

서비스 관점에서 본 영상 인식 AI 클래스의 정의: VOD 오인식 사례를 중심으로

이연주, 이상범
SK 브로드밴드

yunjoo123@sk.com, sb.lee@sk.com

Service Level Definition of Video Recognition AI Class: Focused on VOD Misrecognition Case

Lee Yeonjoo, Lee Sangbeom
SK Broadband

요 약

본 논문은 영상 인식 AI를 서비스로 제공하기에 적절한 클래스별 인식 기준을 정의한다. 기준의 근거로 영상 인식 AI의 VOD 오인식 사례를 제시한다. 오인식을 개선하기 위해 프레임 당 객체 인식보다는 프레임과 프레임 사이의 변화, 즉, 움직임임을 인식하기를 제안한다.

I. 서 론

AI 기술을 개발하는 엔진 개발자와 고객 관점에서 AI 기술을 다루는 서비스 기획자가 기대하는 AI 인식 정확도는 다르다. 개발자는 인식 엔진을 개발하고, 클래스별 precision 과 recall 을 적절한 수준까지 높이는 것을 목표로 하는 반면, 기획자는 기술이 서비스에 자연스럽게 녹아들어 고객이 기술의 존재를 느끼지 못한 채 서비스를 편리하게 사용하길 바란다. 개발자와 기획자의 목표가 다르므로 AI의 정확도에 기대하는 수준 또한 다르다. AI는 100%의 정확도를 보장하지 않는다. 개발자는 AI의 기술적 불완전성을 이해하고 오인식의 가능성을 배제하지 않는다. 반면 기획자는 남아 있는 오인식을 보정하고 기술이 서비스에 100%의 정확도로 노출될 수 있도록 한다. AI가 높은 수준의 정확도를 제공하더라도 기획자는 오인식의 가능성과 사례를 접하는 위치에 있다. 따라서 본 논문은 서비스 관점에서 적절한 AI 인식 기준을 제안하며, 해당 기준을 활용해 사용하기 편리하고 자연스러운 AI 엔진 및 서비스를 개발할 수 있도록 한다.

AI의 한계를 인지한 사람은 AI 서비스의 추천 결과에 적절히 반응할 수 있다. 특히, AI의 오류 발생 상황에 대한 구체적인 이해는 AI와 인간의 협력을 돕는다.[1] 본 논문은 AI의 오류 발생 사례를 제시하여 사용자 및 서비스 운영 담당자의 AI 이해도를 높인다는 의의가 있다.

본 연구의 차별성은 다음과 같다. 첫째, 기존 연구는 AI의 오인식 발생 원인을 개략적으로 설명하며 구체적인 사례를 제시하지 않는다. Bansal, G 외 5인은 AI 오류 범위가 인간의 심성 모형과 일치하지 않아 생기는 문제점

을 지적하며, AI 오인식의 범위를 정의하지만 구체적인 사례는 제시하지 않는다.[1] 둘째, 신경 회로망 개발 논문에서 정의한 오류와 개선 방향은 상세하나 한 가지 분야만을 깊게 다루는 한계가 있다. 예를 들어, CNN 객체 인식 알고리즘의 오인식을 시각화하는 방법을 제안한 연구는 차량 번호판 인식 사례를 다룬다.[2] 오인식의 개선 방법은 명확하지만 다른 분야에는 적용하기 어렵다.

본 논문은 영상 인식 AI를 높은 수준의 서비스로 제공하기 위해 필요한 클래스 별 인식 기준을 정의한다. 기준의 근거로 영상 인식 AI의 VOD 오인식 사례를 제시한다. 오인식 사례를 사전에 고려하여 개발한 AI 엔진은 고객 만족도를 개선하기까지 걸리는 시간을 단축할 수 있을 것이다.

II. 본론

SK브로드밴드는 SK텔레콤의 이벤트 인식 AI 엔진으로 자사 IPTV인 B tv의 VOD를 인식한다. 이벤트란 먹는신, 액션신과 같이 영화, 드라마, 예능 등의 VOD 영상에 자주 나타나는 특정 장면을 의미한다. SK브로드밴드는 인식한 이벤트의 구간을 VOD 재생 중 탐색할 수 있는 INSIDE 서비스를 제공한다. 평균적인 이벤트 인식 정확도는 서비스하기에 충분히 높은 수준이지만, 검수 중 발견한 오인식 사례만을 정리했다.

1. 인식 현황

1) 먹는신 (식사신)

먹는신이란 등장인물이 음식을 먹는 장면을 일컫는다. 먹는신의 첫 번째 오인식 사례는 음식이 아닌 객체가 등장인물의 입에 가까이 위치했을 때 발생한다. 흡연하는 장면, 스마트폰으로 전화하는 장면이 대표적인 예시이다.

두 번째 사례는 영상 내에 음식이 있으나 등장인물이 음식을 먹지 않는 장면이다. 식탁에 음식이나 컵이 놓여있는 장면, 음식을 요리하거나 요리에 관해 설명하는 장면은 먹는신에 해당하지 않는다. 세 번째 사례는 엔진의 인식 기준이 명확하지 않아 발생한 것으로, 물, 음료수, 술 등 액체 형태의 음료를 마시는 장면은 서비스 관점에서 먹는신이 아니라 마시는 신으로 인식해야 한다.

2) 포옹신

포옹신은 두 사람이 껴안는 장면으로 정의한다. 연인이나 가족 간에 포옹 장면이 발생할 가능성이 높다. 포옹신의 오인식은 포옹이 아닌 신체적 접촉으로 인해 발생한다. 어깨동무, 다툼에 의한 신체적 접촉, 가까운 거리에서 대화하는 장면을 포옹신으로 인식한 사례가 있다. 또한, 등장인물이 한 명이지만 특정 객체와 가까이 있어 오인식이 발생할 수 있다. 등장인물이 이불을 끌어안거나, 팔짱을 끼고, 쭈그리고 앉아있는 장면은 포옹신으로 인식하지 않도록 유의해야 한다.

3) 충격신

충격신은 전쟁 및 액션 영화에 자주 나오는 총기류를 활용한 격투 장면을 의미한다. 서비스 관점에서 총을 발포하지 않은 장면은 충격신으로 보기 어렵다. 하지만 총을 겨누거나 들고 있고, 발포가 되지 않은 장면이 충격신으로 인식된 사례가 있다. 총을 오랜 시간 겨누더라도 충격신의 러닝타임 내에서 총알이 총 밖으로 발포되어야 한다.

4) 댄스신

음악에 맞추어 춤을 추는 장면을 뜻한다. 음악이 나오지 않더라도 리듬에 맞게 몸을 움직이면 댄스신으로 취급한다. 싸우는 장면(액션신), 신이 나서 뛰어다니는 장면, 조강하는 장면, 줄넘기하는 장면을 댄스신으로 오인식한 사례가 있다. 댄스신이 다른 격한 움직임과 구분되기 위해선 엔진이 움직임의 패턴을 인식할 필요가 있다.

5) 승마신

말 위에 사람이 타고 이동하는 장면을 뜻한다. 말이 등장하지만, 사람이 타고 있지 않은 장면은 승마신으로 정의하지 않는다. 말이 마차를 끌거나, 말과 사람이 나란히 걷는 장면, 낙타에 짐을 싣고 가는 장면을 승마신으로 오인식한 사례가 있다. 이러한 장면은 서비스 관점에서 승마신으로 분류할 수 없다.

6) 결혼신

한 쌍의 커플이 서양식 결혼식을 치르는 장면으로 정의한다. 여성은 흰색 웨딩드레스를 입고 남성은 턱시도를 입는다. 두 사람은 주변 사람들의 축하를 받으며 결혼식장의 중심을 걷는다. 여성이 웨딩드레스를 입었더라도 결혼식장에 있지 않으면 결혼신으로 분류할 수 없다. 웨딩드레스를 착용해보는 장면이나 웨딩 촬영하는 장면이 대표적인 오인식 사례이다. 또한, 한복이나 중국 사극에 나오는 길고 화려한 복식이 웨딩드레스와 유사해 결혼신으로 오인식한 사례도 있다.

7) 생일신

서양식 생일 파티 장면을 뜻한다. 주로 생크림 케이크에 초를 켜고, 여러 사람이 모여 축하한다는 특징이 있다. 생일이 아니더라도 케이크가 있거나, 향을 피우면 생일신으로 오인식한 사례가 있다. 또한, 생일이 아닌 파티를 생일신으로 인식하기도 한다. 대표적인 예시로 사람들이 술집에서 먹고 마시는 장면이 있다. 케이크나 촛불

등 일부 객체만이 생일신을 인식하는 기준이 되지 않도록 해야 한다.

6) 베드신

베드신은 두 사람의 성관계 장면을 의미한다. 성관계와 관련 없이 침대에 누워있거나 키스신만 있는 장면을 베드신으로 오인식한 사례가 있다. 특히, 인체에 노출이 많은 장면을 베드신으로 인식하는 오류가 자주 발생했다. 어른이 아이를 씻기거나 옷을 벗고 자는 아이들이 나오는 장면이 대표적인 예시이다. 노출의 여부만이 베드신의 인식 기준이 되지 않도록 유의해야 한다.

2. 개선 방향

AI 엔진의 강력한 성능 뒤에는 AI의 원리를 파악할 수 없게 하는 블랙박스라는 특성이 있다.[3] AI가 오인식의 이유를 알려주지 않으므로 오인식 사례로 그 원인을 유추해야 한다. 영상 인식 AI의 오류 사례로 미루어보아, 해당 엔진이 영상 내에 발생하는 동작보다는 객체를 중심으로 이벤트를 인식한 것이 한계로 작동했다.

이벤트의 발생에 수반되는 부수적인 객체와 행위의 등장 여부보다는 중심이 되는 행위의 변화, 즉, 움직임을 정확하게 인식할 필요가 있다. 따라서 프레임 단위로 객체를 인식하기보다 프레임과 다음 프레임 사이의 변화를 인식할 것을 제안하는 바이다.

III. 결론

AI 기반의 서비스는 서비스 출시 후 개선이 이루어지기까지 많은 시간이 소요된다. 기획자는 엔진 인식 결과를 보고 개선할 지점을 찾는다. 이를 반영하기 위해선 AI 학습용 자료를 대량으로 수집해야 하므로 요구사항을 즉각적으로 반영하기 어렵다. 영상 인식 AI 클래스에 대한 엔진 관점의 정의와 서비스 관점 정의의 간극을 줄이기 위해 서비스 관점의 클래스에 대한 명확한 사전 정의가 필요하다.

본 논문을 활용해 서비스 관점에서 자연스러운 AI 인식 기준을 세울 수 있다. 영상 인식 외에도 다양한 분야의 AI 엔진의 오인식 사례와 오류 원인, 그리고 개선 방안에 관한 후속 연구가 진행되어야 할 것이다.

참 고 문 헌

- [1] Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E., "Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance," Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, 7(1), 2-11, 2019.
- [2] Kubota, T., Murata, Y., Uehara, Y., & Nakagawa, A., "A method for visualizing the cause of misrecognition in object recognition using CNN," IEICE Technical Report; IEICE Tech. Rep., 119(193), 99-104, 2019.
- [3] Adadi A., Berrada M., "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," IEEE Access, vol. 6, pp. 52138-52160, 2018.