Passenger Flow Forecast Using Wavelet Neural Network Model

Changkun Chen [1, a], Mengyang Xin [2,b]

[1]School of Highway, Chang'an University, Middle Section of South 2 Ring Rd., Xi'an, China

[2]School of Highway, Chang'an University, Middle Section of South 2 Ring Rd., Xi'an, China

[a]xiamenljs@163.com, [b]xinyuyangtian@msn.cn

Keywords: WNN; passenger flow forecast; hybrid genetic algorithm; morlet wavelet

Abstract. Wavelet neural network (WNN), combining with wavelet analysis and neural network, brings forth a high- accuracy performance in identification and approximation. Passenger flow forecast plays an important role in transit scheduling and an improved WNN model is constructed to actualize dynamic forecast, in which Morlet wavelet is selected as the activation function. Input data series surveyed from No.609 line in Xi'an, China, is pre-processed via a fuzzy operator before transferred to train and test the constructed network. A hybrid genetic algorithm and identical dimension recurrence idea are performed to optimize the structure and shape of WNN dynamically to enhance its forecast accuracy. The result indicates the proposed WNN model can accelerate the convergence speed, improve the global generalization ability and possess the practicality in dynamic transit scheduling.

Introduction

Bus scheduling is one of the most important elements in advanced public transit system, where the quality of the hourly passenger flow forecast with lead time ranging from several minutes to one hour ahead has a significant impact on the efficiency of operation and administration of transit utility. Since many operational decisions, such as bus-dispatching interval, temporary scheduling, and transit service assessment, are based on such forecasts. Many statistical methods, i.e. multiple linear regressions, stochastic time series, general exponential smoothing etc., have been used for passenger flow forecast. However, as a joint-effect result of weather and traffic environment, et al, its transfer regularity always reveals the representatively complexities and nonlinear characters and fails to yield satisfied result. For the past few years, artificial neural networks have been paid a great deal of world-wide attention and are now proposed as powerful computational tool to solve such a complex problem [1].

Recently wavelet neural network, as another fruitful technique combining the merits of neural networks with wavelet analysis, has been widely used in forecast. Since wavelet behaves excellent performance in non-stationary signal analysis and non-linear function modeling, WNN can provide much higher availability of convergence for approximation than ordinary multi-layer networks [2].

Algorithm Design

Wavelet Neural Network. If $\varphi(y)$ is a monotonic increasing function accounting for $|\varphi(y)| \leq M$, $M \in R^+$, in which y stands for a continuous function in a certain bounded subset of closure in R_n, $\forall \varepsilon > 0$, $\exists J \in N$, C_i, θ_j, $W_{ij} \in$ constant, the input data series $f(x_i)$ can be infinitely approximated as [3]:

$$\hat{f}(x_1, x_2, \cdots, x_n) = \sum_{j=1}^{J} C_i \varphi(\sum_{i=1}^{n} W_{ij} x_i - \theta_j) \tag{1}$$

$$Max \left| f(x_1, x_2, \cdots, x_n) - \hat{f}(x_1, x_2, \cdots, x_n) \right| < \varepsilon \tag{2}$$

The principle mentioned above shows any given continuous function can be infinitely approximated by a two-layer perception. The basic idea of wavelet transform is to represent a signal series by a linear combination or an integration of the wavelet basis, $\varphi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\varphi\left(\frac{t-b}{a}\right)$, which is determined by dilation factor a and translation factor b [4]. The basic wavelet, $\varphi(t) \in L^2(R)$, is a kind of square integral function and satisfies $C_\varphi = \int_{-\infty}^{+\infty}|\varphi'(t)|^2 \cdot |t|^{-1}dt < +\infty$, where $\phi'(t)$ stands for the Fourier transformation of $\phi(t)$. The continuous wavelet decomposition is the inner result of input signal and wavelets [4] $W_f(a,b) = \langle f, \varphi_{a,b}\rangle$.

Then the function $f(t)$ can be reconstructed by applying inverse operation and integrating over the transform space as $f(x) = \frac{1}{C_\varphi} \cdot \iint \frac{W_f(a,b) \cdot \varphi_{a,b}(t)}{a^2}dadt$. Let $f(t) \in L^2(R)$ be an arbitrary function with a set of training data $T = \{(t_i, f(t_i)) | i = 1,2,\cdots,N\}$. Then a function $\tilde{f}(t)$ that closely approximates $f(t)$ can be approximated arbitrarily $f(t) = \sum_{j=1}^{N} W_j \cdot \varphi_{a,b}(j)$. Known from the above principle, a function $\tilde{f}(t)$ that closely approximates $f(t)$ can be approximated arbitrarily.
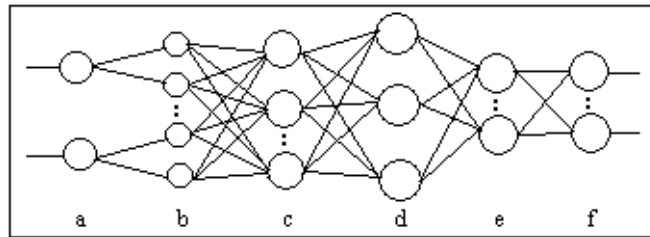
Fig 1. Fuzzy Wavelet Network Structure for Approximation

In Figure 1, the combination of input layer a, membership function layer b, fuzzy regulation layer c, wavelet layer d, hidden layer e and output layer f, leads to a high-accuracy performance [5]. In fuzzy wavelet network, a, b and c are made up of the fuzzy operation, the key content is to deal with the original data through fuzzy reasoning derived from Takagi-Sugeno theory, which is a linear-jointed model from input to output, and states R_i: if x_1 is r_{k1}, x_2 is r_{k2}, ..., x_n is r_{kn}, then $y_k = a_{k0}+a_{k1}x_1+a_{k2}x_2+...+a_{kn}x_n$, where R_k is the kth regulation and r_{kj} ($j=1, 2, ... n$) represents the fuzzy subset with the format of Gauss membership function. The *T-S* principle confirms a certain section $\{(r_{kj}|i \in N\}$ stating the membership grade for input data series $\mu_j(x) = \mu_{r_{k1}}(x_1) \cdot \mu_{r_{k2}}(x_2)\cdots\mu_{r_{kn}}(x_n)$.

If the total fuzzy principle is M, the output computed by fuzzy factors $y' = \sum_{j=1}^{M}\partial_j(x) \cdot g_j(x)$ is yielded subject to $\partial_j(x) = u_j(x)\Big/\sum_{j=1}^{M}u_j(x)$, and transferred to c, a multi-layer perception and the output of wavelet-based neural network y can be represented as [3]:

$$y_i(t) = \sigma\left[\sum_{j=0}^{m}W_{ij}\varphi_{a,b}\left(\sum_{k=0}^{n}W_{jk}x_k(t)\right)\right] \qquad (3)$$

where W_{ij} denotes the connection weight between the output unit j and hidden unit i and W_{jk} is the connection weights between j^{th} input unit and k^{th} each hidden input unit. W_{i0} and W_{j0} respectively represent the bias of the ith hidden unit and j^{th} input unit. σ is the general sigmoid function[6].

The selection of the mother wavelet plays an important role and should depend on particular application. There are a great number of mother wavelets, such as Morlet, Mexican Hat and Meyer [5]. Here Morlet wavelet is chosen to serve as an adaptive basis transfer function.

Optimal Hybrid Genetic Algorithm. Genetic algorithm (GA), which is based on an analogy with natural selection and population genetics, is a kind of search method that was developed by John Holland. One normal application of GA is for search approximate solution to difficult optimization

problems. However, it frequently logs into local minimize if the optimization function has a complex structure [7]. In this paper the optimized objected function of WNN's parameters and shape is constructed to find the optimization solution with a hybrid method, a genetic algorithm combined with a local heuristic search, to offset the deficiency of conventional genetic algorithm.

Network parameters a, b and W can be optimized by minimizing an energy function, which is performed by the least-mean squares theory for data series representation.

$$\begin{cases} MinE(a,b,W) = \dfrac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{m}\left[\tilde{y}_i^j(t) - y_i^j(t)\right]^2 \\ s.t \quad a \in Z^+, b \in R^{m \times P}, W \in R^{P \times n} \end{cases} \tag{4}$$

where $\tilde{f}_i^j(t)$ and $f_i^j(t)$ are the jth desired and obtained output with respect to the ith input data.

Hybrid genetic algorithm involves initialization, mutation, competition and selection. An initial population is selected to initialize the parent generation. In initialization, encoding takes on significant task and each population expressed by encoded chromosome stands for a certain solution to required question. Multi-valued encoding pattern is used to obtain a solution. A chromosome is divided into four segments and the first three segments are made up of the parameter segment, which are encoded in real values to deliver the information about translation factor, telescopic factor and weight factor of hidden unit. Simultaneously, the fourth segment is to encode the structure character in binary, where 1 states it a valid hidden unit, and 0 shows it ineffective [8].

a_1	a_2	...	a_m	b_1	b_2	...	b_m	W_1	W_2	...	W_m

1	0	...	1

Each evolution generation is assigned a fitness value in accordance with the given fitness function $F(a,b,W) = \dfrac{1}{\sqrt{\left[y_i^j(t) - y_i^j(t)\right]^2}}$. Since the most commonly used objective function aims to obtain the least error series between statistical data and forecast value. Offspring is created from their parents by a composite mode. Encoding pattern of cross-gene is firstly judged and linear combination is used in crossover operation for cross-gene encoded by real value. If x_i^g and x_j^g are defined as respective parent individual, satisfying $x_i^{g+1}(P) = \lambda \cdot x_i^g(P) + (1-\lambda) \cdot x_j^g(P)$ and $x_j^{g+1}(P) = \lambda \cdot x_j^g(P) + (1-\lambda) \cdot x_i^g(P)$, where λ is an evolution rate complying with $\beta^{g+1} = 0.99 \cdot \beta^g$ for $g = 0,1,2\cdots$, and subject to $\beta^0 = 0.9$.

Here g and p separately represent generation number of evolution and position of crossover gene. However, conventional single-point crossover is preferred to further evolution, if crossover gene is encoded by binary. Mutation is also treated with different encoding mode and the mutation position encoded by real value yields the following mutation model $x_i^{g+1}(P) = x_i^g(P) + \delta \cdot E(x_{max}) \cdot P_m$, where δ is a stochastic value ranging over [-0.5, 0.5] and $E(x_{max})$ represents cumulative value of squared error derived from the chromosome with maximum fitness when evolving into the gth generation and P_m is a mutation factor to control the mutating operation [7].

If the binary encoding is applied, offspring is generated from their parents by a Gaussian perturbation $x_i^{g+1}(P) = x_i^g(P) + N(0,1) \cdot P_{pi} P_m$, where $N(0,1)$ denotes a normally distributed one dimensional random variable with mean zero and standard deviation. To a total population pop, P_{si} is the ith proportion factor $P_{pi} = F_i \Big/ \sum_{i=1}^{pop} F_i$. To accelerate the convergence speed, a gradient-descending factor P_g is embedded in the search process to optimize new offspring generated from crossover and mutation via a fixed gradient-descending probability [9] $\hat{E}_i^{(P)} = \hat{E}_i^{(P-1)} - P_g \dfrac{\partial E}{\partial a, b, W}\Big|_{(a,b,W)_i = \hat{E}_i^{(P-1)}}$.

Gradient descending factor, which holds the same iterative step-size with evolution process, optimizes the network parameters without remodeling the structure of WNN and characteristic gene of optimized chromosome has to answer for the limited scope, such as dilation factor $a > 0$. Next

evolution group, conserving diversity of population through reproduction and competition, only consists of filial generation derived from crossover and mutation, due to inheriting the unparalleled quality from paternal generation [10]. In hybrid genetic algorithm, competition and selection lead to more adaptive generation through a given proportion, which inevitably causes premature or slow convergence. Here evolving principle is executed through optimum reservation strategy combined with proportion selection. All population is sequenced with ascending sort on the basis of fitness, among which chromosomes with large fitness will be assigned directly into next generation. Otherwise, the above proposed evolving procedure has to be reduplicated via a selection factor

$$P_{si} = (F_{\max} - F_i) \Big/ \sum_{i=1}^{pop} F_i .$$

Selection factor assures the chromosome with low fitness to gain large big probability in next generation. However, evolving population is sometimes logging into a certain individual or some finite ones, which is usually named local convergence and reduces the probability of global optimum. Therefore, distance between two individual is significantly taken into account in selection operation and one individual has to be eliminated correspondingly due to shorter distance [11].

Experiment and Implementation

Input Sample Preparation. The alternation of passenger flow is a random event and affected by a great deal of undetermined factors. However, a long term observation also shows some statistical regularity. K609 bus travels from Xi'an Hi-Tech Industries Development Zone to Tang Paradise covering 25 stations, passenger flow survey have been carried out in November, 2014. In this study, daily amounts of passengers from Nov. 1 to Nov. 22 are used. Input series are endowed corresponding weights matching the fuzzy rule before transferred to wavelet network as shown in Tab 1.

Tab 1. Control Parameters of Fuzzy Operator for input passenger flow data

Input	Fuzzy rule	Division space	Input	Fuzzy rule	Division space
x_1	Low Medium high	[0,6] / (6,12) / [12,18]	x_2	Positive Zero negative	---
x_3	positive & high positive &low zero negative & high negative & low	[3,6] (0,3) 0 (-3,0) [-6,-3]	x_4	positive & high positive & low zero negative & high negative & low	[3,6] (0,3) 0 (-3,0) [-6,-3]

Parameter Definition. With hybrid GA to optimize the WNN, the following control parameters: *POP*=60, hidden unit *M*=16, crossover probability P_c=0.9Z, mutation probability P_m=0.05, gradient descending factor P_g=0.05, total forecast error $e \leq 0.01$, the largest searching step Smax=300. $a \in$ [0.01,12], $b \in$ [-12,12], and $W \in$ [-12,12]. To determine a proper a, b and W, the following simple scheme are proposed on the basis above induced hybrid GA: 1) Pre-processing the input data series through fuzzy reasoning theory; 2) Generating the original population within a logical interval; 3) Utilizing crossover and mutation operations; 4) Creating new offspring via optimum reservation and proportion selection; 5) Sequencing the individuals by fitness and computing the distances between chromosomes; 6) Finishing the search process, if satisfies the stopping rule, or going back to step III; 7) Real-time and dynamic forecast with identical dimension recurrence [12].

Forecast Result. Analysis program of fuzzy wavelet neural network has been practiced to approximate the collected passenger data. With 8 hidden units, 62 evolving generations, the optimum solution is found out to be as $a = 9.6546$, $b = 5.0085$, $W = -5.0432$.

Contrastively, the common GA also searches the satisfied answer with 8 hidden units 246 evolving steps. The total error of the former is 0.0362 and the latter is 0.0428, which means the constructed model in this paper has high efficiency. Fig.2 and Fig.3 give the times of WNN training and result of

network training obtained after about 1183 evolving generation. Meanwhile, the fitness function also meets the accuracy requirement of passenger flow prediction.

From Fig.4, it indicates that WNN model has a higher performance in approximation than BP. However, it displays WNN forecasts more precisely. Let's define the mean standard deviation and mean absolute percentage error MAPE can be expressed by $\mathrm{MSD} = \sqrt{\dfrac{1}{n-1}\sum_{k=2}^{n}\left[x^{(0)}(k)-\hat{x}^{(0)}(k)\right]^2}$

and $\mathrm{MAPE} = \dfrac{1}{n-1}\sum_{k=2}^{n}\left|\dfrac{x^{(0)}(k)-\hat{x}^{(0)}(k)}{x^{(0)}(k)}\right| \times 100\%$. Thus we can obtain MSD = 1437.89 and MAPE = 14.89% for the passenger flow forecast reslt by BP model, and MSD =1481.78 and MAPE = 14.05% for the reslt by the proposed WNN model, respectively, as shown in Fig.5. Obviously, it reaches a biger MSD value and a smaller MAPE value for the forecast result by WNN model, which indicates its higer proficiency and effecyiveness of time series data forecast.
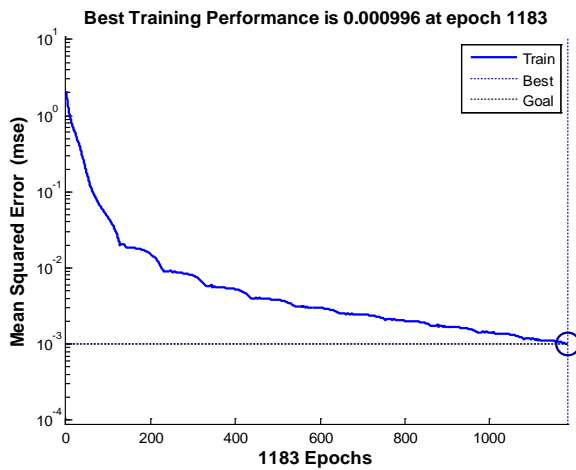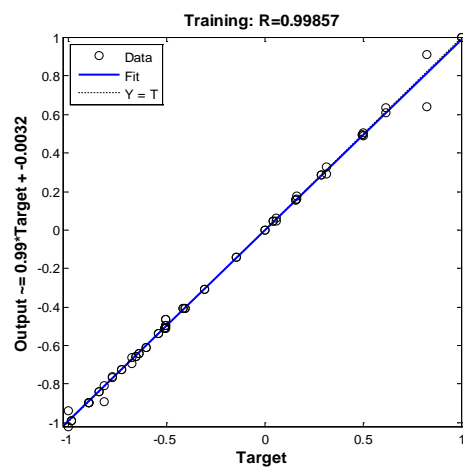

Fig. 2 Times of WNN training

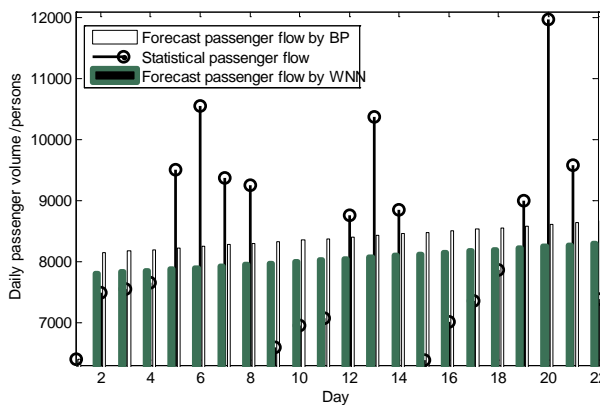**Fig. 3** Result of WNN training
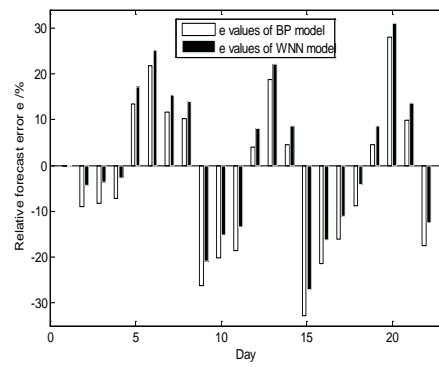

Fig 4. Comparison of passenger flow forecast

Fig 5. Comparison of forecast error *e*

Conclusions

A wavelet-based fuzzy neural network for learning functions from training data, are discussed on the basis of wavelet transform, which combining the excellent performance of fuzzy reasoning model, wavelet and neural network. Hybrid genetic algorithm is introduced to optimize the parameters and structure of constructed WNN model. Experimental result indicates proposed WNN model, optimized by hybrid GA, has fast convergence and high-accuracy ability [13]. Since passenger flow exists in the dynamic system of traffic and transportation, the newer data is supplied with

same-dimension recurrence, the more precise result will be achieved. Plentiful experiences prove the introduction of self-adaptive operation with new data supplement and old one removal can significantly improve the prediction accuracy and realize real-time and dynamic forecast. For future research, it would be of interest to investigate the affecting factors of passenger flow and more optimization algorithms for real-time forecast so as to construct a practical operation system for dynamic public transit scheduling.

References

[1] Wang, Z; Zou, X.F; Dong, H.H. Railway passenger volume forecast during the Spring Festival based on BP network algorithm. *Journal of Beijing Institute of Technology (English Edition)*.2012, *21*, 103-107.

[2] Inoue, Y; Sakaguchi, Y. A wavelet-based method for extracting intermittent discontinuities observed in human motor behavior. *NeuralNetworks*. 2015,*62*, 91-101.

[3] Islam, M.N; Liong, S.Y; Phoon, K.K; et al. Forecasting of river flow data with a general regression neutral network. *IAHS-AISH Publication*.2001, *272*, 285-590.

[4] Bodyanskiy, Y; Lamonova, N; Pliss, I; et al. An adaptive learning algorithm for a wavelet neural network. *Expert Systems*. 2005, *22*, 235-240.

[5] Bazoobandi, H.A; Eftekhari, M. Comparing different versions of differential evolution for training Fuzzy Wavelet Neural Network. Iranian Conference on Intelligent Systems, Teheran, Iran, 2014.

[6] Alsmadi, M.K; Omar, K.B; Noah, S.Az; et al. Performance comparison of multi-layer perceptron (Back Propagation, Delta Rule and Perceptron) algorithms in neural networks. IEEE International Advance Computing Conference, Patiala, India, 2009; pp 296-299.

[7] Peng, S.L; He, Z.Y; Song, Y. A hybrid genetic algorithm for coloring map. *Geomatics and Information Science of Wuhan University*. 2009, 34, 1499-1501.

[8] Kunt, M. M; Aghayan, I; Noii, N. Prediction for traffic accident severity: Comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods. *Transport*. 2011, 26, 353-366.

[9] Lai, J.B; Wang, H.Q; Liu, X.W; et al. WNN-based network security situation quantitative prediction method and its optimization. *Journal of Computer Science and Technology*. 2008, 23, 222-230.

[10] Wang, Y.P. The tourism demand of nonlinear combination forecasting based on time series method and WNN. *International Journal of u- and e- Service, Science and Technology*. 2015, 8, 301-310.

[11] Feng, J.N; Biao, Z.J; Yang, X.B; et al. Wavelet neural network optimization based on hybrid hierarchy genetic algorithm. *Systems Engineering and Electronics*. 2005, *27*, 1112-1114.

[12] Yuen, J.K.K.; Lee, E.W.M.; Lo, S.M. An intelligence-based optimization model of passenger flow in a transportation station. *IEEE Transactions on Intelligent Transportation Systems*. 2013, 14, 1290-1300.

[13] Zhang, X.J; Mao, B.H; Wang, Y.L; et al. Wavelet neural network-based short-term passenger flow forecasting on urban rail transit. *Telkomnika-Indonesian Journal of Electrical Engineering*. 2013, 11, 7379-7385.