

数据库、信号与信息处理

## 一种新的监督潜在语义模型

廖一星

浙江财经学院 信息学院 计算机应用研究所, 杭州 310018

收稿日期 2008-6-27 修回日期 2008-10-6 网络版发布日期 2009-12-4 接受日期

**摘要** Sprinkling方法是一种集成了训练样本类别信息的监督潜在语义模型。但是该方法特征权重采用词频, 降低了文本分类效果, 同时该模型并没有考虑不同样本对分类的贡献能力, 而是认为样本对分类的贡献相同, 另外, 该模型采用多个特征映射一个类别来加强类别知识对分类的贡献。为此, 文章在Sprinkling方法的基础上提出了一种新的监督潜在语义模型。实验结果表明, 该文方法的总体性能优于原始的Sprinkling方法, 在特征数为100时, 获得了最高分类精度, 提高幅度达到1.71%。

**关键词** [文本分类](#) [潜在语义](#) [sprinkling方法](#)

**分类号** [TP181](#)

## New model of supervised latent semantic indexing

LIAO Yi-xing

Department of Information, Zhejiang University of Finance & Economics, Hangzhou 310018, China

### Abstract

Sprinkling method is a supervised latent semantic indexing integrated with the classification information of the training sample. But the method of feature weight is TF which decreases the text classification performance. And this method doesn't consider the contribution ability of different samples. In contrast, this method considers the contribution ability of every sample is same. In addition, this method uses several features corresponding to a class label to boost the contribution of class knowledge to classification. A new supervised latent semantic indexing is proposed based on the sprinkling method. The results show that the new model outperforms the sprinkling method. The new model achieves the highest classification performance when feature number is 1, 100 which is increased 1.71% compared with the original sprinkling method.

**Key words** [text classification](#) [latent semantic](#) [method of sprinkling](#)

DOI: 10.3778/j.issn.1002-8331.2009.33.038

通讯作者 廖一星

### 扩展功能

#### 本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(534KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“文本分类” 的相关文章](#)
- ▶ [本文作者相关文章](#)
- [廖一星](#)