Research and Implementation of an AlgorithmOf the Network Traifc Prediction

Huanqin Li

Zhengzhou Normal University, Zhengzhou Henan 450044, China

Abstract

With the arrival of information age, Internet is developing rapidly while various new network application are emerging one afer another,which led to the emergence of various network problems and brought the network monitoring huge challenges,this paper analyzes the characteristics of network traffics and the traditional forecasting methods, In order to improve the network traffic prediction accurate, this paper proposed a network traffic hybrid prediction method based on wavelet analysis. The hybrid model uses wavelet analysis to decompose network traffic's the linear and nonlinear parts, and then used respectively ARIMA model and the BP neural network model to prediction its, lastly uses wavelet analysis to reconstruction the linear and nonlinear partial results and get mixed models eventually prediction result. Simulation results showed that the hybrid model is better than other network traffic prediction model; it has higher prediction accurate and is an effective analysis method for network optimization control.

Keywords

Network, flow model wavelet decomposition and reconstruction, BP neural network.

1. Introduction

With the rapid development of network technology and increasing of network traffic network congestion is more and more serious, The network administrator has to keep abreast of the current operating status of the network, and take considerable control and management measures to improve the network efficiently. The accurate prediction of network traffic is the prerequisite for network management, Therefore, the network traffic prediction becomes the key of research areas and an issue in the study of network control[1].

Traditional network traffic prediction method has the linear regression model, Poisson model, Matkov models and time series forecasting model, Since the network traffic data is essentially a time series, time series model is the most commonly used in the traditional models, in which autoregressive moving average model (ARIMA) is widely used ,and the predictive performance is the best [2].As the network is more and more widely applied, the type of data network transmission becomes diverse, The changes in network traffics have the features of multi-scale、 highly self-similarity, nonlinear and time-varying, periodic and non-linear ,and so on ,The traditional network traffic prediction methods are based on linear Variation of the model, so the change of network traffic can not be well illustrated and described by these methods , and the prediction is not quite accurate ,which is not suitable for the developing requirements of modern large-scale networks [3]. In recent years, the further development of non-linear prediction theory has attracted attentions of scholars, neural networks and the learning methods as vector machines have emerged which are applied to the network control management. In particular, artificial neural network is a learning method that has a strong linear predictive power, It has the global promotion of best approximation and a good ability to network traffic changes of nonlinearity and uncertainty of the precise description of network traffic, Thus it becomes the main Prediction method [4][5]. But the neural network is a learning method based on empirical risk minimization principle, It cannot fully depict the multi-scale, self-similarity features of network traffic , so the prediction and the actual network traffic are different ,Which provides a wrong direction to the network administrator.To predict network traffic accurately, We must be fully capture its variation, wavelet transform is a multi-resolution approach to multi-scale decomposition of network traffic, It decomposes the network traffic signal from coarse to fine , and distinguish the traffic signals from the nonlinear signal, All the above is a strong above for the analysis of network traffic changes. [6]

To solve the problem of the low accuracy of Network traffic prediction, This paper proposes a network traffic hybrid prediction method based on wavelet analysis after in-depth analysis of network traffic characteristics.Simulation results showed that compared to other network traffic prediction model, hybrid model improves the prediction accuracy of network traffic.

2. Principles of Network Traffic Prediction

Network traffic prediction is developed to use certain prediction model of network traffic, according to collect data on future changes at a certain moment network traffic prediction which directs the network administrator master the network to accurately predict the network traffic, it should predict its linear part and the non-linear part, while the traditional one can only predicts the linear part, so the prediction is not so accurate. Although the BP neural network can predict the non-linear part, It can decompose the network traffic characteristics of changes,but the gap is larger between the prediction and true sometimes . This paper combines the advantages of ARIMA linear predition and the BP neural network nonlinear prediction, It applies the wavelet analysis to decompose the original network traffic and two methods are involves in the prediction.Finally it gets the final prediction.

Affected by the environment, users and network itself,the network has the characteristics of a multi-scale, mutations,etc.The changes are unstable,which can be expressed as the following:

$$z(t) = s(t) + v(t) \tag{1}$$

In the above formula , $z(t)$ is the original sequence for the network traffic, $s(t)$ is for the linear part, and $v(t)$ is the nonlinear part. The mathematical model decomposed by the Wavelet is following :

$$\{s(t), v(t)\} = W(z(t)) \tag{2}$$

Among the Formula , $W()$ is wavelet decomposition operator. Linear part of the mathematical model prediction is :

$$s'(t) = ARIMA(s(t)) \tag{3}$$

Among theFormula, $s'(t)$ is Linear part of the prediction , $ARIMA()$ represents linear prediction models. The mathematical model of the nonlinear part prediction as:

$$v'(t) = BP(v(t)) \tag{4}$$

Among theFormula, $v'(t)$ is Nonlinear part of the prediction, $BP()$ is Nonlinear prediction model. The mixed model of final prediction results reconstructed the mathematical as:

$$z'(t) = W(s(t), v(t)) \tag{5}$$

Among the Formula, $z'(t)$ is the final Predictions, $W()$ is the Wavelet reconstruction operator.

3. The Design of Network Traffic Prediction Model

3.1 Wavelet Decomposition of Network Traffic

The wavelet analysis is a branch of mathematics developed on the basis of the Fourier transform in 80's last century. It can be able to maintain the scale of the object not deformed. the most important specially design of network traffic is self-similarity. It is a random process of deformation in a statistical sense. The wavelet analysis is course to become the methods and tools,with which ideal network traffic analysis and processing areto be done. For the arbitrary function and the signal f (t), the continuous wavelet transform is defined as:

$$Wf(a,b)=\int_R f(t)\overline{\Psi}_{a,b}(t)dt=\frac{1}{\sqrt{a}}\int_R f(t)\overline{\Psi}(\frac{t-b}{a})dt \tag{6}$$

Among the formula, $\Psi_{a,b}(t)$ is a continuous wavelet function.Dilation factor a determines the scale of the observed signal. Shift factor b represents the translational position of the small function on the horizontal axis. Intuitively, the wavelet transforms decompose a one-dimensional signal into approximation signal (aj) after translation in time and scale (or frequencies) on the scaled detail signal (dj). With the increase in the scale j, the original signal f (t) can be multi-scale decomposition after approximation signal and detail of the lossless signal showing. With the increase in the scale j, the original signal f (t) can be showed by multi-scale decomposition after approximation signal and detail of the lossless signal[7].that is shown as formula 7:

$$\begin{aligned} f(t) &= a_j+\sum_{j=0}^{N} d_j \\ &= \sum_k m_x(j,k)\Phi_{j,k}(t)+\sum_{j=0}^{N} n_x(j,k)\Phi_{j,k}(t) \end{aligned} \tag{7}$$

In the formula , N is the number of scales. $m_x(j,k)$, $\Phi_{j,k}(t)$ are the scaling coefficients and mother scaling function respectively, and $n_x(j,k)\Phi_{j,k}(t)$ are wavelet coefficients and mother wavelet function.

Wavelet analysis is a rapidly developing new field in the current mathematics, It also has a double meaning of profound theory and a wide range of application[8]. Compared to the Fourier transform and window Fourier transform, Wavelet transform is a local transformation of the signals in time domain and frequency domain, and thus It can effectively extract from the signal characteristic information[8].

In this paper, as network traffic changes with the nonlinear, multi-scale features, So it can be decomposed by wavelet analysis into different frequency components, namely, the linear part and nonlinear part. Suppose $z(t)$ is obtained the Network traffic based data sets. Apply discrete wavelet to decompose it, thus the network traffic nonlinear coefficient of the scale d_{j+1} can be expressed as.

$$t=0,1,L\ ,N$$

$$d_{j+1}(t)=c_j(t)-c_{j+1}(t) \tag{8}$$

Among them, the decomposition scale factor is:

$$c_{j+1}(t)=\sum_{l=-\infty}^{+\infty} h(l)c_j(t+2^j l) \tag{9}$$

The wavelet transform of Network traffic time series of the original L is resolution:

$$D = \{d_1, d_2, \cdots, d_L, c_L\} \tag{10}$$

Where c_L and d_j represent the network traffic signals respectively.

Wavelet decomposition of network traffic is shown in Figure 1.

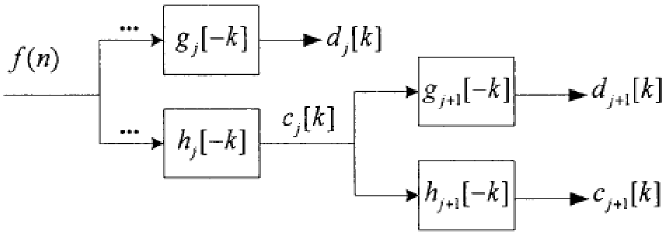

Figure 1 Wavelet decomposition of network traffic

3.2The Linear Part of the Network Traffic Prediction Model (ARIMA)

ARIMA model is is a famous time series prediction method,which proposed by Box andJenkins in the early 1970s. and is also known as the box-jenkins model.it is a generalization of Auto-Regressive and Moving Average Model(ARMA). ARMA is the most commonly used fitting stationary series model, which consists of two parts, since the return (autoregression AR) models and moving average (Moving-average, MA) model. ARIMA model is consists by the AR model, I Differencing or Integration MA model , and has more flexible than ARMA[9].

Let {wt, t = 0,1, ...} represents a random time series, if there is a non-negative integer d , make the formula (11) workable.

$$\nabla^d w_t = x_t \tag{11}$$

In the formula ∇ = 1-B is a differential operator, and B is backward shift operator .

Then The time series is a stationary time series, otherwise they will be a smooth process.

If it is a a stationary time series,we can get the following formula .

$$\Phi_{(B)} = 1 - \Phi_1 B - ....\Phi_q Bq = 1 - \sum_{k=1}^{q} \Phi_k B^k \tag{12}$$

$$\varphi_{(B)} = 1 - \varphi_1 B - ....\varphi_p B^p = 1 - \sum_{k=1}^{p} \varphi_k B^k \tag{13}$$

In the above formulas , $\varphi_{(B)}$ and $\Phi_{(B)}$ are relatively prime, $\varphi_p \Phi_q \neq 0$, $\{\alpha_t\}$ is white noise sequence. SO, {wt, t = 0 , 1 ,… } should satisfy the stochastic differential equation (14).

$$\varphi(B) x_t = \theta(B) \alpha_t \quad (14) \qquad |B| \leq 1$$

We can Claim {wt, t = 0,1, ...} as Autoregressive Integrated Moving Average Model, denote it as ARIMA (p, q, d) model.

In this papar, after wavelet decomposition, the network traffic is divided into two parts, the use of linear high prediction ability of the ARIMA predict its linear part.

ARIMA model is autoregressive moving average model deformation, which includes self-regression model, difference and moving average model of three parts, being more flexible than the ARMA .The Linear data of the network traffic is given as $\{d_1, d_2, \cdots, d_L\}$, First, We should deal with the difference and make it more stable ,and then determine p, q parameters of the ARIMA model, the last one is the model prediction. In this paper, the minimum information criterion (AIC) and the SBC criteria are used to determine the optimal order model, and the conditional least squares estimate model parameters. AIC criterion function is defined as:

$$AIC = n\ln(\sigma_\varepsilon^2) + 2(p+q+1) \tag{15}$$

SBC function is defined as:

$$SBC = n\ln(\sigma_\varepsilon^2) + 2(p+q+1) \tag{16}$$

3.3 The Nonlinear Part Prediction Model of the Network Traffic (BPNN)

BP (Back-Propagation) Neural Network is put forword in 1986 by the team of scientists led by Rumelhart and a McCelland, it is a forwardbroadcasting multiply layer network that trains by the error back propagation algorithm. Also it is currently the most widely used neural network model. BP neural network can learn and store a large number of input - output mode mapping relationship without prior reveal the mathematical equations that describe the mapping between. The learning rule is to use the steepest descent method, adjust the network weights and thresholds by reverse spreading to minimum sum of error square. BP neural network topology include the input layer, hidden layer and output layer.[10]

Bp neural network network model is shown in figure 2.

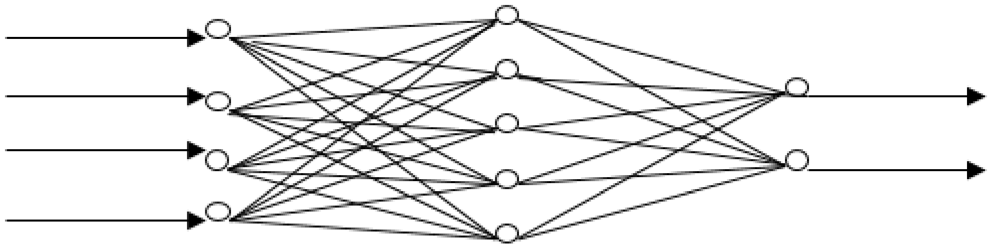

Figure 2 Bp neural network network model

The input layer neurons are responsible for receiving input information from the outside world, and passing the information to the middle layer neurons; the middle layer is the internal information processing one, which is responsible for the information conversion, according to the demand of the information change capacity, the middle layer can be designed as a single hidden layer or multi-hidden layer structure, then the information is passed form the last hidden layer to the output layer neurons, after further processing, a learning process forward propagation is completed, at last the information was output to the outside world by the output layer. When the actual output does not match the expected output ,BP network entere the error back-propagation stage. Errors are back-propagated through the output layer to input layer. In the proccess, the layers weights was Corrected by gradient descent layers, Cycle of information forward propagation and error back-propagation process is ongoing tuning process layers weights,also is neural network learning and training process, This process continues until the network output error is reduced to an acceptable level, or pre-set number of learning far.The BP neural network model is The network model contains the following sections[11].

(1)node output model

hidden node output model is O_j

$$O_j = f(\sum W_{ij} \times X_{i} - q_j) \tag{17}$$

Y_k is output model of output node

$$Y_k = f(\sum T_{jk} \times O_{j} - q_k) \tag{18}$$

In the above ,f is Non-linear effect function, q is Neuron threshold value.

(2) Action function model

Action functions is a function that reflect lower input pulse intensity stimulation to the upper node which is also known as stimulate function which is shown as formula(19)

$$f(x) = 1/(1+e^{-x}) \tag{19}$$

(3)the error calculation model

Error calculation model is a function to reflect the size of the error between the desired output and calculated output of neural network , E_p is the error calculation model

$$E_p = \frac{1}{2}\sum_{j=1}^{m}(t_j^p - y_j^p)^2 \tag{20}$$

In the formula , t_j^p Desired output value of the node y_j^p The node calculation output value.

(4)Self-learning model

The learning process of Neural network is setting the weights array Wij between the lower nodes and the upper node heavy and ajusting the error repeatly. BP network is divided into teacher learning style which Need to set the bar of expectations and the way to learn without a teacher that just need enter the mode.

$$\Delta W_{ij}(n+1) = h \times \Phi_i \times O_j + a \times \Delta W_{ij}(n) \tag{21}$$

h Learning factor, Φ_i The calculation error of the output of node I, O_j caculating output of output node j, a Momentum factor

In this paper ,the quantity layers of BP neural network is L , the output of any node i is O_i the output of the layer $L-1$ of the node i is O_{ip}^{l-1} , the input of the the layer l of the node j is as follow.

$$net_{jp}^{l} = \sum_{i=1}^{n_{l-1}} w_{ij}^{l} \times O_{ip}^{l-1} \tag{22}$$

Among them, w_{ij}^l indicates the network's connection weights.

The output of the the layer l of the node j is as

$$O_{ip}^{l} = f^{(l)}(net_{jp}^{l}) \tag{23}$$

Among them, $f^{(l)}$ represents for the activating function of each node, In this paper, logarithmic sigmoid function is used for neural network activating function.

$$f(x) = \frac{1}{1+e^{-x}} \tag{24}$$

BP neural network is used for the nonlinear part of the training network traffic data set for learning and error feedback, Through the network weights and thresholds, is adaptively adjusted to reduce inaccuracy to the required accuracy, so as to obtain the optimal Network weights, and establish the optimal weights of network traffic on the network[12] .

3.4 The Final Prediction of Network Traffic Wavelet Reconstruction

By the ARIMA model and BP neural network traffic on the network to predict the linear and nonlinear part, the desired results are achieved.The results of linear and nonlinear part are inputted to the wavelet , The final predictions of the network traffic mixed model predictions are to obtain by wavelet reconstruction.

The wavelet reconstruction of Network traffic is wavelet decomposition's anti-process.

$$c_{j-1} = c_j \bar{h} + d_j \bar{g} \quad (j = 0,1,2,\cdots L) \tag{25}$$

Among the Formula, $\bar{h}$ and $\bar{g}$ are the dual operator respectively,

The final prediction result of the network traffic during reconstruction of wavelet is shown in Figure 3.

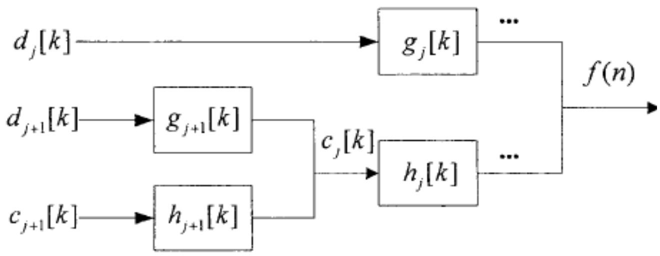

Figure 3: Wavelet reconstruction of network traffic

3.5 The Specific Prediction Process of Network Traffic

Mixed model prediction process is as follows:

(1) The collection of the network traffic data.

(2) Pre-processing the data of network traffic, removing some useless data.

(3) Using wavelet analysis to decompose the network traffic into linear and nonlinear two parts.

(4)Using ARIMA model to predict the linear part of network traffic, thus obtaining a linear variation of network traffic.

(5) Using BP neural network model to predict the nonlinear part of the network traffic , thus Obtaining non-linear changes of network traffic.

(6)Reconstructing the linear part and nonlinear part of network traffic, to obtain hybrid forecasting model for network traffic prediction.

4. Simulation

4.1 Simulation Data

This network traffic data comes from http: / / newsfeed.ntcu.net / ~ news/2008 /, which has an access to 1000 network traffic point, that is shown in Figure 4, The data is divided into two parts: the former 950 data as learning samples, after the data as a test of 50 samples and the lastter considerable pretreatment.

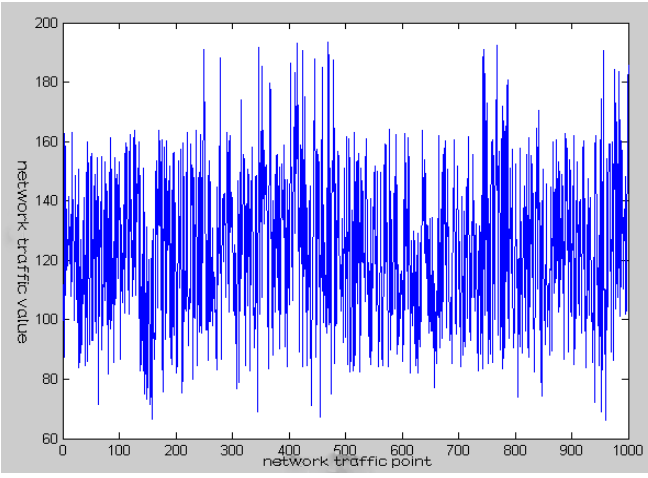

Figure 4: Network traffic data

4.2 Wavelet Decomposition of Network Traffic

Applying Wavelet analysis into network traffic data, decomposition is shown in Figure 4. Which shows that after the Wavelet decomposition. the network traffic is divided into the nonlinear and linear two parts, to facilitate the next step in the ARIMA model and BP neural network.

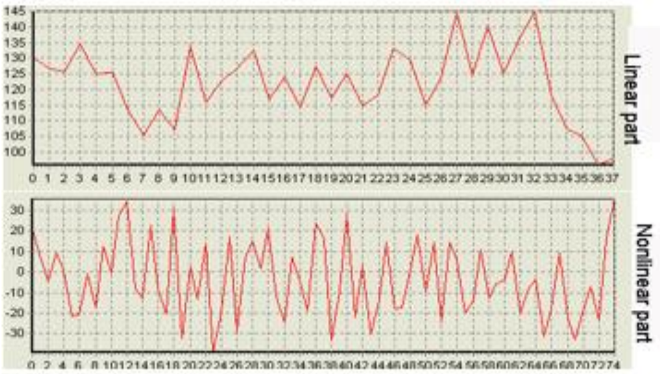

Figure 5 Wavelet decomposition of network traffic

4.3 Results and Analysis of Network Traffic Forecast

ARIMA models and BP neural network model are used as the reference, The prediction of the are shown in Figure 6. Which shows the prediction , This prediction model fit the actual network traffic well, the prediction is accurate. and The ARIMA model of BP neural network prediction accuracy rate is reduced, Which is mainly due to the effective decomposition of wavelet analysis of network traffic, and ARIMA model and BP neural networks variation are used to the linear and nonlinear parts which improves the accuracy. Comparative results show that the hybrid model is an effective, high-accuracy prediction model of network traffic.

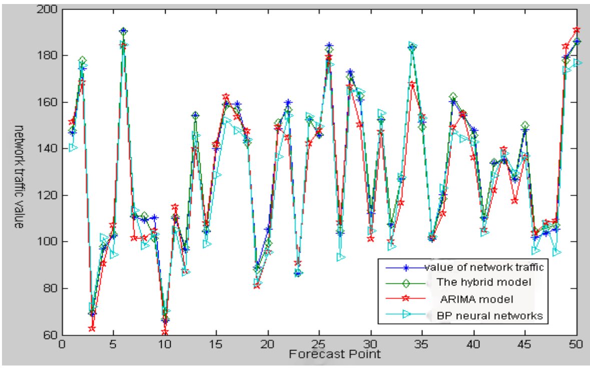

Figure 6 Predictions compared of the three model

5. Conclusion

With the rapid development of computer network,the size and complexity of network are increasingly larger and lager, the likelihood of various problems are greater, network performance is more likely to be affected ,and the difficulty of network management are also increasing, in order to provide quality services to the user, the network maintenance and management is particularly important. In this paper ,a method of network traffic prediction based on wavelet transform, BP nerual network and autoregressive model is proposed.firstly the paper anlysised that.The changes in network traffic have the features of complex, nonlinear, and a high degree of self-similarity. The prediction is not quite accurate using the traditional linear model Secondly,According to the network traffic changes ,this paper proposed a hybrid prediction model based on wavelet analysis, BP nerual network and autoregressive model. At last, the real network traffic data is used to fit the model, the Simulation results show that compared with the simple prediction model, hybrid prediction model improves the prediction accuracy of network traffic, it can well describe the trend of network traffic, which has a wide range of applications.

Reference

[1] R.G.Clegg. A Practical Guide to Measuring the Hurst Parameter[J] .International Journal of Simulation:Systems,Science & Technology, 2006, 7 (2) :3-14

[2]Liu Jiakun ,Jin Zhigang; Traffic prediction and its application using FARIMA models. Journal of electronics and information technology,vol23 no4,apr 2001.

[3] Duan zhi-bin, Sun En-chang, Zhang Yan-hua, Dong Yan .The Prediction of Network Traffic Based on ARMA, Journal of China Academy of Electronics and Information Technology, Vo.l 4 No.4,Aug. 2009.

[4] Feng Hai-liang, Chen d,i Lin Qing-jia, CHEN Chun-xiao, Combined prediction model of Internet traffic based on neural network, Journal of Computer Applications .Vo.l 26 No. 9,Sep.t 2006.

[5] Cheng Guang, Gong Jian, Periodicity Research on Traffic Macro-behavior in a Large-Scale Network. Journal of MINI-MICRO SYSTEMS .Vol.24 No.6 June 2003.

[6] Zou Yue, Chen Qi-yu. Study on Network Traffic Prediction Base on Improved Back Propagation Wavelet Neural Network. Computer Simulation June, 2010. 4.

[7]Yul, K, m C A Novel Shor Teml Load Forecasting Teclinique Using WaveletTransform Analysis .Eleetnema chinesand Power Systems, 2000, 28:537−549.

[8]EricJ.Stollnitz, TonyD.DeRose, and David H.Salesin.Wavelets forcomputer GraPhics : AP--rimer. IEEEComputer Graph, es and APPlicatlons, 15(3):76−84, May 1995(Partl), and15(4): 75−85, Julyl995(PartZ).

[9] Wang Shouren, Chen Zhao guo. Estimation of the order of ARMA model by linear Proeedures. Chinese Ann. of Math.1985,6B:5370.

[10] Zhang daiyuan, The new theory and method of Neural network Beijing, Tsinghua university press, 2007.

[11] Lapdes A,Farber R N Nonlinear signal proeessing Using Neural Networks. Prediction and System Modelling. IEEE Transactions on Neural Networks, 1995, 4(2):169−171.

[12] PeterZ, MinQ.Neural Network Forecasting for Seasonal and trend Time Series, European Joumal of Operational Rcsearch, 2005, 16:501−514.