

방송미디어를 위한 AI 기술 개발 현황

글. 오주현 KBS 미디어기술연구소 AI&DATA팀

작년에 본 지면(방송과기술 7, 8월호)을 통해 인공지능(Artificial Intelligence, 이하 AI) 기술의 개요와 응용 분야, 그리고 기술 도입 방식 등에 관한 방송 미디어 관점의 논의를 소개한 바 있다. 해당 기고에서는 AI가 일부 분야에서 사람의 능력을 뛰어넘게 된 기술 발전 현황과 여러 가지 방송 미디어 응용 분야 즉, 뉴스 프레젠테이션, 품질 관리, 메타 데이터와 검색, 자동 제작, 미디어 포렌식, 광고, 디지털과 모바일 등에 사용될 수 있는 방안에 대해 소개하고, 이러한 AI 기술을 도입하는 방향에 대해서도 함께 논의하였다.

본 고에서는 이와 관련하여 KBS 미디어기술연구소에서 실제로 개발 중인 AI 기술에 대해 소개하고자 한다.

AI 엔진

지금 KBS 미디어기술연구소에서 개발하고 있는 것은 AI 기술을 응용한 특정 시스템이나 서비스라기보다 AI ‘엔진’에 가깝다. 왜 AI 엔진인가? 많은 사람이 궁금해하는 부분이다. 엔진은 기반기술의 집합체인데 기술 기업이 아닌 KBS와 같은 방송사에 그런 기술 경쟁력이 있는가? 자동차로 치면 엔진을 사서 쓰거나, 아니면 완성차를 사거나 빌려서 써야 하는 것 아닌가?

상당 부분 맞는 말이다. AI 엔진도 최첨단 기술의 집약이기 때문이다. 그러나 진입장벽이 매우 높은 자동차 기술과 달리, AI 관련 기술은 이미 많은 부분이 공개되어 사용 가능하다는 점에서 큰 차이가 있다. AI 기술에서는 파이썬(Python)과 텐서플로(Tensorflow), 케라스(Keras)¹⁾ 등 공통 프레임워크와 개발도구를 사용하여 개발된 최신 AI 구현들이 github²⁾ 등을 통해 소스코드와 함께 배포된다. 이와 같이 공개된 기술을 이용하고 개선함으로써 매우 빠른 발전이 이루어질 수 있다. 첨단 AI 관련 논문들은 긴 심사과정을 거치는 저널보다 논문공유 사이트인 arXiv³⁾에 먼저 올라오고, 다른 연구팀에서는 즉각 이를 참조하여 개선 연구가 이루어진다. 딥러닝 시대가 되면서 학계에서는 긴 시간 동안 기존 기술을 연구해온 박사과정보다 석사과정에 갓 들어온 연구원이 더 좋은 결과물과 논문을 내는 경우가 심심찮게 나온다고 한다. 그만큼 진입장벽이 낮아졌다는 이야기다.

1) 파이썬으로 작성된 오픈 소스 신경망 라이브러리로 주로 텐서플로 상위에서 사용

2) 분산 버전 관리 툴인 git을 사용하는 프로젝트를 지원하는 웹호스팅 서비스

3) ‘아카이브’, 수학, 물리학, 천문학, 전산 과학, 계량 생물학, 통계학 분야의 출판 전 논문을 수집하는 웹사이트

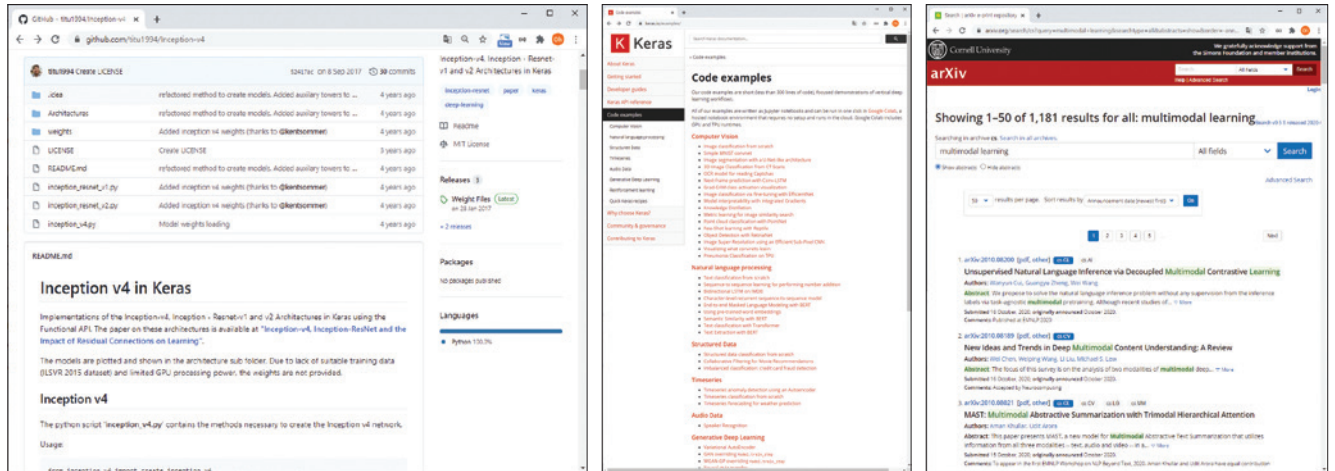


그림 1. 좌로부터 github, Keras, arXiv 사이트

AI 기술의 시초는 1940년대부터 제안된 뉴런(neuron)과 인공신경망(artificial neural network)이다. 그동안 수많은 알고리즘이 제안되고 기술 개선이 이루어짐으로써 과적합(overfitting)과 그래디언트 소실(vanishing gradient) 문제 등을 해결하였고, RNN⁴⁾, GAN⁵⁾ 등의 다양한 형태로 발전할 수 있었음에도, 여전히 AI 기술의 근간은 인공신경망에 있다. 2020년 현재 적어도 학계가 아닌 산업계에서 AI 개발의 핵심은 더 이상 새로운 AI 알고리즘을 개발하는 데 있지 않다. 주어진 인공신경망에 다양한 학습 데이터를 입력하여 원하는 기능을 수행하는 AI 모델을 얻어내는 것으로 충분하다. 누구나 파이썬과 주피터 노트북⁶⁾을 이용하여 원하는 AI 모델을 개발할 수 있다. C++와 STL⁷⁾같은 어렵고 불친절한 도구를 사용하던 개발자에게는 AI 개발에 주로 사용되는 언어인 파이썬 자체가 축복이다. 이에 대해서는 파이썬의 편리함과 높은 생산성을 보여주는 [그림 2]의 인터넷 게시물과 만화가 참고가 될 것 같다.



그림 2. 파이썬의 장점을 보여주는 인터넷(reddit) 게시물과 만화(xkcd.com)

- 4) Recurrent Neural Network, 순환신경망
- 5) Generative Adversarial Network, 적대적 생성 신경망
- 6) Jupyter Notebook, 파이썬 코드를 작성하고 실행해볼 수 있는 웹 기반 환경
- 7) C++에서 다양한 자료구조와 알고리즘을 제공하는 표준 템플릿 라이브러리

[그림 3]에는 AI 개발 예제로 흔히 사용되는 ‘MNIST’ 숫자 이미지 데이터셋을 학습하고 주어진 이미지를 판별(classification)하는 노트북 예제를 나타내었다. 86.4%의 정확도를 가지는 숫자 인식기를 개발하는 데 (중간의 데이터 확인까지 포함하여) 겨우 30라인 남짓한 코딩으로 충분한 것이다. 여기에 구글 코랩⁸⁾을 사용한다면 이 일을 하는 데 필요한 것은 웹브라우저뿐이며, 개발 환경 설정은 물론이고 심지어 AI 계산을 수행할 시스템조차 없어도 된다.

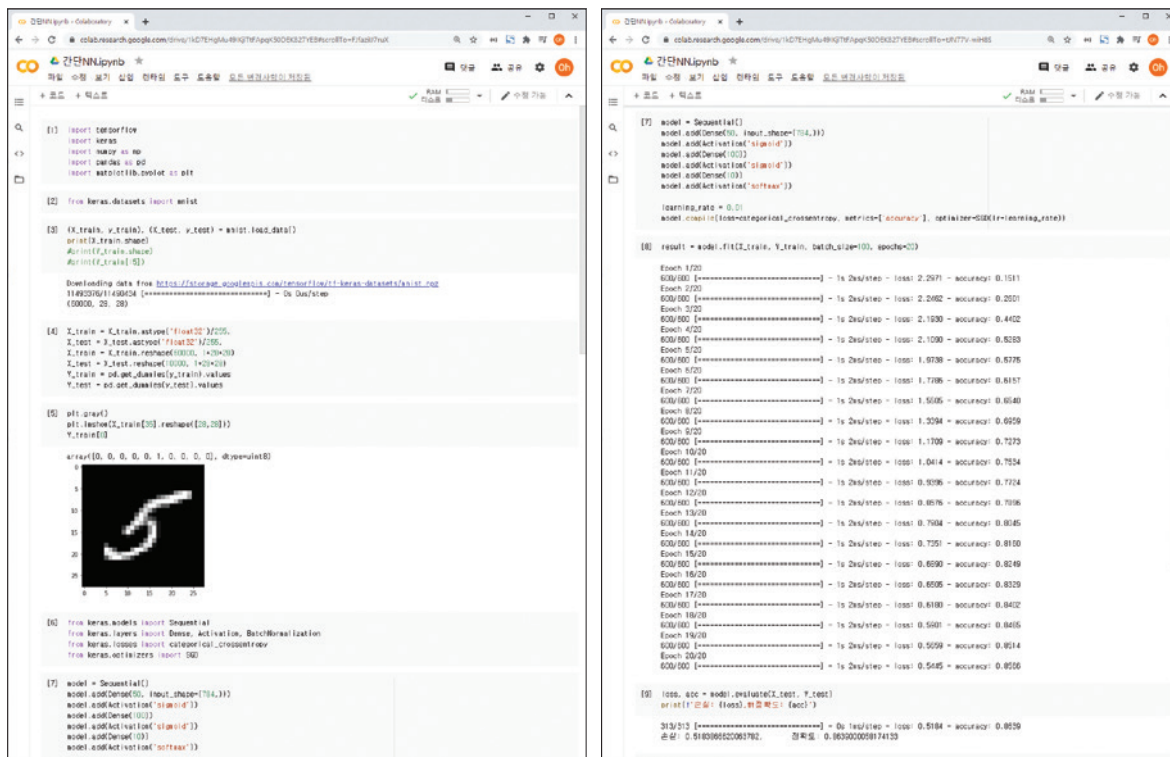


그림 3. 간단한 숫자 인식 예제 노트북

요컨대 지금은 AI 구현에 대한 진입장벽이 그 어느 때보다도 낮은 시대이며, AI 기술이나 알고리즘보다는 AI를 학습시키기 위한 데이터셋이 중요한 시대가 되었다. Kaggle⁹⁾에서는 무려 5만 개의 퍼블릭 데이터셋이 제공된다. 방송사에서 AI 엔진을 인하우스(in-house)로 개발할 수 있는 것도 이러한 이유에서다. 예전에 비해 AI 구현이 용이해진 데다, 퍼블릭 데이터셋 외에도 방대한 분량의 아카이브 콘텐츠를 AI 학습에 사용할 수 있는 방송사는 그만큼 AI 구현에서 유리한 위치에 있기 때문이다.

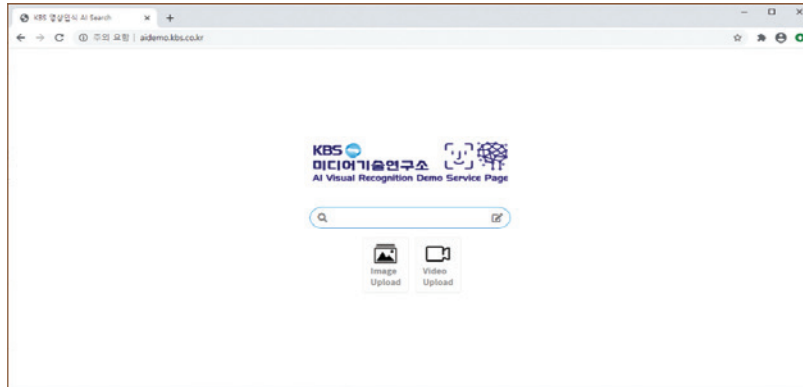
KBS AI 엔진 소개

KBS 미디어기술연구소에서 개발 중인 AI 엔진은 영상으로부터 인물, 사물, 배경 등의 인식이 가능하다. [그림 4]는 개발 중인 AI의 성능을 확인할 수 있는 일종의 개념 검증(POC) 사이트이다(콘텐츠 저작권 등을 감안하여 현재 KBS 내부에서만 접속 가능).

우선 이미지 분석 기능을 확인하기 위해 KBS 드라마 ‘출사표’에 등장한 나나의 사진과 드라마 ‘99억의 여자’ 포스터를 업로드해 보면 [그림 5]와 같은 분석 결과를 얻을 수 있다. 사진의 경우 인물은 ‘나나’, 사물은 ‘침바지’, ‘여자’, ‘얼

8) Google Colaboratory, 구글에서 제공하는 클라우드 기반 주피터 노트북 환경. <https://colab.research.google.com>

9) <https://kaggle.com>, 기업 및 단체에서 데이터와 해결과제를 제공하면 데이터 과학자들이 이를 해결하는 모델을 개발하고 경쟁하는 플랫폼



굴’ 정도를 인식하고, 배경은 ‘무대/실내’일 확률이 41%라고 분석하였다. 포스터의 경우 ‘남자’, ‘여자’, ‘한경’ 등의 사물과 등장인물을 대부분 맞췄지만 정웅인이 ‘박중훈’으로 잘못 인식되었다(변명을 하자면 사람이 뵈을 때도 비슷하게 보이는 면이 있다).

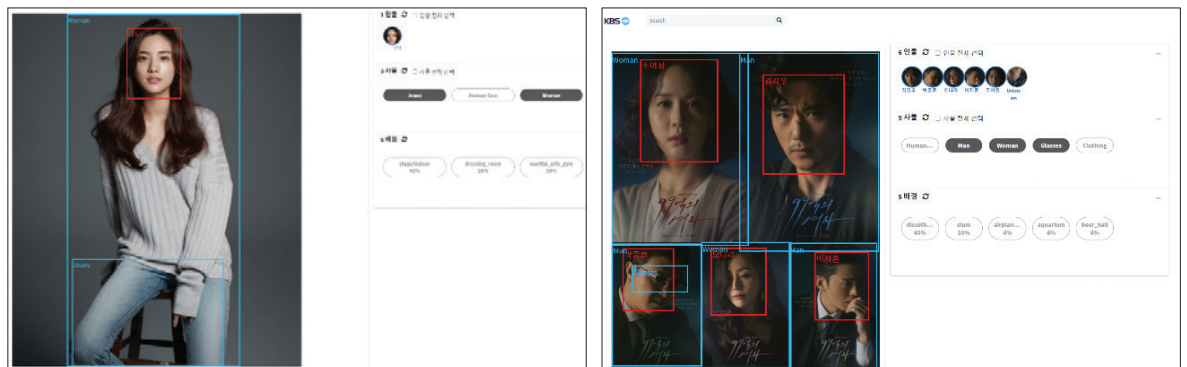


그림 5. 연예인과 드라마 이미지 인식 결과

[그림 6]에서 정치인 사진을 분석한 결과, ‘이낙연’, ‘김부겸’, ‘박주민’ 등의 인물과 ‘남자’, ‘얼굴’, ‘정장’ 등의 사물을 인식하였다. 이어 불인 사진이라 큰 의미는 없지만 배경은 ‘컨퍼런스 센터’로 인식하였다.

개발된 AI 엔진과 검증 사이트는 [그림 7]과 같이 비디오 분석 기능도 제공하고 있다. AI가 분석한 인물, 사물, 배경 정보가 타임라인으로 배치된 결과를 확인할 수 있다. 또한 특정 인물, 사물, 배경이 등장하는 장면을 타임라인 상에서 쉽게 이동하여 확인할 수 있도록 하였다.



그림 6. 정치인 이미지 인식 결과

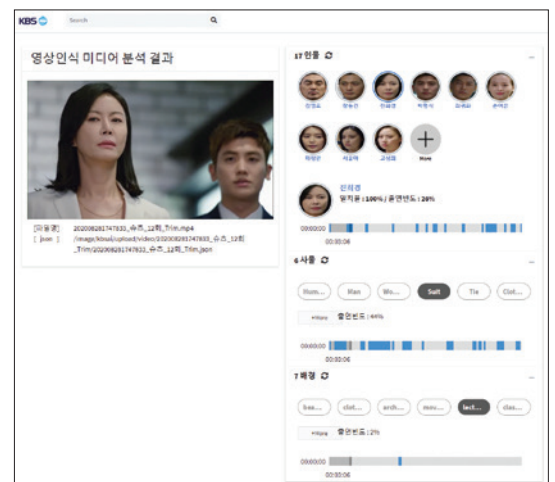


그림 7. 드라마 ‘슈츠’의 짧은 클립 영상 분석 결과



그림 8. '정은경' 인물 검색 결과

소주 광고에 등장한 적이 있는 '아이린'을 검색하면 [그림 9]와 같이 소주병에 붙어있는 얼굴이 검색되기도 한다.

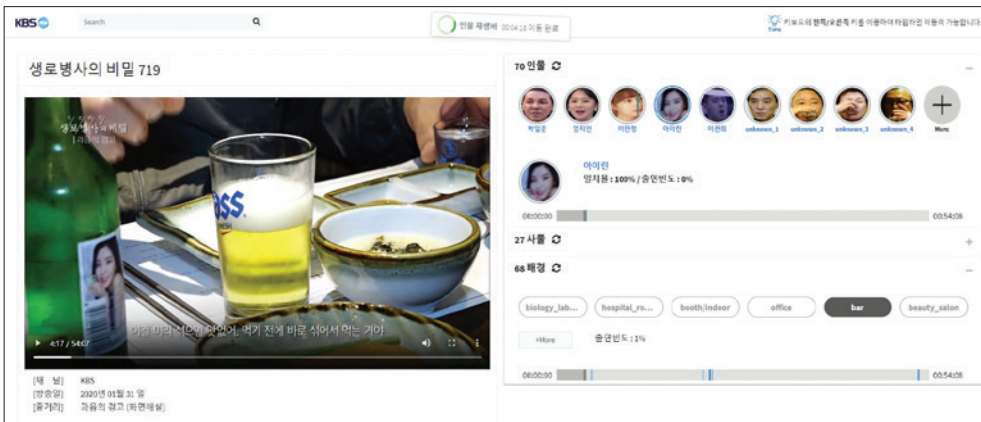


그림 9. '아이린' 검색 결과

이와 같은 인물 인식을 구현하기 위해서 별도의 인물 레이블링 도구를 개발하고 출연 인물 데이터셋을 구축하였으며, 앞으로도 지속적으로 업데이트할 수 있는 방안을 고민 중이다. 사물과 배경에 대한 인식 결과는 '걸어서 세계속으로 512회' 분석 결과를 통해 살펴보고자 한다. [그림 10]과 같이 타임라인 상에서 사물과 배경 인식 결과를 확인할 수 있다.

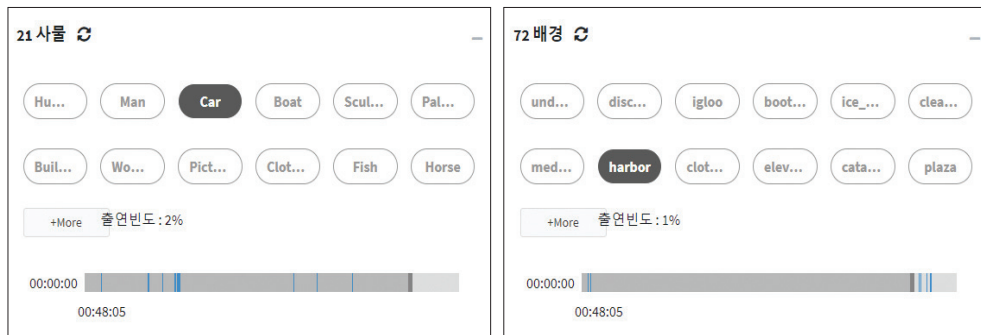


그림 10. 사물과 배경 인식 및 타임라인 표시

인식된 사물과 배경이 등장하는 타임라인으로 이동할 수 있다. [그림 11]은 인식된 몇 가지 사물이 등장하는 장면을 보여준다.

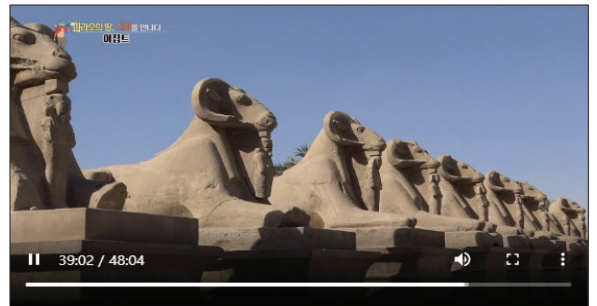
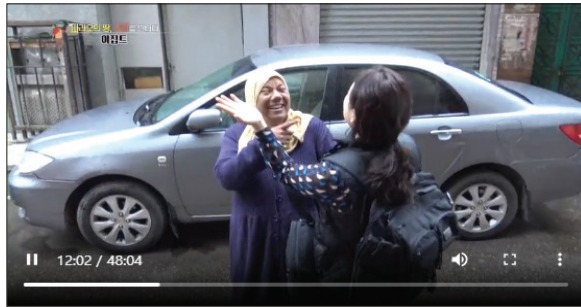


그림 11. AI가 찾아낸 사물 등장 화면, 위로부터 자동차, 조각(sculpture), 말(horse)

[그림 12]는 몇 가지 배경을 인식한 결과를 보여준다. 배경 인식은 인물이나 사물 인식만큼 그 정확도가 높지는 않지만 제작이나 검색에 참고가 될 수 있을 것이다.



그림 12. AI가 분석한 배경, (위) 수중, (아래) 항구

이상으로 미디어기술연구소에서 개발 중인 AI 엔진을 검증 사이트를 통해 알아보았다. 개발 중인 AI 엔진은 메타데이터 입력과 검색 등 AI 분석이 필수적인 분야에서 유용하게 쓰일 수 있을 것이다.

VERTIGO

개발 중인 AI 엔진은 [그림 13]의 KBS 세로영상 제작 시스템 ‘VERTIGO’에도 적용되어 활용 중이다. VERTIGO는 AI 기술을 이용하여 뮤직뱅크 등 영상에서 아이돌을 검출하고 추적하여 ‘직캠’과 같은 영상을 생성한다.



그림 13. 세로영상 제작 시스템 VERTIGO

AI 기반 자막 개선

KBS 미디어기술연구소에서는 장애인방송, 그중에서도 자막방송 서비스를 개선하기 위한 정책과제에 참여하고 있다. 기존 장애인방송 자막을 개선하여 [그림 14]와 같이 화자의 위치를 고려한 자막을 생성하는 데 주안점을 두고 있다. 이와 같은 효과를 위해 멀티모달(multi-modal) AI 기반의 화자 검출 기술을 연구하고 있으며, 이는 향후 AI를 이용한 제작 자동화를 위한 기반 기술이 될 것이다. 이와 함께 [그림 15]와 같이 저장된 자막을 온라인 등에서 2차 활용할 수 있도록 AI 기술을 이용하여 자막을 영상과 동기화하는 연구도 수행하고 있다.



그림 14. 화자의 위치를 반영한 자막 배치(넷플릭스)

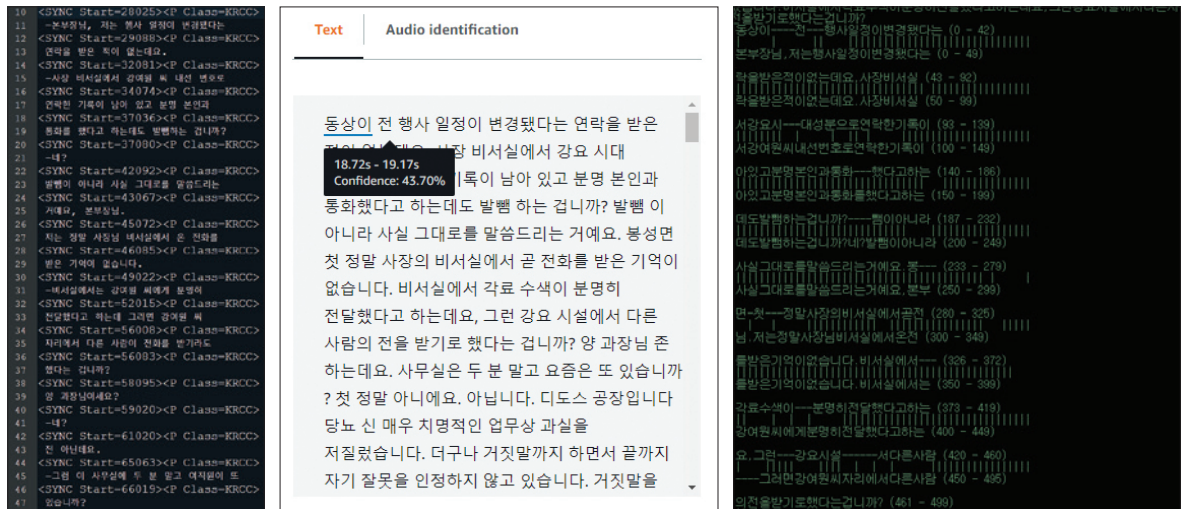


그림 15. 자막 동기화 연구. 좌로부터 폐쇄자막, 음성인식 자막, 자막 정렬(alignment) 결과

결론

이상으로 KBS 미디어기술연구소에서 개발 중인 AI 기술에 대해 알아보았다. 소개한 바와 같이 AI 기술은 더 이상 먼 미래의 일이 아니라 메타데이터 입력이나 자막 동기화 등 많은 인력이 소요되는 단순반복적 업무에 충분히 사용할 수 있는 수준에 도달하였다. 다만 기술의 하이프 사이클(hype cycle)에서 나타나는, 무작정 AI를 도입하고자 하는 현상 또한 경계할 필요가 있는데, 방송을 포함한 미디어 업계에서 AI 기술을 ‘무엇 또는 어디에’ 그리고 ‘어떻게’ 도입할 것인가에 대한 고민을 진지하게 해볼 때라고 생각한다.

KBS 미디어기술연구소 또한 개발 중인 AI 엔진을 기본으로 하여 다양한 제작과 서비스 현장의 요구사항을 받아 구현하고 제공함으로써, AI를 위한 AI가 아닌 현업과 서비스를 위해 필요한 AI를 개발할 수 있도록 할 계획이다. 