

文章编号: 1000-0747(2020)02-0357-09 DOI: 10.11698/PED.2020.02.14

# 基于人工神经网络的注水开发油藏产量预测

NEGASH Berihun Mamo, YAW Atta Dennis

(马来西亚石油大学石油工程系, 斯里依斯干达 32610, 马来西亚)

**摘要:** 针对传统注水开发油藏产量预测方法存在的问题, 提出了基于人工神经网络的预测模型, 阐述了模拟工作流程, 并进行了算例分析。提出了基于流体物理学和测量数据随机组合的特征提取方法, 以提高模型的预测效果。优选贝叶斯正则化算法作为模型的训练算法, 该算法一般耗时较长, 但能对产油量、产气量、产水量等嘈杂数据集进行良好泛化。通过计算均方误差及决定系数、绘制误差分布直方图及模拟数据-验证数据交会图等方式进行模型评价。用 90% 的历史数据训练、验证、测试目标模型结构, 然后用其余 10% 数据进行盲测。研究表明, 提出的流体产量预测模型决定系数超过 0.9, 模拟结果与实际数据吻合程度高, 输入信息少, 计算成本低。图 20 表 8 参 34

**关键词:** 人工神经网络; 机器学习; 特征提取; 贝叶斯正则化算法; 产量预测; 注水开发

**中图分类号:** TE328 **文献标识码:** A

## Artificial neural network based production forecasting for a hydrocarbon reservoir under water injection

NEGASH Berihun Mamo, YAW Atta Dennis

(University Teknologi PETRONAS, Petroleum Engineering Department, 32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia)

**Abstract:** As the conventional prediction methods for production of waterflooding reservoirs have some drawbacks, a production forecasting model based on artificial neural network was proposed, the simulation process by this method was presented, and some examples were illustrated. A workflow that involves a physics-based extraction of features was proposed for fluid production forecasting to improve the prediction effect. The Bayesian regularization algorithm was selected as the training algorithm of the model. This algorithm, although taking longer time, can better generalize oil, gas and water production data sets. The model was evaluated by calculating mean square error and determination coefficient, drawing error distribution histogram and the cross-plot between simulation data and verification data etc. The model structure was trained, validated and tested with 90% of the historical data, and blindly evaluated using the remaining. The predictive model consumes minimal information and computational cost and is capable of predicting fluid production rate with a coefficient of determination of more than 0.9, which has the simulation results consistent with the practical data.

**Key words:** neural networks; machine learning; attribute extraction; Bayesian regularization algorithm; production forecasting; water flooding

引用: NEGASH B M, YAW A D. 基于人工神经网络的注水开发油藏产量预测[J]. 石油勘探与开发, 2020, 47(2): 357-365.

NEGASH B M, YAW A D. Artificial neural network based production forecasting for a hydrocarbon reservoir under water injection[J]. Petroleum Exploration and Development, 2020, 47(2): 357-365.

## 0 引言

现有的大部分油气产量预测方法都要求对油气田的储集层和流体性质进行综合描述, 需要花费大量时间、精力去分析所研究的油气田<sup>[1]</sup>, 预测结果准确性主要取决于模型输入数据, 包括地质资料、岩石物理数据和储集层特征参数等<sup>[2]</sup>。但是, 鉴于储集层的各向异性和非均质性, 数据采集和岩石、流体性质表征难度很大<sup>[3]</sup>。此外, 油气藏模拟的每一步都有诸多不确定性, 既有储集层输入数据的不确定性, 也有实验研究的不

确定性, 会极大地影响油气藏生产动态的预测。油藏工程师一般采用历史拟合法, 运用物质平衡概念, 使用硬数据来验证模型, 但不确定程度依然很高, 有可能影响流体产量预测<sup>[4-5]</sup>。因此, 机器学习方法受到油气行业的关注, 尤其是在快速评估和预测产量方面<sup>[6]</sup>。近年来, 有学者将注入量作为输入数据、产出量作为输出数据, 利用机器学习方法来推断注水开发油藏的注采关系<sup>[7-8]</sup>。

油气行业的主要挑战之一是数据和管理。面对大量数据, 工程技术人员有时难免会忽略有用信

息<sup>[9]</sup>。因此,PDA(油气数据分析)技术受到广泛关注,该技术利用在油气行业中收集的数据来分析、模拟、优化生产运营<sup>[10]</sup>。大数据分析是数据科学的一个分支,涵盖人工智能、数据挖掘、机器学习和模式识别<sup>[11]</sup>。机器学习研究的是计算机基于数据进行学习的能力<sup>[12]</sup>,用于从数据中提取预测模型。机器学习有两种主要类型,分别是监督学习和非监督学习。监督学习通过对数据进行分类或标记开展学习,该算法通过分析训练过的数据得到能够根据同种特征向量预测新算例的模型<sup>[13]</sup>。非监督学习不标注所有输入数据,而监督学习不但标注数据而且对输入、输出数据绘图<sup>[14]</sup>。人工神经网络是油气行业中应用最广泛的机器学习方法之一,它是由生物神经网络启发的一种数学结构形式,用于依赖大量输入数据的函数的近似<sup>[15]</sup>。人工神经网络的名称源自其骨架结构,在其骨架结构中,通过选择输入属性来优化具有多个隐藏层的目标函数。近年来,有研究者应用人工神经网络根据孔隙度、渗透率、饱和度等地质参数预测原油产量,但由于各参数之间关系复杂,预测过程耗时长、费用高<sup>[16]</sup>。人工神经网络也在油气行业的其他领域得到应用<sup>[17-21]</sup>。然而,应用人工神经网络并不是为了取代常规模拟手段及工作流程,而是将其作为一种辅助工具,以根据油气田数据得到更多信息<sup>[22]</sup>。

本文提出模拟油、气、水产量与测试压力、注采数据间关系的人工神经网络,使用特别设计的特征来训练人工神经网络,建立注水开发油藏产量预测模型。进行特征提取时,首先利用实测数据建立能提供信息而不冗余的派生值(特征),以促进后续的学习和泛化步骤。采用的特征提取方法以流体物理学和测量数据随机组合为基础。设计了几个特征,并采用“蛮力”手段在最终的模型中去掉了对模型预测效果没有太大影响的特征。本文将对提出的预测模型进行详细阐述并进行算例分析。

## 1 方法概述

在人工神经网络中,神经信息以权重和偏差的形式储存,权重高的输入项在模型中的作用大<sup>[23]</sup>。经典机器学习方法中将1个数据集分为3个数据子集,即训练数据子集、验证数据子集和测试数据子集<sup>[24]</sup>。训练数据子集用于计算梯度、更新网络权重和偏差。训练过程中要监测验证数据子集的误差。和训练数据子集的误差一样,验证数据子集的误差一般会在训练的初始阶段减小。当网络开始过拟合数据时,验证数据

子集的误差通常又开始增大。网络权重和偏差在验证数据子集误差最小时被保存。测试数据子集用于评价经过训练和验证的模型。在训练过程中不使用测试数据子集误差,但是要用它对不同模型进行比较。可以在整个训练过程中绘制测试数据子集误差图,如果测试数据子集误差达到最小时的迭代次数与验证数据子集误差达到最小时的迭代次数相差很大,那么对数据集的划分可能不合理。本文使用马来盆地1个真实油藏14年的生产数据训练、验证、测试人工神经网络。将数据分成4部分:80%的数据用于训练,5%的数据用于验证,5%的数据用于测试,其余10%的数据用于盲测模型。图1给出了使用人工神经网络模拟注水开发油藏流体产量的工作流程。

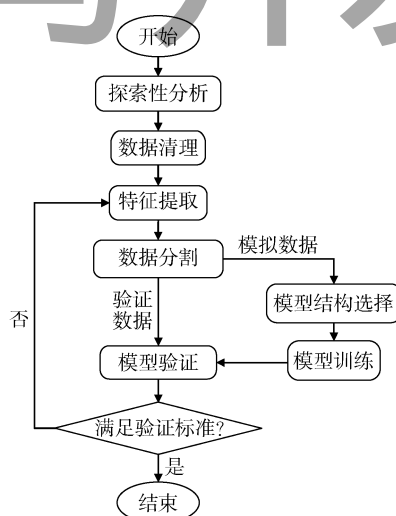


图1 注水开发油藏流体产量模拟工作流程

### 1.1 探索性分析

探索性分析即对数据进行收集和分析。数据的主要来源是Pi-ProcessBook软件,该软件可获得安装在生产井和注入井处的发送器发送的实时数据。利用Pi-ProcessBook软件对这些数据进行分析,并提取某些特征的正态分布直方图。可从分布图中识别可能的离群值,还可识别测量误差。在统计学中,离群值是指不属于某一数据群的数据点,是远离其他值的异常观测值,是分散于其他正常数据之外的观测值。如果数据分布为近似正态分布,那么约68%的数据值与均值的偏差在1个标准差范围内,约95%的数据值与均值的偏差在2个标准差范围内,约99.7%的数据值与均值的偏差在3个标准差范围内。因此,用 $3\sigma$ 规则来检测离群值,根据该规则,任何与均值的偏差达到或超过3个标准差的数据值都被视作离群值。

## 1.2 数据清理

数据清理即发现数据集中的错误以及剔除或修改数据集中的噪声数据。进行数据清理时,首先要了解数据集,例如了解注入井的注入压力与流量之间的关系,然后利用所了解的关系来计算流量计关闭时的注入量,并剔除不准确的数据。又比如,研究海上油田生产日报,确定重大事件和生产设施的关停时间等。本文中,通过使用不同来源的数据进行交叉校验,重新分配、微调日产出量和日注入量数据。另外,用基于窗口的检测方法剔除、替换离群数据,即先将数据划分到固定大小的窗口中,再检测、替换离群数据。

## 1.3 特征提取与扩展

特征提取与扩展即使用现有数据集来创建新的输入特征,使机器学习起作用并提高其性能。López-Iñesta 等<sup>[25]</sup>提出并实施了将特征提取与扩展结合起来提高机器学习效果的想法。特征提取是指将输入数据转换或组合成特征,通过应用该方法可以提高机器学习的效果。本文利用基于流体物理学及测量数据的随机组合提取的特征来改进模型,例如利用注水管汇压力与注水井注入压力之间的关系以及采油井油管头压力、温度与气举量之间的关系。特征中的每项都可以具有物理意义,即描述上述关系。与单独使用某种数据(例如流量、时间)相比,使用根据流体物理学提取的几种数据的组合,有时甚至是随机的组合,也能使模型具有更好的预测效果。随机组合得到的特征中的项不必具有物理意义。此外,可以通过与现有特征相乘或相除得到新的特征。

## 1.4 优化算法选择

本文在训练阶段测试了 Levenberg-Marquardt、比例共轭梯度、贝叶斯正则化等 3 种算法。Levenberg-Marquardt 算法用于求解非线性最小二乘问题,特别是最小二乘曲线的拟合。前人开展了许多用 Levenberg-Marquardt 算法训练神经网络的研究<sup>[26-28]</sup>。SHI Xiancheng 等<sup>[26]</sup>研究发现,Levenberg-Marquardt 算法的主要局限是它只能通过在高斯-牛顿算法和梯度下降法之间进行插值来求局部最小值,当存在大量训练样品和节点权重时模型的准确性会受影响。比例共轭梯度算法对内存的要求低,所用的计算时间短。但它在线性方程系统中表现较好,在具有许多输入特征的系统中可能导致过拟合<sup>[29]</sup>。因此,它不适用于注水开发油藏数据集的训练。贝叶斯正则化算法适用于本文所开展的研究,因为这种算法应用了能够避免过拟合的正则化方法。虽然这种方法耗时较长,但是能对嘈杂数

据集进行很好的泛化<sup>[14]</sup>。贝叶斯正则化算法通过使均方误差与权重的组合达到最小值来更新神经网络的权重和偏差。一旦完成最小化就可以确定正确的组合,得到泛化良好的网络。

## 1.5 模型训练

本文中,90%的数据用于训练、验证和测试模型,其余 10%的数据用于盲测模型。也就是说,90%的数据用于训练模型体系结构、监测训练过程,确定何时停止训练模型,其余数据用于盲测训练后的模型。在机器学习中,用于训练的数据不能再用于测试<sup>[30]</sup>。

## 1.6 模型评价

前人采用过几种模型评价指标<sup>[24, 31-32]</sup>。本文用均方误差(MSE)、 $R^2$ (决定系数)和白度测试对模型进行评价<sup>[33]</sup>。均方误差是输出值与目标值之间的平均平方差。 $R^2$ 值用于衡量输出值与目标值之间的相关性。认为在测试阶段均方误差低、 $R^2$ 值接近 1 的模型是合适的模型。这两个指标能够显示模型是否提取了全部信息或者是否需要进一步调整。白度测试即通过绘制直方图来确定误差分布。对于一个好模型,误差直方图应显示零均值和一定方差。此外,也可以通过绘制交会图来对比模拟数据和验证数据,示例如图 2 所示。如果模拟数据与验证数据交会图位于 45°线上方,说明模型低估了验证数据,反之则说明模型高估了验证数据。理想情况下,模拟数据与验证数据交会图与 45°线重合。

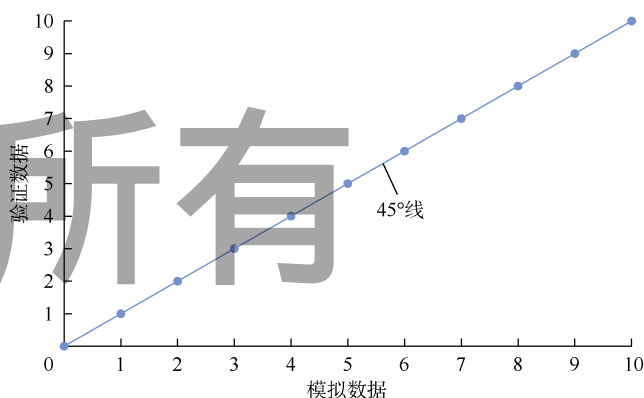


图 2 模拟数据与验证数据交会图示例

## 2 实例应用

### 2.1 油藏和数据集描述

采用马来盆地 1 个油藏的数据,鉴于保密原因,只保留了最小限度的细节。该油藏为饱和油藏,具有小型至中等气顶,中等至强水体,并在一次采油 7 年后通过注水补充能量开发。图 3 所示的 PVT 物性参数表明,油藏流体的泡点压力约为 20.68 MPa。

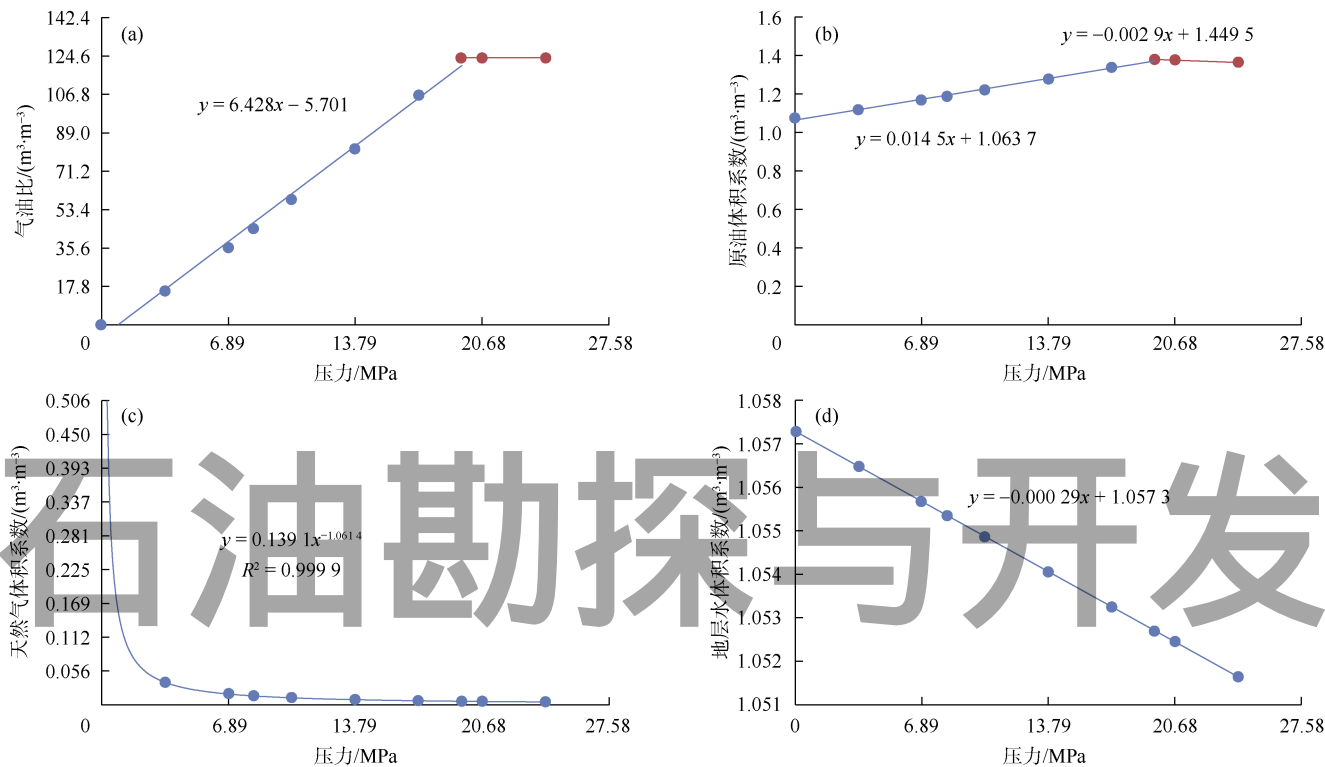


图3 油藏流体 PVT 物性参数

该油藏从 1997 年 7 月至 2004 年 7 月利用气顶+水体自然能量驱动开采。图 4 给出了这一阶段的流体产量剖面。从 2004 年至 2014 年，采用注水开发方式。图 5、图 6 和图 7 分别给出了这一阶段的产油量、产气

量、产水量与注水量对比曲线，可以看出数据非常嘈杂，导致系统无法正确理解和解释。噪声是不可避免的问题，它会影响机器学习应用中的数据收集和数据准备过程。数据噪声由测量工具引起。

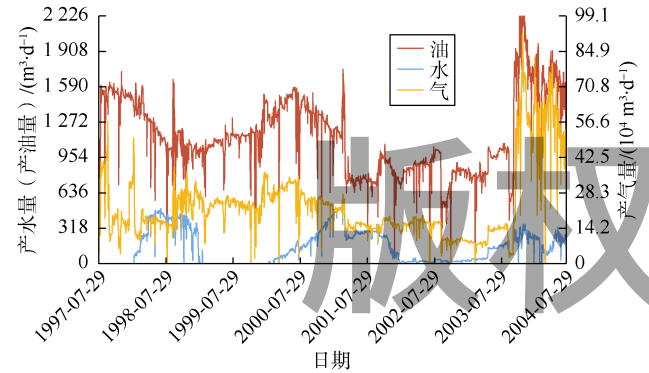


图4 注水前油、气、水产量

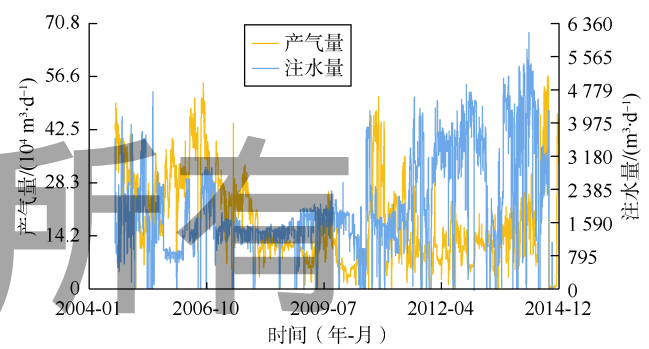


图6 开始注水后油藏产气量

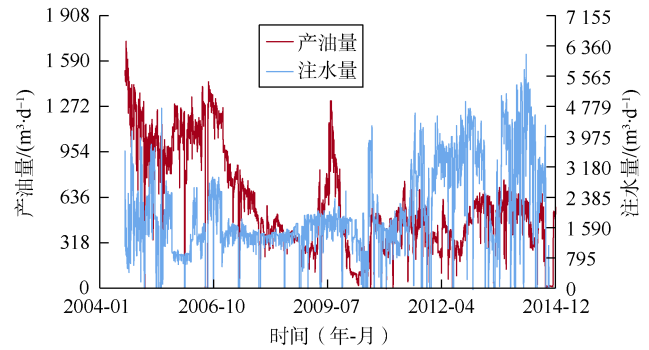


图5 开始注水后油藏产油量

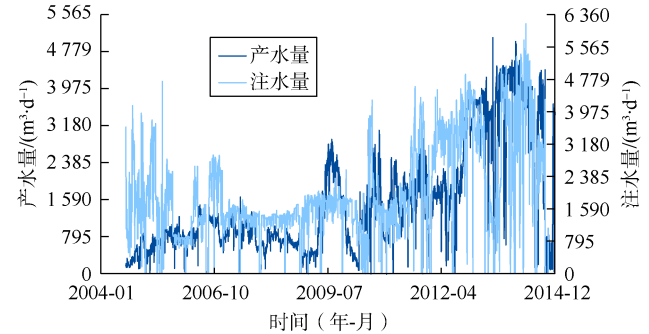


图7 开始注水后油藏产水量

2.2 输入、输出数据设计

为了获得更多有关注水井与采油井间关系的信息，本文开发了一种基于流体物理学的特征提取技术。换言之，使用了更多特征来改善模型预测效果。将这些特征与多孔介质中流体流动的物理原理结合起来，并对各项进行简单试错，直到在验证数据和模拟数据之间得到理想的拟合优度为止。表 1 给出了本文模型的输入项和输出项。表 2 给出了基于物理原理及现有输入项的随机组合提取的 7 个特征（编号 1—7）。根据模型的验证结果对这些特征进行测试后，只将特征 1、特征 2 和特征 5 用于油、气、水产量模拟。表 2 还给出了表征延时产量的特征（编号 8—13）。

表 1 模拟油、气、水产量模型的输入项和输出项

输入项		输出项	
名称	符号	名称	符号
油管头压力（生产井）	THP	产油量	OPR
油管头温度（生产井）	THT	产气量	GPR
气举量（生产井）	GLR	产水量	WPR
套管压力（生产井）	CP		
生产管汇压力	PMP		
注水量	WIR		
注水管汇压力	WIMP		
注水压力	WIP		

表 2 根据现有输入、输出项提取的特征

特征	形式	物理意义
1	$WIMP^{(k-1)}-WIP^{(k-1)}$	1 个时间步前沿注入井井筒压差
2	$THT-GLR$	随机组合
3	$WIMP-WIP$	沿注入井井筒的压差
4	$WIP/WIMP$	井底压力与地面压力之比
5	$THP/GLR$	与将流体带到地面所需的气举量有关
6	$CP-PMP$	摩擦和静液柱造成的压力损失
7	$(CP-PMP)THT$	随机组合
8	$OPR^{(k-1)}$	1 个时间步前产油量
9	$OPR^{(k-2)}$	2 个时间步前产油量
10	$GPR^{(k-1)}$	1 个时间步前产气量
11	$GPR^{(k-2)}$	2 个时间步前产气量
12	$WPR^{(k-1)}$	1 个时间步前产水量
13	$WPR^{(k-2)}$	2 个时间步前产水量

注：k 为当前时间步编号

2.3 模型结构选择

图 8 给出了由 1 个隐藏层、1 个输入层和 1 个输出层组成的神经网络的简单结构。人工神经网络模型中输入参数个数和隐藏神经元个数对模型的最终预测效果有重要影响。然而，最优人工神经网络结构方面的研究非常少。本文设置了不同的隐藏神经元个数，并根据训练和验证结果确定了模型的最优结构。在确定隐藏神经元个数时，为了避免过拟合，很重要的一点是不要有太多参数。模型的过拟合会导致预测失败，

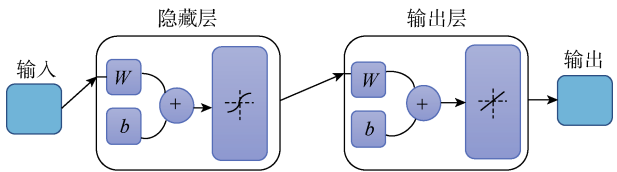


图 8 神经网络的简单结构

这是因为过拟合的模型提取了一些残差。因此，要保持尽可能少的隐藏层数，以避免过拟合。为了确定模型的最优结构，开展了一系列数据集训练与验证工作。

在数据不稳定且需要更高级神经网络的情况下，可以考虑使用更复杂的结构，图 9、图 10、图 11 所示为常用复杂结构示例。

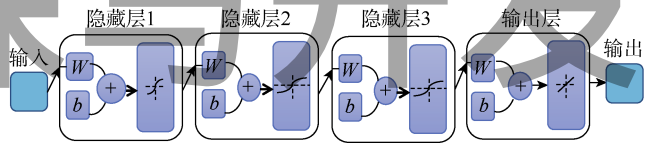


图 9 具有 3 个隐藏层的前馈神经网络结构

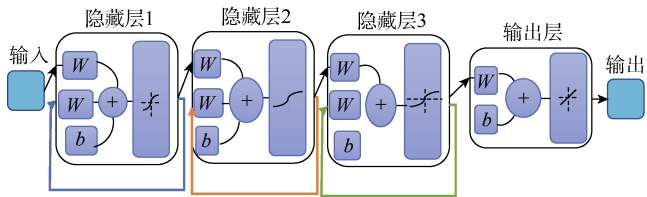


图 10 层递归神经网络结构

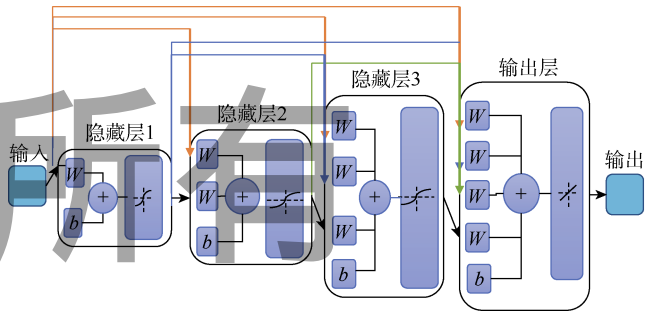


图 11 串级前馈神经网络结构

2.4 模型训练

本文使用贝叶斯正则化算法训练模型。根据验证过程中显示的  $R^2$  值进行比较，当 6 次迭代的验证误差增加时，训练步骤停止。贝叶斯正则化算法可以泛化嘈杂数据集，同时也可以防止过拟合<sup>[14]</sup>。当训练数据集的  $R^2$  值高于测试数据集的  $R^2$  值时，可以判定为过拟合。贝叶斯正则化算法的工作流程参见文献[34]。

2.5 模型评价与产量预测

比较了模型预测数据与油田实际数据，计算了模



型预测数据的均方误差、 $R^2$  值，绘制了误差分布直方图。当均方误差和误差分布直方图的均值趋近于零且  $R^2$  值趋近于 1 时，认为模型的准确性得到了验证。Ramirez A M<sup>[18]</sup>认为，应通过开展误差分析来确定新模型的准确性。因此，将本文模型与几种已知的产量预测公式进行了对比，发现与 Standing、Vazquez-Beggs 等经验公式相比，本文模型相对误差绝对值的平均值最低，为 14.732%。对模型进行验证后，将其用于产量预测。

3 结果与讨论

3.1 产油量模型的结构、训练算法选择及验证

选择带有外部输入的非线性自回归网络（NARX）作为产油量预测模型结构，其中外部输入有  $d$  个时间步的延迟。该类型网络可在给定  $d$  个时间步前的输出值以及其他一系列输入值时预测输出值。NARX 网络的结构如下。

$$y(t+1)=f[y(t),y(t-1),\cdots,y(t-n),$$

$$u(t+1),u(t),\cdots,u(t-m)]$$

(1)

NARX 网络以开环形式建立和训练。开环（单步）训练比闭环（多步）训练更有效。开环训练时可以使用正确的过去输出值，进而得到正确的当前输出值。在训练过程中将训练数据子集提交给网络，并根据训练数据子集的误差调整网络。用验证数据子集测量网络泛化程度，当泛化不再改善时停止训练。测试数据子集对训练过程没有影响，因此可以在训练过程中及训练之后独立测试模型预测效果。表 3 和表 4 分别给出了不同隐藏神经元个数条件下有特征提取和无特征提取的模型的预测效果。可以看出，与无特征提取情况下训练的模型相比，有特征提取情况下训练后模型的预测效果更好。在选择训练算法时，由表 5 可知，贝叶斯正则化算法比 Levenberg-Marquardt 算法和比例共轭梯度算法更好。尽管贝叶斯正则化算法需要的训练时间更长，但是它能很好地泛化难度大、噪声多的数据集，更适宜于处理本文实例中的产油量数据。

表 3 在无特征提取情况下训练后模型的预测效果

隐藏神经元个数	均方误差		$R^2$ 值	
	训练	测试	训练	测试
10	1 076.160	9 832.810	0.963	-0.029
15	806.620	4 318.630	0.972	0.581
20	479.087	5 552.890	0.983	0.034
25	331.670	31 841.000	0.989	0.011
30	368.534	69 404.000	0.756	0.463

表 4 在有特征提取情况下训练后模型的预测效果

隐藏神经元个数	均方误差		$R^2$ 值	
	训练	测试	训练	测试
10	1 036.060	20 094.731	0.964	0.426
15	552.402	64 983.360	0.981	0.160
20	424.630	147 702.370	0.985	0.113
25	332.921	102 965.000	0.989	0.344
30	500.958	29 152.245	0.983	0.947

表 5 不同训练算法的评价

训练算法	均方误差		$R^2$ 值	
	训练	测试	训练	测试
Levenberg-Marquardt	590.440	18 308.971	0.975	0.544
贝叶斯正则化	500.958	29 152.245	0.983	0.947
比例共轭梯度	3 628.976	4 568.945	0.866	0.599

表 6 给出了在有特征提取情况下利用贝叶斯正则化算法训练的产油量模型的最优结构相关信息。

表 6 预测产油量的人工神经网络模型的最优结构相关信息

项目	数值/内容	项目	数值/内容
输入项个数	13	$R^2$ 值（训练）	0.983
隐藏神经元个数	30	$R^2$ 值（测试）	0.947
输出	产油量	遍历次数	946
训练算法	贝叶斯正则化	训练时间	103 s

在模型评价过程中，利用占全部数据 10%的盲测数据（训练过程中未用的数据）模拟产油量。由图 12 所示的交会图可知，有特征提取和无特征提取时模型的  $R^2$  值分别是 0.947 和 0.463，且有特征提取时的模拟产油量与验证产油量交会图与 45°线更接近。这说明本文使用的特征提取方法显著改善了模型的预测效果。由图 13 可知，有特征提取时模型的模拟结果与验证数据非常接近，而无特征提取时则不然。此外，图 14 所示的误差分布直方图呈现正态分布趋势，表明模型的平均误差接近于零，且该误差主要为测量误差。

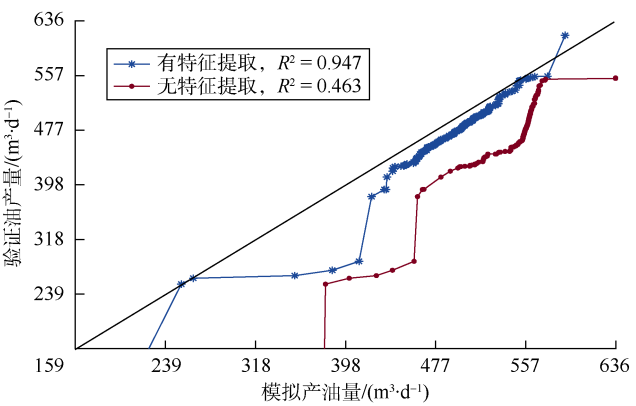


图 12 产油量模拟数据与验证数据交会图

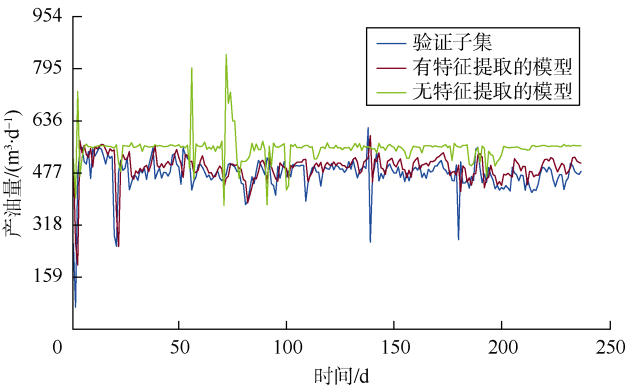


图 13 实际产油量与模拟产油量比较

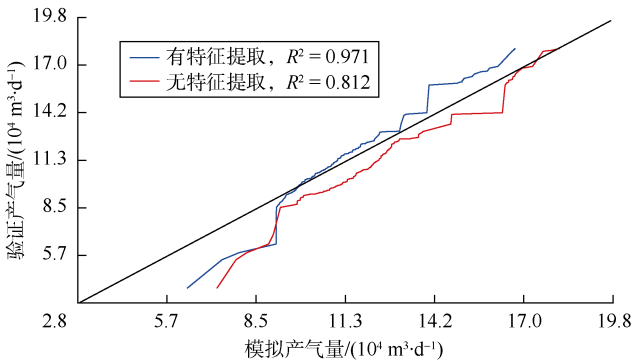


图 15 产气量模拟数据与验证数据交会图

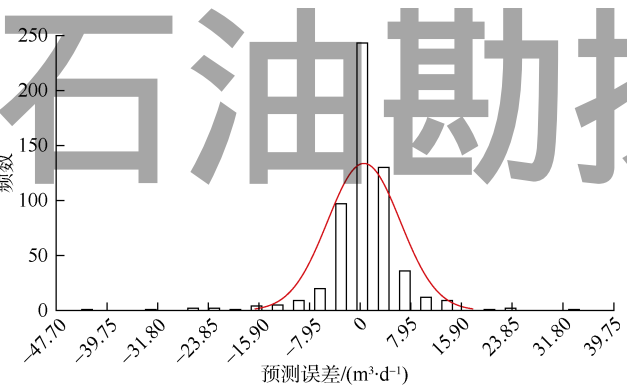


图 14 产油量模型的误差分布直方图

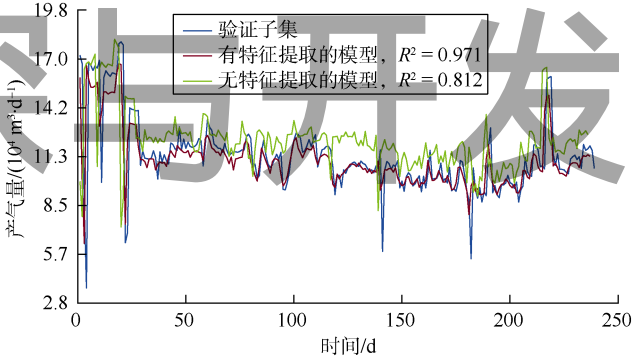


图 16 实际产气量与模拟产气量比较

3.2 产气量模型的结构、训练算法选择及验证

使用与建立产油量模型相同的方法建立产气量预测模型，并使用前文提取得到的相同特征改善模型预测效果。用训练过程中未使用的数据进行模型测试或验证。表 7 给出了在有特征提取情况下产气量模型的最优结构相关信息。

表 7 预测产气量的人工神经网络模型的最优结构  
相关信息

项目	数值/内容	项目	数值/内容
输入项个数	13	$R^2$ 值（训练）	0.949
隐藏神经元个数	15	$R^2$ 值（测试）	0.971
输出	产气量	遍历次数	844
训练算法	贝叶斯正则化	训练时间	96 s

由图 15 所示的交会图可知，有特征提取和无特征提取时模型的  $R^2$  值分别是 0.971 和 0.812；无特征提取时大部分模型预测结果高于实际数据，而有特征提取时多数模型预测结果与实际数据持平。这再次说明本文采取的特征提取方法提高了模型的预测效果。由图 16 可知，与无特征提取时的模型相比，有特征提取时模型的模拟结果与验证数据更接近。此外，图 17 所示的误差分布直方图呈现正态分布趋势，表明模型的平均误差接近于零，且该误差主要为测量误差。

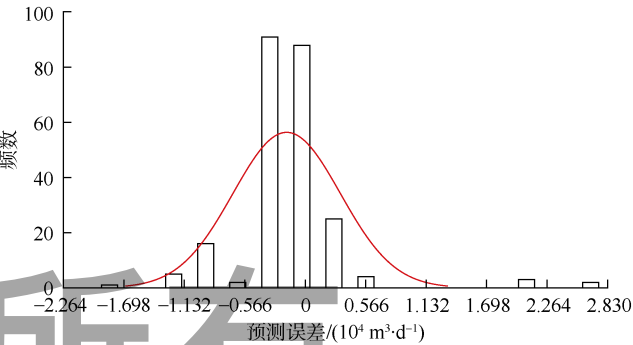


图 17 产气量模型的误差分布直方图

3.3 产水量模型的结构、训练算法选择及验证

使用与建立产油量模型相同的方法建立产水量预测模型。表 8 给出了在有特征提取情况下产水量模型的最优结构相关信息。

表 8 预测产水量的人工神经网络模型的最优结构  
相关信息

项目	数值/内容	项目	数值/内容
输入项个数	13	$R^2$ 值（训练）	0.965
隐藏神经元个数	10	$R^2$ 值（测试）	0.971
输出	产水量	遍历次数	589
训练算法	贝叶斯正则化	训练时间	85 s

由图 18 所示的交会图可知，有特征提取和无特征提取时模型的  $R^2$  值分别是 0.987 和 0.891。由图 19 可

知,与无特征提取时的模型相比,有特征提取时模型的模拟结果与验证数据更接近。此外,图 20 所示的误差分布直方图呈现正态分布趋势,表明误差主要为测量误差。

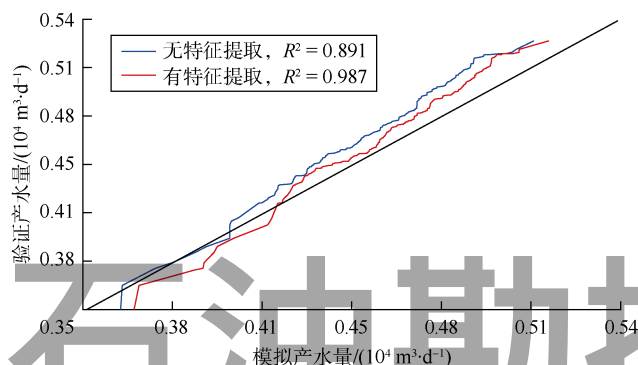


图 18 产水量模拟数据与验证数据交会图

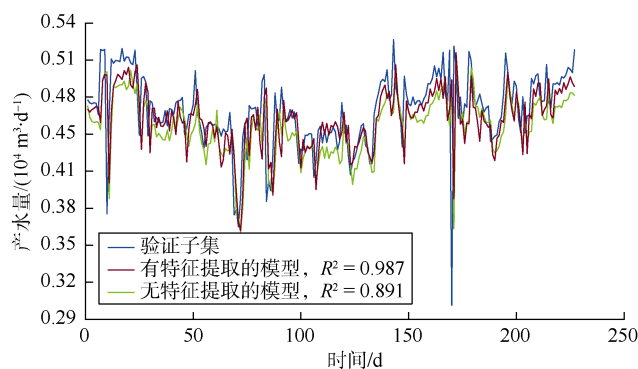


图 19 实际产水量与模拟产水量比较

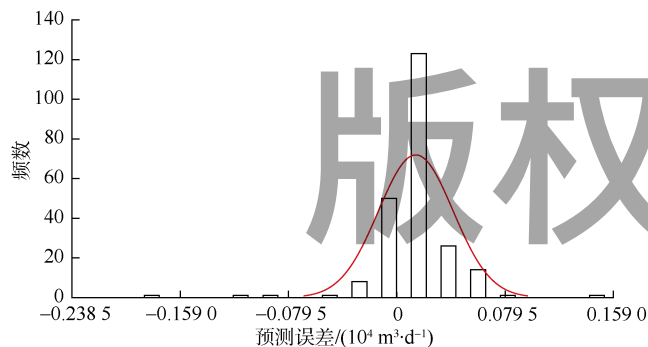


图 20 产水量模型的误差分布直方图

## 4 结论

开发了人工神经网络模型来预测注水开发油藏的产油量、产气量和产水量。与传统预测模型不同,人工神经网络模型在训练所选结构时只需少量容易获得的信息。用基于油藏流体物理学以及变量的随机组合的特征提取方法来改善模型预测效果。研究发现,用贝叶斯正则化算法训练的神经网络结构能在训练和验

证所选神经网络结构的过程中得到最小的均方误差和最高的决定系数。这种算法一般需要较长时间,但是能对产油量、产气量、产水量等嘈杂数据集进行良好泛化。用误差分布直方图来分析模拟数据与验证数据之间的差异,发现每种流体产量预测模型误差均主要为测量误差。

在本文提取的所有特征中,时间步长延迟项可以提高模型的预测效果。这一结论具有重要意义,因为包括油藏数值模拟在内的许多二次采油产量预测方法并没有将过去的注入量作为影响当前产量的因素。

符号注释:

$b$ ——偏差; $d$ ——延迟时间步数; $f$ ——NARX 网络结构; $m, n$ ——时间步个数; $t$ ——时间步序号; $u(t+1), u(t), \dots, u(t-m)$ ——外部输入数据; $W$ ——权重; $y(t), y(t-1), \dots, y(t-n)$ ——观测数据; $y(t+1)$ ——模型预测值; $\sigma$ ——标准差。

参考文献:

- [1] ZHONG H, WU K, JI D, et al. Using least square support vector machines to approximate single phase flow[R]. SPE 185881-MS, 2017.
- [2] NEGASH B M. Artificial neural network and inverse solution method for assisted history matching of a reservoir model[J]. International Journal of Applied Engineering Research, 2017, 12: 2952-2962.
- [3] WEI Chenji, WANG Yuhe, DING Yutao, et al. Uncertainty assessment in production forecasting and optimization for a giant multi-layered sandstone reservoir using optimized artificial neural network technology[R]. SPE 186069-MS, 2017.
- [4] NEGASH B M, VASANT P M, JUFAR S R. Application of artificial neural networks for calibration of a reservoir model[J]. Intelligent Decision Technologies, 2018, 12(2): 1-13.
- [5] NEGASH B M. History matching of the PUNQ-S3 reservoir model using proxy modeling and multi-objective optimizations[R]. Kuala Lumpur, Malaysia: 2016 International Conference on Industrial Engineering and Operations Management, 2016.
- [6] WANTAWIN M, YU W, DACHANUWATTANA S, et al. An iterative response-surface methodology by use of high-degree-polynomial proxy models for integrated history matching and probabilistic forecasting applied to shale-gas reservoirs[R]. SPE 187938-PA, 2017.
- [7] WANG D, NIU D, LI H A. Predicting waterflooding performance in low-permeability reservoirs with linear dynamical systems[R]. SPE 185960-PA, 2017.
- [8] TIAN C, HORNE R N. Inferring interwell connectivity using production data[R]. SPE 181556-MS, 2016.
- [9] SAPUTELLI L. Technology focus: Petroleum data analytics[J]. Journal of Petroleum Technology, 2017, 69(10): 87.
- [10] MOHAGHEGH S D. Shale analytics: Data-driven analytics in unconventional resources[M]. Berlin: Springer International Publishing, 2017.



- [11] MIŠKUF M. Data mining in cloud usage data with Matlab's statistics and machine learning toolbox[R]. Herl'any, Slovakia: 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), 2017.
- [12] ANIFOWOSE F A, LABADIN J, ABDULRAHEEM A. Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization[J]. *Journal of Petroleum Science and Engineering*, 2017, 151: 480-487.
- [13] XIE Y, ZHU C, ZHOU W, et al. Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances[J]. *Journal of Petroleum Science and Engineering*, 2018, 160: 182-193.
- [14] FULFORD D S, BOWIE B, BERRY M E, et al. Machine learning as a reliable technology for evaluating time/rate performance of unconventional wells[J]. *SPE Economics & Management*, 2016, 8(1): 23-39.
- [15] CAO Q. Data driven production forecasting using machine learning[R]. Buenos Aires, Argentina: SPE Argentina Exploration and Production of Unconventional Resources Symposium, 2016.
- [16] LI Y, HAN Y. Decline curve analysis for production forecasting based on machine learning[R]. SPE 189205-MS, 2017.
- [17] TEMIZEL C, NABIZADEH M, KADKHODAEI N, et al. Data-driven optimization of injection/production in waterflood operations[R]. SPE 187468-MS, 2017.
- [18] RAMIREZ A M. Prediction of PVT properties in crude oil using machine learning techniques (MLT)[R]. Buenos Aires, Argentina: SPE Latin America and Caribbean Petroleum Engineering Conference, 2017.
- [19] MOGHADASI L, RANAEE E, INZOLI F, et al. Petrophysical well log analysis through intelligent methods[R]. SPE 185922-MS, 2017.
- [20] ZHENG J, LEUNG J Y, SAWATZKY R P, et al. An AI-based workflow for estimating shale barrier configurations from SAGD production histories[R]. SPE 184984-MS, 2017.
- [21] AZIZI S, AHMADLOO E. Prediction of void fraction for gas-liquid flow in horizontal, upward and downward inclined pipes using artificial neural network[J]. *International Journal of Multiphase Flow*, 2016, 87: 35-44.
- [22] MA Z, LEUNG J Y, ZANON S. Integration of artificial intelligence and production data analysis for shale heterogeneity characterization in steam-assisted gravity-drainage reservoirs[J]. *Journal of Petroleum Science and Engineering*, 2017, 163: 139-155.
- [23] KIM P. MATLAB deep learning: With machine learning, neural networks and artificial intelligence[M]. Berkeley, CA: Apress, 2017.
- [24] ARAS S, KOCAKOÇ İ D. A new model selection strategy in time series forecasting with artificial neural networks: IHITS[J]. *Neurocomputing*, 2016, 174: 974-987.
- [25] LÓPEZ-IÑESTA E, GRIMALDO F, AREVALILLO-HERRÁEZ M. Combining feature extraction and expansion to improve classification based similarity learning[J]. *Pattern Recognition Letters*, 2017, 93: 95-103.
- [26] SHI Xiancheng, FENG Yucheng, ZENG Jinsong, et al. Chaos time-series prediction based on an improved recursive levenberg-marquardt algorithm[J]. *CHAOS SOLITONS AND FRACTALS*, 2017, 100: 57-61.
- [27] XIAN Zhang, AWOTUNDE A A. Improvement of Levenberg-Marquardt algorithm during history fitting for reservoir simulation[J]. *Petroleum Exploration and Development*, 2016, 43(5): 876-885.
- [28] UTAMI E, WIBAWA A, BIYANTO T R, et al. Reservoir zone prediction using logging data - multi well based on levenberg-marquardt method[R]. Langkawi, Malaysia: 2017 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2017.
- [29] SIMM J, ARANY A, ZAKERI P, et al. Macau: Scalable Bayesian factorization with high-dimensional side information using MCMC[R]. Tokyo: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), 2017.
- [30] GUÉRILLOT D R, BRUYELLE J. Uncertainty assessment in production forecast with an optimal artificial neural network[R]. SPE 183921-MS, 2017.
- [31] LU P, LIU H, SERRATELLA C, et al. Assessment of data-driven, machine learning techniques for machinery prognostics of offshore assets[R]. OTC 27577-PT, 2017.
- [32] AMIRIAN E, DEJAM M, CHEN Z, et al. Performance forecasting for polymer flooding in heavy oil reservoirs[J]. *Fuel*, 2018, 216: 83-100.
- [33] NEGASH B M, TUFA L D, RAMASAMY M, et al. System identification based proxy model of a reservoir under water injection[J]. *Modelling and Simulation in Engineering*, 2017(1):1-10.
- [34] HAGAN M T. Neural network design[M]. Boulder, Colorado, USA: Campus Publishing Service, University of Colorado, 2002.
- 第一作者简介: NEGASH Berihun Mamo (1982-), 男, 埃塞俄比亚人, 博士, 马来西亚石油大学石油工程系高级讲师, 主要从事系统模拟与仿真、机器学习、油气藏开发等方面的研究。地址: University Teknologi PETRONAS, Petroleum Engineering Department, 32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia. E-mail: bmamo.negash@utp.edu.my
- 收稿日期: 2019-09-19 修回日期: 2020-01-20
- (编辑 胡苇玮)