# 基于异源多时相遥感数据决策树的作物种植 面积提取研究

# 张旭东,迟道才\*

(沈阳农业大学水利学院,沈阳 110161)

摘要:为了充分发挥不同遥感数据源的优点,采用中高分辨率的 TM 影像覆盖区为典型试验区,由试验区的水稻特征建立基于专家知识的分类规则,结合 C4.5 算法构建分类决策树,然后综合运用 eMODIS 时间序列 NDVI、MOD12Q1、DEM 及其派生的坡度数据进行决策树分类,以辽宁省为例对 2009 年的水稻种植面积进行了提取。结果表明:该方法方便适用,精度高于传统的监督分类、极大似然分类法; 与统计数据相比,结果相对误差为 5.04%,各地级市的遥感结果与同期的统计数据吻合较好,相关系数达到 0.96,能较好地反映辽宁省水稻种植的空间分布状况。

关键词:遥感;归一化植被指数;水稻提取;决策树分类 DOI:10.3969/j.issn.1000-1700.2014.04.013 中图分类号:S127 文献标识码:A

文章编号:1000-1700(2014)04-0451-06

# Mapping Crop Fields by Using Multi-sensor and Multi-temporal Remote Sensing Data with Decision-tree

## ZHANG Xu-dong, CHI Dao-cai\*

(College of Water Conservancy, Shenyang Agricultural University, Shenyang 110161, China)

Abstract: In order to take the advantages of different remote sensing data, we developed a new method to construct the decision tree which was derived from experimental area of rice characteristic to establish the classification rules based on expert knowledge and combined with C4.5 algorithm. The decision tree the high resolution of Landsat thematic mapper (TM) images were adopted as the typical experimental area and eMODIS time-series normalized difference vegetation index (NDVI), MOD12Q1, digital elevation model (DEM). The gradient data were used to extracted the paddy field of Liaoning province in 2009. The results showed that this method was convenient and applicable and the result accuracy was higher than the traditional supervised classification. The province's relative error was 5.04% compared with the statistical data. The remote sensing results of the prefecture-level city were well agreement with the statistical data and the correlation coefficient was 0.96. The extraction results indicated that the method can accurately reflect the space distribution of rice in Liaoning province. Key words: remote sensing; NDVI; paddy fields extraction; decision-tree classification

辽宁省是我国 13 个粮食主产区之一。为了保证国家的粮食安全战略,作物种植面积监控是非常重要的。 与常规的统计方法相比,应用卫星遥感(RS)方法对水稻面积进行估算具有宏观性、综合性、时效性和动态性的 独特优势,其成本相对较低,而结果丰富,不仅可以提取农作物种植面积,而且可实现空间分布的准确定位。但 由于卫星影像的来源非常多,其扫描宽度、空间分辨率和时间分辨率各不相同,研究者根据研究的对象、区域、 目的的不同,发展出了众多的解译和分类方法<sup>[1]</sup>。2004 年以前,NOAA/AVHRR 数据在水稻估产和种植面积的提 取上采用的较多,后来 MODIS(the moderate resolution imaging spectroradiometer)数据几乎取代了 NOAA/AVHRR 数据。程乾等<sup>[2]</sup> 根据南方丘陵山区复杂地形区域水稻的种植易受坡度影响的特性,从数字高程模型(digital elevation models,DEM)中提取坡度信息,采用 DEM 产生的坡度和两个时相 MODIS 影像数据对浙江省水稻种植 面积进行了提取;焦险峰等<sup>[3]</sup>设计了基于分层抽样的中国水稻种植面积遥感监测方法;SAKAMOTO等<sup>[4]</sup>利用 MODIS 数据研究了渭公河三角洲水稻物候的时空分布情况和农作系统;XIAO等<sup>[5-6]</sup>利用 MODIS 的 MOD09A1 产品的 NDVI(normalized difference vegetation index) LSWI 和(enhanced vegetation index)EVI 对中国南部 13 个省

收稿日期:2014-04-03

基金项目:高等学校博士学科点专项科研基金联合资助项目(20112103110007)

作者简介:张旭东(1979-),男,沈阳农业大学讲师,博士研究生,从事灌溉排水及农业节水研究。\*通讯作者 Corresponding author:迟道才(1964-), 男,沈阳农业大学教授,博士,从事灌溉排水及农业节水研究。

和东南亚 13 个国家的水田面积进行提取;孙华生<sup>[1]</sup>利用 MODIS 数据,利用傅里叶低通滤波和小波低通滤波平 滑时间序列 EVI,对我国 2000~2007 年全国的水稻面积进行了提取,得到了全国单季稻、早稻、和晚稻的空间分 布状况。另外,基于 TM(thematic mapper)影像对水稻种植面积的提取,目前也已经取得了丰富的成果。王福民等 <sup>[7]</sup>利用水稻两个不同时期 TM 影像进行穗帽变换,然后根据湿度变量对水稻种植区影像进行分类;吴黎等<sup>[8]</sup>基于 TM 影像采用实地感兴趣区与分类结果建立混淆矩阵,选取满足检验分类精度可行性的端元,比较了波谱角分 类法、最大似然分类法和混合像元线性分解法对沈阳市水稻种植面积中的提取。蔡学良等<sup>[9]</sup>以湖北漳河灌区为 例,将 Landsat ETM+与 MODIS NDVI 数据融合,采用传统的非监督分类提取植被指数变化信息来提取灌区尺 度作物种植结构;邬明权等<sup>[10-11]</sup>提出了两种方法,一是利用样区多时相高分辨率 HJ-1CCD 数据获取水稻移栽 期和生长期的特征,再利用 MODIS 日反射率产品来提取;二是利用高分辨率的 Landsat-ETM+影像和 MODIS 影像融合,再采用传统的分类方法如光谱角分类进行提取。

综上,目前水稻种植面积的提取主要为 MODIS 或 Landsat 系列数据,在方法上经历了从单一时相到两时 相再到时间序列,从单源应用到多源融合应用再到结合水稻生育生长特征及其他信息的发展;在精度上 TM 的 精度或多源融合的要高于 MODIS 的或单源的精度。为了充分发挥不同遥感片源的优点,本研究以辽宁省 2009 年水稻种植面积的提取为例,提出了一种不需数据融合,采用部分 TM 覆盖区为试验区,根据其分类结果挖掘 专家知识规则,并结合 C4.5 算法构建的决策树进行分类方法,为基于异源多时相遥感数据的作物种植面积的 提取提供了一条新的思路。

# 1 材料与方法

#### 1.1 数据准备

1.1.1 TM 数据 选取辽宁省盘锦市 TM5 影像两幅,path 为 120,row 为 31 和 32,日期为 2009 年第 196 天。首先 通过 ENVI 的 Layer Stacking 将各幅 TM 的 7 个波段合成为一个文件;然后将两幅影像通过基于地理坐标的镶嵌 把两幅图像镶嵌成一张图片文件;最后将从中国地理信息基础数据网获得的盘锦市边界做成掩膜,对合成文件进 行裁剪。

1.1.2 MOD12Q1 数据 冉有华<sup>[12]</sup>从类型面积一致性、空间一致性方面对 4 种目前比较流行的全球 1km 土地 覆盖数据集在中国区域的分类精度进行了评价,发现 GLC2000 和 MODIS 土地覆盖数据有较高的整体分类精 度。MODIS 土地覆盖类型产品包括从每年 Terra 星数据中提取的土地覆盖特征,包括 5 种不同分类方案相应的 分类产品:国际地圈生物圈计划(IGBP)方案分 17 类;马里兰大学(UMD)植被方案分 14 类;MODIS 提取叶面积 指数/光合有效辐射分量(LAI/fPAR)方案分 9 类;MODIS 提取净第一生产力(NPP)方案分 9 类;植被功能型(PFT) 分类方案分 12 类。MODIS Terra 数据 1km 土地覆盖类型年合成栅格数据产品包含这 5 种不同的土地覆盖分类 体系。然而,MODIS 的分类却比较粗,它的谷类作物包括水稻、玉米、小麦、大麦、高粱、粟、黍和糜子。

在 MOD12Q1 的 5 种分类方案中, 第 1,2,3,5 种有关于作物的分类, 其中, 第 1,2 种分出的是作物 (Croplands),第 3,5 种分出的是谷类作物(Cereal crops),各种分类结果的作物空间分布基本一致。但将这 4 种分 类结果与 2009 年统计年鉴的数据比较(表 1)可知,若以统计年鉴作物参照,这 4 种分类方法的结果均比统计数 据大,并且相对误差也都比较大,最小也超过 48%。

Table 1 Comparison of crop classification results between wroddi2Q1 and Statistical Tearbook in 2009				
方法 Method	IGBP (类型 1 Type 1)	UMD (类型 2 Type 2)	LAI/fPAR (类型 3 Type 3)	PFT(类型 5 Type 5)
遥感 Remote sensing/×10 <sup>4</sup> hm <sup>2</sup>	814.7	603.1	565.3	818.4
统计 Statistic/×10 <sup>4</sup> hm <sup>2</sup>	406.5	406.5	281.2	281.2
相对误差 Relative error/%	100.4	48.4	101.0	191.0

表 1 2009 年 MOD12Q1 作物分类结果与统计年鉴对比结果

本研究将盘锦市疙瘩楼水库附近基于 MOD12Q1 分类结果和基于 TM 的分类的结果进行比较(图 1)可知, 大多数 TM 分类的水稻基本上叠加在 MOD12Q1 分类的谷类作物之上,但会有线性或不规则的小方块露出,表 明由于 MOD12Q1 本身像素为 1000m×1000m 的原因,将位于稻田之中的较小的道路、田埂、灌排渠系及工程、 部分居民户均划分为了作物面积。这也是 MOD12Q1 比统计年鉴数据大的主要原因。因此,MOD12Q1 的作物分 类结果在对辽宁来说误差是比较大的,但其对作物空间分布 具有一定的参考价值。在研究辽宁水稻面积提取时,为了减 少工作量,将 MODIS 的 MOD1201 产品作为数据源之一。 1.1.3 eMODIS 的 NDVI 数据 NDVI 即为归一化植被指 数,它是遥感监测植被变化、土地利用类型、地表能量分配、作 物产量预测等应用中的一个非常重要的参数<sup>[13]</sup>。NDVI 可由不 同的卫星传感器获得。MODIS 的访问周期和空间分辨率非常 适合对植被的研究。然而,由于标准地图的投影、文件格式、区 间的合成、高纬度"蝴蝶结"效应和产品的延迟性等问题的存 在,MODIS 的优点大大降低。为此,美国地质调查局(USGS)地 球资源观察和科学 (EROS) 中心正式生成了一套基于 MODIS 的名为 "eMODIS (expedited moderate resolution imaging spectroradiometer)"的产品。eMODIS 在生成过程中通过了一致 性检验并对 MODIS 数据包进行取舍。eMODIS NDVI 去除了 云和填充值像素的影响,经过了空间平滑(去除了附加大气噪 音),空间分辨率为 250,500,1000m,时间上以 7d 为周期进行

滚动发布,因此质量较高。eMODIS的合成产品以地球同步轨



图 1 盘锦市疙瘩楼水库附近基于 MOD12Q1 和 基于 TM 的分类比较 Figure 1 Comparison of classification results

between the method based on MOD12Q1 and TM near the Gedalou reservoir of Panjin

道上的标准图像文件格式(Geo-TIFF)存储,以非正弦的映射网格发布,比较适合于地理区域的应用程序。本研究选取 250m 的 eMODIS NDVI 产品的时期覆盖 2009 年的水稻生育期, 儒略日分别为:2009 年第 136~145,146~156,157~166,167~176,177~186,187~196,197~206,207~217,218~227,228~237,238~248,249~258 天共 12 个时间段的平均值。

1.1.4 DEM 数据 数字高程模型采用 SRTM(shuttle radar topography mission)数据,SRTM 地形数据是迄今为止 公开的现势性最好、分辨率最高、精度最好的全球性数字地形数据之一。本研究中使用数据的分辨率为 90m,共 7 张镶嵌在一起,经投影转换并掩膜后为图 2。从数字模型上可知,辽宁省中部的辽河平原、鸭绿江入海口及沿 海地带高程较低且坡度较小,是水稻适宜生长的地区。



Figure 2 Digital elevation model of Liaoning province

### 1.2 方法

1.2.1 基于 TM 试验区的分类 本研究选取盘锦市作为水稻面积提取的试验区。先用 TM 影像对实验区的水稻进行提取,然后根据该结果分析辽宁省水稻 NDVI 的时序特征,再依据该特征对全省的水稻面积进行提取。 根据盘锦市土地利用类型,选择的训练样本类别分为水稻、湿地(主要是芦苇)、水面,居民地、未利用地、旱

-453-

田、林地7类,其中未利用地包括滩涂、裸露的河滩。样本 的要求保证一定的数量,且分布均匀和具备代表性。因为 本研究只关心水稻种植的分布和面积,其他的分类可以在 最终结果中全部划入到非水稻类,但是其他分类的样本也 要满足可分性。可分性用 Jeffries-Matusita,Transformed Divergence 这两个参数来判断,其值在 0~2.0 之间。大于 1.9 表明样本之间可分离性好,属于合格样本;小于 1.8,需 要重新选择样本;小于 1,需要将两类样本合成一类样本。 对试验区采用最大似然法进行分类,经聚类后处理盘锦市 的水稻种植分布如图 3。

由图 3 可知,盘锦市的水稻主要分布在围绕盘山县县 城、双台子区、盘锦市市区和大洼县县城周围,位于辽河、 大辽河、绕阳河、大凌河之间或两岸的平坦地区。水稻一共 有 1252003 个象元,考虑到 TM 卫星影片的分辨率是 30m×30m,因此,盘锦市 2009 年水稻的种植面积为 1.127× 10<sup>4</sup>hm<sup>2</sup>,辽宁统计年鉴<sup>[14]</sup> 2009 年的水稻种植面积为 1.085× 10<sup>4</sup>hm<sup>2</sup>,相对误差为 3.85%。由此可知,基于 TM 影像的分 析比较准确。因此,本研究采用基于 TM 影像提取得到的



图 3 2009 年盈晞巾小柏种植万布至间布向 Figure 3 Spatial patterns of rice planting in 2009 in Panjin

试验区水稻种植面积来挖掘和训练辽宁省水稻的提取规则,并将该规则应用基于异源多时相遥感数据水稻面 积的提取。

1.2.2 基于异源多时相遥感数据的辽宁水稻种植面积的提取 (1)异源多时相遥感空间数据库的构建及训练 样本的获取。采用的覆盖辽宁省的 MOD1201、eMODIS 的 12 时段的 NDVI、DEM 及其派生的坡度数据,将它们 共 15 个波段合成到一个遥感影像文件中,构建出一个包含 15 种特征信息的异源多时相的空间数据库。基于 构建的空间数据库,充分利用试验样本区的分类结果,借助滤波等图像增强工具,结合部分 TM 影像、实地调查 及灌区分布图,采用从高分辨率图中选取感兴趣区(ROI),向数据库导入(Reconcile)再合并(Merge)的办法获取训 练样本。样本共分7类,严格保证各类样本的可分性。(2)基于 C4.5 算法的决策树的构建。决策树分类算法具有 灵活、直观、清晰等特点,而且该算法的分类精度和效率高,不仅适用于数据量小的情况,还适用于数据量大的 情况,在遥感分类问题上表现出极大的优势<sup>15]</sup>。目前,已有一些算法可获取规则并构建决策树,常用的有 C4.5 算法、CART 算法、S-PLUS 算法等。其中,C4.5 算法是 QUINLAN 在 1993 年为了改进 ID3 算法提出的一种新算 法,它不仅可以直接处理离散型属性变量,还可以处理连续型属性变量,分类效率也有了很大的提高<sup>15]</sup>。本研究 采用该算法构建初步的决策树。(3)基于专家知识的规则的构建。专家知识规则的定义是将经验或知识用数学 语言表达的过程,如在辽宁地面坡度大于 3°不大可能种植水稻,海拔高于 250m 的地区也不太可能存在水稻, 因此,在提取中把这两条定义为关于地面坡度和海拔高度的不等式。另外,将基于 TM 试验区的分类结果中的 水稻做为 ROI 前述导入构建的空间数据库,可得到 NDVI 的动态变化规律(图 4),2009 年水稻插秧后到第 161 天,其 NDVI 缓慢增长,然后再到第 191 天为快速增长阶段,从第 191 天到第 232 天,水稻的 NDVI 较高,随后, 其 NDVI 值降低直至生育期结束,因此,可根据该 NDVI 变化建立相应的不等式。将这些相应的专家知识规则 (不等式)添加到 C4.5 算法构建的初步决策树中,最终构建出的决策树叶节点有 57 个,知识规则有 28 条。

## 2 结果与分析

由图 5 可知,辽宁省水稻主要分布在辽宁中部的辽河平原、丹东南部和大连东北部的沿海地区,水稻集中 的地区基本上位于水库下游已形成灌区的地方,如浑蒲灌区、开原灌区、大洼灌区、营口灌区、东港灌区等地。 此外,在水库上游的河谷两岸的低洼地区也有零星分布,但这部分水稻比例较少。ENVI 统计显示,基于遥感的 水稻面积共 5.976×10<sup>e</sup>hm<sup>2</sup>,比统计年鉴的 5.885×10<sup>e</sup>hm<sup>2</sup> 多 5.04%,辽宁省 14 个地级市的遥感结果和统计年鉴 的对比见图 6a,两者的线性相关系数非常高,*R*<sup>2</sup>=0.9633。由图 6b 误差分析可知,各地级市水稻的种植面积越



大,遥感的结果相对误差越小,因此该方法适合于大范围水稻面积的提取。对于种植面积大于 5×10<sup>5</sup>hm<sup>2</sup> 的沈阳、盘锦、铁岭、丹东、辽阳等市,遥感的相对误差小于 18%。用同样的数据源,极大似然法和监督分类法的相对 误差均大于 40%,远高于本研究提出的方法。



Figure 6 Results accuracy analysis of remote sensing

# 3 结论与讨论

本研究以 TM影像覆盖地区为试验样本区,提出了基于异源多时相遥感数据,采用 C4.5 算法和专家知识相 结合的方法构建决策树,进行作物种植面积提取的方法。以辽宁省为例的 2009 年水稻种植面积的提取结果表 明,各地级市的遥感分析结果与统计年鉴结果线性关系紧密,相关系数达到 0.96,全省的相对误差仅为 5.04%, 精度高于传统的极大似然法和监督分类法。用遥感影像提取水稻种植分布具有直观的空间分布特征,为水稻的 估产、农业灌溉用水管理、乃至为大区域其他作物种植面积的监测和分布研究提供了借鉴。

该方法数据容易获取,成本低,精度高,操作性好,充分发挥了 TM 高空间分辨率和 MODIS 高时间分辨率 的优点,避免了对不同分辨率的数据融合以及复杂的波段运算,同时也对异源多时相遥感数据的信息进行了充 分的挖掘和利用,是一种简洁有效的作物种植面积提取方法。C4.5 决策树结构思想简单,算法成熟,稳定性好, 规则的提取采用计算机完成,且易于结合专家知识,人机结合巧妙。但在应用过程中应考虑训练样本和专家知 识经验的可靠性。

#### 参考文献:

- [1] 孙华生.利用多时相 MODIS 数据提取中国水稻种植面积和长势信息[D].杭州:浙江大学,2009.
- [2] 程 乾,王人潮.数字高程模型和多时相 MODIS 数据复合的水稻种植面积遥感估算方法研究[J].农业工程学报,2005,21(5):89-92.
- [3] 焦险峰,杨邦杰,裴志远.基于分层抽样的中国水稻种植面积遥感调查方法研究[J].农业工程学报,2006,22(5):105-110.
- [4] SAKAMOTO T,VAN NGUYEN N,OHNO H,et al.Spatio temporal distribution of rice phenology and cropping systems in the Mekong Delta with special reference to the seasonal water flow of the Mekong and Bassac rivers [J].Remote Sensing of Environment,2006,100(1):1–16.
- [5] XIAO X,BOLES S,LIU J,et al.Mapping paddy rice agriculture in southern China using multi-temporal MODIS images [J]. Remote Sensing of Environment.2005,95(4):480-492.
- [6] XIAO X,BOLES S,FROLKING S,et al.Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images[J].Remote Sensing of Environment,2006,100(1):95–113.
- [7] 王福民,黄敬峰,王秀珍.基于穗帽变换的 TM 影像水稻面积提取[J].中国水稻科学,2008,22(3):297-301.
- [8] 吴 黎,张有智,解文欢,等.基于 TM 的混合分解模型提取水稻种植面积研究[J].农机化研究,2013,35(2):44-47.
- [9] 蔡学良,崔远来.基于异源多时相遥感数据提取灌区作物种植结构[J].农业工程学报,2009,25(8):124-130.
- [10] 邬明权,牛 铮,王长耀.利用遥感数据时空融合技术提取水稻种植面积[J].农业工程学报,2010,26(S2):48-52.
- [11] 邬明权,王长耀,牛 铮.利用多源时序遥感数据提取大范围水稻种植面积[J].农业工程学报,2010,26(7):240-244.
- [12] 冉有华,李 新,卢 玲.四种常用的全球 1km 土地覆盖数据中国区域的精度评价[J].冰川冻土,2009,31(3):490-500.
- [13] 张晓涛.区域蒸发蒸腾量的遥感估算[D].杨凌:西北农林科技大学,2006.
- [14] 辽宁省统计局.辽宁统计年鉴-2010[M].北京:中国统计出版社,2010.
- [15] 夏 双,阮仁宗,佘远见,等.基于 C4.5 算法的遥感影像分类[J].地理空间信息,2012,10(4):89-91.

[责任编辑 亓 国]