

基于改进烟花算法的电力物联网状态可变数据流断点区优化检测仿真

邢巍[†], 李玥, 叶征楠

(国家电网有限公司客户服务中心, 天津 300300)

摘要:电力物联网状态可变数据流断点区检测对于提高数据完整性、质量、处理效率,促进电力物联网智能化发展以及辅助决策支持等方面具有重要的作用,故提出了基于改进烟花算法的电力物联网状态可变数据流断点区优化检测仿真。首先,使用基于自适应窗口滑动的CCA算法分析电力物联网状态可变数据流的变化趋势和特征,从而得到数据流的动态变化情况,为后续的断点检测提供数据支持;其次,设定数据流断点区距离阈值,根据该阈值求取断点区分布密度特征作为数据流断点区的检测标准;最后,根据断点区的分布密度特征等信息,采用改进烟花算法动态调整烟花的爆炸半径、爆炸次数等参数,从而更有效地检测到最精确的可变数据流断点区。实验结果表明,所提方法具有较高的检测精度和检测效率,且可以有效降低物联网的能耗,提升网络性能。

关键词:改进烟花算法;状态可变数据流;断点区检测;CCA算法;分布密度特征

中图分类号:TP393

文献标识码:A

Optimization and Detection Simulation of State Variable Data Flow Breakpoint Area in Power IoT Based on Improved Fireworks Algorithm

XING Wei[†], LI Yue, YE Zhengnan

(State Grid Customer Service Centre, Tianjin 300300, China)

Abstract: The detection of breakpoint areas in the variable data flow of the power Internet of Things plays an important role in improving data integrity, quality, processing efficiency, promoting the intelligent development of the power Internet of Things, and assisting decision-making support. Therefore, an optimized detection simulation of breakpoint areas in the variable data flow of the power Internet of Things based on an improved fireworks algorithm is proposed. Firstly, the CCA algorithm based on adaptive window sliding is used to analyze the trend and characteristics of the variable data flow in the power Internet of Things, in order to obtain the dynamic changes of the data flow and provide data support for subsequent breakpoint detection; Secondly, set the distance threshold of the breakpoint area in the data flow, and use this threshold to obtain the distribution density characteristics of the breakpoint area as the detection standard for the breakpoint area in the data flow; Finally, based on the distribution density characteristics of the breakpoint area and other information, an improved fireworks algorithm is used to dynamically adjust the parameters such as the explosion radius and number of explosions of the fireworks, in order to more effectively detect the most accurate variable data flow breakpoint area. The experimental results show that the proposed method has high detection accuracy and efficiency, and can effectively reduce the energy consumption of the Internet of Things and improve network performance.

Key words: improve fireworks algorithm; state variable data flow; breakpoint area detection; CCA algorithm; distribution density characteristics

1 引言

随着电力物联网技术的快速发展,状态可变量数据流在电力系统中得到了广泛应用^[1]。这些数据流反映了电力设备的运行状态和电力系统的稳定性,对于电力系统的监测、控制和优化具有重要意义。然而,在实际应用中,由于数据流具有可变性和实时性等特点,常常会出现数据断点,导致数据流的不连续和不准确。这不仅会影响电力系统的稳定运行,也会影响对电力系统的优化控制。因此,如何有效地发现和修复电力物联网状态可变量数据流中的断点,提高数据流的连续性和准确性,是当前亟待解决的问题。

武炜杰等^[2]提出一种基于 k 近邻的完全随机森林算法(KCPCForest),该算法利用电力物联网状态可变量数据流中的所有已知类样本的完全随机数展开计算。通过对比叶节点的平均路径长度,将数据划分为正常区和断点区。在此基础上,找出落入断点区范围内的 k 近邻样本,并计算出离散化阈值。利用该阈值与预设的阈值进行比较,如果大于预设阈值,则判断其为新断点区范围。当检测到达规定数量时,调整信息模型,并得出最终的检测结果。该方法未考虑到状态可变量数据流的动态变化特性,检测精度有待检测。胡梦娜等^[3]采用 BiGAN 基本网络框架首先学习真实复杂的状态可变量数据流在物联网中的分布情况;其次根据 GAN 的鉴别器特性采用双向长短期记忆网络捕捉数据流序列分布的时间相关性,在此基础上引入注意力机制,建立数据流断点区域检测模型,通过该模型找到断点区域数据流的相关信息,以此判断断点区域。该方法的计算时间过长,降低了检测效率。

为此,提出了基于改进烟花算法的电力物联网状态可变量数据流断点区优化检测仿真。相比传统的烟花算法,改进烟花算法引入了动态调整策略和变异机制等改进措施,根据数据流的动态变化特性,该算法能够自动调整烟花的“爆炸半径”和“爆炸次数”等参数,使得其能够更好地适应数据流的实时变化,提高搜索效率和寻优能力;优化了计算数据点之间距离或相似度的方式,使得计算更加快速和准确,有助于更高效地判断是否存在断点,提高检测的准确性;引入了基于断点区分布密度的动态调整策略,使得该算法能够根据实际情况,动态调整断点检测的

阈值,确保检测的准确性和可靠性。

1. 数据流相关分析在电力物联网中,数据流是连续的,使用基于自适应窗口滑动的 CCA 算法^[4],可以更好地捕捉和处理电力物联网状态可变量数据流的变化,从而优化断点区检测的效果。

首先定义如下几个参数:

(1)将电力物联网状态可变量数据流定义为允许元素重复出现的无限集合,即:

$$S = \{[s_1, t_1], \dots, [s_i, t_i], \dots, [s_n, t_n]\}, n \rightarrow \infty \quad (1)$$

式中: s 表示数据流标识; t 代表 s 的时间戳。

(2)假设状态可变量数据流依照时间戳 t 的先后顺序流入滑动窗口 w , 那么在任意时间段内各个窗内的数据流都可以表现为如下序列:

$$\begin{cases} W = \{[s_l, t_l], \dots, [s_{l+1}, t_{l+1}], \dots, \\ [s_{u-1}, t_{u-1}], [s_u, t_u]\} \\ K = u - l \end{cases} \quad (2)$$

式中: K 表示窗口内保存的最近标识数量; l 表示进入滑动窗口 w 的时间; u 表示最后进入滑动窗口的标识。

电力物联网状态可变量数据流的处理是以自适应和近似查询为核心的,为此需要计算数据流的两个临界速率以设计不同的窗口滑动策略,进而保证 CCA 的实时性和准确性。

采用基于时间定义的滑动窗口模型^[5],设有效窗口宽度为滑动窗口 w 中元数组 $\text{Size}(w)$ 的最小值,用 EffWin 表示,根据上述计算求得的临界速率和自适应窗口滑动策略,展开以下相关分析流程:

(1)在 t_u 时刻内输入维数为 p 和 q 的电力物联网装填可变量数据流 X 和 Y , 其速率分别为 $R_X(t)$ 和 $R_Y(t)$ 。

(2)初始化窗口宽度 EffWin 、最大等待时间 Δt 以及 CCA 计算时间 T_{CCA} 并按照以下步骤计算出临界速率 R_1 和 R_2 :

①临界速率 R_1 的计算条件为:

$$\text{Size}(w) = \int_{t_u}^{t_u + \Delta t} R(t) dt = \text{EffWin} \quad (3)$$

式中: $R(t)$ 为 t 时刻状态可变量数据流 s 的速率。

②临界速率 R_2 的计算条件为:

$$\int_{t_u}^{t_u + T_{\text{CCA}}} R(t) dt = \text{EffWin}, R_1 < R(t) \leq R_2 \quad (4)$$

(3)此时 $R(t) = \text{MIN}(R_X(t), R_Y(t))$, 选择相应的数据流速率与临界速率比较,选择不同的分析

策略:

①若 $R(t) \leq R_1$, 以 Δt 为步长展开窗口滑动; 在此时下一滑动窗口 w 的开始时间 t_{t+1} 和结束时间 t_{u+1} 可定义为:

$$\begin{cases} \text{Size}(w) = \int_{t_{t+1}}^{t_{u+1}} R(t) dt = \int_{t_u}^{t_u + \Delta t} R(t) dt = \\ \text{EffWin} \\ t_{u+1} = t_u + \Delta t \end{cases} \quad (5)$$

②若 $R_1 < R(t) \leq R_2$, 以 EffWin 为步长展开窗口滑动;

③若 $R(t) > R_2$, 以 T_{CCA} 为步长展开窗口滑动。

(4)若 $\int_{t_i}^{t_u} R(t) dt = \text{EffWin}$, 则执行步骤(4)操作, 反之, 重新步骤(3)操作; 以此计算出最终的 CCA, 输出 t_u 时刻内状态可变量数据流 X 和 Y 之间的相关变化特性。

2 状态可变量数据流断点区优化检测

2.1 断点区距离阈值设定

将电力物联网状态可变量数据流转换为一组与时间相关的信息序列模型^[6]:

$$DS = \{(t_0, y_0), (t_1, y_1), \dots, (t_i, y_i)\} \quad (6)$$

式中: y_i 代表 t_i 时刻到来的数据流对象; 假设任意两个数据流对象为 p 和 o , 且 $p(x_i, y_i), o(x_i, y_i) \in DS$, 则根据欧几里得算法可以获取两个对象之间的距离 $d(p, o)$:

$$d(p, o) = \sqrt{(x_i^2 - x_j^2) + (y_i^2 - y_j^2)} \quad (7)$$

以 ξ 为数据流断点区标准正态分布函数, γ 为断点区检测约束条件, 设定其距离阈值 Z :

$$Z = ((\gamma \pm \xi)/k) + (d(p, o)/(S \times R)) \quad (8)$$

式中: k 为状态可变量数据流断点区置信区间; S 和 R 为断点区的规模和时间序列, 以此得到断点区的距离阈值。

2.2 求取断点区分布密度特征

根据第 2 节分析到的电力物联网状态可变量数据流动态变化特性以及设定的断点区距离阈值, 求取不同分布密度的数据流密度特征, 进而获取不同数据断点区的权值信息并加以标识, 通过分析断点区的分布密度^[7], 可以确定断点高发区和相对稳定区, 以便后续可以有针对性地分配检测资源和能量, 降低能耗。

采用转移矩阵获取数据流断点区数据流节点

的位置分布连接图, 如图 1 所示。

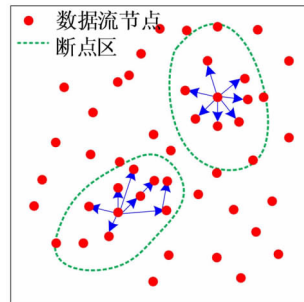


图 1 断点区数据流节点位置分布图

根据图 1, 计算电力物联网平衡状态下的叠加函数和各数据流区域的状态指数, 用一种数据流状态移动矩阵描述:

$$U = \frac{c \pm A}{s} \mp y \quad (9)$$

式中: s 为电力物联网传输数据流信息的目标情况; A 表示数据流动态下的状态移动矩阵; y 代表数据流断点区信号时效频率特征; c 代表高斯白噪声干扰。

设 ϑ 为数据流分布频率区域的特征状态指数, 用 $\vartheta = \beta \mp \eta$ 求出时效频率空间范围内传输数据流分布均匀的特征收敛函数, 其中 β 为传输数据流断点区后方验证概率估算值; η 为数据流在每个区域产生的分布状态函数, 由此选择相应的断点区分布密度特征, 作为断点区检测的标准。

2.3 断点区优化检测

改进烟花算法^[8,9]是一种优化搜索算法, 通过模拟烟花的爆炸和扩散过程来实施全局搜索。相比传统的烟花算法, 改进烟花算法引入了动态调整策略和变异机制等改进措施, 针对电力物联网状态可变量数据流断点区优化检测过程, 该算法主要有以下几个优势:

(1)通过第 2 节分析数据流的动态变化特性, 可以调整烟花的爆炸半径、爆炸次数等参数, 以提高算法的搜索效率和寻优能力。

(2)通过优化计算方式, 在 3.1 节的基础上, 用更快速准确的方式计算数据点之间的距离或相似度, 从而判断是否存在断点。

(3)引入动态调整策略, 根据断点区分布密度特征动态调整断点检测的阈值, 以适应不同情况下的断点检测需求。

具体步骤如下:

(1)初始化烟花种群(即状态可变量数据流)大小 pop; 全局最优解(即电力物联网状态可变量数据流

断点区最优检测结果)为 gBest,最优爆炸火花个体集合为 sparkpBest。

(2)从数据流的第一个数据点开始,依次遍历每个数据点,获取每个烟花(即数据点)的适应度值并以此更新 gBest 的状态。

(3)利用高斯变异增强算法计算种群的多样性,通过越界处理将超过检测范围的烟花个体映射到搜索空间内,加入一个高斯分布函数^[10] $N(0, 1)$,得到如下的计算公式:

$$\begin{cases} \text{gBest} = \text{gBest} \cdot [N(0, 1) + 1] \\ x_i^D = x_{\text{down}}^D + |x_i^D| \% (x_{\text{up}}^D - x_{\text{down}}^D) \end{cases} \quad (10)$$

式中: x_i^D 表示第*i*个个体在第*D*维的位置; x_{up}^D 和 x_{down}^D 为其上下界;%用于表示模运算。

(4)计算烟花的爆炸强度,即第*i*个烟花所产生爆炸火花的总数量。爆炸强度的强弱决定烟花适应度的高低,也同时代表了该点为断点的可能性大小,设定目标适应度值为 $f(x_i)$,爆炸强度计算公式为:

$$S_i = q \frac{Y_{\max} - f(x_i) + \bar{\omega}}{\sum_{i=1}^{\text{pop}} [Y_{\max} - f(x_i)] + \bar{\omega}} \quad (11)$$

式中: q 用于表示爆炸火花数调节常数; $Y_{\max}$ 所代表的是所有烟花个体中适应度值最高的个体; $\bar{\omega}$ 表示极小机器数。

(5)通过数据流断点区的分布密度特征,动态调整烟花的爆炸半径,以第*i*个产生爆炸火花子的烟花种群中最优个体 sparkpBest(*i*)为目标,获取断点区检测敏感度:

$$\begin{cases} \text{EA}(i) = \begin{cases} \text{sparkpBest}(i) - x(i), S_i < \text{AugS} \\ \text{gBest} - x(i), S_i \geq \text{AugS} \end{cases} \\ \text{AugS} = \frac{1}{\text{pop}} \sum_{i=1}^{\text{pop}} S_i \end{cases} \quad (12)$$

式中:AugS所描述的是爆炸火花均值。爆炸半径越小说明能够检测到的断点更多,反之则会漏判断点,因此选择合适的爆炸半径是平衡断点区域检测的关键。

(6)根据烟花个体产生爆炸火花时的位置信息描述新的爆炸半径,在各个维度上自适应调整步长,应用不同的位移偏移^[11]来实现位移操作并继续检测,设定一个烟花爆炸半径 EA(*i*),具体计算公式为:

$$x_{\text{last}}(z) = x(i) + e \cdot \text{EA}(i) \quad (13)$$

式中: $z = 1, 2, \dots, S_i$ 表示第*i*个烟花未爆炸前的位置; $x_{\text{last}}(z)$ 用于描述烟花种群中第*i*个烟

花爆炸后产生的火花中第*z*个爆炸火花的位置; e 指随机数,均匀分布在 $[0, 1]$ 范围内。

(7)对烟花随机选取多个维度实施相应操作,以此产生高斯火花,生成状态可变数据流的断点区域解集 $x_i^D = x_i^D \times N(1, 1)$ 。

(8)根据设定的数据流断点区距离阈值,通过基于欧式距离 $d(x_i, x_j)$ 获取任意两个烟花个体 x_i 和 x_j 间的距离,若超过了设定的阈值则说明两个烟花之间产生了断点:

$$E(x_i) = \sum_{z=1}^K d(x_i, x_z) = \sum_{z=1}^K \|x_i - x_z\| \quad (14)$$

式中: $E(x_i)$ 为 x_i 到其他个体的距离总和; K 所代表产生火花的烟花种群位置集合; $z \in K$ 表示第*z*个位置属于*K*。

采用轮盘赌的方式选取与其他个体距离更远的个体,将其更新为下一代烟花种群个体,并由此得到下一个可能的数据流断点区域,被选取概率^[12]的计算公式为:

$$p(x_i) = E(x_i) / \sum_{z \in K} E(x_i) \quad (15)$$

(9)迭代执行步骤(2)到步骤(8),直至遍历完整个电力物联网状态可变数据流区域,符合终止条件后停止操作,以此实现断点区的全面优化检测。

3 实验与分析

为了验证基于改进烟花算法的电力物联网状态可变数据流断点区优化检测仿真的整体有效性,需要对其展开测试。实验首先分析所提方法的最大适应度波动情况,其次从检测能耗和检测时延两个方面对比所提方法和文献[2]方法、文献[3]方法的检测性能。

3.1 最大适应度值波动

在电力物联网状态可变数据流断点区优化检测中,最大适应度波动可以反映算法对于不同变化特性数据流的识别和处理能力。最大适应度波动较小说明算法的适应性和泛化能力较强。实验选取某一电力企业公司某一段时间段内的含有断点数据流区故障的电力物联网数据作为测试样本,采用所提方法对其断点区展开检测,获取其最大适应度值波动状态,以此验证所提方法的稳定性,结果见图 2。

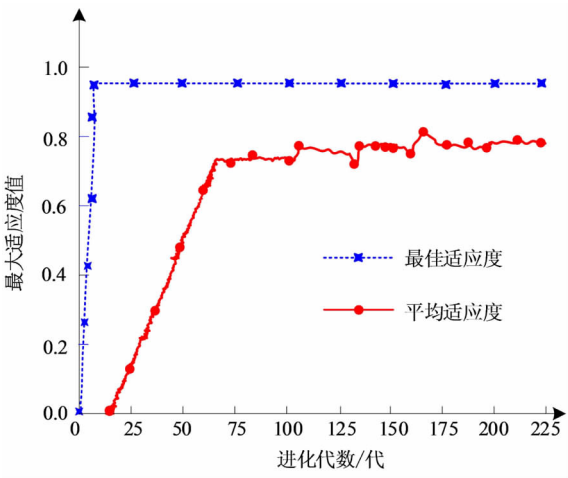


图 2 所提方法最大适应度波动

分析三维图 2 可知,当进化代数随着时间增加而提高时,所提方法的最大适应度也有所提高,但

很快达到平稳状态,此时的进化代数约为 15 代,达到平稳的用时较短,平均适应度也随之稳定,这是因为所提方法在检测前首先对电力物联网状态可变数据流展开了全面分析,对断点区的分布特征也作了深入的研究,经过这几个步骤后,一定程度上提高了所提方法的检测精度和稳定性。

3.2 检测能耗对比

为了进一步评估所提方法的优异性能,选择能耗作为评估指标,能耗指的是断点检测过程中所消耗的能源,例如 CPU、内存、磁盘等硬件设备的能源消耗。对于大规模且动态变化较复杂的电力物联网状态可变数据流断点区检测,能耗是制约物联网性能和效率的重要因素之一。实验在数据流不断增加的条件下,采用所提方法展开测试,并与文献[2]方法和文献[3]方法比较,三种方法的能耗对比结果见图 3。

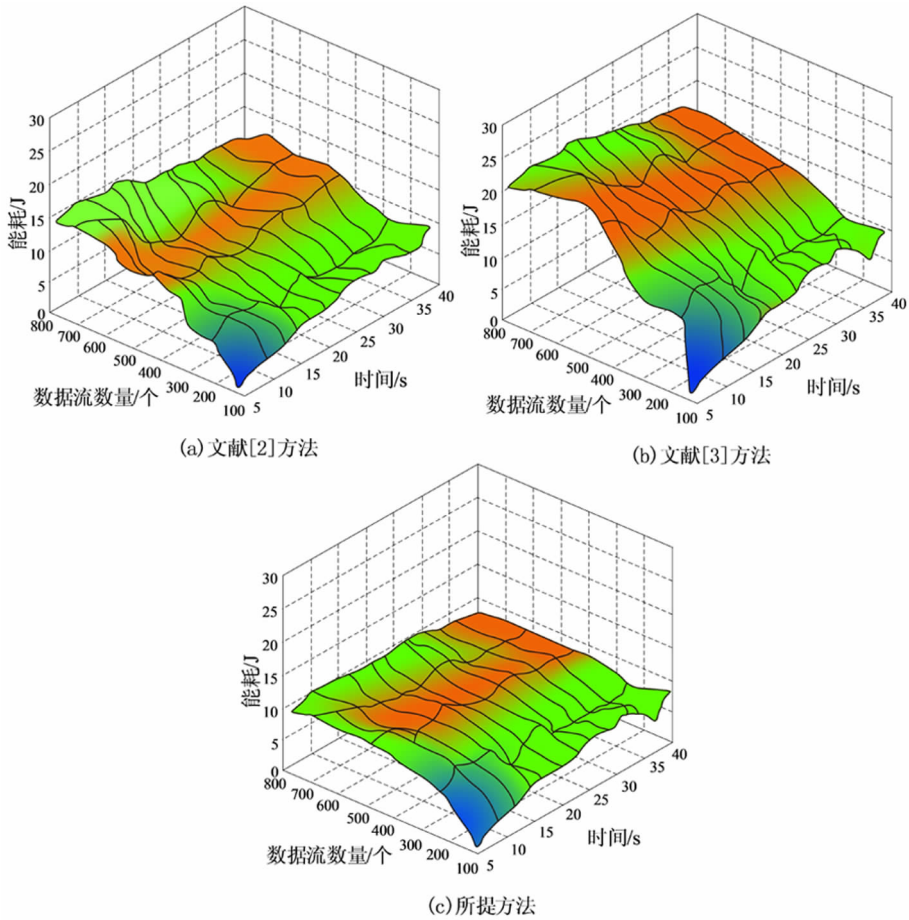


图 3 不同方法检测能耗

通过分析图 3 中的数据可知,状态可变数据流的动态特性和数量增加给三种方法的断点区检测过程均带来了不同程度的影响,相比之下,文献[3]方法的能耗始终高于所提方法和文献[2]方法,说

明该方法的检测性能较差,无法保证电力物联网的性能;文献[2]方法能耗比所提方法的能耗略高一些,说明该方法在数据流断点区检测方面还需进一步增强,只有所提方法能够保持较低的能耗,有效

实现所有数据流断点区的稳定检测,证明了所提方法的性能。

3.3 检测实时性对比

以检测时延为衡量三种检测方法实时性的评估指标,从数据流断点发生到断点区被检测出来的时间延迟对于实时性要求较高的电力物联网数据流断点区检测应用来说,较短的检测时延是有利的。

针对状态可变量数据流在电力物联网中的传输速率多变性,实验在降低和增加数据流传输速率的情况下,测试三种方法的实时检测能力,得到如表 1 所示的对比结果。

表 1 不同方法检测实时性对比			
传输速率较慢情况下时延/(m/s)			
传输速率/ (bt/s)	所提方法	文献[2]方法	文献[3]方法
49.8	0.15	1.05	0.28
50.2	0.18	1.11	0.44
51.1	0.22	1.37	0.61
53.4	0.25	1.55	0.74
55.3	0.29	1.71	1.02
传输速率较快情况下时延/(m/s)			
传输速率/ (bt/s)	所提方法	文献[2]方法	文献[3]方法
97.9	0.08	0.33	0.21
99.8	0.11	0.58	0.28
102.2	0.13	0.77	0.35
103.8	0.16	1.02	0.41
105.4	0.21	1.22	0.55

由表 1 可得,当状态可变量数据流的传输速率较慢时,三种方法需要等待较长时间才能接收到足够的事件数据来实现断点检测,时延也相对增加,相对而言所提方法的时延能更低一些;当状态可变量数据流的传输速率较快时,验证的是数据流断点区检测方法处理和响应事件数据的效率能否及时检测断点区域以减少时延。相比之下,所提方法可以更及时地应对传输速率较快时的数据流断点区检测,时延更低。由此可以充分说明所提方法具有更高的数据流断点区检测效率,能够在降低时延的同时,提高检测的实时性和稳定性。

4 结 论

通过深入研究数据流的特性和断点分布规律,

探索更加准确、高效、适应性强、易于实现的优化检测算法和技术,能够更快地促进电力物联网的发展。所提方法通过对状态可变量数据流的动态变化分析和数据流断点区域的阈值设定和特征求取,为断点区检测提供了各种数据支持;引入了改进烟花算法,在迭代过程中采用一系列动态调整寻找到最为精确的断点区域范围,更利于后续断点修复,保障电力物联网的安全可靠性,具有重要的实际应用意义。

参考文献

[1] 许源佳,吴恒,杨晨,等. 面向状态可变量数据流的集群调度综述[J]. 计算机学报,2022,45(5):973—992.

[2] 武炜杰,张景祥. 有新类的动态数据流分类算法研究[J]. 计算机科学与探索,2021,15(1):132—140.

[3] 胡梦娜,何强,贾俊铖,等. EB-GAN: 基于 BiGAN 的网络流量异常检测方法[J]. 计算机应用与软件,2023,40(6):303—309.

[4] 章壮伟,邢天阳,朱小良. 基于 CCA-DCNN 算法和数字孪生的给水回热系统故障诊断研究[J]. 热能动力工程,2023,38(7):175—184.

[5] 程浩东,韩萌,张妮,等. 基于滑动窗口模型的数据流闭合高效用项集挖掘[J]. 计算机研究与发展,2021,58(11):2500—2514.

[6] 吴明晖,张广洁,金苍宏. 基于多模态信息融合的时间序列预测模型[J]. 计算机应用,2022,42(8):2326—2332.

[7] 任艳平,郑重,江一飞,等. 融合后验概率和密度的不平衡数据欠采样方法[J]. 计算机工程与应用,2022,58(23):268—277.

[8] 于军琪,王福,赵安军,等. 基于改进烟花算法的并联冷机负荷分配优化[J]. 控制与决策,2021,36(11):2618—2626.

[9] 潘钰华,何振学,王克俭. 基于改进烟花算法的 MPRM 逻辑电路面积优化[J]. 兰州大学学报(自然科学版),2021,57(3):416—422.

[10] 汤安迪,韩统,徐登武,等. 使用高斯分布估计策略的改进樽海鞘群算法[J]. 系统工程与电子技术,2022,44(7):2229—2240.

[11] 刘哲锋,朱新革,徐志杰,等. 基础环式风机基础的动态偏移特征研究[J]. 水力发电,2021,47(8):105—109.

[12] 邹耀斌,张进玉,臧兆祥,等. 概率分布双向稀疏化下单一 Tsallis 熵阈值选取方法[J]. 软件学报,2022,33(5):1922—1946.