

doi: 10.7690/bgzd.2026.06.014

基于深度学习的学前教育平台典型数据多目标优化挖掘算法

许萌, 韩鹏

(咸阳职业技术学院师范学院, 陕西 咸阳 712000)

摘要: 针对学前教育方案制定多目标问题, 提出基于深度学习的学前教育平台典型数据多目标优化挖掘算法。采用时间窗口和频繁项集, 提取多源异构教育平台典型数据的内容特征和结构特征, 输入卷积神经网络模型; 在该模型的卷积层中引入多层异构注意力机制, 映射提取的特征结果; 利用批归一化层重构映射结果, 池化层分割重构得到特征结果; 通过模型的全连接层组合分割后典型数据特征, 由 Softmax 分类器进行典型数据划分, 获取优化后的典型数据多目标优化挖掘结果。测试结果表明: 该算法特征提取效果良好, 均方根误差均小于 0.12, 数据挖掘的特异性结果均大于 0.927, 挖掘的典型数据的新颖度结果均大于 91.6%。

关键词: 深度学习; 学前教育平台; 典型数据; 多目标; 优化挖掘; 新颖度

中图分类号: TP183 **文献标志码:** A

Multi-objective Optimization Mining Algorithm for Typical Data of Preschool Education Platform Based on Deep Learning

Xu Meng, Han Peng

(Teachers' College, Xianyang Vocational and Technical College, Xianyang 712000, China)

Abstract: Aiming at the multi-objective problem in preschool education program formulation, this paper proposes a multi-objective optimization mining algorithm for typical data of preschool education platform based on deep learning. Extracting content characteristics and structural characteristics of typical data of a multi-source heterogeneous education platform by adopting a time window and a frequent item set, and inputting a convolutional neural network model; introducing a multi-layer heterogeneous attention mechanism into a convolution layer of the model, and mapping the extracted characteristic results; reconstructing the mapping result by utilizing a batch normalization layer, and segmenting and reconstructing by utilizing a pooling layer to obtain the characteristic results; the typical data features after segmentation are combined by the fully connected layer of the model, and the typical data are divided by the Softmax classifier to obtain the optimized typical data multi-objective optimization mining results. The test results show that the algorithm has a good feature extraction effect, the root mean square error is less than 0.12, the specificity results of data mining are more than 0.927, and the novelty results of typical data mining are more than 91.6%.

Keywords: deep learning; preschool education platform; typical data; multiple objectives; optimize mining; novelty

0 引言

学前教育是教育中的重要阶段, 该阶段对于孩子的思维引导、兴趣培养和开发等具有重要意义。在该阶段中, 老师需充分结合孩子的性格特点, 制定针对性的教育方案。学前教育属于一门综合性学科, 教育机构在实行学前教育计划制定以及执行过程中^[1], 需结合相关教育平台中优秀、典型的数据完成, 参考或者借鉴优秀教师或者优秀机构的教学经验^[2], 获取相关平台中的典型数据, 制定满足需求以及符合教育目标的教学方案^[3]。由于学前教育平台中数据量较大, 并且数据类别较多, 包含课件、视频、图像等典型数据, 在进行数据查找以及挖掘过程中^[4], 耗时较大, 并且在搜索过程中需根据单

项数据需求进行搜索^[5], 降低数据搜索效率; 因此, 为实现学前教育平台典型数据的高效搜索^[6], 需提升数据的搜索效率。文献[7]为快速完成所需数据的挖掘, 构建数据之间的关联规则后, 利用改进的 K -means 算法实现数据挖掘; 但该方法在应用过程中, 无法同时完成多种需求的数据挖掘。文献[8]为保证数据挖掘的精准性, 依据数据样本的密集分布情况, 进行均衡处理以及聚类处理, 获取数据挖掘结果; 但该方法在应用过程中, 对于多源数据的挖掘效果较差。

深度学习作为机器学习领域中的典型方法, 其已通过学习获取数据样本的内在规律^[9], 以此完成数据识别和分类; 因此, 该方法在数据挖掘、个性

收稿日期: 2024-12-07; 修回日期: 2025-01-14

基金项目: 陕西省职业技术教育学会 2023 年度教育教学改革研究项目 (2023SZX207); 咸阳职业技术学院教改项目 (2023SZX212); 咸阳职业技术学院科研基金项目 (2024KJB02)

第一作者: 许萌 (1987—), 女, 陕西人, 硕士。

化推荐等领域中具有较好的应用效果^[10]。笔者为实现学前教育平台典型数据挖掘，提出基于深度学习的学前教育平台典型数据多目标优化挖掘算法，通过该方法实现多源学前教育数据的挖掘，为学前教育计划制定和执行提供可靠依据。

1 学前教育平台典型数据多目标优化挖掘

1.1 学前教育平台典型数据多源特征提取

由于学前教育平台典型数据具有显著的多源异构特性^[11]，因此，在进行学前教育平台典型数据挖掘前，需提取多源异构平台教育典型数据特征提取。文中采用时间窗口和频繁项集提取多源异构教育平台典型数据特征。

如果多源异构学前教育平台中给定数据集用 $V = \{v_i | i = 1, 2, \dots, n\}$ 表示，依据多源异构学前教育平台典型数据之间的类似程度进行数据集划分，形成 C_i 和 C_j 簇，且 $C_i, C_j \subseteq V, i, j = 1, 2, \dots, n$ ，且 $i + j = n$ ，同时满足式(1)的约束：

$$\left. \begin{aligned} C_i \cup C_j &= \emptyset, i \neq j \\ C_i \cup C_j &= V \end{aligned} \right\} \quad (1)$$

如果多源异构学前教育平台典型数据之间的相异度矩阵用 η 表示，其可通过 n 个典型数据中可能存在的 2 个数据之间的相异性描述，该矩阵可采用 $n \times n$ 表示，如果 2 个数据目标分别用 v_1 和 v_2 表示，两者之间的相异性用 $d(v_1, v_2)$ 表示， η 的表达式为：

$$\eta = \begin{pmatrix} 0 & & & \\ d(2,1) & 0 & & \\ \vdots & \vdots & & \\ d(n,1) & d(n,2) & \dots & 0 \end{pmatrix} \quad (2)$$

$d(v_1, v_2)$ 在正常情况下均为非负数，如果 v_1 和 v_2 之间非常接近，则 $d(v_1, v_2)$ 最大程度接近 0。 $d(v_1, v_2)$ 的值越大，表示 v_1 和 v_2 之间的差异越明显，基于此可获取 η 。

依据 η 的结果，提取多源异构学前教育平台典型数据统计序列的特征量，为数据多目标优化挖掘提供可靠依据^[12]。如果多源异构学前教育平台典型数据统计特征值矩阵为：

$$\psi = \eta(R^n, J)^{-1} \quad (3)$$

式中： R^n 为多源异构学前教育平台典型数据检索模糊域； J 为多源异构学前教育平台典型数据分块匹配集。

采用融合的方式完成多源异构学前教育平台典型数据处理^[13]，并计算异构平台数据的聚类中心，

以此获取多源异构学前教育平台典型数据的特征分布域叙述结果 I_{sm} ：

$$I_{sm} = A_{sm} \psi (1 - (2\pi a + D_{sm}) / \xi_{sm}) \quad (4)$$

式中： A_{sm} 为典型数据的加权输出幅值； ξ_{sm} 为典型数据聚类自适应调节参数； D_{sm} 为约束条件。

在上述公式的基础上，依据模糊聚类中心，采用平均加权方法描述多源异构学前教育平台典型数据时间窗口 T ：

$$T = I_{sm}(T_1/T_2) \quad (5)$$

式中 T_1 和 T_2 分别为事件时间和处理事件。

依据式(5)的计算结果，可获得多源异构学前教育平台典型数据融合全局性的寻优返回值，依据该数据获取多源异构学前教育平台典型数据链路增益值，提取多源异构学前教育平台典型数据特征，文中结合学前教育平台典型数据的特点^[14-15]，仅选择数据的内容特征和结构特征作为目标提取特征，以此组成多特征集 X 。

1.2 多源典型数据多目标优化

1.2.1 基于深度学习典型数据多目标优化挖掘模型

笔者为实现多源异构学前教育平台典型数据多目标优化挖掘，该模型结构如图 1 所示。

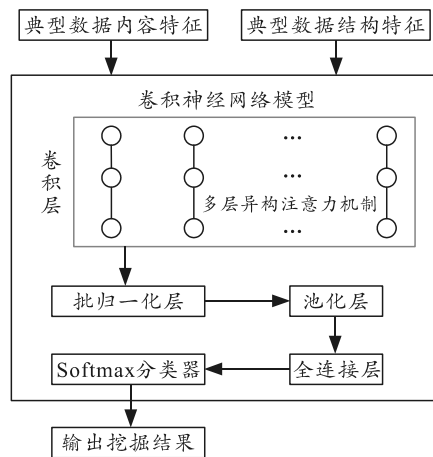


图 1 基于深度学习的典型数据多目标优化挖掘模型结构

该模型是将异构注意力机制融合在卷积层中，利用该机制提升典型数据的多目标分类精度，最后通过全局平均池化层和全连接实现典型数据的分类，输出典型数据挖掘结果。

1.2.2 典型数据的多特征卷处理

将提取的典型数据多特征集 X 输入模型中，通过模型的卷积层进行处理，完成特征映射，实现多特征的深度刻画。卷积层主要是由多个卷积核组成，每个卷积核均为一个 2 维矩阵，可实现多特征集的

过滤和映射，并且采用反向传播完成各个卷积核参数的确定。

如果卷积层是模型的第一层处理层，通过该层完成 X 的映射处理：

$$X_j^l = \sum_{i \in M_j} X_i^{l-1} * K_{ij}^l + b_j^l \quad (6)$$

式中： X_j^l 为典型数据第 j 个特征在第 l 层终端映射结果； X_i^{l-1} 为从 $l-1$ 层中获取的第 i 个输入特征； M_j 为卷积区域； $*$ 为卷积算子； K_{ij}^l 为卷积核； b_j^l 为偏置项。

通过卷积神经网络模型在进行典型数据多目标挖掘时，为更精准完成典型数据多特征描述，引入多层异构注意力机制，该机制能够有效描述不同特征之间的关联性以及连续性，依据特征向量之间存在的关联性，确定需重点关注的特征；即对特征 x_i 而言，其上下文的关联性直接决定需重点关注的特征 x_j 结果，特征向量和其对应的注意力权重计算公式为：

$$\tilde{x}_i = \sum_{j, j \neq i}^m \lambda_{i,j} \times x_j \quad (7)$$

$$\lambda_{i,j} = \exp[f(x_i, x_j)] / \sum_{j'=1}^m \exp[f(x_i, x_{j'})] \quad (8)$$

$$f(x_i, x_j) = \tanh[\theta(x_i \oplus x_j)] \quad (9)$$

$$\sum_{j, j \neq i}^m \lambda_{i,j} = 1 \quad (10)$$

式中： (x_i, x_j) 为特征对； $\lambda_{i,j}$ 为注意力得分； $f(x_i, x_j)$ 为相关性量化函数； θ 为注意力参数矩阵； \oplus 为串联处理。

通过完成 \tilde{x}_i 和 x_i 的连接，形成新的特征向量，实现典型数据转换，获取转换后的典型数据特征 x' 。

1.2.3 典型数据特征归一化重构

笔者为降低模型的初始化敏感程度，同时提升模型训练效率，在卷积层之后添加批归一化层，用于实现模型激活值、梯度传播速度的归一化处理，将 x' 输入归一化层中：

$$x_j'' = (X_j' - \mu_\beta) / (\sqrt{\delta_\beta^2 + \varepsilon}) \quad (11)$$

$$Y_j^l = \gamma x_j'' + \beta \quad (12)$$

式中： x_j'' 为 x' 的归一化处理结果； μ_β 为上层输出的典型数据均值； δ_β^2 为标准差； ε 为极小值； γ 和 β

均为学习参数； Y_j^l 为特征重构结果。

1.2.4 典型数据特征分割

典型数据特征重构后，为降低参数之间的相互依赖性，笔者采用 ReLU 激活函数，同时经由池化层对重构后特征结果进行分割，删除特征中的冗余特征信息，降低典型数据特征维度、减小模型的过拟合风险，并保证有效典型数据的质量。

笔者选择最大池化层完成 Y_j^l 的处理：

$$Y_{j_0}^{l+3} X_{j_0} = \max_{(j_0-1)S \leq r \leq j_0 S} [y_j^l(r)] \quad (13)$$

式中： S 为池化层尺寸； $Y_{j_0}^{l+3}$ 为池化处理后输出的第 j_0 个特征结果； $y_j^l(r)$ 为第 r 个神经元结果。

获取多源异构学前教学平台典型数据的内容特征和结构特征 y_a 和 y_b 。

1.2.5 多源典型数据分类输出

模型通过全连接层完成分割后典型数据特征的组合处理后，依据各个输入特征的返回概率结果，利用 Softmax 分类器进行划分，将其划分至一个互斥类别中，完成多源异构学前教学平台典型数据挖掘。

依据 R^n 中典型数据的权重和偏置项变量完成 y_a 和 y_b 的计算：

$$\left. \begin{aligned} y_a &= f(wy + b) \\ y_b &= \frac{R^n}{n} f\|x\| \end{aligned} \right\} \quad (14)$$

式中： w 为权重； $f\|x\|$ 为数据特征结构相似性； $f(\bullet)$ 为 ReLU 激活函数。

将其计算结果进行映射寻优，获取优化后的典型数据多目标优化挖掘结果：

$$\left. \begin{aligned} \text{sim}(y_a, Q_a) &= (y_a + Q_a) / (y_a \times Q_a) \\ \text{sim}(y_b, Q_b) &= (y_b + Q_b) / (y_b \times Q_b) \end{aligned} \right\} \quad (15)$$

式中 Q_a 和 Q_b 分别为深度学习获取的多源数据之间的内容和结构特征。

在典型数据多目标优化挖掘过程中，以最小损失为学习目标，完成参数更新，文中以交叉损失熵函数作为损失函数：

$$E = - \sum_{n=1}^N \sum_{m=1}^M g_{ab} \ln y_{ab} \quad (16)$$

式中： N 为典型数据样本数量； M 为类别数量； g_{ab} 为样本 a 是否属于类别 b ； y_{ab} 为模型将样本 a 划分为 b 类的概率输出结果，即 Softmax 分类器最终挖

掘输出结果。

2 测试分析

为验证该算法的应用效果，笔者以某学前教育机构作为实例研究对象，该机构主要针对学龄前儿童(4~6岁)思维引导、兴趣培养等进行教育，因此，对孩子的性格情况以及教学目标制定教学计划，需要在多个教育平台网站上进行资源和典型数据的查找和挖掘。在挖掘过程中，所需的典型数据类别较多，因此，为提升典型数据的挖掘效果，利用该算法进行典型数据挖掘。在进行测试过程中，笔者仅以 2 个学前教育平台中各 5 个典型数据集作为测试数据，所有数据集中均包含图像、文件、视频、文字等多种典型数据，数据总量为 5 660 个。

参数设定：深度学习模型卷积核宽度设为 3、4、5，卷积核数量为 256，深度为 145，神经元数量为 130，迭代次数 50，学习率为 0.000 1。

为验证该算法的学前教育平台典型数据多源特征提取效果，随意在 2 个教育平台中各选择一个数据集，采用该算法提取数据集中的典型数据特征，并获取多特征提取结果，如图 2 所示。

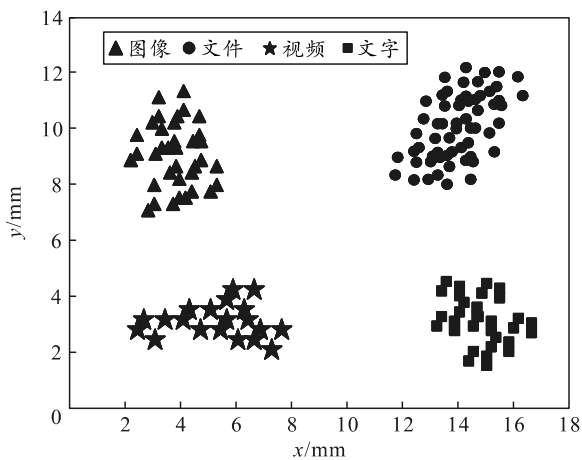


图 2 多特征提取结果

依据图 2 测试结果可知：采用本文中算法进行学前教育平台典型数据多源特征提取后，能够可靠获取不同典型数据的特征分布结果，并且各类特征之间界限清晰，不存在重叠现象。该算法具有较好的典型数据多特征提取能力，可为典型数据多目标优化挖掘提供可靠依据。

为进一步验证该算法的特征提取效果，随机选择视频数据特征和文件数据特征提取为例，以均方根误差 RMSE 为评价指标：

$$RMSE = \sqrt{\sum_{r_{ui} \in N} (r_{ui} - r_{ei})^2 / |N|} \quad (17)$$

式中 r_{ui} 和 r_{ei} 分别为实际结果和提取结果。

依据式(17)计算该算法在进行特征提取时，在不同特征向量维度下，RMSE 的测试结果，允许标准为 0.12，测试结果如表 1 所示。

表 1 均方根误差测试结果

特征向量维度	视频数据特征	文件数据特征
2	0.04	0.06
4	0.06	0.07
6	0.05	0.08
8	0.07	0.05
10	0.03	0.10
12	0.05	0.06
14	0.06	0.09
16	0.07	0.11
18	0.09	0.06
20	0.05	0.07

依据表 1 测试结果可知：采用该算法在进行视频数据特征和文件数据特征提取时，均方根误差结果均小于 0.12，2 种特征提取结果的均方根误差最大值分别为 0.09 和 0.11，满足学前教育平台典型数据多特征提取标准。

为验证该算法的学前教育平台典型数据多目标优化挖掘效果，获取该算法在不同数据类别和不同数据量下，对文件、视频、文字 3 种典型数据的优化挖掘的特异性，特异性主要用于描述正确挖掘比例，其取值范围在 0~1 之间，其值越大，表示挖掘效果越佳，测试结果如图 3 所示。

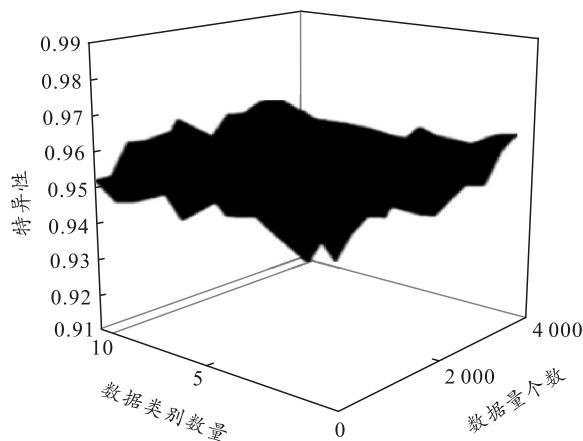


图 3 典型数据挖掘偏差结果

依据图 3 测试结果可知：不同数据类别和不同数据量下，采用该算法同时对文件、视频、文字 3 种典型数据进行挖掘后，特异性结果均大于 0.927。该算法具有精准的学前教育平台典型数据多目标优化挖掘效果，能够较好地完成所需数据的挖掘，为教育计划的制定提供数据依据。

为进一步验证该算法的应用性，笔者采用数据的新颖度 K_q 作为评价指标，用于衡量文中所挖掘的典型数据的新颖水平，其取值结果在 0~100%之间，

其值越大，表示挖掘的典型数据新颖度越佳，挖掘效果越好，计算公式为：

$$\kappa_q = \sum_{i=1}^q h(q, i) / N_i \quad (18)$$

式中： q 为典型数据的挖掘关键词； $h(q, i)$ 为该关键词在时间戳上的文档频率倒数； N_i 为关键词长度。

依据式(18)计算该算法对于图像、文件、视频、文字 4 种学前教育平台典型数据的 κ_q 结果，测试结果如表 2 所示。

表 2 典型数据挖掘新颖度测试结果

关键词长度/Byte	图像/%	文件/%	视频/%	文字/%
2	92.3	93.5	94.4	96.1
4	94.1	94.8	95.5	95.2
6	95.2	96.6	94.9	97.1
8	94.7	97.5	96.1	98.3
10	93.3	98.1	97.6	99.6
12	92.9	96.2	98.4	95.4
14	95.6	94.6	99.3	94.3
16	96.7	93.7	98.1	91.6
18	99.5	99.2	96.2	92.2
20	97.8	97.9	95.7	94.7

依据表 2 测试结果可知：采用该算法进行图像、文件、视频、文字 4 种学前教育平台典型数据优化挖掘后，数据的新颖度结果均大于 91.6%，其中最大值达到 99.6%。由此可见，该算法具有较好的典型数据优化挖掘效果，可较好地获取平台中的典型数据，并且最大程度保证数据的新颖度。

3 结论

笔者提出基于深度学习的学前教育平台典型数据多目标优化算法。对该算法的应用效果进行相关测试的结果表明：该算法能够有效完成多目标典型数据的特征提取，可依据挖掘需求完成多目标数据的可靠挖掘并且挖掘数据的新颖度较好，可为教育计划的制定提供更好的参考数据。

参考文献：

[1] 黄超然, 佟兴, 张召, 等. 面向教育的区块链应用合约架构和数据隐私研究[J]. 华东师范大学学报(自然科学

版), 2022(5): 61-72.
 [2] 李英, 汤庸. 基于关联规则与相似度的数据挖掘算法研究[J]. 华南师范大学学报(自然科学版), 2021, 53(5): 121-127.
 [3] 熊蕾, 彭吉琼, 李铭, 等. 基于深度学习的基层网络数据个性化挖掘算法[J]. 计算机仿真, 2022, 39(1): 318-321, 332.
 [4] 沐燕舟, 王赞, 陈翔, 等. 采用多目标优化的深度学习测试优化方法[J]. 软件学报, 2022, 33(7): 2499-2524.
 [5] 宋菲菲, 隋栋, 周湘贞. 基于深度学习的智能学习资源推荐算法[J]. 南京理工大学学报, 2022, 46(2): 185-191.
 [6] 李瑞峰, 杨海峰, 蔡江辉, 等. 一种基于加权深度森林的离群数据挖掘算法[J]. 小型微型计算机系统, 2022, 43(7): 1426-1431.
 [7] 李璐, 刘鹤, 朱良宽. 基于改进的 K-means 算法的关联规则数据挖掘研究[J]. 小型微型计算机系统, 2021, 42(1): 15-19.
 [8] 黄文秀, 唐超尘, 神显豪, 等. 改进的 k 最邻近算法在海量数据挖掘中的应用[J]. 济南大学学报(自然科学版), 2021, 35(1): 24-28.
 [9] 戴礼灿, 代翔, 崔莹, 等. 基于深度集成学习的社交网络异常数据挖掘算法[J]. 吉林大学学报(工学版), 2022, 52(11): 2712-2717.
 [10] 匡俊攀, 赵畅, 杨柳, 等. 一种基于深度学习的异常数据清洗算法[J]. 电子与信息学报, 2022, 44(2): 507-513.
 [11] 周燕, 肖莉. 基于改进关联聚类算法的网络异常数据挖掘[J]. 计算机工程与设计, 2023, 44(1): 108-115.
 [12] 李娟. 基于 Hadoop 云平台的空间属性数据挖掘技术研究[J]. 南京理工大学学报, 2022, 46(4): 419-426.
 [13] 郭婧, 耿海军, 吴勇. 基于菌群优化的 K 均值聚类算法研究[J]. 南京理工大学学报, 2021, 45(3): 314-319.
 [14] 李锦琬, 包理群, 周彬. 改进果蝇优化算法及其在不平衡数据分类中的应用[J]. 西北师范大学学报(自然科学版), 2021, 57(3): 57-61.
 [15] 张珂, 杨歆豪, 张嘉慧, 等. 基于高次指数平滑动态边界限制的深度学习优化算法[J]. 信息与控制, 2021, 50(6): 685-693.