ARTICLE IN PRESS
JID: NEUCOM [m5G;September 27, 2017;13:31]
Neurocomputing 000 (2017) 1–13
Contents lists available at ScienceDirect
Neurocomputing
journal homepage: www.elsevier.com/locate/neucom

# Fuzzy mixed-prototype clustering algorithm for microarray data analysis

Jin Liu[a,*], Tuan D. Pham[b], Hong Yan[c], Zhizhen Liang[a,*]

[a] School of Computer Science, China University of Mining and Technology, Xuzhou, Jiangsu 221008, China
[b] Department of Biomedical Engineering, Linkoping University, Linkoping 581 83, Sweden
[c] Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong

ARTICLE INFO

*Article history:*
Received 18 November 2015
Revised 21 May 2017
Accepted 10 June 2017
Available online xxx

*Keywords:*
FMP
Microarray data analysis
Fuzzy clustering

ABSTRACT

Being motivated by combining the advantages of hyperplane-based pattern analysis and fuzzy clustering techniques, we present in this paper a fuzzy mix-prototype (FMP) clustering for microarray data analysis. By integrating spherical and hyper-planar cluster prototypes, the FMP is capable of capturing latent data models with both spherical and non-spherical geometric structures. Our contributions of the paper can be summarized into three folds: first, the objective function of the FMP is formulated. Second, an iterative solution which minimizes the objective function under given constraints is derived. Third, the effectiveness of the proposed FMP is demonstrated through experiments on yeast and leukemia data sets.



## 1. Introduction

Despite the fact that the fuzzy *c*-means (FCM) algorithm has been applied in different areas successfully, it has been known that the FCM may perform well only when the data set is of spherical or hyperspherical structure. However, in real world applications, there may be many other different types of data structures in which most of the current clustering algorithms may fail to perform well [1], for instance, linear or hyperplane shaped data clusters. Some techniques are good at linear or non-linear cluster structures detection, i.e., graph-theoretic methods, but there are no explicit prototypes for the clusters, hence it is difficult to further explain the clustering results and perform classification. Furthermore, in some certain research areas, such as image processing and computer vision, clustering algorithms need to consider not only the cluster prototypes but also the geometry of clusters to perform structure segmentation. Last but not least, data samples in real world applications often overlap with each other, i.e., microarray gene expression data [2–4]. For any clustering algorithms, how to take both properties of overlapping and the linear subspace structure of the data samples into consideration is worth investigation.

Since the proposal of support vector machines (SVMs), hyperplanes-based pattern analysis is attracting more and more attention from research community as a result of that the technique provides researchers with great power to handle many different pattern classification problems [5–8]. As one of the most successful classification methods, the SVMs aim to find an optimal separating hyperplane between two different categories of data to perform data classification. Through taking advantage of the kernel trick method, the SVMs are capable of differentiating linearly inseparable data set. The technique is famous for their excellent performance in pattern classification and have been used widely and successfully. Nevertheless, the SVMs are also known for their computation cost during their training process. Estimation of an optimal separating hyperplane is achieved by solving a quadratic programming problem which involves kernel matrix inversion. The training process of SVMs is of complexity on the order of $O(n^3)$, where $n$ is the number of samples in the training set. Recently, many efforts have been devoted to relieve the computational burden of the SVMs while withholding the classification accuracy through adopting the hyperplanes-based approximation [9–14]. Being different from the original SVMs, hyperplanes in these works are adopted to approximate different types of data rather than to split them from each other. The optimal hyperplane minimizes the sum of squared Euclidean distances from one cluster and maximize the sum of squared Euclidean distances from the other cluster. The objective functions are in the form of Rayleigh quotient and the solution can be achieved by generalized eigenvalue decomposition. By this means, the efficiency of these algorithms and the accuracy of classification were reported.

For unsupervised pattern recognition techniques, hyperplane-based clustering algorithms are also attracting research attention

* Corresponding author.
*E-mail addresses:* jinliu@cumt.edu.cn (J. Liu), liang@cumt.edu.cn (Z. Liang).

https://doi.org/10.1016/j.neucom.2017.06.083
0925-2312/

Please cite this article as: J. Liu et al., Fuzzy mixed-prototype clustering algorithm for microarray data analysis, Neurocomputing (2017), https://doi.org/10.1016/j.neucom.2017.06.083

widely. A $k$-planes clustering technique was put forward in [15], where hyperplanes were adopted to represent cluster centers.

The objective of the clustering is to minimize the sum of the squared Euclidean distances between data and their projections on their representative hyperplane. The $k$-planes clustering algorithm iteratively updates the partition matrix and clustering hyperplanes until convergence reached. The $k$-bottleneck hyperplane clustering ($k$-bHPC), which is another hyperplane-based clustering technique, was put forward in [6]. The objective function of $k$-bHPC is the minimum of maximum distance from the data samples to their belonging hyperplane. The clustering algorithm aims to find a group of hyperplanes and a partition matrix which can minimize the given objective function.

There are also some other related works have been put forward recently [16–24]. An extreme learning machine (ELM)-based method for heat load prediction in district heating system was presented in [19]. Nine different ELM predictive models were developed for time horizon from 1 to 24 h ahead. Experiments results were compared with that of genetic programming (GP) and artificial neural networks (ANNs) models. Improvements in predictive accuracy and capability of generalization were demonstrated. In [18], an Expert Multi Agent System (E-MAS) based support vector regression (SVR) was proposed to determine collar dimensions around bridge pier. In [20], a fuzzy clustering approach based on fuzzy distance measurement was presented, and multi-objective mathematical programming was then adopted for further optimization. In [17], a novel density-based fuzzy clustering algorithm based on Active Learning Method (ALM) was presented. In [21], a collaborative clustering framework which combines fuzzy $c$-means (FCM) and mixture mode was presented for mixed data which contains both numerical and categorical attributes.

Being motivated by the useful concepts of combining hyperplane-based data proximation with fuzzy clustering techniques, we presented herein a fuzzy mix-prototype clustering technique in which hyperplanes and hyperspheres are used to form cluster prototypes. The objective function of the proposed clustering technique is the sum of the distances from all of the data points to the clustering hyperplanes, weighted by the degree of the point belonging to the corresponding clusters, and penalized by distances of data samples to cluster mass centers. The proposed fuzzy mix-prototype clustering aims to find a solution to minimize the fuzzy objective function under given constraints. The clustering problem can then be considered as a constrained optimization problem and an iterative solution can then be obtained by using the Lagrangian multiplier method. The solutions are the resulting clusters that minimize the objective function.

The rest of the paper is organized as the follows. Section 2 gives a summary of some related works. Section 3 describes the proposed fuzzy mix-prototype clustering in detail, including formulation of the fuzzy objective function, derivation of an solution and description of the resulting algorithm. In Section 4, we report the experimental results of the proposed method and compare these results with those obtained from some existing methods. Concluding remarks of the proposed approach are addressed in Section 5.

## 2. Related work

Some methods which are closely related to the proposed fuzzy mix-prototype clustering are briefly discussed in the following subsections.

### 2.1. Fuzzy c-means clustering

Fuzzy $c$-means clustering is a kind of soft clustering which allows a data point to belong to more than one cluster [1]. The membership $u_{ij}$ is an continuous value which denotes the degree of data point $\mathbf{x}_i$ belong to cluster $j$, and it is the entry of $i$th row and $j$th column of membership matrix $U$.

FCM uses Euclidean distance to represent the dissimilarity between vectors and the algorithm is derived according to minimization of the following objective function:

$$J_{\mathrm{FCM}} = \sum_{i=1}^{n} \sum_{j=1}^{c} (u_{ij})^m ||\mathbf{x}_i - \mathbf{v}_j||_2^2 \tag{1}$$

where $m \in [1, +\infty)$ denotes the fuzziness degree, $n$ and $c$ denote the number of vectors and cluster centers respectively, $\mathbf{v}_j$ denotes the $j$th cluster center, and $||\mathbf{x}||_2^2$ represents the squared norm 2 of vector $\mathbf{x}$.

The minimization of $J_{\mathrm{FCM}}$ subjects to the following constraints:

$$u_{ij} \in [0, 1], i = 1, ..., n, j = 1, ..., c \tag{2}$$

$$\sum_{j=1}^{c} u_{ij} = 1 \tag{3}$$

$$0 < \sum_{i=1}^{n} u_{ij} < n, j = 1, ..., c \tag{4}$$

By using the Lagrangian multiplier method, necessary conditions for minimizing $J_{\mathrm{FCM}}$ under the given constraints can be derived and the cluster centers and partition matrix can be updated according to

$$\mathbf{v}_j = \sum_{i=1}^{n} (u_{ij})^m \mathbf{x}_i / \sum_{i=1}^{n} (u_{ij})^m \tag{5}$$

$$u_{ij} = \sum_{k=1}^{c} \left( \frac{||\mathbf{x}_i - \mathbf{v}_j||_2^2}{||\mathbf{x}_i - \mathbf{v}_k||_2^2} \right)^{\frac{1}{1-m}} \tag{6}$$

And the algorithm is summarized as

1. Randomly initialize the membership $u_{ij}, i = 1, ..., n; j = 1, ..., c$;
2. Given termination criterion $\varepsilon \in (0, 1)$;
3. Set $t = 0$, iterate:
   (a) update cluster center according to Eq. (5);
   (b) compute $||\mathbf{x}_i - \mathbf{v}_j||$;
   (c) update membership $u_{ij}$ according to Eq. (6);
   (d) if $||U^{(t+1)} - U^{(t)}|| < \varepsilon$ then stop, otherwise continue,

where the fuzzy weighting exponent $m$ is usually chosen as 2.

### 2.2. Kernel FCM

Kernel FCM (KFCM) is a variant of FCM which extends fuzzy clustering into kernel space [25]. The clustering method makes use of kernel transformations to map vectors from the original $p$-dimensional feature space to a kernel space which is of higher dimensionality. Through this mapping, problems that are linearly non-separable in the original feature space become linearly separable in the kernel space, and then fuzzy clustering algorithms can be used to perform data analysis.

Kernel FCM takes advantage of the 'kernel trick' to perform data analysis. The 'kernel trick' is achieved by using a continuous, symmetric, positive semi-definite function which is known as 'kernel function'. By using this kernel function, the inner product between two vectors in the kernel space can be directly computed, without knowing the explicit form of the vectors in the kernel space.

For example, kernel function $K(\mathbf{x},\mathbf{y})$ where

$$K(\mathbf{x},\mathbf{y}) = \phi(\mathbf{x})^T \phi(\mathbf{y}), \tag{7}$$

represents the inner product between two vectors $\mathbf{x}, \mathbf{y}$ in the kernel space, and $\mathbf{x}, \mathbf{y} \in \mathbf{R}^p$ are $p$-dimension vectors, function $\phi(\mathbf{x})$