

DOI:10.16515/j.cnki.32-1722/n.2019.02.008

# 基于神经网络的云顶高反演技术研究分析

孟 恒<sup>1,2</sup>, 杨 忠<sup>2\*</sup>, 郭洪涛<sup>2</sup>

(1. 河海大学(常州校区)机电工程学院, 江苏 常州 213022;  
2. 金陵科技学院智能科学与控制工程学院, 江苏 南京 211169)

**摘 要:**收集整理 2017 年 5 月到 2018 年 3 月的美国 CALIPSO 卫星云参数反演产品和日本“葵花 8”静止卫星云图。研究卫星遥感数据提取与处理技术,提取葵花卫星与美国 CALIPSO 卫星云参数反演产品对应位置、时次数据,制成样本集。运用传统的亮温廓线法和神经网络的 BP、CNN、RNN 三种方法,构建模型对云顶高反演的精度进行研究和比较。研究表明:BP 神经网络模型反演效果较好,RNN 模型次之,CNN 和传统的亮温廓线法反演结果相对较差。

**关键词:**云顶高;神经网络;葵花 8 号卫星;CALIPSO 卫星

**中图分类号:** TP79      **文献标识码:** A      **文章编号:** 1672-755X(2019)02-0034-06

## Research and Analysis on Cloud Top Height Inversion Technology Based on Neural Network

MENG Heng<sup>1,2</sup>, YANG Zhong<sup>2\*</sup>, GUO Hong-tao<sup>2</sup>

(1. Changzhou Campus of Hohai University, Changzhou 213022, China;  
2. Jinling Institute of Technology, Nanjing 211169, China)

**Abstract:** A number of days of US CALIPSO satellite cloud parameter inversion products and Japanese “Himawari-8” geostationary meteorological satellite cloud data are collected and sorted out from May 2017 to March 2018. Research are carried out on the technology of extracting and processing satellite remote sensing data, the corresponding position and time of Himawari satellite and the US CALIPSO satellite data are extracted, and then the sample set is made. The research group use the traditional bright temperature profile method and BP, CNN and RNN methods to construct the model to study and compare the accuracy of cloud top height inversion. The conclusion is that the BP neural network model has better inversion results, the RNN model performs ordinarily, and the CNN and traditional brightness temperature profile inversion results are relatively poor.

**Key words:** cloud top height; neural network; Himawari-8; CALIPSO

云对气候系统的反馈依赖于云顶高、云量和辐射特性,而云顶高研究一直备受关注。当今国内外对云顶高的研究大多以传统的研究方法为主,其包含三大类:成像几何方法<sup>[1]</sup>、单通道反演方法<sup>[2]</sup>、多通道反演

**收稿日期:** 2019-04-18

**基金项目:** 国家自然科学基金项目(51505204);教育部产学合作协同育人项目(201602009006);江苏省产学研联合前瞻研究项目(BY2016012-02)

**作者简介:** 孟恒(1993—),男,江苏连云港人,硕士研究生,主要从事利用深度学习方法构建模型反演云参数的研究。

**通信作者:** 杨忠(1968—),男,江苏南京人,教授,博士,主要从事智能控制理论及应用、光伏发电智能控制等研究。

方法<sup>[3]</sup>。但是传统反演云顶高的方法有两大不足:云顶高反演误差较大;反演的方法所需的条件苛刻或较为繁琐。

因此,本文用传统方法中最经典的亮温廓线法以及神经网络中研究较多的 CNN、RNN、BP 三大神经网络方法,分别对相同的卫星数据进行反演。比较四种方法的精确度,从而建立较优的云顶高反演模型。

## 1 卫星数据结构

本文主要利用葵花 8 号(Himawari-8)气象静止卫星与美国的 CALIPSO 卫星的数据。

### 1.1 葵花 8 号数据

葵花 8 号气象静止卫星数据一共有 16 个通道,主要由可见光、红外以及近红外三种通道构成。本文为了使捕捉的点更为密集,选用 R21 的 06001\_06001 文件,其中 06001\_06001 表示空间的网格数分辨率为 2 km。

### 1.2 CALIPSO 数据

CALIPSO 卫星数据一般由两级数据组成。本文主要利用的是 CALIOP 卫星的二级云产品里的云顶高参数<sup>[4]</sup>,该产品的分辨率为 1 km。

### 1.3 数据提取与匹配

将两种卫星数据分别用 python 编程读取。python 有 netCDF4 与 pyhdf4 两个工具包用于读取葵花卫星的 NC 文件与 CALIPSO 卫星的 HDF 文件,均读取为 TXT 格式。每一行代表一组数据,葵花卫星读取的内容依次是:经度、纬度、1~16 通道的数值,以及其他参数,最后是时间。CALIPSO 卫星读取的内容依次是:时间、经度、纬度、云顶高等一系列参数。

HDF 文件读取出的第一栏里的时间是儒略日数,通过 python 编码将其转化成年月日时分秒。

数据读取完之后就开始进行数据匹配。最主要的条件为:必须是两种卫星数据同经度、同纬度、同时间点,该点才为有效数据。但是对于信息匹配存在两个问题:两种卫星数据的分辨率不同,为 1 km 或 2 km,所以很难有精确重合的点;很难找到完全重合的时间。因此,为了充分利用 CALIPSO 数据,匹配出合适数量的数据。需要以 CALIPSO 的数据为基准,用葵花卫星的数据进行匹配。对于经纬度,各相差在  $0.05^\circ$  即 5 km 的范围之内,可以看作是相同点;对于时间,相差一分钟以内的即可看作是同一时间点。当一条 CALIPSO 数据可以对应于几条葵花卫星的数据时,解决的办法是每匹配一条 CALIPSO 数据,多出的葵花数据分别和该条 CALIPSO 数据的经纬度算直线距离,选取距离最近的点,就可以得到一条 CALIPSO 数据对应一条葵花数据,这样匹配完的数据各近 142 万条。

## 2 反演方法介绍

### 2.1 亮温廓线法

2.1.1 方法介绍 传统的亮温廓线法是先根据亮温算出对应时次和地点的温度,然后依据所得的温度,对大气温度的垂直廓线找出对应温度的高度,从而得到云顶高<sup>[5]</sup>。需要利用 NCEP 数据中的 air. nc 与 hgt. nc 两类文件,并将该类文件读取为 TXT 格式文件。

2.1.2 实现过程 首先根据葵花卫星的数据经度、纬度、时间和第 14 通道的数据(这里的 14 通道为红外通道,其通道的数值就可以近似为云顶温度)。其次在 air. txt 的文件里寻找对应位置和时间点的数据,然后根据葵花卫星第 14 通道的温度值来找 air. txt 里面对应温度数值的层数。再将已有的位置、时间和所对应层的位置,对应 hgt. txt 的位置、时间和层数,该层的值就是要算的云顶高。最后再将已查到的云顶高的值对应于 CALIPSO 卫星的相同位置、时间的云顶高,比较其误差。

其中关于时间、位置、高度、温度的取值问题,直接近似会产生更大的误差,所以需要对其运用插值法。本文主要是运用线性插值法。

2.1.3 误差分析方式 对于误差的表示,本文使用平均绝对误差(MAE)、均方根误差(RMSE)、平均相对误差(MRE)来表示。

$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (2)$$

$$\text{MRE}/\% = \frac{1}{m} \sum_{i=1}^m (|y_i - \hat{y}_i| / y_i) \times 100 \quad (3)$$

## 2.2 BP 神经网络模型

2.2.1 模型构造 本文的数据匹配最后留下的数据为 1 428 128 组,利用 Keras 框架构建模型。

1)构建网络训练样本。对最终筛选出的 1 428 128 组数据先预留 18 128 组数据作为测试集,剩下的 1 410 000 组数据使用 K 折交叉验证法。模型中 K 取 4,所以样本被随机分为 1 057 500 个训练样本和 352 500 个验证样本。2)输入层设计。将红外通道 8~16 通道的数据以及经度、纬度和时间作为输入,神经元的数量为 12。因此输入层的矩阵为(12,1)。输入层设置的节点数量为 21 个,效果最优。另外,对于输入层的数据要实现归一化。3)输出层设计。根据样本数据,输出数据的神经元就一个,就是云顶高的数值,并将输出值进行反归一化。4)隐藏层设计。此模型的隐藏层的层数设置为一层,神经元个数设置为 21 个,激活函数输入、输出以及激活层均设为“ReLU”。

本文运用 Adam 优化算法,并且在每一层的后面还设置了 dropout 层。由公式(4)确定最优的神经元个数<sup>[6]</sup>:

$$\sum_{i=0}^n C_M^i > k \quad (4)$$

其中, $k$  是训练样本的数量, $n$  是输入层神经元个数, $M$  是隐藏层的节点数量,令  $i > M$  时, $C_M^i = 0$ 。

依据本文样本数量可以算出隐藏层的节点数为 16,因此以  $M$  为 16 作为参考值,对于低于或者高于 16 的值均进行尝试,可以得到当  $M$  为 21 时,误差最小,模型最稳定。

2.2.2 实现效果 在运行代码时,还运用了 Tensorboard 来记录训练过程的 loss 值,从而观察精度的变化。

如图 1、图 2 所示,对于训练集 loss 的数值呈下降趋势,刚开始下降很快,在 100 步到 500 步的时候下降缓慢,集中在 0.022 5~0.024 5;而对于验证集的数据,loss 值虽然在图中的波动很大,但是一直集中在 0.036~0.044,可见验证集的 loss 也很小,模型训练的效果不错。

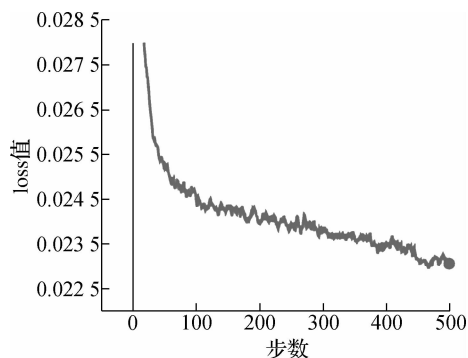


图 1 BP 神经网络训练集的 loss 值变化过程

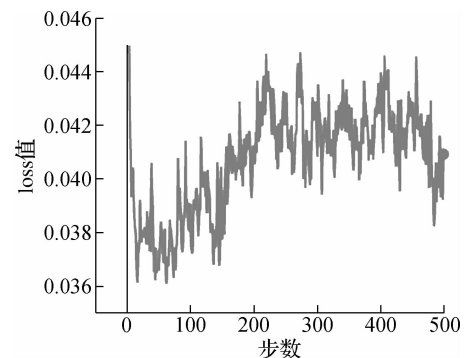


图 2 BP 神经网络验证集的 loss 值变化过程

## 2.3 RNN 神经网络模型

2.3.1 模型构造 首先需要构造的输入数据和 BP 神经网络的输入数据一样。模型主要是 LSTM 模型,优化器是 Adam,激活函数是 ReLU,损失函数是均方误差(Mean Squared Error),神经节点设置为 21。

2.3.2 实现效果 训练集和验证集训练的效果如图 3、图 4 所示。在训练集中,loss 值呈下降趋势,初始下降很快,100 步到 500 步时下降缓慢,集中在 0.015~0.018;验证集的 loss 虽然在图中的波动很大,但是 loss 值一直集中在 0.032~0.047。可见验证集的 loss 也很小,模型训练的效果不错。

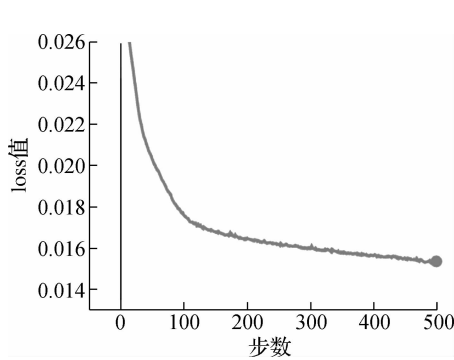


图 3 RNN 神经网络训练集的 loss 值变化过程

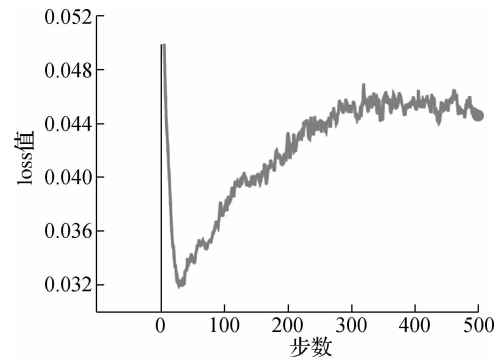


图 4 RNN 神经网络验证集的 loss 值变化过程

## 2.4 CNN 神经网络模型

2.4.1 模型构造 本模型依旧采用 Keras 框架。1)构建网络训练样本。将对最终筛选出的 1 428 128 组数据分成三类样本集,分别是训练集、验证集、测试集。2)输入层设计。经过对数据的观察,红外通道 4~16 通道的数据作为输入,即输入的神经元的个数为 13。另外加上时间、经度、纬度三个变量,共 16 个。设置成  $4 \times 4$  的输入,且对输入层的数据实现归一化。3)输出层设计。根据样本数据,输出数据的神经元就一个,即云顶高的数值,并需对输出数据反归一化。4)隐藏层设计。此模型的隐藏层设置为五层,三层卷积层,两层全连接层。该模型没有用到池化层,因为每个数据输入规格不大。而且对  $4 \times 4$  的周围补 0,形成  $5 \times 5$  的数据。每一层的激活函数都是 ReLU,并且在每一层的后面还设置了 dropout 层,优化器选用了 Adam。

2.4.2 实现效果 训练集和验证集训练的效果如图 5、图 6 所示。在训练集中,loss 的数值呈下降趋势,刚开始下降很快,在 60 步到 200 步的时候下降缓慢,集中在  $0.040 \sim 0.042$ ;但对于验证集的数据,loss 值在图中的波动很大,loss 值一直集中在  $0.100 \sim 0.120$ 。相较于上面两个神经网络模型,CNN 神经网络模型效果较差,即使当训练集的损失较小,在验证集中的效果依旧也不太理想。

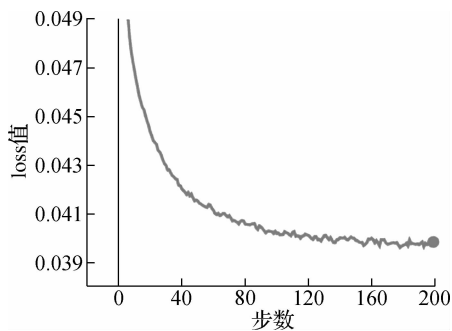


图 5 CNN 神经网络训练集的 loss 值变化过程

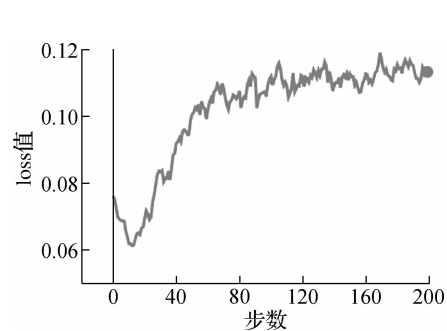


图 6 CNN 神经网络验证集的 loss 值变化过程

## 3 对几种反演云顶高的误差分析

首先,本文的数据对应着 2017 年 5 月—2018 年 3 月的某些天的数据,其反演的平均绝对误差和平均相对误差如图 7、图 8 所示,平均绝对误差和平均相对误差均为 BP 神经网络表现最好,其中平均绝对误差集中在 800 m 左右,平均相对误差集中在 9% 左右。RNN 表现稍差一些,平均绝对误差集中在 1 000 m,平均相对误差集中在 10% 左右。CNN 网络模型和传统的亮温廓线法表现相当平均,且误差较大。

四种方法的平均绝对误差(MAE)、均方根(RMSE)、平均相对误差(MRE)的平均值如表 1 所示。可见,BP 神经网络模型的表现是较优的,RNN 神经网络模型次之,CNN 神经网络模型是效果较差的。其中,CNN 神经网络模型误差和传统的亮温廓线法相当,不具备应用性。比较 BP 神经网络模型和 RNN 神经网络模型可以看出,BP 神经网络模型的误差和均方根误差都较小,说明 BP 神经网络模型更加稳定。

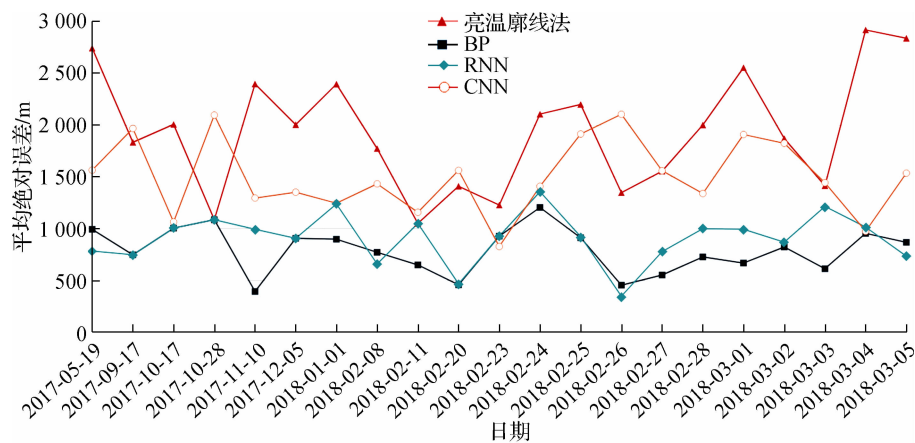


图 7 四种反演模型的平均绝对误差

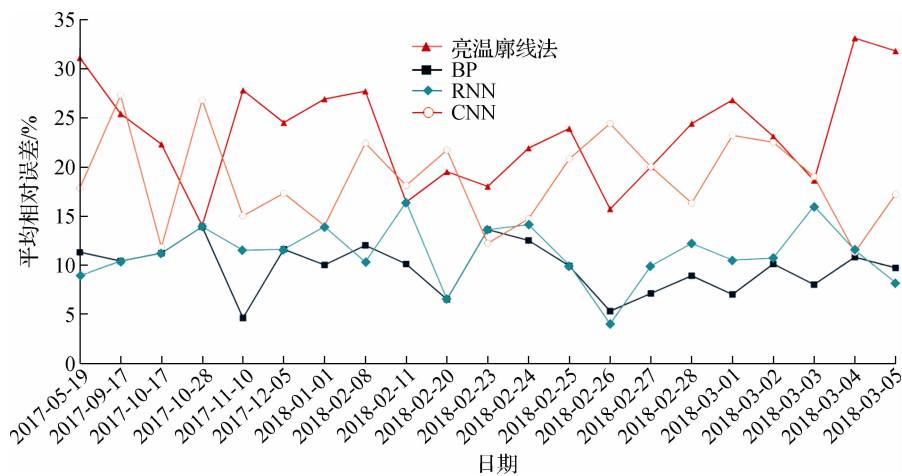


图 8 四种反演模型的平均相对误差

表 1 四种方法的误差对比

方法	平均绝对误差(MAE)/m	均方根误差(RMSE)/m	平均相对误差(MRE)/%
亮温廓线法	1 932.68	3 128.10	23.45
BP 神经网络	791.22	1 635.30	9.71
RNN 神经网络	910.33	1 792.17	11.23
CNN 神经网络	1 501.63	2 815.90	18.74

## 4 结 语

1)云反演的误差降低。从表 1 可以看出,深度学习中的 BP 神经网络与 RNN 神经网络对于传统云顶高反演误差较大的缺点有着明显的改善。测试集对反演云顶高的误差显示,误差最小的神经网络只相当于传统方法误差的三分之一。

BP 神经网络效果较好的原因是本文的卫星输入数据的矩阵为(12,1),输出矩阵为(1,1),其类似于一个十二元多次方程。BP 神经网络利用训练数据不断地调整模型的权值与阈值来使得误差不断减小,使其更趋近于真实值,其实就类似于一个拟合的过程。只是方程的元数与次数不好确定,但因为数据充足以及反复训练,使得训练出的输出结果不断向真实值靠近。而且在构建 BP 模型的同时,也采取了多种方法防止模型过拟合,这样能够训练出一个较好的模型。

2)云反演的条件简单化。就本文所实现的亮温廓线法反演云顶高而言,此方法已经是对反演的条件

要求比较简单的,但还是需要葵花数据、NCEP 数据,并且 NCEP 数据并不是实时数据而是平均数据,所以误差偏大。之前国内外云顶高反演方法的研究,有些是实现方法简单,误差较大,比如亮温廓线法;有些是反演的误差较小,但是反演的条件苛刻并且多是对某一类云有效,比如成像几何法,其必须要在双星同步观测的情况下,而且只是对强风暴云团的反演误差较小;单通道法虽然条件较简单,但是只适用于低云或厚度很厚的云,对于多通道的则是先将云分成透明云、半透明云与不透明云三类,然后才可以进行反演,反演的方法均比较繁琐。

深度学习的方法,规避了需要对云进行分类的环节,因为庞大的数据量足以让模型自己学习不同类的云所应该具备的特征。因此,深度学习中的神经网络方法只需输入葵花数据,依据训练好的模型,对应输出的就是云顶高。

#### 参考文献:

- [1] WANG H, SHENG H, ZHANG Y, et al. Determination of cloud-top height from stereoscopic observation[J]. Progress in Nature Science, 2002, 12(9): 689 - 694
- [2] 朱亚平,程周杰,刘健文,等. FY-2C 静止气象卫星资料对不透明云云顶气压的反演研究[J]. 热带气象学报, 2014, 30(3): 463 - 472
- [3] 林琳,黄思训,杜华栋. 使用 MODIS 数据对云顶高的反演研究[J]. 气象科学, 2006, 26(6): 6655 - 6661
- [4] WINKER D M, PELON J, COAKLEY J A, et al. The CALIPSO mission: A global 3D view of aerosols and clouds[J]. Bulletin of the American Meteorological Society, 2010, 91(9): 1211 - 1229
- [5] 马烁,黄云仙,严卫,等. 三线阵云立体探测技术[J]. 红外与毫米波学报, 2014(2): 164 - 171
- [6] 史峰,郁磊,李洋. MATLAB 神经网络 42 个案例分析[M]. 北京:北京航空航天大学出版社, 2013

(责任编辑:湛 江)