

机器学习与数据挖掘

基于事件本体的Web不良信息挖掘

刘东慧<sup>1,2</sup>,姜薇<sup>1\*</sup>

1.中国矿业大学计算机科学与技术学院,江苏徐州 221000; 2.连云港师范高等专科学校计算机系,江苏连云港 222000

摘要: 为挖掘互联网上的不良信息,本研究借鉴了事件语义分析技术。研究了基于事件本体的Web不良信息挖掘方法,重点是事件本体的构建、文本特征重构。为了验证所提方法的有效性,以信息聚类为例实现了一个基于事件本体的Web不良信息挖掘的原型系统。实验结果表明:基于事件本体和k均值的信息聚类方法的平均准确率为72.1%,较之传统的基于k均值的信息聚类方法的平均准确率提高了5.3%。

关键词: Web挖掘 不良信息 聚类 事件本体 信息检索

Research on Web negative information mining based on event ontology

LIU Dong-hui<sup>1,2</sup>, JIANG Wei<sup>1\*</sup>

1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221000, China; 2. Department of Computers, Lianyungang Teachers' College, Lianyungang 222000, China

Abstract: In order to mine the negative information on the internet, the event based semantic analysis technology was applied. The methods of event ontology based Web negative information mining, especially event ontology construction and text feature reconstitution were studied. Information clustering was used as an example to validate the methods proposed. A prototype system based on event ontology was implemented. Experimental results showed that the average accuracy obtained by the event ontology based and k means method was 72.1%, which increased by 5.3% compared with the traditional k means clustering method.

Keywords: Web mining negative information clustering event ontology information retrieval

收稿日期 2012-05-10 修回日期 网络版发布日期

DOI:

基金项目:

江苏省高校自然科学研究项目(10KJD520008)

通讯作者: 姜薇(1966-),女,江苏徐州人,副教授,硕士生导师,主要从事数据挖掘等方面的研究. E-mail: wjiang@cumt.edu.cn

作者简介: 刘东慧(1980-),女,江苏连云港人,硕士研究生,主要从事概念格,信息检索等方面的研究. E-mail: denghui\_liu@126.com

作者Email:

PDF Preview

参考文献:

本刊中的类似文章

1. 卜德云,张道强.自适应谱聚类算法研究[J].山东大学学报(工学版),2009,39(5): 22-26
2. 许延生,刘兴芳.模糊聚类迭代模型在水资源承载能力评价中的应用[J].山东大学学报(工学版),2007,37(3): 100-104
3. 马志强,常发亮,田伟,赵瑶.彩色图像中的人脸检测方法[J].山东大学学报(工学版),2007,37(4): 19-22
4. 牛新生,叶华,王亮.彩色图像中的人脸检测方法[J].山东大学学报(工学版),2007,37(4): 0-0
5. 赵洪国,张焕水,张承慧.基于RBF神经网络的交通流量预测算法研究[J].山东大学学报(工学版),2007,37(4): 0-0
6. 朱文兴,龙艳萍,贾磊.基于RBF神经网络的交通流量预测算法[J].山东大学学报(工学版),2007,37(4): 23-27

扩展功能

本文信息

Supporting info

PDF(1865KB)

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

Web挖掘

不良信息

聚类

事件本体

信息检索

本文作者相关文章

PubMed

7. 王耘,穆勇,刘庆红.基于灰关联分析的模糊聚类最优划分判定模型[J].山东大学学报(工学版),2006,36(2):86-89
  8. 孙宇清,赵锐,姚青,史斌,刘佳.一种基于网格的障碍约束下空间聚类算法[J].山东大学学报(工学版),2006,36(3):86-90
  9. 雷小锋<sup>1</sup>,庄伟<sup>1</sup>,程宇<sup>1</sup>,丁世飞<sup>1</sup>,谢昆青<sup>2</sup>.OPHCLUS:基于序关系保持的层次聚类算法[J].山东大学学报(工学版),2010,40(5):48-55
  10. 杨立才,赵莉娜,吴晓晴.基于蚁群算法的模糊C均值聚类医学图像分割[J].山东大学学报(工学版),2007,37(3):51-54
-