

Doi: 10.11840/j.issn.1001-6392.2021.03.005

# 基于深度学习的海温观测数据质量控制应用研究

刘玉龙<sup>1</sup>, 王国松<sup>1</sup>, 侯敏<sup>2</sup>, 徐珊珊<sup>1</sup>, 苗庆生<sup>1</sup>

(1. 国家海洋信息中心, 天津 300171; 2. 天津市滨海新区气象局, 天津 300457)

**摘 要:** 基于国家海洋信息中心质量控制后的西太平洋 10 度方区约 100 万站次温盐实测历史调查资料, 对经过 26 种严格质量控制方法的综合海温质量符进行分类分析, 首次将深度学习技术应用于海洋数据质量控制多分类 (multiclass classification) 算法与应用研究。通过人工合成少数类样本和加权损失函数方法减少多数类的频率来降低数据的不平衡, 并构建了多层感知器 (Multi-Layer Perceptron, MLP) 和深度神经网络 (Deep Neural Network, DNN) 两个海温资料质量符分类深度学习模型。分类结果表明本文构建的两个深度分类模型能够较准确快速地识别该海域海温数据质量, 在 20 975 条温盐剖面资料测试集中分类准确率分别达到 99.63% 和 99.69%。海温资料的分类精度评分有着较好的表现, 其中正确数据 (QC1) 和数据缺失 (QC9) 的正确识别率均达 100%。MLP 和 DNN 多分类质量控制模型可大幅降低传统质量控制方法的工作量, 提升海量数据处理速度和分析能力, 为海温观测资料在海洋研究与工程中应用提供参考。

**关键词:** 海温; 质量控制; 深度神经网络; 多层感知器

中图分类号: P731.1; TP181

文献标识码: A

文章编号: 1001-6932(2021)03-0283-09

## Quality control of sea temperature observation data using deep learning neural networks

LIU Yulong<sup>1</sup>, WANG Guosong<sup>1</sup>, HOU Min<sup>2</sup>, XU Shanshan<sup>1</sup>, MIAO Qingsheng<sup>1</sup>

(1. National Marine Data and Information Service, Tianjin 300171, China;

2. Tianjin Binhai NEW AREA Meteorology Administration, Tianjin 300457, China)

**Abstract:** Based on one million records of thermohaline measured data in the western Pacific, the deep learning neural networks is applied to the multiclass classification algorithm of sea temperature observation data. The quality flags of sea temperature, which is from the National Marine Data and Information Service (NMDIS), was classification and analysis after 26 kinds of strict sea temperature quality control methods. In order to reduce the data imbalance, the frequency of most classes is reduced by artificial synthesis of minority samples, down-sampling and cluster sampling, and the deep learning neural networks of quality flags classification of sea temperature observation data are constructed respectively by Multi-Layer Perceptron (MLP) and deep neural network (DNN). The results show that, the two deep classification models constructed in this paper can accurately and quickly identify the quality of sea temperature data. The classification accuracy of the 20 975 temperature and salt profile data on test sets is 99.63% and 99.69%, respectively. The classification accuracy rating has a better performance, and the classification accuracy rating of correct data (QC1) and data missing (QC9) is 100%. The MLP and DNN multiclass classification quality control model can greatly reduce the workload of traditional quality control methods, improve the

收稿日期: 2021-01-08; 修订日期: 2021-03-09

基金项目: 国家重点研发计划 (2016YFC1401905; 2016YFC1402605; 2017YFC1405302); 全球变化与海气相互作用 (GASI-IPOVAI-04); 国家自然科学基金 (41606039; 41776004)

作者简介: 刘玉龙 (1982—), 副研究员, 主要从事海洋资料处理研究。电子邮箱: yulong0631@163.com

通讯作者: 王国松, 博士研究生, 助理研究员, 主要从事数值预报与资料处理研究。电子邮箱: xifengbishu2110@163.com

<http://hytb.nmdis.org.cn>

processing speed and analysis capability of massive data, and provide availability reference for ocean temperature data in marine data research and engineering applications.

**Key words:** sea temperature; quality control; deep neural network; multi-layer perceptron

我国海洋事业的发展涉及政治、经济、民生等各领域,国家防灾减灾救灾体系建设、生态文明建设、“一带一路”倡议等,均对海洋观测数据及服务产品提出了明确需求。高质量、一致、完整、无偏差的海洋历史记录对现代海洋研究至关重要。近年来,在智慧海洋建设的大背景下(姜晓轶等, 2018),各类海洋环境资料受到了空前的关注,其中海水的温度和盐度是描述海水性质的重要物理量(王东晓等, 2004),其时空分布和变化几乎与海洋中所有现象都有密切的联系(李伟等, 2012),为此众多涉及温盐数据的全球或区域海洋观测计划应运而生,如全球Argo浮标观测计划(Oka et al, 2004; Wong et al, 2018)、全球温盐剖面计划(GTSP) (Imis, 2010; Hamilton, 1994; Wilson, 1998; Deng et al, 2016; Levitus et al, 2000, 2013; Locarnini et al, 2013; Reagan et al, 2014)、世界海洋数据集(WOD) (Levitus et al, 2000, 2013; Locarnini et al, 2013; Reagan et al, 2014)、资料浮标合作组(DBCP) (Reul et al, 2014)、全球海洋观测系统(GOOS) (Summerhayes, 2002)、全球海洋表层走航数据项目(GOSUD) (Centurioni et al, 2017)等。海洋剖面观测被广泛用于评估和初始化模型模拟。良好的观测数据集对未来预测地球气候变化情况至关重要。

随着海洋自动化观测仪器的增多,海洋数据量呈几何级上涨,资料质量问题开始突显,因资料质量引发的资料可用性问题严重制约了海洋数据在相关研究中的应用(王辉赞等, 2012)。当今最大的挑战之一是需要构建一套完整的海温数据集,并将其广泛应用于气候研究和服务中。尽管过去几十年来许多研究组织都试图对海洋剖面数据进行维护和质量控制(QC),但全球历史剖面数据库仍然包含相对较大的偏差。针对由于外界环境或者仪器本身的不稳定导致的资料质量问题,各国研究机构均开展了不同程度的质量控制方法研究,但是由于观测仪器的精度和测量方式不同,每种仪器产出资料的处理方式和方法参数都有所不同。例如,针对CTD传感器所携带的电导率、温度、压力三种传

感器由于各自响应时间不同,如不进行调整可能造成传感器不同步,出现明显毛刺,以及由于一些移动漂流浮标在长期运行中无法对电导率传感器进行订正而出现的温盐度漂移误差等。

国际海洋界高度重视观测资料的质量控制。美国国家海洋和大气管理局(NOAA)是国际上较早系统地开展水文气象资料质控研究的机构。英国国家气象局(Met Office)研制了海洋温盐自动质控系统,涵盖常规质控的基本功能。以此为基础,英国国家气象局哈德莱中心(Met Office Hadley Centre)研制了逐月(1900年至今)温盐网格化产品(EN4),其“背景检验”的背景场不是气候态场,而是基于分析矫正方案得到,部分质控采用先进的统计方法实现,后续进一步推广到多变量的质控。此外,欧洲中尺度天气预报中心(ECMWF)发展了一种变分质量控制(VarQC)方案,将质控与分析融为一体。我国也逐步建立了较完善的常规质控系统。国家海洋信息中心作为中国海洋信息资源的管理机构,以及全球海洋与海洋气候资料中心中国中心,已建立业务化的常规海洋环境资料处理及质量控制体系(纪风颖等, 2016; 刘首华等, 2016)。自然资源部第二海洋研究所在Argo实时/延时资料质控方面做出了卓有成效的工作。在海洋温盐数据质量控制方面,主要通过处理原始数据,建立元数据集和标准数据集,通过质量控制系统和专家审核等手段进行数据质量控制,取得显著成效(向先全等, 2015; 杨扬等, 2017; 于婷等, 2013)。中国科学院大气物理研究所基于XBT、CTD、Argo、Glider等所有可搜集到的现场观测数据,利用新的XBT数据偏差订正方案(CH14)和“重采样测试”评估方案等技术,研制了全球海洋温度数据集和全球海洋盐度格点数据产品(Cheng et al, 2016)。

目前海洋行业获取的大量实时和延时温盐数据存在长时间跨度、海区表现特征各异和观测仪器处理及精度差异等问题,这对计算机自动质量控制和专家审核工作提出了更高的要求,传统处理方式对海量多关联数据进行深入挖掘分析的能力存在明显

的不足。海洋数据采集过程逐时、逐分甚至逐秒发生，具有高速更新性特征，传统处理方式对海量和多源数据质量智能快速判别与大数据存在明显的差异，无法有效应对数据巨量并发。对于海洋环境时空变化特征，虽已有很多科研成果，但缺乏基于观测资料的自动化定量分析方式。由于海洋环境具有非常复杂的多尺度变化特征，如何基于算法对一些基本时空特征进行刻画和提取，并辅助用于海洋环境资料的质量控制，是需要解决的关键问题。

目前海洋观测呈现出多源化、立体化、实时化的特征。探索以大数据、人工智能和深度学习为核心的数据处理和海洋现象识别技术逐步成为研究热点。深度学习已经在解决语音感知、目标识别、计算机视觉、自然语言处理等人工智能相关任务中取得了突破性进展 (Lecun et al, 2015)，人工智能方法预测海洋环境信息取得了局部的成果。徐凌宇等 (2018)、郝滢洁等 (2017) 和杜艳玲等 (2017) 分别用 CNN (Convolutional Neural Networks, CNN) 和多层网络模型 DeepEddy 进行海洋中尺度涡旋检测，以期提高涡旋检测的效率和精确度。邹国良等 (2017) 采用深度学习算法来实现海浪等级的划分，经过加权均值滤波去噪，生成与海浪等级对应标签，实现了海浪样本数据集的构建。卢鹏等 (2018) 提出了一种基于信息熵的卷积网络模型 IQA-CNN，进行无参考图像质量评价方法研究。路志英等 (2018) 提出基于物理量参数和深度学习模型 DBNs 的短时强降水天气识别模型。黄冬梅等 (2018) 将 CNN 和深度置信网络 (DBN) 用于海冰的冰水分类，评估不同类型深度学习模型在 SAR 影像海冰分类方面的性能及其影响因素。国际海洋质量控制数据库 (International Quality Controlled Ocean Database, IQuOD) 开发了一种“智能元数据”算法，利用探测器的已知信息 (部署国家、年份、最大深度) 来推断缺失的元数据 (Palmer et al, 2018)，同时完成了基准测试和自动质量控制 (QC) 方法，但深度学习在海洋资料质量控制理论领域还没有开展详细的应用研究。

海洋资料的质控，需针对不同海域，研究水文、气象要素之间的相关性，并结合实际观测分布情况，进行影响质控结果的敏感性等分析。本文在继承海洋资料常规质量控制方法的基础上，考虑到海洋的多要素、海气相互作用的复杂性以及海洋

同步观测资料的数量等多种因素，采用深度学习技术，开展海温资料质量控制多分类算法与应用研究，实现多源观测资料实时快速质量控制判别，提升海量海洋资料处理分析的能力。通过人工合成少数类样本和加权损失函数方法减少多数类的频率来降低数据的不平衡。基于 MLP 和 DNN，构建了两个深度学习模型来进行海温资料质量符判别，并对深度学习的质量控制模型和传统方法质量控制的判别精度进行对比验证分析，对分类精度和分类概率进行了验证分析。

# 1 资料和方法

## 1.1 资料

本文采用 1955—2000 年 CTD 等多源海洋观测仪器获取的 1 044 386 条西太平洋菲律宾和印度尼西亚外海海域的历史观测资料，针对不同仪器获取的海洋温度和盐度资料质量情况分别进行了严格的计算机自动质量控制和人工审核，资料内容包括观测层深度 (Depth)、温度 (Temp)、盐度 (Psal)、温度质量符 (Temp\_QC\_flags) 和盐度质量符 (Psal\_QC\_flags)。温度和盐度质量符定义如表 1 所示，其中单样分析表示观测条件不允许的情况下，不能实现双样或多样分析，而采用单样分析。

表1 温度和盐度质量符定义

| 质量符 | 含义          |
|-----|-------------|
| 0   | 没有质量控制      |
| 1   | 正确数据        |
| 2   | 可能正确        |
| 3   | 可能错误        |
| 4   | 错误数据        |
| 5   | 数据在质量控制时被更改 |
| 6   | 单样分析        |
| 7   | 没进行器差订正     |
| 8   | 插值数据        |
| 9   | 数据缺失        |

## 1.2 特征分析与预处理

从温度资料的剖面分布可以看出 (图 1)，温度和深度两个变量为非线性关系 (pearsonr = -0.59, p = 0)。对 1955—1999 年温度剖面资料质量符分布统计发现 (表 2)，温度剖面资料深度范围为 0~8 000 m，主要集中在 5 000 m 以浅的深度。原始资

料中正确数据 (QC1) 的温度质量符共 843 456 条记录, 可能正确 (QC2)、可能错误 (QC3) 和数据缺失 (QC9) 分别为 20 条、26 条和 884 条记录, 除 QC1 外, 其余质量符仅占全部记录的

0.11%。与目前国际上广泛使用的深度学习数据集相比 (ImageNet), 本文采用的温盐剖面资料具有质量符分布极度不平衡的特点 (Krizhevsky et al, 2012; Russakovsky et al, 2015)。

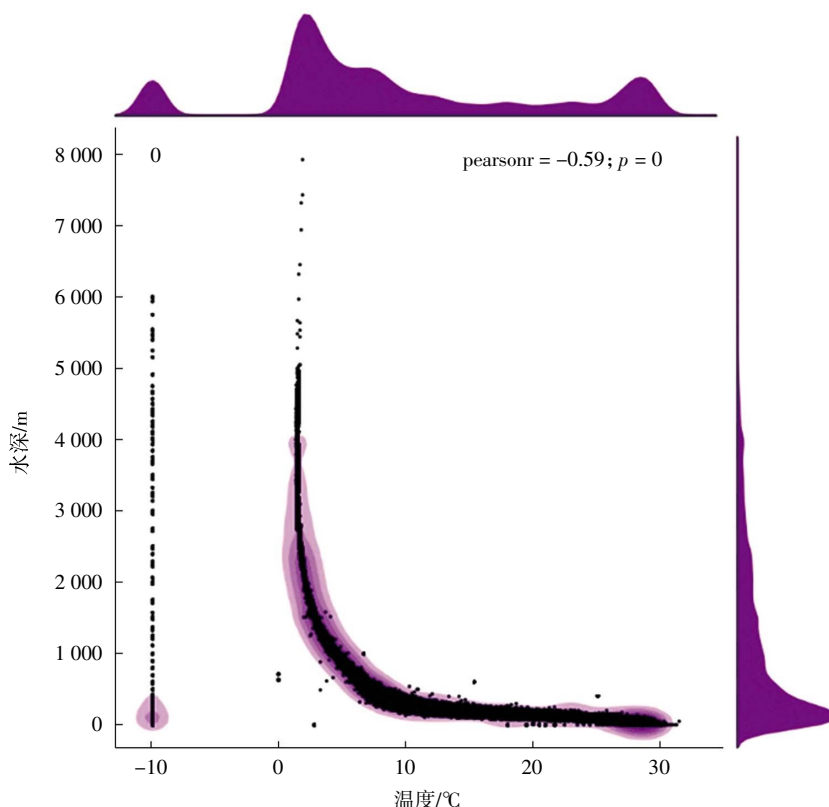


图1 1955—1999年海温资料剖面分布情况 (顶部和右侧填色分别为海温数值和深度的核密度分布)

本文根据温度剖面资料特点构造训练集, 并在训练集数据上进行资料预处理操作, 即对每个数据的特征都进行零中心化, 然后将其数值范围都归一化到  $[-1, 1]$  范围之内 (图 2)。本文资料预处理操作都只在训练集数据上进行计算, 算法训练完毕后再应用到测试集上。

### 1.3 质量符分布不平衡处理

由于温盐剖面资料质量符分布极度不平衡, 本文通过人工合成样本方法来提高少数类分布。依据质量控制方法和质量控制参数, 人工合成部分质量控制记录, 以此来增加训练集中少数类的频率。通过回溯常规质量控制流程中质量符产生流程, 本文采用的温盐剖面资料少数类质量符主要由着陆检验、观测要素范围检验、连续性检验、尖峰检验、密度反转检验、梯度检验和异常值检验产生。根据国家海洋信息中心温盐观测历史资料统计得到的质控参数为依据, 分别人工合成 QC2 和 QC3 质量符

8 153 条和 10 659 条; 以 -9.99 为参数, 人工合成 QC9 质量符 6 080 条。人工合成样本合理地增加了少数类样本分布同时保持原有的海温资料特征, 将分类不平衡性从 1 000 : 1 降至 33 : 1, 同时避免了随机重采样等方法可能引入无意义的甚至有害的新样本。值得一提的是, 以上样本不均衡预处理方法仅在训练集数据上进行, 而验证集和测试集不做处理, 保持和真实海洋环境中海温资料的不均衡分布一致。

上述人工合成样本方法是从样本的层面去克服不平衡, 本文进一步通过构造加权损失函数 (Loss) 方法, 在深度学习模型算法层面来克服样本不平衡。对于每个样本, 计算的损失函数乘上对应的分类权重来计算最终的损失, 其中四种分类计算的损失为:

$$\text{Loss} = 0.1 \times \text{loss1} + 1 \times \text{loss2} + 1 \times \text{loss3} + 0.5 \times \text{loss9}$$

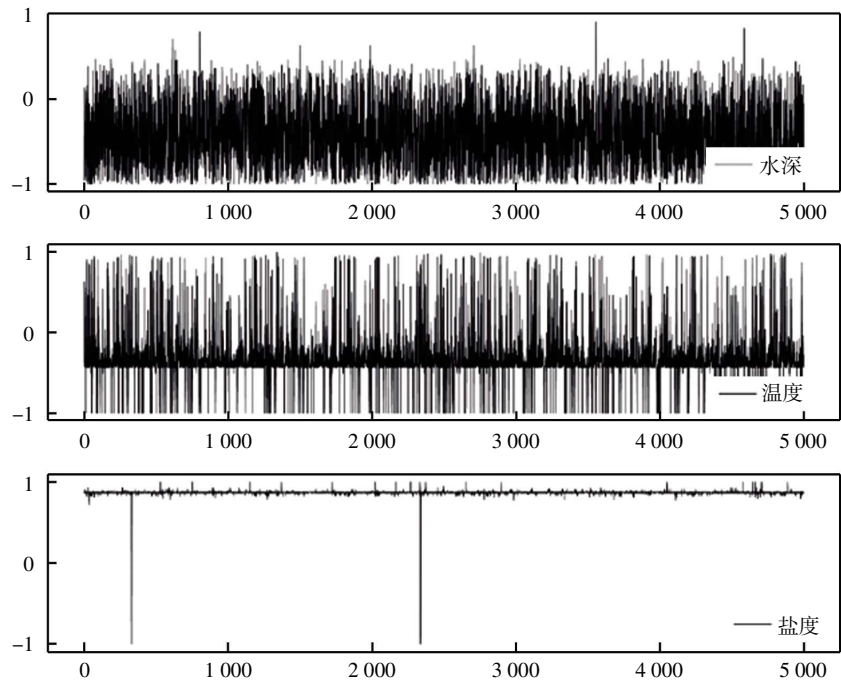


图2 标准化处理后部分输入变量（前5 000条记录）

表 2 1955—1999 年温度剖面资料质量符分布统计

| 温度质量符 | 原始记录数   | 处理后记录数  | 分类权重 |
|-------|---------|---------|------|
| QC1   | 843 456 | 843 456 | 0.1  |
| QC2   | 20      | 8 173   | 1    |
| QC3   | 26      | 10 685  | 1    |
| QC9   | 884     | 6 964   | 0.5  |
| 总计    | 844 386 | 869 278 |      |

2 建立分类模型

深度学习是研究从观测样本出发寻找规律的现代智能技术中的重要方法，利用这些规律对数据情况进行预测或判别。根据学习模式、学习方法以及算法的不同，机器学习存在不同的分类方法。全连接神经网络用于处理向量（1D 张量）数据，输入数据特征中没有特定结构。与二维 CNN 仅查看局部关系不同，全连接神经网络可以映射任意两个输入特征之间的关系，从输入特征空间中学到的是全局模式，最常用于数据分类。CNN 能够查看空间局部模式，其方法是对输入张量的不同空间位置或图像应用相同的几何变换，这样得到的处理结果具有平移不变性（translation invariant），CNN 可以学到模式的空间层次结构（spatial hierarchies of patterns），使得卷积层能够高效利用数据且能够高度模块化。

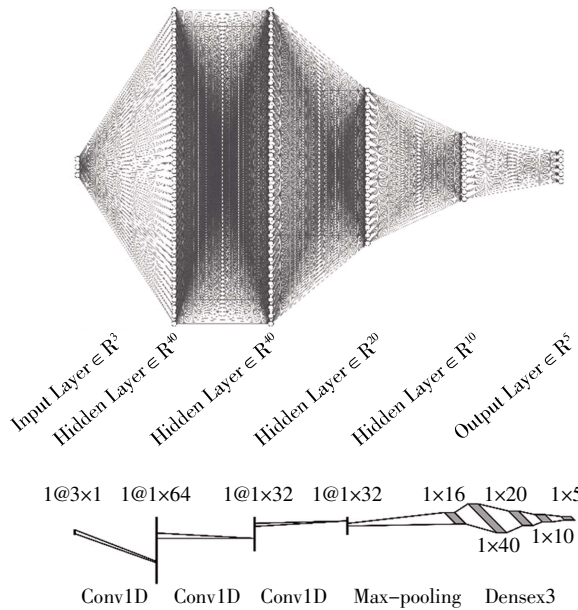


图3 MLP 结构（上）和 DNN 结构（下）

本文基于温盐剖面资料质量符分布不平衡的历史资料特征，分别构建了由多层全连接神经网络构成的 MLP 和由一维 CNN 与多层全连接神经网络构成的 DNN 来进行海温资料质量符的多类判别应用研究（图 3）。两个模型均使用柔性最大值（softmax）作为激活函数，对于每一个输入样本，模型都会输出一个多维向量，其单元个数等于类别个数，模型

将输出在不同输出类别上的概率分布, 概率的总和为 1。采用独热编码 (one-hot) 将分类目标转换为向量, 使用交叉熵 (categorical\_crossentropy) 作为分类损失函数, 同时对不同类对应的损失增加权重, 以平衡这种数据偏差, 改进模型的泛化能力。采用随机失活 (Dropout) 正则技术 (Lecun, Bengio, Hinton, 2015), 在训练中每次更新参数按照一定概率随机断开输入神经元, 防止过拟合, 增强模型的鲁棒性。本文利用平衡后归一化 848 303 条记录作为训练集, 2000 年 20 975 条温盐剖面资料作为测试集, 输入要素为 Depth, Temp 和 Psal, 输出要素为 Temp\_QC\_flags。

### 3 模型评估

MLP 和 DNN 两个深度学习多分类模型训练 10 000 次参数更新迭代后, 模型参数在测试集上准确率超过 92%, 测试数据集上预测损失趋于稳定 (图略)。采用提前终止 (Early stopping) 机制分别保存最优模型参数, 其中 DNN 模型具有很好的鲁棒性和泛化能力。基于构建的模型和最优参数, 对测试的 20 975 条温盐剖面资料的温度资料质量情况进行了分类评估, 得到了模型预测输出的温度资料质量符分类结果在各个类别上的概率分布, 所有类别上的输出概率都不小于 0, 且所有类别上的输出概率和等于 1。

#### 3.1 分类精度验证

测试集在 多分类质量控制模型的 TOP-1 (模型概率分布最大的输出类别且分类正确) 分类结果显示 (图 4): 测试集海温资料在 5 000 m 以浅深度资料分布较为均匀, 具有很好的代表性。两个分类模型 (MLP 和 DNN) 均可以正确地识别出不同深度情况下海温资料的正确数据 (QC1, 红色圆点) 和数据缺失 (QC9, 灰色圆点) 两种质量符, 没有误判情况产生; 对于可能正确 (QC2, 蓝色圆点) 和可能错误 (QC3, 绿色圆点) 两种质量符的分类结果显示, 模型对部分偏离海温分布密集区的孤立值有较好的识别。两个模型在测试集温度质量符的分类准确率分别达到 99.63% 和 99.69%, DNN 模型识别准确度略高于 MLP 模型, 其中 DNN 模型的错误分类主要集中在 1 000 m 以浅区域, 特别是 18~22 °C 的范围, 而 MLP 模型在 4 000 m 左右深

度, 额外把两个 QC1 错误分类为 QC2; 同时在 400 m 左右的一个 25 °C 可能正确值分类为正确。本文构建的 MLP 和 DNN 多分类质量控制模型和最优参数, 可初步实现温度资料质量控制符的准确输出, 为数据质量控制符综合研判提供参考。

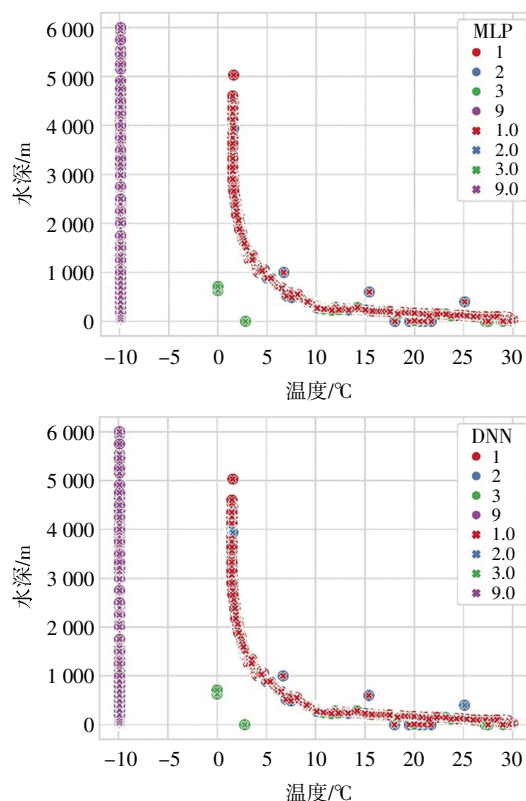


图 4 MLP (a) 和 DNN (b) 测试集 TOP-1 分类结果  
(其中圆点表示质控结果, x 表示模型分类结果。圆点和 x 重合且颜色一致表示模型分类正确, 而颜色不一致表示模型分类错误)

在 多分类任务中, 精准率 (precision), 召回率 (recall) 和 F1 分数 (F1-measures) 可以独立地应用于每个标签 (Everingham et al, 2005)。精准率表示在所有被标记为正的样本中实际为正的样本的概率, 召回率表示分类器在实际为正的样本中被标记为正样本的概率 (Flach et al, 2015)。F1 分数是精准度和召回率的加权调和平均值, 最佳值为 1, 最差分数为 0 (Davis et al, 2006)。本文构建的 MLP 和 DNN 模型在测试集的 TOP-2 分类精度 (模型概率分布最大的 2 个输出类别且分类正确) 情况如表 3 所示, 两个模型对 20 000 条 QC1 和 884 条 QC9 的精准率、召回率和 F1 分数均为 1, 正确识别率达 100%。MLP 和 DNN 分类模型对 QC2 和 QC3 的 F1 分数分别为 0.91 和 0.90, 0.93 和 0.96,

DNN 模型对 QC2 和 QC3 分类结果均优于 MLP 模型，DNN 模型在 QC2 分类预测中 53 个测试样本中有 46 个正确分类，剩余 7 个样本被分类标记为 QC1；对 QC3 分类结果中 38 个测试样本中有 35 个正确分类，剩余 3 个样本被分类标记为 QC1。两个模型（MLP 和 DNN 模型）在 20 975 条测试样本集的微平均（micro-averaging）和加权平均（weighted-averaging）（Tsoumakas et al, 2009；Van Asch, 2013）的综合评分均为 1。宏观平均（macro-averaging）召回率评分分别为 0.91 和 0.93，宏观平均 F1 分数分别为 0.95 和 0.96。宏观平均因强调低精度分类的情况，因此评分偏低。相比于 TOP-1，TOP-2 分类精度对 QC2 和 QC3 的分类结果有很大改善，MLP 和 DNN 模型分别提高了 65.88% 和 46.38%。

表 3 MLP 和 DNN 模型基于测试集 TOP-2 分类精度

| 分类模型         | DNN |             |             | MLP |      |       | 样本数    |
|--------------|-----|-------------|-------------|-----|------|-------|--------|
|              | 精准率 | 召回率         | F1 分数       | 精准率 | 召回率  | F1 分数 |        |
| QC1          | 1   | 1           | 1           | 1   | 1    | 1     | 20 000 |
| QC2          | 1   | <b>0.87</b> | <b>0.93</b> | 1   | 0.83 | 0.91  | 53     |
| QC3          | 1   | <b>0.92</b> | <b>0.96</b> | 1   | 0.82 | 0.90  | 38     |
| QC9          | 1   | 1           | 1           | 1   | 1    | 1     | 884    |
| micro avg    | 1   | 1           | 1           | 1   | 1    | 1     | 20 975 |
| macro avg    | 1   | <b>0.93</b> | <b>0.96</b> | 1   | 0.91 | 0.95  | 20 975 |
| weighted avg | 1   | 1           | 1           | 1   | 1    | 1     | 20 975 |

3.2 分类概率分析

根据模型预测输出的温度资料质量符合分类结果在各个类别上的概率分布统计发现（Urbanowicz et al, 2015；Yang, 2001），在 TOP1 的正确分类概率中，本文构建的两个分类模型正确分类最高分类概率均接近于 1，且比第二高分类概率大两个数量级以上，而错误分类最高分类概率小于 0.97 或者第二高分类概率大于 0.015。根据以上统计结果，采用更加严格质量控制原则，对最高分类概率小于 0.98 或者第二高分类概率大于 0.01 或两个分类模型分类结果不统一的分类结果均标记为可疑结果，标记的可疑结果占测试集资料的 10.79%。

依据模型在不同输出类别上的概率分布向量，利用传统质量控制方法和流程，对标记的可疑结果进行二次质量控制，包括 5 类基础信息检验、12 类剖面特性检验和 3 类气候学检验，以及航迹检

验、剖面一致检验、可视化专家审核等可视化辅助检验，共计 5 级 26 类检验，对可疑结果的质量控制符进行综合研判，得到最终的温度资料质量控制符。基于本文构建的 MLP 和 DNN 多分类质量控制模型和传统质量控制方法流程相结合的方式，可减少干扰因素，大幅降低传统质量控制方法的工作量。

4 结论与讨论

本文采用深度学习技术，开展了海温资料质量控制多分类算法与应用研究，构建了 MLP 和 DNN 两个海温资料质量控制符分类深度学习模型，针对海温数据的质量控制符进行分类特征分析。通过人工合成少数类样本和加权损失函数方法，增加少数类数量、降低多数类的频率来平衡训练数据样本，实现了海温观测数据质量实时快速质量控制。结果表明：深度学习方法可应用于海洋观测数据的质量控制，并有效提升海量海洋数据处理分析能力。

（1）本文构建的 MLP 和 DNN 两个深度分类模型能够较为准确快速地识别海温资料质量，TOP-1 分类准确率在 20 975 条温盐剖面资料测试集中分别达到 99.63% 和 99.69%，其中分类精准率、召回率和 F1 分数，都有着较好的表现，QC1 和 QC9 的正确识别率均达到 100%。

（2）两个深度分类模型 TOP-2 分类精度对 QC2 和 QC3 的分类结果有很大改善，MLP 和 DNN 模型分别提高了 65.88% 和 46.38%，其中 DNN 分类模型对 QC2 和 QC3 的 F1 分数分别为 0.93 和 0.96，MLP 分类模型对 QC2 和 QC3 的 F1 分数分别为 0.91 和 0.90。测试集的微平均和加权平均的综合评分均为 1，宏观平均召回率评分分别为 0.91 和 0.93，宏观平均 F1 分数分别为 0.95 和 0.96。DNN 模型分类结果优于 MLP 模型，具有很好的鲁棒性和泛化能力。

（3）基于本文构建 MLP 和 DNN 多分类质量控制模型和最优参数，可初步实现温度资料质量控制符的多分类输出，达到了较高的准确度，可作为传统质量控制方法的补充结果，为资料质量控制符综合研判提供参考，同时减少大数据量干扰因素，大幅降低传统质量控制方法的工作量。值得一提的是，根据训练数据的特点，本文构建的深度学习方

法模型只适用于本文研究的海域,在其他海域,特别是温跃层、海洋峰或涡旋活跃的海域,需要根据训练数据量和海域特征来构造不同的模型,同时其他输入要素的选择(如降水、台风等)也可能对基于深度学习的海温资料质量控制符分类有重要作用。

深度学习模型的构建和最优参数选取,受输入数据种类、数量和质量的影响较大,因此需要加入更多高质量、长序列、多要素的海洋观测资料数据集。本文中使用 1955—2000 年,菲律宾和印度尼西亚外海海域的多源海洋观测仪器获取的百万条历史观测资料作为数据集进行模型构建,后续将对 2000 年以后,西太平洋和中国近海附近海域的多种海洋观测资料进行分类质量控制。通过应用数据的转换、清洗、抽取和机器学习等技术方法,发现数据质量关联和无效数据,优化质量控制标识,降低质量控制误判和人工审核强度,从而逐步提升海洋水文气象观测资料质量。

本文构建的深度学习方法模型,输入数据为单一深度的温度、盐度、深度和温度质量符四种变量,数据关联挖掘上仅做了温度和盐度的关联检验,同时质量符综合地划分为四类,未开展单项检验判别学习。下一步将主要从三个方面开展深入研究,第一是多要素关联分析研究,目前质量控制方法多是针对一类要素单独地进行质量控制或者局限于一两种要素,而海洋观测要素之间存在相互影响,需要纳入更多高质量、多学科要素进行综合学习,发现数据之间隐藏的问题和关联。第二,观测数据不仅需要判别正误,而且需要判别问题类型。下一步将针对深度递增检验、剖面廓线检验、密度反转检验、剖面尖峰检验等剖面特性和气候范围等开展标识学习,发现错误分类,评估数据质量,提高和改善数据应用价值。最后开展深度学习模型泛化能力研究,下一步工作中将搜集更多海域的数据,对该模型效果在第三方水域进行检验。

海洋环境分析和预报,核心是观测资料。数据驱动的本质决定了它对观测资料有更高的要求,特别是数据的质量控制,海洋水文气象资料联合质控及自动化与智能化是未来的发展方向。软硬件设施及数据获取能力的差异,决定了需针对不同平台开展智能质控算法研究并进行优化。可以预见,随着观测技术及人工智能技术的发展,海洋环境智能质控是未来的发展趋势。虽然深度学习可以避开针对

输出要素形成的复杂机制,但这并不代表可以抛弃海温与输入要素间的物理机制分析,输入要素的种类及质量对模型效果会产生较大影响,后续也将进一步考虑应用场景,继续开展参数优化拓展应用范围,同时通过从理论上分析传统方法的不足,结合物理机制分析来提高质控精度。

## 参 考 文 献

- 杜艳玲, 2017. 基于海洋遥感影像的中尺度涡自动识别及与渔场动态关系研究[D]. 上海: 上海海洋大学.
- 郝滢洁, 2017. 基于卷积神经网络的海洋中尺度涡旋检测算法研究[D]. 青岛: 山东科技大学.
- 黄冬梅, 李明慧, 宋巍, 等, 2018. 卷积神经网络和深度置信网络在 SAR 影像冰水分类的性能评估 [J]. 中国图象图形学报, 23(11): 1720–1732.
- 纪凤颖, 林绍花, 万芳芳, 等, 2016. Argo/GTSPP 与 WOD 数据集及其应用中需注意的若干问题[J]. 海洋通报, 35(2): 140–148.
- 姜晓轶, 潘德炉, 2018. 谈谈我国智慧海洋发展的建议 [J]. 海洋信息, 33(1): 1–6.
- 李伟, 王玉衡, 汪嘉宁, 等, 2012. 2011 年春、夏季黄、东海水团与水文结构分布特征[J]. 海洋与湖沼, 43(3): 615–623.
- 刘首华, 陈满春, 董明娟, 等, 2016. 一种实用海洋浮标数据异常值质控方法[J]. 海洋通报, 35(3): 264–270.
- 卢鹏, 林根巧, 邹国良, 2018. 基于信息熵和深度学习的无参考图像质量评价方法研究[J]. 计算机应用研究, 35(11): 3508–3512.
- 路志英, 任一墨, 孙晓磊, 等, 2018. 基于深度学习的短时强降水天气识别[J]. 天津大学学报, 51(2): 111–119.
- 王东晓, 陈举, 陈荣裕, 等, 2004. 2000 年 8 月南海中部与南部海洋温、盐与环流特征[J]. 海洋与湖沼, 35(2): 97–109.
- 王辉赞, 张韧, 王桂华, 等, 2012. Argo 浮标温盐剖面观测资料的质量控制技术[J]. 地球物理学报, 55(2): 577–588.
- 向先全, 路文海, 杨翼, 等, 2015. 海洋环境监测数据集质量控制方法研究[J]. 海洋开发与管理, (1): 88–91.
- 徐凌云, 张高唯, 江湾湾, 等, 2018. 深度学习神经网络及其在海洋环境信息挖掘预测中的应用[J]. 海洋信息, 33(1): 17–23, 35.
- 杨扬, 苗庆生, 韦广昊, 等, 2017. 海洋站观测资料的质量控制方法及其应用[J]. 海洋开发与管理, 34(10): 109–113.
- 于婷, 刘玉龙, 杨锦坤, 等, 2013. 实时和延时海洋观测数据质量评估方法研究[J]. 海洋通报, (6): 610–614, 625.
- 邹国良, 陈长吉, 郝剑波, 2017. 构建适用于深度学习的海浪样本数据集的并行算法实现及性能优化 [J]. 计算机应用与软件, 34(9): 57–63.
- CENTURIONI L, HORÁNYI A, CARDINALI C, et al. 2017. A global ocean observing system for measuring sea level atmospheric pressure: Effects and impacts on numerical weather prediction [J]. Bulletin of the American Meteorological Society, 98(2): 231–238.
- CHENG L, ABRAHAM J, GONI G, et al. 2016. XBT Science: Assessment of instrumental biases and errors [J]. Bulletin of the American Meteorological Society, 97(6): 924–933.

- DENG Z, WEI G, TING Y U, et al. 2016. The diapycnal mixing in the upper Pacific estimated from GTSP observations [J]. *Acta Oceanologica Sinica*.
- DAVIS J, GOADRICH M. 2006. The relationship between Precision-Recall and ROC curves[C]. *Proceedings of the 23th International Conference on Machine Learning*, 2006.
- EVERINGHAM M, ZISSERMAN A, WILLIAMS C, et al. 2005. The 2005 PASCAL Visual Object Classes Challenge [C]. Springer Berlin Heidelberg.
- FLACH P, KULL M. 2015. Precision-recall-gain curves: PR analysis done right[C].
- HAMILTON D. 1994. GTSP builds and ocean temperature-salinity database [J]. *Earth System Monitor*, 4(4): 4-5.
- IMIS V. 2010. GTSP Real-Time Quality Control Manual [J]. *Ioc manuals & Guides*.
- KRIZHEVSKY A, SUTSKEVER I, HINTON G E. 2012. ImageNet classification with deep convolutional neural networks[C].
- LECUN Y, BENGIO Y, HINTON G. 2015. Deep learning [J]. *Nature*, 521(7553): 436.
- LEVITUS S, ANTONOV J I, BARANOVA O, et al. 2013. The World Ocean Database [J]. *Data Science Journal*, 12.
- LEVITUS S, ANTONOV J I, BOYER T P, et al. 2000. Warming of the World Ocean [J]. *Science*, 287(5461): 2225-2229.
- LOCARNINI R A, MISHONOV A V, ANTONOV J I, et al. 2013. World ocean atlas 2013. Volume 1, Temperature[C]. U.S. Government Printing Office.
- OKA E, ANDO K. 2004. Stability of Temperature and Conductivity Sensors of Argo Profiling Floats [J]. *Journal of Oceanography*, 60(2): 253-258.
- PALMER M D, BOYER T, COWLEY R, et al. 2018. An algorithm for classifying unknown expendable bathythermograph (XBT) instruments based on existing metadata [J]. *Journal of Atmospheric and Oceanic Technology*, 35(3): 429-440.
- REAGAN J, BOYER T P, ANTONOV J I, et al. 2014. Comparison analysis between Aquarius sea surface salinity and World Ocean Database in situ analyzed sea surface salinity[J]. *Journal of Geophysical Research*, 119(11): 8122-8140.
- REUL N, CHAPRON B, LEE T, et al. 2014. Sea surface salinity structure of the meandering Gulf Stream revealed by SMOS sensor [J]. *Geophysical Research Letters*, 41(9): 3141-3148.
- RUSSAKOVSKY O, DENG J, SU H, et al. 2015. ImageNet Large Scale Visual Recognition Challenge [J]. *International Journal of Computer Vision*, 115(3): 211-252.
- SUMMERHAYES C. 2002. The global ocean observing system (GOOS) in 1998[J]. *Elsevier oceanography series*, 66: 57-66.
- TSOUMAKAS G, KATAKIS I, VLAHAVAS I. 2009. Mining multi-label data. *Data mining and knowledge discovery handbook* [M]: Springer, 667-685.
- URBANOWICZ R J, MOORE J H. 2015. ExSTraCS 2.0: description and evaluation of a scalable learning classifier system [J]. *Evolutionary Intelligence*, 8(2-3): 89-116.
- VAN ASCH V. 2013. Macro-and micro-averaged evaluation measures [J]. *Tech. Rep.*
- WILSON J R. 1998. Global Temperature-Salinity Profile Programme (GTSP)-Overview and Future. [J]. *Unesco*.
- WONG A, KEELEY R, CARVAL T, et al. 2018. Argo Quality Control Manual for CTD and Trajectory Data [J].
- YANG Y. A 2001. A study of thresholding strategies for text categorization[C]. *ACM*.

(本文编辑：王少朋)