

基于区域残差和 LSTM 网络的机场延误预测模型

屈景怡, 叶萌, 渠星

(中国民航大学天津市智能信号与图像处理重点实验室, 天津 300300)

摘要: 针对目前民航运输业对机场延误预测高精度的要求, 提出一种基于区域残差和长短时记忆 (RR-LSTM) 网络的机场延误预测模型。首先, 将机场的属性信息、气象信息和相关运行航班信息进行融合; 然后, 利用 RR-LSTM 网络对融合后的机场数据集进行特征提取; 最后, 构建 Softmax 分类器对机场延误分类预测。所提 RR-LSTM 网络模型既能有效提取机场延误数据的时间相关性, 又能避免深层 LSTM 网络的梯度消失问题。实验结果表明, RR-LSTM 网络模型预测准确率可达 95.52%, 取得了比传统网络模型更好的预测效果。其中, 融合机场的气象信息和相关运行航班信息后, 预测准确率可提高约 11%。

关键词: 区域残差网络; 长短时记忆网络; 机场延误预测; 特征提取

中图分类号: TP183

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2019091

Airport delay prediction model based on regional residual and LSTM network

QU Jingyi, YE Meng, QU Xing

Tianjin Key Laboratory of Advanced Signal Processing, Civil Aviation University of China, Tianjin 300300, China

Abstract: Nowadays, the civil aviation industry has a high precision requirement of airport delay prediction, so an airport delay prediction model based on the RR-LSTM network was proposed. Firstly, the airport information, meteorological information and related flight information were integrated. Then, the RR-LSTM network was used to extract the features of the fused airport data set. Finally, the Softmax classifier was adopted to classify and predict the airport delay. The proposed RR-LSTM network model can not only extract the time correlation of airport delay data effectively, but also avoid the gradient disappearance problem of deep LSTM network. The experimental results indicate that the RR-LSTM network model has a prediction accuracy of 95.52%, which achieves better prediction results than the traditional network model. The prediction accuracy can be improved about 11% by fusing the weather information and the flight information of the airport.

Key words: regional residual network, long short term memory network, airport delay prediction, feature extraction

1 引言

随着国民经济和航空运输业的迅猛发展, 空中交通流量日益增加, 但空域资源和地面保障能力不足的问题依然突出, 导致机场延误问题日益严重。

机场延误会对乘客、航空公司、机场运行人员等造成严重影响, 给民航的正常运转带来挑战。而机场延误预测算法的研究, 不仅可以提前预知机场拥堵时间点, 为相关的运行决策提供理论依据^[1-2], 还可以间接反映未来机场航班需求量和容量的关系^[3-4],

收稿日期: 2018-09-19; 修回日期: 2019-02-13

基金项目: 国家自然科学基金资助项目 (No.U1833105); 天津市智能信号与图像处理重点实验室开放基金项目 (No.2017ASP-TJ01); 中央高校基本科研业务费基金资助项目 (No.3122018D006)

Foundation Items: The National Natural Science Foundation of China (No.U1833105), The Open Foundation of Tianjin Key Laboratory of Intelligent Signal and Image Processing (No.2017ASP-TJ01), The Fundamental Research Fund for the Central Universities (No.3122018D006)

为机场建设提供参考, 具有非常重要的现实意义。

针对机场延误的预测问题, 国内外相关学者已经开展了一些研究。已有方法中多采用线性回归、支持向量机、决策树等传统算法^[5-14]。文献[6]基于非线性物理假设和机场动态特性知识, 提出了基于边界条件的机场延误预测方法, 通过对不同机场数据的实验验证, 最高预测准确率为 75.2%。文献[7]针对机场航班数据不断更新的特点, 利用增量式排列支持向量机算法进行延误预警, 其准确率可达 80%。文献[9]利用逻辑回归和决策树这 2 个机器学习模型, 预测机场地面延迟的发生。通过实验对比, 逻辑回归可比决策树取得更好的预测结果, 但其准确率也仅达到 84%。文献[11]尝试利用浅层的人工神经网络 (ANN, artificial neural network) 对机场空域进行延误预测, 但是该模型不能准确学习时间维度上的延误传播特性, 在面对严重且持续性延误状况时, 模型的预测结果并不理想。

虽然国内外学者在机场延误预测方面取得了一定成果, 但仍存在以下 3 个问题。1) 这些方法大多针对小数据集样本进行训练, 面对大量且高维的机场延误数据, 往往需要进行降维处理, 造成原始数据信息的缺失。2) 这些方法没有考虑机场状态信息的时间相关性, 实际上, 前一时刻机场的延误对后续机场延误状态的影响很大。3) 这些方法较少考虑气象信息对延误预测的影响。根据《2017 年全国民航航班运行效率报告》统计结果, 天气原因是造成延误的主要原因之一。

近年来, 面向大数据的深度学习方法在计算机视觉、语音识别、自然语言处理等方面得到了广泛应用, 其性能明显优于传统算法^[15]。在众多深度学习方法中, 循环神经网络^[16] (RNN, recurrent neural network) 将时序的概念引入网络结构设计中, 使其在时序数据分析中表现出更强的适应性。长短时记忆^[17] (LSTM, long short term memory) 网络作为 RNN 的改进, 解决了 RNN 的梯度消失、长期记忆能力不足等问题, 使循环神经网络能够真正有效地利用长距离时序信息。然而, 在实际应用中发现, 由于网络中全连接结构的存在, 使 LSTM 网络无法继续加深, 预测精度无法进一步提高, 并且大规模 LSTM 网络的训练对硬件要求比较高。

为解决上述问题, 本文提出了基于区域残差和 LSTM (RR-LSTM, region residual-long short term memory) 网络的机场延误预测模型。该模型利用

LSTM 网络来充分提取机场延误数据的时间相关性, 并在 LSTM 基础上增加了区域残差网络模块和单向池化模块。其中, 区域残差网络模块是对残差网络 (ResNet, residual network)^[18] 的改进, 以代替 LSTM 层间全连接的功能, 解决深层 LSTM 的梯度消失问题; 单向池化模块的设计是为了保证区域残差网络模块和 LSTM 网络模块的特征维度一致性。此外, 本文提出的基于 RR-LSTM 网络的模型综合考虑了机场的属性信息、气象信息和相关运行航班信息影响, 融合气象和航班信息后, 预测准确率得到了一定程度提高。

2 RR-LSTM 网络结构

图 1(a)为文献[11]的网络结构, 图 1(b)为本文提出的 RR-LSTM 网络结构。通过对比可以看出, 一条待分类机场延误数据在 RR-LSTM 网络中的训练框架主要由 3 个部分组成, 分别是区域残差网络模块、单向池化模块和 LSTM 网络模块。

1) 区域残差网络模块: 分别对不同区域的输入数据进行卷积, 获取输入数据的空间局部特征, 同时保证输入数据的时间顺序。

2) 单向池化模块: 将区域残差网络的输出进行维度转换, 保证网络输入数据在区域残差模块和多层 LSTM 网络中的维度一致性。

3) LSTM 网络模块: 接收单向池化模块的输出作为 LSTM 模块的序列化输入, 获取机场延误数据在时间维度上的依赖关系。

2.1 LSTM 网络

LSTM 网络可以充分利用机场延误信息的时间相关特性。对于 LSTM 网络来说, 其时间步长越多, 记忆能力越强; 但是随着网络深度的增加, LSTM 会出现梯度消失的问题, 文献[19]通过实验验证了增加 LSTM 的层次反而会使其性能下降。为进一步说明此问题, 本文给出深层 LSTM 梯度消失的理论证明。

如图 2 所示, LSTM 网络通过一个记忆单元来存储输入数据的重要特征信息。每一个网络神经元包含核心元素 cell 和 3 个门单元, 分别是输入门、遗忘门和输出门。

LSTM 单元的前向传播过程如式(1)~式(6)所示。其中, i_t^l 、 f_t^l 和 o_t^l 分别为 l 层 t 时刻 LSTM 单元输入门、遗忘门和输出门的输出; b_{l-1}^l 表示 $l-1$ 层网络在 t 时刻的输出, 而 b_{l-1}^{l-1} 则表示 l 层网络在 $t-1$ 时刻的输出; W_i 、 W_f 和 W_o 为 t 时刻 $l-1$ 层网络与

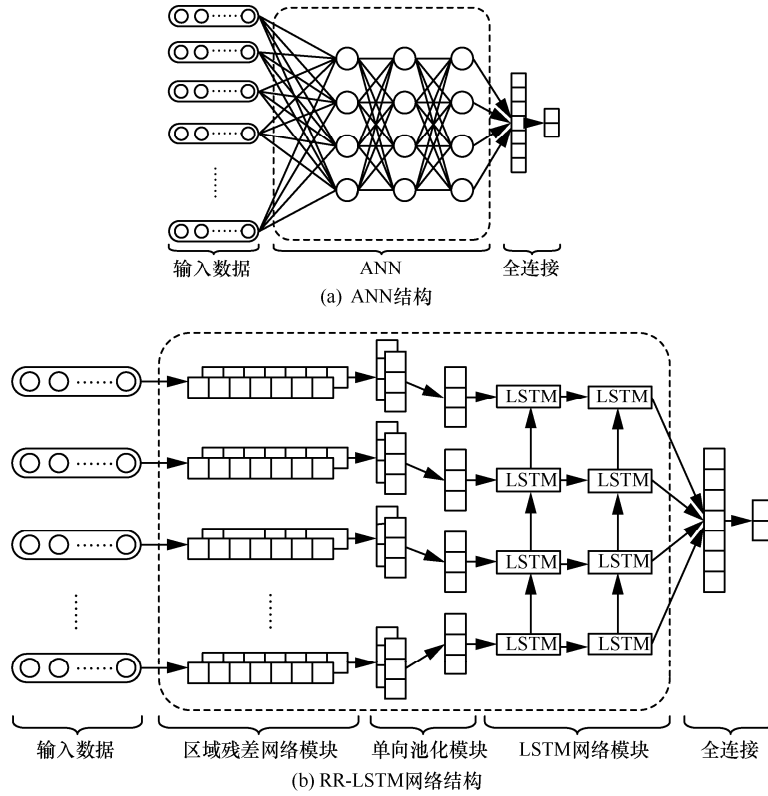


图 1 网络结构

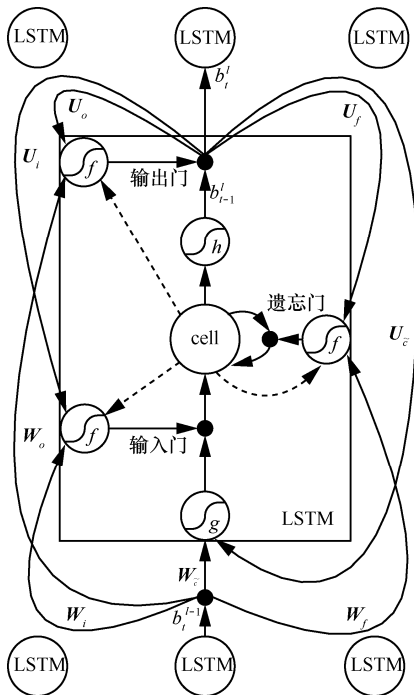


图 2 LSTM 单元结构

该单元输入门、遗忘门、输出门的权重矩阵；同样地， U_i 、 U_f 和 U_o 为 $t-1$ 时刻 l 层网络与该单元各个门的权重矩阵； c_t^l 为核心元素 cell 的输出，

从图 2 和式(4)可以看出， c_t^l 包含以下 2 个部分：第一部分是 $t-1$ 时刻 cell 状态 c_{t-1}^l 和遗忘门输出 f_t^l 的乘积，第二部分是输入门输出 i_t^l 和 \tilde{c}_t^l 的乘积。 W_c 为 t 时刻 $l-1$ 层网络与 cell 连接的权重矩阵， U_c 为 $t-1$ 时刻 l 层网络 cell 连接的权重矩阵。 b_t^l 则为最终 LSTM 的输出。式(1)~式(6)中的 $f(\cdot)$ 、 $g(\cdot)$ 和 $h(\cdot)$ 均为激活函数。

$$i_t^l = f(W_i b_{t-1}^{l-1} + U_i b_{t-1}^l) \tag{1}$$

$$f_t^l = f(W_f b_{t-1}^{l-1} + U_f b_{t-1}^l) \tag{2}$$

$$o_t^l = f(W_o b_{t-1}^{l-1} + U_o b_{t-1}^l) \tag{3}$$

$$\tilde{c}_t^l = g(W_{\tilde{c}} b_{t-1}^{l-1} + U_{\tilde{c}} b_{t-1}^l) \tag{4}$$

$$c_t^l = f_t^l c_{t-1}^l + i_t^l \tilde{c}_t^l \tag{5}$$

$$b_t^l = o_t^l h(c_t^l) \tag{6}$$

LSTM 误差项的反向传播包括 2 个方向：1) 沿时间维度的反向传播，即从当前 t 时刻开始，计算前面每个时刻的误差项；2) 将误差向前一层进行传播。对于后者，目前的相关研究还不多，故根据误差反向传播原理，推导出误差项沿网络层数的反

向传播过程为

$$\delta_i^{l-1} = \frac{\partial J}{\partial b_i^{l-1}} = \delta_{i,i}^l W_i + \delta_{f,i}^l W_f + \delta_{c,i}^l W_c + \delta_{o,i}^l W_o \quad (7)$$

其中, $\delta_{i,i}^l$ 、 $\delta_{f,i}^l$ 、 $\delta_{o,i}^l$ 和 $\delta_{c,i}^l$ 分别为各个门单元的误差项。具体内容如式(8)~式(11)所示。

$$\delta_{i,i}^l = \delta_i^l o_i^l f'(c_i^l) \tilde{c}_i^l i_i^l (1 - i_i^l) \quad (8)$$

$$\delta_{f,i}^l = \delta_i^l o_i^l f'(c_i^l) c_{i-1}^l f_i^l (1 - f_i^l) \quad (9)$$

$$\delta_{o,i}^l = \delta_i^l f(c_i^l) o_i^l (1 - o_i^l) \quad (10)$$

$$\delta_{c,i}^l = \delta_i^l o_i^l f'(c_i^l) i_i^l (1 - (\tilde{c}_i^l)^2) \quad (11)$$

其中, $f(\cdot)$ 为激活函数, $f'(\cdot)$ 为激活函数的导函数, LSTM 单元中的激活函数多采用 tanh 和 sigmoid 函数, 即 $f(\cdot) \leq 1, f'(\cdot) \leq 1$ 。并且权值矩阵中元素的绝对值均小于 1。另外需要解释的是, 由于式(7)中进行相加的四项存在正负号, 加法并不能使 δ_i^{l-1} 的值增大, 因此当 LSTM 层数增加时, 根据链式法则 (δ_i^l 的组成形式和 δ_i^{l-1} 相似), 式(7)中会出现多个小于 1 的项进行连乘, 梯度在向前传播时会不停衰减, 从而导致梯度消失的问题。

为解决深层 LSTM 的梯度消失问题, 根据后续实验结果, 本文提出的 RR-LSTM 网络只保留了两层 LSTM, 单条机场延误数据的特征提取则由深层区域残差网络来代替完成。

2.2 区域残差网络模块

ResNet 中的卷积操作通过权值共享可减少训练参数, 与 LSTM 相比, 更适于数据的空间局部特征提取。通过对传统卷积神经网络结构的改进, ResNet 解决了深层网络结构的梯度消失问题, 具有更强大的特征学习和表达能力, 图 3(a)为文献[18]提出的 ResNet 结构块, 但是常规的卷积操作会破坏输入数据的时间顺序, 无法将 ResNet 结构直接增加至 LSTM 网络中。为此, 本文提出了区域残差网络, 区域残差网络的单个模块结构如图 3(b)所示, 不同层数的网络均由其堆叠而成。

设 X 为输入, 经过批归一化 (BN, batch normalization) [20]和整流线性单元 ReLU[21]操作后, 得到 \tilde{X} 。图 3 (b) 中卷积模块的非线性映射函数 $H(\cdot)$ 的定义如式(12)所示。

$$H(\tilde{X}, \{W\}) = W^{(m_3)} \otimes F(W^{(m_2)} \otimes F(W^{(m_1)} \otimes \tilde{X})) \quad (12)$$

其中, $F(\cdot)$ 表示 ReLU 激活函数和 BN 操作的计算过程; $W^{(m_1)}$ 、 $W^{(m_2)}$ 和 $W^{(m_3)}$ 分别为同一个结构块中 3 个卷积层的权值矩阵, m_1 、 m_2 和 m_3 代表 3 个不同卷积层。单个结构块的非线性映射函数为

$$Y = H(\tilde{X}, \{W\}) + \tilde{X} \quad (13)$$

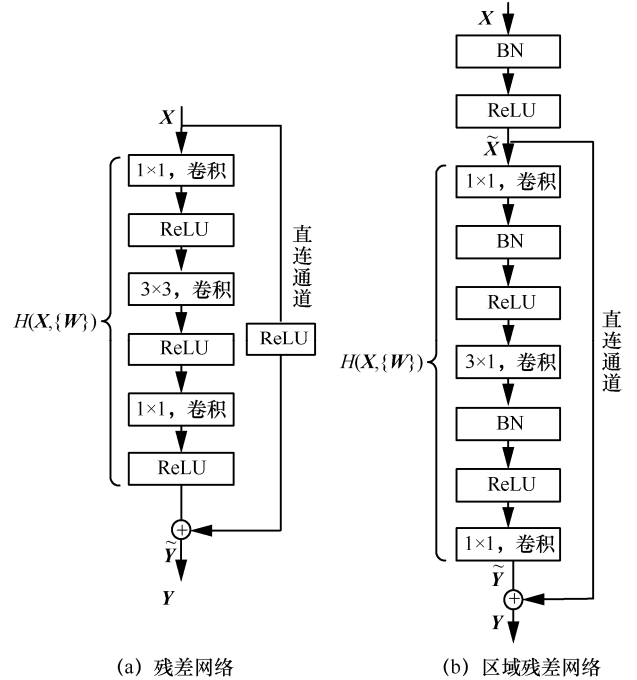


图 3 残差学习结构块

区域残差网络是为了解决 ResNet 中常规卷积操作会破坏数据时序信息的问题。由图 3 可以看出, 其根本区别是根据机场延误的数据集特点, 将残差网络结构块中的卷积核尺寸由 3×3 替换为 3×1。此外, 实际卷积操作时将步长由原来的二维移动变为单向移动, 这样可以使卷积操作仅覆盖单独一个时间区域内的数据, 即单独对输入数据的不同区域进行局部特征提取, 进而实现空间特征深层提取, 同时保留机场延误数据的时序关系。为使网络更加容易训练, 采用了以下 3 个优化方法: 将直通通道中的激活函数 ReLU 去掉, 在每一卷积层前都进行 BN 操作, 对输入进行前置激活[22-23]。

2.3 单向池化

为了将 LSTM 与区域残差模块有效地贯穿融合, 本文设计了单向池化模块, 如图 4 所示, 图中 t 代表时间步长数。该模块首先利用尺寸为 1×1、步长为 1、数量为 128 的卷积核, 对区域残差网络模块的输出特征图进行降维操作, 再进行单向全局池化。池化方程为

$$A_i = (a_{i,1}, a_{i,2}, \dots, a_{i,j}, \dots, a_{i,7}) \quad (14)$$

$$s_i = \text{ave}(A_i) \quad (15)$$

其中， $a_{i,j}$ 表示上一层神经网络输出特征图中坐标为 (i, j) 的神经元， s_i 表示最大池化层中坐标为 i 的神经元， $\text{ave}(\cdot)$ 表示求矩阵中所有元素的均值。

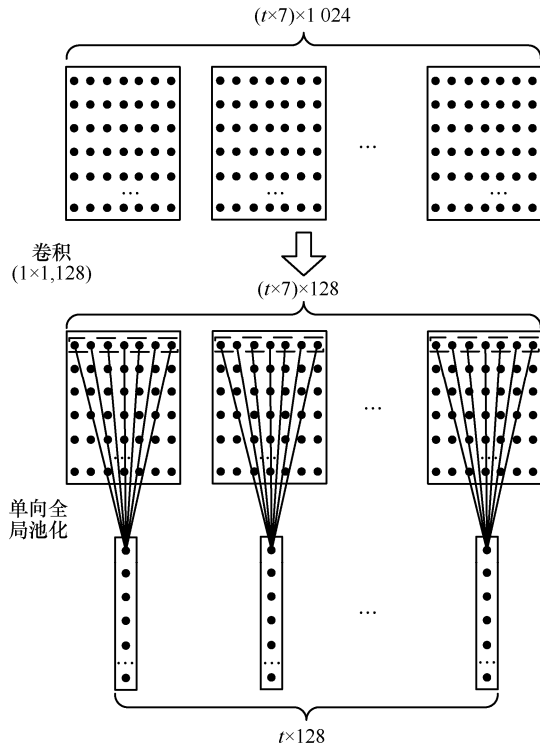


图 4 单向池化模块

本文设计上述单向池化的方式是出于以下 3 点考虑。

1) 区域残差网络的输出特征图维度需要符合 LSTM 输入特征维度的要求，传统方法只是直接进行维度重塑操作，即将三维的数据压缩为二维，再使用全连接的方式进行维度变换。由于全连接的引入，不仅会使计算量增长，而且会增加网络梯度消失问题出现的概率。而单向池化方法通过卷积神经网络的权值共享，可实现维度变换的同时，有效控制网络中增加的训练参数数量。

2) 如图 4 所示，区域残差网络的输出特征图数量为 1 024，若直接进行单向全局池化操作，则会得到维度为 $t \times 1\ 024$ 的数据作为 LSTM 输入，由于 LSTM 网络各单元之间的连接均是全连接方式，在实际实现过程中，会出现网络不收敛的情况，因此在实现维度转换之前进行了 1×1 的卷积处理，以减少特征图的数量。

3) 单向全局池化的特征处理过程是，将不同卷积核输出的特征图，映射到与 LSTM 相连接的不同神经元，这样做的好处是保存了每个卷积核特征表达的独立性。

3 机场延误预测模型

基于 RR-LSTM 网络的机场延误预测模型如图 5 所示，主要包括以下 3 个部分：数据预处理、特征提取和特征分类。特征提取部分介绍详见第 2 节，下面主要对数据预处理和特征分类展开介绍。图 5 给出了各个网络结构的输出特征图维度，以 $(t \times 102) \times 16$ 为例，其中， t 表示循环的时间步长，102 表示每个时刻需要处理的一维向量大小，16 表示输出特征图的个数。

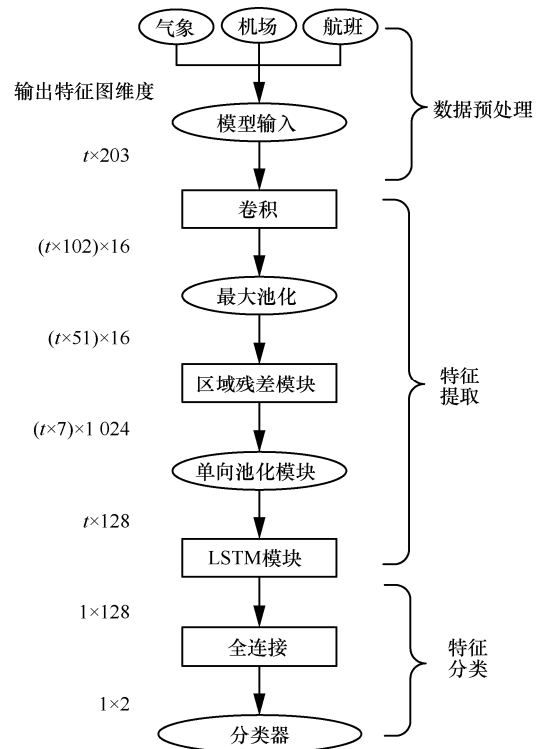


图 5 基于 RR-LSTM 网络的机场延误预测模型

3.1 数据预处理

机场属性数据定义为 A ，包括日期、时间、机场 ID 等 5 个特征属性。

机场运行航班数据定义为 F ，包括起飞日期、起飞/降落机场 ID、起飞/降落城市 ID、航班号、计划起飞/降落时间、起轮/放轮时间、航程等 22 个特征属性。

机场天气数据定义为 W ，包括时间、机场天气状态、风向、风速、能见度、温度、气压等 18 个

特征属性。

数据预处理的步骤如下。

Step1 以日期、航班起飞/降落时间、天气观测时间这 3 个特征作为键值，对机场属性数据 F 、航班数据 A 和天气数据 W 进行融合。

Step2 将原始数据特征属性分为离散性和连续性 2 种。首先对连续性特征进行 min-max 归一化编码^[23]。然后将离散性特征以 100 为阈值分为高低基数 2 类，特征的基数是指该特征类别的穷举数。最后，对低基数离散性数据采用独热编码，即 one-hot 编码^[24]；对高基数离散特征进行 mean-encoder 编码^[25]。

Step3 将编码后的数据 E 按照起飞和降落时间进行排序，然后以长度为 L 、步长为 1 的分割窗口在时间维度上进行滑动分割，得到模型输入 E_t 。分割后每个输入数据的时间序列长度为 L ，对应于机场延误预测模型中的时间步长 t 。

3.2 特征分类

根据模型中机场延误时间的定义，将延误时间在 15 min 以内的情况划分为不延误状态，反之划分为延误状态，从而将机场延误预测问题转化为二分类问题。

如图 5 所示，特征分类部分共有两层：全连接层和 Softmax^[26]分类器层。其中，全连接层输出的特征图维度为 $(1 \times 1) \times q$ ， q 为分类类别数，本文涉及的是二分类问题，所以 $q=2$ 。对于每类有 r 个样本的集合，可以表示为 $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(r)}, y^{(r)})\}$ ，其中， $y^{(r)} \in \{1, 2, \dots, q\}$ ，每一个类别 j 的输出概率值为 $p(y = j | x)$ ，如果用 q 个向量表示 q 个输出概率值，则函数为

$$h_{\theta}(x) = \begin{bmatrix} p(y^{(i)} = 1)x^{(i)}; \theta \\ p(y^{(i)} = 2)x^{(i)}; \theta \\ \vdots \\ p(y^{(i)} = q)x^{(i)}; \theta \end{bmatrix} = \frac{1}{\sum_{j=1}^q e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_q^T x^{(i)}} \end{bmatrix} \quad (16)$$

其中， $h_{\theta}(x)$ 为神经网络的输出， θ 为网络参数， i 为样本序号， $\sum_{j=1}^q e^{\theta_j^T x^{(i)}}$ 为归一化因子。网络训练使用交叉熵作为代价函数，其表达式为

$$l = \frac{1}{r} \left[\sum_{i=1}^r \sum_{j=1}^q 1\{y^{(i)} = j\} \ln \frac{e^{\theta_j^T x^{(i)}}}{\sum_{j=1}^q e^{\theta_j^T x^{(i)}}} \right] \quad (17)$$

为了更好地评估 LSTM 网络在机场延误预测问题上的有效性，本文采用准确率作为评价准则，如式(18)所示。

$$\text{accuracy} = \frac{\sum C}{N} \quad (18)$$

其中， N 为测试集中总的的数据量， C 为模型预测的延误状态与实际延误状态相同的记录数。

4 实验

本节首先介绍了实验环境、参数配置和实验数据，然后基于机场延误数据对比验证了所提模型的性能，并对机场延误数据特征进行了分析。

4.1 实验环境和参数配置

实验所使用计算机的配置如下：处理器为 Intel 至强 Xeon E5-1620，CPU 频率为 3.60 GHz；内存为 16.00 GB；操作系统为 Ubuntu14.04 (64 bit)；图形加速卡为 GeForce GTX TITAN Xp；深度学习开发平台为 Tensorflow 1.3.0。

为了方便重复实验和后续研究，这里给出本文实验中所设置的参数名称与对应的参数值供研究参考：权值初始方式为正交初始化；最大迭代次数为 8.5 万次；训练时的批处理数量为 512，测试时为 128；训练的优化方法为随机梯度下降法；初始学习率为 0.1，训练时分别在迭代 4 万次和 8 万次时进行衰减，每次衰减 10 倍。

4.2 实验数据

本文模型实验数据为 2016 年~2017 年美国亚特兰大机场的航班运行数据和天气数据，其中航班运行数据来自交通运输部 (BTS, bureau of transportation statistics) 的航班准时性数据 (AOTP, airline on-time performance data)，天气数据来自美国国家海洋和大气管理局 (NOAA, national oceanic and atmospheric administration) 提供的本地气候质量监控数据 (QCLCD, quality controlled local climatological data)。将原始数据特征进行预处理后，得到维度为 $1\ 781\ 045 \times 203 \times t \times 1$ 的网络输入数据，其中，1 781 045 为所有数据的数量，203 为编码后每条数据的特征维度， t 为每条数据的时间序列长度，由于计算能力有限，本文只对 $t=30$ 进行了实验验证。

4.3 RR-LSTM 网络结构验证分析

将本文所提网络模型与其他经典神经网络在机场延误数据上进行对比，验证本文所提方法的有效性。

1) ANN。传统前馈型神经网络不同隐藏层之间采取全连接方法。本文设计了层数为 1、2、4、6 的网络结构，对应网络名称为 ANN1、ANN2、ANN4、ANN6。当神经网络深度增加时，模型对计算机性能的要求越来越高，模型的收敛也越来越困难。

2) VGG。VGG 网络作为经典卷积神经网络，包括了反复堆叠 3×3 的小型卷积核和 2×2 的最大池化层，本文基于机场延误数据实现了 VGG16 和 VGG19。对于数据集 R'_t ，由于卷积神经网络没有考虑时间维度上数据的关联性，因此在具体实现策略上，需要将其映射到空间维度上进行卷积，这样的操作必定会丢失一部分数据信息，特别是长时间距离的信息。

3) ResNet。为了增加可对比性，该部分 ResNet 的网络配置与本文所提 RR-LSTM 中的相应模块均一致，实现了 ResNet18、ResNet34 和 ResNet50。与 VGG 相同，ResNet 也需要将其时间的数据映射到空间维度上进行卷积。

4) ANN-LSTM。该网络模型是文献[13]提出的一种模型思想，但是文中的实验数据与本文并不相同，为了保证对比实验的严谨性，需要使用本文数据对模型重新训练。另外在文献[13]的基础上，对隐藏层 ANN 的层数进行了讨论，分别包括 ANN1-LSTM、ANN2-LSTM 和 ANN4-LSTM，其中 LSTM 的层数均为 2 层。

5) VGG-LSTM。与 ANN-LSTM 模型类似，实现了 VGG16-LSTM 和 VGG19-LSTM。

6) RR-LSTM。本文所提网络模型，实现的网络配置有 RR18-LSTM、RR34-LSTM 和 RR50-LSTM。

4.3.1 引入时序信息对预测结果的影响

为验证前续机场延误状态对当前预测结果的影响，本部分将不同神经网络模型在数据集 E 和 E_t 进行了对比实验。根据数据预处理步骤， E 代表单个时刻的数据集， E_t 代表不同时刻数据组成的时间序列集。需要说明的是，由于数据集 E 为单独时刻的特征信息，无法使用 LSTM 网络进行时序关联的学习，因此涉及 LSTM 网络的模型没有对数据集 E 进行训练和测试。

表 1 为 ANN、VGG 和 ResNet 网络在不同数据集 E 和 E_t 上的预测结果对比，可以看出增加数据的时间特征信息可以使模型测试准确率提高 0.91%~1.8%。由此可得，考虑引入机场延误数据的

时序特征能有效提高模型最终的预测精度。

表 1 时序信息引入对分类准确率的影响

网络名称	准确率	
	数据集 E	数据集 E_t
ANN1	87.95%	89.49%
ANN2	88.89%	90.34%
ANN4	88.63%	90.15%
ANN6	87.02%	88.64%
VGG16	87.41%	88.93%
VGG19	88.58%	89.73%
ResNet18	90.05%	91.85%
ResNet34	92.14%	93.05%
ResNet50	92.37%	93.97%

表 2 为不使用 LSTM 模块与使用 LSTM 模块的所有网络结构在数据集 E_t 上的预测结果对比。由表 2 数据可以看出，总体趋势上，增加 LSTM 模块使模型测试准确率提高了 0.67%~1.92%。这表明利用 LSTM 来进一步提取特征的时间维度信息，使网络具有更强大的数据抽象能力。ANN 在层数增加到 6 层时出现了梯度消失的现象，测试准确率随着网络层数的增减不升反降，同样地，ANN4-LSTM 和 ANN6-LSTM 网络与对应的 ANN4 和 ANN6 网络相比，也出现了预测精度下降的情况，这是因为 LSTM 层与层之间是全连接的关系，增加 LSTM 模块也相当于增加了 ANN 的层数。

表 2 LSTM 网络模块对分类准确率的影响

不使用 LSTM 模块		使用 LSTM 模块	
网络名称	准确率	网络名称	准确率
ANN1	89.49%	ANN1-LSTM	90.96%
ANN2	90.34%	ANN2-LSTM	91.01%
ANN4	90.15%	ANN4-LSTM	87.83%
ANN6	88.64%	ANN6-LSTM	84.52%
VGG16	88.93%	VGG16-LSTM	90.85%
VGG19	89.73%	VGG19-LSTM	91.24%
ResNet18	91.85%	RR18-LSTM	93.35%
ResNet34	93.05%	RR34-LSTM	94.96%
ResNet50	93.97%	RR50-LSTM	95.52%

根据以上 2 个实验结果对比，说明了时序性特征的引入以及相应的 LSTM 模块能够有效提高模型的预测性能。

4.3.2 区域残差网络模块性能分析

为证明传统多层 LSTM 网络的梯度消失现象，并验证引入区域残差网络模块对模型的影响，使用传统 LSTM 网络和 RR-LSTM 网络进行实验对比。

首先给出传统 LSTM 网络的预测准确率，实验结果如表 3 所示，其中，LSTM- n 表示该网络包括 n 层 LSTM 和一层全连接。

表 3 传统 LSTM 网络的预测准确率

网络名称	LSTM 层数	全连接层数	网络总层数	准确率
LSTM-1	1	1	2	91.75%
LSTM-2	2	1	3	93.19%
LSTM-3	3	1	4	93.20%
LSTM-4	4	1	5	84.93%

从表 3 中可以看出，对于传统 LSTM 网络，LSTM-3 效果达到最佳，分类正确率为 93.20%，但是与 LSTM-2 相比，只多了 0.01%。通过进一步加深网络，LSTM-4 的测试准确率反而下降到了 84.93%，说明仅增加 LSTM 层数已不能提高模型分类的准确率。

根据以上实验结果分析，将 RR-LSTM 网络中的 LSTM 网络层数均设为 2 层，没有选择分类效果最好的 LSTM-3 是由于 RR-LSTM 网络中已存在大量的卷积神经网络，可以充分学习机场延误数据的空间特征，并且 LSTM-2 网络与 LSTM-3 网络的准确率仅仅相差 0.01%，相对于再增加一层参数较多的全连接 LSTM 网络来说，选择 2 层的 LSTM 网络更为合适。

表 4 给出了 RR-LSTM 网络的预测准确率，其中，RR m -LSTM 表示 m 层区域残差网络、2 层 LSTM 网络、单向池化模块和 2 层全连接。从实验结果可以看出，RR-LSTM 网络随深度的增加预测准确率逐步提升，当网络层数为 54 层时，其预测准确率为 95.52%，可训练网络深度远远超过传统 LSTM 网络。

图 6(a)和图 6(b)分别表示 LSTM 和 RR-LSTM 网络随迭代次数变化的损失值大小。损失值越小，说明模型预测值与真实值大小越接近。由图 6(a)可

知，LSTM-4 和 LSTM-2 收敛后的损失值相差不大，大致分布在 0.2~0.4。但是随着网络层数的增加，LSTM-6 增加到了 0.5 附近。图 6(b)中 RR-LSTM 网络从 22 层、38 层到 54 层，随着模型网络层数的不断增加，损失值不断减少，当网络为 54 层时，损失值分布在 0.1 左右。这说明 RR-LSTM 比 LSTM 模型能够更好地对深层神经网络模型进行训练。

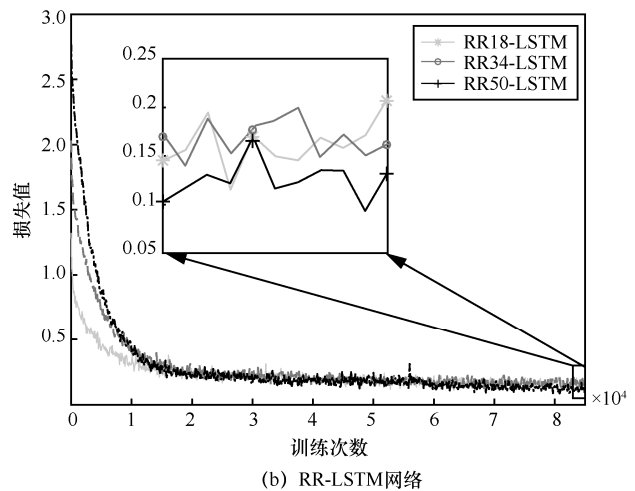
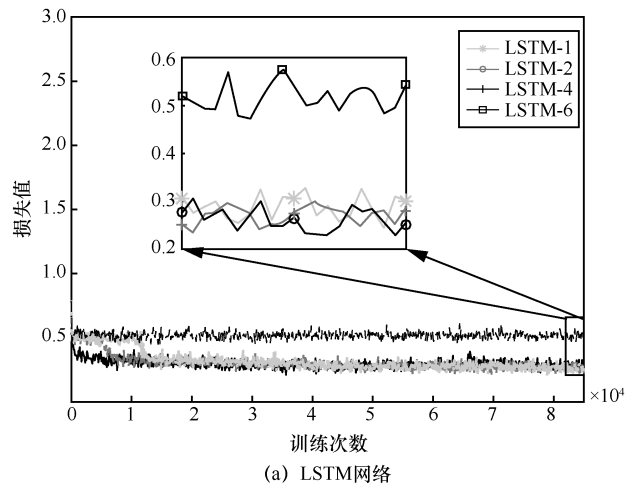


图 6 损失值随迭代次数变化

神经网络模型在反向传导过程中进行隐藏层权重以及误差项的更新，如果网络第一个隐藏层的梯度值随迭代次数变化能一直保持稳定，则表征该

表 4 RR-LSTM 网络的预测准确率

网络名称	卷积层数	LSTM 层数	全连接层数	网络总层数	准确率
RR18-LSTM	18	2	2	22	93.35%
RR34-LSTM	34	2	2	38	94.96%
RR50-LSTM	50	2	2	54	95.52%

网络其他隐藏层均能保证信息有效传递。图 7 为不同 LSTM 和 RR-LSTM 网络结构第一层梯度值随迭代次数的变化曲线。

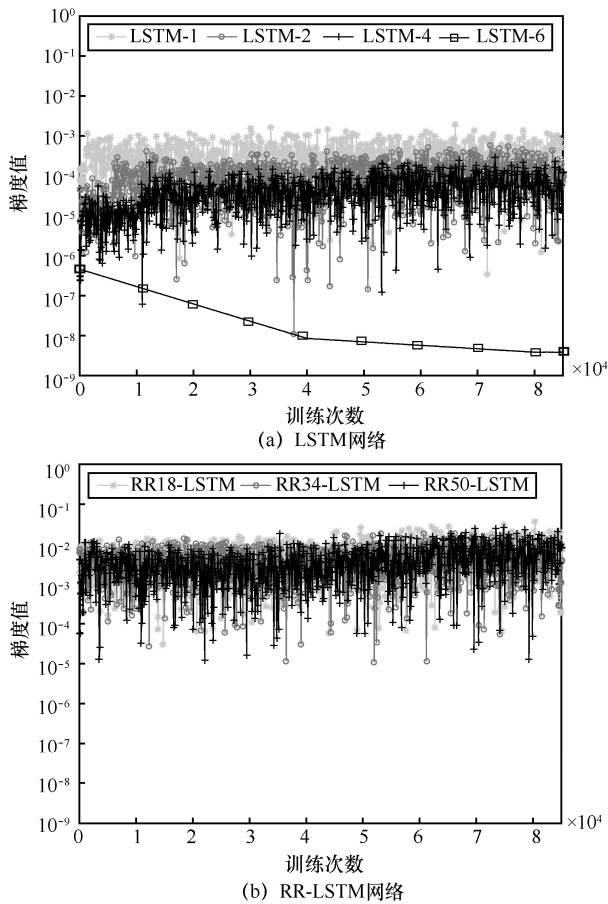


图 7 隐藏层梯度值随迭代次数变化

由图 7(a)中可以明显看出，随着深度的增加，梯度值锐减，当网络增至 6 层时，梯度值减小到 10^{-8} 左右，且仍随着迭代次数逐渐衰减。这说明深层 LSTM 网络在权重更新时存在梯度消失问题，符合上文的理论分析和式(7)~式(11)的推导结论。图 7(b)中的相应梯度值基本稳定在 $10^{-1} \sim 10^{-5}$ ，且随着网络层数的增加，没有出现明显的数值变化。由此可见，RR-LSTM 网络可以进行更深层网络的训练，从而更加完备地表达数据信息。

4.3.3 单向池化结构性能分析

利用图 4 所示的单向池化模块来代替传统的维度重塑方式，可以保证区域残差网络模块输出和 LSTM 模块输入的维度一致性，同时可尽量减少网络中引入参数的数量。使用机场延误数据，根据数据维度转换方式的不同分别训练了 RR18-LSTM、RR34-LSTM 和 RR50-LSTM 这 3 个模型。

从表 5 的数据可以看出，使用单向池化模块比维度重塑方式拥有更高的分类准确率，当网络为 22 层、38 层和 54 层时，准确率分别提高了 0.53%、0.45%、0.46%。

表 5 不同维度转换方式对分类准确率的影响

网络名称	网络层数	测试准确率	
		维度重塑	单向池化
RR18-LSTM	22	92.82%	93.35%
RR34-LSTM	38	94.51%	94.96%
RR50-LSTM	54	95.06%	95.52%

为了验证单向池化模块对模型反向传播过程的影响，图 8 给出了不同模型第一层网络梯度值随迭代次数的变化曲线。可以看出，随着网络层数的增加，单向池化方式的梯度值没有明显变化，而维度重塑方式中，22 层、38 层、54 层网络的梯度值出现了逐一减少的现象。实验结果证明了利用单向池化模块进行维度转换的有效性。

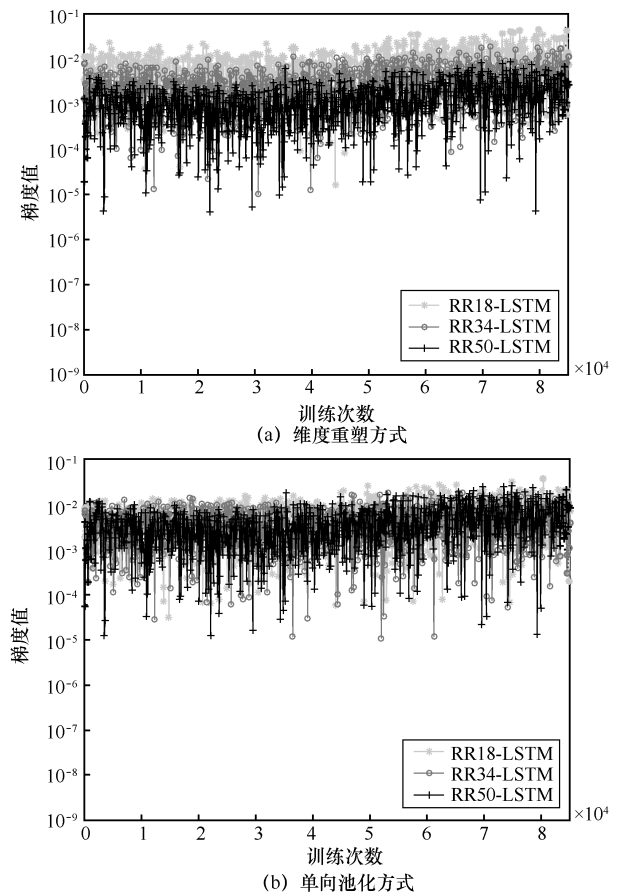


图 8 隐藏层梯度值随迭代次数变化

4.4 机场延误数据特征对预测结果的影响

为研究机场延误数据特征中融合的气象信息 W 和相关运行航班信息 F 对延误结果的影响, 分别将仅考虑机场属性信息的数据集 A 、增加天气因素数据集 $A+W$ 和综合三者信息的数据集 $A+W+F$ 输入 RR-LSTM 模型中进行延误预测, 实验结果如表 6 所示。

表 6 测试准确率

网络名称	准确率		
	数据集 A	数据集 $A+W$	数据集 $A+W+F$
RR18-LSTM	80.54%	90.86%	93.35%
RR34-LSTM	83.06%	92.15%	94.96%
RR50-LSTM	84.19%	93.24%	95.52%

由表 6 可以看出, 与仅考虑机场本身属性相比, 融合天气因素和航班运行数据后, 不同网络结构的预测准确率均增加了 11% 左右。其中, 如果只考虑天气因素预测精度增加约 9%, 进一步增加机场的运行航班数据后, 预测结果准确率又增加了 2%。因此, 综合考虑上述 2 种因素能有效提高模型最终的预测精度。

文献[24]采用与本文同样的数据来研究不同信息对机场延误的影响。但是由于 RR-LSTM 模型具有高效的特征提取及数据处理能力, 因此不需对数据特征按影响力进行降维, 能够尽量全面地考虑可能对延误产生影响的因素, 以增加模型最终的预测性能。

5 结束语

本文通过分析机场延误预测问题的特点, 提出了 RR-LSTM 网络模型, 并且通过大量的实验验证了该模型的有效性, 结论如下。

1) 增加机场延误数据的时序信息, 并且利用 LSTM 网络学习数据的时间依赖关系, 使模型的物理含义更加符合机场延误事件特点, 可得到更高的预测准确率。

2) 当传统多层 LSTM 网路加深时, 模型性能开始变差, 而增加区域残差网络模块进行特征提取时, 模型的性能保持稳定, 证明改进后的网络结构可以保证深度网络的训练。

3) 与传统维度变换方式相比, 单向池化方法可以减少网络训练参数的数量, 进而降低网络梯度消失的概率。

4) 融合气象信息和航班运行信息后, 模型预测准确率均有所提高, 相比之下, 机场内的气象信息对预测结果的影响较大。

本文的预测结果与传统算法相比有较大提升, 有利于进一步改善空管、航空公司等部门的决策, 提升民航综合服务能力。

参考文献:

- [1] 吴薇薇, 孟婷婷, 张皓瑜. 基于机场延误预测的航班计划优化研究[J]. 交通运输系统工程与信息, 2016, 16(6): 189-195.
WU W W, MENG T T, ZHANG H Y. Flight plan optimization based on airport delay prediction[J]. Journal of Transportation Systems Engineering and Information Technology, 2016, 16(6): 189-195.
- [2] NAYAK N, ZHANG Y. Estimation and comparison of impact of single airport delay on national airspace system with multivariate simultaneous models[J]. Transportation Research Record, 2011, 2206(1): 52-60.
- [3] EADS G, KIEFER M, MEHNDIRATTA S. Short-term delay mitigation strategies for san francisco international airport[J]. Transportation Research Record Journal of the Transportation Research Board, 2001, 1744(1):44-51.
- [4] 韩淑敏. 机场运行可预测性分析与优化[D]. 天津: 中国民航大学, 2016.
HAN S M. Analysis and optimization of the airport operational predictability[D]. Tianjing: Civil Aviation University of China, 2016.
- [5] 罗谦, 张永辉, 程华, 等. 基于航空信息网络的枢纽机场航班延误预测模型[J]. 系统工程理论与实践, 2014, 34(S1): 143-150.
LUO Q, ZHANG Y H, CHENG H, et al. Study on flight delay prediction model based on flight networks[J]. Systems Engineering-Theory&Practice, 2014, 34(S1): 143-150.
- [6] DEUTSCHMANN A. Prediction of airport delays based on non-linear considerations of airport systems[C]//The 28th International Congress of the Aeronautical Sciences. 2012: 1-5.
- [7] 徐涛, 丁建立, 顾彬, 等. 基于增量式排列支持向量机的机场航班延误预警[J]. 航空学报, 2009, 30(7): 1256-1263.
XU T, DING J L, GU B, et al. Forecast warning level of flight delays based on incremental ranking support vector machine[J]. Acta Aeronautica et Astronautica Sinica, 2009, 30(7): 1256-1263.
- [8] 张静, 徐肖豪, 王飞, 等. 基于模糊线性回归模型的机场延误性能评估[J]. 交通运输工程学报, 2010, 10(4): 109-114.
ZHANG J, XU X H, WANG F, et al. Airport delay performance evaluation based on fuzzy linear regression model[J]. Journal of Traffic and Transportation Engineering, 2010, 10(4): 109-114.
- [9] MUKHERJEE A, GRABBE S, SRIDHAR B. Predicting ground delay program at an airport based on meteorological conditions[C]//The 14th AIAA Aviation Technology, Integration, and Operations Conference. 2014.
- [10] 郭野晨风, 李杰, 胡明华, 等. 基于简化 WITI 指标的机场延误预测方法[J]. 交通运输系统工程与信息, 2017, 17(5): 207-213.
GUO Y C F, LI J, HU M H, et al. Airport delay prediction method based on simplified WITI index[J]. Journal of Transportation Systems Engineering and Information Technology, 2017, 17(5): 207-213.
- [11] NOBORU T, RYOSUKE K, AKIHIDE S, et al. Prediction of delay due to air traffic control by machine learning[C]//AIAA Modeling and

- Simulation Technologies Conference. 2017: 191-199
- [12] BASPINAR B, URE N K, KOYUNCU E, et al. Analysis of delay characteristics of european air traffic through a data-driven airport-centric queuing network model[J]. IFAC-PapersOnLine, 2016, 49(3): 359-364.
- [13] KIM Y J, CHOI S, BRICENO S, et al. A deep learning approach to flight delay prediction[C]//The 35th Digital Avionics Systems Conference. 2016: 67-72.
- [14] KHANMOHAMMADI S, TUTUN S, KUCUK Y. A new multilevel input layer artificial neural network for predicting flight delays at JFK airport[J]. Procedia Computer Science, 2016, 95: 237-244.
- [15] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [16] TSOI C A, SHAOHUA T. Recurrent neural networks: a constructive algorithm, and its properties [J]. Neuro computing, 1997, 15(3-4): 309-326.
- [17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//The 26th IEEE Conference on Computer Vision and Pattern Recognition.2016: 770-778.
- [19] WANG J, YANG Y, MAO J, et al. CNN-RNN: a unified framework for multi-label image classification[C]//The 29th IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2285-2294.
- [20] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//The 32nd International Conference on Machine Learning. 2015: 448-456.
- [21] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//The 27th International Conference on Machine Learning. 2010: 807-814.
- [22] 屈景怡, 朱威, 吴仁彪. 基于衰减因子的双通道神经网络图像分类算法[J]. 系统工程与电子技术, 2017, 39(6): 1391-1399.
- QU J Y, ZHU W, WU R B. Image classification for dual-channel neural networks based on attenuation factor[J]. Systems Engineering and Electronics, 2017, 39(6): 1391-1399.
- [23] 黄文坚, 唐源. TensorFlow 实战[M]. 北京: 电子工业出版社, 2017.
- HUANG W J, TANG Y. TensorFlow practice[M]. Beijing: Publishing House of Electronics Industry, 2017.
- [24] 吴仁彪, 李佳怡, 屈景怡. 基于双通道卷积神经网络的航班延误预

测模型[J]. 计算机应用, 2018, 38(7): 2100-2112.

WU R B, LI J Y, QU J Y. Flight delay prediction based on dual-channel convolutional neural networks[J]. Journal of Computer Applications, 2018, 38(7): 2100-2112.

[25] MICCI-BARRECA D. A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems[J]. ACM SIGKDD Explorations Newsletter, 2001, 3(1):27-32.

[26] DUAN K, KEERTHI S S, CHU W, et al. Multi-category classification by soft-max combination of binary classifiers[C]//The International workshop on multiple classifier systems. 2003: 125-134.

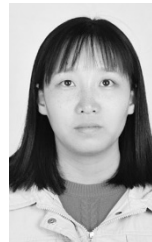
[作者简介]



屈景怡(1978-), 女, 天津人, 博士, 中国民航大学副教授, 主要研究方向为航空运输大数据、神经网络、深度学习。



叶萌(1994-), 女, 河南长葛人, 中国民航大学硕士生, 主要研究方向为航空运输大数据、深度学习。



渠星(1995-), 女, 河北定州人, 中国民航大学硕士生, 主要研究方向为航空运输大数据、深度学习。