

Fuel Cell Health Monitoring Using Self Organizing Maps

Raïssa Onanena^{a,b,c}, Latifa Oukhellou^{*a}, Etienne Côme E.^a, Samir Jemei^c,
Denis Candusso^b, Daniel Hissel^c, Patrice Aknin^a

^aUniversité Paris-Est, IFSTTAR-GRETTIA, 14-20 Boulevard Newton, F-77447 Champs-sur-Marne, France

^bIFSTTAR, FCLAB Institute, Rue Thierry-Mieg, 90010 Belfort Cedex, France

^cFEMTO-ST UMR CNRS 6174, FR FCLAB (FR CNRS 3539), University of Franche-Comte, Rue Thierry-Mieg, 90010 Belfort Cedex, France

latifa.oukhellou@ifsttar.fr

The problem of durability of fuel cell technology is central for its spreading and commercialization. There is therefore a growing need to build accurate diagnosis tools which can give the operating state of the fuel cell during their use. When supervised machine learning approaches are used to build such diagnosis tools, they generally require a large amount of labeled data. Collection and annotation of data can be either difficult to perform or time consuming. In this paper, authors are interested in the monitoring of fuel cells in an unsupervised framework, meaning that no labels are required to learn the diagnosis model. The aim is to build a monitoring tool able to easily visualize the State Of Health of full cells from electrochemical impedance spectroscopy measures, showing thus its evolution from fault free case ("normal" behaviour) to defective classes such as drying or flooding. The proposed approach is based on Self Organizing Maps (SOM) which have shown their performance to solve fault detection and prediction in many industrial systems. By automatically visualizing the data into a two-dimensional space, the interpretation of the results have become easy and instinctive. The approach also allows the clustering of the data into different groups of classes, thus enabling the classification of new observations. Experimental results carried out on real data sets have shown the efficiency of the proposed approach with respect to standard supervised and unsupervised classification approaches.

1. Introduction

The problem of durability of fuel cell technology is central for its spreading and commercialization. There is therefore a growing need to build accurate diagnosis tools which can give the operating state of the fuel cell during their use. Various diagnosis approaches have been proposed for that purpose. These various approaches include model-based approaches (Burford et al. 2004) and gray or black-box model approaches using fuzzy logic (Hissel et al, 2004), neural networks (Nitsche et al, 2004). Non-parametric identification by Markov parameters have also been presented (Tsujioku et al, 2004), a fuzzy-clustering-based FC diagnosis approach (Hissel et al, 2007), as well as a diagnosis approach using bayesian networks (Wang, et al, 2012). As for pattern-recognition-based method authors have proposed in a previous study a supervised machine learning approach for FC diagnosis using Electrochemical Impedance Spectroscopy (EIS) measurements. When using supervised machine learning approaches for diagnosis tools, it is required to have labeled data, i.e. an a priori knowledge about the intrinsic structure of the dataset must be known. Collection and annotation of these labels can be difficult to perform and time consuming, and sometimes the possibility of mislabeling the data is non-negligible. Therefore, it might be an interesting idea to consider the unsupervised learning framework to handle and study data. In this paper, the state-of-health monitoring of PEM fuel cells using an unsupervised learning method is investigated. In unsupervised learning problems, the main goal when considering our task is to discover groups or similarities in the data set and no label is required to learn the diagnosis model. The aim here is to build a monitoring tool able to easily visualize the health of full cells from electrochemical impedance spectroscopy measures, showing thus its evolution from fault free case ("regular" behaviour) to defective classes such as drying or flooding. The proposed approach is based on Self Organizing Maps (SOM)

which have shown their performance to solve fault detection and prediction in many industrial systems such as vehicle cooling system (Svensson et al, 2008) or aircraft engine fleet monitoring (Come et al, 2010). By automatically visualizing the data into a two-dimensional space, the interpretation of the results have become easy and instinctive.

The paper is organized as follows. First of all, in section 2, the database which is used for diagnosis is presented. Then in section 3, the self-organizing maps will be briefly introduced. And finally in section 4, the results applying the Self-Organizing Maps will be presented and discussed.

2. Database

The database that was used in the presented work was built after experimentations on a 20-cell PEMFC stack which has been operated under various operating conditions, comparable to real-life conditions. The stack was assembled with commercial perfluorosulfonic Membrane-Electrodes Assemblies (MEA) graphite bipolar plates and electrodes with 100cm² area. The experiments and EIS measurements have been performed using a 10kW test bench developed in-lab (Wasterlain et al, 2009).

The experimentally generated operating conditions were realized using a 2-level- fractional experimental design. This experimental design deals with the impact of five factors on the operating conditions:

- anodic Relative Humidity (RH_a) [35 %, 75 %]
- cathodic Relative Humidity (RH_c) [35 %, 75 %]
- anodic Stoichiometric Factor (FSA) [1.8, 3.0]
- the cathodic Stoichiometric Factor [2, 3] and
- the FC temperature.

The stack is then operated under these various conditions and faults from drying to flooding are triggered by varying one or several parameters. 11 experimental points have been considered during the experiments where EIS spectra were measured (Wasterlain et al, 2009;).. The current density was fixed to 0.5 A/cm², The whole experiments have enabled us to gather a database of EIS spectra, as presented in Table 1.

Table 1: Experimental points used to generate the dataset

Test Nb.	1	2	3	4	5	6	7	8	9	10	11
RH _a (%)	35	35	35	35	35	75	75	75	75	75	75
RH _c (%)	35	75	75	75	75	35	35	75	75	75	75
FSA	3	1.8	1.8	3	3	1.8	3	1.8	1.8	3	3
FSc	3	1.8	1.8	3	3	1.8	3	1.8	1.8	3	3
T (°C)	60	60	80	80	60	60	80	80	60	60	80
Nb of EIS	20	21	21	21	21	21	21	21	21	21	21

As a pre-processing step, a feature extraction was realized on the data. This step was achieved because of its potential benefits on the performances of the diagnosis tools. These include:

- Facilitating data visualization and understanding.
- Reducing computational times
- Defying the curse of dimensionality.

The extraction approach chosen was to take the real and imaginary parts of the spectra at the following frequencies : 5 kHz ($Re_{(f = 5 \text{ kHz})}$, $Im_{(f = 5 \text{ kHz})}$), 500 Hz ($Re_{(f = 500 \text{ Hz})}$, $Im_{(f = 500 \text{ Hz})}$), 50 Hz ($Re_{(f = 50 \text{ Hz})}$, $Im_{(f = 50 \text{ Hz})}$), 5 Hz ($Re_{(f = 5 \text{ Hz})}$, $Im_{(f = 5 \text{ Hz})}$), 77.5 mHz ($Re_{(f = 77.5 \text{ mHz})}$, $Im_{(f = 77.5 \text{ mHz})}$), 50 mHz ($Re_{(f = 50 \text{ mHz})}$, $Im_{(f = 50 \text{ mHz})}$), (12 features)

Therefore, for the following clustering step, the database is made up of 12 input variables.

3. SOM methodology

A self organizing map (SOM) is a type of Artificial Neural Network (ANN) that is trained within an unsupervised learning framework (no class labels are required). SOM are well-known data mining tools which are frequently used to project high-dimensional data into a low dimensional space that is supposed well suited to capture the structure of the original data. Therefore the map that results to this training shows the different clusters that are present in the data. Like most Artificial Neural Networks, SOMs operate in two modes: training and mapping.

A Self-Organizing Map (SOM) consists of neurons (or units) on a regular low-dimensional grid which defines a neighborhood relation between the prototypes. Each unit t of the grid is represented by a N-

dimensional weight vector or prototype vector $w^i = [w_1^i, w_2^i, \dots, w_N^i]$, where N is equal to the dimension of the input vectors. The neurons are connected to adjacent neurons by a neighborhood relation, which dictates the topology or structure of the map. Self-organization means that after convergence prototypes which are close in the grid will also be close in the original data space. New data can be projected after learning of the SOM by assigning them to the location on the map of the best matching unit (BMU): the closest prototype vector to the data point. The SOM training algorithm resembles vector quantization algorithms such as k-means (Webb, 2011). The important distinction is that in addition to the best matching unit vector, also its topological neighbors on the map are updated: the region around the best matching unit is stretched towards. The aim is to find a suitable set of weights for each unit so that the network estimates the distribution of the input data.

The learning rule responsible for finding a suitable set of weights iterates the two following steps:

Competitive step: given an input vector $x = [x_1, x_2, \dots, x_N]$, the distance between unit t of the SOM and the input vector is calculated by the Euclidian metric. The unit with the smallest distance is the one which closely represents the current input, and thus is considered the winner or the best matching unit (BMU) for that input:

$$\|x - w^{BMU}\| = \min_i \{\|x - w^i\|\} \quad (1)$$

Cooperative step: the weights of the winning unit are moved towards the input. In addition, the winning unit's neighbors are also moved in this direction, but by an amount that decays with the distance of the neighbors from the winning unit. The update rule for the weight vector of unit i is given by the following:

$$w^i(t+1) = w^i(t) + \alpha(t)h_{BMU,i}(t)[x - w^i(t)] \quad (2)$$

where t denotes time, $\alpha(t)$ is the learning rate and $h_{BMU,i}(t)$ is a neighborhood kernel centered on the winner unit. A Gaussian kernel is usually chosen:

$$h_{BMU,i}(t) = e^{-\frac{\|r_{BMU} - r_i\|^2}{2\sigma^2(t)}} \quad (3)$$

where r_{BMU} and r_i are the positions the BMU and unit i respectively on the map, $\sigma(t)$ the neighborhood radius. Both the radius and the neighborhood rate decrease monotonically with time (Vesanto, 2000).

From the initialization phase to the end of the training, the above process is iterated for each input vector of the data set and this, results in a competition between different regions of the input space for units of the map. Dense regions of the input space will tend to attract more units than sparse ones, with the distribution of units in the weight space ultimately reflecting the distribution of the input data in the input space. Neighborhood learning also encourages topology preservation, meaning that the map end up close in the weight space, as well.

4. Results and discussion

4.1 Data clusters

The first experiment aims at demonstrating the ability of SOM to perform a clustering of the dataset. The aim is to identify in the obtained SOM, groups of clusters where each cluster corresponds to an operating class. The training of the SOM was performed using the Matlab SOM toolbox. The learning was done under the default setting: the learning algorithm was the batch algorithm with linear initialization.

Figure 1 presents the results obtained with the classical SOM. The topology of the map was fixed manually: it contains 6 x 12 neurons. And therefore, if each map unit is initially considered as a cluster, it results in a 72 class-clustering. A more in-depth analysis of the map could be done using the U-(matrix) (Utsch, 2003). The U-matrix represents the average distance between the map unit and its closest neighbor; it is subsequently an interesting method for the identification of groups of clusters among the map unit. The scale on the left side of the U-matrix gives information about the distances in the U-matrix. In Figure 1, the redder colors show long distances between the nodes whereas the bluer colors show small distances between the nodes, therefore establishing clusters in the dataset. These figures exhibit 3 or 4 main clusters.

In order to validate the visual assessment, the validation of clusters can be done through criteria such as compactness (members of each clusters should be as close to each other as possible) and separation (the clusters should be widely separated) (Legány et al, 2006). Six validity indexes were tested (see Figure 2), corresponding to the some of the most widely used indices. As it can be seen in Figure 2, this seems to indicate that the number of clusters in the dataset, as indicated in the SOM is likely to be 3 or 4.

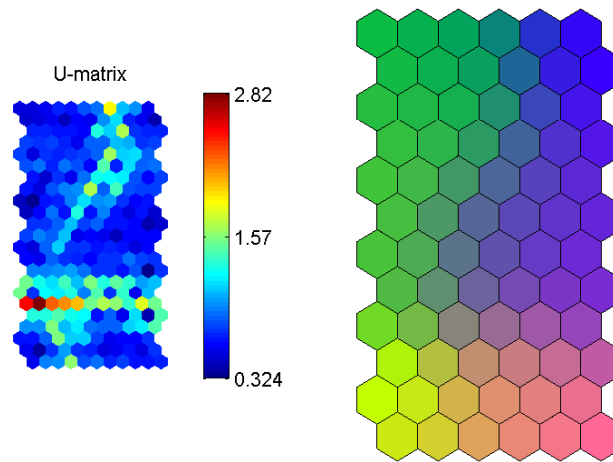


Figure 1: U-matrix obtained when training a SOM on the FC data

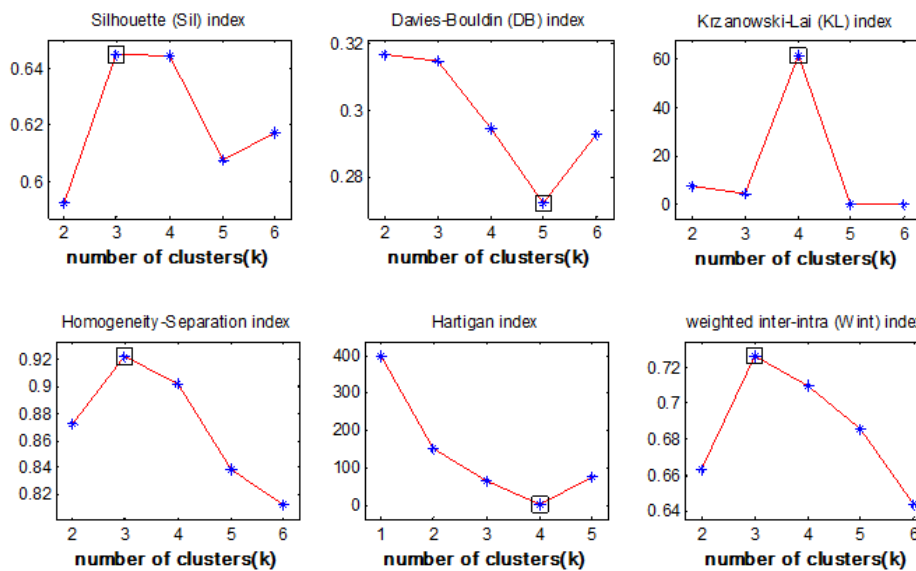


Figure 2: Evaluation of different cluster validity indices to automatically determine the number of clusters

4.2 Data labeling

In this second experiment, the SOM is used to identify mislabeled data. With the experts' advices, preliminary labels were associated with data points given the prior knowledge we have on the operating conditions. These preliminary classes or fault labels are presented in Figure 3. This figure shows the distribution of the data on the map according to this fault labeling. A part of the data seems to be mislabeled. For example, some data labeled as "minor drying" are clustered with data labeled as "flooding". Moreover, it seems that the gravity of the fault grows with the value of the principal components. With the assumption these observations and the assumption that data generated with similar fault operating conditions will belong to the same cluster, labels changes were therefore made accordingly. The new distribution of labels in the data set and on the map is presented in Table 2 and Figure 4. This shows that SOM can be considered as useful tools for visualizing high dimensional data.

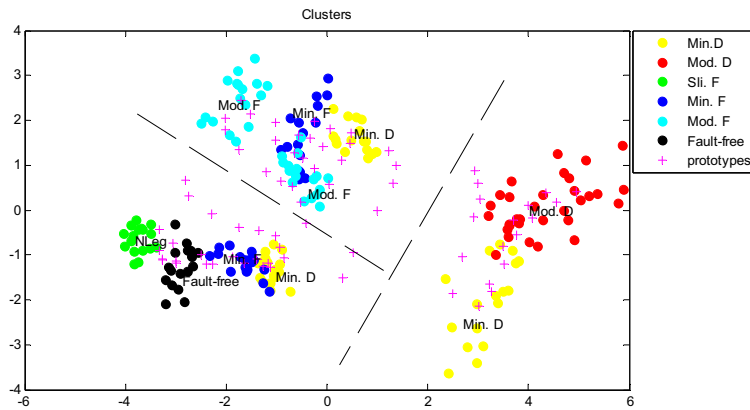


Figure 3: The labelling of the data as initially done by the experts. A group of data point labelled as a drying fault are clustered with data points whose operating conditions are identified as flooding

Table 2: Table of different labels of the data, with the fault gravity

Fault mode	Drying		Fault-free	Flooding		
gravity	moderate (Mod.D)	minor (Min.D)		moderate (Mod.F)	minor (Min.F)	slight (Sli.F)
AssociatedTest Number	1, 6	7	3	2, 4, 8, 10	5, 9	11

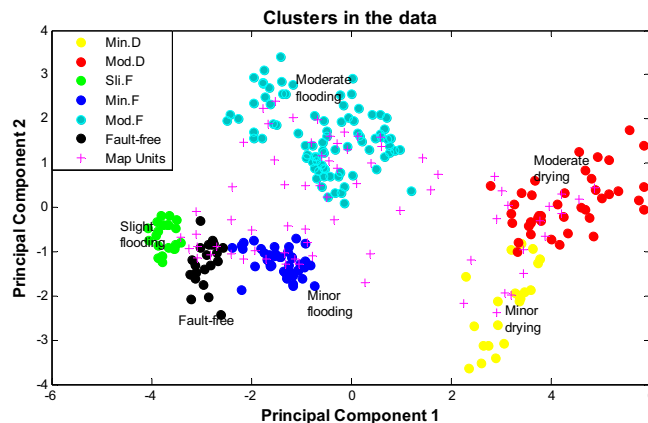


Figure 4: The labelling of the data as finally obtained after the changes of the labels. The modifications were done so that the "new" label corresponds to the labels of its neighbours by a majority vote.

4.3 Fault monitoring

Finally, the SOM has been used for the monitoring of the fuel cell. Two points of the data set, that represent the measurement of the FC operations from flooding conditions to drying conditions are presented on the map. In Figure 5, it is possible to see how the point evolves in the map. The point evolves from top to down, which corresponds to the evolution from flooding operating conditions to drying conditions.

5. Conclusion

In this paper, a non-supervised approach based on Self Organizing Maps is used to monitor fuel cell. We have presented how SOM can be used to visually identify the main clusters in the dataset, then how it could be used to correct eventual expert data mislabeling. Finally experiments are carried out to show that self-organizing map could be used to monitor the State-Of-Health of the fuel cell. This presents an

alternative for supervised learning if the collecting of labels could not be done under difficult to obtain or is time consuming.

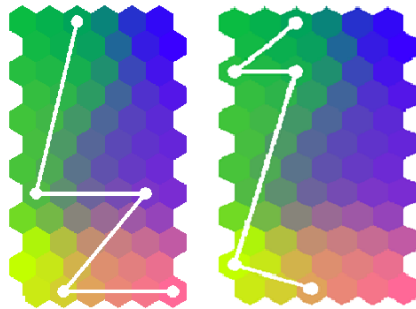


Figure 5: the different labels of the map prototype as learned during the training phase. On the right-hand side, we have the labels on the units after the modification of the mislabelled data

References

- Burford D., Davis T., Mench M., 2004, Real-time electrolyte temperature measurement in an operating polymer electrolyte fuel cell, Proceedings of the Advanced Materials for Fuel Cells and Batteries Symposium, 204th Electrochemical Society Meeting.
- Côme E., Cottrell M., Verleysen M., Lacaille J., 2010. Aircraft engine fleet monitoring using Self Organizing maps and edit distance, 18th European Symposium on Artificial Neural Networks, Bruges, Belgium.
- Hissel D., Péra M. C., Kauffmann J.M., 2004, Diagnosis of automotive fuel cell power generators, Journal of Power Sources, 128(2), pp.239-246.
- Hissel D., Candusso D., Harel F., 2007, Fuzzy Clustering Durability Diagnosis of Polymer Electrolyte Fuel-Cells dedicated to transportation applications, IEEE Transactions on Vehicular Technology (56), pp.2414-2420.
- Kozlowski, J.D., Byington, C.S., Garga, A. K., Watson M. J., Hay T. A., 2001, Model-based predictive diagnostics for electrochemical energy sources, IEEE Aerospace Conference.
- Legány C., Juhász S., Babos A., Cluster validity measurement techniques. Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, pp. 388-393.
- Svensson M., Byttner S., Rognvaldsson T., 2008, Self-organizing maps for automatic fault detection in a vehicle cooling system, 4th International IEEE conference on Intelligent Systems, vol 3, pp. 8-12.
- Tsujioku, Y., Iwase, M., Hatakeyama S., 2004, Analysis and modelling of a direct methanol fuel cell for failure diagnosis, IEEE IECON Conference, Busan, Korea.
- Ultsch A., 2003: U-Matrix: A Tool to visualize Clusters in high dimensional Data, Dept. of Computer Science University of Marburg, Research Report 36.
- Vesanto J., Himberg J., Alhoniemi E., Parhankangas J., 2000, SOM Toolbox for Matlab 5 Documentation. Helsinki University of Technology, Helsinki, Finland.
- Wang K., Pera M.-C., Hissel D., Yousfi Steiner N., 2012, SOFC modelling-based on discrete Bayesian network for system diagnosis use, Proceedings of the 8th Power Plant and Power System Control Symposium (PP&PSC) 2012, Toulouse, France, September 2 – 5.
- Wasterlain, S., Harel F., Candusso D., Hissel D., 2009, Development of a High Voltage Impedance Spectrometer and Diagnosis of Large PEFC stacks, European Fuel Cell Forum, Lucerne, Switzerland.
- Webb A., Copsey K.D., 2011, Statistical Pattern recognition, 3rd edition John Wiley and Sons.