Applied multi-layer clustering to the diagnosis of complex agro-systems

Elisa Roux[1], **Louise Travé-Massuyès**[1] and **Marie-Véronique Le Lann**[1,2]
[1]CNRS, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France
emails: lisa.roux@laas.fr, louise@laas.fr, mvlelann@laas.fr
[2]Univ de Toulouse, INSA, LAAS, F-31400 Toulouse, France

Abstract

In many fields, such as medical, environmental, a lot of data are produced every day. In many cases, the task of machine learning is to analyze these data composed of very heterogeneous types of features. We developed in previous work a classification method based on fuzzy logic, capable of processing three types of features (data): qualitative, quantitative, and more recently intervals. We propose to add a new one: the object type which is a meaningful combination of other features yielding the possibility of developing hierarchical classifications. This is illustrated by a real-life case study taken from the agriculture area[1].

1 Introduction

Nowadays, large scale datasets are produced in various different fields such as social networks, medical, process operation, agricultural/environmental,... Many studies relate to data mining with the intention of analyzing and if possible extracting knowledge from these data. The data classification has to provide a relevant and well-fitted representation of reality. In this context, the issue of representing of data is crucial since the formalisms must be generic yet well suited to every new problem. For machine learning, the concern is to be able to detect adequate patterns from heterogeneous, large, and sometimes uncertain datasets. In diagnosis, the necessity to quickly recognize a problem to provide a sure solution to solve it appears to be essential. One of the main challenges is the necessity to process heterogeneous data (qualitative, quantitative...) and sometimes to merge data obtained in different contexts. We developed a classification method based on fuzzy logic [1] capable of processing heterogeneous data types and noisy data. The LAMDA (Learning Algorithm for Multivariate Data Analysis) method is a classification method, capable to process three types of data: qualitative, quantitative, and intervals [2]. We addressed one of the main difficulties encountered in data analysis tasks: the diversity of information types. Such information types are given by qualitative valued data, which can be nominal or ordinal, mixed with quantitative and interval data. Many situations leading to well-conditioned algorithms for quantitative valued information become very complex whenever there are several data given in qualitative form. In a non-exhaustive list, we can mention, rule based deduction, classification, clustering, dimensionality reduction… During the last decades, few research works have been directed to defy the issue of representing multiplicity for data analysis purposes [3, 11]. However, no standard principle has been proposed in the literature to handle in a unified way heterogeneous data. Indeed, a lot of proposed techniques process separately quantitative and qualitative data. In data reduction tasks for example, they are either based on distance measures for the former type [12] and on information or consistency measures for the later one. Whereas in classification and clustering tasks, eventually only a Hamming distance is used to handle qualitative data [4,11,14]. Other approaches are originally designed to process only quantitative data and therefore arbitrary transformations of qualitative data into a quantitative space are proposed without taking into account their nature in the original space [12,15,16]. For example, the variable *shape* can take values in a discrete unordered set {round, square, triangle}. These values are transformed respectively to quantitative values 1, 2, and 3. However, we can also choose to transform them to 3, 2 and 1. Another inverse practice is to enhance the qualitative aspect and discretize the quantitative value domain into several intervals, then objects in the same interval are labeled by the same qualitative value [17,18]. Obviously, both approaches introduce distortion and end up with information loss with respect to the original data. Moreover, none of the previously proposed approaches combines in a fully adequate way, the processing of symbolic intervals simultaneously with quantitative and qualitative data. Although extensive studies were performed to process this type of data in the Symbolic Data Analysis framework [19], they were focused generally on the clustering tasks [8, 10] and no unified principle was given to handle simultaneously the three types of data for different analysis purposes. In [2], a new general principle, was introduced as "Simultaneous Mapping for Single Processing (SMSP)", enabling the reasoning in a unified way about heterogeneous data for several data analysis purposes. The fact that SMSP together with

[1]This work was supported by the FUI/FEDER project MAISEO involving the companies VIVADOUR, CACG, GEOSYS, METEO FRANCE, PIONEER and laboratories CESBIO, LAAS-CNRS.

LAMDA can process simultaneously these three types of data without pre-processing is one of its principal advantages compared to other classical machine learning methods such as SVM (Support Vector Machine [20]), K-NN [21]. Decision trees are very powerful tools for classification and diagnosis [22] but their sequential approach is still not advisable to process multidimensional data since, by their very nature, they cannot be processed as efficiently as totally independent information [23]. A complete description of the LAMDA method and comparison with other classification techniques on various well known data sets can be found in [24, 25, 26]. Its other main characteristic is the fuzzy formalism which enables an element to belong to several classes simultaneously. It is also possible to perform clustering (i.e. with no a priori knowledge of the number and the class prototypes).

Besides the three existing types, we propose to add another type: the class type which can be processed simultaneously with the three former ones: quantitative, qualitative, intervals thanks to the "SMSP". In this configuration the class feature represents a meaningful aggregation of other features. This aggregation can be defined by a class determined by a previous classification, or the result of an abstraction. This new type gives the possibility to develop hierarchical classifications or to fuse different classifications. It allows an easier representation of many various and complex types of data, like multi-dimensional data, while being realistic and conserving their constraints. In a first part, the LAMDA method is briefly explained. The second part is devoted to the new type of data introduced: the object type. Finally, this new method is exemplified through an agronomical project.

2 The LAMDA method

The LAMDA method is an example of fuzzy logic based classification methods [9]. The classification method takes as input a sample x made up of N features. The first step is to compute for each feature of x, an adequacy degree to each class C_k , $k = 1..K$ where K is the total number of classes. This is obtained by the use of a fuzzy adequacy function providing K vectors of Marginal Adequacy Degree vectors (MAD). This degree estimates the closeness of every single sample feature to the prototype corresponding to its class. At this point, all the features are in a common space. Then the second step is to aggregate all these marginal adequacy degrees into one global adequacy degree (GAD) by means of a fuzzy aggregation function. Thus the K MAD vectors become K GADs. Fuzzy logic[1] is here used to express MADs and GADs, since the membership degree of a sample to a given class is not binary but takes a value in [0,1]. Classes can be known a priori, commonly determined by an expert and the learning process is therefore supervised, or classes can created during the learning itself (unsupervised mode or clustering). Three types of features can be processed by the LAMDA method: quantitative, qualitative and intervals for the MAD calculation [2]. The membership functions μ(x) used by LAMDA are based on the generalization of a probabilistic rule defined on 0, 1 to the [0,1]-space.

2.1 Calculation of MAD for quantitative features

The quantitative type allows the representation of numerical values, assuming that the including space is known as a defined interval. For this type of descriptor, membership functions can be used, such as the Gaussian membership function so that the membership function for the x^{th} sample descriptor to the k^{th} class is:

$$\mu_k^i(x_i) = \exp^{\frac{-(x_i-\rho_k^i)^2}{2\sigma_i^2}} \tag{1}$$

or the binomial membership function:

$$\mu_k^i(x_i) = \rho_k^{i\ x_i}\left(1-\rho_k^i\right)^{(1-x_i)} \tag{2}$$

where:

$\rho_k^i \in$ [0, 1] is the mean of the i^{th} feature based on the samples belonging to the class C_k, $x_i \in [0, 1]$ is the normalized x^{th} feature and σ_i the standard deviation of the i^{th} feature value based on the samples belonging to the class C_k.

2.2 Calculation of MAD for qualitative features

In case of qualitative feature, the possible values of the i^{th} feature forms a set of modalities such as $D_i = Q_1^i, Q_2^i \ldots Q_m^i$ with m the total number of modalities. The qualitative type permits to express by words the different modalities of a criterion.

The frequency of a modality Q_l^i of the i^{th} feature for the class C_k is the quantity of samples belonging to C_k whose modality for their i^{th} feature is Q_l^i [1].So each modality $Q_l^i \in D_i$ has an associated frequency. Let θ_{kj}^i be the frequency of a modality Q_j^i for the class C_k. The membership function concerning the i^{th} feature is:

$$\mu_k^i(x_i) = \left(\theta_{k1}^i\right)^{q_1^i} * \left(\theta_{k2}^i\right)^{q_2^i} * \ldots * \left(\theta_{km}^i\right)^{q_m^i} \tag{3}$$

where q_l^i =1 if $x_i = Q_l^i$ and q_l^i = 0 otherwise, for l=1, ..m.

2.3 Calculation of MAD for interval features

Finally, to take in account the potential uncertainties or noises in data, we can use the interval representation [2]. The membership function for the interval type descriptors is regarded as being the similarity $S(x_i, \rho_k^i)$ between the symbolic interval value for the i^{th} feature x_i and the interval $[\rho_k^{i-}, \rho_k^{i+}]$ which represents the value of the i^{th} feature for the class C_k, so that:

$$\mu_k^i(x_i) = S(x_i, \rho_k^i) \tag{4}$$

Let ω be defined as the scalar cardinal of a fuzzy set in a discrete universe as $\varpi[X] = \sum_{\xi \in V} \mu_X(\xi_i)$.

In case of a crisp interval, it becomes:
$\varpi[X] =$ upperBound(X)- lowerBound(X).
Given two intervals A=[a^-, a^+] and B=[b^-, b^+], the distance is defined as:

$$\delta[A,B] = \max\left[0, \left(\max\{a^-, b^-\} - \min\{a^+, b^+\}\right)\right] \quad (5)$$

and the definition of the similarity measure between two crisp intervals:

$$S(I_1, I_2) = \frac{1}{2}\left(\frac{\varpi[I_1 \cap I_2]}{\varpi[I_1 \cup I_2]} + 1 - \frac{\delta[I_1, I_2]}{\varpi[V]}\right) \quad (6)$$

The similarity combines the Jaccard's similarity measure which computes the similarity when the intervals overlapp, and a second term which allows taking into account the case where the intervals are not straddled.

2.4 Calculation of feature weights

It is possible to determine the relevance of a feature to optimize the separation between classes. The MEMBAS method [8, 9] is a feature weighting method based on a membership margin. A distinguishable property of this method is its capability to process problems characterized by mixed-type data (quantitative, qualitative and interval). It lies on the maximization of the margins between two closest classes for each sample. It can be expressed as:

$$\underset{w_f}{\mathrm{Max}} \sum_{j=1}^{J} \beta_j(w_f) = 1/N \sum_{j=1}^{J} \left\{ \begin{array}{l} \sum_{i=1}^{N} w_{fi}\mu_c^i(x_i^{(j)}) \\ -\sum_{i=1}^{N} w_{fi}\mu_{\tilde{c}}^i(x_i^{(j)}) \end{array} \right\} \quad (7)$$

Subject to the following constraints: $\| w_f \|_2^2 = 1,\ w_f \geq 0$.

The first constraint is the normalized bound for the modulus of w_f so that the maximization ends up with non-infinite values, whereas the second guarantees the nonnegative property of the obtained weight vector. Then can be simplified as:

$$\begin{array}{l} \underset{w_f}{\mathrm{Max}} \ (w_f)^T s \\ \text{Subject to } \ \| w_f \|_2^2 = 1, w_f \geq 0 \end{array} \quad (8)$$

where: $s = 1/N \sum_{j=1}^{J} \{U_{jc} - U_{j\tilde{c}}\}$ and $U_{jc} = \left[\mu_c^1\left(x_i^{(j)}\right), \cdots, \mu_c^N\left(x_i^{(j)}\right)\right]$, $\mu_k^i\left(x_i^{(j)}\right)$ is the membership function of class c (c corresponds to the "right" class for sample $x^{(j)}$, $\tilde{c}$ the closest class evaluated at the given value $x_i^{(j)}$ of the i^{th} feature of pattern $x^{(j)}$. *s* is computed with respect to all samples contained in the data base excluding $x^{(j)}$ ("leave-one-out margin").

This optimization problem has an analytical solution determined by the classical Lagrangian method. Details of the method can be found in [9].

3 The new object type

In order to allow the combination of various data types into one single global object and therefore to support multi-dimensional features, we develop a novel data type. Each feature of an object descriptor can be described by a measured value and an extrinsic object-related weight. A sample GAD calculus formula is then the weighted mean of all MADs:

$$GAD_k^j = \sum\left(MAD_k^{ji} . \tilde{w}_{fi}\right) \text{ for } j=1\ldots J \quad (9)$$

where MAD_k^{ji} = MAD of the j^{th} sample for the i^{th} feature to class *k* and $\tilde{w}_{f_i} \in [0,1]$ = Normalized value of weight w_{f_i} of the i^{th} feature determined by the MEMBAS method, and *J* is the total number of samples which have been classified.

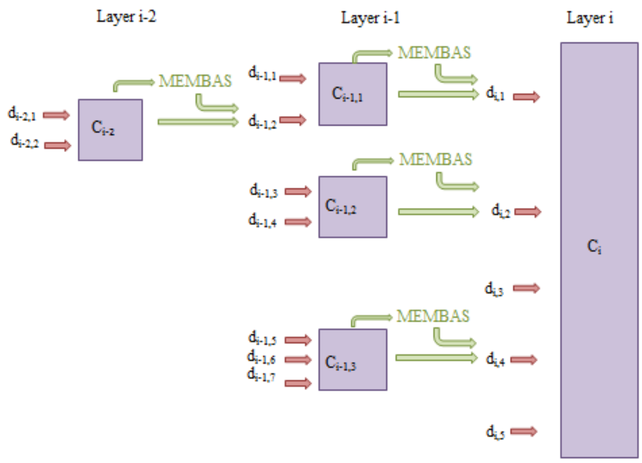

Figure 1: LAMDA architecture

The main advantage of using this new object-oriented data type is to capture the distinct features of a same object as a whole. An object of layer *i-1* is regarded as one single feature for the layer *i* then can be processed as all other descriptors. The weights of the descriptors composing the objects are determined using MEMBAS once the clustering is finished for the layer *i-1*. An object is regarded as being a combination of features, each of which is associated to its weight. In other words, an object regarded as a single entity in reality can be processed as a complex unit. For instance, the weather can be considered as a global concept but also as detailed data (rain, temperature, etc…). All of its features are parts of a same object and are strongly connected together. That realistic consideration implies several distinct clustering layers. The layer *i* concerns the classification of a sample set called *A* and the *i-1* one involves some of their constituent units. Obviously, a second layer of classification is consistent only in case at least one of the sample features is a complex entity. Therefore, for each sample of the set, an object feature becomes itself a whole sample in the layer *i-1* and is compared to the others to constitute a new sample set called *B*. Then a classification of the *B* samples is processed. Once the classification of the *B* samples has been done, its results are used to compute the classification of A. If the samples of the *A* set have *C* complex features, the second classification level implies *C* distinct sample sets B_1, B_2, … B_C thus *C* distinct classifications.

The MEMBAS algorithm [8, 9] can then calculate the weights of every feature for the classes definition. It is applied on the *B* samples so that its involved features become the weighted components of a meaningful object. The complex features of an *A* sample is then a balanced combination of attributes.

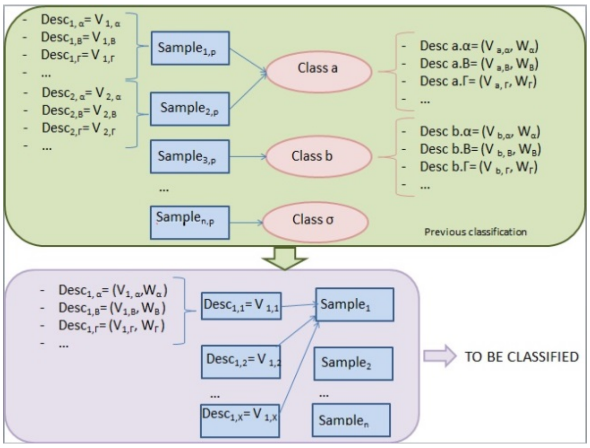

Figure 2: Principle of hierarchical classification

As explained in the Figure 2, the sample $Sample_1$ is described by X features, including the object-type feature $Desc_{1,1}$. $Desc_{1,1}$ is described by $Desc_{1,\alpha}$, $Desc_{1,\beta}$, etc.

To get their respective importance W_α, W_β etc in $Desc_{1,1}$ description, a previous classification is performed regarding $Desc_{1,1}$ as a sample ($Sample_{1,p}$), so that each weight can be calculated using the MEMBAS algorithm [8, 9]. Once the respective weights of each feature are known, objects are automatically instantiated to be involved in the main classification. $Desc_{1,1}$ is then described in line with the obtained weights Wα , Wβ and the known values V1,α , V1,β.

2.5 Evaluation of a classification quality

The comparison of two classifications can be performed by measuring their respective compactness and their separation. Better the classes are compact and separated easier will be the recognition process.

A method to measure the quality of a partition has been proposed by [10]. This index measures the quality partition in terms of classes compactness and separation. This partition index is the *Clusters Validity Index* (*CV*, Eq.(10)) which depends only on the GADs (membership degree of an individual to a class) and not explicitly on data values.

$$CV = \frac{Dis}{N}\cdot D^*_{\min}\cdot\sqrt{K} \qquad (10)$$

where N is the total number of individuals in the data base and K the total number of classes.
Dis represents the dispersion given by:

$$Dis = \sum_{k=1}^{K} 1-\frac{\sum_{j=1}^{J}\delta_{kj}\cdot\exp(\delta_{kj})}{N\cdot GAD_{Mk}\cdot\exp(GAD_{Mk})} \qquad (11)$$

with: $$\delta_{kj} = GAD_{Mk} - GAD_k^j \ \forall j,\ j\in[1,J] \qquad (12)$$

and $$GAD_{Mk} = \max\left[GAD_k^j\right] \qquad (13)$$

$D^*_{\min}$ is the minimum distance between two classes. This distance is computed by using the distance $d^*(A,B)$ between two fuzzy sets *A* and *B* [8] defined by:

$$d^*(A,B)=1-\frac{M[A\cap B]}{M[A\cup B]}=1-\frac{\sum_{j=1}^{J}\min(GAD_A^j,GAD_B^j)}{\sum_{j=1}^{J}\max(GAD_A^j,GAD_B^j)} \qquad (14)$$

The highest value of *CV* corresponds to a better partition.

4 Application to an agronomical project

The agronomical project aims at developing a diagnosis system for an optimized water management system and an efficient distinctive guidance for corn farmers in order to decrease the use of phytosanitary products and the water consumption for irrigation. The project involves two aspects. The first one aims at complementing the benefits of adopting and implementing the cultural profile techniques [28, 29]. In this context, we perform a classification of plots based on various agronomic and SAFRAN meteorological data [30], so that each plot should mostly belong to one particular class whose features are known. Thanks to the provided information stemmed from the classification results, advice can be offered to the corn farmers concerning the corn variety they should sow and the schedule they should follow for an optimized yield. This study includes two steps which are described in figure 3. The first one concerns the clustering of a training set of 50 plots, using the unsupervised LAMDA classification.

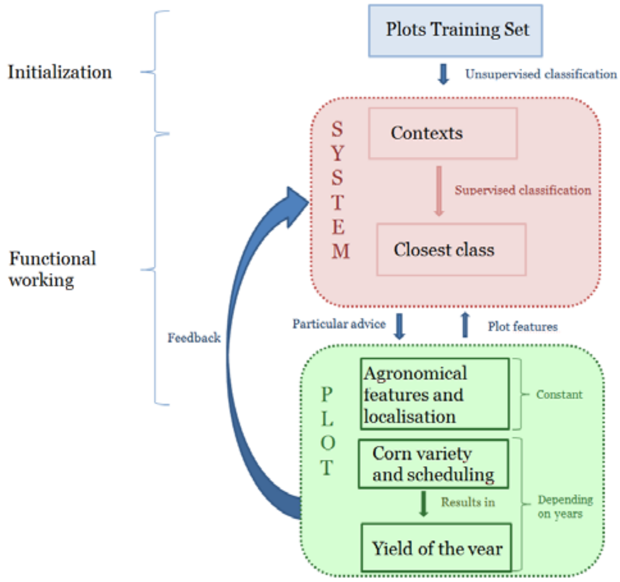

Figure 3: Learning System functioning

The data used for this classification are six distinctive agronomical descriptors, describing the plots' features and that are highly involved in their capacity for yield and water retention, and twenty-one weather features, defining the meteorological class in which the plot is situated. The second part of the project will be repeated annually to update and improve the clustering performed previously by adding new information returned by the farmers after harvest. In the following, only the first part is presented.

Firstly a previous meteorological clustering (A) is required to realize a realistic plot classification since the yield of seedling is highly related to the meteorological conditions. The weather is then regarded as a complex entity so that it is only one of a plot features. It is based on the historical meteorological data of the geographical position corresponding to the studied plot. Those descriptors refer to the temperature, the quantity of rainfall, and the evapotranspiration which occurred during three crucial periods of the year. Each feature is described in several distinctive ways. For instance, one period temperature is evaluated according three types of information. This meteorological clustering is an unsupervised classification based on weather data covering every single days of the determined periods during the fifty last years for all the geolocalized points belonging to the area studied in this project (South-West of France). In the event that the plot is part of the training set (studied area), the weather type of its area is known and the plot classification can be done directly. Otherwise, the weather type is obtained thanks to a supervised classification mode (B') delivering the most appropriate context. In any cases, the weather type is an object-feature. This hierarchical treatment permits to regard each meteorological type as a whole and let the weather contexts follow their natural evolution independently of agronomical variations. Moreover, considering the meteorological features as a single global object permits taking into account the environmental constraints and getting a realistic model. As we can observe in the Figure 4, the meteorological clustering (B) has permitted to divide the area in three sub-areas. The results of clustering (B) and the meteorological supervised classification (B’) have been first performed with every sample of the set and the distribution of the weights between the meteorological features has been determined.

The result of this classification is consistent and so, we can use the obtained classes and weights of the meteorological features (obtained with MEMBAS) as object-features in classification (A). To analyze the benefit of using hierarchical classification, a clustering (A') has been performed by using the twenty-one meteorological features separately and the agronomical features (twenty-seven features taken indistinctly). We can notice that the prototypes of the classes are highly dependent on the meteorological classes for clustering (A) while clustering (A') is mainly influenced by the ground type.

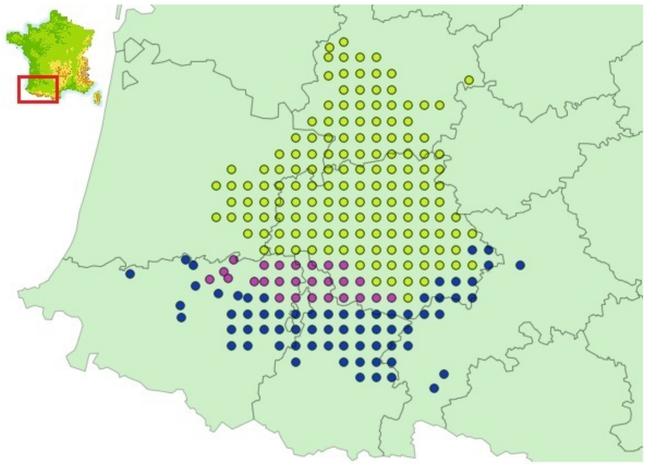

Figure 4: Meteorological sub-areas obtained with classification (B)

To enlighten this, we chose arbitrarily two very close classes containing the similar plots in both clustering. Each class prototype is described by the mean value of its marginal degree memberships (MAD). We represent in Figure 5 these prototype parameters for meteorological features only for both cases (A with diamond and A' with square) with in abscises, the marginal membership degree for class 1 and in ordinate the same marginal membership degree for class 2. For a better quantification of the benefits that the use of the object representation brings, the CV is systematically calculated in order to determine the better partition quality. The results are very encouraging since CV = 0.69 when the meteorological data are regarded as a whole object and 0.2 when they are treated separately. The object type representation enables to multiply by more than 3 this index and therefore the compactness of the obtained partition.

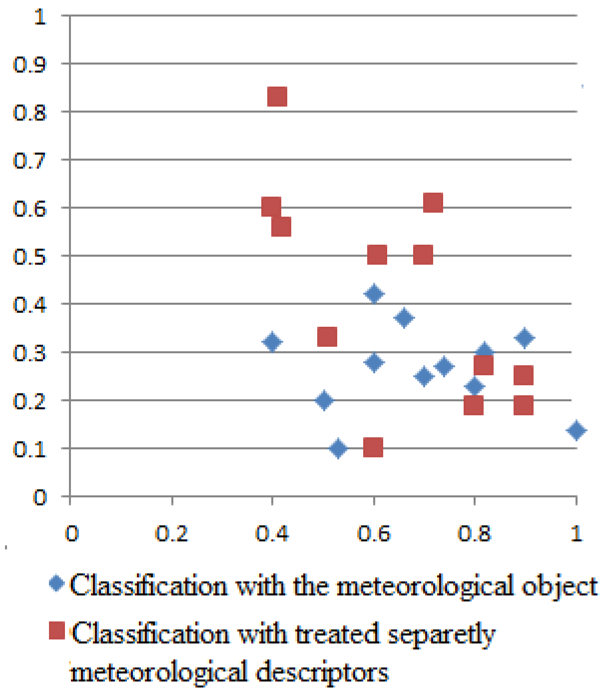

Figure 5: Meteorological prototypes for two close classes in case (A) and (A')

The second aspect of our implication in the project deals with the water utilization of various clusters of farmers with the aim of forecasting the needs of each cluster and adjusting the repartition. From this perspective, we realize an unsupervised classification of a training data-set of 2900 samples described by seven features: distance to the closest waterway, orientation, altitude… Orientation concerns cardinal points and we assume that it is not expressible with different modalities since continuity cannot be represented by qualitative descriptors. It cannot be a number nor an interval because of the cyclic form to be kept. Thus we choose to regard a cluster orientation as an object composed of two descriptors that correspond to the coordinates of its cardinal point in a trigonometric circle base. The orientation of each cluster can take eight different values: N, NE, E, SE, S, SW, W, and NW, which bring us to consider eight different combinations. In accordance with the trigonometrical circle, these eight combinations are respectively: (0,1), $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$, (1,0), $(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$, (0,-1), $(-\frac{\sqrt{2}}{2},-\frac{\sqrt{2}}{2})$, (1,0), $(\frac{\sqrt{2}}{2},-\frac{\sqrt{2}}{2})$.

Once our results are validated by an expert, the classification is experimented twice: firstly treating each descriptor separately and secondly involving the object type. Such as meteorological data in the first example, the CV is calculated in order to determine the better partition quality.

In this case, which implies 2900 samples, CV= 0.08 when abscissa and ordinate are separated, and CV= 0.13 when using an orientation object. As in the first example, these results show a qualitative gain for the partition when the object type is used to express the semantically connected data.

4 Conclusion

This modular architecture allows more flexibility and a more precise treatment of data. As we can notice with the previous agronomical classification, the object approach makes each module able to be managed independently of the others so that they can evolve autonomously, depending on their own specific features and contexts. The object representation permits to preserve multi-dimensionality and makes fusion of datasets easier. A better overview is offered since we can percept the variations of each module distinctively and the evolution of their influences.

As a perspective, an agent-oriented architecture, based on the multi-agents theory [31] will be developed so that each sample could be considered independently of the others. They would be so able to create classes acting simultaneously and comparing themselves to the others, so that the classes definition won't depend on the samples order in the file anymore but will directly result from the samples set definition. This orientation will assure that the classification result of our method is unique and stable for a given samples set. We aim at developing some methods to allow a semantic data processing also.

References

[1] D. Dubois and H. Prade editor. The three semantics of fuzzy sets, Fuzzy sets and systems, vol. 90,N° 2, pp141-150,Elsevier, London, 1997.

[2] L. Hedjazi, J. Aguilar-Martin, M.V. Le Lann and T. Kempowsky, Towards a unified principle for reasoning about heterogeneous data: a fuzzy logic framework, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 20,N°2,pp. 281-302, World Scientific, 2012

[3] R.S. Michalski and R.E. Stepp, automated construction of classifications: Conceptual clustering versus numerical taxonomy, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-5, no. 4 (1980), pp. 396-410.

[4] D.W. Aha, Tolerating noisy, irrelevant and novel attributes in instance based learning algorithms, *Int. Man-Machine Studies* 36 (1992), pp. 267-287.

[5] S. Cost, S. Salzberg, A weighted nearest neighbor algorithm for learning with symbolic features, *Machine learning* **(10)** (1993), pp.57-78.

[6] T. Mohri, T. Hidehiko, An optimal Weighting Criterion of case indexing for both numeric and symbolic attributes, in *D.W. Aha (Ed.)*, Case-based Reasoning: papers from the 1994 workshop. Menlo Park, CA:AIII Press, pp. 123-127.

[7] C. Giraud-Carrier, M. Tony, An Efficient Metric for heterogeneous Inductive Learning Applications in the Attribute-Value Language, *Intelligent systems* (1995) 341–350.

[8] K.C. Gowda, E. Diday, Symbolic clustering using a new similarity measure, *IEEE Trans. SMC* **22(2)** (1992) 368–378.

[9] Q.H Hu, Z.X. Xie, D.R. Yu, Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, *Pattern Recognition* **40** (2007) 3509–3521.

[10] F.A.T. De Carvalho, R.M.C.R. De Souza, Unsupervised Pattern Recognition Models for Mixed Feature-Type Symbolic Data, *Pattern Recognition Letters* **31** (2010) 430–443.

[11] I. Kononenko, Estimating Attributes: Analysis and Extensions of Relief, *Proc. European Conf. Mach. Learning ECML* (1994), pp. 171-182.
[12] K. Kira, L. Rendell, A practical approach to feature selection. *In proced. 9th Int'l Workshop on Machine Learning* (1992), pp. 249-256.
[13] M. Dash, H. Liu, Consistency-based search in feature selection, *Artif. Intell.* 151 (2003) 155–176.
[14] D.W. Aha, Incremental, instance-based learning of independent and graded concept descriptions, *in Proced. Of the 6th int'l Mach. Learning Workshop.* (1989) 387–391.
[15] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, V. Vapnik, Feature Selection for SVMs, *Advances in Neural Information Processing Systems* (2001), pp. 668-674.
[16] T. Cover, P. Hart, Nearest neighbor pattern classification, *IEEE Trans. Inf. Theory 13* (1967), pp.21-27.
[17] H. Liu, F. Hussian, C.L. TAM, M. Dash, Discretization: an enabling technique, *J. Data Mining and Knowledge* Discovery **6(4)** (2002) 393–423.
[18] M.A. Hall, Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning, *Int. Conf. Mach. Learning ICML* (2000), pp. 359-366.
[19] H.H. Bock, Diday E. Analysis of Symbolic Data, Exploratory methods for extracting statistical information from complex data. (Springer, Berlin Heidelberg,2000).
[20] V. Vapnik, The Nature of Statistical Learning Theory Data Mining and Knowledge Discovery ,pp. 1-47, 6, Springer Verlag, 1995
[21] T. Cover and P. Hart, Nearest neighbor pattern classification Information Theory, IEEE Transactions on,13,1,pp. 21-27, 1967
[22] Michie D., Spiegelhalter D.J., Taylor C.C., *Machine Learning, Neural and Statistical Classification,* Ellis Horwood series in Artificial Intelligence, february, 1994
[23] Rakotomalala R., Decision Trees, review MODULAD, 33, 2005.
[24] J. C. Aguado and J. Aguilar-Martin, A mixed qualitative-quantitative self-learning classification technique applied to diagnosis, The Thirteenth International Workshop on Qualitative Reasoning, (QR'99)pp. 124-128, 1999.
[25] L. Hedjazi, J. Aguilar- Martin, M.V. Le Lann, Similarity-margin based feature selection for symbolic interval data, Pattern Recognition Letters, Vol.32, N°4, pp. 578-585, 2012
[26] L. Hedjazi, J. Aguilar-Martin, M.V. Le Lann, and Tatiana Kempowsky-Hamon, Membership-Margin based Feature Selection for Mixed Type and High-dimensional Data: Theory and Applications, Information Sciences, *accepted to be published,* 2015.
[27] C. V. Isaza , H. O. Sarmiento, , T. Kempowsky-Hamon , M.V. Le Lann , Situation prediction based on fuzzy clustering for industrial complex processes, *Information Sciences*, Volume 279, 20 September 2014, pp. 785-804, 2014
[28] Henin S., Gras R., Monnier G., 1969, *Le profil cultural (2e edition),* Masson Ed. Paris.
[29] Gautronneau Y., Gigleux C., 2002, *Towards holistic approaches for soil diagnosis in organic orchards,* Proceedings of the 14th IFOAM Organic World Congress, Victoria, p 34.
[30] Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., ... & Morel, S. (2008). Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France. *Journal of applied meteorology and climatology*, *47*(1), 92-107.
[31] Ferber, J. (1999). *Multi-agent systems: an introduction to distributed artificial intelligence* (Vol. 1). Reading: Addison-Wesley.