THE OPTIMAL CHOICE OF FUZZINESS PARAMETER IN LOAD PROFILING

NURUL FAKHRI BIN OMAR

Bachelor of Electrical Engineering (Honours)
Faculty of Electrical Engineering
Universiti Teknologi MARA
40450 Shah Alam, Selangor

ABSTRACT

Load profiling is an estimated load shape for a group of customers, which can be developed from historical or current day data. Many techniques for load profiling have been reported in the past. The techniques range from conventional method to more sophisticated ones such as fuzzy c-means algorithm. This paper presents a technique for determining the weighting exponent, m, parameter in a fuzzy c-means algorithm using fuzzy clustering validity indexes. The objective of using cluster validity index is to find the best clustering schemes. In this paper, two fuzzy clustering validity indexes used are Non-fuzzy Index (NFI) and Xie-Beni Index (XB). It will give information on the customers cluster pattern in load profile.

1.0 INTRODUCTION

Electric load in distribution system depends on the consumer demands that vary in accordance to their activities. Power companies must produces sufficient energy when demand to meet the requirements of the consumers [1]. Power companies need a load data for produce energy according demand of consumer. The power companies can fix their electricity production by using load profile information. The data also was needed for tariff and pricing. The Fuzzy c-means algorithm was initially developed by Dunn and later generalized by Bezdek [2]. Fuzzy c-means algorithm has been used to get a good cluster for the load data. A fuzzy goal is defined as maximizing the number of data points in a cluster and a fuzzy constraint as the minimizing of the sum of squared errors within a cluster. A decision about weighting exponent, m, is made by taking the intersection of fuzzy goal and constraint such that given m, the fuzzy c-means algorithm produces 'good' clusters. The range of weighting exponent, m, is $1 \leq m \leq \infty$ [2]. There are no theoretical bases for an optimal choice for m, so in this paper proposed to vary a value of m from 1 to 3.0.

2.0 FUZZY C-MEANS ALGORITHM

2.1 The Algorithm

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters [3], [4]. This method is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m}\left\|x_i - c_j\right\|^2 \qquad 1 \leq m < \infty$$

Where;
N = number of load profile
C = number of cluster
m = weighting parameter, in general, m=2
Uij = is the degree of membership of xi in the cluster j
xi = is the profile ith the feeder of measured data,
cj = jth cluster center
||*|| = is any norm expressing the similarity between any measured data and the centre

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by [3], [4]:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{m}}$$

This iteration will stop when

$\max_{ij} \left\{ \left| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right| \right\} < \varepsilon$, where ε is a termination criterion between 0 and 1, where k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m [3], [4].

2.2 Cluster validity indexes

The cluster validity index is often used to indicate the goodness of a partitioning of a data set as obtained by a clustering algorithm [5]. A good cluster validity index should be independent of the number of clusters, thus allowing comparing two clustering with different number of clusters. At the same time, it is desirable that geneses in one cluster have minimum possible distance to each other and maximum distance to the genesis in other clusters [5]. In other words, the clusters are compact and well separated. Two cluster validity measures that fulfill these criteria are the Non fuzzy index and Xie-Beni index.

2.3 Non-fuzzy index

This index was proposed to measure how fuzzy a C-partition is (C>1) [6]. It can be defined as

$$NFI = \frac{C\left[\sum_{C=1}^{C}\sum_{X} \mu_C^2\right] - N}{N(C-1)}$$

Where N is the number of pattern in the dataset and C is the number of clusters.

The optimal number of clusters is solved by maximizing NFI [6].

2.4 Xie-beni index

Xie-Beni proposed one such validity measure, which is defined as the ratio of the total variation J_m to the minimum separation *sep* of the clusters, where J_m and sep can be written [5] as

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2$$

$$sep = n \min_{i,j} \left\| a_i - a_j \right\|^2$$

J_m = objective function

sep= distance between U

$$XB = \frac{\sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^{m} \left\| x_j - c_i \right\|^2}{n \min_{i,j} \left\| c_i - c_j \right\|^2}$$

Note that, when the partitioning is compact and good, the value of σ should be low, while *sep* should be high. Therefore, the Xie-Beni index (XB) should have a low value when the data has been appropriately clustered [5].

There are several other fuzzy cluster validity indices that are available in the literature [5]. The XB index has been found to be better able to indicate the correct number of partitions in the data for a wide range of the choice of the number of clusters.

3.0 METHODOLOGY

In this paper, several steps were taken to achieve the objective of project. The methodologies of the study involve the following:

1. Research about fuzzy c-means, weighting exponent, *m*, and find the load profiling in one area. In fuzzy c-means, the shape of graph will be changed by varying a number of cluster or the weighting exponent, *m*.
2. Find cluster's data by vary a fuzzy c-means parameter (weighting exponent, *m*) using matlab program. All data were recorded for used in clustering validity index. In matlab program, have fuzzy logic toolboxes are used as a tool for solving problems with fuzzy logic. So, this program is suitable and more easily to run a fuzzy logic algorithm.
3. By using Matlab program, plot a graph load profile in per unit versus hours for one day. This graph used for see pattern of load profiling as peak hour or etc.
4. All data were analyzed with Non-fuzzy index and Xie-beni index. A decision about *m* and cluster is made by

comparing the data on the index. The comparison will produce good cluster and suitable value of *m*.

4.0 RESULT AND DISCUSSION

4.1 Non-fuzzy index (NFI)

Data from matlab programming were analyzed and values of *m* against sum of cluster have been obtained. Below is a table of the values;

	m=1.5	m=1.8	m=2.0	m=2.5	m=3.0
C=2	0.7269	0.5578	0.4649	0.2966	0.1949
C=3	0.8158	0.6478	0.5439	0.3456	0.2218
C=4	0.7908	0.6170	0.5090	0.3039	0.1843
C=5	0.7794	0.5864	0.4924	0.2697	0.1588
C=6	0.7803	0.5921	0.4741	0.2618	0.1490

Table 1: Index value of non fuzzy-index

The optimal number of cluster is solved by maximizing NFI.

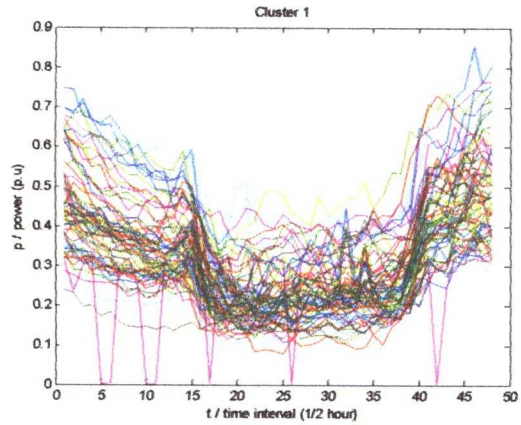

Figure 1: Pattern of cluster 1 on NFI

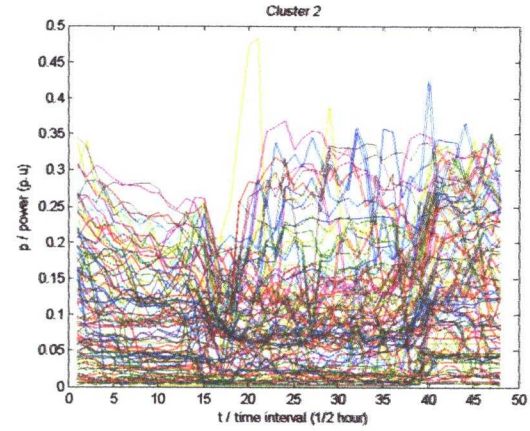

Figure 2: Pattern of cluster 2 on NFI

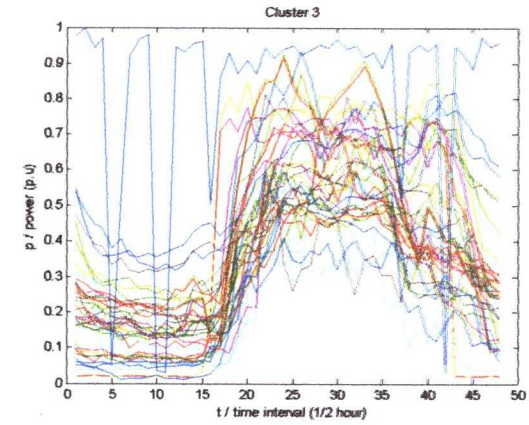

Figure 3: Pattern of cluster 3 on NFI

Figure (1)-(3): load profile for feeder in each cluster with m=1.5 and number of cluster =3

In fuzzy algorithm, a pattern of cluster will be captured using matlab programming with varying value of weighting exponent, *m*. To obtain a good cluster, clustering validity indexes must be calculated with data from the fuzzy algorithm.

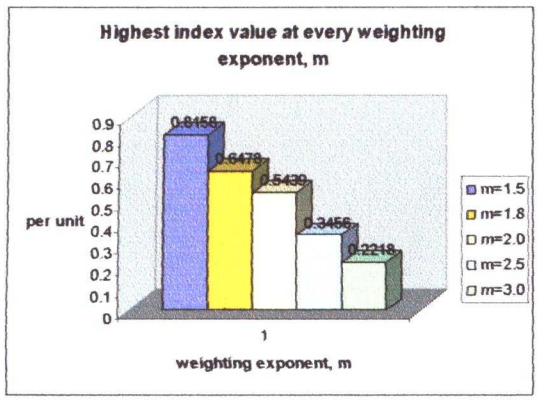

Figure 4: Optimal index value in every weighting exponent

Figure (4) was produced from table (1) where the highest value of index was taken from the highest value at every weighting exponent, *m*. All the highest index was filled at number of cluster=3. Among the index, the highest value was at c=3 and *m*=1.5. By comparing the clusters, the highest value at m=1.5 for all cluster. The index will be decreased when the value of *m* was decreased. From the table (1), the index at a cluster=2 and *m*=1.5, 1.8 and 2.0 was at a lowest value. The index also will decrease if a number of clusters were increase for the all of index data.

Figures (1)-(3) show the pattern of cluster. From an observation, the type of feeder in the cluster can be determine whether industrial, commercial, residential or mix consumer [6]. Example, pattern for cluster 1 (figure (1)), indicate residential consumer.

4.2 Xie-beni index (XB)

Table 2 below shows the values of *m* against sum of cluster using XB formula:

	m=1.5	m=1.8	m=2.0	m=2.5	m=3.0
C=2	0.0594	0.0598	0.0592	0.0558	0.0505
C=3	0.0466	0.0451	0.0434	0.0373	0.0301
C=4	0.0414	0.0399	0.0379	0.0305	0.0218
C=5	0.0397	0.0372	0.0517	0.0255	0.0170
C=6	0.0432	0.0442	0.0437	0.0380	0.0288

Table 2: Index value of XB

The optimal number of cluster is found by solving $\min_{2\le c\le n-1} XB(c)$.

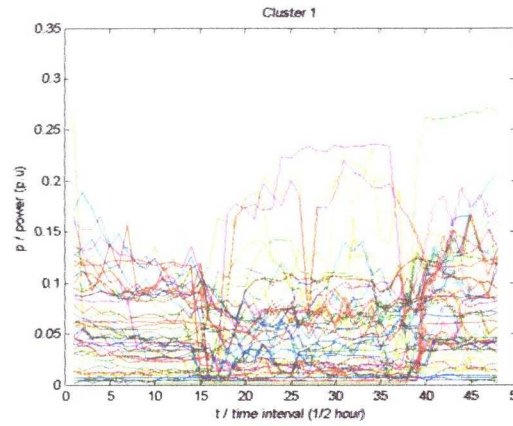

Figure 5: Pattern of cluster 1 on XB

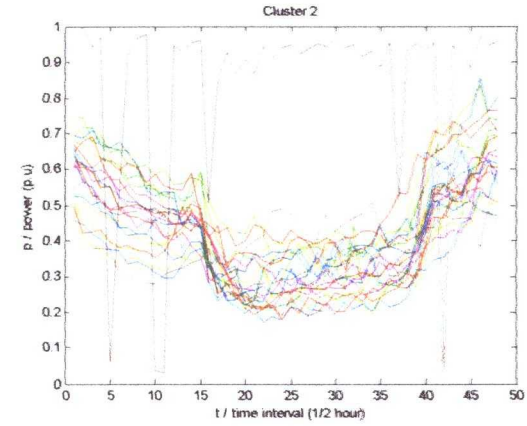

Figure 6: Pattern of cluster 2 on XB

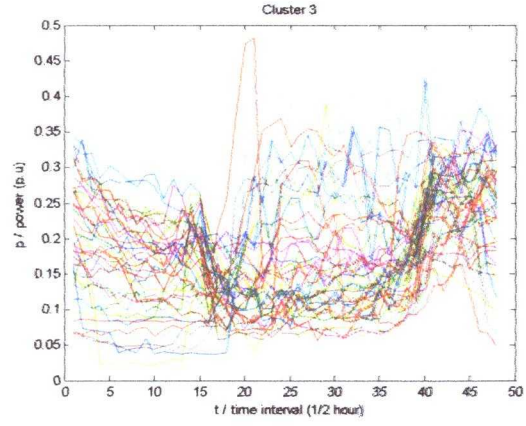

Figure 7: Pattern of cluster 3 on XB

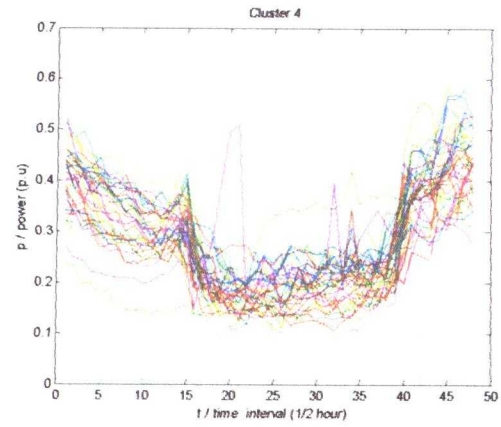

Figure 8: Pattern of cluster 4 on XB

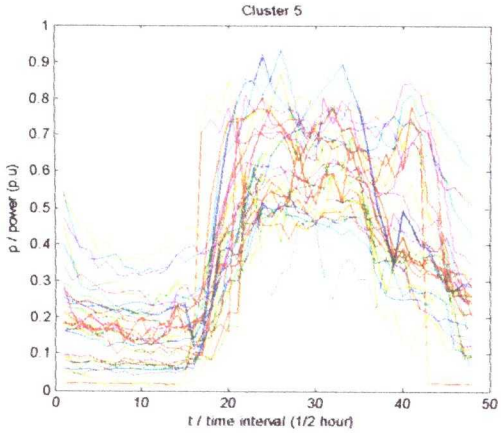

Figure 9: Pattern of cluster 5 on XB

Figure (5)-(9): load profile for feeder in each cluster with m=3.0 and number of cluster =5

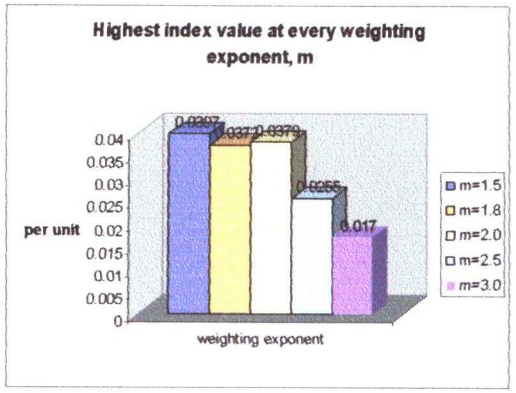

Figure (10): Optimal index value at every weighting exponent

From data in table (2), a graph in figure (10) was plotted. The graph shows the lowest index between weighting exponent and number of clusters. For m=1.5, 1.8, 2.5 and 3.0, the lowest index were at c=5, but for m=2.0, the lowest index was at c=4. Among the lowest index value, the index at m=3.0 and c=5 was a lower index than other. By comparing the clusters, the lowest value at m=3.0 for all cluster. In the table (2), the highest index were at c=2 for all value of m. These indexes will decrease when the numbers of clusters increase except the lowest index.

Figure (5)-(9) were shown the pattern of cluster where the pattern at the optimal number of clusters in xie-beni index. Beside that, type of the cluster can be decide whether industrial, commercial, residential or mix consumer.

In non-fuzzy index, the good clusters was at m=1.5 and c=3. It was because the index value at that case was highest than others. In xie-beni index, the good cluster was at m=3.0 and c=5. The value at that case was a lowest than others.

Non-fuzzy index use only the membership values, u_{ij} partitioned data. Xie-beni index involves both the U matrix and the dataset itself [6]. So, the answers between both indexes were not same.

5.0 CONCLUSION

The weighting exponent is a parameter in Fuzzy c-means algorithm. This parameter can affect the pattern of cluster by varying the value. So, the choice of the value is important to get a good cluster.

An optimal value of parameter must be tested with clustering validity indexes. There were more clustering indexes but in this paper a non-fuzzy and xie-beni index was chosen.

The result obtained is only valid for that particular data only. Even though, there are two different results, it depends on the user. If user need a three clusters only, the non-fuzzy index is suitable but if user need five clusters the xie-beni index is suitable.

6. FUTURE DEVELOPMENT

For the future development, this project can use for analyze data in big area. Tenaga Nasional Berhad (TNB) and Independent Power Producer (IPP) can use this project to get load profiling data, pattern of cluster and etc. So, they

can distribute or produce electricity according demand from consumer.

ACKNOWLEDGEMENT

The author would like to thank Dr. Zuhaina Bt Zakaria for her guidance.

REFERENCES

[1] K.L. Lo and Dr Zuhaina Zakaria, "Fuzzy Classification and Statistical Method for Load Profiling: A Comparison". United Kingdom: University of Strathclyde.

[2] Howon Choe and Jay B. Jordan. "On Optimal Choice of Parameter in a Fuzzy C-Means Algorithm.

[3] J. C. Dunn (1973). "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", *Journal of Cybernetics* 3: 32-57

[4] J. C. Bezdek (1981). "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York.

[5] Sanghamitra Bandyopadhyay. "Simulated Annealing Using a Reversible Jump Markov Chain Monte Carlo Algorithm for Fuzzy Clustering".

[6] Dr Zuhaina Bt Zakaria, "Load Profiling and Load Decomposition for Distribution Feeders" PhD Thesis. University of Strathclyde.

[7] http://www.ieee.org