面向油田大数据的改进 Transformer 缓存预测方法

景瑞林[1] 杨 硕[1] 李 玲[1] 韩 明[1] 杜心萌[2*] 王烈冲[2]
JING Ruilin YANG Shuo LI Ling HAN Ming DU Xinmeng WANG Liechong

摘 要 随着油气勘探开发业务的不断深入，系统产生的数据量持续增长，传统的缓存管理方式在面对高并发、大规模数据访问需求时，逐渐出现实时性差、命中率低等问题。针对这一挑战，文章提出了一种融合稀疏注意力机制、注意力蒸馏与非自回归解码器的改进 Transformer 模型，更准确地预测未来数据访问行为，从而优化缓存策略。同时在模型训练阶段结合学习率调度策略，增强模型训练稳定性。实验结果表明，该模型在不同缓存容量下均优于传统策略与原始 Transformer 模型，有效提升了缓存命中率，具有良好的应用价值与推广前景，为油田场景下的数据缓存管理与调度提供了新的思路。

关键词 油气勘探；缓存策略；Transformer；稀疏注意力

doi：10.3969/j.issn.1672-9528.2026.06.031

0 引言

随着油气勘探和开发工作的不断深入，相关数据呈现出指数级增长的趋势。无论是地震勘探数据、测井数据，还是生产数据，其种类繁多、规模庞大，使得数据存储系统和访问效率面临前所未有的挑战[1]。特别是在智能油田建设和高性能计算平台日益普及的背景下，如何科学地进行数据缓存管理、提升缓存命中率，已经成为影响整个油田信息系统性能的关键问题。

在提升数据处理效率的过程中，缓存管理策略起着关键的作用[2]。目前常见的缓存管理策略有最少使用策略（LRU）、最不经常使用策略（LFU）以及先进先出策略（FIFO）等。LRU 策略主要根据数据的最近使用情况进行判断，会优先淘汰最久未被访问的数据；LFU 关注数据的访问频率，根据数据被访问的频率来决定淘汰顺序，频率越低的内容越容易被移除；FIFO 简单地按照数据进入缓存的顺序进行淘汰。这些方法虽然结构简单、实现成本低，在传统计算场景中应用广泛且表现稳定，但在应对油田数据复杂访问模式时往往力不从心。

在油田大数据环境下，数据访问呈现出明显的时序特征和上下文关联性。传统的静态策略难以准确捕捉这些动态变化规律，无法适应数据访问模式的长期演变，导致对热点数据的识别不够精准，直接影响缓存系统的整体效率[3]。当处理具有高动态性、非线性特征的数据时，这些方法的预测准确性和适应性都明显不足。因此，需要一种能够具备上下文建模能力、支持多步预测的新方法，来更准确地建模数据访问行为，优化缓存管理效能。

近年来，随着深度学习技术的发展，越来越多的研究尝试将神经网络应用于缓存管理中。循环神经网络（recurrent neural network, RNN）[4] 及其变种，如长短期记忆网络（long short-term memory, LSTM）[5] 和门控循环单元（gated recurrent unit, GRU）[6] 等能够处理时间序列数据，学习历史访问序列中的隐含模式。部分研究基于 RNN 进行建模预测，利用 LSTM 预测未来数据访问行为[7]。

基于深度学习的序列建模方法在各类预测任务中取得显著进展[8]，尤其是 Transformer 模型凭借其自注意力机制在自然语言处理和时间序列分析领域[9]展现出优异性能。Transformer 能够并行建模序列中的长期依赖关系，尤其适用于大规模数据场景。例如，已有研究在缓存预测[10]、磁盘 I/O 调度[11]等任务中引入Transformer框架，取得了较优的预测性能。Transformer 能够并行处理序列信息[12]，并捕捉长距离依赖特征，但其在处理长序列时计算复杂度较高[13]，容易忽略局部特征和短期依赖关系[14]，难以在资源受限或实时性要求高的场景中部署，限制了在工业大数据场景中的应用。此外，其采用逐步解码的方式生成多步预测结果，导致推理时间增加，也使得多步预测效率不高，不利于大规模系统应用。

针对上述挑战，本文提出一种面向油田大数据场景改进的 Transformer 缓存预测方法。该方法引入稀疏注意力机制以降低模型复杂度，通过注意力蒸馏压缩特征表达冗余，并采用非自回归解码器实现高效的多步预测。在构建数据访问行为模型的基础上，进一步设计热数据识别与动态缓存替换策

1. 胜利油田分公司数智化管理服务中心 山东东营 257000
2. 中国石油大学（华东）计算机科学与技术学院 山东青岛 266580

略，全面提升系统缓存命中率和数据访问效率。

本文的主要贡献如下：

（1）针对油田数据访问序列具有规模大、时序性强等特点，设计了一种结合稀疏注意力机制和非自回归结构的改进 Transformer 模型，在保证序列建模能力的基础上，降低了模型计算开销，提高了长序列处理效率；

（2）在模型中加入注意力蒸馏模块，对冗余特征进行压缩与筛选，减少无效信息干扰，使特征表达更加紧凑，从而提升模型在复杂场景下的泛化能力和运行稳定性；

（3）在训练过程中引入动态学习率调度方法，根据模型训练状态对学习率进行调整，以提高模型训练效率，加快收敛速度，并增强训练过程的稳定性与鲁棒性；

（4）在真实数据集上开展系统实验，实现热数据智能识别与动态缓存策略，并与多个模型进行对比。实验结果表明，本文提出的模型可有效提升缓存命中率，验证本方法在缓存预测任务中的有效性与先进性。

1 方法

本文提出一种基于改进 Transformer 模型的油田数据缓存预测方法，旨在精准建模数据访问模式、提升缓存命中率并降低访问延迟。整体模型结构如图 1 所示，主要包括输入编码模块、稀疏注意力编码器 [15]、注意力蒸馏模块、非自回归解码器以及输出层。

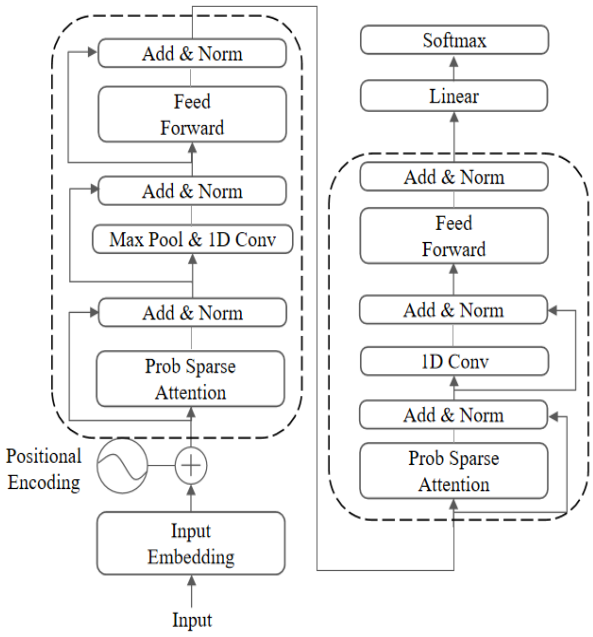

图 1 改进 Transformer 模型结构

1.1 输入嵌入

为了使 Transformer 模型能够有效处理资源访问序列，本研究将离散化后的资源编号序列转化为模型可接受的高维向量形式，并引入动态正弦位置编码以增强模型对访问顺序的感知能力。设输入访问序列为：

$$X=\{x_1,x_2,\cdots,x_L\} \tag{1}$$

式中：L 为时间窗口长度。

本模块采用以下嵌入方式：

（1）值嵌入（Value Embedding）

每个标量输入 x_i 首先通过一个线性变换层嵌入 d 维向量空间中，用公式表示为：

$$\boldsymbol{h}_v(i)=\boldsymbol{W}_e\cdot x_i+\boldsymbol{b}_e \tag{2}$$

式中：$\boldsymbol{W}_e\in\mathbf{R}^{d\times1}$，$\boldsymbol{b}_e\in\mathbf{R}^d$ 为可学习的参数，d 为嵌入维度。

（2）位置嵌入（Position Embedding）

本文采用基于正弦与余弦函数构建的位置编码，具体定义为：

$$\mathrm{PE}(\mathrm{pos},2i)=\sin\left(\frac{\mathrm{pos}}{10\,000^{2i/d}}\right) \tag{3}$$

$$\mathrm{PE}(\mathrm{pos},2i+1)=\cos\left(\frac{\mathrm{pos}}{10\,000^{2i/d}}\right) \tag{4}$$

式中：pos 为序列中的位置，d 为嵌入维度。

（3）嵌入融合

最终每个输入位置的总嵌入表示为值嵌入与位置嵌入的加和：

$$\boldsymbol{h}(i)=\boldsymbol{h}_v(i)+\boldsymbol{h}_p(i) \tag{5}$$

所有嵌入后向量组成嵌入序列：

$$\boldsymbol{H}=\{\boldsymbol{h}(1),\boldsymbol{h}(2),\cdots,\boldsymbol{h}(L)\}\in\mathbf{R}^{L\times d} \tag{6}$$

该序列作为 Transformer 编码器的输入，供模型后续进行序列建模与特征提取。

1.2 稀疏注意力编码器

标准 Transformer 编码器通过多头自注意力机制（Multi-Head Self-Attention）建模输入序列中任意两个位置之间的依赖关系，具有强大的全局建模能力，但其注意力机制的计算复杂度为 $O(n^2)$，其中 n 为序列长度。同时由于油田访问日志往往覆盖数千个对象、序列长度可达上万步，但实际访问集中于少数热点对象。这种全连接注意力计算方式在处理长序列时会带来极大的计算开销和内存占用，同时多数低频对象对预测贡献有限，限制了模型的可扩展性和部署效率。

为了降低复杂度，研究发现自注意力的概率分布具有潜在稀疏性，可以有效减少计算量 [16]。为此，本文引入稀疏注意力机制（Sparse Attention），在计算注意力时限制每个查询向量 $\boldsymbol{Q}_i$ 仅与其点积相似度最高的前 k 个键向量 $\boldsymbol{K}_j$ 计算注意力权重，即最相关的 k 个，而不是所有位置的注意力权重。该策略有效将复杂度降低为 $O(n\cdot k)$，其中 $k<<n$。既捕捉关键上下文又过滤低频噪声，又与油田“热点集中”特点相匹配，采用稀疏注意力可以在保证预测能力的同时大幅降低计算冗

余。

具体而言，首先对每个查询向量 $\boldsymbol{Q}_i$ 与所有键向量 $\boldsymbol{K}_j$ 计算原始注意力得分：

$$S_{ij}=\frac{\boldsymbol{Q}_i\boldsymbol{K}_j^{\mathrm{T}}}{\sqrt{d}} \tag{7}$$

式中：d 为单头注意力的维度，随后对每个 i 仅保留前 k 个最大得分，其余位置置零。记 T_i 为第 i 个查询的 Top-k 索引集合，则有：

$$\tilde{s}_{ij}=\begin{cases}s_{ij}, & j\in T_i\\ 0, & j\notin T_i\end{cases} \tag{8}$$

接着在 Top-k 范围内进行归一化，得到稀疏注意力权重 α_{ij}：

$$\alpha_{ij}=\begin{cases}\dfrac{\exp\left(\tilde{s}_{ij}\right)}{\sum_{l\in T_i}\left(\tilde{s}_{ij}\right)}, & j\in T_i\\ 0, & j\notin T_i\end{cases} \tag{9}$$

最终基于稀疏注意力权重对对应的 Value 向量进行加权求和：

$$\mathrm{Output}_i=\sum\nolimits_{j\in T_i}\alpha_{ij}\boldsymbol{V}_j \tag{10}$$

式中：$\boldsymbol{V}_j$ 为第 j 个键对应的 Value 向量，Output_i 为第 i 个查询的输出特征。保证模型表达能力的同时降低注意力计算的复杂度。

1.3 注意力蒸馏模块

油田访问行为在许多外部因素驱动下存在阶段性变化。例如，在特定采油季节，某些对象的访问需求会短时间剧增，而在某个设备维护期间，这些对象的访问频率又会急剧下降。这类访问模式使得访问序列中包含大量短期关联，易导致模型过拟合。

为了进一步压缩编码器输出中的冗余特征、提升模型泛化性能，本文在 Transformer 编码器后加入了注意力蒸馏模块（Attention Distillation Module）[17-18]，对稀疏注意力矩阵进行压缩与重构，将蒸馏特征和编码器原始输出进行有效融合，如图 2 所示。

该模块以编码器输出的稀疏注意力矩阵 $\boldsymbol{A}\in\mathbf{R}^{n\times d}$ 为输入，其中 n 表示访问序列的长度，d 表示注意力维度。

首先，模块采用一维最大池化操作（MaxPool1D）对注意力矩阵沿时间维度进行压缩，保留局部区域中响应值最大的路径，提取出关键访问模式的显著注意力值。

$$\boldsymbol{A}_{\mathrm{pool}}=\mathrm{MaxPool1D}(\boldsymbol{A}) \tag{11}$$

再通过一维卷积（Conv1D）提取局部结构特征，并进行层归一化处理提取稳健的局部结构特征，增强稳定性：

$$\tilde{\boldsymbol{A}}=\mathrm{Conv1D}\left(\boldsymbol{A}_{\mathrm{pool}}\right) \tag{12}$$

$$\boldsymbol{A}_{\mathrm{distill}}=\mathrm{LayerNorm}\left(\tilde{\boldsymbol{A}}\right) \tag{13}$$

式中：$\tilde{\boldsymbol{A}}\in\mathbf{R}^{n\times d}$ 为压缩后的注意力特征，d 为卷积输出维度。

蒸馏后的特征 $\boldsymbol{A}_{\mathrm{distill}}$ 随后与原始编码器输出 $\boldsymbol{H}_{\mathrm{enc}}$ 进行融合，采用拼接方式对两者特征沿通道维度进行连接，并通过线性映射统一维度，得到融合特征：

$$\boldsymbol{H}_{\mathrm{fusion}}=\mathrm{Linear}\left(\left[\boldsymbol{H}_{\mathrm{enc}};\boldsymbol{A}_{\mathrm{distill}}\right]\right) \tag{14}$$

式中：$\boldsymbol{H}_{\mathrm{fusion}}$ 表示融合后的输出特征。

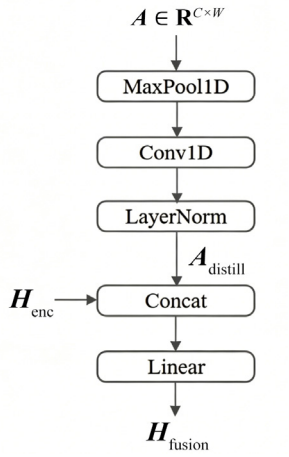

图 2 注意力蒸馏结构

1.4 非自回归解码器

传统 Transformer 解码器通常采用自回归方式进行预测，即每一步输出都依赖前一步结果，导致推理效率较低 [19]，尤其在预测多个未来时间步时存在延迟和误差累积等问题。

针对油田数据服务平台对高效、低延迟预测的需求，本文采用非自回归解码器（Non-Autoregressive Decoder）结构，一次性预测未来多个时间步的数据访问情况。相比逐步解码的方式，这种结构可以有效提高预测效率，减少时间开销，更适用于缓存命中率预测这类序列回归任务。该解码器主要包括以下几个部分：

（1）稀疏注意力层：对输入特征进行筛选，重点关注与当前预测结果相关性较高的信息，在降低计算量的同时保留关键时序特征。

（2）卷积压缩层：利用一维卷积提取局部序列特征，用于捕捉数据访问过程中可能存在的周期变化和短时波动规律，增强模型对时序特征的表达能力。

（3）前馈网络：通过全连接层和非线性激活函数进一步提取深层特征，并输出最终预测结果。

1.5 学习率调度机制

学习率决定了参数更新的步长，是影响模型训练效果的重要参数之一 [20]。学习率设置过大，容易导致训练过程不稳定；设置过小，则会降低模型收敛速度。为了提升模型训练效率与稳定性，本文采用动态学习率调度策略，包括学习率

预热（warm-up）和学习率衰减（Learning Rate Decay）两个阶段。

（1）warm-up 阶段：在训练初期，先使用较小学习率进行训练，再逐渐增加至设定的目标值，可以减小训练初期参数更新步长过大带来的波动，避免梯度不稳定问题，帮助模型更好地适应训练过程。

$$\mathrm{Ir}_{\mathrm{epoch}}=\frac{\mathrm{epoch}}{\mathrm{warmup\_epochs}}\mathrm{Ir}_{\mathrm{base}} \tag{15}$$

$$\mathrm{epoch} < \mathrm{warmup\_epochs} \tag{16}$$

式中：$\mathrm{Ir}_{\mathrm{epoch}}$ 表示当前训练轮次的学习率，epoch 表示当前训练轮次编号，warmup_epochs 表示预设的预热总轮次，$\mathrm{Ir}_{\mathrm{base}}$ 表示基础学习率。

（2）Learning Rate Decay 阶段：warm-up 结束后，使用余弦退火衰减策略（Cosine Annealing）[21-22] 缓慢降低学习率，这种方式可以使模型收敛过程更加平滑，有助于提升后期训练稳定性，并减缓陷入局部最优的风险。

$$\mathrm{Ir}_t=\mathrm{Ir}_{\min}+\frac{1}{2}\left(\mathrm{Ir}_{\mathrm{base}}-\mathrm{Ir}_{\min}\right)\left(1+\cos\left(\frac{t}{T_{\max}}\pi\right)\right) \tag{17}$$

式中：Ir_t 为第 t 步的训练学习率，$\mathrm{Ir}_{\min}$ 为允许的最小学习率，t 为当前全局训练步数，$T_{\max}$ 为余弦退火周期长度，π 为圆周率常数。

该调度机制不仅加快了模型初期的收敛速度，还有效防止陷入局部最优，提升最终预测性能。

2 实验及结果

2.1 数据来源与预处理

实验数据来源于油田真实生产系统，记录了系统内对象访问请求日志，包括访问时间、对象编号、操作类型等信息。原始数据共计 6 789 条访问记录，涉及 1 918 个唯一对象，由于涉及企业内部运营信息，该数据集为非公开数据。

在数据预处理过程中，首先剔除了缺失访问编号、时间戳异常等无效记录。随后根据访问时间对数据按时间顺序排序，保留对象编号作为主特征，用于构造访问序列。将访问日志按时间顺序整理为访问编号序列，并将对象编号转化为整数索引。由于 Transformer 对输入数据范围较敏感，使用归一化将其缩放至 [−1,1] 区间。

此外，对处理后的访问序列进行了频率统计与分布分析，结果显示对象访问频率遵循 Zipf 分布 [23] 规律，即少量高频对象占据大部分访问量，而大多数对象的访问频率较低。如图 3 所示，按照对象编号访问次数从高到低排序，其频率呈现明显的长尾分布。

该分布特性与实际生产环境中“热点数据集中访问”的现象高度一致，为后续的缓存预测模型提供了理论依据。

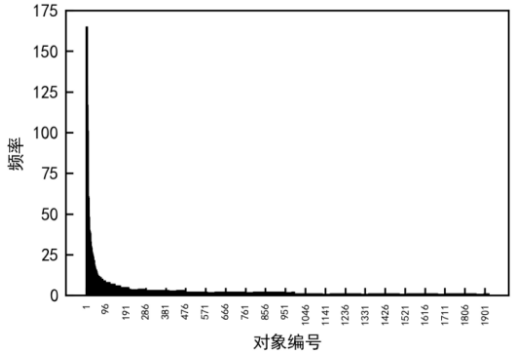

图 3 访问对象编号频率分布图

采用滑动窗口方式构造训练样本，窗口长度设为 20，预测步长为 10，步长为 1。即每 20 个连续访问作为一个输入序列，预测其后 10 个访问编号。通过遍历整个序列生成训练样本，共计构造样本约 6Zipf 分布 700 组，其中 80% 用于训练，20% 用于测试。

2.2 实验评估

为验证本文所提出的 Transformer 缓存预测模型在实际数据场景中的性能表现，本文设计了四组实验。在所有实验中，模型的超参数设置如表 1 所示。

表 1 超参数设置

参数名	参数值
学习率（Learning Rate）	0.001
训练轮数（Epochs）	500
Transformer 层数	4
batch_size	64
多头注意力数量	8
嵌入维度（d_model）	128
Dropout	0.2
输入序列长度（CL）	20
预测步长	10

输入序列的上下文长度固定为 20，预测未来访问步长为 10，实验在不同缓存容量设置下进行评估。所有实验均在实际油田数据集上进行，实验总轮数为 500 轮。

2.2.1 不同缓存容量下的缓存命中率对比

第一组实验用于分析模型在不同缓存容量大小下的性能表现。实验选取了 LRU、LFU、FIFO 三种常用的缓存替换策略，以及原始 Transformer 模型作为对比基线。

实验过程中，模型对未来 10 次数据访问请求进行预测，并根据预测结果提前加载数据，最后统计实际的缓存命中情况。

如图 4 所示，横轴表示缓存容量，纵轴表示缓存命中率。从结果可以看出，当缓存容量从 600 增加到 1 000 时，命中

率整体呈上升趋势，从 68.2% 逐步提升到 75.8%。

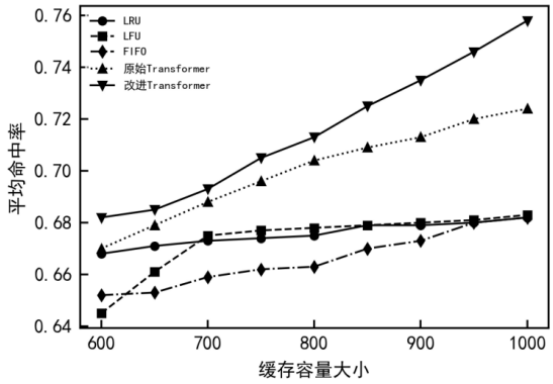

图 4 不同缓存大小下的命中率对比

与传统策略相比，改进 Transformer 模型在不同缓存容量下均取得了更好的结果：在最小容量 600 时，较原始 Transformer 提升 1.2 %；随着容量增大，在 1 000 容量时，提升幅度达 3.4%。

实验结果表明，本文提出的模型在不同缓存容量条件下均具有较好的预测效果。相比传统方法，改进 Transformer 能够更有效地学习数据访问序列中的变化规律，对长期依赖关系的捕捉能力更强，从而能够提升缓存命中率和资源利用效率。

2.2.2 不同上下文长度对性能的影响

第二组实验主要研究输入序列长度（CL）对模型预测性能的影响。在其他实验参数保持不变的条件下，将上下文长度从 20 逐步增加到 80，观察不同历史窗口下缓存命中率的变化情况。

如图 5 所示，在不同窗口长度条件下，缓存容量增加时，缓存命中率整体都会有所提高，说明缓存容量仍然是影响性能的重要因素。另一方面，在固定缓存容量时，随着 CL 不断增大，模型命中率整体呈下降趋势，存在 CL=20 的最优窗口阈值。

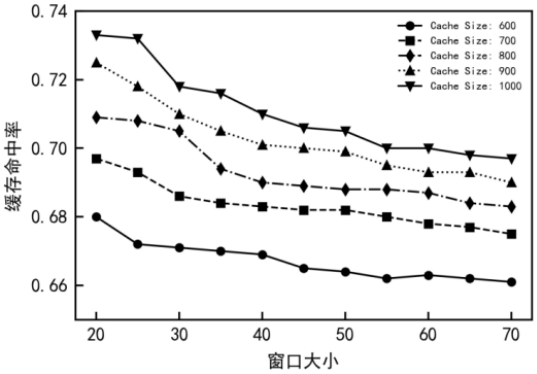

图 5 改进 Transformer 在不同 CL 长度下命中率

造成这一现象的原因主要有两方面。一方面由于稀疏注意力机制和编码器结构限制了长序列建模能力；另一方面，实验数据中的访问周期主要集中在 20 左右区间，因此较短的上下文窗口已经能够覆盖主要时序规律。在缓存预测任务中并非窗口越大越好，而是存在一个与模型结构和数据特征相关的最优上下文长度。

2.2.3 基准模型比较

为进一步验证模型的预测性能，本文选取五种常见的时间序列预测模型进行对比实验。具体包括：

（1）LSTM：长短期记忆网络，一种典型循环神经网络结构，能够学习时间序列中的长期依赖关系；

（2）Seq2Seq：基于编码器 – 解码器结构的序列到序列预测模型；

（3）TCN：一种用于处理时序数据的新型神经网络模型，通过因果卷积建模长期的时序依赖关系 [24]；

（4）N-BEATS：基于全连接网络构建的时间序列预测模型，在多个公开数据集上具有较好的预测性能 [25]；

（5）Transformer：标准全连接注意力机制的 Transformer 架构。

所有模型均在相同的数据集与实验环境下进行训练与测试，超参数调优原则保持一致，以保证对比结果的公平性。

本实验采用平均绝对误差（MAE）、均方根误差（RMSE）和对称平均绝对百分比误差（SMAPE）作为评价指标，计算公式分别为：

$$E_{\mathrm{MA}}=\frac{1}{N}\sum_{i=1}^{N}\left|y_i-\hat{y}_i\right| \tag{18}$$

$$E_{\mathrm{RMS}}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\left|y_i-\hat{y}_i\right|\right)^2} \tag{19}$$

$$E_{\mathrm{SMAP}}=\frac{100\%}{N}\sum_{i=1}^{N}\frac{\left|y_i-\hat{y}_i\right|}{\left(\left|y_i-\hat{y}_i\right|\right)/2} \tag{20}$$

式中：$\hat{y}_i$为预测值，y_i为真实值。

实验结果如表 2 所示，可以看出，相比其他模型，改进 Transformer 缓存预测模型在 MAE、RMSE、SMAPE 三项指标上均取得了较低的误差。该结果表明，改进 Transformer 模型在预测任务中具有更好的预测精度与稳定性。

表 2 不同模型的性能比较

基准模型	E_{MAE}	E_{RMSE}	E_{SMAPE}/%
LSTM	0.232 65	0.290 88	24.35
Seq2Seq	0.325 21	0.426 52	26.95
TCN	0.251 67	0.290 93	24.21
N-BEATS	0.252 72	0.291 94	24.48
Transformer	0.157 65	0.225 01	23.56
本文方法	0.118 65	0.168 57	21.68

由于油田数据访问日志具有热点集中、大部分对象访

问稀疏等特点，部分低频对象的预测误差较大，导致整体 SMAPE 偏高，但仍优于基准模型在相同条件下的表现。

2.2.4 消融实验分析

改进 Transformer 模型包括稀疏注意力机制（sparse attention）、非自回归解码器（non-autoregressive decoder）、注意力蒸馏（attention distillation）和学习率调度器（learning rate scheduler）。为验证各部分对模型性能的影响，本文设计了 5 种网络：

（1）Base 在改进 Transformer 基础上将稀疏注意力机制、非自回归编码器、注意力蒸馏和学习率调度器模块去除，即原始 Transformer；

（2）Base+S 在 Transformer 模型基础上引入稀疏注意力机制；

（3）Base+S+N 添加了非自回归解码器；

（4）Base+S+N+A 添加了注意力蒸馏模块；

（5）Base+S+N+A+L 引入了学习率调度器。

消融结果如表 3 所示。

表 3 不同缓存容量下的消融实验结果

模型变体	600	700	800	900	1 000
Base	67.57	68.05	69.24	69.15	70.02
Base+S	67.66	69.15	70.09	70.52	71.58
Base+S+N	67.88	69.78	70.64	70.74	71.85
Base+S+N+A	66.60	67.87	68.63	69.73	70.09
Base+S+N+A+L	68.25	69.72	71.61	73.69	75.52

表 3 为五个变体模型在不同缓存容量（600~1 000）下的缓存命中率对比实验结果。可以看出 Base 模型在不同缓存容量大小下的命中率较低，说明在去掉核心模块后的模型能力有所下降。

在加入稀疏注意力机制（Base+S）后，模型在各缓存容量下的命中率均有所提高，平均提升约 1.2%，表明稀疏注意力能够在长序列中筛选出更重要的关联信息，从而增强模型对关键时序特征的学习能力。

在此基础上进一步加入非自回归解码器（Base+S+N）后，模型性能继续上升 0.7%。表明非自回归结构摆脱了传统自回归方法对前一步预测强依赖的限制，可以同时完成多步预测，提高了模型的并行处理能力与预测效率。

在继续引入注意力蒸馏模块（Base+S+N+A）后，模型命中率在部分缓存容量下出现了小幅下降，因为当前任务以访问序列建模为主，特征维度相对较低，注意力蒸馏压缩后可能损失部分关键上下文信息，进而影响预测结果。

最后，在加入学习率调度策略（Base+S+N+A+L）后，模型性能再次提升，达到了全组实验中的最优结果。这说明动态学习率调整能够提高模型训练过程中的稳定性，使模型在后期训练阶段获得更好的收敛效果和泛化能力。

实验结果表明，本文提出的改进 Transformer 模型在缓存预测任务中具有较强的性能优势，能够有效提升缓存命中率，该方法在实际油气勘探数据处理场景中具有良好的应用前景。

3 结语

本文针对油气勘探开发过程中数据访问规模大、访问模式复杂以及传统缓存管理效率不足等问题，提出了一种基于改进 Transformer 架构的缓存预测方法。该方法通过引入稀疏注意力机制以降低计算复杂度，结合注意力蒸馏模块提升特征表达能力，并结合非自回归解码结构以实现高效多步预测。同时，本文还设计了动态学习率调度机制，进一步提升模型的训练稳定性与收敛速度。

在真实油田访问日志数据上的实验结果表明，本文方法在不同缓存容量与不同上下文长度下均能够保持较稳定的预测性能。与传统缓存策略以及标准 Transformer 模型相比，本文方法在缓存命中率等指标上取得了更好的结果，并在一定程度上降低了数据访问延迟。消融实验结果也验证了各个模块对模型性能提升的作用。

综上所述，本文所提出的方法对油气勘探场景下的数据缓存管理具有一定的应用价值。未来将进一步扩展该方法的适用性与可落地性，为智能油气信息系统建设提供有力支撑。

参考文献：

[1] 张俊 . 基于 Redis 实现关系型数据库内存化研究 [D]. 成都 : 四川师范大学 , 2021.

[2] 张蒙蒙 , 曹成茂 . 大数据时代 Redis 缓存的性能分析与优化 [J]. 巢湖学院学报 ,2022,24(3):80-87.

[3]Megiddo N, Modha D S. Outperforming LRU with an adaptive replacement cache algorithm[J]. Computer, 2004, 37(4): 58-65.

[4] 李洁 , 林永峰 . 基于多时间尺度 RNN 的时序数据预测 [J]. 计算机应用与软件 , 2018,35(7):33-37.

[5] 谷建伟 , 周梅 , 李志涛 , 等 . 基于数据挖掘的长短期记忆网络模型油井产量预测方法 [J]. 特种油气藏 ,2019,26(2): 77-81.

[6]Fu Rui, Zhang Zuo, Li Li. Using LSTM and GRU neural network methods for traffic flow prediction[C]// 2016 31st Youth Academic Annual Conference of Chinese Association of Au-

tomation (YAC).Piscataway:IEEE,2016: 324-328.

[7] 蔡万升 , 宋曦 . 电网大数据环境中基于 LSTM 的边缘缓存预测 [J]. 电力信息与通信技术 ,2022,20(12):73-80.

[8] 邓田丰 , 李梓铭 , 张少波 , 等 . 基于机器学习和时间序列算法融合的北京地区花粉浓度预测技术研究 [J]. 中国环境科学 , 2025, 45(11):6007-6019.

[9] 陈喜群 , 祝文琪 , 吕朝锋 . 融合轨迹时序与行为修正的车辆冲突风险预测 [J]. 交通运输系统工程与信息 , 2025, 25(4): 219-229.

[10]Kim H, Sun T-J, Huh E-N. T-CacheNet: transformer-based deep reinforcement learning for next-generation internet content caching[C]//Proceedings of the 2024 13th International Conference on Networks, Communication and Computing. NewYork: ACM, 2025: 9-16.

[11] 李晖 . 磁盘存储中基于自适应分区的 I/O 调度算法的研究 [D]. 重庆 : 西南大学 ,2019.

[12]Cai Ling, Janowicz K, Mai Gengcheng, et al. Traffic transformer: capturing the continuity and periodicity of time series for traffic forecasting[J]. Transactions in GIS, 2020, 24(3): 736-755.

[13] 刘钟 , 唐宏 , 王宁喆 , 等 . 融合 RNN 与稀疏自注意力的文本摘要方法 [J]. 计算机工程 ,2025,51(1):312-320.

[14]Yan Jichuan, Hu Lin, Zhen Zhao, et al. Frequency-domain decomposition and deep learning based solar PV power ultra-short-term forecasting model[J]. IEEE Transactions on Industry Applications, 2021, 57(4): 3282-3295.

[15] 高凯 , 刘欣宇 , 胡林 , 等 . 基于稀疏注意力的时空交互车辆轨迹预测 [J]. 汽车工程 ,2025,47(5):809-819.

[16]Zhou Haoyi, Zhang Shanghang, Peng Jieqi, et al. Informer: beyond efficient transformer for long sequence time-series forecasting[PP/OL]. (2021-03-28)[2026-05-28].https://doi.org/10.48550/arXiv.2012.07436.

[17] 吕敬钦 , 胡朗 , 梁炜楠 , 等 .COVID-19 影像诊断的注意力蒸馏对比互学习模型 [J]. 计算机工程 ,2025,51(8):373-382.

[18] 郑筠 , 高朋 . 基于知识蒸馏的卷积神经网络压缩方法 [J]. 沈阳工业大学学报 ,2025,47(3):348-354.

[19] 张煜松 . 基于非自回归方法的图像描述研究及应用 [D]. 北京 : 北京邮电大学 ,2024.

[20] 姚森山 , 庞成鑫 . 基于改进残差网络的光伏逆变器数据异常检测方法 [J]. 太阳能学报 ,2025,46(4):322-330.

[21]Smith L N. A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay[PP/OL]. (2018-04-28)[2026-05-28].https://doi.org/10.48550/arXiv.1803.09820.

[22] 刘国权 , 陈尚良 , 李跃忠 , 等 . 基于 SGD 和余弦退火算法改进 YOLOv3 的高压电力设备目标检测方法 [J]. 东华理工大学学报 (自然科学版),2024,47(3):294-300.

[23] 肖雄 . 工业应用中基于深度学习的缓存策略研究 [D]. 江苏 : 南京邮电大学 ,2023.

[24] 袁和金 , 陈龙 , 李金波 , 等 . 基于 DBSCAN 和 VMD-TCN-LSTM 的变压器油中溶解气体预测方法 [J]. 电力信息与通信技术 ,2025,23(9):28-34.

[25] 李泽龙 , 乔钢柱 , 崔方舒 , 等 . 基于 N-BEATS 和相关向量机的锂电池健康状态混合预测方法 [J]. 中北大学学报 (自然科学版),2025,46(3):316-325.

【作者简介】

景瑞林（1969—），男，山西临汾人，硕士，高级工程师、油田高级专家，研究方向：数据库技术与大数据应用。

杨硕（1982—），男，湖北荆州人，本科，高级工程师，研究方向：数据治理。

李玲（1972—），女，山东潍坊人，硕士，高级工程师，研究方向：油田企业勘探开发信息系统建设。

韩明（1979—），男，山东青岛人，硕士，高级工程师，研究方向：数据技术。

杜心萌（2003—），通信作者（email:15621465376@163.com），女，山东烟台人，硕士，研究方向：时间序列预测研究。

王烈冲（2002—），男，山东聊城人，硕士，研究方向：时间序列预测研究。

（收稿日期：2025-11-19 修回日期：2026-06-09）