

A New Wind Power Forecasting Algorithm Based on Spark

Shaomin Zhang^{1, a}, Peng Chen^{1, b}, Baoyi Wang^{1, c}

¹ School of Control and Computer Engineering, North China Electric Power University,
Baoding, 071003, China

^aemail: zhangshaomin@126.com, ^bemail: 1247631680@qq.com, ^cemail: wangbaoyiqj@126.com

Keywords: Wind Power Prediction; Spark; Parallelization Improvement; Artificial Intelligence

Abstract. With the arrival of the era of big data, wind power forecasting data develops to the massive and multi-dimensional direction. Single computer lack of resources is difficult to solve the problem of wind power forecasting. In this paper, Spark cloud computing platform is introduced to solve the massive data problem that the traditional single machine can not deal with. The combination of BP neural network with the bacterial colony optimization algorithm can avoid the local optimum and improve the accuracy of forecasting results. In the experiment, the real data of a wind farm in Nei Mongol is used as the input data set and Spark cloud computing programming framework improve the parallel algorithm. The results show that the improved algorithm can increase the efficiency of prediction results about 10%.

Introduction

With the growing prominence of energy, China will vigorously develop wind power as a strategic position. The wind power will gradually rise as the main body of the future national energy structure. But the wind power has a great instability, so how to make the wind power forecasting is the problem of the wind power integration [1].

In order to improve the performance of the neural network, Ma Xiuyuan et al. combined the artificial neural network with the particle swarm algorithm to improve the convergence speed of neural network and avoid local optimal solution [2]. Li Manfeng et al. use genetic algorithm and artificial neural network to predict wind speed [3]. Wang Yan et al. proposed a short term wind forecasting method based on online sequential extreme learning machine OS-ELM [4].

In summary, the neural network can be combined with other optimization algorithms to achieve the purpose of wind power forecasting.

At present, the wind power and the relevant meteorological parameters are recorded in the wind farm, which lead to the PB level of the object data, the sampling data, the wind resource and the geographic data. The advantages of the BCO are mainly reflected in the processing of multi-dimensional data input to get the RMSE. It is not able too fast enough to deal with the input parameters of the PB level. So this paper introduces the Spark cloud computing platform, and designs the neural network algorithm optimized by bacterial colony optimizations (BCO-NN) which realized on the Spark platform to complete the wind power prediction [5].

Spark Cloud Computing Platform

Spark is a big data parallel computing framework based on memory, which is based on the memory. So the Spark improves the real-time performance of data processing of large data [6]. Spark adopt a unified technology stack to solve cloud computing issues on big data such as streaming processing, mapping technology, machine learning, NoSQL query and so on. This series of advantages make Spark gradually become the mainstream of cloud computing big data field [7]. Which is showed in figure 1.

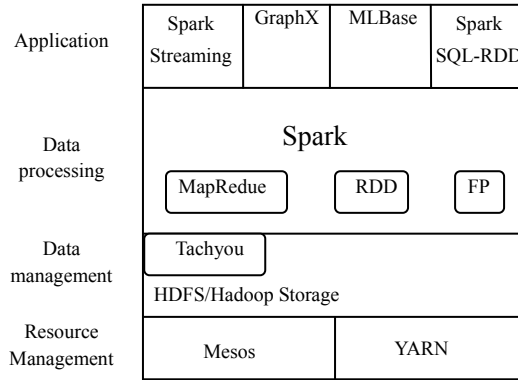


Fig.1. Spark overall architecture diagram

Spark distributes by Spark Streaming. Firstly, Spark Streaming decomposes the flow into a short batch job and then into a section of data according to the size of the size batch. Each section of data are converted into RDD. Then Transformation operation for Streaming DStream becomes Transformation for RDD Spark. Finally, RDD are operated into the intermediate results and stored in memory. All the flow calculation can be superimposed on the results of the middle, or stored to the external device through business needs.

BP neural network

BP neural network is the most widely used neural network. Traditional training method of BP neural network is the use of gradient search method to update neural network. And the parameters of BP neural network are adjusted by the output error of the neural network. The forward and backward propagation of the signal is always present in the training process of the whole neural network until the output error reaches the specified accuracy or the number of iterations scheduled [8]. Three story wind power forecasting neural network topology is showed in figure 2.

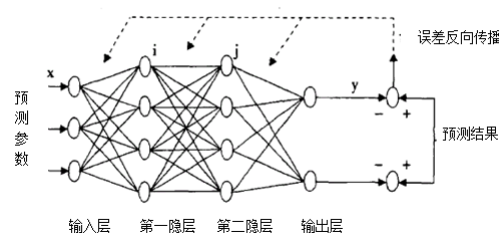


Fig.2. BP neural network topology

Although BP neural network is the most widely used artificial neural network model, BP neural network is also facing a lot of limitations in dealing with optimization problems.

First of all, BP neural network is easy to fall into local optimum for its using the steepest descent method. In addition, the step size and the number of hidden layer nodes are not easy to be determined. This is also the reason that affects the performance of the neural network. Finally, the BP neural network is prone to over fitting.

In this paper, the main problem is to solve the local optimum of neural network. The Spark cloud computing technology and bacterial colony optimization algorithm (BCO) are introduced to improve the accuracy of the neural network to deal with multi-dimensional data and improve the prediction accuracy of wind power. This paper names the algorithm neural network optimized by bacterial colony optimization based on Spark loud computing platform (S-BCO-NN).

S-BCO-NN based on Spark

The BCO algorithm puts forward based on the life cycle of artificial bacteria.

The foraging process of bacterial flora mainly includes the chemotaxis, replication and extinction

and aggregation [9].

The bacterial chemotaxis is divided into two kinds of turning and forward. In the process of feeding, the bacteria in the original direction to find less food, would flip to a new direction. Forward means going along the direction of feed forward flip. The chemotaxis operations can be expressed as:

$$P_i(T) = P_i(T-1) + R_i \times (Ru_{in}) + R \times \Delta(i) \quad (1)$$

$$P_i(T) = P_i(T-1) + R_i \times (Tumb_{in}) + R \times \Delta(i) \quad (2)$$

Here $P_i(T)$ is the position of i -th bacteria at time T ; $Tumb_{in}$ and Ru_{in} is obtained information when bacteria swim and whirl; Δ is a random direction; R_i is bacterial level of competence; R is adaptability of the flora.

Aggregation is that bacteria release an attractive material to attract mates in the process of feeding. If the bacteria found nearby a companion to absorb nutrients, it will release the material with a repulsive force. Through this mechanism, the bacteria group achieves information sharing. The aggregation operation is expressed as

$$J_c(\theta, P(j, k, l)) = \sum_{i=1}^S J_{cc}^i(\theta, \theta^i(j, k, l)) \quad (3)$$

$$= \sum_{i=1}^S [h_{rep} \exp(-\omega_{rep} \sum_{m=1}^P (\theta_m - \theta_m^i)^2)] + \sum_{i=1}^S [-d_{att} \exp(-\omega_{att} \sum_{m=1}^P (\theta_m - \theta_m^i)^2)]$$

Here S is the number of bacteria; $P(j, k, l)$ is the position of the bacteria; d_{att} is the attractive depth; w_{att} is the attractive width; d_{rep} is the repulsive depth; w_{att} is the repulsive width; θ_m^i is the position of i -th bacteria; θ^i is others.

Replication and extinction According to the search ability, each bacterium is given them different energy levels. So we can improve the ability of the whole flora by the ability of the individual to eliminate the ability to replicate the weak.

Neural network are established In each Spark node. And the neural network are trained by these following steps:

Step1 Initialization: The topology of the network are determined according to the input and output sample set of neural network including related parameters. x_{id} means the location of the bacteria; N means the size of the bacteria; N_x is space dimension; K_{max} and K_{min} are the search space.

Step2 Evaluation: The fitness function is used to calculate the fitness of each individual. The most individual is chosen as the global extremum and using Q_{hd} as the marker. And the optimal individual will reproduct in the next step.

Step3 Update extreme: calculate the fitness value of each bacterium f_i , if it is better than the current one Q_{hd} , replace the individual position by P_{hid} . And update the optimal individual.

Step4 Updating weights according to the optimal solution

Step5 Updated bacterial position by Formula (1) and Formula (2).

Step6 Test: If the iteration termination condition has achieved, regard current Q_{hd} as the optimal solution. Otherwise go to Step2.

Finally, the neural network is trained to predict, the forecasting process is shown in Figure 3

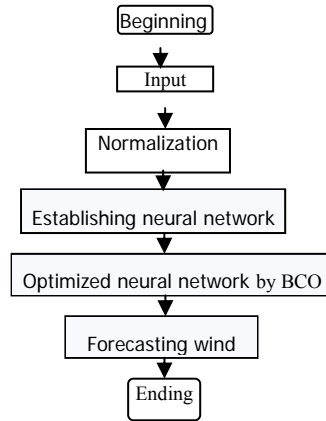


Fig.3. Flow chart of wind power forecast by BCO-NN

Pass the prediction result of each Spark node to the master node. The weights of each neural network are calculated by the of the main nodes [11]. The whole S-BCO-NN algorithm prediction model is shown in Figure 4:

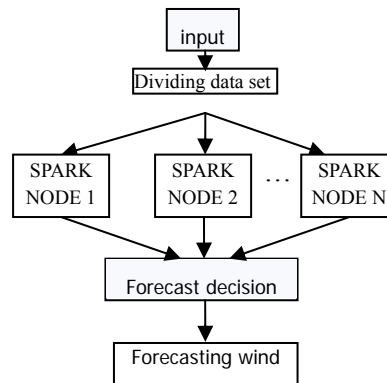


Fig.4. Wind power forecast model based on S-BCO-NN

Steps of the prediction algorithm in Spark platform:

Step1: The wind power forecast of massive data is gotten from the distributed file system of Spark platform. Then Map Parition function of Spark divided the data set into k parts (k is the number of Spark node).

Step2: The map function of each node obtains forecasting result by the neural network algorithm running in the node, which is equivalent to the K neural network.

Step3: The reduce function manipulates prediction from node to get the final result.

Experiment and numerical analysis

The Spark cloud computing platform in the laboratory is composed of 9 nodes, each node is configured as Intel (R) Core (TM) i5-2400 4-core CPU@2.60 GHz, 4 GBRAM, the network bandwidth is 100 Mbit/s. Hadoop version is 0.20.2.

The data sets are selected from a wind power station in Inner Mongolia at 2005 and 2006. They includes average temperature, precipitation, wind speed, the highest temperature, the lowest temperature, the relative amount of radiation, and so on.

The input sample consists of 7 feature vectors. Training set is[Date D, temperature T, rainfall R, evaporation E, radiation F, wind speed W, historical power P].Here $L=[P1, P2, P3, P4, P5, P6, P7]$ means seven days' wind farm power before predict day.

At present, the root mean square error of the wind power prediction result is the standard of measuring the wind power forecast. The RMSE of S-BCO-NN is proved to be more accurate than the traditional wind power forecasting methods.

The formula of RMSE is showed following.

$$S_{RMSE} = P_{cap} \sqrt{\frac{\sum_{i=1}^n (P_{mi} - P_{pi})^2}{n}} \quad (4)$$

Here N is the number of samples; Pcap is the Starting capacity of the wind farm; Pmi is the actual power at i moment; Ppi is forecasting power at i moment.

In addition, the speedup (Speedup) measured under the cloud platform is used to evaluate parallel performance of this method.

$$S_{speedup} = \frac{\text{Execution time stand-alone}}{\text{Execution time based on Spark}} \quad (5)$$

The experimental results of Wind power prediction accuracy based on cloud computing by S-BCO-NN are shown in figure 4. There are two broken lines in the figure. They are actual wind power line and forecasting line by S-BCO-NN algorithm in July 2006.

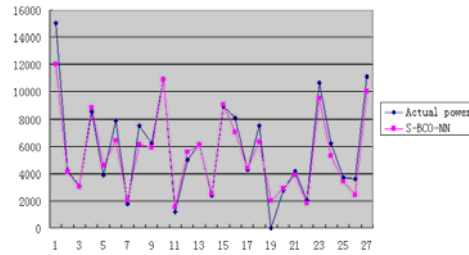


Fig.5. Comparisons of the actual wind power and the predicted ones by S-BCO-NN in July 2006

The RMSE of the result of the S-BCO-NN algorithm is 18.9%. It improves about 10% than existing algorithms [12].

However, with the increase of the input data, the traditional serial processing method is difficult to meet the requirements of wind power forecasting, so this experiment is carried out on the Spark platform.

Speedup is an important criterion to measure the parallel efficiency of a parallel system, Amdahl's law is pointed out following[13].

$$S_{speedup} \leq \frac{N}{NF + (1 - F)} \quad (6)$$

Here N is the number of parallel processors; F is the proportion of serial tasks.

Thus the more excellent parallel computing system, the speedup is more close to N. Then the training data set is extended to 100 times, 200 times, 500 times, times and 1000 times of the original data set. And the performance of the proposed algorithm in the cloud computing platform is measured by the acceleration ratio. This is showed in figure 5.

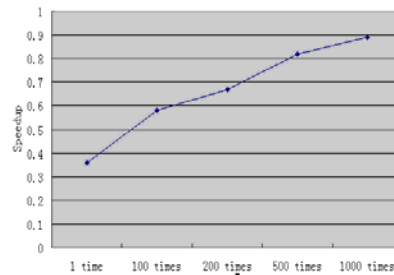


Fig.6. Speedup of S-BC-NN algorithm

It can be seen that with the increase of the amount of data, the acceleration ratio is gradually close to N. When the data volume reached 1000 times the original data, the speedup is more than 6. It shows that the algorithm has good parallel performance on the Spark platform.

Conclusions

Wind power forecasting requires a large amount of data processing. The traditional stand-alone mode to deal with the problem will appear very weak. So this paper introduces the Spark cloud

computing platform and the BCO-NN to predict the wind power. The experimental results show that this algorithm has obvious advantages in the RMSE of the prediction results. The deficiency is that the decision-making method can achieve the best decision in each node weight. It will be further improved in the next time.

Acknowledgement

In this paper, the research was sponsored by the National Natural Science Foundation of China (Project No. 61300040) and Scientific Research Project of Hebei Province (Project No. Z2012077).

References

- [1]ZHU Xingyang, LIU Wenxia, ZHANG Jianhua. Probabilistic Load Flow Method Considering Large-scale Wind Power Integration[J].Proceeding of the CSEE, 2013, 07:77-85+16.
- [2]MA Xiuyuan, DUAN Yufeng, LIU Meng. Prediction of Pressure Drop of Coke Water Slurry Flowing in Pipeline by PSO-BP Neural Network[J]. Proceeding of the CSEE, 2012,05:54-60.
- [3] LI Manfeng,LI Suping,FAN Bo. Research on Solar Collector Simulation Based on Genetic-BP Algorithm[J]. Proceeding of the CSEE, 2012,05:126-130.
- [4]WANG Yan , WANG Zhen , HUANG Minxiang , et al. Ultra-short-term Wind Power Prediction Based on OS-ELM and Bootstrap Method[J]. Automation of Electric Power Systems, 2012, 14:1362-1367
- [5]ZHENG Yingchun.Modification and Application of Bacterial Foraging Optimization Algorithm[D]. South China University of Technology, 2014
- [6]Lizhe Wang,Jie Tao,Rajiv Ranjan,HolgerMarten,Achim Streit,Jingying Chen,Dan Chen. G-Hadoop: MapReduce across distributed data centers for data-intensive computing[J]. Future Generation Computer Systems,2013,293:.
- [7]FENG Lin.Research and Implementation of Memory Optimization Based on Parallel Computing Engine Spark[D]. Tsinghua University,2013.
- [8]Francesco Castellani , Massimiliano Burlando , Samad Taghizadeh , Davide Astolfi , Emanuele Piccioni. Wind Energy Forecast in Complex Sites with a Hybrid Neural Network and CFD based Method[J]. Energy Procedia, 2014, 45:.
- [9]NIU Ben.Bacterial Colony Optimization and Bionic Management[M]. Science Press, 2014, 6.
- [10]ZHANG Haibo, JIANG Liangmin, TAO Wenwei et al. The design and implementation of a practical distributed dynamic power flow calculation system [J]. Automation of Electric Power Systems, 2012,09:67-71+76.
- [11]XIAO Tiwei. Analysis and application of Hadoop distributed computing platform[D].Southwest Petroleum University,2014
- [12]DING Huajie, SONG Yonghua, HU Zechun et al. Probability Density Function of Day-ahead Wind Power Forecast Errors Based on Power Curves of Wind Farms[J]. Proceeding of the CSEE, 2013, 34:136-144+22.
- [13]Lazar Lazić, Goran Pejanović, Momčilo Živković, Luka Ilić. Improved wind forecast for wind power generation using the Etamodel and MOS (Model Output Statistics) method [J]. Energy, 2014, 73:.