



大数据，成就未来



# 基于水色图像的水质分析

2020/3/3

# 目录 CONTENTS

- 01 背景介绍
- 02 数据挖掘目标
- 03 分析方法与过程
- 04 建模及代码展示

01 污水处理

02 环境监测

03 水产养殖

在渔业生产中，从业者可以通过观察水色变化调控水质，以达到维持养殖生态系统中浮游植物、浮游动物，微生物类、等动态平衡的目的。但由于这些多是通过肉眼和经验观察进行判断，存在主观性引起的观察性偏差，使观察结果的可比性、可重复性降低，且受工作时间限制，不易推广应用。

# 数据挖掘目标

通过多年来对水质环境数据的采集和处理，建立环境数据库，得出水产养殖较为完善的生长发育参数和健康养殖量化指标。根据这些量化指标和水产养殖专业知识，基于环境数据库建立起水产健康养殖专家决策系统，对实现水产健康养殖、智能控制和计算机管理具有重要的意义。我们的目标是在水质监测方面，通过计算机视觉处理技术，以专家经验为基础，对池塘水色进行优劣分级，实现对池塘水色的准确快速判别。



# 数据挖掘目标





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







颜色矩：一幅图像的色彩分布可以认为是一种概率分布，图像中任何颜色的分布均可以用它的矩来表示。颜色矩包含各个颜色通道的一阶矩、二阶矩、三阶矩，对于一幅RGB颜色空间图像，具有R、G和B三个颜色通道，所以共具有9个分量。

一阶矩：反应图像的整体明暗度

$$E_i = \frac{1}{N} \sum_{j=1}^N p_{ij}$$

二阶矩：反应图像颜色的分布范围

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^2}$$

三阶矩：反映图像颜色分布的对称性

$$S_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^3}$$





# 分析方法与过程

## 水色图像特征与相应水色类别的部分数据

类别	序号	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
1	1	0.5828229	0.5437737	0.2528287	0.014192	0.0161439	0.04107525	-0.0126431	-0.0160904	-0.0415362
2	1	0.555172	0.5710693	0.3162981	0.0075897	0.0049697	0.00879593	0.00096125	-0.0020346	0.00652788
2	1	0.5555761	0.5792227	0.3312147	0.0065946	0.0050712	0.01071972	-0.0016704	-0.0024116	0.00675625
3	1	0.5366698	0.5205499	0.1913957	0.0075394	0.0056239	0.01267093	-0.0027357	0.0031078	-0.0024782
3	1	0.5257916	0.5223429	0.2570136	0.0078235	0.0052186	0.01144844	0.00488671	0.0010884	-0.0083357
3	2	0.5995602	0.5396726	0.1572549	0.0092801	0.0072411	0.01036641	0.00638896	-0.0038906	-0.0043773
1	2	0.5855547	0.565562	0.2559245	0.007922	0.005304	0.01058632	-0.0040833	0.0021329	0.00532096
1	2	0.5820929	0.5714734	0.2803032	0.0068992	0.0049466	0.01011564	-0.0040232	-0.0023722	0.0056939
2	2	0.5527147	0.552739	0.3271475	0.0104413	0.0083428	0.01466284	-0.0063701	-0.0041057	0.00960346
2	2	0.491916	0.5463674	0.425871	0.0103435	0.0082932	0.01226	0.00928492	0.009663	0.01154853
5	3	0.2115671	0.3355373	0.1119694	0.012056	0.0132956	0.00837975	0.00730481	0.0075034	0.00365034
4	3	0.4377533	0.4312157	0.1896054	0.0114436	0.0091593	0.01149126	-0.0033359	-0.0050138	0.0047221
4	3	0.4618601	0.4627128	0.240098	0.0081545	0.006365	0.01198156	-0.0048741	-0.0023075	-0.0071941
5	3	0.4089634	0.486953	0.1781129	0.0126622	0.0097519	0.01449667	-0.0067208	0.0021676	0.00999232
5	3	0.4411736	0.5151706	0.1954495	0.0122722	0.008599	0.01574276	-0.0069344	-0.0049038	0.00898545



序号	变量名称	变量描述	取值范围
1	R通道一阶矩	水样图像在R颜色通道的一阶矩	0 ~ 1
2	G通道一阶矩	水样图像在G颜色通道的一阶矩	0 ~ 1
3	B通道一阶矩	水样图像在B颜色通道的一阶矩	0 ~ 1
4	R通道二阶矩	水样图像在R颜色通道的二阶矩	0 ~ 1
5	G通道二阶矩	水样图像在G颜色通道的二阶矩	0 ~ 1
6	B通道二阶矩	水样图像在B颜色通道的二阶矩	0 ~ 1
7	R通道三阶矩	水样图像在R颜色通道的三阶矩	-1 ~ 1
8	G通道三阶矩	水样图像在G颜色通道的三阶矩	-1 ~ 1
9	B通道三阶矩	水样图像在B颜色通道的三阶矩	-1 ~ 1
10	水质类别	不同类别表示水中浮游植物的种类多少	1, 2, 3, 4, 5





综合输入变量维度较高，样本量较少，再结合支持向量机具备较高泛化能力的特点，选定支持向量机作为建模算法。对数据采取随机抽取80%作为训练样本，20%作为测试样本的交叉验证方法，以防止模型欠拟合。



# 建模及代码展示

## 结果分析

训练集混淆矩阵

分类准确率：96.91%

预测值 实际值	1	2	3	4	5
1	41	1	1	0	0
2	0	34	0	0	0
3	0	0	59	0	0
4	0	0	1	20	0
5	0	1	0	1	3

测试集混淆矩阵

分类准确率：95.12%

预测值 实际值	1	2	3	4	5
1	7	0	1	0	0
2	0	10	0	0	0
3	0	0	19	0	0
4	0	0	0	3	0
5	0	0	0	1	0



### What's Next?

1. 完成特征构建并用支持向量机(SVM)进行分类
2. 评估模型性能并可视化结果
3. 引入神经网络模型或者卷积神经网络模型进行分类并对比模型性能

Hint: Python Scikit-Learn

# 基于水色图像的水质分析



**THANKS FOR WATCHING**