数字土壤质地制图方法比较
——以黑河张掖地区为例

刘 超[1,2],卢 玲[1],胡晓利[1]
(1. 中国科学院寒区旱区环境与工程研究所,甘肃 兰州 730000;2. 中国科学院研究生院,北京 100049)

摘要:土壤质地类型是陆面过程模型、水文模型和大气模型的重要输入参数。在黑河流域中游地区选取一研究区,利用 200 个实测土壤剖面样点资料及 13 类环境因子组合,分别根据基于支持向量机、决策树和模糊逻辑的理论方法进行土壤质地预测制图。结果表明:支持向量机方法对训练点和验证点的测试精度都偏低,分别为 90%和 55%;基于决策树方法对训练点的验证精度最高,达到 98%,但其对验证点的预测准确率为 57%;基于模糊逻辑方法的综合评价精度最好,对验证点的预测精度达到 64%,对训练样点的验证精度为 74%。从制图结构来说,支持向量机方法弱化了类型环境因子的预测性能,减小了连续环境因子分布的空间差异性,生成的土壤质地图结构相对简单,精度偏低;基于决策树方法对未知环境因子组合预测具有不确定性,生成土壤图斑块相对破碎;模糊逻辑方法可弥补决策树方法的一些不足,生成的土壤质地图既较好地保留了斑块结构的完整性,也较好描述了土壤类型与典型环境因子之间的相关性。因此针对研究区数据特点,采用基于决策树和模糊逻辑相结合的土壤制图方法更有优势。

关 键 词:土壤质地制图;决策树;支持向量机;模糊逻辑;黑河流域
中图分类号:TP 79 **文献标志码**:A **文章编号**:1004-0323(2011)02-0177-09

1 引 言

土壤质地是土壤最基本的物理属性,质地不同的土壤其粗细颗粒组成及含量不同,从而影响土壤持水性及导水性等水力特性以及热传导系数和热容等热力参数;间接影响土壤蒸发、土壤入渗等水文过程。因此,土壤质地类型是陆面过程模型、水文模型和耦合陆面过程的大气模型的重要输入参数,对水分循环和地表通量的模拟具有重要影响[1-5]。目前在数字黑河的参数集制备方面,最急需的是高质量的土地覆盖(或植被图、植被功能型图)和土壤质地图及与它们的类型相关的参数[6]。绝大部分陆面过程模型和分布式水文模型的土壤特性输入都是直接沿用美国农业部(United States Department of Agriculture,USDA)制定的土壤质地分类标准,而目前我国大多数土壤图参照我国的发生分类学或诊断分类学标准,无法直接满足陆面过程模型和分布式水文模型的需求。如何实现我国土壤分类与国际土壤分类标准之间的转换,是一个复杂的问题,史学正等[7]基本实现了我国 1∶100 万数字土壤图发生分类土类与美国系统分类土纲的参比。

在此基础上,近年来国内学者通过建立中国土壤分类学与美国制土壤质地分类标准间的转换关系,制作了几种可供陆面模式和数值模式直接利用的土壤质地分布图。例如,张时煌[8]融合 3 种不同比例尺土壤图制作出可供风沙模式及其他大气模式使用的分辨率为 0.05°的土壤质地资料。高艳红[9]根据土壤亚类剖面数据,寻找土壤发生分类与 USDA 质地分类标准建立对应关系,建立了黑河流域 30 s 分辨率的土壤质地类型分布图。戴永久等[10]根据全国二次土壤调查中质地剖面数据,以中国 100 万土壤图为基础,采用多边形链接方式,建立了中国 30 s 分辨率土壤砂土、粘土、粉土分布图。此外,1981 年世界粮农组织(FAO)/联合国教科文组织(UNESCO)发布了全球 1∶500 万土壤图,1992

收稿日期:2010-10-26;**修订日期**:2011-03-24
基金项目:国家自然科学面上基金项目(40871190),国家自然科学青年基金项目(40901160)资助。
作者简介:刘 超(1985—),女,湖南醴陵人,硕士研究生,主要从事数字土壤制图研究。E-mail:lc287@126.com。

年发布了数字化版本，并根据 FAO/UNESCO 土壤类型划分标准进行了土类的更新[11]。该土壤图在气候变化、陆面过程模拟方面得到了广泛利用，但其分辨率相对较粗，质量欠佳。2009 年 FAO/UNESCO 发布了世界和谐土壤数据库(HWSD)，分辨率提高到 100 万，土壤图精度和土壤属性的丰富程度也有不同程度的提高[12]。

土壤发生分类学与土壤质地分类是完全不同的两种土壤分类体系，直接通过寻找二者间的对应转换关系必然带来不少误差，因此探索如何根据土壤环境数据和土壤质地属性数据，直接生成土壤质地图的数字土壤制图方法很有必要。数字土壤制图始于 20 世纪 70 年代，在 80 年代随着遥感技术、计算机技术、统计和建模的发展而迅速发展[13]。以土壤—景观为代表的模型外推方法，是目前最为主流的、适用于区域范围的土壤定量预测制图方法[14]。该方法基于经典的土壤发生学理论，即著名的 Jenny 方程，将土壤描述为气候、生物、地形、母质和时间的函数。通过 GIS、RS 等地理信息分析技术可以定量获取土壤存在的环境要素信息，根据土壤剖面信息定量提取土壤与环境之间关系的知识，从而进行土壤预测制图。基于土壤—景观模型的制图方法主要有基于统计和基于土壤专家知识两种。其中包括线性回归模型[15]、地统计模型[16]、判别分析[17]、决策树方法[18]、人工神经网络模型[19]、模糊聚类[20]和基于模糊逻辑[21]的制图方法都有广泛应用。线性回归模型和判别分析无法解释土壤与环境发生的复杂相互作用；地统计模型通常需要大量的实测数据为基础，限制了其使用。

本文选取黑河流域中游甘州区和临泽县及青海祁连县部分山区为研究区，根据该区域的环境因子以及野外实测的土壤剖面数据，分别利用决策树模型、支持向量机模糊和模糊逻辑 3 种制图方法，进行土壤质地预测制图，并采用野外验证点对 3 种方法所得土壤质地图精度进行比较，旨在探索出一种适合整个黑河流域的土壤质地制图方法。

2 研究方法

2.1 支持向量机模型

支持向量机是一种基于统计学习理论的机器学习方法，也是解决线性不可分问题的一种重要方法，在解决小样本、高维数、非线性以及局部极小点等问题方面有着较高的性能。支持向量机方法首先把训练数据集非线性地映射到一个高维特征空间(通常是 Hilbert 空间)，随后在高维特征空间中建立一个具有最大隔离距离的最优分离超平面，最后将该映射方法和决策边界应用于待分类数据，即得到分类结果[22]。最优的超平面是通过训练向量的一个子集的加权组合定义的，使用核函数来解决，通过选用合适的核函数，实现从线性不可分空间向高维线性可分特征空间的映射。本研究采用的核函数为高斯径向基核函数，表达方式如下：

$$K(x,y) = \exp(-\|x-y\|^2/\sigma^2) \quad (1)$$

其中：x 代表任一点数值，y 代表经过线性变换后的 x 值，σ 为函数的宽度参数，控制了函数的径向作用范围。同时本研究采用台湾大学林智仁教授[23]开发的一个易于使用的 SVM 识别与回归的软件包 LibSVM 进行制图，它包括多个计算机编程语言的版本，本研究采用的是 Matlab 语言版本。在进行多类分类时，LibSVM 采用一对一的方法，在多分类中每次只考虑两类样本，即对每两类样本设计一个 SVM 模型，因此，总共需要设计 $k(k-l)/2$ 个 SVM 模型，该软件包内部使用径向基核函数(RBF)作为特征空间映射的方法，通过交叉验证来确定径向基核函数的最优参数，避免过度拟合，并寻找最优分离超平面。

2.2 决策树模型

决策树(Decision Tree)[24]是一个类似于流程图的树型结构，其中每个内部节点表示在一个属性上的测试，其分支代表测试的每个结果，而树的每个叶节点就代表一个类别，树的最高层节点是根节点，是整个决策树的开始。决策树算法用于分类和预测制图具有很多优势[25]，能够有效地抑制训练样本噪声和解决属性缺失问题，有效利用连续数据和类型数据，决策树方法生成的决策树或产生式规则集结构简单直观、容易理解，还可以输入到专家系统中加以利用。决策树的生成过程包括两个阶段：树构造(Tree Building)和树剪枝(Tree Pruning)。树剪枝是为了使得到的决策树所蕴含的规则具有普遍意义，防止训练过度，必须适当控制树的生长，进行树的修剪。决策树算法有多种，包括 CART、FACT、CHAID、SPRINT 和 PUBLIC 算法等。当前最有影响的决策树算法是 Quinlan 于 1986 年提出的 ID3 算法[26]和 1993 年提出的 C4.5 算法[27]。

本文采用由美国地质调查局(USGS)开发的基于决策树的土地覆盖制图工具 NLCD Mapping Tools 进行制图，该工具箱是基于 ERDAS Imagine 所开发的。决策树算法由 See 5.0 软件生成，采用的决策树算法

即是在 C4.5 算法基础上改进的 C5.0 算法。

2.3　模糊逻辑

模糊逻辑的推理方法可将类别多边形的非 0 即 1 方式表示为 0～1 的隶属度形式,更能反映土壤的空间渐变分布特征。基于模糊逻辑的制图方法通常将土壤与环境因子的相互关系表达为系列的隶属度函数[28],通常采用高斯型隶属度函数,表达形式如:

$$S_{ij,v} = e^{[(z_0 - z_{ij,v})/\omega]^r \ln(c)} \quad (2)$$

其中:$S_{ij,v}$ 为像元(i,j) 在环境要素 v 下的土壤类型的相似度,$Z_{ij,v}$ 为环境要素 v 在(i,j) 位置的取值,z_0 为该类型土壤的环境要素 v 的典型值,c、r、ω 是参数,用来控制曲线形状;c 通常取为 0.5,r 表示隶属度为 0.5;r 控制曲线的坡度,值越大曲线越陡;ω 控制曲线宽度,表示隶属度为 1 时的环境条件值与隶属度为 0.5 时的环境条件值之差。每个像元的土壤与各土壤类型都有一个相似度,这样所有的隶属度就形成一个 n 维向量,即土壤相似度向量,n 表示土壤类型数,最终取相似度最大的土壤类型为该像元的土壤类型。基于模糊逻辑土壤制图的关键是刻画土壤与环境的隶属度函数,需要两类知识:第一类知识是土壤分布的典型环境条件,即隶属度为 1 时的环境条件;第二类知识是指当环境条件与土壤分布的典型环境条件发生偏离时隶属度的变化情况,即隶属度的类型与宽度[29]。

本文基于模糊逻辑的土壤质地预测制图采用 SoLIM 模型[21,29],是由中国科学院地理科学与资源研究所、美国威斯康星大学地理系和美国达特茅斯大学地理系联合开发的,是基于地理信息系统(GIS)、人工智能(AI)和模糊逻辑的精细土壤普查技术,它将土壤环境关系知识与 GIS 平台下的环境数据结合,可有效地进行土壤空间分布的预测推理(数据来源于国家自然科学基金委员会"中国西部环境与生态科学数据中心"(http://westdc.westgis.ac.cn))。

3　数据获取与处理

3.1　研究区概况

研究区位于黑河流域中游地区(图 1),包括甘州区、临泽县大部分平原地区和青海祁连县部分山区,位于 38.60°～39.42°N,99.84°～100.48°E 之间,面积约5 084 km²。该区域年平均降雨量约为138～800 mm,地形变化起伏大,高程变化范围为 1 343～3 981 m。该区域跨越不同的地形地貌特征区域,临泽的地貌主要为倾斜的洪积扇平原、冲击洪积和剥蚀平原,祁连山区则包括部分现代冰川地貌。根据

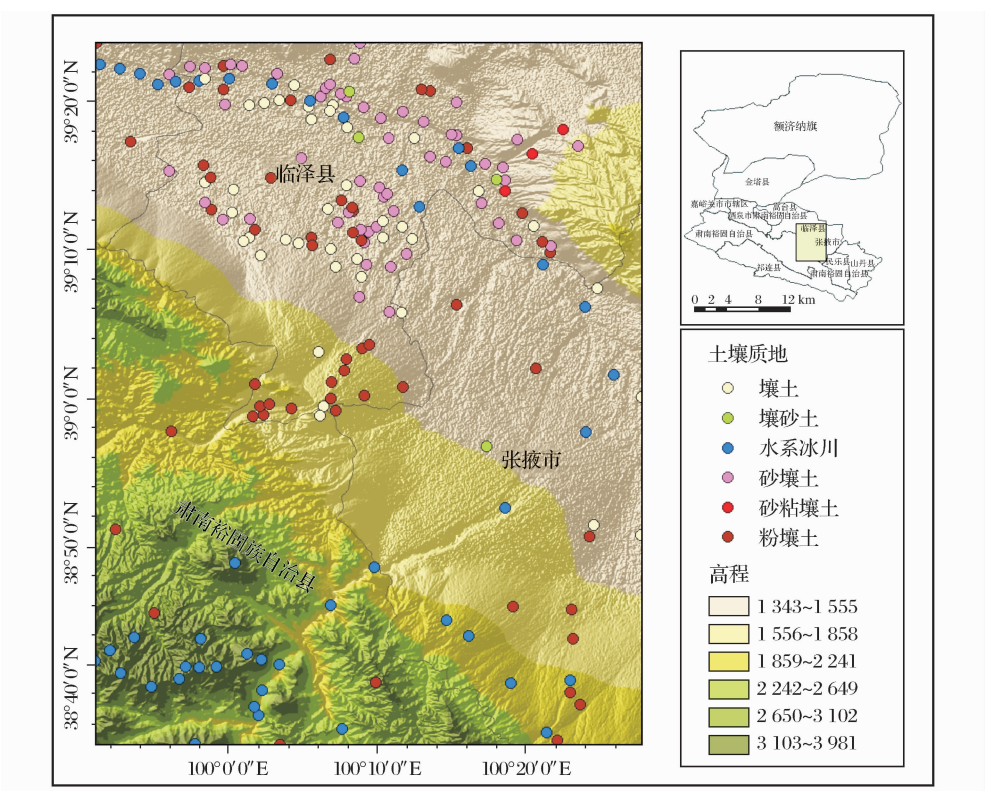

图 1　研究区和土壤剖面分布图

Fig. 1　The study area and the distribution of soil samplings

中国土壤发生分类学，该区域的主要土壤类型为灌漠土、栗钙土、棕钙土、灰钙土和风沙土等。选择此研究区是因为能够获取的野外采样的土壤剖面数据较多，涉及的土壤类型较为全面，分布也较为集中，从而有利于评价土壤—景观模型在不同地形地貌条件下的应用效果，并最终探索出一种适合整个黑河流域的土壤质地制图方法。

3.2 数据

在本研究区一共选取了 200 个土壤剖面数据（图 1），选取其中 129 个土壤剖面作为训练样本点，71 个土壤剖面作为验证点。尽可能在典型环境因子组合条件下选择训练样本点，且训练点分布能反映土壤分布环境条件的差异性和多样化，剩余土壤剖面作为验证点。通过野外采样和室内分析获得土壤的机械组成，即土壤中各种粒级所占的重量百分比，土壤颗粒组成按照美国土壤质地分级标准，包括砂粒（0.05～2 mm），粘粒（＜0.002 mm）和粉粒（0.002～0.05 mm）。根据美制土壤质地分类三角坐标图计算得到土壤质地类型。本区域土壤质地类型一共有 5 种，分别是砂粘壤土、砂壤土、粉壤土、壤土和壤砂土，不同土壤类型样本构成比例如表 1：

表 1 不同土壤类型样本构成比例

Table 1 The proportion of the soil samples for different soil types

土壤类型	砂粘壤土	砂壤土	粉壤土	壤土	壤砂土	冰川和水系
训练点	3	29	37	24	4	32
验证点	0	25	17	15	0	14

本试验区一共选择环境因子数 13 个，包括反映地形的高程数据，以及由高程衍生的坡度、沿剖面曲率和沿平面曲率 3 个地形因子。由于研究区范围较大，覆盖包括平原地区和地势变化较大的山区，仅选择地形因子不能充分刻画土壤与环境的关系，因此本研究中还考虑了地貌、地质、土地利用与土地覆盖和植被 4 个类型因子。考虑到平原区和山区的温度，降雨差异也较大，因而也将降雨、温度和辐射数据作为环境因子来描述土壤质地的空间变化特征。此外，还加入了植被指数和离河道距离两个环境因子。高程数据采用的 SRTM（Shuttle Radar Topography Mission）90m 分辨率高程产品，由美国太空总署（NASA）和国防部国家测绘局（NIMA）联合测量。在 ArcGIS 软件下根据该 DEM 生成坡度和曲率。地貌数据[30]来源于中国西部 1∶100 万数字地貌数据集，再经过重新编码而来；植被类型图[31]采用黑河流域 1∶100 万植被数据集；母质[32]根据全国 1∶50 万地质数据图库经过重新编码而产生；土地利用与土地覆盖[33]采用中国 2000 年 1∶10 万土地利用数据集，进行简单数据预处理，将二级分类合并为一级分类，即林地、草地、耕地等，并将细小图斑合并到邻近大图斑上，以免斑块过于破碎；植被指数采用中国地区长时间序列 SPOT_Vegetation 植被指数数据集[34]，取多年平均值；气温、降雨和辐射数据采用气象再分析资料再经过插值而来。所使用的环境因子列表如表 2，图 2 显示了研究区几种典型的环境因子。

表 2 利用的环境因子

Table 2 Environment factors

环境因子类型	环境因子
母质	地质类型图、地貌类型图
地形	高程、沿剖面曲率、沿等高线曲率、坡度
气候	降雨、气温、太阳辐射
生物	植被类型、植被指数
其他	土地利用与土地覆盖、离河道距离

4 结果及分析

4.1 结果

根据以上土壤剖面数据和 13 类环境因子，分别采用决策树方法、支持向量机方法和模糊逻辑方法对研究区进行土壤预测制图。决策树方法首先利用 See 5.0 软件根据训练样本训练出决策树规则，再根据该规则在 ERDAS Imagine 下进行预测制图；支持向量机方法则首先调用 LibSVM 软件包中的 svmtrain 函数利用训练样本训练出预测模型，再调用 svmpredict 函数根据该模型对未知数据进行预测，本研究中支持向量机核函数采用的是高斯径向基核函数；基于模糊逻辑制图中首先利用决策树方法产生的规则作为土壤存在的典型土壤环境，即根据 See 5.0 软件获得第一类知识，再对该知识规则进行补充或适当修改，获取第二类知识，然后根据该规则在 SoLIM 模型下构造模糊隶属度函数进行土壤制图。图 3 为采用 3 种方法生成的土壤图，表 2 为 3 种土壤图的精度验证结果。

从生成的土壤图（图 3）结构来看，基于支持向量机方法生成的土壤质地详细程度低，基于决策树方法生成的土壤质地结构斑块较为破碎，而基于模糊逻辑方法生成的土壤质地图既较好地保留了斑块结构的完整性，也较好描述了土壤类型与典型环境因子之间的相关性，因而效果相对来说最好。

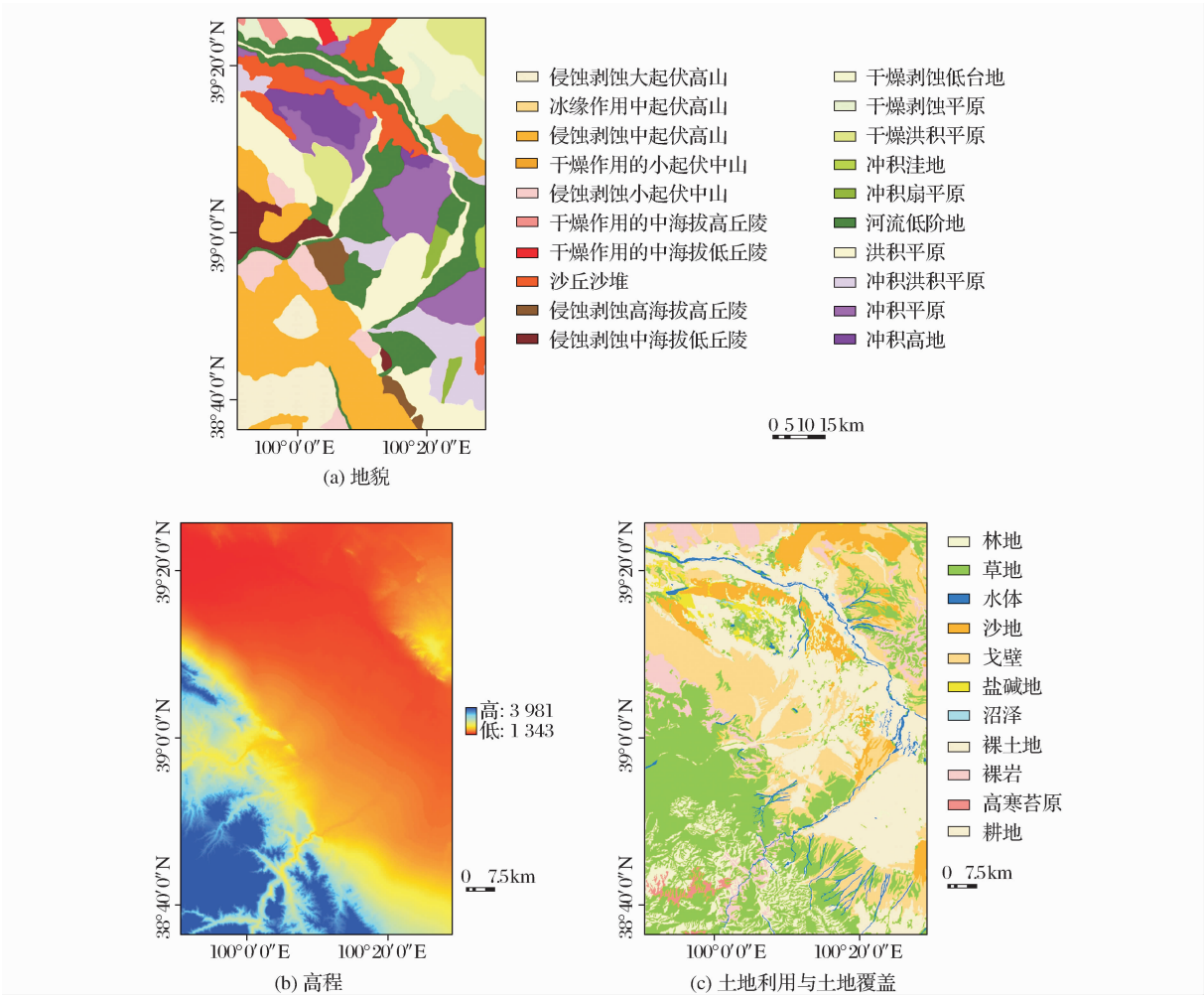

图 2　几种典型环境因子

Fig. 2　Typical environment factors

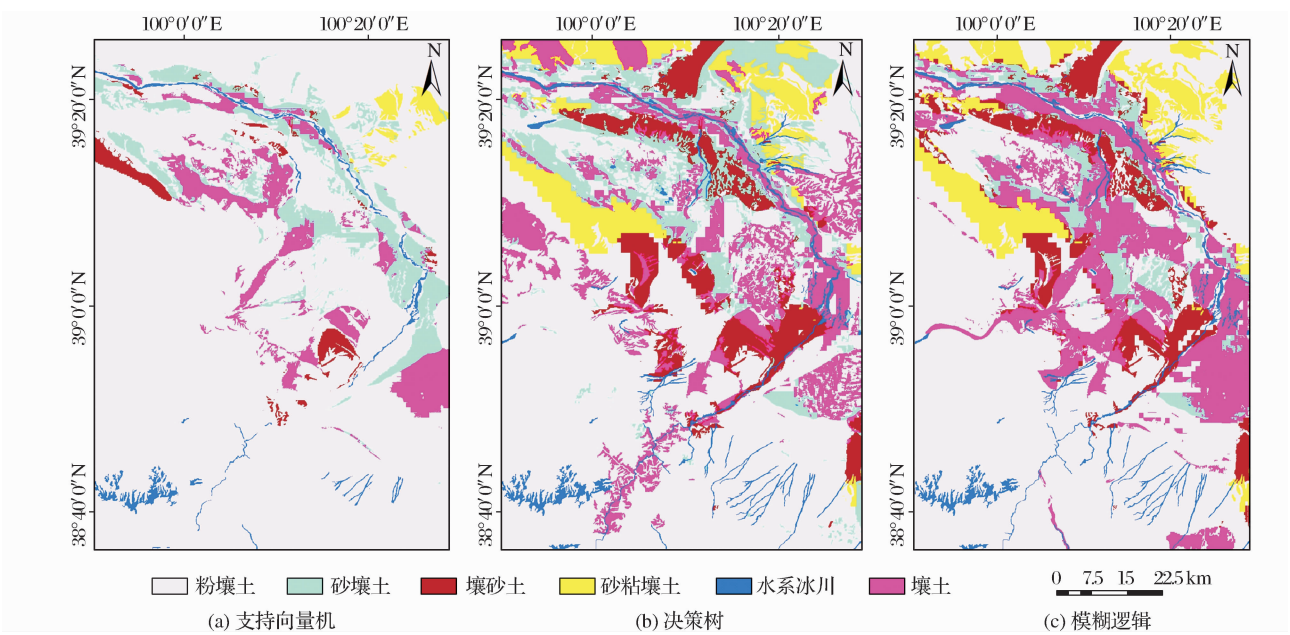

图 3　3 种制图方法分类结果

Fig 3　Soil texture maps derived from different prediction approaches

从土壤类型的预测效果来看，三者都能较好地区分出非土壤的水系和冰川区域，且 3 种土壤图预测的主要土壤类型都为粉壤土，与该区域周边实测剖面土壤质地类型多为粉壤土保持一致。由于基于模糊逻辑和基于决策树方法的规则产生都是根据决策树 C4.5 算法产生，故其斑块格局总体分布相似，但是在基于模糊逻辑制图时加入了一些经验知识，如地貌环境因子中表示沙丘沙堆的分布区域，土壤质地含砂量比其他地貌类型多，同时去掉了部分明显不符合土壤实际情况分布的规则，使生成的土壤质地分布更能反应真实分布情况。在绿洲区土壤质地类型主要随土地利用与土地覆盖、气温和地貌分布变化而变化，且沿河道分布具有较强的分布规律。

表 3 3 种方法生成土壤图精度对比

Table 3 Accuracy comparisons of three soil maps

模型	支持向量机	决策树	模糊逻辑
训练点	0.90	0.98	0.74
验证点	0.55	0.57	0.64

根据本研究区中选取的 129 个训练样点和 71 个测试样点分别进行精度验证，将 3 种方法生成的土壤质地类型图与实测样点的土壤质地类型进行比较。从表 2 中 3 种土壤质地类型预测的精度验证对比来看，支持向量机方法对训练点和验证点的测试精度都偏低，分别为 90％和 55％；基于决策树方法对训练点的验证精度最高，达到 98％，但其对验证点的预测准确率为 57％；基于模糊逻辑方法的综合评价精度最好，对验证点的预测精度能达到 64％，对训练样点的验证精度为 74％。

4.2 结果分析

分析 3 种方法产生精度差异的成因得到：支持向量机方法通常需要对环境因子数据进行预处理，数据统一缩放到[－1,1]或[0,1]之间，以避免一些特征值范围过大而另一些特征值范围过小，以及在训练时为计算核函数而计算内积时引起数值计算的困难。该过程减小了环境因子分布的空间差异性，影响了环境因子与土壤质地分布的协同变化关系。另一方面支持向量机模型只接受数值型输入变量，因此类型环境因子变量需经过特殊预处理过程才能作为输入参数。如表示颜色类型变量红、绿、蓝 3 种颜色，可以表示成三维变量，如(1,0,0)、(0,1,0)、(0,0,1)，或利用间隔一维变量表示，如(0,0.5,1)，试验发现表示成多维变量形式效果更好。但当环境变量类型较多时，会造成数据维数过高，数据冗余大。在本研究实例中，土壤剖面分布不够均匀，数据多集中分布在某几种类型环境因子类型中，如图 3(a)所示，基于支持向量机方法生成的土壤图结构过于简单，信息量小，主要是受该区域训练样本分布和模型特点所限制。

基于决策树方法和模糊逻辑方法不需要对原始数据进行预处理，每一维变量都各自独立，不受制约，也能很好地支持类型数据变量。决策树方法所产生的决策树终端节点中包括全部环境因子组合条件，而由于本研究区中训练样本并不遍布每种环境因子组合，因此使用决策树算法时造成了对未知环境条件预测的不确定性，造成部分区域土壤质地斑块不合理的跳跃分布，如图 3(b)，形成了斑块的破碎化。另一方面，决策树方法容易造成对训练样本的过拟合，而产生对未知预测的不准确，即表 2 中基于决策树方法对训练样本精度可以达到 98％，但对测试样本的预测准确率只有 57％。

基于模糊逻辑方法中第一类知识的获取主要依赖于决策树模型所提取的规则，在此基础上，可以修正一些明显错误的决策规则，并加入部分反映土壤质地类型与环境因子相互关系的专家知识加以补充，如本例中某一决策规则认为土地利用类型若为盐碱地，则土壤质地类型为粉壤土，该规则显然不符合土壤质地分布的实际状况，故将该规则剔除。如图 3(c)，基于模糊逻辑方法生成的土壤图结构更加合理，土壤斑块更加完整；从精度方面来看，模糊逻辑方法对训练点的预测精度偏低，主要原因是其对决策树产生的规则进行了一定简化，详细的多叉决策树转化为规则的过程中会有信息损失；但其预测精度最高，达到 64％，说明其对未知区域的预测效果最好。对训练样点的判别准确率并不是越高越好，对训练样区的预测准确并不代表对其他区域的预测准确，因此，在评价精度时，并不能只考虑训练样本的精度，还需重点考虑验证点所反映的预测精度，并应从地理分布是否合理等方面评价生成的土壤图的合理性。因此综合考虑，根据本研究区的样点分布特点和环境因子数据特点，基于决策树和模糊逻辑相结合的土壤质地制图方法效果较好。

5 结 语

本文选取了黑河流域中游地区的代表性区域，

探索了基于支持向量机、决策树和模糊逻辑 3 种数字土壤制图方法进行土壤质地预测制图。综合来看,支持向量机方法弱化了类型环境因子变量的作用,减小了连续变量的空间分布差异性,使生成的土壤图过于简单,无法反映土壤质地复杂的空间变化;基于决策树方法对未知环境因子组合预测具有不确定性,且容易产生一些不符合土壤质地实际分布的规则,导致土壤图斑块相对破碎,预测准确率偏低;模糊逻辑方法可弥补决策树方法的一些不足,对部分错误规则进行修正,且能生成反映土壤空间渐变分布特征的模糊隶属度图。

使用 3 种方法制作的研究区土壤质地图的预测精度都不太理想,受到几方面影响:① 受土壤样本数量与空间分布的影响。本研究区中土壤剖面样本的空间分布不够均匀,尤其是山区的样点分布过少,有限的训练样点无法全面详细刻画出土壤质地类型与环境的协同变化关系;同时,各土壤类型的土壤剖面分配也不均匀,如砂粘壤土和壤砂土的土壤剖面个数过少,故训练模型时会造成偏差,忽略部分样本分配少的土壤类型的预测。② 受环境因子数据质量影响。土壤质地类型的空间异质性较大,研究中使用的 13 类环境因子中除地形数据分辨率为 90 m,土地利用与土地覆盖图分辨率为30 m外,大部分环境因子数据分辨率都为 1 km,不足以精细刻画环境因子的复杂变化。③ 受到 3 种模型本身的制约。支持向量机需要原始数据进行预处理,该过程损失了部分环境因子包含的丰富的空间分布差异性;决策树方法在树的构造和树剪枝过程中容易对训练样本造成过拟合,而产生不合常理的规则;模糊逻辑方法在构造隶属度函数时,第二类知识获取需要很强的专家经验知识,具有很大的灵活性。3 种方法中模型的训练和预测过程中都需要进行参数调整和参数优化,根据已有的训练样点和验证样点来选择最佳参数,受到剖面质量制约。

因此要提高土壤制图的精度应从以下几方面着手:① 加强对土壤剖面质量的控制及训练样点的筛选工作,能否根据训练样点提取出土壤质地类型分布与环境因子分布的正确关系至关重要,通过设计合适的采样方法,如目的采样设计,提高样点对土壤与环境关系刻画的准确性;② 对环境因子数据的质量控制和筛选,本文利用了 13 种环境因子,在进一步的研究工作中,需要根据研究区域特点对环境因子进行适当筛选,且可以给不同环境因子赋予不同的权重,以突出主导环境因子的控制作用,弱化次要环境因子;③ 对模型参数进行不断调整,以寻找最佳的参数生成更好的预测结果,提高土壤制图精度。

土壤野外采样是一个费资、费时、费力的工作,进行大量采样并不现实。在当前条件下,利用有限的土壤剖面数据,采用决策树和模糊逻辑相结合的数字土壤制图方法,进行黑河流域数字土壤质地制图是一种行之有效的方法,能够满足陆面过程模型、水文模型和大气模型应用需要并提高模拟精度。

致谢:感谢中国科学院寒区旱区环境与工程研究所苏永中研究员、南卓铜研究员和韩旭军博士,南京师范大学李硕老师提供土壤剖面数据,感谢韩旭军博士提供气象再分析资料,感谢李新研究员的悉心指导!

参考文献(References):

[1] Liang Xiao, Dai Yongjiu. A Sensitivity Study of the Common Land Model on Soil Texture and Soil Brightness[J]. Climatic and Environmental Research, 2008, 13(5): 585-597. [梁晓, 戴永久. 通用陆面模式对土壤质地和亮度的敏感性分析[J]. 气候与环境研究, 2008, 13(5): 585-597.]

[2] Gao Y H, Chen F, Barlage M, *et al*. Enhancement of Land Surface Information and Its Impact on Atmospheric Modeling in the Heihe River Basin, Northwest China[J]. Journal of Geophysical Research, 2008, 113(D20): D20S90.

[3] Salazar O, Wesstrom I, Joel A. Evaluation of DRAINMOD Using Saturated Hydraulic Conductivity Estimated by a Pedotransfer Function Model[J]. Agricultural Water Management, 2008, 95(10): 1135-1143.

[4] Bormann H. Towards a Hydrologically Motivated Soil Texture Classification[J]. Geoderma, 2010, 157(3-4): 142-153.

[5] Zhu A X, Mackay D S. Effects of Spatial Detail of Soil Information on Watershed Modeling[J]. Journal of Hydrology, 2001, 248(1-4): 54-77.

[6] Li Xin, Wu Lizong, Ma Mingguo, *et al*. Digital Heihe River Basin2: Data Integration[J]. Advance in Earth Sciences, 2010, 25(3): 306-316. [李新, 吴立宗, 马明国, 等. 数字黑河的思考与实践 2:数据集成[J]. 地球科学进展, 2010, 25(3): 306-316.]

[7] Shi Xuezheng, Yu Dongsheng, Sun Weixia, *et al*. Reference Benchmarks Relating to Great Groups of Genetic Soil Classification of China with Soil Taxonomy[J]. Chinese Science Bul-

letin,2004,49(14):1507-1511.[史学正,于东升,孙维侠,等. 中美土壤分类系统的参比基准研究:土类与美国系统分类土纲间的参比[J]. 科学通报,2004,49(13):1299-1303.]

[8] Zhang Shihuang,Peng Gongbing,Huang Mei. The Feature Extraction and Data Fusion of Regional Soil Textures based on GIS Techniques[J]. Climatic and Environmental Research,2004,9(1):65-79.[张时煌,彭公炳,黄玫. 基于地理信息系统技术的土壤质地分类特征提取与数据融合[J]. 气候与环境研究,2004,9(1):65-79.]

[9] Gao Yanhong,Liu Wei,Cheng Guodong,*et al*. Setup and Validation of the Soil Texture Type Distribution Data in the Heihe River Basin[J]. Plateau Meteorology,2007,26(5):967-974.[高艳红,刘伟,程国栋,等. 黑河流域土壤质地分类数据建立及其模拟效果检验[J]. 高原气象,2007,26(5):967-974.]

[10] Shangguan W,Dai Y,Liu B,*et al*. A Soil Particle-size Distribution Dataset for Regional Land and Climate Modelling in China[Z]. Doi:10.1016/j.geoderma.2011.01.013. Geoderma,2011.

[11] UNESCO. FAO-UNESCO Soil Map of the World[Z]. Paris:UNESCO,1974.

[12] FAO/IIASA/ISRIC/ISSCAS/JRC. Harmonized World Soil Database(Version 1.1)[R]. FAO, Rome, Ltaly and IIASA, Laxenburg,Austria,2009.

[13] Sanchez P A,Ahamed S,Carre F,*et al*. Digital Soil Map of the World[J]. Science,2009,325(5941):680-681.10.1126/science.1175084.

[14] Zhao Liang. Extraction of Quantitative Relationships between Soil Properties and Landscape Factors based on Fuzzy Set Theory and Its Applications in Digital Soil Mapping[D]. Nanjing:Nanjing Agricultural University,2007.[赵量. 基于模糊集理论的土壤—景观定量关系提取及制图应用[D]. 南京:南京农业大学,2007.]

[15] Mcbratney A B,Santos M,Minasny B. On Digital Soil Mapping[J]. Geoderma,2003,117(1-2):3-52.

[16] Goovaerts P. Geostatistical Modelling of Uncertainty in Soil Science[J]. Geoderma,2001,103(1-2):3-26.

[17] Dobos E,Montanarella L,Negre T,*et al*. A Regional Scale Soil Mapping Approach Using Integrated AVHRR and DEM Data[J]. International Journal of Applied Earth Observation and Geoinformation,2001,3(1):30-42.

[18] Bui E N,Moran C J. Disaggregation of Polygons of Surficial Geology and Soil Maps Using Spatial Modeling and Legacy Data[J]. Geoderma,2001,103(1-2):79-94.

[19] Zhao Z,Chow T L,Rees H W,*et al*. Predict Soil Texture Distributions Using an Artificial Neural Network Model[J]. Computers and Electronics in Agriculture,2009,65(1):36-48.

[20] Burrough P A,Macmillan R A,Vandeursen W. Fuzzy Classification Methods for Determining Land Suitability from Soil Profile Observations and Topography[J]. Journal of Soil Science,1992,43(2):193-210.

[21] Zhu A X,Hudson B,Burt J,*et al*. Soil Mapping Using GIS, Expert Knowledge and Fuzzy Logic[J]. Soil Science Society of America Journal,2001,65(5):1463-1472.

[22] Zhang Haoran,Han Zhengzhi,Li Changgang. Support Vector Machine[J]. Computer Science,2002,29(12):135-137.[张浩然,韩正之,李昌刚. 支持向量机[J]. 计算机科学,2002,29(12):135-137.]

[23] Lin C C,Chih-Jen. LiBSVM:a Library for Support Vector Machines[DB/OL]. http://www.csie.ntu.edu.tw/~cjlin/libsvm,2001.

[24] Fan Ming,Meng Xiaofeng. Data Mining:Concepts and Techniques[M]. Beijing:China Machine Press,2001.[范明,孟小峰. 数据挖掘概念与技术[M]. 北京:机械工业出版社,2001.]

[25] Zhou Bin,Wang Fan. Spatial Prediction of Soil Properties based on Decision Tree Modelling[J]. Chinese Journal of Soil Science,2004,4:385-390.[周斌,王繁. 基于决策树模型的土壤性质空间推断[J]. 土壤通报,2004,4:385-390.]

[26] Quinlan J R. Induction of Decision Trees[J]. Machine learning,1986,1(1):81-106.

[27] Quinlan J R. C4.5:Programs for Machine Learning[M]. San Francisco:Morgan Kaufmann,1993.

[28] Zhu A X. A Similarity Model for Representing Soil Spatial Information[J]. Geoderma,1997,77(2-4):217-242.

[29] Zhu Axing. Detail Digital Soil Survey:Models and Methods (in Chinese)[M]. Beijing:Science Press,2008.[朱阿兴. 精细数字土壤普查模型与方法[M]. 北京:科学出版社,2008.]

[30] Zhou Chenghu,Cheng Weiming. 1∶1000 000 Scale Digital Geomorphology Dataset of Western China[DB/OL],2007.[周成虎,程维明. 中国西部 1∶100 万数字地貌数据集[DB/OL],2007.]

[31] China Vegetion Editorial Committee of Chinese Academy of Sciences. 1∶1000 000 Scale China Vegetation Datasets[DB/OL],2001.[中国科学院中国植被图编辑委员会. 中国 1∶100 万植被数据集[DB/OL],2001.]

[32] Mapping Group of Digital Geological Map Database for the PRC. 1∶500 000-Scale Digital Geological Map Database for the PRC[DB/OL]. 1999.[中华人民共和国数字地质图数据库编图组. 中华人民共和国 1∶50 万数字地质图数据库[DB/OL],1999.]

[33] Liu Jiyuan,Zhang Zengxiang,Zhuang Dafang. Study on the Temporal and Spatial Dynamic Remote Sensing Information of LUCC in China in the 1990s[M]. Beijing:Science Press,2005.[刘纪远,张增祥,庄大方. 二十世纪九十年代中国土地利用变化的遥感时空信息研究[M]. 北京:科学出版社,2005.]

[34] Janssens Greet,Ma Mingguo. Long Time Sequenced SPOT_Vegetation Index Dataset of China[DB/OL],2006.[詹森斯·古瑞特,马明国. 中国地区长时间序列 SPOT_Vegetation 植被指数数据集[DB/OL],2006.]

Comparison Analysis on Digital Soil Texture Mapping in an area of Zhangye, Heihe River Basin

LIU Chao[1,2], LU Ling[1], HU Xiao-li[1]

(1. *Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou* 730000, *China*; 2. *Graduate University of Chinese Academy of Sciences, Beijing* 100049, *China*)

Abstract: Soil texture is a key input parameter for the land surface process models, hydrological models and atmospheric models. Many digital soil mapping methods based on the soil-landscape model concept have been widely studied and applied. The general methods include the decision tree algorithm, the support vector machine approach and the fuzzy logic theory. In this paper, a study was conducted to compare the above three different soil mapping methods in an area of Zhangye of Heihe River Basin by integrated 200 ground measured soil samples and 13 types of environmental factors. Meanwhile, different soil texture maps based on the three methods were predicted in the study area respectively. Results show that: the support vector machine model gets relatively low accuracies both for test soil samples and training samples, which are 90% and 55% respectively. The decision tree model gets the highest accuracy of 98% for training datasets among the three methods, but its accuracy for testing data decreases into 57%. The fuzzy logic model gets the highest accuracy of 64% for testing data and a compromise accuracy of 74% for training data. As for the structure characteristics of the texture soil maps, the study finds that the support vector machine model predicts a much simplified soil texture map as it may weak the prediction ability of thematic and continuous environmental factors. The decision tree model often gets unstable predictions of unexpected combinations of environmental factors, resulting in a relatively fragile structure of the soil texture map. The fuzzy logic model predicts the most reasonable soil texture map among them, because it can not only keep the soil texture structure much holistic, but also can illustrate an appropriate relationship between different soil types and different environmental factors. It is suggested that combining the decision tree algorithm with the fuzzy logic theory might be an appropriate method to map the soil texture distribution of Heihe River Basin.

Key words: Soil texture mapping; Decision tree; Support vector machine; Fuzzy logic; Heihe River Basin