

3D Convolutional Neural Network for Regional Precipitation Nowcasting

Kun Wu^{1,3}, Wei Liang², Shuqiang Wang^{1*}

¹Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen Guangdong

²Liaocheng Institute of Water Resource Survey and Design, Liaocheng Shandong

³University of Science and Technology of China, Hefei Anhui

Email: *sq.wang@siat.ac.cn

Received: Sep. 19th, 2018; accepted: Oct. 3rd, 2018; published: Oct. 10th, 2018

Abstract

Accurate regional precipitation forecast has been a very important issue in the field of meteorological services. The goal of short-term rainfall forecasting is to make accurate and timely predictions about the intensity of rainfall in local areas in the short-term future (e.g., 0 - 6 hours). Weather stations can issue emergency urban rainfall alerts and provide effective flood prevention information by integrating the predicted short-term rainfall data with the observed weather forecast meteorological data. In this paper, according to the surrounding historical rainfall data of automatic station detection and weather observation area of different heights above the Doppler radar echo extrapolation figure, we proposed a rainfall prediction model based on the deep learning method. Proposed model is based on 3D Convolution Neural Network, the established network model was applied to the regression problem of the rainfall forecast, use the appropriate index to evaluate the accuracy of model under the high precision of short-term rainfall forecast in a particular area. Through experiments, this model can accurately predict the short-term rainfall over the region. With the experiments under different network structure, the root mean square error of predicted value and observed value is below 6. The training model predicts stability in weather station data throughout the year.

Keywords

Deep Learning, 3D Convolution Neural Network, Rainfall Prediction

基于3D卷积神经网络的区域降雨量预报

吴 昆^{1,3}, 梁 伟², 王书强^{1*}

¹中国科学院深圳先进技术研究院, 广东 深圳

²山东省聊城市水利勘测设计院, 山东 聊城

³中国科学技术大学, 安徽 合肥

*通讯作者。

Email: sq.wang@siat.ac.cn

收稿日期: 2018年9月19日; 录用日期: 2018年10月3日; 发布日期: 2018年10月10日

摘要

准确的区域降水量预报, 在气象服务领域一直是非常重要的问题。短时降雨量预报的目标是在未来短期(0~6小时)内, 对当地区域的降雨强度进行精确和及时的预测。气象站通过预测的短期降雨量数据, 与观测的天气预报气象数据进行整合, 能够发布城市紧急降雨警报, 提供有效的防汛防洪信息。本文根据自动站检测的周边历年降水量数据, 以及气象站观测的区域上空不同高度的多普勒雷达回波外推图, 提出一种基于深度学习方法的降雨预测模型。所提出的模型基于3D卷积神经网络(3D Convolution Neural Network), 将所建立的网络模型应用于降雨预测的回归问题, 并利用合适的指标对模型精度进行评价, 对高精度下特定区域的短时期降雨量进行预测。通过实验, 在不同网络结构下进行分析对比实验预测值与观测值的均方根误差达到了6以下。该方法能够对区域上空未来短期的降雨量进行准确的预测。该训练模型在气象站整年的数据中预测稳定。

关键词

深度学习, 3D卷积神经网络, 降雨预报

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

精确的区域短期降雨量预报, 随着气象预报技术的发展成为了日益受到关注的问题。区域降水量预报的目标是在未来短期(例如 0~6 小时)内, 对当地区域的降雨强度进行精确和及时的预测。气象站通过预测的短期降水量数据, 与观测的天气预报气象数据进行整合, 能够发布城市紧急降雨警报, 提供有效的防汛防洪信息。

对于短期区域降水量预报而言, 天气事件本身具有不可重复性, 每日的降水量作为一个随机事件, 降水过程本身受到区域所处气候带、潮汐、大气环流等气候因素的影响[1]。在随机因素的干扰下, 短时间区域降水量的分布呈现出复杂的变化趋势, 因此需要高时空分辨率和高时效性的数据源来建立预测模型。实时的多普勒雷达回波外推图 6 min 更新的时间分辨率和 1 km 的空间分辨率可以满足这一需求, 由于所需要的预报数值精确度与时间有效性远高于其他传统气象预报任务, 通过多普勒气象雷达回波外推图像进行短时降雨预测, 在防汛和暴雨预警中扮演着重要角色[2] [3]。

降水预报系统的常规预报依赖于光流方法, 气象部门首先根据观测到的雷达回波外推图像通过光流计算对流云的运动, 然后用拉格朗日平流预测未来的雷达回波外推图像[4]。现有的利用光流法预测降水量, 由深度学习的角度来看是基于无监督的学习, 因为光流计算没有利用现有的大量雷达回波数据。降水量预测问题可以归纳为一种时间序列预测问题, 将过去的雷达回波外推图像数据作为输入, 以将来的雷达回波外推图作为固定数量输出的序列, 这种训练与预测问题能够使用有监督的深度学习技术加以解决。

本文的工作主要包含以下两点:

1) 建立短时降雨预测的深度神经网络模型。气象预报的降水过程具有规律较难学习、数据量大以及变化趋势复杂的特点,因其本身具有时序的属性,适合深度神经网络对数据进行训练与学习。通过实时的多普勒雷达图像,根据建立深度神经网络模型对未来短时期的降水量进行预测。

2) 卷积神经网络应用降雨预测学习模型的回归问题。卷积神经网络广泛应用与图像的识别与分类领域,图像多分类任务中输出结果的一般是分类的概率。基于降雨量预测目标建立的网络模型,将目标站点的未来降水量作为标注,训练线性模型,输出结果为降水量的误差值。因此,课题研究所建立的网络模型将应用于降雨预测的回归问题,调整优化网络模型的参数并建立合适的指标对模型精度进行评价。

2. 相关工作

2.1. 降雨量预报研究

在区域内降雨量预测的领域研究中,传统的短期降雨预测模型取得了许多成果。在气象预报发展的早期阶段,孙才志和夏乐天等学者分别用加权马尔可夫链模型对降水量的状况进行了预测,并采取有序聚类方法对降水量的状态进行划分,避免了对降水量状态划分时受主观因素的影响[5]。赵欣和邹良超利用有序聚类的最优图像识别算法、模糊理论和加权马尔可夫链的原理,建立了基于有序聚类的模糊加权马尔可夫预测模型,并对近几年中部某城市的夏季降水量进行预测[6]。

Takasao 和 Nakakita 提出的短期降雨预测模型是基于云物理的概念性降雨模型。该模型利用雷达反馈的三维分布信息估算水汽转化为液态水的转化率,并预测降雨分布的产生,发展,下降和雨带等主要特征[7] [8]。陈森发和张文红等在此基础上,引入了随机误差项,利用随机微分方程理论,建立相应的随机波动模型,使模型能够更好地反映系统的动态变化,但是在提高短期的降雨预测准确度上,容易出现预测结果不稳定的情况[9]。

现代的降水预报模型大致可以分为两类,即基于数字天气预报(Numerical Weather Prediction, NWP)的方法和基于雷达回波外推的方法。对于 NWP 方法,在临近预报时间进行预测需要对大气模型中的物理方程进行复杂而细致的模拟[3] [10]。计算视觉技术尤其是基于光流法的方法,已被证明对精确推断雷达回波外推图像有较好的效果[11]。雷达回波的变分方法的实时光流(ROVER),利用图像序列中像素在时间域上的变化以及相邻帧之间的相关性,结合风向气候因素的影响,容易造成短时间的降水量变化,光谱的帧值快速的减少及消散,使得光流法对于区域降水量的预测有一定的局限性[12]。

基于深度学习的短期降水预报问题,日益成为降雨预测的一个重要方向。郭尚瓚等提出集成多点预测的卷积 MLP 模型进行短时降雨量预测[13]。张继学建立逐日降水量的三层 BP 神经网络预测模型,主要利用 BP 神经网络的线性函数逼近功能[14]。张帅等人运用 RNN 递归神经网络构建模型,通过时间反向传播算法更新网络权重解决长期降水量的预测问题,实验结果较传统方法预测精度高,较好地解决了高维数、非线性和局部极小问题[15]。深度学习在气象预报方面取得了一些成果,但仍需要进一步探索,如无标记数据的特征学习,雷达回波外推图像的图像识别与标注,网络规模与训练精度的平衡问题。

2.2. 卷积神经网络

深度学习的概念源于对神经网络的研究,M Riesenhuber 等人根据细胞的节点功能设计不同节点的分层网络结构,建立人工神经网络的模型[16]。2006年,加拿大多伦多大学教授 Geoffrey Hinton 和学生 Ruslan Salakhutdinov 提出了无监督预训练对权值进行初始化[17],对于每一层特征以低层的表示作为输入,再利用无监督的训练得到隐藏层的抽象的表示方法。通过有监督训练微调与无监督的权值初始化,深层网络训练中梯度消失问题得到解决[18]。在此基础上提出的深度置信网络(Deep Belief Networks, DBN),DBN 的训练方法降低了学习隐藏层参数的难度,经过逐层特征变换,增加了样本的空间特征表

示[19]。Yann Le Cun 提出卷积神经网络(Convolution Neural Network)后,深度学习方法越来越多地用于图像识别领域中[20]。卷积网络模型可以通过神经网络层次结构来学习特征,网络节点组合低层次的特征,抽象形成高层表示的属性与类别,利用无监督学习训练深度神经网络的各层结构,实现自动化的模型构造[21]。近年来随着深度学习在图像识别领域的广泛应用,深度学习模型使用无监督或有监督的方法进行训练,生成的模型在视觉对象识别,自然语言处理和音频对象分类领域获得了突出的成果。

3. 数据与方法

3.1. 实验数据源

我们把多普勒雷达图回波外推图作为样本,把自动气象站的降雨观测数据作为标签。本实验选取2016年3月~6月期间总共71天的自动站经纬度范围内的降雨量站点数据,以及向前的4小时多普勒雷达回波数据。选取原始雷达数据集中15个不同时序、4个不同高度通道的 50×50 尺寸大小的雷达回波图像,以及自动观测站对应X坐标、Y坐标的降雨量标签Hrain,生成训练集和测试集。为了减少内存的占用,提高训练速度,对输入的图像进行了归一化处理,即统一尺寸大小,并舍弃包含缺测值的雷达图像(图1)。

训练集的样本是从共计71天的雷达图回波图中,每次间隔6天采集数据样本。每份样本对当前t时刻的前15个时次(每个时次提取4个不同高度),以及4小时内的所有41个时次(每个时次提取6个不同高度)进行样本提取,共计8215份训练样本。测试集间隔7天采集数据样本,同理提取相同时序和高度的2720份测试样本。

通过对生成的数据集的标签数值进行归纳,降雨量的分布在 $0 \sim 70 \text{ mm/m}^2$ 的范围内,数据集的样本降雨量分布多数在 $0 \sim 60 \text{ mm/m}^2$ 范围内, 2 mm/m^2 的降水量样本量最多达到1500份, $20 \sim 60 \text{ mm/m}^2$ 区间内样本量呈现逐渐减少的趋势。整体的降雨量分布较均衡,测试集的样本在以上的所有样本中随机分配的,共计2720份样本分布在 $0 \sim 60 \text{ mm/m}^2$ 的各个区间中。训练集和测试集的降雨分布统计情况如图2所示。

3.2. 方法描述

本文使用的3D卷积神经网络(3D Convolutional Neural Networks)基于2D卷积神经网络结构,3D卷积神经网络可以在原始输入图像中应用卷积核和滤波器,通过卷积和池化操作,由层次网络结构实现对特征对象的学习。雷达回波外推图是基于三维空间维度(3D)的图像数据,对于雷达探测到的气象云图的特征像素点,既包含了时序信息也包含空间信息,所以基于3D卷积神经网络输入多维图像时表现的更为明显,多维图像能够直接作为神经网络的输入,避免了传统识别算法中复杂的特征提取和数据重建过程。

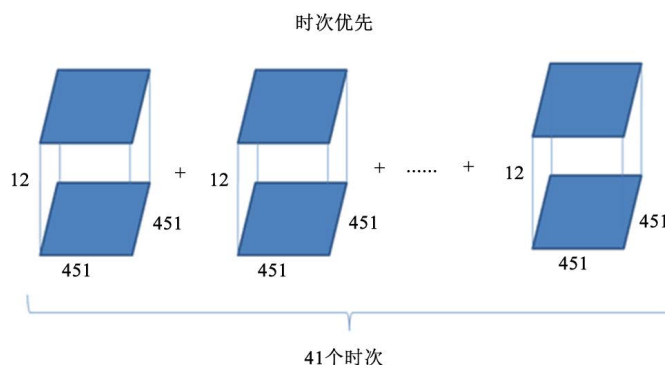


Figure 1. The timing and height of radar data
图1. 雷达图数据的时序与高度

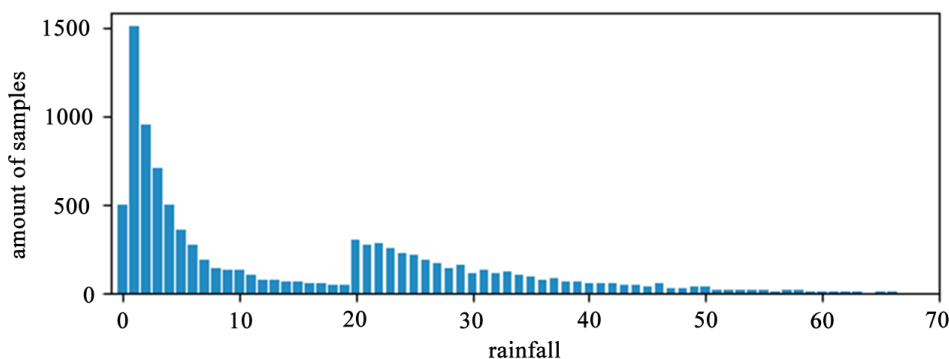


Figure 2. Dataset rainfall distribution
图 2. 数据集降雨量分布

3D 卷积神经网络包含卷积层、池化层、全连接层和 Softmax 层。3D 卷积层根据 2D 卷积神经网络的基础，对卷积神经网络维度进行拓展，网络结构中的每一层的卷积核尺寸、池化层过滤器尺寸均提升到三维。3D 最大池化层同样选取最大池化方法，接受的卷积层输出为一个立方体数据。全连接层中，神经元与邻接层所有神经元相连。全连接层的输入，将特征空间到一个神经元向量，接下来利用矩阵乘法将输入的特征向量，前一层的输入将卷积的 3D 特征向量压平到一个神经元向量。最后输出层是 Softmax 层，最后一个神经向量将计算每个分类的概率。

3D 卷积神经网络的损失函数：对于一个给定的样本，如雷达图 3D 训练样本集合 $(I^{(1)}, y^{(1)})$ $(I^{(1)}, y^{(1)}, \dots, (I^{(n)}, y^{(n)}))$ ，其中 $I^{(m)}$ 是三维输入向量， $y^{(m)}$ 是对应的标签， $\hat{y}^{(m)}$ 代表预测的输出， θ 代表所有参数，损失函数如下：

$$loss = -\frac{1}{N} \sum_{i=1}^N \sum_{c=0}^{C-1} indicator^j = cP(\hat{y}^j = c | I^{(j)}; \theta) \tag{1}$$

$indicator^j$ 表示指示函数， $P(\hat{y}^j = c | I^{(j)}; \theta)$ 表示样本 $I^{(j)}$ 属于分类 c 的预估概率，调节参数使得 $loss$ 最小即输出量精确度高。

多普勒雷达回波图数据集基于时间段分布收集获取的特点，决定了网络模型的特征提取中包含时序与空间信息。基于 3D 卷积神经网络与长短期记忆网络(Long Short-Term Memory, LSTM)，构建融合时序多特征的短期降雨量预测模型。长短期记忆网络(LSTM)通过雷达回波图的时间点多维图像数据和前序时段的特征，对于网络的时序特征进行提取。本文对 3D 卷积神经网络的网络结构进一步的修改，融合 LSTM 单元输出向量的空间状态提取特征，通过重塑向量的形式来表达图像的空间特征信息，实现多维度特征的聚合和节点信息的保存。

3.2.1. 3D 卷积神经网络结构

模型的网络结构由神经层和下采样层交替实现，神经层包含 3D 卷积层，卷积核的尺寸为三维立方体，下采样层包含 3D 池化层，输入层数据一般由特征图按时序堆叠而成。主要包括输入层、卷积层、池化层、全连接层，搭建包含 3 层 3D 卷积层、3 层 3D 池化层的卷积神经网络模型，使用线性整流函数 (Rectified Linear Unit, ReLU) 作为激活函数。具体的网络模型结构和各层参数如图 3 所示。

以 3D 卷积层的第一层 C1 为例，C1 的卷积核大小为 $2 \times 7 \times 7$ ，卷积层的过滤器节点数为 32，即为 32 个卷积核，通过内积运算得到一个特征体上对应位置的一个值，通过特征提取生成 32 个特征体。池化操作时，每张图按池化方式下采样，各特征图间不影响，即池化前有 32 张特征图，池化后还是 32 张。池化操作将高分辨的特征图像转化为分辨率较低的特征图像，通过池化层的处理，减少了最后全连接层

的节点个数。3D池化层P1使用 $2 \times 2 \times 2$ 的最大池化尺寸,则卷积层的输出矩阵维度变为 $32 \times 7 \times 25 \times 25$,后续卷积池化以此类推。

3.2.2.3D 卷积神经网络与 LSTM 时序单元模型

长短期记忆网络(Long Short-Term Memory, LSTM)基于循环神经网络(Recurrent Neural Network, RNN)的扩展[22],是用于处理随时间、空间等因素变化的序列数据的网络结构。LSTM网络能够解决数据模型的长期依赖问题,即数据的基于时间存在依赖关系。例如视频流的帧数据,根据前段的序列数据预测未来帧的时序序列数据,通过对时间点图像的空间特征以向量表达,输入到级联的LSTM单元中,经过LSTM单元的融合输出包含时序的节点信息。

通过对雷达图数据集的归纳,雷达图样本矩阵根据时序的特征,由 x_1 递增到 x_n 排列为 $(x_1, x_2, x_3, \dots, x_n)^T$ 的序列,融合数据标签值 $(y_1, y_2, y_3, \dots, y_n)^T$ 作为输入层的数据。根据时间维度排列的LSTM单元,将网络提取的特征值基于时序进行输入,LSTM单元节点通过门结构控制信息的传递并更新节点的状态,包括输入门、遗忘门和输出门[23]。通过门结构控制时序特征的遗忘与保留,参数的有关的信息被保留下来并赋予权重,与雷达图预测无关的特征信息则被忽略。构建的模型采用LSTM单元用于雷达图数据集的整体结构如图4所示。

通过上述LSTM单元时序序列向量特征提取的过程,基于3D卷积神经网络的网络层结构,在第三个卷积层C3后添加LSTM-Layer层。网络结构分为两个模块:3D-CNN模块和LSTM模块。其中3D-CNN模块的三层卷积层和池化层的参数如下:Conv1($2 \times 7 \times 7$)-Pool2($2 \times 2 \times 2$)-Conv2($2 \times 5 \times 5$)-Pool2($2 \times 2 \times 2$)-Conv3($2 \times 3 \times 3$)-Pool3($1 \times 3 \times 3$)。LSTM模块则由级联的16个单元组成,分别对应输出的特征向量矩阵。基于3D-CNN网络和LSTM单元的网络结构具体如图5所示。

雷达图数据集尺寸为 $50 \times 50 \times 15 \times 4$,按由4小时前到当前的时序排列。模型将输出的4阶向量重塑(Reshape)成 $50 \times 1 \times 64 \times 16$ 的尺寸,并采用16个LSTM节点与输出向量相连接,每个节点输入的向量为 $50 \times 1 \times 64$,利用长短时记忆单元对于节点信息的遗忘和选择,学习基于时序的序列输入向量特征。

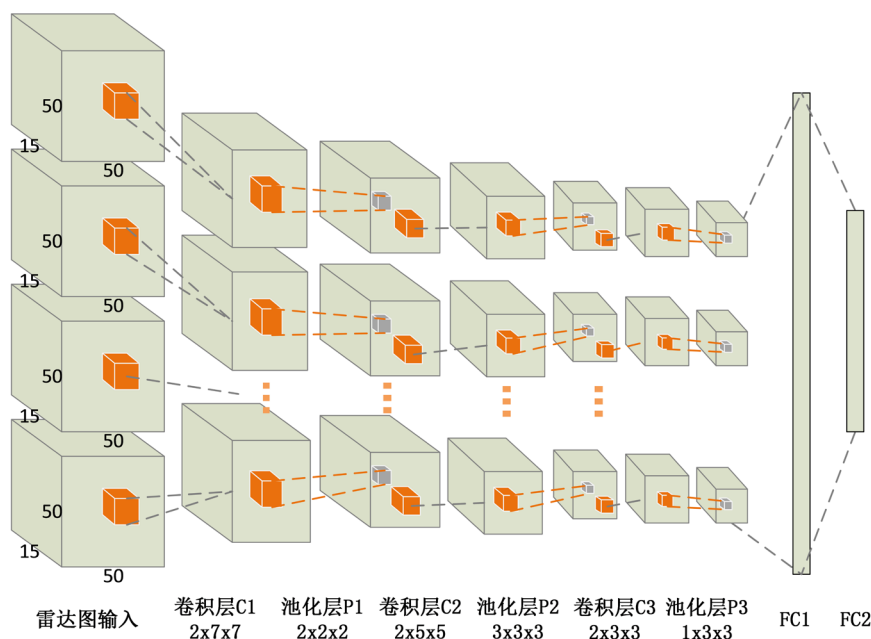


Figure 3. 3D convolution neural network model structure

图3. 3D卷积神经网络模型结构

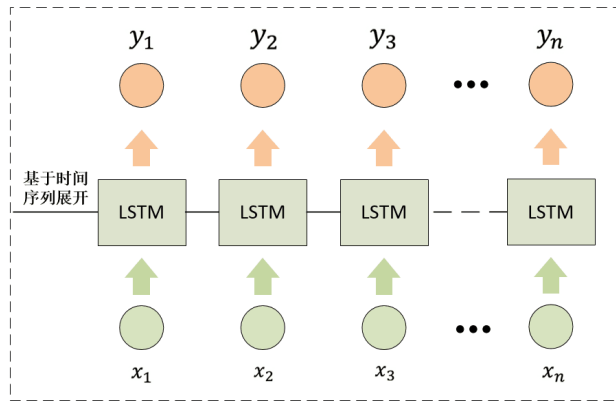


Figure 4. LSTM Unit based on time series
图 4. LSTM 单元时间序列结构图

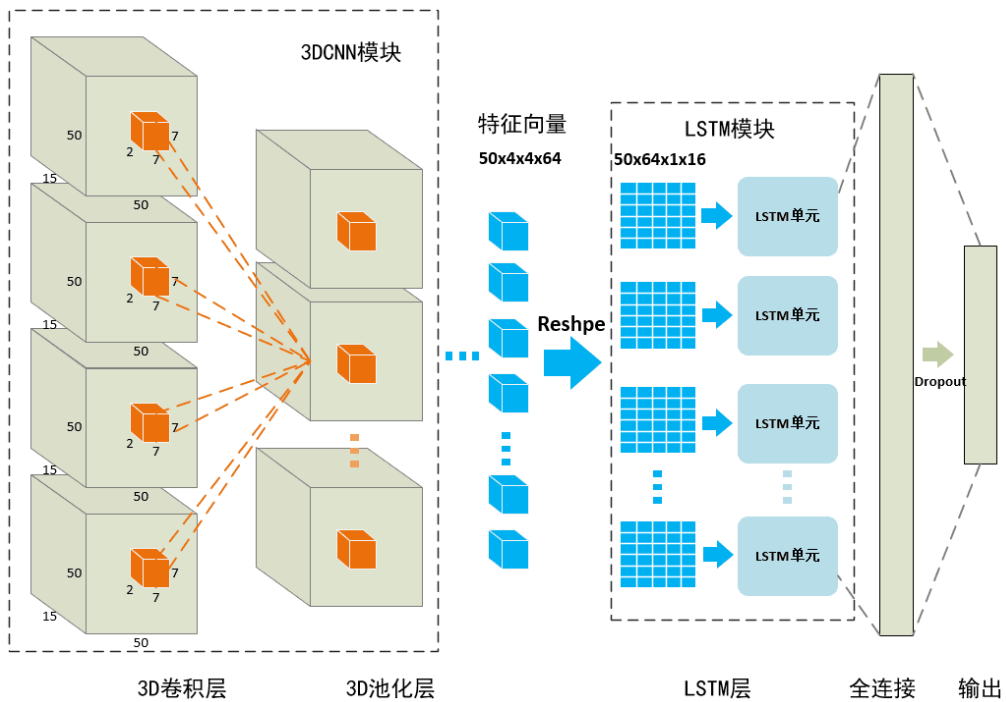


Figure 5. 3D-CNN and LSTM units Network structure
图 5. 基于 3D-CNN 和 LSTM 单元的网络结构

对图像序列在时序上建模,采用 16 个 LSTM 节点,每一个节点都与重塑后的 16 个输出向量集合相连接。LSTM 单元节点提取以时序特征向量为输入的序列图像,在输入门、遗忘门、输出门的信息节点循环计算中,通过时间序列的特征预测下一时段的雷达图降雨量标签值。

4. 实验与结果

4.1. 实验设置

实验采用的操作系统为 Ubuntu16.04, CPU 处理器的型号为 Inter (R) Xeon® CPU E5-1620, 主频 3.5 GHz, 显卡 GPU 型号 NVIDIA Quadro M4000 显存 8 G, CUDA8.0 和 cuDNN5.0 并行计算模型,深度学习语言框架为 Theano (Lasagne)。实验选取不同的卷积核尺寸 Conv3D 对比,使用 Adam 优化算法替代传

统随机梯度下降算法, 基于迭代过程更新网络权重参数并修正偏差, 权重的学习率为 0.001。一阶和二阶估计的指数衰减率分别为 0.9 和 0.999, 输入批次 Minibatch 设置为 80, 分别对 3D-CNN 的模型和 3D-CNN 加入 LSTM 单元后的模型进行实验。

4.2. 实验步骤与评价指标

实验训练过程中, 首先对 3D 卷积神经网络结构搭建与网络权重参数初始化, 将经过时序和高度提取的雷达图像重塑为合适尺寸的样本矩阵 $(x_1, x_2, x_3, \dots, x_n)^T$, 融合数据标签值 $(y_1, y_2, y_3, \dots, y_n)^T$ 作为输入层的数据。经过卷积层的卷积操作提取每批次样本的高阶特征信息, 并通过池化层下采样缩小特征矩阵的尺寸, 生成固定大小的特征图, 最后通过反向传播算法传播误差损失函数, 更新网络权重以及参数, 反复迭代使得梯度下降直至条件收敛。雷达降雨预测算法流程如图 6 所示。

本实验采用的评价标准, 主要采用机器学习领域中回归问题的均方根误差(RMSE)指标进行评价。针对不同降雨的定量预报数值的精确误差值, 预测确切的降水量以评估性能, 减少预报结果的误差。令 X_i 为预测向量, Y_i 为观测值的向量, N 为观测数据大小, RMSE 回归系统的拟合标准差, 计算公式如下。

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^n (Y_i - X_i)^2} \quad (2)$$

4.3. 实验结果

3D 卷积神经网络与 LSTM 单元的模型, 基于多层网络结构的有监督学习方法实现, 在应用过程中需要对图像进行特征提取、传播过程中权值共享, 因此设计合适的网络结构, 优化网络参数是提高特征识别准确率的重要因素。下面主要对卷积神经网络层的识别过程中网络层的深度、卷积核尺寸的大小、是否使用 Dropout 进行研究, LSTM 不同单元设置的影响, 设计多组实验并对结果进行对比分析。

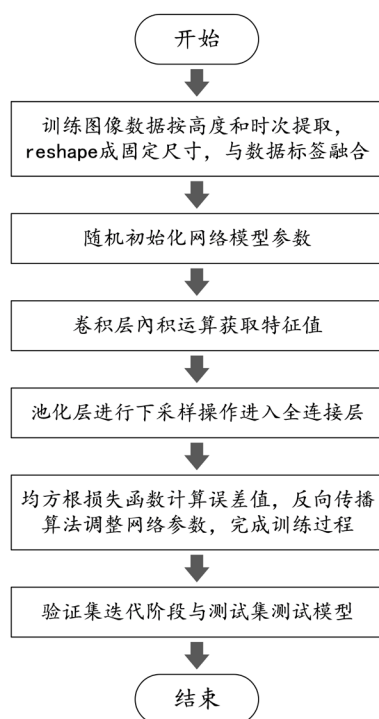


Figure 6. Overall flow chart of training and testing model
图 6. 模型训练与测试总体流程图

4.3.1. 权重参数

实验过程中为了得到较好的实验效果,对学习率的权重参数做出调整。迭代次数 Epoch 训练 100 次,对于输入批次的样本数量的模型学习率 Learning-rate 进行了多组实验,观察训练损失误差 Loss 的数值和趋势曲线,直到训练误差趋于稳定。可以看到当学习率曲线中 Learning-rate = 0.001 时,模型经过训练得到的模型测试性能效果最好。图 7 为设置不同学习率时模型的训练误差 Loss 下降的趋势,其中 a 学习率为 0.001, b 学习率为 0.01, c 学习率为 0.1。

4.3.2. 卷积层对比实验

实验过程中,影响卷积神经网络中卷积层的结构主要包含卷积核的尺寸、卷积层的深度。对于模型的训练过程,在实验 3D 卷积神经网络模型三层卷积核的基础上,设计 4 组不同的卷积核尺寸进行实验,分别为 Conv3D-a, Conv3D-b, Conv3D-c 和 Conv3D-d,网络层的深度和结构保持不变,对应的最大池化层也进行改变,具体的卷积核尺寸和网络结构如表 1 所示。

考虑上述表中不同卷积核的尺寸提取的特征值的区别,对全连接层(FC1)的节点数统一设置为 1024。每个卷积层链接最大池化层,3 层卷积核的过滤器 filter-size 的个数分别为 32, 64 和 128。激活函数采用 RELU,梯度算法优化器均为 Adam 优化算法。根据以上卷积核参数的网络结构,进行训练集与测试集的对比实验,确定最合适的卷积核尺寸结构。本文直接用优化后的卷积核参数进行实验仿真。RMSE 趋势图比较如图 8。

通过实验可以看出,卷积网络结构中 $(2 \times 7 \times 7)$ - $(2 \times 5 \times 5)$ - $(2 \times 3 \times 3)$ 的卷积核尺寸较为合适,Conv 3D-a 的网络收敛速度快, RMSE 的数值达到最低。说明对于气象雷达图数据集,卷积核的尺寸在大小为 $2 \times 7 \times 7$ 并递减的趋势下,卷积滤波通道学习特征图的信息较充分,模型表征能力强,预测回归值的性能最好。

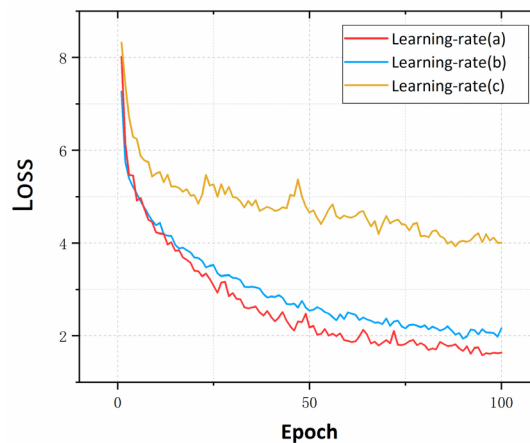


Figure 7. Comparison of training loss with different learning rates
图 7. 不同学习率的训练误差 Loss 比较

Table 1. Different convolution kernel sizes
表 1. 不同卷积核尺寸列表

网络结构	名称	卷积核尺寸	最大池化层
Conv1 + Pool1 + Conv2 + Pool2 + Conv3 + Pool3 + Fc1 + DP + Fc2	Conv3D-a	$(2 \times 7 \times 7)$ - $(2 \times 5 \times 5)$ - $(2 \times 3 \times 3)$	$(2 \times 2 \times 2)$ - $(3 \times 3 \times 3)$ - $(1 \times 3 \times 3)$
	Conv3D-b	$(2 \times 9 \times 9)$ - $(2 \times 7 \times 7)$ - $(2 \times 3 \times 3)$	$(2 \times 2 \times 2)$ - $(2 \times 3 \times 3)$ - $(1 \times 3 \times 3)$
	Conv3D-c	$(3 \times 3 \times 3)$ - $(2 \times 3 \times 3)$ - $(2 \times 3 \times 3)$	$(2 \times 2 \times 2)$ - $(2 \times 3 \times 3)$ - $(1 \times 3 \times 3)$
	Conv3D-d	$(2 \times 5 \times 5)$ - $(2 \times 5 \times 5)$ - $(2 \times 5 \times 5)$	$(2 \times 2 \times 2)$ - $(2 \times 2 \times 2)$ - $(2 \times 3 \times 3)$

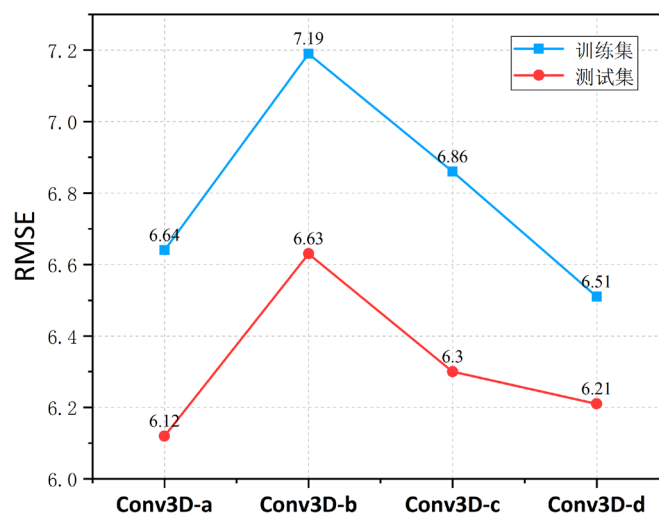


Figure 8. RMSE trend of the model under different convolution kernel sizes
图 8. 不同卷积核尺寸模型的 RMSE 趋势图

4.3.3. 梯度优化算法与 Dropout 层对比

深度学习中运用梯度下降的方法最小化目标函数，算法计算变量在训练过程中的梯度，更新参数对神经网络进行优化。常见的梯度优化算法包括随机梯度下降法(Stochastic Gradient Descent, SGD), Adam (Adaptive Moment Estimation)优化算法[24]。选取上述实验中较好的卷积核尺寸 Conv3D-a, 学习率设置为 0.001, 加入 Dropout 层进行优化算法的比较实验。实验的训练过程中的收敛趋势如图 10(a)所示。根据实验结果可以看出，常用的随机梯度下降法 SGD 在每次参数更新过程中，对相似的样本通过偏导运算求出梯度值，梯度收敛速度快而且波动较大，较大概率进入局部最优解。Adam 算法是自动使用模型学习率的算法，通过梯度的矩阵期望来估计梯度的取值范围，收敛速度较快，训练得到的模型效果更好。

Dropout 是神经网络进行模型参数平均的一种非常有效的方法[25]，通过 Dropout rate 概率随机采样输入的参数化特征实例，丢弃的是在神经网络中隐藏的不可见单元，使得这些选择的单元输出置为 0，从而对于网络参数进行筛选通过避免训练数据分布的协同适应，来减少神经网络中的过拟合。加入 Dropout 层的具体网络模型如图 9 所示。

设置初始的网络结构与加入 Dropout 层的对比实验，Dropout 频率设置为 0.5，迭代次数 100 次在其他网络参数条件相同的情况下进行实验。训练过程中 RMSE 下降的趋势如图 10(b)所示。

通过上图的方式改变全连接层之间的网络结构，在两层全连接层 FC1、FC2 层中间添加了 dropout 层，每次迭代对全连接 FC1 层输出的权重参数进行重新地选择和筛选，再输入到全连接层 FC2 中。当模型过度学习数据集的分布，拟合数据程度过高使得噪音和无效的参数也被输入到网络权重更新过程中。Dropout 随机更新参数的方法能缓解训练过程中的过拟合发生。

4.3.4. LSTM 单元

LSTM 单元对比实验，选用 3DCNN-LSTM 模型的网络结构。根据 3D 卷积神经网络算法的训练与测试流程，对训练集和测试集进行实验。根据实验的数据，选取上述实验中较好的卷积核尺寸 Conv3D-a, 学习率设置为 0.001, LSTM 单元数分别为 8、16、32、64 和 128, 并根据单元数重塑对应的向量矩阵维度。表 2 是 5 组不同 LSTM 单元对比实验的描述。

LSTM 单元对比实验中，4 阶向量 reshape 成 $50 \times 1 \times 64 \times 16$ 的尺寸，并采用 16 个 LSTM 节点与输出向量相连接，每个节点输入的向量为 $50 \times 1 \times 64$ 。后续实验依次类推。

构建的网络模型测试结果误差箱式图如图 11 所示。如上面箱式图的均值线和中位线所示，当趋势逐渐稳定，此时测试样本进行的测试误差 RMSE 在 6~7 左右震荡，在后续 LSTM 模块训练过程中，LSTM 单元数为 16 时误差达到最小值，稳定性也最好。训练损失 Loss 收敛速度快，迭代 100 数值后变动幅度较小。

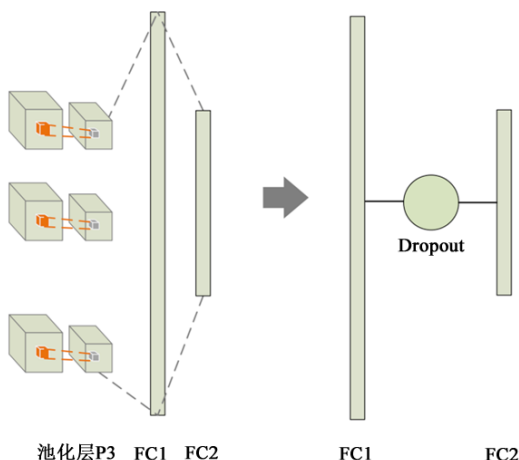


Figure 9. Dropout layer add to the network structure
图 9. 加入 Dropout 层的网络结构

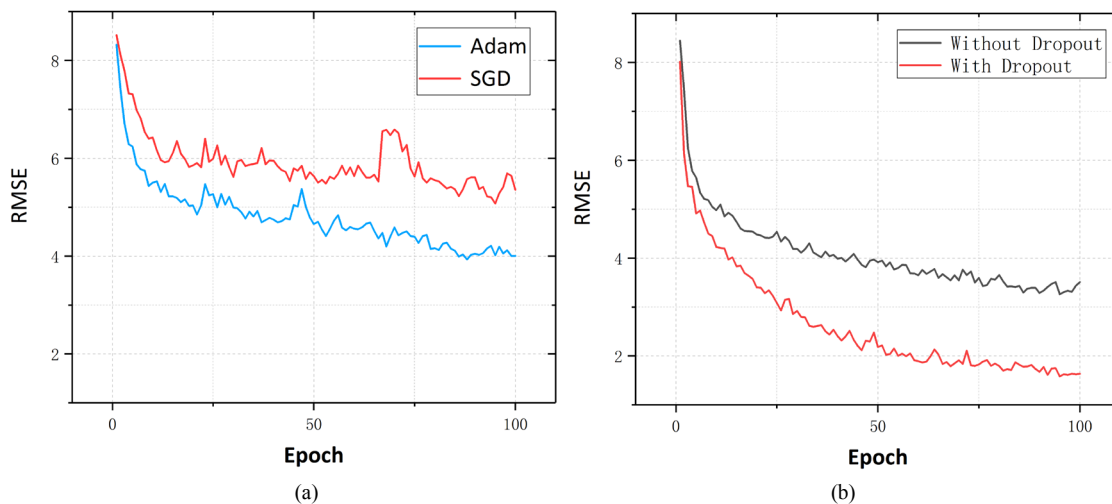


Figure 10. Different gradient algorithm and Dropout layer training loss
图 10. 不同梯度算法和 Dropout 层训练损失

Table 2. Comparison of different LSTM units
表 2. 不同 LSTM 单元对比

网络结构	输入向量	LSTM 节点数量	输出向量
3D CNN (Conv3D-a) - LSTM Layer - Fc1 + DP + Fc2	$50 \times 1 \times 128 \times 8$	8	$50 \times 1 \times 128$
	$50 \times 1 \times 64 \times 16$	16	$50 \times 1 \times 64$
	$50 \times 1 \times 32 \times 32$	32	$50 \times 1 \times 32$
	$50 \times 1 \times 16 \times 64$	64	$50 \times 1 \times 16$
	$50 \times 1 \times 8 \times 128$	128	$50 \times 1 \times 8$

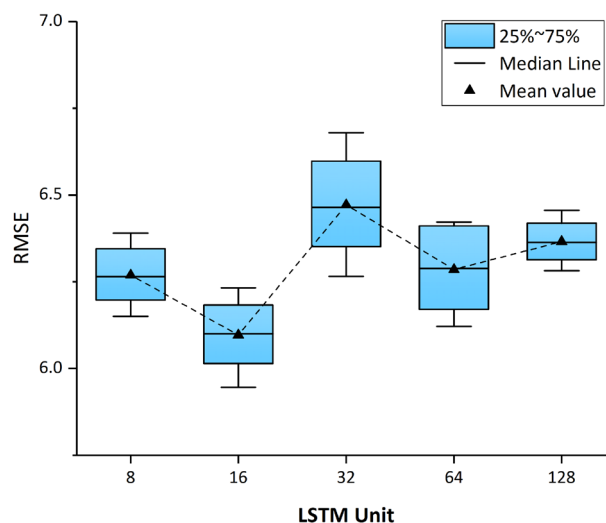


Figure 11. Different LSTM Units RMSE Error value
图 11. 不同 LSTM 单元的 RMSE 误差值

箱式图的均值线和中位线所示,当趋势逐渐稳定时测试样本进行的测试误差 RMSE 在 6~7 左右震荡,在后续的 LSTM 模块训练过程中, LSTM 单元数为 16 时误差达到最小值,稳定性也最好。训练损失 Loss 收敛速度快,迭代 100 数值后变动幅度较小。RMSE 的实验结果说明,在单位 1 小时内 50 × 50 公里的范围内,根据多普勒雷达云图预测降雨量值与实际观测的差值,均方根误差最好达到 5.9 左右。本文上述研究的优化方法中,基于 3D 卷积神经网络和 LSTM 的降雨预报模型达到最高精度。

5. 总结

通过本文研究,能够对区域上空未来短期的降水量进行预测,在不同维度的训练集和不同网络结构下进行分析对比实验,利用预测值与观测值的均方根误差评价模型。通过实验训练预估的模型,针对部分特定区域进行精确的降水量预报,对于提高气象预报工作的准确度,指导公共服务领域的预警机制,具有重要的实际意义和应用前景。本文完成的工作总结包括以下两个方面:

1) 利用自动站检测的周边历年降水量数据,以及气象站观测的区域上空不同高度的雷达回波外推图,结合 3D 卷积神经网络在多维图像处理的特征发现能力,提出了一个构建短时降雨预测的深度学习模型的方案。通过对模型网络结构的调整、网络参数的优化、加入 Dropout 层,解决样本量维度有限,训练过程中存在模型过拟合,特征学习不充分等问题,获得性能方面的优化和提升。

2) 基于 3D 卷积神经网络与长短期记忆网络,融合 LSTM 单元输出向量的空间状态提取特征,通过重塑向量的形式来表达图像的空间特征信息,优化网络结构并构建与时序相关的深度学习网络降雨量预测模型。实验结果说明,在单位 1 小时内 50 × 50 公里的范围内,根据多普勒雷达云图预测降雨量值与实际观测的差值,均方根误差最好达到 5.9 以下。

本文上述研究的优化方法中,基于 3D 卷积神经网络和 LSTM 的降雨预报模型达到最高精度。由于气象数据、模型性能等因素的限制,难免存在不足之处,如降雨时效性局限在一定小时以内,雷达数据包含气象因子较单一等问题,后续工作可以考虑设计多维雷达图像时空序列数据的方案,通过收集温度、湿度、风力等级等气象影响因子,构建包含天气多因素的气象数据模型,完善模型的数据涵盖维度。同时基于 1 小时及其以上的雷达气象数据,建立完整的短期降水量预测模型,解决区域降雨量预测精度不高的问题,对系统进一步优化以提高预报的时效性。

基金项目

本文工作受以下项目资助：深圳市孔雀技术创新项目(KQJSCX20170331162115349)，广东省自然科学基金(2016A030313176)，国家自然科学基金(61502473)。

参考文献

- [1] 张鹏程, 张雷, 王继民. 一种基于深度网络的多环境因素降水量预报模型[J]. 计算机应用与软件, 2017, 34(9): 240-245.
- [2] 张帅, 魏正英, 张育斌. 递归神经网络在降水量预测中的应用研究[J]. 节水灌溉, 2017(5): 63-66.
- [3] Deng, L. (2014) Deep Learning: Methods and Applications. *Foundations & Trends in Signal Processing*, 7, 197-387. <https://doi.org/10.1561/20000000039>
- [4] Sun, D., Roth, S. and Black, M.J. (2010) Secrets of Optical Flow Estimation and Their Principles. *IEEE Computer Vision and Pattern Recognition*, 2432-2439.
- [5] 孙才志, 林学钰. 降水预测的模糊加权马尔可夫模型及应用[J]. 系统工程学报, 2003, 18(4): 294-299.
- [6] 赵欣, 邹良超, 倪林. 基于有序聚类的模糊加权马尔可夫模型在降雨预测中的应用[J]. 江西农业学报, 2009, 21(2): 110-113.
- [7] Takasao, T. and Shiba, M. (1984) Development of Techniques for On-Line Forecasting of Rainfall and Flood Runoff. *Natural Disaster Science*, 6, 83-112.
- [8] Nakakita, E., Ikebuchi, S., Shiiba, M. and Takasao, T. (1990) Advanced Use into Rainfall Prediction of Three-Dimensionally Scanning Radar. *Stochastic Hydrology and Hydraulics*, 4, 135-150. <https://doi.org/10.1007/BF01543287>
- [9] 陈森发, 张文红, 张建坤, 等. 短期降雨预测的随机微分模型[J]. 系统工程学报, 2004, 19(3): 239-243.
- [10] Hayasaka, K., Tagawa, Y., Liu, T., et al. (2016) Optical-Flow-Based Background-Oriented Schlieren Technique for Measuring a Laser-Induced Underwater Shock Wave. *Experiments in Fluids*, 57, 179. <https://doi.org/10.1007/s00348-016-2271-0>
- [11] Sauter, T., Weitzenkamp, B. and Schneider, C. (2010) Spatio-Temporal Prediction of Snow Cover in the Black Forest Mountain Range Using Remote Sensing and a Recurrent Neural Network. *International Journal of Climatology*, 30, 2330-2341. <https://doi.org/10.1002/joc.2043>
- [12] Woo, W.-C. and Wong, W.K. (2017) Operational Application of Optical Flow Techniques to Radar-Based Rainfall Nowcasting. *Atmosphere*, 8, 48.
- [13] 郭尚瓚, 肖达, 袁行远. 基于神经网络和模型集成的短时降雨预测方法[J]. 气象科技进展, 2017(1): 107-113.
- [14] 张继学, 王鹏, 张琳, 等. 神经网络在短期降水预测方面的应用研究[J]. 科技风, 2016(17): 123-124.
- [15] Bengio, Y., Courville, A. and Vincent, P. (2013) Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35, 1798-1828. <https://doi.org/10.1109/TPAMI.2013.50>
- [16] Schmidhuber, J. (2014) Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61, 85-117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- [17] Hinton, G.E. and Salakhutdinov, R.R. (2006) Reducing the Dimensionality of Data with Neural Networks. *Science*, 313, 504-507. <https://doi.org/10.1126/science.1127647>
- [18] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) Learning Representations by Back-Propagating Errors. *Nature*, 323, 399-421.
- [19] Hinton, G.E. (2009) Deep Belief Networks. *Scholarpedia*, 4, 5947. <https://doi.org/10.4249/scholarpedia.5947>
- [20] Lecun, Y., Boser, B., Denker, J.S., et al. (2014) Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1, 541-551. <https://doi.org/10.1162/neco.1989.1.4.541>
- [21] Salakhutdinov, R. and Hinton, G. (2012) An Efficient Learning Procedure for Deep Boltzmann Machines. *Neural Computation*, 24, 1967. https://doi.org/10.1162/NECO_a_00311
- [22] Sundermeyer, M., Schlüter, R. and Ney, H. (2012) LSTM Neural Networks for Language Modeling. *Interspeech*, 601-608.
- [23] D'Informatique, D.E., Ese, N., Esent, P., et al. (2001) Long Short-Term Memory in Recurrent Neural Networks. *École polytechnique Fédérale de Lausanne*, 9, 1735-1780.
- [24] Kingma, D. and Ba, J. (2014) Adam: A Method for Stochastic Optimization.
- [25] Srivastava, N., Hinton, G., Krizhevsky, A., et al. (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15, 1929-1958.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2325-6753，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：jisp@hanspub.org