2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Multiple-Kernel Based Soft Subspace Fuzzy Clustering

Jun Wang, Zhaohong Deng, Yizhang Jiang, Pengjiang Qian and Shitong Wang
School of Digital Media, Jiangnan University Wuxi, China

***Abstract*—Soft subspace fuzzy clustering algorithms have been successfully utilized for high dimensional data in recent studies. However, the existing works often utilize only one distance function to evaluate the similarity between data items along with each feature, which leads to performance degradation for some complex data sets. In this work, a novel soft subspace fuzzy clustering algorithm *MKEWFC-K* is proposed by extending the existing entropy weight soft subspace clustering algorithm with a multiple-kernel learning setting. By incorporating multiple-kernel learning strategy into the framework of soft subspace fuzzy clustering, *MKEWFC-K* can learning the distance function adaptively during the clustering process. Moreover, it is more immune to ineffective kernels and irrelevant features in soft subspace, which makes the choice of kernels less crucial. Experiments on real-world data demonstrate the effectiveness of the proposed *MKEWFC-K* algorithm.**

Keywords—soft subspace clustering; fuzzy clustering; multiple-kernel learning

I. INTRODUCTION

Fuzzy clustering, which partitions data sets into several fuzzy clusters, has been widely applied in many fields. Recently, soft subspace fuzzy clustering has emerged as a hot research topic [1-5]. Under the classical fuzzy clustering framework, it groups data objects in the entire data space but assign different weights to different dimensions of clusters based on the importance of the features in identifying the corresponding clusters. Due to its assigning different vector of feature weights to each cluster, the soft subspace clustering is more suitable for the data sets in which different clusters are correlated with different subsets of features.

Up to now, several soft subspace fuzzy clustering algorithms have been proposed [1-5]. Generally speaking, these algorithms can be unified as the problem of finding the local minimum of the objective function.

$$J(\mathbf{U},\mathbf{W},\mathbf{V})=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^{m}\sum_{h=1}^{s}w_{ih}{}^{\alpha}d_{jih}^{2}+\mathrm{H} \quad (1)$$

under the constraints $\sum_{i=1}^{c}u_{ik}=1$ and $\sum_{h=1}^{d}w_{ih}=1$, where the first term $\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^{m}\sum_{h=1}^{s}w_{ih}^{\alpha}d_{jih}^{2}$ is interpreted as the total weighted distance of each data object to cluster centers and the second term **H** is a penalty term which is often used to optimize the performance of the algorithm. In the first term of Eq.(1), $d_{jih}{}^2$ is a dissimilarity measure of the jth cluster center $\mathbf{z}_j$ and the data point $\mathbf{x}_i$ on the hth feature. In most existing works, it is usually computed as Euclidean distance, i.e. $d_{jih}{}^2=(\mathbf{x}_{ih}-\mathbf{x}_{jh})^2$. In some cases of recent works, it can also be evaluated with other distance functions such as the Minkowski distance function [6], the alternative distance function [2] and the kernelized distance function [7]. As reported in literatures, these soft subspace clustering algorithms can successfully cluster high dimensional data in soft subspaces and achieve more ideal clustering results than some conventional fuzzy space fuzzy clustering approaches.

Although current soft subspace fuzzy clustering algorithms are able to cluster some high dimensional data well, these methods may be ineffective for some complex data. For example, the data item along with a particular feature could exhibit quite complex relationships with the corresponding feature of the cluster center, and thus, the similarities computed with current dissimilarity measure will not reflect the actual relationships between them and this will degrade the performance of the learning algorithm. In this work, the above problem will be investigated by integrating kernel tricks into the framework of soft subspace clustering. Like kernel methods in literatures, a crucial step in the kernelized soft subspace fuzzy clustering is the selection of the best kernels for each feature among an extensive range of possibilities [7]. Unfortunately, it is unclear which kernels are more suitable for a particular feature since this step is often heavily influenced by prior knowledge about the data and by the patterns we expect to discover [8-9].

Instead of a single fixed kernel, multiple-kernel learning methods have been extensively studied in recent years [8, 11-13]. Recent developments in multiple-kernel learning have shown that the construction of a kernel from a number of basis kernels allows for more flexible encoding of domain knowledge from different sources or cues. In this work, multiple-kernel learning will be incorporated into the framework of soft subspace fuzzy clustering and a novel soft subspace fuzzy clustering algorithm named *MKEWFC-K* will be proposed accordingly. *MKEWFC-K* simultaneously locates clusters in kernel space spanned by different kernels and the optimal kernel weights for a combination of a set of kernels. The incorporation of multiple kernels and the automatic adjustment of kernel weights make the similarities between data items along with each feature to be adaptively computed. In this way, the performance of the learning method can be improved effectively.

The rest of the paper is organized as follows. In Section II, we derive the *MKEWFC-K* and, in Section III we report experimental results on both synthetic and real data. The conclusions are given in Section IV.

II. ALGORITHM MKEWFC-K

A. Objective function incorporating multiple kernels

For better modeling and discovering nonlinear relationships among data, kernel methods use a non-linear mapping ϕ that maps the input data from the original feature space to a new space of higher dimensionality, i.e. kernel space [10]. The kernel space could possibly be of infinite dimensionality. Given two data points $\mathbf{x}_i$ and $\mathbf{x}_j$ in the original feature space, the kernel method takes advantage of the fact that the dot product between $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$ in the kernel space can be expressed by a Mercer kernel K given by

$$\mathrm{K}(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j) \tag{2}$$

Kernel methods rely on the kernel function which used to project data in the original space into the high dimensional kernel-induced feature space. A good choice of the kernel function is therefore imperative to the success of the learning problem. However, it is often unclear which kernel is the most suitable for a particular learning task, which has become one of the central problems in kernel methods. Recent studies on kernel methods have shown that constructing the kernel from a number of homogeneous or even heterogeneous kernels can improve the performance of the learning problem effectively, which are often referred as multiple-kernel learning in literatures [11-16]. This provides extra flexibility and allows domain knowledge from possibly different sources to be incorporated. Consider a set of unknown mappings, $\Phi=\{\phi_1, \ldots, \phi_p\}$, each of which, i.e. ϕ_t, maps the hth feature x_{ih} of the input data $\mathbf{x}_i$ as a L_t-dimensional vector $\phi_t(x_{ih})$ in kernel space. Let $\{\mathrm{K}_1, \mathrm{K}_2, \ldots, \mathrm{K}_p\}$ be the Mercer kernels corresponding to these implicit mappings respectively, then we have

$$\mathrm{K}_t(x_{ih}, x_{jh}) = \phi_t(x_{ih})^T\phi_t(x_{jh}) \tag{3}$$

which used to compute the similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$ along with the hth feature in the kernel space. In order to combine these kernels, a new set of independent mappings, $\Phi = \{\Phi_1, \Phi_2, \cdot, \Phi_p\}$, is constructed from the original mappings as follows:

$$\Phi_t(x_{ih}) = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \phi_t(x_{ih}) \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

which converts $\mathbf{x}_{ih}$ into a L-dimensional vector, where $L = \sum_{t=1}^{p} L_t$ and L_t is the dimensionality of ϕ_t. Although the implicit mappings ϕ_ts do not have the same dimensionality, the kernel spaces spanned by Φ_ts have the same dimensionality. In this way, their linear combination can be well defined. $\Phi_t(x_{ih})$s form a set of orthogonal bases in the high dimensional kernel space, i.e.,

$$\Phi_{t_1}{}^T(\mathbf{x}_{ih}) \cdot \Phi_{t_2}(\mathbf{x}_{ih})$$
$$= \phi_t{}^T(x_{ih})\, \phi_t(x_{ih}) = \mathrm{K}_t(x_{ih}, x_{ih}), \text{ if } t_1=t_2 \tag{4}$$
$$\Phi_{t_1}{}^T(\mathbf{x}_{ih}) \cdot \Phi_{t_2}(\mathbf{x}_{ih}) = 0, \text{ if } t_1 \neq t_2$$

Thus, any L-dimensional vector $\Psi(x_{ih}) = \sum_{t=1}^{p} v_t\Phi_t(x_{ih})$, j = 1,2, …, c, in the kernel space can be expressed as a non-negative linear combination of the basis kernels Φ_t and it can be viewed as a mapping which maps the data on each feature to the kernel space.

For kernelized fuzzy clustering algorithms, there are two general categories to perform clustering tasks [9]. The first category retains the prototypes in the feature space during clustering process, which is often denoted as KFCM-F (Kernel-based FCM with prototypes in Feature space). On the other hand, the second category implicitly leaves the prototypes in the kernel space during clustering process and an inverse mapping must be performed to obtain prototypes in the feature space. The algorithms in this category are often denoted as KFCM-K (Kernel-based FCM with prototypes in Kernel space). In this work, we only investigate soft subspace fuzzy clustering with multiple-kernel learning setting under the framework of the second category and denote the proposed algorithm *MKEWFC-K* (Entropy weighting fuzzy clustering with multiple-kernel learning).

Let

$$\Psi(\mathbf{x}) = \begin{bmatrix} \phi_1(\mathbf{x}) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \phi_2(\mathbf{x}) & & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \phi_p(\mathbf{x}) \end{bmatrix} \begin{bmatrix} v_{11} \cdots v_{1p} \\ v_{21} \quad v_{2p} \\ \vdots \quad\; \vdots \\ v_{c1} \cdots v_{cp} \end{bmatrix}^T$$
$$= \begin{bmatrix} \Phi_1(\mathbf{x}) & \Phi_2(\mathbf{x}) & \cdots & \Phi_p(\mathbf{x}) \end{bmatrix} \begin{bmatrix} v_{11} \cdots v_{1p} \\ v_{21} \quad v_{2p} \\ \vdots \quad\; \vdots \\ v_{c1} \cdots v_{cp} \end{bmatrix}^T$$
$$= \begin{bmatrix} \sum_{t=1}^{p} v_{1t}\Phi_t(\mathbf{x}) & \sum_{t=1}^{p} v_{2t}\Phi_t(\mathbf{x}) & \cdots & \sum_{t=1}^{p} v_{ct}\Phi_t(\mathbf{x}) \end{bmatrix}$$
$$= \begin{bmatrix} \Psi_1(\mathbf{x}) & \cdots & \Psi_c(\mathbf{x}) \end{bmatrix}$$

where $\Psi_j(x) = \sum_{t=1}^{p} v_{jt}\Phi_t(x)$, j = 1,2, …, c. By integrating the idea of multiple-kernel learning into the framework of soft subspace clustering, the objective function of *MKEWFC-K* can be constructed as follows:

$$J_{MKEWFC-K}(\mathbf{U}, \mathbf{W}, \mathbf{V}, \mathbf{Z})$$
$$= \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji}{}^m \sum_{h=1}^{s} w_{jh}\left(\Psi_j(x_{ih}) - z_{jh}\right)^T\left(\Psi_j(x_{ih}) - z_{jh}\right)$$
$$+\gamma\sum_{j=1}^{c}\sum_{t=1}^{p} v_{jt}{}^2 \log v_{jt}{}^2 + \eta\sum_{j=1}^{c}\sum_{h=1}^{s} w_{jh}\log w_{jh} \tag{5}$$

subject to

$$\sum_{j=1}^{c} u_{ji} = 1,\ u_{ji} \in [0,1],\ \ m>1$$
$$\sum_{h=1}^{s} w_{jh} = 1,\ j = 1, 2, \ldots, c$$
$$\sum_{t=1}^{p} v_{jt}{}^2 = 1,\ j = 1, 2, \ldots, c$$

where $\mathbf{Z}=[z_{jh}]$ is the cluster center matrix, W = $[w_{jh}]$ is a $c\times s$ feature weight matrix, $\mathbf{V}=[v_{jt}]$ is a $c\times p$ kernel weights matrix and $\mathbf{U}=[u_{ji}]$ is the fuzzy partition matrix.

The main idea of *MKEWFC-K* is to minimize the sum of the within cluster dispersions as well as the negative weight entropy in Eq.(5), which contains three terms, i.e. the within-cluster compactness in the kernel space, the negative entropy of both feature weights and kernel weights. The positive parameters γ and η are used to control the influences of the entropy of both w_{jh} and v_{jt}.

In literatures, the Euclidean norm

$$d_{jih}^2 = \left(x_{ih} - z_{jh}\right)^2 \quad (6)$$

is commonly used for features in soft subspace clustering algorithms. After introducing the mapping Ψ, which is combined with multiple kernels, into the algorithm, the dissimilarity between the jth cluster center $\mathbf{z}_j$ and the ith data point $\mathbf{x}_i$ on the hth feature can be computed as follows:

$$\begin{aligned} d_{jih}^2 &= \left(\Psi_j(x_{ih}) - z_{jh}\right)^T \left(\Psi_j(x_{ih}) - z_{jh}\right) \\ &= \Psi_j^T(x_{ih})\Psi_j(x_{ih}) - 2z_{jh}^T\Psi_j(x_{ih}) + z_{jh}^T z_{jh} \\ &= \sum\nolimits_{t=1}^{p} v_{jt}^2 e_{jih}^{(t)} \end{aligned}$$

where

$$e_{jih}^{(t)} = K_t(x_{ih}, x_{ih}) - 2\frac{\sum_{k=1}^{n} u_{jk}^m K_t\left(x_{kh}, x_{ih}\right)}{\sum_{k=1}^{n} u_{jk}^m} + \frac{\sum_{i1=1}^{n}\sum_{i2=1}^{n} u_{j,i1}^m u_{j,i2}^m K_t\left(x_{i1,h}, x_{i2,h}\right)}{\left(\sum_{i=1}^{n} u_{ji}^m\right)^2}$$

denotes the dissimilarity between data $\mathbf{x}_i$ and the cluster center $\mathbf{z}_j$ on the hth feature in the kernel space induced by the implicit mapping □$_t$. For brevity, d_{ji} is introduced to denote the weighted distance between data $\mathbf{x}_i$ and the cluster center $\mathbf{z}_j$, i.e.

$$d_{ji}^2 = \sum\nolimits_{h=1}^{s} w_{jh} d_{jih}^2 \quad (7)$$

In this way, it is possible to obtain memberships and weights without implicitly evaluating cluster centers. Moreover, the similarities between data items along with each feature can be computed adaptively since v_{jt}^2 can be updated during the clustering process.

B. The proposed algorithm MKEWFC-K

Minimization of the objective function in Eq.(5) with the constraints forms a class of constrained nonlinear optimization problems. The usual method toward optimization of $J_{MKEWFC\text{-}K}$ is to use the partial optimization for **U**, **W**, **V** and **Z**. Similar with FCM-like algorithms, the objective function defined in Eq.(5) can be minimized by iteratively solving the following three minimization problems:

1. Problem P1: Fix **U**, **W** and **Z**, solve the reduced problem $J_{MKEWSC-K}\left(\tilde{\mathbf{U}}, \tilde{\mathbf{W}}, \mathbf{V}, \tilde{\mathbf{Z}}\right)$;

2. Problem P2: Fix **U**, **V** and **Z**, solve the reduced problem $J_{MKEWSC-K}\left(\tilde{\mathbf{U}}, \mathbf{W}, \tilde{\mathbf{V}}, \tilde{\mathbf{Z}}\right)$;

3. Problem P3: Fix **W**, **V** and **Z**, solve the reduced problem $J_{MKEWSC-K}\left(\mathbf{U}, \tilde{\mathbf{W}}, \tilde{\mathbf{V}}, \tilde{\mathbf{Z}}\right)$.

Problem P1 is solved by

$$v_{jt}^2 = \frac{\exp\left(\frac{-\beta_{jt}}{\gamma}\right)}{\sum_{t=1}^{p}\left(\exp\left(\frac{-\beta_{jt}}{\gamma}\right)\right)} \quad (8)$$

where $\beta_{jt} = \sum_{i=1}^{n} u_{ji}^m \delta_{jit}$, $\delta_{jit} = \sum_{h=1}^{s} w_{jh} e_{jih}^{(t)}$.

Theorem 1. Given matrices **U**, **W** and **Z** are fixed, $J_{MKEWFC\text{-}K}$ is minimized if **V** is computed as

$$v_{jt}^2 = \frac{\exp\left(\frac{-\beta_{jt}}{\gamma}\right)}{\sum_{t=1}^{p}\left(\exp\left(\frac{-\beta_{jt}}{\gamma}\right)\right)}$$

where $\beta_{jt} = \sum_{i=1}^{n} u_{ji}^m \delta_{jit}$, $\delta_{jit} = \sum_{h=1}^{s} w_{jh} e_{jih}^{(t)}$.

Proof: The objective function defined in Eq.(5) can be rearranged as:

$$\begin{aligned} & J_{MKEWFC-K}\left(\mathbf{U}, \mathbf{W}, \mathbf{V}, \mathbf{Z}\right) \\ &= \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji}^m \sum_{h=1}^{s} w_{jh} d_{jih}^2 + \gamma\sum_{j=1}^{c}\sum_{t=1}^{p} v_{jt}^2 \log v_{jt}^2 + \eta\sum_{j=1}^{c}\sum_{h=1}^{s} w_{jh}\log w_{jh} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{c}\sum_{t=1}^{p} v_{jt}^2 u_{ji}^m \delta_{jit} + \gamma\sum_{j=1}^{c}\sum_{t=1}^{p} v_{jt}^2 \log v_{jt}^2 + \eta\sum_{j=1}^{c}\sum_{h=1}^{s} w_{jh}\log w_{jh} \end{aligned}$$

where $\delta_{jit} = \sum_{h=1}^{s} w_{jh} e_{jih}^{(t)}$

By forming the Lagrange function with Lagrange multipliers for constraint $\sum_{t=1}^{p} v_{jt}^2 = 1$, we have

$$\begin{aligned} & L_\lambda\left(\mathbf{U}, \mathbf{W}, \mathbf{Z}, \lambda\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{c}\sum_{t=1}^{p} v_{jt}^2 u_{ji}^m \delta_{jit} + \gamma\sum_{j=1}^{c}\sum_{t=1}^{p} v_{jt}^2 \log v_{jt}^2 \\ &+ \eta\sum_{j=1}^{c}\sum_{h=1}^{s} w_{jh}\log w_{jh} - \lambda\left(\sum_{t=1}^{p} v_{jt}^2 - 1\right) \end{aligned} \quad (9)$$

By taking the derivative of Eq.(9) respect to v_{jt}^2 and set them to zero, i.e.:

$$\frac{\partial L_\lambda}{\partial v_{jt}^2} = \sum_{i=1}^{n}\frac{\partial J_i}{\partial v_{jt}^2} + \gamma\frac{\partial\left(\sum_{j=1}^{c}\sum_{h=1}^{s} v_{jt}^2 \log v_{jt}^2\right)}{\partial v_{jt}^2} - \lambda = 0$$

we have:

$$\beta_{jt} + \gamma\left(1 + \log v_{jt}^2\right) - \lambda = 0$$

According to the constraints $\sum_{t=1}^{p} v_{jt}^2 = 1$, we can eliminate and obtain the closed-form solution for v_{jt}^2 as $v_{jt}^2 = \frac{\exp\left(\frac{-\beta_{jt}}{\gamma}\right)}{\sum_{t=1}^{p}\left(\exp\left(\frac{-\beta_{jt}}{\gamma}\right)\right)}$. Thus, the theorem is proved.

Problem P2 is solved by

$$w_{jh} = \frac{\exp\left(\frac{-\alpha_{jh}}{\eta}\right)}{\sum_{h=1}^{s}\left(\exp\left(\frac{-\alpha_{jh}}{\eta}\right)\right)} \quad (10)$$

where $\alpha_{jh} = \sum_{i=1}^{n} u_{ji}^m d_{jih}^2$, d_{jih} is computed with Eq.(6).

Theorem 2. Given matrices **U**, **V** and **Z** are fixed, $J_{\mathrm{MKEWFC\text{-}K}}$ is minimized if **W** is computed as

$$w_{jh} = \frac{\exp\left(\frac{-\alpha_{jh}}{\eta}\right)}{\sum_{h=1}^{s}\left(\exp\left(\frac{-\alpha_{jh}}{\eta}\right)\right)},$$

where $\alpha_{jh} = \sum_{i=1}^{n} u_{ji}^m d_{jih}^2$.

Proof: The objective function defined in Eq.(1) can be rearranged as:

$$J(\mathbf{U},\mathbf{W},\mathbf{Z})$$
$$= \sum_{j=1}^{c}\left(\sum_{i=1}^{n} u_{ji}^m \sum_{h=1}^{s} w_{jh} d_{jih}^2 + \gamma\sum_{t=1}^{p} v_{jt}^2 \log v_{jt}^2 + \eta \sum_{h=1}^{s} w_{jh} \log w_{jh}\right)$$
$$= \sum_{j=1}^{c}\sum_{h=1}^{s} w_{jh} \sum_{i=1}^{n} u_{ji}{}^m d_{jih}^2 + \gamma \sum_{j=1}^{c}\sum_{t=1}^{p} v_{jt}^2 \log v_{jt}^2 + \eta \sum_{j=1}^{c}\sum_{h=1}^{s} w_{jh} \log w_{jh}$$

Similar to FCM-like algorithms, by forming the Lagrange function with Lagrange multipliers for constraint $\sum_{h=1}^{s} w_{jh} = 1$, we have

$$L_\delta(\mathbf{U},\mathbf{V},\mathbf{Z},\delta) = J(\mathbf{U},\mathbf{W},\mathbf{Z}) - \delta\left(\sum_{h=1}^{s} w_{jh} - 1\right) \tag{11}$$

By taking the derivative of Eq.(11) respect to w_{jh} and set them to zero. For each w_{jh}, we have:

$$\frac{\partial L}{\partial w_{jh}} = \frac{\partial J}{\partial w_{jh}} - \delta$$
$$= \sum_{i=1}^{n} u_{ji}{}^m d_{jih}^2 + \eta\left(1 + \log w_{jh}\right) - \delta = 0$$

The solution for w_{jh} is

$$w_{jh} = \exp\left(\frac{\delta}{\eta}\right)\exp\left(\frac{-\sum_{i=1}^{n} u_{ji}{}^m d_{jih}^2 - \eta}{\eta}\right)$$

Considering the constraint $\sum_{h=1}^{s} w_{jh} = 1$, we can eliminate δ and obtain the closed-form solution for the optimal w_{jh} as

$$w_{jh} = \frac{\exp\left(\frac{-\alpha_{jh}}{\eta}\right)}{\sum_{h=1}^{s}\left(\exp\left(\frac{-\alpha_{jh}}{\eta}\right)\right)} \tag{12}$$

in which $\alpha_{jh} = \sum_{i=1}^{n} u_{ji}^m d_{jih}^2$. Thus the theorem is proved.

Problem P3 is solved by

$$u_{ji} = \left(\frac{\lambda_j}{m d_{ji}^2}\right)^{\frac{1}{m-1}} = \frac{d_{ji}^{-\frac{2}{m-1}}}{\sum_{j=1}^{c} d_{ri}^{-\frac{2}{m-1}}} \tag{13}$$

in which d_{ji} is computed with Eq.(7).

Theorem 3. Given matrices **W** and **V** are fixed, $J_{CKS\text{-}EWFC\text{-}K}$ is minimized if **U** is computed as Eq.(13).

Proof: Denoting $d_{ji}^2 = \sum_{h=1}^{s} w_{jh} d_{jih}^2$, the objective function defined in Eq.(5) can be rearranged as:

$$J_{MKEWFC-K}(\mathbf{U},\mathbf{W},\mathbf{V},\mathbf{Z})$$
$$= \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji}^m d_{ji}^2 + \gamma\sum_{j=1}^{c}\sum_{t=1}^{p} v_{jt}^2 \log v_{jt}^2 + \eta\sum_{j=1}^{c}\sum_{h=1}^{s} w_{jh}\log w_{jh}$$

By forming the Lagrange function with Lagrange multipliers for constraints $\sum_{j=1}^{c} u_{ji} = 1$, i=1,2,…,n, we have:

$$L_\zeta(\mathbf{U},\mathbf{W},\zeta) = \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji}^m d_{ji}^2 + \gamma\sum_{j=1}^{c}\sum_{t=1}^{p} v_{jt}^2 \log v_{jt}^2$$
$$+\eta\sum_{j=1}^{c}\sum_{h=1}^{s} w_{jh}\log w_{jh} - \sum_{i=1}^{n}\zeta_i\left(\sum_{j=1}^{c} u_{ji} - 1\right) \tag{14}$$

By setting the gradient of Eq.(14) with respect to u_{ji} and ζ_i to zero, we have:

$$\frac{\partial L_\zeta}{\partial u_{ji}} = m u_{ji}^{m-1} d_{ji}^2 - \zeta_i = 0 \tag{15}$$

and

$$\frac{\partial L_\zeta}{\partial \zeta_i} = \sum_{j=1}^{c} u_{ji} - 1 = 0 \tag{16}$$

Substituting Eq.(15) into Eq.(16), we can eliminate ζ_i and obtain the closed-form solution for u_{ji} as

$$u_{ji} = \frac{d_{ji}^{-\frac{2}{m-1}}}{\sum_{j=1}^{c} d_{ri}^{-\frac{2}{m-1}}}.$$

Thus, the theorem is proved.

Although the cluster centers are in the implicit kernel space and directly evaluating them may not be possible, the above derivation makes the cluster centers be eliminated from the objective function so that *MKEWFC-K* do not need to implicitly evaluate cluster centers, which are potentially not computable.

The proposed algorithm is developed based on the objective function defined in Eq.(5). Since the cluster centers are eliminated from the objective function, *MKEWFC-K* does not need to implicitly evaluate cluster centers, which are potentially not computable for relational data sets. Table I summarizes the *MKEWFC-K* algorithm. The algorithm starts with initializing the memberships and feature weights, and then repeats updating the kernel weights by fixing the memberships and feature members, updating the feature weights by fixing the memberships and kernel weights and updating the memberships by fixing the feature weights and kernel weights. The process is repeated until the amount of change per iteration in the membership matrix falls below a given threshold.

TABLE I THE PSEUDO-CODE OF THE MKEWFC-K ALGORITHM

Algorithm: *MKEWFC-K*
Require: D—the data set, *k*—the number of clusters 1: Randomly initialize membership matrix and initialize **W** with $w_{jh} = 1/s$ 2: for iter = 1: maxIter 3: Update **V** according to Eq.(8); 4: Update **W** according to Eq.(12); 5: Update **U** according to Eq.(13); 6: Calculate the objective function J; 7 if $\lvert J\text{-}J_{old}\rvert < \varepsilon$, break; 8: end for 9: Output **W**, **Z** and **U**.

III. EXPERIMENTAL RESULTS

The proposed algorithm *MKEWFC-K* has been evaluated with a large number of experiments on real data sets and its performance is compared with several classical soft subspace clustering algorithms as well as fuzzy clustering algorithms.

In our experiments, two measures, the rand index (*RI*) [17] and the normalized mutual information (*NMI*) [18] are used for evaluating the performance of the clustering algorithms.

A. Experiments results on UCI data

We tested these algorithms on the data sets which are selected from the UCI repository. For each set, only the extracted feature vectors are available — not the raw data. All these data sets are described by a data matrix of "objects × features". Table II shows the details.

TABLE II. DETAILS OF UCI DATA SETS

Data set	Number of instances	Number of features	Number of clusters
Balance Scale	625	4	3
Breast Tissue	106	9	6
Bupa	345	6	2
Glass	214	9	6
Heart	270	13	2
Iris	150	4	3
Parkinsons	195	23	2
Pima Indians Diabetes	768	8	2
Vehicle	846	18	4
Wine	178	13	3

Kernels are often used to address the problems of ineffective features and similarity measures. As kernel functions are essentially similarity measures for pairs of data, they can be used in many different ways, and multiple kernels sets can be constructed in various ways. Since the feature vectors for data points have been provided in this work, we measured the similarities between data items along with each feature in different nonlinear spaces by mapping data to these spaces with different kernels. In our experiments, the provided feature vectors are substituted into the chosen kernels to calculate pairwise similarities along with each feature. Since optimal kernel choice is still an open problem, in this work, following the strategy of other multiple-kernel learning approaches, we select a set of reasonable kernels that are frequently used by kernel methods. In our experiments, one polynomial kernel with θ=1 and p=2, one linear kernel and several Gaussian kernels were utilized. We varied the bandwidth for Gaussian kernel over {log(0.1), log(0.05), log(0.01), log(0.005), log(0.001), log(0.0005), log(0.0001)} to obtain seven Gaussian kernels. Table III shows the details on the kernels adopted in our experiments. After the kernel matrices were generated for the whole data set, the elements' values were normalized to the range of [0, 1].

TABLE III. BASIS KERNEL INFORMATION

id	kernel type	Parameters settings
K1	Polynomial kernel $k(\mathbf{x}_1,\mathbf{x}_2)=\left(\mathbf{x}_1^T\mathbf{x}_2+\theta\right)^p$	p=1, d=2
K2	Gaussian kernel	σ=log(0.1)
K3		σ=log(0.05)
K4		σ=log(0.01)
K5	$k(\mathbf{x}_1,\mathbf{x}_2)=\exp\left(\frac{\|\mathbf{x}_1-\mathbf{x}_2\|^2}{2\sigma^2}\right)$	σ=log(0.005)
K6		σ=log(0.001)
K7		σ=log(0.0005)
K8		σ=log(0.0001)
K9	Linear kernel $k(\mathbf{x}_1,\mathbf{x}_2)=\mathbf{x}_1^T\mathbf{x}_2$	/

We compared the performance of *MKEWFC-K* with those of AWA [19], EWKM [20], FWKM [21] and FCM [22]. Table IV shows the parameters utilized in the experiments, from which we can observe that a wide range of parameters have been tried. We also run *MKEWFC-K* with only one basis kernel (denoted as *KEWFC-K*) so that the advantage of kernel combination can be fully discovered.

TABLE IV. PARAMETERS USED IN THE EXPERIMENTS

Algorithms	Parameters settings
MKEWFC-K	m=1.05 or 1.2 γ = 1;5;10;50;100;1000 η = 1;5;10;50;100;1000;10000
KEWFC-K	The parameters m, γ and η take values with which *MKEWFC-K* obtained its best performance
EWKM	γ = 1;5;10;50;100;1000;1000
FWKM	β = 1.25; 1.5; 1.75; 2.0; 2.25; 2.5; 3.0
AWA	β = 1.25; 1.5; 1.75; 2.0; 2.25; 2.5; 3.0
FCM	m=1.2; 1.4; 1.6; 1.8; 2.0; 2.2; 2.4; 2.6; 2.8; 3.0

In order to evaluate soft clustering results with *RI* and *NMI* measure, the fuzzy membership degrees were converted to hard assignments by assigning each data to the cluster with the highest membership degree. All the algorithms use data with the same dimensions (specified by the #features attribute in Table II). In the experiment, each algorithm was repeated 10 times under a fixed parameters setting and the average performance was calculated. Different parameters settings for each algorithm were tried in our experiments and the best average performance over all the parameter settings was reported in Table V and Table VI, respectively. The average performance over all the data sets (*aRI* and *aNMI*) of each algorithm as well as its rank is shown in the last row of both tables.

From Table V and Table VI, it can be observed that there is no algorithm always ranked first for all the data sets. However, on average, *MKEWFC-K* performs the best and yields stable performance. For real-world applications, we have no cues in advance as to which algorithm will work best for the given problem. *MKEWFC-K* can help us to obtain more reasonable results in general. From both tables, it can be also observed that *MKEWFC-K* as well as other soft subspace clustering algorithms do not outperform the classical *FCM* algorithm for the case "Balance Scale" and "Glass". We wonder that this is because the inner structures of the clusters within both data sets are so strange and puzzling that the existing soft subspace clustering algorithms cannot work on it well. In our future work, we will further investigate this problem.

TABLE V. BEST AVERAGE RI PERFORMANCE FOR DIFFERENT ALGORITHMS

Data sets		***MKEWFC-K***	***KEWFC-K (MKEWFC-K with single basis kernel)***									***EWKM***	***FWKM***	***AWA***	***FCM***
			K1	***K2***	***K3***	***K4***	***K5***	***K6***	***K7***	***K8***	***K9***				
Balance Scale	mean	0.5879	0.5755	0.5882	0.5834	0.5896	0.5913	0.5903	0.5906	0.5899	0.5875	0.5895	0.5885	0.5901	**0.6243**
	std	0.0142	0.0025	0.0111	0.0142	0.0116	0.0124	0.0142	0.0146	0.0150	0.0164	0.0109	0.0111	0.0117	**0.0612**
Breast Tissue	mean	**0.8100**	0.7433	0.7991	0.7910	0.8011	0.8028	0.7990	0.8009	0.8029	0.7722	0.7140	0.6097	0.7027	0.6729
	std	**0.0227**	0.0262	0.0194	0.0245	0.0189	0.0208	0.0260	0.0272	0.0225	0.0173	0.0216	0.0411	0.0420	0.0080
Bupa	mean	**0.5154**	0.5026	0.5037	0.4994	0.5136	0.5096	0.5122	0.5121	0.5113	0.5031	0.5107	0.4997	0.5044	0.5037
	std	**0.0024**	0.0000	0.0000	0.0009	0.0025	0.0008	0.0000	0.0003	0.0019	0.0000	0.0005	0.0003	0.0033	0.0000
Glass	mean	0.6875	0.6790	0.6901	0.6886	0.6861	0.6844	0.6801	0.6788	0.6771	0.6804	0.6932	0.6664	0.5874	**0.7238**
	std	0.0083	0.0064	0.0046	0.0028	0.0060	0.0083	0.0099	0.0097	0.0087	0.0056	0.0193	0.0275	0.0519	**0.0053**
Heart	mean	**0.6971**	0.6731	0.6952	0.6888	0.6860	0.6906	0.6901	0.6897	0.6948	0.5850	0.6579	0.5708	0.5043	0.5213
	std	**0.0000**	0.0047	0.0044	0.0029	0.0023	0.0032	0.0024	0.0032	0.0025	0.0207	0.0640	0.0171	0.0000	0.0000
Iris	mean	0.9267	0.8923	0.7735	0.8123	0.8820	0.8882	0.9044	0.8879	0.8843	**0.9495**	0.8590	0.9003	0.9342	0.8859
	std	0.0000	0.0000	0.0131	0.0726	0.0730	0.0934	0.0665	0.0627	0.0603	**0.0000**	0.0467	0.0207	0.0484	0.0000
Parkinsons	mean	**1.0000**	**1.0000**	0.6021	0.5975	0.5929	0.5929	0.6027	0.6104	0.6095	**1.0000**	0.6949	0.5427	0.4975	0.6167
	std	**0.0000**	**0.0000**	0.0000	0.0000	0.0000	0.0000	0.0073	0.0160	0.0464	**0.0000**	0.1501	0.0238	0.0000	0.0000
Pima Indians Diabetes	mean	0.6153	**0.6154**	0.5466	0.5466	0.5458	0.5458	0.5465	0.5466	0.5474	0.5592	0.5450	0.5388	0.5390	0.5507
	std	0.0012	**0.0004**	0.0000	0.0000	0.0000	0.0000	0.0003	0.0000	0.0000	0.0008	0.0000	0.0282	0.0000	0.0000
Vehicle	mean	0.6625	0.3698	0.6604	0.6645	0.6622	0.6631	0.6659	0.6597	0.6522	0.5977	0.6135	0.6482	**0.6663**	0.6523
	std	0.0039	0.1066	0.0138	0.0127	0.0043	0.0030	0.0064	0.0178	0.0212	0.0780	0.0140	0.0393	**0.0105**	0.0003
Wine	mean	**0.9318**	0.8972	0.9114	0.9301	0.9036	0.8855	0.8787	0.8834	0.8779	0.9133	0.6864	0.7095	0.6744	0.7156
	std	**0.0000**	0.0000	0.0030	0.0031	0.0000	0.0026	0.0021	0.0064	0.0020	0.0036	0.0080	0.0483	0.0542	0.0097
aRI		**0.7434**	0.6948	0.6770	0.6802	0.6863	0.6854	0.6870	0.6860	0.6847	0.7148	0.6564	0.6275	0.6200	0.6467
		(1)	(3)	(10)	(9)	(5)	(7)	(4)	(6)	(8)	(2)	(11)	(13)	(14)	(12)

TABLE VI. BEST AVERAGE NMI PERFORMANCE FOR DIFFERENT ALGORITHMS

Data sets		***MKEWFC-K***	***KEWFC-K (MKEWFC-K with single basis kernel)***									***EWKM***	***FWKM***	***AWA***	***FCM***
			K1	***K2***	***K3***	***K4***	***K5***	***K6***	***K7***	***K8***	***K9***				
Balance Scale	mean	0.1173	0.0952	0.1168	0.1091	0.1167	0.1191	0.1171	0.1179	0.1164	0.1133	0.1202	0.1173	0.1205	**0.1774**
	std	0.0287	0.0110	0.0227	0.0285	0.0228	0.0243	0.0265	0.0277	0.0279	0.0292	0.0237	0.0188	0.0195	**0.1029**
Breast Tissue	mean	0.4856	0.4794	0.2700	0.2756	0.2734	0.2730	0.2695	0.2695	0.2708	**0.5503**	0.2606	0.3409	0.3050	0.3889
	std	0.0218	0.0181	0.0285	0.0210	0.0281	0.0296	0.0319	0.0319	0.0302	**0.0112**	0.0203	0.0261	0.0291	0.0129
Bupa	mean	**0.0196**	0.0000	0.0093	0.0006	0.0034	0.0009	0.0013	0.0026	0.0031	0.0002	0.0113	0.0134	0.0062	0.0129
	std	**0.0000**	0.0000	0.0011	0.0003	0.0004	0.0006	0.0008	0.0019	0.0017	0.0001	0.0011	0.0000	0.0033	0.0000
Glass	mean	0.3663	0.3418	0.3796	0.3755	0.3508	0.3421	0.3388	0.3371	0.3382	0.3487	0.3721	**0.4308**	0.2224	0.4264
	std	0.0370	0.0187	0.0389	0.0409	0.0320	0.0178	0.0186	0.0171	0.0150	0.0169	0.0359	**0.0247**	0.0684	0.0066
Heart	mean	**0.3062**	0.2648	0.3016	0.2911	0.2875	0.2959	0.2953	0.2946	0.3038	0.1309	0.2827	0.1037	0.0496	0.0304
	std	**0.0000**	0.0075	0.0071	0.0047	0.0039	0.0053	0.0038	0.0050	0.0038	0.0214	0.1047	0.0280	0.0569	0.0000
Iris	mean	0.8513	0.8058	0.5649	0.6366	0.7767	0.7998	0.8081	0.7828	0.7806	**0.8642**	0.7416	0.7882	0.8387	0.7582
	std	0.0000	0.0000	0.0251	0.1205	0.0801	0.0850	0.0588	0.0529	0.0488	**0.0000**	0.0291	0.0292	0.0805	0.0000
Parkinsons	mean	**1.0000**	**1.0000**	0.3212	0.3160	0.3109	0.3109	0.3218	0.3303	0.3130	**1.0000**	0.3206	0.0905	0.1144	0.1265
	std	**0.0000**	**0.0000**	0.0000	0.0000	0.0000	0.0000	0.0082	0.0180	0.0965	**0.0000**	0.0142	0.0611	0.0000	0.0000
Pima Indians Diabetes	mean	0.1306	**0.1381**	0.0096	0.0096	0.0052	0.0052	0.0091	0.0096	0.0139	0.0617	0.0204	0.0619	0.0204	0.0337
	std	0.0021	**0.0007**	0.0000	0.0000	0.0000	0.0000	0.0014	0.0000	0.0000	0.0012	0.0000	0.0435	0.0000	0.0000
Vehicle	mean	0.1836	0.1334	0.1656	0.1658	0.1749	0.1785	0.1648	0.1826	0.1629	0.1214	0.1385	**0.1949**	0.1813	0.1860
	std	0.0425	0.0317	0.0457	0.0250	0.0129	0.0031	0.0501	0.0314	0.0419	0.0057	0.0038	**0.0104**	0.0662	0.0025
Wine	mean	**0.8181**	0.7532	0.7639	0.8023	0.7673	0.7328	0.7204	0.7256	0.7090	0.8056	0.4641	0.4142	0.3428	0.4273
	std	**0.0000**	0.0000	0.0062	0.0070	0.0000	0.0054	0.0058	0.0108	0.0024	0.0054	0.0031	0.0140	0.1433	0.0047
aNMI		**0.4279**	0.4012	0.2903	0.2982	0.3067	0.3058	0.3046	0.3053	0.3012	0.3996	0.2662	0.2556	0.2201	0.2568
		(1)	(2)	(10)	(9)	(4)	(5)	(7)	(6)	(8)	(3)	(11)	(13)	(14)	(12)

B. *Can we always gain from the multiple-kernel version of soft subspace fuzzy clustering?*

As can be seen from Table V and Table VI, with one fixed basis kernel, *KEWFC-K* did not always obtain better performance than classical soft subspace clustering algorithms did. Given a fixed basis kernel, *KEWFC-K* would have better performance for some data sets. However, it could perform worse in other cases. This phenomenon implies that there is no basis kernel that is suitable for all the data sets because data items along with each feature may have various relationships which cannot be evaluated with only one dissimilarity measure.

The integration of multiple-kernel learning into the framework of soft subspace fuzzy clustering helped to improve the clustering results by automatically selecting the effective kernels along with each feature. In this way, we can utilize the multiple-kernel version of soft subspace fuzzy clustering in practice, without wondering which basis kernel is the most suitable one.

Another observation from both Table V and Table VI is that the best average performance of *MKEWFC-K* is always determined by the best results of *KEWFC-K* over different kernels when it run with the same parameters setting as *MKEWFC-K* did, which implies that introducing an effective kernel into the basis kernel sets helps to improve the performance of *MKEWFC-K*.

C. *The scalability of MKEWSC-K with respect to Kernel numbers*

The scalability of *MKEWSC-K* with respect to the kernel numbers is also investigated. In our experiment, *MKEWSC-K* with different kernel numbers was tested by adding one more kernel to the algorithm. Fig. 1 and Fig. 2 show the average time in seconds spent by *MKEWSC-K*, as well as the average *RI* returned by *MKEWSC-K*, on the ten data sets in Table II. The basis kernels in Table III were adopted and details on kernel combination were shown in Table VII.

As can be seen from both figures, the running time of *MKEWFC-K* increased linearly with the increment of the kernel number. On the other hand, the performance of *MKEWFC-K* goes steady although the kernel numbers become larger and larger. Another observation from Fig. 2 is that the addition of an effective kernel always improved the performance of the algorithm while an ineffective kernel affected little on its performance. This is because *MKEWSC-K* is immune to ineffective kernels.

TABLE VII. DETAILS ON KERNEL COMBINATION

Kernel numbers	Kernel information
1	K1
2	K1,K2
3	K1,K2, K3
4	K1,K2, K3, K4
5	K1,K2, K3, K4, K5
6	K1,K2, K3, K4, K5, K6
7	K1,K2, K3, K4, K5, K6, K7
8	K1,K2, K3, K4, K5, K6, K7, K8
9	K1,K2, K3, K4, K5, K6, K7, K8, K9

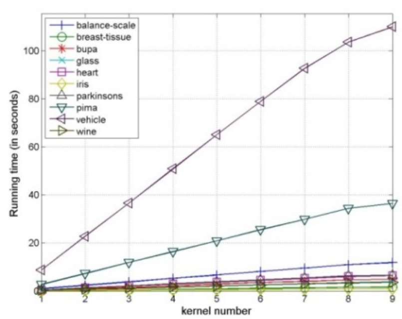

Fig. 1 The relationships between the kernel number and running time of *MKEWFC-K*

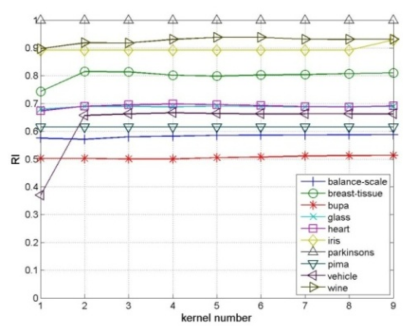

Fig. 2 The relationships between the kernel number and clustering accuracy of *MKEWFC-K*

IV. CONCLUSIONS AND FUTURE WORK

The existing soft subspace fuzzy clustering algorithms often utilize only one distance function to evaluate the similarity between data items along with each feature, which leads to performance degradation for some complex data sets. By introducing the mechanism of multiple-kernel learning, a novel entropy weighting subspace clustering algorithm named *MKEWFC-K* is proposed in this paper. The proposed algorithm can learn distance functions along with each feature adaptively during the clustering process. Moreover, it is easy to implement and provides convincing results that are immune to irrelevant, redundant, ineffective, and unreliable features and kernels. Experiments show that the method effectively incorporates multiple kernels and yields better overall performance. These characteristics make it useful for real-world applications.

This work involves the following aspects: (1) a novel learning criterion integrating the framework of soft subspace fuzzy clustering with multiple-kernel learning is proposed; (2) *MKEWFC-K* is developed with this learning criterion and its properties are investigated; (3) comprehensive experiments are carried out to evaluate the performance of the *MKEWFC-K* algorithm. The findings in this study demonstrate that the proposed *MKEWFC-K* algorithm is more effective in subspace clustering than the existing algorithms in general.

This study will be further extended to improve the performance of the existing subspace clustering algorithms by making use of the mechanism of multiple-kernel learning. For example, the multiple-kernel versions of the existing fuzzy weighting subspace clustering algorithms can be developed. In addition, a theoretical study on the parameter setting of the *MKEWFC-K* algorithm will be conducted, which will be of great importance in providing useful and convenient guidelines for the *MKEWFC-K* algorithm to be more practical in real world applications.

ACKNOWLEDGMENT

This work was supported in part by the Hong Kong Polytechnic University under Grant 1-ZV5V, by the National Natural Science Foundation of China under Grants 61170122, 61272210, 61300151, the Fundamental Research Funds for the Central Universities (JUSRP51321B) and by the Natural Science Foundation of Jiangsu province under Grant BK2011417 and BK20130155.

REFERENCES

[1] Z. Deng, K. Choi, F. Chung, S. Wang, “Enhanced soft subspace clustering integrating within-cluster and between-cluster information,” Pattern Recognition, vol.43, no.3, pp.767–781, 2010.

[2] Yonghui Pan, Jun Wang, Zhaohong Deng. “Alternative soft subspace clustering algorithm,” Journal of Information and Computational Science,vol.10, no.12, pp.3615–3624, 2013.

[3] Jun Wang, Shitong Wang, Fulai Chung, Zhaohong Deng. “Fuzzy partition based soft subspace clustering and its applications in high dimensional data,” Information Sciences, vol.246, no.10, pp.133-154, 2013

[4] Lin Zhu, Longbing Cao, Jie Yang, “Soft subspace clustering with competitive agglomeration,” 2011 IEEE International Conference on Fuzzy Systems, Taipei, Taiwan: 691-698 June 27-30, 2011.

[5] Liuqing Peng and Junying Zhang, "An entropy weighting mixture model for subspace clustering of high-dimensional data," Pattern Recognition Letters, vol. 32, pp.1154–1161, 2011.
[6] Renato Cordeiro de Amorim, Boris Mirkin. "Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering," Pattern Recognition, vol.45, no.3, 45(3): 1061-1075, 2012.
[7] Hong-bin Shen, Jie Yang, Shitong Wang, Xiaojun Liu. "Attribute weighted mercer kernel based fuzzy clustering algorithm for general non-spherical data sets," Soft Computing, vol.10, no.11, pp.1061-1073, 2006
[8] B. Zhao, J. Kwok, and C. Zhang, "Multiple kernel clustering," in Proceedings of the 9th SIAM International Conference on Data Mining (SDM 09), pp. 638–649, 2009.
[9] D. Graves and W. Pedrycz, "Kernel-based fuzzy clustering and fuzzy clustering: A comparative experimental study," Fuzzy Sets and Systems, vol. 161, no. 4, pp. 522–543, 2010.
[10] J. Shawe-Taylor and N. Cristianini, "Kernel Methods for Pattern Analysis," Cambridge University Press, 2004.
[11] H-C Huang, Y-Y Chuang, and Chu-Song Chen. "Multiple kernel fuzzy clustering", IEEE Transactions on Fuzzy Systems, vol.20, no.1, pp.120-134, 2012..
[12] C. S. S. Sonnenburg, G. Ratsch and B. Scholkopf, "Large scale multiple kernel learning," Journal of Machine Learning Research, vol. 7, pp. 1631–1565, 2006.
[13] A. Zien and C. Ong, "Multiclass multiple kernel learning," ICML, 2007.
[14] S. C. A. Rakotomamonjy, F. Bach and Y. Grandvalet, "More efficiency in multiple kernel learning," ICML, 2007.
[15] S. J. J. Ye and J. Chen, "Learning the kernel matrix in discriminant analysis via quadratically constrained quadratic programming," SIGKDD, 2007.
[16] M. Gonen and E. Alpaydin, "Localized multiple kernel learning," ICML, 2008.
[17] W. M. Rand, "Objective criteria for the evaluation of clustering methods," Journal of the American Statistical Association, vol. 66, pp. 846-850, 1971.
[18] S. Zhong and J. Ghosh, "A Comparative Study of Generative Models for Documents Clustering", Proc. SDW Workshop Clustering High-dimensional data and Its Applications, May 2003.
[19] Y. Chan, W. Ching, M.K. Ng, J. Z. Huang. "An optimization algorithm for clustering using weighted dissimilarity measures," Pattern Recognition, vol.37, no.5, pp.943-952, 2004.
[20] Liping Jing, Michael K. Ng, Joshua Zhexue Huang, "An Entropy Weighting k-Means Algorithm for Subspace Clustering of High-Dimensional Sparse Data," IEEE Transactions on Knowledge and Data Engineering, vol.19 no.8, p.1026-1041, 2007.
[21] L. Jing, M. K. Ng, J. Xu, and J.Z. Huang. "Subspace Clustering of Text Documents with Feature Weighting K-Means Algorithm," Proceedings of the Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp.802–812. 2005
[22] J. C. Bezdek, Pattern recognition with fuzzy objective function algorithms. Plenum Press, New York, 1981.