Energy and Buildings 92 (2015) 322–330

Contents lists available at ScienceDirect

Energy and Buildings

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

Short-term load forecasting in a non-residential building contrasting models and attributes

Joaquim Massana*, Carles Pous, Llorenç Burgas, Joaquim Melendez, Joan Colomer

University of Girona, Campus Montilivi, P4 Building, Girona E17071, Spain

ARTICLE INFO

Article history:
Received 27 October 2014
Received in revised form 2 February 2015
Accepted 3 February 2015
Available online 11 February 2015

Keywords:
Load forecasting
Multilayer perceptron
Regression
Neural networks
Support vector machines
Measured data
Mediterranean climate

ABSTRACT

The electric grid is evolving. Smart grids and demand response systems will increase the performance of the grid in terms of cost efficiency, resilience and safety. Accurate load forecasting is an important issue in the daily operation and control of a power system. A suitable short term load forecasting will enable a utility provider to plan the resources and also to take control measures to balance the supply and demand of electricity.

The aim of this paper is to create a method to forecast the electric load in a non-residential building. Another goal is to analyse what kind of data, as weather, indoor ambient, calendar and building occupancy, is the most relevant in building load forecasting. A simple method, tested with three different models, such as MLR, MLP and SVR, is proposed. The results, from a real case study in the University of Girona, show that the proposed forecast method has high accuracy and low computational cost.

1. Introduction

Electricity is the most important resource in the world economy. But, as can be seen in the Lisbon Treaty [20], electricity needs to be more economic and environmentally clean. The classic electric grid has several disadvantages: losses in transport, centralized power generation, high dependence on large generation plants, etc. The purpose of the new electrical power grid, the smart grid, is to manage loads to shape the load curve and use decentralized generation. Thanks to smart grids, the network will be robust, reliable, efficient and dynamic.

Considering these new concepts, one of the most important challenges of the utilities is to adjust power generation to a user's consumption in real time. An overestimation leads to a waste of resources, whilst an underestimation means an increase in the price to cover additional demand. Now, unlike earlier, this adjustment will be made in smaller environments like microgrids. The electric load forecast is, today, the best way to adjust the two sides of the grid.

Almost 40% of the emissions of CO_2 [2] comes from the building sector, therefore, STLF in the building field is fundamental to reduce energy consumption. The load curves of the electric power consumption of cities are different from the building ones. Building load curves present more variability, noise and non-linearity. Thus, the more disaggregated, the more difficult to predict.

Besides, there are many kinds of buildings, like residential or non-residential buildings. The non-residential buildings, as the university administration sector, has daily, weekly and seasonal patterns in their consumption profile. At nights, holidays and weekends, the consumption is extremely low. During the day there are regular patterns based on the user's activities. Within the week there are regular patterns of consumption. During the year there are regular patterns associated with the seasons. Also, at the beginning and at the end of the working day and in the lunch and breakfast times there are transition areas in the consumption. Like most buildings over 10 years old, HVAC is controlled manually with subjective criteria. In addition, the occupancy of the building and the calendar data can be significant in determining consumption.

Another important issue in the electricity forecast is to know which is the key information to be measured and collected. The goal is to obtain the maximum accuracy with the minimum data attributes and instances.

In this paper, meteorological, indoor environment, occupancy, calendar and electric consumption data have been collected in the University of Girona. Then, three different models have been tested to find which of them gives better performance. MLR, MLP and SVR models have been chosen because they are standards with

* Corresponding author. Tel.: +34 646074566.
E-mail addresses: joaquim.massana@udg.edu (J. Massana), carles.pous@udg.edu (C. Pous), joaquim.melendez@udg.edu (J. Melendez), joan.colomer@udg.edu (J. Colomer).

http://dx.doi.org/10.1016/j.enbuild.2015.02.007
0378-7788/

Nomenclature

b	computed parameter
C	complexity parameter
H_{1-6}	6 indoor relative humidity measure points
H_e	outdoor relative humidity measure point
$k(x_i, x)$	kernel function
L_{1-6}	6 indoor luminosity measure points
L_e	solar radiation measure point
N	number of observations
n	norm of the normalized instances
p	grade of the norm
P_{1-2}	2 occupancy measure points
Q_1	first quartile
Q_3	third quartile
T_{1-6}	6 indoor temperature measure points
T_e	outdoor temperature measure point
x_{in}	normalized instance
$\left\|x_p\right\|$	vector norm
$x_1(t), \ldots, x_n(t)$	independent variables of the regression function
y_m	measured output
y_p	predicted output
$y(t)$	dependent variable of the regression function
α	interquartile range
α_i and α_i^*	Lagrangian multipliers
β	random variable of the regression function
$\beta_0, \beta_1, \ldots, \beta_n$	regression coefficients
γ	outlier factor
ϵ	epsilon insensitive loss function
σ	bandwidth of the Pearson width
ω	tailing factor of the curve fitting peak
ANFI	adaptive neural fuzzy inference
ANN	artificial neural network
AR	autoregressive
ARMA	autoregressive moving average
ARX	autoregressive exogeneous
CC	correlation coefficient
CO_2	carbon dioxide
DW	day of the week
EBP	error back propagation
GA	genetic algorithm
GS	grid search
HD	hour of day
HVAC	heating, ventilation, and air conditioning
M	month of the year
MAPE	mean absolute percentage error
MLP	multilayer perceptron
MLR	multiple linear regression
PCA	principal component analysis
PIR	passive infrared receiver
PM	polynomial model
PUK	Pearson VII universal kernel
STLF	short term load forecasting
SVM	support vector machines
SVR	support vector regression
WD	working day
WSN	wireless sensor network

suitable results. The experiment has been carried out using real data collected with a WSN.

Although, in smart building management it is necessary to forecast the load in the future, this paper focuses only on what is essential for the prediction of consumption, thus avoiding possible errors due to the prediction of the variables.

With the advance of computers, drawbacks like difficult parametrization, selection of variables and over-fitting have been solved. Furthermore, the parallel processing and the performance of the present computers will help to decrease the computational time. However, reducing the size of the database will diminish the computational cost. The features of the presented STLF method for non-residential buildings are: high accuracy, low computational requirements, low over-fitting and, minimum data collected and use as simple a model as possible.

The paper starts with related works and follows with background material. Then, the dataset is explained. This is followed by a presentation of the methodology. Next, the results are presented and the method is discussed. Finally, conclusions are shown.

2. Related works

The present state-of-the-art knowledge focuses on STLF in non-residential buildings. The analysis of the papers that follow is organized according to the following aspects: model type (MLR, ANN, etc.), used variables (load, weather, etc.), building type (malls, offices, university campus, simulated buildings, etc.) and climate type (Mediterranean, Oceanic, Continental, etc.). The works are organized in two major blocks: the works with only one tested model and the works with multiple tested models.

In the first major block, there is the case of [14], where PCA is used, through climate data, to create a new climate index Z. Then, this index, with MLR, is used to estimate electricity consumption. Concerning ANN, study [11] proposes to use consumption and weather data to predict load consumption in offices. This study says that variables such as temperature and solar radiation are important while the wind speed or humidity can be omitted. Ref. [7], with synthetic data of consumption and weather, states that the main virtue of ANN is simplicity. In Ref. [16], with the same data type as two previous cases but for a hotel, a type of ANN called ANFI, optimized via GA, is used. This improves ANN performance. In case [4], where the temperature and load are used to predict the consumption of a set of university buildings, a small set of similar days is selected to train a ANN based on the work activity and temperature. Both, in [25] and [18] cases, in addition to consumption and weather data, calendar related data are used to forecast consumption in offices and university buildings, respectively. The work [25] highlights the differences that can be seen in the results between real data and synthetic data, while [18] indicates that the effect of humidity and solar radiation in the prediction is smaller than the effect of the outside temperature. In [17,13], weather, consumption and a variable related to the occupancy of the building are used. Case [17], a library located in China, also presents an ANFI model optimized through GA. On the other hand [13], discusses the difficulty of obtaining data of the real occupancy of a building and which alternatives there are to collect such data in a mall in Hong Kong. Besides, Kwok et al. [13] mention that the data related to the occupancy of a building significantly improves the prediction accuracy of the MLP case presented. Related to SVR papers, the example [28] uses synthetic data of consumption, weather and occupancy and provides a new method of feature selection. Another example is [27] that with the same type of synthetic data as in the previous case, proposes a new SVR model that improves performance.

With regard to papers that contrast several methods, there are three works by the same authors. In the first paper [19], they compare mainly ANN, SVR, AR and PM for the case of a campus by using consumption, weather and calendar data. The conclusion is that the AR with daily work schedules gives the best results. In addition, they conclude that the methods should be simple and not