.

Step 6: Calculation of the information flow and choice of the optimal solution. The calculation equations of information flow are presented in the following (the information flow of weak element clusters and independent element clusters are ignored). The optimal solution is the solution with the lowest *W* value:

Information flow within a cluster:

$$W_i^{(in)} = \frac{1}{2}(n_i - m_i)\sum_{l=0}^{(n_i-m_i)}\sum_{k=0}^{(n_i-m_i)}(d_{m_i+k,m_i+l} + d_{m_i+l,m_i+k}) \qquad (3)$$

Information flow inside a bus cluster:

$$W_b^{(in)} = \frac{1}{2}(n_b - m_b)\sum_{l=0}^{(n_b-m_b)}\sum_{k=0}^{(n_b-m_b)}(d_{m_b+k,m_b+l} + d_{b+l,m_b+k}) \qquad (4)$$

Information flow between two clusters:

$$W_i^{(out)} = \begin{cases} \alpha(N+1)(n_i - m_i + n_j - m_j)\sum_{k=0}^{(n_j-m_j)}\sum_{l=0}^{(n_i-m_i)}(d_{m_j+k,m_i+l}) \\ 0 \end{cases}$$

(5)

Information flow between a bus cluster and a non-bus cluster:

$$W_{b,j}^{(out)} = (n_b - m_b + 1)(n_b - m_b + n_j - m_j) \sum_{k=0}^{(n_b-m_b)} \sum_{l=0}^{(n_j-m_j)} \left(d_{m_b+k,m_j+l} + d_{m_b+l,m_j+k}\right) \quad (6)$$

$$W = \sum_{i=1}^{N} W_i^{(in)} + \sum_{i=1}^{N} \sum_{j=1}^{N} W_{i,j}^{(out)} \quad (7)$$

Where n and m represent the starting and ending position of the elements in any cluster, respectively, n-m represents the scale of the cluster, d represents the value of the cell in the DSM matrix, and α represents the adjustment coefficient, which is 0.8. Additionally, all subscript letters (i, j, k, l, and b) are sequence numbers of clusters and N is the total of all elements.

Different solutions can be obtained by the above steps and the following tree options were obtained:

Option 1: (a3, a7), (a4, a5, a2, a6), (a1, a8)
Option 2: (a2, a4, a3, a7), (a5), (a6), (a1, a8)
Option 3: (a2, a4, a3, a7, a5), (a6), (a1, a8)

Then, according to Eqs. (3)-(7), the information flow of each option can be calculated, as shown in Table 9.

Table 9

Information flow of options

Option	$W^{(in)}$	$W^{(out)}$	$W_b^{(in)}$	$W_b^{(out)}$	W
Option 1	5.89	36.48	0.86	93	136.23
Option 2	8.73	30.72	0.86	86.32	126.63
Option 3	17.74	12.42	0.86	119	150.02

The last column of Table 9 shows that the value of W of option 2 is smallest. Therefore, the best solution is option 2 – (Vivo, Xiaomi, Samsung, Nubiya), (Oppo), (Meizu), and (Huawei, Apple).

4. Results analysis

The DSM clustering model, based on the calculation of emotional scores, is used rarely for text mining. However, this is a helpful method for distinguishing between strong and weak comparative product competitiveness levels. The results of Section 3.6 are shown in Table 10. Cells with the same color are one group, and a total of four groups were obtained.

Table 10

Product DSM clustering results

		Vivo	Samsung	Xiaomi	Nubiya	Oppo	Meizu	Huawei	Apple
		a2	a4	a3	a7	a5	a6	a1	a8
Vivo	a2		0.8		0.8			0.64	0.73
Samsung	a4	0.72		0.48	0.16	0.74		0.64	0.8
Xiaomi	a3	0.66	0.8		0.73		0.8	0.67	0.9
Nubiya	a7			0.69		0.8			0.81
Oppo	a5	0.72		0.8					0.85
Meizu	a6	0.48			0.8			0.67	
Huawei	a1	0.4	0.6	0.84	0.56	0.64	0.64		1
Apple	a8	0.65	0.79	0.75		0.65	0.69	0.72	

Table 10 shows that Huawei and Apple are first echelon companies that belong to a bus cluster, as most cells in the last two rows and columns in the table are filled with scores. Both also have higher scores than others (e.g., cell (3,9) = 0.64 > cell (9,3) = 0.4). The overall evaluation of these two exceeds that of other brands, and both have a strong competitive relationship.

The second echelon companies include Vivo, Xiaomi, Samsung, and Nubiya. Specifically, Vivo mainly compares with Samsung and Xiaomi. Samsung mainly compares with Vivo and Xiaomi. Xiaomi mainly compares with Samsung and Nubiya, while Nubiya compares with Vivo, Samsung, and Xiaomi.

The third and fourth echelon companies are Oppo and Meizu, respectively. Meizu has poor performance, and basically does not have a competitive advantage in any feature. However, the camera function of the phone from Oppo is perceived as remarkable.

The results of the above analysis are quite similar to the actual situation and more accurate than the method of user ratings on the web platform. If DSM is analyzed according to each feature (for a total of 14 features, as shown in Table 3), product clustering results for different features can be obtained. This paper does not list these individually.

However, prior research only used compared emotional scores to analyze the advantages of specific features among the same type of products of different brands. For example, network diagrams were used to display the comparison results, while other studies used radar charts or other visualization techniques [9-13]. However, although DSM is a clustering method applied in the engineering field, it may be a superior tool for dividing products based on compared sentiment scores. The results also show that DSM is suitable for the competitiveness analysis of brands.

5. Conclusion

This paper proposes a DSM clustering model based on comparative sentiment scores between similar products of different brands. The clustering divides brands into four groups according to the strength of their competitiveness (from high to low). The method for sentiment score calculation is superior to conventional methods and the clustering result is consistent with reality.

The limited amount of available data and the insufficient number of competitive pairs affect the validity of the result to some extent. The dictionary-based sentiment analysis may be changed to machine learning to further improve the accuracy of the calculation of sentiment scores. Further improvements can be made at a later stage.

Acknowledgements

This work was partially supported by the special program for guiding local science and technology development by the central government (No. 2019ZYYD012), the Fund of Excellent Youth Scientific and Technological Innovation Team of Hubei University (No. T201818), and by research start-up funding of Hubei University of Education (No. 201817; web course comments mining in data driven education).

R E F E R E N C E S

[1]. *Kasturi Dewi Varathan, Anastasia Giachanou, Fabio Crestani*, Comparative opinion mining: A review. Journal of the Association for Information Science and Technology, 2017, 68(4).

[2]. *Mai M. Eldefrawi, Doaa S. Elzanfaly, Marwa S. Farhan, Ahmed S. Eldin*, Sentiment analysis of Arabic comparative opinions. SN Applied Sciences, 2019, 1(5).

[3]. *Asad Ullah Rafiq Khan, Madiha Khan, Mohammad Badruddin Khan*, Naïve Multi-label classification of YouTube comments using comparative opinion mining. Procedia Computer Science, 2016, 82(2016), 57-64.

[4]. *Xu Jian*, A Survey of Technical Methods for Entity Relationship Extraction. Modern Library and Information Technology, 2008, (8): 18-23.

[5]. *Tang Dunbing, Qian Xiaoming, Liu Jiangang*, Product design and development based on design structure matrix DSM. Science Press, 2009.

[6]. *Nitin Jindal, Bing Liu*, Identifying comparative sentences in text documents. In Proceedings of SIGIR 2006. Washington, USA, 2006: 244-251.

[7]. *Huang Xiaojiang, Wan Xiaojun, Yang Jianwu*, Research on Chinese Comparative Sentence Recognition. Chinese Journal of Information Science, 2008, 22(5): 30-37.

[8]. *Huang Gaohui, Yao Tianfang, Liu Quansheng*, Chinese comparative sentence recognition and relation extraction of CRF algorithm. Application Research of Computers, 2010, 27(6): 2061-2064.

[9]. *Zhang Z, Guo C, Goes P*, Product comparison networks for competitive analysis of online word-of-mouth. ACM Trans-actions on Management Information Systems, 2013, 3(4): 1-22.

[10]. *Wang Fengxia*, Research on comparative sentence recognition and perspective feature extraction method. Taiyuan: Shanxi University, 2013.

[11]. *Zhou Hongzhao, Hou Mingwu, Hou Min, et al.*, Research on Comparative Sentence Recognition and Comparative Feature Extraction Based on Semantic Classification. Chinese Journal of Information, 2014, 28(3): 136-141,149.

[12]. *Wang Hongwei*, *Jiang Wenying*, Gao Song, Hua Jin, Mining comparative opinions for competitive analysis by sentiment analysis: Evidence from online restaurant reviews. Journal of the China society for scientific and technical information, 2015, 34(12): 1259-1269.

[13]. *Wu Chen, Wei Xiangfeng*, Identification and Tendency Analysis of Comparative Sentences in User Evaluation. Computer Science, 2016, 43(6A): 435-439.

[14]. *Steward D V*, The design structure system: a method for managing the design of complex system. IEEE Transactions on Engineering Management, 1981, 28(3): 71-74.

[15]. *Matthias Kreimeyer, Udo Lindemann*, Complexity metrics in engineering design: managing the structure of design processes. Springer, 2011.

[16]. *Jindal N, Liu B*, Mining comparative sentences and relations. The Association for the Advancement of Artificial Intel-ligence, Boston: AAAI Press, 2006, 22: 1331-1336.

[17]. *Huang G H, Yao T F, Liu Q*, Mining Chinese comparative sentences and relations based on CRF algorithm. Application Research of Computers, 2010, 27(6): 2061-2064.