

# 多维注意力机制改进燃气负荷预测<sup>①</sup>



曹辰光, 徐晓钟

(上海师范大学 信息与机电工程学院, 上海 201400)

通信作者: 曹辰光, E-mail: [1046068454@qq.com](mailto:1046068454@qq.com)

**摘要:** 燃气负荷预测是一项让城市安全经济地调配燃气的重要的工作. 目前, 基于注意力机制的 Seq2Seq 模型越来越多地应用于燃气数据预测, 是一种有效的燃气负荷预测方法. 然而, 针对燃气负荷数据这种突变频率高幅度大的特点, 一般基于注意力机制的 Seq2Seq 模型难以提取数据中的多维时间模式信息与应对数据随机突变情况, 在处理影响因素复杂的燃气负荷的预测问题时仍然需要改进. 为此, 本文提出多维注意力机制 Seq2Seq 模型. 一方面设计研究了多层次时间注意力模块, 综合单时间步长、多时间步长的注意力计算提取数据中不同时间模式信息; 另一方面, 设计增加了局部历史注意力模块. 以改进模型中无法区分重要历史信息的缺陷, 使模型在预测时倾向于参考更为重要的历史信息. 本改进模型针对燃气负荷的独特特性, 具有较好的预测表现. 使用国内某市区的燃气消耗数据与 2016 年电工数学建模竞赛的电力负荷数据的实验结果表明, 本改进模型相对于一般注意力机制 Seq2Seq 模型, *MAE* 分别降低了 17% 与 9%.

**关键词:** 燃气负荷预测; 注意力机制; Seq2Seq

引用格式: 曹辰光, 徐晓钟. 多维注意力机制改进燃气负荷预测. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9365.html>

## Improvement of Gas Load Forecasting by Multi-dimensional Attention Mechanism

CAO Chen-Guang, XU Xiao-Zhong

(College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai 201400, China)

**Abstract:** Gas load forecasting is an important task for cities to deploy gas safely and economically. At present, the Seq2Seq model based on the attention mechanism is increasingly utilized in gas data forecasting and is an effective method for gas load forecasting. However, the gas load data have such characteristics as high mutation frequency and large amplitude. The Seq2Seq model based on the general attention mechanism is difficult to extract the multivariate time pattern information in the data and deal with data random mutation. It is still necessary for improving gas load prediction with complex influencing factors. Therefore, this study proposes a multi-dimensional attention mechanism Seq2Seq model. On the one hand, a multi-level time attention module is designed and studied to integrate single-time step and multi-time step attention calculation to extract different time pattern information in the data. On the other hand, the design adds a local history attention module. By improving the model's defect of distinguishing important historical information, the model tends to refer to more important historical information when making predictions. The improved model has better prediction performance for the unique gas load characteristics. The gas consumption data of an urban area in China and the electric load data of the 2016 electrical mathematical modeling competition are taken as examples. The experimental results show that the *MAE* of the improved model is reduced by 17% and 9% respectively compared with the general attention mechanism Seq2Seq model.

**Key words:** gas load forecasting; attention mechanism; Seq2Seq

<sup>①</sup> 基金项目: 上海市科委项目 (115105024)

收稿时间: 2023-07-03; 修改时间: 2023-08-11; 采用时间: 2023-08-24; csa 在线出版时间: 2023-11-17

随着天然气的普及,各大燃气公司逐步推进燃气资源的智能化管理.为实现提前规划未来燃气用量,优化决策方案,降低运营成本的目的.准确的预测未来燃气数据的用量成为国内外学者的研究热点.近年来,受新型冠状病毒以及油价波动的影响,全世界的天然气产量下降,2022年上半年,国内天然气总供应量较历史同期减少了29亿立方米<sup>[1]</sup>.作为有限的清洁能源.准确的预测未来燃气消耗,对可持续发展有着重要的意义.

伴随深度学习模型的发展,Rumelhart等人<sup>[2]</sup>提出了专门处理时间序列任务的循环神经网络(recurrent neural network, RNN)模型,但传统RNN网络会出现梯度消失问题且无法处理数据中的长期依赖,随后,Hochreiter等人<sup>[3]</sup>提出长短期记忆网络(long short-term memory network, LSTM),解决RNN存在的梯度消失问题,同时能够较好地处理数据中的短期依赖与长期依赖.由此,LSTM网络被广泛应用于时间序列预测任务中.

为进一步提高RNN类网络在机器翻译中的性能,Cho等人<sup>[4]</sup>提出Encoder-Decoder架构,该架构使用两个RNN类网络分别作为Encoder与Decoder,输入数据通过Encoder编码成一个固定长度的上下文向量,Decoder依据该向量进行解码输出.由于所有信息都是压缩到一个定长向量中,当面对长输入数据时,就会出现信息丢失问题,模型表现性能迅速下降.于是,Bahdanau等人<sup>[5]</sup>提出基于注意力机制的Encoder-Decoder模型.通过注意力机制筛选最重要的信息,由此构成定长的上下文向量,尽可能降低信息丢失带来的影响.以提高模型的泛化能力.随着注意力模型的成功,在时间序列预测领域中也通过注意力机制来提升模型预测效果.例如,Qin等人<sup>[6]</sup>提出基于双注意力机制模型,通过特征注意力机制来选择最相关的特征数据,利用时间注意力模块获取最相关的隐藏状态信息.Cui等人<sup>[7]</sup>利用自注意力机制来提升模型的短期预测能力.苏雅茜等人<sup>[8]</sup>通过自注意力模块计算不同历史时间点上的数据权重信息来提高传统移动平均算法的预测能力.Li等人<sup>[9]</sup>提出区域注意力机制,通过合并相邻向量的方式来建模不同粒度的依赖.Sehovac等人<sup>[10]</sup>利用注意力机制来提升Seq2Seq模型对电力负荷的预测精度.Yuan等人<sup>[11]</sup>提出时空注意力机制来计算不同输入特征的重要程度与Encoder隐藏层输出中最相关的状态.Huang等人<sup>[12]</sup>设计全局注意力机制与局部注意力机制来处理

动态周期数据中的全局时间依赖与局部时间依赖.Du等人<sup>[13]</sup>结合传统上下文向量与注意力向量设计注意力模块,用于处理数据中长期依赖与隐藏特征.

然而,传统注意力模型计算注意力权重的过程中,每个资源通常是独立看待的,不适用于燃气预测任务——通常燃气负载数据在时间维度上存在着多种时间依赖.直接使用机器翻译领域中的注意力模型可能会导致时间依赖的丢失.虽然,Li等人<sup>[9]</sup>提出的区域注意力机制能够考虑不同的计算粒度,但每次合并的相邻向量的是固定的.不适用于高频率突变的燃气负荷数据.同时,在采用基于注意力机制模型进行时间序列预测时,Decoder是平等地看待每一份输入的历史数据,由于燃气负荷数据在短期内存在的随机突变性,距离预测日期越近的历史燃气负载数据应该提供更多预测依据.

为解决上述提及的问题,本文提出多维注意力机制Seq2Seq模型(multi-dimensional attention Seq2Seq, MA-Seq2Seq),采用LSTM构建Encoder与Decoder.基于区域注意力,设计自适应合并步长计算.提出多层次时间注意力模块,从全局出发,筛选Encoder编码信息中不同维度的时间依赖,生成包含多维时间模式信息的上下文向量.同时,为解决模型无法关注重要历史信息的缺陷,设计局部历史注意力模块,依据短期依赖信息,生成携带关注信息的历史数据,最后Decoder依据多维时间上下文向量与携带专注信息的输入数据进行预测输出.最终,本文使用中国某市区的日燃气消耗数据与2016年电工数学建模竞赛A题的电力负荷数据来验证本文模型的有效性.

## 1 研究方法

### 1.1 符号定义

由于燃气负荷预测本质就是时间序列预测任务,因此,本文中时间序列数据,历史数据,燃气数据等都代指历史燃气负荷数据.设定时间序列数据采集的间隔为每天,则对于 $n$ 天的历史数据表示为 $Y = (y_1, y_2, \dots, y_n)$ ,其中, $y_i \in \mathbb{R}$ .相关特征数据表示为 $X = (x_1, x_2, \dots, x_n)$ ,其中, $X \in \mathbb{R}^{n \times m}$ , $x_t \in \mathbb{R}^{m \times 1}$ , $m$ 表示每次采样时特征数据的个数.Encoder产生的隐藏层输出表示为 $h_i$ , $i \in [1, T]$ , $T$ 表示输出窗口大小.设上一时刻Encoder产生的隐藏状态为 $h_{i-1}$ ,上一时刻Decoder产生的隐藏状态为 $d_{i-1}$ ,长期状态为 $c_{i-1}$ .Encoder与Decoder的隐藏层维度设

置为  $r$ .

### 1.2 注意力机制

注意力机制概念的起源是模拟人类对信息的认知过程——是一个在众多信息中检索重要信息的过程. 2016年, Bahdanau 等人<sup>[5]</sup>首次提出注意力模型来解决 Encoder-Decoder 模型在面对长输入数据时表现性能下降过快的问题——Encoder 会将所有的输入数据都压缩到一个固定长度的上下文向量中, 输入数据过多时就会导致信息丢失. 而注意力模块会选择最为重要的信息来生成上下文向量, 从而尽可能地避免了信息丢失带来的影响. 该模型的成功让后续学者开发出不同版本的注意力模型. 但整体上, 注意力模型是利用对齐函数与分布函数进行数据查询的过程<sup>[14-16]</sup>.

注意力模型中, 通常会涉及 3 个表示向量,  $q, K, V$ , 其中,  $q$  通常代表信息检索的依据, 一般是与模型任务相关的数据向量, 例如, 上一时间步骤中 Decoder 的隐藏状态.  $K$  表示查询内容, 一般代表输入数据, 通常由多个向量组成.  $V$  是与  $K$  相互对应的数据矩阵, 可以是同一份数据. 也可以同一份数据经过不同处理得到  $K$  与  $V$ . 信息检索过程通过对齐函数实现——用于计算  $q$  与  $K$  中每个向量的初始相似程度, 或者能量权重. 将能量权重通过分布函数后便可得到注意力权重, 表示  $K$  中每个向量的重要程度. 分布函数通常用来将能量权重转化成注意力权重, 较为常见的分布函数是 *logistic Sigmoid* 与 *Softmax* 函数, 这两种函数能够确保所得到的注意力权重之和为 1. 这样就可以将注意力权重的含义解释为重要程度的百分占比.

### 1.3 多层次时间注意力

在传统的注意力模型中, 通过对齐函数获取  $q$  与  $K$  之间的初始权重的过程是一个相互独立的过程, 即将  $K$  中的每个向量视作独立的个体与  $q$  进行权重计算, 而在时间序列任务, 采样数据之间是存在着一定的相关性的. 以独立计算方式生成的上下文向量, 可能会导致信息的丢失. 由此, 本文提出多层次注意力模块, 在全局上, 以即联合又独立的方式构建上下文向量. 先用传统的注意力计算方式获取一份上下文向量, 同时, 对同一份  $K, V$  矩阵内的数据进行自适应步长的相邻向量融合处理: 利用算法 1 遍历矩阵, 每次循环计算当前向量与后续向量间的相似度, 将相似度大于设定阈值的相邻向量与当前向量合并, 形成新  $K, V$  矩阵, 随后在新矩阵上进行注意力计算得到另一份上下文向量. 最

后通过一层全连接网络融合两个上下文向量. 模型结构如图 1 所示.

算法 1. 自适应合并粒度算法

```

 $\bar{K}=[]$ 
FOR  $i$  IN  $len(K)$ 
 $S = Similar(k_i, K[i:])$ 
 $k_{sum}=k_i; inner=0$ 
FOR  $j$  IN  $len(S)$ 
IF  $(S[j] \geq \lambda)$ 
 $k_{sum}+=k_{i+j}; inner+=1$ 
ELSE BREAK
END IF
END FOR
 $\bar{K}.append(k_{sum}/(inner+1))$ 
IF  $(inner+1 == len(S))$ 
BREAK
END IF
END FOR
    
```

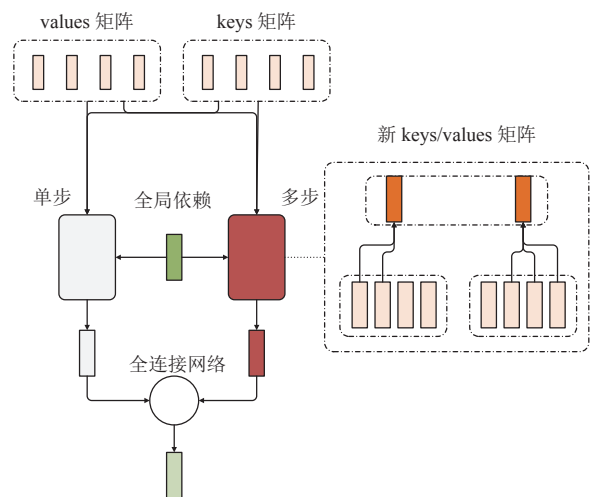


图 1 多层次注意力模块

在构建传统的单步注意力上下文向量的同时, 并行构建多步注意力上下文向量. 图 1 中, values 矩阵与 keys 矩阵通常由 Encoder 产生的隐藏层输出  $h_i$  构成. 为了更好地训练模型, 减少同一份数据的多次使用<sup>[17]</sup>, 对隐藏状态分别进行残差转换与非线性映射. 变换方式如下:

$$k_i = W_k h_i + h_i, 1 \leq i \leq T \quad (1)$$

$$v_i = W_v h_i, 1 \leq i \leq T \quad (2)$$

其中,  $W_k, W_v \in R^{r \times r}$  为可学习参数.  $h_i \in R^{r \times 1}$ .

图 1 中并行执行的两个注意力模块都使用全局依赖作为查询依据, 本文通过上一时间步骤中 Decoder

产生的隐藏状态 $d_{t-1}$ 与长期状态 $c_{t-1}$ 来构建与任务相关的全局依赖. 构建方式如下:

$$q = \frac{W_q(d_{t-1} + c_{t-1})}{2} + \frac{d_{t-1} + c_{t-1}}{2} \quad (3)$$

其中,  $W_q \in R^{r \times r}$ 为可学习参数.  $d_{t-1}, c_{t-1} \in R^{r \times 1}$ .

在单步时间注意力中, 采用传统的注意力计算方式, 独立计算 $q$ 与 $k_i$ 的相似度. 计算过程如下:

$$e_i = qk_i, 1 \leq i \leq T \quad (4)$$

$$a_i = \frac{\exp(e_i)}{\sum_{j=1}^T \exp(e_j)}, 1 \leq i \leq T \quad (5)$$

$$c_1 = \sum_{i=1}^T a_i v_i \quad (6)$$

其中,  $e_i$ 代表初始权重, 即能量权重,  $a_i$ 代表注意力权重. 使用的分布函数为 *Softmax*. 上下文向量 $c_1$ 通过注意力权重与相应的 *value* 值进行加权求和得出.

在计算多步时间注意力模块中, 为让 *values* 矩阵与 *keys* 矩阵中包含不同步长的时间模式信息, 对两个矩阵分别进行自适应步长转化. 如图 1 所示. 对原矩阵中的数据进行遍历, 自适应的选择合并相邻向量的个数. 当某次合并中, 选择了后续所有向量, 则提前结束循环. 自适应选择是通过相似度算法与阈值配合实现, 当相邻向量的相似度超过阈值时, 则认为两者具有同一维度的时间模式需要进行合并. 最终, 新 *key* 值是通过合并后的向量进行平均获得, 对应的 *value* 值则直接使用和向量. 获取新 *keys* 矩阵的算法过程如算法 1.

算法 1 中,  $K$  表示原始的 *keys* 矩阵. *Similar()* 为相似度算法, 用于计算当前 *key* 值与后续 *key* 值是否处于同一维度的时间模式, 本文采用余弦相似度配合 *Softmax* 函数的方式获取相似程度,  $\lambda$  为设定的阈值, 用于控制合并数量. 新 *values* 矩阵的计算过程同 *keys* 矩阵. 随后在新的 *keys* 矩阵与 *values* 矩阵上进行注意力计算, 方式如下:

$$\bar{e}_i = q\bar{k}_i, 1 \leq i \leq \bar{T} \quad (7)$$

$$\bar{a}_i = \frac{\exp(\bar{e}_i)}{\sum_{j=1}^{\bar{T}} \exp(\bar{e}_j)}, 1 \leq i \leq \bar{T} \quad (8)$$

$$c_2 = \sum_{i=1}^{\bar{T}} \bar{a}_i \bar{v}_i \quad (9)$$

其中,  $\bar{T}$  表示新矩阵中向量元素的个数.  $\bar{k}_i, \bar{v}_i$  分别表示新 *keys* 矩阵与新 *values* 矩阵中的元素.  $c_2$  为包含多维

时间模式的上下文向量.

最终合并两个上下文向量 $c_1$ 与 $c_2$ , 确保最终的上下文向量同时包含单步时间信息与多步时间信息, 合并方式如下:

$$c = \tanh(W_1 c_1 + W_2 c_2) \quad (10)$$

其中,  $W_1, W_2 \in R^{r \times r}$ 为可学习参数. 将该上下文向量代替 *Encoder* 原本的隐藏状态向后传递.

在燃气负荷预测中, 受大量外部因素影响, 数据特征存在着复杂的时间模式依赖, 提高了模型预测难度. 通过多层次时间注意力机制来提取不同时间模式特征, 以此优化模型对数据信息的捕捉能力.

#### 1.4 局部历史注意力

在传统的注意力模型当中, 注意力机制模块通常只用来解决在长序列输入中信息丢失的问题, 在时间序列预测任务中, 该部分信息通常为特征数据的信息, 而对应的历史数据 $y = (y_1, y_2, \dots, y_T)$ 是直接输入 *Decoder* 模块, 此时, *Decoder* 在解码预测过程中会平等地看待所有历史数据, 而实际上, 与预测日期越接近的历史数据及与预测日期处于相同位置的数据所提供的信息更具有参考性. 由此引入局部历史注意力模块, 该模块建立在多层次注意力基础之上, 同时考虑近期特征数据与历史数据的情况下生成加权后的历史数据 $\bar{y}$ .

$$e_i = W_q d_{t-1} \tanh(W_k h_{t-1} + W_h [h_i; y_i]), 1 \leq i \leq T \quad (11)$$

$$a_i = \frac{\exp(e_i)}{\sum_{j=1}^T \exp(e_j)}, 1 \leq i \leq T \quad (12)$$

其中,  $W_k \in R^{r \times r}$ ,  $W_h \in R^{r \times (r+1)}$ 为可学习参数,  $[\cdot]$  表示拼接操作.  $h_{t-1}$  为上一时间步骤中 *Encoder* 隐藏状态, 通常为最后一个隐藏层输出或上下文向量. 通过短期特征记忆与当前隐藏层输出和历史数据建立短期变化信息, 在变化信息中查询重要程度较高的历史天数. 最终, 生成携带关注信息的历史数据 $\bar{y} = (a_1 y_1, a_2 y_2, \dots, a_T y_T)$ .

在燃气负荷预测过程中, 历史数据在短期内存在随机突变情况, 这使得一般的注意模型难以在一系列无规律变化中提取有效信息, 同时注意到哪些变化是重要, 哪些变化不那么重要. 通过局部历史注意力模块对数据的处理, 对 *Decoder* 输入窗口中的每份历史负荷数据计算出一个注意力权重, 注意力权重和为 1, 对于预测有重要影响的历史负荷数据将分配更大的权重值, 从而通过 *Decoder* 计算时能够保留更多信息, 使模

型能够察觉哪些变化是重要的, 哪些是可以忽视的. 更好应对数据中的随机事件.

### 1.5 MA-Seq2Seq 模型

为了应对燃气负荷数据中复杂的影响因素, 突变频率高幅度大的特点. 结合多层次时间注意力模块与局部历史注意力模块, 构建 MA-Seq2Seq 模型. 模型结构如图 2 所示.

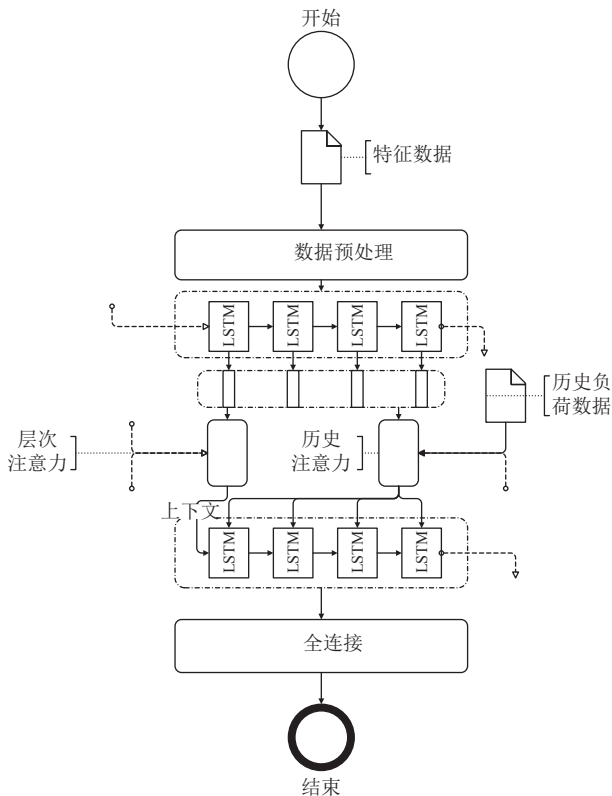


图 2 MA-Seq2Seq 模型

该模型将 Encoder 编码结果通过多层次时间注意力模块生成包含多维时间信息的上下文向量, 同时将该上下文向量作为 Encoder 隐藏状态向后传递. 历史输入数据通过局部历史注意力模块, 生成携带关注信息的历史数据, Decoder 根据两者内容进行解码, 最后, 通过单层全连接网络得到具体的预测结果.

## 2 数据分析

本文使用中国某市区的燃气日使用数据集对模型的有效性进行验证, 数据中包含特征数据有最高温度、最低温度、星期等. 包含 3083 天历史燃气使用数据, 依据二八原则, 将数据的前 80% 作为训练集, 后 20% 作为测试集. 实验数据如图 3 所示.

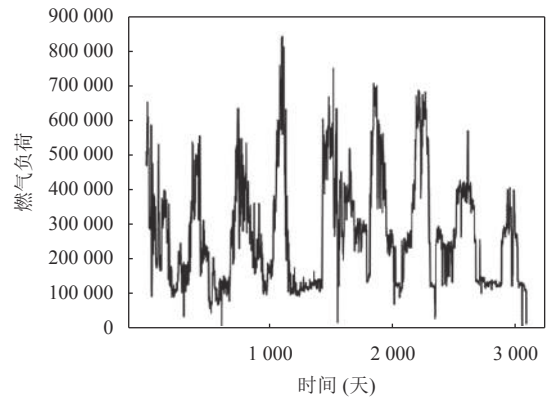


图 3 某市区日燃气消耗数据

### 2.1 训练设置

由于不同数据的取值范围各不相同, 为了让模型更好收敛, 各阶段的输入数据都经过最大最小值法 (Min-Max) 进行处理. 本文模型 MA-Seq2Seq 采用单层 LSTM 构建 Encoder 与 Decoder. 隐藏层的维度设置为 64. 输入窗口  $T$  设置为 14,  $\lambda$  设置为  $len(S)/2$ . 迭代次数为 50 次. 采用 Adam 优化算法. 损失函数采用平均绝对误差 (mean absolute error, MAE). 学习率设置为 0.000 1, 采用 L2 正则化来避免模型过拟合.

### 2.2 评价函数

本文使用平均绝对误差 (mean absolute error, MAE), 平均绝对百分比误差 (mean absolute percentage error, MAPE), 均方根误差 (root mean square error, RMSE) 对模型的测试集数据进行评估, 计算公式定义如下.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (13)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (14)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

其中,  $y_i$  表示真实采样数据, 即一天的燃气消耗.  $\hat{y}_i$  表示模型的预测数据,  $n$  表示测试样本数量.

### 2.3 燃气数据分析

为验证本文模型的 MA-Seq2Seq 的可靠性, 本文使用传统 LSTM 模型, Seq2Seq 模型, 基于注意力的 Seq2Seq 模型进行对比实验. 同时, 为尽可能地控制实验变量, 对比模型参数设置与 MA-Seq2Seq 模型相同.

MA-Seq2Seq 模型在测试集表现如图 4 所示。

从图 4 中可以看出, MA-Seq2Seq 模型在测试集上拥有着不错的表现, 即便在数据突变部分也能有着良好的预测方向. 对比实验结果图如图 5.

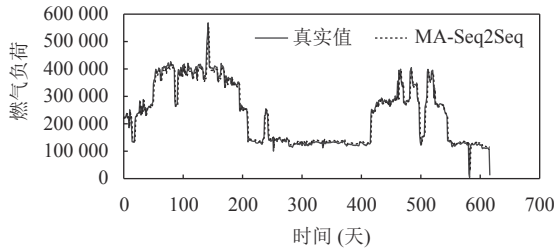


图 4 MA-Seq2Seq 测试结果

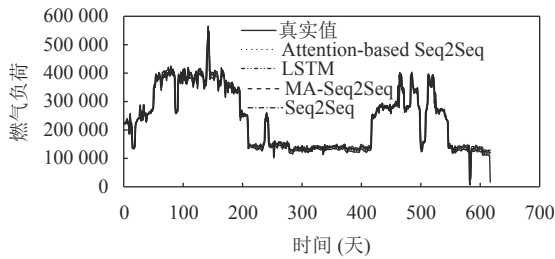


图 5 多模型对比结果

对比实验结果图中可以发现, 在数据缺少较大波动的情况下, MA-Seq2Seq 模型明显比其他模型有更精确的预测结果, 如图 5 中数据段 300-400 之间.

依据表 1, MA-Seq2Seq 模型在 3 个评价指标中都取得最优值, 表明提出的模型相较于常见的燃气预测模型, 能够取得更加精确的预测结果.

表 1 对比实验结果

模型	MAE	MAPE (%)	RMSE
MA-Seq2Seq	<b>12519</b>	<b>8.9</b>	<b>22138</b>
Attention-based Seq2Seq	15160	9.9	23354
Seq2Seq	15411	10.2	23435
LSTM	15192	11.1	23219

### 2.4 可解释性

相对于传统深度学习模型, 注意力模型一大特点是具有可解释性. 就局部历史注意力机制模块而言, 将某份测试数据 (共 14 天输入数据) 的权重信息划分成前 7 天与后 7 天. 由于每次输入两周的数据, 为方便分析, 假定该测试数据预测日期为周一. 结果如表 2 所示.

表 2 前 7 天权重

日期	周一	周二	周三	周四	周五	周六	周日
权重	5.6E-3	3.9E-3	6.4E-4	5.7E-5	1.2E-5	6.5E-6	4.7E-6

从表 2 中可以看出, 对于距离预测日期过于遥远

的数据, 注意力机制会认为与预测日期处于相同位置的历史数据会有更高的参考价值, 当前预测日期为周一, 则同样处于周一位置的数据相较于其他数据而言拥有更高的权重.

表 3 中展现的是与表 2 相反的特性, 由于距离预测日期相对较近, 注意力机制认为前两天的数据会有更高的参考价值, 而是否在历史上处于同样的日期位置则不那么重要.

表 3 后 7 天权重

日期	周一	周二	周三	周四	周五	周六	周日
权重	4.4E-6	5.4E-6	8.4E-6	1.9E-5	1.2E-4	1.1E-2	9.8E-1

结合表 2 和表 3, 总体上注意力机制认为距离预测日期最近的两天历史数据最能够决定预测日期的数据走向. 而这也符合燃气消耗数据的日常规律.

### 2.5 电力数据分析

为验证模型在其他相关数据集上的表现能力, 使用 2016 年电工数学建模竞赛 A 题中电力负荷数据进行实验. 包含的特征数据为最低温度、最高温度、平均温度、相对湿度、降雨量. 总共 1106 份数据, 其中 885 份作为训练数据, 221 份作为测试数据. 同样, 为尽可能控制变量, 所有模型参数与第 2.1 节实验中保持一致. 实验结果如图 6 和图 7.

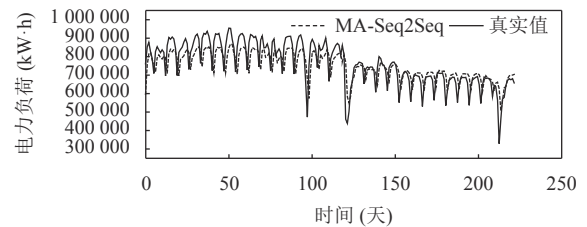


图 6 电力负荷预测结果

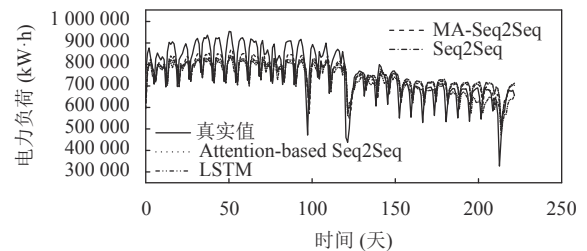


图 7 电力负荷多模型对比结果

从图 6, 图 7 可以直观看出, 相较于其他模型, 本文模型在电力负荷测试集上的表现更加接近负荷数据的真实值. 同时, 表 4 的评价数据显示, 在 MAE, MAPE 及 RMSE 测评上, 本文模型依旧可以取得最优结果.

表4 电力数据测试结果

模型	MAE	MAPE (%)	RMSE
MA-Seq2Seq	<b>59 614</b>	<b>8.3</b>	<b>78 326</b>
Attention-based Seq2Seq	65 462	9.1	83 183
Seq2Seq	64 090	9.0	81 754
LSTM	66 490	9.0	81 476

表5,表6为某份测试数据的历史注意力权重信息,总体上,权重分布与燃气数据实验相似,注意力模块认为与预测日期较近的数据应该占据更多权重.表明本文模型在电力负荷数据集上依旧具备着可行性.

表5 前7天电力数据权重

日期	周一	周二	周三	周四	周五	周六	周日
权重	1.9E-4	2.3E-5	2.7E-6	3.2E-7	1.5E-7	1.1E-7	5.1E-7

表6 后7天电力数据权重

日期	周一	周二	周三	周四	周五	周六	周日
权重	2.9E-6	2.3E-5	2.0E-4	2.4E-3	2.6E-2	1.5E-1	8.2E-1

### 3 结论

本文提出一种多维注意力 Seq2Seq 预测模型,研究设计了多层次时间注意力模块,对 Encoder 隐藏层输出进行多时间步长对齐,综合单时间步长、多时间步长计算,生成包含多维时间模式信息的上下文向量.为 Decoder 解码提供更多支持.设计增加了局部历史注意力模块,结合短期特征编码信息与输入的历史数据,生成携带关注信息的历史数据,以便 Decoder 在预测过程中能够关注更具有参考价值的输入数据.通过两个注意力模块的处理,能够让 Decoder 更具针对性的生成预测内容,实验结果表明 MA-Seq2Seq 模型整体有着良好的预测结果,面对局部突变数据也能有不错的表现结果.同时,为验证模型在相似数据集上的可行性,使用公开电力负载数据进行实验.测试结果显示 MA-Seq2Seq 模型依旧可以取得较高的预测精度.

#### 参考文献

- 1 申洪亮,杨万莉,张强. 2022年中国天然气市场供需预测及未来发展趋势分析. 国际石油经济, 2022, 30(8): 41-50, 100. [doi: 10.3969/j.issn.1004-7298.2022.08.005]
- 2 Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature, 1986, 323(6088): 533-536. [doi: 10.1038/323533a0]
- 3 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9(8): 1735-1780. [doi: 10.1162/neco.1997.9.8.1735]
- 4 Cho K, van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. Proceedings of the 2014

- Conference on Empirical Methods in Natural Language Processing. Doha: Association for Computational Linguistics, 2014. 1724-1734.
- 5 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. Proceedings of the 3rd International Conference on Learning Representations. San Diego, 2015.
- 6 Qin Y, Song DJ, Cheng HF, et al. A dual-stage attention-based recurrent neural network for time series prediction. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: AAAI Press, 2017. 2627-2633.
- 7 Cui YC, Yin B, Li RX, et al. Short-time series load forecasting by Seq2Seq-LSTM model. Proceedings of the 9th IEEE Joint International Information Technology and Artificial Intelligence Conference (ITAIC). Chongqing: IEEE, 2020. 517-521.
- 8 苏雅茜,崔超然,曲浩. 基于自注意力移动平均线的时间序列预测. 南京大学学报(自然科学), 2022, 58(4): 649-657.
- 9 Li Y, Kaiser L, Bengio S, et al. Area attention. Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019. 3846-3855.
- 10 Sehovac L, Grolinger K. Deep learning for load forecasting: Sequence to sequence recurrent neural networks with attention. IEEE Access, 2020, 8: 36411-36426. [doi: 10.1109/ACCESS.2020.2975738]
- 11 Yuan XF, Li L, Shardt YAW, et al. Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. IEEE Transactions on Industrial Electronics, 2021, 68(5): 4404-4414. [doi: 10.1109/TIE.2020.2984443]
- 12 Huang ST, Wang DL, Wu XH, et al. DSANet: Dual self-attention network for multivariate time series forecasting. Proceedings of the 28th ACM International Conference on Information and Knowledge Management. Beijing: Association for Computing Machinery, 2019. 2129-2132.
- 13 Du SD, Li TR, Yang Y, et al. Multivariate time series forecasting via attention-based encoder-decoder framework. Neurocomputing, 2020, 388: 269-279. [doi: 10.1016/j.neucom.2019.12.118]
- 14 Chaudhari S, Mithal V, Polatkan G, et al. An attentive survey of attention models. ACM Transactions on Intelligent Systems and Technology, 2021, 12(5): 53.
- 15 Galassi A, Lippi M, Torrioni P. Attention in natural language processing. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(10): 4291-4308. [doi: 10.1109/TNNLS.2020.3019893]
- 16 Niu ZY, Zhong GQ, Yu H. A review on the attention mechanism of deep learning. Neurocomputing, 2021, 452: 48-62. [doi: 10.1016/j.neucom.2021.03.091]
- 17 Daniluk M, Rocktäschel T, Welbl J, et al. Frustratingly short attention spans in neural language modeling. Proceedings of the 5th International Conference on Learning Representations. Toulon: OpenReview.net, 2017.

(校对责编:孙君艳)