基于SARIMA-LSTM的门诊量预测研究

卢鹏飞[1], 须成杰[2], 张敬谊[1], 韩侣[3], 李静[1]
1. 万达信息股份有限公司, 上海 201112; 2. 复旦大学附属妇产科医院, 上海 200090;
3. 长春理工大学, 吉林 长春 130022

摘要
为了实现更加稳健和精准的门诊量预测，构建了一种基于SARIMA-LSTM的门诊量预测模型。该方法首先使用SARIMA模型对门诊量进行单指标建模，提取门诊量指标蕴含的周期、趋势等信息，然后构建了以节日天数、法定上班天数、平均最高气温等多个相关指标为输入的多对一LSTM模型，对SARIMA模型残差进行进一步学习，实现残差与多个变量间的非线性关系抽取。实证结果表明，构建SARIMA-LSTM混合模型相较5种主流预测方法具有更高的一步预测精度，具有较好的实际应用价值。

关键词
季节性差分自回归滑动平均模型；长短期记忆网络；门诊预测；残差

中图分类号：TP183 文献标识码：A doi: 10.11959/j.issn.2096-0271.2019053

Research on the prediction of outpatient volume based on SARIMA-LSTM

LU Pengfei[1], XU Chengjie[2], ZHANG Jingyi[1], HAN Lyu[3], LI Jing[1]
1. Wonders Information Co., Ltd., Shanghai 201112, China
2. Gynecology Hospital of Fudan University, Shanghai 200090, China
3. Changchun University of Science and Technology, Changchun 130022, China

Abstract
In order to achieve more robust and accurate outpatient volume prediction, a hybrid prediction model based on SARIMA-LSTM was constructed. SARIMA model was used to build a single index model of outpatient volume to extract the cycle, trend and other information contained in outpatient volume index. Then multiple related indexes, including holiday days, legal working days, average maximum temperature, were used as input of a many-to-one LSTM model, in order to further learn the residual of SARIMA model and extract the nonlinear relationship between residual and multiple variables. The empirical results show that the SARIMA-LSTM hybrid model constructed in this paper has higher prediction accuracy than the five mainstream prediction methods, so it has good practical application value.

Key words
seasonal auto-regressive integrated moving average model, long short term memory, outpatient forecast, residual

1 引言

门诊量是医疗机构服务能力的体现，能够直接反映医疗机构当前的运行状况，并为医院经营和资源分配提供重要参考。精准的门诊量预测能够帮助医院管理者对医疗资源进行合理部署和分配，从而保障医院的运营效率。门诊量的波动受多种因素影响，但是门诊时间序列通常存在一定的趋势性和周期性，这为门诊量预测提供了基础。

当前，一些研究者对门诊量预测方法进行了较为深入的研究，传统的时间序列模型包括GM(1,1)、Holt-Winters、自回归滑动平均（auto-regressive moving average，ARMA）模型、季节性差分自回归滑动平均（seasonal auto-regressive integrated moving average，SARIMA）模型等[1-4]。尽管利用ARMA等传统时间序列模型进行门诊量预测是较适宜的方法，但是仍存在一定的限制。导致门诊量波动的影响因素较多，仅依赖门诊量序列本身进行建模无法刻画和解释某些时间段的变化规律。随着人工智能技术的不断发展，研究人员逐渐将机器学习中的监督回归方法应用于门诊量时间序列的预测。例如，李琳等[5]对新疆地区慢性阻塞性肺病的月门诊量构建了LSTM模型并进行1步和12步预测；黄代政等[6]应用三层反向传播神经网络对医院门诊量的多个分量进行了实例验证；桑发文等[7]应用相似日和极限学习机的方法对医院门诊量进行了短期预测探究；张云丽等[8]构造了一种基于灰色预测和RBF神经网络相结合的方法，对门诊量进行建模和预测。相较于传统时间序列模型，上述模型能够对多个输入变量进行特征自动学习，从而获取更多的可用信息。但是，考虑到医疗时间序列指标样本量通常很小，单纯使用机器学习模型对门诊量进行预测的思路也存在一些弊端，如容易过拟合、模型鲁棒性较差等。

为了弥补传统时间序列模型对门诊信息提取不充分的缺陷，本文结合机器学习模型的建模优势，将节日天数、法定上班天数、平均最高气温、平均温差、降雨天数以及挂号人数等相关因素作为模型输入，构建具有时间序列记忆性的长短期记忆网络，对SARIMA模型的残差进行学习，形成SARIMA-LSTM混合模型，更加充分地提取残差信息，从而大幅提升预测精度。

2 理论基础和方法

2.1 SARIMA模型

ARMA模型对平稳序列数据具有良好的建模效果，对于非平稳时间序列，则需要进行d阶差分，形成ARIMA(p,d,q)模型。该模型也称为差分自回归滑动平均模型，其中d为差分项，p、q为延迟参数。然而，对于一些既有季节效应又有长期趋势的时间序列，简单的自回归差分滑动平均（auto-regressive integrated moving average，ARIMA）模型不足以提取其中的季节信息，这时通常需要采用SARIMA模型。综合d阶差分和以s为步长的季节差分运算，SARIMA乘积模型完整结构如下：

$$\nabla_D \nabla^d x_t = \frac{\Theta(B)\Theta_S(B)}{\Phi(B)\Phi_S(B)}\varepsilon_t \tag{1}$$

其中：

$$\begin{aligned}
&\Theta(B) = 1-\theta_1 B-\cdots-\theta_{q_1}B^{q_1}\\
&\Phi(B) = 1-\varphi_1 B-\cdots-\varphi_{p_1}B^{p_1}\\
&\Theta_S(B) = 1-\theta_1 B^s-\cdots-\theta_{q_2}B^{q_2 s}\\
&\Phi_S(B) = 1-\varphi_1 B-\cdots-\varphi_{p_2}B^{p_2 s}
\end{aligned} \tag{2}$$

该乘积模型简记为SARIMA$(p_1,d_1,q_1)\times(p_2,d_2,q_2,s)$。

2.2 LSTM模型

长短时记忆(long short term memory, LSTM)网络是一种结合梯度学习算法的网络，是对循环神经网络(recurrent neural network, RNN)的改进。LSTM通过其独特的记忆模式和遗忘模式使网络可以充分挖掘数据的时间序列特征，学习信号中的时间依赖关系，缓解了传统RNN在训练过程中容易出现的梯度消失、爆炸现象，在众多领域取得了巨大的成功[9]。LSTM与传统RNN的不同点在于LSTM神经元多了3种“门”和1个记忆单元，LSTM结构如**图1**所示。LSTM神经元通过“门”结构有选择性地传递信息，从而达到控制信息的目的。与RNN相同，LSTM神经网络使用反向传播算法进行模型训练。在反向传播进行参数更新时，LSTM误差项的反向传播包括两个方向：一个是沿着时间的方向反向传播，即从当前时刻开始，计算每个时刻的误差项；另一个是将误差项向上一层传播，根据相应的误差项，计算每个权重的梯度。

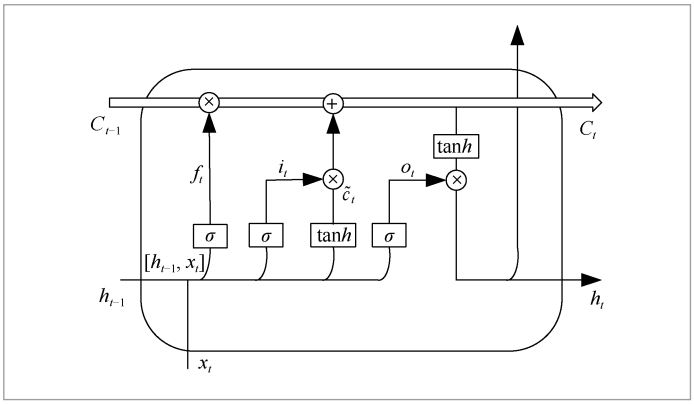


图1 LSTM结构示意

2.3 基于SARIMA-LSTM的门诊量预测模型

在实际应用中，使用SARIMA模型进行单指标建模能够较好地捕获门诊量的历史变化规律，但是仍存在较大的局限性，SARIMA模型无法解释门诊波动的内因，当波动规律较为复杂时，SARIMA模型无法对其进行刻画，因此会有较大的预测偏差。例如，门诊量的变化在某些月份受节假日影响较大，而某些重大节假日(如春节)的时间在月份上不固定，模型利用历史数据抓取的规律在预测某些特殊月份时存在偏差。此外，天气信息(如气温、昼夜温差、降雨时间等)在一定程度上也会影响医院就诊的患者数量，特殊天气造成的门诊量波动效应也无法通过简单的SARIMA模型进行有效解释。在SARIMA模型中，模型无法解释的信息会被归为残差项，为了更加精准地抽取门诊量波动信息，本文构建了LSTM模型，对SARIMA模型的残差指标进行建模，形成SARIMA-LSTM混合模型。在SARIMA-LSTM模型中，LSTM神经网络的输入层不仅包含SARIMA模型的残差指标，还需纳入对门诊量波动有重要影响的协变量，如节假日信息、天气信息以及其他医院综合指标信息。

使用SARIMA-LSTM模型进行门诊量预测的主要步骤如下。

(1)对门诊量数据进行SARIMA建模，使用历史数据$y_i(i=1,\cdots,n)$进行模型参数优化、拟合和检验，根据所得模型计算历史数据估计值$z_i(i=1,\cdots,n)$、历史数据残差$e_i=y_i-z_i(i=1,\cdots,n)$、模型一步预测值$\hat{z}_{n+1}$。

(2)将SARIMA建模残差e_i及多个协变量的一阶滞后项作为模型的自变量，

SARIMA建模残差e_i作为模型的因变量，将其变换为LSTM神经网络的输入值，并进行归一化处理，设定时间窗口n_w，构建多对一的LSTM神经网络一步预测模型，通过模型训练，计算模型一步预测值$\hat{e}_{n+1}$。

（3）计算SARIMI-LSTM模型的一步预测值$\hat{y}_{n+1}=\hat{z}_{n+1}+\hat{e}_{n+1}$。

2.4 模型评估方法

选用均方根误差（root mean squared error，RMSE）和平均绝对百分比误差（mean absolute percentage error，MAPE）这两个指标来评价算法的时间序列预测性能，均方根误差和平均绝对百分比误差的计算方法分别为：

$$\mathrm{RMSE}=\sqrt{\frac{1}{M}\sum_{t=1}^{M}(y_t-\hat{y}_t)^2} \quad (3)$$

$$\mathrm{MAPE}=\frac{1}{M}\sum_{t=1}^{M}\left|\frac{y_t-\hat{y}_t}{y_t}\right| \quad (4)$$

其中，y_t为实际观测值，$\hat{y}_t$为预测值，M表示测试样本个数。

3 实例验证与分析

3.1 数据来源及预处理

医疗指标数据来源于上海某三甲医院，时间范围为2011年1月1日至2018年12月31日，包括门诊量、挂号量2个指标，数据质量良好，无缺失值。节日信息采集于万年日历网站的公开历史信息，包括上班信息和节假日信息。天气信息采集于中国天气网的公开历史信息，包括每一天的最高气温、最低气温和降雨情况。针对门诊量和挂号量，使用累加的方式将其转换为月度数据，即门诊量月度指标和挂号量月度指标；针对上班信息，统计每月法定上班天数，得到上班天数月度指标；针对节假日信息，仅考虑春节、劳动节和国庆节信息，统计春节7天假、劳动节3天假、国庆节7天假在每个月出现的天数，得到节日天数月度指标；针对天气信息，计算每月的平均最高气温月度指标、平均温差月度指标以及降雨天数月度指标。其中，上班天数月度指标和节日天数月度指标信息可预知，因此整体向前移动一个月，例如，时间索引2017年3月对应的上班天数是2017年4月的值，依此类推。通过该方式，LSTM网络在训练过程中能够捕获到所需预测月份的上班和节日信息对该月预测指标的影响，即额外地获取预测当月的上班和节日的信息，从而提升预测精度。

3.2 SARIMA模型构建

3.2.1 门诊量季节性分析

STL（seasonal and trend decomposition using loess）是一种用于分解时间序列的通用且稳健的方法，其中Loess是一种估算非线性关系的方法[10]。针对门诊序列数据，采用STL方法对其进行分解，得到季节效应、趋势效应和剩余波动效应的分解值，即$x_t=S_t+T_t+I_t$，结果如**图2**所示。从**图2（a）**中可见，门诊量存在较为明显的趋势性，在2013年之后呈现上升的态势，因此可能为非平稳序列，需要进行差分运算；从**图2（b）**中可见，门诊量存在稳定的周期性，周期为12个月，根据周期波动规律发现，一个周期内12月医院门诊就诊量最大，2月门诊就诊量最小，12月至3月期间门诊量波动最大。

3.2.2 模型参数搜索

门诊量数据既有季节性成分又有非季节性成分，因此用混合效应的乘积模型

SARIMA$(p_1,d_1,q_1)\times(p_2,d_2,q_2,s)$进行建模，从**图2(b)**中可知季节性参数$s$为12，参数$p,d,q$采用网格搜索法并结合赤池信息量准则(akaike information criterion，AIC)进行确定。网格搜索中参数的搜索范围分别设定为$p_1\in[0,4]$，$d_1\in[0,1]$，$q_1\in[0,5]$，$p_2\in[0,2]$，$d_2\in[0,1]$，$q_2\in[0,2]$。由于参数较多，在参数搜索过程中分两阶段进行搜索，即先构建ARIMA(p_1,d_1,q_1)获得最小AIC值对应的参数$\hat{p}_1,\hat{d}_1,\hat{q}_1$，然后利用最优ARIMA模型参数构建SARIMA$(\hat{p}_1,\hat{d}_1,\hat{q}_1)\times(p_2,d_2,q_2,s)$模型，搜索得到最优的$\hat{p}_2,\hat{d}_2,\hat{q}_2$参数值。在搜索过程中，设定模型必须满足平稳条件和可逆条件，确保模型为平稳可逆模型，若不满足任一条件则自动跳过该参数组合。同时，为了进行下一步模型验证，将2011年1月至2017年12月的月度数据作为模型输入，采用最大似然估计进行模型拟合。两阶段搜索得到的最优参数组合见**表1**，可见最优的参数$(\hat{p}_1,\hat{d}_1,\hat{q}_1)$和$(\hat{p}_2,\hat{d}_2,\hat{q}_2)$分别为(3,1,2)以及(1,1,1)，因此得到的最优模型参数组合为SARIMA$(3,1,2)\times(1,1,1,12)$。同时，参数搜索结果表明，在ARIMA模型加入季节参数之后，模型得到明显优化，AIC值大幅减小。

3.2.3 模型检验

在得到最优的参数组合之后，需要进行模型显著性检验和参数显著性检验。模型的显著性检验主要是检验模型的残差项是否为白噪声序列。SARIMA$(3,1,2)\times(1,1,1,12)$模型的残差信息如**图3**所示，结合**图3(b)**和**图3(c)**，可判断残差接近标准正态分布；**图3(d)**表明残差序列存在较低的自相关，即不存在明显的季节性。此外，使用LBQ(Ljung-Box Q)检验统计量对残差进行白噪声检验，显示1-24阶延迟下的LB统计量的P值均显著大于0.05，因此可以认为该拟合模型的残差序列属于白噪声序列，即该拟合模型显著有效。参数的显著性检验是指检验每一个模型参数是否显著非零，可起到模型精简的作用。模型的参数及其显著性检验信息见**表2**，其中，检验方法为z检验，z为检验值。结果表明除了AR一阶系数不显著非零外，其他系数均通过检验，因此从模型中删除AR一阶滞后自变量，得到最精简拟合模型。

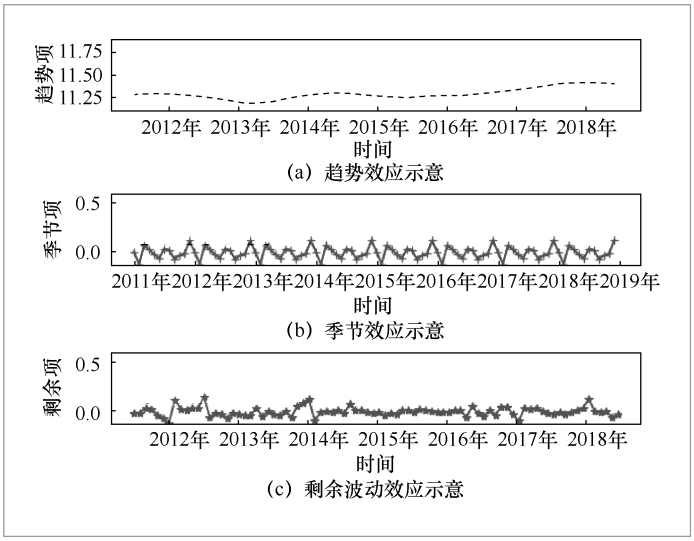


图2 门诊量加法模型分解示意

表1 参数搜索结果

参数	p	d	q	AIC值
第一阶段	3	1	2	1 730.97
第二阶段	1	1	1	1 453.71

3.3 SARIMA-LSTM模型门诊量预测

3.3.1 建模流程与模型结构

在SARIMA-LSTM模型中，首先使用门诊量进行SARIMA建模，流程与上文一致。然后，将SARIMA模型的残差指标、门诊量、挂号量、上班天数、节日天数、平均

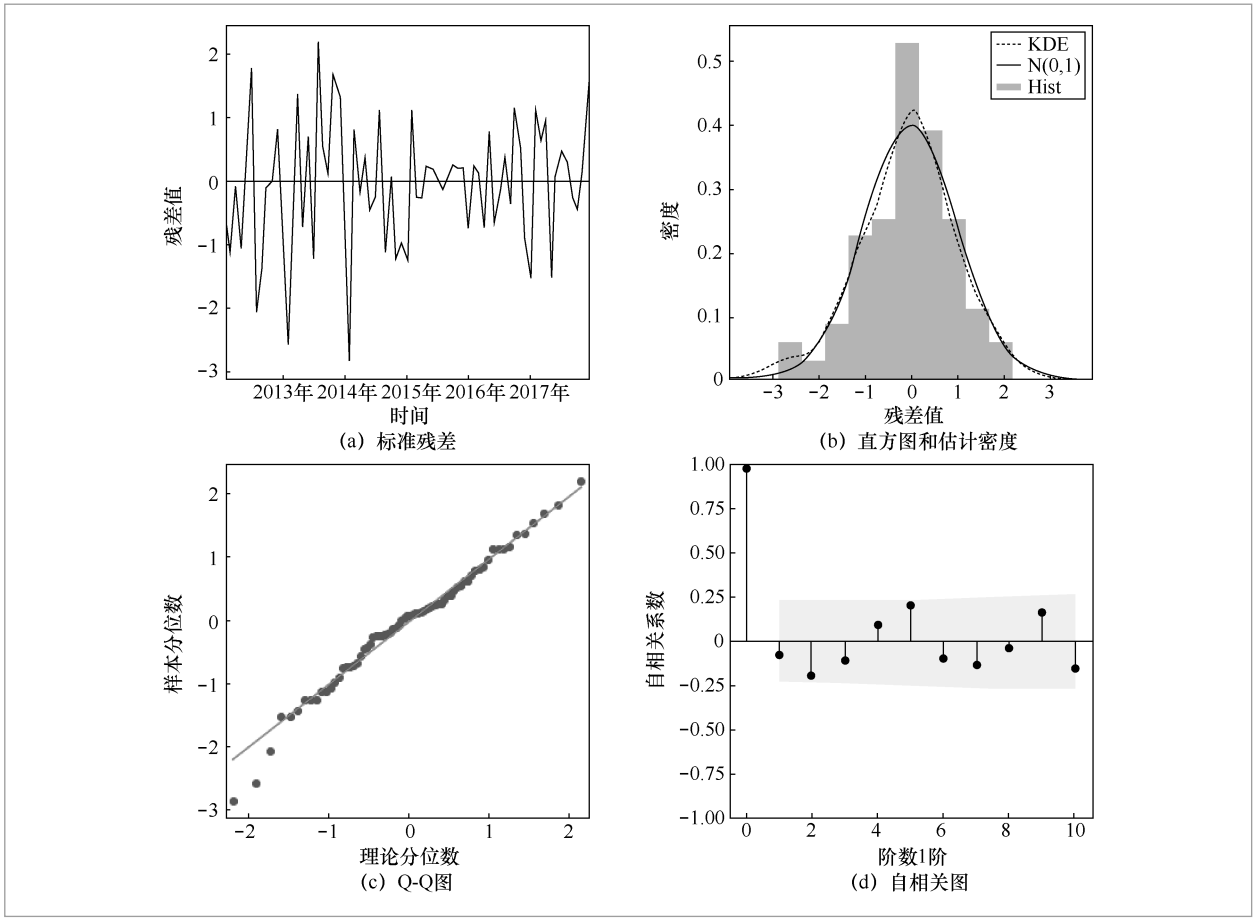


图 3 模型残差信息

表 2 模型参数与参数显著性检验信息

参数	系数值	标准误	*Z*值	*P*值	95%置信下限	95%置信上限
ar.L1	0.020 5	0.107	0.19	0.849	−0.19	0.231
ar.L2	−0.932 2	0.036	−26.014	0	−1.002	−0.862
ar.L3	−0.223	0.106	−2.101	0.036	−0.431	−0.015
ma.L1	−0.275 2	0.046	−5.929	0	−0.366	−0.184
ma.L2	0.989 1	0.099	10.019	0	0.796	1.183
ar.S.L12	0.466 8	0.222	2.105	0.035	0.032	0.901
ma.S.L12	−0.809 5	0.221	−3.662	0	−1.243	−0.376
sigma2	3.94×10^{7}	4.28×10^{-9}	9.19×10^{15}	0	3.94×10^{7}	3.94×10^{7}

最高气温、平均温差以及降雨天数共计8个指标作为LSTM模型的自变量，预测下一个月的SARIMA模型残差指标。最后，将SARIMA模型的一步预测值与预测残差值相加得到下一个月的门诊量预测值，具体如**图4**所示。在进行LSTM建模步骤中，需要对输入变量进行归一化，针对预测值进行反归一化得到门诊量真实值。在LSTM

模型中，包含一个输入层、一个LSTM层和一个输出层。经过多次交叉验证实验，当设定LSTM层单元个数为26，时间窗口宽度为7时，模型泛化能力较强。在模型训练过程中，采用一步滑动预测的方式进行数据转换，得到输入数据维度为77×7×8，输出数据维度为77×1；在模型测试过程中，输入数据维度为1×7×8。由于时间窗口宽度为7，即模型以8个指标的前7个月的真实数据作为自变量，预测第8个月的SARIMA模型残差指标，因此模型在预测过程中能够抓取近7个月的多指标时间序列信息，从而对下一个月进行预测。

3.3.2 SARIMA-LSTM模型训练与预测

在确定LSTM模型结构和数据集之后，设定损失函数为均方误差函数，训练方法为Adam优化算法，进行模型训练。其中，批尺寸设定为77，模型迭代次数为1 000，参数初始化方法为Xaiver方法。然后，训练模型对2018年的12个测试样本进行预测，得到SARIMA模型的12个残差预测值。最后，与SARIMA模型的12个一步预测值累加，得到SARIMA-LSTM模型预测值，如**图5**所示。由**图5**可知，和SARIMA模型预测值相比，该模型的预测值更加准确，尤其在SARIMA模型预测最不准确的2月份，预测值得到大幅调整，与真实值更加接近。同时，调整后的置信区间覆盖了所有月份的真实值，整体变化趋势与真实值波动规律更相符。

3.3.3 模型评估和对比

为了对本文提出的方法进行评估，实验选取几种主流文献中的方法进行对比分析，对比方法包括Holt-Winters、ARIMA、SARIMA、样条回归和LSTM。

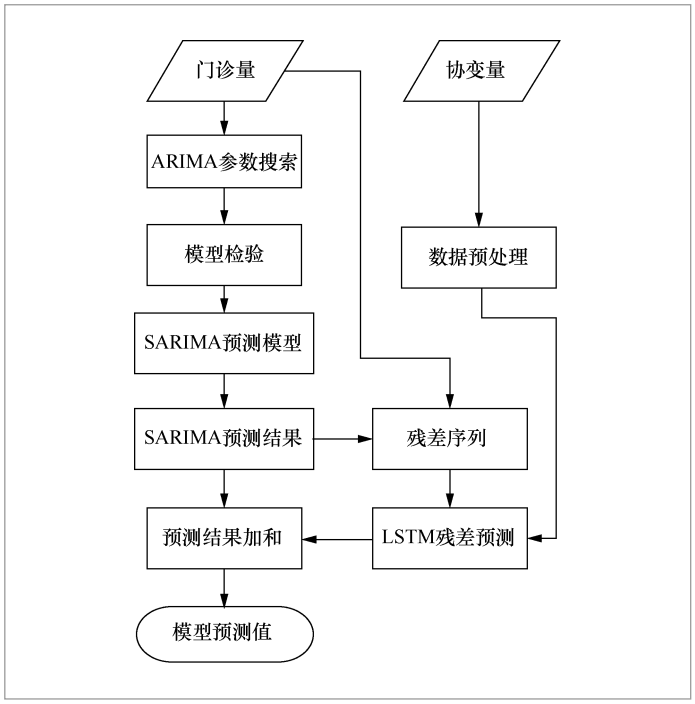


图 4 SARIMA-LSTM 模型整体建模流程

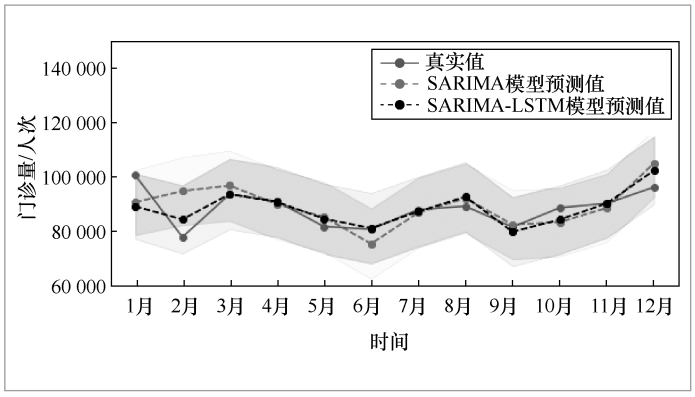


图 5 模型预测结果对比

Holt-Winters方法为3次加法模型，并通过粒子群算法以误差平方和为目标函数，优化了模型初始参数；ARIMA和SARIMA方法与上文建模步骤一致；样条回归模型经过多次逐步回归实验，确定输入变量包括节日天数、法定上班天数、截距项、第19个月节点、第22个月节点、第34个月节点、第37个月节点、第52个月节点且最优的多项

式次数为1，并构建回归预测模型；LSTM方法的输入变量包括本文的7个指标，即门诊量、挂号量、上班天数、节日天数、平均最高气温、平均温差以及降雨天数，多次实验得到最优的网络结构的LSTM层单元个数为60，时间窗宽为7。同样地，分别构建模型，对2018年门诊量进行一步预测，计算得到均方根误差和平均绝对百分比误差，多模型预测结果对比见**表3**，6种方法预测结果对比如**图6**所示。

由对比结果可知，ARIMA模型预测精度最差，该模型未考虑季节因素，因此在几乎所有月份均都有较大的偏差；SARIMA模型加入了季节效应，整体预测精度有所提升，但是对某些月份的预测偏差仍然很大；通过纳入有效的输入变量，样条回归模型在一些特殊月份具有较高的预测精度，但是在二、三季度预测结果较差，难以准确抓取波动规律，在5月、6月、9月预测偏差较大；通过参数优化的方法，Holt-Winters模型能够更好地抓取季节波动信息，由于该方法参数很少且进行了优化，因此在小样本时间序列上能够较好地避免过拟合现象，预测精度较高；LSTM模型纳入了更多输入变量，预测精度也较高，在传统时间序列模型预测结果较差的1月和2月，具有更优的表现，但是由于样本的个数少于模型参数，该模型容易过拟合，模型预测不稳定。尽管SARIMA-LSTM模型中的LSTM模型也容易过拟合，但是其预测值对模型整体预测结果影响较小，仅仅是对SARIMA模型残差的修正，故模型更加稳健。相较于其他几种主流方法，SARIMA-LSTM模型的预测精度大幅提升，RMSE和MAPE分别为4 402.69和3.51%，具有更好的时间序列预测性能。

表 3　多模型预测结果对比

方法	RMSE	MAPE
ARIMA	7 877.29	6.64%
SARIMA	6 733.72	5.67%
样条回归	5 932.29	5.47%
Holt-Winters	5 795.44	5.01%
LSTM	5 235.62	5.13%
SARIMA-LSTM	**4 402.69**	**3.51%**

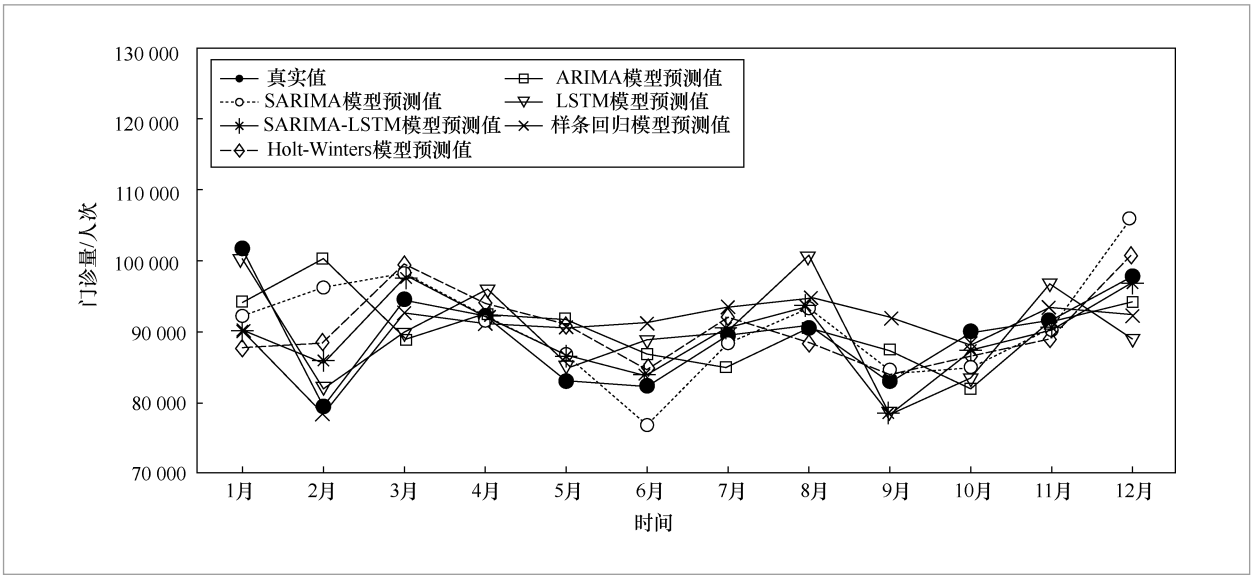


图 6　6 种方法预测结果对比

此外，在构建LSTM模型进行残差预测过程中也存在一些需注意的问题：第一，在进行多指标建模时，训练样本较少且LSTM的参数较多，容易出现过拟合现象，需要通过早停法、dropout等技巧调整模型的迭代次数；第二，SARIMA残差在早期存在异常值，LSTM训练受异常样本影响较大，主要表现为损失函数曲线在迭代过程中有较大的震荡现象，因此在模型训练之前需对异常值进行剔除；第三，当批尺寸较小时，模型训练不稳定，因此可使用所有样本计算一次梯度，同时适当减小LSTM层的单元个数，能够使损失函数曲线更加光滑，易于收敛。

4 结束语

针对医院门诊量时间序列数据，本文设计了一种基于SARIMA-LSTM的预测模型，该模型的主要思想为利用SARIMA模型对门诊量进行预测，之后构建LSTM模型，预测SARIMA模型的残差。与传统的将单一门诊量指标作为输入数据不同，本文的模型输入中还添加了节假日信息、天气信息以及其他医院综合指标信息，从而更加有效地刻画了门诊时间序列变化的内因。通过与Holt-Winters 、ARIMA、SARIMA、样条回归以及LSTM这5种主流医疗指标预测方法进行实例验证对比，发现SARIMA-LSTM模型具有更好的门诊量时间序列预测性能，因此该方法对医院运营管理辅助决策具有一定的参考和应用价值。

参考文献:

[1] YIN S F, WU J H, WANG D, et al. Predicting of the outpatient based on grey model[J]. Lecture Notes in Electrical Engineering, 2014, 270: 383-388.

[2] LI X, MA C, LEI H, et al. Applications of SARIMA model in forecasting outpatient amount[J]. Chinese Medical Record English Edition, 2013, 1(3): 124-128.

[3] WANG X L, CHEN J, SHI T X, et al. Application of winters exponential additive model in outpatient visits prediction[J]. Chinese Journal of Health Informatics and Management, 2016,13(2): 214-216..

[4] LI Y M, WU F, ZHENG C, et al. Predictive analysis of outpatient volumes of a first-class grade a general hospital through ARIMA models[J]. Chinese Medical Record English Edition, 2014, 2(8): 364-367.

[5] LI L, WANG Z, ZHANG X L. The accuracy of monthly outpatient volume prediction based on LSTM deep neural network[J]. China Digital Medicine, 2019, 14(1): 14-17.

[6] HUANG D, WU Z H. Forecasting outpatient visits using empirical mode decomposition coupled with back-propagation artificial neural networks optimized by particle swarm optimization[J]. Plos One, 2017, 12(2): e0172539.

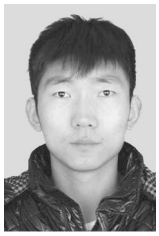
[7] SANG F W, WEI Z, CHEN H, et al. Short-term hospital outpatient amount forecasting based on similar days and extreme learning machine[J]. China Digital Medicine, 2018,13(2): 113-115.

[8] ZHANG Y L, YANG Z S. Grey RBF neural network based forecasting of outpatient capacity in modern hospital[J]. Computer Engineering & Applications, 2010, 46(29): 225-228.

[9] GERS F A, SCHMIDHUBER J A, CUMMINS F A. Learning to forget: continual prediction with LSTM[J]. Neural Computation, 2000, 12(10): 2451-2471.

[10] CLEVELAND R B, CLEVELAND W S. STL: a seasonal-trend decomposition procedure based on loess[J]. Journal of Official Statistics, 1990, 6(1): 3-33.

作者简介



卢鹏飞（1991- ），男，万达信息股份有限公司大数据产品部数据挖掘工程师，主要研究方向为医疗数据挖掘、机器学习、光谱分析。

须成杰（1983- ），男，复旦大学附属妇产科医院信息科工程师，主要研究方向为大数据及人工智能、互联网医疗+物联网、医院信息无纸化管理、医疗可信云计算等。

张敬谊（1974- ），女，博士，万达信息股份有限公司大数据产品部总经理、教授级高级工程师，主要研究方向为并行计算、智能分析、城市信息化等。

韩侣（1992- ），男，长春理工大学硕士生，主要研究方向为统计机器学习、数据挖掘。

李静（1973- ），女，万达信息股份有限公司大数据产品部资深产品经理、高级工程师，主要研究方向为医疗卫生大数据、健康医疗大数据+人工智能。

收稿日期： 2019-09-20

基金项目： 上海市科委民生科技支撑计划专项临床医学科技创新项目（No.17411950500，No.17411950505）

Foundtion Items: Special Clinical Medical Science and Technology Innovation Project of Livelihood Science and Technology Support Plan of Shanghai Science and Technology Commission (No.17411950500, No.17411950505)

《大数据》2019年（第5卷）

总目次

◎专题：健康医疗大数据

◎专题：边缘计算

◎专题：大数据治理

◎专题：大数据的系统结构

◎专题：学术大数据

◎专题：大数据整理

◎研究

◎应用

◎论坛

◎前沿

BIG DATA RESEARCH

Content

2019 (Vol.5)