

Land Use Classification and Vegetation Coverage Inversion Based on Landsat OLI Data in Qingdao

Yanzhen Liu, Yang Liu, Jia Guo

Shandong University of Science and Technology, Qingdao Shandong
Email: 1318928597@qq.com

Received: Jul. 4th, 2019; accepted: Jul. 17th, 2019; published: Jul. 24th, 2019

Abstract

Hierarchical classification method is a method to obtain classification rules by mining spatial data information, and vegetation coverage is an important parameter to characterize surface vegetation coverage. In this paper, based on the 2017 Landsat OLI image data of Qingdao city, land cover in the study area was classified and studied by stratified classification method. First of all, by making the necessary data preprocessing, the spectral characteristics of the image in the study area are analyzed and the image of NDVI, NDWI, NDBI index are calculated, and by using the stratified classification method we classify the land use in the research area. According to the results of land use classification and vegetation coverage inversion in the study area, it is advantageous to the planning of land utilization in Qingdao form, and it is of great importance to protect and improve the ecological environment of Qingdao and to plan the direction of urban development scientifically and reasonably.

Keywords

Landsat OLI Data, Land Use Classification, Hierarchical Classification, Vegetation Coverage

基于Landsat OLI数据的青岛市土地利用分类和植被覆盖度反演

刘艳祯, 刘 杨, 郭 佳

山东科技大学, 山东 青岛
Email: 1318928597@qq.com

收稿日期: 2019年7月4日; 录用日期: 2019年7月17日; 发布日期: 2019年7月24日

摘要

分层分类方法是一种通过挖掘空间数据信息来获取分类规则的方法, 植被覆盖度则是刻画地表植被覆盖的重要参数。本文采用青岛市2017年的OLI影像数据, 采用分层分类方法对研究区进行土地覆盖分类研究。首先对影像进行预处理, 分析影像光谱特征并计算NDVI、NDWI、NDBI等指数, 利用分层分类方法进行土地利用分类, 并根据像元二分模型用NDVI估算研究区植被覆盖度。根据得到的土地利用分类结果和植被覆盖度反演结果, 有利于规划青岛市土地利用形式, 保护和改善青岛市生态环境, 科学合理地规划城市发展方向。

关键词

Landsat OLI数据, 土地利用分类, 分层分类, 植被覆盖度

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来, 农田的缩减和人口的增多使得青岛市人地矛盾越来越突出, 严重制约了青岛市经济发展, 对生态环境造成了很大的威胁。利用遥感技术和GIS技术动态监测研究区的土地利用情况, 可以对研究区土地利用的时空动态变化进行及时有效的了解和分析。常规的遥感图像分类方法有很多, 比如模糊逻辑推理法、掩膜法、主成分分析和多元线性回归法、支持向量机、最大似然监督分类、分层分类方法、决策树分类法等[1]-[11]。2004年, 李彤、吴骅采用决策树分类技术对北京市土地利用覆盖现状进行研究, 得出该方法具有简单、明确、分类结构直观等优点[8]。分层分类法则操作简单, 有良好的稳健性和鲁棒性, 并且可以有效处理大量数据。因此, 本文决定结合以上两种方法, 采用基于决策树的分层分类方法, 利用各个波段的光谱信息, 通过设置阈值逐层地提取各种土地类型。植被覆盖及变化是区域生态系统环境变化的重要指示, 对研究水文、生态、全球变化等都有重要意义。2003年, 李苗苗, 吴炳方等人对已有的像元二分模型参数估算方法加以改进, 并建立了用NDVI定量估算植被覆盖度的模型[12], 本文决定使用该模型, 直接利用植被指数近似估算植被覆盖度。本文的创新之处在于在提取道路时由于考虑到该种土地利用类型混合像元过多不易提取, 决定结合道路矢量文件, 经过波段合成, 最后进行分层分类。结果显示用该方法提取道路的结果要优于传统决策树阈值法。

2. 实验方法

2.1. 基于决策树的分层分类方法

分层分类是把数据按照一定的规则进行划分, 在每个节点利用相应的特征, 使分割后一些准则函数达到最优效果。不同的标准有不同的划分方法和决策树, 但最终都是为了得到最好的分割结果。本文采用的分类方法是建立二叉树。二叉树虽然结构相对简单, 但其识别能力较强。

把影像上的全部地物看作一个根结点T, 先把T分为T1(植被)和T2(非植被)两大类; 然后在每个大类中进一步分类, 比如T1(植被)可分为A(耕地)和B(林地), T2(非植被)可分为C(水体)和T3(非水体); T3可再分为D(交通用地)和T4(非交通用地), T4可再分为E(居民点及工矿用地)和F(未利用土地), 直

到把图像中的全部类别都分出来。最终形成一个树状结构，在树状结构每一个结点位置处可选择不同规则进行更细致的分类。基本原理如图 1 所示。

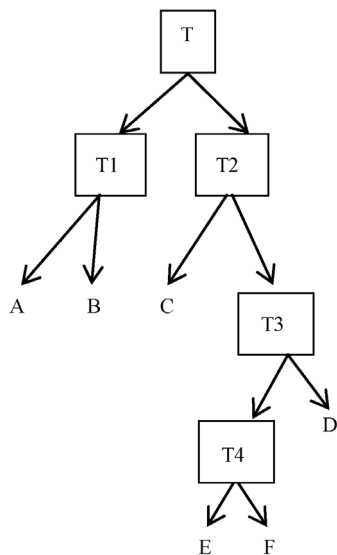


Figure 1. Basic principles of decision tree classification algorithm
图 1. 决策树分类算法基本原理

2.2. 精度评价

在计算机分类中，由于同物异谱或同谱异物、分类系统与数据资料的匹配程度不高等原因，会导致出现分类错误；此外，遥感图像中各类别的大小和形状也会影响分类结果的准确性。造成分类错误有两个主要原因：属性错误，即类别识别错误；位置误差，即所有类型的边界都是不准确的。

本文采用模糊矩阵方法对计算机分类结果的准确性进行评价。通过建立模糊矩阵来计算总体分类精度和 Kappa 系数。其中 p_1 是总体分类精度。

$$K = (p_1 - p_2) / (1 - p_2) \quad (1)$$

假设每一类的真实样本个数分别为 a_1, a_2, \dots, a_0 ，而预测出来的每一类的样本个数分别为 b_1, b_2, \dots, b_0 ，总样本个数为 n ，则有

$$p_2 = (a_1 * b_1 + a_2 * b_2 + \dots + a_0 * b_0) / n^2 \quad (2)$$

2.3. 植被覆盖度反演方法

本文根据像元二分模型的原理利用 NDVI 估算研究区植被覆盖度。

设一个像元中有植被覆盖的面积比例，即该像元的植被覆盖度为 f_c ，可以得到计算植被覆盖度的公式：

$$f_c = (S - S_{soil}) / (S_{veg} - S_{soil}) \quad (3)$$

其中， S 为通过遥感传感器所观测到的信息， S_{soil} 与 S_{veg} 是像元二分模型的 2 个参数。

一个像元的 NDVI 值可以表示为由绿色植被部分所贡献的信息 $NDVI_{veg}$ 与由无植被覆盖(裸土)部分所贡献的信息 $NDVI_{soil}$ 两部分组成。将 NDVI 代入上式，得

$$f_c = (NDVI - NDVI_{soil}) / (NDVI_{veg} - NDVI_{soil}) \quad (4)$$

其中， $NDVI_{soil}$ 为裸土或无植被覆盖区域的 NDVI 值， $NDVI_{veg}$ 则代表完全被植被所覆盖的像元的 NDVI 值，即纯植被像元的 NDVI 值。

3. 实验过程

首先对 Landsat OLI 数据进行预处理, 镶嵌、裁剪出研究区域。然后对裁剪后的 Landsat OLI 数据进行分层分类, 对分类结果进行精度评价, 并反演出研究区植被覆盖度。

3.1. 分类过程

对选取的研究区影像进行解译分析, 利用计算的归一化植被指数(NDVI)、归一化水体指数(NDWI)、归一化建筑物指数(NDBI)、各波段不同地物辐射亮度值等数据作为决策树分类的特征数据。然后应用分层分类, 建立分类模型, 得到初步分类结果。参考已有的土地分类系统, 结合研究的需要, 把研究区分成林地、耕地、水域、交通用地、居民点及工矿用地、未利用土地 6 个类别。

对于植被和非植被, 用 NDVI 值来区分。在本研究中, 由于图像中存在大量混合像元, 因此 $NDVI > 0$ 的未必就是植被, 经过多次实地考察、目视解译、统计分析和精度检验, 最终将阈值确定为 0.23, 即 $NDVI > 0.23$ 的像元被归类为植被。

对植被类型中的耕地和林地, 统计分析各波段 DN 值的最大值、最小值、均值和均方差, 结果表明, 耕地和林地 OLI 第 6 波段的 DN 值有较大差异, 耕地的反射率明显较小, 故可将阈值设为 87, 即 $band6 < 87$ 的像元被归属为耕地。

对于非植被, 首先经过反复实验验证, 用 $NDWI > 0.38$ 作为判别函数来提取水体, 可以得到很好的分离效果。然后在 ArcGIS 中将青岛市道路 shape 矢量文件转换成栅格文件, 进行波段合成, 将转换得到的栅格文件作为 Landsat 影像的一个波段, 记作 $band_{road}$, 则 $band_{road} \neq 255$ 为道路, 对于居民点及工矿用地与未利用土地, 则采用 $NDBI < 0.35$ 作为判别函数来分类, 在这两种地物类型样区进行 DN 值的统计分析, 发现居民点及工矿用地在 OLI 第 5、6 波段的 DN 值之和远小于未利用土地, 将阈值设为 178, 故 $band5 + band6 < 178$ 且 $NDBI < 0.35$ 的像元被归属为居民点及工矿用地, 剩余部分即为未利用土地。具体分类过程如图 2 所示。

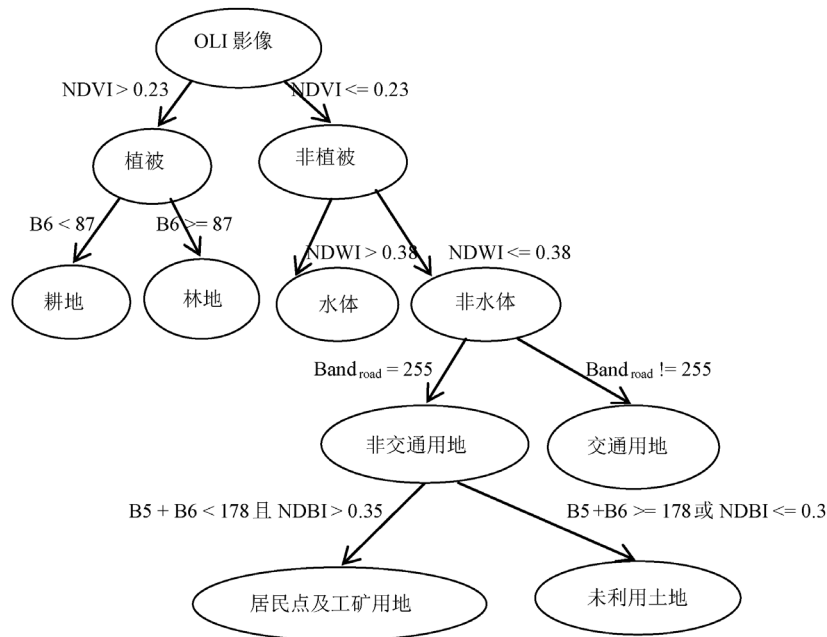


Figure 2. Hierarchical classification map
图 2. 分层分类图

由于分类过程是按像元逐个进行的，在输出的分类结果图中，成片的地物类别中会存在零星少量异类地物像元，且散落分布。为了分类结果实用可靠且做出的图像美观，对上述分类得到的结果还需进行一定的后处理工作。最终的分类结果如图 3 所示。

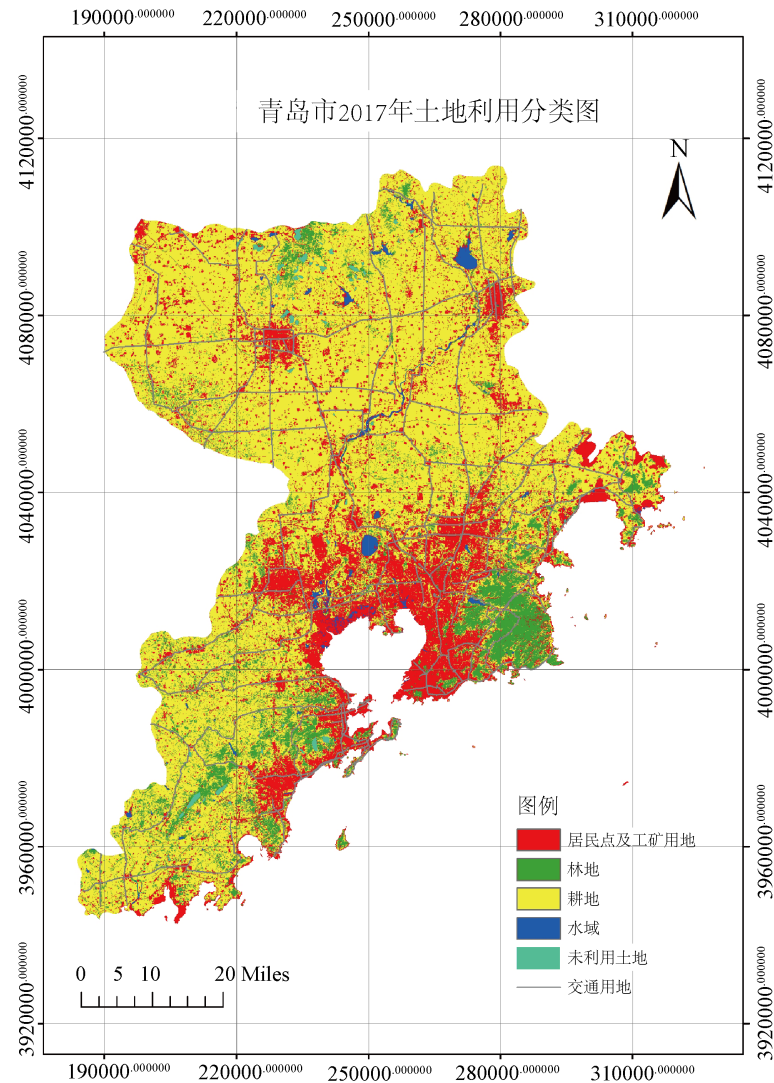


Figure 3. Land use classification results map
图 3. 土地利用分类结果图

如上图所示，研究区内林地约占 16%，主要分布在东部、南部和北部的山地地区；耕地约占 42%，主要分布在境内平原丘陵地区，所占面积最大；居民点及工矿用地约占 29%，主要沿海沿河分布，其中胶州湾附近建筑最密集；未利用土地约占 4%，主要分布在山地、丘陵附近；水域约占 5%，包括境内的河流、湖泊等；交通用地约占 4%，在研究区内纵横交错。

3.2. 精度评价

不同土地利用类型的分类结果与地物真实空间分布情况的一致性程度可以用 Kappa 系数的大小来显示。Kappa 系数的大小范围在-1~1 之间，但通常在 0~1 之间。我们把 Kappa 系数的值分成五组来表示一

致性的程度: 0.00~0.20 表示一致性很低, 0.21~0.40 表示一致性一般, 0.41~0.60 表示一致性中等, 0.61~0.80 表示高度一致和 0.81~1 表示几乎完全一致。

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} x_{+i})} \quad (5)$$

其中, r 为误差矩阵中总列数(即总的类别数); x_{ii} 为误差矩阵中第 i 行、第 i 列上像元数量(即正确分类的数目); x_{i+} 和 x_{+i} 分别为第 i 行和第 i 列的总像元数量; N 为总的用于精度评估的像元数量。

在进行精度评价时, 每类至少有 30~50 个样本点。本研究在研究区内随机选取了 240 个样本点, 并且保证每种地物类型至少有 30 个样本点, 去除边缘过渡地带的样本点, 最终保留 180 个左右样本点进行精度评价。评价依据是原始影像的目视判读和野外实地调查的结果。通过建立混淆矩阵, 根据式 2.1, 最终得到 Kappa 系数约为 0.83。

3.3. 植被覆盖度反演

先需要计算整个图像的 NDVI 值, 然后利用土地分类图提取每个单元的 NDVI 值, 计算每个元素的 NDVI 值的累积频率值。最后, 根据频率累加表, 把土地利用单元的内取频率为 5% 的 NDVI 值作为 $NDVI_{soil}$ 。

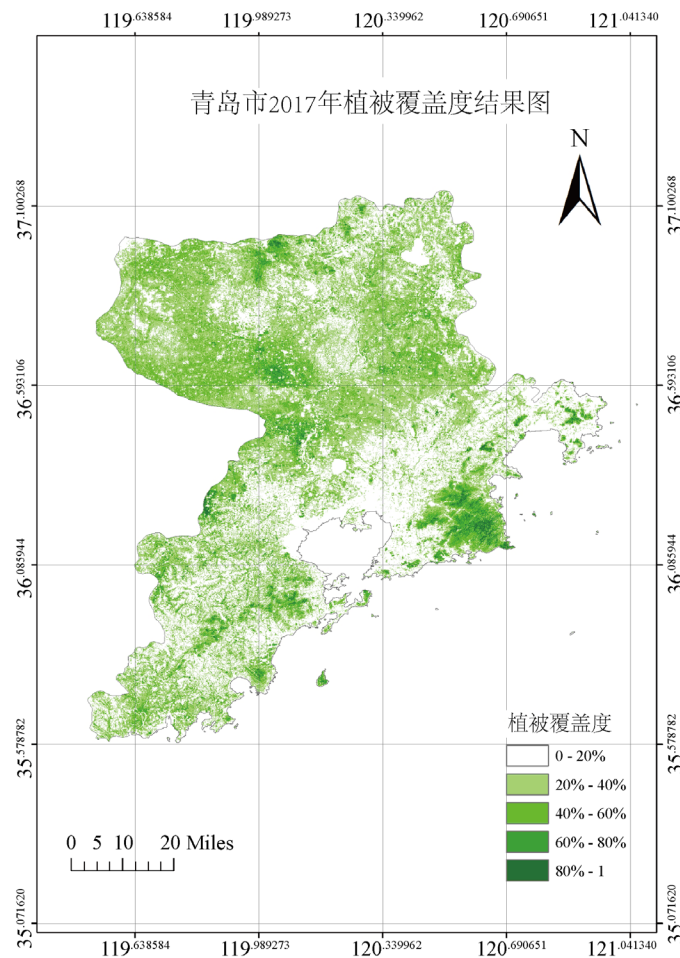


Figure 4. Vegetation coverage inversion results map

图 4. 植被覆盖度反演结果图

从上图 4 可以看出, 青岛市植被大多分布在东部、西南部和北部的山地丘陵地区, 其次是中部的平原和盆地地区, 胶州湾附近植被覆盖度比较低, 出现这种分布状况的原因主要是山上多林地, 而平原丘陵地区种植农作物, 因此这些地区的植被覆盖度都比较高, 而居民点及工矿用地多为建筑物, 植被较少, 另外未利用土地多为裸露的土壤和岩石, 植被覆盖度几乎为零。

4. 结束语

本文以青岛市为研究区, 选择 2017 年 9 月 23 日的 Landsat OLI 影像数据, 应用 ENVI 和 ArcGIS 软件平台, 采用分层分类方法进行土地分类研究并精度评价, 最后得到青岛市植被覆盖度。根据土地利用分类结果和植被覆盖度反演结果, 有利于规划青岛市土地利用形式, 保护和改善青岛市生态环境, 科学合理地规划城市发展方向。

本文在分类过程中存在的主要问题是分类结果中的土地类型没有园地和草地, 并且没有将住宅用地和工矿用地区分开来, 而是统一划分成了居民点及工矿用地。存在上述问题的原因主要是分层分类方法虽然分类结构简单、容易理解, 但该方法的阈值取值随机性很强, 稳定性很差。因此利用分层分类方法对其他地区进行土地利用信息提取时, 必须针对不同地区的遥感影像和土地利用状况修订和调整阈值。而且如果可以结合研究区影像的空间纹理信息等辅助数据, 并对结果进行恰当的分类后处理, 可以进一步完善分类结果并提高分类精度。本文估算植被覆盖度的过程主要存在以下几个问题: $NDVI_{veg}$ 与 $NDVI_{soil}$ 的值不够准确; 用 $NDVI$ 估算植被覆盖度存在一个弱点, 即 $NDVI$ 在植被覆盖度大于 85% 时, 对植被覆盖度的变化将不再敏感。针对上述问题, 在今后的研究中可以尝试采取以下措施: 对于不同类型的土地覆盖, 最好利用实测资料来确定 $NDVI_{veg}$ 和 $NDVI_{soil}$ 值; 图像中某一土地利用类型的总像元个数不能太少。基于遥感技术估算植被覆盖度的方法不需要地面实测数据建模, 易于推广, 从未来发展方面来说, 植被覆盖度的估算方法会变得更加完善。

参考文献

- [1] 张丽苏, 吴嘉平. 分层分区分景相结合的区域土地利用/覆盖分类方法——以浙江钱塘江流域分类为例[J]. 国土资源遥感, 2007, 19(3): 74-77.
- [2] 凌涛, 李鹏, 左小清. 青岛市土地利用演化驱动力因素分析[J]. 安徽农业科学, 2016, 44(13): 257-262.
- [3] Reyes, A., Solla, M. and Lorenzo, H. (2017) Comparison of Different Object-Based Classifications in Landsat TM Images for the Analysis of Heterogeneous Landscapes. *Measurement*, **97**, 29-37.
- [4] Sinha, S., Sharma, L.K. and Nathawat, M.S. (2015) Improved Land-Use/Land-Cover Classification of Semi-Arid Deciduous Forest Landscape Using Thermal Remote Sensing. *The Egyptian Journal of Remote Sensing and Space Science*, **18**, 217-233. <https://doi.org/10.1016/j.ejrs.2015.09.005>
- [5] 明艳芳. 青岛市土地利用信息遥感提取方法[C]//智能信息技术应用学会. Proceedings of 2011 International Conference on Ecological Protection of Lakes-Wetlands-Watershed and Application of 3S Technology (EPLWW3S 2011 V3). 2011: 3.
- [6] 李晓明, 常庆瑞, 张仁慧. 农牧交错地带土地利用分层分类方法研究——以靖边县为例[J]. 干旱地区农业研究, 2010, 28(5): 237-240.
- [7] 李彤, 吴骅. 采用决策树分类技术对北京市土地覆盖现状进行研究[J]. 遥感技术与应用, 2004, 19(6): 485-487.
- [8] 魏信. 基于决策树分类技术的北京城区土地利用分类研究[C]//2010 International Conference on Remote Sensing (ICRS 2010). Proceedings of 2010 International Conference on Remote Sensing (ICRS 2010) Volume 4. 2010: 4.
- [9] 秦臻, 汪云甲, 王行风, 等. 基于 ENVI 的决策树方法在土地利用分类中的应用[J]. 金属矿山, 2011(2): 133-135+140.
- [10] 王光远. 基于决策树的分类方法在土地利用分类中的应用[J]. 中国科技信息, 2013(3): 46-47.
- [11] 郭力娜, 张梦华, 张永彬, 等. 唐山市区土地利用的 Landsat 8 影像分层分类[J]. 测绘科学, 2017, 42(10): 88-94+125.
- [12] 李苗苗, 吴炳方, 颜长珍, 等. 密云水库上游植被覆盖度的遥感估算[J]. 资源科学, 2004, 26(4): 153-159.

知网检索的两种方式：

1. 打开知网首页：<http://cnki.net/>，点击页面中“外文资源总库 CNKI SCHOLAR”，跳转至：<http://scholar.cnki.net/new>，搜索框内直接输入文章标题，即可查询；
或点击“高级检索”，下拉列表框选择：[ISSN]，输入期刊 ISSN：2329-549X，即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版：<http://www.cnki.net/old/>，左侧选择“国际文献总库”进入，搜索框直接输入文章标题，即可查询。

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：gst@hanspub.org