

文章编号: 2095-2163(2020)10-0153-05

中图分类号: TP391.9

文献标志码: A

基于多分类器融合的信息中心网络源数据检测

柯建波

(广东工业大学 华立学院, 广州 511325)

摘要: 为了提高信息中心网络中源数据的提取和检测能力,本文提出一种基于多分类器融合的信息中心网络源数据检测方法,构建统计序列重组模型,采用相空间结构重组方法进行源数据的信息特征挖掘和重构,在高维相空间中实现网络源数据的关联规则挖掘,对提取数据的关联特征集分类融合,构建基于模糊C均值聚类的多模级联分类器,采用数据特征的融合法和分裂法进行数据聚类中心寻优控制,实现信息中心网络源数据检测优化。仿真结果表明,采用该方法进行信息中心网络源数据检测的自适应性较好,检测的准确率较高,提高了信息中心网络源数据分类管理和信息融合能力。

关键词: 信息中心网络; 源数据; 分类器; 融合

Information Center Network Source data Detection based on Multi-classifier Fusion

KE Jianbo

(Huali College, Guangdong University of Technology, Guangzhou 511325, China)

[Abstract] In order to improve the extraction and detection capability of the source data in the information center network, a method for detecting a network source data of an information center network based on a multi-classifier fusion is proposed, and a statistical sequence recombination model of the information center network source data is constructed, the method comprises the following steps of: carrying out information feature mining and reconstruction of source data by adopting a phase space structure recombination method, realizing association rule mining of network source data in a high-dimensional phase space, classifying and fusing an associated feature set of the extracted information center network source data, The multi-mode cascade classifier based on the fuzzy C-means clustering is constructed, and the data clustering center optimization control is carried out by using the fusion method and the splitting method of the data feature, so that the network source data detection and optimization of the information center is realized. The simulation results show that the method has better adaptability to the data detection of the network source data of the information center, and the accuracy of the detection is higher, and the network source data classification management and information fusion capability of the information center network are improved.

[Key words] information center network; source data; classifier; fusion

0 引言

随着网络通信技术的发展,在信息中心网络中的数据规模不断增大,需要对信息中心网络源数据进行有效的识别和检测,提取源数据的关联规则集,根据网络通信的输出特征量进行自适应特征检测和识别,滤除源数据的干扰信息,提高信息中心网络的数据管理能力,研究信息中心网络源数据的优化检测方法,在实现信息中心网络的数据信息集成管理和调度中具有重要意义^[1]。

对信息中心网络源数据的检测是建立在对数据的优化挖掘和融合聚类处理基础上,建立信息中心网络源数据的关联规则分布特征集,结合模糊相关性约束控制方法,进行信息中心网络源数据的自适应特征分离和检测。当前,对信息中心网络源数据的检测方法主要有谱特征检测方法、模糊相关性检

测方法、融合聚类检测方法等。其中,文献[4]中提出一种基于非平稳数据融合的信息中心网络源数据的高效定位挖掘的检测方法,采用点扫描方法在传输链路层进行信息中心网络的传输数据采样,结合特征分离和自适应的信息融合技术,实现信息中心网络的源数据检测和挖掘,但该方法的计算开销较大,实时性不好^[2];一种基于交互式信息融合的信息中心网络源数据检测模型,进行信息中心网络的链路传输信道分析和特征重构,根据数据检测结果进行模糊聚类和挖掘,提高网络数据检测的准确性,但该方法在进行数据检测中的均衡性不好^[3];基于模糊PID信息融合和可达标识的信息中心网络源数据检测方法,采用PID神经网络检测器进行数据检测,结合关联信息挖掘方法,提高数据检测的准确性,该方法在进行大规模数据挖掘的实时性和抗干

作者简介: 柯建波(1981-),男,学士,讲师,主要研究方向:人工智能、计算机网络。

收稿日期: 2020-04-21

扰性不好^[4]。

针对上述问题,本文提出一种基于多分类器融合的信息中心网络源数据检测方法,首先,构建信息中心网络源数据的统计序列重组模型,建立基于模糊C均值聚类的多模级联分类器;其次,采用数据特征的融合法和分裂法进行数据聚类中心寻优控制,实现网络源数据检测优化;最后,进行仿真实验分析。仿真结果说明本文方法在提高信息中心网络源数据检测准确性方面的性能优越。

1 信息中心网络源数据重组模型及特征挖掘

1.1 信息中心网络源数据相空间重组

为了实现对信息中心网络源数据检测设计,首先构建网络源数据的相空间重构模型,结合对原始数据的五元组特征分析结果,进行源节点定位和子图模式分析,建立网络源数据的连通图模式结构,根据源节点模式的匹配结果进行关联规则挖掘和端口结构分析,构建网络源数据的节点匹配模式,根据多分类器融合结果进行信息分类,实现网络源数据的优化检测,实现结构如图1所示。

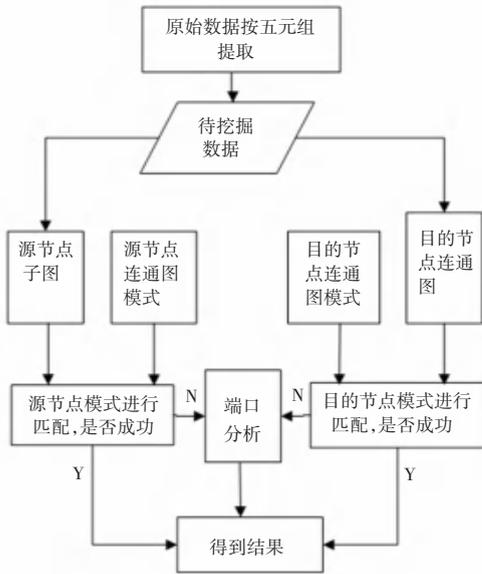


图1 信息中心网络源数据检测的实现结构图

Fig. 1 Realization of information center network source data detection

根据图1所示的信息中心网络源数据检测结构模型,进行检测算法设计,构建网络源数据分布的有限数据集模型为式(1):

$$X = \{x_1, x_2, \dots, x_n\} \subset R^s. \quad (1)$$

信息中心网络中对源数据采样集合中含有 n 个样本,信息中心网络源数据的有限论域内,数据的统计样本序列为 x_i , 对应数据的特征采样序列为 $i =$

$1, 2, \dots, n$, 特征矢量为式(2):

$$x_i = (x_{i1}, x_{i2}, \dots, x_{is})^T. \quad (2)$$

在信息中心网络节点连通图中,构建信息中心网络源数据的高维特征空间分布结构模型,采样相空间重构技术,构建反映信息中心网络源数据关联特征的本体模型,相空间重构的关联权重为式(3):

$$\omega = ((\omega_1, a_1'), (\omega_2, a_2'), \dots, (\omega_n, a_n'))^T, \\ \omega_j \in [0, 1]. \quad (3)$$

在高维相空间中,信息中心网络源数据的嵌入维特征量和主成分特征集为式(4):

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c \mu_{ik}^m (d_{ik})^2. \quad (4)$$

式中, m 为关联权重指数; $(d_{ik})^2$ 为相似度特征集,结合语义本体映射方法源数据的样本特征分布式重构,根据 x_k 与 V_i 的测度距离,得到第 i 个信息中心网络源数据的特征分布子图,由此实现信息中心网络源数据重组,结合重组模型进行特征挖掘和信息融合聚类^[5]。

1.2 数据的特征挖掘

在构建了信息中心网络源数据相空间重构模型的基础上,采用最小均方误差估计方法,进行了信息中心网络源数据的子图特征匹配^[6],构造数据的粗糙集分布模型,在特征估计值 $\hat{f}_x(X, t)$ 与关联规则集 $\hat{f}_\theta(X, t)$ 之间有如下关系,式(5)和式(6):

$$f_x(X, t) = \hat{f}_x(X, t) + \Delta f_x(X, t), \quad (5)$$

$$f_\theta(X, t) = \hat{f}_\theta(X, t) + \Delta f_\theta(X, t). \quad (6)$$

根据信息中心网络源数据的有限论域的特征匹配模型进行信息融合,得到关联规则调度集满足式(7):

$$|\Delta f_x(X, t)| \leq F_x(X, t) \quad |\Delta f_\theta(X, t)| \leq F_\theta(X, t). \quad (7)$$

对信息中心网络源数据的关联规则集进行自动排序,得到数据的特征融合分量描述为式(8)~式(10):

$$\frac{\partial F}{\partial o_k} = \frac{\partial F}{\partial y_k} \cdot \frac{\partial y_k}{\partial o_k}, \quad (8)$$

$$\frac{\partial F}{\partial z_{kj}} = \frac{\partial F}{\partial o_k} \cdot \frac{\partial o_k}{\partial z_{kj}}, \quad (9)$$

$$\frac{\partial F}{\partial w_{ji}} = \sum_{k=1}^m \frac{\partial F}{\partial o_k} \cdot \frac{\partial o_k}{\partial a_j} \cdot \frac{\partial a_j}{\partial n e t_j} \cdot \frac{\partial n e t_j}{\partial w_{ji}}. \quad (10)$$

取 $\sin \theta_p = \theta_p$, $\cos \theta_p = 1$, 进行网络源数据的特征挖掘和属性归集,描述为式(11):

$$\begin{cases} \dot{\hat{x}}_1 = x_3, \\ \dot{\hat{x}}_3 = f_\theta(X, t) + g_\theta(X, t)u(t) + d_\theta(t), \\ \dot{\hat{x}}_2 = x_4, \\ \dot{\hat{x}}_4 = f_x(X, t) + g_x(X, t)u(t) + d_x(t). \end{cases} \quad (11)$$

式中, $X = [\theta, x, \dot{\theta}, \dot{x}]^T$, $f_x(X, t)$, $f_\theta(X, t)$, $g_x(X, t)$, $g_\theta(X, t)$ 分别表示网络的源数据调度的资源匹配集, 采用语义特征融合方法, 在高维相空间中实现网络源数据的关联规则挖掘, 对提取的网络源数据的关联特征集分类融合。

2 数据检测优化

2.1 信息中心网络源数据分类融合

对提取的信息中心网络源数据的关联特征集进行分类融合, 构建基于模糊 C 均值聚类的多模级联分类器进行数据分类融合处理^[7], 当 $\forall i \in S_s$, 得到网络源数据的关联规则项 $\beta_i^c \neq \pm \#$, 在存在冗余数据的扰动下, 可得: $\forall i \in S - S_s, \gamma_i^c \neq \pm \#$, 在特征分布集合 S_s 中, 采用随机信息重组方法进行网络源数据的分类融合, 满足 $\forall i \in S_s, \beta_i^c \neq \pm \#$, 且 $\forall i \in S - S_s, \gamma_i^c \neq \pm \#$ 。对于网络源数据的采样节点 $\forall i \in S_s$, 数据的多分类器融合集为式(12):

$$\begin{aligned} \beta_i^c = & - \sum_{k \in S_s} R_{ik} Q_{kc} - R_{il} \gamma_c = \\ & - \frac{1}{\det(Q')} \left(\sum_{k \in S_s} (-1)^{i+k} \det(Q'_{\setminus ki}) Q_{kc} + \right. \\ & \left. \gamma_c (-1)^{i+1} \det(Q'_{\setminus i}) \right). \end{aligned} \quad (12)$$

信息汇集流分别满足 $\hat{q}_{i+1, i} = c_{i+1, i}$ 和 $\hat{q}_{i, i} = d_{i, i}$, 此时假设 $c_{i, i-1} \leq \min\{c_{i+1, i}, d_{i, i}\}$, 得到关联规则集为式(13)~式(15):

$$K(x_i, x_j) = \langle x_i, x_j \rangle, \quad (13)$$

$$K(x_i, x_j) = (\langle x_i, x_j \rangle + 1)^d, \quad (14)$$

$$K(x_i, x_j) = \exp(\|x_i - x_j\|^2 / 2\sigma^2). \quad (15)$$

式(15)中 $\sigma = 0.707$ 。根据网络源数据的特征分布属性进行分类融合, 得到融合结果为式(16)~式(19):

$$\max_{x_{a,b,d,p}} \sum_{a \in A} \sum_{b \in B} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} V_p, \quad (16)$$

$$\text{s.t.} \sum_{a \in A} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} R_p^{bw} \leq K_b^{bw}(S), b \in B. \quad (17)$$

$$\sum_{a \in A} \sum_{b \in B} \sum_{p \in P} x_{a,b,d,p} R_p^{cp} \leq K_d^{cp}(S), d \in D, \quad (18)$$

$$\sum_{b \in B} \sum_{d \in D} x_{a,b,d,p} \leq D_{a,p}, a \in A, p \in P. \quad (19)$$

在源数据的检测过程中, 得到源数据分类的迭代加速公式为式(20):

$$(1 - \omega)x_i^{(k)} + \omega x_i^{-(k+1)} = x_i^{(k+1)}, \quad i = 1, 2, \dots, n. \quad (20)$$

采用决策树信息融合方法进行网络源数据的多分类器融合, 实现信息融合优化。

2.2 数据检测的多模级联分类

构建基于模糊 C 均值聚类的多模级联分类器, 采用数据特征的融合法和分裂法进行数据聚类中心寻优控制, 令 $p = R(r) \Theta(\theta) Z(z) e^{j\omega t}$, 在置信度 α 下提取信息中心网络源数据的谱特征量为式(21):

$$R_{\tilde{c}}(\tau, \alpha) = E\{\tilde{c}^*(\tau, t) \tilde{c}(\tau, t + \alpha)\}. \quad (21)$$

在云计算环境下的信息中心网络源数据堆栈模型表示为式(22):

$$\frac{\partial J_{MCA}(n)}{\partial f_{ij}(n)} = \frac{1}{4} \frac{\partial \mathfrak{a}}{\partial \mathfrak{e}} \sum_{j=1}^M \frac{\partial J_{R,j}(n)}{\partial f_{ij}(n)} + \sum_{j=1}^M \frac{\partial J_{I,j}(n)}{\partial f_{ij}(n)} \frac{\partial \mathfrak{o}}{\partial \mathfrak{o}}. \quad (22)$$

结合模糊相关性的约束控制方法, 构建信息中心网络源数据的统计特征量, 表示为式(23):

$$e_{R,j} = (|y_{R,j}(n)|^2 - R_{2,R}) \times y_{R,j}(n)^*. \quad (23)$$

以特征提取结果为输入参考量, 得到信息中心网络源数据的分布式差异特征提取结果为式(24):

$$x_i(n) = \sum_{j=1}^M h_{ij}(n) T_{sj}(n) + v_i(n). \quad (24)$$

分别对 $i(i = 1, 2)$ 条网络源数据的检测通道进行信息融合, 滤除干扰信息, 由此确定网络源数据多维特征检测值为式(25):

$$u = [u_1, u_2, \dots, u_N] \in R^{mN}. \quad (25)$$

构建网络源数据统计特征分布模型, 式(26):

$$AVG_X = \frac{1}{m \times n} \sum_{x=1}^n \sum_{y=1}^m |G_X(x, y)|. \quad (26)$$

其中, $G_X(x, y)$ 是网络源数据的特征指向性函数, m, n 分别是网络源数据的嵌入维数和延迟, 结合定量递归分析方法, 实现网络源数据检测, 得到检测输出为式(27)~式(29):

$$WebJaccard(X, Y) = \frac{P(X \cap Y)}{P(X) + P(Y) - P(X \cap Y)}, \quad (27)$$

$$WebOverlap(X, Y) = \frac{P(X \cap Y)}{\min(P(X), P(Y))}. \quad (28)$$

$$WebDice(X, Y) = \frac{2P(X \cap Y)}{P(X) + P(Y)}. \quad (29)$$

其中, $P(X), P(Y)$ 表示网络源数据的联合关联规则分布集, X, Y 为网络源数据的采样延迟, $P(X \cap Y)$ 是网络源数据的互信息函数, X, Y 为网络源数据的模糊特征序列分布集。综上分析, 实现基于多

分类器融合的网络源数据检测。

3 仿真实验与结果分析

为了测试本文方法在实现信息中心网络源数据检测中的应用性能,结合 MATLAB 进行仿真实验,网络源数据库采用 Deep Web 200G,数据样本集为 2 000,测试集为 100,对信息中心网络源数据采集的时间长度为 1 024,离散采样频率为 $f_s = 10 * f_0$ Hz = 10 KHz,数据包的数量设置为 $N = 1 000$,根据上述仿参量设定,进行网络源数据检测,首先进行网络传输数据采样,得到采集的数据波形如图 2 所示。

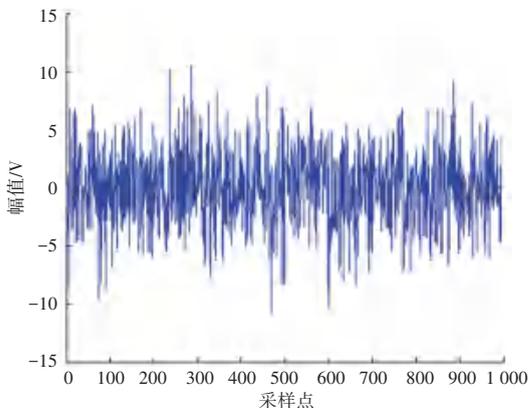


图 2 信息中心网络的数据采样

Fig. 2 Data sampling for information center network

采用相空间结构重组方法进行源数据的信息特征挖掘和重构,得到重构结果如图 3 所示。

在高维相空间中实现网络源数据的关联规则挖掘,对提取的网络源数据的关联特征集分类融合,实现数据的优化检测,得到检测输出如图 4 所示。

分析图 4 得知,采用本文方法能有效实现信息中心网络源数据检测,输出数据的抗干扰性较好,特征分辨能力较强,采用不同方法测试信息中心网络

源数据检测的准确性,得到对比结果见表 1,分析得知,采用本文方法进行信息中心网络源数据检测的准确概率较高。因为本文方法通过对提取的网络源数据进行了关联特征集合并和分类融合,提高了数据检测的精度,相比其他方法在抗干扰性和检测的准确性方面具有优越性。

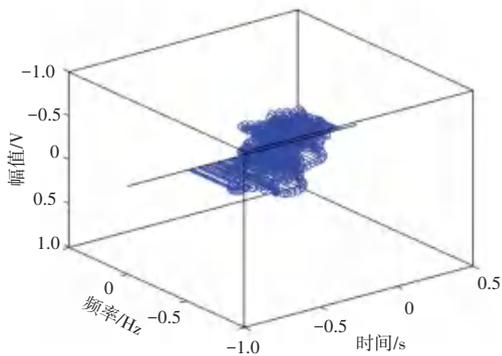


图 3 数据的相空间重构结果

Fig. 3 Phase space reconstruction results of data

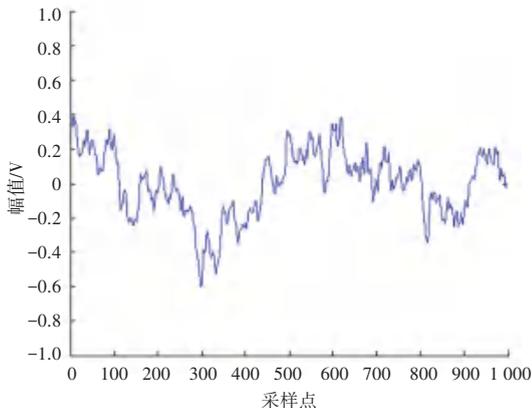


图 4 检测输出

Fig. 4 Detection output

表 1 准确性对比

Tab. 1 Comparison of accuracy

输入信噪比/dB	本文方法	基于非平稳数据融合的网络源数据的高效定位挖掘的检测方法	基于交互式信息融合的网络源数据检测方法
-10	0.873	0.769	0.803
-5	0.921	0.811	0.821
0	0.946	0.825	0.844
5	0.978	0.858	0.878
10	0.993	0.890	0.923

4 结束语

对信息中心网络源数据进行有效的识别和检测,提取网络源数据的关联规则集,滤除网络源数据的干扰信息,提高信息中心网络的数据管理能力,本

文提出一种基于多分类器融合的信息中心网络源数据检测方法。首先,构建网络源数据的节点匹配模式,采用最小均方误差估计方法,进行网络源数据的

(下转第 160 页)