

文章编号: 1004-7271(2009)01-0115-05

· 研究简报 ·

## 基于贝叶斯网的象山港网箱 轮养周期预测模型

滕丽华, 杨季芳

(浙江万里学院生物与环境学院, 浙江 宁波 315100)

**摘要:** 贝叶斯网络具有强大的推理能力, 能与先验知识和数据结合, 进行定性和定量分析, 提供了一条有效的处理预测问题的途径, 首先介绍了以上贝叶斯网络及其特点, 并讨论如何学习贝叶斯网络结构, 然后由专家知识和给定数据, 构造了一个海底网箱养殖的贝叶斯网络预测模型, 该模型能有效的表达网箱养殖环境各个指标之间的因果关系, 进而可以对指定的网箱养殖的移动周期进行预测和决策。实验结果表明, 试验数据显示评价的准确性是 89.7%。以上证明该方法是有可行性的, 表明贝叶斯网络是一种很有前途的预测评价方法。

**关键词:** 象山港; 轮养周期; 网箱养殖; 预测模型; 贝叶斯网

**中图分类号:** S 967.3      **文献标识码:** A

## Study on the net-cages mobile cycle prediction model based on Bayesian network in the Xiangshan Bay

TENG Li-hua YANG Ji-fang

(College of Biological and Environmental Science of Zhejiang Wanli University, Ningbo 315100, China)

**Abstract:** Bayesian network has a powerful ability for reasoning and semantic representation, which combines quantitative analysis with prior knowledge and observed data, and provides an effective way to deal with prediction. Firstly, this paper presented an overview of Bayesian network and its characteristics, and discussed how to learn a Bayesian network structure from given data, and then constructed a model of the Bayesian network for the period of Xiangshan Bay cage culture with expert knowledge and the dataset. The method can be effective expression cage culture environment between the various indicators of a causal relationship and designated by the cage culture of mobile cycle forecasting and decision making. The experimental results based on the test dataset show that evaluation accuracy is 89.7%. All these prove the method is feasible and efficient and indicate that Bayesian network is a promising approach for the prediction.

**Key words:** Xiangshan Bay; culture cycle; cage aquaculture; model; Bayesian network

象山港位于浙江省北部沿海, 南临三门湾, 北靠杭州湾, 东侧为舟山群岛, 是一个独立而完整的生态环境地理单元, 含有海域、港汊、岛屿、滩涂, 依托陆地, 具有特殊的生物群落, 构成了各自的独立的生态

收稿日期: 2008-01-07

基金项目: 宁波市海洋渔业局项目 (甬海办 2005/331-6)

作者简介: 滕丽华 (1974-), 女, 吉林人, 博士研究生, 讲师, 研究方向为环境化学与水污染治理。

通讯作者: 杨季芳, E-mail: th\_98@163.com

系统。港内自然环境优良,水产资源丰富,是多种鱼虾贝藻等海洋生物栖息生长、繁殖和育肥的优良场所,也是浙江省主要的海水养殖基地,有国家“大鱼池”之称<sup>[1]</sup>。

在网箱养殖过程中,大量残饵分布水中,加上鱼类排泄物,使水体富营养化,导致水域自身污染。同时对底质的影响也很明显,残饵等有机物不断积累的过程中,使底质变黑发臭,产生大量硫化物危及鱼类生存,因此,在养殖模式确定后,对网箱的移动周期进行科学的预测,以达到事半功倍的效果,目前国内对海底网箱养殖周期进行预测的模型研究鲜有报道。本文采用贝叶斯网方法进行建模,对指定的网箱养殖的移动周期进行预测和决策。贝叶斯网<sup>[2-3]</sup>是人工智能领域处理不确定性的主要方法之一,广泛应用在现代专家系统、诊断系统及决策支持系统中,其主要优势在于具有坚实的理论基础,能够有效的处理不完整数据,与其它技术相结合能够进行因果分析,能够使先验知识和数据有机的结合。因此,可用贝叶斯网对网箱养殖各个指标进行建模,通过得到的贝叶斯网来预测网箱养殖周期。

## 1 贝叶斯网原理

贝叶斯网是图形表示和概率知识的有机结合,是复杂联合概率分布的图形表示方式。它提供了一种自然的表示因果信息的方法,用来发现数据间的潜在关系。在这个网络中,用节点表示变量,有向边表示变量间的依赖关系,它揭示了领域对象的内在联系<sup>[4]</sup>。贝叶斯网的形式化定义如下:贝叶斯网是一个二元组  $S = \langle G, P \rangle$ ,其中  $G$ 是有向无环图,图中节点与领域知识的随机变量一一对应;网中的有向弧表示变量间的因果关系,从节点  $X$ 到节点  $Y$ 的有向弧的直观含义是  $X$ 对  $Y$ 有直接的因果影响; $P = \{P(X | \text{Parent}(X))\}$ 是局部概率分布的集合,条件概率表示因果影响的强度,其中  $\text{Parent}(X)$ 代表节点  $X$ 的父节点集合。该问题域中变量集合的联合概率分布可以表示成贝叶斯网中的每个节点的条件概率表的乘积,即:

$$P(X_1 \cdots X_n) = \prod_{i=1}^n P(X_i | \text{Parents}(X_i))$$

图 1表示了一个包含 6 个变量的贝叶斯网,右边是变量营火的局部概率分布。

贝叶斯网的通常用法是先利用专家经验或者利用机器学习方法从数据中学习或者二者的结合来构建贝叶斯网,然后通过贝叶斯网的推理来进行某些决策。其中贝叶斯网的推理是指在给定一组证据变量确切值的情况下,计算一组查询变量的概率分布。

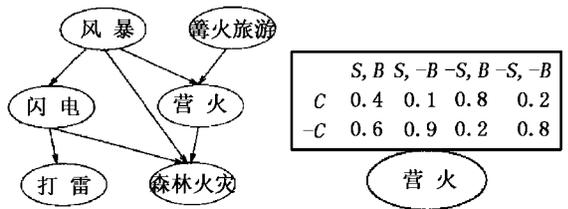


图 1 一个简单的贝叶斯网模型示例

Fig 1 A sample of Bayesian network

## 2 模型构建方法

对于网箱养殖区,网箱养殖环境评价指标体系由  $n$  个指标组成。就可以建立一个  $n+1$  个节点的贝叶斯网络,其中  $n$  个指标对应网中的  $n$  个节点,网箱养殖周期作为另一个节点,将网箱养殖周期分为 5 个级别,分别是 level 1: 2-3年, level 2: 3-4年, level 3: 4-5年, level 4: 5-6年, level 5: 6-7年。将每个指标看成一个节点,首先获得所有指标的取值集合及相应的网箱养殖周期,将此作为训练数据集,某个网箱养殖区的所有指标的一次取值以及网箱养殖周期  $D = \{C_1, C_2, \dots, C_n, H\}$  作为一个训练数据,其中  $C$  表示网箱养殖的第  $i$  个指标的取值,  $H$  表示网箱养殖周期。这个由  $n+1$  个变量组成的贝叶斯网络,反映了  $n$  个指标与网箱养殖周期期间的因果关系。

从网箱养殖周期节点引出的有向弧分别指向其他所有根节点。所有根节点的局部条件概率因此需要修改,此时它们的局部条件概率表示当前各指标在某一级别的条件下,该指标取各个值的概率。这些条件概率可以依靠领域专家来给出。

由于贝叶斯网络处理的是离散变量,所以需要将每个指标的取值离散化。以象山港网箱养殖区为

例,本文获得的原始数据是网箱养殖区在 11个指标下的取值,根据网箱养殖区连续数值属性特点,结合 GB3097—1997《海水水质标准》和 GB11607—1989《渔业水质标准》,采取有监督离散化算法<sup>[5]</sup>对指标的取值进行离散化,如表 1所示。

表 1 网箱养殖水环境指标取值分布情况

Tab. 1 The value distribution of the water environments factors

网箱养殖水环境指标	Low	Normal	Mid	High
投饵量 (kg/d)	<7	7~8	—	>8
水温 (°C)	<15	—	15~28	>29
透明度 (m)	<0.5	—	0.5~1	>1
pH	<6.5	6.5~8	—	>8
DO (mg/L)	<5	—	5~7	>7
COD (mg/L)	<1	—	1~2	>2
总磷 (mg/L)	—	<0.05	0.05~0.07	>0.07
总氮 (mg/L)	—	<1.2	1.2~1.5	>1.5
叶绿素 a (μg/L)	—	<5	5~8	>9
海水交换能力 (cm/s)	<20	20~40	—	>40
微生物降解能力 [μmol/(mL·d)]	<10	10~20	—	>20

### 3 网箱养殖周期的贝叶斯网模型构建与分析

部分样本数据记录,如表 2。

表 2 象山港网箱养殖区 2006年数据

Tab. 2 Part of the training sample data in 2006

水温 (°C)	pH	DO (mg/L)	COD (mg/L)	总磷 (mg/L)	透明度 (m)	总氮 (mg/L)	投饵量 (kg/d)	叶绿素 a (μg/L)	海水交换能力 (cm/s)	微生物降解能力 [μmol/(mL·d)]	网箱养殖周期 (a)
7.2	9.1	9.47	1.65	0.048	0.8	0.87	7.85	0.30	40	15.6	5.0
12.7	8.01	9.70	0.81	0.054	1.2	0.86	7.85	0.20	30	19.6	5.1
13.1	8.03	9.76	0.93	0.059	0.6	0.84	7.88	0.50	20	20.8	5.2
16.1	7.96	8.10	0.79	0.058	0.3	1.23	8.21	0.41	35	13.5	5.3
20.7	7.97	6.80	0.88	0.069	1.4	1.33	8.55	0.63	60	8.2	5.4
22.5	7.94	6.94	0.7	0.059	0.7	1.42	9.56	0.84	50	6.2	5.5
29.3	7.88	6.06	0.99	0.078	0.7	0.70	9.90	4.32	70	11.0	5.6
29.6	7.93	5.64	0.99	0.082	0.9	0.82	9.80	5.04	50	17.2	5.7
27.5	7.90	6.18	0.95	0.063	0.3	0.98	9.80	2.97	35	10.1	5.8
25.1	7.94	5.92	0.8	0.053	0.5	0.88	7.60	3.24	30	10.6	5.9
15.6	8.04	6.24	1.43	0.108	0.8	1.12	6.20	1.01	40	12.9	5.9
9.2	8.76	6.27	1.63	0.089	0.9	1.23	5.60	1.61	25	12.3	6.0

王飞等<sup>[6]</sup>给出了构造贝叶斯网络结构学习的遗传算法,其算法框架为:

- (1) 随机生成若干初始贝叶斯网,并用其作为第 0代群体  $Pop(0)$ ,从中选出一个个体作为最优贝叶斯网结构  $S_0$ ,  $t=0$ ;
- (2) 对于当前前代  $Pop(t)$ 的每一个个体  $S_{tj}$ 计算该个体的适应度函数  $F^{[S_{tj}]}$ ;
- (3) 执行选择,交叉,变异生成子代群体;其中交叉操作是交换两结构某节点的父节点集合,变异是对某结构增加,删除或反转一条边。进而形成新一代群体  $Pop(t+1)$ ;
- (4)  $t=t+1$ ;
- (5) 在新一代群体中找出适应度最高的个体作为当前最优结构  $S_t$ ;
- (6) 如果已经进化了  $g^1$ 代或连续  $g^2$ 代最优网络结构没有变化,则算法结束,否则返回(2)。

通过象山港网箱养殖区 2000—2007年监测数据的结构学习与专家经验知识,我们获得了象山港网箱养殖周期的贝叶斯网模型(图 2),该结构反映了各指标之间的因果关系,条件概率表表示因果影响的强度。

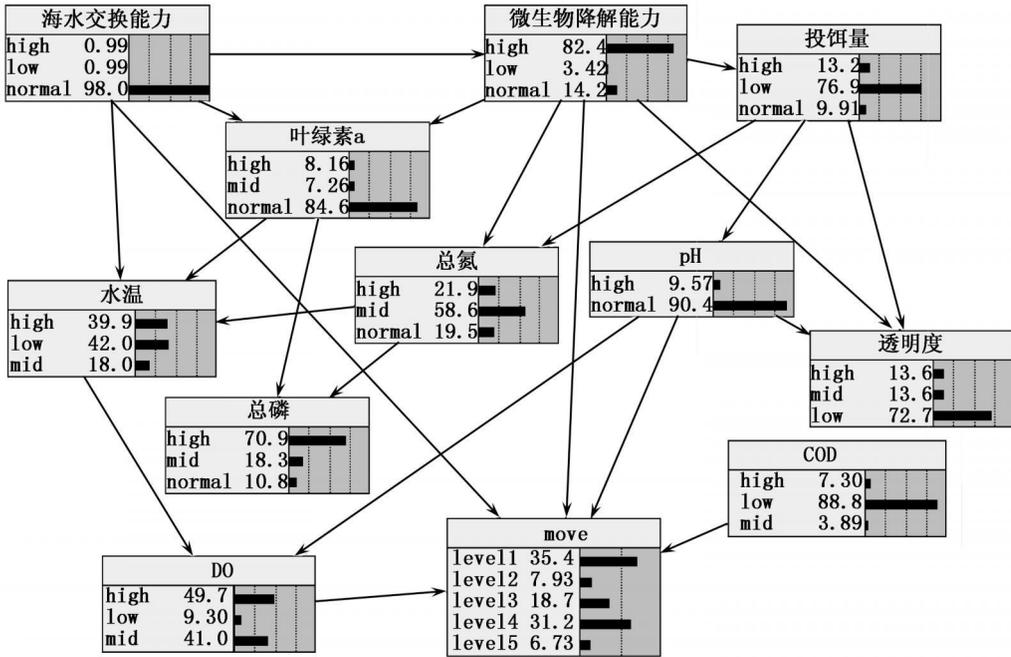


图 2 海底网箱轮养周期的贝叶斯网结构模型  
Fig 2 Bayesian network for the net cages mobile cycle

从网箱养殖周期的贝叶斯网评价模型(图 2)可以看出,溶解氧、海水交换能力、微生物降解能力、COD、pH等 5个变量直接影响网箱养殖周期,属于影响网箱养殖的周期的主导因素,应给予较大的权重;而投饵量,总氮,总磷,透明度,水温、叶绿素 a等 6个变量对网箱养殖周期不产生直接的影响,反映出实际存在的关联性,在判断网箱养殖周期等情况时应赋予较小的权重。因此,只需测定与网箱养殖周期直接相关的溶解氧、海水交换能力、微生物降解能力、COD、pH这 5个变量的值就能通过获得的贝叶斯网得出网箱养殖周期。同时,从图 2中可以看出,各个指标间的因果影响关系和强度,海水交换能力直接影响微生物降解能力,投饵量直接影响水体的 pH值,微生物降解能力、pH、投饵量与透明度,总氮、叶绿素 a、海水交换能力与水温,总氮、叶绿素 a与总磷,水温、pH与溶解氧都存在直接的关联性,表明微生物降解能力对投饵量有一定影响,同时微生物降解能力还受海水交换能力的影响;微生物降解能力和海水交换能力直接影响叶绿素 a的含量。

#### 4 模型验证与推理

基于网箱养殖周期的贝叶斯网预测模型,对其中的 200个数据实例进行预测,其预测精度达 89.7%。

通过 Zhang和 Poole<sup>[7]</sup>给出的贝叶斯网推理算法(变量消去法)来对网箱养殖周期进行预测和决策。该方法的基本思想是:待计算的条件概率为  $P(X_0 | X_E = x_E)$ ,这里  $X_0$  为查询变量的集合,  $X_E$  为给定证据的变量集合,令  $X_H$  为全体变量中  $X_0$  和  $X_E$  之外的变量。该方法的思想是:

$$P(X_0 | X_E) = \frac{P(X_0, X_E)}{P(X_E)} = \frac{\sum_{i \notin \text{QUE}} P(X_H = i | X_0, X_E)}{\sum_{i \notin \text{E}} P(X_H = i | X_E)}$$

该方法进行每次消除一个变量的循环操作,直到得到  $P(X_0 | X_E)$ 为止,它按照以下规则来消去变量:如果 X是代表证据的变量,则用  $\sum_x$  (求和)操作消除条件概率分布中的  $P(X \neq x_e | Pa(X))$ ;如果 X

$=X_0$ , 不进行求和操作, 即不消除  $X$ ; 如果  $X$  为其他情况, 则进行求和操作, 消除变量  $X$ 。以上过程一直进行, 并经过最后的归一化过程, 可以得到  $P(X_0 | X_E)$ 。

例如, 我们以 2007 年 4 月在象山港 XS06 站点的监测数据为例来预测网箱养殖周期, 我们只需获得与网箱养殖周期直接相关的 5 个指标数据即可, 如下表所示:

项目	COD (mg/L)	DO (mg/L)	pH	海水交换能力 (cm/s)	微生物降解能力 [ $\mu\text{mol}/(\text{mL}\cdot\text{d})$ ]
2007.4-XS06数据	1.20	7.71	7.9	35	22.8
离散后数据	mid	high	normal	normal	high

把离散后数据所获得的贝叶斯网模型, 得到该指标下的网箱养殖周期的概率分布如下:

$P=8.47\%$  (网箱养殖周期 = level 1 | COD = mid DO = high pH = normal 海水交换能力 = normal 微生物降解能力 = high)。

$P=1.54\%$  (网箱养殖周期 = level 2 | COD = mid DO = high pH = normal 海水交换能力 = normal 微生物降解能力 = high)。

$P=1.14\%$  (网箱养殖周期 = level 3 | COD = mid DO = high pH = normal 海水交换能力 = normal 微生物降解能力 = high)。

$P=86.02\%$  (网箱养殖周期 = level 4 | COD = mid DO = high pH = normal 海水交换能力 = normal 微生物降解能力 = high)。

$P=2.83\%$  (网箱养殖周期 = level 5 | COD = mid DO = high pH = normal 海水交换能力 = normal 微生物降解能力 = high)。

从上面的网箱养殖周期的条件概率分布推理计算结果可知, 网箱养殖周期等于 level 4 的概率最大为 86.02%, 因此, 我们可以得出网箱养殖周期为 5~6 年。

## 5 结束语

将贝叶斯网络引入到海底网箱养殖中属于首次尝试, 主要利用贝叶斯网络的推理和表达能力来建立科学的象山港海底网箱养殖周期评价体系, 结合机器学习和专家经验建造贝叶斯网, 通过贝叶斯网的推理来进行智能决策。获得的模型清晰、直观地揭示出养殖水环境各指标之间以及指标与网箱移动情况之间的内在因果关系和影响程度, 模型揭示出与网箱养殖周期直接有关的 5 个指标分别为溶解氧、海水交换能力、微生物降解能力、COD、pH。因此, 只需测定与网箱移动直接相关的这 5 个指标的数据, 就可以预测网箱养殖周期的概率。本文建立的贝叶斯网模型能够对网箱养殖周期进行有效预测, 但对模型的优化还需要更多的工作, 有待于在实践中进一步完善。

## 参考文献:

- [1] 蔡燕红, 项有堂. 象山港海水养殖功能区环境质量评价 [J]. 海洋通报, 2002, 21(4): 91-95.
- [2] Rish I, Dechter R. Resolution versus search: Two strategies for SAT [J]. Journal of Automated Reasoning, 2000, 24(1): 225-275.
- [3] Acid S, De C, L, Castellano J G. Learning Bayesian network classifiers: searching in a space of partially directed acyclic graphs [J]. Machine Learning, 2005, 59(3): 213-235.
- [4] Solares C, Sanz A M. Bayesian network classifiers: An application to remote sensing image classification [J]. EAS Transactions on Systems, 2005, 4(4): 343-348.
- [5] 李刚. 知识发现的图模型方法 [D]. 中国科学院软件研究所, 2001.
- [6] 王飞, 刘大有. 基于遗传算法的 Bayesian 网结构增量学习的研究 [J]. 计算机研究与发展, 2005, 42(9): 1461-1466.
- [7] Zhang N, L, Poole D. Probabilistic conflicts in a search algorithm for estimating posterior probabilities in Bayesian networks [J]. Artificial Intelligence, 1996, 88: 69-100.