Pattern Recognition 43 (2010) 767–781

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.de/locate/pr

Enhanced soft subspace clustering integrating within-cluster and between-cluster information

Zhaohong Deng [a,b,*], Kup-Sze Choi [b], Fu-Lai Chung [c], Shitong Wang [a,d]

[a] School of Information Engineering, Jiangnan University, Wuxi, Jiangsu, PR China
[b] Ctr. for Int. Digital Health/School of Nursing, The Hong Kong Polytechnic University, Hong Kong
[c] Department of Computing, The Hong Kong Polytechnic University, Hong Kong
[d] State Key Lab. of CAD&CG, Zhejiang University, PR China

ARTICLE INFO

Article history:
Received 28 December 2008
Received in revised form
18 August 2009
Accepted 8 September 2009

Keywords:
Subspace clustering
Soft subspace
Weighted clustering
Gene expression clustering analysis
Texture image segmentation
ε-insensitive distance

ABSTRACT

While within-cluster information is commonly utilized in most soft subspace clustering approaches in order to develop the algorithms, other important information such as between-cluster information is seldom considered for soft subspace clustering. In this study, a novel clustering technique called enhanced soft subspace clustering (ESSC) is proposed by employing both within-cluster and between-class information. First, a new optimization objective function is developed by integrating the within-class compactness and the between-cluster separation in the subspace. Based on this objective function, the corresponding update rules for clustering are then derived, followed by the development of the novel ESSC algorithm. The properties of this algorithm are investigated and the performance is evaluated experimentally using real and synthetic datasets, including synthetic high dimensional datasets, UCI benchmarking datasets, high dimensional cancer gene expression datasets and texture image datasets. The experimental studies demonstrate that the accuracy of the proposed ESSC algorithm outperforms most existing state-of-the-art soft subspace clustering algorithms.

1. Introduction

Clustering techniques have been studied extensively in the areas of statistics, machine learning, and database communities in the past decades. In most clustering approaches, the data points in a given dataset are partitioned into clusters such that the points within a cluster are more similar among themselves than data points in other clusters [1]. However, conventional clustering techniques fall short when clustering is performed in high dimensional spaces [2]. For example, for any given pair of data points within the same cluster, it is possible that the points are indeed far apart from each other in a few dimensions of the high dimensional space. A key challenge to most conventional clustering algorithms is that, in many real world problems, data points in different clusters are often correlated with different subsets of features, i.e., clusters may exist in different subspaces that are comprised of different subsets of features [3].

Subspace clustering has been proposed to overcome this challenge, and has been studied extensively in recent years [4–28]. The goal of subspace clustering is to locate clusters in different subspaces of the same dataset. In general, a subspace cluster represents not only the cluster itself, but also the subspace where the cluster is situated. The two main categories of subspace clustering algorithms are hard subspace clustering and soft subspace clustering. Hard subspace clustering methods are first studied extensively for the clustering of high dimensional data. This kind of subspace clustering algorithm identifies the exact subspaces for different clusters. The hard subspace clustering algorithms can be divided into bottom-up and top–down subspace search methods [4]. Examples of bottom-up methods include CLIQUE [5], ENCLUS [6] and MAFIA [7]. Common top–down methods are ORCLUS [9], FINDIT [10], DOC [11], δ-Clusters [12], and PROCLUS [8]. Among them, PROCLUS is a representative top–down method. Other hard subspace clustering algorithms also include HARP [13] and LDR [14]. A detailed review of hard subspace clustering algorithms can be found in [4].

While the exact subspaces are identified in hard subspace clustering, a weight is assigned to each dimension in the clustering process of soft subspace clustering to measure the contribution of each dimension to the formation of a particular cluster. In the clustering procedure, each dimension (i.e. feature) contributes differently to every cluster. The subspaces of different clusters can be identified by the value of weights after clustering. Soft subspace clustering can be considered as an extension of the conventional *feature weighting clustering* [15–18], which employs a common weight

* Corresponding author at: School of Information Engineering, Jiangnan University, Wuxi, Jiangsu, PR China.
E-mail address: dzh666828@yahoo.com.cn (Z. Deng).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.09.010

vector for the whole dataset in the clustering procedure. However, it is also distinct in that different weight vectors are assigned to different clusters. From this perspective, soft subspace clustering may thus be referred to as *multiple features weighting clustering*. Soft subspace clustering has recently emerged as a hot research topic, and many algorithms have been reported [19–28]. A detailed review of soft subspace clustering will be provided in Section 2.

Although many soft subspace clustering algorithms have been developed and applied to different areas, their performance can be further enhanced. A major weakness of soft subspace clustering algorithms is that almost all of them are developed based on within-class information only, e.g. the commonly used within-cluster compactness. These algorithms are expected to be improved if more discrimination information, such as between-cluster information [29], is utilized for clustering. Motivated by this idea, we propose and develop an enhanced soft subspace clustering algorithm in this study by integrating the within-cluster and between-cluster information. The major contributions of this paper are as follows.

(1) Unlike most existing soft subspace clustering algorithms, both the within-cluster compactness and between-cluster separation are employed at the same time to develop a new fuzzy optimization objective function, which is used to derive the proposed enhanced soft subspace clustering (ESSC) algorithm.
(2) The properties of this algorithm, including its robustness and its relationship with existing algorithms, are investigated in this paper. To overcome the difficulty of parameter setting, the subspace clustering ensemble strategy is introduced to provide a more robust and stable implementation of the proposed ESSC algorithm.
(3) The performance of the proposed ESSC algorithm was investigated using real and synthetic datasets, including synthetic high dimensional datasets, UCI benchmarking datasets, high dimensional cancer gene expression datasets and texture image datasets.

The rest of this paper is organized as follows. A detailed review of the soft subspace clustering algorithms is provided in Section 2. The proposed ESSC algorithm is presented in Section 3. The integration of within-cluster and between-cluster information is discussed and the properties of the proposed algorithm are analyzed. In Section 4, the parameter setting of the proposed algorithm is put forward and discussed. Experiments and results are reported in Section 5. Finally, conclusions and future work are presented in Section 6.

2. Soft subspace clustering

In this section, the existing soft subspace clustering algorithms are reviewed. The algorithms can be divided into two main categories: *fuzzy weighting subspace clustering* and *entropy weighting subspace clustering*.

2.1. Fuzzy weighting subspace clustering

To our knowledge, the first fuzzy weighting subspace clustering algorithm proposed was the attribute weighting algorithm (AWA) [19]. The objective function of AWA is defined as

$$J_{AWA}=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}\sum_{k=1}^{D}w_{ik}^{\tau}(x_{jk}-v_{ik})^2=\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}^{\tau}\sum_{j=1}^{N}u_{ij}(x_{jk}-v_{ik})^2$$

$$\text{s.t.}\, u_{ij}\in\{0,1\},\sum_{i=1}^{C}u_{ij}=1, 0\le w_{ij}\le 1, \sum_{k=1}^{D}w_{ik}=1, \tag{1}$$

where $\mathbf{V}=[\mathbf{v}_1,\ldots,\mathbf{v}_C]$ and $\mathbf{U}=[u_{ij}]_{C\times N}$ are the cluster center matrix and the hard partition matrix respectively; C, N and D are the numbers of clusters, data and features respectively. A weight vector is assigned to each cluster and a fuzzy weighting w_{ik}^{τ} is assigned to the kth feature of the ith cluster. In order to avoid the difficulty of computation due to zero dispersion of a dimension in a cluster, Eq. (1) is modified to Eq. (2) [24]:

$$J_{FWKM}=\sum_{i=1}^{C}u_{ij}\sum_{j=1}^{N}\sum_{k=1}^{D}w_{ik}^{\tau}((x_{jk}-v_{ik})^2+\sigma)=\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}^{\tau}\Big(\sum_{j=1}^{N}u_{ij}(x_{jk}-v_{ik})^2+\sigma\Big)$$

$$\sigma=\frac{\sum_{j=1}^{N}\sum_{k=1}^{D}(x_{jk}-o_k)^2}{N\cdot D},\quad o_k=\sum_{j=1}^{N}x_{jk}/N \text{ s.t. } u_{ij}\in\{0,1\},$$

$$\sum_{i=1}^{C}u_{ij}=1, 0<\sum_{j=1}^{N}u_{ij}<N, 0\le w_{ij}\le 1, \text{and} \sum_{k=1}^{D}w_{ik}=1. \tag{2}$$

By using Eq. (2), a soft subspace clustering algorithm known as the fuzzy weighting K-means algorithm (FWKM) has been derived. A similar algorithm known as fuzzy subspace clustering (FSC) was also developed in [25]. The objective function of FSC can be formulated as Eq. (3). Detailed analysis of the properties of FSC can be found in [26].

$$J_{FSC}=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}\sum_{k=1}^{D}w_{ik}^{\tau}(x_{jk}-v_{ik})^2+\varepsilon_0\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}^{\tau}=\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}^{\tau}\Big(\sum_{j=1}^{N}u_{ij}(x_{jk}-v_{ik})^2+\varepsilon_0\Big)$$

$$\text{s.t.}\, u_{ij}\in\{0,1\},\sum_{i=1}^{C}u_{ij}=1, 0<\sum_{j=1}^{N}u_{ij}<N, 0\le w_{ij}\le 1, \text{and} \sum_{k=1}^{D}w_{ik}=1 \tag{3}$$

It is clear from Eq. (1)–(3) that a fuzzy weight w_{ik}^{τ} is assigned to the features of different clusters with a fuzzy index τ, which is a common characteristic to all the above algorithms. Hence, they are grouped together here and categorized as *fuzzy weighting subspace clustering algorithms*. The fuzzy weighting w_{ik}^{τ} can be regarded as an extension of the classical weighting w_{ik}. The fuzzy index, τ, is usually set such that $\tau>1$ in order to ensure the convergence of the derived algorithms [19,24,25]. When $\tau\to 1$, the fuzzy weighting w_{ik}^{τ} approaches the classical weighting w_{ik}. Compared with the classical weighting, the fuzzy weighting has better elasticity and, meanwhile, can be easily analyzed using fuzzy optimization techniques.

2.2. Entropy weighting subspace clustering

The latest advance in soft subspace clustering is the introduction of the concept of entropy [27,28]. Unlike fuzzy weighting subspace clustering, the weights in this kind of subspace clustering algorithm are controllable by entropy, and the developed algorithms are therefore referred to as *entropy weighting subspace clustering algorithms*. One example is the entropy weighting K-means clustering algorithm (EWKM) [27], the objective function of which can be formulated as

$$J_{EWKM}=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}\sum_{k=1}^{D}w_{ik}(x_{jk}-v_{ik})^2+\gamma\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}\ln w_{ik}$$

$$\text{s.t.}\, u_{ij}\in\{0,1\},\sum_{i=1}^{C}u_{ij}=1, 0<\sum_{j=1}^{N}u_{ij}<N, 0\le w_{ij}\le 1, \text{and} \sum_{k=1}^{D}w_{ik}=1. \tag{4}$$

Besides EWKM, entropy is also taken into account in the local adaptive clustering (LAC) algorithm for subspace clustering [28]. The objective function of LAC can be expressed as

$$J_{LAC}=\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}X_{ik}+\gamma\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}\ln w_{ik}$$
$$X_{ik}=\Big(\sum_{j=1}^{N}u_{ij}(x_{jk}-v_{ik})^2\Big)/\sum_{j=1}^{N}u_{ij} \tag{5}$$
$$\text{s.t.}u_{ij}\in\{0,1\},\sum_{i=1}^{C}u_{ij}=1,0<\sum_{j=1}^{N}u_{ij}<N0\le w_{ij}\le 1,\text{and}\sum_{k=1}^{D}w_{ik}=1.$$

By comparing Eqs. (4) and (5), it is found that their objective functions are very similar and the only difference is that the effect of cluster size is considered in Eq. (4) but omitted in Eq. (5).

Another example of entropy-based subspace clustering algorithms is the clustering objects on subsets of attributes (COSA) [23], the objective function of which is defined as follows.

$$\begin{aligned}J_{COSA}&=\sum_{i=1}^{C}\frac{\Lambda_i}{n_i^2}\sum_{z(j)=z(j')=i}\left[\sum_{k=1}^{D}(w_{ik}d_{jj'k}+\alpha w_{ik}\log w_{ik})+\alpha\log(D)\right]\\&=\sum_{i=1}^{C}\frac{\Lambda_i}{n_i^2}\sum_{k=1}^{D}\Big(w_{ik}\sum_{z(j)=z(j')=i}d_{jj'k}\Big)+\alpha\sum_{i=1}^{C}\frac{\Lambda_i}{n_i^2}\sum_{k=1}^{D}w_{ik}\log w_{ik}\\&\quad+\alpha\sum_{i=1}^{C}\frac{\Lambda_i}{n_i^2}\sum_{k=1}^{D}\log(D)\end{aligned} \tag{6}$$

As discussed in [27], the weakness of COSA is that the computation process may not be scalable to accommodate large data sets.

In addition to using fuzzy weighting and entropy weighting for soft subspace clustering, another subspace clustering algorithm, also known as local adaptive clustering [22], has also drawn considerable attention. Note that this algorithm is distinct from the entropy-based LAC approach [28] described in Eq. (5) above. To avoid confusion, the entropy weighting subspace clustering LAC is referred to as LAC2 in this paper, whereas the one discussed here is called LAC1. The objective function for LAC1 can be written as

$$J_{LAC1}=\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}\exp(h\cdot X_{ik}),$$
$$X_{ik}=\Big(\sum_{j=1}^{N}u_{ij}(x_{jk}-v_{ik})^2\Big)/\sum_{j=1}^{N}u_{ij}$$
$$\text{s.t.}u_{ij}\in\{0,1\},\sum_{i=1}^{C}u_{ij}=1,0<\sum_{j=1}^{N}u_{ij}<N,\text{ and}\sum_{k=1}^{D}w_{ik}^2=1. \tag{7}$$

Let $h=-1/\gamma$, Eq. (7) can be expressed in the following form

$$J_{LAC1}=\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}\exp(-X_{ik}/\gamma) \tag{8}$$

When $\gamma>0$, the aim of optimization is to maximize J_{LAC1} in Eq. (8).

By inspecting the existing soft subspace clustering techniques, i.e. the fuzzy weighting or entropy weighting approach, it is clear that the within-cluster information (in particular, the within-cluster compactness) is only considered to develop the corresponding algorithms. It is however anticipated that the performance of clustering can be further enhanced by including more discriminative information in the clustering process. Thus, research on the effect of additional discriminative information, such as between-clustering information, on soft subspace clustering is an important study to improve existing soft subspace algorithms. In the next section, this issue is addressed and a novel soft subspace clustering algorithm is presented.

3. Enhanced soft subspace clustering

3.1. Between-cluster separation in weighting subspaces

Among various kinds of between-cluster information, between-cluster separation is an important one that was employed to develop a fuzzy clustering algorithm in an early work by Wu et al. [29]. By considering both the fuzzy within-class compactness and the between-cluster separation as the ranks of the within-class and between-cluster scatter, the FCM-like algorithm called fuzzy compactness and separation (FCS) was proposed in [29]. An evolutionary clustering algorithm was also proposed recently using the within-cluster and between-cluster information [47]. Motivated by these advances, we further study the corresponding subspace clustering technique by introducing the between-cluster separation in the weighting subspace.

As described in [29], the fuzzy within-cluster compactness J_c and the fuzzy between-cluster separation J_s of a dataset containing C clusters can be expressed as follows:

$$J_c=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}^m(\mathbf{x}_j-\mathbf{v}_i)^T(\mathbf{x}_j-\mathbf{v}_i)=tr(S_{FW})$$

$$S_{FW}=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}^m(\mathbf{x}_j-\mathbf{v}_i)(\mathbf{x}_j-\mathbf{v}_i)^T \tag{9}$$

$$J_s=\sum_{i=1}^{C}\left(\sum_{j=1}^{N}u_{ij}^m\right)(\mathbf{v}_i-\mathbf{v}_0)^T(\mathbf{v}_i-\mathbf{v}_0)=tr(S_{FB})$$

$$S_{FB}=\sum_{i=1}^{C}\left(\sum_{j=1}^{N}u_{ij}^m\right)(\mathbf{v}_i-\mathbf{v}_0)(\mathbf{v}_i-\mathbf{v}_0)^T,\mathbf{v}_0=\left(\sum_{j=1}^{N}\mathbf{x}_j\right)/N$$

where the parameter m is the fuzzy index of fuzzy membership, as in most fuzzy framework-based clustering algorithms; S_{FW} and S_{FB} are the fuzzy within-cluster scatter and the fuzzy between-cluster scatter respectively. Correspondingly, in the weighting subspace, the weighting within-cluster compactness J_{c_w} and the fuzzy weighting within-cluster compactness J_{c_fw}, as well as the weighting between-cluster separation J_{s_w} and the fuzzy weighting between-cluster separation J_{s_fw}, are given below.

$$J_{c\_w}=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}^m\sum_{k=1}^{D}w_{ik}(x_{jk}-v_{ik})^2$$

(*the weighting within-cluster compactness*) (11)

$$J_{c\_fw}=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}^m\sum_{k=1}^{D}w_{ik}^\tau(x_{jk}-v_{ik})^2$$

(*the fuzzy weighting within-cluster compactness*) (12)

$$J_{s\_w}=\sum_{i=1}^{C}(\sum_{j=1}^{N}u_{ij}^m)\sum_{k=1}^{D}w_{ik}(v_{ik}-v_{0k})^2$$

(*the weighting between-cluster separation*) (13)

$$J_{s\_fw}=\sum_{i=1}^{C}(\sum_{j=1}^{N}u_{ij}^m)\sum_{k=1}^{D}w_{ik}^\tau(v_{ik}-v_{0k})^2$$

(*the fuzzy weighting between-cluster separation*) (14)

An intuitive explanation of the between-cluster separation in the weighting subspace is given below. As illustrated in Fig. 1, three clusters of data are centered at [0.26, 4.96], [3.67, 7.53] and

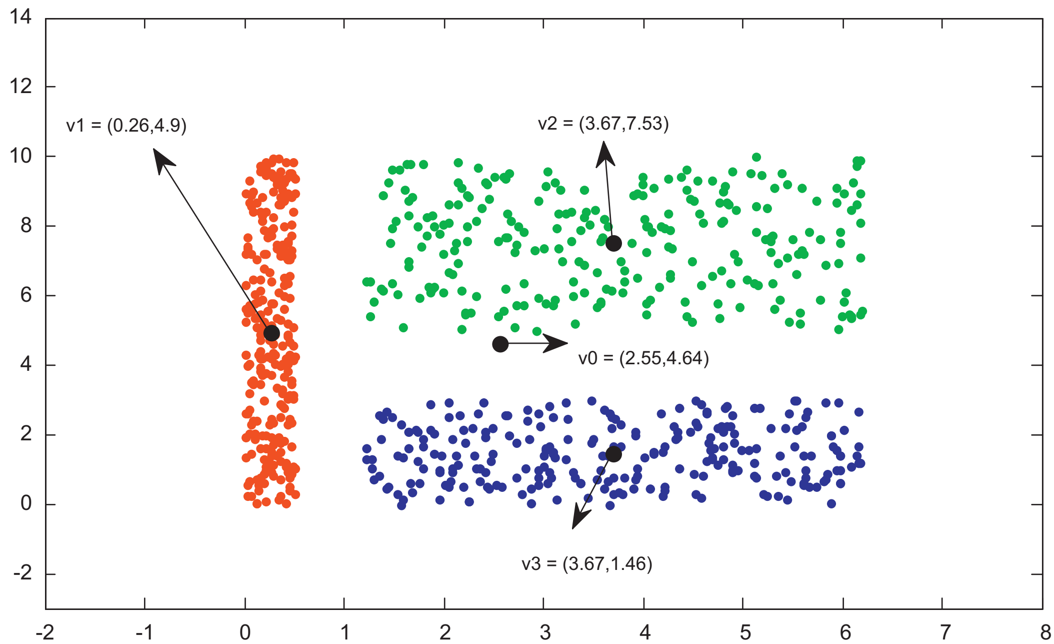

Fig. 1. Dataset containing three clusters located at different subspaces.

[3.67, 1.46] respectively. Suppose the full space is represented with weights $\mathbf{w}_1' = \mathbf{w}_2' = \mathbf{w}_3' = [0.5, 0.5]^T$, and the three subspaces associated with different clusters are represented with weights $\mathbf{w}_1 = [w_{11}, w_{12}]^T = [0.9, 0.1]^T$, $\mathbf{w}_{12} = [w_{21}, w_{22}]^T = [0.6, 0.4]^T$ and $\mathbf{w}_3 = [w_{31}, w_{32}]^T = [0.4, 0.6]^T$. The corresponding between-cluster separations can be expressed as follows:

$$J_{s\_full\ space} = \sum_{i=1}^{3} (w_{i1}' (v_{i1} - v_{01})^2 + w_{i2}' (v_{i2} - v_{02})^2) = 13.14$$

$$J_{s\_subspace} = \sum_{i=1}^{3} (w_{i1}(v_{i1} - v_{01})^2 + w_{i2}(v_{i2} - v_{02})^2) = 15.39$$

In this example, it is obvious that the between-cluster separation in the subspaces is larger than that in the full space. This means that there exist subspaces where the clusters are more scattered and can be clustered more easily when compared to that in the full space. Especially, the full space between-cluster separation and subspace between-cluster separation can also been taken as the separations obtained with different distance metrics. Thus, soft subspace clustering can be studied from the viewpoint of metric learning [49], which deserves detailed investigation in our future work.

As evident from the objective functions of the existing fuzzy and entropy weighting subspace clustering methods, i.e. Eqs. (1)–(6), only the weighting or fuzzy weighting within-cluster compactness is considered for clustering, while the between-cluster separation is not considered. Since the between-cluster separation is an important piece of information for clustering [29], it is valuable to explore new soft subspace clustering approaches by introducing the between-cluster information in soft subspace clustering to improve the performance of existing approaches. In order to achieve this goal, the incorporation of the between-cluster separation in weighting subspaces is investigated in this paper to develop an enhanced soft subspace clustering algorithm.

3.2. Objective function incorporating the between-class information

To incorporate the between-class information, the weighting and fuzzy weighting between-cluster separation can be integrated into the objective functions of existing soft subspace clustering algorithms. The proposed objective function is obtained by extending the objective function of the entropy weighting subspace clustering EWKM algorithm, where the weighting is controllable by the entropy term to a certain extent. Compared with fuzzy weighting subspace clustering, which usually controls the weighting by using a fuzzy index, entropy weighting subspace clustering demonstrates much better adaptive ability [27,28]. In this study, the objective function of EWKM is modified by extending the K-means clustering framework to the fuzzy C-mean (FCM) clustering framework and then introducing the weighting between-cluster separation. The objective function below is thus developed for the proposed ESSC algorithm,

$$\begin{aligned} J_{ESSC}(\mathbf{V}, \mathbf{W}, \mathbf{U}) &= \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^m \sum_{k=1}^{D} w_{ik}(x_{jk} - v_{ik})^2 + \gamma \sum_{i=1}^{C} \sum_{k=1}^{D} w_{ik} \ln w_{ik} \\ &\quad - \eta \sum_{i=1}^{C} \left(\sum_{j=1}^{N} u_{ij}^m \right) \sum_{k=1}^{D} w_{ik}(v_{ik} - v_{0k})^2 \\ &= J_{c\_w} + \gamma \sum_{i=1}^{C} \sum_{k=1}^{D} w_{ik} \ln w_{ik} - \eta \cdot J_{s\_w} \end{aligned}$$

$$\text{s.t.}\ 0 \le u_{ij} \le 1, \sum_{i=1}^{C} u_{ij} = 1, 0 < \sum_{j=1}^{N} u_{ij} < N, 0 \le w_{ij} \le 1,$$

$$\text{and} \sum_{k=1}^{D} w_{ik} = 1 \tag{15}$$

where $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_C]$ is the cluster center matrix, $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_C]$ is the weight matrix, and $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_N]$ is the fuzzy partition matrix. The proposed objective function contains three terms, i.e. the weighting within-cluster compactness, the entropy of weights and the weighting between-cluster separation. The first and second terms are directly inherited from the objective function of EWKM subspace clustering. In this new objective function, the parameters $\gamma\ (\gamma > 0)$ and $\eta\ (\eta > 0)$ are used to control the influences of entropy and the weighting between-cluster separation respectively. The setting of these parameters is discussed in Section 4. Note that while $u_{ij} \in \{0, 1\}$ is used in the

K-means framework-based EWKM algorithm for hard partition, we have extended it here to $0 \le u_{ij} \le 1$ in Eq. (15) for fuzzy partition. For the proposed new objective function, the aim is to minimize the within-cluster compactness in the weighting subspace while maximizing the between-cluster separation.

3.3. The enhanced algorithm: ESSC

The enhanced soft subspace clustering (ESSC) algorithm proposed in this paper is developed based on the objective function defined in Eq. (15). In addition, it is established based on the three basic theorems below.

Theorem 1. *Given that $\mathbf{V}=[\mathbf{v}_1,\dots,\mathbf{v}_C]$ and $\mathbf{W}=[\mathbf{w}_1,\dots,\mathbf{w}_C]$ are fixed, and $m>1$, $\gamma>0$, $\eta\ge 0$, J_{ESSC} in Eq. (15) reaches a local minimum only if **U** satisfies the following conditions*:

$$u_{ij}=\frac{(d_{ij})^{-1/m-1}}{\sum_{i'=1}^{C}(d_{i'j})^{-1/m-1}},\ i=1,\dots,C,\ \ j=1,\dots,N \tag{16a}$$

$$d_{ij}=\sum_{k=1}^{D} w_{ik}(x_{jk}-v_{ik})^2-\eta\sum_{k=1}^{D} w_{ik}(v_{ik}-v_{0k})^2,\quad i=1,\dots,C,\ \ j=1,\dots,N \tag{16b}$$

(The proof is given in Appendix A).

Theorem 1 specifies the necessary conditions for minimizing J_{ESSC} in Eq. (15) when **V** and **W** are fixed, and Eq. (16a) is one of the three basic rules used to develop the ESSC algorithm. However, when η in (16b) is large, d_{ij} and u_{ij} may become negative. This situation contradicts the requirement that $0\le u_{ij}\le 1$ as defined in Eq. (15). It is therefore necessary to impose a constraint on η. For $\mathbf{x}_j$, the fuzzy membership u_{ij} $(i=1,\dots,C)$ to the ith cluster, obtained by Eq. (16a), will be negative when $d_{ij}<0$. When d_{ij} is negative, it means that the first term in Eq. (16b) is less than the second term. In order to obtain a non-negative value of u_{ij}, we must have $d_{ij}\ge 0\,(i=1,\dots,C)$, i.e. $\sum_{k=1}^{D} w_{ik}(x_{jk}-v_{ik})^2 \ge \eta\sum_{k=1}^{D} w_{ik}(v_{ik}-v_{0k})^2$ in Eq. (16b). Hence, the adaptive rule in Eq. (16c) is imposed on η to avoid negative d_{ij} and u_{ij}.

$$\eta=\min\left\{\min_i\Big(\sum_{k=1}^{D} w_{ik}(x_{jk}-v_{ik})^2\Big/\sum_{k=1}^{D} w_{ik}(v_{ik}-v_{0k})^2\Big),\eta\right\}\quad \forall j=1,\dots,N \tag{16c}$$

For $\mathbf{x}_j$, if

$$\eta<\min_i\Big(\sum_{k=1}^{D} w_{ik}(x_{jk}-v_{ik})^2\Big/\sum_{k=1}^{D} w_{ik}(v_{ik}-v_{0k})^2\Big)$$

the value of η is not changed. Otherwise, η is replaced by

$$\min_i\Big(\sum_{k=1}^{D} w_{ik}(x_{jk}-v_{ik})^2\Big/\sum_{k=1}^{D} w_{ik}(v_{ik}-v_{0k})^2\Big).$$

Hence, d_{ij} in Eq. (16b) and u_{ij} in Eq. (16c) will always be positive, and u_{ij} can be used effectively as the fuzzy membership for fuzzy partition.

Theorem 2. *Given that $\mathbf{U}=[\mathbf{u}_1,\dots,\mathbf{u}_N]$ and $\mathbf{W}=[\mathbf{w}_1,\dots,\mathbf{w}_C]$ are fixed, and $m>1$, $\gamma>0$, $0\le\eta<1$, J_{ESSC} reaches the local minimum if and only if **V** meets the following conditions*:

$$v_{ik}=\frac{\sum_{j=1}^{N} u_{ij}^m(x_{jk}-\eta v_{0k})}{\sum_{j=1}^{N} u_{ij}^m(1-\eta)} \tag{17}$$

(The proof is given in Appendix B.)

Theorem 2 specifies the necessary and sufficient conditions for minimizing J_{ESSC} in Eq. (15) when **U** and **W** are fixed. Eq. (17) is the second rule used for developing the ESSC algorithm. It is proved in Appendix B that when $0\le\eta<1$, Eq. (17) is a sufficient and necessary condition to minimize J_{ESSC} when **U** and **W** are fixed, and therefore it is appropriate to set $0\le\eta<1$ in the ESSC algorithm.

Theorem 3. *Given that $\mathbf{U}=[\mathbf{u}_1,\dots,\mathbf{u}_N]$ and $\mathbf{V}=[\mathbf{v}_1,\dots,\mathbf{v}_C]$ are fixed, and $m>1$, $\gamma>0$, $\eta\ge0$, J_{ESSC} reaches the local minimum only if W meets the following conditions*:

$$w_{ik}=\exp\left(-\frac{\sigma_{ik}}{\gamma}\right)\Big/\sum_{k'=1}^{D}\exp\left(-\frac{\sigma_{ik'}}{\gamma}\right) \tag{18a}$$

$$\sigma_{ik}=\sum_{j=1}^{N} u_{ij}^m(x_{jk}-v_{ik})^2-\eta\sum_{j=1}^{N} u_{ij}^m(v_{ik}-v_{0k})^2 \tag{18b}$$

(The proof is given in Appendix C.)

Theorem 3 gives the necessary conditions for minimizing J_{ESSC} in Eq. (15) when **U** and **V** are fixed, and Eq. (18a) is the third rule used for developing the ESSC algorithm. Based on these three rules, the proposed soft subspace clustering algorithm is described in Table 1.

It is noteworthy to point out that, while the FCS algorithm [29] concerns the fuzzy within-cluster compactness and between-cluster separation in the original full space, the proposed ESSC algorithm is indeed an extension of FCS in the weighting space. On the other hand, when $m\to1$ and $\eta=0$, the ESSC algorithm degenerates into the EWKM algorithm [27]. Thus, the EWKM algorithm can be regarded as a special case of the proposed ESSC algorithm.

The three convergence theorems above are obtained by employing a strategy [48] similar to that used in most iteration-based clustering algorithms, such as FCM and MEC algorithms. It can be inferred that the proposed ESSC algorithm, at least along a subsequence, converges to either a local optimal solution or a saddle point of its objective function.

3.4. Robustness analysis based on ε-insensitive function

The robustness of the proposed ESSC algorithm is discussed from the perspective of robust distance in this subsection. One of the most commonly used robust distances is the Vapnik's ε-insensitive distance [30–32], which demonstrates promising robustness against noise and outliers [30–32]. If d is a distance metric, by using the ε-insensitive loss function, the corresponding robust ε-insensitive distance d_ε is defined as

$$d_\varepsilon=\begin{cases} d-\varepsilon & \text{if } d\ge\varepsilon \\ 0 & \text{if } d<\varepsilon\end{cases} \tag{19}$$

Table 1
The enhanced subspace clustering (ESSC) algorithm.

Algorithm ESSC
Input: The number of cluster C, parameter $\gamma>0$, $0\le\eta<1$, threshold $\xi>0$ and the maximal number of iterations *MaxIter*. Randomly initialize cluster centers $\mathbf{V}(0)$ and set the initial weight matrix $\mathbf{W}(0)$ with $w_{ik}=1/D$ and $t=0$.
Repeat:
t=t+1;
Compute the partition matrix $\mathbf{U}(t)$ by (16a);
Compute the cluster center matrix $\mathbf{V}(t)$ by (17);
Compute the weight matrix $\mathbf{W}(t)$ by (18a);
Until $\Vert\mathbf{V}(t)-\mathbf{V}(t-1)\Vert<\xi$ or $t=MaxIter$.

Interestingly, the proposed objective function in Eq. (15) and the derived update rule in Eq. (16a) can be rewritten as

$$J_{ESSC}(\mathbf{V},\mathbf{W},\mathbf{U}) = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m}(\overline{d}_{ij})_{\varepsilon_{ij}} + \gamma \sum_{i=1}^{C}\sum_{k=1}^{D} w_{ik}\ln w_{ik} \tag{20}$$

$$u_{ij} = \frac{((\overline{d}_{ij})_{\varepsilon_{ij}})^{-1/m-1}}{\sum_{i'=1}^{C}((\overline{d}_{i'j})_{\varepsilon_{i'j}})^{-1/m-1}} \tag{21}$$

where $\overline{d}_{ij} = \sum_{k=1}^{D} w_{ik}(x_{jk}-v_{ik})^2$ and $\varepsilon_{ij} = \eta\sum_{k=1}^{D} w_{ik}(v_{ik}-v_{0k})^2$. Eqs. (20) and (21) indicate that J_{ESSC} and u_{ij} can be taken as the robust ε-insensitive distance $(\overline{d}_{ij})_{\varepsilon_{ij}}$ based objective function and update rule, respectively. From the above analyses, we can see that the proposed ESSC algorithm is a robust ε-insensitive distance-based clustering algorithm. Thus, it has similar properties as other robust ε-insensitive clustering algorithms, such as the robust FCM algorithm [30] and the robust MEC algorithm [31].

4. Parameter setting

Like most FCM framework-based algorithms, it is necessary to set the fuzzy index m of fuzzy membership appropriately in the ESSC algorithm. Theoretical results on the study of setting the fuzzy index have been obtained based on the traditional full space fuzzy clustering model [33,34], but the corresponding theoretical study for the subspace fuzzy clustering model is not trivial and requires in-depth investigation. In this study, although the theoretical results of full space fuzzy clustering are directly adopted for the ESSC algorithm, following a considerable number of experiments, we find empirically that parameter setting based on these theoretical results is appropriate in most cases. The theoretical study on the setting of the fuzzy index for the subspace fuzzy clustering model will be a major part of our future research. In the proposed ESSC algorithm, a simple but effective approach is adopted to determine the appropriate value of m [34]. Suppose N and D are the size of the data and the features respectively, if $\min(N,D-1)\geq 3$, m should be within the range as specified by the inequality $1<m\leq \min(N,D-1)/\min(N,D-1)-2$; otherwise, m is set to 2. For high dimensional datasets, $\min(N,D-1)\geq 3$ almost always holds and it is suitable to set $m=\min(N,D-1)/\min(N,D-1)-2$ for simplicity. Based on a large number of experiments, we find that parameter setting based on the above rule is appropriate in most cases.

It is also necessary to set η and γ appropriately in the proposed ESSC algorithm. The parameter η is to maintain a balance between the effect of within-class compactness and that of between-cluster separation on subspace clustering. It can also be regarded as a parameter of the ε-insensitive distance from the perspective of robust distance, as discussed in Section 3.4. According to Theorem 2, η can be selected within the range $0\leq\eta<1$. However, it is difficult to directly choose a suitable value since the domain

Table 2
Parameter setting of six subspace clustering algorithms.

Algorithm	Parameter settings
ESSC	$m=\frac{\min(N,D-1)}{\min(N,D-1)-2}$; $\gamma=1,2,5,10,50,100,1000$; $\eta=0,0.01,0.05,0.1,0.2,0.3,0.5,0.7,0.9$
EWKM, LAC1 and LAC2	$\gamma=1,2,5,10,50,100,1000$
FSC	$\varepsilon_0=0,10^{-10},10^{-5},10^{-4},10^{-3},10^{-2}$; $\tau=1.05,1.2,1.5,2,5,10,20,50$
PROCLUS	$d_{ave}=2,3,\ldots,D$ for synthetic dataset, UCI data sets and texture image data sets and $d_{average}=2^0,2^1,\ldots,2^{\lfloor\log_2(D)\rfloor}$ for gene expression data sets (d_{ave} is the average dimensional number expected to be selected for each cluster; D is the number of whole dimensions)

Table 3
The best clustering results obtained for the synthetic dataset with RI as metric.

	ESSC	EWKM	LAC1	LAC2	FSC	PROCLUS
Mean	0.9248	0.9111	0.9150	0.9123	0.9179	0.7724
Std	0.0645	0.0706	0.0731	0.0637	0.0633	0.0133

Table 4
The best clustering results obtained for the synthetic dataset with NMI as metric.

	ESSC	EWKM	LAC1	LAC2	FSC	PROCLUS
Mean	0.7459	0.7139	0.7276	0.7036	0.7279	0.2460
Std	0.2288	0.2299	0.2229	0.2230	0.2203	0.0730

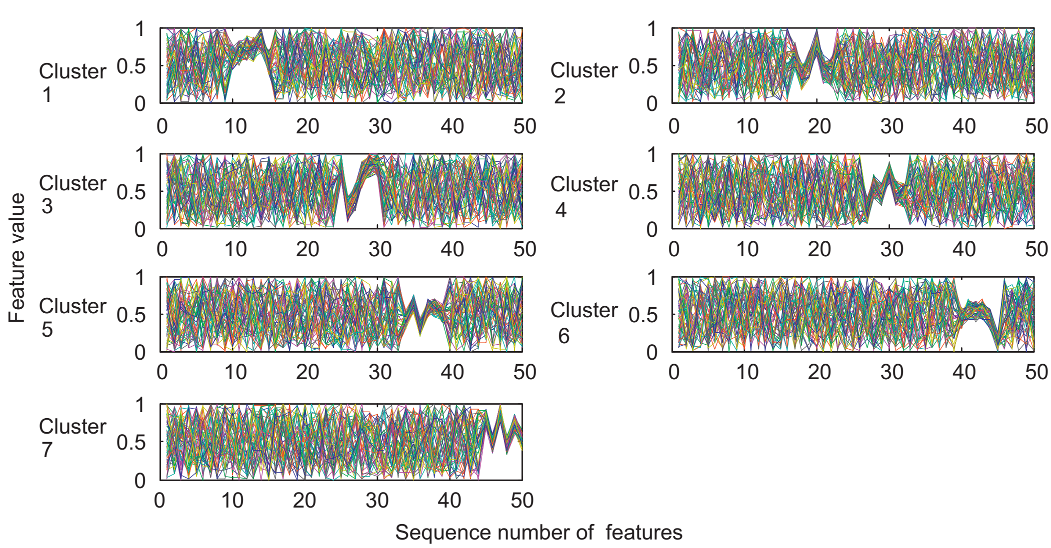

Fig. 2. The distribution of data in seven clusters that locate at different subspaces.

Table 5
Accuracy of subspace detection of different algorithms for the synthetic dataset.

Sequence number of relevant dimensions		Sequence number of detected dimensions in each cluster														
		ESSC			EWKM			LAC1			LAC2			FSC		
Cluster1	10–15	14	11	13	11	14	15	11	12	14	11	12	13	12	11	13
		12	10	15	12	13	10	13	15	10	10	15	14	15	10	14
Cluster2	17–22	20	21	17	21	17	20	21	20	17	21	20	17	21	17	20
		18	19	22	18	19	22	19	18	22	19	18	22	18	19	22
Cluster3	25–30	30	28	27	30	28	27	30	28	27	30	28	25	30	28	29
		25	29	26	**31**	**32**	29	**32**	**31**	29	27	26	29	27	25	26
Cluster4	27–32	32	30	28	30	28	27	30	28	27	**17**	**48**	**12**	30	28	32
		31	27	**35**	32	31	29	32	31	29	**5**	**49**	28	27	31	**35**
Cluster5	34–39	36	39	38	39	38	37	39	35	38	35	34	37	34	36	39
		35	34	37	34	36	35	37	36	34	38	39	36	35	38	37
Cluster6	40–45	45	43	42	45	43	42	45	40	42	40	45	41	41	42	43
		40	44	41	40	44	41	43	41	44	44	43	42	45	40	44
Cluster7	45–50	47	45	46	46	50	48	50	48	47	48	50	47	46	50	48
		50	48	49	45	47	49	46	49	45	49	45	46	45	47	49
Accuracy of subspace detection		41/42			40/42			40/42			37/42			41/42		

Table 6
Running time (seconds) of algorithms performed on the synthetic dataset.

	ESSC	EWKM	LAC1	LAC2	FSC	PROCLUS
Mean	0.1718	0.1437	0.1391	0.1344	0.1344	0.1375
Std	0.0413	0.0161	0.0187	0.0223	0.0237	0.0131

Table 7
The eight UCI datasets.

Dataset	Size of dataset	Number of dimensions	Number of clusters
Australian	690	14	2
Balance-scale	625	4	3
Car1	392	7	3
Glass	214	9	6
Heart	270	13	2
Iris	150	4	3
Vehicle	846	18	4
Wine	178	13	3

knowledge for tuning η for different datasets is not available. The parameter γ in the ESSC algorithm is used to control the influence of entropy in the objective function. Similar to η, an appropriate setting of γ also depends on the domain knowledge of the datasets, which has been discussed in the context of entropy weighting subspace clustering algorithms [28].

In order to overcome this difficulty, the clustering ensemble strategy weighted bipartite partitioning algorithm (WBPA) [28,35] can be introduced to provide a more robust and stable implementation of the proposed ESSC algorithm. The procedure for implementing the WBPA clustering ensemble for the ESSC algorithm can be found in Appendix D. By using the WBPA ensemble strategy for the ESSC algorithm, the weight vector corresponding to the ith cluster obtained by ensemble clustering can be computed by averaging the weight vectors associated with the cluster centers that have been partitioned into the ith cluster in graph partition [35].

5. Performance evaluation

The proposed ESSC algorithm has been evaluated, with a large number of experiments performed on synthetic and real datasets. In this section, the metrics used for performance evaluation and parameter setting are first described. Then, the performance of the algorithm on the synthetic high dimensional datasets, UCI datasets, high dimensional gene expression datasets and synthetic texture image datasets is presented. A detailed comparison with other subspace clustering algorithms is also performed. All the experiments were implemented on a computer with 1.66 GHz CPU and 1 GB RAM.

Table 8
The best clustering results obtained for the eight UCI datasets with RI as metric.

Dataset	ESSC	EWKM	LAC1	LAC2	FSC	PROCLUS
Australian						
Mean	**0.7104**	0.6877	0.6575	0.6576	0.5573	0.5897
Std	0.0218	0.0850	0.0924	0.0972	0.0370	0.0672
Balance-scale						
Mean	**0.6396**	0.5943	0.5935	0.5974	0.5949	0.5624
Std	0.0697	0.0296	0.0157	0.0172	0.0737	0.0245
Car1						
Mean	0.5335	0.4994	0.4901	0.4890	**0.6044**	0.5172
Std	0.0469	0.0363	0.0352	0.0328	0.0428	0.0476
Glass						
Mean	0.6990	0.6610	0.6742	0.6741	0.6534	**0.7044**
Std	0.0284	0.0760	0.0158	0.0144	0.0166	0.0182
Heart						
Mean	**0.6878**	0.6735	0.6510	0.6613	0.6529	0.5770
Std	0.0352	0.0058	0.0570	0.0387	0.0321	0.0511
Iris						
Mean	**0.8786**	0.8785	0.8623	0.8680	0.8370	0.7951
Std	0.0051	0.0027	0.0337	0.0228	0.0740	0.0908
Vehicle						
Mean	**0.6561**	0.6509	0.6532	0.6537	0.6489	0.6518
Std	0.00877	0.0044	0.0026	0.0025	0.0278	0.0134
Wine						
Mean	**0.9475**	0.9310	0.9326	0.9323	0.8827	0.7560
Std	0.0064	0.0054	0.0083	0.0081	0.0055	0.0822

5.1. Performance metrics and experimental setup

Two metrics, the *rand index* (RI) and the *normalized mutual information* (NMI) [36], are used for evaluating the performance of the proposed ESSC algorithm. RI is defined as follows,

$$RI = \frac{f_{00} + f_{11}}{N(N-1)/2}$$

where f_{00} is the number of pairs of data points having different class labels and belonging to different clusters; f_{11} is the number of pairs of data points having the same class labels and belonging to the same clusters; N is the size of the whole dataset. NMI is defined and computed according to the formula below,

$$NMI = \frac{\sum_{i=1}^{C}\sum_{j=1}^{C} N_{i,j}\log N \cdot N_{i,j}/N_i \cdot N_j}{\sqrt{\sum_{i=1}^{C} N_i \log N_i/N \cdot \sum_{j=1}^{C} N_j \log N_j/N}}$$

where $N_{i,j}$ is the number of agreements between cluster i and class j, N_i is the number of data points in cluster i, N_j is the number of data points in class j, and N is the size of the whole dataset. Both RI and NMI take a value within the interval [0,1]. The higher the values, the better the clustering performance.

The ESSC algorithm is compared with five algorithms: four soft subspace clustering algorithms, namely EWKM, FSC, LAC1, LAC2; and one hard subspace clustering algorithm, PROCLUS. Different parameters are used in these six algorithms, and their settings are tabulated in Table 2. The algorithms are applied to synthetic datasets, UCI datasets, gene expression datasets and texture image datasets. The clustering process of the algorithms is repeated 10 times at each setting. In these experiments, all the datasets are preprocessed by normalizing the feature in each dimension into the interval [0, 1]. Furthermore, the maximal number of iterations is set to 20, which is the termination condition for all the soft subspace clustering algorithms.

Table 9
The best clustering results obtained for the eight UCI datasets with NMI as metric.

Dataset	ESSC2	EWKM	LAC1	LAC2	FSC	PROCLUS
Australian						
Mean	0.3617	0.3107	0.3254	0.2578	**0.4279**	0.2575
Std	0.0168	0.1537	0.1457	0.1637	6.2e-17	0.1707
Balance-scale						
Mean	**0.2278**	0.1285	0.1285	0.1345	0.1459	0.0874
Std	0.1407	0.0593	0.0509	0.0326	0.1482	0.0561
Car1						
Mean	**0.1736**	0.1617	0.1504	0.1536	0.1280	0.1449
Std	0.0076	0.0037	0.0184	0.0162	0.0258	0.0196
Glass						
Mean	**0.3505**	0.3460	0.3475	0.3495	0.2404	0.3437
Std	0.0370	0.0243	0.0205	0.0273	0.1138	0.0321
Heart						
Mean	**0.3067**	0.2709	0.2376	0.2520	0.0501	0.1346
Std	0.0471	0.0081	0.0849	0.0605	0.0472	0.0778
Iris						
Mean	**0.7419**	0.7407	0.7183	0.7165	0.6944	0.6461
Std	0.0054	0.0058	0.0476	0.0329	0.1057	0.1323
Vehicle						
Mean	0.1431	0.1274	0.1236	0.1176	0.1497	**0.1773**
Std	0.0373	0.0240	0.0303	0.0289	0.0600	0.0294
Wine						
Mean	**0.8629**	0.8317	0.8346	0.8330	0.7334	0.5257
Std	0.0508	0.0421	0.0566	0.0478	0.0400	0.1436

5.2. Synthetic dataset

In this subsection, a synthetic dataset with controlled cluster structures is used to investigate the performance of the proposed ESSC algorithm. The dataset contains seven clusters located at different subspaces as indicated in the second column of Table 5. For each cluster, the distribution of data points in the irrelevant dimensions is uniform and the distribution of data points in each cluster is shown in Fig. 2. The performance of the ESSC algorithm is compared with that of the other five subspace clustering algorithms by using this dataset. The best clustering results expressed in terms of the means and standard deviations of the RI and NMI values, obtained by executing each algorithm 10 times, are tabulated in Tables 3 and 4 respectively.

The subspace detection results of the five soft subspace clustering algorithms are compared in Table 5. In this table, the n features detected correspond to the top n weights of each cluster, and the *boldface* numbers denote the mistakenly detected sequence numbers of irrelevant features. The running time of these algorithms performed on the synthetic dataset is presented in Table 6.

The experimental results indicate that (1) the clustering accuracy of the ESSC algorithm is better than that of the other algorithms; (2) the accuracy of subspace detection of the ESSC algorithm is competitive with the state-of-the-art algorithms; (3)

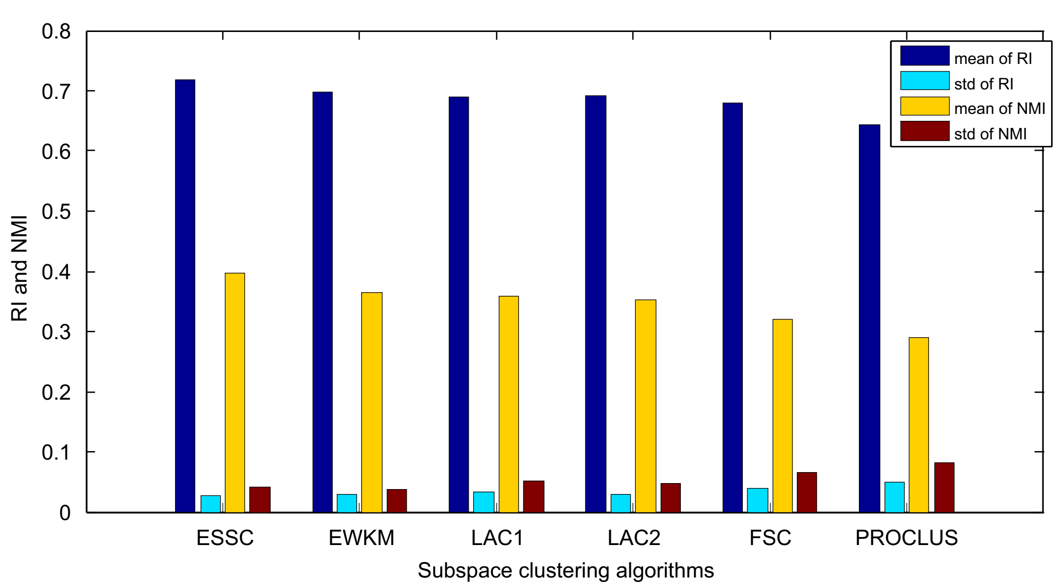

Fig. 3. The averages of the best clustering results obtained for the UCI datasets.

the running time of ESSC is slightly longer than those of the other four soft subspace clustering algorithms for the synthetic dataset.

Table 10
Clustering results by WBPA ensemble strategy for the *Wine* dataset.

Metric	ESSC	EWKM	LAC1	LAC2	FSC
RI					
Mean	**0.9302**	0.9241	0.9264	0.9263	0.7325
Std	0.0034	0.0036	0.0039	0.0037	0.0631
NMI					
Mean	**0.8287**	0.8163	0.8210	0.8188	0.4085
Std	0.0157	0.0220	0.0182	0.0323	0.1709

Table 11
Running time (seconds) of algorithms performed on the eight UCI datasets.

Dataset	ESSC2	EWKM	LAC1	LAC2	FSC	PROCLUS
Australian						
Mean	0.1148	0.0665	0.0633	0.0641	0.0609	0.1598
Std	0.0127	0.0189	0.0139	0.0143	0.0142	0.0119
Balance-scale						
Mean	0.0664	0.0492	0.0516	0.0516	0.0422	0.3802
Std	0.0151	0.0127	0.0114	0.0114	0.0125	0.0211
Car1						
Mean	0.0719	0.0406	0.0430	0.0430	0.0445	0.0977
Std	0.0186	0.0147	0.0133	0.0123	0.0170	0.0117
Glass						
Mean	0.0891	0.0523	0.0570	0.0516	0.0508	0.1719
Std	0.0161	0.0105	0.0154	0.0103	0.0086	0.0239
Heart						
Mean	0.0555	0.0305	0.0344	0.0313	0.0320	0.0904
Std	0.0129	0.0107	0.0120	0.0124	0.0139	0.0087
Iris						
Mean	0.0406	0.0219	0.0242	0.0242	0.0258	0.0662
Std	0.0118	0.0128	0.0107	0.0119	0.0146	0.0121
Vehicle						
Mean	0.2128	0.1453	0.1430	0.1469	0.1414	0.4186
Std	0.0167	0.0161	0.0116	0.0128	0.0212	0.0351
Wine						
Mean	0.0602	0.0352	0.0328	0.0383	0.0305	0.1115
Std	0.0127	0.0133	0.0133	0.0119	0.0119	0.0255

5.3. UCI datasets

The performance of the proposed ESSC algorithm has been evaluated and compared with five existing subspace clustering algorithms using the eight UCI datasets. The datasets are shown in Table 7 [37]. Experiments are conducted with the UCI datasets by running each algorithm 10 times. The best clustering results, expressed in terms of the means and standard deviations of the RI and NMI values, are tabulated in Tables 8 and 9 respectively. It is evident from Table 8 that among the six algorithms, ESSC demonstrates the best performance in the clustering of the eight UCI datasets. The performance of EWKM, LAC1 and LAC2 are comparable or better than that of FSC and PROCLUS. However, even though FSC and PROCLUS are in general inferior to the other four algorithms, they are able to achieve the best clustering performances for the dataset *Car*1 and *Glass* respectively, with RI as the metric. This indicates that there is no single algorithm that is always superior to the others for all the datasets. Similar results can be found in Table 9. By comparing Tables 8 and 9, it is further noticed that the best clustering performance as indicated by RI is not always consistent with that indicated by NMI, i.e. an algorithm showing good clustering performance with a high RI value may not have a high NMI value as well. Therefore, it is necessary to evaluate the performance of a clustering algorithm with different metrics. The averages of the best clustering results obtained from these six algorithms are plotted in Fig. 3.

On the other hand, the effect of adopting the WBPA ensemble strategy in the proposed ESSC algorithm is studied. To evaluate the performance of the ESSC clustering ensemble and to ensure that the comparison is made on equal ground, the WBPA ensemble strategy [28,35] is also implemented for the other four soft subspace clustering algorithms (i.e. EWKM, LAC1, LAC2, FSC). Here, the clustering results obtained with the parameter

Table 13
The five cancer gene expression datasets.

Dataset	Size of dataset	Number of dimensions	Number of clusters	Source
CNS	34	7129	2	[38]
Prostate3	33	12626	2	[39]
Breast	84	9216	5	[40]
DLBCL	88	4026	6	[41]
Lung2	203	12600	5	[42]

Table 12
The 10 features detected in each cluster with the *Wine* dataset.

Algorithm	Cluster	Sequence number of the detected 10 features in each cluster									
ESSC	1	13	4	5	9	2	12	8	1	3	6
	2	8	2	10	1	9	4	6	11	5	3
	3	8	2	6	3	9	11	4	12	5	7
EWKM	1	8	2	10	1	9	6	5	4	11	3
	2	13	1	4	5	3	8	9	12	6	2
	3	2	8	11	9	12	3	4	6	1	7
LAC1	1	8	2	10	1	9	4	6	5	3	12
	2	8	2	9	3	12	11	6	4	5	7
	3	13	4	5	1	2	9	8	3	12	6
LAC2	1	8	2	6	3	9	11	4	5	12	7
	2	2	8	10	1	9	4	6	11	5	3
	3	13	4	5	9	2	12	8	1	3	6
FSC	1	8	2	10	1	9	4	5	6	11	3
	2	8	2	11	6	9	3	4	12	1	7
	3	13	4	5	1	3	9	8	12	6	2

Table 14
The best clustering results obtained for the gene expression datasets with RI as metric.

Dataset	ESSC	EWKM	LAC1	LAC2	FSC	PROCLUS
CNS						
Mean	**0.5805**	0.5538	0.5187	0.5112	0.5591	0.5627
Std	0.0578	0.0430	0.0165	0.0188	0.0358	0.0350
Prostate3						
Mean	**0.8573**	0.7437	0.7288	0.7909	0.5909	0.7854
Std	0.1422	0.1218	0.1126	0.1555	0.1193	0.0948
Breast						
Mean	0.7318	0.7252	0.7344	**0.7573**	0.7381	0.7176
Std	0.0166	0.0189	0.0181	0.0484	0.0368	0.0160
DLBCL						
Mean	**0.8585**	0.7703	0.8074	0.7990	0.7381	0.7355
Std	0.0212	0.0782	0.0514	0.0269	0.0368	0.0180
Lung2						
Mean	**0.7536**	0.5871	0.5894	0.5923	0.5816	0.5526
Std	0.0871	0.0114	0.0058	0.0262	0.0522	0.0232

Table 15
The best clustering results obtained for the gene expression datasets with NMI as metric.

Dataset	ESSC	EWKM	LAC1	LAC2	FSC	PROCLUS
CNS						
Mean	**0.1312**	0.1051	0.0850	0.0744	0.0114	0.0829
Std	0.0870	0.0571	0.0969	0.0715	0.0093	0.0800
Prostate3						
Mean	**0.6889**	0.4761	0.4925	0.6013	0.2138	0.4399
Std	0.2150	0.0613	0.1437	0.2538	0.0798	0.2414
Breast						
Mean	**0.4917**	0.4341	0.4589	0.4634	0.3544	0.3970
Std	0.0994	0.0343	0.0141	0.0554	0.0905	0.1098
DLBCL						
Mean	**0.7526**	0.6023	0.6993	0.6901	0.5006	0.5408
Std	0.0683	0.0720	0.0783	0.0680	0.0839	0.0418
Lung2						
Mean	**0.5427**	0.3433	0.3346	0.3373	0.2150	0.2405
Std	0.0281	0.0294	0.0098	0.0579	0.1395	0.0737

settings in Table 2 are adopted for this study. For each UCI dataset, the clustering ensemble of each algorithm is executed repeatedly 10 times. The experimental results, expressed in terms of the means and standard deviations of RI and NMI, show that the ESSC algorithm usually demonstrates much better performance than the other soft subspace clustering algorithms. For example, from the ensemble clustering results obtained with the *Wine* dataset as shown in Table 10, it is evident that the ESSC algorithm outperforms other soft subspace clustering algorithms. Similar results are also obtained for the other UCI datasets except the *Car*1 dataset. In particular, when the WBPA ensemble strategy is employed, the standard deviations of the clustering results are usually smaller than that obtained by the ESSC algorithm directly. As an example, the standard deviation of the best clustering results obtained by executing the ESSC algorithm alone 10 times are 0.0064 and 0.0508 respectively with RI and NMI as metrics, while they are 0.0034 and 0.0157 respectively when the WBPA ensemble strategy is used. Thus, the ensemble strategy can indeed improve the stability of the proposed ESSC algorithm.

The running time of the six algorithms performed on the UCI datasets is measured and reported in Table 11. We can see that the CPU running time of the ESSC algorithm is much longer than that of the other four K-means framework-based subspace clustering algorithms. This is due to the fact that the ESSC algorithm is based on the FCM framework, and that the between-cluster separation and fuzzy membership are all required in the clustering process.

In Table 12, the subspace detection results of subspace algorithms performed on the UCI *Wine* dataset are reported, where the 10 features detected correspond to the top 10 weights of each cluster. From Table 12, we can see that different subspace clustering algorithms may assign different levels of importance to the same feature. Similar results can be obtained with other UCI datasets.

5.4. Gene expression datasets

In this experiment, five real cancer gene expression datasets [38–42] are used to test the performance of the proposed ESSC algorithm. The datasets are summarized in Table 13. Like many bioinformatics datasets, the cancer gene expression datasets used here contain a small number of samples but a large number of features, suffering from the curse of dimensionality. In terms of

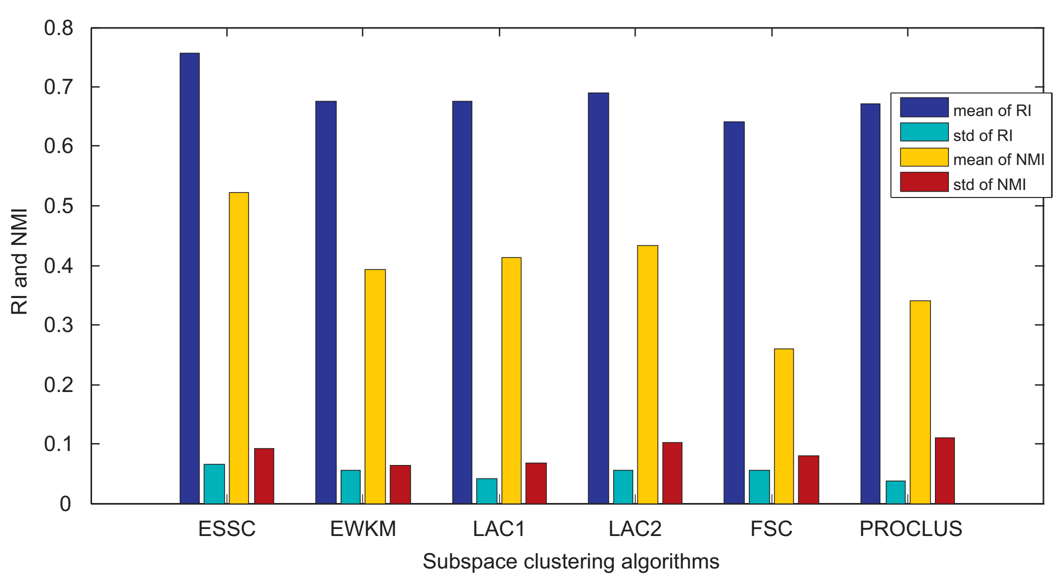

Fig. 4. The averages of the best clustering results obtained for the gene expression datasets.

the means and standard deviations of the RI and NMI values, Tables 14 and 15 show the best clustering results obtained by executing each of the six algorithms 10 times. The averages of the best clustering results obtained by using the five datasets are shown in Fig. 4. The results indicate that the ESSC algorithm is highly competitive among the state-of-the-art algorithms for the cancer gene expression datasets. Similar to the test on UCI datasets, the WBPA ensemble strategy is adopted for all five soft subspace clustering algorithms to evaluate the performance of ESSC ensemble clustering. The clustering results of these five algorithms on the cancer gene expression *DLBCL* dataset are shown and compared in Table 16. The findings obtained here are similar to that in Section 5.3, which demonstrates the advantages of the proposed algorithm once again. The detected subspaces associated with the top 10 weights of each cluster by different subspace algorithms for the *CNS* gene expression dataset are compared in Table 17. It is also found that the sequence number of detected features achieved by these algorithms is diverse. Table 18 shows the running time of the six algorithms performed on the gene expression datasets.

5.5. Texture image datasets

Finally, the proposed ESSC algorithm is experimented with texture image datasets to evaluate its performance in texture image segmentation. The texture image adopted in this experiment is as shown in Fig. 5(a), and is synthesized with seven kinds of texture images (D3, D6, D21, D49, D53, D56, D93) in the Brodatz texture database [43]. The synthesized texture image is resized into 100×100 pixels in our experiment. Gaussian noise with zero mean but different standard deviation is then introduced to this image to generate six noisy texture images, which are shown in Fig. 5(b)–(g). The proposed ESSC algorithm and the other five subspace clustering algorithms are applied for texture image segmentation. The procedure of the segmentation process is as follows.

(1) The Gabor filter presented in [44] is first used to extract the features of the texture images. A filter bank with six orientations (one at every 30°) and five frequencies starting from 0.46 is created. Then, 30 dimensional features are extracted from each pixel of the images by applying the filter bank to these texture images.
(2) A dataset containing data with 30 dimensional features is constructed for each texture image, and the size of the dataset is 10 000.
(3) The high dimensional datasets are clustered with different subspace clustering algorithms. Each of the clusters obtained is regarded as one subsection of the segmented image.

The ideal segmentation result of these texture images is shown in Fig. 5(h), which is used as a reference to quantitatively determine the segmentation performance of the algorithms. Tables 19 and 20 show the best clustering results obtained by running each algorithm 10 times. They are expressed in terms of the means and standard deviations of RI and NMI values. It is evident from these results that the proposed ESSC algorithm demonstrates better image segmentation performance than the other algorithms. Table 21 shows the running time of the six algorithms performed on the texture images shown in Fig. 5(a) and (f).

6. Conclusions

In this study, the enhanced soft subspace clustering algorithm is proposed and developed by considering both within-cluster and between-class information. The work involves the following aspects: (1) a novel objective function integrating the

Table 16
Clustering results by WBPA ensemble strategy for the DLBCL dataset.

Metric	ESSC	EWKM	LAC1	LAC2	FSC
RI					
Mean	**0.7526**	0.7391	0.7385	0.7364	0.6925
Std	0.0150	0.0094	0.0110	0.0113	0.0205
NMI					
Mean	**0.6064**	0.5718	0.5581	0.5715	0.3676
Std	0.0309	0.0302	0.0399	0.0308	0.0609

Table 18
Running time (seconds) of algorithms performed on the gene expression datasets.

Dataset	ESSC	EWKM	LAC1	LAC2	FSC	PROCLUS
CNS						
Mean	3.23	3.16	4.87	2.55	2.19	18.30
Std	0.0875	0.1418	0.1779	0.0570	0.0129	11.48
Prostate3						
Mean	6.34	6.20	5.6727	5.03	4.10	40.86
Std	0.1775	0.5180	0.3107	0.1230	0.0140	29.13
Breast						
Mean	26.87	20.56	21.869	21.70	16.44	67.25
Std	0.0361	0.2458	0.1008	0.1544	0.8906	34.40
DLBCL						
Mean	11.55	9.52	12.01	11.98	9.06	33.21
Std	0.4040	0.1132	70	0.3271	0.4419	14.08
Lung2						
Mean	77.11	58.14	68.95	61.18	55.40	190.63
Std	8.4057	1.6327	3.1358	3.5047	0.8694	63.79

Table 17
The 10 features detected for each cluster with the CNS gene expression dataset.

Algorithm	Cluster	Sequence number of the detected 10 features of each cluster									
ESSC	1	5258	5374	6507	95	768	1497	2517	6536	2114	7028
	2	6353	4480	3275	427	3916	5604	6925	2040	3247	923
EWKM	1	1808	5981	5614	217	130	3874	4545	5269	6248	1808
	2	7023	869	5476	6353	1065	4008	6683	2087	1049	7023
LAC1	1	4503	869	1810	1065	6353	6994	2087	3876	1049	5476
	2	1808	7023	6248	2372	5110	6905	1861	6345	3261	4245
LAC2	1	4008	6248	4503	4606	7023	869	1065	1809	6353	6345
	2	6396	2372	1808	2388	6905	5269	3261	1810	1861	3280
FSC	1	4008	869	4503	7023	3125	6353	1810	1861	1065	5476
	2	1808	6248	6905	2372	3280	1787	4021	6136	1954	3045

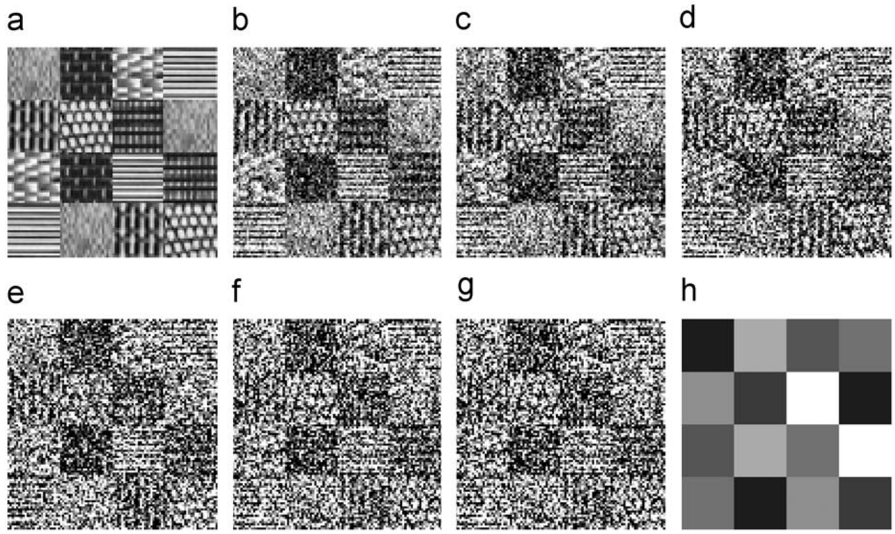

Fig. 5. Texture images: (a) original image; (b)–(g) noisy texture images with different extent Gaussian noise added (σ=0.05, 0.10, 0.15, 0.20, 0.25, 0.30, respectively); (h) ideal segmentation result.

Table 19
The best clustering results obtained for the texture image datasets with RI as metric.

Noise	ESSC	EWKM	LAC1	LAC2	FSC	PROCLUS
σ=0						
Mean	**0.9130**	0.9020	0.9102	0.9077	0.8853	0.84805
Std	0.0164	0.0117	0.0152	0.0150	0.0058	0.0148
σ=0.05						
Mean	**0.9003**	0.8920	0.8858	0.8879	0.8820	0.7884
Std	0.0028	0.0114	0.0108	0.0110	0.0163	0.0139
σ=0.10						
Mean	**0.8774**	0.8657	0.8645	0.8658	0.8552	0.7718
Std	0.0105	0.0123	0.0112	0.0073	0.0100	0.0164
σ=0.15						
Mean	**0.8409**	0.8351	0.8345	0.8351	0.8344	0.7662
Std	0.0040	0.0081	0.0064	0.0049	0.0057	0.0074
σ=0.20						
Mean	**0.8318**	0.8285	0.8288	0.8290	0.8236	0.7542
Std	0.0087	0.0039	0.0060	0.0070	0.0017	0.0120
σ=0.25						
Mean	**0.8264**	0.8185	0.8132	0.8178	0.8156	0.7502
Std	0.0142	0.0090	0.0070	0.0085	0.0071	0.0101
σ=0.30						
Mean	**0.8120**	0.8094	0.8025	0.8085	0.8008	0.7614
Std	0.0041	0.0077	0.0039	0.0052	0.0055	0.0100
Average						
Mean	**0.8574**	0.8502	0.8485	0.8503	0.8424	0.7772
Std	0.0087	0.0092	0.0086	0.0084	0.0074	0.0121

Table 20
The best clustering results obtained for texture image datasets with NMI as metric.

Noise	ESSC	EWKM	LAC1	LAC2	FSC	PROCLUS
σ=0						
Mean	**0.6915**	0.6693	0.6854	0.6788	0.6093	0.5008
Std	0.0314	0.0269	0.0315	0.0329	0.0177	0.0439
σ=0.05						
Mean	0.6366	0.6264	**0.6372**	0.6304	0.6137	0.2829
Std	0.0157	0.0312	0.0166	0.0187	0.0207	0.0664
σ=0.10						
Mean	**0.5339**	0.5288	0.5099	0.5142	0.4843	0.1881
Std	0.0300	0.0356	0.0328	0.0231	0.0409	0.0423
σ=0.15						
Mean	**0.4269**	0.4161	0.4188	0.4139	0.4047	0.1342
Std	0.0058	0.0279	0.0144	0.0190	0.0177	0.0276
σ=0.20						
Mean	**0.3893**	0.3779	0.3809	0.3824	0.3796	0.1043
Std	0.0253	0.0125	0.0161	0.0222	0.0106	0.0203
σ=0.25						
Mean	**0.3565**	0.3535	0.3445	0.3460	0.3448	0.0929
Std	0.0352	0.0309	0.0285	0.0236	0.0190	0.0287
σ=0.30						
Mean	**0.3493**	0.3436	0.3425	0.3403	0.3413	0.1315
Std	0.0080	0.0205	0.0240	0.0140	0.0200	0.0177
Average						
Mean	**0.4834**	0.4737	0.4742	0.4723	0.4540	0.2050
Std	0.0216	0.0265	0.0234	0.0219	0.0209	0.0353

within-cluster compactness and the between-cluster separation is proposed based on the entropy weighting subspace clustering objective function; (2) the enhanced soft subspace clustering algorithm ESSC is developed and the properties are investigated; and (3) comprehensive experiments are carried out to evaluate the performance of the ESSC algorithm. The findings in this study demonstrate that the proposed ESSC algorithm is in general more effective in subspace clustering than the existing algorithms.

This study will be further extended to improve the performance of the existing subspace clustering algorithms by making use of between-class information. For example, enhanced versions

Table 21
Running time (seconds) of algorithms performed on the texture image datasets.

Noise	ESSC	EWKM	LAC1	LAC2	FSC	PROCLUS
σ=0						
Mean	11.69	10.36	8.65	9.72	9.44	13.10
Std	0.3800	0.2021	0.1167	0.1638	0.2208	1.8675
σ=0.25						
Mean	12.07	10.37	8.57	9.65	9.48	13.61
Std	0.4203	0.5632	0.1240	0.2102	0.3811	1.7725

of the existing fuzzy weighting subspace clustering algorithms can be developed. In addition, a theoretical study on the parameter setting of the ESSC algorithm will be conducted, which will be of great importance in providing useful and convenient guiding principles for the ESSC algorithm to be applied to real-world applications.

Acknowledgement

The authors would like to thank the reviewers for their valuable comments which have greatly improved the quality of our manuscript in many ways. This work was supported in part by the Research Grants Council of the HKSAR (nos. PolyU 5145/05E and PolyU 5147/06E) and the National Natural Science Foundation of China (nos. 60773206, 60903100, 60975027 and 90820002) and the Natural Science Foundation of Jiangsu province (no. BK2009067).

Appendix A. Proof of Theorem 1

Given that $\mathbf{V}=[\mathbf{v}_1,\ldots,\mathbf{v}_C]$ and $\mathbf{W}=[\mathbf{w}_1,\ldots,\mathbf{w}_C]$ are fixed, by introducing the Lagrangian multipliers, Eq. (15) is equivalent to the following optimization problem:

$$\Phi_1(\mathbf{U},\boldsymbol{\lambda}^u)=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}^m\sum_{k=1}^{D}w_{ik}((x_{jk}-v_{ik})^2-\eta(v_{ik}-v_{0k})^2) -\sum_{j=1}^{N}\lambda_j^u\left(\sum_{i=1}^{C}u_{ij}-1\right) \tag{A.1}$$

where the superscript u of $\boldsymbol{\lambda}^u$ indicates that it is the Lagrangian multiplier vector with respect to the fuzzy partition matrix $\mathbf{U}=[\mathbf{u}_1,\ldots,\mathbf{u}_N]$. From Eq. (A.1), we can compute the optimal value of u_{ij} by setting $\partial\Phi_1/\partial u_{ij}=0$ and $\partial\Phi_1/\partial\lambda_j^u=0$. Then we obtain

$$\frac{\partial\Phi_1}{\partial u_{ij}}=m\cdot u_{ij}^{m-1}\cdot\sum_{k=1}^{D}w_{ik}((x_{jk}-v_{ik})^2-\eta(v_{ik}-v_{0k})^2)-\lambda_j^u=0 \tag{A.2}$$

$$\frac{\partial\Phi_1}{\partial\lambda_j^u}=\sum_{i=1}^{C}u_{ij}-1=0 \tag{A.3}$$

From Eqs. (A.2) and (A.3), we obtain

$$u_{ij}=\frac{(d_{ij})^{-1/m-1}}{\sum_{i'=1}^{C}(d_{i'j})^{-1/m-1}} \tag{A.4}$$

with

$$d_{ij}=\sum_{k=1}^{D}w_{ik}(x_{jk}-v_{ik})^2-\eta\sum_{k=1}^{D}w_{ik}(v_{ik}-v_{0k})^2$$

This shows that Eq. (A.4) is the necessary conditions for Eq. (15) to reach its minimum when $\mathbf{V}=[\mathbf{v}_1,\ldots,\mathbf{v}_C]$, $\mathbf{W}=[\mathbf{w}_1,\ldots,\mathbf{w}_C]$ are fixed.

Appendix B. Proof of Theorem 2

Given that $\mathbf{U}=[\mathbf{u}_1,\ldots,\mathbf{u}_N]$ and $\mathbf{W}=[\mathbf{w}_1,\ldots,\mathbf{w}_C]$ are fixed, Eq. (15) is equivalent to the following optimization problem:

$$\begin{aligned}\Phi_2(\mathbf{V})&=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}^m\sum_{k=1}^{D}w_{ik}(x_{jk}-v_{ik})^2\\&\quad-\eta\sum_{i=1}^{C}\left(\sum_{j=1}^{N}u_{ij}^m\right)\sum_{k=1}^{D}w_{ik}(v_{ik}-v_{0k})^2\\&=\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}\sum_{j=1}^{N}u_{ij}^m((x_{jk}-v_{ik})^2-\eta(v_{ik}-v_{0k})^2)\end{aligned} \tag{B.1}$$

From Eq. (B.1), we can compute the optimal value of v_{ik} by setting $\partial\Phi_2/\partial v_{ik}=0$. Then we obtain

$$\begin{aligned}\frac{\partial\Phi_2}{\partial v_{ik}}&=w_{ik}\sum_{j=1}^{N}u_{ij}^m(-(x_{jk}-v_{ik})-\eta(v_{ik}-v_{0k}))\\&=-w_{ik}\sum_{j=1}^{N}u_{ij}^m(x_{jk}-\eta v_{0k}-(1-\eta)v_{ik})=0\end{aligned} \tag{B.2}$$

From Eq. (B.2), we obtain

$$v_{ik}=\frac{\sum_{j=1}^{N}u_{ij}(x_{jk}-\eta v_{0k})}{\sum_{j=1}^{N}u_{ij}(1-\eta)} \tag{B.3}$$

This shows that Eq. (B.3) is a necessary condition for Eq. (15) to reach its minimum when $\mathbf{U}=[\mathbf{u}_1,\ldots,\mathbf{u}_N]$, $\mathbf{W}=[\mathbf{w}_1,\ldots,\mathbf{w}_C]$ are fixed. Meanwhile, when $\mathbf{U}=[\mathbf{u}_1,\ldots,\mathbf{u}_N]$, $\mathbf{W}=[\mathbf{w}_1,\ldots,\mathbf{w}_C]$ are fixed, we can obtain the Hessian matrix of Eq. (B.1) with respect to $\mathbf{V}$ as follows.

$$\mathbf{H}^V=\left[\frac{\partial^2 J}{\partial v_{ik}\partial v_{i'k'}}\right]_{CD\times CD}=\begin{cases}(1-\eta)w_{ik}\sum_{j=1}^{N}u_{ij}^m & \text{if } i=i'\ k=k'\\ 0 & \text{otherwise}\end{cases} \tag{B.4}$$

Since $0\le\eta<1$, $w_{ik}>0$, $u_{ij}>0$ and $m>1$, $\mathbf{H}^V$ is positive semi-definite. Thus, according to the optimization theory, we infer that Eq. (B.3) is also the sufficient conditions for Eq. (15) to reach its minimum when $\mathbf{U}=[\mathbf{u}_1,\ldots,\mathbf{u}_N]$, $\mathbf{W}=[\mathbf{w}_1,\ldots,\mathbf{w}_C]$ are fixed.

Appendix C. Proof of Theorem 3

Given that $\mathbf{U}=[\mathbf{u}_1,\ldots,\mathbf{u}_N]$ and $\mathbf{V}=[\mathbf{v}_1,\ldots,\mathbf{v}_C]$ are fixed, by introducing the Lagrangian multipliers, Eq. (15) is equivalent to the following optimization problem:

$$\begin{aligned}\Phi_3(\mathbf{W},\boldsymbol{\lambda}^w)&=\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}\sum_{j=1}^{N}u_{ij}^m((x_{jk}-v_{ik})^2-\eta(v_{ik}-v_{0k})^2)\\&\quad+\gamma\sum_{i=1}^{C}\sum_{k=1}^{D}w_{ik}\log(w_{ik})-\sum_{i=1}^{C}\lambda_i^w\left(\sum_{k=1}^{D}w_{ik}-1\right)\end{aligned} \tag{C.1}$$

where the superscript w of $\boldsymbol{\lambda}^w$ indicates that it is the Lagrangian multiplier vector corresponding to the weighting matrix $\mathbf{W}=[\mathbf{w}_1,\ldots,\mathbf{w}_c]$. From Eq. (C.1), we can compute the optimal value of w_{ik} by setting $\partial\Phi_3/\partial w_{ik}=0$ and $\partial\Phi_3/\partial\lambda_i^w=0$. Then we obtain

$$\frac{\partial\Phi_3}{\partial w_{ik}}=\sum_{j=1}^{N}u_{ij}^m((x_{jk}-v_{ik})^2-\eta(v_{ik}-v_{0k})^2)+\gamma\log(w_{ik})+\gamma-\lambda_i^w=0 \tag{C.2}$$

$$\frac{\partial\Phi_3}{\partial\lambda_i^w}=\sum_{k=1}^{D}w_{ik}-1=0 \tag{C.3}$$

From Eqs. (C.2) and (C.3), we obtain

$$w_{ik}=\exp\left(-\frac{\sigma_{ik}}{\gamma}\right)\Big/\sum_{k'=1}^{D}\exp\left(-\frac{\sigma_{ik'}}{\gamma}\right) \tag{C.4}$$

with

$$\sigma_{ik}=\sum_{j=1}^{N}u_{ij}^m((x_{jk}-v_{ik})^2-\eta(v_{ik}-v_{0k})^2)$$

Thus, Eq. (C.4) is the necessary conditions for Eq. (15) to get its minimum when $\mathbf{U}=[\mathbf{u}_1,\ldots,\mathbf{u}_N]$, $\mathbf{V}=[\mathbf{v}_1,\ldots,\mathbf{v}_C]$ are fixed.

Appendix D. The procedure of the WBPA clustering ensemble for the ESSC algorithm

The procedure for implementing the WBPA clustering ensemble for the ESSC algorithm is described as follows.

(1) Input N data points, cluster number C and index of fuzzy partition m.
(2) Run ESSC algorithm M times with different values of η and γ. Obtain the M partitions: $\mathbf{U}=[u_{ij}]^{(l)}_{C\times N}$, $\mathbf{W}^{(l)}=[\mathbf{w}_1^{(l)},\ldots,\mathbf{w}_C^{(l)}]$ and $\mathbf{V}^{(l)}=[\mathbf{v}_1^{(l)},\ldots,\mathbf{v}_C^{(l)}]$ for $l=1,\ldots,M$.
(3) For each partition $l=1,\ldots,M$
  (a) Compute the posterior probability $p(L_i^{(l)}|\mathbf{x}_j)=u_{ij}^{(l)}$, where L_i $(i=1,2,\ldots,C)$ are the cluster labels.
  (b) Set $\mathbf{P}_j^{(l)}=(p(L_1^{(l)}|\mathbf{x}_j),p(L_2^{(l)}|\mathbf{x}_j),\ldots,p(L_C^{(l)}|\mathbf{x}_j))^T$.
(4) Construct the matrix $\mathbf{A}$ as follows:

$$\mathbf{A}=\begin{bmatrix}(\mathbf{P}_1^{(1)})^T & (\mathbf{P}_1^{(2)})^T & \ldots & (\mathbf{P}_1^{(M)})^T\\ (\mathbf{P}_2^{(1)})^T & (\mathbf{P}_2^{(2)})^T & \ldots & (\mathbf{P}_2^{(M)})^T\\ \vdots & \vdots & & \vdots\\ (\mathbf{P}_N^{(1)})^T & (\mathbf{P}_N^{(2)})^T & \ldots & (\mathbf{P}_N^{(M)})^T\end{bmatrix}_{N\times C\cdot M}.$$

(5) Construct the bipartite graph $G=(V,E)$ where $V=V^C\cup V^I$ is the set of vertices, $V^I=\{\mathbf{x}_j\}$ $(I|=N$, I is the set containing data points), $C=\cup_{l=1}^M V_l^C$, $I^C=\{\mathbf{v}_1^{(l)},\mathbf{v}_2^{(l)}\cdots\mathbf{v}_C^{(l)}\}$ $(C|=CM$, C is the set containing $C\times M$ cluster centers obtained by running ESSC clustering M times), and the connection matrix of the graph is denoted as

$$E=\begin{bmatrix}\mathbf{0} & \mathbf{A}^T\\ \mathbf{A} & \mathbf{0}\end{bmatrix}_{(CM+N)\times(CM+N)}.$$

(6) Run METIS [45] or spectral clustering [46] on the resulting graph G.
(7) Output the resulting C-way partition of N vertices in V^I.

References

[1] A. Jain, M. Murty, P. Flynn, Data clustering: a review, ACM Comput. Surv. 31 (1999) 264–323.
[2] Y. Cao, J. Wu, Projective ART for clustering data sets in high dimensional spaces, Neural Networks 15 (2002) 105–120.
[3] A. Hotho, A. Maedche, S. Staab, Ontology-based text document clustering, in: Proceedings of the IJCAI 2001 Workshop on Text Learning: Beyond Supervision, 2001.
[4] L. Parsons, E. Haque, H. Liu, Subspace clustering for high dimensional data: a review, SIGKDD Explorations 6 (1) (2004) 90–105.
[5] R. Agrawal, J. Gehrke, D. Gunopulos, P. Raghavan, Automatic subspace clustering of high dimensional data for data mining applications, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, 1998, pp. 94–105.
[6] C.H. Cheng, A.W. Fu, Y. Zhang, Entropy-Based Subspace Clustering for Mining Numerical Data, in: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge and Data Mining, 1999, pp. 84–93.
[7] S. Goil, H. Nagesh, A. Choudhary, Mafia: efficient and scalable subspace clustering for very large data sets, Technical Report CPDC-TR-9906-010, Northwest University, 1999.
[8] C. Aggarwal, C. Procopiuc, J.L. Wolf, P.S. Yu, J.S. Park, Fast algorithms for projected clustering, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, 1999, pp. 61–72.
[9] C.C. Aggarwal, P.S. Yu, Finding generalized projected clusters in high dimensional spaces, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, 2000, pp. 70–81.
[10] K.G. Woo, J.H. Lee, Findit: A fast and intelligent subspace clustering algorithm using dimension voting, Ph.D. Dissertation, Korea Advanced Institute of Science and Technology, 2002.
[11] C.M. Procopiuc, M. Jones, P.K. Agarwal, T.M. Murali, A Monte Carlo algorithm for fast projective clustering, in: Proceedings of the ACM SIGMOD Conference on Management of Data, 2002, pp. 418–427.
[12] J. Yang, W. Wang, H. Wang, P. Yu, δ-clusters: Capturing subspace correlation in a large data set, in: Proceedings of the 18th International Conference on Data Engineering, 2002, pp. 517–528.
[13] K.Y. Yip, D.W. Cheung, M.K. Ng, A practical projected clustering algorithm, IEEE Trans. Knowl. Data Eng. 16 (11) (2004) 1387–1397.
[14] K. Chakrabarti, S. Mehrotra, Local dimensionality reduction: A new approach to indexing high dimensional spaces, in: Proceedings of the 26th International Conference on Very Large Data Bases, 2000, pp. 89–100.
[15] G. De Soete, Optimal variable weighting for ultrametric and additive tree clustering, Qual. Quantity 20 (1986) 169–180.
[16] G. De Soete, OVWTRE: a program for optimal variable weighting for ultrametric and additive tree fitting, J. Classification 5 (1988) 101–104.
[17] V. Makarenkov, P. Legendre, Optimal variable weighting for ultrametric and additive trees and k-means partitioning: methods and software, J. Classification 18 (2001) 245–271.
[18] J.Z. Huang, M.K. Ng, H. Rong, Z. Li, Automated variable weighting in k-means type clustering, IEEE Trans. Pattern Anal. Mach. Intell. 27 (5) (2005) 657–668.
[19] Y. Chan, W. Ching, M.K. Ng, J.Z. Huang, An optimization algorithm for clustering using weighted dissimilarity measures, Pattern Recognition 37 (5) (2004) 943–952.
[20] H. Frigui, O. Nasraoui, Unsupervised learning of prototypes and attribute weights, Pattern Recognition 37 (3) (2004) 567–581.
[21] H. Frigui, O. Nasraoui, Simultaneous Clustering and Dynamic Keyword Weighting for Text Documents, in: M. Berry (Ed.), Survey of Text Mining, Springer, Berlin, 2004, pp. 45–70.
[22] C. Domeniconi, D. Papadopoulos, D. Gunopulos, S. Ma, Subspace clustering of high dimensional data, in: Proceedings of the SIAM International Conference on Data Mining, 2004.
[23] J.H. Friedman, J.J. Meulman, Clustering objects on subsets of attributes, J. R. Stat. Soc. B 66 (4) (2004) 815–849.
[24] L. Jing, M.K. Ng, J. Xu, J.Z. Huang, Subspace clustering of text documents with feature weighting k-means algorithm, in: Proceedings of the Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2005, pp. 802–812.
[25] G.J. Gan, J.H. Wu, Z.J. Yang, A fuzzy subspace algorithm for clustering high dimensional data, in: X. Li, O. Zaiane, Z. Li (Eds.), Lecture Notes in Artificial Intelligence, vol. 4093, Springer, Berlin, 2006, pp. 271–278.
[26] G.J. Gan, J.H. Wu, A convergence theorem for the fuzzy subspace clustering (FSC) algorithm, Pattern Recognition 41 (2008) 1939–1947.
[27] L.P. Jing, M.K. Ng, Z.X. Huang, An entropy weighting k-means algorithm for subspace clustering of high-dimensional sparse data, IEEE Trans. Knowl. Data Eng. 19 (8) (2007) 1026–1041.
[28] C. Domeniconi, D. Gunopulos, S. Ma, B. Yan, M. Al-Razgan, D. Papadopoulos, Locally adaptive metrics for clustering high dimensional data, Data Min. Knowl. Discovery J. 14 (2007) 63–97.
[29] K.L. Wu, J. Yu, M.S. Yang, A novel fuzzy clustering algorithm based on a fuzzy scatter matrix with optimality tests, Pattern Recognition Lett. 26 (5) (2005) 639–652.
[30] J.M. Leski, Towards a robust clustering, Fuzzy Sets Syst. 137 (2) (2003) 191–196.
[31] S.T. Wang, F.L. Chung, Z.H. Deng, D.W. Hu, X. Wu, Robust maximum entropy clustering algorithm with its labeling for outliers, Soft Comput. 10 (7) (2006) 555–563.
[32] V. Vapnik, Statistical Learning Theory, Wiley, NewYork, 1998.
[33] J. Yu, M.S. Yang, Optimality test for generalized FCM and its application to parameter selection, IEEE Trans. Fuzzy Syst. 13 (1) (2005) 164–176.
[34] J. Yu, Q.S. Cheng, H.K. Huang, Analysis of the weighting exponent in the FCM, IEEE Trans. Syst., Man, Cybern.—Part B: Cybern. 34 (1) (2004) 164–176.
[35] C. Domeniconi, M. Al-Razgan, Weighted cluster ensembles: methods and analysis, Technical Report ISE-TR-07-06, December 2007.
[36] J. Liu, J. Mohammed, J. Carter, et al., Distance-based clustering of CGH data, Bioinformatics 22 (16) (2006) 1971–1978.
[37] D.J. Newman, S. Hettich, C.L. Blake, et al., UCI repository of machine learning databases, Department of Information and Computer Science, University of California, Irvine, 〈http://archive.ics.uci.edu/ml/〉.
[38] S.L. Pomeroy, et al., Prediction of central nervous system embryonal tumour outcome based on gene expression, Nature 415 (2002) 436–442.
[39] J.B. Welsh, Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer, Cancer Res. 61 (2001) 5974–5978.
[40] C.M. Perou, et al., Molecular portraits of human breast tumours, Nature 406 (2000) 747–752.
[41] A.A. Alizadeh, et al., Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, Nature 403 (2000) 503–511.
[42] A. Bhattacharjee, et al., Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses, Proc. Nat. Acad. Sci. USA 98 (2001) 13790–13795.
[43] T. Randen, Brodatz Texture, 〈http://www.ux.uis.no/~tranden/brodatz.html〉.
[44] V. Kyrki, J.K. Kamarainen, H. Kälviäinen, Simple Gabor feature space for invariant object recognition, Pattern Recognition Lett. 25 (3) (2004) 311–318.

[45] G. Karypis, V. Kumar, A fast and high quality multilevel scheme for partitioning irregular graphs, SIAM J. Sci. Comput. 20 (1) (1998) 359–392.
[46] A.Y. Ng, M. Jordan, Y. Weiss, On spectral clustering: analysis and an algorithm, in: Proceedings of NIPS, 2001, pp. 849–856.
[47] S. Das, A. Abraham, A. Konar, Automatic clustering using an improved differential evolution algorithm, IEEE Trans. Syst., Man Cybern., Part A 38 (1) (2008) 218–237.
[48] J.C. Bezdek, R.J. Hathaway, M.J. Sabin, W. Tucker, Convergence theory for fuzzy c-means: counter-examples and repairs, IEEE Trans. Syst., Man, Cybern. 17 (5) (1987) 873–877.
[49] E.P. Xing, A.Y. Ng, M.I. Jordan, S.J. Russell, Distance metric learning with application to clustering with side-information, in: Proceedings of NIPS, 2003, pp. 505–512.

About the Author—ZHAOHONG DENG received Ph.D. degree from Jiangnan University in 2008. He is currently an assistant professor in the School of Information Technology, Jiangnan University. His research interests include pattern recognition, neurofuzzy systems and their applications. He has published about 20 research papers in international/national journals.

About the Author—KUP-SZE CHOI received his Ph.D. degree in Computer Science and Engineering from the Chinese University of Hong Kong in 2004. He is currently an assistant professor with the School of Nursing and a member of the Centre for Integrative Digital Health, both at the Hong Kong Polytechnic University. His current research interests include computation intelligence, physics-based modeling, computer-assisted surgery and virtual-reality applications in medicine.

About the Author—FU-LAI CHUNG received the B.Sc. degree from the University of Manitoba, Canada, in 1987, and the M.Phil. and Ph.D. degrees from the Chinese University of Hong Kong, in 1991 and 1995, respectively. He joined the Department of Computing, Hong Kong Polytechnic University, in 1994, where he is currently an associate professor. He has published widely in the areas of fuzzy systems, neural networks and pattern recognition. His current research interests include fuzzy data mining, fuzzy neural network modeling and fuzzy techniques for multimedia applications.

About the Author—SHITONG WANG received the M.S. degree in computer science from Nanjing University of Aeronautics and Astronautics, China, in 1987. He visited London University and Bristol University in UK, Hiroshima International University in Japan, Hong Kong University of Science and Technology, Hong Kong Polytechnic University and Hong Kong City University in Hong Kong, as a Research Scientist, for over four years. Since 2000, he is a full professor in the School of Information of Southern Yangtse University, China. His research interests include AI, neuron-fuzzy systems, pattern recognition, and image processing. He has published about 80 papers in international/national journals and has authored seven books.