34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Dimensionality Reduction Based on Fuzzy Rough Sets Oriented to Ischemia Detection

Diana A. Orrego[1], Miguel A. Becerra[2] and Edilson Delgado-Trejos[3]

Abstract— **This paper presents a dimensionality reduction study based on fuzzy rough sets with the aim of increasing the discriminant capability of the representation of normal ECG beats and those that contain ischemic events. A novel procedure is proposed to obtain the fuzzy equivalence classes based on entropy and neighborhood techniques and a modification of the Quick Reduct Algorithm is used to select the relevant features from a large feature space by a dependency function. The tests were carried out on a feature space made up by 840 wavelet features extracted from 900 ECG normal beats and 900 ECG beats with evidence of ischemia. Results of around 99% classification accuracy are obtained. This methodology provides a reduced feature space with low complexity and high representation capability. Additionally, the discriminant strength of entropy in terms of representing ischemic disorders from time-frequency information in ECG signals is highlighted.**

I. INTRODUCTION

Information in electrocardiographic (ECG) signals relating to the discrimination of different functional states, e.g., ischemic events, is usually represented by a large dimensional space, which hinders a proper interpretation of the embedded symbolic physiology in the feature space [1]. Studies on cardiac disease detection show that detecting ischemic episodes accurately in ECG recordings requires maximizing the performance in the classification stage for cases where samples lie near class separation boundaries in the state space [2]. In this regard, a dimensionality reduction must be applied with the aim of obtaining a low-dimensional and compact representation, preserving most of the relevant information of the original data, and thus, maximizing the representation capability as well as minimizing computational cost [3], [4]. Computer-based techniques can be grouped depending on the computational paradigm on which they are based, e.g., rule-based expert systems, artificial neural networks, pattern recognition, among others [5]. Recently, simultaneous feature selection and classification approaches based on sparsity-based arguments (i.e., sparse logistic regression) have been proposed [6], [7], and have yielded notable performance in bioinformatics applications [8]. However, when the objective function is continuously differentiable, and the problem domain set is closed and convex, the technique efficiency strongly depends on the estimation of the step size at each of the optimization iterations. Additionally, another critical issue is present when many real-world classification problems involve data from multiple classes, since this adds noticeable computational cost at the moment of solving large-scale sparse multinomial logistic regressions.

Rough sets (RS) are widely used in feature subset selection and attribute reduction [9]. In most of the existing algorithms, the dependency function is employed to evaluate the quality of a feature subset [10]. Compared with dependency, consistency can reflect not only the size of the decision positive region, but also the sample distribution in the boundary region. Therefore, the consistency measure is able to describe the distinguishing power of an attribute set more finely than the dependency function. However, the consistency value increases or is maintained when a new attribute is added to the attribute set. Moreover, some attributes are introduced into the reduct just for distinguishing a few samples. If we keep these attributes in the final result, the attributes may overfit the data [11]. In [12], a novel interpretation of Yager's entropy was proposed to compute the information of fuzzy indiscernibility relation. Additionally, conditional entropy and relative conditional entropy were proposed to measure the information increment, which can be interpreted as the significance of an attribute in the fuzzy rough set model. So, the independence of an attribute set, reduct, relative reduct in the fuzzy rough set model were redefined. In [13], Shannon's entropy is introduced to measure the information quantity implied in a Pawlak's approximation space. Based on the modified formulas, some generalizations of the entropy are proposed in order to calculate the information in a fuzzy approximation space and a fuzzy probabilistic approximation space, respectively. This inclusion of probability into an approximation space may lead to a tool for randomness, incompleteness, inconsistency and vagueness in real-world applications. Nevertheless, the general problem of these approaches is extracting the most stable feature subset by randomly sampling the data patterns.

In this paper a novel approach of dimensionality reduction is proposed in order to improve the discriminant capability of a representation space where the cardiac dynamics of the ischemic disorders are embedded, taking into account that this kind of physiological disorder has a behaviour that is highly variable within classes and has a low variability between classes. Thus, the goal is to determine a compact representation space with high discriminant power and low

This work was supported by the Research Center of the Instituto Tecnologico Metropolitano of Medellin, Colombia, under grant PM11112.

[1]Diana A. Orrego is with SINERGIA Research Group of the Instituto Tecnologico Metropolitano ITM, Calle 73 No. 76A-354, Medellin, Colombia. dianaorrego@itm.edu.co

[2]Miguel A. Becerra is with GEA Research Group of the Institucion Universitaria Salazar Herrera, Carrera 70 No. 52-49, Medellin, Colombia. migb2b@gmail.com

[3]Edilson Delgado-Trejos is with MIRP Lab of the Research Center of the Instituto Tecnologico Metropolitano ITM, Calle 73 No. 76A-354, Medellin, Colombia. edilsondt@gmail.com

complexity using procedures based on fuzzy rough sets, which must be validated using a simple classifier (e.g., K-NN), such that the performance is not a result of the extraordinary classifier properties but of the representation space quality.

II. MATERIALS AND METHODS

A. Database and representation space

The European $ST-T$ database is intended to be used for evaluating algorithms by analyzing ST and T-wave changes. This database consists of 90 annotated excerpts of ambulatory ECG recordings from 79 subjects. The subjects were 70 men aged 30 to 84, and 9 women aged 55 to 71. Myocardial ischemia was diagnosed or suspected in each subject. Each record is two hours in duration and contains two signals, each sampled at 250 samples per second with 12-bit resolution over a nominal 20 mV input range [14]. For this study, from V4-leads, 1800 representative beats were chosen by a specialist in cardiology: 900 considered normal and 900 beats with evidence of ischemia.

The representation space was generated using wavelet analysis by collecting coefficients corresponding to different families (as shown in Table I) in two ways: *i)* coefficients of the third approximation level and *ii)* decomposition up to level 6 and the 4 coefficients with maximum absolute value in each approximation scale. This is in order to verify the performance of the proposed algorithm in identifying redundant variables, since variation in the representation offered by the different wavelets should not be noticeable. The total extracted features was 840. The observations were separated in two groups of equal size for calibration and validation (each matrix has dimensions 900×840), using a method of random selection.

TABLE I

SPECTRAL REPRESENTATION USING WT

WT	N-order
Daubechies	2-10
Symlet	2-10
Biorthogonal	13, 15, 24, 26, 28, 55, 68
Reverse Biorthogonal	22, 24, 26, 28, 44, 55, 68

B. Dimensionality reduction with fuzzy-rough sets

Fuzzy-rough sets encapsulate the related but distinct concepts of vagueness (for fuzzy sets [15]) and indiscernibility (for rough sets [9]), both of which occur as a result of uncertainty in knowledge [16]. Fuzzy-rough feature selection (FRFS) provides a means by which discrete or real-valued noisy data (or a mixture of both) can be effectively reduced without the need for user-supplied information. Additionally this technique can be applied to data with continuous or nominal decision attributes, and as such can be applied to regression as well as classification datasets [17].

Let $\mathbb{U}$ be a non-empty set of finite objects (the universe of discourse) where $x, y \in \mathbb{U}$ and $\mathbb{A}$ is a non-empty finite set of features where a is a feature in $\mathbb{A}$, $P \subseteq \mathbb{A}$ and $\mathbb{D}$ is a set of decision features. The fuzzy lower and upper approximations can be defined using a Υ-transitive fuzzy similarity relation to approximate a fuzzy equivalence class X [18]:

$$\mu_{\underline{R_P}X}(x) = \inf_y \Psi\left(\mu_{R_P}(x,y), \mu_X(y)\right) \tag{1}$$

$$\mu_{\overline{R_P}X}(x) = \sup_y \Upsilon\left(\mu_{R_P}(x,y), \mu_X(y)\right) \tag{2}$$

where Ψ is a fuzzy implication and Υ a t-norm. R_P is the fuzzy similarity relation induced by the subset of features P:

$$\mu_{R_P}(x,y) = \Upsilon_{a \in P}\left\{\mu_{R_a}(x,y)\right\} \tag{3}$$

μ_{R_a} is the degree to which objects x and y are similar for feature a. The crisp positive region in traditional rough set theory is defined as the union of the lower approximations. By the extension principle [19], the membership of an object $x \in \mathbb{U}$, belonging to the fuzzy positive region can be defined by [20]

$$\mu_{POS_P(\mathbb{D})}(x) = \sup_{X \in \mathbb{U}/\mathbb{D}} \mu_{\underline{R_P}X}(x) \tag{4}$$

An important issue in data analysis is discovering dependencies between attributes. The fuzzy-rough degree of dependency of $\mathbb{D}$ on the attribute subset P can be defined

$$\gamma'_P(\mathbb{D}) = \frac{\sum_x \mu_{POS_P(\mathbb{D})}(x)}{|\mathbb{U}|} \tag{5}$$

A fuzzy-rough reduct R can be defined as a minimal subset of features of the initial attribute set $\mathbb{C}$ such that for a given set of attributes $\mathbb{D}$ preserves the dependency degree of the entire dataset, i.e. $\gamma'_R(\mathbb{D}) = \gamma'_{\mathbb{C}}(\mathbb{D})$ and $\gamma'_{R-\{a\}}(\mathbb{D}) \neq \gamma'_R(\mathbb{D})$ for all $a \in R$ (See Algorithm 1).

Algorithm 1 Quick-Reduct Algorithm [17]

```
Require: ℂ, the set of all conditional features; 𝔻, the set of decision
  features.
  R ← {}; γ'_best = 0; γ'_prev ≠ 0
  while γ'_best ≠ γ'_prev do
    T ← R
    γ'_best = γ'_prev
    for all x ∈ (ℂ − R) do
      if γ'_{R∪{x}}(𝔻) > γ'_T(𝔻) then
        T ← R ∪ {x}
        γ'_best = γ'_T(𝔻)
      end if
    end for
    R ← T
  end while
  return R
Output: R, minimal feature subset
```

C. Proposed procedure

With the aim of obtaining a minimal feature subset, i.e., reduct (R), first of all, the two-class feature matrix with dimensions 1800×840 must be normalized with values in $[0,1]$. By heuristic routines, the parameters associated to the neighbor distance tolerance and the inclusion rate are adjusted. In concordance to Fig. 1, two procedures were developed to analyze the performance for each case in terms of indiscernibility and data overlapping. The former

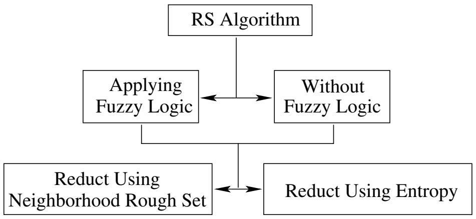

Fig. 1. Rough Set –RS Algorithm with and without fuzzy logic

was carried out including fuzzy logic in the RS algorithm (FRS Algorithm) to obtain the equivalence relation matrix [13], thus, the reduced feature space was achieved by using Algorithm 1 in two different ways according to the evaluation measure: neighborhood distance based on forward greedy strategy [21] and entropy estimation [12]. The latter follows the same procedure without including fuzzy logic while the equivalence relation matrix is obtained by the classic way. The validation tests were developed using a Bayes-based classifier and a K-NN classifier ($K = 5$) with cross-validation (CV) strategy 50-50 and 70-30 for several partitions (30 folds), in order to verify the result consistency and learning capability of the reduced feature space.

III. RESULTS AND DISCUSSION

Table II shows the dimensionality reduction results of the different techniques implemented in this study. The notation RS/FRS–N and RS/FRS–E means Rough Set/Fuzzy Rough Set Algorithm using neighborhood distance and entropy estimation, respectively. Likewise, Table III presents the classification results using the reduced feature space obtained with the different techniques, in concordance to Table II. The best result is obtained by the fuzzy rough set algorithm using the entropy estimation as reduction criterium, a neighbor distance tolerance adjusted to 0.05 and an inclusion rate fixed to 0.5. By comparing the RS–E performance with FRS–E, the differences in the classification consistency are notable, since the spread of the accuracy results is much lower when fuzzy measures are included to obtain the equivalence relation matrix. On the other hand, the robustness of the FRS–E technique is proved when the difference in the classification accuracy corresponding to the cross-validation 70-30 and 50-50 is not significant, which means that the reduced feature space offers high learning capability.

TABLE II

DIMENSIONALITY REDUCTION (FROM 840 FEATURES)

Technique	RS–N	RS–E	FRS–N	FRS–E
Feature number	6	10	4	11

TABLE III

ACCURACY FOR K-NN/BAYES CLASSIFICATION (%)

CV	RS–N	RS–E	FRS–N	FRS–E
70-30	80±3/82±4	95±4/94±5	71±9/71±7	99±0.5/98.5±1
50-50	80±7/81±6	82±4/83±5	74±7/73±9	99±0.5/98.5±1

IV. CONCLUSIONS

A dimensionality reduction approach based on fuzzy rough sets was developed and tested on a high-dimension feature space derived from ECG beats where the cardiac dynamics due to ischemic disorders are embedded. This methodology made it possible to find a reduced feature space with low complexity and high representation capability, which is related to a high learning capability. Despite this cardiac disorder being highly variable within classes and having a low variability between classes, the results showed a high discriminant power according to the classification accuracy rates and a stable consistency level which means reliability in the detection. In this sense, the algorithm performance with and without the inclusion of fuzzy routines in feature selection was compared and the improvement in the classification results achieved by the FRS technique is due to the capability of the fuzzy measures to integrate a relevance concept in a space made up by highly variable intra-class patterns. However, the adjustment of the algorithm parameters, such as the neighbor distance tolerance and the inclusion rate, was not optimized by a criterium function. In this study, the algorithm parameters were obtained using heuristic rules according to the data nature. Thus, the automatic tuning of the algorithm parameters is proposed as future work.

In a clinical decision support system, the two most important issues are accuracy and reliability. In this sense, accuracy is more important than dimensionality. The challenge is to reduce the feature number as much as possible, without affecting accuracy and consistency. From the results it can be inferred that although the gain from FRS-E to RS-E is only one less feature, the representation quality is very different in terms of consistency, since the classification results have a much lower standard deviation in the case of FRS-E, which means that the inclusion of fuzzy logic contributes robustness to the feature selection procedure.

Finally, the results show the discriminant strength that entropy has in terms of representing ischemic disorders from time-frequency information in ECG signals. This could be associated to the complex nature of cardiac rhythms, which can be captured by entropy when any disturbance is presented. In this sense, the geometric analysis of the data based on neighborhood distances is not discriminant in terms of ischemia recognition because these kinds of measurements in the representation space are very sensitive to changes that can be caused by noise or other types of artifacts.

ACKNOWLEDGMENT

We would like to thank the Research Center for its support to the “Semilleros de Investigacion” and the Biomedical Engineering Program of the Instituto Tecnologico Metropolitano –ITM of Medellin, Colombia.

REFERENCES

[1] E. Delgado-Trejos, A. Perera-Lluna, P. Caminal-Magrans, M. Vallverd-Ferrer, and G. Castellanos-Domnguez, “Dimensionality reduction oriented toward the feature visualization for ischemia detection,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 4, pp. 590–598, July 2009.

[2] S. Papadimitriou, S. Mavroudi, L. Vladutu, and A. Bezerianos, "Ischemia detection with a self-organizing map supplemented by supervised learning," *IEEE Transactions on Neural Networks*, vol. 12, no. 3, pp. 503–515, 2001.

[3] E. Delgado-Trejos, G. Castellanos-Domnguez, L. G. Snchez, and J. F. Surez, *Encyclopedia of healthcare information systems*. New York, USA: IGI Global, 2008, vol. II, ch. Feature Selection in Pathology Detection using Hybrid Multidimensional Analysis, pp. 587–593.

[4] G. Castellanos-Domnguez, E. Delgado-Trejos, and J. L. Rodriguez, *Long-Term Biosignal Processing for Cardiac Arrhythmia Detection: An Unsupervised Approach*. Manizales Colombia: Publicaciones Universidad Nacional de Colombia, 2012.

[5] C. Papaloukas, D. I. Fotiadis, A. Likas, A. P. Liavas, and L. K. Michalis, "A knowledge-based technique for automated detection of ischemic episodes in long duration electrocardiograms," *Medical and Biological Engineering and Computing*, vol. 39, p. 105–112, 2001.

[6] J. Liu, J. Chen, and J. Ye, "Large-scale sparse logistic regression," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD'09)*, Paris, France, July 2009, pp. 547–556.

[7] S. Ryali, K. Supekar, D. A. Abrams, and V. Menon, "Sparse logistic regression for whole-brain classification of fMRI data," *NeuroImage*, vol. 51, no. 2, pp. 752–764, 2010.

[8] S. K. Shevade and S. S. Keerthi, "A simple and efficient algorithm for gene selection using sparse logistic regression," *Bioinformatics*, vol. 19, no. 17, p. 2246–2253, 2003.

[9] R. W. Swiniarski and A. Skowron, "Rough set methods in feature selection and recognition," *Pattern Recognition Letters*, vol. 24, no. 6, pp. 833–849, March 2003.

[10] R. B. Bhatt and M. Gopal, "On fuzzy-rough sets approach to feature selection," *Pattern Recognition Letters*, vol. 26, no. 7, pp. 965–975, May 2005.

[11] Q.-H. Hu, H. Zhao, Z.-X. Xie, and D.-R. Yu, *Lecture Notes in Computer Science*, ser. Advances in Knowledge Discovery and Data Mining. Germany: Springer-Verlag Berlin Heidelberg, 2007, vol. 4426, ch. Consistency Based Attribute Reduction, pp. 96–107.

[12] Q.-H. Hu and D.-R. Yu, "Entropies of fuzzy indiscrenibility relation and its operations," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 12, no. 5, pp. 575–589, October 2005.

[13] Q.-H. Hu, D.-R. Yu, Z.-X. Xie, and J.-F. Liu, "Fuzzy probabilistic approximation spaces and their information measures," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 2, pp. 191–201, April 2006.

[14] A. Taddei, G. Distante, M. Emdin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi, "The european *ST-T* database: standard for evaluating systems for the analysis of *ST-T* changes in ambulatory electrocardiography," *European Heart Journal*, vol. 13, pp. 1164–1172, 1992.

[15] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[16] D. Dubois and H. Prade, *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1992, ch. Putting rough sets and fuzzy sets together, pp. 203–232.

[17] R. Jensen and Q. Shen, *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*, ser. IEEE Press Series in Computational Intelligence. Hoboken, NJ: John Wiley & Sons, 2008.

[18] A. M. Radzikowska and E. E. Kerre, "A comparative study of fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 126, no. 2, pp. 137–155, March 2002.

[19] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning - part I," *Information Sciences*, vol. 8, pp. 199–249, 1975.

[20] N. Mac-Parthalin and R. Jensen, "Measures for unsupervised fuzzy-rough feature selection," in *Ninth International Conference on Intelligent Systems Design and Applications (ISDA'09)*, Pisa, Italy, November-December 2009, pp. 560–565.

[21] Q.-H. Hu, Z.-X. Xie, and D.-R. Yu, "Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation," *Pattern Recognition*, vol. 40, no. 12, p. 3509–3521, December 2007.