

学术研究

极大熵球面K均值文本聚类分析

修宇^{1,3}, 王士同^{1,2+}, 朱林¹, 宗成庆²

- 1. 江南大学 信息工程学院, 江苏 无锡 214036
- 2. 中科院自动化研究所 模式识别国家重点实验室, 北京 100080
- 3. 安徽工程科技学院 计算机科学与工程系, 安徽 芜湖 241000

收稿日期 修回日期 网络版发布日期 2007-9-20 接受日期

摘要 提出了一种基于极大熵理论的球面K均值文本聚类算法ME-SPKM。该算法利用了传统文本聚类算法SPKmeans中使用的余弦相似度度量, 进而引入极大熵理论构造了适合文本聚类的极大熵目标函数。对文本数据的实验证明了极大熵球面K均值文本聚类算法取得了比传统文本聚类算法更好的聚类效果。

关键词 [极大熵](#) [文本聚类](#) [球面K均值](#)

分类号

Maximum-entropy sphere K-means document clustering analysis

XIU Yu^{1,3}, WANG Shitong^{1,2+}, ZHU Lin¹, ZONG Chengqing²

- 1. School of Information Engineering, Jiangnan University, Wuxi, Jiangsu 214036, China
- 2. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China
- 3. Department of Computer Science and Engineering, Anhui University of Technology and Science, Wuhu, Anhui 241000, China

Abstract

A maximum-entropy version of the spherical K-means document clustering algorithm ME-SPKM is presented based on the well-known maximum-entropy. The proposed method uses the cosine similarity which is adopted by the typical text clustering algorithm SPKmeans, then constructs a maximum-entropy-based objective function. Experimental results demonstrate that the maximum-entropy spherical K-means ME-SPKM can achieve better clustering results than traditional clustering approaches in text clustering.

Key words [maximum-entropy](#) [document clustering](#) [spherical K-means](#)

DOI:

通讯作者 修宇 [E-mail:wxwangst@yahoo.com.cn](mailto:wxwangst@yahoo.com.cn)

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(871KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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](#)
- ▶ [浏览反馈信息](#)

相关信息

- ▶ [本刊中 包含“极大熵”的相关文章](#)
- ▶ [本文作者相关文章](#)

- [修宇](#)
- [王士同](#)
- [朱林](#)
- [宗成庆](#)