Vol.2 Issue.5, Pg.: 96-101
May 2014

INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

MULTI OBJECTIVE ARTIFICIAL IMMUNE SYSTEM WEIGHTED FEATURE SELECTION FOR CLASSIFICATION

V.Karunakaran[1], Dr. M. Suganthi[2]

[1]*Assistant Professor, Department of Computer Science and Engineering, Karunya University karunakaran@karunya.edu*
[2] *Professor, Department of Electronics and Communication Engineering, Mahindra College of Engineering,*
msuganthib@gmail.com

Abstract

Supervised learning algorithm based on an AIS using clonal selection is already proposed [1]. Even though some AIS for classification have been developed, an artificial immune system for classification with local feature selection model has a unique feature; it includes the local feature selection mechanism. The aim of feature selection is to reduce the dimension of the input vector by the selection of a feature (variable) subset which describes the object in the best manner and ensures the best quality of the learning model. This set can be reduced during an optional apoptosis process. Local feature selection and apoptosis result in data-reduction capabilities. The classifier has only two user-settable parameters controlling the global-local properties of the feature space searching. In order to increase the performance of the system even more, in this paper proposed the weighted feature selection technique. In the weighted feature selection approach, we assign the weight value for each and every features based on their importance. Due to this technique, we can further reduce the features of the classification process. It is used to determine the levels of importance of the features (the strength of individual feature binding). For assigning weights to features, in other words a feature with weights proportional to the identification accuracy of individual features. Here assign the proper weight to each feature is helpful to improve the accuracy of the classification. The proposed system is effective in the feature reduction and has high classification accuracy compared to the existing system.

Keywords*:* Artificial Immune System, Clonal Selection.

1. Introduction

Artificial Immune Systems (AIS) is a diverse area of research that attempts to bridge the divide between immunology and engineering and is developed through the application of techniques such as mathematical and computational modelling of immunology, abstraction from those models into algorithm (and system) design and implementation in the context of engineering. AIS has become known as an area of computer science and engineering that uses immune system metaphors for the creation of novel solutions to problems.. **N**atural Immune systems, as the defence system of animal organisms against pathogens, were the inspiration behind the artificial immune systems (AIS). The interest of researchers is generated by such immune system features as: recognition of antigen (AG) characteristics, pattern memorization capabilities, self organizing memory, adaptation ability, immune response shaping, learning from examples, distributed and parallel data processing, multilayer structure and generalization capability. The application areas for AIS can be summarized as follows [3]:

1) Learning (clustering, classification, recognition, robotic and control applications),

2) Anomaly detection (fault detection, computer, and network security applications),

3) Optimization (continuous and combinatorial).

The most important types of AIS are based on the concepts of negative selection, clonal selection, and the immune network. Paper presents a supervised learning algorithm based on AIS using clonal selection. Even though some AIS for classification have been developed, this model has a unique feature; it includes the local feature selection mechanism.

The design prospective of development of computational tools inspired by nature is termed as biologically inspired computing. An immune system is a naturally occurring event-response system that can quickly adapt to the changing situations. The efficient mechanisms of a biological immune system (BIS) are ability to remember, classify and neutralize the effect of foreign particles. The understanding and investigation on BIS has increased dramatically over the recent few years by several researchers. These leads to development of new algorithms inspired by BIS, under a new branch of computational intelligence known as artificial immune system (AIS). The AIS is emerging as an active and attractive field involving models, techniques and applications of great diversity .It offers powerful and robust information processing capabilities for solving complex problems. Here the objective is to introduce new algorithms inspired by mechanism found in natural immune systems and to develop methodology to apply these algorithms to effectively solve problems of communication and control such as channel equalization and system

The aim of feature selection is to reduce the dimension of the input vector by the selection of a feature (variable) subset which describes the object in the best manner and ensures the best quality of the learning model. In this process irrelevant, redundant and unpredictive features are omitted. Popular feature selection methods are global, i.e., they determine one feature set for all training data. But one can imagine that different features are important in different regions of the input vector space.

2. Artificial Immune System for Classification with Local Feature Selection

The artificial immune system for classification with local feature selection allows the detection of many relevant feature sets (a separate relevant feature set is created for each learning point and its neighbourhood). This method of feature selection is inspired by the binding of an antibody (AB) to an AG, which occurs between amino acid residues forming an epitope and a paratope. Only certain selected residues (so-called energetic residues) take part in the binding. This approach reduces the curse of dimensionality that affects many machine learning methods.

3. Comparison of Existing System with AIS

- The performance of the existing system is low compared to the proposed system
- The classification accuracy of the system is low. Because the feature selection is normally applied in this system
- If the selected features are not efficient in the classification then the result of the classification is not optimal in this work
- The output error is increased in this system and degradation of the performance of the system is increased.

4. Artificial Immune System Weighted Feature Selection for Classification

In the paper introduce the weighted technique for the features. In the feature selection process, we introduced this. It is used to determine the levels of importance of the features (the strength of individual feature binding). For assigning weights to features, in other words a feature with weights proportional to the identification accuracy of individual features. Here assign the proper weight to each feature is helpful to improve the accuracy of the classification.

Vol.2 Issue.5, Pg.: 96-101
May 2014

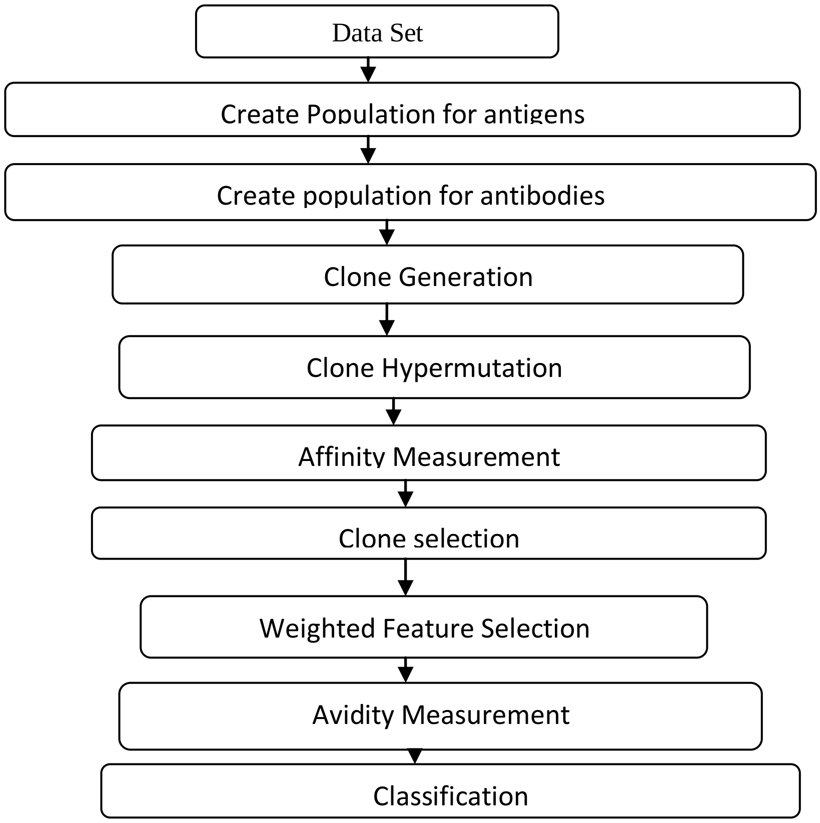

Figure1. Architecture diagram for Artificial Immune system weighted feature selection for classification

Fitness Function: Both normal and weighted features are combined and evaluated simultaneously. Recognition is performed using the feature selected by weight encoded for each features. Identification accuracy, used as the fitness function, is computed on the training set.it is used to select the feature for the next iterations.

Crossover: A set of uniform crossover operations is performed on *parents* to populate a new generation. Crossover operation is same for both normal and weighted features.

Hyper mutation: After crossover, mutation is performed for features by changing one or more weights by a factor of its standard deviation in the previous generation. For feature, mutation is performed by randomly inverting.

The search process is repeated till convergence and terminated when the identification performance of the features in new generation do not improve compared to the performance of classifications in previous generations. At this point optimal weight for each feature (i.e. features giving best recognition accuracy on the training data) are obtained.

$$Fitness = Accuracy * (Totfeatures - selected\ features)/Total\ features + (nmin/K) / Totfeatures) \quad (1)$$

nmin- cardinality of the near-neighbour minority set
K is the number of nearest neighbours

In the feature selection, we assign the weights to the every feature then it will proceed to the classification process. Assigning Weight by SNR feature selection Many researchers reported that SNR feature selection provided the best result for Classification [4-5]. We used this approach in this experiment. SNR is a statistical method that measures effectiveness of feature in identifying a class out of another class.

$$F= \mu 1-\mu 2/\sigma 1-\sigma \qquad (2)$$

The features of data are weighted by SNR score equation 2. These features are used in the terminal set (w1x1 .. ……wnxn,) where w_i is the SNR score of the feature i^{th} and n is the total number of features. Performance of classification is calculated with different parameters as follows

Accuracy: Accuracy can be calculated from formula given as follows

$$\text{Accuracy} = \frac{True\ positive + True\ negative}{True\ positive + True\ negative + False\ positive + False\ negative} \tag{3}$$

Precision: Precision value is calculated is based on the retrieval of information at true positive prediction, false positive. In healthcare data precision is calculated the percentage of positive results returned that are relevant.

$$\text{Precision} = \frac{True\ positive}{True\ positive + False\ positive} \tag{4}$$

Recall: Recall value is calculated is based on the retrieval of information at true positive prediction, false negative. In healthcare data precision is calculated the percentage of positive results returned that are Recall in this context is also referred to as the True Positive Rate. Recall is the fraction of relevant instances that are retrieved

$$\text{Recall} = \frac{True\ positive}{True\ positive + False\ negative} \tag{5}$$

F-measure comparison: F-measure distinguishes the correct classification of document labels within different classes. In essence, it assesses the effectiveness of the algorithm on a single class, and the higher it is, the better is the clustering. It is defined as follows:

$$F=2*precision*recall\ /precision + recall \tag{6}$$

TP (True positive): In a statistical hypothesis test, there are two types of incorrect conclusions that can be drawn. The hypothesis can be inappropriately. A positive test result that accurately reflects the tested-for activity of an analyzed. If the outcome from a prediction is p and the actual value is also p, then it is called a true positive (TP)

$$\text{True positive rate (TPR)} = TP/P \tag{7}$$

$$P= (TP+FN) \tag{8}$$

Where P is the positive. TP is the True Positive

TN (True negative): A result that appears negative when it should not. A true negative (TN) has occurred when both the prediction outcome and the actual value are n is the number of input data.

$$\text{True negative rate (TNR)} = TN/N \tag{9}$$

$$N= (TN+FN) \tag{10}$$

Where, N is the Negative value and TN is the True Negative.

FP (False positive): A result that indicates that a given condition is present when it is not. However if the actual value is n then it is said to be a false positive (FP).

$$\text{False positive rate } (\alpha) = FP / (FP + TN) \tag{11}$$

FN (False negative) :False negative (FN) is when the prediction outcome is n while the actual value is p.

$$\text{False negative rate } (\beta) = FN / (TP + FN) \tag{12}$$

Experimental result

We analyze and compare the performance offered by AIS and AIS with weighted feature selection. For experiment we take two data sets given in table 1 for calculating performance. The performance is evaluated by the parameters such as accuracy, precision, recall and f-measure. Classification result is shown in table 2. Based on the comparison and the results from the experiment show the proposed approach works better than the existing systems for some applications.

Table 1: Dataset used in experiments

Data set	Size	Features	Classes
Wine	178	13	3
Breast Cancer	699	9	2

Table 2: Classification Result

Dataset	**Recall**	**F-measure Comparison**	**Precision**
Wine	78.9	81.3	81.1
Breast Cancer	76.4	80.2	79.2

Conclusion

A new immune inspired general purpose classifier based on the concept of energetic residues was introduced. In order to reduce the number of memory cells, the apoptosis is introduced, but this reduces the classifier efficiency and increases the number of unrecognized test points. The AISLFS classifier is rather stable: no drastic changes in classifier accuracy over a wide range of parameter values were observed. The heuristic trial and error method is recommended to set the parameter values. The AISLFS classifier does quite well in comparison with both other immune inspired classifiers and other classifiers in general. The power of AISLFS is in its unique embedded approach to the local feature selection. The local dimensionality reduction property distinguishes it from other classifier solutions. The data reduction capability of AISLFS is its other important feature. In the proposed work we introduce the weighted technique for the features. In the feature selection process, we introduced this. It is used to determine the levels of importance of the features (the strength of individual feature binding). For assigning weights to features, in other words a feature with weights proportional to the identification accuracy of individual features. Here assign the proper weight to each feature is helpful to improve the accuracy of the classification. In the feature selection, we assign the weights to the every feature then it will proceed to the classification process. The proposed system is very effective than the existing system since the feature reduces according to their importance.

References

1. Grzegorz Dudek ,“An Artificial Immune system for classification with Local Feature Selection “, IEEE Transactions on Evolutionary Computation Vol 16, No 6, December 2012.
2. Supoj Hengpraprohm, “Feature Selection by Weighted-SNR for Cancer Microarray Data Classification” Internaational of Innovative Computing, Information and Control, 2008.
3. S. Hofmeyr and S. Forest, “Architecture for an artificial immune system”, Evolutionary Computing, Vol 7, no 1, PP 1289-1296, 2000.
4. J. Ryu and S.-B. Cho, Gene Expression Classification Using Optimal Feature/Classifier Ensemble with Negative Correlation, *Proc. of the 2002 Int. Joint Conf. on Neural Network*, pp.198 – 203, 2002.
5. U. Markowska-kaczmar and B. Kardas, “Multiclass iteratively refined negative selection classifier”, Applied Soft Computing, Vol 8, no. 2, PP 972-984, March 2008.

Vol.2 Issue.5, Pg.: 96-101
May 2014

6. Dasgupta, D., (1999), Artificial Immune Systems and Their Applications, Ed., Springer-Verlag.
7. Olfa Nasaroui , Fabio Gonzalez and Dipankar Dasgupta, "The Fuzzy Artificial Immune System: Motivations, Basic Concepts, and Application to Clustering and Web Profiling" , *IEEE*, 2002.

A Brief Author Biography

V.Karunakaran *–Currently working as Assistant Professor in the department of Computer Science and Engineering, Karunya University. His Current research interest includes Image processing, Computational Intelligence and Computer Algorithms.*

Dr. M. Suganthi *–Currently working as professor in the department of Electronics and Communication Engineering, Mahendra College of Engineering, Salem. Her current research interest includes Medical Image Processing, Neural Networks and Fuzzy systems.*