Black-Box Fuzzy Identification of a Nonlinear Hydrogen Fuel Cell Model

A.M.A. BERTONE1*, J.B. MARTINS2 and K. YAMANAKA2

Received on November 15, 2016 / Accepted on August 28, 2017

ABSTRACT. A fuzzy identification of the system’s dynamic is developed with a data generated by a hydrogen fuel cell simulator. The data obtained is single input/single output, without having previous knowledge of the system model, and showing nonlinear behavior. The choice of the fuzzy method for identification is based on those particular data features, and the malleability of the mathematical fuzzy technique. The objective of the fuzzy identification is to reach an analytic formula for a better understanding of the cause-effect relationships of the data, followed by its validation. The dynamic system identification process is performed using fuzzy clustering through the Gustafson and Kessel algorithm, followed by a Takagi and Sugeno fuzzy inference method. The \( k \)-fold technique, is the cross validation tool, used to confirm the lack of data over-training. The novelty of this approach covers mathematical and engineering features that makes this study interdisciplinary. For the mathematical contribution, there is a three-dimensional graphic interpretation of the data clustering geometry, obtained through own code computer simulations. Concerning to the engineering context, the novelty is based on the use of the fuzzy approach to the hydrogen fuel cell. Both contributions have no precedent in the literature. The results of the fuzzy identification show high reliability in terms of cross validation, making the fuzzy approach a promising tool for black-box identification. Combining this technique with others will provide powerful instrument for industrial problems.

Keywords: hydrogen fuel cell, fuzzy clustering, identification of dynamical systems, Takagi Sugeno inference method.

1 INTRODUCTION

A fuel cell uses the chemical energy of hydrogen or another fuel to cleanly and efficiently produce electricity. The first fuel cells were invented in 1838 [19]. A century later, NASA used fuel cells for programs searching to generate power for satellites and space capsules. Since then, fuel cells have been used in many other applications, being unique in terms of the variety of their potential applications. They can provide power for systems as large as a utility power station, and as small as a laptop computer. Besides, fuel cells have lower emissions than combustion engines, there are no carbon dioxide emissions, as a consequence, and no air pollutants [23]. These motivations...
have encouraged Avelar [2] to study the functioning of a hydrogen fuel cell with a full-bridge elevator converter, and a single-phase inverter sinusoidal Pulse Width Modulation, manipulation of the pulse width of a square wave to control the power of a particular device. The high reliability of Avelar’s software, allows us to recreate the data, used for identification in this study. The details of Avelar’s simulation software is done in Section 2.

A system identification is a mathematical technique that seeks a model that reproduces, in part, the real dynamic of the data [1]. Among the analysis of linear and nonlinear systems can be included: White box, for physics modeling; gray box, for models with inputs, outputs, and additional information about the system; and black box, only composed by inputs and outputs, without any additional information. This study refers to the latest type, having as only knowledge the input and output data depending on time, besides its nonlinear behavior. There are several methods for the approach of nonlinear systems, among which could be highlighted: neural networks [21], and functional analysis [10], initially conceived by Wiener [22]. The main difference between these methods is the way that the system is considered: global or local [4]. An example of such identification method is the structure NARX (Nonlinear AutoRegressive with eXogenous input) [1, 13]. Local approaches attempt to divide the system into simpler subsystems, generally of the linear type. Fuzzy modeling techniques can be classified as local approximation.

The fuzzy identification used in this research is based in the fuzzy clustering of the data, combined with a Takagi and Sugeno fuzzy inference method. In 1979 Donald E. Gustafson and William C. Kessel published the article [9] presenting a modified Fuzzy C-Means traditional algorithm [5]. Nowadays this algorithm has become known as Gustafson-Kessel (GK), being the basis of a family of fuzzy clustering algorithms as described in [4,7,11,14,16,18].

From the mathematical point of view, at the clustering stage of identification, this study has explored the geometry of the clustering qualitatively, performing graphics that are not common in the literature. At the identification’s stage of fuzzy inference, the main objective is to obtain an analytic formula that is faithful to the dynamics of the hydrogen fuel cell. Besides, it applies the rigorous k-fold method to test the accuracy of the raised model. These two stages, besides their novelties, show the interdisciplinary aspect of this research, which is one of the main contributions. Another innovation is the fuzzy approach for the hydrogen fuel cell model identification. The procedure of these two stages are sketched in Section 3.

The cross-validation method known as k-fold is used for k = 4. This method yields 80.35% of accuracy, based in the Variance Accounted For (VAF) index, a result that demonstrates the ability of the analytic formula obtained to describe the hydrogen fuel cell system dynamic. The codification of the fuzzy clustering and fuzzy identification is developed in C language. The graphics are obtained through the connection between C and the Spyder-Python platform, open source integrated development environment. The graphics can be checked in Section 3. To complete the description of the organization of the work, in section 3, the concept of fuzzy clustering and the Takagi-Sugeno fuzzy inference method are delineated.

The organization of the study finishes with the conclusion expounded in Section 4.
2 HYDROGEN FUEL CELL SIMULATOR

The functioning of a hydrogen fuel cell with an elevator converter full-bridge, and a single-phase inverter sinusoidal PWM, is shown in the Avelar’s work [2] that has created an efficient tool for the study of this type of model. The proposed model was developed for a cell of 1.2 kW and the system identification process used is least squares.

Avelar [2] has developed a model to simulate the hydrogen fuel cell connected to the grid, taking into account the effect of internal temperature. This software is used in to provide the system output, simulating the real plant operation. The main features of the simulator software are:

- the cell is capable to generate voltage between 20V and 50V (continuous).
- the voltage varies with the current drawn at its terminals being raised to 380V before being applied to the inverter module.
- all data is stored in a microcomputer.
- the plant characteristics are different to rise and drop as the heating is faster than cooling.
- a cell polymeric membrane fuel (PEMFC Nexa, manufactured by Ballard) is tested in the laboratory for static and dynamic survey to survey the model.
- all the fuel cell structure is simulated using PSIM® software, an electronic circuit simulation software package, designed specifically for use in power electronics, and motor drive simulations.
- the model presents reasonable performance to represent the plant’s dynamics (current-temperature).

A process scheme applied by Avelar [2] is shown in Figure 1, and the data from this study are generated by the simulator described in [2].

![Figure 1: Alternative energy source based on fuel cell (hydrogen). Source: [2].](image-url)
The type structure is SISO (single input/single output), or only one input and one output. 38,300 are input data (current) with their 38,300 output data (temperature) matching. The current values range from 24.9034A to 1.24136A, which A stands for Ampère unit; temperature values ranging from 0°C to 54.835°C. Input data has been generated in order to traverse the largest possible frequency range, with sampling rate equal 0.2 seconds. The laboratory where the collection is made [2] is shown in Figure 2.

Figure 2: The laboratory where has been obtained the fuel cell hydrogen data. Source: [2].

The gray-box identification studies seek to show important information at the moment of perform the estimation parameters; this modeling is used in this research using data software simulator [2]. It is important to emphasize the efficiency of the simulator demonstrated by tests which, in comparative performances, is shown in Figure 3.

3 RESULTS OF THE FUZZY IDENTIFICATION METHOD FOR THE HYDROGEN FUEL CELL DYNAMIC SYSTEM MODEL

As part of the fuzzy identification technique, the clustering of the data is performed. In general, clustering is a classification of non-supervised of data, forming groups of elements called clusters. The objective of this technique is the organization of the patterns represented by vector, or points, in the multidimensional space, according with a similarity measure. In the process of clustering we have the following components:

1. definition of a similarity between elements. In this case a distance induced by a symmetric matrix, positive is defined to the center of the cluster;

2. defining the criteria for the element belonging or not to the cluster. This measure could be either crisp (classic partition) or fuzzy. In the former case, the measure is a membership function and the pattern is defined by the degree of membership of the element to the cluster.
Figure 3: Comparison between the cell temperature measurement (Stack, T), and simulated (T_{cell}) with digitalized model. Variable parameters are used with the rise and fall of the current after the optimization model. Source: [2].

The fuzzy clustering allows that the elements of the data belong to different groups simultaneously with different membership degrees. Finally, a number c of fuzzy sets is obtained. Other requirements of the clustering process are

- Each element has to have a membership degree of each cluster;
- The sum of all membership degrees of the element to each cluster is exactly one.

Mathematically describing the process of our study, consider the vector

\[ z_j = (z_{j1}, z_{j2}, z_{j3}) \in \mathbb{R}^3, \quad j = 1, \ldots, N, \]

where \( z_{j1} \) represents the input of the system, with \( z_{j2} \) corresponding to the output at the time \( z_{j3} \), and \( N = 38.300 \). In the first step of clustering through the GK algorithm, consider the following elements:

- The Euclidean distance as a measure of similarity;
- A set, \( V_{init} \), of center of cluster which are chosen from the set \( \{z_j\}_{j=1}^{38,300} \) in a way that the centers are well distributed among the data.
With those elements as a starting point, the next purpose of the clustering process is to find the entries of the matrix
\[ U = \begin{bmatrix}
\mu_{11} & \cdots & \mu_{1k} & \cdots & \mu_{1N} \\
\mu_{21} & \cdots & \mu_{2k} & \cdots & \mu_{2N} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
\mu_{c1} & \cdots & \mu_{ck} & \cdots & \mu_{cN}
\end{bmatrix}. \]

For this end, it is formulate an optimization problem given by
\[
\min \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{mj}^m d_{ij}^2
\]
where \( v_i \) is an element of \( V_{ini} \) for \( i = 1, \ldots, c \),
\[
d_{ij}^2 = d^2(z_j, v_i) = (z_j - v_i)^T I (z_j - v_i), \quad i = 1, \ldots, c, \quad j = 1, \ldots, N,
\]
and \( I \) is the identity matrix. The optimization problem has the following constrains:
\[
0 \leq \mu_{ij} \leq 1, \quad i = 1, \ldots, c, \quad j = 1, \ldots, N \quad \text{(degree of membership)};
\]
\[
\sum_{j=1}^{N} \mu_{ij} > 0, \quad i = 1, \ldots, c \quad \text{(non emptiness of the cluster)};
\]
\[
\sum_{i=1}^{c} \mu_{ij} = 1, \quad j = 1, \ldots, N \quad \text{(total degree of membership)},
\]
where \( m \) is known as the fuzzy parameter of the clustering [4]. The number \( m \) determines how much fuzziness is wanted for the clustering. It has been taken for this study \( m = 2 \).

Once the optimization problem (3.1) is solved using the Lagrange Multipliers, the membership function \( \mu_{ij} \) is obtained for each cluster, which, at each element of the data \( z_k \), is given by
\[
\mu_i(z_k) = \left( \frac{d_{ik}^2}{\sum_{j=1}^{N} d_{jk}^2} \right)^{\frac{1}{m}}, \quad i = 1, \ldots, c, \quad k = 1, \ldots, N.
\]

As a convenient choice, the new clusters’ centers are chosen as the center of mass of the membership function \( \mu_{ij}^m \) given by
\[
v_i = \frac{\sum_{j=1}^{N} \mu_{ij}^m z_j}{\sum_{j=1}^{N} \mu_{ij}^m}.
\]
Therefore, a set \( V = \{v_1, v_2, \ldots, v_c \} \), has been constructed \( v_i \in \mathbb{R}^3 \), that represents the new clusters’ centers. Next, considering for each \( i = 1, \ldots, c \) the matrix \( A_i, 3 \times N \) and \( B_i, N \times 3 \), given by

\[
A_i = \frac{\mu_{ij}}{N} (z_j - v_i)^T, \quad B_i = (z_j - v_i),
\]

we define the matrix \( 3 \times 3 \) given by \( F_i = A_i \cdot B_i \), where \( \cdot \) is operator product of matrix. The matrix \( F_i \) is a \( 3 \times 3 \) which entries represent the variance of the memberships of the elements of the cluster \( i \) through \( \mu_{ij}^m \). This matrix is known as the covariance matrix of the cluster \( i \). Finally, the matrix that induces the new distance in the GK process is defined by

\[
M_i = \det(F_i)^{\frac{1}{3}} F_i^{-1},
\]

that determines the new geometry features of the cluster with center \( v_i \). From the matrix \( M_i \), the updated distance measure is defined as

\[
d_{ij}^2 = (z_j - v_i)^T M_i (z_j - v_i), \quad i = 1, \ldots, c, \quad j = 1, \ldots, N.
\]

The procedure can be repeated, ending when an established criteria for stop, \( s \), is set, and the error

\[
\text{error} = \max \left\| U_{ij}^k - U_{ij}^{k+1} \right\|,
\]

is less than \( s \), where \( U_{ij}^k \) represents the entries of the memberships degrees of the data \( z_j \) to the cluster \( i \) at the step \( k \) of the GK algorithm.

After the fuzzy clustering process is finished, the fuzzy inference of Takagi and Sugeno (TS) [20]. An example of the fuzzy identification method is shown in Figure 4, supposing two clusters, \( C_1 \) and \( C_2 \), obtained by the clustering process.

Figure 4: Identification algorithm using clustering and TS for two clusters. Source: by the authors.
Projecting the entries of the matrix $U$ in a plane with the values of the entry as elements in the edge $x$, a membership function is obtained for each cluster, which is named $\mu_1$ and $\mu_2$ respectively for $C_1$ and $C_2$. $x'$ is an input element, associated to the values $\mu_1(x') = \omega_1$ and $\mu_2(x') = \omega_2$. On the other hand, the clustering process is determined by the edge in the direction of the entry data for each cluster, for instance $y_1 = a_1x + b_1$ for cluster $C_1$, and $y_2 = a_2x + b_2$ for cluster $C_2$. The rules established by the Takagi-Sugeno inference method are for this example

$$\text{if } (x', \mu_i(x')) \text{ then } y_i(x') = a_ix' + b_i, \quad i = 1, 2.$$ 

Finally, the defuzzificated output is giving by the sum giving by $\sum_{i=1}^{2} \omega_i y_i(x')$, which is a convex combination of the values $y_i(x')$ by the total degree of membership restriction in (3.3).

The model performance is evaluated by the value in percentage of variance (VAF, Variance Accounted For) which is given by the expression

$$\left(1 - \frac{\text{variance(real output)}-\text{variance(model output)}}{\text{variance(real output)}}\right) \times 100. \quad (3.8)$$

The closer the value (3.8) is to 100%, the better is the performance of the approximation. The decision of the best number of clusters to perform the GK algorithm, is based on the best overall VAF. Once the decision of the number of clusters is made, and a dynamic formula is obtained, next step is to test the efficiency of the approach, through the cross-validation, so call $k$-fold method [6]. The main objective of this method is to verify the model behavior on multiple data groups, to be used for training and testing. In this work a 4-fold is applied for the validation of the data. The procedure is given as follows:

- The data group is divided into four parts, group A, B, C, and D, composing by 9,575 inputs each, from the total of 38,300 data inputs;
- In the first step, the data groups B, C, and D are together for training, and the group A for validation. The VAF rate is calculated corresponding to this simulation. This number is calculated VAF1;
- A new arranging of the groups A,B,C, and D is obtained, making the circulate arrange of the letters, so that, in the second step, the group D is the validation data and the groups A, B, and C are the training data. The calculated validation rate is VAF2;
- By repeating the rearranging of the groups by circulation, the the validation rates VAF3 and VAF4 are obtained;
- The arithmetic average $\frac{\text{VAF1} + \text{VAF2} + \text{VAF3} + \text{VAF4}}{4}$ is the validation rate for the model.

An illustration of the method is shown in Figure 5.
Summarizing the process, the entire algorithm to obtain the fuzzy identification of the model, as has been detailed in this section, follows the steps shown in Figure 6.

The fuzzy clustering, through the GK algorithm is performed for 20 clusters, numbers are chosen in the range [3,30], due to the higher overall VAF=85.87%. It is shown in Figure 7 the three-dimensional graphic of the data (input,output, time), and the centers of the 20 clusters.

One of the clusters and its level of membership degree is shown in Figure 8. Each layer is a surface correspond to a degree of membership. For instance, the outside darkest surface is the level 0.2; the darkest surface in the center is the level 1.
Figure 7: The graphic of the data (input, output, time) together with the centers of the 20 clusters. Source: by the authors.

Figure 8: The graphic of first cluster constructed with the GK algorithm. The minor ellipsoid in black corresponds to the membership degree 1. The scale on the right side specifies the corresponding membership degree of the ellipsoid surface. Source: by the authors.

The 20 clusters projection on the plane (input, output) is shown in Figure 9. The different gray colors indicate the membership degree of the points to the cluster.

Once made the system identification through the algorithm, developed in C, a cross validation technique is applied. The results of the 4-fold validation technique is shown in Table 1.

Table 1: The results of the 4-fold validation technique. Source: by the authors.

<table>
<thead>
<tr>
<th>VAF1</th>
<th>VAF2</th>
<th>VAF3</th>
<th>VAF4</th>
</tr>
</thead>
<tbody>
<tr>
<td>85.87%</td>
<td>84.15%</td>
<td>70.68%</td>
<td>80.69%</td>
</tr>
</tbody>
</table>

The accuracy rate for the fuzzy identification procedures is the arithmetic average

$$\frac{VAF1 + VAF2 + VAF3 + VAF4}{4} = 80.35\%.$$  

The fuzzy identification result compared with the real data is shown in Figure 10.

4 CONCLUSION

In order to identify simulated data of a hydrogen fuel cell dynamics, a fuzzy technique is applied. Through a fuzzy clustering, followed by a Takagi and Sugeno inference method, is generated a model that recreates the dynamics of the single input/single output system, and returns an analytic expression for the relationship between the entries of the data.

It is important to notice that in this type of system identification, black box type, there is no previous information about the model. Furthermore, the behavior that the data show is strongly
nonlinear. Fuzzy technique has the proper pliability for these type of models, which allows to reach faithful formulas, and with capacity to describe the real phenomenon. The identification process that has been applied in this work is divided into three main parts: the fuzzy clustering, a Takagi and Sugeno inference method, and a cross validation of the type $k$-fold. The chosen validation, with $k = 4$, allows the elimination of over training, which reached the value VAF up to 80.35%.

It is important to highlight, as a mathematical contribution of this work, the three-dimensional graphic, which is an interpretation of the geometry of the data clustering. The engineering relevance of the study rests in the use of fuzzy identification in data from a hydrogen fuel cell. The two novel aspects of this proposal are, to the best of our knowledge, uncommon in the literature. Fuzzy methodology has shown good performance for identification tasks, validated by the $k$-fold technique, specially in cases when there is little information about the process, as is the case of this study. The improvement on the codification, for better performance in accuracy and computer speed, is a research currently is progress. Furthermore, a future survey is among the purposes, involving a comparative study with other techniques and its combinations applied to same and different data.

RESUMO. Uma identificação fuzzy da dinâmica de um sistema é desenvolvida com dados gerados por um simulador de uma célula de combustível de hidrogênio. Os dados obtidos são do tipo de entrada única – saída única, sem conhecimento prévio do modelo do sistema, e mostrando um comportamento não-linear. A escolha do método fuzzy para identificação baseia-se nessas características específicas dos dados e na maleabilidade da técnica mate-
mática fuzzy. O objetivo da identificação fuzzy é chegar a uma fórmula analítica para uma melhor compreensão do relacionamento causa – efeito dos dados, seguido de sua validação. O processo dinâmico de identificação do sistema é realizado usando clusterização fuzzy através do algoritmo Gustafson e Kessel, seguido por um método de inferência fuzzy de Takagi e Sugeno. A técnica k-fold, é a ferramenta de validação cruzada, usada para evi-
tar o excesso de treinamento dos dados. A novidade da abordagem proposta abrange carac-
terísticas matemáticas e de engenharia que faz este estudo de caráter interdisciplinar. Como
contribuição matemática, existe uma interpretação gráfica tridimensional da geometria de
agrupamento de dados, obtida através de simulações com codificação própria. Relativo ao
conto de engenharia, a novidade é baseada no uso da abordagem fuzzy para a simulação
da célula de combustível de hidrogênio. Ambas as contribuições não têm precedentes na
literatura. Os resultados da identificação fuzzy mostram alta confiabilidade em termos de
validação cruzada, tornando a abordagem fuzzy uma ferramenta promissora para identifica-
çãode modelos de caixa preta. Combinar esta técnica com outras já estabelecidas irá fornecer
um instrumento poderoso para problemas industriais.

**Palavras-chave:** Célula de Hidrogênio, clusterização fuzzy, identificação de sistemas diná-
micos, método de Takagi Sugeno.

**REFERENCES**


injection in electrical net. PhD Thesis Faculty of Electrical Engineering, FEELT, Federal University
of Uberlândia, MG, Brazil, (2012).


(2002).


the IEEE Control and Decision Conference, (1979), 761–766.

418 DYNAMIC IDENTIFICATION SYSTEM OF A HYDROGEN FUEL CELL THROUGH FUZZY LOGIC


