软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2017,28(11):2925−2939 [doi: 10.13328/j.cnki.jos.005336]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

多源数据融合高时空分辨率晴雨分类*

匡秋明[1,3], 杨雪冰[2,3], 张文生[2,3], 何险峰[3], 惠建忠[1,3]

[1](中国气象局 公共气象服务中心,北京 100081)

[2](中国科学院 自动化研究所,北京 100190)

[3](气象大数据与机器学习联合实验室,北京 100190)

通讯作者: 张文生, E-mail: zhangwenshengia@hotmail.com

摘 要: 高时空分辨率晴雨分类与交通、旅游、农业灌溉及人们日常出行都密切相关,然而“天有不测风云”,“东边日头西边雨”,准确的高时空分辨率晴雨分类是极具挑战性的问题.提出了一种基于多源数据的多视角学习晴雨分类方法,其中,多源数据包括雷达、卫星及地面观测因子及晴雨观测数据.该方法表述如下:首先,依据雷达观测因子构造了 VisCAPPI 视角和 VisPPI 视角,依据葵花卫星资料构造了 VisSat 视角,依据地面观测因子构造了 VisGround 视角;然后,对这 4 个视角特征进行组合获得组合视角 VisCAPPI_PPI,VisRadar_Sat,VisRadar_Groumd,VisSat_Ground,VisRadar_Sat_Ground,应用随机森林机器学习方法分别对这些视角进行样本学习,获得这些视角的晴雨分类模型;最后,对这些视角晴雨分类模型估计进行融合,获得晴雨分类结果.主要贡献在于:(1) 提出了雷达、卫星和地面观测因子多视角构建方法,构建了 VisCAPPI,VisPPI,VisSat 和 VisGround 晴雨分类视角及其组合视角;(2) 提出了一种多视角方法(multi-view weight random forest,简称 MVWRF),能够处理雷达、卫星和地面观测因子多源数据融合晴雨分类问题,提高 1km×1km 和 6min 时空分辨率晴雨分类准确率.在 2016 年 10 月 7 日和 8 日,泉州雷达覆盖的 393 个气象观测站上进行模型训练和测试,结果显示,该方法能够取得较高的晴雨分类准确率和较低的漏报率、空报率,优于对比方法.

关键词: 多源数据;随机森林;多视角;晴雨分类
中图法分类号: TP311

中文引用格式: 匡秋明,杨雪冰,张文生,何险峰,惠建忠.多源数据融合高时空分辨率晴雨分类.软件学报,2017,28(11):2925−2939. http://www.jos.org.cn/1000-9825/5336.htm

英文引用格式: Kuang QM, Yang XB, Zhang WS, He XF, Hui JZ. Fusion of multi-source data for rain/no-rain classification with high spatiotemporal resolution. Ruan Jian Xue Bao/Journal of Software, 2017,28(11):2925−2939 (in Chinese). http://www.jos.org.cn/1000-9825/5336.htm

Fusion of Multi-Source Data for Rain/No-Rain Classification with High Spatiotemporal Resolution

KUANG Qiu-Ming[1,3], YANG Xue-Bing[2,3], ZHANG Wen-Sheng[2,3], HE Xian-Feng[3], HUI Jian-Zhong[1,3]

[1](Public Meteorological Service Center, China Meteorological Administration, Beijing 100081, China)

[2](Institute of Automation, The Chinese Academy of Sciences, Beijing 100190, China)

[3](Joint Laboratory of Meteorological Data and Machine Learning, Beijing 100190, China)

Abstract: High spatiotemporal resolution rainfall estimation is closely related to transportation, tourism, agricultural irrigation and people's daily travel. However, accurate high-resolution rain/no-rain classification is a very challenging problem. This paper proposes a

* 基金项目: 国家自然科学基金(61432008, 61532006, 61472423, 61305018)
Foundation item: National Natural Science Foundation of China (61432008, 61532006, 61472423, 61305018)
本文由复杂环境下的机器学习研究专刊特约编辑张道强教授推荐.
收稿时间: 2017-01-10; 修改时间: 2017-04-11; 采用时间: 2017-06-16

multi-source data based multi-view learning method for rain/no-rain classification. The multiple source data used in this paper include radar, satellite and ground observation factors and rain/no-rain observation data. This method can be summarized as follows. Firstly, VisCAPPI view and VisPPI views are constructed according to the radar observation factors. VisSat view is constructed from the sunflower satellite data. VisGround view is constructed according to the ground observation factors. Secondly, the views of VisCAPPI_PPI, VisRadar_Sat, VisRadar_Groumd, VisSat_Ground, and VisRadar_Sat_Ground are obtained by combining features from preconstructed views. Random forest (RF) classification models are trained from these views using RF method. Finally, the rain/no rain classification results are obtained from the estimated results of these RF classification models. The main contributions of this paper arelisted as follows: (1) Present a method for constructing VisCAPPI, VisPPI, VisSat and VisGround views and their feature combined views for radar, satellite and ground observations; (2) A multi-view weight random forest method (MVWRF) is proposed. Multi-source data of radar, satellite and near surface observations are fused for rain/no-rain classification with temporal resolution of 6-minute and spatial resolution of 1km×1km in virtue of the proposed method. The experimental results show that the proposed method in this paper can obtain high precision of rain/no-rain classification after training and testing on 393 meteorological stations covered by radar in Quanzhou on October 7 and 8, 2016.

Key words: multi-source data; random forest; multi-view; rain/no-rain classification

1 概 述

气象分析研究是一个关乎国计民生的课题,其中,晴雨天气预测是重要的一环[1].晴雨分类对交通、旅游、基建、农业灌溉以及人们的日常生活都有重要影响[2].1km×1km 和 6min 时间间隔高时空分辨率条件下的晴雨分类对许多行业更具服务意义.

提高晴雨预报准确率也是气象预报预测领域中一个十分重要的研究课题,但由于降雨是各种尺度的天气系统共同作用的结果,其形成机制非常复杂,具有显著的非线性、时变性特征,因此,利用传统的统计方法很难揭示其变化规律[3].美国热带测雨任务卫星可以实现全球 3h 间隔的降雨估计,给晴雨分类带来希望,但由于受近地面随时间和位置变化气象条件的影响,近地面晴雨区域的估计尽管非常重要但却非常困难[4].受风速、风向、气压、地形等因素的影响,高时空分辨率晴雨分类准确性不高,估计结果不确定性大,是更具挑战性的研究课题.

为了提高晴雨分类的准确率,在过去的研究中,大量学者将雷达、卫星和地面站观测因子多源数据用于晴雨分类,取得了不少研究成果.

- 卫星观测因子晴雨分类方法

文献[4]利用微波辐射计在雨区与不下雨区域亮温差异进行晴雨分类;文献[5]利用卫星图像识别不降雨云,约 60%的非降雨云能够被识别出来;文献[6]利用卫星上的微波辐射计测量云层中的液态水含量,根据液态水含量进行晴雨分类;文献[7]在微波辐射计因子上应用随机森林机器学习算法进行晴雨分类,优于通用的戈达德数字图表法(GPROF)晴雨分类;文献[8]利用变分法进行晴雨分类.卫星在降雨云区检测方法具有优势,但是卫星观测因子主要反映云顶信息,而真正下雨的区域可能只有云覆盖区域的 1/4.当前卫星观测晴雨分类,微波辐射计因子反演云层液态水含量,进行晴雨分类,对卫星观测空间信息利用不足.

- 雷达观测因子晴雨分类方法

文献[9,10]利用 1.5km,2.5km,3.5km 和 4.5km 高度雷达 CAPPI 数据,依据神经网络机器学习方法进行晴雨分类;文献[2]分析利用 CAPPI 和 PPI 进行晴雨分类,其主要依据阈值来判断晴雨,比如依据 1.5km CAPPI 值判断晴雨.雷达观测的多层空间信息及高时空分辨率对晴雨分类非常有利,但当前雷达晴雨分类仅利用少数几层 CAPPI 进行晴雨分类,而单层 CAPPI 或者少数几层 CAPPI 晴雨分类结果易受降水相态、空中飞行物等干扰而影响准确性.

- 数值模式推断气象因子或者地面自动站观测因子晴雨分类方法

这类方法主要根据降雨形成的气象条件来估计晴雨.早在 1980 年,文献[11]就利用八型图推断 24h 晴雨状况;之后,文献[12]利用 T213 数值模式推断气象要素,借助 KNN 机器学习方法获得 12h 间隔晴雨预报,总体上降低了预报空报率,提高了晴雨预报的 TS 评分和预报准确率;文献[3]利用 T639 数值模式推算 40 个气象要素,借助 SVM 机器学习方法获得 12h 间隔晴雨预报,在北京地区夏季晴雨预报中得到应用;文献[1]利用地面自动化观

测站观测的气温、气压、相对湿度等气象要素进行小时晴雨分类,并且考虑到晴雨样本不平衡问题的处理.但是目前,这类方法很少利用雷达和卫星的数据进行晴雨分类,而雷达和卫星的分辨率要远高于地面观测因子的分辨率.

从晴雨分类方法现状来看,单一利用雷达、卫星或者地面观测气象因子都很难实现较好的高时空分辨率晴雨分类.本文提出一种雷达、卫星和地面观测气象因子多视角融合晴雨分类方法.该方法的物理依据如图 1 所示.在 2016 年泉州雷达覆盖区域上的实验结果表明,该方法在 1km×1km 空间分辨率、6min 间隔时间分辨率晴雨分类上,可以明显提高准确性.

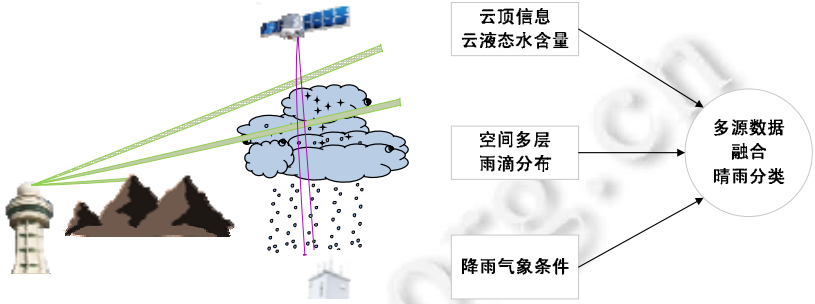

Fig.1 Fusion of radar, satellite and ground observations for rain/no-rain classification

图 1 雷达、卫星和地面观测因子融合晴雨分类

本文的主要贡献如下:

(1) 提出了晴雨分类多视角构建方法.依据雷达观测因子构造了 VisCAPPI 视角和 VisPPI 视角;依据葵花卫星资料构造了 VisSat 视角;依据地面观测因子构造了 VisGround 视角.这 4 个视角都能对 6min、1km×1km 区域进行晴雨分类.

(2) 提出应用空间邻域点特征扩展方法,分别应用在雷达、卫星和地面观测气象因子上,相应地提高了雷达、卫星观测因子单个视角晴雨分类的准确性.

(3) 本文提出了一种多视角权重随机森林方法(MVWRF),实现了雷达、卫星和地面观测气象因子多源数据融合晴雨分类.实验证明,本文所提出的方法能显著提高晴雨分类准确性,优于对比方法.

本文第 1 节说明晴雨分类方法现状,分析当前卫星、雷达及地面观测气象因子晴雨分类所存在的问题,概述本文提出的方法及主要贡献.第 2 节总结多视角相关研究工作进展.第 3 节介绍本文提出的多视角权重随机森林晴雨分类方法.第 4 节介绍实验设计及实验结果.第 5 节介绍进一步实验及结果分析.最后,在第 6 节给出结论和展望,总结全文,并对未来研究方向进行初步探讨.

2 相关工作

由于本文提出一种多视角方法用于雷达、卫星及地面观测因子多源数据融合晴雨分类,因此,本节主要阐述多视角相关研究工作.

文献[13]综述了多视角学习的进展,总结了多视角构建方面的一些主要方法.各种多视角构建方法针对不同问题,各有优劣.这些多视角构建方法与本文在构建晴雨分类多视角构建上的区别与联系如下.

(1) 雷达、卫星和地面观测因子构成了 3 大类具有物理意义的晴雨分类视角.这与文献[14]对图像分块构建多视角有些类似.

(2) 雷达、卫星和地面观察因子视角之间存在时空不一致性,需解决时空不一致性,以方便多视角协同工作.这与一般的多视角方法不同,其原因是:本文采用物理上多视角,而多数方法采用人工生成多视角.

(3) 雷达反射率因子依据物理上水平与竖直投影产生 CAPPI 和 PPI 视角,而且可以对应气象上的 CAPPI 和 PPI 数据产品.

(4) 各视角分类器选择随机森林做为分类,本质上应用了子空间随机投影生成多视角方法.这方面有大量研究工作:文献[15]提出通过随机子空间方法构建决策森林;文献[16]提出随机子空间和不等数量训练样本构建多视角;文献[17]提出综述了高光谱图像,从聚类和随机选择子空间等方面构建多视角.

3 多视角权重随机森林晴雨分类方法

本文主要研究雷达、卫星和地面观测因子多源数据融合晴雨分类,提出一种多视角权重随机森林晴雨分类方法(MVWRF).

选用随机森林作为每一个视角分类模型的原因:(1) 随机森林卓越的分类性能,文献[18]在 121 个数据集上对比了 179 种分类器,随机森林方法取得了最好的分类结果;(2) 随机森林方法具有 Bootstrap 重采样,未被采样的 Out-of-Bag 样本的估计误差是模型泛化误差的无偏估计[19];(3) 随机森林方法易于并行,能处理大数据[20].

该方法流程框图如图 2 所示,包括训练和测试两个过程.

- 训练过程.输入雷达、卫星和地面观测因子以及晴雨训练数据,依据雷达观测数据构建 VisCAPPI 和 VisPPI 晴雨分类视角,依据卫星观测数据构建 VisSat 晴雨分类视角,依据地面观测数据构建 VisGround 晴雨分类视角.在这 4 个视角上分别学习随机森林模型,并获得模型评分.模型评分结合各个视角权重先验,在贝叶斯框架下可以获得各个视角权重.各个视角随机森林模型和视角权重组合到一起形成多视角权重 RF 模型(MVWRF).
- 测试过程.输入卫星、雷达和地面观测因子测试数据,分别构建 VisCAPPI,VisPPI,VisSat 和 VisGround 这 4 个视角,应用多视角权重随机森林模型分别对这 4 个视角进行模型估计,获得 4 个模型估计结果.再对结果进行融合,得出测试样本的最终晴雨分类结果.

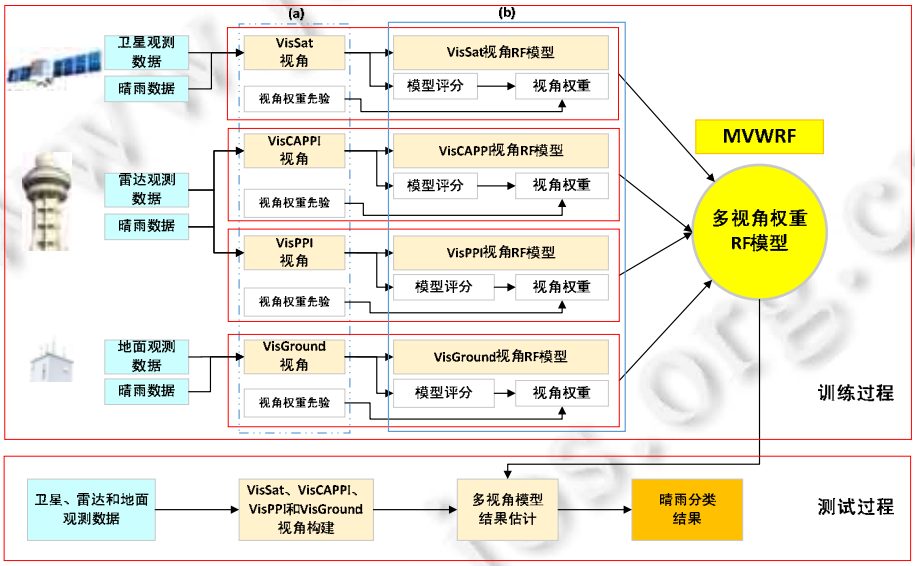

Fig.2 Workflow of multi-view weight random forest method for rain/no-rain classification

图 2 多视角权重随机森林晴雨分类方法流程框图

在这个多视角方法中,关键步骤包括:(a) 由雷达、卫星和地面观测数据生成 VisSat,VisCAPPI,VisPPI 和 VisGround 这 4 个视角及其组合视角;(b) 主导视角和辅助视角随机森林模型学习和视角权重的确定;(c) 晴雨分类结果融合.本节下面的内容详细介绍这 3 个关键步骤.

3.1 雷达卫星及地面观测气象因子视角构建

本节主要介绍雷达、卫星及点观测因子视角构建方法.共有的时空匹配方法包括:反距离加权空间插值[21],

PCHIP 插值方法[22]将 10min 间隔的卫星和地面观测数据转换为 6min 间隔分辨率.此外,本文还将对每个视角特征进行空间扩展[23],增加单视角的有用信息.具体视角构建过程如下.

A. 雷达数据及视角构建

雷达的时空高分辨率使得雷达成为当前高分辨率晴雨分类最合适的手段,因此,雷达观测是晴雨分类的主导视角.按照前面的描述,主导视角需要依据投影变换生成 2 组相辅相成的视角特征.依据气象知识,雷达型号为 CINRAD/SA 多普勒天气雷达(单偏振雷达),获得基本反射率因子.将雷达反射率因子投影到不同海拔高度层上,获得每 6min,1km×1km 分辨率 16 高度层 CAPPI 数据(1km,1.5km,2km,3km,3.5km,4km,4.5km,5km,5.5km,6km,7km,8km,9km,10km,11km,12km),依次组合这些特征,构成网格数据.获取训练自动观测站点的经纬度信息,找到最临近的 4 个网格特征数据,反距离加权插值方法获得自动观测站点对应的 CAPPI 特征数据,除插值生成的 CAPPI 数据之外,再应用空间扩展方法获取临近的 8 个网格点数据作为邻域特征,组合 144 维 CAPPI 特征,获得 VisCAPPI 视角特征.

将雷达反射率因子垂直投影到地面上,每个地理位置上截取 3 层雷达反射率因子,获得每 6min,1km×1km 分辨率 3 高度层 PPI 数据,反距离加权插值方法获得自动观测站点对应的 PPI 特征数据,空间邻域特征扩展获得临近 8 个网格点 PPI 数据,组合 27 维 PPI 特征,构成 VisPPI 视角特征.

B. 卫星数据及视角构建

卫星是识别云、计算云层中体积含水量的重要手段,也是国外晴雨分类研究的重点.以葵花卫星为例,包含 16 个通道的数据,可生成每 10min,2km×2km 分辨率网格数据,反距离加权插值方法获得自动观测站点对应的卫星观测通道特征数据,空间邻域特征扩展获得临近 8 个网格点卫星数据,组合 144 维 PPI 特征,通过时间匹配方法将 10min 的卫星数据插值到 6min,构成 VisSat 视角特征.

C. 地面观测气象因子及视角构建

地面要素包括本站气压、还平面气压、气温、露点温度、相关湿度、水气压等 19 个地面观测数据,通过反距离加权插值获得 1km×1km 格点地面观测数据,空间邻域特征扩展获得临近 8 个网格点地面观测数据,组合 171 维地面观测特征,通过时间匹配方法将 10min 的卫星数据插值到 6min,构成 VisGround 视角特征.

D. 组合视角构建

视角特征组合如图 3 所示.

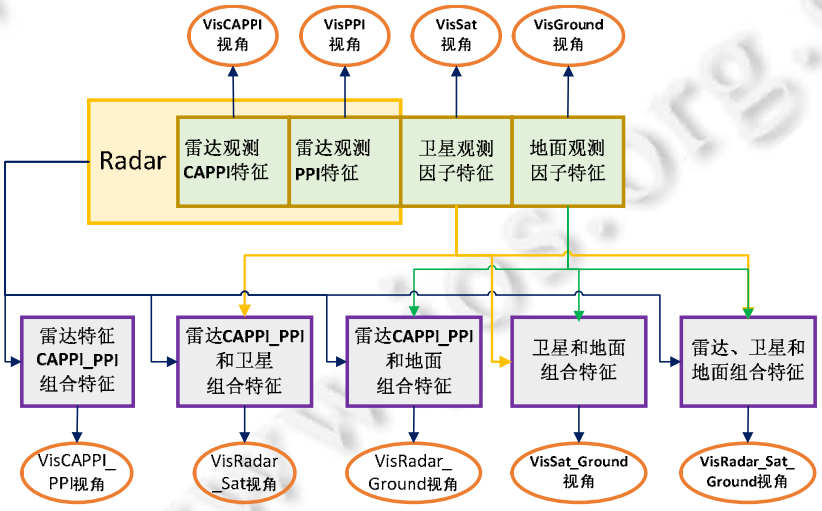

Fig.3 Schematic diagram of generating views from feature combination

图 3 生成视角特征组合示意图

雷达CAPPI和PPI特征组合生成VisCAPPI_PPI视角,雷达CAPPI、PPI和卫星观测特征组合生成VisRadar_Sat视角,雷达CAPPI、PPI和地面观测特征组合生成VisRadar_Ground视角,卫星和地面观测特征组合生成VisSat_Ground视角,雷达CAPPI、PPI、卫星和地面观测特征组合生成VisRadar_Sat_Ground视角.

3.2 各视角晴雨分类分类器学习及视角权重估计

本节主要介绍:(1) 单视角随机森林分类器[19]学习;(2) 视角评价并确定主导视角和辅助视角;(3) 视角权重的确定.

A. 单视角及组合视角随机森林分类器学习

根据雷达观测视角VisCAPPI和VisPPI,卫星观测视角VisSat以及地面观测视角与晴雨观测之间,分别训练4个随机随机森林分类器模型,随机森林构建随机决策树时,节点拆分函数使用Gini指数,随机决策树数目设置为100.依次学习得到的分类模型如下:视角VisCAPPI随机森林分类模型为$H_1(\cdot)$,视角VisPPI随机森林分类模型为$H_2(\cdot)$,视角VisSat随机森林分类模型为$H_3(\cdot)$,视角VisGround随机森林分类模型为$H_4(\cdot)$,视角VisCAPPI_PPI随机森林分类模型为$H_5(\cdot)$,视角VisRadar_Sat随机森林分类模型为$H_6(\cdot)$,视角VisRadar_Ground随机森林分类模型为$H_7(\cdot)$,VisSat_Ground随机森林分类模型为$H_8(\cdot)$,VisRadar_Sat_Ground随机森林分类模型为$H_9(\cdot)$.随机森林分类器学习方法参照文献[19].

B. 多视角模型构建

在确定了单视角随机森林分类器之后,按照组合误差最小的准则,找出总误差最小的4个视角组合,雷达、卫星和地面观测联合多视角分类模型如下:

$$\arg\min_{\Theta} f(\Theta)=\sum_{i=1}^{4}\sum_{j=1}^{N}W_{\Theta(i)}(y_j-H_{\Theta(i)}(x_{ij}))^2 \tag{1}$$

$$W_{\Theta(i)}=(P^i_{prior}\times P^i_{validation})\Big/\sum_{j=1}^{4}(P^j_{prior}\times P^j_{validation}) \tag{2}$$

这里,N表示训练样本个数,x_{ij}表示第i个视角第j个样本的特征,y_j表示晴雨标记,$\Theta(i)$(i=1,2,3,4)表示从单视角及组合视角中随机选择的4个视角的索引,$H_{\Theta(i)}(\cdot)$(i=1,2,3,4)表示从$H_1(\cdot)\overline{\ }H_9(\cdot)$中任选4个视角的分类模型,$W_{\Theta(i)}(\cdot)$表示所选4个视角的视角权重,$P^i_{prior}$($i$=1,2,3,4)表示这所选4个视角分类准确度先验,$P^i_{validation}$(i=1,2,3,4)依次为所选4个视角模型随机森林OOB估计得分,在贝叶斯框架下,$P^i_{prior}\times P^i_{validation}$分别表示这4个视角准确度的后验概率,依据公式(2)计算获得W_i(i=1,2,3,4).根据领域知识和视角权重确定主导视角和辅助视角,保证主导视角数目p大于等于2,辅助视角数目大于q等于1.这里,准确度最好的两个视角被确定为主导视角,另外两个视角被确定为辅助视角.多视角晴雨分类分类器模型训练算法流程见算法1.

算法 **1**. 多视角晴雨分类分类器模型训练算法.

输入:雷达视角训练样本(X_1,X_2,Y)、卫星训练样本(X_3,Y)、地面训练样本(X_4,Y);

雷达视角验证样本(XV_1,XV_2,YV)、卫星训练样本(XV_3,YV)、地面训练样本(XV_4,YV);

视角权重先验概率P^i_{prior}(i=1,2,…,9)和视角模型H_i(i=1,2,…,9).

输出:视角分类模型$H_{\Theta_\min(i)}$(i=1,2,3,4)和视角权重$W_{\Theta_\min(i)}$(i=1,2,3,4).

1. 随机选择4个视角,视角权重先验概率P^i_{prior}(i=1,2,…,9)和视角模型H_i(i=1,2,…,9);
2. for i=1 to 4
3. $H_{\Theta(i)}$=*RandomForestLearning*(X_i,Y)[19];
4. Yt_i= *RandomForestClassify*(H_i,XV_i)[19];
5. 计算第i个视角准确性$P^i_{validation}$(见第5节的准确性计算);
6. 计算后验概率$P^i_{prior}\times P^i_{validation}$;
7. 依据公式(2)计算$W_{\Theta(i)}$;
8. End for

9. 依据公式(1)计算 $f(\Theta)$;

10. 重复步骤 1~步骤 9,遍历各种组合,获得最小的 $f(\Theta\_\min)$;

11. 依据 $\Theta\_\min$,获得 $H_{\Theta_\min(i)}(i=1,2,3,4)$和 $W_{\Theta\_\min(i)}(i=1,2,3,4)$.

这里,验证样本也可以使用 Out-of-Bag 样本.

3.3 晴雨分类结果融合

结果融合动机:(1) 对于双主导视角分类结果一致的估计,确定为最终的估计结果;(2) 对于双主导视角估计结果不一致的情况,由辅助视角投票阈值决定估计结果.

多示例融合晴雨分类过程:首先,应用主导视角学习得到的模型对测试数据进行分类,假定得到测试结果$(T_1,T_2,\ldots,T_p)$,$Y_1=T_1\cap T_2\cap\ldots\cap T_p$(这里,$P$=2,两个主导视角分别为 VisCAPPI 和 VisPPI);如果模型的测试结果一致,则将一致的分类结果作为最终的估计结果.令 $Y_2=T_1\cup T_2\cup\ldots\cup Tp-T_1\cap T_2\cap\ldots\cap Tp$,对于 Y_2,则由辅助视角来确定最终的估计结果.假定辅助视角的测试结果$(TA_1,TA_2,\ldots,TAq)$(这里,q 等于 2,两个辅助视角分别为 VisSat 和 VisGround),辅助视角投票得分最高的类别形成的集合 YA_2 作为 Y_2 的替代估计结果.特别地,针对两类问题,只要样本的辅助视角估计结果在某类别上投票大于设定阈值,即可认为该样本属于这个类别;否则属于另一个类别.晴雨分类结果融合的数学表达如下.

$$f(j)=2\times\sum_{i=1}^{4}r_{ij}H_{\Theta\_\min(i)}(x_{ij}) \tag{3}$$

$$r_{1j}=I_j\cdot W_{\Theta\_\min(1)} \tag{4}$$

$$r_{2j}=I_j\cdot W_{\Theta\_\min(2)} \tag{5}$$

$$r_{3j}=\overline{I}_j\cdot I_{ij}\cdot W_{\Theta\_\min(3)} \tag{6}$$

$$r_{4j}=\overline{I}_j\cdot I_{ij}\cdot W_{\Theta\_\min(4)} \tag{7}$$

这里,N 表示训练样本个数;x_{ij} 表示第 i 个视角第 j 个样本的特征;$H_{\Theta_\min(i)}(\cdot)(i=1,2,3,4)$依次表示 4 个视角的分类模型;$r_{ij}(i=1,2,3,4)$依次表示这 4 个视角模型在测试样本上权重系数,它主要由样本作用系数 I_j, $\overline{I}_j$,I_{ij}和多视角权重系数 $W_{\Theta_\min(i)}(i=1,2,3,4)$决定.样本作用系数确定方法如下.

$$I_j=\begin{cases}1, & H_{\Theta\_\min(1)}(x_{1j})=H_{\Theta\_\min(2)}(x_{2j})\\ 0, & H_{\Theta\_\min(1)}(x_{1j})\neq H_{\Theta\_\min(2)}(x_{2j})\end{cases} \tag{8}$$

$$\overline{I}_j=\begin{cases}0, & H_{\Theta\_\min(1)}(x_{1j})=H_{\Theta\_\min(2)}(x_{2j})\\ 1, & H_{\Theta\_\min(1)}(x_{1j})\neq H_{\Theta\_\min(2)}(x_{2j})\end{cases} \tag{9}$$

$$I_{ij}=\begin{cases}1, & H_{\Theta\_\min(i)}(x_{ij})>0\\ 0, & H_{\Theta\_\min(i)}(x_{ij})=0\end{cases} \tag{10}$$

针对测试样本,晴雨分类结果融合算法流程见算法 2.

算法 **2**. 晴雨分类结果融合算法.

输入:雷达视角测试样本($XT1$,$XT2$)、卫星测试样本 $XT3$、地面测试样本 $XT4$;

4 个视角的随机森林模型 $H_{\Theta\_\min(i)}(\cdot)(i=1,2,3,4)$,权重 $W_{\Theta\_\min(i)}(i=1,2,3,4)$.

输出:晴雨分类结果.

1. 视角 $H_{\Theta_\min(1)}(\cdot)$对第 j 个样本 X_{1j} 分类,获得估计结果 $H_{\Theta\_\min(1)}(X_{1j})$;
2. 视角 $H_{\Theta_\min(2)}(\cdot)$对第 j 个样本 X_{2j} 分类,获得估计结果 $H_{\Theta\_\min(2)}(X_{2j})$;
3. 视角 $H_{\Theta_\min(3)}(\cdot)$对第 j 个样本 X_{3j} 分类,获得估计结果 $H_{\Theta\_\min(3)}(X_{3j})$;
4. 视角 $H_{\Theta_\min(4)}(\cdot)$对第 j 个样本 X_{4j} 分类,获得估计结果 $H_{\Theta\_\min(4)}(X_{4j})$;
5. 依据公式(3)计算,获得 $f(j)$;
6. 如果 $f(j)\geq 0.5$,则判定为下雨;如果 $f(j)<0.5$,则判定为晴.

4 实验设计与实验结果

4.1 实验数据集

本文采用 2016 年 10 月 7 日、8 日距泉州雷达 40km~120km 区域的雷达、卫星和地面观测气象因子作为实验数据.雷达型号为 CINRAD/SA 多普勒天气雷达(单偏振雷达),应用第 3.1 节(A)所述方法构建 VisCAPPI 视角和 VisPPI 视角.雨量站数据来源于距离泉州雷达 40km~120km 的自动观测站,每 6 个连续 1min 雨量观测累积生成 6min 间隔雨量数据.实验中采集了 393 个自动观测站的雨量数据.卫星数据来源于葵花卫星每 10min 间隔 16 通道的观测数据,基于卫星数据,采用第 3.1 节(B)所述方法构建 VisSat 视角.地面要素包括本站气压、海平面气压、气温、露点温度、相关湿度、水气压等 19 个地面观测数据,基于地面观测气象因子,采用第 3.1 节(C)所述方法构建 VisGround 视角.采用第 3.1 节(D)所述方法构建组合视角.数据总体情况见表 1.

Table 1 Data sets

表 **1** 数据集

数据集	数据类型	数据维度	总样本数	晴数据	雨数据	晴雨比
数据集 1	CAPPI 数据	144	3 472	2 521	951	2.65:1
	PPI 数据	27	3 472	2 521	951	2.65:1
	卫星数据	144	3 472	2 521	951	2.65:1
	地面数据	171	3 472	2 521	951	2.65:1
数据集 2	CAPPI 数据	144	62 804	35 972	26 832	1.34:1
	PPI 数据	27	62 804	35 972	26 832	1.34:1
	卫星数据	144	62 804	35 972	26 832	1.34:1
	地面数据	171	62 804	35 972	26 832	1.34:1

4.2 对比方法

本文主要研究卫星、雷达和地面观测因子多源数据融合晴雨分类,因此主要对比卫星观测因子晴雨分类方法、雷达观测因子晴雨分类方法和地面观测因子晴雨分类方法.主要对比方法如下.

a) 典型的卫星晴雨分类方法.应用随机森林机器学习,对微波辐射计信号进行模型学习,进而获得待估计样本的模型估计结果;由于随机森林机器模型分类结果主要取决于强特征的分类精度[24],因此,本文使用卫星的所有通道进行晴雨分类,随机森林模型的随机决策树数量设置为 100,其他参数均采用默认参数.

b) 典型的雷达降雨估计方法.基于组合反射率方法阈值判断方法,基于雷达 CAPPI 方法和 PPI 方法进行晴雨分类.由于神经网络方法结果受网络参数的影响较大,本文中使用随机森林方法代替神经网络方法.随机森林模型的随机决策树数量设置为 100,其他参数均采用默认参数.CAPPI 使用 1.5km,2.5km,3.5km 和 4.5km 高度层 CAPPI 雷达反射率因子,PPI 使用最靠近地面的 3 层 PPI 雷达反射率因子.

c) 典型的地面气象因子晴雨分类机器学习方法.KNN 和 SVM 中使用 19 个地面观测因子进行晴雨分类,KNN 机器学习方法的近邻数目设置为 5,SVM 机器学习方法核函数使用二次函数核.

在实验数据集 1 和数据集 2 上采取 5 折交叉验证方法,每次选择样本的 4/5 样本用于模型训练,剩余样本总数的 1/5 样本用于测试,对交叉验证的测试结果求平均值,得出实验结果.

4.3 评价准则

令 TP 表示测试集中所有观测下雨且模型估计也是下雨的样本数量,FN 表示测试集中所有观测为下雨而模型估计为晴的样本数量,FP 表示测试集中所有观测为晴而模型估计为下雨的样本数量,TN 表示测试集中所有观测为晴且模型估计也是晴的样本数量.评价指标如下.

1) 精确性

$P=TP/(TP+FP)$ 该指标度量模型估计为下雨时,有多大比例真正下雨.

2) 准确性

$A=(TP+TN)/(TP+FN+FP+TN)$ 该指标度量晴雨分类的整体准则性.

3) 召回率

$R=TP/(TP+FN)$ 该指标度量观测下雨有多少被估计正确.

4) F-score

$F=2*P*R/(P+R)$ 该指标是对下雨的平衡度量指标.

5) TS 评分

$TS=TP/(TP+FP+FN)$ 该指标是气象上的 TS 评分度量指标.

6) 漏警概率

$MA=FN/(TP+FN)$ 该指标度量有多少观测下雨被漏报.

7) 虚警概率

$FA=FP/(TP+FP)$ 该指标度量有多少观测晴被误报为下雨.

4.4 实验结果

本节针对本文提出的多视角权重随机森林方法以及上述对比方法,在实验数据集 1 和实验数据集 2 上进行实验,结果如下.

表 2 给出了在实验数据集 1 上本文提出的方法和对比方法的实验结果.从表 1 可以看出:对于高时空分辨率晴雨分类,卫星观测因子晴雨分类方法与基于地面观测因子晴雨分类 KNN 方法相比,准确率提高了 2.95%,召回率提高了 4.99%,漏报率降低了 8.91%,空报率降低了 14.36%;雷达 CAPPI 观测因子晴雨分类方法与卫星观测因子晴雨分类相比,准确率提高了 15.84%,召回率提高了 34.65%,漏报率降低了 53.96%,空报率降低了 76.14%.总体上,雷达多层雷达 CAPPI 晴雨分类结果优于卫星和地面观测因子晴雨分类结果.本文提出的多源数据融合方法晴雨分类结果各项评价指标,均优于对比方法,与对比方法里效果最好的雷达 CAPPI 观测因子晴雨分类方法相比,准确率提高了 1.72%,召回率提高了 5.12%,漏报率降低了 23.33%,空报率降低了 19.05%.

Table 2 Experimental results of different methods on data Set 1

表 **2** 不同方法在数据集 1 的实验结果

晴雨分类方法	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
卫星估计	0.648	0.802	0.609	0.692	0.458	0.391	0.352
组合反射率估计	0.909	0.913	0.759	0.829	0.705	0.241	0.091
CAPPI 估计	0.916	0.929	0.820	0.871	0.762	0.180	0.084
PPI 估计	0.647	0.792	0.539	0.641	0.416	0.461	0.353
地面 KNN 估计	0.589	0.779	0.641	0.703	0.442	0.359	0.411
地面 SVM 估计	0.600	0.729	0.016	0.031	0.016	0.984	0.400
MVWRF	**0.931**	**0.945**	**0.862**	**0.902**	**0.811**	**0.138**	**0.068**

表 3 给出了在实验数据集 2 上本文提出的方法和对比方法的实验结果.

Table 3 Experimental results of different methods on data Set 2

表 **3** 不同方法在数据集 2 的实验结果

晴雨分类方法	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
卫星估计	0.585	0.653	0.643	0.648	0.442	0.357	0.415
组合反射率估计	**0.864**	0.780	0.575	0.662	0.527	0.425	**0.136**
CAPPI 估计	0.811	0.833	0.793	0.812	0.669	0.208	0.187
PPI 估计	0.587	0.651	0.618	0.634	0.431	0.382	0.423
地面 KNN 估计	0.539	0.604	0.498	0.545	0.349	0.502	0.461
地面 SVM 估计	NA	NA	NA	NA	NA	NA	NA
MVWRF	0.829	**0.856**	**0.842**	**0.849**	**0.712**	**0.159**	0.171

从表 3 可以看出:对于高时空分辨率晴雨分类,卫星观测因子晴雨分类方法与基于地面观测因子晴雨分类 KNN 方法相比,准确率提高了 8.11%,召回率提高了 29.12%,漏报率降低了 40.62%,空报率降低了 9.98%;雷达 CAPPI 观测因子晴雨分类方法与卫星观测因子晴雨分类相比,准确率提高了 19.45%,召回率提高了 23.33%,漏

报率降低了 41.74%,空报率降低了 54.94%.总体上,雷达多层雷达 CAPPI 晴雨分类结果优于卫星和地面观测因子晴雨分类结果.本文提出的多源数据融合方法晴雨分类结果各项评价指标,均优于对比方法,与对比方法里效果最好的雷达 CAPPI 观测因子晴雨分类方法相比,准确率提高了 2.76%,召回率提高了 6.13%,漏报率降低了 23.56%,空报率降低了 8.59%.MVWRF 方法在精确率和虚警概率上不如组合反射率晴雨分类,这是因为组合反射率漏警概率很高,也就是将大量难以判断晴雨的情况估计为晴,而晴的样本数量大,从而造成组合反射率估计精确度高、虚警概率低,但是这种大量下雨的情况未被估计出来,服务效果相对较差.

表 3 和表 2 的实验结果表明,卫星观测晴雨分类结果优于或接近地面观测因子晴雨分类结果.这主要是由于相对于卫星观测而言,地面自动化观测点过于稀疏,因此地面观测因子对于高分辨率晴雨分类效果可能不如卫星观测因子.雷达 CAPPI 观测因子晴雨分类结果优于卫星观测和地面观测因子晴雨分类结果,这是由于雷达 CAPPI 更能反映接近地面的低空雨滴分布,而卫星观测主要反映云顶及云综合液态含水量.本文所提出的雷达、卫星和地面观测因子多源数据融合晴雨分类方法综合利用了雷达、卫星和地面观测因子晴雨多源数据融合信息,取得了更好的晴雨分类结果.

5 进一步实验与分析

本节主要分析本文所提出的空间扩展方法对晴雨分类结果的影响,实验不同机器学习方法的晴雨分类效果以及本文所提出的多视角多示例数据融合方法与其他融合方法进行对比实验.

5.1 空间邻域扩展效果对比分析

表 4 给出了在数据集 1 上,雷达、卫星及地面观测因子晴雨分类以及雷达、卫星及地面观测因子空间邻域扩展对比效果.雷达空间邻域扩展准确度提高了 0.76%,召回率提高了 4.86%,漏报率降低了 16.28%,虚警概率增加了 2.44%.卫星空间邻域扩展准确度提高了 1.62%,召回率提高了 9.52%,漏报率降低了 13.30%,虚警概率降低了 3.69%.地面观测空间邻域扩展准确度提高了 0.12%,召回率提高了 0.15%,漏报率降低了 12.44%,虚警概率降低了 3.33%.

Table 4 Experimental results of spatial information extensionon data Set 1

表 **4** 数据集 1 的空间信息扩展实验结果

视角特征	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
最大反射率	**0.918**	0.922	0.782	0.848	0.733	0.215	**0.082**
多点多层 CAPPI	0.916	**0.929**	**0.820**	**0.871**	**0.762**	**0.180**	0.084
组合反射率	**0.909**	0.913	0.759	0.829	0.705	0.241	**0.091**
多点多层 PPI	0.906	**0.931**	**0.834**	**0.879**	**0.767**	**0.166**	0.094
卫星数据	0.648	0.802	0.609	0.692	0.458	0.391	0.352
多点卫星数据	**0.663**	**0.815**	**0.667**	**0.729**	**0.493**	**0.339**	**0.339**
地面观测数据	**0.661**	**0.814**	0.661	0.699	0.472	0.386	**0.330**
多点地面观测数据	0.659	0.813	**0.662**	**0.729**	**0.492**	**0.338**	0.341

表 5 给出了在数据集 2 上,雷达、卫星及地面观测因子晴雨分类以及雷达、卫星及地面观测因子空间邻域扩展对比效果.

Table 5 Experimental results of spatial information extensionon data Set 2

表 **5** 数据集 2 的空间信息扩展实验结果

视角特征	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
最大反射率	**0.847**	0.806	0.666	0.729	0.595	0.334	**0.153**
多点多层 CAPPI	0.827	**0.843**	**0.801**	**0.821**	**0.686**	**0.199**	0.173
组合反射率	**0.864**	0.780	0.575	0.662	0.527	0.425	**0.136**
多点多层 PPI	0.826	**0.836**	**0.781**	**0.807**	**0.671**	**0.219**	0.174
卫星数据	0.585	0.653	0.643	0.648	0.442	0.357	0.415
多点卫星数据	**0.589**	**0.656**	**0.662**	**0.659**	**0.451**	**0.338**	**0.414**
地面观测数据	0.574	0.651	**0.713**	**0.681**	**0.466**	**0.287**	0.426
多点地面观测数据	**0.586**	**0.656**	0.666	0.661	0.453	0.334	**0.414**

雷达空间邻域扩展准确度提高了 4.59%,召回率提高了 20.27%,漏报率降低了 40.42%,虚警概率升高了 13.07%.卫星空间邻域扩展准确度提高了 0.46%,召回率提高了 2.95%,漏报率降低了 5.32%,虚警概率降低了 0.24%.地面观测空间邻域扩展准确度提高了 0.12%,召回率减少了 6.59%,漏报率升高了 16.38%,虚警概率降低了 2.82%.

从实验结果来看,空间邻域扩展主要作用是提高了雷达和卫星的召回率,降雨估计精度明显提高,漏报率降低.对于地面观测因子,提升效果不太明显.

5.2 多种机器学习方法效果对比分析

常用机器学习对比方法有支持向量机、*K* 近邻、朴素贝叶斯、adaboost 和随机森林算法.支持向量机算法默认采用 RBF 核,其他参数为默认设置;*K* 近邻算法近邻数量采用 5 近邻;随机森林算法随机决策树数量设置为 100,其他参数为默认设置.实验数据为雷达、卫星和地面观测因子组合特征向量.

表 6 显示了随机森林、支持向量机、KNN、adaboost 和朴素贝叶斯机器学习方法在数据集 1 上的晴雨分类结果.从实验结果来看,随机森林机器学习方法在晴雨分类准确度和 TS 评分上高于其他机器学习方法,除朴素贝叶斯方法外,其他评价指标与最优的评价指标接近.朴素贝叶斯方法在召回率、*F*-score 和漏警概率上结果好,但是虚警概率是其他机器学习方法的 2 倍以上,将较多难以分辨的晴雨状态分类为雨,从而造成雨的准确性高,而晴的准确性很低.

Table 6 Experimental results of different machine learning methods on data Set 1

表 **6** 数据集 1 的不同机器学习方法实验结果

机器学习方法	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
随机森林	0.905	**0.933**	0.842	0.885	**0.774**	0.158	0.095
支持向量机	NA	NA	NA	NA	NA	NA	NA
KNN	**0.919**	0.928	0.810	0.865	0.756	0.190	**0.081**
adaboost	**0.919**	0.927	0.803	0.861	0.750	0.197	**0.081**
朴素贝叶斯	0.800	0.908	**0.888**	**0.898**	0.727	**0.112**	0.201

表 7 显示了随机森林、支持向量机、KNN、adaboost 和朴素贝叶斯机器学习方法在数据集 2 上的晴雨分类结果.支持向量机选择常用的各种核,在数据集 2 上实验多次迭代不收敛.从实验结果来看,随机森林机器学习方法在晴雨分类各项评价指标上优于或不低于其他机器学习方法.

Table 7 Experimental results of different machine learning methods on data Set 2

表 **7** 数据集 2 的不同机器学习方法实验结果

机器学习方法	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
随机森林	**0.825**	**0.852**	**0.828**	**0.839**	**0.704**	**0.172**	**0.175**
支持向量机	NA	NA	NA	NA	NA	NA	NA
KNN	0.823	0.823	0.747	0.783	0.643	0.253	0.178
adaboost	**0.825**	0.826	0.753	0.788	0.649	0.247	**0.175**
朴素贝叶斯	0.772	0.816	0.808	0.812	0.652	0.193	0.228

表 8 显示了随机森林、KNN、adaboost 和朴素贝叶斯机器学习方法在数据集 1 上的晴雨分类结果.从实验结果来看,除朴素贝叶斯方法外,随机森林机器学习方法在晴雨分类准确度上高于其他机器学习方法,漏报率和空报率低于其他机器学习方法.朴素贝叶斯方法和表 6 的实验结果一样,空报率显著高于其他机器学习方法.

Table 8 Experimental results ofrain/no rain classification using different machine learning methods as basic classifiers on data Set 1

表 **8** 数据集 1 的不同机器学习视角基分类器晴雨分类实验结果

基分类器方法	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
随机森林	**0.931**	**0.945**	0.862	**0.902**	**0.811**	0.138	**0.068**
KNN	0.903	0.932	0.841	0.881	0.772	0.159	0.097
adaboost	0.929	0.938	0.836	0.884	0.786	0.164	0.071
朴素贝叶斯	0.827	0.918	**0.884**	0.900	0.746	**0.116**	0.173

表 9 显示了随机森林、KNN、adaboost 和朴素贝叶斯机器学习方法在数据集 2 上的晴雨分类结果.从实验结果来看,随机森林机器学习方法在晴雨分类各项评价指标上优于其他机器学习方法.支持向量机选择常用的各种核,在数据集 1 和数据集 2 上出现迭代不收敛的情况.

Table 9 Experimental results of rain/no rain classificationusing different machine learning methods as basic classifiers on data Set 2

表 **9** 数据集 2 的不同机器学习视角基分类器晴雨分类实验结果

基分类器方法	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
随机森林	**0.829**	**0.856**	**0.842**	**0.849**	**0.712**	**0.159**	**0.171**
KNN	0.796	0.829	0.800	0.814	0.664	0.201	0.204
adaboost	0.824	0.833	0.767	0.798	0.659	0.233	0.176
朴素贝叶斯	0.768	0.818	0.814	0.816	0.654	0.186	0.230

从这种常用的机器学习方法对比来看,本文选择随机森林机器学习方法可以取得较好的晴雨分类结果.

5.3 雷达卫星地面观测因子融合晴雨分类对比

本节主要从两个方面进行因子融合晴雨分类对比:(1) 不同视角特征组合下晴雨分类对比;(2) 全视角特征下不同视角融合方法晴雨分类效果对比.

A. 视角特征组合对比

本节分析各种不同视角特征相互组合晴雨分类对比结果,模型学习方法为随机森林机器学习方法.随机森林算法随机决策树数量设置为 100,其他参数为默认设置.

表 10 的实验结果表明:从单一视角来看,雷达 VisCAPPI 视角取得了最好的晴雨分类结果;从因子组合来看,雷达、卫星、地面观测全因子组合下取得了最好的晴雨分类结果;从雷达视角来看,与卫星观测因子组合后性能略有上升,与地面观测因子组合后性能严重下降;从卫星视角来看,与雷达观测因子及与地面观测因子组合晴雨分类性能有大幅度上升;从地面观测因子来看,与卫星观测因子组合性能有大幅度提升.雷达双主导视角因子组合 CAPPI_PPI 以及卫星地面因子组合均有不错的性能,这也促成了本文所提出的多视角多示例融合方法能否取得很好的晴雨分类效果.

Table 10 Experimental results of different factors on data Set 1

表 **10** 数据集 1 的因子组合实验结果

视角	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
VisCAPPI	**0.916**	0.929	0.820	0.871	0.762	0.180	**0.084**
VisPPI	0.647	0.792	0.539	0.641	0.416	0.461	0.353
VisSat	0.648	0.802	0.609	0.692	0.458	0.391	0.352
VisGround	0.661	0.814	0.661	0.699	0.472	0.386	0.330
VisCAPPI_PPI	0.913	0.934	0.839	0.884	0.777	0.161	0.087
VisRadar_Sat	0.905	0.934	**0.850**	**0.890**	0.780	**0.150**	0.095
VisRadar_Ground	0.662	0.813	0.656	0.726	0.490	0.344	0.338
VisSat_Ground	0.915	**0.935**	0.842	0.886	**0.781**	0.158	0.085
VisRadar_Sat_Ground	0.905	0.933	0.842	0.885	0.774	0.158	0.095

表 11 的实验结果表明:雷达 VisCAPPI 视角是取得最好晴雨分类效果的单一视角;雷达和卫星因子组合以及卫星和地面因子组合的晴雨分类效果更加显著,在一些评价指标上超过了雷达、卫星和地面观测全因子组合晴雨分类效果;雷达 CAPPI_PPI 因子组合晴雨分类结果与雷达卫星因子组合晴雨分类效果接近,而雷达 VisCAPPI 和 VisPPI 双主导视角正是利用了 CAPPI_PPI 因子信息.此外,卫星和地面观测因子信息正是本文的辅助视角所利用的信息,这也从客观上说明本文所提出的方法适合雷达、卫星和地面观测因子多源数据融合晴雨分类,能够取得更好的晴雨分类结果.

B. 多视角融合方法效果对比分析

多视角融合对比方法有 PCA、子空间学习 LSL[25]、典型相关分析 GCCA、字典学习+PCA、多视角协同表示(RKR[26]).应用多视角方法进行晴雨分类训练和测试的过程如下:首先,应用多视角融合方法对训练数据进

行特征变换;然后,再对变换后的数据进行随机森林机器学习,获得分类模型;最后,对测试数据进行相应的变换,应用获得的分类模型对变换后的数据进行测试,根据测试结果和观测结果得到评价指标.

Table 11 Experimental results of different factors on data Set 2

表 **11** 数据集 2 的因子组合实验结果

视角	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
VisCAPPI	0.811	0.833	0.793	0.812	0.669	0.208	0.187
VisPPI	0.587	0.651	0.618	0.634	0.431	0.382	0.423
VisSat	0.585	0.653	0.643	0.648	0.442	0.357	0.415
VisGround	0.574	0.651	0.713	0.681	0.466	0.287	0.426
VisCAPPI_PPI	**0.829**	0.845	0.804	0.824	0.689	0.197	**0.171**
VisRadar_Sat	0.826	0.851	0.826	0.838	0.703	0.174	0.174
VisRadar_Ground	0.585	0.656	0.668	0.662	0.454	0.332	0.415
VisSat_Ground	0.824	0.851	0.827	**0.839**	0.703	0.173	0.176
VisRadar_Sat_Ground	0.825	**0.852**	**0.828**	**0.839**	**0.704**	**0.172**	0.175

本文所提出的方法如第 3 节所述,其中的随机森林算法参数与多视角融合方法所使用的随机森林算法参数一致,随机决策树数量设置为 100,其他参数为默认设置.表 12 和表 13 的实验结果表明,虽然本文构建了雷达 VisCAPPI 和 VisPPI、卫星 VisSat 以及地面 VisGround 这 4 个视角,但是直接应用多视角融合方法并不能取得好的晴雨分类结果;而本文所提出的多视角融合方法能够取得很好的晴雨分类结果.

Table 12 Experimental results of different multiview methods on data set 1

表 **12** 数据集 1 的多视角方法实验结果

多视角方法	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
Radar_sat_ground	0.905	0.933	0.842	0.885	0.774	0.158	0.095
PCA	0.279	0.657	0.261	0.373	0.178	0.672	0.672
GCCA	0.735	0.772	0.614	0.684	0.433	0.487	0.405
RKR	0.367	0.670	0.480	0.560	0.290	0.418	0.405
MVWRF	**0.931**	**0.945**	**0.862**	**0.902**	**0.811**	**0.138**	**0.068**

Table 13 Experimental results of different multiview methods on data set 2

表 **13** 数据集 2 的多视角方法实验结果

多视角方法	精确度	准确度	召回率	*F*-score	TS 评分	漏警概率	虚警概率
Radar_sat_ground	0.825	0.852	0.828	0.839	0.704	0.172	0.175
PCA	0.472	0.566	0.474	0.516	0.315	0.513	0.514
GCCA	0.705	0.649	0.696	0.671	0.466	0.421	0.415
RKR	NA	NA	NA	NA	NA	NA	NA
MVWRF	**0.829**	**0.856**	**0.842**	**0.849**	**0.712**	**0.159**	**0.171**

6 结论与展望

本文研究雷达、卫星和地面观测多源数据下的晴雨分类问题,概述了国际国内雷达晴雨分类、卫星晴雨分类以及地面观测因子晴雨分类主要方法.在此基础上,本文构建了雷达 VisCAPPI 和 VisPPI 视角、卫星 VisSat 视角和地面 VisGround 这 4 个视角及其组合视角,提出了一种多视角融合晴雨分类方法.在 2016 年 10 月 7 日和 10 月 8 日泉州雷达 131 个自动观测站上测试验证,主要结论如下.

1) 在 10 月 7 日数据集 1 上,5 折交叉验证,结果表明,本文所提出的方法比最好的晴雨分类方法准确率提高了 1.72%,召回率提高了 5.12%,漏报率降低了 23.33%,空报率降低了 19.05%.在 10 月 8 日数据集 2 上,本文所提出的方法比最好的晴雨分类方法准确率提高了 2.76%,召回率提高了 6.13%,漏报率降低了 23.56%,空报率降低了 8.59%.
2) 本文所提出的空间扩展特征构造方法,对于雷达和卫星观测晴雨分类均能取得性能的提升.
3) 随机森林机器学习方法在晴雨分类模型学习上,相对于 SVM、KNN、adaboost 和朴素贝叶斯方法更具优势.

4) 本文在已构建的 4 个视角融合实验时,所对比的多视角融合方法都未能取得晴雨分类效果的提升,本文提出的融合方法可以取得晴雨分类效果的显著提升.

值得说明的是,本文主要研究卫星、雷达和地面观测因子多源数据融合方法所使用的数据集,晴雨样本数量基本接近,在实际应用中,使用大量历史数据进行模型训练时,需研究晴雨训练样本分布不平衡问题,采用合适的不平衡学习方法进行处理.

References:

[1] Wang JH, Liang L, Wang B. Analysis of imbalanced weather data based on branch-and-bound approach. Application Research of Computers, 2016,33(6):1648−1652 (in Chinese with English abstract).

[2] Yoo C, Kang M, Ro Y. Applicability of precipitable water for enhancing radar accuracy on identification of rain and no rain. Journal of Korean Society of Hazard Mitigation, 2015,15(1):111−121. [doi: 10.9798/KOSHAM.2015.15.1.111]

[3] He N, Fu ZY, Zhao W, Wu J, Wu JK, Liao XN. Application of SVM method to summer clear-rain forecast in Beijing region. Torrential Rain and Disasters, 2013,32(3):284−288 (in Chinese with English abstract).

[4] Seto S, Takahashi N, Iguchi T. Rain/No-Rain classification methods for microwave radiometer observations over land using statistical information for brightness temperatures under no-rain conditions. Journal of Applied Meteorology, 2005,44(44): 1243−1259. [doi: 10.1175/JAM2263.1]

[5] Xu LM, Sorooshian S, Gao XG, Gupta HV. A cloud-patch technique for identification and removal of no-rain clouds from satellite infrared imagery. Journal of Applied Meteorology, 2010,38(8):1170−1181. [doi: 10.1175/1520-0450(1999)038<1170:ACPTFI>2.0.CO;2]

[6] Kida S, Shige S, Kubota T, Aonashi K, Okamoto K. Improvement of rain/no-rain classification methods for microwave radiometer observations over the ocean using a 37GHz emission signature. Journal of the Meteorological Society of Japan.Ser.II, 2009,87: 165−181. [doi: 10. 2151/jmsj.87A.165]

[7] Islam T, Rico-Ramirez MA, Srivastava PK, Dai Q. Non-Parametric rain/no rain screening method for satellite-borne passive microwave radiometers at 19~85GHz channels with the random forests algorithm. Int'l Journal of Remote Sensing, 2014,35(9): 3254−3267. [doi: 10.1080/01431161.2014.903444]

[8] Araki K, Murakami M, Ishimoto H, Tajiri T. Ground-Based microwave radiometer variational analysis during no-rain and rain conditions. Scientific Online Letters on the Atmosphere Sola, 2015,11:108−112. [doi: 10.2151/sola.2015-026]

[9] Xiao RR, Chandrasekar V, Liu H, Gorgucci E. Detection of rain/no rain condition on ground from radar data using a Kohonen neural network. In: Proc. of the IEEE Int'l Symp. on Geoscience and Remote Sensing. IEEE, 1998. 159−161. [doi: 10.1109/IGARSS.1998.702834]

[10] Liu H, Chandrasekar V, Gorgucci E. Detection of rain/no rain condition on the ground based on radar observations. IEEE Trans. on Geoscience and Remote Sensing, 2001,39(3):696−699. [doi: 10.1109/36.911127]

[11] Li ZL. A short-term weather forecast method for rain/no rain classification. Journal of Meteorological Research and Application, 1980,(4):25−29 (in Chinese).

[12] Zhou MF, Xiong W, Liu HZ. Forecast experiments of rain/no rain in Guizhou using KNN method. JournM of Guizhou Meteorology, 2010,34(6):3−5 (in Chinese).

[13] Xu C, Tao D, Xu C. A survey on multi-view learning. In: Proc. of the Computer Science. 2013. 1304−5634.

[14] Bickel S, Scheffer T. Multi-View clustering. In: Proc. of the IEEE Int'l Conf. on Data Mining, Vol.4. 2004. 19−26. [doi: 10.1109/ICDM.2004.10095]

[15] Ho TK. The random subspace method for constructing decision forests. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1998,20(8):832−844. [doi: 10.1109/34.709601]

[16] Tao DC, Tang XO, Li XL, Wu XD. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006,28(7):1088−1099. [doi: 10.1109/TPAMI.2006.134]

[17] Di W, Crawford MM. View generation for multiview maximum disagreement based active learning for hyperspectral image classification. IEEE Trans. on Geoscience and Remote Sensing, 2012,50(5):1942−1954. [doi: 10.1109/TGRS.2011.2168566]

[18] Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? Journal of Machine Learning Research, 2014,15(1):3133−3181.

[19] Breiman L. Random forests. Machine Learning, 2001,45(1):5−32. [doi: 10.1023/A:1010933404324]

[20] Río SD, López V, Benítez JM, Herrera F. On the use of MapReduce for imbalanced big data using random forest. Information Sciences, 2014,285:112−137. [doi: 10.1016/j.ins.2014.03.043]

[21] Yang XH, Xie XJ, Liu DL, Ji F, Wang L. Spatial interpolation of daily rainfall data for local climate impact assessment over greater Sydney region. In: Advances in Meteorology. 2015. 1−12. [doi: 10.1155/2015/563629]

[22] Fritsch FN, Carlson RE. Monotone piecewise cubic interpolation. SIAM Journal on Numerical Analysis, 1980,17(2):238−246.

[23] Kuang QM, Yang XB, Zhang WS, Zhang GP. Spatiotemporal modeling and implementation for radar-based rainfall estimation. IEEE Geoscience and Remote Sensing Letters, 2016,13(11):1601−1605. [doi: 10.1109/LGRS.2016.2597170]

[24] Biau G. Analysis of a random forests model. The Journal of Machine Learning Research, 2012,13(1):1063−1095.

[25] Zhang L, Zhu PF, Hu QH, Zhang D. A linear subspace learning approach via sparse coding. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2011. 755−761. [doi: 10.1109/ICCV.2011.6126313]

[26] Wang S. Relaxed collaborative representation for pattern classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2012. 2224−2231. [doi: 10.1109/CVPR.2012.6247931]

附中文参考文献:

[1] 王剑辉,梁路,王彪.基于分支限界的不平衡气象数据晴雨分析.计算机应用研究,2016,33(6):1648−1652.

[3] 何娜,付宗钰,赵玮,吴进,吴剑坤,廖晓农.SVM 方法在北京地区夏季晴雨预报中的初步应用.暴雨灾害,2013,32(3):284−288.

[11] 李志陆.一个短期晴雨天气预报方法.气象研究与应用,1980,(4):25−29.

[12] 周明飞,熊伟,刘还珠.KNN 方法在贵州晴雨预报中的实验.贵州气象,2010,34(6):3−5.

匡秋明(1982－),男,湖南祁东人,博士,主要研究领域为气象大数据,机器学习,人工智能,视频图像处理.

杨雪冰(1991－),男,博士生,主要研究领域为大数据挖掘,机器学习,人工智能.

张文生(1965－),男,博士,研究员,博士生导师,CCF 专业会员,主要研究领域为大数据挖掘,计算机视觉,模式识别,人工智能,人机交互.

何险峰(1957－),男,正研级高工,主要研究领域为气象信息技术,气象物理量诊断计算.

惠建忠(1969－),男,高级工程师,主要研究领域为气象信息技术.