doi：10. 12068/j. issn. 1005－3026. 2026. 20240176

门控驱动的时间与变量信息融合的时间序列预测方法

赵煜辉[1]，郭睿鑫[2]
（1. 东北大学秦皇岛分校 计算机与通信工程学院，河北 秦皇岛 066003；
2. 国网河南省电力公司 濮阳供电公司，河南 濮阳 457000）

摘 要：现有的时间序列预测模型在特征提取过程中忽视了原始数据所蕴含的变量信息与时间信息的有效融合，这将导致原始数据的信息无法被充分利用．本文提出了门控驱动的时间与变量信息融合的时间序列预测模型（GFTV），该模型结合趋势季节分解技术，使用Transformer架构提取原始数据的跨时间特征和跨变量特征，利用一种基于门控单元的门控融合机制将跨时间特征和跨变量特征融合，提高模型对数据信息的获取能力．该模型在ETTh1和Weather数据集上均实现了最优的性能，消融实验验证了GFTV所包含模块的有效性，为时间序列预测领域提供了动态融合时间信息与变量信息的新思路．

关 键 词：时间序列预测；门控融合；Transformer；跨时间特征；跨变量特征；特征融合；趋势季节分解
中图分类号：TP 18 **文献标志码**：A **文章编号**：1005-3026（2026）02-0018-06

Gated-Driven Time Series Forecasting Method Based on Temporal and Variable Information Fusion

ZHAO Yu-hui[1]，*GUO Rui-xin*[2]
（1. School of Computer and Communication Engineering，Northeastern University at Qinhuangdao，Qinhuangdao 066003，China；2. Puyang Power Supply Company，State Grid Henan Electric Power Company，Puyang 457000，China. Corresponding author：GUO Rui-xin，E-mail：1040980690@qq.com）

Abstract：Existing time series forecasting models overlook the effective fusion of variable information and temporal information embedded in raw data during feature extraction，which results in the failure to fully utilize the information of raw data. To address this issue，a gated-driven time series forecasting method based on temporal and variable information fusion（GFTV）was proposed. This model leveraged trend-seasonal decomposition techniques and employed a Transformer architecture to extract both cross-temporal and cross-variable features from the raw data. Subsequently，a gating fusion mechanism based on gating units was used to fuse the cross-temporal and cross-variable features，enhancing the model's ability to capture data information. The model achieved optimal performance on both the ETTh1 and Weather datasets，and ablation studies confirmed the effectiveness of the modules contained in GFTV，providing a novel approach for dynamically fusing temporal and variable information in the field of time series forecasting.

Key words：time series forecasting；gating fusion；Transformer；cross-temporal feature；cross-variable feature；feature fusion；trend-seasonal decomposition

时间序列数据有两个维度，分别是时间维度和变量维度．时间序列数据的时间信息蕴含于时间维度之中，反映数据随时间的变化情况．时间序列数据的变量信息蕴含于变量维度之中，反映同一时间段不同变量之间存在的潜在联系．能否充分提取原始数据的时间依赖和变量依赖极大地影响着预测的精度．

Transformer[1]拥有可直接建模序列不同位置

收稿日期：2024-09-27
基金项目：国家自然科学基金资助项目（72371067）．
作者简介：赵煜辉（1971—），男，黑龙江嫩江人，东北大学副教授．
通信作者：郭睿鑫，E-mail:1040980690@qq.com

间依赖关系的特性,被广泛用于时间序列预测领域[2].以提取变量之间联系的方式作为划分标准,基于Transformer的时间序列预测模型可分为3类.第1类:使用1个嵌入层,实现对变量间关系的编码.这种简单地将变量之间的关系通过1个嵌入层编码,无法有效地提取变量之间的联系,许多基于Transformer架构的时间序列预测模型,如Informer[3],Autoformer[4],FEDformer[5]等,均使用了这种思路.第2类:使用通道独立的思想,不考虑变量之间的关系,本质是将多元时间序列预测简化为单元时间序列预测,PatchTST[6]将这种思路引入基于Transformer的时间序列预测模型中.第3类:专门设计1种模块来提取变量之间的联系特征,如Client[7],iTransformer[8]等.

Informer[3],Autoformer[4]等第1类时间序列预测模型在1个时间点上将数据的多个变量混合至不同的通道之中作为1个词元(token),来混合变量特征,之后使用Transformer的编码层来提取各个token对应时间点的混合信息.这里存在2个问题:①在混合物理含义常完全不同的变量特征时,会有大量冗余特征的出现;②由于传感器等收集数据的设备不一致,再加上设备存在延迟,很难收集到真正意义上同一时刻不同变量的数据.而第2、第3类时间序列预测模型并没有考虑到时间特征与变量特征有效融合的问题,导致模型忽略时间特征与变量特征的相互作用,无法充分利用数据中的全部信息,使信息丢失,进而影响模型的预测精度.

基于以上分析,本文通过分别提取时间序列数据的时间特征与变量特征,再将二者融合来解决上述问题.首先,在提取时间特征时,将多变量时间序列预测转变为单变量时间序列预测,以某时间步的单个变量对应的数据作为token,使用Transformer编码层提取纯粹的时间特征,这样从根本上避免在混合变量信息时时间无法对齐的问题.其次,在提取变量特征时,将时间序列数据的每个变量嵌入1个token,即将同一变量不同时间点的值混合在1个token之中,使模型考虑到更多的局部信息,再使用Transformer编码层来提取变量之间的依赖.最后,使用门控融合机制把提取到的时间信息与变量信息融合,使数据中的时间信息与变量信息充分交互,使模型得到更多有用的信息,进而得到更高的预测精度.

1　GFTV模型的整体框架与构建

使用门控融合机制将从跨时间和跨变量两个角度提取的数据特征融合,可充分提取数据中时间信息与变量信息的交互特征,有助于得到更高的预测精度,因此,本文提出门控驱动时间与变量信息融合的时间序列预测模型(GFTV),其整体框架如图1所示.首先,将输入数据X分解成趋势项$X(Q)$与季节项$X(J)$.对易于预测的趋势项,GFTV选择结构简单的线性层直接将其映射成趋势项预测结果.对较为复杂的季节项,分别提取跨时间特征和跨变量特征,再将提取到的特征进行维度对齐,使用门控融合机制,将融合后的特征输入投影层来生成需要的季节项未来数据.最终,使用门控融合机制自动学习趋势项与季节项预测结果的权重,将二者加权融合,经过线性层映射得到最终的预测结果Y.

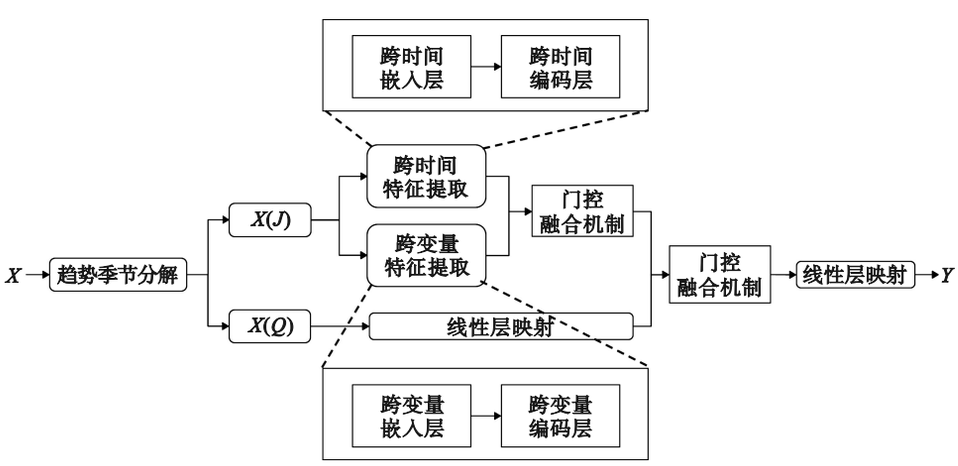


图1　GFTV模型整体框架

Fig. 1　Overall framework of GFTV model

跨变量特征与跨时间特征的提取架构如图2所示.季节项数据$X(J)$在两大编码器中的数据流向,GFTV使用跨变量嵌入层和跨变量编码层提取出变量之间的联系,使用跨时间嵌入层和跨时间编码层得到跨时间特征,图2中N表示编码层层数,B为批量维度,V为变量维度,T为时间维度.

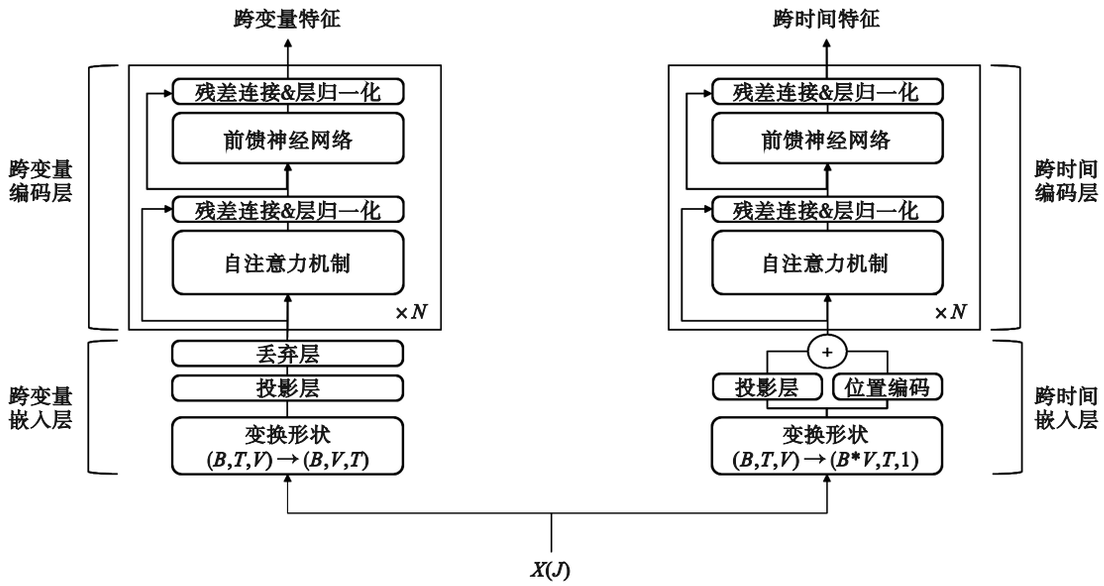


图2 跨变量特征提取和跨时间特征提取架构

Fig. 2 Cross-variable feature extraction and cross-temporal feature extraction architecture

1.1 趋势项季节项分解模块

GFTV使用移动平均分解技术将数据分为趋势项与季节项,从而更清晰地识别时间序列中的规律和变化趋势.通过使用滑动窗口计算时间序列的局部均值,有效去除短期波动成分,生成趋势项.原输入数据减去趋势项数据即得到季节项数据.

$$\left.\begin{aligned} x_t(Q) &= \frac{1}{k}\sum_{i=-\frac{k-1}{2}}^{\frac{k-1}{2}} x_{t+i}, \\ x_t(J) &= x_t - x_t(Q). \end{aligned}\right\} \tag{1}$$

式中:$x_t(Q)$,$x_t(J)$分别为趋势项$X(Q)$和季节项$X(J)$的第t个时间步的数据;k为滑动窗口长度,常设为奇数.

1.2 跨变量特征提取模块

为捕捉变量之间更复杂的非线性依赖关系,本文设计了跨变量特征提取模块,分为跨变量嵌入层及跨变量编码层.跨变量嵌入层将同一变量不同时间步的值嵌入到token中,跨变量编码层采用自注意力机制,以提取变量之间的信息.图2的左边部分为跨变量特征提取模块的架构.

在跨变量嵌入层中,GFTV将季节项数据$X(J)\in\mathbf{R}^{(B\times T\times V)}$的变量维和时间维交换,得到$X'(J)\in\mathbf{R}^{(B\times V\times T)}$.由于变量之间没有先后关系,GFTV去掉了原Transformer模型中的位置编码层,使用投影层将数据的时间维度投影到指定的模型维度d_k上,这样便将单个变量所对应的时间步映射到token中,接着使用丢弃层以降低过拟合的可能,得到的结果即跨变量编码层的输入.

跨变量编码层主要通过自注意力机制和前馈神经网络来学习输入序列中各个变量token之间的依赖关系.自注意力机制通过生成查询、键和值矩阵,以及点积计算来求得不同变量token之间的相关性(即注意力权重),利用所求得的注意力权重来自适应地调整每个token对其他token的关注程度[9].同时,为降低模型的时间和空间复杂度,GFTV引入Informer提出的概率稀疏自注意力机制[3].

1.3 跨时间特征提取模块

为避免不同设备采集到的数据在时间上不严格对齐和混合不同物理意义的变量特征时产生的冗余特征,本文设计跨时间特征提取模块,分为跨时间嵌入层及跨时间编码层.图2的右边部分为跨时间特征提取的架构.

在跨时间嵌入层中,GFTV将输入数据$X(J)\in\mathbf{R}^{(B\times T\times V)}$的变量维度并入批量维度来达到隔离变量的目的,得到结果$X'(J)\in\mathbf{R}^{(B*V\times T\times 1)}$.使用值嵌入和位置嵌入对数据进行编码,再将两种编码结果直接相加,从而将单个变量所对应的某

一时间步映射到 token 中．值嵌入使用投影层将最后一维投影到模型维度 d_k 上，位置编码将数据的位置信息融入模型之中．

跨时间编码层的自注意力机制仅计算代表同一变量不同时刻的各 token 之间的相关性，来提取跨时间维度的特征，而不再考虑变量之间的联系．跨时间编码层模块与跨变量编码层模块的设计是一致的，故不赘述．

1.4　门控融合机制模块

借鉴 LSTM 中门控机制的原理，门控融合机制能够根据输入数据自适应地为各个特征分配不同的权重，从而动态地融合特征．通过这种机制，GFTV 能够灵活、自动地融合跨时间特征与跨变量特征，以及趋势项与季节项．

以跨时间特征与跨变量特征融合为例，图 3 描述了门控融合机制的结构，σ 为门神经元．门神经元接收跨时间特征与跨变量特征，并从中学习 1 组权重，该权重决定了这 2 个特征对整体输出的贡献度．各特征与所对应的权重相乘后相加，得到融合后的总特征．

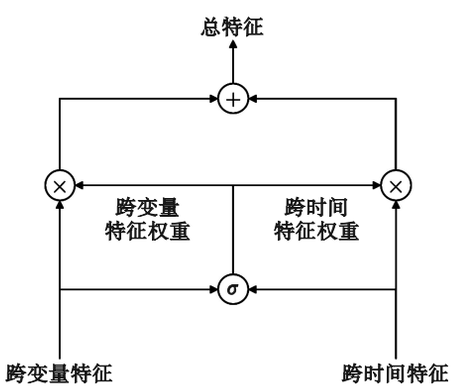


图 3　门控融合机制

Fig. 3　Gating fusion mechanism

门控融合机制计算跨变量特征与跨时间特征的和，并通过线性变换与 Sigmoid 激活函数计算出跨变量特征的权重．由于 GFTV 仅需要为两个特征赋权重来有效融合时间与变量信息，而两个特征的重要性是相对的．依据这一性质，令二者权重之和为 1，此时，只需要求出跨变量特征的权重，再对其取补值即可得到跨时间特征的权重．门控融合机制的公式为

$$\left.\begin{aligned} \boldsymbol{W}_V &= S\left[\boldsymbol{W}\left(F_V + F_T\right)\right], \\ \boldsymbol{W}_T &= 1 - \boldsymbol{W}_V, \\ F &= F_V \boldsymbol{W}_V + F_T \boldsymbol{W}_T. \end{aligned}\right\} \tag{2}$$

式中：F_V 为跨变量特征，F_T 为跨时间特征，F 为总特征，S 为 Sigmoid 激活函数，$\boldsymbol{W}$ 为权重矩阵；$\boldsymbol{W}_V$ 为变量信息权重矩阵；$\boldsymbol{W}_t$ 为时间信息权重矩阵．

2　实验结果与讨论

2.1　数据集与实验设置

使用 ETTh1 和 Weather 两个被时间序列预测领域广泛使用的公开数据集来对 GFTV 的预测精度进行评估．ETTh1 数据集有 17 420 个时间步和 7 个变量，是由国家电网收集的为期两年的中国某县的电力变压器温度数据集，时间步之间间隔 1 h．Weather 数据集有 52 696 个时间步和 21 个变量，时间间隔为 10 min，收集了 2020 年全年的气候数据．这 2 个数据集可在 Autoformer 论文中[4]获取．

使用 5 种不同的预测长度，即 48，96，192，336，720，对模型的预测能力进行评估，采用均方误差（MSE）和平均绝对误差（MAE）评价 GFTV 并与其他基准模型进行对比．

参数设置如下：编码层层数为 2，批处理大小为 4，回溯窗口为 96．针对较小的数据集 ETTh1，注意力头数为 3，模型隐藏状态维度设为 64，全连接层模型维度设为 96；针对较大的数据集 Weather，注意力头数为 4，模型隐藏状态维度为 192，全连接层模型维度设为 336，所有实验均在 12 GB 的 NVIDIA GeForce RTX 4060 Laptop GPU 上完成．

2.2　主要实验及结果；

GFTV 模型与其他对比模型（iTransformer[8]，Client[7]，Autoformer[4]，Stationary[10]，PatchTST[6]，Timesnet[11]）在 ETTh1 和 Weather 数据集上的多步时间序列预测结果分别如表 1 和表 2 所示．最好的结果被加粗标注，由表 1 和表 2 结果可知，GFTV 模型在两大数据集上的预测任务都取得了最佳的性能，这说明使用门控融合机制融合跨时间和跨变量特征、融合趋势项与季节项的方法对模型性能的提高有着相当大的作用．

与专注于提取变量特征的 iTransformer 相比，GFTV 在 Weather 数据集上，均方误差降低 4.9%，与专注于提取时间特征的 PatchTST 相比，GFTV 在 Weather 数据集上，均方误差降低 4.6%．说明在预测过程中，忽略时间信息与变量信息中的任意一方都会导致模型提取到的信息不全面，唯有将二者提取并有效融合才能最大化地利用原始数据所蕴含的信息．

在 ETTh1 数据集上，GFTV 模型在预测时间步为 720 的预测任务上与 PatchTST 模型有一定的差距，这是由于 PatchTST 模型使用 Patch 分割

技术将多个时间点的数据合并为1个时间段,从而帮助模型提取到局部特征并降低了计算复杂度,使PatchTST在更长的预测时间步上具有更好的预测能力.

表1 GFTV与对比模型在ETTh1上的预测结果

Table 1 Prediction results of GFTV and comparison models on ETTh1

预测	GFTV		Timesnet		Stationary		Autoformer		Client		iTransformer		PatchTST	
时间步	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
48	**0.335**	**0.373**	0.365	0.400	0.504	0.471	0.439	0.438	0.347	0.382	0.342	0.380	0.357	0.386
96	**0.377**	**0.398**	0.389	0.412	0.573	0.506	0.491	0.471	0.393	0.407	0.387	0.405	0.396	0.410
192	**0.429**	**0.429**	0.440	0.442	0.592	0.520	0.485	0.476	0.446	0.436	0.441	0.436	0.446	0.435
336	**0.470**	**0.452**	0.492	0.469	0.653	0.550	0.502	0.490	0.485	0.455	0.491	0.462	0.492	0.458
720	0.493	0.481	0.521	0.496	0.642	0.574	0.504	0.506	0.500	0.485	0.508	0.508	**0.480**	**0.471**

注:加粗表示最优值,下同.

表2 GFTV与对比模型在Weather上的预测结果

Table 2 Prediction results of GFTV and comparison models on Weather

预测	GFTV		Timesnet		Stationary		Autoformer		Client		iTransformer		PatchTST	
时间步	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
48	**0.131**	**0.174**	0.138	0.191	0.146	0.195	0.221	0.301	0.132	**0.174**	0.136	**0.174**	0.143	0.182
96	**0.163**	**0.208**	0.168	0.219	0.200	0.243	0.268	0.336	0.166	0.211	0.176	0.214	0.179	0.220
192	**0.210**	**0.251**	0.224	0.264	0.257	0.294	0.354	0.404	0.212	0.252	0.228	0.260	0.225	0.259
336	**0.270**	**0.295**	0.280	0.302	0.295	0.321	0.348	0.384	**0.270**	**0.295**	0.282	0.300	0.278	0.298
720	**0.348**	**0.345**	0.355	0.352	0.405	0.390	0.430	0.435	0.351	0.347	0.358	0.349	0.351	0.346

为直观展示各大模型的预测能力,本章将各模型在Weather数据集上的变量OT的预测结果进行可视化对比,其中OT表示二氧化碳质量分数,如图4所示,图4为GFTV,Stationary,iTransformer,Client的预测结果可视化,回溯窗口与预测长度均设为96.横坐标表示时间步,纵坐标表示预测值与实际值.浅色曲线为各模型预测值,深色曲线为实际值.从图中可以看到,GFTV的两条曲线贴合得最为密切,这说明GFTV模型的预测性能优于其他模型,这也体现了使用门控融合机制融合时间信息与变量信息的有效性.

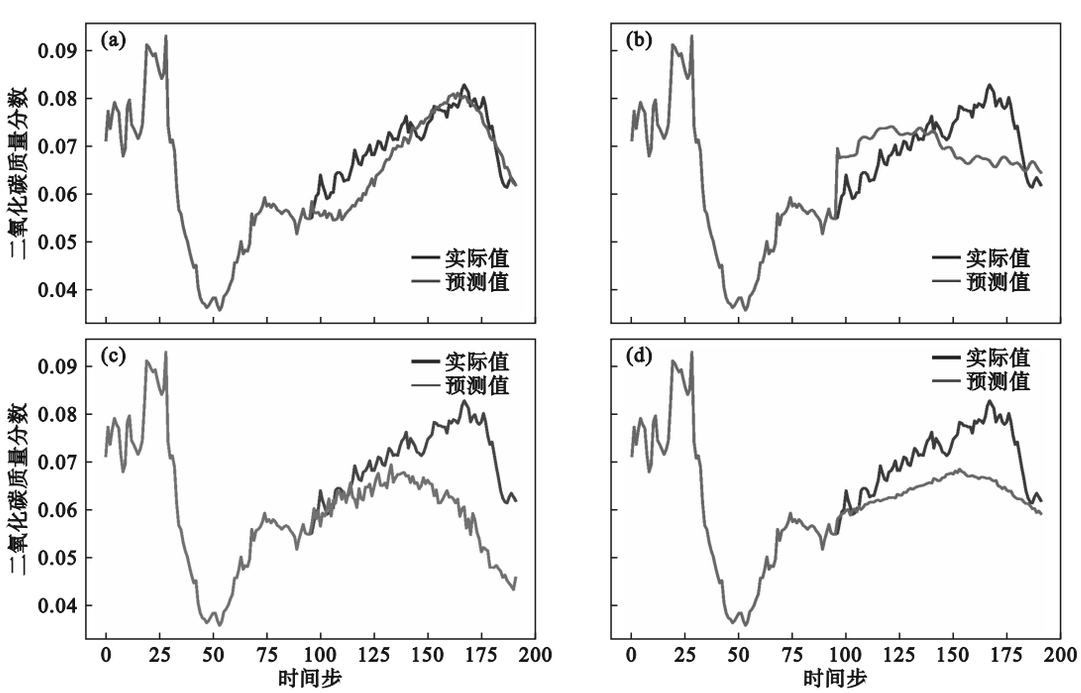


图4 GFTV与其他模型的结果可视化

Fig. 4 Visualization of results of GFTV and other models

(a)—GFTV;(b)—Stationary;(c)—iTransformer;(d)—Client.

2.3　消融实验

为验证GFTV模型各模块的合理性,并评估各模块对最终性能的贡献和影响,本文依次移除GFTV中的各个模块,在ETTh1数据集上设计了一系列的消融实验,实验结果如表3所示.其中,去时间模型表示在对季节项预测时去除跨时间特征模块;去变量模型表示在对季节项预测中去除跨变量特征模块;去时变门控模型在对季节项预测中使用手动调节两特征权重来代替门控融合机制;去趋季门控模型表示去掉趋势季节项融合时的门控融合机制而直接将二者相加的模型;去趋季模型去除了趋势季节分解模块.GFTV取得了最佳的结果,表明GFTV的各模块都有不可或缺的作用.

表3　GFTV与各消融实验模型的预测结果

Table 3　Prediction results of GFTV and ablation study models

预测	**GFTV**		去时间模型		去变量模型		去时变门控模型		去趋季门控模型		去趋季模型	
时间步	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
48	**0.335**	**0.373**	0.342	0.376	0.346	0.376	0.342	0.376	0.343	0.376	0.340	0.376
96	**0.377**	**0.398**	0.385	0.400	0.390	0.400	0.385	0.400	0.386	0.400	0.384	0.401
192	**0.429**	**0.429**	0.430	0.431	0.448	0.436	0.441	0.433	0.441	0.433	0.441	0.435
336	**0.470**	**0.452**	0.474	0.453	0.491	0.457	0.481	0.454	0.481	0.453	0.479	0.454
720	**0.493**	0.481	0.501	0.484	0.505	0.485	0.494	0.478	0.494	0.478	**0.493**	**0.477**

3　结　　语

为了能够动态调整预测模型对时间信息与变量信息的关注权重,本文提出了一种门控驱动的时间与变量信息融合的时间序列预测模型,即GFTV.该模型将原始数据分解为趋势项与季节项,并使用门控融合机制将时间信息与变量信息动态融合.GFTV模型在两个公开数据集上的预测任务中实现了最小的误差,结合与其他模型预测结果的直观对比分析,本文得出GFTV具有更优性能的结论.通过一系列消融实验,证明了GFTV各模块的有效性.本文为时间序列预测领域中时间信息与变量信息的融合提供了一种新颖且有效的解决方案.

参考文献:

[1] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach,2017: 6000-6010.

[2] Zeng A L, Chen M X, Zhang L, et al. Are transformers effective for time series forecasting? [EB/OL]. (2023-06-26)[2024-07-23]. https://doi.org/10.1609/aaai.v37i9.26317.

[3] Zhou H Y, Zhang S H, Peng J Q, et al. Informer: beyond efficient transformer for long sequence time-series forecasting [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, 35(12): 11106-11115.

[4] Wu H X, Xu J H, Wang J M, et al. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting[J]. *Advances in neural information processing systems*,2021, 34:22419-22430.

[5] Zhou T, Ma Z Q, Wen Q S, et al. FEDformer: frequency enhanced decomposed transformer for long-term series forecasting [C]// International conference on machine learning. Baltimore.2022:17-23.

[6] Nie Y Q, Nguyen N H, Sinthong P, et al. A time series is worth 64 words: long-term forecasting with transformers [EB/OL]. (2022-11-27)[2024-07-23]. https://arxiv.org/abs/2211.14730.

[7] Gao J X, Hu W B, Chen Y T. Client: cross-variable linear integrated enhanced transformer for multivariate long-term time series forecasting[EB/OL]. (2023-05-23)[2024-09-23]. https://arxiv.org/abs/2305.18838.

[8] Liu Y, Hu T G, Zhang H R, et al. iTransformer: inverted transformers are effective for time series forecasting [EB/OL]. (2023-10-10) [2024-09-23]. https://arxiv.org/abs/2310.06625.

[9] Li S Y, Jin X Y, Xuan Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting [C]//33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Vancouver. 2019: 1-11.

[10] Liu Y, Wu H X, Wang J M, et al. Non-stationary transformers: exploring the stationarity in time series forecasting[EB/OL]. (2023-11-24)[2024-07-23]. https://arxiv.org/abs/2205.14415

[11] Wu H X, Hu T G, Liu Y, et al. TimesNet: temporal 2D-variation modeling for general time series analysis [EB/OL]. (2023-04-12) [2024-09-23]. https://arxiv.org/abs/2210.02186.