

# 빅 데이터 처리에 의한 유해 적조 분석

김준한, 이동현, 강민석, 윤홍원\*  
\*신라대학교 컴퓨터소프트웨어공학부  
e-mail : hwyun@silla.ac.kr

## Analysis of Harmful Algal Blooms by Big Data Processing

Junhan Kim, Donghyeon Lee, Minseok Kang, Hongwon Yun\*  
\*Department of Computer Software Engineering, Silla University

### 요 약

본 연구에서는 빅데이터 처리 기법을 적용하여 유해 적조의 원인과 영향력을 분석한다. 국립수산물과학원에서 2000년부터 2017년까지 지난 17년간의 정선해양관측 데이터와 유해 적조 데이터를 수집하였다. 수집한 데이터셋을 정제하고 기술통계와 로지스틱 다중 회귀 분석을 통하여 7개 관측 요소의 영향력을 비교 분석한다. 그리고 로지스틱 다중 회귀 분석을 수행하는 GUI를 구현한다.

### 1. 서론

적조(red tide)는 식물 플랑크톤의 대량 번식으로 바닷물의 색깔이 적색, 황색, 적갈색 등으로 변하는 자연 현상을 말한다. 최근에는 기후 변화와 관련하여 고수온, 냉수대, 저염분수 등과 같이 해양 환경이 악화됨으로써 소규모의 적조 발생에도 양식 생물이 폐사하는 막대한 피해를 입히고 있다. 피해의 규모는 해마다 최소 수억에서 많을 때는 수백억에 이르고 있다[1].

적조는 해류와 조류의 소통이 원활하지 않은 해안에서 일조량, 수온, 염분, 영양염류 등 적조 생물의 대량 번식에 알맞은 해양 환경이 조성되면 발생하는 것으로 알려져 있다[2]. 적조를 일으키는 해양 환경의 요소는 일반적으로 알려져 있지만 각 요소가 적조에 미치는 영향력을 분석한 연구는 발견하기 어렵다[3].

우리는 수산 피해를 일으키고 있는 유해 적조의 원인과 영향력을 분석하기 위하여 빅 데이터 처리 방법을 활용하기로 하고 공공 데이터를 수집한다. 본 연구에서는 통계적 방법으로 유해 적조 데이터의 특성을 분석하고 로지스틱 회귀 분석을 통하여 유해 적조에 영향을 미치는 요인을 분석하고 평가한다.

### 2. 자료처리 및 분석방법

국립수산물과학원에서 2000년 1월 1일부터 2017년 12월 31일까지 동해, 서해, 남해, 동중국해에서 관측한 정선 해양 관측 데이터를 내려 받았다[4]. 정선 해양 관측 데이터에는 수온, 염분, 용존산소, 인산염인, 아질산질소, 질산질소, 규산규소의 측정값이 들어 있다. 과거의 유해 적조 데이터는 정형 데이터로 제공되지 않아서 웹 크롤링으로 수

집하였다. 국립수산물과학원에서 한국의 연안과 근해에서 해양 자료를 관측하고 있는 위치는 그림 1에 보이고 있다.

빅 데이터 처리 과정은 데이터의 수집, 정제, 분석과 시각화, 그리고 저장으로 나눌 수 있다[5]. 우리는 데이터의 수집에서 시각화까지는 파이썬 프로그래밍을 사용하고 정제한 데이터는 CSV 파일로 제공하며 데이터베이스에 저장한다.

#### 데이터 수집

데이터의 수집을 위하여 국립수산물과학원에서 정선 해양 관측 데이터(이하, 해양 관측 데이터)를 내려 받았다. 과거의 적조 데이터는 내려 받기를 제공하지 않아서 웹 크롤링으로 수집 하였다.

#### 데이터 정제

해양 관측 데이터에는 비어 있는 값이 많았으며 신뢰할 수 없는 숫자 데이터가 입력되어 있기도 하였다. 신뢰도가 낮은 데이터는 제거 하였다. 분리되어 있는 해양 관측 데이터와 적조 발생 데이터를 병합하였으며 널 값이 있는 레코드는 삭제하였다.

#### 데이터 분석 및 시각화

데이터 분석 과정에서는 파이썬의 데이터 분석용 라이브러리인 팬다스(pandas)를 활용한다. 기술 통계적 방법으로 데이터의 특성을 분석하고, 로지스틱 회귀를 적용하여 적조에 영향을 미치는 요인을 분석한다. 파이썬의 시각화 패키지를 이용하여 통계적 분석 결과를 시각화 한다.

#### 데이터 저장

한국 해역의 해양 관측 데이터와 과거 유해 적조 데이터를 병합한 유해 적조 데이터셋은 CSV 형식의 플랫폼 파일로 공개한다. PyQt를 이용하여 유해 적조 데이터셋을

로지스틱 회귀 분석할 수 있는 GUI를 개발한다. 다양한 검색이 가능하도록 데이터베이스 기반의 애플리케이션을 개발한다.

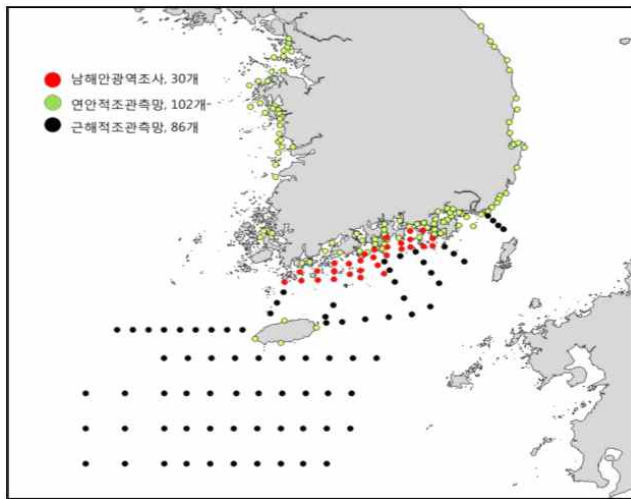


그림 1. 국립수산물과학원 적조 조사 정점도

### 3. 분석 결과 및 평가

우리는 2000년부터 2017년까지 지난 17년간의 유해 적조가 발생한 날짜의 해양 환경 데이터(수온, 염분, 용존산소, 인산염, 아질산질소, 질산질소, 규산규소)를 서로 비교 분석한다. 표 1은 본 연구가 분석할 데이터의 범위를 나타낸다.

표 1. 분석할 데이터의 범위

기간	해역	수심
2000.1.1.-2017.12.31	동해, 남해, 서해, 동중국해	0

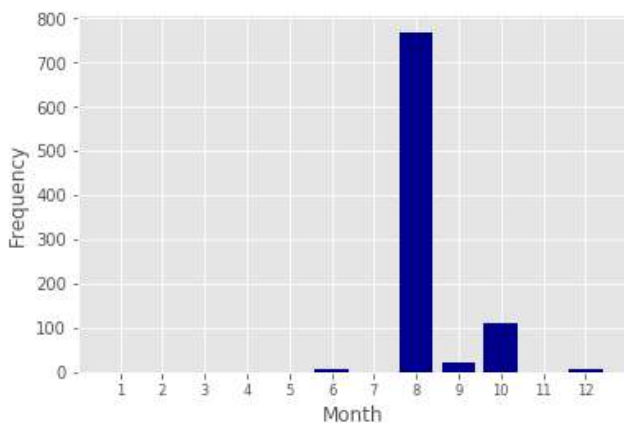


그림 2. 월별 유해 적조 발생 횟수

그림 2는 2000년부터 2017년 사이의 월별 누적 유해 적조 발생 횟수를 나타낸다. 유해 적조가 8월에 집중적으로 발생함을 알 수 있다. 6월에 적조가 처음 발생하며 10월에도 100여건의 적조가 발생하였다.

연중 8월에 유해 적조가 가장 많이 발생하므로 연도별로 8월의 평균 수온을 분석한다. 그 결과는 그림3과 같다. 그림 3의 연도별 유해 적조 발생 횟수와 그림 4의 연도별 해역의 수온 변화를 비교하면, 한국의 전 해역에서 수온이 낮았던 2011년에는 유해 적조 발생 횟수가 현저히 감소하였음을 알 수 있다. 아래의 그림 4부터 그림 10까지의 선 그래프에는 선이 끊기거나 나타나지 않는 기간이 있는데 이것은 관측 데이터가 없기 때문이다.

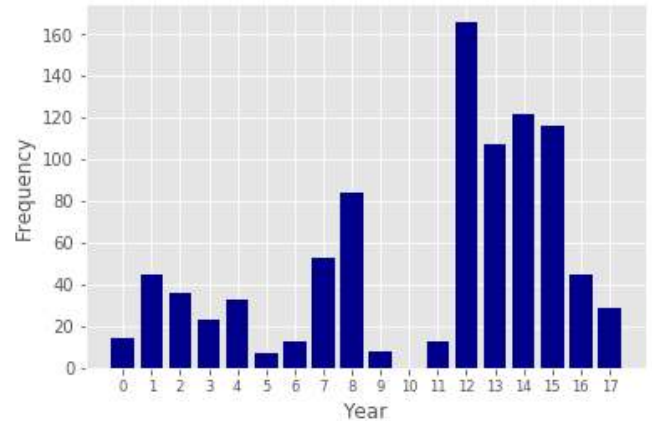


그림 3. 연도별 유해 적조 발생 횟수

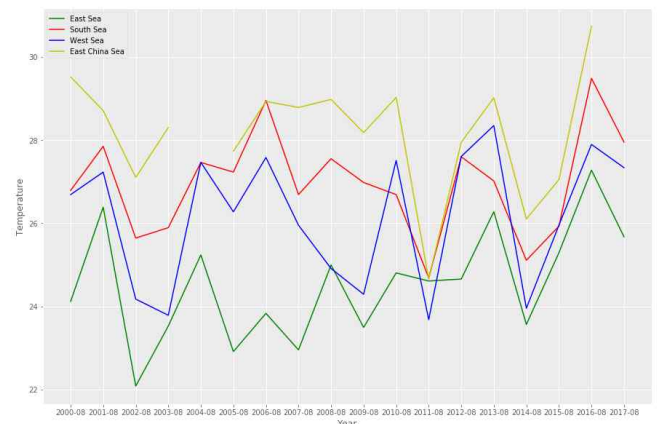


그림 4. 연도별 해역의 수온 변화

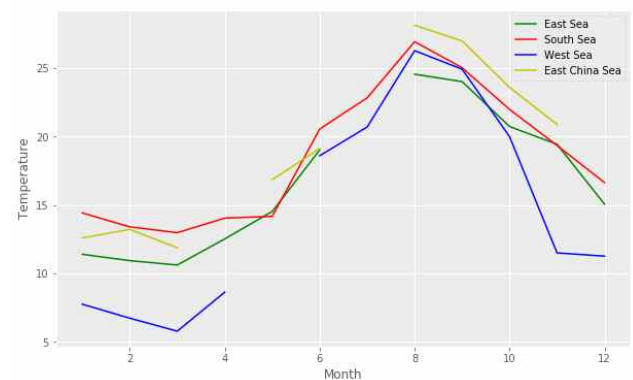


그림 5. 월별 해역의 수온 변화

그림 5는 월별 한국 해역의 수온 변화를 보이고 있다 유해 적조가 가장 많이 발생하는 8월에는 전 해역에 걸쳐

서 수온이 상승하였고 동중국해의 수온이 가장 높게 나타난다. 월별 염분의 변화는 그림 6에 나타나 있다. 남해의 염분은 5월부터 급격하게 감소하여 9월까지 지속적으로 감소하고 있다.

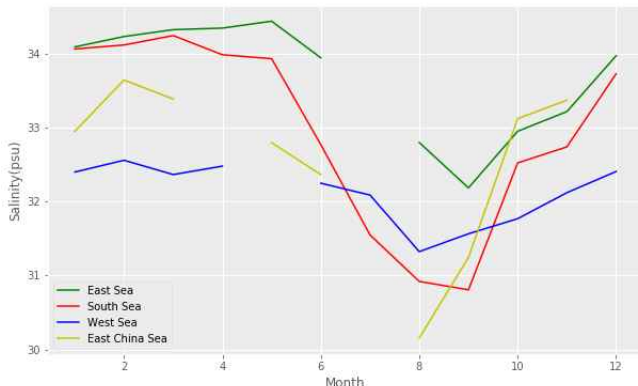


그림 6. 월별 해역의 염분 변화

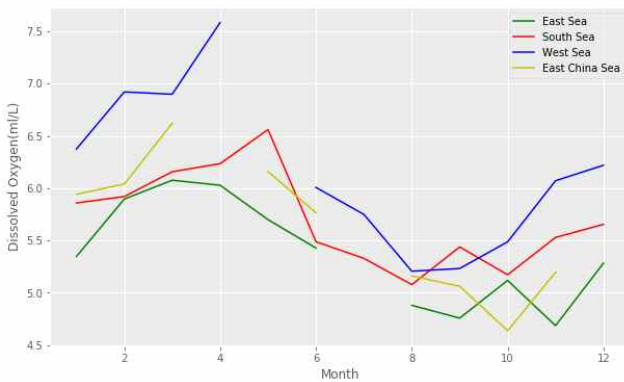


그림 7. 월별 해역의 용존산소 변화

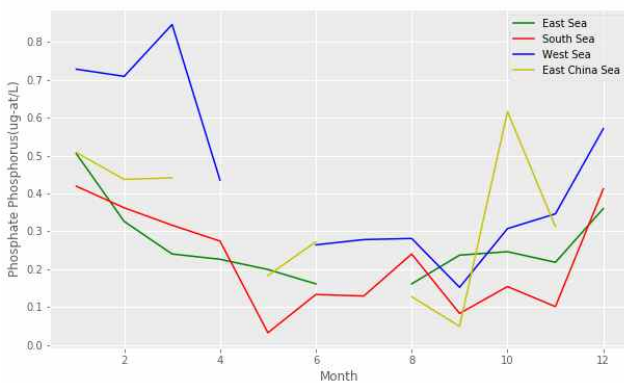


그림 8. 월별 해역의 인산염인 변화

그림 7은 한국 해역의 용존산소의 변화를 월별로 나타내고 있다. 용존산소의 농도는 8월에 모든 한국 해역에서 감소하지만 농도의 차이가 크지 않고 동해가 오히려 가장 낮게 나타난다. 8월에 유해 적조가 가장 많이 발생하는 남해는 9월에 용존산소의 농도가 높게 나타난다.

그림 8은 인산염인의 월별 변화를 나타낸다. 4월부터 9월까지 인산염인의 농도는 전체적으로 낮게 나타나고 있다. 남해는 5월부터 8월까지 인산염인의 농도가 조금씩 상

승하고 있다. 월별 아질산 질소의 농도는 그림 9에 나타난다. 6월에서 9월 사이에 남해와 동중국해의 아질산 질소의 농도가 낮게 나타나고 있다.

그림 10에서 알 수 있는 것처럼 질산질소의 농도는 6월에서 8월 사이에 모든 해역에서 비슷한 수준으로 낮게 나타나다가 8월에는 전 해역이 모두 낮게 나타난다. 그림 11은 각 해역의 월별 규산규소의 변화를 나타낸다. 유해 적조가 주로 발생하는 남해의 월 변화에서 특이함을 발견할 수 있다. 남해의 규산규소의 농도가 1월부터 차츰 감소하다가 5월부터 상승하여 7월에 다른 해역보다 높게 나타난다. 9월에는 전 해역에서 규산규소의 농도가 낮게 나타나고 있다.

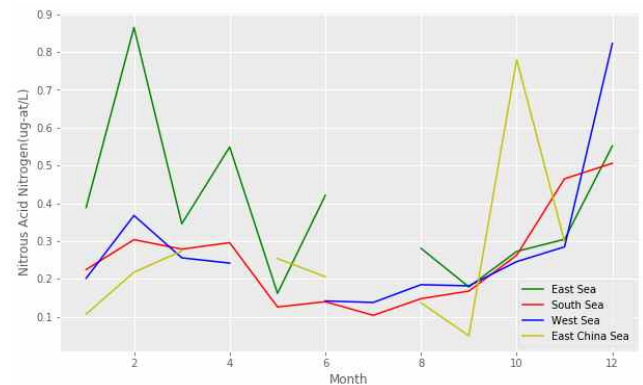


그림 9. 월별 해역의 아질산질소 변화

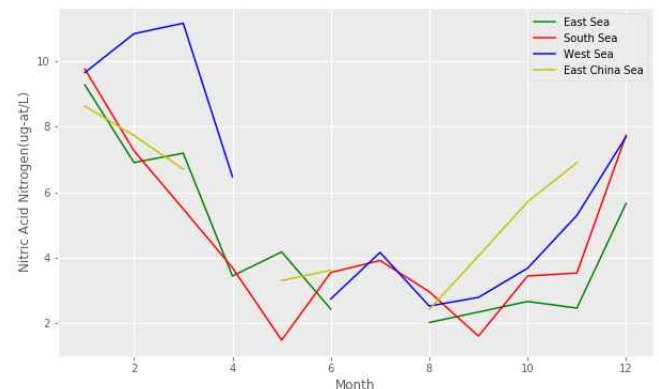


그림 10. 월별 해역의 질산질소 변화

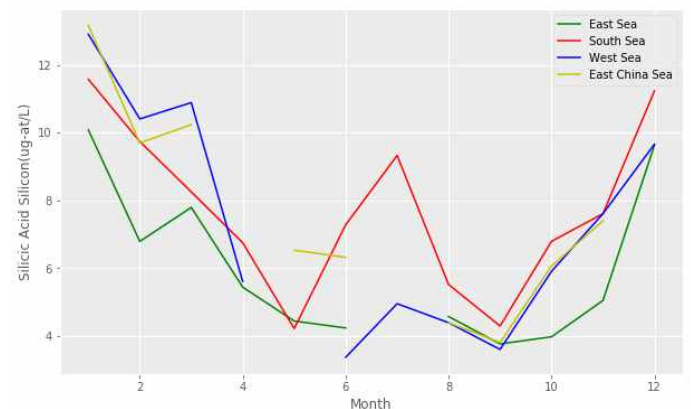


그림 11. 월별 해역의 규산규소 변화

표 2. 로지스틱 다중 회귀 분석의 결과

독립 변수	회귀 계수
수온	0.333
염분	-0.118
용존산소	-0.049
인산염인	0.378
아질산질소	-0.071
질산질소	-0.062
규산규소	-0.019

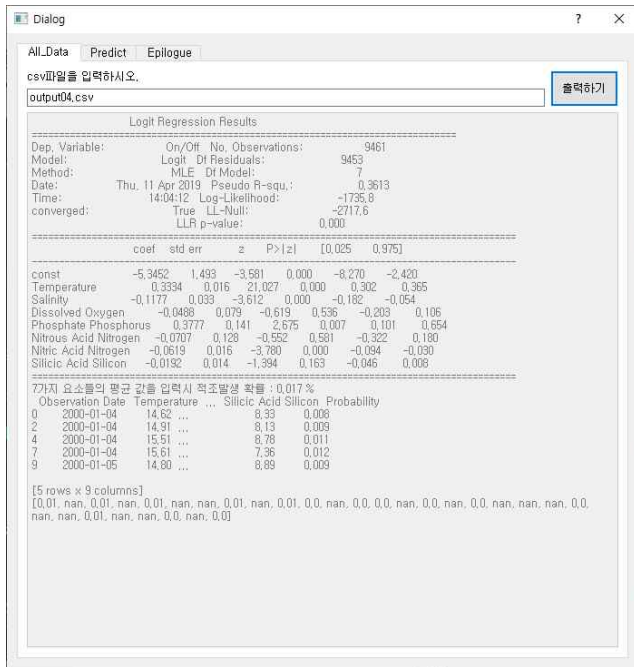


그림 12. 로지스틱 다중 회귀 분석의 구현 화면

유해 적조 데이터셋의 종속변수는 유해 적조의 발생 여부를 나타내는 이진형이므로 로지스틱 회귀 분석을 적용하였고 결과는 표 2와 같다. 분석 결과에 의하면, 수온과 인산염인의 농도가 높을수록 유해 적조의 발생 가능성이 높게 나타난다. 반면에, 염분의 농도가 낮을수록 유해 적조 발생의 가능성이 높아 보인다.

로지스틱 다중 회귀 분석의 결과와 월별 해역의 아질산질소의 농도와 질산질소의 농도를 함께 고려해 보면, 아질산질소와 질산질소의 농도가 낮으면 유해 적조의 발생 가능성이 있다고 보인다. 같은 방법으로 비교하면, 용존산소와 규산규소의 농도가 유해 적조 발생에 미치는 영향력은 낮아 보인다.

우리는 일반 사용자의 편의를 위하여 PyQt를 이용하여 로지스틱 회귀 분석을 위한 GUI를 구현하였으며 그림 12에 나타낸다. 그림 12의 입력 창에서 처리 형식에 맞는 CSV 파일을 입력하면 로지스틱 다중 회귀 분석을 수행하고 결과를 화면에 나타낸다.

#### 4. 결론

본 연구에서는 수산 피해를 입히는 유해 적조의 발생 원인과 영향력을 분석하기 위하여 빅 데이터 처리 과정을 적용하였다. 유해 적조 관련 데이터의 수집을 위하여 국립수산물과학원에서 해양 관측 데이터를 내려 받고 유해 적조 발생 데이터를 웹크롤링 하였다. 수집한 데이터에 들어있는 비어있는 값들과 부정확한 값들을 파이썬 프로그래밍으로 정제하였다. 정제하고 병합한 유해 적조 데이터셋을 파이썬의 팬더스 라이브러리를 사용하여 분석하고 맷플롯립으로 시각화 하였다. 로지스틱 다중 회귀 분석을 수행하였고 이 분석의 결과와 기술통계에서 나온 결과를 비교하였다. 비교 분석한 결과에 의하면, 수온과 인산염인의 농도가 높을수록 유해 적조가 발생할 가능성이 높게 나타났다. 반면에, 염분이 낮을수록 유해 적조의 발생 가능성이 높아 보인다. 아질산질소와 질산질소의 농도가 낮으면 유해 적조의 발생 가능성이 있다고 본다. 용존산소와 규산규소의 농도가 유해 적조 발생에 미치는 영향력은 낮아 보인다.

우리는 유해 적조 데이터셋을 로지스틱 분석할 수 있는 GUI를 구현하였다. 분석에 사용한 데이터는 CSV 파일로 공개하고 데이터베이스에 저장하여 다양한 검색 기능을 구현할 예정이다.

#### 참고문헌

- [1] 고우진 외, “2017-2018년도 한국연안의 적조발생 상황,” 국립수산물과학원, p. 9, 2018.
- [2] 국립수산물과학원 적조정보시스템, <http://www.nifs.go.kr/red/main.red>
- [3] 이문옥, “한국 남해 중부 해역의 장기수질환경변화와 *Cochlodinium polykrikoides* 적조 발생의 특징,” 한국해양 환경공학회지, Vol. 14, No. 1. pp. 19-31, 2011.
- [4] 국립수산물과학원 해양자료센터, <http://www.nifs.go.kr/kodc/index.kodc>
- [5] M. Chen, S. Mao, and Y. Liu, “Big Data: A Survey,” *Mobile Networks and Application*, Vol. 19, pp. 171-209, 2014.