Implementing support vector regression with differential evolution to forecast motherboard shipments

Fu-Kwun Wang [a], Timon Du [b,*]

[a] Department of Industrial Management, National Taiwan University of Science and Technology, Taiwan
[b] Department of Decision Sciences and Managerial Economics, The Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

ABSTRACT

In this study, we investigate the forecasting accuracy of motherboard shipments from Taiwan manufacturers. A generalized Bass diffusion model with external variables can provide better forecasting performance. We present a hybrid particle swarm optimization (HPSO) algorithm to improve the parameter estimates of the generalized Bass diffusion model. A support vector regression (SVR) model was recently used successfully to solve forecasting problems. We propose an SVR model with a differential evolution (DE) algorithm to improve forecasting accuracy. We compare our proposed model with the Bass diffusion and generalized Bass diffusion models. The SVR model with a DE algorithm outperforms the other models on both model fit and forecasting accuracy.

1. Introduction

Taiwanese motherboard manufacturers create 98.5% of the world's desktop motherboards and dominate the global desktop motherboard market (Market Intelligence Center (MIC), 2012). However, this industry's growth rate has slowed due to the trend of replacing desktops with laptops and tablets. In addition, aggressive pricing by laptop/tablet manufacturers has diminished desktop motherboard sales. Forecasting plays an important role in many business activities, such as the volume of demand in order and inventory management, production planning in manufacturing processes, capacity usage in production management, and the diffusion patterns of new products and technological innovations. The market is changing rapidly and a new forecasting model is required. Forecasted results can assist manufacturers in making better decisions on future expansion and investment.

In recent years, the Bass diffusion model (Bass, 1969) has been used successfully to describe the empirical adoption curve for many new products and technological innovations. This model provides good predictions on the timing and magnitude of the sales peaks of the products to which it is applied. Bass, Krishnan, and Jain (1994) proposed a generalized Bass model that included marketing mix variables (e.g., price and advertising variables). This generalized model can produce the best model fit and forecasting performance. Bass (1969) used the ordinary least squares (OLS) method to estimate the parameters of the Bass diffusion model. However, the OLS approach has a bias when estimating continuous

time models. In contrast, Schmittlein and Mahajan (1982) proposed the maximum likelihood estimation (MLE) method to improve the estimation. However, the maximum likelihood formulation considers only the sampling error and ignores all other sources of error, hence the computed standard error estimates may be too optimistic. Many researchers have tried to improve the problem. For example, Srinivasan and Mason (1986) applied the nonlinear least square (NLS) method to obtain valid error estimates. Venkatesan and Kumar (2002) presented genetic algorithms (GAs) to estimate the parameters of the Bass diffusion model. The parameter estimates they obtained from the GAs were consistent with the NLS method. Wang, Chang, and Hsiao (2013) proposed an evolutionary approach based on a GA/particle swarm optimization (PSO) hybrid to obtain the parameter estimates of the modified Bass model. This hybrid evolutionary approach has been successfully applied to real-world engineering design problems (Nagi, Yap, Nagi, Tiong, & Ahmed, 2011; Niu, Liu, & Wu, 2010).

The support vector machine (SVM) developed by Vapnik (1998) is based on statistical learning theory. SVMs have been widely applied in the fields of pattern recognition, bioinformatics, and other artificial-intelligence-related applications. SVMs have also been used to solve nonlinear regression estimation problems, a process known as support vector regression (SVR). SVR models have been used successfully to solve forecasting problems (Cao, 2003; Che, Wang, & Tang, 2012; Chou, Cheng, & Wu, 2013; García, García Villalba, & Portela, 2012; Hong & Pai, 2007; Huang, 2012; Huang, Bo, & Wang, 2011; Jiang & He, 2012; Khashei & Bijari, 2012; Pai & Lin, 2005; Štěpnička, Cortez, Donate, & Štěpničková, 2013). Empirical results have indicated that the selection of the three parameters, including C, ε, and γ, in an SVR model significantly influences its forecasting accuracy. SVRs with evolutionary

* Corresponding author. Tel.: +852 26098569.
E-mail address: timon@cuhk.edu.hk (T. Du).

algorithms (e.g., GA, simulated annealing, PSO, chaos-based PSO, chaos-based firefly, and hybrid) are used to determine appropriate parameter values (Hong, 2009; Kazem, Sharifi, Hussain, Saberi, & Hussain, 2013; Wu, 2010; Wu & Law, 2011).

We implement an SVR model with a differential evolution (DE) algorithm (Price, Storn, & Lampinen, 2006; Storn & Price, 1997) to improve the forecasting performance of motherboard shipments. In addition, we use a hybrid evolutionary algorithm that combines PSO with a quasi-Newton method to improve the parameter estimates of the generalized Bass diffusion model. In the following section, we present the forecasting models. Section 3 explains the SVR model with the DE and hybrid PSO (HPSO) algorithms. Section 4 denotes our use of data on motherboard shipments from Taiwanese firms to demonstrate the application of our proposed forecasting model. Finally, we offer a conclusion and suggestions for future studies in Section 5.

2. Forecasting models

2.1. Bass diffusion and generalized Bass diffusion models

The Bass diffusion model (Bass, 1969) is defined as

$$n(t) = m[F(t) - F(t-1)] + \varepsilon, \tag{1}$$

where $n(t)$ = the sales at time t, m = the number of eventual adopters, $F(t)$ = the cumulative distribution of adoptions at time $t = (1 - e^{-(p+q)t}) / (1 + (\frac{q}{p}) \times e^{-(p+q)t})$, p = the innovation coefficient, q = the imitation coefficient, and ε = the normally distributed random error term with a mean of zero and a variance of σ^2.

The adopter's probability density function $f(t)$ for adoption at time t is derived by

$$f(t) = \left(\frac{(p+q)^2}{p} \times e^{-(p+q)t} \right) \Big/ \left(1 + \left(\frac{q}{p} \right) \times e^{-(p+q)t} \right)^2. \tag{2}$$

Finally, some quantities such as peak sales, peak sales times, sales period inflection points, and forecasts of future sales are of interest in practical applications. The peak sales times can be obtained by differentiating Eq. (2) with respect to t, i.e., $T^* = \ln(q/p)/(p+q)$. The peak sales rate is obtained as $n(T^*) = m \times \frac{(p+q)^2}{4q}$. The inflection point for each sales period can be obtained by differentiating Eq. (2) twice with respect to t and solving for t, which yields $T^*_{left} = \frac{2-\sqrt{3}}{p+q} \times \ln(q/p)$ and $T^*_{right} = \frac{2+\sqrt{3}}{p+q} \times \ln(q/p)$.

However, the Bass diffusion model cannot consider external variables that can affect diffusion. A generalized Bass diffusion model (Bass et al., 1994) was developed to overcome the limitations of the Bass diffusion model. In the generalized Bass model, the mapping function $x(t)$, which describes the current effect of the decision variables on the conditional probability of adoptions at time t, is added to Eq. (1):

$$n(t) = m[F(t) - F(t-1)]x(t) + \varepsilon, \tag{3}$$

where $x(t) = 1 + \beta v(t)/v'(t)$ represents the pricing effect, $v(t)$ and $v'(t)$ represent the absolute price and the rate of price change, respectively, and β reflects the sensitivity to the price change (Bass et al., 1994). In addition, $x(t) = 1 + \beta v(t)$ is found in Jun and Park (1999). The generalized Bass diffusion model has been used to study optimal pricing and advertising policies for single-generation products (Krishnan & Jain, 2000).

2.2. Support vector regression

The SVM is a popular machine learning method of classification, regression, and other learning tasks (Vapnik, 1998). We consider a set of training points, $\{(x_1, y_1), \ldots, (x_l, y_l)\}$, where $x_i \in R^n$ is a feature vector and $y_i \in R$ is the target output. There theoretically exists a linear function to formulate the nonlinear relationship between input and output data. An SVR function is defined as

$$f(x) = w^T \phi(x) + b, \tag{4}$$

where $f(x)$ denotes the forecasting values, $\phi(\cdot)$ is a nonlinear mapping function, and the coefficients w ($w \in R^n$) and b ($b \in R$) are adjustable. Under the given parameters $C > 0$ and $\varepsilon > 0$, the standard form of SVR (Vapnik, 1998) is defined as

$$Min_{w,b,\xi,\xi^*} \frac{1}{2} w^T w + C \sum_{i=1}^{l} (\xi_i + \xi_i^*), \tag{5}$$

with the constraints

$$\begin{aligned} w^T \phi(x_i) + b - y_i &\leq \varepsilon + \xi_i, \\ y_i - w^T \phi(x_i) - b_i &\leq \varepsilon + \xi_i^*, \text{and} \\ \xi_i, \xi_i^* &\geq 0, i = 1, 2, \ldots, l, \end{aligned} \tag{6}$$

where ξ_i^* denotes training errors above ε and ξ_i denotes training errors below $-\varepsilon$.

After the quadratic optimization with inequality constraints is solved, the parameter vector w in Eq. (4) is calculated as

$$w = \sum_{i=1}^{l} (\beta_i^* - \beta_i) \phi(x_i), \tag{7}$$

where β_i^* and β_i are obtained by solving a quadratic program and the Lagrange multipliers. Finally, the SVR function is calculated as

$$f(x) = \sum_{i=1}^{l} (\beta_i^* - \beta_i) K(x_i, x_j) + b, \tag{8}$$

where $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ is the Gaussian radial basis function (RBF) kernel function. Chang and Lin (2011) suggested trying small and large values for C, e.g., 1–1,000, before deciding which are better for the data through cross validation, and finally trying several γs for the better Cs. However, better results are obtainable by practitioner experience.

3. The proposed model and HPSO algorithm

3.1. SVR-DE model

DE is a search heuristic that was introduced by Storn and Price (1997). It has been successfully applied in a wide variety of fields, from computational physics to operations research (Price et al., 2006). DE belongs to the class of genetic algorithms that use the biology-inspired operations of crossover, mutation, and selection on a population to minimize an objective function over the course of successive generations (Mitchell, 1998). DE uses floating-point instead of bit-string encoding on population members, and arithmetic instead of logical operations in mutation. It has several advantages such as its simple structure, ease of use, speed, and robustness (Storn & Price, 1997). Therefore, a DE algorithm can be used to find the best hyperparameters for SVR.

The DE procedure is summarized as follows (Ardia, Boudt, Carl, Mullen, & Peterson, 2011; Mullen, Ardia, Gil, Windover, & Cline, 2011). The variable NP represents the number of parameter vectors in the population. At generation 0, NP guesses the optimal parameter value, and vectors are made using random values between the lower and upper bounds. Each generation involves the creation of a new population from the current population members $x_{i,g}$, where i indexes the vectors and g indexes the generation. This is accomplished using a differential mutation of the population members. A trial mutant parameter vector $v_{i,g}$ is created by choosing three