

数据库、信号与信息处理

基于Isomap的中文短信文本聚类算法

刘金岭

淮阴工学院, 江苏 淮安 223001

收稿日期 2008-8-22 修回日期 2008-11-10 网络版发布日期 2009-12-6 接受日期

摘要 给出的算法思想是首先计算出中文短信的相似度, 再通过使用Isomap方法得到短信在语义空间中的嵌入情况, 然后将短信在低维嵌入上进行聚类分析。该算法克服了短信的传统聚类分析在表示层次上遇到的困难, 也克服了词频统计法不能将内容意思相似的短信聚集在一起的缺点, 实验表明该算法是行之有效的。

关键词 [短信聚类](#) [Isomap算法](#) [语义空间](#)

分类号 [TP311](#)

Chinese short messages text clustering algorithm based on Isomap

LIU Jin-ling

Huaiyin Institute of Technology, Huai'an, Jiangsu 223001, China

Abstract

The calculating way proposed in this paper is to calculate the likeness degree of Chinese message and a message is gottern which embedded in the semantic space by using the Isomap method. This paper analyzes the messages according to the different clustering types in low-dimensional embedding. This algorithm has overcome difficulties in analyzing messages of traditional clustering types on different layers, and it has also overcome weakness of word frequency statistics which can not gather the similar meaning messages together. Experimental result indicates the algorithm is effective.

Key words [short messages clustering](#) [Isomap algorithm](#) [semantic space](#)

DOI: 10.3778/j.issn.1002-8331.2009.34.044

通讯作者 刘金岭 liujinlingg@126.com

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(612KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“短信聚类”的 相关文章](#)
- ▶ [本文作者相关文章](#)
- [刘金岭](#)