Cooperation Controlled Competitive Learning Approach for Data Clustering

Tao Li[1*] Wen-jiang Pei[1] Shao-ping Wang[1] Yiu-ming Cheung[2]

[1]*School of Information Science and Engineering, Southeast University, Nanjing, China*

[2]*Department of Computer Science, Hong Kong Baptist University, Hong Kong, China*

[*]*taoli@seu.edu.cn*

Abstract

Rival Penalized Competitive Learning (RPCL) and its variants can perform clustering analysis efficiently with the ability of selecting the cluster number automatically. Although they have been widely applied in a variety of research areas, some of their problems have not yet been solved. Based on the semi-competitive learning mechanism of Competitive and Cooperative Learning (CCL), this paper presents a new robust learning algorithm named Cooperation Controlled Competitive Learning (CCCL), in which the learning rate of each seed points within the same cooperative team can be adjusted adaptively. CCCL has not only inherited the merits of CCL, RPCL and its variants, but also overcome most of their shortcomings. It is insensitive to the initialization of the seed points and applicable to the heterogeneous clusters with an attractive accurate convergence property. Experiments have shown the efficacy of CCCL. Moreover, in some case its performance is prior to CCL and some other variants of RPCL.

Key Words —— Cooperation Controlled Competitive Learning, Clustering

1. Introduction

In intelligent statistical data analysis or unsupervised classification, clustering analysis is to determine the cluster number and explore the properties of each cluster [1-4]. It is well known that if the cluster number is pre-selected in advance, the most classical clustering algorithm is the *k-means* [5], which is on the basis of minimizing the mean-square-error (MSE) function. However, it suffers from the so called dead unit problem. That is, if the initial positions of some seeds are far away from the inputs in Euclidean space compared to the other seeds, these distant seeds will have no opportunity to be trained and become dead units [6-7]. Moreover, for *k-means* the clusters number k must be appropriately pre-assigned. When k is exactly equal to the true cluster number k^*, *k-means* will work successfully. Otherwise, its performance deteriorates rapidly [8].

Focusing on solve the critical problems encountered by *k-means*, several techniques have been developed. By using the strategy of reducing the winning rate of the frequent winners, *Frequency Sensitive Competitive Learning* (FSCL) solved the dead unit problem successfully [9]. Moreover, as an adaptive version of FSCL, *Rival Penalization Competitive Learning* (RPCL) implemented the automatic selection of the cluster number by gradually driving redundant seeds far away from the input dense regions [10]. As an efficient tool, RPCL has been widely applied to a variety of applications such as neural networks training, image processing, Markov model identification, discrete-valued source separation and so on [10-14]. However some experiments have shown that the performance of RPCL is sensitive to the delearning rate. Although some variants of RPCL such as *Rival Penalization Controlled Competitive Learning* (RPCCL) [15] and *Distance Sensitive RPCL* (DSRPCL) [16] have optimized its performance, they all have some common drawbacks such as unconvergence of redundant seeds, low precision in locating cluster centers, inapplicable to heterogeneous clusters and so on.

Unlike the RPCL and its variants, *Competitive and Cooperative Learning* (CCL) features the attractive seed convergence property, in which seed points not only compete each other for updating to adapt to an input, but also the winner will dynamically select several nearest competitors to form a cooperative team to adapt to the input together [17]. Instead of driving redundant seeds far away from the input dense regions, CCL make them convergent to some cluster centers. Finally, the number of those seed points stayed at different positions is exactly the cluster number. Nevertheless, based on the extensive experiments, we found that the performance of CCL is

978-0-7695-3508-1/08 $25.00 © 2008 IEEE
DOI 10.1109/CIS.2008.174

IEEE computer society

somewhat sensitive to the initialization of seed points. Moreover, its semi-cooperative mechanism is applicable to the homogenous clusters only.

In this paper, we present a new clustering technique named *Cooperation Controlled Competitive Learning* (CCCL). The basic idea is that the learning rate of each seed within the same cooperative team is adjusted dynamically based on the distance between itself and the current input. The cooperation controlled mechanism of CCCL can efficiently prevent the seeds within the same cooperative team from being merged with an excessive speed. Subsequently, those seeds locating at the periphery of the cooperative team can have the opportunity to break away from the current team and join the other teams in successive training. This may give them more chance to wander and search for more appropriate cluster centers. As a result, CCCL is able to overcome most shortcomings of CCL essentially. It is insensitive to the initialization of the seed points and suitable for heterogeneous clusters with an attractive accurate convergence property.

Experimental results of Gaussian mixture clustering have shown that CCCL is applicable to heterogeneous clusters, no matter how the seed points are distributed initially. By contrast, CCL and RPCCL usually fail to give correct clustering result under the same conditions.

The remainder of this paper is organized as follows: Section 2 further investigates the CCL algorithm and elaborates the CCCL in detail. Section 3 shows the performance of CCCL experimentally in comparison with the CCL and RPCCL. Finally, we draw a conclusion in section 4.

2. Cooperation Controlled Competitive Learning Algorithm

The basic idea of the CCL is that seed points not only compete with each other for updating to adapt to an input each time, but also the winner will dynamically select several nearest competitors to form a cooperative team to adapt to the input together without repelling each other [17]. Suppose there are N inputs, $x_1, x_2, \cdots, x_N$, come from k^* unknown clusters. w_c is the winner of input x_i during a certain learning epoch. Then w_c will regards those seeds fallen into the circle centered at itself with the radius $\|w_c - x_i\|$ as the cooperative members. Subsequently, these seeds will cooperate with each other to achieve the learning task. Obviously, the seed points locating in the same cluster will have more opportunity to cooperate than compete and vice versa. The detailed algorithm of CCL is as follows:

Step 1: Pre-specify the number k of clusters with $k \geq k^*$, random initialize the seed points $\{w_j\}_{j=1}^{k}$, and set the winning rate of seed $n_j = 1$ with $j = 1, 2, \cdots, k$.

Step 2: Randomly take an input x_i, calculate the indicator function $I(j \mid x_i)$ by

$$I(j \mid x_i) = \begin{cases} 1, & \text{if } j = \arg\min_{1 \leq r \leq k} r_i \|x_i - w_r\|^2 \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

with the relative winning frequency r_i of w_i defined as

$$r_i = \frac{n_i}{\sum_{j=1}^{k} n_j}, \quad (2)$$

where n_i is the winning times of w_i in the past. Then find out the winner w_c with $I(c \mid x_i) = 1$.

Step 3: The winner w_c regards those seed points fallen into the circle centered at w_c with the radius $\|w_c - x\|$ as the cooperating members. Let the cooperating set $C = \{w_c\}$. Then span C by

$$C = C \cup \left\{ w_j \mid \|w_c - w_j\| \leq \|w_c - x_i\| \right\}. \quad (3)$$

Step 4: Update all members in C by

$$w_u^{new} = w_u^{old} + \eta \left(x_i - w_u^{old} \right), \quad (4)$$

where $w_u \in C$. $\eta > 0$ is the given learning rate with small value. Furthermore, only update n_c by

$$n_c^{new} = n_c^{old} + 1, \quad (5)$$

without uniformly distributing the contribution of this winning to all other n_js. Finally the r_js can be used to estimate the proportion of the data from each cluster.

The above **Step2,3,4** are repeatedly iterated for each input until all seeds converge. CCL enables each extra seed point to finally locate at some of cluster centers. Hence, the number of those seed stayed at different positions is exactly the cluster number. Consequently, CCL can perform clustering analysis without prior knowing the cluster number.

Although CCL have circumvented some problems encountered by RPCL and its variants, based on the experiments results, we have noticed that CCL also has some inherent drawbacks. On one hand, it is somewhat sensitive to the initialization of seed points. On the other hand, CCL is not applicable to heterogeneous clusters. Subsequently, its application area is restricted by its essential drawbacks. Based on the semi-competitive learning mechanism of Competitive and Cooperative

Learning (CCL), this paper presents a new robust cooperation controlled learning algorithm, named Cooperation Controlled Competitive Learning (CCCL), in which the learning rate of each seed points within the same cooperative team can be adjusted adaptively.

In CCCL, the learning rate of each cooperative seeds is in inverse proportion to the distance between the cooperator itself and the current input. That is, the cooperative seeds will have a larger learning rate than the winner if their distance to the current input is smaller than the winner. It can guarantee these seeds to have a large convergent speed. On the contrary, for the other cooperative seeds whose distance to the current input is larger than the winner, a smaller learning rate is appropriate. It can prevent them from moving towards to the current input too rapidly and give them opportunities to join to other cooperative teams which convergent to the different cluster centers. Subsequently, we give out this cooperation controlled mechanism by

$$p(x_t; w_c, w_u) = \begin{cases} \dfrac{f(x_t, w_c)}{f(x_t, w_u)}, & if\ f(x_t, w_u) > \varphi f(x_t, w_c) \\ 1/\varphi, & if\ f(x_t, w_u) \geq \varphi f(x_t, w_c) \end{cases} \quad (6)$$

where η is the pre-assigned learning rate, f is a certain distance measuring function, e.g., Euclidean distance, or more general Mahalanobis distance. $\varphi \in [0.01, 0.1]$ defines the maxim value of $p(x_t; w_c, w_u)$. As shown in Fig.1, for a certain input x_t, we can get the winner seed w_c by Eq.1. Then w_c will regards the shadow circle centered at w_c with the radius $\|w_c - x\|$ as its cooperative region, in which w_u is a random cooperative seed.

When w_u is situated at point C, $f(x_t, w_u) > f(x_t, w_c)$ and $p(x_t; w_c, w_u) < 1$. The learning rate of w_u is smaller than η and gradually attenuated as the distance between w_u and the current input increases. When $f(x_t, w_u) = 2f(x_t, w_c)$, seed point will locate at point A and its learning rate will equal to the minimum value 0.5η. When w_u is located at point B, $f(x_t, w_u) < f(x_t, w_c)$ and $p(x_t; w_c, w_u) > 1$. The learning rate of w_u is larger than η and gradually increases to η / φ as the distance between w_u and the current input decreases. When $f(x_t, w_u) = f(x_t, w_c)$, the learning rate is exact η. Furthermore, if we fix $p(x_t; w_c, w_r) = 1$, CCCL is actually a generalization of the CCL. Hereinafter, we will measure the distance between two points by using the Euclidean distance only, i.e., Eq.6 can be specified as:

$$p(x_t; w_c, w_u) = \frac{\|x_t - w_c\|}{\max(\|x_t - w_u\|, \varphi\|x_t - w_c\|)}, \quad (7)$$

By embedding the cooperation controlled mechanism into cooperator's updating, the CCCL algorithm can be specified as follows:

Step 1: Pre-specify the number k of clusters with $k \geq k^*$, initialize the seed points $\{w_j\}_{j=1}^{k}$, and set the winning rate of seed $n_j = 1$ with $j = 1, 2, \cdots, k$.

Step 2: Randomly take an input x_t from the input data set, calculate $I(j \mid x_i)$ using Eq.1 and find out the winner w_c.

Step 3: Calculate the cooperating set C by Eq.3.

Step 4: Update all members in C by

$$w_u^{new} = w_u^{old} + p(x_t; w_c, w_u)\eta(x_i - w_u^{old}) \quad (8)$$

$$= w_u^{old} + \frac{\|x_t - w_c\|}{\max(\|x_t - w_u\|, \varphi\|x_t - w_c\|)}\eta(x_i - w_u^{old})$$

Step 5: update n_c by Eq.5.

Step 2, 3, 4, 5 are repeatedly iterated for each input until all seed points converge. The condition of break of iteration can also be pre-assigned as $\|w_u^{new} - w_u^{old}\| < \varepsilon$, with ε being a small positive constant.

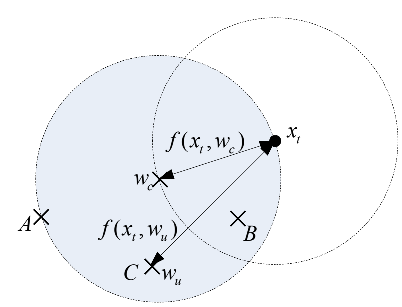


Fig.1. The cooperation controlled mechanism in the CCCL. x_t is the current input, w_c is the winner and w_u is a cooperator. The shadow circle is w_c's cooperative region.

3. Experimental Results

In this section, we will use Gaussian mixture data set to compare the performance of CCCL, CCL and RPCCL respectively. The clusters will be either homogeneous or heterogeneous with the seeds distributed un-uniformly.

3.1 Gaussian mixture clustering with the seeds initialized un-uniformly

We random generated 2,000 data points from a mixture of three 2-dimension Gaussians densities:

$$p(x;\Theta) = 0.3G(x \mid \mu_1, 0.1I) + 0.4G(x \mid \mu_2, 0.1I) + 0.3G(x \mid \mu_3, 0.1I) \quad (9)$$

with $\mu_1 = [1,1]^T$, $\mu_2 = [1,5]^T$ and $\mu_3 = [5,5]^T$ .Where I i s a 2×2 identity matrix, T is a transpose operation of a matrix, and G denotes the Gaussian probability density function of x with the mean μ and co-variance $\sum$ .As shown in Fig.2a, the data from three well-separated clusters. We set the learning rate $\eta = 0.001$ for all three algorithms and $\varphi = 0.5$ for CCCL, and initialized the positions of six seed points $m_1, m_2, \cdots, m_6$ in the input space located at:

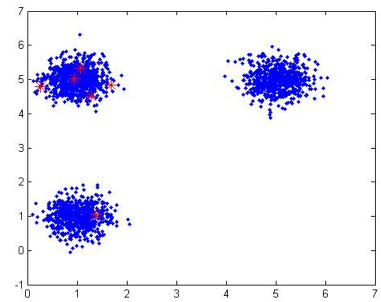
$w_1 = (1.3734, 1.0351)^T$, $w_2 = (0.9392, 5.0324)^T$, $w_3 = (0.2688, 4.7865)^T$, $w_4 = (1.6822, 4.8252)^T$, $w_5 = (1.2882, 4.5142)^T$, $w_6 = (1.0677, 5.3321)^T$,

with one cluster has no seed points at all. As shown in Fig.2b, after 75 CCCL learning epochs, all seed points have been converged to three cluster centers respectively. A snapshot value of convergent seed points are:

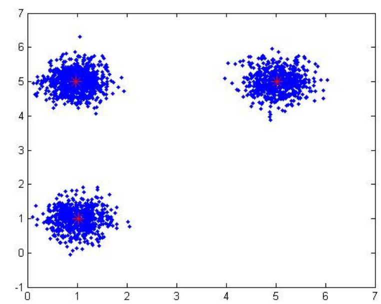
$w_1 = (1.0191, 0.9907)^T$, $w_2 = (0.9808, 4.9944)^T$, $w_3 = (0.9808, 4.9944)^T$, $w_4 = (5.0310, 4.9913)^T$, $w_5 = (5.0310, 4.9913)^T$, $w_6 = (5.0310, 4.9913)^T$,

with their relative winning frequency γ_j s being: $\gamma_1 = 0.3090$, $\gamma_2 = 0.1920$, $\gamma_3 = 0.1920$, $\gamma_4 = 0.1023$, $\gamma_5 = 0.1023$ and $\gamma_6 = 0.1023$.

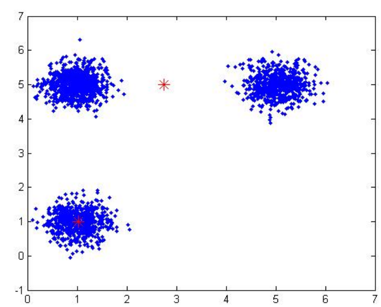
In which we can find that w_1 locates at the first cluster center. w_2 and w_3 converge to the second cluster center. Meanwhile, w_4 , w_5 and w_6 converge to the third cluster center simultaneity. Furthermore, γ_1 is exactly the estimate of the prior probability of $G(x \mid \mu_1, 0.1I)$, the summation of γ_2 and γ_3 is actually an estimate of the prior



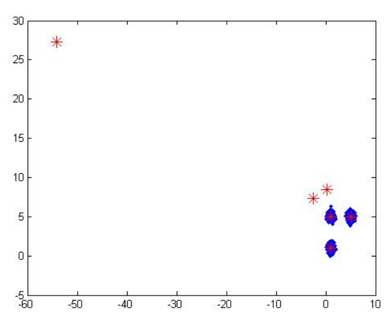
(a)



(b)



(c)



(d)

Fig.2. Gaussian mixture clustering with the seeds initialized un-uniformly. (a) The initial positions of six seed points, (b) the final positions obtained via CCCL, (c) the final positions obtained via CCL, (d) the final positions obtained via RPCCL.

probability of $G(x \mid \mu_2, 0.1I)$. Whereas, the summation of γ_4, γ_5 and γ_6 is the estimate of the prior probability of $G(x \mid \mu_3, 0.1I)$. Obviously, the experimental result is consistent with theory analysis.

But by using CCL, only w_1 was located at a cluster center with the other seed points located at the boundary between the other two clusters. Finally, the CCL let the seed points to:

$$w_1 = (1.0191, 0.9907)^T, \quad w_2 = (2.7437, 4.9905)^T,$$
$$w_3 = (2.7437, 4.9905)^T, \quad w_4 = (2.7437, 4.9905)^T,$$
$$w_5 = (2.7437, 4.9905)^T, \quad w_6 = (2.7437, 4.9905)^T 。$$

After 120 RPCCL epochs, the six seed points learned by RPCCL have been driven to:

$$w_1 = (1.0191, 0.9907)^T, \quad w_2 = (0.9781, 4.9959)^T,$$
$$w_3 = (-54.179, 27.289)^T, \quad w_4 = (5.0310, 4.9913)^T,$$
$$w_5 = (-2.5013, 7.3612)^T, \quad w_6 = (0.2422, 8.4403)^T 。$$

As shown in Fig.2d, the RPCCL has successfully put w_1 and w_4 to the first and the third cluster centers respectively. w_2 was located at the position slightly deviating from the center of the second cluster. Meanwhile, w_3, w_5 and w_6 were driven far away from the input dense regions.

It can be seen that besides CCL, both CCCL and RPCCL can finally work well in this case, but the RPCCL needs more computing cost than CCCL and its locating precision is slightly worse than CCCL.

3.2 Heterogeneous Gaussian mixture clustering

In this experiment, we further compare the performance of CCCL, CCL and RPCCL on the data set, in which the clusters are heterogeneous and seriously overlapped. As shown in Fig.3a. The probability distribution of the input data set is a mixture of four 2-dimension Gaussians densities:

$$p(x;\Theta) = 0.1G(x \mid \mu_1, \varepsilon_1) + 0.5G(x \mid \mu_2, \varepsilon_2) + 0.1G(x \mid \mu_3, \varepsilon_3) + 0.3G(x \mid \mu_4, \varepsilon_4) \quad (13)$$

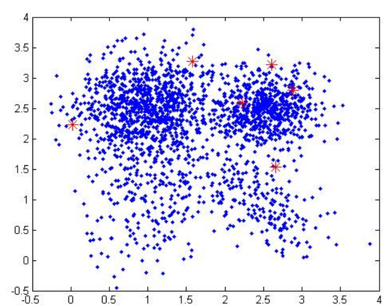
with $\mu_1 = [1.0, 1.0]^T$ $\mu_2 = [1.0, 2.5]^T$, $\mu_3 = [2.5, 1.0]^T$ $\mu_4 = [2.5, 2.5]^T$. Where

$$\varepsilon_1 = \begin{pmatrix} 0.20 & 0.05 \\ 0.05 & 0.30 \end{pmatrix}, \varepsilon_2 = \begin{pmatrix} 0.20 & 0.00 \\ 0.00 & 0.20 \end{pmatrix},$$
$$\varepsilon_3 = \begin{pmatrix} 0.20 & -0.1 \\ -0.1 & 0.20 \end{pmatrix}, \varepsilon_4 = \begin{pmatrix} 0.10 & 0.00 \\ 0.00 & 0.10 \end{pmatrix}.$$

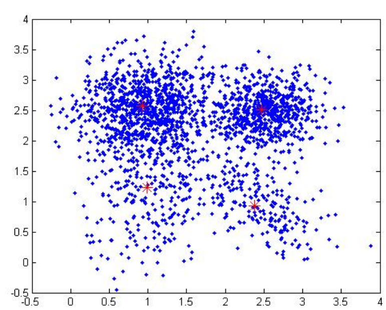
We set the parameters as Section 3.1 and initialized the positions of six seed points $m_1, m_2, \cdots, m_6$ in the input space at:

$$w_1 = (2.2185, 2.5911)^T, \quad w_2 = (1.5739, 3.2708)^T,$$
$$w_3 = (0.0248, 2.2265)^T, \quad w_4 = (2.8808, 2.7922)^T,$$
$$w_5 = (2.6522, 1.5366)^T, \quad w_6 = (2.6059, 3.2119)^T.$$

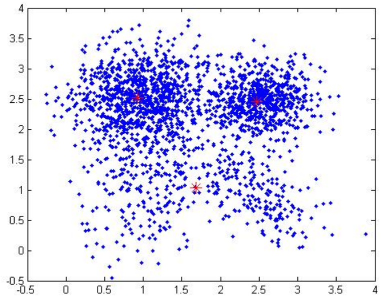
As shown in Fig. 3b, 3c and 3d, After 50 CCCL learning epochs, six seeds have converged to four cluster centers successfully. However, by using CCL, only two dense clusters have been detected with the other two seeds converging to the boundary between the two sparse clusters. In this case, after 200 epochs, RPCCL has driven three seed points out of the input dense regions with the other two seeds located biased from two cluster centers respectively and the last seed located at the boundary between two sparse clusters. The experimental result has shown that CCL and RPCCL are not suitable for heterogeneous clusters at all. Nevertheless, CCCL is applicable to heterogeneous clusters, even if the clusters are overlapped seriously.



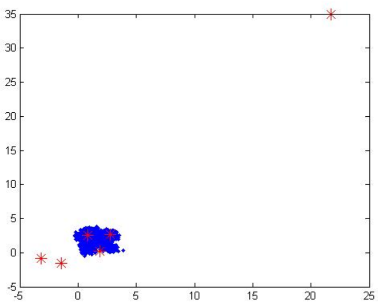
(a)



(b)



(c)



(d)

Fig.3. Gaussian mixture clustering with heterogeneous clusters. (a) The initial positions of six seed points, (b) the final positions obtained via CCCL, (c) the final positions obtained via CCL, (d) the final positions obtained via RPCCL.

4. Conclusion

Based on the semi-competitive learning mechanism of *C*CL, this paper presents an improved competitive algorithm, named Cooperation Controlled Competitive Learning (CCCL), in which the learning rate of each seed points within the same cooperative team can be adjusted adaptively. CCCL is insensitive to the initialization the seed points and applicable to the heterogeneous clusters with an attractive accurate convergence property. Experiments on Gaussian mixture clustering have shown the efficacy of CCCL. Moreover, in some case its performance is prior to CCL and some other variants of RPCL. Next, we will study on the theory analysis of CCCL.

References

[1] E.W. Forgy. Cluster analysis of multivariate data: Efficiency versus interpretability of classifications [J]. Biometrics, 1965, 21(3), 768-769.

[2] L.P. Jing, M.K. Ng, J.Z. Huang. An Entropy Weighting k-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data [J]. IEEE Trans. Knowl. Data Eng., 2007, 19(8):1026-1041.

[3] Z.Y. Zhang, Y.M. Cheung. On Weight Design of Maximum Weighted Likelihood and an Extended EM Algorithm [J]. IEEE Trans. Knowl. Data Eng., 2006, 18(10):1429-1434.

[4] Y.M. Cheung. Maximum Weighted Likelihood via Rival Penalized EM for Density Mixture Clustering with Automatic Model Selection. IEEE Trans. Knowl. Data Eng [J]. 2005, 17(6):750-761.

[5] J.B. MacQueen. Some Methods for Classification and Analysis of Multivariate Observations [C]. Proc. Fifth Berkeley Symp. Math. Statistics and Probability, Berkeley, Calif., USA, 1967, 1, 281-297.

[6] D. E. Rumelhart, D. Zipser. Feature discovery by competitive learning [J]. Cognitive Science, 1985, 9(1):75-112.

[7] S. Grossberg. Competitive learning: from iterative activation to adaptive resonance [J]. Cognitive Science, 1987, 11(1):23-63.

[8] Yiu-ming Cheng. On Rival Penalization Controlled Competitive Learning for Clustering with Automatic Cluster Number Selection [J]. IEEE Trans. Knowl. Data Eng., 2005, 17 (11):1583-1588.

[9] S.C. Ahalt, A.K. Krishnamurty, P. Chen, et al. Competitive Learning Algorithms for Vector Quantization [J]. Neural Networks, 1990, 3(3):277-290.

[10] L. Xu, A. Krzyzak, E. Oja. Rival Penalized Competitive Learning for Clustering Analysis, RBF Net, and Curve Detection [J]. IEEE Trans. Neural Networks, 1993, 4(4):636-648.

[11] Y.M. Cheung, L. Xu. An RPCL-based Approach for Identification of Markov Model with Unknown State Number [J]. IEEE Signal Processing Letters, 2000, 7(10):284-287.

[12] Y.M. Cheung, L. Xu. Rival Penalized Competitive Learning Based Approach for Discrete-Valued Source Separation [J]. Intl. J. of Neural Systems, 2000, 10(6):483-490.

[13] Y.M. Cheung, L. Xu. Rival Penalized Competitive Learning Based Separator on Binary Sources Separation [C]. ICONIP, Kitakyushu, Japan, 1998, 903-906.

[14] Y.M. Cheung, W.M. Leung, L. Xu. Adaptive Rival Penalized Competitive Learning and Combined Linear Predictor Model for Financial Forecast and Investment [J]. Intl. J. Neural System, 1997, 8(5-6):517-534.

[15] Y.M. Cheung. Rival Penalization Controlled Competitive Learning for Data Clustering with Unknown Cluster Number [C]. Proc. 9th Intl. Conf. Neural Information Processing, Vancouver, British Columbia, Canada, 2002, 18-22.

[16] J. W. Ma, T.J Wang. A cost-function approach to rival penalized competitive learning (RPCL) [J]. IEEE Trans. Systems, Man, and Cybernetics, 2006, 36(4):722-737.

[17] Y.M. Cheung. A competitive and cooperative learning approach to robust data clustering [C]. Proc. IASTED Intl. Conf. Neural Networks and Computational Intelligence, Grindelwald, Switzerland, 2004, 131-136.