

网络差异数据的优化挖掘模型仿真分析研究

高 静,董振华,郭 峰

(郑州成功财经学院 信息工程系,河南 巩义 451200)

摘 要: 提出一种基于反距离插值和分布式目标识别的网络差异数据优化挖掘算法. 构建网络差异数据的采集和信号预处理模型,采用反距离插值扰动搜索聚类算法进行网络差异数据的分布源方位估计,将采样数据的协方差矩阵进行时间平均和空间平均,基于分布式目标识别算法实现对网络差异数据的准确挖掘. 仿真结果表明,采用该算法能有效提高对网络差异数据的分布源估计的准确性,对数据的挖掘精度较高,性能优越于传统算法.

关键词: 网络差异数据;挖掘;分布式

中图分类号: TP393

文献标识码: A

文章编号: 1000—7180(2016)07—0136—04

Simulation Analysis of Network Difference Data Optimization Mining

GAO Jing, DONG Zhen-hua, GUO Feng

(Department of Computer and Information Engineering,
Zhengzhou Chenggong University of Finance and Economics, Gongyi 451200, China)

Abstract: A new data mining algorithm based on inverse distance interpolation and distributed object recognition is proposed. Building a network difference data acquisition and signal pre processing model, disturbance search clustering algorithm for network data difference distribution of the source range estimation using the inverse distance weighted interpolation, the sampling data covariance matrix to average the average time and space, based on distributed object recognition algorithm to achieve accurate mining difference data of network. Simulation results show that the proposed algorithm can effectively improve the accuracy of the distributed source estimation of network difference data, and it has higher accuracy and better performance than the traditional algorithm.

Key words: network difference data; mining; distributed

1 引言

在网络环境中,各种差异数据表现形式诸如网络流量序列、网络访问信号、网络多媒体图像和数字音频信息等属性,不同的网络数据的表现属性有一定的差异性,需要对这种差异数据进行有效挖掘,实现对保证网络信息的安全调度和可靠性访问,所以研究网络差异数据的优化挖掘模型具有重要意义^[1-3].

当前,对网络差异数据的挖掘算法是一个基于多层解析构造体制下的挖掘方法,使用系统分层方

算法,把需要挖掘领域的多个数据等效成多层系统构造,经过层和层间的反复计算,达成无损数据的挖掘,获取好的输出数据^[4-7]. 依据类型能分成层次网络差异数据算法、模糊网络差异数据算法、基于密度的网络差异数据算法、分割网络差异数据算法、基于网格的网络差异数据算法等,其中,层次类别算法分为融合法和分裂法^[8-11]. 而模糊聚类算法是采用事物间留有的模糊性进行数据挖掘,其得到的结果都不理想. 传统方法对网络差异数据挖掘采用热点追踪负反馈挖掘算法,由于网络多层系统结构的差异性,对数据挖掘性能不好.

针对上述问题,为了克服传统方法的弊端,本文提出一种基于反距离插值和分布式目标识别的网络差异数据挖掘算法。建立网络差异数据采集及信号预处理模型,采用反距离插值扰动搜索聚类算法进行网络差异数据的分布源方位估计,基于分布式目标识别算法,实现对网络差异数据的精确挖掘,经过仿真实验证明,采用改进的数据挖掘算法其挖掘精度要优于传统的挖掘算法。

2 网络差异数据的采集和预处理模型

2.1 网络差异数据的采集和信号模型构建

网络差异数据是分布在网络环境中的络流量序列、网络访问信号、网络多媒体图像等属性数据集,为了实现对网络差异数据的挖掘,需要首先进行数据集的信号模型构建和数据采集,首先对数据的生成模型进行网络构建,在网络监控、网络流量监测等领域,数据的采集通过数据传感器网络实现,以网络差异流量监测数据为例,进行数据采集和监测主要是依赖于对热量传感器网络的数据监测进行,网络差异流量数据监测的数据采集流程如图 1 所示。

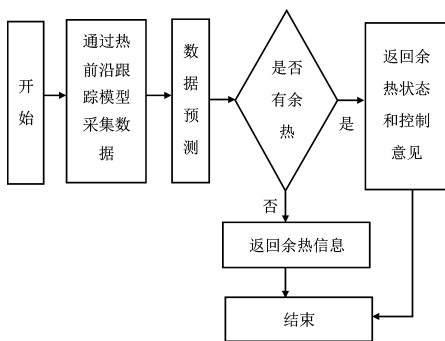


图 1 数据信息采集流程

在网络差异数据挖掘中,对条件表中所有参数进行提取,其优先级为多普勒检测设几条规则:

$$P_i \rightarrow \omega_{i1} * P_{i1}, \omega_{i2} * P_{i2}, \dots, \omega_{im} * P_{im} \quad (1)$$

式中, $1 \leq i \leq m, m \geq 2$, P_i 由 CF_i 和高阶统计量 τ_i 控制,其中 CF_i 为第 i 条网络差异数据特征规则的可信度, τ_i 为数据挖掘的滤波阈值。采用 Welch 法估计出谱的方位特性,当数据挖掘规则各前提真度的加权和 $T_i = \sum_{j=1}^m \omega_{ij} * T(P_{ij}) \geq \tau_i$ 时,用分段加窗把长度为 N 的数据 $X_N(n)$ 分成 L 段,每段长度为 M 。此时,将重复数据滤除和增量编码相结合,得到网络差异数据的真度为

$T(P_i) = CF_i * T_i = CF * \sum_{j=1}^m \omega_{ij} * T(P_{ij})$,若此规则成立的前提为 τ_i 是真规则,考虑到样本中各维数据的权

重,网络差异数据库信息流的波束输出为

$$l(t) = \left(\sum_{m=1}^M u_m \right) \cos(2\pi f_0 t) - \left(\sum_{m=1}^M v_m \right) \sin(2\pi f_0 t) \quad (2)$$

其中,

$$u_m = \cos[2\pi f_0 \tau_m(\theta)]; v_m = \sin[2\pi f_0 \tau_m(\theta)] \quad (3)$$

当某个样本对这一类数据隶属度较大,满足前提 τ 真法则,并设每条规则的前提真度分别为 $T_1, T_2, \dots, T_k$ 。设 n 个样本依据 m 个指标特征值,各个类均由目的基类 CbaseTarget 生成的,目标基类包括了根本的特性函数,以此实现了对网络差异数据的采集和信号模型构建。

2.2 反距离插值扰动搜索聚类预处理

在网络差异数据的采集和信号模型构建的基础上,采用反距离插值扰动搜索聚类算法进行网络差异数据的分布源方位计算,在数据构造设计里,把数据构造区分成点数据、线数据及面数据三样,且使用反距离插值的算法对数据实行插值治理。如果某分布源的网络差异数据信号密度函数能够表示成:

$$g(\varphi) = \sum_{l=1}^m \alpha_l p_{\Phi l} \quad (4)$$

式中, $p_{\Phi l}$ 是高斯分布函数, α_l 是第 l 个高斯分布函数的权系数,采用反距离插值方法提取海量数据的数据特征,然后依据初次提取的数据特征,高斯混合模型,得到一个高斯混合模型,具体形式为

$$g(\varphi) = \sum_{l=1}^m \alpha_l \left(\frac{1}{\sqrt{2\pi}\sigma_l} \exp\left(-\frac{(\theta - \theta_l)^2}{2\sigma_l^2}\right) \right) \quad (5)$$

式中, θ_l, σ_l 分别为第 l 个高斯模型的参数,对特征进行分层分类,设计了点、线、面 3 层结构的数据结构,在数据结构设计的基础上,采用反距离插值的方法对本数据结构性能进行分析,密度函数可以写为

$$g(\varphi) = \frac{\alpha}{\sqrt{2\pi}\sigma_1} e^{-\frac{(\theta - \theta_1)^2}{2\sigma_1^2}} + \frac{1 - \alpha}{\sqrt{2\pi}\sigma_2} e^{-\frac{(\theta - \theta_2)^2}{2\sigma_2^2}} \quad (6)$$

假设两个高斯模型的方差相等,即 $\sigma_1 = \sigma_2$,并且 $\theta_1 = -\theta_2 = \Delta$,为了克服模糊 C 均值算法对噪声和初始值敏感的弱点,本文提出反距离插值扰动搜索聚类算法提高全局搜索能力,当有 K 条规则每条结论的规则相同,设均为 P ,跟着聚类中心的反复迭代处理,依据反距离插值搜索聚类更新模糊分类矩阵,提高数据挖掘能力。

3 算法实现

采用反距离插值扰动搜索聚类算法进行网络差异数据的分布源方位估计,基于分布式目标识别算

法实现对网络差异数据的准确挖掘,多媒体技术下,每种多媒体信息的资源需要的数据存储空间很大,这给数据的传输和处理带来了很大的困难,需要进行差异数据挖掘。

首先对网络差异数据生成一个高斯时间序列 $y(n)$, 设定 $X = (x_1, x_2, \dots, x_D)$ 是空间中的一个点, 网络差异数据存在 q 个分布源, 差异信号矢量 $s(t) = [s_1(t), s_2(t), \dots, s_q(t)]^T$ 和干扰矢量 $n(t)$, $[a_j, b_j]$ 表示最小值和最大值, 满足 $x_j \in [a_j, b_j]$, 其中 N 为观测数据长度, 在目标源之间不相干的情况下, 逐层挖掘方法每产生一层频繁模式集, 挖掘规则的可信度为 $CF_1, CF_2, \dots, CF_k$, 则结论 P 的真度为

$$T(P) = \sum_{i=1}^k CF_i * T_i \tag{7}$$

在上述推理中, 当某一子前提真度未知时, 特点维数特性是离散向量集合, 把数据特点空间维度实行离散化治理, 将其 P 的真度设为 0, 即不真也不假。使用方向矢量和噪声子空间的正交特点来构建算法的空间谱, 把采样数据的协方差矩阵进行时间平均和空间平均, 构造一个 Toeplitz 矩阵 R_T :

$$R_T = \begin{bmatrix} \overline{r(1)}, \overline{r^*(2)}, \dots, \overline{r^*(M)} \\ \overline{r(2)}, \overline{r(1)}, \dots, \overline{r^*(M-1)} \\ \dots, \dots, \dots, \dots \\ \overline{r(M)}, \overline{r(M-1)}, \dots, \overline{r(1)} \end{bmatrix} \tag{8}$$

其中,

$$\overline{r(i)} = \frac{1}{M-i+1} \sum_{m=1}^M R_x(m, m-i+1), i = 1, \dots, M \tag{9}$$

式中, $\sigma_1, \sigma_2, \dots, \sigma_M$ 为网络差异数据修正协方差矩阵奇异值, 对于噪声子空间的重新定义为加权噪声子空间 U_n , 表达式如下:

$$U_n = \left(\frac{1}{\sigma_1} v_1, \frac{1}{\sigma_2} v_2, \dots, \frac{1}{\sigma_M} v_M \right) \tag{10}$$

利用方向矢量和加权后噪声子空间的正交性, 即 $U_n^H b_i(\theta_i) = 0$, 选取 Toeplitz 规律化治理就表示着把不一样的时间平均值再次实行一次阵元间的平均计算, 基于分布式目标辨别, 防止因为差异数据挖掘目标间的相干性造成的矩阵秩的减少, 使得网络差异数据分散个数估算的精确度降低。通过上述分析实现网络差异数据挖掘模型改进。

4 仿真实验与结果分析

为了证明本文设计的网络差异数据挖掘模型的

性能, 需要实行仿真实验证明, 基于计算机 Matlab 实验平台, 其环境是: 主频是 2.8 GHz, Intel Core i5 处理器, 4 GB 内存。网络节点采用 16 个阵元的均匀线列阵, 阵元之间间距为信号的半波长, 将结点所在的平面区域划分为若干个边长为 $l = r/\sqrt{2}$ 的正方形区域, 并采用 K^2 种分布式目标阵列对每个网络差异特征区域进行标定, 由于远场的两个分布式信号源, 网络差异数据分散中心方位角为是 $2^\circ, 4^\circ$; 差异分布源扩展角度为 $1^\circ, 2^\circ$, 信噪比为 5 dB, 阵元数目为 20, 先构建网络差异数据的采集和信号预处理模型, 采用反距离插值扰动搜索聚类算法进行网络差异数据的分布源方位估计, 基于上述仿真环境设计, 进行数据挖掘仿真, 得到解调后 5 个通道的网络差异数据频谱图如图 2 所示。

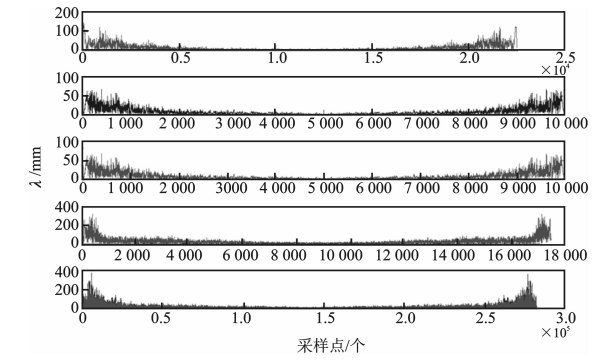


图 2 5 个通道的网络差异数据频谱图

以上述数据为研究样本, 采用反距离插值扰动搜索聚类算法进行网络差异数据的分布源方位估计, 得到仿真结果如图 3 所示。

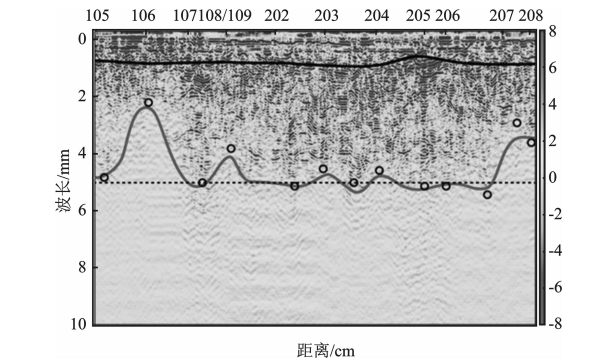


图 3 网络差异数据的分布源估计

为了对比算法性能, 采用本文算法和传统算法, 在不一样干扰环境下, 对每个信噪比作 50 次独立实验, 获取每一个信噪比上不一样算法对网络差异数据分散源的评估误差, 采用本文算法进行差异数据挖掘的精度比传统方法提高 12.68%, 这说明本文

算法在提高网络差异数据挖掘中具有优越的性能.

5 结束语

通过对网络差异数据的准确挖掘保证网络信息的安全调度和可靠性访问. 传统方法对网络差异数据挖掘采用热点追踪负反馈挖掘算法, 由于网络多层系统结构具有差异性, 对数据挖掘性能不好. 本文提出一种基于反距离插值和分布式目标识别的网络差异数据挖掘算法. 采用反距离插值扰动搜索聚类算法进行网络差异数据的分布源方位估计, 基于分布式目标识别算法实现对网络差异数据的准确挖掘, 仿真结果表明, 采用该算法能有效提高对网络差异数据的分布源估计准确性, 挖掘精度较高.

参考文献:

- [1] 欧阳柏成. 基于相对临界耦合预测控制的数据库安全索引[J]. 科技通报, 2014, 30(6): 131-133.
- [2] ZhuQ Y, YangX F, YangL X, et al. Optimal control of computer virus under a delayed model[J]. Applied Mathematics and Computation, 2012, 218 (23): 11613-11619.
- [3] 李骏骁. 多层差异网络深度入侵数据挖掘方法研究[J]. 计算机仿真, 2015, 32(4): 235-239.
- [4] 王慧贤, 靳惠佳, 王娇龙, 等. K 均值聚类引导的遥感影像多尺度分割优化方法[J]. 测绘学报, 2015, 44(5): 526-532.

- [5] 蔡坤, 杨扬. 一种改进的 LIPI 数据挖掘算法的仿真分析[J]. 计算机仿真, 2014, 31(8): 268-272.
- [6] 王恒. 数据挖掘分类技术在电信网络风险预警中的应用[J]. 价值工程, 2013, 19(2): 210-211.
- [7] 吴洁璇, 陈振杰, 张云倩, 等. 多核 CPU 下的 K-means 遥感影像分类并行方法[J]. 计算机应用, 2015, 35(5): 1296-1301.
- [8] 缪裕青, 吴孔玲, 朱晓雁, 等. 一种基于序列末项位置信息的序列模式挖掘算法[J]. 计算机应用研究, 2012, 29(7): 2505-2508.
- [9] 曾志, 王晋, 杜震洪, 等. 一种云格环境下可计算大型电气数据库的云存储资源与服务高效调配机制[J]. 浙江大学学报(理学版), 2014, 41(3): 353-357.
- [10] 黄莉静, 张辉, 王井阳, 等. 基于关系数据库的频繁项集挖掘算法研究[J]. 数学的实践与认识, 2013, 43 (12): 198-203.
- [11] 任锁平. 基于数据挖掘的高职教学质量监控评价系统设计及应用[J]. 电子设计工程, 2015, 23(16): 25-27.

作者简介:

高 静 女, (1983-), 硕士, 讲师. 研究方向为软件工程和计算机网络. E-mail: gao0217jing@163. com.

董振华 女, (1980-), 硕士, 讲师. 研究方向为数据库和计算机网络技术.

郭 峰 男, (1983-), 本科, 讲师. 研究方向为电子商务和网页设计.

(上接第 135 页)

- [5] 刘克明. 基于 ADS 的 S 波段低噪声放大器设计[J]. 数字技术与应用, 2014, 13(3): 143-144.
- [6] 聂星河, 郁健, 程冰, 等. 宽带限幅低噪声放大器设计[J]. 固体电子学研究进展, 2014, 34(5): 445-449.
- [7] 程骏, 李海华. 一种 S 波段低噪声放大器的设计[J]. 电子器件, 2013, 36(2): 206-210.

- [8] 马群. S 波段低温低噪声放大器[J]. 低温与超导, 2013, 41(8): 89-92.

作者简介:

唐 健 男, (1982-), 博士. 研究方向为从事化合物半导体和微波电路设计. E-mail: jtang1982@163. com