Energy 101 (2016) 541–557

Contents lists available at ScienceDirect

Energy

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

Forecasting electric demand of supply fan using data mining techniques

M. Le Cam [a], A. Daoud [b], R. Zmeureanu [a, *]

[a] Center for Zero Energy Building Studies, Department of Building, Civil, and Environmental Engineering, Concordia University, 1515 St. Catherine W., Montreal, Quebec, H3G 1M8, Canada

[b] Laboratoire des Technologies de l'Énergie, Institut de recherche d'Hydro-Québec, 600 Avenue de la Montagne, Shawinigan, Québec, G9N 7N5, Canada

ARTICLE INFO

Article history:
Received 13 July 2015
Received in revised form
15 December 2015
Accepted 9 February 2016
Available online 24 March 2016

Keywords:
Electric demand
Forecast
Data mining
Building Automation System
Clustering analysis

ABSTRACT

This paper presents the application of the process of KDD (knowledge discovery in databases) for the forecasting of the electrical power demand of a supply fan of an AHU (air handling unit). The case study uses trend data from the BAS (Building Automation System), which is recorded every 15 min in an office building. Data mining techniques are used as a preprocessing step in the development of the forecasting model. A clustering analysis detects atypical operations and then partitions the whole dataset into three subsets of typical daily profiles of the supply fan modulation. A hybrid model, combining a closed-loop nonlinear ANN (autoregressive neural network) model and a physical model, forecasts the electric power demand over a horizon of up to 6 h. The optimum architecture of ANN, found by using a Simple Genetic Algorithm, is composed of 13 input neurons, 1 hidden neuron and 23-day training set size, for the cluster corresponding to working days except Mondays. The results show good agreement between the forecasts and measurements of fan modulation, and electric demand, respectively. The fan modulation was forecasted over the testing period with RMSE (Root Mean Squared Error) of 5.5% and CV(RMSE) of 17.6%. The fan electric demand was forecasted with a RMSE of 1.4 kW, CV(RMSE) of 30% over a 6-h time horizon. The sensitivity analysis indicated that the reduction of training data set size from 23 days to 4 or 8 days does not have a negative impact of the value of RMSE.

1. Introduction

1.1. Background

Buildings account for about 41% of primary energy use in the United States [1], and in Canada, the total consumption of commercial and institutional buildings accounts for 12% of the secondary energy use [2]. During peak demand periods, electric utilities face important mismatch issues in power supply and demand. Demand response programs were developed to involve customers, by offering the opportunity to change their electric usage in response to changes in electricity price or incentive programs. HVAC (Heating, Ventilation and Air Conditioning) systems of commercial and industrial buildings present a significant potential for such demand response programs. The current challenging issue is the development and implementation of forecasting tools that could be used for fast assessment of the demand response with a time horizon of a few hours. In this context, DM (Data Mining) techniques can be used for: (1) the extraction of knowledge from large sets of measurements such as trend data recorded by the BAS (Building Automation System); (2) the development of typical daily profiles of target variables (e.g., electric demand of HVAC systems) under real operating conditions, and (3) the selection of most relevant independent variables for the development of accurate forecasting models. Once new measurements become available, the forecasting models can be easily updated.

The proposed approach applies the process of KDD (Knowledge Discovery in Databases) for the purpose of developing a forecasting model based on a regression DM technique.

The terminological confusion between the terms of KDD (Knowledge Discovery in Databases) and DM (Data Mining) is discussed in Refs. [3,4]. DM corresponds to only one step of the process of KDD, in which information is extracted from available observations. When the DM techniques are used within the process

* Corresponding author.
E-mail addresses: radu.zmeureanu@concordia.ca, rmd.zmeureanu@sympatico.ca (R. Zmeureanu).

http://dx.doi.org/10.1016/j.energy.2016.02.061
0360-5442/

of KDD, the final performance is significantly improved [4]. The processing steps of KDD are detailed in Section 2.

The survey conducted by Kurgan et al. [3] presents the historical evolution of the knowledge discovery systems. The concept of process of KDD was first introduced in 1989 [5]. The first set of processing steps of the KDD process was proposed in 1996 by Fayyad et al. [6]. DM techniques can be classified into six groups (the first two for prediction and the rest for description): classification, regression, clustering, summarization, dependency modeling as well as change and deviation detection [6]. Several other processes of KDD emerged in the mid-1990s, such as the CRISP-DM (CRoss-Industry Standard Process for DM) [3]. Most models compared in Ref. [3] present a similar sequence of steps. The most cited is the nine-step model presented by Refs. [6], which was implemented into a software called MineSet™.

1.2. Applications of DM techniques to HVAC-related studies

The application of DM techniques in such HVAC-related studies is still in the early stages. The first reported application of a KDD process to the building engineering field was presented by Buchheit et al., in 2000 [7] for the purposes of extracting trends in the energy performance of an HVAC system, and proving the value of the KDD process and DM techniques for application in building engineering.

DM algorithms have been employed to develop virtual sensors, which derive information from physical sensors, on variables that are either difficult or expensive to measure. Kusiak et al. [8] used four different DM algorithms to develop virtual models of indoor air quality from twelve measured variables. The best predictions were obtained by a NN (neural network) model with the MAE (mean absolute error) of 0.03 °C for air temperature, 6.42 ppm for carbon dioxide and 0.11% for relative humidity.

DM techniques are useful to analyze occupants' behavior and provide recommendations for energy-saving opportunities. Yu et al. tested three types of DM techniques in Ref. [9]. A clustering analysis based on the K-means algorithm was performed on a database gathering information about residential buildings to group the buildings depending on their level of energy use. The occupants' behavior in each cluster can then be studied. A classification analysis based on a decision tree is developed to estimate the cluster attribution of new buildings. The structure of the decision tree also brings useful information to help understand building occupants' behavior. An association rule mining examines association and correlations among the user activities helping in providing recommendations for energy-saving opportunities.

DM techniques were investigated for the purpose of fault detection and diagnosis in HVAC systems. Du et al. [10] used two combined neural networks to detect anomalies in the supply air temperature control loop of the air handling unit. The two neural networks model the controlled variable (supply air temperature) and its most relevant variable (return water temperature). A fault is detected when the combined relative error between the observations and the predictions of the two networks exceeds a threshold value. A historical dataset of measurements including normal and faulty operation is partitioned by the clustering algorithm into groups corresponding to normal operation and faulty variables. The clustering algorithm diagnoses the source of a new detected fault through an adaptive classification pointing at the faulty variable (e.g., cooling coil valve, temperature sensor …).

DM techniques can bring information about the consumption patterns of customers, characterize their demand profiles and classify new consumers. Ramos et al. [11] used four different clustering algorithms to identify typical electric load profiles of existing customers from a database of a utility company. They assessed the quality of typical daily profiles by using eight cluster validity indices. The best partition of daily profiles was obtained by the K-Means algorithm with four clusters, supported by five out of eight validity indices. The energy consumption corresponding to each cluster was obtained by using a rule-based modeling technique. The paper also presented rules for the automatic classification of new consumers according to those typical load profiles.

1.3. Application of DM techniques for the forecasting of building energy use

The performance of DM techniques is significantly improved when the other processing steps of KDD are performed [4]. Only a few studies have applied the complete process of KDD for the purpose of data mining; some are presented in this section.

DM techniques could help for input selection and data reduction, in the development of predictive models. Kusiak et al. [12] developed four inverse predictive models of the energy use by an AHU, by using DM techniques for the parameter selection and dimension reduction. This approach for the selection of independent variables with the highest impact had the potential of resulting in more accurate models. The forward neural network provided the best results compared with the other three models, and was coupled with a single-objective optimization model that was solved by the particle swarm optimization algorithm. Fan et al. [13] used DM techniques to select inputs, from a large set of measurements, for the forecasting models of the whole building next-day energy consumption and peak demand. The clustering analysis was used to identify and remove abnormal building energy consumption data. Eight different predictive models and one ensemble model (composed of the eight predictive models) were developed. The models were optimized using a genetic algorithm. The MAPE (mean absolute percentage error) for the next-day peak power demand by the eight models was between 3.34% and 8.74%, while for the ensemble model was 2.85%. Kusiak et al. [14] used data mining algorithms to select the most appropriate independent variables, from a data set of measurements over three years, for the development of inverse models for the forecasting of the building steam load of the following year. They compared 10 different data mining algorithms, and concluded that a neural network model (MLP Ensemble) gave the best forecasts of daily steam load with the MAPE of 14.44% over the testing data set.

1.4. Forecasting models

The problem of modeling and predicting the energy use in a building can be very complex due to the number and diversity of factors influencing it, such as HVAC system operation, thermal properties of the building envelope, weather conditions, or occupants and their behavior. Diverse predictive models were developed and tested in the past few decades. The most employed forecasting algorithms can be classified into (1) forward, (2) data-driven and (3) hybrid approaches; they are presented in the next paragraphs. Literature surveys of popular predictive models covering the three groups are presented in Refs. [15–17]. The surveys concluded on the need for refining the models from the BAS trend data to perform forecasts at the system level. Another issue is the use of artificial intelligence techniques for predictive models development.

The forward, or white box, models are based on physical principles. They are mainly used at the design phase, for HVAC equipment sizing, for instance. The forward models are implemented into building performance simulation software, such as DOE-2, EnergyPlus. These tools are effective and accurate; they can be calibrated to fit well with measurements if available. These physical models require detailed knowledge about the building; they are