International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

PANTSA Influence in grouping Mixed and Incomplete Data

Yusbel Chávez-Castilla

Abstract*: Obtaining high quality groups and processing mixed and incomplete data (DMI) are still problems in the data clustering. Recently a method was proposed that improves the results obtained by clustering algorithms, the PAntSA; but this was only designed and tested for numerical data. For this reason, this paper analyzes the influence of applying the PAntSA in the performance of DMI restricted clustering algorithms. For this, the results of different algorithms are compared before and after applying the PAntSA. The comparisons made provide experimental evidence that the PAntSA algorithm improves the quality of the groups obtained by traditional DMI clustering methods.

***Keywords* : About four key words or phrases in alphabetical order, separated by commas.**

I. INTRODUCTION

The process of grouping a set of physical or abstract objects into classes with similar objects is called grouping. In this process, a given collection of unlabeled data is taken and a group of groups is created in such a way that the objects belonging to a group are homogeneous with each other [1], also seeking that the heterogeneity between the different groups be highest possible . This technique has acquired great relevance in recent times due to its practical application in the successful solution of dissimilar real-life problems such as: Speech recognition, image segmentation and computer vision, information retrieval and text mining, in computational biology for DNA analysis and many other applications [2-7].

Most of the grouping algorithms have been designed to work only with numerical data or with categorical data, while in a large number of occasions, it is necessary to work with mixed data, that is, with attributes of different types such as: numerical, binary, discreet and categorical. Also on many occasions it is not possible to know the value of a certain attribute, so it is also necessary to develop algorithms to group incomplete data [8-13].

Mixed and incomplete data (DMI) grouping or has been traditionally addressed by following classical paradigms such as hierarchical and particional, but have recently appeared bio-inspired proposals have also had a good performance [14-20]. On the other hand, the grouping obtained by a given algorithm can be improved by another algorithm using an internal validation index. Using this idea in 2010, the bio-inspired PAntSA algorithm [21] (based on an ant tree [22, 23]) was published, which takes the results obtained by a previous grouping algorithm and tries to refine them using the Silhouette index [24] and the definition of an attraction between groups. PAntSA improves the quality of the results obtained by clustering algorithms in numerical data, particularly in the classification of documents; however, as far as we know, there is no study of the influence of PAntSA in the DMI cluster [25-31]. For this reason, in this paper we propose to analyze the influence of PAntSA in improving the results of the DMI clustering algorithms.

From here the rest of the work is organized as follows: In section 2 we show the use of internal validation rates in the DMI cluster. In section 3 we describe the PAntSA algorithm used to improve the groups obtained by other algorithms. In section 4 we present an experimental analysis on the influence of PAntSA in the grouping of DMI and the results obtained. Finally, section 4 gives the conclusions.

II. GROUPING OF DMI AND INTERNAL VALIDATION INDICES

A. Submission of the paper

The grouping of data, as explained above, consists of the unsupervised assignment of objects to unknown groups. This task is more difficult than the supervised classification because the labeled objects are not taken in advance. A consequence of this is that the results of the grouping cannot usually be evaluated with external validation indices such as Entropy and Grouping Error (see [24]) because the correct classification by a human expert is not available . However, the quality of the groups obtained can be evaluated with respect to structural properties of the data through internal validation rates [32-38]. Some of the most commonly used internal indices are indices Dunn and Davies-Bouldin, the Silhouette (Silhouette*)* and Hubert index. For a more detailed description of these indices refer to [39].

The majority of people working in data clustering problems are familiar with the use of indices of internal validation. However, in some recent works other uses have been proposed for this type of measures, specifically in the context of the DMI grouping is its use as an objective function to optimize in different algorithms.

Revised Manuscript Received on December 30, 2019.
***** Correspondence Author
Yusbel Chávez-Castilla*, Computer Science Department, University of Ciego de Ávila, Cuba. Email: fci.unica@gmail.com

This idea has recently been applied in two bio-inspired algorithms: BECA [40] and DELUXE [41].

BECA (*BEe based Clustering Algorithm*) is a method based on ABC metaheuristics , inspired by the natural behavior of bees and proposed by Karaboga in [42]. This algorithm generates *n* initial clusters randomly that constitute the food sources [43-50]. It then generates new sources through the mutation strategy defined in [51] using a heterogeneous dissimilarity to deal with DMI. For the evaluation of food sources, the Dunn index is used as an objective function [24]. Finally, after a number of iterations defined a priori, the food source (grouping) that optimized the objective function is returned.

In the case of DELUXE [41], this is an algorithm developed to group DMI based on FMOA , an optimization technique was recently developed by Xin-She Yang [52] that models the behavior of the fireflies and the flashes of light they emit [46, 53-59]. Like BECA [40], its operation is based on a dissimilarity that allows the treatment of DMI , however unlike this in DELUXE initially a number η of fireflies is formed by selecting *k* random centers and placing the objects around them. Then an iterative process is started comparing in pairs the attractiveness of the different fireflies (clusters) using the Silhouette index [24] as an objective function, in this way, fireflies with less attractive are disturbed towards those of greater attractiveness. The disturbance consists in replacing one of the centers of the groups with another object randomly. To avoid falling into local optimum, a parameter was introduced to the algorithm: the life expectancy of the firefly, allowing replace the fireflies that could not be disturbed for a predetermined number of times.

Silhouette [24] is an internal validation index that combines two key elements to the quality of a given grouping: compactness and separability. This index calculates the average, for all groups, of the width of the silhouette of its points.

III. PANTSA IN THE IMPROVEMENT OF CLUSTERS

The PAntSA algorithm [21] is based on the AntTree algorithm, proposed in [22, 23]. For the best understanding of PAntSA, we will explain in detail first the operation of the AntTree algorithm. This algorithm is a pioneer in the application of the modeling of nest construction by ants, to problems of Artificial Intelligence. The AntTree is based on modeling the ability of ants to build living structures with their bodies [23], to discover, in a distributed and unsupervised manner, a tree structure that organizes a set of data. This hierarchical structure can be interpreted in several ways: as a partition of the data or as a hierarchical structuring of them [23, 60-65].

The fundamental principle of the AntTree is the following: each ant represents a node in the tree that will be constructed (that is, the objects that will be grouped) and there is a function of similarity between two $Sim(i, j)$ objects . On the basis of a fictitious root node to_0, which represents the support on which it is to build the tree, each ant to_i will go gradually setting to the initial node, and successively to ants and fixed, until all The ants are in the structure. All movements and fixations in the structure will depend on the value of $Sim(i, j)$, and on a neighborhood where the ants move. Although Ingaramo et al. they tested the efficiency of the PAntSA [21] in the grouping of documents, and showed that said algorithm is capable of improving clustering algorithms for numerical data, to our knowledge there is no study of whether it is capable of improving algorithm results clustering for mixed data, which is target is and work.

IV. RESULTS AND DISCUSSION

To evaluate the performance of the mixed data PAntSA in an experimental comparison was conducted using algorithms grouping of DMI reported in the literature, by applying the results or PAntSA btenidos for each. For the experiments, 10 mixed and incomplete databases from the repository of the University of California, Irvine (ICU) [66] were used (see table 1).

Table- I: Description of the datasets

Dataset	Cat. Att.	Num. Att.	Classes	Missing	Objects
Autos	10	16	7	Yes	205
Colic	15	7	2	Yes	368
dermatology	1	33	6	Yes	366
heart-c	7	6	5	Yes	303
hepatitis	13	6	2	Yes	155
Labor	6	8	2	Yes	57
lymph	15	3	4	No	148
sponge	44	0	3	Yes	76
tae	2	3	3	No	151
zoo	16	1	7	No	101

The selection of this data set lies in that it is tagged data bases. This enabled to have grouping model (classes) against which to measure the quality of the groups obtained by the algorithms by two index s validation external. The first index used was Entropy, which measures the degree of disorder of the model grouping (*AM)* in the groups obtained by the following equation.

$$E = \sum_{k \in AE} \frac{|k|}{N} \left[\frac{1}{\log(|AM|)} \sum_{m \in AM} \frac{n_k^m}{|k|} \log \frac{n_k^m}{|k|} \right] \quad (1)$$

Where: *AE* represents the grouping to evaluate, $|k|$ is the total objects of group *k*, $|AM|$ it is the total of groups of the model grouping and n_k^m is the total of objects in the group *k* that belongs to the group of *AM*.

In the experimental comparison eight algorithms were used that allow the grouping of DMI and the results obtained by these before and after applying the PAntSA [21] were evaluated . Thus, the k Prototypes [67] and AD2011 [68], the hierarchical HIMIC [69] and the CEBMDC [70] method which is an algorithm that uses a combination of other methods were selected. In addition, AGKA [51], BECA [40] metaheuristics were used.

An important aspect in the design of the experiments are the parameters with which the algorithms will be executed [71-74]. As all algorithms require knowing the number of groups to be formed, the value assigned to this parameter will match, for each database, the number of classes. With this, the classes are taken as a model grouping against which to evaluate the clusters resulting from applying the algorithms. The rest of the parameters were chosen based on the existence of studies that recom sen certain values for better performance. In addition to the common parameters of the different algorithms they were given the same value. This allowed to achieve a certain homogeneity and reduce a possible imbalance in the performance of an algorithm against another by the use of different values for the same parameter. In the case of dissimilarity, the HEOM function [75] was used for all algorithms . The reason for its use was its good results in the treatment of DMI [76-78].

The experiments were conducted as follows: For each database, the different algorithms were applied and Entropy was calculated. Then PAntSA [21] was applied to the result of each algorithm obtaining a new grouping and the Entropy was also calculated to these new results. In each case, the number of classes in each of the databases was established as the number of groups.

In Figures 1 - 6 the results of each are shown algorithm, before and after applying the PAntSA. As can be seen, in many cases Entropy decreases after the application of PAntSA

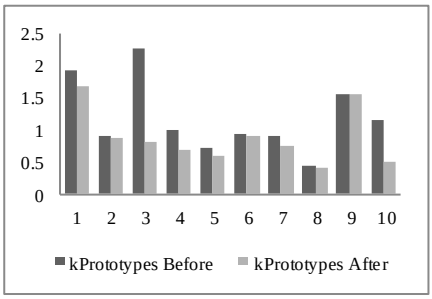

Fig. 1.Entropy results before and after the PAntSA was applied for the kPrototypes algorithm

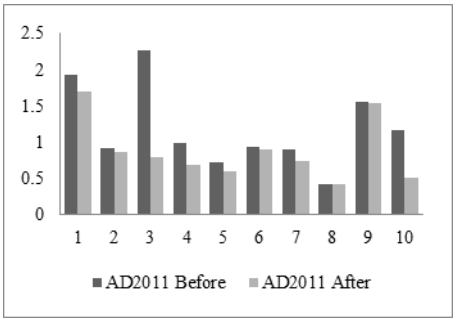

Fig. 2.Entropy results before and after the PAntSA was applied for the AD2011 algorithm

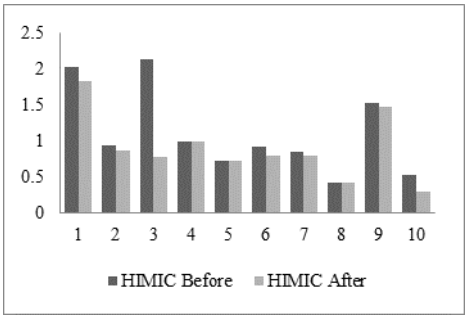

Fig. 3.Entropy results before and after the PAntSA was applied for the HIMIC algorithm

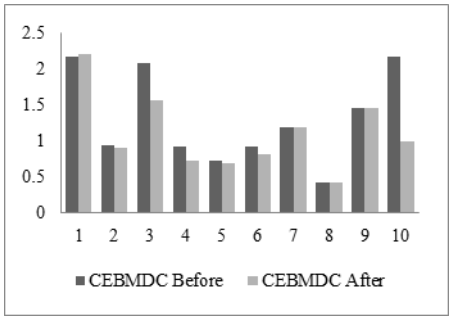

Fig. 4.Entropy results before and after the PAntSA was applied for the CEBMDC algorithm

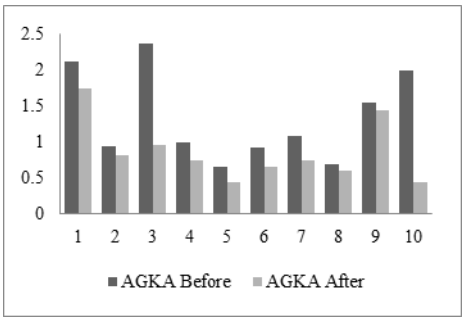

Fig. 5.Entropy results before and after the PAntSA was applied for the AGKA algorithm

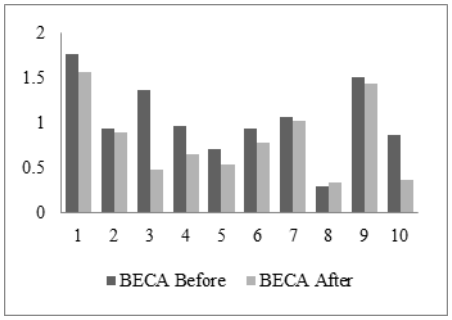

Fig. 6.Entropy results before and after the PAntSA was applied for the BECA algorithm

In most cases, a lower Entropy is obtained after applying the PAntSA. However, to establish whether these differences are statistically significant or not, the Wilcoxon test was applied for two related samples. Table 2 shows the bilateral asymptotic significance obtained by the test.

The Wilcoxon test allows to clarify whether PAntSA effectively improves or not the results obtained by the grouping methods. For this, a 95% confidence level was established. Thus, for probability values greater than 0.05, it is considered that there are no differences in the quality of the groups after applying the PAntSA . For this, the symbol esco is chosen to facilitate understanding. However, for values less than 0.05, it is necessary to determine whether this method improves or worsens the results obtained by the grouping method , using the symbols ☺ and ☹ , respectively.

The symbol ☺ means that the PAntSA significantly improved the results of the corresponding algorithm, while the symbol 😐 means that there were no differences between the results of the algorithm before and after the PAntSA was applied. No evidence was found that PAntSA would worsen the results obtained. As can be seen, PAntSA is able to improve the results of the clusters obtained by all algorithms except for the AD2011 method, which showed no significant differences.

Table- II: Bilateral asymptotic significance of the Wilcoxon test for Entropy of the clusters obtained by the algorithms before and after applying the PAntSA

Algorithms before and after PAntSA	*kPrototypes*	*AD2011*	*HIMIC*	*CEBMDC*	*AGKA*	*BECA*
Sig. Asint.	0.005	1.000	0.008	0.037	0.005	0.009
Decision	☺	😐	☺	☺	☺	☺

In general, as could be seen in the experiments analyzed, PAntSA showed a good level of effectiveness in improving the results of DMI clustering algorithms. Only no differences were obtained using the AD2011 algorithm.

V. CONCLUSION

Obtaining high quality clusters in mixed and incomplete data is especially important. The study carried out allows us to affirm that the results obtained by grouping methods of diverse nature (partitions, hierarchical, bio-inspired and others) can be refined by applying post-processing strategies. The PAntSA algorithm, in all cases improved or maintained the quality of the groups analyzed, and in no case its application implied a detriment to it. On the other hand, the use of internal validation indices, in this case of the Silhouette, opens new lines of research regarding the quality of the clusters, since it is considered that in addition to the compactness and separability properties that this index, other properties can be used to further refine mixed and incomplete data clusters.

REFERENCES

1. Join, A.K., y R. C. Dubes, Algorithms for Clustering Data. 1988, New Jersey, USA: Prentice Hall.
2. Medina-Pérez, M.A., et al. Selecting objects for ALVOT. in Iberoamerican Congress on Pattern Recognition. 2006. Springer.
3. Roman-Godinez, I., I. Lopez-Yanez, and C. Yanez-Marquez. A new classifier based on associative memories. in 2006 15th International Conference on Computing. 2006. IEEE.
4. Villuendas-Rey, Y., et al. Simultaneous features and objects selection for Mixed and Incomplete data. in Iberoamerican Congress on Pattern Recognition. 2006. Springer.
5. Acevedo-Mosqueda, M.E., C. Yáñez-Márquez, and I. López-Yáñez, Alpha–Beta bidirectional associative memories: theory and applications. Neural Processing Letters, 2007. **26**(1): p. 1-40.
6. Guzman, E., et al. Image recognition processor based on morphological associative memories. in Electronics, Robotics and Automotive Mechanics Conference (CERMA 2007). 2007. IEEE.
7. Yáñez-Márquez, C., et al. Using alpha-beta associative memories to learn and recall RGB images. in International Symposium on Neural Networks. 2007. Springer.
8. Ruiz-Shulcloper, J., Pattern Recognition with Mixed and Incomplete Data. Pattern Recognition and Image Analysis, 2008. **18**(4): p. 563-576.
9. Guzmán, E., et al., Morphological transform for image compression. EURASIP Journal on advances in signal processing, 2008. **2008**(1): p. 426580.
10. Moreno-Moreno, P. and C. Yáñez-Márquez. The new informatics technologies in education debate. in World Summit on Knowledge Society. 2008. Springer.
11. Rudas, I.J., et al. Generators of fuzzy operations for hardware implementation of fuzzy systems. in Mexican International Conference on Artificial Intelligence. 2008. Springer.
12. Villuendas-Rey, Y., M. García-Borroto, and J. Ruiz-Shulcloper. Selecting features and objects for mixed and incomplete data. in Iberoamerican Congress on Pattern Recognition. 2008. Springer.
13. Yáñez-Márquez, C., I. López-Yáñez, and G.d.l.L.S. Morales. Analysis and prediction of air quality data with the gamma classifier. in Iberoamerican Congress on Pattern Recognition. 2008. Springer.
14. Acevedo, M.E., C. Yáñez-Márquez, and M.A. Acevedo, Associative models for storing and retrieving concept lattices. Mathematical Problems in Engineering, 2010. **2010**.
15. López-Yáñez, I., et al., Pollutants time-series prediction using the Gamma classifier. International Journal of Computational Intelligence Systems, 2011. **4**(4): p. 680-711.
16. Lopez, S.J., O.C. Nieto, and J.I.C. Oria, Non-parametric modeling of uncertain hyperbolic partial differential equations using pseudo-high order sliding mode observers. International Journal of Innovative Computing, Information and Control, 2012. **8**(3): p. 1501-1521.
17. Villuendas-Rey, Y., Y. Caballero-Mota, and M.M. García-Lorenzo. Using rough sets and maximum similarity graphs for nearest prototype classification. in Iberoamerican Congress on Pattern Recognition. 2012. Springer.
18. Villuendas-Rey, Y., Y. Caballero-Mota, and M.M. García-Lorenzo. Intelligent feature and instance selection to improve nearest neighbor classifiers. in Mexican International Conference on Artificial Intelligence. 2012. Springer.
19. Villuendas-Rey, Y., Y. Caballero-Mota, and M.M. García-Lorenzo. Prototype selection with compact sets and extended rough sets. in Ibero-American Conference on Artificial Intelligence. 2012. Springer.
20. Villuendas-Rey, Y., et al. Nearest prototype classification of special school families based on hierarchical compact sets clustering. in Ibero-American Conference on Artificial Intelligence. 2012. Springer.
21. Ingaramo, D.A., Errecalde, M. L., et al., A general bio-inspired method to improve the short-text clustering task. 2011.
22. Azzag, H., Venturini, G., A clustering model using artificial ants. 2004, Université François-Rabelais,: Tours - France.
23. Azzag, H., Monmarche, N., Slimane, M., Venturini, G., Guinot, C. AntTree: A new model for clustering with artificial ants. in CEC2003. 2003. Canberra, Australia: IEEE Press.
24. Brun, M., et. al., Model-based evaluation of clustering validation measures. Pattern Recognition, 2007: p. 807-824.
25. García-Borroto, M., et al. Using maximum similarity graphs to edit nearest neighbor classifiers. in Iberoamerican Congress on Pattern Recognition. 2009. Springer.
26. García-Borroto, M., et al. Finding small consistent subset for the nearest neighbor classifier based on support graphs. in Iberoamerican Congress on Pattern Recognition. 2009. Springer.
27. Godínez, I.R., I. López-Yánez, and C. Yánez-Márquez, Classifying patterns in bioinformatics databases by using Alpha-Beta associative memories, in Biomedical Data and Applications. 2009, Springer. p. 187-210.
28. Moreno-Moreno, P., C. Yanez-Marquez, and O.A. Moreno-Franco, The new informatics technologies in education debate. International Journal of Technology Enhanced Learning, 2009. **1**(4): p. 327-341.
29. Rudas, I.J., et al. Digital fuzzy parametric conjunctions for hardware implementation of fuzzy systems. in 2009 IEEE International Conference on Computational Cybernetics (ICCC). 2009. IEEE.

30. Zavala, A.H., et al. Parametric operations for digital hardware implementation of fuzzy systems. in Mexican International Conference on Artificial Intelligence. 2009. Springer.
31. Zavala, A.H., et al. VLSI Implementation of a Module for Realization of Basic t-norms on Fuzzy Hardware. in 2009 IEEE International Conference on Fuzzy Systems. 2009. IEEE.
32. Cleofas-Sánchez, L., et al. Hybrid associative memories for imbalanced data classification: an experimental study. in Mexican Conference on Pattern Recognition. 2013. Springer.
33. Villuendas-Rey, Y., M.M. Garcia-Lorenzo, and R. Bello. Support Rough Sets for decision-making. in Fourth International Workshop on Knowledge Discovery, Knowledge Management and Decision Support. 2013. Atlantis Press.
34. Yanez-Marquez, C., et al., BDD-based algorithm for the minimum spanning tree in wireless ad-hoc network routing. IEEE Latin America Transactions, 2013. **11**(1): p. 600-601.
35. Zavala, A.H., et al., Conjunction and disjunction operations for digital fuzzy hardware. Applied Soft Computing, 2013. **13**(7): p. 3248-3258.
36. Aldape-Pérez, M., et al., Collaborative learning based on associative models: Application to pattern classification in medical datasets. Computers in Human Behavior, 2015. **51**: p. 771-779.
37. Guo-Hua, S., et al., Shannon information entropies for position-dependent mass Schrödinger problem with a hyperbolic well. Chinese Physics B, 2015. **24**(10): p. 100303.
38. López-Yáñez, I., et al., Collaborative learning in postgraduate level courses. Computers in Human Behavior, 2015. **51**: p. 938-944.
39. Gurrutxaga, I., Muguerza, J., Arbelaitz, O., Pérez J. M., Martín J. I., Towards a standard metodology to evaluate internal cluster validity indices. Pattern Recognition Letters, 2011. **32**: p. 505-515.
40. Cabrera-Venegas, J.F., BECA algorithm for data clustering (In Spanish). 2012, University of Ciego de Avila.
41. Hernadez-Echemendia, Y., DELUX, DELUXE and DELUXES algorithm for mixed data (In Spanish). 2012, University Máximo Gómez.
42. Karaboga, D., An Idea Based On Honey Bee Swarm For Numerical Optimization. 2005, Technical Report-TR06. Engineering Faculty. Computer Engineering Department, Erciyes University.
43. Salgado, I., et al., Super-twisting sliding mode differentiation for improving PD controllers performance of second order systems. ISA transactions, 2014. **53**(4): p. 1096-1106.
44. Villuendas-Rey, Y. and M.M. Garcia-Lorenzo, Attribute and case selection for nn classifier through rough sets and naturally inspired algorithms. Computación y Sistemas, 2014. **18**(2): p. 295-311.
45. Yáñez-Márquez, C., et al., Emerging computational tools: Impact on engineering education and computer science learning. International Journal of Engineering Education, 2014: p. 533-542.
46. Ortiz-Ángeles, S., et al., Electoral Preferences Prediction of the YouGov Social Network Users Based on Computational Intelligence Algorithms. J. UCS, 2017. **23**(3): p. 304-326.
47. Ramírez-Rubio, R., et al., Pattern classification using smallest normalized difference associative memory. Pattern Recognition Letters, 2017. **93**: p. 104-112.
48. Salgado, I., et al., Output feedback control of a skid-steered mobile robot based on the super-twisting algorithm. Control Engineering Practice, 2017. **58**: p. 193-203.
49. Salgado, I., et al., Adaptive control of discrete-time nonlinear systems by recurrent neural networks in quasi-sliding mode like regime. International Journal of Adaptive Control and Signal Processing, 2017. **31**(1): p. 83-96.
50. Villuendas-Rey, Y., et al., Simultaneous instance and feature selection for improving prediction in special education data. Program, 2017. **51**(3): p. 278-297.
51. Roy, D.K., Sharma, L. K., Genetic k-means clustering algorithm for mixed numeric and categorical datasets. International Journal of Artificial Intelligence & Applications (IJAIA), 2010. **1**(2): p. 23-28.
52. Xin-She. Y., Firefly algorithms for multimodal optimization. Lecture Notes in Computer Sciences, 2009. **5792**: p. 169-178.
53. López-Yáñez, I., L. Sheremetov, and C. Yáñez-Márquez, A novel associative model for time series data mining. Pattern Recognition Letters, 2014. **41**: p. 23-33.
54. Lytras, M.D., et al., The Social Media in Academia and EducationResearch R-evolutions and a Paradox: Advanced Next Generation Social Learning Innovation. J. UCS, 2014. **20**(15): p. 1987-1994.
55. Salgado, I., et al., Proportional derivative fuzzy control supplied with second order sliding mode differentiation. Engineering Applications of Artificial Intelligence, 2014. **35**: p. 84-94.
56. Ferreira-Santiago, A., et al., Enhancing engineering education through link prediction in social networks. International Journal of Engineering Education, 2016: p. 1566-1578.
57. Cerón-Figueroa, S., et al., Instance-based ontology matching for e-learning material using an associative pattern classifier. Computers in Human Behavior, 2017. **69**: p. 218-225.
58. Cerón-Figueroa, S., et al., Instance-based ontology matching for open and distance learning materials. The International Review of Research in Open and Distributed Learning, 2017. **18**(1).
59. García-Floriano, A., et al., Social Web Content Enhancement in a Distance Learning Environment: Intelligent Metadata Generation for Resources. International Review of Research in Open and Distributed Learning, 2017. **18**(1): p. 161-176.
60. Villuendas-Rey, Y., et al., The naïve associative classifier (NAC): a novel, simple, transparent, and accurate classification model evaluated on financial data. Neurocomputing, 2017. **265**: p. 105-115.
61. Antón-Vargas, J.A., et al., Improving the performance of an associative classifier by Gamma rough sets based instance selection. International Journal of Pattern Recognition and Artificial Intelligence, 2018. **32**(01): p. 1860009.
62. Barroso, E., Y. Villuendas, and C. Yanez, Bio-inspired algorithms for improving mixed and incomplete data clustering. IEEE Latin America Transactions, 2018. **16**(8): p. 2248-2253.
63. García-Floriano, A., et al., Support vector regression for predicting software enhancement effort. Information and Software Technology, 2018. **97**: p. 99-109.
64. González-Patiño, D., Y. Villuendas-Rey, and A.J. Argüelles-Cruz, The potential use of bioinspired algorithms applied in the segmentation of mammograms. 2018.
65. Hernández-Castaño, J.A., et al., Experimental platform for intelligent computing (EPIC). Computación y Sistemas, 2018. **22**(1): p. 245-253.
66. Dua, D. and C. Graff. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. 2019.
67. Huang, Z. Clustering large data sets with numeric and categorical values. in 1rst Pacific - Asia Conference on Knowledge discovery and Data Mining. 1997.
68. Ahmad, A., Dey L., A k-means type clustering algorithm for subspace clustering of mixed numeric and categorical data. Pattern Recognition Letters, 2011. **32**: p. 1062-1069.
69. Ahmed, R.A., et. al. HIMIC: A Hierarchical Mixed Type Data Clustering Algorithm. 2005. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.6369&rep=rep1&type=pdf.
70. He, Z., Xu, X., Deng, S. Clustering mixed numeric and categorical data: A cluster ensemble approach. CoRR http://arix.org/abs/cs/0509011, 2005.
71. Serrano-Silva, Y.O., Y. Villuendas-Rey, and C. Yáñez-Márquez, Automatic feature weighting for improving financial Decision Support Systems. Decision Support Systems, 2018. **107**: p. 78-87.
72. Villuendas-Rey, Y., et al., Medical Diagnosis of Chronic Diseases Based on a Novel Computational Intelligence Algorithm. J. UCS, 2018. **24**(6): p. 775-796.
73. Yáñez-Márquez, C., et al., Theoretical Foundations for the Alpha-Beta Associative Memories: 10 Years of Derived Extensions, Models, and Applications. Neural Processing Letters, 2018. **48**(2): p. 811-847.
74. González-Patiño, D., et al., A Novel Bio-Inspired Method for Early Diagnosis of Breast Cancer through Mammographic Image Analysis. Applied Sciences, 2019. **9**(21): p. 4492.
75. Wilson, R.D., Martinez T. R., Improved Heterogeneous Distance Functions. Journal of Artificial Intelligence Research, 1997. **6**: p. 1-34.
76. Villuendas-Rey, Y., Maximal similarity granular rough sets for mixed and incomplete information systems. Soft Computing, 2019. **23**(13): p. 4617-4631.
77. Villuendas-Rey, Y., et al., NACOD: A Naïve Associative Classifier for Online Data. IEEE Access, 2019. **7**: p. 117761-117767.
78. Villuendas-Rey, Y., et al., An Extension of the Gamma Associative Classifier for Dealing With Hybrid Data. IEEE Access, 2019. **7**: p. 64198-64205.

AUTHORS PROFILE

Yusbel Chávez-Castilla obtained his B.S. degree in Computer Science from the University of Ciego de Ávila, Cuba, in 2009, and the M.S. degree on Applied Informatics in 2012, from the same institution. He works as a professor of the Computer Science Department of the Faculty of Informatics, University of Ciego de Ávila. He is currently pursuing the Ph.D. degree on Computer Sciences at the Central University of Las Villas, Cuba. His research interests include image analysis, clustering, bio-inspired algorithms and computational complexity. He is a member of the ACPR (Cuban Association for Patter Recognition) and the Cuban Society of Mathematics, Physics, and Computation.