Evaluating Different Clustering Techniques for Electricity Customer Classification

S. M. Bidoki, N. Mahmoudi-Kohan, *Student Member, IEEE*, M. H. Sadreddini, M. Zolghadri Jahromi, M. P. Moghaddam, *Member, IEEE,*

***Abstract*- In the electricity market, it is highly desirable for suppliers to know the electricity consumption behavior of their customers, in order to provide them with satisfactory services with the minimum cost. Information on customers' consumption pattern in the deregulated power system is becoming critical for distribution companies. One of the suitable tools for extracting characteristics of customers is the clustering technique. Selection of better methods among several existing clustering methods should be considered. Therefore, in this paper, we evaluate the performance of Classical K-Means, Weighted Fuzzy Average K-Means, Modified Follow the Leader, Self-Organizing Maps and Hierarchical algorithms that are more applicable in clustering load curves. The performances were compared by using two adequacy measures named Clustering Dispersion Indicator and Mean Index Adequacy.**

***Index Terms*- Classical K-means, clustering dispersion indicator, clustering technique, electrical load curves, hierarchical algorithms, modified follow-the-leader, self-organizing maps, weighted fuzzy average K-means.**

I. Introduction

CLUSTER analysis is very useful when classification formation pertaining to the data is not available. The main aim of cluster analysis is to find pattern associations by forming groups of patterns such that a pattern in a group is more similar to other patterns in the same group when compared to patterns in other groups. Many clustering approaches have been proposed in the literature to suit various requirements [1].

Clustering techniques have been recently used in clustering load curves [1-8, 10]. These techniques are suitable in defining typical load profile (TLP) of customers. Different applications are available for classification of the load curve of customers. Accurate knowledge of the customers' consumption patterns represents a worthwhile asset for electricity providers in the competitive electricity markets [2]. Classification of loads in terms of their time-varying power consuming behavior is an important task for load forecasting, load data processing, locational customer services, and power system analysis [3].

With the electricity market liberalization, the electricity distribution business looks for better market strategies based on adequate information about the consumption patterns of the electricity customers. A fair insight into the customer's consumption behavior allows the distribution utilities to better address the operation of the distribution infrastructure and its future enhancement, not to mention the ability to design specific tariff options for the various classes of customers in tune with real operation costs [4].

Different methods such as K-Means, Modified Follow the Leader, Self-Organizing Maps, etc. are popular in clustering load curves. One of the important points in the clustering process is the selection of the best method among applicable techniques. For this purpose, there are several adequacy measures such as Mean Index Adequacy (MIA), Clustering Dispersion Indicator (CDI), Least Squares Error, etc.

In this paper, we evaluate and compare the performance of Classical K-Means, Weighted Fuzzy Average (WFA) K-Means, Modified Follow the Leader (MFTL), Self-Organizing Maps (SOM) and Hierarchical algorithms that are more applicable in clustering load curves. For this purpose, MIA and CDI are used as the suitable metrics to evaluate different clustering techniques.

The rest of the paper is organized as follows. Section II describes the clustering background and the adequacy measures. In section III, the clustering methods are discussed. Section IV addresses experimental results and in section V, the conclusions are highlighted.

II. Clustering Technique

A. Clustering process

Classification of customers is obtained by applying a pattern recognition methodology which includes the following steps:

1. Data selection: The first step is the selection of the data. This selection is made according to the voltage level of the consumers by sampling from the customer's load value in some defined time intervals in a period of time. Separate studies must be conducted for different voltage levels [5].
2. Customer classification in whole categories: In order to form the customer classes, the whole set of customers can be preliminarily partitioned into macro-categories defined on the basis of global criteria (e.g., residential, nonresidential, lighting, etc.) [6].
3. Data cleaning and preprocessing: In the cleaning phase, inconsistencies in the data are checked and outliers are removed [5]. If necessary, a preliminary

S. M. Bidoki, M. H. Sadreddini, M. Z. Jahromi are with the Department of Electrical & Computer Engineering, Shiraz University, Shiraz, Iran (e-mail: s.m.bidoki@gamil.com, sadredin@shirazu.ac.ir, zjahromi@shirazu.ac.ir).
N. Mahmoudi-Kohan, M. P. Moghaddam are with the Department of Electrical Engineering, Tarbiat Modares University (TMU), Tehran, Iran (e-mail: nadalimk@yahoo.com, parsa@modares.ac.ir).

978-1-4244-6547-7/10/$26.00 © 2010 IEEE

execution of a pattern recognition algorithm is carried out, in order to track bad measurements or network faults. If there is uncorrected data, it will reduce the number of the useful typical data for a constant number of clusters [6].

4. Load curve clustering and selection of typical load pattern for each class: The clustering methods are applied on customers load curves and finally, the typical load profile (TLP) of each class is selected. TLPs are then used for different programs such as demand side management programs, and designing suitable tariffs. Also, in this step clustering algorithms with better performance are selected by some adequacy measures.

B. Adequacy measure

First, we consider the following distance definitions [2]:

1. The distance between two vectors (e.g., between two members $l^{(i)}$ and $l^{(i)}$, of the set $L^{(k)}$)

$$d\left(l^{(i)}, l^{(j)}\right) = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(l_h^{(i)} - l_h^{(j)})^2} \quad (1)$$

Where, H is the number of attributes of each vector.

2. The distance between a representative load curve $r^{(k)}$ and the subset $L^{(k)}$

$$d\left(r^{(k)}, L^{(k)}\right) = \sqrt{\frac{1}{n^{(k)}}\sum_{m=1}^{n^{(k)}} d^2(r^{(k)}, l^{(m)})} \quad (2)$$

Where, $n^{(k)}$ is the number of vectors in the cluster k.

The adequacy measures can be defined as:

a) Mean Index Adequacy: This is defined as the average distance between the objects in the cluster and the cluster center [2]:

$$MIA = \sqrt{\frac{1}{K}\sum_{k=1}^{K} d^2\,(r^{(k)}, L^{(k)})} \quad (3)$$

This metric is based on the Euclidean distances among the representative load diagrams [7].

b) Clustering Dispersion Indicator: Depending on the distance between the load patterns in the same cluster and (inversely) on the distance between the class representative load profiles [2]:

$$CDI = \frac{\sqrt{\frac{1}{K}\sum_{k=1}^{K}[\frac{1}{2n^{(k)}}\sum_{n=1}^{n^{(k)}} d^2(l^{(n)}, L^{(k)})]}}{\sqrt{\frac{1}{2K}\sum_{k=1}^{K} d^2(r^{(k)}, R)}} \quad (4)$$

This metric merges the information about the compactness of the load diagrams belonging to the same class and (inversely) on the inter-distance among the class representative load diagrams [7].

III. Clustering Methods

A. Classical K-means

K-means algorithm was introduced by J.B. MacQueen in 1967 [8, 9]. Given unlabeled P data vectors with N elements, it groups them into K clusters. The data in a cluster will have similar features or characteristics which will be dissimilar from the data in other clusters. It starts with an initial k-means clustering and then iterates to move each representative point to the centroid of each cluster generated by the current representative points, until a local minimum solution is found [8].

B. Weighted Fuzzy Average (WFA) K-means

One of the problems in K-means clustering to be considered is the fact that outliers in a cluster can unduly affect the center of cluster, when it is obtained as the cluster average (by averaging each component). Medians can be used in place of averages, although they may throw away good points as well as outliers. For solving this problem, a type of fuzzy averaging is used that puts the center prototype among the more densely situated points [9, 10].

Let $\{l^{(1)}, \ldots, l^{(p)}\}$ be a set of P real numbers. To find its weighted fuzzy average, this algorithm initially takes the sample mean $\mu(0)$ and variance σ^2 to start the process. We have extended the formulations which are used in [10] by considering L as a vector with H elements that represents the computation of cluster center as following:

$$w_{(p,h)}^{(r)} = exp\left[-\frac{l_h^{(p)} - \mu_{(h)}^{(r)}}{2\sigma^2}\right] \quad for\ h = 1, \ldots, H \quad (5)$$

$$\mu_{(p,h)}^{(r+1)} = \sum_{(p=1,P)} w_{(p,h)}^{(r)} X_{(p,h)} \quad for\ h=1,\ldots,N \quad (6)$$

This algorithm computes σ^2 on each of three or four iterations and then leaves it fixed. After about five iterations the approximate WFA is sufficiently closed to the true WFA.

C. Modified Follow the Leader (MFTL)

This algorithm does not require initialization of the number of clusters and computes the cluster centers automatically. Euclidean metric in this algorithm is modified by introducing for each distinctive feature a weighting factor $\sigma_j^2/\overline{\sigma}^2$, where σ_j^2 is the variance of the *j-th* feature of all the load patterns in the population and $\overline{\sigma}^2$ is the average value of the variance σ_j^2 for $j = 1, \ldots, J$ [2, 11]. If all load patterns have been normalized to the range (0,1), the impact of high-variance features is amplified in the computation of the weighted Euclidean distance

$$d_i\left(y^{(m)}, r_i^{(k)}\right) = \sqrt{\sum_{j=1}^{J} \sigma_j^2/\overline{\sigma}^2 (y_j^{(m)} - r_{ij}^{(k)})^2} \quad (7)$$

Where $r_i^{(k)}$ designates the *k-th* cluster center for the *i-th* cycle of the clustering process $r_{ij}^{(k)}$ and its *j-th* component.

This procedure is composed of several cycles$i = 1, \ldots, I$. The first cycle ($i = 1$) sets the number K of clusters and the number of $n^{(k)}$ patterns belonging to each cluster $k = 1, \ldots, K$ by using a "follow the leader" approach depending on a distance threshold ρ. All load patterns are presented sequentially. If the minimum weighted Euclidean distance from the center of an existing cluster is lower than ρ, each pattern is assigned to the closest cluster q and the cluster center $r_i^{(q)} = \{r_{ij}^{(q)}, j = 1, \ldots, J\}$ is updated; otherwise, the pattern defines the center of a new cluster. The successive cycles ($i < 1$) may change the number of clusters by reassigning the patterns to the closest cluster and updating the cluster centers until the number a_i of patterns changing clusters in a single cycle is zero. At the *i-th* cycle, the *m-th*

customer belongs to the cluster $h_i^{(m)}$. The maximum number of cycles *I* should be set to a number large enough to allow for clustering stabilization before the *I-th* cycle [11].

D. Self-organizing maps (SOM)

The SOM is an effective software tool for the visualization of high-dimensional data. It converts complex, nonlinear statistical relationships between high-dimensional data items into simple geometric relationships on a low-dimensional display [12]. The basic Self-organizing maps can be visualized as a sheet-like neural-network array. The cells of which become specifically tuned to various input signal patterns or classes of patterns in an orderly fashion [13].

Assume that the input samples are described by a real vector $x(t) \in R^n$ where *t* is the index of the sample, or the discrete-time coordinate. Each node *i* in the map contains a model vector, $m_i(t) \in R^n$, which has the same number of elements as the input vectors. The initial values of the components of the model vectors may even be selected at random [12-14].

The self-organizing maps algorithm creates the ordered mapping as a repetition of the following basic tasks:

- An input vector $x(t)$ is compared with all the model vectors $m_i(t)$. The best-matching unit on the map, i.e., the node where the model vector is most similar to the input vector in some metric is identified. This best matching unit is often called the *winner*.
- The model vectors of the winner and a number of its neighboring nodes in the array are changed towards the input vector according to the Hebbian Learning Law.

The basic idea in the SOM learning process is that, for each sample input vector, the model vector of the winner and the nodes in its neighborhood are changed closer to input vector in the input data space [13].

Adaptation of the model vectors in the learning process may take place according to the following equation [12, 14]:

$$\mathrm{m_i(t+1) = m_i(t) + h_{c(x),i}^{(t)} \cdot \big(x(t) - m_i(t)\big)} \quad (8)$$

Here $h_{c(x),i}^{(t)}$ is called the neighborhood function, and it is like a smoothing kernel that is time-variable. It is a decreasing function of the distance between the *i-th* and *c-th* models on the map grid. For convergence it is necessary that $h_{c(x),i}^{(t)} \to 0$ when $t \to \infty$.

Two simple choices for $h_{c(x),i}^{(t)}$ occur frequently. The simpler of them refers to a neighborhood set of array points around winner node *c*, let their index set be denoted $N_c(t)$.

$$\mathrm{h_{c(x),i}^{(t)} = \alpha(t) \quad , if\ } i \in N_c(t) \quad (9)$$

Where the value of $\alpha(t)$ identified with a learning-rate factor ($0< \alpha(t) <1$). Both $\alpha(t)$ and the radius of $N_c(t)$ are usually decreasing monotonically in time. If the *i-th* node is not in neighborhood of *c*, neighborhood function returns 0.

Another widely applied, smoother neighborhood kernel can be written in term of the Gaussian function,

$$h_{c(x),i}^{(t)} = \alpha(t) \cdot exp\left(-\frac{\|r_i - r_c\|^2}{2\sigma^2(t)}\right) \quad (10)$$

Where $\alpha(t)$ is learning-rate factor, and $\sigma(t)$ defines the width of the kernel; Latter corresponds to the radius of $N_c(t)$ above. Both $\alpha(t)$ and $\sigma(t)$ are some monotonically decreasing functions of time [12, 14].

E. Hierarchical algorithm

In hierarchical clustering, there are initially N singleton clusters, as much as the number of represent load patterns (RLPs). At first, an N×N similarity matrix is built using the Euclidean norm distance criterion. It uses an iterative process to grouping N vectors to binary clusters using linkage criterion. The process is iteratively repeated by merging the clusters of each level into bigger ones at the upper level until all samples are grouped in one cluster. The history of the process is kept in order to form a binary tree structure, whose root is the cluster that contains the whole data set [7].

The linkage criterion measures the similarity between clusters at each level and determines the cluster formation at upper level. In this paper, we use average distance criterion that groups two clusters s and t as the follows [7]:

$$\mathrm{d_s(s,t) = \frac{1}{n^{(s)}n^{(t)}} \sum_{i=1}^{n^{(s)}} \sum_{j=1}^{n^{(t)}} d(x_i^{(s)}, x_j^{(t)})} \quad (11)$$

Where, $d_s(s,t)$ is distance between all pairs of vectors in the two clusters, $n^{(s)}$ and $n^{(t)}$are the number of vectors in *s* and *t*, $x_i^{(s)}$ and $x_i^{(t)}$are the generic vectors in *s* and *t*, and *d(.)* represents Euclidean norm distance.

IV. Experimental Results

For analyzing the performances of the above algorithms, a real world distribution network data with detailed information as following is used: these data includes 127 load curves of non-residential (Commercial and Industrial) customers at 400V level including 45 commercial and 82 industrial customers. These data are the 15-min load values for each individual customer for 24 hours in the year 2004. Then each load curve vector has 96 components. We normalized the load vectors to unit length prior to process them.

We implemented the clustering algorithms by using JAVA and MATLAB. Each algorithm was applied on our data set over several runs. The averages of results for different methods are discussed in the following. The represented methods are performed for the clusters number between 2 to 15.

It is notable about SOM that due to lack of input load curves in our data set, we had to use the same inputs iteratively in a cyclic manner in order to train the implemented SOM. The number of training steps considered as the number of the nodes on the map times 500 [14]. Another point is the radius of neighborhood. We initialized it with a value greater than the half of dimension of the map. We let 500 first training steps be ordering phase and after that we decreased the radius of neighborhood by using linear and Gaussian decreasing functions. The initial learning-rate was set to 0.9 and then inverse-time, linear and exponential decreasing functions were used to decrease its value during training steps. The preliminary results show that the combination of Gaussian and linear functions for decreasing radius of neighborhood and learning-rate, leads to better clustering results. Therefore we applied this combination in our further study.

To compare different methods, we used MIA and CDI measures. It should be noted that small values of adequacy measures for each method show better performance of that method. Figure 1 and 2 shows the results of adequacy measures versus number of clusters for different clustering methods.

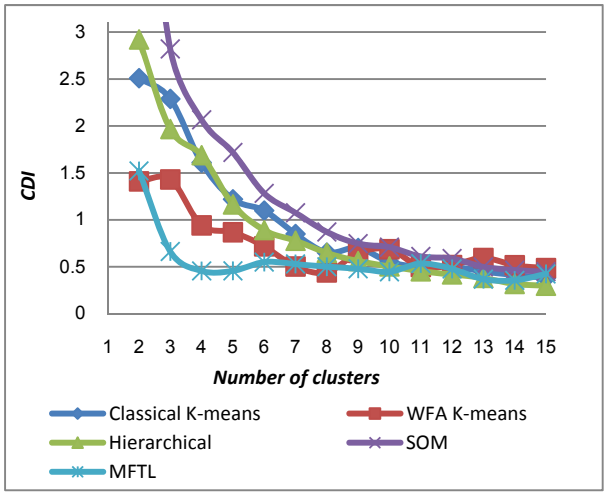

Fig. 1. CDI values for different methods.

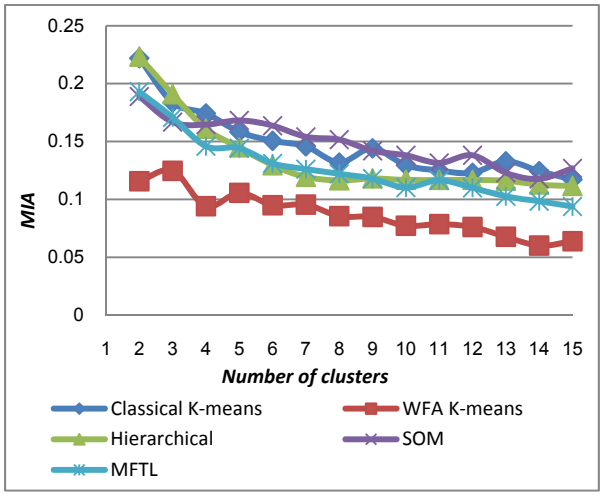

Fig. 2. MIA values for different methods.

As it can be seen in figure 1, the Modified Follow the Leader algorithm has the best Clustering Dispersion Indicator values. Then it can produce more distinctive clusters. In addition, this figure shows that with increasing the number of clusters, the Self-Organizing Maps improves its CDI value and some times get the best result among other methods. However because of its weak throughput when number of clusters is small, it seems not to be a proper algorithm. Then when the purpose of clustering is separating dissimilar input patterns, MFTL algorithm performs well and if the target number of clusters is large enough, SOM algorithm can go well, too.

Figure 2 shows that the Weighted Fuzzy Average K-Means has smallest Mean Index Adequacy values among other methods and the Self-Organizing maps algorithm is in the last place. So when the purpose of clustering is creation of more compact clusters, the best method will be WFA K-Means and the worst is SOM.

As it can be seen in figure 1 and 2, totally WFA K-Means has a reasonable performance that results dense and distinct enough clusters.

V. Conclusions

In this paper, Classical K-Means, Weighted Fuzzy Average K-Means, Modified Follow the Leader, Self-Organizing maps and Hierarchical algorithms that are popular methods in the field of load curve clustering, was discussed. In order to evaluate performance of these clustering algorithms, Mean Index Adequacy (MIA) and Clustering Dispersion Indicator (CDI) were used as adequacy measure.

We used a real world distribution network data in order to analyze the performance of the implemented algorithms. Our data set was included 127 load curves of non-residential customers at 400V level including 45 commercial and 82 industrial customers. These data are the 15-min load values for each individual customer for 24houres in the year 2004.

Results show that the selection of the proper clustering algorithm is somedeal related to the purpose of the clustering. Then, if the clustering is applied for resulting more distinct clusters, Modified Follow the Leader algorithm is the best choice among all evaluated methods. However, if the goal is to have some compact clusters, Weighted Fuzzy Average K-Means seems to be the best algorithm. Generally, Weighted Fuzzy Average K-Means has, on average, better performance than other discussed methods for both purposes.

VI. References

[1] G. Phanendra Babu, M. Narasimha Murty and S. Sathiya Keerthi, "A Stochastic Connectionist Approach for Global Optimization with Application to Pattern Clustering," *IEEE Trans. SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 30, NO. 1,* pp. 10-24 *FEBRUARY 2000.*

[2] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu and C. Toader, "Load Pattern-Based Classification of Electricity Customers," *IEEE Trans. POWER SYSTEMS*, VOL. 19, NO. 2, pp. 1232-1239, MAY 2004

[3] W. Li, J. Zhou, X. Xiong and J. Lu, "A Statistic-Fuzzy Technique for Clustering Load Curves," *IEEE Trans. POWER SYSTEMS,* VOL. 22, NO. 2, pp. 890-891, MAY 2007.

[4] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Electric energy customer characterisation for developing dedicated market strategies," *Power Tech Proceedings, 2001 IEEE Porto Volume 1, Issue , 2001 Page(s):6 pp. vol.1*

[5] V. Figueiredo, F. Rodrigues, Z. Vale and J. B. Gouveia, "An Electric Energy Consumer Characterization Framework Based on Data Mining Techniques," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 596-602, May 2005.

[6] G. J. Tsekouras, N. D. Hatziargyriou and E. N. Dialynas, "Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1120-1128, Aug. 2007.

[7] G. Chicco, R. Napoli and F. Piglione, "Application of Clustering Algorithms and Self Organizing Maps to Classify Electricity Customers," *in Proc. 2003* Vol. 1, Page(s):7 pp.

[8] S. Chunhua, B. Feng, Z. Jianying. T. Tsuyoshi and S. Kouichi, "Privacy-Preserving Two-Party K-Means Clustering via Secure Approximation," in *Proc. 2007 Advanced Information Networking and Applications Workshops 21st International Conf. Vol. 1, pp. 385 – 391.*

[9] S. Nasser, R. Alkhaldi and G. Vert, "A Modified Fuzzy K-means Clustering using Expectation Maximization," in *Proc. 2006 IEEE International Conf., pp. 231-235.*

[10] N. Mahmoudi-Kohan, M. P. Moghaddam, M. K. Sheikh-El-Eslami and S. M. Bidaki, "Improving WFA K-means Technique for Demand Response Programs Applications," accepted for presentation, *IEEE, General Meeting 2009*.
[11] G. Chicco, R. Napoli, P. Postolache, M. Scutariu and C. Toader. "Customer Characterization Options for Improving the Tariff Offer," *IEEE Trans. POWER SYSTEMS, VOL. 18, NO. 1, FEBRUARY 2003*.
[12] T. Kohonen, "SOM Toolbox: Introduction to SOM". Available: *http://www.cis.hut.fi/projects/somtoolbox/theory/somalgorithm.shtml*,
[13] T. Honkela, "Self-Organizing Maps in Natural Language Processing," Thesis for the Degree of Doctor of Philosophy, Helsinki University of Technology, December 1997. Available: *http://www.cis.hut.fi/~tho/thesis/*
[14] T. Kohonen, "Self-Organizing Maps," 3rd Edition, Springer, 2001, pp. 105-138.

VII. BIOGRAPHIES

Seyed Mohammad Bidoki was born in Yazd, May 26, 1985. He received Computer engineering B.Sc. degree from Isfahan University of Technology (IUT), Iran, in 2007. Currently he is M.Sc. student of Software Engineering at Shiraz University, Shiraz, Iran. His main research interests include text mining techniques and applications, pattern clustering and classification and implementing IR systems.

Nadali Mahmoudi Kohan was born in Mehriz, Yazd, Iran, on February 24, 1985. He received Electrical Engineering B.Sc. degree from Isfahan University of Technology (IUT), Isfahan, Iran, in 2007. He is currently M.Sc. student of Electrical engineering at Tarbiat Modares University (TMU), Tehran, Iran. He is also a researcher of Iran Power System Engineering Research Center (IPSERC). His research interests include Demand Response Programs, Electricity market, and distribution network.

Mohammad Hadi Sadreddini is an Associate Professor in the Department of Electrical & Computer Engineering, Shiraz University, Shiraz, Iran. His research interests include Distributed Database Systems and Data Mining.

Mansour Zolghadri Jahromi received Mechanical Engineering B.Sc. degree from Shiraz University, Shiraz, Iran, in 1979, M.Sc. degree in 1982 and Ph.D. degree in 1988 both from University of Bradford, West Yorkshire, England, in Instrumentation & control Engineering. Currently, he is an Associate Professor in the Department of Electrical & Computer Engineering, Shiraz University, Shiraz, Iran. His research interests include Control Systems, Intelligent Control Systems, Instrumentation, Information Retrieval Systems, Computer Networks and Data Communications.

Mohsen Parsa Moghaddam received B.Sc. degree from Sharif University of Technology, Tehran, Iran, in 1980, M.Sc. degree from Toyohashi University of Technology, Japan, in 1985, and Ph.D. degree from Tohoku University, Japan, in 1988 all in Electrical Engineering. Currently, he is an Associate Professor in the Department of Electrical Engineering, Tarbiat Modares University, Tehran, Iran. His research interests include power system planning and control, optimization, and restructuring.