An Enhanced ART2 Neural Network for Clustering Analysis

Jianhong Luo[1], Dezhao Chen[1]
[1] *Zhejiang University, Hang Zhou, 310027, People's Republic of China*
luojianhong@gmail.com

Abstract

The adaptive resonance theory 2 (ART2) neural network exhibits several properties which can be useful in the data mining and which are lacking in most other neural networks. But ART2 has deficiencies that the categories clustered by ART2 are very mutable to slight changes in training conditions. An improved ART2 with enhanced triplex matching mechanism, named as ETM-ART2, is presented to redress the deficiencies. Several tests results show that ETM-ART2 performs better than classic ART2 when applied to clustering tasks. It is an effective improved algorithm and can be applied to a wide variety of problems.

1. Introduction

As one of major data mining methods, cluster analysis plays an important role in analyzing information. Artificial neural networks have been widely applied in clustering analysis. The adaptive resonance theory 2 (ART2) network is suited for tasks that data patterns should be clustered into groups of similar characteristics, and it is comparable to conventional clustering techniques for this purpose [1]. Furthermore, ART2 is a self organizing network capable of dynamic, on-line learning. Compared to other neural networks such as BP neural network, SOM, fuzzy neural network, adaptive resonance theory clarifies how sensory and cognitive processes solve a key problem, called stability –plasticity dilemma. Therefore, ART2 has been applied in many fields [2-7], and clustering analysis is one of an important application. The number of categories, into which samples are classified, is automatically determined by specifying the value of vigilance parameter. Thus, the learning of ART2 neural networks is straightforward. But ART2 suffers from several deficiencies that the categories clustered by ART2 are very sensitive to slight changes in training conditions [2, 8, 9]. For instance, if ART2 network is trained by the same data set but with different input orders, it may exhibit greatly different classification results. But because ART2 does exhibit several properties which can be useful in the clustering task and which are lacking in most other neural networks, many researchers have endeavored to develop extension works to ART2 in order to improve the network's capabilities on the clustering task [10-14]. In reference [12], the first pattern classified to a category is called as the initial template of the category. During the recognition process of ART2, an input pattern should not only be compared with the template of each category but also with its initial template. Only when the comparison results for a category are not lower than an adjustable vigilance parameter in ART2, the input pattern should be classified to the category. The initial pattern of each category is used to enhance the matching criterion and it improves the patterns distinguish capability of ART2. It is called as OP-ART2 in this paper to make comparison tests in the next sections.

In this paper, a concise structural extension of ART2 is studied to redress the clustering deficiencies, which utilizes an enhanced triplex matching mechanism in category discriminating process of ART2 and improves recognition capability for arbitrary sequences of input patterns. It is named as ETM-ART2, and the classic ART2 is named as C-ART2 in this paper. ETM-ART2 is evaluated by several tests and the testing results demonstrate better clustering performance than C-ART2 and OP-ART2, thereby improves the clustering capability of ART2. And it can be applied to a wide variety of problems.

The remainder of this paper is organized as follows. Section 2 introduces ART2 neural work and its' deficiencies in clustering analysis. To redress these deficiencies, ETM-ART2 is presented in Section 3. Evaluation and tests are executed in Section 4.And Section 5 gives a conclusion.

2. ART2 Neural Network

The ART2 network consists of three main components, termed by Carpenter and Grossberg [8] as F1 layer, F2 layer, and the orienting mechanism, see

e-Forensics 2008, January 21-23, 2008, Adelaide, Australia.
© 2008 ICST 978-963-9799-19-6.

Figure1. The F1 layer is the input representation field and the F2 layer is the category representation field.

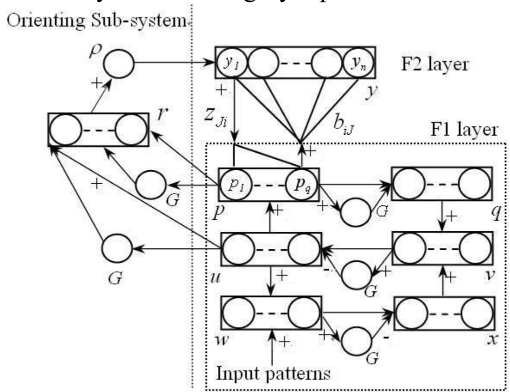

Figure 1. The architecture of ART2 neural network

The nodes of the layer F2 are intended to represent the classes into which input patterns are organized [11]. Each output node represents a class of patterns by storing a template pattern as the top to bottom weights value z_{Ji} into the F1 layer, where $J \in \{1,\cdots,n\}$, $i = 1,\cdots,q$, see Figure 1. The n denotes the number of recognition categories in F2 layer and q denotes the dimensions of input patterns from F1 layer. During a training of ART2, the input pattern first goes through transformations in the F1 layer, and then is passed from the bottom to top weights value b_{iJ} , where $i = 1,\cdots,q$, $J \in \{1,\cdots,n\}$. The result of the bottom to top value is an input signal of F2 layer, and then it activates a competition between all nodes in the F2 layer, each of which represents a stored category template. The goal of this competition is to find a node, which matches with the presented input pattern better than all the other competitors.

The node with the highest activation value in F2 layer is designated as the initial choice of class for the input pattern. The input pattern and the template for the initially chosen class are subjected to a further comparison in the orienting subsystem, if the match is judged to be within the vigilance level, the system gets in resonance state, and the initial choice is taken as final and the template for that winning node is updated according Equation (1), where u_i is the output of i th node of u sub-layer for the input pattern and d is a parameter of ART2. Then the category template is adjusted to represent the current input pattern more closely.

$$z_{Ji} = b_{iJ} = \frac{u_i}{1-d} \quad (0 < d < 1) \tag{1}$$

If the match is judged to be outside of the vigilance level, then the output node of F2 layer with the next highest activation value is designated as the initial match and it is subjected to the same processing in the orienting subsystem to determine if it is an acceptable final choice. This process continues through the available choices, in decreasing order of the activation of the output nodes, and if none of them meets the vigilance level test, then a new output node is recruited as a new class and its template is set to the current input [11].

From the learning cycle of ART2, it is known that the category template is updated to represent the current input pattern more, which contains many elements of recent inputted patterns. With the rest data patterns inputted to the ART2 network, the templates may be "reset and reset" during the learning process. If the serial patterns data input to ART2 in an order that patterns arranged with slight changes, the template may be gradually reset to deflect from the "center" of a category. That is the template can't represent the category of input patterns well. Then the clustering result will be badly affected at this condition [11]. This is a potential disadvantage of ART2 and therefore makes classification errors.

3. ART2 neural network with enhanced triplex matching mechanism(ETM-ART2)

3.1 Architecture of ETM-ART2 neural network

Patterns with a certain extent similarity can be clustered to a category. The similarity between two pattern vectors can be calculated by cosine. See Equation (2), suppose $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are two pattern vectors. In other words, the similarity between two vectors can be denoted as their cross angle ϕ. The angle is littler, the similarity is higher. In ART2, the matching criterion between an input pattern and a chosen category template is determined by an adjustable vigilance parameter. All other things being equal, higher vigilance imposes a stricter matching criterion, which in turn partitions the input data set into finer categories [8].

$$r = \frac{\boldsymbol{v}_1 \cdot \boldsymbol{v}_2}{\|\boldsymbol{v}_1\| \times \|\boldsymbol{v}_2\|} = cos \quad \phi \tag{2}$$

In C-ART2, the category template matching is an important criterion to classify the input patterns. The winner neuron sends its weight vector z^* back to the input layer and only learns the new pattern if it lies within a cone around the weight vector z^* , which represent a category template [15]. The size of the cone is determined by the vigilance parameter. The left part in Figure 2 shows the attentiveness cone. At every vigilance level, the matching criterion is selfscaling and a small mismatch may occur. But the mismatch would also trigger a category template reset in the learning process. So the template may be gradually reset to deflect from the “center” of a category, and it will lead to bad classification errors. Therefore, the template matching criterion is not enough to get good clustering results. Stricter matching criterion is required to decide whether an input pattern should be clustered to the category, in order that template deflection is redressed to reduce clustering errors.

If the similarity between an input pattern and all the current patterns classified into a category is not lower than a vigilance parameter ρ , and the similarity between the input pattern and the category template is also not lower than ρ, then the input pattern may be classified to the category. And the pattern may lie within the cone. See the right part in Figure 2, the cone shows a category with vigilance set to ρ , and the template is z', the input pattern is $\boldsymbol{u}'$. But it is not practical to be compared with all the patterns in a category. In the improved ART2, called as ETM-ART2 in the paper, three significative matching rules are added to enhance the matching criterion instead of that. The two vectors whose similarity is minimum value in a category are likely to lie on the verge of the cone. They are denoted as f_1 and f_2 in Figure2. The barycenter of a category is denoted as ξ. So an input pattern should be compared with these three vectors $\boldsymbol{f}_1, \boldsymbol{f}_2$ and ξ besides the template of a category .If the similarity between an input pattern with ξ is not lower than $\rho/2$, the similarity between the pattern with f_1, f_2 and the category template are all not lower than ρ , then the pattern is clustered to the category. This is an enhanced triplex matching mechanism for the improved ART2, and named as ETM-ART2.

The architecture of ETM-ART2 network is depicted in Figure3. Three sets of mnemonic neural units are added in F2 layer to store the information of two verge patterns $\boldsymbol{f}_1$ and f_2 of each category, and the barycenter ξ of each category. A mnemon vector $\boldsymbol{R}$ is added into Orienting Sub-system to store r_f , which shows the similarity between f_1 and f_2.$\boldsymbol{R}$ is a *n*-

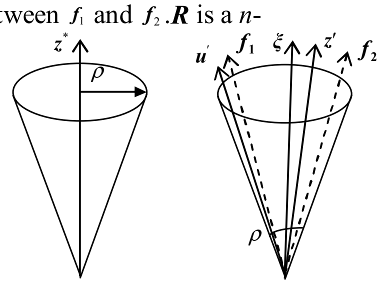

Figure 2. Attentiveness cone of a class

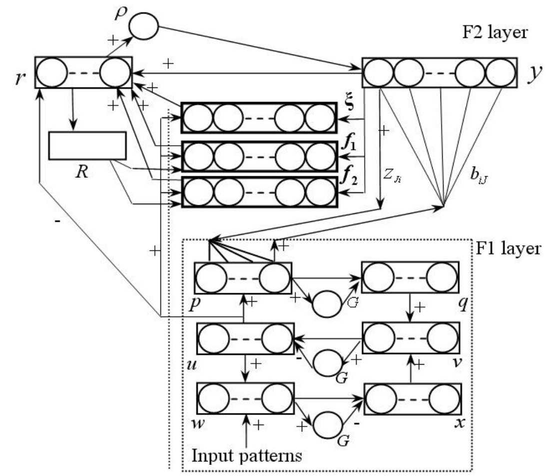

Figure 3 Architecture of ETM-ART2 neural network

dimensional vector, where n is the number of discriminating neural units in the F2 layer. The r_f of each category is stored in each dimension of the $\boldsymbol{R}$ vector.

3.2 Algorithm of ETM-ART2 Neural Network

The important difference between ETM-ART2 and C-ART2 lies in the F2 layer and Orienting Sub-system. The training algorithm of ETM-ART2 is explained as follows.

Step1. An input pattern s is inputted to ETM-ART2, and it first goes through transformations in the F1 layer like C-ART2. The output of p sub-layer is transformed to F2 layer. Find the maximum activation unit y_j after competition learning in F2 layer.

Step2. If it’s the first time for unit y_j to get the maximum activation, a new category that represented by the input pattern s is set up in F2 layer. Update the template and store the information of two verge patterns for the new category: $\boldsymbol{f}_1^{(j)} = \boldsymbol{f}_2^{(j)} = \boldsymbol{\xi}^{(j)} = \boldsymbol{u}^*$, $r_f^{(j)}$ =1, where $\boldsymbol{u}^*$ is the output of u sub-layer and represent the input pattern. Go to Step 5.

Step3. If it’s not the first time for y_j to get the maximum activation in F2 layer. Then compute the similarity between $\boldsymbol{u}^*$ and the *j*-th category template $\boldsymbol{z}^{(j)}$, $\boldsymbol{u}^*$ and $\boldsymbol{f}_1^{(j)}$, $\boldsymbol{u}^*$ and $\boldsymbol{f}_2^{(j)}$, $\boldsymbol{u}^*$ and $\boldsymbol{\xi}^{(j)}$ according to Equations (3)-(6). The f_1, f_2 and ξ are updated after each learning process. And they are also important characteristics of a category during the learning process.

Compare them with the vigilance parameter ρ of ETM-ART2. If $|r_z| \geq \rho, |r_1| \geq \rho, |r_2| \geq \rho$,

$$r_z = \frac{\boldsymbol{u}^* \cdot \boldsymbol{z}^{(j)}}{\|\boldsymbol{u}^*\| \times \|\boldsymbol{z}^{(j)}\|} \tag{3}$$

$$r_1 = \frac{\boldsymbol{u}^* \cdot \boldsymbol{f}_1^{(j)}}{\|\boldsymbol{u}^*\| \times \|\boldsymbol{f}_1^{(j)}\|} \quad (4)$$

$$r_2 = \frac{\boldsymbol{u}^* \cdot \boldsymbol{f}_2^{(j)}}{\|\boldsymbol{u}^*\| \times \|\boldsymbol{f}_2^{(j)}\|} \quad (5)$$

$$r_c = \frac{\boldsymbol{u}^* \cdot \xi^{(j)}}{\|\boldsymbol{u}^*\| \times \|\xi^{(j)}\|} \quad (6)$$

and $|r_c| \geq \rho/2$, the input pattern is classified to the category j and go to next step. Otherwise turn to Step 6.

Step4. If the input pattern is the second pattern of category j, then $\boldsymbol{f}_2^{(j)} = \boldsymbol{u}^*$, $r_f^{(j)} = r_1$, $\boldsymbol{f}_1^{(j)}$ keeps unchanged. Otherwise, $\xi^{(j)}$ is updated, and

$$\left.\begin{array}{l} \boldsymbol{f}_2^{(j)} = \boldsymbol{u}^*,\ r_f^{(j)} = r_1,\quad (r_1 < r_f^{(j)} \ \ and \ \ r_1 < r_2) \\ \boldsymbol{f}_1^{(j)} = \boldsymbol{u}^*,\ r_f^{(j)} = r_2,\quad (r_2 < r_f^{(j)} \ \ and \ \ r_2 < r_1) \end{array}\right\} \quad (7)$$

Therefore, $r_f^{(j)}$ is getting decreasing, which means the similarity between the two verge vectors of category j is getting decreasing. See the cone in the right part of Figure2, ξ shows the center of a category and the two verge vectors are getting closer to the borderline of the cone with more and more patterns be classified to the category.

Step5. ETM-ART2 is resonated and updates weights between F1 layer and F2 layer by Equation (1). Stop the learning for the current pattern *s*, and begin learning for the next pattern.

Step6. The input pattern can't be clustered to the category j. Stop the interaction of unit y_j which should not join the later competition learning. Start a new competition learning process until the clustering result of the input pattern is reached.

In this ETM-ART2 algorithm, the enhanced triplex matching mechanism adds some computation complexity to the network learning, but the time cost for it is very little compared with the whole time cost of the ART2 network.

4. Evaluation and Testing

4.1 Olive oils data set

Clustering analysis for chemical patterns is significant in data mining. A sort of product may contain different percentage composition of elements since they come from different producing area. When faced with large numbers of product data, it is significative to partition them according to their percentage composition for further researches. Olive oil is a typical product of that. Clustering analysis for olive oils data is studied in this section.

This olive oils data set OODS includes 402 Italian olive oils samples come from 6 growing regions based on the chemical analyses of 8 different fatty acids. These fatty acids are: palmitic, palmitoleic, stearic, oleic, linoleic, linolenic, arachidic and eicosenoic. The number of samples from 6 different regions is listed in Table 1.

Table 1. The number of samples of olive oils from different regions in data set OODS

Category NO.	Region	Number of samples
1	North-Apulia	25
2	South-Apulia	181
3	Inland-Sardin	62
4	Coast-Sardini	33
5	West-Liguria	50
6	Umbria	51
total		402

4. 2 Testing results and comparisons

The patterns data in OODS are inputted to ETM-ART2 in random order. Samples that can't be clustered to the category which represents their growing regions are considered as incorrectly classification. Three trials are performed. After each trial, record the number of categories, the number of patterns in each category and the correct rate. They are listed in the Table 2.

To get comparison, the same tests for C-ART2, and tests for OP-ART2 [12] are performed. Patterns are inputted in random order. The clustering results by OP-ART2 are listed in Table 3 and the results by C-ART2 are listed in Table 4. In Table 3, the correct rate is 69.40% in Trial 1 while it is 74.87% in Trial 2. It shows that the clustering results are easily influenced by the input order for C-ART2. From Table 4, the correct rate is improved by OP-ART2. While the influence by input order is reduced a little

From Table 2, it can be seen that ETM-ART2 has good capability in clustering analysis. And the influence by input order is reduced remarkably. The deficiencies exhibited by C-ART2 when applied to clustering tasks can be redressed by ETM-ART2 effectively. Though OP-ART2 works well than C-ART2, it is not as well as ETM-ART2. It is clear that the recognition capability of ETM-ART2 is prominent better than that of C-ART2 and OP-ART2.

Table 2. Testing results for OODS by ETM-ART2 $\rho = 0.2$

Trial NO.	Number of classes	Number of samples in each class	Correct rate
1	6	28 178 66 25 51 54	90.05%
2	6	23 168 72 30 52 57	89.55%
3	6	26 173 64 31 53 55	89.80%

Table 3. Testing results for OODS by C-ART2 $\rho = 0.3$

Trial NO.	Number of classes	Number of samples in each class	Correct rate
1	6	47 115 96 11 42 91	69.40%
2	6	39 132 83 26 41 81	74.87%
3	6	38 120 99 23 39 83	71.89%

Table 4. Testing results for OODS by OP-ART2 $\rho = 0.3$

Trial NO.	Number of classes	Number of samples in each class	Correct rate
1	6	36 141 95 33 82 15	77.86%
2	6	38 139 88 41 71 25	81.84%
3	6	29 162 79 40 62 30	82.58%

5. Conclusions

In this paper, an improved ART2 neural network is developed, which attempts to redress some of the deficiencies exhibited by ART2 when applied to clustering tasks and named as ETM-ART2. Enhanced triplex matching mechanism is proposed to redress template deflection and improve performance of ART2. A concise structural extension is studied based on the classic ART2 neural network. Several tests are employed by ETM-ART2 and the results show high correct rate. It has been demonstrated to perform better than the classic ART2 network and OP-ART2. ETM-ART2 thereby improves the clustering capability of ART2 while preserving the benefits of ART2's self organization and on-line learning characteristics. It can be applied to a wide variety of problems.

6. References

[1] L. I. Burke, "Clustering characteristics of adaptive resonance," *Neural Networks,* pp. 485-491, 1991.

[2] G. P. B. Houshmand, "An efficient neural classification chain of SAR and optical urban images," *International Journal of Remote Sensing,* vol. 22, pp. 1535-1553, 2001.

[3] C. H. Chen, Khoo, L.P. & Yan, W. , "A strategy for acquiring customer requirement patterns using laddering technique and ART2 neural network," *Advanced Engineering Informatics,* vol. 16, pp. 229-240, 2002.

[4] G. Graef, Schaefer, C., "Application of ART2 Networks and Self-Organizing Maps to Collaborative Filtering," *Lecture Notes in Computer Science,* vol. 2266, pp. 296-309, 2002.

[5] T. Obikawa, Shinozuka, J., "Monitoring of flank wear of coated tools in high speed machining with a neural network ART2," *International Journal of Machine Tools & Manufacture,* vol. 44, pp. 1311–1318, 2004.

[6] B. Karthikeyan, Gopal, S., Venkatesh, S. , " ART 2—an unsupervised neural network for PD pattern recognition and classification," *Expert Systems with Applications,* vol. 31, pp. 345-350, 2006.

[7] K. B. Kim, Kim, M., Woo, Y.W., "Recognition of Shipping Container Identifiers Using ART2-Based Quantization and a Refined RBF Network " *Lecture Notes in Computer Science,* vol. 4432, pp. 572-581, 2007.

[8] G. A. Carpenter, Grossberg.S., "ART2: self-organization of stable category recognition codes for analog input pattern," *Applied Optics,* vol. 26, pp. 4919–4930, 1987.

[9] D. T. Pham, Chan, A.B., "Unsupervised adaptive resonance theory neural networks for control chart pattern recognition," *Proceeding of the Institution of Mechanical Engineers,* vol. 215, pp. 59-67, 2001.

[10] G. A. Carpenter, Grossberg, S., Markuzon, N., Reynolds, J.H. & Rosen, D., "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," *IEEE Transactions of Neural Networks,* vol. 3, pp. 698-713, 1992.

[11] T. S. Hussain, Browse, R.A., "ARTSTAR: A Supervised Modular Adaptive Resonance Network Classifier," in *AI'94: Tenth Canadian Conference on Artificial Intelligence*, 1994.

[12] M. J.S., "A new ART2 neural network model," *Patterns Recognization and Artificial Intelligence,* vol. 13, pp. 231-234, 2000 (in Chinese).

[13] Y. Q. Cao, Wu, J.H., "Projective ART for clustering data sets in high dimensional spaces," *Neural Networks,* vol. 15, pp. 105-120, 2002.

[14] Y. P. Xu, " An improved ART2 neural network clustering algorithm," *Computer Applications,* vol. 26, 2006 (in Chinese).

[15] G. A. Carpenter, Grossberg, S., & Rosen, D., "ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition," *Neural Networks,* vol. 4, pp. 493-504, 1991.