A Genetic Algorithm and Support Vector Regression based Hybrid Cost Estimation Model for Feature-based Plastic Injection Products

서 광 규*
*상명대학교 경영공학과

특징기반 플라스틱 사출제품을 위한 유전자 알고리즘과 Support Vector Regression 기반의 하이브리드 비용 평가 모델

Kwang-Kyu Seo*
*Dept. of Management Engineering, Sangmyung University

Abstract

플라스틱 사출 제품은 다양한 가전제품과 하이테크 제품에 널리 사용되고 있다. 그러나 현재의 치열한 경쟁적 비즈니스 환경에서 플라스틱 사출 제품 제조업자들은 고객을 만족시키면서 경쟁력을 얻기 위하여 다른 경쟁자들보다 먼저 새로운 제품을 시장에 출시하고 신제품의 개발기간을 줄이기 위한 노력을 할 여유가 부족하다. 따라서 무한경쟁의 시장에서 살아남기 위해서는 제조업자들은 시장 마켓 점유를 빠르게 올리는 것과 동시에 제품의 가격 경쟁력을 가져야 한다. 특징기반 모델의 구조는 현재 연구에서 3D 제작 도구로서 일반적으로 적용되고 있으며 신제품 개발 엔지니어들이 새로운 제품의 개념을 개발하는 데에도 널리 사용되고 있다. 본 연구에서는 특징기반 플라스틱 사출제품을 위한 유전자 알고리즘과 Support Vector Regression (SVR) 기반의 새로운 하이브리드 비용 평가 모델을 제안한다. 제안하는 하이브리드 모델은 기존의 플라스틱 사출제품의 비용평가절차와 계산을 위해 필요로 하는 변수들을 극적으로 간단하게 하고 줄일 수 있다. 사례연구에서는 제안하는 하이브리드 모델과 기존의 multilayer perceptron networks (MLP) 및 pure SVR과의 비교분석을 통하여 제안모델이 플라스틱 사출 제품의 개발단계에서의 비용평가문제를 해결하는데 효율성과 효과성이 있음을 입증한다.

Keywords : Cost Estimation, Feature-based Model, Plastic Injection Product, Genetic Algorithm, Support Vector Regression, Hybrid Model

1. Introduction

The cost estimation plays a very important role in design and production stages as well as a fairly important role in the company business decision-making. Accurate cost estimation can satisfy the optimization demand at design as well as the customers' satisfaction requirements for minimum cost, highest quality and timely delivery when selecting suppliers[6].

Plastic injection parts in products has become the most widely applied mass-production technology and are used more and more. And with respect to cost estimation, product design development team in the past could only estimate the final product total cost after getting the quotations of the plastic injection parts after the design stage leading to development cost estimation delay.

† 교신저자: 서광규, 충남 천안시 동남구 상명대길 31 상명대학교 공과대학 경영공학과
M・P: 016-718-2682, E-mail: kwangkyu@smu.ac.kr
2012년 7월 20일 접수; 2012년 9월 5일 수정본 접수; 2012년 9월 17일 게재확정

Moreover, rules of thumb of the engineers are often applied as the cost estimation benchmarks, making the results controversial in terms of accuracy. Although calculation by cost model has the advantage of timeliness, only representative values exclusive of indirect tasks cost and raw materials cost are calculated resulting in inadequate estimation accuracy.

Product cost estimation varies widely ranging from standard spare parts manufacturing cost estimation to the cost analysis of the optimized technology and marketing fees of highly customized assembled products with appropriate product estimation models available at stages from product concept design stage to the product design cycle final stages. Therefore, we classified the available cost estimation technologies to choose proper cost estimation models for product price estimation. Zhang, Fuh, and Chan [12] categorized cost estimation techniques into traditional detailed breakdown, simplified-breakdown, group technology-based, regression-based and activity-based cost approaches. Shehab and Abdalla [9] proposed intuitive, parametric, variant-based and generative cost estimating approaches. Cavalieria et al. [1] proved three cost analyses of analogy-based, parametric and engineering approaches. Seo et al.[8] explored an approximate method for providing the preliminary life cycle cost based on ANNs. The review of the existing literature indicates that ANNs are the main and commonly used techniques in cost estimation.

Among the artificial intelligence techniques, support vector machines (SVMs), which were introduced by Vapnik [3], have received much attention with remarkable results. This technique has recently found numerous applications in the fields of machine learning and data mining. Currently, support vector regression (SVR) has been introduced to solve nonlinear regression prediction problems. Although the SVR can be regarded as a neural network, some features of the SVR such as its lesser number of adjustable parameters and faster learning speed, are superior to conventional neural networks. This technique has been applied successfully to solve forecast and prediction problems in many different fields. Although SVR gives good prediction performance and generalization ability, it has two problems such as how to choose the optimal input feature subset, and how to set the best kernel parameters of SVR. These two problems are crucial, because the feature subset selection influences the appropriate kernel parameters and vice versa. Therefore, obtaining the optimal feature subset and SVM parameters must occur simultaneously.

The purpose of this paper is to improve the performance of cost estimation of feature-based plastic injection products by using hybrid techniques based on genetic algorithm (GA) and SVR. To achieve this purpose, a GA and SVR are integrated into a cost estimation model. The selection of the parameters in the SVR technique is optimized by using GA. The proposed hybrid model uses the advantages of GA and SVR to reduce the computational complexity and the time required to design the SVR for cost estimation model of plastic injection products.

2. Research Background

2.1 Genetic Algorithm (GA)

GAs are computational models of evolution. They operate on the basis of a set of candidate solutions. Each candidate solution is called a ‘chromosome’, and the entire set of solutions is called a ‘population’. The algorithm enables iterative transformation from one population of chromosomes to a new population. Each iteration is called a ‘generation’. There are various forms of GAs [4].

In the static population model which is a simple version, the population is ranked according to the fitness of each chromosome. In each generation, two (only two) chromosomes are selected as parents for reproduction. GAs operate iteratively on a population of chromosomes, each of which represents a candidate solution to a given problem, suitably encoded as a string of symbols (e.g., binary code). A randomly generated set of such strings forms the initial population from which the GA starts its search. Three basic genetic operators guide this search: selection, crossover, and mutation. The genetic search process is iterative: each iteration (generation) involves evaluating, selecting, and recombining strings in the population, until a given termination condition is satisfied. Evaluation of each string is based on a fitness function that is problem -dependent. It determines which of the candidate solutions are better.

2.2 Support Vector Regression (SVR)

SVMs were motivated by statistical learning theory; the aim was to solve only the given problem, without solving a more difficult problem as an intermediate step. SVMs are based on the principle of structural risk minimization, which is closely related to regularization theory. This principle incorporates capacity control to prevent over-fitting, thus, it is a partial solution to the problem of bias-variance trade-off. Two key elements in the implementation of SVMs are the techniques of mathematical programming and kernel functions.

SVR is a new training technique based on SVM; however, it requires only the solution of a set of linear equations instead of the long and computationally hard quadratic programming problem involved in the standard SVM. In fact, the SVR works with a least squares cost function [10]. The SVR model can be expressed as equation (1):

$$\min \frac{1}{2} w^T w + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) \tag{1}$$

The SVR model can be rewritten as follows:

$$y = f(x) = \sum_{i=1}^{N} \hat{\alpha}_i K(x, x_i) + \hat{b} \tag{2}$$

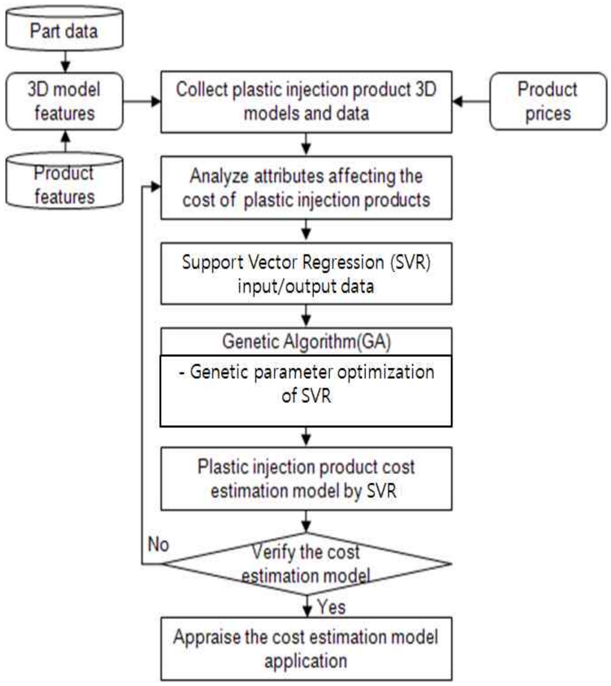
where $K(x, x_i)$ is a kernel function. In this paper, a nonlinear radial basis function (RBF) kernel is considered as equation (2):

$$K(x, x_i) = \exp\left(-\frac{1}{2\sigma^2} \parallel x - x_i \parallel^2\right) \tag{3}$$

where σ is the parameter of the RBF Kernel. In comparison with some other feasible kernel functions, the RBF is a more compact supported kernel and able to shorten the computational burden of training process. It also improves the generalization performance of SVR [7]. To achieve a high level of performance with SVR models, the regularization parameter C in equation (1) which is a cost parameter that controls the amount of overlap and the value of σ from the kernel function have to be set carefully[2].

3. The Hybrid Cost Estimation Model based on GA and SVR for Plastic Injection Products

The proposed hybrid model includes two different techniques, namely GA and SVR. In this model, SVR acts as a supervised learning tool to handle input - output mapping and is focused on cost data characteristics in of plastic injection products, and GA works to optimize the SVR parameters. In fact, the generalization capability and predictive accuracy of the SVR are determined by the problem parameters, including free parameters (i.e., C and σ). To select these parameters, most researchers follow the trial-and-error procedure. They construct a few SVR models based on different parameter sets, then test them on a validation set to obtain optimal parameters. However, this procedure requires some luck and often is time-consuming [2]. This paper solves the above-mentioned shortcomings. Moreover, the k-fold cross validation is utilized to train the SVR in order to reach a more realistic evaluation of the accuracy by dividing the total dataset into multiple training and test sets, and to provide the reliable results. The proposed hybrid model is a computationally efficient combination and is helpful in the cost prediction of plastic injection products. In this model, we could quickly estimate the plastic injection product quotations by input and output variants requested for SVR construction and the training. The plastic injection product quotations data previously collected are used to set up samples for learning and testing to verify the feasibility of applying SVR in quotation estimation. An explanation of the major steps involved in the proposed model is provided as shown in <Figure 1>



<Figure 1> Procedures of the proposed model for plastic injection product cost estimation

3.1 Features Data Collection

As plastic injection products have profits ranging from 20%~40% in different industries as well as the manufacturers' tight control of spare parts features drawings and purchase/quotation data, we cannot get large sum of spare parts drawings from various industries and corresponding quotations from many suppliers of different sizes.

Therefore, this study only discusses the cost estimation model on the basis of notebook computer industry with one supplier's data collected. As we know, a notebook has a various components such as PCB, electric and plastic parts, but we focus on plastic injection parts. The model in this study is mainly to collect the features data with the product design and development team as the main factor considered. We obtained the price quotations of manufacturers quoted or the purchase prices from the buyers according to material numbers. Moreover, we obtained the cost-affecting items from the drawings, and adjusted according to the plastic injection product manufacturing types. The collected reference factors for plastic injection cost estimation are as shown in <Table 1>.

<Table 1> Cost factors

Cost factors	Cost factors adapted in this study
Raw material cost	Product volume, Surface area, cavity quantity, product net weight
Manufacturing cost	projection area, minimum length, width, height and thickness of the product, cavity quantity
Quality control fee	Spare parts quotations or purchase unit price
Selling cost	supplier's cost
Profit	profit rate is about 20%

With the development of CAD/CAM technologies, plastic injection products tend to be streamlined and complex in appearance. However, veteran techno logists in the injection product plants will often appraise on the key items of product benchmarks and specific product design specification items. Hence, this study lists the following parameter drawing software (or feature-based 3D software such as Pro/Engineer or Solid Works) after considering cost-affecting factors data to get the most important item of the product feature model design specifications in addition to referring to the discussion of Wang and Che et al. [11] on the plastic injection product cost estimation parameters. The factors having comparatively bigger effects on cost are found out from plastic injection product as follows:

- volume: the space the product occupies
- material: different materials have different unit prices and mass densities
- product net weight: volume (cm3) * material density (g/cm3)
- surface area: the sum of spare parts surfaces
- number of cavity: number of products for each molding in injection molding process
- projection area: the area of the product in the parting line (cm2), affecting the injection machine selection
- maximum measurements: the minimum length, width and height of the box to contain the product

In accordance with the cost estimation model parameters in <Table 1>, this study collects the relevant data after distinguishing each spare part' features into material features, form features, and molding conditions and parts of the sample data after collection are as shown in <Table 2>. <Table 3> shows the partial list of part feature date of a notebook.

<Table 2> Part features structure

Product	Feature	Attribute
Product  (notebook)	Material features	Material Volume Weight Surface area
	Form features	Length Width Height Thickness
	Molding conditions	Cavity quantity Projection area

<Table 3> Partial list of part feature date of a notebook

P/N	Attributes	Value		
P0001	Material	ABS	PP	PE
	Volume	2488.5	2488.5	2488.5
	Weight	2.9	2.8	2.2
	Surface area	4325.5	4325.5	4325.5
	Length	90.4	90.4	90.4
	Width	41.5	41.5	41.5
	Height	8.5	8.5	8.5
	Thickness	1.5	1.5	1.5
	Cavity quantity	10	10	10
	Projection area	1655.8	1655.8	1655.8
	Price	1.1	1.1	0.9

3.2 Learning Process of the Hybrid Model

In this step, SVR is deployed to handle input-output mapping. The RBF kernel is applied as reasonable choice [5]. By using the k-fold cross validation technique on the training dataset, the SVR training is performed to obtain the prediction model.

Fitness definition: According to [2] the fitness of the training dataset can be obtained easily, but is prone to over-fitting. To handle this problem the k-fold cross validation technique is utilized. In this technique, after randomly dividing the training dataset into k subsets, the regression function is built with a given set of parameters (C_i, σ_i) by using $k-1$ subsets as the training set. The last subset is considered the validation. The above procedure is repeated k times. Consequently, the fitness function is defined as the $MAPE_{CV}$ of the k-fold cross validation technique on the training dataset:

$$\text{Fitness} = \min f = MAPE_{CV} \tag{3}$$

$$MAPE_{CV} = \frac{1}{l}\sum_{j=1}^{l}\left|\frac{y_j-\hat{y}_j}{y_j}\right|\times 100\% \tag{4}$$

where y_j is the actual value; $\hat{y}_j$ is the validation value and l is the number of subsets. The solution with a smaller $MAPE_{CV}$ of the training dataset has a smaller fitness value. In addition, the SVR utilizes free parameters, which are randomly generated and employed by a GA. By using a GA, the proposed model generates a population and then evaluates it according to the SVR. This evolutionary algorithm is simple and straightforward for the SVR. The GA is utilized to search for the SVR parameters concurrently.

3.3 Experimental Results

This study will refer to three test appraisal benchmark data, which are "mean absolute percent error", "root mean squared error", and "coefficient of determination" as follows:

(1) MAPE (mean absolute percent error)

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \tag{5}$$

where N is the number of data

We can learn the diversion rate between the real value and the estimated value by MAPE value. The smaller the value is, the smaller the diversion rate is leading to better results.

(2) RMSE (root mean squared error)

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (6)$$

The smaller the value is, the smaller diversion rate is, yielding better results.

(3) r^2 (coefficient of determination)

$$r^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad (7)$$

where $\overline{y_i}$ is the average of the actual data.

After the aforesaid learning process, the last step is to validate SVR. Take a new injection product design specifications as the test example to learn the difference between the estimated results and the real quotations for further feasibility appraisal of this SVR. If the difference is acceptable to veteran business staff and designing staffs, then the hybrid model can work as the cost estimation model for the plastic injection product.

In this study, we compare multilayer perceptron networks (MLP), pure SVR and the proposed hybrid model.

The network structure and parameters of the MLP experiment are as follows:

- Input: 9 features
- Number of hidden layers: 1
- Number of hidden neurons: 10, 15, 20, 25, 30, 35, 40
- Learning rate: 0.1, 0.3, 0.5, 0.7, 0.9
- Momentum: 0.1, 0.3, 0.5, 0.7, 0.9

When the number of hidden neurons is 35 and the learning rate and momentum are 0.9, the best results of the MLP are obtained.

The inputs of SVR and the proposed model are also 9 and the parameters of the proposed model are optimized by GA whereas those of pure SVR is randomly set.

The more detailed learning processes of the proposed hybrid model are described as below. Before constructing the training cost dataset, the primary cost data should be preprocessed. The experimental data including the training cost data and test cost data are normalized, which can improve the generalization ability of AI prediction techniques. The training cost datasets are utilized to construct training sample sets according to the dimension of the input vector. The k-fold cross validation technique is used in the cost dataset of the plastic injection products. In this paper, the value of k is set to 5. Hence, the training data is divided into 5 subsets, with each subset of the data sharing the same proportion of each subset of data. Four subsets are employed in the training process, while the last one is used in the testing process. The $\mathrm{MAPE}_{\mathrm{CV}}$ in this classification is calculated by averaging the individual error estimation for each run of testing. The advantages of cross validation are that all test subsets are independent and this leads to the robust prediction results. While improvement in the training errors occurs, two parameters of the SVR (i.e., C and σ) are optimized by the GA concurrently. The adjusted parameters possessing the smallest test $\mathrm{MAPE}_{\mathrm{CV}}$ value are selected as the most appropriate parameters in the case studied. Then, the test dataset is utilized to investigate the accuracy of the prediction error.

To set the parameters of the GA, the parameters in the proposed model for the real dataset are experimentally set. The population size is 10 and the number of iterations is fixed as 100. The searching ranges for C and σ are as follows: $C \in [0, 106]$ and $\sigma \in [0.001, 3000]$.

The comparative results between the MLP, pure SVR and the proposed hybrid model are summarized in <Table 4> and the proposed hybrid model outperforms the MLP and pure SVR.

<Table 4> The comparison results

Appraisal indictor	MLP	Pure SVR	The proposed hybrid model
MAPE	0.4957	0.4983	0.4592
RMSE	623.67	645.39	612.95
r2	0.9984	0.9982	0.9996

The main merits of the proposed model are as follows: First, the model implements the structural risk minimization principle which attempts to minimize an upper bound of the generalization error rather than minimize the training error; this critical inherent feature of the SVR leads to a better prediction error than that of other conventional neural network-based techniques. Second, the conventional neural network-based techniques may not reach global solutions. However, in the proposed model the process for training is equivalent to solving a linearly constrained quadratic programming based on the 5-fold cross validation technique. In addition, the solution of the model is near optimal by using the GA.

4. Conclusion

In a competitive environment, the new product shall be introduced to the market earlier to win customers, making the company enjoy competition advantages with more profits, market shares and the image as an industrial leader. Therefore, this study proposed a GA and SVR based hybrid cost estimation model for feature-based plastic injection products. In recent years, support vector regression (SVR), a novel neural network technique based on SVM, has been compared with conventional techniques such as ANNs and nonlinear regression. The SVR technique demonstrated great potential and high performance results. However, to build reliable and robust intelligent prediction models, the parameters of the SVR should be carefully set. This paper focused on the improvement of the SVR model by using the technique of evolutionary computation and k-fold cross validation. Therefore, a hybrid AI modeling approach was proposed for predicting the cost data of plastic injected products. The main advantage of the proposed model is that, unlike conventional neural networks, it is computationally efficient because the SVR training only requires the solution of a set of linear equations. A k-fold CV is utilized to train the SVR leading to reliable and stable results. Moreover, the selection of the SVR parameters is optimized by applying the GA concurrently.

The cost estimation model based on the proposed hybrid approach was used to estimate plastic product quotations and makes it possible for inexperienced engineers to reduce professional judgments of product model design.

For further research, optimizing other kernel functions can be suggested. In addition, to verify and extend the proposed model, results provided by public datasets and real-world applications may be tested in the future.

5. References

[1] Cavalieria, S., Maccarroneb, Pinto, P. R. “Parametric vs. neural network models for the estimation of production costs: A case study in the automotive industry”, International journal of Production Economics, Vol. 91(2) (2004): 165 - 177

[2] Chen, K. -Y., Wang, C. -H., "Support vector regression with genetic algorithms in forecasting tourism demand", Tourism Manage. Vol. 28 (2007): 215 - 226

[3] Cortes, C., Vapnik, V., "Support vector networks", Mach. Learn., Vol. 20(3): 273 - .297 (1995).

[4] Goldberg, D. E., Genetic algorithms in search, optimization, and machine learning. Reading, MA: Addison-Wesley, (1989)

[5] Hsu, C. W., Chang, C .C., Lin, C. J., A practical guide to support vector classification, Technical Report, Department of Computer Science, National Taiwan University, Taipei, Taiwan, (2003)

[6] Kim, W. C., Mauborgne, R. Blue ocean strategy: How to Create Uncontested Market Space and Make the Competition Irrelevant, United States of America: Harvard Business School Publishing Corporation, (2005)

[7] Salgado, D. R., Alonso, F. J., "An approach based on current and sound signals for in-process tool wear monitoring", Int. J. Mach. Tool. Manufact. Vol. 47 (2007): 2140 - 2152

[8] Seo, K. -K., Park, J. -H., Jang, D. -S., Wallace, D., "Approximate estimation of the product life cycle cost using artificial neural networks in conceptual design", Int. J. Adv. Manuf. Technol.

Vol. 19 (2002): 461 - 471

[9] Shehab, E. M., Abdalla, H. S., "Manufacturing Cost Modeling for Concurrent Product Development", Robotics & Computer-Integrated Manufacturing, Vol. 17(4) (2001): 341-353

[10] Vapnik, V., The Nature of Statistical Learning Theory, Springer, London, (1999)

[11] Wang, H. S., Che, Z. H., "A multi-phase model for product part change problems", International Journal of Production Research, Vol. 46(10) (2008): 2797-2825

[12] Zhang, Y. F., Fuh, J. Y. H., Chan, W. T., "Feature-based cost estimation for packaging products using neural networks", Computers in Industry, Vol. 32(1) (1996): 95-113

Author

Kwang-Kyu Seo

Kwang-Kyu Seo is a professor of Management Engineering at Sangmyung University. Dr. Seo received a Ph.D. degree in industrial engineering from Korea University. He is interested in production/operation management, Information system and so on.

Address: Dept. of Management Engineering, Sangmyung University, 31 Sangmyungdae-gil, Dongnamgu, Chonan, Chungnam 330-720, Korea