

基于 SVM 的水稻叶病识别

目录

基于 SVM 的水稻叶病识别.....	1
1 观察数据集.....	2
2 分类预测的初步思路.....	2
3 数据预处理.....	3
4 实验过程.....	5
5 结果.....	7
6 结论.....	8
附录（关于此次实验）.....	9

1 观察数据集

数据集来源：

<http://archive.ics.uci.edu/ml/datasets/Rice+Leaf+Diseases>

共有三类疾病：

1. Bacterial leaf blight (细菌性叶枯病)
2. Brown spot (褐斑病)
3. Leaf smut (叶黑穗病)

每类疾病 40 张 jpg 图片，且大部分尺寸为 3081x897，但不全是，所以在训练和测试过程中需要 `resize`。

因为数据集较小，所以神经网络算法并不适用。选择 **SVM** 进行多分类。

2 分类预测的初步思路

1. 数据预处理
2. 提取若干特征
3. 将数据集分为训练集和测试集。训练集为每类的前 32 张图，测试集为每类的后 8 张图。
4. 训练 **SVM** 进行多分类

3 数据预处理

将数据集文件夹及其文件批量重命名为序号形式，以便批量输入到模型中进行训练和预测。具体地将数据集子文件夹重命名为 1、2、3，并分别将其内图片重命名为“DIR_NUM (PIC_NUM). jpg”。样例如图所示。

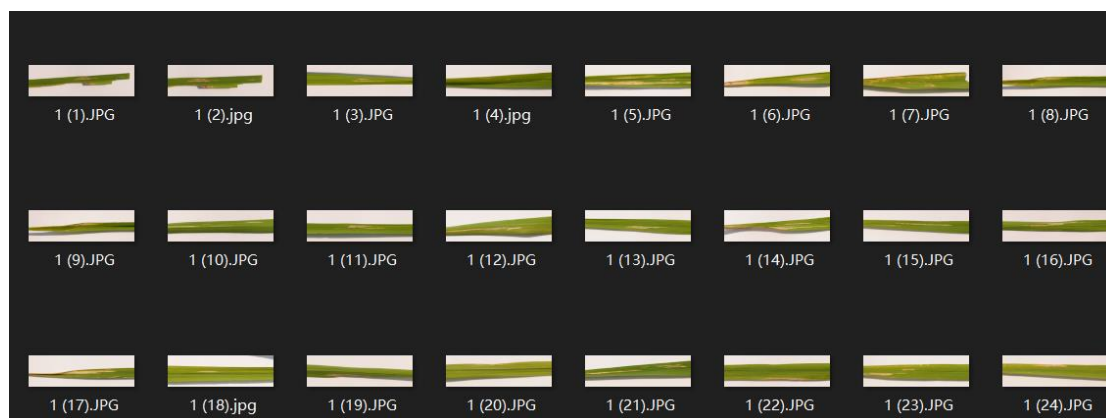


图 数据集子文件夹 1 内图片重命名样例

对图像进行分割，得到患病部分图像，然后将所有患病部分图像 `resize` 成统一大小。

如：

将下图通过一些图形学处理：

原图



图 水稻叶病样例图

得到分割后的患病部分图像：

seg img



图 分离的水稻叶病

然后 `resize` 成统一大小。

4 实验过程

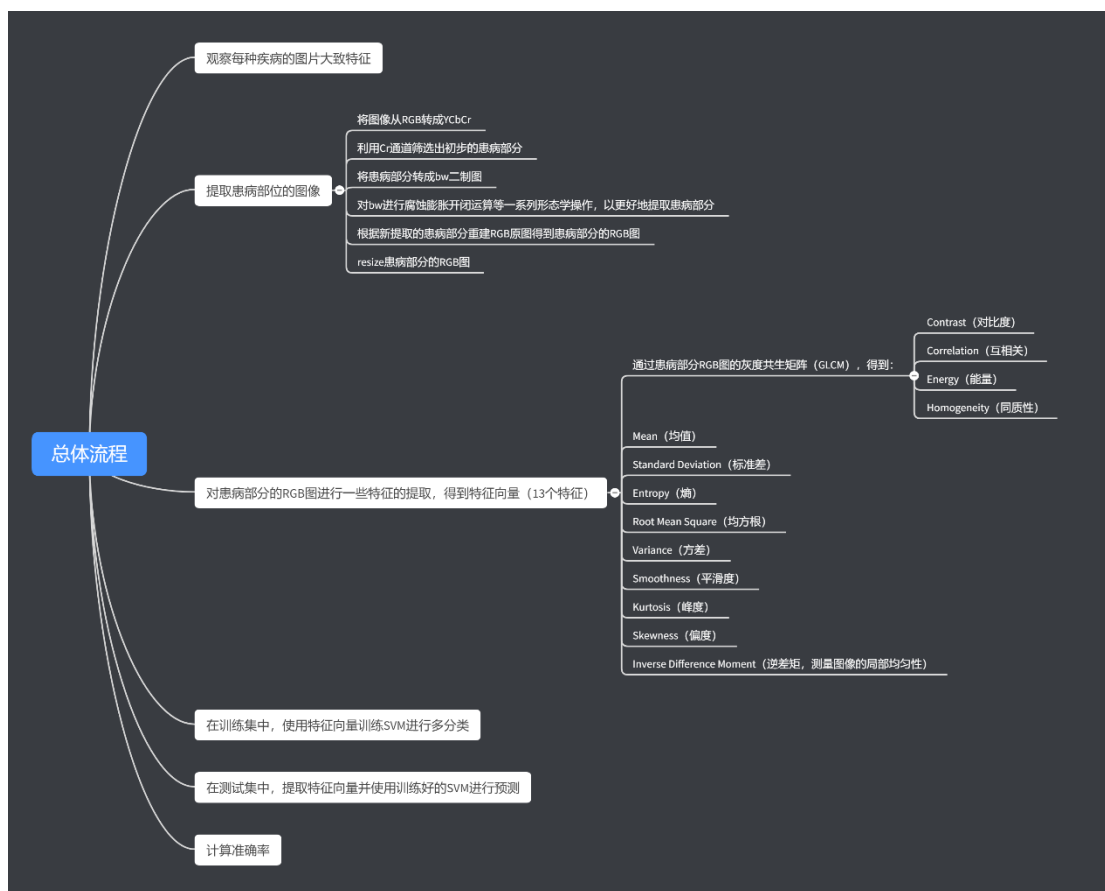


图 思维导图

1. 观察每种疾病的图片大致特征

2. 提取患病部位的图像：

1. 将图像从 RGB 转成 YCbCr，这种颜色模式更适合叶子患病部位的分离。
2. 利用 Cr 通道筛选出初步的患病部分。
3. 将患病部分转成 bw 二制图。

4. 对 `bw` 进行腐蚀膨胀开闭运算等一系列形态学操作，以更好地提取患病部分。
 5. 根据新提取的患病部分重建 RGB 原图得到患病部分的 RGB 图。
 6. `resize` 患病部分的 RGB 图。
3. 对患病部分的 RGB 图进行一些特征的提取，得到特征向量（13 个特征）

通过患病部分 RGB 图的灰度共生矩阵（GLCM），得到：

Contrast（对比度）

Correlation（互相关）

Energy（能量）

Homogeneity（同质性）

Mean（均值）

Standard Deviation（标准差）

Entropy（熵）

Root Mean Square（均方根）

Variance（方差）

Smoothness（平滑度）

Kurtosis (峰度)

Skewness (偏度)

Inverse Difference Moment (逆差矩, 测量图像的局部均匀性)

- 4. 在训练集中, 使用特征向量训练 SVM 进行多分类
- 5. 在测试集中, 提取特征向量并使用训练好的 SVM 进行预测
- 6. 计算准确率。

5 结果

测试集准确率: 75% (18/24) 。

表 正确预测的类别分布

类别	名称	准确占比	同类占比
Class 1	Bacterial leaf blight (细菌性叶枯 病)	7/18	7/8
Class 2	Brown spot (褐斑病)	5/18	5/8
Class 3	Leaf smut (叶 黑穗病)	6/18	6/8

Total			18/24 (75%)
-------	--	--	-------------

```

test accuracy: 75%
Correctly predicted class distribution:
Total right/Total number: 18/24
Class 1: Bacterial leaf blight (细菌性叶枯病), right: 7/18
Class 2: Brown spot (褐斑病), right: 5/18
Class 3: Leaf smut (叶黑穗病), right: 6/18

```

图 运行结果

6 结论

1. 总体来说，预测的准确率还能接受，75%。还有很大的提升空间（在优化患病部位图像提取的情况下）。
2. 可以看出对于第一类患病图像，预测准确率高达 7/8。这是因为在分割患病图像阶段，该类的患病部位提取较为简单且精确。
3. 而第二、三类的预测准确率并不是很高，是因为这两类患病图像特征并不是很好提取，在数据集中可以发现，患病图像的差异较大。
4. 不足：

图像处理阶段：

对于第一类患病图像，分割较容易，但对于后两种患病图像，一些图像的分割效果并不理想，所以影响了准确率。

特征提取阶段：

因为对图像特征的了解有限，选取的特征中有些可能并不

适合本实验，不合适的特征也影响了准确率。

附录（关于此次实验）

此次实验是本学期我们学的数据挖掘课程的最后一个综合实验的选题 1。