Load Profiling and Data Mining Techniques in Electricity Deregulated Market

A. H. Nizar, Associate Member, IEEE Z. Y. Dong, Member, IEEE J. H. Zhao, Student Member, IEEE

***Abstract* — This paper presents load profiles of electricity customers, using the knowledge discovery in databases (KDD) procedure, a data mining technique, to determine the load profiles for different types of customers. In this paper, the current load profiling methods are compared using data mining techniques, by analysing and evaluating these classification techniques. The objective of this study is to determine the best load profiling methods and data mining techniques to classify, detect and predict non-technical losses in the distribution sector, due to faulty metering and billing errors, as well as to gather knowledge on customer behaviour and preferences so as to gain a competitive advantage in the deregulated market. This paper focuses mainly on the comparative analysis of the classification techniques selected; a forthcoming paper will focus on the detection and prediction methods.**

***Index Terms* — Classification, Clustering, Customer Information System, Data Mining, Deregulation, Load Profiles, Non-technical Losses.**

I. Introduction

RESTRUCTURING of electrical supply has begun. Electrical industry deregulation has been advancing in many countries, bringing with it enormous changes in the ways that distribution utilities handle information within their databases on customer information, metering and billing systems. The impact of deregulation will affect the way power utilities handle and manage their business. Utility companies, formerly vertically integrated under one umbrella, will become unbundled, with greater challenges facing them.

It is common practice for power utility companies to record customer data, such as administrative facts, contractual data, billing procedures and consumption recordings, in various databases to support their billing activity. In the deregulated electricity market, information on customer consumption patterns, as well as their payment transaction patterns, is becoming critical for distribution companies. Buried with this vast amount of data are all sorts of information that could make a significant difference to the ways in which power utilities run their business and interact with their current and prospective customers to gain an edge over their competitors. However, the necessary information that exists within the company are too fragmented and complex for a human mind to base efficient conclusions upon. In addition, it is too inaccessible and time consuming to gather, because the information required to make strategic and timely decisions is hidden in complex database systems.

At the same time, power utilities are also suffering from electricity theft because a power system can never be 100% secure from it. Electricity theft can be in the form of fraud (meter tampering), stealing electricity (illegal connections), billing irregularities and unpaid bills[1]. The financial losses resulting from this electricity theft are critical to many electric power organizations. Lost earnings can result in lack of profits, shortage of funds for investment in power system capacity improvement, and the necessity to expand generating capacity to cope with power losses [1].

Besides the problem of non-technical losses, the worldwide deregulation of energy requires power utilities to pay competitive attention to customers, treating them individually and not just over the long term. In order to differentiate themselves from their competitors, power utilities need to segment the market and target the segments with the most effective types of marketing methods. This benefits not only in attracting new customers into the system but also in keeping existing customers from moving to their competitors.

In this paper, we examine current load profiling methods and data mining techniques used by power utilities. The selected methods are compared, analysed and evaluated. The objective of this study is to determine the best load profiling methods and data mining techniques to be used in order to classify and detect non-technical losses in the distribution sector due to faulty meter and billing errors, as well as to gather knowledge on customer behaviour and preferences to gain a competitive advantage as retailers in a deregulated market. For this examination, it is necessary first of all to have a comprehensive understanding of the load profiling methods and data mining techniques used in the power industry.

So this paper reviews the load profiling and data mining techniques in the electricity market to classify customer behaviours, and uses the knowledge gathered to detect and predict non-technical losses due to faulty metering and billing errors.

The paper is organized into sections as follows. A review on load profiling and data mining techniques currently used is presented in the remainder of Section I. The methodological frameworks of load profiling and data mining are elaborated in Section II. Section III describes a case study using company data to demonstrate classification and detection based on the knowledge derived. Section IV provides results and discussion based on the case study. And finally, section V ends the whole paper with the conclusions.

A. H. Nizar is with School of Information Technology and Electrical Engineering, The University of Queensland, St. Lucia, 4072 QLD, Australia (e-mail:nizar@itee.uq.edu.au)

Z. Y. Dong is with School of Information Technology and Electrical Engineering, The University of Queensland, St. Lucia, 4072 QLD, Australia (e-mail:zdong@itee.uq.edu.au)

J. H. Zhao is with School of Information Technology and Electrical Engineering, The University of Queensland, St. Lucia, 4072 QLD, Australia (e-mail:zhao@itee.uq.edu.au)

A. Load profiling and data mining review

Many researches are still in progress to investigate the tools to classify and characterize the behaviour of electricity customers. Load profiling has been identified as one suitable method of dealing with customers without time interval metering equipment. Besides, load profiles also serve as a tool for distribution companies to improve their market strategies and offer new services, as well as to develop new tariffs [2], [3] in the deregulated market. Many different techniques, ranging from conventional to more sophisticated methods [4], [2], [5], have been used for load profile modelling .

Load profiles were used prior to deregulation for many years for tariff formulation. However, since deregulation, the pressure is intensifying and the need for load profiles of electricity customers is becoming crucial, since the knowledge gathered from the load profiling can be used for various purposes.

The load profile has been described as a pattern of electricity usage behaviour of a customer, or a group of customers, over a given period [6]. The process of load profiling is divided into two stages: determining customer classes over a given period, and then allocating the identified customers into each particular category. Basically the model of load profiling can be based on geographical location or categorical type; however this can also depend on the various factors contributing to the study.

The remainder of this section reviews the techniques used in a few studies in determining the customer classes in an electricity market. This includes studies in Taiwan, Portugal, and the UK.

In [3], a study was conducted using load profiling, with an automatic clustering procedure, to characterize customer load patterns and to compute the margins left to a distribution company for fixing dedicated tariffs for different classes of customers. In addition to that, a study [7] to develop dedicated market strategies was also conducted through classification.

In Taiwan, a load survey system has been implemented since 1993 to determine various customer classes of load characteristics [4]. This system served various functions, such as system planning [8], system operation and maintenance, load management [9], rate tariff structure and marketing [2]. This study used statistical analysis to derive the load pattern of each customer pattern and to represent the load behaviour [5].

In Portugal, meanwhile, various data mining techniques have been used to develop electricity customer characterization. Several algorithms were tested on different clustering operations [10]. The selection of the best algorithm was based on the comparative analysis of the performance, using two measures of adequacy: a measure of cluster compactness (MIA) and a measure of cluster separation (CDI).

Most of the research done was to capture the needs arising from deregulation; however there has been very little focus on detecting and predicting non-technical losses in power utilities. Moreover, during the data pre-processing phase, most abnormal data is removed or replaced, using available methods; however, in this study, the abnormal data is going to be treated as outliers, in order to discover abnormalities or irregularities.

B. Objectives of the study

This paper presents electricity customer load profiles based on a knowledge discovery in databases (KDD) procedure, supported by data mining techniques. There are a few studies [11], [12], [13] that have been done in load profiling using data mining techniques to determine the load profiles for different types of customer. These studies were using a combination of unsupervised and supervised learning techniques in customer classification. Both the classification and the clustering methods used have been described to produce different kind of output in order to overcome problems in electricity market. Therefore, this study also embarked into the same techniques that have been used with the objectives to classify and predict customers that potentially will perform non-technical loss activities.

This study used customer historical data collected from a utility company. The objective of this study is to extract knowledge on customer consumption patterns, represented by their load profiles, where the load profiles correspond to customer classes. This knowledge is useful for the power utility to detect any irregularities in customer consumption, to detect non-technical losses due to faulty metering and billing errors. Knowledge of electricity customer behaviour can be extracted from the historical data stored by utility companies, and is represented as load profiles.

Some research work on load profiles in the power industry has already been done for many years, for demand management [14], load management [2], in developing new tariff designs [15], to improve market strategies [16], and as an alternative to metering equipment for small consumers [17]. However, little attention has been paid so far to investigating the load profiling techniques, using customer information systems (CIS) that Utilities Companies have, in order to detect non-technical losses. Some previous studies used data mining techniques to determine customer characteristics, such as [18], [19], [20].

Recently, there are also some research conducted to detect fraud using data mining techniques in electricity market [21], [22] and [23]. However, these studies used different methods to detect the electricity fraud in the power utility. Even though [21] and [22] used decision trees and wavelet based feature extraction and multiple classifiers respectively for detection, our study will focus more on load profiling method to detect the non-technical loss activities in electricity industry.

The next section highlights the methodology used in classifying customers into classes, based on the different factors identified. A significant example of load profiles deviation between public holiday, week day and week end for industry customer is presented.

II. METHODOLOGY

A. Methodology Framework

The data mining techniques used in this study are adopted from [24]. A description of the framework is illustrated in Figure 1. Basically, the framework is divided into 5 phases: problem statement, data collection, data pre-processing, data mining and knowledge representation. In the first phase, the problem identified in this study is to classify the customer electrical behaviour and infer the knowledge to be gathered for future consumption, so that any irregularities due to faulty metering and billing errors can be detected. In the second phase, data is collected from TNB from various categories of customer types: residential, commercial, industrial and mining. The third phase, data pre-processing, contains 3 sub-phases, not necessarily in order: data cleaning, data transformation and data integration. The fourth, data mining, phase concentrates on the load profiling module that classifies customers according to their behaviour and the classification module that focuses more on using the knowledge gathered from load profiling module to infer customers' future consumption. Finally, the last phase is knowledge representation, which concentrates on presenting and interpreting the results derived from the study.

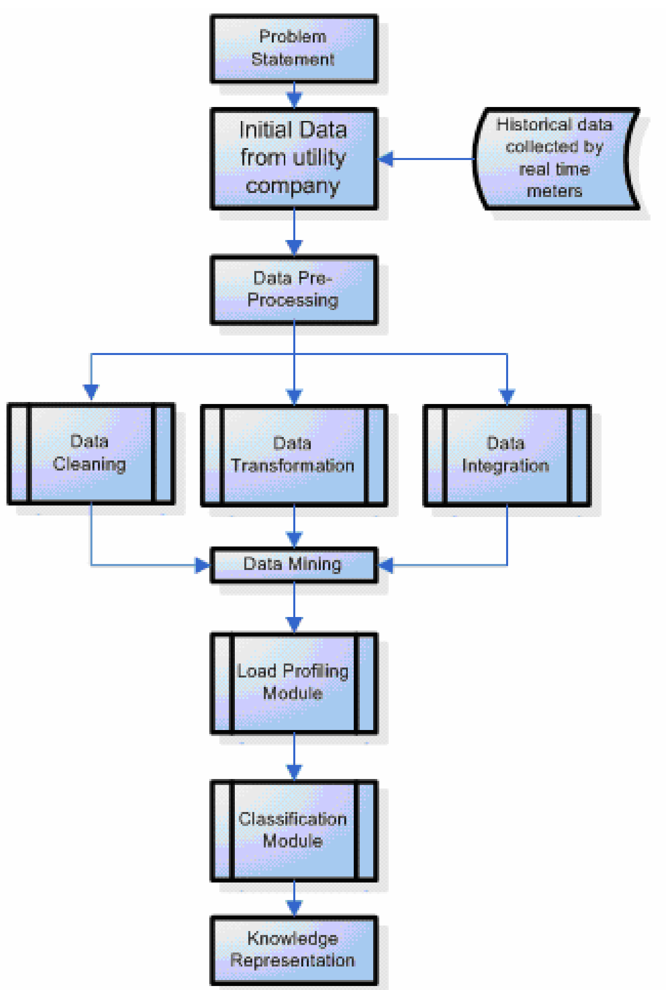

Fig. 1. Methodology Framework [24]

The knowledge of the electricity customer behaviour can be extracted from the historical data stored by utility companies and is represented through load profiles. This study is going to classify the customer behaviour using clustering techniques.

B. Clustering framework

The clustering module's goal is to partition the initial data sample into a set of classes defined according to the load shape of the representative load diagrams of each customer [11]. This study makes comparisons on three clustering methods, which are, k-Means, Cobweb and EM method on the same dataset.

There are few stages involves in the clustering method which are illustrated in the following diagram.

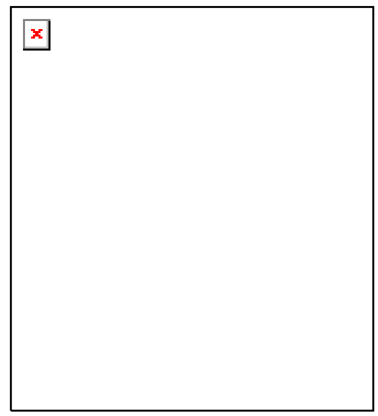

Fig. 2. Load Profiling Framework[24]

C. Classification framework

The classification module goal is to used the knowledge extracted from clustering module and used it to predict the electricity customers whether they will attempt to do any non-technical losses.

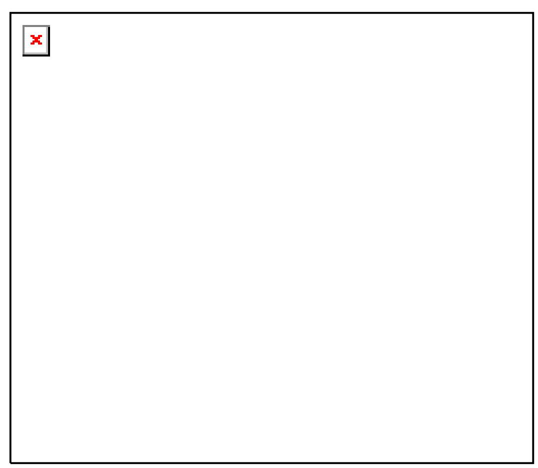

Fig. 3. Classification Framework

III. CASE STUDY ON UTILITY'S CUSTOMERS DATA

This study used customers historical data collected from a utility company. The objective of this study is to extract knowledge on customer consumption patterns, which will be represented by their load profiles where each load profiles represent customer classes. This knowledge is useful for the utility company to detect any irregularities in customer consumption to detect non-technical losses due to faulty metering and billing errors. The study will be divided into two sections where the first section concentrates on clustering module and the second section highlight on classification module.

The data on 200 customers' consumption was collected for duration of period of 6 months including the weekday and week end with the interval of 30 minutes recorded by real times meters on site which gives 48.values a day for each customer for each day. Even though Malaysia is having tropical climate with warm weather throughout the year, but a study to distinguish between hot and rainy seasons can be done to find any significant changes in electricity consumption. Besides that, this study also considered type of customers, customers' activity type as well as the voltage level.

The following table list the various factors involve in determining the customer classes from the dataset.

TABLE 1
FACTORS INVOLVE INDETERMINING THE CLASSES

Factors	Variables
Time of day	Weekday and weekend
Day of the week	Off peak and peak hour
Weather	Hot and rainy
Geographical area	City, suburban and urban
Categorical type	Domestic, commercial, industrial and mining
Voltage level	Low voltage, medium voltage and high voltage

The following figures shows comparison on daily load profiles on public holiday, weekend and week day for an industrial customer.

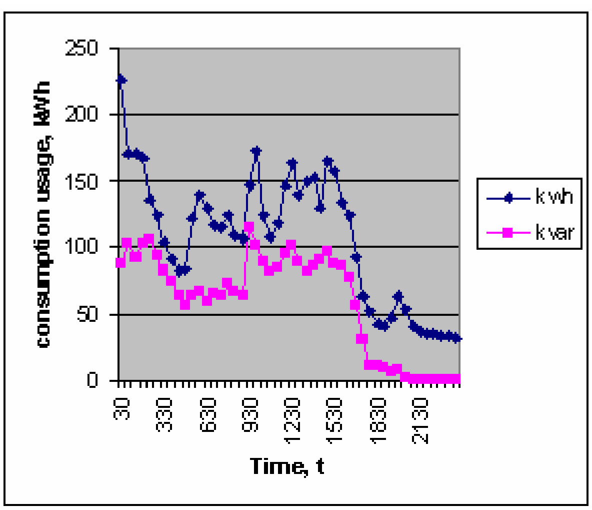

Fig. 4. Load profiles on public holiday

From the above figure, it is shown that the consumption usage is significantly dropped during evening.

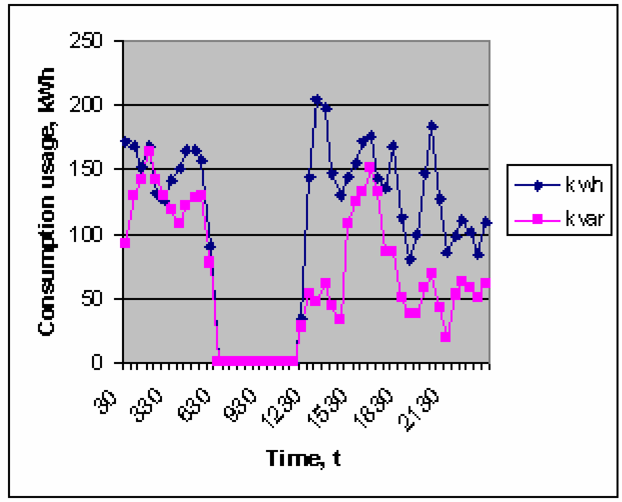

Fig. 5. Load profiles on week end

From the above figure, it is significantly shown that there is no activity conducted in the morning until afternoon on the week end.

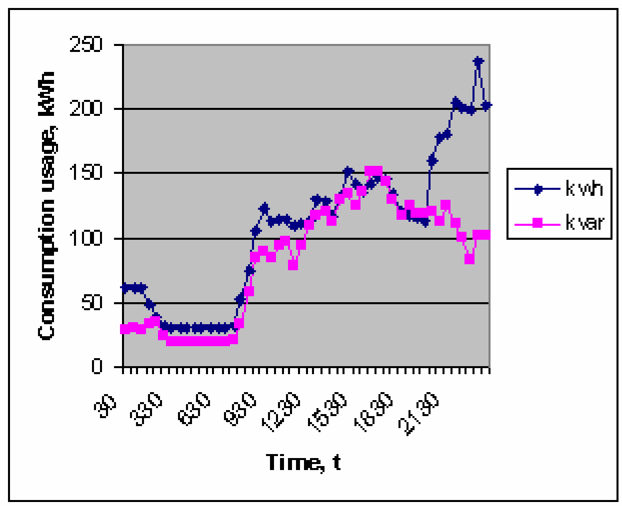

Fig. 6. Load profiles on weekday

From the above figure, it is shown that the consumption usage increased from afternoon until midnight.

Based on the identified factors, the consumption usage for these particular customers' classes will be classified and monitored to see any irregularities or abnormalities. In addition to this, the power utility company can take further action to investigate further on these abnormalities. Moreover, this information can assist power utility in terms of identifying customers to be inspected on site. This effort will significantly reduce the on-site inspection cost to the power utility.

A. Clustering Module

This paper makes comparative analysis on three clustering methods for the purpose on classifying the electricity customers into several groups of clusters. Clustering technique is also known as unsupervised learning because there is no class to be predicted.

1. K-means method

The k-Means is a classic clustering techniques based on iterative distance-based clustering. The k-means algorithm takes the input parameter, k, and partitions a set of n objects into k clusters so that the resulting intra-cluster similarity is high but the inter-cluster similarity is low, which can be measured in regard to the mean value of the objects in a cluster known as cluster's centre of gravity [24]. The objective of this comparison is to determine whether the mean of the set of samples is significantly greater than or significantly less than the mean of another [25]

2. EM method

The EM method is a probability based clustering with the goal to find the most likely set of clusters given the data. This is done where the procedure used for the k-means clustering algorithm and iterate are adopted[24]. This method calculates the cluster probabilities and distribution parameters [25]

3. COBWEB method

The COBWEB method is an incremental clustering where a tree is formed at any stage with instances at the leaves and root node that represents the entire dataset. The tree will consist of the root at the beginning stage and then instances are added one by one until the tree is updated appropriately at each stage and used heuristic evaluation measure called category utility to guide the construction of the tree[24].

IV. Results and Discussions

The following tables present the results gathered from each clustering technique: simple k-means, COBWEB and EM method.

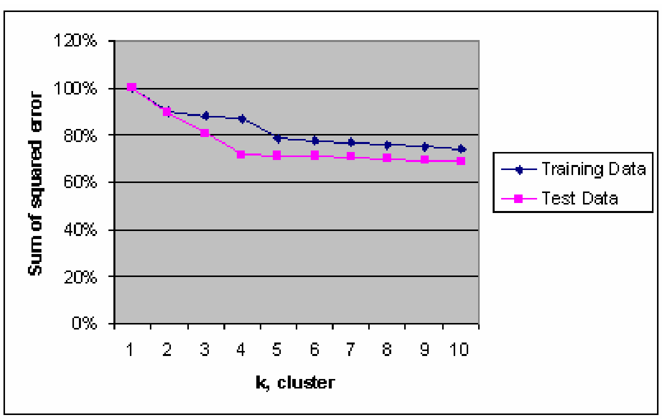

Fig. 7. Results on the Simple-k-means method on number of clusters

The objective of this k-means test is to choose the best cluster center to be the centroid [24], with a minimum total squared distance from each point to its center, once the iteration has stabilized. From the results gathered, k=4 is the best point chosen in this study. Results from the figure 7 and figure 8 have proved the objective of k-means method after being tested on the dataset collected.

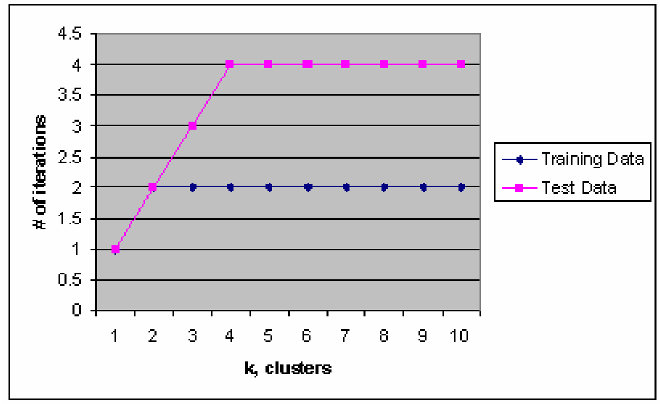

Fig. 8. Results on Simple K-Means for number of iterations

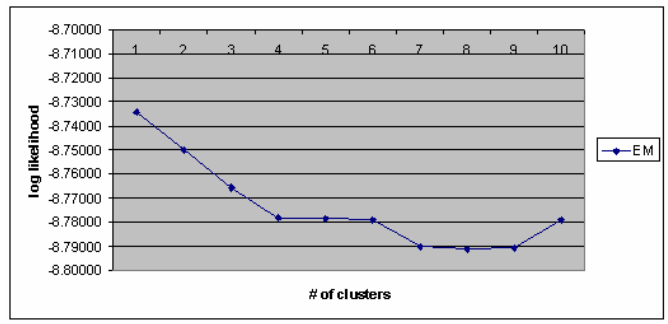

Fig. 9. Results on EM

The log likelihood for the EM test is decreased as the number of clusters increased. The objective of EM is to compare the different final log-likelihood.

TABLE 2
RESULTS ON COBWEB

Acuity	#merges	#split	#clusters
1	1461	1399	6233
2	1461	1399	6233
3	1461	1399	6233

From the table 2, it is shown that the COBWEB method is more stable and consistent on each iteration made.

V. Conclusions

From the results gathered, it is shown that the three clustering techniques gave different results in terms of number of clusters However, the performance on each technique also gives significant impact in terms of cost. The performance time for COBWEB is the longest among all these techniques. While Simple K-means technique performs better in these datasets. Based on the data set available, the electricity customers were classified according to the time of the day factors and the consumption usage characteristics. It is shown that there are significant differences occurred among these three categories. Therefore, in future, the irregularities or abnormalities in this consumption pattern will trigger an alert to the system.

VI. References

[1] T. B. Smith, "Electricity theft: a comparative analysis," *Energy Policy*, vol. 32, pp. 2067-76, 2004.

[2] C. S. Chen, J. C. Hwang, and C. W. Huang, "Application of load survey systems to proper tariff design," *Power Systems, IEEE Transactions on*, vol. 12, pp. 1746-1751, 1997.

[3] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," *Power Systems, IEEE Transactions on*, vol. 18, pp. 381-387, 2003.

[4] C. S. Chen, M. S. Kang, J. C. Hwang, and C. W. Huang, "Implementation of the load survey system in Taipower," presented at Transmission and Distribution Conference, 1999 IEEE, 1999.

[5] C. S. Chen, J. C. Hwang, Y. M. Tzeng, C. W. Huang, and M. Y. Cho, "Determination of customer load characteristics by load survey system at Taipower," *Power Delivery, IEEE Transactions on*, vol. 11, pp. 1430-1436, 1996.

[6] Z. H. Zakaria and K. L. Lo, "Load profiling in the new electricity market," presented at Research and Development, 2002. SCOReD 2002. Student Conference on, 2002.

[7] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Electric energy customer characterisation for developing dedicated market strategies," presented at Power Tech Proceedings, 2001 IEEE Porto, 2001.

[8] C. S. Chen, M. S. Kang, J. C. Hwang, and C. W. Huang, "Temperature effect to distribution system load profiles and feeder losses," *Power Systems, IEEE Transactions on*, vol. 16, pp. 916-921, 2001.

[9] C. S. Chen, M. S. Kang, J. C. Hwang, and C. W. Huang, "Temperature adaptive switching operation for distribution systems," *Power Delivery, IEEE Transactions on*, vol. 16, pp. 694-699, 2001.

[10] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," *Power Systems, IEEE Transactions on*, vol. 20, pp. 596-602, 2005.

[11] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," *IEEE Transactions on Power Systems*, vol. 20, pp. 596-602, 2005.

[12] V. Figueiredo, F. J. Duarte, F. Rodrigues, Z. Vale, C. Ramos, S. Ramos, and J. B. Gouveia, "Determination of Electricity Consumers' Load Profile."

[13] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Emergent electricity customer classification," *Generation, Transmission and Distribution, IEE Proceedings-*, vol. 152, pp. 164-172, 2005.

[14] R. F. Chang and C. N. Lu, "Load profiling and its applications in power market," presented at Power Engineering Society General Meeting, 2003, IEEE, 2003.

[15] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "An approach to customers daily load profile determination," presented at 2002 IEEE Power Engineering Society Summer Meeting, Jul 21-25 2002, Chicago, IL, United States, 2002.

[16] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "A methodology to classify distribution load profiles," presented at Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES, 2002.

[17] M. Kitayama, R. Matsubara, and Y. Izui, "Data mining for customer load profile analysis," presented at Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES, 2002.

[18] F. J. Duarte, F. Rodrigues, V. Figueiredo, Z. Vale, and M. Cordeiro, "Data mining techniques applied to electric energy consumers characterization," presented at Artificial Intelligence and Soft Computing, 14-16 July 2003, Banff, Alta., Canada, 2003.

[19] M. Kitayama, R. Matsubara, and Y. Izui, "Application of data mining to customer profile analysis in the power electric industry," presented at Proceedings of Winter Meeting of the Power Engineering Society, 27-31 Jan. 2002, New York, NY, USA, 2002.

[20] B. D. Pitt and D. S. Kitschen, "Application of data mining techniques to load profiling," presented at Power Industry Computer Applications, 1999. PICA '99. Proceedings of the 21st 1999 IEEE International Conference, 1999.

[21] J. R. Filho, E. M. Gontijo, A. C. Delaiba, E. Mazina, J. E. Cabral, and J. O. P. Pinto, "Fraud identification in electricity company customers using decision tree," presented at Systems, Man and Cybernetics, 2004 IEEE International Conference on, 2004.

[22] R. Jiang, H. Tagaris, A. Lachsz, and M. Jeffrey, "Wavelet based feature extraction and multiple classifiers for electricity fraud detection," presented at Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES, 2002.

[23] R. Mano, R. Cespedes, and D. Maia, "Protecting revenue in the distribution industry: a new approach with the revenue assurance and audit process," presented at Transmission and Distribution Conference and Exposition: Latin America, 2004 IEEE/PES, 2004.

[24] J. Han and M. Kamber, *Data mining : concepts and techniques*. San Francisco, Calif. :: Morgan Kaufmann Publishers,, 2001.

[25] I. H. Witten and E. Frank, *Data mining : practical machine learning tools and techniques*, 2nd ed. ed. San Francisco, Calif. :: Morgan Kaufman,, 2005.

VII. BIOGRAPHIES

Anisah Hanim Nizar currently is a PhD student at School of Information Technology and Electrical Engineering, University of Queensland. Her employment experience included the TNB Research Sdn. Bhd. in Malaysia since 1998 and she has special field of interest in Data Management Technology and Information Systems in electricity market. She received bachelor degree with honours from University Utara Malaysia and her Master degree from University Tenaga Nasional in Malaysia.

Zhao Yang Dong (M'99) obtained his PhD in Electrical and Information Engineering from The University of Sydney, Australia in 1999. He is now a senior lecturer at the School of Information Technology and Electrical Engineering, The University of Queensland, Australia. His research interest includes power system's security assessment and enhancement, electricity market, artificial intelligence and its application in electric power engineering, power system planning and management, and power system stability & control.