

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/57983> holds various files of this Leiden University dissertation.

Author: Cai, F.

Title: Fuzzy systems and unsupervised computing: exploration of applications in biology

Issue Date: 2017/12/12

论文扼要 (Chinese Summary)

本文主要研究了模糊系统在生物信息学领域内的应用。在过去的几十年里，越来越多的科研工作者将目光投入到生物信息学领域。因为在生物学领域，生物实验数据需要被精准而又不失鲁莽地分析；而在计算机领域，对不同类型的数据进行有价值的分析和开发以及设计出高效的处理算法是具有挑战的研究难点。

通常生命科学等领域研究得到的数据不精确或存在缺失，使其具有模糊性和未知性。针对生物数据的特性，模糊系统被应用到相关的研究中。模糊系统是一种较新的启发式算法，相较于其他更加关注计算结果精准度的算法，其更加注重计算结果的重要性。模糊系统是通过在分析数据时提出更多的假设条件，从而有效的简化分析数据中的复杂且非唯一性问题。

本文着重于设计高效可靠的启发式分析算法去挖掘并理解隐藏在生物实验数据中的重要信息。本文设计了一个基于模糊系统的数据分析流程，并在文中不同章节处详细阐述了模糊系统在不同场景的应用及其前景。最后，通过案例分析来展示本文所提出的数据处理的流程。本文亮点如下：

(1) 本文使用模糊算法处理图像背景的噪声。通常，显微成像的数字图像会受到例如电子设备引起的工频噪声，或者实验过程中光照不均带来的叠加伪影的影响。所以，为了解决以上问题，一种新的基于图像形态学和模糊及粗糙系统的图像背景去除算法在本文第二章中被提出。与大多数使用的 EMI 方法以及商业软件中频繁使用的滚球算法相比，本文提出的 DCBC 算法被验证对于传统生物显微图像更具有鲁棒性。

(2) 本文对模糊系统在数据中的特征提取和冗余信息剔除做了学习和研究。这个过程对于数据分析中的表象型数据到经典的统计数据模型的转换是至关重要的：一，该过程可以选择最少的特征去最优地描述整个数据集；二，通过去除冗余的信息来降低数据集的维度，从而一定程度上减少后期分析的计算量。在本文第三章中，一个基于模糊标准函数的“过滤”并“包装”模式的特征提取算法被提出。与此同时，文中还采用了多目标优化函数来帮助原算法在“帕洛托”前沿寻找最优解析解。本文所提出的算法与一般特征提取算法不同之处在于，文中的算法可以给出一组最优解的解集让决策者根据自己的经验、需求选择相对应的可以描述原数据的特征集合。

(3) 本文对模糊系统在信息聚类分析中的应用进行了研究。因为在所有模式提取流程中，即信息概述、信息联合、信息预测和信息聚类等，其中信息聚类是至关重要的一个环节；而且由于生活中的数据大多数是不带有或是带有少量标签信号，所以该聚类分析的方法论在科学研究及日常生产管理中都有着十分

广泛的应用。在本文第四章，通过结合模糊粗糙聚类算法和粒子群算法，提出了一种全新的且对于获得全局最优解非常敏锐的聚类算法。在这个新的算法中，模糊粗糙的概念可以有效地处理原数据中不确定的、模糊的以及重叠覆盖的成分；与此同时，粒子群理论又可以帮助聚类算法找到所有不同聚类结果中的最优解。

(4) 本文基于前几章的研究内容与所设计的算法，整合并提出了一个新的以模糊系统为蓝本的数据处理流程。该处理流程包括图像的预处理（第二章），图像特征提取和选择分析（第三章）以及数据的聚类分析（第四章），并且该分析流程的可行性研究将通过一个特有的案例进行分析。在本文第五章，通过对一维电泳胶体成像数据的分析和研究来进一步了解文中提出的模糊分析系统对癌症细胞（蛋白和 RNA 等）数据处理流程的帮助。对这些电泳胶体成像结果的处理和分析有助于提取，分析和鉴别不同的细胞系（类）在图像中的表达特征。而通过研究图像中细胞蛋白/RNA 的位移和表达变化可以帮助理解并预测案例中关于癌症细胞的发展。

综上所述，本篇论文旨在研究与模式识别相结合的模糊系统在生物信息学领域中的应用。通过对已有的文献的考究，结合模糊逻辑与无监督学习的概念搭建了一个高效且鲁莽的数据处理流程平台，并在文末的案例中得以用来理解研究以生物学或模式识别为导向的问题。