A pareto ensemble based spectral clustering framework

Juanjuan Luo¹ · Huadong Ma¹ · Dongqing Zhou²

Received: 4 July 2020 / Accepted: 3 October 2020 / Published online: 2 November 2020
© The Author(s) 2020

Abstract

Similarity matrix has a significant effect on the performance of the spectral clustering, and how to determine the neighborhood in the similarity matrix effectively is one of its main difficulties. In this paper, a “divide and conquer” strategy is proposed to model the similarity matrix construction task by adopting Multiobjective evolutionary algorithm (MOEA). The whole procedure is divided into two phases, phase I aims to determine the nonzero entries of the similarity matrix, and Phase II aims to determine the value of the nonzero entries of the similarity matrix. In phase I, the main contribution is that we model the task as a biobjective dynamic optimization problem, which optimizes the diversity and the similarity at the same time. It makes each individual determine one nonzero entry for each sample, and the encoding length decreases to $O(N)$ in contrast with the non-ensemble multiobjective spectral clustering. In addition, a specific initialization operator and diversity preservation strategy are proposed during this phase. In phase II, three ensemble strategies are designed to determine the value of the nonzero value of the similarity matrix. Furthermore, this Pareto ensemble framework is extended to semi-supervised clustering by transforming the semi-supervised information to constraints. In contrast with the previous multiobjective evolutionary-based spectral clustering algorithms, the proposed Pareto ensemble-based framework makes a balance between time cost and the clustering accuracy, which is demonstrated in the experiments section.

Keywords Multiobjective evolutionary algorithm · Spectral clustering · Similarity matrix · Pareto ensemble

Introduction

Spectral clustering has become one of the most popular researching fields in the last decades, it has shown impressive results in the practical applications [19,20,28]. The main idea of spectral clustering is to cluster the data into different groups by using the spectrum of the similarity matrix which captures the structure of the data [33]. Hence, the first step is to construct a similarity matrix, which plays an important role in its performance. Therefore, in this paper, we introduce multiobjective evolutionary algorithm into spectral clustering to construct the similarity matrix, and propose a Pareto ensemble-based spectral clustering framework (PESC).

Conventional similarity matrix construction

In recent years, many methods, including traditional methods [44] and other methods [2,20,29,48], have been proposed to build an appropriate similarity matrix. The similarity matrix aims to model the relationship of each sample of a given dataset. Consider a dataset $A = \{a_1, \dots, a_i, \dots, a_N\}$ with N samples, $S \in \mathbb{R}^{N \times N}$ is the similarity matrix, where s_{ij} measures the similarity between sample a_i and a_j . Take the neighborhood of a sample a_i in S into consideration, the traditional methods can be summarized as follows [20]:

1. Fully connected similarity matrix [12,44]. This similarity construction method takes each sample as the neighbor of the other samples, and a nonzero similarity value is assigned to s_{ij} , resulting a full matrix.
2. k -nearest neighbor (kNN) similarity matrix [4,21]. For a sample a_i , all the similarity entries s_{ij} ($j = 1, \dots, N$) are set to 0 except its nearest k neighbors. But the similarity matrix constructed by this method is not symmetric, there are two strategies to solve it. The first one is to set $s_{ij} = s_{ji} \neq 0$ when sample a_i belongs to sample

✉ Juanjuan Luo
ljjinxd@163.com

¹ Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunication, Beijing, China

² Northern Institute of Electronic Equipment of China, Beijing, China

- a_j 's k -nearest neighbors or sample a_j belongs to sample a_i 's k -nearest neighbors. Another way is mutual k -nearest neighbor similarity construction method, it establishes $s_{ij} = s_{ji} \neq 0$ when both sample a_i and sample a_j belong to each other's k -nearest neighbors.
3. ϵ -neighborhood similarity matrix [12,44]. If the distance between the sample a_i and a_j is less than a threshold ϵ , then $s_{ij} = 1$, otherwise $s_{ij} = 0$.

In these methods, the Gaussian Kernel function $K(a_i, a_j) = \exp(-\frac{\text{dist}_{ij}^2}{2\sigma^2})$ is a typical measurement to calculate the value of nonzero similarity entry s_{ij} . In this Kernel function, dist_{ij} represents the distance between sample a_i and a_j (usually the Euclidean distance is adopted), σ is the scale factor to control the width of the neighborhood which plays an important part in inducing meaningful neighborhood structure. Both k -nearest neighbor similarity matrix and ϵ -neighborhood similarity matrix are sparse matrices, resulting that they have a strict demand for the value of parameter k or ϵ . But it is a difficult task to determine the value of these parameters. In recent years, there are some approaches that learn a similarity graph with exactly K (the number of clusters) connected components for graph-based clustering, such as [30,31]. To get this K connected similarity graph, [31] proposed a method that constructing an initial graph with k neighbors and calculating the value of these nonzero entries with a parameter-free approach, however, the value of k still needs to setup in this literature. In both [30,31], a common question is that they should solve the optimization problem with the following form: $\min_x f(x) + \gamma g(x)$, which γ is a parameter sensitive to a different dataset.

The multiobjective evolutionary algorithm based clustering

In recent years, multiobjective evolutionary algorithms (MOEAs) have attracted a lot of researchers' attention for their wide application in real world [5,7]. Many state-of-the-art MOEAs, such as NSGA-II [8], SPEA2 [53], IMOEa [41], and MOEA/D [15,51], have been proposed to handle multiobjective optimization problems (MOPs). They also have been used in the field of data mining successfully, such as clustering [10,18], classification [11,25] and feature selection [27,45]. MOEA-based clustering algorithms focus on using multiple criterions, such as the cluster validity indexes, to capture the characteristics of the data [16,26]. They vary from different aspects, including the type of MOEAs, the encoding schemes, the objective functions, the final Pareto optimal solution selection strategies, and even the evolutionary operators. A detailed survey on this issue has been discussed in the literature [26]. To better understand the proposed framework PESC, we will present some related work in the aspect

of the encoding schemes, the objective functions, and the final Pareto optimal solution selection strategies.

Considering the adopted encoding schemes, an experimental evaluation of cluster representations (prototype-based representation, label-based representation, and graph-based representation) for multi-objective evolutionary clustering has been done in [9]. The advantage of prototype-based representation is that the encoding length is small so that it takes less time to apply the evolutionary operators, nevertheless, this representation scheme is good at dealing with round-shaped data. On the contrary, label-based representation and graph-based representation are not strict to the underlying structure of the clusters, but their coding length is high especially for large-scale dataset.

The cluster validity indexes are usually adopted as the objective functions in MOEAs. The cluster validity indexes, such as overall cluster deviation [14], cluster connectedness [14], Jm [47], XB [50], silhouette index [37], the intracluster entropy H [35,36], and cluster separation $Sep(C)$ [35,36], are the most commonly used objective functions. We have known that the quality of the clustering quite depends on the objective functions adopted, but there is no available analysis on the effect of different objective function until now.

The final Pareto optimal solution selection strategy can be classified into three categories: the independent objective-based approach, the knee-based approach, and the ensemble-based approach. For the first approach, another cluster validity different from objective functions is used to measure the quality of the nondominated solutions. The one with the best quality on this index is selected as the final solution, such approaches can be seen in [22,23]. The knee-based approach aims to select a solution, which the change of one objective value induces the maximum change in the other one, from the Pareto Front (PF) as the final solution, such as MOCK [13,14], StEMO [17].

In this paper, we focus on the ensemble-based final Pareto optimal solution selection strategy. Ensemble learning has been proved to be effective against solving machine learning problems, especially in the practical application, such as SAR image segmentation [52]. As for ensemble-based clustering, there is no explicit correspondence between the labels delivered by different clustering. Its difficulty lies in finding a consensus partition from multiple algorithms or partitions [42]. In [40], three cluster ensemble algorithms, which are cluster-based similarity partitioning algorithm (CSPA), hypergraph partitioning algorithm (HGPA), and meta-clustering algorithm (MCLA), were introduced to complete the ensemble task. Since MOEA can generate a set of nondominated solutions, it provides a suitable way for ensemble learning. The ensemble-based approach assumes that all the nondominated solutions contain some useful information on detecting the structure of the dataset, therefore, all these solutions should be integrated to obtain a single

clustering output. The three cluster-ensemble approaches CSPA, HGPA, and MCLA can be easily applied to MOEA clustering problems. A few researchers have also proposed some MOEA-based cluster-ensemble algorithms. In [24], a multiobjective genetic algorithm-based approach for fuzzy clustering of categorical data which simultaneously optimizes fuzzy compactness and fuzzy separation of the clusters ($MOGA(\pi, sep)$) is proposed. To obtain the final clustering result, a majority voting strategy is implemented on the non-dominated solutions to select the training samples. In practical application, the combination of knee-based approach and ensemble-based approach for recurrent neural networks is successfully used in the prediction of computational fluid dynamic simulations [39] and image identification [1]. In these literatures, not all the individuals in the Pareto set are considered as suitable solutions, only the Pareto-optimal solutions around the knee-point are employed to implement the ensemble task. Besides that, MOEA is also applied to neural network ensembles successfully in [3] by simply combining all the classifiers obtained from MOEA to form the ensemble. In general, all the works have shown that the ensemble-based methods performs better than simply selecting one Pareto solution.

Motivation and contributions

As we know, the existed MOEA-based clustering algorithms are barely related to spectral clustering, except for CSPA. In our previous work [20], we have proposed a sparse representation based spectral clustering framework via MOEAs (denoted as SRMOSC), this work introduces multiobjective optimization into spectral clustering, and constructs the similarity matrix using a sparse representation approach by modeling spectral clustering as a constrained multiobjective optimization problem. Unfortunately, as we mentioned in that work, its space complexity is high, especially when solving large-scale problems. In addition, as mentioned in the above sections, the conventional similarity matrix construction methods usually suffers from parameter tuning difficulty. Motivated by these two aspects, we tried to tackle the above problems in this paper.

It is generally known that MOEA can generate a set of non-dominated solutions for multiobjective optimization problem (MOP). Hence, if the similarity matrix construction problem can be regarded as a MOP and all the nondominated solution can participate in the construction of the similarity matrix, then this problem can be solved in a parallel way and it will be more time saving than the unparallel way. Inspired by this idea, we propose a Pareto ensemble based spectral clustering framework in this paper, whose main procedure falls into two phases, the main contributions are summarized as follows:

1. A “divide and conquer” strategy is proposed to solve the similarity matrix construction problem. In PESC, the main procedure is divided into two phases, in which phase I aims to detect the nonzero entries of the similarity matrix by using MOEA and phase II turns to determine the value of the nonzero entries of the similarity matrix by the ensemble strategies.
2. In phase I, we introduce dynamic multiobjective optimization to spectral clustering by adopting a similarity measurement and a specific designed diversity measurement as objective functions. In addition, a specific initialization is designed to speed up the convergence in this phase.
3. In phase II, three ensemble approaches are proposed to construct the similarity matrix and determine the value of the nonzero entries.

In contrast to the existed MOEA based spectral clustering, PESC makes a balance between the time cost and the clustering accuracy with the “divide and conquer” strategy. Compared with the conventional similarity matrix construction methods, PESC can automatically determine the neighbors of the similarity matrix. It should be noted that the contribution of this work that differs from other Pareto-ensemble algorithms are that a single estimated Pareto-optimal solution can not construct a similarity matrix in this paper. What’s more, using the Pareto-ensemble based framework to construct the similarity matrix results in a reduction of both the time and space complexity.

To give a clear introduction and analysis of PESC, the paper is structured as follows. The description of PESC is shown in “The description of PESC”, including its two main phases. Then the experiment results and analysis are presented in “Experiments and discussion”. At last, a conclusion remark of PESC is given in “Conclusion”.

The description of PESC

In spectral clustering, the key point is to construct a similarity matrix. Assume $A = \{a_1, a_2, \dots, a_N\}$ is the given dataset

with N samples and K clusters, $S = \begin{pmatrix} s_{11} & \dots & s_{1N} \\ \vdots & \ddots & \vdots \\ s_{N1} & \dots & s_{NN} \end{pmatrix}$ is the

similarity matrix to be constructed, then the main procedure of the spectral clustering is as Algorithm 1.

Algorithm 1 Unnormalized spectral clustering**Input:**Dataset A , number of clusters K .**Begin:**Clustering result $A_1, \dots, A_i, \dots, A_K$ with $A_i = \{a_j | v_j \in C_i\}$.

- 1: **Step1:** Construct the similarity matrix S .
- 2: **Step2:** Compute the unnormalized Laplacian L .
- 3: **Step3:** Compute the first K eigenvectors $\{u_1, u_2, \dots, u_K\}$ of L .
- 4: **Step4:** Construct a matrix $Y \in \mathbb{R}^{N \times K}$ whose column vector are $\{u_1, u_2, \dots, u_K\}$. Let $v_i \in \mathbb{R}^K$ be the row vector of Y .
- 5: **Step5:** Cluster $\{v_1, v_2, \dots, v_N\}$ with k -means into clusters $\{C_1, \dots, C_K\}$.

Output:Clustering result $A_1, \dots, A_i, \dots, A_K$ with $A_i = \{a_j | v_j \in C_i\}$.

PESC aims to construct the similarity matrix S with a Pareto ensemble approach, and we give a detailed description of it in this section. The main procedures (see Algorithm 2) are divided into two phases: phase I—the nonzero entries determination phase (Steps1–2) and phase II–Pareto ensemble based weight determination phase (Step3). Hence, this section will be divided into two sections to discuss each issue.

Algorithm 2 The framework of PESC**Input:**

The population size pop ;
 The maximum iteration gen ;
 crossover probability p_c ;
 mutation probability p_m .

Begin:

- 1: **Step1:** Initialization: Generate the initial population P^0 according to the initialization scheme, and evaluate their objective function values;
- 2: **Step2:** Cycling:
- 3: **for** $i = 1 : gen$ **do**
- 4: Implement the crossover and mutation operators to P^{i-1} and generate the offspring population $N P^{i-1}$;
- 5: Combine $C P^{i-1} = P^{i-1} \cup N P^{i-1}$, and evaluate the objective function values;
- 6: Make a nondominated sorting to $C P^{i-1}$, and select pop individuals to P^i according to the proposed diversity preservation strategy;
- 7: **end for**
- 8: **Step3:** Implement the Pareto optimal solutions ensemble strategy to P^{gen} , and construct the similarity matrix S ;
- 9: Apply **Step2-Step5** in **Algorithm 1** to S .

Output:Clustering result $A_1, \dots, A_i, \dots, A_K$ with $A_i = \{a_j | v_j \in C_i\}$.**Phase I: Nonzero entries determination**

Phase I consists of the initialization and the cycling loop of PESC. We distribute this phase into three sections to describe, including the mathematical description of PESC and the evolutionary operators.

Mathematical description of PESC

The proposed algorithm PESC takes advantage of the superiority of MOEA, which can generate a set of solutions, to find

the possible nonzero entries of the similarity matrix. Assume that each individual can find one possible nonzero entry for each sample, the objective functions can be formulated as formulas (1) and (2):

$$\begin{aligned} \min \quad & \begin{cases} f_1(X) = \frac{1}{N} \sum_{i=1}^N \text{sim}(a_i, a_{x_i}) \\ f_2(X) = 1 - \text{DIV}(X) \end{cases} \\ \text{s.t.} \quad & X = \{x_1, \dots, x_i, \dots, x_N\}, \\ & x_i \neq i, \\ & x_i \in \{1, 2, \dots, N\}. \end{aligned} \quad (1)$$

$$\begin{aligned} \text{DIV}(X_i) &= \frac{1}{M} \sum_{m=1}^M \text{div}(X_i, X_m) \\ \text{div}(X_i, X_m) &= \frac{1}{N} \sum_{j=1}^N \mu(x_j^i, x_j^m) \\ \text{where } \mu(c, d) &= \begin{cases} 1, & c \neq d \\ 0, & c = d \end{cases} \end{aligned} \quad (2)$$

where $X = \{x_1, \dots, x_i, \dots, x_N\}$ is the decision vector to be optimized, M is the number of solutions in the current population. For any individual $X = \{x_1, \dots, x_i, \dots, x_N\}$, the i -th gene $x_i = t$ represents the sample a_t connects with the i -th sample a_i , and we denote it as sample a_{x_i} in the following parts. The objective function $f_1(X)$ measures the average similarity between sample a_i and a_{x_i} , where $i = \{1, 2, \dots, N\}$, and $\text{sim}(a, b)$ represents the similarity between samples a and b . Note that different similarity measurements can be adopted as the objective functions in real application, and we use the Euclidean distance in this paper. $\text{DIV}(X)$ measures the diversity of the solution X , $f_2(X)$ is a dynamic objective function since its value changes with the generation. Note that the constraint $x_i \neq i$ should be satisfied, it means that only the similarity between different samples can be measured.

There have been a variety of literatures focus on the diversity of MOEAs [6,46], it has pointed out in many literatures that the decision space should be taken into consideration [43] when estimate the diversity of solutions. The decision space-based diversity estimation usually significantly outperforms the objective space based strategies especially in solving multiobjective optimization problems, such as [34,38,49]. In this paper, the diversity measurement is not only adopted in the objective functions, but also in the diversity preservation strategy. According to formula (2), $\text{DIV}(X_i)$ measures the diversity of the solution X_i in the decision space regarding the character of the similarity matrix construction problem. In theoretical analysis, $\text{DIV}(X_i) \in [0, 1]$. For a solution X_i , the more it differs

from other solutions, the higher its diversity is after this estimation.

Furthermore, unsupervised clustering without any guidance is usually unreliable [32], so it is necessary to extend the above unsupervised clustering model to semi-supervised learning model. Suppose sample a_i and a_j are labeled samples belonging to different clusters, $a_i \in C_l, a_j \in C_m, l \neq m$. The semi-supervised spectral clustering can be modeled as Eq. (3) by adding some constraints involved by these labeled samples:

$$\begin{aligned} \min \quad & \begin{cases} f_1(X) = \frac{1}{N} \sum_{i=1}^N \text{sim}(a_i, a_{x_i}) \\ f_2(X) = 1 - \text{DIV}(X) \end{cases} \\ \text{s.t.} \quad & X = \{x_1, \dots, x_i, \dots, x_N\}, \\ & x_i \neq i, \\ & x_i \neq j, \\ & x_j \neq i, \\ & x_i \in \{1, 2, \dots, N\}. \end{aligned} \quad (3)$$

where the constraints $x_i \neq j, x_j \neq i$ guarantee that samples with different labels do not connect with each other.

Evolutionary operators

According to the objective functions, we use a graph-based representation encoding scheme in PESC. The coding length is N , and the size of the searching space is $(N-1)^N$. To obtain a set of high-quality initial individuals, a specific initialization scheme is designed (see Algorithm 3). This initialization scheme is based on the assumption that the corresponding entry in the similarity matrix prefers to be a nonzero entry if two samples locate in each other's neighborhood. In each individual $X_i = \{x_1^i, x_2^i, \dots, x_j^i, \dots, x_N^i\}$, the value of x_j^i is decided by the Euclidean distance between sample a_j and other samples. In Algorithm 3, the parameter k controls the width of neighborhood, and its value can affect the convergence speed of PESC, it should be larger than $\log(N)$ according to [44]. The adopted crossover and mutation operators are the expansion of classic SBX (simulated binary crossover) and polynomial mutation, respectively. Both of them have good performance on the validity of PESC, which will be demonstrated in the experiment section.

Phase II: Pareto ensemble based weight determination

When the cycling loop finishes, we can obtain a set of nondominated solutions $\{X_1, \dots, X_m, \dots, X_M\}$ where $X_m = \{x_1^m, x_2^m, \dots, x_N^m\}$. The existed Pareto ensemble methods introduced in "Introduction" are not suited for PESC since

Algorithm 3 Initialization

Input:

Dataset A ;
the population size pop .

Begin:

```
1: Step1: Calculate the Euclidean distance matrix of the given dataset;
2: Step2: Individual initialization;
3: for  $i = 1 : pop$  do
4:   for  $j = 1 : N$  do
5:     Take the Euclidean distance into consideration, find one of the
        $k$ -th ( $k \leq \lceil \sqrt{N} \rceil$ ) nearest neighbor (suppose to be sample  $a_m$ )
       of sample  $a_j$ , and set  $x_j^i = m$ ;
6:   end for
7: end for
Output:
The initial individuals  $\{X_1, \dots, X_{pop}\}$ .
```

that they are based on label representation or prototype representation but not the graph-based representation encoding scheme. Besides, the objective functions adopted in PESC make it inappropriate to use those cluster ensemble schemes. In this section, we focus on the task that how to convert these nondominated solutions into a sparse similarity matrix with the proposed methods.

In general, when we convert $\{X_1, \dots, X_m, \dots, X_M\}$ to a sparse similarity matrix S , the following rule should be satisfied: if $x_j^m = j$, it means s_{ij} is a nonzero entry, otherwise, s_{ij} is a zero entry.

For example, if we have a population of 3 individuals $X_1 = \{2 \ 1 \ 4 \ 3 \ 4\}$, $X_2 = \{2 \ 1 \ 5 \ 3 \ 4\}$ and $X_3 = \{2 \ 3 \ 5 \ 3 \ 3\}$, then the nonzero entries of the converted similarity matrix are $\{s_{12}, s_{21}, s_{23}, s_{34}, s_{35}, s_{43}, s_{53}, s_{54}\}$.

However, the value of each nonzero entry can not be determined by phase I, so we design the following three methods to complete this task.

1. Sparse Representation based single objective optimization (denoted as ES1 in the following experiment section):

$$\begin{aligned} \min \quad & \|AS - A\|_2^2 \\ \text{s.t.} \quad & s_{ij} > 0, \quad \text{if } x_j^m = j, \\ & s_{ij} = 0, \quad \text{else.} \end{aligned} \quad (4)$$

This strategy is based on sparse representation which is inspired by our previous work [20], it tries to minimize the reconstruction error $\|AS - A\|_2^2$, where A is not only the over-complete dictionary, but also the measurement matrix. When we use this strategy to obtain the similarity matrix S , which entries are nonzero values have been decided according to the previous rule, implied by the constraints, so what we need to do in this task is to determine the weight of each nonzero entry. A lot of evolutionary algorithms can be adopted here, such as particle swarm optimization (PSO) and Genetic Algorithm (GA). In our

experiment, we adopt simplified PSO to complete this task since its easy implementation and fast convergence.

2. Diversity ensemble (denoted as ES2 in the following experiment section):

$$s_{ij} = \frac{1}{M} \sum_{m=1}^M (1 - DIV(X_m)) \cdot \delta(x_i^m, j)$$

$$\delta(x_i^m, j) = \begin{cases} 1, & x_i^m = j \\ 0, & x_i^m \neq j \end{cases} \quad (5)$$

where M is the number of nondominated solutions.

This similarity construction strategy is related to the diversity of each nondominated solution. Suppose s_{ij} is a nonzero entry, when we try to determine its value, we should find the nondominated solutions which $x_i^m = j$, and calculate their diversity according to (2) first. Then we take the average diversity value as the value of the nonzero entry s_{ij} . Metaphorically speaking, $\delta(x_i^m, j)$ are in charge of finding the solution which agrees to connect sample a_i with a_j , and the solution which has better diversity contributes more to the final weights of s_{ij} . Note that when we we construct S , only the nondominated solutions are considered.

Take the above X_1 , X_2 and X_3 for example here, the value of $div(X_i, X_m)$ is calculated according to formula (2) and shown in Table 1, the diversity $DIV(X_1)$, $DIV(X_2)$, $DIV(X_3)$ are $\frac{4}{15}$, $\frac{1}{5}$, $\frac{1}{3}$, respectively. Therefore, the value of the nonzero entry s_{12} is decided by individual X_1 , X_2 , and X_3 in accordance with formula (5), and $s_{12} = \frac{1}{3}(DIV(X_1) + DIV(X_2) + DIV(X_3)) = \frac{11}{15}$. In the same way, the similarity matrix can be obtained,

$$\text{and } S = \begin{pmatrix} 0 & \frac{11}{15} & 0 & 0 & 0 \\ \frac{22}{45} & 0 & \frac{11}{45} & 0 & 0 \\ 0 & 0 & 0 & \frac{11}{45} & \frac{22}{45} \\ 0 & 0 & \frac{11}{45} & 0 & 0 \\ 0 & 0 & \frac{22}{45} & \frac{11}{45} & 0 \end{pmatrix}$$

3. Kernel function (denoted as ES3 in the following experiment section):

$$s_{ij} = \exp\left(\frac{-\text{dist}_{ij}^2}{2\sigma^2}\right) \quad (6)$$

where dist_{ij} is the Euclidean distance between sample a_i and a_j , σ is the scale factor to control the width of the neighborhood.

Table 1 The value of $div(X_i, X_m)$

$div(X_i, X_m)$	X_1	X_2	X_3
X_1	0	$\frac{1}{5}$	$\frac{3}{5}$
X_2	$\frac{1}{5}$	0	$\frac{2}{5}$
X_3	$\frac{3}{5}$	$\frac{2}{5}$	0

When we compare the performance of the previous strategies, all of them can get satisfying results, but ES1 is time-consuming in contrast with others, ES3 is easy to implement, but we need to choose a suitable value for the scale factor when dealing with different clustering problems. ES2 can get a balance between time complexity and clustering accuracy. The effect of these three ensemble strategies will be discussed in the experiment section.

According to the Step2 of the spectral clustering in Algorithm 1, the similarity matrix has to transform into a Laplacian matrix L which is symmetric and positive semi-definite. So the following methods mentioned in [20] is needed to complete the transformation:

$$s_{ij} = \max\{s_{ij}, s_{ji}\}$$

$$d_{ij} = \begin{cases} 0, & i \neq j \\ \sum_{m=1}^N s_{im}, & i = j \end{cases} \quad (7)$$

$$L = D - S$$

$D \in \mathbb{R}^{N \times N}$ is a diagonal matrix with the diagonal elements d_{ii} whose values equals to the sum of the i -th column of weight matrix S .

Complexity analysis

1. *Space complexity* Part of the memory in our algorithm is used to store the distance matrix among all the samples, and its space complexity is $O(N^2)$. The rest are mainly used to store the population whose space complexity is $O(pop \cdot N)$.
2. *Time complexity* In this algorithm, the main time cost lies in the working cycle of the MOEA. All the time complexity of initialization, crossover, mutation, and evaluation is $O(pop \cdot N)$. The time complexity of the diversity preservation strategy is $O(n_{div} \cdot N \cdot pop)$, where n_{div} is the number of individuals to implement diversity preservation strategy. The time complexity of the updating of each generation relies on the MOEA adopted. In our experiment, PESC is carried out on the basis of NSGA-II, the time complexity of this step is of $O((2pop)^2)$. The time complexity of Step4 (Algorithm 1) also depends on the first K eigenvectors calculation method adopted. Hence, the total time complexity of PESC is simplified as $\max\{O(pop \cdot gen \cdot N), O(N^2)\}$.

Experiments and discussion

PESC is mainly carried out on the basis of NSGA-II, we show the experiment results and give a discussion on them in this section, including the investigation of parameters, the overall

comparison between PESC and traditional or multiobjective clustering algorithms, and the detailed discussion of PESC.

The investigation of parameters

In the experiment section, 11 UCI supervised classification datasets are adopted to test the efficiency of PESC, the number of clusters is fixed to the real number of classes, the clustering accuracy is measured in terms of the percent of instances that are right classified, and all the experiments are obtained from 30 independent runs. The main character of these datasets are shown in Table 2, from which we can see that the number of samples ranges from 150 to 5000. Therefore, we take ‘wine’, ‘balance’, and ‘waveform’ as examples to make an investigation of parameter p_c , p_m , pop and gen , which are shown in Figs. 1, 2 and 3.

Figure 1 shows the clustering accuracy against the parameter p_c from 0.5 to 0.9 with interval 0.1 on dataset “wine”, “balance” and “waveform”. We can observe that when we choose $p_c = 0.9$, three tested datasets get relatively higher clustering accuracy.

On the effect of parameter pop , the scale of the dataset is a major factor to be considered. When the other parameters are fixed ($p_c = 0.9$, the evaluation time is set to 20,000), the effect of pop to the clustering accuracy with pop ranges from 10 to 100 is shown in Fig. 2. An observation, that the value of pop is related to the scale of the dataset, can be derived from this figure. When we obtain a satisfying clustering result, the larger the scale of the dataset is, usually the

higher the value of parameter pop should be. According to literature [44], the number of the neighbors in kNN similarity construction method k is set to $\log(N)$ (N is the number of samples). In contrast with kNN , pop has the similar effect on controlling the size of the neighborhood, but the accuracy of PESC is insensitive to this parameter when the value is large enough as in the observation. When pop is set to a large enough value, the width of the neighborhood will be determined adaptively in the evolutionary process. According to the theoretical analysis and the experimental result, the value of pop should be larger than $\log(N)$. In our following experiment, its value is set to no less than $\sqrt{N}$ to obtain a satisfying result. Take the scale of the tested datasets into consideration, pop is set to 100 according to our experiment (see Fig. 2).

Figure 3 shows the clustering accuracy with respect to the parameter gen . In this figure, $p_c=0.9$, $pop=100$, gen is set to the maximum value 100, and the clustering accuracy is tested in every 5 generations from 0 to 100. We can see that the PESC has converged on the tested dataset since the 20-th iteration from Fig. 3.

In general, take the stability and time complexity of PESC into consideration, the value of parameters pop , gen , p_c and p_m ($p_m = 1 - p_c$) are set to 100, 100, 0.9, 0.1, respectively.

The effect of evolutionary operators

As described in “Evolutionary operators”, we designed a specific initialization strategy which is biased to the neigh-

Table 2 The character of adopted UCI datasets

Dataset	Wine	Glass	Iris	WDBC	Heart	Balance	Vote	<i>E. coli</i>	Thyroid	Image segmentation	Waveform
Number of attributes	13	9	4	30	13	4	16	7	5	19	21
Number of samples	178	214	150	569	270	625	435	336	215	2310	5000
Number of classes	3	6	3	2	2	3	2	8	3	7	3

Fig. 1 The effect of parameter p_c and p_m ($p_m = 1 - p_c$) on the clustering accuracy over 30 runs

Fig. 2 The effect of parameter *pop* on the clustering accuracy over 30 runs

Fig. 3 The effect of parameter *gen* on the clustering accuracy over 30 runs

neighborhood samples. To test its efficiency, we make a comparison between the proposed initialization strategy and the random initialization (see Fig. 4). In the random initialization strategy, x_j^i (the i -th individual is represented as $X_i = \{x_1^i, x_2^i, \dots, x_j^i, \dots, x_N^i\}$) is set to a random value selected from 1 to N . In Fig. 3, *gen* is set to 300, and the clustering accuracy is tested in every five generations, we can see that the convergence speed of random initialization scheme is low especially for large scale dataset. Analyzing Figs. 2 and 4, we can draw the conclusion that the proposed initialization scheme can generate a set of high-quality solutions and speed up the convergence rate.

The similarity matrix comparison

In spectral clustering, the similarity matrix can reveal the relationship between samples. To evaluate the effect of the similarity matrix obtained from PESC, a visualization of the similarity matrix and the corresponding eigenvalues and eigenvectors are shown in Fig. 5. In this experiment, We take ‘wine’ for illustration since that it is very clear to

see the relationship between the different clusters (‘wine’ has 178 samples, 13 attributes, and three categories, with samples 1–59 belonging to category ‘1’, samples 60–130 belonging to category ‘2’, and the rest are in category ‘3’). In Fig. 5, we compare PESC with the sparse spectral clustering algorithm SRMOSC and the traditional similarity matrix construction methods to discuss their performance. In the similarity matrix, the maximum and minimum weights are represented as the white and black pixels respectively for visualization purposes.

When we compare the three ensemble strategies proposed in PESC, we can see that they got similar performance, but the similarity matrices are quite different. The weights obtained by ES1 are lower than that obtained by ES2 and ES3. However, they have the similar properties that the number of nonzero entries is less than that of zero entries and the nonzero entries are mostly distributed as intra-class connections.

In contrast with SRMOSC, ES1 of PESC obtains similar similarity matrix with it. Both of the similarity matrices are sparse, and the values of the nonzero entries are lower

Fig. 4 The convergence analysis of the random initialization strategy

compared with that obtained from the other similarity matrix construction methods. What's more, the time complexity and space complexity of PESC is less than that of SRMOSC, and the corresponding experiment is shown in the next section.

In contrast with the conventional similarity construction methods (fully connected, kNN , mutual kNN , and ϵ -neighborhood similarity matrix construction methods), the proposed algorithm PESC (row 1–row 3) and kNN (row 6) have better performance, but the rest algorithms cannot extract distinguishable information from the eigenvectors to carry out the clustering task exactly. We also can see that it is not available to use the eigenvectors of fully connected, mutual kNN , and ϵ -neighborhood similarity matrices to classify the samples into exact clusters. Furthermore, it is a difficult and time-consuming task to decide the value of parameters in the conventional similarity construction methods, while PESC can determine the neighborhood of a sample automatically in the evolutionary process.

In general, we can obtain the following conclusions from Fig. 5: (1) the similarity matrices obtained from three proposed ensemble strategies are sparse matrices, and all of them can get satisfying clustering result; (2) the nonzero entries are mainly distributed as intra-class connections if the similarity matrix can provide effective distinguishable information.

Experimental comparison between PESC and other algorithms

In this section, we will carry out an overall comparison between PESC and other algorithms, including the comparison between PESC and multiobjective clustering algorithms (SRMOSC and MOCK), the comparison between PESC and other cluster ensemble strategies or algorithms ($MOGA(\pi, sep)$ and CSPA), and the comparison between PESC and other conventional similarity matrix construction methods. Among these comparison algorithms, $MOGA$

(π, sep), CSPA, MOCK, and SRMOSC are all MOEA based clustering algorithms, MOCK is also regarded as a multiobjective cluster-ensemble algorithm since that its initialization is a hybrid of link-based clustering and k -means. Furthermore, the experiments are also extended to semi-supervised clustering with 10% and 20% labeled samples. All these experiments are tested on 11 UCI datasets over 30 independent runs to demonstrate the effectiveness of PESC.

The parameter setting of the comparison algorithms

The value of pop , gen , p_c , p_m are set to 100, 100, 0.7, 0.3 in SRMOSC, MOCK according to [14,20]. CSPA was carried out on the basis of multiobjective genetic algorithm (MOGA), and share the same objective function with $MOGA(\pi, sep)$. So the parameter setting in these two algorithms are the same, which pop , gen , p_c , p_m are set to 100, 100, 0.8 and 0.2, respectively [24]. When constructing the similarity matrix using fully connected kNN and mutual kNN construction methods, the Gaussian Kernel $K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$ is adopted to calculate the similarity. We carry out the experiment with the following values {0.001, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 15} for σ , and choose the one that outputs the best clustering result as the final value. For kNN and mutual kNN methods, k is set to $\log(N)$. ϵ is selected from the value {0.2, 0.3, 0.4, 0.5, 0.6} with the best clustering result in the ϵ -neighborhood method.

The comparison result of unsupervised clustering algorithms

Tables 3 and 4 show the experiment result obtained from PESC against other MOEA based clustering algorithms and the conventional similarity matrix construction algorithms on real-life datasets with no labeled samples, respectively. The best result in these tables is written in bold. In Table 3,

Fig. 5 Visualization of the similarity matrix (column 1), eigenvalues (column 2), and the eigenvectors (columns 3–5) obtained from PESC against other algorithms on dataset ‘wine’

Table 3 Clustering accuracy comparison obtained from pesc against moea based clustering algorithms including cluster ensemble algorithms on real-life datasets

Datasets	PESC			SRMOSC	MOCK	MOGA(π , sep)	CSPA
	ES 1	ES 2	ES 3				
Wine	94.14 (1.61)	94.38 (1.86)	96.20 (1.73)	95.90 (0.93)	68.65 (3.62)	95.03 (1.34)	91.52 (8.63)
Glass	63.79 (6.37)	65.72 (5.45)	61.82 (8.86)	62.45 (2.23)	44.16 (4.46)	60.65 (2.92)	48.38 (1.81)
Iris	93.73 (1.46)	94.26 (1.79)	95.72 (1.72)	92.50 (2.46)	90.10 (0.73)	91.33 (1.65)	71.06 (6.07)
WDBC	97.11 (0.78)	97.42 (0.73)	97.79 (0.79)	94.37 (0.63)	94.52 (0.12)	93.87 (0.46)	70.21 (4.89)
Heart disease	85.21 (8.30)	87.59 (1.53)	87.88 (3.27)	68.54 (5.60)	80.91 (0.52)	81.61 (4.38)	80.74 (0.59)
Balance	87.08 (0.83)	87.37 (0.98)	87.13 (1.73)	66.79 (1.88)	54.88 (5.23)	75.11 (4.25)	65.07 (7.50)
Vote	93.61 (6.15)	95.21 (0.91)	96.86 (0.80)	88.28 (0.60)	65.08 (0.92)	87.79 (0.75)	82.53 (7.03)
<i>E. coli</i>	84.48 (2.27)	85.42 (1.67)	84.48 (2.49)	80.64 (2.42)	64.06 (0.65)	80.82 (1.08)	77.19 (1.24)
Thyroid	93.48 (0.89)	92.96 (1.03)	92.84 (1.56)	92.84 (1.09)	73.53 (2.77)	87.44 (2.04)	75.18 (1.48)
Image segmentation	78.67 (6.29)	78.14 (5.46)	82.12 (6.44)	70.22 (2.89)	57.84 (4.95)	67.37 (3.67)	55.35 (6.52)
Waveform	92.56 (0.51)	93.17 (0.43)	92.56 (0.42)	63.90 (5.40)	69.21 (2.35)	63.96 (5.07)	52.39 (2.18)

Table 4 Clustering accuracy comparison obtained from pesc against conventional similarity construction-based algorithms on real-life datasets

Datasets	PESC			<i>kNN</i>	Mutual <i>kNN</i>	ϵ -neighborhood	Fully connected
	ES 1	ES 2	ES 3				
Wine	94.14 (1.61)	94.38 (1.86)	96.20 (1.73)	96.07 (0.00)	54.61 (6.38)	59.97 (6.93)	64.38 (0.00)
Glass	63.79 (6.37)	65.72 (5.45)	61.82 (8.86)	56.50 (0.94)	44.01 (3.33)	49.95 (1.33)	47.03 (0.94)
Iris	93.73 (1.46)	94.26 (1.79)	95.72 (1.72)	88.36 (0.00)	53.60 (7.46)	68.00 (0.00)	66.80 (0.00)
WDBC	97.11 (0.78)	97.42 (0.73)	97.79 (0.79)	94.90 (0.00)	62.74 (0.00)	69.50 (3.60)	65.40 (2.23)
Heart disease	85.21 (8.30)	87.59 (1.53)	87.88 (3.27)	66.30 (0.00)	56.04 (1.06)	56.24 (2.08)	56.41 (1.21)
Balance	87.08 (0.83)	87.37 (0.98)	87.13 (1.73)	65.44 (1.89)	61.30 (3.98)	65.41 (6.72)	64.74 (2.39)
Vote	93.61 (6.15)	95.21 (0.91)	96.86 (0.80)	88.05 (0.00)	62.49 (1.65)	77.46 (10.62)	63.37 (1.54)
<i>E. coli</i>	84.48 (2.27)	85.42 (1.67)	84.48 (2.49)	78.85 (2.54)	68.36 (5.22)	69.80 (4.81)	64.43 (1.20)
Thyroid	93.48 (0.89)	92.96 (1.03)	92.84 (1.56)	94.18 (1.04)	75.19 (3.65)	71.88 (0.56)	73.77 (1.98)
Image segmentation	78.67 (6.29)	78.14 (5.46)	78.12 (6.44)	65.26 (3.09)	54.19 (4.73)	53.53 (3.53)	37.62 (0.00)
Waveform	92.56 (0.51)	93.17 (0.43)	92.56 (0.42)	52.04 (0.00)	34.03 (0.10)	40.38 (6.09)	63.86 (0.00)

Fig. 6 Time evaluation of PESC against other algorithms

$MOGA(\pi, sep)$ and CSPA are MOEA based cluster ensemble algorithms. We can see that all the three Pareto ensemble strategies proposed in PESC have good performance and outperform other cluster ensemble algorithms, which gives an empirical support of the proposed Pareto ensemble strategies on the evaluated problems. In contrast with the sparse representation based multiobjective spectral clustering algorithm SRMOSC, PESC works better than SRMOSC on most datasets. Although PESC does not make a huge improvement on the clustering accuracy than SRMOSC, it is more time saving in contrast with it. When comparing PESC with multiobjective clustering algorithms MOCK, we can see that PESC outperforms MOCK on most of the tested datasets.

In addition, when we make a comparison between PESC and conventional similarity matrix based spectral clustering algorithms, PESC outperforms fully connected, mutual kNN , and ϵ -neighborhood on all the tested datasets significantly, and performs better than kNN . In contrast with these traditional algorithms, PESC overcomes the difficulty in deciding the width of the neighborhood controlling by the values of parameters in conventional methods, such as k in kNN or mutual kNN , ϵ in ϵ -neighborhood, and the scale factor σ in fully connected similarity matrix construction method.

Take the time cost of all the algorithms into consideration, we give a time evaluation comparison between PESC and other algorithms (see Fig. 6) under the same experiment conditions. In this figure, the vertical axes represents the time cost which is plot in logarithmic scale. The proposed three ensemble strategies are represented by “PESC_ES1”, “PESC_ES2”, “PESC_ES3”, respectively. We can get the following conclusion according to this experiment considering the clustering accuracy: (1) the three proposed ensemble strategies obtained the similar clustering accuracy, however, ES1 is the most time-consuming in these three strategies since that it adopts a simplified PSO algorithm to solve the optimization problem. Hence, we prefer to adopt ‘ES2’ as the ensemble strategy considering both clustering accuracy and time complexity; (2) in contrast with the other MOEA based clustering algorithms (including SRMOSC, MOCK, $MOGA(\pi, sep)$, and CSPA), the time cost of PESC is comparative with $MOGA(\pi, sep)$ and CSPA, and quite lower than that of SRMOSC and MOCK. It demonstrated that the Pareto ensemble strategies not only make improvement in the aspect of time complexity which is a burden in many MOEA clustering algorithms, but also in the aspect of clustering accuracy; (3) when we compare PESC with the traditional similarity matrix construction methods, though the time complexity of PESC is higher than them, it overcomes the difficulty of parameter value determination in constructing the similarity matrix. In general, we can derive that PESC makes a balance between time complexity and clustering accuracy.

Table 5 Semi-supervised clustering with 10% labelled data obtained from pesc against other algorithms on real-life datasets

Datasets	PESC			SRMOSC	semi-MOCK	kNN	Mutual kNN	ϵ -neighborhood	Fully connected
	ES 1	ES 2	ES 3						
Wine	95.57 (1.53)	96.97 (1.91)	97.75 (1.68)	96.40 (0.10)	95.11 (1.19)	96.07 (1.01)	58.11 (11.75)	56.57 (6.77)	62.39 (6.16)
Glass	65.92 (4.84)	66.56 (3.84)	66.71 (6.38)	63.90 (1.76)	60.93 (4.43)	60.44 (3.67)	44.02 (3.33)	48.25 (2.98)	47.12 (2.97)
Iris	94.46 (2.94)	96.04 (1.57)	95.11 (1.91)	93.53 (1.72)	96.80 (3.40)	91.93 (2.92)	64.20 (3.69)	68.00 (0.00)	69.33 (3.67)
WDBC	97.48 (0.78)	97.74 (0.68)	98.08 (0.65)	95.46 (0.64)	95.73 (0.67)	96.40 (0.70)	62.90 (0.53)	67.99 (3.35)	65.14 (1.81)
Heart disease	87.75 (1.27)	88.00 (0.97)	88.96 (1.53)	78.04 (3.91)	81.56 (0.53)	70.89 (9.46)	57.70 (4.00)	59.80 (4.87)	58.44 (2.17)
Balance	89.54 (0.66)	90.14 (0.48)	89.77 (0.75)	78.22 (3.95)	62.86 (7.03)	82.04 (1.62)	67.11 (6.84)	81.42 (1.76)	82.58 (2.09)
Vote	95.32 (0.96)	96.05 (0.70)	95.53 (6.48)	89.79 (0.56)	89.61 (1.56)	90.55 (0.80)	66.59 (5.72)	69.77 (11.41)	63.91 (1.91)
<i>E. coli</i>	84.99 (1.82)	85.46 (1.64)	85.36 (2.64)	81.84 (2.10)	85.71 (1.07)	82.35 (3.86)	76.90 (13.17)	68.10 (10.16)	68.05 (7.84)
Thyroid	93.57 (2.34)	94.16 (2.07)	94.94 (2.83)	93.65 (1.23)	92.79 (2.27)	93.30 (2.27)	78.56 (4.94)	75.35 (2.46)	73.51 (2.47)
Image segmentation	78.44 (4.94)	78.61 (4.94)	80.68 (6.53)	81.79 (2.53)	78.46 (3.55)	79.32 (4.25)	73.76 (6.57)	62.34 (7.70)	44.45 (5.26)
Waveform	92.87 (0.46)	92.96 (0.46)	93.55 (0.46)	75.30 (3.24)	79.22 (1.24)	69.74 (6.13)	34.12 (0.09)	37.38 (1.71)	63.83 (0.27)

Table 6 Semi-supervised clustering with 20% labelled data obtained from pesc against other algorithms on real-life datasets

Datasets	PESC			SRMOSC	semi-MOCK	$k/N/N$	Mutual $k/N/N$	ϵ -neighborhood	Fully connected
	ES 1	ES 2	ES 3						
Wine	96.07 (1.71)	97.47 (0.84)	97.75 (1.68)	96.94 (0.64)	96.38 (0.76)	96.80 (1.00)	58.06 (8.76)	56.80 (5.66)	69.35 (16.44)
Glass	67.02 (4.80)	68.21 (4.60)	67.94 (6.75)	64.60 (1.66)	61.66 (2.44)	63.76 (4.37)	52.06 (7.25)	49.01 (3.57)	50.23 (3.77)
Iris	95.33 (1.67)	96.13 (1.74)	96.63 (0.76)	95.30 (1.63)	97.47 (3.50)	94.73 (2.66)	66.73 (2.12)	68.00 (0.53)	69.63 (4.63)
WDBC	98.04 (0.58)	98.24 (0.47)	98.55 (0.48)	96.63 (0.63)	95.75 (0.75)	97.31 (0.54)	62.89 (0.52)	66.79 (4.23)	67.54 (3.61)
Heart disease	88.85 (1.59)	89.49 (1.60)	91.28 (1.40)	82.42 (2.59)	82.24 (0.61)	75.74 (10.75)	57.28 (0.79)	59.43 (4.62)	60.44 (4.54)
Balance	90.01 (0.71)	90.38 (0.73)	90.06 (1.24)	82.57 (1.56)	65.02 (5.34)	86.50 (1.43)	77.27 (8.96)	85.78 (1.58)	86.73 (1.32)
Vote	96.12 (0.84)	96.81 (0.77)	97.12 (0.74)	91.54 (0.92)	89.87 (0.91)	90.62 (6.91)	66.90 (8.20)	68.31 (8.27)	64.43 (4.04)
<i>E. coli</i>	85.97 (1.74)	86.45 (1.94)	86.79 (1.82)	85.95 (1.44)	85.03 (2.10)	85.55 (2.55)	80.07 (1.75)	70.29 (5.16)	63.88 (7.51)
Thyroid	94.13 (2.42)	95.16 (1.24)	95.07 (1.59)	95.14 (1.07)	93.56 (1.39)	94.39 (3.13)	78.12 (8.43)	73.70 (1.61)	74.23 (3.35)
Image segmentation	78.39 (6.16)	79.13 (5.45)	80.93 (6.93)	90.02 (2.03)	83.30 (2.57)	82.69 (6.80)	81.26 (4.22)	64.17 (4.58)	44.12 (5.01)
Waveform	93.24 (0.27)	93.56 (0.29)	94.04 (0.26)	87.22 (0.53)	85.33 (0.56)	86.23 (0.47)	34.20 (0.15)	42.40 (3.10)	60.54 (0.90)

The comparison results of semi-supervised clustering algorithms

Tables 5 and 6 show the comparison results between PESC and other semi-supervised clustering algorithms with 10% and 20% labeled samples. SRMOSC transforms the semi-supervised information to a constraint to satisfy. Semi-MOCK handles semi-supervised information with an objective function called ‘Adjusted Rand Index’, which is an external measure of clustering quality. In traditional similarity matrix construction methods, all the entries of the labelled samples with the same label are set to the maximum value of its similarity matrix, and the corresponding entries with different labels are set to 0. From these tables, we can see that PESC, no matter which ensemble strategy is adopted, gains better performance than other multiobjective clustering algorithms on most of the tested datasets.

Conclusion

Multiobjective evolutionary algorithm has gained a growing concern in the field of data mining according to the recent literatures. A lot of related works are proposed in the last decades, but few researches focus on the topic of Pareto ensemble or spectral clustering. In this paper, we introduce multiobjective evolutionary algorithm into spectral clustering, and propose a Pareto ensemble based framework for spectral clustering. The main contributions can be summarized as follows.

First and the most important is that we model the similarity matrix construction task as a Pareto ensemble problem, it not only overcomes the difficulty in determining the value of parameters in the conventional methods, but also time-saving than the current existed multiobjective spectral clustering algorithm.

Second, to solve the proposed model effectively, we design a specific diversity preservation strategy for it and also employ this strategy in one of the proposed ensemble strategies. The experiment results derive that this diversity preservation strategy is effective and time-saving than the other ensemble strategies. In addition, a specific neighbor-biased initialization strategy is also proposed, it helps to improve convergence speed and reduce calculation amount during the evolutionary process.

Furthermore, we also extend this model to semi-supervised clustering by transforming the semi-supervised information to some constraints of MOPs, and detailed experiments show the efficiency in handling the related problems.

At last, PESC mainly focuses on the construction of the similarity matrix, so it is easy to extend it to the related field, such as subspace learning.

In general, PESC can obtain a satisfying result in not only the clustering accuracy, but also the time and space complexity in contrast with other multiobjective clustering algorithms. The goal of making a balance between time cost and clustering accuracy is achieved by employing MOEA and ensemble strategy in this paper.

Acknowledgements This work is supported by the National Key R&D Program of China under grant 2018AAA0101200, and the National Natural Science Foundation of China under Grant 61806019.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Albukhanajer WA, Jin Y, Briffa JA (2014) Neural network ensembles for image identification using pareto-optimal features. In: Evolutionary computation, pp 89–96
- Cai D, Chen X (2014) Large scale spectral clustering via landmark-based sparse representation. *IEEE Trans Cybern* 45(8):1669–1680
- Chen H, Yao X (2010) Multiobjective neural network ensembles based on regularized negative correlation learning. *IEEE Trans Knowl Data Eng* 22(12):1738–1751
- Chen WY, Song Y, Bai H, Lin CJ, Chang EY (2010) Parallel spectral clustering in distributed systems. *IEEE Trans Pattern Anal Mach Intell* 33(3):568–586
- Cheng S, Lu H, Lei X, Shi Y (2018) A quarter century of particle swarm optimization. *Complex Intell Syst* 64:227–239
- Chi KC, Yuen SY (2012) A multiobjective evolutionary algorithm that diversifies population by its density. *IEEE Trans Evol Comput* 16(2):149–172
- Coello CAC, Brambila, SG, Gamboa JF, Tapia MGC, G' *textregistered* mez RH (2020) Evolutionary multiobjective optimization: open research areas and some challenges lying ahead. *Complex Intell Syst* 6:221–236
- Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput* 6(2):182–197
- Garcia-Piquer A, Fornells A, Bacardit J, Orriols-Puig A, Golo-bardes E (2014) Large-scale experimental evaluation of cluster representations for multiobjective evolutionary clustering. *IEEE Trans Evol Comput* 18(1):36–53
- Garza-Fabre M, Handl J, Knowles J (2018) An improved and more scalable evolutionary approach to multiobjective clustering. *IEEE Trans Evol Comput* 22(4):515–535
- Gong Z, Chen H, Yuan B, Yao X (2019) Multiobjective learning in the model space for time series classification. *IEEE Trans Cybern* 49(3):918–932
- Hamad D, Biela P (2008) Introduction to spectral clustering. In: Information and communication technologies: from theory to applications. ICTTA 2008. 3rd International Conference on, pp 1–6
- Handl J, Knowles J (2005) Improvements to the scalability of multiobjective clustering. In: IEEE Congress on evolutionary computation, vol 3, pp 2372–2379
- Handl J, Knowles J (2007) An evolutionary approach to multiobjective clustering. *IEEE Trans Evol Comput* 11(1):56–76
- Hui L, Zhang Q, Deng J (2017) Biased multiobjective optimization and decomposition algorithm. *IEEE Trans Cybern* 47(1):52–66
- Karakaya G, Galelli S, Ahipasaoglu SD, Taormina R (2016) Identifying (quasi) equally informative subsets in feature selection problems for classification: a max-relevance min-redundancy approach. *IEEE Trans Cybern* 46(6):1424–1437
- Li L, Yao X, Stolkin R, Gong M, He S (2014) An evolutionary multiobjective approach to sparse reconstruction. *IEEE Trans Evol Comput* 18(6):827–845
- Li X, Wong K (2019) Evolutionary multiobjective clustering and its applications to patient stratification. *IEEE Trans Cybern* 49(5):1680–1693
- Lu H, Zhang R, Li S, Li X (2013) Spectral segmentation via midlevel cues integrating geodesic and intensity. *IEEE Trans Cybern* 43(6):2170–2178
- Luo J, Jiao L, Lozano JA (2016) A sparse spectral clustering framework via multi-objective evolutionary algorithm. *IEEE Trans Evol Comput* 20(3):418–433
- Maier M, Hein M, Von Luxburg U (2009) Optimal construction of k-nearest neighbor graphs for identifying noisy clusters. *Theor Comput Sci* 410(19):1749–1764
- Mukhopadhyay A, Bandyopadhyay S, Maulik U (2009) Analysis of microarray data using multiobjective variable string length genetic fuzzy clustering. In: IEEE Congress on evolutionary computation, 2009. CEC'09, pp 1313–1319
- Mukhopadhyay A, Maulik U (2011) A multiobjective approach to MR brain image segmentation. *Appl Soft Comput* 11(1):872–880
- Mukhopadhyay A, Maulik U, Bandyopadhyay S (2009) Multi-objective genetic algorithm-based fuzzy clustering of categorical attributes. *IEEE Trans Evol Comput* 13(5):991–1005
- Mukhopadhyay A, Maulik U, Bandyopadhyay S, Coello CAC (2014) Survey of multiobjective evolutionary algorithms for data mining: part I. *IEEE Trans Evol Comput* 18(1):4–19
- Mukhopadhyay A, Maulik U, Bandyopadhyay S, Coello CAC (2014) Survey of multiobjective evolutionary algorithms for data mining: part II. *IEEE Trans Evol Comput* 18(1):20–35
- Nguyen BH, Xue B, Andreae P, Ishibuchi H, Zhang M (2020) Multiple reference points-based decomposition for multiobjective feature selection in classification: static and dynamic mechanisms. *IEEE Trans Evol Comput* 24(1):170–184
- Nie F, Li J, Li X (2016) Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), pp 1881–1887
- Nie F, Wang H, Deng C, Gao X, Li X, Huang H (2016) New ℓ_1 -norm relaxations and optimizations for graph clustering, pp 1962–1968
- Nie F, Wang X, Huang H (2014) Clustering and projected clustering with adaptive neighbors. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 977–986
- Nie F, Wang X, Jordan MI, Huang H (2016) The constrained Laplacian rank algorithm for graph-based clustering. In: Thirtieth AAAI Conference on Artificial Intelligence
- Nie F, Xu D, Li X (2012) Initialization independent clustering with actively self-training method. *IEEE Trans Syst Man Cybern Part B* 42(1):17–27

33. Nie F, Zeng Z, Tsang IW, Xu D, Zhang C (2011) Spectral embedded clustering: A framework for in-sample and out-of-sample spectral clustering. *IEEE Trans Neural Netw* 22(11):1796–808
34. Preuss M, Kausch C, Bouvy C, Henrich F (2010) Decision space diversity can be essential for solving multiobjective real-world problems. Springer, Berlin, Heidelberg
35. Ripon KSN, Siddique MNH (2009) Evolutionary multi-objective clustering for overlapping clusters detection. In: *Evolutionary computation. CEC'09. IEEE Congress on*, pp 976–982
36. Ripon KSN, Tsang CH, Kwong S (2006) Multi-objective data clustering using variable-length real jumping genes genetic algorithm and local search method. In: *Neural Networks, 2006. IJCNN '06. International Joint Conference on*
37. Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20(20):53–65
38. Shir OM, Preuss M, Naujoks B, Emmerich M (2009) Enhancing decision space diversity in evolutionary multiobjective algorithms. In: *Evolutionary Multi-Criterion Optimization, International Conference, EMO 2009, Proceedings, , April 7–10, 2009. Nantes*, pp 95–109
39. Smith C, Doherty J, Jin Y (2014) Multi-objective evolutionary recurrent neural network ensemble for prediction of computational fluid dynamic simulations. In: *Evolutionary computation*, pp 61–85
40. Strehl A, Ghosh J (2003) Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *J Mach Learn Res* 3(3):583–617
41. Tan KC, Lee TH, Khor EF (2001) Evolutionary algorithms with dynamic population size and local exploration for multiobjective optimization. *IEEE Trans Evol Comput* 5(6):565–588
42. Topchy AP, Jain AK, Punch WF (2004) A mixture model for clustering ensembles. In: *Siam International Conference on Data Mining, Lake Buena Vista, Florida*
43. Ulrich T, Bader J, Zitzler E (2010) Integrating decision space diversity into hypervolume-based multiobjective search. In: *Genetic and evolutionary computation conference, GECCO 2010, Proceedings, Portland, Oregon*, pp 455–462
44. Von Luxburg U (2007) A tutorial on spectral clustering. *Stat Comput* 17(4):395–416
45. Wan Y, Ma A, Zhong Y, Hu X, Zhang L (2020) Multiobjective hyperspectral feature selection based on discrete sine cosine algorithm. *IEEE Trans Geosci Remote Sens* 58(5):3601–3618
46. Wang H, Jin Y, Yao X (2017) Diversity assessment in many-objective optimization. *IEEE Trans Cybern* 47(6):1510–1522
47. Wang P (1983) Pattern recognition with fuzzy objective function algorithms. *SIAM Rev* 25(3):442–442
48. Wright J, Ma Y, Mairal J, Sapiro G, Huang TS, Yan S (2010) Sparse representation for computer vision and pattern recognition. *Proc IEEE* 98(6):1031–1044
49. Xia H, Zhuang J, Yu D (2014) Combining crowding estimation in objective and decision space with multiple selection and search strategies for multi-objective evolutionary optimization. *IEEE Trans Cybern* 44(3):378–393
50. Xie XL, Beni G (1991) A validity measure for fuzzy clustering. *IEEE Trans Pattern Anal Mach Intell* 13(8):841–847
51. Zhang Q, Li H (2007) MOEA/D: a multiobjective evolutionary algorithm based on decomposition. *IEEE Trans Evol Comput* 11(6):712–731
52. Zhang X, Jiao L, Liu F, Bo L, Gong M (2008) Spectral clustering ensemble applied to SAR image segmentation. *IEEE Trans Geosci Remote Sens* 46(7):2126–2136
53. Zitzler E, Laumanns M, Thiele L (2002) SPEA2: Improving the strength pareto evolutionary algorithm. *Evol Methods Design Optim Control* 3242:95–100

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.