Short-term load forecasting in large scale electrical utility using artificial neural network

Slobodan Ilić\textsuperscript{1,*}, Aleksandar Selakov\textsuperscript{2,*}, Srdjan Vukmirovic\textsuperscript{1,*}, Aleksandar Erdeljan\textsuperscript{1} and Filip Kulic\textsuperscript{1}

\textsuperscript{1}Computer and Control Department, University of Novi Sad, Trg Dositeja Obradovića 6, Novi Sad, Serbia, Europe
\textsuperscript{2}Telvent DMS LLC, Novi Sad, Serbia, Europe

Received 20 July 2012; revised 18 April 2013; accepted 22 August 2013

This paper presents a novel method for short-term load forecasting (STLF), based on artificial neural network (ANN), targeted for use in large-scale systems such as distribution management system (DMS). The system comprises of a preprocessing unit (PPU) and a feed forward ANN ordered in a sequence. PPU prepares the data and feeds them as input to the ANN, which calculates the hourly load forecasts. Preprocessing of the entering data reduces the size of the input space to the ANN, which improves the generalization capability and shortens the training time of the network. Reduced dimension of the input space also diminishes the number of parameters to be set in a training procedure, allowing smaller training set, and thus online usage and adaptation. This is important for a real-world power system where a sufficient set of historical data (training points) may not always be available, for different reasons. Ease of use and fast adaptation are necessary when predictions need to carry out in a large number of nodes in the power grid. Functionality of the proposed method has been tested on recorded data from Serbian electrical utility. Results demonstrate that even with a simple configuration such as this one, fair accuracy can be achieved in forecasting the hourly load. The simplicity and reusability are very important factors for installation of the proposed system in a large-scale DMS, considering the technical requirements (e.g. training data availability, processing power and memory capacity).

Keywords: Load Forecasting, Artificial Neural Networks, Prediction Model, Power Grid

Introduction

Predicting the electricity load from one to several days ahead is referred to as Short-Term Load Forecasting (STLF). Precise STLF is a key requirement for power system managing and operation. Decisions, which are important to the system’s performance such as generator dispatching, unit commitment, reliability analysis, economic calculations, security assessment and maintenance plans, are based on STLF functionality. Load forecasts are very important in open market energy transactions, since profits and market shares are significantly influenced by forecasting errors. Every managing decision of a system operator has to be based on as much reliable data as possible, namely on load forecasting results. STLF is becoming increasingly difficult due to non-stationary character and variability of the time series, resulting from dynamic bidding strategies, time varying electricity price and price dependent loads. Moreover, the constant growth of the size of power grid complicates the processes involved in load forecasting. Therefore, advanced and more sophisticated methods for STLF need to be developed for a modern power system.

This paper introduces a method that tackles the problem of STLF, which is intended for use in a real-world environment, with real-world problems, such as limited data availability. It is based on an Artificial Neural Network (ANN), which has a reduced input space, and is thus eligible for online training and adaptation in systems where a large number of forecasts must run in parallel. This means that many different software (or hardware) instances of the prediction model must exist in the power system, depending on the size of the power system, and all their parameters need to be determined and stored. The type of the ANN used is a Feed Forward (FF) Multi-Layer Perceptron (MLP). The scheme of the models structure is shown in Figure 1. The inputs to the algorithm are the load and temperature measurements recorded for a specific consumer in the power grid. First step in the sequence of the algorithm is data preprocessing, which is in charge of the preprocessing unit (PPU). PPU calculates the integrals of the load and the temperature before it feeds them to the MLP, as inputs. The second step of the algorithm...
is the actual load forecasting, which the MLP accomplishes. The outputs represent the predicted hourly load consumption of the forecasting day.

**ANN in short-term load forecasting**

Artificial Neural Networks are computational models, either software or hardware, which are inspired by biological systems, that is by structure and behavior of the biological neurons. The resemblance to biology however, ceases after this point of inspiration. Feed forward neural networks are mainly used to solve classification and regression problems which are nonlinear, by learning from data. ANN’s are composed of many computing units called neurons. The strengths of the connections between two neurons, called the weights, are true network parameters, and are subject to learning. The formation of neurons is typically in layers, where all neurons in one layer usually pose the same activation function (AF). The nonlinear AFs in the hidden layer of the ANN enable it to be a universal approximator. Nonlinear training is possible because of the differentiability of the HL neurons’ AFs. The important ability of the ANN is its possibility to perform the task of learning. The most elementary learning algorithm is the error back propagation algorithm (EBP). The basic idea behind the EBP is that the error signal terms for HL neurons are calculated by back propagating the error signal terms of the OL neuron.

This section, will give a brief introduction of usage of ANNs in forecasting the load consumption (demand) or any other (financial, weather, biomedical) time series. Because of their generalization capability ANNs are used for prediction in diverse fields such as environmental, social and historical data. It is the matter of debate which parameters are the most correlated to the prediction of behavior of the process in question. Furthermore, it is not always useful to apply the results obtained for any specific application, to a larger class of problems, because any different field of STLF displays different properties of their own (related to the consumer type, the size of the geographical area, the climate zone, etc.). The structure that is the most commonly used for STLF is the Multi Layer Perceptron (MLP). In the next section, we shall give an overview of the proposed methodologies of STLF, and describe how our model relates to some of the accepted structures that use NNs. In section 4, we give a detailed description of the MLP used herein, method in which the test procedure was carried out, together with the test results and their graphical interpretation.

**Related work**

Various methods have been proposed to solve the demanding task of STLF, especially over the past few decades. Time-series techniques represent the early approach to the problem of STLF. Regression methods, Exponential smoothing, Box-Jenkins models, Kalman filter and state space model were also considered in the related field. In recent years, research has converged towards methods that use Artificial Intelligence (AI) such as fuzzy time series, artificial neural networks (ANN), and fuzzy ANNs. The number of different STLF methods and applications increased over the course of the past decades, as has been explained in. This tendency is mainly the result of the constant growth of power industries where many different factors affect the load consumption, and also of the trend to apply newly developed and tested methodologies. The increasing complexity of the processes involved in STLF demands the development of more accurate forecasts, especially in the electricity markets. In order to deal with these complexities, hybrid methods have been developed recently. In fuzzy linear regression is combined with general exponential smoothing. Hybrid algorithms, where a two-stage hybrid network based on support vector machine (SVM) and a self-organizing map (SOM) is developed, or, where a hierarchical structure of ANN is presented, have also been proposed. In a technique based on wavelet decomposition and a neuro-evolutionary
algorithm is developed and tested. ANNs that forecast the integrated load value have been presented in literature\textsuperscript{18}, and they have the capability to predict single load values such as maximum, minimum and the average daily load. Forecasting the load in a smart grid, by means of a bi-level prediction strategy, that combines a feature selection method, and a forecast engine, has been investigated in\textsuperscript{19}. In\textsuperscript{18}, the authors developed a three-layer ANN for solving the STLF problem. The input to the system represents the values, which are defined as follows:

- \(IL\ (d-1)\)-The Load Integral for 24 hours prior to forecasting day
- \(Max T\ (d-1)\)-Max. Temperature for the day prior to forecasting day
- \(Min T\ (d-1)\)-Min. Temperature for the day prior to forecasting day
- \(Max T\ (d)\)-Max. Temperature of the forecasting day
- \(Min T\ (d)\)-Min. Temperature of the forecasting day

In the output layer, there is only one neuron, and it gives the forecasted integrated load, for the forecasting day- \(IL\ (d)\). A variable called the “Integrated Load” defines as the total sum of twenty-four hourly load readings recorded for one day. The value that represents the integrated load implicitly carries the information about the local maximums (peak loads) and local minimums (troughs) for the considered load time series. In recent papers\textsuperscript{19,20,21} the authors demonstrated that minimum and maximum daily temperature affect the integrated load, together with the type of the day considered. In\textsuperscript{20}, a hybrid model is proposed for the load forecasting, which consists of a SOM, and an MLP. In this architecture, at one step, an MLP is used for the mean daily load forecasting. For the oscillation forecasting, the hybrid structure composed by an MLP and a SOM is used. The ANN used for predicting the mean daily load is a three layer MLP.

In\textsuperscript{10}, authors present a detailed review and evaluation of the NN models used to predict the electricity load, with the division in two categories. The first category of NN used has only one output node, typically predicting some characteristic value, i.e. the peak load, the load of the particular hour, or the total load\textsuperscript{22}. The second category usually has 24 output nodes, predicting the load profile of the forecasting day.

Procedure of predicting the load for all 24 hours of the prediction day at once is referred to as \textit{Single-Model Multivariate Forecasting}. This method has two serious drawbacks, according to\textsuperscript{10}. The first one is that NNs have to be very large in order to accommodate a 24 dimensional output vector. Furthermore, if the loads of one or two days previous to the forecasting day are used as time series, the number of parameters to be set in a training procedure will increase to the point where it would be impossible to determine them. The second one is that one year of training data will yield only 365 training points, which is clearly not enough to determine thousands of parameters resulting from a model this complex.

In this paper we also use 24 output nodes, but instead of using time series of the load, or temperature as inputs, we use their integrated values, thus reducing the input space to only three variables (together with the type of the day), and thus decreasing greatly the number of NN parameters. A structure of 24 smaller NNs with only one output node, operated in parallel is also tried out, and the results are compared to the originally proposed structure (with 24 output nodes). The advantage of this method, as claimed by\textsuperscript{10} is that NNs are smaller, and are not likely to be over fitted.

\textbf{Proposed structure}

The structure of the ANN proposed in this paper is motivated by the idea that the simple MLP can be effectively used in predicting the integrated load value of the next day\textsuperscript{18}. Also, in\textsuperscript{22} a simple MLP is used in a number of configurations, one of them having the total load of the previous day as its input. We demonstrate that there is a strong nonlinear correlation between the values of the integrated load on the day prior to forecasting and of the forecasted integrated temperature (i.e. the total environmental energy that describes thermal inertia) on the forecasting day, to the predicted load profile.

The algorithm employed herein comprises of two separate processing units, the first one being the preprocessing unit (PPU), and the second one the actual NN that calculates the load profile of the forecasting day. PPU carries out the task of adjusting the raw measurements of the electric load and temperature. This step is crucial, because it transforms the time series signals to single values enabling a significant reduction of the input space to the forecasting unit, and thus reducing the number of parameters to determine in training procedure. The process of integration could be thought of as some sort of feature extraction, since the integrated values of both the load and temperature carry the information about the total amount of energy demanded from
the power system (consumers) and transmitted to it (environmentally).

This preprocessing is done by integrating the load and temperature data in the following way:

\[ \int_{t=0}^{t=24h} L_{d-1}(t) dt - \text{Previous day integrated load} \quad \ldots (1) \]

\[ \int_{t=0}^{t=24h} T_{d}(t) dt - \text{Forecasting day integrated temperature} \quad \ldots (2) \]

Temperature measurements are not always accessible on an hourly basis, so the interpolation of missing data has to be done. This is achieved by cubical spline interpolation. When there are no temperature measurements at all registered in the electrical utility system, a public weather service has to be consulted instead. The algorithm of calculating the integrated value comes down to summing the hourly values of temperature and load, once they are accessible and pre-arranged.

In this research, the standard of STLF by NN was used, that being the configuration of a MLP with one hidden layer. The structure of the network is feed-forward, and it is fully connected. For the input layer, three nodes were chosen, with the idea of demonstrating that even with such simple configuration, strong correlation to the outputs may be achieved, and those nodes represent:
- Integrated Load of the previous day
- Integrated Temperature of the forecasting day
- The type of the forecasting day

When considering the output layer of the NN two different structures were developed and tested. The first configuration uses 24-dimensional output vector, while the second one uses 24 parallel NNs with only one output node. The result of both configurations is the predicted load profile, and the performance of both structures is presented in section 5. The 24-outputs configuration undergoes only one training procedure, which sets all of its parameters, while the other configuration undergoes 24 different training procedures, one for each predicted hour.

There is no any well-accepted or mathematically approved procedure when determining the number of the hidden neurons. Most authors use the process of trial and error\(^{10}\), while having in mind that if the hidden neurons are too few, the model won’t be able to model the data well, and if there are too many, the model will become too complex, likely leading to over fitting. The criterion was to maximize the generalization capability of the network, and to reduce the errors of forecasting, and the choice was to use five neurons, for both structures (one large NN, and 24 small ones). This configuration of hidden neurons provided the good results regarding least mean square error, while in comparison with other tested structures of the similar complexity there were no significant differences. Moreover, the networks showed resistance to anomalous spikes, due to the integration effect, and had always produced a meaningful output, in all of our tests cases.

The MLP training set consists of input and target vectors. The input vectors are generated by integrating the daily power load measurements of the day before the forecasting day, and by integrating the predicted hourly temperature measurements of the forecasting day. Also, one additional member of the input vector is the type of the forecasting day. A simple way of coding the data type has been adopted, where the workdays are coded as 1, while the weekdays are coded as 0. The holydays are not taken into consideration. The output vectors are obtained by selecting the daily power load measurements for the forecasting day. The input and output vectors are normalized with respect to their minimum and maximum values.

ANN was trained using the modified back-propagation algorithm with momentum and adaptive learning rate\(^{23}\). The weights were updated by the following formula:

\[ \Delta w_{ji}(n) = \eta \delta_j(n) y_j(n) + \alpha \Delta w_{ji}(n-1) \quad \ldots (3) \]

in which \( n \) is the epoch, \( \eta \) is the learning rate, while \( \alpha \) represents the momentum (between zero and one). While the training is performed, the value of the learning rate is changed dynamically according to the current epoch global error. It is either increased or decreased in comparison with the global error of the previous epoch. It is necessary that the system avoid the training abrupt interruption caused by local minimum typical of error surface. The number of epochs was not pre-specified, because the training procedure was performed in Matlab using Neural Network toolbox, which continues to update the ANN parameters, as long as the error on the test subject keeps diminishing, without over-fitting the network.
Testing and results

The model proposed in this paper was implemented in Matlab, using Neural Network toolbox. For MLP part of the algorithm, the feed forward structure of the ANN was used. Activation functions of the hidden layer neurons are of sigmoid type, while those of the output layer neurons are linear\(^1\). The proposed model was tested over a data set, containing historical measurements of electrical load, and of the environmental temperature. The actual values of measurements are obtained for a specific consumer in the power grid, provided by a Serbian electrical utility. It is important to state that the only available data that were recorded for that specific consumer are from the year 2006. This physical limitation further justifies usage of a simple configuration, since the training set can be diminished to as little as 30 days. Each measurement was recorded on an hourly basis, and used as such in the calculation of models performances.

The basic means for model evaluation is the prediction error, which it generates. The error measure which is most commonly employed in the field of STLF, and which was used in the evaluation of the results here presented, is the mean absolute percentage error (MAPE), which defines as:

\[
MAPE = \left( \frac{|x_i - y_i|}{x_i} \right) \times 100 \quad \ldots (4)
\]

where \(x_i\) represent the actual values and \(y_i\) represent the predicted values at time instance - \(i\).

The basic testing procedure is as follows: First a set of 30 days (i.e. 30 training points) prior to the forecasting day is selected. The inputs to the network represent integrated load of the previous day, recorded temperature of the forecasting day, and the type of the forecasting day. The outputs to the network represent the load profile of the forecasting day, for both the structures of one large network, and twenty-four small NNs operating in parallel. It is important to state here that the measurements on which the performance of the network is assessed were unseen in the training procedure, which is also the situation that would be encountered in a real-world situation. The other important thing is to note that the time of the day when the forecasting is performed is the midnight. Since the input to the NN is the type of the next day, all of the predicted values must belong to that day type, leaving the midnight as the only possible choice. This limitation may seem unaccommodating for usage in a large power system, but since the focus here is to demonstrate the feasibility of simple models, we leave it so. The following figures show the systematical results of the proposed model when applied to real life measurements in simulation mode. In Figure 2, the predictions for the first week of February 2006 are shown, obtained by the structure with 24 small NNs operating in parallel. In Figure 3 the same period of prediction is shown but with the large NN structure (24 output neurons) used instead.

In Figures 4 and 5 are the graphs of absolute percentage error (APE) and absolute error (AE) for the prediction period of first week of February 2006, for both methods tested in this paper, respectively.
The information about MAPE and mean absolute error (MAE) is also displayed on those graphs.

Although the structure with 24 NNs operating in parallel may seem like a more appropriate choice, due to resistance of such a small (individual) network to over fitting, we can see from the error analysis that the large structure with 24 output neurons introduces smaller MAPE as well as MAE. Furthermore, in regard of time needed to carry out the training procedure, it is shorter for one NN than for 24 parallel ones.

It is important to mention that recorded temperatures were used in all our test cases, both for training and simulation of predictions. In an industrial application where the system would be installed, the predictions of the temperature would be used instead as input to NN. This would introduce additional error in the performance of the system. However, due to unavailability of such data (predicted hourly temperature for some period in history) the analysis of such behavior has not been investigated.

**Conclusion**

Rapid social and economic development of modern urban areas imposes a need for scientific planning of utility operations and management, where load forecasting is one of crucial points. To improve the accuracy of forecasting, and to preserve the simplicity of the model for large number of consumer types, a simple approach based on ANN has been proposed. The model consists of two units, the first one being a preprocessing stage that calculates the integrals of the load and temperature measurements, while the second one represents a simple ANN that acts as hourly load predictor.

The proposed method was tested with real-life industrial measurements provided by a Serbian electrical utility. It showed the capability of generating acceptable forecasts, while retaining the simple structure of the model. These properties enable for online training and adaptation, which is a necessity for a commercial, real-world application. Moreover, if we apply the algorithm to a large number of consumer types in the power grid, which is its basic intention of use, the averaging effect in the overall forecast will reduce the MAPE even further.

Two different implementations of the proposed method have been presented and compared. The first one consisted of a single NN that has 24 outputs for predicting the load profile, while the second one was accomplished by 24 separate NNs with only one output. It can be seen in the results section, and especially in the APE and AE graphs that the first one of the tested structures provided somewhat better results, and therefore presents a more suitable choice for installation in a power system.

The proposed structure could present potential drawbacks, mainly in regard of the forecasting precision. The model was not fine-tuned to produce perfect forecasts, neither was it its goal in the first place. The accent is mainly on the application of the simple and fast solution to a large software system, such as DMS, where forecastsexecutein a large number of power grid nodes, in a near-real-time fashion.

Finally, it is worth mentioning that the proposed results still have the space for improvement. For example, using evolutional algorithms for ANN training, such as Genetic Algorithm (GA), or Particle Swarm Optimization (PSO), will eliminate the influence of hand set parameters, and probably yield better results.

**Reference**

20 Caciotta M, Giarnetti S & Leccese F, Hybrid neural network system for electric load forecasting of telecommunication station, in *Proceed of XIX IMEKO World Congr Fundamental and Appl Metrol* (Lisbon, Portugal 2009), 657-661.