

로지스틱 회귀모형과 의사결정나무 모형을 이용한 *Cochlodinium polykrikoides* 적조 탐지 기법 연구

박수호* · 김홍민* · 김범규* · 황도현* · 앵흐자리갈 운자야* · 윤홍주**

Study on Detection Technique for *Cochlodinium polykrikoides* Red tide using Logistic Regression Model and Decision Tree Model

Su-Ho Bak* · Heung-Min Kim* · Bum-Kyu Kim* · Do-Hyun Hwang* · Enkhjargal Unuzaya* · Hong-Joo Yoon**

요약

본 연구에서는 기계학습 기법의 한 갈래인 로지스틱 회귀모형과 의사결정나무 모형을 이용하여 인공위성 영상에서 *Cochlodinium polykrikoides* 적조 픽셀을 탐지하는 방법을 제안한다. 학습자료로 적조, 청수, 탁수 해역에서 추출된 수출광량 분광 프로파일(918개)을 활용하였다. 전체 데이터셋의 70%를 추출하여 모형 학습에 활용하였으며, 나머지 30%를 이용하여 모형의 분류 정확도를 평가하였다. 정확도 평가 결과 로지스틱 회귀모형은 약 97%의 분류 정확도를 보였으며, 의사결정나무 모형은 약 86%의 분류 정확도를 보였다.

ABSTRACT

This study propose a new method to detect *Cochlodinium polykrikoides* on satellite images using logistic regression and decision tree. We used spectral profiles(918) extracted from red tide, clear water and turbid water as training data. The 70% of the entire data set was extracted and used for model training, and the classification accuracy of the model was evaluated by using the remaining 30%. As a result of the accuracy evaluation, the logistic regression model showed about 97% classification accuracy, and the decision tree model showed about 86% classification accuracy.

키워드

Red Tide, Ocean Color Remote Sensing, COMS/GOCI, Decision Tree, Machine Learning
적조, 해수색 원격 탐사, 정지 궤도 해상 위성, 로지스틱 회귀 모형, 의사 결정 나무, 기계 학습

1. 서론

적조현상은 해양에서 식물 플랑크톤의 대량 증식으로 해수가 붉게 변색되는 현상이다. 최근 연구결과 해

수의 색을 변화시키지 않는 낮은 생물밀도에서 어패류를 독화시키는 현상이 보고됨에 따라 적조현상의 피해에 초점을 두어 해양에서 식물 플랑크톤 대량증식이 생물에게 물리적 피해를 야기하는 현상을 통틀

* 부경대학교 지구환경시스템과학부
(shbak91@pukyong.ac.kr, funwarm@naver.com, bumkyu1005@nate.com, rupine725@hanmail.net, unuzaya.e@gmail.com)

** 교신저자 : 부경대학교 지구환경시스템과학부

• 접수일 : 2018. 07. 09
• 수정완료일 : 2018. 07. 27
• 게재확정일 : 2018. 08. 15

• Received : Jul. 09, 2018, Revised : Jul. 27, 2018, Accepted : Aug. 15, 2018

• Corresponding Author : Hong-Joo Yoon

Division of Earth Environmental System Science Major of Spatial Information Engineering, Pukyong National University,

Email : yoonhj@pknu.ac.kr

어 유해적조(HAB; Harmful Algal Bloom)라 정의하고 있다[1].

우리나라의 경우 고려사, 조선왕조실록 등 다양한 과거문헌에 등장할 만큼 오래전부터 발생해온 해양현상으로 과학적인 보고로는 1967년이 최초이다[2]. 이후 1980년대까지는 규조류(Diatoms)에 의한 적조가 남해연안 일부 해역에서 일시적으로 발생하였으나[2], 1990년대에 들어 비교적 유해성이 높은 와편모조류(Dinoflagellates)에 의한 적조발생 비율이 급격히 증가하고 있다.

이러한 와편모조류는 편모조류(Phytoplankton)의 일종으로 두 개의 편모를 가지고 운동력을 지니고 있어 한번 대발생을 일으키면 저층의 높은 영양조건에서 지속적으로 영양분을 공급받아 장기간, 넓은 해역에서 적조현상을 지속시킬 수 있다[3]. 우리나라에서는 와편모조류의 일종인 *Cochlodinium polykrikoides* (이후 *C. polykrikoides*)가 최근 20여년 간 매년 여름과 가을철(7~10월)에 발생하고 있으며, 특히 1995년에는 약 750억원의 경제적 손실을 가져온 종이다. *C. polykrikoides*는 무독성 와편모조류로 알려져 있으나, 대발생을 일으켰을 경우 어류의 아가미에 달라붙어 질식사시킴으로써 양식업에 큰 경제적 손실을 입히고 있다. 심할 경우 동해와 서해까지 확산되어 우리나라 전역에 피해를 입힐 수 있는 종이다.

이미 많은 연구를 통해 적조원인생물의 환경생리적 특성들이 파악되었으나 적조현상 자체는 해양환경변화, 기상기후변동, 인간의 활동 등 다양한 요인들에 의해 영향을 받고 있어 적조의 발생을 예측하는 것은 상당히 어려운 일이다[4]. 따라서 이러한 적조의 특성상 적조발생의 사전예측보다는 조기탐지 및 지속적인 모니터링을 통한 신속대응이 효율적이다[4]. 그러나 기존의 적조 모니터링 방법은 주로 선박이나 항공기를 이용한 것으로 인력, 비용, 시간적 측면에서 비효율적일 뿐만 아니라 광범위한 해역에 발생하는 와편모조류에 의한 적조를 모니터링하기에는 제한사항이 많으므로 인공위성을 이용한 원격탐사가 효율적이다[5].

인공위성 원격탐사를 이용할 경우 광범위한 해역을 단시간에 탐지 및 모니터링 할 수 있으며, 적은 인력과 비용으로 시차적인 결과물을 얻을 수 있다는 장점이 있다. 또한 COMS(Communication, Ocean and

Meteorological Satellite)와 같은 정지궤도 위성을 이용할 경우 하루 8회 촬영을 통한 준 연속 모니터링이 가능하여 적조방제 능력을 향상시킬 수 있다.

초기의 인공위성 원격탐사를 이용한 적조 탐지 기법 연구는 주로 해색위성 영상이나 고해상도 육상관측 위성을 통해 추정된 Chlorophyll-*a* 농도와 Chlorophyll-*a* 농도 anomaly를 이용하여 시도되었다[6-9]. 이러한 방법은 적조가 발생하지 않은 일정기간의 Chlorophyll-*a* 농도 평균값을 기준으로 anomaly를 계산하여 특이값을 가지는 픽셀을 추출하는 기법으로 현재 NOAA(: National Oceanic and Atmospheric Administration)에서 공식적인 적조탐지 기법으로 사용되고 있다. 그러나 이 방법은 단순히 식물 플랑크톤의 증식으로 인한 Chlorophyll-*a* 농도 변화를 이용한 것으로 대발생을 일으킨 식물플랑크톤이 유해성인지 무해성인지 구분하지 못한다는 단점을 가지고 있다. 또한 대부분의 해색위성 산출물에 적용되는 Chlorophyll-*a* 산출 알고리즘들은 해수의 색이 식물플랑크톤에 의해서만 결정되는 맑은 해수(Case-1)에 중점을 두고 개발되어 용존 유기물 및 부유물질의 농도가 높은 우리나라 연안해역에서는 그 정확도가 낮다는 단점이 있다[10-12]. 우리나라 연안에서 발생한 적조는 용존 유기물 및 부유물질의 영향으로 짧은 파장의 가시광선에서 흡광이 증가하여 위성에서 추정된 Chlorophyll-*a* 농도가 과대추정된다[12].

최근에는 이러한 Chlorophyll-*a* 농도 기반 탐지 기법의 단점을 개선하고자 해수의 광학적 특성인 IOP(: Inherent Optical Properties)와 AOP(: Apparent Optical Properties)를 이용하여 시도하고 있다[4, 10, 13, 14]. IOP란 해수의 고유한 광특성으로 외부의 광학적 환경에 관계없이 일정한 값을 가지는 흡광계수(Absorption coefficient; a), 산란계수(Scattering coefficients; b), 감쇄계수(Attenuation coefficient; c) 등이 있다. AOP는 해수면의 광학적 환경 특성에 따라 그 값이 변할 수 있는 외형적 광특성으로 수출광량(Water leaving radiance; L_w), 원격반사도(Remote sensing reflectance; R_{rs}) 등이 있다. 이러한 광특성을 파악함으로써 해수 중의 물질 종류와 양을 추정할 수 있다[5]. 그러나 이 방법은 배경이 되는 해양의 가시광선 스펙트럼과 적조의 스펙트럼 형태 차이를 이용한 것으로 탐지 대상이 되는 식물 플랑크톤의 광학

적 특성을 미리알고 있다는 가정하에서 제한적으로 사용이 가능하다. 또한 용존 유기물 및 부유물질의 농도가 높은 해수(Case-2)의 광학적 특성은 식물 플랑크톤 외의 다른 물질(용존 유기물, 부유물질)의 농도에 따라서 크게 변화하기 때문에 실험실에서 얻어진 광학적 특성과 실제 관측되는 광학적 특성 간에 차이가 발생할 수 있다. 뿐만 아니라 충분히 많은 상황에서 얻어진 탐지 대상의 스펙트럼을 얻어낸다 하더라도 배경 해수의 본래 스펙트럼 형태에 따라 변색 패턴이 상이하기 때문에 모든 상황에 대처 가능한 탐지 식을 생산하는 것은 어렵다.

따라서 본 연구에서는 기계학습 기법을 적용하여 위성영상 내에서 적조 픽셀을 탐지하는 방법을 제안하고자 한다.

II. 자료 및 방법

2.1 연구자료

기계학습 모형을 학습시키기 위해 적조, 청수, 탁수 해역에서의 가시광선 및 근적외선 영역 수출광량 분광 프로파일을 활용하였다. 분광 프로파일은 해양위성 센터에서 제공하는 GOCI Level 1B 자료를 통해 생산하였다. 또한 적조해역에서의 분광프로파일을 얻기 위해 국립수산과학원의 적조속보 자료를 활용하였다(그림 1).

GOCI(Geostationary Ocean Color Imager)는 천리안 위성의 해양탐체로 공간해상도는 500m이며, 가시광선 영역의 6개(412, 443, 490, 555, 660, 680nm) 채널과 근적외 영역의 2개(745, 865nm) 채널을 가지고 있다(표 1). Level 1B 영상은 복사보정과 기하보정이 된 영상이며, 이를 GDPS(GOCI Data Processing System)를 이용하여 대기보정 하였다. 대기보정 후 산출되는 Level 2 자료 중 정규화 수출광량(Normalized Water-leaving radiance; $nLw(\lambda)$)을 사용하였다.

표 1. GOCI 밴드 구성
Table 1. Band composition of GOCI

Band	Centroid Wavelength (nm)	Bandwidth (nm)
1	412	20
2	443	20
3	490	20
4	555	20
5	660	20
6	680	10
7	745	20
8	865	40

국립수산과학원에서 제공하는 적조속보 자료는 적조 발생 시 하루에 한번 제공되며, 적조 발생 해역의 위치와 해당 위치에서의 원인생물, 생물밀도, 수온에 대한 정보를 포함하고 있다. 본 연구에서는 과거 적조 발생 위치 정보를 수집하기 위해 적조속보 자료에 포함된 적조발생 해역도를 활용하였다. 적조 발생해역도는 이미지 형태 자료로 정확한 위치정보를 포함하고 있지 않기 때문에 지오레퍼런싱(Georeferencing)을 통해 공간자료화하여 적조발생 해역의 경도, 위도를 추출하였다. 지오레퍼런싱은 좌표값이 부여되지 않은 항공사진 또는 이미지 형태의 지도에 실제계 좌표를 부여하는 방법으로 사진 또는 지도 상의 기준점에 실제 좌표를 입력함으로써 이루어지며, 그 외 모든 좌표는 이 기준점에 대한 상대적인 좌표값으로 입력된다.

2.2 기계학습 모형

기계학습 기법은 회귀분석이나 문턱치 기법(Threshold Method)과 같은 경험적 방법에 비해 비선형적인 모형 설계에 적합하다. 본 연구에서는 기계학습 기법 중 로지스틱 회귀모형(Logistic Regression Model), 의사결정나무 모형(Decision Tree Model)을 활용하여 *C. polykrikoides* 적조 탐지를 시도하였다.

로지스틱 회귀 모형은 일반적인 회귀모형과 마찬가지로 독립변수와 종속변수 사이의 관계를 함수로 나타내는 통계적 모형이다. 독립 변수의 선형 결합을 이용하여 종속 변수를 설명한다는 것은 일반적인 회귀모형과 비슷하지만 범주형 자료가 입력값으로 사용되었을 때, 그 결과가 특정 클래스에 속하게 될 확률로

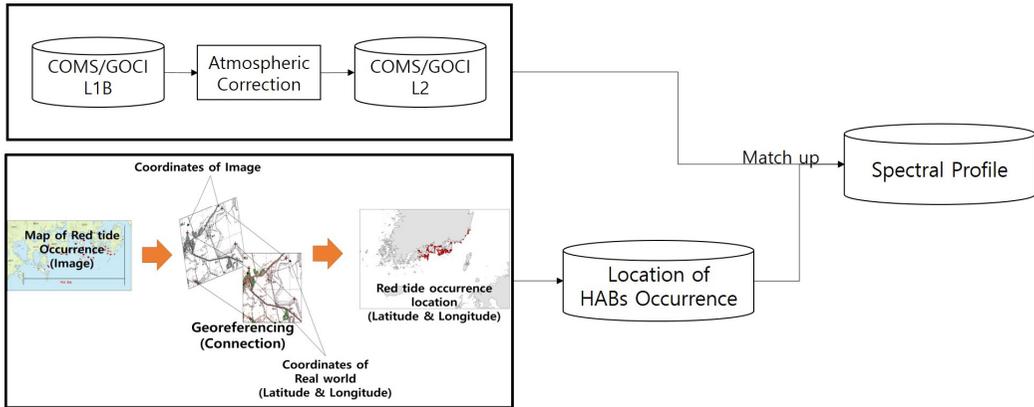


그림 1. 분광 프로파일 생산과정.
Fig. 1 Spectral profile production process.

주어진다는 데서 차이를 보인다. 또한 일반적인 회귀 모형에서는 종속변수의 범위가 $-\infty$ 에서 $+\infty$ 의 값을 가지는데 비해 로지스틱 회귀모형의 경우 S자 형태의 로지스틱 함수(Logistic Function)로 표현되어 0에서 1의 값만 가진다. 로지스틱 회귀모형의 회귀계수는 일반적인 선형회귀계수와 마찬가지로 종속변수와 독립변수들 사이의 관계를 설명하는데 사용된다[15]. 그러나 선형회귀모형의 경우 최소제곱법(Least Squares Method)을 이용하여 회귀계수를 추정하는데 비해 로지스틱 회귀모형의 경우 최대우도법(Maximum Likelihood Estimation Method)을 이용한다는 차이가 있다.

의사결정나무 모형은 의사결정규칙(Decision Rule)을 나무구조로 표현하여 분류(Classification) 또는 예측(Prediction)하는 방법으로 수행 과정이 나무구조에 의한 추론규칙으로 표현되므로 다른 기계학습 모형에 비해 연구자가 과정을 쉽게 이해하고 설명할 수 있다는 장점이 있다[16]. 의사결정나무의 구조는 마디(Node)로 구성되며, 뿌리마디(Root node)에서 시작하여 분리기준(Splitting criterion)에 의해 각 가지가 종단마디(Terminal node)에 이를 때까지 자식마디(Child node)를 생성해나간다.

이 때, 분리기준은 어떤 변수를 사용하며, 그 변수의 어떤 값을 문턱치로 하여 클래스를 분류할 것인지를 결정하는 기준이다. 일반적으로 많이 사용되는 분리기준 알고리즘에는 CHAID(Chi-squared

Automatic Interaction Detection), CART(Classification and Regression Tree), C4.5 등이 있다.

2.3 적조탐지 모형 생성

원격탐사 산출물 생산에 감독학습 기법을 적용하기 위해서는 위치정보 및 해당 위치에서의 속성값이 필요하다. 본 연구에서는 국립수산과학원에서 제공하는 적조발생 해역도를 지오레퍼런싱하여 얻어진 위도 및 경도값을 위치정보로 사용하였으며, 적조발생 해역의 GOCI Level 2 수출광량 분광 프로파일을 속성값으로 사용하였다.

분광 프로파일은 GOCI 영상으로부터 수집된 가시광선 6개 밴드와 근적외선 2개 밴드의 정규화 수출광량을 의미하며, 적조 발생해역, 청수 해역, 탁수 해역에서 각각 306개 픽셀을 무작위로 추출하였다.

로지스틱 회귀모형을 포함한 모든 회귀모형은 어떤 변수를 설명변수로 선택하느냐에 따라 모형의 성능에 큰 영향을 준다. 종속변수에 영향을 줄 수 있는 모든 설명변수를 알고 있다고 하더라도 종속변수의 변화를 설명해주기 위해 필요한 설명변수의 수를 어떻게 설정할 것인가는 매우 중요한 문제이다. 선택가능한 모든 설명변수를 이용하여 모형을 생성할 경우 데이터를 획득하고 분석하며 관리하는데 많은 노력과 시간이 요구될 뿐만 아니라 오히려 모형이 종속변수를 설명하는 능력을 저하시킬 수 있다. 더구나 설명변수의

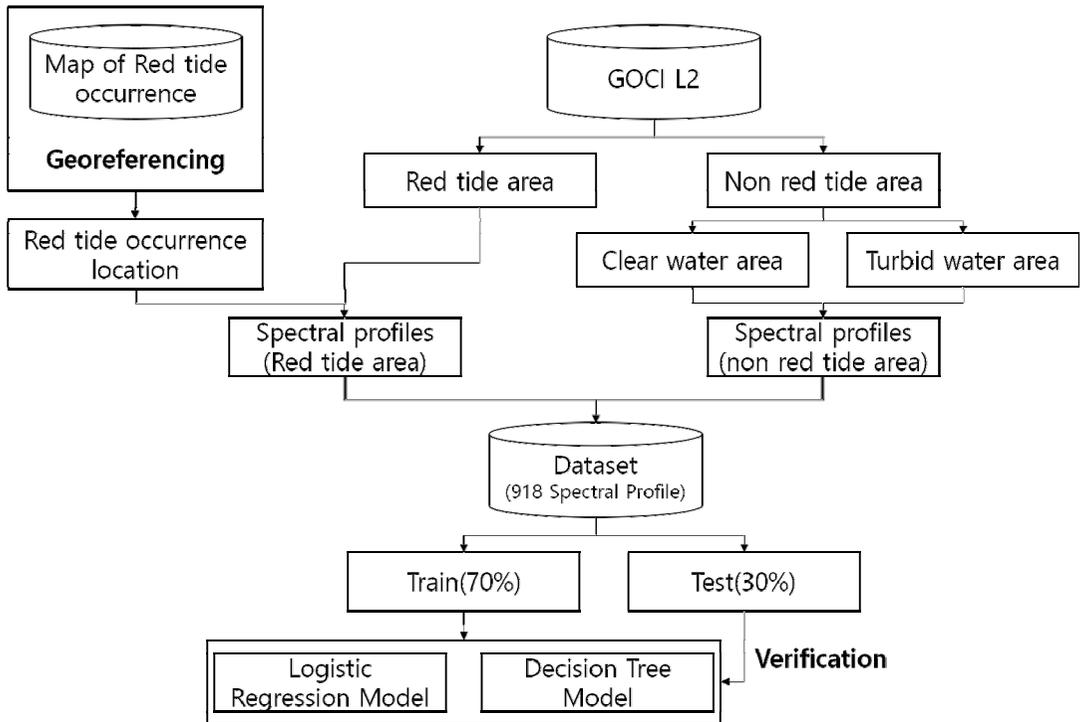


그림 2. 연구 흐름도.
Fig. 2 Flow chart of research

수가 많은 경우 설명변수들 사이에 강한 상관관계가 존재하는 다중공선성(Multicollinearity) 문제가 발생할 수도 있다. 따라서 모형 생성에 필요한 최소한의 설명 변수를 사용하는 변수조합을 찾아내는 것이 관건이다. 설명변수의 조합을 선택하는데 있어서 다양한 방법들이 연구 되어왔는데, 가장 많이 사용하는 방법은 후진 소거법, 전진선택법, 단계적 선택법이 있다.

후진소거법은 모든 설명변수를 포함한 완전모형으로부터 시작하여 일정한 유의수준을 만족시키지 못하는 변수를 찾아내어 순차적으로 제거하는 방법이다. 그러나 모형에서 한번 제거된 설명변수는 다시 모형에 포함될 수 없기 때문에 소거를 수행하는 과정에서 더 좋은 설명변수 조합이 있더라도 선택할 수 없다는 단점이 있다.

전진선택법은 설명변수가 전혀 선택되지 않은 상수항만 가진 회귀모형에서 설명력이 가장 우수한 변수를 모형에 추가하는 방법이다. 후진소거법과 함께 많

이 사용되는 선택법이나 모형에 한번 포함시킨 설명 변수는 도중에 탈락시킬 수 없기 때문에 유의하지 못한 변수가 포함될 가능성이 있다.

단계적 선택법은 후진소거법과 전진선택법의 절충적 방법으로 모형에 포함된 변수도 선택과정에서 언제든지 제거될 수 있다는 것이 특징이다. 변수 선택을 수행하는 각 과정마다 새로운 변수를 추가함과 동시에 이전 단계에서 이미 선택되었던 변수가 새로운 변수 추가로 인해 모형에 기여하는 바가 어떻게 변하는지 평가하여 중요성이 상실될 경우 제거한다.

본 연구에서는 로지스틱 회귀모형 학습을 위해 변수 선택법을 적용하여 입력변수를 축소시켰으며, 이때 단계적 선택법(Stepwise Method)을 사용하였다. 정보기준은 AIC(Akaike Information Criterion)와 SBIC(Schwarz-Bayesian Information Criterion)를 이용하였으며, 각각 만들어진 두 로지스틱 회귀모형의 성능을 평가하여 더 높은 모형을 최종모형으로 선택

하였다.

AIC와 SBIC는 회귀모형의 복잡도에 벌점을 주어 최적화시키는 방법으로 모형 M_k 의 AIC와 SBIC는 식 (1), (2)와 같이 정의할 수 있다[17].

$$AIC = -2 \sum_{i=1}^n \frac{l(y_i, x_i^T \hat{\beta})}{n} + \frac{2k}{n} \quad (1)$$

$$SBIC = -2 \sum_{i=1}^n \frac{l(y_i, x_i^T \hat{\beta})}{n} + \frac{k \log(n)}{n} \quad (2)$$

이 때, $\sum_{i=1}^n l(y_i, x_i^T \hat{\beta})$ 는 로그 우도함수이며, $\hat{\beta}$ 는 β 에 대한 최대우도 추정량, k 는 추정된 모수의 수를 나타낸다. 일반적으로 BIC는 AIC에 비해 더 적은 수의 설명변수 조합을 선택하는 경향이 있다.

의사결정나무 모형은 범주형 자료 학습이 가능한 CART(: Classification and Regression Tree) 모형을 사용하였으며, 가지치기(Pruning)를 통해 과적합을 방지하였다.

전체 데이터셋(918개 분광프로파일)의 70%를 학습 자료로 사용하였고, 나머지 30%를 검증자료 및 모형 평가 자료로 사용하였다(그림 2).

2.4 정확도 평가

본 연구에서는 분류 정확도를 평가하기 위해 오차행렬(Confusion Matrix, 표 2)를 이용한 전체 정확도를 평가측도로 사용하였다.

표 2. 오차행렬
Table 2. Confusion matrix

		Prediction	
		True	False
Reference	True	TP	FN
	False	FP	TN

* TP : True Positive / FP : False Positive
FN : False Negative / TN : True Negative

이 때, 전체 정확도(Accuracy)는 식 (3)과 같이 정의된다.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

III. 결과 및 토의

3.1 모형 생성 결과

로지스틱 회귀모형을 생성하기 위해 단계적 선택법을 적용한 결과 AIC와 SBIC 모두 같은 변수 조합을 선택하였으며, 이 방법을 통해 선택된 설명변수는 $nLw(412)$, $nLw(443)$, $nLw(490)$, $nLw(555)$, $nLw(680)$, $nLw(745)$ 로 총 6개 파장의 정규화 수출광량이였다. 해수에서 식물플랑크톤의 개체 수 증가로 Chlorophyll-a 농도가 증가할수록 440~450nm 파장대와 670~680nm에서 흡수가 일어나며, 560~570nm 파장대에서는 반사 670~680nm 파장대에서는 형광

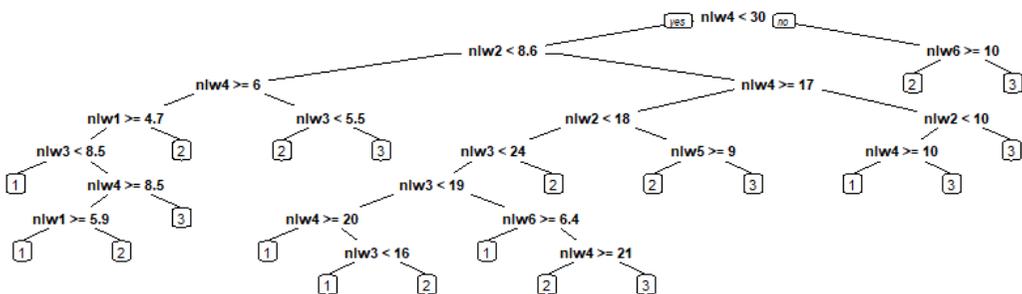


그림 3. 의사결정나무 모형 생성 결과(1: 적조, 2: 탁수, 3: 청수).

Fig. 3 Result of the decision tree model construction(1: Red tide, 2: Turbid water, 3: Clear water)

이 일어난다[10-12]. 이로 인해 적조 발생한 해역에서는 높은 Chlorophyll-*a* 농도로 인해 555nm와 680nm 파장에서 피크(Peak)를 보인다[10]. 변수 선택법을 적용하여 얻어진 6개 파장 중 *nLw*(490), *nLw*(555), *nLw*(680), *nLw*(745)는 이러한 적조 발생 해역에서 관찰되는 피크와 관련된 것들로 선행연구들이 제시한 알고리즘에도 많이 사용되었다[4, 10, 12, 13].

최종적으로 선택된 6개 변수를 이용하여 *C. polykrikoides* 적조 탐지를 위한 로지스틱 회귀모형은 식 (4)와 같다.

$$R = \frac{e^L}{1 + e^L} \quad (4)$$

$$L = -4.411213x_1 + 1.769347x_2 + 2.203324x_3 - 1.758237x_4 - 1.297037x_5 + 20.22928x_6 + 11.62793$$

이 때, x_i 는 각각 *nLw*(412), *nLw*(443), *nLw*(490), *nLw*(555), *nLw*(680), *nLw*(745)을 의미한다.

한편, 의사결정나무 모형은 6개의 변수를 이용하여 적조, 청수, 탁수 해역을 분류하였으며, 이 때 선택된 변수는 *nLw*(412), *nLw*(443), *nLw*(490), *nLw*(555), *nLw*(660), *nLw*(680)이다(그림 3). 로지스틱 회귀모형의 경우 5개의 가시광선 파장과 1개의 근적외선 파장을 선택한 것에 비해 의사결정나무 모형은 6개 모두 가시광선 영역의 파장을 선택하였다.

3.2 정확도 평가 결과

검증용 데이터셋을 이용하여 두 모형의 성능을 평가한 결과 로지스틱 회귀모형은 약 98%, 의사결정나무 모형은 약 86%의 분류 정확도를 나타내었다(표 3과 4). 이는 가장 최근에 제안된 GOCI 적조탐지 알고리즘[10]의 탐지 정확도(52.76%)와 비교해보았을 때, 두 모형 모두 높은 정확도 개선이 있음을 확인할 수 있었다.

로지스틱 회귀모형과 의사결정나무 모형 간의 탐지 특성을 비교해보았을 때, 의사결정나무 모형은 기계학습 기법을 적용하지 않은 선행연구에서도 비교적 쉽게 분류하는 청수와 적조의 분류에서 일부 오류를 보였다. 일반적으로 청수와 적조는 490~555nm 구간의

파장대에서 정반대의 기울기를 가지고 있어 해수의 광학적 특성을 이용한 선행연구들의 알고리즘으로 분리가 가능하다[4, 10, 11, 13].

표 3. 로지스틱 회귀모형의 오차행렬
Table 3. Confusion matrix of logistic regression model

		Detection		
		R	T	C
Reference	R	100	5	0
	T	1	87	0
	C	0	0	82

* R: Red tide / T: Turbid water / C: Clear water
* Total Accuracy=0.98

표 4. 의사결정나무 모형의 오차행렬
Table 4. Confusion matrix of decision tree model

		Detection		
		R	T	C
Reference	R	92	9	2
	T	6	74	9
	C	3	9	71

* R: Red tide / T: Turbid water / C: Clear water
* Total Accuracy=0.86

로지스틱 회귀모형과 선행연구의 모형의 경우 탁수와 적조의 분류에서만 오류를 가지고 있는 것에 비해 의사결정나무 모형은 탁수 뿐만 아니라 청수와 적조 간의 분류에서도 오류를 가지고 있어 결과적으로 모든 유형의 해수에서 적조로 오분류할 수 있는 가능성을 가지고 있다고 볼 수 있다. 이는 학습 데이터셋의 자료 특성이 의사결정나무 모형보다 로지스틱 회귀모형에 더 적합하여 발생한 것으로 판단된다.

의사결정나무 모형은 입력된 변수가 범주형인 경우에는 분류능력이 우수하나, 연속형인 경우 비연속적인 값으로 인식하기 때문에 분리 경계의 주변값에서 오류가 커지는 현상이 발생할 수 있다. 의사결정나무 모형은 학습 과정에서 각 노드의 분리 시 사용된 변수가 연속형인 경우 두 클래스를 나누기 위한 경계값(Threshold)을 설정하게 된다. 이 때, 기준으로 CART의 경우 두 클래스 간의 분산을 최대화하는 값을 사용하기 때문에 클래스 간 분산이 큰 변수를 사용할 경우에는 문제가 없으나, 클래스 간의 분산이 작

은 변수를 사용할 경우 경계값 주변에서 잘못 분류된 사례들이 다수 포함될 수 있다.

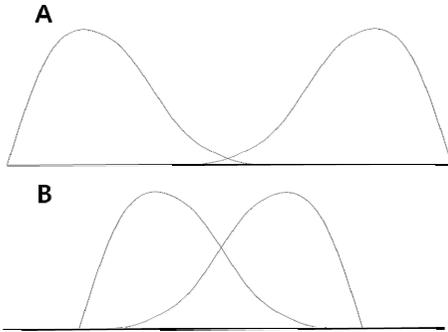


그림 4. 두 클래스 간의 자료 분포 유형.
Fig. 4 The type of distribution of data between two classes.

다시 말해서 기준이 되는 변수 내에서 클래스 간의 분포를 보았을 때, 두 클래스 사이에 공유하는 면적이 작은 경우(그림 4-A) 두 클래스 간의 분산을 최대화하는 경계값을 분류기준으로 사용해도 문제가 없으나, 두 클래스 사이에 공유하는 면적이 큰 경우 경계값 주변에서 오분류 될 확률이 매우 높아진다(그림 4-B).

정규화 수출광량의 경우 각 클래스(적조, 청수, 탁수) 사이에 공유되는 면적이 넓은 사례에 해당한다. 555nm 파장의 정규화 수출광량의 사례를 보았을 때, 적조의 경우 6.02에서 30.20까지 분포하며, 탁수의 경우 4.83에서 58.24까지, 청수의 경우 3.75에서 48.30까지 분포하고 있었다. 이 때, 적조는 자료의 분포 상 청수와 탁수의 부분집합에 해당하여 분류 경계에서 많은 오류를 일으킨 것으로 판단된다.

이러한 문제를 해결하기 위해 학습 데이터셋의 변수 목록을 수정하였다. 기존에 사용된 8개 밴드의 정규화 수출광량에 밴드 간 비율을 추가로 입력하여 의사결정나무 모형을 생성하였다. 밴드 간 비율은 클래스 간에 존재하는 파장대별 기울기 경향 차이를 의미하는 값으로 1을 경계로 아주 큰 분산을 가진다.

수정된 데이터셋을 이용하여 의사결정나무 모형을 생성한 결과 전체 정확도가 기존 약 86%에서 약 97%로 개선되었으며, 청수와 적조 사이에서 분류 오류를 일으켰던 것 또한 개선되어 로지스틱 회귀모형에 준하는 분류 정확도를 보였다(표 5).

표 4. 수정된 데이터셋을 이용하여 학습시킨 의사결정나무 모형의 오차행렬

Table 4. Confusion matrix of decision tree model using modified dataset

		Detection		
		R	T	C
Reference	R	91	5	0
	T	3	84	0
	C	0	0	92

* R: Red tide / T: Turbid water / C: Clear water

* Total Accuracy=0.87

IV. 결 론

본 연구에서는 로지스틱 회귀모형과 의사결정 나무 모형을 활용하여 위성영상에서 적조 픽셀과 일반해수 픽셀을 분류하였다. 두 모형의 성능을 평가해본 결과 선행연구에서 탐지한 결과에 비해 약 30~40%의 정확도 향상이 있음을 확인하였다.

두 모형 모두 광학적 특성을 기반으로 한 기준 모형에 비해 높은 정확도를 보였으나 의사결정나무 모형의 경우 청수와 적조 간의 분류에서 오류를 가지고 있었다. 그에 비해 로지스틱 회귀모형은 청수와 적조 픽셀 사이에서는 오분류가 없었으며, 일부 적조 픽셀과 탁수 픽셀을 혼동하는 것을 제외하면 높은 분류 정확도를 보였다. 이러한 결과는 의사결정나무 모형의 분류 방식에 기인한 것으로 판단된다. 의사결정나무 모형은 연속형 변수를 입력할 경우 변수가 가지는 값을 비연속적인 값으로 인식하여 분류하므로, 클래스 간 분산 차이가 적은 경우 경계값 부근에서 높은 오류를 보일 수 있다. 본 연구에서 사용한 정규화 수출광량의 경우 각 클래스 간에 공유하는 범위가 넓으며, 특히 적조의 경우 청수와 탁수의 부분집합이 되어 로지스틱 회귀모형에 비해 높은 분류 오류를 지닌 것으로 판단된다. 이를 해결하기 위해서는 정규화 수출광량에 비해 클래스간 분산 차이가 큰 변수를 사용해야 하며, 본 연구에서는 밴드 간 비율을 추가로 사용하여 모형을 생성하였다. 그 결과 기존 모형에 비해 정확도 향상을 확인할 수 있었으며, 청수와 적조 픽셀 간의 분류 오류도 개선되었다.

감사의 글

위 논문은 “2018년 한국전자통신학회 봄철학술대회 우수논문”입니다.

본 논문은 2018년 해양수산부 재원으로 한국해양과학기술진흥원의 지원을 받아 수행된 연구임 (“적조피해 최소화를 위한 적조탐지·예측 시스템 구축 및 실증화”, PM60650)

References

- [1] D. Anderson, P. Anderson, V. Bricej, J. Cullen, and J. Rensel, *Monitoring and Management Strategies for Harmful Algal Blooms in Coastal Waters*. Paris: Intergovernmental Oceanographic Commission Technical Series, 2001.
- [2] H. Kim, *Harmful Algal Blooms in the Sea*. Busan: Dasom, 2005.
- [3] Y. Yoon, *Sea rebellion*. Red tide, Paju: Jipmoondang, 2012.
- [4] Y. Kim, Y. Byun, Y. Huh, and Y. Yu, “Detection of *Cochlodinium polykrikoides* Red Tide Using MODIS Level 2 Data in Coastal Waters,” *Korean Society of Civil Engineers Journal of Civil Engineering*, vol. 27, no. 4D, 2007, pp. 535-540.
- [5] Y. Ahn, J. Moon, W. Seo, and H. Yoon, “Inherent Optical Properties of Red Tide Algal for Ocean Color Remote Sensing Application,” *J. of the Korean Society for Marine Environmental Engineering*, vol. 12, no. 1, 2009, pp. 47-54.
- [6] R. Stumpf, M. Culver, P. Tester, M. Tomlinson, G. Kirkpatrick, B. Pederson, E. Truby, V. Ransibrahmanakul, and M. Soracco, “Monitoring *Karenia brevis* blooms in the Gulf of Mexico using satellite ocean color imagery and other data,” *Harmful Algae*, vol. 2, no. 2, 2003, pp. 147-160.
- [7] M. Tomlinson, R. Stumpf, V. Ransibrahmanakul, E. Truby, G. Kirkpatrick, B. Pederson, G. Vargo, and C. Heil, “Evaluation of the use of SeaWiFS imagery for detecting *Karenia brevis* harmful algal blooms in the eastern Gulf of Mexico,” *Remote Sensing of Environment*, vol. 91, no. 3, 2004, pp. 293-303.
- [8] Y. Suh, L. Jang, N. Lee, and J. Ishizaka, “Feasibility of Red Tide Detection Around Korean Waters Using Satellite Remote Sensing,” *J. of Fisheries Science and Technology*, vol. 7, no. 3, 2004, pp. 148-162.
- [9] J. Ishizaka, Y. Kitaura, Y. Touke, H. Sasaki, A. Tanaka, H. Murakami, T. Suzuki, K. Matsuoka, and H. Nakata, “Satellite Detection of Red Tide in Ariake Sound, 1998-2001,” *J. of Oceanography*, vol. 62, no. 1, 2006, pp. 37-45.
- [10] Y. Son, Y. Kang, and J. Ryu, “Monitoring Red Tide in South Sea of Korea(SSK) Using the Geostationary Ocean Color Imager(GOCI),” *Korean J. of Remote Sensing*, vol. 26, no. 5, 2012, pp. 531-548.
- [11] Y. Ahn and P. Shanmugam, “Detecting the red tide algal bloom from satellite ocean color observations in optically complex Northeast-Asia Coastal waters,” *Remote Sensing of Environment*, vol. 103, no. 4, 2006, pp. 419-437.
- [12] Y. Son, J. Ishizaka, J. Jeong, H. Kim, and T. Lee, “*Cochlodinium polykrikoides* red tide detection in the South Sea of Korea using spectral classification of MODIS data,” *Ocean Science Journal*, vol. 46, no. 4, 2011, pp. 239-263.
- [13] S. Bak, H. Kim, D. H. Hwang, H. Yoon, and W. Seo, “Detection technique of Red tide Using GOCI Level 2 Data,” *Korean J. Remote Sensing*, vol. 32, no. 6, 2016, pp. 673-679.
- [14] S. Bak, H. Kim, D. Hwang, S. Oh, and H. Yoon, “Red Tide Detection Technique by Using Multi-temporal GOCI Level 2 Data,” *Internation J. of Grid and Distributed Computing*, vol. 10, no. 10, 2017, pp. 45-56.
- [15] B. Chae, W. Kim, Y. Cho, K. Kim, C. Lee, and Y. Choi, “Development of a Logistic Regression Model for Probabilistic Prediction of Debris Flow,” *The J. of Engineering Geology*,

vol. 14, no. 2, 2014, pp. 211-222.

- [16] Y. Song and B. Chae, "Development to Prediction Techique of Slope Hazards in Geneiss Area using Decision Tree Model," *The J. of Engineering Geology*, vol. 18, no. 1, 2008, pp. 45-54.
- [17] C. Park, Y. Kim, J. Kim, J. Song, and H. Choi, *Data-mining using R*, Seoul: Kyowoo, 2013.

저자 소개



박수호(Su-Ho Bak)

2013년 부경대학교 공간정보시스템
공학과 졸업(공학사)
2017년 부경대학교 공간정보시스템
공학과 졸업(공학석사)

2018년 현재 부경대학교 대학원 지구환경시스템과학부(박사과정)

※ 관심분야 : 해양 원격탐사, GIS



김흥민(Heung-Min Kim)

2013년 부경대학교 공간정보시스템
공학과 졸업(공학사)
2017년 부경대학교 공간정보시스템
공학과 졸업(공학석사)

2018년 현재 부경대학교 대학원 지구환경시스템과학부 공간정보시스템공학전공(박사과정)

※ 관심분야 : 해양 원격탐사, GIS



황도현(Do-Hyun Hwang)

2011년 부경대학교 공간정보시스템
공학과 졸업(공학사)
2013년 부경대학교 대학원 공간정
보시스템공학과 졸업(공학석사)

2018년 현재 부경대학교 대학원 지구환경시스템과학부 공간정보시스템공학전공(박사수료)

※ 관심분야 : 해양원격탐사, GIS



김범규(Bum-Kyu Kim)

2015년 부경대학교 공간정보시스템 공학과 졸업(공학사)

2018년 현재 부경대학교 대학원 지구환경시스템과학부 공간정보시스템공학전공(석사과정)

※ 관심분야 : 해양 원격탐사



엔흐자리갈 운자야

(Enkgjargal Unuzaya)

2014년 몽골 과학기술대학교 정보 및 전기통신기술학과 졸업(공학사)

2018년 현재 부경대학교 대학원 지구환경시스템과학부(석사과정)

※ 관심분야 : 무선 통신, 해양원격탐사, GIS



윤홍주(Hong-Joo Yoon)

1983년 부경대학교 해양공학과 졸업(공학사)

1985년 부경대학교 대학원 해양학과 졸업(공학석사)

1997년 프랑스 그르노블 I 대학교 대학원 위성원격탐사전공 졸업(공학박사)

1999년~2002년 여수대학교 해양공학과 교수

2002년~현재 부경대학교 공간정보시스템공학 교수

2012년~2013년 부경대학교 공간정보연구소 초대 소장

2013년 (사)한국클라우드센터터파크 이사

2014년 한국전자통신학회 부회장

2015년 공간정보 Big Data 센터장

2015년 행정공간정보화연구소 소장

2016년 (사)한국생태공학회 회장

※ 관심분야 : 해양 원격탐사, GIS