Evolutionary Multiobjective Clustering and Its Applications to Patient Stratification

Xiangtao Li and Ka-Chun Wong

***Abstract*—Patient stratification has a major role in enabling efficient and personalized medicine. An important task in patient stratification is to discover disease subtypes for effective treatment. To achieve this goal, the research on clustering algorithms for patient stratification has brought attention from both academia and medical community over the past decades. However, existing clustering algorithms suffer from realistic restrictions such as experimental noises, high dimensionality, and poor interpretability. In particular, the existing clustering algorithms usually determine clustering quality using only one internal evaluation function. Unfortunately, it is obvious that one internal evaluation function is hard to be fitted and robust for all datasets. Therefore, in this paper, a novel multiobjective framework called multiobjective clustering algorithm by fast search and find of density peaks is proposed to address those limitations altogether. In the proposed framework, a parameter candidate population is evolved under multiple objectives to select features and evaluate clustering densities automatically. To guide the multiobjective evolution, five cluster validity indices including compactness, separation, Calinski–Harabasz index, Davies–Bouldin index, and Dunn index, are chosen as the objective functions, capturing multiple characteristics of the evolving clusters. Multiobjective differential evolution algorithm based on decomposition is adopted to optimize those five objective functions simultaneously. To demonstrate its effectiveness, extensive experiments have been conducted, comparing the proposed algorithm with 45 algorithms including nine state-of-the-art clustering algorithms, five multiobjective evolutionary algorithms, and 31 baseline algorithms under different objective subsets on 94 datasets featuring 35 real patient stratification datasets, 55 synthetic datasets based on a real human transcription regulation network model, and four other medical datasets. The numerical results reveal that the proposed algorithm can achieve better or competitive solutions than the others. Besides, time complexity analysis, convergence analysis, and parameter analysis are conducted to demonstrate the robustness of the proposed algorithm from different perspectives.**

***Index Terms*—Clustering, density peaks, multiobjective optimization, patient stratification.**

Manuscript received December 2, 2017; accepted March 16, 2018. Date of publication April 2, 2018; date of current version March 5, 2019. This work was supported in part by the Research Grants Council of the Hong Kong Special Administrative Region under Grant CityU 21200816 and Grant CityU 11203217, in part by the National Natural Science Foundation of China under Grant 61603087, in part by the Natural Science Foundation of Jilin Province under Grant 20160101253JC, and in part by the Fundamental Research Funds of Northeast Normal University under Grant 2412017FZ026. This paper was recommended by Associate Editor S. Mostaghim. *(Corresponding author: Ka-Chun Wong.)*

X. Li is with the Department of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, China (e-mail: lixt314@nenu.edu.cn).

K.-C. Wong is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: kc.w@cityu.edu.hk).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the author.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2018.2817480

I. Introduction

PATIENT stratification is a key challenge for modern medicine to guarantee drug development success until clinical validation. In recent years, the next-generation sequencing technologies have permitted us to sequence massive amounts of different molecular datasets, which have revolutionized the related studies [1]. In particular, those large volumes of sequencing data give us great opportunities to address ground challenges in cancer subtype research [2], since cancer subtypes are connected with different mutations, gene expression profiles, molecular signatures as well as clinical phenotypes [3]. Thus, the goal for patient stratification is to develop efficient computational methods to stratify the patients into subtypes for precise clinical decision making.

In the past, a plethora of data clustering approaches have been proposed for analyzing patient stratification or disease subtyping; for instance, Diaz *et al.* [4] introduced a disease subtyping pipeline, which combines a feature selection approach with well-documented clustering methods. Khakabimamaghani and Ester [5] proposed a Bayesian biclustering method to compare the patient stratification datasets. Chang *et al.* [6] suggested gene interaction networks to enhance the performance of clustering algorithms on gene expression data for subtype-specific blood gene expression signatures. Graim *et al.* [7] introduced a framework for patient stratification, implementing a community detection method to select subtypes out of sparse patient measurements. Wu *et al.* [8] employed unsupervised consensus clustering to analyze robust imaging subtypes and evaluate their clinical relevance. Jaskowiak *et al.* [9] presented fifteen distance metrics along with four clustering algorithms for gene expression microarray datasets. Liu *et al.* [10] proposed an entropy-based consensus clustering (ECC) method, which employed an entropy-based utility function to combine many primary partitions to a consensus one. Unfortunately, most clustering algorithms consider all features to be equally important for clustering. In actuality, various features have diverse effects on clustering. It is one of the reasons why most clustering algorithms may not perform well in the face of high-dimensional molecular data [11], [12]. Therefore, different weights should be assigned to features to quantify their importance. Moreover, for most of the existing clustering algorithms, only one internal evaluation function is considered. It is another reason why

existing clustering algorithms still cannot perform well across multiple datasets.

To quantify the effect of different features, distance weights are often combined into the features. Many weight clustering algorithms have been proposed by integrating the distance weights into the traditional clustering algorithms such as. Yu *et al.* [13] employed the maximum entropy principle to determine these weights automatically for clustering. Nock and Nielsen [14] introduced a generic iterative clustering scheme that is coupled with reweighting scheme to bring improvements over traditional clustering. Modha and Spangler [15] employed an efficient framework for mixing multiple feature weights into the *k*-means (KM) clustering algorithm. De Amorim and Mirkin [16] proposed two modifications of weighted KM clustering algorithm and their competitiveness is experimentally demonstrated. Jing *et al.* [17] suggested a clustering method to solve the problem of clustering complex text data based on a subspace clustering algorithm by computing the feature weights in the KM clustering algorithm. Güneş *et al.* [18] suggested a data preprocessing method called KM clustering-based feature weighting, which connects the *k*-nearest neighbor and decision tree to classify the electroencephalogram sleep. Wang and Wang [19] showed a clustering method in the framework of granular computing that incorporates fuzzy sets, rough sets, and estimates feature weights automatically. Huang *et al.* [20] proposed a KM type clustering algorithm that can automatically compute variable weights; it involves iteratively update variable weights based on data partitioning. Therefore, designing suitable distance weights can mitigate the existence of noisy or redundant features, revealing the actual natural patterns.

Meanwhile, for most of those clustering algorithms, it can be found that they optimize only one cluster validity measure to evolve the clustering algorithm. Given different data properties, a single validity measure usually does not work equally well for all datasets. Therefore, it should be beneficial to optimize multiple validity indices simultaneously to capture different characteristics of the datasets. In the past decade, many multiobjective evolutionary algorithms have been proposed for solving the clustering problem; for instance, Handl and Knowles [21] described an evolutionary approach for multiobjective clustering, which uses the overall deviation and cluster connectedness as the objective functions. Law *et al.* [22] presented a clustering method that uses multiple clustering objective functions simultaneously, which is a two-step process. Maulik *et al.* [23] proposed a multiobjective clustering method to obtain the final solution while considering all Pareto-optimal solutions. Mukhopadhyay *et al.* [24] introduced a multiobjective genetic algorithm-based approach for fuzzy clustering of categorical data, which encodes the cluster modes and simultaneously optimizes fuzzy compactness (CP) and fuzzy separation (Sep) of the clusters. As mentioned, most of the multiobjective clustering algorithms are based on multiobjective genetic algorithms. However, there are some other evolutionary algorithms, such as DE, simulated annealing (SA), particle swarm optimization (PSO), and ant colony optimization (ACO), that have also been employed for multiobjective clustering. Suresh *et al.* [25] developed multiobjective variants of DE on the fuzzy clustering problem, in which two conflicting fuzzy validity indices were optimized at the same time. Armano and Farmani [26] proposed a multiobjective clustering PSO, which can automatically discover the optimal number of clusters. Saha and Bandyopadhyay [27] proposed an SA-based multiobjective optimization method to identify the appropriate amount of clusters as well as the appropriate partitioning. İnkaya *et al.* [28] proposed a clustering algorithm based on ACO, which uses the multiobjective setting for spatial clustering problem. From the above introduction, multiobjective evolutionary clustering algorithm currently enjoys growing interests in the optimization community. However, the field of multiobjective clustering is still in its infancy; future research works are still necessary to develop multiobjective algorithms for clustering.

Recently, a novel clustering by fast search and find of density peaks (CDPs) [29] is proposed based on the idea that cluster centers are identified by a higher density than their neighbors and by a relatively large distance from points with higher densities. This algorithm has been successfully employed in clustering molecular data [30], [31]. However, when analyzing high-dimensional molecular biology data, the algorithm still suffers from the curse of dimensionality because biological experiments often involve multiple conditions and lots of different observed features. As we know, in CDP, the algorithm computes the Euclidean distance to measure feature components in the beginning. With the increase in dimensions, there are less differences in the distances between different pairs of samples. It will lead to the difficulties in computing the effective local density ρ_i and distance δ_i from points of high density. One way to solve the high-dimensional feature space is to assign the weights to features to quantify their importance. Meanwhile, the selection of cutoff distance d_c is introduced based on a heuristic method that the average number of neighbors in a dataset should only be 1%–2% of the entire dataset. However, for different datasets, it needs different values for d_c. Different d_c values also affect clustering results. Therefore, to address those limitations altogether, we propose a novel multiobjective framework based on the CDP, which is called multiobjective CDPs (MOCDPs) for clustering patient stratification data. To alleviate the shortcomings, the population combines instance dissimilarity weights and the related cutoffs to select features and evaluate clustering densities automatically. Five cluster validity indices are selected as the objective functions for intrinsic clustering measurement. Finally, multiobjective DE algorithm based on decomposition (MOEA/D-DE) is adopted to optimize those five objective functions simultaneously to stratify the patients into subtypes in a robust manner.

II. Methods

A. Clustering by Fast Search and Find of Density Peaks

CDPs [29] is a new algorithm that intends to classify elements by their similarity, which has been successfully employed and published on *Science* in 2014. CDP is based on two assumptions that the cluster center is a highly dense data

截断距离是一个启发式的距离，不应该固定他的大小，这个距离一边选择是：让所有的Population中的1%~2%的点落在圈里面

region compared with its neighbors and it is placed at a relatively high distance from other cluster centers. For each data point i, the local density ρ_i and distance δ_i from neighboring high-density data point are calculated as follows.

1) Cut-off Kernel

$$\rho_i = \sum_j \varphi(d_{ij} - d_c) \tag{1}$$

where

$$\varphi(d) = \begin{cases} 1 & d < 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

一个选出聚类中心的算法

where the results of the cut-off kernel are discrete. d_c is the cutoff distance and d_{ij} is the distance between two samples i and j. ρ_i is equal to the number of data points that are closer than d_c to i. The δ_i is determined by finding the minimum distance d_{ij} between the points i and j with the high density. The δ_i can be described as follows:

$$\delta_i = \min_{j:\rho_j > \rho_i} (d_{ij}). \tag{3}$$

Then, CDP takes $\delta_i = \max_j(d_{ij})$ [29] for the point with the highest density. Based on the above, δ_i is much higher than the typical nearest neighbor distance just for points that are local or global maxima in the density. The cluster centers occupy large δ and ρ as compared with other points.

B. Motivation

In the original CDP algorithm, it has two important input parameters: 1) dissimilarity metric and 2) the cutoff distance (d_c). For dissimilarity metric, the CDP uses the Euclidean distance to measure feature components. A major problem with this method is the selection of features. From the CDP algorithm, it cannot select features automatically because the algorithm employs all features fairly and arranges weights of all features equally when evaluating dissimilarity in the clustering process. However, it is unrealistic to assume that all features are equally related to different clustering problems. In fact, many studies indicate that efficient clustering algorithms often happen using feature subsets. Moreover, some features in the original data could be proved to be irrelevant, redundant or even detrimental to clustering. These observations prompt a lot of interests in creating a new feature space over a large number of data. In this paper, we first propose a weighted dissimilarity metric in CDP concerning the following distance function:

对于不同的特征都是平等对待的，这样不好。需要加权

比如计算两个点的距离时候，x要乘以0.7，y要乘以0.3，但是这里的权重是多少合适呢？

$$d_{i,j} = \sqrt{\sum_{l=1}^{D} \omega_l (x_{i,l} - x_{j,l})^2} \tag{4}$$

where $d_{i,j}$ is the distance between x_i and x_j with D-dimensional data and $\omega = \{\omega_1, \omega_2, \ldots, \omega_l\}$ is a weight vector, which is non-negative and $\sum_{l=1}^{D} \omega_l = 1$. From this equation, the distance between two samples is the weighted sum of the differences between two samples. Since no prior information about the sample weights is given, we employ the multiobjective method to optimize the weight to enhance the clustering performance. Meanwhile, the selection of d_c is introduced based on a heuristic method that the average number of neighbors in a dataset should only be 1%–2% of the entire dataset. However, for different datasets, it needs different values for d_c. Different d_c values also affect clustering results. Therefore, we add the d_c as another variable in the weight vector to be optimized.

Another challenge is that CDP usually determines its clustering quality using only one internal evaluation function that is optimized. However, one internal evaluation function cannot perform equally well for all datasets due to different characteristics of the datasets. Therefore, in this paper, we design multiple internal evaluation functions to be optimized simultaneously.

C. Multiobjective Clustering by Fast Search and Find of Density Peaks

In this section, we propose the MOCDP based on decomposition for clustering problems in the context of patient stratification. MOCDP utilizes a decomposition-based framework which is similar to that of the multiobjective evolutionary algorithm based on decomposition (MOEA/D) [32]. This algorithm starts a population with the weights of dissimilarity and cut-off distance for the patient stratification data. After that, it decomposes the proposed multiobjective clustering problem into some scalar subproblems using the tchebychev approach with a set of uniformly distributed weight vectors λ for objective functions. After that, each subproblem is optimized by the MOEA/D-DE using the information only from its neighboring subproblems. Each subproblem has its best solution found so far in the population. Then, the new individual is compared with the original solution with the same weight vector λ and its neighboring subproblems.

1) Multiobjective Optimization: Multiobjective optimization problems consider optimization problems concerning various objective functions, which should be optimized together and that has the conflict with each other [33]. Mathematically, multiobjective optimization problem can be expressed as follows:

$$\begin{aligned} \min \ & f_i(x), i = 1, \ldots, M, x \in X \\ \text{s.t.} \ & g_j(x) \le 0, j = 1, \ldots, J \\ & h_e(x) = 0, e = 1, \ldots, E \end{aligned} \tag{5}$$

where x is a D-dimensional vector having D decision variables. The $g_j(x)$ is the inequality constraint and the $h_e(x)$ is the equality constraint. $f_i(x)$ is the ith objective function and the M is the number of objective functions. Two important concepts of the multiobjective optimization problem are the *dominance* and *Pareto-optimality* [35], [50]. For a minimization problem, these two concepts can be defined as follows.

1) *Definition 1 (Dominance):* A feasible solution $\overrightarrow{x^1} \in X$ is said to *dominate* feasible solution $\overrightarrow{x^2} \in X$ if $f_i(\overrightarrow{x^1}) \le f_i(\overrightarrow{x}^2)$ for all $i \in \{1, \ldots, M\}$ and $f_j(\overrightarrow{x^1}) < f_j(\overrightarrow{x}^2)$ for at least one j, $j \in \{1, \ldots, M\}$.
2) *Definition 2 (Pareto-Optimality):* Based on the definition of *dominance*, the solution $\overrightarrow{x}^*$ is called

Pareto-optimality if it does not exist other solutions to dominate it.

Following that, the set of all Pareto optimal solutions is termed Pareto set; and the set of all Pareto objectives vectors is called Pareto front which depicts the tradeoff curves among multiple competing objectives.

2) Objective Functions: One of the primary research of multiobjective evolutionary algorithm for clustering problem under the context of patient stratification is the choice of suitable objective functions that are to be optimized simultaneously. Without loss of generality, most of the existing multiobjective evolutionary algorithms for clustering problem usually employed different combinations of two cluster validity indices as the objective functions to be optimized simultaneously. Moreover, some multiobjective evolutionary algorithms for clustering problems have also employed more than two cluster validity indices to include multiple characteristics of the evolved clusters, because more objective functions in multiobjective evolutionary algorithms for clustering problem denote more multiple categories of validity indices. In this paper, we propose to consider the clustering problem for patient stratification concerning five objectives functions to make the clustering algorithm to have the ability to learn clustering properties without the ground truth labels. These five objective functions are CP [36], Sep [37], Calinski–Harabasz (CH) index [38], Davies–Bouldin (DB) index [38], and Dunn index [39]. These five objective functions are described in detail in Section S1 in the supplementary material.

3) Multiobjective Algorithm for Clustering Problem: In the previous section, we present five objective functions as the measurement criteria to estimate the clustering problem by MOEA/D-DE. In this section, we propose to combine the multiobjective evolutionary algorithm framework and CDPs to generate the solutions based on those objective functions for clustering the patient stratification datasets. In our algorithm, the tchebychev approach is utilized to decompose the multiobjective clustering problem into a number of scalar clustering optimization subproblems. To be specific, a weight vector $\lambda = \{\lambda_1, \ldots, \lambda_M\}$ with $\sum_{i=1}^{M} \lambda_i = 1$ and $\lambda_i \geq 0$ for all $i = \{1, \ldots, M\}$, is applied. The objective function of a subproblem is stated as follows:

$$g^{te}(x \mid \lambda, z^*) = \max_{1 \leq i \leq M} \left\{ \lambda_i \left| f_i(x) - z_i^* \right| \right\} \tag{6}$$

where $z^* = \{z_1^*, z_2^*, \ldots, z_M^*\}$ is the reference point [that is $z_i^* = \min(f_i(x))$] and z_i^* is the best value found so far for $f_i(x)$ [32]. The objective function of the ith subproblem is $g^{te}(x \mid \lambda_i)$. Suppose there is N weight vector $\{\lambda^1, \ldots, \lambda^N\}$, the patient stratification clustering problem can be decomposed into N single objective patient stratification clustering problems. The detailed steps of the decomposition-based MOCDP framework are described in Algorithm 1.

As outlined in Algorithm 1, MOCDP keeps a population $P = \{P_1, P_2, \ldots, P_N\}$ including the weight of dissimilarity (the dimension of this weight is D) and the cutoff distance (d_c) in each iteration. Then, N subpopulations are built; each of the subpopulation corresponds to a subproblem. After that, each subproblem is associated with different weight vectors λ and then for each $i = \{1, 2, \ldots, N\}$, $B(i) = \{i_1, \ldots, i_T\}$ is defined according to the Euclidean distance between its weight vector and the other weight vectors and then find T closest weight vectors. Next, those five objective functions are computed based on the intermediate clustering results of CDP. At each iteration, the population is first updated by the mutation and crossover operators of DE algorithm based on two parents randomly selected from the current population. In the

Algorithm 1: Pseudocode of MOCDP

Input: input parameters:
1) A patient stratification clustering dataset and the number of samples (S);
2) A stopping criteria;
3) the number of subproblem is N (population size);
4) N weight vector $\{\lambda^1, \cdots, \lambda^N\}$;
5) the size of the neighborhood of each subproblem, denoted as T;

Output: output result:
1) Normalized Mutual Information (NMI);
2) Normalized Rand Index (R_n);

Initialize N individuals to form a population $P = \{P_1, P_2, \cdots, P_N\}$ including the weight of dissimilarity (the dimension of this weight is D) and the cutoff distance (d_c), where $p_{k,j}$ indicates the jth element of P_k.
For each $i = \{1, 2, \cdots, N\}$, $B(i) = \{i_1, \cdots, i_T\}$ is defined according to the Euclidean distance between its weight vector and other weight vectors and then find T closest weight vectors.
Determine those five objective functions for all subproblems based on the tchebycheff approach.
Initial the reference point $z^* = \{z_1^*, z_2^*, \cdots, z_M^*\}$ ($z_i^* = \min(f_i(x))$);
while *a stopping criteria is not satisfied* **do**
  for $k = 1 \rightarrow N$ **do**
    Randomly generate three different indexes r_1, r_2, $r_1 \neq r_2 \neq k$;
    Generate $j_{rand} \in [1, \ldots, D+1]$ is the randomly chosen index.
    for $j = 1 \rightarrow (D+1)$ **do**
      if $rand < CR | j_{rand} = j$ **then**
        $U_{k,j} \leftarrow p_{k,j} + F \cdot (p_{r_1,j} - x_{r_2,j})$;
      else
        $U_{k,j} \leftarrow p_{k,j}$;
    Generate the distance $d(x_i, x_j)$ for every pair (i, j) of samples based on the U_k.
    /*Using cut-off kernel to computing local density ρ and distance δ*/
    $d_c \leftarrow U_{k,D+1}$;
    $\rho \leftarrow \sum_j \varphi(d_{ij} - d_c)$;
    $\delta \leftarrow \min_{j:\rho_j > \rho_i}(d_{ij})$;
    Calculate those five objective functions for each subproblem U_k based on the tchebycheff approach.
    For each objective $m = \{1, \cdots, M\}$, if $z_m^* > f_m(U_k)$, then set $z_m^* = f_m(U_k)$.
    for $j \in B_k$ **do**
      if $g^{te}(U_k | \lambda^k) < g^{te}(P_j | \lambda^k)$ **then**
        $P_j \leftarrow U_k$;
        $f(P_j) \leftarrow f(U_k)$;
  Return the Pareto optimal set $\hat{P}$ including all nondominated individuals with respect to those five objective functions.
Return the best Normalized Mutual Information value NMI$\leftarrow max_{i \in \hat{P}}(\text{NMI}_i)$ and the best Normalized Rand Index R_n.

TABLE I
KEY CHARACTERISTICS OF 35 BENCHMARK DATASETS FOR CLUSTER VALIDITY

No.	Datatset	Tissue	Subject	Gene	Class	No.	Datatset	Tissue	Subject	Gene	Class
1	Alizadeh-2000-v1	Blood	42	1095	2	19	Lapointe-2004-v2	Prostate	110	2496	4
2	Alizadeh-2000-v2	Blood	62	2093	3	20	Liang-2005	Brain	37	1411	3
3	Alizadeh-2000-v3	Blood	62	2093	4	21	Nutt-2003-v1	Brain	50	1377	4
4	Armstrong-2002-v1	Blood	72	1081	2	22	Nutt-2003-v2	Brain	28	1070	2
5	Armstrong-2002-v2	Blood	72	2194	3	23	Nutt-2003-v3	Brain	22	1152	2
6	Bhattacharjee-2001	Lung	203	1543	5	24	Pomeroy-2002-v1	Brain	34	857	2
7	Bitter-2000	Skin	38	2201	2	25	Pomeroy-2002-v2	Brain	42	1379	5
8	Bredel-2005	Brain	50	1739	3	26	Ramaswamy-2001	Multi-tissue	190	1363	14
9	Chen-2002	Liver	179	85	2	27	Risinger-2003	Endometrium	42	1771	4
10	Chowdary-2006	Multi-tissue	104	182	2	28	Shipp-2002	Blood	77	798	2
11	Dyrskjot-2003	Bladder	40	1203	3	29	Singh-2002	Prostate	102	339	2
12	Garber-2001	Lung	66	4553	4	30	Su-2001	Multi-tissue	174	1571	10
13	Golub-1999-v1	Bone marrow	72	1868	2	31	Tomlins-2006-v1	Prostate	92	1288	4
14	Golub-1999-v2	Bone marrow	72	1868	3	32	Tomlins-2006-v2	Prostate	104	2315	5
15	Gordon-2002	Lung	181	1626	2	33	West-2001	Breast	49	1198	2
16	Khan-2001	Multi-tissue	83	1069	4	34	Yeoh-2002-v1	Bone marrow	248	2526	2
17	Laiho-2007	Colon	37	2202	2	35	Yeoh-2002-v2	Bone marrow	248	2526	6
18	Lapointe-2004-v1	Prostate	69	1625	3						

DE algorithm, the scale factor (F) is the mutation scale factor which is used in controlling the strength of differential variation. The crossover rate (CR) controls which and how many components are interchanged in each element of the current population. After that, the offspring individual U is generated. It is noted that only one individual is produced for each subproblem. Then, we adopt the offspring individual as the weight of dissimilarity to recalculate the distance for the patient stratification dataset. The last column of offspring individual is the cutoff distance (d_c). After obtaining the dissimilarity of data and d_c, CDP is implemented to cluster the data based on two assumptions including neighbors with lower local density and that they are at a relatively large distance from any points with a higher local density. After the cluster centers are found, each remaining point is assigned to the same group as its nearest neighbor of higher density. After that, the proposed objective functions can be determined. Then, for each objective $j = \{1, \ldots, M\}$, if $z_j^* > f_j(U_k)$, then set $z_j^* = f_j(U_k)$. Finally, all the neighbors of each subproblem are considered to compare with the current subproblem to find better solutions. For example, if U_k performs better than P_j, the individual P_j is replaced by the offspring individual U_k. By repeating this procedure of all subproblems, each offspring individual is compared with the current subproblem and its neighbors in the original population. Finally, a Pareto optimal set $\hat{P}$ including all nondominated individuals can be obtained.

D. Time Complexity Analysis

In this section, we focus on the time complexity of proposed method. At the beginning of MOCDP algorithm, we first calculate the weight dissimilarity of the data, costing $O(D * (D-1)/2)$, where D is the number of features in the original data. Then, the algorithm randomly selects two solutions for mutation and crossover operators and generates the reference point, costing $O(M)$. Then, the offspring individual is applied to recalculate the distance for the patient stratification clustering dataset as the weight of dissimilarity. The last column of offspring individual is the cutoff distance (d_c). After that, the CDP is used to find the cluster center, costing $O(S^2)$, where S is the number of samples in each patient stratification dataset. For all subproblems, the time complexity is $O(N \cdot (S^2 + D * (D-1)/2))$. After that, five objective functions are calculated. To use the neighborhood information to update the population, it needs $O(MT)$ underlying operations since its major costs lie in the M calculations for T solutions, where T is the number of the neighborhood for each subproblem. In summary, the computational complexity for the neighborhood updating is $O(M \cdot N \cdot T)$. Therefore, the overall time complexity of MOCDP with DE algorithm is $O(N \cdot (S^2 + D * (D-1)/2) + M \cdot N \cdot T)$.

III. EXPERIMENTS

A. Data Sources

In this paper, we employ 35 patient stratification datasets as the benchmark datasets. All these benchmark datasets with label information are obtained from [10]. Table I summarizes these 35 benchmark patient stratification datasets [40]. As described in Table I, the number of samples is varied from 22 to 248 for all experiment datasets; the number of the feature in the original data is varied from 85 to 4553, and the number of clusters is ranged from 2 to 14. From Table I, we also can find that some datasets use the same data source, such as Alizadeh-2000-v2 and Alizadeh-2000-v3 are from the same source with different number of clusters; Golub-1999-v1 and Golub-1999-v2 also share the same source; Yeoh-2002-v1 and Yeoh-2002-v2 are similar; Armstrong-2002-v1 and Armstrong-2002-v2 have the same number of samples; Lapointe-2004-v2 has one more cluster than the Lapointe-2004-v1; and Tomlins-2006-v2 and Tomlins-2006-v1 also have the same data characteristics with Lapointe-2004-v2 and Lapointe-2004-v1.

B. Parameter Setting

In the above section, 35 patient stratification datasets are adopted for the performance comparison. For the MOCDP, the initial population $P = \{P_1, P_2, \ldots, P_N\}$ is randomly produced

within the boundary space. In generating a set of weight vectors, each component of a weight vector takes the values from $\{(0/H), (1/H), \ldots, (H/H)\}$, where H is an integer. Therefore, the number of weight vectors is $N = C_{H+M-1}^{M-1}$, where M is the objective number of our proposed problem and H is set to 5. In this situation, the distributed weight vectors N is 126. The number of the weight vectors in the neighborhood of each weight vector T is 2 (discussion in Section III-J), and the CR is 0.2 and the F is 0.4 (discussion in Section III-J). To provide a reasonable comparison, we set the number of fitness evaluations as the termination criteria instead of the number of generations or CPU times. We set $10N$ objective function evaluation (FE) (discussion in Section III-I) for each dataset. Meanwhile, the proposed algorithms are run for 30 independent times on each dataset for statistical significance. Then, we compute the average result of 30 independent runs and analyze the results on each patient stratification dataset.

C. Other Related Methods From Literature

Several existing clustering methods including KM clustering method [41], agglomerative hierarchical clustering with average-linkage (AL), single-linkage (SL), and complete-linkage (CL) [42], density-based spatial clustering of applications with noise (DBSCAN) [43], link-based cluster ensemble (LCE) [44], approximate SimRank-based (ASRS) methods [45], ECC [10], and CDPs [29], have been adopted here. Those algorithms are chosen to indicate different algorithmic paradigms; for example, KM clustering method, Agglomerative hierarchical clustering with AL, SL, and CL are four simple clustering techniques, which are usually employed by clinical researchers to analyze the data. LCE is the ensemble algorithm. ASRS methods is an efficient variation of the SimRank-based similarity counterpart. ECC uses an entropy-based utility function to fuse many primary partitions to a consensus one that agrees with the basic ones as much as possible. DBSCAN is a density-based clustering algorithm. CDPs discover the centers of clusters by finding of the density peaks.

In other aspect, to demonstrate the effectiveness of multiobjective algorithm, we also compare our proposed algorithm MOCDP with several state-of-the-art multiobjective evolutionary algorithms including nondominated sorting genetic algorithm II (NSGAII) [46], grid-based evolutionary algorithm (GrEA) [47], hypervolume estimation algorithm (HyEA) for multiobjective optimization [48], strength Pareto evolutionary algorithm 2 (SPEA2) [49], and multiobjective PSO (MOPSO) [50]. Those algorithms indicate different algorithmic paradigms to compare with our proposed algorithm. From the perspective of multiobjective evolutionary algorithms, our proposed algorithm MOCDP is a multiobjective algorithm based on decomposition. NSGAII uses the nondominated sorting method and crowding distance strategy. GrEA is a grid-based evolutionary algorithm. HyEA is hypervolume-based evolutionary many-objective optimization algorithm. SPEA2 uses the density estimation technique and enhanced archive truncation method. MOPSO is a multiobjective algorithm with PSO.

For statistically rigorous comparisons, the paired Wilcoxon's signed rank test is computed to perform statistically significant testing between pairs of algorithms. The pair Wilcoxon's signed-rank test is a typical nonparametric statistical hypothesis test method, where three symbols are used: "+," "−," and "≈." The ≈ denotes that there is not any significant difference between the two compared algorithms. The + denotes that our algorithm MOCDP is better than other algorithms. The − indicates our proposed algorithm is inferior to other algorithms.

D. Evaluation Metrics

The clustering results are evaluated by comparing the obtained label of each sample using clustering algorithms with the ground truth label. We adopt the normalized mutual information (NMI) and normalized rand index (R_n) to measure the performance of all compared algorithms. These two metrics are employed to calculate the similarity between the cluster labels and the truth labels. From the results, the clusters with higher values represent better partitions than other results.

NMI is a normalization of the mutual information (MI) score ranged from 0 (no MI) to 1 (perfect correlation), which can be expressed as follows:

$$\mathrm{NMI} = \frac{\sum_{i,j} n_{i,j} \log \frac{n \cdot n_{i,j}}{n_{i+} \cdot n_{+j}}}{\sqrt{\left(\sum_i n_{i+} \log \frac{n_{i+}}{n}\right)\left(\sum_j n_{j+} \log \frac{n_{+j}}{n}\right)}}. \tag{7}$$

Normalized rand index is recommended as the index of choice for measuring agreement between two partitions, which can be computed as follows:

$$R_n = \frac{\sum_{i,j}\binom{n_{i,j}}{2} - \sum_i \binom{n_{i+}}{2} \cdot \sum_j \binom{n_{+j}}{2} \Big/ \binom{n}{2}}{\sum_i \binom{n_{i+}}{2} \Big/ 2 + \sum_j \binom{n_{+j}}{2} \Big/ 2 - \sum_i \binom{n_{i+}}{2} \cdot \sum_j \binom{n_{+j}}{2} \Big/ \binom{n}{2}}. \tag{8}$$

E. Evaluation on Benchmark Patient Stratification Data

In this section, we have employed our proposed algorithm MOCDP to stratify patients into different groups. Thirty-five patient stratification datasets are used to test the performance of MOCDP. We run 30 independent times on each dataset. MOCDP is compared with nine different clustering methods including KM clustering method, Agglomerative hierarchical clustering with AL, SL, and CL, DBSCAN, LCE, ASRS methods, ECC, and CDPs. The experimental results are tabulated in Tables II and III. The last two rows of Tables II and III summarize the statistical results.

1) For the evaluation metrics NMI, MOCDP is the best-performing algorithm. Compared with CDP and MOCDP, we can find that MOCDP is inferior to, equal to and superior to CDP in 0, 1, and 34 patient stratification data, respectively. It demonstrates that multiobjective clustering can enhance the performance of CDP. Meanwhile, the results of Wilcoxon's signed-rank test in the last two rows indicate that MOCDP is significantly better than KM, AL, SL, CL, DBSCAN, LCE, ASRS, ECC, and CDP on 32, 35, 35, 34, 27, 29, 29, 26, and

TABLE II
PERFORMANCE OF DIFFERENT CLUSTERING ALGORITHMS ON BENCHMARK DATASETS BY NMI. FOR EACH ALGORITHM, A WILCOXON'S SIGNED-RANK TEST IS CONDUCTED TO VERIFY WHETHER THE EXPERIMENT RESULTS OF BEST ALGORITHMS ARE BETTER THAN OTHER ALGORITHMS. IF p-VALUE IS LESS THAN 0.05, IT DENOTES THAT THE BEST ALGORITHM OF INTEREST PERFORMS SIGNIFICANTLY BETTER THAN OTHER ALGORITHMS

Dataset	KM	AL	SL	CL	DBSCAN	LCE	ASRS	ECC	CDP	MOCDP
1	0.1108	0.0939	0.0601	0.0939	0.3373	0.4167	0.3516	0.4940	0.4167	**0.5647**
2	0.5663	0.8119	0.0342	0.6972	0.1752	0.5672	0.5877	0.5832	**1.0000**	**1.0000**
3	0.6452	0.6421	0.0960	0.5665	0.1620	0.5300	0.6441	0.6340	0.6656	**0.7729**
4	0.3459	0.0631	0.0263	0.4389	0.4755	0.3459	0.3622	0.3571	0.2111	**0.7128**
5	0.7154	0.1311	0.0780	0.5536	0.5737	**0.7654**	0.7460	0.8154	0.2615	0.6779
6	0.3904	0.4039	0.1014	0.5032	0.3766	0.4186	**0.4941**	0.4585	0.2695	0.4415
7	0.0083	0.0640	0.0640	0.0000	**0.2964**	0.0323	0.0323	0.0502	0.0557	0.2919
8	0.4044	0.0848	0.0848	0.3685	0.5156	0.3688	0.4769	0.3712	0.3886	**0.7292**
9	0.0008	0.0199	0.0199	0.0008	0.0698	0.4980	0.0010	**0.6273**	0.0006	0.4956
10	0.1424	0.0459	0.0459	0.0459	0.1245	0.4739	0.1424	0.8098	0.0459	**0.8597**
11	0.5158	0.1613	0.1613	0.1613	0.3095	0.5743	0.5158	0.5584	0.2672	**0.5868**
12	0.2116	0.0734	0.0734	0.0999	0.4116	0.1647	0.1294	0.2025	0.1400	**0.5045**
13	0.5536	0.2764	0.0684	0.4837	0.4280	0.6085	0.5536	**0.8149**	0.0008	0.7408
14	0.4859	0.2868	0.0803	0.3496	0.3502	0.6870	0.8453	**0.7737**	0.0493	0.6860
15	0.1385	0.0083	0.0083	0.1168	0.1119	0.1411	0.1099	0.1479	0.1583	**0.7025**
16	0.4652	0.0966	0.0941	0.4857	0.5890	0.5109	0.5644	0.6221	0.4901	**0.7950**
17	0.2224	0.0340	0.1680	0.0664	**0.4473**	0.1959	0.0749	0.0768	0.0143	0.4323
18	0.0654	0.0466	0.0466	0.0466	**0.2613**	0.1101	0.1826	0.1281	0.0498	0.2479
19	0.1411	0.0513	0.0406	0.0419	0.1980	0.1357	0.1379	0.1652	0.1374	**0.2549**
20	0.3545	0.3002	0.1181	0.3545	**0.6084**	0.3772	0.3545	0.3545	0.1588	0.5743
21	**0.5083**	0.1293	0.1296	0.2900	0.5065	0.3844	**0.5083**	0.4994	0.2436	0.4443
22	0.1009	0.0406	0.0778	0.2513	**0.4901**	0.3988	0.0406	0.4021	0.1661	0.2968
23	0.7523	0.1600	0.1600	0.1600	0.6432	0.7523	**1.0000**	0.7523	0.2099	0.5556
24	0.0219	0.0521	0.0332	0.0079	**0.4005**	0.0079	0.0079	0.0092	0.1932	0.1996
25	0.5020	0.4150	0.4177	0.5396	0.4179	0.6032	0.5937	0.6171	0.3750	**0.6467**
26	0.5149	0.2176	0.1608	0.3376	0.1953	0.5126	0.5560	0.5277	0.5279	**0.5891**
27	0.2709	0.1349	0.1203	0.1877	0.3906	0.2784	0.2709	0.2911	0.1413	**0.4275**
28	0.0924	0.0288	0.0288	0.0288	0.2435	0.1235	0.1124	0.1280	0.0288	**0.3520**
29	0.0475	0.0675	0.0360	0.0675	0.0953	0.0101	0.0342	0.0100	0.0780	**0.2530**
30	0.5711	0.4919	0.2874	0.5088	0.3067	0.6122	0.6422	0.6276	0.5580	**0.6709**
31	0.4455	0.1158	0.1158	0.4676	0.2642	0.3257	0.3065	0.2558	0.3166	**0.5892**
32	0.2649	0.1281	0.1058	0.1574	0.3229	0.5050	**0.5815**	0.5266	0.2990	0.4784
33	0.0823	0.0823	0.0530	0.0823	0.3743	0.3125	0.3125	0.3125	0.0698	**0.4739**
34	0.0933	0.0070	0.0070	0.0070	0.7517	0.7968	0.0070	0.8564	0.1161	**0.8583**
35	0.3480	0.0691	0.0691	0.0691	0.1029	0.2768	0.4175	0.3815	0.3431	**0.4532**
Avg.	0.3171	0.1667	0.0935	0.2468	0.3522	0.3949	0.3628	0.4355	0.2414	**0.5531**
$-/+/\approx$	32/3/0	35/0/0	35/0/0	34/1/0	27/8/0	29/6/0	29/6/0	26/9/0	34/0/1	N/A

34 patient stratification datasets, respectively. It is significantly worse than KM, CL, DBSCAN, LCE, ASRS, and ECC on 3, 1, 8, 6, 6, and 9 patient stratification datasets, respectively. AL and SL cannot perform significantly better than MOCDP. MOCDP shares the best clustering results with CDP on one patient stratification dataset. For KM clustering method, it can generate better solutions (0.5083) on Nutt-2003-v1 than MOCDP. For DBSCAN, this algorithm can provide better solutions on Bitter-2000, Laiho-2007, Lapointe-2004-v1, Liang-2005, Nutt-2003-v2, and Pomeroy-2002-v1. For ASRS methods, this algorithm can perform better than MOCDP on Bhattacharjee-2001, Golub-1999-v2, Nutt-2003-v1, Nutt-2003-v3, and Tomlins-2006-v2. For ECC, this algorithm can give better solution on three patient stratification datasets including Armstrong-2002-v2, Chen-2002, and Nutt-2003-v2. It is noted that there are specific datasets (Garber-2001 and Gordon-2002) for which the traditional clustering methods yields inferior performance but our proposed algorithms can produce far better solutions, most likely due to our feature weighting under multiobjective setting. Fig. 1 also depicts the clustering performance of different algorithms measured by NMI. Fig. 1(a) visualizes the performance of various clustering methods based on the NMI. Fig. 1(b) shows the graphical results of ANOVA tests which demonstrate the robustness of our proposed algorithm. Therefore, it can conclude that MOCDP has the greater robustness than other methods on most of the patient stratification data.

2) For the evaluation metrics R_n, the experimental results of R_n are summarized in Table III which tabulates the mean values of 30 runs for each method. As observed from Table III, we can conclude several observations.
   a) Our proposed algorithm MOCDP can perform better than nine clustering algorithms on those 35 patient stratification datasets while agglomerative hierarchical clustering with SL shows the worst performance, although SL is commonly used in patient stratification.
   b) ECC can give better solutions than other traditional algorithms while MOCDP is inferior to, equal to and superior to ECC on 0, 11, and 24 patient stratification datasets, respectively.

TABLE III
PERFORMANCE OF DIFFERENT CLUSTERING ALGORITHMS ON BENCHMARK DATASETS BY R_n. FOR EACH ALGORITHM, A WILCOXON'S SIGNED-RANK TEST IS CONDUCTED TO VERIFY WHETHER THE EXPERIMENT RESULTS OF BEST ALGORITHMS ARE BETTER THAN OTHER ALGORITHMS. IF p-VALUE IS LESS THAN 0.05, IT DENOTES THAT THE BEST ALGORITHM OF INTEREST PERFORMS SIGNIFICANTLY BETTER THAN OTHER ALGORITHMS

Dataset	KM	AL	SL	CL	DBSCAN	LCE	ASRS	ECC	CDP	MOCDP
1	0.1242	0.0047	0.0000	0.0047	0.1278	0.4981	0.4306	0.5576	0.4981	**0.5563**
2	0.4027	**0.8519**	-0.0427	0.7483	0.0977	0.41	0.4091	0.4093	**1.0000**	**1.0000**
3	0.4606	0.4387	-0.0129	0.3848	0.0215	0.3235	0.4615	0.4464	0.4443	**0.5202**
4	0.2102	-0.0451	-0.0138	0.4753	0.4890	0.2102	0.2384	0.2296	0.3167	**0.8164**
5	0.7025	0.0009	0.0007	0.4620	0.5242	0.7821	0.7747	**0.848**	0.0989	0.6624
6	0.1790	0.2606	0.0390	0.4662	0.3316	0.204	0.2441	0.2362	0.0470	**0.3405**
7	-0.0157	0.0000	0.0000	-0.0103	0.1614	0.0179	0.0179	0.0429	0.0127	**0.1622**
8	0.3368	0.0056	0.0056	0.3671	0.5911	0.327	0.5183	0.3766	0.5480	**0.8072**
9	-0.0065	-0.0031	-0.0031	-0.0065	0.0001	0.5365	-0.006	**0.6505**	-0.0012	0.4707
10	0.0657	0.0091	0.0091	0.0091	0.0190	0.5273	0.0657	0.8857	0.0091	**0.9238**
11	0.5077	0.0534	0.0534	0.0534	0.1957	0.5779	0.5077	**0.6213**	0.1210	0.5169
12	0.2684	-0.0161	-0.0161	0.1023	0.2735	0.0722	0.1558	0.1084	-0.0395	**0.4471**
13	0.5905	0.2313	0.0243	0.5016	0.5573	0.6347	0.5905	**0.8473**	-0.0104	0.8062
14	0.4720	0.1710	0.0138	0.2440	0.3033	0.6709	**0.8776**	0.8127	-0.0236	0.7001
15	-0.0618	-0.0086	-0.0086	-0.0882	0.1835	-0.0583	-0.0948	-0.0126	0.0332	**0.8139**
16	0.2633	-0.0126	-0.0152	0.1848	0.4671	0.2951	0.3394	0.4433	0.2613	**0.6535**
17	0.2412	0.0838	0.1355	-0.1028	**0.5775**	0.1914	0.1244	0.1284	-0.0649	0.5599
18	0.0615	0.0402	0.0402	0.0402	**0.1961**	0.1037	0.1568	0.1295	0.0635	0.1407
19	0.0801	0.0253	0.0130	0.0375	0.0244	0.0968	0.1018	0.1386	0.1398	**0.2348**
20	0.1573	0.0497	-0.1529	0.1573	**0.6839**	0.1937	0.1573	0.1573	-0.1527	0.3214
21	0.3562	0.0008	0.0005	0.1232	0.2862	0.2483	0.3562	**0.3766**	0.1369	0.2370
22	0.0956	0.0023	0.0000	0.1082	0.3799	0.3055	0.0023	**0.4176**	0.0322	0.1628
23	0.8153	0.1020	0.1020	0.1020	0.7105	0.8153	**1**	0.8153	-0.0288	0.5740
24	0.0533	-0.0645	-0.0366	-0.0459	**0.3707**	-0.0459	-0.0459	-0.0468	0.2545	0.2176
25	0.3111	0.0975	0.1012	0.2549	0.1161	0.4571	0.4748	**0.4979**	0.1491	0.4917
26	0.1580	-0.0021	-0.0029	0.0126	0.0129	0.1782	0.1348	0.2285	0.2530	**0.3306**
27	0.0989	-0.0272	-0.0108	-0.0640	0.2166	0.1	0.0989	0.1212	0.0285	**0.2963**
28	-0.1006	-0.0325	-0.0325	-0.0325	0.3370	-0.0935	-0.1003	-0.0918	-0.0325	**0.4890**
29	0.0259	0.0269	0.0008	0.0269	0.0000	0.0047	0.0245	0.0044	0.0340	**0.2582**
30	0.3409	0.1296	0.0337	0.1821	0.0402	0.4007	0.4175	0.4532	0.3533	**0.5367**
31	0.2436	0.0099	0.0099	0.3178	0.0471	0.1911	0.151	0.1305	0.1632	**0.4555**
32	0.1360	0.0180	0.0122	0.0485	0.1357	0.3369	0.4154	0.3726	0.2115	**0.4244**
33	0.0001	0.0001	-0.0016	0.0001	0.2583	0.3875	0.3875	0.3875	0.0248	**0.4904**
34	-0.1061	-0.0063	-0.0063	-0.0063	0.9037	0.8803	-0.0063	**0.9344**	-0.0487	0.9244
35	0.1900	-0.0011	-0.0011	-0.0011	0.0000	0.1277	0.1866	0.2376	**0.2716**	0.2631
Avg.	0.2188	0.0684	0.0068	0.1445	0.2754	0.3117	0.2734	0.3684	0.1458	**0.5030**
$-/+/\approx$	32/3/0	35/0/0	35/0/0	29/6/0	28/7/0	30/5/0	30/5/1	24/11/0	32/2/1	N/A

c) DBSCAN can provide better solutions on seven patient stratification data while MOCDP is superior to DBSCAN on 28 patient stratification datasets.
d) Agglomerative hierarchical clustering with AL, SL cannot generate any patient stratification comparable to MOCDP while Agglomerative hierarchical clustering with CL only can give one better solution than MOCDP.
e) Compared with CDP and MOCDP, we can find that MOCDP is inferior to, equal to and superior to CDP in 1, 2, and 32 patient stratification data, respectively.
f) For some patient stratification datasets such as Gordon-2002, Khan-2001, Ramaswamy-2001, and Shipp-2002, MOCDP can give the best performance while other algorithms perform unsatisfactorily.
g) Fig. 1 also shows the clustering performance of different algorithms measured by the normalized rand index R_n. Fig. 1(a) demonstrates the performance of various clustering methods based on the normalized rand index R_n. Fig. 1(b) shows the analysis results of ANOVA tests, similar to NMI.

Based on the above analysis, we can conclude that the proposed algorithm has significant advantages in patient stratification. It also demonstrates that our proposed algorithm can enhance the performance of CDP under the proposed multiobjective framework. Meanwhile, the CPU time used by different clustering algorithms on 35 patient stratification datasets are summarized in Table S1 in the supplementary material. From Table S1 in the supplementary material, we can observe that the CPU time consumed by our proposed algorithm MOCDP is higher than other algorithms. The reason is that MOCDP employs a multiobjective evolutionary algorithm framework which is population-based. Therefore, it costs more CPU time for solving patient stratification datasets than other line-search clustering algorithms.

F. Compared With Other Multiobjective Evolutionary Algorithms

This section is devoted to the experimental design for investigating the performance of our proposed algorithm MOCDP from the perspective of multiobjective algorithm. We compare our proposed algorithm MOCDP with five state-of-the-art multiobjective evolutionary algorithms including NSGAII, GrEA,

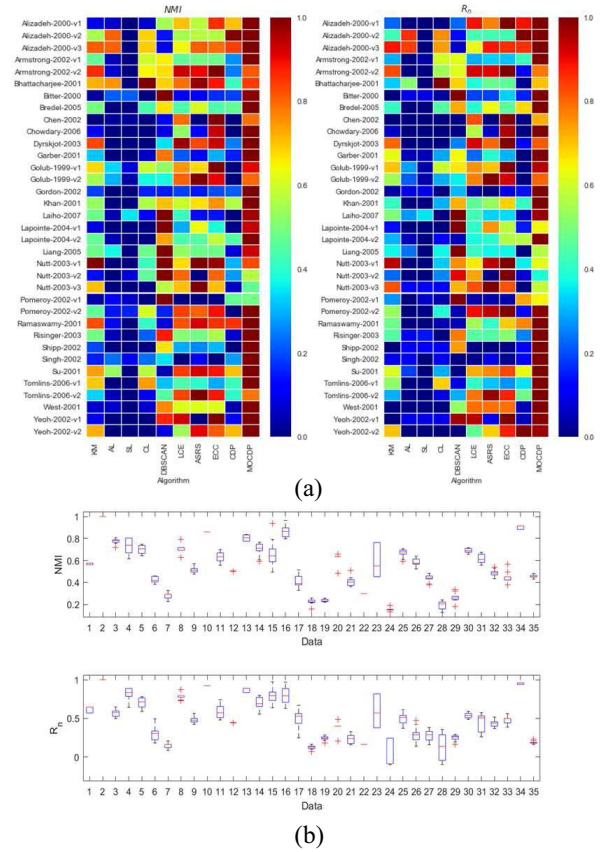


Fig. 1. Performance of MOCDP on 35 benchmark patient stratification datasets. (a) Performance is measured using the NMI and normalized rand index (R_n). (b) Graphical analysis results of ANOVA test.

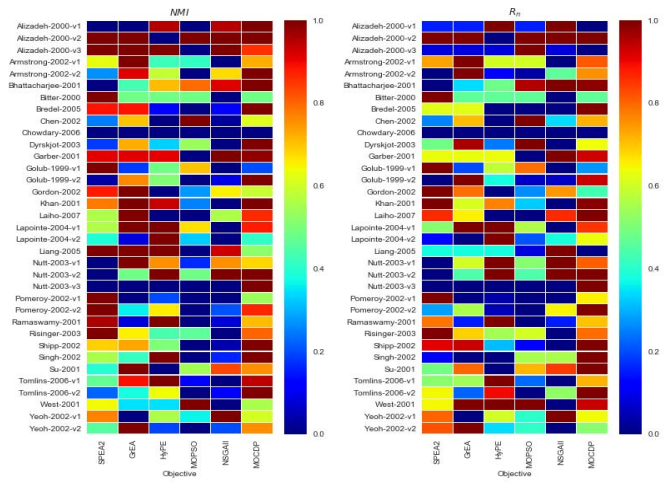


Fig. 2. Performance of MOCDP and other five multiobjective evolutionary algorithms on 35 benchmark patient stratification datasets. The performance is measured using the NMI and normalized rand index (R_n).

HyEA for multiobjective optimization, SPEA2, and MOPSO. Thirty-five patient stratification datasets were adopted to test the performance of MOCDP. Tables S2 and S3 in the supplementary material show the comparative results of those six algorithms. The last two rows of Tables S2 and S3 in the supplementary material summarize the statistical results. All algorithms have been independently run for 30 times on each patient stratification dataset.

1) For the evaluation metrics NMI, we can find that our proposed algorithm MOCDP provides better solutions than other multiobjective evolutionary algorithms as tabulated in Table S2 in the supplementary material. We can conclude several observations.
   a) Our algorithm MOCDP gives the best solutions while MOPSO obtains the worst solution.
   b) Our proposed algorithm MOCDP performs the best on 22, 21, 23, 29, and 25 patient stratification datasets compared with other multiobjective algorithms from the results of Wilcoxon's signed-rank test in the last two rows of Table S2 in the supplementary material.
   c) Our algorithm MOCDP is significantly worse than SPEA2, GrEA, HyPE, MOPSO, and NSGAII on 11, 11, 8, 4, and 7 patient stratification datasets, respectively.
   d) For the SPEA2 algorithm, it can provide the best solutions on eight patient stratification datasets, especially for Golub-1999-v1. For the GrEA algorithm, it can obtain the best solutions on 12 patient stratification datasets while MOPSO can provide best solutions on Chen-2002 and West-2001. For the Garber-2001 and Yeoh-2002-v1, NSGAII gives the best solutions. For the HyPE algorithm, it also gives the best solutions on eight patient stratification datasets.
   e) Fig. 2 depicts the clustering performance of different multiobjective evolutionary algorithms as measured by NMI.
2) For the evaluation metrics R_n, Table S3 in the supplementary material summarizes the experimental results which list the average values of 30 runs for each method. As can be seen in this supplementary material, we can have multiple observations.
   a) Our proposed algorithm MOCDP can provide better solutions than five state-of-the-art multiobjective evolutionary algorithms on those 35 patient stratification datasets while NSGAII provides the worst performance.
   b) SPEA2 can give better solutions than other compared multiobjective evolutionary algorithms while MOCDP is superior to, inferior to, and equal to SPEA2 on 21, 12, and 1 patient stratification datasets, respectively.
   c) MOPSO can provide better solutions on seven patient stratification dataset while MOCDP is superior to MOPSO on 25 patient stratification datasets, which can demonstrate that DE algorithm performs better than PSO.
   d) Compared with GrEA, our proposed algorithm MOCDP is superior to, inferior to, and equal to SPEA2 on 21, 11, and 2 patient stratification datasets, respectively.

e) Compared with HyPE, our proposed algorithm MOCDP also performs better than the others on 22 patient stratification datasets.
f) Fig. 2 shows the clustering performance of different multiobjective evolutionary algorithms as measured by the normalized rand index R_n.

As evidenced by the experimental results, we can claim that our proposed algorithm MOCDP can produce better solutions than other state-of-the-art multiobjective evolutionary algorithms on 35 patient stratification datasets.

G. Effect of Distance Weight

In this section, we compare the MOCDP algorithm with and without the distance weight. Thirty-five patient stratification datasets are also used to test the performance of MOCDP. Thirty independent times are run for each algorithm on those 35 patient stratification datasets. The experiment results in terms of NMI and R_n are summarized in Tables S4 and S5 in the supplementary material. As observed from these tables, MOCDP represents the MOCDP algorithm with the distance weight. MOCDP^{NO} represents the MOCDP algorithm without the distance weight. Table S4 in the supplementary material gives the NMI of MOCDP and MOCDP^{NO}. From Table S4 in the supplementary material, we can find that MOCDP can provide better solutions on 31 patient stratification datasets while MOCDP^{NO} can obtain the better solutions on two patient stratification datasets including Liang-2005 and Singh-2002. For the remaining two patient stratification datasets including Alizadeh-2000-v2 and Chowdary-2006, MOCDP can generate similar solutions with MOCDP^{NO}. The evaluation metrics R_n is summarized in Table S5 in the supplementary material. From this table, we can find that MOCDP algorithm provides better performance than the others on 29 patient stratification datasets. Meanwhile, MOCDP^{NO} can give the best solutions on 5 patient stratification datasets. On Alizadeh-2000-v2 and Chowdary-2006, they also have the same performance. Based on these results, it can be concluded that the MOCDP algorithm with the distance weight is more efficient than the MOCDP algorithm without the distance weight for multiobjective clustering on those 35 patient stratification datasets.

H. Different Objective Function Subsets

In this section, to demonstrate the effectiveness of each combination of five objective functions under the multiobjective framework, we compare our proposed algorithm under 31 different combinations of objective functions. All these objectives are optimized using MOCDPs. In this experiment, our proposed algorithm is compared with those 31 different combinations of objective functions based on 35 patient stratification datasets to evaluate its overall performance. The experimental results are presented in Tables S6–S13 in the supplementary material. Fig. 3 provides the visualization of those results. The pair Wilcoxon's signed-rank test is computed as shown in Tables S6–S13 in the supplementary material. For the evaluation metrics (NMI), it can be seen from Tables S6, S8, S10, and S12 in the supplementary material that our proposed

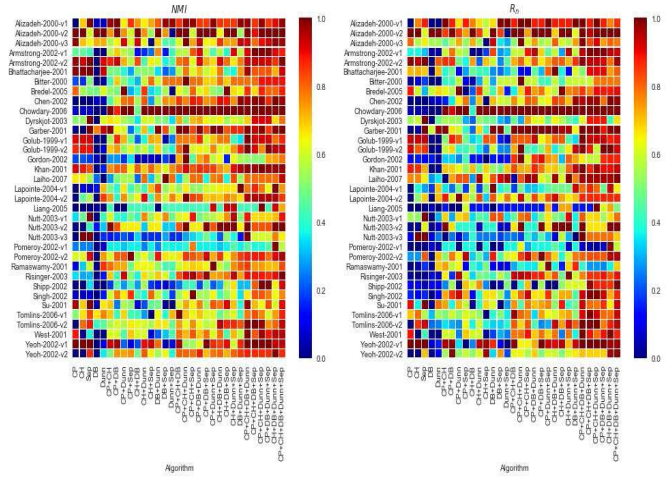


Fig. 3. NMI and R_n performance of all different combinations of objective functions on 35 benchmark patient stratification data sets. The horizontal axis denotes different objective sets while the vertical axis denotes the 35 patient stratification datasets.

algorithm performs better than other different combinations of objective functions such as the combination with one objective function, our proposed algorithm can perform better than those single objectives including CP, Sep, CH index, DB index, and Dunn index on 29, 30, 25, 34 and 32 patient stratification datasets, respectively. CP, Sep, CH, and Dunn perform better than MOCDP on 5, 4, 10, and 2 patient stratification datasets, respectively. The performance of DB objective function cannot provide any result better than that of MOCDP on all patient stratification datasets. For other combinations, five objective functions also provide better solutions than the others. For the evaluation metrics (R_n), the experimental results are summarized in Tables S7, S9, S11, and S13 in the supplementary material and Fig. 3. Based on the observations from Tables S7, S9, S11, and S13 in the supplementary material and Fig. 3, it is pointed out that the performance of our proposed algorithm is significantly superior to different combinations of objective functions such as, for different subsets with three objective functions, our proposed algorithm obtains better R_n value than all subsets with three objective functions on 29, 24, 28, 30, 28, 29, 29, 29, 26, and 29 patient stratification datasets, respectively. On the contrary, those different subsets including three objective functions obtain the better solutions on five objective functions on 4, 8, 6, 3, 4, 4, 4, 3, 6, and 5 patient stratification datasets, respectively. Meanwhile, MOCDP can provide similar results with those subsets under three objective functions on 2, 3, 1, 2, 2, 1, 2, 3, 3, and 1 patient stratification datasets. Therefore, the experimental results of Tables S6–S13 in the supplementary material and Fig. 3 demonstrate that our proposed algorithm MOCDP performs better on most patient stratification datasets than the others.

I. Convergence Behavior

To investigate the full convergence behavior of the proposed algorithm, a convergence analysis based on the number of objective FEs is conducted on those 35 patient stratification datasets. The average values of all those 35 patient stratification datasets are collected against different objective

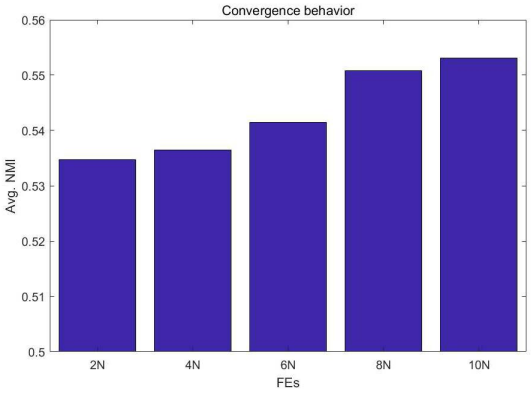


Fig. 4. Convergence behavior of proposed algorithm. The horizontal axis denotes the number of objective FEs while the vertical axis denotes the average value of NMI.

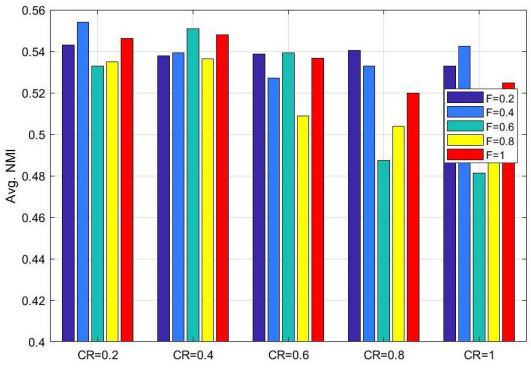


Fig. 5. Parameter analysis on F and CR under $N \times 10$ objective FEs. The horizontal axis denotes different population size while the vertical axis denotes the average NMI.

FEs. The experimental results are summarized in Table S14 in the supplementary material and Fig. 4. We can observe that a monotonically fitness-increasing trend is shared across those 35 patient stratification datasets, which demonstrates its potential with the increasing computing power in the future.

J. Parameter Analysis

1) Effect of Scale Factor and Crossover Rate: This section gives the comparative performance of proposed algorithm (MOCDP) by varying the scaling factor F from 0 to 1 and CR from 0 to 1. F is varied over the set $\{0.2, 0.4, 0.6, 0.8, 1\}$ and CR is varied over the set $\{0.2, 0.4, 0.6, 0.8, 1\}$. Performance at each setting is measured as the average NMI produced over 30 runs. The experimental results after $10N$ objective FEs by taking different values of F and CR are summarized in Fig. 5. As depicted in Fig. 5, we can conclude that the value of F and CR plays an important role on MOCDP. While F is set to 0.4 and CR is set to 0.2, the algorithms can produce statistically better results.

2) Sensitivity of T in MOCDP: To study the sensitivity of the parameter T in our proposed algorithm MOCDP on 35 patient stratification datasets, different settings of T in the MOCDP algorithm are examined over 30 runs. The results are summarized in Table S15 in the supplementary material and Fig. 6 regarding NMI. As clearly shown in Fig. 6,

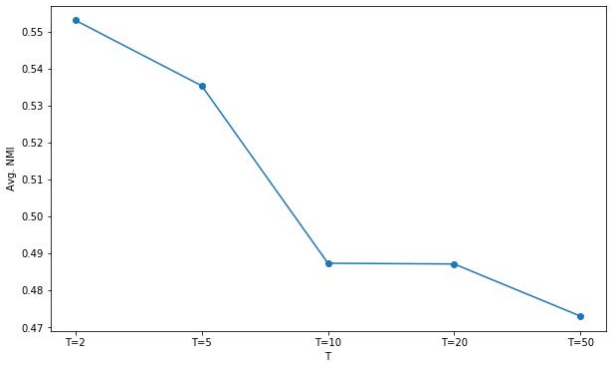


Fig. 6. Average NMI value versus the value of T in MOCDP for 35 patient stratification datasets. The horizontal axis denotes different neighborhood size while the vertical axis denotes the NMI.

MOCDP works very well with $T = 2$ on those patient stratification datasets. The performance of MOCDP is degraded as the parameter T is increased if the algorithm is provided the same number of objective FEs for fair comparisons.

IV. Extended Performance Comparisons With Case Studies

A. Evaluation on Synthetic Data

Fifty-five synthetic gene expression datasets from the published dynamical gene regulation model [10] are generated to further show the performance of our proposed algorithm. All datasets are generated as follows:

$$\begin{aligned} F_i^{m\text{RNA}}(x, y) &= \frac{dx_i}{dt} = m_i \cdot f_i(y) - \lambda_i^{m\text{RNA}} \cdot x_i \\ F_i^{\text{Prot}}(x, y) &= \frac{dy_i}{dt} = r_i \cdot x_i - \lambda_i^{\text{Prot}} \cdot y_i \\ i &= 1, \ldots, n \end{aligned} \tag{9}$$

where m_i is the maximum transcription rate, r_i is the translation rate, and $\lambda_i^{m\text{RNA}}$ and λ_i^{Prot} are the mRNA and protein degradation rates. $f_i(\cdot)$ is the relative activation of gene. For the 55 synthetic gene expression datasets, they are synthesized based on a real human transcription regulation network. Each dataset contains 200 samples, which are divided into four clusters. The number of knock-out genes is varied from 100 to 500. The noise level is varied from 0 to 0.5. Each knock-out genes includes 11 instances for different noise levels.

Nine algorithms including KM clustering method, Agglomerative hierarchical clustering with AL, SL, and CL, DBSCAN, LCE, ECC, CDPs, and MOCDP are applied to cluster 55 synthetic datasets with different settings based on a real human transcriptional regulation network of 2723 genes. The experimental results of NMI and R_n are summarized in Tables S16 and S17 in the supplementary material and Fig. 7 based on 30 independent runs. Our proposed algorithm MOCDP obtains better performance than other clustering algorithms on most of the datasets. Although our proposed algorithm cannot provide the best performance for some cases, the difference between the best one and our proposed algorithm is very small. Meanwhile, compared

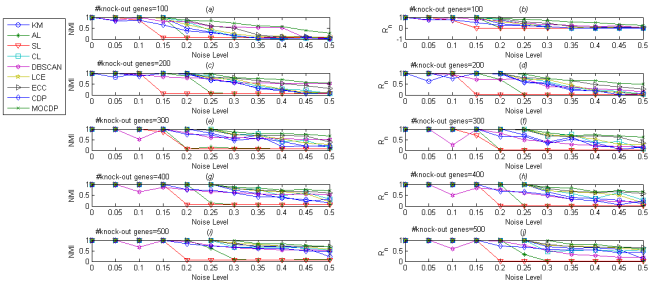


Fig. 7. Performance of different clustering algorithms on the 55 synthetic datasets (based on a real human transcriptional regulation network of 2723 genes). MOCDP has substantial advantages over other methods on the datasets. The number of knock-out genes is varied from 100 to 500. The noise level is varied from 0 to 0.5. Each knock-out genes includes 11 instances at different noise levels.

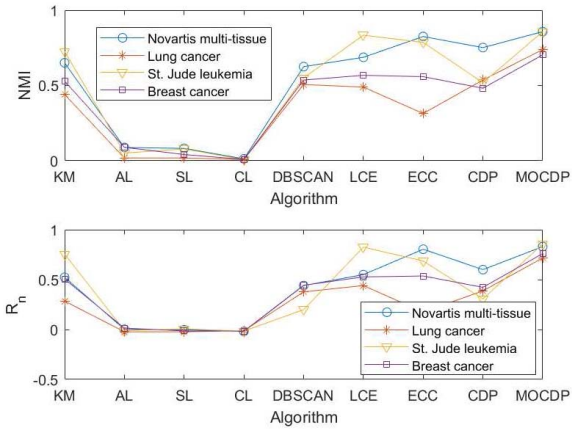


Fig. 8. Performance of different clustering algorithms on those four biomedical datasets. MOCDP has substantial advantages over other methods on the datasets.

with the original CDP algorithm published on Science in 2014, our proposed algorithm can achieve better or similar performance on all synthetic datasets, which demonstrates that we can improve the performance of CDP by transforming the clustering problem into a multiobjective optimization problem.

B. Visualization of Patient Stratification

Visualization plays an important role in analyzing the biological data which number of dimensions is always enormous. In this section, we implement two well-known including *t*-distributed stochastic neighbor embedding (t-SNE) [51] and principal component analysis (PCA) [52] to project the selected feature subsets into two dimensions to visualize the high-dimensional patient stratification data based on 35 patient stratification datasets. The 2-D visualization of all patient stratification datasets is summarized in Fig. S1 in the supplementary material. The axes are in arbitrary units in the projected space. Each point denotes a sample in the patient stratification datasets. It is noted that the t-SNE and PCA algorithms visualize all these datasets without the ground truth labels; the truth label information is added in the form of distinct colors to validate the results. From those figures, we can summarize that the t-SNE can successfully visualize the patient samples into different subgroups compared with PCA.

C. Evaluation on Real-World Medical Datasets

In this section, we choose four real-world datasets including Novartis multitissue with four clusters, Lung cancer with four clusters and St. Jude leukemia with six clusters, and Breast cancer with three clusters. All cancer gene expression data sets are obtained from (http://portals.broadinstitute.org/cgi-bin/cancer/datasets.cgi). The experimental results are summarized in Fig. 8. It can be observed that, the proposed algorithm performs better than other algorithms including KM clustering method, Agglomerative hierarchical clustering with AL, SL, and CL, DBSCAN, LCE, ECC, and CDPs, especially for the Agglomerative hierarchical clustering with AL, SL, and CL. An affinity matrix proposed in [53] of existing classes in the dataset is visualized using the heatmap, which can show

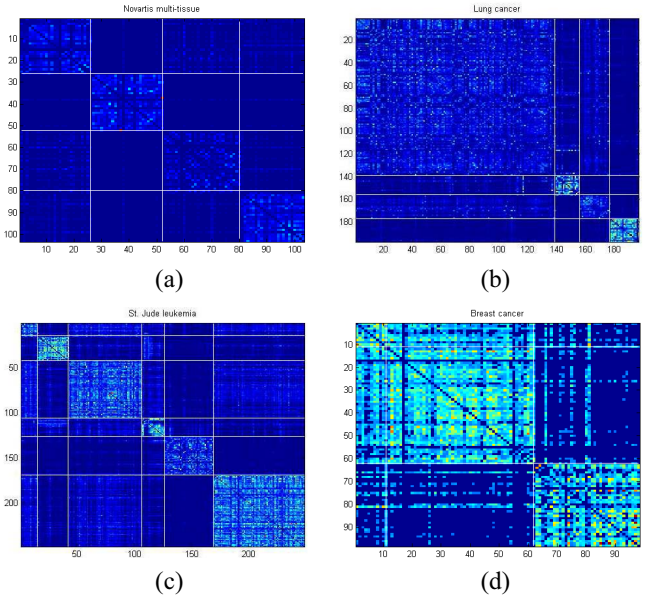


Fig. 9. Affinity matrices of obtained subclasses of Novartis multitissue with four clusters, Lung cancer with four clusters, St. Jude leukemia with six clusters, and Breast cancer with three clusters; the color scheme represents the inner class similarity.

the clusters for different datasets. For the Novartis multitissue with four clusters, the obtained results are comparable with that of other algorithms with the best NMI and R_n among all algorithm; the obtained affinity matrix of real classes is visualized in Fig. 9(a). For the Lung cancer with four clusters, MOCDP can give the best solution. In Fig. 9(b), it can be observed that our method is more reliable to organize the genes into distinct subclasses for Lung cancer. For St. Jude leukemia with six clusters, the superiority over AL, SL, CL, DBSCAN, LCE, ECC, and CDP of our proposed algorithm is illustrated; the related heatmap is shown in Fig. 9(c). For breast cancer with three clusters, MOCDP and other algorithms' clustering results are plotted in Fig. 8. It can be observed that our proposed algorithm (MOCDP) can successfully differentiate different breast cancer types, as shown in Fig. 9(d).

V. Conclusion

In summary, we demonstrate the competitive edges of our proposed multiobjective framework over existing clustering algorithms on patient stratification datasets. In particular, our proposed multiobjective algorithm includes a population of dissimilarity weight and the cutoff distance. Then, we select five cluster validity indices as multiple objective functions to capture multiple characteristics of the evolving clusters, using the MOEA/D-DE to optimize those five objective functions simultaneously. Extensive experiments are conducted on 94 datasets including 55 synthetic datasets based on a real human transcription regulation network, 35 real patient stratification datasets, and four real world medical datasets to analyze and compare our approach with 45 clustering methods, demonstrating the robust and competitive performance of our proposed algorithm for patient stratification across thousands of pair-wise benchmark comparisons.

References

[1] J. W. Davey *et al.*, "Genome-wide genetic marker discovery and genotyping using next-generation sequencing," *Nat. Rev. Genet.*, vol. 12, no. 7, pp. 499–510, 2011.

[2] H. B. Frieboes *et al.*, "An integrated computational/experimental model of tumor invasion," *Cancer Res.*, vol. 66, no. 3, pp. 1597–1604, 2006.

[3] E. Bresin *et al.*, "Combined complement gene mutations in atypical hemolytic uremic syndrome influence clinical phenotype," *J. Amer. Soc. Nephrol.*, vol. 24, no. 3, pp. 475–486, 2013.

[4] D. Diaz, T. Nguyen, and S. Draghici, "A systems biology approach for unsupervised clustering of high-dimensional data," in *Proc. Int. Workshop Mach. Learn. Optim. Big Data*, 2016, pp. 193–203.

[5] S. Khakabimamaghani and M. Ester, "Bayesian biclustering for patient stratification," in *Proc. Pac. Symp. Biocomput.*, vol. 21, 2016, pp. 345–356.

[6] Y. Chang *et al.*, "COPD subtypes identified by network-based clustering of blood gene expression," *Genomics*, vol. 107, nos. 2–3, pp. 51–58, 2016.

[7] K. Graim *et al.*, "Revealing cancer subtypes with higher-order correlations applied to imaging and omics data," *BMC Med. Genomics*, vol. 10, no. 1, p. 20, 2017.

[8] J. Wu *et al.*, "Unsupervised clustering of quantitative image phenotypes reveals breast cancer subtypes with distinct prognoses and molecular pathways," *Clin. Cancer Res.*, vol. 23, no. 13, pp. 3334–3342, 2017.

[9] P. A. Jaskowiak, R. J. Campello, and I. G. Costa, "On the selection of appropriate distances for gene expression data clustering," *BMC Bioinformat.*, vol. 15, no. 2, pp. 1–7, 2014.

[10] H. Liu *et al.*, "Entropy-based consensus clustering for patient stratification," *Bioinformatics*, vol. 33, no. 17, pp. 2691–2698, 2017.

[11] N. Sengee and H. K. Choi, "Brightness preserving weight clustering histogram equalization," *IEEE Trans. Consum. Electron.*, vol. 54, no. 3, pp. 1329–1337, Aug. 2008.

[12] K.-C. Wong, C. Peng, Y. Li, and T.-M. Chan, "Herd clustering: A synergistic data clustering approach using collective intelligence," *Appl. Soft Comput.*, vol. 23, pp. 61–75, Oct. 2014.

[13] J. Yu, M.-S. Yang, and E. S. Lee, "Sample-weighted clustering methods," *Comput. Math. Appl.*, vol. 62, no. 5, pp. 2200–2208, 2011.

[14] R. Nock and F. Nielsen, "On weighting clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1223–1235, Aug. 2006.

[15] D. S. Modha and W. S. Spangler, "Feature weighting in *k*-means clustering," *Mach. Learn.*, vol. 52, no. 3, pp. 217–237, 2003.

[16] R. C. De Amorim and B. Mirkin, "Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering," *Pattern Recognit.*, vol. 45, no. 3, pp. 1061–1075, 2012.

[17] L. Jing, M. K. Ng, J. Xu, and J. Z. Huang, "Subspace clustering of text documents with feature weighting *k*-means algorithm," in *Proc. PAKDD*, 2005, pp. 802–812.

[18] S. Güneş, K. Polat, and Ş. Yosunkaya, "Efficient sleep stage recognition system based on EEG signal using *k*-means clustering based feature weighting," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7922–7928, 2010.

[19] L. Wang and J. Wang, "Feature weighting fuzzy clustering integrating rough sets and shadowed sets," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 26, no. 4, 2012, Art. no. 1250010.

[20] J. Z. Huang, M. K. Ng, H. Rong, and Z. Li, "Automated variable weighting in *k*-means type clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 657–668, May 2005.

[21] J. Handl and J. Knowles, "An evolutionary approach to multiobjective clustering," *IEEE Trans. Evol. Comput.*, vol. 11, no. 1, pp. 56–76, Feb. 2007.

[22] M. H. C. Law, A. P. Topchy, and A. K. Jain, "Multiobjective data clustering," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Washington, DC, USA, 2004, pp. 1–7.

[23] U. Maulik, A. Mukhopadhyay, and S. Bandyopadhyay, "Combining Pareto-optimal clusters using supervised learning for identifying co-expressed genes," *BMC Bioinformat.*, vol. 10, no. 1, p. 27, 2009.

[24] A. Mukhopadhyay, U. Maulik, and S. Bandyopadhyay, "Multiobjective genetic algorithm-based fuzzy clustering of categorical attributes," *IEEE Trans. Evol. Comput.*, vol. 13, no. 5, pp. 991–1005, Oct. 2009.

[25] K. Suresh, D. Kundu, S. Ghosh, S. Das, and A. Abraham, "Data clustering using multi-objective differential evolution algorithms," *Fundamenta Informaticae*, vol. 97, no. 4, pp. 381–403, 2009.

[26] G. Armano and M. R. Farmani, "Multiobjective clustering analysis using particle swarm optimization," *Expert Syst. Appl.*, vol. 55, pp. 184–193, Aug. 2016.

[27] S. Saha and S. Bandyopadhyay, "A new multiobjective simulated annealing based clustering technique using symmetry," *Pattern Recognit. Lett.*, vol. 30, no. 15, pp. 1392–1403, 2009.

[28] T. İnkaya, S. Kayalıgil, and N. E. Özdemirel, "Ant colony optimization based clustering methodology," *Appl. Soft Comput.*, vol. 28, pp. 301–311, Mar. 2015.

[29] A. Rodriguez and L. Alessandro, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[30] C. Wiwie, J. Baumbach, and R. Röttger, "Comparing the performance of biomedical clustering methods," *Nat. Methods*, vol. 12, no. 11, pp. 1033–1038, 2015.

[31] E. A. Josephs *et al.*, "Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage," *Nucleic Acids Res.*, vol. 43, no. 18, pp. 8924–8941, 2015.

[32] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 11, no. 6, pp. 712–731, Dec. 2007.

[33] K. Deb, "Multi-objective optimization," in *Search Methodologies*. Boston, MA, USA: Springer, 2014, pp. 403–449. [Online]. Available: https://scholar.google.co.jp/scholar?hl=zh-CN&as_sdt=2005&sciodt=0%2C5&cites=5259926186684130741&scipsc=&q=Multi-objective+optimization&btnG=

[34] C. A. C. Coello, "Evolutionary multi-objective optimization: A historical view of the field," *IEEE Comput. Intell. Mag.*, vol. 1, no. 1, pp. 28–36, Feb. 2006.

[35] G. R. Zavala, A. J. Nebro, F. Luna, and C. A. C. Coello, "A survey of multi-objective metaheuristics applied to structural optimization," *Struct. Multidiscipl. Optim.*, vol. 49, no. 4, pp. 537–558, 2014.

[36] R. A. Botafogo and B. Shneiderman, "Identifying aggregates in hypertext structures," in *Proc. 3rd Annu. ACM Conf. Hypertext*, San Antonio, TX, USA, 1991, pp. 63–74.

[37] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy C-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 370–379, Aug. 1995.

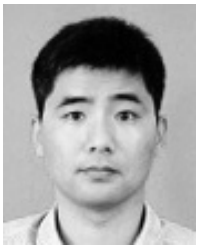
[38] U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1650–1654, Dec. 2002.

[39] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik, "Validity index for crisp and fuzzy clusters," *Pattern Recognit.*, vol. 37, no. 3, pp. 487–501, 2004.

[40] M. C. de Souto, I. G. Costa, D. S. de Araujo, T. B. Ludermir, and A. Schliep, "Clustering cancer gene expression data: A comparative study," *BMC Bioinformat.*, vol. 9, no. 1, p. 497, 2008.

[41] K. Wagstaff, C. Cardie, S. Rogers, and S.Schrödl, "Constrained K-means clustering with background knowledge," in *Proc. ICML*, vol. 1, 2001, pp. 577–584.

[42] D. Jiang, C. Tang, and A. Zhang, "Cluster analysis for gene expression data: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 11, pp. 1370–1386, Nov. 2004.
[43] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. Kdd*, vol. 96, 1996, pp. 226–231.
[44] N. Iam-On, T. Boongoen, and S. Garrett, "LCE: A link-based cluster ensemble method for improved gene expression data analysis," *Bioinformatics*, vol. 26, no. 12, pp. 1513–1519, 2010.
[45] P. Galdi, F. Napolitano, and R. Tagliaferri, "Consensus clustering in gene expression," in *Proc. Int. Meeting Comput. Intell. Methods Bioinformat. Biostat.*, vol. 8623. Naples, Italy, 2015, pp. 57–67. [Online]. Available: https://scholar.google.co.jp/scholar?hl=zh-CN&as_sdt=0%2C5&q=Consensus+clustering+in+gene+expression&btnG=
[46] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.
[47] S. Yang, M. Li, X. Liu, and J. Zheng, "A grid-based evolutionary algorithm for many-objective optimization," *IEEE Trans. Evol. Comput.*, vol. 17, no. 5, pp. 721–736, Oct. 2013.
[48] J. Bader and E. Zitzler, "Hype: An algorithm for fast hypervolume-based many-objective optimization," *Evol. Comput.*, vol. 19, no. 1, pp. 45–76, 2011.
[49] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," in *Proc. Evol. Methods Design Optim. Control Appl. Ind. Problems*, 2001, pp. 95–100.
[50] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 256–279, Jun. 2004.
[51] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[52] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1997.
[53] B. Wang *et al.*, "Similarity network fusion for aggregating data types on a genomic scale," *Nat. Methods*, vol. 11, no. 3, pp. 333–337, 2014.



Xiangtao Li received the B.Eng. degree and the M.Eng. and Ph.D. degrees in computer science from Northeast Normal University, Changchun, China, in 2009, 2012, and 2015, respectively.

He is currently an Associate Professor with the Department of Computer Science and Information Technology, Northeast Normal University. He has published over 50 research papers. His current research interests include intelligent computation, evolutionary data mining, constrained optimization, bioinformatics, computational biology, and interdisciplinary research.

Ka-Chun Wong received the B.Eng. degree in computer engineering and the M.Phil. degree from the United College, Chinese University of Hong Kong, Hong Kong, in 2008 and 2010, respectively, and the Ph.D. degree from the Department of Computer Science, University of Toronto, Toronto, ON, Canada, in 2014.

He is an Assistant Professor with the City University of Hong Kong, Hong Kong. He has solely edited two books published by Springer and CRC Press, attracting 30 peer-reviewed book chapters around the world. His current research interests include bioinformatics, computational biology, evolutionary computation, data mining, machine learning, and interdisciplinary research.

Dr. Wong is merited as an Associate Editor of *BioData Mining* in 2016. He has been on the editorial board of *Applied Soft Computing* since 2016.