

基于特征工程的强降水物理要素提取及分析

钟琦^{1*}, 方祖亮¹, 孙卓², 王秀明¹, 张智察³

¹中国气象局气象干部培训学院, 北京

²河北省气象台, 河北 石家庄

³浙江省气象台, 浙江 杭州

收稿日期: 2021年12月22日; 录用日期: 2022年1月19日; 发布日期: 2022年1月26日

摘要

灾害性强降水事件对人民生命和财产安全具有重大影响, 其精准预报是防灾减灾的重要内容, 但因其突发性强、变化快和强度大的特点, 常用数值模式的预报准确率很低。近年来随着人工智能技术快速发展, 融合物理特征和机器学习的强降水预报成为热点研究。本文利用特征工程提取强降水预报关键物理因子, 发现随着预报精细度提升, 模式直接预报降水的贡献越来越小, 而相关物理特征更具参考价值。对6小时以上累计降水, 中高层的动力维持作用、整层抬升和整层含水量具有较大贡献; 而对于3小时以内降水预报, 近地面层的动力辐合、不稳定能量特征、及影响降水效率的云微物理因子重要性更加凸显, 这与降水系统的多尺度特征和模式可预报性相对应。通过特征工程精简后的物理特征可有效去除冗余提升机器学习模型效率, 为提升强降水的预报准确率提供支撑。

关键词

机器学习, 特征提取, 强降水, 物理要素, 华北地区

Using Feature Engineering to Extract Important Physical Parameters for Heavy Rain

Qi Zhong^{1*}, Zuliang Fang¹, Zhuo Sun², Xiuming Wang¹, Zhicha Zhang³

¹China Meteorological Administration Training Center, Beijing

²Hebei Meteorological Observatory, Shijiazhuang Hebei

³Zhejiang Meteorological Observatory, Hangzhou Zhejiang

Received: Dec. 22nd, 2021; accepted: Jan. 19th, 2022; published: Jan. 26th, 2022

*通讯作者。

文章引用: 钟琦, 方祖亮, 孙卓, 王秀明, 张智察. 基于特征工程的强降水物理要素提取及分析[J]. 计算机科学与应用, 2022, 12(1): 147-157. DOI: 10.12677/csa.2022.121016

Abstract

Hazardous heavy rain events have significant impacts on people's lives and property safety. It's important to make accurate prediction to reduce the loss. However, due to its high intensity and rapid change, the accuracy of operational numerical models is very low. In recent years, with the rapid development of artificial intelligence technology, more and more studies devote on integrating physical understanding and machine learning to improve heavy rain forecast. In this study, the key physical factors of heavy precipitation prediction are extracted by feature engineering methods. It is found that the finer the forecast is, the contribution of forecasted rain is smaller, but the physical features are more important. For the above 6 h rainfall, the dynamics of the middle and high levels, the uplift of the whole layer and the water content of the whole layer contribute greatly; for forecast less than 3 hours, the importance of dynamic uplift at low level, the unstable energy and some microphysical factors are prominent, which is in according to the multi-scale characteristics of the atmosphere. The physical features simplified by feature engineering can effectively remove redundancy and improve the efficiency of machine learning model. The results would help to improve the prediction of heavy rain events.

Keywords

Machine Learning, Feature Extraction, Heavy Rain, Physical Parameters, North China

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

降水预报是天气预报中的重点和难点, 精准的降水预报对于防灾减灾、工农业生产和公众生活等方面具有重要意义。以华北地区为例, “63.8”海河流域暴雨、“75.8”河南暴雨、“96.8”河北暴雨均是发生在华北地区的著名暴雨过程。随着城市化进程加快, 以京津冀为代表的城市群极端暴雨天气频发, 如2004年7月10日北京突发性局地暴雨以及2011年6月23日、2012年7月21日北京特大暴雨天气过程, 特别是2012年7月21日北京特大暴雨, 降水强度大, 持续时间长, 造成北京市79人死亡, 河北省26人死亡、20人失踪, 受灾人口达到190万, 直接经济损失200多亿元。因此加强降水预报研究, 努力提高暴雨预报水平对减小生命财产损失有重大意义。目前降水的预报主要依赖于数值模式, 其每天在世界各地主要气象业务中心运行, 在灾害性天气精细化预报中发挥着不可替代的作用, 但其同时具有不完整性和不精准性的特点, 尤其在具有复杂下垫面条件时和面向强天气影响时预报能力明显不足。机器学习可以方便地进行空间和时间的概括, 处理大量的观测数据和预报变量, 并能学习和集成物理解到模型中。在气象领域, 已有大量研究使用机器学习方法来提高天气预报的准确性, 如应用单个数值预报模式的多个气象要素来构建天气预报模型[1] [2]。门晓磊等[3]基于中国气象局、欧洲、日本气象厅和NCEP的四个数值的近地面2 m处温度预报和实况, 分别使用岭回归、随机森林和深度学习对2 m处温度进行了订正, 这三种方法的预报效果都优于四个数值天气预报模式以及通常的多模式集合平均。在强对流预警方面, Czernecki等[4]通过随机森林方法也实现了对冰雹灾害的精准估计; Hyangsun等[5]在决策树、随机森林、支持向量机等机器学习技术的基础上, 利用各种气象、卫星数据, 实现了强对流形成过程的检测。降水订正方面, 陈浩等[6]选择K近邻(KNN)、多元自回归样条(MARS)、支持向量机、多

项对数线性模型(MLM)和人工神经网络(ANN)等方法,利用海拔、相对湿度、坡向、植被、风速、气温和坡度等因子订正了 ITPCAS (Institute of Tibetan Plateau Research, Chinese Academy of Sciences)和 CMORPH (Climate prediction center morphing technique)两种常用的青藏高原日降水数据集。张晨阳等[7]将基于地形的加权随机森林(terrain-based weighted random forests, TWRF)方法用于雷达定量降水估计,并在此基础上考虑了反射率垂直廓线的特征重要性变化,有效提升了雷达反演降水效果。黄威等[8]采用支持向量机回归方法,基于 ECMWF 的 51 个集合预报和 NCEP 的 21 个集合预报,建立了多模式集成的强降水预报模型。机器学习应用于专业领域的众多研究中特征提取具有关键作用[9] [10] [11] [12]。本文通过集成机器学习中的特征工程方法,综合挖掘多源观测和数值预报等气象信息,遴选面向强降水短临预报的关键要素和时次,以加深对小时尺度强降水特征的物理理解,为强降水发生发展的机理研究和短临预报机器学习订正提供客观支撑。

2. 方法和数据

2.1. 特征工程方法概述

特征工程包括数据清洗、数据预处理、和特征选择等过程。特征工程的目的在于获得更容易表现数据性质的训练集。一般特征越好,灵活性越强,构建的模型越简单,模型的性能越出色,构建模型的计算代价越低。常用的特征工程可分为三大类,过滤式,包裹式和嵌入式(如图 1 所示)。其中过滤式通过设定阈值的评价准则来判识特征与目标量的关系,这种方式可理解性强,计算效率高,但由于与分类器剥离,其对弱相关特征和冗余度的判断并不理想。过滤式结合了特征和分类器算法,搜索选择预测精度较好的特征子集实现改进,但非常耗时。嵌入式直接在机器学习算法中通过权值和贡献度提取重要特征,常能获得比前两类更好的效果,如树模型和梯度提升模型(如 XGBoost)。本文采用了经典的随机森林模型进行研究,并在 3.3 节与相关系数、互信息等特征工程进行集成分析。

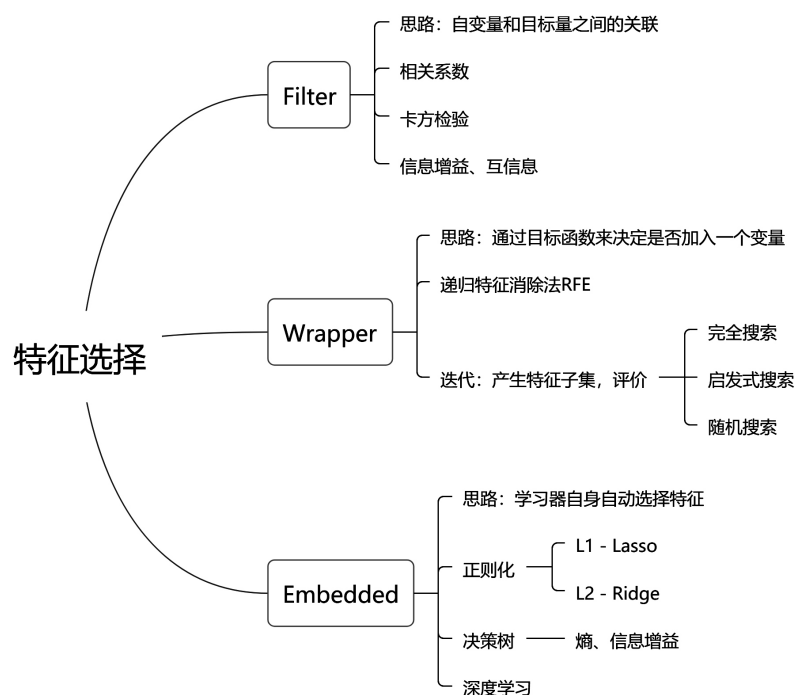


Figure 1. Schematic diagram of feature engineering classification

图 1. 常用特征工程分类示意图

2.2. 随机森林模型及特征重要性度量

随机森林(Random Forest, 以下简称 RF)是一种能够进行回归和分类的机器学习方法,由多个分类或者回归树组成(如图2所示)。RF在以决策树为基学习器构建 Bagging 集成的基础上,还在决策树的训练过程中引入了随机属性选择。具体来说,对基决策树的每个结点,先从该结点的属性集合中随机选择一个包含 k 个属性的子集,然后再从这个子集中选择一个最优属性用于划分。随机森林的基学习器的多样性不仅来自样本扰动,还来自属性扰动,因此使得最终集成的泛化性能可通过个体学习器之间的差异的增加进一步提升。随机森林简单、容易实现、计算开销小,且在很多现实任务中展现出强大的性能,被誉为“代表集成学习技术水平的方法”。

随机森林能够较好地度量每个特征的重要性。通过计算每个特征在随机森林中的每棵树上所做的贡献,取平均,就可以比较出不同特征的贡献大小。通常用基尼指数(Gini index)或者袋外数据(OOB)错误率作为评价指标来衡量。目前 python 中的 sk-learn 库已经实现了用随机森林评估特征重要性,训练好模型后,可直接调用 `feature_importances` 属性得到每个特征的重要性。

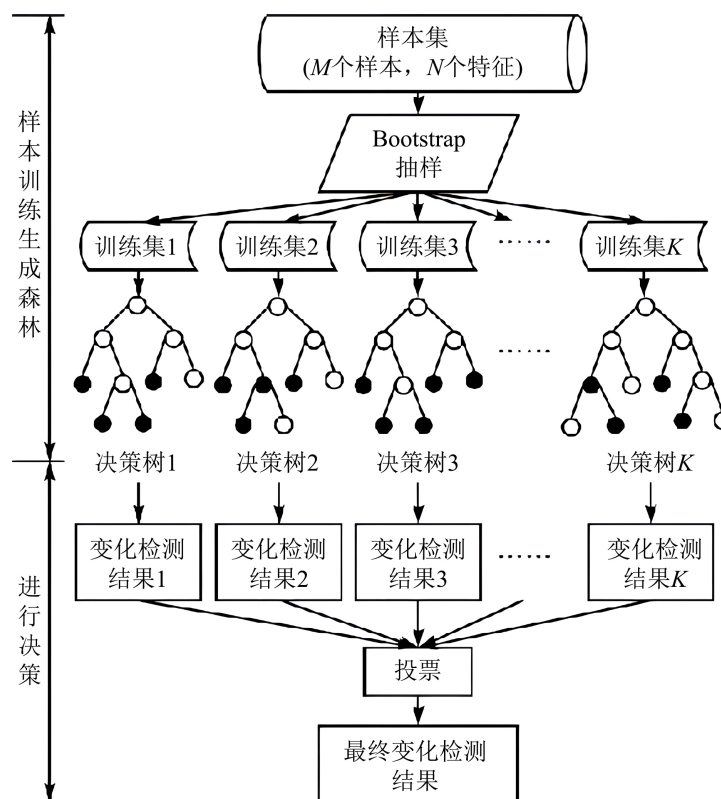


Figure 2. Schematic diagram of Random Forest Model

图2. 随机森林模型示意图

2.3. 数据集及样本平衡

本文使用的多源数据包括加密气象观测资料、高分辨率全球环流模式预报和精细化中尺度数值预报产品资料(以欧洲中期天气预报中心(以下简称 EC)和华东区域模式(以下简称 SMS)预报产品为例)。选取地面观测降水为目标量(因变量),温、压、湿、风等大气要素和综合诊断的动力热力特征量为自变量,进行多源数据的时空匹配融合,构建强降水短临预报物理要素大数据集。根据天气分析预报经验和特征完

备性原则,将众多物理要素归为5类特征,具体为:模式预报的目标量(降水预报, Type 0)、水汽特征(Type 1)、动力抬升特征(Type 2)、热力不稳定特征(Type 3)和维持时间特征(Type 4)。根据天气系统连续和浑沌共存的特性,同时也选取一定历史长度的气象要素作为特征,经前期试验结果选取历史6小时以内的历史观测、模式预报和目标时刻的模式预报作为特征量,预测当前时刻的累计降水。5类特征要素和时次组合共计产生376个特征。

由于强降水属于小样本事件,为获取更多的强降水样本,特选取北方4次典型的强降水事件所在日构建数据集,在个例选取是注意囊括了系统性强降水、局地短时强降水和台风降水等不同降水类型,具体采样覆盖时间为20180804~20180807, 20190802~20190804, 20190810~20190812。图3给出了采样时段内降水分布,降水中心的累积降水超过500 mm,京津冀地区的降水中心量值超过250 mm。基于此样本构建特征X,分别以1小时、3小时和6小时累计降水(以下简称r1、r3和r6)为目标Y,随机选取部分时次共计129万左右的样本进行特征分析,考虑到样本不均衡(即无降水和低量级降水样本数量远远多于大量级降水),因此需要对其进行适当的下采样,得到如表1所示的样本分布。将采样后的样本放到随机森林模型中进行回归训练,获取特征重要性得分。

Table 1. Sample distribution of 3 h rainfall before and after sampling along with thresholds

表 1. r3 样本采样前后不同降水量阈值样本分布

降水阈值(mm)	采样前样本数	采样后样本数
无降水	1,292,440	339,646
0.1	372,259	275,024
1	210,402	210,402
5	80,289	80,289
10	36,165	36,165
15	17,153	17,153
20	9650	9650
30	4035	4035
40	1881	1881
50	957	957
60	533	533
70	311	311
80	170	170
90	89	89
100	37	37

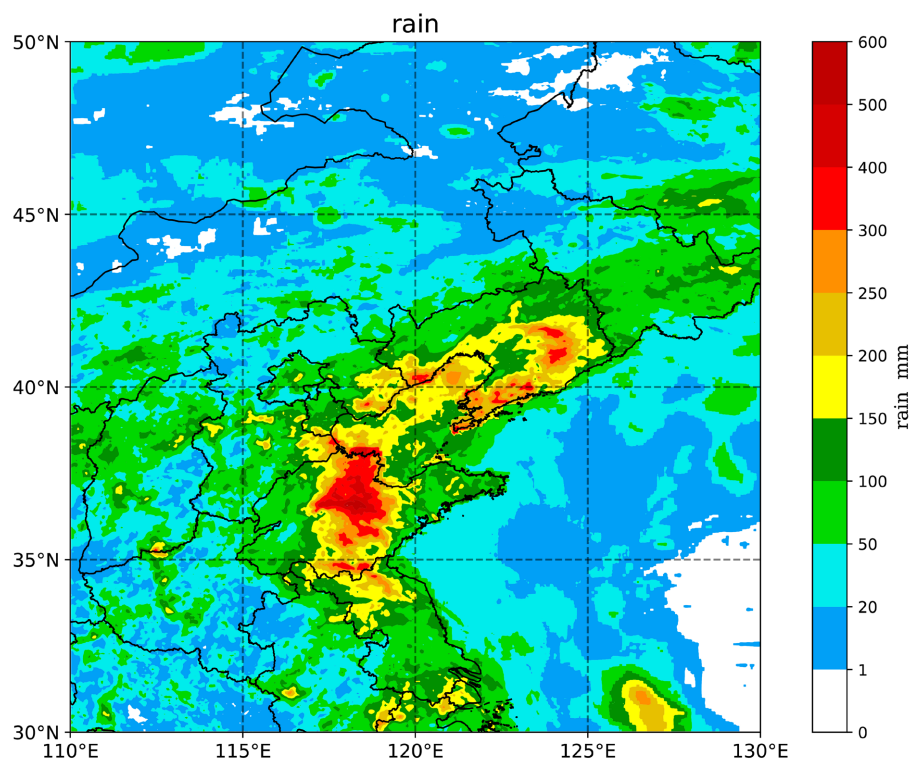


Figure 3. Distribution of accumulated rainfall of selected heavy rain events

图3. 选取强降水事件的累积降水量分布

3. 试验分析

3.1. 特征提取及精简

对原始 376 个特征进行分类贡献度排名, 分别给出以 1 小时/3 小时/6 小时累计降水为目标时 5 个类型特征中得分排名前十名的特征评分值(见表 2 和图 4)。

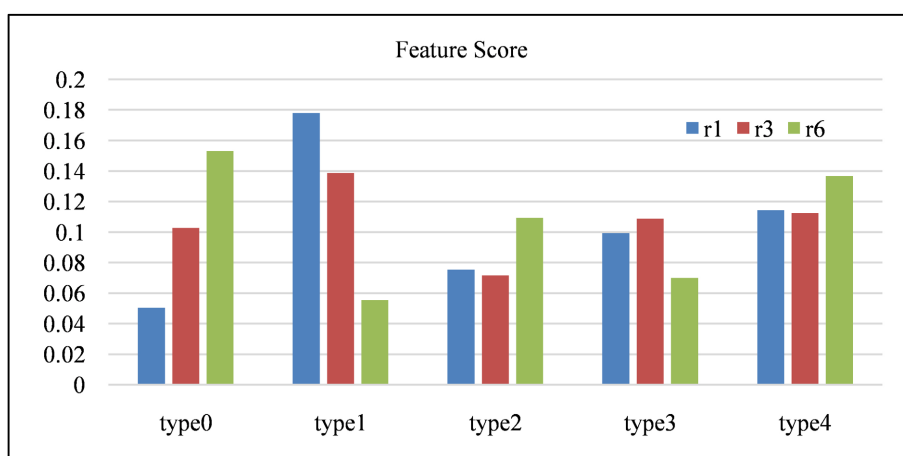


Figure 4. Feature scores of top ten important features in five types with 6 h (orange), 3 h (green) and 1 h (blue) precipitation

图4. 原始特征情况下排名前十特征的分类分值(橙色: 以 r6 为目标量; 绿色: 以 r3 为目标量; 蓝色: 以 r1 为目标量)

Table 2. Feature scores of top ten important features**表 2.** 排名前十特征评分

Top 10	type 0	type 1	type 2	type 3	type 4
r1	0.0503703	0.178047	0.0753751	0.0994199	0.114192
r3	0.10267	0.138791	0.0717084	0.108885	0.11238
r6	0.153034	0.0553356	0.109134	0.0700356	0.136785

图 4 中显示, 随着降水预报精细程度提升(6 小时/3 小时/1 小时累积降水为目标量), 模式直接预报的降水(type 0)贡献度急剧下降, 而相关物理特征, 包括水汽特征(type 1)和热力不稳定特征(type 3)重要性显著增强。这说明, 在越来越精细的小时尺度强降水短临预报中, 模式直接预报的降水可信度迅速减弱, 相反, 物理量预报(尤其是水汽要素)对降水订正有更好的指导意义。从 6 h 累计降水到 1 h 累计降水, 模式直接预报降水贡献越来越小, 与大气的多尺度特征相对应, 也反应了数值模式对多尺度系统的可预报性随精细程度降低。因此, 应用模式预报效果较好的物理特征量订正直接的降水预报具有较大的实际意义。

同时, 从得分直方图(图略)分布中可以看出, 每个类型的前 10 特征中, “马太效应”较为明显, 即排名前三的特征得分之和占据了这类型特征得分总和的 50%以上, 这意味着, 在不牺牲太大准确率的情况下, 可以较为放心的舍弃掉较多的特征, 以此换得较多的计算便利。

3.2. 特征冗余剔除及分析结果

3.2.1. 剔除冗余

原始 376 个特征, 虽然保持了很好的完备性, 但存在冗余。过多的冗余特征会带来如下麻烦: 1) 冗余特征使得机器学习模型更难感知到真正有益于目标识别的重要特征, 因为障碍物变多了; 2) 过多冗余特征使得模型不容易学习到最优的结果; 3) 增大了计算成本, 模型变得复杂。因此很有必要进行冗余特征的剔除和重要特征的选择。这里选择嵌入式特征选择方法, 基学习器采用随机森林。嵌入式的特征选择过程流程是一个完整的模型训练过程。

Step 1: 准备数据集。从所有个例时刻中随机选取十分之一的时刻的样本构建数据集。

Step 2: 考虑到不同时次目标累计降水下, 重要特征可能存在差异(后面的试验也验证了这一点)。分别以 r1、r2 和 r3 为目标, 训练 3 个模型。训练之前需要做降水量级样本平衡(因为无降水样本始终占大多数), 方法为随机剔除 90%的无降水样本。通过随机森林给出的 `feature_importance` 可以定量分析出不同目标下各个特征的得分情况。

Step 3: 将所有特征得分按照 5 个类型(以下简称为 type)进行排序。将 r1, r2 和 r3 中的 type 1 中排名前 15 的特征求并集, 其他类型依次类推, 这样共计得到 5 个 type 的 106 个重要特征。

通过上述步骤既保证了特征的完备性, 也极大降低了特征的冗余程度。根据精简的后的 5 大类 106 个特征, 可以更为仔细的分析出各个类型特征的重要性得分情况。步骤如下:

① 样本选取。从 295 个时刻的样本集按时间顺序每间距一选取一个时次, 选出 144 个时次的样本集。每个时次的样本集中包含 `n_samples` 个样本。如果 `n_sample` 个样本中 r3 (目标 3 小时累计降水) ≥ 1 mm 的样本数量小于 `n_sample/50`, 则认为该时刻不存在有效降水。去除该样本集。通过上述步骤, 将剩下的 `m` 个时刻的样本集的所有样本合并成一个数据集。

② 样本平衡。以 3 小时累计降水(r3)为目标, 根据上述步骤获取的样本数和不同降水量的样本分布情况, 可以看到非降水样本占据了 80%, 另外由于之前站点到格点的插值过程会产生部分 0~0.1 之间的

样本, 因此需要做样本平衡。通过随机采样 1/10 的无降水样本和 r3 在 0~0.1 mm 之间的样本, 即可以得到相对平衡的样本分布(如表 3 所示)。

③ 模型训练。分别以 r1, r2, r3, r6 为目标通过随机森林进行回归学习训练。

Table 3. Sample distribution with different thresholds after sample balance
表 3. 样本平衡处理后的不同降水量样本分布

降水阈值(mm)	采样后样本数
无降水	2,724,136
0.1	1,869,286
1	1,050,201
5	426,027
10	212,528
15	122,916
20	75,308
30	30,543
40	12,124
50	6291
60	3260
70	1769
80	930
90	476
100	228

3.2.2. 分析结果

在保持特征物理完备性的前提下, 按照上述步骤进行冗余特征的精简。得到共计 106 个特征下, 分别以 r1, r2, r3, r6 为目标通过随机森林获取特征量得分统计(图略)。显示结果虽然与图 2 略有不同, 但总体信息类似: 随着预报精细化程度提升(以 r6/r3/r1 为预报目标), 模式直接预报的降水贡献度急剧减小, 而其他物理特征重要性显著增强, 除了与图 2 类似水汽特征和热力不稳定特征贡献增大, 特征精简后动力抬升特征的作用在 1 h 降水预报尤其突出, 维持时间特征重要性也显示出增强。为具体分析高影响的特征因子, 以 3 h 和 1 h 降水预报为例, 给出了分类特征的前十因子排序直方图(图 3 和图 4)。

特征重要性排序显示, 物理要素比时效特征的作用更大, 整体而言越临近时刻的要素预报与目标降水量的相关性越强。直接预报降水特征(Type 0, 图略)显示, EC 模式预报的降水量特征评分远高于其他特征, 其次是地面观测的前一时刻降水和 SMS 中尺度模式预报的降水。以 3 小时累积降水为目标量(如图 5 所示), 在水汽特征中(Type 1), EC 预报的整层可降水量贡献明显高于其他水汽特征, 其次是模式预报的低层水汽含量和水汽通量散度。动力抬升特征中(Type 2), 中低层的风场(低层辐合)作用显著, 但中低层的整层抬升也是值得考虑的因子。在热力特征中(Type 3), EC 模式预报的对流不稳定条件和 SMS 中尺度模式预报的对流有效位能和反射率因子重要性凸显。

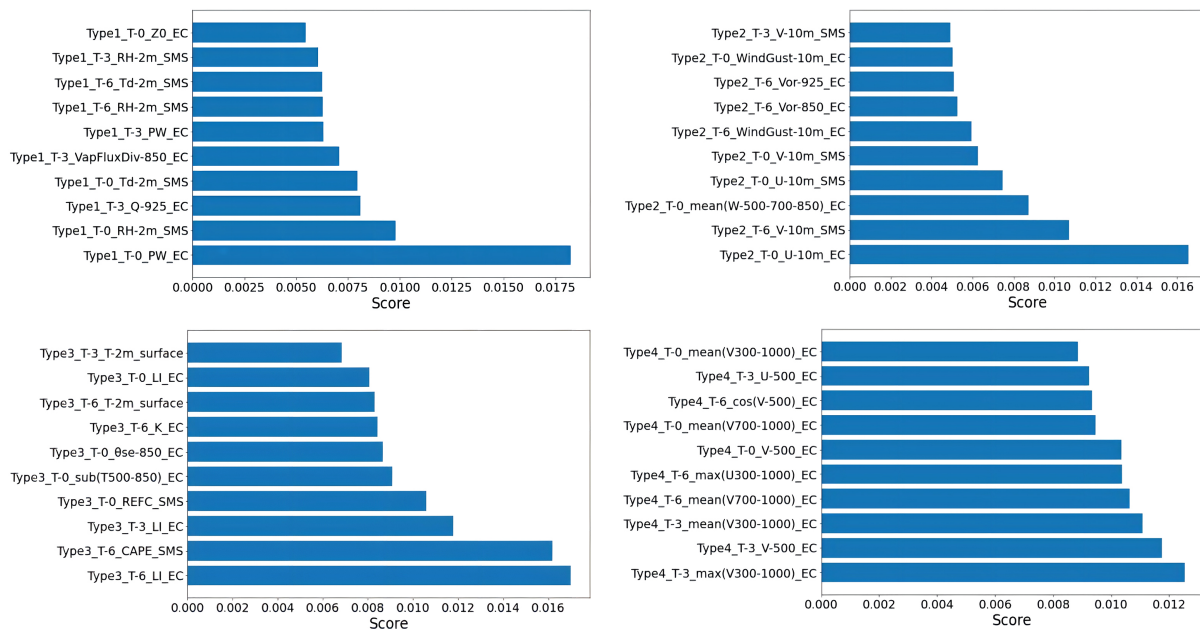


Figure 5. The top ten important physical factors in the five types feature for 3 h precipitation

图 5.3 小时累计降水分类特征贡献前十的物理因子

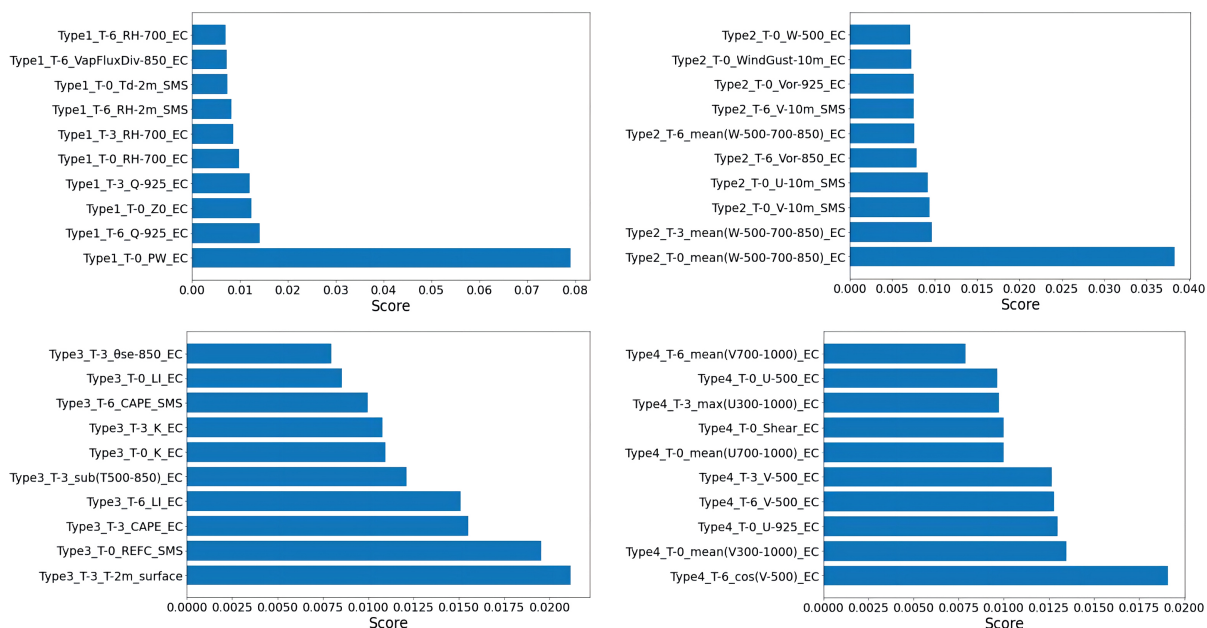


Figure 6. As the same of Figure 5, but for 1 h precipitation

图 6. 同图 5, 但为 1 小时累计降水分类特征

以 1 小时累积降水为目标量进行特征重要性分析(如图 6 所示), 在水汽特征中(Type 1), EC 预报的整层可降水量贡献明显高于其他水汽特征, 其次是模式预报的低层水汽含量, 与 r3 不同, 当精细到逐小时降水时, 与降水强度(云微物理)相关的特征因子, 如零度层高度(Type1_Z0)的重要性位列第三。动力抬升特征中(Type 2), 中底层的整层抬升和近地面层的作用显著, 和 r3 相比, 低层涡度场也是值得考虑的因子。在热力特征中(Type 3), 地面观测的热力条件和中尺度模式预报的反射率因子贡献最大, 其次是 EC

模式预报的对流有效位能和不稳定参数(Li 指数和 K 指数)。总体而言,随着精细化程度提升,热力特征的重要性调整较大,其次是水汽特征。因此,在精细化的强降水预报机器学习订正中应着重考虑这两类特征的提取。

3.3. 特征工程集成提取

为说明上述特征提取的通用性,本节采用特征工程里最常用的 5 种方法进行特征提取集成,包括皮尔逊相关系数法、互信息方法(Mutual Information)、逻辑回归(Logistic Regression, LR)和梯度提升法(GBDT),并与随机森林(RF)结果进行对比。选取各个类型的排名前 15 的特征,对其特征得分进行绝对值平均,如表 4 所示。

Table 4. Importance of the five types of physical features by five methods

表 4. 五种特征工程方法给出的五类物理特征的重要度

Type/Score	相关系数	互信息	LR_L2	RF	GBDT	
r1	Type 0-预报降水	0.0492	0.0779	0.7477	0.0042	0.0065
	Type 1-水汽特征	0.2638	0.0972	1.1563	0.0088	0.0187
	Type 2-动力抬升	0.1217	0.0394	1.6207	0.0035	0.0062
	Type 3-热力特征	0.2611	0.0846	1.1784	0.0063	0.0100
	Type 4-维持时间	0.2186	0.0973	1.8058	0.0048	0.0089
r3	Type 0-预报降水	0.1735	0.0932	1.0637	0.0042	0.0069
	Type 1-水汽特征	0.2060	0.0674	0.9608	0.0088	0.0055
	Type 2-动力抬升	0.1361	0.0706	0.8885	0.0035	0.0038
	Type 3-热力特征	0.1644	0.0618	0.9438	0.0063	0.0043
	Type 4-维持时间	0.1023	0.0839	1.1329	0.0048	0.0042
r6	Type 0-预报降水	0.4769	0.1905	0.4179	0.0130	0.0338
	Type 1-水汽特征	0.2431	0.1270	0.3611	0.0046	0.0044
	Type 2-动力抬升	0.3674	0.1343	0.2932	0.0066	0.0052
	Type 3-热力特征	0.1931	0.1141	0.3494	0.0037	0.0080
	Type 4-维持时间	0.2991	0.1537	0.2547	0.0060	0.0051

表 4 中对每种特征工程方法提取出的特征中最重要的一类特征进行了加粗标注,可以看到不同的提取方法虽然在具体评分上有所差异,但整体提取的特征信息是一致的,在以 r6 为目标量时,模式直接预报的降水可信度很高,是最重要的参考特征,但随着精细程度的提高,其作用急剧下降,水汽特征和动力特征重要性增强,尤其到 1 小时降水预报,多种方法均显示出热力特征的重要性凸显。

4. 总结

近年来,随着人工智能技术的快速发展,结合物理模型和机器学习的专业领域应用得到越来越多关

注。本文利用机器学习对大数据处理的优势,通过特征工程客观提取了不同尺度强降水预报的关键影响因素,获得主要结论如下:

1) 强降水受到大气运动的高维特征影响,通过特征工程客观筛选和先验专家知识判别,精简后的特征可有效去除模型冗余,在不损失机器学习精度和效果的情况下大力提升运行效率。

2) 定时、定点、定量的强降水预报是防灾减灾的重难点之一,当预报较长时间累计降水时(r_6 以上),模式直接预报降水具有较好的参考价值,但当精细到3 h以内时,降水预报可信度急剧减小,而相关物理特征更具参考价值。具有共性的是越接近预报时刻,物理要素的参考价值越大。不同的是对于较长时段的累计降水(r_6),中高层的动力维持作用、整层抬升和整层含水量具有较大贡献,而随着预报精细程度提高(r_1),低层和近地面层的动力作用、热力不稳定特征更加重要,影响降水强度的云微物理相关因子(如零度层高度)作用也更加凸显。

3) 从6 h累计降水到1 h累计,模式直接预报降水贡献越来越小,而相关物理特征如水汽特征,不稳定条件等因子对降水预报具有更大的贡献。这与大气的多尺度特征相对应,也反应了数值模式对多尺度系统的可预报性随精细程度的提升而降低。因此,通过特征工程提取的重要物理特征为提升强降水预报提供了重要支撑。

基金项目

本研究由国家重点研发计划(2018YFC1507606),国家自然科学基金项目(42030611, 42075154, 41505079)资助。

参考文献

- [1] Hernández, E., Sanchez-Anguix, V., Julian, V., Palanca, J. and Duque, N. (2016) Rainfall Prediction: A Deep Learning Approach. In: Martínez-Álvarez, F., Troncoso, A., Quintián, H. and Corchado, E., Eds., *Hybrid Artificial Intelligent Systems. HAIS 2016. Lecture Notes in Computer Science*, Vol. 9648. Springer, Cham. https://doi.org/10.1007/978-3-319-32034-2_13
- [2] 邢彩盈, 张京红, 黄海静. 基于BP神经网络的海口住宅室内气温预测[J]. 贵州气象, 2016, 40(5): 38-42.
- [3] 门晓磊, 焦瑞莉, 王鼎, 等. 基于机器学习的华北气温多模式集合预报的订正方法[J]. 气候与环境研究, 2019, 24(1): 116-124.
- [4] Czernecki, B., Taszarek, M., Marosz, M., et al. (2019) Application of Machine Learning to Large Hail Prediction—The Importance of Radar Reflectivity, Lightning Occurrence and Convective Parameters Derived from ERA5. *Atmospheric Research*, **227**, 249-262. <https://doi.org/10.1016/j.atmosres.2019.05.010>
- [5] Han, H., Lee, S., Im, J., et al. (2015) Detection of Convective Initiation Using Meteorological Imager Onboard Communication, Ocean, and Meteorological Satellite Based on Machine Learning Approaches. *Remote Sensing*, **7**, 9184-9204. <https://doi.org/10.3390/rs70709184>
- [6] 陈浩, 宁忱, 南卓铜, 等. 基于机器学习模型的青藏高原日降水数据的订正研究[J]. 冰川冻土, 2017, 39(3): 583-592.
- [7] 张晨阳, 杨雪冰, 张文生. 气象大数据超短临精准降水机器学习与典型应用[J]. 农业大数据学报, 2019(1): 78-87.
- [8] 黄威, 牛若芸. 基于集合预报和支持向量机的中期强降雨集成预报试验[J]. 气象, 2017, 43(9): 1110-1116.
- [9] 冀慧杰, 倪枫, 刘姜, 等. 基于XGB-BFS特征选择算法的电信客户流失预测[J]. 计算机技术与发展, 2021, 31(5): 21-25.
- [10] 李占山, 刘兆赓. 基于XGBoost的特征选择算法[J]. 通信学报, 2019, 40(10): 101-108.
- [11] 杨长春, 徐筱, 宦娟, 等. 基于随机森林的学生画像特征选择方法[J]. 计算机工程与设计, 2020, 40(10): 2827-2834.
- [12] 王智, 张志强, 谢晓芹, 等. 基于提升树的PM2.5浓度预测模型[J]. 软件, 2018, 39(10): 156-163.