

混合架构下 ORACLE 数据快速查找方法研究

戚 斌

(陕西国防工业职业技术学院, 西安 710300)

摘要: 为了提高对混合架构下 ORACLE 数据的挖掘和查找速度, 提出一种基于频繁项目集关联规则挖掘的数据快速查找方法; 构建 ORACLE 数据的内部关联属性映射关系模型, 在异质网络混合构架模式下, 采用 Graph OLAP 数据仓库模型进行数据库检索的关系维度表, 提取表达 ORACLE 数据属性信息的特征参量, 通过同态标签检索方法, 实现对目标数据的快速查找定位; 仿真结果表明, 采用该方法进行 ORACLE 数据查找的查准率和查全率较高, 计算速度较快, 性能优于传统方法。

关键词: 混合架构; ORACLE 数据; 查找; 数据库

Research on ORACLE Data Fast Search Method in Hybrid Architecture

Qi Bin

(Shaanxi Institute of Technology, Xi'an 710300, China)

Abstract: In order to improve the speed of mining and searching ORACLE data in hybrid architecture, this paper proposes a new method based on frequent itemsets association rule mining. To construct the internal mapping associated attribute relationship model of ORACLE data in the heterogeneous network, hybrid structure Graph OLAP data warehouse model using the database relational dimension table construction, extracting the characteristic parameters of expression feature information of the ORACLE data cable through the method of homomorphic tags fast lookup to locate the target detection data. The simulation results show that using the method of ORACLE data to find the precision and recall of high calculation speed, it has better performance than the conventional method.

Keywords: hybrid architecture; ORACLE data; search; database

0 引言

近年来, 随着流媒体网络技术的飞速发展, 流媒体中携带的 ORACLE 数据呈爆炸式增长, 人们对流媒体中携带的 ORACLE 海量数据的挖掘和查找研究日益盛行。ORACLE 数据作为一种分布式的关联性异构数据, 广泛分布在异构网络数据库中, 对 ORACLE 数据的查找是通过网络搜索引擎实现的, ORACLE 数据中对网络搜索引擎的依赖程度较大, 互联网搜索引擎是万维网中的特殊站点, 通过网络搜索引擎, 结合数据挖掘算法实现对 ORACLE 数据准确查找, 研究数据快速查找方法在数据库检索和搜索方面具有重要应用价值^[1]。

传统方法中, 对混合架构下 ORACLE 数据查找方法主要有数据关联规则挖掘方法^[2]、数据并行调度方法^[3]、特征分解方法和多元线性回归分析方法等^[4], 随着大数据信息处理技术研究的不断深入, 对混合架构下 ORACLE 数据快速查找方法的处理研究受到了相关学者的重视, 并取得了一定的研究成果, 其中, 文献 [5] 中提出一种基于全域子空间分解挖掘的 QoS 预测方法实现 ORACLE 数据快速查找, 采用全域分析方法构建数据分布的特征子空间, 提取数据的点分布特征, 通过数据融合方法实现数据聚类, 提高关联数据的准确查找能力, 但是该方法计算开销较低, 在处理大批量的数据访问问题时的实时性不好。文献 [6] 采用海量散乱点云快速压缩方法进行海量 ORACLE 数据的挖掘和检索, 采用支持向量机进行误差修正, 提高了数据查准率, 提高数据并行计算的性能, 但是当

数据受到混合构架下的不规则和不确定信息干扰时, 数据挖掘的准确性受到限制^[7]。

针对上述问题, 本文提出一种基于频繁项目集关联规则挖掘的数据快速查找方法。构建 ORACLE 数据的内部关联属性映射关系模型, 在异质网络混合构架模式下采用 Graph OLAP 数据仓库模型进行数据库检索的关系维度表构建, 提取表达 ORACLE 数据属性信息的特征参量, 通过同态标签检索方法实现对目标数据的快速查找定位, 最后进行了仿真实验测试, 得出有效性结论, 展示了本文方法在提高数据查找的准确度和速度方面的优越性。

1 ORACLE 数据关联属性映射关系模型构建

1.1 混合架构下 ORACLE 数据的分布属性

为了实现对混合架构下 ORACLE 数据快速查找设计, 需要首先构建 ORACLE 数据关联属性映射关系模型, 在异质网络 Graph OLAP 中, ORACLE 数据需要大量网页信息界面提供接口, 专门用来帮助人们查找存储在其他站点上的信息。通过搜索引擎有告诉用户 ORACLE 数据文件或文档存储在何处^[8]。基于网络搜索引擎将散布于网络上的各相关网页文本集中起来, 实现集中管理和调度。在云存储系统中, 对多数网络搜索引擎来说, ORACLE 数据以页面数 (Page count) 和短文档集 (Snippets) 等两个最重要形式分布在信息资源数据库中, 本文以页面数 (Page count) 和短文档集 (Snippets) 为研究对策, 通过对海量数据的语义信息主题表达设计, 提取数据关联性分布特征, 实现语义信息主题特征匹配, 为 ORACLE 数据库访问和信息检索提供理论基础。基于图结构的 OLAP 模型构建海量数据优化存储和访问算法^[9], 在混合构架下, 采用有向图表示海量 ORACLE 数据的非结构性数据分布

收稿日期: 2017-03-02; 修回日期: 2017-03-26。

作者简介: 戚 斌 (1983-), 男, 陕西西县人, 工程硕士, 讲师, 主要从事软件开发、数据库设计, 高等职业教育方向的研究。

模型，为了增强数据分布和检索的灵活性，在数据查找设计中，放弃了结构化数据库的部分特性，根据大数据的整体几何关系，得到数据分布具有非连续层次性，在进行数据聚类处理中，需要进行面板数据的模板匹配，避免大数据的聚类中心出现偏移，根据上述设计原理，构建混合架构下 ORACLE 数据查找的总体构架模型如图 1 所示。

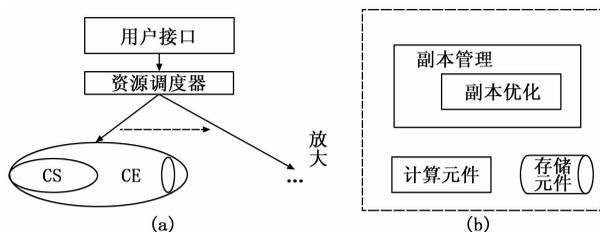


图 1 混合架构下数据查找模型实现结构

为了实现 ORACLE 大数据快速查找，需要进行优化自动聚类处理，采用非线性时间序列分析方法构建大数据的信息流模型，进行大数据时间序列的特征分析和聚类，在大数据信息样本库中，存在映射 $\Phi: X \rightarrow Y$ ，满足：

$$\Phi(x \cdot y) = \Phi(x) \circ \Phi(y) \quad (1)$$

其中： \cdot 是 X 上的运算， \circ 是 Y 上的运算。构建一个状态表 (State Table, S-Table) 表示内部属性与属性之间的一种约束关系，两个数据块的关系模式为 m_i 和 m_j ， $m_i + m_j$ 的数据依赖关系表示为 $T(m_i + m_j)$ ，通过特征参量提取进行聚类特征属性分析和聚类中心搜索，支持项目集的频繁项目由 $T(m_i)$ 和 $T(m_j)$ 生成，即 $T(m_i + m_j) = T(m_i) * T(m_j)$ 。

记关联规则 $X \rightarrow Y$ 的置信度维 $Count(X \rightarrow Y) / Count(X)$ ，对于 F 中的任意数据，将文件 F 划分成 n 个子块，得到频繁 k -项目集 $m_i (1 \leq i \leq n)$ ，然后将每个文件子块分成 k 个基本块，通过实体关系信息维度标识，以实体类型为元素组成的一个有序排列 $m_{i,j} (1 \leq i \leq n, 1 \leq j \leq k)$ ，对大数据信息流进行稀疏迭代协方差估计，每个基本块 $m_{i,j}$ 在数据聚类中心的特征分量为 $T_{i,j}$ ，基于类型的数据分块方法得到 ORACLE 子块的标签信息 T_i 由块指纹和数据类型 $T_{i,j}$ 聚集得到。计算 $N = p * q$ 和 $\varphi(N) = (p-1)(q-1)$ ，服务器端根据数据类型生成随机数 $e (e \in Z_N)$ ，使得 $\gcd(e, \varphi(N)) = 1$ 。则数据块的静态分块 (Static Chunking, SC) 表示为 $pk = (N, e)$ ，存储节点的最优块级为 $sk = (p, q)$ ，通过上述对混合架构下 ORACLE 数据的分布属性分析，为进行数据快速查找提供可靠的数据属性分析基础。

1.2 关联属性映射关系模型

在混合架构下建立 ORACLE 数据的内部关联属性映射关系模型，在异质网络混合构架模式下设计 ORACLE 数据库分布式查询方法，关联属性映射关系表示的是在单位时间内数据库的访问频繁项集^[10]，在此引入切分数据块的 Hash 值概念：

$$hot = AccessNum \times \omega_1 + (Ntime - Ctime) \times \omega_2 \quad (2)$$

式中， $AccessNum$ 为存储节点发出数据查找请求的次数，也是数据库的访问次数， $Ntime$ 为每个存储节点的反馈时间， $Ctime$ 为热点数据索引的时间开销， ω_1 和 ω_2 为权重且 $\omega_1 + \omega_2 = 1$ 。

根据各节点的响应时间，在每个存储节点上发送一个 PutFileReq 请求，相同的数据块通过普通索引表进行信息融合和数据分区聚类，在数据聚类中心中，采用差分进化方法进行

信息检索，对比索引表中的热点数据，用来记录热点数据块的最优块级，对比 Hash 指纹值修正每个特征分布向量 v_i ，选择包含前 m 个最大特征值的量子数据，得到海量 ORACLE 数据云存储的语义本体父概念。通过语义分析，进行 ORACLE 数据特征检索和数据库访问^[11]，其中检索的查准率和查全率是关键，提取大数据信息流的时延尺度特征参量，得到数据块分布的时延尺度为：

$$Cov(X, Y) = E\{[X - E(X)][Y - E(Y)]\} \quad (3)$$

其中： $E\{[X - E(X)]$ 表示 ORACLE 数据分布的自相关协方差矩阵，由此求得大数据信息流的时延尺度的自相关系数表示为：

$$\rho_{xy} = \frac{Cov(X, Y)}{\sqrt{D(X)} \sqrt{D(Y)}} \quad (4)$$

式中， ρ_{xy} 是一个无量纲的量。通过上述分析，构建关联属性映射关系模型表示为：

$$WebJaccard(X, Y) = \frac{P(X \cap Y)}{P(X) + P(Y) - P(X \cap Y)} \quad (5)$$

$$WebOverlap(X, Y) = \frac{P(X \cap Y)}{\min(P(X), P(Y))} \quad (6)$$

$$WebDice(X, Y) = \frac{2P(X \cap Y)}{P(X) + P(Y)} \quad (7)$$

其中： $P(X)$ 、 $P(Y)$ 表示混合架构下 ORACLE 数据关联规则分布的概率密度函数， X 、 Y 为数据存储节点的负载量， $P(X \cap Y)$ 是所有频繁项目集互信息概率密度分布， X 、 Y 为任意数据的时间采样分布。

2 数据快速查找方法实现

2.1 数据库检索的关系维度表构建

在上述进行了 ORACLE 数据关联属性映射关系模型构建和特征分析的基础上，进行数据快速查找方法设计，本文提出一种基于频繁项目集关联规则挖掘的数据快速查找方法。在异质网络混合构架模式下，采用 Graph OLAP 数据仓库模型进行数据库检索的关系维度表构建，采用二叉树模型进行前序遍历^[12]，从普通索引表中挑选出 ORACLE 数据。用二叉树模型中包含关系权重 (Weight)，节点标号表示实体节点，采用 LRU 算法构建全局优先级序列，关系 R 中支持项目集区域分布函数为：

$$E^v(c_1, c_2) = \mu \cdot Length(C) + \nu \cdot Area(inside(C)) + \lambda_1 \int_{inside(C)} |I - c_1|^2 dx dy + \lambda_2 \int_{outside(C)} |I - c_2|^2 dx dy \quad (8)$$

其中： c_1 和 c_2 分别两组 ORACLE 数据序列分布属性的非频繁项目集， $Length(C)$ 表示数据的长度， $Area(inside(C))$ 表示数据分布区域的非空真子集， μ, ν, λ_1 和 λ_2 表示混合构架下所有频繁项目集的自相关系数，均为大于 0 的常数。根据 ORACLE 数据属性间的依赖关系进行特征匹配和融合，得到数据库检索的关系维度计算为：

$$C = Min\{max(C_i)\} \quad (9)$$

$$\sum_{j=1}^p Z_{i,j} = 1, \forall i \in (1, n), \forall j \in (1, n_i) \quad (10)$$

其中： C_i 表示频繁项目集 L 的置信度， $Z_{i,j}$ 表示候选频繁 1-项目集。根据数据的尺度均匀遍历性，挖掘出有价值的关联规则为 $S = \overline{X_1}, \overline{X_2}, \dots, \overline{X_k}, \dots$ ，数据查找的属性分布时间窗口为 $T_1, T_2, \dots, T_K, \dots$ ，通过 Graph OLAP 数据仓库模型得到

一组统计意义上的数据查询值。

2.2 特征参量提取及同态标签检索

在数据库检索的关系维度表构建的基础上,提取表达 ORACLE 数据属性信息的特征参量,通过同态标签检索方法实现对目标数据的快速查找定位,基于数据依赖关系得到数据查找的全局最优向量 $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})$,在混合构架下数据分布属性的聚类权重迭代函数为:

$$\tilde{w}_k^i = \tilde{w}_{k-1}^i \frac{p(z_k/x_k^i) p(x_k^i/x_{k-1}^i)}{q(x_k^i/x_{k-1}^i)} \quad (11)$$

其中: x_k^i 为项目集 X 的支持数, \tilde{w}_{k-1}^i 为初始权重,对指定的数据块进非线性时间序列分析,实现对待查找数据流的结构优化,得到云服务器上存储节点在 $k+1$ 时刻的关联数据块为:

$$x_i(k+1) = x_i(k) + s \left(\frac{x_j(k) - x_i(k)}{\|x_j(k) - x_i(k)\|} \right) \quad (12)$$

对新的数据块 m'_i 进行特征分解处理后,在有限数据集中进行频繁项目集关联规则挖掘,得到挖掘输出的样本 $x_i, i = 1, 2, \dots, n$,根据数据块 m_i 得到标签信息的特征压缩矢量为:

$$x_i = (x_{i1}, x_{i2}, \dots, x_{is})^T \quad (13)$$

提取表达 ORACLE 数据属性信息的特征参量,通过同态标签检索方法实现对目标数据的快速查找定位,数据快速查找的实现步骤描述如下:

1) 经过仿射变换得到 m 个待查找 ORACLE 数据的关联规则特征,将全部 Graph OLAP 数据仓库中的待挖掘数据进行特征分布完整性检测;

2) 计算收到的数据块 m_i 的协方差矩阵 R ;

3) 根据返回的数据完整性证据计算 R 的局部时间特性分布的向量值 $\lambda_1, \lambda_2, \dots, \lambda_n$,以及相应的特征向量值 $\varphi_1, \varphi_2, \dots, \varphi_n$;

4) 基于频繁项目集关联规则挖掘方法进行数据特征检索,将数据分布的特征向量值从大到小进行排列,结果为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$,构建参数联合优化矩阵满足 $A(\varphi_1, \varphi_2, \dots, \varphi_m), m < n$;

5) 利用 $y = A^T x$ 进行傅里叶反变换,在数据聚类中心进行全局优化解搜索,生成一个 len 比特的压缩数据序列,执行模指运算,通过参数联合优化,实现数据快速查找。

3 仿真实验与结果分析

为了测试本文算法在实现混合构架下 ORACLE 数据快速查找中的应用性能,进行仿真实验分析,实验中的硬件环境为: CPU Intel® Core™ i7 - 2600 @ 3.40 GHz,利用 Euca-lyputs 软件构建混合构架云平台,配置了 MIRACL 数据库进行 ORACLE 数据存储,对 3 个 100 MBit 的 ORACLE 数据文件进行特征采样,数据分布的时间窗口系数 τ 为 0.26,特征尺度参数 $a_0 = 1.03$,数据查找的访问带宽 $B = 20$ dB,实验的持续时间 $T = 100$ s,根据上述仿真环境和参数设定,进行数据快速查找实验,得到原始数据采样如图 2 所示。

以图 2 所示的数据样本为研究对象,进行数据分布特征分析,提取表达 ORACLE 数据属性信息的特征参量,通过同态标签检索方法实现对目标数据的快速查找定位,得到特征定位结果如图 3 所示。

从图 3 结果得知,采用本文方法进行数据分布特征定位查找,能准确将需要查找的数据定位在所处空间中,实现数据准确挖掘,图 4 给出了采用本文方法和传统方法进行数据查找的

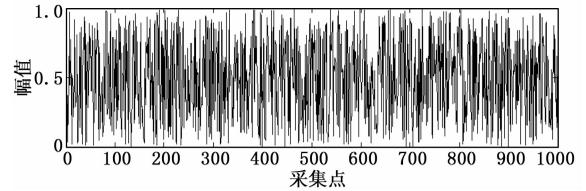


图 2 原始数据采样

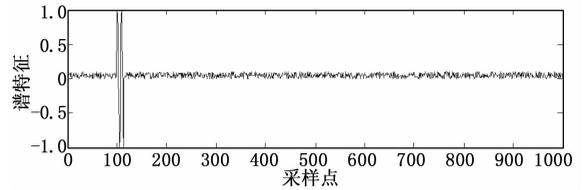


图 3 ORACLE 数据查找的特征定位结果

时间开销。表 1 列出了不同方法进行 100 次实验取均值的数据查找的查准性和查全性对比。

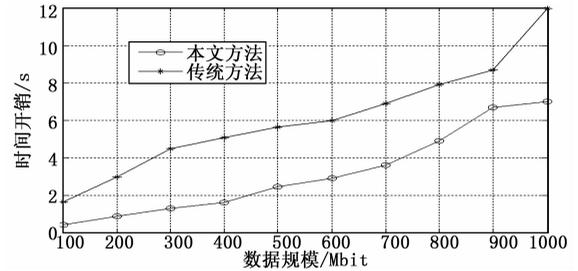


图 4 数据查找时间开销对比

表 1 数据查找的查准性和查全性对比

数据规模/Mbit	本文方法		传统方法	
	查准率/%	查全率/%	查准率/%	查全率/%
100	99.09	92.08	99.94	92.42
200	98.87	86.44	98.76	95.77
300	99.89	89.09	99.33	96.33
400	99.93	93.82	100	92.09
500	96.34	95.44	100	99.35

分析图 4 和表 1 结果得知,采用本文方法进行数据查找的时间开销较小,实现数据快速查找,查准率和查全率较高。

4 结束语

本文研究了 ORACLE 数据优化查找和挖掘问题,提出一种基于频繁项目集关联规则挖掘的数据快速查找方法。构建 ORACLE 数据的内部关联属性映射关系模型,提取表达 ORACLE 数据属性信息的特征参量,通过同态标签检索方法实现对目标数据的快速查找定位。研究得出,采用本文方法进行 ORACLE 数据查找的查准率和查全率较高,计算速度较快,目标数据定位准确,具有较好的应用性能。

参考文献:

[1] 邢淑凝, 刘方爱, 赵晓晖. 基于聚类划分的高效用模式并行挖掘算法 [J]. 计算机应用, 2016, 36 (8): 2202 - 2206.
 [2] 林楠, 史苇杭. 基于多层空间模糊减法聚类算法的 Web 数据库安全索引 [J]. 计算机科学, 2014, 41 (10): 216 - 219.

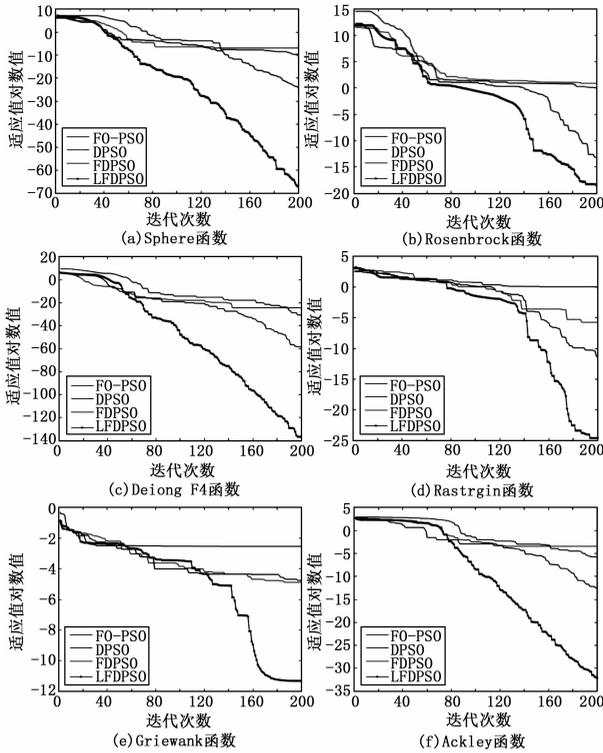


图 LFDPSO 算法在不同的测试函数上的性能比较

参考文献:

[1] Kennedy J, Eberhart R. Particle swarm optimization [A]. Proc IEEE International Conf on Neural Networks [C]. 1995: 1942-1948.

[2] Chen W N, Zhang. A novel set-based particle swarm optimization method for discrete optimization problem [J]. IEEE Transactions on Evolutionary Computation, 2010, 14 (2): 278-300.

[3] Zhang T, Hu T S, Zheng Y, et al. An improved particle swarm optimization for solving believed multiobjective programming problem [J]. Journal of Applied Mathematics, 2012, 2 (4): 1-13.

[4] Ho S Y, Lin H S, Liauh W H, et al. OPSO: Orthogonal particle swarm optimization and its application to task assignment problems [J]. IEEE Transactions on Systems, Man, Cybernetics A: Systems, Humans, 2008, 38 (2): 288-298.

[5] Nie F, Tu T, Pan M, et al. K-Harmonic means data clustering with PSO Algorithm [M]. Springer Berlin Heidelberg, 2012: 67-73.

(上接第 220 页)

[3] 陈志华, 刘晓勇. 云计算下大数据非结构的稳定性检索方法 [J]. 现代电子技术, 2016, 39 (6): 58-61.

[4] 王跃飞, 于 炯, 鲁 亮. 面向内存云的数据块索引方法 [J]. 计算机应用, 2016, 36 (5): 1222-1227.

[5] 张博雅, 胡晓辉. 一种基于全域子空间分解挖掘的 QoS 准确预测方法 [J]. 计算机科学, 2014, 41 (1): 217-224.

[6] 方 芳, 程效军. 海量散乱点云快速压缩算法 [J]. 武汉大学学报 (信息科学版), 2013, 38 (11): 1353-1357.

[7] 张景祥, 王士同, 邓赵红, 等. 融合异构特征的子空间迁移学习算法 [J]. 自动化学报, 2014, 40 (2): 236-246.

[8] 张 涛, 唐振民, 吕建勇. 一种基于低秩表示的子空间聚类改进算法 [J]. 电子与信息学报, 2016, 38 (11): 2811-2818.

[6] Varshney S, Srivastava L, Pandit M. Parameter tuning of statcom using particle swarm optimization based neural network [J]. Intelligent and Soft Computing, 2012, 130 (3): 813-824.

[7] Tillett J, Rao T, Sahin F, et al. Darwinian particle swarm optimization [A]. Indian International Conference on Artificial Intelligence [C]. 2005: 1474-1487.

[8] Gutierrez R E, Rosario J M, Machado J T. Fractional order calculus: basic concepts and engineering applications [J]. Mathematical Problems in Engineering, 2010, 2010: 1-19.

[9] Machado J A T, Jesus I S, Barbosa R, et al. Application of fractional calculus in engineering [J]. Dynamics, Games and Science I, Springer Proceedings in Mathematics, 2011, 1: 619-629.

[10] Pires J A, Moura P B, Oliveira A M, et al. Particle swarm optimization with fractional-order velocity [J]. Nonlinear Dynamics, 2010, 61 (1-2): 295-301.

[11] Couceiro M S, Rocha R P, Fonseca F N M, et al. Introducing the fractional-order Darwinian PSO [J]. Signal, Image and Video Processing, 2012, 6 (3): 343-350.

[12] 陈华, 范宜仁, 等. 一种动态加速因子的自适应微粒群优化算法 [J]. 中国石油大学学报, 2010, 34 (6): 173-176.

[13] Podlubny I. Fractional differential equations [M]. San Diego: Academic Press, 1999.

[14] Lshehawey E, Elbarbary E F, et al. On the solution of the endolymph equation using fractional calculus [J]. Appl. Math. Comput., 2001, 124 (3): 337-341.

[15] Camargo R F, Chiacchio A O, Oliveira E C. Differentiation to fractional orders and the fractional telegraph equation [J]. Math. Phys., 2008, 49 (3): 033-505.

[16] Diethelm K. The analysis of fractional differential equations [M]. Berlin: Springer-Vd: lag, 2010.

[17] Pires E J S, Machado J A T, Oliveira P B M, Cunha, et al. Particle swarm optimization with fractional-order velocity [J]. Nonlinear Dyn, 2010, 61 (1/2): 295-301.

[18] Munkres J R. Topology [M]. 2nd ed. London: Prentice-Hall Inc, 2000.

[19] Guez-Lopez J S R, Romaguera S. The relationship between the Vietoris topology and the Hausdorff quasiuniformity [J]. Topology and Its Applications, 2002, 124: 451-464.

[20] 郭 通, 兰巨龙, 李玉峰, 等. 自适应的分数阶达尔文粒子群优化算法 [J]. 通信学报, 2014, 35 (4): 130-140.

[9] 邓志刚, 曾国荪, 谭云兰, 等. 云存储内容分发网络中的能耗优化方法 [J]. 计算机应用, 2016, 36 (6): 1515-1519.

[10] 张 盛, 鄢 傲, 向忠胜, 等. 基于全网能量均衡的 WirelessHART 图路由算法 [J]. 计算机应用研究, 2014, 31 (5): 15201523.

[11] Mahboubi H, Moezzi K, Aghdam A G, et al. Distributed deployment algorithms for improved coverage in a network of wireless mobile sensors [J]. IEEE Transactions on Industrial Informatics, 2014, 10 (1): 163-174.

[12] Mahboubi H. Distributed deployment algorithms for efficient coverage in a network of mobile sensors with nonidentical sensing Capabilities [J]. IEEE Transactions on Vehicular Technology, 2014, 23 (8): 3998-4016.