Product Form Feature Selection for Mobile Phone Design Using LS-SVR and ARD

Chih-Chieh Yang[1], Meng-Dar Shieh[2], Kuang-Hsiung Chen[1], and Pei-Ju Lin[1]

[1] Department of Multimedia and Entertainment Science, Southern Taiwan University, 1, Nantai Street, Yongkang City, Tainan County 71005, Taiwan
[2] Department of Industrial Design, National Cheng Kung University, 1, University Road, Tainan 70101, Taiwan
scatjay@hotmail.com, mdshieh2010@gmail.com, {khchen,pjlin}@mail.stut.edu.tw

Abstract. In the product design field, it is important to pin point critical product form features (PFFs) that influence consumers' affective responses (CARs) of a product design. In this paper, an approach based on least squares support vector regression (LS-SVR) and automatic relevance determination (ARD) is proposed to streamline the task of product form feature selection (PFFS) according to the CAR data. The representation of PFFs is determined by morphological analysis and pairwise adjectives are used to express CARs. In order to gather the CAR data, an experiment of semantic differential (SD) evaluation on collected product samples was conducted. The LS-SVR prediction model can be constructed using the PFFs as input data and the evaluated SD scores as output value. The optimal parameters of the LS-SVR model are tuned by using Bayesian inference. Finally, an ARD selection process is used to analyze the relative relevance of PFFs to obtain feature ranking.

Keywords: Feature selection, Least squares support vector regression, Automatic relevance determination, Bayesian inference.

1 Introduction

The way a product looks is one of the most important factors affecting a consumer's purchasing decision. Traditionally, the success of a product's design depended on the designers' artistic sensibilities, which quite often did not meet with great acceptance in the marketplace. Many systematic product design studies have been carried out to get a better insight into consumers' subjective perceptions. The most notable research is Kansei engineering (KE) [1]. The basic assumption of KE studies is that there exists a cause-and-effect relationship between product form features (PFFs) and consumers' affective responses (CARs) [2]. Therefore, a prediction model can be constructed from collected product samples and the CAR data. Using the prediction model, the relationship between PFFs and CAR can be analyzed and a specially designed product form for specific target consumer groups can be produced more objectively and efficiently. The construction of the prediction model can be regarded as a function estimation problem (regression) taking PFFs of collected product samples as

C. Stephanidis (Ed.): Universal Access in HCI, Part III, HCII 2011, LNCS 6767, pp. 102–111, 2011.

input data and the CAR data as output values. Various methods can be used to construct the prediction model including multiple linear regressions (MLR) [3] quantification theory type I (QT1) [4], partial least squares regression (PLSR), neural networks (NN) [5] and support vector regression (SVR). Of these methods, SVR's remarkable performance makes it the first choice in a number of real-world applications [6] but it has not been adapted to the product design field according to our knowledge.

It is an important issue in KE [2], that the problem of product form feature selection (PFFS) according to consumers' perceptions has not been intensively investigated. The subjective perceptions of consumers are often influenced by a wide variety of form features. The number of form features could be many and might be highly correlated to each other. Therefore, manual inspection of the relative importance of PFFs and finding out the most critical features that please consumers is a difficult task. In the product design field, critical design features are often arrived at based on the opinions of experts or focus groups. However, the selection of features based on expert opinion often lacks objectivity. Only a few attempts have been made to overcome these shortcomings in the PFFS process. For example, [2] used several traditional statistical methods for screening critical design features including principal component regression (PCR), cluster analysis, and partial least squares (PLS). In the study of [7], a genetic algorithm-based PLS method is applied to screen design variables.

In our previous study [8], a PFFS method for SVM-RFE based on a multiclass classification model of product form design was proposed. Unlike the classification-based feature selection method, such as SVM-RFE, which is based on the consumer's judgment to discriminate the product form design from each other, the regression-based feature selection method, such as ARD used in this study, treats the PFFS problem in a different manner. This study proposes an approach based on least squares support vector regression (LS-SVR), as a variant of the SVR algorithm and ARD to deal with the PFFS problem. A similar backward elimination process of SVM-RFE is used to gradually screen out less important PFF based on the soft feature selection mechanism of ARD. PFFs used as input data and the CAR scores gathered from the questionnaire as output values are used to construct the LS-SVR prediction model. The optimal values of the parameters of the LS-SVR model are determined by Bayesian inference. ARD as a soft-embedded feature selection method is used to determine the relevance of PFFs based on the constructed LS-SVR model. The remainder of the paper is organized as follows: Section 2 presents the outline of the proposed PFFS method. Section 3 introduces the theoretical backgrounds of the LS-SVR algorithm and the ARD method for feature selection. A detailed implementation procedure of the proposed method is given in Section 4. Section 5 demonstrates the results of the proposed method using mobile phone designs as an example. Finally, Section 6 presents some conclusions and discussions.

2 Outline of the Proposed Method for Product form Feature Selection

The procedure of the proposed method for PFFS comprises the following steps:

1. Determine the product form representation using morphological analysis.

2. Conduct the questionnaire investigation for semantic differential (SD) evaluation to gather the CAR data on product samples.
3. Construct the LS-SVR prediction model based on PFFs and the pairwise adjectives.
4. Analyze the relative relevance of PFFs and obtain feature ranking using the ARD selection process.

3 Theoretical Backgrounds

3.1 Least Squares Support Vector Regression

This section briefly introduces the algorithm of LS-SVR. A training data set D of l data points are given, $(x_1, y_1)\ldots(x_l, y_l)$ (Eq. 1), where $x_i \in R^n$ is the input data, $y_i \in R$ is the desired output value. LS-SVR begins with approximating an unknown function in the primal weight space using the following equation $f(x) = w \cdot \phi(x) + b$ (Eq. 2), where $w \in R^k$ is the weight vector and $\phi(\cdot): R^n \to R^k$ is a nonlinear function that maps the input space into a higher dimension feature space. By introducing the quadratic loss function into LS-SVR, one can formulate the following optimization problem in the weight space

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

subject to the equality constraints

$$y_i = w \cdot \phi(x_i) + b + \xi_i,\ i = 1,\ldots,l \tag{4}$$

where $C > 0$ is the regularization parameter, b is a bias term, and ξ_i is the difference between the desired output and the actual output. When the dimension of w becomes infinite, one cannot solve the primal problem in Eq. (3). The dual problem can be derived using the Lagrange multipliers method. The mapping ϕ is usually nonlinear and unknown. Instead of calculating ϕ, the kernel function K is used to compute the inner product of the two vectors in the feature space and thus implicitly defines the mapping function $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ (Eq. 5). The following are three commonly used kernel functions [1]: linear: $x_i \cdot x_j$ (Eq. 6), polynomial: $(1 + x_i \cdot x_j)^p$ (Eq. 7), radial basis function (RBF): $\exp(-\frac{\|x_i - x_j\|^2}{\sigma^2})$ (Eq. 8), where the indices i and j correspond to different input vectors. p in Eq. (7) and σ in Eq. (8) are adjustable kernel parameters. In an LS-SVR algorithm, the kernel parameter and the regularization parameter C are the only two parameters that need to be tuned, which is less than that for the standard SVR algorithm. In the case of the linear kernel, C is the only parameter that needs to be tuned. The resulting LS-SVR prediction function is $f(x)=\sum_{i=1}^{l}\alpha_i K(x,x_i)+b$ (Eq. 9). For detailed deviations of the LS-SVR algorithm the authors refer to the textbook of [2].

3.2 Feature Selection Based on Automatic Relevance Determination

In this study, ARD is used to perform the soft feature selection, which is equivalent to determining the relevant dimensions of the n-dimensional input space using Bayesian inference. The simplest form of ARD can be carried out by introducing the standard Mahalanobis kernel [3] as $K(x_i,x_j)=\exp(-\sum_{d=1}^{n}\frac{\left\|x_i^d-x_j^d\right\|^2}{\sigma_d^2})$ (Eq. 10), where x_i^d denotes the dth element of input vector x_i, $d=1,...,n$. The Mahalanobis kernel can be regarded as a generalized RBF kernel in Eq. (8), which has a separate length scale parameter σ_d for each input space. The influence of the dth feature in the input space can be adjusted by setting larger or smaller values of σ_d. The Bayesian inference of ARD is then implemented by minimizing an objective function and the relevance of each input can be determined automatically according to the optimized σ_d values. However, the objective function often converges very slowly and results in very large values of optimized σ_d when using the standard Mahalanobis kernel. As a consequence, an augmented Mahalanobis kernel with an additional two parameters is adapted as $K(x_i,x_j)=\kappa_a\exp(-\sum_{d=1}^{n}\frac{\left\|x_i^d-x_j^d\right\|^2}{\sigma_d^2})+\kappa_b$ (Eq. 11) [4], where $\kappa_a>0$ denotes the amplitude parameter and $\kappa_b>0$ denotes the offset parameter. The parameters $\{\kappa_a,\sigma_1,...,\sigma_n,\kappa_b\}$ can be determined simultaneously during the optimization process. The ARD process for PFFS combined with Bayesian inference for parameter tuning is described in Section 4.5. For detailed derivations of Bayesian inference for ARD, the authors recommend the textbook of [2].

4 Implementation Procedures

This study implements a method for PFFS by considering the case study of mobile phone design. The detailed procedures are described in the following paragraphs.

4.1 Determination of Product Form Representation

In this aspect of form representation for product design, this study adopted morphological analysis [5]. In fact, morphological analysis in KE studies is the most widely used technique in KE due to its simple and intuitive way to define PFFs. The mixture of continuous and discrete attributes is also allowed. Practically, the product is decomposed into several main components and every possible attribute for each component is examined. For this study, a mobile phone was decomposed into body, function button, number button and panel. Continuous attributes such as length and volume are recorded directly. Discrete attributes such as type of body, style of button etc. were represented as categorical choices. Twelve form features of the mobile phone designs, including four continuous attributes and eight discrete attributes, were used. The list of all PFFs is (X1) length of body, (X2) width of body, (X3) thickness of body, (X4) volume of body, (X5) type of body, (X6) type of function button, (X7) style of function button, (X8) shape of number button, (X9) arrangement of number button, (X10) detail treatment of number button, (X11) position of panel, (X12) shape of panel. Notice that the color and texture information of the product samples were ignored and emphasis was placed on the form features only. The images of product samples used in the experiments described in Section 4.2 are converted to gray image using image-processing software.

4.2 Experiment and Questionnaire Design

A total of 69 mobile phones of different design were collected from the Taiwan marketplace. Three pairwise adjectives including traditional-modern, rational-emotional and heavy-handy were adopted for SD evaluations [6]. In order to collect the CAR data for mobile phone design, 30 subjects (15 of each sex) were asked to evaluate each product sample using a score from -1 to +1 in an interval of 0.1. In this manner, consumers can express their affective response toward every product sample by choosing one of the pairwise adjectives.

4.3 Construction of the LS-SVR Prediction Model

LS-SVR was used to construct the prediction model based on the collected product samples. The form features of these product samples were treated as input data while the average utility scores obtained from all the consumers were used as output values. Since LS-SVR can only deal with one output value at a time, three prediction models were constructed according to the selected adjectives. Since the relationship between the input PFFs and the output CAR score is often nonlinear, the frequently used methods such as MLR [7] and QT1 [8], which dependent on the assumption of linearity, can not deal with the nonlinear relationship effectively. SVR has proved to be very powerful when dealing with nonlinear problems. Compared to the standard SVR algorithm, LS-SVR has higher computing efficiency and fewer parameters that need to be determined.

The performance of the LS-SVR model is heavily dependent on the regularization parameter C and the kernel parameters. In order to balance the trade off between improving training accuracy and prevent the problem of overfitting, conducting a grid search with cross validation (CV) is the most frequently adopted strategy in the literature. In this study, tuning the parameters of the LS-SVR model using Bayesian inference can avoid the time-consuming grid search with CV.

4.4 The ARD Process for PFFS

By using the ARD technique, the relevance of PFFs can be determined in a ‘soft’ way by introducing the augmented Mahalanobis kernel in Eq. (11). The selection mechanism of features is embedded in the Bayesian inference. Small values of σ_d will indicate the high relevance of the dth input; while a large value will indicate that this input is less relevant with the other inputs. As a consequence, a reverse elimination process similar to that of SVM-RFE can be used to prune off the least relevant features with the largest value of σ_d. This process is repeated until all features are removed. The complete feature selection procedure based on LS-SVR and ARD is described as follows:

1. Preprocessing the form features;
2. Optimize the regulating parameter C and the parameter σ of the RBF kernel using Bayesian inference;
3. Use the optimized value of σ obtained in Step (3) as the initial values $\sigma_1 = \sigma_2 = \ldots = \sigma_n$ of the augmented Mahalanobis kernel with initial values of $\kappa_a = 1$ and $\kappa_b = 0$;
4. Start with an empty ranked feature list $R = []$ and the selected feature list $F = [1,...,d]$ with all features;
5. Repeat until all features are ranked:
   (a) Find the feature e with largest value of σ_d using feature list F;
   (b) Store the optimized parameter σ_d for the remaining features;
   (c) Remove the feature e from the feature list F;
   (d) Inset the feature e into feature list R: $R = [e, R]$;
   (e) Re-calculate the parameter C for the training data using the feature list F;
   (f) Re-calculate the initial value of the parameter σ for the training data using the feature list F.
6. Output: Ranked feature list R.

The overall ranking of PFFs can be determined by the resulting feature list R. The relative relevance of each feature in each elimination step can be analyzed by examining the optimized values of σ_d. Note that in each step of the selection procedure, the regulating parameter C and the initial value of parameter σ need to be re-calculated after eliminating the least important feature in order to obtain optimal prediction performance.

5 Experimental Results and Analyses

5.1 Effects of Optimization Algorithms for ARD

Since the Bayesian inference of ARD involves minimizing an objective function, the choice of the optimization algorithm has a great influence on the obtained results. In this study, two algorithms: the steepest descent method and the BFGS method were investigated. In order to examine the effects of the algorithms, the parameters $\kappa_a = 1$ and $\kappa_b = 0$ were fixed to the standard Mahalanobis kernel in Eq. (10). The pairwise adjectives "traditional-modern" were taken as an example and the results of the first

ARD step with the original twelve input features are shown in Figure 1. Notice that the value of the objective function declines quickly in the first few iterations in both algorithms. However, the BFGS method takes fewer (about 1/10) iterations than that of the steepest descent to reach minimization. In addition, the optimized σ_d values obtained by the BFGS method were much larger (about 100 times) than that of the steepest descent. These observations indicate that the BFGS method converges more dramatically than that of the steepest descent. More specifically, the values of σ_d in steepest descent change very steadily after the 100th iterations. The values of σ_d in the BFGS method increase rapidly after only 10 iterations (35th~45th) and the maximum σ_d (least important feature, X11) was about 250 times larger than the minimum σ_d (most important feature, X3). As for steepest descent, the maximum σ_d (X11) was only 25 times larger than the minimum σ_d (X3).

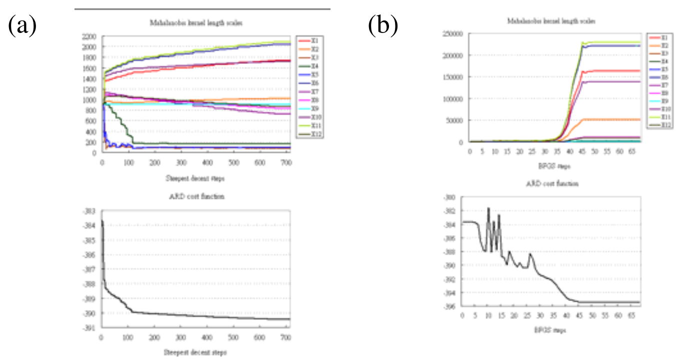

Fig. 1. Step 1 of Bayesian inference on length scales (traditional-modern) using (a) steepest descent method and (b) BFGS method with fixed $\kappa_a = 1$ and $\kappa_b = 0$

After examining the previous analysis using standard a Mahalanobis kernel, the steepest descent method was favored to obtain the feature ranking due to its steady convergence properties. However, it may suffer from not really reaching the minimum of the objective function compared to the BFGS method due to its slow convergence. This problem can be easily overcome by introducing the augmented Mahalanobis kernel with additional parameters in Eq. (11) to accelerate the convergence while still maintaining a steady variation of σ_d. Another issue is when to stop the iteration process in each elimination step. Since most applications of ARD emphasize improving the generalization performance of the prediction model, the iteration of ARD is often set to stop when the training error, e.g. root mean squared error (RMSE), becomes smaller than a pre-defined threshold. However, in the proposed method, the initial values of σ_d were first optimized using the RBF kernel and the initial state of the LS-SVR model had already achieved a relatively low training error. As a consequence, the training error of the LS-SVR did not improve significantly during the ARD iteration process. In this study, another stopping criterion was used.

Concerned that the difference between the form features perceived by the consumers should be kept to a reasonable proportion, the stopping criterion of each elimination is triggered when the proportion of the maximum σ_d to the minimum σ_d exceeds to a pre-defined threshold. The purpose of this study is to analyze the relative relevance of the form features as well as to obtain the feature ranking; how to provide a meaningful comparison between the features is our main concern. In this manner, when the relevance of one feature is small enough, that is, the σ_d of this feature is larger than the other features and thus can be eliminated. In each elimination step, the feature with the largest value of σ_d will be pruned off and the feature ranking obtained.

5.2 Predictive Performance of the LS-SVR Model with ARD

In order to examine the effectiveness of Bayesian inference applied on the LS-SVR model with ARD, the predictive performance is compared with a typical LS-SVR model (RBF kernel) using a grid search with LOOCV and an MLR model. The predictive performance was measured by using RMSE. A grid search with LOOCV is taken using the following sets of values: $C = \{10^{-3}, 10^{-2.9}, ..., 10^{2.9}, 10^{3}\}$ and $\sigma^2 = \{10^{-3}, 10^{-2.9}, ..., 10^{2.9}, 10^{3}\}$. An optimal pair of parameters is obtained from with the lowest RMSE value from the grid search.

Figure 2 shows the predictive performance of the prediction models for the "traditional-modern" pairwise adjective. The blue solid and red dash lines are the original and predictive adjective scores of all the training product samples respectively. For the result shown in Figure 2(a), the parameter C and the parameter σ of the RBF kernel, obtained by Bayesian inference before applying ARD, was $(C, \sigma^2) = (2.4156, 911.09)$. The performance of the model produced the result of RMSE = 0.147. The training result after applying ARD is shown in Figure 2(b). The optimal parameter sets $(C, \sigma^2) = (2.4156, 911.09)$ was used as the initial values of ARD. The performance of the ARD model using the augmented Mahalanobis kernel was more accurate and improved to where the RMSE = 0.036. The improved performance of the ARD model benefits from the soft feature selection mechanism by gradually adjusting the influence of each feature in the Bayesian inference process. As for the result of LS-SVR with LOOCV shown in Figure 2(c), the optimal parameter sets (C, σ^2) were $(25.119, 251.19)$ and produced the result of RMSE = 0.174. The result of MLR shown in Figure 2(d) produced the result of RMSE = 0.197. Among these four constructed models, LS-SVR with ARD gives the best predictive performance. In this study, the posterior distribution of the model parameters can be approximated by the most probable value using Laplace's method. The inverse covariance matrix can be regarded as confidence intervals (error bars) on the most probable parameters. As shown in Figs. 3(a) and 3(b), the black dotted lines denote the upper and lower bound of the error bars. It can be observed that the width of the error bar after applying ARD became more compact than before applying ARD.

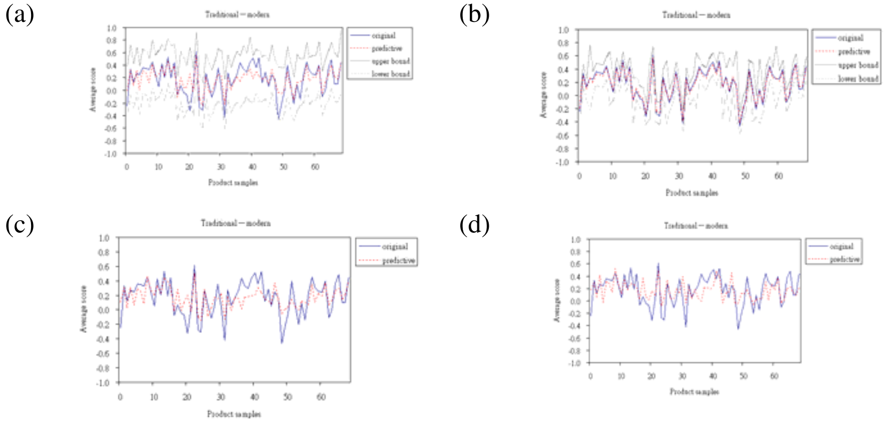

Fig. 2. Performance of the traditional-modern pairwise adjective of (a) LS-SVR with Bayesian inference before ARD, (b) LS-SVR with Bayesian inference after ARD, (c) LS-SVR with LOOCV and (d) MLR

6 Conclusions

In this study, an approach based on LS-SVR and ARD is proposed to deal with the PFFS problem. PFFs are examined by morphological analysis. Pairwise adjectives are used to express CARs. The LS-SVR prediction model is formulated as a function estimation problem taking the form features as the input data and the averaging CAR data as the output value. A backward elimination process based on ARD, a soft-embedded feature selection method, is used to analyze the relative relevance of the form features and obtain the feature ranking. In each elimination step of ARD, the influence of each form feature is adjusted gradually by tuning the length scales of an augmented Mahalanobis kernel with Bayesian inference. Depending on the designated adjective to be analyzed, this proposed method for PFFS provides product designers with a potential tool for systematically determining the relevance of PFFs.

Although the resulting feature ranking obtained by the proposed method based on LS-SVR and ARD, is similar to that of the MLR with BMS, the proposed method gives a higher predictive performance benefits from the soft feature selection mechanism. However, our case study was based on the design of mobile phones 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 main research direction being taken by the authors.

References

[1] Jindo, T., Hirasago, K., Nagamachi, M.: Development of a Design Support System for Office Chairs Using 3-D Graphics. International Journal of Industrial Ergonomics 15, 49–62 (1995)

[2] Han, S.H., Hong, S.W.: A Systematic Approach for Coupling User Satisfaction with Product Design. Ergonomics 46(13/14), 1441–1461 (2003)
[3] Han, S.H., Kim, K.J., Yun, M.H., Hong, S.W., Kim, J.: Identifying Mobile Phone Design Features Critical to User Satisfaction. Human Factors and Ergonomics in Manufacturing 14(1), 15–29 (2004)
[4] Lai, H.H., Lin, Y.C., Yeh, C.H., Wei, C.H.: User-oriented Design for the Optimal Combination on Product Design. International Journal of Production Economics 100, 253–267 (2006)
[5] Hsiao, S.W., Huang, H.C.: A Neural Network Based Approach for Product Form Design. Design Studies 23, 67–84 (2002)
[6] Vapnik, V.N., Golowich, S., Smola, A.: Support Vector Method for Function Approximation, Regression Estimation and Signal Processing. In: Mozer, M., Jordan, M., Petsche, T. (eds.) Advance in Neural Information Processing System 9, pp. 281–287. MIT Press, Cambridge (1997)
[7] Han, S.H., Yang, H.: Screening Important Design Variables for Building a Usability Model. International Journal of Industrial Ergonomics 33, 159–171 (2004)
[8] Shieh, M.D., Yang, C.C.: Multiclass SVM-RFE for Product Form Feature Selection. Expert Systems with Applications 35, 531–541 (2008)