

数据挖掘

基于四层贝叶斯网络的XML文档信息检索模型

章小龙<sup>1</sup>, 衡星辰<sup>2</sup>

- 1. 漳州师范学院计算中心
- 2. 国网信息通信有限公司

**摘要:** 提出了一种新的基于贝叶斯网络对XML文档信息进行查询的模型方法。该模型支持针对XML文档信息的结构化查询。基于XML信息查询的特点, 利用XML数据集中语词、元素和结构化单元的统计信息对模型的拓扑结构和条件概率进行了学习; 结合概率函数的方法, 利用模型的概率推理进程对XML文档和结构化查询条件的相关度进行了估算。最后在基于INEX测试集的实验中证明了该方法的有效性和可靠性。

**关键词:** 贝叶斯网络 XML信息检索 结构化查询 概率函数

XML document information retrieval model based on four-layered Bayesian network

**Abstract:** In this paper, a Bayesian model for XML document information retrieval was proposed, supporting content and structure queries. The topology and conditional probabilities of the Bayesian model were mined from the XML document collection by statistics of terms, elements and structural units contained in the collection. Combined with probability functions, the relevance of a document to a given structured query was obtained by an inference process through a complex dependences network of this model. Experimental results on the INEX corpus of XML documents show the validity and reliability.

**Keywords:** Bayesian network XML information retrieval structured query probability function

收稿日期 2009-03-30 修回日期 2009-06-08 网络版发布日期 2009-10-28

DOI:

基金项目:

无

通讯作者: 衡星辰

作者简介:

作者Email: hengxingchen@sgcc.com.cn

参考文献:

本刊中的类似文章

1. 陈栋梁 王浩 姚宏亮 俞奎. 基于自适应粒子滤波的动态贝叶斯网推理算法[J]. 计算机应用, 2007, 27(2): 369-372
2. 徐建民 白彦霞 吴树芳. 基于同义词扩展的贝叶斯网络检索模型[J]. 计算机应用, 2006, 26(11): 2628-2630
3. 张少中 章锦文 张志勇 韩美君 王秀坤. 面向大规模数据集的贝叶斯网络参数学习算法[J]. 计算机应用, 2006, 26(7): 1689-1691
4. 吕国云 蒋东梅 赵荣椿 蒋晓悦 侯云舒 孙阿利 H. Sahli W. Verhelst. 基于动态贝叶斯网络的音视频连续语音识别和音素切分[J]. 计算机应用, 2007, 27(7): 1670-1673
5. 徐建民 朱松 陈富节. 术语相似度和术语相关度在检索模型中的融合研究[J]. 计算机应用, 2007, (12): 3013-3015

扩展功能

本文信息

- ▶ Supporting info
- ▶ PDF (884KB)
- ▶ [HTML全文]
- ▶ 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

- ▶ 贝叶斯网络
- ▶ XML信息检索
- ▶ 结构化查询
- ▶ 概率函数

本文作者相关文章

- ▶ 章小龙
- ▶ 衡星辰

PubMed

- ▶ Article by Zhang, X.L
- ▶ Article by Heng, X.C

6. 吴俊伟 何良华 方钰.基于动态贝叶斯网络的社交圈归属匹配模型[J]. 计算机应用, 2008,28(12): 3102-3104
  7. 曹卫东 贺国光.连续航班延误与波及的贝叶斯网络分析[J]. 计算机应用, 2009,29(2): 606-610
  8. 郑莉华 陈佳.基于贝叶斯网络的电话费欺诈模型的研究及应用[J]. 计算机应用, 2008,28(2): 511-512
  9. 付晓东 邹平 尚振宏 姜瑛.基于贝叶斯网络的Web服务组合故障诊断[J]. 计算机应用, 2008,28(5): 1095-1097
  10. 肖建琼 冯庆煜.基于贝叶斯网络的自适应系统建构[J]. 计算机应用, 2008,28(5): 1347-1349
  11. 杨喜权 曹雪亚 国岷娜 周建园.基于贝叶斯网络的本体不确定性推理[J]. 计算机应用, 2008,28(5): 1170-1172
  12. 杜智华 王宜伟.一种新颖的基因调控网络结构学习方法 [J]. 计算机应用, 2009,29(06): 1539-1543
  13. 王涛 杨娟 程海川 李博.基于未确知集—贝叶斯网络的构件软件体系可靠性模型 [J]. 计算机应用, 2009,29(06): 1715-1721
  14. 瞿英 李萍 吴祈宗 贾建.基于粗糙集的贝叶斯网络连续变量离散化算法[J]. 计算机应用, 2009,29(09): 2491-2493
-