Volume 4, No. 6, May 2013 (Special Issue)
International Journal of Advanced Research in Computer Science
REVIEW ARTICAL
Available Online at www.ijarcs.info

Design of Electricity Theft Monitoring System in Customer Consumptions

Saurabh Manohar Hadge*, Sarang Giridhar Kamble and Prof. A. D. Raut
Dept. of Information & tech.
J.D.I.E.T., Yavatmal, Maharashtra, India

Abstract**—**This paper proposes a computational technique for the classification of electricity consumption profiles. The methodology is comprised of two steps. In the first one, a C-means-based fuzzy clustering is performed in order to find consumers with similar consumption profiles. Afterwards, a fuzzy classification is performed using a fuzzy membership matrix and the Euclidean distance to the cluster centers. Then, the distance measures are normalized and ordered, yielding a unitary index score, where the potential fraudsters or users with irregular patterns of consumption have the highest scores. The approach was tested and validated on a real database, showing good performance in tasks of fraud and measurement defect detection.

Index Terms**—**Data mining, electricity theft, fuzzy clustering, nontechnical losses.

I. INTRODUCTION

ABNORMAL or irregular consumption is closely related to frauds, measurement errors, undetected consumption, installation problems, illegal electricity connection, and nonpayment. These problems lead to commercial losses (i.e., there are differences between the actual consumption and the estimated consumption (reading) on the bill). Commercial losses in Brazil were about 6% of the whole supplied energy in 2007, summing up 15 TW.h or U.S.$5 billion, considering the average price including taxes [1]. Those frauds have consequences, such as overloading the power-supply system, deterioration in the power-supply quality, and rising energy costs. Unfortunately, there is no doubt that the customers will eventually incur these extra costs.

In order to reduce the commercial losses, the energy utilities (companies) carry out periodic inspections of their facilities.

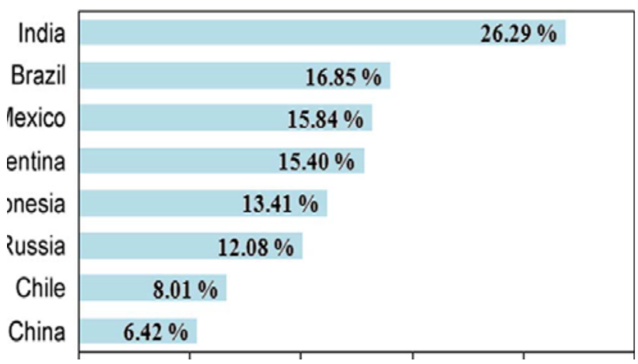
However, these inspections are expensive and do not represent a good tradeoff for these companies. As a consequence, they have applied computational tools, aiming to detect fraudulent patterns and problems.\The detection and identification of frauds in power systems has been initially addressed with statistical techniques [2], [3]. Subsequently, a variety of modern techniques, mostly based on artificial intelligence, has been extensively applied to solve this problem. Several papers report applications of neural networks, fuzzy set, and rough set [4]–[7]. More recent works explore the application of techniques, such as support vector machines and data mining, the latter associated with other intelligent techniques [8]–[11].

There can be many causes for this abnormal behavior, such as measurement error, improperly installed equipment, energy fraud, and error in the database, among others. Besides, companies are interested in detecting, assessing, and recovering unbilled energy revenues in today’s competitive environment.

However, since the number of irregularities (true or false) can be considerable, the implementation of this methodology can be difficult and expensive. Therefore, a reasonable degree of suspicion alone does not justify field inspections. In other words, inspections should consider many aspects, especially, financial return. Thus, inspection should prioritize suspicious cases that lead to higher margins of revenue recovery. Several data-mining-based methodologies were developed to identify abnormalities in energy consumption. Clustering techniques have also been successfully applied in this task, as a primary or secondary step for other algorithms [4], [12], [13].

Furthermore, besides being useful to reduce data-space requirements, it is also an efficient tool to detect natural consumer groups with the same energy consumption profiles. It turns out that clustering is the most suitable technique for modeling and identifying the various energy consumption profiles over an electric utility, especially fuzzy clustering [14]. In this paper, we propose a cluster-based classification strategy that uses a reduced set of five attributes, which can be derived from data commonly available at most utilities. The corresponding two-step unsupervised algorithm does not use dissimilarity matrices, making an attempt to emphasize the natural fuzziness problem by means of an intercluster distance measure. Quality metrics commonly applied in classification systems were used to validate the approach, which performed better than other proposals.

In this context, this paper presents a proposal for identifying suspect profiles of energy consumption compared to regular consumption profiles. The proposed approach determines the distance between the client consumptions and the regular (normal) profile. Fuzzy-based clustering is used to identify subgroups with similar profiles, and the abnormality degree is then calculated.



The main contributions of this paper are as follows.

a. It provides a simple method that requires few attributes as inputs, which is appropriate for most practical distribution systems.
b. It provides good assertiveness to real-life systems.
c. The detection thresholds can be adjusted by users.
d. It can be used by the utility to guide inspections following a criterion of maximum financial return.
e. The proposed algorithm is unsupervised and, therefore, does not depend on rules. It can be applied to any distribution utility.

This paper is organized as follows. Section II analyzes the commercial losses in electric energy systems and indicates the ways to prevent these losses. Section III presents the knowledge discovering process and the methodology for identifying abnormalities in energy consumption. Section IV details how the methodology has been built. Section V shows a case study and its results. Finally, Section VI presents the conclusions of this paper.

II. LOSSES IN POWER SYSTEMS

Losses cannot be avoided during the stages of generation, transmission, and distribution. Moreover, these problems can be caused by sources of various natures. The losses associated with physical laws are called *technical losses*, and should be considered in the planning and dimensioning of distribution systems as well as for maintaining the quality of the supplied energy. The average historical loss is estimated at 15% in Brazil. Considering the year 2004, losses reached a level of 16.85%, as illustrated in Fig. 1, where a comparison between losses of the major developing countries is shown [15].

Using the average value of energy purchasing distributors, estimated at R$80/MW.h in Brazilian currency, the loss rate in Brazil in 2004 was equivalent to U.S.$4 billion. If we consider the average energy sales (R$231.35), it caused financial losses of R$11 billion. The high rate of loss influences the tariff paid by consumers and the entire power-supply system as well, because additional generation and transmission of energy are required to supply the losses. Moreover, there is the environmental cost (pollutant emissions) and the additional infrastructure required to transport the energy excess.

A. Technical Losses:

Technical losses will always arise as the physics of electricity transport means that, no power system can be perfect in its delivery of energy to the end customer. Technical losses are naturally occurring losses (caused by actions internal to the power system) and consist mainly of power dissipation in electrical system components such as transmission lines, power transformers, measurement systems, etc.

Technical losses are possible to compute and control, provided the power system in question consists of known quantities, viz., resistance, reactance, capacitance, voltage, current and power.

These are routinely calculated by utility companies as a way to specify what components will be added to the systems. Loads are not included in the losses because they are actually intended to receive as much energy as possible. Technical losses in power systems are caused by the physical properties of the components of power systems. Example, *I2R* loss or copper loss – in the conductor cables, transformers, switches and generators. The most obvious example is the power dissipated in transmission lines and transformers due to their internal impedance. Technical losses are easy to simulate and calculate; computation tools for calculating power flow, losses, and equipment status in power systems have been developed for some time.

The instantaneous power loss, *Ploss*(*t*) in a transmission line can be expressed as:

$$P\,(t\,)\;P\,(t\,)\;P\,(t\,)\;loss\;source\;load = -\;1$$

Where *Psource*(*t*) is the instantaneous power that the source injects into the transmission line and *Pload*(*t*) is the instantaneous power consumed by the load at the other end of the transmission line. Thus the energy loss, *Wloss*, is given by:

$$= \int (\;)\;b\;loss\;a\;loss\;W\;P\;t\;dt\;2$$

Where *a* and *b* are respectively the starting point and ending point of the time interval being evaluated. It must be noted that a fairly accurate description of *Ploss*(*t*) as a function of time is always needed to make a reliable prediction of *Wloss*.

B. Non-Technical Losses (Commercial Losses):

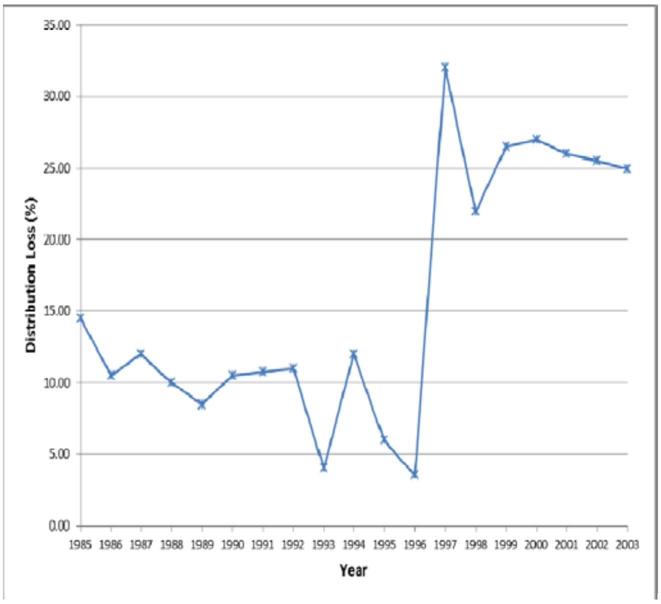
These refer to losses that are independent of technical losses in the power system. Two common examples of sources of such losses are component breakdowns that drastically increase losses before they are replaced and electricity theft. Losses incurred by equipment breakdown are quite rare. These include losses from equipment struck by lightning, equipment damaged by time and neglect. Most power companies do not allow equipment to breakdown in such a way and virtually all companies maintain some form of maintenance policies.

Other probable causes of commercial losses are:

a. Non-payment of bills by customers
b. Errors in technical losses computation
c. Errors in accounting and record keeping that distort technical information.
d. Inaccurate or missing inventories of data on customers

The most prominent forms of commercial losses in Ghana are electricity theft and non-payment of bills. Non-payment, as the name implies, refers to cases where customers refuse or are unable to pay for the electricity used. However, the other forms are not analyzed thoroughly in this project. Nontechnical losses are very difficult to quantify or detect and are more problematic than the other losses.

Non-technical losses can also be viewed as undetected load; customers that the utilities do not know exist. When an undetected load is attached to the system, the actual losses increase while the losses expected by the utilities remains the same. The increased losses will show on the utilities' accounts, and the costs will be passed to customers as transmission and distribution charges. Research has shown that, transmission and distribution costs in Ghana are calculated as part of the customers' bills, while in other countries, customers are usually charged a single flat energy rate that includes all services. This means that, the transmission and distribution losses that increased due to commercial losses would be charged either to the existing customer whose power lines are illegally tapped, or the utility, depending on the method of theft. Both ECG and NED losses range from 24 to 30 percent of power generated, collection rates range from 75 to 85 percent of billing and arrears from government agencies significantly weaken balance sheets [3]. In recent years, ECG has undertaken several measures to reduce losses. Figure 2 shows the total distribution losses of ECG from 1985 to 2003. From the figure, it is observed that losses were relatively low between 1993 and 1996 but increased thereafter.



III. PROCESS OF KNOWLEDGE DISCOVERY IN DATABASES

A well-devised method for discovering irregular profiles involves handling large databases. Therefore, a well-defined methodology should be adopted to obtain good results. Thus, we have used knowledge discovery in databases (KDD) techniques in this work. According to [17], KDD is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Frequently, discovering knowledge in data bases is called data mining. However, data mining is just one of the steps required for discovering knowledge in a database.

A. KDD Phases:

Historically, the notion of finding useful patterns in data has been given a variety of names, including data mining, knowledge extraction, information discovery, information harvesting, data archaeology, and data pattern processing. The term *data mining* has mostly been used by statisticians, data analysts, and the management information systems (MIS) communities. It has also gained popularity in the database field. The phrase *knowledge discovery in databases* was coined at the first KDD workshop in 1989 (Piatetsky-Shapiro 1991) to emphasize that knowledge is the end product of a data-driven discovery. It has been popularized in the AI and machine-learning fields.

KDD refers to the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process. Data mining is the application of specific algorithms for extracting patterns from data. The distinction between the KDD process and the data-mining step (within the process) is a central point of this article. The additional steps in the KDD process, such as data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the results of mining, are essential to ensure that useful knowledge is derived from the data. Blind application of data-mining methods (rightly criticized as data dredging in the statistical literature) can be a dangerous activity, easily leading to the discovery of meaningless and invalid patterns. The data-mining component of KDD currently relies heavily on known techniques from machine learning, pattern recognition, and statistics to find patterns from data in the data-mining step of the KDD process.

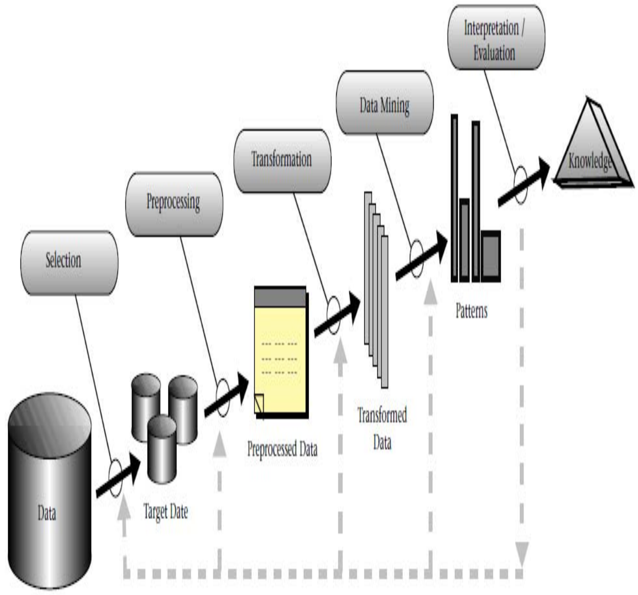


Figure 1. An Overview of the Steps That Compose the KDD Process.

B. Strategies for Data Mining:

Data mining commonly involves some classes of tasks [18]: association discovering, discriminant analysis, summarization, outlier detection, classification, and clustering.We will describe the last two classes.

a. ***Classification:*** This task arranges the data into predefined groups. In other words, we need to discover a function to map each record to a specific class . The algorithm starts using some classified elements in the form , representing the training phase. Then, the new function is applied to the remaining elements, associating it with a corresponding class. The following techniques are commonly used for classification: tree decision, Bayesian networks, Rule-based systems, support vector machines, and artificial neural networks [19]. There are also other methodologies, such as rough sets, fuzzy sets, and genetic algorithms, and each method addresses a specific problem. In this study, the efficiency of a classifier system has been evaluated by two metrics: 1) assertivity and 2) sensitivity. Furthermore, we can use a confusion matrix to indicate the relationship between actual and forecasted classes. Besides, this matrix supports the creation of the concepts of false-positives (FP), false-negatives (FN), true-positives (TP), and true-negatives (TN) [19].

The assertiveness and sensitivity of a classifier system are calculated by (1) and (2), respectively, where represents theexisting positives in the database. Other forms of evaluation can be found in [18] and [19]

CONFERENCE PAPER
"A National Level Conference on Recent Trends in Information Technology and Technical Symposium" On 09th March 2013
Organized by
Dept. of IT, Jawaharlal Darda Inst. Of Eng. & Tech., Yavatmal (MS), India

$$ass = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \quad (1)$$

$$sens = \frac{\mathrm{TP}}{\mathrm{P}} \quad (2)$$

Where is the total of the positive cases.

b. ***Clustering:*** Clustering refers to the grouping of records, observations, or cases into classes of similar objects. A cluster is a collection of records that are similar to each other, and different from records in other clusters [20] or else, a clustering task attempts to segment the entire data into subgroups or clusters, where the similarity of records within a cluster is maximized, whereas the similarity of records outside a cluster is minimized. Thus, the clustering task can be used as a preprocessing technique for classifying data into detected clusters. C-means is one of the most popular algorithms for clustering data, where a record in a data base belongs to only one cluster. On the other hand, a record belongs to more than one cluster with a certain degree (membership degree) in the fuzzy - means (FCM). Considering three matrices , , and , which are the data matrix, center matrix, and fuzzy membership degree matrix, respectively, the FCM attempts to minimize

$$J_f(X,U_f,C) = \sum_{i=1}^{n}\sum_{j=1}^{c}\mu_{ij}^{m}d_{ij}^{2} \quad (3)$$

Where is the number of elements, is the number of attributes for each element, is the number of clusters, is the fuzzification factor, is the membership degree of the th element into the cluster, and is the distance between the th element and the center of the cluster. The minimization is performed by means of a two-step process. In the first one, the center of clusters is constants, and we try to find the optimum value of the membership degrees. Afterwards, the center of the cluster is minimized, keeping the membership degree constant. The optimum values are calculated by using two partial differential equations

$$c_{jk} = \frac{\sum_{i=1}^{n}\mu_{ij}^{m}x_{ik}}{\sum_{i=1}^{n}\mu_{ij}^{m}} \quad (4)$$

$$\mu_{ij} = \frac{d_{ij}^{-\frac{2}{m-1}}}{\sum_{l=1}^{c}d_{il}^{-\frac{2}{m-1}}}, \quad (5)$$

Step 1) random initialization of clustering centers;
Step 2) find out the membership degree of each cluster by means of (5);
Step 3) obtain the new centers by using (4);
Step 4) test a stop criterion, such as if the stop criterion was not met, then return to Step 2).

IV. DESCRIPTION OF THE PROPOSED METHOD

Roughly speaking, most methodologies can be grouped as follows: 1) those with a lot of attributes and 2) those with limited applicability to medium- or high-voltage systems. However, in practice, utilities do not have enough data for adequate monitoring of distribution losses. Indeed, more than half of commercial losses occur in residential areas. Furthermore, some electricity companies have no other data than consumer registering and incoming. The proposed method overcomes those problems by proposing a simple model for identifying abnormal consumption patterns using few attributes. It has been developed based on the following steps:
Step 1) choosing a set of attributes that represents a pattern of energy use;
Step 2) setting the domain of these attributes, creating a regular consumption range;
Step 3) classifying the abnormal patterns after some period of time , taking into account the distance between the abnormal and regular patterns.
These steps are detailed in the following subsections.

A. *Identifying Relevant Attributes:*

Five attributes have been chosen to create a general pattern of power consumption for each client.

a. M6 represents the average consumption (in kilowatt-hours) of a specific client in the last six months;
b. Max6 represents the maximum consumption (in kilowatt hours) of a specific client in the last six months;
c. Dev6represents the standard deviation of the consumption of a specific client in the last six months;
d. Apt6 represents the sum of the inspection remarks in the last six months (notifications reported by regular meter readers regarding the installation conditions)
e. MS6 represents the average power consumption in a residential area where the client has lived in the last six months.

Therefore, each client is represented by a vector with the fiveconsidered attributes. In other words, this vector represents a pattern of consumption. The motivation for choosing them was obtained by analyzing real data from electric utilities and examining some existing fraud detection systems.

Three reasons for using the six-months interval were identified. First, this period was chosen to minimize the effects of adverse events such as vacant properties, season changes, and occupancy increases, among others. Second, an interval greater than six months would create difficulties for running the system in utilities that lack data on power consumption. Finally, it was found that several expert systems have also successfully used this time span to establish relevant rules for the process.

Regarding the selected attributes, each one was included to model specific information. The initial three attributes are correlated, and they have the same data source (monthly consumption data table). Even so, each one makes its own contribution to the energy profile: the variable M6 as an estimate of a typical power consumption of a client; the MAX6 attribute as the peak consumption value and DEV6as the standard deviation consumption in the given period. The last two attributes are mutually independent, providing information on the average power consumption in the neighbourhood (MS6)of a particular consumer group (residential, commercial, single phase, three phase) and the client abnormality degrees attributed to field inspections (Apt6).

Therefore, each attribute provides different kinds of information to the clustering algorithm (or different perspectives of the available information) and complement each other. By means of a cluster-based classification process, one can mutually validate the intrinsic information carried by each variable in order to improve the overall system detection rate.

Practical studies have been performed, which concluded that the minimum size to achieve the expected performance is around five attributes. Notice that the parameters used in the proposed algorithm, such as the month classification interval, the number of attributes, and the clusters, can be easily adjusted to more efficiently capture the peculiarities of other regions or countries.

The number of clusters was based on the system performance with the adopted quality metrics. Tests using an unbalanced training set with a ratio of 10 regular consumers to 1 fraudster have shown that by increasing the number of clusters up to five, higher precision was achieved compared to the 4-clustercase. Therefore, this value was considered appropriate in terms of trade off.

B. Clustering:

The previous step aimed to find out the patterns of consumption for each client. Fuzzy clustering based is done in the initial database in order to detect clients with similar patterns. This task considers the 12 previous months. Therefore, data over a minimum 18-month period should be recorded for each customer prior to the analysis. The algorithm identifies c clusters for n clients, creating the matrix of centres C(c*5) and the matrix of fuzzy membership U0 (n*c). The latter informs the distance between each pattern and the center of each cluster, where a line represents the standard profile of each client.

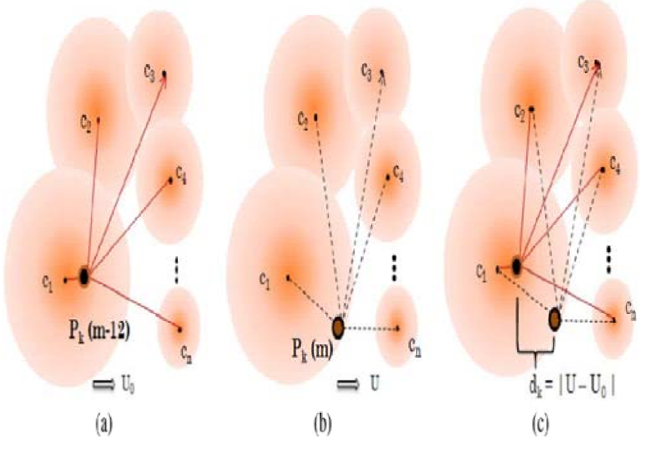


Fig.3. Classification process. (a) Clustering (month m-12), yielding a vector U0 . (b) Current profiles and the current membership vector U. (c) Euclidean distance between the current profile and the last one

C. Classification:

The classification process begins when a standard profile is defined. All in all, the classification process determines the chances of a client having an abnormality. The attributes are calculated regarding the present date (i.e., considering the last 12 months after the clustering process).

So the classification process calculates the current profile of each client in two steps: 1) the current fuzzy membership matrix [(5)] is calculated considering the current profiles and the centres found out in the clustering process. Thereafter, the Euclidean distance between the current fuzzy membership matrix and the last one is obtained, yielding an index that represents the abnormality degree of each client. Finally, the abnormality degree is normalized.

All clients with indices above some given values are considered suspects of irregularities. In this paper, a threshold equal to 0.7 was adopted based on practical tests, which have shown that this value leads to a typical hit rate of 80% (80% irregular customers).

The whole classification process can be seen in Fig. 3. In short, the whole process involves: 1) clustering of the initial database creating the center and the identification of the last fuzzy membership matrix [Fig. 3(a)]; 2) evaluation of the same client in the current month and creates the current membership matrix [Fig. 3(b)]; 3) calculating the Euclidean distance between u and u0, resulting in the client abnormality degree [Fig. 3(c)]

V. CONCLUSION

This paper proposed a methodology based on a fuzzy clustering technique for classifying abnormalities in the profiles of energy consumption. The approach was validated with real data, showing satisfactory performance in the identification of irregularities in residential and commercial customers. Additional tests have been performed with various database distributions at various cut levels and considering several horizons analysis. Rather than a simple task classification process, this problem involves the estimation of relevance in groups, where the degree of membership modeled by fuzzy logic was used, proving to be the mathematical framework ideal for dealing with approximated information.

The applicability of the method is highly feasible because it requires only five variables derived from data commonly available in most distribution utilities. In addition, because the proposed algorithm is unsupervised and, therefore, does not depend on rules, it can be applied in any distribution utility. The formulation of rules covering all possible cases of abnormalities in order to maximize the assertiveness is an unfeasible task, to say the least.

The variables considered in the system are also useful in pre analysis tasks and for spotting certain kinds of hardly detected fraud practices, such as meters having been tampered with from the start of service. The client's average neighbour consumption (MS6), for example, is indeed useful information to detect such types of abnormalities. Almost all electric utilities focus initial investigative attention on customers exhibiting sudden consumption increases or flagged by critical field remarks, such as a broken, tampered fake meter seal, jumpers, and others. A straightforward comparison between attributes MS6 and M6 could also indicate whether a consumer pattern would be compatible with clients having similar consumption categories along their neighbourhoods.

One should emphasize the need for greater integration and cooperation between research centres of fraud in power systems. Indeed, few studies have addressed similar topics.

Also, it is important to focus on issues such as the prediction of financial profit from inspections.

Electricity theft, a common form of commercial losses, involves tampering with meters to distort the billing information or direct connections to the power system. Commercial losses are nearly impossible to measure using traditional power system analysis tools. This is due to the lack of information on both commercial and the legitimate loads in the system, which translates to insufficient inputs for any meaningful loss calculations. Despite the best efforts by utilities, the current results of commercial losses measurements are often inaccurate at best, because the figures rely heavily on the records of detected cases, rather than by actual measurement of the electrical power system. Certainly utilities have some control over the magnitude of commercial losses; but even with their best efforts, some commercial losses will still continue.

VI. REFERENCES

[1]. Federal Court of Audit, "Operational audit report held in National Agency of Electrical Energy, ANEEL, Brazil" Brasília, Brazil, Rep. no. TC 025.619/2007-2, 2007.

[2]. R. J. Bolton and D. J. H, "Statistical fraud detection: A review," Statist. Sci., vol. 17, 2002.

[3]. Y. Kou, C. T. Lu, S. S. , and Y. P. Huang, "Survey of fraud detection techniques," in Proc. IEEE Int. Conf. Netw., Sensing Control, 2004, pp. 21–23.

[4]. A. Nizar, Z. Y. Dong, and P. Zhang, "Detection rules for non technical losses analysis in power utilities," in Proc. IEEE PowerEnergy Soc. Gen. Meeting, 2008, pp. 1–8.

[5]. A. Nizar, Z. Dong, M. Jalaluddin, and M. Raffles, "Load profiling method in detecting non-technical loss activities in a power utility," in Proc. IEEE Int. Power Energy Conf., 2006, pp. 82–87.

[6]. C. Patricio, L. Collazos, J. Pinto, J. C. Junior, E. M. Martins, J. R. Filho, and F. Guimarães, "A tool for pre-selection of power consumers to verify the existence of fraud based on rough sets," presented at the XVI Braz. Autom. Congr., Salvador, Brazil, 2006.

[7]. E. W. Dos-Angelos, O. R. Saavedra, and O. A. C. Cortes, "Intelligent system for identifying frauds in electric power systems based on fuzzy logic," presented at the VII Brazilian Symp. Intell. Autom., Florianópolis, Brazil, 2007.

[8]. J. Nagi, K. Yap, S. K. Tiong, S. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines," IEEE Trans. Power Del., vol. 25, no. 2, pp. 1162–1171, Apr. 2010.

[9]. A. Nizar, Z. Dong, and Y.Wang, "Power utility nontechnical loss analysis with extreme learning machine method," IEEE Trans. Power Syst., vol. 23, no. 3, pp. 946–955, Aug. 2008.

[10]. J. E. Cabral, E. M. Gontijo, J. O. P. Pinto, and J. R. Filho, "Fraud detection in electrical energy consumers using rough sets," in Proc. IEEE Int. Conf. Syst., Man Cybern., 2004, pp. 3625–3629.

[11]. J. Nagi, A. Mohammad, K. Yap, S. Tiong, and S. Ahmed, "Non-technical loss analysis for detection of electricity theft using support vector machines," in Proc. Power Energy Conf., 2008, pp. 907–912.

[12]. F. Fabris, L. Margoto, and F. Varejao, "Novel approaches for detecting frauds in energy consumption," in Proc. 3rd Int. Conf. Netw. Syst. Security, Oct. 2009, pp. 546–551.

[13]. J. Cabral, J. Pinto, and A. Pinto, "Fraud detection system for high and low voltage electricity consumers based on data mining," in Proc. Power Energy Soc. Gen. Meeting, Jul. 2009, pp. 1–5.

[14]. C. Barbosa, M. Vellasco, M. Pacheco, R. Tanscheit, J. Carrilho, Junior, K. Figueiredo, and J. Rocha, "A methodology for detection of irregularities and prevention commercial losses," presented at the XVII Nat. Seminar Electric Energy Distribution, Belo Horizonte, Brazil, 2006.

[15]. IEA, IEA Electricity Statistics. 2004. accessed May 15, 2007. [Online]. Available: asp?PRODUCT=Electricity/Heat

[16]. ABRADEE, "Reflections of insolvency in the area of electric energy," presented at the 5th Seminar Legal Issues, (in Portuguese) Goiânia, Brazil, 2004. Juridicos/apreslcgvseminariojuridicodaabradeeinadimplenc iagoiaslcg. ppt

[17]. U. Fayyad, G. Piatetsky-shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Mag., vol. 17, pp. 37–54, 1996.