A multi-objective PSO approach of mining association rules for affective design based on online customer reviews

Huimin Jiang[a,b], C. K. Kwong[b], W. Y. Park[b] and K. M. Yu[b]
[a]College of Management, Shenzhen University, Shenzhen, People's Republic of China
[b]Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China

Abstract

Affective design is an important aspect of new product development that can enhance customer satisfaction of new products. Previous studies generally conducted customer surveys based on questionnaires and interviews to collect customers' views and preferences of affective design of products. However, the process could be time-consuming and the survey data does not contain much sentiment expression. Presently, a large number of online customer reviews on products can be found on various websites that contain rich information of customer opinions and expectations. However, the generation of useful information based on online customer reviews for affective design has not been addressed in previous studies. In this paper, a methodology for generating association rules for supporting affective design based on online customer reviews is proposed which mainly involves opinion mining of affective dimensions from online customer reviews and association rule mining based on multi-objective particle swarm optimization (PSO). Opinion mining is adopted to analyze online reviews and conduct sentiment analysis for affective dimensions. Based on the mined information and morphological analysis of products, a multi-objective PSO approach is proposed to generate association rules that depict the relationships between affective dimensions and design attributes. A case study was conducted to illustrate the proposed methodology.

Keywords: Affective design; opinion mining; association rule mining; multi-objective PSO

1. Introduction

Affective design is a systematic approach to the analysis of customer reactions to candidate designs (Barnes and Lillford 2009) and has been shown to excite customers' psychological feelings and help improve customer satisfaction in terms of emotional aspects. Products with good affective design can attract customers and influence their choices and preferences (Noble and Kumar 2008). Previous studies commonly conducted customer surveys based on questionnaires and/or interviews to acquire customer affective responses towards products. Such obtained customer survey data can be used to investigate the effects of design attributes and the elements of affective design on customer satisfaction and then model the relationships between design attributes and elements and customer preference. However, there are some limitations in using survey data for affective design. First, conducting the surveys can be quite time-consuming and even expensive especially when interviews are involved; second, only a limited number of affective dimensions are used in surveys which are designed by product designers based on their own opinions and experience. It cannot fully cover all affective dimensions for product development; third, only questionnaire scores or interview answers are recorded from the surveys and the survey data does not contain any sentiment expression which can be easily found in online customer reviews. Hence, customer surveys may not provide complete information for affective design of a product. In recent years, online customer reviews on products have become extremely popular and a huge number of online customer reviews can be found on various e-commerce websites, social media and web-based discussion forums which contain rich information about customers' opinions. Online reviews are different from the survey data obtained from questionnaires or interviews. They are free texts written entirely based on the willingness of customers, out of their own interests, in their own language, and without any pre-defined questions to lead them. Online reviews can provide the opinions of customers towards certain products that cannot be determined in traditional surveys. Online reviews are high-volume and high-velocity information assets which can provide rich information about customers' opinions, views and sentiments on products. Many potential customers are influenced by online reviews in their purchases. Thus, online customer reviews can be a source of valuable information for performing affective design of new products. However, the generation of useful information for affective design based on online customer reviews thus far has not been addressed in previous studies.

In this paper, a methodology for generating associated rules for affective design based on online customer reviews is proposed which mainly involves opinion mining using online customer reviews and association rule mining based on a multi-objective particle swarm optimization (PSO) approach. In the proposed methodology, opinion mining is employed to analyze the online customer reviews, perform sentiment analysis and compute sentiment scores of the affective dimensions. Based on the results of the opinion mining and morphological analysis of products, data sets for association rule mining are generated. Then, a multi-objective PSO approach (Beiranvand et al. 2014) is proposed to mine the association rules for affective design which involves the objectives of maximizing confidence, coverage and interestingness of the rules mining.

The rest of this paper is organized as follow: Related works are presented in Section 2. Section 3 describes the proposed methodology which involves opining mining based on online reviews and multi-objective PSO in association rule mining. A case study on affective design of compact cars based on the proposed methodology is described in Section 4. Finally, the conclusions are given in Section 5.

2. Related works

In the following subsections, related studies on affective design, opinion mining for product design and association rule mining are briefly reviewed.

2.1. Affective design

Affective design attributes such as colors, forms and shapes of a product can evoke the affective responses of customers to products. Nagamachi (1995) proposed Kansei engineering which is a product development methodology that uses quantitative methods to acquire and transform customer affections into design attribute settings using quantitative methods. The framework of Kansei engineering encompasses four tasks (Barnes and Lillford 2009; Nagamachi 1995), namely, definition of the product domain, determination of the dimensions of customer affections, determination of design attributes and attribute options, and evaluation of relationships between customer affections and design attributes.

One important task of the Kansei engineering framework is to evaluate the relationships between defined affective dimensions and design attributes. Previous studies on Kansei engineering have applied various regression analyses. Multiple linear regression was adopted to model the relationships between usability and design elements (Han et al. 2000). The dimensions of usability developed were able to embrace the impression and performance aspects. You et al. (2006) developed customer satisfaction models for automotive interior material using quantification theory I. Based on the models, the significance of design attributes can be identified. Barone et al. (2007) proposed a weighted ordinal logistic regression model for affective design and they showed that logistic regression was more suitable to deal with ordinal affective data and categorical design attributes than statistical linear regression. Seva et al. (2007) proposed multilevel regression to model the relationships in Kansei engineering by which the intra-class correlation for each of the pre-purchase affect was computed and the clustering of affective responses was accounted for. This helps improve the interpretability of affective data. Nagamachi (2008) applied partial least squares analysis in Kansei statistical analysis which provided better evaluation of data projected by principal component analysis. However, the above statistical approaches are unable to address the fuzziness involved in the affective responses and assessments of customers. Various computational intelligence techniques have been attempted to model the relationships. Compared with statistical techniques, computational intelligence techniques are normally more capable of handling the ambiguity of affective data and in modeling the nonlinear relationships between customer affections and design attributes. In general, the models generated based on computational intelligence for affective design can be classified into two types, implicit and explicit models. Previous studies have attempted various computational intelligence techniques to develop implicit models (also named black-box models) for affective design. Lai et al. (2005a) adopted neural networks to model the nonlinear relationships between design attributes and customer affections. Their results showed that the neural network based models outperformed the grey prediction based models in terms of the root of mean square errors. Fuzzy neural networks have been introduced to establish the affective relationships (Tsai et al. 2006). In their study, the values of the root mean square errors based on their proposed approach were able to fall within the acceptable error tolerance. Lau et al. (2006) proposed a fuzzy expert system with gradient descent optimization to develop models which relate affective responses to design attributes in fashion product development and over eighty percent of the prediction accuracy based on their developed models could be achieved. Hsiao and Tsai（2005) proposed neural network-based fuzzy reasoning and genetic algorithm approaches to model the relationships between the design attributes of a new product and the customers' affective image. In their study, a number of diverse solutions could be generated and the most appropriate design solution was identified. Lin et al. (2007) presented a fuzzy logic approach for consumer-oriented product form design and showed that the models developed based on their proposed approach outperformed the neural network-based models in terms of the root of mean square errors. However, no explicit knowledge of the relationships for affective design can be obtained based on implicit models. In recent years, quite a few computational intelligence techniques have been proposed to develop explicit models for depicting the relationships between affective satisfaction and design attributes. Sekkeli et al. (2010) adopted Tanaka's fuzzy regression to develop classification models for customer satisfaction which can address both random and fuzzy types of uncertainties existing in data. Chan et al. (2011) proposed a genetic programming based fuzzy regression approach to develop nonlinear models for relating affective satisfaction and design attributes. In their study, it was shown that their proposed approach outperformed statistical regression, Tanaka's fuzzy regression and Peter's fuzzy regression in modeling affective design in terms of mean validation errors and variance of validation errors. Jiang et al. (2015) introduced rough set and PSO-based adaptive neural fuzzy inference system (ANFIS) to model the nonlinear relationships between affective satisfaction and design attributes. The models developed based on the proposed approach were showed better than those developed based on Tanaka's fuzzy regression, fuzzy least-squares regression and genetic programming based fuzzy regression in terms of mean validation errors and variance of validation errors.

Some previous studies attempted to generate if-then rules for affective design based on customer survey data in order to understand the interactions between customer affections and design attributes. Various approaches were proposed to generate the rules for affective design.

Park and Han (2004) proposed a fuzzy rule-based approach to modelling affective user satisfaction for office chair design. Their study showed that the fuzzy rule-based approach was able to reduce the number of attributes to be considered in affective design. Jiao et al. (2006) proposed associated rule mining to discover affective mapping patterns. In their study, a system was developed to perform the mapping from the dimensions of customer affections to design attributes, and generated association rules for affective design. A multi-objective genetic algorithm approach was proposed for affective product design to generate approximate rules (Fung et al. 2012). The proposed approach can deal with both categorical attributes and quantitative attributes, and can determine the interval of quantitative attributes. The discovered rule sets based on their proposed approach have good support and coverage rates. Owing to the ambiguity of customer affections, rough set theory has been adopted in generating rules for affective design. The basic concept of rough set theory is to formulate an approximation of a crisp set from vague or imprecise information. Various rough set-based rule mining approaches were developed to generate rules for affective design. Nishino et al. (2006) introduced the rough set-based rule mining method based on the variable precision Bayesian rough sets theory into Kansei engineering. Their study showed that the proposed method was very effective to extract design decision rules from human evaluation data. Zhou et al. (2009) employed a rough set-based *K*-optimal rule discovery method in mining rules that performed the mapping between design attributes and multiple dimensions of customer affections. The most important rules according to the rule importance measure were able to be extracted based on their proposed method. However, these approaches are limited to dealing with customer affections in a binary classification and are incapable of measuring the strength of customer affections. Zhai et al. (2009) employed a dominance-based rough set (DRS) for affective rule mining and an approximation to a set of ordinal data was generated. Their results showed that the proposed approach was more suitable for rule mining for affective design compared with the conventional rule-mining approaches. However, previous studies for affective design were conducted based on survey data and no studies have been found thus far about opinion mining from online reviews for affective design.

2.2. Opinion mining for product design

Opinion mining mainly involves two tasks: (1) identification of opinion bearing words/phrases from free texts and (2) determination of sentiment polarity: positive, negative and neutral (Rahmath and Ahmad 2014). Various approaches were attempted to extract customer needs and product attributes using opinion mining from online reviews. Lee (2009) proposed a supervised machine learning approach for sentential-level adaptive text extraction and mining. Their proposed method for analyzing product reviews was shown effective to enhance the traditional methods for assessing user needs of products. Chen et al. (2013) developed an ontology-learning customer needs representation system to extract customer needs. Their study showed that the ontology of customer needs derived from their proposed system contained more semantics than those obtained from the existing ontology learning systems. A combined case-based reasoning and opinion mining approach was proposed to identify latent customer needs (Zhou et al. 2015). Their proposed approach was shown to help improve the discovery process of latent customer needs and reduce designers' mental workload. Zimmermann et al. (2015) proposed a framework for the discovery and polarity monitoring of implicit product features deemed important in customer reviews on different products. Their study showed that the proposed framework performed better than the previous algorithms in terms of cluster quality and execution time. Zhang et al. (2016) proposed an opinion mining extraction algorithm to jointly discover feature and feature-of relations, opinion expressions and their polarities, as well as feature–opinion relations. In their study, the proposed algorithm was found effectively to identify the main elements simultaneously and outperform the baseline methods.

Some previous studies of opinion mining attempted to prioritize product features / customer needs based on various approaches. Jin et al. (2012) employed an ordinal classification approach to rank product features and the performance of the proposed approach in assorted evaluation criteria was found promising. Rai (2012) adopted an unsupervised text mining approach to extract product features from online reviews and identified the importance of product attributes based on three different metrics. The study indicated that the textual content in product reviews had a significant predictive power for elucidating customer preferences. Yang et al. (2016) proposed a combined approach of integrating local context information with global context information. They showed that their approach outperformed all the baseline methods for ranking features. Various approaches were also employed to make comparisons among products based on opinion mining. Xu et al. (2011) proposed a graphical model to extract and visualize comparative relations between products from customer reviews and showed that the proposed method could extract comparative relations more accurately than the benchmark methods. Jin et al. (2016) adopted a Bayesian method to compare products in feature level. Results of their study can help designers to make a reliable justification regarding which product is better than others. Zhang et al. (2010) developed a weighted and directed graph to perform the ranking of products. Their study was found useful for the customers who were interested in specific product features as their proposed method was able to summarize the opinions and experiences of thousands of customers. Some studies of mining opinions from social media for product design were attempted. Tuarob and Tucker (2013) analyzed the sentiment from tweets to predict product sales and product longevity, as well as extract customer sentiment polarities of product features. In their study, sentiment analysis based on tweets was found useful to help predict product sales for 3 months or above for well-known products. Tuarob and Tucker (2015a) proposed an automated approach to discover lead users and latent product features by mining social media networks by which some information about users and product features are able to extract that is found useful in new product development. Tuarob and Tucker (2015b) later proposed a data mining driven methodology for extracting notable product features and customers' opinions about product favorability from large scale social media data. Their study showed that incorporating suggested features into next generation products could result in favorable sentiment from social media users. Lim and Tucker (2016) proposed a Bayesian sampling algorithm to identify product-feature-related keywords from social media data. Their proposed method was shown to provide better F-measure scores than the baseline method and the expert-keyword-selection method.

There are some other related studies. Deck and Trusov (2010) proposed a framework based on natural language processing (NLP) to infer the relative effect of product features on the overall customer satisfaction using mined customer needs from online reviews. In their study, the results obtained based on their proposed framework were found comparable with those obtained from conjoint analysis techniques. Chung and Tseng (2012) developed a business intelligence system based on rough set theory, inductive rule learning, and information retrieval methods for exploring the relationship between online reviews and review ratings. The validation results of their study indicated that the system could achieve high accuracy and coverage related to rule quality, and generate the informative rules with high support and confidence values. Wang and Wang (2014) adopted sentiment analysis to detect product weaknesses based on online customer reviews and showed that their proposed approach outperformed the baseline methods in terms of accuracy. However, the generation of useful information for affective design based on online customer reviews was not addressed in previous studies.

2.3. Association rule mining

Association rule mining is a process of finding particular relations among the attributes/attribute values of a huge database. Some popular conventional algorithms for association rule mining, such as Apriori algorithm (Agrawal and Srikant 1994), the set-oriented algorithm (Houtsma and Swami 1995), and Pincer search (Lin and Kedem 2002), were developed based on the market basket database. To prepare the market basket database, every transaction in the original database is represented as a binary record. For a database with a number of attributes and each attribute containing a number of distinct values, storing of the binary records is one of the limitations of the conventional algorithms. In addition, two parameters, minimal support and confidence, are always required to be determined by decision-makers or by using a trial-and-error method.

In recent years, evolutionary algorithms have been applied to association rule mining, which consist of stochastic-based optimization approaches that provide an efficient method to explore a large search space and discover global optimal solutions (Dehuri et al. 2006). Various types of evolutionary algorithms have been adopted for association rule mining. Yan et al. (2009) proposed a genetic algorithm-based strategy to generate association rules without specifying actual minimum support. They showed that their proposed algorithm significantly reduced the computation costs and generated interesting association rules only. Binary ant colony algorithm was applied to extract classification rules from the trained neural networks (Ozbakir et al. 2009). The proposed algorithm was shown to have better performance than the C4.5, PART, DecisionTable, and NBTree algorithms in terms of classification accuracy and the number of rules. Sarath and Ravi (2013) developed a binary particle swarm optimization based association rule miner and the proposed algorithm was able to yield better rules than a priori algorithm and FP-growth algorithm. Such studies only considered a single-objective function in performing association rule mining, however, simultaneous consideration of several objectives such as predictive accuracy, comprehensibility, and interestingness (Dehuri et al. 2008) is required in association rule mining. Minaei-Bidgoli et al. (2013) adopted multi-objective genetic algorithms to extract association rules. In their study, the proposed method was shown to have better performance than genetic algorithms and rough particle swarm optimization (PSO) in mining association rules. Beiranvand et al. (2014) proposed a multi-objective PSO algorithm in which three objectives were considered simultaneously in generating association rules. Their study showed that multi-objective PSO had a stable behavior in various runs in mining association rules and was better than other methods, including multi-objective differential evolution algorithms, multi-objective genetic algorithms, rough PSO and genetic association rules in terms of confidence, coverage and the number of association rules. PSO has a high degree of stability and has been demonstrated to have fast convergence. It does not rely on the derivative nature of the objective function and can achieve global optimization by comparing objective function values time after time. In the proposed approach, the determination of the threshold values of confidence, coverage and interestingness is not required, which can help enhance the quality of association rule mining.

In previous studies, survey data obtained based on questionnaires and/or interviews many times can be used to generate association rules. However, in this study, the association rules are generated based on opinion mining from online customer reviews. As online reviews are ill-structured and mainly contain texts, the information extracted from the reviews cannot be used directly to generate associated rules. Thus, a method is proposed in the paper to generate proper data sets based on the mined information for association rule mining.

3. Proposed methodology

Two issues were noted from the review of related works. First, market surveys with using questionnaires and/or interviews were adopted in previous studies to collect affective responses of customers/consumers on products. Then, various techniques were employed to develop customer satisfaction models and determine the importance of design attributes/elements for affective design based on the survey data. However, conducting market surveys can be quite time-consuming and even expensive. In addition, customers' sentiment expressions of product affection are difficult to be collected through the surveys. In recent years, a huge number of online customer reviews about product affection can be found on various websites. The reviews contain rich information on customers' opinions and comments about affective design of products that could be highly valuable for designers to perform affective design. However, no studies were found thus far about the use of online reviews for supporting affective design. Second, from the review of related work, numerous studies have been conducted to extract and prioritize product attributes, determine the importance of product attributes, compare and rank products, predict product sales and identify lead users based on opinion mining from online reviews. However, mining association rules for product design based on the results of opinion mining from online reviews was not found thus far in previous studies. To address the above two issues, in this paper, a methodology of mining association rules for supporting affective design based on online customer reviews is proposed which mainly involves opinion mining from online customer reviews and association rule mining based on a multi-objective particle swarm optimization approach. Figure 1 outlines the proposed methodology which involves the following three main processes: morphological analysis based on selected products, opinion mining from online customer reviews, and association rule mining using a multi-objective PSO approach. Details of the latter two processes are described in the following sub-sections while the details of conducting a morphological analysis are provided in Section 4.

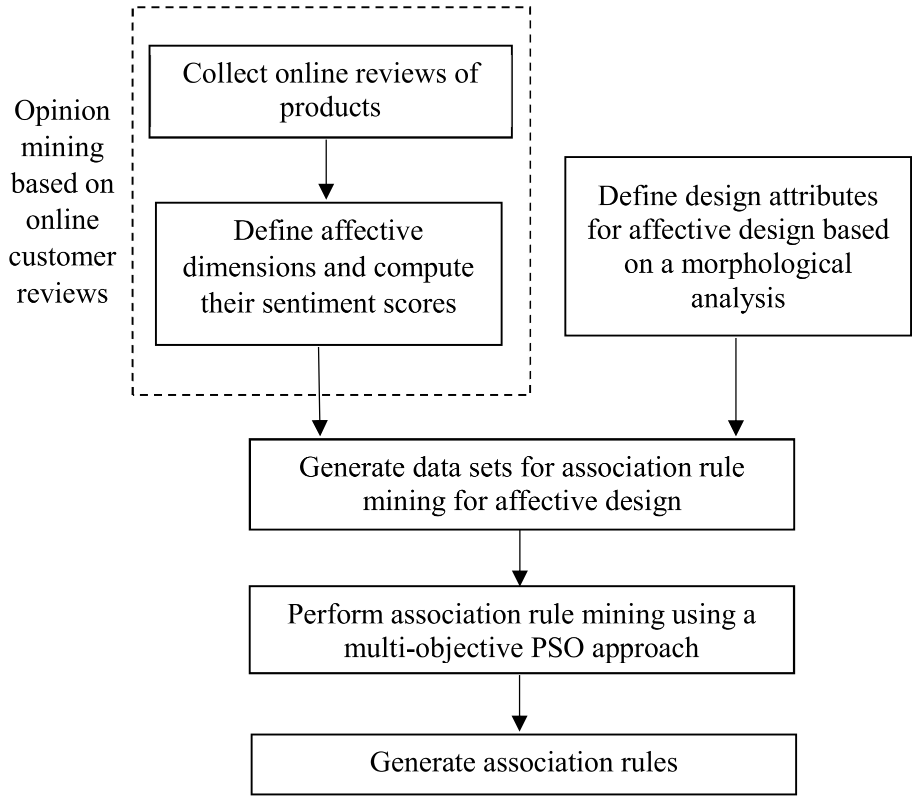


Figure 1. Proposed methodology.

3.1. Opinion mining from online customer reviews

Thousands upon thousands of online reviews for products, especially for popular products, can be found on websites (e.g., Ebay.com, Amazon.com, and Epinions.com). Online product reviews quite often contain three elements: perceived strengths and weaknesses of a product, ratings of the product, and customers' views and comments. First, products to be studied are identified. Their online customer reviews are collected using a web scraping software and stored in Excel files which are utilized as a source of conducting sentiment analysis of the affective dimensions. Opinion mining is then applied to mine affective dimensions from online reviews of the products, and the sentiment scores of individual affective dimensions are computed.

The opinion mining adopted in this study mainly involves five processes. First, the unstructured texts are pre-processed. Clean texts are produced through removing noise such as spaces, punctuations and stop words. For example, the words "a", "an", "is" and "you" are eliminated in the data pre-processing as they do not provide important information on the texts. The words expressing opinions are tagged as either nouns, adverbs, verbs or adjectives using Part-of-Speech (POS) tagging. Most POS taggers are statistical sequence labeling models and the tagging accuracy of the state-of-art POS taggers for English can achieve around 97% (Lu, 2014). Nouns can be treated as product features which include affective dimensions, whereas adjectives and adverbs are considered sentiment expressions of the nouns. Second, relevant sentiment expressions of product features are extracted. Third, incorrect features and redundant features are eliminated using feature pruning. Feature pruning is a process to refine product features by two types of pruning, compactness and redundancy pruning. Compactness pruning is used to check feature phrases and remove those candidate features whose words do not appear together. Redundancy pruning is used to remove redundant features that contain single words. Fourth, the phrases under different categories are grouped using a K-means clustering method. For example, the phrases "sharp picture," "good photos," and "blurred photos" are grouped under the category "photo quality," which is one of the product features of cameras. Finally, sentiment analysis is conducted and the semantic orientation and sentiment scores of opinion words for individual product features are determined using SentiWordNet (Baccianella et al. 2010).

A number of methods and tools are available to conduct a sentiment analysis, such as Python NLTK (Natural Language Toolkit), R (text mining module), RapidMiner, Semantria, Lingpipe, and LIWC 2007 (Linguistic Inquiry and Word Count). In this study, Semantria was chosen to conduct sentiment analysis because of its popularity as a well-known text analysis software tool. Semantria provides Excel add-in that enables the analysis of Excel spreadsheets according to positive, neutral, and negative sentiments. The Semantria Excel add-in conducts an automated sentiment analysis to extract sentiment from online reviews similar to human processing behavior, which contains the above five processes.

3.2. Rule evaluation measures

In this study, three rule evaluation measures are used to mine association rules including confidence, coverage and interestingness. The associated rules are in the form of "IF-THEN" statements and provide good semantic representation in which the meaning is explicated in a structure of semantic features. The conditions associated in the "IF" part is termed as antecedent and those in the "THEN" part is called the consequent, which are referred as A and C, respectively. The rule is represented as R: $A \rightarrow C$. The support count of an association rule is denoted by *SUP*(A→C), which is the number of transactions compatible with both A and C, namely, the number of transactions that contain $A \cup C$, *SUP*(A→C)=*SUP*($A \cup C$). Similarly, *SUP*(A) and *SUP*(C) are the number of transactions compatible with only A and C, respectively. The confidence or predictive accuracy of a rule measures specificity or consistency, which indicates the probability of creating the rule dependent on the antecedent part and is defined as follows:

$$Confidence(A \to C) = \frac{SUP(A \cup C)}{SUP(A)} \tag{1}$$

The coverage is to measure the extent to which the consequent part is covered by the rule and it shows the probability of creating the rule dependent on the consequent part. The coverage of a rule is expressed by (2).

$$Coverage(A \to C) = \frac{SUP(A \cup C)}{SUP(C)} \tag{2}$$

As a part of data mining, association rule mining is required to mine some hidden information and extract those rules that have comparatively less occurrence in the database. Extraction of such surprising rules may be more interesting to users and is difficult to quantify. Interestingness of a rule is to measure how much the rule is surprising for users and is defined as follows (Ghosh and Nath 2004):

$$Interestingness(A \to C) = \frac{SUP(A \cup C)}{SUP(A)} \times \frac{SUP(A \cup C)}{SUP(C)} \times (1 - \frac{SUP(A \cup C)}{|N|}) \tag{3}$$

where $|N|$ represents the total number of transactions in the database. $\frac{SUP(A \cup C)}{|N|}$ shows the probability of generating the rule according to the total records of the dataset and its complement, $1 - \frac{SUP(A \cup C)}{|N|}$, means the probability of not generating the rule.

3.3. Multi-objective PSO approach for association rules mining for affective design

The search algorithm of PSO is based on the social behaviour of a bird flock which works together to find the location of food sources (Kennedy and Eberhart 1995). Each bird in a flock tends to fly to a location using the knowledge of its own experience and the experience of the whole flock. Therefore, the entire flock can converge to the optimal location using swarm intelligence. By analogy, in PSO, every potential solution of the optimization problem can be imagined as being a point in a *D*-dimensional search space. This point is called a “particle”, which looks like a bird in a flock. Particles fly in search space with a certain speed, which is dynamically adjusted according to its own flight experience and its companions’ flying experience. Every particle has a fitness set determined by the values of the objective functions, and knows its current position and its own current best position which has the best fitness set, p_{best}. The p_{best} can be seen as the particle’s own flying experience. In addition, every particle also knows the global best position g_{best}, which has the best value in p_{best}. The g_{best} can be seen as its companions’ flying experience for the particle. Every particle uses the following information to change their current location: (a) the current location; (b) the current speed; (c) the distance between the current location and its own best location; and (d) the distance between the current location and the global best location. The optimization search is achieved by the iteration of the particle swarm which is formed by a group of random initialized particles.

A swarm is composed of m particles flying in the D-dimension at a certain speed. Every particle changes its position based on considering its own historical best position and other particles' historical best position. The position for the ith particle is $x_i = (x_{i1}, x_{i2}, ..., x_{id})$, where $1 \le i \le m$ and $1 \le d \le D$. D is the dimension of the search space which is the number of parameters to be defined in a rule. The speed for the ith particle is $v_i = (v_{i1}, v_{i2}, ..., v_{id})$. The historical best position of the ith particle, which has the best fitness set, is $p_i = (p_{i1}, p_{i2}, ..., p_{id})$. The best position for the whole swarm is $p_g = (p_{j1}, p_{j2}, ..., p_{jd})$, $j \in \{1, 2, \cdots, m\}$. The final result of p_g denotes the optimal parameters of a rule. The process of updating the speed and the position for the particle based on the idea of inertia weight (Shi and Eberhart 1998) is expressed as follows:

$$v_{id}^{k+1} = \omega v_{id}^{k} + c_1 r_1 (p_{id}^{k} - x_{id}^{k}) + c_2 r_2 (p_{jd}^{k} - x_{id}^{k}) \tag{4}$$

$$x_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1} \tag{5}$$

where v_{id}^{k} and x_{id}^{k} are the speed vector and the position vector of the ith particle at the kth iteration, respectively; k is the number of iterations; ω is the inertia weight, which decides the quantity inherited from the current speed of the particle. If it is chosen properly, the particle can have the balanced ability of exploitation and development. c_1 and c_2 are learning factors and are usually set as 2. The values of r_1 and r_2 are randomly chosen from the range $[0,1]$.

3.3.1. The particle representation

In the proposed methodology, one particle represents one candidate rule for which a suitable encoding scheme is required. In general, two approaches can be adopted to encode the particles in multi-objective PSO and both of them were inspired from chromosome representations in genetic algorithms. The first approach is the Pittsburgh approach in which each particle represents a rule set and it is more suitable for classification rule mining. The second approach is the Michigan approach where each particle represents a separate rule and the antecedent and consequent parts are encoded separately, which is more effective for other data mining tasks (Ghosh and Nath 2004; Beiranvand et al. 2014). In this study, the Michigan approach is applied to encode the particles which are initialized and updated during the evolutionary process of the PSO. In association rule mining for affective design, the antecedent part of a rule is the setting of design attributes which are categorical values and the consequent part is the sentiment score of the affective dimension which is a real number. For the antecedent part, each attribute is denoted by two bits in which the first and second bits indicate the state and the value of the corresponding attribute, respectively. If the value of the first bit is 1, the attribute is involved in the antecedent part of a rule; and if it is 0, the attribute does not appear in the rule. If the attribute is present in the rule, the corresponding categorical value of the second bit of each attribute is used to calculate the objectives values. For the consequent part, one bit is designed to denote a level of sentiment of the affective dimension. A ten-point scale from 1 to 10 is used, where 10 means highly positive sentiment and 1 means highly negative sentiment. The number of parameters of a particle is D =2*the number of design attributes + the number of affective dimensions. Table 1 illustrates the structure of a particle designed in this study in which cv_i is a categorical value of design attribute i, $1 \le i \le n$, and n is the number of design attributes.

Table 1. Structure of a particle.

Antecedent part	Consequent part

Design attribute 1		Design attribute 2		…		Design attribute n		Affective dimension
0 or 1	cv_1	0 or 1	cv_2	…	…	0 or 1	cv_n	A scale (1~10)

For example, if the number of design attributes is 4 and the expression of a particle is as follows:

$$x = (1, 3, 0, 2, 1, 5, 1, 1, 2)$$

The corresponding rule is "if x_1 =3, x_3 =5, and x_4 =1, then the level of sentiment of the affective dimension is 2". In this candidate rule, the design attribute x_2 is absent.

3.3.2. Pareto dominance

In the association rules mining, the three rule evaluation measures (1), (2) and (3) are adopted as the three objective functions, respectively, in multi-objective PSO. Based on the results of opinion mining and morphological analysis of the design attributes, data sets for association rule mining are obtained. Based on position vector of particles x_{id}^{k+1}, the values of three objectives (1), (2) and (3) are calculated and recorded as a fitness set for each particle. Pareto dominance is introduced to deal with the multi-objective problem in which solutions are compared with each other. A solution x_1 dominates another solution x_2, if (a) the solution x_1 is no worse than solution x_2 in all objectives and (b) the solution x_1 is strictly better than x_2 in at least one objective. A fitness set is denoted as F $=\{f_1, f_2, f_3\}$ and f_1, f_2, and f_3 are objective functions. In a maximization problem, x_1 dominates x_2 if and only if the following two conditions are satisfied:

$$f_i(x_1) \geq f_i(x_2), \text{ for all } i \in \{1,2,3\} \tag{6}$$

$$f_j(x_1) > f_j(x_2), \text{ for some } j \in \{1,2,3\} \tag{7}$$

In the case of a minimization problem, x_1 dominates x_2 if $f_i(x_1) \leq f_i(x_2)$ for all $f \in F$ and $f_j(x_1) < f_j(x_2)$ for at least one $f \in F$. When a solution is not dominated by any other solutions in the search space, it is called a Pareto optimal solution. In each iteration of the optimization process of multi-objective PSO, every particle compares its current position with its own best position based on Pareto dominance and updates its historical best position p_i. Then, the solutions in p_i are compared with each other using the above two conditions and the global best position p_g is selected.

3.4. Computational procedures of the proposed methodology for mining association rules based on online customer reviews

In the following, the computational procedures of the proposed methodology are as follows:

Step 1: Online reviews of defined products are collected and stored in an Excel file. Opinion mining is then conducted using Semantria as described in Section 3.1. Affective dimensions are defined, and the corresponding sentiment scores are obtained.

Step 2: Design attributes for affective design are identified and morphological analysis of design attributes of the defined products is conducted. Based on the results of the opinion mining and the morphological analysis, data sets for association rule mining are generated.

Step 3: Based on generated data sets and the descriptions in Sections 3.2 and 3.3, multi-objective PSO is used to perform association rule mining. The initialization for a particle swarm is first conducted. Following the structure description of the particles in Table 1, the speed and position of each particle are initialized randomly in the corresponding ranges.

Step 4: In the first iteration, the individual best position p_i and global best position p_g of particles are initialized. The initial position of each particle is used as the initial individual best position. Based on initial positions of the particles and the data sets, the values of three objective functions are calculated for each particle using (1), (2) and (3) as shown in Section 3.2, which are recorded as the initial individual best fitness set p_{best}. Based on the initial p_{best}, particles in p_i are compared with each other using Pareto dominance described in Section 3.3.2. The particle that satisfies (6) and (7) is the non-dominated solution and is selected as the initial best particle. Its position vector is defined as the initial global best position p_g, and its fitness set is defined as the initial global best fitness set g_{best}.

Step 5: The iteration is continued by $k+1 \rightarrow k$. In each iteration, the speed vector v_{id}^{k+1} and the position vector x_{id}^{k+1} for each particle are updated based on (4) and (5), respectively. If the value in v_{id}^{k+1} and x_{id}^{k+1} is beyond the corresponding search range, the value is adjusted as the bound of the range. Then, the fitness set of the ith particle in the $(k+1)th$ iteration, FS_i^{k+1}, is obtained based on the updated position of particles by calculating the values of three objective functions (1), (2) and (3). The FS_i^{k+1} is compared with p_{best} of the ith particle. Based on (6) and (7), if FS_i^{k+1} dominates p_{best}, the p_{best} of the ith particle is set as the value of FS_i^{k+1}, and the ith particle's individual optimal position are updated as $p_i = x_{id}^{k+1}$. The Pareto dominance is then conducted among p_{best} based on (6) and (7). The global best fitness set g_{best} is updated as the non-dominated solution in p_{best} and the number of the best particle is recorded. The global best position p_g is updated as the position of the selected best particle. The values of p_g and g_{best} in each iteration are saved in an external repository.

Step 6: The iteration stops when the pre-defined number of iterations is reached. The values of p_g in the external repository denote the mined optimal rules and g_{best} are the values of confidence, coverage and interestingness of the corresponding rules. Each p_g involves the design attributes appearing in the antecedent part and their optimal settings as well as the level of sentiment of the affective dimensions in the consequent part. Finally, the association rules are generated based on p_g.

4. Implementation

A case study on the affective design of compact cars based on online customer reviews is presented to illustrate the proposed methodology. In this study, six popular compact cars were selected and denoted as A~F, respectively. The online customer reviews of the selected cars available in different websites, including cars.com, thecarconnection.com, edmunds.com, whatcar.com, carbase.my, amazon.com, and carmax.com, were collected and stored in an Excel file. Sentiment analysis was conducted for all the online reviews using the Semantria Excel add-in. Key words were first extracted from online reviews and high frequent key words which are synonymous or related to the same product feature were grouped. For example, the extracted key words "exterior", "look", "hatchback", "wheel", "headlights", "shape", "door", "front", and "beautiful" were grouped as a category "exterior styling" which is one of the affective dimensions of compact cars. Totally, nine common product features were summarized which are comfort, exterior styling, interior design, features, performance, fuel economy, safety, quality and price. "Comfort", "exterior styling", and "interior design" are treated as affective dimensions. The user category analysis of the Semantria Excel was used and the key words related to each affective dimension were the settings of the function "user category" of the Semantria. Based on the settings, the sentiment analysis was repeated. Online reviews were classified into the related affective dimensions and the corresponding sentiment scores were computed. The online reviews belonging to the same category were then collected. In this study, the affective dimension, "exterior styling", is used to illustrate the proposed multi-objective PSO approach of association rule mining for affective design of compact cars and is denoted as y. A total of 1259 online reviews of the six compact cars were collected in which 1013 reviews contained customers' opinions and comments about exterior styling of the compact cars. The range of the computed sentiment scores of "exterior styling" of the six compact cars is [-1.26, 3]. Table 2 shows some contents of the online reviews about exterior styling and their corresponding sentiment polarity and scores.

Table 2. Online reviews related to exterior styling with sentiment polarity and scores.

Exterior styling		
Online reviews	Sentiment score	Sentiment polarity
.... The car exterior styling looks sharp, sporty yet functional. The rear trunk/hatch is very spacious...	0.7283	Positive
I purchased this car for my wife in Evergreen exterior with amber seats. It looks weird in pictures but really good in person. ……. Exterior is sleek and handsome. ……Don't need any fancy wheels to reduce my MPG. Car is relatively quiet.	0.539	Positive
Nice car. I like it. Economic car, saves gas, easy to maintain, LED headlights makes it perfect. Interior and exterior design is okay too	0.498	Neutral
……Initial impression is that the exterior is really catered to young adults. Not suitable for my age …… In addition, my experience with the after-sales is quite unpleasant. Long servicing hours, noise on front suspension ……	-0.6762	Negative
It is really good. It is affordable and a lot of room that five adults can sit comfortably. …and its exterior looks very sporty. It is very reliable and its worth of your money. Overall I would say that it is a very good and reliable car.	0.6655	Positive

Morphological analysis of affective design of the six compact cars was conducted and the results are shown in Table 3. Based on Table 3, the categorical design attributes of affective design of compact cars are represented by numerical data which are then used for association rule mining. The visual shape plays a critical role in affective design as it symbolizes the external appearance of a product (Yadav et al. 2017). This study mainly focuses on the effect of the shape design of compact cars regarding the affective dimension "exterior styling". In the morphological analysis, the shape of the compact cars was decomposed into six design attributes relating to "exterior styling", which are front design, rear design, front light design, rear light design, side front and window profile and they are denoted as x_1, x_2, x_3, x_4, x_5, and x_6, respectively. With reference to the previous studies (Ranscombe et al. 2012; Lai et al. 2005b; Yadav et al. 2017; Chang et al. 2007) and the solicited views of some car's owners, the six attributes were selected to be studied in this case study. Profiles of the six design attributes of the compact cars were captured and various profiles of individual attributes were summarized as alternatives. Table 3 shows the alternative profiles for each design attribute. The number of alternatives is 4 for x_1, x_2, x_3, x_5, and x_6 and is 5 for x_4. Using the 2nd row of the Table 3 as an example, the front design x_1 totally has 4 alternatives among the six compact cars. Based on the defined attribute of the six compact cars, the corresponding alternative number of each design attribute is set as the value of the design attribute. Thus, Table 4 can be generated which shows the settings of the six design attributes of the six compact cars.

Table 3. Morphological analysis on design attributes relating to the affective dimension “exterior styling” of compact cars.

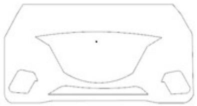
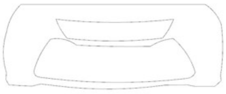
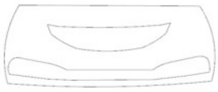
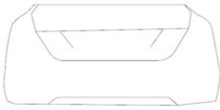
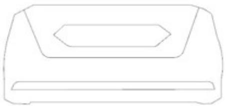
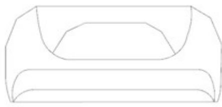
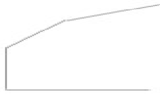
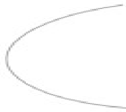
Design Attributes \ Alternative	1	2	3	4	5
Front design (x_1)					
Rear design (x_2)					
Front light design (x_3)					
Rear light design (x_4)					
Side front (x_5)					
Window profile (x_6)					

Table 4. Design attribute settings of six compact cars.

Compact cars	A	B	C	D	E	F
Front design (x_1)	1	3	2	1	4	4
Rear design (x_2)	4	2	4	3	1	1
Front light design (x_3)	4	1	2	4	3	4
Rear light design (x_4)	1	2	3	5	4	5
Side front (x_5)	3	2	4	1	1	2
Window profile(x_6)	4	3	2	1	2	3

Based on the data sets shown in Table 4 and the computed sentiment scores of the affective dimension “exterior styling” for individual compact cars, a multi-objective PSO was introduced to mine the association rules that relate sentiment of “exterior styling” and the design attributes x_1~ x_6. According to the designed structure of a particle shown in Table 1, the number of bits for the antecedent part of a rule in a particle was 12, calculated as 2* the number of design attributes=2*6=12. The number of bits for the consequent part was 1. Therefore, the number of dimensions of the search space for multi-objective PSO was 12+1=13. Thus, the number of parameters to be determined for generating a rule was 13. The search ranges of the first 12 bits of position of particles were [0,1], [1,4], [0,1], [1,4], [0,1], [1,4], [0,1], [1,5], [0,1], [1,4], [0,1], and [1,4], respectively. The range of the sentiment scores of “exterior styling” [-1.26, 3] was divided into ten equal length intervals, [-1.2626, -0.8364], [-0.8364, -0.4101], [-0.4101, 0.0162], [0.0162, 0.4424], [0.4424, 0.8687], [0.8687, 1.295], [1.295, 1.7212], [1.7212, 2.1475], [2.1475, 2.5737], and [2.5737, 3] and were mapped to the levels of sentiment, 1 to 10, respectively. For example, sentiment scores, which fall within the interval [0.4424, 0.8687], were mapped to the level 5. The range of the 13th bit, which is the scale of y, was [1, 10]. The size of the particle swarm was set as 50 and the iteration number was set as 100 which are determined through the repeated operations to make sure that the least number of the iteration and proper search range can be obtained. The inertia weight ω was set as a random value in the range of [0.1, 0.9] and both the learning factors c_1 and c_2 were set as 2. The r_1 and r_2 were random values chosen from the range $[0,1]$ which bring the stochastic state to multi-objective PSO algorithm. The proposed approach was implemented using Matlab software programming language to generate the association rules. Considering that a multi-objective PSO is a stochastic algorithm, different number of runs of the association rule mining based on multi-objective PSO were conducted in order to generate more and better diversity of the association rules. For each run, the generated rules were combined and the repeated rules were removed. Using 60 runs as an example, in all, 40 association rules were summarized. Table 5 shows the settings of the best particles and the corresponding values of confidence, coverage and interestingness. From the table, it can be seen that the values of confidence and coverage are all larger than 0.1, which was usually set as the threshold value of association rule mining in previous studies. The values of interestingness were calculated using (3) and can be used to quantify the interestingness of the generated rules.

Table 5. Settings of the best particles and the values of three objective functions based on 60 runs.

The Settings of the best particles														Confidence	Coverage	Interestingness
	x_1		x_2		x_3		x_4		x_5		x_6		y			
1	0	1	0	2	0	2	1	1	0	2	0	1	6	0.1706	0.2479	0.0411
2	0	4	0	2	0	4	1	4	1	1	0	2	6	0.1533	0.1795	0.0269
3	0	2	1	1	0	2	0	1	0	2	0	1	6	0.1506	0.3077	0.0447
4	0	4	1	3	0	1	0	5	0	2	0	3	3	0.2384	0.4091	0.0906
5	1	1	0	3	0	4	0	2	0	1	0	4	4	0.2373	0.4706	0.0993
6	0	4	0	4	1	3	0	4	1	1	0	1	6	0.1533	0.1795	0.0269
7	0	1	1	2	0	4	0	4	0	2	0	3	5	0.4364	0.2	0.0811
8	0	1	0	4	1	4	0	1	0	3	0	2	4	0.2265	0.5462	0.1078
9	1	3	1	2	0	1	1	2	0	4	0	2	5	0.4364	0.2	0.0811
10	0	1	0	1	1	4	0	3	1	1	0	3	2	0.1225	0.5606	0.0662
11	1	3	1	2	1	1	1	2	1	2	0	2	5	0.4364	0.2	0.0811
12	1	1	0	4	0	1	0	2	1	1	0	1	4	0.2483	0.3151	0.0725
13	0	1	0	1	0	1	0	2	1	2	1	3	5	0.4157	0.3083	0.1141
14	0	1	1	3	1	4	0	2	0	1	0	4	3	0.2384	0.4091	0.0906
15	0	4	0	4	0	1	1	5	0	2	0	1	3	0.2203	0.5057	0.1016
16	0	1	0	2	0	2	1	2	0	1	0	4	5	0.4364	0.2	0.0811
17	1	1	0	4	1	4	0	5	0	1	0	3	5	0.3242	0.425	0.117
18	0	2	0	4	0	3	0	5	0	3	1	3	5	0.4157	0.3083	0.1141
19	0	4	0	4	0	1	0	5	0	3	1	2	4	0.2409	0.2773	0.0624
20	0	2	0	1	0	2	1	5	1	1	0	2	3	0.2384	0.4091	0.0906
21	1	1	0	2	0	2	1	5	0	1	0	1	5	0.2947	0.2472	0.0665
22	1	3	1	2	0	2	0	5	1	2	0	3	5	0.4364	0.2	0.0811
23	1	1	1	3	0	4	0	2	1	1	0	1	4	0.2483	0.3151	0.0725
24	0	3	0	2	0	2	0	4	1	1	0	3	4	0.2437	0.4496	0.098
25	0	4	0	1	1	1	0	2	0	3	0	3	5	0.4364	0.2	0.0811
26	0	1	0	3	0	3	0	1	1	2	0	3	5	0.4157	0.3083	0.1141
27	0	2	1	3	0	4	0	1	1	1	0	1	5	0.2947	0.2472	0.0665
28	1	4	0	3	0	3	0	5	0	4	0	3	5	0.4017	0.2667	0.097

29	0	4	0	1	1	4	1	5	0	1	0	4	4	0.2302	0.3908	0.0817
30	0	2	0	2	1	4	1	5	1	1	0	4	3	0.2384	0.4091	0.0906
31	1	1	0	4	0	3	0	2	0	4	1	1	3	0.2384	0.4091	0.0906
32	1	1	1	3	0	4	0	4	0	3	0	3	5	0.2947	0.2472	0.0665
33	0	4	1	3	0	2	1	5	0	2	1	1	4	0.2483	0.3151	0.0725
34	0	3	1	2	1	1	0	1	0	4	0	1	5	0.4364	0.2	0.0811
35	0	3	1	3	0	4	0	3	0	1	1	1	5	0.2947	0.2472	0.0665
36	0	4	0	4	0	2	1	5	1	1	1	1	5	0.2947	0.2472	0.0665
37	1	1	0	2	0	1	1	1	0	4	1	4	5	0.3765	0.1778	0.0627
38	1	3	0	2	0	1	1	2	1	2	1	3	5	0.4364	0.2	0.0811
39	1	3	0	3	0	1	0	2	1	2	0	4	4	0.2545	0.1765	0.0431
40	1	3	1	2	0	1	1	2	1	2	1	3	5	0.4364	0.2	0.0811

Based on Table 5, the mined association rules are shown in Table 6.

Table 6. Mined association rules based on 60 runs.

Number	IF	THEN
Rule 1	$x_4=1$	$y = 6$
Rule 2	$x_4=4$, and $x_5=1$	$y = 6$
Rule 3	$x_2=1$	$y = 6$
Rule 4	$x_2=3$	$y = 3$
Rule 5	$x_1=1$	$y = 4$
Rule 6	$x_3=3$, and $x_5=1$	$y = 6$
Rule 7	$x_2=2$	$y = 5$
Rule 8	$x_3= 4$	$y = 4$
Rule 9	$x_1=3$, $x_2=2$, and $x_4=2$	$y = 5$
Rule 10	$x_3=4$, and $x_5=1$	$y = 2$
Rule 11	$x_1=3$, $x_2=2$, $x_3=1$, $x_4=2$, and $x_5=2$	$y = 5$
Rule 12	$x_1=1$, and $x_5=1$	$y = 4$
Rule 13	$x_5=2$, and $x_6=3$	$y = 5$
Rule 14	$x_2=3$, and $x_3=4$	$y = 3$
Rule 15	$x_4=5$	$y = 3$
Rule 16	$x_4=2$	$y = 5$
Rule 17	$x_1=1$, and $x_3=4$	$y = 5$
Rule 18	$x_6=3$	$y = 5$
Rule 19	$x_6=2$	$y = 4$
Rule 20	$x_4=5$, and $x_5=1$	$y = 3$
Rule 21	$x_1=1$, and $x_4=5$	$y = 5$
Rule 22	$x_1=3$, $x_2=2$, and $x_5=2$	$y = 5$
Rule 23	$x_1=1$, $x_2=3$, and $x_5=1$	$y = 4$
Rule 24	$x_5=1$	$y = 4$
Rule 25	$x_3=1$	$y = 5$
Rule 26	$x_5=2$	$y = 5$
Rule 27	$x_2=3$, and $x_5=1$	$y = 5$
Rule 28	$x_1=4$	$y = 5$
Rule 29	$x_3=4$, and $x_4=5$	$y = 4$
Rule 30	$x_3=4$, $x_4=5$, and $x_5=1$	$y = 3$
Rule 31	$x_1=1$, and $x_6=1$	$y = 3$
Rule 32	$x_1=1$, and $x_2=3$	$y = 5$
Rule 33	$x_2=3$, $x_4=5$, and $x_6=1$	$y = 4$
Rule 34	$x_2=2$, and $x_3=1$	$y = 5$
Rule 35	$x_2=3$, and $x_6=1$	$y = 5$

Rule 36	x_4=5, x_5=1, and x_6=1	y = 5
Rule 37	x_1=1, x_4=1, and x_6=4	y = 5
Rule 38	x_1=3, x_4=2, x_5=2, and x_6=3	y = 5
Rule 39	x_1=3, and x_5=2	y = 4
Rule 40	x_1=3, x_2=2, x_4=2, x_5=2, and x_6=3	y = 5

From the above rules, some useful information about the affective design of compact cars can be obtained. For example, referring to the Rules 1, 15 and 16, the satisfaction scores of alternatives 1, 2 and 5 of the rear light design (x_4) can be found as 6, 5 and 3 respectively. Considering Rules 24 and 6, it can be inferred based on human reasoning that the satisfaction score of alternative 3 of front light design (x_3) could not be smaller than 6.

Apart from the 60 runs, other numbers of runs of the multi-objective PSO were conducted including 10, 20, 40, 70, 80, 130, 160 and 190 runs in order to generate more association rules. Finally, all the generated association rules were compiled and repeated rules were removed. Based on the finalized rules, the sentiment levels of alternatives of individual design attributes were determined as shown in Table 7. For example, the sentiment level of the 2nd alternative of x_1 (front design) is 6. A higher sentiment level of an alternative of a design attribute indicates higher affective satisfaction of customers on the alternative. From the table, since the 2nd alternative of x_1, the 1st alternative of x_2, the 3rd alternative of x_3, the 1st alternative of x_4, the 4th alternative of x_5, and the 4th alternative of x_6 have the highest sentiment levels as highlighted in the table corresponding to their individual rows, the setting, the 2nd alternative of x_1, the 1st alternative of x_2, the 3rd alternative of x_3, the 1st alternative of x_4, the 4th alternative of x_5, and the 4th alternative of x_6, can be treated as the best one theoretically in terms of affective satisfaction of customers. Designers can make reference to the setting or even adopt it in performing their affective design of new compact cars by which the time for performing the affective design can be shorter and better affective design of compact cars can be created. On the other hand, the differences between the highest and the lowest sentiment levels for the design attributes x_2 to x_6 are 3 which indicates the choice of alternative profiles would largely affect the sentiment scores of the design attributes. For the design attribute x_1, the difference between the highest and the lowest sentiment levels is 2. Thus, the choice of alternative would affect the sentiment scores of design attribute x_1 significantly.

Table 7. The sentiment levels of alternatives of individual design attributes.

Design Attributes \ Sentiment Level \ Alternative Profile	1	2	3	4	5	Difference between the highest and the lowest sentiment levels
x_1	4	6	5	5		2
x_2	6	5	3	5		3
x_3	5	3	6	4		3
x_4	6	5	4	5	3	3
x_5	4	5	6	7		3
x_6	3	4	5	6		3

5. Conclusions

Questionnaires and interviews are commonly used in previous studies to understand affective responses of customers on products. However, it can be quite time-consuming and expensive to conduct the surveys and collect survey data. On the other hand, customers' sentiment expressions of product affection are difficult to be collected through such kind of surveys. Compared with the conventional survey data, online reviews contain rich information on customers' opinions and comments about the affective design of products that could be highly valuable for designers in developing new products. However, the generation of useful information based on online reviews for supporting the affective design of products has not been addressed in previous studies. In this paper, a methodology of mining association rules for supporting affective design based on online customer reviews is proposed which mainly involves opinion mining from online customer reviews and association rule mining based on a multi-objective particle swarm optimization approach. In opinion mining, the sentiment scores of affective dimensions are computed. Based on the results of the opinion mining and morphological analysis of sample products, data sets for association rules mining are generated. Then, a multi-objective particle swarm optimization approach is introduced to mine the association rules that relate the sentiment of affective dimensions and design attributes. Three rule evaluation measures, confidence, coverage and interestingness, are used in the association rules mining. A case study on the affective design of compact cars based on online customer reviews was conducted to illustrate the proposed methodology, and a number of association rules were obtained.

This paper provides a theoretical framework of mining association rules for affective design of products based on opinion mining from online reviews. For the opinion mining adopted in the proposed methodology, some aspects can be explored for improving its results such as the detection of fake reviews and the filtering of extreme reviews. In the proposed methodology, sentiment scores of categories are calculated by averaging the sentiment scores of categories of numerous online reviews and the average scores calculated can be affected by extremely high / low sentiment scores to a certain extent. To overcome this limitation, some techniques such as fuzzy averaging methods can be introduced to aggregate the sentiment scores. The multi-objective particle swarm optimization approach was shown to be effective to discover reliable and interesting rules in this study by performing a trade-off among confidence, coverage and interestingness. However, like most of the other meta-heuristic algorithms, some parameters of multi-objective PSO, such as inertia weight and learning factors, have to be set properly. Since we have implemented PSO for several times in our previous research, we did not take much time to determine a proper setting of the parameters for the multi-objective PSO. However, someone, who does not have experience of implementing PSO, may require a lot of time to determine a proper setting of parameters for PSO through a trial-and-error manner. Nevertheless, some statistically experimental techniques can be employed to help determine the setting. The numbers of runs of multi-objective PSO can affect the results of association rules mining. Thus, we suggest that different number of runs of multi-objective PSO for association rule mining should be conducted in order to obtain more and better diversity of the association rules. Implementation of the proposed methodology yields association rules generated for affective design. Based on the rules, designers are able to identify the alternatives of design attributes easily that appeal customers the most in terms of product affection. It is widely recognized that customer preferences on affective attributes of products could change largely over a short period of time. Thus, the proposed methodology can be implemented again in due course based on the latest online reviews such that updated information for affective design can be obtained. Since the proposed methodology has been implemented in software programs, we only need to change the time periods of crawling online reviews in the programs, execute them using the new time periods, and then obtain the updated association rules for affective design. On the other hand, decision support systems can be developed with using the association rules as well as inference and analysis techniques by which some more useful information of affective design, such as the significance of design attributes and effects of alternatives of design attributes on affective satisfaction, can be obtained.

Future work of this study would involve the development of a rule-based expert system using the generated association rules by which sentiment scores of the affective design of new products can be estimated by reasoning the association rules. On the other hand, to improve the authenticity of online reviews extracted for association rule mining, opinion spam detection can be conducted before performing sentiment analysis through style or content similarity and semantic inconsistency to detect and remove fake reviews.

Acknowledgement
The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. POLYU517113). The work was also partially supported by a grant from The Hong Kong Polytechnic University (Project No. G-YBMX)

References

Agrawal, R., and R. Srikant. 1994. "Fast Algorithms for Mining Association Rules in Large Databases." Proceedings of the 20th International Conference on Very Large Data Bases, Santiago, Chile, 487–499.

Baccianella, S., A. Esuli, and F. Sebastiani. 2010. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining." Proceedings of the 7th Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta, 2200-2204.

Barnes, C., and S. P. Lillford. 2009. "Decision support for the design of affective products." *Journal of Engineering Design* 20(5): 477-492.

Barone, S., A. Lombardo, and P. Tarantino. 2007. "A weighted logistic regression for conjoint analysis and Kansei engineering." *Quality and Reliability Engineering International* 23(6): 689-706.

Beiranvand, V., M. Mobasher-Kashani, and A. A. Bakar. 2014. "Multi-objective PSO algorithm for mining numerical association rules without a priori discretization." *Expert Systems with Applications* 41: 4259-4273.

Chan, K. Y., C. K. Kwong, T. S. Dillon, and 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.

Chang, H. C., H. H. Lai, and Y. M. Chang. 2007. "A measurement scale for evaluating the attractiveness of a passenger car form aimed at young consumers." *International Journal of Industrial Ergonomics* 37(1): 21-30.

Chen, X., C. H. Chen, K. F. Leong, and X. Jiang. 2013. "An ontology learning system for customer needs representation in product development." *The International Journal of Advanced Manufacturing Technology* 67(1-4): 441-453.

Chung, W., and T. L. Tseng. 2012. "Discovering business intelligence from online product reviews: A rule-induction framework." *Expert Systems with Applications* 1(1): 1-10.

Decker, R., and M. Trusov. 2010. "Estimating aggregate consumer preferences from online product reviews." *International Journal of Research in Marketing* 27(4): 293-307.

Dehuri, S., A. Jagadev, A. Ghosh, and R. Mall. 2006. “Multi-objective genetic algorithm for association rule mining using a homogeneous dedicated cluster of workstations.” *American Journal of Applied Sciences* 3(11): 2086–2095.

Dehuri, S., S. Patnaik, A. Ghosh, and R. Mall. 2008. “Application of elitist multiobjective genetic algorithm for classification rule generation.” *Applied Soft Computing* 8(1): 477–487.

Fung, K. Y., C. K., Kwong, K. W. M. Siu, and K. M. Yu. 2012. “A multi-objective genetic algorithm approach to rule mining for affective product design.” *Expert Systems with Applications* 39 (8): 7411-7419.

Ghosh, A., and B. Nath. 2004. “Multi-objective rule mining using genetic algorithms.” *Information Sciences* 163(1-3): 123-133.

Han, S. H., M. H. Yun, K. J. Kim, and J. Kwak. 2000. “Evaluation of product usability: development and validation of usability dimensions and design elements based on empirical models.” *International Journal of Industrial Ergonomics* 26(4): 477–488.

Houtsma, M., and A. Swami. 1995. “Set-oriented mining for association rules in relational databases.” Proceedings of the eleventh international conference on data engineering, Taipei, Taiwan, 25–33.

Hsiao, S. W., and H. C. Tsai. 2005. “Applying a hybrid approach based on fuzzy neural network and genetic algorithm to product from design.” *International Journal of Industrial Ergonomics* 35(5), 411-428.

Jiang, H., C. K. Kwong, K. W. M. Siu, and Y. Liu. 2015. “Rough Set and PSO-based ANFIS Approaches to Modeling Customer Satisfaction for Affective Product Design.” *Advanced Engineering Informatics* 29(3): 727-738.

Jiao, R. J., Y. Zhang, and M. Helander. 2006. “A Kansei mining system for affective design.” *Expert Systems with Applications* 30(4): 658-673.

Jin, J., P. Ji, and Y. Liu. 2012. “Product characteristic weighting for designer from online reviews: an ordinal classification approach.” EDBT-ICDT '12 Proceedings of the 2012 Joint EDBT/ICDT Workshops, Berlin, Germany, 33-40.

Jin, J., Y. Liu, P. Ji, and H. Liu. 2016. “Understanding big consumer opinion data for market-driven product design.” *International Journal of Production Research* 54(10): 3019-41.

Kennedy, J., and R. Eberhart. 1995. “Particle Swarm Optimization.” Proceedings of the 1995 IEEE International Conference on Neural Networks, Perth, Western Australia, 4: 1942-1948.

Lai, H. H., Y. C. Lin, and C. H. Yeh. 2005a. “Form design of product image using grey relational analysis and neural network models.” *Computers & Operations Research* 32(10): 2689-2711.

Lai, H. H., Y. M. Chang, and H. C. Chang. 2005b. “A robust design approach for enhancing the feeling quality of a product: A car profile case study.” *International Journal of Industrial Ergonomics* 35(5): 445-460.

Lau, T. W., P. C. L. Hui, F. S. F. Ng, and K. C. C. Chan. 2006. “A new fuzzy approach to improve fashion product development.” *Computers in Industry* 57(1): 82-92.

Lee, T. 2009. “Automatically Learning User Needs from Online Reviews for New Product Design.” America Conference on Information Systems, San Francisco, California, USA.

Lim, S., and C. S. Tucker. 2016. “A Bayesian sampling method for product feature extraction from large-scale textual data.” *Journal of Mechanical Design* 138(6): 061403.

Lin, D. I., and Z. M. Kedem. 2002. “Pincer-search: an efficient algorithm for discovering the maximal frequent set.” *IEEE Transactions on Knowledge and Data Engineering* 14(3): 553-566.

Lin, Y. C., H. H. Lai, and C. H. Yeh. 2007. “Consumer-oriented product form design based on fuzzy logic: A case study of mobile phones.” *International Journal of Industrial Ergonomics* 37(6): 531-543.

Lu, X. 2014. Computational methods for corpus annotation and analysis. Dordrecht: Springer.

Minaei-Bidgoli, B., R. Barmaki, and M. Nasiri. 2013. "Mining numerical association rules via multi-objective genetic algorithms." *Information Sciences* 233(1): 15-24.
Nagamachi, M. 1995. "Kansei Engineering: A new ergonomic consumer-oriented technology for product development." *International Journal of Industrial Ergonomics* 15(1): 3–11.
Nagamachi, M. 2008. "Perspectives and the new trend of Kansei/affective engineering." *The TQM Journal* 20(4): 290-298.
Nishino, T., M. Nagamachi, and H. Tanaka. 2006. "Variable precision bayesian rough set model and its application to kansei engineering." *Lecture Notes in Computer Science* 4100: 190-206.
Noble, C. H., and M. Kumar. 2008. "Using product design strategically to create deeper consumer connections." *Business Horizons* 51(5): 441-450.
Ozbakir, L., A. Baykasoglu, S. Kulluk, and H. Yapıcı. 2009. "TACO-miner: An ant colony based algorithm for rule extraction from trained neural networks." *Expert Systems with Applications* 36(10): 12295–12305.
Park, J., and S. H. Han. 2004. "A fuzzy rule-based approach to modeling affective user satisfaction towards office chair design." *International Journal of Industrial Ergonomics* 34 (1): 31–47.
Rahmath P. H., and T. Ahmad. 2014. "Fuzzy based sentiment analysis of online product reviews using machine learning techniques." *International Journal of Computer Applications* 99(17): 9-16.
Rai, R. 2012. "Identifying key product attributes and their importance levels from online customer reviews." 38th Design Automation Conference, Chicago, Illinois, USA, 533-540.
Ranscombe, C., B. Hicks, G. Mullineux, and B. Singh. 2012. "Visually decomposing vehicle images: Exploring the influence of different aesthetic features on consumer perception of brand." *Design Studies* 33 (4): 319-341.
Sarath, K. N. V. D, and V. Ravi. 2013. "Association rule mining using binary particle swarm optimization." *Engineering Applications of Artifricial Intelligence* 26(8): 1832-1840.
Sekkeli, G., G. Koksal, I. Batmaz, and O. T. Bayrak. 2010. "Classification models based on Tanaka's fuzzy linear regression approach: The case of customer satisfaction modeling." *Journal of Intelligent & Fuzzy systems* 21(5): 341-351.
Seva, R. R., H. B. L. Duh, and M. G. Helander. 2007. "The marketing implications of affective product design." *Applied Ergonomics* 38(6): 723-731.
Shi, Y., and R. Eberhart. 1998. "A modified particle swarm optimizer." The 1998 IEEE International Conference on Evolutionary Computation, Anchorage, Alaska, 69-73.
Tsai, H. C., S. W. Hsiao, and F. K. Hung. 2006. "An image evaluation approach for parameter-based product form and color design." *Computer-Aided Design* 38(2): 157-171.
Tuarob, S., and C. Tucker. 2013, "Fad or Here to Stay: Predicting Product Market Adoption and Longevity Using Large Scale, Social Media Data." 33rd Computers and Information in Engineering Conference, Portland, Oregon, USA.
Tuarob, S., and C. Tucker. 2015a. "Automated discovery of lead users and latent product features by mining large scale social media networks." *Journal of Mechanical Design* 137(7): 071402-1-11.
Tuarob, S., and C. S. Tucker. 2015b. "Quantifying product favorability and extracting notable product features using large scale social media data." *Journal of Computing and Information Science in Engineering* 15(3): 031003.
Wang, H., and W. Wang. 2014. "Product weakness finder: an opinion-aware system through sentiment analysis." *Industrial management & Data Systems* 114(8): 1301-1320.
Xu, K., S. S. Liao, J. Li, and Y. Song. 2011. "Mining comparative opinions from customer reviews for competitive intelligence." *Decision Support Systems* 50 (4), 743-754.

Yadav, H. C., R. Jain, A.R. Singh, and P.K. Mishra. 2017. "Kano integrated robust design approach for aesthetical product design: a case study of a car profile." *Journal of Intelligent Manufacturing* 28(7): 1709-1727.
Yan, X., C. Zhang, and S. Zhang. 2009. "Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support." *Expert Systems with Applications* 36(2): 3066–3076.
Yang, L., B. Liu, H. Lin, and Y. Lin. 2016. "Combining local and global Information for product feature extraction in opinion documents." *Information Processing Letters* 116(10): 623–627.
You, H., T. Ryu, K. Oh, M. H. Yun, and K. J. Kim. 2006. "Development of customer satisfaction models for automotive interior materials." *International Journal of Industrial Ergonomics* 36(4): 323-330.
Zhai, L. Y., L. P. Khoo, and Z. W. Zhong. 2009. "A dominance-based rough set approach to Kansei Engineering in product development." *Expert Systems with Applications* 36(1): 393-402.
Zhang, H., A. Sekhari, Y. Ouzrout, and A. Bouras. 2016. "Jointly identifying opinion mining elements and fuzzy measurement of opinion intensity to analyze product features." *Engineering Applications of Artificial Intelligence* 47: 122–139.
Zhang K., R. Narayanan, and A. Choudhary. 2010. "Voice of the customers: mining online customer reviews for product feature-based ranking." Proceedings of the 3rd Wonference on Online social networks, Boston, USA.
Zhou, F., R. J. Jiao, and J. S. Linsey. 2015. "Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews." *Journal of Mechanical Design* 137(7): 071401-1-11.
Zhou, F., R. J. Jiao, D. Schaefer, and S. Chen. 2009. "Rough set based rule mining for affective design." International Conference on Engineering Design, Stanford, CA, USA.
Zimmermann, M., E. Ntoutsi, and M. Spiliopoulou. 2015. "Discovering and monitoring product features and the opinions on them with OPINSTREAM." *Neurocomputing* 150 (Part A): 318–330.