Short-term Traffic Flow Forecasting Based on Wavelet Transform and Neural Network

Liwei Ouyang[1,2], Fenghua Zhu[1,3] *(Correspondence Author) Member, IEEE*, Gang Xiong[3,4], *Senior Member, IEEE*, Hongxia Zhao[1], Feiyue Wang[4] *Fellow, IEEE,* Taozhong Liu[5]

1. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

2. School of the Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, 710049, China

3. Cloud Computing Center, Chinese Academy of Sciences, Dongguan, 523808, China

4. Beijing Engineering Research Center of Intelligent Systems and Technology, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

5. Joint Laboratory of Parallel Management & Control and Business Intelligence, Hainan Zhongke Flower Ocean, Cloud Commerce Technology Co. Ltd, Haikou, 570311, China

***Abstract*—A new combinatorial algorithm based on the characteristics of wavelet transform, particle swarm optimization algorithm and BP neural network for short-term traffic flow forecasting was presented in this paper. Firstly, using wavelet transform, we do multi-resolution decomposition of traffic flow data and single branch reconstruction on its original scale, then apply the particle swarm optimized BP neural network to forecast on each reconstruction sequence, at last, add all the forecasting results to the final short-term traffic flow forecasting results. The experimental results show that, comparing to the traditional particle swarm optimized BP neural network and traditional BP neural network, the new algorithm significantly improves the forecasting accuracy of short-term traffic flow, has a broad application prospect and is more suitable for forecasting short-time traffic flow data which contains much more noise or has a serrated signal curve**.

Keywords—intelligent transportation system; short-term traffic flow forecasting; wavelet transform; short-time traffic flow; neural network; particle swarm optimization

I. INTRODUCTION

Short - term traffic flow forecasting is the key link to realize the intelligent transportation system, and its prediction result will directly affect the effect of urban traffic control and induction. The existing predictive models have achieved some good results in simulation and application, and they can be roughly divided into different models based on traditional statistical theory, neural network model and nonlinear theory [1]. However, the traffic system is a random system with strong uncertainty and complexity, a large number of uncertain factors cause short-term traffic flow to present highly complex nonlinear characteristics. For these reasons, it's difficult to improve the precision of single prediction mode or to expand the scope of application. And then, the combined forecasting model with the advantages of different methods become a better choice.

Wavelet analysis is a powerful tool for signal processing and the perfect crystallization of Fourier analysis, functional analysis, spline analysis, harmonic analysis and numerical analysis. It has such abilities like stepwise fine and feature extraction for no stationary random signals [2]. Neural network has a good robustness, fault tolerance, strong self-learning and adaptive ability, is widely used in pattern recognition, function approximation and other fields. How to choose a proper way to combine their advantages to improve the accuracy of traffic flow prediction has been the concern of many scholars [3]. The existing methods often use the wavelet transform to filter the high frequency signal, and then use the neural network to predict left signal to improve the fitting accuracy [4]. Although, the prediction accuracy can be improved to a certain extent in this way, the advantages of the wavelet transform are not fully utilized. He Guo-guang[5] proposed a new idea of wavelet transform pretreatment for this problem, but only gave the theoretical explanation, has not attempted to combine it with the neural network. Therefore, on the basis of the existing work[5], we attempt to combine the wavelet transform with the neural network, consider some useful high frequency signals, and put forward a new traffic flow forecasting idea.

II. ALGORITHM

The algorithm consists of two stages. The first stage uses wavelet transform to decompose and reconstruct the traffic flow data. The second stage uses the neural network to predict the reconstructed traffic data. The algorithm includes the following steps:

(1) Use the wavelet denoising algorithm to remove the noise from the original traffic flow, and get the traffic signal with some high frequency signals.

(2) Use Mallat[6] algorithm to traffic flow signal N-level decomposition, and to decompose the basic signal and the different resolution interference signals.

(3) Perform the single branch reconstruction of the basic signal and the interfering signal at the original resolution, respectively, to obtain N + 1 independent time series.

(4) Use particle swarm optimized BP neural network (PSOBP neural network) to train and predict N + 1 independent time series one by one, then the time series prediction results are obtained respectively.

(5) The algebraic sum of the predicted results of each time series is the predicted result of the real traffic flow.

The above algorithm can be summarized and shown in Fig 1.

In the first stage, the essence of the multi-resolution decomposition of the traffic flow based on the wavelet transform principle is to uniquely decompose a set of traffic signals $\mathrm{v}_i \in V_0$ (V_0 represents the set of original traffic flows, $v_k^N \in V_N$ is the approximate information of $\mathrm{v}_\mathrm{i} \in V_0$ on N-resolution) with multiple sets of information into different information subspaces W_1, W_2, …, W_N, V_N. V_N Corresponding to the basic signal that reflects the trend of the nature of the reaction flow, W_j is a set of missing information between two different resolutions of V_j and V_{j-1}, $V_{j-1} = W_j + V_j$, and corresponding to the interference signal (high frequency signal). After a single branch reconstruction on original scale by using Mallta algorithm, they can obtain the separated approximate information and high frequency information on the original scale. When the N-level decomposition and single branch reconstruction of the signal are done, N + 1 independent time series can be obtained. If they are predicted respectively, then we can make full use of wavelet transform's stepwise fine features, and extract useful information at different frequencies. In this way, the high frequency characteristics of signal can be retained, which are often mistakenly classified as noise in the general algorithm. Therefore, the prediction accuracy can be significantly improved.

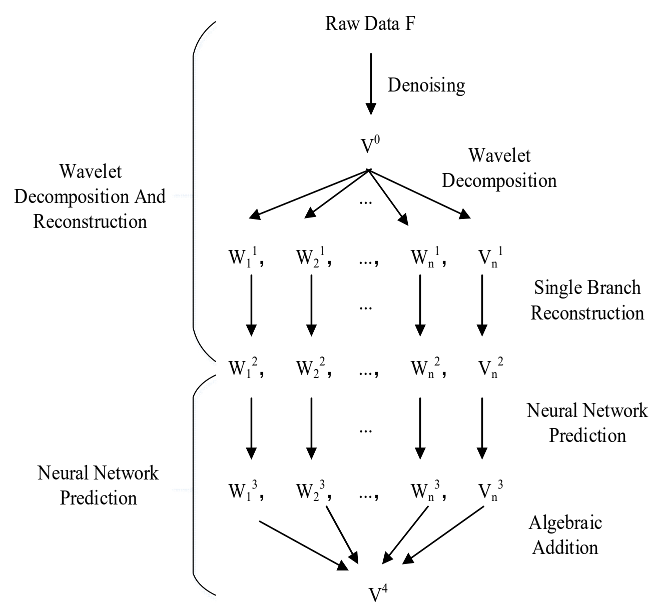

Fig. 1 Algorithm flowchart

In the second stage, we use PSOBP neural network. In this paper, considering the traditional PSO algorithm has some shortcomings as easy to premature convergence, low search accuracy and low iterative efficiency in latter part[7], we introduce mutation operation based on the variation of genetic algorithm[8] to the PSO algorithm. This means some variables will be reinitialized with a certain probability, then the particles can jump out of the current search to the optimal location in a larger space search, so as to avoid them falling into the local extreme point and more likely to find a better solution. In order to obtain higher prediction accuracy, we set the fitness function of the PSO algorithm as the absolute error between the predicted value of the neural network and the real value. The weight threshold with the minimum value of the fitness function, which will be assigned as the initial weight threshold of the BP neural network, can be found by the global search ability of the PSO algorithm with the mutation operator, and then use the BP neural network excellent local optimization and nonlinear approximation capacity for traffic flow prediction.

III. WAVELET DENOISING

The original traffic flow signal often contains a lot of noise, in order to avoid large noise covers the essential signal that can reflect the characteristics of traffic flow, we perform the wavelet denoising on the original signal first.

Considering that the fixed threshold form and the heuristic threshold form remove the noise more thoroughly, the soft threshold processing generally can achieve a more smooth and ideal denoising effect than the hard threshold [9]. In this paper, we use heuristic soft threshold to denoise.

The heuristic method is used to estimate the threshold first. Assume the signal is a discrete time series, k is the time, k = 1, 2,..., n. The threshold obtained by the unbiased likelihood estimation is λ_1, and the threshold obtained by the fixed threshold is λ_2, let:

$$eta = \frac{\|x\|^2 - n}{n} \qquad (1)$$

$$crit = \frac{\left[log(n)/log\,2\right]^{1.5}}{\sqrt{n}} \qquad (2)$$

The heuristic threshold is estimated as follows

$$\lambda_3 = \begin{cases} \lambda_2 & eta < crit \\ min(\lambda_1, \lambda_2) & eta \geq crit \end{cases} \qquad (3)$$

And use the soft threshold to denoise, compare each point's absolute value of the wavelet coefficients ω with the specified threshold value T, if it is less than or equal to T, it becomes zero, otherwise it becomes the difference between the point value and T, the formula is as below:

$$\hat{\omega} = \begin{cases} sgn(\omega)(|\omega| - T) & (|\omega| > T) \\ 0 & (|\omega| \leq T) \end{cases} \qquad (4)$$

$\hat{\omega}$ is the wavelet coefficients obtained after processing.

IV. PARAMETER SETTINGS

In the wavelet transform, the parameters that need to be determined are the wavelet function type and the wavelet decomposition layer. In different application areas, the selection criteria of wavelet function are different, and the factors considered generally include symmetry, orthogonality, tight clipping, smoothness, disappearance matrix order and so on[10]. Considering that the Mallta algorithm is based on dyadic wavelet, it is necessary to have orthogonal wavelet functions such as Coiflet wavelet, Daubechies wavelet, Symlets wavelet and so on. Coiflet has poor compactness and large disappearance matrix orders, it's mainly used in the reconstruction of image. Daubechies is a compacted orthogonal wavelet, but not symmetry. Symlets wavelet constructed by Daubechies wavelet is an orthogonal compact wavelet with approximate symmetric properties, it's an improvement to Daubechies and often used in wavelet denoising. Considering the above factors, this paper chooses SymN as wavelet function for denoising, DbN as wavelet function for wavelet decomposition and single branch reconstruction.

For the selection of the decomposition layer, we generally hope that the wavelet decomposition can not only retain the original data information, but also effectively identify the random error, so the number of wavelet decomposition layer can't be too large or too small, in practice it's difficult to master, and there is no absolute and effective evaluation criteria to help select now. In this paper, considering that the wavelet denoising have to retain a certain degree of high frequency interference information to facilitate the follow-up operation and can't be excessive denoising, the decomposition depth can be less than the best decomposition depth, and for wavelet decomposition and reconstruction, the more the number of layers in the wavelet decomposition and reconstruction, the closer the high frequency information is to the white noise, the less useful information can be obtained, these may cause we mistakenly treat the interference signal in the high frequency information as useful signal, and artificially add the error to the prediction result. Also, because we need to predict all single branch reconstruction signals, more decomposition layers will lead that the data to be processed steep, the algorithm takes too long, the prediction accuracy is not obvious higher, even lower. Based on the above factors, if the best number of decomposed layers can't be found in the algorithm, we can preferentially select smaller decomposition layers

For the PSO neural network algorithm, the parameters need to be discussed are mainly the inertia constant ω, the acceleration constant c_1 and c_2 [7]. This paper does not focus on the influence of the inertia constant on the PSO algorithm, so we select $\omega = 1$. By the literature [11], select $c_1 = c_2 = 2$, the remaining parameters are selected based on the actual data size.

V. EVALUATION CRITERIA

In this paper, the average relative error, root mean square error and equal coefficient are chosen as the evaluation criteria. x_{ireal} is the real traffic flow data after wavelet denoising, and x_{ireal} is the predict traffic flow data.

Average relative error:

$$MRE = \frac{\sum_{i=1}^{N} \left| \frac{x_{ireal} - x_{iforecast}}{x_{ireal}} \right|}{N} \quad (5)$$

MRE is mainly used to measure the overall prediction error, the smaller the MRE is, the higher the prediction accuracy is and the better the prediction effect are.

Root mean square error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(x_{ireal} - x_{iforecast} \right)^2}{N}} \quad (6)$$

RMSE can amplify big errors and reduce small errors, the smaller the RMSE is, the less strange points with great error there are and the better the prediction effect is.

Equal coefficient value:

$$EC = 1 - \frac{\sum_{i=1}^{N} \left(x_{ireal} - x_{iforecast} \right)^2}{\sum_{i=1}^{N} \left(x_{ireal} + x_{iforecast} \right)^2} \quad (7)$$

Equal coefficient value is the geometric characteristic evaluation index that can reflect the fitting degree of the forecasting curve and the actual curve. The higher the value of EC is, the much the predict value of traffic flow is close to the actual observation value, the better the forecasting effect is.

VI. EXPERIMENT AND ANALYSIS

The experimental data is from the Caltrans (California Department of Transportation) PeMS (Performance Measurement System) [12], we get a total of 2016 traffic data from a detector (VDS 313111) recorded at a lane every 5 minutes in the January 30, 2017 00:00 to February 25, 2017, 23:59. Taking into account the difference between the work day and the no-work day, we put a total of 1153 data from Monday to Thursday as the training set, and experiment with the 288 data on Friday as the test set.

Table1 Prediction results of three different algorithms

N=6	Our Algorithm	Particle Swarm Optimized BP Neural Network	BP Neural Network
MRE	0.0230	0.1039	0.1039
RMSE	2.3639	9.2730	9.1860
EC	0.9997	0.9951	0.9952

It can be seen from the above table that the algorithm has a significant superiority compared with the traditional PSOBP neural network and BP neural network, the prediction accuracy is significantly increased, the strange points in prediction result with large error is significantly reduced, and the prediction curve agrees well with the geometric trend of the real curve.

Figure 2 shows the signal curve of the low frequency signal A5 and the high frequency signals D1, D2 and D3 when the number of decomposition layer is 3. It can be seen that the low-frequency signal A5 is the most basic reflection of the real signal, the high-frequency signals D1, D2, D3, although similar to the noise signal image, but we can find they are more stronger when the real signal outline is less clear, it means these high-frequency signals still contain detail and characteristic information of the original signal. As the number of decomposition layers increases, the high frequency signals contain less useful information, lower signal strength, and closer to white noise.

Figure 3 shows the comparison of the prediction curve and the true value curve of three algorithms. It can be seen from Figure 3 that the proposed algorithm has the highest accuracy and the best prediction effect compared with the other two. Even if the experimental data show a strong jagged fluctuations, the new algorithm can fully fit, while the other two can't. It also shows that the traditional algorithm does a large degree of misjudge of useful high-frequency signals in the fitting process and they are the wrong approximation and neglect.

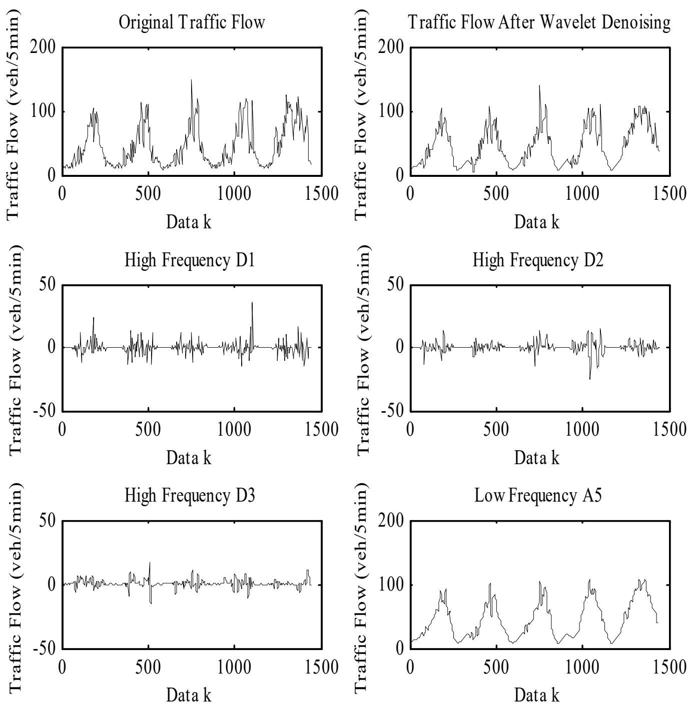

Fig. 2 When decomposition layer is 3, the signal curve of low frequency signal A5 and high frequency signal D1, D2, D3

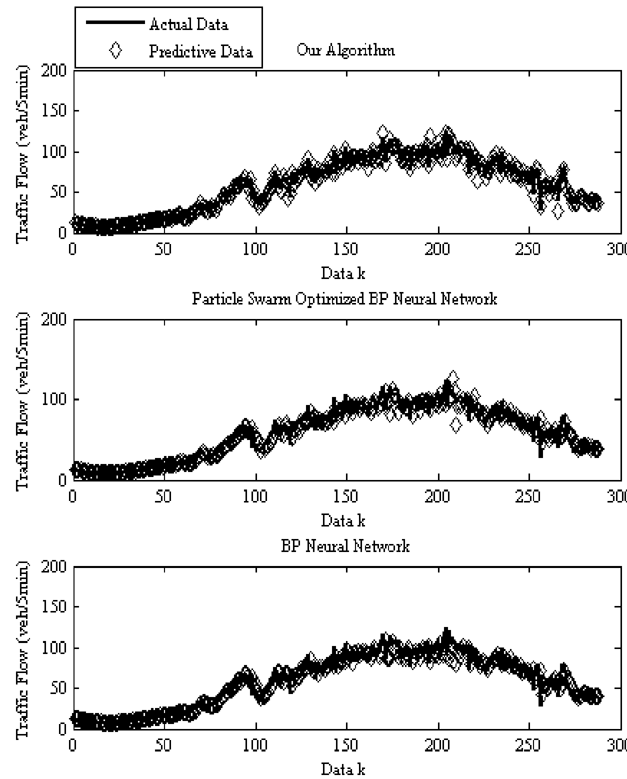

Fig. 3 True value curve and Prediction curve of three algorithms

Table 2 Prediction results of three algorithms for different decomposition levels in wavelet denosing

Algorithm / The Number of Layers: N	**Our Algorithm**		**Particle Swarm Optimized BP Neural Network**		**BP Neural Network**	
	MRE	***RMSE***	***MRE***	***RMSE***	***MRE***	***RMSE***
Original Data	0.0565	3.6122	0.1827	12.2627	0.1839	12.6179
4	0.0313	2.9842	0.1087	9.9825	0.1120	9.3490
5	0.0311	2.8241	0.1071	9.2635	0.1109	9.2513
6	0.0230	2.3639	0.1039	9.2730	0.1039	9.1860
7	0.0275	2.6339	0.0997	9.3526	0.0939	8.8699
8	0.0320	2.9038	0.1029	9.2313	0.1034	9.1566
10	0.0341	2.7149	0.1133	8.4143	0.1164	8.8244
15	0.0879	2.9600	0.2740	8.8238	0.2780	8.5496

Table 2 shows the prediction results of three algorithms, when the number of decomposition layer is different in wavelet denoising. It can be seen from Table 2 that our algorithm has the best prediction effect when the number of decomposing layers is N = 6, and the PSOBP neural network algorithm and the BP neural network algorithm have the best prediction effect when N = 7. When N is increased to 15, the accuracy of our algorithm is significantly reduced, these show that our algorithm is more suitable for predicting traffic flow which preserves some high-frequency "interfering signal" (detail signal) or has the relatively non-smooth signal curve, while the traditional PSOBP neural network algorithm and BP neural network algorithm are more suitable for predicting traffic flow which contains less or no noise or has the relatively smooth signal curve. At the same time, the prediction accuracy of our algorithm is improved obviously compared with the other two under all different levels of decomposition, even if the traffic flow data is not denoised, the prediction result is also quite good. This proves that the algorithm has a wide range of applications, and can get better prediction accuracy, regardless of whether the traffic flow is properly denoised. This feature is especially useful in the practical application.

VII. CONCLUSION

In this paper, based on the methods of wavelet transform, combining the PSOBP neural network, a new short-term traffic flow forecasting algorithm is proposed. Comparing to the traditional PSOBP neural network and BP neural network in experiment, the following conclusions can be obtained:

(1) Compared with the traditional algorithm, our algorithm reduces the error caused by neglecting the high frequency useful information in the training and prediction process of ordinary neural network, significantly improves the prediction accuracy of short-term traffic flow forecasting, and performs well when the traffic flow curve signal is seriously jagged.

(2) Based on its principle, this algorithm is more suitable for short-term traffic flow data with more noise, and retaining detail information. Also, we need pay attention that our algorithm predict each layer of single branch reconstruction signal, and it inevitably increases time-consuming. When the amount of data increases, this feature will be more prominent. Therefore, in practical applications, we can consider to combine our algorithm with cloud computing or other algorithms to carry out large-scale traffic data processing, then time-consuming issues can be solved properly.

ACKNOWLEDGMENT

The work in this paper is supported by the National Natural Science Foundation of China under Grants 61233001, 71232006, 61304201, 61533019, 61773381 and 91520301; 2016 College Students Innovation and Practice Training Program, Chinese Academy of Sciences; Chinese Guangdong's S&T project (2014B090902001, 2014A050503004, 2015B010103001, 2016B090910001).

REFERENCES

[1] Min Guo, Xiang Xiao, and Jinhui Lan, "A summary of the short-time traffic flow forecasting methods," Techniques of Automation and Applications, vol. 28, 2009, pp. 8-9.

[2] Jintai Cui, and Zhengxing Cheng, "An introduction to wavelets," Xi'an: Xi'an Jiao Tong University Press, 1995, pp. 2-24.

[3] Zhe Chen, and Tianjin Feng, "Research advances on combination of wavelet analysis and neural networks," Journal of Electronics, vol. 22, 2000, pp. 496-504.

[4] Yi Liu, and Kun Zhang, "Short-term traffic flow forecasting based on wavelet denoising and GA-Elman neural network," Technology & Economy in Areas of Communications, vol. 17, 2015, pp. 80-85.

[5] Guoguang He, Shoufeng Ma, and Ning Li, "Study on the short-term forecasting for traffic flow based on wavelet analysis," Systems Engineering – Theory & Practice, vol. 9, 2002, pp. 101-106.

[6] Mallat S G, "Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$," Transactions of the American Mathematical Society, vol. 315, 19897, pp. 69-8.

[7] Wei Yang, Qiqiang Li, "Survey on particle swarm optimization algorithm," Engineering Science, vol. 6, 2004, pp. 87-94.

[8] Ratnaweera A, Halgamuge S K, and Watson H C, "Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," IEEE Transactions on Evolutionary Computation, vol. 8, pp. 240-255, Jun 2004.

[9] Ping Li, Weiguo Li, and Zhiguo Wang, "Research on wavelet denoising theory and MATLAB implementation," Modern Business Trade Industry, Vol. 9, 2008, pp. 399-400.

[10] Lvtai Guan, "Wavelet methods and applications," Beijing: Higher Education Press, 2007, pp. 35-37.

[11] Suganthan P N, "Particle swarm optimiser with neighbourhood operator," IEEE Evolutionary Computation. Washington. DC. USA, vol. 3, pp. 1958-1962, July 1999.

[12]http://pems.dot.ca.gov/?report_form=1&dnode=VDS&content=detector_health&tab=dh_timeseries&export=&station_id=313111&s_time_id=1485648000&s_time_id_f=01%2F29%2F2017&e_time_id=1486339140&e_time_id_f=02%2F05%2F2017&html.x=32&html.y=8

[13]F.-Y. Wang, Parallel Control and Management for Intelligent Transportation Systems: Concepts, Architectures, and Applications, IEEE Transactions on Intelligent Transportation Systems, vol. 11, no. 3, pp. 630-638, 2010.

[14]N. Zhang, F.-Y. Wang, F. Zhu, D. Zhao, and S. Tang, DynaCAS: Computational Experiments and Decision Support for ITS, IEEE Intelligent Systems, vol. 23, no. 6, pp. 19-23, 2008.

[15]Y. Lv, Y. Duan, W. Kang, Z. Li, F.-Y. Wang, Traffic Flow Prediction With Big Data: A Deep Learning Approach, IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 865-873, 2015.

[16] Fei-Yue Wang, Xiao Wang, Lingxi Li, Li Li, "Steps toward parallel intelligence", IEEE/CAA Journal of Automatica Sinica, 2016,Vol. 3, No. 4, pp.345-348.

[17]LI Li, LIN Yi-Lun, CAO Dong-Pu, ZHENG Nan-Ning, WANG Fei-Yue. Parallel Learning——A New Framework for Machine Learning. Acta Automatica Sinica, 2017, 43(1): 1-8.