



A REVIEW OF MACHINE LEARNING ALGORITHMS USED FOR LOAD FORECASTING AT MICROGRID LEVEL

Enea Mele

National and Kapodistrian University of
Athens,
Athens, Greece

Abstract:

As load forecasting nowadays is a crucial and integral part of the energy production procedures a large number of forecasting methods has been proposed to address it. However, although there are many forecasting methods which take into account the advances in information, metering and control technologies in order to address the challenges of forecasting problems, the accuracy and efficiency levels required for each type of applications are yet to be determined. Technologies such as machine learning techniques have been proven useful for short-term electricity load forecasting especially in microgrids where a large variety of data should be included in the energy consumption prognosis. In this paper, we present an overview of the main machine learning algorithms applied to electricity load datasets for short-term forecasting such as Support Vector Machines (SVM), k-Nearest Neighbors (kNN), Random Forest and Artificial Neural Networks (ANN) and compare their performance efficiency, capabilities and limitations.

Keywords:

machine learning; short-term forecasting; electricity load profile; renewable energy sources.

1. INTRODUCTION

The penetration of distributed and renewable generation is increasingly growing and new systems such as microgrids are being implemented in the distribution network. This decentralized structure poses new challenges for reliable and quality services of the electricity distribution network and entities. In order to cope with these challenges tools such as short-term load forecasting have become important aspects of the decision making process for purchasing, selling as well as generating electric power, demand side management, and infrastructure development.

Accurate models for electric power load forecasting are essential to the operation and planning of microgrids, especially given the deregulation of the electricity market and the anticipated price fluctuations in all four markets of the target model. Novel techniques and models are taking advantage of the advances in artificial intelligence algorithms which allow for faster convergence, manipulating big data sets and solving more complex problems.

Correspondence:

Enea Mele

e-mail:

dimitrismele@gmail.com



Algorithms that use available data for the training of a model, where the underlying system is more or less treated as a black box, are usually grouped under the term machine learning. Arthur Samuel was the one who coined the term "Machine Learning" in 1959, defining it as "the ability to learn without being explicitly programmed." Machine Learning, at its most basic form, is the practice of using algorithms to parse data, learn from them, and then make a determination or prediction about something in the world.

Many of the recent advances in smart devices are the result of deep learning, a subset of machine learning that uses layers of neural networks with which it is able to develop computer models that are roughly based on the structure of the human brain neurons. When machine learning systems are combined with big data they provide new capabilities and emerging technologies [1].

So instead of hard coding software routines with specific instructions to accomplish a particular task, machine learning is a way of "training" an algorithm to learn by itself. "Training" involves feeding huge amounts of data to the algorithm and allowing the algorithm to adjust itself and improve to perform a given task better and more quickly than expert humans [2], [1].

Machine learning is being used extensively in many problems where big amounts of data are generated, such as weather forecasting. In this work we focus on the electrical load. Long and midterm forecasting has been used for years by all electricity companies using more conventional algorithms, such as linear regression [3] and econometric models [4] with acceptable accuracy. The transition to the smart grid and the emergence of microgrids and distributed generation and storage is re-stating the forecasting problem and setting new requirements. The amount of data generated by the new grid architecture increases dramatically due to distributed generation, smart meters, electrical vehicles, and cannot be handled efficiently by conventional algorithms. Algorithms that are able to classify these data and use them to train systems are required [5]. For the management of microgrids, the short term forecasting problem needs to be solved. In order to attain high accuracy levels, access to a large amount of data such as the day of the week, the month of the year, the respective days and months of past years along with past and future environmental data are required.

This paper aims to establish the state of the art and investigate the latest trends and open challenges in short term forecasting for microgrids. In the following sections, we will present briefly the main families of

algorithms used in these problems, such as support vector machines (SVM), k-Nearest Neighbors (kNN), Random Forest and artificial neural networks (ANN), and their performance with respect to the given problem.

2. MACHINE LEARNING

A more precise formalism regarding machine learning was given by Tom Michell in his book "Machine Learning" that has been repeated a lot since then: "Machine Learning is the process which, a computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ." [6]

Based on this definition it is easy to say that machine learning is closely related to computational statistics, which also focuses on prediction making through the use of computers and has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning tasks may be classified into three broad categories, namely supervised, unsupervised and semi-supervised, depending on whether they use labeled or unlabeled data or both. Unlabeled data usually consist of samples artifacts that are created either naturally or by humans and can be obtained readily through sensors. On the other hand, labeled data typically takes a set of unlabeled data and augments each piece of that unlabeled data with some sort of meaningful "tag," "label," or "class" that is somehow informative or desirable to know. After obtaining a labeled dataset, machine learning models can be applied to the data so that new unlabeled data can be presented to the model and a likely label can be guessed or predicted for that piece of unlabeled data.

Supervised learning

The training data consist of a set of training examples. Each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. More specifically, a supervised learning algorithm takes a known set of input data and known responses to the data (output), and divides them into two sets, one called training set and one called test set. The observations in the training set comprise the



experience that the algorithm uses to learn. The test set is a similar collection of observations that is used to evaluate the performance of the model [7]. For example, a forecasting model could use the past data on the energy consumption of a building to create two sets, one for training and one for testing, use the first to train the algorithm to predict the future consumptions and then compare its results against the second set.

Some of the algorithms that have been used in supervised learning techniques for load forecasting are the Decision Trees, k-Nearest Neighbors (kNN), Naïve Bayes and Support Vector Regression (SVR).

Unsupervised learning

Unsupervised learning can be described as learning without a master. The algorithm is “left on its own” to determine the underlying structure or an input-output relationship using unlabeled data training sets. There is no feedback based on the prediction results, no error or reward signals are issued to evaluate a potential solution. The algorithms must recognize patterns or discover relationships between groups of input data [6]. Unsupervised learning uses procedures that attempt to find natural partitions of patterns as it does not always know what it is searching for, but can independently sort data and find compelling patterns. For example, an unsupervised load forecasting algorithm based on the association feature should be able to generate a pattern able to detect that the energy consumption at a given home is increased in the morning hours but is decreased during the evening and afternoon hours.

Some of the algorithms that are mostly used in unsupervised learning load forecasting are the K-Mean, the k-Nearest Neighbors (kNN) and the Neural Networks.

Semi-Supervised learning

Between supervised and unsupervised learning, there exists the semi-supervised learning. In this case, algorithms are trained on a combination of labeled and unlabeled data which means that the computer is provided with an incomplete training set in which some of the target outputs are missing. Semi-supervised learning is useful for various reasons. First, the process of labeling massive amounts of data for supervised learning is often prohibitively time-consuming and expensive. What’s more, too much labeling can introduce human biases on the model. That means that including lots of

unlabeled data during the training process actually tends to improve the accuracy of the final model while reducing the time and cost spent building it [6], [7].

Another categorization of machine learning tasks is by the desired output. If the output of the model is a class then it is a classification problem, if it is a number then it is a regression problem and if it is a set of input groups, it’s a clustering problem.

In classification, inputs are divided into two or more classes, and the learner must produce a model that assigns unknown inputs to one or more (multi-label classification) of these classes. This is typically accomplished using supervised learning algorithms. In regression, which is also a supervised problem, the outputs are continuous rather than discrete and in clustering, a set of inputs is to be divided into groups. Unlike classification, the groups are not known beforehand, making this typically an unsupervised task.

3. LOAD FORECASTING MACHINE LEARNING ALGORITHMS

Electrical load forecasting algorithms are needed for prediction of the energy demand for the day ahead, a few weeks up to a year or a period of over a year. Traditional clustering and forecasting methods cannot have direct application to microgrids as the required power load is not only several times smaller than in region-wide areas, but also its curve presents a much higher volatility making the development of new techniques that deal with the higher uncertainty of the load behavior necessary [5].

In existing literature, a wild variety of algorithms have been designed to solve specific load forecasting problems with different levels of success. In the conventional grid, which was developed based on the vertically integrated state-owned electricity companies, forecasting was used for long – and midterm forecasting for economic operation and planning. The increasing penetration of renewable energy sources (RES), the interconnections between countries, cross-border electricity trading and wholesale energy trading created the need for next-day forecasting. Power companies have traditionally used simple forecasting models, like linear regression [3] and econometric models [4]. Nowadays, multiple regression models are used for very large systems [8], [9], large metropolitan areas [10], or small areas [11]. Artificial intelligence techniques have also been applied to energy forecasting using either neural networks [12], [13], or fuzzy logic [14], [15], [16].



Lately, efforts concentrate on unsupervised learning neural networks such as Self Organizing Maps (SOMs) [17] - [18], on Neural Networks [19] or on hybrid systems that combine both Self Organizing maps (SOMs) and algorithms such as support vector machines (SVMs) [20] or k-nearest Neighbors algorithm (k-NN) [5].

In the next section we focus on short and very short - term forecasting which is of interest for microgrids. Short term load forecasting (STLF) refers to a time horizon of one day or even less while the very short-term load forecasting (VSTLF) refers to a time horizon of one minute or less.

Table 1. Cycle updates & time horizons of load forecasting

	<i>Cycle Update</i>	<i>Horizon</i>
VSTLF	>= 1 Minute	1 day
STLF	1 Day	2 weeks
MTLF	1 Month	3 Years
LTLF	3 Years	30 Years

Table II summarizes the different time horizons of forecasting techniques with respect to different applications related to operations of the power grid.

Table 2. Applications of electric load forecasting

	<i>VSTLF</i>	<i>STLF</i>	<i>MTLF</i>	<i>LTLF</i>
Purchase and production of electric power	✓	✓	✓	✓
Transmission, transfer and distribution of electric power		✓	✓	✓
Management and maintenance of electric power sources	✓	✓		
Management of the daily electric load demand	✓	✓	✓	✓
Financial and marketing planning			✓	✓

Long-term and medium-term forecasting despite the fact that they are not being used in microgrids because the forecast horizon in microgrids varies only from 1

day to 1 minute, still, they are quite important for other tasks. Long term energy forecasting is important for planning and expanding the power grid while midterm power and energy forecasting is needed for scheduling the fuel supplies, electrical energy imports/exports and the rational exploitation of the water reserves for optimal hydrothermal scheduling. Therefore, a wide range of models have been designed for these two forecast periods. For instance, an application based on fuzzy logic was developed for midterm forecasting which transformed the candidate input variables to differences or relative differences and selected the one to be used through correlation analysis [15].

4. SHORT-TERM & VERY SHORT-TERM FORECASTING

Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression tasks and its formulation is based on the Structural Risk Minimization (SRM) principle, which has been shown to be superior to the traditional Empirical Risk Minimization (ERM) principle, used by conventional neural networks. SRM minimizes an upper bound on the expected risk, where as ERM minimizes the error on the training data. It is this difference which equips SVM with a greater ability to generalize [21].

The advantage of SVM is that it is effective in high dimensional spaces and uses a subset of training points in the decision function making it also memory efficient. However, the algorithm does not directly provide probability estimates, because they are calculated using an expensive five-fold cross-validation.

Support Vector Machines models can be found in many forecasting models for both medium-term forecasting and short-term forecasting or in combination with other algorithms. A case where SVM was compared favourably against other algorithms, such as the auto-regressive (AR) model using the same data for building and testing both models based on the root-mean-square errors between the actual and the predicted data is presented in [22] where SVM allowed the training data set to be increased beyond what is possible using the auto-regressive model or other neural networks methods. Increasing the training data to about two years improved the performance significantly based on the RMSE on testing



data that was not used on training or building the model. When 20000 data points were used for training the SVM model the root-mean-square error was reduced to 0.0028 compared with 0.0376 for the best results obtained using the AR model. In [23], a STLF method based on clustering using support vector machines (SVM) predicts the next day 48 half hourly loads. The daily average load of each day for all the training and testing patterns is first calculated and the patterns are clustered using a threshold value between the daily average load of the testing pattern and the daily average load of the training patterns.

k-Nearest Neighbors

k-Nearest Neighbors (kNN) is a supervised machine learning algorithm used for both classification and regression predictive problems.

kNN is a simple algorithm that stores all available cases and predicts the numerical target based on a similarity measure, e.g., distance, proximity, or closeness functions, as it is based on the assumption that similar data points are close to each other. It has been used as a non-parametric technique, meaning that it does not make any assumptions on the underlying data distribution, since the 1970's. kNN methods are known as non-generalized machine learning methods, since they simply "remember" all their training data and do not have the ability to generalize.

The main advantage of kNN is that it is robust to noisy training data, and effective if the training dataset is large. On the other hand, the user needs to determine the value of K, which corresponds to a positive integer, typically small, and the computation cost is high as it needs to compute the distance of each instance to all the training samples.

kNN models can be found in many forecasting models. For example a kNN algorithm was used to construct a novel hybrid model which comprised of a Wavelet Denoising-Extreme Learning Machine (ELM) and a kNN Regression in order to forecast the short-term electricity load in New South Wales. [24]. In [25], kNN models were used for day ahead load prediction, with only limited temperature forecasts, namely minimum and maximum temperature of a day, as the forecasting input. Recently the combination of kNN with Self Organizing Maps (SOM) to create a novel clustering and forecasting tool at the microgrid level has been proposed. The SOM algorithm carries out the clustering process and when all the data are clustered they serve as input to kNN for forecasting [5].

Random Forests

The Random Forest (RF) technique is used in many forecasting models due to its simplicity and the fact that it can be used for both classification and regression tasks. RF is a supervised learning algorithm which builds multiple decision trees and merges them together to get a more accurate and stable prediction. More specifically, it constructs a multitude of decision trees during training and yields the class that is the mode of the classes (classification) or mean prediction of the individual trees (regression). RF correct the decision trees habit of over fitting their training set because no single tree can learn from all of the instances and explanatory variables. No single tree can memorize all of the noise in the representation.

Its default hyperparameters often produce a good prediction result which make RF a very handy and easy to use algorithm. Its main limitation is the number of trees which it needs to yield these results. As the trees grow, the algorithm speed slows down and it becomes ineffective for real-time predictions.

RF has been widely used in forecasting models. In [26], it was proposed to improve the accuracy of a STLF by combining it with a feature selection method based on the generalized minimum redundancy and maximum relevance. Another paper proposes using RF models for short-term electric load forecasting making use of an ensemble learning method that generates many regression trees and aggregates their results [27].

Artificial Neural Networks

Neural Networks (ANN) are a class of models within the general machine learning literature and also a specific set of algorithms that have revolutionized machine learning. The operating principle of ANNs is based on the functions of the human brain neurons and provide powerful tools for modelling, especially when the underlying data relationship is unknown. Moreover, they can identify and learn correlated patterns between input data sets and corresponding target values [13], [17], [28]. They have been successfully applied in a variety of scientific fields such as mathematics, engineering, medicine, economics, meteorology, psychology, neurology and many other.

The reason that they have been successfully applied in so many scientific fields lies in the fact that they operate in accordance with the four operating principles:



- ◆ The fairly large database that is required, i.e. known inputs should be compared with their corresponding outputs in order to “educate” the network.
- ◆ The comparison of the output value that is produced with the real one and the amendment of the weights in accordance with the “education rule”.
- ◆ The calculated error which decreases with the increasing repetition number, works as a guide. The network is considered trained when the error becomes smaller than the threshold determined in the beginning of the process.
- ◆ The certification that the system is adequately trained when it responds correctly to new samples. The broad spectrum of the learning set is considered a criterion [13], [17], [28].

However, they also present limitations, such as:

- ◆ The fact that they are computationally intensive and therefore hardware dependent as they require processors with parallel processing power, depending to their structure.
- ◆ The completely obscure “black-box” treatment of the system being modeled: when an ANN produces a solution, it does not provide any insight about the why and how the result was obtained and this reduces its trustworthiness. Also, no specific rules apply for the determination of their network structure. Any appropriate network structure is achieved through experience, trial and error.
- ◆ ANNs work with numerical information. This means that all the problems have to be translated into numerical values before they are being introduced to ANN.
- ◆ The fact that the network’s performance is evaluated against a specific preset error value, which is the criterion for ending the training. However, this error value does not necessarily yield the optimum results.

The forecasting models that have been implemented all these years on ANNs are dozens. For example, the model proposed in [29] for STLF in microgrids is based on a three-stage architecture which starts with pattern recognition by a Self-Organizing Map (SOM), a clustering of the previous partition via a k-means algorithm, and finally demand forecasting for each cluster with a MultiLayer Perceptron Model, which is a class of feed-forward ANN, validation was performed with data from

a microgrid-sized environment provided by the Spanish company Iberdrola. In [30], a Self-Recurrent Wavelet Neural Network (SRWNN) is used as a forecast engine in a microgrid using the Levenberg-Marquardt (LM) learning algorithm to train the SRWNN. In order to demonstrate the efficiency of the proposed method, it was tested against actual hourly load data of an educational building supplied by a microgrid.

5. CONCLUSION

We have attempted to give an overview of the most often used machine learning algorithms such as Support machine vectors (SVM), k-Nearest Neighbours (kNN), Random Forest and Artificial Neural Networks (ANN) that have proved to be very efficient for load forecasting at the microgrid level. Machine learning algorithms are often used for short-term forecast or/and are combined with other algorithms creating novel forecasting models for both at the microgrid and metropolitan level.

REFERENCES

- [1] T. V. Ark, "Ask About AI - The Future of Work and Learning," Getting Smart Staff, 2017.
- [2] "Special report: How to implement AI and machine learning," TechRepublic, 2016.
- [3] Z. Mohamed and P. Bodger, "Forecasting electricity consumption in New Zealand using economic and demographic variables," *Energy*, vol. 30, p. 1833–1843, 2005.
- [4] M. Yang and X. Yu, "China’s rural electricity market-a quantitative analysis," *Energy*, vol. 29, p. 961–977, 2004.
- [5] E. Mele, A. Ktena and C. Elias, "Electricity use profiling and forecasting at microgrid level," in *RTU-CON 2018*, Riga, Latvia, 2018.
- [6] T. M. Mitchell, *Machine Learning*, McGraw-Hill Education, 1997.
- [7] P. Jędrzejowicz, "Machine Learning and Agents," in *5th KES International Conference, KES-AMSTA 2011, pp 2-15*, Manchester, 2011.
- [8] T. Haida and S. Muto, "Regression based peak load forecasting using a transformation technique," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1788-1794, Nov. 1994.
- [9] S. Mirasgedis, Y. Safaridis, E. Georgopoulou, D. P. Lalas, M. Moschovits, F. Karagiannis and D. Papakonstantinou, "Models for mid-term electricity demand forecasting incorporating weather influences," *Energy*, vol. 31, pp. 208-227, 2006.



- [10] H. L. Willis and J. E. D. Northcote-green, "Comparison tests of fourteen distribution load forecasting methods," *IEEE Trans. Power App. Syst.*, vol. 103, pp. 1190-1197, 1984.
- [11] H. L. Willis, R. W. Powell, and D. L. Wall, "Load transfer coupling regression curve fitting for distribution load forecasting," *IEEE Trans Power App. Syst.*, vol. 103, pp. 1070-1076, 1984.
- [12] E. Doveh, P. Feigin, D. Greig and L. Hyams, "Experience with FNN models for medium term power demand predictions," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 538-546, May 2002.
- [13] G. J. Tsekouras, N. D. Hatziaargyriou and E. N. Diailynas, "An optimized adaptive neural network for annual midterm energy forecasting," *IEEE Trans. Power Syst.*, vol. 21, no. 1, pp. 385-391, Feb. 2006.
- [14] M. Y. Chow, J. Zhu and H. Tram, "Application of fuzzy multi-objective decision making in spatial load forecasting," *IEEE Trans. Power Syst.*, vol. 13, no. 3, pp. 1185-1190, Aug. 1998.
- [15] C. N. Elias and N. D. Hatziaargyriou, "An Annual Midterm Energy Forecasting Model Using Fuzzy Logic," *IEEE Transactions On Power Systems*, vol. 24, no. 1, pp. 469 - 478, Feb. 2009.
- [16] G. J. Chen, K. K. Li, T. S. Chung, H. B. Sun and G. Q. Tang, "Application of an innovative combined forecasting method in power system load forecasting," *Elect. Power Syst. Res.*, vol. 59, pp. 131-137, 2001.
- [17] J. Llanos, D. Sáez, R. Palma-Behnke, A. Núñez and G. Jiménez-Estévez, "Load Profile Generator and Load Forecasting for a Renewable Based Microgrid Using Self Organizing Maps and Neural Networks," in *WCCI 2012 IEEE World Congress on Computational Intelligence*, Brisbane, Australia, Jun. 2012.
- [18] Z. H. Bohari, H. Azemy, M. N. M. Nasir, M. F. Baharom, M. F. Sulaima and M. H. Jali, "Reliable Short Term Load Forecasting Using Self Organizing Map (SOM) In Deregulated Electricity Market," *Journal of Theoretical and Applied Information Technology*, vol. 79, no. 3, pp. 389-394, Sep. 2015.
- [19] O. A. S. Carpinteiro and A. J. R. Reis, "A Hierarchical Self-Organizing Map Model In Short-Term Load Forecasting," *Journal of Intelligent and Robotic Systems*, vol. 31, no. 1-3, pp. 105-113, May 2001.
- [20] J. Che, J. Wang and G. Wang, "An adaptive fuzzy combination model based on self-organizing map and support," *Energy*, vol. 37, no. 1, pp. 657-664, Jan. 2012.
- [21] V. Jakkula, Tutorial on Support Vector Machine (SVM), School of EECS, Washington State University, 2012.
- [22] M. Mohandes, "Support vector machines for short-term electrical load forecasting," *International Journal of Energy Research*, vol. 26, no. 4, pp. 335-345, Jan. 2002.
- [23] A. Jain and B. Satish, "Clustering based Short Term Load Forecasting using Support Vector Machines," in *2009 IEEE Bucharest PowerTech*, Bucharest, Romania, Jul. 2009.
- [24] W. Li, D. Kong and J. Wu, "A Novel Hybrid Model Based on Extreme Learning Machine, k-Nearest Neighbor Regression and Wavelet Denoising Applied to Short-Term Electric Load Forecasting," *Energies 2017*, vol. 10, no. 5, 2017.
- [25] R. Zhang, Y. Xu, Z. Y. Dong, W. Kong and K. P. Wong, "A composite k-nearest neighbor model for day-ahead load forecasting with limited temperature forecasts," *2016 IEEE Power and Energy Society General Meeting (PESGM)*, Jul. 2016.
- [26] N. Huang, Z. Hu, G. Cai and D. Yang, "Short Term Electrical Load Forecasting Using Mutual Information Based Feature Selection with Generalized Minimum-Redundancy and Maximum-Relevance Criteria," *Entropy 2016*, vol. 18, no. 9, 2016.
- [27] G. Dudek, "Short-Term Load Forecasting using Random Forests, Department of Electrical," in *Intelligent Systems'2014: Proceedings of the 7th IEEE International Conference Intelligent Systems IS'2014, Volume 2: Tools, Architectures, Systems, Applications*, Warsaw, Poland, Sep. 2014, pp. 821-828.
- [28] S. A. Kalogirou, "Artificial neural networks in energy applications in buildings," *International Journal of Low Carbon Technologies*, vol. 1, no. 3, pp. 201-216, Jul. 2006.
- [29] L. Hernández, C. Baladrón, J. M. Aguiar, B. Carro, A. Sánchez-Esguevillas and J. Lloret, "Artificial Neural Networks for Short-Term Load Forecasting in Microgrids Environment," *Energy*, vol. 75, pp. 252-264, Oct. 2014.
- [30] H. Chitsaz, H. Shaker, H. Zareipour, D. Wood and N. Amjady, "Short-term Electricity Load Forecasting of Buildings in Microgrids," *Energy and Buildings*, vol. 99, pp. 50-60, Jul. 2015.