Cluster Validity Analysis and Optimization of Fuzzification Parameter of Fuzzy C-Means for Determination of Typical Load Profiles

Norhasnelly Anuar  
Faculty of Electrical Engineering  
Mara University of Technology  
Penang, Malaysia  
norhasnelly@gmail.com

Zuhaina Zakaria  
Faculty of Electrical Engineering  
Mara University of Technology  
Selangor, Malaysia  
zuhaina@gmail.com

Ismail Musirin  
Faculty of Electrical Engineering  
Mara University of Technology  
Selangor, Malaysia

Nur Fadhilah Jamaludin  
Faculty of Electrical Engineering  
Mara University of Technology  
Penang, Malaysia

Abstract — Information from load profile is useful for electricity suppliers to plan their generation, improving their market strategies and load balancing. Consumers in the new liberalized market have the opportunity of choosing their electricity suppliers between several suppliers and the possibility to access to new products and services from them. Hence they need the knowledge of load profile to help them choose their electricity supplier. On the suppliers’ side, power market becomes competitive and energy commercializers are becoming more interested in the development of new suitable strategies and products to be offered to each of their different user or to find new market opportunities. A lot of efforts have been made to investigate methodologies to form optimal efficiency in determining typical load profiles (TLPs), derived from various clustering and classification techniques. Methodologies proposed in previous work have disadvantages such as time consuming, expensive, poor performance over large scale simulation and produced overlapped data in the obtained TLPs. This project proposes a methodology for determining consumers’ TLPs by using Fuzzy C-means (FCM) clustering method. Initial determination of the number of clusters and fuzzification parameters in FCM greatly influence the resulting clusters. Hence the optimal number of cluster for FCM is obtained through cluster validity analysis and the best value of fuzzification parameter, $m$ of FCM is determined to ensure the optimal result of FCM is obtained.

Keywords - component; Typical load profiles; Fuzzy C-Means; Cluster validity; Fuzzification parameter.

I. INTRODUCTION

In many parts of the world, deregulation and restructuring of electrical industry has begun and is advancing. Knowledge on consumers’ electricity usage becomes very important and crucial. Information gathered from load profiles is useful for companies to improve their market strategies, offer new services, develop tariffs, consumption planning and balance settlements [1-4]. Although information on electricity consumption can be obtained by installing the time interval meter, this technique is not practical and cause unnecessary cost for lower voltage consumers and information from billing is insufficient [5]. Through literature review, few data clustering techniques to obtain Typical Load Profiles (TLPs) has been identified. Some of the reported techniques are K-Means, Fuzzy C-Means (FCM), Hierarchical, Fuzzy K-Means, Follow the Leader and Fuzzy Relation, and Self-Organizing Maps (SOM) [6-11]. These techniques function differently and have their own advantages and disadvantages. This paper shows the determination of TLPs by using FCM. FCM is the most recognized and prominent fuzzy clustering method among other techniques [12] and its advantage is that it handles outliers more efficiently [13]. FCM requires initial determination of the number of cluster, $c$ and fuzzification parameter, $m$ which influence the resulting clusters [14]. To ensure optimal result of FCM clustering, analysis of cluster validity indices and fuzzification parameter $m$ is carried out.

II. FUZZY C-MEANS

FCM is applied in this study as the clustering algorithm to determine the TLPs. Process involves in FCM is minimizing the objective function which represent the distance between cluster point and all the data involve. An objective function measure the overall dissimilarity of data objects within each cluster. FCM minimizes the objective function that represents the distance between a given sample of data and a cluster centre to obtain the best partition of the data set. Each data points belong to several clusters with some degree that is specified by a membership grade. The objective function, $J_m$ is defined as follows,

$$J_m = \sum_{c=1}^{c} \sum_{j=1}^{N} v_{ik}^m \| x_j - v_i \|^2$$  (1)

Where,

- $N$ = number of load profile
- $c$ = number of clusters
- $m$ = fuzzification metre
- $x_j$ = load profile of $k_a$ customer
- $v_i$ = $k_a$ cluster center
- $u_{ai}$ = membership degree of data object $x_k$
Number of cluster, \( c \) and fuzzification parameter, \( m \) need to be predetermined before running the FCM algorithm. Fuzzification parameter, \( m \) is one of the most important parameters in FCM which reflects degrees of membership of each data involve in clustering. FCM updates the membership matrix, \( u_{ik} \) and cluster centre, \( v_i \) iteratively by using the following equation,

\[
\text{Cluster centre}, v_i = \frac{\sum_{k=1}^{N} u_{ik}^m x_k}{\sum_{k=1}^{N} u_{ik}^m} \quad (2)
\]

\[
\text{Membership degree}, u_{ik} = \frac{1}{\sum_{j=1}^{c} \left( \frac{d_{kj}}{d_{ik}} \right)^{m-1}} \quad (3)
\]

The optimal number of cluster and the compactness of clustering result could be determined by using several cluster validity indices or adequacy indices. This project uses cluster validity indices to determine optimal number of cluster, \( c \).

### III. CLUSTER VALIDITY ANALYSIS

Determination of optimal number of clusters or clustering validity is the main problem in FCM [15]. In general, well separated clusters should have a high intra-cluster validity that is low variance among intra-clusters members. However computing intra-cluster similarity alone is not enough to determine the optimal number of clusters. Inter-cluster variance need to be computed as well. Many cluster validity indexes have been proposed in the past work. Some of the popular approaches are nonfuzzy index, Dunn’s index, partition coefficient, partition entropy, Xie Beni index and Davies and Bouldin index [16]. In this paper, three cluster validity indexes are used. The indexes are Xie Beni index, nonfuzzy index and Davies and Bouldin index.

#### A. Xie Beni Index

The Xie Beni index is expressed as,

\[
XB = \frac{\sum_{i=1}^{c} \sum_{j=1}^{k} u_{ij}^m \| x_j - v_i \|^2}{n \min \| v_i - v_j \|^2} \quad (4)
\]

This index measures the compactness and separation of a fuzzy C partition. The optimal number of cluster is given by the minimal number of Xie-Beni index. Xie-Beni index is the most reliable cluster validity index as it provided the best response over a wide range of choices of number of clusters [17].

#### B. Nonfuzzy Index

This index measure how fuzzy a C partition is and is defined by,

\[
\text{NFI} = \frac{c \left( \sum_{i=1}^{c} \sum_{j=1}^{k} u_{ij}^2 \right) - N}{N(c-1)} \quad (5)
\]

Where \( N \) is the number of data in \( c \) clusters. The optimal number of cluster is obtained by finding the maximal value of NFI.

#### C. Davies and Bouldin Index

Davies and Bouldin index represents the ratio of total intra-cluster scatter to inter-cluster separation. A low scatter and high distance between clusters give the optimal number of cluster hence minimal value of this index is desired [18]. Davies-Bouldin index is defined by,

\[
\text{DB} = \frac{1}{c} \sum_{i=1}^{c} R_i \quad (6)
\]

Where \( R_i \) is given by,

\[
R_i = \max_{j \neq i} R_{ij} \quad (7)
\]

And \( R_{ij} \) is given by,

\[
R_{ij} = \frac{\text{var}(c_i) + \text{var}(c_j)}{\| c_i - c_j \|} \quad (8)
\]

### IV. FUZZIFICATION PARAMETER, M OF FCM

Weighting exponents in FCM called fuzzification parameter, \( m \) is one of the most important parameters in FCM which reflects degrees of membership of each data involve in clustering. Paper [19] shows several FCM clustering results with different value of fuzzification parameter. It is concluded that by increasing value of fuzzification parameter in FCM, the load profile produced will become more flat because the degree of membership will belong to all clusters. Calculation by Nikhil R. Pal and James C. Bezdek suggested that the best choice for \( m \) is in the interval \([1.5, 2.5]\) where most common used value of ‘fuzzification’ parameter being reported is \( m = 2 \) [20]. Past works reported that a large value of \( m (>2.5) \) will make FCM to be more robust to noise and outliers while when the value becomes closer to 1, clustering algorithm will behave like a hard clustering where each data will belong to one distinct cluster [21]. As the value of \( m \) increases, the fuzziness will increase and the boundaries between clusters will be less
sharp. In this work, the influence of this parameter \( m \) to clustering result was examined by varying the value from 1.50 to 4.00.

V. RESULTS AND DISCUSSION

A. Cluster Validity Analysis

Figure 1 shows 347 measured load profiles (MLPs) used in this study. MLPs show the changes in electricity usage over a period of time. This data is obtained from Tenaga Nasional Berhad (TNB), the electricity provider in Malaysia. Data is recorded for every 30 minutes, for 24 hours in a day. MLPs are normalized into per unit value by using the peak power of the whole measurement as the normalizing factor. MLPs shown in Figure 1 is the input of FCM algorithm used in this study.

One of the parameter in FCM which is the number of cluster, \( c \) needs to be predetermined. Knowing the properties of load data used in this study, the execution of FCM algorithm is done with the number of cluster, \( c \) set to 3 and is repeated until \( c \) equal to 10. Cluster validity analysis is carried out to measure the quality of clustering results and to assist in deciding the best value of \( c \) that fit the data. Three cluster validity indices are applied in this study which are Xie Beni, nonfuzzy and Davies and Bouldin index. The validity indices are tabulated in Table I. The value of fuzzification parameter \( m \) is set to 2 which is the most common used value as reported.

<table>
<thead>
<tr>
<th>CVI</th>
<th>No of Clusters, c</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>XB</td>
<td>0.038</td>
</tr>
<tr>
<td>NFI</td>
<td>3.081</td>
</tr>
<tr>
<td>DB</td>
<td>0.034</td>
</tr>
</tbody>
</table>

Optimal number of cluster is solved by minimizing the Xie Beni index. From Table I, the smallest index by Xie Beni for \( c \) is equals to 6 clusters. Meanwhile, maximal index of non-fuzzy index is solved to obtain the optimal number of clusters. Table 1 shows that the optimal number of cluster for non-fuzzy index is when \( c \) equals to 3 clusters. Further observation shows that the second maximal value of non-fuzzy index is when \( c \) equals to 6. As for Davies-Bouldin index, the optimal number of cluster is obtained by the minimal value which is when \( c \) equals to 7 clusters. The second minimal value of Davies-Bouldin index is when \( c \) equals to 6 clusters.

Figure 2 shows the result from FCM clustering when \( c \) is equal to 3. The MLPs shown in Figure 1 are clustered into three groups which are C1, C2 and C3. The first cluster, C1 shows low power consumption with almost stable and steady demand throughout the day. The second cluster, C2 shows low power consumption after midnight until 7.30 a.m where the power demand increase steeply until 12.00 p.m where it stabilized and consumption demand drop after 5.00 p.m. The third cluster, C3 shows a stable drop in power consumption after midnight and a sudden drop after 7.00 a.m where the consumption stabilized and starts to increase at 6.30 p.m until midnight. C1 shows consumption pattern for industrial consumer while C2 represents commercial consumer and C3 represents domestic consumer.

Figure 3 shows six average typical load profiles when \( c \) is equal to 6. It can be observed that as the number of cluster, \( c \) increases, the typical load profiles obtained become more ‘crisp’ and dispersed from other classes. Cluster 2, C2 in Figure 2 is being clustered into two clusters, C3 and C6 in Figure 3. Cluster 3, C3 in Figure 2 is
being clustered into three clusters, C1, C4 and C5. C2 in Figure 3 have similar power consumption pattern with C1 in Figure 2 but have different range of power demand. By comparing the pattern of clusters in Figure 2 and Figure 3, 6 is chosen as the optimal number of clusters for this study that best fit the data.

![Figure 3. Average Typical Load Profile when c=6](image)

**B. Optimization of fuzzification parameter, m**

FCM algorithm allows each data to belong to every cluster. Each data points belong to several clusters with some degree of membership grade. Fuzzification parameter, $m$ in FCM reflects degrees of membership of each data involve in clustering. The optimal value of fuzziness parameter can be determined through observation of the cluster validity indexes. Davies-Bouldin index is used to determine the cluster validity. Davies-Bouldin index uses the concepts of dispersion of a cluster and dissimilarity between clusters. Fuzzification parameter, $m$ is varied between 1.5-4.0. The cluster validity indexes for 3 to 9 clusters for Davies-Bouldin index are shown in Table II.

<table>
<thead>
<tr>
<th>No of C</th>
<th>Fuzzification Parameter, $m$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>0.0205</td>
</tr>
<tr>
<td>4</td>
<td>0.0378</td>
</tr>
<tr>
<td>5</td>
<td>0.0567</td>
</tr>
<tr>
<td>6</td>
<td>0.1298</td>
</tr>
<tr>
<td>7</td>
<td>0.0223</td>
</tr>
<tr>
<td>8</td>
<td>0.2232</td>
</tr>
<tr>
<td>9</td>
<td>0.0454</td>
</tr>
</tbody>
</table>

According to Table II, it is observed that $c=3$ is the optimal cluster for when $m=1.5$, 2.0 and 3.5. When $m=2.5$ the results suggest that the optimal $c$ is 6. The validity index suggested that 4 and 7 is the optimal number of cluster when $m=3.0$ and $m=4.0$ accordingly. Since the cluster validity analysis in the previous section shows that the optimal $c$ is 6, hence the optimal value of fuzzification parameter is achieved when $m=2.5$.

**VI. CONCLUSION**

Result analysis of FCM clustering algorithms indicates that the optimal number of cluster for FCM is 6. The 6 typical load profiles represents three main types of consumers which are commercial, industrial and residential. These three main categories are further divided into different level of consumption value. This result has been compared with the list of categories from TNB and it is verified as correct. The database from TNB shows that consumers involved from the tested area are from these 3 categories. Besides number of cluster, another important FCM parameter that could influence the resulting result greatly is the fuzziness parameter, $m$. From this study, the optimal value of fuzziness parameter is 2.5.

**ACKNOWLEDGMENT**

Authors would like to thank Dr Zuhaina Zakaria for the continuous support and assistance in retrieving the measured load profiles used in this study.

**REFERENCES**


