Detection of Irregular Consumption to Load Monitoring in Smart Grids Environment

Bogdan C. Neagu[1], Gheorghe Grigora□[1]

Gheorghe Asachi Technical University, Ia□i, Romania,
bogdan.neagu@tuiasi.ro, phone:+40-232-278683-1161

***Abstract* – Today, power distribution operators are concerned about irregular behavior of electricity consumption in their decision-making process. In the load data, abnormal consumptions may happen due to measurement error, undetected consumption, illegal electricity connection, improperly installed equipment etc. The paper presents a statistical approach which acquires the basic concept of knowledge discovery on databases (KDD)with data mining, focuses on the identification of outliers (irregular consumption) from load curves recorded by Smart Meters on the substation of the power distribution networks. In order to validate the proposed approach, a real database with 300 substations from rural area was used. The results demonstrate that by detection and elimination of irregular consumption, the load curves can be efficiently used by modern distribution operators for a better operation and planning of actually "smart" networks.**

Keywords – data mining, load monitoring, irregular consumption, Smart Grids

I. INTRODUCTION

The Smart Grids is an electrical network that use digital technology to monitor and control the electricity consumption between generation to end users. In accordance with the requirements of the EU, 80% of end users must have devices withy smart measures before 2020[1]. In many data analysis assignment, a large number of variables are being recordedor sampled. One of the first steps towards obtaining a coherent analysis is thedetection of outlaying observations. Although, outliers are often consideredas an error or noise, they may carry important information. Detected outliersare candidates for aberrant data that may otherwise adversely lead to modelmisspecification, biased parameter estimation and incorrect results [2].

The analysis of data represents the starting point for many applications, in the design or operation phase for online control or complex processes. Nowadays, the need to extend the capabilities of human analysis for handling the overwhelming quantity of data that we are able to collect has become increasingly necessary. Since computers have enabled the possibility of storing large amounts of data, it is only natural to resort to computational techniques to help us discover meaningful patterns and massive structures volume data [3].

The load curve plays a fundamental role in operation and planning of power systems. Unfortunately, due to various random factors, the load curves always contain abnormal, deviation, unrepresentative, noisy, strange, anomalous and missing data. It is fundamental for power systems operators to detect and repair anomalous or abnormal data before the use of load curve in planning and modeling process.

In this reason, the paper proposes an approach for detection of outliers using the two stages. In first step, a knowledge discovery on databases technique for extraction of the load curves main indicators computed with information provided by Smart Meters was used. The second stage coincide with the detection of irregular consumption (outliers) from the above computed indicators, using a statistical based data mining technique. The proposed method was tested using a real database with 280 substations that served rural area. The results highlight the ability of proposed approach to be efficiently used by distribution operators in order to take an estimable decision regarding the operation and planning of power systems using only useful load curve characteristic information provided by a large database from smart meters through data mining techniques.

However, the biggest danger of outliers is their impact on statistical results whichcould lead to misleading conclusions. In the business world, decision making whichis based on misleading conclusions can be very costly and devastating for anenterprise. Therefore, it is crucial that outliers be detected first and, subsequently,decide on how to treat them. Outliers might be just mistyped values, but theymay also point to important factors in business processes [4].

II. RELATED RESULTS IN THE LITERATURE

For operation and planning of modern power systems, the extraction of useful load curve information [5] characterized by its indicators (load factor, irregularity factor, loss factor, fill factor, etc.) from some large databases, the clustering or statistical techniques can be used.Outlier data refer that data deviate from most data in the data set. Therefore, the outlier data represent one of the most prominent issues in power consumption (load curve) data. The methods for detection of load curve outlier data can be divided into two main categories, based on data mining and state estimation [6]. Also, in the literature those based on data mining use other approaches i.e. neural network as in [7], fuzzy method [8], clustering techniques [9] and time series analysis [10].

III. DETECTION OF IRREGULAR CONSUMPTION USING A INOVATIVE KNOWLEDGE DATA DISCOVER APPROACH

Data mining is defined as a nontrivial extraction of implicit, previously unknown, and potentially useful information from data" [11]. Data mining is a process that extracts the information and knowledge which is implicit in them, unknown in advance, but potentially useful, from the massive, incomplete, fuzzy, noisy and random data generated in the practical application [12]. The steps for knowledge discovery on databases process are shown in Fig. 1[13].

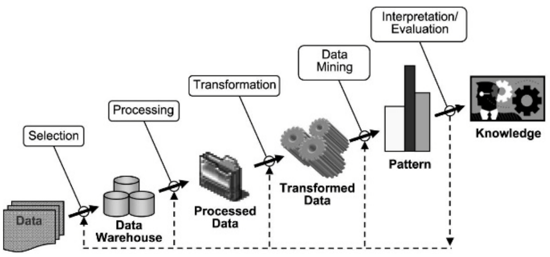

Fig. 1. KDD Process

Through data mining techniques, valuable information can be extracted from large amount of big data in a power system.The representative characteristics of load curves indicators can be obtained by removing the abnormal data (outliers), using a statistical method. The proposed approach is tailored to the Gaussian distribution (Fig. 1) by its implicit assumption that the data can be characterized by its first two statistical moments. This is a situation where the probability density of the normal data is estimated in a simple parametric fashion.

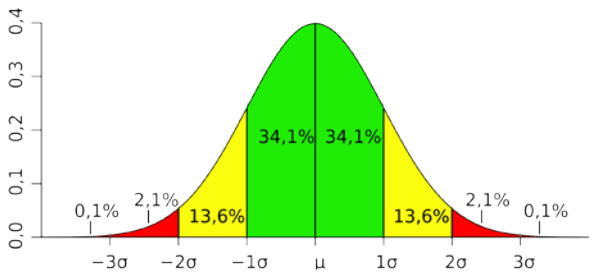

Fig. 2. The Normal(Gaussian) Distribution, [14]

A discordant outlier in a data set is an observation that is surprisingly different from the rest of the data and therefore is believed to be generated by an alternate mechanism to the other data. The Gaussian distribution can be generated by a very large number of measurements, being perfectly balanced if the mean is exactly in the middle (median). The curve has a specific bell shape that might be wider (more spread out) or narrower (closer to the mean). This type of distribution occurs often in data measurement (X). Mathematically, one standard deviation is μ ± σ (68% of entire data), where μ is the arithmetic mean. In accordance with Fig. 2, because two standard deviations have up to 95% of the entire data, the outside values are considered outliers. Because the standard deviation (σ):

$$\sigma = \frac{1}{T_f}\int_0^{T_f} X^2(t)dt - \left(\frac{1}{T_f}\int_0^{T_f} X(t)dt\right)^2 \qquad (1)$$

is a specific distance from the mean value (μ):

$$\mu = \frac{1}{T_f}\int_0^{T_f} X(t)dt \qquad (2)$$

the outliers (Λ) can be estimated using the statistic indicators (1) and (2), with the following relation:

$$\Lambda = \mu \pm \psi \cdot \sigma \qquad (3)$$

where ψ is evaluated after a deeper analysis [12].

IV. RESULTS AND DISCUSSIONS

In these paragraph, a study case was made. Then, a database with 300power transformers placed in the distribution substations supplied by rural distribution feeders with nominal voltage 20 kV was considered. The rated powers of transformers have the primary data presented in Table 1.

Table1.The primary data of transformers from distribution substations.

Type	Number of transformers	Rated Power [kVA]	Active losses	
			Non-load [kW]	Load [kW]
1	9	40	0.230	1.0
2	47	63	0.300	1.5
3	84	100	0.350	2.3
4	92	160	0.525	3.1
5	56	250	0.680	4.4
6	9	400	0.940	6.0
7	1	630	1.250	8.2

All distribution substations are equipped with Smart Meters. From these devices, we collected the load curves, single-phase and tree-phase voltages, power factors etc.

on a low voltage side of transformers. The measurements correspond to one week with a sample step by 10 minutes. An automatic process to extract the characteristics of power curveswas used. Thus, the characteristics of the active, reactive and apparent power curves for every substation were obtained. For the data mining process, only the following factors calculated with the expression from [15] for apparent power were considered due to the strong dependence determined based the correlation analysis: load factor (LF), loss factor (LS), irregularity factor (IF), fill factor (FF). The Pearson coefficients are given in Table 2. For the other load curve characteristics, the values were less than 0.8.

Table2.The Pearson correlation coefficients.

Indicators	LF	LS	IF	FF
LF	1.000	0.985	0.899	1.000
LS	0.985	1.000	0.887	0.986
IF	0.899	0.887	1.000	0.899
FF	1.000	0.986	0.899	1.000

In Fig. 3, the values of loss factor (blue colour), for a characteristic day (Wednesday) presents the outliers (red) computed with (3), using the statistic indicators.

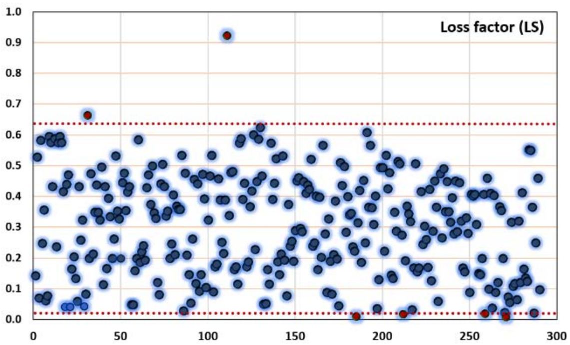

Fig. 3. The loss factor for all substation

Due to some issuesof different reactive power curves a data cleaning was applied. The outliers are indicated in Table 3 for the computed load curve indicators. The value for ψ was established at 2, determined using a statistical processing of database.

Table3.The detection of outliers from load curve indicators.

Characte-ristics	Outliers, *Λ*				
	min	max	No.	μ	σ
LF	0.154	0.796	13	0.505	0.176
LS	0.021	0.637	6	0.308	0.165
IF	0.009	0.611	13	0.301	0.155
FF	0.154	0.856	12	0.505	0.175

By removing the outliers and after the cleaning process, eightpower distribution substations were removed, such that at the final only the characteristics from 290 substations were used. In the following, the analyse was made on the transformer groups. Using the proposed approach, finally can characterize the electric distribution stations from the power curves in transformersviewpoints. Thus, the database was divided in seven transformers categories: 40 kVA (9), 63 kVA (45), 100 kVA (80), 160 kVA (90), 250 kVA (56), 400 kVA (9), and 630 kVA (1). For the accuracy of the database values, in Table 4 the results (without outlier) for the seven transformer categories are presented.

Table4.The statistical results of the load curves indicators for each power transformers category.

Characteristics		LF	LS	IF	FF
40 kVA	μ	0.674	0.519	0.512	0.744
	σ	0.059	0.055	0.066	0.031
63 kVA	μ	0.626	0.491	0.510	0.743
	σ	0.102	0.101	0.203	0.121
100 kVA	μ	0.615	0.483	0.519	0.747
	σ	0.098	0.097	0.187	0.110
160 kVA	μ	0.625	0.489	0.519	0.749
	σ	0.113	0.110	0.174	0.099
250 kVA	μ	0.627	0.493	0.528	0.752
	σ	0.090	0.080	0.172	0.099
400 kVA	μ	0.663	0.505	0.472	0.731
	σ	0.060	0.064	0.145	0.078
630 kVA	μ	0.720	0.552	0.469	0.740

Furthermore, in Fig. 4 are indicate the mean value of the apparent power that flow through all considered transformer categories, with and without outliers. The aforementioned values will be used in post computation of energy losses by distribution operators in the power systems planning process.

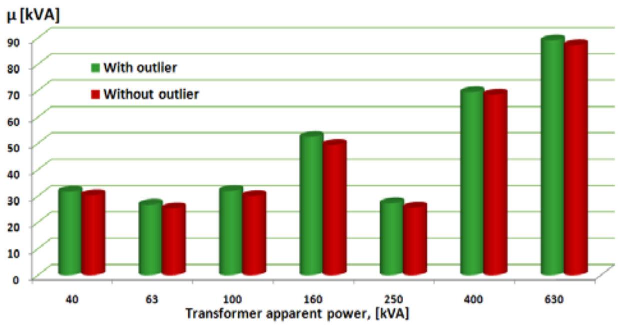

Fig. 4. The mean value for all power transformer categories with and without outliers

V. CONCLUSIONS

The paper presents a new approach that uses a statistical based data mining for detection of irregular consumption using information provided by Smart Meters. A real database with 300 rural substations was tested. The results demonstrate that by outliers eliminating, the proposed approach can be efficiently used by distribution operators for a better operation and planning of power systems. From the results analysis, it can be observed that there are some factors with a more important influence on the detection of the outliers using statistical approach, i.e.load factor, loss factor, irregularity factor and fill factor. Thus, the development planning of the distribution substations knowing only few indicators of the load curves could simplify the work of distribution operators, by proper computation of energy losses

REFERENCES

[1] Levenda, A., Mahmoudi, D., Sussman, G.,*The Neoliberal Politics of "Smart": Electricity Consumption, Household Monitoring, and the Enterprise Form*. Canadian Journal of Communication, v. 40, n. 4, nov. 2015, pp. 615-636.

[2] Gogoi, P., Bhattacharyya, D. K., Borah, B., Kalita, J. K.,*A Survey of Outlier Detection Methods in Network Anomaly Identification*. Computer journal,54(4),2011, pp.570-588.

[3] Abiteboul, S., Arenas, M., Barceló, P., Bienvenu M.,*Research directions for Principles of Data Management*, SIGMOD Record, 45(4), 2016,pp. 5–17.

[4] Zmuk, B., Speeding problem detection in business surveys: benefits of statistical outlier detection methods, Croatian Operational Research Review, no. 8, 2017, pp. 33–59.

[5] Queiroz, L., M., O., Roselli, M.,A., Cavellucci, C., Lyra, C.,*Energy Losses Estimation in Power Distribution Systems*, IEEE Transactions on Power Systems, 27(4), 2012, pp. 1879-1887.

[6] Diksha, M., Raghavendra, T., S., *A Survey on Big Data Energy Based on Smart Grid*, Int. J. of Adv. Res., Id. and Innov. in Tech., 3(1), 2017, pp.594-598.

[7] Rakhshani,E., Sariri, I., Rouzbehi,K., *Application of data mining on fault detection and prediction in boiler of power plant using artificial neural network*, Int. J. Electr. Power Energy Syst., 2009, pp. 473-478.

[8] Cateni,S., Colla, V.,Vannucci, M.,*A fuzzy logic-based method for outliers detection*. in Artif. Intellig. and App., 2007, pp. 605–610.

[9] Loureiro,A., Torgo,L., Soares,C., *Outlier Detection using Clustering Methods: a Data Cleaning Application*, in Proc. of KDNet Symp. on Knowledge-based Syst. for the Public Sector, Bonn, Germany, 2004.

[10] Tang,G., Wu,K., Lei, J.,Bi,Z., Tang.J., *From landscape to portrait: A new approach for outlier detection in load curve data*, IEEE Trans. Smart Grid, 5(4), 2014, pp. 1764-1773.

[11] Jia,N., Wang,J., S., Li,N., *Application of data mining in intelligent power consumption*, Intern. Conf. on Automatic Control and Artifi. Intelligence (ACAI 2012), 2012, pp. 538-541.

[12] NeaguB., C., Grigora□, G.,*Data Mininig Tools in Electricity Distribution Systems*, Acta Electr. J., 56(3), 2015, pp. 209-212.

[13] Gilchrist, M., Mooers, D., Skrubbeltrang, G., Vachon, F., *Knowledge discovery in databases for competitive advantage*. Journal of Management & Strategy, 3(2), 2010, pp. 2-15.

[14] https://kanbanize.com/blog/normal-gaussian-distribution-over-cycle-time/.

[15] Neagu, B., C., Georgescu, G., Gusa, M., D., *Load Curves Characteristics of Consumers Supplied From Electricity Repartition and Distribution Public Systems*, Bul. Instit. Polit. din Iasi,v. LVII (LXI),n.1, 2011, pp.141-157.