TELKOMNIKA, Vol. 11, No. 9, September 2013, pp. 5262~5266
ISSN: 2302-4046

Electricity Consumption Forecasting in the Age of Big Data

Xiaojia Wang
School of Management, Hefei University of Technology, Anhui Hefei 230009
e-mail: tonysun800@sina.com

Abstract

In the age of big data, information mining technology has undergone tremendous change; traditional forecasting mining technology has not been able to solve the information mining problems under a large scale of data. this paper put forward a modeling mechanism of information analysis and mining under the age of big data, the modeling mechanism is, first, construct the model of task decomposition of information by MapReduce tool, then, make data preprocessing and mining operation according to the single task data sheet, use mathematical model, artificial intelligence and other methods to construct the new ideas of information analysis and data mining under the age of big data, finally, a case study presented to demonstrate the feasibility and rationality of our approach.

Keywords: *big data, information mining, electricity consumption, forecasting*

1. Introduction

The age of big data [1] poses a new challenge to the electric power industry development, but it has also brought new opportunities for its development. Production management level of power production, marketing and network operation and maintenance will be improved by using good data management methods and feasible data mining strategy, it will also provide powerful information technology support to Chinese electric power enterprises so as to stand in the advanced ranks in the world.

Under traditional data scale, a lot of methods can be used to forecast the electricity consumption. literature [2] does quantitative research on application scope of GM (1, 1) model and the relation between developing coefficient and prediction accuracy; literature [3] reconstructs the background value using Newton-Cores formula; literature [4] proposed a novel neural-fuzzy short-term load forecasting model based on neural networks and fuzzy logic; literature [5] applied hybrid non-linear models for energy consumption forecasting; literature[6] proposed an adaptive neural–wavelet model for load forecast in the competitive electricity market.

In the age of big data, the environment of electricity consumption forecasting mining has undergone tremendous change, consumption forecasting remains a difficult task because electricity consumption in a region is inevitable influenced by many factors of random nature, such as economic development, population, weather conditions and industrial productions in the region. So, traditional consumption forecasting mining method has not been able to solve the problems under a large scale of data, we must construct a model that can be used to forecast the electricity consumption under the age of big data.

This paper is organized as follows. Section 2 introduces the modeling mechanism of consumption forecasting under the age of big data and present a new forecasting model that can be used for big data. In session 3, the case study is described; finally, the conclusions are presented in session 4.

2. Modeling Mechanism of Consumption Forecasting under the Age of Big Data

2.1. Task Data Decomposition

MapReduce [7] is a programming model and an associated implementation for processing and generating large data sets. Users specify a map function that processes a

Received February 18, 2013; Revised June 9, 2013; Accepted June 21, 2013

key/value pair to generate a set of intermediate key/value pairs, and a reduce function that merges all intermediate values associated with the same intermediate key. Many real world tasks are expressible in this model, as shown in the paper.

The MapReduce program abstracted the data processing task into a series of Map-Reduce operation. Map operation mainly completes the filtering operation of the large scale data, Reduce operation mainly completes the aggregation operation of the large scale data. The input and output data are stored in <key, value> format, the MapReduce framework will make task decomposition automatically and achieve parallel execution. The conceptual model of task decomposition is shown as follows.

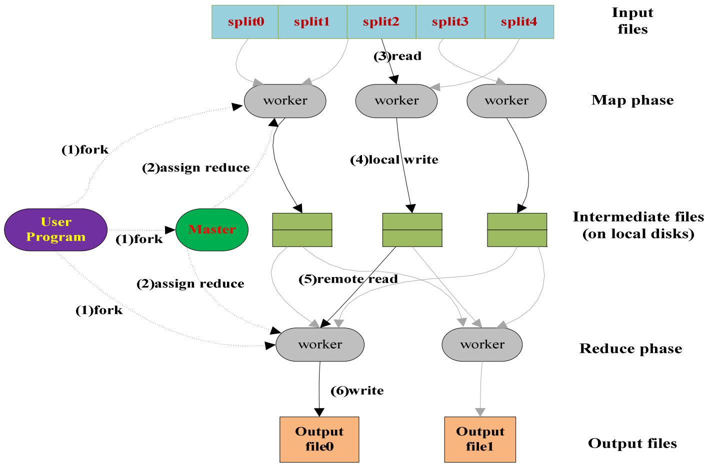

Figure 1. Conceptual model of task decomposition by MapReduce program

2.2. Task Data Mining-consumption Forecasting Modeling

After the decomposition operation, the large scale data sets can be divided into different single small task, but, sometimes, the single task still have large scale, so, we chose neural network algorithm as fundamental method, make a mapping set between each data point in single data set and large training sets consisted by complex neurons. We use support vector machine (SVM) mechanism as an auxiliary method, and optimized the solution of neural network algorithm, improved the hidden layer of neural network.

Step 1: Support J is a big data set, $R(\varpi, \xi_i, \xi_i^*)$ is the zonal region function cover the massive data sets, use SVM algorithm to solve the following problem:

$$\min J = \min R(\varpi, \xi_i, \xi_i^*) = \frac{1}{2} \| \varpi \|^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*) \tag{1}$$

Such that,

$$\begin{cases} d_i - \varpi\phi(x_i) - b_i \le \varepsilon + \xi_i \\ \varpi\phi(x_i) + b_i - d_i \le \varepsilon + \xi_i^* \qquad i = 1, 2, \cdots, N \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{2}$$

Here, ξ_i, ξ_i^* is the slack variable, used to control the linear non-separable boundary. We can obtain a quadratic programming problem:

$$\begin{aligned} \psi(\beta_i, \beta_i^*) = &\sum_{i=1}^{N} d_i (\beta_i - \beta_i^*) - \varepsilon \sum_{i=1}^{N} (\beta_i + \beta_i^*) \\ &- \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\beta_i - \beta_i^*)(\beta_j - \beta_j^*) K(x_i, x_j) \end{aligned} \tag{3}$$

Where $\beta_i, \beta_i^* \in [0, C], i = 1, 2, \cdots, N$, $K(x_i, x_j)$ is Kernel function and $K(x_i, x_j) = \phi(x_i)\phi(x_j)$ $\phi(x_i)$ and $\phi(x_i)$ are images of sample space variable x_i and x_j .

$$f(x) = \sum_{i=1}^{i} \alpha_i \exp(-\|x - c_i\|^2 / 2\sigma_i^2) + b \tag{4}$$

Step 2: Construct the initial structure of neural network by LS-SVR algorithm, use Gauss radial basis kernel function as the hidden layer of neural network, then, we can obtain the new network structure:

$$f(x) = \sum_{i=1}^{i} \alpha_i \exp(-\|x - c_i\|^2 / 2\sigma_i^2) + b \tag{5}$$

Here, c_i is the center of radial basis kernel function that means support vector. l is the number of hidden crunodes, $\|x - c_i\|$ is defined as Euclidean norm. $\sigma = d_m / \sqrt{2l}$ is the width of hidden crunodes, where d_m is the maximum distance between the selected center.

Step 3: In the learning process of RBF network, adjust the center of radial basis function network weight, width and other network parameters, enables the network to achieve optimal performance, this algorithm chose gradient descent algorithm as its training algorithm.

RBF network structure is:

$$J = \frac{1}{2}\sum_{j=1}^{M}(y_j - \hat{y}_j)^2 \tag{6}$$

Here y_j is expected output, $\hat{y}_j$ is network output.

$$\hat{y}_j = \sum_{i}^{t} \omega_{ij} h_i, h_{ji} = \exp(-\frac{\|x_j - c_i\|^2}{\sigma_i^1}) \tag{7}$$

Step 4: Computing the partial derivative ω_{ij}、 c_i、 σ_i of J , we can obtain:

$$\begin{aligned} S_{\omega_{ij}} &= -\frac{\partial J}{\partial \omega_{ij}} = (y_j - \hat{y}_j)h_i \\ S_{c_i} &= -\frac{\partial J}{\partial c_i} = \frac{2h_i(x - c_i)}{\sigma_i^2}\sum_{j=1}^{M}(y_j - \hat{y}_j)\omega_{ij} \\ S_{\sigma_i} &= -\frac{\partial J}{\partial \sigma_i} = \frac{2\|x - c_i\|}{\sigma_i^3} h_i \sum_{j=1}^{M}(y_j - \hat{y}_j)\omega_{ij} \end{aligned} \tag{8}$$

So, the correction formula of parameters is as follows:

$$\begin{aligned} \omega_{ij}(n+1) &= \omega_{ij}(n) + \lambda S_{\omega_{ij}} \\ c_i(n+1) &= c_i(n) + \lambda S_{c_j} \\ \sigma_i(n+1) &= \sigma_i(n) + \lambda S_{\sigma_i} \end{aligned} \tag{9}$$

Here, $1 \le i \le l, 1 \le j \le M$, l is the number of hidden crunodes, M is the output dimension, λ is step length, generally speaking $\lambda = 0.05$.

Step 5: After determine ω_{ij}、 c_i、 σ_i , the RBF forecasting model can describe as follows:

$$y_k = \sum_{i=1}^{m} \omega_{ik} R_i(x), k = 1, 2, \cdots, r \tag{10}$$

Here r is the number of output neuron, ω_{ik} is the network weight.

3. Case Study

Nowadays, big data is the most popular vocabulary in information processing area, but it is not enough to have big data only, the key value is to make these data play business role. At present, there are fewer references about big data processing, from the application object of big data processing; electricity system is a suitable application field. Electricity system is the fundamental, pillar and leading industry, plays an important role in promoting the national information.

Under the background of China's power industry, this paper try to make a forecasting mining study on the electricity consumption big data, so that we can obtain a solution to deal with the big data processing. The specific programs are as follows:

(1) Determine the study object

Taking a certain area of electric power business data as the research object, the time span from January, 2010 to December, 2012. All kinds of business data record total of 9785639 items, the data volume is 1.2PB. Taking into account the original data information is huge and complicated, and the server derived inconvenience, so this paper does not indicate. These business data including structured data and unstructured data, such as the number of users, cost, time length, reason of cut network, style of cut network and so on. In line with the basic requirements in terms of data size and data type of big data.

(2) Separate the study target

Separate the consumption data from the whole business data by using MapReduce technology, separating core code is as follows:

```
Mapper.py
#!/usr/bin/python
import sys
for line in sys.stdin:
for Consumption Data in line.split():
sys.stdout.write("%s\t%d\n"%( Consumption Data,1))
Reducer.py
#!/usr/bin/python
import sys
dict={}
for line in sys.stdin:
(Consumption Data,count)=line.split()
count=int(count)
dict[Consumption Data]=dict.get(Consumption Data,0)+count
for key in dict.keys():
sys.stdout.write("%s\t%d\n"%(key,dict[key]))
```

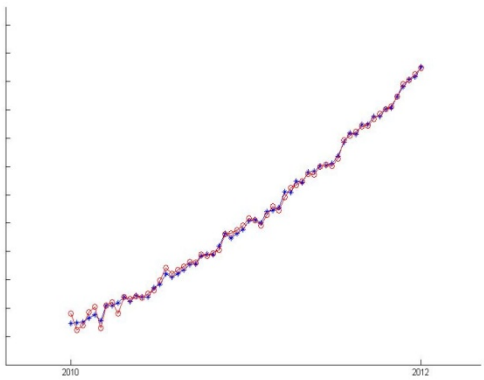

Figure 2. Comparison Chat Electricity Consumption Big Data Mining

4. Conclusion

At present, the data mining study for traditional scale data set has been widely researched and used. Big data age has brought a great change to data mining, especially to extract valuable information from the massive complex data set, is a new challenge. This paper combined the big data and traditional data mining methods by using MapReduce technology, gives a practical way to analysis and mining the large scale data. The prediction model in our paper has good feasibility and portability, can play a significant role in analysis and processing of large scale electricity data set.

Acknowledgements

The authors would like to thank the China Key Laboratory of Process Optimization and Intelligent Decision-Making for their valuable comments and feedback regarding this research study. This paper was supported by the National Natural Science Foundation of China Grant No.71101041 and No.71071045, National 863 project Grant No. 2011AA05A116, Foundation of Higher School Outstanding Talents Grant No. 2012SQRL009 and National Innovative Experiment Program No.111035954.

Finally, we are grateful to the many editors who gave their attention to this paper.

References

[1] Georges Nahon. New Tools for New Computing Challenges, *http://www.nytimes.com.*
[2] Luo Dang, Liu Sifeng, Dang Yaoguo. The Optimization of Grey Model GM (1,1). *Engineering Science.* 2003; 5(8): 50-53.
[3] Li Junfeng, Dai Wenzhan. A New Approach of Background Value-Building and Its Application Based on Data Interpolation and Newton-Cores Formula. *Systems Engineering-Theory & Practice.* 2004; 24(10): 122-126.
[4] Sharaf AM, Tjing Tlie. A novel neuro-fuzzy based self-correcting online electric load forecasting model. *Electric Power Systems Research.* 1995; 34(2): 121–125.
[5] Pao HT. Forecasting energy consumption in Taiwan using hybrid nonlinear models. *Energy.* 2009; 34(10): 1438–1446.
[6] Zhang BL, Dong ZY. An adaptive neural-wavelet model for short term load forecasting. *Electric Power Systems Research.* 2001; 59(2): 121-129.
[7] Jeffrey Dean, Sanjay Ghemawat, MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM*. 2007; 51: 107-114.
[8] Wang XJ, Shen JX, Yang SL. Application research on Gaussian orthogonal interpolation method for electricity consumption forecasting of smart grid. *Power System Protection and Control*. 2010; 21(38): 141-145,151.
[9] Wang XJ, Yang SL, et al. Simulation of Orthogonalization Prediction Based on Grey Markov Chain for Electricity Consumption. *Journal of System Simulation.* 2010; 22(10): 2253-2256.
[10] Xiaojia Wang, Application Research on Electricity Demand Forecasting Based on Gaussian Quadrature Formula, *Procedia Engineering,* 2011;15:5574-5578.
[11] Xiaojia Wang, Shanlin Yang. The Improvements and Applications of Forecasting Method in GM(1,1) Model Based on Combinative Interpolation. *Chinese Journal of Management Science.* 2012; 20(2): 129-134.
[12] Xiaojia Wang. Forecasting Modeling and Analysis of Power Engineering in China Based on Gauss-Chebyshev Formula. *System Engineering.* 2012; 5: 131-136.
[13] JF Liu, ZL Deng. Self-Tuning Weighted Measurement Fusion Kalman Filter for ARMA Signals with Colored Noise. *Applied Mathematics & Information Sciences.* 2012; 6:1-7.
[14] Guo Lei, Li Shixiang. Coal resources demand evaluation based on support vector regression with improved quantum-behaved PSO. *Journal of Convergence Information Technology*. 2012; 7(1): 405-412.
[15] Yin Zihong, Li yuanfu. Structural Deformation Forecasting Based On Support Vector Regression Trained By Particle Swarm Optimization Algorithm. *Advances in Information Sciences and Service Sciences.* 2012; 4(5): 130-136.