

孔嘉鑫, 张昭臣, 张健. 基于多源遥感数据的植物物种分类与识别: 研究进展与展望. 生物多样性, 2019, 27(7), 796–812.
<http://www.biodiversity-science.net/CN/10.17520/biods.2019197>

附录 2 常用物种分类算法的原理及优缺点 (基于 Fassnacht et al, 2016 补充)

Appendix 2 Principles, advantages and disadvantages of commonly used species classification algorithms (Adjusted from Fassnacht et al, 2016)

分类算法	原理	优点	缺点
Species classification algorithms	Principles	Advantages	Disadvantages
支持向量机 Support Vector Machine, SVM	一种非参数监督分类器, 通过对训练样本寻找一个最优决策超平面高维空间, 生成最佳的类分离(Gualtieri & Crompt, 1999)。	无需分布假设; 适于将非遥感变量纳入分类中; 所需训练数据相对较少; 易于获取概率值(Burges, 1998)。	需对多类支持向量机进行优化; 训练计算成本、算法复杂度较高; 应用核函数时参数选择复杂(Zhu & Blumberg, 2002; Mountrakis et al, 2011)。
随机森林 Random Forest, RF	一种非参数算法, 属于 Bagging 算法的一种, 组合多个弱分类器。生成每棵树时所选特征都随机产生, 不需额外剪枝, 通过计算使整体模型具有较高的精确度和泛化性(Chan & Paelinckx, 2008)。	无需分布假设; 对输入变量个数不敏感; 处理高维数据且无需特征选择; 较好的泛化和抗过拟合能力; 直观推导精度和变量重要性; 训练速度快(Ham et al, 2005)。	对噪声数据可能会过度拟合(Richter et al, 2016); 响应类别具有不同数量级时可能存在偏差; 模型自身的偏差。
最大似然分类 Maximum Likelihood Classifiers, MLC	一种监督分类器, 根据先验知识及贝叶斯理论, 选择感兴趣区并计算分类函数, 样本归属于函数概率值最大的类别(Settle & Briggs, 1987)。	简单, 方便实施; 有适于各种估计问题的通用方法; 样本容量较大时几乎无偏差(Richards, 2013)。	假设训练数据服从高斯分布; 对小样本有偏差; 对输入变量的个数敏感; 分类耗时较长; 训练样本数量需大于波段数(Murthy et al, 2003)。
判别式分析 Discriminant Analysis, DA	按照一定标准建立判别函数, 根据研究对象特征值确定判别系数, 进而划分样本的类别归属(Hastie & Tibshirani, 1996)。	无需调整参数, 计算效率较高; 支持多个输入变量; 易于解释类别间的差异; 对模式的归一化或随机化不敏感。	假设训练数据服从高斯分布; 经典判别式分析对不适宜问题和异常值不敏感; 处理多重共线性的能力有限(Suratno et al, 2009)。
k-最近邻分类 k-Nearest Neighbor, KNN	统计某样本在特征空间中 k 个最相邻样本的类别, 该样本被归为有最多样本数的类别并具有该类样本的特性(Cover & Hart, 1967)。	简单, 方便理解、实施; 无需估计参数, 无需训练; 适于对稀有事件进行分类; 适于多分类问题 (Zhu & Basir, 2005)。	受类别样本数的影响较大; 计算量较大; k 参数的选择直接影响结果(Wu et al, 2008)。
人工神经网络 Artificial Neural Networks,	根据大量已知类别的样本建立训练集, 对样本的各种信息进行处理分析, 借助卷积神经网络, 对待分类的	从数据池中提取模式并识别趋势; 处理数学上难以定义的分类问题; 自学习能力较强;	难于训练, 最终结果取决于初始参数; 暗箱操作, 训练过程不透明; 体系结构的通用性较差。

孔嘉鑫, 张昭臣, 张健. 基于多源遥感数据的植物物种分类与识别: 研究进展与展望. 生物多样性, 2019, 27(7), 796–812.
<http://www.biodiversity-science.net/CN/10.17520/biods.2019197>

分类算法	原理	优点	缺点
Species classification algorithms	Principles	Advantages	Disadvantages
ANN	样本进行描述、分类和解释(Krizhevsky et al, 2012)。	高速寻找最优解; 易实施操作(Foody et al, 1995; 毛健等, 2011)。	
光谱角制图 Spectral Angle Mapper (SAM)	基于物理的一种交互式监督分类法。结合光谱维的信息, 利用 n 维角度来匹配样本光谱和参考光谱, 进而确定分类类别(Yuhas et al, 1992)。	无需分布假设; 适于已知足够光谱信息的情况; 对光照不敏感 (Sohn & Rebello, 2002; Mohajane et al, 2017)。	当空间工作有问题时不考虑亚像素值(Fassnacht et al, 2016)。
广义线性模型 Generalized Linear Model (GLM)	通过具体分布参数和指数分布参数间的关系建立模型, 利用极大似然估计求得损失函数, 再用梯度下降优化参数得到最优解。	适于二进制响应变量的建模; 使用二项分布进行分类。	响应变量必须独立; 为避免维度诅咒, 要进行变量选择 (Fassnacht et al, 2016)。
分类和回归树 Classification and regression tree (CART)	基于训练数据集生成决策二叉树, 递归地二分每个特征, 用验证数据集对已生成的树进行剪枝, 并选择最优子树用于分类(张松林, 1997)。	规则易于理解、使用、说明和解释; 可以显示重要字段; 计算量相对不大。	连续字段较难预测; 对有时间顺序的数据需要预处理; 分类时大多根据一个字段进行(Loh, 2011)。
贝叶斯分类算法 Bayesian Classifiers	通过某样本的先验概率, 利用贝叶斯公式计算出其后验概率, 该样本归属于具有最大后验概率的类别 (Williams & Barber, 1998)。	数据的先验知识可合并; 输出结果可解释; 平均风险最小; 方法简单、速度快(Wu et al, 2008)。	先验知识选择的复杂性; 后验结果受先验选择的影响较大; 计算成本高(Fassnacht et al, 2016)。

孔嘉鑫, 张昭臣, 张健. 基于多源遥感数据的植物物种分类与识别: 研究进展与展望. 生物多样性, 2019, 27(7), 796–812.
<http://www.biodiversity-science.net/CN/10.17520/biods.2019197>

附录 2 参考文献

- Burges CJC (1998) A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2, 121–167.
- Chan JCW, Paelinckx D (2008) Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, 112, 2999–3011.
- Cover TM, Hart PE (1967) Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13, 21–27.
- Fassnacht FE, Latifi H, Stereńczak K, Modzelewska A, Lefsky M, Waser LT, Straub C, Ghosh A (2016) Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186, 64–87.
- Foody GM, McCulloch MB, Yates WB (1995) Classification of remotely sensed data by an artificial neural network, issues related to training data characteristics. *Photogrammetric Engineering and Remote Sensing*, 61, 391–401.
- Gualtieri JA, Crompton RF (1999) Support vector machines for hyperspectral remote sensing classification. In: *The 27th AIPR Workshop: Advances in Computer-Assisted Recognition*. Washington, DC, United States, pp. 221–232.
- Ham J, Chen YC, Crawford MM, Ghosh J (2005) Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43, 492–501.
- Hastie T, Tibshirani R (1996) Discriminant analysis by Gaussian mixtures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 155–176.
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 141, 1097–1105.
- Loh W (2011) Classification and regression trees. *Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery*, 1, 14–23.
- Mao J, Zhao HD, Yao QQ (2011) Application and prospect of artificial neural network. *Electronic Design Engineering*, 19(24), 62–65. (in Chinese with English abstract) [毛健, 赵红东, 姚婧婧. 人工神经网络的发展及应用 (2011) 电子设计工程, 19(24), 62–65.]
- Mohajane M, Essahlaoui A, Oudija F, El Hafyani M (2017) Mapping forest species in the Central Middle Atlas of Morocco (Azrou Forest) through remote sensing techniques. *ISPRS International Journal of Geo-Information*, 6, 275.
- Mountrakis G, Im J, Ogole C (2011) Support vector machines in remote sensing, a review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 247–259.
- Murthy CS, Raju PV, Badrinath KVS (2003) Classification of wheat crop with multi-temporal images, performance of maximum likelihood and artificial neural networks. *International Journal of Remote Sensing*, 24, 4871–4890.
- Richards JA (2013) Supervised classification techniques. In: *Remote Sensing Digital Image Analysis: An Introduction*, 5th edn. (ed. Richards JA), pp. 247–318. Springer, Berlin, Heidelberg.
- Richter R, Reu B, Wirth C, Doktor D, Vohland M (2016) The use of airborne hyperspectral data for tree species classification in a species-rich Central European forest area. *International Journal of Applied Earth Observation and Geoinformation*, 52, 464–474.
- Settle JJ, Briggs SA (1987) Fast maximum-likelihood classification of remotely sensed imagery. *International Journal of Remote Sensing*, 8, 723–734.

- 孔嘉鑫, 张昭臣, 张健. 基于多源遥感数据的植物物种分类与识别: 研究进展与展望. 生物多样性, 2019, 27(7), 796–812.
<http://www.biodiversity-science.net/CN/10.17520/biods.2019197>
- Sohn Y, Rebello NS (2002) Supervised and unsupervised spectral angle classifiers. *Photogrammetric Engineering and Remote Sensing*, 68, 1271–1282.
- Suratno A, Seielstad C, Queen L (2009) Tree species identification in mixed coniferous forest using airborne laser scanning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 683–693.
- Williams CKI, Barber D (1998) Bayesian classification with Gaussian processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 1342–1351.
- Wu XD, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou ZH, Steinbach M, Hand DJ, Steinberg D (2008) Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14, 1–37.
- Yuhans RH, Goetz FHA, Boardmann JW (1992) Discrimination among semiarid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm. In: *Summaries of the Third Annual JPL Airborne Geoscience Workshop* (ed. Jet Propulsion Laboratory), pp. 147-149. Jet Propulsion Laboratory Publication, Pasadena, CA.
- Zhang SL (1997) An introduction to the methodology of CART- classification and regression trees. *Volcanology & Mineral Resources*, 18(1), 67–75. (in Chinese with English abstract) [张松林 (1997) CART-分类与回归树方法介绍. *火山地质与矿产*, 18(1), 67–75.]
- Zhu GB, Blumberg DG (2002) Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel. *Remote Sensing of Environment*, 80, 233–240.
- Zhu HW, Basir O (2005) An adaptive fuzzy evidential nearest neighbor formulation for classifying remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 43, 1874–1889.