
UChoo 알고리즘을 이용한 생물 조기 경고 시스템

이종찬* · 이원돈**

Biological Early Warning Systems using UChoo Algorithm

Jong Chan Lee* · Won Don Lee**

요 약

본 논문은 생물 조기 경고 시스템을 구현하기 위한 방법을 제안한다. 이 시스템은 모니터링 데몬을 이용해 간헐적으로 데이터 사건을 생성하고, 이 데이터 집합으로부터 특징 매개변수들을 추출한다. 특징 매개변수는 6개의 변수(x/y 축 좌표, 거리, 절대 거리, 각도, 프랙털 차원)를 가지고 유도된다. 특히 프랙털 이론을 사용해 제안 알고리즘은 입력된 특징들이 독성 환경에 있는지 아닌지의 유기물 특성을 정의한다. 추출된 특징 데이터를 학습하기 위한 적절한 알고리즘을 위해 기계학습 분야에서 널리 쓰이는 확장된 학습 알고리즘(UChoo)을 사용한다. 그리고 본 알고리즘은 특징 집합들이 모니터링 데몬에 의해 주기적으로 추가된다는 BEWS의 특징을 극복하기 위해 확장된 데이터 표현 방법을 이용하는 학습 방법을 포함한다. 이 알고리즘에서 결정트리 분류기는 확장된 데이터 표현에서 가중치 매개변수를 사용하는 부류 분포 정보를 정의 한다. 실험 결과들은 제안된 BEWS가 환경적인 독성을 탐지하는데 이용될 수 있음을 보인다.

ABSTRACT

This paper proposes a method to implement biological early warning systems(BEWS). This system generates periodically data event using a monitoring daemon and it extracts the feature parameters from this data sets. The feature parameters are derived with 6 variables, x/y coordinates, distance, absolute distance, angle, and fractal dimension. Specially by using the fractal dimension theory, the proposed algorithm define the input features represent the organism characteristics in non-toxic or toxic environment. And to find a moderate algorithm for learning the extracted feature data, the system uses an extended learning algorithm(UChoo) popularly used in machine learning. And this algorithm includes a learning method with the extended data expression to overcome the BEWS environment which the feature sets added periodically by a monitoring daemon. In this algorithm, decision tree classifier define class distribution information using the weight parameter in the extended data expression. Experimental results show that the proposed BEWS is available for environmental toxicity detection.

키워드

BEWS, 프랙털 차원, 확장된 데이터 표현, 가중치, UChoo.

Key word

BEWS, Fractal Dimension, Extended Data Expression, Weight, UChoo.

* 종신회원 : 청운대학교 인터넷학과 (jlee@chungwoon.ac.kr)

접수일자 : 2011. 08. 24

** 정회원 : 충남대학교 전기정보통신공학부 컴퓨터전공

심사완료일자 : 2011. 10. 07

I. 서 론

갑작스런 수질의 오염에 따르는 건강의 치명적 위험이 강조되면서, 최근 하천수의 수질모니터링 및 독성평가를 위해 널리 사용될 수 있는 생물 조기 경보 시스템(Biological Early Warning System, BEWS)의 개발이 각광을 받고 있다[1][2][3].

수질을 파악할 때 사용하는 기존의 화학적인 방법은 샘플을 수집하여 분석하는 방법이었다. 이 방법은 샘플을 수집하는 시기의 수질 정보만 알 수 있으며, 사용자가 정의하지 않은 물질을 탐지해 낼 수 없는 방법론적인 한계가 있다. 이에 반해 BEWS는 수질 내 유기물 전체의 생물학적 반응을 지속적으로 감시하여 독성을 탐지한다. 따라서 수질 상태를 실시간으로 파악할 수 있고, 기존의 화학적 방법으로 발견하지 못했던 독성을 감지할 수 있는 등 많은 장점이 있다. 특히 이는 어떠한 환경에서 간헐적인 독성 사건을 찾아내는데 유용하다.

이러한 BEWS가 성공적으로 적용되기 위해서는 수질의 오염 발생을 신뢰성 있게 예측할 수 있어야 한다. 이를 위해 본 논문은 첫째, 수질을 연속적으로 감시하여 독성에 따라 변화하는 생물체의 특징을 추출하는 방법을 제안한다. 둘째, 추출된 데이터를 학습할 수 있는 알고리즘을 제안한다. 이 알고리즘은 기계 학습(machine learning) 알고리즘들[4][5] 중에 널리 쓰이는 C4.5[6]를 본 시스템에 목적에 맞도록 변형하여, 임의의 수질로부터 얻어진 특징 데이터에 독성이 포함되어 있는지를 판별해 내는 목적으로 이용된다.

특징 추출과정은 물고기 모니터 시스템으로부터 물고기의 동작 특징을 얻어 수생 시스템에 대한 BEWS를 구축한다. 이 시스템은 카메라로 물고기의 움직임과 행위를 모니터링하는 과정에서, 깨끗한 물과 오염된 물에서 물고기의 행동을 관찰하며 계속적으로 모아진 데이터를 기록하고 이를 각각 저장한다. 이 시스템은 훈련과 테스트 과정으로 나누는데, 훈련 과정에서는 독성의 정도에 따라 부류(class)가 정해지고 테스트 과정에서 오염과 독성의 존재 여부를 평가하는데 사용한다. 속성들이 생물체의 생리적이고 행위적인 응답을 대표할 수 있어야 하기 때문에, BEWS에서 어떻게 속성들을 추출하느냐가 가장 중요한 문제가 된다. 이를 위해 본 논문에서는 거리, 절대거리, 각도, 프랙털(fractal) 차원을 사용한다.

분류 알고리즘은 결정트리(decision tree), 규칙기저(rule-based), 신경회로망, 베이지 분류기가 있다. 이들 중에 C4.5 등의 결정트리를 이용하는 방법은 짧은 시간에 많은 데이터를 신뢰성 있게 처리하고 이해하기 쉽다는 장점들이 알려져 있어 BEWS의 분류기로 적합하다.

정보와 관련된 분류 알고리즘은 기존의 방대한 데이터는 물론 실시간으로 수집되는 새로운 데이터 또한 계속적으로 반영할 수 있어야 한다. 기존의 전형적인 분류 알고리즘들은 이러한 작업을 수행할 때, 새로운 데이터와 전의 데이터를 모두 사용하여 분류기를 갱신해 왔다. 이러한 방식은 적은 데이터를 추가하기 위해 큰 기존 데이터를 다시 학습해야 하는 비효율적인 측면뿐만 아니라, 만약 기존 데이터를 사용할 수 없는 경우, 더 이상 새로운 데이터를 추가하여 갱신을 할 수 없게 된다는 단점이 있다[7][8].

이를 보완한 것이 점진적 학습법[11][12]이다. 점진적 학습법은 기존의 데이터로 만들어진 규칙에 새로운 데이터로 만들어진 규칙을 합성하여 갱신한다. 만들어진 규칙을 사용하기에 기존의 데이터를 학습할 필요가 없어서 효율적이며, 기존의 데이터를 잃어버릴 경우에도 규칙만 있으면 새로운 데이터로 만든 규칙과 합성하여 분류기를 갱신할 수 있다. 이러한 점진적 학습법을 이용하는데 많은 장점을 가지고 있는 알고리즘으로 C4.5가 있다. 따라서 본 논문은 추출된 특징 데이터를 학습하기 위한 알고리즘으로 C4.5를 데이터 각각에 가중치를 부여할 수 있도록 변형하여 사용하는 방법을 제안한다.

II. 배경

2.1. 프랙털 차원

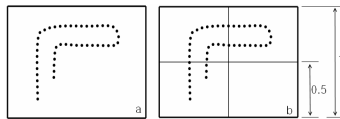
프랙털 기하학에서 프랙털 차원, D 는 객체를 좀 더 작은 비율로 축소함에 따라 프랙털이 얼마나 완벽하게 공간을 채우는지를 나타내는 통계적인 양이다. 이러한 프랙털 차원을 정의하는 방법이 여러 가지가 있는데, 이들 중 박스 수 계산(box-counting) 차원 방법이 실제로 널리 쓰인다[9]. 박스 수 계산 방법은 컴퓨터로 구현하기가 쉽고 복잡한 이미지에 적용될 수 있다는 장점을 가지고 있다. 이는 사용된 비율 요소에 따라 프랙털 차원을 결정한다. 박스 수 계산 프랙털 차원은 다음 식과 같은 근사치인 하우스도르프 차원으로부터 유도된다.

$$D \approx \frac{\log(N(s))}{\log(1/s)} \quad (1)$$

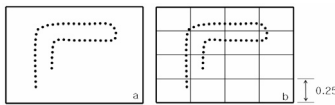
$$D = \lim_{s \rightarrow 0} \log \frac{N(s)}{\log(1/s)}$$

여기서 s 는 박스의 한 면 길이를, $N(s)$ 은 박스의 개수를 의미한다. 이러한 극한이 존재하면 측정된 객체의 박스 수 계산 차원이라 부른다. 이러한 극한은 항상 느리게 수렴하므로 이를 (2) 식과 같은 대체 해를 사용한다.

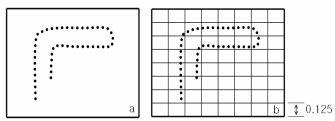
$$\log(N(s)) = D \cdot \log(1/s) \quad (2)$$



(a)



(b)



(c)

그림 1. 차원 값의 산출 예

(a) $s_1=0.5$, $N(s_1)=3$ $D=1.58$ (b) $s_2=0.25$, $N(s_2)=11$ $D=1.73$ (c) $s_3=0.125$ $N(s_3)=19$ $D=1.46$

Fig. 1 The example of fractal dimension

(a) $s_1=0.5$, $N(s_1)=3$ $D=1.58$ (b) $s_2=0.25$, $N(s_2)=11$ $D=1.73$ (c) $s_3=0.125$ $N(s_3)=19$ $D=1.46$

그림 1은 점선이 물고기의 움직임이라고 가정할 때, 박스의 한 면의 길이를 나누는 비율을 2, 4, 8이라 하고 D 값을 산출하는 예를 보이고 있다. 그림 2는 그림 1의 예를 가지고 근사 법을 이용해 얻어진 직선의 기울기가 프랙털 차원을 말한다는 것을 보이고 있다.

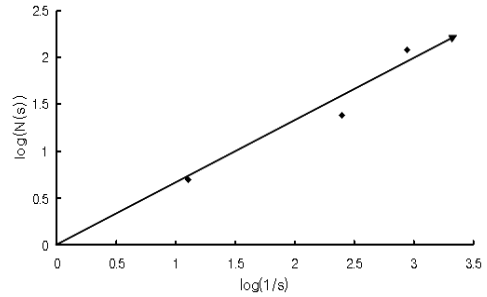


그림 2. 프랙털 차원
Fig. 2 Fractal dimension

2.2. 분류 알고리즘

분류기는 속성 값들이 주어졌을 때 이에 대한 부류를 예측하기 위해 사용하는데, 결정 트리는 분류기에서 널리 이용되는 기법이다. C4.5는 최소의 메모리로 빠르고 좋은 성능을 가진 것으로 알려진 분류 알고리즘이다. UChoo는 이러한 C4.5로부터 발전된 알고리즘이다.

표 1이 C4.5에서 사용되는 일반적인 훈련 데이터라고 한다면, 표 2는 표 1을 UChoo의 표현 방식으로 확장한 예이다. 여기서 각 속성 값들은 0과 1사이의 확률 값으로 채워진다.

표 1. 훈련 데이터
Table. 1 Training Data

Outlook	Temp(°F)	Windy?	부류
Sunny	70	False	Don't Play
Sunny	60	True	Play
Overcast	80	False	Play
Rain	60	True	Don't Play
Rain	70	False	Play
Rain	80	True	Don't Play

예를 들어 표 1의 사건 1에서 Windy?와 부류 값이 소실되었다고 가정한다면, 표 2의 사건 1에서 Windy?는

표 2. 확장된 데이터 표현의 예
Table. 2 The Example of extended data expression

사건 #	가중치	Outlook			Temp(°F)			Windy?		부류	
		sunny	overcast	rain	60	70	80	True	False	Play	Don't Play
1	20	1	0	0	0	1	0	1/2	1/2	2/3	1/3
2	1	1	0	0	1	0	0	1	0	1	0
3	1	0	1	0	0	0	1	0	1	1	0
4	1	0	0	1	1	0	0	1	0	0	1
5	1	0	0	1	0	1	0	0	1	1	0
6	1	0	0	1	0	0	1	1	0	0	1

True와 False에 대해 1/2씩 동일하고, 부류가 Play일 확률은 2/3이고 Don't Play일 확률은 1/3이라고 채워 넣을 수 있다는 의미이다. 또한 각 사건은 가중치 값을 가지게 된다. 이것은 그 사건이 얼마만큼의 중요도를 가지는가를 나타낸다. 일반적인 레코드의 중요도가 1이라고 보았을 때, 가중치 20인 사건은 다른 가중치 1인 레코드의 20개에 해당하는 중요도를 가진 사건이라는 뜻이 된다. 따라서 전체의 사건의 개수와 레코드들의 개수는 서로 다른 값일 수 있다. 표 2에서 전체 레코드는 24개인데 반해, 전체 사건 개수는 5개이다.

이러한 확장된 데이터 표현을 사용하는 UChoo 분류기는 데이터의 일부나 부류가 손실된 불완전한 데이터를 처리할 수 있다는 장점을 가진다. 예를 들어 표 2의 사건 1에서 Windy?가 손실되었다고 했을 때 True와 False에 각각 1/2씩 할당하고 학습을 할 수 있다. 이는 일부분의 손실에 대해 이 사건을 학습에서 제외하는 방법에 비해 정보의 손실을 줄임으로서 보다 정확한 학습을 할 수 있도록 한다.

C4.5 알고리즘에서는 각 속성(attribute, A)들에 대해 정보 이득(gain_ratio) 함수를 계산한다[6]. 정보 이득 함수는 해당 속성 값에 따라 분류할 때 부류들이 얼마나 잘 나누어지는지를 측정하는 값이며, 이들 중 가장 큰 정보 이득을 가진 속성을 선택하여 이 속성 값에 따라 하부 노드로 분기한다. 이와 같은 방법을 모든 부류가 분류될 때까지 반복하면 데이터 집합에 따르는 의사 결정 트리를 얻을 수 있다.

C4.5에서 가중치를 가지고 학습할 수 있도록 변형이 필요하다. 즉 표 2의 확장된 데이터를 학습할 수 있도록 알고리즘이 변경되어야 한다. C4.5와 달라진 정의는 다

음과 같다.

부류의 소속 값 : $C_1(m), C_2(m), C_3(m), \dots, C_{k-1}(m), C_k(m)$

$C_i(m)$ 는 m번째 사건이 C_i 부류에 속한 정도를 나타낸다. 예를 들어 표 2에서 $C_{play}(1)=2/3$ 이다. 여기서 i는 부류 값이고, $\sum_{i=1}^k C_i(m) = 1$ 이다.

속성의 소속 값 : $O_{A_1}(m), O_{A_2}(m), \dots, O_{A_{n-1}}(m), O_{A_n}(m)$

$O_{A_j}(m)$ 는 m번째 사건의 속성 A에서의 속성 값 j가 가지는 값을 말한다. 예를 들어 표 2에서 $O_{Windy?true}(1)=1/2$, $O_{Windy?false}(1)=1/2$ 이다. 여기서 $\sum_{j=1}^n O_{A_j}(m) = 1$ 이다.

Weight(m,S) : 집합 S에서 m번째 사건의 가중치 값
freq(C_i, S) : 집합 S안에서 부류 C_i 에 속해있는 사건들의 개수인데 이 경우에는 (3)식과 같이 나타내 진다.

$$freq(C_i, S) = \sum_{m=1}^{|S|} Weight(m, S) \cdot C_i(m) \quad (3)$$

마찬가지로 freq(C_i, S_{A_j})는 S_{A_j} 안에서 부류 C_i 에 속해있는 사건들의 개수인데 이 경우에는 (4)식과 같이 나타내 진다.

$$freq(C_i, S_{A_j}) = \sum_{m=1}^{|S|} Weight(m, S) \cdot C_i(m) \cdot O_{A_j}(m) \quad (4)$$

$|S_{A_j}|$: S_{A_j} 사건들의 개수. 이 경우에도 위와 마찬가지로

집합 S_{A_j} 에 속해 있는 사건들 각각의 속성의 소속 값 $O_{A_j}(m)$ 에 $Weight(m, S_{A_j})$ 을 곱한 후 그것을 모두 더하여 아래와 같이 계산한다.

$$|S_{A_j}| = \sum_{j=1}^{|S_A|} Weight(m, S_{A_j}) \cdot S_{A_j}(m)$$

따라서 새로 정의된 값들에 의하여 C4.5의 엔트로피 식을 사용하여 그 노드에서 가장 큰 이득 비율(Gain Ratio)를 가진 속성을 결정 할 수 있다. 기존의 boosting은 기본 분류기가 분류하기 어려운 사건을 집중적으로 다음 라운드에 선택하여 예측 확률을 높이는 방식이다. 단점은 다음 라운드에서 선택이 되지 못한 사건들에 대한 정보는 분류기가 알 수 없다는 점이다. UChoo는 사건의 가중치를 부여할 수 있으므로 잘 분류가 되는 사건들은 적은 가중치를 부여한 채 분류기에 포함시킴으로써, 정보의 누락이 줄어들어 분류기의 성능을 높일 수 있다[10].

III. 새로운 BEWS

3.1. 특징 추출

생물 조기 경보 시스템은 주어진 어항에 물고기를 넣고 그 물고기가 행동하는 것을 관찰하여 비정상적인 운동의 경우 경보를 발생시키는 시스템이다. 적절한 특성을 추출하는 문제는 분류를 하기 위해 가장 중요한 부분이다. 이는 분류의 시간을 단축할 수 있음은 물론 성능에도 많은 영향을 미치기 때문이다. 일반적으로 물고기는 매우 빨리 움직인다. 따라서 주어진 T 시간 동안 m 지점을 방문한다고 가정한다면 T를 샘플링 시간이라고 한다. T 시간 동안 물고기의 행동궤적을 수치로 정량화하기 위하여 그 궤적을 나타낼 수 있는 변수(attribute)들을 다음과 같이 정의한다.

- ◆ X/Y 좌표 : 물고기가 이동한 지점의 위치 좌표
- ◆ 거리/절대거리 : 거리는 현재 샘플 시간 t와 다음 시간 t+1 간에 이동한 유클리드 거리를 의미하며, 절대 거리는 현재 위치(t)와 처음 위치(t0) 간의 거리를 의미한다.
- ◆ 각도 : 움직인 각도를 탄젠트 값으로 표시한 값.
- ◆ 프랙털

변수들은 x좌표, y좌표, 거리, 절대 거리, 각도, 프랙털 차원으로 6개이다. 그런 변수들로 이루어진 벡터들로 물고기의 행동이 묘사되고 그 벡터들을 사용하여 규칙을 만들게 된다. 이러한 변수 중의 하나로 프랙털이 있다.

어항에 있는 물고기는 정면으로 관찰되어지므로 2차원의 평면위에서 움직이는 궤적을 보인다. 주어진 시간에 얼마만큼이나 활발하게 행동하느냐는 것을 표현하는 데에 프랙털 변수가 쓰인다. 프랙털 도형에서 프랙털 차원 D를 계산하기 위하여 여러 가지 방법이 제안되었는데 그 중에 상자 수 계산(box counting), 상관관계(correlation) 차원이 일반적으로 널리 사용된다.

본 논문에서는 프랙털 변수의 값을 상자 수 계산 방법에 기초하여 산출한다. 그림 1에서 보는 것과 같이 물고기가 움직인 궤적이 지나가는 상자 수를 재귀적으로 계산하게 되면 그림 2와 같이 프랙털 값에 수렴하게 된다. 만약에 물고기의 행동이 매우 활발하여 주어진 시간에 모든 박스를 다 지나가게 된다면 프랙털 값은 2차원 평면상에서 가질 수 있는 이상적 최대 값인 2가 될 것이다. 그러나 대부분의 경우엔 그보다 작은 0과 2 사이의 값을 가지게 된다. 프랙털 차원 값을 계산 하는 예를 들면 다음과 같다.

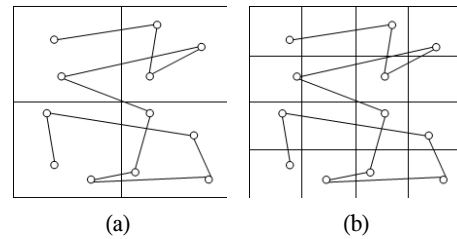


그림 3. 윈도우 분할
Fig. 3 Windows division

그림 3에서 각 포인트는 물고기의 움직임을 나타내고 있다. 여기서 (a)는 윈도우를 4셀로 나누는데 각 셀의 포인트 수는 각각 2, 3, 3, 4이다. 이 경우의 프랙털 차원은 $\log_{12}/\log_4=1.79$ 이다. 같은 방법으로 16, 32셀에 대해 프랙털 차원을 계산한 다음, 이 값의 아크 탄젠트(arc tangent) 값을 계산한다. 마지막으로 이 값을 모두 더한 후 평균값을 구해 최종 프랙털 차원 값으로 정한다. 이로서 프랙털 차원은 시간 T 동안 물고기의 행위 특성을 설명한다.

3.2. BEWS의 구조

제안한 BEWS 시스템은 그림 4와 같이 훈련과정과 테스트과정으로 나누어진다. 두 과정 모두는 물고기의 움직임을 캡처하고 이를 기록하기 위한 모니터 데몬을 필요로 한다.

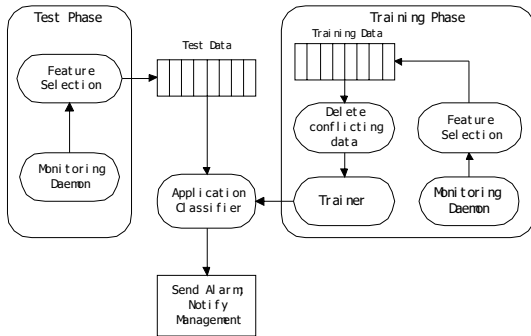


그림 4. BEWS 구조
Fig. 4 BEWS structure

특징 추출은 모니터 디바이스로부터 얻어진 데이터를 분석하기 위해 사용되고 분류기를 위해 새로운 데이터에 추가된다. 새로운 데이터로부터 생성된 분류기는 실시간으로 독성을 검출하는데 적용할 수 있다. 훈련과정에서 데이터는 깨끗한 물과 오염된 물로부터 모아진다. 독성을 물고기가 살고 있는 물에 넣고 물고기의 움직임을 기록한다. 부류들은 부류 0이 깨끗한 물이고 물의 오염된 정도에 따라 부류 값을 할당한다. 테스트과정에서 데이터는 모니터 디바이스로부터 캡처되고 훈련과정에서 얻어진 분류기에 입력된다. 안정된 결과를 위해 몇몇 샘플시간이 포함된 알람시간을 정한다. 각 샘플시간 동안 데이터가 수집된다. 각 알람시간에 출력 값의 평균이 계산되고 이 값으로 환경이 오염되었는지를 결정한다.

IV. 실험

물고기는 최소 2일 정도 간혀진 환경에서 익숙하게 한 후 실험에 사용한다. 독성 환경을 실험하기 위해 서로 다른 양의 중금속(구리, 납, 칼슘이온 용액)들을 물에 주입되었다. 여기서 4개의 부류가 정의되는데, 부류

0은 깨끗한 물을, 부류 1은 0.5ppm, 부류 2는 1ppm, 부류 3은 1.5ppm의 칼슘이온 용액이 각각 투입된 물을 말한다. 물고기는 환경에 적응하기 위해 12시간 동안 깨끗한 물에 있게 한 다음 부류 1의 데이터를 캡처한다. 다음 0.5ppm의 칼슘이온 용액을 깨끗한 물에 투입한 후 10분을 기다리고 부류 1의 데이터를 캡처한다. 1시간 후에 물에 다시 칼슘이온 용액 1ppm을 투여한 후 데이터를 기록한다.

실험은 깨끗한 물과 1ppm의 구리이온 용액을 첨가한 독성이 포함된 물에서 이루어졌다. 총 데이터는 3400개이고, 이들 중 1200개는 깨끗한 물의 데이터이고 나머지는 독성이 포함된 데이터이다. 그림 5는 가로축의 샘플 데이터에 따라 세로축의 특성 값들의 변화를 표시하고 있다. 그림으로부터 1200 데이터에서 독성이 주입된 이후 특성 값들이 변화하고 있는 것을 확인할 수 있다.

다른 실험으로 4개의 부류에 대해 데이터의 수를 균등하게 할당하여 총 24,000개의 데이터를 대상으로 하였다. 이들 데이터에 10-fold cross validation을 적용한다. 즉 데이터를 총 10개의 블록으로 나눈 후 9개의 블록은 훈련 데이터로 사용하고 나머지 블록을 테스트 데이터로 사용하였다. 이러한 방법의 실험을 10번 실행하여 평균 값을 구한다. 그 결과가 표 3에 나타나 있다. 여기서 에러는 물고기가 생물체로서의 한계와 본 논문에서 사용하고 있는 특성 값 이외에 밝혀지지 않은 다른 특성 값이 필요할 수 있음을 의미한다.

표 3. 테스트 에러 비율
Table. 3 Test error rate

테스트 횟수	에러 비율(%)
1	21.67
2	18.91
3	16.51
4	20.68
5	18.04
6	17.26
7	16.79
8	17.55
9	20.90
10	16.95
평균	18.53

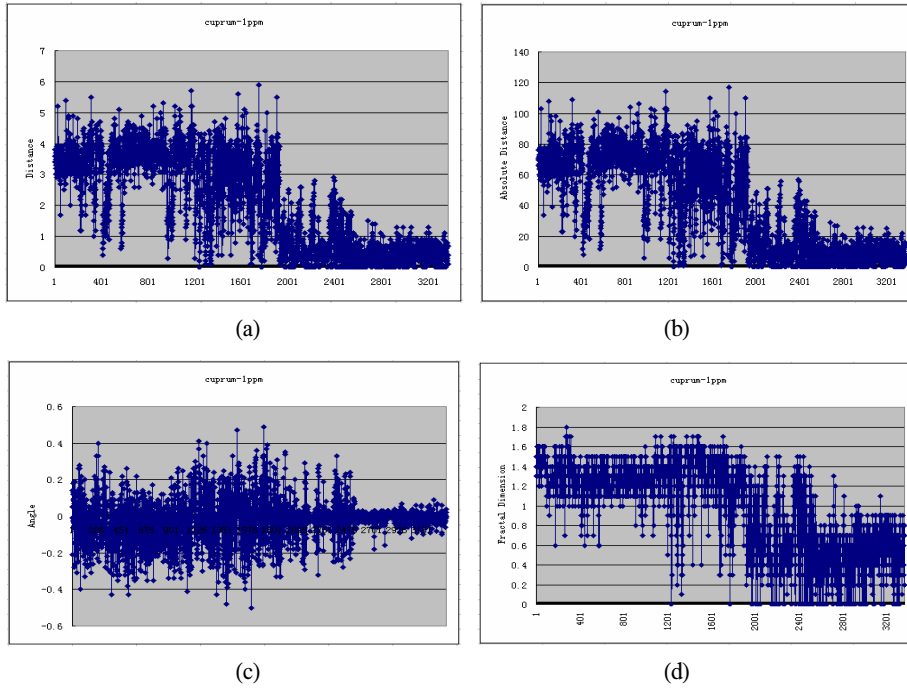


그림 5. 특성 값. (a) 거리 (b) 절대 거리 (c) 각도 (d) 프랙털 차원
 Fig. 5 Feature values. (a) Distance (b) Absolute distance (c) Angle (d) Fractal dimension

독성을 찾아내는 방법으로 테스트 데이터 2400개들에 대해 10-fold cross validation을 적용한 결과가 그림 6에 나타나 있다. 여기서 가로축은 테스트 데이터의 수를 말하고, 세로축은 알람 수준을 말하는 것으로 독성 알람 값이 1이상이면 오염된 것으로 정한다.

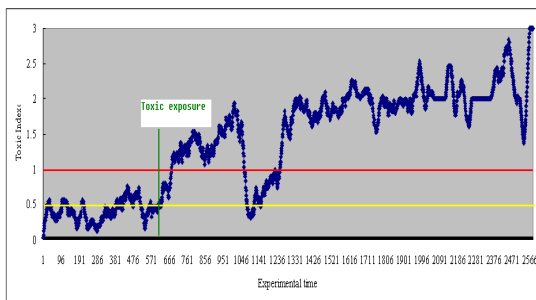


그림 6. 실험 결과
 Fig. 6 Experiment result

V. 결 론

BEWS는 독성을 조기에 찾아내는데 많은 장점들을 가지고 있다. 이 시스템은 화학적 모니터링 기술에 중요한 보완점을 제공한다. 적절한 검색의 한계와 방법적인 한계를 가지고 있는 분석적 화학 기술이 찾아낼 수 없는 물질들을 찾아낸다. 전통적인 독성 테스트나 특정 화학 센서들은 수생의 시스템에서 독성 사건에 실시간에 포괄적인 정보를 제공할 수 없다.

본 논문에서는 새로운 생물적 조기 경보 시스템에 적용될 수 있는 분류 방법을 제안하였다. 이 과정에서 효율적인 특성 추출 방법을 제안하였다. 실험을 통해 오염된 물에서 자라는 물고기의 움직임에 관찰하여 데이터를 만들고 이를 분류기로 분류함으로써 구현될 수 있음을 보였다.

참고문헌

- [1] B. W. Scharf, "A fish test alarm device for the continual recording of acute toxic substances in water", Arch. Hydrobiol. 85, pp250 - 256, 1979.
- [2] P. Schmitz, F. Krebs, U. Urmer, "Development, testing and implementation of automated biotests for the monitoring of the Rhine River, demonstrated by bacteria and algae tests.", Water Sci.Technol. 29 (3), pp215 - 221, 1994.
- [3] W. H. van der Schalie, T. R. Shedd, P. L. Knechtges, M. W. Widder, "Using higher organisms in biological early warning systems for real-time toxicity detection", Elsevier Biosensor & Bioelectronics 16, pp457-465, 2001.
- [4] P.N.Tan, M.Steinbach, V. Kumar, Introduction to data mining, Addison Wesley, 2005
- [5] M. Kantardzic, Data Mining : Concepts, Models, Methods, and Algorithms, Wiley-IEEE Press, 2002.
- [6] J. R. Quinlan, C4.5 : Program for Machine Learning, Morgan Kaufmann, 1993.
- [7] D. H. Kim, D. H. Lee, W. D. Lee, "Classifier using Extended Data Expression", IEEE Mountain Workshop on Adaptive and Learning Systems. pp.154-159, July 2006.
- [8] J.C.Lee, D.H.Seo, C.H.Song, W.D.Lee, "FLDF based Decision Tree using Extended Data Expression", The 6th International Conference on Machine Learning and Cybernetics, Hong Kong, pp3478-3483, August 2007.
- [9] M. J. Turner, J. M. Blackledge, P. R. Andrews, Fractal Geometry in Digital Imaging, Academic Press, pp59-73, 1998
- [10] R. Polikar, "Bootstrap-Inspired Techniques in Computational Intelligence", IEEE Signal Processing Magazine, pp.59-72, July 2007.
- [11] Y. R. Li, D. H. Seo, W. D. Lee "A New Classification Application of Biological Early Warning Systems for Toxicity Detection" CSA2008, pp.239-242, October 2008.

- [12] S. Y. Kim, K. Y. Kwon, W. D. Lee, "A Biological Early Warning System for Toxicity Detection" NCM2009, pp.1157-1160, August 2009.

저자소개

이종찬(Jong Chan Lee)



1988년 충남대학교(학사)
1990년 충남대학교 대학원(석사)
1996년 충남대학교 대학원(박사)
2006년~현재 : 청운대학교
인터넷학과 교수

※ 관심분야 : 신경회로망, 패턴분류, 데이터압축

이원돈(Won Don Lee)



1979년 서울대학교(학사)
1982년 U. of Illinois 대학원(석사)
1986년 U. of Illinois 대학원(박사)
1987년~현재 : 충남대학교
전기정보통신공학부 교수

※ 관심분야 : 신경회로망, 멀티미디어, 데이터마이닝