〈图像处理与仿真〉

基于特征重要性的高光谱图像分类

张因国¹,陶于祥¹,罗小波²,刘明皓¹

(1. 重庆邮电大学 计算机科学与技术学院, 重庆 400065; 2. 重庆市气象科学研究所, 重庆 401147)

摘要: 为了减少高光谱图像中的冗余以及进一步挖掘潜在的分类信息,本文提出了一种基于特征重要性的卷积神经网络(convolutional neural networks, CNN)分类模型。首先,利用贝叶斯优化训练得到的随机森林模型(random forest, RF)对高光谱遥感图像进行特征重要性评估;其次,依据评估结果选择合适数目的高光谱图像波段,以作为新的训练样本;最后,利用三维卷积神经网络对所得样本进行特征提取并分类。基于两个实测的高光谱遥感图像数据,实验结果均表明:相比原始光谱信息直接采用支持向量机(support vector machine, SVM)和卷积神经网络的分类效果,本文所提基于特征重要性的高光谱分类模型能够在降维的同时有效提高高光谱图像的分类精度。 关键词: 高光谱图像;特征重要性;波段选择;卷积神经网络;支持向量机 中图分类号: TP751 文献标识码: A 文章编号: 1001-8891(2020)12-1185-07

Hyperspectral Image Classification Based on Feature Importance

ZHANG Yinguo¹, TAO Yuxiang¹, LUO Xiaobo², LIU Minghao¹

(1.College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;
 2. Chongqing Institute of Meteorological Science, Chongqing 401147, China)

Abstract: To reduce the redundancy in hyperspectral images and further explore their potential classification information, a convolutional neural network(CNN) classification model based on feature importance is proposed. First, the random forest(RF) model obtained by Bayesian optimization training is used to evaluate the importance of hyperspectral images. Second, an appropriate number of hyperspectral image bands are selected as new training samples according to the evaluation results. Finally, the 3D-CNN is used to extract and classify the obtained samples. Based on two sets of measured hyperspectral remote sensing image data, the experimental results demonstrate the following: compared with the original spectral information obtained directly using a support vector machine(SVM) and the CNN classification effect, the proposed hyperspectral classification model based on feature importance can effectively improve the classification accuracy of hyperspectral images while reducing dimensionality.

Key words: hyperspectral image, feature importance, band selection, CNN, SVM

0 引言

近年来,高光谱图像(hyperspectral images, HSIs) 已经在城市制图、环境管理、作物分析等^[1]领域中得 到愈加广泛的应用,随之而来的是对遥感数据的处理 问题,在HSIs领域中,训练样本的数量与可用的光 谱波段数量相比通常是相当有限的,这会导致训练过 程不完全,容易出现过拟合,即休斯现象^[2]。如何高效地处理高光谱数据成为了一个研究热点。当高光谱 图像的波段数目较大时,其波段组合的方式会呈现指数级别增长,造成波段组合数目过于庞大,从而很大 程度上影响到后续算法的运行速度和复杂性。而这一问题可以通过特征选择进行解决,即:通过不同的方 法选取包含信息量大的波段或者特征来降低数据的

收稿日期: 2020-07-21; 修订日期: 2020-09-15.

作者简介: 张因国(1996-), 男,硕士研究生,主要从事遥感图像分类研究。E-mail: S180231026@cqupt.edu.cn。

通信作者: 陶于祥(1966-), 男, 博士, 教授, 研究方向为资源与环境经济学。E-mail: taoyx@cqupt.edu.cn。

基金项目: 国家自然科学基金项目,"城市地表温度降尺度模型及热岛时空演变规律研究"(41871226); 重庆市应用开发计划重点项目 (cstc2014yykfB30003)。

冗余程度。为减少高光谱遥感图像数据的冗余信息, 目前,许多研究学者针对高光谱数据的降维进行了深 入研究并取得了一定成果。降维的方法包括特征提取 与特征选择。大多数特征提取方法如: 主成分分析 (principal components analysis, PCA)^[3]、独立成分 分析 (independent component analysis, ICA)^[4]、奇 异谱分析 (singular spectrum analysis, SSA)^[5]等,都 为线性处理方法。但是高光谱图像数据本质上是非线 性分布的,而且特征提取所得的低维特征会损失原始 高维特征中某些潜在的分类信息,影响最终分类结 果。而特征选择在于根据某种依据,从原始波段集中 选择具有分类价值的波段子集,以作为尽可能保留分 类信息的新特征。因此,在不进行特征变换降维的前 提下,通过对原始高光谱图像的波段进行特征选择是 很有必要的。现在常见的特征选择即波段选择方法有 最佳指数因子法 (optimal index factor, OIF)^[6]、协 方差矩阵行列式法^[7]和自适应波段选择法^[8]等。这些 方法都取得了不错的效果,仍然存在明显缺陷如最佳 指数法选择波段运算量大且获得的波段相关性较高, 以及自适应波段选择算法容易出现波段连续问题等。

近年来像 CNN^[9],生成对抗网络(generative adversarial nets, GAN)^[10]等深度学习模型被更多地用在高光谱图像分类任务当中从而进一步挖掘高光 谱遥感图像的潜在分类信息。与传统 K-means^[11],支 持向量机^[12],核方法^[13],统计算法如逻辑回归^[14]和 贝叶斯模型^[15]等方法相比,深度学习在处理高维数 据时具有更强的抽象能力,通过深层神经网络挖掘数 据中的深层特征和隐含信息,具有更好的特征提取效 果。但由于 HSIs 数据的特殊性,普通的深度学习模 型并不能很好地直接应用在上面。

为减少高光谱遥感图像中的冗余信息以及进一步挖掘潜在的分类信息,本文提出了一种结合特征重要性评估的三维卷积神经网络分类模型,基于两个公开数据集印第安松(Indian Pines)和肯尼迪航天中心(Kennedy Space Center, KSC)数据分别进行实验,并与经典的分类方法(SVM, CNN)进行了比较分析。

1 分类模型

卷积神经网络是一种非常流行的图像分类方法。 由于卷积层滤波器的数量与输入通道的数量成正比, 所以将卷积神经网络运用到高光谱图像相关任务上 时,一种常用方法是预先降低光谱维数,以减少高光 谱图像与普通图像之间的差异。为降低光谱维数并更 好地适应卷积神经网络模型,本文提出了一种基于特 征重要性的高光谱遥感图像的分类方法,算法流程如 图1所示。



图 1 基于特征重要性的高光谱图像分类

Fig.1 Hyperspectral image classification based on feature importance

1.1 特征重要性评估

本文利用随机森林进行特征重要性评估,然后进 行特征选择。随机森林是一种以决策树为基础学习器 的集成学习方法,最早由 Breiman^[16]和 Cutler 等人^[17] 提出。现有研究表明,随机森林对异常值和噪声具有 比较大的容忍度,不容易出现过拟合的现象,而且预 测准确度和稳定性更高^[18]。

而特征重要性评估的主要思想是依据某个特征 在随机森林中每棵树上的贡献值,具体到高光谱图像 处理的任务上可以看作某个波段的贡献值是多少,然 后取平均值,最后将所有波段的贡献值进行比较得出 其特征重要性评估结果。贡献值通常使用基尼指数或 者袋外数据 (out of bag, OOB)来作为评估指标。本 次研究主要使用基尼指数来进行评估。设现有 m 个 特征 Q₁, Q₂, …, Q_j, 要计算出每个特征 Q_j的基尼 指数评分即第 j 个特征在 RF 所有决策树中节点分裂 不纯度的平均改变量。其中 Gini index 指数的计算如 公式(1)所示:

 $GI_m = \sum_{k=1}^{|K|} \sum_{k'\neq k} PmkPmk' = 1 - \sum_{k=1}^{|K|} P^2mk$ (1) 式中: GI 表示基尼指数; K 表示类别数; k 表示有 k 个类别; P 表示概率; k'表示某个类别; Pmk 表示节 点中类别 k 所占的比例,即从节点中随机抽取两个样 本,其类别标记不一致的概率。

特征 Q_j 在节点 m 的重要性, 即节点 m 分支前后的 Gini 指数变化量为:

$$VIM_{im}^{(Gini)} = GI_m - GI_r - GI_c$$
(2)

式中: VIM 为变量重要性评分(variable importance measures), GI_r和 GI_c分别表示分裂后两个新节点的 Gini 指数。如果,特征 Q_j 在决策树 i 中出现的节点 在集合 M 中,那么特征 Q_i 在第 i 颗树的重要性为:

$$VIM_{ij}^{Gini} = \sum_{m \in M} VIM_{jm}^{(Gini)}$$
(3)

假设 RF 中共有 *n* 颗树, 然后把所有求得的重要 性评分归一化处理得公式(4):

$$VIM_{j} = \frac{VIM_{j}}{\sum_{i=1}^{u} VIM_{i}}$$
(4)

式中: u 为树的总数。

针对高光谱图像的特性,本文利用随机森林构建 了一个特征重要性评估模型即波段重要性评估模型, 如图 2 所示。



图 2 特征重要性评估模型



将高光谱数据输入模型后,使用贝叶斯优化选择参数。然后通过训练完成模型中所有决策树汇总过来的信息即可得该高光谱数据集中每个波段的特征重要性信息,根据这些可以在最大程度上保存数据集的原始信息从而选择数量合适的波段组合成新的数据集。

1.2 三维卷积神经网络

随着研究的不断深入,为了获取更令人满意的高 维数据处理结果,一种很有前景的方法被提出来,直 接用 3D-CNN^[19]处理高光谱数据,即用三维卷积同 时处理同样是三维的高光谱图像。与其他 CNN 模型 不同的是,3D-CNN 是直接生成三维特征立方体,它 能更好地提取空间信息。在高光谱分类任务上,其有 了比 1D-CNN 和 2D-CNN 更好的表现^[20]。

如图 3 所示,本文构建的 3D-CNN 模型主要由 2 个卷积层(*C*₁, *C*₂)、2 个池化层(*P*₁, *P*₂)以及 2 个全连接层(*F*₁, *F*₂)组成。首先在卷积输出中,应 用了 ReLU 函数,以从数据中学习更复杂的特征。然 后是池化层对卷积层处理过后的图像进行下采样,本 模型中使用最大池的主要目的是在一个邻域中聚合 多个低级特征以获得局部不变性。此外,它还通过减 小网络的输出规模,降低网络在训练和测试阶段的计 算成本。再当数据从 *P*₂输出时,通过 ReLU 函数在 全连接层*F*₁输出 120 个神经元。*F*₂经过相同的步骤, 输出数量与分类结果相关的神经元,最后通过 softmax 得到分类结果。

2 实验结果与分析

2.1 实验数据

为了验证所提基于特征重要性的高光谱图像的分 类效果,本文采用两个公开的高光谱遥感数据集进行 实验。数据一是从 AVIRIS 传感器中获取的 Indian Pines 影像。拍摄地点在印第安纳州西北部。该影像 大小为 145×145 像素, 波长范围为 0.4~2.5 µm 的 220个波段,共包含农场、森林、高速路、铁路以及 低密度的房屋等 16 个类别。数据二为 1996 年 3 月美 国宇航局在佛罗里达州肯尼迪航天中心(KSC)所获 取数据。KSC 数据从离地面大约 20 km 的高度获得, 空间分辨率为18m。去除吸水率和低信噪比波段后, 对其中的 176 个波段进行研究分析。该影像共包括 13个类别,代表该环境中出现的各种土地覆盖类型。 两个数据集的具体地物类别如图 4 所示: 图 4(a)为 Indian Pines 假彩色合成图像;图 4(b)为 Indian Pines 真实地物类别;图4(c)为KSC假彩色合成图像;4(d) 为 KSC 图像真实地物类别。





(a)Indian Pines 真实图像 (a) Indian Pines real image



(c) KSC 假彩色合成图像 (c) KSC false color composite image



(b) Indian Pines 真实地物类别 (b) Indian Pines real feature category



(d) KSC 真实地物类别 (d) KSC real feature category

图 4 研究所用数据集及其真实地物类别 Fig.4 Research data set and its real feature category

2.2 实验方法

首先,分别对两个数据集的光谱特征进行重要 性评估。图 5 为用所提评估模型得出的两个数据集 波段重要性结果,其中图 5(a)表示 Indian Pines 数据, 图 5(b)表示 KSC 数据。两幅图分别给出了两个数据 中特征重要性前 20 的波段。在数据输入模型前,我 们将其划分为占数据总量 70%的训练集以及占总量 30%的测试集。通过贝叶斯优化法,找出每个参数 的最优值,最终将几个重要的参数确定如下所示(以 Indian Pines 为例): 弱学习器的最大迭代次数 (n estimators),或者说最大的弱学习器个数为 193; 最大特征数(max_features)选择默认数值,即划分 时考虑所有的特征数。内部节点再划分所需最小样 本数(min_samples_split)设置为2,为限制子树继 续划分的条件。如果一个节点的样例小于2,则不 继续尝试选择最优特征来划分。以Indian Pines 数据 集为例:模型第一次输出后,剔除排名靠后的10% 数量的波段,然后向模型中重新输入剩于波段以获 得新排序结果。重复上述步骤,可以最大限度地保 证评估结果的准确性。根据评估结果和实验需求, 可以组合得到波段数量不同的新数据集。



Fig.5 Two datasets of importance for the top 20 bands

然后,将所得数据分为占总量 40%和 60%的训 练集和测试集后输入 3D-CNN 分类模型。经过贝叶 斯优化反复实验,将模型的主要参数设置如下(以 KSC 为例):模型的学习率为 0.01,每一批处理数 据大小(batch size)为100,而所有数据完成的训 练总次数 (epoch) 设置为 500。在选择 3D-CNN 作 为主要分类器的同时,采用了多种不同经典的高光 谱图像分类方法作为实验参考。

2.3 结果与分析

图 6 为本文所提特征选择与 3D-CNN 结合的方 法以及 SVM 模型在 Indian Pines 和 KSC 数据上的





105 144 Number of bands

分类精度,其中图 6(a)为 Indian Pines 分类精度,图 6(b)为 KSC 数据分类精度。图 7、图 8 为在两个数

据集上的分类效果图。选择 SVM 作为主要对比方

法的原因在于,作为经典机器学习模型 SVM 可用

于线性或非线性分类,具有良好的学习能力,可解

3D-CNN 还是 SVM 作为分类器,精度都比在该方

法结合特征选择后得出的结果低,甚至只需要选择

其中的40个波段就比完整选择的效果要好。

0.96

• 0.84

0.92

▲ 0.83

40

不难看出,当选择高光谱数据全部波段(Indian Pines 为 220 个, KSC 为 176 个) 进行实验, 无论

0.97

0.81

— FI+SVM

0.95

0.76

176

决小样本情况下的问题,错误率低。

图 6 不同方法在两个数据集上的分类精度 Fig.6 Classification accuracy of different methods on two datasets

1

0.95

0.9

0.85

0.8 0.8 0.7 0.8 0.7 0.7

0.7

0.65

0.6



(a) SVM



(b) FI -SVM



(c) 3D CNN



(d) FI-3D CNN

图 7 不同方法分类效果图(Indian Pines) Fig.7 Renderings of different methods of classification(Indian Pines)



(a) SVM





Fig.8 Renderings of different methods of classification(KSC)



(d) FI-3D CNN

第42卷 第12期	红外技术	Vol.42 No.12
2020年12月	Infrared Technology	Dec. 2020

另外,可以发现当选择波段较少的时候,SVM 的分类精度持续上升,但是随着波段数的增加,在 波段数分别为110和105时达到了峰值后呈下降趋 势。主要是因为SVM在小样本训练集上能够得到 比其他传统算法好的结果。但是SVM的空间消耗 主要是存储训练样本和核矩阵,借助二次规划来求 解支持向量,而求解二次规划将涉及L阶矩阵的计 算(L为样本的个数),当L数目很大时该矩阵的存 储和计算将耗费大量的机器内存和运算时间。随着 波段数的增加,SVM模型对样本愈加难以实施,所 以导致精度下降。而达到峰值时虽有83%和84%的 精度,仍然要远低于3DCNN得出的结果。这是由 于经典支持向量机模型只给出了二分类的算法,对 于高光谱数据则一般都是通过多个两类支持向量机 的组合来解决多分类的问题。

而 CNN 模型已经在很多实验中证明了其在高 光谱图像分类这一任务上的优越性,本文中使用的 3D-CNN 模型更是如此。因为 3D 卷积核的共享, 对高维数据的处理更加没有压力,结合特征选择方 法分别在波段数为 163 和 144 时充分地挖掘了数据 信息,地物错分和漏分也明显减少,取得了实验过 程中最好的分类结果。

表 1 为本次研究中采用的主要分类方法 FI-3DCNN和多个经典模型在Indian Pines和KSC数 据集上的总体分类精度(overall accuracy, OA)及其 标准差(standard deviation,SD)和Kappa系数。经 定量分析容易发现,整体而言各类模型在KSC数据 集上的表现要好一些。而对于每一个数据集,本文 所提FI-3D CNN模型都最好地完成了分类任务,精 度最高分别达到了98.04%、97.26%左右,Kappa系 数也达到了97.84 和 96.98,再次证明了重要性评估 特征选择方法的有效性。将最佳指数因子法(optimum index factor, OIF)与(1D, 3D)卷积神经网络分别结合,在两个数据集中选取了不同数量的波段对原数据集进行降维,然后利用卷积神经网络模型进行分类。从表1中的最终实验结果可以看出,OIF方法明显是有效的,OIF-CNN的精度在 Indian Pines 和 KSC 数据集上分别比 CNN 要高 1.98%和 1.12%,而 OIF-3D CNN 相较于 3D CNN 则提升了大概 1.8%和 1.64%,Kappa 系数分别提高了 0.87 和 2.8,但却比 FI-3D CNN 相较于 3D CNN 的精度提升(1.99%, 1.77%)还是要低一些。这一结果也让本文提出利用特征选择波段的方法更有说服力。

另外,可以看出相较于 3D CNN,特征选择方 法对 1D CNN 的提升要多一些,其中精度分别提升 了 2.72%、2.25%,而 Kappa 系数分别提高了 2.87 和 3.1。这主要是因为 3D CNN 对于高维数据任务 的适应性很强,已经取得了比较好的结果,所以波 段选择方法对其提升没有那么显著。但就整体而言 其分类精度还是远不及 FI-3D CNN 模型。

3 结论

本文提出了一种新的基于特征重要性的高光谱 分类模型(FI-3D CNN)。该模型主要分为两个部分: 基于特征重要性的波段选择和三维卷积神经网络分 类模型。利用特征重要性进行波段选择的优点在于 它在最大限度地保持数据原始性的同时又降低了数 据的维度,减少了数据的冗余。而作为分类器的 3D 卷积神经网络模型其特有的细粒度特征提取方式, 使得它对图像的处理达到了几近人力的水平。实验 结果表明,本文提出的特征选择方法结合 3D CNN 后取得了比较好的分类结果,而且该方法也能够有

表 1 不同方法在 Indian Pines 和 KSC 上分类精度 Table 1 Classification accuracy of different methods in Indian Pines and KSC

·						
Classification	Indian Pines		KSC			
	OA±SD/(%)	Kappa	OA±SD(%)	Kappa		
SVM	75.86 ± 0.38	79.86	76.34±0.62	80.59		
1D CNN	82.57±0.88	83.31	85.50±0.73	86.18		
3D CNN	96.05±0.25	95.25	95.49±0.83	95.23		
OIF-CNN	84.55±0.33	81.92	86.62±0.36	88.36		
OIF-3D CNN	97.86±0.56	96.12	97.13±0.15	98.03		
FI-SVM	82.34±0.29	86.08	85.38±0.23	88.01		
FI-CNN	85.29±0.13	86.18	87.75±0.12	89.28		
FI-3D CNN	98.03±0.16	97.84	97.26±0.13	96.98		

效地提高经典模型(SVM、CNN等)的分类精度。 特征重要性的评估会直接影响到后面的分类结果, 因此如何最大程度地确保评估方法准确性将是以后 工作的重点。

参考文献:

- YE M C, JI C X, CHEN H, et al. Residual deep PCA-based feature extraction for hyperspectral image[J/OL]. *Neural Computing & Applications*, 2020, 32(7): doi:10.1007/s00521-019-04503-3.
- [2] Donoho D L. High-dimensional data analysis: the curses and blessings of dimensionality[J]. AMS Math Challenges Lecture, 2000, 1: 32.
- [3] Marpu G, Chanussot P R J, Benediktsson J A. Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles[J]. *IEEE Geosci. Remote Sens. Lett.*, 2012, 9(3): 447-451.
- [4] Villa A, Benediktsson J A, Chanussot J, et al. Hyperspectral image classification with independent component discriminant analysis[J]. *IEEE Trans. Geosci. Remote Sens.*, 2011, **49**(12): 4865-4876.
- [5] Zabalza J, REN J, WANG Z, et al. Singular spectrum analysis for effective feature extraction in hyperspectral imaging[J]. *IEEE Geosci. Remote Sens. Lett.*, 2014, 11(11): 1886-1890.
- [6] Chacvez P S, Berlin G L, Sowers L B. Statistical method for selecting landsat MSS retio[J]. *Journal of Applied Photographic Engineering*, 1982, 1(8): 23-30.
- [7] Charles S. Selecting band combination from multispectral data[J]. *Photogrammetric Engineering and Remote Sensing*, 1985, 51(6): 681-687.
- [8] 张爱武,杜楠,康孝岩,等.非线性变换和信息相邻相关的高光谱自适应波段选择[J]. 红外与激光工程,2017,46(5):05308001. ZHANG Aiwu, DU Nan, KANG Xiaoyan, et al. Adaptive band selection for nonlinear transform and information adjacent correlation[J]. Infrared and Laser Engineering, 2017, 46(5): 05308001.
- [9] HU W, HUANG Y, WEI L, et al. Deep convolutional neural networks

for hyperspectral image classification[J]. Journal of Sensors, 2015, 2015; e258619.

- [10] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014, 2: 2672-2680.
- [11] Haut J, Paoletti M, Plaza J, et al. Cloud implementation of the K-means algorithm for hyperspectral image analysis[J]. J. Supercomput., 2017, 73(1): 514-529.
- [12] Melgani F, Lorenzo B. Classification of hyperspectral remote sensing images with support vector machines[J]. *IEEE Trans. Geosci. Remote Sens.*, 2004, **42**(8): 1778-1790.
- [13] Camps-Valls G, Bruzzone L. Kernel-based methods for hyperspectral image classification[J]. *IEEE Trans. Geosci. Remote Sens.*, 2004, 43(6): 1351-1362.
- [14] Haut J, Paoletti M, Paz-Gallardo A, et al. Cloud implementation of logistic regression for hyperspectral image classification[C]//Comput. Math. Methods Sci. Eng. (CMMSE), 2017, 3: 1063-2321.
- [15] Bazi Y, Melgani F. Gaussian process approach to remote sensing image classification[J]. *IEEE Trans. Geosci. Remote Sens.*, 2010, 48(1): 186-197.
- [16] Breiman L. Bagging predictors[J]. Mach Learn, 1996, 24(2): 123-140.
- [17] Cutler A, Cutler D R, Stevens J R. Random Forests[M]//Ensemble Machine Learning, Boston: Springer, 2012: 157-175.
- [18] 李贞贵. 随机森林改进的若干研究[D]. 厦门: 厦门大学, 2013.
 LI Z G. Several Research on Random Forest Improvement[D]. Xiamen: Xiamen University, 2013.
- [19] LI Y, ZHANG H, SHEN Q. Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network[J]. *Remote Sensing*, 2017, 9(1): 67.
- [20] CHEN Y, JIANG H, LI C X, et al. Deep features extraction and classification of hyperspectral images based on convolutional neural networks[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(10): 6232-6251.

更正启事

本刊于 2020 年 11 期刊登的作者名为"池林辉、钱芸生、籍宇豪",题名为"一种提高 FPGA 片间通信 稳定性的校验协议研究"一文中,因失误将位于第 1024 页的公式(2)式体" $V(t) = A_0 \sin\left[\frac{2\pi}{T_0}\left(t + \frac{\varphi(t)}{2\pi f_0}\right)\right]$ "遗漏了,特此更正,并为此向作者及读者致以诚挚的歉意。

《红外技术》编辑部 2020年12月