범용 동물 행동 분석 기술의 최근 동향과 전망

이건석 [1], 최지웅 [1, 2]
[1] 대구경북과학기술원 인공지능전공, [2] 대구경북과학기술원 뇌공학융합센터
gundori0422@dgist.ac.kr, jwchoi@dgist.ac.kr

Toward Generalizable Animal Behavior Understanding

Lee Keon Seok[1], Choi Ji-Woong[1, 2]
[1]The Interdisciplinary Studies of Artificial Intelligence, DGIST,
[2]Brain Engineering Convergence Research Center, DGIST

요 약

뇌질환의 병태생리 규명과 치료 효과의 정량적 평가를 위해 연구자들은 질환 모델 동물을 활용한다. 이러한 동물은 언어로 증상을 표현할 수 없기 때문에, 행동 분석은 뇌기능 변화의 간접적인 지표로서 신경과학 및 생명과학 연구에서 핵심적인 역할을 수행한다. 동물 행동 분석은 실험자의 주관적 판단에 의존한 수동 분석이기 때문에 과도한 시간 소모와 낮은 재현성이라는 한계를 지닌다. 이러한 문제를 극복하기 위해 최근에는 딥러닝 기반의 자세 추정 및 행동 추정 모델들이 활발히 연구되고 있다. 본 논문에서는 두 추정 기법의 한계를 고찰하고, 실험 환경에 구애받지 않는 범용 행동 분석 모델의 필요성과 최근 연구 동향을 정리한다. 아울러, 자세와 행동 간 연계를 통해 보다 정밀하고 유연한 행동 이해를 실현할 수 있는 통합적 방향성을 제안한다.

Ⅰ. 서 론

고령화 사회의 도래와 함께 뇌질환은 인류가 해결해야 할 중대한 과제로 부상하였다. 이에 따라 뇌질환을 심층적으로 규명하기 위한 연구에서 동물 모델의 활용이 활발히 이루어지고 있다. 일반적으로 연구자들은 유전자 조작이나 약물 투여를 통해 뇌질환을 유도한 동물을 질환 모델로 활용한다. 이러한 동물들은 언어로 증상을 표현할 수 없기 때문에, 동물의 행동 변화를 통해 질병의 진행 정도 및 치료 효과를 판단하게 된다. 그러나 연구자가 수동으로 수십~수백 시간에 달하는 동물의 행동을 지속적으로 추적하기에는 한계가 있으므로, 자세를 자동으로 추적하거나 행동을 수치화할 수 있는 다양한 도구들이 병렬적으로 개발되었다. 대표적인 예로 DeepLabCut, B-SOiD 등이 있다. 하지만 이러한 도구들은 실험 조건, 개체 간 특성, 실험실 환경 등에 영향을 받기 때문에 연구자에게 반복적인 파라미터 최적화와 장시간의 학습 데이터 생성이라는 시간 소모적이고 번거로운 작업을 요구한다. 따라서 최소한의 파라미터 수정 및 학습 데이터 생성만으로도 높은 정확도를 유지할 수 있는 범용 모델 및 파이프라인의 개발이 절실하다. 본 논문에서는 자세 및 행동 추정을 위한 도구들과 그 한계를 조망하고, 향후 발전 가능성과 방향성을 제안하고자 한다.

Ⅱ. 본론

2015 년 이후 CNN 의 발전은 컴퓨터 비전 분야에서 딥러닝 기반 접근법의 성능을 비약적으로 향상시켰다. 이와 함께 다양한 데이터셋이 구축되며, 범용 이미지 분류를 위한 사전학습(pre-trained) 모델이 등장하였다. 동물 실험 영상은 확보에 시간이 많이 들고, 라벨링 또한 고비용이기 때문에, 자세 추정처럼 반복적이고 특징이 뚜렷한 작업에 사전 학습된 모델을 적용하려는 연구가 활발히 이루어지고 있다. 예를 들어, 연구자는 전체 프레임 중 일부(약 100~400 장)만 라벨링함으로써 나머지 프레임을 자동으로 라벨링할 수 있는 맞춤형 모델을 구축할 수 있다[1]. 그러나 자세 추정 모델이 개별 프레임을 독립적으로 라벨링하는 방식은, 오류가 발생할 경우 사후 분석의 신뢰성을 저해할 수 있다. 이러한 문제를 해결하기 위해 optical-flow, temporal transformer[2] 등 시공간 정보를 통합하는 접근법이 제시되었으며, 이는 개체 간 구분 능력과 프레임 간 안정성을 향상시켰다.

기존에 사용되던 CNN 의 성능과 범용성을 뛰어 넘을 수 있는 Transformer 기반 비전 모델이 도입되면서 SuperAnimal[3]과 같은 프레임워크가 등장하였다. 아직 해당 모델군에 대한 연구는 초기 단계이나, 다양한 동물 영상 데이터를 수집하여 사전학습 모델을 구축, 배포하려는 시도가 이루어졌다. 그 결과, 라벨 없이도 전문가 수준의 성능을 보이는 zero-shot 모델이 개발되었으며, 최대 27 개의 키포인트를 학습 없이 추정할 수 있게 되었다. 다만, ViT(Vision Transformer) 백본을 사용하는 SuperAnimal 은 한 프레임의 모든 정보를 동시에 처리하는 특성으로 인해, 다중 개체가 등장할 경우 개체 구분이 어려워지는 문제가 있다. 최근에 소개된 TrackFormer, DETR 와 같은 다중 객체 추적(MOT)에 특화된 Transformer 모델을 활용함으로써 개체 수와 상관없이 정밀한 자세 추정이 가능할 것으로 기대된다.

앞서 살펴본 자세 추정 모델은 실험 장치와의 상호작용 분석 등 비교적 단순한 작업에는 효과적으로 적용될 수 있으나, 신경과학적 연구에서는 질환의 경과나 치료 효과를 정량적으로 평가하기 위해 보다 구체적인 미세한 행동의 지속적 추적과 분석이 요구된다. 그러나 행동 추정 역시 연구자가 사전에 정의한 행동 범주와 발생 시점을 수동으로 지정해야 하므로, 본질적으로 노동 집약적인 특성을 지닌다. 더불어 동물 행동에 대한 표준화된 정의가 부재한 상황에서, 분석 결과의 재현성과 정량성이 저해될 수 있다. 이에 따라 행동 추정의 자동화를 위한 다양한 기술적 시도가 이어지고 있다.

행동 추정은 일반적으로 일정 시간창(time-window) 내에서 유의미한 행동 클러스터를 탐색하는 방식으로 수행된다. 추정된 자세 데이터를 저차원 잠재 공간에 투영한 후, 시계열적으로 인접한 자세 시퀀스를 단일 행동 단위로 정의하는 방식이 일반적이며, 이 과정에서 HMM, HDBSCAN 등 시간 기반 군집화 기법이 활용된다. 이러한 방법은 미세한 행동의 변화를 포착하는 데 유리하나, 클러스터 수나 시간창의 길이 등 주요 파라미터에 민감하다는 점에서 실험 조건이나 분석자의 주관에 따른 해석의 편차를 야기할 수 있다.

행동 추정 모델 연구자들은 멀티모달 LLM(V+L)[4]을 도입해 파라미터 민감도를 완화하려는 시도를 하고 있다. 또한, 모델 예측 결과에 대한 연구자의 피드백을 반영하는 continual learning[5] 기법을 통해, 최소한의 라벨링만으로도 높은 성능을 유지하려는 접근이 활발히 이루어지고 있다. 아울러 최근 부상하는 video foundation model 은 다양한 종의 실험 동물에 대해 시공간적 복잡성을 효과적으로 수용하면서도, 소량의 라벨로 질환의 진행과 이에 따른 행동 변화를 정밀하게 추적할 수 있는 잠재력을 보여주고 있다.

Ⅲ. 결론

지난 10 여 년간 동물 행동 분석의 정량화 시도는 자세 추정과 행동 추정이라는 두 축의 병렬적 발전에 의해 이루어져 왔다. CNN 기반의 모델은 소수의 라벨만으로도 고정밀 2D 자세 추정이 가능하다는 점에서 주목받았으나, 개체 수 증가에 따른 오류 누적 문제로 사후 분석이 어려웠다. 이를 극복하고자 시계열 정보를 활용한 모델이 등장하며, 라벨링 부담이 경감되었다. 최근에는 다양한 동물종에서 충분한 키포인트를 추가적인 학습 없이 예측할 수 있는 Transformer 기반의 모델이 제안되며 연구 간 재현성이 개선되었다. 그럼에도 불구하고 다중 개체 혼재 시 자세 추정은 여전히 어려운 문제로 남아 있으며, 시공간 정보 및 해부학적 제약을 통합한 새로운 모델이 요구된다.

행동 추정 자동화를 위한 다양한 시도가 이루어지는 가운데, 멀티모달 LLM 과 continual learning 기법이 점차 도입되고 있다. 이러한 접근은 라벨링 부담과 해석의 주관성을 완화하는 데 기여하고 있으나, 여전히 시간적 연속성과 유연성에 대한 대응에는 한계가 존재한다. 이에 따라, 시공간 정보를 통합적으로 처리할 수 있는 video foundation model 이 이러한 한계를 극복할 수 있는 대안으로 주목받고 있으며, 향후 더욱 유연하고 일반화 가능한 행동 분석 프레임워크로의 발전 가능성을 시사한다.

나아가, 이러한 기술적 축적을 바탕으로 자세 및 행동 추정 모델을 통합한 end-to-end 모델 개발이 본격화되고 있다[6]. 특히, 자세 시계열을 입력으로 활용할 경우 클러스터 안정성과 해상도가 동시에 향상될 수 있음이 보고되었으며, 비디오-언어 모델을 기반으로 행동과 자세를 함께 추론하려는 시도 또한 활발히 이루어지고 있다. 더불어, LiftPose3D 기반의 3 차원 재구성 기술과 SuperAnimal 방식의 다종 전이 학습 기법을 통합할 경우, 다양한 종의 실험 동물에 대해 시공간적 복잡성을 효과적으로 수용하면서도 제한된 라벨만으로 질환의 진행 및 이에 따른 행동 양상의 정밀한 추적이 가능할 것으로 기대된다.

이러한 foundation 모델의 성공적 구축을 위해서는 충분한 양의 데이터셋뿐 아니라, 생후 주령, 유전자형, 약물 투여 시점 등 정제된 메타데이터가 필수적이다. 현재도 흩어져 있는 데이터를 통합하려는 시도가 이어지고 있으나, 공개된 데이터셋은 10 여 종 이하에 불과한 실정이다. 따라서 실험실, 연구자, 프로토콜에 관계없이 보편적으로 적용 가능한 자세, 행동 추정을 위한 foundation 모델 개발을 위해서는 데이터 공유 확대가 필수적이다. 또한, 실시간 분석을 위해서는 양질의 모델 경량화(quantization, knowledge distillation 등) 기술이 함께 요구된다.

경량화 된 자세 및 행동 추정 foundation 모델이 구축된다면, 연구자가 실시간으로 동물의 행동 변화를 모니터링하고, 이를 바탕으로 실험 설계 및 분석 전략을 정교화할 수 있을 것이다. 나아가, 해당 기술이 인간 행동 분석으로 확장될 경우, 개인 맞춤형 치료 기법 개발에도 기여할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 연구는 과학기술정보통신부의 재원으로 한국연구재단의 지원 (RS-2024-00415347, RS-2024-00428887)을 받아 수행된 연구임.

참 고 문 헌

[1] Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nature neuroscience, 21(9), 1281-1289.

[2] Pereira, T. D., Tabris, N., Matsliah, A., Turner, D. M., Li, J., Ravindranath, S., ... & Murthy, M. (2022). SLEAP: A deep learning system for multi-animal pose tracking. Nature methods, 19(4), 486-495.

[3] Ye, S., Filippova, A., Lauer, J., Schneider, S., Vidal, M., Qiu, T., ... & Mathis, M. W. (2024). SuperAnimal pretrained pose estimation models for behavioral analysis. Nature communications, 15(1), 5165.

[4] Xu, T., Zhou, T., Wang, Y., Yang, P., Tang, S., Shao, -K., ... & Yu, J. (2025). MouseGPT: A Large-scale Vision-Language Model for Mouse Behavior Analysis. bioRxiv, 2025-03.

[5] Tillmann, J. F., Hsu, A. I., Schwarz, M. K., & Yttri, E. A. (2024). A-SOiD, an active-learning platform for expert-guided, data-efficient discovery of behavior. Nature Methods, 21(4), 703-711.

[6] Han, Y., Chen, K., Wang, Y., Liu, W., Wang, Z., Wang, X., ... & Wei, P. (2024). Multi-animal 3D social pose estimation, identification and behaviour embedding with a few-shot learning framework. Nature Machine Intelligence, 6(1), 48-61.