POLITECNICO DI TORINO
Repository ISTITUZIONALE

SHAPE: The load prediction and non-technical losses modules

(Article begins on next page)

08 November 2022

SHAPE: THE LOAD PREDICTION AND NON-TECHNICAL LOSSES MODULES

Diego Labate
Enel Distribuzione-Italy
diego.labate@enel.com

Paolo Giubbini
Enel Distribuzione-Italy
paolo.giubbini@enel.com

Gianfranco Chicco
Politecnico di Torino-Italy
gianfranco.chicco@polito.it

Federico Piglione
Politecnico di Torino-Italy
federico.piglione@polito.it

ABSTRACT

Data Analytics applied in the electric sector has been leveraged in the recent years, primarily owing to the introduction worldwide of Automatic Meter Reading/Management and smart grids technologies operated by electric Utilities. Utilities have to face with return of investment from this infrastructure and at the same time have to fulfil the promise of providing better and innovative service to the customer, retailer and distribution network.

This paper deals with the innovative SHAPE Web software platform for Data Analytics applied to the load patterns sourced from the Italian Enel network's smart meters. A previous contribution reported on the customer classification and segmentation modules implemented in the SHAPE platform. This work describes the Load prediction and Non-technical losses modules. The SHAPE Datawarehouse (DW) currently stores four years of progressively updated customer's load patterns.

1. INTRODUCTION

The problem of characterizing the load and predicting the consumption behaviour has been recognized as relevant, and the technological improvement in the metering devices has leveraged various issues in load pattern data management. With the introduction in the recent years of Automatic Meter Reading/Management (AMR/AMM) systems in many countries, there has been a growing interest in developing Data Analytics applications based on load pattern data. This interest has been fed mainly by the ability of the underlying electronic meters technology to record consumption data at a relative low cost, ranging typically from 1 to 60 minutes resolution rather than monthly. The recent smart meters technology allows to record, along with energy, other electrical parameters such as voltages and currents, besides quantities such as temperature, voltage interruptions, voltage variations, etc., driving the interest for decisional processes in knowledge extraction from this huge quantity of data and consequently raising the interest in Data Analytics/Big Data platform for electric Utilities. For a company having millions of customers, the challenge is to manage the continuous flow of data through suitable data mining applications and with appropriate interfaces, in order to enable the company users performing various types of calculations in a reasonable time and with a clear representation of the outcomes.

SHAPE is as innovative, flexible and extensible Data Analytics suite tailored for the electrical domain, allowing analysts build their own analytics workflow in a user-friendly Rich Internet Application (RIA).

Using ad hoc developed advanced time series analysis applied to load patterns sourced from the Enel's smart meters, the SHAPE platform enables the company user to address important tasks through the following main modules:

1. Load pattern datawarehouse (DW) management
2. Basic load pattern analysis
3. Customer load segmentation and classification
4. Load prediction
5. Non-technical losses detection support

A key aspect of SHAPE is the inclusion of all the features concerning the previously indicated modules in a single Web platform managed in an integrated way. This enables the operator performing sequences of tests by calling different procedures running on user-defined sets of data (aggregations) extracted from the SHAPE DW. Figure 1 shows the overall scheme of the Enel SHAPE platform. The general aspects as well as the modules 1,2,3 of the SHAPE platform have been presented in [1].

This paper introduces the Load prediction (LPR) and the Non-technical losses detection (NTLD) modules, well integrated in the Web platform in such a way that the user is able to manage the customer aggregations defined in other modules and subsequently transfer the corresponding load patterns to the LPR module or to the NTLD module for the related analyses.

The next sections of this paper are organized as follows. Section 2 describes the overall architecture of the SHAPE platform. Section 3 illustrates the characteristics of the LPR module and the related sub-modules for load pattern reconstruction and forecasting. Section 4 presents the structure of the NTDL module. The last section contains the concluding remarks.

2. GENERAL ARCHITECTURAL ASPECTS

The SHAPE platform is a Java coded Web application deployed on Apache Tomcat Web server. SHAPE is provided with its own DW (Figure 1) in which load patterns are indexed for retrieval purpose with an efficient Open source technology. For details on DW and specific architectural choice the reader can refer to [1]. All the necessary data such as daily load patterns, technical and commercial related information, temperature data, are provided from different systems data sources. In particular, temperature data are provided on an hourly basis from an external Web service both for historical and forecasted data, the latter up to 10 days
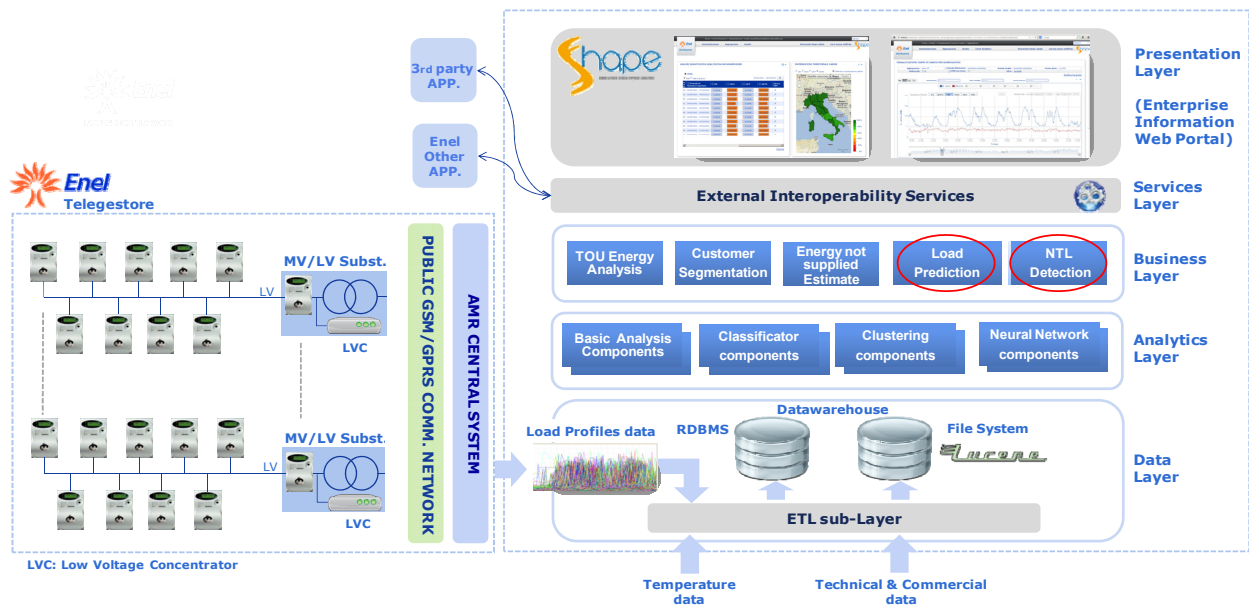
Figure 1 – SHAPE Architecture

ahead. In the SHAPE context, a daily load pattern is a 96-point time series in which each point represents the energy in the last 15-minute interval. The patterns refer to active power consumption and production, and reactive power.

The Analytics layer in SHAPE exploits many-core programming techniques that are particularly effective in the time series data mining area, thanks to inherently high data dimensionality (number of time series on which computation takes place, as well as number of samples per each time series). Most of the implemented algorithms have been carefully crafted using OpenCL^{TM} technology in order to distribute work among nVIDIA$^{®}$ GPU's computation unit while keeping work-item queues always full during both data pre-processing and mining.

The SHAPE DW currently stores four years of progressively updated customer's load patterns collected from the Enel's Telegestore starting from year 2011, as specified in [1].

3. LOAD PREDICTION MODULE

Load prediction is an important aspect in the development of any model for electricity planning. It helps determining which devices operate in a given period, so as to minimize costs and secure demand, and is the basis for establishing the needs for reserve devices in case local failures may occur in the system. Many operational decisions such as economics scheduling of generating capacity or fuel purchase are strongly related to forecasts.

The LPR module deals with short-term and medium-term forecasting of the energy demand at 15 minute resolution and serves two main functionalities, both using the same core prediction algorithm but with a different approach.

The first functionality addresses load pattern reconstruction (see Section 3.A), in which the user reconstructs portion of load patterns happened in the past for which data are not available or have been declared as invalid in the Data cleaning phase. The second functionality (see Section 3.B) addresses load pattern forecasting, in which the data referring to a future unknown portion of the load pattern are calculated by using information on the past evolution of the time series and on possible further exogenous data. The load pattern forecasting functionality, in turn, recalls the load pattern reconstruction functionality as described below.

The algorithm implemented in the LPR module adopts the Extreme Learning Machine (ELM) neural network [2]. ELM has a relatively fast training [3], which makes it appropriate to be embedded in the SHAPE portal, where real-time interaction with the user is needed.

A. Load pattern reconstruction

Missing measurement values existing in the data set are typical, due to temporary lack of meter reachability via Power Line Carrier (PLC) when the AMR system tried to read it through data concentrator (LVC in Figure 1). Though LVC allows some retry mechanism, in particular permanent noisy scenarios, the entire PLC reading procedure could be compromised.

Part of Data cleaning operations on uploaded load pattern time series are carried out during periodic automatic DW update (see ETL sub-layer in Figure 1), whereas missing values are handled by means of the reconstruction functionality in a successive phase, considering it "as a service" for the SHAPE user. This allows the procedure to improve reconstruction accuracy when e.g. in the first year of DW updating the historical data for the training phase are not enough, at the same time leaving the user the choice on which part of the DW needs to be

reconstructed, limiting the processing overload in the upload phase. In the SHAPE platform, the load reconstruction task is applied on *aggregation* object. An aggregation *A* is a collection of customer identifiers and attributes, along with the associated load patterns in a given time interval from t_1 to t_2:

$$A = \{c_1, c_2, \dots c_N, LP(c_1, t_1, t_2) \dots LP(c_N, t_1, t_2)\}$$

where:
- $c_1, c_2, \dots c_N$ are the *N* customers involved in *A* and the associated technical and commercial parameters;
- $LP(c_i, t_1, t_2)$ is the array representing the load pattern in the aggregation period from t_1 to t_2 associated with the customer c_i, for $i = 1, 2, \dots, N$.
The load pattern $LP(A, t_1, t_2)$ corresponding to the aggregation of the *N* customers in *A* is formed in a distinct way for active energy production and consumption, and for reactive energy. The aggregate load pattern is defined as

$$LP(A, t_1, t_2) = \sum_{i=1}^{N} LP(c_i, t_1, t_2)$$

In the load pattern reconstruction functionality, the user takes support from the flexibility allowed in creating the aggregation *A*, as well as from specific indicators related to the coverage of *A* in terms of percentage of days in which the daily load patterns have no missing value [1]. The user-defined aggregation is created through variegate geographic and/or technical/commercial features selection (Query By Example). Alternatively, one can select a preconfigured aggregation *A* across *Italian geographical hierarchy* (district, city, region, …, country), *Enel's district organization hierarchy*, or along the real *Electrical Enel's network hierarchy*, ranging from single customer to meters belonging to an individual LV feeder, up to aggregations of MV/LV substations. The implemented procedure reconstructs each customer's load pattern belonging to *A* as schematically represented in the box 1 of Figure 2, achieving $\widehat{A} = Reconstruct(A)$ ($\widehat{A} = R(A)$ in Figure 2) to be saved as reconstructed aggregation used in successive analyses.
An example indicating the principles of load pattern reconstruction for 48 successive quarters of hour is reported in Figure 3. The reconstructed portion of the load pattern is created by comparing 20 surrogate load patterns generated by the prediction procedure and choosing in an automatic way the one whose data better fit the connection with the first points of the real load pattern after the time interval to be reconstructed [4]. In Figure 3 the surrogate data are shown also after the reconstruction time period. In practice, in the implemented procedure the pattern reconstruction stops when the real data are available again.
The user will visualize the aggregations *A* and $\widehat{A}$, deciding on whether or not to store permanently the

reconstructed single load patterns in the DW, where each new pattern conventionally replaces the initial missing one.
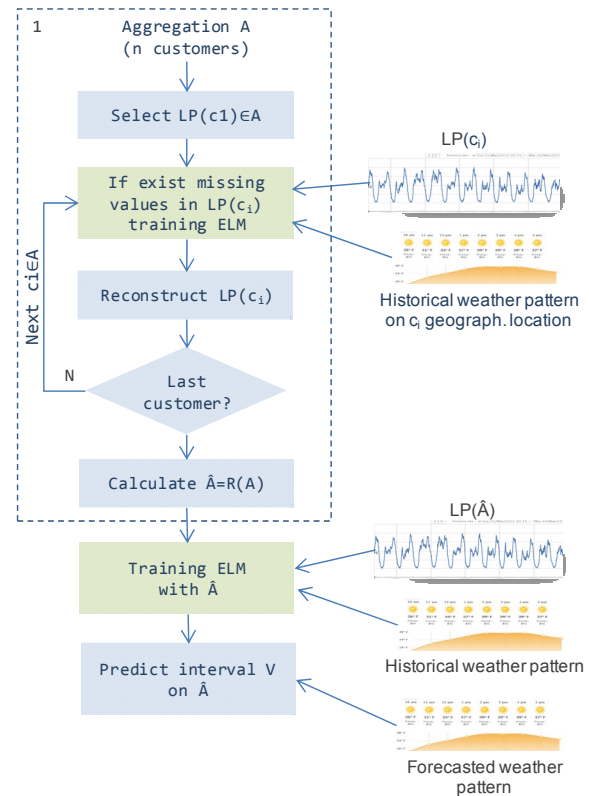

Figure 2 – Load reconstruction and forecasting task (for brevity, the time interval limits t_1 and t_2 are omitted).
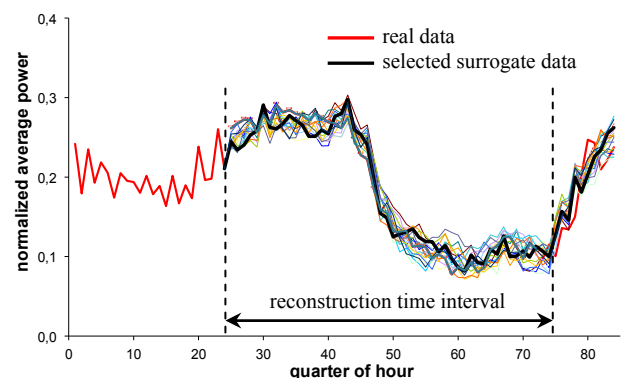

Figure 3 – Principle of load pattern reconstruction with selection of the fittest surrogate data pattern among 20 alternatives.

By this way, a future aggregation instanced on the same reconstructed customers and period will not need to be reconstructed again. Reconstructed energy samples are expressly marked in the file system structure in order to make it possible for the user to recall original data as well as to make them visually recognizable. In particular, the complete reconstructed time series is appropriate in running the functionality of Section 3.B in the short term and in the long term forecasting, as described below.

B. Load pattern forecasting

In a dedicated Web form, the SHAPE user specifies the time interval V (Figure 2) to be forecasted and the customer's load aggregation on which the forecasting procedure has to be executed. Forecasting is applied directly on the aggregate load pattern **LP**(*A*) rather than on single LPs belonging to *A*, allowing for better prediction accuracy due to an averaging effect from the number of customers involved.

Forecasting **LP**(*A*) needs to take into account that, in turn, a single LP belonging to *A* could have some missing values. In this case, the load curve **LP**(*A*) and the forecasted one could have values not representing the reality, besides the forecasting uncertainty. For this purpose, the forecasting procedure first executes the reconstruction step as described in Section 3.A, enabling the formation of a conventional reconstructed aggregation $\widehat{A}$ used to make prediction on the corresponding conventional load pattern **LP**($\widehat{A}$). The entire process is described in Figure 2. Note that the reconstruction phase (box 1) could be applied to a period external to the one of the aggregation. The length of the period in which the LPs have to be reconstructed also depends on how far the regressors needed for the ELM training are found in the **LP**(*A*) series. In addition, the choice of the regressors may depend on the presence of anomalous days, needing to extend the search of valid regressors to previous time periods.

Figure 4 shows the values of the classical forecast performance indicators MAPE (Mean Absolute Percentage Error) WMAPE (Weighted Mean Absolute Percentage Error) [3] obtained on a data set composed of 41 LV residential customers with contract power 3 kW and load pattern time series represented with 15-min time step. The prediction is made by taking the data with a moving window for the next quarter of hour, using as regressors the lags -1, -2, -3, -95, -96, and -97 on real data. The training set is of 3 weeks, with the test set on the fourth week. The forecasting algorithm is ELM with 50 hidden units. The customer data have been aggregated by starting from a single load pattern and adding progressively the other patterns, one at a time. Hence, the first prediction refers to one load pattern, and the last one uses the aggregation of the 41 load patterns. This example is particularly challenging, because it refers to a group of residential customers (including residents and non-residents, some of the latter having very irregular consumption, also with periods of null consumption) and the step of 15 min makes the load pattern data less smooth than in cases with longer time steps (e.g., hourly load patterns). The evolution of the forecasting errors depends on the sequence of load aggregations. Figure 4 also indicates that the MAPE values are very high when the number of aggregate patterns is low and the aggregate consumption may contain very small values, leading to high entries in the MAPE definition. The use of WMAPE

mitigates these high entries and provides a more appropriate formulation of the forecasting errors. By increasing the level of aggregation, the aggregate LP becomes smoother and the minimum power value in the aggregation period increases, making forecasting easier.

Figure 4 – Forecasting performance indicators for different levels of load aggregation.

4. NON-TECHNICAL LOSSES DETECTION MODULE

The NTLD module addresses various aspects referring to the non-metered energy that is delivered to the distribution systems at the substation level. Among these aspects, the main point relates to energy thefts corresponding to energy adsorbed by the customers but partially or totally not measured by the company's smart meters. This may happen as a result of meter tampering or by bypassing it (Figure 5).

Unlike Technical losses, the Non-technical losses represent an avoidable financial loss for the Utility and the other actors involved, although it has been estimated that a significant percentage of the energy consumed by the fraudulent users would not be consumed if those users would have to pay for that energy [5]. Suitable indicators can be defined by using the experience of the domain expert (coded as a set of rules introduced in a supervised procedure) or unsupervised indicators calculated from the data set in an automatic way.

The rationale of the non-technical losses detection module is based on the construction of a set of relevant indicators, each of which highlights a specific aspect of the effect that non-technical losses may have on the electricity consumption pattern. As the individual indicators are set up to recognize different characteristics of the perpetrated frauds, clearly there will not be any individual pattern exhibiting high values for all the indicators.

Thanks to the availability of historical load patterns corresponding to real anomalies or to the evolution in time of the consumption from fraudsters during the fraud period, a supervised indicator has also been implemented. With the aim to identify an overall ranking of the customers, the supervised and unsupervised indicators values have been combined in a unique indicator, named IPPNT.

Referring to the supervised indicator (named "CLASS") a classification tree algorithm has been trained externally to the platform for most commercial macro-categories. At the end of the training phase, almost each macro-category had its own classification model deployed in the SHAPE Analytic Layer. Scoring each new customer with the supervised indicator consists of selecting the macro-category model the customer under analysis belongs to, and predicting the class label F (possible fraudulent) towards NF (estimated as non-fraudulent) for each of its daily load patterns. At the end of the scoring procedure, the customer indicator "CLASS" corresponds to the percentage of classification outputs with label F vs. the number of daily LPs included in the period under analysis:

$$CLASS = \frac{\#F}{\#daily\ LPs} \cdot 100$$

Non-technical losses detection interfaces

The SHAPE user starts the analysis by establishing the customer aggregation **A** to be ranked with reference to the fraud indicators. Each indicator can be enabled/disabled during the calculation and in the final results representation.

Once the fraud analysis has been executed on **A**, the output is represented by all the fraud indicators and the related values for each customers visualized in a coloured or numerical list. The customers are overall ranked in the descending order of probability of being an abnormal/fraudulent consumer by using the global index IPPNT that combines the estimated indicators (Figure 6). This overall ranking is of particular interest both to verify a specific customer or in order to improve the success rate of the periodical in-field inspection campaigns towards a pre-allocated company budget in a selected geographical zone.
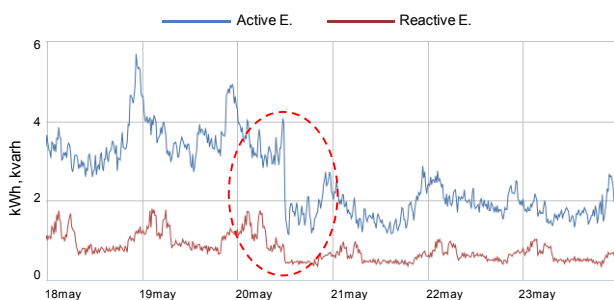

Figure 5 – Example of tampering executed on the smart meter's measurement block from 20 May.

5. CONCLUSIONS AND FUTURE WORK

The SHAPE platform is a fully integrated software tool aimed at answering relevant business questions for the electrical utilities, concerning load pattern analysis, classification, prediction and non-technical losses detection. This paper has presented the LPR module (short-term and medium-term forecasting for any customer aggregation) and the NTLD module (supporting the company to identify the more probable fraudulent customers from their metered data). Future developments refer to extended parallelization of the implemented Analytics layer and to improving scalability performance.
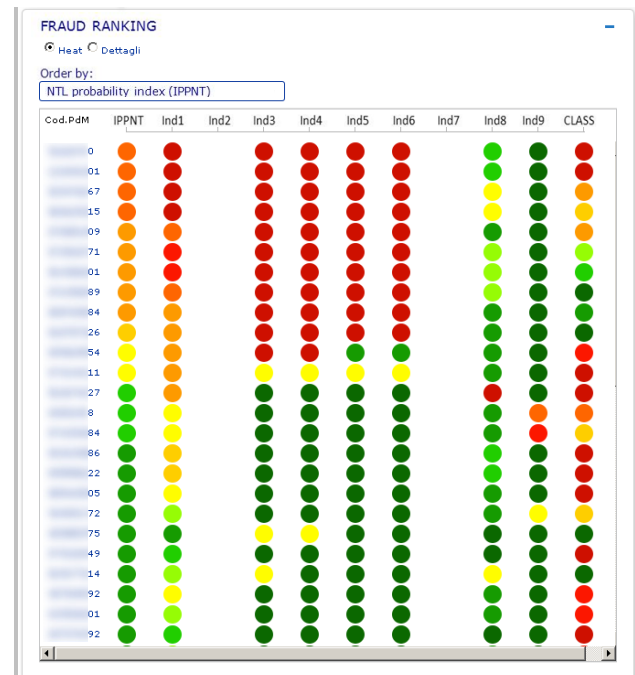

Figure 6 – Example of customers' ranking with reference to the overall fraud indicator IPPNT.

6. ACKNOWLEDGMENTS

REFERENCES

[1] D.Labate, P.Giubbini, G.Chicco, M.Ettorre, SHAPE: A new Business Analytics web platform for getting insights on electrical load patterns, CIRED Workshop - Rome, 11-12 June 2014, paper 0354.

[2] H. Guang-Bin, Z. Qin-Yu and S. Chee-Kheong, Extreme Learning Machine: Theory and Applications, Neurocomputing, vol. 70, 2006, pp. 489-501.

[3] S. Ding, H. Zhao, Y. Zhang, X. Xu and R. Nie, Extreme learning machine: algorithm, theory and applications, Artificial Intelligence Review, Springer, 2013.

[4] A.G. Rusu, Analysis and forecasting of data time series in electrical systems, Doctoral thesis, Politecnico di Torino, 2007.

[5] P. Antmann, Reducing technical and Non-Technical losses in the Power Sector, Background paper for the World Bank group Energy sector strategy, July 2009.