

논문 2012-49SP-1-5

# 앙상블 학습을 이용한 적조 발생 예측의 성능향상

(Enhancing of Red Tide Blooms Prediction using Ensemble Train)

박 선\*, 정 민 아\*\*, 이 성 로\*\*\*

(Sun Park, Min A Jeong, and Seong Ro Lee)

## 요 약

적조란 유해조류의 일시적인 대 번식으로 바다를 적색으로 변화시키며 양식장의 어패류를 집단 폐사 시킬 뿐 아니라 연안 환경 및 바다 생태계에 악영향을 미치는 자연 현상이다. 적조에 의한 양식어업의 피해는 매년 발생하고 있으며 매년 적조방제에 많은 비용을 소비하고 있다. 이 때문에 적조 발생을 미리 예측할 수 있으면 적조에 대한 피해 및 방제 비용을 최소화시킬 수 있다. 본 논문은 앙상블 학습을 이용한 적조발생 예측 방법을 제안한다. 제안방법은 앙상블 학습의 bagging과 boosting 방법을 이용하여서 적조를 예측의 성능을 향상시킨다. 실험결과 제안방법은 단일 분류기에 비하여서 더 좋은 적조 발생 예측 성능을 보였다.

## Abstract

Red tide is a natural phenomenon temporary blooming harmful algal with changing sea color from normal to red, which fish and shellfish die en masse. It also give a bad influence to coastal environment and sea ecosystem. The damage of sea farming by a red tide has been occurred each year which it cost much to prevent disasters of red tide blooms. Red tide damage and prevention cost of red tide disasters can be minimized by means of prediction of red tide blooms. In this paper, we proposed the red tide blooms prediction method using ensemble train. The proposed method use the bagging and boosting ensemble train methods for enhancing red tide prediction and forecast. The experimental results demonstrate that the proposed method achieves a better red tide prediction performance than other single classifiers.

**Keywords** : 적조 발생(red tide blooms), 적조 예측(red tide prediction), 앙상블(ensemble), 분류기(classifier)

## I. 서 론

적조(red tide)란 유해적조 생물의 일시적인 대량 번식으로 바다물의 색깔이 적색, 황색, 적갈색으로 변하는

자연현상으로써, 연안 환경 및 생태계에 악영향을 미치며 양식장의 어패류를 집단 폐사시키면 수산업에 피해를 일으킨다. 국내의 적조에 의한 피해는 90년대 이후로 적조의 주원인 생물이 수산물과 인간에게 유독한 와편모조 변화되면서 양식장 등 수산업에 큰 피해를 주고 있다. 특히 95년에는 역대 최대인 764억 원의 피해를 보였으며, 매년 여름 우리나라 남해안에서 발생하고 있으나 점차 동해안 및 서해안으로 영역이 증가하는 추세에 있다. 그러나 최근 3년간은 수산업에는 큰 피해가 없었으나 매년 방제 및 적조 예방에 많은 예산을 소모하고 있다. 최근에는 지속적인 적조에 대한 피해가 없으나 언제 해양환경이 변해서 적조 피해가 국내의 수산업에 다시 발생할 지는 아무도 예측하지 못하고 있다. 국내에서는 60여 종의 적조 생물들이 있으며, 이중 수

\* 정회원-교신저자, 목포대학교 정보산업연구소  
(Research Institute of Information Science and Engineering, Mokpo National University)

\*\* 정회원, 목포대학교 컴퓨터공학과  
(Dept. of Computer Engineering,  
Mokpo National University)

\*\*\* 정회원, 목포대학교 정보전자공학과  
(Dept. of Information & Electronics,  
Mokpo National University)

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

접수일자: 2011년6월30일, 수정완료일: 2011년9월8일

산업에 피해를 미치는 유해적조가 총 7종이 있고, 이들은 어패류를 치사시키는 적조생물, 패류를 독화시키거나 식중독을 일으키는 적조생물 등으로 구분할 수 있다. 우리나라 연안에서 출현하여 어류나 패류를 직접 치사시키는 독성물질을 생성하는 종은 *coclodinium polykoides*, *heterosigma akashiwo*, *gymnodinium mikimotoi*, *gyrodinium sp* 등 4종이 있으며, 굴, 홍합 등 패류를 독화시켜 독화된 패류를 사람이 먹으면 마비성 패독증을 일으키는 종은 *alexandrium tamarense*, *dinophysis acuminata*, *nitzschia pungens f. multiseries* 등 3종이 있다. 이들 중에서 국내에서는 특히 *coclodinium polykoides*에 의한 양식업의 피해가 가장 크다<sup>[1~4]</sup>.

적조에 의한 양식장의 피해가 매년 발생함에 따라서 적조에 대한 많은 연구가 진행되고 있다. 그러나 적조에 대하여서 연구되는 분야는 적조원인 생물의 생리적 특성 및 환경변화에 대한 생물의 반응이 주로 연구되고 있으며, 이들 적조의 생화학적 특성 및 환경 영향에 대한 부분은 상당히 연구되어 활용하는 수준에 있다<sup>[5]</sup>. 특히 적조에 대한 연구 분야 중에서 적조발생 예측연구는 예측 결과를 분석하여서 필요시 가두리 양식장의 이동 및 방재 준비를 통하여서 적조에 의한 양식장의 피해를 최소화 시킬 수 있을 것이다. 그러나 적조 발생의 예측에 대한 국내의 연구는 아직 미흡한 편에 있다<sup>[1~4]</sup>.

국내에서 발생하는 적조는 서해, 남해안 등 발생 범위가 다양하고, 발생해역으로부터 해류를 따라서 빠르게 이동하는 특성을 가지고 있다. 이러한 적조의 다양한 발생 범위와 빠른 확산 때문에 직접 적조를 탐지하여서 적조의 피해를 대비하는 것에는 한계가 있다<sup>[5]</sup>. 적조 피해는 미리 적조 발생을 예측하여서 최소화 시킬 수 있다. 적조 발생은 과거에 발생한 적조 정보와 적조 발생 지역의 현재의 해양 환경 정보를 비교하여서 예측이 가능하다. 다음은 국내에서 연구된 적조생물의 특성 및 환경영향에 대한 연구이다. 이문옥 등의 연구에서<sup>[6][7]</sup>는 적조 생물에 따라 다르지만 규조류나 편모조류는 1일 1회 2분열을 하므로 10일 정도 지나면 적조를 인식할 수 있는 적조기준밀도(1,000cells/ml)에 도달하는 것을 보고, 적조발생 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.

김용민 등<sup>[8]</sup>의 연구에서는 국내에 가장 피해를 많이 준 유해적조인 *coclodinium 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<sup>[2~4]</sup>.

이러한 적조와 환경인자의 관계에 대한 연구를 기반으로 본 논문에서는 적조발생 10일전의 해양환경자료를 이용하여서 적조 발생 예측의 성능을 향상할 수 있는 방법을 제안한다. 제안방법은 통영지역에서 발생한 과거의 *coclodinium p.*의 적조발생정보와 적조예측일 이전 10일간의 해양환경자료에 양상블 학습을 이용하여 적조 발생 예측의 성능을 향상시킨다. 본 논문에서는 양상블 학습에 의사결정트리와 Adaboost(adaptive boosting) 분류기를 이용하였다.

본 논문의 구성은 다음과 같다. II장에서는 적조발생 예측에 대한 관련연구와 양상블 분류기에 대하여 알아보고, III장에서는 제안방법을, IV장에서는 적조발생예측에 대한 실험 평가 결과를, V장에서는 결론 및 향후 연구 방향을 제시한다.

## II. 관련연구

### 2.1 적조예측에 대한 관련연구

다음은 국내외에서 연구되는 적조 발생의 예측에 대한 관련연구이다. 송병호 등의 연구<sup>[8]</sup>는 사례기반 추론을 이용한 적조 예측 모니터링 시스템을 구현 및 설계하였다. 이들의 제안방법은 추론을 위해서 기존의 적조 발생 자료로 사례베이스를 생성하고, kNN알고리즘을 이용하여 가장 유사성이 높은 사례를 검색하였다. 그러나 이들이 사용한 kNN은 매우 단순한 방법이며, 수치형 영역의 사례기반 처리에는 제한이 있는 형태이다. 또한 이들의 방법은 단순히 기존 적조발생 자료와 입력 자료와의 유사도를 기반으로 유사도가 가장 높은 적조 발생 자료를 선택하여서 입력 자료가 적조인지를 판별하는 데 그치고 있다. Fdez-riverola 등의 연구<sup>[9]</sup>는 적조를 위한 예측 시스템을 제안하였다. 이들의 방법은 신경망과 퍼지의 혼합방법에 기반을 둔 사례기반 방법을 이용하였다. 이들의 방법은 새로운 입력 자료가 들어오면 사례베이스로부터 사례를 검색하여 재사용하고, 새로운 해결 방안이 필요하면 새로운 해결 방법에 맞도록 수정하여서 새로운 사례를 만들어 문제를 해결하고 사례베이스에 저장 한다. 즉 구체적인 방법으로는 사례

검색단계에서 자기 조직화 특징 지도(SOFM, self organized feature map) 신경망을 이용하여서 사례를 검색하고, 재사용 단계에서는 RBF(radial bassis function) 신경망을 이용하여 문제에 적합한 사례를 사용하며, 문제에 적합한 사례가 없으면 사례수정단계에서 퍼지를 이용하여 새로운 사례 및 해결 방법을 만든다. 마지막으로 사례유지단계에서는 퍼지 와 RBF/자기 조직화 특징 지도 신경망을 사용하여 새로운 사례베이스를 구축하였다<sup>[9]</sup>. 본 논문의 저자들의 이전 연구<sup>[2~4]</sup>에서는 신경망과 SVM(support vector machine)을 이용하여서 적조 발생을 예측할 수 있는 방법을 제안하였다. 적조를 예측할 수 있도록 적조자료를 전처리하여서 모형화 하였으며, 모형화된 자료를 기반으로 적조 발생 예측을 위하여 역전파 신경망, 회귀 신경망, SVM 분류기를 각각 이용하였다. 제안방법의 비교 실험 결과 SVM 분류기를 이용한 방법 가장 높은 정확률을 보였다<sup>[2]</sup>. 이 방법은 적조의 단순한 발생만을 예측하는 이원 분류에 기반을 두고 있다. 저자들의 또 다른 제안방법은 적조 발생 및 발생 적조의 생물 밀도를 예측하기 위하여 퍼지 추론을 사용하는 적조 발생 예측 방법을 제안하였다. 이 방법은 이전 저자들이 제안한 방법에 비하여 더 좋은 성능을 보였으며, 적조 생물의 생물 밀도를 어느 정도 예측이 가능 하였다<sup>[3]</sup>. 또한 저자들은 나이브베이스 분류기와 퍼지 추론을 이용하여서 적조 발생 예측의 성능을 향상 시키는 방법을 제안하였다<sup>[4]</sup>.

2.2 앙상블 학습

앙상블 학습은 분류 및 예측 분류자의 성능 향상을 위해서 제안된 방법이다. 앙상블 학습은 기본 분류기(base classifiers) 집합인 앙상블을 구성하고, 앙상블의 복수의 학습결과를 하나의 강한 분류기(a single strong classifier)와 결합하여 최종 결과를 계산한다. 앙상블을 구성하는 이유는 단일 분류자만으로 학습이 어려운 복잡한 패턴들을 여러 개의 하위 패턴으로 나누어 효과적으로 학습할 수 있다. 또한 단일 분류자 보다 앙상블 학습결과가 더 높은 정확률을 제공한다<sup>[10~11]</sup>.

앙상블 학습에가 가장 일반적으로 사용되는 방법은 bagging과 boost가 있다<sup>[10~11]</sup>. bagging은 n개의 학습 자료에서 복원추출 방법으로 k개의 bootstrap 분석용 자료를 생성하여서 각각의 분류기를 구성한다. 최종 결과는 각각의 분류기의 결과와 강한 분류기인 bagging의 결합함수를 이용하여서 식(1)과 같이 계산된다.

$$C(x_i) = \frac{1}{k} \left( \sum_{j=1}^k \alpha_j C_j(x_i) \right) \tag{1}$$

여기서  $C(x_i)$ 는 i번째 관측 값에 대한 최종 결과, k는 bootstrap의 개수,  $\alpha_j$ 는 j번째 분류자의 정확률,  $C_j(x_i)$ 는 i번째 학습벡터에 대한 j번째 분류기의 학습결과이다.

boost 학습방법 중에서 가장 일반적으로 사용되는 방법은 AdaBoost방법<sup>[10~11]</sup>이 있다. AdaBoost방법은 분류기를 순차적으로 생성하고, 학습 자료를 이전의 분류기의 결과를 기반으로 추출된 각 결과 값을 사용한다. 처음 학습 자료의 가중치는 동일한 상태에서 시작하여야 한다. 분류기에 의해서 오분류된 분류 값에는 높은 가중치를 주고, 정 분류된 분류 값에는 낮은 가중치를 부여함으로써 분류 값의 학습 자료에 대한 가중치를 재조정한다. 이러한 과정을 통하여서 가중치가 재조정된 분류 값들을 통하여서 새로운 학습 자료가 생성되어서 가중치가 증가한 학습 자료가 많이 선택됨으로써 분류하기 힘든 학습 자료를 더 잘 분류하도록 한다<sup>[10~11]</sup>.

III. 제안방법

저자들의 이전 연구<sup>[2~4]</sup>에서는 신경망, SVM, 나이브 베이스, 퍼지 등의 분류기를 이용하여서 적조발생을 예측하였다. 본 논문에서는 이전 적조발생 예측 방법의 성능을 향상시키기 위하여 앙상블 학습을 이용한 방법을 이용한다.

본 논문에서 제안한 적조 발생 예측 과정은 그림1과 같이 전처리, 분류기 생성 단계로 구성된다. 전처리 단계에서는 과거 적조발생시의 해양환경자료를 학습에 의한 분류기 생성과 평가에 적합한 자료로 가공한다. 분류기 단계에서는 학습 자료와 앙상블 학습을 이용하여서 분류기를 생성하고, 생성된 분류기와 입력 자료를

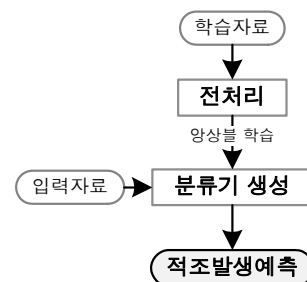


그림 1. 제안된 적조 발생 예측 블록도  
Fig. 1. Block diagram of proposed red tide blooms prediction.

이용해서 적조의 발생을 예측한다. 본 논문에서는 양상 블 학습에 가장 많이 사용되는 의사결정트리를 기본 분류기로 사용하는 bagging과 AdaBoost 부스팅 방법을 이용한다.

3.1 전처리

그림 1의 전처리 단계에서는 분류기 생성을 위한 학습 자료와 적조 예측을 위한 입력 자료를 가공한다. 분류기 생성을 위한 학습 자료는 2002년부터 2007년까지 6년간의 통영지역에서 발생한 적조생물인 coclodinium p.의 발생자료 및 이 지역의 수온, 기온, 강수량을 이용한다. 적조 예측을 위한 평가 자료로는 2007년부터 2010년까지 4년간의 같은 통영 지역의 수온, 기온, 강수량을 이용하였으며, 분류기의 예측 정보를 평가하기 위하여 같은 기간 동안의 적조 발생자료를 이용하여 가공하였다. 이 중에서 적조발생정보는 국립수산과학원의 적조정보시스템<sup>[1]</sup>으로 부터, 수온정보는 해양수산연구 정보포털<sup>[2]</sup>의 연안정지관측정보로부터, 기온 및 강수량은 기상청<sup>[3]</sup> 관측 자료로부터 수집하였다.

표 1은 학습자료 중 적조가 적게 발생한 2005년도의 적조 생물밀도이다. 표 2는 표 1의 적조가 발생하기 전의 10일간의 수온, 기온, 강수량 정보이다. 표3은 적조 발생 건수, 적조 발생 전 수온 및 기온의 일수, 강수 횟수를 나타낸다<sup>[2-4]</sup>.

학습 자료와 입력 자료를 위한 전처리로 수온 및 기온은 식(2)과 같이 적조발생 전 10일간의 평균값을 계산한다. 또한 강수량은 식(3)과 같이 적조발생 전 10일간의 총 강수량을 계산한다. 식(2)과 식(3)에 의한 10일 평균수온, 10일평균기온, 10일총강수량은 학습 자료의 학습과 입력 자료의 분류예측을 위해 사용되고, 표 1의 생물밀도는 학습 자료의 분류학습과 입력 자료의 분류 예측 평가에 사용된다. 분류기에 사용되는 출력함수는 -1과 1을 출력하기 때문에 최종적으로 표 1의 생물밀도

표 1. 2005년도 coclodinium p. 생물 밀도<sup>[3-4]</sup>  
Table 1. The density of a coclodinium p. in the year 2005<sup>[3-4]</sup>.

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

-1과 1로 변화시킨다. 즉, 생물밀도가 1보다 큰 값을 가지면 적조가 발생하는 것으로 여기고 1로, 생물밀도가 1보다 작은 값을 가지면 적조가 발생하지 않는 것으로 여기고 -1로 변화시킨다. 표 4의 수온 10일평균, 기온 10일평균, 강수량 10일 총량, 생물밀도 등이 분류기 생성을 위한 학습 자료와 적조발생 예측과 평가를 위한 입력 자료의 변수를 정의한 것이다<sup>[2-4]</sup>.

표 2. 적조 발생 전 10일간 해양환경<sup>[3-4]</sup>  
Table 2. Sea environment during 10 days before red tide blooms<sup>[3-4]</sup>.

월	일	수온		기온		강수량		
		C°	10일평균	C°	10일평균	량	10일총량	횟수
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	.	.	.	.
	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	.	.	.	.	
8	1	24.6	.	27.6	.	4	.	.
	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	.	.	.	.
31	25	.	29.9	.	.	.	.	
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<sup>[3-4]</sup>  
Table 3. Statistic data of red tide and sea environment.<sup>[3-4]</sup>

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

표 4. 학습자료 변수의 사례 정의<sup>[2,4]</sup>  
Table 4. Variables of training data that define a case<sup>[2-4]</sup>.

변수	단위	입
수온	C°	$D_1^W, D_2^W, D_3^W, \dots, D_{N-1}^W, D_N^W$
기온	C°	$D_1^T, D_2^T, D_3^T, \dots, D_{N-1}^T, D_N^T$
강수	량	$D_1^R, D_2^R, D_3^R, \dots, D_{N-1}^R, D_N^R$
생물밀도	개체수	$D_N^{RT}$

$$10\text{일평균수온/기온} = \frac{\sum_{i=1}^N D_{N-i}}{N} \quad (2)$$

$$10\text{일총강수량} = \sum_{i=1}^N D_{N-i} \quad (3)$$

여기서  $N$ 은 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 앙상블을 이용한 분류기 생성

대표적인 앙상블 학습 방법에는 bagging과 boosting이 있다. Brieman은 bagging과 같은 앙상블 학습은 불안정한 학습 분류기의 성능을 크게 향상시킬 수 있으나, 안정 적인 학습 분류기에서는 효과적인 성능 향상이 보이지 않는다고 주장 하였다. 즉, 의사결정트리와 같은 불안정한 분류기는 학습 자료의 변화에 민감하게 반응하여 의사결정트리 구조가 변경되는 등 분류기에 큰 영향을 주어서 좋은 성능향상을 보인다. 그러는 반면, 신경망이나 SVM 학습 분류기는 학습 자료가 변화하더라도 유사한 분류기를 반복적으로 생성하므로 분류기 사이의 상관관계가 높아져 뚜렷한 성능 향상을 보이지 않는다<sup>[11, 14]</sup>. 또한 boosting의 경우 분류하기 힘든 관측 값에 대해서 높은 가중치를 주어 분류를 잘 유도하기 때문에 일반적으로 bagging과 비교하였을 때 boosting이 매우 효과적으로 오차를 감소시킨다고 알려졌다<sup>[15]</sup>.

본 논문에서는 이러한 이전 연구결과를 기반으로 적조 발생 예측에 bagging에는 의사결정트리를 이용한 앙상블 학습을, boosting에는 일반적으로 좋은 성능을 보이는 AdaBoost방법을 적용하여 분류기를 생성한다.

#### 3.2.1 bagging 앙상블 학습에 의한 의사결정트리 분류기 생성

본 논문에서는 bagging 앙상블 학습을 사용되는 기본 분류기로 의사결정트리 방법 중에서 가장 많이 사용하는 필수 방법인 CART(classification and regression tree) 방법<sup>[16]</sup>을 사용한다. CART 방법은 전체 자료집합으로부터 시작하여 반복해서 두개의 자식 노드를 생성하기 위해 모든 예측 변수를 사용하여 자료 집합의 부분집합을 분리함으로써 의사결정트리를 생성한다. CART에서는 속성 선택을 위한 기준으로 식(4)과 같은

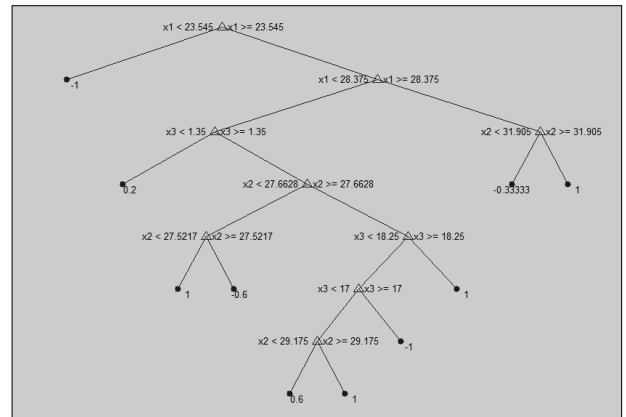


그림 2. 학습 자료와 CART 방법에 의한 분류기  
Fig. 2. The constructed classifier using CART method and training data.

엔트로피 매트릭스를 사용한다.

$$I_G(f) = \sum_{i=1}^m f_i(1-f_i) = \sum_{i=1}^m (f_i - f_i^2) = \sum_{i=1}^m f_i - \sum_{i=1}^m f_i^2 = 1 - \sum_{i=1}^m f_i^2 \quad (4)$$

여기서  $f_i$ 는  $i$ 번 자료가 분기되는 분류이다.

다음 그림 2는 2002년부터 2007년까지 6년간의 자료를 전처리한 학습 자료와 식(4)의 CART 앙상블 학습 방법을 이용하여서 의사결정트리 분류기를 생성하여 도식한 것이다.

본 논문에서는 적조 발생의 예측을 위해서 의사결정트리를 기본 분류기로 하여서 앙상블의 bagging 학습을 한다. 적조 발생 예측을 위해 생성하는 앙상블 학습 모형은 다음과 같다. 먼저 전처리된 6년간의 적조발생 및 관련 해양환경 학습 자료로부터 40개의 bootstrap 표본을 추출하고 각 bootstrap에 CART 분류기 모형을 적합 한다. 본 논문에서는 bootstrap 표본의 개수를 40개로 결정한 이유는 1부터 200까지 표본을 증가 시키면서 실험결과 40일 때의 정확률이 가장 높았기 때문이다. 다음으로 식(1)을 이용하여서 40개의 분류기를 결합한다. 식(1)의 결과 40개의 분류기 중에서 가장 큰 값을 가지는 분류기를 최종 분류기로 정한다.

#### 3.2.2 AdaBoost에 의한 분류기 생성

AdaBoost 학습방법<sup>[10~11]</sup>은  $n$ 개의 학습 자료와  $k$ 개의 기본 분류기와 구성된 앙상블  $C = \{C_1, C_2, \dots, C_k\}$ 와  $k$ 번째 분류자의 오류률( $e_k$ )인 식(5)과 같은 단순평균으로 계산된다.

$$e_k = \frac{1}{n} \sum_{i=1}^n L(C_k(x_i), y_i), L(C_k(x_i), y_i) = \begin{cases} 1 & C_k(x_i) \neq y_i \\ 0 & C_k(x_i) = y_i \end{cases} \quad (5)$$

여기서,  $x_i$ 는  $i$ 번째 학습 자료의 벡터이고,  $y_i$ 는  $i$ 번째 학습 자료의 분류이며,  $C_k(x_i)$ 는  $k$ 번째 학습벡터  $x_i$ 에 대한  $k$ 번째 분류기의 학습결과이다.  $L()$ 은 분류평가 함수로 분류 결과가 맞으면 0이고 맞지 않으면 1이다.

$k+1$ 번째 분류기에서  $i$ 번째 학습 자료에 부여되는 가중치는 다음 식(6)으로 조정되어 오분류된 학습 자료에 더 높은 가중치가 부과된다.

$$w_{k+1}(i) = w_k(i) \exp(\alpha_k L(C_k(x_i), y_i)), \alpha_k = \ln\left(\frac{1 - e_k}{e_k}\right) \quad (6)$$

식(5)을 이용하여서 오분류된 학습 자료에 높은 가중치가 부여되며  $k+1$ 번째 분류기의 학습 자료를 구성할 때 가장 높은 오분류 자료가 포함될 가능성이 높기 때문에 boost 방법은 오분류 자료에 초점을 맞춘 학습을 진행할 수 있다.  $i$ 번째 학습 자료의 최종 결과는 강한 분류기인 boosting 분류기를 통하여서 bagging의 식(1)과 같이 계산된다.

#### IV. 실험 및 평가

본 논문에서는 국립수산과학원의 적조정보시스템<sup>[1]</sup>으로 부터 가져온 통영지역의 2007년부터 2010년 4년 동안 발생한 coclodinium p. 적조경보 및 주의보의 자료를 이용하여 제안방법의 성능을 평가하였다. 또한 같은 년도 및 적조발생 해역의 해양 환경자료를 해양수산연구정보포털<sup>[12]</sup>의 연안정지관측정보로부터 획득한 수온정보를 이용하였으며, 기상청<sup>[13]</sup> 관측 자료로부터 수집한 기온정보와 강수정보도 이용하였다. 다음 표 5는 실험에 사용한 입력 자료의 속성표이다<sup>[2-4]</sup>.

표 6은 통영지역에서 4년간 발생한 적조 coclodinium

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

구분	적조발생수	수온/기온 일수	강수 횟수
2007	17	33	14
2008	23	47	10
2009	4	23	5
2010	11	51	14

표 6. 평가결과

Table 6. Result of experiment of evaluation.

구분	GRNN	SVM	DT
정확률	0.8545	0.9091	0.9455

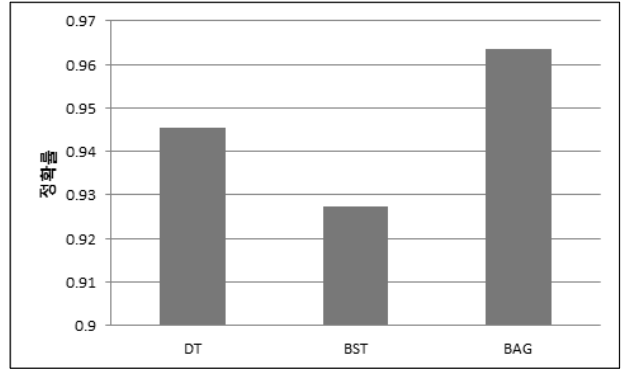


그림 3. 제안방법의 정확률 비교 결과

Fig. 3. Result of comparison of precision of proposed methods.

p.에 대한 적조발생예측의 정확률 평가 결과이다. 본 논문에서의 정확율은 이원분류의 평가에 많이 사용하는 미시평균 정확율(microaveraged precision)<sup>[17]</sup>을 사용하였다. 여기서 SVM는 SVM<sup>[2, 6]</sup>을 나타내고, GRNN 회귀 신경망<sup>[2, 6]</sup>, DT는 의사결정트리<sup>[10]</sup>를 나타낸다. 평가결과를 DT 평균 정확률이 GRNN에 비해서 9.1%이고 SVM에 비해서는 3.64%가 더 높다.

다음 그림 3은 제안 방법의 비교 결과이다. 그림3에서 DT는 의사결정트리이고, BST는 AdaBoost 앙상블 학습방법이며, BAG은 의사결정트리를 기본 분류기로 사용한 bagging 앙상블 학습방법이다. 각각의 방법의 정확률은 DT가 0.9455, BST가 0.9273, BAG가 0.9635이다. 그림 3에서 보는 것과 같이 AdaBoost 앙상블 학습방법인 BST가 단일 분류기인 의사결정트리 DT보다 더 낮은 성능을 보이고 있다. 이는 적조발생 예측의 자료특성이 의사결정트리를 이용한 분류기에 더욱 적합한 것으로 보인다. 또한 의사결정트리를 기본 분류기로 bagging 앙상블의 정확률이 가장 높은 값을 갖는 것은 의사결정트리에 의한 앙상블 학습이 적조발생 예측의 성능을 향상 시키는 것을 알 수 있다.

#### V. 결 론

양식장에 대한 적조의 피해가 증가하면서 적조에 대한 많은 연구가 수행되고 있으나, 대부분 생물학적 관

점에서 연구되고 있다. 특히 적조 발생을 자동으로 예측할 수 있으면 양식장의 적조방제준비와 가두리양식장의 이동을 통하여서 적조 피해를 감소시킬 수 있다. 본 논문은 앙상블 학습에 의한 적조발생 예측의 성능 향상에 대한 연구이다. 본 논문에서 의사결정트리를 기본 분류기로 이용한 bagging 앙상블 학습과 AdaBoost 앙상블 학습을 이용하여 미래에 발생할 적조 예측의 정확률을 향상시키는 방법에 대하여 제안하였다. 제안방법은 통영지역의 2002년부터 2007년 동안 발생한 coclodinium p. 적조 정보와 같은 지역의 해양 환경 정보를 이용하여 학습하였으며, 2007년부터 2010년 동안 4년간의 정보를 이용하여서 제안방법을 평가하였다. 평가결과 의사결정트리를 기본 분류기로 사용한 bagging 앙상블 학습방법이 적조발생 예측의 성능을 향상시켰다. 본 논문에서는 단순히 적조가 발생 여부만을 예측하는 이원분류에 그치고 있다. 이 때문에 적조 예측과 함께 적조 생물의 발생 밀도도 효율적으로 예측할 수 있는 방법에 대한 연구가 필요하다.

## 참 고 문 헌

- [1] 국립수산과학원 적조정보 홈페이지  
<http://portal.nfrdi.re.kr/redtide/index.jsp>
- [2] 박선, 이진석, 이성로, “신경망과 SVM을 이용한 적조발생 예측”, 대한전자공학회 논문지 2011년 9월호 게재예정, 2011.
- [3] 박선, 이성로, “퍼지 추론을 이용한 적조 발생 예측”, 한국정보처리학회 논문지 2011년 8월호 게재예정, 2011.
- [4] 이정훈, 박선, 최명수, 김영주, 민산원, 정민아, “나 이브베이스 분류자와 퍼지 추론을 이용한 적조 발생 예측의 성능향상”, 2011년도 한국통신학회 하계 종합학술발표회, Vol. 45, pp.158, 2011.
- [5] 김용민, 변영기, 허용, 유기윤, “MODIS Level 2 Data를 이용한 Cochlodinium Polykrikoides 적조 탐지”, 대한토목학회논문지, pp.535-540, 2007.
- [6] 이문옥, 김평주, 문진한, “진해만의 해양환경이 적조발생에 미치는 영향”, 한국해양환경공학회 2006년도 춘계학술대회 논문집, pp.177-183, 2006.
- [7] 이문옥, 김평주, “진해만의 해양환경과 적조발생의 특징”, 2006 대한토목학회 정기학술대회, pp.2173-2176, 2006.
- [8] 송병호, 정민아, 이성로, “사례 기반 추론을 이용한 적조 예측 모니터링 시스템 구현 및 설계”, 한국통신학회논문지, 제35권 제12호, pp.1819~1826, 2010.
- [9] F. Fdez-Riverola, J. M. Corchado, “FSIRT: Forecasting System for Red Tides”, Applied Intelligence 21, pp.251~264, 2004.
- [10] J. Han, M. Kamber, “Second Edition, Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers, 2006.
- [11] 김명중, “유전자 알고리즘을 이용한 분류자 앙상블의 최적 선택”, 지능정보논문지, 제16권 제4호, pp. 99~112, 2010.
- [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>”, 2011.
- [14] L. Breiman, “Bagging predictors”, Machine learning, vol. 24, no. 2, pp.123-140, 1996.
- [15] 정연해, 어수행, 문호석, 조형준, “앙상블기법을 이용한 다양한 데이터마이닝 성능향상 연구”, 한국통계학회논문집, 제17권, 제4호, pp. 561-574, 2010.
- [16] Breiman, L., J. Friedman, R. Olshen, and C. Stone. “Classification and Regression Trees”, Boca Raton, FL: CRC Press, 1984.
- [17] Soumen, C. “mining the web discovering knowledge from hypertext data”, Morgan Kaufmann, 2003.

— 저 자 소 개 —



**박 선**(정회원)-교신저자  
 1996년 전주대학교 전자계산학과 학사 졸업.  
 2001년 한남대학교 정보산업대학원 정보통신학과 석사 졸업.  
 2007년 인하대학교 컴퓨터정보공학과 박사 졸업.

2008년~2009년 호남대학교 컴퓨터공학과 전임강사.  
 2010년 전북대학교 전기전자정보인력양성사업단 박사후과정.  
 2011년~현재 목포대학교 정보산업연구소 연구교수.

<주관심분야 : 정보검색, 데이터마이닝, 데이터베이스, 해양생물 IT정보융합>



**이 성 로**(정회원)  
 1987년 고려대학교 전자공학과 졸업  
 1990년 한국과학기술원 전기및 전자공학과 석사  
 1996년 한국과학기술원 전기및 전자공학과 박사

1997년 9월~현재 목포대학교 공과대학 정보전자공학과 교수  
 <주관심분야 : 디지털통신시스템, 이동 및 위성통신시스템, USN/텔레메틱스응용분야, 임베디드시스템>



**정 민 아**(정회원)  
 1992년 전남대학교 전산통계학사 졸업  
 1994년 전남대학교 전산통계 석사 졸업  
 2002년 전남대학교 전산통계 박사  
 2005년~현재 목포대학교 컴퓨터공학과 부교수

<주관심분야 : 데이터베이스/데이터마이닝, 생체인식시스템, 무선통신응용분야(RFID, USN, 텔레메틱스), 임베디드시스템>