

使用流程（案例）

一、数据准备与案例数据说明

【\Data】文件夹为配套的案例原始数据，数据说明见表 1：

表 1 数据说明

数据准备	数据格式	案例数据
研究区边界	面矢量 (.shp)	\Data\point\study_range.shp
研究区历史滑坡数据	点矢量 (.shp)	\Data\point\landslide_point.shp
DEM	30m 栅格 (.tif)	\Data\big_factor\dem.tif
断层	线矢量 (.shp)	\Data\big_factor\faults.shp
岩性	30m 栅格 (.tif)	\Data\big_factor\lithology.tif
道路	线矢量 (.shp)	\Data\big_factor\roads.shp
河流	线矢量 (.shp)	\Data\big_factor\rivers.shp
NDVI	30m 栅格 (.tif)	\Data\big_factor\NDVI.tif
逐月降雨量	NC4 文件 (.nc4)	\Data\big_factor\rainfall*.nc4

注：所有数据均需投影至 UTM 投影坐标系下。

【\Case】文件夹为使用工具箱生成的论文数据，本手册所有案例数据以 ArcGIS 软件中生成为例，数据说明如下：

- ① \Case\big_factor 文件夹：存放原始数据及未批量裁剪的各影响因子数据。
- ② \Case\point 文件夹：存放研究区矢量数据及滑坡与非滑坡点矢量文件。
- ③ \Case\factors 文件夹：存放裁剪后的研究区各因子层数据。
- ④ \Case\dataset 文件夹：存放未进行因子筛选生成的块数据集，包括滑坡 (\landslide) 与非滑坡样本 (\non-landslide)。
- ⑤ \Case\IGR_dataset 文件夹：存放经过因子筛选后生成的块数据集，包括滑坡 (\landslide) 与非滑坡样本 (\non-landslide)。
- ⑥ \Case\model 文件夹：存放生成的不同参数得到的模型及模型评价结果。
- ⑦ \Case\predict 文件夹：存放生成的待预测影像及预测得到的易发性图。

二、SVM-LSM 流程（以吴起县为例）

以陕西省延安市吴起县为例，使用开发的工具箱进行滑坡易发性评价。整体流程图如图 1 所示。首先，收集研究区的历史滑坡数据，并根据研究区滑坡成因

等因素选取合适的滑坡影响因子用于后续研究。注意：在使用中应确保所有数据均处于 UTM 投影坐标系下，否则可能会出现某些未知的错误。由于研究区域内基本没有断层分布，且不受断层影响，因此在该案例中未使用距断层的距离。

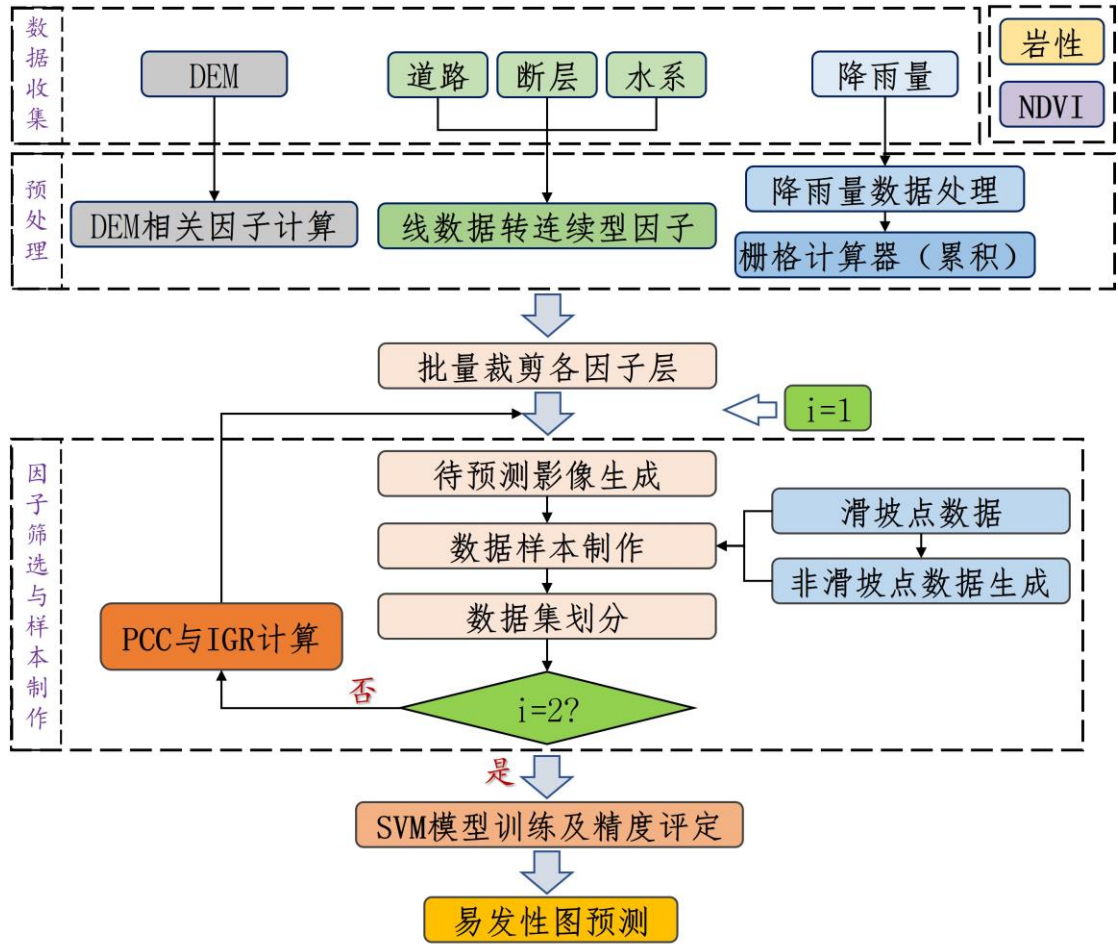


图 1 SVM-LSM 流程图

1、dem 相关因子计算

使用“1 影响因子制作”工具箱内的“1 dem 相关因子计算”工具，根据研究区 DEM 数据自动计算其他地形因素，如坡度、坡向、曲率、平面曲率、剖面曲率、地形起伏度、地表粗糙度、地形湿度指数（TWI）等。注意：DEM 数据必须为 UTM 投影坐标系。计算平面曲率时必须计算坡向；计算剖面曲率、地表粗糙度或地形湿度指数时必须计算坡度。

【输入】

- ① DEM 数据（UTM 坐标系）：【\Case\big_factor\dem.tif】
- ② 工作空间：【\Case\big_factor】
- ③ 计算因子选择：全选。并分别命名【slp, asp, cur, plancur, profilecur, SroughnessC,

relief, TWI】

④ 环境设置：并行处理因子设置为“0”。

【输出】

坡度【\Case\big_factor\dempro\slp.tif】

曲率【\Case\big_factor\dempro\cur.tif】

平面曲率【\Case\big_factor\dempro\plancur.tif】

剖面曲率【\Case\big_factor\dempro\profilecur.tif】

地形起伏度【\Case\big_factor\dempro\relief.tif】

地表粗糙度【\Case\big_factor\dempro\SroughnessC.tif】

地形湿度指数（TWI）【\Case\big_factor\dempro\TWI.tif】

为方便后续批量裁剪，建议将 dempro 文件夹内数据移动至【\Case\big_factor\】文件夹下。



图 2 操作步骤 1

2、线数据转连续型因子

使用“1 影响因子制作”工具箱内的“2 线数据转连续型因子”工具，将研究区线矢量数据自动转换为连续型栅格数据，如道路转换为距道路的距离、断层转换为距断层的距离、水系转换为距水系的距离等。其中，生成的栅格数据分辨

率为 30m, 转换原理采用欧氏距离。注意: 线矢量数据必须为 UTM 投影坐标系。

【输入】

- ① 线矢量数据: 研究区道路数据(UTM 坐标系下)。【\Case\big_factor\roads.shp】
研究区水系数据(UTM 坐标系下)。【\Case\big_factor\rivers.shp】
- ② 文件保存路径: 【\Case\big_factor】
- ③ 环境设置: 并行处理因子设置为“0”。

【输出】

距道路的距离 【\Case\big_factor\roads.tif】

距水系的距离 【\Case\big_factor\rivers.tif】



图 3 操作步骤 2

3、降雨量数据处理（月）

使用“1 影响因子制作”工具箱内的“3 降雨量数据处理（月）”工具，本工具针对从 NASA (<https://gpm.nasa.gov/>) 下载的逐月降雨量数据（.nc4 格式）数据转换为 30m 的 tif 栅格数据。转换后仍为逐月的.tif 栅格数据。

【输入】

- ① 降雨量数据（.nc4）所在文件夹路径: 【\Case\big_factor\rainfall】

② 输出坐标系：【WGS_1984_UTM_Zone_49N】

【输出】

与.nc4 文件对应的逐月降雨量.tif 文件。【\Case\big_factor\rainfall\UTM_30m】

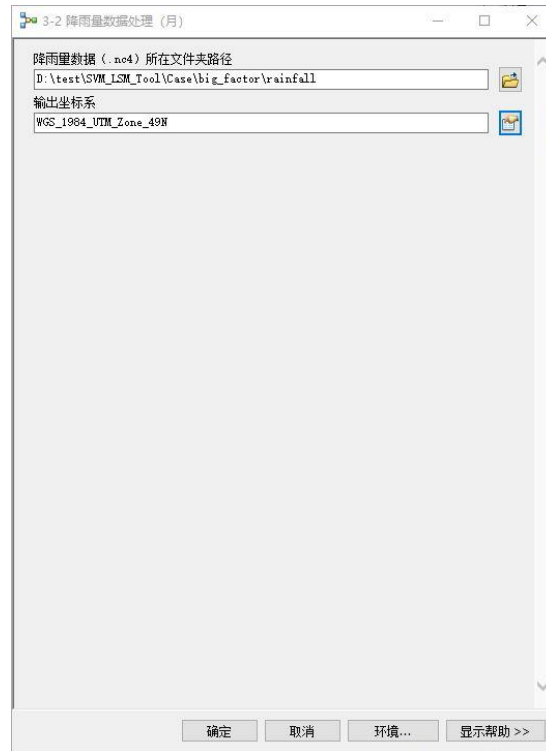


图 4 操作步骤 3

4、年尺度降雨量数据

使用栅格计算器，将所有.tif 累加（严格使用栅格计算器操作，切勿键盘输入“+”等），得到年降雨量，命名为 rainfall.tif

【输入】

转换后的月降雨量栅格数据。

【输出】

年尺度降雨数据 【\Case\big_factor\rainfall.tif】

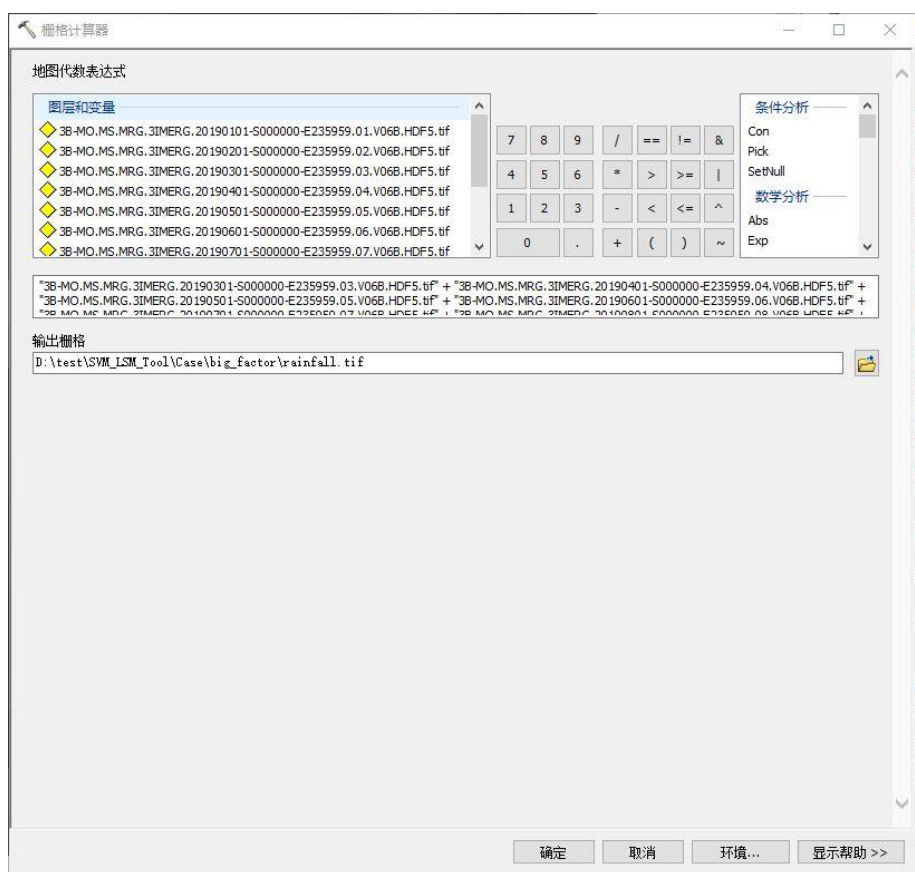


图 5 操作步骤 4

5、批量裁剪各因子层

使用“1 影响因子制作”工具箱内的“4 批量裁剪各因子层”工具，本工具用于根据研究区矢量数据批量裁剪各因子层栅格数据，得到研究区因子层数据。本工具仅需给定栅格因子所在文件夹，自动迭代选择.tif 文件进行裁剪。注意：矢量数据和栅格数据必须是 UTM 投影坐标系，栅格数据分辨率保持一致。

【输入】

- ① 研究区矢量数据（UTM 坐标系）：【\Case\point\study_range.shp】
- ② 输入栅格影像所在文件夹（栅格文件 UTM 坐标系）：【\Case\big_factor】
- ③ 使用输入要素裁剪几何：不选
- ④ 保持裁剪范围：选中
- ⑤ 输出数据文件夹：【\Case\factors】
- ⑥ 环境设置：并行处理因子设置为“0”。

【输出】

裁剪后的研究区因子层栅格数据。【\Case\factors】

【 dem.tif, slp.tif, asp.tif, cur.tif, plancur.tif, profilecur.tif, rivers.tif, roads.tif, lithology.tif, SroughnessC.tif, relief.tif, rainfall.tif, NDVI.tif, TWI.tif】



图 6 操作步骤 5

6、非滑坡点数据生成

使用“2 数据集制作与因子筛选”工具箱内的“1 非滑坡点数据生成”工具，本工具用于在研究区矢量数据图层范围内生成非滑坡点数据。原则：在给定滑坡样本点一定缓冲区范围外随机选取相同数量的非滑坡样本点。注意：研究区矢量数据和滑坡点矢量数据必须是 UTM 投影坐标系，得到的非滑坡点矢量数据默认与滑坡点矢量数据坐标系保持一致。

【输入】

- ① 滑坡点要素 (UTM 坐标系)：【\Case\point\landslide_point.shp】
- ② 距滑坡点距离：1000m。
- ③ 研究区矢量文件 (UTM 坐标系)：【\Case\point\study_range.shp】
- ④ 点数：789。
- ⑤ 输出位置：【\Case\point】
- ⑥ 输出坐标系：不选，默认与滑坡点矢量数据坐标系保持一致。

【输出】

研究区非滑坡样本矢量数据。【\Case\point\non_landslide_point.shp】

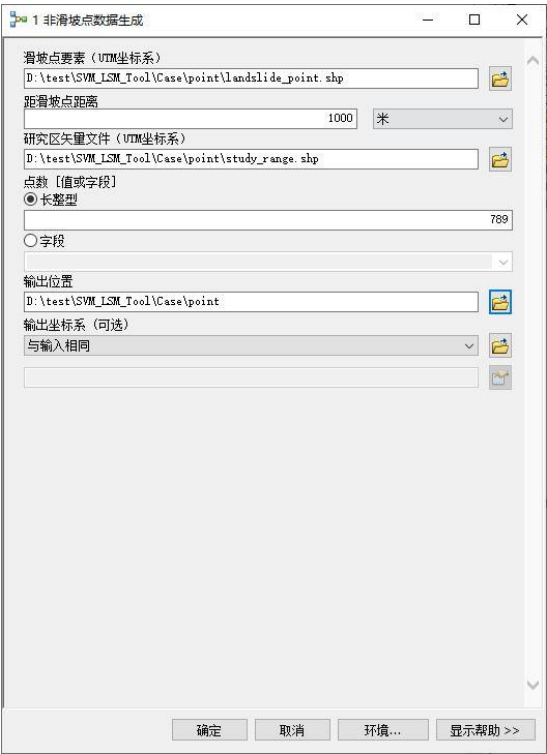


图 7 操作步骤 6

7、待预测影像生成

使用“3 模型训练及预测”工具箱内的“1 待预测影像生成”工具，此工具根据各因子层栅格数据生成待预测影像多通道栅格数据。用于后续数据样本制作及易发性图预测。注意：所有因子栅格数据（.tif）均在同一 UTM 投影坐标系下。得到的多通道栅格数据默认与各因子栅格数据（.tif）坐标系保持一致。

【输入】

- ① 各因子层所在路径：【\Case\factors】
- ② 因子层叠加顺序：【'dem', 'slp', 'asp', 'cur', 'plancur', 'profilecur', 'rivers', 'roads', 'lithology', 'SroughnessC', 'relief', 'rainfall', 'NDVI', 'TWT'】
- ③ 待预测影像保存文件：自动在所选文件夹下创建“Factors_所选因子层数量_mapping.tif”的栅格文件。【\Case\predict】

【输出】

研究区待预测影像多通道栅格数据。【\Case\predict\Factors_14_mapping.tif】



图 8 操作步骤 7

8、数据样本制作

使用“2 数据集制作与因子筛选”工具箱内的“2 数据样本制作”工具，此工具根据矢量点数据生成多通道块样本栅格数据。原则：利用矢量点数据(.shp)制作缓冲区并逐个要素裁剪多通道栅格数据(.tif),得到各要素的单个裁剪结果并以“FID”值命名。注意：矢量点数据(.shp)和多通道栅格数据(.tif)均在同一 UTM 投影坐标系下。得到的多通道块样本栅格数据默认与多通道栅格数据(.tif)坐标系保持一致。

本步骤执行两次，分别为滑坡样本与非滑坡样本。

【输入】

- ① 输入矢量点要素：第一次【\Case\point\landslide_point.shp】，第二次【\Case\point\non_landslide_point.shp】
- ② 缓冲区距离：120m。
- ③ 多通道因子层数据：【\Case\predict\Factors_14_mapping.tif】
- ④ 裁剪后数据存放目录：【\Case\dataset】
- ⑤ 样本标签类别 (landslide or non-landslide)：第一次 landslide，第二次 non-landslide。

- ⑥ 使用输入要素裁剪几何：未选中
- ⑦ 保持裁剪范围：选中

【输出】

研究区滑坡与非滑坡数据样本。本例中为 15*8*8 大小的数据。

【\Case\dataset\landslide, \Case\dataset\non-landslide】



图 9 操作步骤 8-1



图 10 操作步骤 8-2

9、数据集划分

使用“2 数据集制作与因子筛选”工具箱内的“3 数据集划分”工具，此工具根据滑坡点数量及上一步生成的数据样本，按照测试集比例划分训练集和测试集，并以 txt 保存划分结果。

【输入】

- ① 样本所在目录：【\Case\dataset】
- ② 滑坡点个数：789。
- ③ 测试集比例（如 0.3）：0.3。

【输出】

【\Case\dataset.txt】所有样本路径及标签。

【\Case\datasettrain.txt】所有训练样本路径及标签。

【\Case\datasettest.txt】所有测试样本路径及标签。

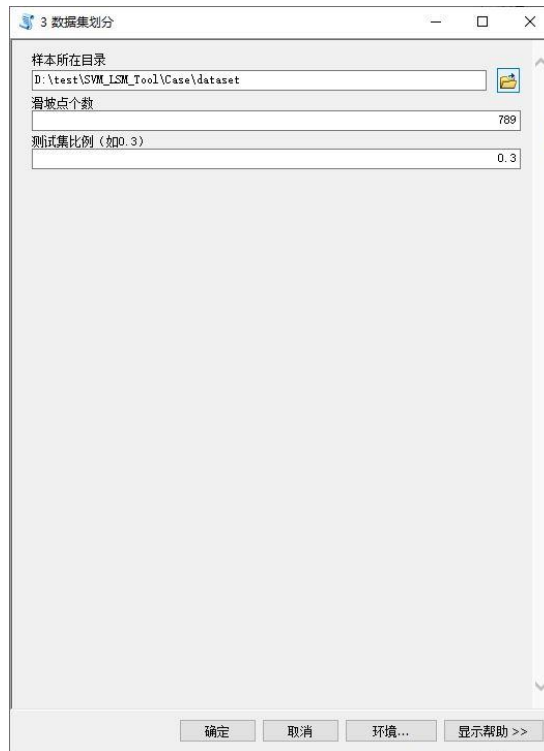


图 11 操作步骤 9

10、PCC 与 IGR 计算

使用“2 数据集制作与因子筛选”工具箱内的“4 PCC 与 IGR 计算”工具，此工具基于生成的数据样本与【\Case\dataset.txt】文件，计算各影响因子层的皮尔逊相关系数与信息增益比。皮尔逊相关系数代表的因子层之间的相关性大小，其值位于[-1, 1]之间，相关性较大的因子应考虑剔除。信息增益比代表各因子层对滑坡发生的贡献大小，其值大于 0 说明对滑坡发生有贡献，值越大，贡献越大。

【输入】

- ① 数据集所在目录：【\Case\dataset】
- ② 结果保存路径：【\Case】
- ③ 输入因子层顺序：【'dem', 'slp', 'asp', 'cur', 'plancur', 'profilecur', 'rivers', 'roads', 'lithology', 'SroughnessC', 'relief', 'rainfall', 'NDVI', 'TWT'】

【输出】

在【\Case\】文件夹下生成【PCC.txt, PCC.png, PCC.svg, IGR.txt, IGR.png, IGR.svg】六个文件。

【*.txt】PCC 或 IGR 结果 txt 文件。

【*.png】PCC 或 IGR 结果图 png 文件。

【*.svg】PCC 或 IGR 结果图可编辑矢量文件。

综合考虑 PCC 与 IGR 结果，剔除坡度和地形起伏度两个冗余因子，其余 12 各用于后续研究。

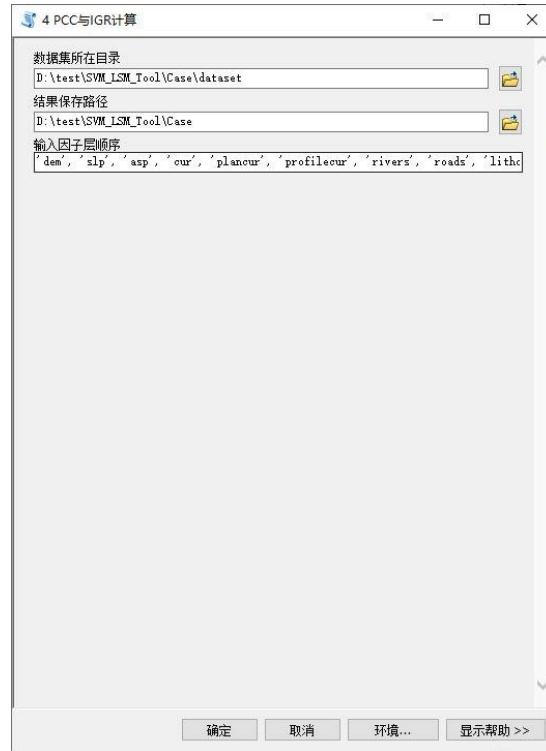


图 12 操作步骤 10

11、更新待预测影像及样本

根据上一步筛选结果，以信息增益比递减的顺序对筛选后的 12 个影响因子进行排序，重新制作待预测影像及数据样本生成。

重复 7、8、9 步骤。其他选项保持不变，除了：

第 7 步中因子层叠加顺序更新为【'lithology', 'plancur', 'profilecur', 'NDVI', 'TWI', 'asp', 'SroughnessC', 'rivers', 'dem', 'roads', 'rainfall', 'cur'】

第 8 步中多通道因子层数据更新为 【\Case\predict\Factors_12_mapping.tif】，裁剪后数据存放目录更新为 【\Case\IGR_dataset】

第 9 步中样本所在目录更新为 【\Case\ IGR_dataset】

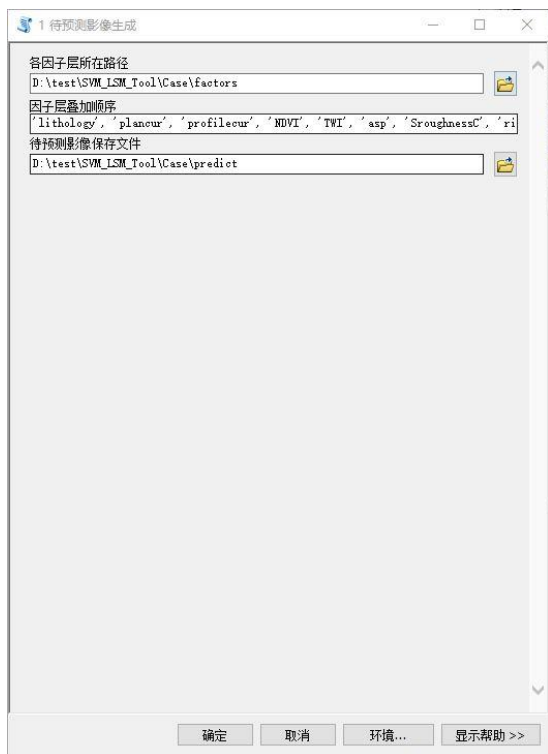


图 13 操作步骤 11-1



图 14 操作步骤 11-2



图 15 操作步骤 11-3

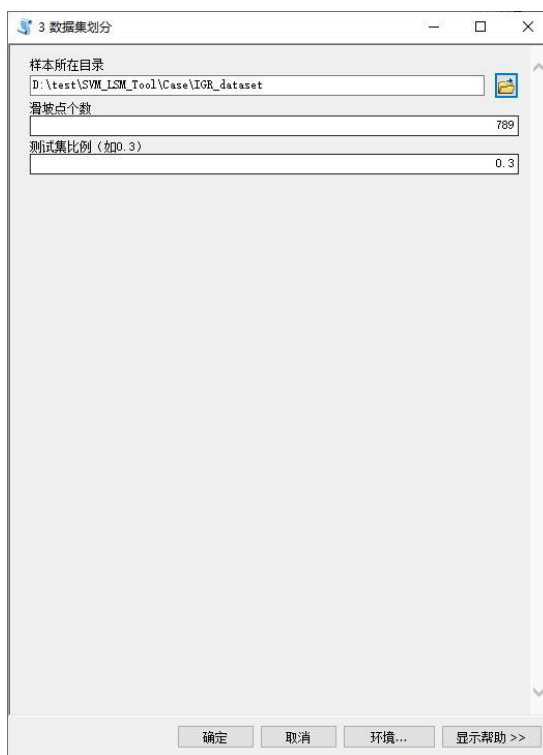


图 16 操作步骤 11-4

12、SVM 模型训练及精度评定

使用“3 模型训练及预测”工具箱内的“2 SVM 模型训练及精度评定”工具，此工具用于生成各组参数下的 SVM 模型并给出模型精度的评定结果。SVM 核

函数本工具默认为径向基核函数（RBF），需要调节的参数为 **gamma** 与惩罚因子 **C**，使用的调参方法为网格搜索算法。

【输入】

- ① 数据集样本所在目录：【\Case\IGR_dataset】
- ② 模型保存路径：该路径下会以“**g_g 值_C_C 值**”新建不同参数文件夹存放结果。【\Case\model】
- ③ **gamma** 参数取值：0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.5, 0.8, 1, 2, 5。
- ④ **C** 参数取值：0.01, 0.02, 0.05, 0.08, 0.1, 0.2, 0.5, 0.8, 1, 2, 5。
- ⑤ 数据集行数：8。
- ⑥ 数据集列数：8。
- ⑦ 数据集通道数：12。

【输出】

【\Case\model】：存放各组参数训练的模型及精度评定指标文件夹。

【\Case\model\parameter_result_txt.txt】存放各组参数下的 SVM 模型在测试集上的 AUC 值、测试集精度 Test_acc、训练集精度 Train_acc 及两者之差。用于最优模型的选择。

【\Case\model\parameter_result_png.png】txt 文件绘图展示。图中，圆形大小代表 AUC 值，圆形越大，AUC 值越高，模型精度越好；圆形颜色代表在训练集和测试集上的精度之差，超过 0.5 的以 0.5 计，差值越大，表明模型过拟合程度越高，泛化性能越差。

以 **gamma** 为 0.02，**C** 为 2 为例：

【\Case\model\g_0.02_C_2】：gamma 为 0.02，**C** 为 2 的 SVM 结果存放路径。

【\Case\model\g_0.02_C_2\SVM_g_0.02_C_2.model】gamma 为 0.02，**C** 为 2 下的 SVM 模型。

【\Case\model\g_0.02_C_2\SVM_train_result_txt.txt】gamma 为 0.02，**C** 为 2 下的 SVM 模型在训练集上的预测结果。

【\Case\model\g_0.02_C_2\SVM_test_result_txt.txt】gamma 为 0.02，**C** 为 2 下的 SVM 模型在测试集上的预测结果。

【\Case\model\g_0.02_C_2\evaluate_result.txt】gamma 为 0.02，**C** 为 2 下的

SVM 模型在测试集上的各类评价指标，如混淆矩阵，准确度，精度，F1 值，AUC 值等。

【\Case\model\g_0.02_C_2\ROC.png】gamma 为 0.02，C 为 2 下的 SVM 模型在测试集上的 ROC 曲线及 AUC 值。

根据【\Case\model\parameter_result_txt.txt 与 parameter_result_png.png】选取 AUC 值较高且差值较小的模型作为最优模型用于易发性图预测。

本例中最优模型为 g_0.02_C_2。

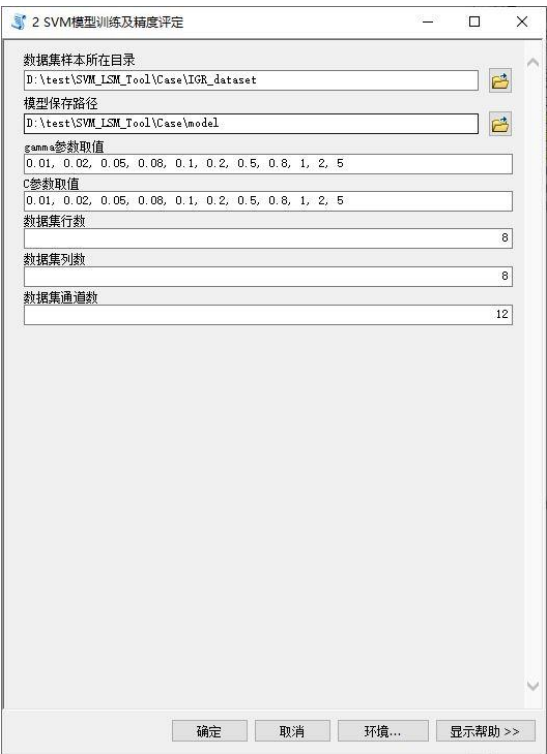


图 17 操作步骤 12

13、易发性图预测

使用“3 模型训练及预测”工具箱内的“3 易发性图预测”工具，单进程或多进程自由选择。此工具根据上一步得到的最优参数模型对研究区滑坡易发性图进行预测。得到的易发性图坐标系与输入的待预测影像一致。

【输入】

- ① 待预测影像路径：【\Case\predict\Factors_12_Mapping.tif】
- ② 研究区矢量边界：【\Case\point\study_range.shp】
- ③ 模型路径：【\Case\predict\model\g_0.01_C_0.8\SVM_g_0.01_C_0.8.model】
- ④ 易发性图保存路径：【\Case\predict\LSM.tif】

- ⑤ 数据集行数：8。
- ⑥ 数据集列数：8。
- ⑦ （多进程参数）pythonw.exe 所在路径：【C:\Python27\ArcGIS10.8】

【输出】

研究区滑坡易发性图。【\Case\predict\LSM.tif】



图 18 操作步骤 13



图 19 操作步骤 13