

Clinical Assessment Using an Algorithm Based on Fuzzy C-Means Clustering

Alfonso A. Guijarro-Rodríguez^{1(✉)}, Lorenzo J. Cevallos-Torres¹, Miguel Botto-Tobar^{1,2}, Maikel Leyva-Vazquez¹, and Jessica Yepez Holguin¹

¹ Universidad de Guayaquil, Guayaquil, Ecuador
{alfonso.guijarror,lorenzo.cevallost,miguel.bottot,
maikel.leyvav,jessica.yepezh}@ug.edu.ec

² Eindhoven University of Technology, Eindhoven, The Netherlands
m.a.botto.tobar@tue.nl

Abstract. The Fuzzy c-means (FCM) algorithms define a grouping criterion from a function, which seeks to minimize iteratively the function up to until an optimal fuzzy partition is obtained. In the execution of this algorithm each element to the clusters is related to others that belong in the same n -dimensional space, which means that an element can belong to more than one clusters. This proposal aims to define a fuzzy clustering algorithm which allows the patient classifications based on the clinical assessment of the medical staff. In this work 30 cases were studied using the Glasgow Coma Scale to measure the level of awareness for each one which were prioritized by triage Manchester method. After applying the FCM algorithm the data is separated data into two clusters, thus, verified the fuzzy grouping in patients with a degree of membership that specifies the level of prioritization.

Keywords: Fuzzy logic · Clinical assessment · Fuzzy grouping · Triage · Glasgow

1 Introduction

Fuzzy c-means (FCM), are algorithms that operate with a non-supervised grouping criteria, their aim is to find out hidden patterns from a given set [1]. This algorithm for its processing defines an objective function which seeks to minimize the function in an iterative way in order to obtain an optimal fuzzy partition, therefore, the elements of the same class could be similar to others.

In the medical field fuzzy logic has being handled for many years ago, where enough evidence has been found in experimental programs aiming to improve the accuracy degree among them for decision-making, staff selection, and the clinical assessment process, where ambiguity and uncertainty are mostly handled [1].

This article presents a fuzzy grouping method for patient classification at the moment of coming to the doctor medical appointment [2]. GCS is a method to assesses the level of consciousness through nomenclatures, assessing the state of

alertness and the cognitive status of a person [2]. The Manchester Triage System model determines the level of urgency based on the information obtained from a patient, allowing its prioritization and classification according to the severity of his/her case [3].

From the sample considered in this research, an evaluation with the GCS method on each of one of 30 patients was performed in order to prioritize them and group them according to their level of severity, and regarding to the fuzzy grouping, the FCM algorithm was applied.

The paper is organized as follows: Sect. 2 discusses the fuzzy logic. Section 3 presents the research method. Section 4 describes the results obtained, and finally, Sect. 5 presents our conclusions and suggest areas for further investigation.

2 Fuzzy Logic

Fuzzy logic is considered as a technique to developing decision-making process using ambiguous variables [2], representing uncertainty mathematically, is worth to note that the ambiguous terms are not the logic, rather the objective to be analyzed [4]. Other definitions consider the fuzzy logic as a form of multi-valued logic that deals with approximate and non-precise reasoning [2].

This concept was proposed by Lofti Zadeh, who posed the fuzzy sets theory, and afterwards, he published the fuzzy logic theory and approximate reasoning in 1974. However, the diffuse logic unlike traditional logic has an unconcrete limit, it proposes a range of possibilities of veracity between (0-1), it may exist that an assertion is partially true or partially false. As a multivalued logic uses continuous values between zero (pretends completely facts) and one (totally certain) [3]. One example of information that uses the fuzzy logic is as follow: slightly low temperature, half-wet clothes, low speed, etc. making use of fuzzy terms for representing ambiguous data.

In classical sets elements belong or not to an established set, whereas in fuzzy sets the elements are related to a linguistic value, which is a word or phrase written in an artificial language. For instance, a linguistic variable could be the position of a balloon and its linguistic values could be: almost nothing, little, balanced, large, huge. The sentences can be formed between linguistic variables and linguistic terms. In this way the traditional logic can be observed as a particular event of the fuzzy logic (see Fig. 1).

2.1 Fuzzy Sets

Fuzzy sets are those elements that have a certain belonging degree, where an element can have a value of one or zero of the total elements. In [5] they define that a fuzzy set is a mathematical tool to address uncertainties that are free of difficulties. Unlike classical sets, where an element can just belong to a single class or group, whereas the elements of a fuzzy set, an individual may belong to more than one group in certain way. For the manipulation of fuzzy sets, two fundamental elements are applied.

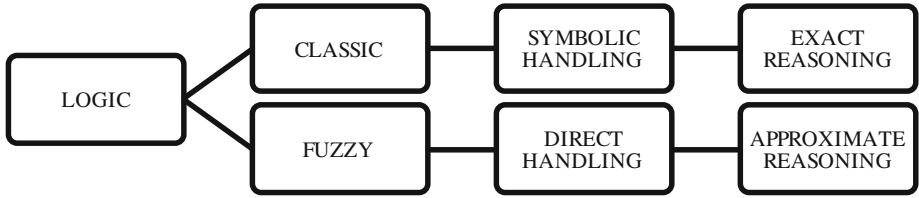


Fig. 1. Differences between classical and fuzzy logic.

2.2 Universe of Discourse

It is defined as all possible values that a variable can take, indicating all the elements that are classified by a belonging degree, it is represented by X . A clear example would be the classification of a group of people according to their height. In this case, the totality would be all the people that make up the group.

2.3 Membership Function

Also called belonging function $\mu(x)$, it is a curve that determines the membership degree of the set components, the graphs of the function are values $\mu(x) \in [0,1]$.

In medical environments, many people handle fuzzy terms that include some imprecision, those terms, are not handled by doctors and information systems directly. When a patient is interviewed, the obtained information is classical logic, these data are subtly fitted into two sets: “with pain”, “without pain”. Applying the theory of diffuse sets the degrees of pain belonging are determined, such as for example: “strong pain”, “moderate pain”, “half pain”, “low pain”.

2.4 Fuzzy Driver

It is a complete and diffuse decision-making system that uses fuzzy sets theory [4]. It includes three important steps: fuzzification, the inference, and defuzzification. The [6] fuzzification aims to transform the input data into output linguistic variables. The rule or inference engine is carried out in the grouping of input data according to output data. The final step is the defuzzification, its procedure is to transform the diffuse data into an arithmetical value (See Fig. 2).

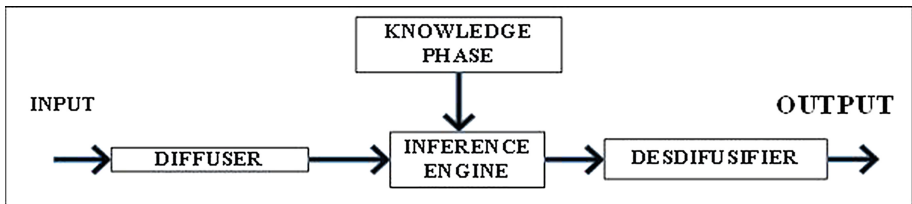


Fig. 2. Fuzzification and Defuzzification.

2.5 Fuzzy Grouping

The clustering of data is based on the assumption that a population of objects can be subdivided into smaller, homogeneous subgroups internally for one or more functions [7]. Clustering is an elements associating method sharing similar properties. Many diverse techniques [8] have been developed in order to discover similar groups in large data sets, of which hierarchical and partitional techniques are being widely practiced. The group is much more natural than conventional set theory, since the objects that are at the borders of these groups, may not necessarily be forced to belong to one of them [9].

3 Research Method

In order to carry out the clinical assessment process to diffuse grouping, we used the FCM algorithm formulated by Dunn and Bezdek. The case study data were assessed, applying three aspects of level of SGA awareness to the 30 patients with brain injury traumatism. According to obtained results from the assessment scale, we identified patients with a high degree of consciousness and patients who are completely unconscious, which ones were classified using the Manchester triage system method. For the data processing, Orange Canvas tool [10] was used to apply the clustering algorithms and to observe the separated results in two clusters

3.1 Glasgow Coma Scale (GCS)

GCS [11] is considered as an instrument with high sensitivity for the assessment of patients with brain damage, the level of consciousness of a person is measured through three parameters: the ocular, verbal and motor response. The ocular response has a maximum value of 4 and a minimum score of 1, the maximum score is obtained when the patient responds by spontaneously opening the eyes. To assess the motor response, it gives the following scores: 6 if the patient captures the orders, 5 if the patient locates the pain, 4 if the patient responds in a retreat to the response to the pain, 3 if he/she makes an abnormal flexion, 2 if he/she makes an abnormal extension and 1 if there are responses absences. Regarding to verbal assessment the values are as follow: 5 if the answer is oriented, 4 if the patient responds confused, 3 if responds inconsistently, 2 if speech incomprehensible, if no answer a score of 1. In the out-of-hospital setting, GCS is an important tool for decision-making and classification, where the initial score acts as an important prognostic indicator following traumatic brain injury [12]. The Nomenclatures of the GCS are shown in Table 1.

The maximum value obtained on the scale is 15, which it indicates that the patient is conscious, and on the other hand, the minimum value is 3, the patient is unconscious.

Table 1. Parameters of GCS with their nomenclatures.

LEVEL	ANSWER	VALUE
Opening Eyes	Spontaneous, flashing	4
	For verbal stimulation, speech	3
	For pain only	2
	No response	1
Verbal	Oriented	5
	Answer to confuse questions	4
	Inappropriate words	3
	Incomprehensible speech	2
	No response	1
Motor response	Obey orders for movement	6
	Pain stimulus movement	5
	Pain response is withdrawn	4
	Bending in response to pain	3
	Extension in response to pain	2
	No response	1

3.2 Process of the Manchester Triage System (MTS)

Triage has emerged as a method to optimize care and minimize damage caused by overcrowding by identifying patients who need immediate care [13]. The MTS is a method of prioritization and classification that was raised in the English city Manchester by professionals in emergency services. Nowadays, this method of prioritizing patients in the emergency room is applied in public hospitals, which is implemented at the IESS: Teodoro Maldonado Carbo hospital and at the University Hospital of the Guayaquil city. In [14], the classification levels of the MTS are presented. The highest level means that a patient is probably suffering from a life-threatening illness, so immediate attention is provided. Such patients are assigned the red color category. Urgent conditions are assigned the orange color category, they are able to wait a time limit of 10 min. Less urgent patients are assigned a yellow color category. Non-urgent patients, green category with time limit of 2 hours. Other patients are given a blue triage level, which means they can wait for 4 hours. The MTS process will then be displayed in Table 2.

In general terms, this method attempts to provide diagnosis, elimination or clinical priority.

3.3 FCM Clustering Method

It is one of the most used algorithms in diffuse grouping. It assigns to each data a membership value within each cluster and therefore, specific data may belong to more than one cluster [15]. A fuzzy partition determines the intervention of each

Table 2. Classification of the Manchester Triage System.

Priority	Attention	Time	Color
1	Immediate	0 min	Red
2	Very urgent	10 min	Orange
3	Urgent	60 min	Yellow
4	Standard	120 min	Green
5	Non-urgent	240 min	Blue

sample in the clusters using membership functions that take values of [0,1]. FCM performs a soft distribution known as soft partition, such partition elements are hierarchically integrated to all clusters. A mathematically smooth partition is defined:

$$X = \text{data set}$$

$$xi = \text{element corresponding to } X$$

$P = \{C_1, C_2, \dots, C_c\}$ soft partition X , whether only if the following is true:

$$\forall xi \in X \quad \forall C_j \in P \quad 0 \leq U_{c_j}(xi) \leq 1 \tag{1}$$

For all Xi which belongs to X , for all cluster C_j which belongs to P , the belonging coefficient $U_{c_j}(xi)$ must be between [0-1].

$$\forall xi \in X \quad \exists C_j \in P \quad \text{such that} \quad U_{c_j}(Xi) > 0 \tag{2}$$

For all xi which belongs to X , there is a cluster which belongs to P , such that the belonging coefficient $U_{c_j}(xi)$ is greater than zero. FCM minimizes function:

$$J_m = \sum_{i=1}^N \sum_{j=1}^C U_{ij}^m \|x_i - c_j\|, 1 \leq m \leq \infty \tag{3}$$

known as the objective function where:

$m =$ is some arithmetic value greater than 1.

$U_{ij} =$ is the belonging degree x_i in the group j .

$x_i = i$ of the measured data.

$c_j =$ cluster dimension center.

$\| * \| =$ similarity between measured data and center.

The calculation is done updating the belonging coefficients U_{ij} , and the centroids C_j . A diffuse partition $\{C_1, C_2, \dots, C_k\}$ can be a local minimum of the objective function J^m only if the following conditions are fulfilled [15].

3.4 FCM Grouping Algorithm

Step 1.

- Initialize the centroids.
- Choose the number of cluster C .
- Choose the exponent m .
- Initialize the array of partition U randomly.

Step 2. Calculate the belonging functions of n clusters, through the use of:

$$\cup_{cj}(x) = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - v_i\|^2}{\|x_i - v_j\|^2} \right)^{\frac{1}{m-1}}} \tag{4}$$

Step 3. Sum each of belonging in both clusters which complies with:

$$\sum_j \cup_{ci}(x_i) = 1 \quad \forall x_i \in X \tag{5}$$

In case of error, the belonging grade should be calculated once more.

Step 4. Update the n centroids $C = C_j$.

$$v_i = \frac{\sum_{k=1}^N \cup_{ci}(x)^m .x}{\sum_{k=1}^N \cup_{ci}(x)^m} \tag{6}$$

Note: Once the number of clusters and the value of m have been defined, the centroids and the core initialized the then FCM algorithm performs 2 steps: first, the belonging functions are calculated using the equation in step 2, and second, the prototypes are updated by means the equation in step 4.

3.5 Case Study

There is a table evaluating the Glasgow valuation process and triage prioritization of 30 patients with head trauma. These data were stored in an Excel table, then it was taken to the orange application [10], where the evaluation was performed (see Table 4).

Once the results were obtained, a close value to 1 indicated that the patient is consent and did not need immediate attention, and on the other hand, a close value to 0 showed that the patient is unconscious and needed immediate attention. Depending on the case we should expect a time limit according to the Manchester Triage system. It is desired to separate the data set into two groups (clusters) to visualize those patients with different characteristics. For this case, centroids or prototypes are initially defined.

Data: The cluster quantity is 2, the parameter m is 2, centroids: $v1 = (0.3, 0.6)$ and $v2 = (0.9, 0.6)$.

Note: We designed and worked with a table according to the Manchester triage system in fuzzy level (Table 3) in order to classify the patients according to the result obtained in the Glasgow scale assessment.

As indicated above we grouped those patients who have a belonging degree close to 1 in a cluster which meant that they did not need immediate attention, and on the other hand, the patients with a lower degree of belonging in another cluster, needed immediate attention.

Table 3. Manchester Triage System with fuzzy levels.

Fuzzy prioritation	Color	
[0–0.20]	Red	Immediate attention
[0.21–0.47]	Orange	Immediate attention
[0.48–0.65]	Yellow	Non-immediate attention
[0.66–0.80]	Green	Non-immediate attention
[0.81–100]	Blue	Non-immediate attention

4 Results

Table 4 shows the results of the Glasgow evaluation of 30 patients with craniofemoral trauma, we applied 3 parameters in order to measure the level of consciousness, identifying those who need immediate attention or not, resulting in random values between [0-1].

Table 5 shows the final results of the evaluation process, the sum is the result of the clinical assessment for each one of the 30 patients, for that reason was generated an approximation value, indicating the priority of the Clinical assessment by means of an algorithm based on clustering Fuzzy c-means.

For this case study, the first iteration of the algorithm, the belonging coefficient in the first element is calculated. An important elements to consider in the fuzzy clustering algorithm are the centroids. This value allows to represent all the possible characteristics belonging to a diffuse set. It is worth noting that the value to be considered as centroid oscillates between 0 and 1. Therefore, for this research we considered referential values $v1 = (0.3, 0.6)$ and $v2 = (0.9, 0.6)$, and the following formula was applied:

$$U_{c1}(X1) = \frac{1}{\sum_{j=1}^2 \left(\frac{\|x_i - v_j\|^2}{\|x_i - v_j\|^2} \right)^2} \tag{7}$$

The sum of the centroids $v1 = (0.3, 0.6)$ is calculated. The first point is subtracted with centroid $v1$. $0.67 - 0.3 = 0.37$, and $0.53 - 0.6 = 0.07$. With that same point we subtracted the second centroid $v2 = (0.9, 0.6)$ to obtain the values of 0.2 and 0.07.

$$\|x1 - v1\|^2 = 0.37^2 + 0.07^2 = 0.1369 + 0.0049 = 0.1418 \tag{8}$$

$$\|x1 - v2\|^2 = 0.23^2 + 0.07^2 = 0.0529 + 0.0049 = 0.0578 \tag{9}$$

$$U_{c1}(x1) = \frac{1}{\frac{0.1418}{0.1418} + \frac{0.1418}{0.0578}} \tag{10}$$

$$U_{c1}(x1) = \frac{1}{1 + 2.4533} \tag{11}$$

$$U_{c1}(x1) = 0.2896 \tag{12}$$

Table 4. Glasgow valuation.

Patient	Eye response	Verbal response	Motor response
1	0,27	0,07	0,33
2	0,07	0,33	0,13
3	0,07	0,13	0,13
4	0,13	0,33	0,07
5	0,20	0,20	0,33
6	0,07	0,20	0,07
7	0,07	0,20	0,20
8	0,07	0,13	0,33
9	0,20	0,13	0,07
10	0,13	0,20	0,27
11	0,27	0,07	0,13
12	0,07	0,07	0,13
13	0,07	0,20	0,13
14	0,20	0,13	0,07
15	0,20	0,13	0,40
16	0,07	0,20	0,27
17	0,20	0,27	0,20
18	0,27	0,13	0,40
19	0,27	0,20	0,27
20	0,27	0,13	0,40
21	0,07	0,33	0,40
22	0,13	0,33	0,40
23	0,07	0,33	0,07
24	0,07	0,13	0,33
25	0,27	0,27	0,33
26	0,27	0,07	0,07
27	0,27	0,27	0,40
28	0,20	0,33	0,40
29	0,27	0,07	0,07
30	0,13	0,07	0,27

Similarly, the values of the other membership functions are calculated, beside the complete results are shown in Table 6. For the calculation of the belonging values of both clusters, we used a source code of the FCM algorithm implemented in C# [16].

Table 5. Final results.

Patient	Health status	Priority
	Glasgow	Medical attention
1	0,67	0,53
2	0,53	0,60
3	0,33	0,53
4	0,53	0,53
5	0,73	0,33
6	0,33	0,67
7	0,47	0,33
8	0,53	0,47
9	0,40	0,67
10	0,60	0,19
11	0,47	0,47
12	0,27	0,80
13	0,40	0,67
14	0,40	0,87
15	0,73	0,53
16	0,53	0,40
17	0,67	0,93
18	0,80	0,60
19	0,73	0,53
20	0,53	0,67
21	0,80	0,28
22	0,87	0,12
23	0,47	0,12
24	0,53	0,87
25	0,87	0,52
26	0,40	0,98
27	0,93	0,05
28	0,93	0,43
29	0,40	0,09
30	0,47	0,25

Table 6 shows the belonging values obtained in 2 cluster by compiling the program, and it provided as a result the approximation values of the clinical assessment in 30 patients.

The grade of belonging 1 indicates the maximum of belonging, whereas 0 indicates that the data does not belong to the cluster, thus, the data with greater

Table 6. Values of belonging for each data/cluster.

Priority - Medical attention	Cluster 1	Cluster 2
0,53	0,2896	0,7104
0,60	0,7213	0,2787
0,53	0,9827	0,0173
0,53	0,7104	0,2896
0,33	0,2831	0,7169
0,67	0,9827	0,0173
0,33	0,8488	0,1512
0,47	0,6878	0,3122
0,67	0,9448	0,0552
0,19	0,5000	0,5000
0,47	0,8150	0,1850
0,80	0,9144	0,0856
0,67	0,9448	0,0552
0,87	0,7957	0,2043
0,53	0,1512	0,8488
0,40	0,6557	0,3443
0,93	0,3970	0,6030
0,60	0,0385	0,9615
0,53	0,1512	0,8488
0,67	0,7104	0,2896
0,28	0,2418	0,7582
0,12	0,2941	0,7059
0,12	0,3844	0,6156
0,87	0,6251	0,3749
0,52	0,0216	0,9784
0,98	0,7187	0,2813
0,05	0,3026	0,6974
0,43	0,0654	0,9346
0,09	0,6538	0,3402
0,25	0,6700	0,3300

belonging in cluster 1 are 2, 3, 4, 6, 7, 8, 9, 11, 12, 13, 14, 16, 20, 23, 26, 29, 30 while the remaining data has a greater belonging in cluster 2. These values were represented using Open Source Orange Canvas tool [10]. In Fig. 3, the data were loaded into the program by means of the widget field. Widget Data Table allowed to visualize the simulated data, and the widget Scatter Plot showed the results of the dispersion of the data. The first cluster brings together patients who are aware and do not need medical attention thus they would wait for a time limit, while the second cluster brings together all patients who need immediate care.

Based on the sample applied with the Manchester triage system, 17 patients obtained non-immediate priority, which had to wait the time limit according to

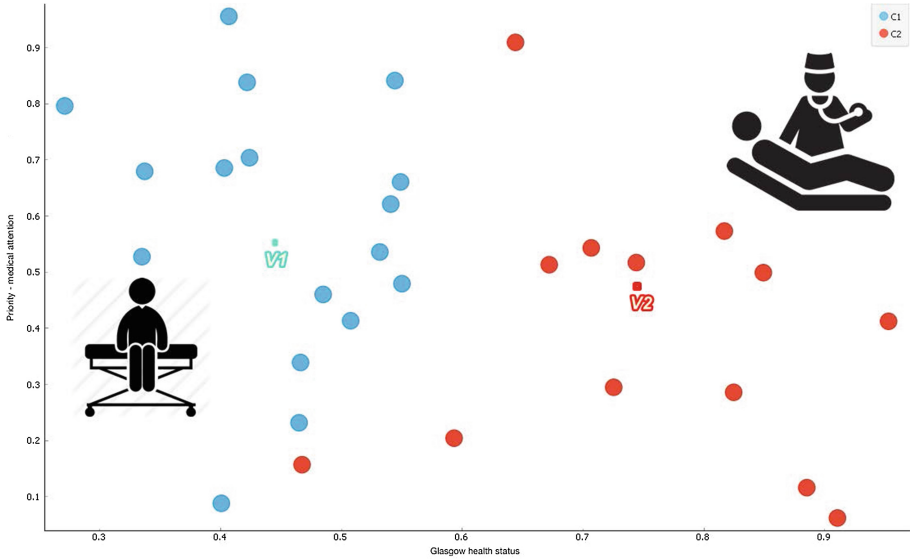


Fig. 3. Dispersion of data and in Orange Canvas [10] separated in 2 cluster.

their level of urgency, meanwhile the remaining patients were considered as cases of immediate attention.

According to the experts, the case study allowed us to glimpse the fuzzy grouping, as well as a proposed classification of the patients in the emergency room of the University Hospital, which indicates that the prioritization and the classification of the patient has improved significantly through the use of Fuzzy c-means algorithm.

5 Conclusions

The clinical assessment work was performed with five groups of 30 patients in approximately 30 days in the emergency room of the University Hospital. Afterwards, a case study was generated with 30 patients from a total of 150, with the intention of generalizing the method and making the respective suggestions of the use Fuzzy c-means algorithm, the results obtained allows to segment patients groups form high to low priority.

After the application of the Fuzzy c-means algorithm to the clinical evaluation process, an effective contribution was obtained by allowing the medical staff and their work team to obtain a better data visualization, therefore, they can make better decisions regarding prioritization of the patient.

References

1. Gudino-Penalosa, F., González-Mendoza, M., Mora-Vargas, J.: Uso de técnicas de agrupamiento en la clasificación de estilos de aprendizaje. *Res. Comput. Sci.* **95**, 135–146 (2015)
2. Hazelzet, J.A.: Can fuzzy logic make things more clear? *Crit. Care* **13**(1), 116 (2009)
3. Horgby, P.J.: Risk classification by fuzzy inference. *GENEVA Pap. Risk Insur. Theory* **23**(1), 63–82 (1998)
4. Lim, C.H., Vats, E., Chan, C.S.: Fuzzy human motion analysis: a review. *Pattern Recogn.* **48**(5), 1773–1796 (2015)
5. Çelik, Y., Yamak, S.: Fuzzy soft set theory applied to medical diagnosis using fuzzy arithmetic operations. *J. Inequalities Appl.* **2013**(1), 82 (2013)
6. Velarde, E.S.S., Sotelo-deÁvila, A.A., Rico-Asención, I.O., Alemán, N.R., González, R.S., Ramírez-Sotelo, M.G., Cabrera-Llanos, A.I.: Fuzzy-state machine for triage priority classifier in emergency room. In: Jaffray, D. (ed.) *World Congress on Medical Physics and Biomedical Engineering*, June 7–12, 2015, Toronto, Canada, pp. 1488–1491. Springer, Cham (2015)
7. Fu, L., Medico, E.: Flame, a novel fuzzy clustering method for the analysis of dna microarray data. *BMC Bioinf.* **8**(1), 3 (2007)
8. Sandra, A.K., Sarkar, A.K.: Application of fuzzy logic and clustering techniques for pavement maintenance. *Transp. Infrastruct. Geotechnol.* **2**(3), 103–119 (2015)
9. Plata, E.C., Plata, C.C.: Análisis de los algoritmos de agrupamiento borroso para detectar asimetría de información. *Economía y Administración* **45**(70), 7–22 (2008)
10. Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., et al.: Orange: data mining toolbox in python. *J. Mach. Learn. Res.* **14**(1), 2349–2353 (2013)
11. Muñana-Rodríguez, J.E., Ramírez-Elías, A.: Escala de coma de glasgow: origen, análisis y uso apropiado. *Enfermería universitaria* **11**(1), 24–35 (2014)
12. Heim, C., Schoettker, P., Gilliard, N., Spahn, D.R.: Knowledge of glasgow coma scale by air-rescue physicians. *Scand. J. Trauma Resuscitation Emerg. Med.* **17**(1), 39 (2009)
13. Azeredo, T.R.M., Guedes, H.M., de Almeida, R.A.R., Chianca, T.C.M., Martins, J.C.A.: Efficacy of the manchester triage system: a systematic review. *Int. Emerg. Nurs.* **23**(2), 47–52 (2015)
14. Forsgren, S., Forsman, B., Carlström, E.D.: Working with manchester triage-job satisfaction in nursing. *Int. Emerg. Nurs.* **17**(4), 226–232 (2009)
15. Porras, J.C.C., Laverde, R.M., Diaz, J.R.: Tecnicas de logica difusa aplicadas a la minería de datos. *Sci. Tech.* **3**(40), 1–6 (2008)
16. Jurado, J., Penagos, D.: Fuzzy C Means Presentation (2013)