

文章编号: 2095-2163(2020)05-0152-04

中图分类号: TP399

文献标志码: A

# 基于机器学习的钻井液体系优选方法分析

董萃莲, 董海峰, 闫红丹

(西安石油大学 计算机学院, 西安 710065)

**摘要:** 由于钻井液体系种类繁多, 各种各样的地质层环境或多或少都会影响钻井液体系的选择, 石油工程师在如此多的钻井液体系中选择出一种合适的钻井液效率较低。故考虑将机器学习应用于钻井液体系优选中, 本文简单介绍机器学习在钻井液体系优选中的方法理论及采用的模型。首先介绍对初始数据集处理的3种方法的基本理论, 分别是主成分分析、灰色关联度分析及奇异值分解; 然后介绍这三种方法分别与BP神经网络结合的优选模型; 最后对本文提出的钻井液体系优选方法进行总结概括。

**关键词:** 机器学习; 钻井液体系优选; 主成分分析; 灰色关联度分析; 奇异值分解; BP神经网络

## Analysis of optimization method of drilling fluid system based on machine learning

DONG Cuilian, DONG Haifeng, YAN Hongdan

(College of Computer Science, Xi'an Shiyou University, Xi'an 710065, China)

**[Abstract]** Due to the wide variety of drilling fluid systems, various geological environments can more or less affect the selection of drilling fluid systems, and it is less efficient for petroleum engineers to choose a suitable drilling fluid among such a large number of drilling fluid systems. Therefore, machine learning is considered to be applied in the optimization of drilling fluid system. This paper briefly introduces the method theory and model of machine learning in the optimization of drilling fluid system. Firstly, the basic theory of three methods for initial data set processing is introduced, namely principal component analysis, grey relational degree analysis and singular value decomposition. Then the optimal model combining these three methods with BP neural network is introduced. Finally, the optimization method of drilling fluid system proposed in this paper is summarized.

**[Key words]** machine learning; optimization of drilling fluid system; principal component analysis; grey correlation analysis; singular value decomposition; BP neural network

## 0 引言

钻井液是为了保证安全、快速、优质钻井工程的关键性因素, 不同的地质环境可能对钻井液的选择有所不同, 以往通过石油工程师在一定的实验分析下, 通过经验等方式得出钻井液种类<sup>[1]</sup>。钻井液种类繁多, 形成了各种各样的体系。由于地质环境的复杂性往往会给石油工程师的选择带来各种各样的问题, 为了避免石油工程师在钻井液种类选择时所遇到的种种问题, 利用机器学习来对此问题进行探讨。

机器学习近年来发展迅猛, 越来越多的算法被学者们应用于各个领域, 比如: 人脸识别、语音识别、天气预测、股票预测、房价预测等等。而在预测优选领域, BP神经网络模型也越来越成熟, 被广泛应用于各种预测中。但在实际应用中, 由于各种问题所需要训练的模型输入变量有所不同, 有些研究对象的输入特征变量可能比较多或者该类型问题比

较复杂等, 如果还单一性的采用BP神经网络模型来进行预测研究, 可能会出现某些不可控的问题, 同时也会增加模型的预测难度、降低预测准确性, 使得模型预测性能大大地降低<sup>[2]</sup>。因此, 需要对此类问题进行前期的相应处理, 借用组合模型来避免此类问题的发生。

## 1 问题描述

钻井在探测与开掘石油、天然气资源时是至关重要的环节, 同时也是勘测和开发石油天然气的一个重要手段。钻井液在安全、快速和优质钻井中起着重要的作用。由于油气层数据信息非常复杂, 不确定性较高, 某些钻井液对地质层参数的影响还不能用精确数值来量化, 具有一定的模糊性和非线性。通常情况下, 石油工程师会在一定实验分析的基础上, 采用经验公式等方法来进行分析解释, 效率低、预测结果可靠性误差大。为了避免这种方法存在的缺陷问题, 通过之前选择钻井液体系的历史数据信

**作者简介:** 董萃莲(1993-), 女, 硕士研究生, 主要研究方向: 管理信息系统; 董海峰(1965-), 男, 博士, 高级工程师, 副教授, 硕士生导师, 主要研究方向: 信息管理系统、办公自动化、数据库开发等; 闫红丹(1993-), 女, 硕士研究生, 主要研究方向: Web服务推荐、QoS预测、软件工程。

收稿日期: 2020-01-15

息进行分析,引入机器学习的方法,来构建钻井液体系优选的模型,从而在较少和一定模糊数据的情况下完成钻井液体系的优选。

在对钻井液体系进行优选时,不仅要考虑保护储集层,还要考虑地质层、工程要求、环保等各方面的因<sup>[3]</sup>。主要考虑这些影响因素:井别、井眼最高温度、最高地层孔隙压力、井斜角、储层孔隙平均半径、储层水敏性(水敏)、储层盐敏性(盐敏)、储层碱敏性(碱敏)、储层岩石润湿性、储层PH值、井眼盐层及盐水层状况(盐侵度)、井眼石膏层及水泥层状况、地层粘土含量、地层粘土膨胀率、环境类别、钙离子浓度,利用以往数据集来作为训练集,训练确定优选模型来进行后续优选,输入上述16种特征参数,通过训练好的优选模型得出最终的相应优选钻井液体系种类。从而达到解决钻井液优选的问题。

## 2 算法分析

由于钻井液体系优选的参数达到16个,考虑到BP神经网络优选模型高维输入参数可能使得优选模型的规模过于庞大,引起训练效果低下、准确度低、训练时长剧增等问题。故考虑对其先进行前期的处理,分析以下3种方法来避免此种问题的发生。

### 2.1 主成分分析法

主成分分析主要通过降维的思想,将原始的特征集,通过其矩阵之间的相关关系找出各个特征不相关的、可代表所有特征的、比原始特征少的特征项<sup>[4-5]</sup>。从而达到用少数特征项全权代表全体原始特征项的目的。

在本文钻井液体系的优选中可构建模型为:假设有 $x$ 个样本,每个样本有16个特征项,分别为 $CS_1, CS_2, CS_3, \dots, CS_{16}$ ,则原始特征项矩阵为(1):

$$CS = \begin{bmatrix} CS_{11} & CS_{21} & \dots & CS_{x1} \\ CS_{12} & CS_{22} & \dots & CS_{x2} \\ \vdots & \vdots & \dots & \vdots \\ CS_{16} & CS_{26} & \dots & CS_{x16} \end{bmatrix} = (CS_1, CS_2, \dots, CS_{16}). \quad (1)$$

由上特征项矩阵,可得到该体系综合特征项线性组合。

$$\begin{cases} T_1 = A_{11}CS_{11} + A_{12}CS_{12} + \dots + A_{16}CS_{16}, \\ T_2 = A_{21}CS_{21} + A_{22}CS_{22} + \dots + A_{26}CS_{26}, \\ \dots & \dots & \dots & \dots \\ T_x = A_{x1}CS_{x1} + A_{x2}CS_{x2} + \dots + A_{x16}CS_{x16}. \end{cases}$$

线性组合式中的组合系数 $A$ 满足该项组合系数平方和为1。

主成分分析算法的具体步骤如下:

假如有 $x$ 条数据,有16个特征项

(1)将原始数据集组成 $x$ 行16列的矩阵 $CS$ ;

(2)将 $CS$ 的每一列(即一个特征项属性)进行均值化操作(在此采用零均值化),也就是减去该列的平均值;

(3)计算出协方差矩阵;

(4)计算出协方差矩阵的各个特征值及相对应的特征向量 $R$ ;

(5)把特征向量 $R$ 按照相对应的特征值的大小顺序由上至下组成矩阵,通过实验观察选取合适的前 $M$ 项,组成新矩阵 $CSS$ ;

(6) $CSS$ 即为降维处理后的数据集。

### 2.2 奇异值分解法

奇异值分解是通过数学中矩阵的因式分解,通过求出奇异值个数和奇异值的同时,得到一个重新排列的,可用于代替原始数据矩阵的相似矩阵,将该数据集的奇异值按大小顺序排列,得出其重要程度表,从而根据实际需求而确定留下重要程度大的特征项,舍去不重要的特征参数,从而达到减少原始数据集项的目的,找出原始数据集中的主要影响项<sup>[9-10]</sup>。

奇异值分解算法具体步骤如下:

(1)获取矩阵 $CS$ (即需要降维的原始数据集);

(2)计算出 $CS^TCS$ 和 $CS CS^T$ ;

(3)计算出 $CS^TCS$ 和 $CS CS^T$ 的特征向量 $\lambda_i$ 与 $\lambda_j, v_i$ 与 $u_i$ ;

(4)依据原始正交基映射后还是正交基的性质,得出 $CSv_i = \sigma_i u_i$ ,由此可求出奇异值 $\sigma_i$ ;

(5)根据 $CS = U\Sigma V^T$ ,可以得到 $CS$ 的奇异值分解, $U\Sigma$ 即为所求的数据集;

(6)根据奇异值的大小,通过实验观察,确定需要留下的奇异值个数,即可得到相应的降维数据集。

### 2.3 灰色关联度分析法

灰色关联度分析法是通过确定性数据集列(即结果)和若干个比较数据列(即特征参数项)的几何形状的相似程度来决断它们之间的紧密关系,从而找出它们之间的关联程度,该方法一般用于分析、研究各个特征因素对结果的影响程度大小,在此将其应用于寻找16个特征参数影响钻井液体系选择结果的影响程度之中,从而确定16个特征参数中最重要且最终保留,用于进行后续训练的特征参数<sup>[13~14]</sup>。

灰色关联度分析算法步骤如下:

(1)确定分析数列:确定反映钻井液体系优选

的参考数列和影响钻井液体系选择的比较数列。

①参考数列(即为钻井液选择结果项);

②比较数列(即为16个特征参数项)

$$Z_i = Z_i(k) \mid k = 1, 2, \dots, n; i = 1, 2, \dots, m.$$

(2)对变量进行无量纲化操作,通过对比较数列进行均值化处理,达到无量纲化的目的:

$$z_i(k) = z_i(k) \div \bar{z}_i, k = 1, 2, \dots, n; i = 0, 1, 2, \dots, m.$$

(3)根据公式(2)计算出关联系数。

$$\zeta_i(k) = \frac{\min_i \min_k |g(k) - z_i(k)| + \rho \max_i \max_k |g(k) - z_i(k)|}{|g(k) - z_i(k)| + \rho \max_i \max_k |g(k) - z_i(k)|}.$$

(2)

在此将  $\Delta_i(k) = |g(k) - z_i(k)|$ 。  $\rho$  为分辨系数,一般取值为0.5,将公式(2)化简为公式(3):

$$\zeta_i(k) = \frac{\min_i \min_k (\Delta_i(k)) + 0.5 \max_i \max_k \Delta_i(k)}{\Delta_i(k) + 0.5 \max_i \max_k \Delta_i(k)}.$$

(3)

(4)根据公式(4)计算出关联度。

$$R_i = \frac{1}{n} \sum_{k=1}^n \zeta_i(k), k = 1, 2, \dots, n.$$

(4)

(5)对关联度  $R$  进行排序,依据排序大小找出对应的重要项,确定留下多少项特征参数。

## 2.4 BP神经网络优选

BP神经网络的主要思想是在输入历史数据集样本之后,用反向传播的算法思想对BP神经网络的权值和偏差进行反复多次性的调整训练<sup>[15]</sup>,最终使得到的向量同用户希望的向量尽可能的接近,当学习模型输出的误差小于事先指定的误差时,则训练结束,保存此时的权值和偏差值,以备后续的训练学习使用<sup>[16]</sup>。

神经网络模型主要分为3部分:输入层、隐含层和输出层。在此处使用的神经网络模型为16个输入节点,1个输出结点,隐含层借用经验公式(5)来确定,  $N$  为输入神经元的个数,  $M$  为输出层神经元的个数,  $A$  为  $[1, 10]$  之间的常数。

$$L = \sqrt{N + M} + A.$$

(5)

BP神经网络优选算法的主要步骤如下:(算法分为两大主干部分)

(1)输入信号的前向传播,输入数据通过输入层传至隐含层,最后到达输出层。

(2)误差反向传播,从输出层传至隐含层,在从隐含层传至输入层,依次调整隐含层至输出层的权重和偏置量,输入层至隐含层的权重和偏置量。

## 3 优选模型

对于钻井液体系优选,主要采用如下算法模型思维图所示方法进行构建模型,如图1所示。

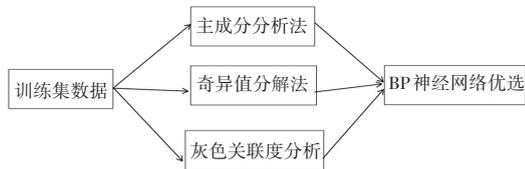


图1 算法模型思维图

Fig. 1 Algorithmic model mind map

对于钻井液体系优选的研究主要采用如算法模型思维图所示的方式进行优选模型的构建,采用3种方式,分别是主成分分析法+BP神经网络优选、奇异值分解法+BP神经网络优选以及灰色关联度分析+BP神经网络优选。具体的模型训练过程如下钻井液体系优选流程图,如图2所示。

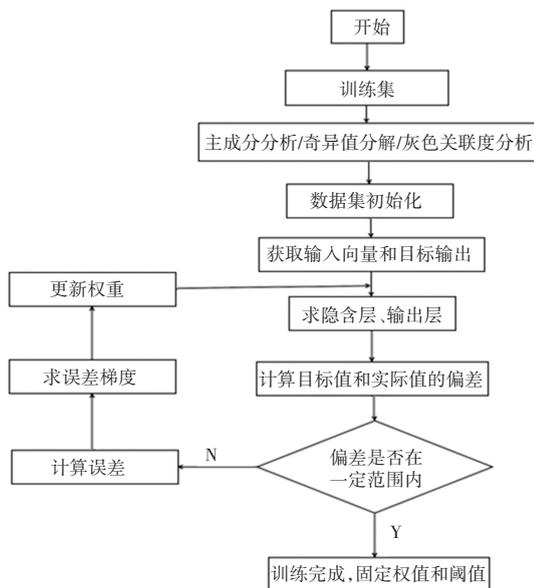


图2 钻井液体系优选流程图

Fig. 2 Flow chart of drilling fluid system optimization

## 4 结束语

本文主要研究了如何将机器学习应用于钻井液体系优选中,以解决石油工程师选取钻井液时所遇到的问题,由于钻井液的选择涉及的影响因素颇多,故在此借用三种方法对历史钻井液体系数据集进行处理,结合实验将数据集处理,得到最合适的结果,将处理后的数据集用作钻井液体系优选的训练集,从而解决了钻井液优选的问题。经实验证明,利用该方式进行优选可以提高训练的准确度,减少优选训练时间,减轻了石油工程师繁杂的工作量,同时为快速、安全、优质钻井提供保障。

(下转第157页)