SELECTED PAPER

Classification Function of Radial-Basis Adaptive Resonance Theory Map and Its ApplicationYoko Enosawa¹, Haruna Matsushita², Yasuhiro Tanaka³, Naoaki Kuji⁴ and Toshimichi Saito¹

¹Dept. of Electrical & Electronics Eng.
Hosei University, Tokyo 184-8584, Japan
E-mail: tsaito@hosei.ac.jp

²Dept. of Reliability-based Information Systems Eng.
Kagawa University, Kagawa 761-0396, Japan
E-mail: haruna@eng.kagawa-u.ac.jp

³Tanaka Women's Clinic
5-25-1 Okusawa, Tokyo 158-0083, Japan
E-mail: info@tanaka-wc.com

⁴Dept. of Obstetrics & Gynecology, School of Medicine
Keio University, Tokyo 160-8482, Japan
E-mail: naoaki@z7.keio.jp

Abstract

We present a radial-basis adaptive resonance theory map and its application to classification problems. In order to classify input data, the system generates radial-basis categories. Two key parameters can control the number and size of the categories. Performing numerical experiments for two types of traditional expectant mothers, we have investigated the effects of key parameters on the classification functions.

1. Introduction

The adaptive resonance theory networks (ART-nets) are typical unsupervised learning systems having a feature extraction function of objective environments [1]. In order to improve the algorithm performance, several variants have been studied: e.g., hypersphere ART-nets [2], ellipsoid ART-nets [3] and radial-basis ART-nets [4]. The adaptive resonance theory maps (ART-maps) are supervised learning systems based on the ART-nets and have flexible classification functions [5]-[7]. There are many applications of the ART-nets and ART-maps, including image-processing, adaptive anomaly detection and combinatorial optimization problems [7]-[9].

We present a radial-basis ART-map (RB-ARTM) and its application to classification problems. The RB-ARTM is based on a standard ART-maps and radial-basis ART-nets. In order to classify input data with class information, the RB-ARTM generates categories characterized by center, radius, and class. The learning algorithm includes two key parameters: vigilance parameter ρ and distance parameter k . These parameters can control the number and size of categories. We investigate the algorithm performance in basic numerical experiments for two example problems: classification of typical benchmarks, and classification of abdomen shapes of expectant mothers.

2. Radial-Basis ART-Map and Learning Algorithm

We define the RB-ARTM for classification problems. Let M denote the number of classes. Let S_0 be in an N -dimensional input space. For simplicity, we explain the case of $M = 2$ and $N = 2$. The i -th input P_i is characterized by its position $\mathbf{X}_i$ and class C_i :

$$\begin{aligned} P_i &= (\mathbf{X}_i, C_i) \\ \mathbf{X}_i &\equiv (X_i, Y_i) \in S_0, C_i \in \{-1, 1\}, i = 1 \sim N_p \end{aligned} \quad (1)$$

where N_p is the number of inputs. The j -th category $\mathbf{W}_j$ at discrete time t is characterized by a circle at the center $\mathbf{x}_j \equiv (x_j, y_j)$ with radius r_j and class c_j :

$$\begin{aligned} \mathbf{W}_j(t) &= (\mathbf{x}_j(t), r_j(t), c_j(t)) \\ \mathbf{x}_j &\in S_0, c_j \in \{-1, 1\}, j = 1 \sim N_c(t) \end{aligned} \quad (2)$$

The number of categories $N_c(t)$ varies in the algorithm. After

Figure 1: Learning algorithm: A cross denotes an input datum. A circle is a category. (a) Preservation, (b) Creation (A point on the input is a newly created category whose radius is zero), (c) Before enlargement, (d) After enlargement.

learning, each input datum is classified into either one of the

two classes. For this classification, we define a key parameter: *base class* C_b . C_b is defined as the class consisting of more data than another class.

Step 1 (Initialization): Let $t = 0$, $N_c(0) = 1$ ($j = 1$), and $r_j(0) = 0$. The category center $\mathbf{x}_j(0) \in S_0$ is initialized using a random number.

Step 2 (Comparison of base class): Input P_i is applied. If P_i has the same class as the base class C_b , namely, $C_i = C_b$, then go to Step 3. Otherwise, go to Step 6.

Step 3 (Winning selection): If P_i is included in any category shown in Figure 1(a), then go to Step 6. Otherwise, we find the winning category W_j that is the closest category to P_i :

$$T_j = \min_j (T_j(\mathbf{X}_i)), T_j(\mathbf{X}_i) = \|\mathbf{X}_i - \mathbf{x}_j(t)\| - k \times r_j(t), \quad (3)$$

where T_j is the choice function of the j -th category, $\|\cdot\|$ is the Euclidean distance, and $k \in [-1, 1]$ is the distance parameter. If $T_j > \rho$, then go to Step 4 where $\rho \in [0, \sqrt{2}]$ is the vigilance parameter. Otherwise, namely, $T_j \leq \rho$, go to Step 5.

Step 4 (Birth of a new category, Figure 1(b)): Let $N_c(t) = N_c(t) + 1$ and $j = j + 1$, and let $N_c(t)$ be the suffix of the new category: $W_{N_c(t)} = (\mathbf{X}_i, 0, C_i)$. A new category is born at the same position of input P_i . The radius of the new category is zero. The class of the new category is the same as that of input P_i . Go to Step 6.

Step 5 (Category enlargement, Figures 1(c) and 1(d)): Enlarge the winning category W_j to the input P_i .

Step 6: Let $t = t + 1$, go to Step 2 and repeat until the time limit $t = N_p$.

Classification: After learning, each input datum is classified into either one of the two classes in accordance with the class of the categories created by the RB-ARTM. If the datum is included in one or more categories, its class is regarded as the base class C_b . The datum included in no category is classified into another class that is different from C_b .

3. Numerical Experiments

In order to consider the algorithm performance, we apply the algorithm to two kinds of examples. Here, we use input datum composed of 2 classes, and the base class C_b is chosen from the 2 classes depending on the number of data in each class. The algorithm can classify the input data according to the minimum number of created categories.

3.1 Example 1

In order to evaluate the performance of the RB-ARTM, we have performed basic numerical experiments for 3 classification problems: Annulus (Figure 2), Checkerboard (Figure 3(a)), and Spirals (Figure 3(b)). The dimension of the input space is $N = 2$ and the number of classes is $M = 2$.

Figure 2: Example of simulation result of Annulus: The triangles denote class +1 data, and the squares denote class -1 data. (a) Classification problem, (b) 1,000 training samples used as input data, (c) Learning result (The number of created categories is 8 where $C_b = -1$, $k = 0$, and $\rho = 0.2$), (d) Classification results of evaluating 40,000 testing samples (The black points means error results. The coincidence rate is 83.7%.)

Figure 3: Two classification problems: The dimension of the input space is $N = 2$, and the number of classes is $M = 2$. (a) Checkerboard, (b) Spirals

The Annulus shown in Figure 2(a) is centered at (0.5, 0.5) between radii 0.4 and 0.2. The Checkerboard shown in Figure 3(a) has 9 squares of alternating classes. The Spirals shown in Figure 3(b) have 3 cycles each and a radius of 0.15.

We introduce two kinds of data: training sample and testing sample. The training sample is used to create categories, namely, the inputs for the RB-ARTM. The testing sample is used to evaluate the simulation results. We define the coincidence rate as the measure. The coincidence rate is obtained from classes with the correct answer. We compare the created class with the class of testing samples. The created class is regarded as the correct answer if it is the same class as the testing sample. If the testing sample belongs to multiple categories, the class is regarded as the base class C_b . Otherwise,

Figure 4: Relationships between the number of categories created by the RB-ARTM and the classification accuracy: (a) Annulus, (b) Checkerboard, (c) Spirals

the class is regarded as another class.

Figure 2 shows an example of simulation results for the Annulus. As shown in Figure 2(b), 1,000 training samples are selected randomly. The triangles denote class +1 data, and the squares denote class -1 data. The base class C_b is class -1. Figure 2(c) shows a learning result. The number of categories shown in the circles is 8, where $k = 0$ and $\rho = 0.2$. To measure the coincidence rate, 40,000 testing samples are selected randomly, as shown in Figure 2(d). The black points mean erroneous results. The coincidence rate is 83.7%.

In order to investigate the relationships between the number of created categories and the coincidence rate, we carried out simulations. For each simulation, we used different parameter sets: the vigilance parameter $\rho \in \{0, 0.01, 0.02, \dots, 1.41\}$ and the distance parameter $k \in \{-1, 0, 1\}$, namely, 426 kinds of parameter sets. Depending on the parameter value, the number of categories varies. The Annulus, Checkerboard, and Spirals consist of 100 training

samples. The coincidence rate has been calculated as the average of 50 trials for each parameter set, using 40,000 testing samples.

Figure 4 shows the simulation results of 3 classification problems. The horizontal axis is the number of categories created by the RB-ARTM, and the vertical axis is its coincidence rate. We summarize the key points in these results of the numerical experiments.

(1) The use of many categories does not always make the coincidence rate better. The coincidence rate is more unstable than in the case of fewer categories.

(2) Only two parameters, vigilance parameter ρ and distance parameter k , are important for the RB-ARTM. The performance of RB-ARTM can be controlled by these two parameters.

(3) The simulation results of the RB-ARTM strongly depend on the type of input data.

3.2 Example 2

We have applied the algorithm to a classification problem of practical data in bstetrics. We have basic data sets of 46 expectant mothers: abdomen shapes at 38 or 39-th week of pregnancy, and the sex of newborns.

Figure 5: Example shape of abdomen

Table 1: Principal component analysis

n -th principal component	1st	2nd	3rd	4th
Eigenvalue	2.80	0.83	0.24	0.13
eigenvector				
M_b	-3.89E-06	-6.27E-05	-2.83E-07	1.00E+00
M_r	4.57E-06	-1.12E-11	1.00E+00	-2.83E-07
M_f	1.00E+00	-4.04E-08	-4.57E-05	3.89E-06
M_l	4.02E-08	1.00E+00	2.71E-11	6.27E-05
Contribution ratio	70.0	20.7	5.95	3.32

Figure 5 shows an example abdomen shape. For simple quantification of the shapes, we introduce 4 feature points: back M_b , forward M_f , left M_l , and right M_r points. These points are defined using the polar coordinates (r, θ) around the center of gravity $G_p = (X_g, Y_g)$: M_b is the closest point to G_p for $\theta \in [-\pi/2, \pi/2]$, M_f is the farthest point from G_p for $\theta \in [\pi/2, 3\pi/2]$, M_l is the closest point to G_p for $\theta \in [\pi, 3\pi/2]$, and M_r is the closest point to G_p for $\theta \in [\pi/2, \pi]$.

Figure 6: Simulation result of the RB-ARTM for $k = 0$ and $\rho = 0.2$: The 22 squares and the 24 triangles are data for boys and girls, respectively. The solid and the dotted circles are categories for the boys and girls, respectively.

In order to reveal the internal structure of the data, we have applied the principal component analysis (PCA). The PCA is a simple true eigenvector-based multivariate analyses [10]. We can see that the 1st and 2nd principal components have a cumulative contribution ratio of 90.7% and these two components are close to M_f and M_l . For simplicity, we have selected M_f and M_l as characteristic quantities and have normalized them:

$$m_f = \frac{M_f - M_{fmin}}{M_{fmax} - M_{fmin}}, \quad m_l = \frac{M_l - M_{lmin}}{M_{lmax} - M_{lmin}} \quad (4)$$

where (M_{fmin}, M_{lmin}) and (M_{fmax}, M_{lmax}) are the minimum and maximum points of (M_f, M_l) , respectively. Figure 6 shows the 44 datums on the m_f vs m_l plane. The data are used as the training sample for the RB-ART and are labeled in 2 classes: the boy data are indicated by the 22 square points and the girl data are shown by the 24 triangular points. Figure 6 shows the learning result of the RB-ARTM. After trial and errors, the distance and vigilance parameters are fixed as $k = 0$ and $\rho = 0.2$. We can see that m_f of the girl categories has larger average and variance than m_f of the boy categories.

4. Conclusions

We studied supervised learning of the RB-ARTM and investigated the effects of two key parameters, the vigilance parameter ρ and the distance parameter k , on the classification capability of RB-ARTM. The results of basic numerical experiments for two example problems, the classification of typical benchmarks and abdomen shapes of expectant moth-

ers, confirmed that many created categories does not always make the coincidence rate better.

Future topics include the following: application to multi dimensional data, analysis of the learning process, analysis of parameter effects, and more detailed comparison with other classification methods.

References

- [1] G. A. Carpenter and S. Grossberg: A massively parallel architecture for a selforganizing, neural pattern recognition machine, *Computer Vision, Graphics, and Image Processing*, Vol. 37, pp. 54–115, 1987.
- [2] G. C. Anagnostopoulos and M. Georgiopoulos: Hyper-sphere ART and ARTMAP for unsupervised and supervised, incremental learning, in *Proc. IEEE-INNS Int'l. Joint Conf. Neural Networks*, pp. 59–64, 2000.
- [3] G. C. Anagnostopoulos and M. Georgiopoulos: Ellipsoid ART and ARTMAP for incremental clustering and classification, in *Proc. IEEE-INNS Int'l. Joint Conf. Neural Networks*, pp. 1221–1226, 2001.
- [4] M. Ohki, H. Torikai and T. Saito: Basic characteristics of a simple radial-basis ART network and its application, in *Proc. Int'l. Conf. Cogni. Neur. Syst.*, 83, 2004.
- [5] G. A. Carpenter, S. Grossberg and D. B. Rosen: Fuzzy ART: fast stable learning and categorization of analog patterns by an adaptive resonance system, *Neural Networks*, Vol. 4, pp. 759–771, 1991.
- [6] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds and D. B. Rosen: Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps, *IEEE Trans. Neural Networks*, Vol. 3, pp. 698–713, 1992.
- [7] O. Parsons and G. A. Carpenter: ARTMAP neural networks for information fusion and data mining: map production and target recognition methodologies, *Neural Networks*, Vol. 16, pp. 1075–1089, 2003.
- [8] Y. Liao, V. R. Vemuri and A. Pasos: Adaptive anomaly detection with evolving connectionist systems, *Journal of Network and Computer Applications*, Vol. 30, pp. 60–80, 2005.
- [9] M. Takanashi, H. Torikai and T. Saito: An approach to fusion of growing self-organizing maps and adaptive resonance theory maps, *IEICE Trans. Fundamentals*, Vol. E90-A, pp. 2047–2050, 2007.
- [10] H. Harold: Simplified calculation of principal components, *Psychometrika*, Vol. 1, pp. 27–35, 1935.