

本期目录 | 下期目录 | 过刊浏览 | 高级检索

[打印本页] [关闭]

人工智能及识别技术

基于非负矩阵分解的双重约束文本聚类算法

马慧芳¹, 赵卫中², 史忠植³

(1. 西北师范大学数学与信息科学学院, 兰州 730070; 2. 湘潭大学信息工程学院, 湖南 湘潭 411105; 3. 中国科学院计算技术研究所智能信息处理重点实验室, 北京 100190)

摘要: 提出一种基于非负矩阵分解(NMF)的双重约束文本聚类算法。在正交三重NMF模型中, 加入文本空间的成对约束信息和词空间的类别约束信息, 将不同的特征词项进行分类。利用迭代规则对原始的词-文档矩阵进行分解, 获得文本聚类结果。与多种传统半监督文本聚类算法的对比结果表明, 该算法具有较高的聚类精度, 能提供更准确和有效的聚类结果。

关键词: 半监督聚类 非负矩阵分解 成对约束 类别约束

Dual-constraints Text Clustering Algorithm Based on Non-negative Matrix Factorization

MA Hui-fang¹, ZHAO Wei-zhong², SHI Zhong-zhi³

(1. College of Mathematics and Information Science, Northwest Normal University, Lanzhou 730070, China; 2. College of Information Engineering, Xiangtan University, Xiangtan 411105, China; 3. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Non-negative Matrix Factorization(NMF) with dual constraints method for document clustering is proposed. It is based on NMF model with adding of pair-wise constraints on documents and categorization constraints of the words. Iterative rules obtained from the original word-document matrix are decomposed to get document clustering results. Compared with a variety of popular semi-supervised clustering algorithm, the method for document clustering can effectively improve the accuracy of document clustering, and can provide more accurate and efficient clustering results.

Keywords: semi-supervised clustering Non-negative Matrix Factorization(NMF) pairwise constraint category constraint

收稿日期 2011-03-15 修回日期 网络版发布日期 2011-12-20

DOI: 10.3969/j.issn.1000-3428.2011.24.054

基金项目:

国家自然科学基金资助项目(61105052, 61163039); 西北师范大学青年教师科研能力提升计划基金资助项目“面向Web的主题建模关键技术研究”(NWNLU-LKQN-10-1); 湘潭大学博士启动基金资助项目(10QDZ42)

通讯作者:

作者简介: 马慧芳(1981—), 女, 副教授、博士, 主研方向: 数据挖掘, 文本聚类, 机器学习; 赵卫中, 讲师、博士; 史忠植, 研究员、博士生导师

通讯作者E-mail: mahuifang@yeah.net

参考文献:

- [1] Lee D, Seung H. Learning the Parts of Objects by Non-negative Matrix Factorization [J]. Nature. 1999, 401(6755): 788-797 

本刊中的类似文章

扩展功能

本文信息

Supporting info

PDF(409KB)

[HTML] 下载

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

半监督聚类

非负矩阵分解

成对约束

类别约束

本文作者相关文章

马慧芳

赵卫中

史忠植

PubMed

Article by Ma, H. F.

Article by Diao, W. Z.

Article by Shi, Z. Z.

1. 叶军.基于投影梯度及下逼近方法的非负矩阵分解[J]. 计算机工程, 2012,38(3): 200-202
2. 王亮, 王士同.基于成对约束的动态加权半监督模糊核聚类[J]. 计算机工程, 2012,38(01): 148-150
3. 钟将, 刘龙海, 梁传伟.基于成对约束的主动半监督文本聚类[J]. 计算机工程, 2011,37(13): 183-186
4. 郭立;张守志;汪卫;施伯乐.一种增量式非负矩阵分解算法[J]. 计算机工程, 2010,36(4): 66-68
5. 王成章, 白晓明.基于Fisher块对角LNMF的彩色人脸识别[J]. 计算机工程, 2010,36(16): 24-26
6. 吴毓龙;袁平波.密度敏感的距离测度在特定图像聚类中的应用[J]. 计算机工程, 2009,35(6): 210-212
7. 宿韬;张强;魏小鹏;周昌军.基于NMF分组策略的人脸识别[J]. 计算机工程, 2009,35(4): 199-200
8. 于红芸;姜涛;关键.一种新的SAR图像目标检测算法[J]. 计算机工程, 2009,35(24): 213-215
9. 张磊;冯晓森;项学智.基于非负矩阵分解的中文文本主题分类[J]. 计算机工程, 2009,35(13): 26-27,5
10. 周昌军;张强;魏小鹏;.基于NMF图像重构的人脸识别[J]. 计算机工程, 2008,34(3): 217-219

文章评论

反馈人	<input type="text"/>	邮箱地址	<input type="text"/>
反馈标题	<input type="text"/>	验证码	<input type="text" value="6304"/>
<input type="text"/> 			