

The Prediction and Improvement of Liquid Hydrocarbons Production Based on the Machine Learning Algorithm

Xinzhou Yang*, Zhenghe Yan, Ruiqiao Luo, Peng Yang, Shenglai Tang

CNOOC China Limited, Shenzhen Branch, Shenzhen Guangdong
Email: *yangxzh15@cnooc.com.cn

Received: Oct. 1st, 2020; accepted: Nov. 18th, 2020; published: Dec. 15th, 2020

Abstract

The natural gas extracted from the offshore gasfield is transported to the onshore gas treatment terminal through the submarine pipeline. After a series of treatment processes, propane, butane, light hydrocarbon and condensate are produced. All these liquid hydrocarbons production is not only related to the temperament component of each gas field, but also affected by the working condition of the land terminal equipment. Firstly, by analyzing the technological process of the land terminal, the precipitation of all kinds of hydrocarbons production was analyzed through clustering to find out the key equipment that had an impact on the recovery efficiency. Accurate calculation of liquid hydrocarbons production output is achieved by machine learning method. Finally, using the machine learning model, we can predict the production of each liquid hydrocarbon under the different working conditions, and provide the basis for the optimization direction of the process flow.

Keywords

Gas, Terminal Gas Byproduct, Abnormal Detection, Machine Learning, Prediction and Improvement

*通信作者。

基于机器学习实现海上气田陆地终端液态产品产量预测与挖潜

羊新州*, 闫正和, 罗睿乔, 杨 鹏, 唐圣来

中海石油深海开发有限公司, 广东 深圳

Email: *yangxzh15@cnooc.com.cn

收稿日期: 2020年10月1日; 录用日期: 2020年11月18日; 发布日期: 2020年12月15日

摘 要

海上开采出来的天然气通过海底管线输送到陆地天然气处理终端, 经过一系列处理工艺后, 生产量丙烷、丁烷、轻烃和凝析油等副产品。副产品的产出不光与各气田本身的气质组分相关, 同样受到陆地终端设备工况的影响。笔者首先通过分析陆地终端的工艺流程, 归纳影响终端副产品的关键流程。然后将各类副产品的析出情况通过聚类分析, 找出对各类副产品回收效率有影响的关键设备, 筛选出相应的异常值, 进行异常标注。再结合设备工况的标注信息, 通过机器学习方法实现对液态产品产量的精准计算。最后挖掘生产潜力, 预测在各设备完好条件下各液态产品的产量, 为工艺流程的优化方向提供基础。

关键词

天然气, 终端液态产品, 异常标记, 机器学习, 潜力挖掘

Copyright © 2020 by author(s), Yangtze University and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

南海某天然气陆地处理终端处理来自海上 5 个气田产量的天然气, 装置设计天然气处理能力达到 80 亿方/年。上岸天然气经过脱碳、脱水等处理之后, 进入液烃分馏单元, 实现重烃组分分离之后, 外输干气。分离的重烃组分经过多级分馏, 产出不同种类液烃产品。常规副产品的预测往往是通过简单的线性回归进行计算, 忽视了设备故障、参数波动、环境变化等因素的影响, 预测结果与实际副产品的产量存在一定差异。机器学习是一种用于设计复杂模型和算法并以此实现预测功能的方法, 它由数据分析习得, 而不依赖于规则导向的程序设计, 它能够基于对现有结构化数据的观察, 自行识别结构化数据中的模型, 并以此来输出对未来结果的预测。通过建立可靠的机器学习模型, 使计算机从数据中自动分析出相关规律, 并利用这些规律对未来的变化进行预测, 该方法已经在油气行业的各个方向进行了广泛的应用[1]-[10]。本文通过建立机器学习模型来学习各海上气田产量与终端各对应副产品产出的相关关系, 识别出影响副产品产出的关键设备, 继而对关键设备进行异常标注, 以提高副产品产量预测的准确度。

2. 终端液态产品处理设备异常标记

多个海上气田开采出的天然气在海上中心平台进行简单的脱水处理后，通过海底管道输送到终端，由海管登陆的天然气首先在天然气进站预处理单元经调压系统稳压后，进入段塞流捕集器进行气液分离。分出的凝液去凝析油稳定单元处理，生产稳定凝析油，稳定后的凝析油进入储罐储存，可装船或装车外输，亦可通过管道外输至精细化工；分离出的气相脱除夹带的微量汞和机械杂质后进入脱碳单元，脱碳后的湿净化气的 CO_2 含量控制在 2.8% 以下，随后进入两套并联的脱水、制冷、分馏单元处理，生产干气、丙烷、丁烷、液化石油气及稳定轻烃产品，干气通过天然气管道外输至广东管网，丙烷、丁烷、液化石油气、轻烃等液态产品进入储罐储存，可装船或装车外输，相关工艺流程见图 1 所示，其中，可以将副产品划分为 2 类，一为凝析油产品，一为丙烷、丁烷、轻烃等的液态烃类产品。

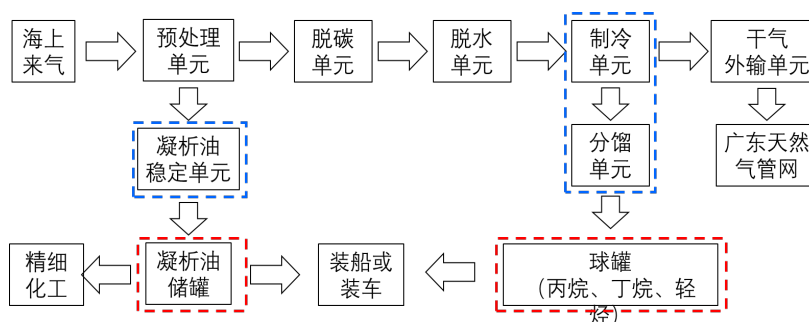


Figure 1. Schematic diagram of processing flow of natural gas processing terminal
图 1. 天然气处理终端处理流程示意图

聚类算法是指将一堆没有标签的数据自动划分成几类的方法，属于无监督学习方法[11][12][13]。一般的天然气藏在一定时间范围内，副产品的产量与天然气的产量比是一个相对稳定的值，当天然气产量一定时，对应的副产品产量接近一个定值，当副产品产量明显偏离这个定值时，则判断为设备出现了异常。本文通过收集海上各平台和陆地终端生产数据，经过聚类分析，划分出副产品产量的异常点。

本文采用 K 均值聚类算法，其步骤是，将各类副产品析出系数分为 2 组，随机抽取 2 个对象作为初始的聚类中心，然后将每个对象与 2 个中资聚类中心之间的距离，把每个对象分配给距离它最近的聚类中心。聚类中心以及分配给它们的对象就代表一个聚类。每分配一个样本，聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。终止条件可以是没有(或最小数目)对象被重新分配给不同的聚类，没有(或最小数目)聚类中心再发生变化，误差平方和局部最小。本文将该方法应用于终端及上游气田凝析油和液态烃类产品析出系数的异常判断，通过分析该终端及上游气田 19 年 1 月至 2020 年 5 月生产数据，判断结果如图 2 和图 3 所示。其中，在 2019 年 12 月，烃类产品的析出能力发生显著降低，通过对比设备工况，当时制冷单元设备出现异常，制冷效率为达到设定值，导致烃类产品产出量减少。可见，通过 K 均值聚类算法可以有效识别副产品的析出能力异常。

3. 终端液态产品产量预测

根据上述分析和现场对处理设备的异常标记，应用气田与终端的生产数据作为训练数据，用以预测 2020 年 1 月至 2020 年 5 月的副产品的产量。建模过程如图 4 所示。

3.1. 数据预处理

从图 1 可知，天然气从海上各气田采出输送到终端后，整个处理流程中，除了脱碳系统有化学反应

以外，其余工艺基本以物理变化为主。上岸天然气经过脱碳、脱水等处理之后，进入液烃分馏单元，实现重烃组分分离之后，外输干气。分离的重烃组分经过多级分馏，产量不同种类液烃产品。因此本次建模的初始特征选择海上5个气田的井口气产量和终端设备异常标记值，用以预测终端各副产品的产量。

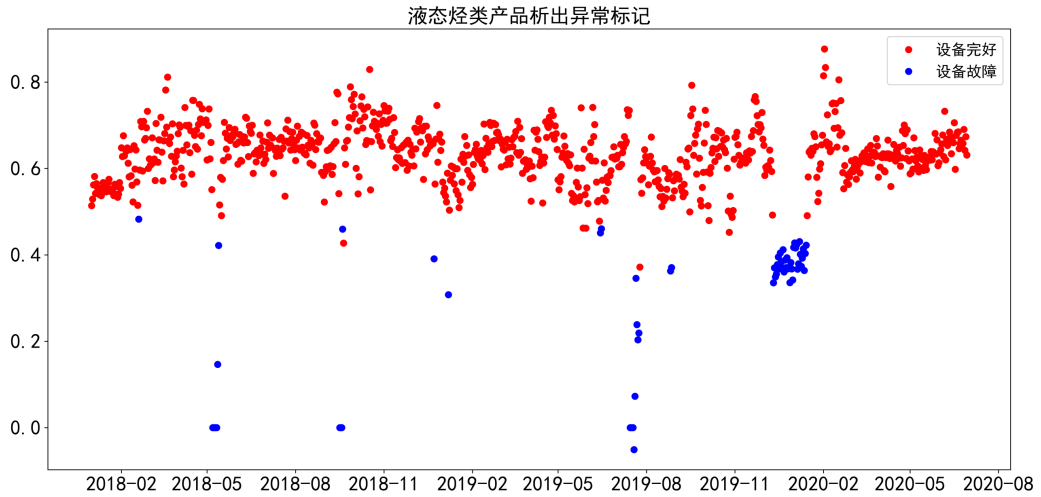


Figure 2. Scatter diagram: abnormal judgment of liquid hydrocarbon products precipitation
图 2. 液态烃类产品析出异常判断

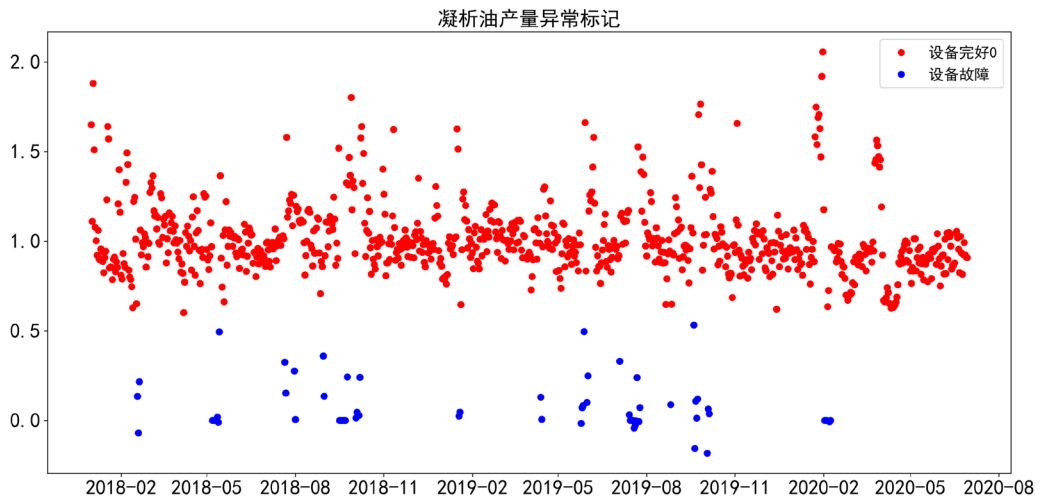


Figure 3. Scatter diagram: determination of condensate output anomaly
图 3. 凝析油产量异常判断

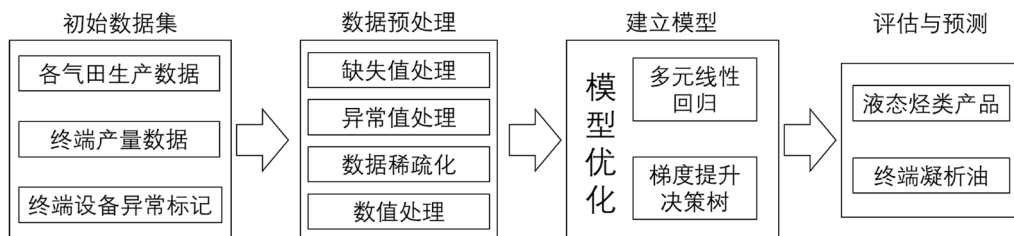


Figure 4. Process: the modeling process of predictive liquid hydrocarbon and condensate
图 4. 副产品预测建模预测流程

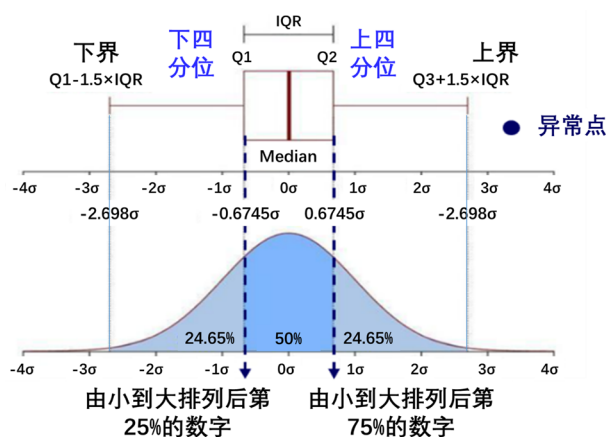


Figure 5. Curve: apply box diagrams for outliers determination
图 5. 应用箱型图进行异常值处理

应用箱型图来对各气田产量进行异常值判断，如图 5 所示。箱型图，主要包含六个数据节点，将一组数据从大到小排列，分别计算出他的上边缘，上四分位数 Q3，中位数，下四分位数 Q1，下边缘，还有一个异常值。它也可以粗略地看出数据是否具有对称性，分布的分散程度等信息，特别可以用于对几个样本的比较，无需对数据进行正态分布要求。异常值在箱型图中被定义为小于 $Q1 - 1.5IQR$ 或大于 $Q3 + 1.5IQR$ 的值，经过异常值处理后的气田群产量数据集如表 1 所示。

Table 1. Data set after outlier processing
表 1. 异常值处理后数据集

日期	A 气田井口气	B 气田井口气	C 气田井口气	D 气田井口气	E 气田井口气	烃类产品设备异常标记	凝析油处理单元异常标记
19-12-1	135.85	1010.97	230.99	68.4	121.55	1	1
19-12-2	134.68	1166.41	229.89	68.6	121.73	1	1
19-12-3	134.95	1181.02	229.39	68.89	122.07	1	1
19-12-4	134.86	1181.33	228.69	69.27	122.63	1	1
19-12-5	134.53	1179.38	225.34	69.51	123.19	1	1
19-12-6	132.64	1145.34	150.85	70.76	123.49	1	1
19-12-7	145.29	1065.17	212.96	69.81	125.68	1	1
19-12-8	157.24	1152.35	232.7	72.53	132.79	1	1
19-12-9	145.49	1126.21	225.67	69.71	127.7	1	1
19-12-10	116.41	1134.7	221.21	70.35	130.27	1	1
19-12-11	105.89	1113.94	210.57	70.48	129.76	0	1
19-12-12	99.75	1114.03	209.86	70.46	131.49	0	1
19-12-13	113.86	1090.89	211.12	70.51	131.07	0	1
19-12-14	137.78	1064.99	211.52	70.56	129.46	0	1
19-12-15	136.71	1056.99	211.39	68.67	128.31	0	1

为进一步提取数据集的相关特征，我们对各气田产量结合异常标记进行稀疏化处理，通过稀疏化实现特征的自动选择，去除影响最终预测的无用特征，即将无用特征对应的权重设置为 0，以凝析油数据集为例，数据稀疏化后处理结果见表 2。

Table 2. Modeling data set for predicting hydrocarbon product output after data sparsity
表 2. 数据稀疏化后预测烃类产品产量建模数据集

日期	A 气田 烃类设 备正常	B 气田 烃类设 备正常	C 气田 烃类设 备正常	D 气田 烃类设 备正常	E 气田 烃类设 备正常	A 气田 烃类设 备异常	B 气田 烃类设 备异常	C 气田 烃类设 备异常	D 气田 烃类设 备异常	E 气田 烃类设 备异常	终端产 品方
19-12-1	135.85	1010.97	230.99	68.4	121.55	0	0	0	0	0	1887.88
19-12-2	134.68	1166.41	229.89	68.6	121.73	0	0	0	0	0	1932.58
19-12-3	134.95	1181.02	229.39	68.89	122.07	0	0	0	0	0	2026.62
19-12-4	134.86	1181.33	228.69	69.27	122.63	0	0	0	0	0	2018.06
19-12-5	134.53	1179.38	225.34	69.51	123.19	0	0	0	0	0	2046.66
19-12-6	132.64	1145.34	150.85	70.76	123.49	0	0	0	0	0	1963.27
19-12-7	145.29	1065.17	212.96	69.81	125.68	0	0	0	0	0	1948.23
19-12-8	157.24	1152.35	232.7	72.53	132.79	0	0	0	0	0	2038.64
19-12-9	145.49	1126.21	225.67	69.71	127.7	0	0	0	0	0	1918.94
19-12-10	116.41	1134.7	221.21	70.35	130.27	0	0	0	0	0	1610.21
19-12-11	0	0	0	0	0	105.89	1113.94	210.57	70.48	129.76	1103.66
19-12-12	0	0	0	0	0	99.75	1114.03	209.86	70.46	131.49	1177.78
19-12-13	0	0	0	0	0	113.86	1090.89	211.12	70.51	131.07	1102.86
19-12-14	0	0	0	0	0	137.78	1064.99	211.52	70.56	129.46	1095.33
19-12-15	0	0	0	0	0	136.71	1056.99	211.39	68.67	128.31	1148

最后，选择 2019 年的生产数据作为训练集，2020 年的生产数据做测试集，进行机器学习建模。

3.2. 机器学习建模

将终端凝析油产量与终端液烃产品分开进行预测，对上述产品的预测即为一个回归问题，假设训练集样本 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ ，需要回归 (x_1, x_2, \dots, x_N) 与 (y_1, y_2, \dots, y_N) 之间的关系。用于回归的机器学习模型有很多，常用的有多元线性回归、支持向量回归、决策树回归、神经网络等，本次选取多元线性回归与梯度提升决策树进行回归预测。

在多元线性回归分析中，如果有两个或两个以上的自变量，其方程形式为：

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon \tag{1}$$

其中自变量的观测值为 $(1, x_{11}, \dots, x_{1p}), (1, x_{21}, \dots, x_{2p}), \dots$ ，对应的因变量观测值为 y_1, y_2, \dots ，通过引入矩阵来表示：

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \tag{2}$$

$$y = \beta X + \epsilon \tag{3}$$

多元回归即是需要求得相应的参数 β ，使得训练集的预测值与真实的回归目标值之间的均方误差最小。

提升树是迭代多棵回归树来共同决策，提升树模型可以表示为决策树的加法模型，计算方法如下：

- 1) 初始化提升树模型 $f_0(x) = 0$
- 2) 对 $m = 1, 2, \dots, M$
 - a) 计算残差
 - b) $r_{mi} = y_i - f_{m-1}(x_i), i = 1, 2, \dots, N$
 - c) 拟合残差 r_{mi} 学习一个回归树，得到 $T(x; \Theta_m)$
 - d) 更新 $f_m(x) = f_{m-1}(x) + T(x; \Theta_m)$
- 3) 得到回归问题提升树 $f_M(x) = \sum_1^m T(x; \Theta_m)$

分别应用上述 2 个方法对 2019 年的生产数据进行训练，然后通过 2020 上半年的气田生产数据预测终端液烃产品和终端凝析油的产量，结果如图 6、图 7 所示。

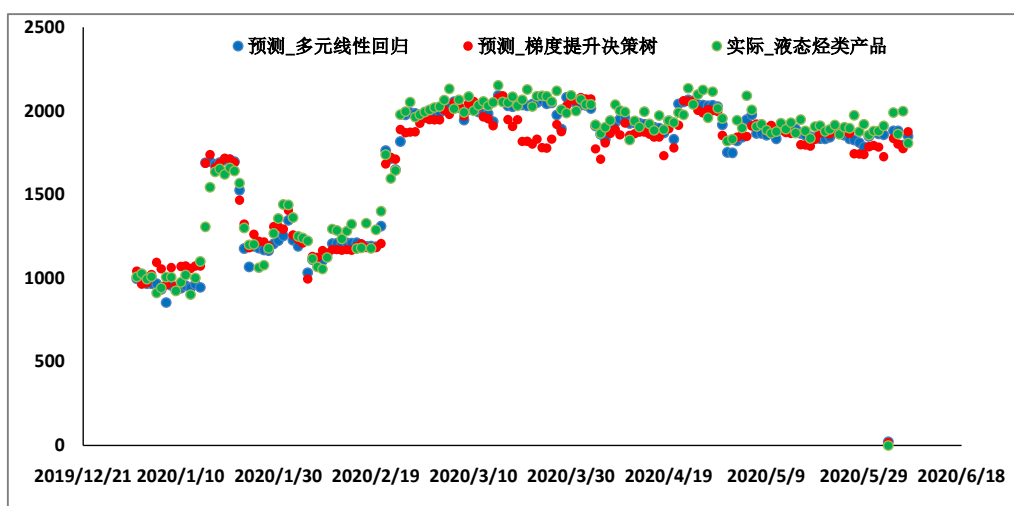


Figure 6. Scatter diagram: comparison between prediction of terminal hydrocarbon products and actual production

图 6. 2020 年预测终端烃类产品与实际对比

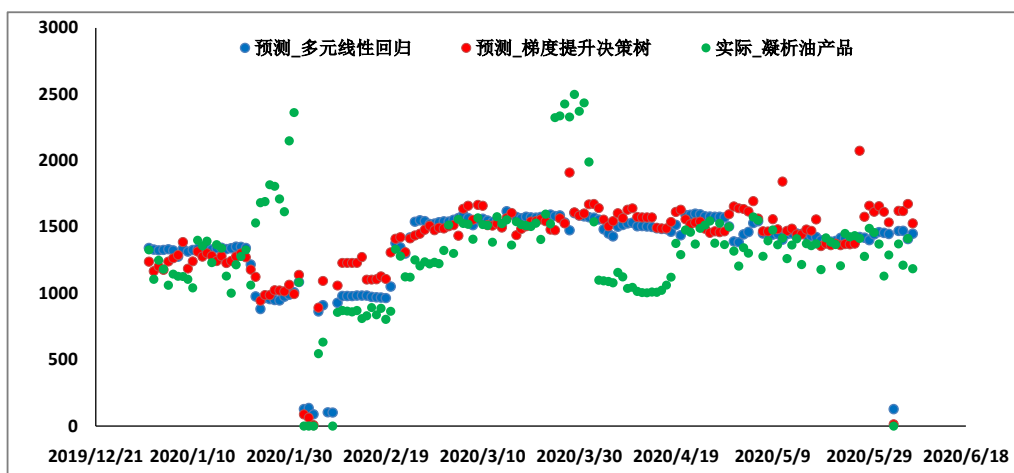


Figure 7. Scatter diagram: comparison between prediction of condensate oil and actual production

图 7. 2020 年预测终端凝析油与实际对比

由图 6 可以看出,无论是采用多元线性回归还是梯度提升决策树回归,预测的结果与终端烃类产品实际产量相一致。相比较来说,采用多元线性回归的方法更能准确的反应实际的产量波动。由图 7 可知,部分计算结果与终端凝析油的产量存在一定的差异,经过与现场核实,该波动主要是由于现场操作的变化,主动降低了凝析油稳定单元的液位导致的。截止到 2020 年 6 月,预测的累积凝析油产量,液态烃类产品产量与实际产量对比见图 8。

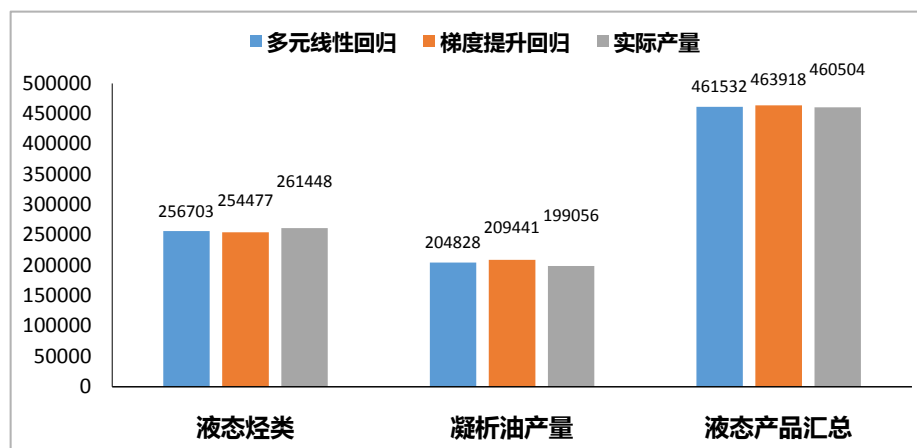


Figure 8. Histogram: comparison between prediction of cumulative production and actual cumulative production

图 8. 预测累积产量与实际累积产量对比

4. 终端液态产品产量潜力挖掘

根据对终端设备异常统计,2020 年液态烃类处理设备的异常已经出现了 15 天,每年凝析油处理设备的异常约为 10 天,因此,我们可以在预测模型中人为的修改设备的异常标记,假设在设备运作完好的条件下,结合年度的产量计划,预测最优运行条件下的液态烃类产品与凝析油的产量,各气田的年度产量计划如表 3 所示。

Table 3. Annual production plan for each gas field

表 3. 各气田年度产量计划

气田	A 气田	B 气田	C 气田	D 气田	E 气田
产量计划亿方	4.02	36.11	6.59	1.74	3.27

分别计算两种条件下终端各液态产品,计算结果如表 4 所示,可以看出,最优运行条件下比现有的液态产品产量总计大约增加了 2.26 万方,约为 14 万桶,按 30 美元/桶原油计算,可增加经济效益 420 万美元。

Table 4. Optimal production of liquid products under operating conditions

表 4. 最优工况条件下液态产品产量

	现有运行条件	最优运行条件
凝析油万方	43.42	44.81
液态烃类万方	59.57	60.45
液态产品汇总万方	102.99	105.26

5. 结论

1) 通过聚类分析进行析出比异常判断可以迅速定位出终端设备的异常情况, 结合后续对终端相关设备的异常标记, 建立机器学习模型, 可以精确预测各类液态产品产量。

2) 通过训练好的机器学习模型, 可以进行终端液态产品生产潜力挖掘, 预测在设备运行最优条件下液态产品产量, 预测后的结果可以为后续进行设备升级改造, 或者进行工艺优化提供决策依据。

参考文献

- [1] 武男, 陈东, 孙斌, 等. 基于分类方法的煤层气井压裂开发效果评价[J]. 煤炭学报, 2018, 43(6): 1694-1700.
- [2] 刘才俊, 胡洁, 宋述刚. 一种自适应变异策略的集体决策优化算法[J]. 长江大学学报(自然科学版), 2019, 16(10): 89-92+98.
- [3] 田冷, 何顺利, 顾岱鸿, 等. 改进 BP 神经网络模型在长庆气田产能预测中的应用[J]. 石油天然气学报, 2008, 30(5): 106-109+380.
- [4] 田亚鹏, 鞠斌山. 基于遗传算法改进 BP 神经网络的页岩气产量递减预测模型[J]. 中国科技论文, 2016, 11(15): 1710-1715.
- [5] 孙敬, 刘德华, 董春艳. 大数据挖掘技术评价致密气藏水平井产能[J]. 特种油气藏, 2016, 23(5): 74-77+154.
- [6] 刘占良, 石万里, 孙振, 等. 人工神经网络在气井管理及动态预测中的应用[J]. 天然气工业, 2014, 34(11): 62-65.
- [7] 吴永平, 昌伦杰, 郑广全, 等. 低渗裂缝性气藏产能分类方法[J]. 天然气地球科学, 2013, 24(6): 1220-1225.
- [8] 蒋裕强, 李成勇, 李志军. 基于模糊综合评判和 BP 神经网络的气井产能预测新模型[J]. 油气田地面工程, 2009, 28(10): 5-7.
- [9] 童凯军, 单钰铭, 李海鹏, 等. 支持向量回归机在气井产能预测中的应用[J]. 新疆石油地质, 2008, 29(3): 382-384.
- [10] 王建国, 何顺利, 兰朝利, 等. 低渗透气藏无阻流量的另一种求法[J]. 石油钻采工艺, 2006, 28(6): 61-64+4-5.
- [11] 王宏安, 陈国明. 基于 K 均值聚类的油气管道漏磁缺陷标记方法[J]. 科学技术与工程, 2020, 20(21): 8643-8646.
- [12] 高东升, 张辉, 雷霄, 等. 基于聚类差异分析的边底水气藏产水预测新方法[J]. 特种油气藏, 2017, 24(3): 109-113.
- [13] 王娟, 梁立星, 武璠, 等. 数据挖掘技术在数字化油气藏研究与决策中的应用[J]. 钻采工艺, 2018, 41(3): 107-109.