

- >> 首页
- >> 被收录信息
- >> 投稿须知
- >> 模板下载
- >> 信息发布
- >> 常见问题及解答
- >> 合作单位
- >> 产品介绍
- >> 编委会/董事会
- >> 关于我们
- >> 网上订阅
- >> 友情链接

#### 友情链接

- >> 中国期刊网
- >> 万方数据资源库
- >> 台湾中文电子期刊
- >> 四川省计算应用研究中心
- >> 维普资讯网

## 结合LSA的中文谱聚类算法研究

### Research of Chinese spectral clustering with LSA

摘要点击: 29 全文下载: 10

[查看全文](#) [查看/发表评论](#) [下载PDF阅读器](#)

中文关键词: [文本聚类](#) [潜在语义分析](#) [奇异值分解](#) [谱聚类](#)

英文关键词: [text clustering](#) [LSA](#) [SVD](#) [spectral cluster](#)

基金项目:

作者

[熊忠阳](#), [暴自强](#), [李智星](#), [张玉芳](#)

单位

[\(重庆大学 计算机学院, 重庆 400044\)](#)

中文摘要:

传统的文本谱聚类需要的文本相似矩阵依赖于向量空间模型, 忽略了词与词之间的语义关系, 存在词频维数过高、计算代价高等问题。针对这些问题, 提出了一种基于潜在语义分析 (latent semantic analysis, LSA) 的文本相似矩阵构造方法, 利用奇异值分解 (singular value decomposition, SVD) 降维, 在低维的语义空间表示文本, 以此来提高同类文本间的语义相似度, 并进行了相关对比实验。在该实验中, 改进方法的聚类效果要好于传统的方法, 从而验证了改进方法的有效性和可行性。

英文摘要:

Traditional text samples similarity matrix for spectral cluster heavily rely on the vector space model which ignores the semantic relationship among terms. It will give rise to problems such as curse of dimensionality, feature redundancy and high computing cost. To solve the problems above, this paper proposed a new method based on LSA to solve it, which used SVD to lowering rank of matrices. The experimental results turn out that the new method enhances the cluster accuracy and less the data-process elapsed time.

您是第2826912位访问者

主办单位: 四川省计算机研究院 单位地址: 成都市武侯区成科西路3号

服务热线: 028-85249567 传真: 028-85210177 邮编: 610041 Email: arocmag@163.com

蜀ICP备05005319号 本系统由北京勤云科技发展有限公司设计