

硕士学位论文

基于群体智能的基因表达数据双聚类研究

**RESEARCH ON BICLUSTERING OF
GENE EXPRESSION DATA BASED ON
SWARM INTELLIGENCE**

凡振豪

西南大学

2019 年 12 月

国内图书分类号：TM301.2
国际图书分类号：62-5

学校代码：10635
密级：公开

工学硕士学位论文

基于群体智能的基因表达数据双聚类研究

硕士研究生：凡振豪

导 师：欧灵副教授

申 请 学 位：工学硕士

学 科：计算机软件与理论

所 在 单 位：计算机与信息科学学院

答 辩 日 期：2019 年 12 月

授予学位单位：西南大学

Classified Index: TM301.2

U.D.C: 62-5

Dissertation for the Master's Degree in Engineering

RESEARCH ON BICLUSTERING OF GENE EXPRESSION DATA BASED ON SWARM INTELLIGENCE

Candidate:	FAN Zhenhao
Supervisor:	OU Ling
Academic Degree Applied for:	Master of Engineering
Specialty:	Computer Architecture
Affiliation:	College of Computer and Information Science
Date of Defence:	December, 2019
Degree-Conferring-Institution:	Southwest University

摘 要

高通量基因微阵列技术的出现,产生了大量的基因表达数据。这些数据在追踪生物过程,基因规则发现以及病理分析中有着至关重要的作用。基因表达数据的双聚类是指,找出在某些条件子集下包含一致表达波动的基因子集。双聚类可以看作是一种优化问题。针对表达数据高维度,高冗余的特点,许多群智能算法被用于双聚类中来。本文对粒子群算法、量子粒子群算法,萤火虫算法以及布谷鸟算法在四个数据集上进行了比较研究。实验结果表明了量子粒子群算法与布谷鸟算法具有更好的全局寻优能力。最后基于基因本体工程,对双聚类结果给出了生物学解释。

关键词: 群智能算法; 基因表达数据; 双聚类; 多目标优化

Abstract

The advent of high-throughput gene microarray technology has generated a large amount of gene expression data. These data play a vital role in tracking biological processes, discovering genetic rules, and analyzing pathology. Biclustering of gene expression data refers to finding a subset of genes that contain consistent expression fluctuations under certain conditional subsets. Biclustering can be considered as an optimization problem. In view of the high dimensionality and high redundancy of expression data, many swarm intelligence algorithms are used in biclustering. In this paper, particle swarm algorithm, quantum particle swarm algorithm, firefly algorithm and cuckoo algorithm are compared and researched on four data sets. The experimental results show that the quantum particle swarm algorithm and cuckoo algorithm have better global optimization capabilities. Finally, based on the gene ontology engineering, a biological explanation is given for the results of the biclustering.

Keywords: Swarm intelligence Algorithm, Gene Expression Data, Biclustering, Multiple-optimistic

目 录

摘 要	I
ABSTRACT	III
第 1 章 绪论	1
1.1 研究背景及意义	1
1.2 基因表达数据的研究现状	1
1.3 本文的研究内容及组织结构	1
第 2 章 相关算法概述	2
2.1 双聚类算法	2
2.1.1 双聚类算法的发展现状	2
2.1.2 CC 算法	2
2.1.3 FLOC 算法.....	2
2.1.4 双聚类结果评价	2
2.2 群智能算法	2
2.2.1 群智能算法的发展现状	2
2.2.2 布谷鸟算法	2
2.2.3 萤火虫算法	2
2.2.4 细菌觅食算法算法	2
2.3 本章小结	2
第 3 章 基于 CS-FA 的双聚类算法	3
3.1 算法描述	3
3.2 实验结果及分析	3
3.3 本章小结	3
第 4 章 基于 BFO 的多目标优化双聚类算法	4
4.1 算法描述	4
4.2 实验结果及分析	4
4.3 本章小结	4
第 5 章 总结与展望	5

第 1 章 绪论

- 1.1 研究背景及意义
- 1.2 基因表达数据的研究现状
- 1.3 本文的研究内容及组织结构

第 2 章 相关算法概述

2.1 双聚类算法

2.1.1 双聚类算法的发展现状

2.1.2 CC 算法

2.1.3 FLOC 算法

2.1.4 双聚类结果评价

2.2 群智能算法

2.2.1 群智能算法的发展现状

2.2.2 布谷鸟算法

2.2.3 萤火虫算法

2.2.4 细菌觅食算法算法

2.3 本章小结

第 3 章 基于 CS-FA 的双聚类算法

3.1 算法描述

3.2 实验结果及分析

3.3 本章小结

第 4 章 基于 BFO 的多目标优化双聚类算法

4.1 算法描述

4.2 实验结果及分析

4.3 本章小结

第 5 章 总结与展望