

의사결정나무를 이용한 생물의 행동 패턴 구분과 인식

Classification and Recognition of Movement Behavior of Animal based on Decision Tree

이승태, 김성신
부산대학교 전기공학과

Sengtai Lee, Sungshin Kim
School of Electrical and Electronic Engineering, Pusan National University
E-mail : youandi@pusan.ac.kr

ABSTRACT

본 논문에서는 생물의 2차원영상에서 4가지의 특징을 추출한 다음 약품에 대한 생물의 행동 패턴 반응에 대하여 의사결정나무를 적용하여 패턴의 인식 및 분류를 하였다. 생물의 행동패턴을 대변하는 물리적인 특징인 속도, 방향전환 각도, 이동거리에 대하여 각각 중간이상속도비율, FFT(Fast Fourier Transformation), 2차원 히스토그램 면적, 프렉탈, 무게중심을 사용하여 특징을 추출하였다. 이렇게 추출된 4가지의 특징변수들을 사용하여 의사결정나무 모델을 구성한 다음 생물의 약품 첨가에 대한 반응을 분석하였다. 또한 결과에서는 기존의 생물의 행동패턴 구분에 쓰였던 전형적인 기법(conventional methods)보다 본 연구에서 적용한 의사결정나무가 생물의 행동패턴이 가지는 물리적 요소에 대한 독해력을 가짐을 보임으로써 특정환경에서 이동행동에 대한 분석을 용이하게 하고자 하였다.

Key words : 의사결정나무, 행동 패턴, 패턴 인식, 특징 추출, 2차원 영상

1. 서 론

환경 데이터는 매우 복잡하고, 불규칙적이며, 많은 오차를 가진다. 이러한 환경 데이터는 상당히 비선형적이며, 고차원적인 관계식을 가진다. 일반적으로 사용된 탐색적이거나 통계적인 모델링 기법으로는 이러한 환경 데이터에서부터 의미론적인 특정 패턴을 발견하기란 어렵다 [1, 2, 3]. 수질 모니터링과 생물의 행동 특성 파악은 생물 관리와 위험 관리를 다루는데 상당한 중요성을 가진다 [4, 5]. 이러한 행동패턴 분석을 위하여 바람, 페로몬, 상대습도, 자

당공급에 따른 바퀴벌레 같은 곤충의 연속적인 운동 궤적 추적을 조사하기 위한 방법으로 서보미터 같은 방법들이 있다. 본 연구에서는 행동패턴을 구별하기 위하여 약품을 첨가하지 않은 정상상태에서의 송사리의 행동 궤적과 약품을 첨가한 비정상상태에서의 송사리의 행동 궤적을 영상 획득 시스템으로 얻어 일정한 패턴으로 분류한 다음 의사결정 나무를 이용하여 생물의 행동 패턴 변화를 분석하고자 하였다. 실시간 패턴분류를 통하여 일정 시간 경과 후에 약품의 첨가 여부를 알아봄으로써 생물 행동 패턴 구분으로 환경 데이터를 분석하였다.

II. 본 론

2.1 생물 행동 데이터

본 실험에서 사용된 송사리는 일본의 나고야 대학의 메다카센터에서 분양되어 한국화학 부설 안정성평가연구소(대전광역시 소재)에서 계대번식 시킨 것을 분양 받아서 실험실 조건(수온 25℃)에서 충분한 양의 먹이와 산소를 공급하여 1개월 이상 기른 것을 실험에 사용하였다. 일반적인 송사리는 체장이 짧고 유전적으로 균일성 있는 다수의 개체를 유지하고 있다. 또한 생태계에 널리 분포하며, 환경오염에 의해 가장 먼저 영향을 받는 지표종으로 인정되어 있다. 또한 온도, DO, pH등의 환경변화에 대한 내성이 매우 강하다. 본 논문에서는 이러한 생물의 행동을 관찰함으로써 특정 패턴을 구분 및 인식하여 행동에 미치는 원인을 찾고자 하였다.

관찰 수조 정면에 설치된 CCD 카메라를 통해 들어온 영상은 매 0.25초 간격으로 포착되어 컴퓨터의 비디오 보드에서 디지털 신호로 변환되었다. 그 후, 물고기 위치 추적 프로그램에 의해 관찰 영역 내에서 송사리의 위치는 x, y의 2차원 좌표 값으로 기록되었다. 좌표 값으로 기록된 송사리의 시, 공간적 위치 데이터는 물고기 위치 재생 프로그램을 이용하여 약품 처리와 약품 비처리 상태에서의 행동 변화를 연속적으로 재생 관찰하였다.

2.2 생물 행동 패턴 구분

영상 획득 시스템으로부터 얻어진 생물 궤적 데이터는 너무나 많은 다양성과 복잡성을 포함하고 있다. 이렇게 복잡한 행동들을 구별하기 위하여 본 논문에서는 우선 4가지 패턴들을 선택하였다. 생물의 행동 관찰에서 크게 구별할지을 수 있는 4가지 패턴은 Active shaking, Active smooth, Inactive shaking, Inactive smooth가 있다. 그림 2는 영상 획득 시스템으로부터 얻어진 생물의 궤적을 1분 동안 표시한 것이다.

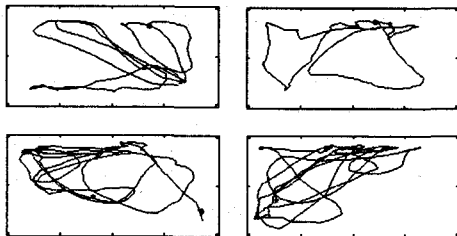


그림 1. 1분 동안의 생물의 궤적

이렇게 다양한 생물의 행동 궤적을 분류하기 위하여 행동 궤적에 대한 특징을 추출하기 위

하여 계량적 특성화를 먼저 하였다.

2.2.1 속도 비율

속도를 비율로 계산하였다. 특정 패턴이 가지는 물리적인 속도에 관한 정보가 드러날 수 있도록 하였다. 그리하여, 특정 패턴에서 속도 비율이 높다는 것이 나타난다면 그 특정 패턴이 많이 나타나는 개체에서는 큰 이동속도를 가진다는 것을 유추할 수 있도록 하였다. 속도에 대한 식은 다음과 같다.

$$S = \sqrt{(x_{n+1} - x_n)^2 + (y_{n+1} - y_n)^2} \quad n=1,2,3,\dots$$

여기서 x_n 과 y_n 은 각각 x축 좌표와 y축 좌표를 나타낸다. 속도 비율식은 다음과 같다.

$$S_{ratio} = \frac{\text{Number of samples above A2}}{\text{Number of samples in one minute}} \times 100(\%)$$

여기서 A2는 평균속도를 나타낸다. 위의 식은 평균속도를 넘어서는 속도비율을 나타낸다.

2.2.2 각도 변화에 대한 FFT

송사리의 행동 궤적 추적에서 각도의 변화도 하나의 변수로 사용하였다. H로 표현된 샘플 시간 사이의 각도 변화 계산식은 다음과 같다.

$$H = \arctan\left(\frac{y_{n+1} - y_n}{x_{n+1} - x_n}\right), \quad n=1,2,\dots$$

푸리에 변환은 시간영역의 파형 변화를 주파수 영역에서 분석할 수 있게 해주는 유용한 변환법이다 [91]. 각도 변화에 대한 에너지 값을 계산하기 위하여 고속 푸리에 변환을 사용하였다. 주어진 이산 신호 $x[n]$ 에 대한 계산식은 다음과 같다.

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j(2\pi kn/N)}, \quad k=0,1,\dots,N-1.$$

각도 변화에 대한 고속 푸리에 변환을 행한 후에, 다음의 식을 사용하여 각도 변화 중간값(A3)를 넘어서는 샘플 수에 대한 비율을 계산하였다.

$$P = \sqrt{\sum_{i=1}^k x_i^2}$$

여기서 x_i 는 각도 변화 중간값(A3)를 넘어서는 진폭을 나타낸다. 본 논문에서는 사용한 A3는 20이다. 송사리의 행동 분류에 대한 2번째 특징 변수로써 이러한 고속푸리에 변환에 대한 에너지 값(P)를 사용하였다.

2.2.3 x축과 y축에 대한 히스토그램

히스토그램은 2차원 그래프에 있는 도형의 모양을 1차원 그래프로 나타내는 방법이다. 본 논문에서는 2차원 공간에 있는 송사리의 이동 경로를 파악하기 위하여 히스토그램을 이용하였다. x축에 대한 히스토그램과 y축에 대한 히스토그램을 각각 계산한 후에 곱하여 송사리가

이동한 경로에 대한 면적을 구하였다. 계산된 면적은 송사리가 넓게 움직이는지 또는 좁은 영역에서 움직이는 지에 관한 정보를 알려준다.

2.2.4 프랙탈 차원

프랙탈 차원은 구면이나 정육면체나 그 밖의 입체들은 3차원이고, 정사각형이나 삼각형은 2차원이며, 직선과 곡선은 1차원이고, 점들은 0차원이다. 그러나 프랙탈은 얼마나 많이 구부러져 있는가에 따라 1차원과 2차원 사이의 어느 차원이나 될 수 있다. 프랙탈 차원을 하나의 특징 변수로 사용함으로써 생물이 지나간 궤적에 대한 분석을 하고자 하였다.

2.2.5 무게중심

생물의 행동 특성을 나타내는 또 다른 특징 변수로써 무게중심을 사용하였다. 송사리라는 개체의 특성상 물 속에 있는 송사리는 호흡과 먹이 공급을 위하여 자주 수면으로 올라오는 행동을 보인다. 이에 본 논문에서는 송사리의 수면에 있는 자취를 기록하고 분류하기 위하여 무게중심 값을 사용하게 되었다. 무게중심 좌표에 대한 식은 다음과 같다.

$$C = \sum_j \sum_j (x_j, y_j)$$

그림 3은 Active shaking 패턴에 대한 특징 추출 한 결과를 나타낸 것이다.

2.3 의사결정 나무에의 적용

의사결정 나무는 귀납적 학습방법에 속하는 알고리즘으로 수집된 모델 추정용(training) 데이터로부터 순환적 분할(recursive partitioning) 방식을 이용하여 나무를 구축하는 기법으로, 구축되어진 의사결정 나무는 속성의 분리기준을 포함하는 내부마디(Internal nodes)와 최종분류를 의미하는 잎(leaves)으로 구성된다[6, 7].

분석의 목적과 데이터 구조에 따라 적절한 분리기준(split criterion)과 정지규칙(stopping rule)을 지정하여 나무모델을 만든다. 목표변수가 이산형인 경우에는 목표변수의 각 범주에 속하는 빈도(frequency)에 기초하여 분리가 일어나며, 이를 분류나무(classification)를 구성한다고 한다. 분류를 위한 나무구조에서 대표적인 분리기준으로는 Chi-squared test, 지니 계수(Gini Index), 엔트로피(Entropy)를 들 수 있다. 본 논문에서는 CART(classification and regression tree)라는 이진분리(binary split)를 수행하는 알고리즘을 사용하여 엔트로피 함수를 사용하여 불순도(impurity)를 측정하였다. 다음은 엔

트로피 함수를 나타낸다.

$$i(N) = -\sum_j P(\omega_j) \log_2 P(\omega_j)$$

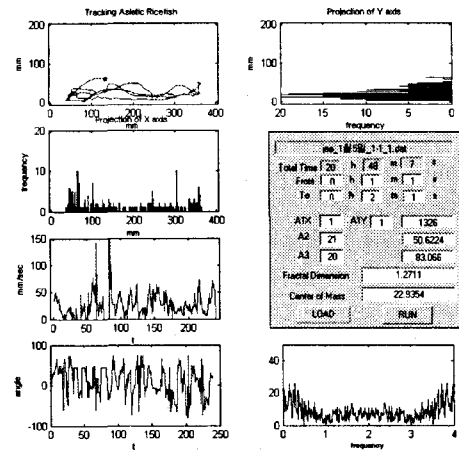


그림 2. Active shaking 패턴 분석의 예

$i(N)$ 은 노드에서의 불순도를 나타내고, $P(w_j)$ 는 카테고리 w_j 에서 노드N에 속하는 것을 비율로 나타낸 것이다. 변수선택 후에 분기의 정도를 구하기 위하여 불순도를 낮추기 위하여 다음의 공식을 사용하여 분기점을 구한다.

$$\Delta i(N) = i(N) - P_L i(N_L) - (1 - P_L) i(N_R)$$

생물의 행동에 대한 궤적으로부터 추출한 이상 5가지의 특징 변수들을 가지고 의사결정나무 모델을 구성하였다. 의사결정나무 결과를 IF-THEN 규칙으로 나타내고 트리 형태의 분류로 표현하기 위하여 Matlab 6.1을 이용하여 프로그램 하였다. Active shaking, Active smooth, Inactive shaking, Inactive smooth의 4가지 대표적인 패턴을 생물 전문가의 도움을 받아 구분을 하였고, 이 4가지 패턴에 들지 않는 Not determined라는 패턴을 더하여 구성하였다.

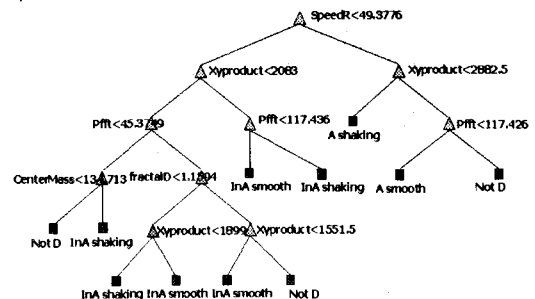


그림 3. 의사결정나무 모델 구성 결과

각각의 패턴을 구별하기 위한 특징으로 속도 비율, FFT, x축과 y축에 대한 히스토그램, 프렉탈 차원, 무게중심을 사용하였다. 의사결정나무 모델 구성을 위한 트레이닝 데이터로는 각각의 패턴 별로 30개씩을 추출하였다. 그림 3은 의사결정나무 모델 구성 결과를 보여준다.

III. 결 과

의사결정나무 모델을 구성하여 나온 로직을 사용하여 실제 약품 첨가 데이터와 약품을 첨가하지 않은 데이터에 적용하여 보았다. 그림 4는 약품을 첨가하지 않은 송사리의 궤적에 의사결정나무 로직을 적용하여 Inactive shaking 패턴과 Inactive shaking 패턴을 제외한 나머지 패턴들이 차지하는 비율을 비교한 것이다. 약품을 첨가하지 않은 총 18개의 데이터에서 Inactive shaking 패턴의 비율을 평균 13.5%를 나타냈다. 송사리의 총 이동궤적 중 처음부터 1시간 30분동안의 행동 변화를 관찰하였다.

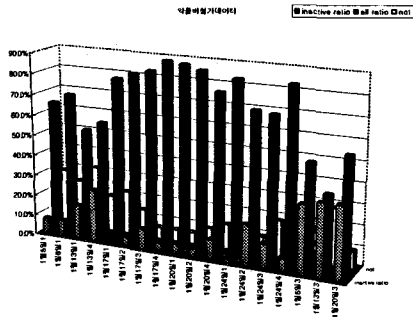


그림 4. 약품 비첨가 데이터에 대한 의사결정나무 분석 결과

그림 5는 약품을 첨가한 송사리의 궤적에 의사결정나무 로직을 적용한 결과이다. 약품을 첨가한 총 18개의 데이터에서 Inactive shaking 패턴의 비율은 평균 45%를 나타냈다.

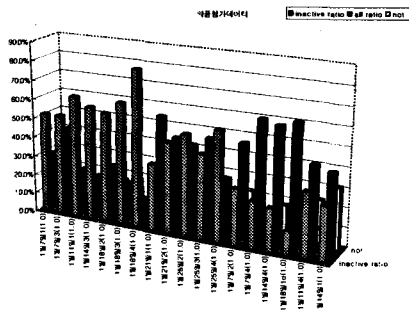


그림 5. 약품 첨가 데이터에 대한 의사결정나무 분석 결과

IV. 결 론

본 논문에서는 충분한 관찰을 바탕으로 송사리의 행동을 구별 지을 수 있는 방법으로 속도, 각도변화, 이동면적, 프렉탈 차원, 무게중심으로 변수를 구성하여 패턴을 분류하고자 하였다. 이렇게 추출한 특징들을 가지고 의사결정나무를 이용하여 약품 처리한 송사리와 약품 처리하지 않은 송사리에서의 행동 패턴 변화를 알아보았다. 본 연구의 내용을 보다 많은 실험 데이터에서 적용하여 본다면 더 정확한 행동 구분 패턴을 얻을 수 있을 것이라 생각된다. 보다 나아가서는 실시간으로 생물의 패턴을 구별함으로써 생물의 특성연구에 도움이 되리라 생각된다.

감사의 글 : 본 연구는 한국과학재단 목적기초 연구(R01-2003-000-10714-0)지원으로 수행 되었음.

V. 참고문헌

- [1] Breiman, L., J. H. Friedman, R. A. Olshen, and C. G. Stone. "Classification and Regression Trees", Wadsworth International Group, 1984
- [2] Ripley, B. D. "Pattern recognition and neural networks". Cambridge University Press. 1996.
- [3] Richard, O. D., Peter, E. H., David, G. S. "Pattern Classification 2nd edn.". Wiley Interscience. 1996.
- [4] Dutta, H., Marcelino, J., Richmonds, Ch. "Brain acetylcholinesterase activity and optomotor behavior in bluefills, *Lepomis macrochirus* exposed to different concentrations of diazinon". Arch. Intern. Physiol. Biochim. Biophys. Volumn:100. Issue:5. 1993. page(s):331-334
- [5] Lemly, A. D., Smith, R. J. "A behavioral assay for assessing effects of pollutants of fish chemoreception". Ecotoxicology and Enviornmental Safety. Volumn: 11 Issue: 2. 1986. Page(s):210-218
- [6] Quinlan, J. R. "Discovering rules by induction from large collections of examples". Edinburgh University Press. 1979
- [7] Quinlan, J. R. "Induction of decision trees" Machine Learning, volumn:1 Issue:1. 1986. page(s):81-106