

论文

统计显著性标记的聚类分析算法与网络实现

张文军,冯永军,古德祥

中山大学昆虫学研究所与生物防治国家重点实验室;中山大学昆虫学研究所与生物防治国家重点实验室;中山大学昆虫学研究所与生物防治国家重点实验室 广州 510275 ;广州 510275 ;广州 510275

摘要:

聚类分析方法应用广泛,但过程及结果缺乏可靠的统计学检验,数学上不严格.另外,用于聚类分析的数据分布类型复杂多样,往往无法确定,而经典统计检验方法设定了各种统计前提和假设,应用依据不足.鉴于此,本研究用随机化方法对分类进行统计显著性检验,建立了具有统计显著性标记的聚类分析算法,用于对若干个样品进行有显著性标记的聚类分析.该算法包括数据加权与规范化,计算距离测度,系统聚类,及随机化统计检验等过程.在该算法中,有14种距离测度、5种系统聚类方法、3种数据规范化方法及指标加权与否可供选择.随机化检验不需统计前提和假设,适用于各种统计问题.算法用Java语言网络化实现,包含6个类和一个HTML文件.可通过网络在多种Java兼容的浏览器上实现算法共享.以水稻田无脊椎动物多样性的调查数据,对该算法进行了对比分析,给出了选择距离测度的一些原则.

关键词:

ALGORITHM AND IMPLEMENTATION OF A CLUSTER ANALYSIS WITH STATISTIC SIGNIFICANCE

Zhang Wenjun Feng Yongjun Gu Dexiang (Research Institute of Entomology and State Key Laboratory for Biological Control, Zhongshan University, Guangzhou 510275)

Abstract:

Algorithms of cluster analysis are not always statistically tested. An algorithm of cluster analysis with statistic significance test to classification was developed in this paper. It was made of four parts, i.e., data weighting and standardization, calculation of distance measures, hierarchical clustering, and randomization statistic test. Fourteen distance measures and five methods of hierarchical clustering were provided in the algorithm to be choosed. The algorithm was implemented as the network program with Java language, which made of 6 Java classes and a HTML file and can be loaded and run on Java-enabled web browsers. The algorithm was tested with investigation data of rice invertebrate diversity. Principles for choosing distance measures was explained.

Keywords:

收稿日期 修回日期 网络版发布日期

DOI:

基金项目:

通讯作者:

作者简介:

本刊中的类似文章

扩展功能

本文信息

Supporting info

PDF(412KB)

[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

引用本文

Email Alert

文章反馈

浏览反馈信息

本文关键词相关文章

本文作者相关文章

PubMed