World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:9, No:5, 2015

Forecasting of Grape Juice Flavor by Using Support Vector Regression

Ren-Jieh Kuo, Chun-Shou Huang

***Abstract*—**The research of juice flavor forecasting has become more important in China. Due to the fast economic growth in China, many different kinds of juices have been introduced to the market. If a beverage company can understand their customers' preference well, the juice can be served more attractive. Thus, this study intends to introducing the basic theory and computing process of grapes juice flavor forecasting based on support vector regression (SVR). Applying SVR, BPN, and LR to forecast the flavor of grapes juice in real data shows that SVR is more suitable and effective at predicting performance.

***Keywords*—**Flavor forecasting, artificial neural networks, support vector regression, grape juice flavor.

I. Introduction

IN today's economy, organizations face a very competitive environment and have to employ a number of strategies to gain competitive advantage. Effective decisions should be made to manage the process of production, distribution and consumption. Current decision development utilizes a variety of forecasting techniques to analyze the nature of juice flavor. Therefore, the study of juice flavor forecasting is of critical importance.

Recently, juice beverage market shows a rapid growth, with the rise of health conscious. Providing healthy and delicious drinks such as vitamin and mineral rich drinks now become the core to attract consumers. Thus, the companies put more spending and material resources in the research and development (R&D) to find consumers' favorite juice flavor. Establishing new business opportunities without considering consumers' demands will cause many problems, such as loss of goodwill, decrease in profit, customer satisfaction and loyalty. However, if company misses the opportunities to understand customer preference, it will waste human, financial, and material resources. The information on customers' favorite flavor can be used to develop predictive models, and to reduce R&D cost. In addition, the customer favorite flavor forecast also may affect the company's marketing strategy. Since market and consumer tastes change at any time. Thus, grapes juice flavor forecasting is essential to successful, R&D, marketing and policy making. In the previous years, there weren't many companies using certain forecasting methods to predict customer juice flavor. But now, due to the development of forecasting methods, lots of companies, including beverage companies, start to use certain methods to predict customer favorite juice flavor. By employing a forecasting method, it is easier for the companies to predict their customer demands, but sometimes it is difficult to guarantee that the forecasting result is accurate due to noise in the past data. Therefore, grade juice flavor forecasting is regarded as one of the most challenging tasks.

They demonstrated that different ANNs can be used to tackle diverse business-related problems [1]. They surveyed more than 100 related published articles that focus on neural and neuro-fuzzy techniques applied to forecasting stock markets [2]. They found that neural network is a widely accepted tool for studying and evaluating behaviors in the stock market. Even if the ANN provides a great deal of promise, the algorithm also suffers from a number of shortcomings such as number of parameters, difficult to obtain a stable solution, and the risk of model over-fitting [3]-[6].

Support vector regression (SVR) -developed through statistical learning theory- has been widely applied in the nonlinear regression estimation. Basically, SVR adopts the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error [7]. Structural risk minimization principle allows researchers to obtain better generalization from limited-size datasets. SVR has been success-fully applied in different problems of time series prediction such as demand forecasting [8]-[10], traffic flow forecasting [11]-[13], and financial time series forecasting [3]-[6], [14]-[16].

Additionally, few studies focus on forecasting in the juice industry, especially for grapes juice flavor. Therefore, this study intends to develop a SVR model that improves the forecasting accuracy of grapes juice flavor and return on investment. Juice flavor data contain correlations between forecasting variables. Grapes juice flavor may be affected by some underlying factors like juice components. This study focuses on grapes juice flavor forecasting for beverage manufacturers and retailers in juice drinks industry.

The rest of this paper is organized as follows. Section II gives a brief introduction to the SVR and BPN, while the experimental results and related discussions are presented in Section III. Finally, the concluding remarks are made in Section IV.

II. Methodology

A. Support Vector Regression (SVR)

SVR is a novel neural network algorithm technique based on statistical learning theory that has received increasing attention for solving nonlinear regression estimation problems. SVR is

R.J. Kuo is the Dean of College of Management, National Taipei University of Technology, Taiwan, ROC.

C.S. Huang is a PhD candidate at Department of Industrial Management, National Taiwan University of Science and Technology, Taiwan (e-mail: cshcsh1107@gmail.com).

derived from the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error [7]. It has been successfully applied in different time series prediction problems such as production value forecasting, traffic flow prediction, and financial time series forecasting [5], [14], [17], [18]. The SVR model can be expressed as (1) [7]:

$$y = w\phi(x) + b \tag{1}$$

where w is a weight vector, b is bias, and $\psi(x)$ is a kernel function in which a non-linear function is used to transform the non-linear input into linear mode in high-dimension feature space. Under traditional regression analysis, coefficients are found by minimizing the square error, which can be considered empirical risk based on a loss function. Vapnik [7] introduced the so-called ε-insensitivity loss function to SVR. It can be expressed as:

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

where ε is defined as the region of ε-insensitivity; when the predicted value falls into the band area, the loss is zero. In contrast, if the predicted value falls outside the band area, then the loss is equal to the difference between the predicted value and the margin. When empirical risk and structure risk are considered together, the SVR model can be constructed to minimize the following quadratic programming problem.

$$minimize: R(w, \xi, \xi^*) = \frac{1}{2}\| w \|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

subject to:

$$\begin{cases} y_i - w\phi(x_i) - b \leq \varepsilon + \xi_i \\ w\phi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0\,; \quad i = 1, ..., n \end{cases} \tag{3}$$

where i=1,2,…,n is the number of training data; (ξ_i+ξ_i*) is the empirical risk; $1/2\|w\|^2$ is the structure risk preventing over-learning and lack of applied universality; and C is a modifying coefficient representing the trade-off between empirical risk and structure risk. Equation (3) is a standard quadratic programming problem. After selecting an appropriate modifying coefficient (C), band area width (ε), and kernel function (φ), the optimum value of each parameter can be resolved though Lagrange functions. The general form of the SVR-based regression function can be written as [7]:

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x_j) + b \tag{4}$$

where α_i and α_i* are Lagrangian multipliers and satisfy the equality $\alpha_i\alpha_i$*=0; φ (x_i, x_j) is the kernel function. Any function that meets Mercer's condition can be used as the kernel function. According to Cherkassky and Ma's [19] proposal, radial basis function (RBF) is suitable for solving most forecasting problems. SVR performance is mainly affected by the setting of parameters C and ε [19]. There are no general rules governing the choice of C and ε. They proposed an analytic parameter selection method for SVR modeling that is based on sketching the structure of training data and using a trial-and-error approach to determine the best parameter values [14]. This study employs this analytic method to determine the best set of C and ε parameter values.

B. Back-Propagation Neural Network (BPN)

Artificial neural network (ANN) is a system derived through models of neurophysiology. In general, it consists of a collection of simple nonlinear computing elements whose inputs and outputs are tied together to form a network. The learning algorithms of ANNs can generally be divided into three different types: supervised, unsupervised, and hybrid learning. ANNs are nonparametric data driven approaches which can capture nonlinear data structures without prior assumption about the underlying relationship in a particular problem. ANNs are more general and flexible modeling and analysis tools for forecasting applications in that not only can they find nonlinear structures, they also can model linear processes [16].

BPN is the most popular neural network training algorithm, which has a simple architecture but powerful problem-solving ability [2], [20]. The neural network architecture is organized into nodes and the types of connections permitted. The nodes in the input layer receive input signals from an external source, while the nodes in the output layer provide the target output signals. Any layers between the input and output layers are called hidden layers. Since one hidden layer network is sufficient to model any complex system with the desired degree of accuracy [21], the BPN model designed in this study will have only one hidden layer. The neurons obtain inputs from initial inputs or interconnections and generate outputs using a non-linear transfer function.

BPN uses the gradient descent training algorithm to minimize error (the difference between the desired output and the network output) and adjusts interconnection weights during the training process. For the gradient descent algorithm, the step size (otherwise known as the learning rate) must be specified first. The learning rate is a crucial aspect of the BPN model because lower learning rates tend to slow down the learning process before convergence, while higher learning rates may cause network oscillation and an inability to converge. BPN performance is mainly affected by the setting of the network topology, i.e., the number of nodes in each layer and the learning rates. There are no general rules governing the choice of network topology. Its selection is usually based on the trial-and-error (or cross-validation) method. In this study, the optimal network topology of the BPN model is determined using the trial-and-error method. Chauvin and Rumelhart [21], and Haykin [22] descripted more details about how to determine the appropriate network topology (the number of layers, the number of nodes in each layer, and the appropriate learning rates).

III. Experimental Results

A. Dataset and Forecasting Variables

The data collection period for juice amounts is from August 2010 to January 2011, giving a total of 21,810 grapes juice flavor about the degree of taste sample in the entire China market. We used 5 cross validation analysis. The 17,448 data points (about 80% of the total number of sample points) are used as the training sample, while the remaining 4,362 data points (about 20% of the total number of sample points) are employed as the holdout and used as the testing sample for measuring out-of sample forecasting ability. To build the forecasting models, the eight indicators depicted in Table I, which are chosen on the basis of subjective judgment, expert opinions and industry knowledge are employed as forecasting variables. Note that Brix, Acid and PH value indicators are also adopted in this study since they are most widely used measures in grapes juice flavor forecasting. The variables X1 to X8 are input variables and the variable Y is the output variable.

TABLE I
List of Forecasting Variables in Building Forecasting Models

Variables	Description
Y	Degree of all flavor
X1	Degree of fragrance flavor
X2	Degree of color flavor
X3	Degree of sweet flavor
X4	Degree of sour flavor
X5	Degree of fruit flavor
X6	Brix
X7	Acid
X8	PH value

This study utilizes four commonly used performance criteria to evaluate the accuracy of the three forecasting models: root mean squared error (RMSE), mean absolute deviation (MAD), and mean absolute percentage error (MAPE). The definitions of these criteria can be found in TABLE II. RMSE, MAD, and MAPE measure the deviation of the forecast value from the actual value. Hence, the smaller the indicated value is, the greater the accuracy.

TABLE II
Performance Measures and Their Definitions

Metrics	Calculation [a]
RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(T_i - A_i)^2}{N}}$
MAD	$MAD = \frac{\sum_{i=1}^{N}\lvert T_i - A_i \rvert}{N}$
MAPE	$MAPE = \frac{\sum_{i=1}^{N}\left\lvert \frac{T_i - A_i}{T_i} \right\rvert}{N} \times 100\%$

a: Note that T and represent the actual and predicted value, respectively; n is the total number of data.

The input variable data employed in this study include a variety of measurements reflecting a range of units, the magnitude of the absolute value can vary significantly for different variables. If the original values of input variables are applied directly in modeling neural network models, including SVR, BPN and LR specifications, the large value input variables may overwhelm smaller valuable inputs in the neural network learning process, leading to the neural network learning saturation problem.

Saturation means that continued learning does not lead to improved network performance, and is often indicative of over-fitting and detrimental to generalization [23], [24]. Moreover, many studies have reported that linear scaling can improve the performance of neural networks [20], [25]-[27]. Therefore, to avoid the saturation problem and produce accurate forecasts, the raw input variable data employed in this study are linear-scaled into the range of [-1.0, 1.0] before the neural network models are constructed.

B. Forecasting Results

To compare the results derived from three forecasting models in the frameworks, this section discusses the forecasting models constructed using the SVR, BPN, and LR methodologies to predict total juice flavor amounts.

The parameters setting in both algorithms are important. In the BPN model, the input layer has eight nodes as the eight forecasting variables are all adopted. As there are no general rules for determining the appropriate number of nodes in the hidden layer, the number of hidden nodes to be tested is set at 14, 15, 16, 17, and 18. The network has only one output node: The forecast juice flavor amount. As lower learning rates tend to give the best network results [21], learning rates of 0.001, 0.005, 0.01, and 0.1 are tested during the training process. The network topology with the minimum testing RMSE is considered the optimal network. Table III reports the best results of testing all models from cross-validation. From the forecasting result, BPN can be observed that the {8-17-1} topology {8-17-1} indicates there are 8 nodes in the input layer, 17 nodes in the hidden layer, and one node in the output layer with a learning rate of 0.01 gives the minimum testing RMSE and it is the best topology setup for the BPN model.

TABLE III
The Parameter Selection Results of the All Models from Cross-Validation

Methods	Training RMSE	Testing RMSE
SVR($\varepsilon=2^{-7}$, $C=2^{9}$)	0.5454	0.4227
BPN{8-17-1} learning rate=0.01	0.5388	0.5341
LR	0.8648	0.5766

In constructing the SVR model, all of the predictor variables are used. Table III shows the best results of testing the SVR model with different combinations of parameter sets. It can be seen that the parameter set ($\varepsilon=2^{-7}$, $C=2^{9}$) gives the best result (minimum testing RMSE) and it is the best parameter setup for the SVR model.

For comparing the forecasting models performance in K-fold cross validation. In the experiment, the data are separated in 5 partitions. Then, there are 80% data for training, and 20% for testing. The results are shown in Table IV. Performance is

evaluated using MAD, RMSE, and MAPE. Compared from SVR, BPN and LR of grapes juice flavor forecasting testing data, MAD average is 0.3327, RMSE average is 0.4577 and MAPE average is 0.0946 from SVR model are the best. It indicates that SVR model has the best performance than BPN and LR model.

TABLE IV
THE COMPARISON OF 5-CROSS VALIDATION FORECAST AVERAGE RESULTS FROM THE THREE OF GRAPES JUICE FLAVOR FORECASTING MODEL

Methods	Fold	MAD	RMSE	MAPE
SVR	1	0.3104	0.4381	0.0894
	2	0.3313	0.4393	0.0943
	3	0.3950	0.5478	0.1101
	4	0.3132	0.4227	0.0890
	5	0.3136	0.4408	0.0903
	Avg.	0.3327	0.4577	0.0946
BPN	1	0.3421	0.4511	0.1663
	2	0.3306	0.4651	0.0929
	3	0.3743	0.5462	0.1042
	4	0.3334	0.5341	0.1097
	5	0.3050	0.4384	0.0887
	Avg.	0.3710	0.4869	0.1123
LR	1	0.3130	0.4416	0.0900
	2	0.4674	0.6340	0.1307
	3	0.4174	0.5778	0.1187
	4	0.4205	0.5766	0.1185
	5	0.3167	0.4250	0.0898
	Avg.	0.3870	0.5310	0.1095

Based on the findings discussed above, it can be inferred that the SVR methodology is suitable for constructing forecasting models for beverage manufacturers and retailers due to its excellent variable screening and model interpretation credentials.

IV. CONCLUSIONS

Grapes juice flavor forecasting is crucial for business strategy, R&D, and customer service of beverage manufacturers and retailers. The forecasting results are obtained by comparing SVR, BPN and LR models. The experimental results show that the SVR model gives the lowest degree of prediction error and outperforms the other two comparison methods. The results also indicate that SVR represents a good alternative to grapes juice flavor forecasting method for beverage companies in China. Therefore, the grapes juice flavor which has been predicted by SVR model can help companies adjust their marketing strategy to attain better performance in the future.

V. REFERENCES

[1] A. Vellido, P. J. Lisboa, and J. Vaughan, "Neural networks in business: A survey of applications (1992–1998)," *Expert Systems with Applications,* vol. 17, no. 1, 51-70, 1999.

[2] G. S. Atsalakis, and K. P. Valavanis, "Surveying stock market forecasting techniques–Part II: Soft computing methods," *Expert Systems with Applications*, vol. 36, no. 3, 5932–5941, 2009.

[3] L. Cao, "Support vector machines experts for time series forecasting," *Neurocomputing*, 51, 321-339, Apr. 2003

[4] L. Cao, and F. E. Tay, "Financial forecasting using support vector machines," *Neural Computing & Applications*, 10(2), 184-192, 2001.

[5] L.J. Cao, and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *Neural Networks, IEEE Transactions on*, vol. 14, no. 6, 1506-1518, 2003.

[6] F. E. Tay, and L. Cao, "Application of support vector machines in financial time series forecasting," *Omega*, vol. 29, no. 4, 309-317, 2001.

[7] V. N. Vapnik, V. N. "An overview of statistical learning theory," *Neural Networks, IEEE Transactions on*, vol. 10, no. 5, 988-999, 1999.

[8] C. J. Lu, and Y. W. Wang, "Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting," *International Journal of Production Economics*, vol. 128, no. 2, 603-613, 2010.

[9] K. Y. Chen, and C. H. Wang, "Support vector regression with genetic algorithms in forecasting tourism demand," *Tourism Management*, vol. 28, no. 1, 215-226, 2007.

[10] E. E. Elattar, J. Goulermas, and Q. H. Wu, Q. H. "Electric load forecasting based on locally weighted support vector regression. Systems, Man, and Cybernetics," *Part C: Applications and Reviews, IEEE Transactions on*, vol. 40, no.4, 438-447, 2010.

[11] W. C Hong, "Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm," *Neurocomputing*, vol. 74, no. 12, 2096-2107, 2011.

[12] W. C. Hong, Y. Dong, F. Zheng, and C. Y. Lai, "Forecasting urban traffic flow by SVR with continuous ACO," *Applied Mathematical Modelling*, vol. 35, no. 3, 1282-1291, 2011.

[13] W. C. Hong, Y. Dong, F. Zheng, and S. Y. Wei, "Hybrid evolutionary algorithms in a SVR traffic flow forecasting model," *Applied Mathematics and Computation*, 217(15), 6733-6747, 2011.

[14] C. J. Lu, T. S. Lee, and C. C. Chiu, "Financial time series forecasting using independent component analysis and support vector regression," *Decision Support Systems*, vol. 47, no. 2, 115-125,2009.

[15] K. J. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no.1, 307-319,2003.

[16] X. Liang, H. Zhang, J. Xiao, and Y. Chen, "Improving option price forecasts with neural networks and support vector regressions.," *Neurocomputing*, vol. 72, no. 13, 3055-3065, 2009.

[17] M. Castro-Neto, Y. S. Jeong, M. K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert Systems with Applications*, vol. 36, no. 3, 6164-6173, 2003.

[18] P. F. Pai, S. L. Yang, and P. T. Chang, "Forecasting output of integrated circuit industry by support vector regression models with marriage honey-bees optimization algorithms," *Expert Systems with Applications*, vol. 36, no. 7, 10746-10751, 2009.

[19] V. Cherkassky, and Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks*, vol. 17, no. 1, 113-126, 2004.

[20] G. Zhang, B. Eddy Patuwo, B., & Y. Hu, "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, vol. 14, no. 1, 35-62, 1998.

[21] Y. Chauvin, and D. E. Rumelhart, "*Back-propagation: Theory, architectures, and applications*," NJ: Lawrence Erlbaum, Hillsdale, 1995, pp. 1-35.

[22] S. Haykin, "*Neural networks: A comprehensive foundation*," Prentice Hall PTR upper Saddle River, 1994.

[23] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network," *Information Theory, IEEE Transactions on*, vol. 44, no. 2, 525-536, 1998

[24] Y. Lee, S. H. Oh, M. W. Kim, "An analysis of premature saturation in back propagation learning," *Neural networks*, vol. 6, no. 5, 719-728,1993.

[25] S. F. Crone, J. Guajardo, and R. Weber, "*The impact of preprocessing on support vector regression and neural networks in time series prediction*," Paper presented at the DMIN, 2006.

[26] S. F. Crone, S. Lessmann, and R. Stahlbock, "The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing," *European Journal of Operational Research*, vol. 173, no. 3, 781-800, 2006.

[27] Z. Tang, and P. A. Fishwick, "Feed-forward neural nets as models for time series forecasting," *ORSA Journal on Computing*, vol. 5, no. 4, 374-385,1993.