

文章编号: 2095-2163(2024)02-0177-06

中图分类号: TP183

文献标志码: A

# 基于 SVM-LSTM-ATTE 组合模型的高校录取分数线预测

秦信芳<sup>1</sup>, 魏嘉银<sup>1</sup>, 姚林<sup>2</sup>, 卢友军<sup>1</sup>, 干霞<sup>1</sup>, 来小孟<sup>1</sup>

(1 贵州民族大学 数据科学与信息工程学院, 贵阳 550025; 2 贵州民族大学 人事处, 贵阳 550025)

**摘要:** 高校录取分数线预测受到许多动态因素影响,传统的 SVM 算法在预测高校录取分数线方面存在难以衡量输入特征序列对目标特征的影响程度,而注意力机制可以动态分配权重给重要特征,且基于注意力机制的长短期记忆网络(LSTM-ATTE),在高校录取分数线方面有显著成效。综合考虑多种因素,利用 SVM 可以把高维空间中的非线性问题转化为线性问题,以及 LSTM-ATTE 能解决时间序列数据长期依赖的问题,提出将 SVM 模型与 LSTM-ATTE 方法相结合的组合模型,来构建高校录取分数线预测模型,并通过网格搜索寻找组成 LSTM-ATTE 最优参数组合。实验证明,本文提出的组合模型在误差范围为 3 分的情况下,相比于其他模型准确率最高提升了 13%。

**关键词:** 长短期记忆网络;支持向量机;注意力机制;网格搜索

## University admission score prediction based on SVM-LSTM-ATTE model

QIN Xinfang<sup>1</sup>, WEI Jiayin<sup>1</sup>, YAO Lin<sup>2</sup>, LU Youjun<sup>1</sup>, GAN Xia<sup>1</sup>, LAI Xiaomeng<sup>1</sup>

(1 School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China;

2 Personnel Office, Guizhou Minzu University, Guiyang 550025, China)

**Abstract:** The prediction of university admission score thresholds is influenced by many dynamic factors. Traditional SVM algorithms have difficulties in measuring the impact of input feature sequences on the target feature. In contrast, attention mechanisms can dynamically allocate weights to important features, and the Long Short-Term Memory network with attention mechanism (LSTM-ATTE) has shown significant effectiveness in predicting university admission score thresholds. Taking into account various factors, the SVM can transform non-linear problems in high-dimensional space into linear problems, and the LSTM-ATTE can address the problem of long-term dependencies in time series data. This paper proposes a composite model that combines the SVM model with the LSTM-ATTE method to construct a university admission score prediction model and uses grid search to find the optimal parameter combination for LSTM-ATTE. Experiments show that the proposed composite model has an accuracy rate that is up to 13% higher than other models, with an error range of 3 points.

**Key words:** LSTM; SVM; attention mechanisms; grid search

## 0 引言

高校录取分数线是衡量学生是否能够进入高校的重要标准,通过分析历年的高校录取分数线数据,不仅可以为教育政策制定者和高校提供决策依据,还可以为考生报考院校提供重要的数据支撑。高校录取分数线作为学生选拔的重要依据,对学生的发展和未来职业道路具有重要影响。

高校录取分数线预测是指在知晓高校的最低投档分数线、平均分、最低投档位次、录取人数等信息

的基础上,对该高校的录取分数进行预测,精准的高校录取分数线是为考生进行院校推荐的基础。由于每年高考试题难度和政策都可能发生变化,对于依赖历史数据的预测模型构成了挑战。新的政策和难度调整可能会导致录取分数线的不确定性,使预测模型的准确性受到一定局限,并且高中毕业生人数的波动,也会导致不同年份的竞争激烈程度发生变化,从而影响录取分数线的预测。与此同时,投档位次的波动范围也是难以把握的,因此仅考虑单一因素进行高校分数线的预测,其结果往往是预测精度

**基金项目:** 贵州省科技计划项目(黔科合基础[2018]1082,黔科合基础[2019]1159); 贵州省教育厅自然科学研究项目(黔教技[2022]015,黔教技[2022]047,黔教技[2023]012,黔教技[2023]061,黔教技[2023]062)。

**作者简介:** 秦信芳(1996-),女,硕士研究生,主要研究方向:海量数据统计与分析。

**通讯作者:** 魏嘉银(1986-),男,博士,副教授,硕士生导师,CCF 会员,主要研究方向:算法设计与分析。Email:weijiayin05@sina.com

收稿日期: 2023-11-10

不高且适用性差。

高校录取分数线的预测是一个复杂的问题,受到多种动态因素的影响。为了提高预测准确性,需要不断更新模型,及时调整参数,以适应不断变化的环境和政策。综合考虑多种数据源和因素,提出将SVM模型与LSTM方法相结合的组合模型来构建高校录取分数线预测模型。针对上述研究问题,采用2017-2023年各高校在贵州的投档数据,对高校录取分数线的预测方法进行了深入研究。

## 1 相关工作

基于分数线的预测一般是依据往年各校的投档分数,通过分数线与年份的对应关系,以此确定两者之间的相互依赖关系。近年来,诸多学者在高考录取分数线方面做了大量研究。

李诚等<sup>[1]</sup>考虑到高校录取分数线受到的影响因素较多且准确率不高的情况,使用支持向量机(SVM)理论,并结合高考录取模式,建立了针对高校录取分数线预测的SVM模型,通过实验结果证明了该预测方法的预测效果较为理想,其预测的平均误差为7.6分,验证了SVM在高校录取分数线预测中的可行性。为了提高分数线预测模型的预测精度,郭孝文等<sup>[2]</sup>提出了一种符合分数线变化规律的反向传播神经网络分数线预测模型,并对该模型进行了改进。结果表明,该模型能有效提高分数线预测精度。王泽卿等<sup>[3]</sup>采用BP神经网络进行各高校最低投档位次以及最低投档分数线的预测,并在此基础上运用多特征方法计算院校的录取概率。通过实验验证,证明了基于BP神经网络的高校录取分数预测算法在不同误差线下表现优于其他算法,不仅在平均录取率方面取得提升,而且在平均位次上也取得显著改善。王宏利等<sup>[4]</sup>使用长短期记忆网络并结合线上百分位作为录取测度,构建预测高校高考录取分数模型。该模型通过实验,证明了基于LSTM方法能较为准确的预测高校录取分数的未来趋势,从而提高录取分数预测的准确度。胡如明<sup>[5]</sup>提出基于注意力机制的LSTM网络模型(LSTM-ATTE),该模型利用注意力机制,在输入时间序列之间建立长距离的依赖关系,能够动态生成不同连接的权重。经实验证明,在高校高考分数预测方面,LSTM-ATTE模型的预测结果与实际值之间分数在5分之内的准确度能够达到93.56%。

## 2 问题描述

由于高校录取分数受到许多因素影响,只有充

分考虑历年分数、排名、省控线、招生人数等因素以及相互之间的非线性影响<sup>[4]</sup>,才能更为精确地预测高校高考分数线。为此,在考虑上述因素后,通过融合多个特征以构建以下几个相关特征:最低分线上百分位、最高分线上百分位、平均分线上百分位、最低位次线上百分位、高校最低位次对应参加高考人数标准差。

### 2.1 线上百分位

线上百分位作为衡量高校录取分波动的指标,具备有效解决因不同考试难度引起分数波动的优势。设 $s = p(r)$ 为将排名反向映射为分数的函数,其中 $r$ 为院校最低位次, $s$ 为院校最低录取分数线,则线上百分位的计算公式如下:

$$a_l = \begin{cases} \frac{p(l_1) - p(l)}{p(l_1)} \times 100\%, & l > l_1 \\ \frac{p(l_2) - p(l)}{p(l_2) - p(l_1)} \times 100\%, & l_1 > l > l_2 \end{cases} \quad (1)$$

其中, $p(\cdot)$ 表示将位次映射为分数的函数; $l_1$ 表示本科一批分数线; $l_2$ 表示本科二批分数线; $a_l$ 表示根据给定条件计算出的一个百分比值。最低分线上百分位、最高分线上百分位、平均分线上百分位、最低位次线上百分位在计算时,只要将公式(3)中的 $l$ 替换为最低分、最高分、平均分、最低位次,即可计算得到。

### 2.2 人数标准差

高校的最低投档位次反映了其在全国高考人数中的相对位置,而参加高考人数的标准差可用于度量高校录取学生的位置分布离散程度。采用标准差作为度量工具的好处在于,其有助于减小个别极端数据对整体结果的影响,确保结果的稳定性,计算公式为

$$d = \sqrt{(e_i - v)^2} \quad (2)$$

其中, $e_i$ 为各院校第 $i$ 年的最低位次在该年参加高考人数的占比; $v$ 为近 $n(n=6)$ 年内该校理科最低位次在参加高考人数占比的平均值; $d$ 为高校最低投档位次与当年参加高考人数之间差值的标准差。

## 3 SVM-LSTM-ATTE 组合预测模型

### 3.1 支持向量回归模型

支持向量机(Support Vector Machine, SVM)<sup>[6]</sup>是一种广泛用于分类和回归问题的监督学习算法,其主要目标是找到一个能够在特征空间中划分不同

类别的超平面,同时最大化分类间隔,以获得良好的泛化性能。支持向量回归 (Support Vector Regression, SVR)<sup>[7]</sup> 是 SVM 的回归方法,通过找到最好的拟合数据函数,从而预测位置数据的输出值,其目标是找到一个尽可能平滑的函数,使得函数预测的结果与真实值之间的误差最小。SVR 的基本回归模型可以表示为

$$y = f(x) = \mathbf{w}^T \boldsymbol{\varphi}(x) + b \quad (3)$$

其中,  $y$  是目标变量;  $\boldsymbol{\varphi}(x)$  是输入特征经过函数映射后的高维特征向量;  $\mathbf{w}$  是权重向量;  $b$  是偏置。

SVR 的目标是最小化模型预测与真实目标值之间的误差,同时确保在容忍误差范围内使这些误差最小化,其目标函数表示为

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + c \sum (y - f(x))^2 \quad (4)$$

其中,  $\|\mathbf{w}\|^2$  是权重质量的范数平方;  $c$  是正则化参数;  $\sum (y - f(x))^2$  是误差项和。

SVR 允许一些样本点出现超平面的偏差范围内,引入松弛变量来阻止这种偏差。松弛变量的引入允许模型承受一些噪声和丢失群点。在 SVR 中,引入松弛变量后的目标函数可以表示为:

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + c \sum (\xi + \xi^*) \quad (5)$$

$$\begin{cases} y - f(x) \leq \varepsilon + \xi \\ f(x) - y \leq \varepsilon + \xi^* \\ \varepsilon, \xi^* \geq 0 \end{cases} \quad (6)$$

其中,公式(5)引入了松弛变量  $\xi$  和  $\xi^*$ , 分别对应于样本点落在超平面上方和下方的误差;公式(6)是 SVR 的约束条件,  $\varepsilon$  是容差范围。

### 3.2 LSTM-ATTE 预测模型

#### 3.2.1 LSTM 神经网络

循环神经网络 (Recurrent Neural Network, RNN)<sup>[8]</sup> 是一类神经网络,具有内部状态或记忆,用以处理序列数据。然而,传统的 RNN 在处理长序列时,容易出现梯度消失或梯度爆炸的情况,导致难以捕捉长期依赖关系。为了克服这个问题,长短时记忆网络 (Long Short-Term Memory, LSTM)<sup>[9]</sup> 被引入。LSTM 是一种特殊的 RNN,其引入了更复杂的结构,以实现长期依赖关系的捕捉。LSTM 的独特之处在于它具有 3 种门控制机制:输入门、遗忘门和输出门,这些门允许网络更精细地控制其内部状态,从而有效地处理长序列,具体表示如下:

输入门值:

$$i_t = \sigma(W_i[x_t, h_{t-1}] + b_i) \quad (7)$$

候选值:

$$\tilde{C}_t = \tanh[W_c[x_t, h_{t-1}] + b_c] \quad (8)$$

更新记忆细胞:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, h_t = o_t \odot \tanh(C_t) \quad (9)$$

遗忘门:

$$f_t = \sigma(W_f[x_t, h_{t-1}] + b_f) \quad (10)$$

输出门:

$$o_t = \sigma(W_o[x_t, h_{t-1}] + b_o) \quad (11)$$

$$h_t = \text{LSTM}(x_t, h_{t-1}) \quad (12)$$

其中,  $x_t$  是  $t$  时刻的输入信息;  $h_{t-1}$  是上一时刻的隐藏状态;  $f_t$  为  $t$  时刻遗忘门的输出信息;  $W_f$  为遗忘门权重系数;  $b_f$  为遗忘门偏置系数;  $\sigma$  为 sigmoid 激活函数;  $W_i, W_o, W_f, W_c, b_i, b_o, b_f, b_c$  是可学习参数;  $C_{t-1}$  是上一时刻的细胞状态。

这些门的结合,使 LSTM 网络能够更好地处理长期依赖性,允许其存储和检索距离较远的信息,从而解决了传统 RNN 的短期记忆问题,LSTM 的基本模型如图 1 所示。

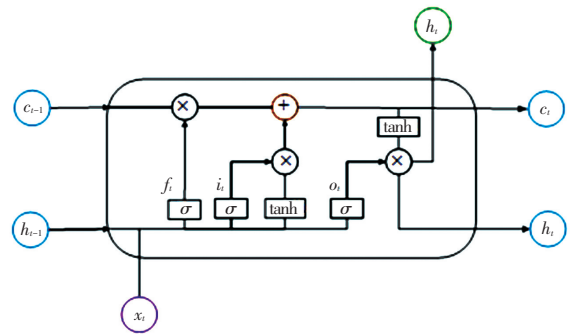


图 1 LSTM 模型结构

Fig. 1 Structure of LSTM

#### 3.2.2 LSTM-ATTE 预测模型

LSTM-ATTE 模型主要由输入层、LSTM 层、注意力机制模块、全连接层和输出层组成。在  $t$  时刻的 LSTM-ATTE 模型示意如图 2 所示,引入注意力机制对经 LSTM 层处理后的时间序列数据进行加权处理,以便更关注重要的时间步骤,再通过全连接层进行预测,其具体实现过程如下:

输入层:输入层接收原始的数据  $X = [x_{i,j}]_{N \times m}$ , 输入数据维度为  $N \times m$ ,即每个样本有  $m$  个特征。

$$X_t = [X_1, X_2, \dots, X_N] \quad (13)$$

其中,  $X_i$  表示第  $i$  所院校的  $m$  个特征,在  $t$  时刻将  $X$  输入到输入层。

LSTM 层: 设  $H = \text{LSTM}(X)$  为 LSTM 的输出,



$LSTM(\cdot)$  是 LSTM 层的计算过程。

注意力层:注意力机制的作用是对 LSTM 层的输出进行加权处理,以便更加关注输入序列中的重要部分。注意力机制通过计算每个时间步的注意力权重,将重要的时间步加权求和,从而生成一个上下文向量  $O$ ,该向量将用于后续层的输入,以提取更有信息量的特征。注意力机制可以帮助模型更好地理解输入序列中的关键信息,提高模型的性能和泛化能力,其计算过程如式(14)所示:

$$\begin{cases} h_t = LSTM(X_t, h_{t-1}) \\ a_t = \text{softmax}(W_a * h_t + b_a) \\ o = \sum a_t * h_t \\ y = \tanh(W_y * [o, h_t] + b_y) \end{cases} \quad (14)$$

其中,  $X_t$  是  $t$  时刻的输入特征;  $a_t$  是注意力权重;  $W_a$  和  $b_a$  是全连接层的权重和偏置;  $*$  是矩阵乘法。

输出层:LSTM-ATTE 层的输出是通过注意力机制,对输入序列的隐藏状态进行加权求和得到的上下文向量  $O$ ,并将上下文向量与最后一个隐藏状态  $h_t$  进行拼接,然后通过全连接层进行处理得到预测结果。最后的输出可以通过式(15)表示:

$$y = [y_1, y_2, \dots, y_{t-1}, y_t] \quad (15)$$

其中,  $y$  是模型的最终输出,表示对整个输入序列的预测结果,  $y_t$  是模型对输入序列  $X_t$  的预测结果。

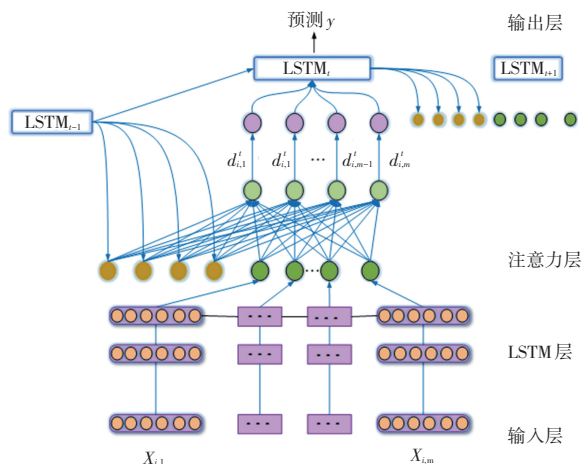


图2 LSTM-ATT 模型结构

Fig. 2 Structure of LSTM-ATT module

### 3.3 基于 SVM-LSTM-ATTE 组合预测模型

SVR 模型可以处理非线性的时间序列问题,其可以使用核函数将数据映射到高维特征空间中,从而可以更好地捕捉数据之间的非线性关系,并且由于 SVR 是基于支持向量的方法,其只依赖于少数支

持向量样本,因此对于离群值的影响较小。但是,在遇到预测方面的问题时,SVR 难以解释预测结果,无法衡量特征变量对目标变量的影响程度。

结合了注意力机制的 LSTM 模型可以学习和关注输入特征中的重要部分,  $\text{Attention}^{[10]}$  可以提供对不同时间步的不同权重,从而更加集中地处理序列中的关键信息。但是, LSTM-ATTE 模型的性能高度依赖各种参数的选择和调优,如 LSTM 单元的数量、注意力机制的类型和超参数等。

基于上述问题,本文结合 SVR 和 LSTM-ATTE 模型组成 SVM-LSTM-ATTE 模型,其高校录取分数线预测的组合模型框架描述如下:

首先,初步构建 LSTM-ATTE 时使用创建 GridSearchCV<sup>[11]</sup> 对象,对 LSTM-ATTE 神经元个数、dropout 丢弃率、全连接层的神经元个数等超参数进行搜索,GridSearchCV 执行交叉验证,以选择产生最小均方误差(MSE)的超参数组合,从而确定 LSTM-ATTE 模型的最佳超参数组合。

然后使用 SVR 模型对输入的数据训练 SVR 模型,并通过不断调整 SVR 的超参数以获得最佳性能。最佳的 LSTM-ATTE 模型和 SVR 模型分别训练,并用于预测测试集的目标值。

最后,组合模型将两种模型的预测结果结合在一起,该模型使用两个模型的预测结果作为输入,并采用线性支持向量机回归,以最小化均方误差来优化预测,从而产生更准确的最终预测。

## 4 实验结果分析

### 4.1 实验数据

实验所用数据均来自贵州省招生考试院,选择 2017-2023 年全国各高校在贵州的投档数据,每条数据包含了每所院校对应的年份以及该院校的最低投档分数、最低投档位次、最高投档分数、投档数、平均投档分数、线差、录取人数、录取率等特征。

### 4.2 评价指标

#### 4.2.1 平均绝对误差

平均绝对误差 (Mean Absolute Error, MAE) 是衡量模型预测结果准确性的指标之一,平均绝对误差越小,则模型预测的准确度越高。计算公式为

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (16)$$

#### 4.2.2 均方根误差

均方根误差 (Root Mean Square Error, RMSE)

能够准确地反映预测值与实际测量值之间的绝对误差偏差水平,数值越小表示模型的预测精度越高。计算公式为

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (17)$$

#### 4.2.3 R 平方

$R^2$  ( $R$ -squared,  $R^2$ ) 用于评估回归模型性能的统计指标,其提供了有关模型对数据拟合程度的信息。当  $R^2$  接近 1 时,表示回归模型能够很好地解释目标变量的方差,即模型所预测的值与实际观测值非常接近,意味着模型对数据的拟合程度非常高,大部分方差都能够由模型解释。计算公式为

$$R^2 = \frac{\sum_{i=1}^n w_i (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n w_i (y_i - \bar{y})^2} \quad (18)$$

#### 4.2.4 准确率

准确率 (*Accuracy*) 是用来衡量预测结果与真实结果差异的指标,其值越接近 1 表明预测结果越准确。计算公式为

$$Accuracy(\leq e) = \frac{(|y_i - \hat{y}_i| < e)}{|y_i - \hat{y}_i|} \quad (19)$$

### 4.3 结果分析

为比较所提 SVM-LSTM-ATTE 模型的有效性,实验选取 SVR 模型、BP 模型、LSTM 模型、LSTM-ATTE 模型来验证所提模型的有效性,并采用 *MAE*、*RMSE*、 $R^2$  和 *Accuracy* 作为模型性能的评价指标,分别选取 3 个区间分值段进行验证,如图 3 所示。为保证模型的可比性,实验均在自建数据集上进行模拟,各算法中的 LSTM 部分网络的拓扑结构、优化器、迭代次数均保持一致,各模型预测结果见表 1。

由表 1 可知,其他模型与本文提出的 SVM-LSTM-ATTE 模型的  $R^2$  相差不大,其结果都接近于 1,表明 5 个模型较好地拟合了数据信息。

误差范围在 3 分以内的性能评估中,SVM-LSTM-ATTE 模型在平均绝对误差和均方根误差的结果优于另外 4 个模型,其准确率为 0.70,表明该模型在三分以内的预测值和真实值之间的误差较小。BP 模型在所有评估指标下的准确性相对较低,仅为 0.57,LSTM 和 LSTM-ATTE 预测准确率较为接近,而 SVM-LSTM-ATTE 模型在 3 分以内的准确率评估中表现最好,相比于最差的 BP,准确率提升了

0.13。

表 1 各模型预测结果

Table 1 Prediction results of different models

Model	Index	≤3	≤5	≤10
SVR	<i>MAE</i>	1.18	1.49	2.13
	<i>RMSE</i>	1.36	1.95	3.04
	$R^2$	0.99	0.998	0.997
BP	<i>Accuracy</i>	0.59	0.79	0.93
	<i>MAE</i>	1.33	1.87	2.50
	<i>RMSE</i>	1.38	2.89	3.28
LSTM	$R^2$	0.99	0.998 2	0.997
	<i>Accuracy</i>	0.57	0.79	0.94
	<i>MAE</i>	1.48	1.81	1.98
LSTM-ATT	<i>RMSE</i>	1.70	2.13	3.12
	$R^2$	0.998	0.997	0.995
	<i>Accuracy</i>	0.66	0.77	0.94
SVM-LSTM-ATT	<i>MAE</i>	1.43	1.63	3.09
	<i>RMSE</i>	1.31	2.11	3.10
	$R^2$	0.997	0.996	0.990
LSTM-ATTE	<i>Accuracy</i>	0.68	0.86	0.95
	<i>MAE</i>	1.54	1.56	2.11
	<i>RMSE</i>	1.38	1.92	3.02
SVM-LSTM-ATTE	$R^2$	0.999	0.998 7	0.998 1
	<i>Accuracy</i>	<b>0.70</b>	<b>0.88</b>	<b>0.96</b>

在 5 分以内的准确性评估中,SVR、BP 和 LSTM 的准确率都低于 0.8,只有 SVM-LSTM-ATTE 和 LSTM-ATT 模型预测准确率表现最好,分别达到了 0.86 和 0.88,最高的 SVM-LSTM-ATTE 相比于最低的 LSTM,其误差在 5 分以内的准确率提高了 0.11。

误差在 10 分以内的评估中,5 个模型的准确率相差不大,都超过了 0.9。其中 SVM-LSTM-ATTE 模型的准确率最高,为 0.96,最低的 SVR 模型为 0.93。两者相比,SVM-LSTM-ATTE 模型提高了 0.03。

综上所述,SVR 在处理大规模数据集时可能面临计算复杂度和内存消耗的问题,而 BP 神经网络虽然具备一定的通用性,但在处理序列数据方面可能不如 LSTM、LSTM-ATTE 和 SVM-LSTM-ATTE 模型有效。

即 SVM-LSTM-ATTE 模型在所有评估指标下表现最佳,通过预测结果可知,SVM-LSTM-ATTE 模型在分数线预测方面具有一定的可靠性。

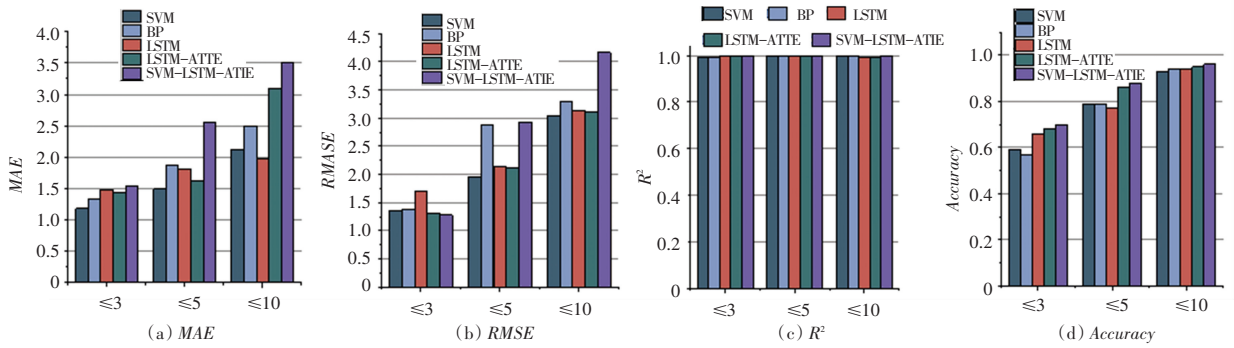


图3 模型评估结果

Fig. 3 Evaluation results

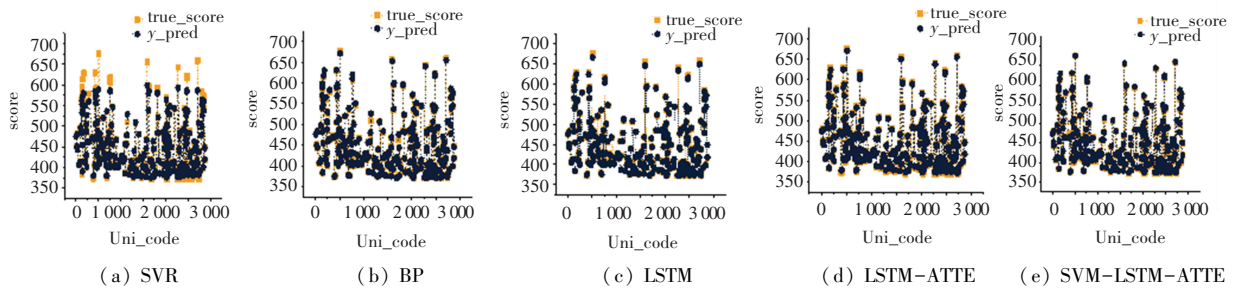


图4 模型拟合效果

Fig. 4 Model fitting effect

选取2023年在贵州投档的200所院校进行验证,效果如图4所示。可以看出,LSTM-ATT模型的预测值与真实值之间的拟合效果最好。总而言之,SVM-LSTM-ATTE将SVM和LSTM-ATTE结合起来,利用SVM的线性特性和LSTM-ATTE的非线性建模能力,可以更好地处理具有复杂关系的数据,继承了SVM的鲁棒性和泛化能力,从而提高模型的整体性能以及模型的预测准确率。

## 5 结束语

针对高校录取分数线精准预测问题,本文提出一种融合多特征的录取分数线预测模型,该模型将注意力机制应用于LSTM模型中,能够有效捕捉序列之间的交互信息,从而提升LSTM预测时间序列数据的准确性,并提出将SVM和LSTM-ATTE组合构成SVM-LSTM-ATTE组合预测模型,该模型既保留了SVR处理小样本数据的优点,又解决了对序列数据权重分配不均的问题。实验结果证明,SVM-LSTM-ATTE模型在误差为3分的范围内的准确率最高提升了0.13,在误差为5分的范围内提高了0.11,在误差为10分的范围内提升了0.03,证明了SVM-LSTM-ATTE在高校录取分数线预测方面的有效性。

## 参考文献

- [1] 李诚,张宏烈,王崇. 基于SVM的高校录取分数预测模型[J]. 高师理科学刊,2016,36(12):22-24.
- [2] 郭孝文,梁向阳. 改进的BP神经网络在分数线预测中的应用[J]. 西安工业大学学报,2018,38(3):286-292.
- [3] 王泽卿,季圣鹏,李鑫,等. 基于分数线预测的多特征融合高考志愿推荐算法[J]. 计算机科学,2022,49(S2):254-260.
- [4] 王宏利,边帅,孙全亮,等. 基于LSTM的高校高考录取分数预测研究[J]. 经济师,2020(2):160-162.
- [5] 胡如明. 基于深度学习的高考分数预测模型与算法研究[D]. 武汉:武汉工程大学,2023.
- [6] 丁世飞,齐丙娟,谭红艳. 支持向量机理论与算法研究综述[J]. 电子科技大学学报,2011,40(1):2-10.
- [7] SMOLA A J, SCHÖLKOPF B. A tutorial on support vector regression[J]. Statistics and Computing, 2004, 14: 199-222.
- [8] YU Y, SI X, HU C, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. Neural Computation, 2019, 31(7): 1235-1270.
- [9] SMAGULOVA K, JAMES A P. A survey on LSTM memristive neural network architectures and applications[J]. The European Physical Journal Special Topics, 2019, 228(10): 2313-2324.
- [10] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017,30:5998-6008.
- [11] AHMAD G N, FATIMA H, ULLAH S, et al. Efficient medical diagnosis of human heart diseases using machine learning techniques with and without grid search CV[J]. IEEE Access, 2022, 10: 80151-80173.