

训练样本对农作物遥感分类的精度影响研究

潘洪涛,王 轩,王晓飞

(黑龙江大学 电子工程学院,黑龙江 哈尔滨 150080)

摘 要: 为了比较好地研究和分析训练样本的数量和质量对农作物分类精度的影响,选取黑龙江省海伦市作为研究所需要的实验区,以 Landsat 8 遥感影像作为数据源,利用最大似然、神经网络、支持向量机 3 种分类方法分别去研究训练样本数量与质量对分类精度的影响,并且对 3 种分类方法进行了多次实验。最终的研究结果表明:(1) 在训练样本质量相对恒定下,同一种分类方法对相同数量的训练样本的响应程度以及不同分类方法对训练样本数量的响应程度是不同的,并且分类精度存在不同程度的波动,随着训练样本数量的增加,这种波动会减小,当训练样本的数量达到一定程度,分类精度的均值将趋于相对稳定;(2) 在训练样本数量恒定下,同一种分类方法以及不同种分类方法对相同质量等级的训练样本的响应程度是不同的;同一种分类方法对不同质量等级的训练样本响应程度也是不同的。

关键词: 遥感图像分类; 农作物分类; 训练样本; 分类精度

中图分类号: TP751.1 **文献标志码:** A **DOI:** 10.3788/IRLA201746.S126003

Study on the effect of training samples on the accuracy of crop remote sensing classification

Pan Hongtao, Wang Xuan, Wang Xiaofei

(School of Electrical Engineering, Heilongjiang University, Harbin 150080, China)

Abstract: In order to study and analyse the influence the number of and quality of the training samples on the classification accuracy better, Helen city in Heilongjiang Province was chosen as the research required experimentation area, using Landsat 8 remote sensing images as the data source, the effects of the number and quantity of training samples on the classification accuracy were studied respectively by using the maximum likelihood, neural network and support vector machine three kinds of methods, and several experiments were made on these three kinds of classification methods. The final result shows that: (1) when the training sample quality is relatively constant, the degree of response of the same classification method to the same number of training samples as well as the degree of response of the different classification methods to the number of training samples are different, and the classification accuracy has different degree of volatility, with the increase of the number of training samples, the volatility will decrease, when the number of training samples reaches a certain degree, the mean of classification accuracy will tend to be relatively stable; (2) when the number of training samples is

收稿日期:2017-06-05; 修订日期:2017-07-09

基金项目:国家重点研发计划(2016YFB0502502);黑龙江省教育厅科学技术研究项目(12531493)

作者简介:潘洪涛(1975-),男,讲师,硕士,主要从事遥感图像处理与应用方面的研究。Email:1998013@hlju.edu.cn

通讯作者:王晓飞(1977-),男,副教授,博士,主要从事遥感图像处理与应用方面的研究。Email:nk_wxf@hlju.edu.cn

constant, the same classification methods as well as the different classification methods have different degree of response to the training samples of the same quality grade; the degree of response of the same classification method to the different training samples quality level is also different.

Key words: remote sensing image classification; crop classification; training samples; classification accuracy

0 引言

农业遥感在遥感领域具有广泛的应用并且也是该领域最重要的技术之一。在现代农业发展过程中,快速、及时并且准确的获得农作物的相关信息是至关重要的,不同农作物之间的有效识别与分类,直接影响着社会的经济、粮食的安全、生态功能以及相关政策的制定^[1-2],其分类后结果精度的准确性和可靠性对信息的处理和分析有着重要的影响。因此,在农作物遥感分类过程中如何获取比较高的分类结果精度,成为农作物遥感分类过程中研究的核心问题之一。影响农作物遥感分类精度的因素很多,其中训练样本是影响遥感分类精度的一个重要的因素,训练样本对农作物分类精度影响的两个方面:数量和质量。陶秋香等^[3]研究了植被高光谱分类中训练样本的选择方法。薄树奎等^[4]利用统计学相关理论,在基于面向对象方法下得到了确定训练样本数量方法的相关依据。赵慧等^[5]探讨了训练样本、辅助数据以及分类技术对土地利用/覆盖分类精度的影响。朱秀芳等^[6]在其研究过程中得到不同的分类器分类的结果精度随着训练样本数量的不断的增加而增加。阎静等^[7]采用神经网络对不同样本点分类得到的结果进行了研究与分析。Arai 基于最大似然法,提出了一种纯化训练样本的方法,通过研究表明:进行纯化后的训练样本得到的分类精度有了大幅提高^[8]。Peddle 等^[9]将多光谱图像与研究区域的纹理信息以及 DEM 结合起来综合应用,在分类之后得到的分类结果精度会有大幅度提高。Nelson 和 Van 等利用统计学方法得出了遥感影像分类过程中需要的最小训练样本数量的计算公式^[10-11]。Serra^[12]等通过 ISOMN 和 CLSMIX 两种不同的分类方法对农作物进行分类,得到的分类精度都比较高。

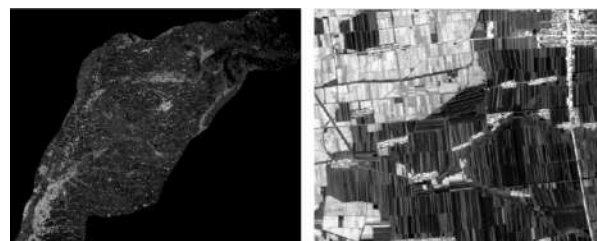
文中针对农作物分类过程中训练样本与分类精度之间的关系问题,选择典型实验区,以 Landsat8 遥感影像作为数据源,结合 Google Earth 影像数据和

野外 GPS 数据,选取不同数量的训练样本以及不同质量等级的训练样本,然后采用最大似然、神经网络、支持向量机 3 种分类方法进行反复实验,以期找到训练样本与分类精度之间的关系以及训练样本对不同分类方法敏感性程度,为遥感影像分类提供一定的实验依据。

1 研究区的选择与数据预处理

1.1 研究区的选择

海伦市是黑龙江省绥化市下辖的一个县级市,地处黑龙江省中部(126°14′-127°45′E,46°58′-47°52′N),地形为丘陵、漫岗,平均海拔 239 m。海伦市境内资源丰富,被称作是世界三块黑土地之一,素有“粮仓”之称。西部平原的主要粮食作物有大豆、玉米以及水稻等,也种植大量的经济作物如亚麻和烤烟等。北部山区主要盛产木材、中药以及山产品,它是国家重要的商品粮基地县,为文中的研究提供了理想的实验条件。文中选取 Landsat8 遥感影像中海伦市农作物特征信息比较明显的西南地区某区域(原始影像行带号为 118/27,影像获取时间为 2014-09-17,影像空间分辨率为 30 m,大小为 500 行×500 列)作为实验区,实验区域影像在整副海伦市遥感影像上的位置如图 1 所示。



(a) 研究区在海伦市影像中的位置 (b) 研究区域影像(500×500)
(a) Location of the study area in Helen's image (b) Study area image (500×500)

图 1 研究区 OLI 影像

Fig.1 OLI image of the study area

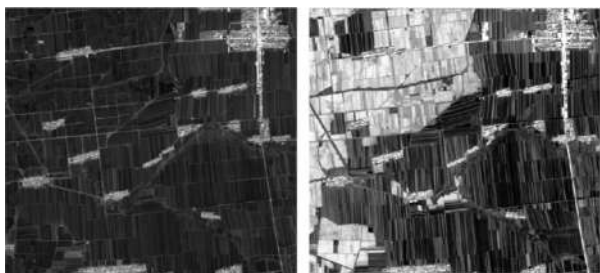
1.2 数据预处理

执行监督分类之前,文中对数据源做预处理,首先对原始影像做辐射校正以及大气校正,然后对遥感影像做几何校正,在几何校正过程中,利用 Google Earth 历史无偏移影像对该影像进行几何精校正,也就是利用 Google Earth 影像上的 GCP 控制点去校正该影像,最后将几何精校正过后的影像利用所研究区域的矢量边界裁剪出所需要研究的区域,从而获得满足研究需要的一幅遥感影像。

2 研究的方法

2.1 训练样本的选择方法

在执行监督分类的过程中,一般情况下,每一种地物类别所选取的训练样本数据应该包括 $10n$ (其中 n 为特征维数)个以上的样本点,这样才能有足够的数来保证计算某些算法中所需要的方差-协方差矩阵^[13]。通过查阅资料与分析,得到在所选取的研究区内主要种植的农作物有水稻、玉米、大豆。在 9 月中旬水稻呈现黄色,玉米的收割期在 10 月 1 日前后,大豆的收割期在 9 月初。图 2 是不同时间段研究区影像图,从图 2(a)中可以看出 3 种农作物处于生长期,通过与图 2(a)的对比以及对相关资料的分析研究,



(a) 2014 年 7 月 12 日研究区影像图 (b) 2014 年 9 月 17 日研究区影像图

(a) Study area image on July 12, 2014 (b) Study area image on September 17, 2014

图 2 不同时间段研究区影像图

Fig.2 Image of the study area in different time periods

可以从图 2(b)中分析得到呈现黄色的为水稻,呈现绿色的为玉米,此时大豆已经收割完毕,所以呈现褐色的为收割大豆后的耕地。为了研究需要最终将研究区分为包含水稻、玉米、耕地、其他类 4 类不同类型的数据区图像,利用已有的野外 GPS 和 Google Earth 影像数据对研究区域进行反复的目视纠正和

处理,最终得到研究区内总共像元 250 000 个,其中包含混合像元的水稻样本有 60 206 个,包含混合像元的玉米样本有 61959 个,包含混合像元的耕地样本有 46 031 个,包含混合像元的其他类样本有 81 804 个。

训练样本的分离度检验是在训练样本选取结束后进行的,文中主要是通过 J-M 距离和转换分离度来衡量训练样本(ROI)的可分离性,训练样本的 J-M 距离和转换分离度如表 1 和表 2 所述,从表中可以看出,训练样本的分离度能够满足分类实验的要求。

表 1 训练样本的 J-M 距离对比表

Tab.1 J-M distance table of training sample

Type	Rice	Corn	Land	Others
Rice	-	1.999 9	1.997 9	1.865 9
Corn	1.999 9	-	1.999 6	1.967 7
Land	1.997 9	1.999 6	-	1.852 2
Others	1.865 9	1.967 7	1.852 2	-

表 2 训练样本的转换分离度对比表

Tab.2 Conversion separation table of training sample

Type	Rice	Corn	Land	Others
Rice	-	2.000 0	2.000 0	2.000 0
Corn	2.000 0	-	2.000 0	2.000 0
Land	2.000 0	2.000 0	-	2.000 0
Others	2.000 0	2.000 0	2.000 0	-

2.2 训练样本数量的控制方法

文中是在控制训练样本质量的前提下,研究与分析训练样本数量对分类精度的影响。在所研究的区域内,不同地物类别的面积存在一定程度的差异,为了能够使小面积的不同物种也能够获取足够多的样本量,所以文中在选取所需要的样本的时候,以不同地物类别为依据,采用分层抽样的方法抽取不同数量(0.05%、0.1%、0.2%、0.5%、1%、2%、5%、10%)的样本数据作为训练样本,为了尽可能地减小随机误差对分类精度的影响,文中采用最大似然、神经网络、支持向量机 3 种分类方法对每个数量下选取的训练样本均做 10 次重复实验,然后求得 10 次实验结果的平均值作为每个训练样本数量下最终的分类结果精度。

2.3 训练样本质量的控制方法

文中抽取各个地物类别数量的 1%作为恒定样

本数量, 由于不同地物类别的面积存在一定程度的差异, 为了能够使小面积的不同物种也能够获取足够多的样本量, 采用分层抽样的方法抽取各个地物类别纯净的训练样本和混合的训练样本不同数量的组合 (20%+80%、40%+60%、60%+40%、80%+20%、100%+0%) 去研究训练样本的质量对分类精度的影响, 为了研究方便, 在下文叙述中将 5 种不同的数量组合统称为 5 个质量等级, 依次为 1~5 质量等级, 为了尽可能避免随机误差对分类结果的影响, 文中采用最大似然、神经网络、支持向量机 3 种分类方法在各个不同质量等级下均做 10 次实验, 然后求得 10 次实验结果的平均值作为每个不同质量等级下最终的分

2.4 精度评价的指标和方法

文中采用制作地表真实感兴趣区来获取验证训练样本。为了统一误差, 采用随机抽样的方法, 选取一组各个类别总样本量的 10% 作为验证样本, 利用 Kappa 系数和总体分类精度两个指标对分类结果进行评价。

(1) 总体分类精度 (Overall Accuracy): 是由所有被准确分类的像元数与像元总数的比值计算得到的, 地物类别的真实图像或者真实的 ROI 对像元的真实分类具有一定的限定性。所有地表真实参考源总的像元数就是像元总数。

(2) Kappa 系数: 它也是计算分类精度的一种方法。它是离散型并且多元化的综合分析方法, 像元数对它没有影响, 它的结果完全可以利用混淆矩阵中的值得到, 这种优势是其他精度评价方法不具备的。其公式^[14]为:

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

式中: r 为混淆矩阵中的总列数; x_{ii} 为混淆矩阵中第 i 行、第 i 列上的值, 也就是被正确分类的数目; x_{i+} 、 x_{+i} 分别代表的是第 i 行和第 i 列上总像元的数量; N 代表的是混淆矩阵中参与统计的所有像元数目。

3 研究结果与分析

3.1 训练样本数量对农作物遥感分类精度的影响

图 3~图 5 是不同数量的训练样本下, 3 种分类

方法 10 次实验结果的总体分类精度和 Kappa 系数图。表 3 是 3 种分类方法在不同样本量抽样比例下 10 次实验得到的分类结果精度的均值。

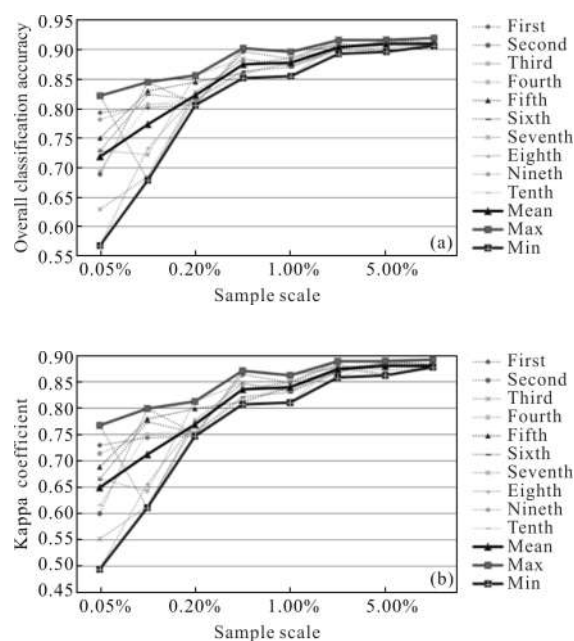


图 3 最大似然分类的总体精度和 Kappa 系数图

Fig.3 Overall classification accuracy and Kappa coefficients of the maximum likelihood classification

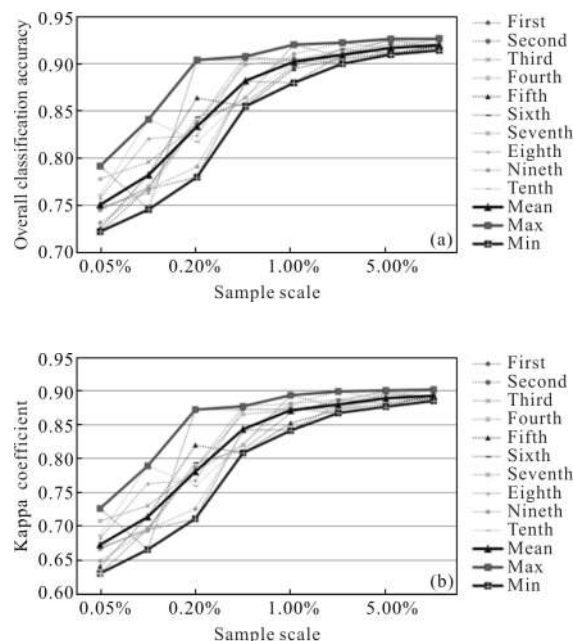


图 4 神经网络分类的总体精度和 Kappa 系数图

Fig.4 Overall classification accuracy and Kappa coefficients of the neural network classification

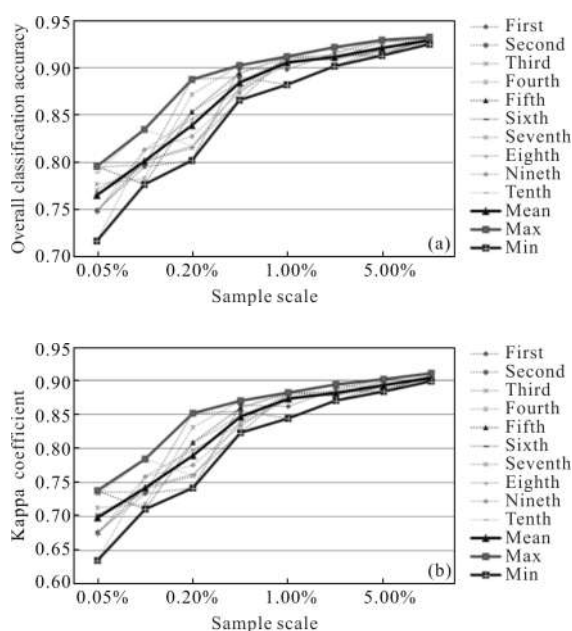


图 5 支持向量机分类的总体精度和 Kappa 系数图

Fig.5 Overall classification accuracy and Kappa coefficients of the support vector machine classification

表 3 3 种不同的分类方法在不同样本量抽样比例下分类结果精度的均值

Tab.3 Mean of the accuracy of the classification results by using three classification

Sample ratio	Maximum likelihood classification		Neural network classification		Support vector machine classification	
	Overall classification accuracy	Kappa coefficient	Overall classification accuracy	Kappa coefficient	Overall classification accuracy	Kappa coefficient
0.05%	71.81%	0.648 3	74.98%	0.672 0	76.46%	0.697 2
0.1%	77.26%	0.711 3	78.14%	0.713 1	80.06%	0.741 2
0.2%	82.21%	0.768 4	83.29%	0.779 9	83.85%	0.788 8
0.5%	87.42%	0.835 3	88.14%	0.842 5	88.37%	0.846 1
1%	87.72%	0.839 4	90.16%	0.870 4	90.50%	0.872 5
2%	90.35%	0.873 9	90.90%	0.878 9	91.09%	0.881 2
5%	90.89%	0.880 1	91.62%	0.888 3	92.06%	0.892 6
10%	90.95%	0.880 9	91.92%	0.892 0	92.86%	0.903 3

从图 3~图 5 和表 3 可以得出以下结论。

对于同一种分类算法而言：

(1) 在抽取的训练样本数量相同的情况下,每个分类方法 10 次分类的 Kappa 系数和总体分类精度都存在不同程度的波动,当训练样本的数量比较少的情况下,10 次实验得到的 Kappa 系数和总体分类

精度的波动范围是比较大的,但是随着训练样本的不断增多,Kappa 系数和总体分类精度的波动范围随之逐渐减小并且平均值是趋于相对稳定的,如果在实验的过程中以某一次的分类结果作为最终结果的话,那么得到的分类结果将会存在不确定性和随机误差,所以在实际的操作过程中,要取多次分类结果的平均值作为最终的结果,这样可以尽可能地消除随机误差并且使得最终结果具有很好的可靠性。

(2) 10 次实验分类结果的 Kappa 系数和总体分类精度的极差值(最大值减去最小值)随着抽取的训练样本比例的增加而逐渐减少,可以看出在研究的过程中,增加训练样本的数量可以降低单次分类引起的随机误差,并且可以使得分类的结果更加稳定。

对于不同的分类算法而言：

(1) 相同的抽样样本在不同的分类算法下得到的 Kappa 系数和总体分类精度也是存在很大差异,从表 3 可以看出在相同的抽样比例下,支持向量机分类的 Kappa 系数和总体分类精度的平均值是最高的(Kappa 系数范围为 0.697 2~0.903 3,总体分类精度范围为 76.46%~92.86%),其次是神经网络分类(Kappa 系数范围为 0.672~0.892,总体分类精度范围为 74.98%~91.92%),最后是最大似然法分类(Kappa 系数范围为 0.648 3~0.880 9,总体分类精度范围为 71.81%~90.95%),这 3 种分类算法相对来说分类的精度都是比较高的。

(2) 在抽取的训练样本数量相同的情况下,10 次实验分类结果的 Kappa 系数和总体分类精度的波动程度是有差异的,波动相对较小的是支持向量机分类算法(抽取的样本量超过 2%时,Kappa 系数稳定在 0.87 以上,总体分类精度在 90%以上),可以得到的是使用支持向量机分类算法可以得到比最大似然法分类和神经网络分类更高的分类精度并且其稳定性也是最好的。如果将 10 次实验的分类结果看作是不同的操作员进行的实验的话,那么就可以得到支持向量机受人为因素影响是比较小的。

(3) 随着训练样本数量的不断增加,支持向量机的 Kappa 系数和总体分类精度的平均值有不断增加的趋势,而最大似然法分类和神经网络分类的 Kappa 系数和总体分类精度的均值都是首先不断增加,当训练样本的数量达到一定程度的时候,其 Kappa 系数和总体分类精度的均值则趋于稳定,其

中最主要的原因是支持向量机相对于一般的分类方法而言,它对训练样本的响应程度是不同的,而对于基于统计的传统分类方法而言,当训练样本的数量达到统计所需要的数量并且能够精确的描述研究的区域内所有像元的分布规律的时候,不断地增加训练样本的数量对分类的精度是没有很大影响的,但是对于基于统计学习的小样本机器的支持向量机分类算法则不同,它分类的关键是寻找最优超平面并且计算出各类之间的距离。在支持向量机分类算法中对分类精度起影响的是支持向量的那部分像元^[15],支持向量的那部分像元会随着训练样本的不断增多而增加,所以支持向量机分类算法的 Kappa 系数和总体分类精度的平均值会随着训练样本数量的增加而不断增加。

3.2 训练样本质量对农作物遥感分类精度的影响

图 6~图 8 是在不同质量等级的训练样本下,3种

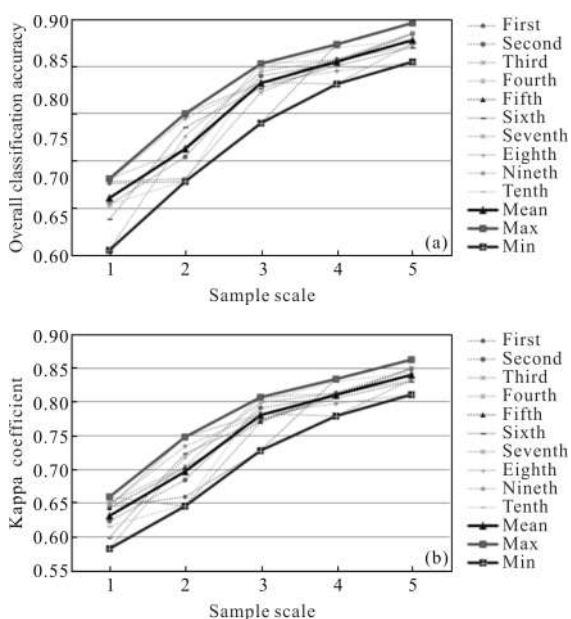


图 6 最大似然分类的总体精度和 Kappa 系数图

Fig.6 Overall classification accuracy and Kappa coefficients of the maximum likelihood classification

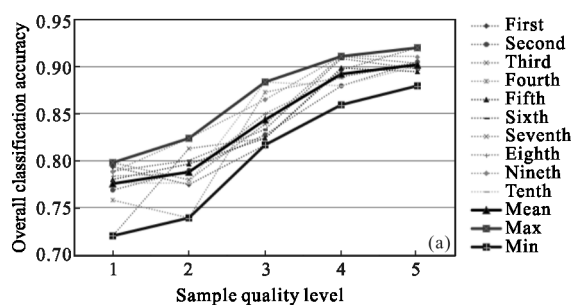


图 7 神经网络分类的总体精度和 Kappa 系数图

Fig.7 Overall classification accuracy and Kappa coefficients of the neural network classification

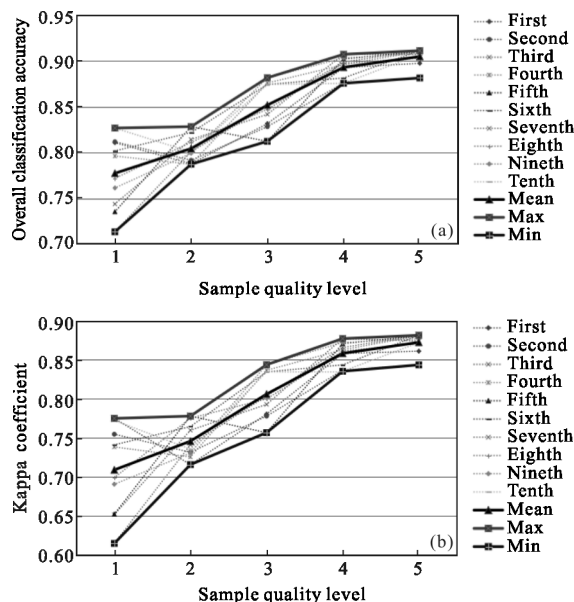


图 8 支持向量机分类的总体精度和 Kappa 系数图

Fig.8 Overall classification accuracy and Kappa coefficients of the support vector machine classification

分类方法 10 次实验结果的总体分类精度和 Kappa 系数图。表 4 是在不同质量等级下,最大似然法、神经网络法、支持向量机 3 种不同的分类方法 10 次实验得到的分类结果精度的均值。

从图 6~图 8 和表 4 可以看出,在分层随机抽样过程中,抽取相同数量不同质量等级的训练样本得到的 Kappa 系数和总体分类精度存在差异并且都是介于最大值和最小值之间的,说明相同样本数量下不同质量等级的训练样本对遥感分类精度有影响,通过对图 6~图 8 和表 4 分析可以得到以下结论:

(1) 相同质量等级的训练样本下,同一种分类方法对其的响应程度是不同的,从图中可以看出在同

一质量等级的训练样本下,采用同一分类方法 10 次实验得到的分类结果精度存在一定程度的波动。

(2) 同一分类方法对不同质量等级的训练样本得到的 Kappa 系数和总体分类精度是不同的,从图6~图 8 可以看出,最大似然法分类、神经网络分类、支持向量机分类 3 种分类方法随着训练样本质量等级从 1~5(纯净训练样本数量逐渐增加)变化,它们的总体分类精度和 Kappa 系数的均值逐渐增加,说明在训练样本的选取过程中,纯净训练样本数量越多,分类之后得到 Kappa 系数和总体分类精度也相对越高。

(3) 不同分类算法对相同质量等级的训练样本得到的 Kappa 系数和总体分类精度是不同的,从表4可以看出,在第 1 质量等级下最大似然法分类得到的总体分类精度和 Kappa 系数的均值分别为 71.05%和 0.630 8,神经网络法分类得到的总体分类精度和 Kappa 系数的均值分别为 77.51%和 0.688,支持向量机分类得到的总体分类精度和 Kappa 系数的均值分别为 77.76%和 0.708 9,其他质量等级下亦是如此,说明在同一质量等级下,支持向量机分类结果的精度相比较于其他两种分类方法而言更高一些,其次是神经网络法分类,最后是最大似然法分类。

表 4 三种不同的分类方法在不同质量等级下分类结果精度的均值

Tab.4 Mean of the accuracy of the classification results by using three classification methods under different quality levels

Quality level	Maximum likelihood classification		Neural network classification		Support vector machine classification	
	Overall classification accuracy	Kappa coefficient	Overall classification accuracy	Kappa coefficient	Overall classification accuracy	Kappa coefficient
1	71.05%	0.630 8	77.51%	0.688	77.76%	0.708 9
2	76.23%	0.696 3	78.75%	0.710 1	80.46%	0.745 8
3	83.19%	0.779 9	84.31%	0.790 2	85.2%	0.806 1
4	85.45%	0.809 8	89.18%	0.855 6	89.32%	0.858 4
5	87.72%	0.839 4	90.16%	0.870 4	90.5%	0.872 5

4 结果与讨论

4.1 结论

文中主要研究了 Landsat8 遥感影像中训练样本

的数量和训练样本的质量对农作物遥感分类精度的影响,在研究的过程中,利用已有的野外 GPS 和 Google Earth 影像数据,在研究区中选取所需要的不同数量和不同质量等级的训练样本集,分别采用最大似然、神经网络、支持向量机 3 种分类方法对选取的训练样本进行分类,最后利用 Kappa 系数和总体分类精度对分类结果进行精度评价分析,最终得到以下结论:

(1) 通过对所实验区域的研究和分析,得出训练样本数量对不同分类方法分类精度的影响:1) 3 种不同的分类方法对相同数量的训练样本的响应程度是有差异的;2) 在相同数量的训练样本下,同种分类方法对其的响应程度也是有差异的,并且分类精度存在一定程度的波动,随着样本量的增加,波动程度会相应地减少;3) 随着训练样本数量的增加,3 种不同分类方法精度的均值都随之增加,当训练样本的数量达到一定比例时,神经网络分类和最大似然分类精度的均值趋于稳定,而支持向量机分类精度的均值出现随着训练样本的增加而增加的趋势,所以在实际的应用中,应该多次实验取平均值作为最终分类的结果;4) 在 3 种分类方法的比较之下,支持向量机的稳定性较好并且分类精度最高,所以在实际的分类过程中,应该首先考虑选取支持向量机分类方法进行分类。

(2) 通过对实验区域的研究和分析,得出不同训练样本质量等级对不同分类方法分类精度的影响:1) 3 种不同的分类方法对相同质量等级的训练样本响应程度是不同的,其中支持向量机的分类精度相比较于其他两种分类方法而言更高一些,其次是神经网络法分类,最后是最大似然法分类;2) 在相同质量等级的训练样本下,同一种分类方法对其的响应程度也是不同的,分类精度存在一定程度的波动;3) 随着训练样本质量等级从 1~5 (纯净训练样本数量逐渐增加) 变化,3 种不同分类方法精度的均值逐渐增加,所以在实际的分类过程中尽可能选取纯净的训练样本作为研究的样本。

4.2 讨论

文中主要研究的是训练样本的数量和训练样本的质量对农作物分类结果的影响,通过研究得到了在 Landsat8 遥感影像中,训练样本的数量和质量对分类结果影响的结论,但是由于时间和研究水平的

限制,还有许多地方需要改进和完善,今后还需要从以下几个方面进行研究:

(1) 文中仅在 Landsat8 遥感影像上进行了研究与分析,对于其他类型的遥感影像比如 MODIS 等,本文并没有进行研究与分析,是否其他类型的影像也具有相同的规律与结论,还需要在未来学习中进一步研究与分析。

(2) 文中仅在最大似然法、神经网络法、支持向量机 3 种分类方法下,得到了训练样本的数量和质量对分类精度的影响,对于其他的分类方法是否有相同的规律,还有待进一步研究和分析。

(3) 针对某一精度需求,如何才能快速并且准确的得到训练样本的数量,是否能够建立一个训练样本数量的计算公式,在训练样本的选取过程中,能否得到一个判别方法尽可能的选取与实际地物类型相一致的训练样本,因为训练样本选取好与不好对分类精度的影响很大,如果样本选取的不准确,则训练出来的样本就不能很好的描述研究区域的特征,所以在分类的过程中就很难得到正确的分类精度,所以在训练样本的选取过程中必要时需要实地的考察。

(4) 对于文中通过目视选取的方法,在不同物种边界处得到混合像元与纯净像元组合从而得到不同质量等级的训练样本,相比较通过其他方法得到的不同质量等级的训练样本对分类精度的影响有何差异,值得进一步研究和分析。

(5) 如果选取的影像质量不能符合分类的要求,那么训练样本的选取就极为困难,也给分类精度带来了很大的误差,所以能否建立一个标准并且完善的方法对待分类的影像进行评价分析,以便为后续训练样本的选取和分类方法的选择提供依据,使得分类结果更加准确。

参考文献:

- [1] Ding Ling, Tang Ping, Li Hongyi. Dimensionality reduction and classification for hyperspectral remote sensing data using ISOMAP[J]. *Infrared and Laser Engineering*, 2013, 42(10): 2702–2711. (in Chinese)
- [2] Duan Yunsheng, Zhang Dongyan, Huang Linsheng, et al. Comparison of hyperspectral and imagery characteristics of freezing stress and normal wheat [J]. *Infrared and Laser Engineering*, 2015, 44(7): 2218–2223. (in Chinese)
- [3] Tao Qiuxiang, Zhang Lianpeng, Li Hongmei. The methods for selecting training samples in vegetation classification based on hyperspectral remote sensing [J]. *Remote Sensing for Land & Resources*, 2005, 2(64): 33–45. (in Chinese)
- [4] Bo Shukui, Ding Lin. The effect of the size of training sample on classification accuracy in object-oriented image analysis [J]. *Journal of Image and Graphic*, 2010, 15(7): 1106–1111. (in Chinese)
- [5] Zhao Hui, Wang Yunjia. Research on the factors affecting the classification accuracy of ETM remote sensing image land cover/use [J]. *Remote Sensing Technology and Application*, 2012, 27(4): 600–608. (in Chinese)
- [6] Zhu Xiufang, Pan Yaozhong, Zhang Jinshui, et al. The effects of training samples on the wheat planting area measure accuracy in TM Scale (I): the accuracy response of different classifiers to training samples [J]. *Journal of Remote Sensing*, 2007, 11(6): 826–837.
- [7] Yan Jing, Wang Wen, Li Xiangge. Extracting the rice planting areas using an artificial neural network [J]. *Journal of Remote Sensing*, 2001, 5(3): 227–231. (in Chinese)
- [8] Arai K. A supervised thematic mapper classification with a purification of training samples [J]. *International Journal of Remote Sensing*, 1992, 13(11): 2039–2049.
- [9] Hui Wenhua. TM image classification based on support vector machine [J]. *Journal of Earth Sciences and Environment*, 2006, 28(2): 93–95.
- [10] Wu Jianping, Yang Xingwei. Purification of training samples in supervised classification of remote sensing data [J]. *Remote Sensing for Land & Resources*, 1996, 27(1): 36–41. (in Chinese)
- [11] Nelson R F, Latty R S, Mott G. Classifying northern forests using Thematic Mapper simulator data [J]. *Photogrammetric Engineering and Remote Sensing*, 1984, 50(5): 607–617.
- [12] Serra P, Pons X. Monitoring farmers' decisions on Mediterranean irrigated crops using satellite image time series [J]. *International Journal of Remote Sensing*, 2008, 29(8): 2293–2316.
- [13] Zhang Shaojia. Remote sensing classification by combing multiple classifiers [D]. Changsha: Southeast University, 2010. (in Chinese)
- [14] Liu Li, Yu Qiang. A study on a classification method of remote sensing combined stratified classification with supervised classification [J]. *Forest Inventory and Planning*, 2007, 32(4): 37–39. (in Chinese)
- [15] Zhang Hua. Study on reliable classification methods based on remotely sensed image [D]. Xuzhou: China University of Mining and Technology, 2012. (in Chinese)