基于多核学习的浮游生物图像分类研究

答辩人 王 如 晨
导师 姬光荣教授、郑海永副教授
专业 信号与信息处理

中国海洋大学
信息科学与工程学院

2017 年 5 月

内容提要

下一节内容

课题背景

浮游生物包括浮游植物和浮游动物两大类。

课题背景

传统浮游生物丰富度监测

网采、泵采和瓶采 $\Longrightarrow$ 人工分类计数

工作量大、速度慢

需要丰富的专业知识

课题背景

传统浮游生物丰富度监测

网采、泵采和瓶采 $\Longrightarrow$ 人工分类计数

工作量大、速度慢

需要丰富的专业知识

浮游生物图像自动识别系统

浮游生物图像自动识别系统 { 浮游生物图像采集
浮游生物图像分类

国内外研究现状

浮游生物图像分类

年份	研究者	分类方法	实验结果
1996	Culverhouse	利用细胞的形状纹理特征进行分析	3 种浮游植物
		采用人工神经网络进行甲藻分类	72%
1998	Tang 等	不变矩和傅里叶描述子描述形状纹理	6 种浮游动物
		用改进的学习矢量量化网络进行分类	95%
2006	Hu 等	灰度共生矩阵描述灰度特征	7 种浮游生物
		用支持向量机训练分类器	72%
2007	Sosik 等	大小、形状、几何等特征	22 种浮游植物
		采用支持向量机进行分类	88%
2012	Mosleh 等	提取藻类图像的形状纹理特征	5 种浮游植物
		采用人工神经网络进行分类	93%

存在的问题

- 采用特征种类单一，不能全面描述浮游生物的形态特征。
- 简单的特征串联，不能充分利用每种特征中包含的信息。
- 适用的浮游生物种类较少，适用范围窄。

存在的问题

- 采用特征种类单一，不能全面描述浮游生物的形态特征。
- 简单的特征串联，不能充分利用每种特征中包含的信息。
- 适用的浮游生物种类较少，适用范围窄。

研究思路

- 分析浮游生物的形态特征从多角度进行特征描述。
- 采用多核学习依据每种特征在分类过程中的贡献度进行特征融合。
- 提高分类系统分类准确率和泛化能力。

下一节内容

下一节内容

数据集构建

1. 伍兹霍尔海洋研究所（WHOI）[1]用 FlowCytobot 采集的浮游生物图像

22 类	共 6600 张

2. ZooScan 系统[2]采集的浮游生物图像

20 类	共 3771 张

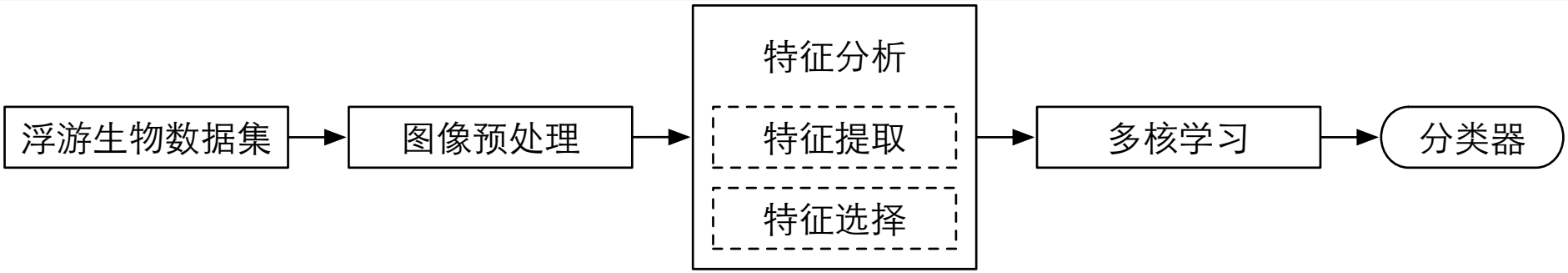
3. Kaggle 竞赛[3]中使用的浮游生物图像

38 类	共 28748 张

[1] http://aslo.org/lomethods/free/2007/0204a1.html
[2] http://www.zooscan.obs-vlfr.fr//rubrique.php3?id_rubrique=33?lang=en
[3] https://www.kaggle.com/c/datasciencebowl/data

下一节内容

基于多核学习的浮游生物分类系统



①图像预处理

1. 图像分割[4] 对未分割的图像进行分割

相位一致性 → 边缘检测 → 形态学处理（闭运算、膨胀、细化） → 提取最简边缘

2. 去除悬浮颗粒等杂质 开运算，去除小连通区域

[4]Sosik H M, Olson R J. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. Limnol. Oceanogr. Methods, 2007, 5(204):e216.

②浮游生物特征分析

特征提取

① 几何灰度特征

周长、面积、体态比、灰度平均值、灰度标准差 $\cdots$ 共 43 个。

② 粒子测度

用来计算二值图像中目标区域的大小分布情况。

$$F_G(\lambda) = 1 - \frac{v(\psi_\lambda(G))}{v(G)} \qquad \psi_\lambda(G) = G \circ \lambda T$$

②浮游生物特征分析

③ 纹理特征

- 变差函数
- Gabor 滤波器
- 局部二值模式
- 二元梯度轮廓

④ 局部特征

- 内距离形状上下文
- 方向梯度直方图
- 尺度不变特征变换

②浮游生物特征分析

③ 纹理特征

- 变差函数
- Gabor 滤波器
- 局部二值模式
- 二元梯度轮廓

④ 局部特征

- 内距离形状上下文
- 方向梯度直方图
- 尺度不变特征变换

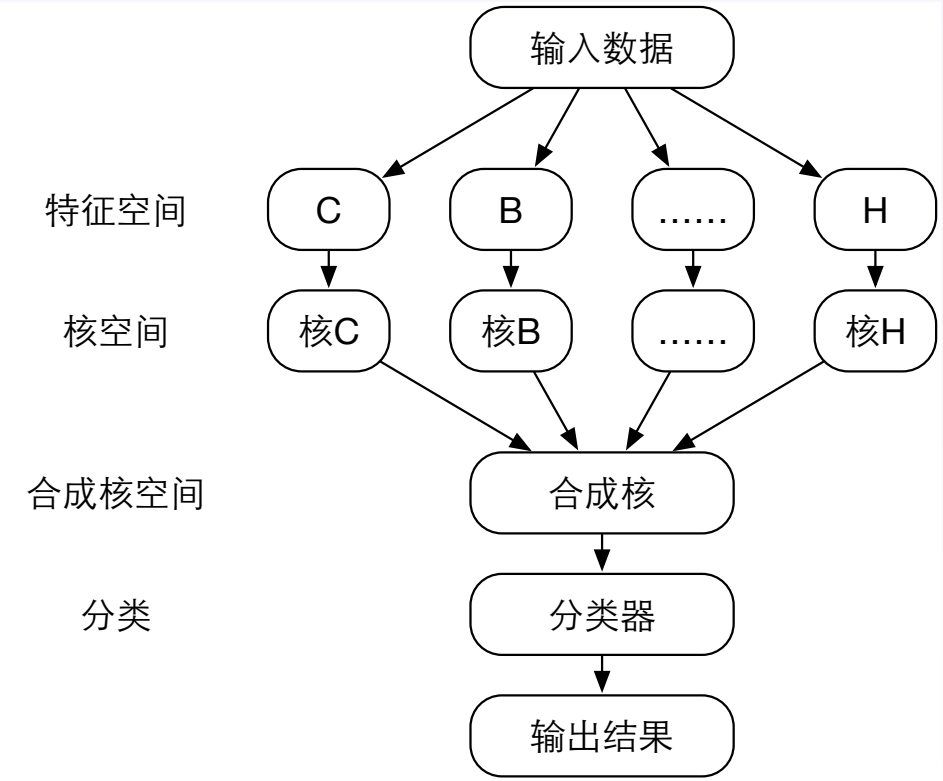
特征选择

从特征集合中选取有用的特征子集，去除冗余特征，降低特征维数。

③多核学习

机器学习中常用的支持向量机是单核学习算法。

多核学习用多个核函数的组合代替单个核函数。



下一节内容

混淆矩阵

一种特定的矩阵用来呈现算法性能的可视化工具。

		预测结果	
		正样本	负样本
实际结果	正样本	a	b
	负样本	c	d

- 真阳性率（True positive rate，也称召回率 Recall）：

$$TPR = \frac{a}{a+b}$$

- 阳性预测值（Positive predictive value，也称为命中率 Precision）：

$$Precision = \frac{a}{a+c}$$

F-Measure

F-Measure是一种综合评价指标。

当 Recall 和 Precision 出现矛盾时，就可以采用该方法进行评价。

$$F = \frac{(\alpha^2 + 1)P * R}{\alpha^2(P + R)}$$

当 $\alpha = 1$ 就得到 F1-Measure：

$$F1 = \frac{2 * PR}{P + R}$$

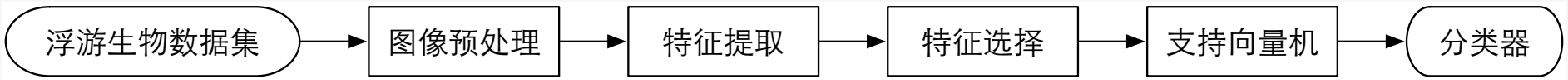
下一节内容

对比实验

1. 基准实验
2. 特征对比实验
3. 基于多核学习的浮游生物图像分类实验

①基准实验

根据 Sosik 等人在 2007 年提出的浮游植物自动分类方法[5]和 ZooScan 系统[6]设计浮游生物分类基准系统。



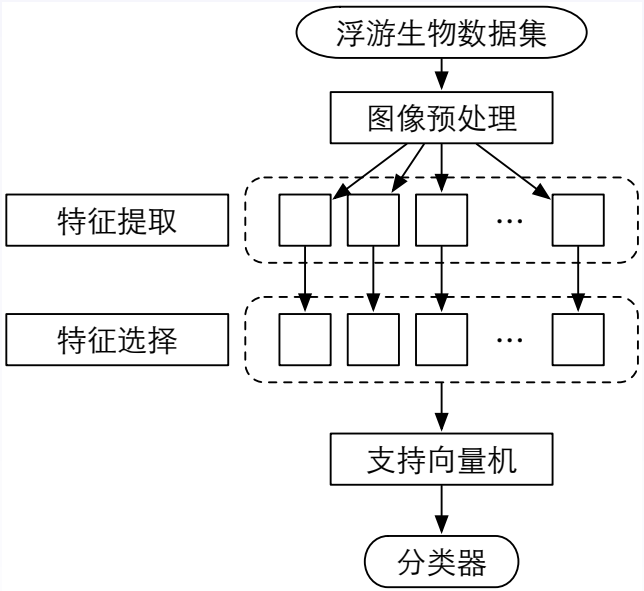
基准实验结果

	WHOI 数据集	ZooScan 数据集	Kaggle 数据集
F-Measure	0.8832 (0.8792)	0.8212 (0.7947)	0.7690

[5]Sosik H M, Olson R J. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. Limnol. Oceanogr. Methods, 2007, 5(204):e216.

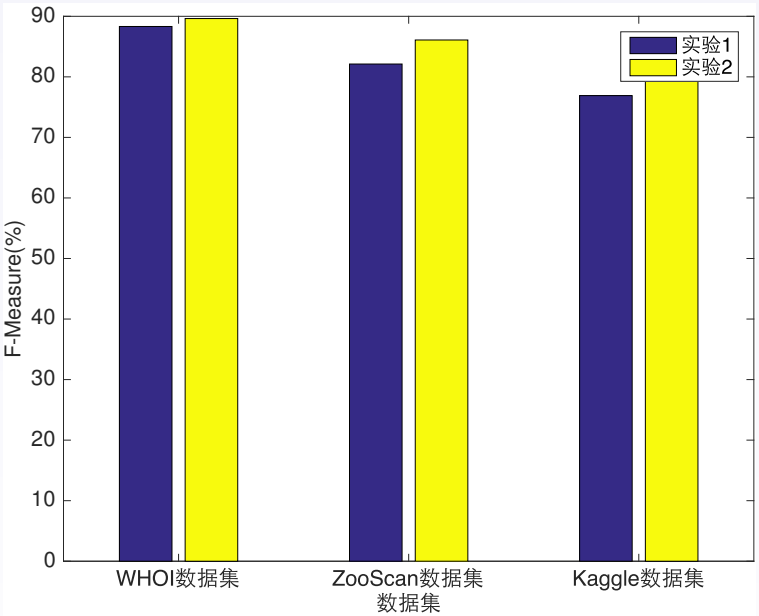
[6]Gorsky G, Ohman M D, Picheral M, et al. Digital zooplankton image analysis using the zooscan integrated system. Journal of Plankton Research, 2010, 32(3):285–303.

②特征对比实验



特征对比实验结果

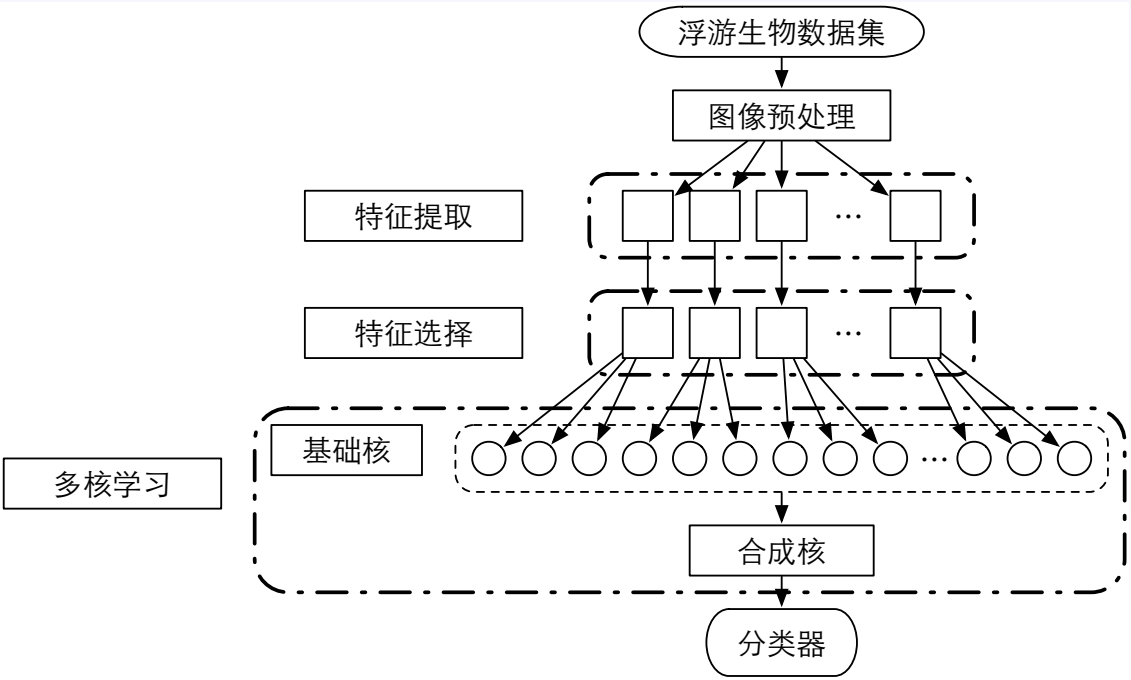
数据集	C	高斯核函数 F-Measure	多项式核函数 F-Measure	线性核函数 F-Measure
WHOI 数据集	1	0.8428	0.8901	0.8652
	10	0.8897	0.8949	0.8817
	100	0.8963	0.8848	0.8637
ZooScan 数据集	1	0.8174	0.8322	0.8087
	10	0.8609	0.8446	0.8475
	100	0.8562	0.8351	0.8202
Kaggle 数据集	1	0.7910	0.7891	0.7307
	10	0.8304	0.8131	0.7890
	100	0.8260	0.7964	0.7802



对比实验①②

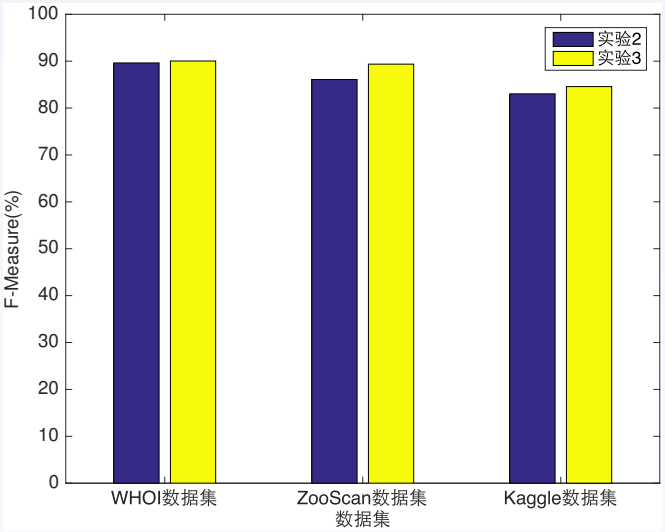
数据集	F-Measure 提高量
WHOI 数据集	0.0131
ZooScan 数据集	0.0397
Kaggle 数据集	0.0641

③基于多核学习的浮游生物图像分类实验



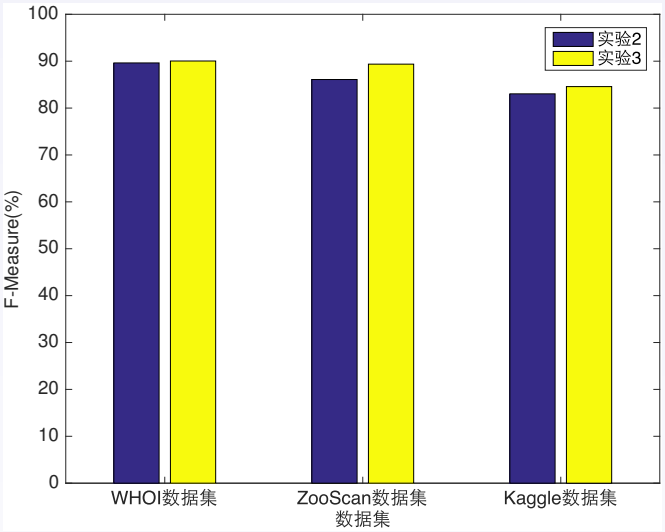
实验结果

数据集	C	F-Measure
WHOI 数据集	1	0.8973
	10	0.8992
	100	0.9004
ZooScan 数据集	1	0.8699
	10	0.8937
	100	0.8924
Kaggle 数据集	1	0.8205
	10	0.8458
	100	0.8428



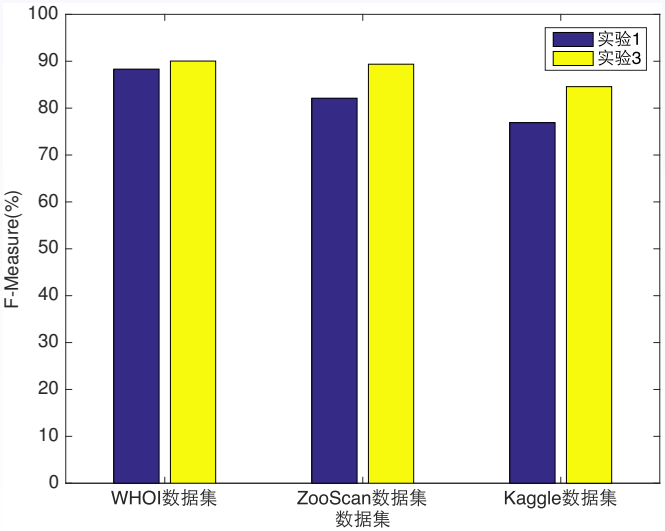
对比实验②③

数据集	F-Measure 提高量
WHOI 数据集	0.0041
ZooScan 数据集	0.0328
Kaggle 数据集	0.0154



对比实验②③

数据集	F-Measure 提高量
WHOI 数据集	0.0041
ZooScan 数据集	0.0328
Kaggle 数据集	0.0154



对比实验①③

数据集	F-Measure 提高量
WHOI 数据集	0.0172
ZooScan 数据集	0.0725
Kaggle 数据集	0.0768

下一节内容

总结与展望

总结

- 分析浮游生物形态特征，从多角度对浮游生物进行描述。
- 提出基于多核学习的浮游生物自动分类系统。
- 收集构建不同的浮游生物数据集，设计对比实验评价分类性能。

展望

- 提高分类系统的计算效率。
- 针对数据集不均衡问题进行进一步研究。

谢谢！

Ruchen Wang
Ocean University of China
2017.05