A demand forecast model using a combination of surrogate data analysis and optimal neural network approach

H.C.W. Lau a,⁎, G.T.S. Ho b, Yi Zhao b

a CInIS & School of Management, University of Western Sydney, Australia
b Department of Industrial and System Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

ABSTRACT

As rough or inaccurate estimation of demands is one of the main causes of the bullwhip effect harming the entire supply chain, we have developed a mathematical approach, the minimum description length (MDL), to determine the optimal artificial neural network (ANN) that can provide accurate demand forecasts. Two types of simulated customer and one practical demand are employed to validate the capability of the MDL method. Since stochastic factors hidden in the demand data disturb the prediction, the surrogate data method is proposed for identifying the characteristics of the demand data. This method excludes demands that are totally stochastic when forecasting. We demonstrate how optimal models estimated by MDL are consistent with the dynamics of demand data identified by the surrogate data method. The complementary approach of the surrogate data method and neural network constitutes a comprehensive framework for making various demand predictions. This framework is applicable to a wide variety of real-world data.

1. Introduction

The operation of the supply chain has undergone significant changes during the past decade [3]. Within the supply chain, enterprises usually adopt a strategy of having low inventories at all levels, including the end sale retailers, in order to reduce the their costs [22,27]. Meanwhile, the retailer sectors have to face uncertainties emerging in their supply chains. Customer demands depend on uncertain, stochastic factors, which make it difficult for supply chain participants to give an accurate estimation of future demands. This issue would be further extended by variation amplification, known as bullwhip effect, and make the parties involved getting lost in inventory management by receiving faulted notification. Obviously, the extra stock keeping results in excess production at the upstream levels, since the producer aims to fulfill the over-estimated demand. Likewise, underestimated demand causes upstream players not producing enough quantity fulfilling actual demand. Both scenarios lead to inefficiency in supply chain management. Thus, the challenge for a participant in the supply chain is to determine the appropriate quantity in terms of accurate demand forecast.

As the bullwhip effect has been recognized as a forecast-driven problem through supply chains, it is necessary to develop advanced techniques for forecasting customer demand and extend the visibility of customer demands as far as possible. However, doing an accurate and truthful demand forecasting is not straightforward and its difficulty is that customer demands depend on many environmental factors while the pattern hidden in those demands is unknown or is too complicated for managers along supply chain to understand.

In this paper, an intelligent system based on the MDL-optimal neural network for "learning" the underlying pattern and predicting future demands is developed. Neural networks are considered as the primary and most popular technique for demand forecasting in supply chain management, and in particular the multi-layer feed-forward neural network is able to approximate any nonlinear or linear function under certain conditions [2,13,20]. Moreover, the high-degree of freedom in the neural network architecture provides the potential to model any function but, unfortunately, also results in a very high probability of overfitting [38]. So the crucial issue in developing a neural network is the generalization of the network; however, being ignored in published works. An alternative novel approach is taken to determine the optimal neural network considering generalization and defined as the MDL-optimal neural network for demand forecasting.

Furthermore, there is little in the literature that focuses on studying the nature of customer demand. Most articles about demand forecasting usually maintain that the given data should be predicted by using their approaches no matter whether the data is stochastic, linear or nonlinear and no matter whether the forecasting techniques are suitable for modeling the data or not. Actual demands depend on a lot of stochastic elements, which very probably result in the demands becoming completely stochastic. Meanwhile, some kinds of demand data in which deterministic patterns dominate appear to be random, and then people are very likely to ignore investigation of those demands. To address this issue, we exploit the surrogate data method and examine the dynamics of the specific simulated...
customer demands. We notice that the results of the surrogate data method are also confirmed by the MDL-optimal model. We, therefore, attempt to make use of the hypothesis testing with surrogate data method confirming the modeling technique so as to provide a comprehensive solution to problems of customer demand prediction in terms of identifying data feature, and selecting optimal network setting from one to another model. The contribution of the proposed techniques is also validated according to the performance in accuracy and flexibility in comparison with commonly used mathematical approaches in experiments.

The rest of the paper is organized as follows: Section 2 contains a brief review of techniques of demand forecasting. This is followed by a description of the developed optimal neural network, as well as the surrogate data method for identifying data characteristics. In Section 4 the application of proposed technique is described and benchmarked with existing approaches through simulated models and a practical study. In the last section a conclusion on the approaches that have been developed is given and suggestions for future research are provided.

2. Literature review

Uncertainty in a supply chain can be defined as unpredictable events that affect its planned performance [24]. Demand uncertainty, caused by inaccurate forecasting, feeds into the information exchange network, which in turn results in the bullwhip effect [30]. Such an effect may decrease the ability of the chain to meet the expected target for the delivery of products and services.

The bullwhip effect, also known as the Forrester or whiplash effect, is one of the key areas of research in supply chain management (SCM) applications and its’ typical cases could be found in the commercial operation of Campbell’s Soup [15], HP and Proctor & Gamble [25], and a garment supply chain [12]. High inventory levels and poor customer services along the supply chain constitute symptoms of the bullwhip effect [9].

In terms of management science techniques, Yao, [39] Paik, and SeungKuk [31] identified demand forecasting as one of the significant variables for the bullwhip control, and Miyaoaka found that improved forecasting could reduce fluctuations in manufacturing production levels [29]. It is of vital importance as it has close relationship with reordering system [6], inventory cost [26], decision making [28], profit maximization [18], etc. Gurnani et al. [18] found that retailers have to strike a balance between an accurate demand forecast and unit cost uncertainty before selling season starts. Different situations require different forecast models or methods. Sani and Kingsman [34] compared five forecasting methods in terms of cost, service level and then both of them and found that they perform slightly different form each other. To forecast telephone demand in Australia and Chile, different techniques were also adopted by Bhattacharyya and Wellenius respectively [5]. Referring to above works, it can be concluded that optimal forecasting methods have to be designed or selected carefully [14].

To address this far-reaching factor, the moving average and the single exponential smoothing methods have been used by Graves [17] and Chen et al. [7,8]. Multiple correlations and a regression equation with weights can also be applied for optimal prediction [10]. It is a conventional technique in demand forecasting [33]. According to a comparative study, ANN can produce better predictions than with the Multiple Regression method [16].

As neural networks have been developed rapidly and are used widely in operations management [23,32,35]. Demand forecasting problems deploying this artificial intelligent technique are studied following classification, simulation, and decision making. Hill et al. [21] stated that neural networks can perform significantly well in forecasting tasks. Possible reasons for poor forecast of classical decomposition are indentified and Hansen and Nelson [19] concluded that use of neural networks is a way out. More studies in comparing and combining traditional and neural network based forecasting methodology suggest that neural network can offer some improvement in performance and its feasibility of cooperation [11]. In this paper, we focus on how MDL-optimal neural networks approach for tackling linear and nonlinear simulated customer demand forecasting.

3. Framework of the MDL-optimal neural network model

Simulation logic, along with a flowchart, is shown in Fig. 1. To verify that the program does actually perform as intended, the conceptual model is divided into three parts: demand generation, forecasting and calculation of prediction accuracy. The neural network used in this paper is the three-layer feedforward neural network, with a single hidden layer, sigmoid activation functions, and one linear output, as illustrated in Fig. 1. Given the input vector \([x_{t-1}, x_{t-2}, ..., x_{t-d}]\) the transfer function of the neural network, \(f\) can be mathematically expressed as

\[
f(x_{t-1}, x_{t-2}, ..., x_{t-d}) = b_0 + \sum_{i=1}^{k} v_i \phi \left( \sum_{j=1}^{d} o_{i,j} x_{t-j} + b_j \right)
\]

where \([v_i, o_{i,j}, b_j]\) are weights and biases respectively, \(\phi\) is the tan-sigmoid transfer function, \(k\) is the number of neurons, and \(d\) is the number of inputs. The Levenberg–Marquardt algorithm is used to train the neural network.

Overfitting has long been recognized as an endemic problem to neural networks having a number of parameters. The biological nature for ANN (i.e. a massive highly connected array of nonlinear excitatory “neurons”) promotes the construction of neural networks with a large number of neurons. Correspondingly the resulting models easily become overfitted. Therefore, the adequate generalization of ANN for a specific application is a primary element to ensure successful application in practice. We, therefore, utilize a novel approach, MDL, to directly determine the optimal neural network (i.e. the number of neurons in the neural network) with the focus on prediction accuracy.

3.1. The minimum description length

The MDL principle is based on the trade-off strategy to estimate the optimal model according to the minimum of total description length. The description length of a model is composed of two parts: the cost describing both the model parameters, and its prediction errors.

Definition 1. \(N(s)\) is the description length of parameters of a neural network, whose neuron number is \(s\).

Definition 2. \(E(s)\) is the description length of the prediction error of the same model.

The total description length with respect to this model, denoted by \(D(s)\), is then given by the sum of both parts. The MDL principle states that the optimal model is the one that minimize \(D(s)\).

Let \(\{d_i\}_{i=1}^{N}\) be customer demands of \(N\) time units (for example weeks) and \(f(d_{i-1}, d_{i-2}, ..., d_{i-k}; A)\) be a prediction of the neural network given of the previous \(k\) inputs and neural network parameters \(A\) with respect to \(s\) neurons. The prediction error at the \(i\)th time unit is then given by \(e_i = f(d_{i-1}, d_{i-2}, ..., d_{i-k}; A) - d_i\). So the description length of the neural network \(f(\cdot; A)\) is given by the description length of all the parameters \(A\) [40]:

\[
N(s) = L(A_s) = \sum_{j=1}^{s} \ln \frac{\gamma}{\delta_j}
\]
where γ is a constant. (δ₀, δ₁, ..., δₖ) are defined as the solution of
\[
Q \begin{bmatrix}
δ₀ \\
δ₁ \\
δ₂ \\
\vdots \\
δₖ
\end{bmatrix} = \frac{1}{δ_{j-1}}
\]
(3)

where Q is the second derivative of E(s) and (⋅)ₗ denotes the jth element of the vector (⋅) [40].

E(s) is the negative logarithm of the likelihood of the errors \( e = \{e_i\}_{i=1}^{N} \) under the given probability distribution of those errors. With the assumption that these errors follow the standard Gaussian distribution the description length of model prediction errors is approximated by [40]

\[
E(s) = \frac{N}{2} + \ln \left( \frac{2\pi}{N} \right)^{N/2} + \ln \left( \sum_{i=1}^{N} e_i^2 \right)^{N/2}.
\]
(4)

In this paper we further study the prediction of customer demands by using the related environmental processes. Since the end customer demand is related to many environmental factors the end customer demand can be regarded as an unknown function with respect to those factors. Given the time series of K environment factors at the jth time unit as the inputs, \( \{p_i\}_{i=1}^{K} \), the neural network gives its prediction, \( f(p_1, p_2, ..., p_{K}; A_i) \) and correspondingly the prediction error at the jth time unit \( e_i = f(p_1, p_2, ..., p_{K}; A_i) - d_i \). We then employ equations for M(s) and E(s) to compute the description length of the neural network in this scenario.

### 3.2. Surrogate data method

The surrogate data method was suggested and standardized by Theiler [37]. The rationale for surrogate data hypothesis testing is to generate an ensemble of surrogate data (surrogates in short) that preserve certain properties of the original data (i.e. consistent with some null hypotheses) [36].

There are three typical null hypotheses, NH0 (the data is random noise), NH1 (the data is linearly filtered noise) and NH2 (the data is a static monotonic nonlinear transformation of linearly filtered noise) [42]. Correspondingly, Algorithm 0, Algorithm 1, and Algorithm 2 produce surrogates that are consistent with these hypotheses. One then applies some test statistic to both the surrogates and the original data. If the test statistic value for the data is out of the distribution formed by values estimated for the surrogates, the given hypotheses is rejected as being the likely origin of the data. If the statistical value for the data is in the distribution formed from surrogates, this suggests that the data is consistent with the given hypotheses.
For example, we test the demand data against the hypothesis of NH1. That is, we wish to examine whether the demand data we are interested in originates from the linear process. If so, the linear model is more suitable to model such data. So the surrogate data method acts as a guide to the next modeling technique. We then notice whether the optimal neural network selected by the MDL is equivalent to a linear model so as to further validate the capability of our model selection technique. If the given data is consistent with the nonlinear process nonlinear modeling techniques, like neural networks should outperform the linear counterparts. However, nonlinear models are sensitive to data sets and more easily become overfitted than do linear models. We then observe whether the MDL-optimal model can avoid overfitting and can provide an accurate prediction. This is discussed at length in the next section.

4. Experimental results

First of all, the customers follow a linear demand process with seasonal swings. The customer demands in the simulation model are generated using the following formula [41],

\[ D_t = (\text{base}_t + \text{slope}_t) \times \left( \frac{\text{season}_t + \text{period}_{2\pi} + t}{\text{season}_t} \right)^2 + \text{noise} \times \text{rand}\_\text{normal}(\cdot). \]  

where \( D_t \) is the demand at the \( t \)th week, \( \text{rand}\_\text{normal}(\cdot) \) is a standard normal random generator between zero and one, \( \text{noise} \) represents the amplitude of the contaminated normal random noise, and \( \text{base}, \text{slope}, \text{period} \) and \( \text{season} \) are variable coefficients concerned with the time, a week.

The same form is also used by Bayraktar et al. [4] but they set \( \text{base}, \text{slope} \) and \( \text{season} \) with constants. The underlying tendency in their simulated data is easy to follow so we replace them with time-variant coefficients and add stochastic factors into the generation of these coefficients. The fluctuation of \( \text{base} \) and peaks of \( \text{slope} \) randomly appear, and the variable, \( \text{season} \), is added with strong random noise, all of which are tried to reflect the practical factors. Finally, the generated customer demand is more difficult to forecast.

Base, slope, season and period are specified by the simulated data, as shown in Fig. 2. These four data sets contaminated with observational noise simulate practical environmental factors related to the customer demand. Then the demand data per week is generated using the given formula above. In order to examine the suitability of MDL approach to forecast demand, some forecasting techniques, MSE-optimal neural network, exponential smoothing and multiple regressions, would be taken as comparison.

4.1. Identification of stochastic customer demands

Prior to the prediction of the simulated customer demands, it is necessary to determine whether the given data is predictable or not. It is not significant to make a prediction of customer demands which are consistent with stochastic noise. The surrogate data method with the hypotheses of NH0 and NH1 is used to test the dynamic property of the given demand data (i.e. predictability and linear property). The results are presented in Fig. 3.

100 surrogates of the original customer demand are generated, which are consistent with NH0 and NH1. One popular statistical criterion, complexity [42], is applied to the given simulated data. Green bars in Fig. 3 show the probability distribution of statistical values for all the surrogates; and the red star is the complexity of the original demand. Note that the complexity of the original time series is not shown in the top panel of Fig. 3, as its value is far away from the range of the surrogates’ results (i.e. 0.99–1.10).

Gaussian distribution is employed to fit the distribution of statistical values for surrogate data. With this known distribution, we can determine the confidence level of rejecting the given hypothesis. The central limit theorem proves that a large number of independent surrogates are distributed approximately normally. The normal distribution is appropriate for modeling the distribution of statistical values of surrogates as 100 surrogates are generated independently.

We observe that the complexity of the original demand data is even larger than the mean of the fitted Gaussian distribution plus its three-time standard deviation. It, therefore, suggests that the given customer demand is inconsistent with random noise, with almost 100% confidence probability. Furthermore, in the bottom
panel the statistical value of the original demand data is in the center of the distribution of the surrogates. This result indicates that the given customer demand is consistent with linear dynamics. Thus both conclusions indicate that the given data itself is not of a stochastic character and is suitable for prediction by linear modeling techniques.

4.2. Prediction of customer demands with historical demands

Various neural networks are employed with one to twenty neurons to forecast the demand data generated by the formula above, of which the first 400 points are selected to train the network, and the rest are the test data. Typical predictions obtained by the

![Fig. 3. Application of the surrogate data method with hypotheses of NH0 (top panel) and NH1 (bottom panel) to the simulated demand data.](image_url)

![Fig. 4. Prediction of various neural networks of the customer demands by using historical demand data.](image_url)
neural networks with one, five, seven, ten, sixteen, and nineteen neurons are listed in Fig. 4, where the red curves are the prediction and the blue curves are the original demand. The neural network with only one neuron predicts exactly the future customer demand. Note that some predicted values are negative, as shown in Fig. 4(c) and (d) since they are obtained automatically by the given neural network. The negative values indicate that the neural network has become overfitted. The description length curve of all twenty neural networks is shown in Fig. 5. The neural network candidate with one neuron minimizes the description length and thereby this neural network is the MDL-optimal model. The neural network with one neuron is equivalent to a linear model. It conforms to the previous conclusion taken by the surrogate data method that the linear model is suitable for predicting the given simulated demand.

Here the standard deviation of the test data is set as the threshold for assessing the accuracy of the prediction. It means that if the deviation between the original data and its prediction is lower than the threshold, this prediction is regarded as an accurate one; otherwise, it is an inaccurate prediction. Finally, a curve of the prediction accuracy for all twenty models is achieved, as shown in the bottom panel of Fig. 5. Again, the MDL-optimal model gives the most accurate prediction.

As a comparison, we illustrate the mean square errors of the training and test data, as presented in Fig. 6. The mean square errors of the training data keep decreasing while the mean square errors of the test data keep increasing, with fluctuations. The presence of mean square errors indicates that the neural networks with more neurons are inclined to become overfitted but it cannot estimate the optimal neural network. The mean square errors in the test data can be regarded as evidence of overfitting, rather than a criterion for model selection.

The performance among selected forecasting techniques is given in Table 1, with negligible mean-square errors for the four methods, it is found that neural network approach got slightly better prediction comparing to those of exponential smoothing and multiple regression with overall accuracy about 0.8.

4.3 Prediction of customer demands with environmental factors

The records of base, slope, season and period are fed to the neural network to give the corresponding demand data. Obviously, the generation of demand data is a simple linear function with respect to these parameters. Using the environmental factors related to the practical customer demand is the best way to forecast such demand.
Fig. 7 gives the description length curve of twenty neural networks and their prediction accuracy. The description length curve shows that for the same type of simulated demand data, the MDL-optimal neural network is also the neural network candidate with one neuron, which achieves the highest prediction accuracy. The curve expression in Fig. 7 and calculated mean square error also show that overfitting occurs in the network models with more than one neuron. Hence, the MDL-optimal model can provide an accurate prediction for the linear demand data in both scenarios.

Under the effect of environmental factors, computations, again, Table 2 shows that MDL-optimal neural network performed the best among the four specified demand forecasting methods. It is deserved to mention that prediction accuracy with environmental factors of MDL-optimal neural network is about 37% and 20% higher than those of exponential smoothing and multiple regressions respectively.

4.4. Prediction of nonlinear customer demands

We now introduce a chaotic time series, the Ikeda map, which appears to be periodic. The equation of the Ikeda map is given by

\[
\begin{align*}
\dot{x}_{n+1} &= 1 + \mu(x_n \cos t_n - y_n \sin t_n) \\
\dot{y}_{n+1} &= \mu(x_n \sin t_n + y_n \cos t_n)
\end{align*}
\]

where \(\mu = 0.7\) and \(t_n = 0.4 - 6/(1 + x_n^2 + y_n^2)\).

It is known that with the parameters above, the data generated is nonlinear. Certainly, other known nonlinear data can also be used. Here we choose the x-component data of the Ikeda map denoted by \(\text{Ikeda}_x\). We, therefore, simulate the nonlinear demand by using the formula, \(D_t = \text{base} + \text{slope} + \text{season} + \text{Ikeda}_x\).

The base, slope and season are the same as those of the previous function, and are contaminated with stochastic noise. The demand that is generated is also a linear function with respect to these four coefficients, but the data itself is nonlinear. The simulated nonlinear demand data is tested by the surrogate data method, as shown in Fig. 8. The dashed curves are a Gaussian distribution fitted to the

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Prediction accuracy of customer demands with historical demands.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MDL-optimal NN</td>
</tr>
<tr>
<td>MSE</td>
<td>(1.75 \times 10^{-3})</td>
</tr>
<tr>
<td>Prediction accuracy</td>
<td>0.88</td>
</tr>
</tbody>
</table>
distribution of two kinds of surrogates. In the figure, the statistical value of the original simulated demand data is 0.721, which is lower than the mean of complexity of surrogates minus their standard deviation. We, therefore, conclude that the given demand data is not random noise, and not suitable for linear models.

We also generate 500 points of this system, of which 400 points are selected as training data and the rest are used as testing data. Typical predictions obtained by neural networks with one, five, nine, thirteen, sixteen, and twenty neurons are presented in Fig. 9, where the denotation of curves is the same as that in Fig. 4. The description length of these twenty neural network candidates is plotted in the top panel of Fig. 10 and the bottom panel shows the accuracy of the predictions.

Referring to Fig. 10, the MDL-optimal neural network is the neural network candidate with five neurons, which achieves the highest prediction accuracy of all. From Fig. 9, we observe that the MDL-optimal neural network accurately predicts the demand of the next customer. The neural network with one neuron cannot capture the underlying dynamics of this demand data although the amplitude of prediction is lower than the original (i.e. the variance of the prediction is also lower than that of the original).

While comparing all results of prediction accuracies, the accuracy of MDL-optimal neural network is as high as 0.92 which is also approximately 1.4 times of those of exponential smoothing and multiple regressions as shown in Table 3.

We also implement the prediction for this nonlinear demand data by using the four coefficients. As we expected, the description length curve shows the neural network with one neuron is the optimal model in this case. For the sake of brevity, we do not show these figures. In addition, we replace the time series of the Ikeda map with the periodic data and then repeat the prediction in the same way. We notice that the MDL-optimal model as well as the other models fails to follow the future tendency of the demand data. This suggests that one should be cautious when using environmental factors which are possibly related to customer demands when making

![Graph](image-url)

**Fig. 7.** Description length and prediction accuracy of the twenty neural network candidates, using environmental factors.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Prediction accuracy of customer demands with environmental factors.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MDL-optimal NN</td>
</tr>
<tr>
<td>MSE</td>
<td>$3.33 \times 10^{-3}$</td>
</tr>
<tr>
<td>Prediction accuracy</td>
<td><strong>0.85</strong></td>
</tr>
</tbody>
</table>
predictions. It is quite difficult to identify the environmental factors that affect customer demands, and wrong employment of irrelevant environmental records would result in poor predictions. In contrast, prediction using historical demand data is preferable but it is suggested that the data characteristics be investigated prior to the prediction by using the surrogate data method, as discussed in this section.

4.5. Prediction of practical demand data

Here we employ the practical demand data, Monthly gasoline demand Ontario gallons in millions from 1960 to 1975, to validate the proposed modeling technique and the framework [1]. We follow the same procedure as in previous cases. The surrogate data method,
as shown in Fig. 11, suggests, with a weak confidence level that the practical demand is consistent with the linear process. There are 192 data points, of which 102 points are selected to train neural networks and the rest are used as testing data. Typical predictions obtained by various neural networks including the MDL-optimal one are presented in Fig. 12. We observe that the MDL-optimal neural network overwhelms the other candidates. Fig. 13 exhibits the description length of these twenty neural network candidates and their prediction accuracy which suggests that MDL is the optimal model.

According to Table 4, we can find that MDL-optimal neural network, still, performed the best with prediction accuracy 0.84. It is of 31%, 50% and 55.6% higher correspondingly. It is noticed that the large scale historical demand causes relatively large MSEs. In conclusion, MDL-optimal neural network shows its high accuracy and adoptability in different perspectives mentioned.

5. Conclusion

This study has provided an analysis of the impact of the neural network forecasting technique when dealing with a linear demand structure with seasonal swings, and also with a non-linear demand structure. Although earlier researchers examined analytically similar demand forecasting problem [4] they did not consider the dynamic properties of the demand data in their prediction. The complementary approach developed here examines the characteristics of several demand processes by using statistical hypothesis testing in order to exclude totally stochastic demands from being used for the purpose of prediction.
Based on the simulation analysis, this study notes a highly significant finding that the method of description length can be adapted so that it can be used to select the optimal neural network that is consistent with the demand structure identified by the surrogate data method. The MDL-optimal neural networks give accurate predictions for typical demand data outperforming its counterparts. The surrogate data method and the MDL method confirm each other by their findings. The proposed framework of both methods gives an insight into various demand predictions. Here we do not consider the ordering policy in the supply chain. Neither has an up-to replenishment policy been considered, that is, one in which the stocks in the retail sectors are kept “up to” a certain level by replenishing products consumed by customers (i.e., their demands). Obviously, the MDL method that has been developed, as well as the MDL-optimal neural network are also applicable to other replenishment policies in those scenarios by ensuring that predictions are accurate.

This study may further be extended as a way of assessing the impact of the bullwhip effect on the performance measures of the supply chain (e.g., total inventory cost and service level of the chain). Given the fact that bullwhip effect has a deteriorating impact...
on the operation cost of the whole chain, the direct relationship between the bullwhip effect and the performance of the proposed prediction techniques is an interesting area for future research. In addition, the structural configuration of the prediction system is suggested as a way to integrate various modules/techniques so as to enhance the efficiency of the whole system.

Acknowledgment

The authors wish to thank the Research Committee of the Hong Kong Polytechnic University and the CInIS research group of University of Western Sydney for their support of this project.

References


Table 4

<table>
<thead>
<tr>
<th>Prediction accuracy of practical demand data.</th>
<th>MDL-optimal NN</th>
<th>MSE-optimal NN</th>
<th>Exponential smoothing</th>
<th>Multiple regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>$1.48 \times 10^4$</td>
<td>$1.03 \times 10^4$</td>
<td>$4.76 \times 10^4$</td>
<td>$6.02 \times 10^4$</td>
</tr>
<tr>
<td>Prediction accuracy</td>
<td><strong>0.84</strong></td>
<td>0.64</td>
<td>0.56</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Fig. 13. Description length of twenty neural networks and their prediction accuracy.


H.C.W. Lau received his M.Sc. degree from Aston University, Birmingham U.K., in 1981 and the Ph.D. degree from the University of Adelaide, Adelaide, Australia, in 1995. He is currently an Associate Professor with the Department of Industrial and Systems Engineering, Hong Kong Polytechnic University, involved in research and teaching activities. His current research areas cover manufacturing, data management, workflow automation, and artificial intelligence applications.

G.T.S. Ho received his Bachelor degree and PhD at the Hong Kong Polytechnic University. He is currently a lecturer in the Department of Industrial and Systems Engineering at the Hong Kong Polytechnic University. His research interests include supply chain management, integrated quality enhancement systems and artificial intelligence applications.

Yi Zhao received his M.Eng. degree from Zhejiang University, Hangzhou, China, in 2003 and the Ph.D. degree from Hong Kong Polytechnic University, Hong Kong, China, in 2007. He is currently a Postdoctoral Fellow at Hong Kong Polytechnic University. His research interests include time series analysis, nonlinear system modeling, and neural networks.