

Comparison of Accuracy of Different Supervised Classification Methods Based on Landsat 8 Images

Kou Zhang, Wujun Xi*, Zhengbiao Zhu

School of Geography and Tourism Management, Chuxiong Normal University, Chuxiong Yunnan
Email: 3030282201@qq.com, *xwj@cxtc.edu.cn, 1005107349@qq.com

Received: Jul. 15th, 2020; accepted: Jul. 26th, 2020; published: Aug. 3rd, 2020

Abstract

Supervised classification is a common method of remote sensing image classification. In order to explore the accuracy of different supervised classification methods, we used neural network, spectral angle, maximum likelihood, minimum distance, spectral information divergence, Mahalanobis distance, parallelepiped, support vector machine, binary coding methods to classify and precision analysis with Landsat 8 images in Lufeng county, Yunnan Province. The results show that: from the perspective of overall accuracy, the accuracy of the neural network, support vector machine, and maximum likelihood method is high; from the accuracy of different land types, the maximum likelihood method has the highest classification accuracy for grassland and cultivated land, the neural network classification method has the highest classification accuracy for forest land and unused land, and the support vector machine method has the highest classification accuracy for urban and rural industrial and mining residential land and water.

Keywords

Supervised Classification, Classification Accuracy, Landsat 8

基于Landsat 8影像的不同监督分类方法精度对比分析

张 蕊, 席武俊*, 诸正彪

楚雄师范学院, 地理科学与旅游管理学院, 云南 楚雄
Email: 3030282201@qq.com, *xwj@cxtc.edu.cn, 1005107349@qq.com

收稿日期: 2020年7月15日; 录用日期: 2020年7月26日; 发布日期: 2020年8月3日

*通讯作者。

摘要

监督分类是遥感图像分类的常用方法。为了探究不同监督分类方法的精度,分别使用神经网络、波谱角、最大似然、最小距离、光谱信息散度、马氏距离、平行六面体、支持向量机法、二进制编码等9种算法监督分类方法对云南省禄丰县Landsat 8影像进行分类及精度分析,结果表明:从总体精度来看,神经网络分类法、支持向量机法、最大似然法的精度较高;从各地类分类精度来看,最大似然法对草地以及耕地的分类精度最高,神经网络分类法对林地和未利用地的分类精度最高,支持向量机法对城乡工矿居民用地和水域的分类精度最高。

关键词

监督分类, 分类精度, Landsat 8

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

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



Open Access

1. 引言

在关于遥感技术的研究中,计算机图像分类研究是较重要的内容,在进行图像分类时,监督分类是普遍适用的一种分类法[1],该方法需要进行学习和一定的知识经验才能进行分类,但能在进行土地利用分类时准确确定地物的类别和属性。因此,关于监督分类方法的研究在科技发展中也在不断创新。并且随着遥感技术在土地利用监测方面应用越来越广泛,学者们对监督分类的研究主要集中在分类方法的对比和优化上。

监督分类算法有多种,有基于传统统计分析学的,基于神经网络的,基于模式识别等多种分类方法,在目前的遥感分类应用中,用得较多的是传统的模式识别分类方法,诸如最小距离法,平行六面体法、最大似然法、等监督非监督分类法[2]。而一些不常用的方法,例如波谱角法,二进制编码法、光谱信息散度法等未有过多涉及。不同分类方法因其分类原理不同,得到的分类结果也不同,在对图像进行土地利用时,各分类方法的优缺点不同,对各地物的分类精度也不同,本文在借鉴国内外关于图像监督分类的研究的基础上,对比9种监督分类方法,分析出精度较高的分类方法,得到对不同地物分类精度较高的分类方法。所得结论可以为合理选择分类方法提供参考,为以后更科学、精确地研究各地区的土地利用现状及其动态变化发挥作用。

2. 研究区概况

该实验的研究区是云南省禄丰县,该县的位置在 $24^{\circ}51' \sim 25^{\circ}30'N$ 和 $101^{\circ}38' \sim 102^{\circ}25'W$ 之间,处于滇中地区,楚雄彝族自治州东部,连双柏县和玉溪市易门县,西倚楚雄市、牟定县,北邻元谋县、武定县,是云南省楚雄彝族自治州的下辖县。该县总面积为3536平方千米,土地类型主要有耕地、水域、建设用地、草地、林地、未利用地、其他用地等。禄丰县大部分地区属于低纬高原,境内重山叠翠,河谷纵横,海拔1300~2754米。县境坐落在长江、红河两大水系上游分水岭地段。该县的气候是中亚热带季风气候。年均气温 $16.2^{\circ}C$,年平均降雨量930~950毫米,年无霜期322天,处于滇中湿润区与半湿润区的交汇地带,水分较充足,森林覆盖率较高,达70%。禄丰县的行政区划如图1所示。

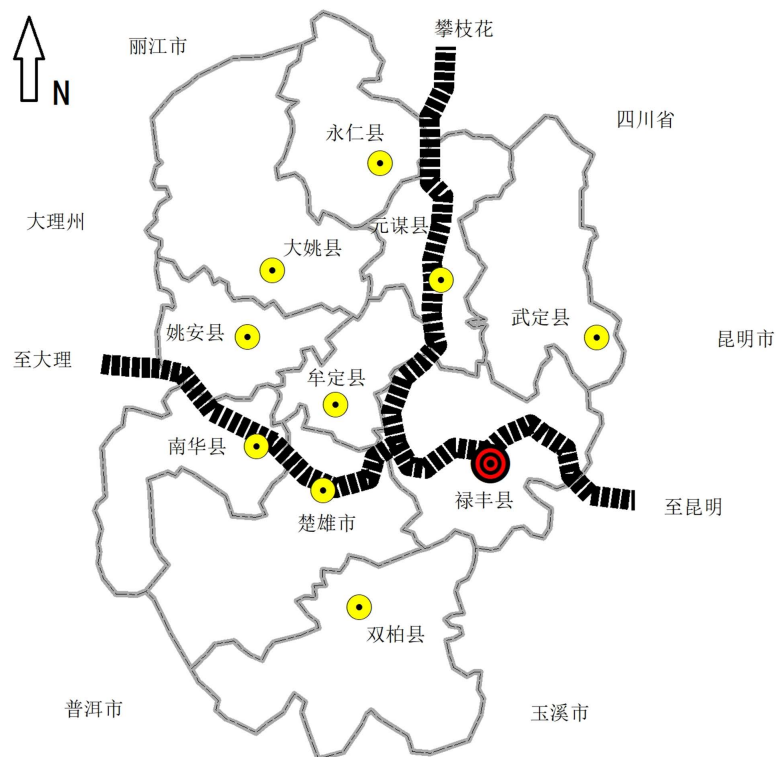


Figure 1. Location map of Lufeng County
图 1. 禄丰县方位图

3. 数据获取与处理

3.1. 数据源

本次实验用到的影像是来自美国的 Landsat 8 遥感卫星, 该卫星是在 2013 年发射的, 它有 OLI 陆地成像仪和红外传感器, 该卫星影像是多波段、高分辨率, 且遥感图像精度较高, 可以满足该次试验的精度要求。在地理空间数据云中获取本次实验数据, 对数据进行筛选, 选取了条带号和行列号为(130, 43)和(129, 43)的 2018 年 Landsat 8 OLI_TIRS 的两景影像数据。具体信息见表 1。

Table 1. Data sources
表 1. 数据源

时间	传感器	波段及分辨率
2018/4/25	OLI-TIRS	1-7,9_OLI 多光谱波段(30 米)
2018/3/1	OLI-TIRS	1-7,9_OLI 多光谱波段(30 米)

3.2. 数据处理

用 ENVI 遥感处理平台对获取的遥感影像进行一系列预处理步骤(大气校正、波段组合、影像拼接裁剪、图像增强)以便后续进行分类, 预处理后的结果图如图 2 所示, 从图中可以目视可分辨出城乡工矿用地、林地、耕地、水域等其他地类, 且研究区内没有云层覆盖, 方便选取训练样本, 符合影像分类的要求。

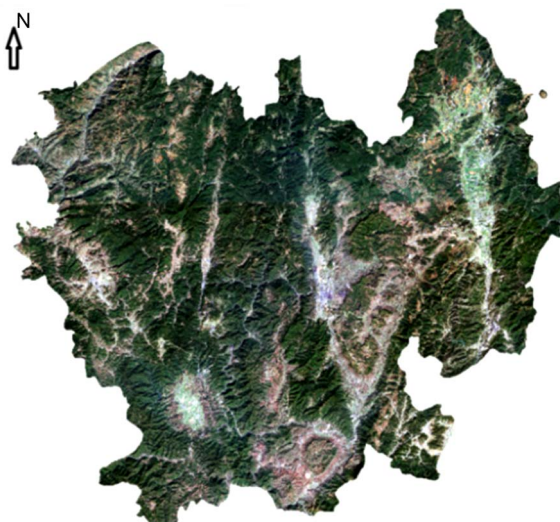


Figure 2. Pretreatment results
图 2. 预处理结果图

3.3. 建立训练样本

从 ENVI 软件中打开预处理后的影像数据, 以标准假彩色 432 波段组合显示, 通过目视可以辨别城乡工矿用地、耕地、林地、草地、水域、未利用地。从 RGB 彩色图像上获取 6 种 ROI 以创建感兴趣区, 分别为城乡工矿用地、耕地、林地、草地、水域、未利用地, 每一种感兴趣区对应的颜色分别为红、黄、绿、梅红、深蓝、天蓝, 并且在图像中画出相符的像元区域。把训练样本完整选择好, 选择一个分类样本, 再下载一幅研究区的谷歌 5 m 的高清影像图, 目视解译地物类别, 对照选择一个检验样本, 根据以上步骤选择一个用作检验的样本。样本点选择在研究区位置如图 3 所示。

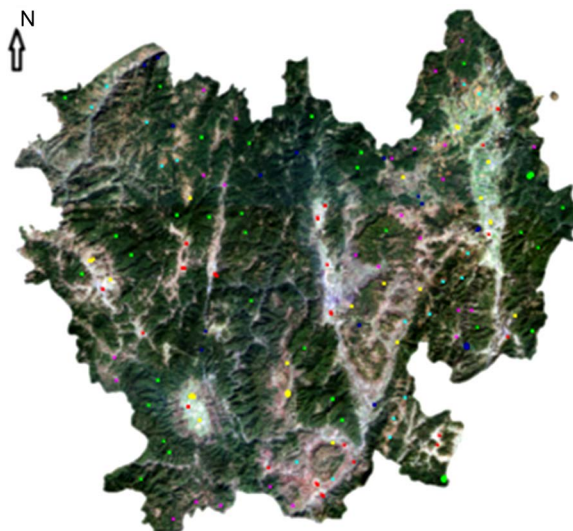


Figure 3. Sample location map
图 3. 样本位置图

样本值如表 2 所示, 接着对样本的可分离性进行计算, 得到两个参数值 Jeffries-Matusita 和 Transformed Divergence, 这两个参数的值在 0~2.0 之间, 这两个参数值是决定该样本是否可用的参考数值, 参数值越

靠近 2 说明样本的可分离性好, 可分离性好样本才合格可用。该次试验选择的分类样本和检验样本的可分离性均大于 1.9, 样本可用。

Table 2. Values of various samples
表 2. 各类样本数值

	样本数(point)					
	城乡工矿用地	未利用地	林地	草地	水域	耕地
分类样本	4183	1921	4543	1106	863	2145
检验样本	3323	767	4299	612	677	1969

4. 研究方法

借助遥感图像处理平台 ENVI, 对遥感影像进行预处理, 用多种监督分类方法进行土地利用分类, 为了保证所得数据的可比性, 选用同一景遥感影像以及同一训练样本进行试验, 得出分类结果, 对分类结果进行评价分析得到精度报表, 对所得各个方法的精度数据制作成表格进行统计分析, 对比择优出精度较高的分类算法。

4.1. 图像监督分类算法

监督分类又被称为“训练分类法”, 是用被确认类别的样本像元来识别未知类别像元的过程[3]。本研究所用的方法有以下几种: 平行六面体分类法、神经网络法[4]、波谱角法[5]、最小距离分类法、马氏距离法[5]、最大似然分类法[6]、支持向量机分类[7]、光谱信息散度分类法、二进制编码分类法。

4.2. 精度评价指标

分类结果进行精度验证后会得到对应的精度报表, 精度报表是通过矩阵的形式展现分类成果与检验资料之间的对应关系。从误差矩阵中可以得到精度系数, 各地类分类精度, 以及 Kappa 指数, 然后根据精度数值进行对比分析各种分类方法的优良。

1) Kappa 指数

该数值均小于等于或等于 1 (等于 1 时表示极好的分类效果), 数值越大表示分类质量好, 精度高, 数值该数值来验证该分类所有类别的准确性, 数值计算复杂, 但现已发展成为遥感影像分类的主要精度评价方法[8]。以下是它的计算公式

$$k_{\text{hat}} = \frac{n \sum_{i=1}^n x_{ii} - \sum_{i=1}^n (x_{i+} x_{+i})}{N^2 - \sum_{i=1}^n (x_{i+} x_{+i})} \quad (4-1)$$

式(4-1)中, n 是混淆矩阵中的总列数(即总的类别数); x_{ii} 是混淆矩阵中第 i 行、第 j 列上的像元数量(即正确分类的数日); x_{i+} 和 x_{+i} 分别是第 i 行和第 j 列的总像元数量; N 是总的用于精度评估的像元数量。 x_{ii} 表示应属于 j 类的像素被分到 i 类的像元总数。

2) 制图精度

它是指: 该种分类方法将整幅图像的像元是正确地划分到某一类型的像元数, 再和该类型的实际参考点总数相比成一个百分比, 也就是分类结果分到实际相属类别的百分比。该指标可以反映地物是否被准确分类, 同时可以用来比较各种分类方法的好坏, 制图精度与漏分误差互补, 制图精度越低, 漏测误差越高[9]。

3) 用户精精度

是指被准确划分到正确类别的像元总数值，然后再和整个影像中被划分到该类别的像元总数值相比得到一个百分数。这个指标是用来体现结果可信度的数值。用户精度与错分误差互补，用户精度越低，多测误差越高[10]。

5. 实验过程及结果

5.1. 执行监督分类

对预处理好的遥感图像进行图像分类，导入之前选择好的分类样本，分别选择上述 9 种方法进行分析。为了使结果具有对比性，使用了相同影像、相同训练样本、相同分类体系进行土地利用分类，最终分类结果如图 4 所示。

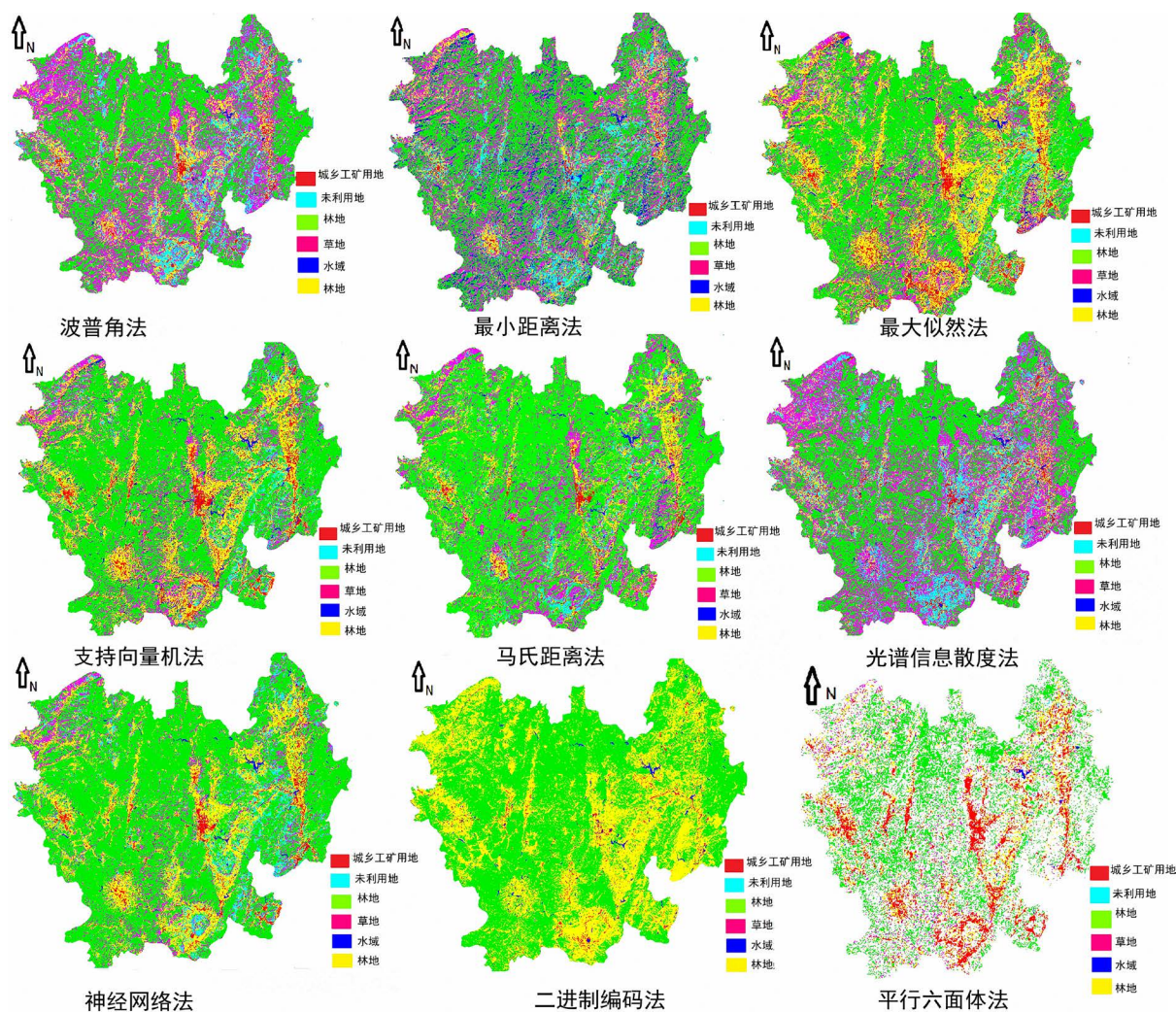


Figure 4. Classification results

图 4. 分类结果图

从以上这些分类图中，可以看出不同的分类方法得到的结果不同，不同的分类器，得到的各地类的面积也不同。经过目视对比结果图，可以初步看出平行六面体法的分类效果最差，漏分的很多地方，城

乡工矿用地和耕地被错分的面积较大,水域和未利用地以及草地没有准确的划分处理,错分的地方较多。最小距离分类法的分类结果也很差,图中大部分为草地和林地,划分得不准确,而且城乡工矿用地和耕地的面积划分得很不准确,错分的面积很大。

其次是二进制分类法,从分类结果图中可以看出图中只有五种地类,草地没有被分出来,而图中红色部分,也就是城乡工矿用地,该地物类型被错分和漏分的地方也较多。对比来看,最大似然,支持向量机的分类效果较好,在图中水域、城乡工矿用地、林地、耕地、未利用地、草地 6 种类别均可见,分类效果相比较好,划分得相对准确。

5.2. 分类结果精度分析评价

再次验证之前分类的准确性,对结果进行评价分析,确定结果的准确性和可比性。使用混淆矩阵进行评价分析,原理是通过将真实地表图像与处理后的图像进行比较来显示混淆矩阵中分类结果的准确性。再次选取一个训练样本用做精度验证,通过计算机计算得到精度报表,得到各分类方法的 kappa 系数如表 3 所示。

Table 3. Kappa coefficient of each classification method
表 3. 各分类方法 kappa 系数表

监督分类方法	Kappa 系数
光谱信息散度	0.5932
二进制分类法	0.5092
平行六面体法	0.3600
支持向量机	0.9259
最大似然法	0.9306
最小距离法	0.6569
神经网络法	0.9246
马氏距离法	0.8646
波谱角法	0.7517

表 3 中的数据显示,不同的算法得到的分类结果不同,从表中数据对比可得,kappa 系最高的是最大似然法的最高,平行六面体分类法的系数最低。由此得出,最大似然法和神经网络法、支持向量机的分类精度较高,其次为波谱角和马氏距离法,精度相对较低的 2 种方法是平行六面体法和二进制码分类法。

在区分地物方面,不同的分类方法得到的分类结果中,各地类的分类精度也不同。表 4 中数据显示的是各地类的漏分误差,表中的数值可以判断各地物类别的分类准确度,也可以判断各分类法的优势与缺点。它与制图精度互补,它的数值越低,也就代表误差越小,分类精度就高。从表 4 中可以看出:最大似然法对草地信息的提取较准确,精度较高,对耕地信息的提取上,也是最大似然法精度高,对林地和未利用地信息的提取,神经网络分类法的精度最高,对城乡工矿居民用地和水域信息的提取上,支持向量机的精度最高。

Table 4. Table of error coefficients for classification of ground features**表 4.** 各地物类别漏分误差系数表

分类器	漏分误差%					
	草地	耕地	林地	城乡工矿居民用地	水域	未利用地
光谱信息散度	14.54	69.53	3.19	39.09	12.41	76.01
平行六面体	78.45	75.37	53.92	39.51	35.75	74.45
支持向量机	12.42	10.06	1.91	2.98	3.40	20.99
最大似然	1.47	5.33	3.47	4.30	4.87	21.51
最小距离	68.44	53.97	17.34	13.83	7.91	10.93
神经网络	26.47	12.75	1.20	6.54	6.27	2.54
马氏距离	8.99	22.8	23.91	24.07	10.49	17.34
二进制	100.00	37.38	1.16	71.02	11.82	100.00
波谱角	12.58	61.91	5.05	12.01	33.09	4.61

表 5 的数据表示是各分类法对研究区不同地物的错分率，从表中可以看出各分类方法对各类地物的错分率。总体来看，支持向量机法、神经网络法在地物分类上的错分率相对较低，光谱信息散度法、最小距离法、波谱角分类法、马氏距离法的总体错分率都较高。

Table 5. Misclassification rate of different classification methods for different features in the study area**表 5.** 各分类法对研究区不同地物的错分率

分类器	错分率%					
	草地	耕地	林地	城乡工矿居民用地	水域	未利用地
光谱信息散度	59.17	17.92	12.29	30.18	34.76	82.99
平行六面体	17.37	13.70	0.50	25.28	0.46	0.51
支持向量机	10.22	13.19	2.83	4.98	1.65	0.98
最大似然	18.29	14.14	0.62	3.67	1.38	0.99
最小距离	96.46	12.80	0.22	0.33	30.29	33.08
神经网络	13.96	13.05	5.47	1.23	4.33	4.95
马氏距离	56.25	24.45	3.77	3.33	5.90	26.87
二进制	0.00	65.54	24.42	31.85	42.26	0.00
波谱角	56.15	25.30	8.25	4.51	8.11	42.77

由此，通过对各个分类方法不同精度系数以及土地利用分类结果图的对比分析可得以下结论：

1) 平行六面体的 Kappa 系数较低，从分类结果图来看，分类效果较差，耕地、未利用地和草地的漏分率较高。各地物的分类精度都较低，主要是由于平行六面体所选类别均值决定校准差阈值，尺度又

由标准差阀决定, 所选择类别的划分以及均值的计算导致它分类结果较差, 精度较低。因此, 在进行图像分类时, 不建议用平行六面体法。

2) 二进制编码分类法的 $Kappa$ 系数也较低, 用二进制编码进行土地利用的分类结果图效果也不理想, 分类结果图效果较差, 耕地和水域的错分率较高, 草地没有被准确的划分出来, 原本是草地的地方被划分成了林地, 一些原本是未利用地被划分成了耕地, 由此, 在进行图像分类时, 二进制编码同样不适用。

3) 最大似然法的 $Kappa$ 系数相对较高, 体现在分类结果错分、漏分的地方较少, 误差小, 各地类的分类精度都较高, 尤其是草地的分类精度, 比较适合用于提取地物信息。该方法的分类原理是计算给定像元属于训练样本的可能性, 并将该像元分类为最相似的类别, 因此分类精度较高。它的土地利用分类结果比较接实际, 适用于土地利用及土地覆盖方面的地类分类之中。

4) 支持向量机法的 $Kappa$ 系数较高, 从数值来看, 效果好, 错分、漏分率较低, 其中, 林地的分类精度最高。该方法的精度高是由于它的算法特别, 会将各类别之间的差距最大化, 该方法操作简单, 具有较好的推广性和较高的分类精度。

5) 神经网络法的 $Kappa$ 系数较好, 该方法的分类效果与上述两种分类方法相差不大, 分类效果均比较理想, 漏分和错分的地方较少, 各地类分类精度高, 在对林地提取时, 神经网络分类法是较好的选择。由于神经网络分类方法是非线性的, 可以处理复杂的数据并准确的识别模型。所以分类效果较好。但神经网络的操作过程不同于其他的分类方法, 操作起来比一般的分类复杂, 但分类精度比一般的分类方法精度高。

6) 马氏距离法的精度数值不是很高, 且草地的错分率偏高。马氏距离法与最大似然法有些相似, 同样是距离分类器, 操作简单, 分类起来较快, 但分类精度一般。

7) 波谱角分类法得到的结果, 精度不高, 其中, 草地和耕地被错分的地方都很多, 这两种地类的精度都较低。在结果图中看出, 很多城乡工矿用地被错分为了未利用地。对于对分类精度要求高的图像分类, 波谱角分类法不适用。

8) 光谱信息散度的分类精度差, $Kappa$ 系数为 0.5932。分类效果较差, 这种方法适用于高光谱图像, 它的原理是对比高光谱图像中两种相不同像元的相似度, 它也是根据相似度确定归属。在它的分类结果中可以看出。耕地没有很好的被划分出来, 很多耕地被错分为了未利用地。

9) 最小距离法的精度数值也较差, 各地类分类精度不高, 草地、未利用地错分、漏分率高。最小距离法也是一种距离分类器。在分类结果图中可以看出, 草地、耕地、城乡工矿用地都划分得不准确。对于对分类精度要求高的图像分类, 最小距离分类法同样不适用。

6. 结语

遥感图像的分类精度一直是学者们研究的热点问题, 在对图像进行分类时, 要考量原始图像的精度以及要进行分类的类别等因素。本实验用 Landsat 8 卫星影像进行土地利用分类, 可以达到很好的分类效果。用同一景影像以及同一训练样本进行分类, 保证了结果的可比性, 分类方法不同, 得到的结果也大不相同, 得到的结论是: 最大似然法分类效果最好, 精度高, 该方法操作起来也相对简单, 速度快。此外, 还有神经网络分类法和支持向量机法这两种分类方法的精度也较高。平行六面体法以及二进制编码分类法的精度相对而言较低, 得到的分类结果不理想, 由此在进行图像分类时, 建议使用最大似然法, 它的分类效果较好。

从各地类的分类精度来看, 最大似然法对草地信息的提取较准确。在耕地信息的提取上, 同样是最大似然法效果好, 对林地和未利用信息的提取, 神经网络分类法的精度最高。对城乡工矿用地和水域信息的提取, 支持向量机的精度最高。

因此, 在进行图像分类和特征信息提取时, 可以根据不同的需要选择相适的分类方法, 从而有效提高分类结果的准确性, 得到理想的效果。

基金项目

本项目研究得到云南省大学生创新创业训练计划项目“基于 RS 和 GIS 的富源县土地利用动态研究”资助。

参考文献

- [1] 朱建华, 刘政凯, 俞能海. 一种多光谱遥感图像的自适应最小距离分类方法[J]. 中国图象图形学报, 2000, 5(1): 22-24.
- [2] 吴健平. 遥感数据监督分类中训练样本的纯化[J]. 国土资源遥感, 1996, 26(1): 36-41.
- [3] 童威. 基于 ENVI 的遥感影像监督分类方法的研究[J]. 科技创新与应用, 2019(23): 12-15.
- [4] 徐涵秋, 唐菲. 新一代 Landsat 系列卫星: Landsat8 遥感影像新增特征及其生态环境意义[J]. 生态学报, 2013(33): 3249-3257.
- [5] 李石华, 王金亮, 毕艳, 等. 遥感图像分类方法研究综述[J]. 国土资源遥感, 2005(2): 1-6.
- [6] 席武俊. ENVI 5.3 遥感实验教程[M]. 昆明: 云南民族出版社, 2017.
- [7] 靳文戟, 刘政凯. 多类别遥感图像的复合分类方法[J]. 环境遥感, 1995, 10(4): 298-302.
- [8] Cohen, J. (1960) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, **20**, 37-46. <https://doi.org/10.1177/001316446002000104>
- [9] 邓书斌, 陈秋锦, 杜会建, 等. ENVI 遥感图像处理方法[M]. 北京: 高等教育出版社, 2014.
- [10] 夏文韬, 王莺, 冯琦胜, 等. 甘南地区 MODIS 土地覆盖产品精度评价[J]. 草业科学, 2010, 27(9): 11-18.