Expert Systems with Applications 41 (2014) 2144–2156

Contents lists available at ScienceDirect

Expert Systems with Applications

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

# Smart meter monitoring and data mining techniques for predicting refrigeration system performance

Jui-Sheng Chou [a,*], Yu-Chien Hsu [a], Liang-Tse Lin [a]

[a] Department of Construction Engineering, National Taiwan University of Science and Technology, 43, Sec. 4, Keelung Rd., Taipei 106, Taiwan

ARTICLE INFO

*Keywords:*
Refrigeration management
Smart meter
Monitoring experiment
Data mining
Performance diagnosis

ABSTRACT

A major challenge in many countries is providing sufficient energy for human beings and for supporting economic activities while minimizing social and environmental harm. This study predicted coefficient of performance (COP) for refrigeration equipment under varying amounts of refrigerant (R404A) with the aids of data mining (DM) techniques. The performance of artificial neural networks (ANNs), support vector machines (SVMs), classification and regression tree (CART), multiple regression (MR), generalized linear regression (GLR), and chi-squared automatic interaction detector (CHAID) were applied within DM process. After obtaining the COP value, abnormal equipment conditions can be evaluated for refrigerant leakage. Analytical results from cross-fold validation method are compared to determine the best models. The study shows that DM techniques can be used for accurately and efficiently predicting COP. In the liquid leakage phase, ANNs provide the best performance. In the vapor leakage phase, the best model is the GLR model. Experimental results confirm that systematic analyses of model construction processes are effective for evaluating and optimizing refrigeration equipment performance.



## 1. Introduction

Data mining (DM) techniques have formed a branch of applied Artificial Intelligence (AI) since the 1960s, several major kinds of data mining methods such as generalization, characterization, classification, clustering association, evolution, pattern matching, data visualization and meta-rule guided mining (Liao, Chu, & Hsiao, 2012). Prediction is one of major data mining functions used in the applications (Köksal, Batmaz, & Testik, 2011). Its broad applications include marketing, healthcare, civil engineering, and many others (Chen, Chiang, Wu, & Chu, 2013; Chou, 2009; Kao, Chen, & Chou, 2011; Koyuncugil & Ozgulbas, 2012; Küçüksille, Selbaş, & Şencan, 2009; Moreno Sáez, Sidrach-de-Cardona, & Mora-López, 2013). In addition, Köksal et al. (2011) comprehensively reviewed DM applications in manufacturing industries (Köksal et al., 2011). However, DM is rarely applied in the energy field, particularly to support energy efficiency.

Taiwan, which imports 99.4% of its energy needs, has already begun replacing conventional meters with smart meters. Taiwan Power Company plans to install 1 million smart meters for its customers before 2015. To achieve this target, smart meter devices will be installed in 10 thousands households in 2012, and installations will gradually increase to 1 million devices during 2013 to 2015 (Lee, 2011). The aim of this policy is to improve energy efficiency and reduce carbon emissions in Taiwan. Based on the overseas experience in similar projects, policy makers have predicted that the devices will reduce carbon dioxide emissions in the future. Moreover, in many countries, current policies for reducing emissions coupled with growing public awareness of increased utilities price have increased the use of smart meters as monitoring tools (Bennett, Stewart, & Beal, 2013; Li, Fang, Mahatma, & Hampapur, 2011; Usman & Shami, 2013). Therefore improving refrigeration system performance *via* smart meter and data mining is an important research issue.

As energy conservation and carbon reduction have recently become an important issue, investigation of large consumption energy systems has been prioritized (Bektas, Ekici, & Aksoy, 2011; Kalogirou, 2000; Oğuz, Sarıtas, & Baydan, 2010; Rodger, 2014; Soyguder & Alli, 2009). One of the facilities being studied is refrigeration systems for preserving food and for air conditioning, which constantly consume energy. Refrigeration systems are ubiquitous and can be found in many locations, including factories, households, offices, etc. Despite their wide spread use, the performance of these systems has not been fully investigated (Ahmed, Korres, Ploennigs, Elhadi, & Menzel, 2011; Ozgoren, Bilgili, & Babayigit, 2012; Şahin, 2011). Thus, developing an appropriate methodology for predicting refrigeration system performance based on refrigerant conditions is imperative.

Although many studies of this problem have been performed, available studies using data mining approach of such energy systems are still rare. Notably, the electrical properties of refrigerant amount used in vapor compression refrigeration systems can only

* Corresponding author. Tel.: +886 2 2737 6321; fax: +886 2 2737 6606.
*E-mail address:* jschou@mail.ntust.edu.tw (J.-S. Chou).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.09.013

be determined through experiments. This study designed laboratory experiments to achieve this goal. All experimental data were retrieved from smart meters, which were also used to monitor electricity usage and user behavior. The DM techniques were used as analytical tools to predict the coefficient of performance (COP) under different refrigerant amounts.

The DM techniques compared in this study included artificial neural networks (ANNs), support vector machines (SVMs), classification and regression tree (CART), multiple linear regression (MLR), generalized linear regression (GLR), and chi-squared automatic interaction detector (CHAID) techniques. The indicators used to evaluate model performance were mean absolute percentage error (MAPE), root mean square error (RMSE), mean absolute error (MAE) and correlation coefficient. Cross-fold validation was also performed to ensure a balanced view and to avoid bias from data.

The rest of this paper is organized as follows. Research findings from previous studies are summarized in Section 2. Section 3 presents the data mining methodology used in this work. Section 4 describes the experimental design and monitoring system. Section 5 discusses the model implementation and analytical outcomes, i.e., model settings, cross-fold validation and the analysis results. The last section gives concluding remarks.

## 2. Literature review

A major challenge in many countries is providing sufficient energy for human beings and for supporting economic activities while minimizing social and environmental harm. Therefore, the social and scientific importance of electrical power system load forecasting has increased (Liu & Yi, 2010; Metaxiotis, Kagiannas, Askounis, & Psarras, 2003; Pao, 2006). In recent years, various electricity load forecasting methods have become highly advanced (Ahmed et al., 2011; Mateo et al., 2013). Current forecasting tools include: regression based models (Taylor & Buizza, 2003), time series models (Saab, Badr, & Nasr, 2001), artificial intelligence techniques (Jun & Chuntian, 2008; Park, Kim, Kim, Jo, & Yeo, 2010; Tso & Yau, 2007), fuzzy logic method (Bin, Chuangxin, & Yijia, 2004), nonlinear approach (Pao, 2006), Adaptive Network Based Inference System (Bektas Ekici & Aksoy, 2011), and fuzzy Bayesian network (Penz, Flesch, Nassar, Flesch, & de Oliveira, 2012).

Küçüksille et al. (2009), for example, presented ten modeling techniques within data mining process for the prediction of thermo-physical properties of refrigerants (R134a, R404a, R407c and R410a) (Küçüksille et al., 2009). Another study by Şahin (2011) reviewed the literature on performance analysis of refrigeration systems in the electrical engineering field and indicated that ANN model is slightly better than ANFIS for R134a whereas ANFIS model is slightly better than ANN for R404a and R407c (Şahin, 2011). Similarly, Chengmin, Yufeng, and Lijun (2012) proposed three modeling techniques, namely stepwise regression, decision tree and neural networks, for the prediction of electricity energy consumption. The prediction results showed that decision tree and neural networks were best models in the summer and winter phase, respectively (Chengmin et al., 2012).

Ying and Pan (2008) further applied the adaptive network based fuzzy inference system (ANFIS) model to forecast regional electricity loads in Taiwan and reported that the ANFIS model has better forecasting performance compared to the regression model, ANN model, support vector machines with genetic algorithms model, recurrent support vector machines with genetic algorithms model and hybrid ellipsoidal fuzzy systems for time series forecasting model (Ying & Pan, 2008). Similarly, Hosoz, Ertunc, and Bulgurcu (2011) suggested that the ANFIS modeling technique could be used successfully for predicting the performance of an R134a vapor-compression refrigeration system with a cooling tower (Hosoz et al., 2011). In another attempt to predict thermodynamic properties of refrigerant by using data mining, including linear regression, pace regression, sequential minimal optimization, M5 model tree, M5'Rules and back propagation neural network (BPNN) models were applied to determine the specific volume of different refrigerant (Şencan, 2007). The study showed that the BPNN model had the best prediction performance.

In the refrigeration industry, system performance is expressed in terms of COP, which is defined as the ratio of change in heat at the "output" to the supplied work. Therefore, the literature has increasingly focused on predicting and utilizing COP. Bechtler, Browne, Bansal, and Kecman (2001) used a dynamic neural network model to predict relevant vapor-compression liquid chillers such as COP or compressor work input. Application of the model in two dynamic processes in two different chillers showed that the model accurately identified all process characteristics (Bechtler et al., 2001). Chengmin et al. (2012) used COP as one of indicators to measure the phase cost and environment impact of different central heating systems in city of Tianjin China (Chengmin et al., 2012). Ozgoren et al. (2012) utilized $COP_{cooling}$ as a performance prediction evaluation of solar absorption refrigeration system (Ozgoren et al., 2012). Both studies confirm COP as a refrigeration system performance indicator.

Moreover, Swider (2003) predicted COP using input variables that are readily available to the operating engineer. The comparison results showed that ANNs have higher generalization abilities and consistently obtain better results compared to regression models (Swider, 2003). Arcaklioğlu, Erişen, and Yilmaz (2004) used ANNs to predict the performance of a vapor compression heat pump when using various R12/R22 ratios refrigerants mixtures. The study showed that ANNs can accurately forecast COP and rational efficiency RE for a heat pump, which was confirmed by $R^2$ values equal to 0.9999 (Arcaklioğlu et al., 2004). Leung, Tse, Lai, and Chow (2012) built ANNs-based model using occupancy space electrical power demand usage to predict building cooling load and produced satisfactory results (Leung et al., 2012). Küçüksille, Selbaş, and Şencan (2011) used data mining to determine the thermodynamic properties of refrigerants, including enthalpy, entropy and volume of alternative refrigerants R134a, R404a, R407c and R410a (Küçüksille et al., 2011).

Above studies confirmed that data mining methodology effectively obtains refrigerant properties under varying temperature and pressure. An important benefit of this methodology is the reduced research time needed to compute vapor compression for a refrigeration system (Küçüksille et al., 2009). The literature shows that a comprehensive review of refrigeration system models can greatly aid users in evaluating refrigeration equipment condition. To increase the energy efficiency and predict performance of refrigeration systems, this study therefore investigated six data mining methods to predict refrigeration system performance. A feature selection process for correlating input and output variables was also used to compare modeling efficiency based on the parsimony rule, *i.e.,* the smallest number of input variables that obtain an adequate output accuracy.

## 3. Research methodology

The data mining and artificial intelligence based approaches use computer system programs to solve problems by emulating human brain processes. The DM methods are developed recently in many fields, including generalization, characterization, classification, clustering, association, evolution, pattern matching, data visualization and meta-rule guided mining (Liao et al., 2012; Şen, Uçar, & Delen, 2012; Tsai, 2012; Yeh & Lien, 2009). Therefore, the use of DM- and AI-based models can predict and improve the future refrigeration system performance.

This work used numerical predictor nodes and six data mining techniques, ANNs, SVMs, CART, MLR, GLR, and CHAID, for