Computers in Industry 108 (2019) 1–11

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

A proposal for Kansei knowledge extraction method based on natural language processing technology and online product reviews

Yiru Jiao[a], Qing-Xing Qu[b,*]

[a] Department of Management Science and Engineering, School of Management, Harbin Institute of Technology, Harbin 150001, PR China
[b] Department of Industrial Engineering, School of Business Administration, Northeastern University, Shenyang 110167, PR China

ARTICLE INFO

Article history:
Received 2 November 2018
Received in revised form 9 February 2019
Accepted 12 February 2019
Available online 22 February 2019

Keywords:
Natural language processing (NLP)
Online product reviews
Sentiment analysis
Kansei engineering
Product design

ABSTRACT

With the rapid development of the economy, product design has gradually shifted to emotional design that focuses on satisfying users' emotional needs. Kansei engineering is the commonly used method in product emotional design, the first and vital stage of which needed to be addressed is the acquisition of Kansei knowledge. Considering the development of natural language processing technology and online shopping, a computerized method to extract Kansei knowledge from online product reviews is firstly proposed in this article, and a relational extraction method to establish the relationship between product features and user perceptions is further provided. This article analyzes and extracts the Kansei words of 10 mice respectively using the proposed computerized method, taking the mouse as the case study. Then three evaluation indicators including diversity, effectiveness, and concentration are defined to assess the method, which evaluates the superiority with the advantage of 19.03% in diversity, 6.91% in effectiveness, 22.18% in the concentration and 8.9 times higher in the total score compared with traditional method. Furthermore, taking the best-selling mouse for example, the relational extraction method is applied to extract the relationship between the user concern and the user attitude, establish the relational table, draw Kansei knowledge tree, and finally model connection between product features and user perceptions. By utilizing natural language processing technology and integrating Kansei engineering, linguistics and computer science, it could be considered that the results of this article can accelerate the traditional user survey process, clarify users' emotional needs, guide the adjustment of product design, and assist the user-centered product emotional design.

1. Introduction

The key to successful new product development is that its design meets users' needs. With the increasing improvement of the economy, users' needs of products gradually not only stay in the using value, but also in the rich psychological and emotional needs contained in products, which have attracted more and more attention [1]. In the market environment where product functions are becoming more and more homogeneous, whether users' understanding of the product meets their potential emotional needs is playing a more important role in product competition and becomes a significant manifestation of product design value [2]. As a vital factor to meet high-level spiritual needs, emotion has become a new research goal in product design [3].

For designers, it is difficult to straightly obtain users' emotion about the product. To satisfy users' needs, the physical elements of the product need to be linked to users' perception about the product, which is the problem what emotional design meets and solves [4]. One of the techniques of emotional design is Kansei engineering developed in Japan in the year 1970. The Kansei that is user's impression and emotion in English means users' feelings about the product including the design, size, color, function, feasibility, price, and so on [5]. Kansei engineering is considered to be the most reliable and useful methodology to handle users' emotional needs and has spread out in the world [6].

In the early stages of Kansei engineering, questionnaire surveys that consists of a list of Kansei attributes and product attributes are commonly used to obtain commodity properties and user experience [7]. Traditional methods provide a high-quality result, but they have weakness of small-scale and one-time, leading limitations in data scale, data updating and collection efficiency [8]. With the development of computer and internet technology, the function of the computer network as information exchange media has been promoted rapidly and the trend of online shopping

* Corresponding author at: Department of Industrial Engineering, School of Business Administration, Northeastern University, NO. 195, Chuangxin Road, Hunnan District, Shenyang, 110167, PR China.
E-mail address: quqingxing@gamil.com (Q.-X. Qu).

https://doi.org/10.1016/j.compind.2019.02.011
0166-3615/

constantly impacts the consumption pattern, making it undergo a significant transformation in volume and structure [9]. Nowadays, people can choose their favorite products in the online shopping mall, browse other consumers' comments on the products, and also leave their opinions. These data constitute an information network linked commodity properties and user experience, which buries rich information treasures. Therefore, mining information from online product data has drawn much concern.

Methods proposed in this paper for extracting Kansei knowledge from online product reviews bases on Kansei engineering, linguistics and computer technology, starting from deep semantic and syntactic structure to explore user's emotional needs to assist emotional design. It analyzes the users' perceptual understanding from online product reviews and digs out the users' emotional needs, which can greatly improve the efficiency of research related to the product emotional design. As shown in Fig. 1, at first, we should select research objects of the suited product from O_1 to O_n. Single reviews r_1 to r_m under each product O_1 to O_n form the review vectors R_1 to R_n, and then users' product images excavated from which integrate Kansei knowledge vectors K_1 to K_n for the product O_1 to O_n.

In this article, the 2nd section introduced some related work about Kansei engineering and sentiment analysis. In Section 3, this article explains what product is suitable for and introduces the extracting methodology. Section 4 experiments it with the mouse and evaluates the results. Section 5 discusses its possible significance and limitations. At last, Section 6 summarizes the work results of this paper.

2. Related work

2.1. Kansei engineering

Kansei engineering is a comprehensive subject that integrates design science, ergonomics, and engineering. It requires designers to establish a bridge between users' psychological feeling and product design solution [10] (Fig. 2). The first of the four points concerning the technology Nagamachi proposed in 1995 is how to grasp the users' Kansei knowledge about the product [11]. How to accurately measure, locate and capture them is the primary problem to be solved in product emotional design, and also the premise to ensure that the product can be accepted and satisfied emotionally by consumers. As the emotional needs of users for products are implicit, imprecise and ambiguous [12], how to accurately express users' perceptual understanding of products and how to integrate users' emotional needs into product design, has become the focus and difficulty for academic research and product development.

Users' emotional needs for products are based on product features on the one hand and functional experience on the other [13]. It is an information processing process to transform users' emotional needs into cognitive images of products. It mainly relies on users' observation, compared with personal experience and psychological structure, and then makes a judgment by reasoning

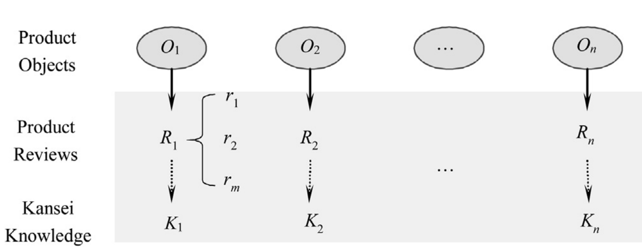

Fig. 1. Grabbing reviews and extracting Kansei knowledge.

and thinking [14]. The traditional modeling method depends on experiments involving humans and on subsequent analyses, requiring a lot of time and effort [15]. In the early stage, the extraction of Kansei knowledge was mainly based on psychological experiments, which were collected through comprehensive methods such as questionnaire survey, interview combined with verbal protocol analysis, semantic differential method, concept sketch, and image scale method [16]. Later, with the development of various scientific measurement instruments, researchers measured users' physiological data with the help of electroencephalogram (EEG), eye tracker, functional magnetic resonance imager (fMRI), and other devices, so as to judge the image of users' perception according to EEG signals, eye movement signals and blood oxygen content in the brain region [4].

2.2. Sentiment analysis

Sentiment analysis, also known as opinion mining, is a type of problem in natural language processing (NLP). The term sentiment analysis first appeared in a paper by Nasukawa and Yi [17], and Dave et al. [18] presented the idea of exploring the concept. Its goal is to analyze people's views, sentiments, evaluations, attitudes, and emotions on entities and their attributes from the text [19]. Therefore, a variety of research, including e-learning [20], health care [21], economy, and management [22,23], have applied sentiment analysis.

Some studies have been conducted for online review emotion analysis and ranking the products by online reviews [24–26]. On the one hand, many of them mainly focused on polarity classification, determining whether the text expresses a positive or negative (or neutral) opinion, which lost the specifics of emotions the text delivered, so that is not enough for Kansei engineering. On the other hand, they are mainly based on the relevant research on word segmentation, which is a one-dimensional and flat method.

Language is a system where all terms are interdependent and where the value of one is the result of the simultaneous presence of the others. The expression and understanding of natural language are stratified, and the semantics are the true expressions buried under the statement. Wanner [27] experimented and found that when people use language, no matter what they read or hear, as long as there is no special requirement, they can only remember its meaning, that is, the underlying proposition rather than its original words. It proved that the surface form of a sentence is stored separately from the deeper meaning. The current research on the understanding of emotions in the text is still conducted on part-of-speech tagging [28,29], which is far from the process of human brain understanding and limited by the bag of words. In this way, even the most efficient algorithms perform poorly, if not properly trained or when contexts and domains change [30]. Moreover, this method has high requirements for the integrity of the word library, and it is even more helpless for Internet neologism and non-standard words.

How to carry out sentiment analysis based on semantics is still a subject discussed by many researchers. In recent years, such researches are gradually setting sentiment analysis as interdisciplinary fields between mere natural language processing and natural language understanding by gradually shifting from syntax-based techniques to more and more semantics-aware frameworks, which consider both conceptual knowledge and sentence structure [31].

2.3. Kansei mining

Mining Kansei knowledge through natural language processing technology is not a brand-new idea never being considered. There

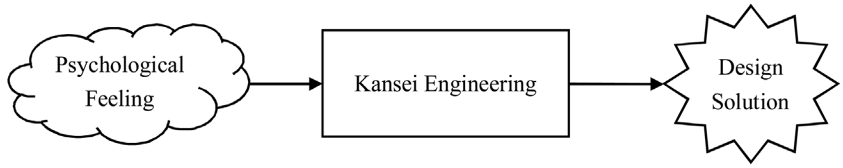

Fig. 2. Kansei engineering system (KES) adapted from Nagamachi.

have been some studies tried mining emotional information for Kansei design, they mainly concentrated on improving the data scale and collecting speed, such as using online questionnaires [8], evaluating words with word classes and dictionaries [15], and extracting affective information through text mining [32]. All of them mainly considering quantity weakness such as data scale and emotion dictionary, are still inadequate to promote the efficiency of both quantity and quality.

As studies about sentiment mining overviewed in Section 2.2, emotions in sentences are not isolated as a word and another word but related to the context, which is just the current research direction of natural language processing. Therefore, the first major concern of this study is to extract Kansei knowledge from the emotional phrases decomposed based on the context of review texts (including emotions and their subjects). Furthermore, this paper has noticed the connection between emotions and the subjects corresponding to them when analysis of the extracted results in the current research is still mainly performed by frequency statistics and structured the connection by generating Kansei knowledge tree. To sum up, the research results of this paper contributes to obtaining users' concerns about the product and their attitudes towards various concerns, so as to draw the users' image of the product.

3. Methodology

3.1. Suitability determination

Similarly with us, Wang et al. [33] have extracted affective words from online reviews. Their extraction is firstly based on the words collected from published articles in the field of Kansei design and secondly on probabilistic part-of-speech tagger, which is limited by the research products with existing studies. In this paper, the extraction is based on the semantics of the reviews rather than the existing words base. Therefore, it is necessary to determine whether the research product is suitable for the method proposed in this paper before the extraction.

It is not suitable for all the products to extract Kansei knowledge from online reviews. For some products like daily necessities, the use value of them accounts for a large proportion of the product value, so that the perceptual experience has little influence on the popularity of the product. Some products like office supplies are less customized and have very vague design features, for them, it is difficult to carry out product design from the perspective of Kansei engineering. However, for some products highly customized like assembled computers, it is difficult to reflect users' specific understanding of products in online reviews. Also, some products' value is reflected in the interaction with users, such as books and other service products. In this situation, users have varied different emotional experience of the products so that the statistical significance of which is not obvious.

3.2. Knowledge extraction

In conventional deep learning for sentiment mining, distinguishing between facts and opinions may be one of the most important subtasks [34]. In this article, considering designers needs of not only emotional evaluations but also product characteristics, we extract users' perception of products including both objective facts and psychological opinions. We design five main steps for the extraction of single Kansei knowledge (as shown in Fig. 3): corpus cleaning, Dependency Parsing (a technique in NLP), emotional phrase searching, negative emotion modifying, and frequency counting.

3.2.1. Corpus cleaning

First of all, the product reviews need to be crawled. Usually, the original reviews are mixed with a large number of invalid and unclear information, in which there is a limited amount that can be useful for consumer or business analysis. Therefore, the original reviews need to be cleaned to improve the sample quality. The cleaning rules are as follows.

- Eliminate repeated reviews. There are three cases of repeated reviews. The first one is the duplicate comments when the same user buys more than one piece. The second is the "water army" behavior of different users, which creates some trend; The third is the system's default comment.
- Filter simple reviews. A simple review is a word or phrase that simply expresses a general emotion, such as "good", "like it", "trash", and so on. Although it expresses users' overall evaluation of the product, it has nothing to do with the specific image extraction studied in this paper.
- Delete product information reviews. Product information review such as the product brand, model and so on is a kind of meaningless comments found in this paper when observing the data collected from product web pages, which are junk information to be deleted.

3.2.2. Emotional phrases searching utilized dependency parsing

In cleaned product reviews, there are generally three common forms. In the first type, firstly, the advantages and disadvantages of

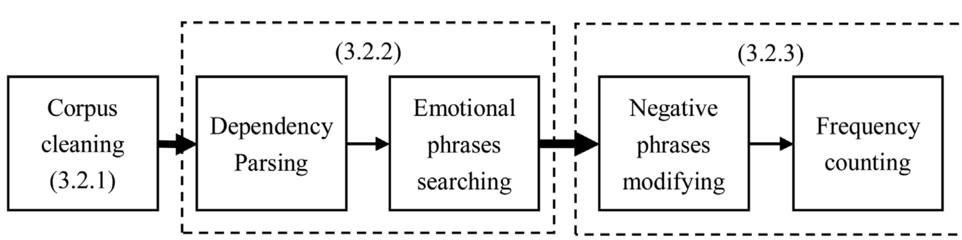

Fig. 3. Five steps for single Kansei knowledge extraction.

the product are briefly and concisely explained, and then the detailed information is described. The second type lists user's viewpoints, which directly describes the detailed cognitive information according to the advantages and disadvantages. The third type is mixing with both positive and negative comments with no regular format. The first two types belong to the semi-structured review structure, which is the user's consumption attitude towards the product attributes by clear emotional inclination. While the third is free and disorderly, whose emotional tendency and product attributes to be evaluated are confused or even unknown, with the lowest degree of structure and the most difficult to deal with.

Although the three types of reviews have a lower or higher structured degree, the analysis object of extracting the Kansei knowledge is the emotional phrase containing the knowledge. In other words, the level of emotional analysis reduces from the paragraph to the phrase, which is not affected by the paragraph structure. Therefore, the way of dealing with the reviews in this paper is first to divide the review paragraphs into emotional phrases, and then extract the Kansei knowledge.

To accomplish the work here are many famous NLP platforms such as GATE [35], UIMA, and NLTK [36]. These platforms have a common problem: they all emphasize the architecture of the system itself and lack precise language analysis technology, especially the analysis technology of non-morphological language [37]. In the past, grammar has been dominant in the expression and processing of natural language [38]. However, in this paper, the extraction of Kansei knowledge is not to find out the adjectives in the sentence directly, but to analyze the dependent phrase containing the user's emotion. Therefore, this paper selects the language technology platform (LTP) launched by the research center of computing and information retrieval of Harbin Institute of Technology, whose architecture is shown in Fig. 4 [39]. Compared with other toolkits for NLP, LTP provides a high order graph-based method for dependency parsing [40]. Many researches have applied LTP to identifying features [41], mining opinions [42,43], and detecting sentiments [44–47].

The key technology of LTP is dependency parsing (DP), which is the core part used in this paper to identify the Kansei knowledge. The DP reveals the language structure by analyzing the dependencies between the components in the unit of review, transforms the review statement from a linear sequence to a structural dependency parsing tree, and reflects the syntactic relationships between the words in the sentence through the relationship markers on the dependency arc [48]. Visually, DP identifies the composition of each phrase in the review sentence and analyzes the relationship between the components, so that we can find the emotional phrase from the phrases with the relationship markers of "VOB" (verb-object) and "FOB" (fronting-object).

3.2.3. Negative phrases modifying and frequency counting

With the LTP, we first decompose the cleaned original reviews and realize the hierarchical representation of its connotation. Then we analyze sentence structure and deep semantics and extracts emotional phrases in the statements that contain Kansei knowledge.

In this step, considering the distinction between negative prefixes and negative words is very vague in unstructured languages like Chinese, we need to adjust these situations to keep the intact meaning of the words or phrases extracted.

It has a vaguer structure that Chinese compared with English, which distinguishes segmentation between Chinese and English. "Dislike" and "don't like" are two different word-using in English, but in Chinese, they are the same and can be segmented into "no/not" and "like". Because of that, sentence segmentation may decompose negative meaning into two independent parts of a negative, and a positive, leading the extraction ignores negative opinions. This step makes modifications of negative emotion phrases, which will transfer the negative in the phrase to the adjective itself so that negative opinions can be retained. Fig. 5 shows some examples.

Finally, the analysis results of Kansei knowledge extraction are presented in the form of frequency. The extracted phrases are broken down into two aspects of words: user concern and user attitude. The user concern is the basic attribute of the user's evaluation and is equivalent to the dimensions of that, while the user attitude is the degree of each dimension that describes the user's emotional experience. Its extracted results are presented in the form of "frequency word (user concern or attitude)." Through user concern and user attitude these two aspects and their frequency, we can model the modeling Kansei knowledge.

3.3. Relational extraction of Kansei knowledge

Just as the Kansei knowledge in this article is not only about users' emotion or attitude, but also features of the product that users concern, it is able to set up a complete understanding of the product.

After extraction above, although users' perceptual cognition and evaluation of product characteristics can be obtained from frequency analysis, the connection between them is ignored. Users' concerns and attitudes are simply separated in this way. However, in the actual process of product emotional design, designers need

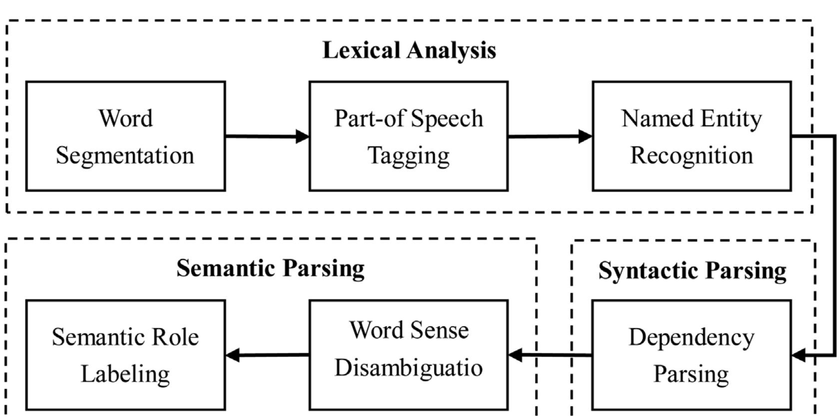

Fig. 4. The architecture of LTP.

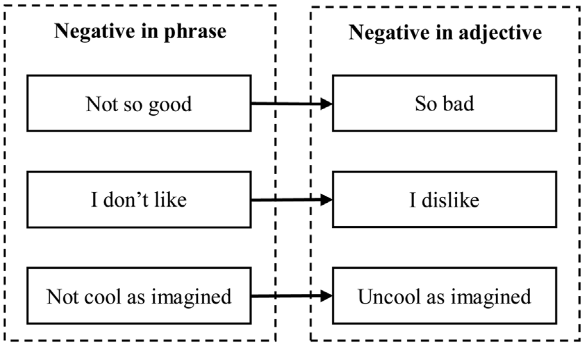

Fig. 5. Shift the negative from phrases to adjectives.

not only some overall product characteristics and some emotional evaluations but also to establish a network corresponding to the relationship between product design features and users' emotional needs.

Therefore, we design the relational extraction method, paying more attention to establishing the connection between design features (user concern) and user emotions (user attitude). The further method extracts the relationship between user concern and user attitude based on the former extraction (Fig. 6). It consists of two main steps; one is seeking the parts that contain the relationship of users' concerns and attitudes (such a part is not necessarily an emotional phrase that contains a specific Kansei knowledge) from online product reviews, the next is establishing the relationship table and the Kansei knowledge tree.

The former extraction method mainly focuses on analyzing the meaning of the commentary text to look for the emotional phrase to extract the Kansei knowledge from the sentence semantics. By contrast, the relational extraction focuses on "connection" rather than "image", whose result is the relationship between users' concerns and their attitudes towards each concern.

By relational extraction of Kansei knowledge, we can complete the relationship modeling of product design elements and user perceptual evaluation, and then make users' emotional needs clear and guide the adjustment of design features, to achieve user-centered product emotional design.

4. Case study

4.1. Objects and metrics

In order to evaluate the effectiveness of the proposed method, an experiment is conducted by using information collected from *jd.com*, a Chinese B2C website (https://www.jd.com/). Product's selection is under these three rules: 1) non-consumables and disposable products, with relatively high quality of user comments; 2) main brands of the product are relatively concentrated, which facilitates the horizontal comparison between products; 3) relatively few changes in product parameters, which are convenient for vertical comparison of products. We choose the mouse for the experiment and select the top ten mice in descending order of sales volume. For each mouse, we crawl 1000 original reviews.

Because the words extracted in this article are not contrasted with known patterns, no standard database can be used to verify whether the result is "correct". Therefore, we only evaluate and compare the performance of methods by the indicators defined by ourselves. This paper defines diversity, effectiveness, and concentration three indicators shown in Table 1 for interpretation.

- Diversity is represented by the cumulative frequency of the top 10 items to evaluate the scope of the range of analysis that extracted Kansei knowledge can present. To transform it into a benefit indicator, we subtract this value from 1. The larger the value of this indicator, the higher the Diversity of the evaluation method.
- Effectiveness, the ratio of the words consisted of more than one Chinese character in all words (a Chinese word containing one single character is generally considered to express a vague and general content), is used to evaluate the possibility that the extracted Kansei knowledge can be effectively utilized. The larger the value of this indicator, the higher the Effectiveness of the evaluation method.
- Concentration is obtained by the number of the first 80% items in the cumulative frequency, which is used to evaluate the reflection capacity of extracted Kansei knowledge. The larger the value of this indicator, the more concentrated the Kansei knowledge extracted by this method, the more it can reflect users' emotional needs overall.

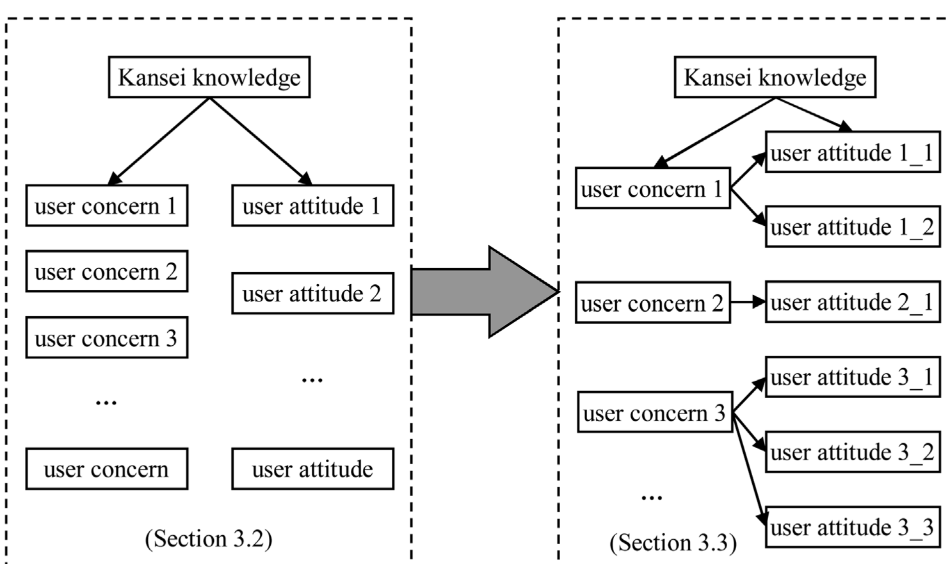

Fig. 6. From single extraction to relational extraction.

Table 1
Evaluation indicators of extraction.

Name	Definition	Value	Characteristics reflected
Diversity (D)	1 minus the cumulative frequency of the top 10 items.	Larger	Higher
Effectiveness (E)	Percentage of words consisting of more than one character.	Larger	Higher
Concentration (C)	The number of entries in the top 80%.	Higher	Higher

4.2. Extraction results

Due to a large amount of data, the intact extracted results are no longer presented here. Table 2 presents partial statistical results, the count and frequency of user concern and attitude are cross-linked with the 10 mice, respectively showing the first three extractions in frequency.

Most of the words presented in the table do not point to some Kansei knowledge. Because a high frequency means that the word has a strong generalizing ability about product features, while a more detailed picture needs to be drawn from the richer overall extraction. The 10 mice can be classified into three categories according to their manufacturer's positioning and ranged by descent selling as follows:

- For the game: Logitech G502, Logitech G102, Razer 2000, Logitech G903
- For fashion: Apple Wireless 2, MI
- For the office: Logitech M330, Logitech M545, Rapoo MT750, Logitech M590

We selected the most popular mouse among the three categories to make their user concern and user attitude wordle respectively. Wordle was proposed by Rich Gordon in 2006, which generates "keyword clouds" or "keyword rendering" to simplify large amounts of text information so that readers can get the main idea of the text at a glance. As shown below, Fig. 7 (a) and (b) are for Logitech G502, Fig. 7 (c) and (d for Apple Wireless 2, and Fig. 7 (e) and (f) for Logitech M330.

Through observing the wordle, we can get such information. For Logitech G502, a mouse for the game, user concern of it mainly focus on game, chicken, hand-feeling, which is consistent with the manufacturer's positioning of it as known. The main emotional experiences of it are heavy, comfortable, big, expensive, etc. For positive emotions, it needs to make clear that this is the prominent advantage of this product, and for negative emotions, we need to improve them. According to Apple wireless 2, a typical mouse for fashion, its user concern is relatively dispersed compared with the other two categories, mainly are brand, quality, habit, system, speed, and price. Its brand effect is obvious, and it is not difficult to find that it has no definite goals in specific application scenarios. At the same time, users' attitudes towards the mouse are focused on good, expensive and convenient, with little relation to the technical features of the product itself. Lastly, considering Logitech M330 for office, its wordles most embody the product features. The user concern focuses on voice, mute, battery, feel and time, while the user attitude focuses on good, small, comfortable and new, whether the manufacturer's position or people's general knowledge of it is well suited to it.

4.3. Comparison and evaluation

In previous research on Kansei knowledge extraction, some studies use noun extraction to extract the product features from reviews [49], and similarly, some studies extract adjectives in reviews as Kansei words [50]. In order to evaluate the effect and results of the proposed method, we directly extract nouns and adjectives from the same corpus as the proposed method using and

Table 2
Statistical results of user's concern and attitude of different products.

Brand	Type	User concerns				User attitudes			
		Chinese	English	Count	Frequency	Chinese	English	Count	Frequency
Logitech	G502	鸡	Chicken	64	4.01%	重	Heavy	71	11.01%
		游戏	Game	63	3.94%	好	Good	56	8.68%
		问题	Problem	51	3.19%	大	Big	41	6.36%
Apple	Wireless 2	苹果	Apple	52	6.27%	好	Good	56	15.86%
		正品	Real	31	3.73%	贵	Expensive	28	7.93%
		电脑	Computer	25	3.01%	不错	Not bad	18	5.10%
Logitech	G102	游戏	Game	82	5.69%	好	Good	71	11.60%
		手	Hand	45	3.13%	小	Small	56	9.15%
		问题	Problem	45	3.13%	大	big	49	8.01%
Logitech	M330	声音	Voice	118	11.37%	好	Good	62	14.94%
		静音	Silence	44	4.24%	小	Small	54	13.01%
		电池	Battery	34	3.28%	大	big	32	7.71%
Razer	Razer 2000	雷蛇	Razer	27	4.09%	好	Good	44	15.77%
		感觉	Feeling	21	3.18%	大	big	28	10.04%
		手	Hand	20	3.03%	不错	Not bad	16	5.73%
MI	MI	小米	MI	90	8.33%	好	Good	63	15.79%
		蓝牙	Bluetooth	63	5.83%	方便	Convenient	23	5.76%
		电脑	Computer	41	3.80%	不错	Not bad	20	5.01%
Logitech	G903	问题	Problem	54	4.47%	好	Good	62	12.55%
		鼠标垫	Mouse mat	39	3.23%	大	Big	34	6.88%
		游戏	Game	32	2.65%	新	New	23	4.66%
Logitech	M545	时间	Time	31	3.53%	好	Good	56	15.22%
		问题	Problem	27	3.08%	小	Small	31	8.42%
		感觉	Feeling	20	2.28%	大	Big	25	6.79%
Rapoo	MT750	问题	Problem	52	5.33%	大	Big	54	14.36%
		蓝牙	Bluetooth	32	3.28%	好	Good	43	11.44%
		游戏	Game	31	3.18%	不错	Not bad	20	5.32%
Logitech	M590	蓝牙	Bluetooth	39	4.61%	好	Good	47	14.55%
		问题	Problem	38	4.49%	小	Small	23	7.12%
		电脑	Computer	35	4.14%	大	Big	18	5.57%

Fig. 7. (a) Logitech G502 user concern Fig. 7 (b) Logitech G502 user attitude, (c) Apple wireless2 user concern Fig. 7 (d) Apple wireless2 user attitude, (e) Logitech M330 user concern Fig. 7 (f) Logitech M330 user attitude.

name it “directly extract” to mark, and accordingly name the proposed method “based on DP (Dependency Parsing)”.

We list a score table giving 1 score to the method with high performance and 0 score to the other method, to obtain the total score of the application of the two methods in these ten products. As shown in Table 3, it can be seen from the average value comparison of each index and the score comparison of the two methods, that compared with the direct words extraction, the method proposed in this paper has an advantage of 19.03% in diversity, 6.91% in effectiveness and 22.18% in centrality. The total score of the method proposed in this paper is 8.9 times higher than that of direct words extraction.

In order to make it more intuitive, Fig. 8 visualizes the data table. It could be found that: There is no obvious difference when comparing the products horizontally, so that it is not applicable to the comparison of different products extraction; From the view of

Table 3
Evaluation and comparison of extraction.

Brand	Type	Method	User concern			User attitude			Score
			D	E	C	D	E	C	
Logitech	G502	Based on DP	24.91%	84.63%	260	51.16%	65.48%	53	5.5
		Directly extract	33.09%	74.63%	164	56.53%	65.35%	53	0.5
Apple	Wireless 2	Based on DP	26.14%	86.09%	172	48.44%	70.80%	45	6
		Directly extract	32.95%	76.57%	132	54.50%	68.64%	43	0
Logitech	G102	Based on DP	25.14%	84.23%	213	48.86%	65.85%	53	5
		Directly extract	33.48%	76.41%	141	56.59%	66.23%	47	1
Logitech	M330	Based on DP	31.79%	85.79%	173	57.83%	65.38%	31	6
		Directly extract	38.01%	76.81%	104	67.96%	63.76%	28	0
Razer	Razer 2000	Based on DP	25.45%	84.14%	177	53.76%	65.22%	37	6
		Directly extract	37.01%	73.11%	140	62.97%	62.02%	34	0
MI	MI	Based on DP	31.30%	86.55%	141	46.62%	68.99%	53	5
		Directly extract	35.67%	75.66%	112	58.99%	69.92%	42	1
Logitech	G903	Based on DP	23.41%	85.01%	206	44.33%	67.36%	56	6
		Directly extract	31.32%	77.22%	173	50.30%	64.89%	54	0
Logitech	M545	Based on DP	21.41%	86.04%	176	53.26%	63.92%	36	5
		Directly extract	36.52%	76.28%	108	65.03%	66.37%	35	1
Rapoo	MT750	Based on DP	27.05%	86.69%	158	51.06%	69.92%	48	6
		Directly extract	32.46%	76.43%	142	60.25%	66.95%	41	0
Logitech	M590	Based on DP	27.78%	84.26%	155	47.68%	66.35%	40	4
		Directly extract	33.78%	75.13%	128	57.56%	66.81%	41	2
Average value /Total score		Based on DP	26.44%	85.34%	183.1	50.30%	66.93%	45.2	54.5
		Directly extract	34.43%	75.82%	134.4	59.07%	66.09%	41.8	5.5

Diversity (red column for user concern and blue column for user attitude), data extracted by the method based on DP can provide more scopes and perspectives for analysis, and user concerns contain more information than user attitudes; From the view of Effectiveness (yellow column for user concern and green column for user attitude), data extracted by the method based on DP has more space for effective use in Kansei knowledge analysis, in which user concerns have more obvious advantages than user attitudes; From the view of Concentration (dark green line for user concern and pink line for user attitude), the Kansei knowledge formed of data extracted by the method based on DP can more fully reflect the emotional needs of users, with a remarkable advantage on the user concerns.

4.4. Relational extraction for Kansei knowledge tree

The user concerns and user attitudes of the product have been obtained and evaluated, the relationship between them will be established in this section to generate the product's Kansei knowledge tree.

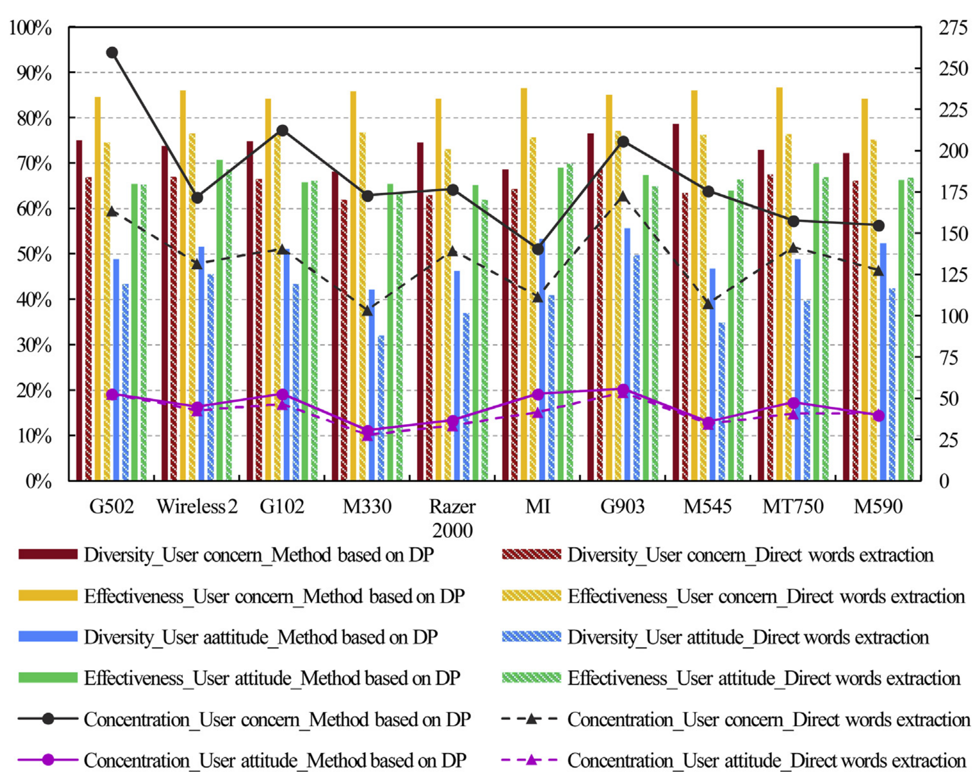

Fig. 8. Comparison and visualization of extraction.

Table 4
Classification of user concern in the relational extracted Kansei knowledge of Logitech G502.

Product	Aspect	Attribute	User concern
Product (859)	Overall (295)	Overall (16)	whole (7), collectivity (3)
		Product (213)	product (6), thing (23), goods (11), mouse (179)
		Feeling (66)	feeling (35), question (23), character (3), praise (5)
	Features (251)	Brand (19)	Logitech (8), Razer (11)
		Usage (27)	habit (9), time (18)
		Performance (24)	performance (9), price (15)
		Appearance (19)	appearance (9), configuration (4), shape (2), beauty (4)
		Function (88)	function (18), capability (2), mode (9), game (33), chicken (23), website (3)
		Service (40)	logistics (16), courier (2), express (3), support (7), buddy (3), attitude (9)
		Manufacturing (34)	packing (16), box (5), workmanship (5), quality (8)
	Details (313)	Roller (27)	wheel (10), roller (11), pulley (6)
		Key (47)	key (35), side key (6), left key (6)
		Hand feel (124)	hand (40), hand feel (82), grip (2)
		Light (13)	lamp (3), light (6), breathing lamp (4)
		Speed (21)	speed (17), top speed (2), DPI (2)
		Weight (49)	weight (13), volume (4), a counterweight (11), bulk (17), the center of gravity (4)
		Sound (9)	sound (9)
		Shell (9)	metal (2), lid (2), scratch (5)
		Accessories (9)	software (6), introduction (3)
		Other (5)	tail (3), mouse cable (2)

Taking the best-selling mouse Logitech G502 as the example, there are 143 user concerns extracted from the 1000 reviews of the product. After eliminating those with unclear or repeated meanings, there are 10 (divided into three categories) in the overall level of products, 26 (divided into 7) in the level of features, 30 (divided into 10) in the level of details. Table 4 shows the classification, and each item is attached the number of the user attitudes about it in parentheses.

The extracted result generates a tree structuring the product Kansei knowledge. Therefore, this article defines such a tree structure, the trunk for an analysis product object, the branches for a furcated structure composed of user concerns and their classification, the leaves for the user attitudes of its branch (user concern), and the size of a leaf depends on the frequency of the user attitude. Call this tree structure the Kansei knowledge tree. Fig. 9 draws a schematic diagram of the Kansei knowledge tree of Logitech G502 by 4 local areas of product, aspect, attribute, and user concern.

5. Discussion

In this paper, we proposed a computerized method to automatically capture Kansei words and a relational extracting method to construct a Kansei knowledge tree for product emotional design from online reviews using natural language processing technology. For comparison, we produced Table 5 to illustrate recent articles on extracting Kansei knowledge and their research aims, technology, tools and study case.

The main advantage of this study compared with these previous studies is that, first, we draw a Kansei knowledge tree to model the connection between product features and user perceptions. We consider not only the quantity improvement (data scale and data updating) [8] of the user survey, but also the quality improvement and establishing the connection between product attributes and user emotions. Second, our research is based on sentiment analysis and utilize a suitable toolkit to explore users' emotional needs, which makes better use of reviews data than other text processing techniques in emotions extraction [33]. And the third, the most difference, is that this study is based on the semantics of the sentence itself, rather than by comparing words to the dictionary [15] and the historical database [32], which helps to discover the Kansei knowledge that was not discovered in the past.

From a theoretical sense, considering the rapid development of information technology, the method proposed in this paper extracts the Kansei knowledge from online product reviews to break through the traditional study method of an offline user survey, which utilizes the natural language processing technology and is an auxiliary method for Kansei design research.

From the practical sense, many famous enterprises, such as Ford, Philips, Mazda, and Nike, attach great importance to Kansei design and apply Kansei design strategies to enhance the soft functions of products. The method proposed in this paper uses computer technology to extract Kansei knowledge, which can be conveniently and quickly applied to relevant applications of product design, thus improving the stability and accuracy of user survey and accelerating the efficiency of research and development.

There are still some limitations in the research. Firstly, when establishing the Kansei knowledge tree, this article is through manual classification and manual tree drawing. Indeed, for designers, it is not enough to assist designing only by data sheets. Therefore, the method in this paper still needs to be combined with the word classification technology of artificial intelligence and drawing the tree of products directly through the computer, to realize a more convenient extraction. Secondly, this paper has not applied the actual questionnaire survey, and physiological experiment instead extracts the Kansei knowledge of product only through online reviews. It fails to test each other among the three methods, while the weakness of extraction from existed text is that it cannot obtain those users' cognition of the product which needed to generate in a long time, which is the emotional demand hidden in users' subconscious.

6. Conclusions

By integrating Linguistics, Kansei engineering, and computers, with the mouse as an experimental object, this paper puts forward a method for Kansei knowledge extraction from online product reviews, and then further proposes a relational extraction method that generates the relationship of product design features and user perceptual evaluation. The method starting from deep semantic and syntactic structure extracts pure Kansei knowledge through NLP, finds emotional phrases and extracts knowledge based on the Dependency Parsing, whose extracted results are presented in the form of the frequency distribution. The further relational extraction method is to correlate the users' concerns with attitudes in Kansei

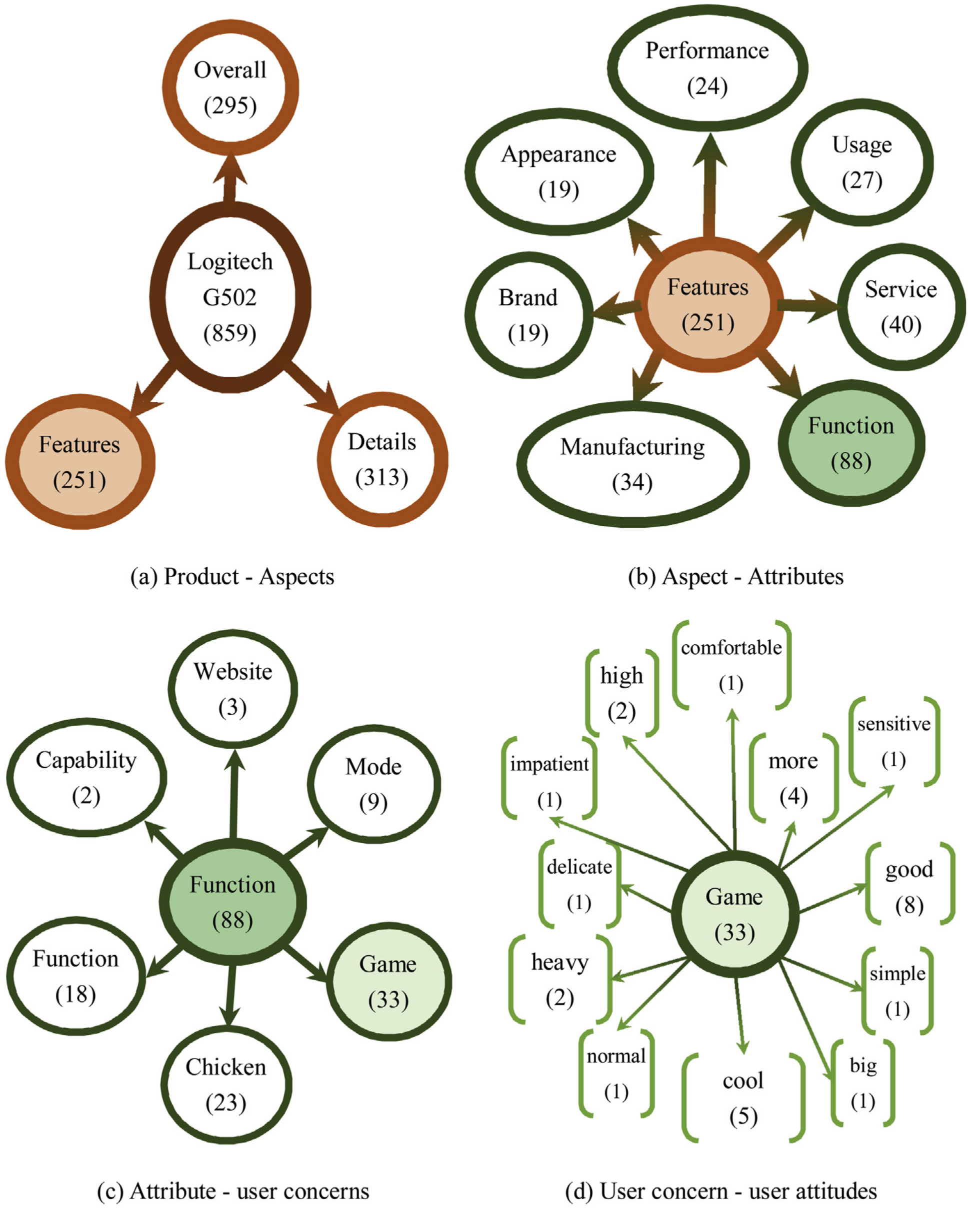

Fig. 9. Kansei knowledge tree of Logitech G502.

knowledge, the extracted result of which is presented as a Kansei knowledge tree, focusing on the establishment of the relationship between product design features and user perceptual evaluations, which modeled the relationship between product design elements and user emotional needs. Therefore, the efficiency of the user survey can be accelerated, users' emotional needs can be clarified, and product design features can be adjusted to realize the goal of assisting the user-centered product emotional design.

Table 5
Comparison of articles on extracting Kansei knowledge.

Article	Aim to	Based on	Make use of	Recognize by	Case study as
Li et al. [8]	large scale, data updating	machine learning	online questionnaires	related literature of affective design	smartwatch
Yamada et al. [15]	automatically Kansei evaluation	text mining	Japanese dictionaries, part-of-speech data	a Bayes extension of Latent Dirichlet	wristwatch
Wang, Li, Liu, et al. [32]	unstructured text, the gap between affective response and intentions	text mining, fuzzy set theory	attributes and words used in the literature	synonyms and antonyms found by WordNet	online dataset
Wang, Li, Tian, et al. [33]	large scale, real-time, summarize information	natural language processing	probabilistic part of speech tagging	attributes and words used in the literature	Hasbro Furby
This study	large scale, summarize information, establish a relationship	natural language processing, sentiment analysis	Dependency Parsing in language technology platform (LTP)	the sentence themselves	mice after analysis

Acknowledgments

The authors would like to thank the Center for Information Retrieval at Harbin Institute of Technology for providing the LTP software. This work is supported by the National Natural Science Foundation of China (Grant NO. 71471033 & 71771045) and the "Double First-Class" Disciplines Construction Project of Northeastern University (Grant No. 02050021940101). We thank all the participants for carrying out the study. Furthermore, we thank the editor and anonymous reviewers for their valuable comments and advice.

References

[1] H.M. Khalid, Embracing diversity in user needs for affective design, Appl. Ergon. 37 (4) (2006) 409–418.
[2] F. Guo, W.L. Liu, F.T. Liu, H. Wang, T.B. Wang, Emotional design method of product presented in multi-dimensional variables based on Kansei Engineering, J. Eng. Des. 25 (4-6) (2014) 194–212.
[3] J. Vieira, J.M.A. Osório, S. Mouta, P. Delgado, A. Portinha, J.F. Meireles, J.A. Santos, Kansei engineering as a tool for the design of in-vehicle rubber keypads, Appl. Ergon. 61 (2017) 1–11.
[4] V. Sharma, P. Karla, A. Kumar, Customer perception assessment technique kansei engineering: a review, Int. J. Sci. Res. 2 (7) (2013) 237–240.
[5] Y. Matsubara, M. Nagamachi, Hybrid Kansei engineering system and design support, Int. J. Ind. Ergon. 19 (2) (1997) 81–92.
[6] Y.-C. Lin, H.-H. Lai, C.-H. Yeh, Consumer-oriented product form design based on fuzzy logic: a case study of mobile phones, Int. J. Ind. Ergon. 37 (6) (2007) 531–543.
[7] C. Llinares, A.F. Page, Kano's model in Kansei Engineering to evaluate subjective real estate consumer preferences, Int. J. Ind. Ergon. 41 (3) (2011) 233–246.
[8] Z. Li, Z. Tian, J. Wang, W. Wang, G. Huang, Dynamic mapping of design elements and affective responses: a machine learning based method for affective design, J. Eng. Des. (2018) 1–23.
[9] S. Chin, K.-Y. Kim, Facial configuration and BMI based personalized face and upper body modeling for customer-oriented wearable product design, Comput. Ind. 61 (6) (2010) 559–575.
[10] M. Nagamachi, Kansei engineering as a powerful consumer-oriented technology for product development, Appl. Ergon. 33 (3) (2002) 289–294.
[11] M. Nagamachi, Kansei engineering: a new ergonomic consumer-oriented technology for product development, Int. J. Ind. Ergon. 15 (1) (1995) 3–11.
[12] H.M. Khalid, M.G. Helander, A framework for affective customer needs in product design, Theor. Issues Ergon. Sci. 5 (1) (2004) 27–42.
[13] K. Fung, C. Kwong, K.W. Siu, K. Yu, A multi-objective genetic algorithm approach to rule mining for affective product design, Expert Syst. Appl. 39 (8) (2012) 7411–7419.
[14] J. Barlow, D. Maul, Emotional Value: Creating Strong Bonds With Your Customers, Berrett-Koehler Publishers, 2000.
[15] A. Yamada, S. Hashimoto, N. Nagata, A text mining approach for automatic modeling of Kansei evaluation from review texts, Paper Presented at the International Conference on Kansei Engineering & Emotion Research, (2018) .
[16] K.-H. Chai, X. Xiao, Understanding design research: a bibliometric analysis of Design Studies (1996–2010), Des. Stud. 33 (1) (2012) 24–43.
[17] T. Nasukawa, J. Yi, Sentiment analysis: capturing favorability using natural language processing, Paper Presented at the Proceedings of the 2nd International Conference on Knowledge Capture, (2003) .
[18] K. Dave, S. Lawrence, D.M. Pennock, Mining the peanut gallery: opinion extraction and semantic classification of product reviews, Paper Presented at the Proceedings of the 12th International Conference on World Wide Web, (2003) .
[19] J. Zhao, K. Liu, L. Xu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, MIT Press, 2016.
[20] A. Ortigosa, J.M. Martín, R.M. Carro, Sentiment analysis in Facebook and its application to e-learning, Comput. Hum. Behav. 31 (2014) 527–541.
[21] B. Desmet, V. Hoste, Emotion detection in suicide notes, Expert Syst. Appl. 40 (16) (2013) 6351–6358.
[22] N. Oliveira, P. Cortez, N. Areal, Stock market sentiment lexicon acquisition using microblogging data and statistical measures, Decis. Support Syst. 85 (2016) 62–73.
[23] H. Yuan, W. Xu, Q. Li, R. Lau, Topic sentiment mining for sales performance prediction in e-commerce, Ann. Oper. Res. (2017) 1–24.
[24] J.G. Breslin, D. O'Sullivan, A. Passant, L. Vasiliu, Semantic Web computing in industry, Comput. Ind. 61 (8) (2010) 729–741.
[25] N. Cummins, S. Amiriparian, S. Ottl, M. Gerczuk, M. Schmitt, B. Schuller, Multimodal bag-of-words for cross domains sentiment analysis, Paper Presented at the 2018 IEEE International Conference on Acoustics, Speech, and Signal Processing (2018).
[26] P.D. Turney, M.L. Littman, Measuring praise and criticism: inference of semantic orientation from association, ACM Trans. Inform. Syst. (TOIS) 21 (4) (2003) 315–346.
[27] E. Wanner, On Remembering, Forgetting, and Understanding Sentences: a Study of the Deep Structure Hypothesis, Mouton De Gruyter, 1974.
[28] S.S. Htay, K.T. Lynn, Extracting product features and opinion words using pattern knowledge in customer reviews, Sci. World J. 2013 (2013).
[29] D.T. Santosh, K.S. Babu, S. Prasad, A. Vivekananda, Opinion mining of online product reviews from traditional LDA topic clusters using feature ontology tree and sentiwordnet, Int. J. Educ. Manage. Eng. (IJEME) 6 (6) (2016) 34–44.
[30] E. Cambria, B. White, Jumping NLP curves: a review of natural language processing research, IEEE Comput. Intell. Mag. 9 (2) (2014) 48–57, doi:http://dx.doi.org/10.1109/MCI.2014.2307227.
[31] E. Cambria, Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2) (2016) 102–107, doi:http://dx.doi.org/10.1109/MIS.2016.31.
[32] W. Wang, Z. Li, L. Liu, Z. Tian, E. Tsui, Mining of affective responses and affective intentions of products from unstructured text, J. Eng. Des. (2018) 1–26.
[33] W. Wang, Z. Li, Z. Tian, J. Wang, M. Cheng, Extracting and summarizing affective features and responses from online product descriptions and reviews: a Kansei text mining approach, Eng. Appl. Artif. Intell. 73 (2018) 149–162.
[34] I. Chaturvedi, E. Cambria, R.E. Welsch, F. Herrera, Distinguishing between facts and opinions for sentiment analysis: survey and challenges, Inf. Fusion 44 (2018) 65–77, doi:http://dx.doi.org/10.1016/j.inffus.2017.12.006.
[35] H. Cunningham, GATE, a general architecture for text engineering, Comput. Hum. 36 (2) (2002) 223–254.
[36] E. Loper, S. Bird, NLTK: the natural language toolkit, Paper Presented at the Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics-Volume 1, (2002) .
[37] Y. Jin, W. Xiong, C. Wang, Feature selection for chinese text categorization based on improved particle swarm optimization, Paper Presented at the Natural Language Processing and Knowledge Engineering (NLP-KE), 2010 International Conference on, (2010) .
[38] X. Yang, J. Su, J. Lang, C.L. Tan, T. Liu, S. Li, An Entity-Mention Model for Coreference Resolution With Inductive Logic Programming. Proceedings of Acl-08: Hlt, (2008) , pp. 843–851.
[39] W. Che, Z. Li, T. Liu, Ltp: a chinese language technology platform, Paper Presented at the Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations, (2010) .
[40] S. Sun, C. Luo, J. Chen, A review of natural language processing techniques for opinion mining systems, Inf. Fusion 36 (2017) 10–25, doi:http://dx.doi.org/10.1016/j.inffus.2016.10.004.
[41] Z. Hai, K. Chang, J.J. Kim, C.C. Yang, Identifying features in opinion mining via intrinsic and extrinsic domain relevance, IEEE Trans. Knowl. Data Eng. 26 (3) (2014) 623–634, doi:http://dx.doi.org/10.1109/TKDE.2013.26.
[42] Z. Hai, K. Chang, G. Cong, One seed to find them all: mining opinion features via association, Paper Presented at the 21st ACM International Conference on Information and Knowledge Management, CIKM 2012 (2012).
[43] J. Liao, S. Wang, D. Li, X. Li, FREERL: fusion relation embedded representation learning framework for aspect extraction, Knowledge Based Syst. 135 (2017) 9–17, doi:http://dx.doi.org/10.1016/j.knosys.2017.07.015.
[44] J. Cao, K. Zeng, H. Wang, J. Cheng, F. Qiao, D. Wen, Y. Gao, Web-based traffic sentiment analysis: methods and applications, Ieee Trans. Intell. Transp. Syst. 15 (2) (2014) 844–853, doi:http://dx.doi.org/10.1109/TITS.2013.2291241.
[45] W. Che, Y. Zhao, H. Guo, Z. Su, T. Liu, Sentence compression for aspect-based sentiment analysis, IEEEACM Trans. Audio Speech Lang. Process. 23 (12) (2015) 2111–2124, doi:http://dx.doi.org/10.1109/TASLP.2015.2443982.
[46] L. Zhao, J. Jia, L. Feng, Teenagers' stress detection based on time-sensitive micro-blog comment/response actions, 4th IFIP TC 12 International Conference on Rtificial Intelligence in Theory and Practice, IFIP AITP 2015 Held as Part of International Federation for Information Processing, WCC 2015 vol. 465 (2015) 26–36.
[47] Y. Zhou, T. Lu, T. Zhu, Z. Chen, Environmental incidents detection from chinese microblog based on sentiment analysis, 2nd International Conference on Human Centered Computing, HCC 2016 vol. 9567 (2016) 849–854.
[48] W. Che, Z. Li, Y. Li, Y. Guo, B. Qin, T. Liu, Multilingual dependency-based syntactic and semantic parsing, Paper Presented at the Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, (2009) .
[49] J. Jin, P. Ji, R. Gu, Identifying comparative customer requirements from product online reviews for competitor analysis, Eng. Appl. Artif. Intell. 49 (2016) 61–73.
[50] M.-D. Shieh, Y.-E. Yeh, C.-L. Huang, Eliciting design knowledge from affective responses using rough sets and Kansei engineering system, J. Ambient Intell. Humaniz. Comput. 7 (1) (2016) 107–120.