Deep Learning Based Consumer Classification for Smart Grid

Kálmán Tornai[1(✉)], András Oláh[1], Rajmund Drenyovszki[2], Lóránt Kovács[2], István Pintér[2], and János Levendovszky[3]

[1] Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary
tornai.kalman@itk.ppke.hu
[2] GAMF Faculty of Engineering and Computer Science, Pallas Athene University, Kecskemet, Hungary
[3] Department of Networked Systems and Services, Budapest University of Technology and Economics, Budapest, Hungary

Abstract. Classification of different power consumers is a very important task in smart power transmission grids as the different type of consumers may be treated with different conditions. Furthermore, the power suppliers can use the category information of consumers to forecast better their behavior which is a relevant task for load balancing.

In this paper, we present performance results on the classification of consumers using deep learning based classification scheme in smart grid systems. The results are compared with existing classification methods using real, measured power consumption data.

We demonstrate that consumer classification performed by neural networks can outperform existing, traditional tools as in several cases the correct class assignment rate is greater than 0.97.

Keywords: Classification methods · Consumer classification · Deep learning · Softmax layer network

1 Introduction

Smart power transmission grids efficiently integrate renewable energy sources and can manage the balance between the supply and demand adaptively. The new capabilities inhere in the integrated, intelligent measurement system. In addition, with two-way communication not only the measurements can be collected but the endpoints also can be controlled.

The use of smart meters in the network implies that a vast amount of data is being acquired. These data have to be processed in order to obtain relevant information about the status of the network or the consumer behavior.

The classification of consumers is the basic tool to recognize category changes, consumption behavior changes, or irregularities of the grid. This information can be used (i) for using different pricing for consumers with different behavioral

E.T. Lau et al. (Eds.): SmartGIFT 2017, LNICST 203, pp. 132–141, 2017.
DOI: 10.1007/978-3-319-61813-5_13

patterns [13]; (ii) for estimating the future consumption better; (iii) in load balancing as the consumers belonging to different classes can be estimated well; (iv) to extract further relevant information about the behavior of different consumer categories.

In this paper, we introduce results on using Deep Learning techniques of recognition of consumer class. The performance of the deep learning based solution is compared with other, existing solutions on real consumption measurements. It will be shown that the proposed method outperforms the existing implementations as it can provide more accurate results compared to other methods. Furthermore, the proposed method is more flexible in the sense that it can handle different scales of the input data as well.

The paper will be organized as follows: in Sect. 2, the existing classification methods will be briefly reviewed. In Sect. 3, the new method will be discussed. In Sect. 4, the performance of the proposed method will be described, finally in Sect. 5, conclusions will be drawn.

2 Related Works

In the followings, based on the review of Zhou et al. [8], the existing classification approaches are summarized. The following reviewed methods are the most commonly used solutions for classifying or clustering power consumption data. The performance of following methods will be compared with the proposed, deep learning based method in Sect. 4.

2.1 K-Nearest Neighbor

Using the k-nearest neighbor (kNN) method [21], the decision is made based on the distance between data elements. For a new object, the distances between the existing objects and new instance have to be calculated. The class assignment problem is solved by seeking the class with smallest (average) distance. The performance of the method is highly influenced by the applied metric [19]. In most cases, the Euclidean distance is used, but special problems require more complicated metrics such as dynamic time warping distance [3,11]. This method is applicable to sequences with non-numerical values, but the data has to be transformed or sequence alignment has to performed. In the case of power consumption classification this method can be applied as it is simple and efficient enough, however it is sensitive to noise and the performance depends on the applied metric highly.

Fuzzy k-NN classification is similar to k-NN classification, but objects have a membership degree for classes. It has the advantage that the algorithm does not assign the object to a class. For example, in a case where a new object lies between two classes this fact is reported by the algorithm and creation of new class or other solutions can be considered [20]. This method has better average performance, however, it can easily happen that a new consumer cannot be assigned to the correct class for sure as the fuzzy membership is not significant enough.

2.2 Artificial Neural Networks

The Artificial Neural Networks (ANN), especially recurrent neural networks (RNN) can model the properties of the time series of classes [15]. These models are constructed for all classes using training sets. With Multilayer Perceptron network (MLP) the data set classification task is solved by using hyperplanes for set separation [6].

Kernel methods such as Support Vector Machines (SVMs) can be used to reduce the number and to extract the features. The kernel-based methods are commonly used to process biological data [10].

Different types of ANNs have been successfully deployed to solve the of power consumer classification [9]. However, there are also certain disadvantages of applying such structures as the performance is affected by weights of network connections the initialization parameters, the order of training samples. Further optimization is possible and new structures can also be deployed, such as introduced in this work.

2.3 Hidden Markov Model Based Classification

The model-based classification methods construct a model for all classes using training sets. The incoming data is classified upon the best fitting model [14]. One of the most popular statistical models is Hidden Markov Model [7]. A trained HMM can reflect the probabilistic relations of the values within the time series thus an HMM represents the structure of the time series. The optimal parameters of an HMM can be found by using the Baum-Welch algorithm.

2.4 Forecast Based Classification

Our previously introduced solution is the forecast based classification method (FBM). This method [18] exploits the different statistical properties of the power consumption time series. For each consumer class, a feed-forward neural network (FFNN) [5] is trained to forecast the time series of the specific class with low error rate. Hence the trained forecaster is able to approximate time series of the same class with low error rate and has a superior error rate in case of time series belonging to different consumer classes.

Newly arriving consumers are classified as follows: each of previously trained forecaster is evaluated on the new time series and the forecast error is calculated as the difference between the forecast and real value of time series. The mean of the forecast error will be used as a decision variable to decide the class where the sequence belongs to.

The advantage of this method is that it utilizes well the temporal property of time series, furthermore short term forecasting is also performed as part of the algorithms. However, this method is highly influenced by the capabilities of the applied forecast method, which affect the performance of the solution. (As the selection and training of the best forecast method is not trivial.)

3 Consumer Classification Using Deep Learning Techniques

Deep learning methods and neural networks are hot topics as the amount of recorded (measured) data increase data day by day. However, the convolutional neural networks (CNN) were invented and first demonstrated by Fukushima Kunihiko in 1980 [4], but they were not applied in practice until 2011 when Yan LeCun applied this method successfully in many different problems [17]. It has been demonstrated that in certain problems they are able to reach a classification accuracy that is comparable to the human performance, or in some cases it can surpass it. In most cases, a series of CNNs are used and a softmax layer is added to determine the resulting class.

Autoencoders (or Diabolo network) [16] with softmax layers are also used for classification. The aim of autoencoders is to learn a representation (model or encoding) for a set of data while the dimensionality of the data is reduced. In this paper, we applied autoencoders with softmax layer to solve the power consumption data classification task.

3.1 Autoencoder Network

The structure of an autoencoder network is a feedforward (without any feedback) neural network (FFNN) with the following properties:

- The input and output layer has the same number of artificial neurons.
- The network has one or more hidden layers between the input layer and output layers.

The network is used for unsupervised learning as the purpose of the output layer is to reconstruct the input data, while the inner layers try to reduce the dimensionality. The structure is demonstrated by Fig. 1.

Autoencoders consists of two parts: (i) encoder and (ii) decoder, which can be described by the following transformations $\Psi : \mathcal{T} \rightarrow \mathcal{C}$ and $\Phi : \mathcal{C} \rightarrow \mathcal{T}$. Theses transformation should be defined such that

$$\arg\min_{\Phi,\Psi} \left\| \mathbf{T} - (\Psi \circ \Phi)\, \mathbf{T} \right\|^2 . \tag{1}$$

Thus the reconstruction error is minimized. The dimensionality of $\mathcal{C}$ should be less than the dimensionality of $\mathcal{T}$ in order to extract features from the original data.

Int the neural networks sigmoid activation function

$$\varphi_S(u) = \frac{1}{1 + e^{-\alpha u}}, \tag{2}$$

or rectified linear activation functions are used:

$$\varphi_R(u) = \begin{cases} u, \text{if } u > 0 \\ 0, \text{if } u \leq 0 \end{cases} \tag{3}$$

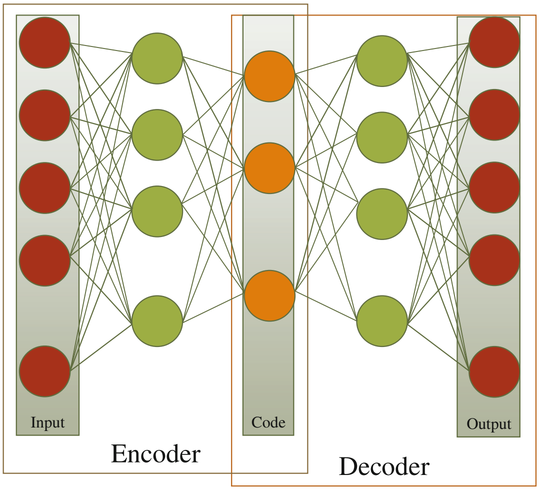

Fig. 1. Structure of autoencoder network

The training of the network can be performed with any of backpropagation algorithms used in feedforward neural networks. In our test, we have used Scaled conjugate gradient backpropagation and Levenberg–Marquardt backpropagation [2] as the first one can be used efficiently on GP-GPU implementation and the execution time of the latter one is often the shortest among of several methods in the case of CPU based simulation.

3.2 Soft Max Layer

This layer consists of artificial neural networks having the soft max function as activation function (or normalized exponential):

$$P\left(\mathcal{C}_r|u\right) = \frac{e^{a_r}}{\sum_{j=1}^{k} e^{a_j}}, \tag{4}$$

where $0 \leq P\left(\mathcal{C}_r|u\right) \leq 1$ and

$$\sum_{j=1}^{k} P\left(\mathcal{C}_j|u\right) = 1.$$

In previous equation $P\left(u|\mathcal{C}_j\right)$ denotes the conditional probability of the sample given class r and $P\left(\mathcal{C}_j\right)$ is the prior probability of the class, and

$$a_r = \ln\left(P\left(u|\mathcal{C}_j\right) P\left(\mathcal{C}_r\right)\right).$$

The softmax function can be considered the multi-class generalization of logistic sigmoid function.

3.3 Proposed Structure

The properties and parameters of the architecture of the neural networks proposed in this paper to solve the power consumer classification problem are overviewed in this section.

Two autoencoders have been applied sequentially in order to reduce the dimensionality of the input time series highly. The number of outputs of the encoder has been determined as the proportion to the original number of inputs. We have investigated the performance of the network using the different ratio of compression.

The encoder of autoencoders has been implemented with rectified linear activation functions and in the output layer, linear activation function was used.

4 Performance Analysis

In this section, the test environment, test data and the performance of the proposed method is introduced and compared to existing solutions.

4.1 Test Environment

The tests of the algorithms have been carried out in Matlab environment [12]. All test has been repeated for several times to have averaged results. Thus the extremely good and bad results are eliminated. The outlier performance values are caused by randomly chosen learning parameters of artificial neural network based algorithms. The available data was randomly split into training data set, test data set and for training purposes validation dataset, with the following ratios of 0.45, 0.45 and 0.1 respectively.

The performance of the proposed method can be evaluated by comparing the results of the method with the known information provided in the database we used for evaluation. (The class assignments made by the database providers are considered as correct solutions.) Hence the performance metrics is the number of correctly classified time series divided by the total number of time series. All performance results are the evaluation of the algorithms on the same test set.

4.2 Real, Measured Consumption Data

Real measurements were obtained from two different sources. First of them is a rather small database. The database was obtained from a large Central-European electricity distribution company, where the power consumption time series was classified into 8 classes by company experts. The consumption was measured at 150 different sites for one year in 2009. Furthermore, as the actual data is trade secret it has been normalized by the company and personal information was removed as well. While the class assignment provided by the power distribution company is unambiguous, the actual classes are overlapping thus automated classification cannot achieve 100% precision.

The second database is "Commercial and Residential Hourly Load Profiles for all TMY3 Locations in the United States" [1]. This dataset contains hourly load profile data for 16 different commercial building types (based on the DOE commercial reference building models) and residential buildings (based off the Building America House Simulation Protocols). The hourly load profiles are available for overall TMY3 locations in the United States. Due to the size of the database, we have reduced the number of classes from 16 to 8 (Hospital, Large office, Medium office, Primary school, Secondary school, Small office, Warehouse, Supermarket) and then the number of sites reduced from 936 to 100 randomly chosen sites.

4.3 Performance Results

We have investigated the influence of actual coding rates of the two autoencoders to the performance of the correct classification capabilities. Several test results with different parameters show that maximal performance can be achieved in the case when the number of nodes of the first encoder's code layer is higher than 90% of the number of the nodes of the input layer. In the case of the second layer the number of nodes has less impact on the performance, however, the best results have been achieved at having the number of code nodes 35% of the input nodes. The results are summarized by Fig. 2. In the following performance test the previous parameters have been used.

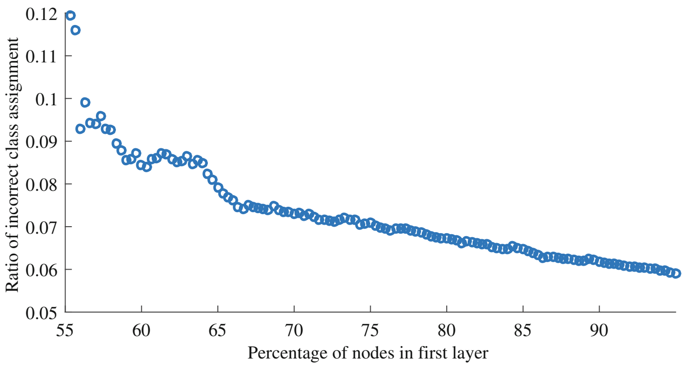

Fig. 2. Performance results in function of the number of nodes in autoencoder network

Results on European Database. As the actual classes are overlapping in some cases we have investigated the performance of the method in four disjoint classes and all classes separately. Figure 3 indicates that the Deep Learning based solution outperforms the other methods in both cases. The correct class assignment ratio is over than 98% in the first case and 95% in the latter case. The classes which are mostly confused by the solutions are investigated using several statistical parameters and the principal components are also compared. It has

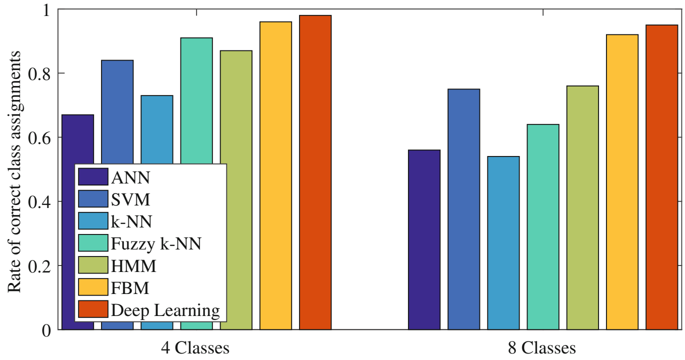

Fig. 3. Performance results of different classification methods executed on the European measurements

been found that the three of the eight classes has several overlapping parameters, as the classification may be acceptable even the class assignment made by algorithms does not match to the manual assignment made by experts. (However in the figure the performance results show the case when we do not take into consideration the overlaps, and if the algorithmic result does not match to the manual assignment the class assignment is considered as a wrong solution.)

Results on US Database. The comparison of the performance of deep learning method to other algorithms is summarized by Fig. 4. The recorded data came from all across the US thus the geographic location may have an influence on the behavior of consumers, as a result, the load profiles may vary between different climate zones. As a result, we have investigated the performance of the

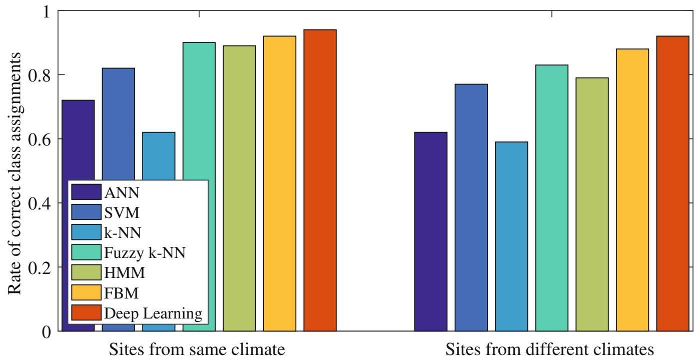

Fig. 4. Performance results of different classification methods executed on the American measurements

algorithms using data sets from the same climate zone (humid subtropical) and using data sets from different climate zones as well.

Results indicate that both in the case of same and different climate zones the deep learning based solution outperforms the other solutions. However, the difference is smaller between the algorithms compared to the European database. It may be explained by the normalization of the European data. Furthermore, the geographic location has only a small influence on the performance of the algorithms.

The algorithm made incorrect class assignments in case of the following pairs of classes: Secondary Schools and Primary Schools; Warehouses and Supermarkets. As the members of these pairs of classes can be considered similar, the number of incorrect class assignments can be acceptable.

5 Conclusion

In this paper, we have introduced performance results on real electricity consumer data of deep learning based power consumption classification scheme. The proposed scheme is capable of classifying the power consumers at very high level in most of the cases.

In addition, the results of the scheme have been compared to existing classification methods, and it has been proven experimentally that the performance of the deep learning method is better than the other available solutions. A further advantage of the solution that is can be efficiently used on data which are previously normalized, and the novel method is more tolerant to deviations of the data set, such as it is capable of recognizing classes which are spanning over different climate zones.

In the future, we are going to investigate whether this scheme can be modified to cluster or rather to automatically categorize the power consumers and it also is going to be tested on additional databases.

Acknowledgment. This publication/research has been supported by PPKE KAP 16-71009-1.2-ITK Grant. This source of support is gratefully acknowledged.

References

1. Commercial and residential hourly load profiles for all TMY3 locations in the united states. http://en.openei.org/datasets/dataset/commercial-and-residential-hourly-load-profiles-for-all-tmy3/locations-in-the-united-states. Accessed 30 Aug 2016
2. Bishop, C.: Pattern Recognition and Machine Learning. Springer, New York (2006)
3. Chan, S., Kao, B., Yip, C., Tang, M.: A brief survey on sequence classification. In: Proceedings Eighth International Conference on Database Systems for Advanced Applications (DASFAA 2003), pp. 119–124. IEEE (2003)
4. Fukushima, K.: Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. **36**(4), 193–202 (1980)

5. Haykin, S.: Neural Networks, A Comprehensive Foundation, 3rd edn. Pearson, Prentince Hall, Upper Saddle River (2008)
6. Kaefer, F., Heilman, C.M., Ramenofsky, S.D.: A neural network application to consumer classification to improve the timing of direct marketing activities. Comput. Oper. Res. **32**(10), 2595–2615 (2005)
7. Kim, M., Pavlovic, V.: Sequence classification via large margin Hidden Markov Models. Data Min. Knowl. Disc. **23**(2), 322–344 (2011)
8. le Zhou, K., Yang, S.L., Shen, C.: A review of electric load classification in smart grid environment. Renew. Sustain. Energy Rev. **24**, 103–110 (2013)
9. Lo, K.L., Zakaria, Z.: Electricity consumer classification using artificial intelligence. In: 39th International Universities Power Engineering Conference, UPEC 2004, vol. 1, pp. 443-447, September 2004
10. Lodhi, H., Saunders, C., Shawe-Taylor, J., Cristianini, N., Watkins, C.: Text classification using string kernels. J. Mach. Learn. Res. **2**, 419–444 (2002)
11. Maimon, O., Rokach, L. (eds.): Data Mining and Knowledge Discovery Handbook. Springer, New York (2005)
12. MATLAB, version 8.6.0 (r2015b). The MathWorks Inc., Natick, Massachusetts (2015)
13. Noble, P.M., Gruca, T.S.: Industrial pricing: theory and managerial practice. Mark. Sci. **18**(3), 435–454 (1999)
14. Rani, S., Sikka, G.: Recent techniques of clustering of time series data: a survey. Int. J. Comput. Appl. **52**(15), 1–9 (2012)
15. Ruck, D., Rogers, S., Kabrisky, K., Oxley, M., Suter, B.: The multilayer perceptron as an approximation to an optimal Bayes estimator. IEEE Trans. Neural Netw. **1**(4), 296–298 (1990)
16. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by backpropagating errors. Nature **323**, 533–536 (1986)
17. Sermanet, P., LeCun, Y.: Traffic sign recognition with multi-scale convolutional networks. In: Proceedings of International Joint Conference on Neural Networks (IJCNN 2011) (2011)
18. Tornai, K., Kovács, L., Oláh, A., Drenyovszki, R., Pintér, I., Tisza, D., Levendovszky, J.: Classification for consumption data in smart grid based on forecasting time series. Electr. Power Syst. Res. **141**, 191–201 (2016)
19. Xing, Z., Pei, J., Keogh, E.: A brief survey on sequence classification. ACM SIGKDD Explor. Newslett. **12**(1), 40–48 (2010)
20. Zakaria, Z., Lo, K.L., Sohod, M.H.: Application of fuzzy clustering to determine electricity consumers' load profiles. In: IEEE International Power and Energy Conference, pp. 99–103, November 2006
21. Zhang, B., Srihari, S.N.: Fast k-nearest neighbor classification using cluster-based trees. IEEE Trans. Pattern Anal. Mach. Intell. **26**(4), 525–528 (2004)