Evaluation of dependence on cluster representation in ARTMAP classifiers

Norbert Kop o
Dept. of Cognitive and Neural Systems
Boston University
677 Beacon St., Boston, MA 02215
And
CIG, KKUI, FEI TU Košice
Letná 9, 04001 Košice, Slovak Rep.
kopco@bu.edu

Peter Sin ák
Computational Intelligence Group
Department of Cybernetics and AI
Faculty of EE and Informatics
Technical University
Letná 9, 04001 Košice
Slovak Republic
sincak@tuke.sk

Štefan Molnár
Department of Cybernetics and AI
Faculty of EE and Informatics
Technical University
Letná 9, 04001 Košice
Slovak Republic
molnars@neuron-ai.tuke.sk

***Abstract* – This paper presents an analysis of performance of several types of the ARTMAP neural network. The performance of the networks is analyzed on the task of classification of satellite images obtained by remote sensing of the Earth. The analysis is concentrated on the dependence of classification accuracy on the difference in cluster type preferably identified by each of the classifiers. Three types of ARTMAP classifier are compared: fuzzy ARTMAP, Gaussian ARTMAP, and Extended Gaussian ARTMAP. The main difference among these classifiers is in the way they determine/represent individual clusters in feature space. Best results are obtained for Extended Gaussian ARTMAP, a modification of the Gaussian ARTMAP neural network that preferably identifies Gaussian-distributed clusters.**

I. INTRODUCTION

During the past several years, satellite remote sensing of the Earth has become one of the main sources of data for many geographical applications, e.g., land-use and vegetation maps generation [1]. On one hand, this development causes that there are more and more high-quality images easily available for all kinds of users and applications. On the other hand, to be able to process the huge amounts of currently available data with appropriate quality of performance, there is a need for highly accurate, automated, and preferably autonomous systems for image data analysis and processing. Traditional approaches to this task are based on statistical classification and pattern analysis methods, the most important being the Maximum Likelihood classifier [2]; and on various rule-based algorithms [3]. The statistical methods have several useful properties, the most important one being their optimal behavior if several assumptions are fulfilled. The disadvantages of these methods include their poor performance if the assumptions are not fulfilled, as well as their large complexity (in terms of training time and memory requirements).

Recently, several alternative approaches have been proposed to the problem of classification of remotely sensed images. Among these approaches, artificial neural networks play a significant role [4], and among them the Error Back-propagation algorithm [5] is the predominant method. There are also several reports of application of the *Adaptive Resonance Theory (ART)* neural networks in this domain [6, 7, 8], most of them applying fuzzy ARTMAP

Fig. 1 Original image. Highlighted areas were classified by expert (A – urban area, B – barren fields, C – bushes, D – agricultural fields, E – meadows, F – forests, G – water)

neural network. In [7], the authors report results of comparative analysis of ARTMAP, fuzzy ARTMAP, and Gaussian ARTMAP neural networks, when applied to the classification of remotely sensed imagery. The present article is a continuation of that study, introducing a new class of ARTMAP neural networks, called Extended Gaussian ARTMAP networks, and comparing the performance of this neural network to the performance of the previously analyzed systems. In this comparison the goal is to answer the question of "What is the best assumption about distribution of remotely sensed data," distinguishing between hyper-rectangle data clusters preferred by fuzzy ARTMAP, zero-covariance Gaussian-distributed data preferred by Gaussian ARTMAP, and arbitrary Gaussian distributions preferred by Extended Gaussian ARTMAP.

TABLE I
CLASSES DEFINED IN THE IMAGE

Class	Label
Urban Area	A
Barren Fields	B
Bushes	C
Agricultural Fields	D
Meadows	E
Forests	F
Water	G

The data for this study come from a Landsat Thematic Mapper image of the city of Košice in Eastern Slovakia. The whole image consists of 368,125 7-dimensional pixels, out of which an expert assigned 6,331 pixels into seven thematic categories, shown in TABLE I. The regions of the image included into the training and testing data set are shown in Fig. 1.

II. ARTMAP NEURAL NETWORKS

ARTMAP neural networks belong to the class of neural networks called Adaptive Resonance Theory (ART), a theory of cognitive information processing in human brain [9]. Based on this theory, a whole family of neural network algorithms was developed. These neural networks were shown to give a very good performance in applications involving clustering, classification, and pattern recognition. When compared to statistical and other neural-network-based clustering/classification algorithms, these networks usually obtain very good classification accuracy, while securing proven stability and high level of compression in the system.

An overview of the process of development of ARTMAP neural networks can be found in [8]. From the point of view of this study, the currently available ARTMAP classification systems can be divided into two groups. First, systems based on (or systems that are a modification of) fuzzy ARTMAP algorithm (e.g., ARTMAP-IC, ART-EMAP, etc., see [10]). All these systems share the property that they prefer data clusters distributed into hyper-rectangles in feature space. In these systems the basic properties of the original ARTMAP design (stability, proven convergence, fast on-line learning) are preserved, but they also have well-known disadvantages, e.g., noise sensitivity and tendency to category proliferation. The other group is based on the Gaussian ARTMAP neural network [11]. In this group of networks, preferably identifying Gaussian-shaped clusters, the stability and fast on-line learning properties of the fuzzy ARTMAP networks is traded for emphasis on ability of the system to generalize and for its decreased sensitivity to noise in the input data.

III. ANALYZED ARTMAP CLASSIFIERS

Structurally, every ARTMAP network (fuzzy ARTMAP or Gaussian ARTMAP) can be divided into two parts. The first part, represented by an ART module, dynamically generates units, each identifying a single data cluster in feature space. This part can be used autonomously for cluster analysis of a given data set. The second part serves to identify each of the clusters found in the data with one of the classes defined on the data set.

A detailed description of fuzzy ARTMAP (FA), first of the algorithms analyzed in this study, can be found in many previously published studies. For a description directly related to processing of data from remote sensing, the reader is referred to publications [6, 7, and 8]. From the point of view of this study, the most important property of this system is that the subsystem identifying clusters in feature space preferably identifies the clusters in which patterns are distributed as hyper-rectangles.

The second algorithm, Gaussian ARTMAP (GA), is described in detail, e.g., in [7 and 11]. Its main feature is that it preferably identifies clusters with Gaussian distribution, in which the co-variance (off-diagonal) coefficients in the co-variance matrix describing the cluster are fixed to zero. This restriction was imposed on the Gaussian ARTMAP system for computational purposes, the reason being that with this kind of representation each cluster-identifying node is described by $2*M+1$ parameters, where M is the dimensionality of feature space. This memory requirement is only slightly worse than memory requirements of fuzzy ARTMAP networks ($2*M$). But it is much lower than that of the Extended Gaussian ARTMAP (as described in the next paragraph).

IV. EXTENDED GAUSSIAN ARTMAP

The Extended Gaussian ARTMAP (EGA) was developed independently by the authors of this study, and in [12]. The main difference between this algorithm and the standard Gaussian ARTMAP algorithm is in the way the nodes identifying clusters in feature space are described. In contrast to GA, in EGA each cluster-identifying node j (also called a category) is described by a full Gaussian distribution. I.e., each category is defined by an M-dimensional vector μ_j, describing the mean value in each dimension; by a full MxM-dimensional covariance matrix Σ_j, and by a scalar number n_j, in which the number of patterns coded by a given node is stored. This last number is equivalent to the *a priori* probability of the given category. Thus, to represent an M-dimensional input I, each category requires M^2+M+1 components.

During the process of training, each new pattern is assigned to the cluster where it belongs with the highest probability. This *a posteriori* probability of category j given input I is defined as:

$$P(j \mid I) = \frac{p(I \mid j)P(j)}{p(I)}. \quad (1)$$

Each category is defined by a fully described (non-separable) Gaussian distribution, which includes the mean values μ_j and co-variance matrices Σ_j, so the conditional density of I given category j from Eq. 1 is defined as

$$p(I \mid j) = \frac{1}{(2\pi)^{\frac{M}{2}} \left|\Sigma_j\right|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\left(I - \mu_j\right)' \Sigma_j^{-1}\left(I - \mu_j\right)\right] \quad (2)$$

where Σ^{1} is inversion of the co-variance matrix and $|\,\Sigma\,|$ is the determinant of the co-variance matrix. The *a priori* probability of category j in Eq. 1 is

$$P(j) = \frac{n_j}{\sum_{j'=1}^{N} n_{j'}} \quad (3)$$

where N is the number of categories.

For each new training pattern the winning category is determined by first computing the Bayes discrimination function [13] for each category j, based on Eqs. 1-3:

$$g_j(I) = \log\left((2\pi)^{\frac{M}{2}} P(I \mid j)P(j)\right) = \\ = -\frac{1}{2}\left[\left(I - \mu_j\right)' \Sigma_j^{-1}\left(I - \mu_j\right)\right] - \frac{1}{2}\log\left|\Sigma_j\right| + \log(P(j)) \quad (4)$$

and then determining the non-reset category with the highest value of the discrimination function:

$$J = \arg\max_j \left(g_j(I)\right) \quad (5)$$

which represents the category to which a given pattern belongs with the highest probability. In all ART network there is a criterion of degree of *match* between a given input and the proposed category J. In EGA (similarly to GA) this *match* criterion is defined by first computing the measure of *match*:

$$g_j^{'}(I) = \log\left((2\pi)^{\frac{M}{2}} P(I \mid j)\right) = \\ = -\frac{1}{2}\left[\left(I - \mu_j\right)' \Sigma_j^{-1}\left(I - \mu_j\right)\right] - \frac{1}{2}\log\left|\Sigma_j\right| = \\ = g_J(I) - \log(P(J)). \quad (6)$$

and then comparing this value to a *vigilance* parameter ρ. If the condition

$$g_J^{'}(I) > \rho \quad (7)$$

holds, the state of *resonance* occurs. If condition 7 does not hold, the node J is reset and a new winner is determined by Eq. 5. Finally, the class K predicted by category J is determined

$$K = \Omega(J) \quad (8)$$

where $\Omega(\ )$ maps category J to its predicted class K. If the predicted class K is the correct one for a given input pattern, then the parameters of the winning category J are updated. Otherwise, a *match tracking* occurs. This mechanism is implemented by a temporary increase of the value of the *vigilance* parameter ρ to a value which forces the system to reset the current winner and thus to a choice of a new winner. The value of the vigilance parameter ρ is increased to a value of

$$\rho = g_J^{'}(I) \quad (9)$$

and it is reset to its original value only after a new pattern has been presented.

Once a category has been found which correctly predicts the class to which the current input pattern belongs, the parameters of this category are updated according to the following equations

$$n_J^{new} = n_J^{old} + 1 \\ \mu_J^{new} = \left(1 - \left(n_J^{old}\right)^{-1}\right)\mu_J^{old} + \left(n_J^{old}\right)^{-1} I \\ \Sigma_{Jk,l}^{new} = \left(n_J^{new}\right)^{-1}\left(n_J^{old}\Sigma_{Jk,l}^{old} + n_J^{old}\mu_{Jk}^{old}\mu_{Jl}^{old} + I_k I_l\right) - \\ - \left(n_J^{new}\right)^{-2}\left(n_J^{old}\mu_{Jk}^{old} + I_k\right)\left(n_J^{old}\mu_{Jl}^{old} + I_l\right). \quad (10)$$

Each newly created category is in the first step of the training algorithm initialized. Its mean values μ_j are set to the values of the present input I, and its co-variance matrix Σ_j is set to the values of $\gamma^2 E$, where γ^2 is the initial variance and E is a diagonal matrix.

During testing, the EGA behaves the same way as during the training stage most of the time. It follows exactly Eqs. 4-7. The difference between training and testing is only in the method of determination of the class, to which an unknown testing pattern is assigned. Here, similarly to the GA algorithm, the unknown pattern is assigned to the class with the highest cumulative probability over the whole network, defined by

$$K = \arg\max_k \left(\sum_{j \in \Omega^{-1}(k)} \exp\left(g_j(I)\right)\right) \quad (11)$$

where $\Omega^{1}(k)$ defines the set of categories j mapped to the output class k.

The different ways of mapping of the winning category J to an output class K, described in Eqs. 8 and 11, are equivalent in that either of them can be used during the training and/or during the testing phase. In the present simulations, these equations were used as described above.

The final mechanism for improvement of classification performance, used in most ARTMAP neural networks, is the *voting* strategy. This strategy consists in training several independent neural networks on the same training set with the training patterns presented to each network in a different order. Then, in the testing phase, an unknown pattern is presented to each of the networks and the decision of each of these networks contributes a vote to the final decision of the system concerning the output class for a given input pattern. In GA and EGA, this strategy is usually implemented as follows

$$K = \arg\max_k \left(\sum_{v=1}^{V} \sum_{j \in \Omega_v^{-1}(k)} \exp\left(g_{v,j}(I)\right) \right) \quad (12)$$

where V is the number of the EGA networks participating in the *voting*. Eq. 12 does not define the only possible method of *voting* in GA networks, but this is the way *voting* was implemented in the following simulations.

V. SIMULATIONS

The goal in the present simulations was to compare the three classification methods (fuzzy ARTMAP, Gaussian ARTMAP, and Extended Gaussian ARTMAP) in terms of their classification accuracy achieved on the image data from remote sensing of the Earth. The results should suggest the most suitable method of cluster identification for the image data used here. Another goal is to present the Extended Gaussian ARTMAP as a new method for classification of remotely sensed data.

The data set, described in SECTION I, was split into two equal-sized subsets, the training and the testing set. Five permutations of the training set were generated to analyze the sensitivity of the examined systems to the ordering of the data, and to evaluate the usefulness of the *voting* strategy for improvement of classification accuracy in these systems.

All the simulations were run with the following values of the network parameters, obtained by a simple cross-validation technique (ρ - baseline vigilance/ similarity, β - learning rate, and γ- initial std. deviation in GA and EGA): FA ($\rho = 0.8, \beta = 1$), GA ($\rho = 0.0, \beta = 1, \gamma = 0.5$), and EGA ($\rho = 0.0, \beta = 1, \gamma = 0.5$).

VI. RESULTS AND DISCUSSION

The classification performance of the three analyzed systems is compared in terms of the Weighted Percent of Correctly Classified patterns (weighted PCC, described in [8]) in TABLE II. The table shows the accuracy of

TABLE II
PERFORMANCE (IN WEIGHTED PCC) FOR THE THREE COMPARED MATHEDS ON PERMUTATIONS OF THE TRAINING SET AND FOR VOTING

	Set #1	Set #2	Set #3	Set #4	Set #5	Voting
Fuzzy ARTMAP	93.72	91.48	90.82	90.82	92.16	93.95
Gauss. ARTMAP	93.90	93.57	93.49	94.24	93.09	94.04
Ext. Gauss. ARTMAP	93.82	94.14	94.14	94.07	93.88	94.23

classification of the testing set, obtained by each system after training on a permutation of the training set. Also shown is the performance of each system obtained using the *voting* strategy. The results show that the highest classification accuracy is obtained by the Extended Gaussian ARTMAP neural network. This is the case both for training on individual permutations of the training set and for the system using the *voting* strategy. The performance of EGA is only slightly better than that of the GA algorithm, both in simulations with and without *voting*. However, there is a significant difference in performance of EGA (or GA) when compared to the performance of the fuzzy ARTMAP algorithm, especially in the case without voting. The difference among the algorithms is not so significant when *voting* is used. This suggests that, when compared to the other two algorithms, the fuzzy ARTMAP algorithm is much more sensitive to the ordering of the input patterns.

More insight into behavior of the three systems can be obtained by analysis of the confusion matrices shown for FA, GA, and EGA in TABLES III, IV, and V, respectively.

TABLE III
CONFUSION MATRIX FOR FUZZY ARTMAP NN WITH VOTING (WEIGHTED PCC = 93.95). EACH ITEM IN THE TABLE GIVES THE PER CENT OF PIXELS FROM A GIVEN ACTUAL CLASS (COLUMN) CLASSIFIED INTO GIVEN PREDICTED CLASS (ROW). THE TOTAL FOR EACH ACTUAL CLASS (BOTTOM ROW) GIVES PER CENT OF PATTERNS IN THE TESTING SET THAT BELONG TO THE CORRESPONDING ACTUAL CLASS. THE TOTAL FOR EACH PREDICTED CLASS HAS ANALOGOUS MEANING

Predicted	Actual Class							
Class	A	B	C	D	E	F	G	Total
A’	**88.84**	0.87	1.57	0.00	0.00	0.00	2.64	2.53
B’	2.68	**98.97**	0.00	0.00	0.00	0.00	0.00	36.56
C’	5.80	0.16	**79.55**	0.11	0.00	2.47	3.25	5.31
D’	0.00	0.00	0.52	**96.33**	0.00	8.45	0.00	28.66
E’	0.00	0.00	2.27	0.00	**100.00**	0.00	0.00	6.60
F’	1.34	0.00	12.76	3.56	0.00	**88.30**	0.61	15.39
G’	1.34	0.00	3.32	0.00	0.00	0.84	**93.51**	4.96
Total	2.24	0.00	5.72	28.37	6.48	15.39	4.93	100.00

TABLE IV
CONFUSION MATRIX FOR GAUSSIAN ARTMAP NN WITH VOTING (WEIGHTED PCC = 94.04). FORMAT OF TABLE AS DESCRIBED IN TABLE III

Predicted	Actual Class							
Class	A	B	C	D	E	F	G	Total
A’	**91.52**	0.35	1.05	0.11	0.00	0.00	1.22	2.34
B’	0.00	**99.48**	0.00	0.00	0.00	0.00	0.00	36.68
C’	5.80	0.00	**87.24**	0.11	0.00	4.55	4.46	6.07
D’	1.34	0.00	0.00	**97.22**	0.00	7.21	0.00	28.72
E’	0.00	0.16	0.00	0.00	**100.00**	0.00	0.00	6.54
F’	0.00	0.00	8.92	2.57	0.00	**88.04**	1.22	14.85
G’	1.34	0.00	2.80	0.00	0.00	0.19	**92.90**	4.80
Total	2.24	0.00	5.72	28.37	6.48	15.39	4.93	100.00

TABLE V
CONFUSION MATRIX FOR EXTENDED GAUSSIAN ARTMAP NN WITH VOTING (WEIGHTED PCC = 94.23). FORMAT OF TABLE AS DESCRIBED IN TABLE III. THE METHOD OF GENERATION OF THIS TABLE WAS DIFFERENT FROM TABLES III AND IV

Predicted	Actual Class							
Class	A	B	C	D	E	F	G	Total
A’	**88.73**	0.00	5.63	1.41	1.41	1.41	1.41	unavail.
B’	0.60	**99.40**	0.00	0.00	0.00	0.00	0.00	unavail.
C’	0.00	0.00	**87.85**	0.00	0.55	8.84	2.76	unavail.
D’	0.00	0.00	0.00	**96.21**	0.00	3.79	0.00	unavail.
E’	0.00	0.00	0.00	0.00	**100.00**	0.00	0.00	unavail.
F’	0.00	0.00	2.46	7.39	0.00	**90.14**	0.00	unavail.
G’	0.00	0.00	3.21	0.64	0.00	0.00	**96.15**	unavail.
Total	2.24	0.00	5.72	28.37	6.48	15.39	4.93	100.00

TABLE V was generated by a method slightly different from TABLES III and IV, so only the values on diagonals are comparable among these tables. Despite this inconsistency, it can be seen from the TABLES that there are significant differences among the systems when their performance on individual classes is analyzed. The most significant difference is in classification performance on classes A, C, and G. This result suggests that one could create a hierarchical system combining information from all three tested systems to obtain a structure with performance superior to that of any of the individual systems.

When system dynamics and properties of EGA vs. GA are compared, one can conclude that the number of internally generated categories was on average 15-20% lower in EGA than in GA. On the other hand, the computations in EGA are much more complex and, in this case, each category is represented by a more complex structure (a full co-variance matrix has to be stored, updated, inverted, and a determinant of it has to be computed in every iteration). So the time and memory requirements of the EGA system are larger than those for the GA system.

Because of the way EGA represents each category (means, full co-variance matrix, and *a priory* probability) it can be expected that this system would be extremely slow when tested on high-dimensional data. However, results of this study show that this method can be successfully used on large data sets with low dimensionality.

VII. CONCLUSION

In this study, a new neural network algorithm, called Extended Gaussian ARTMAP, has been introduced. It was shown that despite several disadvantages, EGA is a suitable algorithm for classification of remotely sensed images. Also, it can be concluded that the algorithms that prefer data structured into Gaussian-distributed clusters outperform algorithms that expect data distributed in clusters of hyper-rectangular shape. EGA algorithm is suitable for classification of large low-dimensional data sets. For high-dimensional databases GA is the preferable algorithm. When either of these algorithms is used, the *voting* strategy can be omitted most of the time, which is not the case for the fuzzy ARTMAP algorithm.

VII. ACKNOWLEDGMENT

The authors would like to thank Prof. Howard Veregin from University of Minnesota for his help with preparation of the data used in this study.

VIII. REFERENCES

[1] P.M. Atkinson and A.R.L. Tatnall, "Neural networks in remote sensing," *Int. J. Remote Sensing,* vol. 18, 4, 1997, 711-725.

[2] J. Richards, *Remote sensing digital image analysis: An introduction.* Springer Verlag: Berlin, 1993.

[3] B.G. Lees and K. Ritman, "Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments," *Environmental Management*, vol. 15, 1991, pp. 823-831.

[4] C.H. Chen, "Trends on information processing for remote sensing," in *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 3, Aug. 3-8 1997, pp. 1190-1192.

[5] P. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, Ph.D. thesis, Harvard University, 1974.

[6] G.A. Carpenter, M.N. Gjaja, S. Gopal, and C.E. Woodcock, "ART neural networks for remote sensing" Vegetation classification from Landsat TM and terrain data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 2, 1997, pp. 308-325.

[7] N. Kop o, P. Sin ák, and H. Veregin, "Extended Methods for Classification of Remotely Sensed Images Based on ARTMAP Neural Networks," in *Computational Intelligence – Theory and Applications (Lecture Notes in Computer Science 1625), Proceedings of International Conference "The 6-th Fuzzy days,"* Dortmund, Germany, May 1999, pp. 206-219.

[8] P. Sin ák, H. Veregin, and N. Kop o, "Conflation techniques in multispectral image processing," *Geocarto Int*, Jun. 1999.

[9] S. Grossberg, "Adaptive pattern classification and universal recoding, I: Feedback, expectation, olfaction, and illusions," *Biological Cybernetics,* vol. 23, 1976, pp. 187-202.

[10] G.A. Carpenter, B.L. Milenova, and B.W. Noeske, "Distributed ARTMAP: a neural network for fast distributed supervised learning," *Neural Networks,* vol. 11, no. 5, Jul. 1998, pp. 793-813.

[11] J.R. Williamson, "Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps," *Neural Networks,* vol. 9, 1996, pp. 881-897.

[12] R.K. Cunningham, *Learning and recognizing patterns of visual motion, color, and form,* Unpublished Ph.D. thesis, Boston University, Boston, MA: 1998.

[13] R. Duda and P. Hart, *Pattern Classification and Scene Analysis,* Wiley, New York: 1973.