Electrical Power and Energy Systems 80 (2016) 171–178

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

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

Clustering based day-ahead and hour-ahead bus load forecasting models

Ioannis P. Panapakidis *

Department of Electrical Engineering, Technological Education Institute of Thessaly, 41110 Larisa, Greece

ARTICLE INFO

Article history:
Received 3 August 2015
Received in revised form 28 November 2015
Accepted 2 January 2016
Available online 10 February 2016

Keywords:
Bus load forecasting
Load profiling
Machine learning
Time-series clustering

ABSTRACT

The importance of Short-Term Load Forecasting (STLF) in power systems planning and management is reflected by the plethora of the related researches. STFL is a popular technical field in the power systems community and already counts many years of research activities and applications. The vast majority of the studies focus at the aggregated system load. Little attention is placed at small size loads, i.e. in buses of the transmission and distribution systems. Since there is a continuous advancement of smart grids technologies involving small size loads, bus STLF is a potentially important tool in smart grid applications. In contrast to system load, bus load presents a high level of stochasticity. Thus a robust forecaster should be able to capture and simulate the attributes of bus loads. The scope of the study is to develop bus forecasting models for day-ahead and hour-ahead load predictions. The models are based on Artificial Neural Networks (ANNs). Using a clustering methodology, the forecasting accuracy of the ANNs is enhanced leading to the formulation of hybrid forecasting models that are characterized by high level of parameterization and efficiency. The developed models are tested on a set of buses covering urban, sub-urban and industrial loads.

Introduction

Power system planning uses a set of tools for the purpose of covering the demand subject to the various techno-economic and environmental constrains. Among the fundamental tools is STLF, a technical field that has gathered the interest of the power systems research community over the last years [1]. A STLF tool is essential for the operation of policy makers, regulators, utilities, retailers, aggregators and others. For instance, retailers require accurate predictions in order to increase their profitability, competiveness and level of influence in the day-ahead market decisions. As for the utilities, high prediction errors lead to increased operation cost of the electricity grid. If the prediction leads to lower estimated demand, there is an inefficient utilization of the power capacity. Thus, the needs of operation reserves will be lower and the operation planning will employ costly marginal power units [2].

Due to the special characteristic of each application, there is no universally acclaimed model that fits every forecasting problem. Taking into account the requirements of each application, various models should be compared in order to determine the one that will present robust performance. However, there are some basic properties that a forecasting model should present: low computation burden, capability of simulation of the human expertise, flexibility, interpretability and exploitation of the results [3]. According to the related literature, computational intelligence techniques like Artificial Neural Networks (ANNs) are becoming more popular mainly due to their capacity of tracking the nonlinear relationship between the demand and the factors that influence it [4]. The models proposed in the literature can be classified into two general classes: *trend methods* and *similar-day approaches* [5]. In the first class, the demand is treated as a function of time, which is obtained as the function that fits best the data under examination. Time is considered as the factor that determines the values of the demand. The prediction is accomplished via a projection based on the available historical load values. In the second class, the aim is to track similarities between the current and historical load values. Apart from the load, these similarities include exogenous variables and time related factors. This class refers to computational intelligence models like ANNs, support vector machines, fuzzy logic and others. For the purpose of determining the optimal parameters of the model, a training set is formed. After the model is trained sufficiently, it is applied on a separate data set, namely the test set. If the forecasting efficiency is not acceptable, the training set is modified and the model is trained again.

For the purpose of lowering the forecasting errors, many researches proposed hybrid models, involving a clustering algorithm with a forecaster. The clustering is used to capture the attributed of the data, such as outliers, periodicities and special

* Tel.: +30 2410 684303; fax: +30 2410 684325.
E-mail address: ipanap@ee.auth.gr

http://dx.doi.org/10.1016/j.ijepes.2016.01.035
0142-0615/

relationships between the data. Through the clustering, the training set is portioned into homogenous clusters. The patterns within the same cluster display higher similarity compared with the rest patterns of the training set. A number of different forecasters are trained with the sets corresponding to the different clusters. At least theoretically, this approach leads to better trained forecasters since they are trained with more correlated data. The forecaster may refer either to time series models or computational intelligence models, i.e. Support Vector Machines (SVMs) or ANNs. The Self-Organizing Map (SOM) combined with SVM is used for peak load forecasting in long-term horizon [6]. The SOM is used to categorize the training data in [7]. For each cluster, separate support vector regression models are applied to predict the daily peak load. In [8] the SOM and the support vector regression are combined with an adaptive fuzzy rule-based forecasting system. The hybrid model is used in a series of cases that involve different magnitudes of time periods for the training and test sets. Ref. [9] utilizes a functional linear regression model for peak forecasting. The peak load is treated as the scalar response of the regression model. The previous day‘s load curve is the functional regressor. Through a functional clustering technique, the analysis of the paper reaches to a solution to the problem of modeling the seasonality effect during each period under study. For each output cluster, an estimation of a certain functional linear coefficient takes place. Authors of [10] combine the deterministic annealing clustering together with ANNs focusing on peak load forecasting. Ref. [11] employs the wavelet transform in order to decompose the clustered load data in four signal components. The forecasting is held for each component separately and the forecasted load is obtained by the reconstruction of the components. The K-means algorithm is combined with the k-NN algorithm to group the clustered data into 4 patterns for Mondays, other weekdays, Saturdays, and Sundays [12]. For each day type class, a time series model is employed. In [13] the Particle Swarm Optimization algorithm is used to define the optimal parameters of the support vector regression models. The clustering is done with the Fuzzy C-means algorithm.

Apart from clustering the load time series employed in the literature, the cluster labels have been regarded in [14,15]. Ref. [14] proposes the novel Pattern Sequence Similarity (PSF) approach. The K-means algorithm is employed for the purpose of grouping and labeling the samples from the dataset. Next, the pattern sequence prior to the day to be predicted is obtained. This sequence is searched in the historical data and the prediction is calculated by averaging all the samples immediately after the matched sequence. In [15] the PSF is combined with ANN. The PSF produces an initial prediction and next this prediction is refined by the ANN. The authors examine a series of variations of the PSF-ANN model that differ in the types of inputs. A hybrid system involving the SOM and ANN is proposed in [16] for both load and price forecasting. The SOM is used to obtain the cluster pattern sequence of the historical time series and the ANN is used to forecast the topological coordinates of next cluster label. According to the topology relations extracted from previous series clustering, next cluster label is predicted as that of the nearest nonzero hits pattern to the forecasted coordinates. Finally, most researchers have proposed models built on feed-forward neural networks trained by modification of the basic back-propagation algorithm. These types of networks have displayed robust learning capabilities, a fact that makes them suitable in prediction problems. An alternate to these networks are Radial Basis Function (RBF) networks. Authors of [17] are assessing five different learning algorithms for RBFs to advance their performance in the load forecasting of the ISO-New England market. Also, the paper proposes a modified version of an existing training algorithm.

While STLF literature is rich and diverse, no sufficient attention is placed in bus loads [18]. A contemporary power system includes a large number of buses of the transmission and distribution systems. Bus load forecasting refers to the consumption forecasting at a given bus. It can on the optimal scheduling of decentralized generation, network congestion studies and others. Bus load differs compared to the aggregated load of the system. Bus load is characterized by high complexity due to the presence of more atypical patterns, nonlinearity and volatility. Due to these reasons, forecasting accuracy is considerably lower. The literature focusing in bus load forecasting involves mainly ANNs solely applied or combined with other models like ARIMA [19–23]. Referring to the use of clustering in small loads, authors of [24] highlight the need to develop forecasters for microgrids environment. The focus of interest is a city in Spain. The load varies between 7 and 39 MW. The proposed model uses the combination of the SOM, k-means and multilayer perceptron.

The present study focuses on bus STLF. The data set under examine refer the hourly load values of 10 buses located in the area of Thessaloniki, North Greece. The available data set covers the period between 01/01/2011 and 31/10/2015. The buses under study are located within the prefecture of Thessaloniki, supplying the main city and the sub-urban areas. Among them almost 80% of the data are used for training set and the rest 20% is used as test set. The training set is used to define the optimal ANN configuration via a detailed parametric analysis. The test set is employed for the models comparison. Since there is an absence of studies focusing of buses of the Greek system, the first task is to test the ANN that has been proposed for the Greek interconnected system, hereafter referred to as Model A. Next using a clustering process, the training set is clustered into a specific number of clusters. For each cluster, a different ANN is implemented, leading to a hybrid model, hereafter referred to as Model B, that combines unsupervised machine learning via the clustering stage and supervised learning via the ANN. The proposed model differs from the other hybrid models of the literature in terms of the clustering pattern used; Model B refers to the clustering of patterns that apart from the historical load values include all the other variables used for the ANN training. For instance, in [24] the patterns used for clustering (SOM and K-means) differ from those used as inputs in the ANN. By using the same patterns both in the clustering and forecasting phases, the hybrid model operates in a more synergistic manner; no separate data bank is need for the clustering and forecasting stage. Moreover, in [24] the selection of the model in the test stage is done only by using the type of the day and season. In the models proposed in this study, the selection involves also temperature values and the most correlated historical loads. Also, two models are proposed for hour-ahead forecasting. To the best of the author‘s knowledge, this is the first study that proposes hybrid model for day-ahead predictions. Hour-ahead forecasts are potentially essential in smart grid paradigms with prosumers that bid load reductions hourly, in distribution system lines congestion management and operational planning of distributed generation and storage technologies.

Day-ahead load forecasting models

Model A

Due to the absence of previous researches on bus STLF within the Greek system, a benchmark model is constructed, in order to check the robustness of the hybrid model. It refers to a slight variation of the ANN proposed for the total load of the interconnected system [25,26]. Apart from exogenous variables like temperature, load presents high correlation with its past values [3]. The first task is to examine the short-term periodicity of the load using the correlation coefficient [27]. Fig. 1 presents the relationship between