An Improved Method for Multi-objective Clustering Ensemble Algorithm

Ruochen Liu, *Member, IEEE*, Yong Liu, Yangyang Li, *Member, IEEE*

Abstract—In this paper, we present a cluster algorithm which is an improvement of the multi-objective clustering ensemble algorithm (MOCLE), which is denoted as IMOCLE for short. First, we introduce a new clustering objective function to measure the individual difference in the optimization process so as to remain the diversity of the population. Then, a clustering ensemble technique is applied to MOCLE to obtain more competitive individual. The proposed algorithm can also ensure good partitions not be eliminated. The performance of the proposed algorithm has been compared with MOCLE over a suit of gene datasets. The experimental results show that, the superiority of the proposed method in terms of capability found the optimum number of clusters, and accuracy

Keywords: *multi-objective clustering; cluster ensemble; diversity*

I. INTRODUCTION

The goal of cluster analysis is to discover the relations between given objects, and assign them to corresponding groups. Clustering algorithms have been widely used in solving real problems.

Most of the traditional cluster algorithms (e.g. partitional clustering) gets the clustering result by optimizing one or more different objective functions (clustering criterions). There are several problems that are needed to be solved for traditional clustering algorithms. The most obvious one is that there is not a precise definition of a cluster [1]. Different data sets have different structures, and there is not a definition of cluster for all kind of data. So different cluster criterions are proper for different shaped clusters [2]. For example, the widely used clustering algorithm k-means is biased towards spherically shaped clusters, because it mostly uses Euclidean metric for computing distance between points. Meanwhile, for a given clustering algorithm, it may lead to different results with different parameters. In other words, some algorithms are sensitive to their parameter settings [1] [3]. Most of the time, we have no prior information to decide how parameters are selected.

For the first problem mentioned above, it is necessary to apply several different clustering criterions to describe the clusters. Multi-objective clustering algorithms which optimize

This work was supported by the National Natural Science Foundation of China (No. 60803098), Research Fund for the Doctoral Program of Higher Education of China (No. 20070701022); the Provincial Natural Science Foundation of Shaanxi of China (2010JM8030, No. 2010JQ8023), and the Fundamental Research Funds for the Central Universities (No.K50511020014 No.K50510020011).

All authors are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an 710071, China. (Phone: +86-029-88202661; email: aliang3399@gmail.com).

several cluster criterions simultaneously are proposed. Many multi-objective optimization (MOO) clustering algorithms have been developed recently. In [4], Handl and Knowles proposed an MOO clustering algorithm, namely, MOCK, which can automatically detect the appropriate partitioning from a data set having either the hyperspherical shaped clusters or the well-separated clusters. In [5], a symmetry based multi-objective clustering technique (VAMOS) is developed, which is based on the simulated annealing (SA) algorithm. Faceli and Marcilio have presented a multi-objective clustering algorithm which refers to both multi-objective methods and cluster ensemble techniques in the optimization process (MOCLE) [1]. Experiment results have shown that multi-objective clustering algorithms have obviously advantages over the algorithms optimizing one single cluster criterion.

For the second problem mentioned, one algorithm may result in different partitions with different parameters setting, clustering ensemble methods can be applied to get a combined partition from the partitions obtained with different parameter settings. Some work on combining multiple clustering algorithms can be found in [6-8]. The cluster-based similarity partitioning algorithm (CSPA) which is proposed in [9], is based on a co-association matrix, and METIS which is a software package for partitioning unstructured graphs and hyper-graph [10]. In the hyper-graph partitioning algorithm(HGPA) [9], the combination is dealt with as a problem of partitioning a hypergraph. In this hypergraph, the clusters of the base partitions are represented as hyperedges. The hypergraph is partitioned by cutting a minimal number of hyperedges [9, 11]. The meta-clustering algorithm (MCLA) algorithm is one of the most popular methods which is proposed in [9]. In MCLA, each cluster is represented by a hyper-edge, like HGPA. MCLA is composed of the following steps: (1) constructing the meta-graph, (2) partitioning the metagraph, and (3) computing cluster members.

Multi-objective clustering ensemble algorithm (MOCLE) is a novel multi-objective algorithm combined with cluster ensemble methods. In [9], MOCLE is applied to the clustering analysis of microarray data with cancer. In our study of the algorithm MOCLE, we have found that some good solutions sometimes will be eliminated in the optimization process, which is cause by the objective functions. Considering this, in order to improve the solutions obtained, we present a new objective function. The function refers to the diversity of the population in the optimization process. It does good to save the good solutions which are not well evaluated by cluster criterions, meanwhile, it can ensure a diverse population for the ensemble method applied.

The remaining part of this paper is organized as follows: Section II introduces the background and related work. Section III describes the problems existed in MOCLE and presents our method of improving MOCLE. Section IV gives the experimental results. Section V concludes this paper.

II. BACKGROUND AND RELATED WORK

The aim of clustering algorithm is to infer properties of a given data set in an unsupervised way and divide the data objects into groups so that objects in the same group are more similar than ones in different groups [12, 13, 14]. The result of a clustering algorithm on a given data is a partition of the objects. Each object should be assigned to only one cluster.

A. Traditional cluster algorithms

There are two main different groups of traditional clustering algorithms: hierarchical clustering and partitional clustering. Hierarchical clustering algorithms are divided into two groups by different clustering steps. The first groups are called agglomerative clustering algorithms. These algorithms start with each objective as a cluster, and then similar clusters are grouped into one. The other groups start with all the objects as one clusters. Then the only one cluster is divided into smaller clusters. The clusters obtained are then divided recursively. Three well known hierarchical algorithms are: single linkage, complete linkage and average linkage [15, 16]. They differ from each other in their ways distance between two clusters is calculated. In complete linkage (CL), the distance inter-cluster distance is calculated by the farthest distance of a pair of points in two clusters. Different from CL, the average link age (AL) takes the average distance between all the objects in one cluster and all the objects in another cluster.

The most well known partitional cluster algorithm is the k-means algorithm, which is a partitional iterative algorithm that optimizes the function between clusters and their representatives using a pre-defined number of clusters [17][18]. A defect of the k-means algorithm is that is sensitive to the selection of the initial centroids [17]. Although the point mentioned above, it is still one of the most widely used algorithm for clustering for its easy implementation and high efficiency.

Recently, biological information process system has inspired natural computation such as evolutionary algorithm, immune computation, particle swarm algorithm and so on. It provides new solutions for clustering analysis. Especially, many clustering algorithms based on evolutionary computation have been proposed.

B. Multi-objective clustering and cluster ensembles

Traditional clustering algorithms, such as k-means [13], optimizes only one clustering criterion (e.g., compactness of the clusters) and are often very effective in this purpose. However, they fail for data in accordance to a different criterion. One alternative is to combine several clustering criterions to improve the quality of the final solution. Nowadays, there are two main approaches that address the use of multiple clustering criteria: cluster ensemble and the multi-objective clustering [1].

The application of multi-objective optimization techniques proves to be an alternative and promising direction [19]. MOO simultaneously optimize several different cluster criterions instead of one. MOO may get a better result than the algorithm that optimizes only one criterion. Many cluster criterions have been developed to find partitioning consisting of compact and well-separated clusters. MOO clustering algorithms result in a set of Pareto optimal solutions, none of which is better than another considering the criterions used.

The success of ensemble methods for supervised learning has motivated the development of ensemble methods for unsupervised learning[9].

The ensemble problems can be described as follows[9] [20]. Let Z be a data set and let $P = \{P_1, \dots, P_L\}$ be a set of partitions on Z . partitions are obtained by applying different clustering algorithms or an algorithm with different parameter settings. The goal is to find a single partition based on the information contained in the set P .

A clustering ensemble algorithm contains two main parts . First, a set of partitions are obtained. These partitions are usually named ensemble members. Then, ensemble members are combined via a consensus function [11][21]. There are many ways acting as the consensus functions, such as methods based on co-association, graph partitioning, voting based methods and so on. The goal of an ensemble of clustering solutions is to find a consensus partition that optimally summarizes the ensemble and to obtain a clustering solution with improved accuracy and stability compared to the individual members of the ensemble [22]. Cluster ensemble methods seem to get better solutions than a single cluster algorithm for the datasets with complex structures. Different algorithms result in different clustering solutions. These results represent different structures of a data set. Cluster ensembles merge all these different structures into one using the information provided by the different structures different algorithms obtained. In order to get a good solution by using cluster ensemble method, two aspects of the partitions to be combined are need to be taken into account. First, partitions should have a high accuracy rate corresponding to the true cluster label. Second, partitions used in ensemble algorithms should be diverse. Many recent studies have been concentrated on the issue of constructing a set of accurate and diverse ensemble members (basic partitions).

C. The combination of multi-objective clustering algorithm and cluster ensemble

Recently, Faceli and Marcilio have presented a multi-objective clustering algorithm (MOCLE), which refers to both multi-objective methods and cluster ensemble techniques in the optimization process. The main steps of MOCLE are as follows: First, initial populations are obtained by applying several different clustering algorithms on a given data set. Then several objective functions are optimized in the evolution process with a special crossover which combines two parents using cluster ensemble technique. In the end, a set of solutions are obtained [11].

III. SOME PROBLEMS IN MOCLE AND OUR WORK

A. A brief introduction to MOCLE

In [11], MOCLE is applied to the cluster analysis of the gene expression data. Our improved method is an improved version of MOCLE, and we will have a simple introduction for MOCLE.

- Data transform

Data sets used in MOCLE are transformed in two widely used ways in clustering [23] [24]. One is based on the z-score formula (standardization) and the other (normalization) involves the use of the maximum and minimum values on the features of the data.

- Initial population

Four different algorithms are used to get initial population such as k-means, CL, AL, and SPC (spectral clustering) .

The size of the initial population is related to the true cluster number of the dataset which is not used in the optimization process and just used in the obtaining of initial population to decide the size of the population. Initial population is generated with K in the range $[c-2, c+2]$, where c is the true cluster number of a given data set and K is the cluster number applied to the four different algorithms to get the clustering result. If $c-2 < 2$, the minimum of the range is set to 2. For each K , four different algorithms AL, CL, KM and SPC are used to get the basic partitions. Each algorithm is implemented with two different proximity indices: Pearson correlation and Euclidean distance. For the Euclidean version, experiments are performed with the data set in three different ways: original data, standardized data and normalized data (data transformation is introduced above).

- String representation

In MOCLE, the strings are made up of integer numbers which represent the coordinates of the cluster label of a data point. For example, for the data set $X=\{x_1, x_2, x_3, x_4\}$, string $\{1, 2, 2, 3\}$ represents three clusters: x_1 belongs to cluster 1, x_2 and x_3 belong to cluster 2, x_4 belongs to cluster 3.

- Objective functions

The multi-objective algorithm applied in MOCLE is the non-dominated sorting genetic algorithm II (NSGAII) [25].

In MOCLE, two measures are used as objective functions: overall deviation and connectivity [1] [19] are applied. A version of connectivity is implemented with Pearson correlation. So there are three objective functions applied in MOCLE [11].

- Crossover and mutual operation

In MOCLE, cluster ensemble is applied to combine two parent partitions. First, two parents are selected by using binary tournament. k^1 and k^2 are the respective numbers of the two parents π^1 and π^2 . The number of cluster k^c in the

resulting ensemble partitioning is randomly picked in the interval $[k^1, k^2]$. Then the ensemble algorithm is applied to the two parent partitions to get one offspring partition with k^c clusters. With this operator, the partitions are combined in pairs, iteratively, during the evolution process [1].

The cluster ensemble algorithm applied in MOCLE is the meta-clustering algorithm (MCLA) [9].

MOCLE does not apply a mutation operator in order to restrict the search space to the base partitions and its combination.

The steps of MOCLE are executed until the iteration times reach as much as 50 for the solutions are not obviously improved

- Criterion of measuring the algorithm

To measure the success of the algorithm in recovering the true partition of the data sets, we use the corrected Rand (CR) [26] as MOCLE does. The value of CR is ranging from -1 to 1. The larger the CR is, the better the partition is.

B. Our work

We have given an introduction of MOCLE in details. In [11], many experiments have been done to show the advantages of the algorithm. MOCLE applies four different clustering algorithms to produce its initial population. It results in a diverse population. In our experiment, we have found that the best partitions obtained by the optimization process are not better than the best partition in the initial population for all the six data sets tested in [11], which can be also found in the experiment result shown in [11]. Take data set “Armstrong” for example, the best solution in the initial population is 0.7038 considering the value of CR, however, the best solution after optimization process becomes 0.5146.

We have found that the special crossover operator can produce some good partitions. But why the good partitions and those of best partition in the initial population do not exist in the final result? We think it is related to the objective functions. We also take “Armstrong” for example. Assuming that the best solution in the initial population exists in the final population, it will not be selected as the Pareto solution, because it is donated by other solutions. In other words, the three clustering objective functions applied in MOCLE are not so good to evaluate the good partitions. This may also lead to good solutions produced by the special crossover to be eliminated in the optimization process. This phenomenon is also tightly related to the data sets. The six data sets tested in MOCLE are gene expression data which have very high dimensions. Although filtered, the data are still very high in dimension. The ordinary clustering criterions sometimes can not describe the overall properties of data with high dimension. If we improve the objective functions, the problem may still exist.

In our study, in order to analyze the property of the whole population in the optimization process, we compare all the individuals to the true label of the tested data set. It is needed

to say that the true label is just used to analyze the population but it is not used in the optimization process. We have found that although some good partitions exist in the population in the optimization process, most of the partitions are not so good. That is to say if we calculate the similarity between one partition and the other partitions in the current population, and then obtain the average value. The result for the good partition must be small, because such partitions are just a small part of the population and they are not similar to the most partitions in the population. The average similarity for partitions that are not so good should be big, for they are most part of the population and more similar to each other. Meanwhile, the average similarity for very poor partitions is small. Motivated by this way of thinking, we present an objective function referring to the similarity of one partition to others. We call it similarity. Here we write it Sim for short. The definition is as follows:

$$Sim(\pi_i) = \frac{1}{n} \sum_{j=1}^n S(\pi_i, \pi_j)$$

where π_i is the partition to be evaluated. $S(\pi_i, \pi_j)$ is the agreement between the two partitions. We randomly select n data to form the current population. Then we calculate the average agreement between the partition to be evaluated and the selected partitions. For the last results are measured by calculating CR between solution and the true cluster label of a given data set, here $S(\pi_i, \pi_j)$ is the value of CR between the partition π_i and π_j . We randomly select n partitions because those very poor partitions have the similar property with good partitions in similarity, which can increase the random possibility. Similarity reflects the similarity of one partition with others in the current population. Our goal is to minimize similarity. There are two aspects to explain the reasons. First, as mentioned above, the smaller value indicates that the partition may be a good partition corresponding to the true label of a given data set, since most of the partitions in the population are not good corresponding to the true label, and they are more similar to each other. If the other three cluster criterions can not evaluate the good partition very well, it is a good way to save the good partitions by calculating the value of similarity. Another intention we apply such an objective function is that it can ensure a diverse population. Since similarity reflects the agreement of a partition to some others, one partition with smaller similarity exists in a higher possibility. MOCLE applies a special crossover operator by using clustering ensemble method. A diversity population is needed to produce good partition by applying clustering ensemble methods. The other objective functions applied in the optimization process can ensure the “accuracy” of the partitions, and similarity ensures a diverse set of partitions. The “accuracy” and diverse partitions fulfill the requirement of the cluster ensemble algorithm applied.

Normally, clustering criterions reveal the structure property of a given data set or reflect the relations of data objects to their neighbor objects. Multi-objective clustering algorithms optimize several clustering criterions. The

clustering criterions used in optimization process should contain not only ones reflect the inner property of one partition but also the criterions reflect the relation between different partitions. Similarity is such an objective function. It is a complement to ordinary cluster criterions.

In a word, we apply a new function objective function, namely, *similarity*, together with the three objective functions applied in MOCLE, which ensure good partitions in the optimization process.

IV. EXPERIMENT

A. Data sets

Eleven microarray data sets are included in our analysis. They were available in [27-37]. Most microarray data sets are not very large, but they are high in dimensionality. From Table 1, we can also find this phenomenon. The largest data set tested in our experiment is Yeoh which was also the largest one used in [11]. All the data sets are filtered as they are in [11]. As Table 1 illustrates, the first column is the name of the data sets, the second column is the number of objects, the third column is the number of classes (class), the fourth is the number of dimension(d) and the last one is the dimensionality after feature selection (Filtered d).

Table 1. The details of datasets

Data set	n	class	d	Filtered d
Armstrong [27]	72	2	12582	1081
Dyrskjot [28]	40	3	7129	1203
Chowdary [29]	104	2	22283	182
Gordon [30]	181	2	12533	1626
Golub [31]	72	2	7129	1877
West [32]	49	2	7129	1198
Laiho [33]	37	2	22883	2202
Chen [34]	179	2	22699	85
Alizadeh [35]	42	2	4022	1095
Bredel [36]	50	3	41472	1793
Yeoh [37]	248	6	12625	2526

From Table 1, we can see that the number of objects in each data set is not big. However, the data sets are all high in dimensionality. Although, all the data sets are filtered by feature selection, the dimension is still high.

B. Experiment parameter settings

All the parameter settings in our experiments are the same as the ones in the MOCLE including the size of the initial population, the number of nearest neighbors (or most correlated objects for the connectivity by correlation). To calculate the connectivity and connectivity with correlation was set to 5% of the number of objects in the data sets. To calculate the similarity we apply, 10% of the number of the current population is randomly selected, for we have found 10% of the population is enough.

Table 2 illustrates parameter setting in IMOCLE for each data set. Here, s stands for the number of partitions in the initial population. N stands for the number of nearest neighbors for the connectivity. k interval represents the numbers of clusters used to generate the population. For the

size of population in the optimization is changing, we have not shown the number to calculate the similarity.

Table 2. Parameter values for IMOCLE

Data set	s	N	k interval
Armstrong	48	4	[2,4]
Dyrskjot	64	2	[2,5]
Chowdary	48	6	[2,4]
Gordon	48	10	[2,4]
Golub	64	4	[2,5]
West	48	3	[2,4]
Laiho	48	2	[2,4]
Chen	48	9	[2,4]
Alizadeh	48	3	[2,4]
Bredel	64	3	[2,4]
Yeoh	80	13	[4,8]

C. Result and discussion

The result is average results of 10 runs of MOCLE and IMOCLE. We calculate CR between each solution and the true label of the data set. Then we selected the largest CR. After that we calculate the mean and standard deviation of the CR of the partitions over the 10 runs.

- Best partitions in initial population

Like MOCLE, the perspective of the initial population is analyzed by calculating CR between each partition in the initial population and the true label of the tested given data set. Different from MOCLE, we just show the best CR in the initial population. The goal is to have a comparison to the best CR for MOCLE after the optimization process.

Now, let us focus our attention to the result obtained with the initial population. Table 3 illustrates the mean (M) and its standard deviation (SD) of the best CR in the initial population.

Table 3. CR for best partitions in the initial population

Data set	CR	
	M	SD
Armstrong	0.7037	0.0000
Dyrskjot	0.6114	0.0451
Chowdary	0.9238	0.0000
Gordon	0.9727	0.0000
Golub	0.6522	0.0407
West	0.5613	0.0000
Laiho	0.5145	0.0000
Chen	0.7116	0.0232
Alizadeh	0.4269	0.0411
Bredel	0.5342	0.0272
Yeoh	0.4303	0.0354

From Table 3, although most of the data sets are high in dimensionality we can find that good solutions ($CR > 0.5$) exist in the initial population for most of the data sets, such as Armstrong, Dyrskjot, Chowdary, Gordon, Golub, Laiho, Chen Bredel and West. The best CR for the Gordon even reaches as much as 0.9727. The result shows that using different algorithms to get the basic partitions can result in some good solutions in the population.

- The result for MOCLE

Table 4 illustrates the mean (M) and standard deviation (SD) of the best CR and the corresponding cluster number K in MOCLE.

Table 4. Best CR for MOCLE

Data set	CR		K	
	M	SD	M	SD
Armstrong	0.5124	0.0000	3.0	0.0000
Dyrskjot	0.4741	0.1164	5.0	0.0000
Chowdary	0.9238	0.0000	2.0	0.0000
Gordon	0.9766	0.0189	2.0	0.0000
Golub	0.9724	0.0532	2.0	0.0000
West	0.4523	0.0813	3.2	0.0813
Laiho	0.3275	0.0000	3.0	0.0000
Chen	0.5808	0.0830	3.8	0.4629
Alizadeh	0.0630	0.1356	3.3	0.6749
Bredel	0.6052	0.0805	2.3	0.4830
Yeoh	0.1755	0.0000	8.0	0.0000

We compare the result of MOCLE with the best initial CR for the initial population (Table 3). As we highlight the better results by using boldface in Table 4. For Golub, Gordon and Bredel, and MOCLE can get partitions better than the best partitions in the initial population. This is because the objective functions applied are proper to evaluate the data set, and the special crossover MOCLE is applied can produce good partitions. For the remaining data sets, the best CR partition in the initial population is larger than the best CR for the MOCLE. That is to say, in most of the case MOCLE obtain the partitions that are not so good. In this case, the effect of the special crossover operator is decreased. The goal of the crossover operator is to find good partitions, there has been some good partitions existing in the initial population, however, good partitions like them do not exist in the solutions obtained by MOCLE. It indicates that even if good partitions are produced by the special crossover operator, they will be eliminated in the optimization process.

MOCLE find the large number of clusters for the most of the time. The reason may be mostly related to the evaluated criterions CR.

- The results for IMOCLE

Now, let us focus our attention to the result obtained by IMOCLE.

Table 5: Best CR for IMOCLE

Data set	CR		K	
	M	SD	M	SD
Armstrong	0.7496	0.1403	2.4	0.8433
Dyrskjot	0.5839	0.0762	4.5	0.5270
Chowdary	0.9258	0.0133	2.0	0.0000
Gordon	0.8957	0.0783	2.0	0.0000
Golub	0.9780	0.0528	2.0	0.0000
West	0.5028	0.0210	3.3	0.6749
Laiho	0.5196	0.1161	2.6	0.6992
Chen	0.7081	0.0339	2.9	0.3162
Alizadeh	0.3188	0.0667	3.9	0.3162
Bredel	0.6040	0.0548	2.1	0.3126
Yeoh	0.4051	0.0576	6.4	1.3333

First, we compare the result obtained by IMOCLE (Table 5) to the result obtained by MOCLE (Table 4). IMOCLE get a larger *CR* than MOCLE except Gordon and Bredel. The result illustrates the fact that IMOCLE have a better performance than MOCLE on obtaining good partitions.

We also compare the result in Table 5 with those of the best *CR* partition in the initial population (Table 3). For Armstrong, Chowdary, Golub, Laiho and Bredal, IMOCLE can obtain a larger *CR* than the best *CR* in the initial population for our application of the new objective function, similarity. For the rest data sets, IMOCLE still can not reach the largest *CR* value in the initial population as MOCLE, but the value are improved than MOCLE.

From what mentioned above, we can find that the result has been improved due to the application of the new objective function, for the objective function, similarity, can ensure good partitions exist in a high possibility and it can also do good to keep a more diversity population, which is very important for the application of clustering ensemble method.

D. Analysis of parameters

In IMOCLE, we apply a new objective function which is named similarity. In the optimization process, we randomly select n individuals to calculate the similarity of each partition. In our experiment, we set n to the 10% of the size of current population. We test the effect of the selection proportion through increasing selection proportion from 10% to 1 for the data sets: Armstrong and Chowdary

Fig.1 result for Armstrong data set

Fig.2 result for Chowdary data set

From Fig.1, we can find that the selection proportion has a great influence on the result obtained by IMOCLE for the Armstrong data set. When the selection proportion reaches 40%, the largest average *CR* is obtained. This may be mostly related to the number of good partitions existing in that current population in the optimization process. Form Fig.2, we can find that although the select proportion has influence on the last result for the Chowdary data set, the value of *CR* ([0.9238-0.9258]) floats not so heavily. When the selection proportion is 0.1(10%), the largest average *CR* is obtained. If we have a comparison between the two figures, we can get a conclusion that different selection proportion is needed to get the largest *CR* for different data sets. Up to now, we have not obtained the prior information to decide which selection proportion can lead to the best result, but what we are sure is that setting the parameter to 10% can get better result than MOCLE in most conditions. On another aspect, the smaller the selection proportion is, the less time is cost on calculating the objective function. In a word, 10% is a good choice to calculate the objective function, similarity.

V. CONCLUSION AND FUTURE WORK

In this paper, we present an improved method for the multi-objective clustering ensemble method. In the method, we apply a new objective function named similarity. To minimize the value of similarity have two aspects affecting the evolution process. First, it helps to save the good partitions that the other three clustering criterions can not evaluate very well. Second, it can ensure a diversity population in the optimization process. Since clustering ensemble method is applied in the algorithm, a diversity set of partitions are needed. Complementing each other, the four objective functions ensure that high quantity of the partitions in the optimization process. .

Much work is needed to be done. Because of the application of the new objective function, similarity, the number of the Pareto solutions is increased. The reason is that the objective function keeps a diversity set of partitions. We have not shown the number of the Pareto solutions in the

paper, for our goal is to find as good partitions as possible. Further work on how to select a single solution from the Pareto solutions is needed. More work on the analysis of the parameters of the new objective function is also needed.

REFERENCES

- [1] K. Faceli, A. Carvalho, M. de Souto, Multi-objective clustering ensemble, *International Journal of Hybrid Intelligent Systems*, Special Issue: Ensemble and Integration Approaches 4 (3) (2007) 145-156.
- [2] M. Law, A. Topchy and A.K. Jain, Multi-objective data clustering, in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, volume 2, 2004, 424-430
- [3] A. Jain and R. Dubes, *Algorithms for Clustering Data*, PrenticeHall, 1988.
- [4] J. Handl, J. Knowles, An evolutionary approach to multi-objective clustering, *IEEE Transactions on Evolutionary Computation* 11 (1) (2007) 56-76.
- [5] S. Saha, S. Bandyopadhyay: A symmetry based multiobjective clustering technique for automatic evolution of clusters, *Pattern Recognition* 43 (2010) 738-751.
- [6] H. G. Ayad, M.S. Kamel, On voting-based consensus of cluster ensembles, *Pattern Recognition* 43 (5) (2010) 1943-1953.
- [7] X. Wang, C. Yang, J. Zhou, Clustering aggregation by probability accumulation, *Pattern Recognition* 42 (5) (2009) 668-675.
- [8] S. Vega-Pons, J. Correa-Morris, J. Ruiz-Shulcloper, Weighted partition consensus via kernels, *Pattern Recognition* 43 (8) (2010) 2712-2724.
- [9] A. Strehl, J. Ghosh, Cluster ensembles-a knowledge reuse framework for combining multiple partitions, *Journal of Machine Learning Research* 3 (2002) 583-618.
- [10] G. Karypis, V. Kumar, Multilevel algorithms for multi-constraint graph partitioning, in: *Supercomputing '98: Proceedings of the 1998 ACM/IEEE Conference on Supercomputing (CDROM)*, IEEE Computer Society, Washington, DC, USA 1998, pp. 1-1.
- [11] K. Faceli, A. Carvalho, M. de Souto. Multi-objective clustering ensemble for gene expression data analysis, *Neurocomputing* 72(2009)2753-2774.
- [12] J. Handl, J. Knowles, D. Kell, Computational cluster validation in post-genomic data analysis, *Bioinformatics* 21 (15) (2005) 3201-3212.
- [13] A. Jain, R. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
- [14] R. Xu, D. Wunsch, Survey of clustering algorithms, *IEEE Transactions on Neural Networks* 16 (3) (2005) 645-678.
- [15] A.K. Jain, M.N. Murty, and P.J. Flynn, "Data clustering: A review," *ACM Computing Surveys* 31 (3) (1999), 264-323.
- [16] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, Wiley, second edition, 2001.
- [17] J. McQueen, Some methods of classification and analysis of multivariate observations, in: *5th Berkeley Symposium in Mathematics, Statistics and Probability*, 1967.
- [18] M.H. Law, A. Topchy, A. K. Jain, Multiobjective data clustering, in: *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2004, pp. 424-430.
- [19] J. Handl, J. Knowles, An evolutionary approach to multi-objective clustering, *IEEE Transactions on Evolutionary Computation* 11 (1) (2007) 56-76.
- [20] S. Thadjitodorov, L. IKuncheva, L. PTodorova, Moderate diversity for better cluster ensembles, *Information Fushion* 7(2006)264-275.
- [21] L.I. Kuncheva, S.T. Hadjitodorov, L.P. Todorova, Experimental comparison of cluster ensemble methods, in: *Proceedings of FUSION 2006*, 2006.
- [22] H.G. Ayad, M. SKmal, On voting-based consensus of cluster ensemble, *Pattern Recognition* 43(2010)1943-1953.
- [23] A.Jain, R. Dubes, *Algorithms for Clustering Data*, Prentice-Hall, EnglewoodCliffs, NJ, 1988.
- [24] G.W. Milligan, M.C. Cooper, A study of standardization of variables in cluster analysis, *Journal of Classification* 5 (1988) 181-204.
- [25] K. Deb, A. Pratap, S. Agarwal and T.Meyrivan, A fast and elitist multi-objective genetic algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation* 6(2) (2002), 182-197.
- [26] G.W. Milligan, M.C. Cooper, A study of the comparability of external criteria for hierarchical cluster analysis, *Multivariate Behavioral Research* 21 (1986) 441-458.
- [27] S.A. Armstrong, J.E. Staunton, L.B. Silverman, R. Pieters, M.L. den Boer, M.D. Minden, S.E. Sallan, E.S. Lander, T.R. Golub, S.J. Korsmeyer, MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia, *Nature Genetics* 30 (1) (2002) 41-47.
- [28] L. Dyrskjot, T. Thykjaer, M. Kruhoffer, J.L. Jensen, N. Marcussen, S. Hamilton-Dutoit, H. Wolf, T.F. Orntoft, Identifying distinct classes of bladder carcinoma using microarrays, *Nature Genetics* 33 (1) (2003) 90-96.
- [29] D. Chowdary, J. Lathrop, J. Skelton, K. Curtin, T. Briggs, Y. Zhang, J. Yu, Y.Wang, A. Mazumder, Prognostic gene expression signatures can be measured in tissues collected in RNA later preservative, *Journal of Molecular Diagnostics* 8 (1) (2006) 31-39.
- [30] G.J. Gordon, R.V. Jensen, L.L. Hsiao, S.R. Gullans, J.E. Blumenstock, S. Ramaswamy, W.G. Richards, D.J. Sugarbaker, R. Bueno, Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma, *Cancer Research* 62 (17) (2002) 4963-4967.
- [31] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard and M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J. R. Downing, M.A. Caligiuri, C.D. Bloomfield, E.S. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science* 286(5439)(1999)531-537.
- [32] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J.A. Olson, J. R.Marks, J.R. Nevins, Predicting the clinical status of human breast cancer by using gene expression profiles, *Proc Natl Acad Sci U S A* 98(20)(2001) 11462-11467.
- [33] P. Laiho, A. Kokko, S. Vanharanta, R. Salovaara, H. Sammalkorpi, H. Jarvinen, J-P. Mecklin, T.J. Karttunen, K. Tuppurainen, V. Davalos, S. Schwartz, D. Arango, M.J. Makinen, L.A. Aaltonen, Serrated carcinomas form a subclass of colorectal cancer with distinct molecular basis, 26(2) (2007)312-320.
- [34] Chen, Xin, Cheung, Siu Tim, So, Samuel and Fan, Sheung Tat, Barry, Christopher, Higgins, John, Lai, Kin-Man, Ji, Jiafu, Dudoit, Sandrine, Ng, Irene O.L, van de Rijn, Matt, Botstein, David, Brown, Patrick O, Gene Expression Patterns in Human Liver Cancers, 13(6)(2002) 1929-1939.
- [35] A. Alizadeh, M.B. Eisen, R. E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J.C. Boldrick, H. Sabet, T. Tran, X. Yu, J.I. Powell, L. Yang, G.E. Marti, T. Moore, J. Hudson, L. Lu, D.B. Lewis, R. Tibshirani, G. Sherlock, W.C. Chan, T.C. Greiner, D. D. Weisenburger, J.O. Armitage, R. Warnke, R. Levy, W. Wilson, M.R. Grever, J.C. Byrd, D. Botstein, P.O. Brown, L.M. Staudt, Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, 403(6769)(2000) 503-511.
- [36] Markus Bredel, Claudia Bredel, Dejan Juric, Griffith R Harsh, Hannes Vogel, Lawrence D Recht, Branimir I Sikic, Functional network analysis reveals extended gliomagenesis pathway maps

and three novel MYC-interacting genes in human gliomas, *Cancer Res*, 65(19)(2005) 8679-8689.

- [37] E.J. Yeoh, M.E. Ross, S.A. Shurtleff, W.K. Williams, D. Patel, R. Mahfouz, F.G. Behm, S.C. Raimondi, M.V. Relling, A. Patel, C. Cheng, D. Campana, D. Wilkins, X. Zhou, J. Li, H. Liu, C.-H.

Pui, W.E. Evans, C. Naeve, L. Wong, J.R. Downing, Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling, *Cancer Cell* 1 (2) (2002) 133–143.