

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Kohonen Maps Combined to Fuzzy C-means, a Two Level Clustering Approach. Application to Electricity Load Data

Khadir M. Tarek and Benabbas Farouk

Laboratoire de Gestion Electronique de Documents (LabGED)

*Department of Computer Science, University Badji Mokhtar, Annaba,
Algeria*

1. Introduction

In a deregulated electricity market, load forecasting is nowadays of paramount importance to estimate next day load resulting in energy save and environment protection. Electricity demand is influenced (among other things) by the day of the week, the time of year and special periods and/or days such as religious and national events, all of which must be identified prior to modelling. This identification, known as day type identification, must be included in the design stages either by segmenting the data and modelling each day type separately or by including the day type as an input, which implies data classification and cluster creation.

Data classification consists in regrouping objects of a similar data set into homogenous classes. Two main types of classifications exist: supervised and unsupervised classification. Supervised classification is based on a set of objects L of known classes, called training set, with the main goal being to identify candidate objects into their belonging classes. Where, unsupervised classification consists in partitioning a set of data D into sub-sets of similar attributes called classes or clusters (Halgamuge, 2005). Unsupervised classification is termed *clustering*, and will be so in the remaining of the chapter.

For clustering means, conventional research usually employs multivariate analysis procedures. However, it was found that clustering the data directly, becomes computationally heavy using statistical method as the size of the data set increases (Jain & Dubes, 1988; Xu & Wunsch, 2005). Despite this fact, many linear approaches such as Principal Component Analysis (PCA) (Jolliffe, 2002) and K-means were and remain, extensively used for classification and clustering purposes.

Nonlinear classification and clustering approaches stand as a strong alternative in order to treat the complexity and visualisation problems issued from large multidimensional data sets. In recent years, due to their high performance in engineering, Artificial Neural Networks (ANN), more specifically Self Organising Maps (SOM), and fuzzy logic are now being used as alternate statistical tools. Combining both paradigms in a two-level approach may be profitable to reduce significantly the computational cost as shown in (Khadir et. al., 2010) where SOM and K-means were combined for time series clustering.

The idea behind the two level approach is to reduce the data dimensions using SOM finding the data quantization prototypes in the first stage. The second stage, then focuses on clustering the already obtained prototypes using, in this case, Fuzzy C-means (FCM) (Vesanto & Alhoniemi, 2000).

2. The Kohonen self-organizing map

The Kohonen self-organizing map (SOM) is an unsupervised classification method, which transforms a set of complex data to one or two dimensional vectors with a simple geometric relationships, and preserving the most important initial data metrics during the display, i.e. the close dataset of the input space will have close representations in the output space and thus will be classified in the same cluster or nearby clusters (Kohonen, 1990, Dreyfus *et al.*, 2004). The self organizing map is suitable for data survey because it has prominent visualization properties; it is also a very effective tool for visualizing and exploring multidimensional data (Himberg, 2000; Vesanto, 1999). SOM has two layers, the input and the Kohonen or output layer, Figure 1.

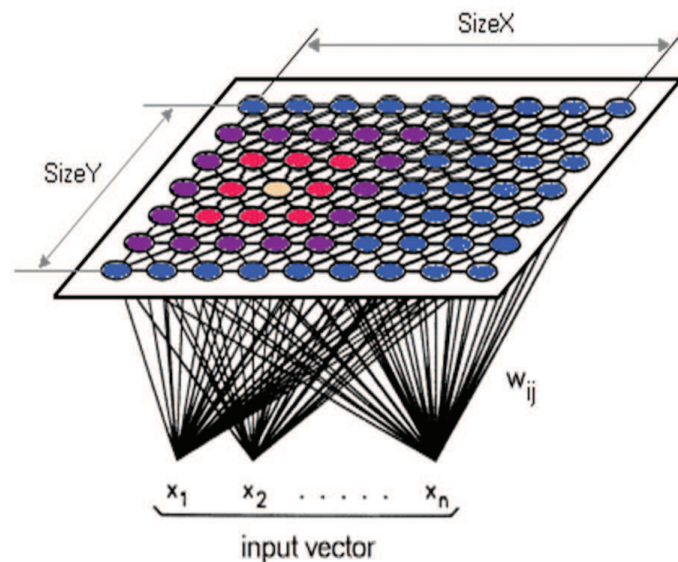


Fig. 1. Two dimensional Kohonen map

The network consists in a grid of output nodes connected to the inputs via a set of weights. When presented with the k^{th} input vector $P_k \in \mathbb{R}^{1 \times n}$, the network calculates the activation of each node using P_k as:

$$a_{i,j,k} = W_{i,j} P_k \quad (1)$$

where $a_{i,j,k}$ and $W_{i,j}$ are the activation of, and weight ($\in \mathbb{R}^{1 \times n}$) connecting P_k to, node i, j respectively. P_k is said to be mapped onto the node with the highest activation. After several inputs have been presented, similar inputs are mapped to the same or adjacent nodes, i.e., within a small neighbourhood. A neighbourhood of size N_c around node i, j is defined as nodes $i \pm N_c$ to $j \pm N_c$. P_k for the current study is formed in two steps.

Each neuron of the topological layer is completely connected to the input layer neurons $W_i = (W_{i1} \dots W_{in})$, the weight vectors of these connections form the referent or prototype associated to each neuron, it has the same dimension as the input vectors. In each training

step, one sample vector \mathbf{x} from the input data set is chosen and a similarity measure is calculated between it and all the weight vectors of the map. The Best-Matching Unit (BMU), is the unit whose weight vector has the greatest similarity with the input sample P . The similarity is usually defined by means of a distance measure; typically Euclidian distance. The use of neighbourhood concept introduces the topological constraints in the final SOM geometry.

The weights may or may be not, initialised randomly. In some cases they are initialised around the mean of the inputs as the inputs are all similar and thus restricted to a small portion of the space.

The neurons of the Kohonen map learn to recognize groups of similar input vectors. Thus, the neuron whose weight vector is closer to the input vector is then updated to be even closer. The result is that the winning neuron is more likely to win the competition next time if similar vector is presented, and less likely to win when a very different input vector is presented. The training stage stops when any of the following conditions are met: the maximum number of epochs is reached, the performance has been minimized to the goal, or maximum amount of time has been exceeded.

During training the inputs are presented one by one and the weights of the triggered node (the node to which the inputs is mapped) and nodes in its neighbourhood are updated as in equation (2).

$$W_{i,j}(m+1) = W_{i,j}(m) + \alpha(m) [P_k - W_{i,j}(m)] \quad (2)$$

Where α is the adaptation gain, with $0 < \alpha < 1$, and m is the iteration number. This has the effect of increasing the activation of the triggered node and its neighbours. In a single iteration all the inputs are presented and the weights adapted. After several iterations, the neighbourhood size is reduced by one and so on until zero, i.e., the triggered node only is adapted.

3. Fuzzy C-means clustering

K-means and conventional clustering techniques are referred to as hard or crisp clustering, which means that each object is assigned to only one cluster. For fuzzy clustering, this restriction is relaxed, and the object can belong to all identified clusters with a certain degree of membership (Bezdek, 1981). This is particularly useful when the boundaries among clusters are not well separated and ambiguous. FCM is one of the most popular fuzzy clustering algorithms (Szilágyi, 2009), it attempts to find the most characteristic point in each cluster, which can be considered as the "centroid" of the cluster and then, the grade of membership for each object in the clusters. Such aim is achieved by minimizing an objective function. A commonly used objective function is: membership weighted within cluster error defined as follows:

$$J_m(U, V) = \sum_{i=1}^n \sum_{j=1}^c (u_{i,j})^m \|x_j - v_i\|^2 \quad (3)$$

Where n is the total number of patterns in a given data set and c is the number of clusters; $X = \{x_1, x_2, \dots, x_n \in \mathbf{R}^s\}$ and $V = \{v_1, v_2, \dots, v_n \in \mathbf{R}^s\}$ are the feature data and $U = [u_{ij}]_{c \times n}$ is a fuzzy

partition matrix composed of the membership grade of pattern x_j to each cluster i . $\|x_j - v_i\|$ is the Euclidean norm between x_j and v_i .

The weighting exponent m is called the fuzzifier which influences the clustering performances of the FCM (Cannon, 1986; Bezdek and Pal, 1988; Yu et al., 2005). The cluster centroids and the respective membership functions that solve the constrained optimization problem in (3) are given by the following equations:

$$v_i = \frac{\sum_{j=1}^n (u_{ij})^m x_j}{\sum_{j=1}^n (u_{ij})^m}, \quad 1 \leq i \leq c, \quad (4)$$

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{\|x_j - v_i\|^2}{\|x_j - v_k\|^2} \right)^{1/(m-1)} \right]^{-1}, \quad 1 \leq i \leq c, 1 \leq j \leq n. \quad (5)$$

Equations (4) and (5) constitute an iterative optimization procedure. The goal is to iteratively improve a sequence of sets of fuzzy clusters until no further improvement in $J_m(U, V)$ is possible.

3.1 The fuzzy C-means clustering algorithm

The FCM algorithm is executed in the following steps:

Step 1: Given a pre-selected number of cluster c , a chosen value of m , initialize memberships u_{ij} of x_j belonging to cluster i such that

$$\sum_{i=1}^c u_{ij} = 1. \quad (6)$$

Step 2: Calculate the fuzzy cluster centroid v_i for $i = 1, 2, \dots, c$ using Eq. (4).

Step 3: Employ Eq. (5) to update the fuzzy membership u_{ij} .

Step 4: If the improvement in $J_m(U, V)$ is less than a certain threshold (ε), then halt; otherwise go to step 2.

Numerous FCM variants and other fuzzy clustering algorithms have appeared as a result of the intensive investigation on the distance measure function, the effect of weighting exponent on fuzziness control (Hoppner, 1999; Eschrich et al., 2003).

3.2 Cluster validity and validity indices in fuzzy environment

The FCM clustering algorithm has been widely used to obtain a fuzzy c -partition. This algorithm requires the user to predefine the number of cluster (c) and the fuzzier parameter m . Since it is not always possible to know these parameters in advance, different fuzzy partitions are obtained for different values of c . An evaluation methodology is required to validate each of the fuzzy c -partitions and, once the c -partitions are established, an optimal partition (or optimal number of clusters) may be considered. This quantitative evaluation is the subject of cluster validity. The cluster validity may be an indicator on partition quality.

Consequently, the cluster validity index can also be used to search for the optimal number of clusters when the number of clusters in data set or prototypes is not known in advance. Among the most used validity indices, and the one used in the remainder of the chapter, one may cite the following:

The Partition Coefficient (PC): measures the amount of overlapping between clusters. It is defined by (Bezdek, 1981) as follows:

$$PC = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n u_{ij}^2 \quad (7)$$

The Partition Entropy (PE): measures only the fuzziness of the Partition, similarly to the Partition Coefficient (PC) (Bezdek, 1981).

$$PE = -\frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n u_{ij} \log u_{ij} \quad (8)$$

The Partition Index (SC): It is the sum of the individual cluster validity measures normalized through division by the fuzzy cardinality of each cluster (Bensaid, *et al.*, 1996).

$$SC = \sum_{i=1}^c \frac{\sum_{j=1}^n (u_{ij})^m \|x_j - v_i\|^2}{N_i \sum_{k=1}^c \|x_j - v_i\|^2} \quad (9)$$

The Xie and Beni's index: aims to quantify the ratio of the total variation within clusters and the separation of clusters (Xie and Beni, 1991).

$$XB = \frac{\sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2}{n \min_{i \neq j} \|x_j - v_j\|^2} \quad (10)$$

Both *PC* and *PE* possess monotonic evolution tendencies with *c*, and involve only the membership values. This may have some drawbacks cited as: Their monotonous dependency on the number of clusters, their sensitivity to the fuzzier parameter, *m* and the lack of direct connection to the geometry of the data or prototypes, since they do not use the data itself. *SC* is useful when comparing different partitions having equal number of clusters. A lower value of *SC* indicates a better partition. Xie and Beni's validity function involves the membership values and the data or prototypes. The validity *XB* index focuses on two properties: compactness and separation.

4. A two level clustering approach

The number of prototype vectors resulting from SOM clustering may be large, especially when dealing with highly multidimensional time series applications. Only one classification level can then be revealing. A high level is interesting because it provides more detailed quality analysis and less compresses the dataset if we summarize all days by representatives

of a small class's number (Rousset, 1999). It also can be very difficult to attribute some units of the input vector to a given cluster given by the map. The problem lies in the selection of some clusters border, where a clear distinction between two clusters is impossible. A second clustering stage becomes then useful to remove ambiguity and validate the SOM results.

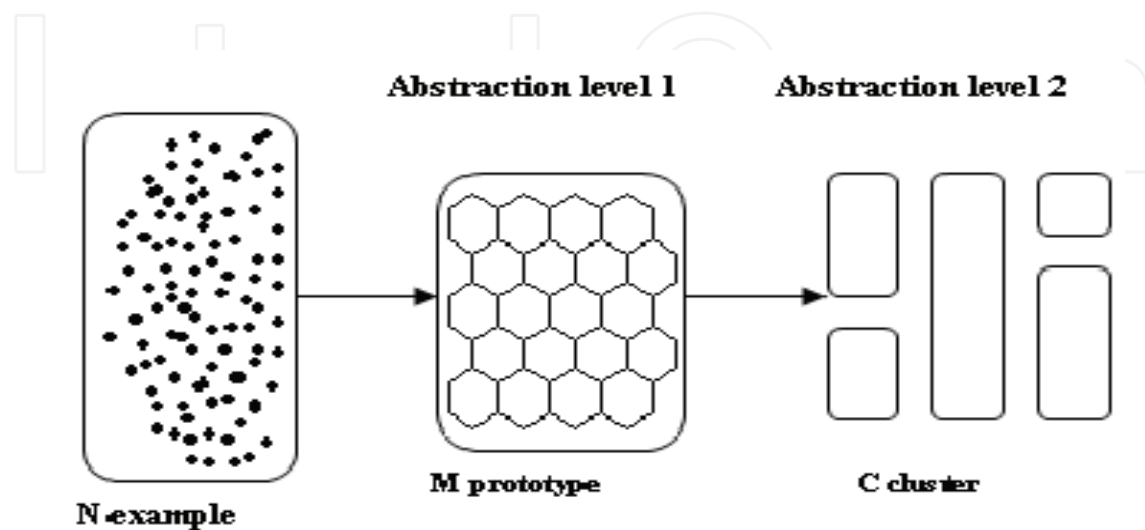


Fig. 2. First abstraction level is obtained by creating a set of prototypes vectors using the SOM. Clustering of the SOM creates the second abstraction level

The approach used in this chapter, is depicted in Fig. 2, the first abstraction level is achieved by creating a set of prototypes using SOM. These prototypes are then clustered in the second abstraction level using the fuzzy c-means clustering algorithm (Section 3). It was noticed that clustering a large multidimensional time series data using only fuzzy c-means or k-means is computationally heavier than the two-level clustering approach. Another advantage of this approach is noise reduction (Vesanto & Alhoniemi, 2000), as the prototypes are local averages of the data and therefore less sensitive to random variations than the original data.

5. Day type identification of electricity load

As explained in Section 1, it is proven that the day types or daily consumer's habits for different periods of time, such as working days, weekends, special holidays, etc affect heavily the load shape (Fay, 2004). Different prediction models may then be designed for each day type. A rigorous study of the load data is, therefore of paramount importance prior to any modeling stages.

5.1 Overview of Algerian electricity load

Electrical demand in Algeria from 01/01/2000 to 31/12/2004 is shown in Fig. 3. As can be seen there is an upward trend in the data reflecting increasing economic activity over this period, also the seasonal aspect of the time series is clearly highlighted.

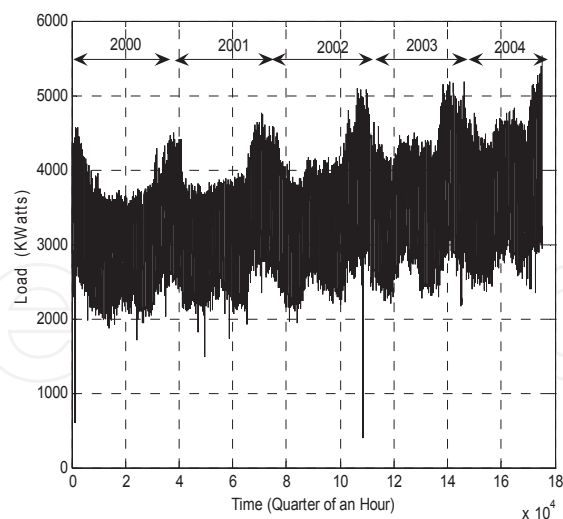


Fig. 3. Algerian electricity load 2000-2004

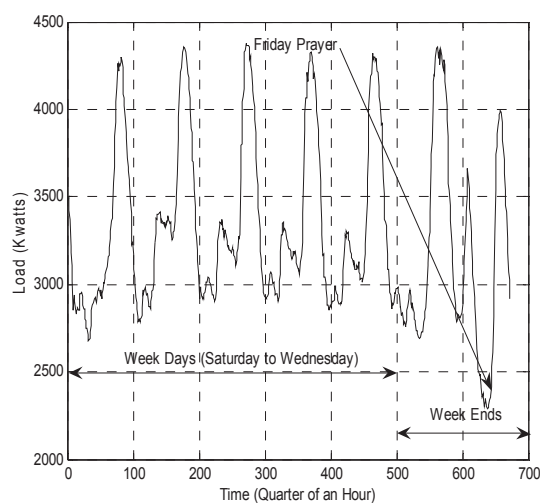


Fig. 4. Weekly load

Daily load data can be disaggregated into distinct groups (called day-types) each of which has common characteristics. As can be seen in (Fig. 4.) there is, for example, an obvious difference between the shapes of the load on a typical weekend day, such as Friday and a working day such as Saturday or Sunday due to decreased economic activity and the weekly religious prayer on Friday. Note that in Algeria the weekend was on Thursdays and Fridays in those years. Furthermore, there is a distinct difference between the shape of a typical winter day and summer day.

In addition to time, seasonal and economic activity, the regional aspect may strongly affect load characteristics. As such, the identification study has to address concerns for the load of four distinctive regions: Algiers (the capital), Oran, Hassi Messaoud and In Amenas. The two first regions are located in the north-west and north-center of the country. Their climate is Mediterranean with a complete seasonal cycle. The size of the population and the economic development are increasing significantly. The last two regions are located in the south of the country (Sahara). Their climate is very hot all year long; and the population

density is less than one inhabitant per Km² which decreases considerably any economic activity. Fig. 5 shows the Saturday 18th of March 2000, normalized load for all four regions. As it can be seen, the differences in characteristics, concerning the early, midway and late peaks are clearly highlighted.

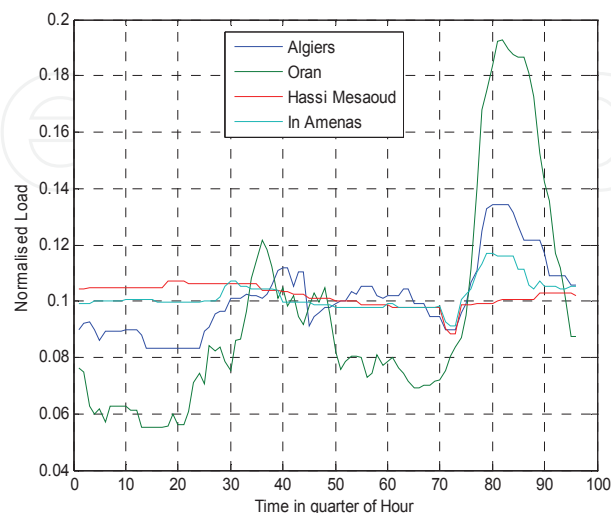


Fig. 5. Regional load differences

5.2 Day type identification using Kohonen maps

The existence of several different day types has been shown by several researchers (Bretschneider *et al.*, 1999; Hsu and yang, 1991; Muller and Petrisch, 1998). However, the level of desegregation in day-type selection is, to a large extent, subjective and dependant on the judgment of the forecaster. As pointed out by (Hubele and Cheng 1990), the application of a separate load forecasting model for different seasons (for example summer, autumn, winter and spring) has the advantage that the models do not need to incorporate seasonal information.

Further desegregation of the load by day of the week (for example Summer Sunday, Winter Sunday, Summer Monday etc.) reduces further the amount of information that the model needs to incorporate. Such approaches have been implemented successfully by (Srinivasan *et al.*, 1999) and (Mastorocostas *et al.*, 1999), to mention but a few. Where a single model is used for all the data, the day-type information is often incorporated as an additional input (two examples are (Chen *et al.*, 1992) and (Lertpalangsunti and Chan, 1998)). In either case the day-types must, however, be identified. The selection of day-types can be guided by analytical techniques. The self-organising feature map or Kohonen map (Kohonen, 1990) would appear ideal for day-type identification as the number and similarity between day-types is not known *a priori*. The Kohonen map can be implemented for day-type identification in several different ways (Fay and Ringwood, 2003; Hsu and yang, 1991; Muller and Petrisch, 1998) however differences in the results are insignificant in most cases thus the algorithm used by Hsu and Yang (Hsu and yang, 1991) was chosen.

For the present trials, the full years of data from 2003 and 2004 for two characteristic regions were used. In order to investigate the regional aspect of electricity load, data from Oran and Algiers representing northern large populated cities and Hassi Messaoud and In -Amenas southern ones were chosen. The Kohonen map was trained using the following parameters,

an initial neighborhood size of $N_c=1$, adaptation gain equal to 0.003, a total number of iteration $m=10$ and a grid size 18×18 (324) in total.

Initially, the daily load curve is extracted from each day to give a set of load curves that have a minimum value of zero and a maximum value of one (Hsu and yang, 1991).

$$Y'(i)_k = \frac{Y(i)_k - \min Y_k}{\max Y_k - \min Y_k} \quad i = 1, \dots, 24 \quad (11)$$

where $Y'(i)_k$ and $Y(i)_k$ are the i^{th} elements (hour) of the load curve $Y'_k \in R^{1 \times 24}$, and actual load $Y_k \in R^{1 \times 24}$ of day k respectively. The load curves are then normalised to give them unity length:

$$P(i)_k = \frac{Y^i(i)_k}{\left(\sum_{j=1}^{24} Y_k^j\right)^{1/2}} \quad i = 1, \dots, 24 \quad (12)$$

where $P(i)_k$ is the i^{th} element of P_k . The weights are initialised as:

$$W_{i,j} = \left\| \left[\left(\mu_p(1), \dots, \mu_p(24) \right) + 5\mu \left[\left(\rho_p(1), \dots, \rho_p(24) \right) \right] \right] \right\| \quad (13)$$

where $\mu_p(1)$ and $\rho_p(1)$ are the sample mean and standard deviation of $P(i)$ over all k , μ is a uniformly distributed random number between -0.5 to 0.5 and $W_{i,j}$ is normalised to unit length as in (Hsu and yang, 1991). Weight update is then done following equation (14) repeated below for clarity:

$$W_{i,j}(m+1) = W_{i,j}(m) + \alpha(m) \left[P_k - W_{i,j}(m) \right] \quad (14)$$

Fig. 6 shows the triggered nodes identified for the years starting from 2000 until 2004. We notice that they are located in the map at the coordinates i, j with: i between 0 and 17 and j between 10 and 20.

All the years from 2000 to 2004 for Algiers.

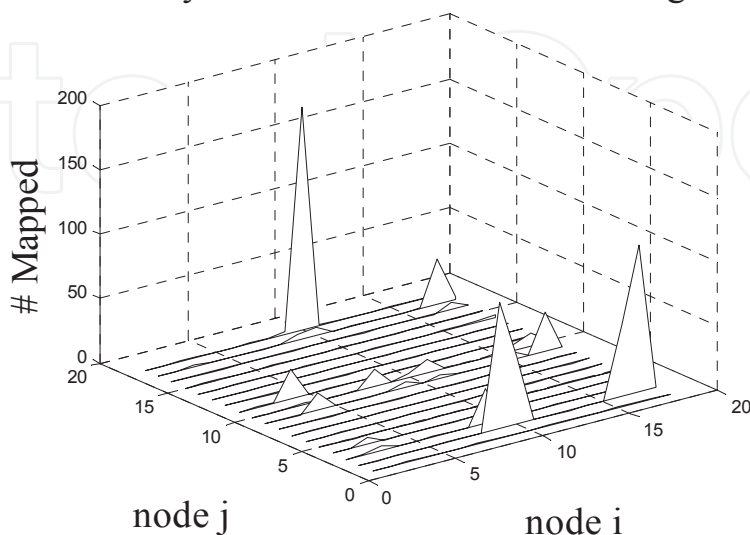


Fig. 6. Kohonen map results for Algiers load

Fig. 7 shows respectively the triggered nodes identified for 2003 until 2004 for Algiers, Oran, Hassi Messaoud and In Amenas. Clearly, according to their regional characteristics, northern cities exhibit similar behaviour.

It can be seen, Fig. 8, that week days activate roughly the same map nodes where, the weekend activate different nodes for Algiers load. This is also true for Fridays which is the weekly prayer occurring from 12 to 2:30. Thursday and Friday are the day of the weekend in Algeria.

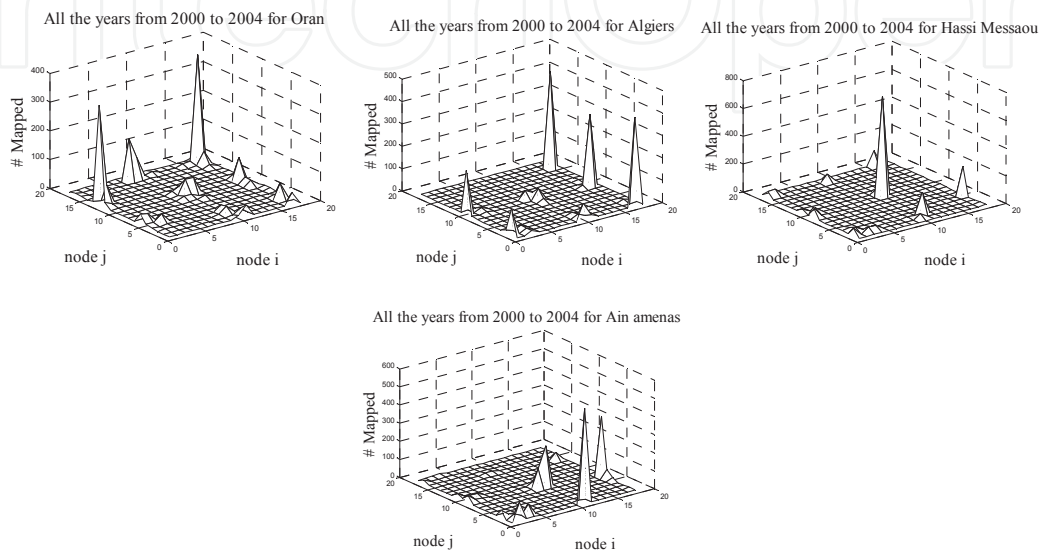


Fig. 7. Nodes triggered for All the years from 2003 to 2004 for (Algiers, Oran, Hassi Messaoud and In-Amenas)

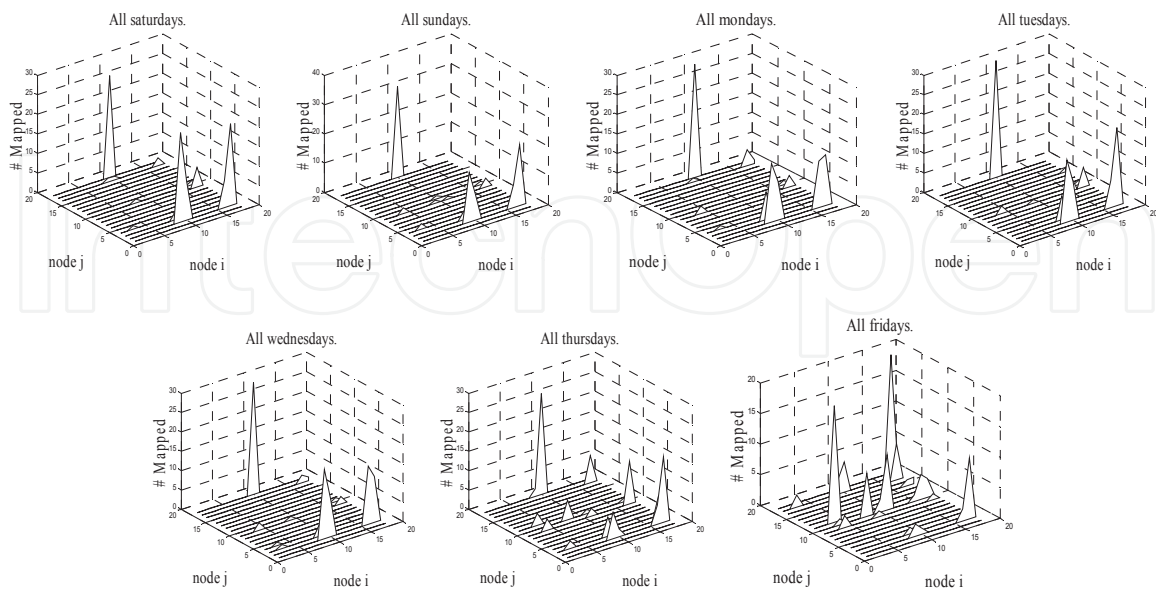


Fig. 8. Nodes triggered for working days (Saturday to Wednesday) and Week days (Thursday and Friday) loads for the region of Algiers

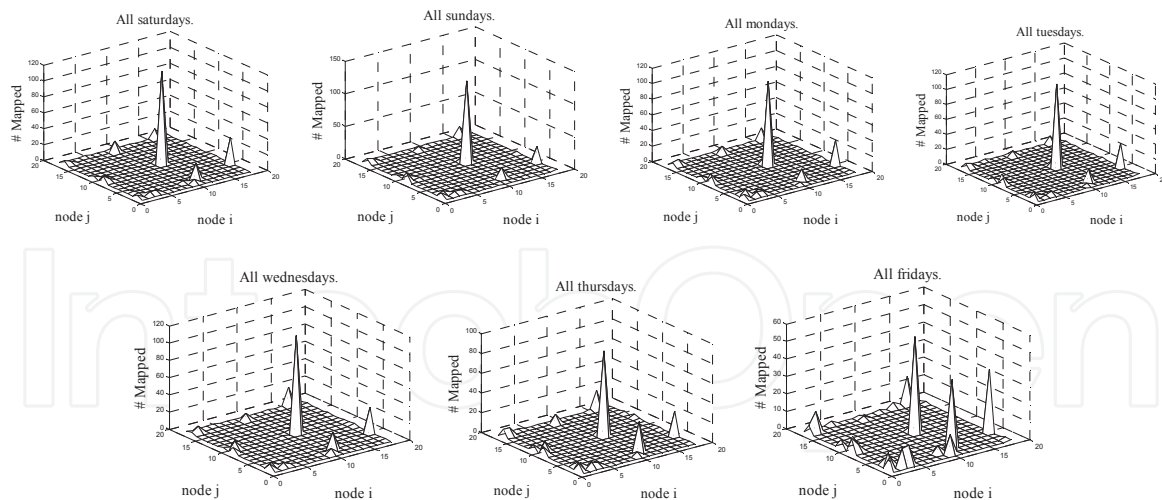


Fig. 9. Nodes triggered for working days (Saturday to Wednesday) and Week days(Thursday and Friday) loads for Hassi Messaoud

Weekdays and weekends however appear differently on the map. The nodes that are triggered from Saturday to Wednesday occupy the same parts of the grid but Thursday and Friday (weekends) loads; trigger different parts of the grid showing the difference between these day types. The figure shows the difference between these days for Algiers load, where the disparate distribution of Fridays appears clearly, and is heavily dependant on seasonal effect.

For southern regions, nodes triggered from Saturdays to Fridays for Hassi Messaoud and In-Amenas are roughly the same, showing no difference between weekdays and weekend, characteristic of southern cities, with small economic activity and population (Fig. 8).

The seasonal effect is clearly shown for northern cities, Fig. 10 for Algiers where peaks appear along the longitudinal axe of the SOM with respect to monthly (seasonal) load. Fig. 11 shows seasonal day-type identification for a southern region. The number of visually seen clusters may be identified as 4 clusters for Algiers and Oran and 3 clusters for Hassi Messaoud and two clusters for In-Amenas (Benabbas *et al.*, 2008).

In the experiments, the cluster validity indices were tested for Algiers and Hassi Messaoud prototypes in Section 5.3.

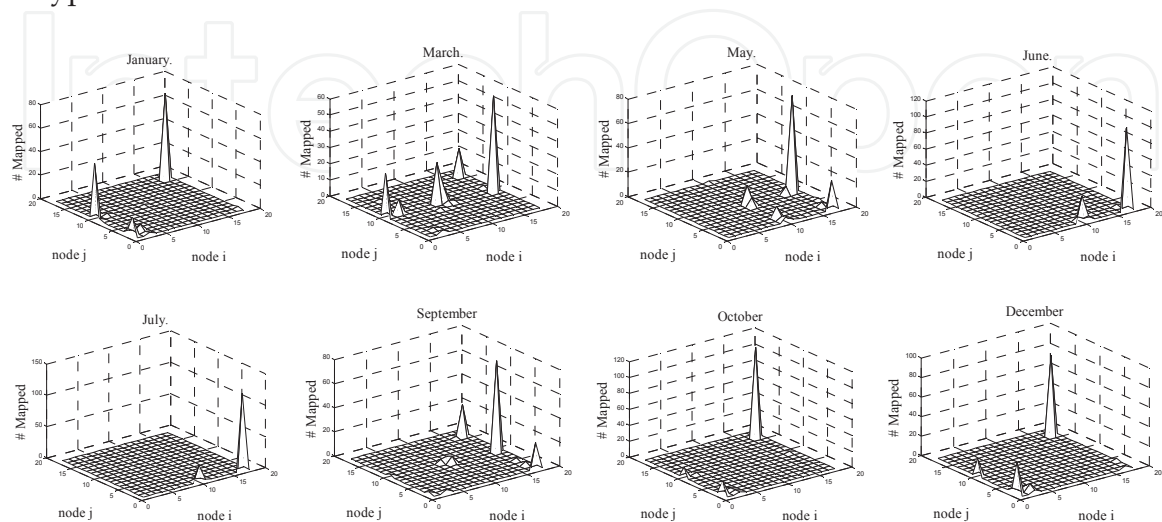


Fig. 10. Seasonal day-type identification for Algiers

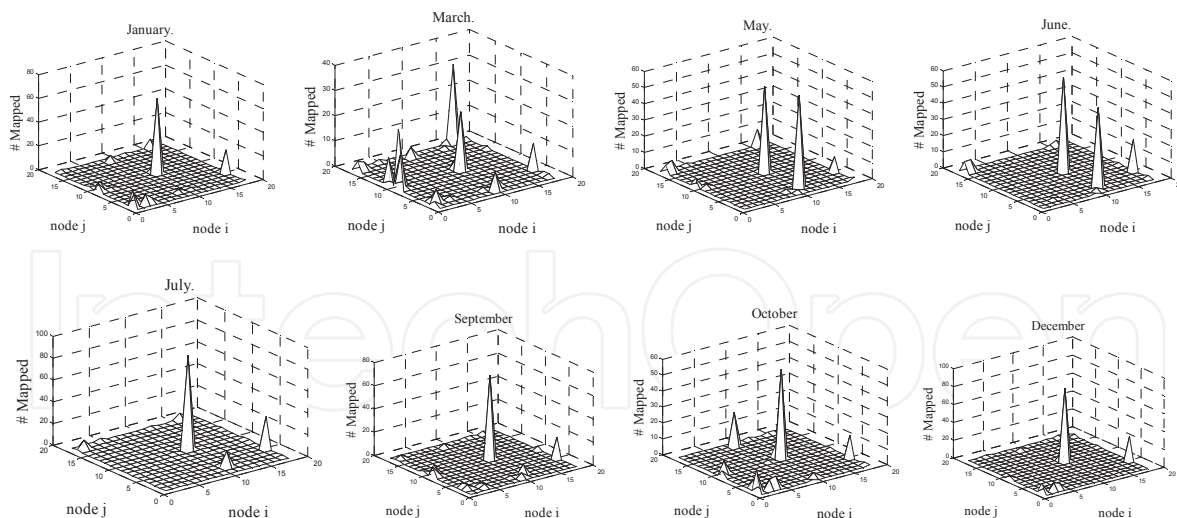


Fig. 11. Seasonal day-type identification for Hassi Messaoud

5.3 Fuzzy C-means clustering of the SOM prototypes

Table 1 lists the results of validity index for $c=2,3,\dots,13$. For each $c \geq 2$, index values were computed for each of the four validity index considered (PC, PE, SC and XB given in equations 7, 8, 9 and 10 respectively). The optimal number of clusters is found to be, 4 for Algiers and 3 for Hassi Messaoud.

As a second stage clustering technique, FCM is used to assign the already identified items to different clusters of electricity load (Jang et al., 1997). The fuzziness index m was set to a value of 2; the maximum number of iteration was set to 100. These values were selected after a number of simulations and were found most suited for clustering the data obtained from the first clustering stage. The minimal amount of improvement was initially set to $1e-5$. The membership's value in each group or cluster indicates the probability for a day to belong to that specific cluster. A sample of the membership's values of each of the 731 days for each of the 4 group is presented in Table 2. The group which has the highest membership's value among the 4 groups is the representative group for that day.

DataSet	PC	PE	SC	XB
Algiers	2	2	4	4
Hassi Messaoud	3	3	3	3

Table 1. Optimal cluster number according to each validity indices

For the first day corresponding to the 1st January 2003, membership's values for the 4 groups are 0.0812, 0.0353, 0.2641 and 0.6192 respectively for groups 1 to 4. The sum of these values should be equal to 1. The representative group for the day represented by item 1, is group 4 (having the maximum membership's value of 0.6192). Similarly, all other days were analysed and grouped. It is observed that none of the 4 groups is empty. This may be due to the advantage of Fuzzy cluster analysis which allows each data set to have partial membership in all clusters. Table 3 shows the results of fuzzy clustering in three clusters for Hassi Messaoud.

Item number	C1	C2	C3	C4
1	0.0812	0.0353	0.2641	0.6192
2	0.1012	0.0388	0.3516	0.5083
3	0.0691	0.0341	0.2005	0.6960
4	0.2288	0.0535	0.5690	0.1485
5	0.7004	0.0675	0.1813	0.0505
.....				
.....				
727	0.3198	0.4956	0.1209	0.0635
728	0.4382	0.4055	0.1080	0.0481
729	0.5729	0.3079	0.0864	0.0326
730	0.4382	0.4055	0.1080	0.0481
731	0.6134	0.2451	0.1013	0.0401

Table 2. Obtained results for Algiers (four clusters)

Item number	C1	C2	C3
1	0.0695	0.1092	0.8211
2	0.0412	0.0476	0.9110
3	0.2916	0.1350	0.5733
4	0.0319	0.0658	0.9022
5	0.1279	0.1038	0.7682
6	0.0379	0.0938	0.8681
.....			
.....			
726	0.0700	0.2241	0.7058
727	0.0522	0.8123	0.1353
728	0.0737	0.3281	0.5980
729	0.0438	0.1481	0.8079
730	0.0379	0.0938	0.8681
731	0.0773	0.7364	0.1861

Table 3. Obtained results for the southern city of Hassi Messaoud (three clusters)

In Fig. 12.a and Fig.13.a the items data for Algiers and Hassi Messaoud (northern and southern city respectively) are presented according to their membership degrees in a graphical interpretation. The horizontal axis in each subfigure represents the prototypes obtained by SOM, and the vertical axis in each subfigure represents their associated memberships.

The visualization is obtained using PCA (Principal Component Analysis). The results presented in 3D and 2D are given for Algiers (Fig. 12.b) and (Fig. 12.c) and Hassi Messaoud (Fig.13.b) and (Fig. 13.c.), respectively. X and Y, being respectively, the first and the second component.

The number of identified clusters is clearly shown, in 3-D Fig. 12(b) and 13(b) respectively for Algiers and Hassi Messaoud. For a clearer interpretation, Fig. 12(c) and 13(c) shows the clusters in a 2-D representation. Fig. 12(a) and 13(a), show the membership values for each identified cluster.

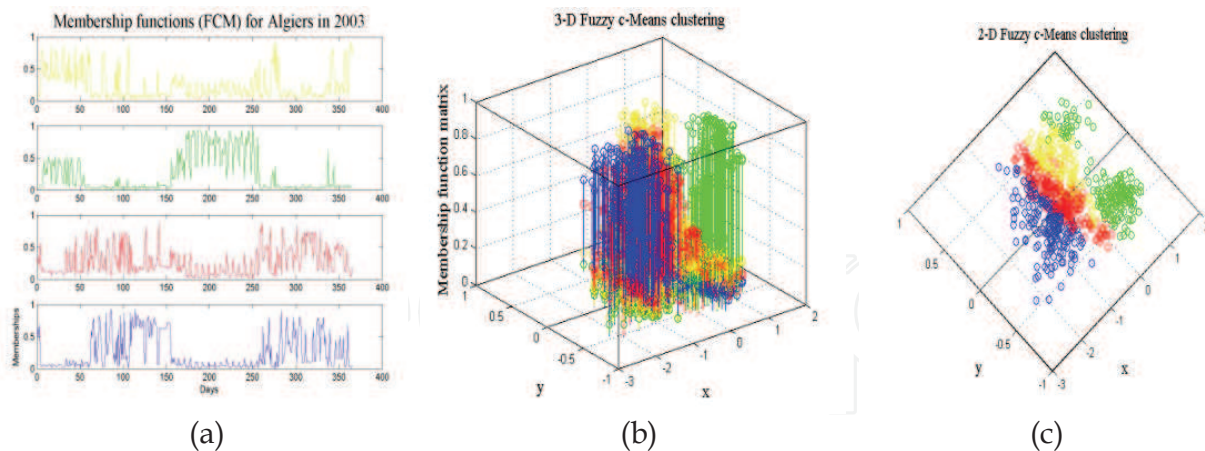


Fig. 12. Graphical results of Fuzzy Clustering for Algiers. (a) Four cluster from top to bottom (C1 to C4). The horizontal axis in each subfigure represents the prototypes obtained by SOM and the vertical axis in each subfigure represents the membership value. (b) and (c) are the visualisation of clusters using PCA respectively in 3D and 2D dimensions. X and Y are respectively the first and second component

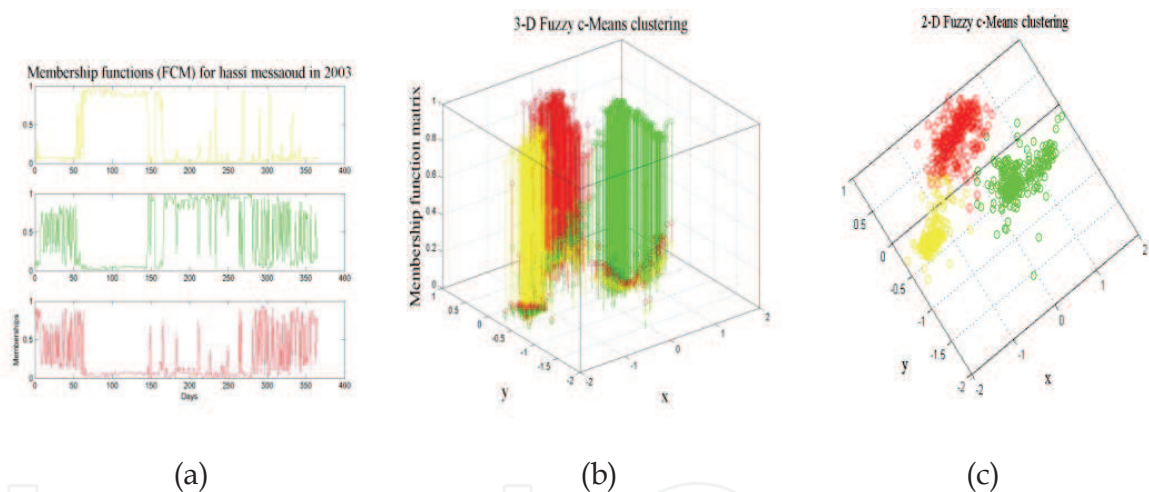


Fig. 13. Graphical results of Fuzzy Clustering for Hassi Messaoud. (a) Three cluster from top to bottom (C1 to C3). The horizontal axis in each subfigure represents the prototypes obtained by SOM and the vertical axis in each subfigure represents the membership value. (b) and (c) are the visualisation of clusters using PCA respectively in 3D and 2D dimensions. X and Y are respectively the first and second component

The day with the highest membership value in a group is the representative day for that group. The representative day for each class for Algiers and Hassi Messaoud are summarised respectively in Table 4 and Table 5.

Table 6 and Table 7 shows respectively the period and classes associated with each cluster coloured and visualized in figure 12 and 13.

Table 8 and 9 indicate for each region and for each day type the level of memberships to the clusters. For example for Algiers, The day corresponding to 01 of July 2003 belongs with a confidence degree of 90,55% to cluster C2 (Summer season) , 5,55 % to cluster C1(Winter season) 2,6 % to cluster C3 (Autumn season) and 1,3 % to cluster C4 (Spring season).

The day corresponding to the 31th January, a winter day belongs with a confidence degree of 84,14 % to cluster C1 (Winter season), 8,87 % to cluster C2 (Summer season), 5,26 % to cluster C3 (Autumn season) and 1,73 % to cluster C4 (Spring season).

Representative Day	30 December 2003	27 march 2003	21 May 2003	25 june 2003
Membership value	0.9099	0.926	0.8741	0.9352

Table 4. Representative day per class for Algiers

Representative Day	01 April 2003	27 May 2003	21 October 2003
Membership value	0.9899	0.9881	0.9462

Table 5. Representative day per class for Hassi Messaoud

Season	Classes	Overlapping	Period
winter	C1 (yellow)	C2,C3,C4	Dec, Jan, Fev
Summer	C2 (green)	C1,C3,C4	July et August
Autumn	C3 (red)	C1,C2,C4	September, October and November
spring	C4 (blue)	C1,C2,C3	March April May June

Table 6. Classes identified for Algiers

Season	Classes	Overlapping	Period
Autumn	C1 (yellow)	C2,C3	Feb March April
Summer	C2 (green)	C1,C3	May June July August
Spring	C3 (red)	C1,C2	Sept Oct Nov Dec, Jan

Table 7. Classes identified for Hassi Messaoud

Day	C1	C2	C3	C4
01/07/2003	0.0555	0.9055	0.026	0.013
31/01/2003	0.8414	0.0887	0.0526	0.0173
22/01/2003	0.7928	0.0799	0.0953	0.032
21/10/2003	0.1645	0.056	0.5202	0.2594
10/08/2003	0.208	0.6406	0.1072	0.0442
10/05/2003	0.102	0.0497	0.3682	0.4802

Table 8. Membership values per class per day for Algiers

Day	C1	C2	C3
31/01/2003	0.07	0.2373	0.6927
20/12/2003	0.0717	0.2314	0.6969

Table 9. Membership values per class per day for Hassi Messaoud

6. Conclusion

This chapter presents day type identification for regional Algerian electricity load using a combination of two of the most popular artificial intelligence paradigms, for instance Kohonen Maps and Fuzzy logic. The Kohonen map analysis allows a rough visual identification of the different existing classes, while the C-means clustering approach identifies existing clusters as well as providing each new day to identify, with a membership value. The choice of the optimal number of clusters is based on four fuzzy cluster validation measures (PC, PE, SC and XB). Four clusters are clearly identified for Algiers and three clusters for Hassi Messaoud, two large representative cities, respectively from northern and southern part of the country.

Fuzzy clustering methods allow objects to belong to several clusters simultaneously, with different degrees of membership. This is more natural than hard clustering, as objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial memberships.

This is an important feature for day type identification to increase sensitivity, allowing a Neuro-Fuzzy forecasting approach, as existing clusters are already identified as potential models for each studied region (city). The subsequent combination between forecasting models may be given by the already determined membership functions.

7. References

- Benabbas, F., Khadir, M.T., Fay, D., Boughrira, A. (2008). Kohonen Map Combined to the K-Means Algorithm for the Identification of Day Types of Algerian Electricity Load, *In Proceedings of The International Conference of Computer Information Systems and Industrial Management Applications*, Ostrava, Czech Republic.
- Bensaid, A.M., Hall, L.O Bezdek, J.C., Clarke, L.P., Silbiger M.L. and Arrington J.A. (1996). Validity-guided (Re)clustering with applications to image segmentation, *IEEE Transactions on fuzzy systems*, 4(2) , 112-123.
- Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum, NY.
- Bezdek, J.C. and Pal, N.R. (1988). Some new indexes of cluster validity, *IEEE Transactions on systems, Man and Cybernetics – Part B*, 28, 301-315.
- Bretschneider, P., Rauschenbach, T., and Wernstedt, J., (1999). Forecast using an adaptive fuzzy classification algorithm for load, *6th European Congress on Intelligent Techniques and Soft Computing*, 3, 1916-1919.
- Cannon, R.L., Dave J.V. and Bezdek J.C. (1986). Efficient implementation of the fuzzy c-means clustering algorithm. *IEEE Trans. Pattern Anal. Machine Intell.* 8, 248-255.
- Chen, S.T. Yu, D.C. Moghaddamjo, A.R. (1992). Weather sensitive short-term load forecasting using non-fully connected artificial neural network, *IEEE Transactions on Power Systems*, 7 (3), 1098-1104.
- Dreyfus G. , Martinez, J.M., Samuelides, M., Gordon, M.B., Badran, F., Thiria, S. and Hérault, L. (2004). *Reseaux De Neurones : Méthodologie Et Application*, Eyrolles ISBN : 2-212-11464-8, France.
- Eschrich, S., Ke, J., Hall, L., and Goldof, D. (2003). Fast accurate fuzzy clustering through data reduction, *IEEE Transactions on Fuzzy syst*, 11(2), 262-270.

- Fay, D., Ringwood, J.V., Condon, M., and Kelly, M. (2003). 24-hour electrical load data -a sequential or partitioned time series? *Journal of Neurocomputing*, 55(3-4), 469-498.
- Fay, D. (2004). *A strategy for short-term load forecasting in Ireland*, Ph.D Thesis, Dept. of Electronic Engineering, Dublin City University, Ireland.
- Halgamuge, S.K. (2005). *Classification and Clustering for Knowledge Discovery (Studies in Computational Intelligence)*, Wang, L. (Eds), Springer, 3540260730, The Netherlands.
- Hoppner, F., Klawonn, F., and Kruse, R. (1999). *Fuzzy Cluster Analysis : Methods for Classification, Data Analysis, and Image Recognition*, New York, Wiley.
- Himberg, A. (2000). SOM Based Cluster Visualization and Its Application for False Coloring, *Proceedings of International Joint Conference on Neural Networks (IJCNN)*, 3, 587-592.
- Hsu, Y.Y. and Yang, C.C. (1991). Design of artificial neural networks for short-term load forecasting Part I: Self-organising feature maps for day type identification, *IEEE Proceedings-C*, 138(5), pp 407-413.
- Hubele, N.F. and Cheng C.S., (1990). Identification of seasonal short-term forecasting models using statistical decision functions, *IEEE Transactions on Power Systems*, 5 (1), 40-45.
- Jain, K. and Dubes, R.C. (1988). *Algorithms for clustering data*, Prentice-Hall, Englewoods Cliffs.
- Jang J. S. R., Sun C. T. and Mizutani, E. (1997). *Neuro-fuzzy and soft computing*, p (426-427), Prentice Hall.
- Jolliffe, I.T., (2002). *Principal Component Analysis*, Springer, 0387954422, New York.
- Khadir M. Tarek, Khedairia Sofiane and Benabbas Farouk, *Kohonen Maps Combined to K-means in a Two Level Strategy for Time Series Clustering Application to Meteorological and Electricity Load data*, Book Title: Self-Organizing Maps, edited by George K Matsopoulos, InTech, Austria, pp 63-80, 2010.
- Kohonen, T. (1990). The self-organising map, *Proceedings IEEE*, 78 (9). 1464-1480.
- Lertpalangsunti N., and Chan C.W., (1998). An architectural framework for the construction of hybrid intelligent forecasting systems: application for electricity demand prediction, *Engineering Applications of Artificial Intelligence*, 11, pp 549-565.
- Muller, H. Petrisch, G. (1998). *Energy and load forecasting by fuzzy neural networks*. In: H. Jurgen, H.J. Zimmermann eds., *Proceedings, European Congress on Intelligent Techniques and Soft Computing*, Aachen, Germany, September. Aachen: Elite foundation, 1925-1929.
- Rousset, P. (1999). *Applications des algorithmes d'auto-organisation à la classification et à la prévision*. Ph.D thesis, University of Paris I, France.
- Szilágyi, S., Szilágyi, L., Iclánzan, D., Dávid, L., Frigy, A. and Benyó Z. (2009). Intensity in homogeneity correction and segmentation of magnetic resonance images using a multi-stage fuzzy clustering approach, *International Journal on non-standard computing and artificial intelligence*, 19(5), 513-528.
- Srinivasan D., Tan S.S., and Chang, E.K. (1999). Parallel neural network-fuzzy expert system for short-term load forecasting: system implementation and performance evaluation, *IEEE Transactions on Power Systems*, 14(3), 1100-1106.
- Vesanto, J. (1999). SOM-Based Data Visualization Methods, *Intelligent Data Analysis*, 3(2), 111-126.

- Vesanto J. and Alhoniemi, E. (2000). Clustering of the Self-Organizing Map, *IEEE Transactions on Neural Networks, special issue on data mining*, 11(3), 586–600.
- Xie, X. L. and Beni, G. A. (1991). Validity measure for fuzzy clustering, *IEEE Transactions on Pattern analysis and Machine intelligence*, 13(8), 841-847.
- Xu, k.C.R. and Haynes, L. (2001). A new data clustering and its applications. *Proceeding of SPIE-the international society for optical engineering*, 4384, 1-5.
- Xu, R. and Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645-678.
- Yu, J., Cheng, Q. and Huang, H. (2004). Analysis of the weighting exponent in the FCM and its application to parameter selection. *IEEE Transactions on Fuzzy systems*, 13(1), 164 - 176.

IntechOpen



Self Organizing Maps - Applications and Novel Algorithm Design

Edited by Dr Josphat Igadwa Mwasiagi

ISBN 978-953-307-546-4

Hard cover, 702 pages

Publisher InTech

Published online 21, January, 2011

Published in print edition January, 2011

Kohonen Self Organizing Maps (SOM) has found application in practical all fields, especially those which tend to handle high dimensional data. SOM can be used for the clustering of genes in the medical field, the study of multi-media and web based contents and in the transportation industry, just to name a few. Apart from the aforementioned areas this book also covers the study of complex data found in meteorological and remotely sensed images acquired using satellite sensing. Data management and envelopment analysis has also been covered. The application of SOM in mechanical and manufacturing engineering forms another important area of this book. The final section of this book, addresses the design and application of novel variants of SOM algorithms.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Khadir M. Tarek and Benabbas Farouk (2011). Kohonen Maps Combined to Fuzzy C-means, a Two Level Clustering Approach. Application to Electricity Load Data, Self Organizing Maps - Applications and Novel Algorithm Design, Dr Josphat Igadwa Mwasiagi (Ed.), ISBN: 978-953-307-546-4, InTech, Available from: <http://www.intechopen.com/books/self-organizing-maps-applications-and-novel-algorithm-design/kohonen-maps-combined-to-fuzzy-c-means-a-two-level-clustering-approach-application-to-electricity-lo>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen