DOI: 10.13382/j.jemi.B2104288

探空温度传感器误差预测技术研究*

杨加春[1] 王彦明[2,3,4] 李庆军[1] 贾克斌[3,4] 刘鹏宇[3,4]

(1. 天津华云天仪特种气象探测技术有限公司 天津 300392;2. 石家庄邮电职业技术学院 石家庄 050021;
3. 北京工业大学信息学部 北京 100124;4. 北京工业大学 计算智能与智能系统北京市重点实验室 北京 100124)

摘 要:随着气候诊断、气候变化、天气预报等学科的深入开展,对探空温度传感器的测量精度提升到了 0.1 ℃的量级要求,而由于太阳辐射、升空速度、入云出云等因素的干扰,引起的测量误差可达 3℃甚至更高,已成为制约气象探测精度提升的主要障碍。针对此问题,首先通过三维建模及流体力学分析,得到了温度传感器最优的设计方案,从传感器形态设计上实现了测量误差最小化。然后对历史气象探测数据进行分析和汇总,构造出国内首个基于真实环境的、包含 900 000 条探测记录的高空气象探测数据集,以解决仿真环境与真实环境存在偏差的问题。最后,将 Morlet 小波作为深度神经网络的激活函数,并将支持向量机、XGBoost、深度神经网络、线性回归相融合,构造出一个针对探空温度传感器测量误差的预测模型。经过本文所提出的误差预测模型,平均误差从 0.817 降低到了 0.008,均方误差从 0.878 降低到了 0.068,标准差从 0.458 降低到了 0.204,拟合系数 R^2 为 0.93,使温度传感器的测量精度得到显著提升,更有利于气象学科相关内容的展开。

关键词: 气象探测;温度传感器;误差预测;特征工程;模型融合

中图分类号: TP212.1;TN06 **文献标识码**: A **国家标准学科分类代码**: 460.4020

Research on error prediction technology of radiosonde temperature sensor

Yang Jiachun[1] Wang Yanming[2,3,4] Li Qingjun[1] Jia Kebin[2,3] Liu Pengyu[2,3]

(1. Tianjin Huayuntianyi Special Meteorological Detection Technology Co., Ltd, Tianjin 300392, China;
2. Shijiazhuang Posts and Telecommunications Technical College, Shijiazhuang 050021, China;
3. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;
4. Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124, China)

Abstract: With the development of disciplines such as climate diagnosis, climate change and weather forecasting, the measurement accuracy of the sounding temperature sensor has increased to the order of 0.1℃. Due to the interference of factors such as solar radiation, lift-off speed, cloud in and out of the cloud, the sensor measurement error can reach 3℃ or even higher, which has become the main obstacle restricting the improvement of meteorological detection accuracy. Aiming at this problem, firstly, the optimal design scheme of the temperature sensor is obtained through three-dimensional modeling and fluid mechanics analysis. The measurement error is minimized from the sensor morphology design. Then, by analyzing and summarizing historical meteorological observation data, the first domestic high-altitude observation dataset containing 900 000 detection records based on the real environment was constructed to solve the problem of deviation between the simulated environment and the real environment. Finally, Morlet wavelet is used as the activation function of the deep neural network. The support vector machine, XGBoost, deep neural network, and linear regression are combined to construct a prediction model for the measurement error of the sounding temperature sensor. After the error prediction model proposed in this paper, the average error is reduced from 0.817 to 0.008, the root mean square error is reduced from 0.878 to 0.068, the standard deviation is reduced from 0.458 to 0.204, and the fitting coefficient R^2 is 0.93. The measurement accuracy of the temperature sensor has been significantly improved, which is more conducive to the development of relevant content of the meteorological discipline.

Keywords: aerological sounding; temperature sensor; error prediction; feature engineering; model fusion

收稿日期: 2021-05-14 Received Date: 2021-05-14

*基金项目:国家重点研发计划(2018YFC1506202)项目资助

0　引　言

近年来,随着全球工业化进程的加快,对环境的污染也显著增加,由此造成了雾霾、酸雨、全球变暖等极端现象频发。根据世界气象组织、联合国减灾署和世界银行集团等 16 个国际机构联合发布的《2020 年气候服务状况报告》,在近 50 年里,全球共发生 1.1 万多起由于天气、气候和降水导致的自然灾害,造成了约 200 万人死亡和 3.6 万亿美元的损失[1]。同时报告指出,由于气候变化的复杂性,极端天气现象发生的频率、程度以及严重性均有所增加,预计到 2030 年,全球因自然灾害而需要救援的人数将比 2018 年多 50%,每年因此造成的开销约为 200 亿美元。世界气象组织指出,准确地气象预警系统可以有效地减少灾害风险,若在正确的时间、正确的地点做出相应的措施,可减少财产损失,保护生命安全。

进行气象预警系统构建时,各海拔下的温度数据能够反映大气热力过程,是进行气候预测、气候变化估计和自然灾害判断等方面的研究基础[2]。高空温度、湿度、气压等气象数据的精确测量,可以有效地提升天气种类的预测精度,特别是对于局部地区的天气预测,可以提升到小时的预测精度[3];进一步,相关数据的综合分析可以揭示上空云的种类、雾的结构,从而对大气的状态进行推断,以预测空气中存在的云状及云量,预测雾、霜等天气产生及消退时间[4]。

为确保探空仪测量数据的准确性及探测过程的通风性,温度传感器通常直接暴露在探空仪主体上方,使温度传感器直接受到外界因素的干扰。当太阳光照射到温度传感器表面时,其表面材料会吸收相应的辐射能量,从而使温度传感器表面温度上升,其效果相当于给温度传感器加热,使其测量值高于大气的温度数值[5-7]。因此,为了获得真实的大气温度数值,有必要研究各个外部环境对温度传感器测量结果的影响,以对探空仪测量的数据进行修正,从而最大限度地获取高空中真实的大气温度。

在 1986 年,美国航天局率先公布了在温度传感器模型构建方面的研究成果,实现了对探空仪热敏电阻误差的初步估计[8]。在所构建的模型中,通过使用参数 H 来表示热敏电阻的对流传热系数,以此来计算探空仪温度传感器所受到的环境干扰误差。在气象探测过程中,探空仪所遭受的气流方向、热敏电阻材料、引线粗细等因素均会对温度传感器测量结果造成影响,并且所构建的模型也只针对特定型号、特定尺寸、特定安装方式的温度传感器,适用面较为狭窄,很难推广到不同场景、不同类别下的探空传感器中[9]。

在 2003 年,Ruffieux 等[10]在真空实验箱内利用压力泵和隔热板搭建了一个简易的低气压辐射风洞,通过控制光源的亮度实现了探空仪温度传感器辐射误差研究。进一步,根据相关实验数据,构建了温度传感器误差预测方程,实现了传感器误差修正。

实际操作过程中,在仪器内部设计和产生 1 000 ~ 10 hPa 的风洞具有一定的难度,在气压低于 1 hPa 时更是存在许多技术难点。在太阳辐射仿真方面也存在很大的困难,要在风洞中仿真出太阳辐射的真实情况,不仅需要寻找光谱特性逼真、稳定性强、均匀性高的光源,同时也要避免仪器内壁对光源能量反射所造成的影响,否则可能会造成其他辐射误差[11]。风洞实验在技术、场地、经济等方面的投入巨大,目前尚无成熟的、可靠的、经济实用的测试设备提供给相关人员使用。

随着计算机科学的发展,通过流体动力学对探空温度传感器进行建模,成为了目前传感器误差预测方面的主要研究方法。刘清惓等人针对太阳辐射导致温度传感器测量结果存在偏差的问题,通过计算流体动力学方法对传感器从地面到 32 km 高空进行了热数值模拟分析[12]。在仿真分析过程中,充分考虑了外部环境对传感器测量误差的影响并通过构建对流——太阳辐射耦合边界条件,得出了不同涂层反射率、不同电阻尺寸、不同海拔高度下太阳辐射强度与温度传感器测量误差之间的关系曲线。

通过流体力学仿真得到的实验数据与实际测量结果之间的准确性未得到验证,所得结论也只是保留在理论阶段,由仿真数据得到的传感器形态设计方案、测量误差预测方程并未在实际中得到应用,还需进一步在实际环境中对仿真结果进行检验。

本文结合探空仪放飞过程中采集到的真实数据,开展高空动态环境下温度传感器测量误差预测方法研究。

1）探空仪温度传感器误差分析

从探空传感器形态设计和高空环境两个方面对传感器测量误差进行分析。在传感器形态设计方面,利用流体动力学仿真方法建立起传感器涂层反射率、传感器尺寸、传感器引线夹角与传感器测量误差之间的关系,以此来指导设计最佳的传感器结构

2）高空探测数据集构建

与传统探空传感器误差预测方法不同,本文实验数据与探空仪制造商合作,对探空仪放飞过程中采集到的时间、气压、温度、湿度、风速、经度、纬度等 15 个与高空气象探测相关的特征数据作为基础。通过对采集到的原始数据进行分析,在相关理论的基础上利用特征工程方法构造出未能直接采集的、与温度传感器测量误差有关的特征变量,实现气象探测数据集的扩充,从而完成气象探测数据的构建,为温度传感器误差预测构建提供数据基础。

3）高空综合修正模型设计

为实现±0.01 ℃的温度传感器测量精度，将深度神经网络（DNN）、小波函数、支持向量机（SVM）、XGBoost、线性回归相融合，设计了一个针对高空探测温度传感器测量误差的综合预测模型。进一步，以实际气象探测过程中采集的数据为基础，对所提出的方法进行评估以验证模型的有效性。

1 探空传感器测量误差仿真分析

1.1 温度传感器仿真模型构建

为在不同的结构和参数下量化温度传感器的测量误差，通过使用三维实体建模软件 Pro/Engineer 建立珠状热敏电阻传感器的物理模型，其模型如图 1 所示。

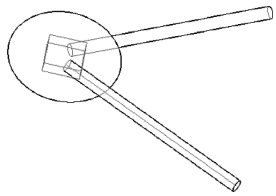


图 1 热敏电阻物理模型

Fig. 1 Physical model of thermistor

进一步，通过计算流体动力学方法对所建立的温度传感器模型进行四面体网格划分，并进行相应的多物理场耦合传热分析[13]。最终建立的数值计算模型如图 2 所示。

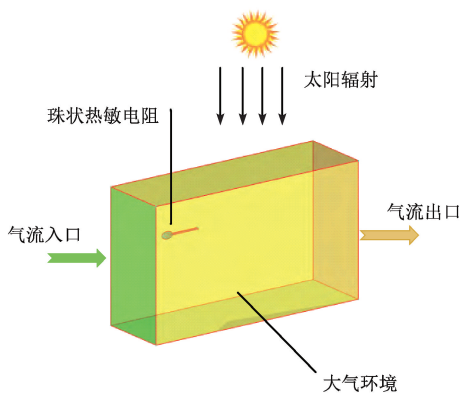


图 2 数值计算模型

Fig. 2 Numerical calculation model

在图 2 的数值计算模型中，长方体的左侧部分为输入气流，来模拟探空仪的上方大气状态；右侧部分为输出气流，来模拟探空仪的下方大气环境；中间部分将温度传感器置于其中，代表探空仪所处的大气状态，并辅助以太阳辐射来模拟太阳光照射对温度传感器测量结果的影响。仿真过程中，珠状热敏电阻在太阳辐射和大气热交换环境中的温度场分布如图 3 所示。

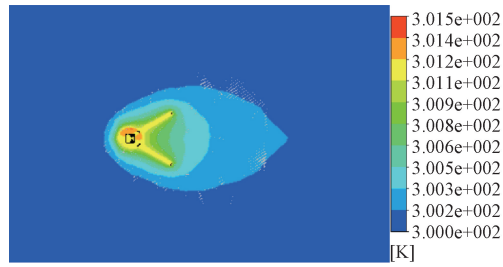


图 3 传感器温度场分布

Fig. 3 Sensor temperature field distribution

由图 3 可以看出，在太阳光照射下，珠状热敏电阻的温度环境高于周围空气中的温度，由此产生的温度差值即为太阳辐射误差[14]。随着空气尾流的流动，温度传感器周围的温度已逐步接近大气温度，与实际情况相符，证明了仿真环境搭建的有效性与真实性。

为了量化传感器在不同形态设计方面对测量结果的影响，本文从珠状热敏电阻的体积、引线夹角、引线直径三方面进行了相关仿真分析。

1.2 传感器体积与辐射误差的关系

通过对珠状热敏电阻从地面到 40 km 高空进行数值仿真分析，并将传感器体积设计为 0.6~1.6 mm^3，步长为 0.2 mm^3，得到了不同电阻下的辐射误差对应关系[10]，仿真结果如图 4 所示。

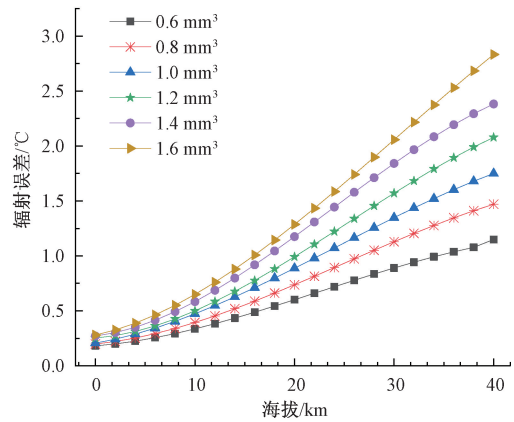


图 4 体积与辐射误差关系

Fig. 4 Diagram of volume and radiation error

由图 4 可以看出，在近地面，不同电阻尺寸之间的测量误差差距较小。随着海拔高度的增加，不同尺寸电阻间的测量误差显著增大，最大可相差 1.5 ℃。由图 4 可知，电阻尺寸越小所带来的的辐射误差也就越小，同时对传感器的制造流程、制造工艺要求也就越高。

1.3 传感器体积与辐射误差的关系

为获得探空仪在气象探测过程中由于引线夹角的不

同而对传感器测量结果的影响，通过将引线夹角设置为 0°、60°、120°、180°四种状态，来进行珠状热敏电阻数值仿真分析，得出了引线夹角与温度场分布如图 5 所示。

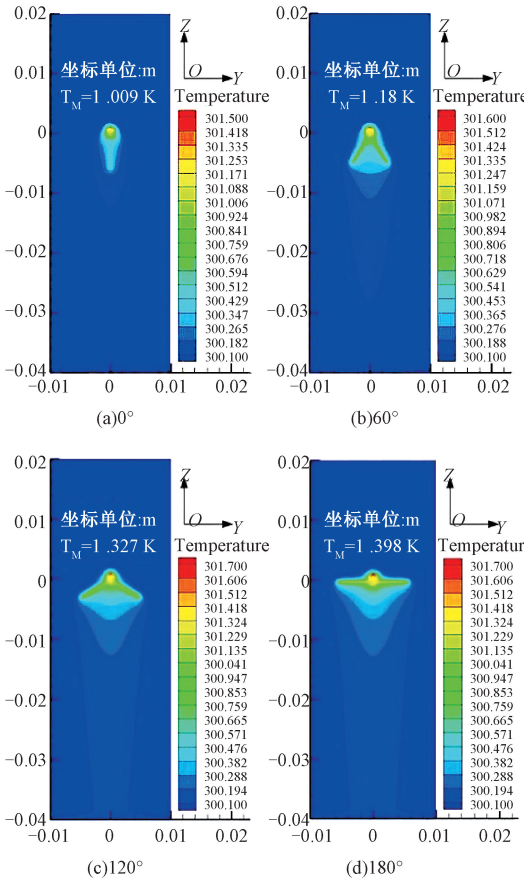


图 5　引线夹角与温度场分布

Fig. 5　Distribution of angle diameter and temperature field

图 5 中，不同的颜色代表了不同的温度值，对比图 5(a)~(d)可以发现，引线夹角越大，太阳辐射造成的影响也越来越大，其尾部带来的温度变化区域也显著增多。当引线夹角为 0°时，对应的太阳辐射误差最小为 1.009 K，当引线夹角为 180°时对应的太阳辐射误差最大为 1.398 K，不同引线夹角引起的误差可达 0.389 K。

1.4　引线直径与辐射误差的关系

在引线设计方面，温度传感器的测量误差不仅与其夹角有关还与所使用的引线直径有关。为探索引线直径对测量误差的影响，设计了引线直径为 0.1、0.2、0.3 mm 的 3 种引线状态，不同引线直径下的温度传感器温度场分布如图 6 所示。

对比图 6(a)~(c)可以看出，随着引线直径的增加，传感器周围的温度也越来越高，对周围的影响也越来越大。所以，在进行探空仪引线焊接时，在保证供电电流的情况下，应使用尽可能细的引线，以减少引线带来的测量误差。

2　传感器数据处理与数据集构建

2.1　数据来源

本文实验数据依托于国家重点研发计划——高精度高空多参数监测传感器研发及应用，在实际环境中开展了探空仪系统对比实验。在某次放飞过程中，探空仪飞行路线如图 7 所示。探空仪从湖南省长沙市(28.109 23，112.786 40)飞行到湖南省娄底市(27.997 8，111.832 5)，飞行距离 100 km 左右，飞行时间约 1.5 h。探空仪数据传输系统每秒向地面接收站发送一次传感器采集数

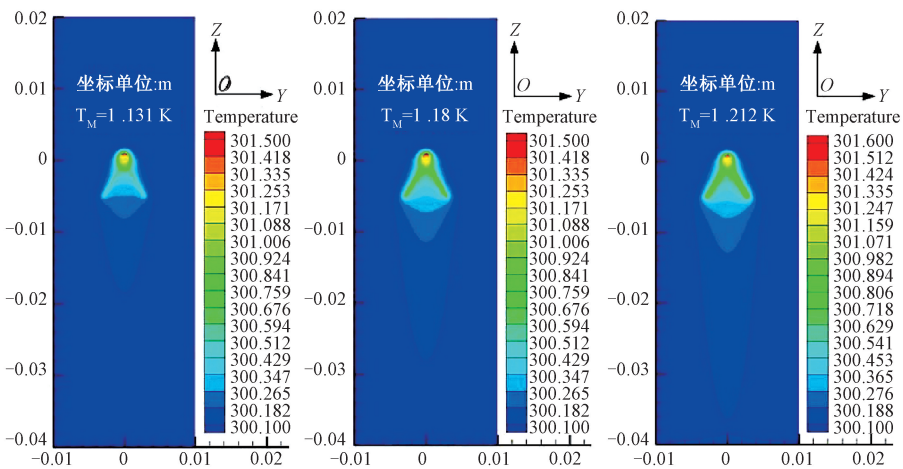


图 6　引线直径与温度场分布图

Fig. 6　Distribution of wire diameter and temperature field

图 7 探空仪飞行轨迹图

Fig. 7 Radiosonde flight trajectory diagram

据,在一次飞行过程中可以收集到 6 000 左右条实验数据。

探空仪回传的数据如表 1 所示,共有时间、气压、温度、湿度等 15 个与大气观测相关的特征数据,并附有对应条件下气压、温度、湿度的标准值。

表 1 探空仪数据表

Table 1 Radiosonde data

参数	示例	参数	示例
时间	18:58:05	经度	112.787 9°
测量气压	992.59 hPa	纬度	28.109 5°
测量温度	30.94℃	方位角	76.4°
测量湿度	68%	仰角	1.6°
露点	24.34℃	标准时	39 483
海拔	119.3 m	辐射	0.02 Wh
风速	5.35 m/s	标准气压	992.5 hPa
风向	128.1°	标准温度	31.2℃
距离	0.148 m	标准湿度	65%

与历史实验数据汇总后发现,总计有 150 次放飞记录,每次放飞平均可以采集到 6 000 条 15 个维度的气象数据。在此基础上,本文构建了一个基于真实观测数据的探空传感器数据集,该数据集共有 900 000 条气象观测数据,其中每一条都有 15 个与气象探测有关的特征变量。本文后续研究内容将在此数据集中展开,通过对采集的各个气象特征数据进行分析与建模,来实现温度传感器测量误差的预测,从而提升探空温度传感器的测量精度,以便为后续的天气预测、气候诊断提供准确的数据支撑。

2.2 气象探测特征构建

为实现探空温度传感器的误差预测,特征工程是关键且不可缺少的一步。探空仪所测量的高空环境数据往往不够全面,不能全方位地反映与温度传感器测量误差相关的全部因素[15]。为实现高精度地误差预测,需要利用数据挖掘方法对原始数据进行分析,以构造出与温度传感器测量误差相关的环境因素。特征工程的作用就是在原始数据中提取和构造出与传感器测量误差相关的变量,这些变量对模型最终的预测效果有着至关重要的作用。

本文在原始数据的基础上,通过气象学、地理学相关知识,构建出太阳高度角、湿球效应、升空速度等特征来进一步刻画温度传感器的测量误差。

1)太阳高度角

太阳高度角是影响传感器测量结果的原因之一,但探空仪并不能实现对太阳高度角的测量,因此需要通过特征工程的方法来对太阳高度角进行计算。太阳高度角与所处的时间、经度、纬度有关,其计算公式如式(1)所示[16]。

$$\sin\varphi = \sin\alpha \times \sin\delta + \cos\alpha \times \cos\delta \times \cos t \tag{1}$$

式中:φ 为太阳高度角。

经过计算后,在 11:52 时进行气象探测时,太阳高度角变化如图 8 所示。

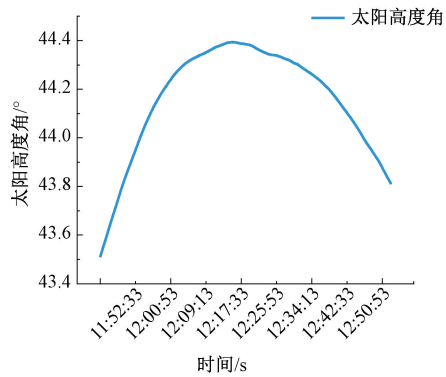


图 8 太阳高度角变化

Fig. 8 Variation of solar altitude angle

由图 8 可以看出,在 12:00 左右,太阳高度角先增加再减少,变化趋势与实际情况相一致,从而验证了计算的有效性。

2)湿球效应

由氢气球携带的探空仪从地面升至 40 km 的高空进行气象要素测量,由于云层处在 3~6 km,在上升过程中会不断地穿入、穿出各级云层[17]。云是漂浮在大气中小水滴的聚合体,在云层穿越过程中,会使温度传感器表面粘有水滴。水滴在蒸发时会吸收温度传感器表面的热量,从而导致温度传感器测量值偏低,此种现象在气象学中称为湿球效应,所带来的误差叫做水污染误差。

在某次探测过程中,海拔高度与温度传感器测量误差、湿度传感器测量值关系如图 9 所示。图 9 中下坐标轴为温度传感器测量误差,上坐标轴为湿度传感器测量值,纵轴为海拔高度。

从图 9 可以看出,在海拔 2 000 m 处,湿度传感器先变高后降低,说明探空仪穿越了云层,而此之后,温度传感器的测量误差也发生了巨大地变化,从而验证湿球效应对温度传感器的测量误差有着显著的影响。

本文试图通过构建穿云特征来解决湿球效应带来的传感器测量误差,当探空仪未穿云时,所测量的湿度值在短时间内不会有大幅度的变化;当探空仪穿云时,所测量

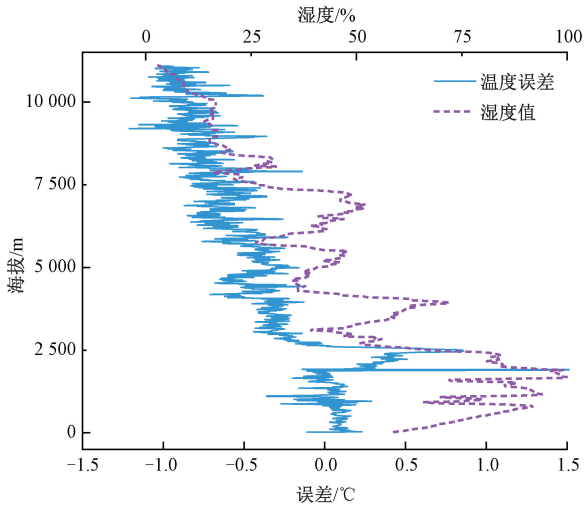


图 9　湿球效应

Fig. 9　Wet bulb effect

的湿度值在短时间内会发生很大的改变。通过对数据的分析与汇总可以发现,在无云情况下,5 s 内湿度值变化幅度小于 6%。因此可以通过判断湿度值的变换幅度来对是否穿云进行判断,从而对温度传感器测量误差的预测提供决策依据[18]。

3)升空速度

升空速度是与温度传感器测量误差相关的另一因素,探空仪升空速度计算公式如式(2)所示。

$$v = \frac{\Delta s}{\Delta t} \tag{2}$$

式中:Δt 为时间间隔;Δs 为对应的距离。为了反映探空仪飞行过程中的实际情况,Δt 应取的尽可能小,以接近瞬时速度[19]。

经计算后,探空仪升空速度与海拔高度如图 10 所示。从图 10 可以看出,在气象探测过程中,由于各种环境因素的干扰,升空速度较为分散,且在短时间内变化剧烈,当以此信息作为输入时,可能被误认为是噪声干扰,从而不能充分地利用此特征信息。

鉴于图 10 升空速度较为分散的情况,可以对其进行数据分桶操作,将一段范围内的升空速度都统一为一个固定值,使模型进一步获取相关信息。根据数据分布情况,本文将升空速度以 2 m/s 为一个单位进行数据分桶,分桶后结果如图 11 所示。

从图 11 可以看出,数据分桶操作相当于对数据进行了离散化操作,连续变化的升速值变为 2、4、6 等几个离散数值,其数值结果不仅更加方便存储,而且增加了数据的鲁棒性。误差预测模型也可以更好地捕捉特征的变化趋势,有助于传感器测量误差的提升。

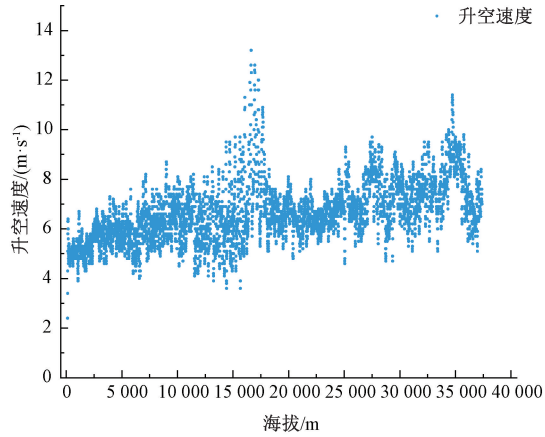


图 10　升空速度

Fig. 10　lift-off velocity

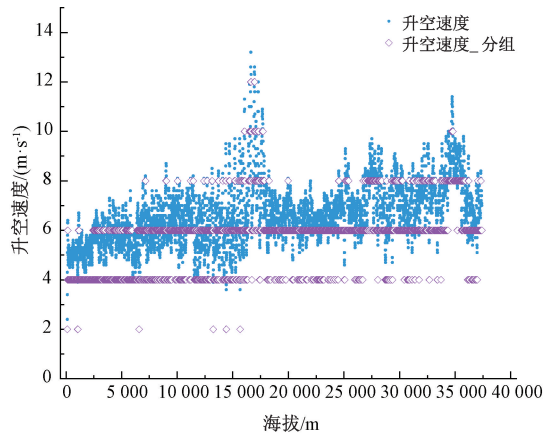


图 11　升空速度分组

Fig. 11　Blast off velocity grouping

3　误差预测模型构建与实验结果分析

3.1　模型融合误差预测方法

为进一步提升误差的预测精度,可以将多个单一模型进行整合,采用集成学习的方法来获得更加准确的预测结果。在集成学习中,模型融合是一种被广泛应用且行之有效的方法。由于单一预测模型实现的原理不尽相同,其在数据空间分布上的学习也略有差异,当模型之间差异越大时,通过模型融合方法对单一预测模型进行组合就有可能会产生更优的预测结果。

本文设计的传感器误差预测模型如图 12 所示,在模型设计时使用了基于不同原理的预测模型,有基于特征加权的线性回归模型,基于超平面的回归向量机(SVR)模型、基于神经元的深度神经网络模型以及基于决策树的 XGBoost 模型[20]。不同原理的预测模型会学习不同的数据特征以及数据之间的交叉关系,但也会忽略部分

数据特征。通过模型融合的方法对多个单一模型的预测值进行整合,可以把各个模型提取较好的特征部分给抓取出来,同时舍弃它们各自表现不好的部分,以有效地优化预测结果[21]。

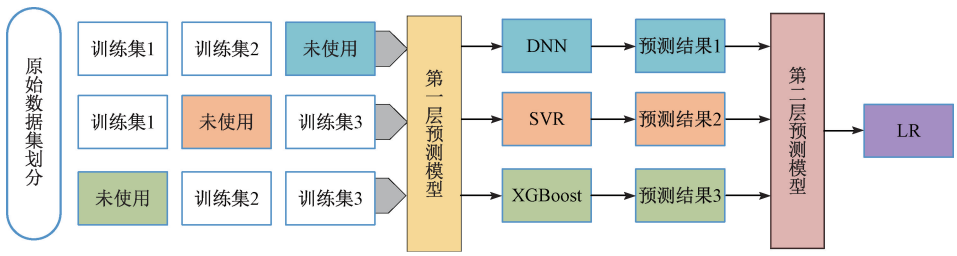


图 12 融合模型设计

Fig. 12 Fusion model design

模型训练时,首先采用 3 折交叉验证将数据集中 900 000 条训练数据划分为训练集 1、训练集 2 和训练集 3,将其中的两部分作为训练集,另外一部分作为测试集。通过对训练数据进行无重复的三次划分,分别对其中每一部分进行一次预测,而预测的模型就是由其余两部分训练得到的,最终预测的结果还是完整的数据集数据,并将此部分数据送入到第二次预测模型中去进行预测。

3.2 误差预测模型参数调优

1)LR 参数调优

为使所构造的线性回归模型在训练集和测试集上都能实现相似地误差预测能力,防止过拟合的发生。在模型训练时采用了 5 折交叉验证的方法,其原理如图 13 所示[22]。

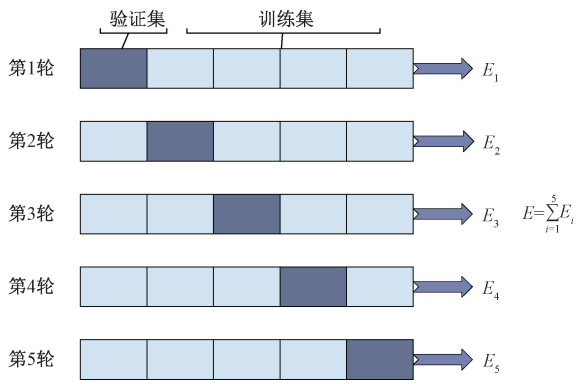


图 13 五折交叉验证示意图

Fig. 13 Five fold cross validation schematic diagram

五折交叉验证即将训练集中 630 000 条数据平均分成 5 份,选取其中的 4 份做训练集,对模型参数进行调整,选取其中的 1 份做测试集,对构建出的模型进行检验,以此来重复进行 5 次。在这 5 次划分中,每次具有不同的训练集和验证集,实现了训练数据的充分利用。在线性回归模型训练过程中,每折验证数据所对应的均方根误差如表 2 所示。

表 2 LR 模型交叉验证得分

Table 2 Table of LR model cross validation score

	cv1	cv2	cv3	cv4	cv5
MSE/hPa	0.41	0.35	0.47	0.49	0.52

从表 2 可以看出,在 5 折交叉验证中,每折对应的均方误差(MSE)极为接近,平均均方误差为 0.448,说明线性回归模型对数据进行了准确地预测,并没有造成预测结果偏差较大的过拟合现象。

在构建上述线性回归模型时,损失函数采用的是均方根误差函数,为了进一步提高模型的泛化能力,在均方根误差的基础上引入了正则化项。常用的有 L1 和 L2 正则化,L1 正则化将权值向量 $\boldsymbol{w}$ 中各个元素的绝对值相加,L2 正则化将权值向量中各个元素进行平方相加,再求出对应的平方根。

将两种正则化方法加入到线性回归模型中,当训练完成后,通过测试集数据对训练结果进行检验,最终结果如表 3 所示。进一步,为了对所构造的线性回归模型进行可视化展示,各特征变量权重系数对比如图 4 所示。

表 3 正则化得分对比

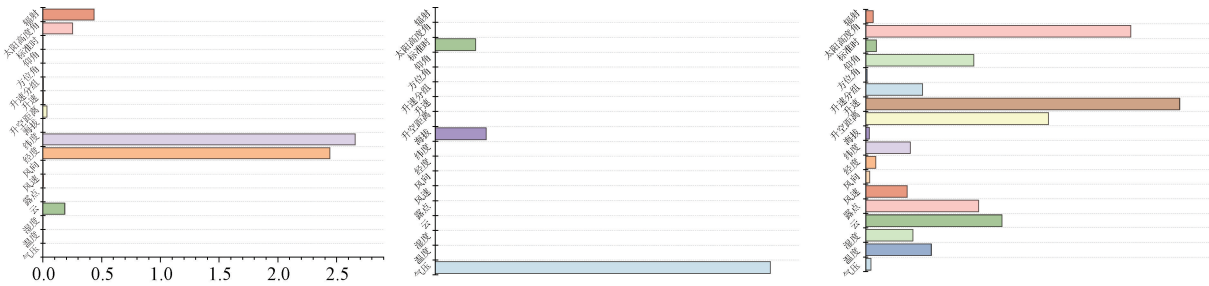
Table 3 Comparison table of regularization scores

	LR	LR+L1	LR+L2
MSE/hPa	0.448	0.521	0.973

从表 3 可以看出,在均方误差这一指标上,预测精度 LR> LR+L1> LR+L2,从图 14 各模型特征权重对比可以发现,LR+L1 模型对数据集中各个特征数据进行了筛选,将气压值赋予很高的权重。相比之下,LR+L2 模型充分使用了各个特征的数据,使最终的预测结果不是由某一两个特征来决定的,增强了模型的泛化能力和抗干扰能力。因此,在 LR 模型构建上,选择了 LR+L2 模型用作温度传感器误差值预测。

2)SVR 参数调优

当使用 SVR 模型进行温度传感器误差预测时,需调



(a) LR　　(b) LR+L1　　(c) LR+L2

图 14　LR 正则化模型权值对比

Fig. 14　Comparison diagram of LR regularization model weight

整的超参数有 Kernel、Gamma 和 C 三个变量,各超参数含义及备选项如表 4 所示[23]。

表 4　SVR 模型超参数表

Table 4　Hyper parameter table of SVR model

超参数	备选项
Kernel	核函数,可选 Linear、Poly、RBF 等
Gamma	核函数系数
C	惩罚系数

为了找出 SVR 模型的最优参数值,采用网格搜索法对上述 3 个参数值进行求解。所谓网格搜索法即通过循环遍历,对每一种参数组合进行尝试,以获得最高地误差预测精度。在进行网格搜索时,核函数待搜索项为 Linear、Poly 以及 RBF,核函数系数 Gamma 范围为 0~10,惩罚系数 C 范围为 0~100。

经网格搜索后,模型的最优参数如表 5 所示。在此参数下,训练集 $MSE=0.315$,测试集 $MSE=0.323$。训练集与训练集接近,所训练的模型不仅保证了传感器误差预测的精度而且预防了过拟合现象。

表 5　SVR 模型最优参数表

Table 5　Table of optimal parameters of SVR model

Kernel	Gamma	C
RBF	0.1	30

3) DNN 参数调优

根据 DNN 网络基本结构,选用 5 层网络进行训练,一个输入层、3 个隐藏层、一个输出层[24]。由数据集中的数据可知,所用到的特征有时间、经度、纬度、海拔高度等 19 个与传感器测量误差相关的特征,所以输入层有 19 个神经元。为了能实现较好的误差修正能力,隐含层设计为 20—30—10。经过深度神经网络预测后,使其输出对应的预测误差,所以输出层有 1 个神经元,即网络结构为 19—20—30—10—1。

对于探空温度传感器误差预测,由于其有较高的预测精度需求,传统的 Sigmoid 函数无法高精度地进行预测。因此,本文将 Morlet 小波作为网络隐藏层的激活函数,以构造出一种增强型深度神经网络误差预测模型,以实现更好地获取特征值之间的作用关系,从而进一步提升温度传感器测量误差的预测精度[25]。

为验证不同激活函数对 DNN 模型的影响,将相同的数据输入到模型中,经过训练使网络达到最终收敛,统计其在测试集上的表现,实验结果如表 6 所示。

表 6　激活函数均方误差对比

Table 6　Table of activation function MSE

	修正前	DNN	DNN+Morlet
MSE/hPa	0.879	0.297	0.215

从表 6 可以看出,两种深度神经网络模型都能有效地对温度传感器测量误差进行预测,使 MSE 显著降低。相比之下,采用小波函数作为激活函数比传统 Sigmoid 函数预测精度更高,MSE 进一步降低了 0.082,使温度传感器达到了更高的测量精度。

4) XGBoost 参数调优

XGBoost 是基于树的集成学习算法,通过构建多个决策树来实现模型的整体预测。在 XGBoost 模型构建过程中,每次新增一颗决策树来拟合之前预测的残差值,通过此种方式迭代,使模型最终预测值不断地接近于真实值[26]。

XGBoost 在训练过程中需调整的参数有分类器个数、最大树深、最小节点权重、正则化系数等诸多变量。为了找出上述参数的具体数值,可以通过网格搜索法对不同参数进行测试,并通过交叉验证的方式对不同的参数值进行判断,从而找出适应于传感器误差预测的模型参数值,实现温度传感器测量误差的准确预测。

在网格搜索法中,首先对分类器数目与最大树深进行探索。在分类器数目上,搜索范围为 10~1 000,在最大树深上,搜索范围为 1~5。在网格搜索过程中通过 5 折交叉验证方法来对待搜索参数进行验证。网格搜索结束后,分类器数目、最大树深与 MSE 的热力图如图 15 所示。

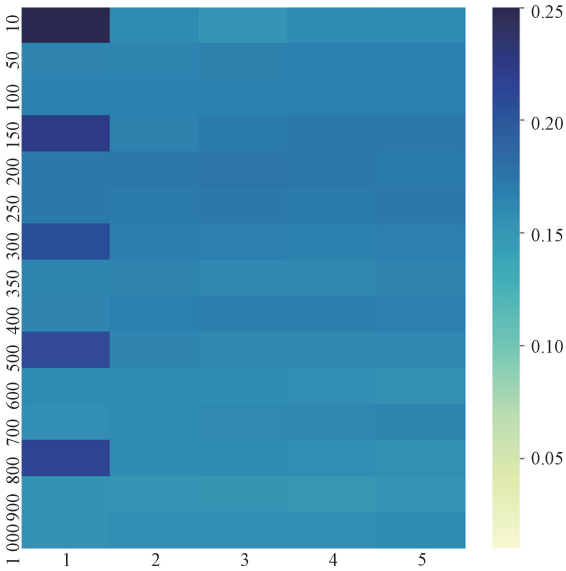


图 15 分类器数据与最大树深热力图

Fig. 15 The heat map of classifier data and maximum tree deep

图 15 中，x 轴为最大树深，y 轴为分类器数目，不同的颜色代表了不同的预测误差。从图 15 可以看出，当分类器数目为固定值时，MSE 随最大树深的增加而减少。通过网格搜索结果可以发现，当分类器数目为 400、最大树深为 5 时，对应的 MSE 最小。

3.3 气象特征贡献度分析

为了比较不同模型对不同特征的选择情况，将各个模型的特征权重进行了归一化处理，特征重要性对比如图 16 所示。

在 LR 模型中，各特征的重要性可用其对应的权重来表示；在 SVR、XGBoost 模型中，可用通过特征变量的使用次数或带来的总增益来进行排序；而在 DNN 模型中，并没有分裂次数、信息增益等衡量指标，可以在训练前对每个特征进行随机数填充，训练完成后观测 MSE 变化，将 MSE 的变化率作为特征重要性因素；在融合模型中，由于采用了多个模型级联预测，可以将 SVR、XGBoost、DNN 模型的权重按 LR 模型权重系数进行加权相加，即可当做融合模型的特征权重。

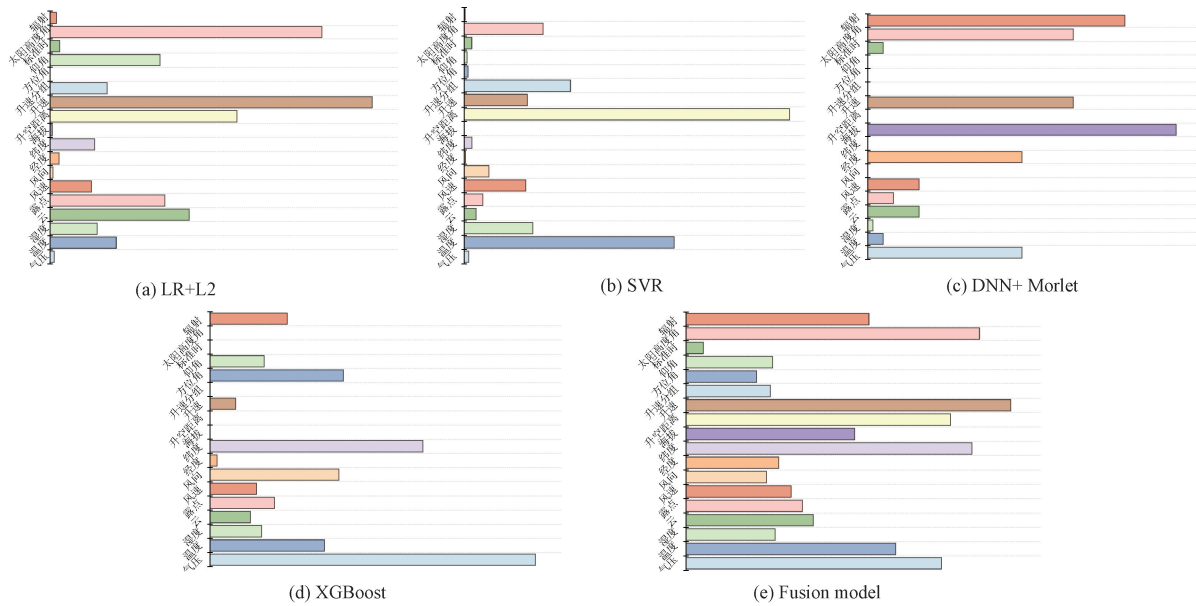


图 16 模型特征重要性对比

Fig. 16 Comparison diagram of the importance of model features

从图 16 可以看出，在 LR 模型中，最重要的前 3 个特征为升空速度、太阳高度角、升空距离；在 SVR 模型中，最重要的前 3 个特征为升空距离、温度、升速分组；在 DNN 模型中，最重要的前 3 个特征为海拔、太阳辐射强度、太阳高度角；在 XGBoost 模型中，最重要的前 3 个特征为气压、纬度、方位角；在融合模型中，最重要的前 3 个特征为升速、太阳高度角、纬度。对比各个模型的重要性分布可以得出，基于不同预测原理的模型对数据特征有不同的侧重，而融合模型方法可以充分地综合各特征数据间的影响，从而实现更加准确地预测。

本文所使用的数据集共有 19 个特征变量，其中通过特征工程的方法构造出了太阳高度角、升空速度、升空速度分组、穿云 4 个特征变量。从图 16 所示的特征重要性分布可得，在所提出的融合模型中，升速、太阳高度角分别占据了第 1、第 2 重要性特征。通过特征工程所构造的 4 个特征，排序最低的为穿云。与其他影响因素相比，穿

云在整个气象探测过程中占据了很小一部分的比例,只有在特定的海拔高度范围下才会出现,从而在整体的特征重要性排序上处于靠后位置。

3.4　各误差预测模型结果对比

1)评价指标对比

为了对比不同模型在温度传感器误差预测上的效果,从平均误差、标准差、均方误差、R^2 四个方面进行评价,其结果如表 7 所示。从表 7 可以看出,在衡量预测误差均值的平均误差上,LR 模型获得了最佳的效果,说明 LR 模型可以找到一条直线,使分布在两侧的误差值相互抵消;在衡量预测误差离散程度的标准差、衡量误差整体情况的均方根误差上,XGBoost 模型获得了最好的效果,说明 XGBoost 可以通过多次树结构划分,使模型预测值充分接近于真实值。

通过对比各个弱学习器与本文提出的融合模型在各个评价指标上的数值可以发现,通过模型融合方法,使平均误差、标准偏差、MSE 在预测数据中均得到了降低,同时 R^2 也得到了提升,说明了融合模型对传感器误差的预测精度更高,可以获得更加准确地气象探测数据,从而有助于后续的气候诊断、天气预测。

表 7　预测误差对比

Table 7　Comparison table of prediction errors

	平均误差/hPa	标准差/hPa	MSE/hPa	R^2
预测前	0.817	0.458	0.878	
LR	0.074	0.364	0.573	0.80
SVR	0.126	0.351	0.315	0.81
DNN	0.104	0.336	0.215	0.83
XGBoost	0.115	0.229	0.178	0.88
融合模型	0.008	0.204	0.068	0.93

2)图形可视化对比

为了进一步揭示不同模型的预测效果,选取了某次气象观测过程中的数据,传感器预测误差分布对比如图 17 所示。图 17 中 x 轴为传感器的测量误差,y 轴为对应的海拔高度,紫色为传感器测量误差值,蓝色为模型对应的预测值。

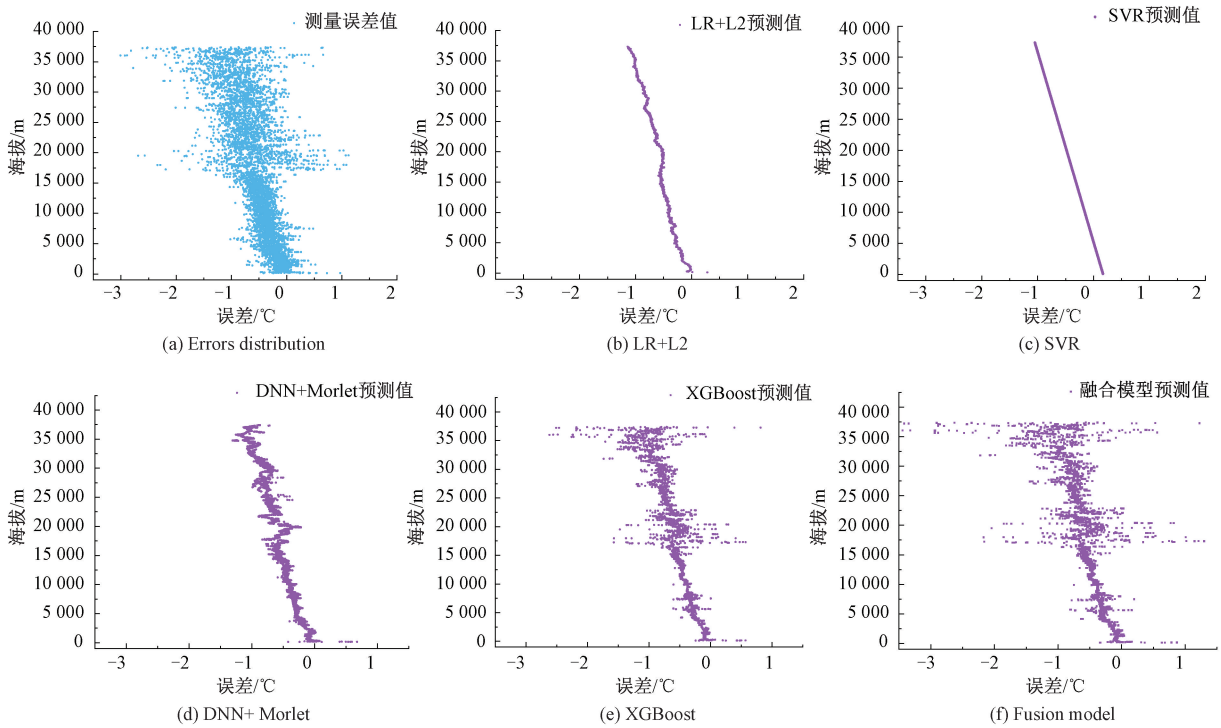


图 17　各模型预测效果对比

Fig. 17　Comparison diagram of prediction effect of each model

从图 17 的预测值与真实值的分布情况可以看出,基于不同原理的模型所带来的预测值分布较为不同。在基于超平面的 SVR 模型中,其预测分布近似为一条直线;在基于多层感知机的 DNN 模型中,其预测值可以刻画出误差分布的大体趋势;在基于决策树原理的 XGBoost 模型中,其预测值分布较为分散,可以充分地接近真实值所处的位置。从各图 17 对比中可以看出,本文所提出的融合模型在预测值与真实值的分布上最为接近,可以在结合各个模型预测的基础上实现更为精确的传感器误差预测[27]。

4 结 论

各海拔下的温度数据能够反映大气热力过程和运动轨迹,是进行气候预测、气候变化估计和自然灾害判断等方面的基础。如何准确高效地实现温度传感器的误差预测是气象探测领域研究的重点和难点。本文针对温度传感器误差预测模型展开研究,摒弃了传统的计算机仿真的方法,以国际探空仪系统对比大赛为基准,构造出了一个包含 900 000 条数据的高空气象探测数据集。在每一条数据中,共有温度、湿度、太阳高度角、升空速度等 19 个与气象观测有关的测量数据,解决了仿真环境与实际环境存在偏差的问题,更有利于实际工程应用。鉴于传感器误差预测问题中深度神经网络预测精度低的问题,提出了一种适应于传感器测量误差的增强型深度神经网络预测模型。在该模型中,将 Morlet 小波作为隐藏层的激活函数,从而使模型的非线性映射能力显著增加。经过训练对比,可以发现在各个评价指标上所提出的增强型深度神经网络模型均优于传统模型,并在预测结果可视化中可以明显观察到两者的预测差别。基于残差思想设计并实现了融合误差预测模型,以最大限度地提升传感器测量误差的预测精度。在支持向量机、深度神经网络、XGBoost 预测的基础上,通过线性回归模型对各个预测器的结果进行加权,将特征较好的部分给抓取出来,同时舍弃它们各自表现不好的部分。由实验结果可以看出,所提出的融合模型在各个评价指标上均优于单一预测模型。经模型预测后,MSE 从 0.878 降低到了 0.068,标准差从 0.458 降低到了 0.204,不仅使传感器测量误差得到了显著降低而且误差分布也进一步集中。

本文针对气象探测过程中温度传感器测量误差问题展开了研究,在目前研究成果的基础上,未来工作可以从以下 3 点展开。

1) 所构造的数据集仅为部分气象观测站、部分观测时间下的测量数据。对于不同区域的传感器测量数据,由于高空环境的不同,造成的特征权重可能会略有差异,可以进一步构建一个全国各个气象站,以实现各个观测时间下的探空数据,完成对数据集的补充。

2) 所提出的融合模型仅实现了温度传感器的误差修正,在气象探测过程中,湿度传感器、气压传感器的测量值也存在一定的误差,可以进一步对模型优化,引入预测能力更强的深度学习模型来实现温度、湿度、气压传感器的综合同步修正。

3) 在系统设计方面,可以在现有传感器误差预测系统上进行升级,设计一个基于 B/S 架构的应用平台。各个气象观测站均在此平台上接收存储数据,从而构造出一个气象大数据平台,来进一步指导相关学科的开展。

参考文献

[1] World meteorological organization. 2020 state of climate services[R]. WMO, 2020.

[2] 杨杰, 刘清惓, 戴伟, 等. 用于气象观测的阵列式温度传感器流体动力学分析与实验研究[J]. 物理学报, 2016, 65(9): 128-139.
YANG J, LIU Q J, DAI W, et al. Fluid dynamic analysis and experimental study of a temperature sensor array used in meteorological observation [J]. Acta Physica Sinica, 2016, 65(9): 128-139.

[3] 周方明, 王刚, 汪林, 等. 寒冷地区高速公路冰雪检测与预警[J]. 交通运输系统工程与信息, 2013, 13(3): 178-182.
ZHOU F M, WANG G, WANG L, et al. Freeway traffic ice-detecting and warning method in cold region [J]. Journal of Transportation Systems Engineering and Information Technology, 2013, 13(3): 178-182.

[4] 戴伟. 探空温度传感器的数值分析与优化设计[D]. 南京:南京信息工程大学, 2012.
DAI W. Numerical analysis and optimization design of the sounding temperature sensor[D]. Nanjing: Nanjing University of Information Science & Technology, 2021.

[5] 吴翼凡, 刘清惓, 杨杰, 等. 基于 BP 神经网络的温度传感器辐射误差修正[J]. 现代电子技术, 2018, 41(24): 43-47.
WU Y F, LIU Q J, YANG J, et al. Numerical analysis and optimization design of the sounding temperature sensor [J]. Modern Electronics Technique, 2018, 41(24): 43-47.

[6] 侯天浩, 行鸿彦, 刘洋. 多旋翼无人机在气象探测中的现状与展望[J]. 电子测量与仪器学报, 2019, 33(10): 7-15.
HOU T H, XING H Y, LIU Y. Current status and prospects of multi-rotor UAVs use in meteorological detection [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(10): 7-15.

[7] MARTIN A, WEISSMANN M, REITEBUCH O, et al. Validation of Aeolus winds using radiosonde observations and numerical weather prediction model equivalents[J]. Atmospheric Measurement Techniques, 2021, 14(3): 2167-2183.

[8] SCHMIDLIN F J, LUERS J K, HUFFMAN P D. Preliminary estimates of radiosonde thermistor errors[J]. National Aeronautics and Space Administration, 1986, 15(3): 123-136.

[9] 杨加春, 钱媛, 徐洁, 等. 华云高精度导航探空仪在甘肃庆阳干旱气候夏季的温湿数据对比分析[J]. 气象科技进展, 2018, 8(6): 71-76.
YANG J CH, QIAN Y, XU J, et al. Analysis of temperature and humidity data of Huayun high accuracy navigation radiosonde in arid climate summer in qingyang gansu province[J]. Advances in Meteorological Science and Technology, 2018, 8(6): 71-76.

[10] RUFFIEUX D, JOSS J. Influence of radiation on the temperature sensor mounted on the swiss radiosonde[J]. Journal of Atmospheric and Oceanic Technology, 2003, 20(11): 1576-1582.

[11] 唐志共, 许晓斌, 杨彦广, 等. 高超声速风洞气动力试验技术进展[J]. 航空学报, 2015, 36(1): 86-97.
TANG ZH G, XU X B, YANG Y G, et al. Research progress on hypersonic wind tunnel aerodynamic testing techniques[J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(1): 86-97.

[12] 刘清惓, 戴伟, 杨荣康, 等. 探空温度测量太阳辐射误差的流体动力学分析[J]. 高原气象, 2013, 32(4): 1157-1164.
LIU Q J, DAI W, YANG R K, et al. Fluid dynamics analysis on solar radiation error of radiosonde temperature measurement[J]. Plateau Meteorology, 2013, 32(4): 1157-1164.

[13] 戴伟, 刘清惓, 杨杰, 等. 探空温度传感器的计算流体动力学分析与实验研究[J]. 物理学报, 2016, 65(11): 163-173.
DAI W, LIU Q J, YANG J, et al. Computational fluid dynamics analysis and experimental study of sounding temperature sensor[J]. Acta Physica Sinica, 2016, 65(11): 163-173.

[14] 杜太行, 崔景瑞, 孙曙光, 等. 基于混合式仿真模型的交流接触器动态特性研究[J]. 电子测量与仪器学报, 2019, 33(10): 187-194.
DU T X, CUI J R, SUN SH G, et al. Study on dynamic characteristics of AC contactor based on hybrid simulation model[J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(10): 187-194.

[15] 郭亮, 董勋, 高宏力, 等. 无标签数据下基于特征知识迁移的机械设备智能故障诊断[J]. 仪器仪表学报, 2019, 40(8): 58-64.
GUO L, DONG X, GAO H L, et al. Feature knowledge transfer based intelligent fault diagnosis method of machines with unlabeled data[J]. Chinese Journal of Scientific Instrument, 2019, 40(8): 58-64.

[16] 李文, 赵永超. 地球椭球模型中太阳位置计算的改进[J]. 中国科学院大学学报, 2019, 36(3): 363-375.
LI W, ZHAO Y CH. The improvement in solar position calculations in the ellipsoid model of the earth[J]. Journal of University of Chinese Academy of Sciences, 2019, 36(3): 363-375.

[17] 王彦明, 贾克斌, 刘鹏宇, 等. 基于增强型 BP 网络的气象传感器标校方法[J]. 中国测试, 2020, 46(12): 105-111.
WANG Y M, JIA K B, LIU P Y, et al. Calibration method of meteorological sensor based on enhanced BP network[J]. China Measurement & Test, 2020, 46(12): 105-111.

[18] 高振斌, 潘星, 田丰, 等. 基于热敏电阻的热式液体流量传感器[J]. 仪表技术与传感器, 2014(9): 12-15.
GAO ZH B, PAN X, TIAN F, et al. Thermal liquid flow sensor based on thermistor[J]. Instrument Technique and Sensor, 2014(9): 12-15.

[19] 梅玉航. 基于光电经纬仪测角信息直接计算速度的方法[J]. 电子测量技术, 2020, 43(16): 128-131.
MEI Y H. A method of direct calculating velocity based on angle measuring information of film theologize[J]. Electronic Measurement Technology, 2020, 43(16): 128-131.

[20] 史佳琪, 张建华. 基于多模型融合 Stacking 集成学习方式的负荷预测方法[J]. 中国电机工程学报, 2019, 39(14): 4032-4042.
SHI J Q, ZHANG J H. Load forecasting based on multi-model by stacking ensemble learning[J]. Proceedings of the CSEE, 2019, 39(14): 4032-4042.

[21] YANG J, WANG Y, LIU P, et al. Error prediction method for temperature sensor based on multi model fusion[C]. 2020 IEEE International Conference on Artificial Intelligence and Big Data, 2020: 183-187.

[22] CHEN J. Data-driven sensor deployment for spatiotemporal field reconstruction[J]. Instrumentation, 2019, 6(3): 28-38.

[23] 陈家瑞, 陈忠孝, 秦刚, 等. 基于 PSO 算法与 SVR 算法在企业直流配电网短期负荷预测的研究[J]. 国外电子测量技术, 2020, 39(12): 70-73.
CHEN J R, CHEN ZH X, QIN G, et al. Research on short-term load prediction of enterprise DC distribution network based on PSO algorithm and SVR algorithm[J]. Foreign Electronic Measurement Technology, 2020, 39(12): 70-73.

[24] WANG Y, JIA K, LIU P, et al. Nonlinear correction based on depth neural network [C]. 2020 IEEE International Conference on Artificial Intelligence and Security, 2020: 289-301.

[25] 贾克斌, 王彦明, 杨加春, 等. 基于神经网络的气压传感器非线性校正[J]. 北京工业大学学报, 2021, 47(1): 40-49.
JIA K B, WANG Y M, YANG J CH, et al. Nonlinear correction of pressure sensors based on neural network[J]. Journal of Beijing University of Technology, 2021, 47(1): 40-49.

[26] 王燕, 郭元凯. 改进的 XGBoost 模型在股票预测中的应用[J]. 计算机工程与应用, 2019, 55(20): 202-207.
WANG Y, GUO Y K. Application of improved XGBoost model in stock forecasting[J]. Computer Engineering and Applications, 2019, 55(20): 202-207.

[27] 谭光韬, 张文文, 王磊. 气体传感器阵列混合气体检测算法研究[J]. 电子测量与仪器学报, 2020, 34(7): 95-102.
TAN G T, ZHANG W W, WANG L. Research on mixed gas detection algorithm of gas sensor array[J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(7): 95-102.

作者简介

杨加春,2004 年于北京科技大学获得学士学位,2007 年于北京科技大学获得硕士学位,现为天津华云天仪特种气象探测技术有限公司高级工程师,主要研究方向为高空气象仪器的应用开发、高空探测数据处理。
E-mail:yangjiachun@ cnhyc. com

Yang Jiachun received his B. Sc. degree from University of Science and Technology Beijing in 2004, M. Sc. degree from University of Science and Technology Beijing in 2007. Now he is a senior engineer at Tianjin Huayuntianyi Special Meteorological Sounding Co., Ltd. His main research interests include development of sounding instruments and data processing of sounding.

王彦明(通信作者),2018 年于河北大学获得学士学位,2021 年于北京工业大学获得硕士学位,主要研究方向为大数据处理。
E-mail:wangyanming@ emails. bjut. edu. cn

Wang Yanming (Corresponding author) received his B. Sc. degree from Hebei University in 2018, M. Sc. degree from Beijing University of Technology in 2021. His main research interests include big data and data mining.