A Support Vector Regression Based Prediction Model of Affective Responses for Product Form Design

Chih-Chieh Yang
Department of Multimedia and Entertainment Science, Southern Taiwan University, Tainan County, Taiwan 71005
Meng-Dar Shieh
Department of Industrial Design, National Cheng Kung University, No. 1, University Road, Tainan, Taiwan 70101

Abstract

In this paper, a state-of-the-art machine learning approach known as support vector regression (SVR) is introduced to develop a model that predicts consumers' affective responses (CARs) for product form design. First, pairwise adjectives were used to describe CARs toward product samples. Second, the product form features (PFFs) were examined systematically and then stored either as continuous or discrete attributes. The adjective evaluation data of consumers were gathered from questionnaires. Finally, prediction models based on different adjectives were constructed using SVR, which trained a series of PFFs and the average CAR rating of all the respondents. The real-coded genetic algorithm (RCGA) was used to determine the optimal training parameters of SVR. The predictive performance of the SVR with RCGA (SVR-RCGA) is compared to that of SVR with 5-fold cross-validation (SVR-5FCV) and back-propagation neural network (BPNN) with 5-fold cross-validation (BPNN-5FCV). The experimental results using the data sets of mobile phone and electronic scooter show that the performance of SVR is better than BPNN. Moreover, the RCGA for optimizing training parameters for SVR is more convenient than time-consuming CV for practical usage in product form design.

Keywords: Kansei engineering; Product form design; Support vector regression; Genetic algorithm; Neural network.

1. Introduction

The basic assumption for modeling consumers' affective responses (CARs) is that there exists a cause-and-effect relationship between CARs and the product form features (PFFs); that is, specific PFFs will produce different subjective feelings (Han

& Hong, 2003). Therefore, by analyzing the relationship between CARs and the PFFs in a systematic way, a prediction model can be constructed to facilitate product development. With the aid of the prediction model, an especially designed product form that targets specific consumer groups can be produced more objectively and efficiently instead of only relying on the designers' intuition and experience.

The crux to constructing such a prediction model is how to deal with the inter-attribute correlations that exist between product attributes and how to reconcile the nonlinear properties of these attributes (Shimizu & Jindo, 1995; Park & Han, 2004). There have been some attempts to define the relationship between the PFFs. The most noted research was by Kansei engineering (Nagamachi, 1995). The most adapted techniques in the product design field such as multiple regression analysis (Park & Han, 2004) and quantification theory type I (Jindo et al., 1995) depend heavily on an assumption of linearity and, therefore, cannot deal effectively with nonlinear relationships. In addition, prior to establishing a mathematical model, data simplification and variable screening is often needed to obtain better results (Han et al., 2000). Fuzzy regression analysis (Shimizu & Jindo, 1995) and other methods suffer from the same shortcomings (Park & Han, 2004).

To deal with the nonlinearity of many-to-many mapping between variables, neural network (NN) is a good candidate for building the prediction model. A few researches have illustrated the use of NN in the product design field. For example, Hsiao & Huang (2002) demonstrated the ability of NN to deal with nonlinear relationships between the PFFs. In later research by Hsiao & Tsai (2005) NN was used as part of a hybrid framework for a product form search. However, NN suffers from a number of shortcomings. NN is considered a "black-box" necessitating numerous control parameters and it is difficult to obtain a stable solution. Another drawback of NN, which is shared by all types of black-box models, is that the data of the resulting model and its parameters are difficult to interpret. In addition, NN follows the empirical risk minimization (ERM) approach, which is commonly employed by conventional machine learning methods. In the ERM approach, a measure of the prediction error, such as the root mean squared error (RMSE), pertaining to the training set outputs, is minimized. Since the ERM is based exclusively on the training set error, it does not guarantee that the resulting model will give a good generalization performance.

Vapnik (1995) developed a new kind of NN algorithm called support vector machine (SVM). SVM follows the principle of structural risk minimization (SRM), seeking to minimize an upper bound of the generalization error rather than minimize the training error (the principle followed by NN). SVM has been shown to provide better performance than traditional learning techniques (Burges, 1998). SVM's remarkable performance with respect to sparse and noisy data makes it a first choice in a number of real-world applications such as pattern recognition (Burges, 1998) and bioinformatics (Scholkopf et al., 2003). SVM is also known for its elegance in solving nonlinear problems with the "kernels" technique, which automatically carries out a nonlinear mapping to a feature space. With the introduction of an ε-insensitive loss function, SVM can be extended to solve function estimation problems. This is known as support vector regression (SVR). The properties of SRM equip the SVR model with a greater potential for generalizing the input-output relationship learnt during the training process. SVR has also been shown to exhibit excellent performance which benefits from their roots in SVM (Vapnik et al., 1997).

Despite being endowed with a number of attractive properties, SVR has yet to be applied widely in the field of product design. In this paper, SVR has been introduced for the purpose of developing a model that effectively predicts CARs. The remainder of the paper is organized as follows: Section 2 gives an introduction to SVR. Section 3 presents the proposed prediction model of CAR for product form design. Section 4 demonstrates the experimental results using mobile phone and electronic scooter as examples. Section 5 presents some brief conclusions. Finally, several suggestions for future research to extend this study are described in Section 6.

2. Theoretical backgrounds

2.1. Support vector regression

In this section, a description of the basic idea and formulation of SVR are reviewed. A data set, D of l training samples are given,

$$(x_1, y_1) \ldots (x_l, y_l) \tag{1}$$

where $x_i \in R^n$ is the input data, $y_i \in R$ is the desired output value. The objective of the SVR model is to identify the regression function, $y = f(x)$, which accurately predicts an output value that corresponds to a new set of data points. The SVR begins by introducing a loss function to minimize the regression risk. Various loss functions such as the linear ε-insensitive loss function, quadratic ε-insensitive loss function and Huber loss function can be used to construct different SVR models. In this study, standard SVR with a linear ε-insensitive loss function L is used:

$$L = (f(x) - y) = \begin{cases} |f(x) - y| - \varepsilon, & |f(x) - y| \geq \varepsilon, \\ 0, & otherwise \end{cases} \tag{2}$$

where ε is a precision parameter representing the radius of the tube located around the regression function, $f(x)$. Then, the SVR problem is regarded as the solution to

$$\text{minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{3}$$

$$\text{subject to } \begin{cases} y_i - w\cdot\phi(x_i) - b_i \le \varepsilon + \xi_i,\ \xi_i \ge 0 \\ w\cdot\phi(x_i) + b_i - y_i \le \varepsilon + \xi_i^*,\ \xi_i^* \ge 0 \\ i = 1,\dots,l \end{cases} \tag{4}$$

where $\phi(x)$ is the high dimensional feature space that is nonlinearly mapped from the input space x. The constant C is a regulation parameter. ξ_i is the upper training error (ξ_i^* is lower), subject to the ε-insensitive tube:

$$|y - (w\cdot\phi(x) + b)| \le \varepsilon \tag{5}$$

By introducing the Lagrange multipliers α_i^*, α_i and a kernel function K, the optimization problem in Eq. (3) can be rewritten as

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x, x_i) + b,\ 0 \le \alpha_i^*, \alpha_i \le C \tag{6}$$

In Eq. (6), α_i and α_i^* satisfy the equalities $\alpha_i\alpha_i^* = 0$, $\alpha_i \ge 0$ and $\alpha_i^* \ge 0$ where $i = 1, 2, \dots, l$ and are obtained by maximizing the dual function of Eq. (6), which has the following form:

$$\begin{aligned} \text{maximize } W(\alpha_i, \alpha_i^*) = &\sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) - \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) \\ &- \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) \end{aligned} \tag{7}$$

with constraints

$$\begin{aligned} &\sum_{i=1}^{n}(\alpha_i - \alpha_i^*), \\ &0 \le \alpha_i \le C\ , i = 1, 2, \dots, l, \\ &0 \le \alpha_i^* \le C\ , i = 1, 2, \dots, l. \end{aligned} \tag{8}$$

Finally, the regression function can be obtained as

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x_i, x) + b. \tag{9}$$

The training points with corresponding α_i and α_i^* equal to zero have no influence on the regression function solution. If these points are removed from the training set, the solution obtained would be still the same (Thissen et al., 2003). This characteristic is known as the “sparseness” of the solution and enables the SVR model to be defined as a combination of a relatively small number of input vectors.

The mapping ϕ is usually nonlinear and unknown. Instead of calculating ϕ, the kernel function K is used to compute the inner product of two vectors x_i and x_j in the feature space $\phi(x_i)$ and $\phi(x_j)$, that is, $K(x_i, x_j) = \phi(x_i)\cdot\phi(x_j)$. The elegance of using the kernel function is that one can deal with feature spaces of arbitrary dimensionality without having to compute the map $\phi(x)$ explicitly. Any function satisfying Mercer's condition can be used as the kernel function. The following are three commonly used kernel functions:

$$\text{linear: } K(x_i, x_j) = x_i \cdot x_j \,. \tag{10}$$

$$\text{polynomial: } K(x_i, x_j) = (1 + x_i \cdot x_j)^{\rho},\ \rho > 0 \,. \tag{11}$$

$$\text{radial basis function (RBF): } K(x_i, x_j) = \exp(-\left\| x_i - x_j \right\| / \sigma^2) \tag{12}$$

Here, ρ and σ are adjustable kernel parameters. The kernel parameter should be carefully chosen as it implicitly defines the structure of the high dimensional feature space $\phi(x)$ and thus controls the complexity of the final solution. In addition, the performance of the SVR model is heavily dependent on the regulation parameter C, the width of the tube ε and the parameter of the chosen kernel function. From the implementation point of view, training SVR is equivalent to solving a linearly constrained quadratic programming (QP) with the number of variables twice that of the input data dimension.

3. Prediction model of consumer's affective responses for product form design

This study aims to construct the prediction model of CARs for product form design. First, CARs are described using pairwise adjectives. Second, the PFFs displaying the properties of sparseness and mixed attribute types are used to represent the product samples. The evaluation data of consumers were gathered using questionnaires. Finally, SVR prediction models according to different adjectives are constructed using the PFFs of product samples and the collected consumer evaluation data. To obtain the best training model of SVR, the real-coded genetic algorithm (RCGA) is used to determine the optimal training parameters. For comparison purpose, back-propagation neural network (BPNN) is also used to construct the prediction model.

3.1. Describing the affective responses of consumer with pairwise adjectives

There are some differences in the affective responses of professional product designers and consumers (Hsu et al., 2000). Different kinds of adjectives must be used for them to describe their affective responses. This problem is often neglected in most researches. The framework proposed by Lamb & Kallal (1992) is adapted in this study to bring out two types of adjectives to describe affective responses: aesthetic adjectives for product designers and expressive adjectives for consumers

Aesthetic adjectives are more suitable for designers to describe their affective responses of the PFFs since the role of product designers is to manipulate form elements to make them aesthetically appealing. These kind of adjectives are closely related to aesthetic or style principles. Most aesthetic adjectives can be interpreted using certain aesthetic principles such as harmony, balance, uniformity, etc. These adjectives are often closely related to product attributes such as color, material and texture. For example, Chen (1997) has successfully combined these kind of adjectives (Table 1) with computerized PFFs.

Expressive adjectives are more suitable for consumer to express their sensations toward product samples. Due to the properties of high diversity and lack of easy interpretation, the expressive adjectives are more useful for the analysis of consumer preferences. These kind of adjectives can be found in the research of Hsu et al. (2000); Chuang & Ma (2001); and Hsiao & Tsai (2005) and examples of expressive adjectives are shown in Table 2. Notice that if consumers were asked to express their sensation using aesthetic adjectives, the answer would be ambiguous and not very satisfying. The expressive adjectives also provide us with an excellent tool for studying consumer preferences and help to determine new product positioning.

In the research of product form design, either single or pairwise adjectives can be used to describe CARs. In our experience, if consumers cannot distinguish the chosen adjectives very well, the accuracy of the prediction model will decrease dramatically. Since relationships such as relevancy, dependency, redundancy, cause/effect, and similarity often exist among adjectives (Han & Hong, 2003), pairwise adjectives are more suitable for describing CARs. Adjectives in each pair have similar concepts but are opposite to each other. For example, the adjectives “masculine-feminine” is based on the concept of gender and the two adjectives are polar opposites. This situation is very common when adjectives are used to describe CARs.

<Insert Table 1 about here>

<Insert Table 2 about here>

3.2. Representing sparse and mixed product form features

Two important properties are considered for PFF representation problems in this study. First, the form feature vector is often sparse. There often exist large amounts of features to represent a product form design, and each product sample does not necessarily occupy all the PFFs. The number of active or non-zero features in a feature vector is lower than the total number of features. This situation is very common in PFF representation (Kwahk & Han, 2002). Second, the PFFs can be either of the “continuous” or “discrete” type. Continuous attributes such as length and proportion often have some kind of scale or can be measured and the variable domain is continuous and without interruption. Discrete attributes denote categorical choices among a fixed number of variables, such as textures, the material used in the parts, etc.

SVR requires that each data sample be represented as a vector of real numbers. Continuous attributes can be processed without any problem. The discrete attributes, which are in fact nominal variables, were pre-processed using one-of-n encoding, instead of encoding as integer numbers. Taking a three-category attribute “circle, rectangle, triangle” for example, it can be represented as (0,0,1), (0,1,0), and (1,0,0). If the number of values in an attribute is not too great, this coding is more stable than using a single number to represent a discrete attribute (Hsu et al., 2003).

In addition, the scaling of attribute variables before applying SVR is very important. The main advantage is that it avoids attributes in greater numerical ranges dominating those in smaller numerical ranges. Another advantage is that it avoids numerical difficulties during calculation. Because kernel values usually depend on the inner products of feature vectors, e.g. linear kernel and polynomial kernel, large attribute values might cause numerical problems. Each attribute is normalized to zero mean and unit variance. In this study, 69 mobile phones and 137 electronic scooters were collected from the Taiwan market. Their PFFs were represented and processed by the previously mentioned method. For example, the complete list of 12 PFFs of mobile phone design is shown in Table 3. The dimension of the input features after pre-processing encoding equals to 31 (4 continuous variables $[X_1 \sim X_4]$ plus 29 dummy variables $[X_5 \sim X_{12}]$). Notice that the color and texture information on the product samples were ignored and were emphasized on the form features only.

<Insert Table 3 about here>

3.3. Questionnaire investigation for adjective evaluation

For adjective evaluation, twenty-two pairwise adjectives were adapted from the research of Hsu et al. (2000) as shown in Table 2. To keep things as simple as possible the number of adjective words used to conduct the experiment should not exceed seven. This should assure a reasonable consistency (Miller, 1956). The pairwise adjective “traditional-modern” is chosen for adjective evaluations.

To collect CARs' data for product form design, 30 subjects, 15 males and 15 females, were asked to evaluate product samples using a score from -1 to +1 in an interval of 0.1. Notice that if consumers were asked to evaluate the whole set of product samples at the same time, the experimental results would inevitably produce more errors. As a consequence, each respondent was asked to evaluate only one-third of the total samples. The presentation orders of the products were also randomized to avoid any systematic effects (Han & Hong, 2003). All of the subjects' evaluation scores for each product sample were averaged to get a final utility rating. To collect the evaluation data in a more effective way, a user-friendly questionnaire interface was designed, as shown in Fig. 1. The evaluation data of each subject could be recorded directly thus simplifying the post-processing procedure of the data.

<Insert Fig. 1 about here>

3.4. Constructing the support vector regression prediction model

SVR was used to construct a prediction model based on the PFFs of the collected product samples and the average CAR ratings obtained from all the respondents. Since SVR can only deal with one output value at a time, each prediction model need to be constructed according to different adjectives. The training scheme of a single prediction model based on SVR is depicted in Fig. 2. First, input product samples consisting of a series of the PFFs are fed into the training model. These input samples are mapped into feature space by the map function ϕ. Then, using the kernel function, dot products are computed with the images of the training samples under the map ϕ. The weights in the SVR represent the knowledge acquired from the product samples. Finally, the dot products are added up using the weights in terms of the Lagrange multipliers, α_i and α_i^*. This, plus the constant term b, yields the final predictive output value. The process described here is very similar to the regression in NN, only in the case of SVR the weights in the input layer are a subset of the training samples. The following procedure was used to train the product samples:

(1) Transform data to the format of SVR;
(2) Normalize the continuous attributes to a zero mean and unit variance and deal with the discrete attributes using one-of-n encoding;
(3) Choose a suitable kernel function;
(4) Use the RCGA to optimize the best parameter C, ε and the kernel parameter of a corresponding kernel function;
(5) Use the best parameter set to train the whole training set.

<Insert Fig. 2 about here>

3.5. Optimizing training parameters of SVR using real-coded genetic algorithm

Since the number of product samples is small, it is important to obtain best generalization performance and reduce the overfitting problem. The most frequently technique to obtain optimal training parameters is to conduct a grid search combined with *k*-fold CV (Hsu et al., 2003). Taking RBF kernel in Eq. (12) as example. If a grid search is taken using the following sets of values: $C=\left\{10^{-3},10^{-2},...,10^{5}\right\}$, $\varepsilon=\left\{10^{-5},10^{-4},...,10^{3}\right\}$, $\sigma^{2}=\left\{10^{-3},10^{-2},...,10^{5}\right\}$. Thus 8×8×8＝512 combinations need to be tried in the grid search. Although a denser gird may obtain better result, it also increases the loading of computation rapidly.

In this study, in order to overcome the shortcomings of grid search, the RCGA is adapted to optimize the parameters of SVR model. The rationale for applying genetic algorithm is to make use of its evolutionary computation and avoid exhaustive search on all possible combination of training parameters. Since the training parameters are numerical quantities in continuous domain, the real-coding, instead of traditional binary-coding, is more suitable for the chromosome representation. A fitness function for assessing the performance of each chromosome is assigned by its 5-fold CV error. The RMSE was used to evaluate the performance of each chromosome. The RMSE is defined as follows:

$$RMSE=\sqrt{\frac{1}{l}\sum_{i=1}^{l}(y_i-y_i')^2}$$

where y_i and y_i' denote the predicted result and the measured value respectively. Therefore, the objective of RCGA is to obtain the solution which minimizes the RMSE of 5-fold CV error.

The detailed setting of RCGA is described as follows. The population size is set to 10 times of the length of the chromosome. For example, RBF kernel have three parameters (C,ε,σ^2), thus the population size is set to 30. For most of the parameter optimization problem in this study, the number of generation is set to 50 which can achieve satisfactory results. The simulated binary crossover (SBX) (Deb & Agrawal, 1995) and the polynomial mutation (Deb & Goyal, 1996) are used as crossover and mutation operator. The SBX and polynomial mutation operators are calculated based on the polynomial probability distribution. An extra parameter called distribution index (DI) needs to be determined. Here, both DI values for the two operators are set to 1. Typically, a large crossover rate (>0.9) can achieve better results, while the mutation rate is set to less than 0.1. In order to maintain tendency toward global optimum, a large replacement rate (>0.9) is assigned to enforce a chromosome among the children to be replaced by a better chromosome among its parent generation. The optimal parameters obtained by RCGA are used to build the final prediction model using all product samples.

3.6. Constructing the back-propagation neural network prediction model

In this study, a three-layered fully connected NN is used for comparison purpose. Before training BPNN, the input form features were pre-processed and normalized as described in Sect. 3.2. Take the mobile phone dataset as example, the specific network structure of 31 (input nodes) × 32 (hidden nodes) × 1 (output) was selected for each CAR prediction model. The network is initialized with small random weights. The sigmoid activation function is used in each node, allowing the network to perform nonlinear mapping between the input form features and output CAR value. The back-propagation algorithm is used for the training of the network, which continuously comparing the network output with the expected value and adjusting the weights and biases of the network. In order to avoid the problem of overfitting and maximize generalization performance, the strategy of early-stopping with 5-fold cross-validation (CV) is adopted. The rationale for using CV to train NN is to stop training at the point when the error (RMSE) in CV starts to increase rapidly, because the best generalization performance had been reached. The RMSE of CV was checked once every one-hundredth of total training epochs. The training parameters such as learning rate, learning rate reduced factor, learning rate minimum bound, and number of epochs are considered accordingly. The optimal training parameters obtained by 5-fold CV are used to construct the BPNN prediction model using all product samples. In the present work, BPNN with 5-fold CV is conducted by using the CAFE software package.

4. Experimental results

4.1. Analysis of the optimization process using RCGA

In order to obtain best performance and reduce the overfitting of the training model, the RCGA was used to determine optimal parameters. Optimal parameters can be obtained by calculating the solution which minimizes the RMSE of 5-fold CV error.

The results of two different kernel functions including polynomial and RBF kernel are compared. For each kernel function, at least 10 different initial random populations are used to test the effect of the RCGA optimization process. With proper setting of RCGA described in Sect. 3.5, the obtained results of different initial populations are similar. Fig. 3 shows the optimization process of RCGA for polynomial kernel, taking the mobile phone design as example. The range of parameters C, ε and ρ in a real-coded chromosome are $\{10^{-3},10^{5}\}$, $\{10^{-5},10^{3}\}$ and $\{1,10\}$. In each generation, the maximized, median and minimized CV errors were plotted and marked in red plus, black dot, blue cross, respectively. In the first 10 generations, the minimized CV error reduces slowly from 0.068 to 0.063. In the 14th generation, the minimized CV error reduces rapidly to 0.057. After 16th generation, the minimized CV error converges to a relatively low value about 0.054. In the 50th generation, the minimized RMSE of CV error was 0.054 and the obtained optimal parameter set (C,ε,ρ) was $(2475.28, 0.142, 4.98)$. During the optimization process, the tendency of the median CV error indicates that the RCGA is capable to converge to a global optimum.

The optimization process of RCGA for RBF kernel is shown in Fig. 4. The range of parameters C, ε and σ^2 in a chromosome are $\{10^{-3},10^{5}\}$, $\{10^{-5},10^{3}\}$ and $\{10^{-3},10^{5}\}$. The minimized CV error in the initial generation was 0.068, which is similar to that of polynomial kernel. However, the minimized CV error reduces rapidly in the first 8 generations and reaches a much lower value within the range of 0.047 to 0.049 in the following generations. In the 50th generation, the minimized RMSE of CV error was 0.048, which is much smaller than that of polynomial kernel (0.054), and the optimal parameter set (C,ε,σ^2) was $(4309.78, 0.194, 416.85)$. The tendency of the median CV error indicates a faster convergence to a global optimum compared to polynomial kernel.

<Insert Fig. 3 about here>

<Insert Fig. 4 about here>

4.2. Comparison of predictive performance for different kernel functions

The optimal parameters of polynomial and RBF kernel obtained from the RCGA process were used to construct two prediction models of CAR for the adjective “traditional-modern”. The predictive performance of these two models for mobile phone design is compared. Fig. 5 shows the predictive performance of the model constructed with polynomial kernel using the optimized parameter set $(C,\varepsilon,\rho)=(2475.28, 0.142, 4.98)$. The RMSE of the training model using polynomial kernel is 0.181. It can be observed that the predictive adjective scores (drawn in red dash line) do not fit well to the original scores (drawn in blue solid line) of the 69 mobile phone samples. The predictive performance of the training model constructed with RBF kernel using the optimized parameter set $(C,\varepsilon,\sigma^2) = (4309.78, 0.194, 416.85)$ is shown in Fig. 6. The RMSE of the training model using RBF kernel is 0.078, which is better than that of the model using polynomial kernel. Since the predictive adjective scores fit well to the original adjective scores, this training model constructed using RBF kernel is more suitable, compared to the training model constructed with polynomial kernel, for the purpose to predict CARs.

<Insert Fig. 5 about here>

<Insert Fig. 6 about here>

4.3. Comparison of predictive performance for SVR-RCGA, SVR-5FCV and BPNN-5FCV models

The predictive performance of the SVR with RCGA (SVR-RCGA), SVR with 5-fold cross-validation (SVR-5FCV) and back-propagation neural network (BPNN) with 5-fold cross-validation (BPNN-5FCV) were compared using the data sets of mobile phone and electronic scooter evaluated using the pairwise adjective “traditional-modern”. The RBF kernel is used for the SVR-RCGA and SVR-5FCV models and the training parameters were optimized using the method described in Sect. 3.5. For BPNN-5FCV, the performance is closely related to learning rate. A practical guide for training the BPNN model is to start with large values of learning rate and learning rate reduced factor and observe the RMSE of CV as gradually reducing these two parameters. A trial-and-error procedure is still needed to obtain reliable performance without overfitting. From the results shown in Table 4, SVR-RCGA performs best with the lowest RMSE (0.078 for mobile phone and 0.072 for electronic scooter) compared to the other two models for both data sets. Comparing SVR-RCGA and SVR-5FCV, the performance of the former is only slightly better. However, the computation time needed for SVR-RCGA is much less than that of SVR-5FCV benefited from the heuristic searching capacity of genetic algorithm. In general, the performance of SVR is more superior to that of BPNN. Moreover, the trial-and-error procedure for tuning BPNN is cumbersome and heavily relies on the researchers’ experiences.

<Insert Table 4 about here>

5. Conclusions

In this study, SVR is used to develop the prediction model of CARs. The prediction model can be constructed using the PFFs as input data and the adjective evaluation score gathered from questionnaires as output value. The optimal training parameters of the prediction model were determined by the RCGA, which minimizes the RMSE of 5-fold CV error of the training data. The performance of SVR-RCGA using two different kernel functions, including polynomial and RBF kernel, were compared. Using the data set of mobile phone design, for RCGA constructed with polynomial kernel, the optimal parameter set (C, ε, ρ) was $(2475.28, 0.142, 4.98)$ and the RMSE of the model is 0.181. For SVR-RCGA constructed with RBF kernel, the optimal parameter set $(C, \varepsilon, \sigma^2)$ was $(4309.78, 0.194, 416.85)$ and the RMSE of the model is 0.078. Therefore, RBF kernel has better performance for SVR-RCGA. Also, according the experimental results using the data set of mobile phone and electronic scooter, the RCGA is capable of determining the optimal training parameters very effectively. Compared to optimize training parameters using a grid search combined with 5-fold CV (SVR-5FCV), the RCGA is more computationally efficient. The resulting model of SVR-RCGA has better predictive performance compared to that of SVR-5FCV and BPNN-5FCV. In a consequence, SVR-RCGA with RBF kernel is more suitable to predict CARs for product form design.

6. Future research

The case study in this study was based on the design of mobile phone and electronic scooter and used a relatively small amount of PFFs. The form features of other products such as consumer electronics, furniture, automobiles, etc., may have to consider different characteristics. A more comprehensive study of different products is needed to verify the effectiveness of the proposed method. Also, inducing more complex product attributes other than form features, such as color, texture and material information on the product samples, is the authors' main research direction. In addition, the collection of the CAR data is often accomplished by tedious questionnaire experiments. A web-based platform providing natural browsing environment which mimic e-commerce website is capable to automatically gather more evaluated responses for large amounts of product samples.

In this study, the predictive performance of SVR using two standard kernel functions such as polynomial kernel and RBF kernel were compared. In fact, selection of kernels is still an ongoing research branch in the area of SVR (Wang et al., 2003). Since every single kernel function has different properties and generalization performance, the advantages of different kernel functions can be combined by using their mixtures (Smits & Jordaan, 2002). In addition, constructing specially designed kernel function can incorporate prior knowledge (Scholkopf et al., 1998). For example, hierarchy of components in product form (Kwahk & Han, 2002) can be considered as local structure to emphasize local correlations (Hotta, 2004). These issues are very interesting when considering the application of product form design.

References

Burges, C. (1998). A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2).

Chen, K. (1997). Form language and style description. Design studies, 18, 249-274.

Chuang, M. C. & Ma, Y. C. (2001). Expressing the expected product images in product design of micro-electronic products. International Journal of Industrial Ergonomics, 27, 233-245.

Deb, K. & Agrawal, R. B. (1995). Simulated binary crossover for continuous search space. Complex Systems, 9, 115-148.

Deb, K. & Goyal, M. (1996). A combined genetic adaptive search (GeneAS) for engineering design. Computer Sciences and Informatics, 26(4), 30-45.

Han, S. H. & Hong, S. W. (2003). A systematic approach for coupling user satisfaction with product design. Ergonomics, 46(13/14), 1441-1461.

Han, S. H., Yun, M. H., Kim, K. J. & Kwahk, J. (2000). Evaluation of product usability: development and validation of usability dimensions and design elements based on empirical models. International Journal of Industrial Ergonomics, 26, 477-488.

Hotta, K. (2004). Support vector machine with local summation kernel for robust face recognition. Proceedings of the 17th International Conferences on Pattern Recognition.

Hsiao, S. W. & Huang, H. C. (2002). A neural network based approach for product form design. Design Studies, 23, 67-84.

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.

Hsu, C. W., Chang, C. C. & Lin, C. J. (2003). Technical Report, A practical guide to support vector classification, Department of Computer Science & Information Engineering, National Taiwan University, Taiwan.

Hsu, S. H., Chuang, M. C. & Chang, C. C. (2000). A semantic differential study of designers' and users' product form perception. International Journal of

Industrial Ergonomics, 25, 375-391.

Jindo, T., Hirasago, K. & Nagamachi, M. (1995). Development of a design support system for office chairs using 3-D graphics. International Journal of Industrial Ergonomics, 15, 49-62.

Kwahk, J. & Han, S. H. (2002). A methodology for evaluating the usability of audiovisual consumer electronic products. Applied Ergonomics, 33, 419-431.

Lamb, J. M. & Kallal, M. J. (1992). A conceptual framework for apparel design. Clothing & Textiles Research Journal, 10(2), 42-47.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review, 63, 81-97.

Nagamachi, M. (1995). Kansei Engineering: a new ergonomic consumer-oriented technology for product development. International Journal of Industrial Ergonomics, 15(1), 311-346.

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.

Scholkopf, B., Guyon, I. & Wetson, J. (2003). Statistical learning and kernel methods in bioinformatics. Artificial Intelligence and Heuristic Methods in Bioinformatics 183. Amsterdam, The Netherlands, IOS Press Amsterdam: 1-21.

Scholkopf, B., Simard, P., Smola, A. & Vapnik, V. (1998). Prior knowledge in support vector kernels. Advances in Neural Information Processing Systems. M. Jordan, M. Kearns and S. Solla. Cambridge, MIT Press. 10.

Shimizu, Y. & Jindo, T. (1995). A fuzzy logic analysis method for evaluating human sensitivities. International Journal of Industrial Ergonomics, 15, 39-47.

Smits, G. F. & Jordaan, E. M. (2002). Improved SVM regression using mixtures of kernels. Proceedings of IJCNN'02 on Neural Networks.

Thissen, U., R. Van Brakel, A. P. De Weijer, Melssen, W. J. & Buydens, L. M. C. (2003). Using support vector machines for time series prediction. Chemometrics and Intelligent Laboratory Systems, 1-2, 35-49.

Vapnik, V. N. (1995). The nature of statistical learning theory. New York, Springer.

Vapnik, V. N., Golowich, S. & Smola, A. (1997). Support vector method for function approximation, regression estimation and signal processing. Advance in neural information processing system 9. M. Mozer, M. Jordan and T. Petsche. Cambridge, MA, MIT Press: 281-287.

Wang, W., Xu, Z., Lu, W. & Zhang, X. (2003). Determination of the spread parameter in the Gaussian kernel for classification and regression. Neurocomputing, 55, 643-663.

Fig. 1. An interface of questionnaire for adjective evaluation.

Fig. 2. Training scheme of the SVR model.

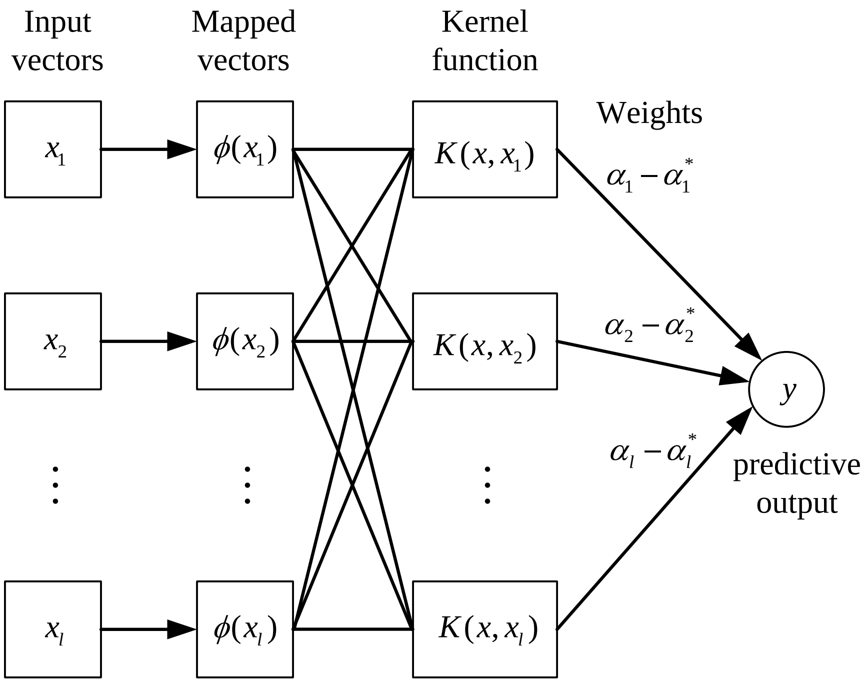


Fig. 3. The RCGA optimization process of polynomial kernel for mobile phone design .

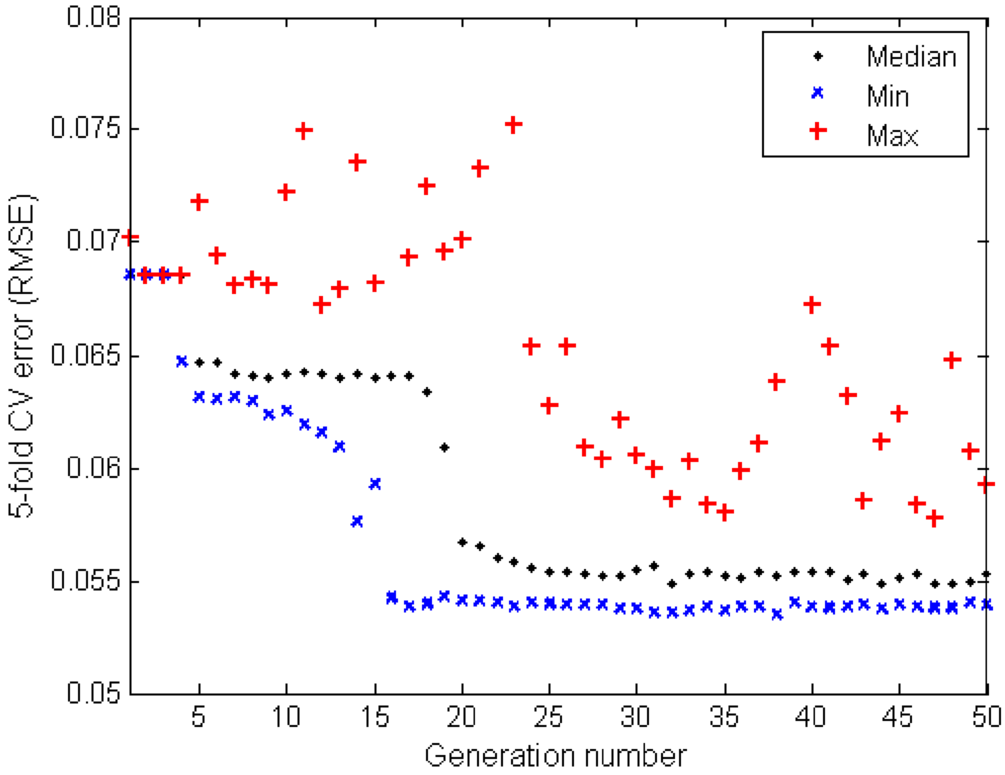


Fig. 4. The RCGA optimization process of RBF kernel for mobile phone design.

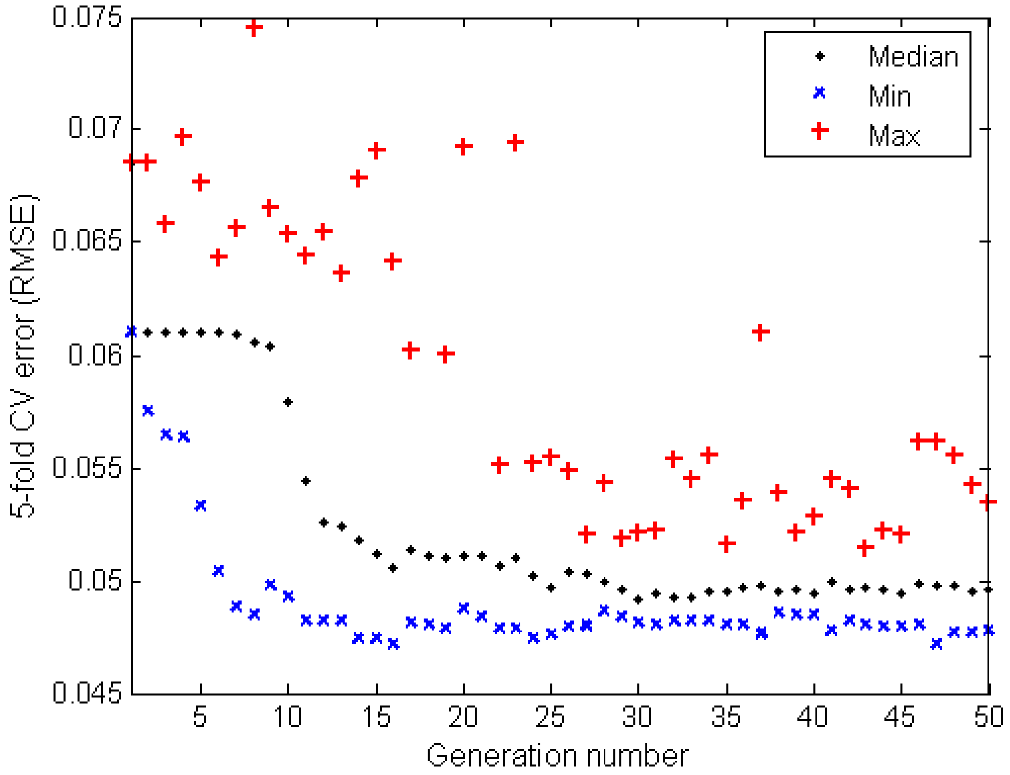


Fig. 5. Predictive performance of polynomial kernel for mobile phone design.

Fig. 6. Predictive performance of RBF kernel for mobile phone design.

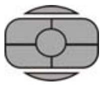
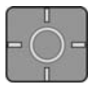
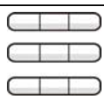
Table 1. Aesthetic adjectives adapted from Chen (1997).

01. harmonious-contrasting	02. homogeneous-heterogeneous	03. geometric-biomorphic	04. pure-impure
05. simple-complex	06. balanced-unstable	07. monollthic-fragmentary	08. static-dynamic
09. uniform-multiform	10. functional-decorative	11. subtle-bold	12. single-multiple

Table 2. Expressive adjectives adapted from Hsu et al. (2000).

01. traditional-modern	02. hard-soft	03. old-new	04. heavy-handy
05. obedient-rebellious	06. nostalgic-futuristic	07. coarse-delicate	08. masculine-feminine
09. rational-emotional	10. hand made-hi tech	11. childish-mature	12. unoriginal-creative
13. simple-complicated	14. conservative-avant grade	15. standard-outstanding	16. common-particular
17. plain-luxurious	18. decorative-practical	19. inert-active	20. personal-professional
21. obtuse-brilliant	22. discordant-harmonious		

Table 3. Complete list of twelve product form features for mobile phone design.

	Form features	Type	Attributes			
Body	Length (X_1)	Continuous	None			
	Width (X_2)	Continuous	None			
	Thickness (X_3)	Continuous	None			
	Volume (X_4)	Continuous	None			
	Type (X_5)	Discrete	 Block body (X_{51})	 Flip body (X_{52})	 Slide body (X_{53})	
Function button	Type (X_6)	Discrete	 Full-separated (X_{61})	 Partial-separated (X_{62})	 Regular-separated (X_{63})	
	Style (X_7)	Discrete	 Round (X_{71})	 Square (X_{72})	 Bar (X_{73})	
Number button	Shape (X_8)	Discrete	 Circular (X_{81})	 Regular (X_{82})	 Asymmetric (X_{83})	
	Arrangement (X_9)	Discrete	 Square (X_{91})	 Vertical (X_{92})	 Horizontal (X_{93})	

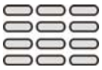
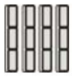
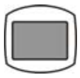
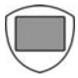
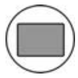
	Detail treatment (X_{10})	Discrete	 Regular seam (X_{101})	 Seamless (X_{102})	 Vertical seam (X_{103})	 Horizontal seam (X_{104})
Panel	Position (X_{11})	Discrete	 Middle (X_{111})	 Upper (X_{112})	 Lower (X_{113})	 Full (X_{114})
	Shape (X_{12})	Discrete	 Square (X_{121})	 Fillet (X_{122})	 Shield (X_{123})	 Round (X_{123})

Table 4. Performance comparison (RMSE) of SVR-RCGA, SVR-5FCV, and BPNN-5FCV.

	Mobile phone	Electronic scooter
SVR-RCGA	0.078	0.072
SVR-5FCV	0.086	0.081
BPNN-5FCV	0.137	0.113