An Algorithm of Fuzzy Collaborative Clustering based on Kernel Competitive Agglomeration

Cui-Fang GAO
School of Science, Jiangnan University, Wuxi 214122, PR China
Cuifnag_gao@163.com

Xiao-Jun WU
School of Computer Science and Technology, Jiangnan University, Wuxi 214122, PR China

Philip YU
Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam, Hong Kong

Abstract—Kernel-based clustering generally maps the observed data to a high dimensional feature space and can usually achieve preferable classification by enlarging the difference among samples. Competitive kernel clustering creates a competitive environment by means of hierarchical method in which clusters compete for samples based on cardinalities in kernel space. Collaborative clustering implementing on several subsets can be processed by one objective function, which improves the clustering performance by sharing partition matrices among different subsets. In this paper an improved algorithm of collaborative competitive kernel clustering analysis (CCKCA) is proposed, in which the mechanism of collaboration is introduced into competitive kernel clustering. Exploiting the advantages of basic algorithms, CCKCA makes full use of the knowledge of collaborative relation among different subsets based on kernel competitive clustering. The results obtained on the benchmark datasets show that CCKCA can achieve approving clustering performance.

Index Terms—kernel methods, competitive clustering, kernel fuzzy c-means clustering, collaborative clustering

I. INTRODUCTION

Clustering is commonly viewed as a powerful tool for classification and data analysis, the main task of clustering is to separate N samples from a dataset into C groups according to a certain similarity measure, each group is called a cluster. Objective function-based clustering algorithms, which are widely used in the field of pattern recognition[1], deal with the optimization problem by deriving the iterative solution of objective function with constraints. Fuzzy clustering reflects the ambiguous characteristic of data points in their objective function and develops to be a popular and useful method. The fuzzy C-means clustering (FCM) [2] is exactly such typical algorithm which has undergone comprehensive studies and developed a variety of modified version. Clustering by competitive agglomeration (CA) [3] stems from FCM which suggests an additively competitive term in the objective function of FCM, and thus it can automatically control the number of clusters, but CA perform clustering directly in the original observed space even on nonlinearly complicated datasets. An alternative method for nonlinear classification is to map the observed data to a high dimensional feature space with a smooth and continuous nonlinear function [4,5]. After the data is mapped into the feature space, difference among samples can be extended and the structure of dataset can be simplified, and consequently it might be possible to find clearer separation in the high dimensional space. The primary advantage of kernel methods is that nonlinear problems can be transformed into linear problems, which is applicable for more complicated datasets.

Many algorithms perform on a complete dataset with all features, but a meaningful and feasible idea is separating the dataset into several subsets which can describe the group of samples in different feature spaces, and then concerning the interaction of partition matrices based on these subsets. Collaborative clustering [8-11] introduces a concept of collaborative mechanism which improved the clustering performance by exchanging partition information among multiple subsets. The modified objective function with an additional term expressed the effect of collaborative relation is used to guide the overall optimization activities. Knowledge of these relationships will be beneficial in discovering the meaningful structure and making the results more stable. The intensity of collaboration between different subsets is established through a matrix of interaction coefficients [8], and the final collaborative membership matrix can reflect a total effect of all datasets from multiple feature spaces, therefore it usually generate relative accurate partition matrix. Based on above researches, here the kernel method is introduced into competitive clustering algorithm, and inspired by the idea of collaboration, we propose the new collaboratively competitive kernel clustering algorithm (CCKCA) in order to obtain preferable performance by accommodating both the advantages of each individual techniques.

This paper is organized as follows: Section 2 briefly describes the clustering by competitive agglomeration
A. Clustering by Competitive Agglomeration

CA is defined as follows:

$$J_{CA} = \sum_{k=1}^{C} \sum_{i=1}^{N} (u_{ik})^2 d^2(x_i, v_k) - \beta \sum_{k=1}^{C} \sum_{i=1}^{N} u_{ik}$$

Subject to the constraints:

$$\sum_{k=1}^{C} u_{ik} = 1 \quad i = 1, 2, ..., N$$

where $C$ is the number of clusters, $v_k = (v_{k1}, v_{k2}, ..., v_{kl})$ is the $k$th cluster center, $u_{ik}$ denotes the membership degree of the $ith$ sample belonging to the $k$th cluster, $d^2(x_i, v_k)$ indicates the square distance between the $ith$ sample and the $k$th cluster center. $N_i = \sum_{k=1}^{C} u_{ik}$ is the cardinality of the $k$th cluster expressed as the sum of membership degrees of the $k$th cluster.

The choice of $\beta$ is important in CA since it reflects the importance of the second term (competition component) relative to the first term (FCM function) in Eq. (1). In paper [3], $\beta$ was chosen as the function of $t$:

$$\beta(t) = \eta_{\beta} \exp(-\nu t) \sum_{k=1}^{C} \sum_{i=1}^{N} (u_{ik})^2 d^2(x_i, v_k)$$

where $t$ is the iteration number, $\beta$ varies with $t$ during the clustering process. $\eta_{\beta} \exp(-\nu t)$ is an exponential decay function in which the initial value $\eta_{\beta}$ will effect on the convergent speed of CA. The algorithm can starts with a large value of $\beta$ so that the competition term of the objective function dominates in the early iterations. Then the value of $\beta$ decreases slowly to help CA to seek the approximate "optimum" cluster number. When it is closed to the appropriate partitions, $\beta$ becomes small and the first term will be emphasized.

B. Kernel Methods

Clustering by kernel methods has the advantage which performs clustering in the mapped space. Suppose $\phi(\cdot)$ is a nonlinear mapping function to achieve $\phi : x \rightarrow \phi(x)$, where $x$ denotes an element in the input space and $\phi(x)$ is the corresponding element in the high dimensional mapped space. With kernel methods, the calculation of scalar inner product can be transformed into kernel calculation:

$$|\phi(x_i) - \phi(v_k)|^2 = |\phi(x_i) - \phi(v_k) + \phi(v_k) - \phi(v_k) - \phi(x_i)|^2$$
$$= K(x_i, x_i) + K(v_k, v_k) - 2K(x_i, v_k)$$

where $K(x, y) = \phi(x) \cdot \phi(y)$ is a predefined kernel function, which can be represented with a dot product in the high dimensional mapped space. If the commonly used Gaussian kernel function was adopted:

$$K(x, y) = \exp\left(-\frac{||x - y||^2}{\sigma^2}\right)$$

then $K(x, x) = 1$, and then Eq. (4) becomes:

$$\|\phi(x_i) - \phi(v_k)\|^2 = 2(1 - K(x_i, v_k))$$

where $\| \cdot \|$ can be viewed as a kernel-induced metric, paper [4] gives the theoretical proof that the similarity defined in Eq. (6) is a kernel-induced metric in the original data space.

If the distance $d^2(x_i, v_k)$ in CA algorithm takes the metric as Eq. (6), the competitive kernel clustering will be developed, and then Eq. (1) and Eq. (3) become correspondingly:

$$J_{Kernel-CA} = \sum_{k=1}^{C} \sum_{i=1}^{N} (u_{ik})^2 \|\phi(x_i) - \phi(v_k)\|^2 - \beta \sum_{k=1}^{C} \sum_{i=1}^{N} u_{ik}$$

$$\beta(t) = \eta_{\beta} \exp(-\nu t) \sum_{k=1}^{C} \sum_{i=1}^{N} (u_{ik})^2 d^2(x_i, v_k)$$

III. COLLABORATIVELY COMPETITIVE KERNEL CLUSTERING ANALYSIS (CCKCA)

Collaborative fuzzy clustering implementing on several subsets can be processed by one objective function, which introduces a novel collaborative concept and improves the clustering performance by communicating partition matrices among multiple feature subsets. Collaborative clustering consider $p$ subsets of data described as different features and every subset concerns the same group of samples, so the number of samples in each subset is the same while the dimensionality of the patterns could be different.

The result of clustering obtained on each subset can be presented in the form of a partition matrix and prototypes. The collaboration relationship between these subsets is established through a matrix of interaction coefficient, see Fig.1, where a specific subset is identified as a symbol notation, for example $\Lambda$, $\Gamma$ and $\Phi$ (subsets of $\Lambda$, $\Gamma$, $\Phi$, and so on, are given serial number as 1, 2, ..., $p$).
Collaborative coefficients should assume nonnegative values, which describe the interaction intensity. The higher value of the interaction coefficient is, the stronger collaboration between the corresponding subsets, and also the stronger collaboration effect on the final partition matrix.

A. Description of CCKCA

In this section, we introduce the collaborative scheme into competitive kernel clustering to develop a new algorithm CCKCA, in which the interaction of multiple partitions serving as an additional collaborative function is defined based on a kernel-induced metric and thus an integrated objective function is reformulated. The new collaborative function is defined to be compatible with the basic competitive kernel clustering and all terms of the objective function can achieve a good cooperation. If we don’t doing so, even the collaborative concept is established, the optimization problem will incur difficulty in case of different terms performing in different feature spaces. The objective function of CCKCA is defined as follows:

\[
J'_{CCKCA} = \sum_{k=1}^{C} \sum_{i=1}^{N} (u_{ik}^\Gamma)^2 (d_{ik}^\Gamma)^2 - \beta \sum_{k=1}^{C} \sum_{i=1}^{N} (u_{ik}^\Gamma)^2 + \sum_{\phi \in [\Phi]} \sum_{i=1}^{N} \sum_{r=1}^{N_\phi} \alpha^{(\phi, r)} (u_{ir}^\phi - u_{ir}^\phi)^2 (d_{ir}^\phi)^2 \tag{9}
\]

where \( \phi \) denotes the collaborative coefficient.

Subject to the same constraints as Eq. (2). Here super script notation (i.e. \( \phi, \Gamma \)) is used to identify the specific subset, \( \alpha^{(\phi, \Gamma)} \) denotes the collaborative coefficient between subset \( \Gamma \) and subset \( \phi \). The symbol notation in bracket (i.e. \([ \phi, \Gamma ]\)) indicates the serial number of the specific subsets. Where the distance between the \( kth \) cluster center and the \( ith \) samples in subset \( \Gamma \) is defined as the kernel-induced metric:

\[
(d_{ik}^\Gamma)^2 = \| \phi(x_i) - \phi(v_k) \|^2 = 2(1 - K(x_i, v_k))
\]

As described above, CCKCA consists of three components. The first term in eq. (9) is the kernel clustering algorithm with alternative new metric, the second term is the competitive function stem from CA, and the third term is the new additive collaborative function expressed as the membership subtraction for detecting the difference among subsets. When searching for the optimal value of the objective function, CCKCA establishes some interaction and allows all subsets engage in the form of communication and reconciliation of their partitions.

To minimize the objective function of \( J_{CCKCA} \) in Eq. (9), Lagrangian multiplier method is used to get the following iterative equations of membership values and cluster centers (refer to the Appendices for the detailed derivations).

Iterative equations of membership values is given as:

\[
u_{ir}^\Gamma = \frac{1}{\sum_{\phi \in [\Phi] \setminus [\Gamma]} \phi_{ir}^\phi} \sum_{\phi \in [\Phi] \setminus [\Gamma]} \phi_{ir}^\phi
\]

Iterative equations of cluster centers is obtained as:

\[
v_{ir}^\Gamma = \frac{1}{\sum_{i=1}^{N} (u_{ir}^\phi)^2 K(x_i, v_{ir}^\phi)} \sum_{i=1}^{N} (u_{ir}^\phi)^2 K(x_i, v_{ir}^\phi) K(x_i, v_{ir}^\phi) x_i^\Gamma
\]
through a simultaneous optimization of the partition matrices with the values of interaction coefficient according to the iterative scheme. CCKCA algorithm can be summarized as follows:

**Input:** given \( p \) data sets and the samples are expressed as feature vectors.

**Output:** \( C \) clusters of input data.

**Step1:** Initialization: specify the maximum number of clusters, initialize parameter \( \sigma^2 \) of kernel function and collaborative coefficient matrix \( \mathbf{A}^{r,\phi} \), randomly generate initial vectors \( \mathbf{V}^r \), and averagely generate membership matrix \( ([J]_{12} \ldots p \ [DF]_{12} \ldots p \ [IF]_{12} \ldots p \ [DF]) \).

**Step2:** Normalize the original samples in each subset, obtain independent partition matrices for each subset using the competitive kernel clustering algorithm.

**Step3:** reprocess independent partition matrices of different data sets for cluster matchup.

**Step4:** Compute the values of \( \beta \) using Eq. (8).

**Step5:** Update the membership values using Eq. (10).

**Step6:** Compute the cardinalities of all clusters:

\[
N_s = \sum_{i=1}^{N} n_i \quad (s=1,2,\ldots, C).
\]

**Step7:** Update the cluster centers using Eq. (11).

**Step8:** Repeat Step4 to Step7, until the termination condition is met:

\[
|J^{CCKCA}(U^{C})-J^{CCKCA}(U^{I-1})| \leq \varepsilon \quad (\varepsilon \text{ is a predefined small positive constant}).
\]

**B. The Convergent Property of CCKCA**

**Theorem 1**

Let \( \beta \) and \( N_s^t \) be fixed, \( U^r = [u^r_s]_{s \in [1, N]} \) for \( s \in [1, C] \), \( t \in [1, N] \), \( \| \phi(x^l_s) - \phi(v^r_s) \| > 0 \), then \( U^r \) is a local minimum of \( J^{CCKCA}(U^{C}) \) if and only if \( u^r_s \) is computed from eq. (10).

**Proof:** The only-if part has been proven by the derivation of \( u^r_s \). To prove its sufficiency, we examine the Hessian matrix \( H(U^r) \) of the Lagrangian equation of \( J^{CCKCA}(U^{C}) \).

The partial derivative of \( J^{CCKCA}(U^{C}) \) with respect to \( u^r_s \) is given by

\[
\frac{\partial J^{CCKCA}}{\partial u^r_s} = 2u^r_s (d^r_s)^2 - 2\beta N^r_s + 2 \sum_{[\phi]|\phi|\phi} \mathbf{A}^{r,\phi} (u^r_s - u^r_i \phi) (u^r_s - u^r_i \phi) - \lambda = 0
\]

where \( N^r_s = \sum_{i=1}^{N} u^r_i \)

then, the second partial derivative of \( J^{CCKCA} \) is:

\[
\frac{\partial^2 J^{CCKCA}}{\partial u^r_s} = \frac{\partial}{\partial u^r_s} \left[ \frac{\partial J^{CCKCA}}{\partial u^r_s} \right] = 2(d^r_s)^2 + 2 \sum_{[\phi]|\phi|\phi} \mathbf{A}^{r,\phi} (d^r_s)^2 > 0 \quad \text{if} \quad s = k, t = i
\]

\[= 0 \quad \text{otherwise}
\]

That is to say \( H(U^r) = h_{u,u}(U^r) \) is a diagonal matrix and Hessian \( H(U^r) \) is positive definite, consequently eq. (10) is a sufficient condition for minimizing \( J^{CCKCA}(U^{C}) \).

**IV. NUMERICAL TESTS AND DISCUSSIONS**

The performances of CCKCA are tested on iris and wine datasets respectively, which can be obtained from UCI Machine Learning Repository [13]. Iris dataset contains three classes of fifty samples each, and every sample is denoted as a vector with 4 features. Wine dataset contains 178 samples of 13 features each, this version is the chemical analysis of three kinds of wines. To facilitate description and comparison, here we use the abbreviation CKCA for the competitive kernel clustering algorithm and HC-FCM for the collaborative clustering in paper [8], some values of primary parameters are shown in table I. The three algorithms HC-FCM, CKCA, CCKCA are separately implemented on the above two datasets, their comparative results are shown in table II. It should be noted that there is no collaboration in CKCA, whereas HC-FCM and CCKCA are collaborative clustering and based on the same feature subsets.

<table>
<thead>
<tr>
<th>datasets</th>
<th>features of subsets</th>
<th>values of parameters</th>
<th>clustering accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>iris</td>
<td>1, 3, 4</td>
<td>( \alpha = 0.5 )</td>
<td>87.3</td>
</tr>
<tr>
<td></td>
<td>2, 3, 4</td>
<td>( \sigma = 2.5 ), ( \eta = 0.6 )</td>
<td>91.3</td>
</tr>
<tr>
<td>wine</td>
<td>1, 3, 5, 7, 9, 11, 13</td>
<td>( \alpha = 0.3 )</td>
<td>96.6</td>
</tr>
<tr>
<td></td>
<td>2, 4, 6, 8, 10, 12</td>
<td>( \sigma = 5 ), ( \eta = 0.2 )</td>
<td>97.2</td>
</tr>
</tbody>
</table>

Our primary intention is to objectively examine the effectiveness of collaboration on the basis of communication. As the clustering results shown in table II, the performance of CCKCA has obtained acceptable improvement compared with the algorithms of CKCA in which no collaboration has been established. Although a relatively high accuracy on wine dataset and a lower one on iris dataset, here an undeniable implication is presented: the collaborative optimization activities perform its functionality in consistency. In other words, the collaboration has led to improvements of clustering performance.

Fig. 3 to Fig. 5 visualized the clustering results of HC-FCM, CKCA and CCKCA in 2-dimensional space (by linear map method we can get 2D data from the original 4D input [19]). The iris data set contains 3 classes of 50 instances each, one class is linearly separable from the other two, but the latter two are nonlinear and many samples are overlapping, so it is difficult for them to separate from each other. It can be seen, compared to the original distribution in fig. 2, the separation results by
CCKCA is the best among the three algorithms.

We use the results of CKCA without collaboration as points of reference, and find that the centers as well as membership matrices are affected by the collaboration. Fig.6 reveals the changes of the membership degree between CKCA and CCKCA. It can be seen that there are universal effect on most samples as the results of collaboration, especially some samples in cluster 2 or cluster 3 change their membership values significantly. Obviously all these changes are caused by the collaboration mechanism and resultingly CCKCA can induce more exact result.

The reasonable change of membership incurred by the collaboration is signed, a positive value will enhances the membership of sample belonging to the specify cluster, while a negative one will reduces the membership. Table III lists some specific samples that change to the proper cluster by increasing or reducing their membership values. So a commendable advantage of CCKCA is to adjust the inaccurate memberships by collaboration so as to achieve exact partition.

Three centers of iris dataset obtained by CKCA and CCKCA are comparatively shown in Table IV. Under the same initial condition, the two algorithms gave different cluster centers. It can be found that CCKCA suggested much larger inter-cluster distance of separation especially between the centers of class2 and class3. So centers obtained by CCKCA are meaningful and as well they are better representatives of the overall dataset structure compared to that by CKCA.

| TABLE III |
| Adjustments of Collaboration Imposing on The Final Membership Degree Values |

<table>
<thead>
<tr>
<th>samples</th>
<th>classes</th>
<th>CKCA membership degree values</th>
<th>CCKCA membership degree values</th>
<th>the adjusted values</th>
</tr>
</thead>
<tbody>
<tr>
<td>52</td>
<td>3</td>
<td>0.0996 0.3247 0.5757</td>
<td>2.0.0926 0.4945 0.4128</td>
<td>-0.0070 0.1698 -0.1628</td>
</tr>
<tr>
<td>66</td>
<td>3</td>
<td>0.1005 0.3185 0.5810</td>
<td>2.0.0940 0.5354 0.3706</td>
<td>-0.0065 0.2170 -0.2105</td>
</tr>
<tr>
<td>76</td>
<td>3</td>
<td>0.0898 0.3983 0.5119</td>
<td>2.0.0861 0.6561 0.2758</td>
<td>-0.0217 0.3578 -0.2361</td>
</tr>
<tr>
<td>77</td>
<td>3</td>
<td>0.0935 0.3951 0.5113</td>
<td>2.0.0393 0.7425 0.2182</td>
<td>-0.0542 0.3474 -0.2931</td>
</tr>
<tr>
<td>86</td>
<td>3</td>
<td>0.1677 0.3590 0.4733</td>
<td>2.0.1132 0.5172 0.3697</td>
<td>-0.0545 0.1582 -0.1036</td>
</tr>
<tr>
<td>122</td>
<td>2</td>
<td>0.0951 0.6237 0.2812</td>
<td>3.0.0658 0.4387 0.4954</td>
<td>-0.0292 -0.1850 0.2142</td>
</tr>
<tr>
<td>139</td>
<td>2</td>
<td>0.0700 0.4810 0.4490</td>
<td>3.0.0424 0.4125 0.5451</td>
<td>-0.0276 -0.0685 0.0961</td>
</tr>
<tr>
<td>150</td>
<td>2</td>
<td>0.0764 0.4850 0.4386</td>
<td>3.0.0370 0.3328 0.6302</td>
<td>-0.0395 -0.1522 0.1916</td>
</tr>
</tbody>
</table>
Fig. 7 and Fig. 8 report the changed distribution of the membership degree obtained by CKCA and CCKCA respectively (where memberships of class 1 are denoted by real line, class 2 by broken line and class 3 by dash dot line). It can be seen that the partition structure of CCKCA is more clear-cut contrasting with CKCA in which no collaboration was considered, and therefore CCKCA can achieve relative accurate results.

The experimental results presented above are all based on the parameters in Table I, but the effect of CCKCA driven by the collaboration mechanism may depend on the multiple subsets with different collaborative coefficients or feature subsets. Here the performance of CCKCA is studied by doing experiments with various essential parameters. Table V shows the comparative results of iris dataset with different values of $\alpha$, and Table VI shows the comparative results with different subsets.

The coefficient $\alpha$ implies a certain level of intensity of collaboration and keeps the balance of optimization guided by the collaboration. Table V indicates that higher values of $\alpha$ lead to more changes in the membership values. If the values of $\alpha$ are within the interval of $[0.01, 5]$, CCKCA can achieve some quantifiable enhancement of clustering accuracy, the highest accuracy can be obtained especially in the interval of $[0.1, 1]$. While there is no obvious impact of the collaboration evidence when $\alpha$ less than 0.01, and the membership degree matrices of subsets are on the verge of stableness.

The results in Table VI suggested some different results of CCKCA on different features subsets involving in the collaboration. Since all subsets will act upon each other while running the optimization, clustering accuracy depends on the collaborative activities between subsets. It is not the case that more subsets engaged in collaboration will improve the performance of CCKCA from the results of experiments.

### Table IV.

<table>
<thead>
<tr>
<th>classes</th>
<th>Centers by CKCA</th>
<th>Centers by CCKCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>class1</td>
<td>(-1.0362, 0.7581, -1.2897, -1.2465)</td>
<td>(-1.0241, 0.7816, -1.2929, -1.2493)</td>
</tr>
<tr>
<td>class2</td>
<td>(-0.0026, -0.7383, 0.3438, 0.2431)</td>
<td>(0.0522, -0.7021, 0.3491, 0.2305)</td>
</tr>
<tr>
<td>class3</td>
<td>(0.9887, 0.0023, 0.9344, 1.0245)</td>
<td>(1.0562, -0.0190, 0.9922, 1.0989)</td>
</tr>
</tbody>
</table>

Table V.

<table>
<thead>
<tr>
<th>coefficients</th>
<th>clustering accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha &lt; 0.01$</td>
<td>88</td>
</tr>
<tr>
<td>$\alpha \in [0.01, 0.1]$</td>
<td>89.3</td>
</tr>
<tr>
<td>$\alpha \in [0.1, 1]$</td>
<td>92</td>
</tr>
<tr>
<td>$\alpha \in [1, 5]$</td>
<td>91.3</td>
</tr>
<tr>
<td>$\alpha &gt; 5$</td>
<td>88.7</td>
</tr>
</tbody>
</table>

### Table VI.

<table>
<thead>
<tr>
<th>subset 1</th>
<th>subset 2</th>
<th>subset 3</th>
<th>subset 4</th>
<th>clustering accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 3, 4</td>
<td>2, 3, 4</td>
<td>--</td>
<td>--</td>
<td>91.3</td>
</tr>
<tr>
<td>2, 3, 4</td>
<td>1, 2, 3</td>
<td>--</td>
<td>--</td>
<td>89.3</td>
</tr>
<tr>
<td>1, 2, 3</td>
<td>1, 3, 4</td>
<td>2, 3, 4</td>
<td>--</td>
<td>93.3</td>
</tr>
<tr>
<td>1, 2, 4</td>
<td>1, 3, 4</td>
<td>2, 3, 4</td>
<td>--</td>
<td>91.3</td>
</tr>
<tr>
<td>1, 2, 3</td>
<td>1, 3, 4</td>
<td>1, 2, 4</td>
<td>2, 3, 4</td>
<td>90.7</td>
</tr>
</tbody>
</table>

V. Conclusions

We developed a new clustering algorithm named CCKCA that introduced collaborative mechanism into competitive clustering based on kernel method. The new algorithm is essentially the optimization which was collaboratively processed with an objective function on multiple feature subsets in kernel feature space. Its improved performance has been shown on benchmark datasets. An important aspect of CCKCA is that it is implemented on different subsets and therefore its effectiveness will keep a close relationship with the number of subsets and the combination of features in each subset. So a great challenge in the future involves searching the optimal relevant features in a subset for a general preferable performance. Another important aspect of CCKCA is that it maybe achieves potential better performance when using different algorithm on individual subset according to its specific characteristics of data structure, rather than using the same competitive kernel clustering algorithm on all subsets.
ACKNOWLEDGEMENTS

This research was partially supported by Program for Natural Science Foundation of China (Grant No.: 60973094, 11202084), Specialized Research Fund for the Doctoral Program of Higher Education (Grant No.: 20120093120016), Fundamental Research Funds for the Central Universities (Grant No.: JUSRP211A23, JUSRP31103).

REFERENCES


Cuifang Gao received her B.S. degree from Sun Yat-Sen University, Guangzhou, PR China in 1998 and M.S. degree in 2007, and Ph.D. degree in 2011 both from Jiangnan University, Wuxi, PR China. Her current research interests are pattern recognition and computational intelligence.

Xiaojun Wu received his B.S. degree in mathematics from Nanjing Normal University, Nanjing, PR China in 1991 and M.S. degree in 1996, and Ph.D. degree in Pattern Recognition and Intelligent System in 2002, both from Nanjing University of Science and Technology, Nanjing, PR China, respectively. He was a visiting researcher in the Centre for Vision, Speech, and Signal Processing (CVSSP), University of Surrey, UK from 2003 to 2004. He has published more than 100 papers. His current research interests are pattern recognition, computer vision, fuzzy systems and intelligent systems.

APPENDICES

A. Derivations of Eq. (10)

The objective function of CCKCA is transformed into

\[ J_{\text{CCKCA}}^i = \sum_{k=1}^{C} \sum_{i=1}^{N} \left( u_{ik}^r \right)^T d_{ik}^r - \beta \sum_{k=1}^{C} \sum_{i=1}^{N} \left( u_{ik}^r \right)^T d_{ik}^r \]

where \( \alpha^{(r, \phi)} \) denotes the collaborative coefficient between subset \( r \) and subset \( \phi \).

Taking into account the constraints \( \sum_{i=1}^{C} u_{ik}^r = 1 \), with \( \lambda \) denoting the Lagrangian multiplier, we can obtain the following Lagrangian function:

\[ \lambda \left( \sum_{i=1}^{C} u_{ik}^r - 1 \right) + \sum_{k=1}^{C} \sum_{i=1}^{N} \left( u_{ik}^r \right)^T d_{ik}^r \]

\[ \sum_{k=1}^{C} \sum_{i=1}^{N} \left( u_{ik}^r \right)^T d_{ik}^r \]
\[ J_{ccxct}^r = \sum_{s=1}^{C} \sum_{l=1}^{\mathcal{N}} \left( u_{st}^r \right)^2 \left( d_{ls}^r \right)^2 - \beta \sum_{s=1}^{C} \sum_{l=1}^{\mathcal{N}} \left( u_{st}^r \right)^2 - \sum_{s=1}^{C} \sum_{l=1}^{\mathcal{N}} \lambda_s \left( \sum_{s=1}^{C} u_{st}^r - 1 \right) \]  

(A.2)

In subset \( \Gamma \), for \( t \in \{1, \ldots, \mathcal{N} \} \) and \( s \in \{1, \ldots, C \} \), taking the derivative of \( J_{ccxct} \) with respect to \( u_{st}^r \) and setting the result to zero, we have:

\[ \frac{\partial J_{ccxct}^r}{\partial u_{st}^r} = 2u_{st}^r \left( d_{ls}^r \right)^2 - 2\beta N_t^r + \sum_{s=1}^{C} \sum_{l=1}^{\mathcal{N}} \alpha^{(r, \phi)} \left( u_{st}^r - u_{st}^0 \right) \left( d_{ls}^r \right)^2 - \lambda_s = 0 \]  

(1)

where \( N_t^r = \sum_{s=1}^{C} u_{st}^r \)  

(A.3)

Solving (A.3) for \( u_{st}^r \), we have:

\[ u_{st}^r = \frac{\lambda_s + 2\beta N_t^r + \sum_{s=1}^{C} \sum_{l=1}^{\mathcal{N}} \alpha^{(r, \phi)} \left( u_{st}^r - u_{st}^0 \right) \left( d_{ls}^r \right)^2}{2 \left( d_{ls}^r \right)^2 + \sum_{s=1}^{C} \sum_{l=1}^{\mathcal{N}} \alpha^{(r, \phi)} \left( d_{ls}^r \right)^2} \]  

(A.4)

Setting \( \varphi_s^r = \sum_{s=1}^{C} \alpha^{(r, \phi)} \psi_s^r \) and \( \psi^r = \sum_{s=1}^{C} \alpha^{(r, \phi)} \)

Eq. (A.4) can be simplified as:

\[ u_{st}^r = \lambda_s + 2\beta N_t^r + \frac{\varphi_s^r \left( d_{ls}^r \right)^2}{2 \left( d_{ls}^r \right)^2 + \left( 1 + \psi^r \right)} \]  

(A.5)

Considering the constraints:

\[ \sum_{s=1}^{C} \frac{\lambda_s + 2\beta N_t^r + \varphi_s^r \left( d_{ls}^r \right)^2}{2 \left( d_{ls}^r \right)^2 + \left( 1 + \psi^r \right)} = 1 \]  

(A.6)

Solving (A.6) for \( \lambda_s \), we have:

\[ \lambda_s = \frac{1 - \sum_{s=1}^{C} \left( \beta N_t^r + \varphi_s^r \left( d_{ls}^r \right)^2 \right) \left( 1 + \psi^r \right)}{\sum_{s=1}^{C} \left( d_{ls}^r \right)^2 \left( 1 + \psi^r \right)} \]  

Substituting \( \lambda_s \) into (A.5), we can obtain the final membership equation shown as:

\[ u_{st}^r \left( d_{ls}^r \right)^2 \left( 1 + \psi^r \right) = \frac{1 - \sum_{s=1}^{C} \left( \beta \tilde{N}_t^r + \varphi_s^r \left( d_{ls}^r \right)^2 \right) \left( 1 + \psi^r \right)}{\sum_{s=1}^{C} \left( d_{ls}^r \right)^2 \left( 1 + \psi^r \right)} + \frac{\beta N_t^r}{\sum_{s=1}^{C} \left( d_{ls}^r \right)^2 \left( 1 + \psi^r \right)} + \frac{\varphi_s^r \left( d_{ls}^r \right)^2}{\sum_{s=1}^{C} \left( d_{ls}^r \right)^2 \left( 1 + \psi^r \right)} \]  

(A.7)

B. Derivations of Eq. (11)

With Gaussian kernel function \( K(x_s, v_k) = \exp\left(-\frac{\|x_s - v_k\|^2}{\sigma^2}\right) \), we have \( \|\phi(x_s) - \phi(v_k)\|^2 = 2(1 - K(x_s, v_k)) \)

The objective function of CCKCA is transformed into:

\[ J_{ccxct}^r = 2 \sum_{s=1}^{C} \sum_{l=1}^{\mathcal{N}} \left( u_{st}^r \right)^2 \left( 1 - \exp\left(-\frac{\|x_s^r - v_k^r\|^2}{\sigma^2}\right)\right) - \beta \sum_{s=1}^{C} \sum_{l=1}^{\mathcal{N}} \left( u_{st}^r \right)^2 - 2 \sum_{s=1}^{C} \sum_{l=1}^{\mathcal{N}} \alpha^{(r, \phi)} \left( u_{st}^r - u_{st}^0 \right) \left( 1 - \exp\left(-\frac{\|x_s^r - v_k^r\|^2}{\sigma^2}\right)\right) \]  

(B.1)

In subset \( \Gamma \), for \( k \in \{1, \ldots, C \} \), taking the derivative of \( J_{ccxct}^r \) with respect to \( v_k^r \) and setting the result to zero, we have:
\[
\frac{\partial J_{C^2C^4}}{\partial V_i^T} = -\frac{4}{\sigma^2} \sum_{j=1}^{N} \left( u_j^T \right)^{\phi \in \mathcal{N}} \left[ \exp \left( \frac{-\|x_i^T - v_j^T\|^2}{\sigma^2} \right) \right] (x_i^T - v_j^T) + \frac{4}{\sigma^2} \sum_{\phi \in \mathcal{F} \setminus \mathcal{F} \cap \mathcal{G}} \left[ \alpha^{(r, \phi)} \sum_{j=1}^{N} (u_j^T - u_j^0) \right] \left[ \exp \left( \frac{-\|x_i^T - v_j^T\|^2}{\sigma^2} \right) \right] (x_i^T - v_j^T) = 0
\]

Eq. (B.2) can be simplified as:

\[
\sum_{j=1}^{N} \left( u_j^T \right)^{\phi \in \mathcal{N}} K(x_i^T, v_j^T) (x_i^T - v_j^T) + \sum_{\phi \in \mathcal{F} \setminus \mathcal{F} \cap \mathcal{G}} \left[ \alpha^{(r, \phi)} \sum_{j=1}^{N} (u_j^T - u_j^0) \right] K(x_i^T, v_j^T) (x_i^T - v_j^T) = 0
\]  
(B.3)

Solving (B.3) for \( v_i^T \), we have:

\[
v_i^T = \frac{\sum_{j=1}^{N} \left( u_j^T \right)^{\phi \in \mathcal{N}} K(x_i^T, v_j^T) x_i^T + \sum_{\phi \in \mathcal{F} \setminus \mathcal{F} \cap \mathcal{G}} \left[ \alpha^{(r, \phi)} \sum_{j=1}^{N} (u_j^T - u_j^0) \right] K(x_i^T, v_j^T) x_i^T}{\sum_{j=1}^{N} \left( u_j^T \right)^{\phi \in \mathcal{N}} K(x_i^T, v_j^T) + \sum_{\phi \in \mathcal{F} \setminus \mathcal{F} \cap \mathcal{G}} \left[ \alpha^{(r, \phi)} \sum_{j=1}^{N} (u_j^T - u_j^0) \right] K(x_i^T, v_j^T)} \]  
(B.4)

© 2013 ACADEMY PUBLISHER