International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-1, November 2019

Using Extended Compact Sets to Cluster Educational Data

Carmen F. Rey -Benguría

Abstract*: *The pedagogical orientation to the families of children with behavioral deficiencies needs a differentiated approach, accorded to the particular characteristics of each family. To accomplish this task, the personnel in charge of family orientation in the Schools for children with affective-behavioral maladies, diagnose the families, to obtain the peculiarities of the familiar dynamics. After that, they need to obtain groups of similar families, in order to carry out a better orientation. The data of the familiar dynamics are mixed an incomplete, and the desired group number is not given. In this paper we extend the · 0-compact sets structuralization to cluster the families according to their characteristics. We found five groups, each of them with distinctive familiar dynamics. With the five groups, the family orientation specialists designed a personalized strategy with a more coherent and adequate orientation.

Keywords* : *clustering, education, family orientation, mixed data.

I. INTRODUCTION

In Cuba, the Education Ministry includes specialized educational schools for dealing with children with special educational needs. Among them, there are schools for Blind children, Depths children, Down syndrome children and Schools for children with affective-behavioural maladies (SABM). In this kind of schools, the adequate orientations to the family of the children play a key role in correcting the deficiencies, and in effectively insert the children into society. That is why the personnel in charge of the family orientation process in the SABM School of Ciego de Ávila characterize the familiar dynamics of each family, and then proceed to design a personalized strategy for each group of families with similar dynamics.

As known, clustering is one of the major tasks in Artificial Intelligence. It is devoted to finding a natural structure in a data set [1, 2]. Unlike its supervised counterpart, clustering lacks of information about class labels, or any other predefined distribution of instances in a particular data set [3]. Clustering algorithms use instance descriptions and instances dissimilarities in order to group together very similar instances, in a manner that it accomplishes the maxima of "higher cohesion and low coupling", meaning that within-group similarity must be as high as possible, and between-group similarity as low as possible.

Despite the challenges attached to clustering data, the need of structuralizing data in several domains imposes a must be for clustering techniques. Several domains have instances with different attribute types [4, 5]. For example, a customer description can include simultaneously attributes such as age (integer), sex (nominal), salary (real), educational degree (nominal), employed (Boolean), etc. [6-9]. In the particular case of SABM families, the attribute time in the school is numeric, and others such as type of crisis are nominal. This type of instance description is per say a challenge for any algorithm. The lacking of a metric space makes impossible the definition of a sum operator and also the scalar multiplication. In the same way, numeric attributes often have a large amount of values, each with low frequency. It makes frequency-based solutions developed for categorical data impracticable for numeric data.

In addition, the presence of missing values in instances descriptions makes it more complex for any classification procedure [3, 10-12]. Taking into account that dependencies may occur among attributes, estimating missing values is not always a feasible solution. On the other hand, there are many types of missing values, each of them with particular characteristics. That's why some authors have considered the best solution for clustering mixed datasets is to develop algorithms that are able to manage the absence of information, as well as mixed data types [13, 14].

In our research, we addressed the issue of clustering mixed data types social data, also including absences of information. We do not impose any restriction to the nature of the attributes. We consider clustering algorithms which obtain a partition C={c1,…, ck} of the input instances O, so that no instance belongs simultaneously to more than one group.

The article has the following structure: section 2 reviews some of the existing algorithms for free clustering of mixed and incomplete data. Section 3 explains the characteristics of the data of SABM families, as well as the dissimilarity used to compare the descriptions of the families. Section 4 explains the new Clustering algorithm based on Extended Compact Sets. Section 5 is devoted to the results obtained in SABM data. Also, we offer conclusions and future work

Revised Manuscript Received on November 30, 2019.
* Correspondence Author
Carmen F. Rey-Benguría*, Educational Center "José Martí", University of Ciego de Ávila, Cuba. Email: carmenrb2008@gmail.com

II. PREVIOUS WORKS

Although there is not a formal definition of clustering, there is a consensus on the scientific community that clustering consists on obtain several groups of instances, such that instances belonging to the same group are highly similar, and not similar to instances in other groups [15]. Clustering algorithms can be roughly classified according the type of data they handle (figure 1), but also by taking into account if the desired number of groups is given or not to the algorithm (figure 2). When we know a priori the desired number of groups, we make reference to the restricted clustering. If not, we make reference to free clustering.

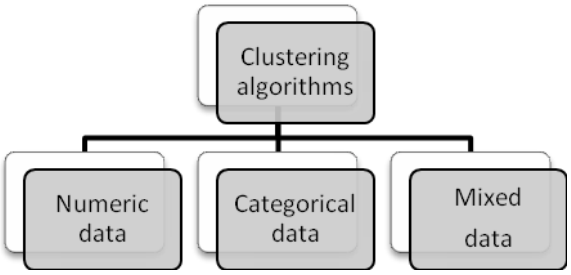

Fig. 1.Classification of clustering algoritms according to the type of data handled

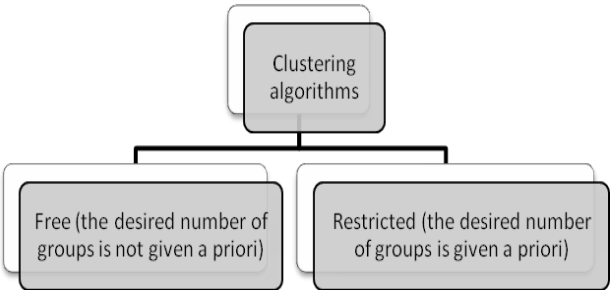

Fig. 2.Classification of clustering algoritms according to the dessired number of clusters

For a better understanding of the clustering problem, we briefly define the unsupervised classification problem as Ruiz-Shulcloper and Montellano-Ballesteros [16]:

Given the descriptions $I(O_i) = (X_1(O_i), \ldots, X_n(O_i))$, i =1,…,m, of instances $O_1, \ldots, O_m$ from a given universe Ω, in terms of a set of attributes $\Re = \{X_1, \ldots, X_n\}$. Each attribute X_j has associated a set of admissible values, $D_j \subseteq 2M_j$, j = 1,…,n, $X_j(O_i) \in D_j$. Usually the information about instances is given in a matrix form MI = $(X_j(O_i))$m x n with m rows (instance descriptions) and n columns (values of each attribute) [17, 18].

Let be $I(O_i) \in D_1 \times \ldots \times D_n$, i=1,…,m. the Cartesian space is the initial representation space (IRS) of instances. The nature of attributes will be, simultaneously, of any kind (qualitative, nominal, ordinal, numeric, etc.) [10, 19, 20]. We do not suppose any algebraic or topological structure over the IRS, that is, we do not assume any algebraic or logical operation over D_j, nor did any metric (distance) define a priori. The problem of the unsupervised classification (clustering) consists on finding the inner structure of a set of instance descriptions over the Initial Space of Representation, that is, to find a family $\zeta = \{C_1, \ldots, C_c\}$, where $C_i \subseteq$ MI, such that $\bigcup_{i=1}^{c} C_i = MI$. Sometimes, it is necessary that $C_i \cap C_j = \phi$, $i \neq j$ (partition).

By taking that into account, the acquisition of the inner structure of data depends of the selection of the IRS, of the definition of a similarity (dissimilarity) function Γ defined between instance descriptions and also of a clustering criterion Π, expressing the ways we can use Γ [16, 21]. This criterion will be responsible for determining which instances belong to which clusters. The previous expresses that the selection of the representation space, the dissimilarity function and the clustering criterion will be crucial to solve the clustering problem, and must be based on a solid knowledge of the particular problem we want to model.

There are several algorithms proposed for clustering mixed and incomplete data [1, 2, 13, 22, 23]. However, they need as a parameter the desired number of clusters. One of the most relevant free clustering algorithms are hierarchical agglomerative clustering algorithms, based on single-linkage, complete-linkage and average-linkage [24]. The hierarchical agglomerative clustering algorithms start with one instance at each cluster, and then merging at each stage, the two most similar clusters, until all instances are in a single cluster (algorithm 1). They can be easily extended to handle mixed and incomplete data by redefining the dissimilarity between clusters, as shown in table I, and using a dissimilarity function for mixed data to compare the instances. Such dissimilarities can be found in [25].

Algorithm # 1 Hierarchical Agglomerative Clustering (HAC)	
Input:	MI: matrix of instances d: dissimilarity function among groups
Output:	C: hierarchy of clustering at each step
Steps:	1. Each instance is considered as a group (leaf nodes of the tree C) 2. Merge the two less dissimilar groups, according to the dissimilarity function d 3. Repeat step 2, until all instances are in the same group

Fig. 3.Hierarchical Agglomerative Clustering

Table- I: Dissimilarity functions for clustering algorithms

Algorithms	Between groups dissimilarity
Single-Link	$d(C_i, C_j) = \min_{\substack{x \in C_i \\ y \in C_j}} \{d(x,y)\}$
Average-Link	$d(C_i, C_j) = \sum_{\substack{x \in C_i \\ y \in C_j}} d(x,y) \Big/ \lvert C_i \rvert * \lvert C_j \rvert$
Complete-Link	$d(C_i, C_j) = \max_{\substack{x \in C_i \\ y \in C_j}} \{d(x,y)\}$

Another algorithms for clustering mixed data without defining a priori the number of clusters are the structuralizations based on the β_0-simmilarity criteria (connected, compacts and strong compacts) proposed by [16]. In the β_0-simmilarity based algorithms, the clusters are formed by the instances satisfying the definitions according to the selected criterion. In the following we will address these algorithms.

A. Clustering based on β_0-connected groups

The β_0-connected structuralization consists on associate two instances if its similarity is greater than β_0. According to that, each instance has a minimum within group similarity of at least β_0, and a maximum without group similarity greater than β_0. Formally, the groups are formed according to the following definition:

Definition 1. A subset $NU_r \neq \varnothing$ of MI is a β_0 *-connected group if and only if*:

a) $\forall O_i, O_j \in NU_r \; \exists O_{i_1}, \ldots, O_{i_q} \in NU_r [O_i = O_{i_1} \wedge O_j = O_{i_q} \wedge \forall p \in \{1, \ldots, q-1\}, \Gamma(O_{i_p}, O_{i_{p+1}}) \geq \beta_0]$

b) $\forall O_i \in MI \; [(O_j \in NU_r \wedge \Gamma(O_i, O_j) \geq \beta_0) \Rightarrow O_i \in NU_r]$

c) Every instance β_0 -isolated is a β_0 *-connected* (degenerated).

Algorithm # 2 β_0 -connected clustering	
Input:	MI: matrix of instances β_0: similarity threshold Γ: similarity function between instances
Output:	C: resulted clustering
Steps:	1. For each pair of instances, if its similarity is greater or equal to β_0, assign them to the same group, according to the definition 1. 2. Add to C each group of instance created at step 1.

Fig. 4.β_0-connected clustering

Although the algorithm is defined in terms of a similarity function, it can be defined also in terms of dissimilarities (grouping together instances with dissimilarities lower than β_0). To illustrate the β_0-connected clustering in terms of dissimilarities, we will use the set of instances in figure 5 (nine points in R^2) and we will use the Euclidean distance as a dissimilarity function among them.

In figure 6 we show the resulted clustering of the instances of figure 7, using the β_0-connected structuralization with a $\beta 0$ threshold equal to 0.5.

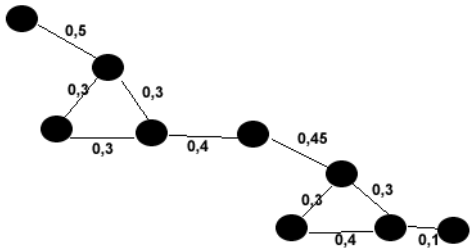

Fig. 5.Points in 2D, and distances among them.

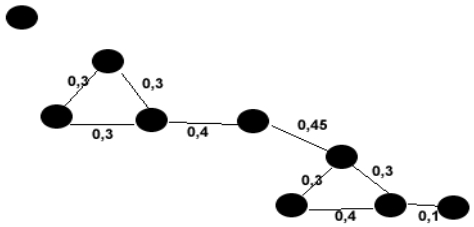

Fig. 6.Results of the β_0-connected clustering.

B. Clustering based on β_0-compact groups

Good The β_0-compact structuralization consists on associate each instance with its most similar instance, if their similarity is greater than β_0. So, a compact group is a connected component of a maximum similarity group. Formally, the groups are formed according to the following definition:

Definition 2. A subset $NU_r \neq \varnothing$ de MI is a β_0 *-compact group* if and only if:

a) $\forall O_j \in M \quad [O_i \in NU_r \wedge (\max_{\substack{O_t \in MI \\ O_t \neq O_i}} \{\Gamma(O_i, O_t)\} = \Gamma(O_i, O_j) \geq \beta_0 \vee \max_{\substack{O_t \in MI \\ O_t \neq O_j}} \{\Gamma(O_j, O_t)\} = \Gamma(O_j, O_i) \geq \beta_0)] \Rightarrow O_j \in NU_r$

b) $\forall O_i, O_j \in NU_r \; \exists O_{i_1}, \ldots, O_{i_q} \in NU_r [O_i = O_{i_1} \wedge O_j = O_{i_q} \wedge \forall p \in \{1, \ldots, q-1\}$

$[\max_{\substack{O_t \in MI \\ O_t \neq O_{i_p}}} \{\Gamma(O_{i_p}, O_t)\} = \Gamma(O_{i_p}, O_{i_{p+1}}) \geq \beta_0 \vee \max_{\substack{O_t \in MI \\ O_t \neq O_{i_{p+1}}}} \{\Gamma(O_{i_{p+1}}, O_t)\} = \Gamma(O_{i_p}, O_{i_{p+1}}) \geq \beta_0]]$

c) Every isolated instance is a β_0 *-compact group* (degenerated).

Algorithm # 3 β_0 -compact clustering	
Input:	MI: matrix of instances β_0: similarity threshold Γ: similarity function between instances
Output:	C: resulted clustering

Steps:	3. Create a maximum β_0-similarity graph. 4. Add to C each connected component of the graph created at step 1.

Fig. 7.β_0-compact clustering

In figure 8 we show the resulted clustering of β0-compact structuralization, with a β0 threshold equal to 0.5.

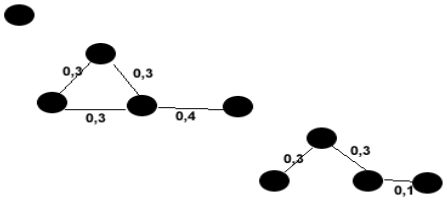

Fig. 8.Results of the β_0-compact clustering.

C. Clustering based on β_0-strong compact groups

The β_0-compact structuralization consists on associate each instance o with its most similar instance m, if the similarity between o and m is greater than β_0, and if the similarity of m with respect to o is also maximum. A strong compact group is a strong connected component of a maximum similarity group. Formally, the groups are formed according to the following definition:

Definition 3. A subset $NU_r \neq \varnothing$ of *MI* is a β_0 *-strong compact group* if and only if:

a) $\forall O_j \in MI\,[O_i \in NU_r \wedge \max_{\substack{O \in MI \\ O \neq O_i}} \{\Gamma(O_i,O)\} = \Gamma(O_i,O_j) \geq \beta_0\,]$

$\Rightarrow O_j \in NU_r$

b) $\exists O_i \in NU_r, \forall O_j \in NU_r, \exists\, O_{i_1},\ldots,O_{i_q}$

$\in NU_r[O_i = O_{i_1} \wedge O_j = O_{i_q} \wedge \forall p<q[\max_{\substack{O \in MI \\ O \neq O_{i_p}}} \{\Gamma(O_{i_p},O)\}$

$= \Gamma(O_{i_p}, O_{i_{p+1}}) \geq \beta_0\,]]$

c) There is no NU_r' that fulfils a) and b) such that $NU_r \subset NU_r'$.

d) Every instance β_0 -isolated is a β_0 *-strong compact group* (degenerated).

Algorithm # 4 β_0 -strong compact clustering

Input:	MI: matrix of instances β_0: similarity threshold Γ: similarity function between instances
Output:	C: resulted clustering
Steps:	5. Create a maximum β_0-similarity graph. 6. Add to C each strong connected component of the graph created at step 1.

Fig. 9.β_0-strong compact clustering.

In figure 10 we show the results for the β_0-strong compact clustering.

The β_0-similarity based clustering algorithms allow using a similarity (dissimilarity) function according to the particular problem we want to solve. However, setting an adequate value of β_0 is complex in practice, and this value is crucial to obtain high quality clusters.

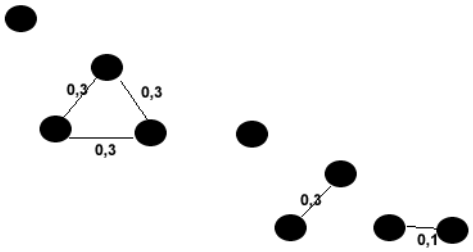

Fig. 10. Results of the β_0-strong compact clustering, with β_0 = 0.5.

To solve the mentioned drawbacks, several researchers have used compact sets without setting the β_0 threshold, only connecting each instance to its most similar instance, regarding the similarity value [26, 27].

Although compact set have been useful in Artificial Intelligence tasks, the definition of maximum similarity implies that each instance is only connected to its most similar one, not taking into account other instances that might be slightly less similar, but also very similar to the analysed instance [28-32].

That is why we extended the compacts sets, by relaxing its definition, and use this extension to clustering the families of the Schools for children with affective-behavioural maladies SABM.

III. CHARACTERIZATION OF THE SABM FAMILIES TO DESIGN A HIGH QUALITY FAMILY ORIENTATION PROCESS

Several researchers have considered the usefulness of supervised and unsupervised classification of social data [33-41]. Our research follows such path.

The process of clustering data of the families with children at the School for children with affective-behavioural maladies in Ciego de Avila city was carried out in three stages. At the first stage, the personnel in charge of the family orientation process characterize every family, by using attributes describing the familiar dynamics and the relation of the family with the School. At this stage, the specialists determine which elements or parameters could be used to diagnose the current dynamic in the family, the way the family accept the inclusion of the child in the School, etc.

At the second stage, we define in conjunction of the specialists, the way to determine the similarity between two family descriptions. By doing this we determine the similarity functions that models the likeness between families, and we establish relations of similarity and dissimilarity among families. At the third stage, we cluster the families by using the new developed algorithm. In this section, we will explain the first and second stages of the process.

A. First stage. Family characterization

To characterize the families, we take into account the attributes described in table II.

This attributes measure the attitude of the family to the inclusion of a child in the School for children with affective-behavioural maladies SABM, as well as the peculiarities of the family dynamic. It is important to mention that to obtain a successful depart of children from the School, they need a huge support of the family, to correct the deficiencies, and effectively insert the children into society.

Table- II: Attributes used to characterize the families

Att.	Name	Values	Description
att1	Impact or shock	yes, no, ?[a]	If exists impact or shock in the family when they are notified that a child will be included at the Centre
att.2	attitude	acceptance, rejection, ?	The attitude adopted to the inclusion of a child in the Centre
att.3	Embracing change	O, R, G, A, ?	How the family adopts the change, if they oppose (O), they resists (R), they have resignation (G) or they agree (A)
att.4	Guilty	yes, no,?	If there is or not guilty feelings in the family
att.5	Emotional clime	positive, negative, ?	The kind of emotional clime, if it is positive or negative
att.6	communication	Functional, dysfunctional, ?	The kind of communication that prevail in the family
att.7	handling	Control, decontrol, ?	The way the family handles the fact of including a child into the Centre
att.8	relations	good, bad,?	The way the interpersonal relations are developed into the family
att.9	emotional crisis	D, O, I, F, no, ?	The kind of emotional crisis, by demoralization (D), disarranging (O), frustration (F) , impotence (I) or no crisis
att.10	self estimation	high, low, ?	The way the self estimation of the family is
att.11	consciousness	yes, no,?	If there is or not consciousness of the reality the family faces
att.12	linkage	good, bad, ?	If there is or not a favorable link with the Centre
att.13	hopes to the future	optimistic, pessimistic,?	The hopes the family has to the future
att.14	time	integer	The time (in months) the child is at the Centre

[a] The ? symbol represents an unknown datum, it cannot be obtained by the specialist

B. Second stage. Definition of the similarity function to compare the families

To compare the families, we design a similarity function to compare the families (equation 1). We take into consideration the criteria of the personnel in charge of the family orientation process at the SABM. The similarity function is defined as the opposite of the sum of the dissimilarities among all attributes. For each attribute but the third (embracing change), if two instances have the same values, there is a maximum similarity. Otherwise, if the values are known and different, there is a minimum similarity [42-44]. If only one of the values is known, there is an average similarity. Let be two families, f_i and f_j:

$$S(f_i, f_j) = 1 - D(f_i, f_j)$$

$$\text{where } D(f_i, f_j) = \sum_{k=1}^{14} D_k(f_i, f_j)$$

$$D_k(f_i, f_j) = \begin{cases} 0 & if \quad X_k(f_i) = X_k(f_j) \\ 0{,}5 & if \quad X_k(f_i) = ? \vee X_k(f_j) = ? \\ 1 & if \quad X_k(f_i) \neq X_k(f_j) \end{cases} \quad (1)$$

$$k \in [1,14], k \neq 3$$

In case of the third attribute, “embracing change”, the different attribute values have a peculiar meaning, at their similarity depends of each value combination. The embracing change attribute defines the attitude the family adopts to face the fact that one of the family members, a child, will be intern into the SABM. By taking this into consideration, the specialists consider the values opposition and resistance as very similar attitudes (with dissimilarity value of 0.2, bidirectional). Also, they judge the acceptance and resignation, although different, have things in common and they assigned the dissimilarity value of 0.4, also bidirectional.

On the contrary, the attitudes of resistance and resignation have a non-symmetric similarity, because a family which has resistance to change is similar to a family which has resignation to change, in the sense that both families are incapable to avoid change, but a family which has resignation to change is less similar to a family which has resistance to change, because the family with resignation is acquiescence, do not fight back, etc. Taking that into consideration, the specialists decided to assign the following dissimilarity values: (resistance vs. resignation equal to 0.4), and (resignation vs. resistance equal to 0.8). In table III we show the comparison matrix of values for the attribute “embracing change”.

Table- III: Comparison matrix of the values for the attribute “embracing change”. Each cell shows the dissimilarity values of the pair (row vs. column)

Value	Opposition (O)	Resistance (R)	Resignation (G)	Agreement (A)
Opposition (O)	0	0.2	0.8	1
Resistance (R)	0.2	0	0.4	0.8
Resignation (G)	0.8	0.8	0	0.4
Agreement (A)	1	0.8	0.4	0

C. Third stage. Clustering families

Taking into consideration the nature of the problem we want to solve (mixed and incomplete data, very delicate situation of social sciences, the clustering of the families will decide the kind of family orientation strategy for every group, non-symmetric dissimilarity function, etc.) we decide to use the clustering based on compact sets. In the next section we address the extensions we made to the compact sets clustering [16], as well as the new proposed algorithm.

IV. CLUSTERING ALGORITHMS BASED ON EXTENDED COMPACT SETS

The compact sets clustering algorithm, described before, despite handles mixed and incomplete data, as well as non-symmetric similarities, depends of the selection of the β_0 parameter. As stated before, several researchers have used the β_0 value equal to 0, allowing each instance to connect to its most similar instance in the dataset. However, the compact set only takes into consideration the most similar instances, and discards other instances that might be highly similar. In this research we extended the maximum similarity graph [6, 8, 9, 45, 46] to a dynamic maximum similarity graph, that is, every instance is connected dynamically to its most similar instances. We also developed a procedure for obtained the dynamic compact sets.

Algorithm# 5. Dynamic compact set clustering

Inputs:	MI: matrix of instances Γ: dissimilarity function ε: difference ratio
Output:	C: resulted clustering
Steps:	1. M ← MI 2. For each instance o_i in MI a. $\Gamma Min_i = 0$ b. $Max_i=0$ c. InstanceSearch(MI, Γ, G, o_i, ΓMin_i, Max_i) 3. Add to C each connected component of the graph obtained at step 2.

Fig. 11. Dynamic compact set clustering.

Algorithm# 6. InstanceSearch

Inputs:	MI: matrix of instances Γ: dissimilarity function G: dissimilarity graph o_i: instance to analyse ΓMin_i: minimum dissimilarity value for instance Max_i: maximum permitted difference
Steps:	1. If $MI \neq \phi$ 2. Find o_i nearest instance (*no*) in MI a. If $(\Gamma(o_i, no) - \Gamma Min_i) > Max_i$ i. Connect o_i and *no* in graph G ii. $Max_i=\varepsilon*\Gamma(o_i, no)$ i. $MI \leftarrow MI - \{no\}$ b. Else return c. InstanceSearch(MI, Γ, G, o_i, ΓMin_i, Max_i) 3. Else return

Fig. 12. Algorithm for instance searching

We obtain five groups, each of them with families with similar characteristics. The first group was formed by families which child will be interned at the School, but they have not been internet yet. All of them present rejection and opposition to the inclusion of the child at the School, and also have a dysfunctional communication. The second group was formed by pessimistic families, with bad relations with the SABM, diminished self estimation and a dysfunctional communication. The third group has optimistic families, with good relations with the School, but with a diminished self estimation and a dysfunctional communication. The fourth group has optimistic, functional families, with an elevated self estimation and good relations with the School. The last group has families with good relations with the Centre, but which are pessimistic and have a dysfunctional communication.

Taking into account the peculiarities of the founded groups, the personnel in charge of the family orientation process at the SABM design a personalized pedagogical strategy, to give a better orientation to each group of families.

V. CONCLUSION

Cuban Special Education has focused its efforts to obtain an integral educational process, in which intervene the family and the school, as the main educational contexts for the development of the child personality. Among that, an adequate orientation process to the families with children with affective-behavioural maladies plays a key role at Cuban Specialized Schools for the treatment of these children. In this investigation, we developed a novel clustering algorithm based on compact sets to obtain groups of families. By taking into consideration the peculiarities of the families at each group, the sociologists and pedagogues in charge of the family orientation process were able to design a personalized strategy to give a better orientation, and to get a faster journey of the kids for the School for children with Affective-Behavioural Maladies, in the Cuban city of Ciego de Avila.

REFERENCES

1. Shih, M.-Y., J.-W. Jheng, and L.-F. Lai, *A two-step method for clustering mixed categorical and numeric data.* Tamkang Journal of Science and Engineering, 2010. **3**(1): p. 11-19.
2. Ahmad, A. and L. Dey, *A k-means type clustering algorithm for subspace clustering of mixed numeric and categorical data.* Pattern Recognition Letters, 2011. **32**: p. 1062-1069.
3. 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.
4. 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.
5. 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.
6. Medina-Pérez, M.A., et al. *Selecting objects for ALVOT*. in *Iberoamerican Congress on Pattern Recognition*. 2006. Springer.
7. Villuendas-Rey, Y., et al. *Simultaneous features and objects selection for Mixed and Incomplete data*. in *Iberoamerican Congress on Pattern Recognition*. 2006. Springer.
8. García-Borroto, M., et al. *Using maximum similarity graphs to edit nearest neighbor classifiers*. in *Iberoamerican Congress on Pattern Recognition*. 2009. Springer.

9. 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.
10. 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.
11. 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.
12. 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.
13. Naija, Y., et al. *Extension of Partitional Clustering Methods for Handling Mixed Data*. in *IEEE International Conference on Data mining Workshops ISDMW*. 2008.
14. Roy, D.K. and L.K. Sharma, *Genetic k-means clustering algorithm for mixed numeric and categorical datasets.* International Journal of Artificial Intelligence & Applications (IJAIA), 2010. **1**(2): p. 23-28.
15. Brun, M., et al., *Model-based evaluation of clustering validation measures.* Pattern Recognition, 2007. **40**: p. 807-824.
16. Ruiz-Shulcloper, J. and J.J. Montellano-Ballesteros. *A new model of fuzzy clustering algorithms*. in *3th European Congress on Fuzzy and Intelligent Technologies and Soft Computing*. 1995. Aachen, Germany.
17. Villuendas-Rey, Y., et al., *Simultaneous instance and feature selection for improving prediction in special education data.* Program, 2017. **51**(3): p. 278-297.
18. 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.
19. 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.
20. 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.
21. Martínez-Trinidad, J.F. and J. Ruiz-Shulcloper. *Fuzzy semantic clustering*. in *4th European Congress on Intelligent Techniques and Soft Computing*. 1996. Aachen, Germany.
22. Ahmad, A. and L. Dey, *A k-means clustering algorithm for mixed numerical and categorical data.* Data & Knowledge Engineering, 2007. **63**: p. 503-527.
23. Roy, D.K. and L.K. Sharma, *Genetic k-means clustering algorithm for mixed numeric and categorical datasets.* International Journal of Artificial Intelligence & Applications (IJAIA), 2010. **1**(2).
24. Jain, A.K. and R.C. Dubes, *Algorithms for clustering data*. 1988, New Jersey, USA: Prentice Hall.
25. Wilson, R.D. and T.R. Martinez, *Improved Heterogeneous Distance Functions.* Journal of Artificial Intelligence Research, 1997. **6**: p. 1-34.
26. Villuendas-Rey, Y., et al., *Simultaneous features and objects selection for Mixed and Incomplete data.* Lecture Notes in Computer Science, 2006. **4225**: p. 597-605.
27. Villuendas-Rey, Y., et al., *Selecting features and objects for mixed and incomplete data.* Lecture Notes in Computer Science, 2008. **5197**: p. 381-388.
28. García-Floriano, A., et al., *Support vector regression for predicting software enhancement effort.* Information and Software Technology, 2018. **97**: p. 99-109.
29. González-Patiño, D., Y. Villuendas-Rey, and A.J. Argüelles-Cruz, *The potential use of bioinspired algorithms applied in the segmentation of mammograms.* 2018.
30. Hernández-Castaño, J.A., et al., *Experimental platform for intelligent computing (EPIC).* Computación y Sistemas, 2018. **22**(1): p. 245-253.
31. 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.
32. 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.
33. 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.
34. Acevedo, M.E., C. Yáñez-Márquez, and M.A. Acevedo, *Associative models for storing and retrieving concept lattices.* Mathematical Problems in Engineering, 2010. **2010**.
35. 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.
36. López-Yáñez, I., et al., *Collaborative learning in postgraduate level courses.* Computers in Human Behavior, 2015. **51**: p. 938-944.
37. 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.
38. 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).
39. 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.
40. 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.
41. 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.
42. Villuendas-Rey, Y., *Maximal similarity granular rough sets for mixed and incomplete information systems.* Soft Computing, 2019. **23**(13): p. 4617-4631.
43. Villuendas-Rey, Y., et al., *NACOD: A Naïve Associative Classifier for Online Data.* IEEE Access, 2019. **7**: p. 117761-117767.
44. Villuendas-Rey, Y., et al., *An Extension of the Gamma Associative Classifier for Dealing With Hybrid Data.* IEEE Access, 2019. **7**: p. 64198-64205.
45. 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.
46. 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.

AUTHORS PROFILE

Carmen F. Rey-Benguría obtained her Bachelor degree on Preschooler Pedagogy and Psychology at Hersen State Pedagogical Institute in Saint Petersburg, Russia. Her MSc and Ph.D. (2002 and 2005, respectively) degrees on Pedagogical Sciences were received at Universidad de Oriente, Cuba. Currently, she is with Universidad de Ciego de Ávila, Cuba, as the head of the Ph.D. program in Educational Sciences of the Educational Center "José Martí". Her research interests include communication, psychological strategies, artificial intelligence in society and machine learning. She had received the Ministry of Education award to the best post-graduate teacher, and the University of Ciego de Ávila award for the best Ph.D. program coordination.