倪康，古今用，申乾坤. 基于PCA-GA算法的贵州省水库标准化管理因子研究[J]. 水利水电技术(中英文)，2025，56(S1)：569-576.
NI Kang，GU Jinyong，SHEN Qiankun. Research on standardized management factors of reservoirs in Guizhou Province based on PCA-GA algorithm[J]. Water Resources and Hydropower Engineering，2025，56(S1)：569-576.

基于PCA-GA算法的贵州省水库标准化管理因子研究

倪　康，古今用，申乾坤
(贵州省水利科学研究院，贵州 贵阳　550002)

摘　要：依赖专家对标对表的创建工作中，存在专家在扣分区间内自由把控扣分程度的现象，针对所造成的因子重要性、所占总分比例、扣分区间一致，但其差异化不一致的问题，利用主成分分析法(PCA)以离散度作为表征指标，重构《贵州省水库标准化管理评价标准》(简写为《评价标准》)，建立特征矩阵作为输入变量，利用GA遗传算法改进BP神经网络进行拟合验证。结果表明，《评价标准》因子离散程度越高，越能代表创建实际情况；GA算法通过建立解释目标与解释适应度核函数之间的映射关系，进一步提升了BP神经网络的泛化能力和精准度，改进后的GA-BP神经网络算法模型的拟合准确率达98.78%；重构后的《评价标准》拟合R^2高达0.953，较重构前提升了0.107，有着更好的拟合精度。目前从因子离散度角度出发对水利工程达标创建标准解构、重组的研究相对较少，研究结论能够较好的辅助贵州省水库标准化管理创建、后续《评价标准》的修编等工作。

关键词：水库标准化；主成分分析法；离散度；遗传算法；重构标准
DOI：10.13928/j.cnki.wrahe.2025.S1.087
中图分类号：TV-9　　**文献标志码**：A　　**文章编号**：1000-0860(2025)S1-0569-08

Research on standardized management factors of reservoirs in Guizhou Province based on PCA-GA algorithm

NI Kang，GU Jinyong，SHEN Qiankun
(Guizhou Water Resources Research Institute，Guiyang　550002，Guizhou，China)

Abstract：In the creation of benchmark tables relying on experts，there exists freely controlling the degree of deduction within the deduction interval. In response to the problem of inconsistent differentiation caused by the importance of factors，the proportion of total scores，and the consistency of deduction intervals，principal component analysis (PCA) is used to reconstruct the "Evaluation Criteria" and establish a feature matrix as input variables. GA genetic algorithm is used to improve the BP neural network for fitting verification. The result indicate that the higher the degree of dispersion of the factors in the Evaluation Criteria，the more representative they are of the actual situation of creation；The GA algorithm further enhances the generalization ability and accuracy of the BP neural network by establishing a mapping relationship between the explanatory target and the explanatory fitness kernel function. The improved GA-BP neural network algorithm model has a fitting accuracy of 98.78%；The fitted R^2 of the "Evaluation Criteria" after reconstruction is as high as 0.953，which has improved by 0.107 compared to before

收稿日期：2024-09-28
基金项目：机理-数据协同驱动下降雨诱发堆积层滑坡变形预测预警关键技术研究(KT202401)
作者简介：倪　康(1993—)，男，工程师，硕士，主要从事水利水电工程研究。E-mail：283196165@qq.com

reconstruction and has better fitting accuracy. At present, there is relatively little research on the deconstruction and reorganization of water conservancy engineering standardization from the perspective of factor dispersion. The research conclusion of this article can effectively assist in the creation of standardized management for reservoirs in Guizhou Province and the subsequent revision of the "Evaluation Criteria".

Keywords: reservoir standardization; principal component analysis; dispersion; genetic algorithm; refactoring standards

0　引　言

规范和加强水利工程标准化工作，是满足保障国家水安全，提升水旱灾害防御能力的需求，随着全球水资源压力的持续增大，水利工程的有效管理显得尤为重要[1]。随着计算机技术和大数据的进步，数学、机器学习方式已经被应用到标准化体系研究当中。2024，闫彭彭[2]以实际工程为例，依据法规、规范，建立了一种严格遵循法规、规范的多层次模糊综合评价模型，实现了按 SL 214—2015《水闸安全评价导则》评定的水闸安全类别。赵勇、李溦等[3]利用分区评价建立了“双指标、四要素、三等级”的水资源短缺评价体系，并据此开展全国 345 个地市水资源短缺评价。构建算法模型与评价标准体系间的拟合验证模型，对于提升评价标准的适用性、优化修编具有重大意义。

目前，算法模型应用于拟合、预测等方面的有三类：统计模型、确定性模型和混合模型[4]。统计模型利用实测数据，通过统计方法进行建模和分析。刘昱等[5]基于主成分分析法、层次分析法和综合得分法，构建多层次多指标的模拟精度评价指标体系能够反映出与实测序列相同的水沙变化情况。然而，该类模型的准确性高度依赖于输入数据的质量，数据的误差或缺失将对预测和评估结果产生不利影响。确定性模型基于物理和数学原理[4]，其主要优点在于概念明确，可以更好地与结构性态相联系。冯凡等[6]等分别利用 3 种典型的数理统计模型(证据权模型、信息量模型和逻辑回归模型)进行上述因素与历史滑坡间的相关性分析，并应用于当地的黄土滑坡敏感性评价。准确确定性模型模拟结果需要大量真实、精确的输入数据，且确定性模型在处理未知或非线性动态行为时也存在一定的局限性。混合模型相对于统计模型和确定性模型而言，在建模时采用数值分析法来分析数据，采用优化算法来拟合计算结果[7]。洪小萍等[8]基于博弈论组合赋权 TOPSIS 算法，对 2019—2021 年 17 家充电桩上市公司的创新绩效进行综合评价，得出了政策支持和税收优惠政策激发充电桩企业创新活力这一结论。但该类模型的预测效果也受限于统计模型和确定性模型的局限，尤其是在数据量不足或数据质量差的情况下。

综上所述，本文基于《贵州省水库标准化管理评价标准》(以下简称《评价标准》)的创建结果，采用主成分分析法(PCA)计算因子离散度，解构《评价标准》，采用遗传算法(GA)对 BP 神经网络模型进行优化，找出离散度对《评价标准》的影响，基于 PCA-GA-BP 神经网络算法模型重构《评价标准》，建立一个可以指导后续贵州省水库标准化管理创建的评估解释模型，同时能对《评价标准》应用及修订提供一定的参考价值。

1　研究内容

1.1　现状问题

贵州省水利科学研究院编制的《贵州省水库标准化管理评价标准》2023 年指导 33 座水库完成标准化管理创建，《评价标准》采用扣分制，水库创建最终得分是行业各专家执行《评价标准》的结果。《评价标准》31 项评估因子为，1 工程面貌与环境、2 挡水建筑物、3 泄水建筑物、4 输(引)水建筑物、5 金属结构与机电设备、6 管理设施、7 标识标牌、8 注册登记、9 责任制、10 工程划界、11 保护管理、12 安全鉴定、13 防汛组织、14 防汛物料、15 应急预案、16 安全生产、17 雨水情测报、18 工程巡查、19 安全监测、20 维修养护、21 调度运用、22 工程效益、23 管理体制、24 标准化工作手册、25 规章制度、26 经费保障、27 精神文明、28 档案管理、29 信息化平台建设、30 自动化监测预警、31 网络安全管理。各个因子评分特征矩阵为 $\boldsymbol{X}_j=[\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \cdots, \boldsymbol{x}_i]$，$i \in 1, 2, \cdots, 33$。本文 33 座水库总体样本特征矩阵为 $\boldsymbol{D}=[\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3], \cdots, \boldsymbol{X}_j$，$j \in 1, 2, \cdots, 31$。

问题来源：《评价标准》评价内容共 31 项因子，各扣分要点已明确，但专家依然有在扣分区间内自由把控扣分程度的权力。

问题表现形式：按照《评价标准》要求评审专家应对每一项因子严格实行扣分制度，但当问题细化到因子扣分区间上时，基本取决于专家对因子的工作经验与理解程度，总体上表现为整改难、复杂程度高扣分严格；反之，扣分宽松。造成了因子重要性、所占

总分比例、扣分区间一致，但其差异化不一致，形成了差异化、同质化两类因子。

1.2　研究内容

通过需建立的现状问题与《评价标准》31 项因子间 x_i 的联系，研究差异化、同质化因子对贵州省水库标准化管理创建评估结果的影响，重构《贵州省水库标准化管理评价标准》。

1.3　研究技术路线

本文从因子间差异程度入手，以差异大小寻找划分差异性因子、同质化因子并利用算法进行研究，研究的技术路线如下。

(1)引入离散度表征因子间的差异程度。在统计学中，离散度与数据间的差异程度之间存在着直接且密切的关系。离散度是指数据集中各个数据点之间的差异程度，它可以用来衡量数据的分散程度，即离散度大差异性明显、离散度小差异性不明显(同质化)。

(2)利用方差量化离散度。在统计学中，方差(Variance)是描述一组数据与其均值(平均数)偏离程度的量化指标。方差越大，表明这组数据偏离平均数越大，即离散度越大；反之，方差越小，表明这组数据分布比较集中，各数据偏离平均数越小，即离散度越小。方差计算利用主成分分析法(PCA)进行，并划分离散度区间。

(3)解构《评价标准》划分特征矩阵。利用主成分分析法(PCA)计算因子的方差及方差累计贡献率 VCR(Variance Contribution Rate)来指定各离散区间特征矩阵成分，形成 PCA 特征矩阵 d_1、d_2、d_3、d_4。

(4)搭建 GA-BP 神经网络验证模型。利用遗传算法(GA)优化 BP 神经网络，以 PCA 特征矩阵为输入变量，搭建 GA-BP 神经网络算法模型，找出离散度大小与拟合结果的关系，从而得出差异化、同质化因子对水库创建结果的影响程度。

(5)PCA-GA-BP 神经网络重构《评价标准》验证。利用主成分分析法(PCA)剔除同质化因子重构《评价标准》，基于 PCA-GA-BP 神经网络算法模型验证重构后的《评价标准》拟合精度。

2　利用主成分分析划分离散度特征矩阵

2.1　主成分分析(PCA)法

主成分分析(PCA)法是一种通过解释方差进行数据降维和特征提取的算法，PCA 通过计算数据各个主成分对应的方差提取特征值，最后选择特征值大于或累计贡献率达到一定阈值方差贡献率 VCR(Variance Contribution Rate)来指定主成分数量，PCA 主要步骤分为 4 步。

步骤 1：数据标准化，计算 $X_j=x_1, x_2, x_3, \cdots, x_i$ 的均值 μ、方差 σ。x_i 进行数据标准化，标准化公式为公式(1)。

步骤 2：构建协方差矩阵，协方差矩阵 S(公式(2))反映了原始变量之间的相关性，其主对角线上的元素是各变量的方差，非对角线上的元素是变量之间的协方差，公式为

$$x'_i=\frac{(x_i-\mu)}{\sqrt{\sigma}} \tag{1}$$

$$S=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^{\mathrm{T}} \tag{2}$$

步骤 3：提取特征值和特征向量，对相关系数矩阵进行特征值分解，得到特征值和对应的特征向量，特征值从大到小排列，前几个特征值对应的特征向量表示主要的主成分。

步骤 4：按照累积方差贡献率确定主成分数量，根据特征值的大小，保留累计方差贡献率达到预定阈值(如 80%~95%)的特征值所对应的特征向量，方差贡献率即每个主成分方差占总方差的比例，这些特征向量即为所需的主成分。

2.3　离散度特征矩阵

利用 PCA 计算各特征向量 X_j 的方差大小以观察其离散程度，计算结果降序排列如图 1、图 2 所示。

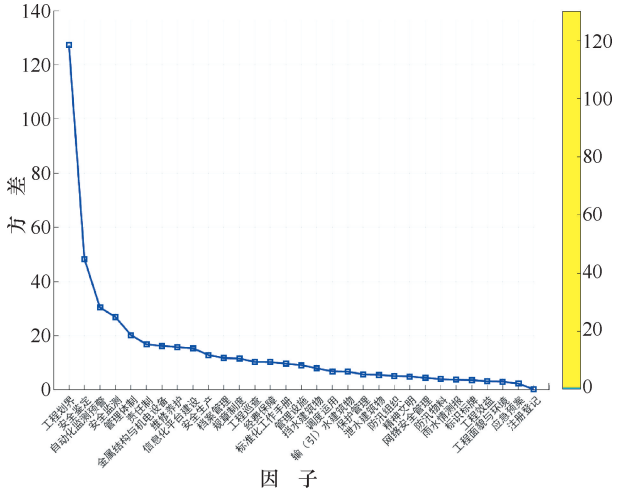

图 1　特征向量方差降序分布

从图 1、图 2 计算结果可知，利用 PCA 计算各特征向量 X_j 的方差，将因子重新降序排列解构划分成 4 个离散区间[>16]、[8~16]、[4~8]、[0~4]，由此构建 4 个离散程度逐步增大的分段特征矩阵 d_1、d_2、d_3、d_4，各特征矩阵包含的特征向量数量及排序如下

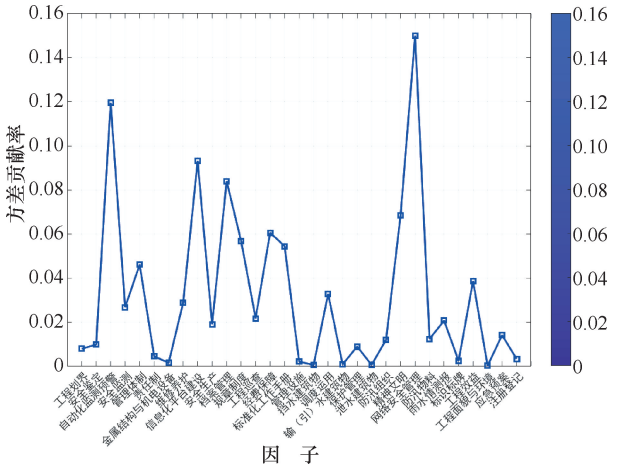

图 2　特征向量方差贡献率

$$\begin{aligned} \boldsymbol{d}_1 &= [\boldsymbol{X}_{10}, \boldsymbol{X}_{12}, \boldsymbol{X}_{30}, \boldsymbol{X}_{19}, \boldsymbol{X}_{23}, \boldsymbol{X}_9, \boldsymbol{X}_5] \\ \boldsymbol{d}_2 &= [\boldsymbol{X}_{20}, \boldsymbol{X}_{29}, \boldsymbol{X}_{16}, \boldsymbol{X}_{28}, \boldsymbol{X}_{25}, \boldsymbol{X}_{18}, \boldsymbol{X}_{26}, \boldsymbol{X}_{24}, \boldsymbol{X}_6] \\ \boldsymbol{d}_3 &= [\boldsymbol{X}_2, \boldsymbol{X}_{21}, \boldsymbol{X}_4, \boldsymbol{X}_{11}, \boldsymbol{X}_3, \boldsymbol{X}_{13}, \boldsymbol{X}_{27}, \boldsymbol{X}_{31}, \boldsymbol{X}_{14}] \\ \boldsymbol{d}_4 &= [\boldsymbol{X}_{17}, \boldsymbol{X}_7, \boldsymbol{X}_{22}, \boldsymbol{X}_1, \boldsymbol{X}_{15}, \boldsymbol{X}_8] \end{aligned} \tag{3}$$

3　基于 GA-BP 神经网络算法验证

3.1　构建 GA-BP 神经网络验证模型

3.1.1　BP 神经网络模型原理

BP 神经网络(Back Propagation Neural Network)，即反向传播神经网络，是一种基于误差反向传播算法(Back Propagation Algorithm)的多层前馈神经网络，由输入层、隐藏层 m(可以包含多个层次)和输出层组成。每一层包含若干个神经元(或称为节点、单元)，层与层之间通过权重 w_n 连接。输入层负责接收外部输入信号，隐藏层对输入信号进行非线性变换，输出层则生成最终的输出结果$\hat{y}$ 为

$$\hat{y} = sign(w_1x_1 + w_1x_1 + \cdots + w_nx_n - t) \tag{4}$$

式中，$sign$ 为符号函数，作为输出神经元的激活函数，当参数为正时输出+1，参数为负时输出−1；t 为偏置因子；x_1，x_2，…，x_n 为输入属性值。

输入链权值公式如下

$$w_j^{k+1} = w_j^k + \lambda(y_i + \hat{y}_i^k)x_{ij} \tag{5}$$

式中，w^k 是第 k 次循环后第 i 个输入链上的权值；参数 λ 为学习率；x_{ij} 是训练样本 x_i 的第 j 个属性值；w^{k+1} 是更新权值，等于旧权值 w^k 加上一个正比于预测误差 $(y - \hat{y})$ 的项。

3.1.2　构建基于 GA-BP 神经网络算法评估预测模型

利用遗传算法(GA)对 BP 神经网络进行优化，是通过模拟生物进化过程，将实际问题的变量进行编码形成染色体，随机产生一定数目的初始种群，种群每个个体是带有特征的染色体。初始种群产生后，按照优胜劣汰的原理，在每一代，根据问题域中个体的适应度大小选择个体，并借助于自然遗传学的遗传算子进行组合交叉和变异，产生出代表新的解集的种群，后代种群比前代更加适应环境，末代种群中的最优个体经过解码，可以作为问题近似最优解。构建贵州省水库标准化管理 GA-BP 神经网络评估预测模型主要步骤分为 6 步。

步骤 1 确定最优隐藏层数。设定初始隐含层数为 m，31 项“评估内容”为输入特征数据，则输入层节点数 $n=31$，按照隐含层数经验公式缩小 m 取值范围，评估总分 l 为输出层，$l=1$。再用穷举法对不同的隐含层数 m 逐一训练，选取均方误差最小、回归 R^2 值最大时的节点数 m 为最优隐含层。隐含层数经验公式见公式(4)，其中 a 为 0~10 的整数，取值如下

$$n - 1 < m < \sqrt{l + n} + a \tag{6}$$

步骤 2 染色体编码 BP 神经网络权值。将 BP 神经网络目标问题与染色体位串结构之间建立联系。根据 BP 网络结构，计算所需优化的权值和阈值的总数，设计染色体位串的长度，使其能够包含所有权值和阈值的编码。染色体编码空间采用二进制编码，利用二进制串表示变量十进制数值，输入变量 $X_i \in [a_{\min}, a_{\max}]$，若染色体编码的串长精度为 l，总则共能够产生 2^l 种不同的编码，编码精度 δ，公式如下

$$l = \sum_{i=1}^{n} l_i \tag{7}$$

$$\delta = \frac{a_{\max} - a_{\min}}{2^{li} - 1} \tag{8}$$

式中，串长精度 l 为需要的小数点后精度，数据集 X_i 值域至少可以分为 $(a_{\max} - a_{\min}) \times 10^l$ 份，值域满足 $2^{li-1} < (b_i - a_i) \times 10^l < 2^{li}$。

步骤 3 建立适应度函数与目标函数的映射关系。适应度函数是用来衡量个体优劣，度量个体适应度的函数。适应度函数值越大的个体越好，反之，适应度函数值越小的个体越差。目标函数有正有负，为此在 GA-BP 神经网络解释模型中建立适应度函数 $f(x)$ 和 BP 神经网络目标函数 $K(Y_{ij})$ 的映射关系，以较好地解决实际问题，针对本文误差最小化问题研究适应度函数一般采用公式(9)，c 为目标函数界限的保守估计值，公式如下

$$f(x) = \frac{1}{1 + c + K(Y_{ij})},\ c \geqslant 0,\ c + K(Y_{ij}) \geqslant 0 \tag{9}$$

步骤 4 约束条件处理。在 GA 遗传算法中必须对

约束条件、最大进化次数、最优种群数量进行处理，本次利用可变搜索空间法进行研究，通过明确搜索空间上、下限，获取首次循环的最优解集，确定各阶段最优解集的极值，并设置宽度 W_{width}。决策变量的搜索空间的上限、下限随着逐次循环次数 k 的增加而改变，即各阶段的搜索范围不断减小，$X_{\max 1}$、$X_{\min 1}$ 分别为第一次循环开始时决策变量搜索空间的上、下限。

步骤 5 遗传算子模拟。遗传算法中包含 3 个模拟生物基因遗传操作的遗传算子：选择(复制)、交叉(重组)和变异(突变)。选择(复制)根据个体的适应度值占全部个体适应度值之和的比例来选择个体；交叉操作的基本思想是通过对两个个体之间进行某部分基因的互换来实现产生新个体的目的；变异操作是指将个体染色体编码串中的某些基因座的基因值用该基因座的其他等位来替代，从而形成一个新的个体。

步骤 6 输出解释结果。GA-BP 神经网络解释模型输出解释结果的条件有两个：遗传操作中连续多次前后两代群体中最优个体的适应度相差在某个任意小的正数 ε 所确定的范围内；达到遗传操作的最大进化代数。满足任何一个条件，搜索结束输出解释值

$$0 < | K_{new} - K_{old} | < \varepsilon \tag{10}$$

式中，K_{new} 为新产生的群体中最优个体的适应度；K_{old} 为前代群体中最优个体的适应度。

3.2　输入假设特征矩阵及验证矩阵

特征矩阵：将本文 PCA 计算的离散特征矩阵 $\boldsymbol{d}_1$、$\boldsymbol{d}_2$、$\boldsymbol{d}_3$、$\boldsymbol{d}_4$ 作为输入变量，分别验证分析。

验证矩阵：2023 年贵州省依据《评价标准》31 项特征因子完成 33 座水库标准化管理创建，验证矩阵为各个水库创建评估结果 $\boldsymbol{Y}$，$i \in 1, 2, \cdots, 33$，$j \in 1, 2, \cdots, 31$。

$$Y = \left[\sum_{j}^{1} \boldsymbol{x}_{1j}, \ \sum_{j}^{1} \boldsymbol{x}_{2j}, \ \cdots \sum_{j}^{1} \boldsymbol{x}_{ij} \right] \tag{11}$$

3.3　参数设置方案

编程语言为 Matlab，GA-BP 神经网络解释模型最优参数实验步骤分为 4 步，步骤 1 确定一般参数. 以查文献法，根据相关研究结论，确定 70% 数据用于训练、30% 数据用于解释，训练次数设为 1 000，最大进化次数 60，交叉概率 0.8，变异概率 0.2 适合模型进行解释，反复试验 10 次，确定基础 BP 神经网络训练过程中隐藏层节点数为 15；步骤 2 确定最优种群数量。利用可变搜索空间法明确搜索空间上限 $X_{\max 1}$ 为 100、下限 $X_{\min 1}$ 为 10，设置宽度 W_{width} 为 5，逐次循环寻找最优的种群数量；步骤 4 确定最优解。建立寻优时间加权解释误差，所得最小值确定为模型解释最优解。

3.4　拟合验证

输入矩阵 PCA 主成分分析法分段的 $\boldsymbol{d}_1$、$\boldsymbol{d}_2$、$\boldsymbol{d}_3$、$\boldsymbol{d}_4$ 特征矩阵，每个特征向量包含的样本数量为 33。按照参数设置方案各输入矩阵最优种群数量及最优解的计算结果如图 3—图 6 所示及表 1 所列。

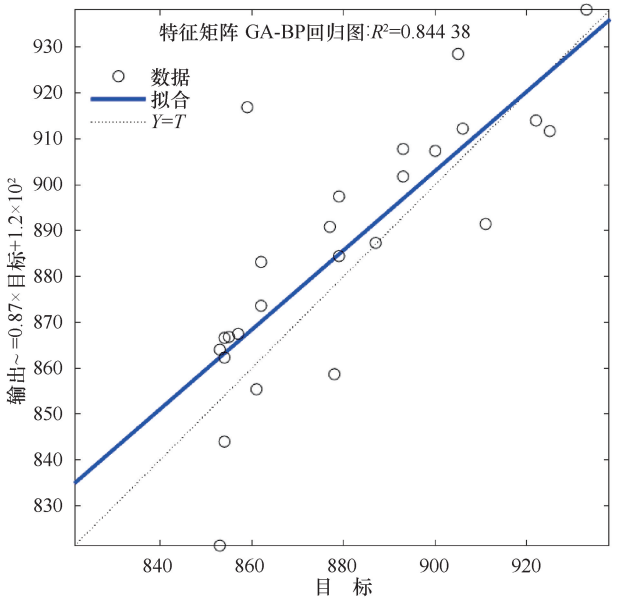

图 3　$\boldsymbol{d}_1$ 特征矩阵拟合 R^2 结果

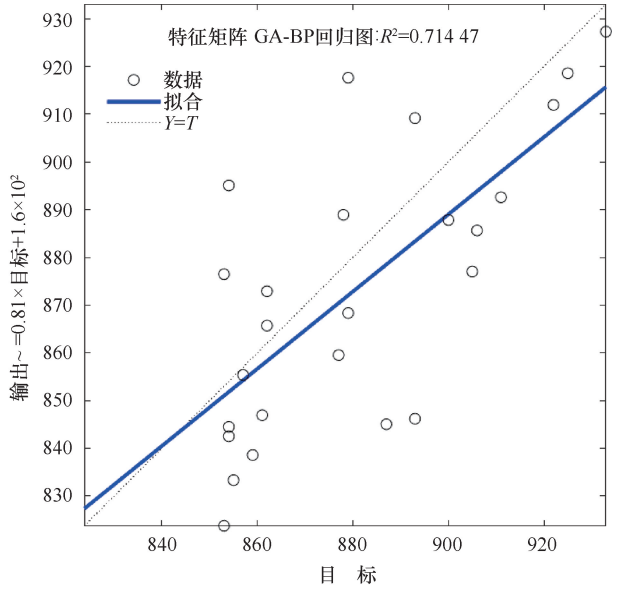

图 4　$\boldsymbol{d}_2$ 特征矩阵拟合 R^2 结果

RMSE、*MAE* 值越小，回归系数 R^2 越趋近于 1，模型拟合效果越好。从拟合结果可知，按照离散程度从大到小划分的特征矩阵 $\boldsymbol{d}_1$、$\boldsymbol{d}_2$、$\boldsymbol{d}_3$、$\boldsymbol{d}_4$ 其拟合精度逐步下降；$\boldsymbol{d}_3$、$\boldsymbol{d}_4$ 拟合回归系数 R^2 仅为 0.396、0.422，同质化因子特征矩阵无法与水库创建评估结果 $\boldsymbol{Y}$ 拟合精度低；特征矩阵所包含的特征因子数量

表 1　特征矩阵验证拟合结果

输入矩阵	特征因子数量/个	最优种群/个	离散区间	*MAE*	*RMSE*	回归系数
$\boldsymbol{d}_1$	7	30	>16	0.121	0.151	0.844
$\boldsymbol{d}_2$	9	35	8~16	0.196	0.238	0.714
$\boldsymbol{d}_3$	9	35	4~8	0.224	0.293	0.422
$\boldsymbol{d}_4$	6	25	0~4	0.238	0.284	0.396

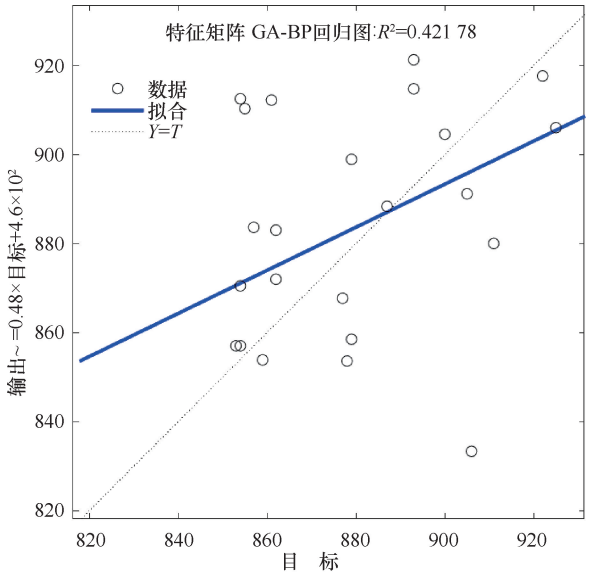

图 5　$\boldsymbol{d}_3$ 特征矩阵拟合 R^2 结果

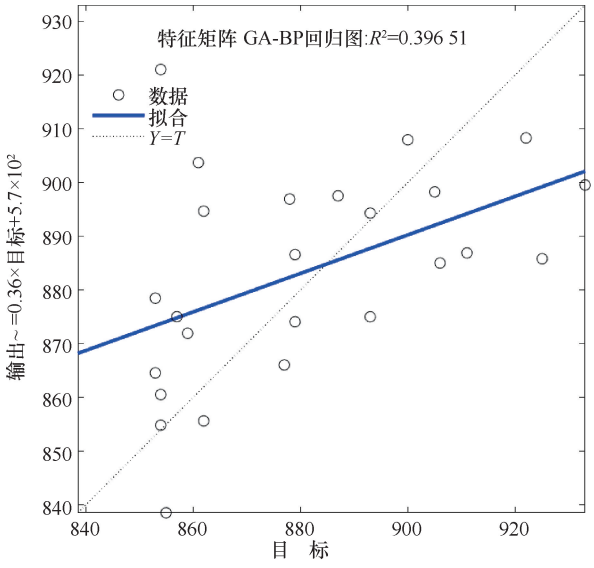

图 6　$\boldsymbol{d}_4$ 特征矩阵拟合 R^2 结果

多少与拟合精度无相关性，与特征矩阵离散程度呈正相关，特征矩阵离散程度越高，拟合精度越高。

综上所述，说明同质化因子无法与水库创建评估结果 $\boldsymbol{Y}$ 建立良好的拟合关系，影响较低。差异化因子对水库创建评估结果 Y 影响较大。

4　基于 PCA-GA-BP 神经网络重构《评价标准》验证

剔除离散度小同质化的因子，同时降低因子间的互相影响是重构《评价标准》主要办法。利用 Pearson 相关分析进行降低因子间的互相影响，重构后的《评价标准》利用 PCA 计算方差累计贡献率 VCR 验证主成分数量，最后输入《评价标准》重构特征矩阵，基于 PCA-GA-BP 神经网络算法模型验证重构特征矩阵拟合精度。

4.1　因子分析

鉴于差异化、同质化因子与创建评估结果拟合精度的验证结论，说明创建评估结果是专家对因子差异化评估的结果。为此，同时对差异化特征矩阵 $\boldsymbol{d}_1$、$\boldsymbol{d}_2$，同质化特征矩阵 $\boldsymbol{d}_3$、$\boldsymbol{d}_4$ 进行 Pearson 相关分析，利用 Pearson 相关性可以实现因子的两两对比，两个因子间相关性越强，相关系数越趋近于 1，方格色彩趋近黄色；反之相关系数越低，越趋近于蓝色。计算结果如图 7、图 8 所示。

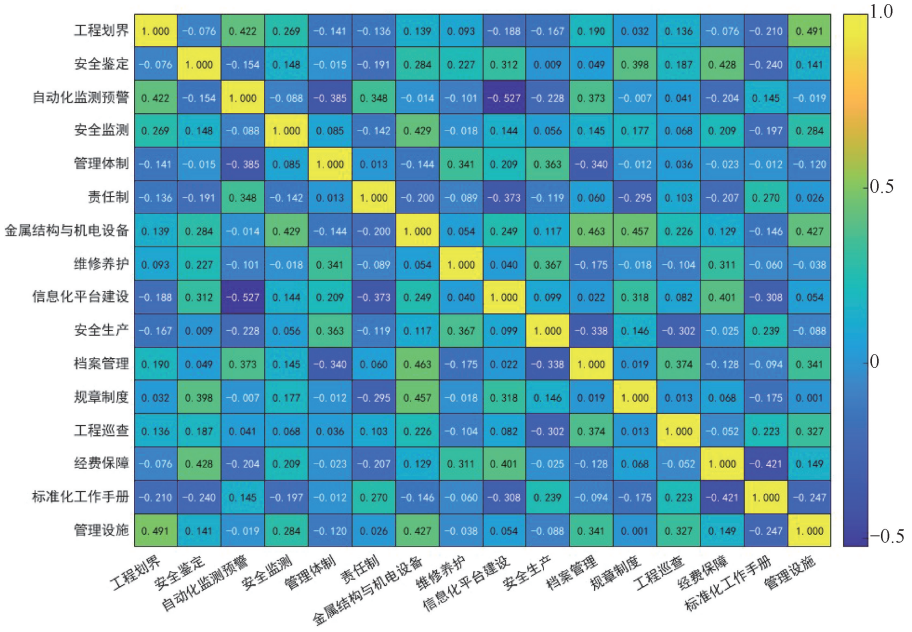

图 7　差异化因子 Pearson 相关性热图

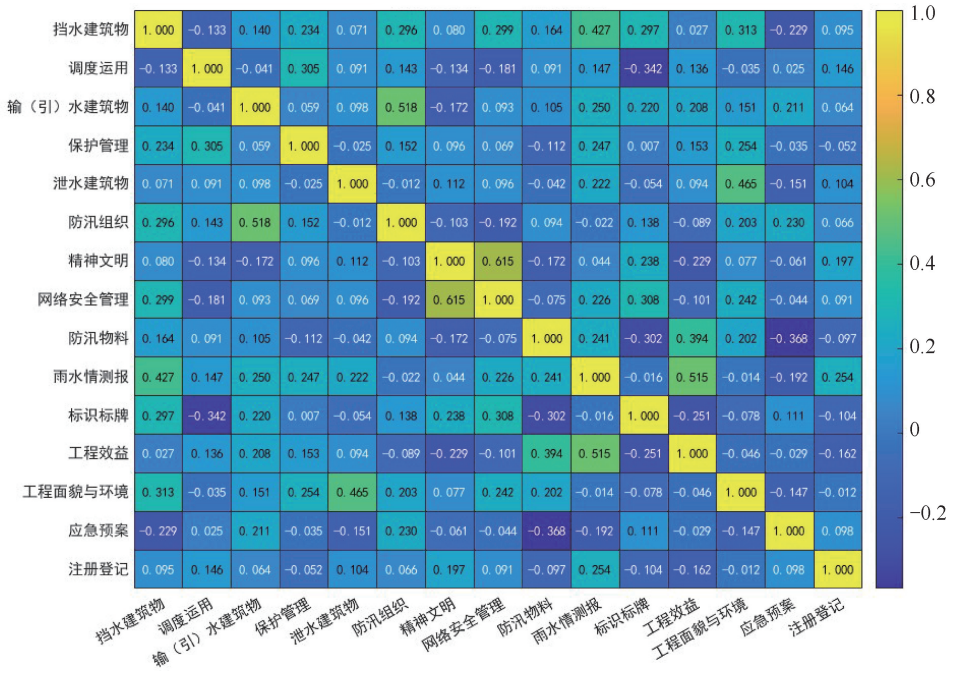

图 8　同质化因子 Pearson 相关性热图

4.1.1　差异化因子分析

据图 7 计算结果，差异化因子间相关性较低，均在 0.4 以下，说明差异化因子间相互影响较小。鉴于其拟合精度，为保证主要信息量，16 项差异化因子全部纳入重构《评价标准》。

4.1.2　同质化因子分析

据图 8 计算结果，为尽可能保留《评价标准》主要信息，将相关性 0.5 以上的 6 项同质化因子按离散度大小排列，用离散度相对大的 3 项因子去替代解释离散度小的因子；图 8 中有 5 项因子与其他因子重复相关，利用这 5 项去替代解释另外的因子，剔除重复选中的因子后，从同质化因子中选出 5 项纳入重构《评价标准》。

4.2　重构《评价标准》验证

《评价标准》重构矩阵 $\boldsymbol{V}$ 包括 16 项差异化因子、5 项同质化因子。重构矩阵包含的因子序列 $\boldsymbol{V}=\boldsymbol{X}_{10}$，$\boldsymbol{X}_{12}$，$\boldsymbol{X}_{30}$，$\boldsymbol{X}_{19}$，$\boldsymbol{X}_{23}$，$\boldsymbol{X}_{9}$，$\boldsymbol{X}_{5}$，$\boldsymbol{X}_{20}$，$\boldsymbol{X}_{29}$，$\boldsymbol{X}_{16}$，$\boldsymbol{X}_{28}$，$\boldsymbol{X}_{25}$，$\boldsymbol{X}_{18}$，$\boldsymbol{X}_{26}$，$\boldsymbol{X}_{24}$，$\boldsymbol{X}_{6}$，$\boldsymbol{X}_{21}$，$\boldsymbol{X}_{4}$，$\boldsymbol{X}_{13}$，$\boldsymbol{X}_{27}$，$\boldsymbol{X}_{17}$，累计 21 项，据特征向量方差贡献率计算结果显示，重构特征矩阵累计方差贡献率达 83.16%，方差解释度完全满足要求。以重构矩阵 $\boldsymbol{V}$ 为输入变量与验证矩阵 $\boldsymbol{Y}$ 进行拟合，同时输入重构前总体样本矩阵 $\boldsymbol{D}$ 进行对比验证，拟合结果如图 9—12 所示。

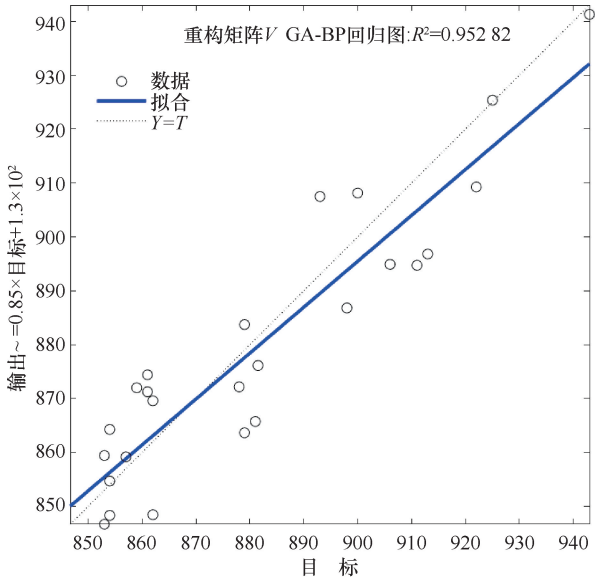

图 9　重构矩阵拟合 R^2 结果

RMSE、*MAE* 值越小，模型的预测误差越小。利用遗传算法 GA 改进 BP 神经网络模型后，各特征矩阵的拟合结果误差均在 0.3 以下，特别是 $\boldsymbol{D}$、$\boldsymbol{V}$ 矩阵误差均下降至 0.15 以下，拟合可信度高。根据拟合结果，可以看出重构《评价标准》其因子组成的特征矩阵 $\boldsymbol{V}$ 与水库创建评估结果实际值的拟合效果较好，其拟合精度为 0.953，优于 $\boldsymbol{d}_1$、$\boldsymbol{d}_2$、$\boldsymbol{d}_3$、$\boldsymbol{d}_4$ 及总样本矩阵 $\boldsymbol{D}$。

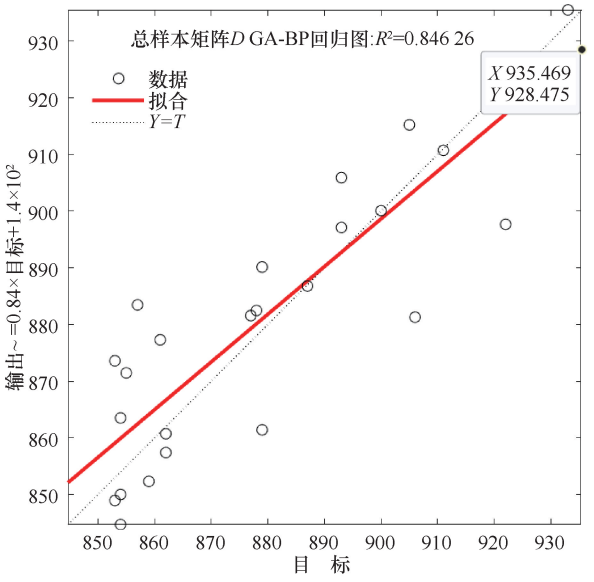

图 10　总样本矩阵拟合 R^2 结果

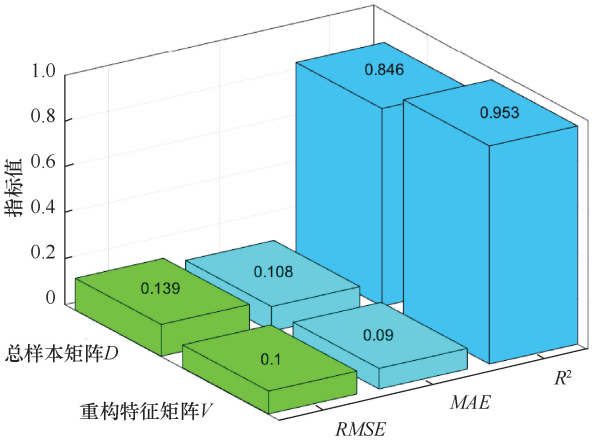

图 11　重构矩阵与总样本矩阵指标对比

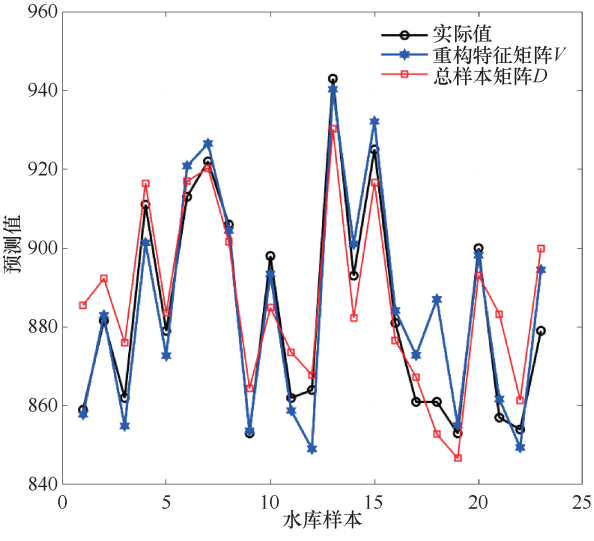

图 12　重构矩阵、总样本矩阵预测值与实际值对比

5　结　论

对标对表评分创建工作已经成为水利行业高质量发展的主要方式，这种形式不可避免存在专家在扣分区间内自由把控扣分程度的现象，针对所造成的因子重要性、所占总分比例、扣分区间一致，但其差异化不一致的问题，利用主成分分析法(PCA)以离散度作为表征指标，重构《评价标准》建立特征矩阵作为输入变量，利用 GA 遗传算法改进 BP 神经网络进行拟合验证，形成的主要结论如下。

(1)因子的离散程度直接影响创建评估结果，因子离散程度越高，越能代表创建实际情况，专家在采用扣分制的《评价标准》时应当严格扣分，体现因子差异化。

(2)剔除 10 项同质化因子后，重构《评价标准》的拟合精度 R^2 高达 0.953，较重构前提升了 0.107，精度上具有优越性，对后续《评价标准》的修编提供了一定的参考价值，也给贵州省需要创建标准化管理的水库提供了重点创建方向。

(3)引入遗传算法(GA)进行 BP 神经网络优化，构建了基于 GA-BP 神经网络算法拟合模型，通过改进目标函数与模型核函数的映射关系，避免了回归模型容易陷入局部最优的共性问题，很好地捕捉了特征矩阵 $\boldsymbol{X}_j$ 中的基本规律，提升了模型在数据集上的泛化能力、在精度上的优越性。

参考文献：

[1]　王伟立. 水利工程中水文水资源标准化管理研究[J]. 水上安全，2024(16)：34-36.
[2]　闫彭彭. 基于行业标准的水闸安全评价方法研究[J]. 水科学与工程技术，2024(4)：53-56.
[3]　赵勇，李溦，何凡，等. 中国水资源短缺标准与分区评价[J]. 中国水利，2024(15)：13-19.
[4]　赵二峰. 大坝安全的监测数据分析理论和评估方法[M]. 南京：河海大学出版社，2018.
[5]　刘昱，于坤霞，李鹏，等. 黄河中游典型流域水文统计模型精度集合评价[J]. 人民黄河，2023，45(4)：20-27.
[6]　冯凡，唐亚明，卢全中，等. 数理统计模型在黄土滑坡敏感性评价中的应用[J]. 甘肃科学学报，2019，31(3)：68-76.
[7]　顾冲时，吴中如. 大坝与坝基安全监控理论和方法及其应用[M]. 南京：河海大学出版社，2006.
[8]　洪小萍，孔莉. 中国充电桩行业创新绩效评价研究：基于博弈论组合赋权 TOPSIS 算法的分析[J]. 价格理论与实践，2024(4)：34-39.

(责任编辑　王　璐)