

基于密度的孤立点分析算法融合下的异常心理沙盘检测

王泽贤¹, 牛小飞¹, 蒋志方^{2,3}, 赵迅冕¹, 王耀辉¹, 宋泓远¹

(1. 山东建筑大学 计算机科学与技术学院, 山东 济南 250101; 2. 山东大学 软件学院, 山东 济南 250100;
3. 聆心云(山东)智能科技有限公司, 山东 济南 250098)

摘要: 全球精神疾病患者数量庞大, 精神健康问题已成为全球性的公共卫生挑战, 及时发现并提前干预治疗对患者恢复心理健康具有重要现实意义。沙盘游戏是一种常用的心理问题识别和治疗工具, 仅仅通过心理专家对反映不健康心理的异常沙盘进行检测, 费时费力, 且效率低下, 利用数据挖掘技术自动识别异常心理沙盘迫在眉睫。提出融合基于密度的孤立点分析算法以检测异常心理沙盘, 对沙盘数据进行预处理, 提取区分正常沙盘和异常沙盘的特征, 将DBSCAN算法进行改进并与LOF算法进行融合以提高异常心理沙盘检测性能。在“聆心云数字化心理健康服务平台”提供的数据集上的实验表明, 融合基于密度的孤立点分析算法能够识别出更多反映不健康心理的异常沙盘。

关键词: 心理沙盘; 异常检测; DBSCAN算法; LOF算法

DOI: 10.11907/rjdk.241733

扫描二维码阅读全文:

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2025)010-0127-07



Abnormal Psychological Sand Tables Detection Based on Density-based Outlier Analysis Algorithm Fusion

WANG Zexian¹, NIU Xiaofei¹, JIANG Zhifang^{2,3}, ZHAO Xunmian¹, WANG Yaohui¹, SONG Hongyuan¹

(1. College of Computer Science and Technology, Shandong Jianzhu University, Ji'nan 250101, China; 2. School of Software, Shandong University, Ji'nan 250100, China; 3. Lingxinyun(Shandong) Intelligent Technology Co., Ltd, Ji'nan 250098, China)

Abstract: The number of people suffering from mental illness all over the world is huge, mental health problems have become a global public health challenge, timely detection and early intervention treatment is of great practical importance for patients to restore their mental health. Sandplay is a commonly used tool for identifying and treating psychological problems. It is time-consuming and inefficient to detect abnormal sand tables reflecting unhealthy psychology only by psychologists. It is imminent to automatically identify abnormal psychological sand tables using data mining techniques. Fusing density based outlier analysis algorithms is proposed to detect abnormal psychological sand tables. The sand tables data is preprocessed. The features that distinguish normal sand tables from abnormal ones are extracted. The fused algorithm improves the DBSCAN algorithm and integrates it with the LOF algorithm to improve the performance of detecting abnormal psychological sand tables. Experiments were conducted on the dataset provided by the "LingXinYun Digital Mental Health Service Platform". Experimental results show that the fusing density based outlier analysis algorithm identifies more abnormal sand tables reflecting unhealthy psychology.

Key Words: psychological sand table; abnormal detection; DBSCAN algorithm; LOF algorithm

收稿日期: 2024-09-25

基金项目: 山东省自然科学基金项目(ZR2021MF006)

作者简介: 王泽贤(2000-), 男, 山东建筑大学计算机科学与技术学院硕士研究生, 研究方向为数据挖掘、心理健康辅助诊断; 牛小飞(1979-), 女, 博士, CCF会员, 山东建筑大学计算机科学与技术学院教授、硕士生导师, 英国Ulster大学兼职博士生导师, 研究方向为数据挖掘、机器学习、心理健康辅助诊断; 蒋志方(1961-), 男, 博士, CCF会员, 山东大学软件学院副教授、硕士生导师, 研究方向为人机交互与虚拟现实、数据挖掘可视分析、数字化心理健康辅助诊疗等; 赵迅冕(2000-), 男, 山东建筑大学计算机科学与技术学院硕士研究生, 研究方向为数据挖掘、网页作弊检测、心理健康辅助诊断; 王耀辉(2000-), 男, 山东建筑大学计算机科学与技术学院硕士研究生, 研究方向为数据挖掘、心理健康辅助诊断; 宋泓远(2001-), 男, 山东建筑大学计算机科学与技术学院硕士研究生, 研究方向为数据挖掘、心理健康辅助诊断。本文通讯作者: 牛小飞。

0 引言

根据世界卫生组织的调查数据,全球约有十亿人正受到精神疾病的困扰。《2023 年度中国精神心理健康》蓝皮书显示,我国成年抑郁风险检出率达 10.6%,且学生群体心理健康问题愈发凸显,其中高中生抑郁检出率超过 40%,初中生抑郁检出率超过 30%,小学生抑郁检出率也超过了 10%^[1]。这些数据表明,抑郁症发病群体呈现年轻化趋势,社会需要更加关注青少年心理健康问题。

目前,在我国社会及学校领域,筛查心理不健康人员所采用的主要方法是借助心理测评量表,如心理健康量表 SCL-90、抑郁自评量表 SDS 等。这类量表是为了评定个体在感觉、情绪、人际关系、饮食睡眠等方面的心理健康症状而设计,包含若干问题,如感到胸闷、不能控制地大发脾气、难以入睡、轻生等,测评者针对每个问题从没有、很轻、中等、偏重、严重 5 个选项中选 1 个。这种方法简单方便,但缺点是有些测评者为了掩盖自己不健康的心理状态,不如实回答问题,导致无法被筛选出。

沙盘游戏是一种常用的心理问题识别和治疗工具,在心理健康教育方面有重要作用^[2-5]。它能够帮助沙游者表达内心世界、探索情感和解决问题。图 1 为两个沙游者所做的沙盘,反映了他们不同的心理状态。心理专家可以从沙游者使用的沙具、建造沙盘具体过程、沙具与沙具之间的距离摆位等多个角度进行全面分析,更好地理解沙游者的内心状况,从而有效地识别心理不健康的沙游者。

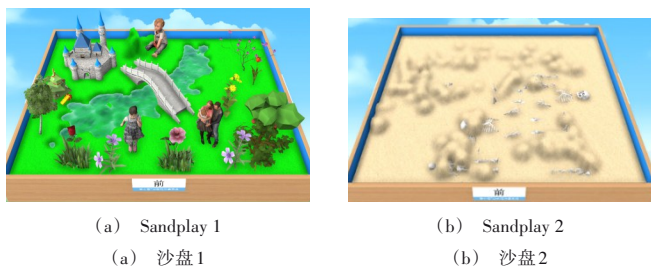


Fig. 1 Example of sandplay

图 1 沙盘游戏示例

通过心理专家对心理沙盘进行识别,虽然能够较为准确地检测出反映不健康心理的沙盘,但是费时费力,且效率低下。数据挖掘技术中的孤立点分析或异常检测算法可以从大量复杂的数据集中检测出与正常数据行为或特征属性差别较大的异常数据模式^[6]。通过孤立点分析或异常检测算法自动检测心理沙盘可以极大减少检测时间和人力成本。

常见的异常检测算法有:基于距离的异常检测算法、基于聚类的异常检测算法、基于密度的异常检测算法等^[7]。基于距离的异常检测算法主要依赖于计算数据点之间的距离以识别异常点,其优点在于不需要假设数据的分布状态,简单易懂,适用于多种数据类型,但其在处理高

维海量数据时效率较低。基于聚类的异常检测算法通过将数据集划分为多个簇以识别异常点,其优势在于无需标签和先验知识,便能够有效处理不同类型的数据,但聚类算法的效果受特定聚类方法的影响较大,对超参数的选择也较为敏感。基于密度的异常检测算法认为正常样本的邻域密度与其周围样本的邻域密度相似,而异常点的邻域密度明显低于周围样本。该类算法能够有效检测局部异常点,并量化数据点的异常程度,但该类算法在不同数据分布情况下的适用性会受到限制。

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) 是一种基于密度的聚类算法,广泛用于异常检测^[8]。DBSCAN 算法不需要预先指定簇的数量,在处理数据结构时非常灵活,且能够识别任意形状的簇,适用于复杂的数据分布,且 DBSCAN 算法对噪声和异常值具有鲁棒性。LOF (Local Outlier Factor) 算法专注于识别局部异常,在处理数据密度不均匀的场景中表现出色,且 LOF 为每个数据点分配一个异常因子,提供了异常程度的量化指标^[9]。此外,LOF 算法不需要对数据进行假设或预处理,能够处理高维数据,且适应性强,适合于复杂的实际应用场景。

本文研究将孤立点定义为能反映不健康心理的沙盘,即异常心理沙盘,提出融合基于密度的孤立点分析算法检测异常心理沙盘,通过对 DBSCAN 算法进行改进,并与 LOF 算法结合,改进了异常检测算法的局部异常检测能力和全局密度识别能力。实验结果表明,融合算法可以更有效地检测出异常心理沙盘。

1 相关工作

在基于距离的异常检测算法方面,Ying 等^[10]提出了一种基于改进的 K-Nearest Neighbor (KNN) 算法和自动标记样本集的高效日志异常检测方法,使用平均加权技术对 KNN 算法进行了改进,实现了在不平衡样本上的异常检测;Shuai 等^[11]提出了一种局部密度 KNN 算法,减少了特征空间中的局部密度偏差,提高了异常检测效果。

在基于聚类的异常检测算法方面,Zhang 等^[12]提出了一种基于快速密度峰值聚类异常因子的异常检测算法,采用密度和距离乘积的方法自动选择聚类中心进行异常检测,提高了异常检测算法的有效性和时间效率。Chen 等^[13]提出了一种大规模数据聚类算法,提高了计算的效率,为在大规模数据集上聚类进行异常检测提供了前提。

在基于密度的异常检测算法方面,Li 等^[14]提出了一种基于密度距离决策图的局部和聚类异常值检测,将密度比和密度提升距离相结合,绘制密度距离决策图,并计算两个指标的乘积作为最终异常值得分,有效提高了异常检测效果。Hu 等^[15]提出了一种基于局部核密度估计和上下文回归的异常检测算法,设计了一种局部核回归估计器和一

种分层策略,提高了异常检测算法的有效性。Qian等^[16]提出了一种基于相对密度的多密度DBSCAN算法,通过相对密度提取数据集中的低密度数据点,然后在已识别的低密度数据点中找到自然集群,最后通过使用DBSCAN对剩余数据点进行聚类,有效提高了聚类效果。Wei等^[17]提出了基于相对密度的密度峰值聚类算法,结果表明算法在簇间密度相差较大的数据集上能够准确识别聚类中心和分配剩余样本点。Jin等^[18]提出了一种基于DBSCAN的变尺度超立方体加速密度空间聚类的带噪应用异常检测方法,设计了超立方体分割的遍历搜索策略,实现了多维工业数据的异常检测。Zhang等^[19]描述了LOF算法在异常检测中的应用研究,结果显示算法根据节点参数较准确地检测出异常节点,从而避免了用户对异常节点的访问。Prabhakar等^[20]提出了基于LOF算法的EEG坏通道检测方法,使用LOF算法能够适应任何类型的EEG,通过实验得出,使用LOF算法能够获得性能好的指标。Yu等^[21]基于LOF算法设计了一种用于数据挖掘中异常检测的流式算法,有效克服了流数据异常检测中经常存在的伪装问题,提高了异常检测效果。

2 基于密度的心理沙盘检测算法

2.1 沙盘数据及特征提取

电子沙盘原始数据记录了沙盘创作者在构建沙盘作品的过程中所涉及的操作行为,主要包括沙子操作、所用沙具及对沙具的操作等。系统中可供使用的操作包括挖沙子、堆沙子、平整地形等10种无需具体沙具的操作和创建沙具、移动沙具、缩放沙具等12种需要具体沙具的操作。系统中可供使用的沙具有警察、士兵、城堡等549个沙具。因为沙子和沙具操作次数可以一定程度地反映沙游者的心理状态,所以本文将沙具操作和具体沙具进行结合,统计出沙子或沙具的不同操作次数并将其作为特征,共得到6 598个操作次数相关特征。

沙具是沙盘游戏的重要工具,不同类型的沙具具有不同的特征和象征意义,能够帮助心理咨询师更好地了解沙盘创造者的内心世界和精神情感^[22]。比如,象征着“欢乐、勇气、好运、桃李芬芳”的桃树和象征着“爱国情怀,戎装理想”的士兵这种沙具常被用来代表开心、积极等正面情感和体验;象征着“愤怒心理、负罪感”的阎罗王和象征着“阴影,创伤经历”的恶魔常被用来代表恐惧、痛苦等负面情感和体验。本文将象征负面情感和体验的沙具称之为“负面情感沙具”,简称为“负面沙具”。一个沙盘中负面沙具的使用频次也一定程度地反映了沙盘创作者的心理状态。因此,本文也将负面沙具的使用频次作为一个特征。

综上所述,本文共提取出6 599个特征,其中某些特征列全部为0,被认为是无用的特征列,将这些特征列全部删除以初步降维,得到1 593维特征,降低了约75%。为了便

于实现聚类结果可视化,本文使用TSNE降维方法将维度降到二维。

数据集中不同特征值的取值范围不同对最终结果影响较大,本文对所有数据进行规范化处理,又因为负面沙具的使用次数对该沙盘是否为异常沙盘的影响较大,且样本中负面沙具使用频次99.8%位于20次以下,故利用式(1)将所有数据规范化到[0,20]区间。

$$X_{norm} = \frac{(X - X_{min}) * 20}{X_{max} - X_{min}} \quad (1)$$

其中, X 代表原始数据, X_{min} 代表原始数据中的最小值, X_{max} 代表原始数据中的最大值, X_{norm} 代表规范化后的数据。

2.2 DBSCAN算法

DBSCAN是一种典型的基于密度的空间聚类算法。该算法通过对数据点在空间上的密度进行分析,将具有足够密度的区域划分为簇,将数据点分为核心点、边界点和噪声点从而实现聚类。其中,核心点为满足参数(ϵ , MinPts)要求的数据, ϵ 为半径,MinPts为一个核心点的 ϵ -邻域包含点的最小个数;边界点不是核心点,但落在某个核心点的邻域内;噪声点是那些既不是边界点也不是核心点的数据。该算法利用密度相连点的最大集合来定义簇,从而避免了对簇的个数和形状的预设。

DBSCAN算法聚类过程如下:首先,根据给定参数(ϵ , MinPts)从给定数据集中找出所有核心对象;其次,以任一核心对象为出发点,找出其密度可达的样本生成聚类簇;再次,算法会选择下一个未被访问的核心对象,并以它为起点重复上述过程,直到所有核心对象均被搜索完成为止;最后,DBSCAN算法将生成一组聚类簇,每个簇包含核心对象及密度可达的样本点。同时,算法还可以将未被访问的样本点,即不属于任何聚类簇的样本点标记为噪声点,这些点既不是核心对象,也不在核心对象的邻域内。

2.3 LOF算法

LOF算法是一种基于密度的离群点检测方法。该算法评估数据点的异常程度,通过比较该点邻域密度与周围数据点邻域密度的比值以确定簇中的离群点。为实现检测,LOF算法利用基于密度的聚类的概念,在聚类空间中给每个数据点分配一个依赖于邻域密度的离群因子,对于每个数据点而言,可以通过其离群因子大小判断数据点是否为离群点。

LOF算法实现过程如下:首先,计算每个点 p 的 k -近邻距离 k -distance(p),即数据点 p 到第 k 个最近邻的距离,确定其 k -近邻集合 $N_k(p)$;其次,利用式(2)计算每个点的可达距离,其中 $q \in N_k(p)$, $d(p, q)$ 为数据点 p 与 q 之间的直接距离;再次,计算局部可达密度,如式(3)所示;最后,利用式(4)计算每个点的局部离群因子,将其作为每个数据点的异常分数。

$$reach_{dist_k}(p, q) = \max \{k - distance(q), d(p, q)\} \quad (2)$$

$$lrd_k(p) = \frac{1}{\sum_{q \in N_k(p)} \frac{reach_{disk_k}(p, q)}{|N_k(p)|}} \quad (3)$$

$$OF_k(p) = \frac{\sum_{o \in N_k(p)} \frac{lrd(o)}{lrd(p)}}{|N_k(p)|} \quad (4)$$

2.4 融合算法

本文使用集成学习的方法,将DBSCAN算法进行改进并与LOF算法进行结合,以提高异常心理沙盘的检测性能。

对于DBSCAN算法的改进,先使用DBSCAN算法进行聚类得到异常簇和单独噪声点,再使用欧氏距离计算出异常簇中的点和噪声点与主体簇中心的距离,将该距离作为该点的异常分数。

融合算法具体步骤如下:①使用DBSCAN算法进行聚类并且求出异常点和异常点分数;②使用LOF算法求出离群点和局部离群因子,将局部离群因子作为异常点分数;③使用最小—最大规范化对DBSCAN算法的异常点分数进行处理,将其范围规范在LOF算法的异常点分数范围内;④将DBSCAN算法和LOF算法分别得到的异常点取交集,按照DBSCAN算法规范化后的异常点分数和LOF算法得到的异常点分数由大到小排序。重复点取异常分数排名靠前的点,得到最终异常点及异常分数。

算法伪代码如下:

3 实验与结果分析

3.1 数据集

本文实验数据来源于“聆心云数字化心理健康服务平台”的5 697条电子心理沙盘数据。由于操作数较少的沙盘数据难以提供足够的信息以进行准确的异常检测,本文将沙盘数据操作数少于10个的数据样本进行剔除,最终保留了4 800个有效心理沙盘数据。

3.2 参数设置

3.2.1 改进的DBSCAN算法半径 ε 设置

在DBSCAN算法中,半径控制着一个核心点的邻域大小。如果两个点之间的距离 $\leq \varepsilon$,则这两个点被认为在同一簇中。若邻域值过小,则大多数点将被标记为噪声点,会导致聚类粒度过细,从而出现很多小簇。如果邻域值过大,则所有点可能被视为同一簇中的点,这会导致聚类粒度过粗,噪声点过少。

k -近邻算法是一种基于实例分类的机器学习经典算法, k -近邻算法用于显示每个点与其第 k 个最近邻点之间的距离,其中 k 值是事先选择的参数。图2为 k 值与半径 ε 值的对应图,其中横轴代表 k 值,纵轴代表半径 ε 值。

本文采用 $k = [1, 100]$ 以计算DBSCAN算法的半径 ε 值,使用网格搜索法以调整合适的半径 ε 值,并且选择准

算法:异常沙盘检测算法

输入:样本集合 $D = \{x_1, x_2, x_3, \dots, x_n\}$,改进的DBSCAN算法参数 $(\varepsilon, MinPts)$,LOF算法参数邻域 k 。

输出:异常点集合 C

```

1 FOR EACH  $p \in D$  DO
2 确定 $p$ 的 $k$ 近邻 $N_k(p)$ ;
3 计算 $p$ 的可达距离 $reachdist(p, q), q \in N_k(p)$ ;
4 计算 $p$ 的局部可达密度 $lrd(p)$ ;
5 END
6 FOR EACH  $p \in D$  DO
7 计算 $p$ 的局部离群因子 $LOF(p)$ ,将其作为每个数据点的异常分数;
8 END
9 初始化核心对象集合: $O \leftarrow \emptyset$ ;
10 FOR  $j = 1, 2, \dots, n$  DO
11 确定样本 $x_j$ 的邻域 $\varepsilon$ ;
12 IF  $|N_\varepsilon(x_j)| \geq MinPts$  THEN
13 将样本 $x_j$ 加入核心对象集合: $O \leftarrow O \cup \{x_j\}$ ;
14 END
15 END
16 初始化聚类簇数: $m = 0$ ;
17 初始化未访问样本集合: $\Gamma = D$ ;
18 WHILE  $\Gamma \neq \emptyset$  DO
19 记录当前未访问样本集合: $\Gamma_{old} \leftarrow \Gamma$ ;
20 随机选取一个核心对象 $o \in O$ ,初始化队列 $Q = \{o\}$ ;
21  $\Gamma \leftarrow \Gamma \setminus \{o\}$ ;
22 WHILE  $Q \neq \emptyset$  DO
23 取出队列 $Q$ 的首个样本 $q$ ;
24 IF  $|N_\varepsilon(x_i)| \geq MinPts$  THEN
25 令 $\Delta = N_\varepsilon(q) \cap \Gamma$ 
26 将 $\Delta$ 中的样本加入队列 $Q$ ;
27  $\Gamma \leftarrow \Gamma \setminus \Delta$ 
28 END
29 END
30  $m \leftarrow m + 1$ ,生成聚类簇 $C_m = \Gamma_{old} \setminus \Gamma$ ,最大的簇命名为 $C_0$ 
31  $O \leftarrow O \setminus C_m$ 
32 FOR  $j = 1, 2, \dots, m$  DO
33 取出 $C_j$ 中的每一个数据点,计算其与最大簇的欧氏距离,作为异常点分数。
34 END
35 将改进的DBSCAN算法的异常点分数规范化到LOF算法的异常点分数范围。
36 将两个算法得到的异常点取交集,并且将分数规范化后重新排序,得到最后的异常点及异常分。

```

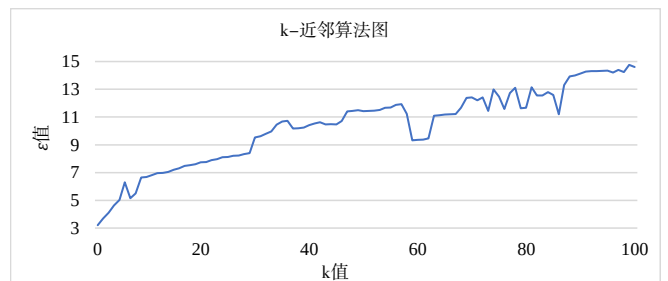


Fig. 2 K-Nearest algorithm

图2 k -近邻算法

确率最高的半径 ε 值作为实验所用的半径 ε 值。如图 3 所示,当 $k=5$ 时,所得算法效果最好,此时 $\varepsilon = 5.036$ 。

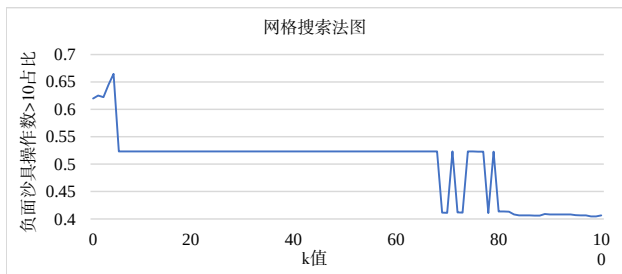


Fig. 3 Grid search method

图 3 网格搜索法

3.2.2 改进的 DBSCAN 算法 $MinPts$ 设置

在使用 DBSCAN 算法时, $MinPts$ 的值需要根据 k 值确定,通常使用 $MinPts$ 的取值为上述 k 值加 1。由于以上步骤已确定最优邻域值 ε ,可在 ε 固定的情况下,选择不同的 $MinPts$ 值: $MinPts \in [2, 101]$ 。由实验可得 $k=5$,故 $MinPts = 6$ 。

3.2.3 LOF 算法 $n_neighbors$ 设置

参数 $n_neighbors$ 对 LOF 算法结果有很大影响,同样使用网格搜索方式以调整 LOF 算法的参数 $n_neighbors$ 。将 $n_neighbors$ 值从 2-100 进行遍历,并在每个 $n_neighbors$ 值下计算模型的平均异常得分,如图 4 所示。选取得分最高的 $n_neighbors$ 值作为最优参数值,根据结果,确定 $n_neighbors$ 取值为 11。

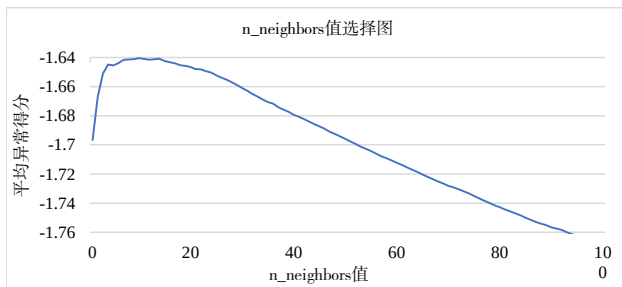


Fig. 4 Selection of parameter $n_neighbors$

图 4 参数 $n_neighbors$ 的选择

3.3 实验结果与分析

根据心理专家提供的部分沙具象征意义信息,筛选出包含负面情感的沙具,并计算沙具的情感得分。大于零的值表示正面情感、零表示中性情感、小于零的值表示负面情感,沙盘中负面沙具操作数与其情感值的乘积作为沙盘的负面情感得分,负面情感分值越高说明越异常。本文所提算法会对所识别出的异常沙盘计算异常分数,并依此进行排序,然后根据排名前 100 和排名后 100 的沙盘中负面沙具操作数比例和负面情感值比例对算法结果进行评估。此外,本文与改进的 DBSCAN 算法、LOF 算法、孤立森林算法^[23]和 LSCP 算法^[24]进行了对比实验,以验证本文算法的有效性。

3.3.1 改进的 DBSCAN 算法性能分析

改进的 DBSCAN 算法将沙盘聚成 31 簇,聚类效果如图 5 所示。

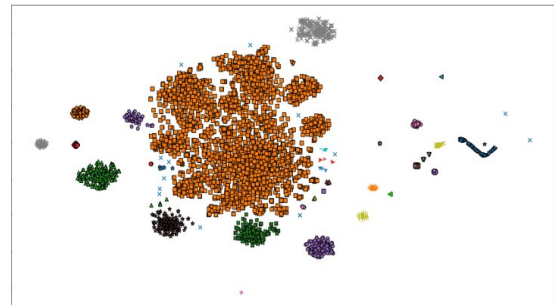


Fig. 5 DBSCAN algorithm clustering effect

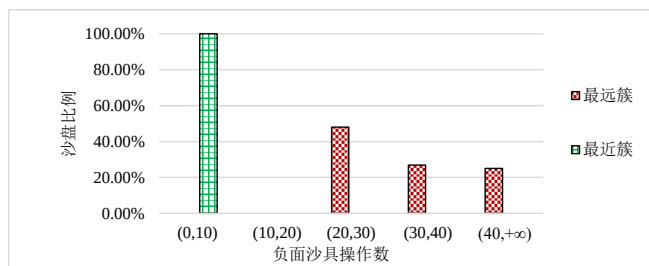
图 5 DBSCAN 算法聚类效果

对图 5 进行分析可得,标签为 -1 的为最大一簇,得到的沙盘为正常沙盘;在最大一簇周围的簇以及孤立点与最大簇的距离由沙盘的异常程度决定,距离最大簇最远的孤立点与最大簇的距离为 3.58,沙盘中含有负面沙具操作数 145 个,沙盘负面情感得分为 -116.29,此孤立点沙盘如图 1 (右) 所示;与最大簇距离最远的一个簇中共有 104 个沙盘,沙盘中含有负面沙具操作数和负面情感得分所占比例如图 6 所示,其中有 25% 的沙盘负面沙具操作数超过 40,超过 17.3% 的沙盘负面情感值低于 -40。距离最大簇最近的孤立点与最大簇的距离为 1.09,沙盘负面情感得分为 0,沙盘中含有负面沙具操作数 0 个,此孤立点沙盘如图 7 所示;与最大簇距离最近的 11 个簇中共有 138 个沙盘,沙盘中含有负面沙具操作数和负面情感得分所占比例如图 6 所示,这些沙盘负面沙具操作数都低于 10,并且超过 90% 的沙盘负面情感值高于 -20。综上所述,使用改进的 DBSCAN 算法对沙盘进行聚类,异常得分高的沙盘存在的负面沙具较多,沙盘的负面情感值也较高。因此,通过对比可以初步得出,使用改进的 DBSCAN 算法对心理沙盘进行聚类能够筛查出明显存在心理健康问题的沙盘。

3.3.2 不同算法检测异常心理沙盘性能比较

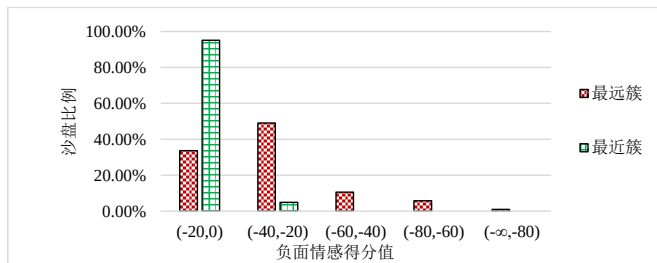
表 1 给出了改进的 DBSCAN 算法、LOF 算法、孤立森林算法、LSCP 算法和融合算法排名前 100 沙盘中负面沙具操作数比例,LOF 算法都低于 20;孤立森林算法大多低于 30,有 20% 的高于 40;LSCP 算法集中于 (20, 40) 之间,有 30% 的高于 40;融合算法有 89% 的数大于 20,有 11% 的在 (10, 20) 之间。对比发现,负面沙具操作次数大于 40 的比例中,融合算法比改进的 DBSCAN 算法高出 6%、比 LOF 算法高出 31%,比孤立森林算法高出 11%、比 LSCP 算法高出 1%。这说明融合后的算法可以检测出更多使用负面沙具操作数超过 40 的异常程度更大的心理沙盘。

表 2 给出了改进的 DBSCAN 算法、LOF 算法、孤立森林算法、LSCP 算法和融合算法排名前 100 沙盘中负面情感得分值比例,LOF 算法 96%、孤立森林算法 86%、LSCP 算法 90% 高于 -40;融合算法有 22% 的低于 -40,有 78% 的



(a) DBSCAN algorithm negative sand tool operation number

(a) DBSCAN算法负面沙具操作数



(b) DBSCAN algorithm negative sentiment score sand table ratio

(b) DBSCAN算法负面情感得分值沙盘比例

Fig. 6 DBSCAN algorithm negative sand tool operation ratio and negative emotion score sand table ratio

图6 DBSCAN算法负面沙具操作数比例和负面情感得分值沙盘比例



Fig. 7 Sand table of isolated points closest to the largest cluster

图7 距离最大簇最近的孤立点沙盘

在(0,−40)之间。对比发现,负面情感得分值低于−80的比例中,融合算法比改进的DBSCAN算法高出3.04%、比LOF算法高出4%、比孤立森林算法高出2%、比LSCP算法高出4%,这说明融合后的算法可以检测出更多负面情感得分值低于−80的异常程度更大的心理沙盘。此外,如图8所示,融合算法排名后100的异常沙盘中有97%的负面情感得分值在(0,−20)之间。

综上所述,本文融合算法能够筛选出更多异常程度较大的心理沙盘。

Table 1 The proportion of negative sand tool operations in the top 100 sand tables ranked by different algorithms

表1 不同算法排名前100沙盘负面沙具操作数比例 (%)

负面沙具操作数	(0,10)	(10,20)	(20,30)	(30,40)	>40
改进的DBSCAN算法	0	0	48.08	26.92	25
LOF算法	31	69	0	0	0
孤立森林算法	62	5	10	3	20
LSCP算法	0	0	43	27	30
融合算法	0	11	34	24	31

Table 2 The proportion of negative sentiment score values in the top 100 sand tables ranked by different algorithms

表2 不同算法排名前100沙盘负面情感得分值比例 (%)

负面情感得分值	(−20,0)	(−40,−20)	(−60,−40)	(−80,−60)	(−∞,−80)
改进的DBSCAN算法	33.65	49.04	10.58	5.77	0.96
LOF算法	87	9	3	1	0
孤立森林算法	57	29	5	7	2
LSCP算法	82	9	9	0	0
融合算法	33	45	11	7	4

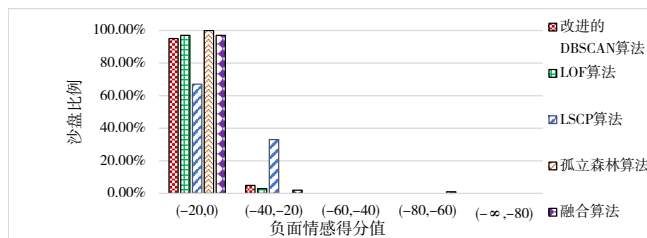


Fig. 8 Negative sentiment score ratio of the last 100 sand tables

图8 排名后100沙盘负面情感得分值比例

4 结语

本文对沙盘数据进行预处理、提取特征,并融合基于密度的异常检测算法DBSCAN和LOF算法对心理沙盘进行分析,求得心理沙盘的异常分数,找出异常心理沙盘,有效地识别出心理不健康的沙游者,为心理专家的诊断提供有效的参考价值。实验结果表明,融合后的算法可以检测出更多使用负面沙具操作数超过40、负面情感得分低于−80的异常程度更大的心理沙盘。

未来工作包括如下:提取沙具操作时间、相对位置等特征;利用基于距离、概率统计或深度学习等其他类型的孤立点分析算法对心理沙盘进行分析,进一步提升异常沙盘检测效果。

参考文献:

- [1] China National Narcotic Drugs Association. The blue book on China's mental health in 2023 has been released[EB/OL]. <https://mp.weixin.qq.com/s/CJS1FY5-yHKn2l-YXdmLRQ>.
中国麻醉药品协会.《2023年度中国精神心理健康》蓝皮书发布[EB/OL]. <https://mp.weixin.qq.com/s/CJS1FY5-yHKn2l-YXdmLRQ>.
- [2] CHENG Z H, CAI F, ZENG G, et al. Clinical efficacy of sandplay psychotherapy in patients with depression[J]. China Journal of Practical Medicine, 2021, 16(19): 203-205.
成芝花,蔡凤,曾干,等.沙盘游戏心理治疗在抑郁患者中的临床疗效分析[J].中国实用医药,2021,16(19):203-205.
- [3] LI Y Y, ZHOU P, JIANG Z B. Efficacy of sandplay therapy in the treatment of psychological disorders in children with cerebral palsy[J]. China Journal of Practical Medicine, 2023, 18(14): 159-161.
李栢望,周平,蒋志斌.沙盘游戏疗法治疗脑性瘫痪儿童心理障碍的疗效观察[J].中国实用医药,2023,18(14):159-161.
- [4] DUAN Y N, MAN Y C. Exploration of the application of group sandplay in mental health education in colleges and universities[J]. Journal of Science

- and Education, 2023(27): 130-132.
- 段颖娜, 满园春. 团体沙盘游戏在高校心理健康教育中的应用探索[J]. 科教导刊, 2023(27): 130-132.
- [5] ZHAO M S. Application of psychological sand tables in mental health education in primary schools[J]. Electronic Journal of the New Education Era (Student Edition), 2022(42): 24-26.
- 赵明珠. 心理沙盘在小学心理健康教育的应用[J]. 新教育时代电子杂志(学生版), 2022(42): 24-26.
- [6] HAN J W. Data mining: concepts and techniques (3rd Edition)[M]. Beijing: China Machine Press, 2022.
- 韩家炜. 数据挖掘: 概念与技术(原书第3版)[M]. 北京: 机械工业出版社, 2022.
- [7] KONG L C, LIU G Z. Overview of outlier detection algorithms[J]. Computer Science, 2024, 51(8): 20-33.
- 孔翎超, 刘国柱. 离群点检测算法综述[J]. 计算机科学, 2024, 51(8): 20-33.
- [8] FANG M B, WEI K W, LONG C. DBSCAN: density-based spatial clustering of applications with noise [J]. Journal of Nanjing University (Natural Sciences), 2012, 48(4): 491-498.
- [9] BREUNIG M M, KRIEGER H P, NG R T, et al. LOF: identifying density-based local outliers[C]//Proceedings of the 2000 ACM Sigmod International Conference on Management of Data, 2000: 93-104.
- [10] YING S, WANG B M, WANG L, et al. An improved KNN-based efficient log anomaly detection method with automatically labeled samples [J]. ACM Transactions on Knowledge Discovery from Data, 2021, 15(3): 1-22.
- [11] SHUAI L Y, DONG M M, WAI K W. REB: reducing biases in representation for industrial anomaly detection [J]. Knowledge-Based Systems, 2024, 290: 111563.
- [12] ZHANG Z P, LI S, LIU W X, et al. A new outlier detection algorithm based on fast density peak clustering outlier factor[J]. International Journal of Data Warehousing and Mining (IJDWM), 2023, 19(2): 1-19.
- [13] CHEN Y W, HU X L, FAN W T, et al. Fast density peak clustering for large scale data based on KNN [J]. Knowledge-Based Systems, 2020, 187: 104824.
- [14] LI K S, GAO X, JIA X, et al. Detection of local and clustered outliers based on the density-distance decision graph [J]. Engineering Applications of Artificial Intelligence, 2022, 110: 104719.
- [15] HU W M, GAO J, LI B, et al. Anomaly detection using local kernel density estimation and context-based regression [J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 32(2): 218-233.
- [16] QIAN J X, ZHOU Y, HAN X M, et al. MDBSCAN: a multi-density DBSCAN based on relative density [J]. Neurocomputing, 2024, 576: 127329.
- [17] WEI Y, ZHANG Z J, HE K L, et al. Density peak clustering algorithm based on relative density [J]. Computer Engineering, 2023, 49(6): 53-61.
- 位雅, 张正军, 何凯琳, 等. 基于相对密度的密度峰值聚类算法[J]. 计算机工程, 2023, 49(6): 53-61.
- [18] JIN F, WU H, LIU Y, et al. Varying-scale HCA-DBSCAN-based anomaly detection method for multi-dimensional energy data in steel industry [J]. Information Sciences, 2023, 647: 119479.
- [19] ZHANG G, GAI Y. Application of local outlier factor algorithm (LOF) in anomaly detection [J]. Network Security Technology and Application, 2020(11): 49-50.
- 张戈, 盖赞. 局部离群因子算法(LOF)在异常检测中的应用研究[J]. 网络安全技术与应用, 2020(11): 49-50.
- [20] PRABHAKAR V K, MARCO B, EUGENIO P, et al. Adaptable and robust EEG bad channel detection using local outlier factor (LOF)[J]. Sensors, 2022, 22(19): 7314-7314.
- [21] YU K Q, SHI W, SANTORO N. Designing a streaming algorithm for outlier detection in data mining—an incrementa approach [J]. Sensors, 2020, 20(5): 1261.
- [22] YAO J H. Common factors and symbolic significance in long-term sand-play game counseling for children with autism[J]. Psychological Monthly, 2021, 16(22): 215-217.
- 姚晶宏. 孤独症儿童长程沙盘游戏咨询中的共同因素及象征意义[J]. 心理月刊, 2021, 16(22): 215-217.
- [23] LIU D T, TING K M, ZHOU Z H. Isolation forest[C]//2008 Eighth IEEE International Conference on Data Mining, 2008: 413-422.
- [24] ZHAO Y, NASRULLAH Z, HRYNIEWICKI M K, et al. LSCP: locally selective combination in parallel outlier ensembles[C]//Calgary: SIAM International Conference on Data Mining (SDM), 2019: 585-593.

(责任编辑: 孙娟)