DASVDD 모형을 통한 반려동물 센서 데이터 이상치 탐지

박정현[1], 고준혁[1], 김시웅[1] 문남미[2]
[1]호서대학교 컴퓨터공학과 석사과정
[2]호서대학교 컴퓨터공학부 교수

jh.park970609@gmail.com, junhyeok970306@gmail.com
kimsiung990811@gmail.com, nammee.moon@gmail.com

Detection of outliers in pet sensor data through DASVDD

JeongHyeon Park[1], JunHyeok Go[1], SiUng Kim[1], Nammee Moon[1]
[1]Dept. of Computer Science and Engineering, Hoseo University

요 약

이상치는 주로 저빈도로 발생하기 때문에, 이상치 탐지 분야에서는 정상 데이터만을 이용한 비지도 기반 학습 모델을 사용하는 방법들이 제안되었다. 따라서, 본 논문에서는 반려동물 센서 데이터를 이용해 비지도 기반 모델인 DASVDD을 활용하여 이상치를 탐지한다. 하지만 데이터셋에 이상치가 존재하지 않아 반려동물이 고빈도로 보여주는 A행동군(서다, 앉다, 엎드리다, 눕다, 걷다), 저빈도로 보여주는 B행동군(킁킁대다, 먹다)으로 분리하여 학습을 진행한다. 모델의 성능은 ROC-AUC을 기준으로 79.05%의 성능을 보여주는 것을 확인하였다.

1. 서론

이상치는 특정 데이터 집합에서 보통의 패턴에서 벗어나는 관측치로 정의되며, 이러한 이상치를 탐지하는 방법을 일반적으로 이상치 탐지라고 한다[1]. 그러나 이상치는 일반적으로 주어진 데이터 집합에서 저빈도로 발생하는 특성을 가지고있어, 일반적인 상황에서의 수집이 어려운 경우가 많다. 또한, 생체 신호와 관련된 데이터 수집 시 비윤리적인 문제도 발생할 수 있다.

이러한 문제를 극복하기 위해 이전에 진행된 이상치 탐지 연구에서는 정상 데이터만을 활용하는 비지도 학습 기반의 모델들이 제안되었다[2]. 비지도 학습 모델은 주어진 데이터 집합에서 일반적인 패턴을 학습하고, 이 패턴과 일치하지 않는 관측치를 이상치로 식별한다. 이러한 방법은 저빈도로 발생하는 이상치에도 적용 가능하며, 생체 신호와 같이 민감한 데이터 수집에서의 비윤리적 문제를 최소화하는 데 도움이 된다.

본 연구에서는 반려동물 센서 데이터를 비지도 학습 모델인 DASVDD(Deep Autoencoding Support Vector Data Descriptor)로 모델을 구성한 뒤, 해당 모델을 기반으로 이상치 탐지를 진행하고자 한다.

2. 이상치 탐지

2.1. 데이터셋 소개

실험에는 9축 IMU(가속도, 각속도, 지자계)로 수집된 반려동물 센서 데이터셋를 사용한다[3]. 데이터는 총 10마리의 반려동물을 대상으로 50Hz 주기로 수집되었으며 <표 1>과 같이 7가지의 행동군(서다, 앉다, 엎드리다, 눕다, 걷다, 킁킁대다, 먹다)으로 분류되어있다.

본 연구에서는 반려동물이 고빈도로 보여주는 행동인 A행동군(서다, 앉다, 엎드리다, 눕다, 걷다)을 정상 데이터로 가정하고 저빈도로 나타나는 B행동군(킁킁대다, 먹다)을 이상치로 가정하여 학습을 진행한다.

<표 1> 반려동물 데이터셋

행동군	행동	개수
A	서다	3,323
	앉다	5,257
	엎드리다	4,377
	눕다	2,385
	걷다	1,370
B	킁킁대다	640
	먹다	1,230

2.2. 데이터셋 전처리

가. 결측치 보간

한가지의 행동 시퀀스에 결측치가 10% 이상인 경우 데이터의 신뢰성이 보장될 수 없다고 판단하여 해당 시퀀스를 제거한다. 나머지 결측치가 10% 이하인 스퀀스에 대해서는 선형 보간을 통해 결측치를 해결한다.

나. 데이터 스케일링

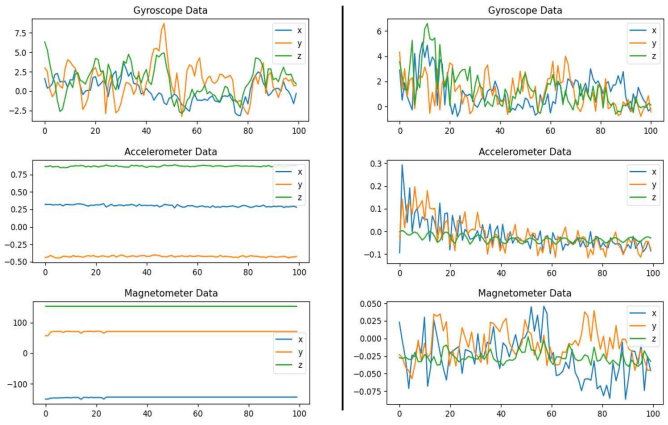
다음으로 데이터의 균일한 스케일을 위해 표준화(Standardization)를 수행한다. 표준화는 각 변수의 평균을 0으로, 표준 편차를 1로 조정하여 데이터 분포를 균일하게 만드는 작업이다. 이를 통해 변수 간의 상대적인 중요도를 균형 있게 유지하며, 학습의 정확도와 신뢰도를 향상시킨다.

다. FFT

다음은 주파수 영역으로 데이터를 변환하기 위해 FFT를 활용한다. FFT를 통해 데이터는 주파수 도메인으로 변환되어, 원 데이터의 시간 영역에서 파악하기 어려운 주기적 패턴이나 주파수 특성을 뚜렷하게 드러낸다. 이를 통해 더 정교한 학습이 가능하다.

라. 센서 데이터 시퀀스

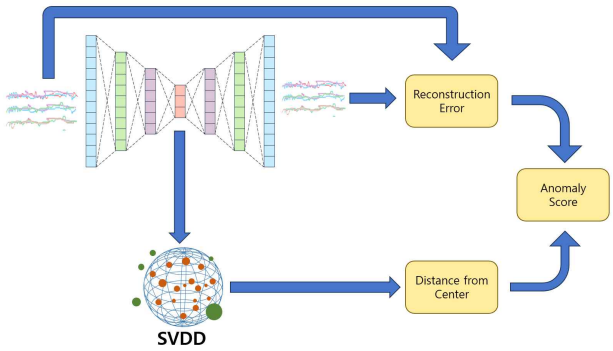
앞서 언급한 전처리 과정 이후, 센서 데이터를 100(2초)의 길이로 시퀀스를 재구성한다. 최종적으로 학습에 사용되는 데이터의 입력 형태는 (9, 100)이다. 전처리 전과 후의 센서 데이터 파형은 (그림 1)과 같다.



(그림 1) 데이터 시퀀스 파형 전처리 전(좌) 전처리 후(우)

2.3. 비지도 기반 학습 모델

비지도 기반 학습 모델은 AE(AutoEncoder)와 SVDD(Support Vector Data Descriptor)를 결합한 DASVDD를 사용한다[4]. DASVDD는 (그림 2)와 같이 AE를 통해 입력 데이터를 저차원의 잠재 공간으로 인코딩해 잠재 벡터를 생성한다. 인코딩된 잠재 벡터를 SVDD에 적용하여 초구 형태의 경계로 모델링한다. 이후 모델링된 SVDD는 초구 경계를 기반으로 이상치를 식별한다.



(그림 2) DASVDD 구조

2.4. 실험 설계 및 평가 방법

비지도 기반 학습 모델의 학습 데이터로는 정상 행동 데이터만을 사용해야 한다. 하지만 연구에서 사용하는 반려동물 행동 데이터셋에는 정상과 비정상으로 구분되어 있지 않다. 때문에, 고빈도로 발생하는 A행동군 데이터를 학습 데이터로 사용한다. A행동군 데이터로 학습된 DASVDD 모델에 대한 성능 검증을 위해 테스트 데이터를 사용한다. 학습과 테스트 데이터의 개수는 <표 2>와 같다.

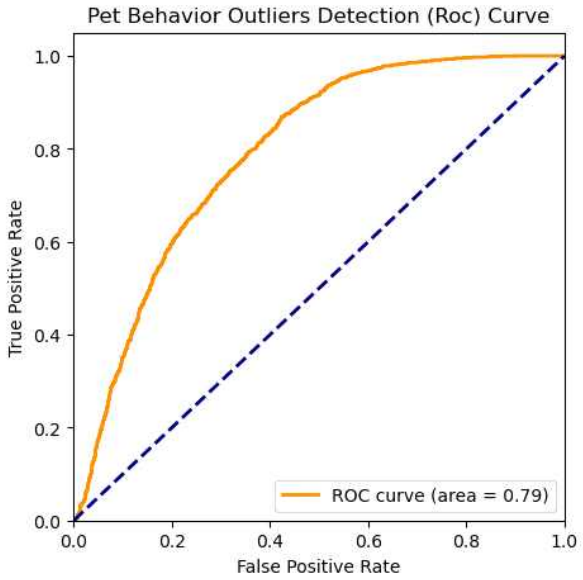
<표 2> 반려동물 데이터셋

	행동군	개수
Train	A	13,369
Test	A	3,343
	B	1,870
총		18,582

성능 평가는 이상치 탐지 모델의 성능을 정량화하는데 사용되는 ROC-AUC(ROC Area Under the Curve)를 사용하여 수행한다[5]. ROC는 다양한 이상치 감지 임계값에 대한 모델의 민감도(TPR: True Positive Rate)와 1에서 모델의 특이도(FPR: False Positive Rate)를 나타낸다. ROC-AUC는 이러한 ROC 커브의 밑면적을 구한 수치이다.

2.5. 실험 결과

학습 데이터를 사용하여 DASVDD 모델 50회 학습을 진행한다. 학습된 DASVDD 모델에 테스트 데이터로 검증을 진행한다. 다음 (그림 3)은 테스트 데이터에 대한 ROC Curve 그래프로 ROC_AUC 기준 79.05%의 이상치 탐지 성능을 나타냈다.



(그림 3) 모델 ROC 커브

3. 결론

DASVDD 모델을 기반으로 반려동물 센서 데이터 이상치 탐지를 진행하는 방법을 제안하였다. 실험에 사용한 데이터에 이상치가 존재하지 않아 반려동물이 고빈도로 행동하는 A행동군, 저빈도로 보여주는 B행동군으로 분리하여 학습을 진행하였다. 학습된 모델의 이상치 탐지의 ROC_AUC는 79.05%를 보여주었다. 하지만 실제 이상치가 아닌 정상 행동을 분리했다는 한계점이 존재한다. 향후 실제 반려동물의 이상치 센서 데이터를 확보함으로써 실제 환경에서 발생하는 다양한 이상치 탐지를 진행 할 수 있을 것이다. 이러한 이상치 탐지를 통하여 반려동물에게 발생하는 위험 상황을 탐지해 사고를 방지할 수 있도록 도움이 될 것이다.

ACKNOWLEDGEMENT

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF- 2021R1A2C2011966).

참고문헌

[1]E. M. Knorr and R. T. Ng, “Finding intensional knowledge of distance-based outliers,” in Proceedings of 25th InternationalConference on Very Large Databases, 1999

[2]Bernhard Schölkopf, Robert Williamson, Alex Smola, John Shawe-Taylor, and John Platt. 1999. Support vector method for novelty detection. In Proceedings of the 12th International Conference on Neural Information Processing Systems (NIPS'99). MIT Press, Cambridge, MA, USA, 582 - 588.

[3]Kim J, Moon N. Dog Behavior Recognition Based on Multimodal Data from a Camera and Wearable Device. Applied Sciences. 2022; 12(6):3199. https://doi.org/10.3390/app12063199

[4]Hadi Hojjati, Narges Armanfard, “DASVDD: Deep Autoencoding Support Vector Data Descriptor for Anomaly Detection”, arXiv:2106.05 410 , https://doi.org/10.48550/arXiv.2106.05410

[5]Campos, G.O., Zimek, A., Sander, J. et al. On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. Data Min Knowl Disc 30, 891 - 927 (2016). https://doi.org/10.1007/s10618-015-0444-8