

앙상블 학습을 이용한 적조 발생 예측의 성능 향상

박선, 강일우, *배진수, **김광순, 정민아, 이연우, 이성로

목포대, *세종대, **연세대

sunpark@mokpo.ac.kr, srlee@mokpo.ac.kr

Enhancing of Red Tide Blooms Prediction using Ensemble Train

Sun Park, Kang Il Woo, *Jinsoo Bae, **Kim Kwang Soon, Min A Jeong, Yeon Woo Lee, Seong Ro Lee

Mokpo National Univ., *Sejong Univ., **Yonsei Univ.

요약

적조란 유해조류의 일시적인 대 번식으로 바다를 적색으로 변화시키며 연안 환경 및 바다 생태계에 악영향뿐만 아니라 양식장의 어패류를 집단 폐사시키는 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 최소화 시킬 수 있다. 본 논문은 앙상블 학습을 이용한 적조 발생 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.

I. 서론

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 외편모조 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 변화해서 적조 피해가 국내의 수산업에 다시 발생할 지는 아무도 예측하지 못하고 있다. 국내에서는 60여 종의 적조 생물들이 있으며, 이중 수산업에 피해를 미치는 유해적조가 총 7종이 있고, 이들은 어패류를 치사시키는 적조생물, 패류를 독화시키거나 식독증을 일으키는 적조생물 등으로 구분할 수 있다. 우리나라 연안에서 출현하여 어류나 패류를 직접 치사시키는 독성물질을 생성하는 종은 *coccolodinium polykoides*, *heterosigma akashiwo*, *gymnodinium mikimotoi*, *gyrodinium sp* 등 4종이 있으며, 굴, 홍합 등 패류를 독화시켜 독화된 패류를 사람이 먹으면 마비성 패독증을 일으키는 *alexandrium tamarense*, *dinophysis acuminata*, *nitzschia pungens f. multiseriata* 등 3종이 있다. 이 들 중에서 국내에서는 특히 *coccolodinium polykoides*에 의한 양식어업의 피해가 가장 크다[1][2][3][4]. 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[5]. 특히 적조에 대한 연구 분야 중에서 적조 발생 예측연구는 예측 결과를 분석하여서 필요시 가두리 양식 장의 이

동 및 방제 준비를 통하여서 적조에 의한 양식장의 피해를 최소화 시킬 수 있을 것이다. 그러나 적조 발생의 예측에 대한 국내의 연구는 아직 미흡한 편에 있다[1][2][3][4].

국내에서 발생하는 적조는 서해, 남해안 등 발생 범위가 다양하고, 발생 해역으로부터 해류를 따라서 빠르게 이동하는 특성을 가지고 있다. 이러한 적조의 다양한 발생 범위와 빠른 확산 때문에 직접 적조를 탐지하여서 적조의 피해를 대비하는 것에는 한계가 있다[5]. 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[6][7]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, 적조 발생 전 10일 동안의 해양환경조건이나 기상조건이 우리나라 적조 발생의 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[5]의 연구에서는 국내에 가장 피해를 많이 준 유해적조인 *cochloclodium p.*의 발생은 주로 고수온과 저염분에 의해 지배되는 것으로 보이고, 적조 발생 전 집중 호우가 관측된 후 염분농도가 급격히 감소한 경우에 많이 발생하는 것을 보였다[2][3][4]. 이러한 적조와 환경인자의 관계에 대한 연구를 기반으로 본 논문에서는 적조 발생 10일전의 해양환경자료를 이용하여서 적조 발생 예측의 성능을 향상할 수 있는 방법을 제안한다. 제안방법은 통영지역에서 발생한 과거의 *cochloclodium p.*의 적조 발생 정보와 적조 예측일 이전 10일간의 해양환경자료에 앙상블 학습을 이용하여 적조 발생 예측의 성능을 향상시킨다. 본 논문에서는 앙상블 학습에 의사결정트리와 Adaboost 분류기를 이용하였다.

II. 본론

저자들의 이전 연구[2][3][4]에서는 신경망, SVM, 나이브베이스, 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 이용해서 적조의 발생을 예측한다. 본 논문에서는 앙상블 학습에 가장 많이 사용되는 의사결정트리를 기본 분류기로 사용하는 bagging과 AdaBoost 부스팅 방법을 이용한다.

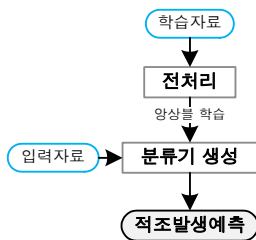


그림 1. 제안된 적조 발생 예측 블록도

Fig. 1. Block diagram of proposed red tide blooms prediction

앙상블 학습에 가장 일반적으로 사용되는 방법은 bagging과 boost가 있다[8][9]. 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.

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

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

본 논문에서는 bagging 앙상블 학습을 사용되는 기본 분류기로 의사결정 트리 방법 중에서 가장 많이 사용하는 필수 방법인 CART(classification and regression tree) 방법[10]을 사용한다. CART 방법은 전체 자료집합으로 부터 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 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 i번 자료가 분기되는 분류이다.

III. 결론

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

트리를 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 제안 방법을 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.

ACKNOWLEDGMENT

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

참고 문헌

- [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, “나이브베이스 분류자와 퍼지 추론을 이용한 적조 발생 예측의 성능향상”, 2011년도 한국통신학회 하계종합학술발표회, 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] J. Han, M. Kamber, “Second Edition, Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers, 2006.
- [9] 김명중, “유전자 알고리즘을 이용한 분류자 앙상블의 최적 선택”, 지능정보논문지, 제16권 제4호, pp. 99-112, 2010.
- [10] Breiman, L., J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Boca Raton, FL: CRC Press, 1984.와는 다를 수 있다.