Supporting Fuzzy-Rough Sets in the Dendritic Cell Algorithm Data Pre-processing Phase

Zeineb Chelly and Zied Elouedi

LARODEC, Institut Supérieur de Gestion de Tunis, Tunisia
zeinebchelly@yahoo.fr, zied.elouedi@gmx.fr

Abstract. The Dendritic Cell Algorithm (DCA) is an immune algorithm based on the behavior of dendritic cells. The DCA performance relies on its data pre-processing phase which includes two sub-steps; feature selection and signal categorization. For an automatic data pre-processing task, DCA applied Rough Set Theory (RST). Nevertheless, the developed rough approach presents an information loss as data should be discretized beforehand. Thus, the aim of this paper is to develop a new DCA feature selection and signal categorization method based on Fuzzy Rough Set Theory (FRST) which allows dealing with real-valued data with no data quantization beforehand. Results show that applying FRST, instead of RST, is more convenient for the DCA data pre-processing phase yielding much better performance in terms of accuracy.

Keywords: Evolutionary Algorithm, Fuzzy Rough Set Theory, Feature Selection.

1 Introduction

One of the emerging algorithms within the set of evolutionary computing algorithms is the Dendritic Cell Algorithm (DCA) [1]. DCA is inspired by the function of the natural dendritic cells. DCA has the ability to combine a series of informative signals with a sequence of repeating abstract identifiers, termed "antigens", to perform anomaly detection. To achieve this and through the pre-processing phase, DCA selects a subset of features and assigns each selected feature to a specific signal category; either as "Danger Signal" (DS), "Safe Signal" (SS) or as "Pathogen-Associated Molecular Pattern"(PAMP). The resulting correlation signal values are then classified to form an anomaly detection style of two-class classification. Technically and for an automatic DCA data pre-processing task, Rough Set Theory (RST) [2] was introduced leading to the development of various rough-DCA algorithms; namely RST-DCA [3], RC-DCA [4] and QR-DCA [5]. Based on the RST concepts and to perform feature selection, the rough DCA algorithms keep only a set of the most informative features, a subset termed *reduct*, that preserve nearly the same classification power of the original dataset; and to assign each selected feature to its specific signal category, the algorithms are based on the RST *reduct* and *core* concepts. The main difference between the algorithms is that RST-DCA assigns the same attribute

M. Lee et al. (Eds.): ICONIP 2013, Part II, LNCS 8227, pp. 164–171, 2013.

to both SS and PAMP signals. However, RC-DCA and QR-DCA assign different attributes for SS and PAMP. Another main difference, is that QR-DCA looks for a trade-off between generating good classification results and preserving the lightweight of the DCA algorithm. More details about the algorithms can be found in [3][4][5]. Nevertheless, in all these crisp rough algorithms, the use of RST as a pre-processor technique is reliant upon crisp datasets; the feature values of the input dataset have to be discretized beforehand. Consequently, important information may be lost as a result of quantization [6]. This information loss may influence the rough-DCA algorithms, RST-DCA, RC-DCA, QR-DCA, feature selection process by generating an incorrect set of selected attributes; as a consequence, this will misguide the algorithms categorization phase by categorizing the features to erroneous signal categories. As a result, this will influence the algorithms classification process by generating unreliable classification results. To overcome the RST applicability restriction, Fuzzy Rough Set Theory (FRST) was introduced in [7] as it provides the means of data reduction for crisp and real-value attributed datasets which utilizes the extent to which values are similar. FRST encapsulates the related but distinct concepts of vagueness (for fuzzy sets) and indiscernibility (for rough sets), both of which occur as a result of uncertainty in data; a method employing fuzzy-rough sets can handle this uncertainty. Therefore, in this paper, we propose to develop a novel DCA version based on a new feature selection and signal categorization technique. Specifically, our new model, named FLA-DCA, is based on the framework of fuzzy rough set theory for data pre-processing and more precisely on the use of the fuzzy lower approximation (FLA); to guarantee a more rigorous data pre-processing phase. The main contributions of this paper are to introduce the concept of FRST in the DCA data pre-processing phase and to show how FRST can be applied to search for the convenient features to retain and how it can be appropriate for the categorization of each selected feature to its right type of signal. This will be achieved by avoiding the information loss already discussed and by keeping the attribute values unchanged with no need for a quantization process beforehand.

2 The Dendritic Cell Algorithm

DCA is a population based system, with each agent in the system is represented as a cell. Each cell has the capacity to collect data items, termed *antigens*. Formally, the DCA initial step is the automatic data pre-processing phase where feature selection and signal categorization are achieved. More precisely, DCA selects the most important features, from the initial input database, and assigns each selected attribute to its specific signal category (SS, DS or PAMP). Once data pre-processing is achieved and after calculating the values of the safe, PAMP and DS signals [1], DCA adheres these three signal categories and antigen to fix the context of each object (DC) which is the step of *Signal Processing*. In fact, the algorithm processes its input signals (already pre-categorized) in order to get three output signals: costimulation signal (Csm), semi-mature signal (Semi) and mature signal (Mat) [1]. A migration threshold is incorporated into the DCA in order to determine the lifespan of a DC. As soon as the Csm exceeds the migration threshold; the DC ceases to sample signals and antigens. The migration state of a DC to the semi-mature state or to the mature state is determined by the comparison between cumulative $Semi$ and cumulative Mat. If the cumulative $Semi$ is greater than the cumulative Mat, then the DC goes to the semi-mature context, which implies that the antigen data was collected under normal conditions. Otherwise, the DC goes to the mature context, signifying a potentially anomalous data item. This step is known to be the *Context Assessment* phase. The nature of the response is determined by measuring the number of DCs that are fully mature and is represented by the Mature Context Antigen Value (MCAV). $MCAV$ is applied in the DCA final step which is the *Classification* procedure and used to assess the degree of anomaly of a given antigen. The closer the $MCAV$ is to 1, the greater the probability that the antigen is anomalous. By applying thresholds at various levels, analysis can be performed to assess the anomaly detection capabilities of the algorithm. Those antigens whose $MCAV$ are greater than the anomalous threshold, which can be automatically generated from the input data, are classified as anomalous while the others are classified as normal.

3 Fuzzy-Rough Sets for Feature Selection

1) Basic Concepts: In the same way that crisp equivalence classes are central to rough sets [2], *fuzzy equivalence classes* are central to the fuzzy-rough set approach [7]. For typical applications, this means that the decision values and the conditional values may all be fuzzy. The concept of crisp equivalence classes can be extended by the inclusion of a fuzzy similarity relation S on the universe, which determines the extent to which two elements are similar in S. The fuzzy lower and fuzzy upper approximations become $\mu_{\underline{R_P}X}(x) = inf_{y \in U} I(\mu_{R_P}(x,y), \mu_X(y))$ and $\mu_{\overline{R_P}X}(x) = sup_{y \in U} T(\mu_{R_P}(x,y), \mu_X(y))$. In the presented formulae, I is a fuzzy implicator and T is a t-norm. R_P is the fuzzy similarity relation induced by the subset of features P: $\mu_{R_P}(x,y) = \bigcup_{a \in P}\{\mu_{R_a}(x,y)\}$ where $\mu_{R_a}(x,y)$ is the degree to which objects x and y are similar for feature a. A fuzzy similarity relation can be constructed for this purpose, defined as: $\mu_{R_a}(x,y) = max(min(\frac{(a(y)-(a(x)-\sigma_a))}{(a(x)-(a(x)-\sigma_a))}, \frac{((a(x)+\sigma_a)-a(y))}{((a(x)+\sigma_a)-a(x))}), 0)$ where σ_a is the standard deviation of feature a. The fuzzy lower approximation contains information regarding the extent of certainty of object membership to a given concept. The fuzzy upper approximation contains information regarding the degree of uncertainty of objects. The couple $< \underline{P}(X), \overline{P}(X) >$ is called a *fuzzy-rough set.*

2) Reduction Process: To search for the optimal subset of features, the fuzzy-rough reduct, the fuzzy positive region has to be calculated. Formally, in the traditional RST, the crisp positive region is defined as the union of the lower approximations. By the extension to the fuzzy principal, the membership of an object $x \in U$ belonging to the *fuzzy positive region* can be defined by: $\mu_{POS_{R_P(Q)}}(x) = sup_{X \in U/Q} \mu_{\underline{R_P}X}(x)$. Object x will not belong to the fuzzy positive region only if the fuzzy equivalence class it belongs to is not a constituent of the fuzzy positive region. Using the definition of the fuzzy positive region, the *fuzzy-rough dependency function* can be defined as follows: $\gamma'_P(Q) = \frac{\sum_{x \in U} \mu_{POS_{R_P(Q)}}(x)}{|U|}$. As with crisp rough sets, the dependency of Q on P is the proportion of objects that are discernible out of the entire dataset. In the present approach, this corresponds to determining the fuzzy cardinality of $\mu_{POS_{R_P(Q)}}(x)$ divided by the total number of objects in the universe. A Fuzzy-Rough QuickReduct algorithm can be constructed for locating a fuzzy-rough reduct based on this measure. According to Fuzzy-Rough QuickReduct algorithm, the fuzzy dependency degree of the addition of each attribute to the current fuzzy reduct candidate, R, (initially empty) is calculated, and the best candidate is chosen. This process continues until the fuzzy dependency of the subset equals the fuzzy dependency degree (consistency) of the entire dataset, i.e., $\gamma'_R(D) = \gamma'_C(D)$. A worked example on how to compute a fuzzy-rough reduct using the Fuzzy-Rough QuickReduct algorithm, based on the fuzzy lower approximation, can be found in [6].

4 FLA-DCA: The Solution Approach

4.1 The FLA-DCA Signal Selection Process

For antigen classification, our learning problem has to select high discriminating features from the original input database which corresponds to the antigen information dataset. We may formalize this problem as an information table, where universe $U = \{x_1, x_2, \ldots, x_N\}$ is a set of antigen identifiers, the conditional attribute set $C = \{c_1, c_2, \ldots, c_A\}$ contains each feature of the information table to select and the decision attribute D of our learning problem corresponds to the class label of each sample. As FLA-DCA is based on the standard DCA concepts, except for the data pre-processing phase, and since DCA is applied to binary classification problems; then our developed FLA-DCA will be, also, applied to two-class datasets. Therefore, the decision attribute, D, of the input database of our FLA-DCA has binary values d_k: either the antigen is collected under safe circumstances reflecting a normal behavior (classified as normal) or the antigen is collected under dangerous circumstances reflecting an anomalous behavior (classified as anomalous). The condition attribute feature D is defined as follows: $D = \{normal, anomalous\}$. For feature selection, FLA-DCA computes, first of all, the fuzzy lower approximations of the two decision concepts d_k, for all attributes c_i and for all objects x_j; denoted by $\mu_{\underline{R_{c_i}}\{d_k\}}(x_j)$. Using these results, FLA-DCA calculates the fuzzy positive regions for all c_i, for each object x_j, defined as $\mu_{POS_{R_{c_i}}(D)}(x_j)$. Based on these calculations and to find the fuzzy-rough reduct, FLA-DCA starts off with an empty set and moves to calculate the fuzzy dependency degrees of D on c_i, defined as $\gamma'_{c_i}(D)$; as presented in Section 3. The attribute c_m having the greatest value of fuzzy-rough dependency degree is added to the empty fuzzy-rough reduct set. Once the first attribute c_m is selected, FLA-DCA adds, in turn, one attribute to the selected first attribute and computes the fuzzy-rough dependency degree of each obtained attributes' couple $\gamma'_{\{c_m,c_i\}}(D)$. The algorithm chooses the couple having the greatest fuzzy-rough dependency degree. The process of adding each time one attribute to the subset of the selected features continues until the fuzzy-rough dependency degree of the obtained set of features equals the fuzzy-rough dependency degree, $\gamma'_C(D)$, of the entire database.

4.2 The FLA-DCA Signal Categorization Process

Our method has to assign, now, for each selected attribute, included in the generated fuzzy-rough reduct, its definite and specific signal category; either as SS, as DS or as a PAMP signal. The general guidelines for signal categorization are based on the semantic of each signal category:

Safe signals : They certainly indicate that no anomalies are present.
PAMPs : They usually means that there is an anomalous situation.
Danger signals : They may or may not show an anomalous situation, however the probability of an anomaly is higher than under normal circumstances.

Based on the immunological definitions stated above, it is clear that both PAMP and SS are positive indicators of an anomalous and normal signal while the DS is measuring situations where the risk of anomalousness is high, but there is no signature of a specific cause. This problem can be formulated as follows: Based on the semantics of the mentioned signals, a ranking can be performed for these signals. More precisely, both SS and PAMP are more informative than DS which means that both of these signals can be seen as indispensable attributes; reflecting the first and the second ranking positions. To represent this level of importance, our method uses the first obtained couple of features through the fuzzy-rough reduct generation. On the other hand, DS is less informative than PAMP and SS; reflecting the last and third ranking position. Therefore, our method applies the rest of the fuzzy-rough reduct attributes, discarding the two first selected attributes that are chosen to represent the SS and PAMP signals, to represent the DS. More precisely, our method processes as follows:

As FLA-DCA has already calculated the fuzzy-rough dependency degree of each attribute c_i a part, $\gamma'_{c_i}(D)$, FLA-DCA selects the first attribute c_m having the greatest fuzzy-rough dependency degree to form the SS as it is considered the most informative first feature added to the fuzzy-rough reduct set. With no additional computations and since FLA-DCA has already computed the fuzzy-rough dependency degrees of each attributes' couple $\gamma'_{\{c_m,c_i\}}(D)$ when adding, in turn, one attribute c_i to the selected first attribute c_m that represents the SS, FLA-DCA chooses the couple having the greatest dependency degree. More precisely, FLA-DCA selects that second attribute c_r having the greatest $\gamma'_{\{c_m,c_r\}}(D)$ among the calculated $\gamma'_{\{c_m,c_i\}}(D)$; to form the PAMP signal. Finally, the rest of the reduct attributes are combined and affected to represent the DS as it is less than certain to be anomalous. Once signal categorization is achieved, FLA-DCA processes its next steps as the DCA does [1].

5 Experimental Setup

To test the validity of our FLA-DCA fuzzy-rough model, our experiments are performed on two-class, real-valued attributes, databases from [8]. In [5], a comparison between the rough DCA algorithms is performed and it was shown that QR-DCA outperforms both RST-DCA and RC-DCA in terms of classification accuracy. Another characteristic of QR-DCA is that it takes less time to process than RC-DCA and RST-DCA. Therefore, in what follows we will compare our new fuzzy-rough developed approach, FLA-DCA, to the QR-DCA crisp rough algorithm.

Table 1. Description of Databases

Database	Ref	♯ Instances	♯ Attributes
Sonar	SN	208	61
Molecular-Bio	Bio	106	59
Spambase	SP	4601	58
Cylinder Bands	CylB	540	40
Ionosphere	IONO	351	35
Sick	Sck	3772	30

For data pre-processing, FLA-DCA and QR-DCA uses FRST and RST, respectively. For both approaches, each data item is mapped as an antigen, with the value of the antigen equal to the data ID of the item. For all DCA algorithms, a population of 100 cells is used. To perform anomaly detection, a threshold which is automatically generated from the data is applied to the MCAVs. The MCAV threshold is derived from the proportion of anomalous data instances of the whole dataset. Items below the threshold are classified as class one and above as class two. The resulting classified antigens are compared to the labels given in the original datasets. For each experiment, the results presented are based on mean MCAV values generated across a 10-fold cross validation. We evaluate the performance of the mentioned DCA methods in terms of number of extracted features, running time, sensitivity, specificity and accuracy which are defined as: $Sensitivity = TP/(TP + FN)$; $Specificity = TN/(TN + FP)$; $Accuracy = (TP+TN)/(TP+TN+FN+FP)$; where TP, FP, TN, and FN refer respectively to: true positive, false positive, true negative and false negative.

6 Results and Analysis

In this Section, we aim to show that applying FRST, instead of RST, can avoid the information loss caused by the mandatory step of data quantization. Furthermore, we aim to show that by leaving the attribute values unchanged, our proposed fuzzy-rough FLA-DCA algorithm is able to select fewer features than the crisp rough QR-DCA approach, leading to better guide the FLA-DCA algorithm classification task. This is confirmed by the results presented in Table 2. For instance, from Table 2, we can notice that our new fuzzy-rough DCA model,

Table 2. Comparison Results of the Rough DCA Approaches

Database	Specificity(%)		Sensitivity(%)		Accuracy(%)		Time(s)		♯ Attributes	
	DCA		DCA		DCA		DCA		DCA	
	QR	FLA	QR	FLA	QR	FLA	QR	FLA	QR	FLA
SN	92.79	96.90	89.19	95.49	90.86	96.15	7.79	9.41	22	10
Bio	79.24	92.45	77.35	86.79	78.30	89.62	5.25	8.47	19	9
SP	98.67	99.89	99.17	99.77	98.87	99.84	1976.05	2071.8	11	8
CylB	97.75	98.39	97.00	97.00	97.46	97.85	12.68	18.96	7	5
IONO	96.88	99.11	96.03	98.41	96.58	98.86	15.88	30.72	22	9
Sck	97.65	99.12	96.53	96.96	97.58	98.99	510.05	602.8	22	14

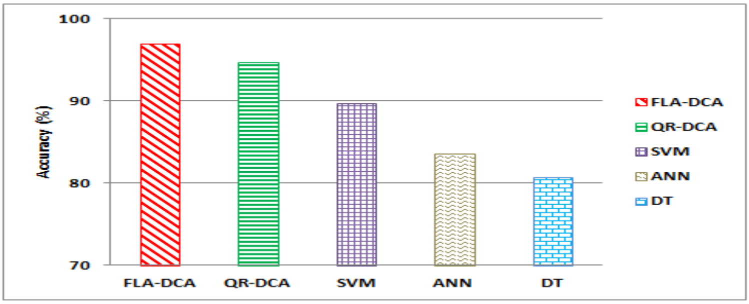

Fig. 1. Classifiers' Average Accuracies

FLA-DCA, selects fewer features than the crisp rough DCA model, QR-DCA. This is explained by the fact that FLA-DCA, by applying the Fuzzy-Rough QuickReduct algorithm, incorporates the information usually lost in crisp discretization by utilizing the fuzzy lower approximation to provide a more informed technique. For instance, applying FLA-DCA to the SN database, the number of selected attributes is reduced by more than 50% (10 features) in comparison to the number of features selected by QR-DCA, which is set to 22. Our FLA-DCA new approach performs much better than traditional RST on the whole, in terms of both feature selection and classification quality. For instance, when applying the algorithms to the SN dataset, the classification accuracy of FLA-DCA is set to 96.15%. However, when applying QR-DCA to the same database, the accuracy is set to 90.86%. Same remark is observed for the specificity and the sensitivity criteria. When comparing the results in terms of running time, we can notice that the time taken by our FLA-DCA to process is a bit longer than the time needed by QR-DCA to function. This is explained by the fact that our FLA-DCA incorporates the fuzzy component in comparison to QR-DCA.

We have, also, compared the performance of our FLA-DCA to other classifiers which are the Support Vector Machine (SVM), Artificial Neural Network (ANN) and the Decision Tree (DT). The comparison made is in terms of the average of accuracies on the databases presented in Table 1. Fig.1 shows that that the highest classification accuracy is noticed for our fuzzy-rough DCA new model, FLA-DCA.

7 Conclusion and Future Work

A new hybrid DCA classification model based on fuzzy rough set theory is proposed in this paper. Our fuzzy-rough model ensures a more rigorous data pre-processing when dealing with datasets with real-valued attributes. Results show that our FLA-DCA is capable of performing better its classification task than the crisp rough model and other classifiers. As future work, we aim to boost the DCA data pre-processing phase by extending the application of FRST to databases with missing data.

References

1. Greensmith, J., Aickelin, U., Twycross, J.: Articulation and clarification of the dendritic cell algorithm. In: Bersini, H., Carneiro, J. (eds.) ICARIS 2006. LNCS, vol. 4163, pp. 404–417. Springer, Heidelberg (2006)
2. Pawlak, Z.: Rough sets. International Journal of Computer and Information Science 11, 341–356 (1982)
3. Chelly, Z., Elouedi, Z.: RST-DCA: A dendritic cell algorithm based on rough set theory. In: Huang, T., Zeng, Z., Li, C., Leung, C.S. (eds.) ICONIP 2012, Part III. LNCS, vol. 7665, pp. 480–487. Springer, Heidelberg (2012)
4. Chelly, Z., Elouedi, Z.: RC-DCA: A new feature selection and signal categorization technique for the dendritic cell algorithm based on rough set theory. In: Coello Coello, C.A., Greensmith, J., Krasnogor, N., Liò, P., Nicosia, G., Pavone, M. (eds.) ICARIS 2012. LNCS, vol. 7597, pp. 152–165. Springer, Heidelberg (2012)
5. Chelly, Z., Elouedi, Z.: QR-DCA: A new rough data pre-processing approach for the dendritic cell algorithm. In: Tomassini, M., Antonioni, A., Daolio, F., Buesser, P. (eds.) ICANNGA 2013. LNCS, vol. 7824, pp. 140–150. Springer, Heidelberg (2013)
6. Jensen, R., Shen, Q.: New approaches to fuzzy-rough feature selection. IEEE Transactions on Fuzzy Systems 17, 824–838 (2009)
7. Dubois, D., Prade, H.: Putting rough sets and fuzzy sets together. Kluwer Academic Publishers, Dordrecht (1992)
8. Asuncion, A., Newman, D.J.: UCI machine learning repository (2007), `http://mlearn.ics.uci.edu/mlrepository.html`