ESANN'2006 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2006, d-side publi., ISBN 2-930307-06-4.

Immune Network based Ensembles

Nicolás García-Pedrajas[1] and Colin Fyfe[2]

1- Dept. of Computing and Numerical Analysis
University of Córdoba (SPAIN)
e-mail: npedrajas@uco.es

2- the Dept. of Computing
University of Paisley
e-mail: colin.fyfe@paisley.ac.uk

Abstract. This paper presents a new method for constructing ensembles of classifiers based on Immune Network Theory, one of the most interesting paradigms within the field of Artificial Immune Systems. Ensembles of classifiers are a very interesting alternative to single classifiers when facing difficult problems. In general, ensembles are able to achieve better performance in terms of learning and generalization error.
We construct an Immune Network that constitutes an ensemble of classifiers. Using a neural network as base classifier we have compared the performance of this ensemble with five standard methods of ensemble construction. This comparison is made using 35 real-world classification problems from the UCI Machine Learning Repository. The results show a general advantage of the proposed model over the standard methods.

1 Introduction

Classifiers ensembles[1] are receiving increasing attention in recent research in the machine learning community, due to their interesting features. They are a powerful tool especially when facing complex problems. For a detailed descriptions of ensembles the reader is referred to [2].

In most cases, classifiers in an ensemble are designed independently or sequentially, so the advantages of interaction and cooperation among the individual classifiers are not exploited. In this paper we show how the interdependent evolution of the classifiers that make up the ensemble by means of an Artificial Immune Network is able to obtain very good performance.

2 Artificial Immune System

Artificial Immune System is a recent paradigm in the growing field of bioinspired algorithms that mimics the immune system of animals. The immune system [3] is a complex system that enables a mechanism by which certain dangers to the organism can be identified. These dangers can be roughly classified as those which arise from dysfunction within the organism's own cells and those which are due to the action of exogenous pathogens.

The adaptive immune system is believed to be continually creating antibodies in a somewhat random fashion: it is rather as though it is exploring the space of antigens always on the lookout for new dangers to the organism

Jerne's immune network theory [4] is one of the most interesting fields within both natural and artificial Immune Systems. This proposes that the various parts of the immune system itself recognise other parts of the same system and indeed affect the production or suppression of other parts of the system. A positive reaction between components can lead to cell proliferation and activation while a negative response can lead to cell tolerance and suppression.

3 Immune Network Ensemble

Many of the features of Immune Networks are very appropriate to the design of committee machines, such as classifier ensembles. In the immune network the immune cells must cooperate to defend the individual; in classifier ensembles the different classifiers must develop different behaviors in order to collaborate to solve the given problem. The model we present in this paper can be applied to any committee machine regardless of the classifier used.

In order to develop an Immune Network to design a classifier ensemble, the first step is the definition of the immune elements to use. We will have only two immune elements: antibodies and antigens, as these are the basic elements of any Immune Network. Each antibody is a classifier potentially usable to form the ensemble.

The antigens represent the problem to be solved. However, the way the problem is represented by the antigens can be subject to many different approaches. We have a set of training patterns and each antigen must represent a different view of the classification problem. We have developed our model using three different definitions for the antigen:

1. A bootstrapping sample from the training set, as in *bagging* [5].

2. One of the most interesting alternatives to ensemble construction using resampling of patterns is the Random Subspace Method [6]. This method was proposed for constructing a decision forest by randomly selecting subspaces from the original dataset, and very good results were reported. Random subspace method has been successfully applied to different problems [7].

   This method can be straightforwardly adapted to our model. In this formulation each antigen is a random subspace of the original space of variables.

3. Each antigen represents a random non-linear projection. A similar method has been used in [8] for ensemble construction. This non-linear projection is constructed using a random sample with replacement from the population. The idea is to have different non-linear projections that favour the classification of different subsets of patterns.

The affinity measure defines the relationship between antigens and antibodies. In our model it must measure if the antibody is able to solve the problem that the antigen represents. We have selected the most simple and straightforward measure: the affinity measure between an antibody, Ab_i, (a classifier),

and an antigen, Ag_j, (a view of the training data), $\phi(Ab_i, Ag_k)$, is the number of patterns that are correctly classified by the antibody with the view of the dataset represented by the antigen.

Immune Systems rely on hypermutation as the only operator to obtain new solutions. So, this operator must be able to create many different networks. We will consider that hypermutation consists of performing both parametric and structural mutation. Each of them will be performed with a certain probability. Structural mutation consists of the addition or deletion of a connection or node. Parametric mutation consists of the addition to every connection weight of a small quantity normally distributed, $r \in N(0, \sigma)$, where $\sigma = 1$ in our experiments.

3.1 Stimulation value

In the framework of immune network theory each antibody is represented by a *paratope* (π) that recognises other molecules, and an *idiotope* (ι) that is recognised by other molecules. The idiotope is termed *epitope* (ϵ) within the framework of the clonal selection theory. In order to obtain the idiotope of an antibody, Ab^{ι}, we set in turn every input to 1 and the rest of the inputs to 0, and get the output of the antibody.

To measure to which extent one antibody recognises other antibody we compute the Euclidean distance between their idiotopes, $\|Ab_i^{\iota} - Ab_j^{\iota}\|$. The similarity measure, $\tau(Ab_i, Ab_j)$, is given by:

$$\tau(Ab_i, Ab_j) = 1 - \frac{\|Ab_i^{\iota} - Ab_j^{\iota}\|}{D_{max}}, \tag{1}$$

where D_{max} is the maximum Euclidean distance between two idiotopes.

The key aspect of our model is the stimulation level achieved by each antibody within the immune network. This value must reflect the interaction of the antibody with the antigen population and the rest of antibodies.

The stimulation value is composed by two terms. The first one, aff_i, measures the affinity of the antibody to the population of antigens:

$$\text{aff}_i = \sum_{\forall k \in P(Ag): \phi(Ab_i, Ag_k) > \epsilon} \phi(Ab_i, Ag_k), \tag{2}$$

where $\phi(Ab_i, Ag_k)$ is the affinity value between antibody Ab_i and antigen Ag_k, $P(Ag)$ is the population of antigens, and ϵ is a recognition threshold ($\epsilon = 0.5$ in our experiments).

A second term, net_i, measures the interaction between the antibody and the rest of the antibodies of the population. This is based on two premises:

1. Each antibody only interacts with the antibodies that recognize the same antigens.

2. The strength of the interaction is proportional to the level of self-recognition of the antibodies.

In this way an *"immune subnetwork"* is created by each antigen. The antibodies that recognize the antigen above a certain threshold ϵ belong to the subnetwork. Each antibody of the subnetwork interacts with the rest of the antibodies of the subnetwork. The given function $\tau(Ab_i, Ab_j)$ measures the level of self-recognition among two antibodies. From each immune subnetwork to which the antibody belongs, the antibody receives a level of stimulation. The value of this stimulation for the subnetwork of Ag_j antigen, $\text{net}(Ab_i, Ag_j)$ is given by:

$$\text{net}(Ab_i, Ag_j) = \sum_{k \in \text{subnet}(Ag_j)} \tau(Ab_i, Ab_k)(\phi(Ab_i, Ag_j) - \phi(Ab_k, Ag_j)). \quad (3)$$

The value of stimulation from the interaction is given by the addition of stimulation within each subnetwork to which the antibody belongs:

$$\text{net}_i = \sum_{\forall j: i \in \text{subnet}(Ag_j)} \text{net}(Ab_i, Ag_j). \quad (4)$$

The underlying idea is that an antibody must recognize antigens that are not recognized by other antibodies. Otherwise, its level of recognition must be high, in order not to receive negative stimulation.

The stimulation value of antibody Ab_i, s_i, is given by $s_i = \text{aff}_i + \text{net}_i$

4 Experiments

In order to test the performance of our model on solving classification tasks, we need to compare it with other widely used ensemble methods. We have made the experiments with five standard methods for creating ensembles of classifiers. For testing the validity of the proposed model we have selected 35 datasets from the UCI Machine Learning Repository.

We perform a single significance test for every pair of algorithms. This test is a sign test on the win/draw/loss record of the two algorithms across all datasets. Our base learner is a MLP neural network trained using a standard back-propagation algorithm.

Table 1 shows the results in terms of test error for the five standard methods and the three immune networks. Table 2 shows the comparison of the different models as explained above. In the table the win/draw/loss (column against row) record is labelled s. The row labelled p is the result of the two-tailed sign test on the win-loss record.

The table also presents the mean of errors across all datasets and the *geometric mean error ratio* of every pair of algorithms. The row labelled $\dot{r}$ shows the geometric mean of the error ratio *column/row*. A value below 1 indicates a general advantage of the algorithm corresponding to the column to the algorithm corresponding to the row.

The table shows that the best standard method based on resampling is AdaBoost, although the difference is not significant for all the problems. Random

Table 1: Summary of test errors for the standard and the immune methods.

Dataset	Standard ensemble methods					Immune Network		
	None	Bagging	Arc-x4	AdaBoost	Sub-std	Sampling	Subspace	NLP
Audio	0.2137	0.2060	0.1554	0.1304	0.2036	0.1500	0.1572	0.3214
Autos	0.2301	0.2111	0.2503	0.2307	0.1883	0.2118	0.2392	0.1373
Balance	0.1167	0.1165	0.0190	0.0611	0.0788	0.0128	0.0282	0.0641
Breast-c	0.2981	0.3023	0.2953	0.2859	0.2944	0.2902	0.2845	0.2817
Card	0.1293	0.1252	0.1364	0.1312	0.1308	0.1302	0.1302	0.1070
Derma	0.0355	0.0337	0.0370	0.0366	0.0456	0.0355	0.0289	0.0111
Ecoli	0.1202	0.1338	0.2222	0.2048	0.1417	0.1355	0.1309	0.0952
Gene	0.1108	0.1098	0.1006	0.1052	0.0849	0.0946	0.0905	0.1097
German	0.2493	0.2487	0.2537	0.2524	0.2476	0.2484	0.2560	0.2520
Glass	0.2862	0.2868	0.2862	0.2679	0.2717	0.2566	0.2528	0.2679
Glass-g2	0.1950	0.1983	0.1650	0.1858	0.1675	0.1600	0.1600	0.1000
Heart	0.1461	0.1333	0.1490	0.1348	0.1265	0.1176	0.1117	0.1265
Heart-c	0.1645	0.1346	0.1623	0.1566	0.1197	0.1079	0.1000	0.1053
Hepatitis	0.1193	0.1175	0.1447	0.1412	0.1290	0.1263	0.1158	0.1053
Horse	0.3007	0.2857	0.2828	0.2722	0.2758	0.2572	0.2857	0.2088
Hypo	0.0427	0.0422	0.0255	0.0258	0.0304	0.0290	0.0286	0.0467
Ionos	0.0747	0.0743	0.0816	0.0805	0.0874	0.0851	0.0936	0.1034
Labor	0.0714	0.0714	0.0714	0.0714	0.0785	0.0714	0.0714	0.0714
Liver	0.3089	0.3023	0.3267	0.3008	0.2756	0.2752	0.2954	0.2733
Lymph	0.1378	0.1477	0.1495	0.1441	0.1297	0.1351	0.1135	0.1351
Page-bk	0.0388	0.0389	0.0392	0.0469	0.0434	0.0471	0.0447	0.0556
Pima	0.2014	0.2043	0.2359	0.2000	0.2083	0.2008	0.2000	0.2083
Post-op	0.2864	0.2955	0.3682	0.3197	0.2818	0.2662	0.2727	0.2273
Primary	0.5385	0.5321	0.5242	0.5321	0.5155	0.5258	0.5286	0.6071
Promoters	0.2308	0.2308	0.2308	0.2308	0.1231	0.1692	0.1538	0.1154
Satimage	0.1399	0.1401	0.1187	0.1263	0.1280	0.1298	0.1286	0.1418
Segment	0.0798	0.0786	0.0546	0.0666	0.0581	0.0648	0.0548	0.0312
Sick	0.0390	0.0383	0.0299	0.0346	0.0401	0.0301	0.0305	0.0626
Sonar	0.1788	0.1766	0.1779	0.1788	0.1702	0.1519	0.1615	0.1923
Soybean	0.0692	0.0710	0.0671	0.0667	0.0988	0.0859	0.0706	0.1118
Vehicle	0.1995	0.1976	0.1940	0.1779	0.1630	0.1619	0.1640	0.1564
Vote	0.0574	0.0593	0.0614	0.0660	0.0685	0.0630	0.0667	0.0741
Vowel	0.5616	0.5622	0.4861	0.5307	0.4286	0.4649	0.4818	0.4481
Waveform	0.1178	0.1195	0.1238	0.1156	0.1088	0.1174	0.1139	0.1352
Yeast	0.3950	0.4022	0.5102	0.4046	0.4170	0.4102	0.4135	0.4151

Subspace shows the best performance among the standard methods, with a behaviour slightly better than AdaBoost. Regarding the immune ensembles, the model based on wagging is able to improve the performance of the five standard models, with a difference that is statistically significant. Subspace based antigens also obtain good results. They are able to improve the results of the standard methods, with the exception of arcing and random subspace method. Furthermore, the geometric mean-error ratio is below 0.93 for all the standard methods.

The results of the antigens based on non-linear projections are very interesting. On the one hand, for many problems the results are very good, improving the performance of the standard methods very significantly. On the other hand, the performance in some other problems is very poor. The source of this variance of the results is not clear and we are performing further experiments investigating how this kind of antigen behaves.

5 Conclusions

In this paper we have shown how Immune Networks can be used for evolving ensembles of neural networks. The obtained ensembles have a very good performance in terms of generalisation error. These ensembles outperform significantly ensembles obtained with widely used standard methods. These results are relevant as the benchmark used includes many different real-world problems.

These results are promising enough to justify further development of the

Table 2: Comparison of the five standard methods and the three immune ensembles. Win/draw/loss record (row s) of the algorithms against each other and p-value of the sign test (row p), and the geometric mean of the error ratio (row $\dot{r}$).

							Immune Network		
Algorithm		None	Wagging	Arc-x4	AdaBoost	Sub-std	Sample	Subsp.	NLP
$\bar{\epsilon}$		0.1867	0.1837	0.1868	0.1805	0.1703	0.1663	0.1674	0.1687
None	s		19/2/14	18/3/14	22/3/10	25/0/10	28/1/6	27/0/8	21/0/14
	p		0.4869	0.5966	0.0501	0.0167	0.0002	0.0019	0.3105
	$\dot{r}$		0.9801	0.9346	0.9505	0.9247	0.8549	0.8598	0.8770
Wagging	s			15/2/18	18/3/14	23/0/12	25/0/10	28/0/7	21/0/14
	p			0.7283	0.5966	0.0895	0.0167	0.0005	0.3105
	$\dot{r}$			0.9536	0.9697	0.9434	0.8722	0.8772	0.8948
Arc-x4	s				21/2/12	22/0/13	26/0/9	22/0/13	22/0/13
	p				0.1628	0.1755	0.0060	0.1755	0.1755
	$\dot{r}$				1.0169	0.9893	0.9147	0.9199	0.9383
AdaBoost	s					20/0/15	25/0/10	25/0/10	19/0/16
	p					0.4996	0.0167	0.0167	0.7359
	$\dot{r}$					0.9729	0.8995	0.9046	0.9227
Sub-std	s						24/0/11	22/0/13	19/1/15
	p						0.0410	0.1755	0.6076
	$\dot{r}$						0.9245	0.9298	0.9484
Sampling	s							17/3/15	15/2/18
	p							0.8601	0.7283
	$\dot{r}$							1.0057	1.0258
Subspace	s								15/1/19
	p								0.6076
	$\dot{r}$								1.0200

application, making use of many features of the natural immune network theory that have not been considered in this initial model.

Our current research is focused on two different ideas. Firstly, we are working in a new definition of the stimulation function that takes into account the diversity of the antibodies more explicitly. Secondly, in the present definition the population of antigens is static, nevertheless, natural pathogens do mutate. Using this feature we are developing a population of antigens subject to modification as the population of antibodies develops.

References

[1] M. P. Perrone and L. N. Cooper. When networks disagree: Ensemble methods for hybrid neural networks. In R. J. Mammone, editor, *Neural Networks for Speech and Image Processing*, pages 126–142. Chapman – Hall, 1993.

[2] S. Dzeroski and B. Zenko. Is combining classifiers with stacking better than selecting the best one? *Machine Learning*, 54:255–273, 2004.

[3] D. Dasgupta. *Artificial Immune Systems and their Applications*. Springer-Verlag, 2000.

[4] K. N. Jerne. Towards a network theory of the immune system. *Annals of Immunology*, 125C:373–389, 1974.

[5] E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36(1/2):105–142, July/August 1999.

[6] T. K. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

[7] R. Munro, D. Ler, and J. Patrick. Meta-learning orthographic and contextual models for language independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning*, 2003.

[8] N. García-Pedrajas and C. Fyfe. Nonlinear *"boosting"* projections for ensemble construction. *Journal of Machine Learning Research*, 2005. (submitted).