Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 1651 - 1666, 2005

DESIGN OF DYNAMIC NEURAL NETWORKS TO FORECAST SHORT-TERM RAILWAY PASSENGER DEMAND

Tsung-Hsien TSAI
Ph. D. candidate
Department of Transportation and
Communication Management Science
National Cheng Kung University
1, Da-Shiue Rd., Tainan, 70101 Taiwan
Fax: +886-6-275-3882
E-mail: thtsai@mail.tcm.ncku.edu.tw

Chi-Kang LEE
Professor
Department of Transportation and
Communication Management Science
National Cheng Kung University
1, Da-Shiue Rd., Tainan, 70101 Taiwan
Fax: +886-6-275-3882
E-mail: cklee@mail.ncku.edu.tw

Chien-Hung WEI
Professor
Department of Transportation and
Communication Management Science
National Cheng Kung University
1, Da-Shiue Rd., Tainan, 70101 Taiwan
Fax: +886-6-275-3882
E-mail: louisewei@mail.ncku.edu.tw

Abstract: This paper develops two dynamic neural network structures to forecast short-term railway passenger demand. The first neural network structure follows the idea of autoregressive model in time series forecasting and forms a nonlinear autoregressive model. In addition, two experiments are tested to eliminate redundant inputs and training samples. The second neural network structure extends the first model and integrates internal recurrent to pursue a parsimonious structure. The result of the first model shows the proposed nonlinear autoregressive model can attain promising performance and most cases are fewer than 20% of Mean Absolute Percentage Error. The result of the second model shows the proposed internal recurrent neural network can perform as well as the first model does and keep the model parsimonious. Short-term forecasting is essential for short-term operational planning, such as seat allocation. The proposed network structures can be applied to solve this issue with promising performance and parsimonious structures.

Key Words: Dynamic neural network, Short-term forecasting, Railway Transportation, Revenue Management

1. INTRODUCTION

Short-term forecasting is essential for the following planning because it can offer future demand information as the base for resource allocation. There have been many successful applications of short-term forecasting in the literature, such as short-term traffic forecasting (Fu and Rilett, 2000), short-term load forecasting (Mohammed et al., 1995), short-term flood forecasting (Chiang et al., 2004), short-term passenger demand forecasting (Sun et, al., 1997) and short-term wind forecasting (More and Deo, 2003). These researches show effective forecasting models are the fundamental of short-term operational planning.

Short-term forecasting in railway field is also a vital issue because it can offer short-term demand information to operators for decision making. For example, railway operators can establish booking limits of served OD pairs according to forecasting results. Furthermore, operators can monitor booking demand against the forecast to conduct appropriate marketing strategies to obtain better revenues. For example, if reservation is overestimated, then adequate marketing strategies must be triggered to avoid too many empty seats. In contrast, if reservation is underestimated, then adequate marketing strategies must be triggered to avoid too many rejections. Above concept is the basis of seat allocation, which is a vital sub-system in a revenue management system (RM), and effective forecasting can assist correct seat allocation. A quantitative benefit of forecasting in a simulated RM environment reported by (Lee, 1990) is 10% increase in forecast accuracy can obtain 0.5% to 3% increase in revenues for high demand flights. RM is popular in service industries which products are perishable, such as airlines, hotels and restaurants. RM can be seen as a state-of-art tool to control capacity to maximize revenue (Boyd and Bilegan, 2003). The benefit of RM to revenues in airlines was reported to be in the range of 3% to 7% annually (Pinchuk, 2002). As a result, if effective forecasting models are available, then appropriate booking policies can be established and the improvement of revenues can be expected.

Many papers have compared different methods to forecast short-term road traffic volume, speed, and travel time. (Vlahogianni et al., 2004) concluded that non-parametric models outperform parametric models, for example, non-parametric regression and artificial neural networks. In this study, we focus on the model construction of artificial neural networks. (Ishak et al, 2003) applied three dynamic neural networks to forecast short-term traffic speed. (Lingras and Mounford, 2001) applied time delay neural network (TDNN) to forecast inter-city traffic flow. (Yun et al., 1998) also used three dynamic neural networks to forecast traffic volumes. These researches apply the concept of external dynamics and tapped-delay line to turn network structures from static to dynamic and all of the papers showed promising performances of dynamic neural networks. However, there is no consistency about which kind of design is the best for short-term forecasting. In the study, we propose two dynamic neural network structures. The first neural network structure is

based on the idea of autoregressive model, which is belonged to the design of external dynamics, in time series forecasting. Moreover, two experiments are tested to eliminate redundant input variables and training samples. It is because too many redundant input variables might contain noises and also make models become complicated. Too many redundant training samples slow down training speed and only contribute trivial to forecasting performance. The second model is to modify the network structure in the first phase and pursue network parsimony without scarifying forecasting performance. It is because one of the main goals of modeling is to provide a simplified but acceptable accurate view of a complicated situation. In addition, a simple model structure is helpful during model construction and beneficial for the following maintenance. The organization of the paper is as follows. The second section shows characteristics of data applied in the study which is offered by Taiwan Railway Administration. The third section introduces the design of network structures for short-term railway passenger demand forecasting. The forth section shows the result of an empirical study. Conclusions are briefed in the last section.

2. DATA ANALYSIS

Instinctively, passenger demand may be affected by both temporal and space factors. For example, historical passenger demand pattern is important to explain current passenger demand pattern in terms of temporal aspects. Passengers who cannot get the tickets of their desired schedules might convert to adjacent schedules which can be turned into a space aspect in a network structure. In the previous research (Tsai et al, 2003), we have shown the space factors do not affect forecasting performance significantly. Other exogenous variables, such as price and competition in the market, are constant in the short-term. As a result, we treat the problem as a univariate time series forecasting problem. That is, the forecasting target is passenger demand of a specific OD pair on a specific schedule and possible explanatory variables are its historical patterns.

Two-year sales data is collected from Taiwan Railway Administration (TRA). The data of passenger demand in year 1999 is applied for model construction and the data of year 2000 is used to verify forecasting performance. Figure 1 shows the distribution of raw data. From the result of Figure 1 and priori knowledge, the data has the following features. First, there was no policy which may cause significant changes of demand in the research period. Second, extreme values are apparent in the figure. It is necessary to modify these extreme values because we cannot find any appropriate explanations. Third, the data has two temporary peaks and one lasting peak. These peaks make the data become non-stationary. The starts of two temporary peaks and a lasting peak at point 35, point 86 and point 188 represent winter vacation, spring vacation and summer vacation respectively. Forth, there

are many turning points in the data. This phenomenon shows the difficulty of the research problem and the justification to use sophisticated methods. Fifth, there are usually four general factors to influence the distribution of data. They are trend, seasonality, cycle and randomness. For short-term railway passenger demand, cycle is not as obvious and important as that in economic data. Trend is also hard to recognize from the figure 1. Seasonality, which means weekly pattern in the study, is significant according to prior knowledge.

Figure 2 shows the frequency graph of passenger volumes. We can see that both year 1999 and 2000 are normal-like distribution. As a result, the modification of outliers will not change the original characteristics of passenger demand. Figure 3 shows the variation of each day-of-week in 1999. We can see that the central 80% of data fluctuates between 50 and 125. Tuesday has the minimal variation, and Thursday has the maximal variation. The period from Thursday to Saturday is peak within a week. Figure 4 shows the variation of each day-of-week in 2000. We can find that central 80% of data varies between 56 and 126. Friday has the minimal variation and Wednesday has the maximal variation. The peak within a week happens on Wednesday, Friday and Saturday. The difference between Figure 3 and Figure 4 shows the change of passenger behavior over time. Figure 5 shows the distribution of autocorrelation values (ACF) in 1999. ACF values can reveal the relationship between current and past passenger volumes, and can be used to select vital inputs from historical data. The top 5 ACF values in Figure 5 are (t-7), (t-14), (t-21), (t-2), (t-1) respectively.

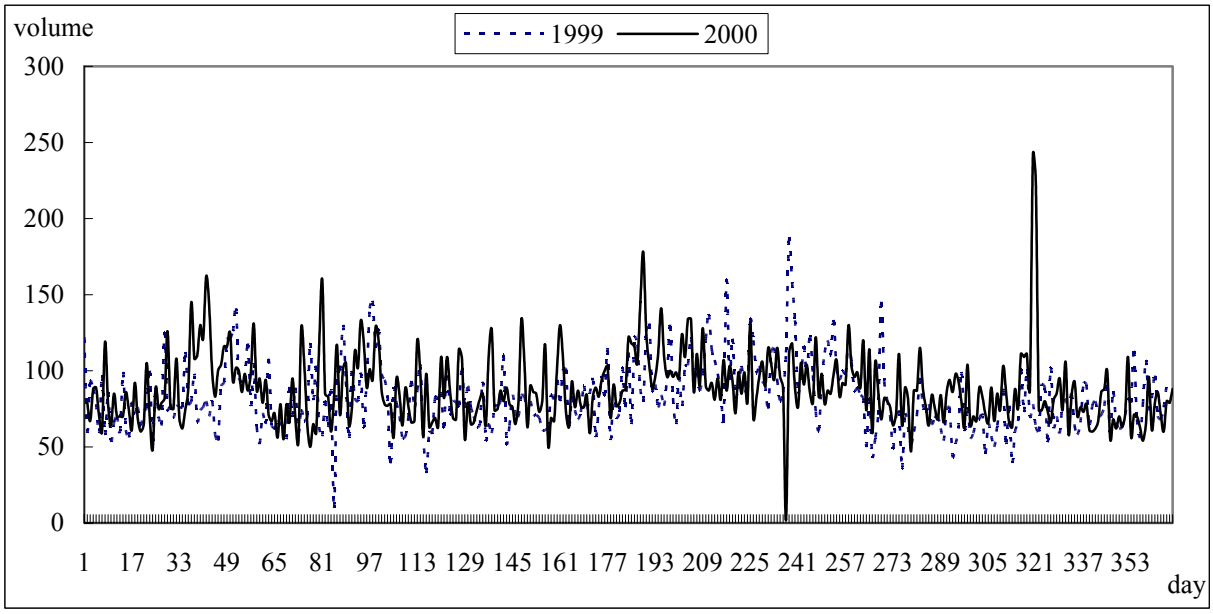


Figure 1. Data Distribution of 1999 and 2000

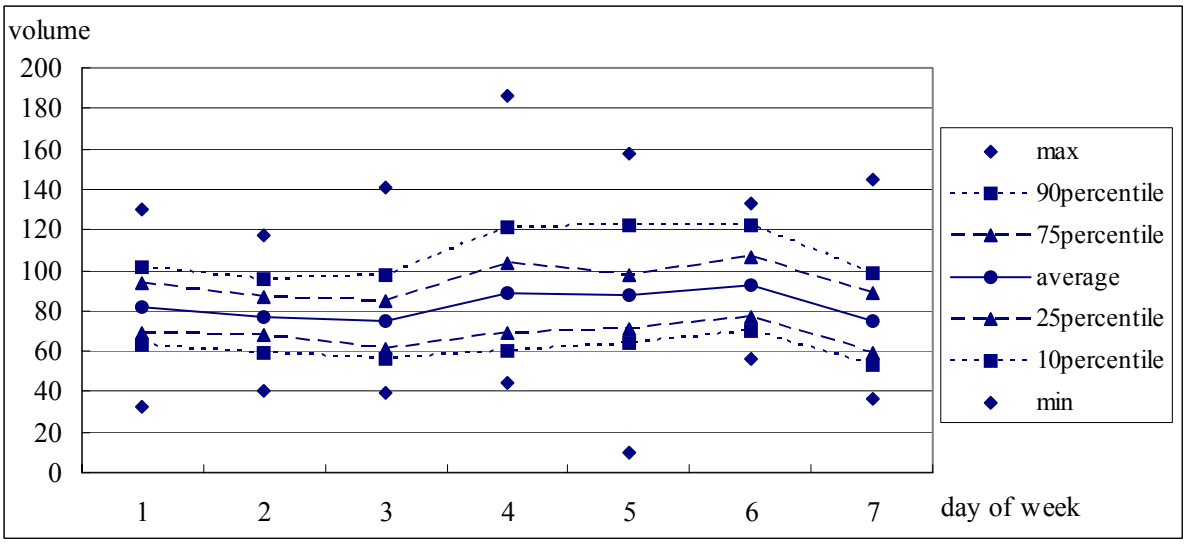


Figure 2. Frequency Graph of 1999 and 2000

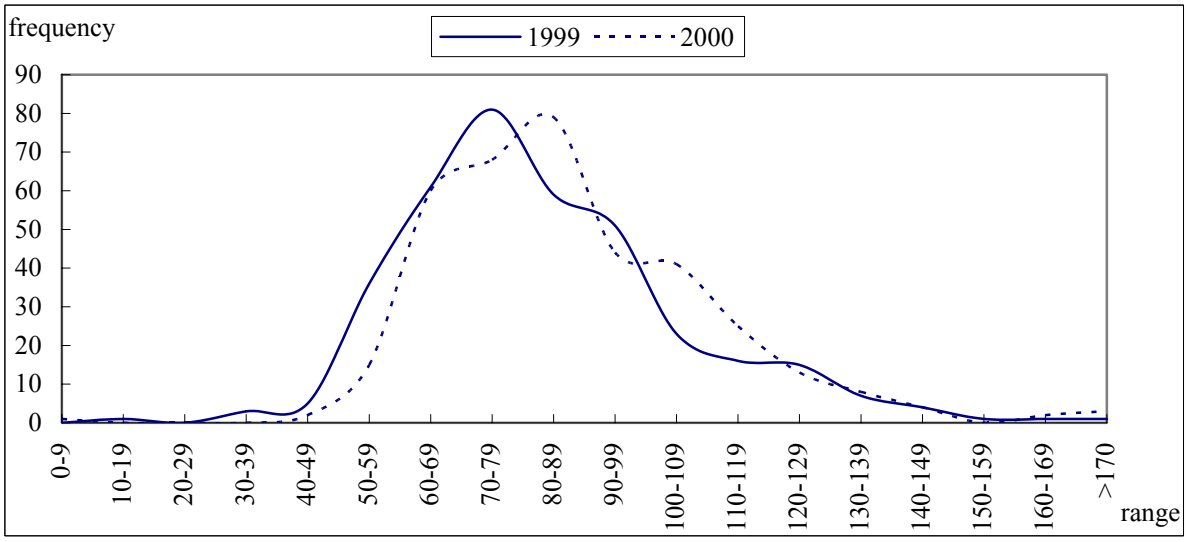


Figure 3. Variation Based on Day-of-Week in 1999

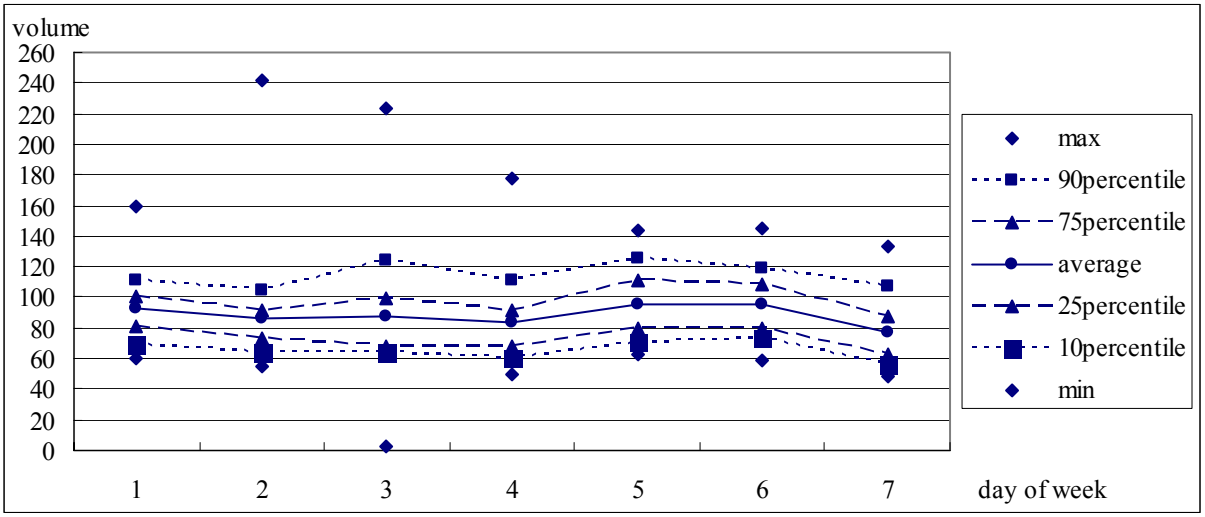


Figure 4. Variation Based on Day-of-Week in 2000

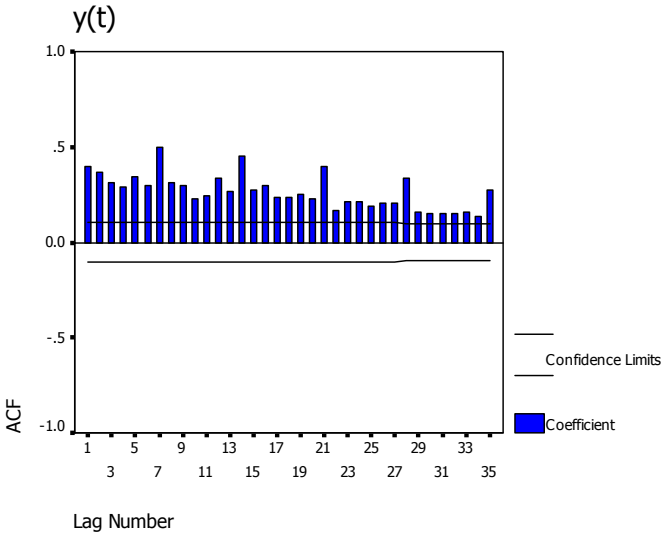


Figure 5. ACF Values of 1999

3. DESIGN OF NETWORK STRUCTURE

3.1 Multi-layer Feed-forward Neural Network (MLFN)

ANN has already been applied in transportation field. The most common network structure for short-term forecasting is MLFN which is a parallel distributed processing network. Parallel distributed processing network is formed by many basic units called neurons. Information processing takes place through the interactions of a large number of neurons can solve difficult tasks. This is the idea of learning tasks via different angles via neurons and the interactions between neurons. In this research, we adopt MLFN as the basic model structure because of five major reasons. First, MLFN is a data-driven model and can learn the relationship between inputs and outputs from data without any assumption during model construction. This is beneficial because wrong model specification may lead to bad performance and it is a good idea to let data speak for itself. Second, MLFN is a nonlinear approximator which is appropriate for complex problems such as passenger behavior. In addition, MLFN is also a universal approximator, which means MLFN can approximate any kind of unknown functions. Third, MLFN is formed by many simple units and parallel connections. It is advantageous because information can be received by different units via connections, and distortion of some units would not cause too many damages. Forth, MLFN has been shown to have better performance than other parametric or non-parametric models in short-term traffic forecasting (Vlahogianni et al., 2004). It is a good option for model selection. Fifth, MLFN is flexible to extend to other kinds of neural network structures. MLFN has the most appealing properties so that many improvements and developments on

this network structure are still going. Figure 6 shows the whole network of MLFN applied in the study. The first layer is called input layer which just simply receives external information. The second layer is called hidden layer which can be seen as a feature extractor. If a network can capture the information of a task via the feature extractor as possible as it can, then it can be expected to have satisfactory performance. The third layer is called output layer which yields the final network output.

3.2 Dynamic Neural Network

Dynamic neural network is the extension of static neural network via the consideration of time. The proposed dynamic models are developed based on static MLFN. In general, dynamics can be expressed by using a tapped-delay line, external dynamics and internal dynamics (Nelles, 2001). Tapped-delay line approach uses a sequence of delay to express dynamics and forms time-delay neural network (Yun et al., 1998; Lingras and Mountford, 2001). External dynamics approach uses the historical information of output itself to show dynamics and forms autoregressive type neural network (Campolucci et al., 1999; Tsai et al., 2003). The internal dynamics approach realizes a nonlinear state space model without information about the true process state (Yasdi, 1999; Ishak et al., 2003). In this study, we use the technique of external and internal dynamics.

3.2.1 Non-linear Autoregressive Neural Network (NAR)

The first proposed network structure is the most straightforward network structure in time series forecasting. In the study, the forecasting target is railway passenger volume. If inputs consist of historical passenger volume, then the model becomes a NAR structure, as shown in Figure 7. There are two important issues for this model structure. The first issue is how many input variables we should adopt. This is the advantage of NAR because we can get a lot of historical variables from data. However, it is also the disadvantage of NAR because too many input variables make models become complicated and harm to learning efficiency. Fortunately, several fashions can be used to select input variables such as ACF introduced in section two. We also add two dummy variables to enhance the effect of vacations and national festivals. The second issue is how many samples we should adopt during model construction. Although it is good to include all available samples when training, however, it may not be suitable for short-term forecasting. It is because the change of passenger behavior is so fast that too many (far) samples may make the model misunderstand the trend. These two issues are further discussed in the empirical study.

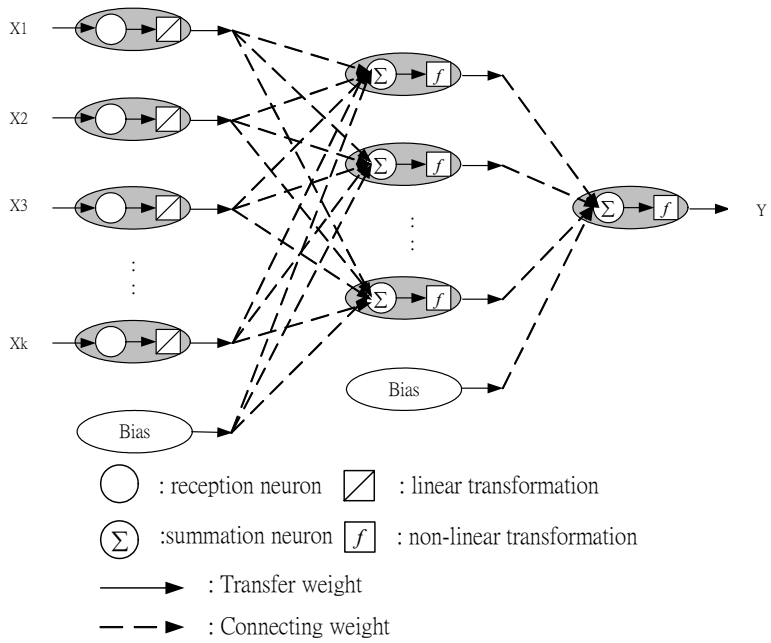


Figure 6. MLFN Network Structure

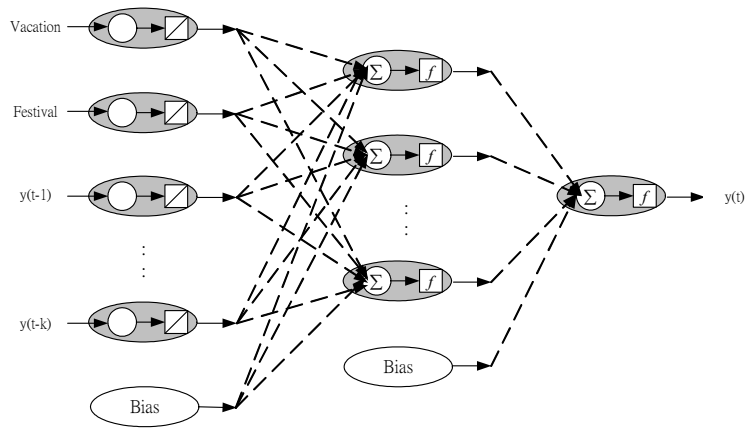
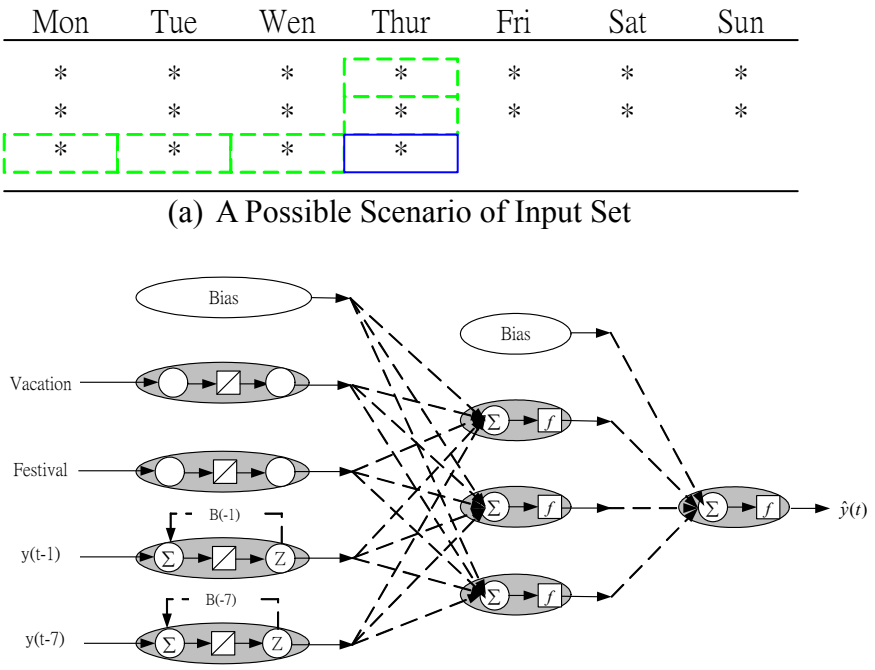


Figure 7. Non-linear Autoregressive Model Structure

3.2.2 NAR With Internal Dynamics (NARWID)

In order to attain generality, we make out-of-sample validation by using the data of 2000 which is not used during model construction. We split the data into twelve forecast targets and construct twelve individual models via the concept of moving window data learning method. The method includes the recent one-month data and excludes the farthest one-month data after each forecast. The problem of this model structure is we may not get a constant model structure. For example, we may get different input sets in different models. If we can design a network structure to eliminate this problem and make the network structure become constant, then it is helpful for model construction and the following practical use. The condition is we cannot sacrifice forecasting performance when applying the designed network structure in comparison with NAR. According to the idea of ACF

shown in Figure 5, we can find that the most important inputs are the previous few days within 7 days for the trend and the previous few days with the same day-of-week for the seasonality. An example is shown in Figure 8(a) which Thursday with solid line is a sample in the forecast target and other days with broken line are inputs. Each forecast target may get different input sets and it depends on the data used during model construction. In order to keep input sets constant without missing possible historical inputs, the concept of internal dynamics is adopted. The idea of internal dynamics comes out the network structure as Figure 8(b). The reason why the proposed network structure can contain historical information is shown in equation (1) and (2). Equation (1) shows if input $y(t-1)$ applies an internal recurrent, then the new variable will consist of not only variable $y(t-1)$ but also the variables of previous lags to form recent trend through a weight structure. Equation (2) shows if input $y(t-7)$ applies an internal recurrent, then the new variable will consist of not only variable $y(t-7)$ but also the variables of previous lags with the same day-of-week to form seasonality through a weight structure. In addition, if the value of weight is less than 1, then the farther lag will get less weight. Through the design, the network can maintain a constant model structure and also includes historical information. Back-propagation learning algorithm (BPN) and real time recurrent learning algorithm (RTRL) are applied to estimate connecting weights of NAR and NARWID respectively. These two algorithms are based on steep descent method and initial seeds might influence the quality of solutions. As a result, extra efforts should be taken to ensure the stability of solutions. In the study, multi-start strategy is adopted to observe the stability of results.



(a) A Possible Scenario of Input Set

(b) NARWID Network Structure

Figure 8. NAR With Internal Dynamics

$$\begin{aligned} z(t) &= y(t-1) + \alpha z(t-1) \\ &= y(t-1) + \alpha\left[y(t-2) + \alpha z(t-2)\right] \\ &= y(t-1) + \alpha y(t-2) + \alpha^2 z(t-2) \\ &= y(t-1) + \alpha y(t-2) + \alpha^2 y(t-3) + \ldots \end{aligned} \quad \text{, where } \alpha \text{ is the weight} \tag{1}$$

$$\begin{aligned} z(t) &= y(t-7) + \beta z(t-7) \\ &= y(t-7) + \beta\left[y(t-14) + \beta z(t-14)\right] \\ &= y(t-7) + \beta y(t-14) + \beta^2 z(t-14) \\ &= y(t-7) + \beta y(t-14) + \beta^2 y(t-21) + \ldots \end{aligned} \quad \text{, where } \beta \text{ is the weight} \tag{2}$$

3.3 Performance Measurement

In the study, we use forecasting error which is the most common criteria in the literature to evaluate forecasting performance. Several variants based on forecasting error can be reached such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and so on. (Lewis, 1982) offered a benchmark and declared if a model's MAPE is fewer than 20%, then it can be categorized as good forecast. We take this benchmark in the study. The formula of MAPE is shown in equation (3) where *d(t)* is actual output at time *t*, *o(t)* is network output at time *t*, and *N* is the number of samples for model validation.

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|d(t) - o(t)|}{d(t)} * 100\% \tag{3}$$

Because of the adoption of multi-start strategy, we can also observe the variation of forecasting results. It should be noted that small variance is expected for stable results. Moreover, paired-t tests are implemented to see whether the difference of average and variance of MAPE between models is significant or not. If the difference is not significant, then we should choose a simpler network structure. Another criterion is to compare the number of connecting weights which is related to the complexity of model structure. In summary, we can observe MAPE between forecast targets within a model structure and use pair-t tests of average and variance of MAPE and the number of connection weights between different model structures.

4. EMPIRICAL STUDY

In this section, we try to answer several hypotheses. First, we start from the most straightforward fashion in time series forecasting and apply a plenty of lag variables. We want to observe whether NAR with a plenty of lag variables can attain promising

performance or not. Second, we apply ACF function for input selection before implementing NAR. We want to observe whether a simple model structure can attain similar or even better performance in comparison with that of the first model structure. Third, we use less training samples during model construction to observe whether a model with less training samples can still get promising results in short-term forecasting. Forth, we apply the proposed NARWID structure to observe whether NARWID can perform as similar as NAR does. If the result is positive, then NARWID is better than NAR because of its simpler model structure.

4.1 NAR With a Plenty of Lag Variables (NARLAG)

In this model structure, we use 28 lag variables arbitrary with two dummy variables as input set for each forecast target. The number of training samples is one year. Table 1 shows performance measurement and this model can be seen as the basic model for the comparison of next model structure. From table 1, we can see that performances are fewer than 20% of MAPE except for Feb, Mar and Apr. Overall speaking, the model can be regarded as good forecast according to Lewis' criterion. However, this model structure takes a lot of efforts to train because of enormous inputs and connecting weights.

Table 1. Performances of NARLAG

Month	Jan	Feb	Mar	Apr	May	Jun
Average MAPE	0.168	0.310	0.292	0.220	0.157	0.181
Variance	1.26638E-05	0.0007	0.0031	5.21E-05	6.12E-05	8.63E-05
# of weights	513	513	513	513	513	513
Month	Jul	Aug	Sep	Oct	Nov	Dec
Average MAPE	0.192	0.175	0.176	0.153	0.139	0.190
Variance	0.0032	0.0005	9.89E-05	1.81E-05	2.96205E-05	0.0004
# of weights	513	513	513	513	513	513

4.2 NAR With ACF Input Set (NARACF)

In this model structure, we try to eliminate some redundant variables which may cause noise. The fashion of ACF introduced in section two is applied to select inputs for each forecast target and a temporal variable enters the input set if it passes ACF t-test. Table 2 shows the result of input selection for each case. We can find that input sets are not constant. The most important variables are $y(t-1)$, $y(t-7)$, $y(t-14)$ and two dummy variables. Table 3 shows Feb gets a lot of improvement because MAPE falls below 20% from 31%. Jul and Aug also get some improvement and other months keep similar results as NARLAG. Moreover, pair-t tests of average (*t value = 2.188*) and variance (*t value = 1.879*) of MAPE

also show significant difference between these two model structures at 90% significant level (*t value = 1.796*). The number of weights also gets significant improvement in this model structure. The consequence shows cautious selections of input sets perform better than that with arbitrary selection which may include many redundant variables.

Table 2. ACF Input Sets

	T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-14	T-21	Vacation	Holiday
Jan							@	@		@	@
Feb							@	@	@	@	@
Mar	@	@					@	@		@	@
Apr	@						@	@		@	@
May	@						@	@		@	@
Jun	@						@	@		@	@
Jul	@					@	@	@		@	@
Aug	@	@	@	@	@	@	@	@		@	@
Sep	@	@	@	@	@	@	@	@		@	@
Oct	@	@	@		@		@	@		@	@
Nov	@						@	@		@	@
Dec	@						@	@		@	@

Table 3. Performances of NARACF

Month	Jan	Feb	Mar	Apr	May	Jun
Average MAPE	0.14	0.198	0.280	0.224	0.139	0.193
Variance	1.80878E-07	2.92E-05	2.06E-05	6.71E-06	2.14E-06	6.21E-05
# of weights	19	22	32	22	22	22
Month	Jul	Aug	Sep	Oct	Nov	Dec
Average MAPE	0.140	0.140	0.161	0.158	0.141	0.182
Variance	6.83E-06	0.0002	0.0002	1.14E-06	8.94E-07	1.18E-06
# of weights	32	72	72	51	22	22

4.3 Test of Training Samples

In this part, we maintain the structure of NARACF but try to apply less training samples during model construction. (Tsai et al., 2003) have implemented this concept and proposed an ANN approach. Here, we apply the result of the approach and use the previous month and one-year-ago month dated before the forecast target during model construction. Table 4 shows the final result and we can see almost all months stay at the same level of accuracy as the model with one-year training samples does. In addition, pair-t tests of average (*t value = -0.644*) and variance (*t value = 0.22*) of MAPE show there is no significant difference between the model with two-month training data and that with one-year training data at

90% significant level (*t value = 1.796*). As a result, it is beneficial to use two-month training data and ignore redundant training samples because small size of training samples helps a lot in training speed and data management.

Table 4. Performances of NARACF With Two-month Training Samples

Month	Jan	Feb	Mar	Apr	May	Jun
Average MAPE	0.143	0.224	0.272	0.216	0.126	0.161
Variance	2.36075E-06	2.44E-05	0.000308	5.74E-06	8.44E-07	3.92E-06
# of weights	19	22	32	22	22	22
Month	Jul	Aug	Sep	Oct	Nov	Dec
Average MAPE	0.137	0.149	0.182	0.186	0.143	0.198
Variance	2.55E-05	2.23E-05	3.03E-05	1.77E-05	4.35E-06	7.24E-06
# of weights	32	72	72	51	22	22

4.4 NAR With Internal Dynamics (NARWID)

In this part, we try to design a network structure which can not only attain performance as similar as NARACF does but also maintain a constant input set. This idea is helpful for three reasons. First, we do not need to select input variables before running the model and it is good for model construction. Second, a constant input set is good for data management because we can just keep the data we need from a lot of historical data. Third, the model is more automatic than models which need to decide input sets in advance and it is convenient for practical use.

In NARACF, we can find that the most important variables which happen in almost all months are $y(t-1)$, $y(t-7)$, $y(t-14)$ and two dummy variables. Other temporal variables have irregular appearance. In order to keep the input set constant without losing other time factors, we design NARWID with abovementioned five variables which three historical variables have internal recurrent. Table 5 shows the result of NARWID. We can find that the performances stay at a similar level in comparison with table 4. The pair-t test of average of MAPE (*t value = 0.401*) shows there is no significant different between NARWID and NARACF at 90% significant level (*t value = 1.796*). NARWID really attains the goal we set up. If we observe the pair-t test of variance of MAPE (*t value = -2.181*), the result shows there is no significant different between NARWID and NARACF at 95% significant level (*t value = 2.201*). It should be noted that the variance of MAPE in NARWID is bigger than that in NARACF. However, the difference is not large enough to pass the test at 95% significant level.

Table 5. Performances of NARWID

Month	Jan	Feb	Mar	Apr	May	Jun
Average MAPE	0.147	0.249	0.243	0.222	0.146	0.148
Variance	0.0002	0.0001	0.0004	0.001	0.0002	0.0002
# of weights	25	25	25	25	25	25
Month	Jul	Aug	Sep	Oct	Nov	Dec
Average MAPE	0.148	0.134	0.154	0.192	0.164	0.16
Variance	0.0002	1.76E-05	0.0001	4.48E-05	3.55E-05	3.47E-05
# of weights	25	25	25	25	25	25

4.5 Discussion

In this research, we design two model structures to forecast short-term railway passenger demand via dynamic neural networks. We find that ACF is a good tool to select temporal inputs before running neural network models. However, different forecast targets may get different input sets. In order to solve this issue and make the input set constant, NARWID is developed and functions as good as NARACF does in the empirical study. There are three following issues. First, even though the test of variance of MAPE in NARWID does not pass at 95% significant level, we should still be careful about the variance. We can see the average of MAPE between NARWID and NARACF is very similar. That means, NARWID may get both better results and worse results than NARACF does. It will be a little troublesome when using the forecast because we do not know the forecast belongs to better results or worse results. This is the problem of local minima in NARWID. Other learning algorithms can be applied to solve this problem and ensure the stability of solutions. Second, we can find the performances of Feb, Mar and Apr are over 20% of MAPE. If we compare the network output and actual passenger demand, we can find the forecast somehow misses the trend, especially for vacations. Short-term passenger behavior changes dramatically and the situations might be more serious in these months so that the historical data cannot capture the trend. An alternative for this issue is to make the forecast become adaptive. In the study, we apply the weights obtained during model construction and make single-ahead forecast for a month at once. We can make the model rerun and get adaptive connecting weights when the forecast error is over a preset threshold so that the model can keep up with the recent trend. Third, other possible designs which have potential to upgrade performance can be tried. For example, other internal dynamics such as the concept of Elman recurrent neural network and internal recurrent in hidden layer can be integrated in NARWID.

5. CONCLUSIONS

In this study, we design two dynamic neural network structures to forecast short-term railway passenger demand. NARACF applies ACF for input selection and performs better than NARLAG. NARACF eliminates the problem of redundant variables. Another experiment about the size of training samples shows the model with two-month training samples performs as good as that with one-year training samples. This experiment eliminates the problem of redundant training samples. However, ACF selects different inputs for different forecast targets. In order to make the input set constant, we design NARWID which can maintain the constant input set without scarifying forecasting performances.

ACKNOWLEDGEMENTS

The authors are grateful to National Science Council for the financial support of NSC93-2211-E-006-052.

REFERENCES

Boyd, E.A. and Bilegan, I.C. (2003) Revenue management and E-commerce, **Management Science, Vol. 49, No. 10,** 1363-1386.

Campolucci, P., Uncini, A., Piazza, F. and Rao, B.D. (1999) On-line learning algorithms for locally recurrent neural networks, **IEEE transactions on neural networks, Vol. 10, No. 2,** 253-271.

Chiang, Y-M., Chang, L-C. and Chang, F-J. (2004) Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling, **Journal of hydrology, Vol. 290,** 297-311.

Fu, L. and Rilett, L.R. (2000) Estimation of time-dependent, stochastic route travel times using artificial neural networks, **Transportation planning and technology, Vol. 24,** 25-48.

Ishak, S., Kotha, P. and Alecsandru, C. (2003) Optimization of dynamic neural networks performance for short-term traffic prediction. Proceedings of the Transportation Research Board 82nd Annual Meeting, Washington, DC.

Lee, A.O. (1990) **Airline reservations forecasting: probabilistic and statistical models of the booking process**. Ph. D. dissertation, MIT, USA.

Lewis, C.D. (1982) **Industrial and Business Forecasting Method**. Butterworth, London.

Lingras, P. and Mountford, P. (2001) Time delay neural networks designed using genetic algorithms for short term inter-city traffic forecasting. In L. Monostori, J. Vancza and A. Moonis (eds.), **IEA/AIE 2001.** Springer, Berlin.

Mohammed, D., Park, D., Merchant, R., Dinh, T., Tong, C. and Azeem, A. (1995) Practical experiences with an adaptive neural network short-term load forecasting system, **IEEE transaction on power systems, Vol. 10, No. 1,** 254-265.

More, A. and Deo, M.C. (2003) Forecasting wind with neural networks, **Marine structures, Vol. 16,** 35-49.

Nelles, O. (2001) **Nonlinear system identification.** Springer, Germany.

Pinchuk, S. (2002) Revenue management's ability to control marketing, pricing and product development, **Journal of revenue and pricing management, Vol. 1, No. 1,** 76-86.

Smith, B.L. and Demetsky, M.J. (1994) Short-term traffic flow prediction: Neural network approach, **Transportation Research Record 1453,** 98-104.

Sun, X., Brauner, E. and Hormby, S. (1997) A large scale neural network for airline forecasting in revenue management. In Yu, G. (eds.), **Operations Research in the Airline Industry**. Kluwer Academic Publishers, Boston.

Tsai, T-H., Lee, C-K. and Wei, C-H. (2003) Artificial Neural Networks Based Approach for Short-term Railway Passenger Demand Forecasting, J**ournal of the Eastern Asia Society for Transportation Studies, Vol. 4,** 221-235.

Vlahogianni, E.I., Golias, J.C. and Karlaftis, M.G. (2004) Short-term traffic forecasting: Overview of objectives and methods, **Transport reviews, Vol. 24, No. 5**, 533-557.

Yasdi, R. (1999) Prediction of road traffic using a neural network approach, **Neural computing and applications, Vol. 8,** 135-142.

Yun, S.Y., Namkoong, S., Rho, J.H., Shin, S.W. and Choi, J.U. (1998) A performance evaluation of neural network models in traffic volume forecasting, **Mathematic Computing Modelling, Vol. 27, No.9-11,** 293-310.