# 一种用于识别共调控基因的AFS新型模糊双聚类算法

Kaijie Xu

摘要—识别共调控基因及其转录因子结合位点（转录因子结合位点）是理解转录调控的关键步骤。除了有效的实验室检测外，人们还开发了各种用于检测共表达基因的双聚类算法。将双聚类方法应用于基因表达数据时，可发现基因子群：它们在待识别的实验条件子集下具有相似的表达模式。本文利用公理模糊集（AFS）理论，为基因表达数据构建两个模糊划分矩阵，并据此提出一种用于识别共调控基因的新型模糊双聚类算法。具体而言，首先通过 AFS 的子偏好关系理论，将基因表达数据转换为两个模糊划分矩阵。其中一个矩阵以基因为论域、条件为概念；另一个则以基因为概念、条件为论域。共调控基因（双聚类）的识别同时在这两个划分矩阵上进行。随后，基于划分矩阵定义了一种新型的基于模糊集的相似性准则，并设计了一种循环优化算法，以发现表达水平上显著的双聚类。上述过程保证了生成的双聚类具有比传统双聚类方法提取出的双聚类更显著的表达值。最后，本文在公开的真实微阵列数据集上，将所提方法与三种著名的双聚类算法进行了性能评估。实验结果与理论分析一致，表明所提算法无需任何基因表达数据的先验知识，即可有效检测共调控基因。

索引词—双聚类、公理模糊集（公理模糊集）、基因表达、共调控基因、划分矩阵。

## I. 引言

基因表达聚类能够让人不受拘束地探索数据，而不至于淹没在数千个单独的基因之中 [1]。传统的（全局）聚类方法只分析所有实验条件下的基因，或只分析所有基因在各条件下的情况。在实际的许多细胞过程中，许多基因会在某些特殊条件下定期共表达（共同调控）[2]，[3] 但在不同条件下表现各异。因此，挖掘局部共表达的有价值模式，成为发现遗传通路的重要目标；这些通路在全局聚类时并不十分明显 [4]。这就是所谓的双聚类问题。双聚类扩展了传统的聚类技术，旨在寻找在表达模式相似的基因（全部）子群，这些基因在应用于基因表达数据时，所对应的是尚待识别的实验条件子集。

双聚类能够从大量基因表达数据中发现有价值的共调控基因模式；与传统聚类方法相比，这些模式更有助于识别协同发挥作用的基因。双聚类模型衡量基因和条件子集内部的一致性。该模型能够有效揭示基因或条件参与多重通路的情况，其中一些通路只有在更一致的通路占主导地位时才会被发现 [5]。一致性得分通常通过构建涉及相关基因和条件的对称函数来定义，因此，双聚类是同时对基因和条件进行聚类的过程。Cheng 和 Church [6]. 提出的均方残差（MSR）是一种定义方法，最初用于经过对数变换并加上加法逆元的基因表达数据。此外，MSR 也是双聚类中最常用的指标，许多双聚类算法都是基于该指标开发的。

迄今为止，人们已开发出各种算法来尝试解决双聚类问题。文献中广受关注的双聚类算法包括 Cheng and Church（CC）算法 [6],灵活重叠簇（FLOC） [7], Plaid [8],保序子矩阵（OPSM） [9], 迭代签名算法（ISA） [10], 保守基因表达基序（xMOTIFs）[11], 和 BiMax [12] 。新兴算法，如贝叶斯双聚类 [13], 最大相似度双聚类算法（MSB）[14], 和定性双聚类算法（QUBIC）[15]，尚未得到广泛研究。总之，截至目前，双聚类研究仍处于起步阶段。由于缺乏能够可靠解决一般双聚类问题的有效且高效的算法，这种聚类策略的真正潜力尚未得到充分发挥。

在现有的双聚类算法中，CC和FLOC算法目前被认为是处理基因表达数据最有效的工具。CC算法是最早出现、研究最多的算法，新兴算法大多基于CC算法的思想。CC算法使用双聚类的均方残差（MSR）作为相似性度量，通过贪心方式提取满足同质性约束的双聚类。它随机生成行簇和列簇，然后改进双聚类以使MSR值最小化。每次只识别一个双聚类，随后用随机数替换该双聚类，再识别下一个簇[16]。在此基础上，为改进通用的CC算法，Yang等人提出了另一种著名方法FLOC [7]，，其中增加了一个

K. Xu 任职于中国西安 710071 西安电子科技大学电子工程学院（电子邮件：kjxu@xidian.edu.cn）。

引入了一种函数，用于处理缺失数据并发现重叠双聚类[17]。后续研究表明，均方残差（MSR）仅适用于识别某些类别的共表达基因，不足以检测其他转录共调控基因。随后提出了另一种著名算法 QUBIC [15]，它能够以更通用的形式解决双聚类问题。事实证明，QUBIC 算法能够有效且高效地识别其他双聚类算法无法识别的所有统计显著的双聚类，是识别共调控基因的更有用工具。

基于公理模糊集（AFS）理论[18], ，本文提出了一种用于识别共调控基因的新型双聚类模型。AFS理论为如何将数据转换为模糊集（隶属函数）并实现其模糊逻辑运算提供了一种方法，从而为表示人类知识和模拟人类认知过程提供了灵活而强大的工具。近年来，AFS理论受到了越来越多的关注[18]。AFS理论将随机性的不确定性与模糊性的非精确性视为一个统一且连贯的过程，因此隶属函数可由观测数据确定。在AFS中，模糊集（隶属函数）及其逻辑运算依据原始数据的分布（AFS结构和AFS代数），通过一致的算法以客观且自动的方式确定；这与传统模糊集有很大不同，传统模糊集的隶属函数通常凭借个人直觉给出，逻辑运算则通过某种三角范数（t-范数）实现。其中的对象属性可以是各种数据类型或子偏好关系，甚至可以是人类直觉描述；不需要距离函数和目标函数，也不需要关于数据集的任何先验知识。对于维度很高且包含大量基因的数据，仅凭个人直觉定义隶属函数，并定义基于距离的函数来实现模糊逻辑运算，是不可能或非常困难的。因此，本文基于AFS理论设计了一种双聚类算法，用于发现共调控基因。

从设计角度来看，公理模糊集的子偏好关系理论仅依据原始基因表达数据的分布构建模糊隶属度（**划分**）矩阵，无需使用任何距离度量，也不需要有关基因表达数据的先验知识。具体而言，在所提出的方案中，首先选取一个参考基因。随后，将基因视为论域、将条件视为概念，利用子偏好关系理论构建一个模糊划分矩阵 [18]。类似地，将基因视为概念、将条件视为论域时，也可以构建另一个模糊划分矩阵。基于这两个划分矩阵，我们定义了一种基于模糊集的相似性准则，用于度量特定条件下共调控基因之间的相似性。随后，我们设计了一种循环优化算法来发现双聚类（共调控基因）。据我们所知，这是首次提出此类基于模糊集的相似性准则，也是首次将其用于解决双聚类问题。此外，我们还提出了一种基于模糊 C 均值（FCM）[19-20] 聚类的方法，用于选取若干参考基因。在真实世界基因表达数据上开展的实验研究表明，与几种用于基因表达分析的知名方法相比，所提出的方法取得了更好的性能。

简而言之，本文的主要贡献是提出一种新的双聚类算法，用于发现共调控基因。该算法不需要关于基因表达数据的任何先验知识。据我们所知，先前研究中尚未提出过本文方法的构想。

本文的概念结构如下。双聚类的新相似性定义和所提方法的原理将在第二节中介绍。第三节讨论性能指标。第四节介绍实验设置，并分析已完成的实验。第五节总结全文。

## II. A 模糊双聚类

### *A.* 问题定义

一种常用的基因表达分析微阵列数据可视化方法，是将数据集表示为一个矩阵，其中行代表基因，列代表条件（反之亦可），矩阵中的每个元素表示某个基因在特定条件下的表达值。因此，在微阵列数据集中识别出在待识别条件下具有相似表达模式的基因组，等同于寻找具有相似属性的子矩阵。

设 为一个微阵列表达数据矩阵，其中包含一组基因和一组条件 ,，其中 表示基因 在条件 下的表达值。 双聚类本质上是一个子矩阵 ,，其呈现某种相似趋势，可表示为 ,，其中和 分别是基因和条件的子集。设 为一个参考基因，我们的目标是找到一组与 相关的基因（共调控基因、双聚类）。 当参考基因未知时，我们可以枚举矩阵中的所有基因，或随机选取若干基因作为参考基因子集。类似的思想也用于 [21]。

### *B.* 相似性定义

考虑上述基因表达数据矩阵。首先，我们使用AFS理论构建两个模糊划分矩阵[22]，用于定义本文中的相似性矩阵。其中一个矩阵以基因为论域、以条件为概念，另一个矩阵以基因为概念、以条件为论域。假设这两个模糊划分矩阵为和。基于AFS的隶属度计算方法如下

[18, 23]：

$$\mu(x) = \sup_{i \in I} \left\{ \frac{\mathcal{M}\left[A_i(x)\right]}{\mathcal{M}(x)} \right\} \tag{1}$$

其中，是 $x \in X$ e, $X$ 是在 $\sigma$-代数上的有限正测度，是一个非空

指标集 [23]。

对于一个基因表达数据矩阵和给定的参考基因，定义，。显然，可以刻画第 $i$ 个基因与参考基因在第 $j$ 个条件下的相似性；越小，相似性越大，反之亦然。

此外，被用作另一种相似性度量，以共同发现局部共调控基因，这就是所谓的所提出的双相似性。

设为一个基因相似性矩阵，为的一个双聚类（子矩阵）。对于列，我们将 $j$ 列在中的差异性得分定义为该列的极差，即：

$$\mu(I,j)=\sum_{i\in I}\left[\max\left(\mu_{I,j}\right)-\min\left(\mu_{I,j}\right)\right] \tag{2}$$

差异性 ity score of $U_{Gg^*}(I,J)$ is

$$\mu(I,J)=\frac{1}{J}\sum_{(i\in I),j=1}^{J}\left[\max\left(\mu_{I,j}\right)-\min\left(\mu_{I,j}\right)\right] \tag{3}$$

设矩阵，且为的一个双聚类（子矩阵）。对于行，我们将 $i$ 行在中的差异性得分定义为该行的极差，即：

$$u(i,J)=\sum_{j\in J}\left[\max\left(u_{i,J}\right)-\min\left(u_{i,J}\right)\right] \tag{4}$$

差异性 ity score of is

$$u(I,J)=\frac{1}{I}\sum_{(j\in J),i=1}^{I}\left[\max\left(u_{i,J}\right)-\min\left(u_{i,J}\right)\right] \tag{5}$$

考虑性得分较低，并且基因在所有条件下的表达值相似。然而，许多基因会在某些特定条件下规律地共表达，换言之，挖掘局部共表达有价值模式比全局聚类更有意义。因此，我们使用和来联合发现局部共表达有价值模式。对于一个双聚类，如果和均较低，那么其中的基因在这些条件下共表达。

基于上述分析，我们将介绍所提出的双聚类发现算法。

*C.* 双聚类发现

该算法本质上是一种贪心算法，以整个基因表达数据矩阵作为初始双聚类。在发现双聚类的过程中，我们基于上述相似性矩阵定义了一种新颖的双相似性准则。使用该双相似性准则，可以增强同一双聚类中基因的共调控。

令和分别为和的差异度分数，我们的目标是发现一个差异度分数较小的双聚类，例如

和，其中和是该……的尺度因子

双聚类的列和行。因此，我们可以将该双聚类称为一个（$\alpha\beta$）-双聚类。通过按照某些特定规则删除和添加行与列，可以生成一个优秀的双聚类。示意图如下：

算法1（节点删除）

输入：和，基因表达数据矩阵的两个模糊划分矩阵，以及，两个待查找双聚类的列和行的尺度因子。

输出：，一个（$\alpha\beta$）-双聚类，它是的子矩阵，行集为，列集为，其差异度分数分别不大于，。

初始化：和被初始化为基因表达数据中的基因集和条件集，并且；参考基因由用户给定；列和行的最大可接受差异度分数为：，。

迭代：

1. 计算对所有，对所有，和，。如果且，则返回。
2. 找出值最大的列

$$\mu(I,j)=\sum_{i\in I}\left[\max\left(\mu_{I,j}\right)-\min\left(\mu_{I,j}\right)\right] \tag{6}$$

和行 $i\in I$ with largest

$$u(i,J)=\sum_{j\in J}\left[\max\left(u_{i,J}\right)-\min\left(u_{i,J}\right)\right] \tag{7}$$

移除该列 umn if

$$\frac{\alpha\mu(I,j)}{\mu(N,M)}>\frac{\beta u(i,J)}{u(N,M)} \tag{8}$$

否则，通过更新或来移除该行。

显然，删除节点后，子矩阵的行和列不相似度评分都会降低。然而，得到的 ($\alpha\beta$)-双聚类可能不是极大的，也就是说，在不增加差异度分数的情况下，仍可添加某些行和列。因此，我们设计了另一种算法来优化双聚类。

算法2（节点添加）输入：，一个实数矩阵的子矩阵

值，并表示一个 ($\alpha\beta$)-双聚类。

输出：，和使得和具有以下性质

$$\mu(I',J')\le\mu(I,J)\ \&\ u(I',J')\le u(I,J) \tag{9}$$

迭代：

1. 对所有，并重新计算和，并在满足时添加这些列。

$$\mu(I,J)\le\frac{\mu(N,M)}{\alpha}\ \&\ u(I,J)\le\frac{u(N,M)}{\beta} \tag{10}$$

2. 计算和，并在满足条件时添加行

$$\mu(I,J)\le\frac{\mu(N,M)}{\alpha}\ \&\ u(I,J)\le\frac{u(N,M)}{\beta} \tag{11}$$

3. 如果什么都没有以及作为和 ngs added in the iteration, return the final $I$ $J$ $I'$ $J'$

显然，执行节点添加算法后，行不相似度评分和列不相似度评分都不会增加。有时，添加操作带来的评分下降幅度可能大于任何删除操作。

*D.* 参考基因的选择

在上述算法中，我们感兴趣的参考基因是预先已知的。当参考基因未知时，我们应选择若干基因作为参考基因。在某些情况下，所选参考基因与双聚类的质量密切相关。此外，我们通常希望双聚类的规模（共调控基因和条件的数量）尽可能大（即在更多条件下发现更多共调控基因）。换句话说，如果一个基因在更多条件下有更多相似基因，那么它就更适合作为参考基因。基于此，我们提出一种选择参考基因的方法。

首先，我们计算所有基因在每个（$j$个）条件下的模糊相似度矩阵，并得到 个相似性矩阵，例如， $M$

$$\boldsymbol{S}_j=\left[\boldsymbol{s}_1^{(j)},\boldsymbol{s}_2^{(j)},\mathrm{L}\ ,\boldsymbol{s}_i^{(j)}\ \mathrm{L}\ ,\boldsymbol{s}_N^{(j)}\right]$$
$$\boldsymbol{s}_i^{(j)}=\left[s_1^{(ji)},s_2^{(ji)},\mathrm{L}\ ,s_k^{(ji)},\mathrm{L}\ ,s_N^{(ji)}\right]^T \tag{12}$$
$$i=1,2,\mathrm{L}\ ,N;\ \ j=1,2,\mathrm{L}\ ,M;\ \ k=1,2,\mathrm{L}\ ,N$$

其中$T$为.....stands for...the 令为...for the，transpose operat 我们将其转换为如下的隶属度矩阵：

$$\boldsymbol{\Phi}_i^{(j)}=\left[\varphi_{k1}^{(ji)}\ \ \varphi_{k2}^{(ji)}\right]^T\in R^{2\times N}$$
$$\varphi_{kc}^{(ji)}=\frac{\left\|s_k^{(ji)}-v_c^{(ji)}\right\|^{\frac{-2}{m-1}}}{\sum_{h=1}^{2}\left(\frac{1}{\left\|s_h^{(ji)}-v_c^{(ji)}\right\|}\right)^{\frac{2}{m-1}}} \tag{13}$$
$$c=1,2;\ \ k=1,2,\mathrm{L}\ ,N$$

其中 是$c$的第.....个元素（聚类原型）； $m$ 是模糊指数（模糊系数）； 表示欧氏距离。 是个体 属于聚类$c$的隶属度，并满足以下条件

$$\sum_{c=1}^{2}\varphi_{kc}^{(ji)}=1\quad\text{for }k=1,2,\mathrm{L}\ ,N \tag{14}$$

然后，我们构造一个函数，用于计算第i个基因在所有M个条件下的大模糊相似度的均值。

$$\varpi_i=\frac{\sum_{j=1}^{M}\sum_{k=1}^{N}s_k^{(ji)}sign\left[\varphi_{k1}^{(ji)}-0.5\right]}{\delta\left\{\sum_{k=1}^{N}sign\left[\varphi_{k1}^{(ji)}-0.5\right]\right\}+M\sum_{k=1}^{N}sign\left[\varphi_{k1}^{(ji)}-0.5\right]} \tag{15}$$

其中 是单位脉冲响应函数。通常，我们认为具有较大 值的基因更适合作为参考基因

$\delta$ $\varpi$

因为如前所述，它们在更多条件下具有更多相似的基因。

## III. 性能指标

为了评估所提出算法的性能

简要讨论了两种常用的性能指标。

*A.* 方差指标

给定一个基因表达数据矩阵，其中包含 一组行（基因）和一组列（条件），双聚类是一个子矩阵，它是 的 部分。 是数据矩阵中 对应于第$i$行和第$j$列的值。 我们用 表示双聚类中第$i$行的均值，用 表示双聚类中第$j$列的均值，并用表示双聚类中所有元素的均值。这些值定义如下：

$$a_{iJ}=\frac{1}{|J|}\sum_{j\in J}a_{ij} \tag{16}$$

$$a_{Ij}=\frac{1}{|I|}\sum_{i\in I}a_{ij} \tag{17}$$

$$a_{IJ}=\frac{1}{|I||J|}\sum_{i\in I,\,j\in J}a_{ij}=\frac{1}{|I|}\sum_{i\in I}a_{iJ}=\frac{1}{|J|}\sum_{j\in J}a_{Ij} \tag{18}$$

每个双聚类的方差 is used to evaluate the quality o

$$VAR(I,J)=\sum_{i\in I,\,j\in J}\left(a_{ij}-a_{IJ}\right)^2 \tag{19}$$

该值越低，双聚类的质量越好 the better the qua 完美的双聚类是方差等于零的子矩阵。

*B.* 平均波动度指标

平均波动度（MFD）用于评估基因在每种条件转变下的变化趋势。双聚类的平均波动度定义为

$$MFD(I,J)=\sqrt{\frac{1}{|I||J|}\sum_{i\in I,\,j\in J}\left(\Theta_{ij}-\frac{1}{|I|}\sqrt{\Theta_{ij}}\right)^2} \tag{20}$$

其中

$$\Theta_{ij}\in\boldsymbol{\Theta}$$
$$\boldsymbol{\Theta}=180\arctan\left(\boldsymbol{\Delta}^{\dagger}\boldsymbol{\Xi}\right)/\pi \tag{21}$$

$$\boldsymbol{\Delta}=\frac{diag\left\{\max(a_{1j})-\min(a_{1j}),\mathrm{L}\ ,\max(a_{ij})-\min(a_{ij}),\mathrm{L}\right\}}{M-1} \tag{22}$$
$$i=1,2,\mathrm{L}\ ,N;\ \ j=1,2,\mathrm{L}\ ,M$$

$$\boldsymbol{\Xi}=[\mathrm{L}\ ,a_{ij}-a_{i(j-1)},\mathrm{L}\ ]\in R^{N\times(M-1)} \tag{23}$$
$$i=1,2,\mathrm{L}\ ,N;\ \ j=2,\mathrm{L}\ ,M$$

其中，上标 † 表示矩阵 [25] 的 Moore广义逆，即 Penrose。显然，对于一个双聚类，如果其中的基因（行）在每种条件转变下具有相似的变化趋势，其平均波动度就会相对较小。此外，如果双聚类中的所有基因（行）在每种条件转变下的变化趋势完全相似（或相同），其平均波动度将为零。

特别是，对于仅包含单行（或单列）的“双聚类”，其方差和 MFD 指标也为零，但这种 “ 双聚类 ” 没有意义。为了公平比较

算法的性能，我们将剔除结果中仅包含一行或一列的双聚类。

## IV. 实验研究

在以下实验中，我们将所提出的模糊双聚类（FBC）方法与 CC、FLOC 和 QUBIC 方法进行性能比较。这些方法是常用于基因表达分析的两种知名双聚类方法。实验采用了两个知名且可公开获取的真实微阵列数据集 Yeast（http://arep.med.harvard.edu/biclustering/yeast.matrix）和 Gordon 2002 [26]，它们是双聚类中最常用的数据集。方差 [24] 和平均波动度均作为评价指标，下面将对此进行简要讨论。

### *A.* 实验

这些方法用于寻找 100 个双聚类。表 I 简要列出了实验中所用参数的取值。这些方法重复运行 10 次；实验结果以均值和标准差表示。实验结果绘于图 1 至图 2。显然，所提出的算法通过将波动度趋势更相似的基因归为一组，能够有效发现高质量的双聚类。与 CC、FLOC 和 QUBIC 聚类方法相比，所提出的方法具有明显优势。平均改进幅度为 25%，最小改进幅度为 15%，而最显著情况下的改进幅度可达 35%。

表 I. 实验中使用的数据集和参数

| 数据集 | | 酵母 | Gordon-2002 |
|---|---|---|---|
| 数量 | 基因 | 2,884 | 1,626 |
| | 条件 | 17 | 181 |
| MSR阈值 | | 300 | 3,000 |
| α | | 5.0 | 5.5 |
| $\beta$ | | 1.8 | 3.0 |

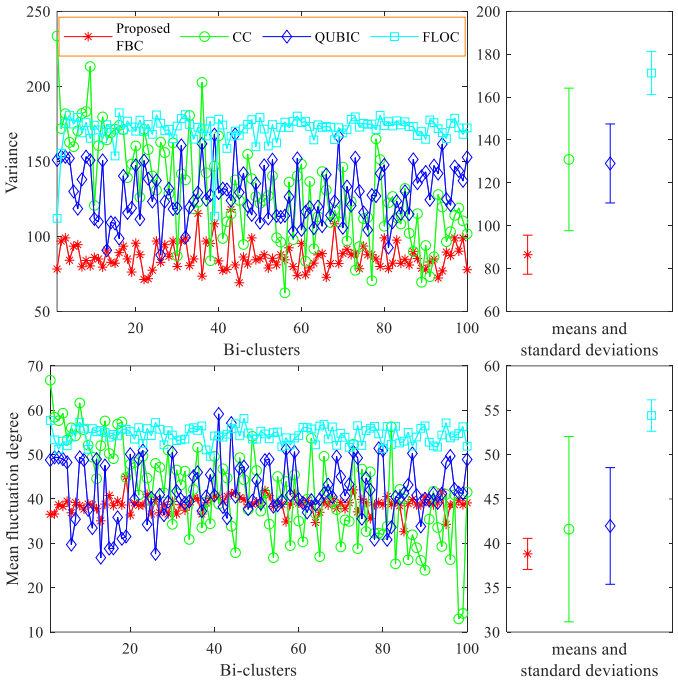


图 1. 各方法在酵母数据集上的性能比较

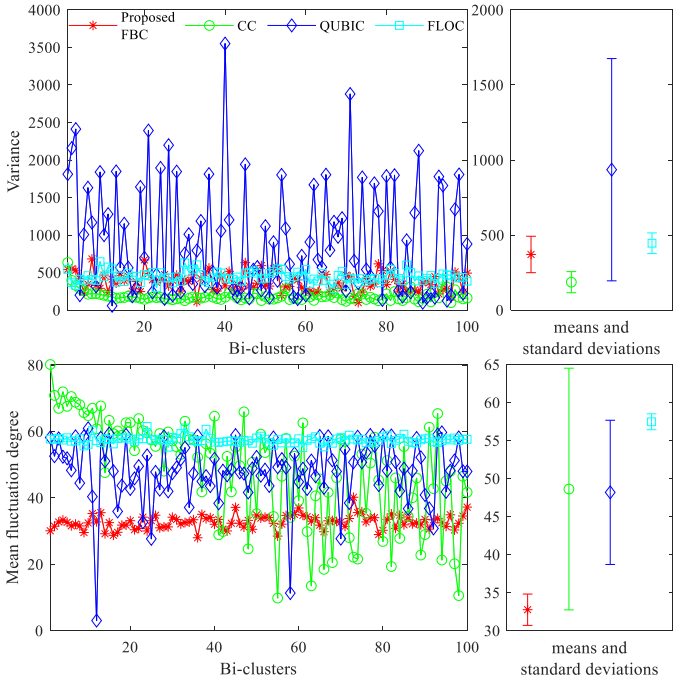


图 2. 各方法在 Gordon 2002数据集上的性能比较

## V. 结论

本研究设计了一种基于AFS理论的双聚类算法，用于识别共调控基因。在设计过程中，我们引入了AFS的子偏好关系理论，构建两个模糊隶属度矩阵，以定义一种基于模糊集的相似性准则。基于该相似性准则，我们设计了一种循环优化算法，用于发现双聚类（共调控基因）。我们进行了理论分析，并开展了一系列全面的实验。理论和实验结果均予以呈现，以验证所提方法的有效性。实验结果表明，所提方法

在发现双聚类方面优于现有算法，并展现出在基因表达研究中的卓越性能和巨大开发潜力。据我们所知，本研究方案是首次提出的，能够稳步提升双聚类的性能。目前，我们已完成理论分析，并开展了一系列全面的实验。未来研究可进一步探索一些实际实验，这将是一个有趣的方向。

## 参考文献


[1] P. Dhaeseleer，“基因表达聚类如何工作？”《自然·生物技术》，第23卷，第12期，第1499–1501页，2005年12月。[2] L. Pattini、R. Sassi和S. Cerutti，“通过系统医学的多尺度方法剖析心力衰竭”，IEEE 生物医学工程汇刊，第61卷，第5期，第1593–1603页，2014年5月。[3] R. M. Mulqueen、D. Pokholok和S. J. Norberg，“单细胞中DNA 甲基化谱的高度可扩展生成”，《自然·生物技术》，第36卷，第5期，第1499–1501页，2018年4月。[4] D. Mishra、K. Shaw和S. Mishra，“基因表达网络发现：一种基于模式的双聚类方法”，载于《2011年国际通信、计算与安全会议》论文集。ACM，印度奥里萨邦鲁尔克拉，第307–312页，2011年2月。[5] J. Yang、H. Wang和W. Wang，“表达数据上的增强双聚类”，载于《第三届 IEEE 生物信息学与生物工程研讨会》论文集，贝塞斯达，马里兰州，美国，美国，第321–327页，2003年3月。[6] Y. Cheng和G. M. Church，“表达数据的双聚类”，载于《分子生物学智能系统会议（ISM）》论文集，第93–103页，2000年1月。[7] J. Yang、H. X. Wang和W. Wang，“一种用于分析基因表达谱的改进双聚类方法”，《国际人工智能工具杂志》，第14卷，第5期，第771–789页，2005年10月。[8] L. Lazzeroni和A. Owen，“用于基因表达数据的 Plaid 模型”，《中华统计学志》，第12卷，第1期，第61–86页，2000年5月。[9] A. Ben-Dor、B. Chor和R. Karp，“发现基因表达数据中的局部结构：保序子矩阵问题”，《计算生物学杂志》，第10卷，第3期，第373–384页，2003年2月。[10] S. Bergmann、J. Ihmels和N. Barkai，“用于分析大规模基因表达数据的迭代特征签名算法”，《统计、非线性与软物质物理》，第67卷，第1期，第1–18页，2003年3月。[11] T. M. Murali和S. Kasif，“从基因表达数据中提取保守基因表达模体”，《太平洋生物计算研讨会》，第77–88页，2003年1月。[12] A. Prelic、S.Bleuler和P. Zimmermann，“基因表达数据双聚类方法的系统比较与评估”，《生物信息学》，第22卷，第9期，第1122–1129页，2006年5月。[13] C. Gao、I. C. McDowell和S. Zhao，“通过贝叶斯双聚类构建情境特异性和差异性基因共表达网络”，《计算生物学》，第12卷，第7期，第1–39页，2016年7月。[14] X. W. Liu和L.S. Wang，“计算基因表达数据的最大相似度双聚类”，《生物信息学》，第23卷，第1期，第50–56页，2007年1月。[15] G. Li、Q. Ma和H. Tang，“QUBIC：一种用于基因表达数据分析的定性双聚类算法”，《核酸研究》，第37卷，第15期，第e101页，2009年7月。[16] M. P. Shruthi和K. E. Saravana，“双聚类综述”，《国际科学与技术创新研究期刊》，第3卷，第5期，第2349–6010页，2016年10月。[17] B. Khalid和K. Allab，“利用自组织映射对连续数据进行双聚类”，《神经计算与应用》，第22卷，第7期，第1551–1562页，2013年6月。[18] X. D. Liu、W. J Jia和Y.G. Wang，“公理模糊集理论框架下的知识发现与语义学习”，《Wiley 跨学科评论：数据挖掘与知识发现》，第8卷，第5期，第1268–1292页，2018年9月。[19] C. Lian、S. Ruan和T. Denœux，“利用自适应距离度量进行空间证据聚类以实现FDG-PET 图像中的肿瘤分割”，IEEE 生物医学工程汇刊，第65卷，第1期，第21–30页，2018年1月。[20] K. J. Xu、W. Pedrycz、Z. W. Li和W. K. Nie，“构建虚拟空间以提升模糊聚类的分类性能”，IEEE模糊系统汇刊，doi：10.1109/tfuzz.2018.2889020，2018年。[21] X. W. Liu和L.S. Wang，“计算基因表达数据的最大相似度双聚类”，《生物信息学》，第23卷，第1期，第50–56页，2007年1月。[22] K. J. Xu、W. Pedrycz、Z. W. Li和W. K. Nie，“基于高斯核的高精度信号子空间分离算法”，IEEE工业电子学汇刊，第66卷，第1期，第491–499页，2019年1月。[23] Y. Ren、M. Song和X. Liu，“基于 AFS理论的模糊聚类新方法”，《国际信息与系统科学杂志》，第3卷，第2期，第307–325页。2007年1月。[24] S. C. Madeira和A. L. Oliveira，“生物数据分析的双聚类算法：综述”，IEEE/ACM计算生物学与生物信息学汇刊（TCBB），第1卷，第1期，第24–45页，2004年1月。[25] D. Stanev和K. Moustakas，“任务空间中受约束的肌肉骨骼系统仿真”，IEEE 生物医学工程汇刊，第65卷，第2期，第307–318页，2017年10月。[26] X. Li和K. C. Wong，“进化多目标聚类及其在患者分层中的应用”，IEEE控制论汇刊，DOI：10.1109/TCYB.2018.2817480。