

马宪敏, 崔元全, 李放. 基于可穿戴式纳米生物传感器的人体运动数据挖掘算法[J]. 智能计算机与应用, 2024, 14(8): 220-224. DOI:10.20169/j.issn.2095-2163.240835

基于可穿戴式纳米生物传感器的人体运动数据挖掘算法

马宪敏¹, 崔元全², 李放³

(1 黑龙江外国语学院 东语系, 哈尔滨 150025; 2 哈尔滨师范大学 人事处, 哈尔滨 150025;

3 黑龙江外国语学院 信息工程系, 哈尔滨 150025)

摘要: 针对当前人体运动数据挖掘算法无法对实时数据进行采集与分析, 导致人体运动数据挖掘正确率较低且时间较长的问题, 提出基于可穿戴式纳米生物传感器的人体运动数据挖掘算法。首先, 利用可穿戴式纳米生物传感器采集人体运动数据, 将采集到的数据转换为二进制数据形式, 并对转换后的数据进行清洗与补位处理; 最后, 使用萤火虫算法对 K 均值聚类方法进行优化, 利用优化后的 K 均值聚类方法对清洗与补位后的数据进行聚类处理。实验结果表明, 所提算法的召回率平均值为 97.12%, 数据挖掘正确率平均值为 98.42%, 为运动员生理指标的实时监测与分析提供重要的数据基础。

关键词: 纳米生物传感器; 人体运动; 数据挖掘; 数据清洗; K 均值聚类算法; 数据采集

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2024)08-0220-05

Data mining algorithm for human motion based on wearable nano biosensor

MA Xianmin¹, CUI Yuanquan², LI Fang³

(1 Department of Eastern Language, Heilongjiang International University, Harbin 150025, China;

2 Human Resources Office, Harbin Normal University, Harbin 150025, China;

3 Department of Information Engineering, Heilongjiang International University, Harbin 150025, China)

Abstract: To address the problem that current human motion data mining algorithms cannot collect and analyze real-time data, which leads to low correct rate and long time for human motion data mining, we propose a human motion data mining algorithm based on wearable nano-biosensor. Firstly, the human motion data is collected using the wearable nano-biosensor. Secondly, the collected data are converted into binary data form, and the converted data are cleaned and complemented. Finally, the K-mean clustering method is optimized using the firefly algorithm, and the optimized K-mean clustering method is used to cluster the cleaned and patched data. The results show that the average recall rate of the proposed algorithm is 97.12% and the average correct data mining rate is 98.42%, which provides an important data base for real-time monitoring and analysis of athletes' physiological indexes.

Key words: nano biosensors; human motion; data mining; data cleaning; K-means clustering algorithm; data acquisition

0 引言

当前, 世界各国均掀起了运动热潮, 各类智能手环及智能手机运动软件也得以广泛发展^[1-3]。由于此类智能设备仅能对人体的运动时间、运动速度及心率等数据进行采集, 对于专业运动员的训练工作帮助不大^[4-5]。随着可穿戴式纳米生物传感器的发展, 大量新型可穿戴式纳米生物传感器, 可对运动员的心脏实时数据、呼吸实时数据进行监测, 使运动员生理指标的实时监测成为可能。为获取生理指标实

时数据背后的关键信息, 需要采用数据挖掘算法对人体运动数据进行聚类分析^[6-7]。

在以往的研究中, 大量的研究人员提供多种数据挖掘算法, 但对于人体运动数据的挖掘工作而言, 还存在一定程度上的不足^[8-10]。例如, Zong X 等^[11]研究了基于神经网络的人体运动数据挖掘算法, 构建了基于离散动态建模技术的数据挖掘模型, 并结合多层传感器实现了从运动特征和类型方面收集数据信息。但是在实际应用中发现, 这种算法存在计算效率低的问题。Mohamed M E 等^[12]研究了基于

基金项目: 2023 年度黑龙江省哲学社会科学研究规划项目 (23JYB252); 黑龙江省教育科学规划重点课题 (GJB1422539)。

作者简介: 马宪敏 (1979-), 女, 博士研究生, 教授, 硕士生导师, 主要研究方向: 大数据, 软件工程; 李放 (1980-), 女, 硕士, 副教授, 主要研究方向: 大数据, 软件工程。

通讯作者: 崔元全 (1979-), 男, 硕士, 副研究员, 主要研究方向: 大数据, 教育管理。Email: 12389135@qq.com

收稿日期: 2023-05-24

自然语言处理与决策树的数据挖掘算法。利用传感器采集运动员相关数据, 以此实现数据挖掘。但此部分数据不具备实时性, 导致算法在人体运动数据的分析中出现异常。Liu Meiling 等^[13] 研究优化 K 均值的数据挖掘算法, 利用熵值法对数据对象的属性赋权来修正对象间的欧氏距离, 通过比较初聚类的赋权类别目标价值函数, 以此实现数据挖掘, 但是此算法的应用过程还需进行优化。Li Hua 等^[14] 设计了改进 K 均值和 LSTM 的数据挖掘算法, 选取了 9 个典型特征, 但该算法无法直接应用到体育数据中。Wang Guangtong 等^[15] 提出了一种应用粗糙熵的数据挖掘算法, 根据数据的实际表现和数据特征, 利用数据挖掘技术发现股市数据中的异常数据, 并采用基于密度和距离的孤立点检测方法对获取的异常数据进行分析, 以获取其隐含的有用信息。此算法计算能力较高但计算时间过长。

针对上述问题, 本文提出了一种基于可穿戴式纳米生物传感器的人体运动数据挖掘算法。首先, 利用这种传感器采集人体运动数据, 并通过数据清洗和聚类处理, 实现有效的数据挖掘, 为了提高聚类的精准度, 采用了萤火虫算法优化 K 均值聚类方法。实验结果表明, 所设计的算法能够快速、精确地挖掘人体运动数据, 并为体育训练提供决策信息, 推动了体育训练技术的发展。

1 人体运动数据挖掘算法

针对当前人体运动数据挖掘算法无法对实时数据进行采集与分析, 导致人体运动数据挖掘正确率较低且时间较长的问题, 提出基于可穿戴式纳米生物传感器的人体运动数据挖掘算法。具体来说, 通过对采集到的数据进行分析 and 挖掘, 可以实现以下目的:

- (1) 评估运动员的身体状态和运动表现: 通过分析心率、步数、速度、肌肉活动等数据, 可以评估运动员的身体状态和运动表现, 以确定训练强度和频率, 提高训练效果。
- (2) 监测和预警运动损伤和疲劳: 通过分析运动员的肌肉活动和能量消耗等数据, 可以监测和预警运动损伤和疲劳, 以避免过度训练和运动损伤的发生。
- (3) 优化训练和竞赛策略: 通过分析运动员的姿势、步态、动作等数据, 可以优化训练和竞赛策略, 提高运动员的竞技水平和表现。
- (4) 个性化训练和康复方案: 通过分析运动员的身体状态和运动表现, 可以为不同的运动员制定个性化的训练和康复方案, 以满足不同运动员的需求和要求。

求和要求。

(5) 科学决策和管理: 通过对大量运动数据进行挖掘和分析, 可以为教练、医生和管理人员提供科学的决策和管理依据, 以改善训练和竞赛的质量和效果。

由此可见, 基于可穿戴式纳米生物传感器的人体运动数据挖掘的目的, 是为了提高运动员的竞技水平和身体健康状况, 优化训练和竞赛策略, 实现个性化训练和康复方案, 为科学决策和管理提供依据。

1.1 可穿戴式纳米生物传感器数据采集

传统体育数据分析研究过程中, 由于无法获取到精准的实时数据, 导致分析结果可靠性较低^[16-17]。在本研究中, 使用可穿戴式纳米生物传感器作为人体运动数据的采集设备, 可获取运动员的实时运动数据。可穿戴式纳米生物传感器架构如图 1 所示。

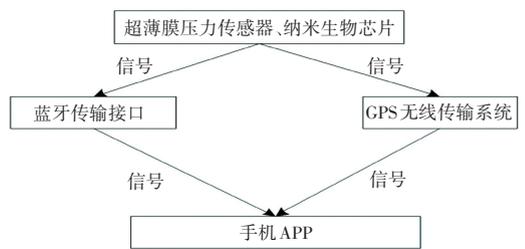


图 1 可穿戴式纳米生物传感器架构

Fig. 1 Wearable nanosensor architecture

由图 1 可见, 可穿戴式纳米生物传感器架构由超薄膜压力传感器、纳米生物芯片、蓝牙传输端口、GPS 无线传输系统以及手机 APP 组成。将超薄膜压力传感器、纳米生物芯片安装至运动员穿戴的弹性运动服装中采集人体运动信号, 将信号通过蓝牙传输端口、GPS 无线传输系统传输至手机 APP 中, 以此保证数据的采集质量与效率。

可穿戴式纳米生物传感器的人体运动数据采集流程如图 2 所示。

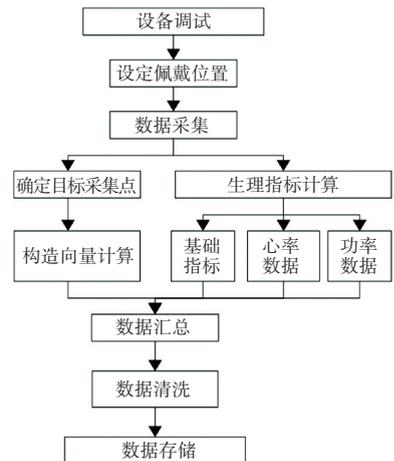


图 2 数据采集流程

Fig. 2 Data collection process

可穿戴式纳米生物传感器作为一种智能运动监测设备,可在运动员的运动过程中获取运动员的多种生理指标,为后续的数据挖掘提供数据基础。在此设备进行数据采集的过程中,还需对其进行相应的设定。假设可穿戴式纳米生物传感器佩戴在人体四肢处,将其作为数据采集点,此点可表示为 $q(x, y, z)$ 。在数据采集的过程中,需要将其构造向量导入数据采集结果中,此构造向量可以表示为:

$$\begin{cases} \bar{o}_1 = q_1 - q_0 \\ \bar{o}_2 = q_2 - q_1 \\ \bar{o}_3 = q_3 - q_2 \\ \bar{o}_4 = q_4 - q_3 \end{cases} \quad (1)$$

式中: q_0, q_1, q_2, q_3 以及 q_4 表示可穿戴式纳米生物传感器佩戴点, $\bar{o}, \bar{o}_1, \bar{o}_2, \bar{o}_3$ 表示不同的构造向量。

使用此公式,对人体运动过程中的基本生理指标进行采集,部分生理指标无法采用数据的形式进行体现,需要将其转化为数据信息,才能完成后续的分析工作。因此研究中选取下述指标进行转化处理:

1) 实时心率

$$g = \frac{\alpha}{(r_1 - r_2)} = \frac{(e_1 - e_2)}{2} \quad (2)$$

式中: r_1, r_2 分别表示此时段前一时间节点与此时段的心跳数据, α 表示心率计算标准系数, e_1, e_2 分别表示此时段前一时间节点及此时段的血压数据。

2) 运动总功率

假设运动员的实时生理指标数据拟合函数 $f(a)$ 在某恒定区间内连续,则根据此拟合函数可得到运动员的运动总功率,具体计算结果可表示为

$$\int_j^i f(a) = f(j) - f(i) \quad (3)$$

式中: $f(i)$ 表示区间端点 1 的拟合函数计算结果, $f(j)$ 表示区间端点 2 的拟合函数计算结果。

根据上述公式,对此两项生理指标进行整理与分析后,将其与可穿戴式纳米生物传感器采集结果导入同一数据库中,为后续的人体运动数据挖掘提供基础。

1.2 人体运动数据整理与分析

由于可穿戴式纳米生物传感器采集到的原始数据类别较为丰富且量纲不同,为了更好地完成数据挖掘过程,需要对采集到的数据进行预处理,将数据整合为二进制数据的形式。

假设采集到的原始数据为 $E = (1, 2, 3, \dots, n)$,

生理指标数据 v 的二进制位数 β_v 可表示为

$$\beta_v = (\beta_{v_1}, \beta_{v_2}, \dots, \beta_{v_n}) \quad (4)$$

式中: n 表示生理指标的个数。在此部分处理完成后,设定 H 为数据 3 挖掘候选数据集,对各生理指标的基础支持度展开计算,则有:

$$\sup(H) = \sum_{i=1}^n N_i \quad (5)$$

式中:支持系数 N_i 的计算公式如下:

$$N_i = \begin{cases} 1, & \beta_v \cap H = H \\ 0, & \beta_v \cap H \neq H \end{cases} \quad (6)$$

根据上述公式,将数据整合为二进制数据的形式,导入数据库中。可穿戴式纳米生物传感器采集到的原始数据中或存在异常数据,需要对此部分数据进行清洗,对此数据中的数据进行二次分析。假设数据为完整数据,则 $\beta_v = 1$; 若数据为异常数据,则 $\beta_v = 0$ 。发现异常数据后,使用插值法对数据进行补位处理。数据库全部数据分析完成后,进行人体运动数据挖掘。

1.3 人体运动数据挖掘

在上文基础上,引用数据挖掘技术中的 K 均值聚类分析方法,对数据进行挖掘处理,构建人体运动数据挖掘算法^[18]。

假设此部分数据可划分为 k 个子类,在子类划分的过程中需要进行多次迭代计算,则各子类之间的变差可表示为

$$U = \sum_{i=1}^n \sum_{w \in L_i} \text{dist}(w, l_i)^2 \quad (7)$$

式中: U 表示数据类别变差, l_i 表示数据簇代号, w 表示各数据簇的质心, $\text{dist}(w, l_i)$ 表示观测数据与数据簇质心的欧式距离。将此公式作为基础,构建人体运动数据挖掘算法。

根据体育数据的特征,对其进行完善与优化。由于人体运动数据采集量相对较大,在计算的过程中难免出现计算时间较长的问题。为此,在此研究中使用萤火虫算法对其展开优化。

假设各原始数据的核心信息为 C , 此时存在:

$$C_i = f(A_i) \quad (8)$$

式中: A_i 表示基础信息值, $f(A_i)$ 表示体育信息数据挖掘目标函数。根据此公式,确定原始数据之间的吸引力,则有:

$$\mu = \mu_0 * \exp(-vp_{ij}) \quad (9)$$

式中: μ_0 表示数据之间的初始吸引力, p_{ij} 表示两原始数据之间的欧式距离, v 表示不同数据之间的吸引力吸收系数。当两数据的基础信息量不同时,数

据的聚类挖掘过程可表示为

$$g_{i+1} = g_i + \mu(g_j - g_i) + \eta(rand - \frac{1}{2}) \quad (10)$$

式中: g_i 与 g_j 表示两体育数据的空间位置, μ 表示数据吸引度, η 表示随机常数, $rand$ 表示从 0 到最大随机数的任意整数。

将公式(10)与公式(7)相结合,使其完成数据挖掘过程。至此,基于可穿戴式纳米生物传感器的人体运动数据挖掘算法设计完成。

2 实验结果分析

2.1 数据集

本文实验选择 Actions 数据集与实测数据集作为数据来源。Actions 数据集中包含 7 种运动数据,分别为跳水、滑雪、单板滑雪、平台跳水、竞走、艺术体操以及蹦床,含有 10 000 条数据。实测数据集为时空数据检测数据集,此数据集来自 30 位志愿者的人体运动实测数据。数据可分为 6 类人体运动,共计 5 000 条数据。将两数据集中的数据融合在测试数据中,并将其随机划分为 10 组,每组中均需要包含此两数据集中数据。划分结果见表 1。

表 1 算例测试数据集
Table 1 Example test dataset

组别序号	数据量	组别序号	数据量
XC1	1 500	XC6	1 500
XC2	1 500	XC7	1 500
XC3	1 500	XC8	1 500
XC4	1 500	XC9	1 500
XC5	1 500	XC10	1 500

选取测试组 XC1 - XC5 作为训练组, XC6 - XC10 为测试组。使用所提算法对上述数据进行数据挖掘分析,用以验证所提算法的应用效果。

利用可穿戴式纳米生物传感器采集到的人体运动数据类型包括以下几个方面:

(1) 心率:通过测量心电图(ECG)信号,可以获得运动员的心率数据,以评估运动状态和身体健康状况,数据量大小为 15.61 GB。

(2) 步数和步频:通过加速度计和陀螺仪等传感器,可以获得运动员的步数和步频数据,以评估运动员的步态和步行姿势,数据量大小为 10.36 GB。

(3) 距离和速度:通过 GPS 和加速度计等传感器,可以获得运动员的距离和速度,以评估运动员的运动强度和速度,数据量大小为 7.89 GB。

(4) 姿势和运动轨迹:通过陀螺仪和加速度计

等传感器,可以获取运动员的姿势和运动轨迹数据,以评估运动员的姿态和动作是否正确,数据量大小为 8.96 GB。

(5) 肌肉活动和耗能:通过肌电图(EMG)传感器等技术,可以获取运动员的肌肉活动和能量消耗数据,以评估运动员的肌肉状态和运动强度,数据量大小为 5.96 GB。

(6) 血氧和血压:通过血氧和血压传感器等技术,可以获取运动员的血氧和血压数据,以评估运动员的身体状况和运动负荷,数据量大小为 11.34 GB。

2.2 实验指标

实验选取文献[11](算法 1)、文献[12](算法 2)、文献[13](算法 3)、文献[14](算法 4)以及文献[15](算法 5)等算法,与所提方法进行数据挖掘分析。根据预设的数据挖掘算法评定指标,确定各种算法的优劣性,保证所提算法的应用效果符合当前设计要求。实验评定指标大致可分为人体运动数据挖掘的召回率、人体运动数据挖掘正确率、人体运动数据采集的精度等 3 类,共计 3 个指标。通过对指标的融合分析,确定所提算法的计算性能。

人体运动数据挖掘的召回率计算公式:

$$D_1 = \frac{S_0}{S_{all}} \times 100\% \quad (11)$$

式中: S_{all} 表示应召回数据, S_0 表示实际召回数据。

人体运动数据挖掘正确率计算公式:

$$D_2 = \frac{S_1}{S_0} \times 100\% \quad (12)$$

式中: S_1 表示召回数据中数据挖掘处理正确的数据。

人体运动数据采集精度计算公式:

$$D_3 = \frac{S_1}{S_{all}} \times 100\% \quad (13)$$

2.3 实验结果

不同算法对于人体运动数据挖掘的召回率,比较结果见表 2。对表 2 中数据进行分析可以看出,针对数据挖掘过程中的数据召回率,所提算法与其他 5 种算法的召回率大致相同,即便存在差异也没有超过 4.0%。通过分析可以看出,所提算法的召回率平均值为 97.31%,分别比算法 1、算法 2、算法 3、算法 4 和算法 5 高出了 1.53%、1.15%、0.87%、1.21%和 1.18%。实验中所提算法召回率相对较高,整体较为稳定,数据分析处理能力相对较为稳定,具有一定的可行性。

对表 3 中数据进行分析可以看出,在数据挖掘正

准确率方面,所提算法与其他5种算法开始出现差异。其中,所提算法的数据挖掘正确率平均值为98.42%,分别比算法1、算法2、算法3、算法4和算法5高出了3.6%、3.12%、3.36%、3.59%和3.66%。所提算法的数据挖掘正确率明显高于其他算法,说明此方法具有较高的数据挖掘分析能力。其他5种算法挖掘正确率算法满足了人体运动数据的正确率要求,但整体偏低,还需对这些算法的此类性能进行优化。

表2 数据召回率对比结果

Table 2 Comparison results of data recall rate %

组别序号	所提算法	算法1	算法2	算法3	算法4	算法5
XC6	98.85	96.96	95.17	96.25	96.63	95.85
XC7	95.57	95.14	96.15	96.35	96.69	96.12
XC8	96.90	96.14	95.81	96.47	95.21	96.18
XC9	97.37	95.59	96.74	96.67	95.13	96.14
XC10	97.89	95.05	96.91	96.48	96.84	96.34
平均值	97.31	95.78	96.16	96.44	96.10	96.13

表3 数据正确率对比结果

Table 3 Comparison results of data accuracy %

组别序号	所提算法	算法1	算法2	算法3	算法4	算法5
XC6	98.23	95.91	95.99	94.21	94.12	95.02
XC7	98.71	94.99	95.75	95.47	94.44	95.33
XC8	98.12	94.18	94.78	94.66	95.24	94.27
XC9	98.27	94.57	94.45	95.68	95.55	94.66
XC10	98.75	94.42	95.55	95.27	94.82	94.51
平均值	98.42	94.82	95.30	95.06	94.83	94.76

对表4中数据进行整理分析可以看出,其他5种算法的人体运动数据采集精度明显不如所提算法。其中,所提算法的人体运动数据采集精度平均值为98.06%,分别比算法1、算法2、算法3、算法4和算法5高出了3.5%、4.12%、4.44%、4.06%和4.13%。由此可知,所提算法的人体运动数据采集精度优于其他算法。

表4 数据采集精度对比结果

Table 4 Comparison results of data acquisition accuracy %

组别序号	所提算法	算法1	算法2	算法3	算法4	算法5
XC6	98.50	94.72	93.13	94.17	94.06	93.25
XC7	98.00	94.06	93.66	93.05	94.28	93.04
XC8	97.98	94.05	93.07	93.27	93.92	94.92
XC9	97.68	94.91	95.32	94.28	93.86	94.09
XC10	98.15	95.04	94.54	93.35	93.87	94.36
平均值	98.06	94.56	93.94	93.62	94.00	93.93

3 结束语

针对当前体育数据分析研究过程中出现的问题,应用可穿戴式纳米生物传感器获取原始数据并设定新型数据挖掘算法。经过算例分析证实此算法具有较高的应用价值。该算法可以应用在运动员身体状况评估、运动损伤和疲劳监测、运动姿态和动作评估、个性化训练和康复方案制定、科学决策和管理等方面,

以提高运动员的竞技水平和身体健康状况,优化训练和竞赛策略,实现个性化训练和康复方案,以及为科学决策和管理提供依据。虽然研究取得了一定成果但此次设计仅针对实时数据采集问题进行优化,未对数据挖掘过程中的其他问题进行完善。未来研究中,还需对体育数据分析工作中的其他环节进行优化,以保证此算法的应用效果符合数据分析要求。

参考文献

- [1] 董英辉. 基于数据挖掘的体育运动即时数据自动化采集系统设计[J]. 自动化与仪器仪表, 2022(10): 155-160.
- [2] 李佳佳, 丁伟, 王伯伟, 等. 基于随机森林的民俗体育对身体指标影响评估方法[J]. 南京大学学报(自然科学), 2021, 57(1): 59-67.
- [3] 孙宇航, 周建钦, 张学锋. 基于加速度传感器的人体运动模式识别[J]. 计算机系统应用, 2020, 29(6): 196-203.
- [4] 程翔, 郭端, 伍晓明, 等. 可穿戴式足底压力监测系统的应用及进展[J]. 机器人外科学杂志(中英文), 2024, 5(2): 143-147.
- [5] 叶彩仙, 胥立军. 基于多可穿戴传感器的用户人体复杂行为主动识别[J]. 传感器世界, 2024, 30(2): 28-33.
- [6] 孟令飞, 章碧玉, 张学兵, 等. 可穿戴式汗液传感器在运动监测中的研究进展[J]. 传感器与微系统, 2024, 43(2): 1-6.
- [7] 孙希. 基于STM32的智能穿戴式人体生理参数监测仪[J]. 仪表技术, 2024(1): 26-29.
- [8] 黎王安, 张朝晖, 赵小燕, 等. 基于体域网的可穿戴式短跑运动状态监测[J]. 物联网技术, 2024, 14(1): 4-6.
- [9] 林淑洁, 张立秀, 陈香萍, 等. 可穿戴设备在压力性损伤预防中的研究进展[J]. 军事护理, 2023, 40(12): 93-95.
- [10] 徐自远, 姜宇. 基于无线网络的穿戴式人体健康监测系统[J]. 自动化技术与应用, 2023, 42(9): 58-62.
- [11] ZONG X, ZHANG C, WU D. Research on data mining of sports wearable intelligent devices based on big data analysis[J]. Discrete Dynamics in Nature and Society, 2022, 69(1): 1-9.
- [12] MOHAMED M E, BOUIKHALENE B, SAFI S. Automatic CV processing for scientific research using data mining algorithm[J]. Journal of King Saud University - Computer and Information Sciences, 2020, 32(5): 561-567.
- [13] LIU Meiling, ZHANG Beixian, LI Xi, et al. An optimized k-means algorithm based on information entropy[J]. Computer Journal, 2021, 64(7): 1130-1143.
- [14] LI Hua, HU Bo, LIU Yubo, et al. Classification of electricity consumption behavior based on improved k-means and LSTM[J]. Applied Sciences, 2021, 11(16): 7625-7625.
- [15] WANG Guangtong, MIAO Jianchun. Design of data mining algorithm based on rough entropy for us stock market abnormality[J]. Journal of Intelligent & Fuzzy Systems, 2020, 39(4): 1-9.
- [16] 陈达理, 刘雪红. 基于自适应样本熵的穿戴式传感器人体活动识别[J]. 武汉理工大学学报, 2022, 44(8): 91-96.
- [17] 徐甲栋, 陈强, 徐一雄, 等. 基于MEMS传感器的实时跌倒检测系统设计[J]. 传感器与微系统, 2022, 41(7): 77-80.
- [18] CHEN Z S, ZHANG X, PEDRYCZ W, et al. K-means clustering for the aggregation of HFLTS possibility distributions: N-two-stage algorithmic paradigm[J]. Knowledge - Based Systems, 2021, 227: 107230.