

人工智能及识别技术

基于LSI和SVM的文本分类研究

刘美茹

(哈尔滨铁道职业技术学院计算机教研室, 哈尔滨 150086)

收稿日期 修回日期 网络版发布日期 2007-8-1 接受日期

摘要 文本分类技术是文本数据挖掘的基础和核心, 是基于自然语言处理技术和机器学习算法的一个具体应用。特征选择和分类算法是文本分类中两个最关键的技术, 该文提出了利用潜在语义索引进行特征提取和降维, 并结合支持向量机(SVM)算法进行多类分类, 实验结果显示与向量空间模型(VSM)结合SVM方法和LSI结合K近邻(KNN)方法相比, 取得了更好的效果, 在文本类别数较少、类别划分比较清晰的情况下可以达到实用效果。

关键词 [特征提取](#) [潜在语义索引](#) [支持向量机](#)

分类号 [TP391](#)

DOI:

通讯作者:

作者个人主页: 刘美茹

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(120KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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](#)

相关信息

- ▶ [本刊中 包含“特征提取”的 相关文章](#)
- ▶ 本文作者相关文章
 - [刘美茹](#)