On the Value of Distribution Network Topology Information in the Identification of End-user Phase Consumption: A Graph Neural Network Approach

Terezija Matijašević ^{1*}, Tomislav Antić¹, Tomislav Capuder¹ ¹ University of Zagreb Faculty of Electrical Engineering and Computing, Department of Energy and Power Systems Zagreb, Croatia (* Corresponding Author)

ABSTRACT

End-users are transiting towards more active, integrating new low-carbon (LC) technologies and bringing unpredictability to low-voltage (LV) distribution networks. Although smart meters have a great potential in increasing the observability, they are mostly being employed only for billing purposes, leaving many other possibilities unexploited, further complicating the many analyses required for effective operational planning and real-time (RT) operation. Detection of phase consumption of end-users is significantly difficult, due to the nonlinear relationships between obtained phase voltage measurements and aggregated end-user consumption. Machine learning (ML) is increasingly used for these and similar problems, and therefore, in this paper, a neural network (NN) - based model is developed to detect end-user consumption in an LV distribution network available from voltage measurements and aggregated end-user consumption. Furthermore, the influence of topology on the output values of the model is investigated and a graph neural network (GNN) - based model is created that considers both the structure and data of the distribution network elements. Both models are tested on the real-world LV distribution network with more than 150 end-users. The results showed the effectiveness of both models in determining the distribution of end-user consumption, with the GNN-based model showing significantly better results. Such a model can help the energy utilities to overcome this time-consuming problem and lay a good foundation for further analyzes required to enable operation and planning of distribution networks.

Keywords: smart meters, distribution network, phase consumption, machine learning, graph neural network

1. INTRODUCTION

With the integration of low carbon (LC) technologies and an increasing number of active endusers, especially prosumers, distribution systems are becoming more complex in terms of planning and operation. In recent years, the application of machine learning (ML) algorithms has significantly increased due to their potential in resolving the problems related to the number and complexity of the collected data.

The authors in [1] consider different ML models for the forecast of electricity prices and additionally extend the current state-of-the-art by considering previously unused predictive features. Electricity load was forecasted by using the recurrent extreme learning machine model in [2]. ML algorithms are also used for the detection of unwanted events in a network, e.g., detection of faults [3] or power quality disturbances [4]. The other potential of ML algorithms is used in the detection of energy thefts in smart distribution networks using end-users' consumption patterns [5]. As mentioned before, smart distribution networks require the processing of a high number of complex data. Therefore, it is important to accurately detect the needed set of features used in different ML-based predictions [6] but also, to detect false data in smart grids so the digital communication and other important aspects of the smart distribution networks planning, and operation would not be compromised [7].

In this paper, we use ML methods for the distribution of three-phase connected end-users' consumption among the phases with the application of real-world data from smart meters installed with customers of the Croatian distribution network. Even though phase detection is a well-known problem, it was mostly solved only for the single-phase users at the substation level or is based on physical devices such as PLC [8] or Phasor Measurement Units [9].

The contributions of this paper are:

- An NN-based model for the distribution of end-user phase consumption.
- An extension of the previously developed model by integrating the distribution network topology information.

The proposed tool is of great significance to energy utilities as any further analysis of the distribution network depends on the correct identification of the end-user phase connection and consumption among phases. The rest of the paper is organized as follows: Section II presents the methodology of the proposed approach. Section III contains results for both of the proposed models, and the conclusions are drawn in Section IV.

2. PROPOSED METHODOLOGY

A significant number of smart meters installed in the active, smart distribution networks have great potential to increase observability, but most smart meters are still used mostly for billing purposes, and for this reason, they only collect total end-user consumption. In some cases, smart meters are modified so they can collect processing data, such as voltage magnitude. Such aggregated consumption prevents further analyses needed for the planning and RT operation of unbalanced distribution networks.

Even though most end-users' devices are singlephase, end-users themselves can be single-phase or three-phase connected to a network, which causes a problem in the distribution of aggregated consumption among the phases. The advantages of NN overcome the complexity caused by the nonlinear relationships between phase consumption and voltage at each phase, and therefore, in this paper, two NN-based models for determining the distribution of end-user's consumption by individual phase are developed. The first model implements the standard NN configuration, which consists of the input layer, hidden layer, and output layer. The hidden layer consists of a fully connected layer (FCL) that is deeply connected with its preceding layer, i.e. the neurons of its layer are connected to every neuron of its preceding layer.

The input dataset is a dataset U, which contains the values of voltage magnitudes, while the output dataset is set P with the values of the active power of each enduser. The influence of reactive power and voltage angle is out of the scope of this paper, which leaves room for the implementation of the proposed model in Q-U regulation and integration of LC units connected by converters.

Input and output datasets are obtained as a result of time simulations of power flows and represent the state of the observed LV network over a period of time. Therefore, the dimensions are given by *TxNx3*, where *T* stands for the total number of time steps in the considered time period, *N* accounts for the total number of observed end-users and 3 represents the number of phases in an LV distribution network.

The second model is an extension of the first model, which, in addition to the initial input dataset, applies the admittance matrix of the observed LV distribution network, which contains information about the network elements and allows the display of the network in graph form. A graph can be written as G = (V, E), where V represents a set of vertices/nodes, and E is a set of edges/lines. In graph G, v_i is the i^{th} node and e_{ij} is the edge from the i^{th} node to the j^{th} node.

Therefore, to enable the processing of network data, it is necessary to integrate additional layers that are suitable for working with data in the form of graphs. GNNs are the best solution for the implementation of the graph data, and therefore the second model is based on GNNs (*Fig. 1*). Unlike the first model, this one uses graph convolutional layers along with FCLs to establish the relationship between input and output datasets. Each layer *I* in GNNs can be denoted as

$$H^{(l+1)} = f(H^{(l)}, A) = \sigma(AH^{(l)}W^{(l)})$$
(1)

where *H* represents node features, *A* is an adjacency matrix that represents the graph connection $(A_{ij} = [0, 1])$, *W* is a weight matrix for the *I*th layer, and σ is an activation function. Since this equation does not include the feature vector of the observed node (i.e., it includes features for all of a given node's neighbors), we extended layer representation proposed in [10], by adding an identity matrix to the adjacency matrix.

Besides the P and U datasets, the second model applies the dataset M, which corresponds to an admittance matrix with dimensions $3K \times 3K$, where K is the total number of nodes in the LV distribution network. This admittance matrix can be classified as a weighted adjacency matrix which includes continuous values corresponding to the magnitudes of complex elements of the admittance matrix. Therefore, dataset M cannot be used as a direct input to the GNN model and therefore it is necessary to create a Laplacian matrix L as:

$$L = I - D^{-\frac{1}{2}}MD^{-\frac{1}{2}}$$
 (2)

where I is the identity matrix and D is the degree matrix.



Fig. 1. NN design for the second model

3. RESULTS

The models described in Section 2 are applied to a real-world LV distribution network of 151 residential consumers and a total of 245 nodes. Datasets **P** and **U** are generated using the *pandapower* power flow simulation tool. Since there are several nodes in the network that do not contain information about voltage measurements (e.g., connectors, network cabinets), it is necessary to ensure that the model does not generate predictions for them.

For the proposed models to be applicable to distribution networks containing both three-phase and single-phase connected end-users, an LV distribution network containing only single-phase connected residential consumers is applied for the training of both models, while for testing a model, a network containing 66 three-phase end-users and 85 single-phase end-users, connected to phases A, B, and C is utilized.

Due to the focus on the application of the proposed models on both types of networks, it is necessary to add a constraint to the loss function that will ensure that the obtained phase consumption is equal to the aggregated consumption of each end-user.

3.1 NN model

The datasets for the first model consist of 1342 different time steps, of which 1100 time steps make up the training set and the rest is contained in the validation set. The best hyperparameters of the model are determined by validation and are presented in *Table 1*.

Table 1: NN	hyperparameters
-------------	-----------------

neurons (hidden layer)	8
act.function (hidden layer)	tanh
act.function (output layer)	linear
batch size	100
optimiser	RMSprop
learning rate	0.0001
epochs	1500

The performance of the model is assessed using the test dataset collected from the LV distribution network containing both three-phase and single-phase residential consumers, collected in 48 different time steps. There are 66 three-phase connected end-users and 85 end-users connected to phases A, B, and C.

Furthermore, the validation loss of this model is 0.014, while the mean squared errors (MSEs) on the test set are 0.025030 (for a larger test set with 1342 time steps) and 0.029671 (for a smaller test with 48 time steps). The distribution of end-user consumption per phase determined by this model for a part of the larger test set is presented in *Fig. 2.*





Fig. 2. Distribution of end-user consumption, NN model

3.2 Graph neural network

The datasets for the second model consist of 672 different time steps, of which 500 time steps make up the training set and the rest is contained in the validation set. The best hyperparameters are determined by validation and are presented in *Table 2*.

Table 2: GNN	hyperparameters
--------------	-----------------

	1
neurons GConv1, GConv2	3
act.function GConv1, GConv2	tanh
neurons (FCL)	3
act.function (FCL)	tanh
act.function	linear
(output layer)	
batch size	50
optimiser	RMSprop
learning rate	0.0001
epochs	500

The performance of the model is also tested using an LV distribution network that includes both threephase and single-phase connected end-users, whose phase distribution is described in Subsection 3.1. Moreover, the validation loss of the GNN model is 0.0059. Due to the complexity and time-consuming nature of the model, the test dataset is limited in size and cannot reach the same number of time steps as the test set in the first NN model. For this reason, the test datasets contain one and 48 time steps with an MSE of 9.38513 x 10⁻¹⁹ and 0.00621, relatively. A small error that could be considered as insignificant shows that the developed model with the application of network topology successfully mitigates the problem of unavailable end-user phase consumption, which later leads to better planning and operation of the distribution network.

The distribution of the consumption of residential consumers per phase for the smaller test set is presented in *Fig. 3.*



Fig. 3. Distribution of end-user consumption, GNN model

4. CONCLUSION

The integration of smart meters potentially increases the network's observability but, in most cases, smart meters measure only limited set of values, important for the billing purposes. In this paper, two models that accurately determine phase consumption from aggregated end-user consumption are created. The first model is based on NNs and seeks to find a relationship between voltage measurements and enduser consumption measurements. The second model is upgraded compared to the first, and it includes the topology of the network itself in the form of a graph, which leads to the implementation of more complex GNN algorithms.

The proposed models are examined on a real-world LV distribution network with more than 150 residential consumers. Since the created models should enable the estimation of consumption in unbalanced distribution networks with three-phase and single-phase connected residential consumers, the models are tested for cases of a network with models of both single-phase and three-phase connected end-users.

The GNN model based on network topology shows significantly better results compared to the simple NN model. Thus, the GNN-based model has an MSE error of 9.38513 x 10⁻¹⁹ and 0.00621 for smaller and larger test sets, while the NN-based model results in an error of 0.029671 and 0.02503 for smaller and larger test sets. The MSE error value is almost negligible in both cases, but the GNN model is still the preferable model due to its employment of network topology. However, in cases where the network topology is unknown or there are errors in the network data, the NN model can be utilized as a great alternative.

Further development should go in the direction of implementing the model on distribution networks of different sizes which can consequently enable further analysis for energy utilities, including consumption forecasts, reservation of flexibility services, and other analyzes required for RT operation of smart distribution networks.

ACKNOWLEDGEMENT

This work has been supported in part by the European Structural and Investment Funds under KK.01.2.1.02.0042 DINGO (Distribution Grid Optimization) and in part by Croatian Science Foundation (HRZZ) and Croatian Distribution System Operator (HEP ODS) under the project IMAGINE – Innovative Modelling and Laboratory Tested Solutions for Next Generation of Distribution Networks (PAR-2018-12).

REFERENCE

- [1] L. Tschora, E. Pierre, M. Plantevit, and C. Robardet, "Electricity price forecasting on the day-ahead market using machine learning," *Appl. Energy*, vol. 313, p. 118752, May 2022, doi: 10.1016/J.APENERGY.2022.118752.
- [2] Ö. F. Ertugrul, "Forecasting electricity load by a novel recurrent extreme learning machines approach," *Int. J. Electr. Power Energy Syst.*, vol.

78, pp. 429–435, Jun. 2016, doi: 10.1016/J.IJEPES.2015.12.006.

- [3] M. S. Coutinho et al., "Machine learning-based system for fault detection on anchor rods of cable-stayed power transmission towers," *Electr. Power Syst. Res.*, vol. 194, p. 107106, May 2021, doi: 10.1016/J.EPSR.2021.107106.
- [4] H. Erişti, Ö. Yildirim, B. Erişti, and Y. Demir, "Automatic recognition system of underlying causes of power quality disturbances based on S-Transform and Extreme Learning Machine," Int. J. Electr. Power Energy Syst., vol. 61, pp. 553– 562, Oct. 2014, doi: 10.1016/J.IJEPES.2014.04.010.
- [5] S. K. Gunturi and D. Sarkar, "Ensemble machine learning models for the detection of energy theft," *Electr. Power Syst. Res.*, vol. 192, p. 106904, Mar. 2021, doi: 10.1016/J.EPSR.2020.106904.
- [6] S. Salcedo-Sanz, L. Cornejo-Bueno, L. Prieto, D. Paredes, and R. García-Herrera, "Feature selection in machine learning prediction systems for renewable energy applications," *Renew. Sustain. Energy Rev.*, vol. 90, pp. 728–741, Jul. 2018, doi: 10.1016/J.RSER.2018.04.008.
- [7] R. Nawaz, R. Akhtar, M. A. Shahid, I. M. Qureshi, and M. H. Mahmood, "Machine learning based false data injection in smart grid," *Int. J. Electr. Power Energy Syst.*, vol. 130, p. 106819, Sep. 2021, doi: 10.1016/J.IJEPES.2021.106819.
- [8] M. Lisowski, R. Masnicki, and J. Mindykowski, "PLC-Enabled Low Voltage Distribution Network Topology Monitoring," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6436–6448, Nov. 2019, doi: 10.1109/TSG.2019.2904681.
- [9] E. Dusabimana and S. G. Yoon, "A Survey on the Micro-Phasor Measurement Unit in Distribution Networks," *Electron. 2020, Vol. 9, Page 305*, vol. 9, no. 2, p. 305, Feb. 2020, doi: 10.3390/ELECTRONICS9020305.
- T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks," Sep. 2016, Accessed: Mar. 11, 2022.
 [Online]. Available: https://arxiv.org/abs/1609.02907