Energy 87 (2015) 41–48

Contents lists available at ScienceDirect

Energy

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

Fault diagnosis for a solar assisted heat pump system under incomplete data and expert knowledge

Zengkai Liu, Yonghong Liu*, Dawei Zhang, Baoping Cai, Chao Zheng

College of Mechanical and Electronic Engineering, China University of Petroleum, Qingdao, Shandong, 266580, China

ARTICLE INFO

Article history:
Received 4 December 2014
Received in revised form
17 March 2015
Accepted 21 April 2015
Available online 27 May 2015

Keywords:
Fault diagnosis
Bayesian network
Parameter learning
Incomplete data
Incomplete expert knowledge
Solar assisted heat pump system

ABSTRACT

Fault diagnosis for a solar assisted heat pump (SAHP) system in the presence of incomplete data and expert knowledge is discussed in this article. A method for parameter learning of Bayesian networks (BNs) from incomplete data based on the back-propagation (BP) neural network and maximum likelihood estimation (MLE), which is called BP-MLE method, is presented. The BP neural network is utilized to impute the missing data and the complete data sets are addressed with MLE to obtain the parameters of BN. A method for parameter estimation under incomplete expert knowledge based on BP neural networks and fuzzy set theory is also presented, which is called BP-FS method. Similarly, the missing information is imputed by the trained BP neural network. Fuzzy set theory is employed to quantify the parameters of BN based on complete qualitative expert knowledge. The presented methods are applied to parameter learning of diagnostic BN for a SAHP system with incomplete simulation data and expert knowledge. The developed BN can perform fault diagnosis with complete or incomplete symptoms.

1. Introduction

A (SAHP) solar assisted heat pump system integrates a heat pump with solar collectors to take advantage of solar energy as an evaporating heat source, which can achieves high coefficient of performance [1]. Until now, there are plenty of theoretical and experimental researches on SAHP systems. Mohanraj et al. [2] develop an (ANN) artificial neural network to predict the performance of a direct expansion SAHP and the reported results demonstrate that the proposed method is acceptable. Liang et al. [3] present a new SAHP system with flexible operational modes to improve the performance of the heating system and the developed system validates the established mathematical model. Li and Yang [4] investigate the application of the SAHP system for hot water production in Hongkong. Chow et al. [5] describe a case study with a new design of SAHP for indoor swimming pool space- and water-heating purpose.

Since failures in pump system will cause the occurrence of abnormal operating and degradation in performance, fault diagnosis of the system is beneficial to the energy saving and operating cost saving. Some fault diagnosis methods have been developed by the researchers. Zhao et al. [6] present a new fault detection and diagnosis method based on support vector regression and the exponentially-weighted moving average control charts for centrifugal chillers of building air-conditioning systems. Chen and Lan [7] propose a fault detection method based on principle component analysis model to detect the faults in air-source heat pump water chiller/heaters. Zogg et al. [8] develop a model-based fault diagnosis system for commercial heat pumps, which is based on parameter identification and vector clustering techniques. Zhao et al. [9] present a three-layered diagnostic Bayesian network to make use of more useful information of the chiller concerned and expert knowledge. Cai et al. [10] present a multi-source information fusion based fault diagnosis method for ground-source heat pump system to increase the diagnostic accuracy. Najafi et al. [11] develop diagnostic algorithms for air handling units using machine-learning techniques.

Recently, (BNs) Bayesian networks for fault diagnosis have been widely developed in variety of fields including electrical power systems [12], telecommunications networks [13,14], rotating machinery [15], airplane engine [16] and others. A BN is a directed acyclic graph composed of nodes and arcs among the nodes. In a BN, nodes denote random variables and the directed arcs mean the conditional dependencies among variables [17].

A BN consists of parameters and structure, which can be defined by expert knowledge or obtained by machine learning with data

* Corresponding author. Tel.: +86 0532 869833032; fax: +86 0546 8393620.
E-mail address: liuyhupc@163.com (Y. Liu).

http://dx.doi.org/10.1016/j.energy.2015.04.090
0360-5442/

sets. The former can reflect experts' knowledge, but the process is difficult and time-consuming. Besides, it is not sure if the network presented by the experts is really the most proper model. Although learning from data is able to overcome the problems of the former, it is not always available because the data cannot be ready at all the time the BN is constructed. In addition, the change of environments is not considered [18]. Generally, there are two kinds of Bayesian learning problems: parameter learning and structure learning. For parameter learning, one is to learn BN parameters when its structure is known, and the other is to learn both structure and parameters at the same time. This paper focuses on BN parameter learning with known BN structure. The parameter learning is divided into two categories. If the data is complete, (MLE) maximum likelihood estimation or Bayesian estimation method can be used to learn parameters. However, in real world, to find the complete data for learning is difficult for various reasons. Some of the variables may be difficult or even impossible to observe. The presence of missing data leads to analytical intractability and complex computation compared with the complete data scenario. Because the incomplete sample sets might reduce the accuracy of parameters. The easiest way to deal with the incomplete data is to delete it directly. To simply discard the incomplete data sets, the relevant information may be deleted [19]. In addition, throwing away data can lead to estimates with larger standard errors due to the reduced sample size. Rather than deleting the incomplete data sets, another approach is to impute the missing values. This method keeps the full sample size, which can be advantageous for bias and precision [20]. Many methods have been proposed to learn BN parameters when the data is incomplete. The most common learning algorithm is (EM) expectation maximization [21]. Recently, many researchers have proposed some new approaches for parameter learning under incomplete data. Pernkopf et al. [22] present a genetic based EM algorithm for learning Gaussian mixture models form multivariate data and the algorithm is less sensitive to the initialization compared to the standard EM. Majdi-Nasab et al. [23] propose new approaches based on genetic algorithms, simulated annealing and EM for parameter learning of the mixture Gaussian model. Huda et al. [24] present a hybrid algorithm for estimation of the hidden Markov model in automatic speech recognition using a constraint-based evolutionary algorithm and EM and the presented algorithm overcome the problem of EM converging to a local optimum. Liao et al. [25] propose a learning algorithm incorporating qualitative constraints of domain knowledge on some of the parameters into the learning process and this algorithm is able to regularize the otherwise ill-posed problem, limit the search space, and avoid local optima.

ANN is a powerful tool in the modeling of nonlinear multivariate systems. It is able to capture the complicated nonlinear relationships between inputs and outputs by proper training. Among ANN methods, (BP) back-propagation neural network is the most widely used training algorithm. BP neural network is multilayer feed-forward that is trained by the error BP algorithms. Classical BP neural network has three layers, namely input layer, hidden layer and output layer. Input layer receives and distributes the input pattern. Hidden layer establishes the nonlinearities of the input and output relationship. Output layer produces the output pattern [26]. BP neural network has provided effective solutions to quality prediction, prediction of the mechanical properties, prediction of various stock indices and oil reservoir prediction and so on [27–30].

Fuzzy sets were introduced to represent/manipulate data and information processing nonstatistical uncertainties [31]. To mathematically represent uncertainty and vagueness, it is a formalized tool for dealing with the imprecision intrinsic to many problems. Fuzzy set theory has been widely used in different fields of application including risk assessment, rock mass classification, radiation therapy, decision support system and pattern recognition [32–36].

In the real system, incomplete data is a common phenomenon, which could be caused by a sudden mechanical breakdown, hardware sensor failure or data acquisition system malfunction, etc [37]. Another increasing common source for the incomplete data problem is the integration of communication networks and the subsequent potential for data losses and packet dropouts [38]. Sensor failures are only one of the possible reasons that lead to incomplete data. The data sets for learning BN parameters refer to statistical samples of the equipment or systems. There might be unreliable data samples in the complete data sets due to the undetected sensor faults. However, preventive maintenances and repairs actions of the equipment can greatly reduce the existence of various faults in practice. Besides, there is a large quantity of samples. Therefore, unreliable samples account for only a tiny proportion of the whole data sets. It is rational that most of the data sets can be considered to be reliable. Great efforts should be paid to collect the sample sets. It is possible that there might be unreliable samples due to the various faults. But even if there are no these faults, human error might also lead to some unreliable samples. Therefore, uncertainty is unavoidable in the data sets. Fortunately, Bayesian network is a very powerful tool in uncertainty representation and reasoning. In order to delete the samples with wrong value or outlier, some data analysis methods in statistics can be used for data preprocessing, such as expert judgments, Chebyshev's theorem, distance-based clusters, pattern recognition et al. In addition, experts might not be able to provide complete qualitative knowledge, because they are not familiar with the concerned issues. In this paper, fault diagnosis of a SAHP system in the presence of incomplete data and expert knowledge is discussed. Based on BP neural network and MLE, BP-MLE method is presented for parameter learning of BN from incomplete data. The BP neural network is used to impute the missing data and then the complete data sets are addressed with MLE to obtain the parameters of BN. Based on the BP neural network and fuzzy set theory, BP-FS method is proposed for parameter estimation under incomplete expert knowledge. Similarly, the missing information is imputed by the trained BP neural network. Fuzzy set theory is used to quantify the parameters of BN based on the complete qualitative expert knowledge.

Firstly, the missing information is reconstructed to be complete data sets. Then all the data sets are used to determine the parameters of the developed diagnostic Bayesian network. Finally, based on the developed Bayesian network, fault diagnosis can be performed by using observed symptoms. Due to the powerful reasoning capacity of Bayesian network, it can perform fault diagnosis based on complete or incomplete information. The reminder of this paper is organized as follows. Section 2 provides a description of the proposed methods. The presented methods are applied to parameter estimation of a SAHP system in Section 3. Section 4 performs fault diagnosis based on the developed Bayesian networks. Section 5 summarizes the paper.

2. The proposed methods

2.1. BP-MLE method under incomplete data

Fig. 1 shows the flow chart of the proposed parameter learning method under incomplete data. The incomplete data is composed of complete and incomplete sample sets. The complete sample sets are utilized to train BP neural networks. When a BP network is trained, parameters such as the number of hidden layers, hidden