

논문 2012-49CI-2-1

## 적조현상의 실험적 예측성능 비교

### ( Experimental Performance Comparison for Prediction of Red Tide Phenomenon )

허 원 지\*, 원 재 강\*\*, 정 용 규\*\*\*

( Won-Ji Heo, Jae-Kang Won, and Yong-Gyu Jung )

#### 요 약

최근 지구적인 기상 변화와 함께 태풍, 집중 호우 등으로 인한 재산 및 인명 피해가 크게 늘어가고 있다. 각 분야에서 재난을 소재로 한 여러 주장이 있어 왔다. 특히 영화까지 출시될 정도로 앞으로 이 세계에 불어 닥칠 기상변화에 촉각을 곤두세우고 있다. 해상에서 기상현상은 이런 피해를 줄이고 장기적으로 기상의 변화를 예측하려는 요구가 대두되고 있다. 본 연구에서는 적조현상을 다중선형 회귀분석과 인공신경망 기법을 통해 미리 예측하는 방법을 제시한다. 적조현상이 발생하는 원인에 위험수치를 보일시에 센서로 감지하여 기상 예측관에 데이터를 실시간으로 전송해줄 수 있는 유비쿼터스 기반의 맞춤형 해양 정보 서비스 시스템의 활용방안을 통하여 기상재해의 피해를 최소화할 수 있을 것으로 기대한다.

#### Abstract

In recent years global climate change of hurricanes and torrential rains are going to significantly, that increase damages to property and human life. The disasters have been several claimed in every field. In future, climate changes blowing are keen to strike released to the world like in several movies. Reducing the damage of long-term weather phenomena are emerging with predicting changes in weather. In this study, it is shown how to predict the red tide phenomenon with multiple linear regression analysis and artificial neural network techniques. The red tide phenomenon causing risk could be reduced by filtering sensor data which are transmitted and forecasted in real time. It could be ubiquitous driven custom marine information service system, and forecasting techniques to use throughout the meteorological disasters to minimize damage.

**Keywords** : red tide phenomenon, multiple linear regression analysis, artificial neural networks, phytoplankton, phosphorus, nitrogen

#### I. 서 론

오늘날 지구 자연환경의 변천에 따른 기후변화나 재난이 바다에서 비롯되는 것을 주목해 보면 해양과학의 중요성은 더욱 강조될 수밖에 없다. 바다가 지구표면의

10분의 7을 차지하고 있는 걸 보면 우리가 살고 있는 행성을 지구(地球)가 아니라 해구(海球)라 불리도 될 듯하다. 그러나 전 지구를 둘러싸고 있는 바다는 아주 위험한 상태에 있다. 그 중 대표적인 것이 해양에서 수질 오염의 한 현상으로 나타나는 적조현상이다. 지금까지 적조현상을 일으키는 동물성 플랑크톤과 식물성 플랑크톤에 관한 연구가 많았다. 장마기가 됐을 때 수온과 pH에 따라 플랑크톤의 수가 변화하는 것에 대한 연구가 진행되었다. 그러나 이것은 환경오염에 대한 국민 의식이 개선되고 환경보호 단체들의 노력으로 문제의 심각성을 일반 국민들에게 널리 알리는 데는 일정한 성과를 거두었지만 기업의 무책임성 때문에 구체적인 결실을

\* 정회원, 을지대학교 의료산업학부 의료전산학전공 (Major of Medical Computing, Eulji University)

\*\* 정회원, 경기대학교 전자계산학과 (Department of Computer Science, Kyonggi University)

\*\*\* 평생회원-교신저자, 을지대학교 의료IT마케팅학과 (Department of Medical IT marketing, Eulji University)

접수일자: 2011년1월31일, 수정완료일: 2012년3월6일

못 거두고 있는 실정이다. 따라서 본 논문에서는 적조 현상을 유발하는 물질들을 다중선형회귀분석 모델과 인공 신경망 기법에 맞추어 분석하여 보고 성능 비교 실험한 후 USN 알고리즘으로 적조현상을 예측 가능하게 하는 모델을 연구해 본다.

## II. 관련 연구

### 2.1 적조현상(Red Tide Phenomenon)

적조현상이란 해양 식물성 플랑크톤이 일시에 대량 번식하거나 집적함으로써 바닷물 색이 붉게 변색되는 현상으로 오랜 과거부터 자연적으로 발생되었고 현재는 연안환경의 부영양화로 특정해역에 주기적으로 발생되고 있다.

이것은 아름다운 대자연 본연의 미관을 해치고 심각한 환경오염을 유발한다. 특히, 1960년대 산업화, 공업화가 진행되면서부터 농토에서 흘러 들어오는 표면수는 다량의 영양성분과 비료성분을 포함하고 있어서 질소나 인과 같은 무기 영양 염류에 의해 성장이 제한되고 있던 식물성 플랑크톤의 성장을 돕는다.

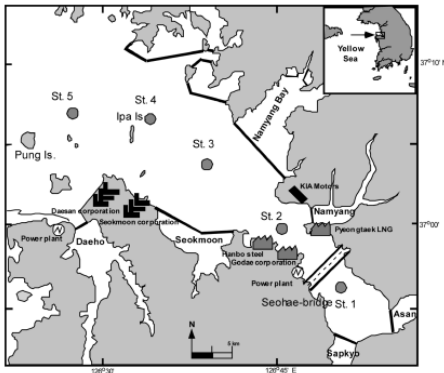


그림 1. 아산 만의 적조현상  
Fig. 1. map of the sampling sites in Asan Bay.

### 2.2 다중 회귀 분석

독립변수가 하나이며 독립변수와 종속변수간의 관계가 선형인 것을 단순회귀분석이라 한다. 그러나 회귀분석을 실제에 적용할 때 하나의 독립변수만으로 종속변수를 충분히 설명할 수 없는 경우가 대부분이다. 따라서 두 개, 혹은 그 이상의 독립변수를 이용하여 회귀분석을 할 경우가 많은데 이를 다중회귀분석(multiple regression model)이라 한다. 독립변수가 n개인 다중회

귀분석의 일반적 형태는 다음 식과 같다.

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n \tag{1}$$

여기서 Y는 종속변수,  $\beta_n$ 는 회귀계수,  $x_n$ 는 독립변수를 나타낸다. 독립변수  $x_n$ 의 회귀계수  $\beta_n$ 은 다른 독립변수들  $x_n$ 이 1단위 증가할 때, 이에 따른 종속변수 Y의 변동 폭을 나타낸다.

### 2.3 인공 신경망

인간의 뇌는 주로 뉴런이라 불리는 신경 세포로 구성되어 있으며 축색돌기라 부르는 섬유 가닥들로 다른 뉴런과 연결되어 있다.

인간의 두뇌 구조와 유사하게, ANN은 노드들과 방향을 가진 링크들의 상호 연결된 집합으로 구성된다. 그림 2와 같이 ANN은 입력층, 출력층, 그리고 은닉층으로 구성되어 있으며, 입력 값을 이용하여 인공 뉴런의 출력 값을 계산한다. ANN은 X1부터 Xn까지 입력값에 각각 W1부터 Wn까지의 가중치를 곱하고 그 모든 합이 변형함수를 통해 임계치가 초과되면 출력값이 발생하게 된다.

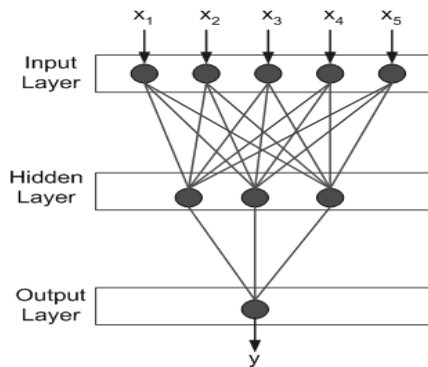


그림 2. 인공신경망의 구성  
Fig. 2. artificial neural network configuration.

### 2.4 모델 비교

다중선형회귀분석은 standard Coefficients를 사용하여 각각의 설명 변수가 종속변수에 얼마나 영향을 미치는지를 가능할 수 있게 해주며 이 설명 변수들을 사용해 종속변수를 가능한 정확히 예측할 수 있게 해주기 때문에 매우 유용한 분석방법이다. 반면에 인공신경망 기법은 복잡하고 다양한 자료를 쉽게 해결해 준다는 점, 질적 변수와 양적 변수에 관계없이 모두 분석이 가

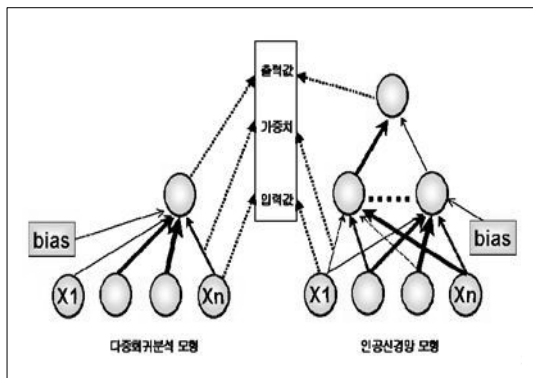


그림 3. 다중선형회귀분석과 인공신경망 모델의 비교  
Fig. 3. Multiple linear regression analysis and artificial neural network model.

능하다는 점, 입력 변수들 간 비선형 조합이 가능하여 예측력이 우수하다는 점을 장점으로 꼽는다. 그러나 결과에 대한 분류와 예측 결과만을 보여주기 때문에 결과 생성의 원인과 이유를 설명하기가 어렵다.

### III. 실험

#### 3.1 실험대상

과거의 데이터를 토대로 수치가 일정 임계값을 넘어 가면 공장 유입 내에 폐수가 흘러들어가는 위치에 센서를 장착하여 센서가 울릴 수 있도록 하는 방법으로 데이터를 수집하는 것이 보편적인 방법이다. 그림 4는 해양 정보가 관측된 후 이 정보를 토대로 맞춤형 해양 정보 서비스를 제공할 때까지의 과정을 나타내는 순서도를 도식한다.

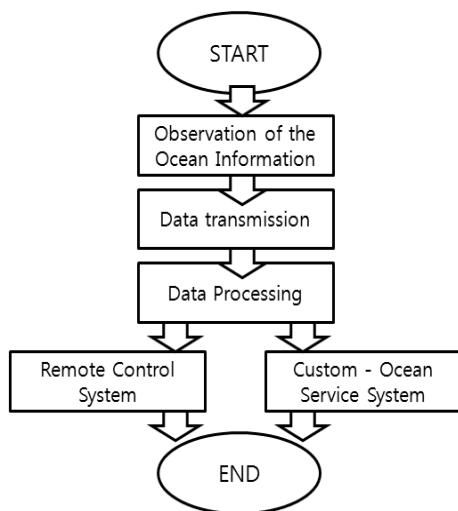


그림 4. 데이터 수집과정  
Fig. 4. data collection process.

해양수산부에서 1995년부터 2009년까지 조사 발표된 해수수질 실태보고의 데이터를 사용하였다. 본 데이터는 서해, 남해, 동해의 각 해역별로 수온, 염분, PH, 용존산소, 화학적산소요구량, 용존무기질소, 총질소, 용존무기인, 총인, 규산규소, 부유물질, 클로로필의 수치로 구성되어 있다.

#### 3.2 전처리

본 실험에서는 각 연안별 해수 수질 실태를 나타내는 15년간의 데이터를 이용하였고, 데이터의 품질을 유지하기 위하여 전처리과정을 거쳤다. Null데이터는 누락값을 추정하여 속성이 연속형인 경우에는 가장 가까운 이웃의 평균 속성 값을 사용하였다. 속성이 범주적이라면 가장 공통적으로 발생하는 속성 값을 사용하여 데이터의 신뢰성을 높였다. 또한 Over Range 데이터는 과감히 제거하였다.

#### 3.3 실험방법

본 연구에서는 위 표 1을 기준으로 하여 실험데이터의 적조 수치를 점검하는 실험을 하였으며 두 가지 모델에 대한 성능을 비교하였다. 마이너 틀은 XLMiner를 사용하였으며 적조현상의 최대 원인 물질인 총 인&질소와 다른 물질들 간의 상관관계를 알기 위해 각각을 Output으로 하여 실험데이터를 구성하였다.

표 1. 총 인과 수온, pH, 용존산소DO, 화학적 산소요구량COD, 부유물질SPM 과의 상관관계  
Table 1. the relationship between Total phosphorus water temperature, pH, dissolved oxygen DO, chemical oxygen demand COD, suspended solids, SPM.

Mining Schema						
Selected Variables	수온(°C)	pH	DO(mg/l)	요구량	SPM(mg/l)	N(mg/l)
Variable Type	Input	Input	Input	Input	Input	Output
Inputs Normalised	No					
Constant term present	Yes					

표 2. 총 질소와 수온, pH, 용존산소DO, 화학적 산소요구량COD, 부유물질SPM 과의 상관관계  
Table 2. the relationship between total nitrogen and water temperature, pH, dissolved oxygen DO, chemical oxygen demand COD, suspended solids, SPM.

Mining Schema						
Selected Variables	수온(°C)	pH	DO(mg/l)	요구량	SPM(mg/l)	총인T-P(mgP/ml)
Variable Type	Input	Input	Input	Input	Input	Output
Inputs Normalised	No					
Constant term present	Yes					

3.4 교차검증(10-fold-validation)

k-묶음 교차검증 기법은 학습 시 주어진 데이터를 학습 데이터(training dataset)와 검증 데이터(validation dataset) 로 나누어 검증 데이터에 대한 최적의 모델을 선정하는 방법이다.

전체 집합을 k개로 나눈 뒤 하나를 다른 것들과 비교하여 전체적으로 특이한 집합이 없는지 확인하는 방식이다. 수집된 최초의 오리지널 샘플을 k개의 서브샘플로 나눈다. 하나의 서브 샘플은 모델의 테스트를 위한 검증 데이터로 두고 남은 k-1개의 서브 샘플들은 학습 데이터로 사용된다. 모든 서브샘플들이 검증 데이터로 정확히 한번씩 사용될 때까지 k개의 서브 샘플들은 cross-validation 프로세스 동안 k번 반복된다. 프로세스의 각 스텝마다 각 부분으로부터 나온 k개의 결과는 하나의 평가 지표로 만들기 위해 평균을 구하며, 이를 이용해 검증을 수행할 수 있다. 10-fold cross-validation이 일반적으로 사용되며 이 방법의 장점은 모든 데이터들이 학습과 검증에 모두 사용된다는 것이다. 또한 모든 데이터는 전부 단 한번만의 validation이 수행된다.

본 실험에서는 간단하면서도 우수한 성능을 나타내는 교차검증을 사용하였다. 학습 데이터와 검증 데이터의 비율은 60대 40으로 하였다. 훈련 집단의 데이터를 통해 유효 집단의 데이터를 예측할 수 있다.

표 3과 표 4는 전체 데이터를 10분위로 나누어 속성에 대한 각 집단의 평균을 구하였다. 이것은 각 집단이 전체 데이터에 비해 총 인 및 총 질소가 평균적으로 어떤 수치를 기록하며 평균으로부터 얼마만큼 고르게 분포되어 있는지 확인할 수 있다. 이렇게 함으로써 데이터의 특징을 파악하고 충분하지 못한 데이터의 단점을 보완할 수 있다. 더불어 적조현상을 좌우하는 물질인

표 3. 훈련 집합의 10 분위와 평균(총 인)  
Table 3. Total Phosphorus of 10 quartile and average.

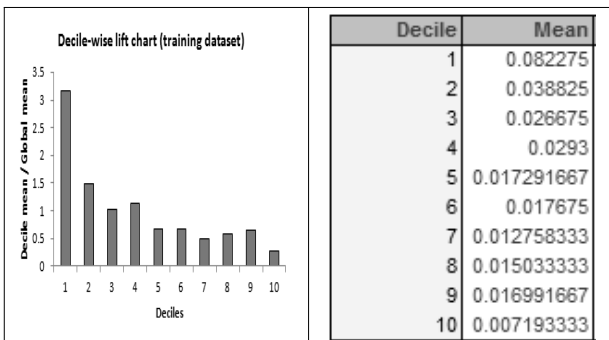
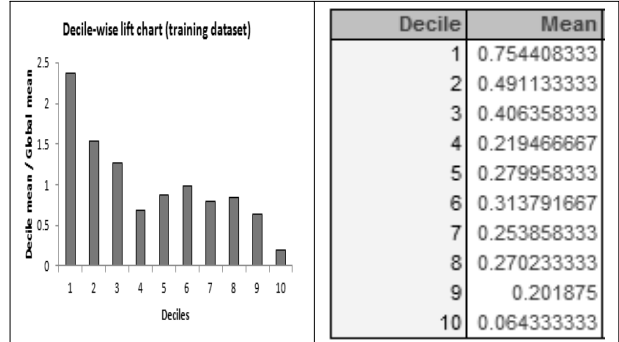


표 4. 훈련 집합의 10 분위와 평균(총 질소)  
Table 4. training set of 10 quartile and average. (total nitrogen)



인 또는 질소와 다른 요소들 간의 상관관계를 밝히는 데도 일조한다.

IV. 실험결과 및 고찰

4.1 다중선형회귀분석 실험결과

적조 현상을 발생시키는 가장 근본적인 물질은 인과 질소이다. 따라서 인과 질소를 유발해내는 물질들과 그들간의 관계를 파악하여 그 물질들을 빠르게 예측해 낸다면 결국 적조를 예방할 수 있다. 총 인과 용존산소 DO, 화학적산소요구량COD는 서로 양의 상관관계에 있다. 총 인의 수치를 낮추어 적조현상을 피하고 싶다면 용존산소DO와 화학적산소요구량COD의 양을 줄여주면 된다.

마찬가지로, 질소에 대해 같은 실험을 반복한 결과, 총 질소와 pH, 용존산소DO, 화학적산소요구량COD가 서로 양의 상관관계에 있다는 것을 알 수 있다. 총 질소의 부하를 막기 위해서는 pH, 용존산소DO, 화학적산소요구량COD를 감소시키면 된다.

표 5. 실험결과  
Table 5. Experimental Results.

Model	
Input Variables	Coefficient
Constant Term	0.47879499
수온(°C)	-0.00627712
pH	-0.05982778
용존산소DO(mg/l)	0.01415796
화학적산소요구량CDO(mg/l)	0.01249883
부유물질 SPM(mg/l)	-0.00070362

표 6. 실험결과

Table 6. Experimental Results.

Model	
Input Variables	Coefficient
Constant Term	-1.61637437
수온(°C)	-0.03316385
pH	0.19489463
용존산소DO(mg/l)	0.06935303
화학적산소요구량CDO(mg/l)	0.17410143
부유물질SPM(mg/l)	-0.00294186

회귀분석을 통한 적조현상 예측을 통해 각 물질들 간 상관성을 파악할 수 있어 각 설명 변수가 종속 변수에 얼마나 영향을 미치는지를 알아낼 수 있었다. 이것은 종속변수를 정확하게 예측해 주기 때문에 적조 현상을 감지하고 그것을 유발하는 물질을 빠르게 파악해 낼 수 있는 장점을 가진다.

4.2 인공신경망 실험결과

인공신경망은 은닉층을 4개로 각각 25의 가중치를 적용하여 실험하였다.

인공신경망을 통한 적조현상 예측을 통해 복잡하고 다양한 자료를 쉽게 해결하였다. 이 기법은 입력 변수들 간 비선형 조합이 가능하기 때문에 결과 도출에는 유리하지만 적조현상과 같은 어떤 특정한 현상의 원인이 필요한 경우에서, 결과 생성의 원인과 이유를 설명하기가 어렵다는 단점이 있었다.

표 7. 실험 결과 - 은닉층과 출력값

Table 7. Experimental Results - hidden layer and output.

Hidden	Input Layer					
	수온(°C)	pH	용존산소	화학적산소	부유물질	Bias Node
Node # 1	-0.98183544	-0.74627755	-0.14241772	-0.38843751	0.239795188	0.840709706
Node # 2	-0.87747591	0.743924538	0.064424304	0.898111145	0.572872827	0.718261899
Node # 3	1.055086622	-0.85405214	-0.22382259	-0.56229728	0.564301645	-0.12238797
Node # 4	-0.39022717	0.924314252	0.781345916	0.738658255	0.279977976	0.009641982
Node # 5	-0.10542538	0.837325757	0.65968907	-0.49284011	0.622152009	-0.63476183
Node # 6	0.43343803	0.180755822	-0.18038973	-0.80474247	0.934845481	0.985454731
Node # 7	-0.69643639	0.56653933	0.278450052	0.838261899	-0.58764574	0.7798427
Node # 8	0.123327448	-0.98060231	-0.54571424	-0.53475377	-0.97471215	0.478694823
Node # 9	-0.60370035	-0.18948859	-1.00289773	-0.77123916	-0.09815884	0.760363017
Node # 10	0.689362057	-0.29572333	0.334796182	-0.85846703	-0.67509125	-0.01164942

4.3 고찰

다른 여러 분류기들의 성능을 비교하여 주어진 데이터 집합에 대하여 어떤 분류기가 더 잘 작동하는지 정하는 것은 종종 유용하다. 본 논문에서는 다중선형 회귀분석과 인공신경망 기법으로 적조현상을 미리 예측하

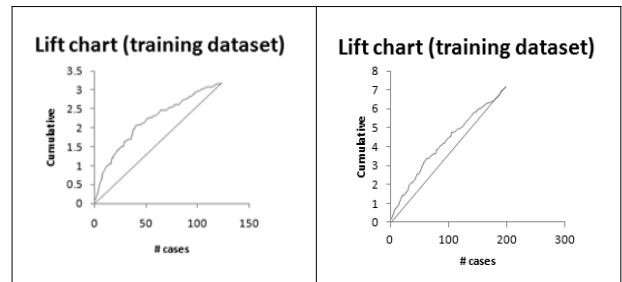


그림 5. 다중선형 회귀분석과 인공 신경망의 훈련 집단 비교(총 인)

Fig. 5. total Phosphorus of Multiple linear regression analysis and artificial neural networks.

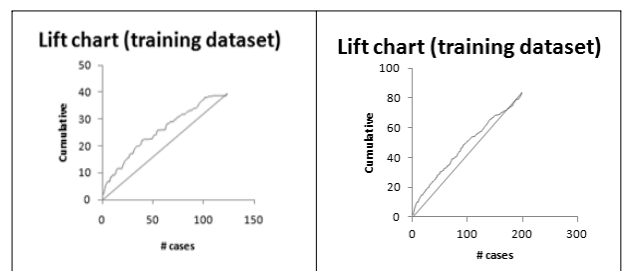


그림 6. 실험결과 - 다중선형 회귀분석과 인공 신경망의 훈련 집단 비교(총 질소)

Fig. 6. total nitrogen of multiple linear regression analysis and artificial neural network training compared to group.

기 위한 실험을 하였는데 적조현상을 일으키는 물질들의 상관관계를 쉽게 파악할 수 있는 다중선형 회귀분석이, 블랙박스로 숨겨진 레이어로 감추어진 인공신경망에 비해 훨씬 뚜렷한 인과관계를 보였음을 확인할 수 있었다.

다중선형 회귀분석의 경우, 훈련 집단의 그래프에서 총 인 및 총 질소의 수치가 평균인 빨간 색의 선보다 상위로 많이 벗어났음을 확인할 수 있다. 이는 다중선형 회귀분석이, 은닉층의 복잡성으로 인해 상대적으로 예측에 시간이 오래 걸리는 인공신경망보다 과거의 데이터를 학습한 후 새로운 데이터를 예측하였을 때 더 눈에 띄는 결과를 도출함을 인지할 수 있음을 의미한다.

V. 결론

지구 표면적의 약 71%를 차지하고 있는 바다는 천연 자원의 보고이며, 지구 생태계의 모체이다. 그러나 환경 오염과 생태계의 파괴는 전 지구적인 차원에서 진행되며 날로 그 심각성이 더해 가고 있다. 이것은 적조현상의 형태로 표출되며 본 논문에서는 이를 조명하였다.

실험에서는 적조현상을 다중선형 회귀분석과 인공신경망 기법의 비교를 통해 자연재해를 미리 예측하는 방법을 제시하였다. 적조현상의 원인을 감지할 수 있는 다중선형 회귀분석이, 결과에 대한 분류만을 보여주어 결과 생성의 원인과 이유를 설명하기가 어려운 인공신경망보다 적조 탐지에 유리할 것으로 판단되었다. 이 기법을 통한 적조현상의 예측은 긴급사태를 대비하는데 유용하다. 관측 상황에 대한 해양 정보 시스템의 예보를 통해 지역 당국은 적조현상에 대한 감시와 지역 보건문제를 준비할 수 있으며, 패류 양식업자에게도 수확 시기를 조절할 수 있는 기회를 제공한다. 적조현상은 한반도 전역에 걸쳐 반복적으로 발생하는 문제인데다, 이를 해결할만한 충분한 자원을 확보하지 못했기 때문에, 적조현상의 심각성에 대해 미리 정보를 가지고 사전에 대책을 준비한다면 보유하고 있는 자원을 효율적으로 사용할 수 있다. 이로써 해양 수질 오염을 막고 생태계를 보존하여 국가적인 경제적 손실을 방지한다.

현대적인 과학 기술의 발전과 더불어 이러한 예측은 상당한 효용 가치를 가질 것으로 기대된다. 바다는 경제적 측면, 식량 자원 측면 그리고 환경 측면에서 굉장한 이득을 가져다주기 때문이다. 바다는 미래의 주요한 식량 공급지 역할을 하고 수송로 기능을 수행하면서 중요한 교통로가 되어 왔다. 따라서 앞으로도 우리 인류는 바다를 지키고 가꾸어야 하는 데 많은 관심과 노력을 기울여야 한다. 그러나 본 실험은 통계적 기법이나 간단한 마이닝 기법을 활용하였으므로 다음 연구에서는 적조현상을 앙상블 기법을 이용하여 예측하고자 한다.

## 참 고 문 헌

- [1] 수중세계 1999/09/10호
- [2] 병들어가는 호수- 적조현상, 물고기의 폐죽음, 식수오염, 김상중, 동아사이언스 과학동아 제5권 제6호(통권 제54호), 1990,6,pp 106-113
- [3] 김범준, 김형수, 최승안, 심명필, 다중회귀분석을 이용한 충주댐 홍수조절 효과분석, 2006 대한토목학회 정기학술대회, 1pp
- [4] 이두별, 박철, 양성렬, 신용식, 아산만 해역에서 장마기 전후 중형동물플랑크톤 군집의 변화, 3pp
- [5] 조남훈, 교차검증을 이용한 SVM 전력수요예측, 대한전기학회, 전기학회논문지 A, 제55A권 제11호 2006.11, page(s): 485-491
- [6] 용환승 외 6명, INTRODUCTION TO DATA MINING, 인피니티북스, 241 - 250

- [7] 권혁춘, 이병걸, 이창선, 고장우, 로지스틱회귀분석 기법과 인공신경망기법을 이용한 제주지역 산사태 가능성분석, 한국지형공간정보학회, 한국지형공간정보학회지, 제19권 제3호 2011.9, page(s): 33-40
- [8] 김주훈, 김경탁, 윤강훈, 다중선형회귀모형에 의한 홍수량예측 적용, 대한토목학회, 2005년도 대한토목학회 정기 학술대회 2005.10, page(s): 2672-2675
- [9] 신택수 홍태호 공역, Data Mining for business intelligence, SciTech
- [10] 이우영, 송병호, 김갑기, 최조천, 이경효, 이연우, 사례 기반 추론을 이용한 적조 예측 시스템 구현, 대한전자공학회 2010년 하계종합학술대회, 2010.6
- [11] 박선, 김경준, 이진석, 이성로, 신경망과 SVM을 이용한 적조 발생 예측, 전자공학논문지제48권제5호, 2011.9

## 저 자 소 개



허 원 지(정회원)

2009년~현재 을지대학교

의료산업학부

<주관심분야: 의료정보시스템, 의사결정지원시스템>



원 재 강(정회원)

1999년 강릉대학교 (이학사)

2002년 경기대학교 (이학석사)

2009년 경기대학교 (이학박사)

2000년~현재 경기대학교

외래강사

<주관심분야: 워크플로우, BPM, 데이터마이닝>



정 용 규(평생회원)

1981년 서울대학교 이학사

1994년 연세대학교 공학석사

2003년 경기대학교 이학박사

1999년~현재 을지대학교 의료IT

마케팅학과 교수

<주관심분야: 임상데이터마이닝, 의료정보시스템, 전자거래표준>