考虑边界区域的地表覆盖分类精度评价方法

刘梦^{①②}, 曹鑫^{①*}, 李阳^①, 陈晋^{①②}, 陈学泓^①

① 北京师范大学地表过程与资源生态国家重点实验室, 北京 100875;

② 北京师范大学全球变化与地球系统科学研究院, 北京 100875

* 通讯作者, E-mail: caoxin@bnu.edu.cn

收稿日期: 2015-09-18; 接受日期: 2016-06-06; 网络版发表日期: 2016-09-26

国家高技术研究发展计划项目(编号: 2013AA122802)资助

摘要 地表覆盖作为地表状况的直接反映, 与全球物质能量循环、气候变化以及人类社会经济活动等密切相关。地表覆盖分类数据通常利用遥感图像进行分类得到, 其数据质量对全球环境监测及其变化研究与决策支持等过程有重要的影响。因此, 对分类结果进行精度评估是地表覆盖分类的重要步骤。由于影像上不同地物类型边界处往往存在较多的混合像元, 使得分类时处于边界处的像元分类精度相对低于内部像元, 直接导致了类内自身的分类精度异质性增大, 分类精度评估的不稳定性也随之增大。本文基于传统的分层抽样(stratified sampling, SS), 提出了一种考虑边界区域的分层抽样方法(stratified sampling considering edges, SSCE)。理论推导及实验结果表明, SSCE对总体精度(overall accuracy, OA)和 $kappa$ 估计的准确性和稳定性均优于 SS 方法, 主要结论为: 在样本点总数较少的情况下, SSCE对分类精度估计的准确性高于 SS 方法; 在相同的容许误差下, SSCE方法所需的样本数小于 SS 方法; 分类图像中类内边界区域与内部区域分类精度的差异越大、其面积比例越均衡, SSCE 方法越能准确地进行精度估计。因此, 本文提出的 SSCE 方法能够以更少的样本获取代价、更高的准确性和稳定性对地表覆盖数据进行分类精度的评估。

关键词 地表覆盖, 分类精度评估, 分层抽样, 边界区域, 内部区域

1 引言

地表覆盖是自然过程与人类社会经济活动过程相互作用的产物, 其变化直接影响到地表水分、能量和碳循环。全球地表覆盖的空间分布与动态变化制

图已成为全球变化研究、地理国情监测、可持续发展规划、土地资源管理等领域不可或缺的基础信息(Hansen等, 2013; Turner等, 1995; Zell等, 2012)。遥感数据分类是生产全球地表覆盖制图产品的唯一可行方法(Chen等, 2015), 但由于地物在不同区域、不同时

中文引用格式: 刘梦, 曹鑫, 李阳, 陈晋, 陈学泓. 2016. 考虑边界区域的地表覆盖分类精度评价方法. 中国科学: 地球科学, 46: 1472–1481, doi: 10.1360/N072015-00276

英文引用格式: Liu M, Cao X, Li Y, Chen J, Chen X H. 2016. Method for land cover classification accuracy assessment considering edges. Science China Earth Sciences, 59: 2318–2327, doi: 10.1007/s11430-016-5333-5

相的遥感图像上普遍存在“同物异谱”及“异物同谱”的现象,而训练样本、先验知识以及分类器本身的不完备性都可能导致不同程度的分类错误(Foody, 2002),从而直接影响到地表覆盖数据产品的实际应用.因此,如何对分类产品进行科学的精度评估是这些数据产品能否合理应用并理解其不确定性来源的基础.

目前分类图像精度评估主要基于混淆矩阵(error matrix)计算总体精度(overall accuracy, OA)和kappa等精度评价参数(Cohen, 1960; Liu等, 2007; Olofsson等, 2014).由于遥感图像覆盖范围较大,一般难以获得准确而完整的地面参考数据,分类精度评估需要通过抽样构建混淆矩阵,然后对分类图像总体进行评价.常用的抽样方法包括简单随机抽样、集群抽样、系统抽样和分层抽样等(Stehman, 2009),其中分层抽样方法应用最为广泛(廖安平等, 2014; Olofsson等, 2013; Stehman, 2012).分层抽样(stratified sampling, SS)对各层进行独立的随机抽样,保证每层数据均能采集到一定的样本量(Särndal等, 2003),从而能够考虑不同层的分类精度差异,降低精度评价的不确定性.然而,现有分层抽样法使用时,并未考虑同一层(同一地表覆盖类型)边界区域分类精度(混合像元较多)低于内部区域的事实(Hammond和Verbyla, 1996; Sweeney和Evans, 2012),即同一种地表覆盖类型内分类精度具有异质性,从而导致分类精度将出现10~20%的高估(Hammond和Verbyla, 1996; Stehman等, 2003).针对这一问题, Sweeney等(2012)对影像进行网格化处理,在各网格内随机选择子样区来对同一层内的边界区域进行考虑,从而降低了其精度评价的高估现象;然而该方法对子样区代表性的要求较高,限制了其实用性.李蓓蓓等(2010)也使用过网格化的方式来对中国东北部的耕地数据进行评估,但并未考虑到边界的问题.

本文在传统的分层抽样方法的基础上,提出了一种考虑边界区域的分层抽样方案.通过理论推导,证明新方法对传统分层抽样方法在精度评估准确性和稳定性等方面具有一定优势;通过在样本总数、容许误差、精度评估稳定性、分类图像性质等方面的实验,对考虑边界区域的分层抽样方法和传统分层抽样方法进行了比较,从而确定新方法相对于传统分层抽样法的优势和适用范围.

2 方法

2.1 分层抽样及其样本分配

2.1.1 分层抽样法(SS)

当总体的不同单元之间的差异很大时,不宜直接进行随机采样,可以将总体划分为若干个层(strata)再进行分层抽样(金勇进等, 2008).在分层时,需要按照层内差异小、层间差异大的原则进行.分层抽样法(stratified sampling method)主要通过以下两个步骤进行: (1) 按照一定的原则将图像划分为不同的层,并在各层中进行简单随机抽样.分层时可以按照不同的地物类型进行分层,将每个类型都划分为单独的层,也可以按照地理位置进行分层(Stehman等, 2003). (2) 根据各层简单随机采样的结果对总体进行估计,构建总体混淆矩阵,计算出OA和kappa系数来表征图像的分类精度.值得注意的是,在以类型为分层依据的分层抽样过程中,类型数目与层的数目相同,并且类型与层在位置上是完全重合的,每一种地物类型都是一个单独的层.所有的层之间互不重合,层的并集构成完整的分类图像.而对于以地理位置为分层依据的分层抽样而言,空间上的每一个区域都是单独的层,比如对我国的人口密度进行抽样,可以按照行政区划来分层.进行分层抽样,建立总体混淆矩阵(表1;以4类为例),计算出OA和kappa来反映影像的分类效果.

表1中N表示分类图像上像元总数, $\hat{N}_{ij} = N_i \frac{n_{ij}}{n_i}$

表示根据采样所得到的在分类图像上属于类型i而在参考图像上属于类型j的像元数的估计值(其中 n_i 表示第i层的样本点数目, n_{ij} 表示在分类图像上属于类型i而在参考图像上属于类型j的样本数目), N_i 表示分类

表 1 总体混淆矩阵

		参考图像				
		类别1	类别2	类别3	类别4	行和
分类图像	类别1	$\hat{N}_{11}$	$\hat{N}_{12}$	$\hat{N}_{13}$	$\hat{N}_{14}$	N_1
	类别2	$\hat{N}_{21}$	$\hat{N}_{22}$	$\hat{N}_{23}$	$\hat{N}_{24}$	N_2
	类别3	$\hat{N}_{31}$	$\hat{N}_{32}$	$\hat{N}_{33}$	$\hat{N}_{34}$	N_3
	类别4	$\hat{N}_{41}$	$\hat{N}_{42}$	$\hat{N}_{43}$	$\hat{N}_{44}$	N_4
	列和	$\hat{M}_1$	$\hat{M}_2$	$\hat{M}_3$	$\hat{M}_4$	N

图像上第*i*类的实际像元总数, $\hat{M}_j$ 表示参考图像上根据采样结果所估计出的第*j*类的像元总数. 根据总体混淆矩阵(表1)计算OA和kappa(Cohen, 1960):

$$OA = \frac{1}{N} \sum_{i=1}^L \hat{N}_{ii}, \quad (1)$$

$$kappa = \frac{N \sum_{i=1}^L \hat{N}_{ii} - \sum_{j=1}^L \hat{M}_j N_j}{N^2 - \sum_{j=1}^L \hat{M}_j N_j}, \quad (2)$$

式中, L 表示地表覆盖类型数.

2.1.2 分层抽样的样本分配

在利用分层抽样方法建立混淆矩阵时, 需要确定分类图像的各层所需的样本数目. 常用的样点分配方法包括按面积比例分配、均匀分配、按权重分配、以及将各层层内方差与面积比例相结合的内曼最优分配等(Särndal等, 2003; Stehman, 2012). 内曼最优分配方法在进行样点分配时, 以各层面积比例与其自身标准差的乘积作为权重, 能够保证OA的估计值具有最小的方差, 各层所需样点的计算公式(Stehman, 2012)为

$$n_h = n \frac{W_h S_h}{\sum_{h=1}^L W_h S_h}, \quad (3)$$

式中, n_h 表示各层分配的样本数, n 表示样本点总数, W_h 表示各层占影像总面积的比例, S_h 表示各层对应的标准差. 样本点总数 n 可以通过预先设置出参数的估计值与真实值之间的容许误差 d , 并结合各层标准差计算出来(金勇进等, 2008):

$$n = \left(\frac{z_{1-\frac{\alpha}{2}}}{d} \right)^2 \left(\sum_{h=1}^L W_h S_h \right)^2, \quad (4)$$

式中, d 表示抽样之前预先人为设定的容许误差(或称预期偏差), $z_{1-\frac{\alpha}{2}}$ 表示正态分布的分位数, 其中 α 的取值一般为0.05.

2.2 考虑边界的分层抽样法(SSCE)

2.2.1 SSCE 流程

分层抽样(SS)以不同层具有不同性质而且各层性质均一为原则(Khan等, 2015; Särndal, 2003), 然而,

在地表覆盖分类中同一层中的边界区域与内部区域在分类精度上存在差异, 使得同一类型的性质并不均一. 本文在传统分层抽样法的基础上, 提出一种考虑边界的分层抽样法(stratified sampling considering edges method, SSCE), 通过设置每类的边界区域和内部区域作为新增的层, 从而对边界区域进行单独抽样. SSCE的基本步骤为: (1) 划分边界像元与内部像元, 按照8邻域搜索的方式(Sweeney和Evans, 2012)进行, 如果当前像元8邻域内的像元都与当前像元具有相同的类型, 那么当前像元就属于内部像元; 否则就划分为边界像元. (2) 结合“类型”与“是否边界”两个依据进行分层, 将两个分层依据同等对待, 在分层过程中同时使用. 以3类为例, 首先根据类型可以将图像划分为3个相互独立的层; 然后在各个层内根据边界与非边界的判断, 将每层继续划分为边界像元层和内部像元层; 最终, 分类图像应被划分为6个平等的层, 包括类型1的边界与内部、类型2的边界与内部和类型3的边界与内部. (3) 使用内曼最优分配方法来计算出每个层内所需的样本点数目, 并在各个层中进行随机抽样. 首先对图像进行预采样, 计算出各层自身的标准差 S_h . 然后确定抽样过程中所需要的样本点总数(可以人为给定也可以根据式(4)计算得到), 再结合内曼最优分配的原理(式(3))计算出各层所需的样本点数目, 在各层中进行简单随机采样. 根据采样结果构建混淆矩阵, 计算出OA和kappa等参量.

使用内曼最优分配时, 需要已知各层的标准差. Olofsson等(2014)根据先验知识给出各层可能的方差, 以此来计算抽样的样点数目. 由于通常难以预先设定各层的方差, 因此需要对各层的方差进行估计. 本文使用对整幅分类图像进行预抽样的方式来估计各层的方差. 首先设置每层的样本点数目均为30个, 根据预抽样获得的分类精度以及样点数目计算出各层的层内方差, 进而获得各层的标准差. 根据标准差的大小, 结合式(3)和(4)计算出各层在正式抽样时所需的抽样点数目. 考虑边界的分层抽样法(SSCE)的操作流程如图1所示.

抽样完成之后, 可对总体精度参数进行估计. 以抽样点属于内部像元或是边界像元作为区分, 得到边界区域和内部区域各自的混淆矩阵. 对于内部区域的抽样点来说, 可用下式对每一类中内部区域的未知参数进行估计:

图1 考虑边界的分层抽样法(SSCE)流程图

$$\hat{N}_{ij} = N_i \frac{n_{ij}}{n_i}, \quad (5)$$

式中, N_i 表示分类图像上第*i*类内部区域的像元总数, n_{ij} 表示在内部区域分类为第*i*类而在参考图像上属于第*j*类的抽样点数目, n_i 表示内部区域在第*i*类上布设的抽样点数目. 同理, 对于边界区域的抽样点来说, 可用下式对每一类中边界区域未知参数进行估计:

$$\hat{N}E_{ij} = NE_i \frac{ne_{ij}}{ne_i}, \quad (6)$$

式中, NE_i 表示分类图像上第*i*类边界区域的像元总数, ne_{ij} 表示在边界区域分类为第*i*类而在参考图像上属于第*j*类的抽样点数目, ne_i 表示边界区域在第*i*类上布设的抽样点数目.

将每一类的内部区域混淆矩阵与边界区域混淆矩阵各自对应参数的估计值进行加和, 得到总体估计的混淆矩阵(表1). 根据式(1)和(2), 计算出OA和kappa的估计值.

2.2.2 SSCE 与 SS 的比较

在分类为*L*类的影像上, 第*h*类的面积比例为 W_h , 层内方差为 S_h^2 , 第*h*类内部区域占图像总面积的比例为 W_{hi} , 第*h*类内部区域的标准差为 S_{hi} , 第*h*类边界区域占图像总面积的比例为 W_{he} , 第*h*类边界区域的标准差为 S_{he} ; 第*h*类内部区域占该类的面积比例为 W'_{hi} , 第*h*类边界区域占该类的面积比例为 W'_{he} , 正态分布的分位数为 $z_{1-\frac{\alpha}{2}}$, 其中 α 的取值一般为0.05. 在相同的容许误差*d*下, SS方法所需样本总数的计算(金勇进等, 2008)与SSCE方法所需样本总数的计算如下:

$$n_{ss} = \left(\frac{z_{1-\frac{\alpha}{2}}}{d} \right)^2 \left(\sum_{h=1}^L W_h S_h \right)^2, \quad (7)$$

$$n_{ssce} = \left(\frac{z_{1-\frac{\alpha}{2}}}{d} \right)^2 \left[\sum_{h=1}^L W_h (W'_{hi} S_{hi} + W'_{he} S_{he}) \right]^2, \quad (8)$$

式中, n_{ss} 和 n_{ssce} 分别表示SS方法和SSCE方法各自的样本总数. 根据方差合成原理(Chen等, 2011), 总体的方差可以由各个部分的方差合成得到, SS方法的第*h*层方差可以表示如下:

$$S_h^2 = W'_{hi} S_{hi}^2 + W'_{he} S_{he}^2 + W'_{hi} (\bar{y}_{hi} - \bar{y}_h)^2 + W'_{he} (\bar{y}_{he} - \bar{y}_h)^2 > W'_{hi} S_{hi}^2 + W'_{he} S_{he}^2. \quad (9)$$

根据式(9)中的不等式关系知, 确定了 $(W'_{hi} S_{hi}^2 + W'_{he} S_{he}^2)$ 与 $(W'_{hi} S_{hi} + W'_{he} S_{he})^2$ 之间的相对大小, 就可以确定 $(W'_{hi} S_{hi} + W'_{he} S_{he})$ 与 S_h 之间的大小关系:

$$(W'_{hi} S_{hi}^2 + W'_{he} S_{he}^2) - (W'_{hi} S_{hi} + W'_{he} S_{he})^2 = W'_{hi} W'_{he} (S_{hi} - S_{he})^2 > 0. \quad (10)$$

由此可见, $n_{ssce} < n_{ss}$, 即在相同的容许误差下, 使用SSCE方法需要的样本点数目更少.

在使用相同数目样本的情况下, SS方法在OA估计值上的方差(金勇进等, 2008)与SSCE方法在OA估计值上的方差之间的差值可表示为

$$\text{Var}(\text{OA}_{ss}) - \text{Var}(\text{OA}_{ssce}) = \frac{1}{n} \left(\sum_{h=1}^L W_h S_h \right)^2 - \frac{1}{n} \left[\sum_{h=1}^L (W'_{hi} S_{hi} + W'_{he} S_{he}) \right]^2, \quad (11)$$

式中, $\text{Var}(\text{OA}_{ss})$ 和 $\text{Var}(\text{OA}_{ssce})$ 分别表示SS方法和SSCE方法各自对OA进行估计时的方差, 方差越小表明估计相对越准确. 由式(9)和(10)可得 $(W'_{hi} S_{hi} + W'_{he} S_{he}) < S_h$, 即有 $(W'_{hi} S_{hi} + W'_{he} S_{he}) < W_h S_h$. 因此, 式(11)展开之后可以简化为

$$\text{Var}(\text{OA}_{ss}) - \text{Var}(\text{OA}_{ssce}) > \frac{1}{n} \sum_{h=1}^L W_h^2 \left[S_h^2 - (W'_{hi} S_{hi} + W'_{he} S_{he})^2 \right]. \quad (12)$$

由式(10)可得 $(W'_{hi} S_{hi}^2 + W'_{he} S_{he}^2) > (W'_{hi} S_{hi} + W'_{he} S_{he})^2$, 可以对上式进行变换, 得到不等式:

$$\text{Var}(\text{OA}_{ss}) - \text{Var}(\text{OA}_{ssce}) > \frac{1}{n} \sum_{h=1}^L W_h^2 \left[S_h^2 - (W'_{hi} S_{hi}^2 + W'_{he} S_{he}^2) \right]. \quad (13)$$

根据式(9)中的方差合成原理, 可以对式(13)进行处理, 得到:

$$\begin{aligned} & \text{Var}(\text{OA}_{\text{SS}}) - \text{Var}(\text{OA}_{\text{SSCE}}) \\ & > \frac{1}{n} \sum_{h=1}^L W_h^2 \left[W_{hi}' (\bar{y}_{hi} - \bar{y}_h)^2 + W_{he}' (\bar{y}_{he} - \bar{y}_h)^2 \right] \\ & = \frac{1}{n} \sum_{h=1}^L W_{hi} W_{he} (\bar{y}_{hi} - \bar{y}_{he})^2, \end{aligned} \quad (14)$$

式中, $\bar{y}_{hi}$ 表示第 h 类内部区域的分类精度; $\bar{y}_{he}$ 表示第 h 类边界区域的分类精度; $\bar{y}_h$ 表示第 h 类整体的分类精度. 式(14)直接反映了SSCE方法在总体精度的估计上具有更小的方差, 因此对分类精度的估计更为准确.

通过以上的推导过程证明: SSCE方法在样本数量需求和精度估计指标的稳健性上, 相比传统SS方法具有一定优势, 但其步骤相对较多. 因此, 需要确定其具体适用条件. 当边界区域面积较大, 且边界区域与内部区域分类精度差异越大时, 传统SS方法的假设条件越难满足, SSCE方法越具有优势. 据此, 本文定义了一个参量SWDA (sum of area weighted square difference of accuracy between edge pixels and interior pixels), 作为选择SSCE方法的判断标准, 即以每类边界区域与内部区域面积比例的乘积作为权重, 对内部区域与边界区域分类精度差值平方进行加权求和. SWDA计算如下:

$$\text{SWDA} = \sum_{h=1}^L W_{hi} W_{he} (\bar{y}_{hi} - \bar{y}_{he})^2, \quad (15)$$

式中, W_{hi} 为第 h 类内部区域占图像总面积的比例, W_{he} 为第 h 类边界区域占图像总面积的比例, $\bar{y}_{hi}$ 表示第 h 类内部区域的分类精度; $\bar{y}_{he}$ 表示第 h 类边界区域的分类精度. SWDA指标可用于表征分类图像的特征, 即分类图像上边界区域与内部区域间面积比例的均衡性以及分类精度差异的大小. 但与一般的景观指数(Li等, 2005)不同, SWDA不仅仅只是刻画图像的复杂度, 而是将分类精度信息也融合到指数中. SWDA值越大, 表明分类图像上类内边界区域与内部区域的分类精度差值越大而且面积比例越均衡.

3 实验数据

在对抽样方案的效果进行测试时, 需要有对应于分类图像的完整参考图像(地表真实数据), 从而将抽样所得的估计值与真实的OA和kappa值进行比较.

为评价方法的优劣, 将OA和kappa估计值与真实值(通过全图计算)进行比较, 用模拟值与真实值进行比较是常用的精度验证方式(朱鹏和宫鹏, 2014), 计算出均方根误差(RMSE)如下:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\hat{\alpha}_i - \alpha_i)^2}{N}}, \quad (16)$$

其中, $\hat{\alpha}_i$ 表示参数(OA或者kappa)的估计值, α_i 表示参数真实值, N 表示测试次数.

全球30m地表覆盖分类产品(GlobeLand30)数据服务网站(<http://www.globallandcover.com>)提供了2000和2010年全球地表覆盖数据, 在2010年的分类数据集中选取了京津冀、陕西、印度和巴西等地区共20幅分类结果作为地表真实参考图像(图2), 并获取了同时期对应区域的Landsat TM影像. 由于真实的地表覆盖数据难以得到, 且使用高分影像进行解译的工作量巨大. 因此这里在对方法进行实证时, 以GlobeLand30的分类结果作为分类的真实值(即参考图像). GlobeLand30在全局尺度上的分类精度达到了80%以上, 其分类结果相对来说是可靠的. 考虑到影像分类时, 分类错误的分布是具有空间相关性的(Congalton, 1988), 因此, 在参考图像上随机生成椒盐噪声(Chen等, 2010)来获得分类图像的方式并不符合实际. 本文中采用最大似然分类方法(maximum likelihood classification, MLC)来生成分类图像, 作为评价对象.

对于每幅地表真实参考图像, 分别从各个类别中随机选择500个内部像元作为原始的训练样本, 使用MLC进行分类图像的生成. 在生成分类图像时, 将原始的训练样本设置成5种不同的组合, 即训练样本错误率为0%、5%、10%、15%和20%. 训练样本的不同错误率通过将一定比例训练样本设置为其他类型实现. 在生成固定比例错误的训练样本时, 以10%的错误率及森林类型为例, 其训练样本有90%来自于森林一类的内部像元(认为内部像元是分类正确的), 10%的样本来自于其他类型的像元. 训练样本完全随机生成, 不进行人为干预. 在这5种梯度下, 结合Landsat TM影像即可生成5幅带有分类误差的分类图像. 总共得到100组参考图像与分类图像的组合, 这100组图像都被用于对SSCE和SS方法进行比较和

分析.

以图2g中河北省廊坊市的一幅影像为例来说明分类图像的生成结果. 使用不同错误梯度的训练样本结合最大似然分类生成5幅分类图像(图3), 分类图像的相关精度参数如表2所示. 廊坊市的参考图像中存在极少量的水体像元, 为简便起见, 将这部分水体

赋值为人造地表, 以讨论两种地表覆盖类型下分类误差随错误样本变化的情况. 表2表明, 随着训练样本中错误率的增大, 分类图像的精度(OA和 $kappa$)出现明显的降低. 通过这种方式生成的分类图像既保证了图像上分类误差出现的真实性, 也同时实现了图像分类质量的控制.

图 2 从 GlobeLand30(2010 年)选取的 20 幅参考分类影像

图3 Landsat TM 影像、参考分类图像与不同错误率的分类结果

表 2 根据训练样本错误控制实现的分类图像精度差异

精度	0%错误 样本	5%错误 样本	10%错误 样本	15%错误 样本	20%错误 样本
OA	0.8558	0.8515	0.8408	0.8308	0.7939
kappa	0.6707	0.6596	0.6288	0.5934	0.4869

4 实验结果与分析

4.1 样点总数的比较

根据2.2.2节的讨论,在样本总数相同的情况下,SSCE方法比SS方法估计出的分类精度会更加准确.

通过以下实验进行验证:控制抽样样本总数在100~1000之间变化,以100作为步长,分别通过SSCE方法和SS方法来估计OA和kappa.图4为两种方法对OA(图4a)和kappa(图4b)的估计值与参考图像真实值之间的RMSE随样本总数变化的折线图.随着样本点总数的增加,两种方法对于OA和kappa估计值的RMSE都在逐渐减小,样本数目越大,两种方法之间RMSE的差距越小,当样本点总数达到900以上时,两种方法OA的RMSE均接近0.01,kappa的RMSE均接近0.02.表明通过增加抽样样本数量可以提高精度估计的准确性,当样本数目增大到一定程度时,不论哪

图4 OA 与 kappa 的 RMSE 随样点总数的变化

种方法都能比较好地对分类精度进行估计. 更为重要的是, 可以看到SSCE方法的RMSE均小于SS方法, 而且, 当样本数目较小时, SSCE方法优势越明显, 证明新方法在小样本下的精度估计更为稳健.

4.2 容许误差的比较

式(4)表明容许误差 d 直接控制着抽样所需的样本总数 n , d 越小, n 越大. 通过设置容许误差 d 在0.1~0.01之间均匀变化, 步长为0.01, 计算出在各种容许误差下SSCE与SS方法各自所需的样本点总数(图5). 随着容许误差的减小, 两种方法所需样本点数目均呈现出近似指数型的增长(图5a), 在容许误差 d 为0.1时, 两种方法之间抽样点数目差值为3, 当容许误差 d 为0.01时, 二者样本数的差值可达到350(图5b), 表明对于相同的容许误差而言, 使用SSCE方法所需的样本点数目要小于SS方法. 而且, 当所希望的容许误差较小时, SS方法需要的样本点数目会明显高于SSCE方法, 其抽样的花费代价也会更高, 而这时SSCE方法优势就更为显著.

4.3 精度评估稳定性的比较

根据上述讨论, 如果容许误差 d 设置得过小, 会导致所需的样点数目剧烈增长. 为了保证方法的估计效果, 同时使得样点总数在可接受的范围内, 我们设置容许误差为0.04(Olofsson等, 2014), 对生成的100组数据每组重复测试50次, 总共测试5000次. SSCE与SS方法分别估计出的OA和 $kappa$ 与真实值的散点图如图6所示. 从图6可以看出, 两种方法对于OA和 $kappa$ 的估计值都在真实值附近波动, 散点的

分布都接近于 $y=x$ 直线.

通过计算OA和 $kappa$ 估计值相对于真实值的RMSE, 对SSCE与SS方法进行定量比较, 结果如表3所示. SSCE方法估计的OA及 $kappa$ 与真实值的RMSE都低于SS方法. 对于OA的估计, SSCE方法的RMSE比SS方法减少了13.87%; 而对于 $kappa$ 的估计而言, SSCE方法的RMSE相对减少了15.70%, 说明SSCE方法可以提高精度估计的稳定性.

4.4 分类图像性质对精度估计准确性的影响分析

对上面5000次计算所得的SWDA进行统计, 其取值范围在0.000332~0.031859之间, 从0开始以0.0025作为间隔, 得到SWDA取值的13个区间. 在每个区间内, 计算出SSCE与SS方法各自的OA和 $kappa$ 估计值相对于真实值的RMSE. 以SWDA取值区间的组中值为横坐标, 以对应区间内SS的RMSE与SSCE的RMSE的差值为纵坐标绘制散点图(图7). 图7中纵坐标均大于0, 即SS对于OA和 $kappa$ 估计值的RMSE均大于SSCE, 表明SSCE在精度估计上的准确性更高. 随着SWDA取值的增大, SS与SSCE在OA(图7a)和 $kappa$ (图7b)上的RMSE的差值均表现出了明显的增大趋势, 表明SWDA取值越大时, SSCE相对于SS在分类精度评估上就越具有优势. 而且SS与SSCE在 $kappa$ 上的RMSE差值增大的速率比在OA上更大, 表明 $kappa$ 对于SWDA的变化更为敏感. 当SWDA取值较小时, 分类图像上内部像元区域与边界像元区域分类精度差异较小, 面积比例也不均衡, SS与SSCE在OA及 $kappa$ 上的RMSE差值均较小, 使用这两种抽样方法在分类精度评估上的效果相当; 而当SWDA

图5 抽样点总数及其差值随预期偏差 d 的变化

图6 真实值与估计值的比较

表3 抽样方案的比较

	SSCE	SS	SS-SSCE	(SS-SSCE)/SS
$RMSE_{OA}$	0.0196	0.0227	0.0031	13.87%
$RMSE_{kappa}$	0.0360	0.0427	0.0067	15.70%

取值越大时,考虑边界的必要性增大,使用SSCE更能够相对提高参数估计的准确性,从而获得更为准确的分类精度评估结果。

5 讨论与结论

在传统分层抽样(SS)的基础上,提出了考虑边界的分层抽样方法(SSCE),将边界区域设置成为新的层,并使用内曼最优分配来确定各层所需的样本点数目。通过理论推导和数据实验,表明SSCE对精度(OA和kappa)估计的准确性和稳定性均优于SS方法,并得出以下结论:(1)在样本点总数较少的情况下,SSCE对精度估计的准确性明显高于SS方法;当样本

图7 SS与SSCE的OA及kappa的RMSE差值随SWDA组中值的变化

点数目增大到一定程度时, SSCE与SS方法对精度评估的效果相当. 因此, SSCE方法可花费较少成本获取一定数量的检验样本, 即能保证精度评估的准确性. (2) 在相同的容许误差下, SSCE方法所需的样本数小于SS方法. 因此, 在相同的精度估计准确性要求下, 使用SSCE方法需要样本点数目较小而且更为稳健. 随着容许误差的减小, 所需要的样本点总数呈现出指数型的增长. 在实际使用中, 只需要容许误差满足需求即可, 从而避免出现样本量“爆炸”的问题. (3) 分类图像的性质, 即类内边界区域与内部区域分类精度的差异、以及其面积比例的差异, 决定了SSCE方法相对于SS方法的优势. 分类图像中类内边界区域与内部区域分类精度的差异越大、其面积比例越均衡, 则SSCE方法越能准确地进行精度估计. 在实际应用中, 可以通过预抽样的方式来获得分类图像上每一类内部区域与边界区域分类精度的差异, 计算初步的SWDA值, 从而决定是否有必要使用SSCE方法. 如果SWDA的值相对较大(例如大于0.01), 那么使用SSCE方法就能够更有效地进行精度评估.

需要指出的是, 本文中提出的考虑边界的分层抽样方法在分层的过程中, 层的数目会随着分类的类型数目增加以倍数增长. 当分层的数目过多时, 每一类边界像元相对就会减少, 这样类内边界区域与内部区域的面积比例易于失衡, 从而可能导致SSCE方法丧失相对于SS方法的优势. 如果能够根据实际需要确定哪些类型需要对边界区域与内部区域进行区分, 以此适当减少分层数目, 将会进一步提高新方法的适用性和稳定性.

参考文献

- 金勇进, 杜子芳, 蒋妍. 2008. 抽样技术. 北京: 中国人民大学出版社. 66
- 李蓓蓓, 方修琦, 叶瑜, 张学珍. 2010. 全球土地利用数据集精度的区域评估——以中国东北地区为例. 中国科学: 地球科学, 40: 1048–1059
- 廖安平, 陈利军, 陈军, 何超英, 曹鑫, 陈晋, 彭舒, 孙芳蒂, 宫鹏. 2014. 全球陆表水体高分辨率遥感制图. 中国科学: 地球科学, 44: 1634–1645
- 朱鹏, 宫鹏. 2014. 全球陆表湿地潜在分布区制图及遥感验证. 中国科学: 地球科学, 44: 1610–1620
- Chen J, Chen J, Liao A, Cao X, Chen L, Chen X, He C, Han G, Peng S, Lu M, Zhang W, Tong X, Mills J. 2015. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS-J Photogramm Remote Sens*, 103: 7–27
- Chen J, Zhu X, Imura H, Chen X. 2010. Consistency of accuracy assessment indices for soft classification: Simulation analysis. *ISPRS-J Photogramm Remote Sens*, 65: 156–164
- Chen X H, Chen J, Jia X P, Somers B, Wu J, Coppin P. 2011. A quantitative analysis of virtual endmembers' increased impact on the collinearity effect in spectral unmixing. *IEEE Trans Geosci Remote Sensing*, 49: 2945–2956
- Cohen J. 1960. A coefficient of agreement for nominal scale. *Educ Psychol Meas*, 20: 37–46
- Congalton R G. 1988. Using spatial autocorrelation analysis to explore the errors in maps generated from remotely sensed data. *Photogramm Eng Remote Sens*, 54: 587–592
- Foody G M. 2002. Status of land cover classification accuracy assessment. *Remote Sens Environ*, 80: 185–201
- Hammond T O, Verbyla D L. 1996. Optimistic bias in classification accuracy assessment. *Int J Remote Sens*, 17: 1261–1266
- Hansen M C, Potapov P V, Moore R, Hancher M, Turubanova S A, Tyukavina A, Thau D, Stehman S V, Goetz S J, Loveland T R, Kommareddy A, Egorov A, Chini L, Justice C O, Townshend J R G. 2013. High-resolution global maps of 21st-century forest cover change. *Science*, 342: 850–853
- Khan M, Rao D, Ansari A H, Ahsan M J. 2015. Determining optimum strata boundaries and sample sizes for Skewed population with log-normal distribution. *Commun Stat-Simul C*, 44: 1364–1387
- Li X, He H S, Bu R, Wen Q, Chang Y, Hu Y, Li Y. 2005. The adequacy of different landscape metrics for various landscape patterns. *Pattern Recognit*, 38: 2626–2638
- Liu C, Frazier P, Kumar L. 2007. Comparative assessment of the measures of thematic classification accuracy. *Remote Sens Environ*, 107: 606–616
- Olofsson P, Foody G M, Herold M, Stehman S V, Woodcock C E, Wulder M A. 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sens Environ*, 148: 42–57
- Olofsson P, Foody G M, Stehman S V, Woodcock C E. 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens Environ*, 129: 122–131
- Särndal C, Swensson B, Wretman J. 2003. *Model Assisted Survey Sampling*. New York: Springer Science & Business Media. 418
- Stehman S V. 2009. Sampling designs for accuracy assessment of land cover. *Int J Remote Sens*, 30: 5243–5272
- Stehman S V. 2012. Impact of sample size allocation when using stratified random sampling to estimate accuracy and area of land-cover change. *Remote Sens Lett*, 3: 111–120
- Stehman S V, Wickham J D, Smith J H, Yang L. 2003. Thematic accuracy of the 1992 National Land-Cover Data for the Eastern United States: Statistical methodology and regional results. *Remote Sens Environ*, 86: 500–516
- Sweeney S P, Evans T P. 2012. An edge-oriented approach to thematic map error assessment. *Geocarto Int*, 27: 31–56
- Turner B L, Skole D L, Sanderson S, Fischer G, Fresco L, Leemans R. 1995. Land-use and land-cover change. In: IGBP, ed. *IGBP Global Change Report No.35*. Stockholm: International Geosphere- Biosphere Programme
- Zell E, Huff A K, Carpenter A T, Friedl L A. 2012. A user-driven approach to determining critical earth observation priorities for societal benefit. *IEEE J-Stars*, 5: 1594–1602