

实训项目 14、基于专家知识的决策树分类

1、概述

基于知识的决策树分类是基于遥感影像数据及其他空间数据，通过专家经验总结、简单数学统计和归纳方法等，获得分类规则并进行遥感分类。分类规则易于理解，分类过程也符合人的认知过程，最大的特点是利用多源数据。

专家知识决策树分类的步骤大体上可分为四步：知识（规则）定义、规则输入、决策树运行和分类后处理。难点是规则的获取，可以来自经验总结，如坡度小于 20 度是缓坡等；也可以通过统计的方法从样本中获取规则，如 C4.5 算法、CART 算法、S-PLUS 算法等。

本课程以 Landsat TM5 影像和这个地区对应的 DEM 数据为例，学习基于专家知识决策树分类。数据存放在“\12.基于专家知识决策树分类\数据”文件夹内，影像和 DEM 经过了精确配准。

2、步骤

2.1 规则获取

根据经验和专家知识获取如下规则：

Class1（缓坡植被）： $NDVI > 0.3, slope < 20$

Class2（朝北陡坡植被）： $NDVI > 0.3, slope \geq 20, 90 \leq aspect \leq 270$

Class3（朝南陡坡植被）： $NDVI > 0.3, slope \geq 20, aspect < 90$ 或 $aspect > 270$

Class4（水体）： $NDVI \leq 0.3, b4 = 0$

Class5（裸地）： $NDVI \leq 0.3, b4 \geq 20$

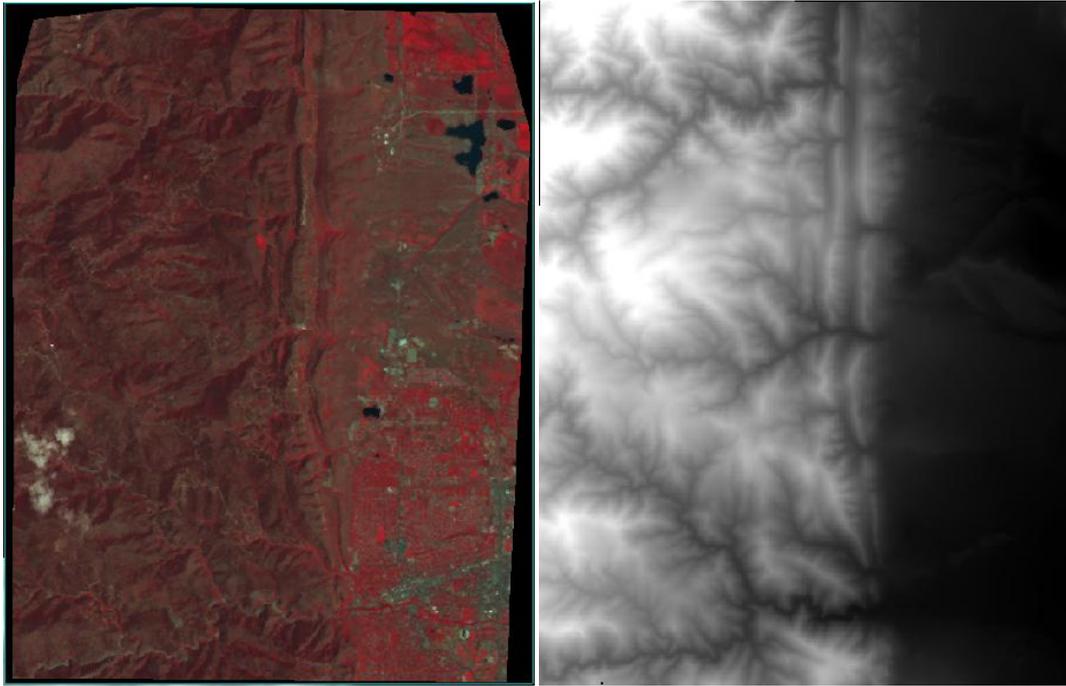
Class6（无数据区，背景）： $NDVI \leq 0.3, b4 = 0$

注：其中，NDVI 为归一化植被指数；slope 为坡度；aspect 为坡向；bN 代表第 N 个波段。

2.2 制作决策树

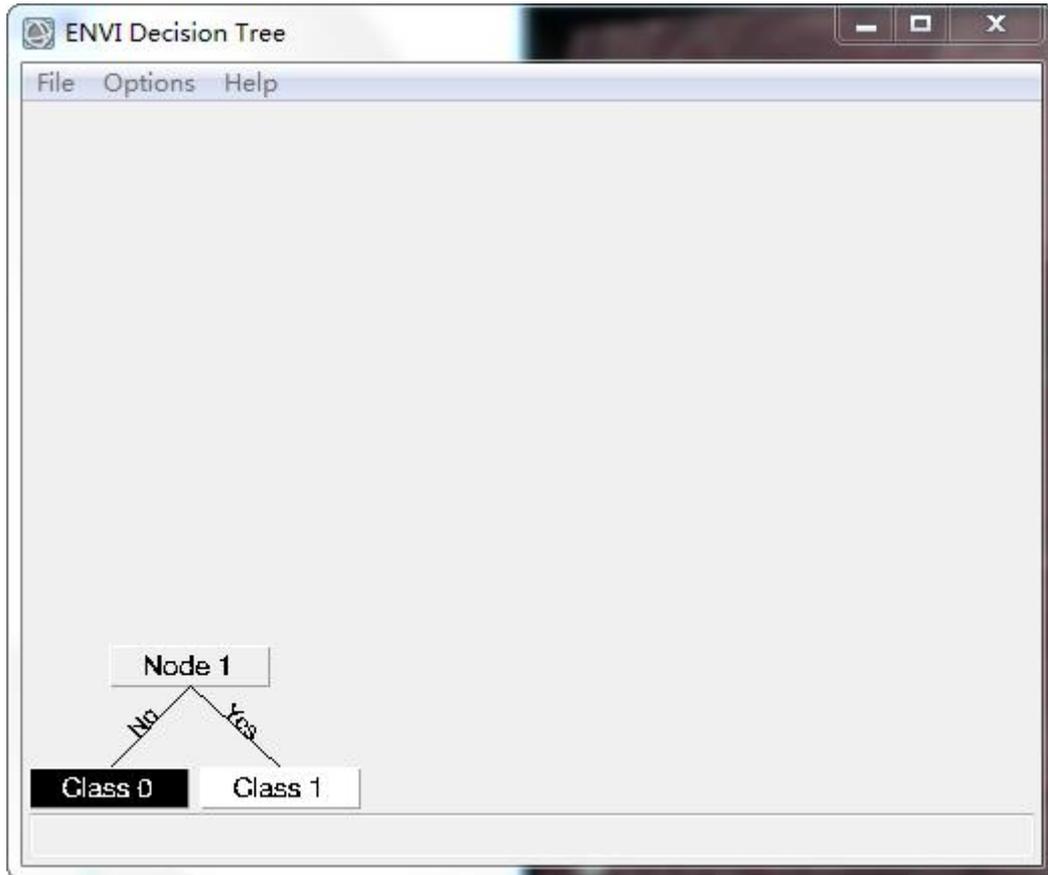
（1）首先打开待分类数据及其他多源数据。打开 File > Open，选择数据文件夹内的 boulder_tm.dat 和 boulder_dem.dat；

注：boulder_tm.dat 为待分类图像，boulder_dem.dat 为 DEM 数据。

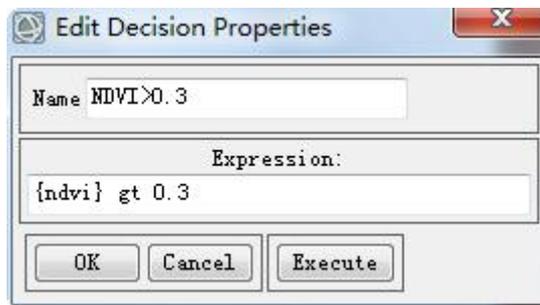


(2) 打开新建决策树工具，路径为 Toolbox/Classification/Decision Tree/New Decision Tree，如下图所示，默认显示一个节点和两个类别；

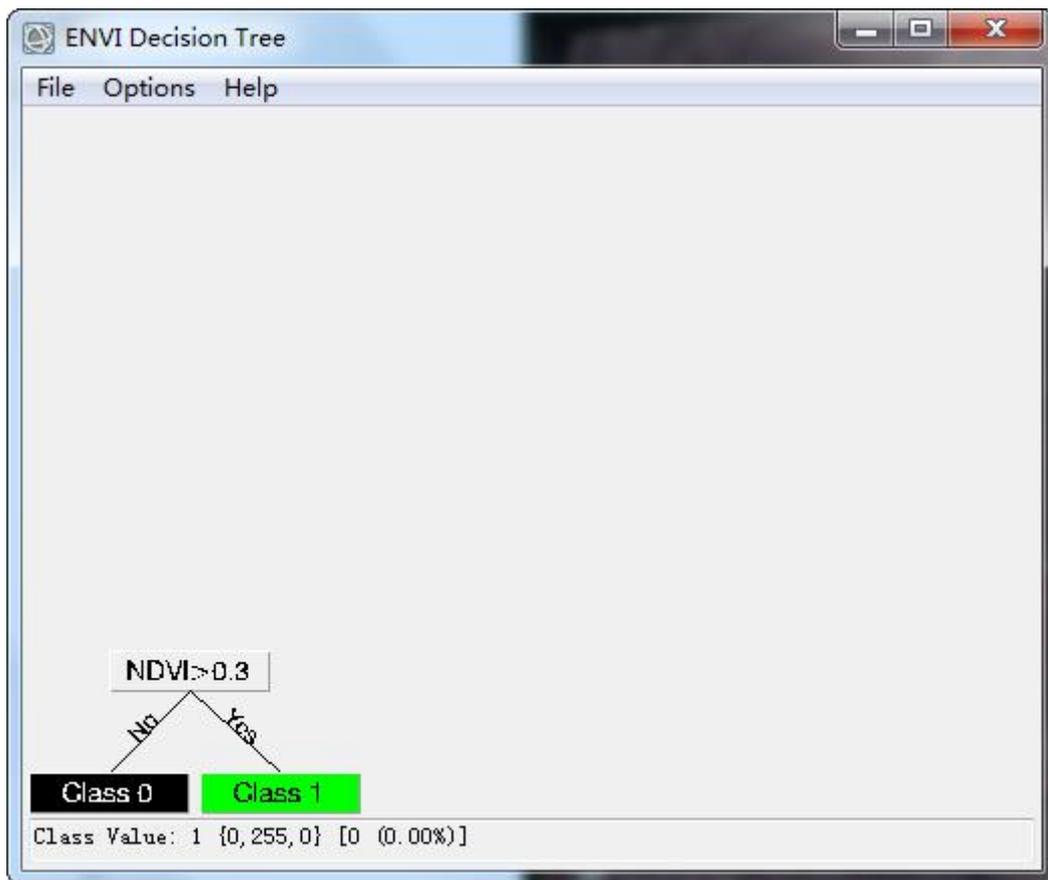




(3) 首先按照 NDVI 来区分植被与非植被。单击节点 Node 1，在弹出的对话框内输入节点名 (Name) 和条件表达式 (Expression)，如下图所示；



(4) 点击 OK 后，在弹出的 Variable/File Pairings 对话框内需要为 {ndvi} 指定一个数据源，如下图所示。点击面板中显示 {ndvi} 的表格，然后选择 boulder_tm.dat 即可。
注：因为所选数据具有波长信息，ENVI 自动根据波长识别红波段与近红外波段，如果没有波长，需要手动指定这两个波段。



(5) 在进行条件表达式 (Expression) 编写时, 需要符合 IDL 的语法规则, 包括运算符和函数名。常用的运算符和函数如下表所示。

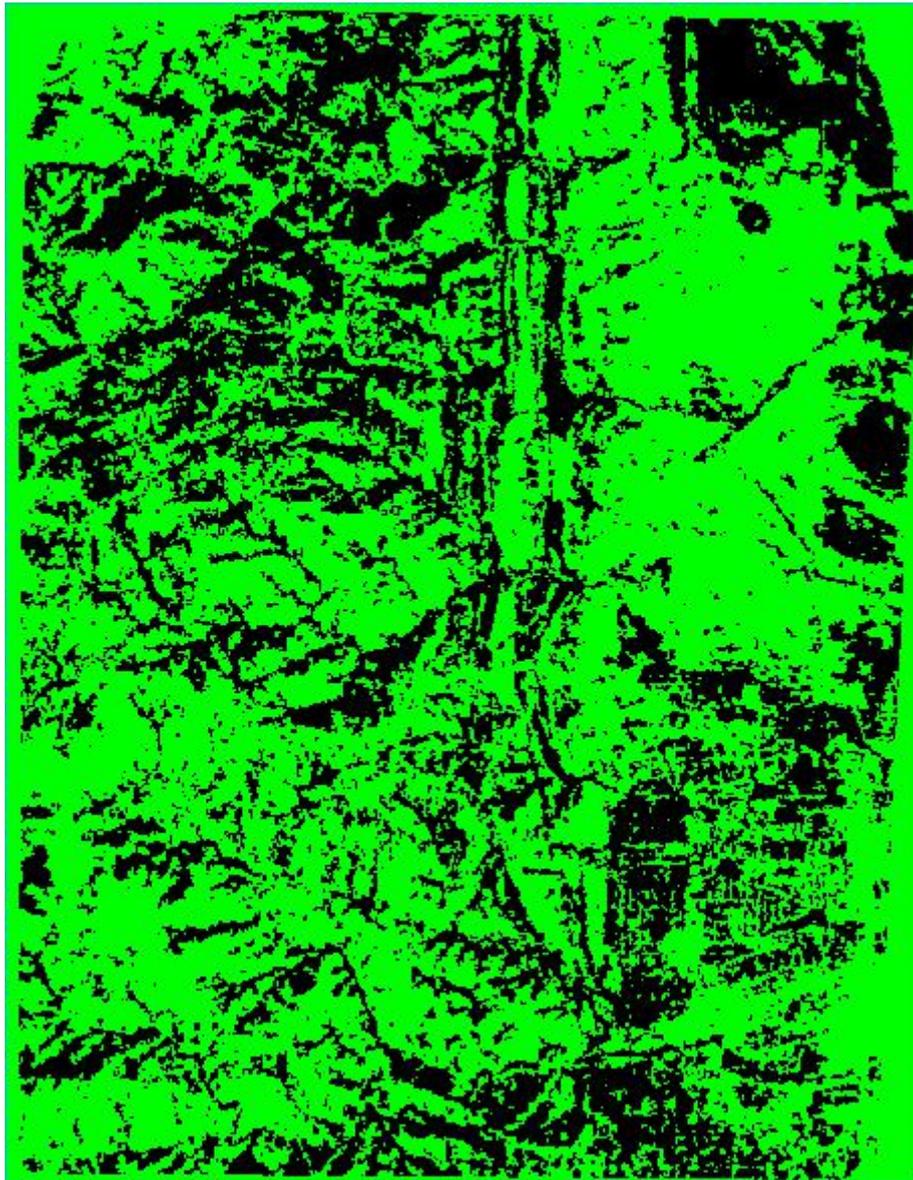
表达式	部分可用函数
基本运算符	+、-、*、/
三角函数	正弦 $\text{Sin}(x)$ 、余弦 $\text{cos}(x)$ 、正切 $\text{tan}(x)$ 反正弦 $\text{Asin}(x)$ 、反余弦 $\text{acos}(x)$ 、反正切 $\text{atan}(x)$ 双曲线正弦 $\text{Sinh}(x)$ 、双曲线余弦 $\text{cosh}(x)$ 、双曲线正切 $\text{tanh}(x)$
关系/逻辑	小于 LT、小于等于 LE、等于 EQ、不等于 NE、大于等于 GE、大于 GT and 、 or 、 not 、 XOR 最大值 ($>$)、最小值 ($<$)

其他符号	指数 (^)、自然指数 exp
	自然对数 alog(x)
	以 10 为底的对数 alog10(x)
	取整——round(x)、ceil(x)、fix(x)
	平方根 (sqrt)、绝对值 (abs)

(6) ENVI 决策树分类器中的变量是指一个波段或作用于数据的一个特定函数。如果为波段，需要命名为 bN，其中 N 为 1~255 的数字，代表数据的某一个波段；如果为函数，则变量名必须包含在大括号中，即{变量名}，如{ndvi}。如果变量被赋值为多波段文件，变量名必须包含一个写在方括号中的下标，表示波段数，比如{pc[1]}表示主成分分析的第一主成分。支持特定变量名，如下表所示，用户也可以通过 IDL 编写自定义函数。

变量	作用
slope	计算坡度
aspect	计算坡向
ndvi	计算归一化植被指数
tascap[n]	穗帽变换，n 表示获取的是哪一分量。
pc[n]	主成分分析，n 表示获取的是哪一分量。
lpc[n]	局部主成分分析，n 表示获取的是哪一分量。
mnf[n]	最小噪声变换，n 表示获取的是哪一分量。
lmnf[n]	局部最小噪声变换，n 表示获取的是哪一分量。
stdev[n]	波段 n 的标准差
lstdev[n]	波段 n 的局部标准差
mean[n]	波段 n 的平均值
lmean[n]	波段 n 的局部平均值
min[n]、max[n]	波段 n 的最大、最小值
lmin[n]、lmax[n]	波段 n 的局部最大、最小值

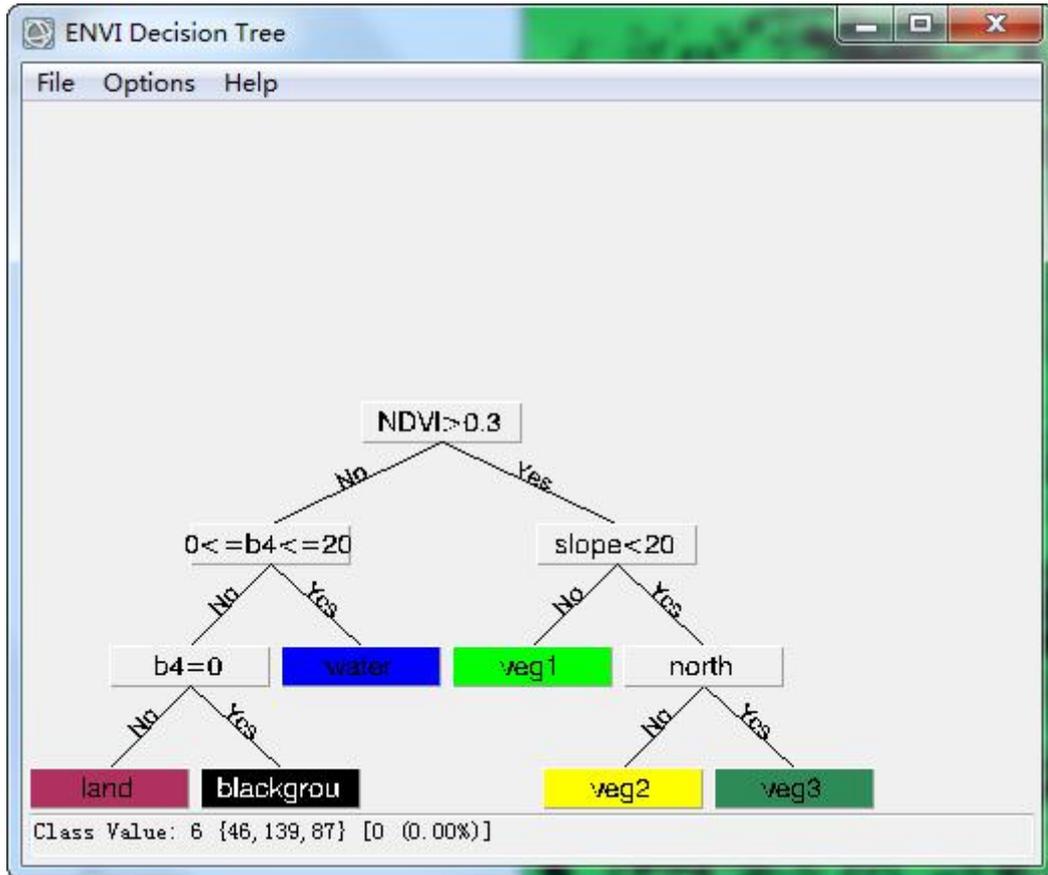
(7) 第一层节点根据 NDVI 的值划分为植被和非植被，如果不需要进一步分类的话，这个影像就会被分成两类：class0 和 class1。



(8) 对 NDVI 大于 0.3，也就是 class1，根据坡度划分成缓坡植被和陡坡植被。在 class1 图标上右键，选择 Add Children。单击节点标识符，打开节点属性窗口，Name 为 Slope<20，在 Expression 中填写：{Slope} lt 20。

节点名	表达式
ndvi>0.3	{ndvi} gt 0.3
$0 \leq b4 \leq 20$	b4 le 20 and b4 ge 0
b4 = 0	b4 eq 0
slope<20	{slope} lt 20
north	{aspect} lt 90 and {aspect} gt 270

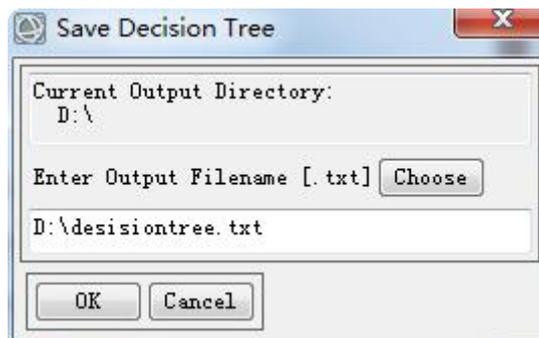
(9) 同样的方法，将所有规则输入，末节点图标右键 Edit Properties，可以设置分类结果的名称和颜色，最后结果如下图所示。



注：可以选择菜单 Options > Show Variable / File Pairings 进行参数与变量的数据源设定。结果如下图所示。



(10) 保存决策树。

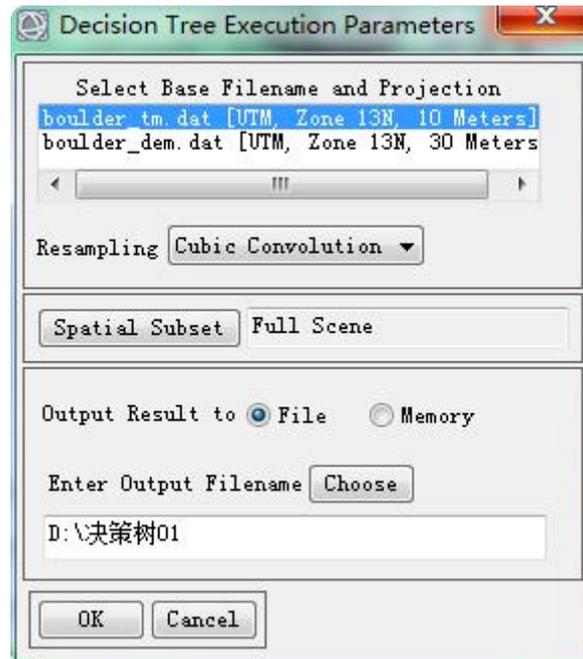


2.3 执行决策树

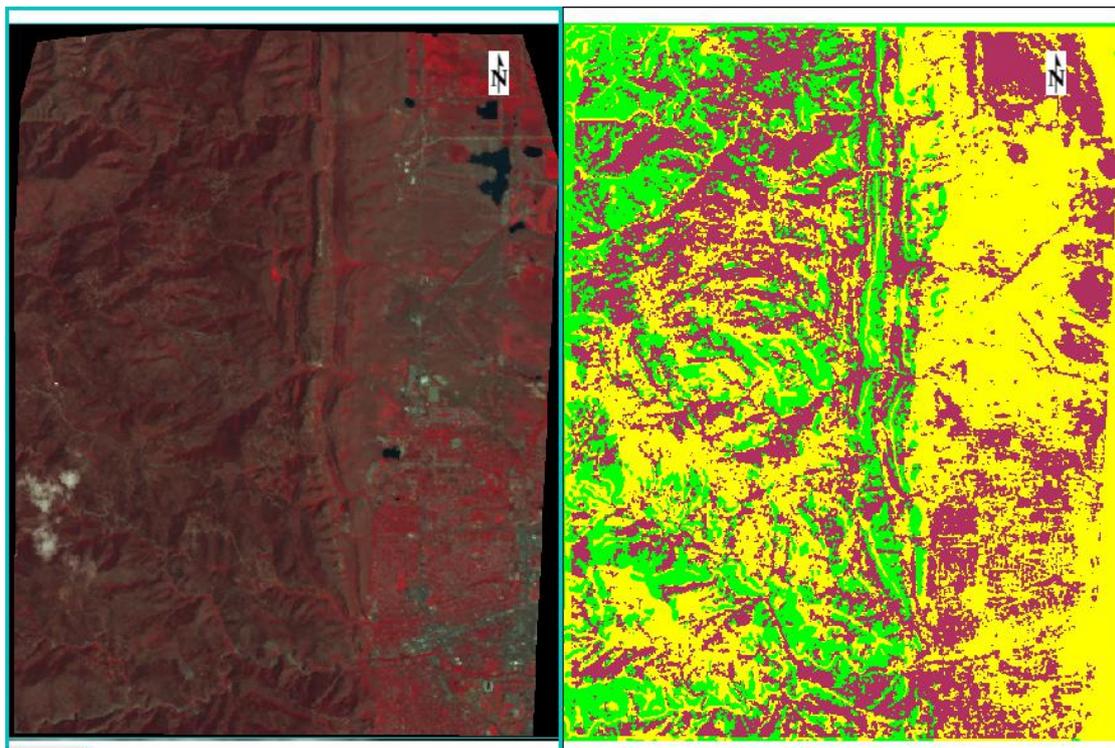
(1) 选择 Options > Execute，可以执行决策树。由于使用了多源数据，各个数据可能拥有不同的坐标系、空间分辨率等。在弹出的 Decision Tree Execution Parameters 对话框（如图）中，需要选择输出结果的参照图像，这里选择 boulder_tm.dat，即输出的分类结果的坐标系和空间分辨率等信息与 boulder_tm.dat 相同。

(2) 选择输出路径和文件名，点击 OK 即可。

注：此步骤中可以选择空间范围裁剪。



(3) 如果 ENVI 没有自动打开结果文件，可以手动打开分类结果。如下图所示。



3、实训要求

- (1) 每人以给定的数据为基础，按照本实训指导书的步骤，进行决策树分类。
- (2) 统计各类地物的面积（单位：平方米），所占百分比。
- (3) 以文字+截图的形式，简要撰写实训报告，包括主要步骤和结果，问题与解决办法，意见和建议等。

4、提交结果

- (1) 最终分类结果：1份，envi 标准格式。
- (2) 实训报告：1份，doc 格式。