

论文与报告

一种用于蛋白质结构聚类的聚类中心选择算法

黄旭¹, 吕强^{1,2}, 钱培德^{1,2}

1. 苏州大学计算机科学与技术学院 苏州 215006;
2. 江苏省计算机信息处理技术重点实验室 苏州 215006

收稿日期 2010-9-7 修回日期 2010-12-27 网络版发布日期 接受日期

摘要

提出一种对蛋白质结构聚类中心进行选择的算法。聚类是蛋白质结构预测过程中必不可少的一个后处理步骤,而目前在蛋白质结构预测中常用的属性阈值(Quality threshold, QT)聚类算法依赖于由经验得出的聚类半径;其他聚类算法,如近邻传播(Affinity propagation, AP)聚类算法也存在影响聚类分布的参数。为克服对主观经验参数的依赖,本文提出一种聚类中心选择算法(Exemplar selection algorithm, ESA),用于对不同参数下的聚类结果进行分析,从而选择最佳聚类中心,进而确定聚类半径等经验参数。该算法在真实蛋白质结构数据集上进行了实验,在未知经验参数情况下选择出最佳聚类中心,同时也为不同聚类算法寻找适合相应数据集的客观聚类参数提供了支持。

关键词 [蛋白质结构](#) [聚类](#) [属性阈值](#) [近邻传播](#) [聚类中心选择](#)

分类号

An Exemplar Selection Algorithm for Protein Structures Clustering

HUANG Xu¹, LV Qiang^{1,2}, QIAN Pei-De^{1,2}

1. School of Computer Science and Technology, Soochow University, Suzhou 215006;
2. Jiangsu Provincial Key Laboratory for Computer Information Processing Technology, Suzhou 215006

Abstract

This paper proposes an exemplar selection algorithm (ESA) for protein structures clustering, which is a necessary post-processing step for protein structure prediction. The widely-used quality threshold (QT) algorithm in protein structure prediction depends on clustering radius derived from experience, which also affects clustering distribution in other widely-used clustering algorithms such as affinity propagation (AP). The proposed exemplar selection algorithm can analyze clustering results, choose the best exemplar, and confirm clustering parameter such as clustering radius. Experimental results on real protein structure predictions confirm the effectiveness of our exemplar selection algorithm, which can choose the best exemplar with no experience parameter, and can find the best parameter fitting for data set.

Key words [Protein structure](#) [clustering](#) [quality threshold \(QT\)](#) [affinity propagation \(AP\)](#) [exemplar selection](#)

DOI: 10.3724/SP.J.1004.2011.00682

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(539KB\)](#)
- ▶ [\[HTML全文\]\(OKB\)](#)
- ▶ [参考文献\[PDF\]](#)

参考文献

服务与反馈

- ▶ [把本文推荐给朋友](#)
- ▶ [加入我的书架](#)
- ▶ [加入引用管理器](#)
- ▶ [复制索引](#)

Email Alert

相关信息

- ▶ [本刊中包含“蛋白质结构”的相关文章](#)
- ▶ 本文作者相关文章

- [黄旭](#)
- [吕强](#)
- [钱培德](#)
-

通讯作者

作者个人主页 [黄旭¹; 吕强^{1,2}; 钱培德^{1,2}](#)