

# 基于模糊聚类的多传感器数据融合算法优化

谢宇威, 林传峰

(浙江大学 信息技术中心, 杭州 310027)

**摘要:** 针对受传感器本身误差和外界干扰的影响, 不同传感器获取的数据可能存在不确定性和不一致性的问题, 为有效消除数据之间的矛盾和冲突, 提升数据融合效果, 提出一种基于模糊聚类的多传感器数据融合算法. 首先, 采用 D-S(Dempster-Shafer)证据理论进行数据初步融合, 计算异类数据之间的距离并确定对应的信任函数, 对不同传感器数据进行校正和协调, 以提高数据的一致性. 其次, 引入模糊聚类方法对多传感器数据初步融合结果进行优化, 将数据点分组为具有相似特征的簇, 确定初始聚类中心. 最后, 利用模糊聚类算法对数据分组, 提高数据融合结果的准确性和鲁棒性. 实验结果表明, 该算法在多传感器数据融合中拟合优度和分片接收率较高, 且能量总体消耗较低, 整体性能优异.

**关键词:** 模糊聚类; 多传感器; 数据融合; D-S 证据理论

**中图分类号:** TP393 **文献标志码:** A **文章编号:** 1671-5489(2025)05-1462-06

## Optimization of Multi-sensor Data Fusion Algorithm Based on Fuzzy Clustering

XIE Yuwei, LIN Chuanfeng

(Information Technology Center, Zhejiang University, Hangzhou 310027, China)

**Abstract:** Aiming at the problem that the data obtained by different sensors might be uncertain and inconsistent due to the influence of sensor's own error and external interference, in order to effectively eliminate the contradictions and conflicts between data and improve the data fusion effect, we proposed a multi-sensor data fusion algorithm based on fuzzy clustering. Firstly, the D-S (Dempster-Shafer) evidence theory was used for preliminary data fusion, the distance between heterogeneous data was calculated and the corresponding trust function was determined, and different sensor data were corrected and coordinated to improve data consistency. Secondly, we introduced fuzzy clustering method to optimize the preliminary fusion results of multi-sensor data, grouped data points into clusters with similar features, and determined the initial clustering center. Finally, we used fuzzy clustering algorithm to group data and improve the accuracy and robustness of data fusion results. The experimental results show that the proposed algorithm has high fitting goodness and shard reception rate in multi-sensor data fusion, and overall energy consumption is low, with excellent overall performance.

**Keywords:** fuzzy clustering; multi-sensor; data fusion; D-S evidence theory

收稿日期: 2024-11-04.

第一作者简介: 谢宇威(1988—), 男, 汉族, 硕士, 工程师, 从事信息化建设、教育信息化、信息化服务和人工智能的研究.

E-mail: xieyuwei@zju.edu.cn.

基金项目: 国家自然科学基金(批准号: 61561018).

不同传感器因其物理原理、工作环境及性能差异, 所采集的数据通常存在冗余、矛盾甚至冲突, 从而影响数据的有效利用及准确解析. 为克服这些不足, 提升数据处理的精度和效率, 多传感器数据融合<sup>[1-2]</sup>算法应运而生. 多传感器数据融合旨在通过集成来自不同传感器的信息, 实现数据的互补和增强<sup>[3-5]</sup>, 进而形成对监测对象或环境更全面、准确和可靠的理解, 该算法能有效消除数据间的冗余和冲突, 目前在智能监控和医疗诊断等领域<sup>[6-7]</sup>已成为实现高精度感知和决策的关键技术之一.

目前, 对多传感器数据融合的研究已取得了许多成果. 例如, 梁海波等<sup>[8]</sup>构建了加权正则化极限学习机(weighted regularized extreme learning machine, WRELM)数据融合模型, 并引入鸟群算法实现了数据融合. 但该方法未充分考虑传感器节点受老化及干扰的影响, 数据采集偏差较大, 影响数据融合的精确性. 刘康等<sup>[9]</sup>利用四分位离散度方法剔除异常数据后, 采用测度算子和自适应融合算法为各传感器分配权重, 利用综合权重进行加权处理, 得到融合结果. 但由于传感器之间相互影响, 且存在数据冗余问题, 导致数据融合结果不理想. 梅武军等<sup>[10]</sup>通过 K-means 聚类算法处理异步问题, 设计融合滑动窗口机制, 确保数据流的动态变化始终保持在预设的容差阈值内, 以此实现传感器信息融合. 但如果初始聚类中心选择不当, 则会导致算法收敛到局部最优解, 影响融合结果的准确性和稳定性. Liu 等<sup>[11]</sup>从不同来源生成多个测量值, 通过线性和非线性的方式测量潜在变量, 同时捕获测量值中包含的公共信息, 通过考虑公共信息与潜在变量之间的非线性映射关系, 实现数据融合. 在实际应用中, 数据通常具有实时性和动态性, 该方法需能实时处理不断变化的测量值, 如果无法及时响应数据的变化, 则可能会导致数据融合结果的滞后或失效.

基于上述研究, 本文提出一种基于模糊聚类的多传感器数据融合算法优化方法, 旨在通过改进和优化现有的数据融合算法, 解决多传感器数据在融合过程中面临的实时性和准确性等挑战, 提高数据处理效果.

# 1 多传感器数据融合算法优化

## 1.1 多传感器数据初步融合

由于传感器本身的误差和外界干扰, 因此不同传感器获取的数据可能存在不确定性. 采用 D-S (Dempster-Shafer) 证据理论进行数据初步融合可对这些数据进行校正和协调. D-S 证据理论在应对不确定性问题上具有显著优势, 但在证据间冲突激烈时可能引发悖论. 基于此, 有效处理悖论成为该问题的研究重点. 同时, 传感器节点受老化及干扰影响, 采集数据偏差大, 甚至可能出现零值, 导致在基本概率赋值过程中 mass 函数值偏低或为零. 基于上述分析, 需聚焦于两大优化方向: 一是有效管理融合过程中的信息冲突; 二是简化 mass 函数的计算流程, 以降低复杂度, 从而确保数据融合结果的精确性.

先对各传感器节点收集的数据进行整理与归并, 并将异常值剔除; 然后对异类数据进行融合决策处理, 最终得出多传感器数据的初步融合结果. 多传感器数据融合过程如图 1 所示.

通过欧氏距离有效识别并移除传感器数据中的异常值. 基本思路是: 同类数据间距离越小, 相似度越高. 获取两个数据间距离和的均值, 并设定对应的阈值, 将异常值删除.

设传感器的数量为  $n$ , 采集到的同类数据表示为  $U = \{u_1, u_2, \dots, u_n\}$ , 其中  $u_i$  表示第  $i$  类数据,  $i = 1, 2, \dots, n$ ,  $u_i$  和  $u_j$  之间的距离  $d(u_i, u_j)$  表示为

$$d(u_i, u_j) = \left[ \frac{1}{n-1} \sum_{i,j=1}^n (u_i - u_j)^2 \right]^{1/2}. \tag{1}$$

利用式(1)可计算出  $d(u_i, u_j)$ , 将其组成距离矩阵为

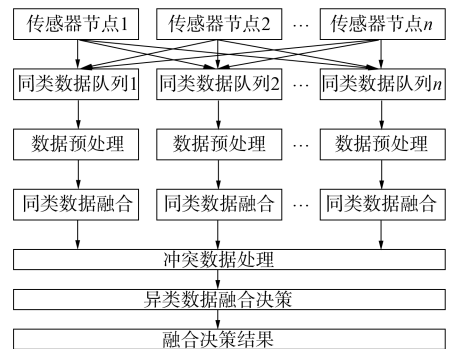


图 1 多传感器数据融合过程

Fig. 1 Multi-sensor data fusion process

$$D = \begin{pmatrix} d_{11} & \cdots & d_{1j} & \cdots & d_{1n} \\ \vdots & & \vdots & & \vdots \\ d_{i1} & \cdots & d_{ij} & \cdots & d_{in} \\ \vdots & & \vdots & & \vdots \\ d_{m1} & \cdots & d_{mj} & \cdots & d_{mn} \end{pmatrix}, \tag{2}$$

其中  $D$  主要用来描述与其他数据之间的相似程度,  $D$  中的元素越小相似程度越高. 相似程度对应的计算公式为

$$G(D) = \sum_{i=1}^n \sum_{j=1}^n d_{ij}. \tag{3}$$

对  $G(D)$  进行归一化处理, 强化同类数据间距离与相似度的正相关性, 即距离越近则相似度越高. 完成归一化相似度计算后, 可直接根据此结果构建同类数据的信任函数  $I_{DSUP(u_i)}$  为

$$I_{DSUP(u_i)} = 1 - G(D). \tag{4}$$

同类数据的信任函数值  $I_{DSUP(u_i)}$  主要用于反映数据的真实性,  $I_{DSUP(u_i)}$  值越大则数据越可靠. 在同类数据中, 异常值会导致信任函数值偏低, 且异常值间的  $I_{DSUP(u_i)}$  趋于接近; 相反, 正常值则具有较高的信任值, 且分布相对集中. 因此, 设定阈值  $\theta$ , 当  $I_{DSUP(m)} - \theta \leq I_{DSUP(u_i)} \leq I_{DSUP(m)} + \theta$  时, 则保留  $I_{DSUP(u_i)}$  值并且剔除异常值, 进而得到正确的同类数据, 其中  $I_{DSUP(m)}$  为  $I_{DSUP(u_i)}$  的中位数.

将筛选得到的数据进行均值计算, 得到数据  $\bar{U}_i$ , 用

$$D(\bar{U}_i) = \sum_{i=1}^n \left[ \frac{1}{S_{i\max} - S_{i\min}} \left( \bar{U}_i - \frac{\Delta S_i}{2} \right)^2 \right]^{1/2} \tag{5}$$

计算各类数据  $\bar{U}_i$  到每个相应等级  $S_i$  之间的距离,  $D(\bar{U}_i)$  取值越小, 则说明越靠近该等级. 其中  $S_{i\max}$  和  $S_{i\min}$  分别表示不同类数据等级特征值的最大值和最小值,  $\Delta S_i/2$  表示每类数据每个等级区间的中间值. 基于式(5), 可得异类数据间的信任函数  $I_{MSUP(u_i)}$  为

$$I_{MSUP(u_i)} = 1/D(\bar{U}_i). \tag{6}$$

对式(6)获取的信任函数  $I_{MSUP(u_i)}$  进行归一化处理, 并利用 D-S 证据理论进行融合, 则可得多传感器数据初步融合结果为

$$m(B_i) = \frac{I_{MSUP(u_i)}}{\sum_{i=1}^n I_{MSUP(u_i)}}. \tag{7}$$

### 1.2 多传感器数据融合优化

在多传感器数据融合过程中, 初步融合了来自不同传感器的信息, 提高了数据的一致性和可靠性. 但初步融合可能仍存在噪声和不完整的数据, 导致融合结果不精确. 因此, 引入模糊聚类方法对多传感器数据融合算法进行优化<sup>[12-13]</sup>. 模糊聚类方法通过将数据点分组为具有相似特征的簇, 能进一步优化数据融合效果, 提高融合结果的准确性和鲁棒性. 操作步骤如下.

1) 将多传感器数据集经过主成分分析处理, 计算得到数据点间的相似性矩阵  $H$ , 用于捕捉数据的主要特征及其相互关系, 计算公式为

$$H = (h_{ij})_{n \times n}, \tag{8}$$

其中  $h_{ij}$  表示数据点  $i$  和  $j$  之间的相似度.

2) 对密度半径  $r_c$  进行初始化处理, 按下式计算数据点密度  $\beta$  和聚类半径  $\phi$ :

$$\begin{cases} \beta = \sum_j e^{-(h_{ij}/h_c)}, \\ \phi = \min_{1 \leq j \leq n} \{h_{ij}\} \times r_c, \end{cases} \tag{9}$$

其中  $e$  为常数,  $h_c$  表示截断距离. 根据  $\beta$  和  $\phi$  计算结果, 删除偏差较大的奇异点, 将剩余点作为初始聚类中心.

3) 设定聚类半径  $\phi$  后, 用迭代爬山策略最大化 Fitness 值, 从而确定最优解. 其基本思想是从

一个初始解开始, 通过不断尝试改进解的邻域寻找更优的解. 在每一步中, 算法都会评估当前解的邻域, 并选择使目标函数(Fitness 函数)值最大的解作为下一步的起点. Fitness 对应的计算式为

$$\text{Fitness} = \frac{\text{Fitness}_2}{\text{Fitness}_1}, \tag{10}$$

其中  $\text{Fitness}_1$  和  $\text{Fitness}_2$  分别表示不同轮数的 Fitness 值.

4) 在聚类前期, 对参数进行初始化处理.

5) 计算聚类中心和权重向量.

① 聚类中心的计算. 设定隶属度  $z_{ik}$  和权重向量  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$  保持不变, 在确保损失达到最小的情况下, 对聚类中心  $p_i = \{p_{i1}, p_{i2}, \dots, p_{in}\}$  进行更新:

$$p_{ij} = \frac{\sum_{k=1}^n z_{ik} x_{kj}}{\sum_{k=1}^n z_{ik}}, \tag{11}$$

其中  $x_{kj}$  表示数据点.

② 按下式计算权重向量  $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ :

$$\omega = \left[ \sum_{i=1}^n \sum_{j=1}^c z_{ik} (x_{kj} - p_{ij})^2 \right]^{1/2}. \tag{12}$$

6) 按下式计算隶属度矩阵:

$$z_{ik} = \frac{1}{\sum_{j=1}^c \left[ \sum_{i=1}^c \omega (x_{kj} - p_{ij})^2 \right]^2}. \tag{13}$$

7) 判断  $z_{ik}$  的变化量是否低于  $\alpha$ , 如果是, 则直接终止操作; 反之, 则跳转至步骤 3).

8) 在评估聚类有效性时, 若某一类别包含的数据点数量少于总数据点的一定比例, 则认为该类别无效, 并视其内部所有点为潜在噪声点.

9) 剔除潜在噪声点后, 确定正常类别的类中心作为数据融合输出, 实现多传感器数据融合, 达成数据融合优化的目标.

## 2 实验及结果分析

### 2.1 实验数据采集

在进行实验验证时, 选择的应用场景是自动驾驶车辆的环境感知, 选择激光雷达(LiDAR)、毫米波雷达、摄像头作为主要的传感器类型. 样本数据采集过程如下: 首先, 确保所有传感器都已正确安装并校准到车辆上, 以便它们能准确捕捉周围环境的信息. 激光雷达(LiDAR)通过发射激光束并测量其反射回的时间构建周围环境的 3D 点云图, 其扫描频率为 10 Hz, 扫描角度为 360°水平扫描, 测距范围为 150 m. 毫米波雷达则主要用于检测车辆前方的障碍物, 其探测距离为 250 m, 探测角度为水平 ±45°. 摄像头则负责捕捉车辆周围的图像信息, 选择多个摄像头以覆盖不同的视野范围, 包括前向、侧向和后向.

在实验数据采集过程中, 利用 GPS 时间戳确保所有传感器都能同步工作, 以便后续的数据融合处理. 同时, 记录每个传感器的数据输出格式和采样时间戳, 以便在后续数据处理中能准确地将它们对应. 将上述采集到的数据划分为 4 个数据集, 每个数据集中均包含 2 500 条数据, 以此进行实验测试.

### 2.2 实验结果分析

将加权正则化极限学习机和 K-means 聚类算法作为对比方法与本文算法进行对比分析. 选取拟合优度作为测试指标进行实验分析, 其值越大说明算法对数据的拟合程度越高, 能更准确地反映数据的真实情况, 即多传感器数据融合性能越优, 实验测试结果如图 2 所示. 由图 2 可见, 在 3 种测试方法中, 利用本文算法对多传感器数据进行融合处理, 可显著提升拟合优度, 说明本文算法获取的融合结果更接近真实情况, 具有较好的融合性能, 整体融合精度较高. 本文算法利用 D-S 证据理论进行初步的数据融合, 通过计算异类数据之间的距离和确定信任函数初步校正和协调数据, 有效减少了数据间的不一致性和冲突. 这一步骤为后续的模糊聚类处理提供了更准确和一致的数据基础.

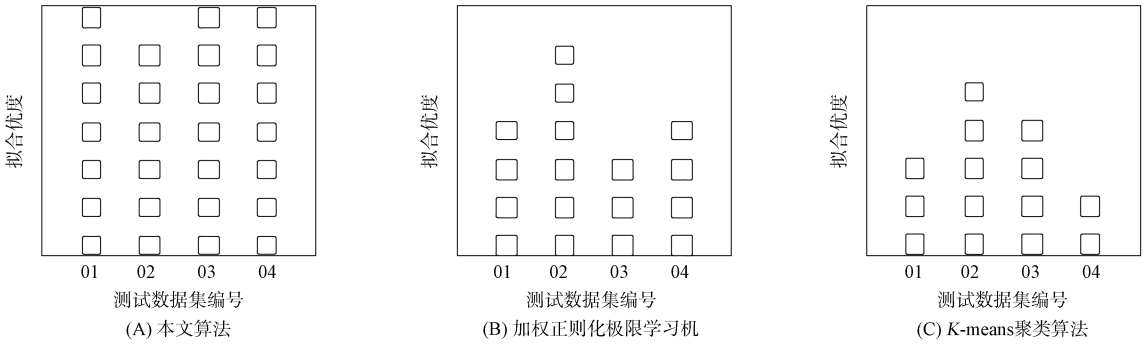


图 2 不同算法的拟合优度测试结果比较

Fig. 2 Comparison of goodness of fit test results for different algorithms

为进一步验证本文算法的优越性, 将分片接收率  $f_r$  作为测试指标. 分片接收率表示在数据融合过程中, 成功接收到的数据分片的比例. 如果数据融合方法能高效地整合来自不同源的数据, 并确保这些数据分片在传输过程中保持稳定性和完整性, 则分片接收率将会较高, 从而验证了数据融合方法的有效性和可靠性.  $f_r$  计算公式为

$$f_r = \frac{\text{节点真实接收分片数量}}{\text{节点实际发送分片数量}} \quad (14)$$

图 3 为不同分片数量下不同方法分片接收率的测试结果. 由图 3 可见, 当采用不同方法进行融合处理后, 随着分片量的持续增加, 各方法的分片接收率也呈上升趋势. 对比分析表明, 本文算法的分片接收率最高, 说明其可以捕获更多的传感器数据, 有效确保数据的完整性和丰富性, 减少数据在传输过程中因干扰造成的丢包情况, 整体融合性能得到显著优化. 本文算法引入模糊聚类方法对初步融合后的数据进行进一步优化, 模糊聚类不仅能将数据点分组为具有相似特征的簇, 还能通过模糊隶属度表达数据点属于各簇的不确定性, 从而在处理含有噪声和不确定性的多传感器数据时具有更高的灵活性和鲁棒性, 提高融合结果的准确性和可靠性.

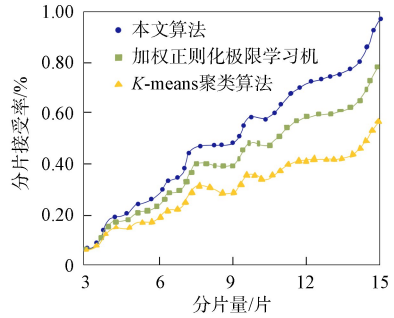


图 3 不同分片数量下不同方法分片接收率的测试结果

Fig. 3 Test results of shard reception rates of different methods under different shard quantities

表 1 列出了不同测试时间下不同方法的能量总体消耗测试结果. 由表 1 可见, 在 3 种测试方法中, 本文算法的能量总体消耗明显低于另外两种算法, 说明利用本文算法对多传感器数据进行融合处理时, 可显著降低能量总体消耗, 能高效地整合来自多个传感器的数据, 减少了冗余计算和通信开销.

表 1 不同测试时间下不同方法的能量总体消耗测试结果

Table 1 Overall energy consumption test results of different methods at different testing time

时间/s	能量总体消耗/J		
	本文算法	加权正则化极限学习机	K-means 聚类算法
200	8.56	10.34	12.29
400	10.11	12.77	13.78
600	11.75	14.85	15.66
800	12.34	16.59	17.30
1 000	13.80	17.11	18.69
1 200	14.92	21.28	22.97
1 400	16.03	23.62	25.02
1 600	17.05	25.27	27.36
1 800	18.99	28.58	29.40
2 000	21.34	30.46	31.89

综上所述,为显著提升数据融合结果的准确性,本文结合模糊聚类算法提出了一种多传感器数据融合算法.实验结果表明,该算法可显著提升拟合优度和分片接收率,降低能量总体消耗,获取更精准且满意的多传感器数据融合结果,保证其融合性能得到优化和完善,全面增强网络的整体性能.

### 参 考 文 献

- [1] 孙雷,吴庆宪,王玉惠,等.基于改进D-S证据理论的多源空战信息博弈融合方法[J].信息与控制,2022,51(5):566-572.(SUN L, WU Q X, WANG Y H, et al. Air Combat Information Game Fusion Method Based on Improved D-S Evidence Theory [J]. Information and Control, 2022, 51(5): 566-572.)
- [2] 胡楚阳,柳先辉,赵卫东.基于规则链的网络协同制造数据融合方法研究[J].计算机科学,2022,49(增刊2):219-225.(HU C Y, LIU X H, ZHAO W D. Data Fusion Method of Network Collaborative Manufacturing Based on Rule Chain [J]. Computer Science, 2022, 49(Suppl 2): 219-225.)
- [3] 付亚平,闻姐,王巧云.基于FFT-COSS的光纤F-P传感器解调算法[J].东北大学学报(自然科学版),2022,43(11):1529-1535.(FU Y P, WEN D, WANG Q Y. A Demodulation Algorithm for Fiber-Optic F-P Sensor Based on Fast Fourier Transform and Cosine Similarity [J]. Journal of Northeastern University (Natural Science), 2022, 43(11): 1529-1535.)
- [4] 李忠利,马理想,韩冲,等.多传感器融合的割草机器人障碍物检测方法[J].江苏大学学报(自然科学版),2024,45(2):160-166.(LI Z L, MA L X, HAN C, et al. Obstacle Detection Method of Lawn Mowing Robot Based on Multi-sensor Fusion [J]. Journal of Jiangsu University (Natural Science Edition), 2024, 45(2): 160-166.)
- [5] 侯鹏帅,周大鹏,杨大鹏,等.基于多传感器融合的无人舰载机舰面定位研究[J].东北大学学报(自然科学版),2024,45(1):1-9.(HOU P S, ZHOU D P, YANG D P, et al. Research on Positioning of Unmanned Carrier-Based Aircraft Based on Multi-sensor Fusion in Ship's Surface Environment [J]. Journal of Northeastern University (Natural Science), 2024, 45(1): 1-9.)
- [6] 卢艳军,张前,张晓东,等.基于光流传感器的位置信息融合方法[J].探测与控制学报,2022,44(3):51-56.(LU Y J, ZHANG Q, ZHANG X D, et al. Location Information Fusion Method Based on Extended Kalman Filter [J]. Journal of Detection & Control, 2022, 44(3): 51-56.)
- [7] 谢李祥,邢传玺,张东玉,等.浅海多节点环境参数测量数据融合方法研究[J].声学技术,2022,41(2):274-281.(XIE L X, XING C X, ZHANG D Y, et al. Research on Data Fusion Method of Multi-node Environmental Parameter Measurement in Shallow Sea [J]. Technical Acoustics, 2022, 41(2): 274-281.)
- [8] 梁海波,成刚,张志东,等.多传感器检测管道缺陷数据融合方法[J].激光与光电子学进展,2023,60(4):316-324.(LIANG H B, CHENG G, ZHANG Z D, et al. Data Fusion Method for Multi-sensor Detection of Pipeline Defects [J]. Laser & Optoelectronics Progress, 2023, 60(4): 316-324.)
- [9] 刘康,何明浩,韩俊,等.基于多传感器的雷达对抗侦察数据融合算法[J].系统工程与电子技术,2023,45(1):101-107.(LIU K, HE M H, HAN J, et al. Data Fusion Algorithm for Radar Countermeasures and Reconnaissance Based on Multi-sensor [J]. Systems Engineering and Electronics, 2023, 45(1): 101-107.)
- [10] 梅武军,郑军,金杰,等.基于滑动聚类的多传感器异步信息融合方法[J].仪器仪表学报,2022,43(6):109-117.(MEI W J, ZHENG J, JIN J, et al. A Sliding-Clustering-Based Method for Multi-sensor Asynchronous Information Fusion [J]. Chinese Journal of Scientific Instrument, 2022, 43(6): 109-117.)
- [11] LIU Y W, SUN X X, ZHONG W X, et al. B-Scaling: A Novel Nonparametric Data Fusion Method [J]. The Annals of Applied Statistics, 2022, 16(3): 1292-1312.
- [12] 杜秀丽,姜晓虎,孙晨瞳,等.基于方向性多重假设检验和信息熵的函数型数据聚类新方法[J].南京师大学报(自然科学版),2022,45(4):1-9.(DU X L, JIANG X H, SUN C T, et al. A New Functional Data Clustering Method Based on Directional Multiple Hypothesis Test and Information Entropy [J]. Journal of Nanjing Normal University (Natural Science Edition), 2022, 45(4): 1-9.)
- [13] 刘丽华.档案管理中文本数据的增量多模态聚类方法[J].重庆大学学报,2022,45(5):147-156.(LIU L H. Incremental Multi-modal Clustering Methods for Text Data in Archives Administration [J]. Journal of Chongqing University, 2022, 45(5): 147-156.)