

基于人工神经网络的钻井机械钻速预测模型的分析与建立

刘胜娃¹ 孙俊明² 高 翔² 王 敏²

(中石油川庆钻探工程有限公司 西安 710018)¹ (西北工业大学计算机学院 西安 710072)²

摘 要 长庆钻井公司的各部门在过去的十几年里积累了海量的各类钻井数据,随着企业云平台的建设完成和投入运营,目前已经利用数据集成技术将各部门的数据进行了汇总和统一规范管理。挖掘这些积累的宝贵数据可以为科学地制定钻井方案提供参考。机械钻速的准确预测对于科学配置钻井资源、降低钻井成本有重要作用。文中介绍了一种基于人工神经网络技术的定向井机械钻速预测模型实现方法,由钻井领域专家确定网络的输入和输出,由数据工程师设计网络拓扑结构和网络训练。实验结果表明,在数据量较充足、数据质量较高的条件下,采用神经网络构建的预测模型的预测准确度较高,完全能够满足用户的需求。

关键词 大数据,人工智能,神经网络

中图法分类号 TP391 文献标识码 A

Analysis and Establishment of Drilling Speed Prediction Model for Drilling Machinery
Based on Artificial Neural Networks

LIU Sheng-wa¹ SUN Jun-ming² GAO Xiang² WANG Min²

(Department of CNPC Chuanqing Drilling Engineering Company,Xi'an 710018,China)¹

(School of Computer Science,Northwest Polytechnical University,Xi'an 710072,China)²

Abstract Various departments of Changqing Drilling Company have accumulated a great deal of drilling data in the past ten years. With the completion of the construction and operation of the enterprise cloud platform,data integration technology has been adopted to collect and standardize the data of various departments. Mining these accumulated valuable data can provide reference for making drilling plan scientifically. Accurate prediction of penetration rate plays an important role in scientific allocation of drilling resources and reducing drilling cost. This paper proposed a novel method based on neural networks for predicting the penetration rate of directional wells. The network inputs and outputs are determined by drilling experts. The network topology and network training are designed by data engineers. The experimental results show that the prediction model constructed by neural networks has high accuracy and can meet the needs of users under the condition of sufficient data and high data quality.

Keywords Big data,Artificial intelligence,Neural network

1 引言

与世界产油大国相比,中国是一个贫油国家,经过多年来的勘探,我国石油的浅层与中层的储量已基本探清,已剩余不多并正在开采,日渐枯竭,亟需要向深井开采发展^[1-2]。深井开采成本高昂,改进深井钻井技术、提高钻井效率、提升钻井速度是减少开采成本的重要手段和措施^[3-4]。机械钻速是钻井工程中的一个重要指标,它直接与开采成本、开采时间相关,如果能够准确地预测出机械钻速,就能估算出钻进时间,从而优化配置资源,提高人员和设备的使用效率,从而达到降低开采成本的目的。

国内外有很多专家学者研究了机械钻速的预测方法,著名学者李昌盛^[5]提出基于多元回归分析的钻速预测方法;西南石油大学的范翔宇教授提出了采用数理统计等办法预测机

械钻速的方法;Njobuenwu 等^[6]国外专家提出了钻井固体对钻井速度和性能的影响;Rashidi 等^[7]外国学者都提出了各自的与地层性质和钻头结构相关的钻速方程。传统方法确实给机械钻速的提高带来了可见的影响,但是这些预测方法大多依靠经验,或者是根据一些理论的钻井公式进行推理,或者是依靠一些如控制变量法等实验,通过实验找寻影响机械钻速的最主要因素。经过多年的钻井工作以及国内外调查发现,机械钻速是受钻井液、钻井地区地质地形、钻头尺寸、泥饼等诸多特征条件共同影响的,如果仅靠经验来获取钻速,往往不能获取最优化的钻速;靠控制变量法实验获取单个或者若干关键因素,不仅费时费力,而且得到的结果也往往不是最优的。随着云平台、高性能计算技术的发展,沉寂已久的人工神经网络和深度学习技术重新焕发青春。目前,人工神经网络已经在钻井工程中的岩性识别、钻头优选、钻井液优化等方面

本文受到陕西省科技计划项目(2018GY-048),中石油川庆钻探公司科研项目(CZ2017-18)资助。

刘胜娃(1963—),男,高级工程师,主要研究方向为软件设计,E-mail:zjsxslw@cnpc.com.cn;孙俊明(1992—),男,硕士生,主要研究方向为计算机应用;高翔(1974—),男,博士,副教授,CCF 会员,主要研究方向为计算机应用,E-mail:gaoxg@nwpu.edu.cn(通信作者);王敏(1975—),女,博士,副教授,主要研究方向为计算机应用。

得到了广泛的应用和推广^[8-9]。因此,我们考虑是否能通过机器学习的方法找到影响特定地区机械钻速的真正的主因,进而准确地预测钻速,达到提高钻速减少钻井周期的目的?

长庆钻井公司的信息化建设已逾 10 年,收集了产区大量区块的地层地质信息,例如钻井液配方、钻具组合、钻井周期、机械钻速等相关数据,为长庆钻井的大数据研究提供了坚实的基础。本文利用机器学习技术进行分析研究,找出影响长庆产区中影响机械钻速的关键因素,建立预测机械钻速的模型,为科学指导生产提供有力的支持。

过去两年里,长庆钻井公司的大数据工程整合了各部门的数据,打通了信息孤岛,统一了数据规范,提高了数据质量,为机械钻速的影响因素分析和预测提供了坚实的数据基础^[10]。本文就是在前期整合好的数据基础上,尝试采用人工神经网络等技术和领域知识相结合的方法,实现特定区块地层的机械钻速的准确预测^[11]。实践中,通过与领域专家头脑风暴的方式从众多的数据字段中筛选出了最有可能影响机械钻速的数据字段^[12-13],形成神经网络的输入层,经过多次训练和迭代,构建出钻速模型,并用真实的历史数据进行了测试。本文也使用其他机器学习的方法进行了对比,结果表明使用 BP 神经网络方法预测的机械钻速的准确度更高。

2 BP 神经网络简介

BP(Back Propagation)神经网络是一种按误差逆传播算法,通过使用最速下降法和反向传播的学习规则不断迭代,调整网络的权值和阈值,从而训练出误差平方和最小的多层前馈神经网络。

BP 神经网络的训练包括正向传播和逆向反馈两个过程。正向传播时,从输入层输入训练样本,经过各隐层逐级计算后,传到输出层,然后比对输出层的输出与期望的正确结果,如果偏差超过给定阈值,则转入误差的逆向反馈阶段。在反馈时,将输出结果按照特定形式通过隐层反向传递给输入层,各层的所有单元都将分摊到部分误差,以此获得各神经元的误差信号,根据误差信号修正各级神经元的权值。

3 钻井机械钻速预测模型的建立

3.1 数据预处理

神经网络需要在互相独立的数据集上进行训练、评估和测试,以防止过拟合,从而使该神经网络对其他未知数据也能进行准确预测。训练模型通过训练数据学习得到不同的参数,并使用评估数据进行评估。性能最优的模型再使用测试数据进行测试,通过与带正确标识的数据进行比较,评定神经网络模型的性能^[14]。在该钻井机械钻速预测模型中,数据集采用 10 折交叉验证,分配为:训练数据占 90%,评估数据占 10%。

由于一个预测模型仅适用于一个特定区块,因此我们选择了长庆南梁区块 220 口定向井的 2927 条 excel 数据构成数据集。部分数据如图 1 所示,包含项目部、钻机型号、钻头类型、钻头型号、钻头尺寸、钻压、泵压、转速、排量、密度、粘度、失水、泥饼、含砂量、PH 值、下入井深、起出井深、纯钻时间等 20 个输入参数字段,以机械钻速作为输出参数字段。

钻头尺寸	钻压	泵压	转速	排量	密度	粘度	失水	泥饼	含砂	PH	下入井深	起出井深	纯钻时间	机械钻速
346	20	6	60	32	1	28			7	0	373	6	62.17	
346	20	5	60	32	1	28			7	0	373	6	62.17	
222.2	100	8	60	32	1.05	68	5	0.5	0.1	11	1,700	2,074	12	31.17
346	20	5	60	32	1	28			7	0	383	4	95.75	
311	30	3	180	29	1.01	31	22	0.1	0.1	7	0	620	17	36.47
311	20	8	70	30	1.01	27			7	0	615	20	30.75	
311	40	8	60	30	1.01	30			7	0	614	17	36.12	
346	60	9	60	32	1	29			0.1	7	0	623	19	32.79
206.4	120	10	60	32	1.05	58	6	0.3	0.1	11	1,940	2,303	18	20.17

图 1 220 口定向井钻井部分数据

不同字段的取值范围差异极大,例如钻头尺寸、钻压、泵压、含砂量等属性的数据相对于其他输入属性字段的数据特别大或特别小的奇异样本数据,这就不可避免地造成取值范围大的数据对整体的影响力较大,这是不合理的。于是需要对每个字段都进行规范化处理,从而解决不同评价指标具有不同量纲的问题。我们采用了归一化的规范化方法^[15-16],其中采用最大-最小标准化方法将各个字段的数据转换到[0, 1]之间。最大-最小标准化方法的计算公式如下:

$$Y=(X-\min A)/(\max A-\min A)$$

按照公式进行计算,归一化后的数据取值范围为(0,1]。该数据还包含了一些像钻头类型、钻头型号等用文字描述的字段,对于这些字段,我们用小于 10 的整数取代。

4 机械钻速预测模型神经网络的训练

4.1 输入输出层的设计

神经网络训练的的目的是找到合适的权值集合使得目标函数达到最小值,目标函数设计的是预测结果与实测值之间的差值。在机械钻速预测模型中,在和各部门领域专家进行了数轮的头脑风暴讨论后,归纳出 excel 表格中的钻头尺寸、钻压、泵压等 11 个字段作为网络的输入,仅有机械钻速这 1 个数值作为网络的输出。因此,输入层的节点数为 11,输出层的节点数为 1。

4.2 隐含层的设计

在 BP 神经网络中,输入层和输出层层数和节点数是确定的,但是隐含层层数和节点数是不定的。一般来说,增加隐含层层数能减少网络的误差并且提高网络的精度,但是同样也导致了网络的复杂和过拟合现象的出现。通常情况下,在 1 层隐含层的 BP 神经网络中,设置好隐含层中的节点数也能很好地起到平衡过拟合和保证精度的效果^[17]。在本次实验中,设置了 1 层隐含层,并且通过如下经验式(1)来确定节点数:

$$h=\sqrt{m+n}+a$$

其中, h 为隐含层节点的数目, m 和 n 分别为输入层和输出层节点的数目, a 为调节常数。本文采取试凑的方法,根据经验公式设置不同参数。经过多次对比实验,最后 h 为 7 时有更好的效果。

4.3 网络的训练

BP 神经网络的训练过程,就是调整网络中神经元的权重和偏置的过程。该训练过程分为以下两个步骤。

(1)前向输入

最初,神经网络的权重和偏置是随机初始化的,将一个随机的 $[-1,1]$ 区间内的实数赋值给一个神经元的权值,将一个

$[0,1]$ 区间内的随机实数赋值给一个偏置,初始化完成后就可以进行前向输入。本次数据包含 11 个属性,因此可以在输入层定义 11 个神经元,神经元 N_i 用于表示数据记录上第 i 个的属性值 X_i ,其权值为 w_i 。输入层以外的各层的输入值均是前一层的输入值加权累加计算后加上偏置,相当于给每个结点的输入值做变换后输出,如图 2 所示。

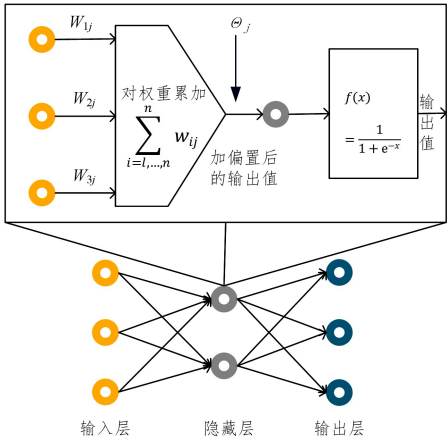


图 2 前向输入逐层传播流程图

BP 神经网络采用了连续可微的且具有阈值特性的 Sigmoid 函数作为神经元的触发函数：

$$f(x)=1/(1+e^x)$$
(2)

前向输入的输出层的转换公式为：

$$O_j=f(I_j)=1/(1+e^{I_j})$$
(3)

其中, $I_j=\sum_{i=0}^jw_ix_i$ 。前向输入的过程中需对隐藏层、输出层的所有神经元都按式(3)计算输出值。

(2) 逆向反馈

逆向反馈就是从输出层开始逆向传递的过程,通过多次迭代来修正神经网络中的神经元权值和偏置值。每次迭代根据神经网络的输出与带标识训练数据的正确结果之间的差异进行参数调整,缩小差异的过程就是神经网络的训练过程。输出层计算公式为：

$$E_j=O_j(1-O_j)(T_j-O_j)$$
(4)

其中, E_j 表示 j 个神经元的误差, O_j 表示第 j 个结点的输出, T_j 是训练数据的标准值。中间的隐藏层不直接与训练数据进行比较,而是所有误差的加权累计,计算公式为：

$$E_j=O_j(1-O_j)\sum E_kW_{jk}$$
(5)

其中, E_j 表示 j 个神经元的误差, E_k 代表下一层的结点 k 的误差率, W_{jk} 代表当前结点 j 到下一层的结点 k 的权值。计算完误差率后,就可以更新权值和偏置,权值计算公式为：

$$\Delta W_{ij}=\lambda E_jO_i$$
(6)

$$W_{ij}=W_{ij}+\Delta W_{ij}$$
(7)

其中, λ 表示学习速率,为 $[0,1]$ 之间的一个数,决定了神经网络的收敛速度,设置得越大收敛越快,但容易陷入局部最优解,设置得太小则耗费时间太长,效率太低,因此需要折中考虑。偏置的计算公式为：

$$\Delta \theta_j=\lambda E_j$$
(8)

$$\theta_j=\theta_j+\Delta \theta_j$$
(9)

神经网络的训练过程在达到终止条件后结束。终止条件有两个,满足其一即可:1)事先设置好训练次数,达到次数后终止训练;2)设定差异的阈值,当输出结果和预期结果的差异

小于给定阈值时则终止训练。

4.4 交叉验证

为了保证模型不会因为原始数据随机分配训练集和测试集造成随机误差,我们进行了 10 折交叉验证。将数据集平均分成 10 份,取其中 9 份作为训练集,剩余 1 份作为测试集进行一次实验;然后依次从训练集中拿出一份作为新的测试集,将原来的测试集放到训练集当中。这样依次进行 10 组实验。取绝对平均误差最小的模型作为最后的模型,取 10 次实验结果的绝对平均误差的平均值来评估该模型。

5 机械钻速预测模型评估

5.1 训练方法的选择

在模型的训练初期,我们分别选用了如多元线性回归、支持向量回归等常见回归方法来创建模型并进行对比实验,实验结果如表 1 所示。最终,根据表 1 结果,综合考虑各种方法的可决系数 R^2 ,选择 R^2 最高的 BP 神经网络作为本次模型的创建方法。

表 1 训练方法对比

训练方法	R^2
BP 神经网络	0.88
多元线性回归	0.70
支持向量回归	0.85
岭回归	0.75
主成分回归	0.67

5.2 BP 神经网络方法下的结果评估

我们用预处理的 2927 条数据对神经网络进行了训练和验证,在硬件配置为 intel i7-4500u,主频 3.0 GHz,内存 16 GB 的主机上进行实验。网络参数设置中,设置学习率为 0.01,最大训练次数为 1200 Epochs。训练轨迹如图 3 所示,大约在经过了 1070 Epochs 后就达到了预期要求。

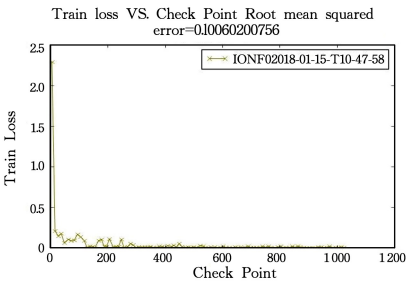


图 3 机械钻速预测模型的训练轨迹

图 4 显示了机械钻速预测模型神经网络的实际执行效果,图中圆圈代表训练数据中的机械钻速实测值,菱形代表训练完成后的神经网络的预测值,图中横坐标为井号,纵坐标为钻速,可见预测的准确度非常高,拟合效果非常好。

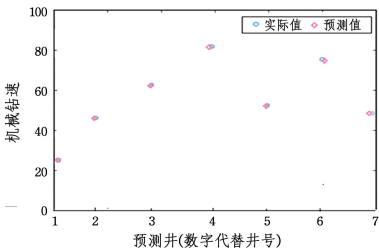


图 4 机械钻速预测模型神经网络的实际执行效果

将数据集进行 10 次随机划分,每次都按 9 : 1 的比例将

数据集划分为训练集和评估集,再进行训练和预测,得出该模型的 R^2 为0.88。

5.3 模型的实际泛化效果

我们采用此方法对南梁井区、罗1井区、镇90井区、靖中北、靖南、桃2井区几个区块的数据进行了训练和仿真预测,如表2所列。例如,在南梁区块选择了2927口井作为训练数据,另外剩余的1000口井作为验证数据。在最初的300轮左右的迭代中,损失误差较大和训练耗时较长,随着迭代次数的不断增加,训练损失误差率逐渐收敛,训练时间同时显著缩短。训练结束后,绝对平均误差为4.25。各组训练结果如表2所列,精确度已经基本可以满足用户的实际需求。

表2 基于神经网络建立预测模型的不同区块定向井训练结果

区块数据组	平均绝对误差	R^2
南梁井区	4.25	0.88
罗1井区	5.70	0.85
镇90井区	4.52	0.87
靖中北	5.44	0.82
靖南	4.77	0.84
桃2井区	4.76	0.84

结束语 长庆钻井公司在建成企业云平台,完成了数据集成后,多年来积累的钻井大数据就像一个待挖的金矿,需要我们去进一步发掘数据背后隐藏的经验知识,为钻井提供科学的指导。本文仅仅是针对机械钻速预测的课题研究,此外还有许多类似课题,如钻井液配方、钻具组合、安全生产管理等,需要我们继续采用新兴的大数据和人工智能技术去深入研究,为提高钻井效率和提升效益提供强有力的支持。

参 考 文 献

[1] ZHANG N N, WANG Q, WANG J J, et al. Characteristics of oil and gas discoveries in recent 20 years and future exploration in the world[J]. China Petroleum Exploration, 2018, 23(1): 44-53.

[2] HU W R, BAO J W. Development trends of oil industry and China's countermeasures[J]. Journal of China University of Petroleum (Edition of Natural Science), 2018, 42(4): 1-10.

[3] Yi C, Li G S, SHEN Z H. Experimental study on drilling rate improvement by self-resonating cavitating jet[J]. Natural Gas Industry, 2006, 26(5): 52-54.

[4] YARALIA O. Assessment of relationships between drilling rate index and mechanical properties of rocks[J]. Tunnelling & Underground Space Technology Incorporating Trenchless Technol-

ogy Research, 2013, 33(33): 46-53.

[5] LI C S. Study of Method for Predict Rate of Penetration Based of Multiple Regression Analysis[J]. Science Technology & Engineering, 2013, 13(7): 1740-1744.

[6] NJOBUENWU D O, WOBO C A. Effect of drilled solids on drilling rate and performance[J]. Journal of Petroleum Science & Engineering, 2007, 55(3): 271-276.

[7] RASHIDI B, HARELAND G, WU Z. Performance, simulation and field application modeling of rollercone bits[J]. Journal of Petroleum Science and Engineering, 2015, 133: 507-517.

[8] BI X L, YAN T, TAO L J. Research on optimizing bit by neural network in Qingshen Oilfield[J]. Journal of Harbin Engineering University, 2006, 27(z1): 111-114.

[9] LI J, LI K, WANG B. An Evaluation Model of Drilling Safety Based on Combined Rough Set and Neural Network[J]. Journal of Southwest Petroleum University, 2017, 39(5): 120-128.

[10] ZHONG J B, SHI Y J, ZHANG H T, et al. Construction and Application of Log Data Transmission and Cooperative Evaluation System Based on Network[J]. Well Logging Technology, 2016, 40(3): 356-363.

[11] GUAN Z C, SHENG Y N, XU Y Q, et al. Dynamic risk assessment method of drilling based on PSO optimized BP neural network[J]. Journal of Safety Science & Technology, 2017, 13(8): 5-11.

[12] GUO Z M, ZHOU A Y. Research on Data Quality and Data Cleaning: a Surve[J]. Journal of Software, 2002, 13(11): 2076-2082.

[13] PIAO J S, YIN K, FAN L M. Neural Network Method in the Penetration Rate Prediction of Percussive Drilling[J]. Journal of Jilin University, 2009, 39(5): 882-886.

[14] REICH Y, BARAI S V. Evaluating machine learning models for engineering problems[J]. Artificial Intelligence in Engineering, 1999, 13(3): 257-272.

[15] ZHU Q S, ZHOU D D, HUANG W. Application Research of Preprocess in BP Neural Network Sample Data [J]. World Sci-Tech R & D, 2012, 34(4): 624-626.

[16] HINTON G, VINYALS O, DEAN J. Distilling the Knowledge in a Neural Network[J]. Computer Science, 2015, 14(7): 38-39.

[17] WANG R B, XU H Y, LI B, et al. Research on Method of Determining Hidden Layer Nodes in BP Neural Network[J]. Computer Technology and Development, 2018, 28(4): 31-35.

(上接第588页)

[11] MCDONOUGH K, KOLMANOVSKY I. Integrator resetting for enforcing constraints in aircraft flight control systems[C]// AIAA Guidance, Navigation, and Control Conference. 2015: 2015-1995.

[12] FALANGAS E T. Performance Evaluation and Design of Flight Vehicle Control Systems[M]. Hoboken: John Wiley & Sons, 2015.

[13] 刘剑威, 吴传利, 徐国亮. 反舰导弹典型攻击模式及运动仿真研究[J]. 指挥控制与仿真, 2006, 28(5): 47-50.

[14] MHLANGA F S, PERRY E L, KIRCHNER R. On Adapting a Military Combat Discrete Event Simulation with Big Data and

Geospatial Modeling Toward a Predictive Model Ecosystem for Interpersonal Violence[J]. Journal of Information Systems Applied Research, 2015, 8(1): 42-55.

[15] ZHAO X, YANG M, XING Z, et al. Structure and content enhancement to Military Scenario Definition Language[C]// 2012 IEEE Symposium on Robotics and Applications (ISRA). 2012: 379-382.

[16] HUBER D, KALLFASS D. Applying data farming for military operation planning in nato MSG-124 using the interoperation of two simulations of different resolution[C]// 2015 Winter Simulation Conference (WSC). IEEE, 2015: 2535-2546.