Fuzzy Mechanisms for Unified Reasoning about Heterogeneous Data

L. Hedjazi[1,2], J. Aguilar-Martin [3], M-V. Le Lann [1,2] and T. Kempowsky[1,2]

[1]CNRS ; LAAS ;7, avenue du Colonel Roche ;F-31077 Toulouse, France
[2] Université de Toulouse ;UPS, INSA,INP, ISAE ; LAAS ;F-31077 Toulouse, France
[3]Universitat Politècnica de Catalunya, grup SAC; Rambla de Sant Nebridi, 10, E-08222 Catalunya, spain
{lhedjazi, aguilar, mvlelann, tkempows}@laas.fr

Abstract

Human knowledge about monitoring process variables is usually incomplete. To deal with this partial knowledge many types of representation other than the quantitative one are used to describe process variables (qualitative, symbolic interval). Thus, the development of automatic reasoning mechanisms about the process is faced with this problem of multiple data representations. In this paper, a unified principle for reasoning about heterogeneous data is introduced. This principle is based on a simultaneous mapping of data from initially heterogeneous spaces into only one homogeneous space based on a relative measure using appropriate characteristic functions. Once the heterogeneous data are represented in a unified space, a single processing for various analysis purposes can be performed using simple reasoning mechanisms. An application of this principle within a fuzzy logic framework is performed here to demonstrate its effectiveness. We show that simple fuzzy reasoning mechanisms can be used to reason in a unified way about heterogeneous data in two well known machine learning problems.

Introduction

Qualitative reasoning is taken in this paper as the mechanism of reasoning about qualitative information. Such information is given by qualitative valued data, which can be nominal or ordinal, mixed with quantitative and interval data. We address here one of the main difficulties encountered in qualitative reasoning: the diversity of types of information. The focus of the present paper is to propose a unified principle to establish various reasoning mechanisms using simultaneously three types of data: pure quantitative, symbolic interval and pure qualitative modalities. During the last decades, few research works have been directed to defy the issue of representation multiplicity for data analysis purposes (Michalski and Stepp 1980; Aha 1992; Cost and Salzberg 1993; Mohri and Hidehiko 1994; Giraud-Carrier and Tony 1995; Gowda and Diday 1992; Hu et *al.* 2007; De Carvalho and De Souza 2010; Kononenko 1994). 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 (Kira and Randell 1992) and on information or consistency measures for the later one (Dash and Liu 2003). Whereas in classification and clustering tasks, eventually only a Hamming distance is used to handle qualitative data (Aha 1992; Kononenko 1994; Aha 1989). 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 (Kira and Randell 1992; Weston *et al.* 2001; Cover and Hart 1967). 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 (Liu and Hussian 2002; Hall 2000). 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 (Bock and Diday 2000), they were focused generally more on the clustering tasks (Gowda and Diday 1992; De Carvalho and De Souza 2010) and no unified principle was given to handle simultaneously the three types of data for different analysis. In this paper a new general principle, introduced here as "Simultaneous Mapping for Single Processing (SMSP)", enables the reasoning in a unified way about heterogeneous data for several data analysis purposes. To illustrate the possible use of the proposed approach, two methodological applications are considered: dimensionality reduction and classification. This paper is organized as follows: in the second section the SMSP principle is introduced. In section 3, an example of simultaneous mapping of mixed variables into a common space is presented within the framework of fuzzy logic. Section 4 presents the use of this principle to establish a reasoning scheme for a variable selection task of heterogeneous data. Another reasoning mechanism using this principle, devoted to the classification of heterogeneous data is presented in section 5. In each of these three sections, an application example on real-world heterogeneous data sets from the UCI Repository database is given to exhibit the performance of the proposed method. The final section gives conclusions.

SMSP Principle

As the name indicates it, SMSP principle is based initially on an *appropriate* simultaneous mapping of heterogeneous data into a unified space. This mapping can be obtained by using a characteristic function for each type of data to bring them into a homogeneous space. These functions can be designed in such a way that they express a relative measure, as for example the measure of the appropriateness of each variable value of individuals to existing partitions, based on available prior knowledge. For instance, in a fuzzy sets theory framework, this measure is technically synonymous with the terms of membership measure which is a number of the real unit interval I.

In this paper, the SMSP principle is applied within a fuzzy set theory framework to reason about heterogeneous data. Once suitable membership functions that characterize the adequacy to each class are chosen according to a variable type, a fuzzy partition of variables can be performed based on empirical data. It results that the initial data is mapped into a homogeneous space isomorphic to an n^{th} dimensional unit cube I^n. Therefore, a unique and simple fuzzy reasoning mechanism can be used to reason about the resulting data whatever its original type. It will be shown thereafter that it enables to perform a wide variety of analysis (classification, dimensionality reduction, clustering…).

Homogeneous Space of Variables

Fuzzy partition of variables

Basically we consider the three above mentioned types of data: Quantitative variables, Symbolic intervals and Qualitative variables.

Definition 1. Fuzzification of variables

Let $D=\{x_n, C_k\}_{n=1}^{N} \in X \times C$ be a dataset, where $x_n=[x_{n1},x_{n2},...,x_{nm}]$ is the n^{th} individual (item) and N is the total number of individuals. Each individual is represented by *m* variables possibly of different types (quantitative, qualitative or symbolic interval), and C_k is the class label assigned to each individual in the pre-established partitions: $k=1,2,\ldots,l$.

Namely, let $\{ mff_1^i,\ldots, mff_l^i \}$ represent the *l* fuzzy sets that form a fuzzy partition for the i^{th} variable.

The membership function mff_k^i *is* defined in the i^{th} rank X_i of the i^{th} variable:

$$mff_k^i = \mu_k^i(x_i / i^{th} \text{ parameter of class } C_k); \; k=1,2,\ldots,l \qquad (1)$$

Where μ_k^i is the membership function of the i^{th} variable to the class C_k.

The membership function will be written using a parameter θ_{ki} representing the i^{th} prototype of class C_k as:

$$\mu_k^i(x_i) = f_i(x_i, \theta_{ki}) \qquad (2)$$

In the following, we define the particular membership functions that will be used here for the three different types.

Quantitative type variables. When the variable is quantitative, its numerical values are normalized within the interval $[x_{i\min}, x_{i\max}]$, where the bounds can be the extremes of a given dataset or independently imposed. This linear re-scaling of the variable into the interval [0,1] is performed by:

$$x_i = \frac{\hat{x}_i - \hat{x}_{i\min}}{\hat{x}_{i\max} - \hat{x}_{i\min}} \qquad (3)$$

Where $\hat{x}_i$ is the i^{th} variable measured value and x_i is its normalized one.

In this work the binomial membership function has been chosen:

$$\mu_k^i(x_i) = \varphi_k^{i\,1-x_i}\left(1-\varphi_k^i\right)^{x_i} \qquad (4)$$

where φ_k^i represents for each class *k* the mean value of the i^{th} variable of the m_k individuals belonging to this class :

$$\varphi_k^i = \frac{1}{m_k}\sum_{j=1}^{j=m_k} x_i^j \qquad (5)$$

Interval type variables. The membership function for interval type variables is chosen as the similarity between the symbolic interval value of the i^{th} variable x_i and the interval $\rho_k^i = \left[\rho_k^{i-}, \rho_k^{i+}\right]$ representing the class C_k as:

$$\mu_k^i(x_i) = S(x_i, \rho_k^i) \qquad (6)$$

Symbolic interval variables are extensions of pure real data types, in the way that each variable may take an interval of values instead of a single value (Gowda and Diday 1992). In this framework, the value of a quantity *x* is expressed as a closed interval $[x^-, x^+]$ whenever only an incomplete knowledge is available about it; representing the knowledge that $x^- \le x \le x^+$ (Kuipers 1994).

Let us consider the global *domain U*, continuous or discrete, in which the intervals are defined. *U* must be a compact subset of the real line R.

The *measure* ϖ of an interval *X* is given by its length: $\varpi[X] = upper.bound(X) - lower.bound(X)$.

Given 2 intervals $A=[a^+, a^-]$ and $B=[b^+, b^-]$, their *distance* ∂ is defined as:

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

Definition 2. Similarity measure between two intervals

The similarity measure is defined here as:

$$S(A,B)=\frac{1}{2}\left(\frac{\varpi\left[A\cap B\right]}{\varpi\left[A\cup B\right]}+1-\frac{\partial\left[A,B\right]}{\varpi\left[U\right]}\right) \qquad (7)$$

It is worthwhile to note that the function $S(A,B)$ fulfils the following properties :

(i) $0 \leq S(A,B) \leq 1$;

(ii) $S(A,B) = 1$ if and only if A equals to B ;

(iii) $S(A,B) = S(B,A)$.

These properties are commonly used to characterize a general similarity measure (Li and Wu 2008).

Lets consider that m_k individuals have been assigned to class C_k, this class will have as prototype a vector whose components are the intervals obtained by the mean bounds:

$$\rho_k^{i-} = \frac{1}{m_k}\sum_{j=1}^{m_k} x_i^{j-} \text{ , and } \rho_k^{i+} = \frac{1}{m_k}\sum_{j=1}^{m_k} x_i^{j+} \qquad (8)$$

Where x_i^{j-} is the i^{th} variable lower bound of the j^{th} sample and x_i^{j+} is its upper bound. Consequently, the resulted class prototype for the r interval variables is given by the vector of intervals:

$$\rho_k = \left[\rho_k^1, \rho_k^2, ..., \rho_k^r\right]^T \qquad (9)$$

Qualitative type variables. For qualitative variables, the possible values of the i^{th} variable form a set of modalities:

$$D_i = \left\{Q_1^i, \dots Q_j^i, \dots Q_{Mi}^i\right\} \qquad (10)$$

The membership function of qualitative variable x_i is specified as:

$$\mu_k^i\left(x_i\right) = \left(\Phi_{k1}^i\right)^{q_{i1}} * \dots * \left(\Phi_{kMi}^i\right)^{q_{iMi}} \qquad (11)$$

Where Φ_{kj}^i is the frequency of modality Q_j^i in the class C_k

and
$$q_j^i = \begin{cases} 1 & if \quad x_i = Q_j^i \\ 0 & if \quad x_i \neq Q_j^i \end{cases}$$

Therefore, the class prototypes are represented by $\Omega_k^i = \left[\Phi_{k1}^i, \dots, \Phi_{kj}^i, \dots, \Phi_{kMi}^i\right]$

Common membership space

Thus, having a quantitative space R^v , a qualitative space R^q and an interval space R^r, for each class the resulting membership space is R^m with $m= v+q+r$ which is the total number of variables. In case of dichotomy problems only one R^m space is necessary as the other can be obtained by complementary of membership.

Definition 3. Membership Degree Vector

A Membership Degree Vector (MDV) of dimension m, can be associated for a given individual x_n to each class as follows:

$$U_{nc_k} = \left[\mu_k^1\left(x_{n1}\right),\ \mu_k^2\left(x_{n2}\right), ...,\ \mu_k^m\left(x_{nm}\right)\right]^T ;\quad k = 1,2,...,l \qquad (12)$$

Where $\mu_k^i(x_{ni})$ (i.e. $\mu_k^i(x_i = x_{ni})$), is the membership function of class C_k evaluated at the given value x_{ni} of the i^{th} variable of individual x_n .

MDV is a m^{th} dimensional image of individual x_n with respect to the considered class. All the components of the MDV are positive numbers in the unit interval [0,1], therefore U_{nc_k} can be considered as a discrete fuzzy subset and the function $\psi(U_{nc_k}) = \sum_i \mu_k^i(x_{ni})$ represents its scalar cardinality (power or sigma count) as defined by (Zwick *et al* 1987).

SMSP for heterogeneous variable selection

Theory

Variable selection is defined as the problem of choosing a small subset of variables that ideally is necessary and sufficient to describe the target concept (Kira and Randell 1992). Generally, in classical **dimensionality reduction** methods, the variable relevance is estimated in the space assumed to be quantitative (Kira and Randell 1992; Weston *et al*. 2001). This restrictive assumption requires that other variable types must be transformed arbitrary, without taking any consideration about their original space.

The margin concept extensively studied in statistical learning theory (Vapnik 1998) is used here to perform a heterogeneous variable selection task. A Membership Margin is introduced in this section to accomplish heterogeneous variable selection task based on the SMSP principle.

Definition 4. Membership Margin (MM).

Let us consider class c , and its complement $\tilde{c}$. We assume that the n^{th} individual $x_n=[x_{n1},x_{n2},...,x_{nm}]$ is labeled by class c .

The membership margin for individual x_n is given by:

$$\beta_n = \psi\left(U_{nc}\right) - \psi\left(U_{n\tilde{c}}\right) \qquad (13)$$

Where U_{nc} and $U_{n\tilde{c}}$ are the MDVs of individual x_n respectively to classes c and $\tilde{c}$, computed with respect to all individuals contained in D excluding x_n ("leave-one-out margin") and $\psi(U_{nc})$ is the scalar cardinality of the discrete fuzzy set U_{nc}. Individual x_n is considered correctly classified if $\beta_n > 0$.

Definition 5. Weighted adequacy of an individual

Given a vector of positive weights $W_f = \left[w_{f1}, \dots w_{fm}\right] \in R^m$, the weighted adequacy of the n^{th} individual is defined by the cardinality of the new fuzzy set that takes into account the weight of each variable adequacy. It is given by the scalar product:

$$\Psi(U_{nc_k}/W_f) = W_f^T U_{nc_k} = \sum_i w_{fi}\, \mu_k^i(x_{ni}) \qquad (14)$$

These weights are non-negative numbers expressing the discriminative power of the fuzzy sets between existing classes. Therefore, a weighted membership margin for individual x_n can be defined as:

$$\beta_n = \psi(U_{nc}/W_f) - \psi(U_{n\tilde{c}}/W_f) \qquad (15)$$

The basic idea to determine the adequacy weights is to scale variable memberships in the membership space by minimizing the leave-one-out error.
This problem can be written as an optimization problem in the membership space:

$$\underset{w_f}{\text{Min}} \quad \sum_{n=1}^{N} I\left(\beta_n\left(w_f\right) < 0\right) \qquad (16)$$

Where $\beta_n(w_f)$ is the x_n^{th} margin computed with respect to w_f and I is an indicator function. To solve the above stated problem, an objective function has been used so that the averaged membership margin in the resulted weighted membership space is maximized:

$$\underset{w_f}{\text{Max}} \sum_{n=1}^{N} \beta_n(w_f) = \sum_{n=1}^{N} \{\sum_{i=1}^{m} w_{fi}\mu_c^i\left(x_{ni}\right) - \sum_{i=1}^{m} w_{fi}\mu_{\tilde{c}}^i\left(x_{ni}\right)\}$$
$$\text{s.t. } : \|w_f\|_2^2 = 1 \ , \ w_f \geq 0 \qquad (17)$$

The first constraint has been introduced to bound the weight vector w_f whereas the second one ensures its non negativity.

To solve the stated optimization problem, the well known Lagrangian optimization method can be used and yields an analytical solution, whose closed-form expression is given by:

$$w_f^* = \frac{v^+}{\left\|v^+\right\|} \qquad (18)$$

With $v^+ = [\max(v_1,0), \ldots, \max(v_m,0)]^T$

It is interesting to note that, by minimizing the objective function (16) within a membership margin framework, this mechanism avoids the combinatorial search and enables to reason about data regardless the number of variables.

Application example

Ljubljana Prognosis Dataset. The dataset used here concerns Ljubljana Prognosis dataset which deals with breast cancer prognosis; it contains a total of 286 patients where 201 among them have not relapsed after five years and 85 have relapsed (Murphy and Aha 1995). Patients with missing data were excluded from this study (9 patients). All patients are described by 9 variables (6 qualitative and 3 interval type):

- Menopause: >40, <40, pre-menopause.
- Ablation Ganglia: yes, no.
- Malignancy Degree (Grade): I, II, III
- Breast right, left
- Quadrant: sup. left, inf. left sup. right, inf. right, center.
- Irradiation: yes, no
- Age: 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99
- Tumor Size: 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59.
- Invaded Nodes: 0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, 24-26, 27-29, 30-32, 33-35, 36-39.

Experimental setup and results. The proposed reasoning tool in this section is used to find the set of important factors for this problem. Moreover, 50 random quantitative variables were added also to assess the robustness of this mechanism against irrelevant variables.

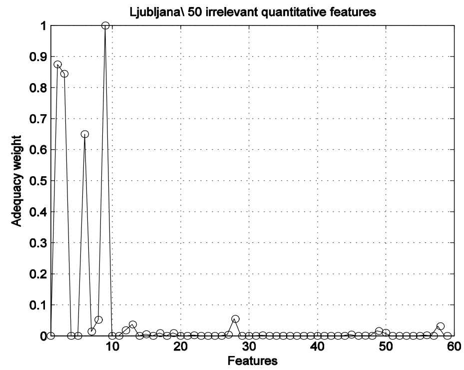

Fig. 1. Adequacy weights of variables

In figure 1, it can be observed that the order of the most relevant variables appears to be: Invaded Nodes (interval), Ablation ganglia (qualitative), Malignancy Degree (qualitative), Irradiation (qualitative). Moreover, the proposed mechanism succeeds to identify the 50 added irrelevant variables by assigning them approximately zero weights (they correspond to the last 50 variables in figure 1).

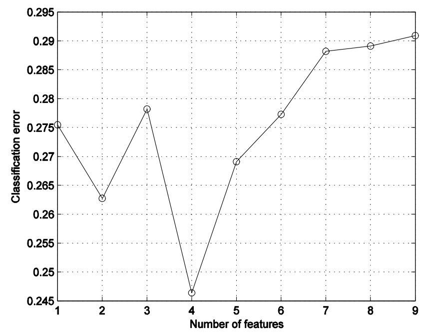

Fig. 2. Classification error as function of the top ranked variables

In order to validate the importance of the four chosen variables to improve the performance of the classification task, the fuzzy reasoning tool of classification proposed in the next section can be used. Moreover, to eliminate any statistical variation, this operation is repeated randomly 20 times and the accuracy of classification has been taken as the averaged error over all the runs. In figure 2 the obtained classification error as function of the top ranked variables has been plotted. It can be observed that the minimal classification error corresponds to the resulted four top ranked variables.

SMSP for classification of heterogeneous data

Theory

Classification is considered as one of the fundamental problems in machine learning. Duda and Hart (2000) define it as the problem of assigning a physical object or event to one of several pre-specified categories. In this section, we illustrate the problem of heterogeneous data classification as a reasoning problem in a common space based on the SMSP principle. Indeed, once the different types of variables have been mapped into a common space it is possible to establish a unified reasoning scheme for a classification purpose. The performance of the proposed approach is illustrated on a real-world problem. This approach is based on using the resulted fuzzy partitions in the previous section, to establish a fuzzy inference engine. In this approach, for each class C_k only a single fuzzy if-then rule is generated:

R_k: If x_1 is A_1 and x_2 is A_2...and x_m is A_m then class C_k

where the antecedent fuzzy sets A_i correspond to membership functions $\mu_k^i(x_i)$ for each class C_k defined previously according to the type of the variable x_i. It must be noticed here that the set of variables x_i used to evaluate each fuzzy if-then rule can possibly be of mixed types (quantitative, qualitative or interval-valued).

Then, the truth value of the consequent of each rule is determined by a fuzzy logic implication function which consists in a linear interpolation between a t-norm and a t-conorm. Finally, the individual is assigned to the class corresponding to the maximum membership obtained using the following fuzzy inference engine:

$$R^* = \arg\max_{R_k} \left\{ \alpha\gamma\left[\mu_k^1(x_1),\cdots,\mu_k^n(x_n)\right] + (1-\alpha)\beta\left[\mu_k^1(x_1),\cdots,\mu_k^n(x_n)\right] \middle| k = 1,\dots,l \right\}$$

where γ and β are dual fuzzy aggregation functions that combine memberships, given by the components of the MDV U_{nc_k}, of an individual $x_n=[x_{n1},x_{n2},\dots,x_{nm}]$ to a class C_k. The parameter α, called *exigency* allows to adjust the compensation between the union and the intersection operators which can be pre-specified by the user or estimated based on the training data.

Without the unification of the space of variables, this simple inference mechanism could not be applied, and the influence of the different types of variables would not be balanced.

Application example

Heart Dataset. The dataset used in the application of the proposed fuzzy reasoning mechanism for a classification purpose concerns Heart disease (Murphy and Aha 1995). It contains 270 observations described by 13 heterogeneous variables of patients diagnosed whether having a heart disease. The thirteen variables include 7 of quantitative type and 6 of qualitative type which are respectively:

- Age in years
- Resting blood pressure (in mm Hg on admission to the hospital)
- Serum cholesterol in mg/dl
- Maximum heart rate achieved
- Oldpeak = ST depression induced by exercise relative to rest
- The slope of the peak exercise ST segment
- Number of major vessels (0-3) colored by fluoroscopy
- Sex: male, female.
- Chest pain type : typical angina, atypical angina, non-angina pain, asymptomatic
- Fasting blood sugar > 120 mg/dl: yes, no.
- Resting electrocardiographic results: normal, having ST-T wave abnormality, showing probable or definite left ventricular hypertrophy.
- Exercise induced angina: yes, no.
- Thal: normal, fixed defect, reversible defect

Experimental setup and results. In order to validate the proposed methodology for classification on this heterogeneous dataset, cross-validation was performed. It consists of partitioning the dataset on two subsets: training and test. The training subset of patients is used to perform the fuzzy partition of variables and generate the rules used in the fuzzy engine inference, whereas the test subset is used to assess the performance of the resulted mechanism on unseen patients. The same procedure of cross-validation and statistical variation elimination as in section 4, is adopted here. To further demonstrate the effectiveness of the proposed methodology, a comparison on the same dataset with the well known k-Nearest Neighbor (Cover and Hart 1967) method is performed. It must be noticed here that this classification method is suitable for pure quantitative data whereas a random transformation of other data types is required. The obtained test error rate with both methods is shown in Table1. It can be observed that the proposed fuzzy reasoning mechanism enhances significantly the performance of the classification task.

Table1.Classification error for Heart dataset

Method	Training subset size	Test subset size	Test-Accuracy
Proposed approach	80%	20%	**82.13**
k-NN	80%	20%	69.9

Conclusion

In this paper a unified principle is introduced to cope with the problem of data heterogeneity. This principle is based on a simultaneous mapping of data from initially heterogeneous spaces into only one homogeneous space using appropriate characteristic functions. Once the heterogeneous data are represented in a unified space, only a single processing for various analysis purposes such as machine learning tasks can be performed. It has been shown that applying this principle within a fuzzy logic framework can significantly improve the mechanism performance to reason about heterogeneous data. Firstly, a simultaneous mapping is performed based on an appropriateness measure of variables values to each class using suitable membership functions according to each type of variable (quantitative, qualitative, interval). Then, simple fuzzy reasoning mechanisms were proposed to deal, in unified way, with heterogeneous data either for classification or variable selection tasks. For each application task, a validation on real-world problem was performed using heterogeneous datasets from the UCI repository. The proposed methodology leads to meaningful results and improves significantly tasks performance.

References

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

Aha, D.W. 1992. Tolerating noisy, irrelevant and novel attributes in instance based learning algorithms. *Int. Man-Machine Studies* 36: 267-287.

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

Mohri, T., and Hidehiko, T. 1994. 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, 123-127.

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

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

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

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

Kononenko, I. 1994. Estimating Attributes: Analysis and Extensions of Relief. *Proc. European Conf. Mach. Learning ECML*, 171-182.

Kira, K., and Rendell, L. 1992. A practical approach to feature selection. *In proced. 9th Int'l Workshop on Machine Learning*, 249-256.

Dash, M., and Liu, H. 2003. Consistency-based search in feature selection. *Artif. Intell.* 151: 155–176.

Aha, D.W. 1989. Incremental, instance-based learning of independent and graded concept descriptions. *In Proced. Of the 6th int'l Mach. Learning Workshop.* 387–391.

Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., and Vapnik, V. 2001. Feature Selection for SVMs. *Advances in Neural Information Processing Systems*, 668-674.

Cover, T., and Hart, P. 1967. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13: 21-27.

Liu, H., Hussian, F., TAM, C.L., and Dash, M. 2002. Discretization: an enabling technique. *J. Data Mining and Knowledge Discovery* 6(4): 393–423.

Hall, M.A. 2000. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. *Int. Conf. Mach. Learning ICML*, 359-366.

Bock, H.H., and Diday E. 2000. *Analysis of Symbolic Data, Exploratory methods for extracting statistical information from complex data*. Springer, Berlin Heidelberg.

Atkeson, C.G., Moore, A.W., and Schaal, S. 1997. Locally Weighted Learning. *Artificial Intelligence Rev.,* 11(15): 11–73.

Kuipers, B. 1994. *Qualitative Reasoning: Modeling and simulation with Incomplete Knowledge*. The MIT Press, Cambridge, Massachusetts, London.

Li, Y., and Wu, Z-F. 2008. Fuzzy feature selection based on min-max learning rule and extension matrix. *Pattern Recognition* 41: 217-226.

Zwick, R., Carlstein, E., and Budescu, D.V. 1987. Measures of similarity among fuzzy concepts: A comparative analysis. *Int'l J. Approx. Reason.* 1(2): 221-242.

Duda, R.O. Petter, E.H. and David, G.S. 2000. *Pattern Classification*, 2nd edition: John Wiley & Sons.

Murphy, P., and Aha, D. 1995. *UCI repository of machine learning databases, Dept. Information and Computer Science*. University of California. Irvine(http:// www.ics.uci.edu/~mlearn/ MLRepository.html).

Vapnik, V.N. 1998. *Statistical Learning Theory*. John Wiley & Sons.