

数据库、信号与信息处理

模糊策略下的搜索文本聚类分析技术

万红新¹, 彭云²

1.江西科技师范学院 数学与计算机科学学院, 南昌 330013

2.江西师范大学 计算机信息工程学院, 南昌 330022

收稿日期 2008-6-30 修回日期 2008-9-12 网络版发布日期 2009-12-4 接受日期

摘要 在现有的搜索文本中, 存在大量的不确定文本结构和内容, 使得常规的聚类算法难以实现, 并且文本搜索的结果没有进行类聚, 造成搜索结果集合数据量非常庞大。提出了基于模糊集的文本搜索的聚类分析的方法, 通过模糊技术对异构数据进行处理, 可以改善算法实现的时间和空间的复杂度, 减少文本处理的维度, 提高算法的鲁棒性, 对算法的实现给出了实例分析。通过与其他聚类算法的实测数据的比对分析, 验证了算法实现的精确性和效率性。

关键词 [聚类分析](#) [文本挖掘](#) [模糊集](#) [隶属函数](#)

分类号 [TP18](#) [TP311](#)

Technique of searching text clustering analysis based on fuzzy set

WAN Hong-xin¹, PENG Yun²

1.Mathematics & Computer Science College, Jiangxi Science & Technology Normal University, Nanchang 330013, China

2.College of Computer Information and Engineering, Jiangxi Normal University, Nanchang 330022, China

Abstract

There are a large number of non-certain and non-structure contents in the web searching text. It is difficult to cluster the searching text by some normal classification methods. Because the searching text is not clustered, the searching result of the text is very enormous. A technique of searching text clustering analysis based on fuzzy set is proposed, and the algorithm has been described in detail by example. It can improve the algorithm complexity of time and space, decrease the dimensions of the algorithm, which should increase the robustness of the algorithm. To check the accuracy and efficiency of the algorithm, the comparative analysis of the sample and test data is provided.

Key words [clustering analysis](#) [text mining](#) [fuzzy set](#) [membership function](#)

DOI: 10.3778/j.issn.1002-8331.2009.33.044

通讯作者 万红新 wanhongxin@126.com

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(504KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

相关信息

▶ 本刊中 [包含“聚类分析”的相关文章](#)

▶ 本文作者相关文章

· [万红新](#)

· [彭云](#)