

学校编码: 10384

密级_____

学号: 31120101151290

廈門大學

硕士学位论文

大黄鱼疾病诊断与预警系统初步构建

Preliminary establishment of large yellow croak disease

diagnosis and pre-warning system

蔡晓鹏

指导教师姓名: 苏永全 教授

专业名称: 海洋生物技术

论文提交日期: 2013年5月

论文答辩时间: 2013年6月

2013年5月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

()1.经厦门大学保密委员会审查核定的保密学位论文，于
年 月 日解密，解密后适用上述授权。

()2.不保密，适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

声明人(签名)：

年 月 日

目录	
摘要	I
Abstract	III
缩略语中英文对照表	1
第一章 绪论	1
1.1 大黄鱼养殖现状	1
1.2 刺激隐核虫病	2
1.2.1 刺激隐核虫	2
1.2.2 刺激隐核虫病的防治方法	3
1.3 疾病防治-诊断与预警	4
1.3.1 疾病诊断	4
1.3.2 疾病预警	5
1.4 分类方法	6
1.5 随机森林算法	7
1.5.1 随机森林算法的定义	7
1.5.2 随机森林算法	7
1.5.3 随机森林算法的优点	8
1.5.4 随机森林算法的应用	9
1.6 目的意义	10
第二章 大黄鱼流行病学调研报告	11
2.1 调研目的	11
2.2 调研任务	11
2.3 调研成果总结	11
2.4 调研附录：大黄鱼常见病害	13
2.4.1 寄生虫病	13
2.4.2 细菌性疾病	15
2.4.3 其他原因导致的疾病	17

第三章 大黄鱼病诊断系统	20
3.1 引言.....	20
3.1.1 构建系统目的.....	20
3.1.2 项目背景.....	20
3.2 软件概述.....	20
3.2.1 目标.....	20
3.2.2 功能.....	21
3.2.3 性能.....	21
3.3 运行环境.....	21
3.3.1 硬件.....	21
3.3.2 支持软件.....	21
3.3.3 开发工具.....	21
3.4 终端用户使用说明.....	21
3.4.1 系统图标.....	22
3.4.2 系统欢迎界面.....	22
3.4.3 系统主界面.....	23
3.4.4 部分主要常用界面的详细介绍.....	24
3.5 意义及创新点.....	27
第四章 基于多元线性分析的大黄鱼刺激隐核虫病预警模型	29
4.1 材料.....	29
4.1.1 采样地点.....	29
4.1.2 数据采集.....	30
4.2 方法.....	31
4.2.1 各参数的月变化趋势分析.....	31
4.2.2 各影响因子与疾病情况的相关分析.....	31
4.2.3 各参数间单因素方差分析.....	31
4.2.4 各参数对疾病等级的判别分析.....	31
4.3 结果分析.....	32

4.3.1 各影响因子的月变化趋势分析.....	32
4.3.2 各影响因子与疾病情况的相关分析.....	32
4.3.3 各参数间单因素方差分析.....	35
4.3.4 各参数对疾病等级的判别分析.....	38
4.4 讨论.....	43
第五章 基于随机森林算法的大黄鱼刺激隐核虫病预警模型	45
5.1 材料.....	45
5.1.1 采样地点.....	45
5.1.2 数据采集.....	45
5.2 方法.....	45
5.2.1 数据预处理.....	46
5.2.2 全因子模型构建.....	46
5.2.3 降维模型构建.....	47
5.3 结果分析.....	48
5.3.1 <i>ntree</i> 取值的确定	48
5.3.2 全因子模型构建.....	50
5.3.3 降维模型构建.....	51
5.4 讨论.....	55
第六章 基于随机森林算法的大黄鱼刺激隐核虫病预警系统	58
6.1 引言.....	58
6.2 系统概述.....	58
6.2.1 目标.....	58
6.2.2 功能.....	59
6.2.3 性能.....	59
6.3 运行环境.....	60
6.3.1 硬件设备.....	60
6.3.2 支持软件.....	61
6.3.3 数据结构.....	61

6.4 使用说明.....	62
6.4.1 安装与初始化.....	62
6.4.2 具体功能详解.....	62
第七章 总结.....	69
参考文献.....	71
在学期间发表的论文及成果.....	76
致谢.....	78

厦门大学博硕士学位论文摘要

Contents

Chinese abstract	I
English abstract	III
List of abbreviations	VI
Chapter 1 Introduction	1
1.1 Review in large yellow croak aquaculture	1
1.2 <i>Cryptocaryon irritans</i> infection	2
1.2.1 <i>Cryptocaryon irritans</i>	2
1.2.2 Control and treatment for <i>Cryptocaryon irritans</i> infection	3
1.3 Disease control-diagnosis, pre-warning and prediction	4
1.3.1 Disease diagnosis	4
1.3.2 Disease pre-warning and prediction-water quality prediction	5
1.4 Classification	6
1.5 Random forest algorithm	7
1.5.1 Definition of random forest	7
1.5.2 Random forest algorithm	7
1.5.3 Advantage of random forest	8
1.5.4 Application of random forest	9
1.6 Research objectives and significance	10
Chapter 2 Epidemiological survey for large yellow croak diseases	11
2.1 Objectives	11
2.2 Tasks	11
2.3 Conclusions	11
2.4 appendix: common large yellow croak disease	13
2.4.1 Parasitosis	13
2.4.2 Bacteriosis	15
2.4.3 Diseases caused by other reasons	17
Chapter 3 Large yellow croak disease diagnosis system	20

3.1 Introduction	20
3.1.1 Objectives	20
3.1.2 Backgrounds	20
3.2 Software Overview	20
3.2.1 Goals	20
3.2.2 Fuction	21
3.2.3 Performance	21
3.3 Operating environment.....	21
3.3.1 Hardware.....	21
3.3.2 Support software	21
3.3.3 Development tools	21
3.4 Instructions	21
3.4.1 Icons.....	22
3.4.2 Welcome screen	22
3.4.3 The main interface	23
3.4.4 Part of the common interface illustrated in detail.....	24
3.5 Significance and innovations	27
Chapter 4 Pre-warning models for large yellow croak marine white spot disease based on multiple linear analysis.....	29
4.1 Materials.....	29
4.1.1 Sampling sites	29
4.1.2 Date sampling	30
4.2 Methods.....	31
4.2.1 Monthly variation analysis for each parameter	31
4.2.2 Correlation analysis	31
4.2.3 Variance analysis	31
4.2.4 Discriminant analysis.....	31
4.3 Results	32

4.3.1 Monthly variation analysis for each parameter	32
4.3.2 Correlation analysis	32
4.3.3 Variance analysis.....	35
4.3.4 Discriminant analysis.....	38
4.4 Discussion	43
Chapter 5 Pre-warning models for large yellow croak marine white spot disease based on random forest algorithm	45
5.1 Materials	45
5.1.1 Sampling sites	45
5.1.2 Data sampling	45
5.2 Methods.....	45
5.2.1 Data preprocessing.....	46
5.2.2 Full-factor model construction.....	46
5.2.3 Reduced-dimension modle construction.....	47
5.3 Results	48
5.3.1 Determination of <i>n_{tree}</i>	48
5.3.2 Full-factor model construction.....	50
5.3.3 Reduced-dimension modle construction.....	51
5.4 Discussion	55
Chapter 6 Pre-warning system for large yellow croak marine white spot disease based on random forest algorithm	58
6.1 Introduction	58
6.2 Overview of the system.....	58
6.2.1 Goals	58
6.2.2 Function	59
6.2.3 Performance	59
6.3 Operating environment.....	60
6.3.1 Hardware.....	60

6.3.2 Support software	61
6.3.3 Data structure	61
6.4 Instructions	62
6.4.1 Installation and initialization.....	62
6.4.2 Specific function	62
Chapter 7 Conclusion	69
References	71
Papers published	76
Acknowledgments	78

厦门大学博硕士学位论文摘要库

摘要

大黄鱼 (*Larimichthys crocea*, Richardson, 1846)是中国近海特有的主要经济鱼类和当前最主要海水养殖鱼类之一,也是中国八大优势出口水产品之一。目前全国年育苗量近 20 亿尾,养殖产量超过 7 万吨,直接产值超过 30 亿元。但是,近年来随着鱼用抗生素等药物的滥用,养殖密度的提高,水域生态环境日益恶化,各种大黄鱼疾病频繁爆发。病害问题给大黄鱼养殖产业带来了巨大的经济损失。因此,如何有效的防治大黄鱼疾病的爆发,对该产业可持续健康发展具有重大的意义。

目前大部分大黄鱼养殖户对大黄鱼疾病的认识程度不高,为提高广大大黄鱼养殖户对大黄鱼疾病的认识水平,使其在所养殖大黄鱼得病时能够较快、较准确的判断所患病类型,并及时采取治疗措施,本研究经过长期的调研,最终研制开发出:大黄鱼病诊断系统。通过本系统用户可以学习大黄鱼各种疾病的基本知识以及治疗方案,同时只需提供大黄鱼患病的身体各个部位的症状,本系统也能智能判别对应的疾病类型,并给出其有效的治疗方案。本系统使用操作简单,用户经过简单培训即可使用。

疾病诊断和治疗只是在疾病爆发后的补救措施,并未能从根本上避免疾病的爆发,而疾病的预警预报是解决该问题的潜在的有效手段。水质被认为是影响人或动物的健康的最主要因素之一,也是影响水产养殖业可持续健康发展的关键因素,本研究通过传统分类的多元线性方法以及人工智能分类的随机森林算法,分析 11 个水质参数和大黄鱼最严重的刺激隐核虫病疾病等级的关系,并建立水质对该病的预警模型。剔除影响不显著的 5 个水质因子后,分别建立了 1-6 维模型,按维度上升其准确率从 50.5% 升至 68.8%,综合考虑不同维度模型的性能,最终建立了水温、溶解氧、化学耗氧量对该病的 3 维预警预报模型,其表达式为:

$$Y_1 = -85.123 + 3.136 \times X_1 + 14.414 \times X_2 - 5.370 \times X_3 \quad (1)$$

$$Y_2 = -90.051 + 3.395 \times X_1 + 14.178 \times X_2 - 4.430 \times X_3 \quad (2)$$

$$Y_3 = -88.515 + 3.367 \times X_1 + 13.995 \times X_2 - 4.046 \times X_3 \quad (3)$$

$$Y_4 = -73.768 + 3.028 \times X_1 + 12.876 \times X_2 - 4.484 \times X_3 \quad (4)$$

其中 Y_1 、 Y_2 、 Y_3 、 Y_4 分别为疾病等级 1、2、3、4 即未发病、少量发病、中等程度发病、大面积发病；

X_1 - X_3 分别为水温 ($^{\circ}\text{C}$)、溶解氧 (mg/L)、化学耗氧量 (mg/L)；

实际应用中将上述 3 个参数数值带入上述 4 个公式分别计算, 分别得出 Y_1 、 Y_2 、 Y_3 、 Y_4 值, 并比较大小, 其中最大的值所对应的疾病等级则为本次判别预测的疾病等级, 本模型的综合预测准确率为 66.4%, 预测性能较为一般。

随机森林算法则按照参数的影响权重从高到低排列, 分别建立了 2-11 维的模型, 按维度的上升, 其预测准确率从 81.0% 升至 90.6%, 综合考虑不同维度的性能, 最终建立了水温、溶解氧、化学耗氧量的 3 维预警预报模型, 其综合预测准确率为 83.6%, 预测性能较好。

值得一提的是, 两个模型所分析的对该病影响最大的 3 个水质参数均为水温、溶解氧、化学耗氧量, 进一步说明这 3 个参数对该病的影响的重要性。对比两个模型, 可发现随机森林模型的预测性能与多元线性模型的性能较大幅度的提升 (预测准确率约提高 17%), 证明了本研究所建立随机森林模型是预测大黄鱼刺激隐核虫病的理想预测模型。但是与多元线性模型对比, 随机森林模型没有可视的数学公式, 其预测需要较强的专业知识, 推广性较差。为此, 本研究建立了大黄鱼刺激隐核虫病预警预报系统, 将随机森林算法植入其中, 用户可通过简单的培训就可掌握。通过该系统用户可单个或批量将自己所测水质参数输入系统, 系统会自动建模, 输出所预测的疾病等级, 用户可通过预测结果, 采取相应预防对策。再者, 随机森林是机器学习的一种方法, 有较强的自我学习能力, 该系统需要用户的真实反馈所预测数据的实际值, 使得系统数据库将越来越完善, 最终系统将会用更好的预测性能回馈用户。

关键词: 大黄鱼; 疾病诊断; 预警; 多元线性模型; 随机森林算法

Abstract

Large yellow croak (*Larimichthys crocea*, Richardson, 1846), one of the major mariculture species in southeastern coastal area of China and one of China's top 8 leading exports of aquatic products, has exceeded annual output of 70,000 metric tons with total economic value of some RMB 3.0 billion. In recent years, along with the abuse of antibiotics and other drugs and the increase in the number and density of farmed species, water quality and the ecological environment had gradually deteriorated, as well as frequent outbreak of diseases. Therefore, prevention and prediction of diseases has great significant importance to the sustainable development of large yellow croak farming industry.

At present, the majority of the farmers are lacking of scientific knowledge about disease. In order to develop their understandings, help them with accurate diagnose and make timely treatment, we finally developed the "Large Yellow Croak Disease Diagnosis System" through research of long period of time. Users could manipulate the software and learn basic disease knowledge as well as understand related treatments by getting systematic training. By providing samples of symptom, the system is also able to identify the disease.

However, disease diagnosis and treatment are just remedial measures which could not prevent the disease and reduce economic losses. Disease pre-warning and prediction becomes a potential method to solve this problem. Water quality is considered to be the most important factor affecting the health of aquatic animals, as well the sustainable key element of aquaculture industry. In this study, we introduced both the traditional classification method-multiple linear (ML) analysis and the artificial intelligence classification method-random forest (RF) algorithm to analyze and predict the effect between 11 water quality parameters on the most common and severe large yellow croak disease-marine white spot disease. In ML analysis, 5

insignificant parameters were wiped out, then 6 (1 to 6-dimension) disease prediction models were respectively built, with which the accuracy increased from 50.5% to 68.8% along with the increasing dimensions. After comprehensive evaluation for the performance of each model, we finally chose the 3-dimension model: the water temperature (WT), the dissolve oxygen (DO), the chemical oxygen demand (COD), as parameters to enhance our disease prediction system. The formulas were expresses al followed:

$$Y_1 = -85.123 + 3.136 \times X_1 + 14.414 \times X_2 - 5.370 \times X_3 \quad (1)$$

$$Y_2 = -90.051 + 3.395 \times X_1 + 14.178 \times X_2 - 4.430 \times X_3 \quad (2)$$

$$Y_3 = -88.515 + 3.367 \times X_1 + 13.995 \times X_2 - 4.046 \times X_3 \quad (3)$$

$$Y_4 = -73.768 + 3.028 \times X_1 + 12.876 \times X_2 - 4.484 \times X_3 \quad (4)$$

Y_1 to Y_4 indicated severity rate 1 to 4, which represented healthy conditions, small-scale disease, medium-scale disease and large-scale disease, respectively.

X_1 to X_3 represented WT ($^{\circ}$ C), DO (mg/L) and COD (mg/L), respectively.

In practice, the above 3 parameter measured could be put into the 4 formula to calculate Y_1 , Y_2 , Y_3 , Y_4 , the maximum value among Y_1 to Y_4 represented the severity rate as predicted. The total prediction accuracy was 66.4%, which indicated a general performance of this model.

In RF analysis, effect weight to the disease severity rate of each parameter was calculated, where the parameters was selected according to the weight (from high to low). We built 10 (2 to 11-dimension) models which respectively represent the increasing in number of dimensions, and the prediction accuracy had increased from 81.0% to 90.6% After comprehensive evaluation for the performance of each model, we finally chose the 3-dimension prediction model: the WT, DO, COD to disease severity rate model. The total prediction accuracy was 83.6%, which showed that the performance of this model was in good condition.

To speak of, the top 3 effect weight parameters the ML and RF model analyzed were the same, which furthermore demonstrated the importance of these 3 parameters.

Comparing ML with RF model, it was found that the RF performed much better than the ML model (the prediction accuracy was higher up to about 17%), and concluded the RF model built in this study was ideal for large yellow croak white spot disease forecast. Whereas there is no visual mathematical formula in RF model, therefore, it needs intensive specialized knowledge which result in promoting difficulties in farmers. To this end, we combined software techniques to build a related prediction system into which the RF model was planted. The users could input the measured water quality data by a single or batch, and the system would built model itself and finally forecast disease severity rate. The users can take related measures in advanced to prevent disease occurrence by knowing the predicted result. Moreover, RF is an machine learning tools, which has great self-learning property. It need the users to feedback the real severity rates that are updated, to make improvements if the prediction result was sometimes in error.

Key word: Large yellow croak; Disease diagnosis; Pre-warning; Multiple linear model; Random forest algorithm

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

廈門大學博碩士論文摘要庫