Applied Energy 205 (2017) 926–940

Contents lists available at ScienceDirect

Applied Energy

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

# Fault detection and operation optimization in district heating substations based on data mining techniques

Puning Xue[a], Zhigang Zhou[a], Xiumu Fang[a], Xin Chen[a], Lin Liu[a], Yaowen Liu[c], Jing Liu[a,b,*]

[a] School of Municipal and Environmental Engineering, Harbin Institute of Technology, Harbin 150000, China
[b] State Key Laboratory of Urban Water Resource and Environment, Harbin Institute of Technology, Harbin 150000, China
[c] Heilongjiang LONG DIAN Electric Co. Ltd., Harbin 150000, China

## HIGHLIGHTS

- A data-mining-based method is used to analyze district heating operational data.
- Clustering analysis can identify the seasonal and daily operating patterns.
- Association rules can help to understand the substation regulation strategies.
- A malfunction of a secondary remote flow meter is detected accurately.
- An energy-inefficient operation strategy in a substation is detected and modified.

## ARTICLE INFO

*Keywords:*
District heating substation
Data mining
Automatic meter reading system
Fault detection
Operation optimization

## ABSTRACT

The present generation of district heating (DH) technologies will have to be further developed into the 4th generation to fulfil the important role in future smart energy systems. At present, automatic meter reading systems have been installed in DH systems. These systems make hourly or even minutely meter readings available at low cost. However, the sheer quantity and complex of the data poses a challenge at various levels for traditional data analysis approaches. Data mining is a promising technology and is used to automatically extract valuable knowledge hidden in large amounts of data. To investigate the potential application of descriptive data mining techniques in DH systems, this study proposes a method based on descriptive data mining to improve the energy performance of DH substations. The proposed method consists of five steps: data cleaning, data transformation, cluster analysis, association analysis, and interpretation/evaluation. Data cleaning and transformation are implemented to improve data quality and transform data into forms that are appropriate for mining. Cluster analysis is performed to identify distinct operating patterns of substations. Based on each pattern, association analysis is then adopted to discover the unsuspected knowledge in the form of rules. Interpretation/evaluation is performed to select and interpret potentially useful rules. To demonstrate its applicability, the proposed method is used to analyze the datasets obtained from an automatic meter reading system at two substations in the DH system in Changchun, China. This application reveals that the method can effectively extract potentially useful knowledge and thereby provide essential guidance for the fault detection and operation optimization of DH substations.

## 1. Introduction

District heating (DH) is an efficient method of supplying heat to consumers and is widely used all over the world, particularly in northern China. Over the last 10–15 years, the Chinese heating sector has grown faster than other countries [1]. According to [2], the DH area coverage in China was 5 billion $m^2$ in 2001, but significantly grew to 12 billion $m^2$ in 2013. Thus far, DH consumes 24% of the total annual energy cost in China. Unlike other countries, coal is still the primary fuel used for heating in China. Approximately 40% of the air pollution in China originates from coal dust [3]. A number of studies indicate that sulfide and nitrogen oxide emission from Chinese urban DH are two of the main pollution sources [2]. Increasing haze seriously affects the lives, health and activities of the people. With the acceleration of urbanization, DH is facing the conflict between rapid growth in demand and environmental problems in China. To address this situation,

* Corresponding author at: School of Municipal and Environmental Engineering, Harbin Institute of Technology, No. 73, Huanghe Road, Nangang District, Harbin 150000, China.
*E-mail address:* liujinghit0@163.com (J. Liu).

http://dx.doi.org/10.1016/j.apenergy.2017.08.035
Received 12 April 2017; Received in revised form 20 June 2017; Accepted 8 August 2017
Available online 20 August 2017
0306-2619/ 

increasing energy efficiency of DH systems is one of the most important and essential strategies. It includes a two-fold process: the use of affordable energy sources and the improvement of current energy management procedures and infrastructures. Future smart energy systems present a promising solution for this challenge, and DH system will become an integrated part of the operation of smart energy systems. In the smart energy systems, smart electricity, thermal and gas grids are combined and coordinated to achieve an optimal solution for each individual sector, as well as for the overall energy system [4,5]. Compared to other heating systems, DH is flexible in using various forms of energy [6]. Therefore, DH can play an important role in future smart energy systems. Furthermore, the present generation of DH technologies will have to be further developed into the 4th generation to fulfil the important role in smart energy systems [7–9]. One of the challenges is that the smart energy systems will require more information exchanges between interfaces in the energy system [10]. With the development of Information and Communication Technology (ICT), automatic meter reading systems have been installed in DH substations. This makes hourly or even minutely meter readings of heat, flow, pressure, and temperature available at low cost. These meter readings can be used for fault detection, control optimization, and identification of heat load patterns in DH substations [11–14]. These advances will signify a shift in the perception of who owns these data and who benefits from them. However, the large amount of DH-related data is rarely fully analyzed because of the poor quality of meter readings and the lack of effective data analysis techniques. In other words, the sheer quantity and complex of the data poses a challenge at various levels for traditional data analysis approaches. To create future applications in smart energy systems, advanced technology is needed to analyze the enormous amounts of data collected from automatic meter reading systems.

An emerging and promising technology known as data mining is a powerful and versatile tool used for automatically extracting the valuable knowledge hidden in large amounts of data [15]. Hand et al. [16] defines data mining as follows: "The analysis of large observation datasets to find unsuspected relationships and to summarize the data in novel ways so that owners can fully understand and make use of the data." Data mining is an interdisciplinary field that covers multiple subjects, such as statistics, machine learning, and pattern recognition [17]. In general, data mining tasks can be divided into two main categories, i.e., predictive methods (supervised learning) and descriptive methods (unsupervised learning). Predictive methods aim to use some variables to predict unknown or future values of other variables. Description methods aim to find human-interpretable patterns hidden in the data, such as the associations, correlations, and clusters [18].

Increasing data awareness of users has led to the development of solutions based on data mining. Over the past several decades, data mining has been successfully applied in economics, retail, telecommunication, and financial services [18]. Recently, efforts have also been made to introduce it into the building field that is a well-fit application area for data mining. In DH research area, data mining techniques have been widely used for predicting heat load. Commonly used methods include support vector machine (SVM), support vector regression (SVR), artificial neural networks (ANNs), autoregressive integrated moving average (ARIMA), adaptive neuro-fuzzy inference system (ANFIS), and extreme learning machine (ELM). In [19], the load forecast models were generated using four techniques such as SVM, regression tree, feed forward neural network, and multiple linear regression. The results showed that SVM had the least normalized root mean square error for a forecast horizon of 24 h. Shamshirband et al. [20] also developed nine short-term multi-step ahead predictive models using SVM coupled with a discrete wavelet transform algorithm. And compared the models with genetic programming (GP) and ANNs models. The results showed that an improvement in predictive accuracy and capability of generalization could be achieved by the SVM model. A short-term multistep-ahead predictive model was also developed by Al-Shammari et al. [21] using SVM with firefly algorithm. Protić et al. [22] designed and tested numerous models based on SVR. The results showed that the models could be a reliable tool for short term prediction of consumers' heat load. Sandberg et al. [23] applied nonlinear autoregressive neural network with external input to forecast heat demand of a commercial building. The predictive accuracy was 96% on an hourly basis for the period of a whole year. Xie et al. [24] developed a predictive model based on Elman neural networks and analyzed the impacts of direct solar radiance and wind speed on the model performance. Kato et al. [25] proposed a thermal load prediction method which used a recurrent neural network to deal with the dynamic variation of heat load and its characteristics. Ediger and Akar [26] applied ARIMA to estimate the primary energy demand. Grosswindhager et al. [27] used seasonal autoregressive integrated moving average (SARIMA) model for online short-term forecast of system heat load. Fang and Lahdelma [28] also developed a SARIMA model, and the advantage of the model was that it perused the high accuracy for both long-term and short-term forecast by considering both exogenous factors and time series. In [29], Shamshirband and Petkovic et al. utilized ANFIS to predict the heat load for individual consumers in DH systems. In [30], ANFIS was also applied to select the most relevant variables for predictive models of consumer heat load. In [31,32], three various methods including ELM, ANNs, and GP were employed to develop heat load predictive models.

Despite of the encouraging research outcomes, previous studies mainly used predictive data mining methods. The main focuses were placed on capturing the complex and nonlinear relationships between inputs and outputs. Compared to predictive methods, descriptive methods are more flexible in application, since they do not involve a training process and are not guided by pre-defined targets [33]. The literatures [34–36] demonstrated that the descriptive data mining techniques were useful for building sector, especially for building automation system (BAS). However, the potential of descriptive methods has not been fully investigated in DH systems. In this regard, this study proposes a method based on descriptive data mining techniques to mine unsuspected knowledge in the datasets derived from an automatic meter reading system.

The intention of this study is to let the data tell the story and then using domain knowledge to interpret, select and apply the discovered knowledge. Two popular data mining techniques, cluster analysis and association analysis, are integrated in this method. To demonstrate the applicability of this method, it is applied to analyze the datasets obtained from an automatic meter reading system at two substations in the DH system located in Changchun, China. It should be mentioned that R, an open source software, is applied to implement all of the data mining algorithms used in this study. The results show that the proposed method can effectively extract potentially useful knowledge, thereby providing essential guidance for the fault detection and operation optimization of DH substations.

## 2. Description of DH substations

The datasets used for this case study are collected from a DH system located in Changchun (43°89′N, 125°32′E), a large provincial capital city in northeast China. Changchun has a temperate continental monsoon climate, and its winters are cold and long. Over the course of the year, there are five months where the monthly average temperature reaches below 0 °C. The monthly average temperature is −15.1 °C in January, the coldest month. In addition, the duration of the heating period can extend up to 169 days. As a result, DH systems are essential for building a comfortable and healthy indoor environment for the local residents.

The subject DH system, an indirect hot-water heating system with a regional coal-fired boiler plant as heat source, is a conventional heating system commonly used in northern China. Because of the large scale and complicated structure of the DH system, there are usually hydraulic