Pattern Recognition Letters 000 (2017) 1–9

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Distributed electricity load forecasting model mining based on hybrid gene expression programming and cloud computing

Song Deng[a,*], Changan Yuan[b], Lechan Yang[c], Liping Zhang[a]

[a] Nanjing University of Posts and Telecommunications, 9 Wenyuan Road, Nanjing and 210023, China
[b] Guangxi Teachers Education University, 175 Mingxiu East Road, Nanning and 530023, China
[c] Nanjing University, 163 Xianlin Road, Nanjing and 210093, China

ARTICLE INFO

Article history:
Available online xxx

MSC:
41A05
41A10
65D05
65D17

Keywords:
Load forecasting
Artificial fish swarm
Gene expression programming
Cloud computing

ABSTRACT

Load forecasting is an important part of power grid management. Accurate and timely load forecasting is of great significance to formulate economical and reasonable power allocation plan, improve safety and economy of power grid operation and improve power quality. In this paper, in order to find electricity load forecasting model, we propose an electricity load forecasting function mining algorithm based on artificial fish swarm and gene expression programming (ELFFM-AFSGEP). On the basis, distributed load forecast model mining based on hybrid gene expression programming and cloud computing (DLFMM-HGEPCloud) is proposed to solve the problem of massive electricity load forecasting. In order to better solve global electricity load forecasting model, error minimization crossover is introduced into DLFMM-HGEPCloud. The performance of the proposed algorithm in this paper is evaluated with a real-world dataset, and compared with GEP and some published algorithms by using the same dataset. Experimental results show that our proposed algorithm has an advantage in average time-consumption, average number of convergence, forecasted accuracy and excellent parallel performance in speedup and scaleup.

1. Introduction

Electricity load forecasting plays an important role for power plant operators, power suppliers and traders, and energy managers in decision making [8]. From the point of view of the forecasting, the power load forecasting includes the prediction of future power demand and the prediction of future electricity consumption and the forecasting of the load curve. The main task is to predict the future distribution of power load and spatial distribution, power system planning and operation to provide a reliable basis for decision-making. The load forecasting is highly related to power system operations such as dispatch scheduling, preventive maintenance plan for the generators, and the reliability evaluation of the power systems. In addition, accurate power load data is crucial to power price forecasting in the electricity market. Accurate load forecasting is conducive to improve the safe operation of the power grid, the economy and power quality. Load forecasting is an important part of power grid management. Accurate and timely load forecasting is of great significance to formulate economical and reasonable power allocation plan, improve safety and economy of power grid operation and improve power quality.

However, with the rapid development of informationization and intelligent degree of power grid and the increasing demand of power demand factors, the increasing size of power load data, the first-tier cities in the peak period, faced with millions of records of the size of the power data collection, One year of data storage scale will grow from the current GB level to the terabyte level, and even the PB level [23,36–39]. Power load forecasting will face challenges and difficulties. First, the increasing mass of load data makes it possible for existing centralized load forecasting algorithms to become more difficult in terms of timely analysis and processing. Secondly, the existing power load forecasting algorithms mainly include statistical and artificial intelligence methods [1,2,7,13,25,32,34]. The load forecasting algorithm based on the statistical method is better in predicting the linear sequence of the power load, but the shortcomings of the inflexibility of the structure lead to the weak ability of the algorithm to predict the nonlinear load sequence, and subjective of the statistical method is too strong. The load forecasting algorithm based on artificial intelligence can deal with the nonlinear relationship between load and related factors, but this kind of algorithm is easy to lead to over-fitting and fall into local optimum, and the training time is longer and the efficiency is lower. In this paper, we mainly introduce the gene expression programming (GEP)[9,10] to automatically mine the function form of load forecasting model, and ulti-

* Corresponding author.
E-mail address: dengsong@njupt.edu.cn (S. Deng).

https://doi.org/10.1016/j.patrec.2017.10.004
0167-8655/

mately achieve the purpose of accurate and timely prediction. The major contributions of our work are listed as follows:

(1) The power load forecasting function model mining algorithm based on a single GEP, is easy to produce premature phenomena in the population. In order to solve this problem, this paper combines the artificial fish swarm algorithm with strong local and global optimization ability, and proposes an electricity load forecasting function mining algorithm based on artificial fish swarm and gene expression programming (ELFFM-AFSGEP).
(2) With the rapid development of informationization of distribution network, the power load data show the characteristics of distribution, mass and high dimension. The centralized mining model of power load forecasting function will undoubtedly increase the safety, privacy and data storage problem, but also reduce the efficiency of power load forecasting. In order to solve prediction of massive power load data, based on the ELFFM-AFSGEP algorithm, this paper puts forward distributed load forecast model mining based on hybrid gene expression programming and cloud computing (DLFMM-HGEPCloud).
(3) Meanwhile, In order to solve the global electricity load forecasting model in DLFMM-HGEPCloud algorithm, it is not possible to solve the effective merging of nonlinear function model based on statistical method. Therefore, this paper proposes a global model generation algorithm for electricity load forecasting based on error minimization crossover (GMGELF-EMC).

The remainder of this paper is organized as follows. In Section 2, we briefly describe the related works. In Section 3, we introduce electricity load forecasting function mining algorithm based on artificial fish swarm and gene expression programming. In Section 4, we analyze distributed electricity load forecasting model mining algorithm gene expression programming and cloud computing. Experimental results are provided in Section 5 and we conclude this paper in Section 6.

2. Related work

2.1. Load forecasting

In the past decades, many methods have been proposed to improve load forecasting accuracy by many researchers. Taylor used exponentially weighted method to forecast short-term load [34]. Infield et al. proposed optimal smoothing for trend removal in short term electricity demand forecasting by using Kalman filters [15]. Huang et al. used ARMA model identification to forecast short-term load including non-Gaussian process considerations [13]. Experimental results show the effectiveness of the method by using the utility data on a practical system. Wi et al. proposed holiday load forecasting method by using fuzzy polynomial regression with weather feature selection and adjustment [40]. However, these methods cannot properly reflect the nonlinear relationship of the short-term load series and need to depend on a priori knowledge and a lot of subjective factors. Hence, some artificial intelligence techniques were used to load forecasting. Luis et al. applied artificial neural network to short-term load forecasting in micro grid environments [25]. Li et al. proposed an annual electric load forecasting method by applying a least squares support vector machine with a fruit fly optimization algorithm [20]. The experimental results showed that the method proposed in this paper combined with fruit fly optimization algorithm outperformed other alternative methods. In view of strong non-linear learning capability of support vector regression (SVR), Fan et al. present a hybrid SVR model for electric load forecasting which combined with the empirical mode decomposition method and auto regression [7]. Salgado et al. applied ID3 decision tree and expert systems to the load forecasting. The experimental results showed that the proposed model in this paper was better than other methods by using real load data in southeast of Brazil [30]. Sheikhan et al. proposed two hybrid models for load forecasting algorithm. These models included combination of genetic algorithm and ant colony optimization for feature selection and multi-layer perceptron for hourly load prediction. Experimental results showed that the proposed hybrid model performed better in load prediction of 24-h ahead [32]. Chen presented a collaborative fuzzy-neural method for long-term load forecasting. The experimental results showed that the proposed method improved both the precision and accuracy of long term load forecasting by using Taiwan case [2]. Jin et al. proposed short-term electricity load forecasting algorithm based on grey correlation contest modeling [16]. A hybrid PSO-SVM method was applied to short-term load forecasting during periods with significant temperature variations in city of Burbank [31]. Li et al. proposed a hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm [21]. Abdoos et al. applied hybrid intelligent method to short term load forecasting. The methods included Wavelet Transform, Gram-Schmidt and Support Vector Machine. Compared with some reported algorithms, comparative results showed the priority of the proposed method in aspect of prediction accuracy [1]. Geng et al. proposed a load forecasting model which combined the support vector regression with chaotic cloud simulated annealing. Experimental results indicated that the proposed mode in this paper yields more accurate forecasting results than other models [11]. Li et al. proposed a new short-term load forecasting algorithm by using wavelet transform, extreme learning machine and improved artificial bee colony algorithm. The simulation results showed that the proposed model can obtain superior results in contrast to other models [22]. In order to accurately predict power load, Chen et al. proposed a hybrid forecasting algorithm based on empirical mode decomposition, seasonal adjustment, particle swarm optimization and least squares support vector machine [3]. Appropriate feature selection is crucial for accurate load forecasting. Koprinska et al. proposed correlation and instance based feature selection for electricity load forecasting [19]. Khwaja et al. presented improved short-term load forecasting using bagged neural networks [18]. In particular, these electricity load forecasting models based on artificial intelligence algorithms can describe the nonlinear relationship between load and its dependent factors. However, the efficiency of the nonlinear function model mined by these electricity load forecasting algorithms was low. Meanwhile, in order to improve efficiency of load forecasting algorithms, we should preprocess the load datasets by feature selection [12,44]. In essence, load forecasting can be seen as a classification. Good classification algorithms can improve efficiency and accuracy of load forecasting [41,43].

2.2. Gene expression programming

Recently, gene expression programming (GEP) was proposed as a new function model mining algorithm [9]. Compared with traditional genetic algorithms (GA) and Genetic Programming (GP), GEP has advantages in terms of convergence speed and ability to solve complex problems. At present, research on GEP focused on symbolic regression, function finding, combinatorial optimization and prediction. In symbolic regression and function mining, Peng et al. proposed an improved GEP algorithm named S_GEP, which is especially suitable for dealing with symbolic regression problems [27]. Zhao et al. treated image registration as a formula discovery problem, and proposed two-stage gene expression programming and the improved cooperative particle swarm optimizer used to identify the registration formula for the reference image and the floating image [42]. In combinatorial optimization, Sabar et al. proposed

Please cite this article as: S. Deng et al., Distributed electricity load forecasting model mining based on hybrid gene expression programming and cloud computing, Pattern Recognition Letters (2017), https://doi.org/10.1016/j.patrec.2017.10.004