TELKOMNIKA, Vol. 11, No. 12, December 2013, pp. 7309~7315
e-ISSN: 2087-278X

Ensemble of Differential Equations using Pareto Optimal for Traffic Forecasting

Bin Yang[1], Yuehui Chen[2], Mingyan Jiang*[1]
[1]School of Information Science and Engineering, Shandong University, Jinan, P.R. China 250100
[2]Computational Intelligence Lab. School of Information Science and Engineering, University of Jinan, 106 Jiwei Road, 250022 Jinan, P.R. China
*Corresponding author, e-mail: jiangmingyan@sdu.edu.cn

Abstract

The formal and empirical proof is that the ensemble of the learning models performs better than the single one. In order to construct the ensemble of the system of ordinary differential equations (ODEs), the two problems (diversity and accuracy of ODEs) are considered. In the paper, we estimate experimentally the model ensemble using multi-objective optimization. This paper presents a pareto optimal approach for identifying a family of the additive tree models which are used to reconstruct and identify the system of ordinary differential equations to predict the small-time scale traffic measurements data. We employ the tree-structure based evolution algorithm and particle swarm optimization (PSO) to evolve the architecture and the parameters of the additive tree model. The small-scale traffic measurements data is used to test ODE ensemble, and experimental results reveal that the proposed method is feasible and efficient for forecasting the time series.

Keywords: *multiobjective optimization, Pareto optimal approach, the additive tree models, ordinary differential equations, small-scale traffic*

1. Introduction

The reasonable mathematical models based on the observed time series data are used to provide system analysis and prediction in every area [1-7]. Mathematical modeling is the art of translating problems from an application area into tractable mathematical formulations, whose theoretical and numerical analysis provides insight, answers, and guidance useful for the originating application [1]. The system of differential equations can describe the dynamic properties of a system, which changes with time quite well and predicts the future states of the system very conveniently. In our previous research [8], we used a hybrid evolutionary method to forecast the traffic measurements data, in which the tree-structure based evolution algorithm and particle swarm optimization were employed to evolve the architecture and the parameters of the additive tree models for the system of ordinary differential equation identification. The result demonstrated that the ODE was a powerful predictor model in the discovery of sciential laws for dynamic data

The formal and empirical proof is that the ensemble of the learning models performs better, and more robust/reliableof than the single one. Moreover a single learner is impossible to achieve because of the fact that if a model does not perform well, it will not perform well whereas with an ensemble of models, there is always a chance that some subset of the models will work even if other are not performing well at all [9]. In order to construct the ensemble of the learning model, the two problems are considered. One is the diversity and the other is the accuracy of the learners that comprise the ensemble. In general one objective can just be improved at the expense of at least another objective. The aim is to search a trade-of of diversity and accuracy in a multi-objective evolutionary setup. Recently, many researches focused on the construction of ensemble using artificial neural networks (ANNs) called neuro-ensemble. Kottathra et al firstly proposed that the neural network learning problem was a multiobjective problem. The two objectives were the mean square error and number of hidden units in the network, respectively [10]. But the author could not generate the pareto-optimal set. Wang et al. [11] used the multi-objective neural network approach to reconstruct the image. The authors used the weighted sum method to combine the two objectives which were the

Received April 7, 2013; Revised June 25, 2013; Accepted August 14, 2013

smoothness of the image and the cross entropy real and reconstruction data. Abbass [12-14] proposed pareto-frontier differential evolution (PDE) and the memetic Pareto artificial neural network (MPANN) algorithm which were based on differential evolution for continuous optimization. The ensemble was formed from all networks on the Pareto frontier. Chandra et al. [9] proposed DIVACE (Diverse and Accurate Ensemble Learning Algorithm) which combined good ideas from Negative Correlation Learning (NCL) [15] and Memetic Pareto Artificial Neural Network (MPANN) [12]. The author formulated the ensemble learning problem as a multi-objective aiming at finding a good trade-off between diversity and accuracy.

The additive tree model is a new representation scheme for the system identification especially the reconstruction of polynomials and the identification of linear/nonlinear systems proposed by us [16]. This model is robust, and easy to be analyzed by traditional techniques whose computational complexity is similar to the GP, so we use the additive tree models to identify and reconstruct the system of ordinary differential equation. In this paper, we propose a Pareto optimal approach for identifying a family of the additive tree models to predict the small-time scale traffic measurements data. In this approach, to achieve a trade-of of diversity and accuracy, we take in ideas from NCL algorithms which can be seen as one of the recent well-tested work and the negative correlation penalty function is used to quantify diversity as the second objective of the multi-objective optimization. We employ the tree structure based evolution algorithm and particle swarm optimization (PSO) to evolve the architecture and the parameters of the additive tree models. The partitioning [17] is used in the process of identification of structure of system. The ODEs in the pareto-optimal set by the approach are integrated to predict the traffic data.

2. Research Method

2.1. Representation of Additive Tree Model

Two instruction/operator sets I_0 and I_1 are used to generate the additive tree in this approach (Figure 1).

$$\begin{aligned} I_0 &= \{+2, +3, ..., +N\} \\ I_1 &= F \bigcup T = \{*, /, \sin, \cos, \exp, r\log, x, R\} \end{aligned} \quad (1)$$

Where F = {*, /, sin, cos, exp, rlog*} and T = {x,R} are function and terminal set. +*N*, *, /, *sin, cos, exp, rlog, x,* and *R* denote the addition, multiplication, protected division ($\forall x, y \in R$: when $y = 0$, $x/0 = 1$), sine, cosine, exponent, protected logarithm($\forall x \in R$, $x \neq 0$: $rlog(x) = log(abs(x))$ and $rlog(0) = 0$), system inputs, and random constant number, taking *N*, 2, 2, 1, 1, 1, 1, 0 and 0 arguments respectively [16]. *N* is an integer number (the maximum number of an ODE terms), I_0 is the instruction set and the root node, and the instructions of other nodes are selected from the instruction set I_1. Note that if the right-hand side of ODEs is the polynomial, the instruction set I_1 can be defined as $I_1 = \{*2, *3, ..., *n, x_1, x_2, ..., x_n, R\}$.

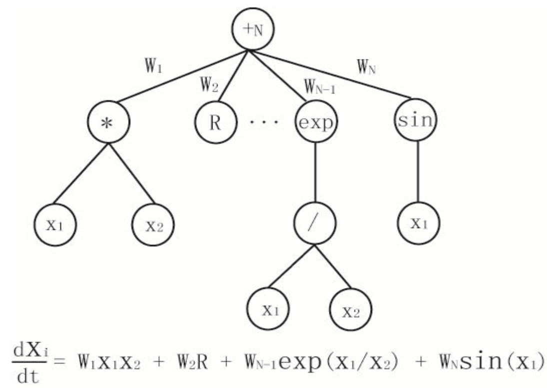

Figure 1. Example of ODEs in the Form of the Additive Tree Model

2.2. Parametes Optimization of Models using PSO

According to the Figure 1, we check all the parameters contained in each equation, namely count their number n_i (i=1, 2,...,N, N is the number of the equations).

According to the number of parameters of each tree model, the particles are randomly generated initially. Each particle x_i represents a potential solution. A swarm of particles moves through space; with the moving velocity of each particle represented by a velocity vector v_i. At each step, each particle is evaluated and keeps track of its own best position, which is associated with the best fitness it has achieved so far in a vector $Pbest_i$. The best position among all the particles is kept as *Gbest* [18, 19]. A new velocity for particle *i* is updated by:

$$v_i(t+1) = v_i(t) + c_1 r_1 (Pbest_i - x_i(t)) + c_2 r_2 (Gbest(t) - x_i(t)) \quad (2)$$

Where c_1 and c_2 are positive constant and r_1 and r_2 are uniformly distributed random number in [0,1]. Based on the updated velocities, each particle changes its position according to the following equation:

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (3)$$

2.3. Negative Correlation Learning

The additive tree model is evolved to identify the ODE to predict the time series. The ODE ensemble, i.e. the additive tree model ensemble, not only need the accurate ODEs, but also need be uniformly distributed on the pareto optimal front, i.e. diversity is catered for. Remarkably, the optimization process should lead to convergence to the pareto optimal front while at the same time maintaining as diverse about distribution of solutions as possible on the pareto front [9]. So we take in ideas from NCL algorithms which can be seen as one of the recent well-tested work. NCL method treats the diversity as a separate objective and the negative correlation penalty function is used to quantify it.

In NCL, for each ODE *i* in the ensemble, the accuracy is defined as followed.

$$(Minimise) Accuracy_i = \frac{1}{N} \sum_{k=1}^{N} (\overline{y}_k^i - y_k)^2 \quad (4)$$

Where N is the number of train data, $\overline{y}_k^i$ is the output of the ODE i at the time point k, and y_k is the actual output.

The negative correlation penalty function which is used as the second objective in NCL, is described as followed. Let N be the number of training data point and let there be M members in the ensemble.

$$(Minimise) Diverse_i = \sum_{k=1}^{N} (\overline{y}_k^i - e_k) \left[\sum_{j=1, j \neq k}^{N} (\overline{y}_j^i - e_k) \right] \quad (5)$$

Where e_k is the output of the ensemble at the time point k, $\overline{y}_k^i$ is the output of the ODE i at the time point k. The above equation indicates the difference from the other members in the ensemble.

2.4. Summary of Pareto Optimal Approach

Integrating with NCL method, to achieve a trade-of of diversity and accuracy, pareto additive tree model evolution (constructing ensemble of ODEs using multi-objective optimization) algorithm is described in detail as followed.

(1) Create the initial population randomly (size M, structures and their corresponding parameters of the additive models);

(2) Apply PSO to all the individuals in the population;

(3) Evaluate all the individuals in the population using the two objectives as described in subsection 2.3, and label the non-dominated set. In the phrase of the training, we take all the population as the ensemble, but for testing, we only use pareto set as the ensemble.

(4) Delete the dominated individuals in the population.

(5) To produce the child, we use the two operators until population size is *M*. 1) Crossover operator: select randomly the two non-dominated individuals from the population, and according to the predefined crossover probability *Pc* select one nonterminal node in the hidden layer for each additive tree randomly, and then swap the selected subtree. 2) Three mutation operators: (a) for a child, randomly select one terminal node in the tree and replace it with another terminal node, which is generated randomly; (b) randomly select one terminal node in the tree and replace it with another terminal node, which is generated randomly; (c) randomly select a function node in a tree and replace it with a terminal node selected in the set *T*.

(6) Apply PSO to all the individuals in the population;

(7) If maximum number of generations is reached or a satisfactory solution is found, then stop; otherwise go to step (3). Finally, we obtain the pareto set including the diverse and accuracy enough members to predict the time series.

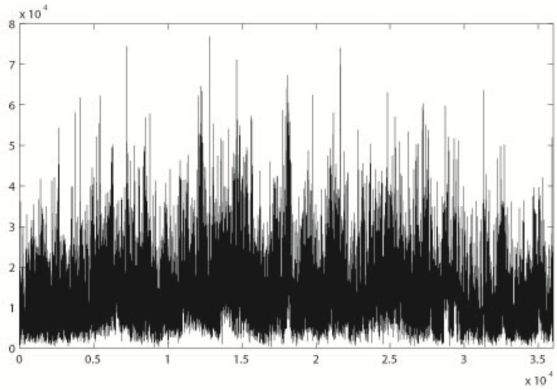

Figure 2. Actual Traffic Measurements Data

3. Results and Analysis

To test the effectiveness of the proposed method, we uses the TCP traffic data, which is published by the Lawrence Berkeley Laboratory. This traffic data contain an hour's worth of all wide-area traffic between Digital Equipment Corporation and the rest of the world. The data package used in this paper is DEC-Pkt1, and the time stamps have millisecond precision (http://ita.ee.lbl.gov/).The traffic data aggregated with time bin 0.1s, that is the arrived package's amount within the 0.1s time interval, are shown in Figure 2.

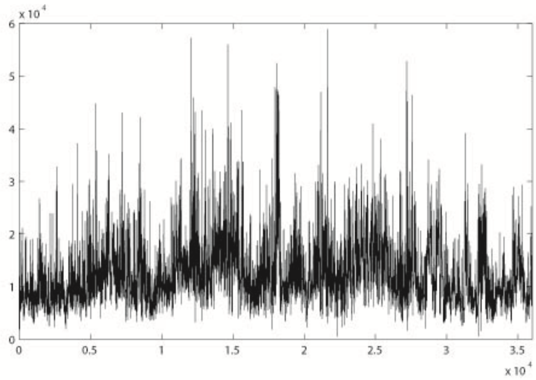

Figure 3. Filtered Traffic Measurements Data

In general, the traffic measurements can be considered as a sum of a regular process and a stochastic part which are related to the high-frequency noise. The elimination of the noise may simplify the analyzed time series, so we apply the wavelet soft threshold noise reduction method to this data. The difference between the original time series and the filtered signal, corresponds to the noisy component. Figure 2 presents the original traffic series, Figure 3 presents the corresponding filtered signal and Figure 4 presents the noisy component. Then the filtered traffic measurements data are normalized to the interval [0, 1] with following formula.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

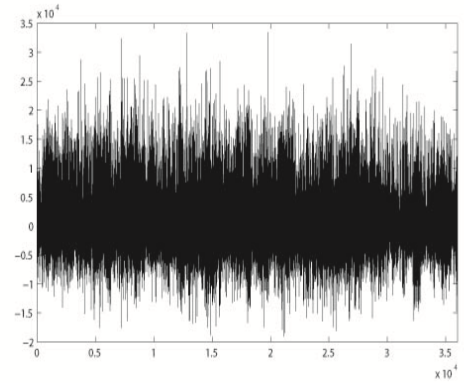

Figure 4. Noisy Component

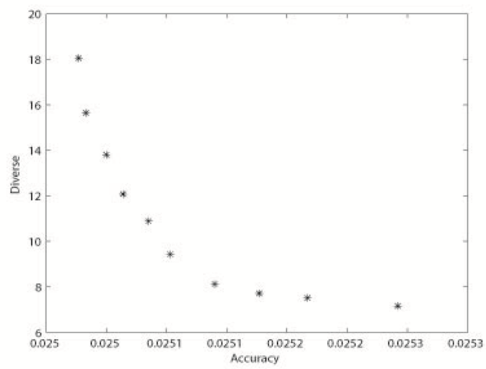

Figure 5. Pareto-optimal Set Obtained using our Approach

The past research proves that the network traffic time series possess of the nonlinear nature [20]. The length of this traffic measurements data is 36000. The front 33000 data points are used as the training set and the last 3000 data points are used as the test set. The population size is set as 30, and the maximum number of generations is set as 50. The ordinary differential equation is solved using the approximate forth-order Runge-Kutta method. We use the 10 input variables to construct an ordinary differential equation model. Namely we use the front 10 variable to predict the current variable. The used instruction set I_0 = {+2, +3, +4, +5, +6, +7, +8} and I_1 = {*, /, log, exp, rlog, sin, cos, X_1, X_2, X_3,…, X_{10}}. We use Equation (7) as the ensemble method of the predictors (where the α_k are chosen to minimize the root mean square error between the ODE outputs and the desired values, and in this paper, the optimal weights of the ensemble predictor are optimized by using PSO algorithm). Through the experiment, we can obtain the Pareto optimal solutions for prediction network traffic measurements data whose two objective values are shown in Figure 5. The time series predicted by the ODEs ensembles is shown in Figure 6 along with that of the target. And the difference of actual network traffic time series data and the predicted ones is illustrated in Figure 7. From Figure 6 and Figure 7, we can clearly see that the system of ordinary differential equation ensembles using Pareto optimal approach can effectively predict the traffic data, and the result is well, the error is very low. And from Figure 8, which depicts the statistical histogram of the absolute difference between the actual time series and the predicted data by our method, it can be clearly seen that the prediction error of the ODE model ensemble to the traffic data mainly concentrates on the vicinity of zero.

In our previous research [8], we have demonstrated that the system of ODE was more powerful predictor model compared with the feedforward neural network. To test the validity of the Pareto optimal approach, we employ the tree-structure based evolution algorithm and particle swarm optimization to evolve the architecture and the parameters of the additive tree models for identification of the system of ODE.We select the 10 better ODE models in the term of accuracy. The ensemble of 10 better ODE models (simple ensemble) are used to predict the traffic data and the results are listed in Table 1 compared with our result. As evident, the prediction performance of ODE ensemble using pareto optimal approach is better.

$$f = \sum_{k=1}^{N} \alpha_k f_k(x) \tag{7}$$

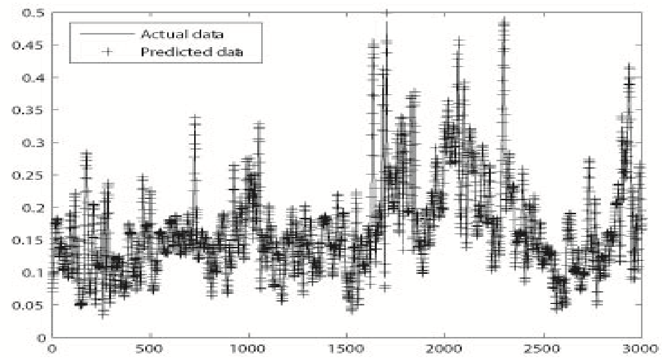

Figure 6. Comparision of Actual Time Series and Predicted Ones

Table 1. The Results among our Method, Simple Ensemble and Single ODE

	our pareto optimal	simple ensemble	single ODE
RMSE for testing data	0.0087956	0.0119124	0.012454

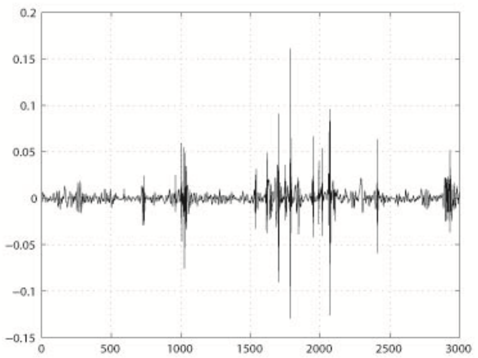

Figure 7. Errors of Actual Time Series and Predicted Ones

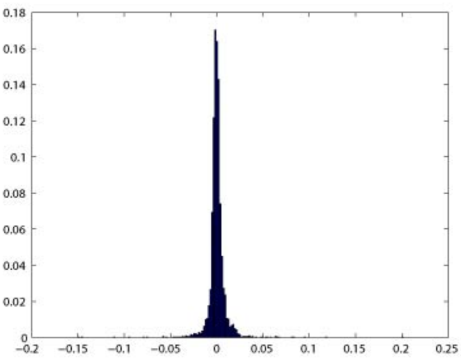

Figure 8. The Statistical Histogram of the Predicted Errors

4. Conclusion

In this paper, a Pareto optimal approach for identifying a family of the additive tree models to predict the small-time scale traffic measurements data is proposed. In this approach, to achieve a trade-of of diversity and accuracy, the negative correlation penalty function is used to quantify diversity as the second objective of the multi-objective optimization. Tree-structure based evolution algorithm and particle swarm optimization are employed to evolve the architecture and the parameters of the additive tree models for system of ordinary differential equation identification. The experiment results clearly illustrate that the ODE model ensemble using Pareto optimal approach can effectively predict the traffic measurements data, the prediction accuracy is well, the prediction error mainly concentrates on the vicinity of zero and the Pareto optimal approach we proposed is very effective by comparison.

Acknowledgements

This research is partially supported by the Natural Science Foundation of China (60573065), the Natural Science Foundation of Shandong Province (No.ZR2010FM040), the Natural Science Foundation of Province (Y2007G33), the Shandong Distinguished Middle-aged and Young Scientist Encourage and Reward Foundation of China (BS2009SW003), and the Key Subject Research Foundation of Shandong Province.

References

[1] Neumaier A. Mathematical Model Building, Chapter 3 in: Modeling Languages in Mathematical Optimization (J. Kallrath, ed). Boston: Kluwer. 2004.
[2] Kou G, Peng Y, Chen ZX, Shi Y. Multiple criteria mathematical programming for multi-class classification and application in network intrusion detection. *Information Sciences*. 2009; 179(4): 371-381.
[3] Sundar S, Naresh R, Misra AK, Shukla JB. A nonlinear mathematical model to study the interactions of hot gases with cloud droplets and raindrops. *Applied Mathematical Modelling*. 2009; 33(7): 3015-3024.
[4] Sadrnezhaad SK, Ferdowsi A, Payab H. Mathematical model for a straight grate iron ore pellet induration process of industrial scale. *Computational Materials Science*. 2008; 44(2): 296-302.
[5] Guin RPF. Pear drying: Experimental validation of a mathematical prediction model. *Food and Bioproducts Processing*. 2008; 86(4): 248-253.
[6] Braghin F, Lewis R, Dwyer-Joyce RS, Bruni S. A mathematical model to predict railway wheel profile evolution due to wear. *Wear*. 2006; 261(11-12): 1253-1264.
[7] Cuartas-Uribe B, Vincent-Vela MC, lvarez-Blanco S, Alcaina-Miranda MI, Soriano-Costa E. Prediction of solute rejection in nanofiltration processes using different mathematical models. *Desalination*. 2006; 200(1-3): 144-145.
[8] Chen YH, Yang B, Meng QF, Zhao YO, Abraham A. Time-series forecasting using a system of ordinary differential equations. *Information Sciences*. 2010; 181(1): 106-114.
[9] Chandra A, Yao X. *DIVACE: Diverse and Accurate Ensemble Learning Algorithm*. In Proc. 5th Intl, Conference on Intelligent Data Engineering and Automated Learning. Exeter. 2004; 3177: 619-625.
[10] Kottathra K, Attikiouzel Y. A novel multicriteria optimization algorithm for the structure determination of multilayer feedforward neural networks. J*ournal of Network and Computer Applications*. 1996; 19: 135-147.
[11] Wang Y, Wahl F. *Multi-objective neural network for image reconstruction*. IEEE Proceedings-Visions, Image and Signal Processing. 1997; 144(4): 233-236.
[12] Abbass HA. *A memetic pareto evolutionary approach to artificial neural networks*. Proceedings of the 14th Australian Joint Conference on Artificial Intelligence (AI01). Berlin. 2001; 1-12.
[13] Abbass HA. Speeding up back-propagation using multiobjective evolutionary algorithms. *Neural Computation*. 2003; 15(11): 2705-2726.
[14] Abbass HA. *Pareto neuro-evolution: Constructing ensemble of neural networks using multi-objective optimization*. In Proceedings of the the IEEE Congress on Evolutionary Computation (CEC2003). Canberra. 2003; 3: 2074- 2080.
[15] Liu Y, Yao X. Ensemble learning via negative correlation. *Neural Networks*. 1999; 12(10): 1399-1404.
[16] Chen YH, Yang J, Zhang Y, dong JW. Evolving Additive tree models for System Identification. *International Journal of Computational Cognition*. 2005; 3(2): 19-26.
[17] Bongard J, Lipson H. Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Science*. 2007; 104(24): 9943-9948.
[18] Zhou TP, Sun W. MPPT Method of Wind Power Based on Improved Particle Swarm Optimization. *TELKOMNIKA*. 2013; 11(6): 3206-3212.
[19] Yan XS, Wu QH, Liu HM. Orthogonal Particle Swarm Optimization Algorithm and Its Application in Circuit Design. *TELKOMNIKA*. 2013; 11(6): 2926-2932.
[20] Shang P, Li X, Kamae XS. Chaotic analysis of traffic time series. *Chaos Solitons Fractals*. 2005; 25: 121-128.