

基于变分自编码器与对比学习的油井产液量预测方法

杨晶, 查文舒, 李道伦*

合肥工业大学 数学学院, 安徽 合肥 230009

DOI:10.61369/ASDS.2026050008

摘 要 : 为解决油井生产时序数据伴随噪声干扰、动态演化具备非线性规律, 以及传统深度学习模型特征挖掘能力不足的现实问题, 本文结合变分自编码器 (VAE)、监督对比学习与门控循环单元 (GRU) 结构, 构建 VAE-CLP 混合时序预测模型。具体而言, 首先搭建变分特征重构网络, 整合井底压力、注水量、历史产液量等多维时序信息, 完成高维数据向低维连续潜在空间的映射转换。模型引入重参数化采样方案, 搭配 KL 散度完成约束调控, 让潜在变量贴合标准正态分布规律, 以此强化模型泛化能力。其次, 在该潜在空间中创造性地引入监督对比学习机制, 促使工况相近的样本在特征空间中自动聚集、差异样本相互远离, 强化模型对油田复杂生产状态的特征分辨效果。最后, 将解码器生成的重构时序序列导入双层 GRU 结构, 深度挖掘多变量协同演化里的长短时序关联, 实现油井未来产液量的高精度预测。选取油田现场多口油井的实测数据开展交叉验证, 该模型在测试样本中能够维持稳定运行状态, 整体实用价值与适配性表现优异。

关 键 词 : 产液量预测; 变分自编码器; 门控循环单元; 对比学习

Prediction Method of Oil Well Liquid Production Based on Variational Autoencoder and Contrastive Learning

Yang Jing, Zha Wenshu, Li Daolun*

School of Mathematics, Hefei University of Technology, Hefei, Anhui 230009

Abstract : To address the practical challenges including noise interference in oil well production time series data, nonlinear dynamic evolution characteristics, and inadequate feature mining capability of traditional deep learning models, this paper constructs a hybrid time series prediction model VAE-CLP by combining Variational Autoencoder (VAE), supervised contrastive learning and Gated Recurrent Unit (GRU). Specifically, a variational feature reconstruction network is firstly built to integrate multi-dimensional time series information such as bottom-hole pressure, water injection rate and historical liquid production, and realize the mapping of high-dimensional data to a low-dimensional continuous latent space. The model adopts the reparameterization sampling method and introduces KL divergence for constraint control, which enables latent variables to fit the standard normal distribution and strengthens the model's generalization ability. Secondly, a supervised contrastive learning mechanism is innovatively embedded into the latent space to aggregate samples with similar working conditions and separate heterogeneous samples in the feature space, so as to improve the model's ability to distinguish complex oilfield production states. Finally, the reconstructed time series generated by the decoder is imported into a two-layer GRU structure to deeply mine long-term and short-term temporal correlations in multivariate co-evolution, achieving high-precision prediction of future oil well liquid production. Cross-validation is performed with field measured data of multiple oil wells. The model maintains stable operation on test samples and presents outstanding practical value and adaptability.

Keywords : liquid production prediction; variational autoencoder; gated recurrent unit; contrastive learning

引言

近年来, 人工智能技术持续迭代升级, 机器学习方法在油气田产量预测领域得到广泛应用^[1-5]。吴新根等人^[6]最早将人工神经网络引入油田产量预测场景, 验证了机器学习方法在该领域的可行性。李道伦等^[7]依托神经网络隐式处理技术, 完成缺失时段测井数据的补

作者简介:

杨晶, 合肥工业大学数学学院, 硕士研究生, 研究方向为油井产液量预测;

查文舒, 合肥工业大学数学学院, 硕士生导师, 研究方向为油气流动模型与数值模拟方法。

通讯作者: 李道伦, 合肥工业大学数学学院, 博士生导师, 研究方向为油气大数据技术。

齐与推演,并进一步建立了隐式曲线与神经计算相结合的时序演化分析框架。Fan 等^[8]在考虑人为干预因素的前提下,结合 ARIMA 模型与长短期记忆网络,完成页岩气井产量相关推演工作。樊冬艳等^[9]基于时间序列相似性分析,采用 GRU 和 LSTM 网络对页岩气井产量进行预测,结果表明强相关序列能有效提高预测精度。

传统循环神经网络存在特征挖掘深度不足,模型解释力度偏弱等局限性,研究者开始探索更灵活的网络结构与跨领域技术融合路径^[10-14]。Hu 等人^[15]于2024年结合时间卷积网络与 KAN 网络搭建混合模型,利用特征选择机制,强化产量预测结果贴合油田物理机理解释能力。就算法对比层面而言,Ibrahim 等人^[16]在2022年对比多类回归算法,验证深度神经网络适配大规模井网数据的分类处理。杨莉等人^[17]以多变量时序数据为基础搭建预测框架。模型一方面挖掘深层特征完成时序推演,另一方面引入 LightGBM 搭建回归模块,设计超前参数递归策略。该方案脱离已知输入条件,依旧可以保障产量预测的精准程度。何佑伟等^[18]依托 K 近邻算法优化油井生产数据,完成降噪与降维处理。数据整体质量得到优化,模型预测输出效果同步加强。侯梦瑶等^[19]开展特征权重排序分析,确定压裂液用量、储层厚度两类指标,左右非常规气田产能变化,为后续特征筛选工作提供理论支撑。

深度学习模型虽在产量预测上取得显著成效,但在工程实践中仍有严峻挑战^[20-23]。首先,地质条件存在区域差异,注采调配方案各区块并不统一,单井产液状态容易产生非线性波动。不同油井的产能衰减规律具备独立特征,整体差异化表现突出。油气生产数据普遍呈现非平稳特质,海量样本清洗与标签标注消耗大量人力成本,高价值特征的提取工作随之增加难度。另外,常规监督学习多选用单点回归损失函数,无法捕捉样本之间的内在关联。特征空间划分界限模糊,模型泛化适应能力受到限制。生产制度出现调整或是地质环境产生极端变化,数据分布随之偏移,预测模型极易出现精度大幅下滑的问题。

针对油井产液量预测中数据高维性、噪声干扰及特征表征能力不足的问题,本文提出了一套融合变分自编码机制与监督对比学习的油井产液量高精度混合预测方法(VAE-CLP)。首先,依托全连接层搭建变分特征编码器^[24],将多变量的高维时序输入降维投影至一个平滑的概率潜在空间中,并同步利用重参数化采样为隐变量注入合理的统计随机性,以此大幅增强模型面对未见工况的泛化能力。其次,创新性地引入基于产液量相似度阈值的监督对比学习损失^[25],在潜在表征层嵌入了以真实产液量相似度为阈值的监督对比学习损失(Contrastive Loss),该机制强制具有相近目标产量的工况序列在隐空间内相互靠拢、差异悬殊的样本相互排斥,构建鲁棒特征空间。最终,模型利用 VAE 解码网络重构出高质量去噪序列,并将其送入由深层门控循环单元^[26]及全连接模块构成的时序预测网络中,实现对未来产液量的精准捕捉与预测。

一、方法概述

(一) 数据预处理

油井现场生产阶段,传感器异常、定期修井作业、注采制度调整等多种条件,都会干扰原始动态监测数据,致使数据样本存在大量缺失内容与持续性零值片段。将未经清洗的异常数据直接接入 VAE-CLP 混合模型,会打乱潜在空间的连续分布状态,造成 GRU 时序预测模块出现梯度弥散问题,同时误导模型拟合不符合现场机理的错误规律。

未投产作业、产能枯竭关井两类场景,容易让产液量序列首尾生成大范围连续零值,该类片段不具备实际生产参考价值。为剔除上述无效数据,本文对油井原始产液量序列做标准化定义,设定符号 $LPR = \{y_1, y_2, \dots, y_T\}$ 代表完整时序样本。检索序列首尾两端首次、末次非零有效数据节点,依托节点位置划定单井真实生产范围,以此界定绝对有效开采区间 $[t_{start}, t_{end}]$:

$$t_{start} = \min \{t | y_t \neq 0, t \in [1, T]\} \quad (1)$$

$$t_{end} = \max \{t | y_t \neq 0, t \in [1, T]\} \quad (2)$$

现场油井进入有效开采区间后,间歇性关井作业依旧会让时序曲线伴随零散零值波动。时序建模环节需要保障模型全程运行平稳,本文搭配序列长度指标、数据有效率指标,搭建双重阈值

约束筛选机制。

最小有效时间步约束:完成区间截断处理后的时序样本,统一核验全域时间步总长 L ,对应判定准则如下:

$$L = t_{end} - t_{start} + 1 \geq L_{min} \quad (3)$$

本次实验统一划定 L_{min} 基础数值为50。时序样本实测长度达不到该临界标准,就无法完整刻画油井全周期产能衰减演化规律,直接剔除该类低可用度样本。

有效率约束:为兼顾对正常短期修井停工的容忍与对工况极不稳定劣质井的排除,定义生产有效率 η ,如公式(5)所示,即截断区间内非零值所占比例:

$$\eta = \frac{1}{L} \sum_{t=t_{start}}^{t_{end}} \Pi(y_t \neq 0) \geq \eta_{threshold} \quad (4)$$

其中, $\Pi(\cdot)$ 为指示函数,条件成立取1,否则取0。设定有效率阈值 $\eta_{threshold} = 0.8$ 。若 $\eta < 0.8$ 则判定该井数据质量不达标,予以剔除。

经上述双重约束筛选,初始原始数据集即可迭代适配为规整度高、干扰项少的标准化优质样本集合。本实验从达标油井中选取8口作为研究对象。

为消除特征间量纲差异,筛选后的合规样本全部接入 Z-score 标准化算法完成归一化处理^[25]。该类标准化算法适配 VAE 概率生成底层数理逻辑,贴合模型固有机理,大幅降低编码器特征自主

学习难度，同步加快模型全局迭代收敛速率，适配油气时序建模全流程场景。具体数理转换公式如下：

$$x_{scaled} = \frac{x - \mu}{\sigma} \quad (5)$$

其中， x 为原始输入特征值； μ 为该特征序列的算术平均值； σ 为该特征序列的标准差； x_{scaled} 为无量纲化后的特征值。

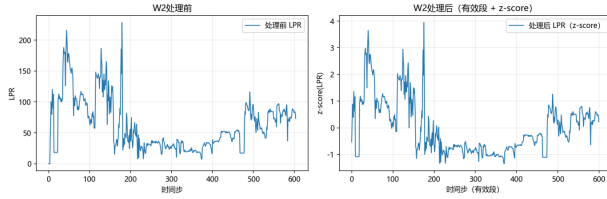


图1: W_2 处理前后数据对比

如图1所示， W_2 井原产液量时序起伏幅度偏高，整体量级分布跨度偏大。依次开展有效生产区间筛选、Z-score 全域归一化两类前置操作，能够剔除全段无效异常数值，抹平不同维度特征之间的量纲偏移。时序整体波动幅度收拢至统一标准区间，数据全域分布均衡性显著提升，为后续预测建模提供了更鲁棒的输入特征。

(二) VAE-CLP 模型整体结构

针对油井生产数据普遍存在的高噪声、非线性以及工况分布不均等问题，本章提出了一种名为 VAE-CLP (Variational Autoencoder - Contrastive Latent Prediction) 的混合预测模型。

该模型的核心设计理念在于先分布重构，后时序建模。传统预测模型直接依托带噪原始数据完成拟合训练，极易导致过拟合问题，面对非常规复杂工况适配能力偏弱。如图2所示，VAE-CLP 利用变分自编码器 (VAE) 将复杂的原始数据空间映射到一个连续、平滑的标准正态分布潜在空间 (Latent Space)。通过这一过程，模型能够强制过滤掉与正态分布不符的随机噪声，学习到数据背后的鲁棒特征表示。随后，利用门控循环单元 (GRU) 捕捉长短期时序依赖，从而在保证预测精度的同时，显著增强模型对不同数据类型的泛化适应能力。

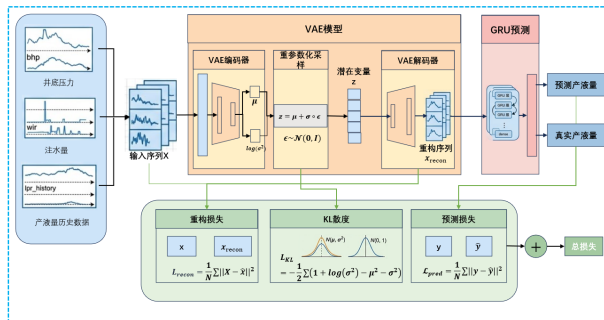


图2: VAE-CLP 模型架构示意图

VAE-CLP 模型采用级联式深度学习架构，主要由变分去噪模块和时序预测模块两部分组成。整体数据流向如下：输入层：接收原始的多变量时间序列数据 X (包含井底压力、注水量等特征)。变分特征重构网络：由编码器 (Encoder) 将输入映射为潜在分布参数 (μ, σ) ，借助重参数化技巧采样得到潜在变量 z 。在此阶段引入对比学习约束，优化潜在空间结构。解码器 (Decoder)

随后将 z 还原为去噪后的重构序列 $\hat{x}$ 。时序预测网络：GRU 单元接收重构序列 $\hat{x}$ ，利用其门控机制提取时间演变规律。输出层：基于 GRU 的隐状态输出最终的日产液量预测值 $\hat{y}$ 。

(三) 基于 VAE 的特征降维与序列重构

作为整个模型的前置处理模块，变分自编码器 (VAE) 的主要任务是对原始含噪数据进行非线性降维与分布平滑。

1. 序列编码

编码器采用多层感知机 (MLP) 结构，旨在推断后验分布 $p(z|x)$ 。为了使模型具备处理各种类型数据的能力，编码器输出两个特征向量：均值向量 μ 和对数方差向量 $\log(\sigma^2)$ ：

$$[\mu, \log(\sigma^2)] = f_{enc}(X) \quad (6)$$

式中， $f_{enc}(\cdot)$ 代表由多层全连接网络构成的深度编码器。

2. 重参数化采样

为使随机采样过程可导并支持端到端的反向传播训练，模型引入重参数化技巧。本研究添加一个服从标准正态分布的辅助噪声变量 $\epsilon \sim N(0, I)$ ，将潜在变量 z 的生成过程转化为确定性变换：

$$z = \mu + \sigma \circ \epsilon \quad (7)$$

这一步确保了模型的泛化能力。引入 ϵ 相当于在训练时向潜在空间注入微小扰动，解码器能学习到更加连续、鲁棒的特征流形。

3. 潜在空间监督对比学习

标准 VAE 算法仅能完成基础数据平滑重构作业，无监督运行模式容易让无关背景特征混入潜在空间 Z ，大幅降低特征表征区分能力。本文依托专属监督对比学习机制优化潜在变量 z ，提高模型预测精度。

机制运行以实测产液量差值作为强监督信号，重塑潜在空间整体拓扑结构。同批次工况样本纳入统一比对范围，设定 i 和 j 为任意两组对比样本。核算两组样本实测产液量差值，数值低于预设阈值即 $y_i - y_j < \delta$ ，判定两组样本互为正向匹配样本对，其余情况统一判定为负样本对。

本研究利用改进后的 InfoNCE 损失函数优化潜在特征间的余弦相似性：

$$L_{contrast} = -\sum_i \log \frac{\sum_{j \in P(i)} \exp(\text{sim}(z_i, z_j) / \tau_{temp})}{\sum_{k \in A(i)} \exp(\text{sim}(z_i, z_k) / \tau_{temp})} \quad (8)$$

式中， $\text{sim}(\cdot)$ 表示余弦相似度， $P(i)$ 为正样本集合。最小化该损失可使生产规律相近的工况在潜在空间内相互聚拢，同时使差异较大的工况彼此远离，形成物理意义一致的特征空间

4. 序列解码

解码器与编码器结构对称，负责将潜在变量 z 映射回原始数据空间，生成重构序列 $\hat{x}$ ：

$$\hat{x} = f_{dec}(z) \quad (9)$$

式中， $f_{dec}(\cdot)$ 代表由多层全连接网络构成的解码器函数。

(四) GRU 预测单元

时序建模环节依托 GRU 单元完成深层规律挖掘，模型将去噪

重构序列 $\hat{\mathbf{x}}$ 直接输入门控循环单元。相较于传统 LSTM 网络，GRU 整体参数体量更小，迭代收敛效率更突出，适配油井生产时序非线性强、波动节奏复杂的全域演化场景。GRU 依靠更新门 z_t 和重置门 r_t 控制信息流动：

$$z_t = \sigma(W_z \cdot [\hat{x}_t, h_{t-1}]) \quad (10)$$

$$r_t = \sigma(W_r \cdot [\hat{x}_t, h_{t-1}]) \quad (11)$$

$$\tilde{h} = \tanh(W \cdot [\hat{x}_t, r_t \odot h_{t-1}]) \quad (12)$$

$$h_t = (1z_t) \odot h_{t-1} + z_t \odot \tilde{h} \quad (13)$$

上式中，t 时刻的重构特征输入记为 $\hat{\mathbf{x}}_t$ 。GRU 擅长保留长期趋势并滤除短期波动。取最后一时步的隐状态 h_t 经由后置全连接层完成特征加权融合，直接输出对应时段油井日产量最终预测结果。

（五）损失函数

VAE-CLP 模型采用多任务联合损失函数，协同优化分布重构、特征提取及时序预测三部分。各损失定义如下：

1. 重构损失 (L_{recon})：实验选取均方误差 (MSE) 作为核心计量准则，使重构序列 $\hat{\mathbf{x}}$ 保留原始数据中的关键信息。

$$L_{recon} = \frac{1}{N} \sum ||X - \hat{x}||^2 \quad (14)$$

2. KL 散度 (L_{KL})：该项指标作为提升模型跨工况泛化性能的核心关键，量化计量自建潜在特征分布与标准正态分布 $N(0, I)$ 之间的偏离程度。最小化 KL 散度使隐变量分布趋近规整的正态分布，从而抑制过拟合，并使模型具备对未经历过工况的适应能力，对应计算公式如下：

$$L_{KL} = -\frac{1}{2} \sum (1 + \log(\sigma^2) - \mu^2 - \sigma^2) \quad (15)$$

3. 预测损失 (L_{pred})：用于衡量最终产液量预测值与真实值的差异。

$$L_{pred} = \frac{1}{N} \sum ||y - \hat{y}||^2 \quad (16)$$

4. 对比损失 ($L_{contrast}$)：如 2.3.3 节所述，负责调整潜在空间的结构。得到总损失函数 L_{total} 为：

$$L_{total} = L_{recon} + \beta L_{KL} + \lambda L_{pred} + \gamma L_{contrast} \quad (17)$$

通过调节超参数 β, λ, γ ，模型能够在数据保真度、分布规范性和预测准确性之间取得最佳平衡。

二、实验与分析

（一）数据集

本实验所用数据集来自多口油井实际生产历史时序记录。每口井记录井底压力、注水量与产液量三项关键连续动态变量，采样周期为月，时间跨度若干年。原始数据经异常值检测、缺失值

补齐及标准化预处理后，选用其中的 $W_1, W_2, W_3, W_4, W_5, W_6, W_7, W_8$ 八口井作为实验对象（下标为井号）。采用滑动窗口构造训练样本，窗口长度设置为 6 个月，即用前 6 个月数据预测下月产液量。按时间顺序以 5:1:4 比例划分训练集、验证集与测试集。

（二）实验设置

模型超参数：序列长度取 6，批量大小 32，VAE 隐层维度 64，潜在空间 16，GRU 隐层 128，层数 2，Dropout 率 0.2，学习率 0.0003。优化器选 Adam，实验在 PyTorch 框架上完成。

模型主要的超参数设置如表所示。

表：模型网络参数表

网络层	描述	数据维度
输入层 (Input Layer)	输入信号	(32, 6, 3)
编码器展平层 (Encoder Flatten)	形状重置 (Reshape)	(32, 18)
编码器全连接层 (Encoder FC)	全连接 + ReLU (Fully-connected + ReLU)	64
Z 采样层 (Z Sampling)	重参数化采样 (Reparameterization)	(32, 16)
解码器全连接层 (Decoder FC)	全连接 + ReLU (Fully-connected + ReLU)	64
解码器形状重置 (Decoder Reshape)	形状重置 (Reshape)	(32, 6, 3)
GRU 预测层 (GRU)	时序预测，2 层	(128, 1)
预测器全连接层 1 (Predictor FC1)	全连接 + ReLU + Dropout(0.2)	128
预测器输出层 (Predictor FC2)	全连接 (Fully-connected)	1

（三）评价指标

为衡量预测曲线与真实曲线的拟合程度及其解释力，该研究将拟合优度 R 平方（拟合 R^2 ）与均方根误差 (RMSE) 共同作为评价指标：RMSE 越小、拟合 R 平方越大，说明预测效果越好。其中拟合 R 平方的定义如下：

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (18)$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (19)$$

其中，样本真实值记为 y ，样本均值记为 $\bar{y}$ ，预测值记为 $\hat{y}$ ，样本量记为 n 。RMSE 越小表示预测精度越高， R^2 越接近 1 表示模型拟合效果越好。

（四）实验结果

1. VAE-CLP 模型实验结果

为全面评估 VAE-CLP 模型适配各类生产工况的泛化性能，本文选取实验集中 8 口目标油井开展逐井验证。如图 3 所示，模型对各井产液量的拟合优度 R^2 整体表现优异，指标最低值为 83.45%，最高值达到 95.11%，多数测试井该项数值稳定维持在 90% 以上。依托 VAE 潜在空间分布重构方法并结合对比学习完成模型优化，可让模型适配各类未知复杂测试井数据，同时具备稳定的数据预测与自适应分析能力。

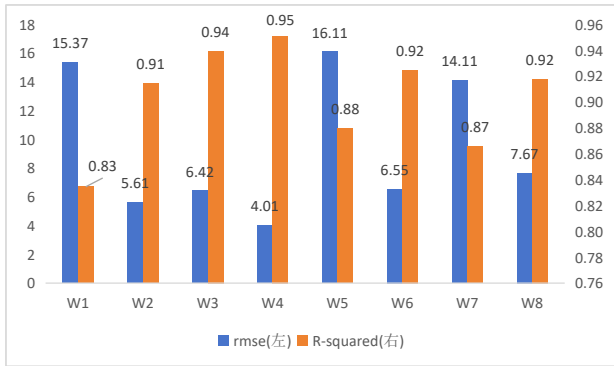


图3: 井产液量预测 RMSE 和 R²值

此外，各测试井均方根误差（RMSE）集中落在 4~20 m³/d 区间。 W_4 井 RMSE 数值达到区间最低，仅为 4.01 m³/d； W_1 井预测难度在所有样本中最高，其 RMSE 峰值也仅 15.37 m³/d，整体误差波动幅度保持平稳。各油井 RMSE 数值出现分化，一方面是部分测试井在地层能量层面出现剧烈波动，同时注水量参数也经过多次调整。井底压力和历史产液量时序数据，容易受现场工况影响混入大量动态噪声。依托 VAE 的去噪机制，模型面对极端工况依旧可以把预测精度稳定控制在合理区间。

图4展示了模型在所有测试集上的真实产液量与预测产液量对比散点图，从图中可以直观地观察到，大量测试样本的散点紧密、均匀地分布在理想预测线附近，极少出现严重的偏离点，这表明模型的预测值与真实值具有高度的一致性，其 R^2 高达 0.9639，说明该混合模型能够解释超过 96% 的产液量数据变异，极好地捕捉到了油井生产的动态演化规律；同时，全局均方根误差（RMSE）控制在了 12.01 m³/d 的较低水平，表明模型在具体数值的预测上具有极高的绝对精度。

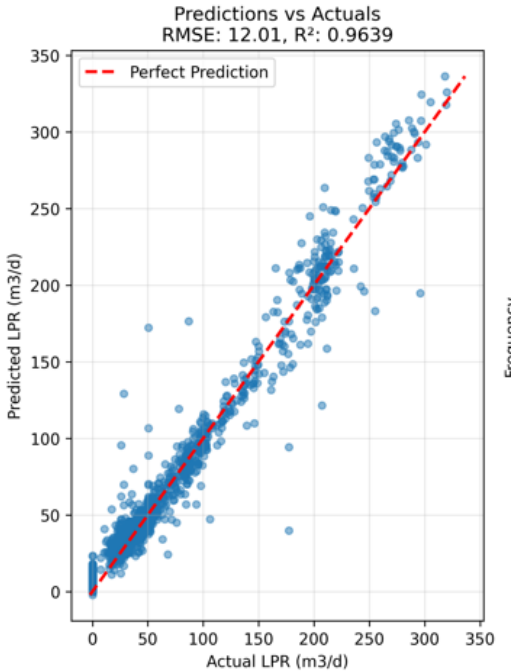


图4: 真实产液量与预测产液量对比散点图

本文选取了其中两个井，绘制其产液量随时间变化的实际值与预测值对比图，如图5所示，预测时间跨度约为1998年至2016

年。从整体趋势层面分析，两口油井的预测曲线和实际曲线在变化走向与波动周期上高度契合，能够体现模型对油井产液量时序变化具备良好拟合效果。

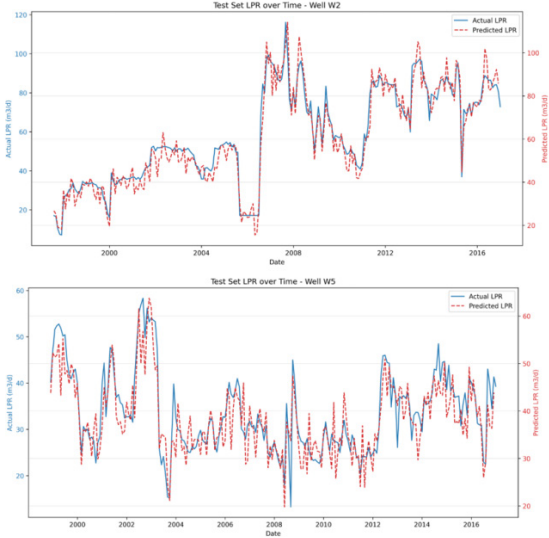


图5: W_2 与 W_3 预测值与真实值曲线图

2. 对比试验

为充分验证 VAE-CLP 模型具备的性能优势与实际有效性，本研究选取四类经典时序预测主流模型开展对照实验。参与对比的模型包含长短期记忆网络（LSTM）、门控循环单元（GRU）以及时间卷积网络（TCN）^[27]，以及融入特征约束机制构建的 GRU 模型（GRU-constration）。各模型在测试集上的预测误差（RMSE）对比结果如图6所示。

为保证实验对比具备公平性，四类基线模型统一采用本文设定的滑动时间窗口，直接引入井底压力、注水量、历史产液量等原始时序特征，完成端到端预测流程。LSTM、GRU 依靠循环结构与门控机制，挖掘数据内部长期依赖关系；TCN 借助因果膨胀卷积（Dilated Causal Convolutions）提取多尺度局部特征。但这些模型均不具备本文提出的变分潜在空间重构去噪（VAE）与工况物理一致性约束（对比学习）能力。

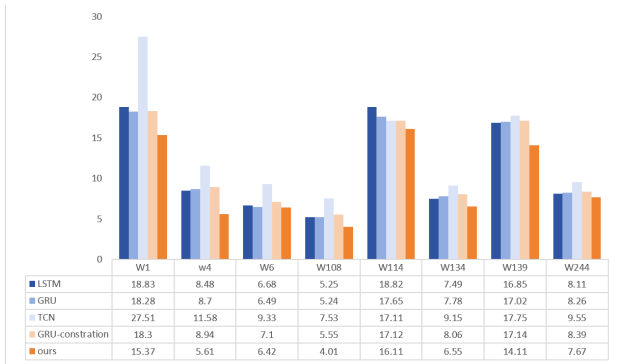


图6: 不同模型下实验结果

实验结果表明，全部8口独立测试井中，本文所提 VAE-CLP 混合模型均实现了最低预测误差，整体精度全面优于其他四种基

线算法。但针对预测难度极大的复杂工况井（如 W_1 和 W_2 井），传统时序模型极易受高频非线性噪声干扰，导致特征提取失效且预测偏差显著；本文模型得益于 VAE 核心模块强大的概率重构与去噪机制，能够有效滤除瞬态波动并提取核心演化趋势。GRU 和 LSTM 模型在 W_1 井的 RMSE 分别为 18.28 和 18.83，效果偏差，在其余水井上 RMSE 相差不大，比较稳定。这充分证明本文模型在应对剧烈波动的时序数据时具有极强的鲁棒性。

TCN 模型在基础对比中整体表现较差， W_1 井的 RMSE 高达 27.51。其膨胀卷积结构对局部瞬态噪声过于敏感，容易放大异常跳变。LSTM 与 GRU 表现相对接近，例如 W_4 井上分别为 5.25 和 5.24，但二者缺乏对原始特征的非线性降维与分布平滑机制，整体预测精度受噪声干扰所限，难以进一步提升。

从 GRU-constration 模型的表现来看，仅在 GRU 基础上加入约束机制（未融合 VAE 分布重构），其在多数井（如 W_3 、 W_5 ）上的预测误差与基础 GRU 基本持平，并未带来显著性能增益。本文模型通过“VAE 去噪重构 + 对比学习聚类”的深度协同，各井预测误差均显著低于 GRU-constration 模型。这印证了一点：处理高噪非平稳时序数据，仅靠单一时序网络与简单表层数值约束远远不够。要让模型真正具备应对复杂变工况的泛化能力与鲁棒性，需深入挖掘数据底层演化规律：借助变分自编码器提取剥离瞬态噪声的深度特征，结合对比学习机制进一步学习具有强判别性的鲁棒特征表征，才能实现预测精度与模型鲁棒性的双重提升。

综上，本文构建的模型成功解决传统深度学习算法应用于油

气产量预测时，易受噪声干扰、特征表征模糊的核心问题，在跨井泛化能力与预测精度两方面均呈现最优表现。

三、结论

本文提出一种基于变分自编码器与对比学习（VAE-CLP）的油井产液量混合预测方法，旨在解决未知、复杂工况下不同油井生产数据噪声大且分布不一致的问题。

本文设计基于 VAE 的变分特征重构网络，引入 KL 散度约束潜在变量，使其服从标准正态分布。该网络可从不同油井数据分布中提取鲁棒不变特征，同时平滑单井内部因注采波动产生的动态差异。潜在空间内，本文额外设计监督对比学习机制，通过自适应调整相似工况样本间距，强化模型对物理一致性特征的自主学习能力。重参数化采样技巧的应用，可在潜在空间生成多样化合理工况样本，让模型训练过程中更聚焦核心生产演变规律，大幅提升鲁棒性与泛化能力。经 VAE 重构后的高质量时间序列，输入 GRU 网络，完成长短期时序建模与产液量预测任务。

依托油田多井实际生产数据集开展测试，VAE-CLP 模型面对各类未知工况，均可以保持较高拟合优度与稳定的误差水平。相较于现有常规深度学习时序算法，该模型在实际应用的有效性、与场景适配性上更具优势。后续研究将维持模型现有泛化稳定性，进一步探索深度学习框架与油藏物理机理的深度融合路径，在变工况条件下进一步提升产量预测精度。

参考文献

- [1] 樊宏恩, 张蕾, 米兰, 等. 人工智能在全球油气工业领域的应用现状与前景展望 [J]. 石油钻采工艺, 2021, 43(4): 405-419.
- [2] 匡立春, 刘合, 任义丽, 等. 人工智能在石油勘探开发领域的应用现状与发展趋势 [J]. 石油勘探与开发, 2021, 48(1): 1-11.
- [3] 李阳, 廉培庆, 薛兆杰, 等. 大数据及人工智能在油气田开发中的应用现状及展望 [J]. 中国石油大学学报 (自然科学版), 2020, 44(4): 1-11.
- [4] 胡晓东, 涂志勇, 罗英浩, 等. 拟合函数-神经网络协同的页岩气井产能预测模型 [J]. 石油科学通报, 2022, 7(3): 394-405.
- [5] 梁潇. 基于双向递归神经网络模型的产量预测方法研究 [J]. 北京石油化工学院学报, 2025, 33(04): 41-47. DOI: 10.19770/j.cnki.issn.1008-2565.2025.04.008.
- [6] 吴新根, 葛家理. 应用人工神经网络预测油田产量 [J]. 石油勘探与开发, 1994, 21(3): 75-78.
- [7] 李道伦, 刘旭亮, 查文舒, 等. 基于卷积神经网络的径向复合油藏自动试井解释方法 [J]. 石油勘探与开发, 2020, 47(3): 1-8.
- [8] Fan D, Sun H, Yao J, et al. Well production forecasting based on ARIMA-LSTM model considering manual operations[J]. Energy, 2021, 220: 119708.
- [9] 樊冬艳, 杨灿, 孙海, 等. 基于时间序列相似性与机器学习方法的页岩气井产量预测 [J]. 中国石油大学学报 (自然科学版), 2024, 48(03): 119-126.
- [10] Sun J, Ma X, Kazi M. Comparison of decline curve analysis DCA with recursive neural networks RNN for production forecast of multiple wells[C]//SPE western regional meeting. SPE, 2018: D041S012R009.
- [11] 高亚军, 唐力辉, 王振鹏, 等. 基于循环神经网络和数据差分处理的油田产量预测方法 [J]. 中国海上油气, 2023, 35(3): 126-136.
- [12] 张昀, 陈彦润, 陈晓刚, 等. 数据驱动的气井井筒积液与产量预测模型研究及应用 [J]. 天然气与石油, 2025, 43(01): 9-19.
- [13] Sagheer A, Kotb M. Time series forecasting of petroleum production using deep LSTM recurrent networks[J]. Neurocomputing, 2019, 323: 203-213.
- [14] 石博玮, 郑恒, 李军亮, 等. 基于机器学习算法的压裂参数优化与产量预测 [J]. 科学技术与工程, 2025, 25(25): 10663-10670.
- [15] Hu Y, Xin X, Yu G, et al. Deep insight: an efficient hybrid model for oil well production forecasting using spatio-temporal convolutional networks and Kolmogorov-Arnold networks[J]. Scientific reports, 2025, 15(1): 8221. DOI: 10.1038/S41598-025-91412-2.
- [16] Ibrahim N M, et al. Well Performance Classification and Prediction: Deep Learning and Machine Learning Approaches [C]. Italian National Conference on Geoscience and Petroleum Engineering, 2022.
- [17] 王洪亮, 李宁, 李欣, 等. 基于数据驱动的致密油藏产能预测方法 [J]. 特种油气藏, 2025, 32(02): 82-88.
- [18] 何佑伟, 贺质越, 汤勇, 等. 基于机器学习的页岩气井产量评价与预测 [J]. 石油钻采工艺, 2021, 43(4): 518-524.
- [19] 侯梦瑶, 潘晓甜, 张春晓, 等. 可解释机器学习在油气产量预测中的研究进展 [J]. 石化工业应用, 2023, 42(10): 5-8, 25.
- [20] 梁潇. 基于双向递归神经网络模型的产量预测方法研究 [J]. 北京石油化工学院学报, 2025, 33(04): 41-47. DOI: 10.19770/j.cnki.issn.1008-2565.2025.04.008.
- [21] 刘保磊, 张心怡. 油气产量预测方法与技术研究进展 [J]. 中国石油勘探, 2025, 30(05): 128-144.
- [22] Chen Y, Chen P, Wang Y, et al. From clustering to federated learning: A collaborative approach to optimizing data-driven models for the production monitoring of unconventional natural gas wells[J]. Gas Science and Engineering, 2025, 137: 205580.
- [23] 杨莉, 周子希, 王婷婷, 等. 基于 CNN-GRU-LightGBM 模型的单井产量预测方法 [J]. 科学技术与工程, 2024, 24(18): 7606-7614.
- [24] Kingma PD, Welling M. Auto-Encoding Variational Bayes[J]. CoRR, 2013, abs/1312.6114.
- [25] R. H. S. C. Y. L. Dimensionality reduction by learning an invariant mapping[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, 21735-1742. DOI: 10.1109/CVPR.2006.100.
- [26] Dey R, Salemt F M. Gate-variants of Gated Recurrent Unit (GRU) neural networks[J]. IEEE, 2017: 1597-1600. DOI: 10.1109/MWSCAS.2017.8053243.
- [27] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint arXiv:1803.01271, 2018.