水色图像水质项目分析报告

# 项目介绍

## 项目背景

在水产养殖的水体中，由于水池内很多有机物的发酵反应，会让水体中生成许多的还原性中间产物，例如硫化氢、亚硝酸盐以及氨氮等等，这些产物对于很多养殖鱼类来说会产生较大的不良影响，同时在水体中存在很多水生生物，特别是一些微生物如细菌或寄生虫，这些相对来说容易造成危害的微生物绝大部分是条件致病菌，在水体稍微恶化的情况之下，很多鱼类产品由于自身抵抗能力降低会导致致病菌的毒性左右更大，从而对水产品形成更大的伤害。因此我们必须要控制好水体中的有毒物质。而对于养殖场来说，对水质的检测是很大一笔投入，这其中包括了仪器的购置，人员的分配，时间的消耗都是不小的投入。当然，对于一些很有经验的检测员来说分辨水质很容易，但也是很耗时耗力的，毕竟人的精力很有限。而对于许多小型的水产养殖散户来讲这笔投入甚至是入不敷出的**。**

## 项目目标

尝试使用类似与人肉眼观察的方法——图像分析处理，来对水质进行分辨，以降低这个方向的资金投入。具体的描述为利用给出的已经分类的数据，利用图像处理技术实现水质自动评价。

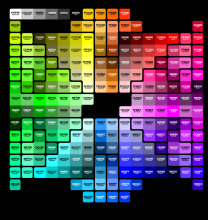
## 数据来源

来自某水产养殖所提供的203张照片，由有经验的渔民分为了5类水质。其中2\_15.jpg图片有损坏，不对其保留，进行删除操作。

# 操作原理

## RGB

RGB是从颜色发光的原理来设计定的，通俗点说它的颜色混合方式就好像有红、绿、蓝三盏灯，当它们的光相互叠合的时候，色彩相混，而亮度却等于三者亮度之总和，越混合亮度越高，即加法混合。

[](https://baike.baidu.com/pic/RGB/342517/0/ac2fc3c479211ad038db4933?fr=lemma&ct=single)

红、绿、蓝三盏灯的叠加情况，中心三色最亮的叠加区为白色，加法混合的特点：越叠加越明亮。

红、绿、蓝三个颜色通道每种色各分为256阶亮度，在0时“灯”最弱——是关掉的，而在255时“灯”最亮。当三色灰度数值相同时，产生不同灰度值的灰色调，即三色灰度都为0时，是最暗的黑色调；三色灰度都为255时，是最亮的白色调。

RGB 颜色称为加成色，因为您通过将 R、G 和 B 添加在一起（即所有光线反射回眼睛）可产生白色。加成色用于照明光、电视和计算机显示器。例如，显示器通过红色、绿色和蓝色荧光粉发射光线产生颜色。绝大多数可视光谱都可表示为红、绿、蓝 (RGB) 三色光在不同比例和强度上的混合。这些颜色若发生重叠，则产生青、洋红和黄。

## 颜色矩

颜色矩是一种简单有效的颜色特征表示方法，有一阶矩(均值,mean)、二阶矩(方差,variance)和三阶矩(斜度,skewness)等，由于颜色信息主要分布于低阶矩中，所以用一阶矩，二阶矩和三阶矩足以表达图像的颜色分布，颜色矩已证明可有效地表示图像中的颜色分布。

### 一阶颜色矩

### 二阶颜色矩

### 三阶颜色矩

## SVM

支持向量机（Support Vector Machine, SVM）是一类按监督学习（supervised learning）方式对数据进行二元分类的广义线性分类器（generalized linear classifier），其决策边界是对学习样本求解的最大边距超平面（maximum-margin hyperplane）。

SVM使用铰链损失函数（hinge loss）计算经验风险（empirical risk）并在求解系统中加入了正则化项以优化结构风险（structural risk），是一个具有稀疏性和稳健性的分类器。SVM可以通过核方法（kernel method）进行非线性分类，是常见的核学习（kernel learning）方法之一。

SVM被提出于1964年，在二十世纪90年代后得到快速发展并衍生出一系列改进和扩展算法，在识别、文本分类等模式识别（pattern recognition）问题中有得到应用。

## 图像特征

主要包括颜色特征、纹理特征、形状特征和空间关系特征等。与几何特征相比，颜色特征更为稳健，对物体的大小和方向不敏感，表现出较强的鲁棒性。由于本案例中的水色图像是均匀的，所以主要关注颜色特征。颜色特征是一种全局特征，描述图像或者图像对应景物的表明性质。一般颜色特征是基于像素点的特征，所以图像区域的像素点都有自己的贡献。

在利用图像颜色特征进行分析的研究中，实现方法上已经有啦很多研究成果，主要采用直方图法和颜色矩方法。

其中直方图法反映颜色分布，即哪些颜色及其出现概率，优点是适合描述难以自动分割的图像和不需要考虑物体空间位置的图像；缺点是无法描述颜色的局部分布及每种色彩所处的空间位置，也就是无法描述图像中的某一具体的对象和物体。

其中颜色矩法的数字基础为图像中的任何颜色分布均可以用它道德矩来表示。根据概率论，随机变量的概率分布可以由其各阶矩唯一描述和表示。一个图像的颜色分布完全可以看做是一种概率分布，图像就可以由各阶矩描述。颜色矩包含各个颜色的一阶矩、二阶矩和三阶矩，一个RGB图像有R、G、B三个通道，共9个分量。

直方图法产生的特征维数一般会大于颜色矩产生的，为了避免过多变量干扰分类结果，使用颜色矩法。

# 处理过程

## 图像切割

**图像切割代码**

def Division(x):

M, N = x.size  
 box = [M / 2 - 50, N / 2 - 50, M / 2 + 50, N / 2 + 50]  
 p = x.crop(box)  
 return p

在项目实践的具体过程中，选用了图像分割的办法来提取图像特征，由二分图像看出，盛水的容器干扰了特征分析，所需提取有意义的图像进一步缩小，最终定位到了100\*100的像素大小。

## 特征提取

提取切割后的图像颜色一、二、三阶矩，作为颜色特征，共计9个字段，并加上水质图像的种类编号，得到数据集。如表1-1所示。

**表1-1**



# 数据建模

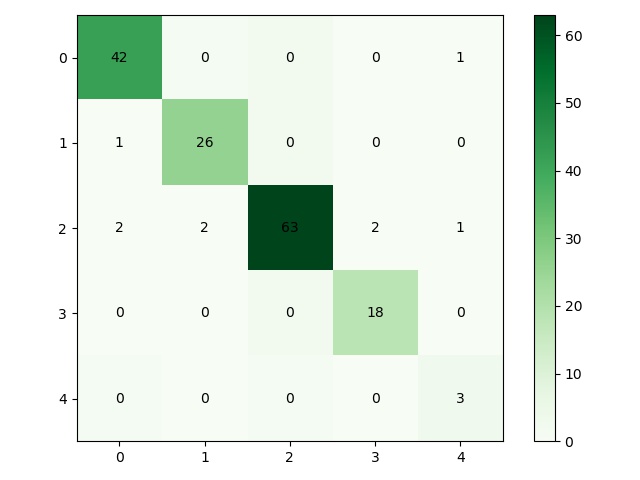
使用传统机器学习中8:2的比例，划分训练集与测试集，使用支持向量机进行分类。在数据计算中，因为特征数据在[-1,1]之间，所以可以使用放大模型数据的方法来提高模型的精度。

**建模代码**

from sklearn.svm import SVC

model = SVC().fit(x\_train, y\_train)

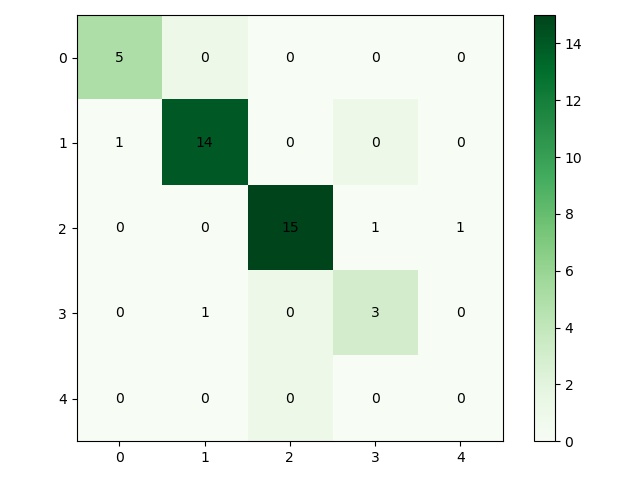
在使用训练集建模之后，将模型保存，并调用模型，输出混淆矩阵，将混淆矩阵可视化：



计算分类准确率为94.4%，因此判断，模型可以用来进行水质评价。

# 结论-水质评价

将训练集导入模型，并可视化混淆矩阵：



计算出分类准确率为90.5%，说明此模型的分类水平良好，可以用来进行水质评价。

## 附录

——《水质对水产养殖技术的影响》——赵东华 2017-02-02

——百度百科

——「王大阳\_」

原文链接：https://blog.csdn.net/weixin\_43746433/java/article/details/100246546