基于 SCAD 先验知识算法的脑卒中预测模型研究

郭　茜[1],赵否曦[2],彭宇翔[3],支亚京[1],汪　华[1],王　娟[1],刘国强[4*]
(1. 贵州省气象数据中心,贵州 贵阳 550002;
2. 贵州省疾病与预防控制中心,贵州 贵阳 550002;
3. 贵州省气象台,贵州 贵阳 550002;
4. 贵州省气象局,贵州 贵阳 550002)

摘　要:该文提出了基于平滑削边绝对偏离(SCAD)先验和注意力机制的脑卒中发病率预测模型。通过深入研究,针对气象特征复杂多变等因素,设计了气象特征注意力模块,以提高模型的特征表达能力,并利用多头注意力机制,有效捕捉气象因子与脑卒中发病率的权重关联信息,降低了无用信息干扰,使得训练过程中更关注对脑卒中发病有显著影响的气象因子。为了在本研究中数据规模有限的情况下,进一步提升模型的预测性能,该文结合了 SCAD 特征筛选方法和先验知识方法,设计了 SCAD 先验知识模块,帮助模型更快收敛,降低模型对数据规模的依赖。该算法相比于多个对比模型在各项评价指标上都有所提升,其中相比于基线算法中表现最好的 Xgboost 模型,MSE 降低了 0.064,R^2 评分提高了 0.054。此外,对该算法进行了消融分析,验证了该算法设计的核心模块的作用,以及对模型性能提升的贡献。实验结果表明,基于该方法设计的模型提高了脑卒中发病率的预测精度,适用于贵州山区脑卒中发病规律的目标识别。

关键词:气象因子;脑卒中发病率;注意力机制;SCAD;先验知识
中图分类号:TP181;R743.3;P49　　**文献标识码**:A　　**文章编号**:1673-629X(2025)05-0136-09
doi:10.20165/j.cnki.ISSN1673-629X.2025.0005

Research on Stroke Incidence Prediction Model Based on SCAD Prior Knowledge Algorithm

GUO Xi[1],ZHAO Fou-xi[2],PENG Yu-xiang[3],ZHI Ya-jing[1],WANG Hua[1],
WANG Juan[1],LIU Guo-qiang[4*]
(1. Guizhou Meteorological Data Center,Guiyang 550002,China;
2. Guizhou Center for Disease Control and Prevention,Guiyang 550002,China;
3. Guizhou Provincial Meteorological Observatory,Guiyang 550002,China;
4. Guizhou Meteorological Bureau,Guiyang 550002,China)

Abstract:An attention mechanism and Smoothly Clipped Absolute Deviation (SCAD) prior-based stroke incidence prediction model is developed. Through in-depth research,the meteorological feature attention module is designed to improve the feature expression ability of the model,and the multi-head attention mechanism is used to effectively capture the weight association information between meteorological factors and stroke incidence,reduce the interference of useless information,and make the training process pay more attention to meteorological factors that have a significant influence on stroke incidence. In order to further improve the prediction performance of the model in the case of limited data scale,we combine SCAD feature screening method and prior knowledge method to design SCAD prior knowledge module to help the model converge faster and reduce the model's dependence on data scale. The proposed algorithm has improved in different evaluation metrics when compared to numerous comparative models. The MSE is decreased by 0.064 for each of them when compared to the baseline algorithm's top-performing Xgboost model,and the R^2 score is raised by 0.054 for each. The suggested algorithm is also subjected to an ablation analysis,which verifies the function of the core modules created for the algorithm in this chapter and its contribution to the enhancement of model performance. In conclusion,the proposed model can aid in stroke incidence prevention by properly predicting the stroke incidence based on meteorological data.

Key words:meteorological factors;stroke incidence;attention mechanism;smoothly clipped absolute deviation (SCAD);prior knowledge

收稿日期:2024-07-30　　**修回日期**:2024-12-02
基金项目:贵州省科技基础研究计划(黔科合基础—ZK[2022]一般 244);贵州省科技支撑计划(黔科合支撑[2023]一般 165)
作者简介:郭　茜(1989-),女,硕士,高级工程师,研究方向为气象大数据应用研究;通信作者:刘国强(1984-),男,硕士,研究方向为健康气象服务数据融合。

0　引　言

随着脑卒中发病率、死亡率的不断上升,人们对脑卒中疾病危害的认识逐步加深,并发现该病的发病率存在季节性特点。近年来智能医疗领域的研究已经深入应用了机器学习等算法,其中对脑卒中发病风险的预测成为研究的热点[1]。因此,借助深度学习算法建立准确的预测模型,在发病风险较高时发出预警,能够有效地帮助人们预防该疾病的发生,最大限度地降低脑卒中的发病率,以减轻患者及家庭承受的压力和经济负担[2]。

针对气象特征的复杂多变,以及脑卒中发病与气象特征呈现弱相关等因素,该文提出基于平滑削边绝对偏离(SCAD)先验和注意力机制的脑卒中发病率预测算法。通过向深度学习模型中引入气象特征注意力机制,使得模型可以更好地关注对脑卒中发病有显著影响的气象特征,将筛选后的结果作为先验知识。在一定程度上降低了对数据量的需求,并在有限数据规模的情况下,提升了模型准确率和泛化能力。

主要贡献如下:

(1)研究并设计了基于 SCAD 先验和注意力机制的脑卒中发病率预测模型。增强了模型的特征表达能力,提升了对脑卒中发病率预测的准确率。

(2)对数据集进行了训练集的数据筛选,并且在此基础上利用引进多头注意力机制的算法,对模型进行了多次对比实验和消融分析,以更好地应对具有复杂度的脑卒中发病识别任务,对模型的有效性及关键组件性能提升得到了有效的验证。

1　基础理论

1.1　SCAD 算法

目前,传统特征筛选法已成功应用于不同领域的模型,具体包括:逐步回归法、前进法、后退法、最优子集法等等。这些方法的基本思想是将特征的所有子集构成的子模型进行比较,然后根据某种信息准则(如 AIC[3]、BIC[4]、AICC[5]等准则)选出最优子模型。Tibshirani[6]在研究中开拓性地提出了最小绝对值压缩和选择算子(Lasso)[7-9]惩罚估计改进方法。Fan 和 Li[9]提出了一种新的特征筛选方法—平滑剪切绝对偏差惩罚(Smoothly Clipped Absolute Deviation,SCAD),并且该惩罚函数同时具备优质惩罚函数的三大优势:稀疏性、无偏性和连续性。SCAD 惩罚函数定义如公式 1、公式 2 所示。

$$p_\lambda=\begin{cases}\lambda|\beta|, & |\beta|\leqslant\lambda\\ 0-\dfrac{\beta^2-2a\lambda|\beta|+\lambda^2}{2(a-1)}, & \lambda<|\beta|\leqslant a\lambda\\ \dfrac{(a+1)\lambda^2}{2}, & |\beta|>a\lambda\end{cases}\quad(1)$$

求得 SCAD 惩罚函数的一阶导数为:

$$p'_\lambda(\beta)=\lambda\{I(\beta\leqslant\lambda)+\frac{a\lambda-\beta}{(a-1)\lambda}+I(\beta>\lambda)\}\quad(2)$$

式中,β 代表特征向量,$a>2$ 代表调节参数,$\lambda\geqslant 0$ 代表惩罚参数,并且 $p_\lambda(0)=0$。在公式 2 中,I 代表示性函数,当满足某种性质时,其值为 1,否则为 0。调节参数 a 和惩罚参数 λ 分别控制了特征筛选的平滑性和稀疏性,当 a 越大时,惩罚函数曲线越平缓,使得特征筛选越趋向于平滑,同时减少了估计误差。当 λ 越大时,对于一些系数较小的特征,其估计量会被惩罚为 0,从而控制了特征的稀疏性。当公式 2 结果为 3.7 时,在贝叶斯风险下接近最优。

1.2　注意力机制

注意力机制可以更好地捕捉局部信息和全局信息,卷积神经网络(CNN)[10]和循环神经网络(RNN)[11]等通常只能在局部区域或序列上进行特征提取,而注意力机制能够在全局范围内对输入特征进行加权,从而更好地捕捉全局信息。针对气象特征十分复杂多变,且与脑卒中发病呈弱相关性,通过引入注意力机制中的点积注意力机制,可以增强模型的特征表示能力,帮助模型关注对脑卒中发病有显著影响的气象因子。通过输入 3 个特征经过线性映射,分别是 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$,代表查询、键和值矩阵,表示键的维数,点积注意力机制的公式如下:

$$\text{Attention}_{\text{Dot}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_k}})\boldsymbol{V}\quad(3)$$

接着计算 $\boldsymbol{Q}$ 与所有特征的 $\boldsymbol{K}$ 的内积,从而得到注意力分数,该分数代表两个特征之间的相关程度。注意力分数经过输入端、网络头部、骨干网络最后输出。在分类任务中使用 softmax 函数进行归一化,形成回归的概率分布,降低了数据维度,可以得到每个气象因子特征真正的权重,该权重与 Value 进行逐个相乘加权求和,得到最终的输出特征。相比传统模型,加入点积注意力机制后其表示能力显著增强。

1.3　Transformer 模型

Transformer 模型是由 Vaswani 等人在 2017 年所提出的一种完全基于注意力机制的模型[11-12]。基于自注意力机制的设计的深度架构在并行性、最大路径长度方面都比卷积神经网络、循环神经网络更有优势。

Transformer 的架构由自注意层、前馈层和标准化层组成。自衰减层允许模型从序列中不同时间点之间的关系中学习,促进捕获复杂的时间模式。自衰减块决定了序列中每个元素相对于电流的重要性元素这个过程是通过为序列中的每个标记生成查询($\boldsymbol{Q}$)、键($\boldsymbol{K}$)和值($\boldsymbol{V}$)向量来实现的。通过这种方式,更多

相关的令牌对当前的令牌编码的贡献更大,从而产生更准确和更有意义的数据编码。

2 脑卒中发病率预测算法

该文结合了 SCAD 特征筛选方法和先验知识的思想,设计了 SCAD 先验知识模块,以帮助模型更快地收敛,旨在有限的数据规模下,进一步提升模型的预测性能,降低模型对数据规模的依赖。此外,由于气象特征复杂多变,并且与脑卒中发病呈弱相关性,需要模型具有更强的特征表达能力,并且能够在学习过程中关注到对脑卒中发病有显著影响的气象特征,才能达到优秀的预测性能。因此,对 Transformer 的骨干网络进行了改进,并设计了气象特征注意力模块,以解决上述问题。基于先验知识模块脑卒中预警模型的结构如图 1 所示。

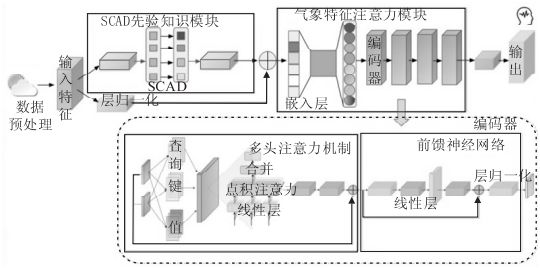


图 1 基于 SCAD 先验知识和注意力机制的脑卒中发病率预测算法网络结构

2.1 算法网络结构

文中算法的骨干网络基于 Transformer[13] 模型改进,保留了编码器(Encoder)部分,并去除了解码器(Decoder)部分。将 Transformer 模型应用在不同的实际问题中,脑卒中算法模型的建立主要基于创新改造后的 Transformer 的模型架构[14-15],如图 2 所示。根据气象因子的变化预测脑卒中的发病率,需要建立预测模型,模型的输入是经过预处理后的气象特征,输出为一个发病率预测值。与序列到序列的建模不同,模型的输出结果是一个预测值。

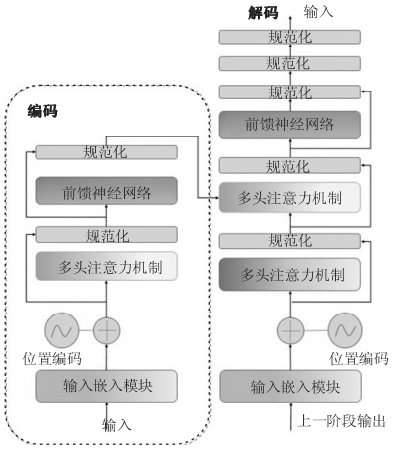


图 2 改进后的 Transformer 模型模块结构

Transformer 模型显示了机器反应的最新结果和反应预测结果。训练的变压器在大型和有噪声的数据集上的结果是一个模型。解码器会对序列进行掩码处理,先前的位置会生成相应后面位置的输出,以此相互依赖,编码后得到的表示可以用来表示预测或分类,以此减少信息流动的损失,也降低了计算成本。此外,解码器部分还需要对输入序列进行一次编码和一次解码,这样会增加信息的损失和噪声。去掉解码器可以直接使用编码后的表示进行预测或分类,减少信息流动的损失。

气象数据经过预处理后的结果作为模型的输入。与此同时,经过预处理后的气象特征也输入到 SCAD 先验模块中,该模块使用 SCAD 算法对气象特征进行筛选,经过 SCAD 处理后,每个特征会得到一个系数,该系数代表了该气象因子与脑卒中发病的关联性,关联性弱的气象特征的系数会被衰减或变为 0。将 SCAD 筛选后的权重矩阵作为先验知识,与 Transformer 模型输入特征的权重矩阵共同构成输入到编码器的输入向量。在反向传播更新参数的过程中,分别对每个特征的初始化矩阵参数和 SCAD 先验知识权重矩阵参数进行求偏导以计算梯度,并分别更新对应的参数,具体见公式 4 和公式 5。

$$y = W_{\text{prior}}\chi + W_0\chi \tag{4}$$

$$\nabla_{xy} = \nabla_{x_0}y + \nabla_{x\,\text{prior}}y \tag{5}$$

将 SCAD 先验知识和输入特征映射到高维向量空间中,离散的特征转换为连续的实数向量。最后,通过线性层将经过注意力机制模块后的结果通过降维得到一维的结果,这个结果作为最终的预测输出,即脑卒中发病率的预测值。

2.2 SCAD 先验知识模块

该文设计了 SCAD 先验知识模块(SCAD Prior Knowledge Module),将其简称为 SCAD-PKM,模块结构如图 3 所示。深度学习在临床应用方面存在两个限制因素:缺乏可解释性和需要大量数据驱动。注意力机制具有高度的适应性和灵活性,能够在不同的层进行多次叠加,每次叠加都可以关注不同的信息。当使用足够多数据用于训练时,理论上注意力模块叠加的越深,模型的效果就越好。

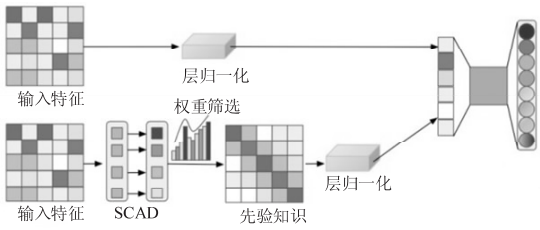


图 3 SCAD-PKM 模块结构

受到机器学习的启发,在深度学习中,除了输入模

型中作为训练集的数据之外,还可以额外集成知识,即先验知识。SCAD-PKM 模块,通过向模型中加入先验知识的方式,帮助模型在已有数据集上更快地收敛,从而在一定程度上降低了模型对数据量的依赖,同时先验知识比用纯数据驱动的模型用了多维协作学习策略,通过一维卷积实现跨维信息交互。

在实现 SCAD 算法时,先求解无惩罚项的对数似然函数,并根据该结果来设置初始值,然后再去迭代加上 SCAD 惩罚项的损失函数,从而减少带惩罚项的对数似然函数较为复杂的计算。通过实验验证,一个合适的初始点能够大大减少牛顿高斯迭代的计算量。上述做法有效地加快了 SCAD 的收敛速度。

将 SCAD 筛选后的权重向量记为 $\boldsymbol{X}_{\text{prior}}$,接着对其和预处理完毕后的特征输入向量分别进行层归一化,然后再将它们串联,作为文中算法注意力模块的嵌入层(Embedding)的输入。这种串联的方式能够将 SCAD 筛选得到的权重作为先验知识,与原始的输入特征相结合,每个特征变量对应于一个原始的向量和一个先验知识向量,两个向量共同参与模型后面的训练和学习,如下所示。

$$\boldsymbol{X} = [\boldsymbol{X}_0; \boldsymbol{X}_{\text{prior}}] \tag{6}$$

加入 $\boldsymbol{X}_{\text{prior}}$ 后,$\boldsymbol{X}_{\text{prior}}$ 对模型的学习起到了一定的指导作用,可以帮助模型更快地收敛。

2.3　气象特征注意力模块

本模型设计了气象特征注意力模块(Meteorological Characteristics Attention Module),将其简称为 MCAM,模块结构如图 4 所示。该模块利用对模型混合特征的加权聚合,提出了一个混合加权聚合算子,并讨论了它的几个重要性质,使模型能够调整对不同气象特征的关注度,获得更有代表性的气象特征表示,从而使得模型的注意力高效集中在对脑卒中发病有显著影响的气象特征。

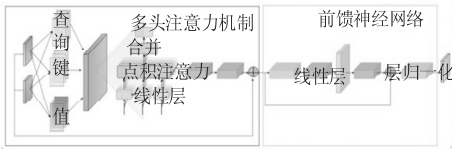


图 4　MCAM 模块结构

(1)Embedding 层。

Embedding 层的作用是将数据特征集合映射到向量空间,进而把数据进行向量化的过程。在 MCAM 中,输入由两部分组成,第一部分是数据预处理后得到的 DataFrame 格式的原始气象特征数据,另一部分是经过 SCAD-PKM 模块处理后得到的先验知识,这两部分作为独立的变量参与模型的学习。本模型的 Embedding 层通过稀疏的方式处理气象特征,同时融合了大量基本特征,将气象因子的特征升为 256 维高阶特征向量。经过该层后,模型将获得预测输入矩阵的嵌入向量,随后需要通过线性映射(Liner Projection)来将嵌入向量转换为矩阵:查询 $\boldsymbol{Q}$、键 $\boldsymbol{K}$、值 $\boldsymbol{V}$。$\boldsymbol{W}_q, \boldsymbol{W}_k, \boldsymbol{W}_v$ 是三个参数矩阵。该矩阵被作为输入参数传入到编码器 Encoder 模块中。

(2)多头注意力机制层。

文中算法的多头注意力机制层是由多个自注意力机制结构组成,其中自注意力机制采用缩放点积注意力机制的形式。对 Query 有显著影响的 Key 会获得较高的注意力得分,并通过求得注意力得分与值(Value)相乘后的加权和来获得最终的输出,作为模型所关注的注意力。

在进行点积运算时,当查询向量的维度 d_k 较大时,则会导致一些结果的值过大或过小,此时较大的值通过 softmax 函数输出的结果就会更加接近于 1,其余的值经过 softmax 函数后就会更加趋近于 0,这种情况下得到的结果的分布就会更加向两端靠拢,导致模型计算梯度时得到的梯度值很小。为了使模型具有突出显著特征的能力,利用一个维度增强注意(DEA)模块,即插即用模块,它被嵌入到缩放点积注意力机制中,以确保注意力得分不会过大或过小,从而缓解 softmax 函数的梯度消失问题。

为了计算模型中的输出向量,分别将矩阵 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 进行顺序合并,从而客观地得到输出向量的值,$\boldsymbol{X} = [\boldsymbol{X}_0; \boldsymbol{X}_{\text{prior}}] x_i q_i k_i v_i$。因此,缩放点积注意力机制的矩阵计算过程如图 5 所示,输入时包括大小为 d_k 的 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$,以及大小为 d_v 的 $\boldsymbol{V}$。通过计算 $\boldsymbol{Q}$ 与所有 $\boldsymbol{K}$ 点积的值,并将所得到的结果除以均方根 $\sqrt{d_k}$,最终通过输入 softmax 函数取得最终权重的值为 Y_{attn}。

$$Y_{\text{attn}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \sum_{i=1}^{n} v_i \frac{\exp\left(\frac{k_i^{\mathrm{T}} q_i}{\sqrt{d_k}}\right)}{\sum_{j=1}^{n} \exp\left(\frac{k_i^{\mathrm{T}} q_i}{\sqrt{d_k}}\right)} \tag{7}$$

$$Y_{\text{attn}}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = V_{\text{softmax}}\left(\frac{\boldsymbol{K}^{\mathrm{T}}\boldsymbol{Q}}{\sqrt{d_k}}\right) \tag{8}$$

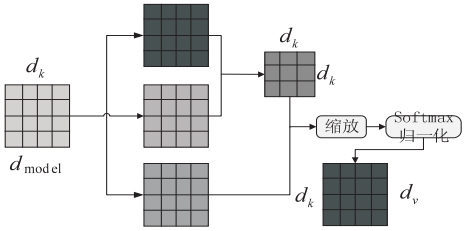


图 5　缩放点积注意力机制矩阵计算过程

多头注意力机制将输入的 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 矩阵分别进行线性映射,将它们投影到低维空间,每个注意力的头都包含了一个自注意力机制,捕捉输入序列在不同子空间中的信息,并且将计算得到的结果对应到每个头。

在计算了多个头的结果后得到注意力输出,同时将输出的结果进行串联合并。随后,使用 $\boldsymbol{W}^0$ 权重矩阵对最终结果进行线性变换,对向量进行降维,增强模型的表达能力,一定程度上解决了气象特征复杂多变和气象特征与脑卒中发病弱相关的问题。帮助模型更好地理解输入特征,同时,线性并行计算增加了预测模型计算的速度,多头注意力机制的计算过程表示为:

$$M(\boldsymbol{Q}、\boldsymbol{K}、\boldsymbol{V}) = \text{concat}(\text{head}_1, \text{head}_2, \cdots, \text{head}_n)\boldsymbol{W}^0 \tag{9}$$

$$\text{head}_i = Y_{\text{attn}}(\boldsymbol{Q}\boldsymbol{W}_i^Q, \boldsymbol{Q}\boldsymbol{W}_i^K, \boldsymbol{Q}\boldsymbol{W}_i^V) \tag{10}$$

利用 LayerNorm 函数对结果进行归一化处理后,将多头注意力机制层的结果作为输入,放到前馈神经网络中。其公式如下所示:

$$y = \text{LayerNorm}(\chi + Y_{\text{attn}}(\chi)) \tag{11}$$

$$\text{LayerNorm}(\chi) = \frac{a(\chi - \mu)}{\sigma + \varepsilon} + b \tag{12}$$

残差连接将当前层的输入加上自身的输出,从而保留输入信息,解决因网络加深而导致的模型退化问题。层归一化其实就是数据标准化的一个过程,将输入的矩阵参数进行处理后使其限定在均值为 0,方差为 1 的范围内,这样加快了梯度下降求最优解的速度,同时提高了结果的精度。

(3)前馈神经网络层。

经过多头注意力机制层的处理后,模型得到了一个经过编码的特征表示。为了增强对特征的非线性表达能力,特征需要通过前馈神经网络进一步加工和处理,其单元之间的连接不会形成一个循环,中间使用激活函数进行非线性变换。其中, W_1W_2 为权重参数, b_1b_2 为偏置参数。

$$\text{FFN}(\chi) = \max(0, \chi\boldsymbol{W}_1 + b_1)\boldsymbol{W}_2 + b_2 \tag{13}$$

并且为了防止过拟合,在层归一化前也加入了一层 Dropout 操作,最终前馈神经网络经过归一化处理完毕后,经过一个线性变换层降为 1 维,该结果就是模型预测的脑卒中发病率。

2.4 损失函数

考虑到气象特征的复杂多变性以及气象因子与脑卒中发病之间的弱相关性,选取 Log-Cosh 损失函数,从而避免了大误差的影响。针对预测值和真实值的差异较小的情况,Log-Cosh 损失函数将近似 MSE 损失函数,保证了平滑的特性。Log-Cosh 损失函数公式如下所示:

$$L_{\text{Log-Cosh}}(y, \hat{y}) = \sum_{i=1}^{n} \log[\cos(\hat{y} - y)] \tag{14}$$

3 实验分析

3.1 数据集

本研究使用的数据包括贵州省疾控中心提供的 2017 年至 2020 年脑卒中患者的病例数据,涵盖了性别、年龄、发病时间和诊断时间等关键信息,共计 62 318 条;以及来自贵州省气象局提供的当地气象站的逐日气象数据,包括逐日的气温(平均值、最高值、最低值、24 小时变温)、气压(平均值、最高值、最低值、日较差)、平均相对湿度、日照时数、水气压、平均风速、云量、能见度、地表温度等多项气象指标。

3.2 数据处理

该文采用 Pandas 处理数据,首先对数据进行读取,将表格型数据读取为 DataFrame 形式。使用 Pandas 的解析函数 read_excel( ) 从 Excel 的 XLS 文件中读取表格数据。然后使用 Pandas 等 concat( ) 函数将读取到的病例数据和气象数据分别整合,使对象在轴上进行堆叠,完成数据的合并。在完成数据清理后,为了构建实验的数据集,需要将病例数据和气象数据进行关联,完成气象数据和病例数据的数据融合。最后,需要为数据集进行打标签处理,其中打标签的逻辑如公式 15 所示。

$$\text{label} = \frac{d_c}{y_c} \tag{15}$$

其中,label 代表标签(发病率), d_c 代表一个日期的发病人数, y_c 代表当前年份的年总发病人数。

使用 Pandas 的 groupby( ) 函数按照年份和日期进行分组,然后使用 count( ) 函数分别计算出年总发病人数和日发病人数,根据公式 15 进行计算,从而完成对每个样本的打标签处理。公式 14 是 Log-Cosh 损失函数,如何确定数据样本,有待进一步说明。

3.3 实验设置

文中算法使用经过预处理后的数据集,其中 2017-2019 的数据作为训练集,共包含 46 738 个样本。将 2020 年的数据作为本轮实验的测试集,共包含 14 681 个样本。在训练过程中,每个训练批(Batch)输送 256 个样本至模型中。训练轮次为 1 万次。epochs 采用 Adam 优化器来更新网络参数,优化训练过程中学习率的调整,Dropout 层的丢弃率为 0.3。文中算法的损失阈值 loss_threshold 设为 0.5×10^{-9} ,训练过程中,满足损失小于 loss_threshold 或达到最大训练轮次 epochs 时停止训练。

3.4 评价指标

该文研究的是根据气象因素预测脑卒中发病率问题,因此本研究预测的结果是一个连续变量(脑卒中发病率),所以本研究属于回归问题。为了综合评估模型的准确率和泛化能力,选取的评价指标包括 MSE、MAE、R^2 score。

均方误差(MSE)用于评价模型的准确性,是实际值与预测值差值的平方的平均值,其对应于平方误差

的期望。MSE 如公式 16 所示。该指标的取值范围为 MSE 最优值为 0,最差值为 + ∞。

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{16}$$

平均绝对误差(MAE)是预测值与真实值之间误差的绝对值取平均,用于异常值代表损坏的部分,如公式 17 所示。MAE 最优值为 0,最差值为 + ∞。MAE 并没有惩罚太多的训练异常值(L1 范数以某种方式平滑了所有可能的异常值的误差),从而为模型提供了一个通用的和有界的性能度量。

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{17}$$

R^2 反映了预测值与真实值间的拟合程度,该指标对应于拟合优度。总平方和(SST)是由具体的因素变化引起的,代表真实数据相对均值的离散程度;误差项平方和(SSE),组内方差,表示实验误差大小的偏差平方和。计算结果为真实数据和预测数据之差的平方和越接近于 1,代表模型的拟合程度越好。R^2 的取值可以是负值,但是这意味着回归表现不佳。当回归模型不解释响应数据在其平均值周围的可变性时,R^2 可以为 0。脑卒中发病决定系数的正值在[0,1]区间内,1 表示完美的预测。

$$\mathrm{SST} = \sum_{i=1}^{n}(y_i - \hat{y})^2 \tag{18}$$

$$\mathrm{SSE} = \sum_{i=1}^{n}(f(x_i) - y_i)^2 \tag{19}$$

$$R^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} \tag{20}$$

3.5　实验结果分析

3.5.1　预测性能对比实验

通过 MSE、MAE、R^2 三种误差评价指标,与现有的七种算法(多元线性回归、岭回归、SVM、随机森林、Lightgbm、Xgboost、BP 神经网络)进行对比。不同算法在测试集上的实验结果如表 1 所示。

表 1　模型预测性能对比

算法	MSE	MAE	R^2 score
文中算法	**0.874**	**0.272**	**0.747**
多元线性回归	1.102	0.284	0.55
岭回归	1.104	0.281	0.547
SVM	2.381	0.717	-0.572
随机森林	1.015	0.273	0.608
Lightgbm	0.987	0.278	0.650
Xgboost	0.938	0.275	0.693
BP 神经网络	0.986	0.269	0.629

由实验结果可以看出,提出的基于 SCAD 先验和注意力机制的脑卒中发病率预测模型,相比于其他基线算法,在 MSE、MAE、R^2 等指标上均取得了最佳结果。表明针对研究的任务,文中算法建立的模型具有更高的准确性,拟合更良好,模型的泛化性能更佳。

为了能更直观地观察预测值和真实值的分布情况,在本实验中,绘制了每个算法对训练集和测试集中的每个样本的预测值,并和标签(发病率)的真实值放在一起构成散点图,如图 6 所示。其中,圆形图标代表真实的发病率,菱形图标代表模型预测的发病率。左侧 Train DataSet 代表该算法在训练集上的实验结果,右侧 Test DataSet 代表该算法在测试集上的实验结果。8 种机器学习模型的预测值(菱形)与实际值(圆形)之间的拟合程度可以由图 6 所示。

通过以上各图可以直观地看出不同模型之间性能的差异。基于 SCAD 先验和注意力机制的脑卒中发病率预测算法所建立的模型,在训练集和测试集上的预测值都十分接近真实值,达到了最好的预测效果。

3.5.2　滑动窗口实验

气象因素诱发脑卒中发病的机制复杂,剧烈天气变化和脑卒中发病存在密切关联。而文中算法模型中考虑的气象特征主要为上述数据集中的气象因素。为了分析各个气象因素在文中算法设计中与脑卒中发病的暴露-反应关系,同时能够更好地提高模型预测的性能,选择滑动窗口的方式,通过对脑卒中发病前的不同时间周期的各个气象特征做加权平均处理,并将处理结果输入模型中,同时控制模型的网络结构和参数不变,进行对比实验。实验中滑动窗口 S_w 大小取值为 3 天、5 天、7 天、11 天、15 天、30 天。当 S_w = 1 时为文中算法的初始情况。S_w 实验结果如表 2 所示。

表 2　滑动窗口实验结果

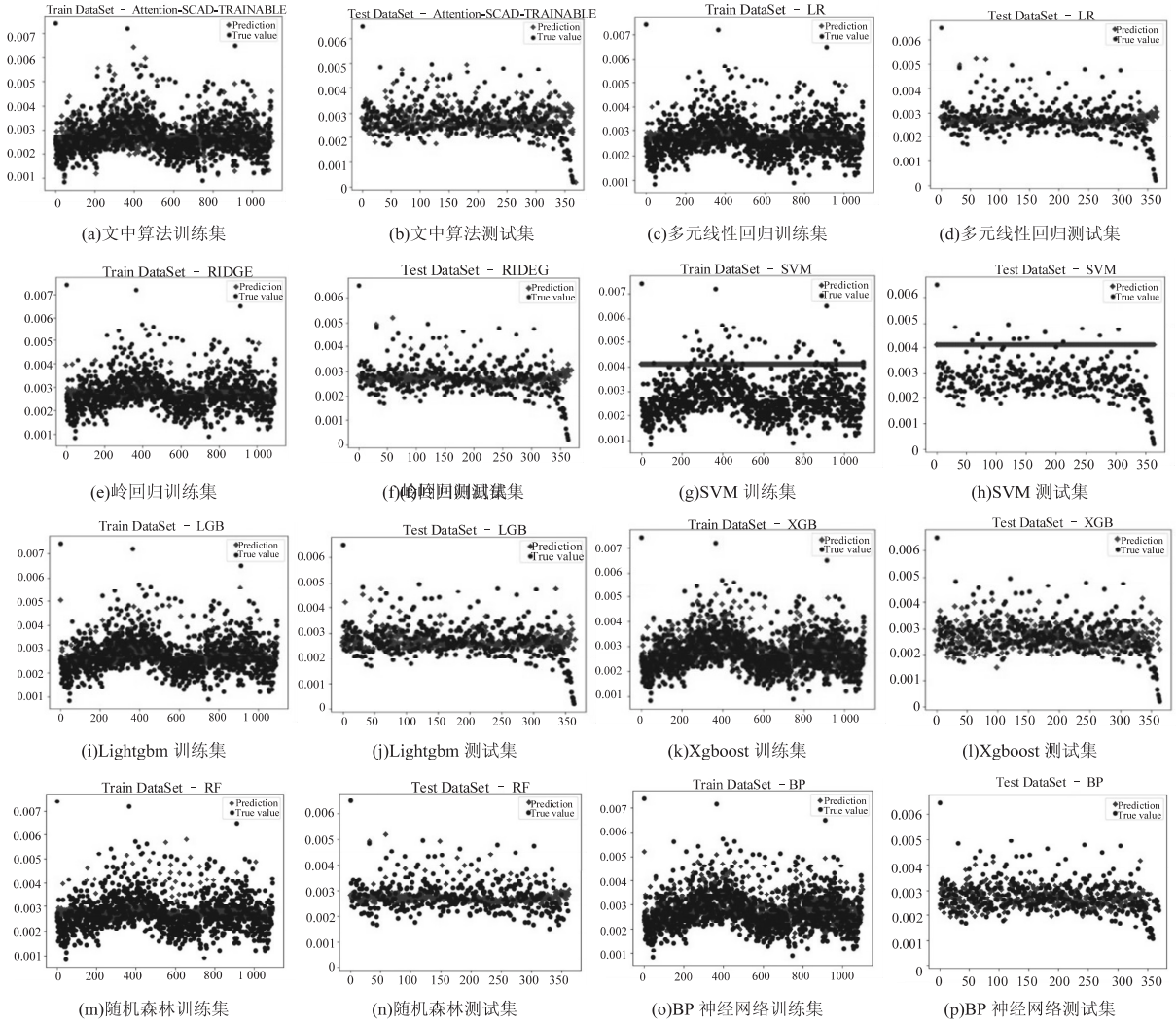
S_w	MSE	MAE	R^2 score
1	0.874	0.272	0.747
3	0.878	0.291	0.743
5	0.858	0.285	0.764
7	0.845	0.269	0.778
11	0.851	0.287	0.763
15	**0.828**	**0.286**	**0.789**
30	0.839	0.298	0.780

从实验结果中可以看出,当滑动窗口 S_w 取 15 时,即考虑脑卒中患者发病当日和前 14 天的气象指标和气象变化时,模型的预测性能表现最为准确,在 MSE、R^2 等指标上均取得了最优的结果,而当滑动窗口 S_w 取 15 时,MAE 指标取得了最优结果。

3.6　消融分析实验

本节分析了文中算法所建立的模型中核心模块的有效性,从而验证了文中算法提出的网络结构及其改

进点的可靠性。



(a)文中算法训练集 (b)文中算法测试集 (c)多元线性回归训练集 (d)多元线性回归测试集

(e)岭回归训练集 (f)岭回归测试集 (g)SVM 训练集 (h)SVM 测试集

(i)Lightgbm 训练集 (j)Lightgbm 测试集 (k)Xgboost 训练集 (l)Xgboost 测试集

(m)随机森林训练集 (n)随机森林测试集 (o)BP 神经网络训练集 (p)BP 神经网络测试集

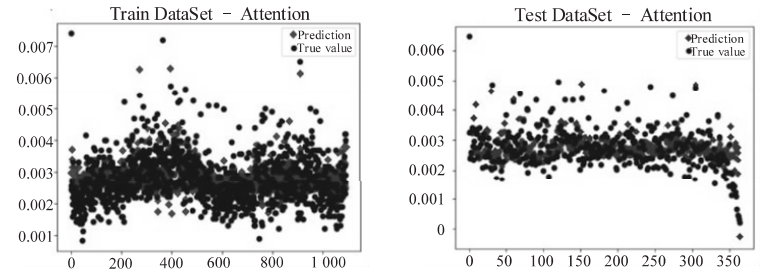
图 6 文中算法与基线算法在训练集和测试集上预测值和真实值的分布图

本节旨在验证 SCAD 先验模块帮助模型更快收敛并提升模型预测性能方面的作用。为此,本节设计三种不同的变体网络,分别为 No_SCAD_Attn、SCAD_Attn、SCAD_Attn_Train。No_SCAD_Attn 网络为文中算法完全删除掉 SCAD 先验知识模块,只采用原始的输入特征进行训练,后续网络结构不变。SCAD_Attn 网络则将气象特征通过 SCAD 算法筛选后得到的特征向量与原始的特征向量输入到 Embedding 层中,分别得到它们所对应的权重矩阵,随后将二者相加,相当于原始特征向量在自身的权重矩阵上增加了一个常数,作为一定的先验知识。而 SCAD_Attn_Train 为文中算法采用的 SCAD 先验知识自适应学习的方式。将经过 Embedding 层后得到的 SCAD 先验知识权重矩阵和原始特征的权重矩阵串联,相当于每个特征对应于两个变量,SCAD 先验知识也参与到训练中并更新参数。实验结果如表 3 所示。

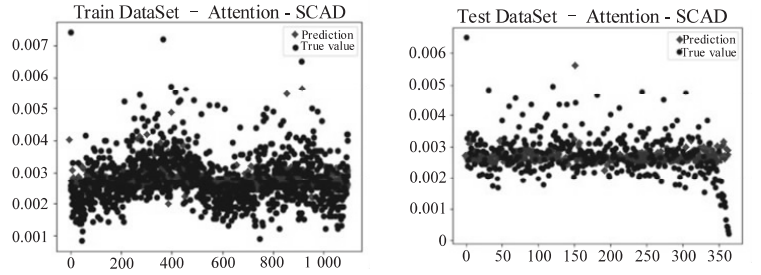
表 3 SCAD 先验知识模块消融实验结果

变体网络	MSE	MAE	R^2 score	Cost/s
No_SCAD_Attn	0.913	0.282	0.715	2 338.6
SCAD_Attn	0.927	0.281	0.709	**1 724.3**
SCAD_Attn_Train	**0.874**	**0.272**	**0.747**	1 980.5

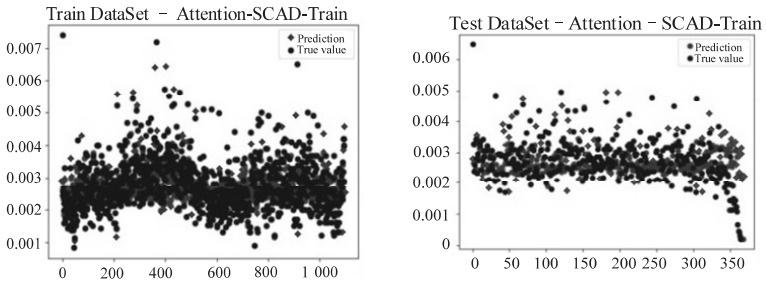
实验结果表明,SCAD_Attn_Train 相比其他两种方式,在各个指标上都有了提升。而 SCAD_Attn 在 MAE 上有了提升,但在其他指标上性能下降。此外,实验还对三种不同网络训练中的耗时做了统计。结果表明,加入 SCAD 先验知识模块,显著地缩短了训练时间,帮助模型更快地收敛。本节同样绘制了每个网络在训练集和测试集中每个样本的预测值和真实值的散点图,以直观地观测模型的预测性能,如图 7 所示。



(a)No_SCAD_Attn



(b)SCAD_Attn



(c)SCAD_Attn_Train

图 7　SCAD 先验知模块消融实验预测值与真实值分布图

3.7　气象特征关联分析

通过上述实验,当气象诱发脑卒中发病模型选择本文算法设定滑动窗口 $S_w = 15$,同时,选取 SCAD_Attn_Train 先验知识消融实验训练数据时,得到脑卒中发病率与各个气象要素相关系数,如表 4 所示。发病当日至前 15 日气象特征与脑卒中相关性,从模型表现中可以看出,气压与脑卒中发病影响较明显;发病当日至前 14 日平均本站气压、最高本站气压和最低本站气压与发病率呈较显著的负相关,其中发病当日相关系数最小,分别为-0.19、-0.17 和-0.2,表明随着气压升高,发病数逐渐降低。前 1 日至前 14 日最高气温、平均湿度与发病率呈正相关,发病率与当日日变温呈正相关;前 1-15 日的日照时数、能见度、日较差和最低温度均与脑卒中发病率呈负相关,与其余气象特征相关性不显著。总体上说,通过文中相关算法以及实验设计研究,脑卒中发病与气压和湿度相关性较显著,当 24 小时平均气温升高、前 1 日-15 日相对湿度增大时,贵州省脑卒中发病人数增多,前 1 日气压降低、前 1-15 日照时数减少、能见度降低、前 14 日气温日较差变小、以及前 15 日最低气温较低时,脑卒中发病率增高。

上述相关性分析表明,气温、气压、平均相对湿度、日照时数等多项气象特征与脑卒中发病有一定的相关性,其中,冷热效应针对脑卒中发病影响明显。

表 4　发病率与各气象特征相关系数

	日变温	最低气温	最高气温	平均湿度	日照时数	平均气压	最高气压	最低气压
当日	0.12	-0.12	0.13	0.11	-0.09	-0.19	-0.17	-0.2
1 日	0.13	-0.11	0.15	0.12	-0.10	-0.13	-0.12	-0.14
2 日	0.14	-0.12	0.13	0.11	-0.12	-0.11	-0.1	-0.13
3 日	0.12	-0.13	0.11	0.12	-0.13	-0.11	-0.1	-0.12
4 日	0.12	-0.10	0.10	0.14	-0.10	-0.12	-0.11	-0.14

续表 4

	日变温	最低气温	最高气温	平均湿度	日照时数	平均气压	最高气压	最低气压
5 日	0.11	-0.14	0.10	0.13	-0.11	-0.15	-0.13	-0.12
6 日	0.10	-0.13	0.15	0.12	-0.11	-0.12	-0.12	-0.13
7 日	0.09	-0.11	0.16	0.11	-0.12	-0.12	-0.11	-0.11
8 日	0.10	-0.11	0.14	0.12	-0.13	-0.12	-0.12	-0.11
9 日	0.10	-0.12	0.13	0.10	-0.14	-0.13	-0.13	-0.12
10 日	0.08	-0.15	0.11	0.11	-0.13	-0.14	-0.13	-0.13
11 日	0.09	-0.12	0.12	0.11	-0.13	-0.14	-0.12	-0.13
12 日	0.09	-0.12	0.10	0.10	0.15	-0.14	-0.11	-0.12
13 日	0.08	-0.11	0.10	0.10	-0.14	-0.12	-0.10	-0.12
14 日	0.07	-0.12	0.10	0.12	-0.14	-0.11	-0.11	0.10
15 日	0.08	-0.11	0.10	0.11	-0.15	-0.10	-0.09	0.10

4 结束语

结合了 SCAD 特征筛选方法和先验知识的思想，设计了 SCAD 先验知识模块，以帮助模型更快地收敛，旨在有限的数据规模下，进一步提升模型的预测性能，降低模型对数据规模的依赖。此外，由于气象特征复杂多变，并且与脑卒中发病呈弱相关性，需要模型具有更强的特征表达能力，并且能够在学习过程中关注到对脑卒中发病有显著影响的气象特征，才能达到优秀的预测性能。因此，对 Transformer 的骨干网络进行了改进，并设计了气象特征注意力模块，以解决上述问题。设计了多个对比和消融实验，验证了该算法在模型预测性能上的提升。实验结果表明，该算法相比其它基线算法，在预测性能和泛化能力上都取得了优异的表现。在消融实验中，证明了算法设计的核心模块的作用，以及对模型性能提升的贡献。综上所述，该预测模型能够根据气象因素较为准确地预测脑卒中发病率，对于医疗部门针对脑卒中发病干预做出了应有的贡献。

参考文献：

[1] 韩沛文. 人工智能在脑卒中风险评估中的应用[J]. 中国新通信，2019，21(4)：68-69.

[2] 白 良，陈 娅. 如何防治脑卒中：脑卒中的分类[J]. 特别健康，2019(23)：54-55.

[3] 陈鸥宇，刘怡俊，叶武剑，等. 基于深度学习和 MFCC 特征的脑卒中预测方法[J]. 信息与电脑：理论版，2019(3)：141-143.

[4] 蒋望东，林士敏，鲁明羽. 基于 BIC 测度和混合遗传算法的 BNC 结构学习[J]. 计算机技术与发展，2007，17(3)：84-87.

[5] 陈笑屹. 深度梯度提升模型及其在脑卒中预测中的应用[D]. 石家庄：河北地质大学，2019.

[6] 李洁洁，张雁儒，李 昊，等. 机器学习在脑卒中预测中的研究进展[J]. 河南医学研究，2022，31(20)：3832-3835.

[7] 郑杰生，谢彬瑜，吴广财，等. 一种基于 Lasso 回归的微服务性能建模方法[J]. 计算机技术与发展，2020，30(12)：216-220.

[8] 刘 洋. 基于机器学习的脑卒中预测模型的研究及其应用[D]. 哈尔滨：东北林业大学，2022.

[9] 杨俊赛. 基于机器学习算法的进展性缺血性脑卒中预测模型的构建和比较分析[D]. 南昌：南昌大学，2024.

[10] 许必宵，宫 婧，孙知信. 基于卷积神经网络的目标检测模型综述[J]. 计算机技术与发展，2019，29(12)：87-92.

[11] 杨祎玥，伏 潜，万定生. 基于深度循环神经网络的时间序列预测模型[J]. 计算机技术与发展，2017，27(3)：35-38.

[12] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is all you need[J]. Advances in Neural Information Processing Systems，2017，30：5998-6008.

[13] LIN T，WANG Y，LIU X，et al. A survey of transformers[J]. AI Open，2022，3：111-132.

[14] 张晓飞，宋其江. 基于 RF-RFECV 和 Stacking 集成学习的脑卒中预测研究[J]. 智能计算机与应用，2024，14(5)：252-256.

[15] 张玉瑶，程学林，尹天鹤. 基于深度学习和矩阵分解的推荐算法[J]. 计算机技术与发展，2021，31(7)：21-27.