LOAD CURVE ESTIMATION FOR DISTRIBUTION SYSTEMS USING ANN

J. N. FIDALGO
Department of Electric Engineering and Computers
Faculty of Engineering of Porto University
Rua Dr. Roberto Frias, nº 378, 4200-465, Porto
PORTUGAL
jfidalgo@inescporto.pt  http://www.fe.up.pt

Abstract: - Loads estimation is becoming each time more fundamental for an efficient management and planning of electric distribution systems. Among the factors that contribute to this need of more efficiency are the increasing complexity of these networks, the deregulation process and the competition in an open energy market, and environment preservation requirements. However, the only information generally available at MV and LV levels is essentially of commercial nature, i.e., monthly energy consumption, hired power contracts and activity codes. In consequence, distribution utilities face the problem of estimating load diagrams to be used in planning and operation studies. The typical procedure uses measurements in typical classes of consumers defined by experts to construct inference engines that, most of the times, only estimate peak loads. In this paper, the definition of classes was performed by clustering the collected load diagrams. Artificial Neural Networks (ANN) were then used for load curve estimation. This article describes the adopted methodology and presents some representative results. Performance attained is discussed as well as a method to achieve confidence intervals of the main predicted diagrams.

Key-Words: - load estimation, clustering, artificial neural networks, distribution networks

1 Introduction

During last decade, one has witnessed not only the growth of transmission and distribution power networks but especially how they got more and more complex. Particularly, distribution systems have been continuously spreading and its complexity is also increasing not only because of new operation alternatives but also in consequence of the advent of deregulation. The competition in an open energy market also demands a higher energetic efficiency due to the will to explore the existing infra-structures to the limits and postpone investments as much as possible. At the same time, environment preservation and the need for energy efficiency have also become more demanding. In this framework, load curve estimation is becoming each time more fundamental to an efficient management and planning of electric distribution systems.

Consumptions’ estimation studies have been carried out by several investigators [1-3]. Some distribution utilities performed these studies, modeling consumers’ behavior for planning purposes, and using inference processes usually based on linear regression. The main purpose is frequently the estimation of peak power.

In the present project, the network planning needs are accessed by a Load Curve Management Module that estimates hourly load diagrams for LV and MV clients. The project included measurement campaigns, model development and testing. A short description of the measurement campaigns can be found in section 3.

The present approach includes three main procedures:

a) definition of consumers’ classes (clustering);
b) inference of load diagrams of MV/LV public stations or individual MV consumers;
c) estimation of error bars providing a image of hourly consumption’s variability.

This paper is organized as follows. Section 2 clarifies the goal of Load Curve Estimation. Section 3 describes the measurement campaigns. Sections 4, 5 and 6 present a detailed description on each of information processing phases a), b) and c). Finally, a few results are presented and commented.

2 Load curve estimation

One must stress that the present work should not be viewed as a traditional short-time load forecasting. Two sorts of arguments lead us to distinguish the two techniques: first, load forecasting is usually based on a time series analysis and, second, its goal consists generally of predicting loads in a short-time basis (next hours or, at maximum, a few days in advance).

In the proposed load curve estimation approach, ANN inputs consist mainly on commercial...
characteristics of consumers (hired power, activity codes, monthly energy consumption). On the other hand, the estimation period can be extended to several months or years, provided that some forecasts exists for monthly energy consumption. It is particularly adapted for planning of distribution network evolution. As an example, suppose that power distribution company aims at installing a new MV/LV station that will fed a given set of consumers; the question is: what is the load diagram expected for this station on a given year period?

3 Measurement campaigns

Initial modeling contributed to define the scope of the measurement campaigns. These campaigns were implemented taking into account the need for hourly base diagrams, considering two year periods, Winter and Summer, and also two groups, one for weekends and holidays and another for working days. Samples were taken accordingly in order to cover a large spectrum of possibilities.

EDP has carried out the measurement campaigns, collecting consumptions’ evolution data, in order to implicitly characterize consumers’ behavior. A large spectrum of possible load curves is quintessential to represent the whole universe of consumers. To accomplish this purpose, two measurement campaigns were carried out (one in the Summer and another in the Winter). Load curve recorders (LCR) were installed in a variety of consumers (238 MV clients and 267 LV clients) located in the neighborhoods of 5 different Portuguese cities. Observation areas include urban, semi-urban and rural types. The power of a consumer or group of consumers was registered every 15 minutes during a period of two or three weeks. The peak loads as well as the date and hour were also registered.

At the end of the observation period, the LCR transferred the information gathered directly to a PC. This information was completed with the information available in EDP databases, such as commercial characteristics (hired power, monthly energy consumption, peak loads, activity codes) and additional information concerning the network tree-structure. The basic idea consists of implementing a mechanism for load curve estimation of MV/LV public stations and MV individual consumers and pass this information to the network analysis tool. This device will aggregate consumptions estimates following the network tree-structure, and assessing load curves at secondary substation, primary substation and distribution centers.

4 Characterization and objectives

In this project, all phases were addressed in a innovative way, based on the use of neural networks (NN).

4.1 Clustering

The definition of classes was performed by clustering collected load diagrams, in order to avoid biasing introduced by preconceived ideas about the way consumers behave. Each clustering training pattern contains 24 elements - the registered power consumptions a each hour of a given day. This training set was presented to a Kohonen clustering tool, in order to obtain different load evolution classes. Results were compared with other classification tools.

4.2 Load Curve Estimation

When dealing with MV networks, the main issue is to get load diagrams in any point of the network, to be used later for planning purposes (Fig. 1). The available information from utility data base consists mainly of commercial data and energy consumptions. One intends to estimate load diagrams essentially for MV/LV public stations and MV individual clients. It was also decided to aggregate LV consumers, dependent from public stations, in order to evaluate their accumulated load diagram. This will avoid the need for the characterization of each LV individual consumer, reducing the size of the data base needed for future studies. Furthermore, there are no imperative knowledge requirements for a single LV client.

The data obtained from the measurements campaigns was divided following the season (Summer or Winter), the weekday (workday or weekend) and the type of consumer (LV or MV). Other available data of LV consumers are the monthly energy consumption, the activity code and hired power. All those consumers are fed by MV/LV public substations. For MV consumers, the accessible activity code, peak power, hired power and energy measures for different tariffs are known. These curves and values must be available in several points of the network, for instance in a MV/LV public station or in a feeder. Additional parameters are evaluated for each load curve (peak power use, load factor, loss factor, etc.), which help the characterization of single or aggregated consumer behavior.
4.3 Estimation of confidence intervals

Load consumption is always characterized by a considerable variability. For similar conditions (season of the year, workday or weekend, and so on), a given consumer or a set of consumers might present two quite different diagrams for two consecutive days. Fortunately, in general, one can observe some kind of behavior pattern, and load curves obtained for similar circumstances define a kind of fuzzy diagram.

On the present work, we propose to train an auxiliary ANN to learn load curves dispersion (error bounds distribution). These error bars will depend not only on the type and number of consumers aggregated in a given public station, but also on the hour of the day. This way, more complete information on load diagrams dispersion is obtained. Results in section 7.2 show this feature with some detail.

5 Models

MV consumers’ modeling needed two different types of analysis: MV clients and MV/LV public stations. In fact, a lot of public stations have no load measurement at all, and it is not possible to infer directly its hourly consumption. Moreover, as a result of the different behavior according to time and season of the year, the analysis was divided into Winter/Summer and week/weekend day cases. The establishment of the partial models for all the mentioned cases was followed by the development of a integration procedure to cope adequately with intermediate situations.

The modeling process and the subsequent data handling are applied to each referred cases.

5.1 Natural classes

One of the fundamental steps of this approach was the identification of natural classes from registered diagrams instead of defining a priori the classes. For that purpose, two different methods were tried: fuzzy clustering [4] and self-organized neural networks (Kohonen maps [5]).

In the LV case study, the best clustering performance was obtained when load diagrams were separated into six classes both with Kohonen and fuzzy clustering. Kohonen prototypes obtained are shown in Fig. 3. The fuzzy prototypes achieved are similar to Kohonen’s. This leads to an independent confirmation of the results obtained before.
5.2 Operational classes

Although valuable, the natural classes are not useful for operational purposes. In fact, they are defined directly from diagrams and not from consumers’ characteristics available in commercial data base. After obtaining the cluster’s prototypes one must induce the relation between classifications and commercial data (tariff class, hired power and monthly energy consumption), in order to generalize the classification of consumers for which only commercial data is known, (what constitutes the real future operational conditions).

Some experiences were carried out for determining a good combination of clustering (number of classes based on load evolution) and inference of classification rules (based on commercial data). This inference process was based on the observation of the distribution of classes’ members on the 3D space of commercial data (tariff class, hired power and monthly energy consumption).

The best classification performance was obtained with four classes (Figure 4) and the following classification rules:

- D Night consumers
- A Domestic consumers (Tc=0), low hired power (Pc≤6.6kW) and low energy consumption (E<600kWh)
- B Industrial consumers (Tc=4)
- C Other consumers

The comparison Fig. 4 to Fig. 3 show the match of class D to 6, B to 1, A to 3 to and C to the rest. Operational classes were effectively used in the inference process described in next section.

6 ANN implementation

ANN is the basic tool used in this work for inference purposes. All ANN were trained with the Adaptive Backpropagation (ABP) training algorithm [6]. The ABP is based on the classic backpropagation but uses an individual adaptive learning rate for each weight, which provides a much faster learning process. The stop training criterion was based in the well known cross validation principle, which fights against overfitting.

For the MV clients case, an ANN is used to estimate their consumption curves directly from commercial data. For the Public Stations case, the available data for each one of these stations only comprises:
- Number of LV consumers of each operational class;
- Total energy consumption of each class.

The daily diagram estimation in an hourly base (p0, p1, ..., p23) is made using a back-propagation neural network, as the one represented on Figure 5. ni and Ei are respectively the number of consumers and the total energy (monthly) of class i (the indices’ i=0..3 relate to the 4 classes, A a D previously described).

To train this ANN, 2000 patterns were generated from the data file derived from measurement campaigns. Each pattern was generated to include from 80 to 160 LV consumers, randomly selected from basic samples.

For each pattern, the 8 ANN values were settled from the classification of prototype elements, followed by energy counting and sum for each class. The values from the outputs have equivalence in the 24 time intervals from the aggregated diagram of the consumers belonging to each pattern. From the 2000 generated patterns, 1500 were taken out to train the ANN and the remaining were used for testing.

7 Results

This section begins with the presentation of some illustrative examples where one’s compare inferred load diagrams with the real thing. After, we propose a method for assessing error bandwidth around the predicted load curve. The idea is to obtain a measure of confidence intervals.
7.1 Load diagrams inference

Figure 6 presents pattern examples of the test set, comparing the real diagram (real) and ANN outputs (NN). Examples shown referred to LV consumers, summer and workdays.

Figure 7 shows similar results but for winter workdays. Global results show that the ANN is capable of estimating the test set diagrams presenting an rms error around 10%. This may be considered a good result, if we take into account the arbitrariness inherent to loads behavior.

7.2 Confidence intervals

Despite the good quality of approximation achieved (Fig. 6 and 7), it is always desirable to access the confidence intervals in order to provide a characterization of the accuracy of such estimates. It would be interesting to obtain a given bandwidth around ANN estimated load diagram in such way that the probability of a real load curve be inside that error bounds is, let’s say, 0.9.

There has been some interesting work in the area of confidence interval prediction for ANN [7-10]. In most of those studies, authors assume Gaussian or t-student distributions and estimate output variance as a function of inputs variance and of input/output transferring function, using Bayes rule.

Here, we used an auxiliary ANN (called ANNd) to learn load curves’ dispersion (error bounds distribution) depending not only on the type and number of consumers aggregated in a given public station, but also on the hour of the day, as shown in Fig. 8. This figure presents aggregated load diagrams of a given public station on different week days (lines) as well as ANN estimation (circles).

The analysis of a large amount of figures like this one has shown that there is a pattern on the errors spreading. If there is not, we can only evaluate average errors. We can observe that consumptions’ dispersion is not homogeneous, that is, the same consumer or group of consumers does not present the same uncertainty around a medium load curve for all the hours of the days. For instance, the dispersion before 7 a.m. is usually smaller than at (e.g.) 11:00.

Inputs of ANNd are the same of ANN1. Its outputs are the absolute values of the differences between ANN1 outputs and load consumption curves. This way, ANNd produces an error dispersion measure of diagrams estimated by ANN1.

It must be stressed that there are two kinds of errors: a) errors arising from ANN implementation limitations and b) errors (called dispersion errors) related with the nature of predicted values. We shouldn’t expect that ANNd learn the approximation errors that ANN1 couldn’t learn, but we hope that
Fig. 9 is a superposition of Fig. 8 with the 90% confidence intervals (that corresponds to 2.5 times the output of ANNd, as one can see in Fig. 10. This figure characterizes the relation between what we have called bandwidth factor and the percentage of hourly consumptions within the bands. These bands represent a measure of the confidence interval of the estimated load curve. Presented results contribute to confirm that adopted tools are the most suitable for the proposed objectives.

9 Conclusions

This article proposes a new inference mechanism to estimate MV/LV load curve estimation. As an additional result, confidence intervals were derived, using a second ANN to access the error of the first one. This approach has the capital advantage of including all kind of errors inherent to the load estimator ANN. Confidence intervals are useful to represent the uncertainty of the estimated diagrams. The results obtained support the adopted approach, showing that this methodology constitutes a powerful tool especially for distribution planning.

8 Further developments

Now that this approach has been extensively tested in real operation, one arrive to the conclusion that new restrict measurements campaigns should be made to complete information on a few specific type of consumers that weren’t properly represented in some particular year periods. Another interesting issue that may be included is weather influence, specially the temperature. This well known effect is used in most forecast techniques, and its inclusion in this work may lead to better results also in estimated load curves.

A new project is presently under study where one intents to integrate geographical information to improve load curve estimation performance.

References: