A product affective properties identification approach based on web mining and domain ontology in a crowdsourcing environment

Danni Chang[a], Carman Lee[b]
[a] *School of Media & Design, Shanghai Jiao Tong University, Shanghai, China*
[b] *Department of Industrial and Systems Engineering, Hong Kong Polytechnic University, Hong Kong, China*

Abstract. Precisely understanding the values and perceptions of consumers has long been recognized as an essential factor for product success. Customers' emotional experience, as the basis for the formation of human values and judgment, has been considered carefully in product design to ensure customer satisfaction and increased product competitiveness. Therefore, affective product design has attracted more and more research attention in order to develop products that satisfy customer feelings as an aspect of product quality. However, conventional product design places more attention on functional attributes and relies on surveys and user experiments to collect the evaluations of users/customers. That, in effect, leads to issues in which i) consumers can only express their feelings and opinions on the design attributes specified by assigners; ii) abundant online consumer resources are neglected, especially when online channels (e.g. a crowdsourcing platform) have become popular choices for consumers to post their opinions; and iii) a lack of further prioritization and re-construction of affective design properties that provide useful references for design practice. The contribution of this study is in developing a product affective properties identification approach in a crowdsourcing environment. Particularly, web mining and text mining are deployed to capture online product review resources and perform preliminary computation. A product design knowledge hierarchy has been established to help find design properties, while sentiment analysis was undertaken to identify consumers' sensations. With the help of domain ontology providing semantic relations to connect design properties and corresponding human affection, product affective properties can be identified. Furthermore, the identified product affective properties can be prioritized based on a joint consideration of design and affective performance, so as to provide designers with practical references for future design improvement and support decision making. To illustrate the proposed approach, a pilot study on iPhone 7 was conducted, in which the influential affective properties have been identified and ranked.

Keywords. Product affective property, data mining, product design knowledge hierarchy, domain ontology, prioritization.

1. Introduction

Nowadays, the increasingly competitive market has elicited an urgent need to develop successful products that can satisfy the increasing consumer expectations and demands [46]. Apart from basic functions and economic considerations, affective aspects of products are also of great concern to consumers [41]. Design attributes, such as color and form can provoke feelings, and influence the overall perception of a product. Therefore, affective product design (APD) was advocated to develop products that satisfy customer feelings as an aspect of product quality [4, 40]. In this regard, in the relevant literature, it is indicated that products with deliberate affective design can help improve consumer satisfaction and further promote product success [11]. Therefore, good affective features can sharpen the competitive edge of products, and a precise understanding of product affective properties appears particularly important and deserves in-depth investigation.

Regarding previous affective design studies, most research focus was placed on the analysis of emotions [15, 16, 20] and the establishment of quantitative models linking emotions and design attributes [4, 5, 6]. In terms of data acquirement methods, surveys and user experiments are widely adopted. As in [20, 23], questionnaires are used as the main method to collect consumer responses. However, the design of a survey, questionnaire or user experiment, to some extent, imposes the constraints of user involvement and user freedom in presenting their sensations on any design properties.

Moreover, with the rapid development of the Internet, Web 2.0 provides an open platform to connect enterprises with worldwide consumers. Consumers have more convenient channels in which to contribute their opinions. In particular, crowdsourcing, which is an important method for drawing on large numbers of people to contribute their opinions [28, 45, 46], has become an important way to effect consumer responses. Taking Proctor & Gamble as an example, the use of a crowdsourcing platform "InnoCentive" to collect product problems and possible solutions from Internet users has helped them increase the problem solving rate to 30%. For more examples, Wikipedia, Amazon's Mechanical Turk and iStockPhoto.com take advantage of the tremendous numbers of Web users that are willing to contribute their knowledge and ideas. Crowdsourcing appears to be a promising way to solicit consumer responses and is studied as the main data source in this work. Considering that large numbers of consumers' comments are collected via crowdsourcing and are often presented by textual data, data mining techniques, which are efficient in dealing with big data and effective for textual analysis are considered.

Therefore, to fully consider users' opinions and discover possible important affective design properties from abundant online consumer resources, this paper aims to investigate affective design properties based on users/consumer responses acquired by crowdsourcing platforms. Furthermore, the identified affective properties are re-organized into a systematic structure according to their design and affective importance in order to provide designers with strategic reference for further improvement.

To make the explanation clearer, the concepts of product affective properties and affective performance are defined.

Definition 1 Product Affective Properties. Design aspects or product features which can provoke users'/customers' emotions. If there exists a cause-and-effect relation (in a qualitative or quantitative manner) between design properties and consumers' affectations, the design properties are identified as product affective properties.

Definition 2 Affective Performance of Product Properties. It relates to the success of product properties in influencing consumers' emotions and can be measured against standards such as polarity and intensity.

2. Related Works

In this section, existing related works are analyzed from three aspects. Firstly, the concept of crowdsourcing is examined to reveal the current use of interactive Internet platforms by consumers in order to contribute their ideas and opinions on the product design process. As an efficient way to collect large amounts of consumer responses, the focus is on the crowdsourcing environment will be focused as the source of consumer responses in this work. Then, the research status of product affective design is thoroughly reviewed, and potential promising research directions can be identified accordingly. Finally, data mining techniques, which are powerful in dealing with a large amount of data, are considered in APD. Based on the analysis of existing studies, the challenges and possible contributions of this work are summarized and highlighted.

2.1. A Brief Overview of Crowdsourcing

The term "crowdsourcing" was first coined by Howe in the article 'The Rise of Crowdsourcing', in 2006 [28]. Basically, crowdsourcing can be schematically depicted as in Fig. 1. The employer/assigner (right side) defines the task requirements and submits the tasks to a mediator, viz. the online crowdsourcing platform. Online workers/providers (left side) who are interested in this task can work on it and submit their solutions to the mediator after completion. These solutions are then forwarded to the employer who rewards the participants if their solutions are approved.

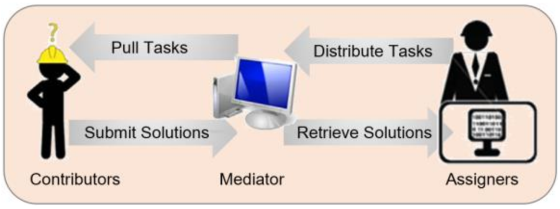


Figure 1 A typical crowdsourcing scheme

Conventionally, crowdsourcing practices rely on the power of the massive work force, such as the Great Wall and Oxford English Dictionary. Nowadays, the Internet is developing rapidly, and modern crowdsourcing has transferred mainly to the Internet. Especially in the product design realm, the dependence on consumers/users is increasing, from the traditional internal development to the current user-centered design. It simultaneously requires more convenient channels for communication and interaction with customers/users. Compared with general websites, crowdsourcing platforms not merely enable Internet users to post their comments freely, but are also more targeted, having more Internet users with particular interest or knowledge of the product. Thus crowdsourcing may have higher possibilities to receive useful customer information. For this reason, crowdsourcing has been practiced in various design-related websites, such as Dell's *Ideastorm* for acquiring comments and suggestions for all Dell products from Internet users, product review websites gathering users' comments and *DesignBoom* collecting innovative design concepts. Therefore, crowdsourcing has assisted in developing a highly interactive environment for user/customer interaction and is an important source of consumer/user responses.

2.2. ***Affective Product Design***

There have been many studies directed at APD, from different perspectives. Considering the basic assumption for APD that there exists a cause-and-effect relationship between consumers' affective responses and product features, existing studies can be generally classified into the following categories:

- *Identification and classification of consumers' affectation*; this is to study consumers'/users' emotions provoked by products. Regarding particular techniques, the semantic deferential (SD) method is frequently used to investigate customers' perception of products [1]. By studying product semantics, customers' subjective feelings about a product can be discerned and quantized on a Likert-type scale. Other assessment of emotions can be seen in the use of Conjoint analysis and Quality Function Deployment (QFD). Considering the ambiguity and subjectivity of consumers' affectations, the selection and classification methods for Kansei words (or emotional adjectives) are a research hotspot. K-means clustering, affinity diagrams and design structure matrixes have been used to achieve this purpose [2, 3, 20].

- *Understanding of emotion-related design features*; basically, this research area can be classified into two kinds: one is to identify the features which can evoke users' emotional responses, and the other is to classify and measure such emotion-related features. For the first type, empirical study is a useful way and is widely adopted by researchers. For example, an empirical survey was conducted to investigate the influence of website emotional design features, visual appeal and ease of use in regard to users' perceptions of usefulness, trust, as well as the intention to use websites [15]. In addition, an empirical experiment was performed to examine the emotions of children influenced by game design [16]. For the second type, quantitative Kansei models were developed to classify emotional features [20]. Moreover, the indexing of consumers' subjective and emotion-driven opinions is advocated to unify the "perceived value" of designers and consumers from the affective perspective [18].

- *Modelling the relationship between affective responses and design attributes*; for qualitative relations, hypothesis testing is often used to identify the effects of design factors on affective responses [19]. However, most studies put research focus on the quantitative relations between the design attributes/factors and the affective responses. For example, a methodological framework combining user tests and statistical analysis was established to build links between user's emotional responses and the geometrical features of the products [12]. A fuzzy regression model was proposed to model the relationship between customer satisfaction and product design parameters [21]. Moreover, Kansei engineering is a notable way of translating consumers' psychological feelings about a product into perceptual design attributes [4]. A number of Kansei models have been accordingly developed to improve the association accuracy or applicability of Kansei approaches. To illustrate, fuzzy logic and rough set theory have been applied to cope with the uncertainty and ambiguity of affective responses [5, 6]. In addition, statistical methods, especially regression algorithms and artificial intelligence techniques, such as neural networking, rule mining and genetic algorithms, are also widely studied to model the relationship between affective responses and specific design attributes [7, 8]. Recently emerging studies, have placed more attention in tackling the nonlinearity problem of the relationships [47].

- *Design assessment and decision making*; this relates to the estimation of customer satisfaction based on affective responses, so as to support the specification, classification and prediction of the design attributes [9]. Factor analysis and the hypothesis-testing approach are often used to examine the affective performance of the design attributes [10]. Kansei engineering is also utilized to help with the classification of products for facilitating decision making in practical industrial design cases [23]. In this respect, pre-purchase affectations are more emphasized in order to identify those affectations which can influence the purchase decision [11, 27].

- *Integration of affective design and other design concerns*; for instance, an artificial intelligence (AI)-based methodology for integrating affective design, engineering, and marketing for defining design specifications of new products has been proposed by which the concerns of the three processes can be considered simultaneously in the early design stage [13]. In addition, [27] hypothesized that product attributes influence users' emotions and that the relationship is moderated by the adherence of these product attributes to purchase criteria, and further hypothesized that the emotional experience of the user influences purchase intention.

Although APD has been studied from different perspectives, the methods to acquire consumer data mainly rely on user tests including questionnaires, surveys and interviews. Online platforms have become a popular way for consumers to present their comments, and thus contain considerable consumer resources, have been neglected. On the other hand, the design attributes used for such user tests are specified by assigners, rather than consumers. Specifically, product appearance such as color, and size is much more focused in APD [42, 43].In [17], color preference was studied in emotional design.[22] aimed to demonstrate that product attributes related to form are relevant in eliciting intense emotion and usability perception in mobile phones. [25] intended to identify the roles of visual merchandising inside a car showroom as stimuli that attract customers. In addition, [26] presented a neural network based approach for modeling consumers' affective responses in product form design. Due to the major attention on product form design, other product features or design properties might have been neglected.

Furthermore, customers' affective evaluations elicited by product form directly lead to the issue that pre-purchase emotions is more focused. However, many essential product quality aspects, such as functionality, usability and safety, cannot be perceived unless customers use the product. Therefore, post-purchase or post-use affection is also important and deserves more research attention.

In summary, existing APD studies, in effect, impose constraints on consumers in expressing their feelings and opinions on any design attribute at any user phase using any preferable words.

2.3. ***Data mining in affective product design***

Intelligent techniques have been widely applied in APD. On the one hand, advanced quantitative models are developed to deal with the uncertainty and ambiguity in consumers' responses. For instance, rough set (RS) and particle swarm optimization (PSO) based-ANFIS approaches were proposed to model customer satisfaction for affective design and to further improve the modeling accuracy [24]. A genetic algorithm (GA)-based rule-mining method was proposed to discover a set of rules relating design attributes with customer evaluation based on survey data in order to facilitate APD [8]. On the other hand, intelligent computation is helpful for processing consumer data efficiently. Especially with the development of the Internet, the amount of consumer data involved in product design process is increasing massively.

Data mining, which is powerful in dealing with mass data, has attracted much attention. Actually, data mining is a generic term that covers such techniques as clustering [29, 30], association rule generation [31] and neural network' [32, 33, 34, 35]. Clustering is mainly used for classification based on distances (similarities) between different concepts or designs. In APD, clustering is often applied to i) classify Kansei tags; and ii) classify product candidates [20, 23]. Association rule generation is employed to find regularities between products in large-scale transaction data. It is possible to cope with both numerical data and textual information. For example, rule-mining is utilized to discover a set of rules relating to design attributes based on customer evaluation so as to improve APD [8]. Neural networking consists of an interconnected group of artificial neurons, and processes information through a connectionist approach to computation. In this regard, a neuro network model was developed enabling single users' responses to different products to be predicted, and showed that the mathematical models based on neural networks achieved highly accurate predictions [26].

Besides, data mining is capable of coping with qualitative data, since online consumer responses are mostly in qualitative format (e.g. textual comments) [36]. For example, web mining provides an effective way to discover textual patterns and to extract Web content; and text mining shows advantages in analyzing textual information and deriving high-quality information from the text. Regarding specific applications, a framework was developed by He [37] to improve user experience with case-based reasoning systems using text mining and Web 2.0. In addition, a text mining system, built on ontology to deal with the diagnosis data in the automotive domain, was proposed by Rajpathak [38]. Moreover, text mining was applied in opinion polarity classification that helped decrease the sensitivity of ambiguous terms [39].

2.4. ***Summary***

Based on a thorough analysis of existing related work, it was found that APD has become an important research area in the product design realm. However, challenges still exist, hindering the further exploitation of online consumer resources from the affective perspective. In detail,

1) Previous approaches commonly adopted surveys and questionnaires to collect consumers' affective responses, and neglected the massive online resources; 2) the product attributes/design properties are specified by assigners, which restricts users' choices to comment on other attributes; 3) affective design is often considered for product form based on pre-purchase emotion; 4) existing research is heavily inclined to affective aspects, and neglects the integration of design concerns, viz. the incorporation of design knowledge to further prioritize and re-construct affective design properties, so as to benefit designers with more practical advice.

To tackle the problems above, this work aims to contribute a product affective properties identification approach which is able to 1) make use of crowdsourced consumer responses (where free comments including pre- and post-use experience can be obtained) to identify important affective design properties; 2) deal efficiently with a large amount of data in qualitative format; and 3) further prioritize and represent affective design properties in terms of their comprehensive performance in both affective and design aspects.

3. Research Methodology

The overall framework of research methodology is outlined in Fig. 2. Generally, it can be divided into three stages. Stage 1 is to capture product review data from a crowdsourcing product design environment (i.e. online resources) and perform basic text processing to extract useful textual tokens. In Stage 2, the textual tokens are examined to identify product affective properties. For this purpose, a product design knowledge hierarchy (PDKH) is constructed to provide design considerations in order to identify design properties, and ontology is utilized to assist in semantic analysis and sentiment analysis, so as to investigate the emotions associated with the product properties. In Stage 3, a prioritization process is deployed to estimate the importance of different affective design properties. As a result, affective design properties can be ranked according to their comprehensive importance in terms of design and consumers' affectations, and a new representation of product properties can be mapped out which can be useful reference for designers in regard to design evaluation and decision making. Furthermore, the properties with strong affectations will be retained for potential use in future design.

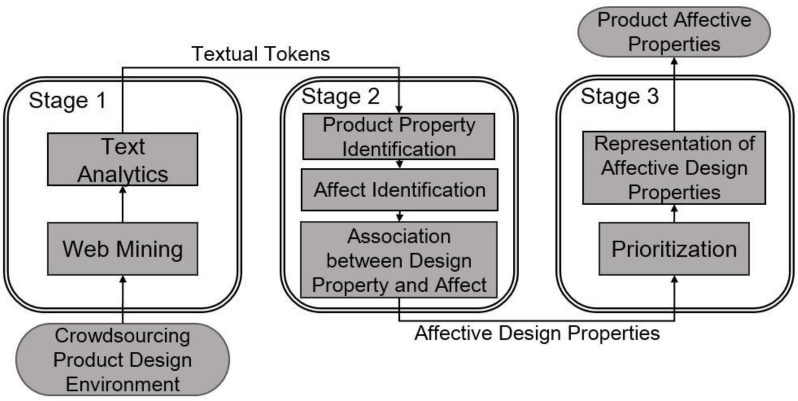


Figure 2. Overall Framework of Research Methodology.

3.1. *Stage 1: Extraction of Online Consumer Responses*

As mentioned above, with the rapid development of the Internet, there are more and more online platforms and convenient access for consumers to offer their understanding and impressions of products. Customers/users can give any comments on any design aspects which can arose their interest. Especially for crowdsourcing, it creates channels to collect opinions and solutions from the large number of Internet users and is studied as an important data source of consumer responses. Therefore, the development or selection of a proper product design crowdsourcing platform should be the primary consideration. Actually, there are already some online platforms which can function as crowdsourcing platforms, such as Wikipedia. In regard to product design, pervasive sensor networks, internet services and social media, especially online product review websites, which actually contain abundant consumer responses, can provide a wide range of sources of customer responses. In this work, the development of a crowdsourcing platform is beyond the research scope. Therefore, existing crowdsourcing platforms, such as online product review websites, are treated as the main source of consumer responses.

Step 1: ***Content Extraction***. Web mining is applied to crawl web logs of crowdsourcing websites (e.g. product review websites) to extract meaningful content (i.e. consumer responses/comments). Considering the data formats contained in such websites are varied (e.g. textual data and graphics, which are mainly used in Web 2.0), textual data is mainly the focus in this study, since product reviews are often expressed in text, and posting texts is a comfortable and preferred way for consumers. Extracted content is collected as response documents. In particular, the content from one webpage is stored as one document, and all documents are treated as the corpus.

Step 2: ***Text Processing***. Text mining is applied to deal with the captured contents. Necessary operations include the discovery of textual patterns, tokenization to divide the documents into textual tokens (including steps like setting filter stop words and filter tokens by length, transforming cases into a certain pattern) and further pre-processing of the textual data (e.g. calculating TF-IDF, correlation, and similarity in a quantitative manner). Considering the further processing for identifying the affective design properties, N-gram generation is employed to capture short phrases, so as to fully retain the semantic context and ensure the accurate identification of the affective product properties.

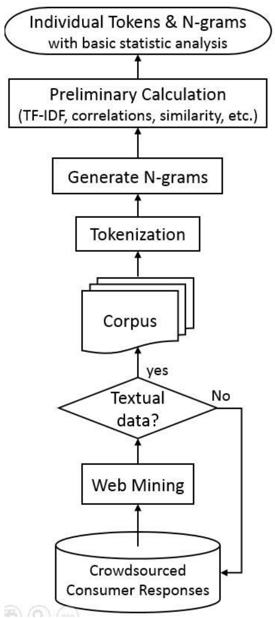


Figure 3. Workflow of Stage 1

*3.2. **Stage 2: Identification of Product Affective Properties***

The extracted word tokens (including individual tokens and n-grams) are examined from the perspectives of design and affectation, since the importance of product properties in design and consumer affectations is not the same. For example, the cpu chip of a smartphone is very important from the design perspective; however, it is not a property directly attracting consumers' emotional responses. Therefore, design and affective aspects should be jointly considered to ensure the accurate capture of important emotional properties. In particular, tokens (or n-grams) representing product attributes or specifications are identified as design property tokens (or n-grams). Tokens implying affectations, such as adjectives or nouns which have been defined as containing emotions in referring to some lexical database, are identified as affective tokens. Through identifying the linguistic connections between affective tokens and design property tokens, product affective properties can be found.

Step 3*: *Establishment of PDKH.

From the design perspective, a product design knowledge hierarchy (PDKH) is outlined to assist in the identification of design-related tokens. With reference to [48], product design (as functionality) can be considered layer by layer from abstract ideas to concrete specifications (as shown in Fig. 4). There are mainly five layers, i.e., product objectives, functional specifications, interaction design, interface design and sensory design. Following this structure, design can be developed in detail for each level. In particular, a hierarchical structure can be referred to for depicting the design knowledge at each level. As shown in Fig. 5, design knowledge can be analyzed from: what aspects should be considered; what kind of properties should be equipped; to what specifications should be set at each design level. In general, Fig. 4 presents the product design flow, while Fig. 5 provides the knowledge representation structure for each design stage. With the help of Figs 4 and 5, design knowledge can be fully examined and systematically presented.

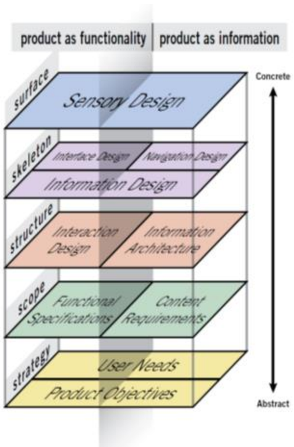


Figure 4. Five Layers of Design based on User

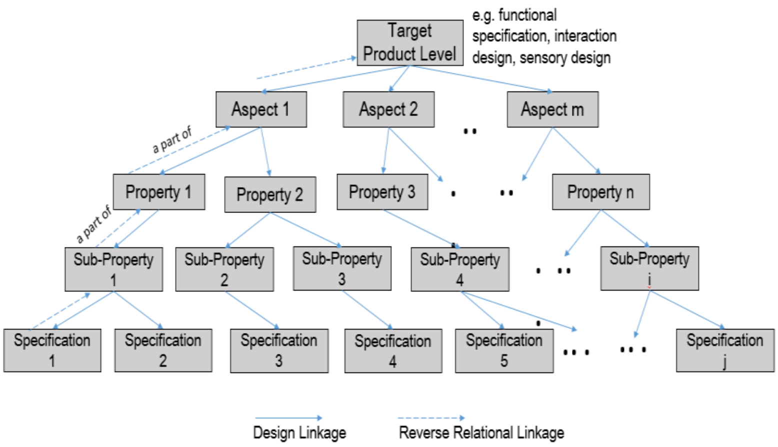


Figure 5. Product Design Knowledge Hierarchy

Experience [48]

For different products, the details of PDKH should be specified accordingly. Then the tokens extracted from the crowdsourced responses can be analyzed with reference to PDKH. For this purpose, a lexical database is needed to provide the Synset (semantic) relations and Word (lexical) relations. If tokens can be associated with the design knowledge in PDKH (e.g., subordination, similar, synonymous), the tokens can be identified as design tokens. Therefore, design properties can be found out from each document.

Step 4: Affective Identification.

For affective tokens, an electronic lexical database is necessary to provide the definition, lexical categories (e.g., nouns, verbs, adjectives and adverbs), semantic relations, word relations to help with the affective identification. Especially, a sentiment dictionary (e.g., The Dictionary of Affect in Language), which defines the sentiment contained by different words, is needed as important reference for polarity analysis and sentiment measurement. At this step, two main actions are executed, i.e., sentiment analysis to examine the affect (in the current study, polarity analysis is more empathized), and preliminary computation to outline the general sentiment trend.

Step 4.1: Sentiment analysis; In fact, emotions should be understood in context. The more fully the whole context is understood, the more exact is the estimation of the emotion or mood. Therefore, sentiment in this work is analyzed from two levels, i.e., sentence-level (local context) and document-level (whole context). The sentiment of certain words (or concepts in the lexical database) is estimated mainly in the local context, and whole context is considered to further amend the sentiment analysis results.

Firstly, polarity analysis is deployed to determine the sentiment polarity of each document. Polarity confidence can be calculated accordingly and treated as the sentiment score in whole context (denoted as SS_W and $SS_W \in [0, 1]$). With the help of the lexical database and sentiment dictionary, word tokens which contain affectations can be identified. Correspondingly, n-grams containing the affective tokens can be treated as affective n-grams. Afterwards, the sentiment of the sentences containing these affective tokens/n-grams can be analyzed, and the sentiment score in local context (denoted as SS_l and $SS_l \in [0, 1]$) can be obtained. Therefore, the sentiment score can be estimated by combining SS_W and SS_l.

If the polarity of the whole context is consistent with the local context, the sentiment is strengthened. Considering that the sentiment is measured mainly based on local context, the correction effect caused by the whole context can be treated as equivalent to modal adverbs, and thus a square operation can be applied [49].

$$SS = SS_l \bullet (1+|SS_W|^2)$$

If the polarity of the whole context is opposite to the local context, the sentiment is weakened.

$$SS = SS_l \bullet (1-|SS_W|^2)$$

Step 4.2: Preliminary computation; Based on the sentiment analysis results in the previous step, basic statistical analysis can be executed to show the general quantitative trend. For example, the ratio of the positive and negative responses, the correlations and similarities between different responses or different tokens. Descriptive statistics can be considered to describe the quantitative characteristics of the collected response. In general, this step aims to depict the rough affective performance of the product.

Through sentiment analysis, tokens containing affectations can be identified and treated as affective tokens.

Step 5: Establishment of Associations between Design Properties and Affect.

In steps 3&4, product property tokens and affective tokens have been identified. The subsequent consideration is to identify the connections between them, so as to further identify the product affective property.

Generally, one complete consumer response is treated as one document, so the xth document can be denoted as D_x. The identified token of D_x is denoted as T_{xy}. Assume the total number of identified tokens of D_x is m. Then D_x can be denoted as:

$$D_x : \{T_{x1}, T_{x2}, T_{x3} \cdots T_{xm}\}$$

The corpus consists for all documents. Assume the total number of documents is n:

$$\phi : \{D_1, D_2, D_3, \cdots D_x \cdots D_n\} = \{(T_{11}, T_{12}, T_{13} \cdots T_{1m_1}), (T_{21}, T_{22}, T_{23} \cdots T_{2m_2}), \cdots (T_{n1}, T_{n2}, T_{n3} \cdots T_{nm_n})\}$$

Domain ontology is leveraged to examine the semantic relations between affective tokens (*AT*) and product property tokens (*DT*). Thus it is easy to show that the *AT* and *DT* of xth documents are both subsets of D_x:

$$AT_x \subset D_x ; \quad DT_x \subset D_x$$

A relationship set U is defined to bridge affective tokens $at_i \in AT$, $\mathrm{AT} = \{at_i | i = 1, \dots, I\}$ and design property tokens $dt_j \in DT$, $\mathrm{DT} = \{dt_j | j = 1, \dots, J\}$. As shown in Fig. 6, the elements in *AT* are related to *DT* based on U, and an association matrix can be thusly obtained. The value of the elements of U is from the set {0, 1}. If *AT* and

DT are connected with each other, the *u* is 1; if not, *u* is 0. If an association relationship exists in a proper semantic context, the *dt* with associated *at* can be regarded as an affective product property.

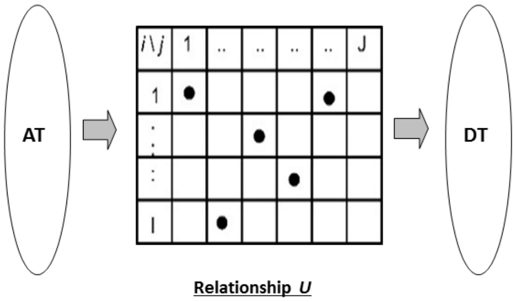


Figure 6. Association relationships between *AT* and *DT*.

Moreover, the local context and n-grams are also referred to identify U. If *AT* and *DT* are in the same local context (i.e., sentence-level context) or the same n-grams, they can be treated as connected with each other.

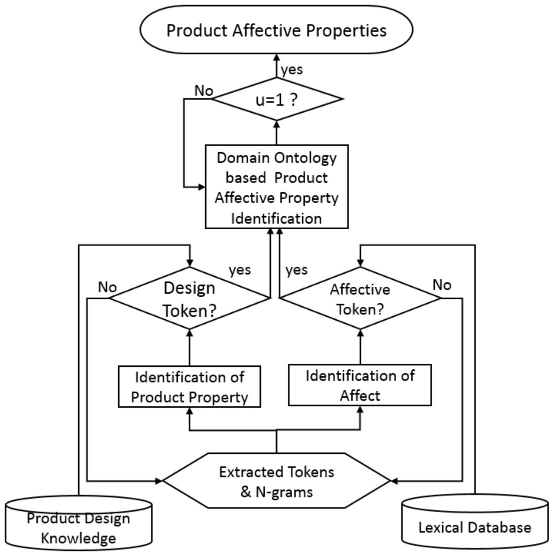


Figure 7. Workflow of Stage 2

3.3. ***Stage 3: Prioritization of Product Affective Properties***

The identified product affective properties are further prioritized according to their design importance and affective intensity. However, for the same design property, different consumers may have different emotional responses. Therefore, the effect on design properties is analyzed in their corresponding documents first, and then all the emotions on the same design properties are integrated to achieve the overall affective performance. Through combining the affective performance and design importance, the final priority of product affective properties can be obtained. Based on the priority, the product affective properties can be re-organized and ranked accordingly.

Step 6: Prioritization.

The priority of product affective properties is estimated from two perspectives: design and affect. The Design importance (*DI*) is calculated based on the total occurrence (i.e., $\mathrm{TO}_{dt_{xi}}$, *i*th design token in *x*th document), term frequency–inverse document frequency ($TF-IDF_{dt_{xi}}$, see Equation 1) and *PDKH* priority (i.e., priority subject to *PDKH*, $\mathrm{f}(h_{PDKH})_{dt_{xi}}$, see Equation 2) of the design property tokens. For $\mathrm{TO}_{dt_{xi}}$ and $TF-IDF_{dt_{xi}}$, they are used to measure if the property dt_{xi} is of concern to consumers.

$$TF-IDF_{dt_{xi}} = tf(dt_{xi}, D_x)\cdot idf(dt_{xi}, \phi) = (0.5 + 0.5\frac{f_{dt_{xi},D_x}}{\max_{dt_{xi}} f_{dt_{xi},D_x}})\cdot \log\frac{N}{n_{dt_{xi}}} \tag{1}$$

where f_{dt_{xi},D_x} is the number of times that design token dt_{xi} occurs in document D_x, N is total number of documents in the corpus ϕ, $n_{dt_{xi}}$ is the number of documents where the design token dt_{xi} appears.

According to [55], the earlier design stage is crucial to the final product quality and product life cycle cost, thus the design tokens at the more abstract levels should occupy heavier importance. With reference to [50-52], the depth function is useful to assign importance to the hirarchical structure.

$$f(h_{PDKH})_{dt_{xi}} = \frac{e^{\beta h_{PDKH}} - e^{-\beta h_{PDKH}}}{e^{\beta h_{PDKH}} + e^{-\beta h_{PDKH}}}; \beta > 0 \tag{2}$$

where h_{PDKH} is the depth of the design token to the top level (i.e. the most concrete level) in PDKH, so the more abstract design leve dt_{xi} is located, $f(h_{PDKH})_{dt_{xi}}$ is heavier; β is a smoothing factor, $\beta > 0$.

The more often dt_{xi} is discussed, it indicates a greater more importance to consumers. Moreover, if it only appears in one particular document, it implies that the design property is not of wide concern by consumers and may contain high risk of personal bias; thus, less importance should be assigned to this design property. Therefore, the design importance of dt_{xi}, i.e., $DI_{dt_{xi}}$ can be calculated using Equation 3.

$$DI_{dt_{xi}} = TO_{dt_{xi}} \cdot (1 - TF - IDF_{dt_{xi}})\cdot f(h_{PDKH})_{dt_{xi}} \tag{3}$$

For the same design property dt_i, the design importance estimated in different documents can be integrated to achieve the overall design importance to the corpus:

$$DI_{dt_i} = \sum_{x=1}^{N_{dt_i}} DI_{dt_{xi}} = \sum_{x=1}^{N_{dt_i}} TO_{dt_{xi}} \cdot (1 - TF - IDF_{dt_{xi}})\cdot f(h_{PDKH})_{dt_{xi}} \tag{4}$$

where N_{dt_i} is the total number of documents containing the design property dt_i.

Likewise, the affective intensity is estimated based on the polarity (i.e., $p_{at_{xj}}$, if positive, then $p_{at_{xj}} = +1$; if negative, then $p_{at_{xj}} = -1$), the total occurrence of affective token (i.e., $\mathrm{TO}_{at_{xj}}$), $TF-IDF_{at_{xj}}$ (see Equation 5) and sentiment score

$SS_{at_{xj}}$ (numerial value without polarity concern, see Equation 6).

$$TF-IDF_{at_{xj}} = tf(at_{xj}, D_x)\cdot idf(at_{xj}, \phi) = (0.5+0.5\frac{f_{at_{xj},D_x}}{\max_{at_{xj}} f_{at_{xj},D_x}})\cdot \log\frac{N}{n_{at_{xj}}} \tag{5}$$

where f_{at_{xj},D_x} is the number of times that affective token at_{xj} occurs in document D_x, N is total number of documents in the corpus, $n_{at_{xj}}$ is the number of documents where the design token at_{xj} appears.

As mentioned in Section 3.2, sentiment score is estimated based on the joint consideration of local and whole contexts in order to identify the exact emotional intensity of the affective token.

$$SS_{at_{xj}} = \begin{cases} SS_{lat_{xj}} \cdot (1+|SS_{wat_{xj}}|^2), & if \quad p_{lat_{xj}} = p_{wat_{xj}} \\ SS_{lat_{xj}} \cdot (1-|SS_{wat_{xj}}|^2), & if \quad p_{lat_{xj}} \neq p_{wat_{xj}} \end{cases} \tag{6}$$

where $SS_{lat_{xj}}$ is the sentiment score in local context (sentence-level), and $SS_{wat_{xj}}$is the sentiment score in the whole context (document-level).

Thus, the affective intensity of affective token at_{xj}, i.e., $AI_{at_{xj}}$, can be calculated using Equation 7.

$$AI_{at_{xj}} = p_{at_{xj}} \cdot TO_{at_{xj}} \cdot (1-TF-IDF_{at_{xj}}) \cdot SS_{at_{xj}} \tag{7}$$

For the same design property dt_i, the associated affect in the whole corpus can be accumulated in two ways. One is the sum of absolute values of all associated affectations to reflect the total intensity of the whole emotion attached to the property (see Equation 8). Another is the algebraic sum of all associated affectations including their polarities in order to determine the overall polarity of the design property (see Equation 9).

$$|AI| \ for \ dt_i = \sum_{x=1}^{N_{dt_i}} \sum_{j=1}^{n_{at_{xi}}} AI_{at_{xj}} = \sum_{x=1}^{N_{dt_i}} \sum_{j=1}^{n_{at_{xi}}} TO_{at_{xj}} \cdot (1-TF-IDF_{at_{xj}}) \cdot SS_{at_{xj}} \tag{8}$$

where N_{dt_i} is the total number documents containing dt_i, $n_{at_{xi}}$ is the total number of affective tokens which are associated to design property dt_i in D_x.

$$AI \ for \ dt_i = \sum_{x=1}^{N_{dt_i}} \sum_{j=1}^{n_{at_{xi}}} AI_{at_{xj}} = \sum_{x=1}^{N_{dt_i}} \sum_{j=1}^{n_{at_{xi}}} p_{at_{xj}} \cdot TO_{at_{xj}} \cdot (1-TF-IDF_{at_{xj}}) \cdot SS_{at_{xj}} \tag{9}$$

where N_{dt_i} is the total number documents containing dt_i, $n_{at_{xi}}$ is the total number of affective tokens which are associated to design property dt_i in D_x.

One product affective property represents one design property and the associated affectation. Thus an affective productproperty ($P_{xi,xj}$) can be denoted as $at_{xj} \cdot dt_{xi}$. A weighted calculation is introduced to integrate *DI* and *AI* in order to achieve the overall priority (i.e., the priority of design property *i* with affect j, $OP_{xi,xj}$, Equation 10).

$$OP_{xi,xj} = \omega_d DI_{dt_{xi}} + \omega_a \left| AI_{at_{xj}} \right| \tag{10}$$
$$\omega_d + \omega_a = 1;\ 0 < \omega_d < 1\ \&\ 0 < \omega_a < 1$$

where w_d is the weight of design importance, w_a is the weight of affective intensity; |AI| is considered, since no matter positive or negative responses, the more the absolute value, the more it is concerned by consumers.

For one product affective property, the overall priority (OP) to the corpus can be accumulated as:

$$OP_i = \sum_{x=1}^{N_{dt_i}} \sum_{j=1}^{n_{dt_i}} OP_{xi,xj} = \sum_{x=1}^{N_{dt_i}} \sum_{j=1}^{n_{dt_i}} (\omega_d DI_{dt_{xi}} + \omega_a \left| AI_{at_{xj}} \right|) \tag{11}$$

where N_{dt_i} is the total number documents containing dt_i, $n_{at_{xi}}$ is the total number of affective tokens which are associated to design property dt_i in D_x.

Step 7: Product Affective Property Representation.

Every product affective property can be denoted as "design property (*DI*, *AI, OP*)". Generally, a higher *OP* means a higher integrated priority, thus more attention should be paid. In particular, for *AI*>0, it represents the product properties with a positive consumer affectation; therefore, such properties could be the strength of this design. On the contrary, for *AI*<0, it indicates the product properties cannot satisfy the consumers and maybe a weakness in the product. As proposed in Fig. 8, normalized *DI* and *AI* can be used to segment product properties into different sections according to their affective performance. Strength and weakness can be easily recognized, and valuable references as to which properties, to what degree, should be improved can be achieved.

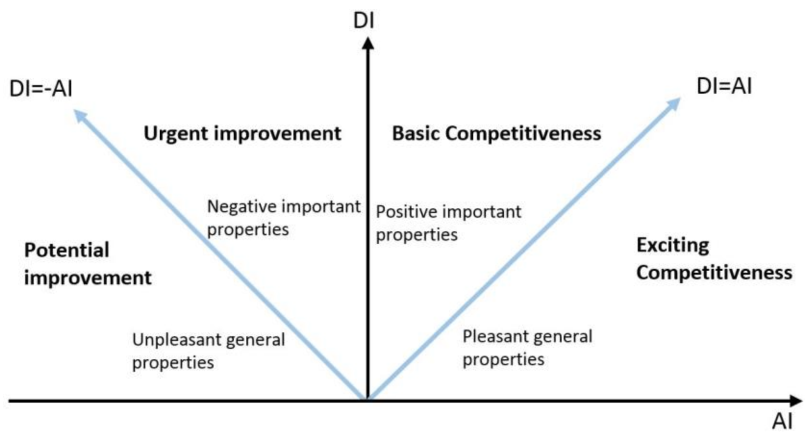


Figure 8. Strategic segmentation of product design properties according to *DI* and *AI*

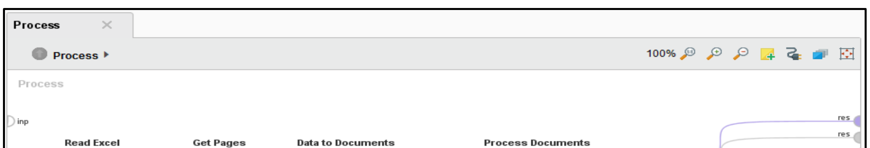
4. A Pilot Study

A pilot study on iPhone 7 was conducted to demonstrate the proposed approach, since mobile phone design is often used for case studies in APD research. To facilitate the comparison between the proposed approach and other methods, the cellphone is used in the pilot study. As the development of crowdsourcing platform is not the focus of this study, existing crowdsourcing websites are considered as data sources. Amongst them, Amazon and CNET, which are two important product review platforms that often post topics for collecting comments from Internet users, are selected as the crowdsourcing product review resources. Therefore, review posts and comments under the topic of iPhone 7 are targeted.

Normally, the comments are presented in a combination of textual descriptions and product pictures. Considering that texts are widely preferred by Internet users to express their opinions, and image processing is beyond the scope of this work, textual content is the focus and non-textual content is removed. Moreover, considering that the qualities of the posted reviews vary greatly (viz. some are very rough and grammatical mistakes frequently appear) and not all of them contain sufficient valuable information, a filtering process is deployed. Too rough responses (less than 150 words) are removed, and responses with too many grammatical mistakes are removed, as well. Hence, 57 documents were captured in total in this preliminary study. The post dates range from 2016 October to 2017 May. The complete content of one review webpage is treated as one document, and the collection of all documents is treated as the corpus.

4.1. ***Content Extraction and Text Analytics***

The first stage is actually a data mining process, where web mining and text mining are applied to extract useful textual information from the crowdsourced responses. For this purpose, a powerful tool *RapidMiner* is used to crawl the product review webpages and perform basic text analysis. The main processes are deployed as *Read Excel*, *Get Pages*, *Data to Documents*, and *Process Documents* (Fig. 9). In particular, the target web links are collected and the URLs are recorded in excel. The *Read Excel* operator is used to read excel and recognize the details in each cell. The *Get Pages* operator is functional in crawling web sites, extracting HTML content and verifying if the content value matches the expected value type. Each design document is regarded as one example set. *Data to Documents* is utilized to convert the example sets to an object collection where each row represents each example document. Having the organized object collection, the *Process Document* operator is responsible for specifically coping with the example sets and is executed by the sub-processes of *Tokenize*, *Transform Cases*, *Filter Stopwords (by English)*, *Filter Tokens (by length)* and *Generate n-Grams (terms)*. Therein, *Tokenize* is used to separate the textual content into individual word tokens. *Transform Cases* is to convert the word tokens into a consistent case and avoid repetitive count of the same words, so that the text analysis is not case-sensitive. *Filter Stopwords (by English)* and *Filter Tokens (by length)* are employed to control the tokenization process and restrict the representation of the tokens. For example, if the value of *Filter Tokens (by length)* is set as five, that means the words longer than five letters are represented as the first 5 letters. To fully and correctly understand consumers' emotions, their original expressions are extremely important and should be retained. Therefore, *Generate*



n-Grams (terms) (in this study, N is set as 3) are deployed to generate semantically meaningful short phrases.

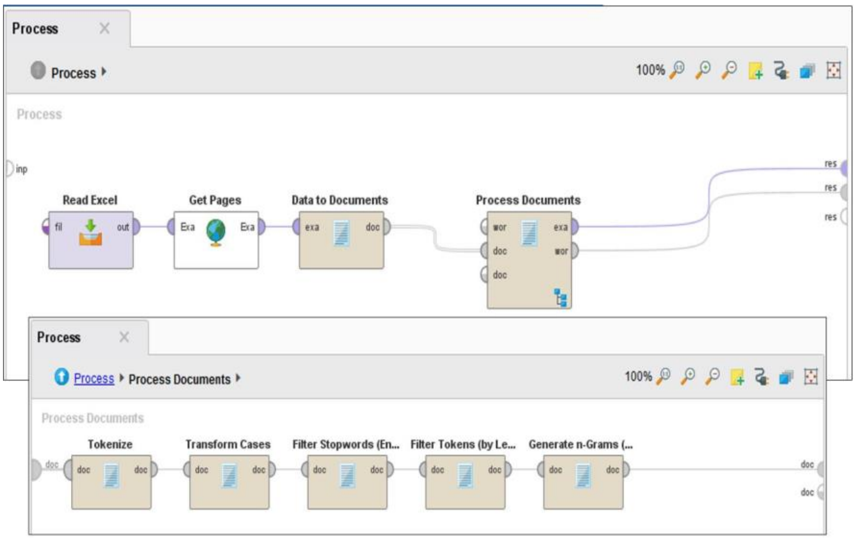


Figure 9. Extraction of Online Consumer Responses with Rapid Miner

In addition, *Correlation Matrix*, *Data to Similarity* and *Clustering* can also be considered to provide simple analysis of the captured documents (Fig. 10). However, the correlations and similarities in these operators are calculated, mainly based on numerical vectors such as TF-IDF, rather than the affective performance.

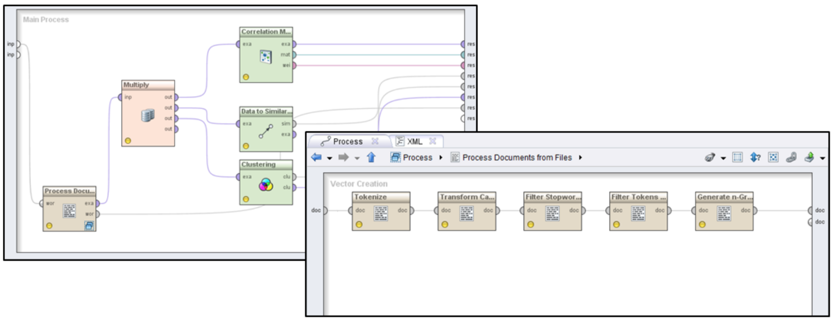


Figure 10. Preliminary Analysis of Online Consumer Responses with Rapid Miner

14426 attributes (including single word tokens, 2-gram terms and 3-gram terms) were obtained.

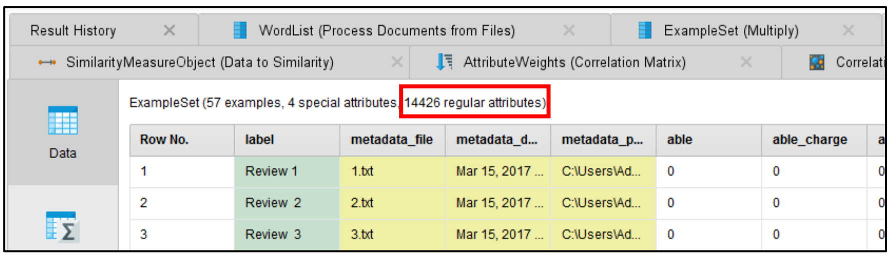


Figure 10. *RapidMiner* extraction results.

The table below presents a list of examples of the extracted tokens (including single tokens, 2-gram tokens and 3-gram tokens)

Table 1. Examples of extracted tokens

absurd	loss	love
absurd_mounts	loss_headphone	love_android
absurd_mounts_cash	loss_headphone_jack	love_android_good
accept	loss_price	love_apple
accept_card	loss_price_pods	love_apple_products
accept_card_calls	lost	love_button
accept_card_phone	lost_android	love_button_took
accept_want	lost_android_root	love_fastest
accept_want_updated	lost_destroyed	love_fastest_device
access	lost_destroyed_expensive	love_improvements
access_apps	lost_easier	love_improvements_came
access_apps_basic	lost_easier_send	love_iphone
access_framerate	lost_focus	love_iphone_reason
access_framerate_resolution	lost_focus_beautiful	love_larger
access_tech	lost_time	love_larger_screen
access_tech_relied	lost_time_included	love_nonsense
accessories	lost_truth	love_nonsense_people
accessories_adaptors	lost_truth_matter	love_portrait
accessories_adaptors_integrated	Lots	love_portrait_mode
accessories_companies	lots_people	love_review
accessories_companies_iphones	lots_people_care	love_review_totally
	etd.	love_tech
edt.		love_tech_phone

4.2. *Identification of Product Affective Properties*

For the identification of product property tokens, PDKH is referred to in examining if these word tokens are related to design knowledge of a Smartphone. An electronic lexical database *WordNet* is introduced to provide lexical and semantic references. By the use of *WordNet*, the definitions, lexical categories (e.g., nouns, verbs, adjectives and adverbs), semantic relations and word relations can be recognized.

Generally, the extracted word tokens are marked as *Concepts*, *Entities* and *Others*. The tokens tagged as *Concepts* and *Entities* are examined to determine if they belong to Smartphone design knowledge. A set of proper semantic relations is selected based on domain ontology to narrow the connections down to certain relations with relatively higher importance. Referring to previous ontology-related studies, "*synonymous*", "*Meronym*" (a part of) and "*Hypernym*" (a kind of) are frequently studied, and thus used in this study to identify the associations between the word tokens and PDKH. Two main kinds of associations are considered: 1) whether these tokens belong to any corresponding design levels, and 2) whether these tokens are semantically related to the tokens which have been confirmed to be design property tokens. To explain, Fig. 12 shows the association process of tokens extracted from Document 1 based on PDKH. With the help of these two kinds of relations in PDKH, tokens which represent design knowledge can be identified and treated as design tokens.

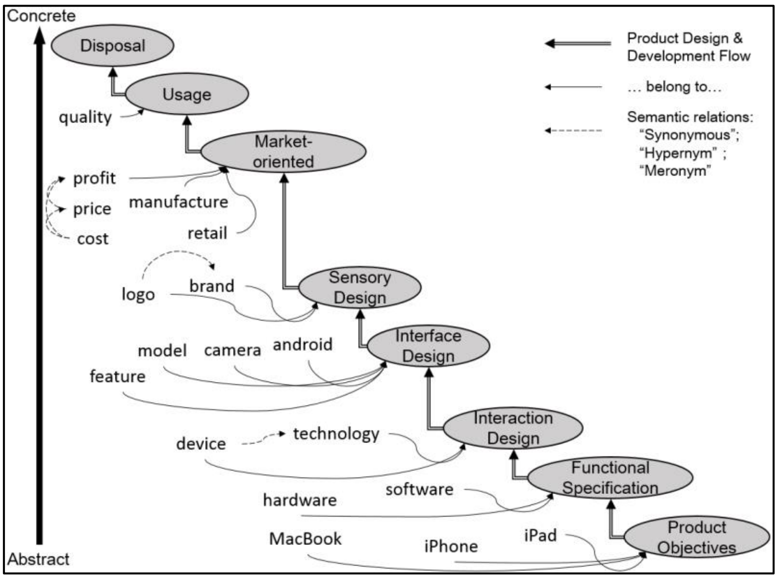


Figure 12. Association of Individual Tokens of Document 1 referring to PDKH

For sentiment analysis, Text Analysis API, a package of Natural Language Processing, Information Retrieval and Machine Learning tools for extracting meaning and insight from textual and visual content with ease, is introduced to analyze the emotions contained in the collected responses. In general, API can help to analyze the emotions in every sentence and calculate the average sentiment score of all sentences in the document. Therefore, different APIs with different sentiment dictionaries may lead to different results. For this reason, three APIs, i.e., *AYLIEN API for documents*, *AYLIEN API for social media*, and *Meaningcloud API*, are applied to take advantage of multiple sentiment dictionaries in order to achieve relatively accurate estimation of the emotions. Denoting p_{at_j} as the polarity of token AT_j, the polarities are identified by the three APIs p_{ADat_j}, p_{ASat_j}, and p_{MCat_j}, respectively. The polarity is from the set of {positive, negative, neutral}. The sentiment scores estimated by the three APIs are SS_{ADat_j}, SS_{ASat_j} and SS_{WCat_j}. The sentiment analysis results by these three tools can be processed as follows:

a. If $p_{ADat_j} = p_{ASat_j} = p_{MCat_j}$, then $p_{at_j} = p_{ADat_j} = p_{ASat_j} = p_{MCat_j}$, and the sentiment score SS_{at_j}=Avg (SS_{ADat_j}, SS_{ASat_j}, SS_{WCat_j});
b. If $p_{ADat_j} = p_{ASat_j} \neq p_{MCat_j}$, then $p_{at_j} = p_{ADat_j} = p_{ASat_j}$, and SS_{at_j}=Avg (SS_{ADat_j}, SS_{ASat_j}); that is to say, p_{at_j} adopts the polarity of the majority, and SS_{at_j} is the average of the sentiment scores of the majority;
c. If $p_{ADat_j} \neq p_{ASat_j} \neq p_{MCat_j}$, then expertise is incorporated to identify the p_{at_j}, and SS_{at_j} adopts the value in which the polarity is consistent with expert opinion.

Based on the sentiment analysis by *AYLIEN API for documents*, *AYLIEN API for social media*, and *Meaningcloud API*, the overall sentiment analysis results can be achieved (Table 2). As shown in Fig. 13 a & b, although there are more positive responses than negative ones, the difference is not significant. It indicates that consumers still have dissatisfaction and doubts on this product, which further enhance the need to investigate the emotions in regard to specific design properties.

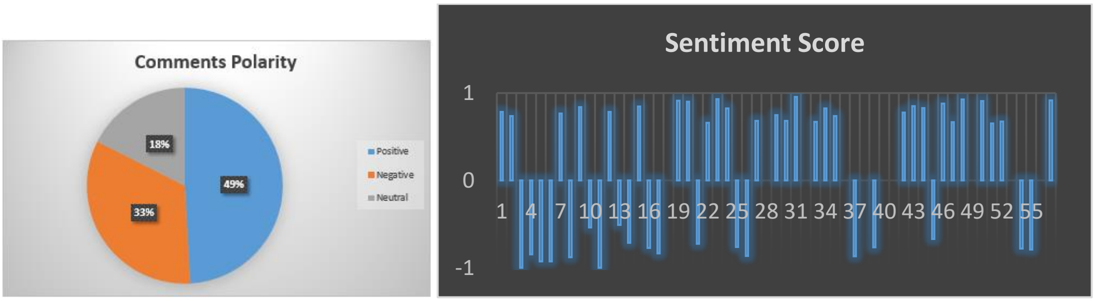


Figure 13. a) The proportion of documents with different polarities; b) Sentiment scores of all documents

Table 2. Sentiment Analysis Results by Different APIs

Sentiment Analysis by APIs		Mean	Standard Error	Number of Documents
Meaningcloud API	Positive	0.86	0.0072	31
	Negative	-0.86	0.0113	13
AYLIEN API for social media	Positive	0.43	0.1312	18
	Negative	-0.79	0.0357	25
AYLIEN API for documents	Positive	0.87	0.0285	32
	Negative	-0.75	0.0527	17
Overall	Positive	0.80	0.0183	28
	Negative	-0.80	0.0304	19

With the help of APIs, affective tokens can be identified. Two types of affective tokens are examined, i.e., one kind is the affective tokens containing emotions, and another kind involves the concepts/entities which have been identified with emotions in their local context. Some examples of identified affective tokens are listed below:

Table 3. A list of examples of identified affective tokens

iPhone (N, R1)	Microsoft Corporation (P, R2)	host (P+, R2)	chip (P+, R2)
Apple (N, R1)		compass (P, R2)	language (N, R3)
Macbook (N, R1)	LG (P+, R2)	apple (P, R2)	battery (Neu, R3)
Company (P, R1)	RAM (P+, R2)	report (P, R2)	sim (N, R4)
Love (P, R1)	Samsung (N, R2)	variant (P, R2)	button (N, R4)
High-end (P, R1)	Apple apps (P, R2)	telephone (P, R2)	telephone (N, R5)
Innovation (P, R1)	S-AMOLEDS (P+, R2)	ecosystem (N, R2)	cover (P, R5)
Product (N, R1)	SD (P+, R2)	competition (P, R2)	accessory (P, R5)
Model (P, R1)	FHD (N, R2)	dollar (P, R2)	Internt (N, R6)
iCloud (P, R2)	Amazon (P+, R2)	camera (P+, R2)	Wifi (N, R6)
iPhone 5s (P, R2)	Google maps (P, R2)	system (P+, R2)	iPhone (N, R7)
honor (P, R2)	update (P, R2)	standard (P+, R2)	screen (P, R7)
ios (N, R2)	application (P, R2)	plug (P, R2)	Camera (P, R7)
Android (P+, R2)	stock (P)	interest (P)	Music (P, R7)

Chrome (P, R2)	feature (P) storage (P)	license (P)	Speaker (N, R8)

* The first denotation in brackets represents the polarity (P is positive, N is negative, and NEU is neutral); the second represents the documents from which the token is extracted.

Afterwards, the relationships between design tokens and affective tokens are examined with joint consideration of the semantic relations and lexical reference, and the design tokens which are successfully associated with affective tokens are regarded as product affective properties. As results, more than 400 product affective properties have been identified. Some examples are presented in Table 4.

Table 4. Examples of Identified Product Affective Properties (N-grams).

Product Affective Property Examples		
amazed_smoothness	beautiful_designed_products	cheap_headphone
amazed_smoothness_controls	beautiful_device	cheap_water_resistance
amazing_cpu	beautiful_device_look	cheaper_amoled_screen
amazing_cpu_wasted	bigger_battery	chip_powerful
amolded_screen_icing	bigger_battery_play	connector_expandable
amolded_display_pleasing	bigger_battery_speaker	decent_cell_phone
battery_heavy_use	bigger_screen	decent_headphone
battery_issue_timed	buds_excellent	device_simpler
battery_life_excellent	buds_excellent_durable	device_simpler_frustrating
battery_life_great	camera_fantastic	device_audiophile_old
battery_life_lower	camera_fantastic_work	disappointed_fast_charging
battery_life_shorter	camera_faster	dish_pretty_good
battery_life_terrible	camera_faster_processor	display_pleasing
etc…	etc…	etc…

4.3. ***Prioritization of Affective Design Properties***

The *DI* and *AI* of each product affective property are calculated using Equations 1-9, and overall priority of product affective properties can be computed using Equations 10 and 11 (in this study, *DI* and *AI* are assigned with the same weight, namely, 0.5). To enable a direct comparison and integration of *DI* and *AI*, the calculated results of *DI* and *AI* by Equations 1-9 are normalized into the range of [0, 1], and denoted as N(*DI*) and N(*AI*). *AI* is the aggregated intensity of all affectations, namely, the sum of the absolute values of all positive and negative affectations. The final polarity is determined by the algebraic sum of all affectations including their polarities. If the algebraic sum is positive, the final polarity is positive, and vice versa. Moreover, the ratio of algebraic sum to *AI* will determine the intensity level of final polarity. If the ratio is significant (e.g. larger than 20%), it means the final polarity is much stronger than other polarities, and the polarity can be presented by two positive/negative signs.

Considering the large number of product affective properties extracted from online consumer responses, the properties with higher priorities are listed in Table 5 below. Individual design properties are used to represent all the relevant n-gram product affective properties.

Table 5. Product Affective Properties with Higher Priority.

Affective Design Properties	Count	DI	N(DI)	AI	N(AI)	OP	Polarity*
Camera	118	86.26	1.000	29.57	0.963	**0.982**	(++)
Headphone, headphone-jack, headphones, jack	36	26.58	0.255	30.63	1.000	**0.627**	(+)
Battery	44	32.86	0.333	22.02	0.700	**0.517**	**(-)**
Sound, sounds	30	21.74	0.194	19.89	0.626	0.410	(++)
Charging,charge, charged, charger	42	30.98	0.310	14.94	0.453	0.382	**(--)**
Look, looks	58	25.97	0.247	13.53	0.404	0.326	(++)
Screen	18	13.33	0.089	11.23	0.324	0.207	(++)
Size	28	19.74	0.169	7.49	0.194	0.181	(++)
Bluetooth	21	14.96	0.109	7.54	0.196	0.153	**(--)**
Adapter, adapters	21	15.46	0.116	7.33	0.188	0.152	**(--)**
Plug	27	23.31	0.214	4.29	0.082	0.148	(++)
Music	22	19.03	0.160	5.66	0.130	0.145	(++)
Tech, technologies, technology	25	10.96	0.059	4.81	0.101	0.080	(++)
Lightning	10	7.33	0.014	1.95	0.001	0.007	**(-)**
Software	14	6.20	0.000	1.93	0.000	0.000	(++)

*+ means there are stronger positive responses than negative ones for the property; - means there are stronger negative responses for the property; the number of + and – implies how much this affect is stronger than the other, for example, ++ means there are significantly stronger positive responses about the property.

According to the results, it was found that the market performance of iPhone 7 is generally good. In particular, the camera, headphone, and battery have the heaviest priorities and significant attention should be paid during the design and improvement process. Especially for the camera, it is the most frequently mentioned property and has received intense emotional feedback from consumers, indicateing the necessity to pay attention to camera design so as to ensure users' satisfaction. For positive design properties, camera, screen, technology and appearance are the traditional strengths of iPhone and still receive strongly positive comments from consumers in this study. Therefore, the advantages of iPhone 7 on these positive properties should be maintained in future generations.

On the other hand, the battery and charging-related issues are the main pain points of iPhone 7. Referring to the market performance of iPhone 6, the battery problem has been widely complained about and discussed among consumers, so the battery and charging performance of iPhone 7 is the focus, and questioned about by consumers. For *adapters*, it can be considered jointly with "*Lightning*" which is also negative, since a new change for iPhone 7 is the "*Lightning to Headphone Jack Adapter*". The new change brings some inconvenience to consumers, since they need one more adapter and cannot charge and use headphones simultaneously. Moreover, *Bluetooth* is also commented on negatively, which indicates that an improvement on this traditional function may be needed.

In summary, the usual strengths of iPhone have been further proven, and unsatisfactory properties of iPhone 7 are examined. Since the comments analyzed in this pilot study were posted during the half year after iPhone 7 was newly launched, it indicates that not all new changes were well accepted by consumers. Therefore, the analysis on the important affective properties appears particularly important to identify the negative properties and hints at possible improvement directions. Furthermore, the prioritization provides clear and practical reference for designers as to which property should be paid more attention and how to deploy a further design process.

4.4. ***Comparison with Existing APD Studies***

In this section, two groups of comparisons are executed: i) properties extracted from online resources vs. properties evaluated by surveys; and ii) product affective properties identified based on joint consideration of affect and design vs. based on pure sentiment analysis.

Group 1: *Affective design properties extracted by the proposed approach vs. design properties used in existing APD studies*

Referring to Table 6, the proposed approach was compared with other APD studies of on mobile phones. Through the comparison between the design properties extracted from crowdsourced consumer responses and the design attributes used in existing APD studies (as shown in Table 6), it can be seen that the proposed approach is able to discover more design properties which are of concern by consumers and have provoked consumers' emotions, while existing APD studies are concerned with very limited design attributes. Moreover, these properties cover different product design aspects ranging from hardware, software to marketing issues. It suggests the stronger capability of the proposed approach in unearthing all the potential design concerns significantly influencing consumers' emotion.

Group 2: *Prioritization of product affective properties based on the joint considerations of design and affective concerns vs. based on only sentiment analysis*

The prioritization based on affective intensity and the joint consideration of design and affectation are presented in Table 7.

Table 7. Prioritization based on affective concern vs. based on the joint consideration of affective and design concerns

Prioritization based on affective intensity (AI)	**Prioritization based on the overall priority (OP)**
Headphone, headphones	Camera, cameras
Camera, cameras	Headphone, headphones
Battery, batteries	Battery, batteries
Sound, sounds	Sound, sounds
Screen, screens	Screen, screens
Charging, charge, charged,	Charging, charge, charged,

charger	charger
Plug	Plug
Music	Tech, technologies, technology
Tech, technologies, technology	Music
Adapter, adapters	Adapter, adapters
Look, looks	Bluetooth
Lightning	Look, looks
Bluetooth	Lightning
Size	Size
Software	Software

To understand the difference between the two prioritization results, a focus group was organized. 5 design experts including researchers and engineers with more than 5 years' experience in the industrial design area were involved. They were invited to give evaluation on the design properties according to their expertise and user interaction experience. Generally, the ranking of some design properties is consistent, thus the comparison focuses on the properties with different priorities. In particular, an AHP process is employed. The fundamental AHP scale (1: Equal importance, 3: Moderate importance, 5: Strong importance) is used to compare the properties with different priorities. 1 was assigned to the property with lower priority as basis evaluation score.

Table 8. A brief summary of AHP evaluation results

Camera vs. Headphone	3:1	Camera is one crucial part of iPhone and undoubtedly one of the most cared about functions of users. Therefore, it should be assigned with more priority.
Technology vs. Music	3:1	Technology is one important design aspect. Ofentimes, more effort is devoted to the development of innovative technologies. Moreover, the new technology is always an essential factor to appeal to fans or consumers.
Look vs. Lightning vs. Bluetooth	5:1:3	Industrial design is one success factor of iPhone, thus the appearance/form design is undoubtedly important, no matter for designers or consumers; bluetooth relates to multiple functions such as file transfer and connecting to car bluetooth and other wearable electronics, thus could be assigned with more priorities compared with lighting.

With expertise incorporation, it was found that expert opinions are more consistent with the prioritization results by the proposed approach. It indicates that the prioritization based on an integrated consideration of both design and affective concerns is relatively more reasonable. It is promising to provide useful guidance for designers in the design process. Therefore, compared with product property analysis methods from only design perspective, the proposed approach is user/consumer-centered and contains a higher potential to achieve satisfying products.

Compared with pure sentiment analysis, the proposed approach involves design concerns and is able to give more comprehensive and rational analysis results. Compared with manual processing of consumers' emotional feedback, the data mining process of the proposed approach is more efficient and scientific.

Table 6. Comparison of the properties extracted by the proposed approach and properties evaluated by surveys

	General Impression		Hardware								Software						Marketing		
	Appearance/ Form	Technology	Screen	Sensors	Audio / Output	Battery	Camera	Storage	SIM Card	Accessory	Interface	Phone	Internet Connectivity	Text Input	Email Text Massage	Third-party Application	Payment	Service	delivery
The proposed approach	√	√	√	√	√	√	√	√	√	√		√	√			√	√	√	√
Yang, 2011 [3]	√																		
Seva et al., 2011 [10]	√																		
Chan et al., 2011, [47]	√																		
Fung et al., 2012, [8]	√																		
Bhandari et al., 2017 [19]											√					√			
Jiang et al., 2015 [24]	√		√																

5. Discussion and Conclusions

This work aims to develop a product affective property identification approach. For this end, a web- and text-mining process is deployed to make use of online product review resources; capture useful consumer responses and perform textual analysis. Afterwards, domain ontology and electronic lexical database are utilized to provide semantic relations and lexical reference to identify design-related and affect-related word tokens, and furthermore, to assist in the association between design tokens and affective tokens. The design properties which are related with affective tokens are regarded as affective design properties. The design importance and affective intensity of the affective properties are estimated, and overall priority of these properties can be accordingly achieved. Through a pilot study, it has been demonstrated that the proposed approach is capable of capturing more possible affective design properties, and a clear and practical reference in terms of the priorities of different design properties can be achieved so as to facilitate decision making and product improvement.

However, there are still some limitations of this work. For example, the identification of affectation is based on sentiment analysis, which actually relies on the recognition and measurement of affective words (which have been defined and tagged in existing lexical databases or sentiment dictionaries). If consumers' statements do not include obvious affective words, the feeling may not be detected. Moreover, since the focus of this work is the identification of affective design properties, consumers' emotions are not investigated in very specific types, such as happiness, sadness and confusion. In future research, consumers' emotions will be further studied.

In conclusion, this study explores a product affective properties identification approach based on data mining and is promising in contributing to i) the discovery of all possible important affective design properties through taking advantage of abundant online consumer responses; ii) an efficient computation method to deal with large among of online data, iii) the integration of design knowledge into affective design so as to achieve more comprehensive and rational understanding of customer emotions, and iv) prioritization of affective design properties, which can be practically useful reference to facilitate product design and improvement.

Acknowledgement

This work was supported by Shanghai Pujiang Talent Program, project no. 17PJC054.

References

[1] C. Osgood, G. Suci, & P. Tannenbaum, (1967). The measurement of meaning. Urbana-Champaign, IL: University of Illinois Press.

[2] A.M. Lokman, K.A. Kamaruddin, (2010). Kansei Affinity Cluster for Affective Product Design, 2010 International Conference on User Science Engineering (i-USEr).
[3] C.-C. Yang, (2011). Constructing a hybrid Kansei engineering system based on multiple affective responses: Application to product form design, Computers & Industrial Engineering, 60: 760–768.
[4] Nagamachi, M., 2002. Kansei engineering as a powerful consumer-oriented technology for product development. Appl. Ergon. 33, 289–294.
[5] Park, J., & Han, S. H. (2004). A fuzzy rule-based approach to modeling affective user satisfaction towards office chair design. International Journal of Industrial Ergonomics, 34, 31–47.
[6] L.-Y Zhai, L.-P. Khoo, Z.-W. Zhong, (2009). A rough set based decision support approach to improving consumer affective satisfaction in product design, International Journal of Industrial Ergonomics, 39: 295–302.
[7] Hsiao, S. W., & Tsai, H. C. (2005). Applying a hybrid approach based on fuzzy neural network and genetic algorithm to product form design. International Journal of Industrial Ergonomics, 35, 411–428.
[8] K.Y. Fung, C.K. Kwong, K.W.M. Siu, K.M. Yu, (2012). A multi-objective genetic algorithm approach to rule mining for affective product design, Expert Systems with Applications, 39: 7411–7419.
[9] C.-C. Yang, M.-D. Shieh, (2010).A support vector regression based prediction model of affective responses for product form design, Computers & Industrial Engineering, 59: 682–689.
[10] R.R. Seva, K.G.T. Gosiaco, Ma. C. E.D. Santos, D.M.L. Pangilinan, (2011). Product design enhancement using apparent usability and affective quality, Applied Ergonomics, 42: 511-517.
[11] R.R. Seva, H.B.-L. Duh, M.G. Helander, (2007). The marketing implications of affective product design, Applied Ergonomics, 38: 723–731.
[12] Weihua Lu, Jean-François Petiot, (2014). Affective design of products using an audio-based protocol: Application to eyeglass frame, International Journal of Industrial Ergonomics 44: 383-394.
[13] Kwong C.K., Jiang H., Luo X.G, (2016). AI-based methodology of integrating affective design, engineering, and marketing for defining design specifications of new products, Engineering Applications of Artificial Intelligence, 47: 49–6.
[15] Supavich (Fone) Pengnate, Rathindra Sarathy, (2017). An experimental investigation of the influence of website emotional design features on trust in unfamiliar online vendors, Computers in Human Behavior 67: 49-60.
[16] R. Gennaria, A. Melonioa, D. Raccanellob, M. Brondino, G. Dodero, M. Pasinib, S. Torello, (2017). Children's emotions and quality of products in participatory game design, International Journal of Human – Computer Studies, 101: 45–61.
[17] Lilia R. Prado-León, (2015). Color preferences in household appliances: data for emotional design, Procedia Manufacturing, 3: 5707 – 5714.
[18] Li-Chieh Chen, Po-Ying Chu, (2012). Developing the index for product design communication and evaluation from emotional perspectives, Expert Systems with Applications, 39: 2011–2020.
[19] Upasna Bhandari, Tillmann Neben, Klarissa Chang, (2017). Wen Yong Chua, Effects of interface design factors on affective responses and quality evaluations in mobile applications, Computers in Human Behavior, 72: 525-534

[20] Yuexiang Huang, Chun-Hsien Chen, Li Pheng Khoo, (2012). Kansei clustering for emotional design using a combined design structure matrix, International Journal of Industrial Ergonomics, 42: 416-427.
[21] Salman Nazari-Shirkouhi, Abbas Keramati, (2017). Modeling customer satisfaction with new product design using a flexible fuzzy regression-data envelopment analysis algorithm, Applied Mathematical Modelling, 50: 755–771
[22] Rosemary R. Seva, Katherine Grace T. Gosiaco, Ma. Crea Eurice D. Santos, Denise Mae L. Pangilinan, (2011). Product design enhancement using apparent usability and affective quality, Applied Ergonomics, 42: 511-517
[23] Yuexiang Huang, Chun-Hsien Chen, Li Pheng Khoo, (2012). Products classification in emotional design using a basic-emotion based semantic differential method, International Journal of Industrial Ergonomics, 42: 569-580.
[24] Huimin Jiang, C.K. Kwong, K.W.M. Siu, Y. Liu, (2015). Rough set and PSO-based ANFIS approaches to modeling customer satisfaction for affective product design, Advanced Engineering Informatics, 29: 727–738
[25] Nur Hazwani Zolkifly, Shahrel Nizar Baharom, (2016). Selling Cars through Visual Merchandising: Proposing Emotional Design Approach, Procedia Economics and Finance, 37: 412–417
[26] Jose A. Diego-Mas, Jorge Alcaide-Marzal, (2016). Single users' affective responses models for product form design, International Journal of Industrial Ergonomics, 53: 102-114
[27] Rosemary R. Sev, Henry Been-Lirn Duh, Martin G. Helander, (2007). The marketing implications of affective product design, Applied Ergonomics, 38: 723–731.
[28] J. Howe,Theriseofcrowdsourcing,WiredMag.14.06(2006)1–5.
[29] Y.Q. Lv, K.M. Lee, Z. Wu, Fuzzy theory applied in quality management of distributed manufacturing system: A literature review and classification, Eng. Appl. Artif. Intel. 24 (2) (2011) 266–277.
[30] I. Sousa, D. Wallace, Product classification to support approximate life-cycle assessment of design concepts, Technol. Forecast. Soc. 73 (3) (2006) 228–249.
[31] C.Y. Lee, B. Song, Y.T. Park, Design of convergent product concepts based on functionality: An association rule mining and decision tree approach, Expert Syst. Appl. 39 (10) (2012) 9534–9542.
[32] D. Cao, Z. Li, K. Ramani, Ontology-based customer preference modeling for concept generation, Adv. Eng. Inform. 25 (2) (2011) 162–176.
[33] J.L. Chen, C.-W. Liau, A simple life cycle assessment method for green product conceptual design. Environmentally Conscious Design and Inverse Manufacturing, Proceedings EcoDesign, Second International Symposium, 2001, pp. 775–780
[34] M.I. Mazhar, S. Kara, H. Kaebernick, Remaining life estimation of used components in customer products: life cycle data analysis by Weibull and artificial neural networks, J. Oper. Manag. 25 (6) (2007) 1184–1193.
[35] J.-H. Park, K.-K. Seo, A knowledge-based approximate life cycle assessment system for evaluating environmental impacts of product design alternatives in a collaborative design environment, Adv. Eng. Inform. 20 (2) (2003) 147–154.
[36] W. Yan, C.-H. Chen, Y. Huang, W. Mi, A data-mining approach for product conceptualization in a web-based architecture, Comput. Ind. 60 (1) (2009) 21–34.
[37] W. He, Improving user experience with case-based reasoning systems using text mining and Web 2.0, Expert Syst. Appl. 40 (2) (2013) 500–507.
[38] D.G. Rajpathak, An ontology based text mining system for knowledge discovery from the diagnosis data in the automotive domain, Comput. Ind. 64 (5) (2013) 565–580.

[39] S.-T. Li, F.-C. Tsai, A fuzzy conceptualization model for text mining with application in opinion polarity classification, Knowl.-Based Syst. 39 (2013) 23–33.
[40] Lai, H., Chang, Y., & Chang, H. (2005). A robust design approach for enhancing the feeling quality of a product: a car profile case study. International Journal of Industrial Ergonomics, 35, 445–460.
[41] Jordan, P.W., 1998. Human factors for pleasure in product use. Applied Ergonomics 29, 25-33.
[42] Chuang, M.C., Chang, C.C., Hsu, S.H., 2001. Perceptual factors underlying user preferences toward product form of mobile phones. International Journal of Industrial Ergonomics 27, 247-258.
[43] Crilly, N., Moultrie, J., Clarkson, J., 2004. Seeing things: consumer response to the visual domain in product design. Design Studies 25 (6), 547-577.
[44] Khalid, H., 2006. Embracing diversity in user needs for affective design. Applied Ergonomics 37 (4), 409-418.
[45] Chang, D., Chen, C.-H., and Lee, K.M. (2014). A crowdsourcing development approach based on a neuro-fuzzy network for creating innovative product concepts. Neurocomputing. 142: 60–72.
[46] N. Cross, (2000). Engineering design methods: Strategies for product design (3rd ed.). Chichester, UK: Wiley.
[47] K. Y. Chan , C. K. Kwong , T. S. Dillon & K. Y. Fung, 2011. An intelligent fuzzy regression approach for affective product design that captures nonlinearity and fuzziness, Journal of Engineering Design, 22(8): 523–542.
[48] J.J. Garrett, 2011, The Elements of User Experience: User-centered Design for the Web and Beyond (2rd Edition), New Riders, Pearson Education, USA
[49] M. Negnevitsky, 2011, Artificial Intelligence: A Guide to Intelligent Systems (3rd Edition), Pearson Education Limited, UK.
[50] M. Ehrig, P. Haase, M. Hefke, N. Stojanovic, Similarity for ontologies – a comprehensive framework, in: ECIS, 2005.
[51] W. Song, C.H. Li, S.C. Park, Genetic algorithm for text clustering using ontology and evaluating the validity of various semantic similarity measures, Expert Syst. Appl. 36 (5) (2009) 9095–9104.
[52] X. Wang, M. Hauswirth, T. Vitvar, M. Zaremba, Semantic web services selection improved by application ontology with multiple concept relations, in: Proceedings of the 23rd Annual ACM Symposium on Applied Computing, 2008, pp. 2237–2242.
[53] R. H., Hayes, Wheelwright, S. C., and Clark, K. B., 1988, Dynamic manufacturing. New York: Free Press.