

引用格式:林倩瑜. 云服务环境下的大数据多标签属性分类技术 [J]. 微电子学与计算机, 2019, 36(2): 101-104. [LIN Q Y. Big data multi-label attribute classification technology in cloud service environment [J]. Microelectronics & Computer, 2019, 36(2): 101-104.]

云服务环境下的大数据多标签属性分类技术

林倩瑜

(集美大学 诚毅学院, 福建 厦门 361021)

摘 要: 云服务环境下的大数据具有多标签属性, 导致数据访问过程中的分类性不好, 为了降低大数据分类的误分率, 提出一种基于极限学习机优化神经网络的云服务大数据分类方法。采用高维相空间重构模型对随机采样的云服务大数据先验特征进行信息重排, 采用关联规则挖掘方法提取数据的标签属性特征量, 采用多维标签属性的主成分分析方法进行大数据的主特征分析, 将主特征量输入到神经网络分类器中, 结合极限学习机进行神经网络分类的加权控制, 提高数据分类的自适应性, 实现云服务大数据优化分类。仿真结果表明, 采用该算法进行云服务大数据分类的准确性较高, 误分率较低, 在云组合服务中提高了数据的召回能力。

关键词: 云服务环境; 大数据; 分类; 神经网络; 极限机学习

中图分类号: TP391 **文献标识码:** A **文章编号:** 1000-7180(2019)02-0101-04

Big Aata Multi-label Attribute Classification Technology in Cloud Service Environment

LIN Qian-yu

(Chengyi University College, Jimei University, Xiamen 361021, China)

Abstract: Big data in the cloud service environment has multi-label attributes, which leads to poor classification in the process of data access. In order to reduce the misclassification rate of big data classification, a classifying method for cloud service big data based on the neural network optimization of extreme learning machine is proposed. The high-dimensional phase space reconstruction model is used to rearrange the information of the random sampling cloud service big data's prior feature, and the association rule mining method is used to extract the label attribute feature quantity of the data. The principal component analysis method of multidimensional label attribute is used to analyze big data's principal feature. The principal feature quantity is input into the neural network classifier, and the weighted control of neural network classification is carried out by combining with the extreme learning machine. Improve the self-adaptability of data classification, realize cloud service big data optimization classification. The simulation results show that the proposed algorithm has higher accuracy and lower misclassification rate in cloud service big data classification, and improves the recall ability of data in cloud composition service.

Key words: cloud service environment; big data; classification; neural network; limit machine learning

1 引言

随着云服务技术的发展, 采用高性能的云组合服务技术进行网络信息、物联网信息以及传感器信息的

智能管理, 能实现对大数据的优化调度和集成控制, 提高云服务的智能化管理和经营水平. 在进行云组合服务中的大数据管理和调度中, 需要融合高性能的计算机信息处理技术, 通过安全优质的数据组合分类算

法,提高对大规模集成云组合服务数据的自适应分类管理能力^[1].研究云服务环境下的大数据分类技术,并应用在云制造、云计算以及多目标的云规划设计中,为用户和企业提供最佳的优质云服务.

传统方法中,对大数据的分类方法主要有支持向量机分类方法、BP 神经网络分类方法、模糊神经网络分类方法等^[2,3],采用特征提取和标签定位技术对大数据标签序列提取和特征分析,实现云服务环境下的大数据分类.文献[4]中提出一种基于相关功率平特征分析后的云服务组合数据分类方法,反复调整聚类中心来进行大数据的聚类优化,结合PID 神经网络训练进行数据分类挖掘,该方法由于计算开销较大,数据分类的实时性不好^[5].针对上述问题,本文提出一种基于极限学习机优化神经网络的云服务大数据分类方法.

2 云服务大数据的数据结构及先验信息处理

2.1 云服务大数据的存储结构模型

本文研究云服务环境下的大数据分类算法,需要给出云服务环境下的大数据标签序列分布结构模型,进行云服务环境下的大数据特征提取和数据聚类处理,构建数据存储结构模型,按照五元组(srcIP、proto、dstIP、srcPort、dstPort)进行云服务环境下的大数据的自适应调度^[6],如图1所示.

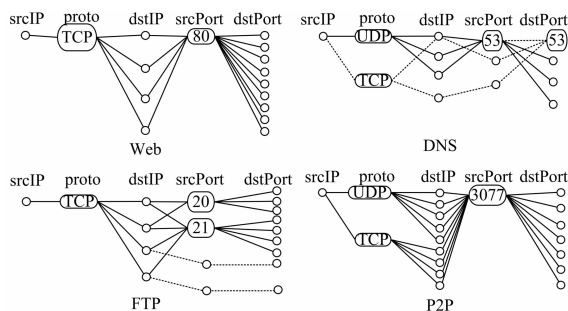


图1 云服务环境下的大数据调度模型

根据图1所示的大数据采样和调度模型,进行数据存储结构分析,假设随机采样云服务大数据时间序列采样的先验知识表示为 $\{x_i\}_{i=1}^N$,原始的云服务大数据具有线性相关性,按照序列 $y'(n)$ 的秩进行云服务大数据的相空间重构,结合特征序列分布进行信息重排,采用替代数据法进行云服务大数据的自适应调度,得到云服务大数据原始序列的采样冲激响应序列为 $\{\delta x_i(j_k) = x_{j_k} - x_i \mid k \in 1, 2, \dots, N_b\}$,构建随机采样数据的判别随机采样量产生 N_b 个 m 维向量,采用高维相空间重构模型云服务大数据原始序列,得到随机采样的云服务大数据的关联

规则特征量分布为:

$$\max_{x_{a,b,d,p}} \sum_{a \in A} \sum_{b \in B} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} V_p \quad (1)$$

$$s. t. \sum_{a \in A} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} R_p^{bw} \leq K_b^{bw}(S), b \in B \quad (2)$$

采用拟线性高斯分布模型分析云服务大数据的存储结构,采用广义最小二乘估计(GLS)对云服务大数据进行回归分析,在云存储空间中提高数据的分类准确性^[7].

2.2 大数据的多标签属性特征分析

采用高维相空间重构模型对随机采样得到的云服务大数据先验特征进行信息重排^[8],结合因子分析,得到云服务大数据随机采样序列的因子主成分概率密度置信域为:

$$g(x_i, y_i \mid \mu_k, \delta_k^2) = \prod_{k=1}^k \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right\} \quad (3)$$

上式中, α_k 为云服务大数据随机采样序列在全局渐进稳定状态下的矢量场, μ_k 为标签属性的关联系数,根据先验排序准则,可得云服务大数据的统计特征量分布为: $c_1 = 0; c_2 = \sigma^2; c_k = 0 (k \geq 3)$.对云大数据的样本数据通过极限学习方法进行自适应聚类处理,采用模糊C均值聚类方法,进行云服务大数据的特征分解和信息聚类,将初始聚类中心解 $X^{(0)}$ 划分为 N 部分,为 $X^{(1)}, X^{(2)}, \dots, X^{(N)}$,即 $X^{(0)} = \bigcup_{i=1}^N X^{(i)}$,按照大数据重排序列 $y'(n)$ 的秩来对原始的云服务大数据进行多标签属性归类,得到随机采样的云服务大数据的拟合模型表示为:

$$x_n = \varphi_0 + \sum_{i=1}^p \varphi_i x_{n-i} + \sum_{j=0}^q \theta_j \eta_{n-j} \quad (4)$$

其中 $\{\eta_i\}$ 为均值为0,方差为 σ^2 的标准正态分布标签信息, $\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_p$ 称为云服务大数据的时间的采样关联系数, $\theta_1, \theta_2, \dots, \theta_q$ 称为滑动时间窗口的平均系数,由此实现了对大数据的多标签属性特征分析,结合主成分特征提取方法,进行数据分类算法研究^[9].

3 数据分类算法的优化设计

3.1 大数据的主特征提取及多维标签属性分析

在提取数据的标签属性特征量,采用多维标签属性的主成分分析方法进行大数据的主特征分析的基础上,进行数据分类算法的改进设计,本文提出一种基于极限学习机优化神经网络的云服务大数据分

类方法,采用标准的极限学习机进行数据聚类中心的自动更新和分类标签属性识别,在分类的邻域范围 R_d^i 内得到云组合服务数据分类的极限学习机学习迭代式为:

$$P_{ij}(t) = \frac{L_j(t) - L_i(t)}{\sum_{k \in N_i(t)} L_k(t) - L_i(t)} \quad (5)$$

在合适的惯性权重值训练下,得到 k 时刻云服务大数据信息流的状态可以记为:

$$x_k = \sum_{i=1}^N \tilde{x}_k^i \tilde{w}_k^i \quad (6)$$

对高维重构后的云服务大数据信息流通过关联规则挖掘方法提取数据的标签属性特征量,采用多维标签属性的主成分分析方法进行大数据的主特征分析^[10],随机采样云服务大数据时间序列在 t_{n+1} 时刻和 t_n 时刻的状态可以记为:

$$\hat{x}_k = \sum_{i=1}^m \frac{n_i}{N} x_k^i, \sum_{i=1}^m n_i = N \quad (7)$$

假设云服务大数据样本集训练的神经网络共有 N 个神经元,大数据主成分输出特征提取结果为:

$$KL = \sum_{i=1}^{m+a} \frac{1}{N} \ln \frac{1}{N w_d^i(H)} + \sum_{i=1}^{N-m-a} \frac{1}{N} \ln \frac{1}{N w_d^i(H)} + 0 \quad (8)$$

$$\sum_{i=1}^m n_i = \sum_{i=1}^N = N \quad (9)$$

分析得到神经元的比例元集 $\{x_L^i\}$,采用组合聚类方法,结合多标签属性特征聚类技术,进行数据的模糊分类。

3.2 数据分类输出

对重排的云服务大数据信息流采用神经网络训练方法进行主成分特征提取,极限学习机算法为一个无监督学习过程,学习迭代式为:

$$x_i(k+1) = x_i(k) + s \left(\frac{x_j(k) - x_i(k)}{\|x_j(k) - x_i(k)\|} \right) \quad (10)$$

式中, $\|\vec{x}\|$ 表示 $\vec{x}$ 的范数. 由于极限学习机 i 的适应度 $f(x_i)$ 优于云数据分类标签属性的全局适应度 $f(p_g)$,调整极限学习机学习范围, $r_d^i(k)$ 表示 k 时刻第 i 个极限学习机节点进行云服务大数据的神经网络分类加权系数,且 $0 < r_d^i(k) < r_s$,得到云服务大数据分类的模糊学习迭代式为:

$$r_d^i(k+1) = \min\{r_s, \max\{0, r_d^i(k) + \beta(n_i - |N_i(k)|)\}\} \quad (11)$$

式中, β 表示在大数据分类全局搜索的关联特征量变, n_i

表示极限学习机的适应度值,迭代次数达到预设最大值时,极限学习机在进行大数据分类学习的收敛度达到最优,此时得到输出优化训练的权重需满足

$$w_b^i > \frac{K d_{ij}^{\max}}{N} \quad (12)$$

式中, N 为极限学习机的学习次数, $d_{ij}^{\max}$ 为分类标签 i 与分类标签 j 之间的最大欧式距离. 采用多标签属性检测方法进行分类训练的误差修正,得到分类结果输出为:

$$x_O^i = x_S^i + K d_i^{\max} (x_L^i - x_S^i) \quad (13)$$

式中, $K = 1 / \|x_L^i - x_S^i\|$,可见,采用多标签属性进行分类的权重学习,能提高的云服务大数据分类全局收敛性。

4 仿真实验分析

为了验证本文方法在实现云服务环境下大数据优化分类中的性能,进行仿真实验,实验采用 Matlab 7 设计,采样的云服务环境大数据的样本总数为 2 000,训练集为 100,分类的标签属性根据云服务的对象,分别设定为教育、制造、医疗等方面的云服务标签属性,根据标签属性的种类数,给出神经网络分类器的输出层节点数为 3,输入层节点数为 12,极限学习机学习维数为 $m=40$,相空间重构的维数为 $D=30$,数据分类的惯性权重 $\omega=0.24$,神经网络训练的搜索范围 $-5.56 \leq x_d \leq 6$,大数据分类的自适应训练函数选择为 $f_1(x) = \sum_{i=1}^n x_i^2$,总的迭代次数为 10 000 次,根据上述仿真环境和参数设定,进行云服务环境下的大数据多标签属性分类,得到采样的大数据样本如图 2 所示。

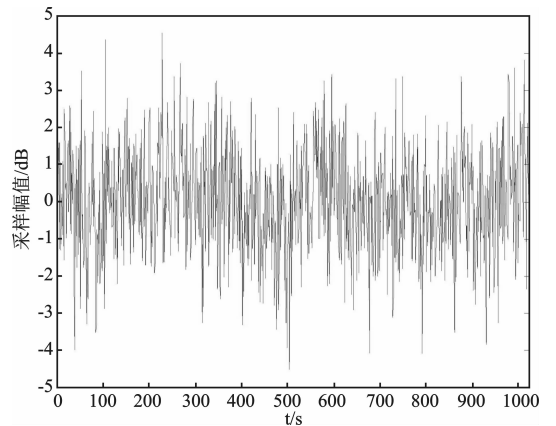


图2 大数据样本时域波形

以图 2 采样的大数据为测试集,通过关联规则挖掘方法提取数据的标签属性特征量,采用多维标

签属性的主成分分析方法提取大数据的主特征,将特征量输入到本文设计的极限学习机神经网络分类器中,实现数据分类,得到对不同标签属性的数据分类输出如图 3 所示.

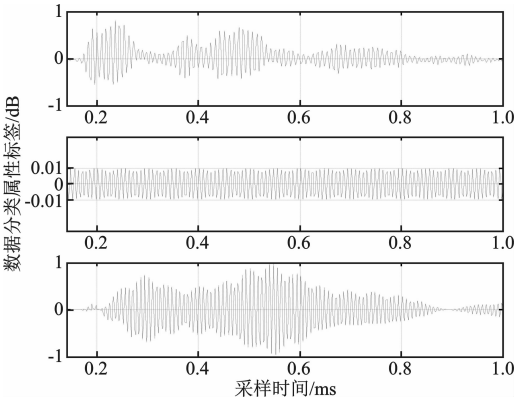


图 3 多标签属性的数据分类输出

分析图 3 得知,采用本文方法能对不同标签属性的数据进行有效分类,提高了数据分类输出的准确性.测试不同方法进行数据分类的误分率,得到对比结果如图 4 所示.

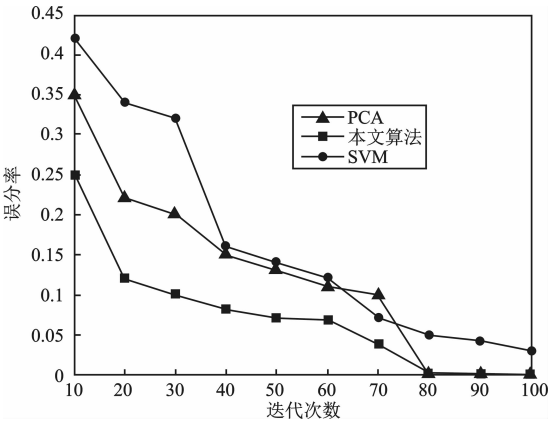


图 4 性能对比

分析可知,采用本文算法进行云服务大数据分类的准确性较高,误分率较低,相比传统方法有显著改善.

5 结束语

在进行云组合服务中的大数据管理和调度中,需要融合高性能的计算机信息处理技术,提高对大

规模集成云组合服务数据的自适应分类管理能力,本文提出一种基于极限学习机优化神经网络的云服务大数据分类方法.采用高维相空间重构模型对随机采样得到的云服务大数据先验特征进行信息重排,结合多维标签属性的主成分分析方法进行大数据的主特征分析,应用极限学习机进行神经网络分类的加权控制,提高数据分类的自适应性,实现云服务大数据优化分类.研究得知,本文方法大数据分类的多标签属性标记能力较好,分类的误分率较低.

参考文献:

[1] 王伟,胡长武,郭栋,张等.一种面向云构软件的云操作系统[J]. 计算机科学,2017,44(11):33-40.

[2] 陈志旺,黄兴旺,陈志兴,等.区间多目标优化非支配排序云模型算法[J]. 计算机工程与应用,2017,53(22):143-149.

[3] 唐小川,罗亮.基于析因设计的大数据相关关系挖掘算法[J]. 计算机应用,2018,38(9):2507-2510.

[4] 郭华平,董亚东,毛海涛,等.一种基于逻辑判别式的稀有类分类方法[J]. 小型微型计算机系统,2016,37(1):140-145.

[5] 郭华平,周俊,郭长安,等.面向非平衡类问题的k近邻分类算法[J]. 计算机应用,2018,38(4):955-959.

[6] 刘健,常莲.基于Bohmann机的物联网数据分类方法研究[J]. 激光杂志,2016,37(10):116-120.

[7] 毛文涛,田杨阳,王金婉,等.面向贯序不均衡分类的粒度极限学习机[J]. 控制与决策,2016,31(12):2147-2154.

[8] 吴磊,原鹏,丁维龙.智能家居网关与云服务器数据同步协议的研究[J]. 计算机技术与发展,2018,28(09):151-155.

[9] 熊丽荣,王玲燕,黄玉柱.一种联合LTR和社交网络的Top-k推荐方法[J]. 小型微型计算机系统,2018,39(12):2577-2584.

[10] 李永刚,张治中,李龙江.一种WLAN与eHRPD系统间负载均衡算法[J]. 微电子学与计算机,2017,34(1):44-47.

作者简介:

林倩瑜 女(1985-),硕士,实验师.研究方向为数据挖掘,神经网络. E-mail:18005926008@163.com.