

[基于自动机器学习的水色图像水质评价]

[文档副标题]



2020-4-6

马倩

目录

**一．背景介绍**2

1.背景2

2.项目目标2

**二．数据说明**2

**三．分析方法&过程**3

**四．数据预处理**3

1.专家样本4

2.图像切割4

3.特征提取4

**五．模型构建**6

1.决策树6

2.SVM6

**六．结果分析**7

**七．水质评价**8

1. 背景介绍

1.背景：

在很火热的一档综艺节目《最强大脑》中，水哥凭借“微观辨水”一战成名，现场观众依次将同一水源倒入520个同款水杯中，然后从520杯水中挑选一杯供水哥无触碰地观察，然后把这杯水放回原处，结果水哥正确辨识则挑战成功。正所谓，念念不忘，必有回响。

水产养殖业是我国国民经济的一个重要组成部分，在养殖过程中，无污染的水源对养殖起着至关重要的作用。而后水产养殖的企业找到公司，希望可以像水哥一样通过观察水色变化来调控水质。从而维持养殖水体生态系统中浮游植物、微生物类、浮游动物等合理的动态平衡，有经验的渔业从业者能够通过水体的颜色来判断水质的好坏，从而判断是否适合动植物生长。但在生活中，仅凭肉眼和经验对水体观察，总会存在一定的误差。而且还得日积月累才能造就这样本领。并且人与人的对颜色的辨识度偏差，识别能力差异，经验的非文字化传授性，所以这种方法实用性不够强。随着计算机视觉技术图像的发展，为了水产养殖业的可持续化发展，利用机器视觉代替人工视觉已经在很多领域应用。所以也可以给水产养殖业一定的技术支持。

2.项目目标

(1)进行探究不同的水质，它们的不同点在哪里，我们可以根据什么特征来识别不同水质；

(2)利用这些特征数据去总结，学习出一个处理系统或者模型出来，使它具备水质类别划分的能力。

具体可以分为以下几个步骤：

* 提取出基于图片的水质关键特征
* 基于关键特征构建水质评价模型
* 使用自动机器学习实现参数、算法的自动选择，并将其跟手动调节的模型进行比较

1. 数据说明

* 图像来源和获取：将所有采样的全部用一个桶装，在拍照时将其放入预先搭建的“小黑屋”，排除光源及其他因素的影响，而后将获得的水源放置在固定位置。保证每次水源接收到的光源都是一样的，并确保拍照前，水面平静，没有波纹。从而对不同的水源进行拍照。
* 图像数据量：共203张水样图片（共五个类别标签）
* 原数据样本：如图1

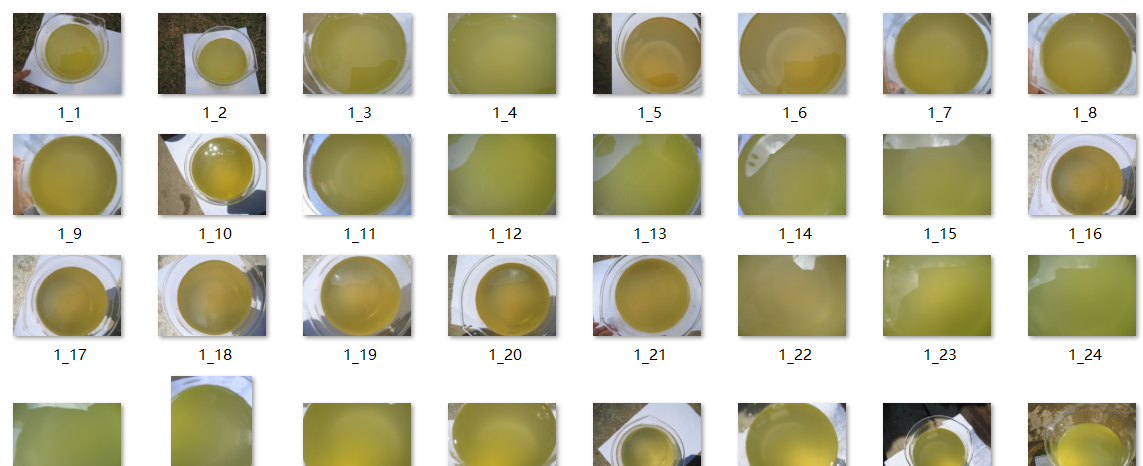


图1

图像详见:/image/

1. 分析方法&过程

通过控制变量法来采取了203张水样图像，由于样本图像维度过大需要对其进行特征提取。通过提取图像本质的一些关键指标，从而达到自动识别分类的目的。显然特征提取的结果直接影响图像识别分析的结果。

基于水色图像特征提取的水质评价流程，步骤如下：

从采集的图像中通过选择抽取和实时抽取形成建模数据和增量数据；

* 对建模数据和增量数据进行数据预处理（图像切割&特征提取）；
* 按照专家样本对数据进行分类；
* 构建模型；
* 利用构建好的模型对水质进行评价。

1. 数据预处理
2. 专家样本：

|  |  |
| --- | --- |
| 水色 | 水质类别 |
| 浅绿色（清水或浊水） | 1 |
| 灰蓝色 | 2 |
| 黄褐色 | 3 |
| 茶褐色（姜黄，茶褐，红褐，褐中带绿等） | 4 |
| 绿色（黄绿，蓝绿，油绿，墨绿） | 5 |

2.图像切割:

* 选取原始图像中中间固定位置（中心位置100\*100像素）

切割示例：将如图2所示的水质图像切割如图3所示

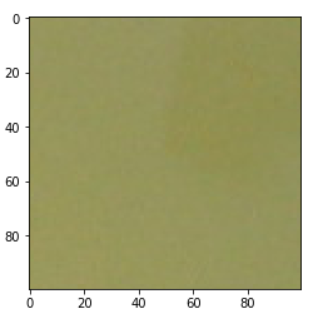
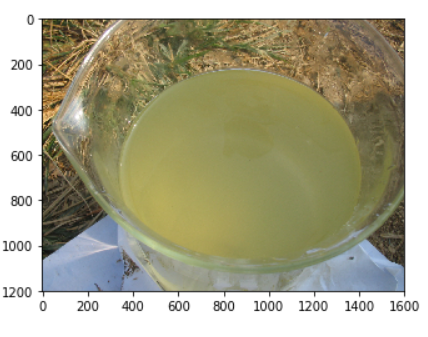


图2 图3

详见/image/

3.特征处理：

图像特征主要包括颜色、纹理、形状和空间关系特征。

本案列利用图像的颜色特征进行图像处理识别分类，主要分为颜色直方图和颜色矩。颜色直方图产生的特征维数一般大于颜色矩的特征维数。为了更好的对样本分类，所以采用颜色矩来提取水质图像的关键特征。

（1）一阶颜色矩

一阶颜色矩采用一阶原点矩，反映图像的整体明暗程度。

Ei表示第i个颜色通道的一阶颜色矩，对于RGB颜色的空间图像，i=1.2.3，Pij表示第i个像素点的第j个颜色通道的颜色值。

（2）二阶颜色矩

二阶颜色矩采用的是二阶中心矩的平方根，反映图像颜色的分布范围。

𝟈表示第i个颜色通道的二阶颜色矩，Ei表示第i个颜色通道的一阶颜色矩

（3）三阶颜色矩

三阶颜色矩采用的是三阶中心距的立方根，反映图像颜色分布的对称性。

S表示第i个颜色通道的三阶颜色矩，Ei表示第i个颜色通道的一阶颜色矩

因为我们获取的是彩色图像，所以一共由R、G、B三个颜色通道组成。由上述颜色矩公式便可从每张图像中提取9个特征。

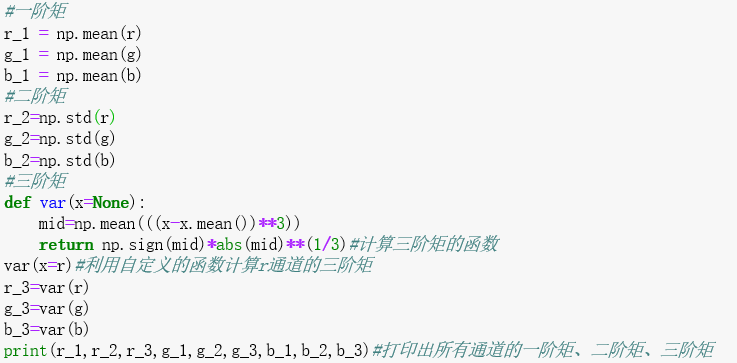


图4

在python中通过如上步骤便可获取一张图像的9个特征，并针对所有水色图像进行如上操作。（如图4）

具体代码详见：/code/任务1代码

获取的数据示例:（如下表1）

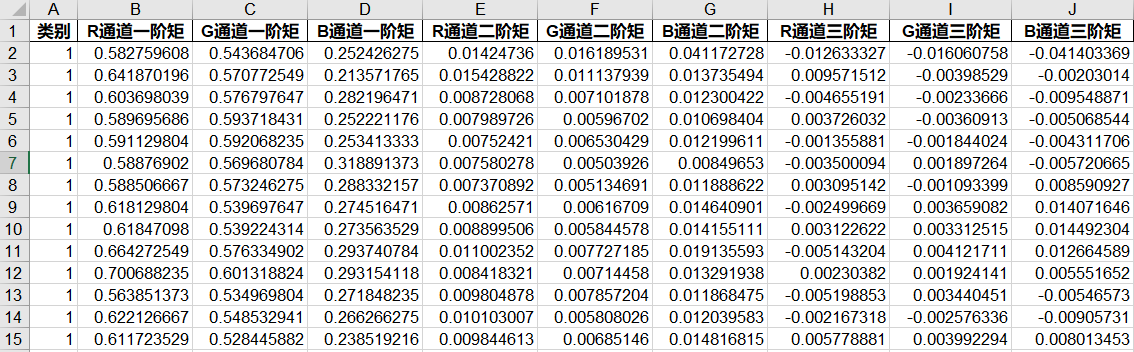


表1

数据详见：/result/tu

参数说明：'类别'表示根据不同水质类别中浮游植/动物的种类数量而划分的，即上文提到的专家样本中的5个类别；'R通道一阶矩','G通道一阶矩','B通道一阶矩','R通道二阶矩','G通道二阶矩','B通道二阶矩','R通道三阶矩','G通道三阶矩','B通道三阶矩'为图像特征提取中的9个特征，其中R、G、B三通道的一阶矩的取值范围皆为[0,1]，因为三阶矩是取立方根，所以其范围为[-1,1]。

1. 模型构建

在我们获得的特征提取数据中进行抽样，抽取其中的80%作为训练集，20%作为测试集，用于对水质评价进行检验。

首先输入决策树算法进行训练模型，得到的分类准确率为60%。

1.决策树

代码示例：（如图5）



图5

具体代码详见：/code/决策树

2.SVM

代码示例：（如图6）

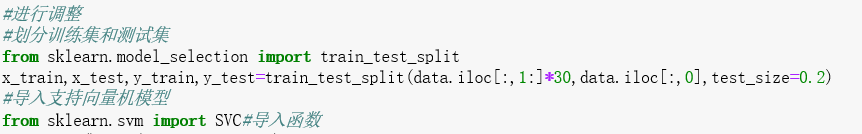


图6

具体代码详见：/code/SVM

所以，先直接输入SVM模型进行训练，因为其特征值范围在[0,1]，所以训练结束发现其区分度较小。结果展示：（如图7）

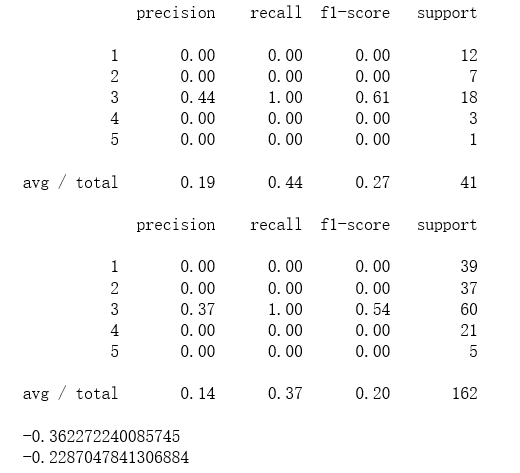


图7

因为直接输入模型结果不理想，所以猜想将所有的特征同意乘以一个适当的常数K，从而达到模型最优。最后根据测试集的准确率经过反复实验，选取K值为30。然后建立支持向量机模型

代码示例：（如图8）

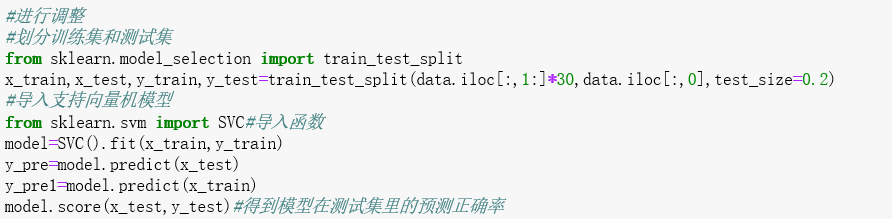


图8

具体代码详见：/code/ SVM

生成混淆矩阵：（如图9）



图9

具体代码详见：/code/ SVM

1. 结果分析

模型构建结束，通过训练集对样本回判，得出以下混淆矩阵（如表2）

得到分类准确率为97%，结果显示分类结果较好，可以应用该SVM模型于水质评价。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| 1 | 40 | 0 | 2 | 0 | 0 |
| 2 | 0 | 35 | 0 | 0 | 0 |
| 3 | 0 | 0 | 60 | 0 | 0 |
| 4 | 0 | 0 | 4 | 15 | 0 |
| 5 | 0 | 1 | 0 | 1 | 4 |

表2

表格详见：/resul/train （属于随机函数生成，重复实验将的到的结果会有所不同）

1. 水质评价

将所有的测试样本代入构建好的支持向量机模型中，得到结果从而预测水质类型。得出以下混淆矩阵，如下表3，分类的准确为93%。说明该支持向量机模型对于输入样本分类效果较好。所以该模型成立，可以适用导水质自动评价系统，实现水质的自动评价分类。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| 1 | 9 | 1 | 1 | 0 | 0 |
| 2 | 0 | 5 | 0 | 0 | 0 |
| 3 | 0 | 0 | 20 | 0 | 0 |
| 4 | 0 | 0 | 0 | 4 | 0 |
| 5 | 0 | 1 | 0 | 0 | 0 |

表3

表格详见：/resul/te （属于随机函数生成，重复实验将的到的结果会有所不同）