

Modeling customer satisfaction for product development using genetic programming

Abstract

New product development involves several processes in which product planning is the first one. Several tasks normally are required to be conducted in the product planning process and one of them is to determine settings of design attributes for new products. Facing with fierce competition in marketplaces, companies try to determine the settings such that the best customer satisfaction of new products could be obtained. To achieve this, models relating customer satisfaction to design attributes need to be developed first. Previous research has adopted various modelling techniques to develop the models, but those models are not able to address interaction terms or higher order terms in relating customer satisfaction to design attributes, or they are the black-box typed models. In this paper, a method based on genetic programming (GP) is presented to generate models for relating customer satisfaction to design attributes. The GP is first used to construct branches of a tree representing structures of a model where interaction terms and higher order terms can be addressed. Then an orthogonal least squares algorithm is used to determine the coefficients of the model. The models thus developed are explicit, and consist of interaction terms and higher order terms in relating customer satisfaction to design attributes. A case study of a digital camera design is used to illustrate the proposed method.

Keywords: Genetic programming, interaction terms, higher order terms, customer satisfaction, design attributes

1. Introduction

Development of new products especially for consumer products normally involves the processes of product planning, industrial design, embodiment and detail design, and prototyping. The product planning process needs to be performed at the outset of new product development which mainly involves three major tasks; identifying customers and markets to be targeted, defining new products to be developed and determining settings of design attributes of the new products. This paper is aimed at addressing the third task. One of the key issues of the third task is how settings of design attributes of new products can be determined such that a high degree or even optimal customer satisfaction of new products can be obtained. To achieve this, we need to model the relationship between customer satisfaction and design attributes first. However, the modeling process is quite complex as the relationships to be modelled could be highly non-linear and fuzzy, and substantial interactions among design attributions would exist.

Quite a number of studies have been attempted to investigate the modelling of the relationship between customer satisfaction and design attributes / elements. Chen et al. [1] developed a prototype system for affective design in which Kohonen's self-organizing map neural network was employed to consolidate the relationship between design attributes and customer satisfaction. Park et al. [2] adopted the fuzzy rule-based approach to build models relating customer satisfaction to design attributes. Hsiao et al. [3] proposed a method that enables an automatic product form search or product image evaluation by means of a neural-network-based fuzzy reasoning genetic algorithm. The neural-network-based fuzzy reasoning algorithm was applied to establish relationships between the input form parameters and a series of adjectival image words. Liu *et al.* [4] proposed a fuzzy model to examine a customer satisfaction index in e-commerce. They considered a method to calculate the index based on a 5-level quantity table using fuzzy techniques. However, the model they developed was implicit. Lin *et al.* [5] proposed a fuzzy logic model to determine the consumer-oriented mobile phone form design.

From their experimental results, they reported that the fuzzy model outperformed two Neural-Network-based models in terms of the root of mean square errors. Grigoroudis and Siskos [6] developed the MULTicriteria Satisfaction Analysis (MUSA) method for measuring and analyzing customer satisfaction. MUSA is a preference disaggregation model based on the working principles of ordinal regression analysis. Using the survey data, MUSA aggregated individual judgments into a collective value function so as to quantify customer satisfaction. The model assumed that global or overall customer satisfaction was solely measured with respect to a number of customer attributes. This implies that the customer satisfaction model that considers each customer attribute independently was ignored. Grigoroudis *et al.* [7] further applied the MUSA method to measure user-perceived web quality. You *et al.* [8] developed the customer satisfaction models for automotive interior material, using quantification I analysis. Based on the models, the related design variables and preferred design features were examined. Hence, significant design variables and their value in affecting customer satisfaction were identified. Again, the models were generally implicit. Han *et al.* [9] developed a variety of usability dimensions including both subjective and objective aspects, and attempted to evaluate product usability based on statistically regressed models which modeled functional relationships between design attributes and customer satisfaction.

Additionally, various techniques have been attempted to model the functional relationships between customer requirements and design attributes / engineering characteristics in QFD. A multiple linear regression method, which considers non-linear coefficients, were attempted to model the functional relationships in QFD [10]. However, the model is in a polynomial form, and the order of the polynomials generated is user-defined, thus an optimal model could not be generated. Fung *et al.* [11] introduced fuzzy logic to develop fuzzy rule based models for relating customer requirements with engineering characteristics. To address the fuzziness of the modeling, quite a few previous studies have adopted fuzzy set theory on

modeling the relationship. Kim and Park [12] suggested a fuzzy regression approach to estimating the functional relationships in QFD. Chen et al [13] proposed another fuzzy regression approach, based on asymmetric triangular fuzzy coefficients, to develop models for the functional relationships in QFD. The use of non-linear programming to develop fuzzy regression models for modeling the functional relationships in QFD was proposed by Chen [14]. However, the above approaches can only be used to develop models with linear terms, and generation of interaction terms and/or higher order terms of models can not be addressed.

Artificial neural networks [15] have been used to develop nonlinear models especially in manufacturing processes, such as resistance spot welding [16] and transfer molding [17]. These networks have the capability to transform a non-linear mathematical model into a simplified black-box structure, and have the advantage of learning and generalization abilities, as well as nonlinearity. However, the existing neural network approaches normally require a large number of data sets to develop models, which are usually not available in QFD. Also due to their ‘black-box’ nature, behavior of the functional relationships in QFD cannot be made known easily.

Previous research has found that genetic programming (GP) can be used to generate models in a polynomial form in which interaction terms and higher order terms can be considered [18, 19]. Madar et al. [20] and Rodriguez-Vazquez et al. [21] have demonstrated how the GP can be used to generate models with interaction terms and higher order terms. Lakshminarayanan et al [22] has also applied the GP to the modeling of chemical systems by using a small number of data sets. As it is widely recognized that the behaviour of the relationships between customer satisfaction and design attributes is non-linear and the number of data sets for developing those models available in product planning is small many times [23, 24, 25], in this paper, the GP is proposed to develop models for relating customer satisfaction and design attributes. General outcomes of the GP are used to construct the structures of models based on a tree representation. Then, an orthogonal least squares algorithm [26, 27] is introduced

to estimate contributions of branches of the tree so as to determine the coefficients of the models. To evaluate the effectiveness of the proposed method, a case study of digital camera design was conducted. In the case study, a lead user survey of fifteen competitive digital cameras was performed first. Based on the survey results, the GP method was introduced to develop models for relating customer satisfaction to design attributes. The results of the modelling are then compared with those based on fuzzy regression and statistical regression.

2. The GP approach to modelling the relationships between customer satisfaction and design attributes

In new product development, models of relating customer satisfaction to design attributes can be described as follows:

$$y_i = f_i(x_1, x_2, \dots, x_n) \quad (1)$$

where y_i , $i = 1, 2, \dots, m$, is the degree of satisfaction of the i -th dimension of customer satisfaction; x_j , $j=1, 2, \dots, n$, is the setting of the j -th design attribute; and the f_i is a function of the relationships.

In this research, the GP is proposed to generate the models. The pseudocode of the GP used in this research is shown below:

```

t=0
Initialize  $\Omega(t)=[\theta_1(t), \theta_2(t), \dots, \theta_{POP}(t)]$ 

//  $\Omega(t)$  is the population of the  $t$ -th generation.

//  $\theta_i(t)$  is the  $i$ -th individual of  $\Omega(t)$ .

Assign parameters in all  $\theta_i(t)$  by orthogonal least square method

Evaluate all  $\theta_i(t)$  according to a fitness function

while (Terminational condition not fulfilled) do {

    Parent Selection  $\Omega(t+1)$ 

```

Crossover $\Omega(t+1)$

Mutation $\Omega(t+1)$

Assign parameters in all $\theta_i(t+1)$ by orthogonal least square method

Evaluate all $\theta_i(t+1)$

$\Omega(t) = \Omega(t+1)$

$t = t+1$

}

The GP starts with creating a random initial population of individuals $\Omega(t)$ with *POP* individuals $\theta_i(t)$, while $t=0$. Each individual $\theta_i(t)$ is in the form of a tree structure that can be represented in a polynomial function (2) for modeling the functional relationship:

$$y_i = p_0^i + \sum_{j_1=1}^n p_{j_1}^i x_{j_1} + \sum_{j_1=1}^n \sum_{j_2=1}^n p_{j_1 j_2}^i x_{j_1} x_{j_2} + \dots \sum_{j_1=1}^n \dots \sum_{j_n=1}^n p_{j_1 \dots j_n}^i \prod_{k=1}^n x_{j_k}, \quad (2)$$

where y_i is the degree of satisfaction of the i -th customer satisfaction, and x_{j_k} is the j_k -th design attribute.

Then the coefficients of each individual $\theta_i(t)$ are determined by applying an orthogonal least squares method [26, 27]. All individuals are evaluated according to a defined fitness function, which is used to evaluate the goodness-of-fitness of modeling the functional relationship between customer satisfaction and design attributes. The parent selection process uses the goodness-of-fitness of each individual to determine the selection of potential individuals for performing crossover or mutation. Finally, the new individuals with the determined coefficients are evaluated using the fitness function in order to create a new population $\Omega(t+1)$. The process continues until the pre-defined termination condition is met. Major aspects of applying the GP on modeling the relationships are described below:

2.1 Model Representation

The population member of the GP represents the function f_i as shown in equation (2). In the GP, one of the most popular methods to represent structures is hierarchical trees, which are composed of functions F and terminals T [18]. The set of functions F contains the arithmetic operations, +, - and *, of function (2), thus F is represented as $F = \{ +, -, * \}$. The set of terminals $T = \{ \mathbf{x}, \mathbf{p} \}$ contains a design attribute set $\mathbf{x} = \{ x_1, x_2, \dots, x_n \}$ of the function (2) and a coefficient set $\mathbf{p} = \{ p_0, p_1, p_2, \dots, p_{n_t} \}$ of the function, where n_t is the number of terms of the function. A potential solution is depicted as a tree with branches, which consists of operations (internal nodes of the tree) F from the function set and arguments (terminal nodes of the tree) from the terminal set T . For example, Figure 1 shows an example of a hierarchical tree that expresses the following polynomial, which consists of interaction term and higher order terms:

$$(x_1 * x_1) - (x_2 * x_2) + (x_1 * x_2 * x_4)$$

which is equivalent to:

$$x_1^2 - x_2^2 + x_1 \cdot x_2 \cdot x_4$$

The coefficient set $\mathbf{p} = (p_0, p_1, p_2 \text{ and } p_3)$ is determined after determining the structure of the polynomial, where the number of coefficients of the polynomial is 4. Therefore the completed function can be represented as follows.

$$p_0 + p_1 \cdot x_1^2 - p_2 \cdot x_2^2 + p_3 \cdot x_1 \cdot x_2 \cdot x_4.$$

In this research, the coefficients p_0, p_1, p_2 and p_3 are determined using an orthogonal least square algorithm [26, 27], which has been demonstrated to be effective in determining coefficients in a linear-in-parameters model generated by the GP [21]. Details of the orthogonal least square algorithm can be found in [24, 27].

2.2 Fitness function

The fitness function is based on the mean absolute error (MAE), which reflects the differences between the calculated degree of customer satisfaction based on the model and the actual degree of customer satisfaction based on the data sets. The MAE of the j -th individual can be calculated based on (3).

$$\text{MAE}_j = 100\% \times \frac{1}{n_{train}} \sum_{k=1}^{n_{train}} \left| \frac{y(k) - F_j(\mathbf{x}(k))}{y(k)} \right|, \quad (3)$$

where F_j is the model represented by the j -th individual, $y(k)$ is the degree of customer satisfaction of the k -th data set. $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_n(k)]$ is a setting of design attributes of the k -th data set, and n_{train} is the number of data sets used for developing the model. MAE in (3) is commonly known as an indicator of training errors of a model that reflects how well the model fits the data sets. However, a model for relating customer satisfaction and design attributes may contain a lot of unnecessary and complex terms. A complex or over-parameterized model with a large number of parametrical terms reduces the transparency and interpretation of the model. To prevent the GP from generating models that are too complex, a fitness function is designed to balance the tradeoff between the reduction of complexity and model accuracy. In this research, penalty terms are introduced into the fitness function of the GP [27] and the fitness of the j -th individual denoted as:

$$\text{fitness}_j = \frac{-MSE_j}{\left(1 + \exp(c_1(L_j - c_2))\right)}, \quad (4)$$

where fitness_j is the fitness value, L_j is the number of arithmetic operations of the model represented by the j^{th} individual, and c_1 and c_2 are both penalty terms.

2.3 Crossover and mutation

Like other evolutionary algorithms, the two main evolutionary operators are crossover and mutation. The crossover operation produces a pair of offspring that inherit characteristics from both parents by selecting a random node in each of the hierarchical tree structures of the parents (as shown in Figure 2a) and exchanging the associated sub-expressions of the hierarchical tree structures (as shown in Figure 2b). Because of the dynamic representation used in GP, the parents are typically of a different size, shape and content. The process of mapping the genotype onto the phenotype does not correspond to a one-to-one relationship. Therefore the resulting offspring can be expressed by more than one different tree structure and can allow diversification of the population.

Mutation is performed by randomly selecting a node that can be an internal or terminal node, and by replacing the associated sub-expression with a randomly generated sub-expression. For example, Figure 3 shows that the arithmetic operation of a minus is selected and is mutated to a sum.

2.4 Selection and convergence

After the operations of crossover and mutation, individuals from the current population with relatively better fitness defined in (4) are selected to serve as parents for the next generation. The roulette-wheel approach, which is one of the most common selection methods used for selecting individuals to perform reproduction operations in evolutionary algorithms [29], is used for the selection of individuals. The fitness of the j^{th} individual is assigned a value $fitness_j$, and the selection probability value, $prob_j$, is defined as:

$$prob_j = \frac{fitness_j}{\sum_{j=1}^{Popsiz} fitness_j} \quad (5)$$

where Popsiz is the population size of the GP. Equation (5) shows that the individual with a larger fitness value has a higher probability of being selected.

After the selection, the population evolves and improves iteratively until a stopping condition is met. In this research, the stopping criterion is met when the number of generations is equal to a pre-defined number of generations. Otherwise, the GP goes on to the next evolutionary iteration.

3. An Illustrative Example

An example of digital camera design [24] is used to illustrate the GP approach to modeling the relationship between customer satisfaction and design attributes as defined in (2). In this section, two dimensions of customer satisfaction are used for illustrative and validation purposes, which are “Photo quality” (CS_1), and “Take distant image” (CS_2). Degrees of customer satisfaction of them are denoted as y_1 and y_2 respectively. The associated design attributes of the two dimensions of customer satisfaction are “Max. Resolution Support” (DA_1), “Optical Zoom” (DA_2), “Aperture Exposure Control” (DA_3), “LCD size” (DA_4), “Storage Media Support” (DA_5), and “Weight” (DA_6). Value settings of them are denoted as x_1 , x_2 , x_3 , x_4 , x_5 and x_6 respectively. A lead user survey of 15 competitive digital cameras was conducted. Table 1 shows the survey results and values of the design attributes of the corresponding competitive digital cameras.

The proposed GP approach to modeling the relationship between customer satisfaction and design attributes was implemented using MATLAB and a prototype system was developed. The GP parameters were set as shown in Table 2 with reference to [20]. Figure 4 shows the number of dimensions of customer satisfaction, number of design attributes and number of competitive products inputted to the prototype system for this case study. Figure 5 shows the input of the survey results. After 100 generations of the GP, the two models were generated as shown in Tables 3 and 4 respectively for CS_1 and CS_2 . The convergence plots, which illustrate

the progresses of the runs of the GP, are shown in Figure 6 and Figure 7 respectively for CS₁ and CS₂.

Since the GP is a stochastic method, different results could be obtained from different runs. To evaluate its overall performance, 30 runs on the GP were carried out. For each modelling, the mean of the 30 runs was calculated. To compare the effectiveness of the GP in modeling customer satisfaction with those of statistical linear regression (LR) [30] and fuzzy regression (FR) [23, 31], the same data sets were used to develop customer satisfaction models based on LR and FR. Table 3 and Table 4 show the models developed for CS₁ ‘photo quality’ and CS₂ ‘take distant image’ based on the three methods respectively. It can be seen that the interaction terms exist in the CS₁ model developed based on the GP and second order terms exist in the CS₂ model developed based on the GP. No interaction and/or higher order terms are found in models developed based on LR and FR.

The *N*-fold cross validation was used to evaluate the GP approaches as compared with statistical linear regression and fuzzy regression in modeling customer satisfaction. The trials of the cross validation were repeated 30 times. For each trial, 13 of the 15 data sets were used for model training while the remaining 2 data sets were used for validating the trained models. The two measures, training error and validation error, were used to evaluate the effectiveness of the GP approach. The training error is used to reflect how well the developed models can fit the training data sets. The validation error is used to reflect how well the developed models can predict a response.

The 15 training errors based on the three methods, LR, FR and GP, are shown in Figures 8(a) and 8(b). The *x*-axis of the figures indicates the data sets which were used for testing. For example, the (1,2) means that the 1st and 2nd data sets were used for testing, and the rest of the data sets, 3rd to 15th data sets, were used for model training. It can be seen from the figures that the GP yields the smallest number of training errors in modelling both the CS₁ ‘photo quality’

and CS₂ ‘take distant image’. Figures 9(a) and 9(b) show the validation errors based on the three methods for the CS₁ ‘photo quality’ and CS₂ ‘take distant image’ respectively. It shows that the models based on GP yields the smallest number of validation errors in both the CS₁ and CS₂. Means and variances of the thirty validation errors of the three methods are shown in Table 5 and Table 6 respectively for CS₁ and CS₂ from which it can be found that the GP yields the smallest number of validation errors. Also from the two tables, it can be found that the means of the thirty training errors based on the GP are the smallest as compared with those based on LR and FR. T-test was used to evaluate the significance of the improvement. Table 7 shows the t-values of the T-test from which it can be found that all the t-values in Table 7 are higher than 1.89. Based on the normal distribution table, if the t-value obtained is higher than 1.89, it can be said that the performance of GP is better than LR and FR with a 97% of confidence level in all benchmark functions.

The developed models can be incorporated into an optimization model with an objective of maximizing customer satisfaction. By solving the model, an optimal / near optimal setting of design attributes can be obtained. Of course, design teams could pre-define degree of satisfaction of some dimensions of customer satisfaction in view of different design scenarios. For example, if the degrees of satisfaction of ‘photo quality’ and ‘take distant image’ are set as 4 and 3 respectively, the corresponding models become as follows.

$$0.2905 + 0.7342 \cdot x_1 + 0.6727 \cdot x_2 - 0.6727 \cdot x_4 + 0.1526 \cdot x_1 \cdot x_3 - 0.1526 \cdot x_1 \cdot x_2 = 4$$

$$0.3703 + 1.1359 \cdot x_2 - 0.0645 \cdot x_4 + 0.0645 \cdot x_5 - 0.0645 \cdot x_2^2 - 0.0645 \cdot x_5^2 = 3$$

These two models can be treated as two constraints of the optimization model. By solving the model, an alternative setting of design attributes can be obtained.

4. Conclusions

In this paper, a GP method has been proposed to develop models for relating customer satisfaction and design attributes, based on a small size of survey data sets. In the proposed method, the GP is used to construct structures of models based on a tree representation, in which interaction terms and higher order terms can be generated. Then an orthogonal least squares algorithm is used to estimate the contribution of each branch of the tree so as to identify the coefficients of the models. Since interaction terms and/or higher order terms can be introduced to the branches of the trees in GPs, models with interaction terms and higher order terms can be generated. To evaluate the effectiveness of the proposed method, a case study of a digital camera was carried out. Models based on the GP for relating customer satisfaction to design attributes were developed. The effectiveness of the models was compared with that of the models based on statistical linear regression and fuzzy regression. Results of the comparison show that the models developed based on GP yields fewer training errors and fewer validation errors.

Future work would involve the determination of an optimal setting of design attributes for a new product. To achieve this, the models for relating customer satisfaction and design attributes are developed based on the GP approach. Then, the models are incorporated into the formulation of an optimization model with the objective ‘maximizing customer satisfaction’. By solving the optimization model, an optimal setting of design attributes for the new product can be obtained. In addition, since uncertainty due to fuzziness is unavoidable in modelling the relationship between customer satisfaction and design attributes, future work will involve integrating GP with fuzzy theory for the modelling.

Acknowledgement

The work described in this paper was financially supported by the Research Grant Council of Hong Kong, China (Project no. PolyU 5184/07E)

References

- [1] Chen, C.H., Khoo L.P., Yan W., 2006. An investigation into affective design using sorting technique and Kohonen self-organizing map. *Advances in Engineering Software*, 37, 334-349.
- [2] 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.
- [3] 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.
- [4] Liu, X., Zeng, X., Xu, Y., Koehl, L., 2007. A fuzzy model for customer satisfaction index in e-commerce. *Mathematics and Computers in Simulation*, 77, 512-521.
- [5] Lin, Y. C., Lai, H. H., Yeh, C. H., 2007. Consumer-oriented product form design based on fuzzy logic: A case study of mobile phones. *International Journal of Industrial Ergonomics*, 37, 531-543.
- [6] Grigoroudis, E., Siskos, Y., 2002. Preference disaggregation for measuring and analyzing customer satisfaction: The MUSA method. *European Journal of Operational Research*, 143, 148-170.
- [7] Grigoroudis, E., Litos, C., Moustakis, V. A., Politis, Y., Tsironis, L., 2008. The assessment of user-perceived web quality: Application of a satisfaction benchmarking approach. *European Journal of Operational Research*, 187, 1346-1357.
- [8] You, H., Ryu, T., Oh, K., Yun, M. H., Kim, K. J., 2006. Development of customer satisfaction models for automotive interior materials. *International Journal of Industrial Ergonomics*, 36, 323-330.
- [9] 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.
- [10] Dawson, D., and Askin, R. G. (1999). Optimal new product design using quality function deployment with empirical value functions. *Quality and Reliability Engineering International*. 15, 17-32.
- [11] Fung, R. Y. K., Popplewell, K., Xie, J., 1998. An intelligent hybrid system for customer requirements analysis and product attribute targets determination. *International Journal of Production Research*, 36, 13-34.
- [12] Kim, K., Park, T., 1998. Determination of an optimal set of design requirements using house of quality. *Journal of Operations Management*, 16, 569-581.

- [13] Chen, Y., Tang, J., Fung, R.Y.K., Ren, Z., 2004. Fuzzy regression-based mathematical programming model for quality function deployment. *International Journal of Production Research*, 42(5), 1009-1027.
- [14] Chen, Y., Chen, L., 2006. A non-linear possibilistic regression approach to model functional relationships in product planning. *International Journal of Advanced Manufacturing Technology*, 28, 1175-1181.
- [15] Simpson P.K., *Artificial neural systems*, New York: Pergamon Press, 1989.
- [16] Li H.L., Chou T. and Chou C.P., Optimization of resistance spot welding process using Taguchi method and a neural network, *Experimental Techniques*, vol. 31, no. 5, pp. 30-36, 2007.
- [17] Tong K.W., Kwong C.K., Yu K.M., Intelligent process design system for the transfer moulding of electronic packages, *International Journal of Production Research*, vol. 42, no. 10, pp. 1911-1931, 2004.
- [18] Koza, J., 1992. *Genetic Programming: On the Programming of Computers by Means of Natural Evolution*, MIT Press: Cambridge.
- [19] Koza, J., 1994. *Genetic Programming II: automatic discovery of reusable programs*, MIT Press.
- [20] Madar, J., Abonyi, J. and Szeifert, F., 2005. Genetic programming for the identification of nonlinear input – output models, *Industrial and Engineering Chemistry Research*. 44, 3178 – 3186.
- [21] Rodriguez-Vazquez, K., Fonseca, C.M., Fleming P.J., 2004. Identifying the structure of nonlinear dynamic systems using multiobjective genetic programming. *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans*, 34(4), 531 – 545.
- [22] Lakshminarayanan, S., Fujii, H., Grosman, B., Dassau, E., Lewin, D.R., 2000. New product design via analysis of historical databases, *Computers and Chemical Engineering*. 24, 671-676.
- [23] Fung, R.Y.K., Chen, Y.Z., Tang, J.F., 2006. Estimating the functional relationships for quality function deployment under uncertainties. *Fuzzy Sets and Systems*, 157, 98-120.
- [24] Kwong, C.K., Chen, Y., Bai, H., Chan, D.S.K., 2007. A methodology of determining aggregated importance of engineering characteristics in QFD. *Computers and Industrial Engineering*, 53(4), 667-679.
- [25] Xu, D., Yan, H.S., 2006. An intelligent estimation method for product design time. *International Journal of Advanced Manufacturing Technology*, 30, 601-613.
- [26] Billings, S., Korenberg, M., Chen, S., 1988. Identification of nonlinear outputaffine systems using an orthogonal least-squares algorithm. *International Journal of Systems Science*, 19, 1559-1568.

- [27] Chen, S., Billings, S., Luo, W., 1989. Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, 50, 1873-1896.
- [28] McKay, B., Willis, M.J., Barton, G.W., 1997. Steady-state modeling of chemical processes using genetic programming. *Computers and Chemical Engineering*, 21(9), 981-996.
- [29] Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. United States of America: Addison Wesley Longman, Inc.
- [30] Seber, G.A.F., 2003. *Linear regression analysis*, Wiley.
- [31] Chen, Y.Z., Fung, R.Y.K., Tang, J.F., 2005. Fuzzy expected value modeling approach for determining target values of engineering characteristics in QFD. *International Journal of Production Research*, 43(17), 3583-3604.

Table 1 The survey results of the fifteen competitive digital cameras

Brands of digital cameras	Customer satisfaction (CS)		Design attributes (DAs)					
	CS ₁ (y ₁)	CS ₂ (y ₂)	DA ₁ (x ₁)	DA ₂ (x ₂)	DA ₃ (x ₃)	DA ₄ (x ₄)	DA ₅ (x ₅)	DA ₆ (x ₆)
A	4	3	5	3	3	1.85	2	323
B	1	1	2	1	0	1.6	3	90
C	4	5	4	10	3	2.2	2	337
D	5	5	3.2	10	3	1.5	2	441
E	3	3	3.1	3	0	1.5	1	180
F	2	3	4.23	3	0	1.5	1	165
G	2	3	4	3	0	1.5	2	185
H	5	4	6.3	6	3	1.8	3	590
I	2	3	3.24	3	0	1.6	2	150
J	2	3	5.1	3	0	2.5	1	200
K	3	3	1.95	3	3	1.5	2	210
L	5	3	5	3	3	2.5	1	705
M	5	5	8	7	3	1.8	3	906
N	2	3	1.92	3	2	2	1	300
O	4	4	4	4	3	1.8	1	490

Table 2 Parameter setting of the GP

Population size	50
Maximum number of evaluated individuals	5000
Generation gap	0.9
Probability of crossover	0.5
Probability of mutation	0.5
Probability of changing terminal via non-terminal	0.25
Penalty factors in the fitness function (4)	$c_1=0.5$ and $c_2=30$

Table 3 Developed models for CS₁ 'photo quality'

Methods	Models
Linear regression model	$y_1 = 2.07 + 0.120 \cdot x_1 + 0.113 \cdot x_2 + 0.315 \cdot x_3 - 0.639 \cdot x_4 - 0.210 \cdot x_5 + 0.00389 \cdot x_6$
Fuzzy regression model	$y_1 = (2.2503, 1.8253) + (-0.3736, 2.5993) \cdot x_1 + (0.4667, 1.8900 \times 10^{-11}) \cdot x_2 + (0.4992, 1.3983 \times 10^{-11}) \cdot x_3 + (-2.0806 \times 10^{-2}, 3.9336 \times 10^{-11}) \cdot x_4 - (7.3416 \times 10^{-1}, 2.9729 \times 10^{-11}) \cdot x_5 + (1.4758, 1.8957 \times 10^{-11}) \cdot x_6$
GP based model	$y_1 = 0.2905 + 0.7342 \cdot x_1 + 0.6727 \cdot x_2 - 0.6727 \cdot x_4 + 0.1526 \cdot x_1 \cdot x_3 - 0.1526 \cdot x_1 \cdot x_2$

Table 4 Developed models for CS₂ ‘take distant image’

Methods	Models
Linear regression model	$y_2 = 2.87 + 0.210 \cdot x_1 + 0.313 \cdot x_2 + 0.153 \cdot x_3 - 0.569 \cdot x_4 - 0.454 \cdot x_5 + 0.00042 \cdot x_6$
Fuzzy regression model	$y_2 = (2.0940, 2.8422 \times 10^{-14}) + (6.1867 \times 10^{-1}, 0) \cdot x_1 + (1.5433, 0) \cdot x_2 + (1.0822 \times 10^{-1}, 0) \cdot x_3 + (-3.6483 \times 10^{-1}, 0) \cdot x_4 + (-5.6080 \times 10^{-1}, 2.8765) \cdot x_5 + (-6.6674 \times 10^{-2}, 0) \cdot x_6$
GP based model	$y_2 = 0.3703 + 1.1359 \cdot x_2 - 0.0645 \cdot x_4 + 0.0645 \cdot x_5 - 0.0645 \cdot x_2^2 - 0.0645 \cdot x_5^2$

Table 5 Means and variances of training and validation errors of developed models for CS₁

‘photo quality’

		Training error	Validation error
Linear regression model	Mean	6.419	23.852
	Variance	1.6305	84.218
Fuzzy regression model	Mean	8.0269	31.478
	Variance	1.3078	142.75
GP based model	Mean	3.1228	17.395
	Variance	1.8009	59.825

Table 6 Means and variances of training and validation errors of developed models for CS₂ ‘take distant image’

		Training error	Validation error
Linear regression model	Mean	29.836	19.753
	Variance	58.852	145.83
Fuzzy regression model	Mean	40.078	22.917
	Variance	107.15	256.92
GP based model	Mean	19.964	15.616
	Variance	78.531	126.32

Table 7 T-values of training and validation errors of developed models for CS₁ and CS₂

	CS ₁		CS ₂	
	Training error	Validation error	Training error	Validation error
LR-GP	9.7463	2.9468	4.6132	1.9735
FR-GP	15.235	5.4195	8.0849	2.0427

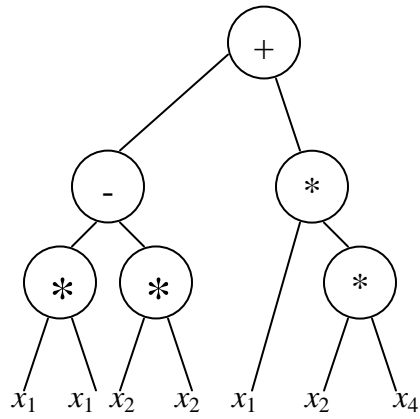


Figure 1 An example of a hierarchical tree

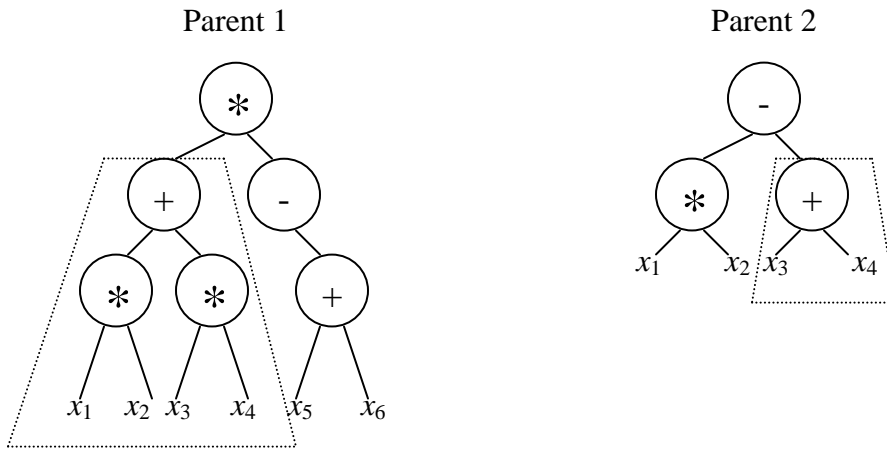


Figure 2(a) Random selection of a sub-expression before crossover

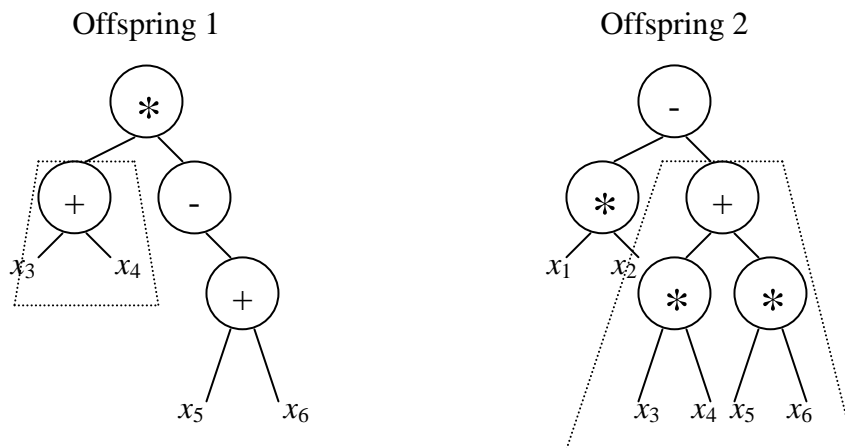


Figure 2(b) Offspring produced by the crossover

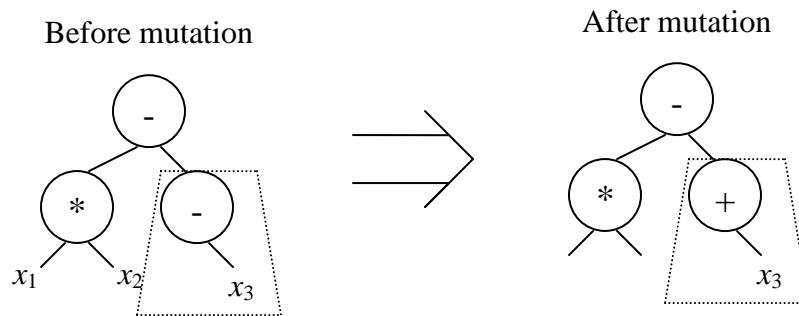


Figure 3 Offspring produced by the mutation

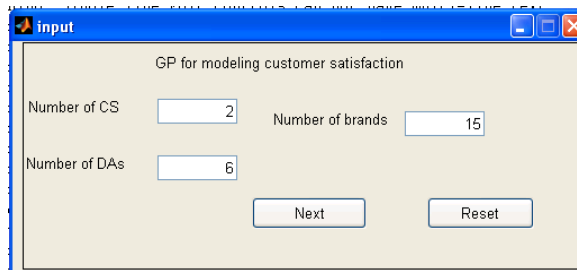


Figure 4 An input interface of the prototype system

		Customer satisfaction				Design attributes			
		y1	y2	x1	x2	x3	x4	x5	x6
Brands		4	3	5	3	3	1.8	2	32
		1	1	2	1	0	1.8	3	90
		4	5	4	10	3	1.8	2	33
		5	6	3.2	10	3	1.8	2	44
		3	3	3.1	3	0	1.8	1	18
		2	3	4.2	3	0	1.8	1	16
		2	3	4	3	0	1.8	2	18
		5	4	6.3	6	3	1.8	3	59
		2	3	3.2	3	0	1.8	2	19
		2	3	5.1	3	0	2.5	1	20
		3	3	1.9	3	3	1.8	2	21
		5	3	5	3	3	2.5	1	70
		5	5	8	7	3	1.8	3	90
		2	3	1.9	3	2	2	1	30
		4	4	4	4	3	1.8	1	49

Figure 5 Input of survey results of competitive digital cameras

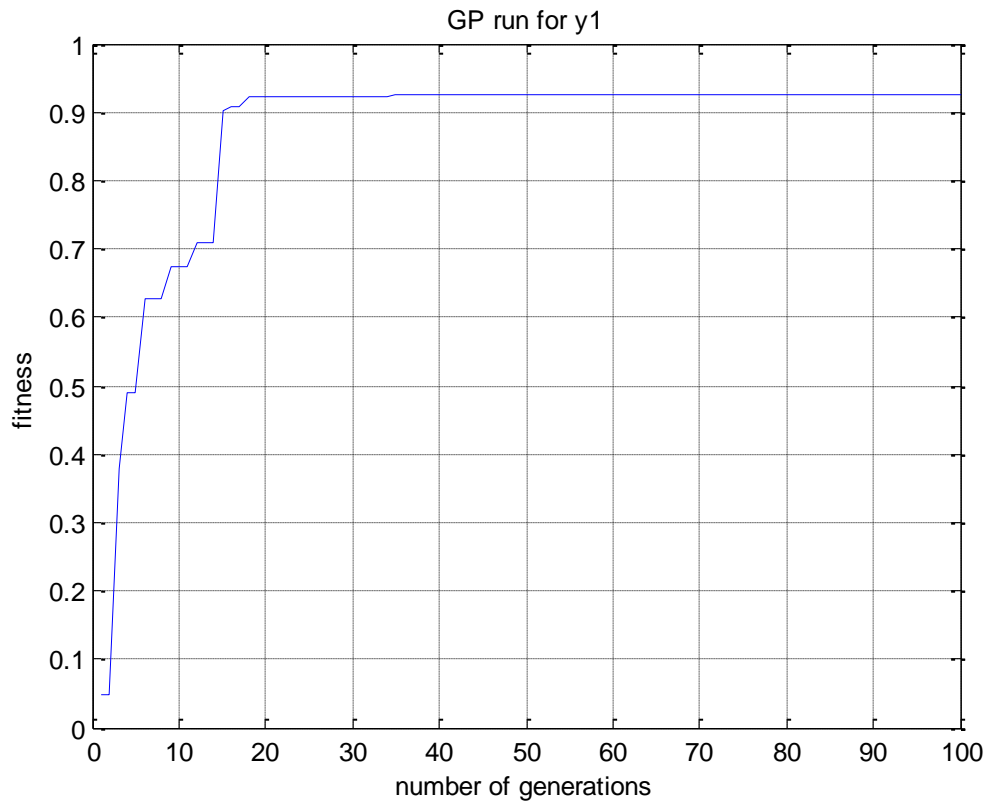


Figure 6 Convergence plot of the GP run for CS₁ 'photo quality'

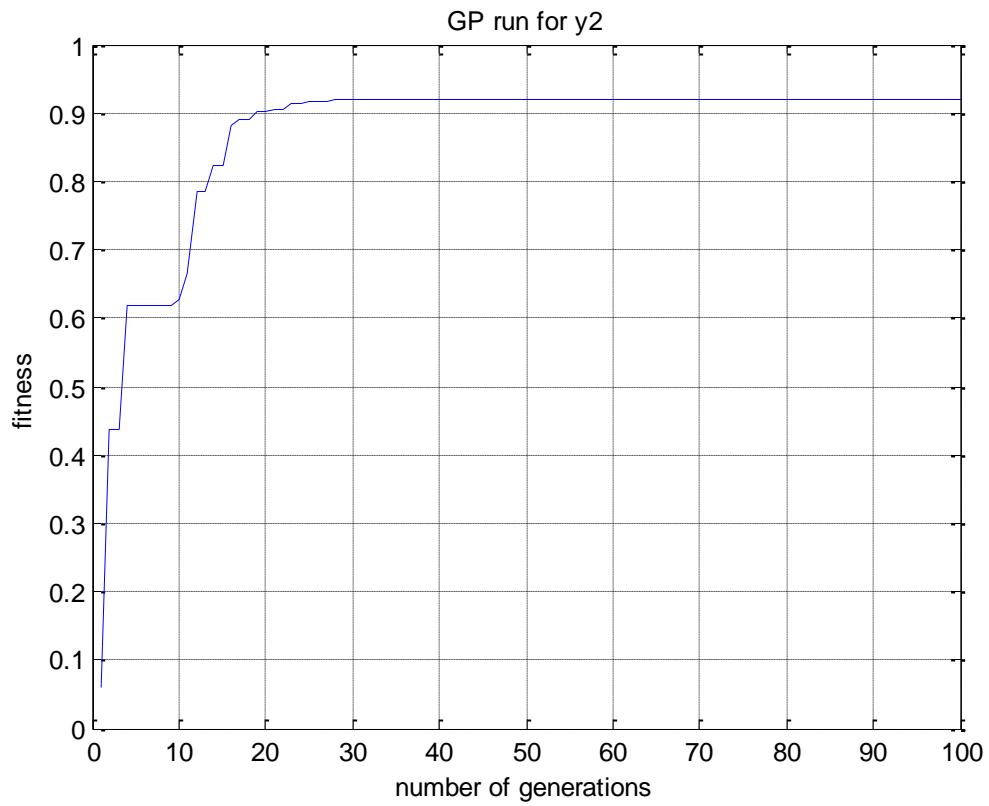


Figure 7 Convergence plot of the GP run for the CS₂ ‘take distant image’

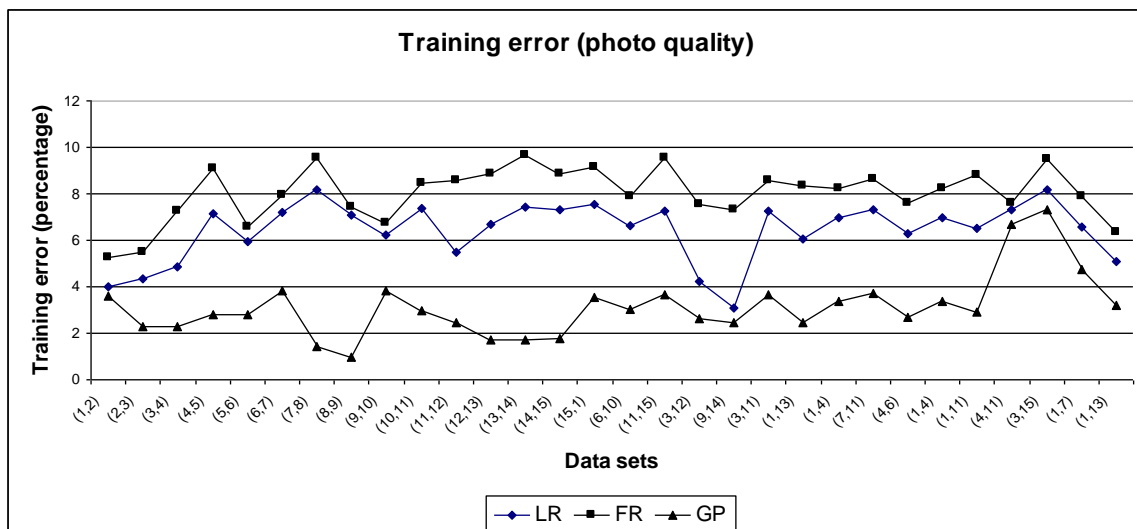


Figure 8(a) Training errors of CS₁ ‘photo quality’

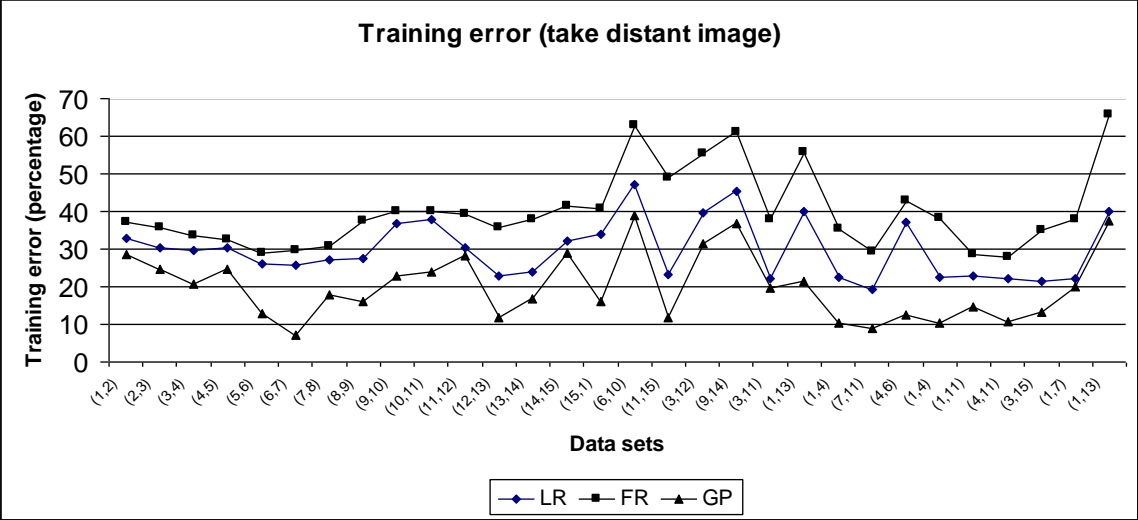


Figure 8(b) Training errors of CS₂ ‘take distant image’

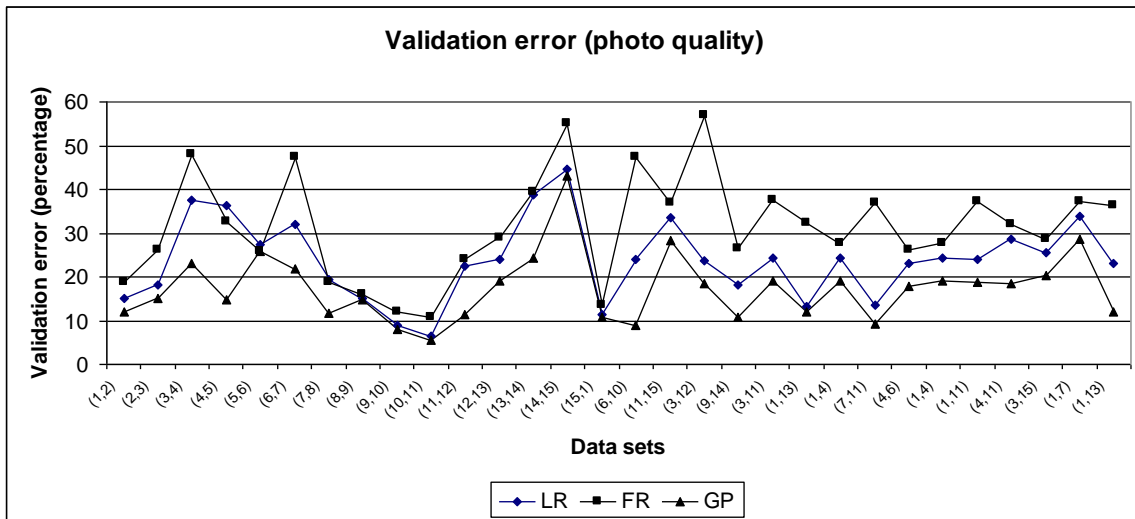


Figure 9(a) Validation errors of CS₁ 'photo quality'

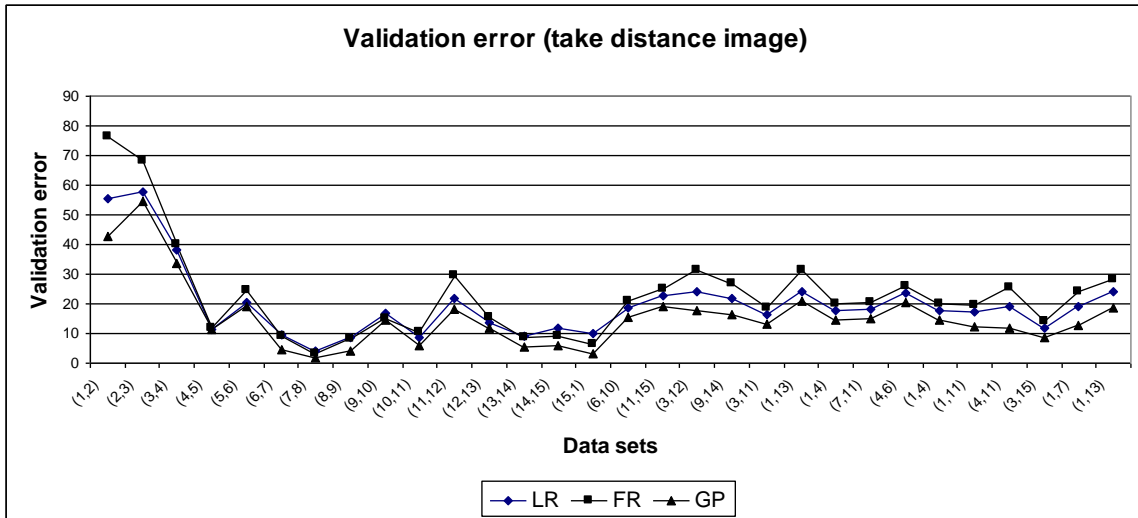


Figure 9(b) Validation errors of CS₂ 'take distant image'