
나이브베이스 분류자와 퍼지 추론을 이용한 적조 발생 예측의 성능향상

박선* · 이성로**

Enhancing Red Tides Prediction using Fuzzy Reasoning and Naive Bayes Classifier

Sun Park* · Seong Ro Lee**

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소
지원사업으로 수행된 연구임(2011-0022980)

요 약

적조란 유해조류의 일시적인 대 번식인 자연현상으로 어패류를 집단 폐사 시킨다. 적조에 의한 양식어업의 피해는 매년 발생하고 있다. 이 때문에 적조 발생을 미리 예측할 수 있으면 적조에 대한 피해를 최소화 시킬 수 있다. 적조발생 예측시 나이브베이스 분류자를 이용하면 좋은 예측결과를 얻을 수 있다. 그러나 나이브베이스를 이용한 결과는 단순한 발생 여부 만을 판별 할뿐 발생하는 적조가 어느 정도 증가 할지는 알 수 없다. 본 논문은 퍼지 추론과 나이브베이스 분류자를 이용한 새로운 적조발생 예측 방법을 제안한다. 제안방법은 적조 발생 예측의 정확률을 향상시키면서 적조생물 밀도의 증가율을 예측할 수 있다.

ABSTRACT

Red tide is a natural phenomenon to bloom harmful algal, which fish and shellfish die en masse. Red tide damage with respect to sea farming has been occurred each year. Red tide damage can be minimized by means of prediction of red tide blooms. Red tide prediction using naive bayes classifier can be achieve good prediction results. The result of naive bayes method only determine red tide blooms, whereas the method can not know how increasing of red tide algae density. In this paper, we proposed the red tide blooms prediction method using fuzzy reasoning and naive bayes classifier. The proposed method can enhance the precision of red tide prediction and forecast the increasing density of red tide algae.

키워드

적조 발생, 적조 예측, 퍼지 추론, 나이브베이스

Key word

red tide blooms, red tide prediction, fuzzy reasoning, naive bayes

* 정회원 : 목포대학교 정보산업연구소 전임연구교수

(교신저자, sunpark@mokpo.ac.kr)

접수일자 : 2011. 04. 29

심사완료일자 : 2011. 06. 10

** 정회원 : 목포대학교 정보전자공학과 교수

I. 서 론

적조(red tide)란 유해적조 생물의 일시적인 대량 번식으로 바다물의 색깔이 적색이나 황색으로 변하며 수산업에 피해를 일으키는 현상이다. 국내의 적조에 의한 피해는 90년대 이후로 양식장 등 수산업에 큰 피해를 주고 있으며, 특히 95년에는 역대 최대 피해를 보였다. 이후 매년 적조에 의한 피해가 발생하고 있으나, 최근 3년간은 수산업에는 큰 피해가 없었다. 그러나 언제 적조의 피해가 국내의 수산업에 다시 발생할지는 아무도 예측하지 못하고 있다. 국내에서는 60여 종의 적조 생물들이 있으며, 이중 수산업에 피해를 미치는 유해적조가 총 9종이 있다. 국내에서는 특히 *cocladinium polykrikoides*에 의한 양식업의 피해가 가장 크다[1][2][3].

이러한 적조에 의한 수산업의 피해가 매년 발생함에 따라서 적조에 대한 많은 연구가 진행되고 있으나, 주로 연구되는 분야는 적조원인 생물의 생리적 특성 및 환경 변화에 대한 생물의 반응으로 상당부분 연구되어 활용하는 수준에 있다[4]. 특히, 적조의 발생 시기를 예측하여 대비를 하면 적조에 의한 수산업의 피해를 최소화 시킬 수 있을 것이다. 그러나 적조 발생의 예측에 대한 국내의 연구는 아직 미흡한 편에 있다[1][3].

다음은 국내외에서 연구되는 적조 발생의 예측에 대한 관련연구이다. 참고문헌[5]에서는 사례기반 추론을 이용한 적조 예측 모니터링 시스템을 구현 및 설계하였다. 이들의 제안방법은 추론을 위한 사례베이스를 생성하고, kNN(k nearest neighbor) 알고리즘을 이용하여 가장 유사성이 높은 사례를 검색하였다. 여기서 kNN 알고리즘 방법은 k개의 자료들과 가장 가까이에 있는 범주를 그 자료들의 분류 범주로 정하는 방법이다. 그러나 이들의 방법은 단순히 기존 적조발생 사례를 기반으로 입력자료가 발생적조인지만을 분류하는데 그치고 있다. Fdez-riverola 등은 적조를 위한 예측 시스템을 제안하였다. 이들의 방법은 신경망과 퍼지의 혼합방법에 기반을 둔 사례기반 방법을 이용하였다. 사례검색단계에서는 자기 조직화 특징 지도(SOFM, self organized feature map) 신경망을, 재사용 단계에서는 RBF(radial basis function) 신경망을, 사례수정단계에서는 퍼지를, 사례유지단계에서는 퍼지, RBF 및 자기 조직화 특징 지도 신경망을 사용하였다[6]. 저자의 이전 연구[2]에서는 역전파 신경망

(BPNN, back propagation neural network)과 회귀 신경망 (GRNN, generalized regression neural network)을 이용하여 적조 발생을 예측할 수 있는 방법을 제안하였다. 또한 저자들은 적조 발생 및 발생 적조의 생물 밀도를 예측하기 위하여 퍼지 추론을 사용하는 적조 발생 예측 방법을 제안하였다. 이 방법은 역전파 신경망, 회귀 신경망, SVM(support vector machine)과 비교하여 더 좋은 성능을 보였으며, 적조 생물의 생물 밀도를 어느 정도 예측이 가능하나 후처리 단계가 없기 때문에 본 논문에서 제안한 방법에 비해서 성능이 떨어진다[3].

국내에서 발생하는 적조는 동해, 서해, 남해안 등 발생 범위가 다양하고, 발생 해역으로부터 해류를 따라서 빠르게 이동하기 때문에 직접 적조를 탐지하여 빠른 대처를 통한 피해를 최소화하기에는 한계가 있다[4]. 이러한 문제를 해결하기 위해서 과거에 발생한 적조 정보와 적조발생 지역의 현재의 해양 환경 정보를 이용하여 적조 발생 예측하면 적조에 의한 수산업의 피해를 최소화 시킬 수 있다. 참고문헌[7][8]에서는 적조 생물에 따라 다르지만 규조류나 편모조규는 1일 1회 2분열을 하므로 10일 정도 지나면 적조를 인식할 수 있는 적조기준밀도(1,000cells/ml)에 도달하는 것을 보고, 적조 발생 전 10일 동안의 해양환경조건이나 기상조건이 우리나라 적조발생의 해명에 중요하다고 보았다. 이러한 전제하에서 각 해역별로 적조가 발생하기 전 10일 동안의 표층 수온, 강수량 및 일사량을 조사하여 적조가 어떠한 환경에서 발생하는지를 조사하였다. 김 등[9]의 연구에서는 국내에 가장 피해를 많이 준 유해적조인 *cocladinium p.*의 발생은 주로 고수온과 저 염분에 의해 지배되는 것으로 보이고, 적조 발생 전 집중 호우가 관측된 후 염분농도가 급격히 감소한 경우에 많이 발생하는 것을 보였다[2][3].

이러한 적조와 환경인자의 관계에 대한 연구를 기반으로 본 논문에서는 적조발생 10일전의 해양환경자료를 이용하여서 적조 발생과 적조 생물의 발생한 생물 밀도를 예측할 수 있는 방법을 제안한다. 제안방법은 통영 지역에서 발생한 과거의 *cocladinium p.*의 적조발생정보와 적조예측일 이전 10일간의 해양환경자료에 퍼지 추론과 나이브베이스 분류자를 이용하여 적조 발생 예측의 성능을 향상시킨다.

본 논문의 구성은 다음과 같다. 2장에서는 퍼지추론과 나이브베이스 분류자에 대하여 알아보고, 3장에서

는 제안방법을, 4장에서는 적조발생예측에 대한 실험 평가 결과를, 5장에서는 결론 및 향후 연구 방향을 제시 한다.

II. 관련연구

2.1 퍼지추론

퍼지추론(fuzzy reasoning)은 애매하게 표현된 지식과 정보를 사용해 다른 정보를 찾아내는 방법이다. 퍼지이론에서 추론은 몇 개의 퍼지명제에서 연역적으로 각각 하나의 근사적인 퍼지명제를 유도하는 것을 기본으로 한다. 이 때문에 퍼지추론 또는 근사적 추론(approximate reasoning)이라 한다[3][10].

퍼지집합은 단위구간 $[0, 1]$ 사이의 실수 값을 멤버십(소속) 정도로 취하는 원소들로 구성된 집합으로 식(1)과 같이 나타낸다.

$$A = \{(x, \mu_A(x)) | x \in X, \mu_A(x) \in [0,1]\} \quad (1)$$

여기서 A 를 X 의 퍼지부분집합 또는 퍼지집합이라 하고, μ_A 는 전체집합 X 에 대한 멤버십함수(membership function)이며, 퍼지집합 A 의 $\mu_A(x)$ 값은 $x \in X$ 에 있어서 멤버십 값(membership value) 또는 등급(grade)으로 원소 x 가 퍼지집합 A 에 속하는 정도이다.

퍼지추론은 사실 A 와 규칙 $A \rightarrow B$ (A 이면 B 이다)일 때, 결론 B 가 사실임을 얻는 분리규칙(modus ponens)에 기초를 두며, 퍼지조건문으로 멤버십을 이용한 함위규칙은 식(2)과 같다.

$$\mu_A \rightarrow B(x, y) = \mu_A(x) \rightarrow \mu_B(y) \quad (2)$$

퍼지추론의 합성규칙은 퍼지규칙 $A \rightarrow B$ 는 퍼지관계로 식(3)의 결론 B' 는 퍼지집합 A' 와 퍼지관계 $A \rightarrow B$ 와의 합성(\circ)에 의해 얻는다.

$$B' = A' \circ (A \rightarrow B) \quad (3)$$

퍼지추론을 위한 추론규칙은 IF-THEN 형식으로 기술된다[3][10].

2.2 나이브베이스 분류자

나이브베이스 분류자는 나이브베이스 알고리즘에 기반을 두고 있다. 나이브베이스 알고리즘은 자료를 분류하기 위하여서 식(4)과 같이 베이스 정리를 이용하여서 주어진 자료에 대한 분류표시의 사후확률 값(posterior probability)을 추정하여서 이루어진다[11].

$$\Pr(c_j|d_i) = \frac{\Pr(c_j)\Pr(d_i|c_j)}{\Pr(d_i)} \quad (4)$$

여기서 $\Pr(c_j)$ 는 전체 자료집합에서 임의로 추출한 자료가 분류표시 c_j 에 속할 사전확률 값, $\Pr(d_i|c_j)$ 는 분류표시 c_j 에 속하는 자료집합에서 임의로 추출된 자료가 d_i 일 확률, $\Pr(d_i)$ 는 전체자료집합에서 임의 추출한 자료가 d_i 일 확률 값을 의미한다.

식(4)의 사후확률 값을 추정할 수 있도록 모델 ΘNB를 추정하고 나면 자료 분류자 $f_{\theta_{NB}}$ 는 입력된 자료에 대한 모든 분류표시에 대한 사후확률 값을 계산한 후에 가장 큰 사후확률 값을 가지는 분류표시를 반환한다. 이를 다음 식(5)과 같이 나타낼 수 있다.

$$\begin{aligned} f_{\theta_{NB}}(d_i) &= \operatorname{argmax}_{c_i \in C} \Pr(c_j|d_i) \\ &= \operatorname{argmax}_{c_i \in C} \frac{\Pr(c_j)\Pr(d_i|c_j)}{\Pr(d_i)} \end{aligned} \quad (5)$$

III. 제안방법

본 논문에서는 퍼지추론과 나이브베이스 분류자를 이용하여서 적조 발생 생물밀도 및 발생을 예측한다. 저자들의 이전 연구[2]에서는 단지 적조 발생을 예측할 수 있으나 발생하는 생물 밀도는 예상하지 못하였다. 또 다른 연구[3]에서는 적조발생 생물의 밀도 및 발생을 예측 할 수 있었으나, 예측한 적조생물 밀도의 오차 범위가 커지는 문제점을 가지고 있었다. 본 논문에선 이전[2][3] 연구의 문제점을 해결하기 위하여서 퍼지 추론과 나이브베이스 분류자의 양상별 방법을 이용하여 적조 발생의 생물 밀도와 발생을 예측하는 방법을 제안하였다. 실험 결과 적조발생 예측에 사용된 기법인 신경망과 SVM에 비하여서 더 높은 정확률을 보인 나이브베이스 분류

자를 이용하였다. 여기서 사용된 신경망은 역전파와 회귀신경망으로 역전파 신경망은 2개의 입/출력 층으로 구성되고, 학습과정에서 입력 층과 출력 층의 연결강도를 이용하여 구한 출력 값과 목표 값의 차이인 오차를 하위 처리단위로 되돌려 보냄으로써 오차를 감소시키는 방향으로 연결강도를 조정하여 분류기를 생성한다. 회귀신경망은 역전파 신경망과 비슷하게 학습하여 분류기를 생성하나 4개의 층으로 구성되는 차이가 있다. SVM은 학습 자료를 상위 차원으로 변형하고, 상위 차원으로부터 최적의 초평면을 검색하여서 분류기를 생성하는 방법이다. 나이브베이스 분류자가 더 높은 정확률을 보이는 것은 학습 자료인 과거의 적조발생자료가 나이브베이스 분류자의 특성에 더욱 부합되기 때문인 것으로 예상된다.

본 논문에서 제안한 적조 발생 예측 과정은 그림1과 같이 전처리, 퍼지추론, 나이브베이스 분류자, 후처리 단계로 구성된다. 전처리 단계에서는 과거 적조발생시의 해양환경자료를 퍼지추론에 적합한 학습 자료로 가공 및 정규화 한다. 퍼지추론 단계에서는 학습 자료를 이용하여 퍼지추론 규칙을 생성하고 생성된 규칙을 이용하여 적조발생시 적조생물 밀도를 예측한다. 나이브베이스 분류자 단계에서는 적조발생을 예측한다. 후처리 단계에서는 예측된 적조 생물 밀도와 발생을 조합하여서 적조 발생을 예측한다.

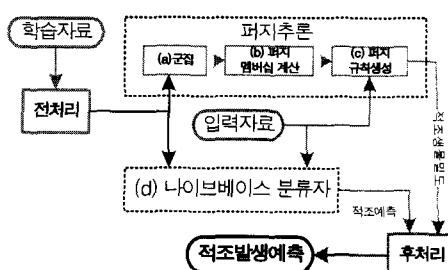


그림 1. 제안된 적조 발생 예측 블록도
Fig. 1. Block diagram of proposed red tide blooms prediction

3.1 전처리

그림1의 전처리에서는 적조발생예측을 위해서 학습 자료를 가공한다. 학습 자료는 2002년부터 2007년까지 6년간의 통영지역에서 발생한 적조생물인 *cocloidinium p.*의 발생자료 및 이 지역의 수온, 기온, 강수량을 이용한

다. 이 중에서 적조발생정보는 국립수산과학원의 적조 정보시스템[9]으로부터, 수온정보는 해양수산연구정보포털[12]의 연안정지관측정보로부터, 기온 및 강수량은 기상청[13] 관측 자료로부터 수집하였다. 표1은 학습자료 중 적조가 적게 발생한 2005년도의 적조 생물밀도이다. 표2는 표1의 적조가 발생하기 전의 10일간의 수온, 기온, 강수량 정보이다. 표3은 적조발생 건수, 적조 발생 전 수온 및 기온의 일수, 강수 횟수를 나타낸다[2][3].

표 1. 2005년도 *cocloidinium p.* 생물 밀도[3]
Table 1. The density of a *cocloidinium p.* in the year 2005[3].

일자	생물밀도	적조율
2005-07-29	640	0.64
2005-07-30	300	0.30
2005-08-04	6500	6.30
2005-09-08	5500	5.50
2005-09-11	1200	1.20

표 2. 적조 발생 전 10일간 해양환경[3]
Table 2. Sea environment during 10 days before red tide blooms[3]

월	일	수온		기온		강수량	
		C°	10일평균	C°	10일평균	월	10일총량
7	19	25.7	.	31.1	.	.	.
	20	26	.	29.4	.	.	.
	21	25.4	.	26.8	.	.	.
	22	24.3	.	29.6	.	.	.
	23	25.5	.	29	.	.	.
	24	26.8	.	30.4	.	.	.
	25	27.2	.	30.7	.	.	.
	26	27.5	.	29.4	.	.	.
	27	26.3	.	28.8	.	.	.
	28	26.5	.	27.6	.	.	.
8	29(적)	26.4	26.1	28.2	29.3	5.5	0 0
	30(적)	25.5	26.2	26.5	28.9	.	5.5 1
	31	26	.	27	.	.	.
	1	24.6	.	27.6	.	4	.
9	2	23.5	.	26.5	.	49.5	.
	3	25.8	.	28.1	.	.	.
	4(적)	27.5	25.9	30	28.2	.	59 3
	29	26	.	28.1	.	.	.
	30	26.5	.	26	.	.	.
10	31	25	.	29.9	.	.	.
	1	26.8	.	29.4	.	4.5	.
	2	28.5	.	30.2	.	.	.
	3	27.6	.	28.7	.	.	.
	4	24.8	.	25.4	.	.	.
	5	25.5	.	24.1	.	.	.
	6	24.7	.	23.2	.	50.5	.
	7	25.7	.	29	.	.	.
	8(적)	25.2	26.1	27.8	27.4	.	55 2
	9	25	.	24.9	.	84.5	.
	10	26.2	.	27.9	.	0.5	.
	11(적)	25.8	26	26.1	27.1	0.1	140 4

표 3. 적조 발생 및 해양환경 통계자료[3]
Table 3. Statistic data of red tide and sea environment[3]

구분	적조발생수	수온/기온 일수	강수 횟수
2002	27	41	16
2003	23	35	10
2004	21	33	9
2005	5	31	8
2006	3	13	5
2007	17	33	17

학습 자료를 위한 전처리로 수온 및 기온은 식(6)과 같이 적조발생 전 10일간의 평균값을 계산하고, 강수량은 식(7)과 같이 적조발생 전 10일간의 총 강수량을 계산한다. 이들 중에서 10일평균수온, 10일평균기온, 10일총강수량은 입력 학습 자료이고, 표1의 생물밀도는 출력 학습 자료이다. 표4의 수온 10일평균, 기온 10일평균, 강수량 10일총량, 생물밀도 등이 학습 자료를 위한 학습 자료의 변수를 정의한 것이다. 분류자에 사용되는 출력함수는 최소값 0과 최대값 1을 출력하기 때문에 최종적으로 입력 자료를 분류자에 적용하기 위해서는 0과 1사이의 값으로 정규화 시킨다[2][3].

표 4. 학습자료 변수의 사례 정의[2]
Table 4. Variables of training data that define a case[2].

변수	단위	일
수온	C°	$D_1, D_2, D_3, \dots, D_{N-1}, D_N$
기온	C°	$D_1, D_2, D_3, \dots, D_{N-1}, D_N$
강수	량	$D_1, D_2, D_3, \dots, D_{N-1}, D_N$
생물밀도	개체수	D_N

$$10\text{일 평균수온}/\text{기온} = \frac{\sum_{i=1}^N D_{N-i}}{N} \quad (6)$$

$$10\text{일 총강수량} = \sum_{i=1}^N D_{N-i} \quad (7)$$

여기서 N은 총 전처리 일수를 나타내고, 학습자료 변수를 구분하기 위하여서 D^W 는 수온, D^T 는 기온, D^R 은 강수량, D^{RT} 는 적조생물밀도를 각각 나타낸다.

3.2 퍼지 추론

퍼지 추론을 이용하여서 적조발생을 예측하기 위해서는 그림1(a)의 군집, 그림1(b)의 퍼지 멤버십 계산, 그림1(c)의 퍼지 규칙 생성단계로 구분된다. 즉, 적조예측을 위한 분류기를 생성하기 위해서 학습 자료를 식(8)을 이용하여서 유사한 특성의 군집으로 분류하며, 군집된 자료들을 식(9)를 이용하여 퍼지 멤버십을 계산하고 군집된 유사특성에 따라서 IF-THEN 형식으로 기술한다. 적조를 예측하기 위해서는 검증자료와 식(10)을 이용하여서 적조의 생물밀도를 예측한다.

3.2.1 군집

퍼지규칙을 생성하기 위해서는 학습 자료를 각 특성에 맞는 집단으로 구분할 필요가 있다. 이를 위해서 본 논문은 퍼지추론에서 많이 사용하는 감산 군집(subtractive clustering)방법을 이용한다. 감산 군집방법은 자료집합으로부터 유사한 특징을 갖는 집단을 추정하여 군집하는 방법이다. 본 논문에서는 군집추정을 위해서 Chiu의 방법[14]을 이용하며 다음 식(8)를 이용하여서 군집의 잠재적인 중심을 추정하여 군집에 이용한다[3].

$$P_i = \sum_{j=1}^n e^{-\frac{4}{r_{i,5}^2} \|x_i - x_j\|^2} \quad (8)$$

여기서, P_i 는 i번째 잠재 군집의 중심이고, r 은 주변 반지름으로 정의된 양의 계수, x_i 는 i번째 군집에 속하는 잠재적 자료, x_j 는 j번째 학습 자료이다.

3.2.2 퍼지 멤버십 계산

본 논문에서는 퍼지 멤버십 함수로 가우시안 함수(gaussian function)를 사용한다. 기본적인 퍼지 멤버십 함수는 삼각형(triangular), 가우시안(gaussian), 사다리꼴형(trapezoid)이 있다. 이들 중에서 본 논문은 세밀한 멤버십을 표현하기 위해서 가우시안 멤버십 함수를 이용한다. 가우시안 멤버십 함수[10]는 다음 식(9)과 같이 계산할 수 있으며 계산된 멤버십은 그림2와 같이 범종형 곡선을 나타낸다[3].

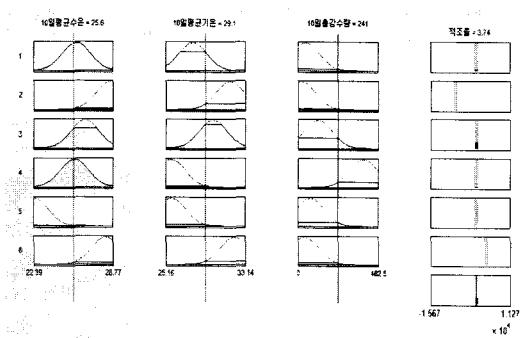


그림 2. 학습자료의 군집에 대한 분류규칙
Fig. 2. Classification rules of clustering of training data

$$f(x) = e^{\frac{-(x-c)^2}{2\sigma^2}} \quad (9)$$

여기서 x 는 학습 자료이고, σ 는 가우시안 곡선의 폭을 결정하는 상수이며, c 는 가우시안 곡선의 중심을 결정하는 상수이다. 본 논문에서는 $\sigma=2$, $c=5$ 를 사용해서 가우시안 곡선의 폭은 2로 곡선의 중심은 중앙에 위치하도록 초기 값을 설정하였다.

3.2.3 퍼지 규칙 생성

본 논문에서는 퍼지 규칙 생성을 위하여 퍼지 추론에서 많이 사용하는 sugeno 퍼지 모델[15]을 이용한다. sugeno 퍼지 모델의 출력 규칙의 최종 출력 값인 적조율은 식(10)과 같이 계산한다[3].

$$\text{final out} = \frac{\sum_{i=1}^N w_i z_i}{\sum_{i=1}^N w_i} \quad (10)$$

여기서 w 는 퍼지 규칙에 대한 가중치로 식(11)과 같으며, z 는 적조율 멤버십에 대한 군집, N 은 규칙의 수이다.

$$w_i = \prod_{k=1}^n P_k^i(x_k) \quad (11)$$

여기서 w_i 는 i 번째 규칙의 가중치이고, $f()$ 는 식(5)의 가우시안 멤버십 함수이다.

퍼지 규칙을 생성한 다음에 적조 예측 전 10일간의 적조발생지역의 환경정보를 입력하여 적조 발생 및 발생 생물밀도를 예측한다.

3.3 나이브베이스 분류자

본 논문에서는 나이브베이스 분류자를 이용하여 적조발생을 예측한다. 나이브베이스 분류자의 학습을 위하여서 사전확률 $\Pr(c_j)$ 와 $\Pr(d|c_j)$ 의 계산은 주어진 적조 학습 자료에 대하여서 모형의 매개 변수의 추정은 MAP(maximum a posteriori hypothesis)가설을 이용한다. $\Pr(c_j)$ 의 계산은 식(12)과 같이 학습자료 TD(training data) 중에 c_j 에 속하는 자료의 수로써 쉽게 계산할 수 있다 [11]. 여기서 TD는 표4에 정의된 수온, 기온, 강수, 생물밀도이다.

$$\Pr(c_j) = \frac{c_j \text{에 존재하는 자료의 수}}{|TD|} \quad (12)$$

3.4 후처리

본 논문의 적조발생 예측에 대한 실험결과 퍼지 추론에 비하여 나이브베이스 분류자를 이용한 예측 결과가 더 높은 정확률을 보인다. 이는 퍼지 추론의 분류자에 비해서 나이브베이스 분류자가 적조발생 예측의 학습자료로 더 적합한 것으로 예측된다. 이 때문에 후처리에서는 나이브베이스 분류자의 예측결과를 기준으로 이용하여서 퍼지 추론 단계에서 예측된 적조 발생 생물밀도와 예측결과를 조정해 준다. 예를 들어 나이브베이스 분류자의 결과가 적조 발생이나 퍼지 추론이 적조가 발생하지 않는 결과를 예측하면 퍼지 추론의 결과를 나이브베이스 분류자의 결과에 맞추어 적조 발생으로 맞추며 생물밀도는 1로 설정한다. 이와 반대의 경우는 생물밀도는 0으로 설정한다.

IV. 실험 및 평가

본 논문에서는 국립수산과학원의 적조정보시스템[9]으로부터 가져온 통영지역의 2008년부터 2010년 3년 동안 발생한 *coccolodinium p.* 적조경보 및 주의보 38건의 자료를 이용하여 제안방법의 성능을 평가하였다. 또한 같은 년도의 해양수산연구정보포털[12]의 연안정지판측

정보로부터 121일간의 수온정보를 이용하였으며, 기상청[13] 관측 자료로부터 수집한 121간의 기온정보와 29건의 강수정보도 이용하였다. 다음 표5는 실험에 사용한 입력 자료의 속성표이다[2][3].

표 5. 적조 발생 및 해양환경 속성표 [3]
Table 5. Property table of red tide and sea environment [3]

구분	적조발생수	수온/기온 일수	강수 횟수
2008	23	47	10
2009	4	23	5
2010	11	51	14

표6은 통영지역에서 3년간 발생한 적조 *cocladinium* p.에 대한 적조발생예측의 정확률 평가결과이다. 여기서 SVM(support vector machine)은 SVM[7]을 나타내고, BPNN(back propagation neural network)은 역전파 신경망[7]을 나타내며, GRNN(generalized regression neural network)은 회귀 신경망[7], FR(fuzzy reasoning)은 퍼지 추론[3], NB(naive bayes)는 나이브베이스 분류자[11]를 나타낸다. 평가결과를 NB 평균 정확률이 BPNN에 비해서 16.3%, GRNN에 비해서 13.7%이고 SVM에 비해서는 5.8%가, NB에 비해서는 3.3%가 더 높다[2][3].

표 6. 평가결과
Table 6. Result of experiment of evaluation

구분	BPNN	GRNN	SVM	FR	NB
정확률	73.7%	76.3%	84.2%	86.7%	90%

표7과 표8은 후처리 전후의 관계로 실제 적조 발생 생물 밀도와 예측결과 값의 일부분을 나타낸 것이다. 표7과 표8에서 FR은 퍼지 추론을 사용하여 예측된 적조발생 생물 밀도로 0보다 작으면 적조가 발생하지 않은 것을 나타내고, NB는 나이브베이스 분류자를 나타낸 것으로 0이면 적조가 발생하지 않은 것을, 1이면 적조가 발생한 것을 나타낸다.

표7에서 6~10번은 퍼지 추론이 잘못 예측한 값이고, 4번은 나이브베이스 분류자가 잘못 예측한 값이다. 표7을 후처리하여서 표8과 같이 보정한다. 즉, 표7의 4번은 0으로, 6~10번은 1로 보정하면 표8과 같은 결과가 나온다. 표6과 표8을 통해서 적조발생 예측의 정확율이 90%인 발생 생물밀도를 예측할 수 있다.

표 7. 후처리 전

Table 7. Before postprocessing

번	FR	NB	적조생물밀도률
1	15	1	0.72
2	13.8	1	0.92
3	12.3	1	0.60
4	10.3	0	0.90
5	0.656	1	2.65
6	-17.7	1	0.95
7	-11.7	1	0.95
8	-8.21	1	1.10
9	-5.26	1	0.98
10	-4.45	1	0.98
11	0.295	1	1.80
12	5.81	1	1.50

표 8. 후처리 후

Table 8. After postprocessing

번	FR	NB	적조생물밀도률
1	15	1	0.72
2	13.8	1	0.92
3	12.3	1	0.60
4	0	0	0.90
5	0.656	1	2.65
6	1	1	0.95
7	1	1	0.95
8	1	1	1.10
9	1	1	0.98
10	1	1	0.98
11	0.295	1	1.80
12	5.81	1	1.50

V. 결론

수산업에 대한 적조의 피해가 증가는 적조에 대한 많은 연구의 필요성을 가지고 수행하고 있으나, 대부분 생물학적 관점에서 연구되고 있다. 이 때문에 적조정보를 자동으로 처리할 수 있는 방법에 대한 연구는 미흡한 실정이다. 특히 적조 발생을 예측할 수 있으면 미리 대비하여 수산업에 대한 적조의 피해를 감소시킬 수 있다. 본 논문은 국내 적조연구에서 취약한 분야인 자동화에 의한 적조발생 예측의 성능 향상 연구이다. 본 논문에서 퍼지 추론과 나이브베이즈 분류자를 이용하여 미래에 발생할 적조 예측의 정확률을 향상시키는 방법에 대하여 제안하였다. 제안방법은 적조의 발생 예측 뿐 아니라 발생 생물의 생물밀도률도 어느 정도 예측할 수 있다. 제안

방법은 통영지역의 2002년부터 2007년 동안 발생한 *cocladinium p.* 적조 정보를 이용하여 학습하였으며, 2008년부터 2010년간의 정보를 이용하여서 제안방법을 평가하였다. 평가결과 3년간 평균 예측 정확률이 90%로 좋은 결과를 보였다. 본 논문에서 평가한 지역은 통영지역에 한정되어 있으나 앞으로는 국내 전 해역과 다양한 적조생물을 대상으로 실험을 진행할 예정이다. 또한 예측 정확률과 생물밀도률을 더욱 높일 수 있는 방법에 대해서 연구할 것이다.

참고문헌

- [1] 국립수산과학원 적조정보 홈페이지
<http://portal.nfrdi.re.kr/redtide/index.jsp>
- [2] 박선, 강일우, 이연우, 정민아, 오일환, 이성로, “신경망을 이용한 적조발생 예측”, 제21회 통신정보 합동 학술대회 논문집, pp.321-324, 2011.
- [3] 박선, “퍼지 추론을 이용한 적조 발생 예측”, 한국정보처리학회논문지 2011년 8월호 게재 확정, 2011.
- [4] 김용민, 변영기, 허용, 유기윤, “MODIS Level 2 Data를 이용한 *Cochlodinium Polykrikoides* 적조 탐지”, 대한토목학회논문지, pp.535-540, 2007.
- [5] 송병호, 정민아, 이성로, “사례 기반 추론을 이용한 적조 예측 모니터링 시스템 구현 및 설계”, 한국통신학회논문지, 제35권 제12호, pp.1819~1826, 2010.
- [6] F. Fdez-Riverola, J. M. Corchado, “FSfRT: Forecasting System for Red Tides”, Applied Intelligence 21, pp.251~264, 2004.
- [7] 이문옥, 김평주, 문진한, “진해만의 해양환경이 적조발생에 미치는 영향”, 한국해양환경공학회 2006년도 춘계학술대회 논문집, pp.177-183, 2006.
- [8] 이문옥, 김평주, “진해만의 해양환경과 적조발생의 특징”, 2006 대한토목학회 정기학술대회, pp.2173-2176, 2006.
- [9] 국립수산과학원 적조정보시스템,
[“http://portal.nfrdi.re.kr/redtide/index.jsp”](http://portal.nfrdi.re.kr/redtide/index.jsp), 2011.
- [11] 한학용, “패턴인식 개론 MATLAB 실습을 통한 입체적 학습”, 한빛미디어, 2009.
- [10] 오성권, “프로그래밍에 의한 컴퓨터지능(퍼지, 신경회로망 및 진화알고리즘을 중심으로)”, 내한출판사,

2002.

- [12] 해양수산연구정보포털,
[“http://portal.nfrdi.re.kr/page?id=pr_index”](http://portal.nfrdi.re.kr/page?id=pr_index), 2011.
- [13] 기상청, [“http://www.kma.go.kr/index.jsp”](http://www.kma.go.kr/index.jsp), 2011.
- [14] S. L. Chiu, “Fuzzy Model Identification based on Clustering Estimation”, Journal of Intelligent and Fuzzy System, Vol. 2, pp.267-278, 1994.
- [15] M. Sugeno, “Industrial Applications of Fuzzy Control”, Elsevier Science Publishers, pp.260-268, 1985.

저자소개



박선(Park Sun)

1996년 전주대학교(학사)
2001년 한남대학교
정보산업대학원(석사)
2007년 인하대학교 대학원(박사)

2008~2009년 호남대학교 컴퓨터공학과 전임강사
2010년 전북대학교 전기전자정보 인력양성사업단
박사후과정
2011년~현재 목포대학교 정보산업연구소
전임연구교수

※관심분야: 정보검색, 데이터마이닝, 데이터베이스,
해양생물 IT정보융합



이성로(Lee Seong Ro)

1987년 고려대학교(학사)
1990년 한국과학기술원(석사)
1996년 한국과학기술원(박사)
1997년 9월~현재 목포대학교
정보전자공학과 교수

※관심분야: 디지털통신시스템, 이동 및 위성통신
시스템, USN/텔레미티ックス응용분야, 임베디드시스템