

全球地图第一版中国土地覆盖产品的生成

张永红¹ 张继贤¹ 郭 健¹ 曹银璇¹

(1 中国测绘科学研究院,北京市海淀区北太平路 16 号,100039)

摘 要:介绍了全球地图第一版中国土地覆盖分类产品的遥感数据源、数据处理方法及结果,详细介绍了非监督聚类 and 决策树分类相结合的区域级土地覆盖分类方法及处理步骤,并以 2003 年全年的 MODIS/TERRA 16 d 合成产品 MOD43B4 为主要遥感数据源,采用先进行影像聚类,然后以聚类结果为处理单元进行多时相多特征决策树分类,得到了覆盖全国、约 1 km 分辨率的土地覆盖分类结果。采用室内精度评定和外业验证两种方法对分类结果进行了检验,生成的中国土地覆盖分类产品的精度为 89.14%。

关键词:MODIS; 土地覆盖; 影像分类; 决策树; 非监督聚类

中图法分类号:P285.23

1 全球测图项目简介

1992 年底,日本国土交通省倡导了全球测图(global mapping)概念,其基本思想是通过国际合作发展全球尺度的地理信息。1996 年 2 月成立了全球测图国际指导委员会(ISCGM),其职责是促进和协调全球测图项目^[1]。全球测图项目的目的是提供一组对公众公开、质量已知和经过检验、具有一致性规范的数据(称为全球地图)。全球地图共有 8 层数据,即行政区划、水系、交通、居民点等 4 层矢量数据,以及地形、土地覆盖、土地利用、植被等 4 层栅格数据^[2]。

1997 年至 2000 年,全球测图项目各成员国测绘机构按 ISCGM 的要求利用现有地理信息开发了第 0 版(Version 0.0)全球地图。第 0 版全球地图中许多国家的土地覆盖数据都是直接使用 IGBP(International-Geosphere—Biosphere Programme)DISCover 土地覆盖产品。IGBP DISCover 是美国地质调查局、内布拉斯加林肯大学和欧洲委员会联合研究中心三家单位合作,以 1992-04~1993-03 月间的 AVHRR 影像为主要数据源开发的 1 km 分辨率全球土地覆盖数据集^[3]。为了提高全球地图数据的现势性和精度,

2003 年的 ISCGM 大会决定生产第 1 版(Version 1.0)的全球地图,主要是第一版的土地覆盖数据和植被数据。中国国家测绘局作为全球测图项目的成员单位,负责中国境内第一版全球地图数据的生产。

2 数据情况

生产全球地图第 1 版土地覆盖数据的主要遥感数据源是 2003 年全年的 MODIS/TERRA 16 d 合成产品 MOD43B4。MOD43B4 是经过大气校正和严格 BRDF 校正的第 3 级 MODIS 产品。原始 MOD43B4 产品为正弦曲线投影,并按每景影像覆盖经纬度 10°×10°范围分片。为了覆盖全中国,ISCGM 将东经 71°~137°、北纬 15°~55°之间的全部 MOD43B4 产品镶嵌在一起,并重新投影至经纬度系统,分辨率为 32.374 8″(约为 1 km)。最后获得的数据包括 23 期 MOD43B4 镶嵌影像,第 1 期影像的数据获取时间从 2003-01-01~01-16,依次类推,第 23 期影像的数据获取时间从 2003-12-19~12-31。每期影像由 MODIS 第 1 至第 7 光谱波段的经过大气和 BRDF 校正的反射率组成。

除此以外,在分类中还用到其他辅助数据,主

要包括:1999 年出版的《1 : 50 万全国土地利用图》^[4],2001 年出版的《中国 1 : 100 万植被图集》^[5],以及若干零星分布的 2002 年至 2003 年间获取的 Landsat ETM⁺ 和 SPOT5 影像。这些数据的主要作用是产生训练样本、进行精度评价以及在分类过程中用作参考。

全球地图第 1 版土地覆盖数据的分类系统采用联合国粮农组织开发的 LCCS(land cover classification system)系统^[6],这样可以使不同国家生产的土地覆盖产品具有兼容性和可比性。根据亚洲国家的状况,以 LCCS 2.0 为基础,确定了如下由 20 个地类组成的土地覆盖分类系统,如表 1 所示。

表 1 全球测图第 1 版土地覆盖分类系统
Tab. 1 Land Cover Classification Legend of Version 1.0 Global Map

序号	土地覆盖类型	说 明
1	常绿阔叶林	由常绿阔叶树木组成,树冠的覆盖率在 100%~40%,高度范围为 3~30 m
2	落叶阔叶林	由落叶阔叶树木组成,树冠的覆盖率在 100%~40%,高度范围为 3~30 m
3	常绿针叶林	主要由常绿针叶树木组成,树冠的覆盖率为 100%~40%,高度范围为 3~30 m
4	落叶针叶林	主要层由落叶针叶树木组成,树冠的覆盖率为 100%~40%之间,高度范围为 3~30 m
5	混交林	上述四种情况的混合
6	开放树木	主要层由开放树木组成,树冠的覆盖率为 40%和 20~10%之间,植被的开放度可进一步规定
7	灌木林	主要层由郁闭的灌木(树冠的覆盖率超过 60%)或开放的灌木(树冠的覆盖率在 20%~60%之间)组成,高度范围为 3~0.03 m
8	草地	主要层由郁闭的草本(冠盖覆盖率超过 60%)或开放的草本(冠盖的覆盖率在 20%~60%之间)组成,高度范围为 3~0.03 m
9	有稀树/灌木的草本植被	主要层由郁闭的草本(冠盖覆盖率超过 60%)或开放的草本(冠盖的覆盖率在 20%至~0%之间)组成,高度范围为 3~0.03 m,第二层由稀树组成,第三层由稀少灌木丛组成
10	稀疏草本或稀疏灌木	主要层由稀疏草本植被或稀疏灌木丛组成,冠盖覆盖率为 20%与 1%之间
11	农作物	除水稻外的其他草本农作物覆盖地区
12	稻田	种植水稻的地区
13	作物与自然植被镶嵌体	农作物与自然植被共生的地区(考虑到中国的实际情况,这种地区主要还是以农作物为主,因此将其归入“农作物”类别)
14	水性植被(盐性)	主要层由郁闭的树木(冠盖覆盖率超过 60%)或木本植物(冠盖覆盖率在 20%~60%之间)组成,水质为盐性的,总溶解固体浓度在 1 000~10 000 ppm 之间;或者,水质为盐性的,总溶解固体浓度超过 10 000 ppm,(只在海南岛等地有少量存在,但在 1 km 分辨率影像上难以反映,故最后分类结果中没有考虑此类)
15	湿地	主要植被地区每年至少两月覆盖有超过 4%的植被,环境主要受水存在时期的影响,例如,一年中水存在超过三个月,或者一年中水存在不足三个月,但在 75%的汛期存在
16	裸岩	地表由岩石、砂砾等覆盖
17	裸土	地表由沙、土覆盖
18	居民地	包括城市和农村居民点
19	雪和冰	由冰、雪覆盖的陆地
20	水体	由自然或人造水体覆盖的地区

注:第 13 和第 14 类没有包含在最终的土地覆盖分类中。

3 全国土地覆盖分类方法

对于区域或全球尺度的土地覆盖分类来说,所用的分类方法主要有非监督分类方法、决策树法和基于神经网络的 Fuzzy ARTMAP 方法等。前面提到的 IGBP DISCover 即是采用非监督分类方法完成的,但其中使用了大量基于人工干预的后处理。Fuzzy ARTMAP 方法因为处理高维数据太慢而没有在大尺度的土地覆盖分类中大量使用。决策树方法具有直观简单、精度高和速度快的优点^[7],而且无需考虑输入数据的概率分布,具有广泛的适应性^[8]。

本文提出了一种决策树和非监督分类相结合

的方法来进行全中国的土地覆盖分类。基本思想是首先将 MODIS 影像先进行非监督聚类,然后以非监督分类得到的图斑为处理单元,利用时间序列的 MODIS 影像构建决策树,将每一图斑指派到 20 个土地覆盖类别中。非监督聚类曾经被用于 IGBP DISCover 全球土地覆盖分类^[3]及 Global Land Cover 2000 项目中的中北美洲土地覆盖分类中^[9],但是单纯使用这种方法需要大量的人工后处理。本文提出的方法其实是一种面向对象的分类思想,决策树分类处理的基本单元是非监督聚类得到的图斑,即对象,而非单个像元。这样处理的优势有二:首先,分类的结果中椒盐噪声大大减少;其次,决策树规则的阈值可以方便、准确地被确定。

3.1 掩模生成

首先使用 2002 年更新的全国 1 : 25 万数字线划图(DLG, 矢量格式)提取居民地, 然后将其转化为 1 km 分辨率的栅格影像, 配准叠加在 MODIS 影像上, 形成居民地的掩模。因为混合像元的关系, 直接从 MODIS 影像上分类出居民地将会有很大误差, 而且也会给后续决策树分类带来影响。

3.2 非监督聚类

以 2003-08-13~2003-08-28 的第 15 期 MODIS 影像为数据源, 采用 ISODATA 非监督分类算法进行影像聚类, 设定输出类别为 60 类。T15 影像对应着植被最茂盛的时期, 在该影像上, 植被区、无植被区以及不同植被覆盖度的区域能得到明显的区分。设定 60 个输出类别数目, 可以保证具有非常细微的光谱特征差异的斑块保持在不同的初始类别中。否则, 如果初始聚类数目过少, 可能导致从一开始有些类别就混合在一起, 或者导致一些细小的类别被淹没。

在初始聚类的结果基础上, 进行了基于多时相均值的进一步聚类。这样做的主要目的是得到更有意义的图斑。另一方面是考虑到由于我国地域广大, 各种植被的最绿期在不同地域对应的时间并不一致。进一步聚类的对象是初始聚类的输出图斑, 对每一图斑, 考察其在第 13、14、15、16 期影像上的均值, 再次进行非监督聚类, 设定输出的类别数为 40, 插页 IV 彩图 1 表示了这一过程。

3.3 基于决策树的多时相 MODIS 影像分类

为了充分利用多时相信息, 在决策树分类中选用了 4 期的 MODIS 影像, 除上述的 T15 外, 另外还用了 2003-04-7-04-22 的影像(记为 T7), 以及 06-26~07-11 的影像(记为 T12)、11-01~11-16 的影像(记为 T20)。

3.3.1 分类特征选择

本文研究的决策规则分类特征由 MODIS 前 7 个通道的反射率及以下两个指数构成。

增强型植被指数 EVI (enhanced vegetation index)。EVI 是 Heute 针对 MODIS 数据的特点发展的一种新型植被指数^[7]。EVI 克服了 NDVI 容易饱和的缺点, 同时对残留的大气气溶胶和土壤背景的影响进行了校正^[11], 能更准确地反映地表植被生物量的大小。

MODIS EVI 的计算公式如下:

$$EVI = 2.5 * (\rho_2 - \rho_1) / (\rho_2 + c_1 \rho_1 + c_2 \rho_3 + L)$$

(1)

其中, L 为冠盖背景(通常是土壤)调节参数, 这里

取 1; 参数 c_1 和 c_2 分别对应大气气溶胶对蓝光和红光的抵抗效应, 通过蓝波段(ρ_3)的辐射来修正大气气溶胶对红波段(ρ_1)的影响, 分别取 6 和 -7.5。这里 ρ 要求是已经过大气校正的反射率数据^[12]。

水分指数 NDWI (normalized difference water index)^[13]:

$$NDWI = (\rho_2 - \rho_5) / (\rho_2 + \rho_5)$$
 (2)

对于植被来说, 无论含水多少, 对于近红外光都有强烈的反射, 而在短波红外波段, 随着植被含水量的增加, 其反射显著下降。因此, 植被的 NDWI 都是正值, 并随着植被含水量的增加而变大。对于裸地来说, 在短波红外波段比近红外波段有更强的反射, 因此, 土壤的 NDWI 为负值, 但同样随着湿度的增加 NDWI 变大。

3.3.2 决策树分类规则的建立

以《1 : 50 万全国土地利用图》和《中国 1 : 100 万植被图集》为主要参考数据, 在经过聚类后的 MODIS 影像上人工选择了比较纯净的图斑作为训练样本。一共选择了 18 个土地覆盖类型的训练样本。经过对训练样本的学习、归纳, 形成了区分各个类别的规则。以下规则是串行处理的, 即前面被提取出的像斑将以掩模的形式被排除, 不参加后面的规则判断。

1) 冰雪和水体。冰雪在可见光和近红外波段具有极大的反射率, 而水体在近红外和短波红外波段强烈吸收入射光, 通过对相应波段内的反射率设置阈值就可以与其他地类区分开来。

2) 裸地。裸地主要从 T15 影像的 EVI 和近红外及短波红外波段反射率加以区分。裸地又分为裸土和裸岩。裸土在短波红外波段具有较高的反射率, 与其他地类能很好地区分。除水体之外, 湿地和裸岩在 MODIS 波段 2 上具有最低的反射率。而裸岩和湿地可以通过 EVI 值的比较进一步分离, 湿地具有较裸岩更大的 EVI。

3) 稀疏草本或稀疏灌木。通过短波红外波段和 EVI 值的结合可完成稀疏草本或稀疏灌木地类的提取。采用 T15 影像的波段 6 反射值, 在提取出裸土之后, 稀疏草本或稀疏灌木与草地有部分混淆。进一步, 通过 EVI 可以将草地区分开, 因为在 T15 影像上, 草地有更高的 EVI 值。

4) 湿地。通过对样本的分析, 在 T15 影像上湿地具有最小的第 5 波段反射率, 组合这几个条件构成了准确提取湿地的规则。

5) 水稻。水稻的区分采用了处于水稻生长早期(4 月中下旬)的 T7 影像。此时水稻茎叶中包含了大量水分, 因此, 水稻具有较大的 NDWI

值。部分常绿阔叶林与水稻有相似的 NDWI 值,进一步的分离通过比较 T7 影像第 4 波段的反射率加以实现。

6) 有稀树/灌木的草本植被。有稀树/灌木的草本植被的分类规则是基于 T12 影像的 EVI 和 MODIS 第 3 波段反射率构建的。

7) 灌木。灌木林的提取采用的是 T15 影像的 EVI 值和红光波段来完成的。灌木林的 EVI 值与落叶针叶、常绿针叶和常绿阔叶林等有重叠,但在 MODIS 第 1 波段上,灌木林比林地有更大的反射率。

8) 农作物。农作物的提取是通过 T15 影像的近红外波段及 EVI 值来实现的。T15 影像对应着农作物的成熟期,此时农作物具有较大的近红外波段反射率。

9) 林地。林地分为常绿阔叶林、落叶阔叶林和常绿针叶林、落叶针叶林及混交林。首先采用 T20 影像利用落叶林的季候特征将落叶林与常

绿林和混交林分开。落叶针叶林和落叶阔叶林的区分采用植被茂盛的 T15 影像的 EVI 值与近红外波段进行,阔叶林比针叶林在近红外波段具有明显更强的反射率。常绿阔叶林与常绿针叶林和混交林的区分是通过 T20 时相的 EVI 值和短波红外反射率实现的。常绿阔叶林的冠层含水量通常比常绿针叶林和混交林高,这一差别在 11 月初的 T20 影像上更加明显。因此,在 T20 时相短波红外反射率上,常绿阔叶林显著低于常绿针叶林和混交林。常绿针叶林与混交林的区分是通过 T15 影像的 EVI 值和近红外波段组合实现的。

10) 草地。草地覆盖面积比较广,类型多样,难以用单一条件将其全部选出。主要的草地类型选用植被比较茂盛的 T15 时期影像与其他植被类型进行区分。另外两个次要的草地类型分别是与灌木林及有稀树/灌木的草本植被两个地类相区分的。

以上形成的规则如表 2 所示。

表 2 决策树分类规则
Tab. 2 Decision Rules of the Decision Tree Classifier

土地覆盖类型		决策规则
冰雪		$\rho_1(T15) \geq 0.4$
水体		$\rho_2(T15) \leq 0.1$
裸土		$\rho_6(T15) \geq 0.35, EVI(T15) < 0.1$
裸岩		$0.1 \leq \rho_2(T15) \leq 0.2, 0.02 \leq EVI(T15) \leq 0.082$
稀疏灌木或稀疏草本		$\rho_6(T15) \geq 0.306, 0.088 \leq EVI(T15) \leq 0.15$
湿地		$0.11 \leq \rho_5(T15) \leq 0.18, 0.2 \leq EVI(T15) \leq 0.36, NDWI(T15) > 0$
水稻		$0.08 \leq \rho_7(T20) \leq 0.145, 0.024 \leq NDWI(T7) \leq 0.12$
有稀树/灌木的草本植被		$0.12 \leq EVI(T12) \leq 0.25, 0.1 \leq \rho_3(T12) \leq 0.24$
	灌木	$0.34 \leq EVI(T15) \leq 0.41, \rho_1(T15) \geq 0.037$
	庄稼	$0.08 < \rho_1(T20) < 0.16, \rho_2(T15) > 0.28, EVI(T15) > 0.41$
	落叶针叶林	$0.05 \leq EVI(T20) \leq 0.178, 0.02 \leq \rho_1(T15) \leq 0.04, 0.17 \leq \rho_2(T15) \leq 0.26, 0.3 \leq EVI(T15) \leq 0.45$
落叶阔叶林		$0.05 \leq EVI(T20) \leq 0.178, \rho_1(T20) < 0.9, 0.42 < EVI(T15) < 0.55, 0.06 < \rho_2(T15) < 0.33$
常绿阔叶林		$0.18 \leq EVI(T20) \leq 0.34, \rho_1(T20) < 0.09, 0.1 < \rho_5(T15) < 0.18$
常绿针叶林		$0.39 < EVI(T15) < 0.563, \rho_1(T20) < 0.09, 0.22 < \rho_2(T15) \leq 0.3$
混交林		$0.4 < EVI(T15) < 0.574, 0.32 \leq \rho_2(T15) \leq 0.342, \rho_1(T20) < 0.09$
草地	草地 1	$0.153 \leq EVI(T15) \leq 0.23, 0.239 \leq \rho_6(T15) \leq 0.43$
	草地 2	$0.22 \leq EVI(T15) \leq 0.32, 0.26 \leq \rho_2(T15) \leq 0.35$
	草地 3	$0.3 \leq EVI(T15) \leq 0.41, 0.051 \leq \rho_1(T15) \leq 0.124$

4 分类结果及精度评价

基于非监督聚类图斑的决策树土地覆盖分类的结果如封三彩图 2 所示。为了对此结果进行精度评定,在全国范围内随机产生了 2 000 个采样点。采样点上的土地覆盖分类参考 2002~2003 年间获取的 SPOT5 和 Landsat ETM⁺ 多光谱影像,并结合 1 : 500 000 土地利用图和 1 : 1 000 000 植被图进行解译。在没有 SPOT5 和

ETM⁺ 影像可用的情况下,采用了 Google Earth 上下载的大量影像。从这 2 000 个点的比较结果进行统计,得到总体分类精度为 89.14%,Kappa 系数为 87.76%。

在进行室内精度评价的同时,也开展了外业验证。在北京周边地区选择了 22 个实地考察点,如封三彩图 3 所示。外业调查人员携带 GPS 到每一个外业点采集了土地覆盖地面实况数据。由于交通原因,最终有 3 个考察点没能到达。封三彩图 4 显示了其中一个点上的土地覆盖分类及实

地照片对照结果。在 19 个外业考察点上有 16 个点的土地覆盖分类结果与地面实况调查是一致的,分类正确率为 84%。由于经费限制,目前无法在全国范围内开展更多的实地验证,但这个结果也能部分支持室内精度评价。

5 结 语

第 1 版全球地图产品已于 2008 年 6 月 5 日正式发布,本文完成的中国土地覆盖分类结果是其中重要的组成部分。本文提出的非监督聚类 and 决策树分类相结合的 MODIS 影像土地覆盖分类方法融入了面向对象的处理思想,能有效改善基于像元处理导致的椒盐噪声,同时充分利用了决策树分类简洁、高效、能容纳多维特征的优势。通过该方法在本项目的成功应用,证明了该方法在区域及全球尺度、多时相遥感数据分类中的有效性。下一步将对形成的分类方法进行改进,重点完善如何快速确定最优的决策规则等,将其应用于最新的 MODIS 数据以生产更新的土地覆盖产品。

参 考 文 献

[1] 苏山舞. 全球地图——人类的共同财富[J]. 测绘通报,2000(2):1-3

[2] 王桂芝. 全球测图项目及其应用浅谈[J]. 地理信息世界,2005, 3(1): 28-33

[3] Loveland T R, Reed B C. Development of a Global Land Cover Characteristics Database and IGBP DIS-Cover from 1 km AVHRR Data[J]. International Journal of Remote Sensing, 2000, 21(6/7): 1 303-1 330

[4] 国土资源部全国土地资源调查办公室. 1:50 万全国土地利用图[M]. 北京:科学出版社,1999

[5] 中国科学院. 1:100 万中国植被图集[M]. 北京:

科学出版社,2001

[6] Di Gregorio A. Land Cover Classification System—Classification Concepts and User Manual for Software Version 2[C]. FAO Environment and Natural Resources Service Series, Rome, 2005

[7] Strahler A, Doug M, Jordan B, et al. MODIS Land Cover Product Algorithm Theoretical Basis Document (ATBD) Version 5. 0[OL]. http://modis.gsfc.nasa.gov/data/atbd/land_atbd.php/, 1999

[8] Friedl M A, Brodley C E. Decision Tree Classification of Land Cover from Remotely Sensed Data[J]. Remote Sensing of Environment, 1997, 61: 399-409

[9] Latifovic R, Zhu Z L, Cihla J, et al. Land Cover Mapping of North and Central America-Global Land Cover 2000[J]. Remote Sensing of Environment, 2004, 89: 116-127

[10] Huete A, Justice C, Liu H. Development of Vegetation and Soil Indices for MODIS-EOS[J]. Remote Sensing of Environment, 1994, 49: 224-234

[11] 王正兴,刘闯,Huete A. 植被指数研究进展:从 AVHRR-NDVI 到 MODIS-EVI[J]. 生态学报, 2003, 23: 979-987

[12] Gao B C. NDWI-a Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space[J]. Remote Sensing of Environment, 1996, 58: 257-266

[13] Huete A, Didan K, Miura T, et al. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices[J]. Remote Sensing of Environment, 2002, 83: 195-213

第一作者简介:张永红,博士,研究员。主要研究领域包括土地利用土地覆盖及其变化、合成孔径雷达影像的极化及干涉处理、森林及植被燃烧遥感监测等。
E-mail:yhzhang@casm.ac.cn

The Production of China Land Cover Data for Version 1.0 Global Map

ZHANG Yonghong¹ ZHANG Jixian¹ GUO Jian¹ CAO Yinxuan¹

(1 Chinese Academy of Surveying and Mapping, 16 Beitaping Road, Beijing 100039, China)

Abstract: This paper describes the production of China land cover data for version 1.0 Global Map with the emphasis on the developed land cover classification methodology—a combination of unsupervised image clustering and supervised decision tree classification. The 23 MODIS/TERRA 16 d composites (MOD43B4 product) acquired over the whole year of 2003 are the