

基于深度学习的强对流高分辨率临近预报试验^{*}

郭瀚阳^{1,2} 陈明轩² 韩雷¹ 张巍¹ 秦睿² 宋林焯²

GUO Hanyang^{1,2} CHEN Mingxuan² HAN Lei¹ ZHANG Wei¹ QIN Rui² SONG Linye²

1. 中国海洋大学, 青岛, 266100
 2. 北京城市气象研究院, 北京, 100089
 1. *Ocean University of China, Qingdao 266100, China*
 2. *Institute of Urban Meteorology, CMA, Beijing 100089, China*
- 2018-07-19 收稿, 2018-12-13 改回.

郭瀚阳, 陈明轩, 韩雷, 张巍, 秦睿, 宋林焯. 2019. 基于深度学习的强对流高分辨率临近预报试验. *气象学报*, 77(4):715-727
Guo Hanyang, Chen Mingxuan, Han Lei, Zhang Wei, Qin Rui, Song Linye. 2019. High resolution nowcasting experiment of severe convections based on deep learning. *Acta Meteorologica Sinica*, 77(4):715-727

Abstract Nowcasting and early warnings of severe convective weather play an extremely important role in the prevention of meteorological disasters. In meteorological services, it becomes one of the most difficult and hot research topics because of its requirement for high accuracy and fine temporal-spatial resolution. Deep learning method is trying to solve the problem of strong convective nowcasting with high spatial and temporal resolution. First, the strong convective nowcasting is abstracted into a sequence of prediction problems containing both time and space. Then, the long sequence and high temporal resolution weather radar network mosaic data are fed to train the model using the encoder-decoder based on the improved recurrent neural network algorithm. The end-to-end neural network trained by radar echo data is then used to predict evolution characteristics of the radar echoes at 6 min intervals in the next one hour. By comparing the traditional extrapolation method with the deep learning algorithm, it is found that the deep learning method can effectively learn the intrinsic correlation of the data features in the high temporal-spatial resolution sequence, construct the abstract deep features through the multi-layer neural network and effectively capture the motion state of radar echoes. The calculation of prediction hit rate, false alarm rate, critical success index shows that compared with the traditional extrapolation forecast method, deep learning method can improve the nowcasting accuracy of strong convections, and demonstrates a wide application prospect.

Key words Deep learning, Nowcasting, Sequence prediction, Radar echo

摘要 强对流天气临近预报、预警在气象灾害防御中具有极为重要的地位。在气象业务中, 因对强对流天气临近预报、预警准确率和时、空分辨率的极高要求, 使其成为业务难点和研究热点之一。对于高时、空分辨率强对流临近预报问题, 尝试用深度学习方法来解决。首先将强对流临近预报抽象成同时包含时间和空间的序列预测问题; 然后基于改进的循环神经网络算法形成的自编码模型, 使用京津冀地区长序列、高时空分辨率天气雷达组网拼图数据进行模型训练; 最后利用基于历史 0.5 h 雷达回波拼图数据训练得到的端到端神经网络, 预报未来 1 h 内的逐 6 min 回波演变特征。通过基于传统外推算法的临近预报方法与深度学习算法的临近预报方法进行对比, 发现使用的深度学习方法可以有效“学习”到高时、空分辨率序列雷达数据特征的内在关联, 通过多层神经网络构造出抽象的深层特征, 能够有效捕捉到雷达回波的演变规律和运动状态。通过计算雷达回波预报的命中率(POD)、虚警率(FAR)、临界成功指数(CSI)等检验表明, 相较传统外推临近预报方法, 在强对流

^{*} 资助课题: 国家自然科学基金项目(41575050、40605015、41605031、41875049)、国家自然科学基金应急管理项目(41741013)、国家重点基础研究发展计划 973 项目(2011CB241503)、北京市科技计划课题(Z171100004417008)、山东省自然科学基金(ZR2016DM05)。

作者简介: 郭瀚阳, 主要从事深度学习与气象应用研究。E-mail: hyguo@ium.cn

通信作者: 陈明轩, 主要从事临近预报研究。E-mail: mxchen@ium.cn

回波临近预报准确率上有较明显提高。

关键词 深度学习, 临近预报, 序列预测, 雷达回波

中图法分类号 P435

1 引言

临近预报通常是指对某一区域未来短时间内的降雨量或强对流天气进行预测,在气象灾害防御中具有极为重要的地位。一方面,对于经常致灾的强对流天气,虽然进行了几十年的研究,但是因其演变规律极其复杂,依然有很多问题没有解决,包括强对流天气的准确预报、预警。另一方面,在目前的气象业务服务中,对强对流天气临近预报预警的准确率、时空分辨率、时效性都有极高要求。因此,如何获得快速、精准的强对流天气临近预报,成为气象业务的难点和研究热点之一(Sun, et al, 2014)。

陈明轩等(2004)、俞小鼎等(2012)、程丛兰等(2013)分别对强对流临近预报的方法进行了详细阐述,目前主要有两种方法:基于雷达回波的外推技术和数值天气预报模式。雷达回波外推技术是业务临近预报的主要方法,是指根据天气雷达探测到的回波数据来确定回波体的移动速度和方向。回波外推主要使用单体质心法(Lai, 1999)和交叉相关法(Han, et al, 2009)两种。单体质心法将目标简化为一点,适合跟踪预测大而强的回波目标,而当回波体发生融合分裂时,预报准确率迅速降低。交叉相关法则是计算两个临近时刻的空间优化相关系数,然后为所有的雷达回波建立拟合关系。它可以有效跟踪层状云降雨系统,但对于回波变化快的强对流过程,追踪准确度较低。总的来说,雷达回波外推在临近预报方面还是存在一定缺陷。数值天气预报是指根据大气实际情况,在一定的初始和边界条件下,通过大型计算机做数值计算,求解描述天气演变过程的流体力学和热力学方程组,预测未来一定时段的大气运动状态和天气现象的方法。数值预报虽然已经在气象领域得到广泛应用,但也有其自身的局限性(Weisman, et al, 2010)。特别是在临近预报应用方面准确率低,不够成熟,且需要复杂的物理方程计算,在精细化预测上难以满足准确性和实时性的需求。

鉴于上述两种方法存在的不足,本研究尝试使用深度学习方法来试验强对流临近预报。深度学习

作为当前炙手可热的人工智能研究方向,已经在技术和理论上取得重大突破(Lecun, et al, 2015),在很多领域展现出优秀的的能力,例如在计算机视觉、语音识别、自然语言处理、自动驾驶等领域的应用中,深度学习技术均已达到较高水平。这也为其他领域的创新发展提供了新的途径。

相对于传统的机器学习算法,深度学习的优势是能够通过学习算法对海量数据进行自动学习,从而挖掘到数据的内在特征及其蕴含的物理规律,而且不需要非常专业的相关领域的先验知识。深度学习本质上是堆叠多个隐藏层的深度神经网络,强调模型深度,通过逐层特征非线性变换组合,将原始特征映射到高维度的特征空间。由于模型的层次、参数很多,容量足够大,因此,模型擅长于表示大规模数据(海量数据),自动学习到其中有效的特征结构和物理规律,从而构建出高鲁棒性的机器学习系统。目前深度学习主要包括监督学习与无监督学习两种。例如,卷积神经网络(Convolutional Neural Networks, CNN)和循环神经网络(Recurrent Neural Networks, RNN)属于监督学习模型,而深度置信网络(Deep Belief Networks, DBN)是无监督学习模型。在大数据的时代背景下,深度学习在利用大数据获得更好的学习特征,在挖掘数据潜在的、丰富的信息关系和物理规律方面均有着很强的优势。

作为数据驱动型的技术,深度学习在解决天气领域的问题上同样具有很大的潜力,例如可以用来解决短时临近降雨问题。气象领域拥有大量高时空分辨率气象观测的历史数据,如雷达回波和地面观测站点资料,根据具体问题构建数据样本和对应标签,在计算密集型的服务器上使用这些数据来训练端到端的深度神经网络,虽然训练过程耗时较长,但训练好的神经网络模型则可以直接部署在业务平台上进行实时的预测产品的推送,甚至做到时间尺度为秒级的预测。

目前深度学习应用于临近预报的成果较少,如何利用深度学习来提高短时临近预报的效果仍是一个开放性问题。郭尚瓚等(2017)提出将多层感知器(Multilayer Perceptron, MLP)用于短时降雨预测,

用来预测某区域 36 min 内的降雨概率。基于卷积长短期记忆单元(Convolutional Long-Short Term Memory, ConvLSTM)首次提出使用改进的循环神经网络预测降水(Shi, et al, 2015),通过在长短期记忆单元(Long-Short Term Memory, LSTM)的模块(Hochreiter, et al, 1997)内部嵌入卷积计算来学习空间特征,Shi 等(2017)又在此基础上进行了改进。受其启发,本研究使用基于自编码(Sutskever, et al, 2014)的卷积 GRU(Convolutional Gated Recurrent Unit, ConvGRU)网络,利用历史 0.5 h 雷达回波数据来预测未来 1 h 的回波移动情况。试验结果表明,基于深度学习的方法比传统预测方法在预测精确度上有较大的提高。

目前应用到对流天气预报的深度网络模型包括三种:一种是卷积神经网络,通过将输入的格点天气要素看成是图像的形式,通过图像滤波器进行特征学习,充分考虑到空间结构的相关,但缺点是缺乏处理序列数据的能力,只适合处理“定长”数据。另一种是循环神经网络,这种经常应用于自然语言处理的模型通过一种自回归结构,能够灵活地处理序列数据,在时间维度上进行有效的学习,缺点也较为明显,和多层感知机一样,只能把输入特征用一维的向量进行表征,这样就丢失了格点数据固有的空间特征,学习能力较弱。最后一种是将上述两种模型进行不同形式的结合,能同时学习到空间和时间的特征,更加适合解决对流天气预报的问题,卷积长短期记忆单元模型和本研究的卷积 GRU 模型均属于这一类网络。

首先将临近预报问题转化为序列预测问题,并使用京津冀地区长序列天气雷达组网拼图资料训练由循环神经网络卷积 GRU 组成的自编码模型(Encoder-Decoder)。通过利用历史 0.5 h 雷达回波拼图数据训练得到的端到端神经网络,预报未来 1 h 内京津冀地区逐 6 min、1 km 分辨率的回波强度及回波形态。最后通过与传统外推临近预报算法进行检验对比分析,挖掘深度学习在临近预报方面的潜在优势。

2 模型

2.1 临近预报问题的模型构建

从机器学习的角度来看,临近预报是一种基于时间和空间范畴的序列预测问题。假设要持续地观

测某一个动态系统中 P 个变量(例如一段视频流中含有 RGB 通道即 3 个变量),那么每个变量在某一时刻可以由 $M \times N$ 的二维格点数据来表示(例如视频流中截取的某一帧数据的某一通道是由二维像素点组成)。考虑到变量这一维度,在任意时刻收集的 P 个变量可以看作一个三维的张量 $\mathbf{X} \in \mathbf{R}^{P \times M \times N}$ 。而从时间维度来说,不断地按照固定时间间隔来收集观测数据,即可以得到一组序列数据 $\{X_1, X_2, \dots, X_t\}$ 。序列预测问题由此可以抽象成在给定长度为 J 的历史观测序列数据,如何预测未来的长度为 K 的序列

$$\tilde{X}_{t+1}, \dots, \tilde{X}_{t+K} = \operatorname{argmax}_{X_{t+1}, \dots, X_{t+K}} P(X_{t+1}, \dots, X_{t+K} | \tilde{X}_{t-J+1}, \tilde{X}_{t-J+2}, \dots, \tilde{X}_t) \quad (1)$$

临近预报问题可以看作上述序列预测问题的一种情况,即使用历史(包括当前时刻)的气象观测资料,如雷达回波、温度、湿度、风向、风速等观测量来预测未来某一区域天气现象的演变过程。例如,在实际业务场景中,基于雷达回波外推技术的临近预报使用的是雷达回波数据,通常是逐 5—10 min 更新一次,通过用最近的若干帧雷达回波图像或数据来计算运动矢量,然后用来外推预测未来几个小时的雷达回波或降雨特征。

按照上述时间序列预测的建模方式,没有任何人工特征选取的过程,直接使用逐 6 min 的京津冀雷达回波拼图数据。利用历史 0.5 h 雷达回波图(连续 5 帧)来训练神经网络,并预测未来 1 h 雷达回波图(连续 10 帧)。使用深度学习模型的思路来解决临近预报问题。大致分为两个过程,首先构建模型的训练集进行模型的训练,其中,每个样本就包含 15 帧连续的雷达数据,也就是一个时间序列。前 5 帧作为模型的输入,模型从中进行特征学习,然后自动生成 10 帧的雷达图,而模型的任务就是输出的 10 帧与样本序列中的后 10 帧尽可能的一致,通过训练过程不断地学习。充分学习的模型就只需要根据输入的 5 帧雷达回波图给出未来 10 帧雷达回波图的合理预测,从而达到临近预报雷达回波特征即强对流演变特征的目的。

2.2 基于卷积的循环神经网络

对于序列预测问题,机器学习领域通常使用循环神经网络来解决,循环神经网络利用过去一段时间内的某些特征来预测未来一段时间内的某种状

态,是一类相对比较复杂的预测建模问题,其依赖于与事件发生的先后关系,挖掘时间维度的特征。相比于普通神经网络各层神经元节点相互独立的特点,循环神经网络的每一个隐藏层节点则同时依赖于当前的输入和上一时刻的中间状态,这样符合处理时间序列的规律,即处理新的输入的同时,也具备了记忆之前计算结果的功能。其中长短期记忆单元和 GRU(Gated Recurrent Unit)是最为常见的两种循环神经网络模型。它们都是可以有效解决资料序列间的长期依赖问题(Bengio, et al, 1994),长短期记忆单元(Hochreiter, et al, 1997)通过细胞记忆单元作为中间特征状态的“寄存器”,然后使用输入门、遗忘门和输出门这 3 个门限单元来控制信息流动,使得细胞记忆单元可以长时间记忆资料特征信息,即有用的信息不会因预测序列长度增大而丢失,并且会有选择地进行“更新”和“遗忘”。具体来说,每当有新的资料输入时,如果此时输入门处于激活状态,新的资料信息就会累积到细胞记忆单元。如果遗忘门处于关闭状态,那个过去的细胞状态就会被遗忘。同样地,更新后的细胞记忆单元通过输出门得到最终的隐藏层状态。使用这种存储单元和门控单元来控制信息流的一个优点是:反向传播(Rumelhart, et al, 1986)的梯度将被捕获在记忆单元中,可以有效防止其过快地消失,从而解决梯度弥散问题。2014 年,Cho 等(2014)提出了长短期记忆单元的一种简化版本——GRU,它只包含两个门:更新门和重置门,移除了输出门及隐层细胞状态。Chung 等(2014)通过多组对比试验发现,GRU 虽然参数更少,模型简单,但是在多个任务上都和长短期记忆单元有相近甚至更优的表现。

与长短期记忆单元相比,基于卷积的长短期记忆单元(卷积长短期记忆单元)具有更强的学习能力。通过把长短期记忆单元的矩阵乘法替换成卷积操作,卷积长短期记忆单元能够同时考虑到资料的空间和时间结构属性并进行特征学习,克服了长短期记忆单元用向量表征时资料空间结构信息丢失的问题。在卷积长短期记忆单元中,所有的序列输入 $\{X_1, X_2, \dots, X_t\}$ 、细胞状态 $\{C_1, C_2, \dots, C_t\}$ 、隐藏层输出 $\{H_1, H_2, \dots, H_t\}$ 以及控制信息流动的门控单元 $\{i_t, f_t, o_t\}$ 都是三维的张量 $\mathbf{R}^{P \times M \times N}$, P 表示输入变量的个数或者隐藏层的特征数量, M 和 N 代表输入的空间维度。和长短期记忆单元相同,在时间序

列迭代计算的过程中,卷积长短期记忆单元的内部状态是由当前时刻的输入和上一时刻的内部状态共同决定的。卷积长短期记忆单元的计算式为

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f) \quad (3)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (4)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \odot C_t + b_o) \quad (5)$$

$$H_t = o_t \odot \tanh C_t \quad (6)$$

式中, σ 代表 sigmoid 激活函数,用来将门限单元的值映射到 0—1, $*$ 和 \odot 分别代表卷积和点乘操作。假设卷积长短期记忆单元的内部状态被看作是某个移动物体的表征,则内部状态对应的卷积核越大,其输出的感受野越大,就可以跟踪快速移动的物体,反之则只能跟踪缓慢移动的物体。而上述门控单元输入门(i_t)、遗忘门(f_t)以及输出门(o_t)可以控制记忆单元(C_t)的信息流动。通过这种机制,反向传播的梯度可以像长短期记忆单元一样,在记忆单元中持续保留而不会产生梯度弥散的问题。

相对于长短期记忆单元,GRU 更加简洁高效,去掉了记忆单元和一个门控单元,在速度上的表现要优于长短期记忆单元,训练收敛时间更短(Chung, et al, 2014),且学习能力和长短期记忆单元相当。所以,文中使用基于卷积 GRU 进行建模

$$Z_t = \sigma(W_{xz} * X_t + W_{hz} * H_{t-1} + b_z) \quad (7)$$

$$R_t = \sigma(W_{xr} * X_t + W_{hr} * H_{t-1} + b_r) \quad (8)$$

$$H'_t = \tanh(W_{zh} * X_t + R_t \odot (W_{hh} * H_{t-1}) + b_h) \quad (9)$$

$$H_t = (1 - Z_t) \odot H'_t + Z_t \odot H_{t-1} \quad (10)$$

卷积 GRU 网络不再使用记忆单元,使用重置门(R_t)控制资料状态信息的“遗忘”,并和新的资料输入 X_t 组合成新的候选状态 H'_t ,最后使用更新门(Z_t)控制 H'_t 和上一时刻的状态 H_{t-1} 融合成新的隐藏状态(H_t)。使用下文即将介绍的自编码循环网络结构,通过多层卷积 GRU 模块的堆叠可以形成更加复杂的结构,有助于学习多层次特征,形成学习能力更强的模型。

2.3 序列预测问题的自编码模型

对于序列问题,通常用循环神经网络来解决,而

循环神经网络也可以作为基本单元组成更复杂的网络结构。短时临近预报可以看作是时、空序列的预测问题,而输入序列和输出序列长度不一定会相等,时间间隔也不一定相等,因此,使用一个循环神经网络模块同时针对输入、输出序列进行建模不合理,两者不能等同对待。文中使用自编码模型来解决短时临近预报这种等间隔的多步预测问题(Sutskever, et al, 2014)。一般来说,自编码模型主要包含编码阶段和解码阶段,多为对称结构,传统的自编码器的目的在于重建输入数据。例如,可以使用编码的结果对原始输入进行表征,实现输入数据的压缩。而本研究使用自编码结构主要进行数据生成,达到预测未来的目的。

自编码模型包括两个子模块(图1),编码模块将输入的所有时序数据映射成固定长度的隐藏层状态,即将所有输入序列的信息用最后一次输入的隐藏层状态来表征。然后用编码的最后一步的隐藏层状态来初始化解码模块的隐藏层状态。对于解码模块来说,初始化隐层状态后,就相当于加载了编码模块特征学习后的信息。首先用编码输入的最后一帧作为解码的第1个输入来得到关于未来的预测输出,然后解码后续的计算过程总是根据上一步的输出不断迭代生成新的预测结果,是一个自回归的计算过程,即当前时刻的输出又反过来作为下一时刻的输入,采取这种方法的优点是可以产生变长的预测序列。简言之,自编码模型首先通过编码将所有历史资料信息先编码成固定大小的向量,然后再根据不同的时刻解码有助于当前时刻预测的资料信息,这样可以避免“全连接”(所有的输入均与每一个输出相连接)方式带来的参数过量,一定程度地避免

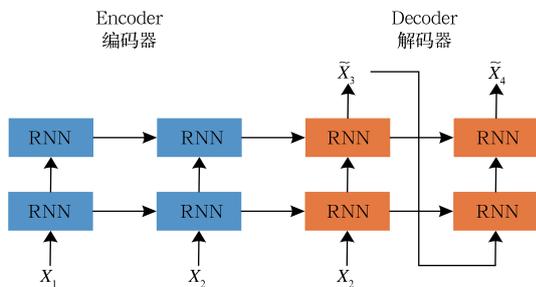


图1 两层循环神经网络组成的自编码模型网络
Fig.1 An Encoder-Decoder network composed of two layers

模型的过拟合,同时也符合预测的物理逻辑规律。编码和解码可以分别使用多层堆叠的循环神经网络模块(长短期记忆单元、卷积 GRU 等)来学习分层数据特征,即底层模块的输出作为高层模块的输入,这也是深度学习模型的一大优势所在,通过对底层特征进一步的非线性组合映射成高维特征,增加模型的学习能力,使其具有较强的时、空特征的表征能力,可以用来解决类似于强对流短时临近预报的问题。

3 数据

使用的数据包括京津冀地区 2010—2017 年 6 部新一代天气雷达(北京 S 波段 BJRS、天津 S 波段 TJRS、石家庄 S 波段 SJZRS、秦皇岛 S 波段 QHDRS、张北 C 波段 ZBRC、承德 C 波段 CDRC, 图 2)的三维拼图(CAPPI)数据。选取海拔高度 2.5 km 的雷达回波数据用于表征强对流回波的主要演变特征,覆盖京津冀地区(800 km × 800 km),水平分辨率 0.01° × 0.01°。雷达拼图数据流间隔为 6 min,逐日 240 帧数据,每一帧 800 × 800 分辨率。

首先对雷达拼图数据进行预处理,将雷达回波

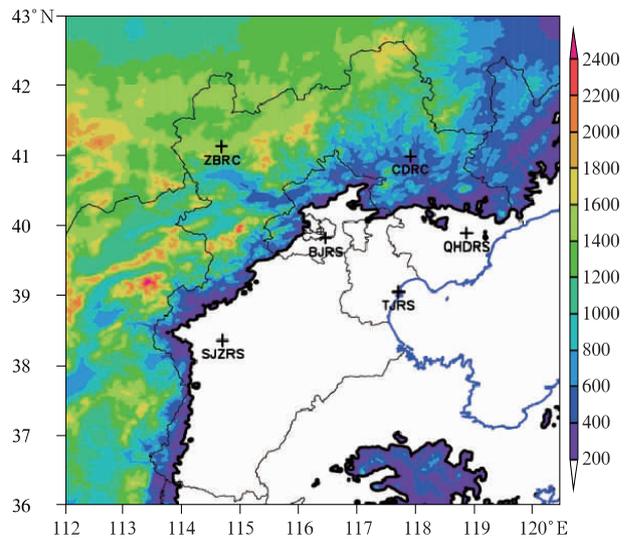


图2 研究区域地形
(色阶,单位:m;雷达资料由京津冀地区6部雷达收集,+为雷达位置)

Fig.2 The domain of this study
(color-shaded areas represent different altitudes (unit: m). Radar data are collected from 6 radar stations located in Beijing-Tianjin-Hebei region, where "+" symbols represent radar locations)

映射到 $[0, 1]$ 。主要选取京津冀地区暖季(6—9月)发生的所有强对流个例组成模型所需的数据集。其中,856 d 数据作为训练集;61 d 作为交叉验证集;最后 162 d 作为测试集检验模型效果。需要强调的是,训练集和验证集是从 2010—2016 年的数据中选取的,而 2017 年的数据单独构成最终的测试集,因为 2017 年数据并没有参与模型训练和调参,可以客观地验证模型的泛化和迁移能力以及临近预报效果(表 1)。

模型的任务是利用历史 0.5 h 的雷达回波资料来预测出未来 1 h 内逐 6 min 的雷达回波。雷达回波图作为模型训练输入的同时,也作为样本的标签来检验预测效果。所以对原始的序列数据使用长度为 15 的滑动窗口以步长为 1 进行滑动采样,得到包含 15 帧雷达图的序列样本集,其中 5 帧作为模型输入,10 帧作为模型标签,即模型预测的目标。这样的条件设定也符合真实的业务场景,因为在预测未来时刻的雷达回波之前,天气雷达总会给出近期的雷达回波探测资料。

表 1 数据集的详细信息

Table 1 Dataset details

	训练集	验证集	测试集
年份	2010—2016	2010—2016	2017
天数(d)	856	61	162
资料样本数	126558	8679	34977

4 试验

预测雷达回波的形状及运动趋势本身是个很有挑战性的任务。因为雷达回波反映的是强对流的瞬时演变特征,其本身没用固定的形状,且处于不断快速变化中,并且发展的过程中不断融合和分裂(雷暴合并及分裂)。试验的目标是利用历史 5 帧雷达回波图去预测未来 10 帧,即用历史 0.5 h 数据预报未来 1 h 内逐 6 min 回波演变。模型必须同时学习到雷达回波的时间和空间特征结构及演变规律,包括运动的趋势和周围的背景信息,才能准确预测出回波的移动速度和方向。文中的深度学习模型直接预测的是雷达回波图,这一点与基于雷达回波的外推技术相似,所以试验部分会使用一种交叉相关外推算法(CTREC),与卷积 GRU 的预测结果进行比较。陈明轩等(2007)、郑永光等(2013)对 CTREC 算法进行了深入研究,在此不再赘述。文中用

CTREC 算法与卷积 GRU 模型进行 0.5 h 和 1 h 的临近预报对比分析。但需要强调的是,CTREC 等外推技术通常是由两部分组成,先通过历史雷达回波数据计算出对应的雷达回波运动矢量,再用矢量图去做外推预测。而文中使用的卷积 GRU 是端到端的模型,直接使用蕴含强对流演变规律的“历史图”预测出未来的雷达图(强对流未来演变特征),预测效率更高。

4.1 试验方案

卷积 GRU 采用平均绝对误差(Mean Absolute Error, MAE)与平均平方误差(Mean Squared Error, MSE)作为训练模型的损失函数, $MAE = \frac{1}{n} \sum_{i=1}^n |y - \tilde{y}|$, $MSE = \frac{1}{n} \sum_{i=1}^n |y - \tilde{y}|^2$, 其中, \tilde{y} 为模型的预测值,而 y 为对应的标签,即真实值, n 为样本数。平均绝对误差和平均平方误差都是机器学习领域常用的损失函数,用来衡量预测结果和实况的相似度。如图 3 所示,首先是训练模型,假设当前时刻为 t ,模型输入的 5 帧雷达回波数据为 $\{X_{t-4}, X_{t-3}, \dots, X_t\}$,模型的输出为未来的 10 帧雷达回波图 $\{X_{t+1}, X_{t+2}, \dots, X_{t+10}\}$,对应的雷达回波实况图为 $\{\tilde{X}_{t+1}, \tilde{X}_{t+2}, \dots, \tilde{X}_{t+10}\}$,然后根据每一帧的所有元素计算的损失函数求和,最终计算出平均每帧雷达回波预测图所具有的损失函数值,并通过训练过程中使损失函数值不断降低。训练模型使用常用的反向传播算法 Adam(Kingma, et al, 2014)进行训练学习。训练完成后,模型就可以根据某一时刻的 5 帧输入,给出未来 10 帧回波图的合理预测,即图 3

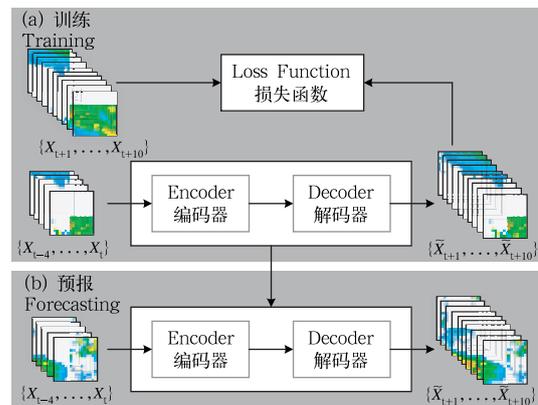


图 3 试验整体流程

(包括训练过程和预报过程)

Fig. 3 Flowchart of the experiment,

which contains training and forecast processes

中的预报阶段。训练过程的参数细节如下:初始学习率 10^{-3} , 学习率惩罚因子 0.5, 即每当验证集的损失函数值停止下降时, 将学习率减半。批量大小设为 4, 最终训练迭代数为 10 万次。试验使用四层卷积 GRU 组成的自编码模型网络, 其隐藏层节点分别为 64、64、128、128。卷积核尺寸分别对应为 5×5 、 5×5 、 3×3 、 3×3 。使用 2010—2016 年雷达拼图数据进行模型训练, 最终在 2017 年雷达拼图数据上进行预报检验。

4.2 检验标准

为了检验模型预测结果, 引入多个评价指标。在气象领域中, 命中率 (POD)、虚警率 (FAR) 和临界成功指数 (CSI) 经常被用于评估预报效果 (黄伟等, 2009; 程丛兰等, 2013; Chen, et al, 2017)。这些指标都是基于二分类混淆矩阵 (表 2) 计算得出。因为模型最后输出的是雷达回波, 属于回归问题, 不是 0/1 分类结果, 需要首先定义一组回波阈值 (dBz) 把最终的预测结果分解成多个二分类问题 (大于阈值的点设为正例, 反之设为负例), 然后根据二分类混淆矩阵计算评分。假设检验的雷达回波阈值为 K dBz, TP 表示观测不小于 K 且预测也不小于 K 的次数, FN 表示观测不小于 K 而预测小于 K 的次数, FP 表示观测小于 K 而预测不小于 K 的次数, TN 表示观测小于 K 且预测小于 K 的次数。本试验使用的阈值分别为 10、20、30、40 dBz。评分标准

计算式为

$$CSI = \frac{TP}{TP + FN + FP} \quad (11)$$

$$POD = \frac{TP}{TP + FN} \quad (12)$$

$$FAR = \frac{FP}{TP + FP} \quad (13)$$

表 2 二分类问题下的混淆矩阵

Table 2 The obfuscation matrix under the binary classification problem

	预测为正类	预测为负类
观测为正类	True Positive (TP)	False Negative (FN)
观测为负类	False Positive (FP)	True Negative (TN)

4.3 试验结果检验

深度学习模型卷积 GRU 和交叉相关外推算法 CTREC 在测试集上的结果如表 3、4 所示。可以看出, 在 0.5 和 1 h 预测中, 卷积 GRU 在每一个阈值的命中率、临界成功指数均高于 CTREC 算法, 而虚警率却更低。说明卷积 GRU 可以学习到雷达回波数据集中有效的强对流演变时、空特征, 从不同时间序列输入数据中也能学习到雷达回波的运动趋势和强对流的演变趋势, 进而准确预测出雷达回波的移动位置和形状变化。

4.4 试验个例分析

为了更直观分析预测效果, 从测试数据集中选

表 3 0.5 h 预测评分

Table 3 Half-hour prediction scores

回波阈值 (dBz)	命中率		虚警率		临界成功指数	
	卷积 GRU	CTREC	卷积 GRU	CTREC	卷积 GRU	CTREC
10	0.70	0.52	0.19	0.31	0.60	0.38
20	0.76	0.55	0.28	0.32	0.59	0.35
30	0.64	0.39	0.44	0.47	0.43	0.32
40	0.40	0.35	0.49	0.54	0.29	0.22

表 4 1 h 预测评分

Table 4 1 h prediction scores

回波阈值 (dBz)	命中率		虚警率		临界成功指数	
	卷积 GRU	CTREC	卷积 GRU	CTREC	卷积 GRU	CTREC
10	0.64	0.52	0.25	0.29	0.53	0.37
20	0.69	0.48	0.37	0.38	0.50	0.35
30	0.52	0.36	0.55	0.61	0.31	0.23
40	0.22	0.18	0.64	0.79	0.16	0.11

取3组个例进行分析。卷积GRU模型本身是通过历史的5帧雷达图而生成未来的10帧预测。如图4所示,0—1 h,模型始终能较为准确捕捉到不同强度回波的演变特征和移动位置,说明模型能够“学习”并“理解”强对流天气的演变规律,从训练集中学习到的强对流演变特征能有效应用于“未来”新场景,有很好的泛化能力。而且,对于输入雷达图中存

在的“噪声”(雷达杂波),卷积GRU能有效识别并去除(实况图中左上部分的噪点),这要归功于模型的卷积操作算法固有的先进滤波特性(Zhang, et al, 2018)。值得注意的是,图中右下方刚进入研究区域的雷达回波,卷积GRU同样也能持续准确地跟踪新生回波的移动,这是因为在海量数据的训练集中,模型也会遇到很多次类似的情形,所以可以学

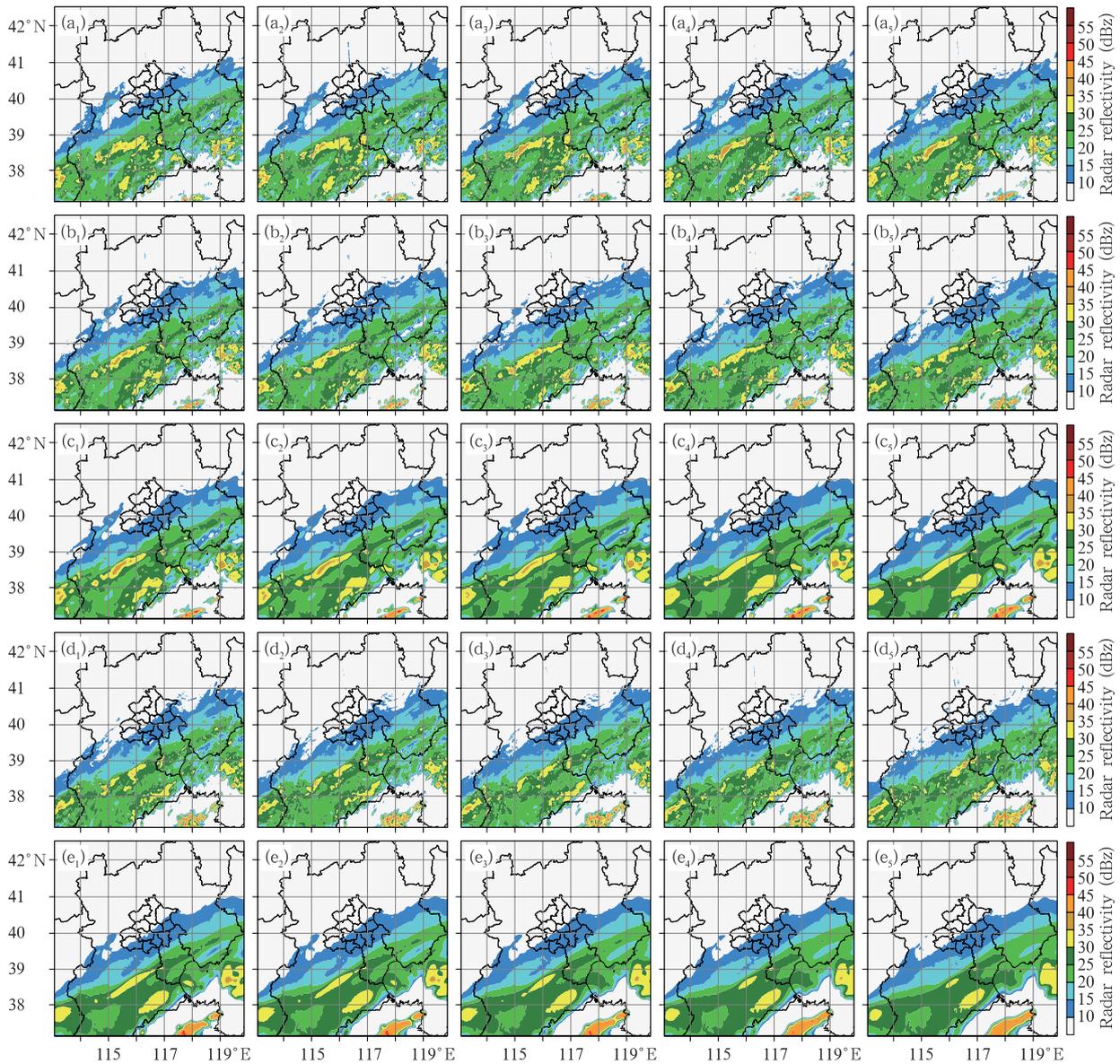


图4 逐6 min的雷达回波(5帧)

(a. 模型输入, b. 6—30 min 实况, c. 6—30 min 预测, d. 36—60 min 实况, e. 36—60 min 预测;
第一帧输入为2017年5月22日10时41分(北京时间,下同))

Fig. 4 Radar echo maps at 6 min intervals (5 frames)

(a. model input, b. 6—30 min ground truth, c. 6—30 min prediction, d. 36—60 min ground truth,
e. 36—60 min prediction respectively. The first frame input is at 10:41 BT 22 May 2017)

习到这种在边界上出现的雷达回波的移动趋势并给出合理的预测结果。但存在的问题是,随时间的推移,不同强度回波的边缘细节逐渐丢失,一个可能的原因就是解码器网络不断迭代产生预测输出,预测的难度逐渐加大,这导致预测值的置信度逐渐降低并反馈到下一时刻的输入,导致预测的不确定性不断累加,在图中的表现就是边缘细节慢慢丢失,变得模糊。

从图 5 中可以看出,在强对流天气中,卷积 GRU 模型能准确追踪到对流风暴的移动趋势,强回波位置和发展形势与对应的雷达观测实况较为接近。但同样存在一些不足,和实况图相比,预测的雷达回波图中,不同强度回波的边缘轮廓较为模糊,丢失一些细节。例如实况图上是几个相近位置的不同雷暴回波单体,而卷积 GRU 很难将其区分开,预测图中它们融合成了一个更大的回波,产生虚报的问题。出现这种现象,一个可能的原因是这种天气预测问题本身固有的不确定性(强对流的非线性快速演变),随着时间的推移,不确定性快速传播,几乎不可能准确预测出未来回波的准确位置。所以模型为了减少预测误差,就会倾向于产生模糊的预测以提高预测的“命中率”,CTREC 算法预报的结果虽然包含更多的细节,看起来更加锐利,但这些细节更多的是从原始雷达回波“拷贝”过来的,而回波一直处在动态的变化之中,这种保留细节的方式会带来更多的虚警(“误报率”增大),并且其追踪的回波位置并不准确(预报位置有些超前)。总的来说,卷积 GRU 模型对不同强对流回波区域的 0.5 和 1 h 临近预报与实况更吻合。

再来比较卷积 GRU 模型与 CTREC 算法在一次飊线过程中的预测效果(图 6),京津冀地区出现一次明显的飊线过程,卷积 GRU 模型在 0.5 和 1 h 预测中均可准确地追踪飊线的移动位置和演变特征,其形状和位置与雷达观测实况基本吻合。而 CTREC 算法的追踪位置与实况存在偏差。且卷积 GRU 模型对飊线北部的较强回波单体的追踪也较为准确,CTREC 算法的预测与实况相比明显偏南、偏强。但在 1 h 预测中卷积 GRU 模型存在细节丢失的现象,预测的回波强度也存在偏弱的问题,需要在之后的试验中尝试解决。从试验结果看出,卷积

GRU 模型对类似于飊线这样组织性较强的对流系统的 1 h 预报有更明确的参考意义。

从试验结果来看,卷积 GRU 模型对强对流临近预报具有较好效果,可以有效给出未来 0.5 和 1 h 的雷达回波位置及形状预报。但卷积 GRU 的预测结果更加平滑,主要是两个原因造成的:一个是预测任务本身的特点,本身存在较大的不确定性,且随时间递增不确定性会不断增大;另一个是损失函数引起的,文中使用的平均绝对误差和平均平方误差的损失函数会倾向于产生模糊结果,例如当平均平方误差作为损失函数来驱动模型训练时,随着预测时间的推移,预测难度增大,模型无法学习到有效的特征进行合理预测时,模型就会倾向于用样本概率分布的期望值作为预测结果,因为这种预测策略会使得平均平方误差的损失函数值最小,同时也会产生平滑的预测结果。总的来说,卷积 GRU 模型的预报结果与 CTREC 算法的预报结果相比,要更接近于实况观测,预报结果具有较好指示意义,对临近预报有一定帮助。

5 结 论

选取 2010—2017 年每年 6—9 月京津冀地区 6 部新一代天气雷达逐 6 min 回波拼图数据,使用深度学习方法构建了一个高时效、易拓展的临近预报模型——卷积 GRU。试验及检验结果表明,卷积 GRU 的几项检验评分结果都要优于传统的雷达回波区域跟踪外推临近预报方法,能有效预测未来 1 h 的雷达回波演变特征。研究结果说明深度学习模型在特征选择和特征学习方面的优势,在处理雷达回波等高时、空分辨率气象观测海量“大数据”时,可以在其丰富的数据分析中,更加有效地提取出有用的时、空气象信息以及所蕴含的中小尺度天气演变规律和物理模型,通过多层神经网络将输入的关联物理量叠加组合成更高级更抽象的特征物理量,从而提高了模型预测的能力和准确度。

卷积 GRU 模型作为一种人工智能技术,在大数据时代背景下,能够满足气象工作人员对强对流回波的 1 h 临近预报需求。气象学家长期以来一直在追踪天气数据,并利用这些数据进行预测,在这方面,深度学习方法可以提供新的思路,用来帮助解决

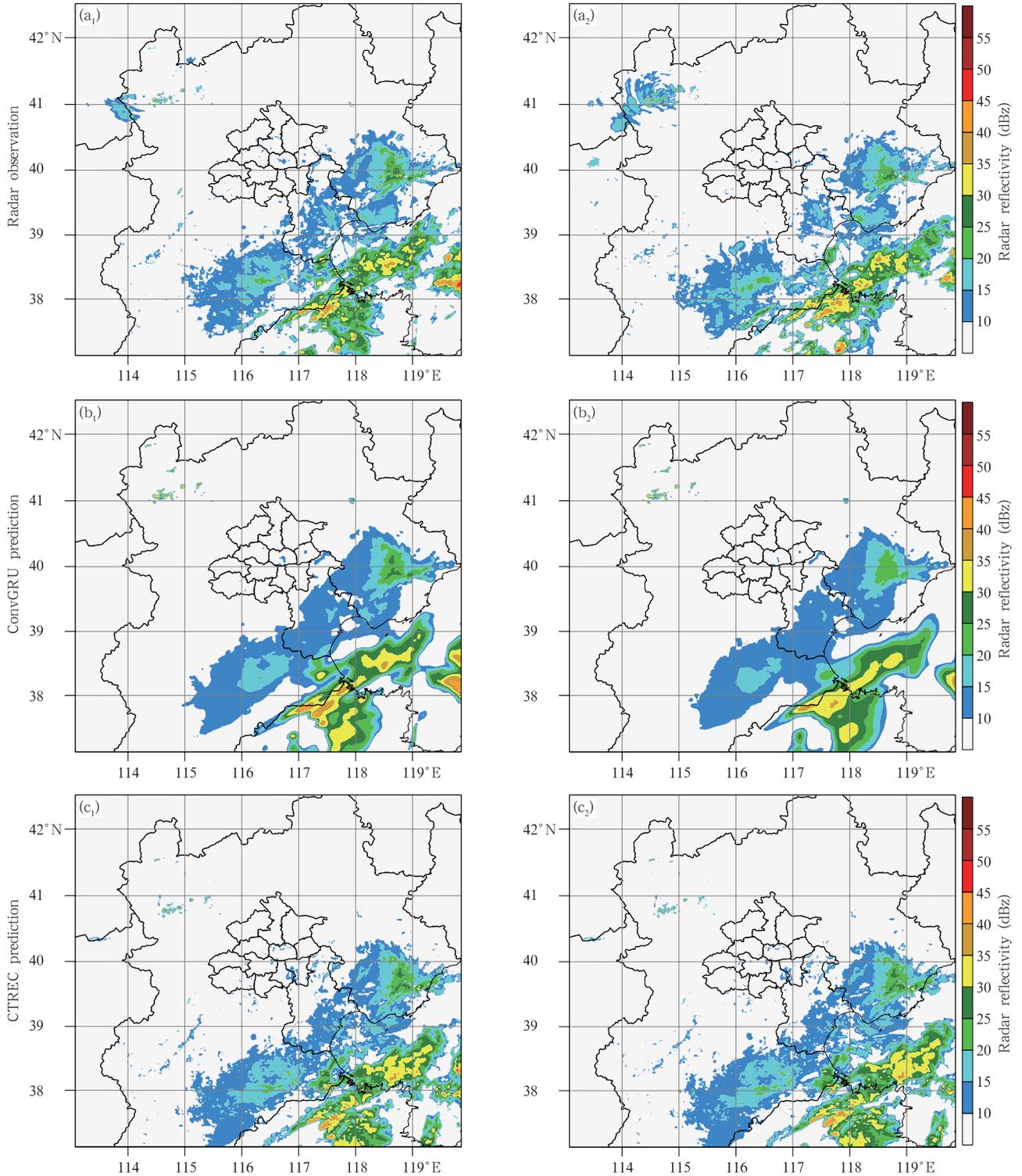


图5 卷积GRU模型(b)和CTREC算法(c)在2017年8月23日09时41分的
临近预报与雷达观测实况(a)的对比

(a_1-c_1 . 0.5 h, a_2-c_2 . 1 h)

Fig. 5 Comparison between radar observations (a) and nowcasts by ConvGRU model (b)
and CTREC algorithm (c) at 09:41 BT 23 August 2017

(a_1-c_1 . 0.5 h, a_2-c_2 . 1 h)

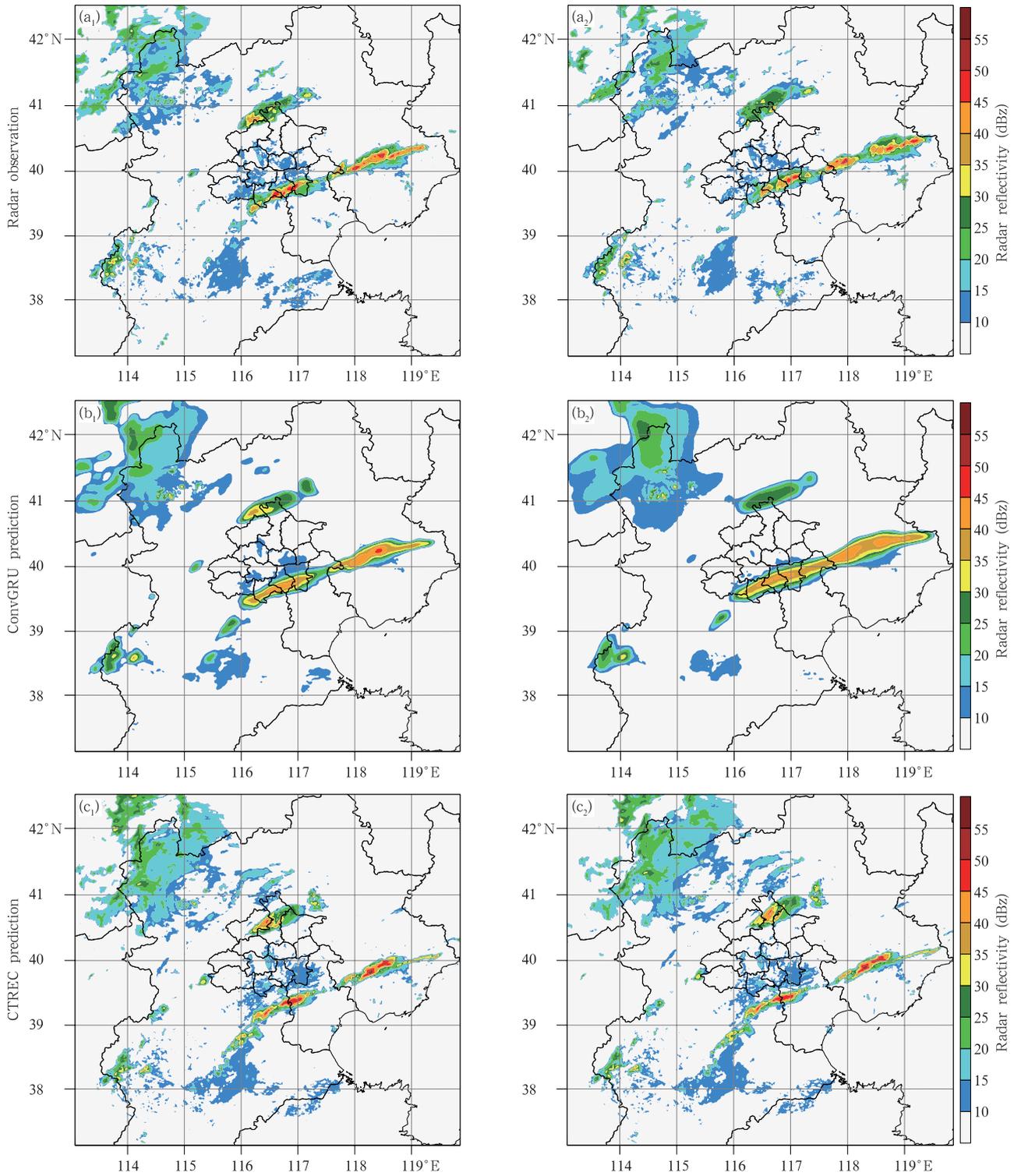


图6 卷积GRU模型(b)和CTREC算法(c)在2017年8月1日22时47分的
临近预报与雷达观测实况(a)的对比
(a_1-c_1 . 0.5 h, a_2-c_2 . 1 h)

Fig. 6 Comparison between radar observations and nowcasts by ConvGRU model
and CTREC algorithm at 22:47 BT 1 August 2017
(a_1-c_1 . 0.5 h, a_2-c_2 . 1 h)

一些更困难的天气分析和预报难题,例如在强对流天气短时临近预报方面,深度学习具有很大潜力。

需要指出的是,目前深度学习在气象领域的应用仍旧处于探索阶段,尚有较大的改进空间。例如,从个例分析对比可以看出,卷积 GRU 相较于 CTREC 的预报结果较为模糊,这必然会对卷积 GRU 最后的检验评分有一定的正面影响,如何在不损失模型检验评分的情况下,尽可能的让卷积 GRU 也产出“锐利”的预测图,是接下来研究的重点,会使用另一种深度学习网络模型——“生成对抗网络模型”(Goodfellow, et al, 2014)集成到当前的模型中,来尝试使模型输出精细的预测结果。在临近预报方面,如何有效使用雷达、地面、卫星等多源数据进行融合预测,如何使用三维数据进行预测,将预测时间延长至 2 h 以上,或是利用在线学习构建实时在线训练和推理系统,不断将新生成的观测数据纳入神经网络模型训练中,使得预测效果随着时间的推移不断改善。另外,如何将气象上的物理概念模型与深度学习算法有机结合,提升预报的准确度并给出预报结果的合理物理解释,也值得深入研究。

毋庸置疑,深度学习技术在气象领域有着广阔的应用前景,可以为气象领域增添新的活力。随着深度学习技术的不断发展,将来会有更多的成果加快推进深度学习技术在气象领域的研究应用。

参考文献

- 陈明轩, 俞小鼎, 谭晓光等. 2004. 对流天气临近预报技术的发展与研究进展. 应用气象学报, 15(6): 754-766. Chen M X, Yu X D, Tan X G, et al. 2004. A brief review on the development of nowcasting for convective storms. J Appl Meteor Sci, 15(6): 754-766 (in Chinese)
- 陈明轩, 王迎春, 俞小鼎. 2007. 交叉相关外推算法的改进及其在对流临近预报中的应用. 应用气象学报, 18(5): 690-700. Chen M X, Wang Y C, Yu X D. 2007. Improvement and application test of TREC algorithm for convective storm nowcast. J Appl Meteor Sci, 18(5): 690-700 (in Chinese)
- 程丛兰, 陈明轩, 王建捷等. 2013. 基于雷达外推临近预报和中尺度数值预报融合技术的短时定量降水预报试验. 气象学报, 71(3): 397-415. Cheng C L, Chen M X, Wang J J, et al. 2013. Short-term quantitative precipitation forecast experiments based on blending of nowcasting with numerical weather prediction. Acta Meteor Sinica, 71(3): 397-415 (in Chinese)
- 郭尚瓚, 肖达, 袁行远. 2017. 基于神经网络和模型集成的短时降雨预测方法. 气象科技进展, 7(1): 107-113. Guo S Z, Xiao D, Yuan H Y. 2017. A Short-term rainfall prediction method based on neural networks and model ensemble. Adv Meteor Sci Technol, 7(1): 107-113 (in Chinese)
- 黄伟, 余晖, 梁旭东. 2009. GRAPES-TCM 对登陆热带气旋降水的预报及其性能评估. 气象学报, 67(5): 892-901. Huang W, Yu H, Liang X D. 2009. Evaluation of GRAPES-TCM rainfall forecast for China landfalling tropical cyclone in 2006. Acta Meteor Sinica, 67(5): 892-901 (in Chinese)
- 俞小鼎, 周小刚, 王秀明. 2012. 雷暴与强对流临近天气预报技术进展. 气象学报, 70(3): 311-337. Yu X D, Zhou X G, Wang X M. 2012. The advances in the nowcasting techniques on thunderstorms and severe convection. Acta Meteor Sinica, 70(3): 311-337 (in Chinese)
- 郑永光, 林隐静, 朱文剑等. 2013. 强对流天气综合监测业务系统建设. 气象, 39(2): 234-240. Zheng Y G, Lin Y J, Zhu W J, et al. 2013. Operational system of severe convective weather comprehensive monitoring. Meteor Mon, 39(2): 234-240 (in Chinese)
- Bengio Y, Simard P, Frasconi P. 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Networks, 5(2): 157-166
- Chen M, Bica B, Tüchler L, et al. 2017. Statistically extrapolated nowcasting of summertime precipitation over the Eastern Alps. Adv Atmos Sci, 34(7): 925-938
- Cho K, Van Merriënboer B, Gulcehre C, et al. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv:1406.1078
- Chung J, Gulcehre C, Cho K H, et al. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv:1412.3555
- Goodfellow I J, Pouget-Abadie J, Mirza M, et al. 2014. Generative adversarial nets//27th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2672-2680
- Han L, Fu S X, Zhao L F, et al. 2009. 3D convective storm identification, tracking, and forecasting: An enhanced TITAN algorithm. J Atmos Oceanic Technol, 26(4): 719-732
- Hochreiter S, Schmidhuber J. 1997. Long short-term memory. Neural Comput, 9(8): 1735-1780
- Kingma D P, Ba J. 2014. Adam: A method for stochastic optimization. arXiv:1412.6980
- Lai E S T. 1998. TREC application in tropical cyclone observation. ESCAP/WMO Typhoon Committee Annual Review, 867-875
- LeCun Y, Bengio Y, Hinton G. 2015. Deep learning. Nature, 521(7553): 436-444
- Rumelhart D E, Hinton G E, Williams R J. 1986. Learning representations by back-propagating errors. Nature, 323(6088):

533-536

- Shi X J, Chen Z R, Wang H, et al. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting//28th International Conference on Neural Information Processing Systems. Montreal, Canada; MIT Press
- Shi X J, Gao Z H, Lausen L, et al. 2017. Deep learning for precipitation nowcasting: A benchmark and a new model//31st Conference on Neural Information Processing Systems. Long Beach, CA, USA
- Sun J Z, Xue M, Wilson J W, et al. 2014. Use of NWP for nowcasting convective precipitation: Recent progress and challenges. *Bull Amer Meteor Soc*, 95(3): 409-426
- Sutskever I, Vinyals O, Le Q V. 2014. Sequence to sequence learning with neural networks. *arXiv:1409.3215*
- Weisman M L, Davis C, Wang W, et al. 2010. Experiences with 0 - 36 h explicit convective forecasts with the WRF-ARW model. *Wea Forecast*, 23(3): 407
- Zhang K, Zuo W M, Zhang L. 2018. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans Image Process*, 27(9): 4608-4622