

基于深度SE-DenseNet的航班延误预测模型

吴仁彪 赵婷 屈景怡*

(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航班延误高精度预测的需求, 该文提出一种基于深度SE-DenseNet的航班延误预测模型。该模型首先将航班信息、相关机场延误信息和天气信息进行数据融合; 其次, 利用改进后的SE-DenseNet算法对融合后的航班数据集进行自动特征提取; 最后, 构建Softmax分类器进行航班离港延误等级的预测。该文提出的SE-DenseNet结构融合了DenseNet和SENet二者的优势, 既能加强深层信息的传递, 避免梯度消失, 又可以实现特征提取过程中的特征重标定。实验结果表明, 数据融合后, 预测准确率较只考虑航班属性提高约1.8%; 算法改进后可以有效提升网络性能, 模型最终准确率达93.19%。

关键词: 航班延误预测; SE-DenseNet; 数据融合; 特征重标定

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2019)06-1510-08

DOI: 10.11999/JEIT180644

Flight Delay Prediction Model Based on Deep SE-DenseNet

WU Renbiao ZHAO Ting QU Jingyi

(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 prediction demand of flight delays, thus a flight delay prediction model based on the deep SE-DenseNet is proposed. Firstly, flight data, associated airport delay information and meteorological data are fused in the model. Then, the improved SE-DenseNet algorithm is used to extract feature automatically based on the fused flight data set. Finally, the softmax classifier is used to predict the delay level of flight. The proposed SE-DenseNet, combing the advantages of DenseNet and SENet, can not only enhance the transmission of deep information, avoid the problem of vanishing gradients, but also achieve feature recalibration by the feature extraction process. The results indicate that after data fusion, the accuracy of the model is improved 1.8% than only considering the characteristics of the flight itself. The improved algorithm can effectively improve the network performance. The final accuracy of the model reaches 93.19%.

Key words: Flight delay prediction; SE-DenseNet; Data fusion; Feature recalibration

1 引言

近年来, 航班延误频发, 给我国民航运输业带来了严峻的挑战。对航班延误实现精准预测可以为相关部门决策提供参考, 以便及时治理可控延误, 降低航空公司的损失, 对民航的高效安全运营具有实际意义^[1]。针对航班延误预测问题, 国内外学者已展开了大量研究, 多采用神经网络、决策树、支

持向量机等传统的机器学习算法^[1-5]。文献[2]引入多级输入层神经网络算法处理航班数据中的名义变量, 实现最终的延误预测。文献[3]结合航班自身数据特点构建了基于C4.5决策树方法的延误预测模型, 实验结果较贝叶斯网络有一定提高。文献[4]在考虑空运需求与机场容量冲突的条件下, 采用支持向量机进行延误预警研究。但是, 上述研究均只对航班以及起降机场信息建模, 未考虑相关机场间的延误波及性, 而机场间的延误是相互影响的。同时, 以上模型也较少考虑天气因素对航班延误的影响, 气象信息对航班延误预测也十分重要。

此外, 已有研究成果大多针对小数据集样本进行训练, 面对大量且高维的机场、航空公司以及管制部门数据, 往往需要进行降维处理, 这样会造成

收稿日期: 2018-07-02; 改回日期: 2018-11-16; 网络出版: 2018-12-04

*通信作者: 屈景怡 qujingyicauc@163.com

基金项目: 国家自然科学基金联合基金(U1833105), 天津市智能信号与图像处理重点实验室开放项目(2017ASP-TJ01)

Foundation Items: The National Natural Science Foundation of China (U1833105), The Tianjin Key Laboratory of Advanced Signal Processing Open Project (2017ASP-TJ01)

部分信息的丢失。因此，本文提出基于大数据的深度学习方法进行延误预测，从高维数据中自动学习特征，通过提升数据隐藏价值的利用率，进而提高延误预测的准确率。目前，卷积神经网络(Convolutional Neural Networks, CNN)在计算机视觉、语音识别等方面均取得了较大突破，其性能明显优于传统的机器学习算法^[6-12]。文献[8]提出的DenseNet较传统CNN不仅网络性能更优，而且能有效解决深层网络训练时的梯度消失现象，但是却并没有从特征维度层面着手，自动学习每个特征通道的重要程度，大大降低了特征提取过程的效率。而文献[9]提出的SENet的主要思想正是通过建模特征提取过程中特征通道的相互依赖关系，重新标定特征权重，进而提高网络性能。

因此，针对上述问题，本文在深入研究CNN各种变形的基础上，借鉴DenseNet和SENet的设计思想，提出一种能融合二者优势，网络性能更优、更高效的深度SE-DenseNet算法。然后，结合航班信息，综合考虑相关机场延误信息以及气象因素对航班的影响，构建基于深度SE-DenseNet的航班延误预测模型，并通过实际数据集对模型有效性进行验证。

2 SE-DenseNet

2.1 网络描述

传统的DenseNet^[8]网络结构如图1(a)所示，每层网络均以稠密连接的方式连接到后续所有层。本

文提出的SE-DenseNet即是在每个DenseNet结构块的3×3卷积层后增加SE模块，如图1(b)所示。通过这种融合机制，网络既可以实现原始输入信息的无损传输，又可以自动学习全局信息去获得每个通道的重要程度，然后根据重要程度去增强有益特征并抑制无用特征，从而实现特征通道的自适应校准。

表1分别列出40层的SE-DenseNet和DenseNet网络配置信息，其中 k 表示卷积层的特征通道数， γ 为SE操作的维度变换率。由表1可以看出SE-DenseNet与传统的DenseNet相比有以下几点创新：(1)在每个结构块的卷积变换之后设计“SE”模块，对每个特征通道的信息进行权重自动校准；(2)在第1个卷积层后未设置最大池化层，防止由池化操作导致低级特征丢失；(3)在结构块之间的转换层中去掉平均池化操作，只保留1×1卷积，在大量减少计算参数的条件下保留全局信息，以增强整个神经网络的鲁棒性。

2.2 SE模块

SE模块是一个计算单元，结构如图1(b)所示。图中 \mathbf{X}_1 表示对应层输入特征矩阵，表示经卷积变换后的非线性特征映射， \mathbf{X}_2 表示该层最终的输出矩阵，图中 M', H', C', M, H, C 分别代表特征矩阵 $\mathbf{X}_1, \mathbf{X}_2$ 的3维信息。则SE-DenseNet单个结构块里第 L 层网络经卷积变换后的特征映射 U 可表示为

$$U = \mathbf{W}^L \otimes f(\text{BN}(\mathbf{W}^{(L-1)} \otimes f(\text{BN}([\mathbf{X}_0 \mathbf{X}_1 \cdots \mathbf{X}_{L-1}])))) \quad (1)$$

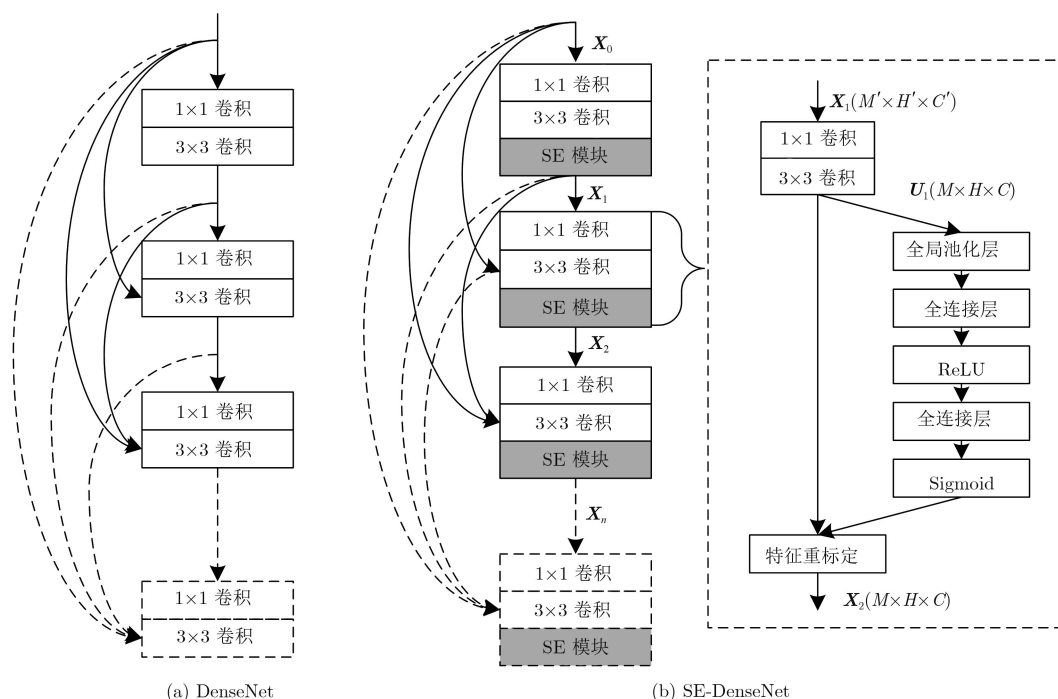


图1 网络结构示意图

表1 SE-DenseNet网络配置信息表

网络层	SE-DenseNet		DenseNet	
	矩阵维度	结构配置	矩阵维度	结构配置
卷积层	$n \times n$	$[3 \times 3, 2k]$	$n \times n$	$[3 \times 3, 2k]$
池化层			$n/2 \times n/2$	3×3 最大池化
结构块1	$n \times n$	$\begin{bmatrix} 1 \times 1, 4k \\ 3 \times 3, k \\ SE(k/\gamma, k) \end{bmatrix} \times 6$	$n/2 \times n/2$	$\begin{bmatrix} 1 \times 1, 4k \\ 3 \times 3, k \end{bmatrix} \times 6$
转换层1	$n \times n$	$[1 \times 1, 0.5k]$	$n/2 \times n/2$	$[1 \times 1, 0.5k]$
			$n/4 \times n/4$	2×2 平均池化
			:	
分类层	1×1	全局平均池化, 全连接, softmax		

其中, $[X_0 X_1 \dots X_{L-1}]$ 表示采用稠密连接的方式将前面所有层的特征映射都作为输入。 $W^{(L-1)}$ 和 W^L 依次表示 $1 \times 1, 3 \times 3$ 卷积权值矩阵, $BN(\cdot)$ 表示对每个隐藏层的输出数据进行批归一化处理^[10], $f(\cdot)$ 为ReLU激活函数^[11], 本文卷积操作均用 \otimes 表示。

SE模块主要包含3个步骤: 首先, 为了建模特征通道间的相互依赖关系, 提出利用全局平均池化将空间信息压缩成一个信道描述符, 如式(2)所示; 其次, 为了标定上述压缩操作中聚合的特征权重, 利用全连接层与sigmoid激活函数组成的门控机制学习多个通道间的非线性交互, 如式(3)所示; 最后, 利用激活值 s 重新缩放转换 U 实现对原始特征的权重标定, 如式(4)。

$$z_c = F_{sq}(u_c) = \frac{1}{M \times H} \sum_{i=1}^M \sum_{j=1}^H u_c(i, j) \quad (2)$$

$$s = F_{ex}(z, W) = \sigma(W_2 f(W_1 z)) \quad (3)$$

$$y_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (4)$$

其中, u_c 和 z_c 分别表示卷积变换、全局特征压缩变换后第 c 个通道的输出映射, $\sigma(\cdot)$ 表示sigmoid函数, $W_1 \in R^{C/\gamma \times C}$ 和 $W_2 \in R^{C \times C/\gamma}$ 分别表示2个全连接层的权值矩阵, y_c 表示经权重标定后第 c 个通道的输出矩阵, 则SE-DenseNet单个结构块里第 L 层网络最终的输出矩阵 $Y = [y_1 y_2 \dots y_c]$ 。

2.3 网络反向传播

SE-DenseNet模型的训练主要通过反向传播(Back Propagation, BP)算法^[12]将训练样本的误差信息传回到隐藏层, 实现隐藏层之间权重矩阵的不断更新迭代, 直至网络收敛。现根据BP算法, 以SE-DenseNet的第1个网络结构块为例对各隐藏层间的梯度值进行推导。假设该结构块包含3组非线性变换, 每组变换有两个卷积层, 一个SE模块, 如图2所示, 则该结构块内各隐藏层的误差项计算如式(5)–式(13)

$$\delta^{k9} = \partial J / \partial U^{k9} \quad (5)$$

$$\delta^{k8} = \delta^{k9} * W^{k9} \otimes f'(U^{k8}) \quad (6)$$

$$\delta^{k7} = \delta^{k8} * W^{k8} \otimes f'(U^{k7}) \quad (7)$$

$$\delta^{k6} = \delta^{k7} * W^{k7} \quad (8)$$

$$\delta^{k5} = \delta^{k6} * W^{k6} \otimes f'(U^{k5}) \quad (9)$$

$$\delta^{k4} = \delta^{k5} * W^{k5} \otimes f'(U^{k4}) \quad (10)$$

$$\delta^{k3} = \delta^{k4} * W^{k4} + \delta^{k7} * W^{n1} \quad (11)$$

$$\delta^{k2} = \delta^{k3} * W^{k3} \otimes f'(U^{k2}) \quad (12)$$

$$\delta^{k1} = \delta^{k2} * W^{k2} \otimes f'(U^{k1}) \quad (13)$$

其中, $\delta^{k1}, \delta^{k2}, \dots, \delta^{k9}$ 分别表示该结构块内对应各

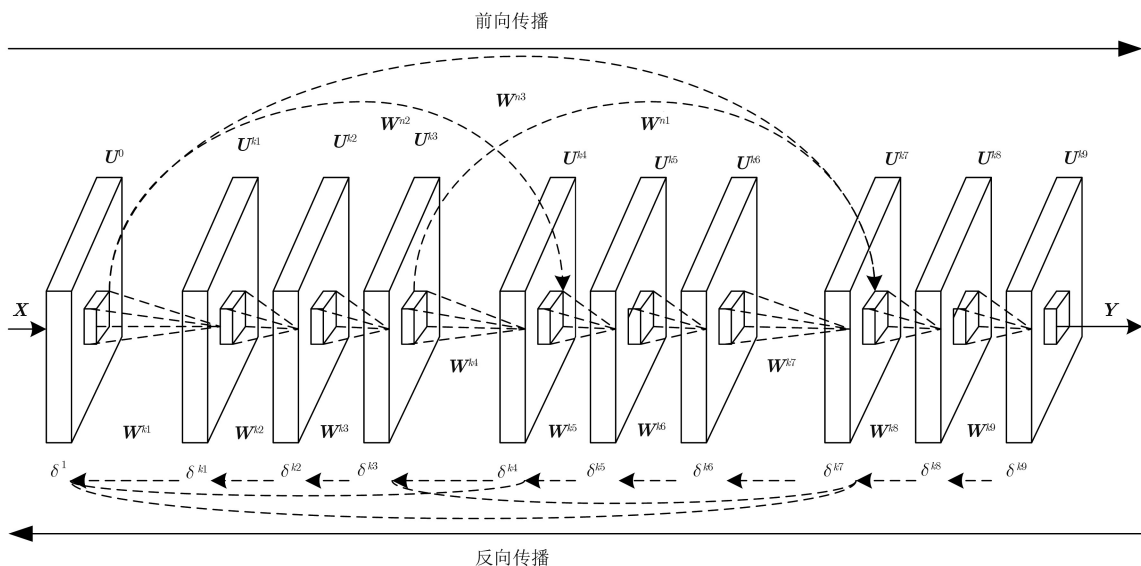


图2 SE-DenseNet的训练

层的误差项, $U^{k1}, U^{k2}, \dots, U^{k9}$ 为各层的输出特征映射, W^{n1} 表示 $k3$ 与 $k7$ 层之间的权值矩阵, $\partial J/\partial U^{k9}$ 表示损失函数 J 对最后一层网络输出特征映射的导数, “*”表示把卷积操作 \otimes 中的卷积核进行翻转。根据BP算法, 其余结构块的误差项也可以由式(5)–式(13)同理推出。

结合式(5)–式(13)以及SE-DenseNet的设计方式, 网络第1个隐藏层的梯度值可表示为

$$\begin{aligned} \partial J/\partial W^1 = & (\delta^{k1} * W^{k1} + \delta^{k3} * W^{n2} \\ & + \delta^{k7} * W^{n3}) \otimes A^0 \end{aligned} \quad (14)$$

其中, W^{n2} 和 W^{n3} 表示 $k3$ 和 $k7$ 层与第1层之间的权值矩阵, A^0 表示输入特征矩阵。可以看出, 第1个隐藏层的梯度信息不仅包括其下一层反向传播的梯

度权值, 还包括该结构块中每组非线性变换的梯度信息, 例如 $k3$ 和 $k7$ 层。这样的设计能使隐藏层梯度值一直维持在一个稳定的范围, 缓解因网络加深而导致的梯度消失现象。

因此, SE-Densenet不仅能够通过“特征重标定”策略, 高效利用多维特征信息, 更能通过网络自身的反向传导机制, 减缓各隐藏层误差项的衰减, 保证梯度权值信息的稳定, 增强深层网络的学习和表达能力, 从而进一步提升网络性能。

3 基于SE-DenseNet的航班延误预测

航班延误预测模型结构如图3所示, 特征提取部分介绍详见第2节, 下面主要对数据预处理和分类预测部分展开描述。

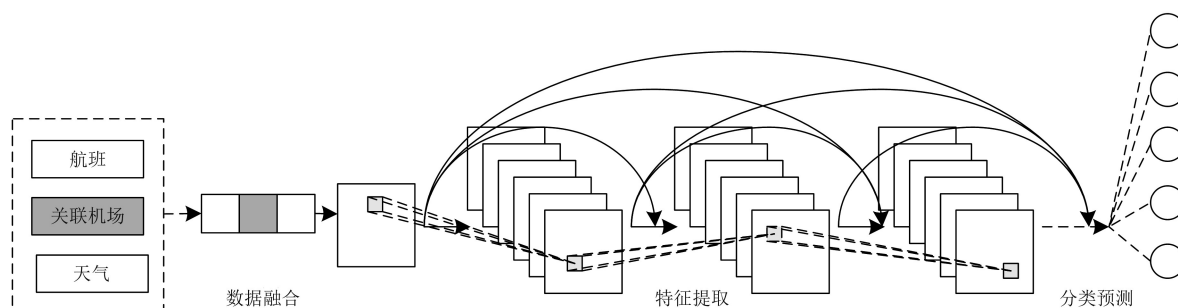


图3 航班预测模型总体结构

3.1 数据集介绍及预处理

本文所使用的数据集为美国交通运输统计局(Bureau of Transportation Statistics, BTS)提供的2016, 2017年的航班数据(Airline On-Time Performance data, AOTP)和美国国家气象数据中心(National Climatic Data Center, NCDC)提供的本地质控气象数据(Quality Controlled Local Climatological Data, QCLCD)。其中, 模型所需特征属性定义如下:

定义1 航班信息 F_O 主要包括起飞日期、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、轮挡/轮起时间、航程等24个属性。

定义2 相关机场延误属性 F_A : 选取对目标机场航班延误影响较大的机场作为相关机场。本文主要研究离港延误, 目标机场即为起飞机场。经过分析及专家知识, 按影响程度高低依次选取相关机场的离港平均延误时间、离港延误飞机架次、目标机场方向离港航班架次、目标机场延误总时间4个延误指标。

定义3 气象数据 F_W 为某航班起降机场对应时刻的天气状态, 包括风向、风速、能见度、温度、气压、观测站所在机场、观测时间等12个特征。

综合上述特征属性定义, 需要进行数据预处理。首先, 分别对原始航班信息和气象数据进行数据清洗; 其次, 选取目标机场的前十大相关机场, 根据定义2分别对其进行延误信息统计; 然后, 为研究特征 F_A 和 F_W 对延误的影响, 利用python数据分析技术将上述3种特征进行数据融合; 此外, 在利用SE-DenseNet进行特征提取时, 直接输入无序的离散数据会对网络内部训练机制产生误导, 混淆特征的重要程度, 离散数据采用按变量出现频次依次赋值的编码方式。而对于连续性数据, 通过min-max归一化的方式将特征值映射到 $[0, 1]$, 以去除特征间的量纲差异。最后, 将处理后的数据转换成适宜输入到CNN中的特征矩阵。

数据融合流程如下:

(1) 将 F_O 中的时间属性按照其所在州进行州时转换, 保证全部时间的时区一致;

(2) 选定 F_O 中的起飞机场、计划起飞时间为关联主键key1, 选定 F_W 中的观测站、观测时间为key2;

(3) 将 F_O 和 F_W 中具有相同key的数据执行merge操作, 构建融合数据 W ;

(4) 分别选定 F_A 和 W 中的日期以及起飞机场为关联主键key3, key4;

(5) 将上述二者中具有相同key的数据执行merge操作, 构建总特征 F 。

3.2 分类预测

分类预测模块主要包含3层网络结构: 全局池化层、全连接层、Softmax分类器^[13]。通过3层网络结构的组合操作, 将 $n \times n \times k$ 维度大小的矩阵转为对应每个类别出现的概率。其中, Softmax分类器按照表2延误时间 T 的划分方式进行延误等级预测^[3]。

表2 航班延误等级划分

延误等级	延误时间 T (min)
0 (未延误)	$T \leq 15$
1 (低度延误)	$15 < T \leq 60$
2 (中度延误)	$60 < T \leq 120$
3 (高度延误)	$120 < T \leq 240$
4 (重度延误)	$T > 240$

4 实验结果分析

实验计算机硬件配置为Inter®Xeon®处理器, 16 GB内存, GPU加速显卡为NVIDIA Tesla P100; 软件运行环境为64位Ubuntu14.04操作系统下搭建的Pytorch深度学习框架。

4.1 实验设置

本文实验使用的航班数据共3955056条, 特征维度为81, 最终经矩阵变换输入网络中的数据总维度为 $3955056 \times 9 \times 9$ 。此外, 由于真实世界中不延误的航班比例较大, 数据集每一类的样本都是不均匀的。在训练阶段采用文献^[14]提出的数据平衡策略进行小批量抽样, 以补偿类别的不均匀分布。

航班延误预测模型采用表1的网络配置信息进行训练, 为探索网络深度对模型性能的影响, 不同层数的配置可同理推出。本文模型超参数配置如下: 使用文献^[15]中的权值初始化策略; 为当前GPU设置随机种子, 以保证每次的训练结果一致; 模型优化器采用引入Momentum的随机梯度下降方法, 其中动量因子设为0.9, 权重衰减系数为0.0001, 学习率为0.1; 训练时批处理数量为256, 总迭代次数为 1×10^5 次。

4.2 维度变化比率 γ 对分类性能的影响

式(3)中的全连接权重 \mathbf{W}_1 和 \mathbf{W}_2 通过引入维度变化比率 γ 调控特征标定过程中的特征通道数目, 进而对模型中“SE”模块的计算成本以及网络性能产生影响。为了研究这种关系, 本文基于40层的SE-DenseNet模型对不同的 γ 值进行实验, 如表3所示。结果表明, 随着模型容量的增加不能直接提高

表3 维度变换比率 γ 对分类性能的影响

比率 γ	准确率(%)	模型大小(MB)
4	93.05	15.46
8	93.15	15.41
12	93.14	15.39
24	93.06	15.36

网络性能, 综合考虑准确性和模型复杂度, 当 $\gamma=12$ 时, 性能最佳。因此, 后续实验 γ 统一设为12。

4.3 SE-DenseNet结构验证分析

为验证改进后的SE-DenseNet在整体性能方面优于传统的DenseNet, 对两种网络进行对比。实验结果分别从算法复杂性、分类准确率、损失值大小、隐藏层梯度值4方面进行分析。

4.3.1 算法复杂性

算法复杂性主要体现在运行该算法所需要的计算资源的多少。模型参数越多, 所需计算资源越多, 算法则越复杂。表4列出了22, 40, 58层的SE-DenseNet和DenseNet的模型参数。从表4中可以看出, 模型参数至多仅有0.36%的增长。因此, 在现有网络结构中嵌入SE模块而导致的参数和计算量的增长可忽略不计, 其算法复杂性也与改进前基本持平。

表4 算法改进前后模型参数对比

网络层数	DenseNet	SE-DenseNet	参数增长比率(%)
22	1608469	1614256	0.36
40	3978167	3989741	0.29
58	7116085	7133446	0.24

4.3.2 分类准确率

深度网络主要通过测试集的最终准确率验证模型拟合程度, 即所有预测正确的航班数据之和与测试集总数之比, 表示为

$$w = \sum_{i=0}^{q} TP_i / N \quad (15)$$

其中, N 表示测试集总数, q 为延误等级数, TP_i 表示实际延误等级为 i 的航班中, 被预测正确的航班条数。

表5列出了不同层数的SE-DenseNet和DenseNet在数据集 F 上的准确率大小。可以看出, SE-DenseNet每一层的准确率均高于DenseNet, 当网络达到58层时, 准确率为93.19%。为进一步验证改进后算法模型的稳定性及可训练深度, 分别对88, 100, 124层的网络进行训练, 如表6所示。结果表明, 深度SE-DenseNet模型性能一直保持稳定, 并

表5 算法改进前后分类准确率对比(%)

网络层数	DenseNet	SE-DenseNet
16	91.86	92.33
22	92.07	92.69
34	92.28	92.80
40	92.57	93.14
52	92.69	93.17
58	92.72	93.19

表6 深层SE-DenseNet准确率

网络层数	SE-DenseNet(%)
88	93.19
100	93.17
124	93.19

且针对此数据集，模型从58层开始，准确率一直保持在93.19%左右。

4.3.3 损失值

网络一般通过训练集的损失值大小判断模型最终是否收敛，损失值越小，说明模型预测值与真实值大小越接近。图4(a)和4(b)分别表示22, 40, 58层的DenseNet和SE-DenseNet随迭代次数变化的损失值大小。可以看出，随着迭代次数的增加，模型逐渐趋于收敛，损失值最终减小到一个固定的范围。

DenseNet的损失值减小至0.2左右，改进后的SE-DenseNet最终可减小至0.1以下，说明SE-DenseNet较传统的DenseNet能更好地训练深层模型。

4.3.4 隐藏层梯度值

深度模型在反向传导的过程中第1个隐藏层的梯度值随迭代次数变化能一直保持稳定，则表征该网络内部各隐藏层均能保证信息的有效传输，模型具有较强的表达能力。图5(a)和5(b)分别表示22, 40, 58层的DenseNet和SE-DenseNet中第1个隐藏层对应的梯度平均值变化曲线。可以看出，改进后的SE-DenseNet隐藏层的梯度值能一直保持稳定，恰好符合上文的理论分析和梯度值计算式(5)–式(14)的推导结论。DenseNet变化曲线也与文献[8]中指出的能有效解决梯度消失现象的理论一致。

4.4 航班延误影响因素分析

为研究相关机场延误属性 F_A 和气象信息 F_W 对航班延误的影响，分别将仅考虑航班信息的数据集 F_O 、只融合气象信息 F_W 的数据集 W 以及综合考虑三者信息的数据集 F 输入到SE-DenseNet模型中进行延误预测，实验结果如表7所示。

由表7可以看出，与仅考虑航班信息相比，综合考虑三者信息后预测准确率在各层均增加1.8%左右。其中，如果只考虑天气因素预测精度增加约1%，增加相关机场延误信息后准确率能增加0.8%。

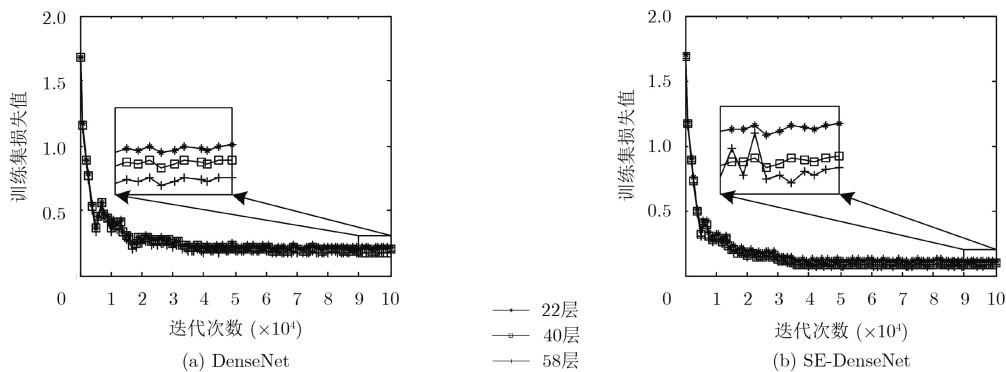


图4 改进前后网络结构损失值对比

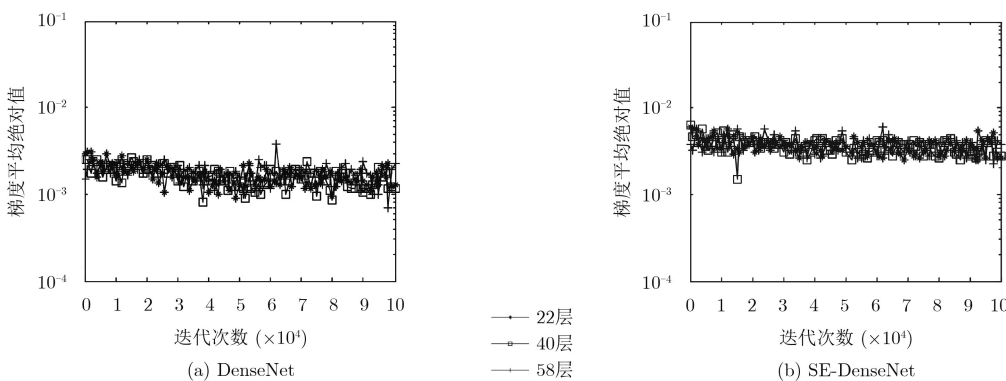


图5 改进前后网络结构梯度值变化对比

表7 不同影响因素预测准确率对比(%)

网络层数	数据集 F_O	数据集 W	数据集 F
16	90.64	92.06	92.33
22	90.96	92.15	92.69
34	91.06	92.29	92.80
40	91.09	92.33	93.14
52	91.24	92.35	93.17
58	91.36	92.38	93.19

文献[16]采用与本文同样的数据集研究气象信息对航班延误的影响,发现气象特征影响程度由高到低依次为风向、风速、能见度、温度、气压以及天空状况等。但由于SE-DenseNet模型具有高效的特征提取及数据处理能力,因此不对数据特征按影响力大小进行降维,尽量全面考虑可能对延误产生影响的因素,以提高模型最终的预测性能。

4.5 不同航班延误预测模型对比

为验证基于大数据的深度学习方法在数据处理以及预测精度方面与传统算法相比有较大优势,分别对不同的航班延误预测模型^[2-4]进行结果对比,如表8所示。对比可知,基于深度学习的SE-DenseNet模型预测准确率有较大提升。

表8 不同预测模型结果对比(%)

模型算法	预测准确率
C4.5决策树	78.05
支持向量机	80.00
人工神经网络	86.30
SE-DenseNet	93.19

5 结束语

本文提出了一种基于深度SE-DenseNet的航班延误预测模型,并且通过大量的实验验证了模型的有效性,结论如下:首先,在航班信息的基础上融合相关机场延误信息和气象数据后预测精度能提高近1.8%,其中,如果只融合天气数据,预测准确率提高1%。因此,二者都能对延误产生影响,但天气因素影响较大。其次,改进后的SE-DenseNet既能缓解深层网络梯度消失的问题,又能实现特征通道的自适应权重标定,最终达到提升网络整体性能的目的。最后,基于大数据的深度学习方法较传统算法可在无需特征降维的情况下,充分挖掘航班、相关机场以及天气因素的隐藏价值,达到更高的预测精度。但是,由于民航相关部门数据的保密限制,本文并未使用国内数据进行建模。目前正在与

国内空管部门积极洽谈合作,今后的研究重点将转为针对我国民航具体情况的分析研究。

参考文献

- [1] 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. doi: [10.1016/j.ifacol.2016.07.060](https://doi.org/10.1016/j.ifacol.2016.07.060).
- [2] KHANMOHAMMAD S, TUTUN S, and KUCUK Y. A new multilevel input layer artificial neural network for predicting flight delays at JFK airport[C]. *Complex Adaptive Systems*, Los Angeles, USA, 2016: 237–244. doi: [10.1016/j.procs.2016.09.321](https://doi.org/10.1016/j.procs.2016.09.321).
- [3] 程华, 李艳梅, 罗谦, 等. 基于C4.5决策树方法的到港航班延误预测问题研究[J]. *系统工程理论与实践*, 2014, 34(S1): 239–247. doi: [10.12011/1000-6788\(2014\)s1-239](https://doi.org/10.12011/1000-6788(2014)s1-239).
CHENG Hua, LI Yanmei, LUO Qian, *et al.* Study on flight delay with C4.5 decision tree based prediction method[J]. *System Engineering – Theory & Practice*, 2014, 34(S1): 239–247. doi: [10.12011/1000-6788\(2014\)s1-239](https://doi.org/10.12011/1000-6788(2014)s1-239).
- [4] 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. doi: [10.3321/j.issn:1000-6893.2009.07.014](https://doi.org/10.3321/j.issn:1000-6893.2009.07.014).
XU Tao, DING Jianli, GU Bin, *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. doi: [10.3321/j.issn:1000-6893.2009.07.014](https://doi.org/10.3321/j.issn:1000-6893.2009.07.014).
- [5] MANNA S, BISWAS S, KUNDU R, *et al.* A statistical approach to predict flight delay using gradient boosted decision tree[C]. 2017 International Conference on Computational Intelligence in Data Science, Chennai, India, 2017: 1–5. doi: [10.1109/ICCIDS.2017.8272656](https://doi.org/10.1109/ICCIDS.2017.8272656).
- [6] KIM Y J, CHOI S, BRICENO S, *et al.* A deep learning approach to flight delay prediction[C]. 35th Digital Avionics Systems Conference, Sacramento, USA, 2016: 1–6. doi: [10.1109/DASC.2016.7778092](https://doi.org/10.1109/DASC.2016.7778092).
- [7] LECUN Y, BENGIO Y, and HINTON G E. Deep learning[J]. *Nature*, 2015, 521(7553): 436–444. doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [8] HUANG Gao, LIU Zhuang, and WEINBERGER K Q. Densely connected convolutional networks[C]. 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, USA, 2017: 2261–2269. doi: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
- [9] HU Jie, SHEN Li, and SUN Gang. Squeeze-and-excitation networks[OL]. <https://arxiv.org/pdf/1709.01507.pdf>, 2018.4.
- [10] IOFFE S and SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal

- covariate shift[C]. 32nd International Conference on Machine Learning, Lille, France, 2015: 448–456.
- [11] NAIR V and HINTON G E. Rectified linear units improve restricted boltzmann machines[C]. 27th International Conference on Machine Learning, Haifa, Israel, 2010: 807–814.
- [12] RUMELHART D E, HINTON G E, and WILLIAMS R J. Learning representations by back-propagating errors[J]. *Nature*, 1986, 323(9): 533–536. doi: [10.1038/323533a0](https://doi.org/10.1038/323533a0).
- [13] DUAN Kaibo, KEERTHI S S, CHU Wei, *et al.* Multi-category classification by soft-max combination of binary classifiers[C]. 4th International Workshop on Multiple Classifier Systems, Guildford, United Kingdom, 2003: 125–134.
- [14] SHEN Li, LIN Zhouchen, and HUANG Qingming. Relay backpropagation for effective learning of deep convolutional neural networks[C]. European Conference on Computer Vision, Amsterdam, The Netherlands, 2016, 467–482. doi: https://doi.org/10.1007/978-3-319-46478-7_29.
- [15] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]. 15th IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1026–1034. doi: [10.1109/ICCV.2015.123](https://doi.org/10.1109/ICCV.2015.123).
- [16] 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): 2100–2106. doi: [10.11772/j.issn.1001-9081.2018010037](https://doi.org/10.11772/j.issn.1001-9081.2018010037).
- WU Renbiao, LI Jiayi, and QU Jingyi. Flight delay prediction based on dual-channel convolutional neural networks[J]. *Journal of Computer Applications*, 2018, 38(7): 2100–2106. doi: [10.11772/j.issn.1001-9081.2018010037](https://doi.org/10.11772/j.issn.1001-9081.2018010037).
- 吴仁彪: 男, 1966年生, 教授, 博士, 研究方向为自适应信号处理、现代谱分析及其在雷达、卫星导航、空管中的应用.
- 赵 婷: 女, 1994年生, 硕士生, 研究方向为航空运输大数据、深度学习.
- 屈景怡: 女, 1978年生, 副教授, 博士, 研究方向为航空运输大数据、深度学习、神经网络.