DBpia

사례 기반 추론을 이용한 적조 예측 모니터링 시스템 구현 및 설계

정회원 송 병 호*, 정 민 아**, 이 성 로***

A Design and Implementation Red Tide Prediction Monitoring System using Case Based Reasoning

Byoung-Ho Song*, Min-A Jung**, Sung-ro Lee*** *Regular Members*

요 약

적조 현상에 대한 판별, 예측 분석을 위한 시스템은 현재 개발이 아주 미흡한 상태이고 현재의 적조원인에 대한 연구는 화학 및 생물학적 원인의 규명에 대해 그 초점이 맞추어져 있어 지능적인 의사 결정 알고리즘을 갖는 시스템 구현이 필요하다. 본 논문에서는 사례 기반 추론 기법을 이용하여 적조 현상에 관한 사례를 지식 베이스로 구축하고 추론하는 시스템을 설계하였다. 가장 유사한 사례 추천을 위해 KNN 알고리즘을 이용하였고 적조 사례 베이스를 구축하기 위하여 375 건의 데이터를 입력 받아 실험하였다. 학습 데이터로부터의 영향을 최소화하고 신뢰성을 확보하기 위해 10-Fold 교차검증을 수행한 결과 적조 사례에 대한 평균 정확도는 약 84.2%를 나타냈고 유사도 분류 k 개수가 5인 경우에 최적의 수행 결과를 나타냈다. 또한, 추론된 결과를 이용하여 적조 모니터링 시스템을 구현하였다.

Key Words : Red Tide, Case Based Reasoning, Expert System, Knn, SVM

ABSTRACT

It is necessary to implementation of system contain intelligent decision making algorithm because discriminant and prediction system for Red Tide is insufficient development and the study of red tide are focused for the investigation of chemical and biological causing. In this paper, we designed inference system using case based reasoning method and implemented knowledge base that case for Red Tide. We used K-Nearest Neighbor algorithm for recommend best similar case and input 375 EA by case for Red Tide case base. As a result, conducted 10-fold cross verification for minimal impact from learning data and acquired confidence, we obtained about 84.2% average accuracy for Red Tide case and the best performance results in case by number of similarity classification k is 5. And, we implemented Red Tide monitoring system using inference result.

Ⅰ. 서 론

적조현상이란 바다의 플랑크톤이 일시에 대량번식하거나 집적됨으로서 바닷물의 색이 붉은색과 황녹색으로 변하는 현상으로서 오랜 옛날부터 자연적으로 발생하였다. 그러나 최근에는 연안도시와 임해공업단지 등 임해산업의 발달과 여기서 발생된 각종 오염물질이 연안 해역으로 유입되어 적조현상의 발생빈도가

※이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소 지원사업으로 수행된 연구임(2010-0028295)
※이 논문은 2008년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. D00280).
* 목포대학교 중점연구소(cssstar@empal.com), ** 목포대학교 컴퓨터공학과 (majung@mokpo.ac.kr),
*** 목포대학교 정보전자공학과 (srlee@mokpo.ac.kr)
논문번호 : KICS2010-06-250, 접수일자 : 2010년 6월 3일, 최종논문접수일자 : 2010년 11월 23일

www.dbpia.co.kr

늘어나거나 그 기간도 길어짐으로서 연안의 어민에게 대량 피해를 주고 있는 현상을 말한다.

1990년부터 거의 매년 우리나라 연안에는 유해성 적조의 발생으로 양식어장 피해가 빈번히 발생하고 있다. 이러한 유해성 적조 생물은 기온, 수온, 염분, 일사량, 일조시수, 영양염류, 강수량, 바람, 조류 등의 주요한 해양 환경 인자에 의해 그 발생과 소멸이 영향을 받게 된다[1].

따라서, 적조의 발생과 진행을 연구하는 해양학자는 이러한 적조 발생에 관한 다양한 환경 인자를 조사하고 이들 자료에 의해 적조의 생물학적 활동을 분석하였다.

적조 연구는 적조 생물의 종을 분류하고 이들의 생활사를 분석하는 연구에서부터 적조의 번성과 이동을 예측하는 모델, 적조를 긴급히 방제하는데 효과적인 방안의 연구 등 다양한 연구가 추진되어 왔다. 국내외 관련 연구를 살펴보면 연안이용도가 높은 일본에서는 수산 피해를 일으키는 적조의 발생기구 규명과 방제기술 개발연구를 1950년대부터 수행하고 있고, 미국과 유럽의 경우 생물공학적인 기법에 의한 적조 제어 기술개발 연구를 추진하고 있다.

국내에서도 해양연구원에서 해양환경 관측 및 개선을 위한 기반 기술 연구 중 적조생물 원격탐사 기술개발을 하고 있으며 기상인자를 활용한 연안 적조예측 기술 개발 등 많은 전문가들이 활발한 연구 활동을 하고 있다.

그러나 적조 판별, 예측 분석을 위한 시스템은 현재 개발이 아주 미흡한 상태이고 기존의 시스템은 단순히 모니터링 하는 데 그치고 있으므로 적조의 특성을 고려한 지능적인 의사 결정 알고리즘을 갖는 시스템 구현이 필요할 것이다.

의사결정을 지원하기 위해서는 전문가 시스템, 신경망, 사례기반추론과 같은 인공지능 기술이 활용되고 있으나 기존의 방법들은 해 추적 과정의 검토가 힘들고 모델 구축 과정에 전문적 지식을 요구하는 경우가 많아 과거의 경험으로부터 효과를 볼 수 있는 추론 기법인 사례기반추론을 이용한 방법들이 연구되고 있는 추세이다.

사례기반추론은 지속적으로 발생하는 과거 경험들을 통하여 유사한 문제를 해결하는 분야에 적합하고, 복잡한 문제라도 과거에 얻은 경험 사례와 같다면 특별한 추론 없이 해결책을 바로 얻을 수 있으며, 유사도 함수를 이용하여 가장 유사한 사례를 찾기 때문에 해결책을 구하는데 많은 시간이 요구되는 문제에 효율적이다[2,3].

본 논문에서는 현재 사례와 가장 유사한 k개의 사례를 추천해 주는 KNN(K-Nearest Neighbor) 알고리즘을 이용한다.

KNN 알고리즘은 이미 알려진 개체들을 훈련 집합의 형태로 메모리에 기억한 다음 그 중 유사한 개체를 선택하여 선택된 개체의 값에 따라 새로운 개체의 값을 예측하는 방식의 분류(classification) 알고리즘이다.

이에 본 논문에서는 적조로 인한 어업 피해를 최소화하기 위해 국내 연안에서 발생한 적조의 발생 범위, 적조 생물, 해양환경요소를 사례 기반에 의한 적조 분석 시스템을 구축하고자 한다. 입력된 데이터는 사례베이스를 통해 가장 유사한 과거 사례를 추천해 주고 새로운 사례가 발생한다면 사례 베이스에 저장되고 분석된 결과를 토대로 적조 모니터링 시스템을 구현한다.

본 논문은 다음과 같이 구성되었다. 2장에서는 관련 연구 동향을 알아보고 3장에서는 입력 사례 데이터 측정 및 사례베이스 구축, 4장은 시스템 구현 결과를 보여주고 5장에서는 성능 평가에 대해서 기술한다. 6장에서는 결론과 향후 연구 방향을 제시한다.

Ⅱ. 관련 연구

2.1 적조 현상

해수 속으로 질소(N)와 인(P)등의 영양염류가 과다하게 유입되고 해수의 온도가 21-26도를 나타내면 해수 중에 특정 조류가 이상 증식하여 바닷물의 색깔이 빨간색으로 변하게 되는데 이를 적조현상이라 한다. 적조현상은 세계의 모든 연안 수역에서 널리 발생하며, 특히 일본의 세토 내해, 미국의 캘리포니아 연안, 동남아시아 연안과 북해 연안에서 자주 발생하고 있다. 최근 국내 남해 연안과 서해 및 동해 남부 연안 수역에서 자주 발생하는 적조현상은 계절에 관계없이 상습적으로 발생하고 있으며, 특히 적조생물은 규조류에서 편모조류로 바뀌어 가고 고밀도화 되어 가는 경향을 보이고 있다. 적조가 1981년까지는 7, 8월에 발생하였으나 이후에는 4월에도 적조가 발생, 점점 빨리 발생되는 추세를 보였고, 어떤 해역에서는 11월까지도 적조현상이 관찰되는 등 발생기간의 폭이 넓어지고 있고, 어류를 직접 폐사시키는 유해 적조가 발생하고 있다. 적조생물은 광합성 활동에 필요한 일조량이 충분하고 해수의 온도가 증식에 알맞으면 빠른 속도로 증식한다. 이렇게 증식된 유해성 적조 생물은 기온, 수온, 염분, 일사량, 일조시수, 영양염류, 강수량, 바람, 조류 등의 주요한 해양 환경 인자에 의해 그 발

www.dbpia.co.kr

생과 소멸이 영향을 받게 된다[1]. 특히, 기온, 강수량, 일조시수, 일사량, 바람 등의 기상인자는 적조가 발생 가능한 조건을 형성하는데 매우 중요한 인자로 작용한다. 적조 관련 연구를 살펴보면 적조 생물의 종을 분류하고 이들의 생활사를 분석하는 연구에서부터 적조의 번성과 이동을 예측하는 모델, 적조를 긴급히 방제하는데 효과적인 방안의 연구 등 다양한 연구가 추진되어 왔다. Tassan[4], 유신재[5]등은 탁한 해역에서 원격탐사기술에 의해 적조의 공간분포를 파악하고자 하였고, 정종철[6]은 다양한 센서에 의한 적조관측기법을 제시하였다. 그러나 , 적조 발생에 영향을 미치는 환경 인자를 사례베이스로 구축하고 이를 통해 적조의 정보를 예측 분석하는 체계는 제시되지 못하고 있는 실정이다. 이에 본 논문에서는 국내 연안에서 발생한 적조의 발생 범위, 적조 생물, 해양환경요소들에 대한 사례 기반 적조 분석 시스템을 구축한다.

2.2 사례기반추론

사례기반추론은 과거의 어떤 문제를 해결하기 위해 사용했던 경험을 바탕으로 하여 주어진 새로운 문제를 해결하는 방법으로서 분류 및 예측 문제 모두에 효과적으로 적용 가능하다. 사례기반추론은 두 개의 기본 사상에 기반 하는데 하나는 유사한 문제는 유사한 해법을 가진다는 것이고, 다른 하나는 한번 발생한 문제는 자주 발생할 수 있다는 것이다. 따라서 과거에 현재의 문제와 유사한 문제가 존재하였고 그것이 어떻게 해결됐는지를 안다면, 과거의 경험을 바탕으로 현재 문제의 해결책을 추론할 수 있다는 것이다. 사례기반추론의 문제 해결방식은 인간의 문제 해결 방식과 유사하기 때문에 그 결과를 이해하기 쉽고, 새로운 사례를 단순히 저장하는 것만으로도 추가적인 작업 없이 학습이 진행된다는 장점을 가진다. 사례기반추론은 그림 1과 같이 사례 검색, 사례 재사용, 사례 수정, 사례 유지 단계를 거쳐 새로운 문제를 해결한다[6].

사례 검색(Retrieve) 단계에서는 해결해야 할 새로운 문제와 가장 유사한 과거 사례를 추출하고, 사례 재사용(Reuse) 단계에서는 검색을 통해 찾아진 유사 사례들의 해법을 현재 문제 해결을 위해 사용하는 것이다. 사례 수정(Revise) 단계에서는 도출된 해를 평가하고 평가결과를 바탕으로 사례를 재교정하며 사례 유지(Retain) 단계에서는 앞의 3단계를 거쳐 발생한 새로운 사례를 기존의 사례 Library에 저장한다. 사례기반추론은 적용 대상 시스템에 대한 모델링 없이도 학습 가능하기 때문에 적용 분야에 관한 사전 정보가 없는 경우에도 적용할 수 있고 출력 클래스가 이산형

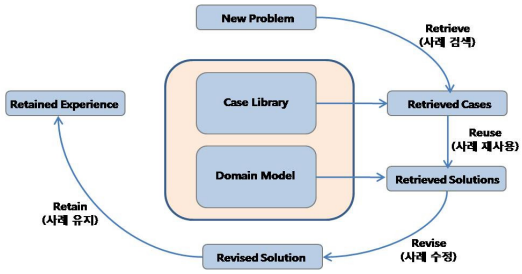

그림 1. 사례 기반 추론 과정
Fig. 1. Case based reasoning process

이거나 연속형인 경우 모두 처리할 수 있는 장점이 있다. 이처럼 다양한 분야에서 사례기반추론 모형의 활용이 증대되면서 그 성능 개선에 대한 요구 역시 점점 증가하고 있는데 사례기반추론 모형은 가장 유사한 사례를 얼마나 신속하게 추출하는가를 그 주요 성능 지표로 한다. 따라서 그 성능을 개선하기 위해서는 어떠한 방법을 통해 사례기반추론의 두 중요한 알고리즘 요소인 사례의 인덱싱(case indexing)과 사례의 추출(case retrieving)을 수행할 것인가를 결정하는 것이 매우 중요하다. 사례기반 추론 시스템에 대한 선행연구를 살펴보면 Limam et al.은 새로운 비지니스 프로세스 재설계를 위한 사례기반 추론 기법의 사용을 제시하였다[7]. Bartlmae & Riemenschneider 는 데이터베이스에서 지식발견 프로젝트 수행 시에 얻어지는 경험 지식들을 관리하기 위한 경험 운영 시스템 접근법과 사례기반추론 기법에 기초한 지식관리 프레임워크를 제시하였다. 실제로 사례기반추론은 최근 몇 년간 문제 해결과 학습의 도구로 많이 사용되고 있으며 다양한 분야에서 성공적인 사례기반 추론 시스템이 개발되고 있다.

2.3 유사도 측정을 위한 Knn 알고리즘

사례 베이스로부터 유사도를 측정하기 위한 대표적인 검색 방법으로는 귀납적 검색과 최근접 이웃 검색이 있다. 귀납적 검색은 사례를 가장 잘 구분시켜주는 속성들을 찾아서 이 속성들을 사용하여 유사 사례를 검색 하는 방법이다. 귀납적 검색은 사례의 검색 및 구성을 위해 의사결정나무 형태의 구조를 사용한다. 최근접 이웃 검색은 현재 문제의 유사 사례 검색을 위해 현재 문제를 표현하는 사례와 사례 베이스에 있는 모든 사례와의 유사도를 측정한 후 유사도가 높은 일정 개수의 사례들을 검색하여 유사 사례를 찾는 방법이다. 최근접 이웃 검색에서 일반적으로 가장 많이 사용되는 방법은 현재 사례와 가장 유사한 k개의 사례를 검색해 주는 KNN 알고리즘이다. 본 논문에서는

www.dbpia.co.kr

KNN 알고리즘을 사용하여 유사 사례를 검색하였다. KNN 알고리즘은 새로운 사례를 분류, 예측함에 있어 가장 가까운 하나의 사례에만 의존하는 것이 아니라 가장 가까운 K개의 사례들로부터 투표를 취하는 방법을 사용하기 때문에 예측의 정확성이 높아지게 된다. 또한, 입력 데이터로부터 모든 사례에 대하여 유사성을 계산하여 유사성이 가장 높은 사례를 선택한다. 그림 2는 kNN 알고리즘의 수행 과정을 나타낸다. KNN 알고리즘에 대한 선행 연구를 살펴보면 KNN 알고리즘은 최초로 Cover 와 Har에 의하여 각각 독립적으로 제안되었다[8]. 이 후 Smith와 Medin 등에 의하여 KNN 알고리즘은 논리학적으로 그 타당성을 인정받았지만 실제 알고리즘을 위한 모델은 개발되지 않은 상태였다[9]. 이후 Aha, Kibler and Albert에 의하여 몇 개의 개체중심 학습(instance-based learning,IBL) 알고리즘이 개발되었다[10]. KNN 알고리즘은 여러 분야에서 적용되고 있으며 그 성능을 검증 받고 있는 기계학습 방법 중의 하나이다.

그림 2. kNN 알고리즘 수행 과정
Fig. 2. kNN Algorithm Execute Process

2.4 SVM 알고리즘

본 논문에서는 사례기반추론 모델의 유용성을 확인을 위해서 분류 문제에 대표적으로 많이 사용되어온 기법인 SVM을 적용한 모델과 성능을 비교 하였다. SVM 알고리즘은 두 그룹을 잘 분리시키는 분류 초평면을 찾는 방법이다[11]. SVM의 기본 원리는 선형 분리가 가능한 문제에서부터 출발한다. d-차원에서 입력데이터 X_i가 주어졌을 때 학습데이터의 출력으로 -1과 +1처럼 이진 값으로 구분되는 문제를 고려한다. 두 집합을 분류하기 위한 모델을 정의하기 위하여 그림 3과 같은 선형 식별함수인 초평면 (hyperplane)을 정의할 수 있다. 여기에서 Support Vector란 분류 규칙을 결정 짓는 경계와 밀접한 연관이 있는 표본을 의미한다. SVM은 단순히 분류 평면을 찾는다거나 표본 에러를 최소화하는 작업을 하는 것이 아니라 분류 여백(Margin)을 최대화 한다.

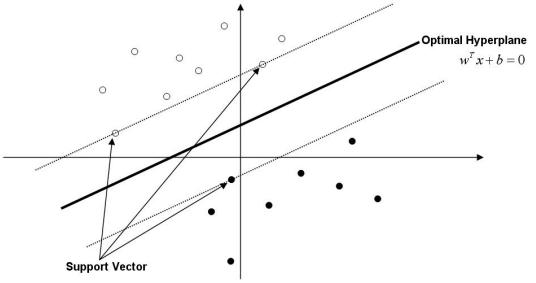

그림 3. 최적화 초평면과 서포트 벡터
Fig. 3. Optimal Hyperplane and Support Vector

Ⅲ. 사용 데이터 및 사례베이스 구축

3.1 사례 베이스 사용 데이터

본 연구에 사용된 데이터는 국립수산과학원에서 제공하는 적조정보시스템에서 획득한 2008년 1월부터 2009년 12월까지 2년간의 적조 발생 데이터로써 발생일시, 발생 지역, 생물 밀도, 수온, 원인 생물, 염분, 질소, 인 데이터로 구성되어 있다. 본 데이터는 관계형 데이터베이스로 구성되어 있으며 크게 적조발생정보, 적조속성정보, 적조생물정보의 세 개의 테이블로 구성되어 있다. 데이터의 개체-관계도는 그림 4와 같다.

테이블 구성을 살펴보면 적조발생정보 테이블은 발생일시, 발생 장소, 발생 좌표값으로 구성되어 있다. 적조생물정보 테이블은 적조 발생시 원인이 되는 적조 생물에 대한 정보로써 발생 일시, 발생 장소, 원인생물, 생물 개체수로 구성되어 있다. 적조속성정보 테이블은 적조 발생에 영향을 미치는 속성으로 구성되어 있으며 발생 일시, 발생 장소, 수온, 염분, 질소 함유량, 인 함유량으로 구성되어 있다.

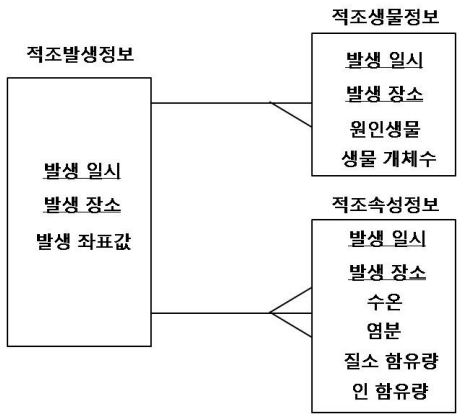

그림 4. 데이터 개체 관계도
Fig. 4. Data Entity Relationship Diagram

www.dbpia.co.kr

3.2 데이터 모델링

본 논문에서는 총 375 건의 데이터 중 262건의 데이터를 훈련데이터 집합(Training Data Set)으로 사용하였으며, 112건의 데이터를 최적 모델 선정을 위한 검증용 데이터 집합(Validation Data Set)으로 사용하고, 나머지 한 건의 데이터(실제 측정되어 입력된 데이터)를 입력 사례 데이터 집합(Test Data Set)으로 사용하였다. 훈련데이터 집합은 SVM의 경우에는 모델의 구축에 사용하였고 사례기반추론의 경우에는 사례베이스 구축에 사용하였다. 본 적조 사례기반추론 시스템의 모델 구축을 위해 사용한 입력 속성은 표 1과 같다.

사례기반추론 모델의 예측성능은 해를 생성하기 위해 참조되는 최근접 이웃의 수와 유사도 산출에 사용되는 속성들의 가중치로부터 많은 영향을 받게 된다. 본 실험에서 생물 개체수, 수온, 염분, 질소 함유량, 인 함유량은 수치형 데이터로써 각각의 속성은 적조 발생에 미치는 영향이 비슷하여 속성의 가중치는 동일하게 '1' 로 주었다. 원인 생물의 경우 속성 유형이 범주형이므로 속성값이 같으면 1을, 다르면 0을 부여하였다.

표 1. 데이터 입력 속성
Table 1. Data Input Attribute

입력 데이터	속성 유형
원인 생물	범주형
생물 개체수	수치형
수온	수치형
염분	수치형
질소 함유량	수치형
인 함유량	수치형

3.3 유사도 측정

입력 데이터로부터 모든 사례에 대하여 유사성을 계산하여 유사성이 가장 높은 사례를 선택하기 위하여 사례간의 유사도를 계산하는 방법은 다음과 같다. 새로운 입력 사례 N 과 사례 베이스에 있는 사례 O 간의 총 유사도 $S(N, O)$는 식 (1)과 같이 속성 i 별로 유사도 점수인 $f(N_i, O_i)$를 구하고, 각 속성의 가중치를 곱한 후 이를 총합하여 계산한다.

$$S(N,O)=\frac{\sum_{i=1}^{n} f(N_i,O_i)\times W_i}{\sum_{i=1}^{n} W_i} \quad (1)$$

여기에서, n은 속성의 개수, $f(N_i, O_i)$ 는 사례 N 과 O 의 i 속성 간의 유사도 점수, W_i 는 i 속성의 가중치를 의미한다. 사례간의 유사도는 식 (1)에 의하여 0에서 1사이의 실수 값으로 표현되는데, 0에 가까울수록 두 사례의 유사성이 낮다는 것을 의미하고, 1에 가까울수록 유사성이 높다는 것을 의미한다. 즉, 사례를 검색할 때 유사성이 1에 가장 가까운 사례를 추천한다. 속성 간의 유사도 점수인 $f(N_i, O_i)$ 는 속성의 유형(수치형, 범주형)에 따라 달라지는 데 본 논문에서는 입력 데이터가 모두 수치형이기 때문에 유사도 점수는 식 (2)와 같이 계산한다.

$$f(N_i,O_i)=1-\frac{a_i-b_i}{\max_i} \quad (2)$$

여기에서 a_i 는 N_i 값, b_i 는 O_i 값, $\max_i$ 는 사례베이스에 있는 i번째 속성 값 중 최대값을 의미한다. 일반적으로, 사례기반추론 시스템의 성능은 해를 생성하기 위해 참조되는 최근접 이웃의 수 k와 유사도 산출에 사용되는 속성들의 가중치로부터 많은 영향을 받게 된다. 본 논문에서는 최근접 이웃의 수 k를 증가시키면서 실험을 수행함으로 시스템을 최적화 하고자 한다. 사례 베이스에 있는 모든 사례에 대하여 총 유사도를 구한 다음 총 유사도가 가장 큰 사례부터 내림차순으로 정렬하여 우선순위 별로 적조 모니터링 시스템에 전송한다.

Ⅳ. 시스템 구현 결과

본 논문에서는 사례 기반 추론을 이용하여 적조 현상을 예측하기 위해서 사례 베이스를 구축하였다. 적조 데이터베이스는 적조발생정보, 적조속성정보, 적조생물정보의 세 개의 테이블로 구성되어 있으며 입력 데이터는 사례 검색, 사례 재사용, 사례 수정, 사례 유지 단계를 거쳐 구축된 사례 베이스에서 가장 유사한 과거 사례를 추론하여 진단하고 진단된 결과는 모니터링 시스템으로 전송된다. 그림 5는 제안된 시스템의 구성도이다.

본 논문에서는 델파이를 이용하여 적조 모니터링 시스템을 구현하였으며 표 2는 시스템 구현 환경을 나타낸다.

구현된 적조 모니터링 시스템은 발생 일자별, 발생 장소별 판단 결과를 모니터링 할 수 있다. 카운터에 나와 있는 숫자는 매분마다 데이터베이스에 새로운

www.dbpia.co.kr

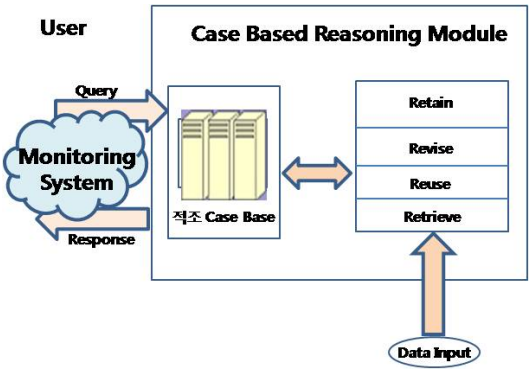

그림 5. 시스템 구성도
Figure 5. System block diagram

표 2. 시스템 구현 환경
Table 2. system implement environment

	항목	종류
소프트웨어	운영체제	Windows XP
	사용언어	Delphi
	DBMS	MSSQL
하드웨어	DB서버	Sqlserver 2000
	서버	Pentium(R) P.core2 Duo 1.6

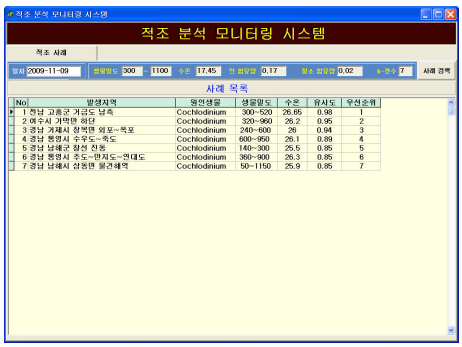

그림 6. 모니터링 시스템 구현 결과
Figure 6. Result of Monitoring System implementation

데이터가 추가 되었는가를 확인하고 새롭게 추가된 데이터가 있다면 시스템에서 판단할 수 있도록 매분 단위로 체크한다. 또한 유사 사례 k 개의 개수를 조절하여 검색되는 결과를 확인할 수 있다. 그림 6은 모니터링 시스템의 구현 결과이다.

V. 성능 평가

본 논문에서는 사례 베이스 구축 및 입력 데이터생성을 위해서 국립수산과학원에서 제공하는 적조정보 시스템에서 획득한 2008년 1월부터 2009년 12월까지 2년간의 적조 발생 데이터를 사용하였다. 사례 베이스를 구축하기 위하여 총 375 건의 데이터를 입력 받아 실험하였다. 본 논문에서 제안한 사례기반추론 모델의 유용성을 확인하기 위해서 분류 문제에 대표적으로 많이 사용되어온 기법인 SVM을 적용한 모델과 그 성능을 비교 평가하였다. 실험 결과에 대해 학습 데이터로부터의 영향을 최소화하고 신뢰성을 확보하기 위해 10-Fold 교차검증을 수행하였다. 실험 데이터는 사례베이스 구축용 데이터와 검증용 데이터로 구분되며 각각 7:3의 비율로 사용하였다. 본 실험에서 생물 개체수, 수온, 염분, 질소 함유량, 인 함유량은 수치형 데이터로써 각각의 속성은 적조 발생에 미치는 영향이 비슷하여 속성의 가중치는 동일하게 '1' 로 주었다. 원인 생물의 경우 속성 유형이 범주형이므로 속성값이 같으면 1을, 다르면 0을 부여하여 사례베이스를 만든 후에 Leave-One-Out의 방식으로 시스템을 최적화하였다. 표 3은 실험을 통해 얻은 검증용 데이터의 SVM 모델과 사례기반추론 모델의 데이터 정확도이다. SVM 모델에 대한 평균 정확도는 약 81.9%를 보였고 사례기반추론의 평균 정확도는 약 84.2%를 보임으로써 본 논문에서 제시한 사례기반추론 모델이 보다 더 높은 성능을 보였다.

또한 최적의 사례기반추론 시스템을 구축하기 위해서 유사도 분류 k 개수를 조정해 가면서 실험을 수행

표 3. 검증용 데이터 성능 평가
Table 3. Performance evaluation of data

Fold 번호	SVM	사례기반추론
1	82.8	85.2
2	81.6	87.4
3	80.3	83.3
4	79.7	82.9
5	78.8	81.2
6	83.3	86.5
7	83.7	85.8
8	85.1	84.4
9	81.3	82.7
10	82.8	83.1

표 4. k 개수에 따른 정확도
Table 4. Accuracy by number of k

k 개수	평균 정확도(%)
1	78.3
3	82.4
5	85.3
7	84.6
9	81.1

www.dbpia.co.kr

하였고 표 4는 k 개수에 따른 평균 정확도(단위 :%)에 대한 실험 결과이다. 실험 결과 k의 개수가 아주 낮거나 높은 경우에는 정확도가 상대적으로 감소함을 알 수 있다.

그림 7은 k 개수의 따른 정확도의 변화를 보여준다. k 개수가 5인 경우에 평균 정확도가 가장 높음을 알 수 있다.

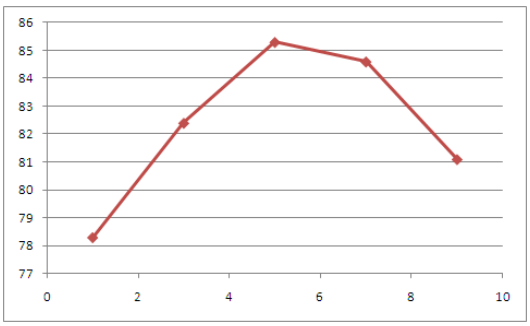

그림 7. k 개수에 따른 정확도 변화
Fig. 7. Accuracy change by number of k

Ⅵ. 결 론

우리나라 연안에는 유해성 적조의 발생으로 양식어장 피해가 빈번히 발생하고 있다. 적조 현상에 대한 판별, 예측 분석을 위한 시스템은 현재 개발이 아주 미흡한 상태이고 대부분의 적조 현상에 대한 연구는 화학 및 생물학적 원인의 규명에 대해 그 초점이 맞추어져 있어 지능적인 의사 결정 알고리즘을 갖는 시스템 구현이 필요하다. 이에 본 논문에서는 사례 기반 추론 기법을 이용하여 적조 현상에 관한 사례를 지식베이스로 구축하고 추론하는 시스템을 설계하였다. 사례 베이스를 구축하기 위하여 375 건의 데이터를 입력 받아 10-Fold 교차검증을 수행하였다. 또한, 사례기반추론 모델의 유용성을 확인하기 위해서 SVM을 적용한 모델과 그 성능을 비교 평가한 결과 SVM을 적용한 모델은 약 81.9%의 평균 정확도를 보였고 사례기반추론의 평균 정확도는 약 84.2%를 보임으로써 본 논문에서 제시한 사례기반추론 모델이 정확도 측면에서 보다 더 높은 성능을 보였다. 최적의 사례기반추론 시스템을 구축하기 위해서 유사도 분류 k 개수를 조정해 가면서 실험을 수행한 결과 k 개수가 아주 낮거나 높은 경우에 정확도가 감소함을 알 수 있었다. 적조 사례기반 시스템은 적조 현상에 대해 의사 결정을 지원함으로써 적조 현상 발생에 대해 효과적인 대처가 가능할 것이다. 향후에는 적조 현상에 대한 특징 선별 및 분류 알고리즘 기술과 해수 자동채취 제어 모듈 등에 대해서 연구가 필요할 것이다.

참 고 문 헌

〔1〕 윤홍주, 서영상, 정종철, 남광우, “한국 연안의 적조형성과 기상인자간의 통계적 해석”, 한국해양정보통신학회 제8권, 제4호, pp.9261~932, 2004.

〔2〕 G. Morcous, H. Rivard, A. M. Hanna, “Case-Based Reasoning System for Modeling Infrastructure Deterioration”, Journal of Computing in Civil Engineering, Vol.16, No.2, pp.104-114, 2002.

〔3〕 Gavin Finnie, Zhaohao Sun, “Similarity and metrics in case-based reasoning”, International Journal of Intelligent System, Vol.17, pp.273-287, 2002.

〔4〕 Tassan, S., “An Algorithm for the Detection of the White-Tide Phenomenon in the Adriatic Sea Using AVHRR data”, Remote Sen, ENviron.45 29-42, 1993.

〔5〕 Yoo, S. J. and J. C., Jeong, “Detecting Red Tides in Turbid Waters”, Journal of the Korean Society of Remote Sensing, Vol.15, No.4, pp. 321-327, 1999.

〔6〕 정종철, “AVHRR과 Landsat TM자료를 이용한 적조 패취관측”, 한국환경영향평가학회, 제10권 제1호, pp.1-8, 2001.

〔7〕 Limam. S., Marir. F., & Reijers. A., “Case-based reasoning as a technique for knowledge management in business process redesign.”, Electronic Journal on Knowledge Management, Vol.1. No.2, pp.124-133, 2003.

〔8〕 T. M. Cover and P. E. Hart, “Nearest Neighbor Pattern Classification,” IEEE Transactions on Information Theory, Vol.13, 1967.

〔9〕 E. E. Smith and D. L. Medin, “Categories and Concepts,” Cambridge, MA: Harvard University Press, 1981.

〔10〕 D. Aha, D. Kibler and M. Albert, “Instance-based Learning Algorithms,” Machine Learning, 6(1) pp.37-66, 1991.

〔11〕 Zhuang, D., Zhang, B., Yang, Q., Yan, J., Chen, Z., & Chen, Y. 2005. “Efficient Text Classification by Weighted Proximal SVM.”

www.dbpia.co.kr

Proceedings of the Fifth IEEE International Conference on Data Mining: 538-545.

송 병 호 (Byoung-Ho Song) 정회원

1998년 2월 조선대학교 전산통계학과
2000년 2월 조선대학교 전산통계학과 석사
2008년 2월 조선대학교 전산통계학과 이학박사
2009년 12월~현재 목포대학교 중점연구소 연구교수

<관심분야> 해양통신, 인공지능, 무선통신응용분야(RFID, USN), 생체인식시스템

정 민 아 (Min-A Jung) 정회원

1992년 2월 전남대학교 전산통계학과
1994년 2월 전남대학교 전산통계학과 석사
2002년 2월 전남대학교 전산통계학과 박사
2005년 3월~현재 목포대학교 컴퓨터공학과 조교수

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

이 성 로 (Seong-ro Lee) 정회원

1987년 2월 고려대학교 전자공학과
1990년 2월 한국과학기술원 전기및전자공학과 석사
1996년 8월 한국과학기술원 전기및전자공학과 박사
2005년 3월~현재 목포대학교 정보공학부 정보전공학전공 부교수

<관심분야> 디지털통신시스템, 이동 및 위성통신시스템, USN/텔레미틱스응용분야, 임베디드시스템, 생체인식시스템

www.dbpia.co.kr