DOI:10.11918/202410036

采用 FD 约束迁移生成虚拟样本的动液面预测

王 通[1],胡安邦[1],韩 莹[2]

(1.沈阳工业大学 电气工程学院,沈阳 110870;2.辽宁工程技术大学 电气与控制工程学院,辽宁 葫芦岛 125105)

摘 要:为改善因石油开采过程复杂多变,难以获取完备油井生产数据信息导致的动液面预测模型效果不佳问题,提出一种基于模糊决策(fuzzy decision,FD)约束迁移的变分自编码器-生成对抗网络(variational autoencoder-generative adversarial network,VAE-GAN)模型,用于生成油井动液面预测建模虚拟样本。首先,利用信息较为完备的相似工况源域油井数据,基于迁移学习的 VAE-GAN 生成虚拟样本,并采用最大均值差异使虚拟样本特征分布向待测目标域油井迁移;其次,为使虚拟样本具有源域油井的丰富信息特征并符合目标域油井生产规则,设计了模糊判别约束模块用于辨别虚拟样本参数间生产特征,进一步约束虚拟样本生成;最后,将所得虚拟样本与目标域少量真实样本混合,构建动液面预测模型。仿真结果表明:本文所提方法生成的虚拟样本可以有效补充待测油井的生产数据,相较于采用待测油井真实数据构建的预测模型,其构建的预测模型在满足油田生产关键参数测量相对误差 5% 以内要求的基础上,平均绝对误差下降 29.589%,均方根误差下降 26.386%。所提方法可以有效解决油井生产数据信息不完备问题,能够降低动液面预测模型的输出误差。

关键词:动液面;虚拟样本;模糊决策;迁移学习;变分自编码器;生成对抗网络

中图分类号:TE319 **文献标志码**:A **文章编号**:0367-6234(2026)09-0183-13

Dynamic liquid level prediction using virtual samples generated by FD-constrained transfer

Wang Tong[1], Hu Anbang[1], Han Ying[2]

(1. School of Electrical Engineering, Shenyang University of Technology, Shenyang 110870, China;
2. Faculty of Electrical and Control Engineering, Liaoning Technical University, Huludao 125105, Liaoning, China)

Abstract: To solve the problem that the dynamic fluid level prediction model is ineffective due to the complex and changeable oil production process and the difficulty of obtaining complete oil well production data information, a variational autoencoder-generative adversarial network (VAE-GAN) model based on fuzzy decision constraint transfer is proposed to generate virtual samples for the dynamic liquid level prediction modeling of oil wells. Firstly, the oil well data in the source domain with relatively complete information is used to generate virtual samples through a VAE-GAN based on transfer learning. Meanwhile, the maximum mean discrepancy is used to transfer the feature distribution of virtual samples toward the target domain to be measured. Secondly, to ensure that the virtual samples possess the rich information characteristics of the source domain oil wells while conforming to the production patterns of the target domain, we design a fuzzy discriminant constraint module to distinguish the differences between the virtual samples and the data in the target domain and to further constrain the generation of the virtual samples. Finally, the obtained virtual samples are mixed with a small number of real samples in the target domain to construct the dynamic liquid level prediction model. The simulation results demonstrate that the virtual samples generated by the proposed method can effectively augment the production data of the target oil well. Compared to the prediction model constructed using actual data from the target oil well, the model developed with this method achieves a 29.589% reduction in mean absolute error and a 26.386% reduction in root mean square error, while ensuring that the relative error of key oil field production parameters remains below 5%. The proposed methodology effectively tackles the issue of incomplete information in oil well production data, thus mitigating the output error associated with the dynamic fluid level prediction model.

Keywords: dynamic liquid level; virtual sample; fuzzy decision; transfer learning; variational autoencoder; generative adversarial network

收稿日期:2024-10-16;**录用日期**:2024-11-30;**网络首发日期**:2025-01-10
网络首发地址:https://link.cnki.net/urlid/23.1235.t.20250109.1659.002
基金项目:国家自然科学基金(62203197)
作者简介:王 通(1976—),男,副教授,硕士生导师
通信作者:胡安邦,961062671@qq.com

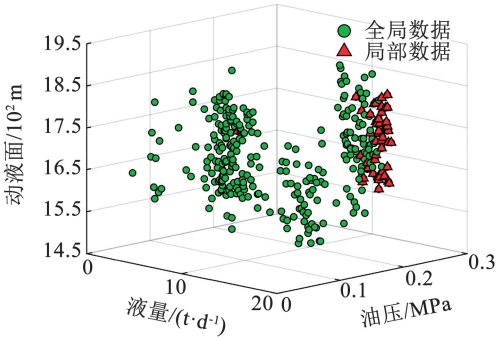
石油作为国家重要的战略资源,确保其安全、稳定生产具有重要战略意义。在石油开采工作中,油井动液面是反映油层供液能力和井下排供关系的重要参数,通过动液面数据可以分析油井运行状态,合理制定油井生产工艺[1]。相较于操作复杂、成本较高的动液面声波测量法[2],基于数据驱动的测量方法因其操作简便、时效性强、灵活性高等优势逐渐在油井动液面测量中得到应用。Leng 等[3]通过提取动态和静态特征实现对动液面的预测。Jia 等[4]经特征融合并结合机理公式实现多油井的动液面预测。侯延彬等[5]利用机理分析提取多尺度状态特征进行建模以实现动液面预测。Tian[6]利用核主成分分析法消除输入变量间冗余性以提升动液面预测精度。然而,油井生产过程的复杂性和不可重复性致使油井生产数据存在样本缺失或相似数据量较大但信息不足等问题,导致基于数据驱动的测量模型在油井动液面应用中泛化性和鲁棒性较差,极大地制约了模型的现场应用推广。

近些年,针对预测模型建模样本信息不完备问题,主要采用数据增强、迁移学习、集成学习、虚拟样本等方法进行处理。其中,虚拟样本生成(virtual sample generation, VSG)[7]技术通过捕捉和模仿原始数据的分布特征,生成具有相似特性的样本来扩充数据集,在增强数据多样性的同时,能够提高预测模型的泛化性和精确性。其主流方法有变分自编码器(variational autoencoder, VAE)[8]、生成对抗网络(generative adversarial network, GAN)[9]以及基于两者的改进应用。Zhao 等[10]针对工业过程数据的动态时变问题,利用多尺度卷积神经网络替代 VAE 中编码器,有效地增强了虚拟样本的动态适应性。Oliveira 等[11]在属性信息约束下构建修改语法 VAE,使生成数据能够反映输入分子的性质。Yoon 等[12]提出了一种用于时间序列数据生成的 TimeGAN,通过结合无监督范式的灵活性与监督训练的可控性生成真实数据。Gao 等[13]针对数据分布的不平衡问题,采用基于梯度惩罚的 Wasserstein-GAN 来补充数据集,以提高模型精度。Lee 等[14]结合域适应方法与卷积变分自编码器,将大量源域知识迁移到目标域,生成虚拟样本数据用于预测模型构建。Zhu 等[15]针对不同类型数据间的生成问题,结合迁移学习思想提出了 CycleGAN,其利用循环一致性损失函数生成与目标域数据类型一致的虚拟样本。Muhammad 等[16]利用迁移学习提高生成模型的重构性能,通过大量源域数据生成目标域数据。孙玲俊等[17]将 GAN 模型置于源域进行预训练以提取跨域底层特征,并生成多样化目标域虚拟样本。由上述研究可知,在不同领域下,通过虚拟样本生成技术解决建模样本数据不足的问题已取得了显著进展。但油井生产过程复杂多变,难以获取待测油井完备的生产信息,导致生成模型输出的虚拟样本多样性不足,且部分虚拟样本不符合待测油井生产参数间的机理关系,使得基于虚拟样本构建的动液面预测模型性能无法得到有效提高。

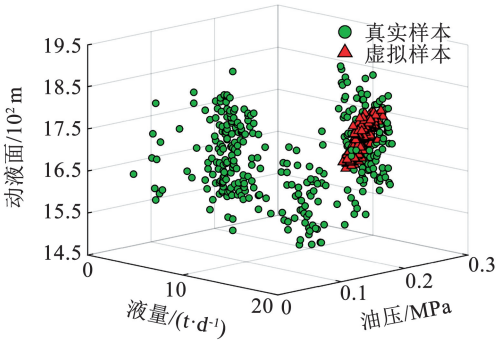
针对上述问题,本文提出一种基于模糊决策(fuzzy decision, FD)约束迁移的虚拟样本生成方法,并在此基础上进行油井动液面预测。首先,为解决待测油井生产数据信息量不足的问题,提出基于结合迁移学习的变分自编码器-生成对抗网络(variational autoencoder-generative adversarial network, VAE-GAN)模型,将相似工况源域油井数据的丰富信息迁移至待测目标域油井,并利用最大均值差异(maximum mean discrepancy, M_{md})生成与待测油井数据相匹配的虚拟样本;其次,为增加生成虚拟样本的多样性并符合待测油井生产特征,提出基于 FD 的判别约束模块,通过模糊决策判别样本参数间生产规则,进一步约束生成模型输出符合待测目标域油井生产特征的虚拟样本;最后,采用基于 FD 约束迁移的 VAE-GAN 模型生成虚拟样本,构建动液面预测模型,并选用辽河油田某油井生产数据验证其有效性。

1 动液面建模数据分析

传统油井动液面预测模型建模数据来源于油井生产可测量参数,但仅拥有待测油井当前生产历史数据无法使动液面预测模型捕捉到该油井全局生产信息,待测油井部分参数(液量、油压和动液面)的全局与局部数据分布如图 1(a)所示。



(a) 全局与局部数据



(b) 真实与虚拟样本

图 1 油井生产数据分布

Fig. 1 Distribution of well production data

为获得更充足的建模样本，通过传统 VSG 方法捕捉并学习待测油井局部数据间特征，生成含有待测油井生产信息的虚拟样本。待测油井全局样本与基于局部数据所生成虚拟样本分布如图 1(b)所示。同时，根据王通等[18]采用的油井动液面与辅助参数间机理关系，发现所生成虚拟样本中存在异常数据，即虚拟样本与真实油井参数间的生产特征不一致，如图 2 中标记所示，无法用于构建动液面预测模型。

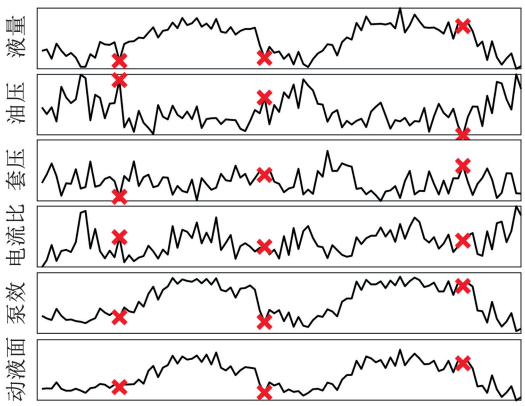
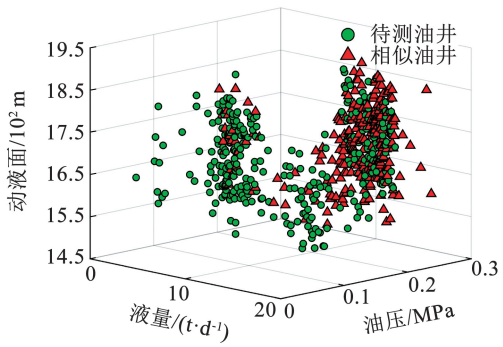


图 2　虚拟样本异常数据

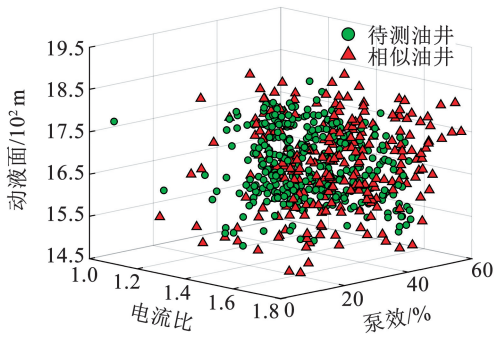
Fig. 2　Abnormal data of virtual samples

综上，传统 VSG 方法无法获取待测油井局部数据以外的生产信息，生成虚拟样本难以反映待测油井实际生产的完整状况。同时部分虚拟样本不符合待测油井生产特征，无法解决动液面预测模型效果不理想的问题。

在油田生产中，存在与待测油井位于相同油藏环境且有相似生产工况历史过程的其他油井，两者数据分布如图 3 所示。



(a) 液量、油压与动液面



(b)　电流比、泵效与动液面

图 3　相似油井数据分布

Fig. 3　Data distribution of similar wells

由图 3 可知，工况相似的不同油井间全局生产数据存在重叠，同时两者生产参数间的机理关系相似。表明两者在开采过程中具有类似的生产规律和特征，可以通过学习相似工况油井数据中丰富的生产信息，帮助待测油井补充预测模型建模所缺失的生产数据信息。在此基础上可以对生成过程进行约束，使基于相似工况油井数据生成的虚拟样本符合待测油井的实际生产特征，以提高动液面预测模型的准确性和泛化性。

2　VAE-GAN

VAE-GAN 无监督生成模型采用 VAE 代替 GAN 中的生成器部分，使整个模型同时具有 VAE 可控生成的能力和 GAN 生成高质量数据的性能[19-21]，其基本结构如图 4 所示。

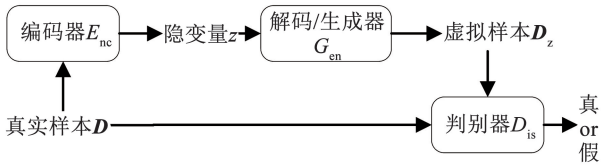


图 4　VAE-GAN 结构

Fig. 4　Structure of VAE-GAN

采用 VAE-GAN 对待测油井动液面预测模型建模样本 $\boldsymbol{D}$ 进行扩充过程中，VAE 通过对潜在变量 $p(z)$ 施加先验，使编码器 E_{nc} 正则化，并经过重构获得后验分布 $q(z|\boldsymbol{D})$，其中先验正则化项 L_{KL} 的表达式为

$$L_{KL}=D_{KL}(q(z|\boldsymbol{D})\parallel p(z)) \tag{1}$$

式中：$q(z|\boldsymbol{D})$ 服从 $E_{nc}(\boldsymbol{D})$ 分布，即 $q(z|\boldsymbol{D})\sim E_{nc}(\boldsymbol{D})$；$D_{KL}$ 为 KL(Kullback-Leibler)散度。

GAN 的目标是使判别器 D_{is} 能够区分出真实样本 $\boldsymbol{D}$ 和虚拟样本 $\boldsymbol{D}_z$，使解码/生成器 G_{en} 生成虚拟样本分布能够拟合真实分布。其目标函数是最大化/最小化二元交叉熵(binary cross entropy，BCE)，即

$$L_{GAN}=E_{\boldsymbol{D}\sim P_{\boldsymbol{D}}}[\log D_{is}(\boldsymbol{D})]+E_{z\sim P_G}[\log(1-D_{is}(G_{en}(z)))] \tag{2}$$

式中：E 为期望，$D_{is}(\boldsymbol{D})$、$D_{is}(G_{en}(z))$ 分别为判别器 D_{is} 对真实样本和虚拟样本的判断结果。

采用判别器 D_{is} 对重构误差 L_{llike}^{Dis} 进行表示。

$$L_{llike}^{Dis}=-E_{z\sim q(z|\boldsymbol{D})}[\log p(D_{is}(\boldsymbol{D})|z)] \tag{3}$$

式中 $p(D_{is}(\boldsymbol{D})|z)$ 为在隐变量 z 确定的条件下，判别器 D_{is} 对 $\boldsymbol{D}$ 的输出概率分布。

采用三重准则[22]训练 VAE-GAN，获得模型总损失函数 L 为

$$L=L_{VAE}+L_{GAN}=L_{KL}+L_{llike}^{Dis}+L_{GAN} \tag{4}$$

式中 L_{VAE} 为 VAE 部分的损失函数。

3 基于 FD 约束迁移的虚拟样本生成

针对待测油井生成的虚拟样本多样性及信息量不足,且部分样本与待测油井生产规则和特征不匹配,无法满足待测油井动液面预测模型的建模需求,本文结合油田生产现状,提出一种基于 FD 约束迁移的油井动液面预测模型建模虚拟样本生成方法,整体框架如图 5 所示。其中 $\boldsymbol{D}^{\mathrm{t}}$ 为待测油井生产数据,$\boldsymbol{D}^{\mathrm{s}}$ 为相似工况油井生产数据,σ 为标准差,μ 为均值,$N(0,1)$ 为标准正态分布,F_{dt} 为模糊决策树,$L^{\mathrm{t}}_{\mathrm{VAE}}$、$L^{\mathrm{s}}_{\mathrm{VAE}}$ 为 VAE 损失,$L^{\mathrm{t}}_{\mathrm{KL}}$、$L^{\mathrm{s}}_{\mathrm{KL}}$ 为 KL 损失,$L^{\mathrm{t}}_{\mathrm{MSE}}$、$L^{\mathrm{s}}_{\mathrm{MSE}}$ 为重构损失,L_{Dis}、L_{MMD}、L_{F} 和 L_{G} 分别为判别、迁移、FD 和对抗损失。

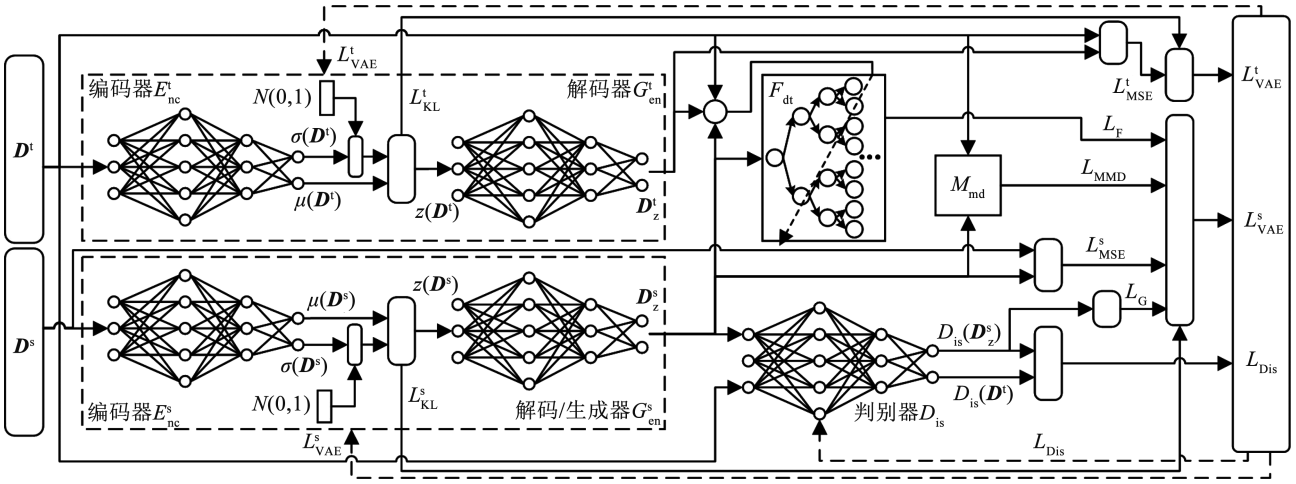


图 5 模型整体框架

Fig. 5 Overall framework of model

首先,通过编码器 $E^{\mathrm{s}}_{\mathrm{nc}}$、解码/生成器 $G^{\mathrm{s}}_{\mathrm{en}}$ 组成源域 VAE 并结合判别器 D_{is}、M_{md} 构建基于迁移学习的 VAE-GAN 模型。通过生产信息丰富的相似工况源域油井数据经源域 VAE 生成虚拟样本,利用判别器 D_{is} 约束至待测目标域油井,同时采用 M_{md} 缩小两域之间的样本分布差异。其次,采用模糊决策树(fuzzy decision tree,F_{dt})和基于编码器 $E^{\mathrm{t}}_{\mathrm{nc}}$、解码器 $G^{\mathrm{t}}_{\mathrm{en}}$ 的目标域 VAE 构建 FD 判别约束模块,进一步增强对源域 VAE 的约束能力。同时,目标域 VAE 的存在为 FD 判别约束模块提供更丰富的待测目标域油井生产信息,增强模块判别能力。最后,在基于迁移学习的 VAE-GAN 模型的基础上添加 FD 判别约束模块,构建基于 FD 判别约束的 VAE-GAN 模型。通过对抗训练和反向传播优化整个模型,生成符合待测目标域油井生产规则和特征的虚拟样本,进而用于构建待测油井动液面预测模型。

3.1 基于迁移学习的虚拟样本生成

为解决 VAE-GAN 使用待测油井生成虚拟样本的不足,利用相似工况油井中丰富的生产信息,提出一种基于迁移学习的 VAE-GAN 虚拟样本生成模型,结构如图 6 所示。

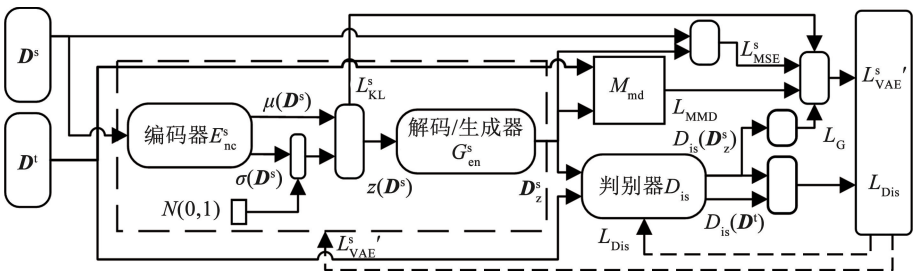


图 6 基于迁移学习的 VAE-GAN 模型结构

Fig. 6 Structure of VAE-GAN model based on transfer learning

设待测油井为目标域油井,其生产数据为 $\boldsymbol{D}^{\mathrm{t}}=[\boldsymbol{X}^{\mathrm{t}},\boldsymbol{y}^{\mathrm{t}}]$,样本数为 n^{t};基于 $\boldsymbol{D}^{\mathrm{t}}$ 生成的虚拟样本 $\boldsymbol{D}^{\mathrm{t}}_{\mathrm{z}}=[\boldsymbol{X}^{\mathrm{t}}_{\mathrm{z}},\boldsymbol{y}^{\mathrm{t}}_{\mathrm{z}}]$,样本数为 n^{t};选取信息丰富的相似工况油井为源域油井[23],其生产数据为 $\boldsymbol{D}^{\mathrm{s}}=[\boldsymbol{X}^{\mathrm{s}},\boldsymbol{y}^{\mathrm{s}}]$,样本数为 n^{s};基于 $\boldsymbol{D}^{\mathrm{s}}$ 生成的虚拟样本 $\boldsymbol{D}^{\mathrm{s}}_{\mathrm{z}}=[\boldsymbol{X}^{\mathrm{s}}_{\mathrm{z}},\boldsymbol{y}^{\mathrm{s}}_{\mathrm{z}}]$,样本数为 n^{s}。其中 $\boldsymbol{X}^{\mathrm{t}}$、$\boldsymbol{X}^{\mathrm{t}}_{\mathrm{z}}\in\mathbf{R}^{n^{\mathrm{t}}\times m}$ 和 $\boldsymbol{X}^{\mathrm{s}}$、$\boldsymbol{X}^{\mathrm{s}}_{\mathrm{z}}\in\mathbf{R}^{n^{\mathrm{s}}\times m}$ 为预测模型输入数据,m 为其维度;$\boldsymbol{y}^{\mathrm{t}}$、$\boldsymbol{y}^{\mathrm{t}}_{\mathrm{z}}\in\mathbf{R}^{n^{\mathrm{t}}\times 1}$ 和 $\boldsymbol{y}^{\mathrm{s}}$、$\boldsymbol{y}^{\mathrm{s}}_{\mathrm{z}}\in\mathbf{R}^{n^{\mathrm{s}}\times 1}$ 为油井动液面数据,且 $n^{\mathrm{t}}\gg m$,$n^{\mathrm{s}}\gg m$。

源域 VAE 的输入样本为 $\boldsymbol{D}^{\mathrm{s}}$,编码器 $E^{\mathrm{s}}_{\mathrm{nc}}$ 通过均

值 $\mu(\boldsymbol{D}^{\mathrm{s}})$ 和标准差 $\sigma(\boldsymbol{D}^{\mathrm{s}})$ 构造后验分布 $q(z|\boldsymbol{D}^{\mathrm{s}})$，解码/生成器 $G_{\mathrm{en}}^{\mathrm{s}}$ 从 $q(z|\boldsymbol{D}^{\mathrm{s}})$ 中采样潜在变量 $z(\boldsymbol{D}^{\mathrm{s}})$ 生成虚拟样本 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$。为使 $q(z|\boldsymbol{D}^{\mathrm{s}})$ 接近潜在分布 $p(z)$，通过最小化 KL 散度减少两者间的差距。其中存在先验假设 $p(z)=N(0,1)$，损失函数为

$$L_{\mathrm{KL}}^{\mathrm{s}}=\frac{1}{2}\sum_{i=1}^{r}(-\log\sigma^2+\mu^2+\sigma^2-1)=D_{\mathrm{KL}}(q(z|\boldsymbol{D}^{\mathrm{s}})\|p(z)) \tag{5}$$

式中 r 为隐变量 z 的维度。

为使生成模型能够拟合目标域油井数据 $\boldsymbol{D}^{\mathrm{t}}$ 分布，选用 M_{md} 衡量 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 与 $\boldsymbol{D}^{\mathrm{t}}$ 间的差异。通过最小化 M_{md} 指导源域 VAE 训练，生成符合 $\boldsymbol{D}^{\mathrm{t}}$ 数据分布的虚拟样本。M_{md} 的表达式为

$$M_{\mathrm{md}}(\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}},\boldsymbol{D}^{\mathrm{t}})=\left\|\frac{1}{n^{\mathrm{s}}}\sum_{i=1}^{n^{\mathrm{s}}}\phi(\boldsymbol{d}_{\mathrm{z},i}^{\mathrm{s}})-\frac{1}{n^{\mathrm{t}}}\sum_{j=1}^{n^{\mathrm{t}}}\phi(\boldsymbol{d}_j^{\mathrm{t}})\right\|_H^2 \tag{6}$$

式中：$\boldsymbol{d}_{\mathrm{z},i}^{\mathrm{s}}\in\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$，即 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}=[\boldsymbol{d}_{\mathrm{z},1}^{\mathrm{s}},\boldsymbol{d}_{\mathrm{z},2}^{\mathrm{s}},\cdots,\boldsymbol{d}_{\mathrm{z},n^{\mathrm{s}}}^{\mathrm{s}}]^{\mathrm{T}}$；$\boldsymbol{d}_j^{\mathrm{t}}\in\boldsymbol{D}^{\mathrm{t}}$，即 $\boldsymbol{D}^{\mathrm{t}}=[\boldsymbol{d}_1^{\mathrm{t}},\boldsymbol{d}_2^{\mathrm{t}},\cdots,\boldsymbol{d}_{n^{\mathrm{t}}}^{\mathrm{t}}]^{\mathrm{T}}$；$\phi$ 为映射函数，选用高斯核函数 $k(x,y)=\exp(-\|x-y\|^2/2\sigma^2)$；$H$ 为特征空间。

M_{md} 的迁移损失函数为

$$L_{\mathrm{MMD}}=M_{\mathrm{md}}(\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}},\boldsymbol{D}^{\mathrm{t}}) \tag{7}$$

采用判别器 D_{is} 与基于 $\boldsymbol{D}^{\mathrm{s}}$ 的源域 VAE 进行对抗训练，即利用 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 与 $\boldsymbol{D}^{\mathrm{t}}$ 训练判别器 D_{is}，使其有效判别出两者差异，进而约束源域 VAE 生成。通过判别结果 $D_{\mathrm{is}}(\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}})$ 和 $D_{\mathrm{is}}(\boldsymbol{D}^{\mathrm{t}})$ 计算其损失函数。

$$L_{\mathrm{Dis}}=E_{\boldsymbol{D}^{\mathrm{t}}\sim P_{\mathrm{t}}}[\log D_{\mathrm{is}}(\boldsymbol{D}^{\mathrm{t}})]+E_{\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}\sim P_{\mathrm{G}}}[\log(1-D_{\mathrm{is}}(\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}))] \tag{8}$$

源域 VAE 与判别器 D_{is} 的对抗损失函数为

$$L_{\mathrm{G}}=E_{\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}\sim P_{\mathrm{G}}}[\log(1-D_{\mathrm{is}}(\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}))] \tag{9}$$

同时，源域 VAE 需保证经重构得到的虚拟样本尽可能保留 $\boldsymbol{D}^{\mathrm{s}}$ 中丰富的生产信息，选用均方误差（mean squared error，MSE）计算其重构损失函数 $L_{\mathrm{MSE}}^{\mathrm{s}}$，表达式为

$$L_{\mathrm{MSE}}^{\mathrm{s}}=\frac{1}{n^{\mathrm{s}}}\sum_{i=1}^{n^{\mathrm{s}}}\|\boldsymbol{d}_i^{\mathrm{s}}-\boldsymbol{d}_{\mathrm{z},i}^{\mathrm{s}}\|^2 \tag{10}$$

式中 $\boldsymbol{d}_i^{\mathrm{s}}\in\boldsymbol{D}^{\mathrm{s}}$，即 $\boldsymbol{D}^{\mathrm{s}}=[\boldsymbol{d}_1^{\mathrm{s}},\boldsymbol{d}_2^{\mathrm{s}},\cdots,\boldsymbol{d}_{n^{\mathrm{s}}}^{\mathrm{s}}]^{\mathrm{T}}$。

在 L_{G} 的基础上，结合 L_{MMD}、$L_{\mathrm{MSE}}^{\mathrm{s}}$ 和 $L_{\mathrm{KL}}^{\mathrm{s}}$ 优化源域 VAE，使 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 在保留 $\boldsymbol{D}^{\mathrm{s}}$ 丰富的特征信息的同时能够符合 $\boldsymbol{D}^{\mathrm{t}}$ 的数据分布。基于 $\boldsymbol{D}^{\mathrm{s}}$ 的源域 VAE 总损失函数为

$$L_{\mathrm{VAE}}^{\mathrm{s}}{}'=w_{\mathrm{g}}^{\mathrm{s}}L_{\mathrm{MSE}}^{\mathrm{s}}+w_{\mathrm{m}}L_{\mathrm{MMD}}+w_{\mathrm{d}}L_{\mathrm{G}}+w_{\mathrm{k}}^{\mathrm{s}}L_{\mathrm{KL}}^{\mathrm{s}} \tag{11}$$

式中 $w_{\mathrm{g}}^{\mathrm{s}}$、$w_{\mathrm{m}}$、$w_{\mathrm{d}}$、$w_{\mathrm{k}}^{\mathrm{s}}$ 为权重系数。

3.2　基于模糊决策的判别约束模块

虽然基于迁移学习的 VAE-GAN 模型丰富了目标域虚拟样本信息，但所生成 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 的部分样本参数不符合目标域油井生产规则。F_{dt} 通过结合模糊理论，学习数据内在规则以提供分类决策[24-25]，使其在面对复杂多变的油井生产环境时可以有效处理生产参数间的不确定性和模糊性。因此，本文设计基于 FD 的判别约束模块，利用 F_{dt} 学习目标域数据和虚拟样本的内在关联，生成基于油井生产特征的模糊规则，实现对虚拟样本的二次判别，进而约束基于迁移学习的 VAE-GAN 模型，生成更符合目标域油井实际生产特征的虚拟样本。FD 判别约束模块训练过程如图 7 所示。其中 FCM 为模糊 C 均值聚类算法，A_j 为 $\boldsymbol{D}^{\mathrm{t}}$、$\boldsymbol{D}_{\mathrm{z}}^{\mathrm{t}}$、$\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 的特征标签，$\boldsymbol{S}_j$ 为 $\boldsymbol{D}^{\mathrm{t,s}}$ 对应的特征列，$\boldsymbol{U}$ 为模糊隶属度矩阵。

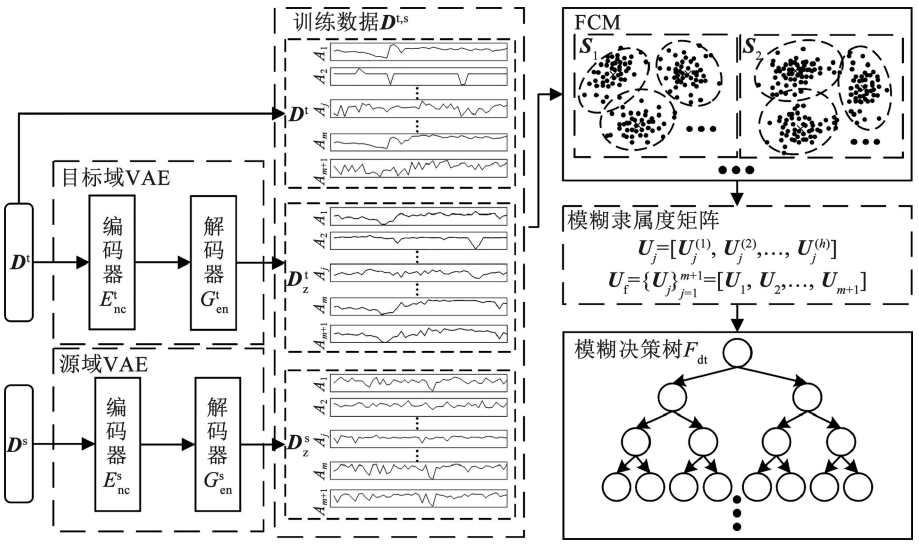


图 7　FD 判别约束模块训练过程

Fig. 7　Training process of FD discriminative constraint module

为解决由 $\boldsymbol{D}^{\mathrm{t}}$、$\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 训练的 FD 判别约束模块因 $\boldsymbol{D}^{\mathrm{t}}$ 样本量不足而导致模块训练不充分，对 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 的判别效果不佳的问题，将 $\boldsymbol{D}^{\mathrm{t}}$ 作为输入，利用 VAE 模型生成含有目标域特征的虚拟样本 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{t}}$，增加训练所需样本数量并增强目标域样本生产特征。基于 $\boldsymbol{D}^{\mathrm{t}}$ 的目标域 VAE 模型与基于 $\boldsymbol{D}^{\mathrm{s}}$ 的源域 VAE 模型结构相同，其总损失函数 $L_{\mathrm{VAE}}^{\mathrm{t}}$ 由 KL 散度 $L_{\mathrm{KL}}^{\mathrm{t}}$ 和重构损失函数 $L_{\mathrm{MSE}}^{\mathrm{t}}$ 构成，表达式为

$$L_{\mathrm{VAE}}^{\mathrm{t}} = w_{\mathrm{g}}^{\mathrm{t}} L_{\mathrm{MSE}}^{\mathrm{t}} + w_{\mathrm{k}}^{\mathrm{t}} L_{\mathrm{KL}}^{\mathrm{t}} \tag{12}$$

式中：$w_{\mathrm{g}}^{\mathrm{t}}$、$w_{\mathrm{k}}^{\mathrm{t}}$ 为权重系数；$L_{\mathrm{KL}}^{\mathrm{t}}$、$L_{\mathrm{MSE}}^{\mathrm{t}}$ 与式(5)、式(10)计算过程相同，但计算数据为 $\boldsymbol{D}^{\mathrm{t}}$。

将 $\boldsymbol{D}^{\mathrm{t}}$、$\boldsymbol{D}_{\mathrm{z}}^{\mathrm{t}}$ 和 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 设为 FD 约束判别模块训练集 $\boldsymbol{D}^{\mathrm{t,s}}$，并划分特征列 $\boldsymbol{S}_j \in \mathbf{R}^{(n^{\mathrm{s}}+2n^{\mathrm{t}})\times 1}$。

$$\boldsymbol{D}^{\mathrm{t,s}} = [\boldsymbol{D}^{\mathrm{t}}, \boldsymbol{D}_{\mathrm{z}}^{\mathrm{t}}, \boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}]^{\mathrm{T}} = [\boldsymbol{S}_1, \boldsymbol{S}_2, \cdots, \boldsymbol{S}_{m+1}] \tag{13}$$

采用 FCM 聚类算法[26]对 $\boldsymbol{D}^{\mathrm{t,s}}$ 进行特征模糊化处理，以获得模糊隶属度。计算每个特征列 $\boldsymbol{S}_j$ 中每个数据点 $d_{ij}^{\mathrm{t,s}}$ 对聚类中心 $c_j^{(v)}$ $(v=1,2,\cdots,h)$ 的模糊隶属度 $u_{ij}^{(v)}$，即聚类范围，其表达式为

$$u_{ij}^{(v)} = \frac{1}{\sum_{b=1}^{h}\left(\dfrac{\| d_{ij}^{\mathrm{t,s}} - c_j^{(v)} \|}{\| d_{ij}^{\mathrm{t,s}} - c_j^{(b)} \|}\right)^{\frac{2}{l-1}}} \tag{14}$$

式中：$\| d_{ij}^{\mathrm{t,s}} - c_j^{(v)} \|$、$\| d_{ij}^{\mathrm{t,s}} - c_j^{(b)} \|$ 分别为数据点 $d_{ij}^{\mathrm{t,s}}$ 至聚类中心 $c_j^{(v)}$、$c_j^{(b)}$ 的欧氏距离，l 为模糊因子。

聚类中心 $c_j^{(v)}$ 通过式(15)进行更新，直至 $c_j^{(v)}$ 的变化小于容忍度或迭代次数达到最大。

$$c_j^{(v)} = \frac{\sum_{i=1}^{n^{\mathrm{s}}+2n^{\mathrm{t}}}[(u_{ij}^{(v)})^l d_{ij}^{\mathrm{t,s}}]}{\sum_{i=1}^{n^{\mathrm{s}}+2n^{\mathrm{t}}}(u_{ij}^{(v)})^l} \tag{15}$$

获得每个特征列 $\boldsymbol{S}_j$ 的模糊隶属度矩阵 $\boldsymbol{U}_j = \{\boldsymbol{U}_j^{(v)}\}_{v=1}^{h}$，将所有矩阵合并得到最终的模糊隶属度矩阵 $\boldsymbol{U}_{\mathrm{f}} = \{\boldsymbol{U}_j\}_{j=1}^{m+1}$。

在树的构造阶段，通过对分类回归树(classification and regression tree，CART)[27]进行模糊扩展来构建 F_{dt}，根据得到的模糊隶属度进行度量模糊化。模糊度量设为模糊基尼系数 $G_{\mathrm{ini}}(\boldsymbol{D}^{\mathrm{t,s}})$ 和模糊基尼增益 $G_{\mathrm{ain}}(\boldsymbol{D}^{\mathrm{t,s}}, \boldsymbol{S}_j)$，表达式分别为：

$$G_{\mathrm{ini}}(\boldsymbol{D}^{\mathrm{t,s}}) = \sum_{k=1}^{3}[p_k(1-p_k)] \tag{16}$$

$$G_{\mathrm{ain}}(\boldsymbol{D}^{\mathrm{t,s}}, \boldsymbol{S}_j) = G_{\mathrm{ini}}(\boldsymbol{D}^{\mathrm{t,s}}) - E_{\mathrm{gini}}(\boldsymbol{D}^{\mathrm{t,s}}, \boldsymbol{S}_j) \tag{17}$$

式中：$p_k = |\boldsymbol{U}_{\mathrm{f}}^{C_k}|/|\boldsymbol{U}_{\mathrm{f}}|$ 为 $\boldsymbol{D}^{\mathrm{t,s}}$ 被标记为 k 类的模糊概率，C_k $(k=1,2,3)$ 为标签，$|\boldsymbol{U}_{\mathrm{f}}^{C_k}|$、$|\boldsymbol{U}_{\mathrm{f}}|$ 分别为 $\boldsymbol{D}^{C_k}$、$\boldsymbol{D}^{\mathrm{t,s}}$ 中模糊隶属度的和，$\boldsymbol{D}^{C_k}$ 是 $\boldsymbol{D}^{\mathrm{t,s}}$ 的模糊子集；$E_{\mathrm{gini}}(\boldsymbol{D}^{\mathrm{t,s}}, \boldsymbol{S}_j)$ 为 $\boldsymbol{D}^{\mathrm{t,s}}$ 按 A_j 分类后的模糊基尼系数，其表达式为

$$E_{\mathrm{gini}}(\boldsymbol{D}^{\mathrm{t,s}}, \boldsymbol{S}_j) = \sum_{v=1}^{h}\left[\frac{|\boldsymbol{U}_j^{(v)}|}{|\boldsymbol{U}_j|} G_{\mathrm{ini}}(\boldsymbol{U}_j^{(v)})\right] \tag{18}$$

FD 判别约束模块通过计算所得模糊基尼增益 $G_{\mathrm{ain}}(\boldsymbol{D}^{\mathrm{t,s}}, \boldsymbol{S}_j)$ 完成对模糊规则的构建。统计每个分类结果并利用投票制决定最终分类所属标签，完成对输入数据 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 的判断。

获得概率集合 $P^{\mathrm{t}} = P(C_1|\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}})$、$P_{\mathrm{z}}^{\mathrm{t}} = P(C_2|\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}})$ 和 $P_{\mathrm{z}}^{\mathrm{s}} = P(C_3|\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}})$，并取概率 P^{t} 与 $P_{\mathrm{z}}^{\mathrm{t}}$ 之和作为模糊决策树 F_{dt} 判别输出，其表达式为

$$F_{\mathrm{dt}}(\boldsymbol{D}_{\mathrm{z}}^{\mathrm{t}}) = 1 - P_{\mathrm{z}}^{\mathrm{s}} = \{p(C_1|\boldsymbol{d}_{\mathrm{z},i}^{\mathrm{s}}) + p(C_2|\boldsymbol{d}_{\mathrm{z},i}^{\mathrm{s}})\}_{i=1}^{n^{\mathrm{s}}} \tag{19}$$

FD 判别约束模块的目的是约束源域 VAE 生成的 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 向目标域逼近，选取 BCE 函数计算损失函数 L_{F}，并最小化 L_{F} 以优化源域 VAE 的参数。损失函数 L_{F} 表达式为

$$L_{\mathrm{F}} = E_{\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}} \sim P_{\mathrm{s}}}[\log(1 - F_{\mathrm{dt}}(\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}))] \tag{20}$$

3.3 总损失函数

基于 FD 判别约束的 VAE-GAN 模型的训练过程即为极大极小零和博弈过程。判别器 D_{is} 通过不断优化以最大化 L_{Dis}，同时 FD 判别约束模块通过训练更新以提升对 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 的识别能力，两者的训练目的均是降低判别输出 $D_{\mathrm{is}}(\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}})$ 和 $F_{\mathrm{dt}}(\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}})$。而源域 VAE 的训练目的与之相反，通过不断优化以最小化 L_{G}、L_{MMD}、$L_{\mathrm{KL}}^{\mathrm{s}}$、$L_{\mathrm{MSE}}^{\mathrm{s}}$ 和 L_{F}，使 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 与目标域油井真实生产数据更相似，提高判别输出 $D_{\mathrm{is}}(\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}})$ 和 $F_{\mathrm{dt}}(\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}})$，从而与前两者形成对抗。

为确保基于 $\boldsymbol{D}^{\mathrm{s}}$ 的源域 VAE 生成 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 更符合目标域油井生产特征，在 $L_{\mathrm{VAE}}^{\mathrm{s}}{}'$ 的基础上添加 L_{F}，对其进一步约束，保证 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 蕴含源域油井 $\boldsymbol{D}^{\mathrm{s}}$ 数据丰富信息的同时，更符合目标域油井生产特征。结合 FD 判别约束模块的源域 VAE 总损失 $L_{\mathrm{VAE}}^{\mathrm{s}}$ 的公式为

$$L_{\mathrm{VAE}}^{\mathrm{s}} = L_{\mathrm{VAE}}^{\mathrm{s}}{}' + w_{\mathrm{f}} L_{\mathrm{F}} \tag{21}$$

式中 w_{f} 为权重系数。

判别器 D_{is} 的目标与基于迁移学习的 VAE-GAN 保持一致，即清晰地分辨 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 与 $\boldsymbol{D}^{\mathrm{t}}$ 间的区别，L_{Dis} 仍采用式(8)计算获得。

3.4 模型执行步骤

基于 FD 约束迁移的 VAE-GAN 模型执行具体步骤如下：

步骤 1 输入源域数据 $\boldsymbol{D}^{\mathrm{s}}$、目标域数据 $\boldsymbol{D}^{\mathrm{t}}$；同时初始化编码器 $E_{\mathrm{nc}}^{\mathrm{s}}$ 和 $E_{\mathrm{nc}}^{\mathrm{t}}$、解码/生成器 $G_{\mathrm{en}}^{\mathrm{s}}$、解码器 $G_{\mathrm{en}}^{\mathrm{t}}$ 和判别器 D_{is} 的超参数，源域 VAE-GAN 最大迭代次数，目标域 VAE 最大迭代次数，FD 判别约束模块迭代条件 Epoch_FD，模糊标签 C^{t}、$C_{\mathrm{z}}^{\mathrm{t}}$、$C_{\mathrm{z}}^{\mathrm{s}}$，损失函数权重 w_{m}、w_{d}、$w_{\mathrm{g}}^{\mathrm{s}}$、$w_{\mathrm{k}}^{\mathrm{s}}$、$w_{\mathrm{g}}^{\mathrm{t}}$、$w_{\mathrm{k}}^{\mathrm{t}}$。

步骤 2　判断当前迭代次数是否达到目标域 VAE 最大迭代次数，如未达到，执行步骤 3，否则目标域 VAE 训练完成，执行步骤 8。

步骤 3　$\boldsymbol{D}^{\mathrm{t}}$ 经 $E_{\mathrm{nc}}^{\mathrm{t}}$ 转换为隐变量 $z(\boldsymbol{D}^{\mathrm{t}})$，更新 $E_{\mathrm{nc}}^{\mathrm{t}}$ 的 $L_{\mathrm{KL}}^{\mathrm{t}}$。

步骤 4　$z(\boldsymbol{D}^{\mathrm{t}})$ 通过 $G_{\mathrm{en}}^{\mathrm{t}}$ 获得目标域虚拟样本 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{t}}$，更新 $L_{\mathrm{MSE}}^{\mathrm{t}}$。

步骤 5　通过 $L_{\mathrm{KL}}^{\mathrm{t}}$ 和 $L_{\mathrm{MSE}}^{\mathrm{t}}$ 计算 $L_{\mathrm{VAE}}^{\mathrm{t}}$。

步骤 6　利用 Adam 算法，基于损失函数 $L_{\mathrm{VAE}}^{\mathrm{t}}$ 优化编码器 $E_{\mathrm{nc}}^{\mathrm{t}}$ 和解码器 $G_{\mathrm{en}}^{\mathrm{t}}$。

步骤 7　返回步骤 2。

步骤 8　判断当前迭代次数是否达到源域 VAE-GAN 最大迭代次数，如未达到，执行步骤 9，否则执行步骤 20。

步骤 9　$\boldsymbol{D}^{\mathrm{s}}$ 经 $E_{\mathrm{nc}}^{\mathrm{s}}$ 转换为隐变量 $z(\boldsymbol{D}^{\mathrm{s}})$，更新 $E_{\mathrm{nc}}^{\mathrm{s}}$ 的 $L_{\mathrm{KL}}^{\mathrm{s}}$。

步骤 10　$z(\boldsymbol{D}^{\mathrm{s}})$ 通过 $G_{\mathrm{en}}^{\mathrm{s}}$ 获得源域虚拟样本 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$，更新 $L_{\mathrm{MSE}}^{\mathrm{s}}$。

步骤 11　判别器 D_{is} 判别 $\boldsymbol{D}^{\mathrm{t}}$ 和 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$，更新 L_{Dis}、L_{G}。

步骤 12　利用 Adam 算法，基于 L_{Dis} 优化判别器 D_{is}。

步骤 13　判断当前迭代次数是否为 Epoch_FD 的倍数，如未达到，执行步骤 15，否则执行步骤 14。

步骤 14　使用 $\boldsymbol{D}^{\mathrm{t}}$、$\boldsymbol{D}_{\mathrm{z}}^{\mathrm{t}}$、$\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 训练更新 FD 判别约束模块。

步骤 15　将 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 输入 FD 判别约束模块，计算 L_{F}。

步骤 16　由 $\boldsymbol{D}^{\mathrm{t}}$ 和 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 计算 M_{md} 迁移损失 L_{MMD}。

步骤 17　结合 L_{G}、$L_{\mathrm{MSE}}^{\mathrm{s}}$、$L_{\mathrm{MMD}}$ 和 L_{F} 计算源域 VAE 的 $L_{\mathrm{VAE}}^{\mathrm{s}}$。

步骤 18　使用 Adam 算法，基于 $L_{\mathrm{VAE}}^{\mathrm{s}}$ 优化编码器 $E_{\mathrm{nc}}^{\mathrm{s}}$ 和解码/生成器 $G_{\mathrm{en}}^{\mathrm{s}}$。

步骤 19　返回步骤 8。

步骤 20　输出最终生成虚拟样本 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$，结束。

4　基于虚拟样本的动液面预测

极限梯度提升（extreme gradient boosting，XGBoost）[28] 方法是一种分布式梯度提升树，通过集成多个基学习器提升模型的性能，具有快速的训练过程和良好的预测效果。本文采用基于 FD 约束迁移的 VAE-GAN 模型生成虚拟样本，通过 XGBoost 构建动液面预测模型，将 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 作为训练数据，$\boldsymbol{X}$ 作为测试输入数据。

XGBoost 的目标是最小化 L_{MSE}（MSE 损失函数）和正则化项 $\Omega(f_k)$ 之和，其损失函数为：

$$L_{\mathrm{XGB}} = \sum_{i=1}^{n^{\mathrm{s}}} L_{\mathrm{MSE}}(y_{\mathrm{z},i}^{\mathrm{s}} - \hat{y}_{\mathrm{z},i}^{\mathrm{s}}) + \sum_{k=1}^{K} \Omega(f_k) \tag{22}$$

$$\Omega(f_k) = \gamma K + \frac{1}{2}\left(\lambda \sum_{j=1}^{K} \omega_j^2\right) \tag{23}$$

式中：$y_{\mathrm{z},i}^{\mathrm{s}} \in \boldsymbol{y}_{\mathrm{z}}^{\mathrm{s}}$，$\hat{y}_{\mathrm{z},i}^{\mathrm{s}}$ 为训练数据预测输出，f_k 为第 k 棵树，K 为树的数量，ω 为叶子权重，γ、λ 为权重系数。

XGBoost 是基于加法模型的梯度提升框架，其模型输出为多个决策树的预测值之和，表达式为

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \tag{24}$$

式中：$x_i \in \boldsymbol{X}$；$\hat{y}_i \in \hat{\boldsymbol{y}}$，$\hat{\boldsymbol{y}}$ 为输出动液面。

$\boldsymbol{X}$ 经基于虚拟样本 $\boldsymbol{D}_{\mathrm{z}}^{\mathrm{s}}$ 构建的 XGBoost 模型，获得动液面预测结果 $\hat{\boldsymbol{y}}$。

5　实验结果与分析

5.1　数据集

实验选取辽河油田某作业区某油井 150 组数据作为目标域待测油井数据，抽取前 50 组作为目标域油井训练数据，剩余 100 组作为目标域油井测试数据，油井动液面预测模型的输入参数为油井液量、油压、套压、电流比、泵效，输出参数为动液面[29]。选取相似油井为源域油井，间隔抽取相似工况生产数据 300 组，目标域及源域油井部分生产数据见表 1。

表 1　源域与目标域油井部分生产参数

Tab. 1　Partial production data of wells in source domain and target domain

来源	序号	液量/($\mathrm{t\cdot d^{-1}}$)	油压/MPa	套压/MPa	电流比	泵效/%	动液面/m
源域	1	11.7	0.3	0.5	1.590	47.8	1 676
	2	11.6	0.3	0.5	1.586	47.3	1 789
	3	11.0	0.3	0.5	1.595	44.9	1 758
	4	11.5	0.3	0.5	1.590	46.9	1 704
	5	11.1	0.3	0.5	1.622	45.3	1 603
目标域	1	12.8	0.3	0.5	1.592	48.7	1 709
	2	12.5	0.3	0.5	1.750	47.6	1 609
	3	11.9	0.3	0.5	1.840	45.3	1 703
	4	12.0	0.3	0.5	1.840	45.7	1 744
	5	11.8	0.3	0.5	1.666	44.9	1 737

5.2　合理性仿真及分析

5.2.1　参数设置

为验证本文所提虚拟样本生成模型的合理性，通过仿真实验分析对抗网络中各个模块的作用。将不同模块组合为不同生成模型，组合方式见表 2。其中√表示模型所采用模块。

表 2 各模块组合方式

Tab. 2 Combination mode of each module

模型	源域 VAE-GAN	M_{md}	FD	目标域 VAE
a	√			
b	√	√		
c	√		√	
d	√		√	√
e	√	√	√	
f	√	√	√	√

各生成模型的迭代次数为 1 000,判别器 D_{is} 与解码/生成器 G_{en} 训练次数比为 2∶1;每迭代 50 次更新 FD 判别约束模块;设 $\boldsymbol{D}^{t}$ 的模糊标签为 $C^{t}=C_1=1$,$\boldsymbol{D}_z^{t}$ 的模糊标签为 $C_z^{t}=C_2=0.5$,$\boldsymbol{D}_z^{s}$ 的模糊标签为 $C_z^{s}=C_3=0$;设置树的最大深度为 5,拆分节点所需最小样本数为 2;每次更新时 FCM 算法最大训练代数为 300,模糊因子为 2,容忍度为 1×10^{-4},聚类中心数为 5,并从对应特征列随机抽取 5 个样本作为初始聚类中心;生成模型损失函数的权重因子由交叉验证选取。

实验中目标域 VAE 和源域 VAE 的编码器 E_{nc}、解码/生成器 G_{en} 结构相同,与判别器 D_{is} 均采用全连接神经网络构建,隐藏层激活函数采用 PReLU,解码/生成器 G_{en} 和判别器 D_{is} 的输出层采用 Sigmoid。相关参数设置见表 3。

表 3 模块参数设置

Tab. 3 Module parameter settings

模块类别	网络层数	输入层节点	隐藏层节点	输出层节点
编码器 E_{nc}	4	36	64 128	256
解码/生成器 G_{en}	4	256	128 64	36
判别器 D_{is}	5	36	64 128 64	1

5.2.2 建立模糊规则

选取油井生产数据中液量、油压、套压、电流比、泵效、动液面作为模糊特征属性,记为 A_1、A_2、A_3、A_4、A_5、A_6。利用 3.3 节中的 FCM 算法进行特征模糊化并采用模糊 CART 算法构建 F_{dt},选取最终构建的 F_{dt} 进行分析,其聚类中心值见表 4。

表 4 聚类中心值

Tab. 4 Cluster center values at different stages

聚类中心	液量/(t·d^{-1})	油压/MPa	套压/MPa	电流比	泵效/%	动液面/m
1	0.502	0.564	0.020	0.627	0.579	0.878
2	0.627	0.704	0.100	0.446	0.746	0.667
3	0.561	0.143	0.052	0.278	0.281	0.789
4	0.442	0.636	0.260	0.372	0.837	0.345
5	0.209	0.789	0.166	0.520	0.673	0.549

获得每个特征列的隶属度矩阵,分别为 $\boldsymbol{U}_1\sim\boldsymbol{U}_6$,其中 $\boldsymbol{U}_1$ 如下:

$$\boldsymbol{U}_1=\begin{bmatrix}0.162 & 0 & 0.838 & 0 & 0\\ 0 & 0 & 0 & 0.934 & 0.017\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0.522 & 0 & 0.478 & 0 & 0\\ 0 & 0.205 & 0.773 & 0 & 0\end{bmatrix}_{400\times5} \tag{25}$$

合并隶属度矩阵 $\boldsymbol{U}_1\sim\boldsymbol{U}_6$,构成最终模糊隶属度矩阵 $\boldsymbol{U}_f$,用于计算模糊基尼系数和模糊基尼增益,$\boldsymbol{U}_f$ 见式(26)。

所构建的 F_{dt} 如图 8 所示,其中叶节点上所示标签即为分类结果,$G_1\sim G_{48}$ 为模糊基尼增益,T 为正确,F 为错误。

$$\boldsymbol{U}_f=\begin{bmatrix}\overbrace{0.162}^{\boldsymbol{U}_1} & \cdots & 0 & \overbrace{0}^{\boldsymbol{U}_2} & \cdots & 0 & \overbrace{0}^{\boldsymbol{U}_3} & \cdots & 0.898 & \overbrace{0.053}^{\boldsymbol{U}_4} & \cdots & 0.923 & \overbrace{0.015}^{\boldsymbol{U}_5} & \cdots & 0.981 & \overbrace{0.378}^{\boldsymbol{U}_6} & \cdots & 0\\ 0 & \cdots & 0.017 & 0.434 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0.535 & \cdots & 0.310 & 0 & \cdots & 0.112\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0.522 & \cdots & 0 & 0.694 & \cdots & 0 & 0.255 & \cdots & 0 & 0.129 & \cdots & 0.813 & 0 & \cdots & 0.114 & 0.267 & \cdots & 0\\ 0 & \cdots & 0 & 0 & \cdots & 0.124 & 0.937 & \cdots & 0 & 0 & \cdots & 0.330 & 0 & \cdots & 0 & 0 & \cdots & 0.822\end{bmatrix} \tag{26}$$

从根节点到叶子节点的每条路径都是一条模糊规则,将构建的 F_{dt} 转换成模糊规则,见表 5。

由构建的 F_{dt} 和模糊规则可知,每个样本参数在不同阈值下被充分分析,形成基于多特征比较与交互的决策路径,说明 F_{dt} 不仅捕捉到各个参数在不同条件下对结果的影响,还揭示了多个参数共同作用下形成的生产规则,从而为 FD 判别约束模块的构建提供有力支持。

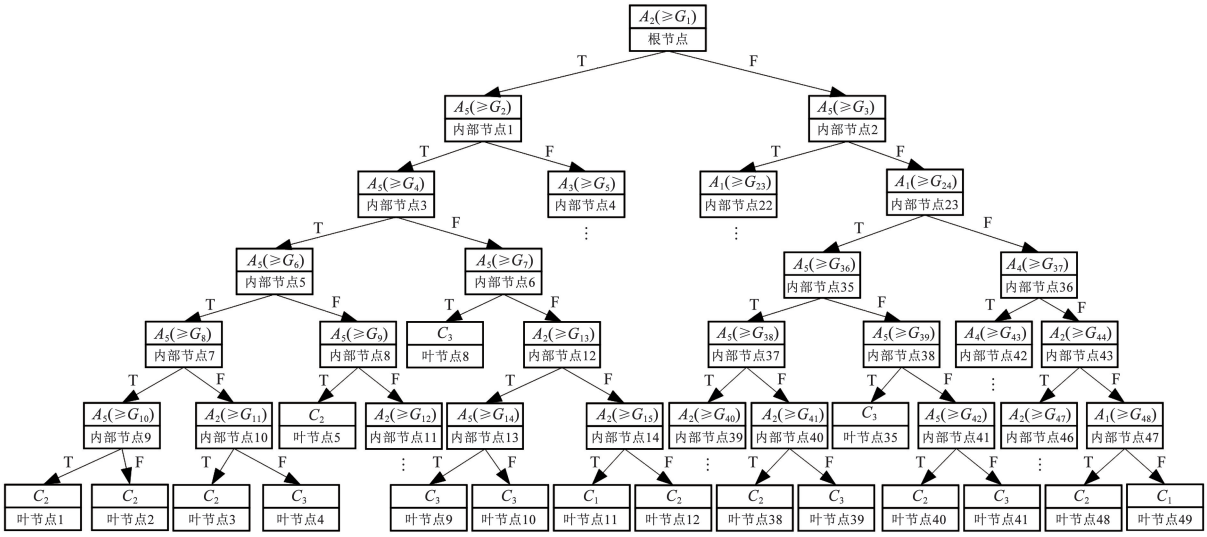


图 8　模糊决策树

Fig. 8　Fuzzy decision tree

表 5　模糊规则

Tab. 5　Fuzzy classification rules

序号	模糊规则	结果
1	$A_2(\geqslant G_1),A_5(\geqslant G_2),A_5(\geqslant G_4),A_5(\geqslant G_6),A_5(\geqslant G_8),A_5(\geqslant G_{10})$	C_z^t
2	$A_2(\geqslant G_1),A_5(\geqslant G_2),A_5(\geqslant G_4),A_5(\geqslant G_6),A_5(\leqslant G_8),A_5(<G_{10})$	C_z^t
3	$A_2(\geqslant G_1),A_5(\geqslant G_2),A_5(\geqslant G_4),A_5(\geqslant G_6),A_5(<G_8),A_2(\geqslant G_{11})$	C_z^t
⋮	⋮	⋮
8	$A_2(\geqslant G_1),A_5(\geqslant G_2),A_5(<G_4),A_5(\geqslant G_7)$	C_z^s
9	$A_2(\geqslant G_1),A_5(\geqslant G_2),A_5(<G_4),A_5(<G_7),A_2(\geqslant G_{13}),A_5(\geqslant G_{14})$	C_z^s
10	$A_2(\geqslant G_1),A_5(\geqslant G_2),A_5(<G_4),A_5(<G_7),A_2(\geqslant G_{13}),A_5(<G_{14})$	C_z^s
⋮	⋮	⋮
38	$A_2(<G_1),A_5(<G_3),A_1(\geqslant G_{24}),A_5(\geqslant G_{36}),A_5(<G_{38}),A_5(\geqslant G_{41})$	C_z^t
39	$A_2(<G_1),A_5(<G_3),A_1(\geqslant G_{24}),A_5(\geqslant G_{36}),A_5(<G_{38}),A_5(<G_{41})$	C_z^s
40	$A_2(<G_1),A_5(<G_3),A_1(\geqslant G_{24}),A_5(<G_{36}),A_5(<G_{39}),A_5(\geqslant G_{42})$	C_z^t
⋮	⋮	⋮
48	$A_2(<G_1),A_5(<G_3),A_1(<G_{24}),A_4(<G_{37}),A_2(<G_{44}),A_1(\geqslant G_{48})$	C_z^t
49	$A_2(<G_1),A_5(<G_3),A_1(<G_{24}),A_4(<G_{37}),A_2(<G_{44}),A_1(<G_{48})$	C^t

5.2.3　结果分析

利用模型 a ~ f 分别生成 300 组虚拟样本，采用主成分分析[30]将各生成模型生成的虚拟样本与 100 组目标域测试数据降维至二维空间进行对比，结果如图 9 所示。

由图 9 可以看出：仅采用少量目标域油井训练数据的模型 a 所生成的虚拟样本虽与目标域油井测试数据的分布趋势一致，但样本过于集中，说明所生成虚拟样本包含的生产信息不足。具有 M_{md} 迁移损失补偿的模型 b 所生成的虚拟样本有很好的分散性，但仍有部分样本偏离目标域。模型 c 生成的虚拟样本与目标域油井测试数据的分布趋势接近，且重叠样本较多，证明基于 $\boldsymbol{D}^t$ 和 $\boldsymbol{D}_z^s$ 的 FD 判别约束模块在源域 VAE-GAN 生成过程中起到一定约束作用，但仍有部分样本分布差异较大。模型 d 生成的虚拟样本分布与模型 c 类似，且样本更为分散，表明添加目标域 VAE 在一定程度上提升了 FD 判别约束模块的判别效果。基于 $\boldsymbol{D}^t$ 和 $\boldsymbol{D}_z^s$ 的 FD 判别约束模块与 M_{md} 损失共同作用时，模型 e 所生成的虚拟样本分布较模型 b 有明显的约束效果，且与模型 c、d 相比更为发散，但仍存在局部样本分布较为集中的现象。相较于其他组合，本文模型生成的虚拟样本在分布和趋势上均与目标域油井测试数据更为匹配，其基本被约束在目标域数据范围内，并表现出良好的聚散性。说明结合 M_{md} 损失和基于 $\boldsymbol{D}^t$、$\boldsymbol{D}_z^s$ 和 $\boldsymbol{D}_z^t$ 的 FD 判别约束模块能有效指导生成模型生成待测目标域油井虚拟样本，验证了所提出虚拟样本生成模型设计的合理性。

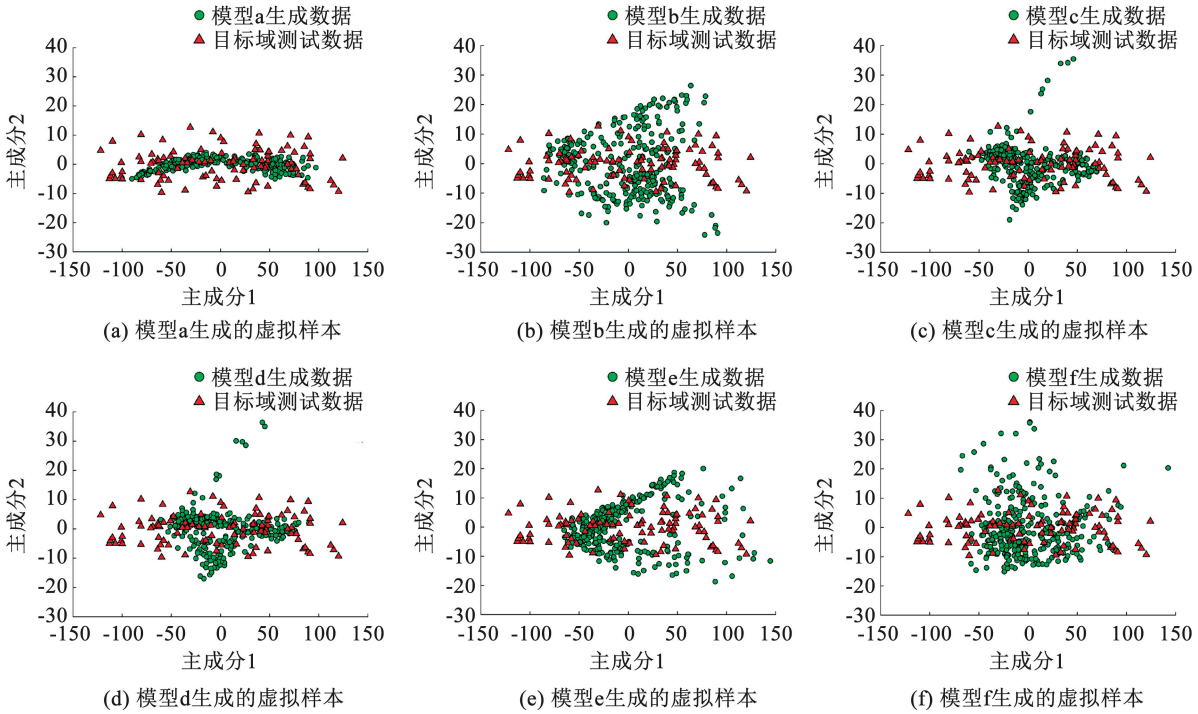


图 9　各模型生成的虚拟样本与目标域测试数据分布对比

Fig. 9　Comparison of virtual samples generated by each model and test data distribution in the target domain

为进一步定量观测上述各生成模型生成虚拟样本的质量，对所构建油井动液面预测模型进行分析。根据油井生产规律，需扩充虚拟样本数量为真实样本数量的 5 ~ 7 倍，因此，各生成模型生成 300 组虚拟样本并分别拼接 50 组目标域训练样本组成训练集 1 ~ 6。XGBoost 超参数经交叉验证确定，采用平均绝对误差（mean absolute error，E_{MA}）、均方根误差（root mean squared error，E_{RMS}）为预测模型评估指标，其表达式分别为：

$$E_{MA} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{27}$$

$$E_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{28}$$

各模型生成虚拟样本训练的预测模型对 100 组目标域测试数据的预测结果如图 10 所示，评估指标结果见表 6。

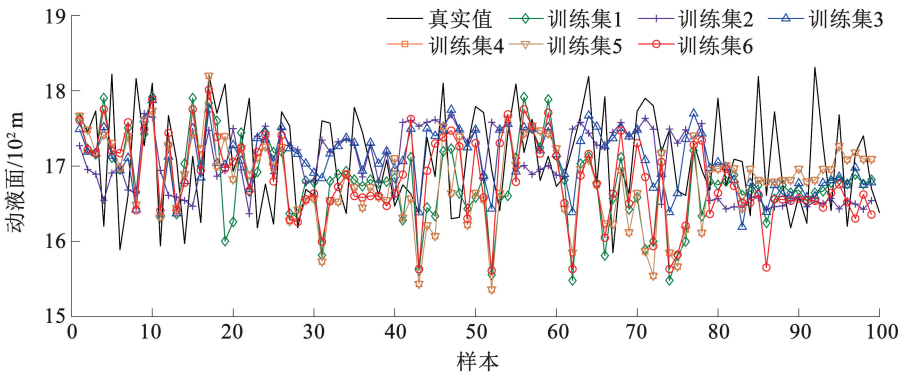


图 10　基于不同训练集的动液面预测模型预测结果

Fig. 10　Prediction results of the dynamic liquid level prediction model based on different training sets

由图 10 和表 6 可知：基于训练集 2 ~ 4 的动液面预测模型性能相较训练集 1 构建的预测模型性能均有所不足，说明仅从数据分布或规则约束角度出发生成的虚拟样本难以充分约束源域数据向目标域进行有效迁移。相较训练集 3 所构建的预测模型，基于训练集 4 的预测模型性能更佳，说明添加目标域 VAE 能有效提升 FD 判别约束模块的约束能力。基于训练集 5、6 的预测模型性能优于训练集 2 所构建的模型，说明 FD 判别约束模块对虚拟样本的二次判别效果优于仅引入 M_{md} 损失的 VAE-GAN 生成模型的一次判别效果。基于训练集 6 的动液面预测模型具有最小输出误差和最佳预测性能，说明本文

生成模型结合了 M_{md} 损失和基于 $\boldsymbol{D}^{t}$、$\boldsymbol{D}_{z}^{s}$ 和 $\boldsymbol{D}_{z}^{t}$ 的 FD 判别约束模块的优势，能获得目标域真实样本中不具备的生产信息，生成符合目标域油井真实生产数据特征的虚拟样本，从而提升动液面预测模型性能。

表 6　评估指标结果

Tab. 6　Results of evaluation metrics

训练集序号	训练集来源	E_{MA}/m	E_{RMS}/m
1	模型 a + 目标域训练集	66.136	84.228
2	模型 b + 目标域训练集	69.097	86.933
3	模型 c + 目标域训练集	69.294	85.668
4	模型 d + 目标域训练集	67.121	83.819
5	模型 e + 目标域训练集	66.068	82.721
6	模型 f + 目标域训练集	**57.766**	**72.308**

5.3　有效性仿真及分析

5.3.1　参数设置

为验证本文所提出的用于动液面预测模型建模的虚拟样本生成模型的有效性，分别将 50 组目标域训练数据、300 组源域生产数据和本文生成模型生成的 300 组虚拟样本单独作为训练集 1 ~ 3；将 VAE、GAN、合成少数过采样技术（synthetic minority over-sampling technique，SMOTE）[31]、VAE-GAN 以及本文生成模型分别生成的 300 组虚拟样本与 50 组目标域训练数据拼接构成训练集 4 ~ 8，进行对比仿真实验。同时为验证本文生成模型所生成虚拟样本对提升其他预测模型性能的有效性，采用训练集 1 ~ 3 和 8 构建随机配置网络（stochastic configuration network，SCN）[32] 预测模型补充对比实验。SMOTE 中邻近点数经交叉验证设为 6，生成样本数为 300；VAE、VAE-GAN 和本文生成模型的模块参数设置见表 3，GAN 参数设置见表 7，隐藏层激活函数均为 PReLU，输出层均为 Sigmoid。生成模型、XGBoost 预测模型超参数设置及评估指标与上一实验保持一致；SCN 预测模型中最大隐藏节点数为 10，容忍度为 0.01，最大候选节点数为 100。

表 7　GAN 参数设置

Tab. 7　GAN parameter settings

模块类别	网络层数	输入层节点	隐藏层节点	输出层节点
生成器 G_{en}	5	36	64 128 64	36
判别器 D_{is}	5	36	64 128 64	1

5.3.2　结果分析

基于不同训练集构建不同动液面预测模型的测试结果如图 11 所示，评估指标结果见表 8。

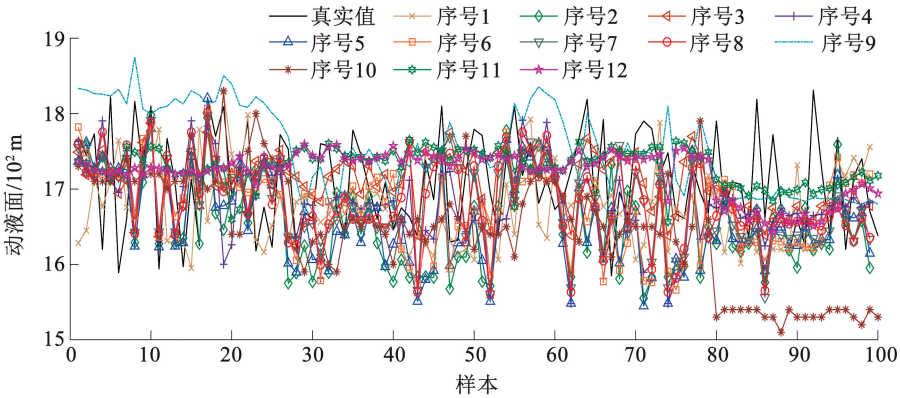


图 11　基于不同训练集的各动液面预测模型预测结果

Fig. 11　Prediction results of each dynamic fluid level prediction model based on different training sets

由图 11 和表 8 可知：基于源域数据训练集 1 构建的 XGBoost 预测模型性能优于采用目标域少量数据训练集 2 构建的预测模型，说明源域数据中蕴含着与目标域相似的生产信息，印证了将迁移学习引入动液面虚拟样本生成过程的可行性。基于训练集 3 ~ 8 构建的 XGBoost 预测模型评估指标均优于采用训练集 1、2 构建的预测模型，说明采用 VSG 模型扩充目标域油井数据可以降低动液面预测模型的输出误差。其中，本文生成模型生成的训练集 3 相较于训练集 2 所构建预测模型的 E_{MA} 和 E_{RMS} 分别下降 27.517%、24.367%，表明所提生成模型生成的虚拟样本能有效提升预测模型的性能。而基于混合样本训练集 8 构建的 XGBoost 生成模型输出误差优于仅采用虚拟样本训练集 3 的预测模型，相较于训练集 2 所构建预测模型的 E_{MA} 和 E_{RMS} 分别下降 29.589%、26.386%，说明混合后的样本可以提供更丰富的生产信息，进一步提升预测模型的性能。同时，与 VAE、GAN、SMOTE 及 VAE-GAN 相比，基于本文生

成模型的训练集 8 所构建 XGBoost 预测模型的预测性能最佳,E_{MA}分别下降 12.331%、15.161%、16.091%、12.656%,E_{RMS} 分别下降 13.531%、16.359%、16.464%、14.152%。说明传统 VSG 模型在样本有限情况下难以充分学习待测油井的生产信息,而本文生成模型通过迁移相似工况油井的生产信息来补充待测油井,使生成的虚拟样本能有效地提升预测模型的性能。此外,基于本文生成模型生成虚拟样本构建的 SCN 预测模型与 XGBoost 预测模型性能相近,表明本文生成模型生成虚拟样本在 SCN 上同样有效。但相较于 XGBoost 预测模型,SCN 预测模型输出预测曲线拟合程度略显不足,说明其在动液面预测任务中性能逊于 XGBoost。

表 8 评估指标结果

Tab. 8 Results of evaluation metrics

序号	训练集	预测模型	训练集来源	E_{MA}/m	E_{RMS}/m
1	1	XGBoost	源域	70.920	87.631
2	2		目标域	82.041	98.226
3	3		本文生成模型	**59.466**	**74.291**
4	4		VAE + 目标域	65.891	83.623
5	5		GAN + 目标域	68.089	86.450
6	6		SMOTE + 目标域	68.844	86.559
7	7		VAE-GAN + 目标域	66.136	84.228
8	8		本文生成模型 + 目标域	**57.766**	**72.308**
9	1	SCN	源域	69.957	89.251
10	2		目标域	92.152	113.588
11	3		本文生成模型	**61.600**	**75.154**
12	8		本文生成模型 + 目标域	**58.165**	**73.764**

6 结 论

本文提出一种基于 FD 约束迁移的动液面预测建模虚拟样本生成模型,并通过合理性和有效性仿真实验进行系统验证,得出结论如下:

1)合理性实验表明,采用 FD 判别约束模块和 M_{md}迁移损失能够有效引导具有源域油井丰富生产信息的虚拟样本向目标域待测油井迁移,使之更符合目标域油井生产特征。

2)有效性对比实验表明,相较于通过少量真实样本和 VAE、GAN、SMOTE、VAE-GAN 生成虚拟样本构建的预测模型,采用本文所提方法构建的预测模型的 E_{MA}分别下降了 29.589%、12.331%、15.161%、16.091%、12.656%,E_{RMS} 分别下降了 26.386%、13.531%、16.359%、16.464%、14.152%,说明本文所提生成模型生成的虚拟样本可以有效提升动液面预测模型性能。

参考文献

[1]王通, 高宪文, 刘文芳. 自适应软测量方法在动液面预测中的研究与应用[J]. 化工学报, 2014, 65(12): 4898
Wang Tong, Gao Xianwen, Liu Wenfang. Adaptive soft sensor method and application in determination of dynamic fluid levels[J]. CIESC Journal, 2014, 65(12): 4898. DOI:10.3969/j.issn.0438 - 1157.2014.12.034

[2]Andsager R L, Knapp R M. Acoustic determination of liquid levels in gas wells[J]. Journal of Petroleum Tecnology, 1967, 19(5): 601. DOI:10.2118/1640 - PA

[3]Leng Chunyang, Jia Mingxing, Zheng Haijin, et al. Dynamic liquid level prediction in oil wells during oil extraction based on WOA-AM-LSTM-ANN model using dynamic and static information[J]. Energy, 2023, 282: 128981. DOI: 10.1016/j.energy.2023.128981

[4]Jia Mingxing, Leng Chunyang. Multi-well dynamic liquid-level prediction method of pumping well based on dynamic and static information feature fusion neural network[J]. Transactions of the Institute of Measurement and Control, 2023, 45(6): 1099. DOI: 10.1177/01423312221126006

[5]侯延彬, 高宪文, 李翔宇. 采油过程多尺度状态特征生成的有杆泵动态液面预测[J]. 化工学报, 2019, 70(增刊2): 311
Hou Yanbin, Gao Xianwen, Li Xiangyu. Prediction for dynamic liquid level of sucker rod pumping using generation of multi-scale state characteristics in oil field production[J]. CIESC Journal, 2019, 70(S2): 311. DOI:10.11949/0438 - 1157.20190352

[6]Tian Zhongda. Kernel principal component analysis-based least squares support vector machine optimized by improved grey wolf optimization algorithm and application in dynamic liquid level forecasting of beam pump[J]. Transactions of the Institute of Measurement and Control, 2020, 42(6): 1135. DOI: 10.1177/0142331219885273

[7]许小伟, 韦道明, 严运兵, 等. 基于改进 ACGAN 的永磁同步电机数据扩张方法[J]. 哈尔滨工业大学学报, 2023, 55(10): 114
Xu Xiaowei, Wei Daoming, Yan Yunbing, et al. Data expansion method of permanent magnet synchronous motor based on improved ACGAN[J]. Journal of Harbin Institute of Technology, 2023, 55(10): 114. DOI:10.11918/202203045

[8]Kingma D P, Welling M. Auto-encoding variational bayes[C]//2nd International Conference on Learning Representations, ICLR 2014. Banff: ICLR, 2014: 184

[9]Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139. DOI:10.1145/342262

[10]Zhao Chen, Keyak J H, Cao Xuewei, et al. Multi-view information fusion using multi-view variational autoencoder to predict proximal femoral fracture load[J]. Frontiers in Endocrinology, 2023, 14: 1261088. DOI:10.3389/fendo.2023.1261088

[11]Oliveira A F, Dasilva J L F, Quiles M G. Molecular property prediction and molecular design using a supervised grammar variational autoencoder[J]. Journal of Chemical Information and Modeling, 2022, 62(4): 817. DOI:10.1021/acs.jcim.1c01573

[12]Yoon J, Jarrett D, Schaar M. Time-series generative adversarial networks[C]//Proceedings of the 33rd Annual Conference on Neural Information Processing Systems, NeurIPS 2019. Vancouver: Neural Information Processing Systems Foundation, 2019: 5508. DOI:10.5555/3454287.3454781

[13] Gao Xin, Deng Fang, Yue Xianghu. Data augmentation in fault diagnosis based on the Wasserstein generative adversarial network with gradient penalty[J]. Neurocompting, 2020, 396: 487. DOI: 10.1016/j.neucom.2018.10.109

[14] Lee S M, Park S Y, Choi B H. Application of domain-adaptive convolutional variational autoencoder for stress-state prediction[J]. Knowledge-Based Systems, 2022, 248: 108827. DOI:10.1016/j.knosys.2022.108827

[15] Zhu Junyan, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// Proceedings of the 16th IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 2223. DOI: 10.1109/ICCV.2017.244

[16] Muhammad Y, Feng Jinchao, Shahzad A, et al. GAN-TL: generative adversarial networks with transfer learning for MRI reconstruction[J]. Applied Sciences-Basel, 2022, 12(17): 8841. DOI:10.3390/pp12178841

[17] 孙玲俊, 毛经宇, 刘坤. FAGAN: 一种基于模型迁移与注意力机制的太阳电池缺陷图像对抗生成方法[J]. 太阳能学报, 2023, 44(9): 78
Sun Lingjun, Mao Jingyu, Liu Kun. FAGAN: an adversarial generation method of solar cells defect image based on model transfer and attention mechanism[J]. Acta Energiae Solaris Sinica, 2023, 44(9): 78. DOI: 10.19912/j.0254-0096.tynxb.2022-0754

[18] 王通, 陈延彬. 基于改进生成对抗网络的动液面建模数据扩充[J]. 电子测量与仪器学报, 2023, 37(2): 99
Wang Tong, Chen Yanbin. Dynamic liquid level modeling data augmentation based on improved generative adversarial networks [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(2): 99. DOI:10.13382/j.jemi.B2205928

[19] Lin Mansheng, Teng Shuai, Chen Gongfa, et al. Transfer learning with attributes for improving the landslide spatial prediction performance in sample-scarce area based on variational autoencoder generative adversarial network [J]. Land, 2023, 12(3): 525. DOI:10.3390/land12030525

[20] Li Zilu, Peng Xiangang, Cui Wenbo, et al. A novel scenario generation method of renewable energy using improved VAEGAN with controllable interpretable features[J]. Applied Energy, 2024, 363: 122905. DOI:10.1016/j.apenergy.2024.122905

[21] Luo Yuxuan, Wang Xizhao, Pourpanah F. Dual VAEGAN: a generative model for generalized zero-shot learning[J]. Applied Soft Computing, 2021, 107: 107352. DOI: 10.1016/j.asoc.2021.107352

[22] Larsen A B L, Sonderby S K, Larochelle H, et al. Autoencoding beyond pixels using a learned similarity metric[C]//Proceedings of the 33rd International Conference on Machine Learning. New York: IMLS, 2016: 1558

[23] 王通, 高宪文, 刘文芳. 基于改进即时学习算法的动液面软测量建模[J]. 东北大学学报(自然科学版), 2015, 36(7): 918
Wang Tong, Gao Xianwen, Liu Wenfang. Soft sensor for determination of dynamic fluid levels based on enhanced just-in-time learning algorithm[J]. Journal of Northeastern University (Natural Science), 2015, 36(7): 918. DOI: 10.3969/j.issn.1005-3026.2015.07.002

[24] 翟俊海, 王熙照, 张素芳. 基于模糊积分的多模糊决策树融合[J]. 计算机研究与发展, 2009, 46(3): 470
Zhai Junhai, Wang Xizhao, Zhang Sufang. Integration of multiple fuzzy decision trees based on fuzzy integral[J]. Journal of Computer Research and Development, 2009, 46(3): 470

[25] 张晓丹, 赵海, 王刚, 等. 不确定信息的模糊决策融合算法[J]. 东北大学学报(自然科学版), 2004, 25(7): 657
Zhang Xiaodan, Zhao Hai, Wang Gang, et al. Fusion algorithm for uncertain information by fuzzy decision tree [J]. Journal of Northeastern University(Natural Science), 2004, 25(7): 657

[26] 魏文军, 张轩铭, 杨立本. 基于模糊聚类和改进 Densenet 网络的小样本轴承故障诊断[J]. 哈尔滨工业大学学报, 2024, 56(3): 154
Wei Wenjun, Zhang Xuanming, Yang Liben. Fault diagnosis of small sample bearings based on fuzzy clustering and improved Densenet network[J]. Journal of Harbin Institute of Technology, 2024, 56(3): 154. DOI:10.11918/202206075

[27] 张俊玉, 胡家豪, 黄嵩. CART 决策树方法在煤电厂节能降耗中的应用[J]. 控制与决策, 2021, 36(5): 1232
Zhang Junyu, Hu Jiahao, Huang Song. Application of CART decision tree model in reducing coal consumption incoal power plant [J]. Control and Decision, 2021, 36(5): 1232. DOI:10.13195/j.kzyjc.2019.1272

[28] 周洲, 焦文玲, 任乐梅, 等. 蚁群算法分配权重的燃气日负荷组合预测模型[J]. 哈尔滨工业大学学报, 2021, 53(6): 177
Zhou Zhou, Jiao Wenling, Ren Lemei, et al. Combined forecasting model of gas daily load based on weight distribution of ant colony algorithm[J]. Journal of Harbin Institute of Technology, 2021, 53(6): 177. DOI:10.11918/201911084

[29] 王通, 段泽文. 基于模糊评估自适应更新的油井动液面软测量建模[J]. 化工学报, 2019, 70(12): 4760
Wang Tong, Duan Zewen. Soft sensor modeling for dynamic liquid level of oil well based on fuzzy inference adaptive updating[J]. CIESC Journal, 2019, 70(12): 4760. DOI:10.11949/0438-1157.20190729

[30] Asante-Okyere S, Shen Chuanbo, Ziggah Y Y, et al. Principal component analysis (PCA) based hybrid models for the accurate estimation of reservoir water saturation [J]. Computers & Geosciences, 2020, 145: 104555. DOI:10.1016/j.cageo.2020.104555

[31] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique [J]. Journal of Artificial Intelligence Research, 2002, 16: 321. DOI: 10.1613/jair.953

[32] 代伟, 张政煊, 杨春雨, 等. 基于 SCN 数据模型的 SISO 非线性自适应控制[J]. 自动化学报, 2024, 50(10): 2002
Dai Wei, Zhang Zhengxuan, Yang Chunyu, et al. Adaptive control of SISO nonlinear system using data-driven SCN model[J]. Acta Automatica Sinica, 2024, 50(10): 2002. DOI:10.16383/j.aas.c210174