

## 예쁜꼬마선충의 수영 행동 영상과 기계학습 모델을 이용한 수질 오염 물질 구분 방법

강승호<sup>1</sup> · 정인선<sup>2</sup> · 임형석<sup>3\*</sup>

### A Method for the Classification of Water Pollutants using Machine Learning Model with Swimming Activities Videos of *Caenorhabditis elegans*

Seung-Ho Kang<sup>1</sup> · In-Seon Jeong<sup>2</sup> · Hyeong-Seok Lim<sup>3\*</sup>

<sup>1</sup>Associate Professor, Department of Information Security, Dongshin University, Naju, 58245 Korea

<sup>2</sup>PostDoc, Department of Software Engineering, Chonnam National University, Gwangju, 61186 Korea

<sup>3\*</sup>Professor, Department of Software Engineering, Chonnam National University, Gwangju, 61186 Korea

#### 요 약

예쁜꼬마선충(*Caenorhabditis elegans*)은 염기서열이 완전히 밝혀진 동물로 유전자 기능 분석, 동물 행동 연구 등 다양한 연구 분야에 사용되는 대표적인 생물 종이다. 그동안 선충을 이용해 물의 오염 여부를 판별하기 위한 바이오 모니터링 시스템에 대한 여러 연구들이 있었다. 본 논문은 하천의 수질 오염의 원인이 되는 화학물질을 식별하기 위해 선충의 수영 행동이 활용 가능한 지를 보여주기 위해 기계학습 기반의 바이오 모니터링 시스템을 제안한다. 선충의 수영 행동을 대표하기 위해 선충을 대상으로 가지 길이 유사성(Branch Length Similarity) 엔트로피를 계산한다. 그리고 BLS 엔트로피의 조합인 BLS 엔트로피 프로파일을 클러스터링 알고리즘을 사용해 몇 가지 패턴으로 유형화하여 데이터 집합을 만든다. 0.1ppm 농도의 포름알데히드, 벤젠, 톨루엔이 첨가된 아레나에서 선충의 수영 행동을 촬영하고 개발한 히든 마코프 모델(Hidden Markov Model: HMM)의 성능을 검증한다.

#### ABSTRACT

*Caenorhabditis elegans* whose DNA sequence was completely identified is a representative species used in various research fields such as gene functional analysis and animal behavioral research. In the mean time, many researches on the bio-monitoring system to determine whether water is contaminated or not by using the swimming activities of nematodes. In this paper, we show the possibility of using the swimming activities of *C. elegans* in the development of a machine learning based bio-monitoring system which identifies chemicals that cause water pollution. To characterize swimming activities of nematode, BLS entropy is computed for the nematode in a frame. And, BLS entropy profile, an assembly of entropies, are classified into several patterns using clustering algorithms. Finally these patterns are used to construct data sets. We recorded images of swimming behavior of nematodes in the arenas in which formaldehyde, benzene and toluene were added at a concentration of 0.1 ppm, respectively, and evaluate the performance of the developed HMM.

**키워드** : 바이오 모니터링 시스템, 예쁜꼬마선충, 히든 마코프 모델, 기계학습, 수질 오염

**Keywords** : Bio-monitoring system, *Caenorhabditis elegans*, Hidden Markov model, Machine learning, Water pollution

Received 9 June 2021, Revised 11 June 2021, Accepted 23 June 2021

\* **Corresponding Author** Hyeong-Seok Lim(E-mail:hslim@chonnam.ac.kr, Tel:+82-62-530-3426)

Professor, Department of Software Engineering, Chonnam National University, Gwangju, 61186 Korea

**Open Access** <http://doi.org/10.6109/jkiice.2021.25.7.903>

print ISSN: 2234-4772 online ISSN: 2288-4165

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서론

현대 산업과 농업의 발달은 인간 생활의 편의와 함께 다양한 오염 물질의 발생을 초래했다. 특히 배출된 오염 물질의 최종 종착지가 하수나 바다라는 점에서 수질 오염에 대한 관심은 최근 더해지고 있다. 특히 물 부족 국가에 속하는 대한민국의 경우 수질 오염으로 인한 피해는 국가 안보와도 직결되어 있다는 점에서 안보 차원에서 다루어져야 할 사안이기도 하다.

수질 오염은 오염 물질의 종류나 농도에 따라 다양한 형태를 가지기 때문에 이를 오염 원인의 발생 즉시 판별하기가 쉽지 않다. 특히 오염 물질의 종류나 농도 등을 가능한 이른 시간 안에 정확히 특정 지어야 피해 규모를 최소화하고 적절한 대처와 함께 책임 소재를 명확히 할 수 있다.

오랫동안 대부분의 수질 오염 판별에 사용되고 있는 화학적 방법은 오염 물질이 배출된 이후 생화학적 산소 요구량(BOD)이나 화학적 산소요구량(COD)을 전문가가 측정 장비를 이용해 오염 여부를 판단한다는 점에서 측정할 수 있는 오염 물질의 종류가 제한적이며 많은 시간과 경비를 요구하고, 전문가만이 평가가 가능하다는 한계가 있다.

따라서 이러한 화학적 측정 방법의 단점을 극복하기 위한 새로운 측정 방법의 필요성이 제기되었다. 이 중 특정한 화학물질에 대해 정확하고 예민하게 반응하는 능력이 있는 어류[1-3]나 깔다구[4], 물벼룩[5]을 검지 소자(bio-indicator)로 사용하여 모니터링하고 경보를 발생하는 바이오 모니터링 시스템에 대한 연구가 활발히 진행되어 왔다.

한편 디지털 영상 장비 및 영상 처리 기술의 발전으로 생물 종의 행동을 촬영해 특징을 추출하고 이를 기계학습 방법과 연계해 수질 오염을 측정하는 연구들도 제시되었다. [1]은 살충제 성분인 0.1mg/l 농도의 다이아지논을 투여하기 전후의 물속에서 메다카(*Oryzias Latipes*)의 수영 행동을 촬영하고 자기 조직화 맵(SOM: Self-organizing map)과 2차원 빠른 푸리에 변환(2D FFD)을 이용해 수영 행동의 변화를 확인하였다. [2]는 표백제가 투여된 물속에서 제브라피시(*Danio rerio*)의 수영 행동을 이용해 수질 변화 검출 연구를 수행하였고, [6]은 독소 물질인 황산구리에 대한 블랙웜(*Lumbriculus variegatus*)의 몸길이 변화를 이용해 수질 오염 여부를

판단하는 연구를 수행하였다.

특히 완전한 염기서열 및 유전자 구조가 밝혀진 예쁜꼬마선충(*Caenorhabditis Elegance*)은 유전자 수준에서 행동 수준까지 다양한 연구 대상이 되고 있으며, 환경 화학물질의 영향 및 검지 소자로서의 가능성에 대한 연구가 많은 관심을 불러일으키고 있다 [7-9]. [9]는 촬영된 예쁜꼬마선충의 수영 행동을 대상으로 Hidden Markov Model(HMM)을 사용하여 농도 0.1ppm인 포름알데히드(Formaldehyde)의 오염 여부를 판별하는 연구를 수행하여 선충을 이용한 바이오 모니터링 시스템의 개발 가능성을 보여줬다.

다만, [9]를 비롯한 대부분의 연구는 특정 농도에서의 특정 환경 물질의 투여 전후를 비교하는 연구에 한정되었다. 따라서 환경 물질(오염 물질)의 종류에 따른 생물 종의 행동 변이에 대한 연구는 미비한 상태이며 이를 구분하기 위한 기계학습 연구도 저자들이 조사한 바로는 찾아보기 힘들었다.

본 연구는 [9]의 연구를 확장해 포름알데이드 이외에 대표적인 수질 환경 물질인 벤젠(Benzene), 톨루엔(Toluene)을 포함하고 예쁜꼬마선충의 수영 행동을 이용해 이들 상호 간의 구분이 가능한지를 목표로 하였다. 각 환경 물질에 의해 오염된 물에서 동영상 촬영하고 다양한 영상 처리 기술을 활용한 다음 가지길이 유사성(Branch Length Similarity: BLS) 엔트로피를 이용해 특징을 추출한 후 HMM을 이용해 학습하였다.

본 논문의 구성은 다음과 같다. 2장에서 특징 추출 방법을 설명하고 3장에서는 분류기로 사용한 HMM을 설계한다. 4장에서 실험 및 성능 분석을 제시하고 5장에서 결론을 맺는다.

## II. 특징 추출

본 연구에서는 총 21개체의 예쁜꼬마선충(*C. elegans*)이 사용되었는데, 20°C 인큐베이터에서 대장균을 먹이로 채운 페트리 접시에서 배양하였다. 너비 4.0mm, 깊이 2.0mm인 원형 아레나(circular arena)에 증류수를 채우고 0.1ppm 농도의 포름알데히드, 벤젠, 톨루엔을 별도로 각각 7개의 아레나에 투여한 후 총 21개의 아레나에 1개체씩을 넣고 CCD 카메라를 이용해 1시간씩 촬영하였다(그림 1). 그리고 동영상은 1초당 4프레임씩 촬영

하여 개체당 총 14,400 프레임의 영상을 확보하였다.

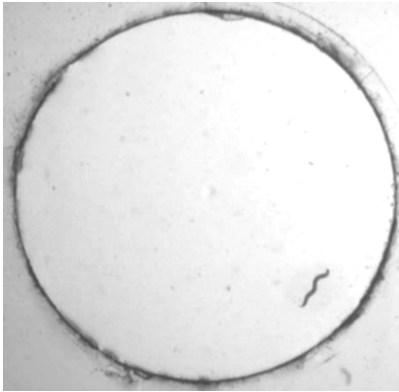


Fig. 1 An image of *C. elegans* swimming in an arena

선충의 수영 행동을 대표하기 위해 [9]와 동일하게 BLS 엔트로피를 사용하였다. BLS 엔트로피는 단위 가지 네트워크(Unit Branch Network: UBN)를 이용해 계산되는데, UBN은 주어진 그래프에서 노드들 중 한 점을 루트 노드로 하고 다른 노드들과 직접 에지(가지)로 연결된 스타 트리 형태의 네트워크를 말한다(그림 2).

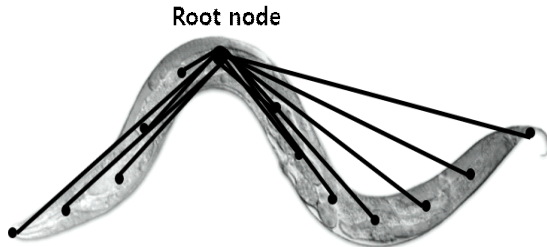


Fig. 2 A UBN composed of the 13 points placed at equal intervals along the length of *C. elegans*

선충의 길이 선을 따라 13개의 등 간격 점을 선택하고 특정 노드를 루트 노드로 삼아 UBN을 구성한 후 BLS 엔트로피를 아래 식과 같이 계산한다[9].

$$s = -\frac{1}{\log(n)} \sum_{i=1}^n p_i \log(p_i) \quad (1)$$

여기서,  $p_i = \frac{l_i}{\sum_{k=1}^n l_k}$  이며,  $l_i$ 는  $i$ -번째 가지의 길이를 나타내며,  $n$ 은 가지의 개수이다.

13개 점 각각을 루트 노드로 한 UBN 13개를 대상으

로 구해진 각각의 BLS 엔트로피의 단순 결합을 BLS 프로파일이라 하고 이를 특정 영상 프레임에 있는 선충의 수영 행동에 대한 특징 벡터로 활용하였다. 즉, 영상 프레임을 대표하는 입력 벡터는 13개의 BLS 엔트로피로 구성된 13차 벡터가 된다. 그림 3에 8개의 샘플 데이터를 예시하였다.

0.946595	0.929395	0.928039	0.935754	0.952005	0.960711	0.963673	0.960809	0.951073	0.940532	0.935309	0.938904	0.95282
0.94732	0.929703	0.928211	0.935902	0.951659	0.960878	0.96509	0.962449	0.951595	0.940621	0.935018	0.938573	0.952609
0.950054	0.932253	0.92917	0.935804	0.950783	0.959414	0.967408	0.963402	0.952141	0.938743	0.93336	0.937042	0.953956
0.951533	0.934629	0.931861	0.938885	0.953376	0.962343	0.967435	0.963181	0.952332	0.940484	0.935218	0.939385	0.955989
0.950098	0.933605	0.930971	0.938166	0.953235	0.961941	0.967957	0.964588	0.954313	0.943238	0.937527	0.941334	0.95723
0.951749	0.93511	0.931917	0.940712	0.952654	0.963602	0.967385	0.963316	0.952568	0.941844	0.936524	0.940787	0.957002
0.952522	0.9347	0.93015	0.93771	0.950071	0.963076	0.96782	0.96593	0.955324	0.943386	0.937826	0.940757	0.956019
0.951935	0.934831	0.930979	0.939207	0.951747	0.965291	0.968356	0.965624	0.956108	0.943844	0.938132	0.941638	0.957035

Fig. 3 Samples of 13 dimension BLS entropy vectors obtained from a experiment using formaldehyde

BLS 프로파일로 대표되는 특징 벡터는 벡터의 개별 요소값들이 실수로써 무한대의 벡터 공간의 특성을 가짐을 알 수 있다. 무한대의 패턴을 패턴 분류에 직접 사용하는 것은 불가능하다. 따라서 클러스터링 기술을 활용해 특징 벡터를 군집화하여 몇 개의 패턴으로 나눠 최종 특징으로 사용한다.

사용한 클러스터링 알고리즘은 인공신경망의 하나인 자기조직화맵(SOM: Self Organizing Map)을 활용하였다. SOM은 특징 벡터로 표현된 데이터의 분포 공간상의 위상을 보존하는 특성을 가지고 있어서 무한개의 특징 벡터를 유한개의 패턴 군집으로 분류하는데 적합한 성질을 가진 비지도학습 유형에 속하는 기계학습 방법이다. 본 연구에서는 [9]와 같이 14×14 크기의 평면 맵(planar map)을 SOM의 기본 맵으로 사용하였다. 그러나 14×14 크기의 평면 맵은 총 196가지의 패턴을 제공하므로 이 역시 시간 데이터의 특징으로 활용하기에는 패턴이 지나치게 많다. 따라서 이 196가지 패턴을 k-means 클러스터링 알고리즘을 이용해 적절한 개수의 패턴으로 다시 군집하고 각각의 패턴에 최종 특징 라벨인 숫자 기호를 부여하였다. 본 논문에서는 5, 6, 7개의 패턴을 사용해 실험하였다. 결국 패턴 수가 7인 경우 연속된 6개의 프레임에 나타난 꼬마선충의 수영 행동은 1~7까지의 수로 구성된 [1, 1, 2, 4, 3, 3] 같은 6차원 벡터가 된다.

[9]에 따르면 클러스터링을 활용해 얻은 패턴은 선충의 모양(shape)에 따라 선 형태부터 원형까지 선충의 구부러진 정도에 따라 7개 정도로 구분된다고 한다.

SOM과 k-means 클러스터링 기술을 혼합해 사용하는 이유는 SOM을 이용해 데이터 분포를 보다 벡터 공간의 위상을 유지하며 구분 짓는 동시에 k-means 클러스터링을 통해 이들을 원하는 개수의 특징으로 군집화하여 사용할 수 있기 때문이다.

구현에는 파이썬으로 된 Somoclu 모듈[10]과 scikit-learn의 KMeans 모듈[11]을 사용하였다. 또한 맵 유형 이외에 SOM의 훈련에 사용된 radius, scale 등의 파라미터는 Somoclu의 기본값을 그대로 사용하였다.

### III. HMM 기반의 모니터링 시스템

선충의 수영 행동 동영상에 대상으로 오염 물질 분류 문제를 해결하기 위해 시계열 데이터 사용에 적합한 대표적인 기계학습 모델인 히든 마코프 모델(HMM: Hidden Markov Model)을 사용하였다[12]. HMM은 일반적으로 5-튜플  $\langle N, M, A, B, \pi \rangle$ 로 특징 지어지며 그 의미는 표 1과 같이 정의된다.

Table. 1 The meaning of elements in 5-tuple

	meaning
N	the number of hidden states
M	the number of observation symbols in a state
A	the probability distribution of the state transition
B	the probability distribution of the observation symbol
$\pi$	the probability distribution of the initial state

모델 파라미터 중 히든 상태 수(N)는 각각 5, 6, 7개로 개수를 달리하여 실험하였고, SOM과 k-means 클러스터링의 결과인 패턴 수에 해당하는 관찰 심볼 수(M)도 히든 상태와 동일한 개수를 사용해 실험하여 최적 파라미터를 선택하였다.

각 영상 프레임 속 선충의 수영 행동을 반영하는 패턴들은 관찰 시간에 해당하는 프레임 수 차원의 입력 벡터로 조합하여 학습 데이터로 사용하였다.

3가지 오염 물질에서 얻은 학습 데이터 각각을 이용해 그에 상응하는 3가지 HMM을 대상으로 각 50회 반복하여 학습하였다. 테스트 데이터를 대상으로 한 오염 물질 판별은 학습된 3가지 모델에 각각 테스트 샘플을 입력 데이터로 이용해 각각의 확률을 구한 다음 가장 높

은 확률을 보이는 HMM의 라벨을 예측치로 사용하였다.

HMM 구현에는 파이썬으로 구현된 hmmlearn 모듈을 사용하였다[13]. 표 1에 정의된 파라미터 이외의 학습에 필요한 파라미터는 hmmlearn 모듈의 기본값들을 사용하였다.

### IV. 실험 및 성능 분석

앞에서 언급한 대로 0.1ppm 농도의 포름알데히드, 벤젠, 톨루엔이 투여된 아레나에서 각각 7개체에 대해 1초당 4 프레임의 60분 동영상을 촬영하였다. 초기 적응 과정과 후반부 실험 개체의 사망 등으로 인한 움직임 없는 영상 부분을 제외하기 위해 초기 10분 뒷부분 20분을 제외한 30분 영상만 사용하였다. 따라서 개체 당 실제 영상 프레임 수는 7,200 프레임이고, 각 화학 물질 당 7개 체이므로 전체 실험 데이터는 151,200 프레임으로 구성되었다.

생물 종을 활용한 모니터링에서 수질 오염 여부를 판단하는 가장 중요한 요소 중 하나는 오염 여부 판별에 소모되는 관찰 시간이다[9]. 마찬가지로 환경 물질의 종류를 구별하는 모니터링 문제에서도 관찰 시간은 가장 중요한 요소 중 하나이다. 따라서 학습과 테스트에 사용한 관찰 시간을 30초, 60초, 90초, 120초, 150초, 180초로 다양하게 설정하고 실험하였다.

관찰 시간은 프로파일의 길이에 비례하는 것으로, 30초 관찰 시간을 사용하는 경우  $30 \times 4 = 120$  프레임에 해당하는 BLS 프로파일(100차원의 입력 벡터)을 사용하게 된다. 실제 전체 입력 벡터 집합은 30분 데이터로부터 관찰 구간 사이에 오버랩 구간 20%를 두고 잘라내 구성한다. 한편 전체 데이터의 80%를 훈련 데이터로 사용하고 나머지 20%는 테스트 데이터로 사용하였다. 각 환경 물질의 관찰 시간에 따른 입력 데이터 수는 아래 표 2와 같다.

Table. 2 The composition of experiment data set

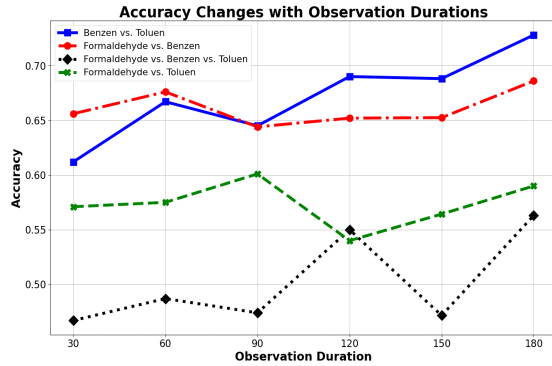
observation time (seconds)	30	60	90	120	150	180
Input vector length	120	240	360	480	600	720
the number of training data	419	209	139	104	83	69
the number of test data	105	53	35	27	21	18

**Table. 3** The best accuracy with the number of patterns and observation time for classification problems

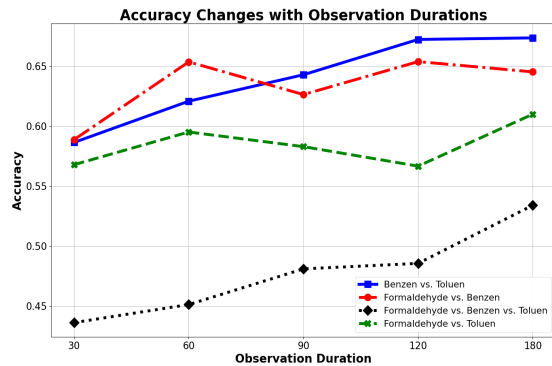
	Formaldehyde vs. Benzen	Formaldehyde vs. Toluene	Benzen vs. Toluene	Formaldehyde vs. Benzen vs. Toluene
accuracy	0.69	0.66	0.75	0.56
the number of patterns	7	5	7	7
observation time (sec)	180	180	180	180

환경 물질 분류 문제는 3가지 환경 물질을 쌍으로 비교한 이진 분류 문제 3가지(Formaldehyde vs. Benzen, Formaldehyde vs. Toluene, Benzen vs. Toluene)와 3가지 모든 환경 물질을 분류하는 다중 분류 문제로 구성하였다. 이들 각각의 분류 문제를 패턴 수, HMM의 히든 상태 수, 관찰 시간을 변경해 가며 실험하고 최적인 상태에서 20회 실시한 후 평균 낸 정확도(accuracy)를 표 3에 나타내었다. 다만, 실험 패턴 수와 상태 수는 5, 6, 7개로 동일한 개수를 사용하였다. 즉 패턴 수가 5인 경우 상태 수도 5로 하여 실험하였다.

표 3에서 알 수 있듯이, 포름알데이드와 톨루엔 실험 이외의 모든 실험 결과에서 [9]와 동일하게 7개의 패턴에서 가장 높은 정확도를 보여 주었다. 그리고 관찰 시간은 모두 180초에서 가장 높은 정확도를 보여주었다. 이와 관련해 네 가지 분류 문제에 관찰 시간이 정확도에 미치는 영향을 6가지 시간을 이용해 실험하고 그 결과를 그림 4에 나타내었다. 이때 패턴과 히든 상태는 7로 고정하여 실험하였다. 포름알데이드와 톨루엔 실험을 제외하고는 모두 180초에서 가장 높은 정확도를 보여주었으며 뚜렷하진 않지만 대체로 관찰 시간에 비례해 정확도가 높아짐을 알 수 있다. 패턴과 히든 상태를 5로 한 실험 결과인 그림 5에서도 확인할 수 있듯이 포름알데이드와 톨루엔 실험의 경우도 최고 정확도를 보여 주었던 패턴 수 5, 상태 수 5에서는 180초에서 가장 높은 정확도를 보여 주었다. 따라서 충분한 관찰 시간을 활용해야 높은 정확도의 판별이 가능함을 알 수 있다.



**Fig. 4** The accuracy of the monitoring system over the test data with observation time varying from 30 to 180 seconds under state number 7 and pattern number 7



**Fig. 5** The accuracy of the monitoring system over the test data with observation time varying from 30 to 180 seconds under state number 5 and pattern number 5

## V. 결론

본 논문에서는 예쁜꼬마선충의 수영 행동을 이용해 3가지 환경 오염 물질을 구분할 수 있는 HMM 기반의 바이오 모니터링 시스템을 구현하였다. 시스템은 먼저 3가지 각각 다른 오염 물질 학습 데이터를 이용해 학습한 3가지 HMM을 생성한 후 테스트 샘플을 대상으로 가장 높은 확률을 보이는 HMM의 예측을 최종 예측으로 결정한다.

2가지 환경 오염 물질을 분류하는 문제에 대해서는 70% 정도의 정확도를 보여주었으나 3가지 모든 환경 오염 물질을 분류하는 데는 56% 정도의 정확도를 보여주었다. 이러한 정확도는 현장에서 바로 사용 가능한 수

준의 바이오 모니터링 시스템으로는 부족한 정확도라고 생각한다.

다만, 본 논문이 실험을 통해 확인해 보고자 한 것은 이전 연구와 달리 오염 물질의 존재 여부가 아니라 오염 물질 간의 차이를 예쁜꼬마선충의 수영 행동과 기계학습을 이용해 분류 가능한지이다. 비록 높은 수준의 정확성을 보여주지는 못했지만 무작위 결정(2가지 오염 물질의 경우 50%, 3가지 오염 물질의 경우 33%)에 비해서는 의미 있는 수준의 정확성을 보여주고 있어, 앞으로 성능 개선을 위한 추가 연구가 이루어지면 현장에서 사용 가능한 수준의 성능도 확보할 수 있을 것으로 판단된다.

#### ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea under Grant NRF-2020R111A3071599.

#### References

- [ 1 ] Y. S. Park, N. I. Chung, K. H. Choi, E. Y. Cha, and S. K. Lee, "Computational characterization of behavioral response of medaka (*Oryzias latipes*) treated with diazinon," *Aquatic Toxicology*, vol. 71, no. 3, pp. 215-228, 2005.
- [ 2 ] L. O. Teles, M. Fernandes, J. Amorim, and V. Vasconcelos, "Video-tracking of zebrafish (*Danio rerio*) as a biological early warning system using two distinct artificial neural networks: Probabilistic neural network (PNN) and self-organizing map (SOM)," *Aquatic Toxicology*, vol. 165, pp. 241-248, 2015.
- [ 3 ] M. Fernandes, J. Amorim, V. Vasconcelos, and L. O. Teles, "Resilience assessment of a biological early warning system based on the locomotor behavior of zebrafish (*Danio rerio*)," *Environmental Science and Pollution Research*, vol. 23, pp. 18858-18868, Jun, 2016.
- [ 4 ] S. Villa, D. N. Valeria, P. Tanita, B. Francesco, M. Francesco, F. Antonio, and L. Valeria, "Comparison of the behavioural effects of pharmaceuticals and pesticides on *Diamesa zernyi* larvae(Chironomidae)," *Environmental Pollution*, vol. 238, pp. 130-139, 2018.
- [ 5 ] Q. A. V. Le, S. S. Sekhon, L. Lee, and J. H. Ko, "Daphnia in water quality biomonitoring - "omic" approaches," *Toxicology and Environmental Health Sciences*, vol. 8, no. 1, pp. 1-6, 2016.
- [ 6 ] C. W. Ji, Y. S. Park, Y. Cui, H. Wang, I. S. Kwak, and T. S. Chon, "Analyzing the Response Behavior of *Lumbriculus variegatus* (Oligochaeta: Lumbriculidae) to Different Concentrations of Copper Sulfate Based on Line Body Shape Detection and a Recurrent Self-Organizing Map," *International Journal of Environmental Research and Public Health*, vol. 17, no. 8, pp. 2627, 2020.
- [ 7 ] P. Wittkowski, P. Marx-Stoelting, N. Violet, V. Fetz, F. Schwarz, M. Oelgeschläger, G. Schönfelder, and S. Vogl, "Caenorhabditis elegans As a Promising Alternative Model for Environmental Chemical Mixture Effect Assessment-A Comparative Study," *Environmental Science & Technology*, vol. 53, no. 21, pp. 12725-12733, 2019.
- [ 8 ] M. E. Hurtado-Díaz, R. Estrada-Valencia, E. Rangel-López, M. Maya-López, A. Colonnello, S. Galván-Arzate, S. V. Verstraeten, C. Karasu, I. Túnez, M. Aschner, and A. Santamaría, "Thallium Toxicity in *Caenorhabditis elegans*: Involvement of the SKN-1 Pathway and Protection by S-Allylcysteine," *Neurotoxicity Research*, vol. 38, no. 2, pp. 287-298, May, 2020.
- [ 9 ] I. S. Jeong, S. R. Lee, I. Song, and S. H. Kang, "A biological monitoring method based on the response behavior of caenorhabditis elegans to chemicals in water," *Journal of Environmental Informatics*, vol. 33, no. 1, pp. 47-55, 2019.
- [ 10 ] Somoclu module implemented in python for Self-organizing Map [Internet]. Available: <https://somoclu.readthedocs.io/en/stable/#>.
- [ 11 ] scikit learn module for KMeans clustering algorithm [Internet]. Available: <https://scikit-learn.org/stable/>.
- [ 12 ] S. Pattnaik, A. K. Nayak, and S. Pattnaik, "A Semi-supervised Learning of HMM to Build a POS Tagger for a Low Resourced Language," *Journal of Information and Communication Convergence Engineering*, vol. 18, no. 4, pp. 207-215, 2020.
- [ 13 ] Hmmlern module implemented in python for Hidden Markov model [Internet]. Available: <https://hmmlern.readthedocs.io/en/latest/>.



**강승호(Seung-Ho Kang)**

1994년 8월 전남대학교 전산학과 이학박사  
2003년 8월 전남대학교 전산학과 이학박사  
2009년 8월 전남대학교 전산학과 이학박사  
2013년 9월 ~ 현재 동신대학교 정보보안학과 교수  
※관심분야: 알고리즘, 기계학습, 정보보안



**정인선(In-Seon Jeong)**

2001년 2월 여수대학교 전자계산학과 학사  
2006년 2월 전남대학교 전산학과 석사  
2011년 2월 전남대학교 전산학과 박사  
2017년 8월 ~ 현재 전남대학교 박사후연구원  
※관심분야: 알고리즘, 기계학습



**임형석(Hyeong-Seok Lim)**

1983년 2월 서울대학교 컴퓨터공학과 학사  
1985년 2월 한국과학기술원 전산학과 석사  
1993년 8월 한국과학기술원 전산학과 박사  
1987년 9월 ~ 현재 전남대학교 소프트웨어공학과 교수  
※관심분야: 알고리즘, 그래프 이론, 바이오인포매틱스