

大型网络异常数据库的快速数据定位模型仿真

朱保锋¹, 苏小玲²

(1 河南教育学院 信息技术系, 河南 郑州 450046;

2 河南工业大学 信息科学与工程学院, 河南 郑州 450001)

摘 要: 在对大型网络异常数据库快速数据定位的研究过程中, 采用当前的算法进行快速数据定位时无法精确地提取特定的异常数据特征, 存在数据定位精确度低的问题. 对此提出基于改进多子群萤火虫算法的大型网络异常数据库的快速数据定位建模方法. 先组建大型网络异常数据库的实体模型, 并以组建的模型为依据详细地分析异常数据形成的原因, 同时提取网络异常数据库中异常数据的特征, 融入多子群萤火虫算法, 获得大型网络异常数据库中特定异常数据特征, 并根据其特征建立大型网络异常数据库的快速数据定位模型. 实验仿真证明, 基于改进多子群萤火虫算法的大型网络异常数据库的快速数据定位建模方法定位精确度高, 适应性强.

关键词: 数据库; 定位; 多子群萤火虫算法

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2016)02-0140-04

Rapid Data Location Model Simulation of Abnomal Database of Large Network

ZHU Bao-feng¹, SU Xiao-ling²

(1 Department of Information Technology, Henan Institute of Education, Zhengzhou 450046, China;

2 School of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China)

Abstract: On rapid data of large-scale network abnormal database localization in the process of research, using the current algorithm when positioning rapid data cannot be accurately extract the specific characteristics of abnormal data exist the problem of low positioning accuracy data. Is proposed based on improved fertility group of firefly algorithm of large fast positioning data modeling method of network anomaly database. To form a large network anomaly database entity model, and on the basis of the model to form a detailed analysis of the causes of the formation of abnormal data, extracting network anomaly characteristics of the abnormal data in the database at the same time, the fertility group of firefly algorithm, to obtain large network anomaly characteristics of the specific abnormal data in the database, and according to its characteristics to establish a large database of power network fault positioning rapid data model. Experimental simulation show that the firefly algorithm based on improved fertility group of large-scale network abnormal database orientation of rapid data modeling method of high precision, strong adaptability.

Key words: database; positioning; many children of firefly algorithm

1 引言

随着网络科学技术的快速发展, 人类对网络的应用也越来越多, 网络数据库也随之产生^[1-3]. 与此同时, 大型网络系统的组建和网络数据库平台的逐步完善, 形成了一定的网络异常数据库, 也使网络异

常数据库信息实现了智能化管理^[4-6]. 然而, 随着时间与信息的逐渐积累, 大型网络异常数据库存储的异常数据会因为各种原因产生一些特定的异常数据, 如何有效地对这些特定的异常数据进行定位成为了该领域亟待解决的主要问题, 而大型网络异常数据库的快速数据定位建模方法则是解决这一问题

的有效途径,引起了很多专家与学者的重视^[7].

文献[8]提出了基于粒子群算法的大型网络异常数据库的快速数据定位建模方法.该方法通过定义异常数据库查询执行代价,融合粒子全局寻优的性能实现对大型网络异常数据库的快速数据定位建模.该方法较为简单,但是存在查询误差较大的问题.文献[9]重点研究了基于分层拓扑模型的大型网络异常数据库的快速数据定位建模方法.该方法依据建立的异常数据库分层拓扑模型对异常数据库的信息状态进行计算,建立了大型网络异常数据库的快速数据定位模型.该方法计算时间矩,但是在进行快速数据定位时对特定异常数据提取的性能受限,存在数据定位准确度不高的问题.文献[10]提出基于神经网络算法的大型网络异常数据库的快速数据定位建模方法.该方法通过建立神经网络较为强大的拓扑模型来完成对大型网络异常数据库的快速数据定位建模,虽然适应性较强,但是存在计算较为繁琐、耗时长弊端.

针对上述问题,文中提出基于改进多子群萤火虫算法的大型网络异常数据库的快速数据定位建模方法.

2 大型网络异常数据库的快速数据定位建模优化方法

2.1 网络异常数据特征的提取

在大型网络异常数据库的快速数据定位建模过程中,先组建大型网络异常数据库的实体模型,并以组建的模型为依据详细地分析异常数据形成的原因,同时提取网络异常数据的特征.具体的步骤如下所示.

(1)组建大型网络异常数据库的实体模型.为分析网络异常数据形成原因提供基础依据.

由 (E_i, E_j, d, t) 代表大型网络数据库中包含的四元组, R 代表四元组间的信任关系, $A = \{A_1, A_2, \dots, A_m\}$ 表示异常数据分类属性, 选取异常数据当前时刻查询得到的信任值的最大值可以得出, 查询间隔较长的信任值对当前的影响就会相对较小, 则时间减弱函数的计算公式如下:

$$f(k) = \begin{cases} f(k-1) - \frac{1}{n} \\ 1 \end{cases} \quad (1)$$

则在 $k-1$ 时刻内的大型网络异常数据库中集成的异常数据状态方程, 可以表述为

$$p_i(t) = \sum_{n=1}^N \frac{A}{r} e^{-jkr} R_{in} \frac{1}{r} e^{-jkr} \quad (2)$$

$$p_i(t) = \frac{A}{r^2} \sum_{n=1}^N e^{-j2kr} a_{in} e^{j\varphi_{in}} \quad (3)$$

式中, A 代表网络异常数据库查询指向性异常数据强度, r 代表异常数据长度, K 代表异常数据出现时的波数, f 代表异常数据产生的瞬时频率, c 代表指向性数据库传递权重.

在获取由 $\{x_{k-1}^i, w_{k-1}^i\}$ 代表的 $k-1$ 时刻的大型网络异常数据库集成的异常数据的状态后验分布后, 依据由 B_j 代表的异常数据库各存储节点反馈系数, 组建大型网络异常数据库实体模型, 利用下式表述:

$$p(y|\alpha, \theta) = \sum_{k=1}^k \alpha_k p_k(y|\mu_k, \sum_k) \quad (4)$$

(2) 对大型网络异常数据库实体模型进行分析, 提取特定异常数据的特征, 为大型网络异常数据库的快速数据定位建模提供依据.

利用相空间重构和非线性时间序列分析方法分析异常数据形成的原因, 同时提取网络异常数据的特征. 具体的步骤如下所示.

假设 $\{x_1, x_2, \dots, x_N\}$ 代表大型网络异常数据库异常数据的时间序列, m 代表嵌入维数, τ 代表延迟时间间隔, $p(x_0)$ 代表异常数据的后验概率估计, 可以得到异常数据的组成形式, 如下:

$$p_{ij}(k) = \frac{[l_j(k) - l_j(k)] \eta_{ij}(k)}{\sum_{j \in N_i(k)} [l_j(k) - l_j(k)] \eta_{ij}(k)} \quad (5)$$

式中, $l_j(k)$ 代表编号为 j 的第 k 个异常数据出现的概率.

依据上式得到异常数据组成形式后, 对提取特征公式进行简化, 用下式表述:

$$E_{q \times r} = Y_{q \times U} W_{u \times u} Y_{u \times r}^T P_{ij}(k) \quad (6)$$

式中, $Y_{q \times U}$ 代表异常数据维度矩阵, $W_{u \times u}$ 代表异常数据均衡概率.

则利用下式可提取大型网络异常数据库的异常数据特征:

$$E_n = \sum_{i=1}^n y_i \varphi_i \quad (7)$$

式中, y_i 代表大型网络异常数据库的异常数据个数.

综上所述, 在大型网络异常数据库的快速数据定位建模过程中, 先组建大型网络异常数据库的实体模型, 并以组建的模型为依据详细地分析了异常

数据组成形式,同时提取网络异常数据的特征,为实现大型网络异常数据库中特定异常数据的快速定位奠定基础.

2.2 大型网络异常数据库的快速数据定位建模的实现

在上述获得的异常数据特征的基础上,融合多子群萤火虫算法,对大型网络异常数据库中的特定异常数据进行查询并定位. 先将萤火虫群分为多个子群,各子群最优萤火虫通过信息交流找到大型网络异常数据库的数据查询最优方案,建立大型网络异常数据库的快速数据定位模型. 具体的步骤如下所示.

特定异常数据查询问题是搜索到一条最小的异常数据进行查询. 假设在大型网络异常数据库的特定异常数据定位建模过程中,有 n 个关系 $R = \{r_1, r_2, \dots, r_n\}$ 进行计算,特定异常数据的查询计划集合由 s 代表,则由 $\text{cost}(s)$ 代表 s 的相关联代价,由 t_m 代表中间的特定异常数据的查询代价的总和,则网络异常数据库的特定异常数据查询的最优模型,如下:

$$\text{Mincost}(s) = \sum_{m=1}^{n-1} \text{cost}(t_m) \quad (8)$$

在获得查询最优时的模型之后,建立大型网络异常数据库的快速数据优化定位模型. 具体的步骤如下所示.

(1) 初始化参数,由 m 代表萤火虫数目,由 n 代表萤火虫子群数,在每个子群中分别设置光强吸收系数由 λ 代表. 且设置收敛条件,在网络异常数据库的数据查询优化求解的空间内随机生成每个萤火虫的初始位置,并由 $x_{ij}(i = 1, 2, \dots, n)$ 代表.

(2) 确定萤火虫的位置进化方向,公式如下:

$$r_{ij} = \|X_i - X_j\| \quad (9)$$

(3) 计算由 β_{ij} 代表的每个子群萤火虫的吸引度,计算公式如下:

$$\beta = \beta_0 e^{-\lambda r_{ij}^2} \quad (10)$$

(4) 更新萤火虫的空间位置,计算公式如下:

$$x_i = x_i + \beta(x_j - x_i) + a(\text{rand} - 1/2) \quad (11)$$

(5) 获得最优萤火虫位置对应的网络异常数据库中特定异常数据的具体位置进行定位,其表达式如下:

$$\text{Mincost}(s) = \sum_{m=1}^{n-1} \text{cost}(t_m) \quad (12)$$

综上所述,通过获取大型网络异常数据库中异常数据的特征,引入多子群萤火虫算法,对异常数据

库中的特定异常数据进行定位,可精确地建立大型网络异常数据库的快速数据定位模型. 实现建模之后,需要进步进行仿真试验,验证模型的性能.

3 实验仿真证明

实验 1: 分别采用改进算法、粒子群算法、神经网络算法进行大型网络异常数据特征的提取,并将不同算法提取异常数据特征的精确度进行对比,用来衡量不同算法的有效性,对比结果见表 1.

表 1 不同算法提取故障数据特征精确度对比 %

故障数据类型	改进算法精确度	粒子群算法精确度	神经网络算法精确度
1 类	96	87	79
2 类	98	88	76
3 类	97	86	73

从表 1 可以看出改进算法提取的异常数据特征的精确度要高于其他几种算法,这主要是因为改进算法先组建大型网络异常数据库的实体模型,并以组建的模型为依据详细地分析了异常数据的组成形式,同时提取网络异常数据的特征,因此保障了改进算法提取异常数据特征的精确度.

实验 2: 分别采用改进算法、粒子群算法、神经网络算法进行大型网络异常数据库的快速数据定位实验,将不同算法的收敛速度和最优查询计划执行时间进行对比,用来衡量不同算法的有效性. 对比结果见表 2 和表 3.

表 2 不同算法的查询收敛速度对比

关系个数/个	改进算法迭代次数/次	粒子算法迭代次数/次	神经网络算法迭代次数/次
10	120	150	178
15	135	165	176
20	230	360	466
45	320	420	488
65	355	430	472

表 3 不同算法的执行时间对比

关系个数/个	改进算法执行时间/s	粒子群算法执行时间/s	神经网络算法执行时间/s
10	30.17	57.34	62.14
15	55.12	83.14	94.14
20	175.05	180.1	175.23
45	225.13	280.91	380.21
65	205.5	305.1	345.8

假设由 $n = 20$ 代表网络异常数据库特定异常数据查询优化问题,改进算法迭代 230 次时,便得到了网络异常数据库特定异常数据的最优查询结果,相应时间为 175.05 s,从表 3 和表 4 可以看出,改进算法的有效性优于其他几种算法,这是因为改进算法融合多子群萤火虫算法,将网络异常数据库特定异常数据定位问题转换为特定异常数据查询优化问题,将萤火虫群分为多个子群,各子群最优萤火虫通过信息交流找到大型网络异常数据库的特定异常数据查询最优方案,有效地建立了大型网络异常数据库的快速数据定位模型。

4 结束语

针对采用当前的算法进行快速数据定位时无法精确地提取异常数据的特征,存在特定异常数据定位精确度低的问题,提出基于改进多子群萤火虫算法的大型网络异常数据库的快速数据定位建模方法。先组建大型网络异常数据库的实体模型,并以组建的模型为依据详细的分析异常数据的组成形式,同时提取异常数据的信号特征,在融合多子群萤火虫算法,将网络异常数据库的特定异常数据定位问题转换为议程数据库特定异常数据查询优化问题,将萤火虫群分为多个子群,各子群最优萤火虫通过信息交流找到大型网络异常数据库异常数据查询最优方案,建立大型网络异常数据库的快速数据定位模型。实验仿真证明,基于改进多子群萤火虫算法的大型网络异常数据库的快速数据定位建模方法定位精确度高,适应性强。

参考文献:

- [1] 王春红. 大型图像数据库的快速检索方法研究仿真[J]. 计算机仿真, 2014, 26(5):246-249.
- [2] 董建丽. 电力系统和通信网络交互影响下的连锁故障分析[J]. 中国新通信, 2013, 21(1):7-11.
- [3] 舒征宇, 邓长虹, 黄文涛, 等. 小世界电力网络故障传播过程与抑制策略[J]. 电网技术, 2013, 37(3):861-867.
- [4] 舒征宇, 马俊民, 邓长虹, 等. 小世界电网的故障传播特性分析与运行极限匹配模型研究[J]. 电力系统保护与控制, 2013, 25(7):29-35.
- [5] 韩雪, 齐园. 基于自感应主元素分析的电网故障诊断系统[J]. 计算机测量与控制, 2013, 21(4):889-891.
- [6] 邓德祉. 浅谈船艇电力网故障分析与保护[J]. 中国外资月刊, 2013, 25(14):323,325.
- [7] 郑圣, 张阳, 郑贤舜, 等. 电力故障现场反馈拍照功能设计与效益分析[J]. 科技与企业, 2014, 31(4):113.
- [8] 魏金蓉, 周亚玲. 电力电缆故障测试技术及应用的概述[J]. 科技资讯, 2013, 25(18):117.
- [9] 李素若. 电力网络抗毁性能的量化研究与仿真[J]. 计算机仿真, 2014, 31(1):93-95.
- [10] 郭万东. 10kV 电力线路故障原因分析及预防对策[J]. 西铁科技, 2013, 25(3):20-21.

作者简介:

朱保锋 女,(1978-),硕士,讲师.研究方向为计算机网络技术. E-mail:zbf_teacher@163.com

苏小玲 女,(1979-),硕士研究生,讲师.研究方向为计算机网络安全.