The Medical Applications of Attribute Weighted Artificial Immune System (AWAIS): Diagnosis of Heart and Diabetes Diseases

Seral Şahan[1], Kemal Polat[1], Halife Kodaz[2], and Salih Güneş[1]

[1] Selcuk University, Eng.-Arch. Fac. Electrical & Electronics Eng.
42031-Konya/Turkey
{seral, sgunes, kpolat}@selcuk.edu.tr
[2] Selcuk University, Eng.-Arch. Fac. Computer Eng.
42031-Konya/Turkey
hkodaz@selcuk.edu.tr

Abstract. In our previous work, we had been proposed a new artificial immune system named as Attribute Weighted Artificial Immune System (AWAIS) to eliminate the negative effects of taking into account of all attributes in calculating Euclidean distance in shape-space representation which is used in many network-based Artificial Immune Systems (AISs). This system depends on the weighting attributes with respect to their importance degrees in class discrimination. These weights are then used in calculation of Euclidean distances. The performance analyses were conducted in the previous study by using machine learning benchmark datasets. In this study, the performance of AWAIS was investigated for real world problems. The used datasets were medical datasets consisting of Statlog Heart Disease and Pima Indian Diabetes datasets taken from University of California at Irvine (UCI) Machine Learning Repository. Classification accuracies for these datasets were obtained through using 10-fold cross validation method. AWAIS reached 82.59% classification accuracy for Statlog Heart Disease while it obtained a classification accuracy of 75.87% for Pima Indians Diabetes. These results are comparable with other classifiers and give promising performance to AWAIS for that kind of problems.

1 Introduction

A new artificial intelligence area named as Artificial Immune Systems (AISs) is going forward gradually. There are many AIS algorithms in which recognition and learning mechanisms of immune system were modeled. As a representation method of immune system cells, shape-space approach is used in many of the AIS classification algorithms. Shape-space model, which was proposed by Perelson and Oster in 1979 [1], is used as a representation mechanism modeling the interactions between two cells in the immune system.

In the systems that use a distance criterion as a similarity metric, some shape-space related problems may exist in case of irrelevant attributes [2], [3]. One attribute value

C. Jacob et al. (Eds.): ICARIS 2005, LNCS 3627, pp. 456 – 468, 2005.

in shape space can cause two data in the same class to be distant from each other and therefore to be recognized and classified by different system units. If that attribute is irrelevant for class discrimination process, the algorithm may result in erroneous classes.

In our previous study [4], it was aimed to reach higher classification accuracy by assigning weights to important attributes in classification. This was done with some modifications to affinity measures of AISs and then a system named AWAIS (Attribute Weighted Artificial Immune System) has come into existence. In that paper, we had conducted the performance analyses of AWAIS for chainlink and two-spirals datasets which are commonly used machine learning benchmarks and for wine dataset representing a real-world problem. For all of those problems, the performance of AWAIS was very satisfactory and promising. In this paper we carried applications of AWAIS for real-world situations further and used AWAIS as a classifier in medical domain to diagnose diseases. The problems dealt in this study are Heart Disease and Diabetes Diagnosing problems via classification. Used datasets were taken from UCI Machine Learning Repository carrying the names Statlog Heart Disease and Pima Indians Diabetes respectively [5]. A form of k-fold cross validation which is a very commonly used method was used to evaluate classification accuracies more reliably in the experimental studies. The obtained classification accuracies were 82.59% and 75.87% for Heart Disease and Diabetes respectively.

This paper is organized as follows. In the second section of this paper, the background information is given including natural and artificial immune systems, shape-space representation and curse of dimensionality problem. The third section is reserved for introduction of AWAIS. The used datasets and k-cross validation are all given in fourth section of the paper under the title of method. Results and discussions about these results were given in section five which is then followed by the conclusion in section six.

2 Background

2.1 Natural and Artificial Immune Systems and Shape-Space Representation

The natural immune system is a distributed novel-pattern detection system with several functional components positioned in strategic locations throughout the body [6]. Immune system regulates the defense mechanism of body by means of innate and adaptive immune responses. Between these, adaptive immune response is much more important for us because it contains metaphors like recognition, memory acquisition, diversity, self-regulation…etc. The main architects of adaptive immune response are Lymphocytes, which divide into two classes as *T* and *B* Lymphocytes (cells), each having its own function. Especially *B* cells have a great importance because of their secreted antibodies (*Ab*s) that takes very critical roles in adaptive immune response. For detailed information about immune system refer to [7].

AISs emerged in the 1990s as a new computational research area. AISs link several emerging computational fields inspired by biological behavior such as Artificial Neural Networks and Artificial Life [8].

Among the studies conducted in the field of AIS, *B* cell modeling is the most encountered representation type. Different representation methods have been proposed in that modeling. Among these, shape-space representation is the most commonly used one [1].

The shape-space model (*S*) aims at quantitatively describing the interactions among antigens (*Ag*s), the foreign elements that enter the body like microbe,…etc., and antibodies (***Ag-Ab***). The set of features that characterize a molecule is called its *generalized shape*. The ***Ag-Ab*** representation (binary or real-valued) determines a distance measure to be used to calculate the degree of interaction between these molecules. Mathematically, the generalized shape of a molecule (m), either an antibody or an antigen, can be represented by a set of coordinates $m = \langle m_1, m_2, \ldots m_L \rangle$, which can be regarded as a point in an L-dimensional real-valued shape-space ($m \in S^L$)[6]. In this work, we used real strings to represent the molecules. Antigens and antibodies were considered of same length L. The length and cell representation depends upon the problem.

2.2 Problems with Euclidean Distance as an Affinity Measure

The shape-space representation gives a good model of interactions in immune system but because the affinities between Abs to Ags are calculated based on a distance criterion, some problems exist like other distance-based approaches. The distances between instances are calculated based on all attributes of the instances and so distance can be dominated by irrelevant attributes [2]. To illustrate this, let us think the two points of a same class shown in Fig. 1. Again we will assume second attribute is not so important for class determination. The first and the third attributes of these two points are the same but the difference in second attribute value results the two data point to be apart from each other. If we take second attribute into account in a same degree with other two attributes, a possible wrong decision about the class of points can be done. This is also a problem because each attribute value is squared while determining Euclidean distance as stated in [9].

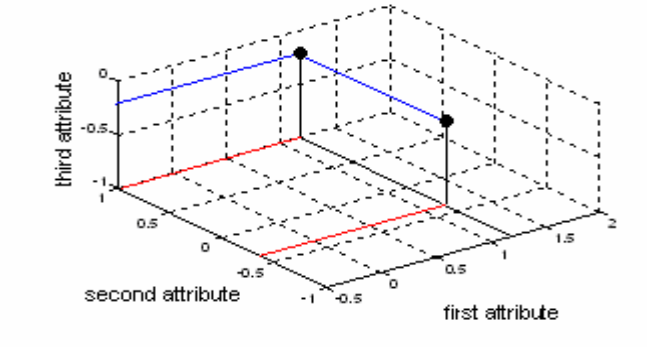


Fig. 1. Two points in same class illustrating curse of dimensionality problem

This difficulty, which arises when many irrelevant attributes present can be solved by using attribute weights. Using weighted attributes is a common way to get rid of the problem like this as done in ML algorithms like in [10], [11] and etc.

3 AWAIS (Attribute Weighted Artificial Immune System)

As mentioned before, most of network-based AIS algorithms use shape-space representation and the problem stated above related with irrelevant attributes inevitably appeared in turn affects the system performance. The AWAIS algorithm proposed for minimizing the effect of this problem is a supervised Artificial Immune System based on attribute-weighted distance criterion. The supervision in the algorithm shows itself while determining the weights of attributes and during the process of developing memory cells in the training by taking the class of the input data into account. AWAIS is a two-stage classification system in which attribute weights of each class are formed in one level and a training procedure with these weights takes place at the other.

3.1 Attribute Weighting

In most real valued shape-space representations, the distance between two points is calculated by the Euclidean distance criteria (Eq. (1)):

$$D = \sqrt{\sum_{i=1}^{L} (ab_i - ag_i)^2} . \quad (1)$$

Where ***ab*** and ***ag*** are the two points in the shape-space represented by a vector respectively and *L* is the length of these vectors. According to this formula, all of the attributes have same effect in determining distance. However, there are such data sets that some attributes of them have no effect on the class of data while some other attributes are more important in determining class. So, if it is assigned higher weights to the attributes that are more important in determining one class and if these weights are used in calculation of distance, it can be prevented to make a misclassification of the two distant data according to the Euclidean norm in the same class [2]. Starting from this point, the used attribute weighting depends on the following base: if one attribute doesn't changing very much among the data of one class, this attribute is one of the characteristic attributes of related class and it must have a higher weight than others [12].

The applied attribute weighting procedure in the AWAIS is as follows:

(1) Normalization of each attribute in data set between 0-1.
(2) Determine the antigens of each class→ Ag_class_j (j:1,….n, n: number of class)
(3) For each class do:
    For $Ag\_class_{(LxNc)}$ to be a matrix that involves the antigens of that class;
    (L: attribute num., Nc: ag num. of that class);

(3.1) For i^{th} attribute do:(i:1,…..L)
Evaluate standard deviation of i^{th} attribute with Eq. (2):

$$\text{std\_dev}_i = \sqrt{\frac{1}{Nc}\sum_{k=1}^{Nc}\left(Ag_{k,i} - \text{mean}\,(Ag_i)\right)^2}. \quad (2)$$

Here $Ag_{k,i}$ is the i^{th} attribute of k^{th} Ag in j^{th} class; $mean(Ag_i)$ is the mean of i^{th} attribute of all Ags in j^{th} class.
Calculate the weights as follows:

$$w_{j,i}=1/std\_dev_i \ , \ (i=1,\ldots L;\ j=1,\ldots n) \quad (3)$$

(3.2) normalize the weights of j^{th} class.

The calculated w_{nxL} matrix is a normalized weight matrix involving the weights of each attribute for each class and this matrix is used in distance calculations of the training algorithm of AWAIS.

Here, in the attribute weighting procedure, a means of normalization of attributes for each class by standard deviation is performed. By doing so, each class has its own set of attribute weights.

3.2 AWAIS Training Algorithm

The training procedure of the algorithm conducts the following steps:

(1) For each Ag_i do :(i: 1,…N)
(1.1) Determine the class of Ag_i. Call memory Abs of that class and calculate the distances between Ag_i and these memory Abs with Eq. (4):

$$D = \sqrt{\sum_{k=1}^{L} w_{j,k}\left(Ab_{j,k} - Ag_{i,k}\right)^2}. \quad (4)$$

Here $Ab_{i,k}$ and $Ag_{i,k}$ are the k^{th} attribute of Ab_j and Ag_i respectively; $w_{j,k}$ is the weight of k^{th} attribute that belongs to the class of Ab_j.
(1.2) If the minimum distance among the calculated distances above is less than a threshold value named as suppression value ($supp$) then return to step 1.
(1.3) Form a memory Ab for Ag_i:
At each iteration do:
(1.3.1) Make a random Ab population with $Ab=[Ab\_mem\ ;\ Ab\_rand]$ and calculate the distances of these Abs to Ag_i.

(1.3.2) Select *m* nearest *Abs* to Ag_i; clon and mutate these *Abs* (*Ab_mutate*).
(1.3.3) Keep the *m* nearest *Abs* in the *Ab_mutate* population to Ag_i as *Ab_mem* temporary memory population.
(1.3.4) Define the nearest *Ab* to Ag_i as *Ab_cand*, candidate memory *Ab* for Ag_i and stop iterative process if the distance of *Ab_cand* to Ag_i is less that a threshold value named as stopping criterion (*sc*).
(1.3.5) Concatenate *Ab_cand* as a new memory *Ab* to memory matrix of the class of Ag_i.
(1.4) Stop training.

The mutation mechanism in the algorithm which is used in many AIS algorithms and named as *hypermutation* is performed proportional to distance between two cells (Eq. (5)):

$$Ab_{j,k}' = Ab_{j.k} \pm D_{j,i} * (Ab_{j.k}) \tag{5}$$

Here $Ab_{j.k}'$ is the new value and $Ab_{j.k}$ *is the old value of* k^{th} attribute of j^{th} *Ab*. $D_{j,i}$ stands for the distance between Ag_i and Ab_j.

The used affinity measure is no more a pure Euclidean Distance and the attribute weights are used in distance criteria. The classes of memory *Abs* in the AWAIS after training are known with the aid of a labeling vector that contains the information about which memory *Abs* belong to which class.

After memory Antibodies is formed by this training procedure, test samples are presented to these Antibodies and the classes of these samples are determined by using k-nearest neighbor method.

4 Method

4.1 Statlog Heart Disease and Pima Indians Diabets Datasets

The Statlog Heart disease dataset was taken from UCI Machine Learning Respiratory [5]. 270 samples belong to patients with heart problem while the remaining 150 samples are of healthy persons. The samples taken from patients and healthy persons include 13 attributes which are: 1. age, 2. sex, 3. chest pain type (4 values), 4. resting blood pressure, 5. serum cholestoral in mg/dl, 6. fasting blood sugar > 120 mg/dl, 7. resting electrocardiographic results (values 0,1,2), 8. maximum heart rate achieved, 9. exercise induced angina, 10. oldpeak = ST depression induced by exercise relative to rest, 11. the slope of the peak exercise ST segment, 12. number of major vessels (0-3) colored by flourosopy, 13. thal: 3 = normal; 6 = fixed defect; 7 = reversable defect This dataset has 13 attributes and 2 classes. The class information is included in the dataset as 1 and 2 regarding absence and presence of disease respectively.

The other used data set for Diabetes problem was also taken from the same database and it is named as Pima Indians Diabetes [5]. This dataset contains 768 samples taken from healthy and unhealthy persons. 500 of these samples belong to persons

with no diabetes problem while the remaining 286 sample are of persons with diabetes. The class information contained in this data set is given by 0 for healthy persons and by 1 for diabetic patients. The number of attributes in samples is 8. These attributes are: 1. Number of times pregnant, 2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test, 3. Diastolic blood pressure (mm Hg), 4. Triceps skin fold thickness (mm), 5. 2-Hours serum insulin (mu U/ml), 6. Body mass index (weight in kg/(height in m)^2), 7. Diabetes pedigree function, 8. Age (years).

4.2 K-Fold Cross Validation

In this study, the classification accuracies for the datasets were measured according to the Eq. (6):

$$accuracy(T) = \frac{\sum_{i=1}^{|T|} assess(t_i)}{|T|}, t_i \in T$$
$$assess(t) = \begin{cases} 1, & if\ classify(t) = t.c \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where T is the set of data items to be classified (the test set), $t \epsilon T$, $t.c$ is the class of the item t, and classify(t) returns the classification of t by AIRS.

For test results to be more valuable, k-fold cross validation is used among the researchers. It minimizes the bias associated with the random sampling of the training [13]. In this method, whole data is randomly divided to k mutually exclusive and approximately equal size subsets. The classification algorithm trained and tested k times. In each case, one of the folds is taken as test data and the remaining folds are added to form training data. Thus k different test results exist for each training-test configuration. The average of these results gives the test accuracy of the algorithm [13]. We used this method as 10-fold cross validation in our applications.

5 Results and Discussion

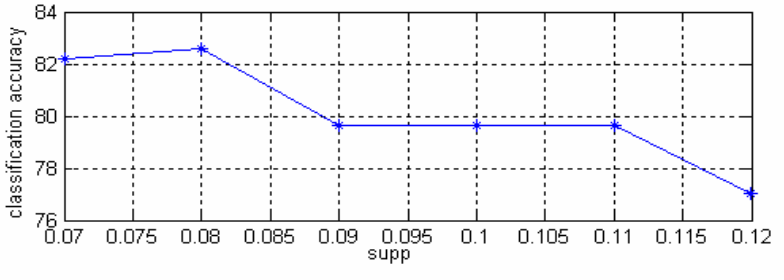
5.1 Results for Statlog Heart Disease

Whereas AWAIS has a number of parameters that affect the classification performance of the algorithm, the key parameter to adjust in AWAIS algorithm is *supp* parameter since it determines the number of memory Abs so the classification accuracy. The other parameters in the system were found to have little effect on classification accuracy, they would rather affect the classification time of the algorithm. The number of best memory cells selected in each iteration, which is given by m, was selected as 25. An Ab population that consists of 100 members was used. This population consists of 25 best Abs from previous iteration and 75 randomly generated Abs. The percentage of memory Abs was chosen experimentally and from these experimentations it was found that if there were less number of memory Abs, the algorithm had tended to be more like a random search algorithm while if the percentage of memory Abs were chosen to be high, the population had dominated with memory Abs and the algorithm had converged to the best individual. But this had no serious effect on classification accuracy, it only affected the classification time of the algorithm.

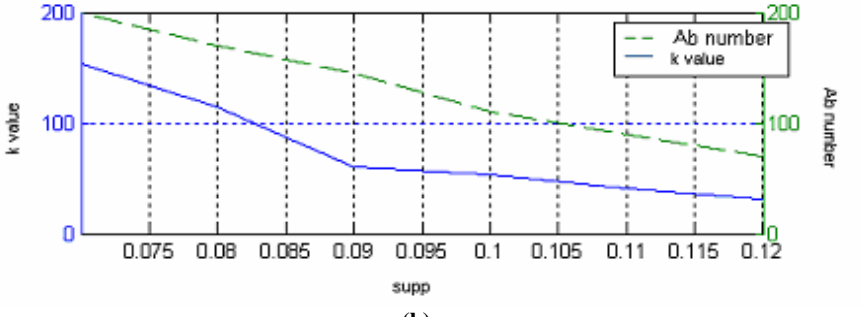
The value of supp parameter is selected in the [0,1] range. If this value is selected too high, the number of Abs will be too low and in contrary if this value is too low, there will be more memory Abs. The number of memory Abs highly affects the classification performance. Besides of supp parameter the k value for k-nn also affects the classification accuracy. Because the number of memory Abs is different for each supp value, the k value was changed for each supp value to obtain highest classification accuracy.

The classification accuracy with respect to the supp parameter is plotted in Fig. 2 (a) with respect to the supp parameter. The number of memory Abs and variation of k-value for which the highest accuracy was obtained with respect to the supp value are shown in Fig. 2 (b).

The dashed line in Fig. 2 (b) shows the memory Ab number for each supp value and the straight line represents the k-values at which highest classification accuracies were obtained for each supp value. As stated above, the number of Abs grows with decreasing supp while this growing results in higher classification accuracy to a degree. The maximum classification accuracy was obtained for 0.08 value of supp parameter as 82.59%. For this value of supp parameter, the number of memory Abs was about 160. Also, as can be seen from the Fig. 2 (b), k value increases with decreasing supp value as proportional to the number of memory Abs.



(a)



(b)

Fig. 2. (a) obtained classification accuracies with respect to the supp parameter, (b) k value and Ab number versus supp parameter (for Statlog Heart Disease)

Table 1. Classification accuracies obtained by AWAIS and other classifiers for the Statlog Heart Disease dataset

Author	Method	Accuracy (%)
WEKA, RA	Naive-Bayes	83.60
Our Study (2005)	**AWAIS**	**82.59**
Newton Cheung (2001)	Naive Bayes	81.48
Newton Cheung (2001)	BNND	81.11
Newton Cheung (2001)	C4.5	81.11
Newton Cheung (2001)	BNNF	80.96
Robert Detrano	Logistic regression	77.00
WEKA, RA	K*	76.70
WEKA, RA	IB1c	74.00
WEKA, RA	1R	71.40
WEKA, RA	T2	68.10
ToolDiag, RA	MLP+BP	65.60
WEKA, RA	FOIL	64.00
ToolDiag, RA	RBF	60.00
WEKA, RA	InductH	58.50

The classification accuracy obtained by AWAIS is shown in Table.1 with accuracies obtained for the same problem with other classifiers in Literature [14]. 10-fold cross validation was used in all of the classifiers in the table. The table shows that, AWAIS is the second best performed classifier after the study of WEKA group with respect to the classification accuracy. This promising result gives the way for AWAIS to be used in real-world problems as other classifiers. Whereas AWAIS couldn't reach the highest accuracy for the problem, the obtained accuracy is good for an AIS algorithm that uses a distance criterion as an affinity measure. When we look at the table, the highest accuracy was obtained by Naïve-Bayes classifier and Bayes classifiers are known to be optimal classifiers for some kind of problems. It is promising to see that an AIS algorithm performs comparable with a good classifier and if appropriate algorithm formulation for the problem at hand is constructed, even an over-performed AIS algorithm can be found.

5.2 Results for Pima Indians Diabetes

As for the Statlog Dataset, supp parameter was adjusted to obtain highest classification accuracy. Fig. 3 (a) shows the classification accuracy with respect to the supp parameter. The number of memory Abs and k value for corresponding supp parameter values are presented in Fig. 3 (b).

The maximum classification accuracy was obtained for 0.06 value of supp parameter as 75.87%. For this value of supp parameter, the number of memory Abs was about 230.

The dashed line in Fig. 3 (b) shows the memory Ab number for each supp value and the straight line represents the k-values at which highest classification accuracies were obtained for each supp value. Again, k value increases with decreasing supp value as proportional to the number of memory Abs.

The classification accuracy obtained by AWAIS for Pima Indians Diabetes dataset is shown in Table.2 with accuracies obtained for the same problem with other classifiers in Literature [14]. Indeed, Pima Indians Diabetes is a popular medical classification dataset among Machine Learning researchers. So many studies have been conducted related with this classification problem. Whereas high classification accuracies were reached, the accuracies couldn't go above 80% with cross validation scheme. This is also a reason why researchers are dealing with this problem. Because of the vast amount of classifiers were used for this dataset, only some of them are reported in the Table 2 and this is enough to have an opinion how is the performance of AWAIS is.

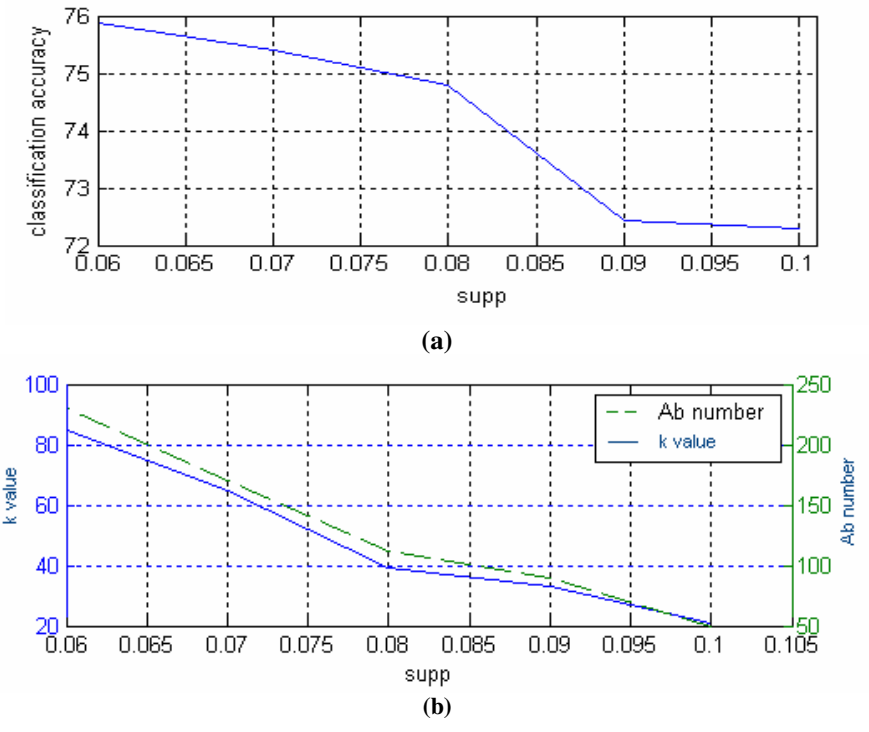


Fig. 3. (a) obtained classification accuracies with respect to the supp parameter, (b) k value and Ab number versus supp parameter (for Pima Indians Diabetes dataset)

According to the table, AWAIS is comparable with other classifiers. Whereas the classification accuracy obtained by AWAIS is also less than those of some classifiers, it was satisfactory to see AWAIS as an average classifier for this dataset if we remember that this dataset is hard to classify. Also, as can be seen from the table, an other AIS, AIRS, was applied by Watkins for this problem and a classification accuracy of 74.10% was obtained [16]. It is good to see that AWAIS has been overperformed to another AIS algorithm for this problem. Also, as it can be seen from the

Table 2. Classification accuracies obtained by AWAIS and some other classifiers for the Pima Indians Diabetes datasets

Method	Accuracy (%)	Reference
Logdisc	77.70	Statlog
IncNet	77.60	Norbert Jankowski
DIPOL92	77.60	Statlog
Linear Discr. Anal.	77.50	Statlog
SMART	76.80	Statlog
GTODT (5xCV)	76.80	Bernet and Blue
kNN,k=23,Manh,raw,W (3xCV)	76.7±4.0	WD-GM,feat. Weigh.
kNN,k=1:25,Manh,raw	76.6±3.4	WD-GM
ASI	76.60	Ster & Dobnikar
Fisher discr. Analysis	76.50	Ster & Dobnikar
MLP+BP	76.40	Ster & Dobnikar
AWAIS (10xCV)	**75.87**	**Our study**
MLP+BP	75.8±6.2	Zarndt
LVQ	75.80	Ster & Dobnikar
LFC	75.80	Ster & Dobnikar
RBF	75.70	Statlog
NNEE (2004) [15]	75.57	Y.Jiang &Z.-H.Zhou
NB	75.5-73.8	Ster & Dobnikar
kNN, k=22, Manh	75.50	Karol Grudzinski
MML	75.5±6.3	Zarndt
SNB	75.40	Ster & Dobnikar
BP	75.20	Statlog
SSV DT	75.0±3.6	WD-GM
kNN, k=18,Euclid,raw	74.8±4.8	WD-GM
ILAS	74.80	Jaurne et. al
CART DT	74.7±5.4	Zarndt
CART DT	74.50	Statlog
DB-CART	74.40	Shang & Breiman
ASR	74.30	Ster & Dobnikar
AIRS (13xCV) [16]	74.10	Watkins
SSV DT	73.7±4.7	WD-GM
C 4.5 DT	73.00	Statlog
CART	72.80	Ster & Dobnikar
C 4.5 DT	72.7±6.6	Zarndt
Kohonen	72.70	Statlog
Bayes	72.2±6.9	Zarndt
C 4.5 (5xCV)	72.00	Bernet and Blue

table, another algorithm with distance criterion, k-nn algorithm, was used for this problem with manhattan distance. The obtained results for this algorithm are better than AWAIS for k=23 with 3xCV and for k=1:25. The classification accuracy is better but in this case the classification time of AWAIS is probably less than k-nn because the used system units in k-nn algorithm are the raw training data whereas for AWAIS the memory Abs are used for classification which are far less than training data. Besides of this, for k-nn, the used distance criterion is Manhattan instead of Euclidean. If we conduct the comparison of AWAIS with k-nn algorithm in the same

context, we must take the classification result of k-nn with Euclidean distance which was found to be 74.8% according to the table. Than it is straightforward to say that the use of Manhattan distance can improve the performance of algorithm which is also stated in [9].

6 Conclusions

Shape-space representation, especially used in many network-based AIS algorithms is a means of representing immune system units as system units and this representation scheme also defines the interactions of the system units with the environment by means of distance criteria. A problem caused by irrelevant attributes based on distance criteria appeared in the distance-based classification systems affects the classification performance in negative manner especially for nonlinear data sets. A system named as AWAIS had been proposed for minimizing these negative effects by weighting attributes and using these weighted attributes in calculation of distances. In this paper the real-world performance of this system was analyzed through two medical classification problems. These are heart and diabetes diseases classification problems and the used datasets for these problems were taken from UCI machine learning repository carrying names Stotlog Heart Disease and Pima Indians Diabetes respectively. These datasets were taken especially for their hardness in classification. In performance evaluation of our system for regarding problems, 10 fold cross validation scheme was used.

AWAIS has performed very well for used datasets by reaching 82.59% and 75.87% classification accuracy for Statlog Heart Disease and Pima Indians Diabetes respectively. These results are not the highest ones among other classifiers applied to corresponding datasets so far but the place of AWAIS among them is satisfactorily high. Besides, as stated above, the used datasets are among the medical classification datasets that are hard to classify. Furthermore, another AIS, AIRS, had been applied for Pima Diabetes dataset and the classification accuracy of AWAIS is higher than it's.

Without a doubt, the weighting procedure that was adapted in AWAIS is not the only way for weighting. This weighting procedure assumes that the attribute values in one class doesn't change so much. However, there can be lots of practical problems that don't obey this assumption and for these problems AWAIS may not do so much. Also, weighting each attribute independently seems to be ineffective for datasets that have highly correlated attributes. There are lots of methods to determine the relevancy of attributes in literature and they can be used for determining attributes weights. The use of these methods may add to the performance of AWAIS more than our method. But the key point in AWAIS is to eliminate the negative effects of taking into account all attributes in calculation of distances due to the possible disadvantages of this scheme. It has not strictly stated that the AWAIS can be used with only its weighting procedure. Also, instead of Euclidean distance, Manhattan distance can be used as affinity measure and this opens the way for further studies in this context.

Acknowledgements

This work is supported by the Coordinatorship of Selçuk University's Scientific Research Projects Grant.

References

1. Perelson, A.S. and Oster, G. F., Theoretical Studies of Clonal Selection: Minimal Antibody Repertoire Size and Reliabilty of Self-Nonself Discrimination, Journal of Theoretical Biology, Vol. 81(1979) 645-670
2. Mitchell, T.M., Machine Learning, The McGraw-Hill Companies Press (1997) 234-236
3. Hart, E., Immunology as a Metaphor for Computational Information Processing: Fact or Fiction?, Doctor of Philosophy, Artificial Intelligence Applications Institute, Division of Informatics, University of Edinburgh (2002) 28-34
4. Şahan, S., Kodaz, H., Güneş, S., Polat, K., A New Classifier Based on Attribute Weighted Artificial Immune System, Lecture Notes in Computer Science, Vol. 3280. Springer-Verlag, Berlin Heidelberg (2004) 11-20
5. Blake, C.L., and Merz, C.J., UJI Reporsitory of Machine Learning Databases, (1996) http://www.ics.uci.edu./~mlearn/MLReporsitory.html
6. De Castro, L.N. and Von Zuben, F. J, Artificial Immune Systems: Part I-Basic Theory and Applications, Technical Report, TR-DCA 01/99 (1999) 3-14
7. Abbas, A.K., Lichtman, A.H., Pober, J.S.,Cellular and Molecular Immunology, 4th ed., W.B. Saunders Company (Ed.), 2000
8. De Castro, L.N and Timmis, J., Artificial Immune Systems: A New Computational Intelligence Approach, Springer-Verlag Press (2002)
9. Freitas, A. and Timmis., J., Revisiting the Foundations of Artificial Immune Systems: A Problem Oriented Perspective, Timmis, J., Bentley, P., and Hart, E. (editors), Proceedings of the 2nd International Conference on Artificial Immune Systems, volume 2787 of Lecture Notes in Computer Science, Springer, (2003) 229-241
10. Emam, K. E., Benlarbi, S., Goel., N., and Rai, S. N., Comparing case-based reasoning classifiers for predicting high risk software components, The Journal of Systems and Software, Vol. 55. Elsevier, (2001) 301-320
11. Liao, T.W., Zhang, Z., and Mount, C.R., Similarity measures for retrieval in case-based reasoning systems, Applied Artificial Intelligence, Vol. 12. (1998) 267-288
12. Cherkassky, V. and Mulier, F., Learning From Data, John Wiley & Sons, Inc. (Ed.), 1998
13. Delen, D., Walker, G., Kadam, A., Predicting breast cancer survivability: a comparison of three data mining methods, Artificial Intelligence in Medicine, in Pres (Sep. 2004)
14. http://www.phys.uni.torun.pl/kmk/projects/datasets.html#Sheart (last modified 2002)
15. Jiang, Y. and Zhou, Z.-H., Editing training data for kNN classifiers with Neural Network ensemble, Lecture Notes in Computer Science, Vol. 3173. Springer-Verlag, Berlin Heidelberg (2004) 356–361
16. Watkins, A., AIRS: A Resource Limited Artificial Immune Classifier, Master Thesis, Mississippi State University 2001