Class Majority in Designing Fuzzy Local Approximation NNs for Overlapping Data in Pattern Classification

M.R. Widyanto, K. Kawamoto, and K. Hirota

Department of Computational Intelligence & Systems Science, Tokyo Institute of Technology
G3-49, 4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan
e-mail: {widyanto,kawa, hirota}@hrt.dis.titech.ac.jp

***Abstract* – To deal with a problem of overlapping data in pattern classification, a class majority concept in designing hidden units of Fuzzy Local Approximation NNs, that is the Fuzzy Similarity based Self-Organized Network inspired by Immune Algorithm (F-SONIA), is proposed. For each cluster formed by the adaptive clustering of F-SONIA, number of vectors that belong to the same class is calculated. Therefore fractions of all classes are known. One class should have a single class majority in the cluster. Then, clusters with no single majority are broken down into two clusters. In experiments, the proposed method is compared with the standard F-SONIA and the well-known Back Propagation NNs. Experiments show that the proposed method improves classification performance of vowel classification by 5 %.**

I. INTRODUCTION

Neural networks (NNs) have been widely studied and applied in various areas. Haykin [1] divides neural networks types into global approximation and local approximation NNs. In global approximation NNs, responses to input vectors are given globally as multiplication of hidden unit weights and the value of inputs. For pattern classification, however, it makes the networks difficult to find relation between input vectors and pattern characteristics. Back-Propagation NNs [2] and Recurrent Networks [3] are belong to global approximation NNs.

Local approximation NNs defines the relation between input vectors and hidden unit weights by distance between those two vectors. This mechanism enables the networks to more easily find relation between input vectors and hidden unit weights. LVQ [4,5], RBF [6], SONIA [7] uses Euclidean distance to define similarity relation between input vectors and hidden unit weights, while F-SONIA [8] uses fuzzy similarity to define the relation. Widyanto et al. [8] shows that the use of fuzzy similarity rather than Euclidean distance enables the networks to easily find the relation between input vectors and pattern characteristics, and therefore the recognition accuracy is improved and the learning time is reduced.

In F-SONIA, however, if an input vector is grouped to a cluster with vectors that belong to a different class, the input vector will be incorrectly classified by the output layer since the output layer weights are adjusted to those majority vectors with different class. In overlapping data, this condition is unavoidable, so that the task is to minimize the condition. To improve the classification performance of F-SONIA in overlapping data by minimizing the above condition, a class majority concept of hidden unit initialization of F-SONIA is proposed.

In Sec. 2, the hidden unit initialization of F-SONIA and the problem with overlapping data is mentioned. In Sec. 3, the class majority concept in designing F-SONIA is proposed. Sec. 4 describes the experiments with vowel classification and IRIS data.

II. OVERLAPPING DATA IN F-SONIA

The hidden unit initialization of F-SONIA [8] is similar to that of SONIA [7] where an adaptive clustering method inspired by immune algorithm [10] is applied to form clusters as basis for hidden units. The adaptive clustering method is summarized as follows.

Let $\boldsymbol{s}_m$ ($m = 1,\ldots,M$, $M \in \mathbb{N}$) be a input to be learned. In initialization process, the first hidden neuron(t_1 , $\boldsymbol{w}_{H1}$) is created with $t_1 = 0$ and $\boldsymbol{w}_{H1}$ is taken arbitrarily from available input vectors. Variable t_1 shows the number of input vectors associated with the first hidden unit and $\boldsymbol{w}_{H1}$ represents the weight vector of the first hidden unit.

For $m = 1$ with respect to M , repeat the below procedure until all input vectors associated with a corresponding hidden unit.

1. For j = 1 to N_H (number of hidden unit), calculate the distance between the *m-th* input and weight vector of the *j-th* hidden unit

$$d_{mj} = \sqrt{\sum_{i=1}^{N_I} \left(s_{mi} - w_{Hij} \right)^2} \qquad (\in \mathbb{R}_+), \tag{1}$$

where s_{mi} is the *i-th* component of vector $\boldsymbol{s}_m$, w_{Hij} is the *i-th* component of vector $\boldsymbol{w}_{Hj}$, and N_I is the number of input dimension.

2. Select the closest distance

$$c = \text{argmin}_j \, d_{mj} \quad (\in \mathbb{N}). \tag{2}$$

3.a. If the shortest distance d_{mc} is below a stimulation level $sl \in (0,1)$ then update the followings

$$t_c = t_c + 1, \tag{3}$$

$$\boldsymbol{w}_{Hj} = \boldsymbol{w}_{Hj} + h \cdot d_{mc}, \tag{4}$$

where $h \in (0,1)$ is a learning rate.

b. If the shortest distance d_{mc} is bigger than stimulation level then update the following

$$N_H = N_H + 1, \tag{5}$$

and create a new hidden unit ($t_{N_H}, \boldsymbol{w}_{HN_H}$) with $t_{N_H} = 0$, and $\boldsymbol{w}_{HN_H} = \boldsymbol{s}_m$. After that, reset the value of t_k $(k = 1,..., N_H)$ to 0, the value of m to 1, and repeat the procedure.

In the above adaptive clustering procedure, the value of stimulation level should be decided by trial and error process. The other disadvantage is that, if an input vector is grouped to a cluster with vectors that belong to a different class, then the input vector will be classified incorrectly by the output layer. This is because the output weights are adjusted to those majority vectors with different class. In overlapping data, this condition is unavoidable, so that the task is to minimize the condition.

If numbers of two different classes are almost the same then the output layer, which is trained by the Back-Propagation algorithm [2], will be difficult to adjust the output weights related to the hidden unit. This condition is illustrated by Fig. 1.

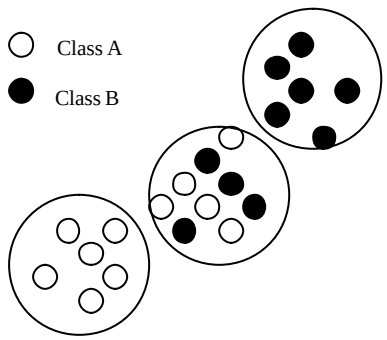

Figure 1. An example result of Adaptive Clustering in F-SONIA

In Fig. 1 the adaptive clustering procedure gives three clusters. In the middle cluster, vectors that belong to class A and class B are grouped together. This condition confuses the learning process of output weights whether to adjust the weights to class A or class B. Frequently this condition falls into incorrect classification.

III. Class Majority Concept

In order to minimize the condition describe in Sec. 2 which occurs in overlapping data, a class majority concept of hidden unit initialization of F-SONIA is proposed. Each cluster in F-SONIA should have a single class majority. It means that a number of vectors that belongs to a certain class should be big enough comparing to vectors that belong to different classes.

Ideally, a cluster in F-SONIA should consist of members that belong to one class. However this condition is almost impossible for overlapping data. If the clusters are forced to only consist of one class, it makes size of the cluster becomes very small. Consequently, there will be a huge number of clusters and the hidden units. It makes overheads in computation and the networks tend to be over fitted. So the purpose of the class majority concept is to minimize the errors while still maintaining the generalization of the networks.

The class majority in hidden unit initialization of F-SONIA is described as below

1. Calculate the centers of vectors that belong to the same class. For k = 1 to K (number of classes)

$$\boldsymbol{w}_k = \left(\sum_{l=1}^{B_k} \boldsymbol{s}_l \right) / B_k. \tag{5}$$

where B_k is the number of vectors that belong to class k.

2. Select the closest distance between those centers to be the stimulation level (sl)

$$sl = \min \left\| w_m - w_n \right\| \quad (m, n \in 1..K, m \neq n) \; (\in \mathbb{N}). \tag{6}$$

3. Use the centers as initial hidden unit for the adaptive clustering (t_k, $\boldsymbol{w}_k$) $(k = 1..K)$ where t_k = 0.
4. Apply the adaptive clustering (in Sec.2) using stimulation level and initial hidden units as given in step 2 & 3.
5. For each cluster formed by the adaptive clustering procedure, number of vectors that belongs to the same class are calculated. Therefore fractions of class in the cluster are known. For j = 1 to N_H, for k = 1 to K

$$f_{jk} = B_{jk} / t_j \quad (\in R_+), \tag{7}$$

where f_{jk} fraction of vectors belong to class k in cluster j, B_{jk} is number of vectors belong to class k in cluster j, and t_j shows the number of vectors belong to cluster j.

6. Calculate the biggest fraction in the clusters. For j = 1 to N_H,

$$f_j = \max f_{jk} \quad (k=1...K) \, (\in \mathbb{R}_+). \tag{8}$$

7. Clusters whose biggest fraction below a class majority threshold (mt) are broken down by the following procedure. For $j = 1$ to N_H, if $f_j < mt$ then

a. Calculate the center of vectors that belong to the majority class

$$e = \arg\max f_{jk} \quad (k=1..K)(\in \mathbb{N}), \tag{9}$$

$$w_j = \left(\sum_{n=1}^{B_{je}} s_n \right) / t_j, \tag{10}$$

where e is an index of class that has the biggest fraction, B_{je} is number of vectors that belong to class e in cluster j, and t_j shows the number of vectors belong to cluster j.

b. Calculate the center of other vectors that do not belong to the majority class. Add this center as new hidden unit

$$N_H = N_H + 1, \tag{11}$$

$$w_{N_H} = \left(\sum_{k=1}^{K} \sum_{b=1}^{B_{jk}} s_b \right) / t_j \quad (k \neq e). \tag{12}$$

where B_{jk} is number of vectors that belong to class k in cluster j.

After that, the hidden units are ready to be used for training process in which Back-Propagation algorithm [2] is used to train the output layer, and the fuzzy similarity [9] is used to measure the similarity relation between input vectors and hidden unit weights. Corresponding to Fig.1, the result of the class majority hidden unit initialization of F-SONIA is given by Fig. 2.

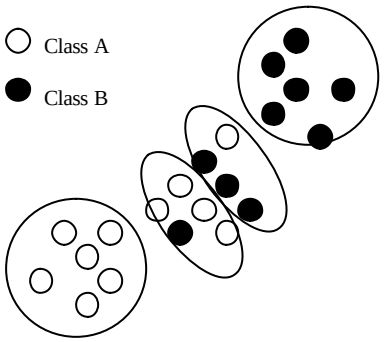

Figure 2. Example results of Class Majority Initialization

In Fig. 2, the proposed class majority hidden unit initialization creates two smaller clusters in which a single majority exists for each cluster. This minimizes the errors resulted by incorrectly grouped vectors as described in Sec. 2. The training process of output layer is easier to adapt the weights related to a hidden unit to a single class. It will minimize the errors, thus improving the classification performance. The other advantage is that there is no more need to set the value of stimulation level by trial and error process. The above procedure decides the appropriate stimulation level parameter based on the closest distance between class centers.

IV. EXPERIMENTS ON VOWEL & IRIS DATA

To test classification performance of the proposed method, a benchmark vowel recognition data [11] is used in experiment. The vowel recognition data is a set of two-dimensional (2-D) vowel data formed by computing the first two formants F1 and F2 from samples of ten vowels spoken by 67 speakers. The vowels are head, hid, hod, had, hawed, heard, heed, hud, who`d, hood. The data set is used because there is significant overlapping between vectors corresponding to different vowel plane (see Fig. 3). The data consists of a training set, containing 338 vectors, and a testing set, containing 333 vectors.

The performance of the proposed class majority F-SONIA is compared with conventional F-SONIA [8] and BP NNs [2]. For every data set, performances of those methods were observed through learning phase for 300 iterations and testing phase. In learning phase, the learning time required is recorded, meanwhile in testing phase the recognition accuracy is measured by using training and testing data as inputs. For each experiment, the learning & testing phases are carried out for 10 times, and the classification accuracy is the average of these trials. The experiments have been done using Matlab 6.1 [12] C/C++ compiler under Microsoft Windows XP operating system on Pentium 4 2.2 GHz processor with 256 Megabytes memory. For BP NNs, Matlab neural network toolbox is used.

The proposed class majority F-SONIA with majority threshold 0.6 gives 45 hidden units, and the stimulation level parameter of the conventional F-SONIA was set to give comparable number of hidden units which is 44 hidden units. For a balance comparison, the hidden neuron number of BP NNs is set to be 45 and other parameters are set as advised by the Matlab neural network toolbox [12]. The fuzziness of F-SONIA and class majority F-SONIA is set to be 0.04. The class majority F-SONIA improves classification accuracy for 5% (see Table 1). Meanwhile the average learning time of Class Majority SONIA is almost the same with the standard SONIA. This shows that the proposed method improves the classification accuracy of F-SONIA for overlapping data. Fig.3. also shows the classification plane resulted by class majority F-SONIA.

TABLE I. VOWEL CLASSIFICATION RESULTS

Method	Training (%)	Testing (%)	Learning Time (seconds)
Class Majority F-SONIA	80.18	79.28	67.83
Conventional F-SONIA	74.56	73.87	67.77
BP NNs	64.50	64.56	6.65

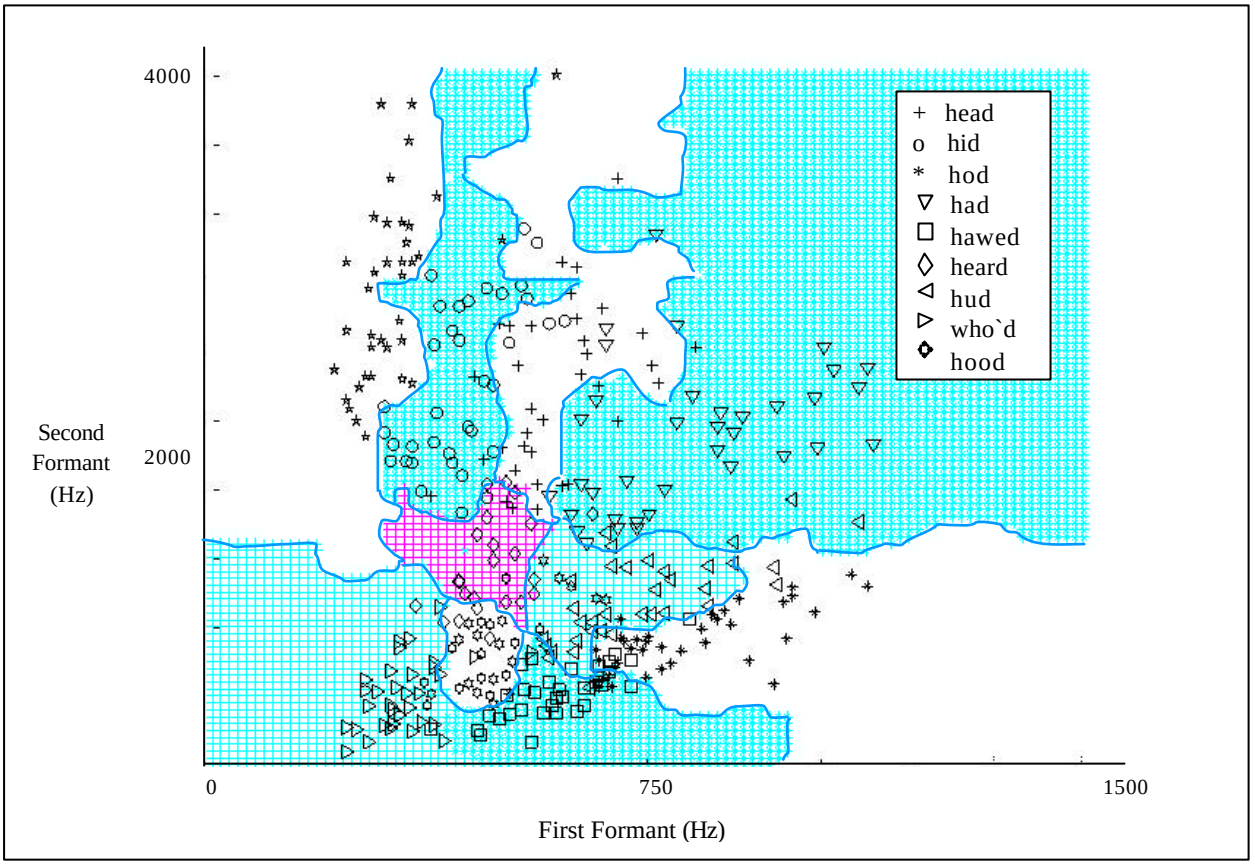

Figure 3. Overlapping in Vowel Recognition Data & Classification Plane with Class Majority

To see how the classification performance of class majority F-SONIA works for separable data that do not have many overlapping data, the proposed method is tested with IRIS benchmark data [13]. The database was created by Fisher in 1936 and has been widely used in subsequent research in pattern classification. The data set contains three classes of 50 instances each, where each classes refers to a type of Iris plant. The three classes: Iris sentosa, Iris versicolor, and Iris virginica. There are four numeric attributes and no missing value. The four attributes are: sepal length in cm., sepal width in cm., petal length in cm., and petal width in cm. (cm. = centimeter). One of the three classes is well separated from the other two, which are not easily separable. The 150 examples contained in the IRIS data set were randomly split to form the training and testing sets, each containing 75 examples. Table 2 is a brief statistical analysis of the sample attributes.

TABLE II. STATISTICAL ANALYSIS OF IRIS PLANT DATA

Attribute	Min	Max	Mean	Standard Deviation
Sepal length	4.3	7.9	5.84	0.83
Sepal width	2.0	4.4	3.05	0.43
Petal length	1.0	6.9	3.76	1.76
Petal width	0.1	2.5	1.20	0.76

Statistical analysis of the Iris plants sample attributes (in centimeter).

Like previous experiment, the performance of the proposed class majority F-SONIA is compared with conventional F-SONIA [8] and BP NNs [2]. All the experiment settings are the same with previous experiment with vowel data. The proposed class majority F-SONIA gives 7 hidden units, and the stimulation level parameter of the conventional F-SONIA is set to give comparable number of hidden units. For a balance comparison, the hidden neuron number of BP NNs is set to be 7 and other parameters are set as advised by the Matlab neural network toolbox [12].

Table 3 shows the classification accuracy of class majority F-SONIA, conventional F-SONIA, and BP NNs. It can be seen that the class majority F-SONIA does not significantly improves the classification accuracy since IRIS data is more separable comparing to vowel data and do not have many overlapping feature vectors between classes.

TABLE III. IRIS PLANT CLASSIFICATION RESULTS

Method	Training (%)	Testing (%)	Learning Time (seconds)
Class Majority F-SONIA	97.33	96.00	12.83
Conventional F-SONIA	96.00	94.67	12.78
BP NNs	86.67	88.00	1.25

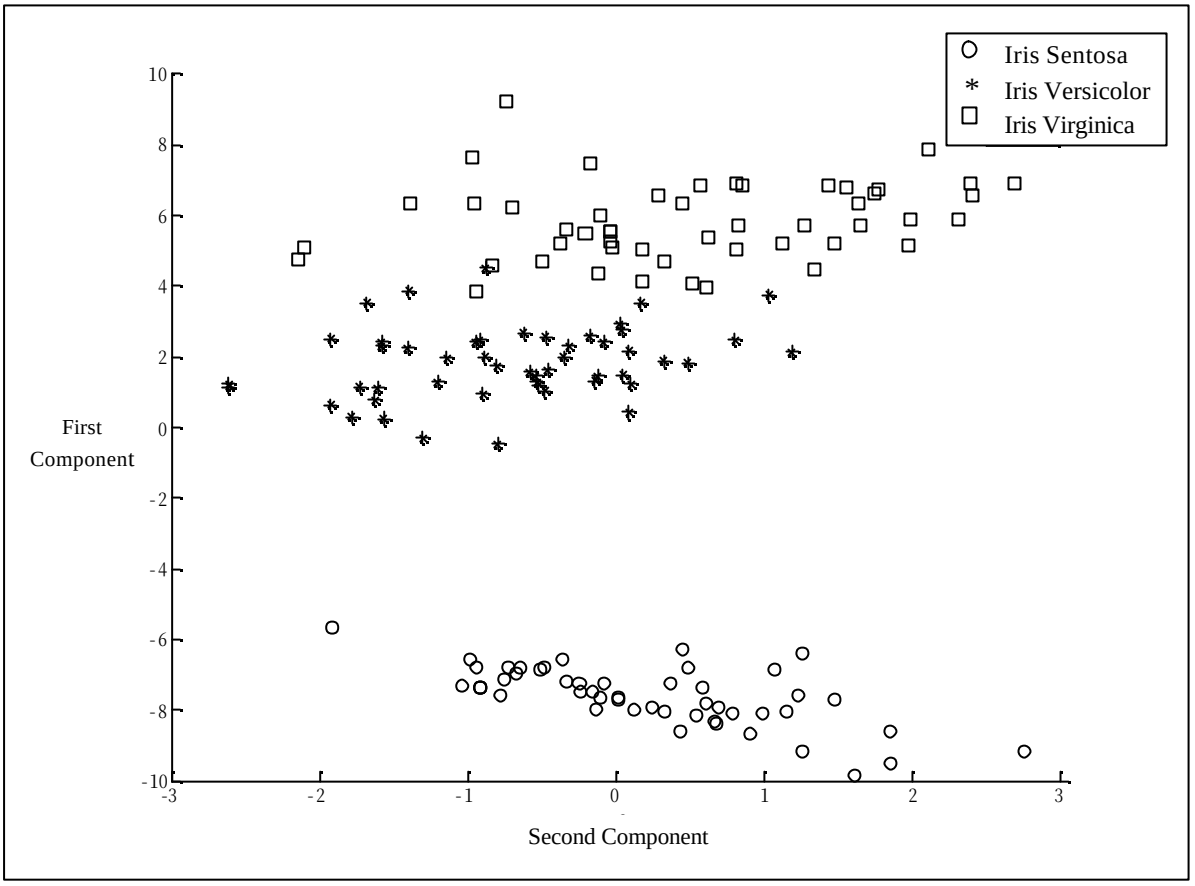

Figure 4. Two dimensional IRIS Plant Data

To see how separable the IRIS plant data, the Matlab one-way Multivariate Analysis of Variance (MANOVA) [12] is used to reduce the four dimensional data into two dimensional data (see Fig. 4). MANOVA tool works as dimensional reduction of data so that it is easy to see the separable condition of IRIS plant data. Small number of vectors from Iris versicolor and Iris virginica overlap each other meanwhile Iris sentosa is very separable to the other class. This condition is the reason why the proposed class majority F-SONIA did not improve classification performance significantly comparing to conventional F-SONIA.

V. CONCLUSIONS

A class majority hidden unit initialization of a Fuzzy Similarity based Self-Organized Network inspired by Immune Algorithm (F-SONIA) is proposed. Each cluster in F-SONIA should have a single class majority. This mechanism minimizes the errors resulted by incorrectly grouped vectors. The training process of output layer is easier to adapt the weights related to a hidden unit to a single class. It will minimize the errors therefore improve the classification performance.

From the experiments, the proposed method improves the classification accuracy of the conventional F-SONIA by 5% for vowel recognition which contains overlapping data. This shows that the introduction of class majority concept minimizes the errors resulted by incorrectly grouped vectors. The training process of output layer is more easily adapt the weights related to a hidden unit to a single class therefore the classification accuracy is improved.

For the IRIS plant data, the class majority F-SONIA does not significantly improve the classification accuracy because the IRIS plant data is more separable comparing the vowel recognition data. For data that contain overlaps e.g. voice & ultrasonic data, the proposed class majority F-SONIA is very promising to be implemented.

REFERENCES

[1] S. Haykin, *Neural Networks*, Prentice Hall, 1999.

[2] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by error propagation”, Nature, (323), pp. 533-536, 1986.

[3] Elman, J. L., "Finding structure in time," *Cognitive Science*, vol. 14, 1990, pp. 179-211.

[4] T. Kohonen, “Learning vector quantization for pattern recognition”, *Technical Report TKK-F-A601*, Helsinki University of Technology, Finland, 1986.

[5] J. Kangas, and T. Kohonen, “Developments and applications of the self-organizing map and related algorithms,” *Elsevier Mathematics and Computers in Simulation*, (41), pp. 3-12, 1996.

[6] J. E. Moody and C. J. Darken, “Fast learning in networks of locally-tuned processing units”, *Neural Computation*, vol. 1, pp. 281-294, 1989.

[7] M. R. Widyanto, “A self-organized network inspired by immune algorithm”, *Master Theses*, Tokyo Institute of Technology, Tokyo, Japan, 2003.

[8] M. R. Widyanto et. al., “Improvement of three mixture fragrance recognition using fuzzy similarity based self-organized network inspired by immune algorithm, in *Proceeding of International Symposium on Intelligent Systems*, Korea, 2003.

[9] L. A. Zadeh, "Similarity relations and fuzzy orderings", *Information Sciences*, (3), pp. 177-200, 1971.

[10] J. I. Timmis, Artificial immune systems : a novel data analysis technique inspired by the immune network theory, *PhD. Dissertation*, University of Wales, Aberystwyth, 2001.

[11] R. P. Lippmann, “Pattern classification using neural networks,” *IEEE Community Magazine*, vol. 27, pp. 47-54, 1989.

[12] H. Demuth, and M. Beale, "Neural networks toolbox", *Matlab 6.1 User Guide*, Mathworks, 2001.

[13] E. Anderson, “The IRISes of the Gaspe peninsula,” *Bull. America IRIS Soc.*, vol. 59, pp. 2-5, 1939.