

量子自注意力神经网络的时间序列预测



陈欣, 李闯*, 金凡

(中金金融认证中心有限公司, 北京 100176)

摘要 在“量子-经典”混合模式下, 设计了多头量子自注意力神经网络预测模型 (MQSAPN) 用以进行时间序列预测, 模型包括多头量子自注意力模块以及变分量子线路预测模块两部分。通过对输入数据按时间步长分别进行量子态编码以及 K 、 Q 、 V 的计算, 借鉴已有研究使用高斯函数进行自注意力系数的估计方式, 将量子自注意力特征提取后的数据再次编码到变分预测线路中, 经过线路演化及测量, 最终获取预测结果。完整流程与模型搭建均采用 VQNet 框架实现。在天气学变量的时间序列预测任务中, 该模型表现出与经典多头自注意力模型预测模型以及长短期记忆单元网络模型相当的预测精度。此外, 相对于同样是量子机器学习的 data-reuploading 变分线路而言, 在近乎同等规模线路深度与参数量的前提下, 表现出更高的预测精度, 这也进一步验证了引入量子自注意力机制的有效性。值得指出的是, 作为预测部分的变分线路会随着输入数据量的增多 (如时间窗加长、特征变量规模增加等), 其参数量与线路深度也会显著增加, 尽管多层 QSA 能够较好地进行特征表达, 但依然有可能因遇到“贫瘠高原”困难而成为整个网络的瓶颈。

关键词 量子计算; 量子机器学习; 自注意力机制; 时间序列预测

中图分类号 TP18; P45 文献标志码 A DOI 10.12178/1001-0548.2022340

Research on Time Series Prediction via Quantum Self-Attention Neural Networks

CHEN Xin, LI Chuang*, and JIN Fan

(China Financial Certification Authority, Beijing 100176, China)

Abstract A Multi-head Quantum Self-Attention Predict Network (MQSAPN) is designed in hybrid manner, which could be used in time-series forecasting. MQSAPN comprises two components, one is the Multi-head Quantum Self-Attention (MQSA) model, and the other is the predicting Variational Quantum Circuits (pVQC). When fed with sequential inputs, the MQSA firstly computes the key, query, and value vectors corresponding to all time steps through the variational circuits, and then according to exist studies, the attention is estimated via Gaussian function. With residual link on input and multi-head features, the output of MQSA were pushed to pVQC part, which was encoded into quantum circuit again, and the prediction would be ultimately calculated out by measurements on observables. The prediction results of MQSAPN numerical experiments on atmospheric variables indicate the effectiveness of quantum self-attention, by comparison with the results of a data-reuploading VQC model with almost same amount of parameters. The accuracy of predicting is close to classic multi-head transformer model and LSTM net. To be noted, as input time window extends or the more features are adopted, the number of parameters of pVQC will also increases correspondingly, which makes the pVQC part become the bottleneck of the whole model due to ‘barren plateau’ problems during training process.

Key words quantum computing; quantum machine learning; self-attention; time series prediction

时间序列预测作为序列建模的重要方面, 广泛存在于金融、能源、交通、气象、公共卫生等领域。并且随着机器学习/深度学习方法的迅猛发展, 序列建模在高可用 (high available) 场景中 (如在线机器翻译、实时语音识别、自然语言处理、计算机视觉等领域) 已得到了较广泛的应用。另一方面, 随着量子计算领域近来在“中等含噪

量子计算” (Noisy Intermediate-Scale Quantum, NISQ) 取得了突破性进展^[1-2], 使用几十至数百量子比特 (qubit) 运行特定类型的量子算法成为可能, 如变分量子本征求解器算法 (Variational Quantum Eigensolver, VQE)^[3]、量子近似优化算法 (Quantum Approximate Optimization Algorithm, QAOA)^[4] 以及量子机器学习 (Quantum Machine

收稿日期: 2022-09-30; 修回日期: 2022-11-30

作者简介: 陈欣, 博士, 高级工程师, 主要从事机器学习与量子计算等方面的研究。

*通信作者 E-mail: lichuang@cfca.com.cn

Learning, QML) [5] 等通过“量子-经典”混合方式实现的算法。这些算法一方面利用量子叠加 (superposition) 与纠缠 (entanglement) 的特性可以实现高并行的处理速度 (个别算法可达到指数级别加速), 另一方面通过量子编码实现数据的并行存储。文献 [6] 认为 QML 通过将原始特征映射到希尔伯特空间, 能够实现从经典方法难以采样的复杂分布中稳定采样, 因而其具有经典方法不具备的优势 [7]。此外, 文献 [8] 从核方法 (kernel method) 的角度, 通过对比预测误差界限 (error-bounding), 认为在量子比特数较少时, 量子机器学习与经典机器学习算法的差异主要取决于训练数据的特性, 而随着构建模型的量子比特数量逐渐增多, 量子算法的优势才会更容易显现。

从序列预测的一般方法而言, 通过选取特征变量的历史时期 (早于预测时间点) 数据作为输入, 通过建立模型, 实现对未来时间点的预测。较早时期的统计建模方法, 如自回归滑动平均模型 (Autoregressive Integrated Moving Average Model, ARIMA) [9], 通过提取差分信息建立线性模型可以实现对平稳序列的预测。考虑到基于统计方法的预测模型对数据自身的统计特性有较为严格的要求 (如平稳性、各态历经性), 并且模型对于多变量间的非线性拟合能力有限。因此, 能够保留较长时期的数据信号并对变量间非线性关系具有较好拟合能力的深度学习方法, 成为近年来的主流方法。递归神经网络 (RNN) [10] 在时序建模方面表现出色, 但由于其存在梯度消失和爆炸的问题 [11], 使其在使用较长历史数据进行训练时遇到困难。进一步地, 长短期记忆模型 (Long Short-Term Memory, LSTM) [12] 以及门控单元网络 (Gated Recurrent Unit, GRU) 通过实现网络内部结构的“门控”机制 [13], 将输入数据进行多次变换、连接与状态再输出, 从而实现较长时期输入信号更好地“捕捉”并且较大幅度缓解梯度消失与爆炸的问题 [14]。而现有的 RNN 网络架构存在对全局长时间依赖 (long term dependency) 信息提取不足及难以并行的缺点, 新提出的 Transformer 框架依据自注意力机制 (Self-Attention, SA) [15], 则可以较好地解决这两方面困难。该框架在自然语言处理方面表现突出 [16], 同时也在计算机视觉方面 (如视觉分类、目标检测等) 取得了较好的效果 [17-18], 这也进一步展现了自注意力机制具有更强的可拓展性以及整合输入数据点之间的相关性并增强数据表达的能力。

另一方面, 在量子算法领域, 近期已有越来越多的适用于 NISQ 量子硬件特性的量子神经网络方法 (Quantum Neural Network, QNN) [19] 被提出, 如量子卷积神经网络 [20]、量子自然语言处理方法 (Quantum NLP, QNLP) [21-22] 等, 而基于变量子线路进行时间序列预测的研究也开展得较早 [23-24]。在金融领域, 量子时间序列预测方法也应用在对股票等资产价格的预测方面 [25]。文献 [26] 采用量子线路进一步“增强”经典深度学习的方法, 通过将输入的文本嵌入信息转化到描述量子态的希尔伯特空间中, 量子部分的计算可以在 NISQ 硬件上真实运行, 同时该研究也认为现有的量子-经典混合架构对于构建复杂的序列处理模型 (如 LSTM, Transformer) 来处理 NLP 领域的问题是可行的, 并且该研究也初步实现了量子-经典混合的 Transformer 架构。类似地, 文献 [27] 提出并实现了 Quantum LSTM (QLSTM) 的量子经典混合架构, 其主要思想是在基本 LSTM 单元 (LSTM Cell) 中引入变量子线路 (Variational Quantum Circuit, VQC) 对输入数据的中间结果进行多次转换, 即将量子线路演化的测量结果作为门 (记忆门、输出门、遗忘门) 间的交换数据。该模型在预测动力系统时间序列任务中可以达到与经典 LSTM 精度相当且收敛更快的特点。进一步地, 文献 [28] 提出了一种线路和结构更加简化的量子自注意力神经网络 (Quantum Self-Attention Neural Network, QSANN), 该研究的优势在于计算自注意力系数 (或相似度系数) 时, 通过量子线路测量得到 K 与 Q 的数值后输入高斯函数直接计算结果, 而不是按照一般自注意力机制的内积运算。从而在文本分类任务中, QSANN 利用高斯函数可以有效提取词向量间的潜在联系, 还可以显著提升训练效率, 其分类精度与经典自注意力网络相当, 相较于其他 QNLP 模型具有显著提升。

鉴于 QSANN [28] 在文本分类方面所表现出的优越性能, 本研究主要针对 QSANN 的自注意力机制编码部分进行多头扩展, 并通过构建变分线路实现多变量的时间序列预测模型, 采取“单步”预测方案, 在典型天气学要素变量数据上进行验证和测试。

1 自注意力机制与模型搭建

1.1 自注意力机制

自注意力机制 (Self-Attention, SA) [15] 其实现机制完全不同于以往的 RNN 或 CNN, 其最初是用来进行序列信息提取, 由于其实现简单且易于并行

的特性, 使得其在 NLP 及 CV 领域得到广泛应用。简而言之, 自注意力机制的基本思想是通过提取序列数据内部各个时间点特征向量间的相关性, 即通过 Query 与 Key 计算相关性 α 值, 由 α 值与 Value 计算得到自注意力输出 (如图 1 所示), 最终通过残差连接并重复进行多层计算, 达到逐渐增强输入数据信息表达的目的, 在支持分布式训练的条件下, 可以有效提升模型在下游任务 (如分类、预测、检测等) 中的准确性。

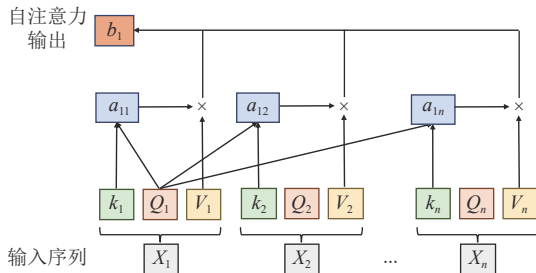


图 1 经典自注意力机制计算流程

1.2 量子自注意力机制

量子自注意力机制相比于经典自注意力机制中计算 K 、 Q 、 V 矩阵形式变换, 在采用量子线路实现过程中, 将可训练权重转换为变分量子线路中的含参量子门参数, 再通过 K 与 Q 值构建序列输入数据间的注意力矩阵 (或相似度矩阵), 接下来与经典 SA 相同, 对相似度矩阵进行归一化 (一般采用 softmax 方式) 后对各个时间点的数据特征进行加权平均, 得到自注意力的结果输出, 最终通过残差连接获取单层自注意力网络层的输出。文献 [28] 采用“量子-经典”混合的方式, 将输入单时间步长的特征向量进行编码 (角度编码) 后再经过含参量子线路的演化, 通过对量子线路的可测量变量进行测量, 即可完成 K 、 Q 、 V 的计算。必须指出的是, 如果将量子态的内积作为量子自注意力机制中的相似度, 则需要 $O(3N \times T)$ 数量的量子比特且线路复杂度极高。作为适应当前 NISQ 硬件的方法, 文献 [28] 采用高斯核函数计算自注意力系数。考虑到将高斯函数作为核函数, 其自身具有拟合高维度数据的能力, 并且相对于计算 Key 与 Query 的内积, 使用高斯函数可以直接简化量子线路的演化流程。

2.3 模型搭建

本研究提出的多头量子自注意力预测网络 (Multi-Head Quantum Self-Attention Prediction Network, MQSAPN) 包含两个主要部分, 即多头量子自注意力模块与变分量子线路预测模块, 其中多头计算

部分主要是在模型构建时对基本 QSA 网络层输出进行多次联结, 如图 2 所示。一般认为多头注意力机制可以从不同层面对输入信息潜在的注意力关联性进行提取, 从而增强信息的表达能力^[29]。尤其在 NLP 领域, 使用多头自注意力机制在解释嵌入词向量在输入句子中所代表的词性、指代、语义关联等方面有较好的表现^[30]。

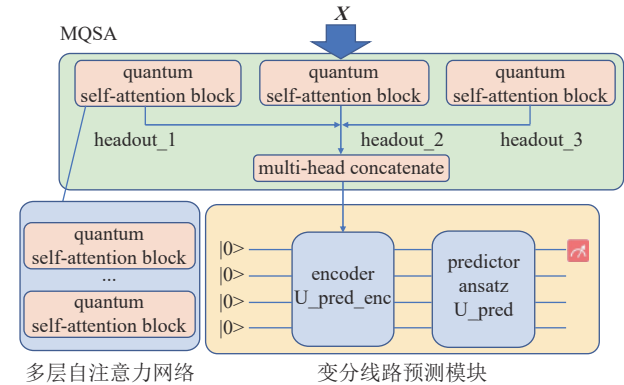


图 2 多头量子自注意力网络架构

具体来说, QSA 网络层按以下方式构建。以具有两个特征变量时间长度为 n 的序列输入为例。计算 Q 与 K 的线路, 如图 3 所示, 线路包括量子编码和变分参数线路两部分。输入数据 X 为:

$$X = [x_0, x_1, \dots, x_{n-1}] \quad x_i \in \mathbb{R}^2 \quad (1)$$

经过编码线路演化得到编码量子态 $|\psi_0\rangle$, 即:

$$|\psi_0\rangle = U_{\text{enc}}(x_i)|0^2\rangle \quad (2)$$

编码后的量子态分别经过对应参数线路演化, 最终按照 Pauli-Z 门进行 Z 方向的测量获得测量期望:

$$\langle Z_K \rangle = \langle \psi_0 | U_K^\dagger(\theta_K) Z_0 U_K(\theta_K) | \psi_0 \rangle \quad (3)$$

$$\langle Z_Q \rangle = \langle \psi_0 | U_Q^\dagger(\theta_Q) Z_0 U_Q(\theta_Q) | \psi_0 \rangle \quad (4)$$

假设 V 为包含两个特征变量的情形, 计算 V :

$$\langle Z_{V_0} \rangle = \langle \psi_0 | U_V^\dagger(\theta_V) Z_0 U_V(\theta_V) | \psi_0 \rangle \quad (5)$$

$$\langle Z_{V_1} \rangle = \langle \psi_0 | U_V^\dagger(\theta_V) Z_1 U_V(\theta_V) | \psi_0 \rangle \quad (6)$$

在 i 时刻, 计算得到 KQV:

$$K_i = \langle Z_K \rangle_i \quad (7)$$

$$Q_i = \langle Z_Q \rangle_i \quad (8)$$

$$V_i = [\langle Z_{V_0} \rangle, \langle Z_{V_1} \rangle]_i \quad (9)$$

按照文献 [28] 的处理, 自注意力系数则采用高斯函数进行计算:

$$\alpha_{i,j} = e^{-(Q_i - K_j)^2} \quad (10)$$

$$\alpha'_{i,j} = \frac{\alpha_{i,j}}{\sum_{j=0}^{n-1} \alpha_{i,j}} \quad (11)$$

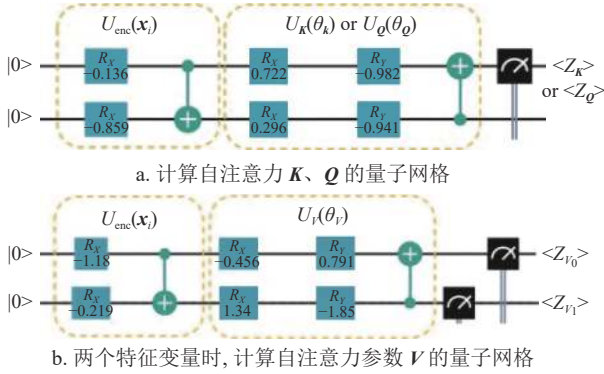


图 3 计算 K 、 Q 以及 V 的量子线路

按式 (11) 对自注意力系数进行归一化后, 通过残差连接计算得到量子自注意力层的输出, 得到

“单头” (single head) 注意力计算输出 ($head_{out}$), 本研究将量子自注意力层进行“多头” (multi-head) 拼接, 如图 4 所示, 作为变分线路预测模块的输入特征:

$$ch_{out} = \begin{bmatrix} head_{out_1} \\ head_{out_2} \\ head_{out_3} \end{bmatrix} \quad (12)$$

MQSAPN 的变分线路: 1) 将 MQSA 输出拼接后的多头特征通过编码线路 (U_{pred_enc}) 进行量子编码, 如式 (13) 所示; 2) 编码量子态经过参数线路 U_{pred} 演化并通过测量哈密顿量的期望得到对应输入序列在下一时刻的预测值, 如式 (14) 所示:

$$|\psi_s\rangle = U_{pred_enc}(ch_{out})|0^m\rangle \quad (13)$$

$$\langle Z_{pred} \rangle = \langle \psi_s | U_{pred}^\dagger(\theta_{pred}) Z_0 U_{pred}(\theta_{pred}) | \psi_s \rangle \quad (14)$$

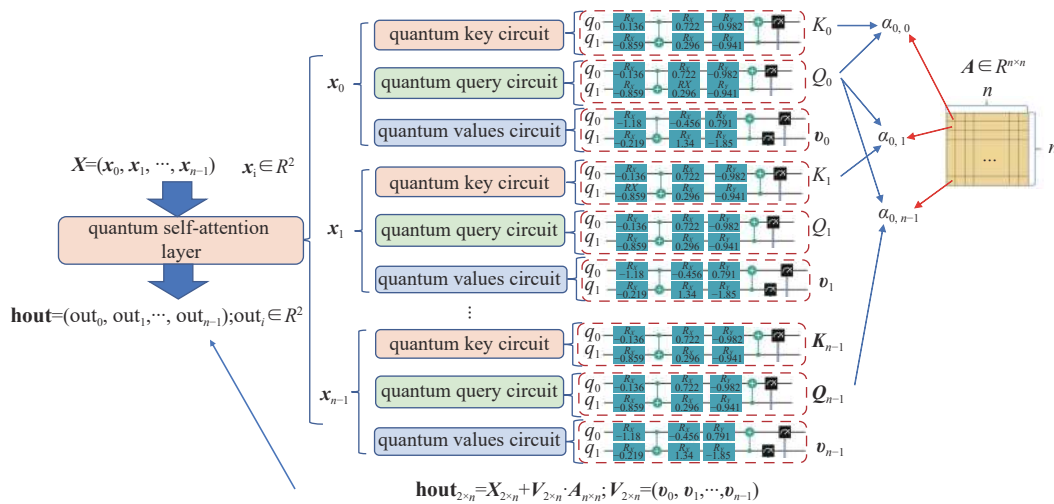
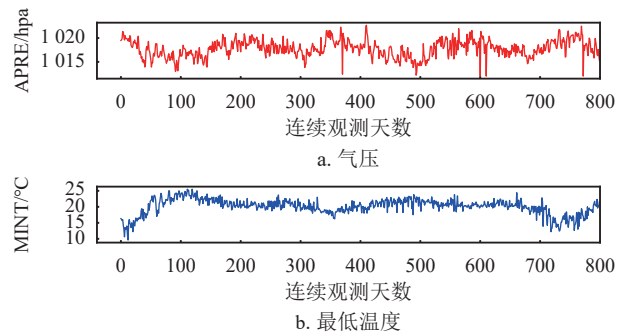


图 4 single-head 量子自注意力网络层

2 数据与实验

针对典型天气学变量数据实施预测实验, 数据来源于气象站连续 1 000 天的真实观测结果, 考虑到现有基于观测数据的统计建模预报的手段已成为天气要素预报的重要途径, 并且基于“数据驱动 (data driven)”的天气要素预报, 已成为独立于动力学模式的另一重要手段^[31-32]。从数据特性而言, 天气学变量数据 (气温、湿度、气压、风速) 间存在高度非线性关联, 因而, 天气学变量预测任务可以用来验证 QSA 是否具有抽取变量间非线性关系的能力。该数据集共包含 5 个天气学变量, 即气压 (Air Pressure, APRE)、最低温度 (Minimum

Temperature, MINT)、最高温度 (Maximum Temperature, MAXT)、相对湿度 (Relative Humidity, RH)、风速 (Wind Speed, WS)。训练数据 800 天, 测试数据 200 天, 如图 5 所示。



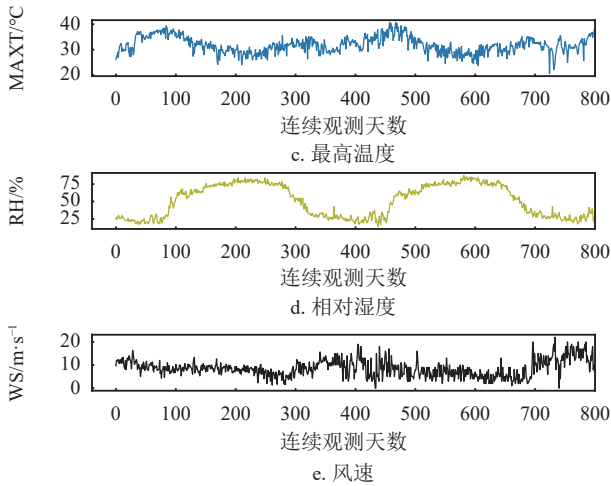


图 5 原始数据（训练时段）变化序列

2.1 数据预处理与评估

首先针对原始数据进行“离群值”检测和处理。设定 15 天滑动时间窗口，以正负 3 倍标准差作为检测条件，对检测出的离群值与窗口平均值再进行加权平均后插补回训练数据集。此外，对处理离群值后的数据集进行归一化，使得变量的变化范围在 [0, 1] 区间内，不再具有物理量纲。同样在评估阶段，对测试集数据进行归一化后输入模型进行预测，并对预测值进行反归一化再与测试集原始数

据进行评估。归一化采用“最大-最小值”方式，设定上（下）限按照训练集上各变量最大（最小）值的 5% 增加（减少）。预测结果评估以模型精度（accuracy）作为指标：

$$acc = 1 - \sqrt{\frac{1}{n} \sum_{i=1}^n E_i^2} \quad (15)$$

$$E_i = \frac{|\text{actual} - \text{predict}|}{\text{actual}} \quad (16)$$

2.2 特征选择

考虑到数据集中 5 个特征变量所刻画物理特性以及实际的数据表现，本研究对训练数据各变量进行相关分析。考虑到不同物理属性天气学变量与给定物理变量信号间的相关性差异，将待预报的 5 个变量分别按动力学与热力学属性进行分类，一般同类型物理变量的信号特征更具一致性，对应较为一致的天气学演变过程。此外，数据相关性在不同的 100 天时间内表现出较为相似的特征，如图 6 所示，也进一步验证了依据物理特征进行分类的合理性。因此，针对各个变量的预测任务选择相同物理特性的变量作为特征输入也更为合适。本研究中各变量的预测特征选择如表 1 所示。

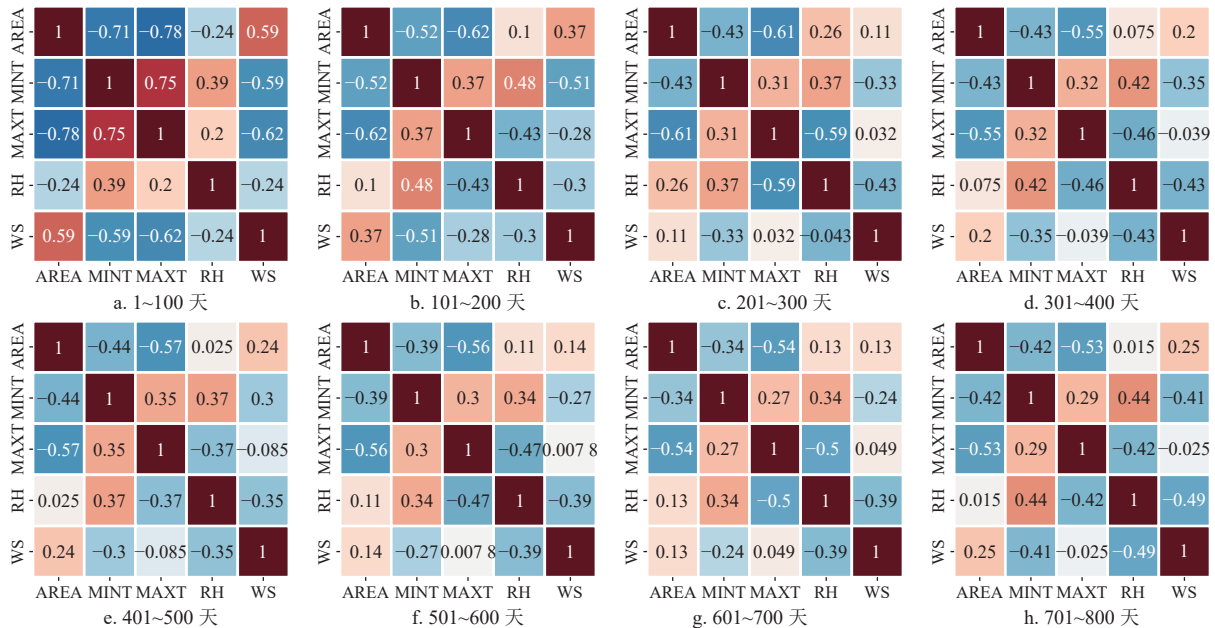


图 6 训练数据集上 5 个变量间连续 800 天的相关系数热图

表 1 目标变量与特征变量

目标变量	特征变量
APRE	APRE, WS
MINT	MINT, MAXT
MAXT	MAXT, MINT
RH	RH, MAXT
WS	WS, APRE

2.3 预测实验

采用本源量子 VQNET 量子机器学习框架^[33]实现 MQSAPN 算法流程并在该框架内的初始化的量子模拟器中完成训练。量子自注意力网络训练实验设计如表 2 所示，同时，本研究采用经典多头自注意力网络（Classic Multi-Head Self-Attention,

CMHSA) 以及长短期记忆网络 (LSTM) 设计了相同的训练方案 (精度结果见表 3)。预测方案如图 7, 将选为特征变量序列的前 n 天 ($n=4$ 或 8) 数据作为输入特征向量。每一时刻的输入向量按照前面计算 K 、 Q 、 V 的量子线路进行量子态编码和演化。

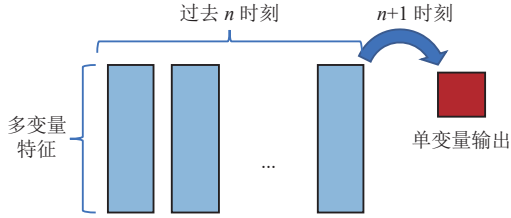


图 7 时间序列预测方案

量子自注意力网络中 QSA 层数为 L , 多头计算数量为 H , 则模型中可训练参数与所有线路中的含参量子门一一对应, 如式 (17) 所示。总数据量 N_t , 取决于输入数据时间窗长度, 如 $n=4$ 时, 总时间长度为 200 的数据, 可训练数据量为 $N_t=196$; $n=8$ 时, $N_t=192$ 。训练方案中使用 Adam 优化器用以更新模型参数, 学习率设置为 $0.02 \sim 0.075$, 损失函数采用均方根误差 (Mean Squared Error, MSE) 式 (19), 预测值 y_p 则为测量量子变分线路的哈密顿量测量值 (式 (14)、式 (20))。

$$\Theta = \{^{(l)}\theta_K^{(h)}, ^{(l)}\theta_Q^{(h)}, ^{(l)}\theta_V^{(h)}, \theta_{\text{pred}}\} \quad (17)$$

$$1 \leq l \leq L \quad 1 \leq h \leq H \quad (18)$$

$$L(\Theta, D) = \frac{1}{N_t} \sum_{d=1}^{N_t} ({}^{(d)}y_p - {}^{(d)}y_t)^2 \quad (19)$$

$${}^{(d)}y_p = \langle Z_{\text{pred}} \rangle \quad (20)$$

此外, 预测变分线路模块中通过角度编码将多头计算的结果编码到量子态后, 考虑到 Hardware Efficient Ansatz (HEA) 线路模块易于在硬件实现^[34] (如 NISQ 硬件), 本研究采用 HEA 作为基本模块单元, 采用 4 个量子比特、以输入时间窗 4 为例, 变分线路如图 8 所示, 共包含两个基本单元 (depth=2)。

为验证模型搭建过程中输入数据时间长度、量子自注意力网络层数及是否选择采用多头方式对最终预测结果的影响, 共设计 6 组实验以及 1 组 Data-ReupLoading (DRL) 实验 (见表 2)。在同样数据集设定下, 首先, 采用经典 Transformer 的自注意力编码模型进行训练和预测评估, 输入数据时间窗长度分别为 4 天和 8 天, 为了评估量子自注意力机制的引入对预测效果产生的影响, 本研究利用量子振幅编码^[35] 将时间窗口长度为 8 天输入数据编码到量子态, 采用 DRL 方式构造的变分量子线路^[36], 其基本线路模块由 DRL 线路层构成, 如图 8 所示, DRL 单层内部的 HEA 结构采用与 MQSAPN 中的预测变分线路相同的基本模块 (图 7)。DRL 结构的优势在于随着量子线路的加深, DRL 层除接受上一层演化后的量子态之外, 输入数据也会被再次编码进入线路, 进而保持输入信息的“新鲜性”, 可有效缓解量子变分线路优化过程中出现的“贫瘠高原”问题^[37]。

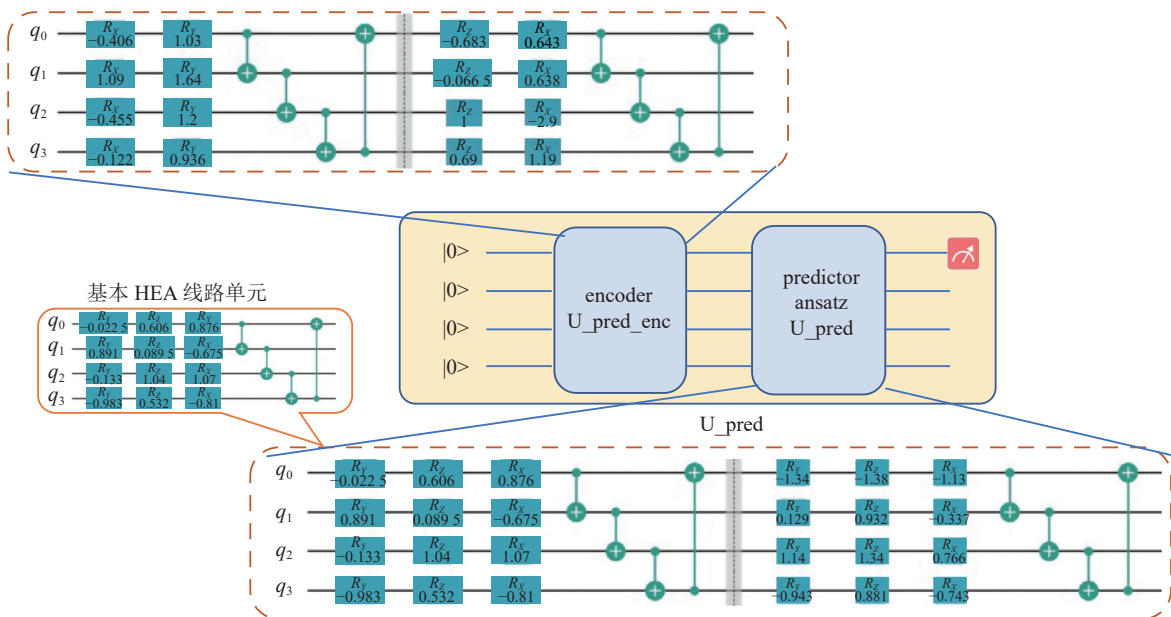


图 8 采用 4 量子比特构建预测模块 (角度编码线路与变分预测线路)

表 2 时间序列预测实验设计方案

训练方案	SH/MH	QSA层数	线路门数	参数门数	最大深度	时间窗/天
t4-layer3-sh	Single-Head	3	81	52	14	4
t4-layer6-sh	Single-Head	6	126	88	14	4
t8-layer3-sh	Single-Head	3	137	96	26	8
t4-layer1-mh	Multi-Head	1	108	48	14	4
t4-layer3-mh	Multi-Head	3	204	96	20	4
t8-layer3-mh	Multi-Head	3	168	96	30	8
DRL-8	-	-	130	48	40	8

从实验结果来看(表 3), 采用 MQSAPN 预测精度与经典多头注意力模型以及 LSTM 网络的预测精度在各个预测变量的表现相当。与文献 [28] 中使用 QSANN 在文本分类任务中与经典 Transformer 进行比较, 结论相似。并且相比于直接采用 DRL 变分子量子线路(如图 9 所示)的预测精度而言, MQSAPN 各个变量(除气压外)在测试集上精度均有较为明显的提升, 这直接表明了 QML 中引入自注意力机制对于时间序列预测问题的有效性, 也验证了量子自注意力机制自身在数据信息提取上的优势。方案(1)(2)与(4)(5)的预测结果表明, 使用更多的 QSA 神经网络层对预测结果精度的提升非常有限。另外, 本研究采用“多头”计算的方式对量子自注意力机制进行扩展, 尽管在 NLP 领域, 采用“多头”方式能够以多维度切入的方式, 提升输入数据的信息表达能力, 然而本研究的预测实验结果中, 时间窗为 4 的方案(5)相比方案(1), 在最低温度与风速的预测效果有一定的提升, 但“多头”方式在时间窗

为 8 的实验结果中, 即方案(6)相对于方案(3), 并没有展现出明显的优势。这可能与本研究中预测特征数目相对较少有一定关系, 并且单一时间点数据的嵌入维数较低, 多维度数据特征之间可被提取的数据关联非常有限, 因而并未完全体现出多头计算的显著优势。

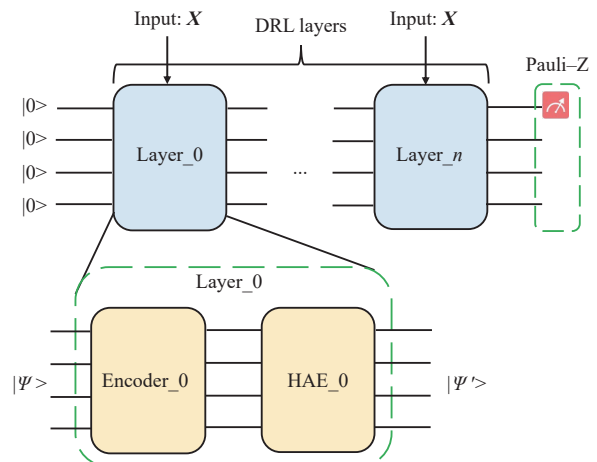


图 9 Data-reuploading 变分子量子线路

表 3 测试集预测精度

序号	方案	气压	最低温度	最高温度	相对湿度	风速
(1)	t4-layer3-sh	0.998 9	0.936 5	0.931 4	0.913 9	0.515 7
(2)	t4-layer6-sh	0.998 9	0.938 1	0.930 3	0.930 9	0.532 6
(3)	t8-layer3-sh	0.998 8	0.941 3	0.932 9	0.931 1	0.548 6
(4)	t4-layer1-mh	0.998 8	0.940 3	0.926 7	0.927 8	0.504 1
(5)	t4-layer3-mh	0.998 8	0.941 0	0.930 2	0.923 7	0.554 0
(6)	t8-layer3-mh	0.998 8	0.943 1	0.931 2	0.917 2	0.511 3
(7)	DRL-8	0.998 7	0.928 8	0.926 0	0.922 7	0.426 7
(8)	CMHSA-t8	0.998 7	0.940 1	0.932 6	0.923 6	0.553 1
(9)	CMHSA-t4	0.998 7	0.942 1	0.931 1	0.925 4	0.543 4
(10)	LSTM-t4	0.998 8	0.940 2	0.928 8	0.933 0	0.445 3
(11)	LSTM-t8	0.998 9	0.940 3	0.930 8	0.928 0	0.512 2

此外, 就时间窗长度的影响而言, 尽管对于一般 NLP 问题, 虽然经典 Transformer 模型可通过位

置编码(position encoding)的方式实现对全局位置信息的捕获, 并且全局信息往往对单时间点词向

量在下游任务上的表现起较重要的作用, 但文献 [16] 提出类似于卷积神经网络的局部信息注意力机制用以提取局部信息, 同时考虑到 Transformer 本身存在的信息存储瓶颈, 该研究认为 Transformer 能够通过局部信息的“增强”来提升模型效果。本研究中, 设定输入特征变量时间窗长度分别为 4 天和 8 天, 比较方案 (1) (3) 及方案 (4) (6) 的预测结果, 时间窗长度为 8 天的预测精度更高, 一方面, 扩充更多时间点的信息输入增加了总体输入信息对预测值的相关程度, 提高了可预测性; 另一方面, 在实际操作中, 输入信息的扩充使得可训练的变分线路参量也相应增多, 进而模型的拟合能力也会得到相应的提升。但必须要指出的是, 在方案 (6) 中, 预测变分线路的深度以及参数量大幅增加, 从结果上看, 相对湿度变量的预测精度甚至要低于 DRL 线路的结果, 而不同于 DRL 的是, 在 MQSAPN 的预测模块中变分线路并未引入 re-uploading 机制, 因此推测因为“贫瘠高原”的困难一定程度上造成了模型收敛效果不佳。

3 结束语

本研究在“量子-经典”混合的模式下, 设计了多头量子自注意力神经网络预测模型, 主要包括多头量子自注意力模块以及变分量量子线路预测模块两部分。完整流程与模型搭建均采用本源量子-VQNet 框架实现。本研究可以确定在序列预测任务上引入 QSA 具有一定优势。其次, 进行“多头”扩展的 QSA 神经网络模型在预测精度上具有一定的提升但并不显著, 可能与本研究构建的特征输入维度较低有关, 推测在更高嵌入维度下会有较好表现。另外, 选择较长的时间窗作为输入有更高精度的表现。最后, 作为预测功能的变分线路模块依然可能因为“贫瘠高原”的问题困扰而影响模型的最终表现。未来在改进预测线路模块的设计时, 也要提升 KQV 各自数值的线路设计。

参考文献

- [1] BROOKS M. Beyond quantum supremacy: The hunt for useful quantum computers[J]. *Nature*, 2019, 574(7776): 19-22.
- [2] PRESKILL J. Quantum computing in the NISQ era and beyond[J]. *Quantum*, 2018, 2: 79.
- [3] FEDOROV D A, PENG B, GOVIND N, et al. VQE method: A short survey and recent developments[J]. *Materials Theory*, 2022, 6(1): 1-21.
- [4] MOLL N, BARKOUTSOS P, BISHOP L S, et al. Quantum optimization using variational algorithms on near-term quantum devices[J]. *Quantum Science and Technology*, 2018, 3(3): 030503.
- [5] BIAMONTE J, WITTEK P, PANCOTTI N, et al. Quantum machine learning[J]. *Nature*, 2017, 549(7671): 195-202.
- [6] HAVLÍČEK V, CÔRCOLES A D, TEMME K, et al. Supervised learning with quantum-enhanced feature spaces[J]. *Nature*, 2019, 567(7747): 209-212.
- [7] BOIXO S, ISAKOV S V, SMELYANSKIY V N, et al. Characterizing quantum supremacy in near-term devices[J]. *Nature Physics*, 2018, 14(6): 595-600.
- [8] HUANG H Y, BROUGHTON M, MOHSENI M, et al. Power of data in quantum machine learning[J]. *Nature Communications*, 2021, 12(1): 1-9.
- [9] BOX G E, PIERCE D A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models[J]. *Journal of the American Statistical Association*, 1970, 65(332): 1509-1526.
- [10] ELMAN J L. Finding structure in time[J]. *Cognitive Science*, 1990, 14(2): 179-211.
- [11] PASCANU R, MIKOLOV T, BENGIO Y. Understanding the exploding gradient problem[EB/OL]. [2022-08-20]. <https://arxiv.org/pdf/1211.5063v1.pdf>.
- [12] GRAVES A. Supervised sequence labelling with recurrent neural networks[M]. [S.l.]: Springer, 2012.
- [13] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[EB/OL]. [2022-08-25]. <https://arxiv.org/pdf/1412.3555.pdf>.
- [14] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. *IEEE Transactions on Neural Networks*, 1994, 5(2): 157-166.
- [15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5998-6008.
- [16] LI S, JIN X, XUAN Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: [s.n.], 2019: 5243-5253.
- [17] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2021: 10012-10022.
- [18] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]//International Conference on Learning Representations. Vienna: [s.n.], 2021.
- [19] FARHI E, NEVEN H. Classification with quantum neural networks on near term processors[EB/OL]. [2022-08-26]. <https://arxiv.org/pdf/1802.06002.pdf>.
- [20] WEI S, CHEN Y, ZHOU Z, et al. A quantum convolutional neural network on nisq devices[J]. *AAPPS Bulletin*, 2022, 32(1): 1-11.
- [21] WIEBE N, BOCHAROV A, SMOLENSKY P, et al.

- Quantum language processing[J]. [2022-09-10]. <https://arxiv.org/pdf/1902.05162.pdf>.
- [22] MEICHANETZIDIS K, GOGIOSO S, DE FELICE G, et al. Quantum natural language processing on near-term quantum computers[EB/OL]. [2022-09-12]. <https://arxiv.org/abs/2005.04147v1>.
- [23] AZEVEDO C R, FERREIRA T A. The application of qubit neural networks for time series forecasting with automatic phase adjustment mechanism[C]//Encontro Nacional de Inteligência Artificial. [S.l.]: [s.n.], 2007: 1112-1121.
- [24] DASKIN A. A walk through of time series analysis on quantum computers[EB/OL]. [2022-09-18]. https://www.engineeringvillage.com/app/doc/?docid=cpx_fb76050180f2128638M640b10178163134.
- [25] EMMANOULOPOULOS D, DIMOSKA S. Quantum machine learning in finance: Time series forecasting[EB/OL]. [2022-09-22]. <https://arxiv.org/abs/2202.00599>.
- [26] DI SIPIO R, HUANG J H, CHEN S Y C, et al. The dawn of quantum natural language processing[C]//ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.]: IEEE, 2022: 8612-8616.
- [27] CHEN S Y C, YOO S, FANG Y L L. Quantum long short-term memory[C]//ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). [S.l.]: IEEE, 2022: 8622-8626.
- [28] LI G, ZHAO X, WANG X. Quantum self-attention neural networks for text classification[EB/OL]. [2022-09-25]. <https://arxiv.org/abs/2205.05625>.
- [29] VOITA E, TALBOT D, MOISEEV F, et al. Analyzing multi-head selfattention: Specialized heads do the heavy lifting, the rest can be pruned[EB/OL]. [2022-09-28]. <https://arxiv.org/abs/1905.09418v2>.
- [30] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[EB/OL]. [2022-10-03]. <https://arxiv.org/pdf/1810.04805.pdf>.
- [31] 杨璐, 南刚强, 陈明轩, 等. 基于三种机器学习方法的降水相态高分辨率格点预报模型的构建及对比分析[J]. *气象学报*, 2021, 79(6): 1022-1034.
- YANG L, NAN G Q, CHEN M X, et al. The construction and comparison of high resolution precipitation type prediction models based on three machine learning methods[J]. *Acta Meteorologica Sinica*, 2021, 79(6): 1022-1034.
- [32] 朱复成. 中尺度天气数值模拟及近年来国外的进展[J]. *气象科技*, 1987(4): 1-7.
- ZHU F C. Recent developments on mesoscale system simulation[J]. *Meteorological Science and Technology*, 1987(4): 1-7.
- [33] CHEN Z Y, XUE C, CHEN S M, et al. Vqnet: Library for a quantumclassical hybrid neural network[EB/OL]. [2022-10-12]. <https://arxiv.org/pdf/1901.09133.pdf>.
- [34] WIERSEMA R, ZHOU C, DE SEREVILLE Y, et al. Exploring entanglement and optimization within the hamiltonian variational ansatz[J]. *PRX Quantum*, 2020, 1(2): 020319.
- [35] SCHULD M, KILLORAN N. Quantum machine learning in feature hilbert spaces[J]. *Physical Review Letters*, 2019, 122(4): 040504.
- [36] ÉREZ-SALINAS A, CERVERA-LIERTA A, GIL-FUSTER E, et al. Data re-uploading for a universal quantum classifier[J]. *Quantum*, 2020, 4: 226.
- [37] MCCLEAN J R, BOIXO S, SMELYANSKIY V N, et al. Barren plateaus in quantum neural network training landscapes[J]. *Nature Communications*, 2018, 9(1): 1-6.

编辑 叶芳