

# KAKAO AI REPORT

vol.14  
2018 AUTUMN

# KAKAO AI REPORT

Vo1.14

발행일 | 2018년 9월 28일

발행처 | (주)카카오

발행인 | 여민수, 조수용

편집인 | 김대원

편집장 | 최은필

편집 | 김수정, 심지은, 연지은

디자인 | 허진아

이메일 | kakaoaireport@kakaocorp.com

브런치 주소 | <https://brunch.co.kr/magazine/kakaoaireport>

본 <카카오 시리포트>의 판권은 주식회사 카카오가 소유하고 있으며, 카카오는 <카카오 시리포트>를 상업적 목적으로 이용하는 것을 금지합니다. 또한 저작권법에 따른 정당한 범위 내에서 <카카오 시리포트>를 이용·가공·인용하는 경우라도 반드시 출처를 밝혀주시기 바랍니다.

## contents

<b>preface</b>	<카카오 시리포트> 14호를 내며	03
<b>Special Topic</b>	지식그래프: 카카오미니와 검색 적용 소개	08
if (kakao) dev 2018	눈으로 듣는 음악 추천 시스템	14
	이미지로 이미지 검색하기	24
	딥러닝을 활용한 뉴스 메타 태깅	30
	딥러닝을 이용한 실시간 인코딩 효율 최적화	38
	카카오 봇 플랫폼 소개	48
<b>In-Depth</b>	개황   디지털 사회로의 변화와 스포츠	56
<b>Sports &amp; AI</b>	경기   스포츠 경기력 향상을 위한 AI 활용 방안	64
	분석   스포츠 경기 분석 전문가와 AI의 만남	70
	판정   로봇 심판과 판정 알고리즘의 의미 있는 도전	78
	도핑   스마트 도핑 검사의 도입과 발전 과제	84
	보도   스포츠 저널리즘의 변화와 AI의 활용	92
	챗봇   만능 스포츠봇의 등장과 발전 방향	98
<b>Tech Insider</b>	Apache S2Graph 기반 머신러닝 모델 환경 구축 <span>AI in Kakao</span>	108
	딥러닝이 탐구하지 못한 언어와 5가지 태스크	116
	2018 ICML을 통해 살펴보는 AI 연구 동향	126
	2018 CVPR 논문 동향 및 주요 연구 소개	138
	카카오 크루들의 커피 주문 방식	146

# 〈카카오 시리포트〉 14호를 내며

매월 찾아뵙다 3개월 만에 인사를 드립니다. 그간 〈카카오 시리포트〉 편집진은 개편 작업을 진행했습니다. 〈카카오 시리포트〉가 지향해온 정신과 방향성은 유지하되 더 다양한 생각과 정보를 보다 쉽고 편하게 전달해드리는 것으로 방향을 모색하였습니다. 인쇄본을 접하신 분들은 먼저 사뭇 달라진 디자인과 마주하셨을 겁니다. 디자인 변화 작업은 겉과 판형뿐 아니라 내부 편집에도 적용했습니다. 기존 2단 구성 편집을 1단 구성으로 바꾸어 독자 여러분들이 더욱 편하게 읽으실 수 있도록 변경하였습니다.

이번 호의 첫 기획으로는 지난 9월 4일 진행된 카카오 개발자 콘퍼런스 ‘if (kacao) dev 2018’ 중 인공지능(AI) 관련 내용입니다. 현장에 오시고 싶었지만 사정상 그러지 못했던 분들, 현장에 오셨더라도 다시 한 번 발표 내용을 곱씹어보고 싶으신 분들을 위해 준비했습니다. 이를 위해 편집진은 대학(원)생 기자단을 꾸렸습니다. 다시 한 번 이 자리를 빌려 멋진 글을 보내준 기자단 여러분들께 감사드립니다. 지면 사정상 이번 호에 담지 못한 콘퍼런스 내용은 브런치를 통해 소개해드릴 예정입니다(<https://brunch.co.kr/magazine/kakaoaireport>).

평창 동계올림픽, 러시아 월드컵을 거쳐 최근 성황리에 종료된 자카르타 팔렘방 아시안 게임까지, ‘스포츠의 해’라고 해도 과언이 아닐 정도로 수많은 국제적 이벤트가 있었던 한 해였습니다. 그래서 이번 호 기획 기사의 테마로 ‘AI와 스포츠’를 선정했습니다. 경기력부터 심판, 도핑 검사, 그리고 언론보도와 챗봇까지 스포츠와 관련된 일련의 활동에 가미된 AI의 양상과 형태를 살펴봤습니다. 이밖에 머신러닝 모델을 실제 서비스에 적용하기 위한 환경 조성 이야기, 딥러닝 기술이 아직 정복하지 못한 언어와 주어진 과제, 2018 ICML & CVPR의 학회 연구 동향, 그리고 카카오 크루들이 챗봇을 이용해 사내 카페를 이용하는 방법을 담아냈습니다.

몇 달간의 리뉴얼 호를 기획하고 준비 과정을 거치면서 사회 곳곳의 여러 목소리들을 들었습니다. 〈카카오 시리포트〉는 AI에 대한 다양한 생각과 기술을 공유하는 장으로서, 시가 건전한 방향으로 세상을 바꾸는 데 작은 도움이 될 수 있도록 앞으로도 노력하겠습니다. 다음 계절의 초입에 또 다른 〈카카오 시리포트〉로 인사드리겠습니다. 감사합니다.

2018년 9월 28일  
카카오 시리포트 편집진 드림

# Special Topic

if (kakao) dev 2018

지식그래프: 카카오미니와 검색 적용 소개	08
눈으로 듣는 음악 추천 시스템	14
이미지로 이미지 검색하기	24
딥러닝을 활용한 뉴스 메타 태깅	30
딥러닝을 이용한 실시간 인코딩 효율 최적화	38
카카오 봇 플랫폼 소개	48



## if (kakao) dev 2018

카카오는 지난 9월 4일 'if (kakao) dev 2018'이란 이름의 AI 관련 콘퍼런스를 개최했습니다. 이번 콘퍼런스에서는 카카오의 AI 기술 및 비즈니스 현황을 중심으로 블록체인, 은행, 페이 등 카카오에서 개발하고 있는 다양한 기술 및 서비스와 관련한 내용들을 다루었습니다.

이번 행사에서는 27개의 세션이 열렸습니다. 다채로웠던 행사였던 만큼 볼거리가 많았습니다. <카카오 시리포트> 편집진은 총 6명으로 구성된 대학(원)생 기자단을 꾸려 이 행사를 취재했습니다. 이는 많은 '개발자들과 AI 생태계를 만들고, 미래 기술을 공유하고 나누는'는 콘퍼런스의 취지에 부합하기 위한 선택의 일환이었습니다. 사전 공모 방식으로 선발된 대학(원)생 기자단은 각각 2개 세션의 취재를 맡았습니다. 개인별 취재 결과 중 하나는 이번 호에, 다른 하나는 <카카오 시리포트> 브런치에 게재됩니다. 본문에 삽입된 슬라이드 자료는 'if kakao 2018' 개발자 콘퍼런스 발표자료에서 인용하였습니다. (출처: <https://if.kakao.com>)

# 지식그래프: 카카오미니와 검색 적용 소개

인터넷을 검색하다 보면 답답한 순간들이 많다. 찾아내고 싶은 것이 있는데 목적과는 다르게 검색 결과는 엉뚱한 결과들만 나열한다. 예컨대 이효리 남편이 소속된 그룹에 어떤 멤버들이 있는지 궁금할 때 이것을 검색으로 찾는다고 가정해보자. 가장 직관적인 검색 방식은 '이효리 남편이 소속된 그룹의 멤버'라고 검색하는 것이지만, 우리가 만족할 만한 결과가 나오지 않는다. 다음으로 생각해볼 수 있는 방식은 '이효리 남편'을 검색하여 '이상순'이라는 결과값을 얻은 후, 다시 '이상순의 그룹'을 검색해보는 것이다. 그리고 이에 대한 결과값인 '롤러코스터'를 클릭하여 '조원선'이 해당 그룹의 멤버임을 알아낸다. 이렇듯 우리는 원하는 정보를 찾아내기 위해 적잖은 불편을 감수해야 한다.

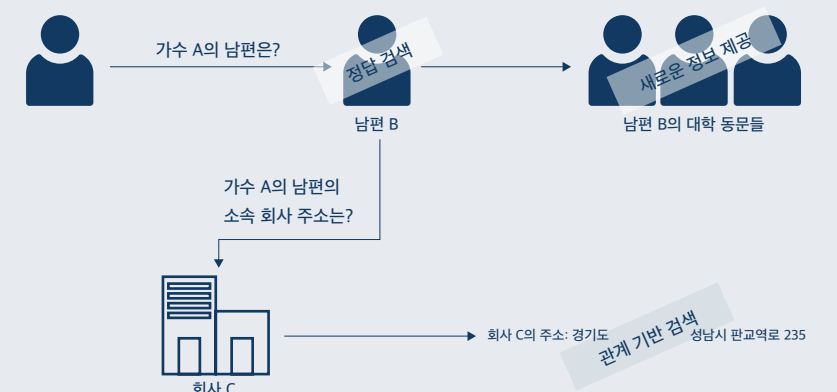
이 글에서는 이러한 검색의 불편함을 해소할 수 있는 '카카오 지식그래프와 적용'에 대해 다루어보고자 했다.

## 지식그래프의 의미와 편리성

지식그래프란 관련 있는 정보와 지식들을 서로 연결해놓은 것을 의미한다. 예를 들어 기존에는 '아이유'의 직업, 성별, 소속사, 데뷔 연도 등의 기본 정보와 방송 출연, 공연 등이 각각 다른 문서에 분절되어 저장된 상태였다면, 지식그래프에서는 이들을 '아이유'와 관련 있는 것으로 모두 엮어놓는다. 하나의 생각이 꼬리를 물고 관련된 다른 생각으로 이어지는 마인드맵처럼 지식들을 이어놓은 것으로 이해할 수 있다. 따라서 지식그래프에서의 '검색어'는 단순히 독립된 텍스트를 넘어 주변 정보들과 관계를 지니는 의미어로 확장된다.

이렇듯 지식그래프의 연결된 특성은 사용자의 검색을 편리하게 해준다. 앞에서 언급한 예제의 경우 지식그래프를 활용한다면, '이효리 남편이 소속된 그룹의 멤버'라고만 검색하여도 '조원선'이라는 결과값을 곧장 얻을 수 있다. 아래 사진에서와 같이 이효리 - 이효리의 남편 이상순 - 이상순의 그룹 롤러코스터 - 롤러코스터의 멤버 정보가 유기적으로 연결되어 있기 때문이다. 기존 검색보다 사용자가 훨씬 편리하게 원하는 정보를 찾을 수 있게 된 것이다.

[그림 1] 지식그래프의 연결 특성을 이용한 관계 기반 검색 예시



컨퍼런스 발표 | 남기훈 kane.nam@kakaocorp.com

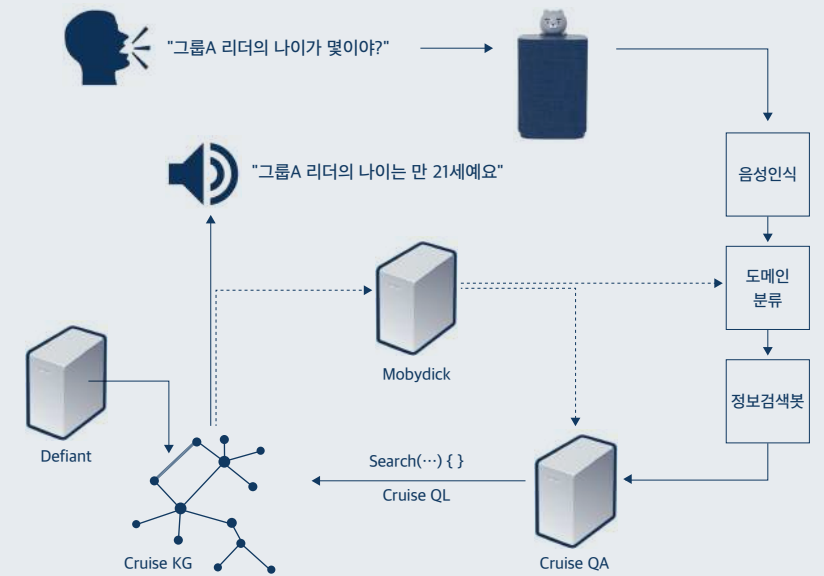
글 | 김동현 dobedh@gmail.com 서울대학교에서 교육상담을 전공하였습니다. 사람에게 관심이 많아 학부에서는 심리학을, 대학원에서는 상담을 공부하였습니다. 만드는 것에 대한 열정과 사람에 대한 관심으로 세상에 기여하고자 합니다. 보다 성장하는 기획자가 되기 위해 노력하고 있습니다.



이것은 사람이 컴퓨터와 소통할 때, '이렇게 말하면 컴퓨터가 이해하겠지?'라는 다소 불편한 단계를 생략할 수 있게 된 것으로 볼 수 있다. 컴퓨터의 처지를 고려하지 않고 사람의 표현 방식으로 이야기를 하더라도 컴퓨터가 사용자의 의도를 적척 파악하는 '대화형' 검색이 가능해지는 것이다. 카카오의 지식그래프에는 지금까지 약 12억 건 정도의 그래프가 연결되어 있으며, 그 수가 늘어날수록 이러한 대화형 검색이 용이해질 것으로 보인다.

**카카오미니에서의 지식그래프 활용**

[그림 2] 카카오미니의 정보 검색 흐름 및 지식그래프 시스템 구조



대화형 상호작용을 가능케 하는 지식그래프는 카카오의 AI 스피커인 카카오미니에도 적용되고 있다. 카카오미니는 음성으로 사용자와 상호작용이 일어나기에 지식그래프의 관계 기반, 의미 기반 검색의 장점을 활용하는 데 적합한 서비스라 생각된다.

예를 들어 카카오미니에게 “트와이스 리더의 나이가 몇이야?”라고 물었다고 가정해보자. 만약 지식그래프를 활용하지 않았다면 카카오미니는 해당 명령어를 인식하지 못했을 가능성이 크다. “트와이스 리더가 누구야?”, “지효 나이” 등을 묻는 단계를 거쳐야만 “지효의 나이는 만 21세예요”라는 답을 줄 수 있을 것이다. 하지만 현재 카카오미니는 지식그래프 기술을 적용하고 있기에 단 하나의 질문만으로도 “트와이스의 리더의 나이는 만 21세예요”라는 답을 할 수 있는 것이다. [그림 2]에서 ‘Cruise KG’에 해당되는 것이 바로 지식그래프로, 이는 트와이스 리더-지효-나이-만 21세 등 일련의 연결된 정보 그래프로 구성되어 있다.

### 지식그래프 적용의 어려움

이렇듯 지식그래프는 기술적으로는 훌륭하지만, 이를 카카오의 서비스에 적용하는 데는 나름의 어려움이 존재한다. 사용자들이 지식그래프를 활용하는 검색 방식에 익숙지 않기 때문이다. 예컨대 “아이유 동문인 연예인 알려줘”라고 검색을 하면 결과가 나오지 않을 것으로 예상한 사용자들이 기존의 검색 방식을 고수하고 있다는 것이다.

이와 같은 방식은 사용자 입장에서는 시행착오를 통해 터득한 최상의 검색 방식이기에 이를 바꾸기가 쉽지 않다. 혹시나 하는 마음에 입에서 나오는 말 그대로 검색했다가 원하는 결과를 얻지 못한 경험이 누구나 한 번씩은 있었을 것이다. 지식그래프라는 유용한 검색 도구가 있음에도 사용자들이 기존의 검색 방식에서 탈피하기 쉽지 않은 이유이다. 카카오를 비롯해 지식그래프를 활용하는 기업들이 사용자의 검색 방식을 어떻게 변화시켜야 할지 고민해야 할 필요가 있어 보인다.

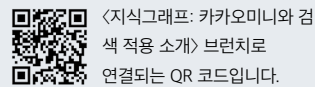
### 지식그래프에 대한 짧은 소감

해당 콘퍼런스에 참여한 후 지식그래프에 대한 몇 가지 생각들이 떠올랐다.

첫째, 지식그래프가 마치 사람의 인지구조와 유사하다는 느낌을 받았다. 스위스의 철학자이자 심리학자인 장 피아제(Jean Piaget)에 따르면 인간은 도식(schema)을 바탕으로 세상을 이해한다고 한다. 여기서 도식이란 인간이 세상을 바라보는 틀이며<sup>1</sup>, ‘날아가는 것=새’와 같이 일련의 지식들이 구조화된 것이라 볼 수 있다. 이에 지식그래프가 이러한 인간의 도식과 상당히 유사하다는 느낌을 받았다. 또한 뇌에서 하나의 뉴런은 보통 1만여 개의 다른 뉴런과 연결되어 네트워크를 이루는데<sup>2</sup>, 지식그래프가 이러한 뉴런의 체계와도 유사하다는 인상을 받았다. 기계를 통해 인간의 뇌를 모방하고자 했던 목표<sup>3</sup>가 현실화되고 있는 것이 아닐까.

둘째, 지식그래프는 단순한 검색을 넘어 시가 관계의 대상이 될 수 있도록 기여할 것이라는 생각이 들었다. 지식그래프 덕분에 사용자는 자신에게 가장 친숙한 표현을 사용하여 컴퓨터의 반응을 얻을 수 있게 되었다. ‘컴퓨터를 이해시키는 말을 한다’에서 ‘컴퓨터가 내 말을 이해한다’로 관계가 전환되는 것이다. 그리고 이렇게 전환된 관계로 인해 시는 단순히 편의성만을 제공하는 것을 넘어 ‘관계의 대상’이라는 가치 있는 역할을 할 것으로 생각한다.

\*1 참고 | 신중호(2011), 《교육심리학: 교육 실제를 보는 창》, 피어슨에듀케이션코리아  
\*2 참고 | 강봉균(2013), <뇌 이해하기>, 한국분자-세포생물학회  
\*3 참고 | 김대식(2016), 《김대식의 인간 vs 기계: 인공지능이란 무엇인가》 동아시야





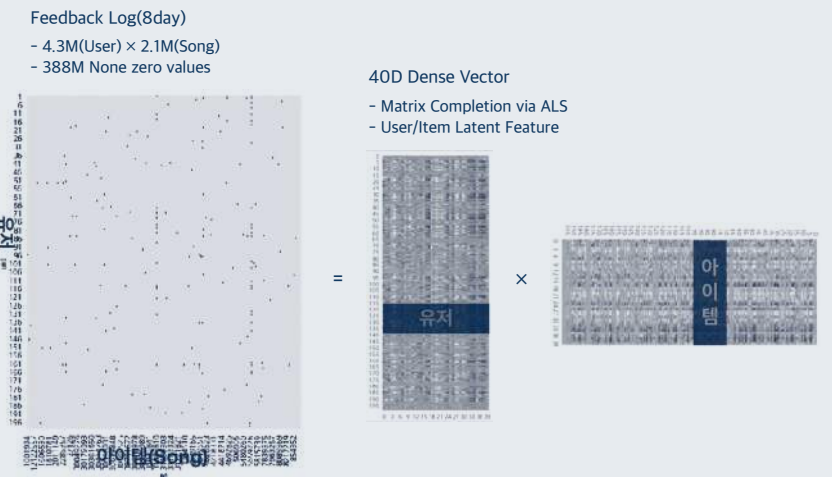
# 눈으로 듣는 음악 추천 시스템

이번 글은 음악 추천 시스템에서 임베딩(embedding)한 음원 벡터(song vector)를 장르와 인기도, 발매 연도의 관점에서 고찰해보는 발표 내용을 담았다. 실제 추천팀에서는 하루 1억 5000만 건의 스트리밍(streaming) 데이터, 5600만 개의 사용자 플레이리스트 데이터, 18만 개의 디제이(DJ) 플레이리스트 데이터를 협업 필터링(collaborative filtering, CF) 방식으로, 3000만 곡의 음원 및 메타 정보는 콘텐츠 기반 접근 방식(contents based filtering, CBF)을 통해 다양한 음원 벡터를 모델링하고 있다.

## 음악 추천 시스템에 담긴 추천 알고리즘: CF와 CBF

[그림 1] 협업 필터링 방법 중 대표적으로 쓰이는 행렬분해(matrix factorization) 방법

CF의 대표 기술인 Matrix Factorization



[그림 1]과 같이 협업 필터링 방식 중 성능이 뛰어나다고 알려진 행렬분해(matrix factorization, MF)를 이용해 유저(user)와 아이템(song)의 잠재 특징(latent feature)들을 원본보다 밀도 있는 특정 차원의 벡터들로 뽑아낼 수 있는데, 카카오에서는 40차원의 벡터를 사용한다. 또한 계산상의 이점을 얻기 위해 잘 알려진 ALS(alternating least square) 기법을 사용해 매트릭스 컴플리션(matrix completion)을 진행한다. 협업 필터링을 통해 얻는 음원 벡터는 여기서 바로 아이템에 대해 얻은 잠재 특징(latent feature) 벡터이고, 각 벡터들은 사용자들의 소비 특성에 기반해 유사한 속성을 가진 음원들의 벡터 간 거리를 가깝게 표현했을 것이라 기대된다.

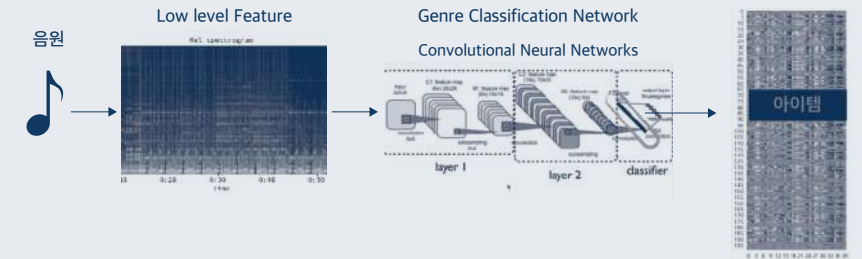
컨퍼런스 발표 | 최규민 pi.314@kakaocorp.com

글 | 김태훈 th.kim@snu.ac.kr 서울대학교 음악오디오연구실에서 음악을 비롯한 콘텐츠 추천 시스템을 연구하고 있습니다. 추천 시스템의 발전은 창작하는 사람과 향유하는 사람 사이의 피드백 선순환을 일으켜 양질의 콘텐츠 생산에 기여할 수 있다고 믿고, 기술로써 그에 조금이나마 보탬이 되기 위해 노력 중입니다.



[그림 2] 콘텐츠 기반 음원 벡터를 뽑는 과정

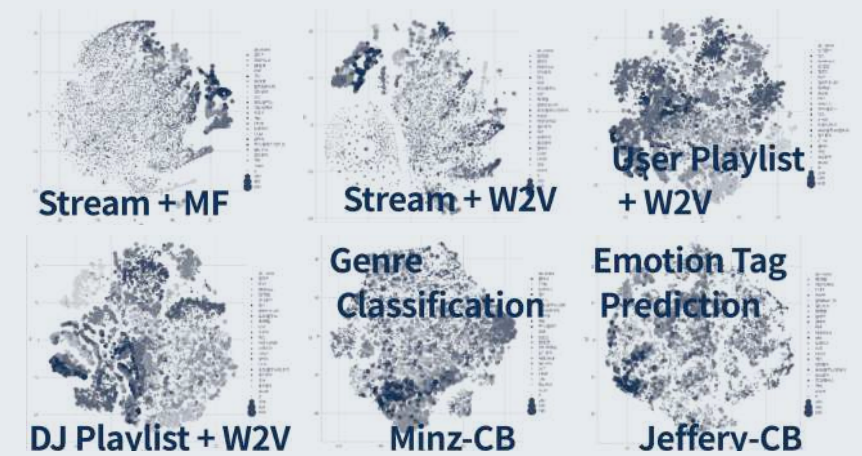
CBF: 장르 Classification Model



음원의 오디오 특성을 이용해 음원 벡터를 구성할 수도 있는데, 카카오에서는 음원의 멜-스펙트로그램(Mel-Spectrogram)과 같은 낮은 수준 특성(low-level feature)을 뽑아 입력으로 사용하고, 가지고 있는 메타 정보를 이용해 분류 태스크(장르 혹은 태그)를 진행하는 딥 뉴럴 네트워크(deep neural network, DNN)를 구성한다. 해당 네트워크로 학습을 시킨 후 특정 레이어(layer)를 가져오면 각 음원의 콘텐츠 기반 특성으로부터 클래스를 분류하는 데 사용된 벡터를 얻을 수 있다. 이것을 콘텐츠 기반 접근을 통해 얻은 음원 벡터로 사용하는 것이다(그림 2 참조).

음악 추천을 위해 사용하는 음원 벡터와 그 특징

[그림 3] 카카오 추천팀에서 사용하는 6가지 음원 벡터



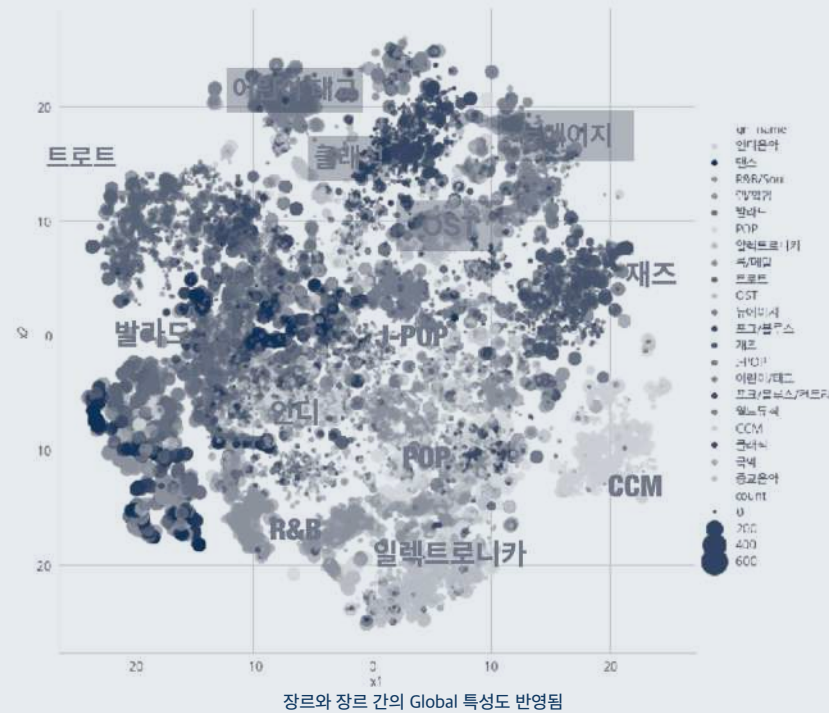
카카오는 크게 6가지의 음원 벡터를 가지고 있다. 각각은 [그림 3]에서 볼 수 있는 것처럼 스트리밍 데이터 + 행렬분해, 스트리밍 데이터 및 사용자가 직접 만든 플레이리스트를 Word2Vec(word to vector, W2V)으로 처리한 것, 전문 DJ의 플레이리스트를 Word2Vec으로 처리한 것, 그리고 장르 분류 모델을 이용한 콘텐츠 기반의 음원 벡터, 감정 태그를 분류하는 모델로부터 얻을 수 있는 음원 벡터이다. 각기 다른 데이터와 방법론을 거쳐 만들어진 음원 벡터이다 보니 특정 쿼리(query) 곡에 대한 유사곡 추천 결과 양상도 조금씩 다르게 나타난다.

음원 벡터를 통해 유사한 곡이 추천되는 과정을 눈으로 듣는 과정은 다음과 같다. 먼저 1만 6000개의 곡을 샘플링(가용 가능한 전체 곡에서 상위 1000곡+나머지 랜덤 샘플링 1만5000곡)한 뒤, 각 방법론에 의해 뽑힌 음원 벡터를 t-분포 확률적 임베딩(t-distributed Stochastic Neighbor Embedding, t-SNE)으로 차원을 낮춰 평면에 시각화한다. 거기에 각 음원의 메타 정보(장르, 발매 연도, 스트리밍 순위 등)를 매핑(mapping)해서 살펴보고, 멜론 안에서 각 곡이 소비되는 정도에 따라 4가지로 분류(light, medium, heavy, extreme)으로 분류했으며 뒤로 갈수록 대중적으로 많이 소비되는 곡을 뜻한다)해 별도의 그림을 그려보기도 하였다.

스트리밍 데이터 기반의 행렬분해(Stream + MF) 방법론을 통해 뽑아낸 음원 벡터의 특징은 헤비(heavy)한 정도의 인기를 가진 노래의 장르들은 잘 식별되고, 상위 1000곡에 속하는 노래들을 인기도로 구분된 분포를 보면 일정한 패턴(상위 1000곡끼리 모여 있는 등)이 나타나는 데 반해 노래가 발매된 연도의 경우 뚜렷한 패턴이 나타나지 않는다는 것이다.

[ 그림 4 ] Stream + W2V를 통해 뽑아낸 음원 벡터의 시각화 - 장르별 군집화 관찰

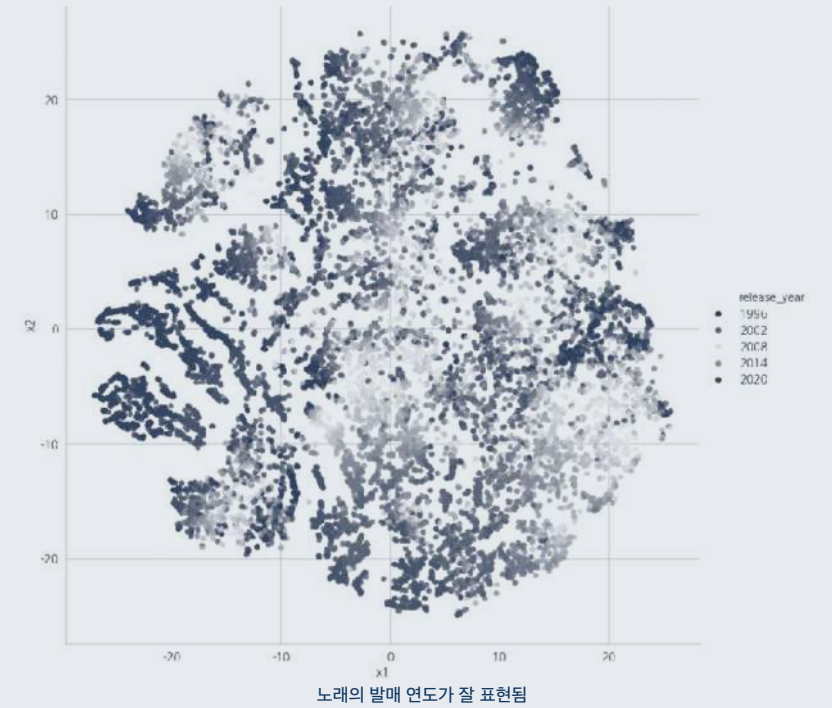
Stream + W2V: 노래가 장르별로 군집화가 더 잘됨



스트리밍 데이터를 가지고 Word2Vec 방법론을 통해 뽑아낸 음원 벡터(Stream + W2V)의 특징은 [그림 4]와 같이 장르별 군집화가 잘된다는 점이다. 이 방법은 행렬분해의 경우와 달리, 인기도에 큰 영향을 받지 않고 비슷한 장르는 동일 공간에 표현되는 특성을 보였고, 장르와 장르 간의 연관성도 어느 정도 반영된 모습을 보여주었다. 노래의 발매 연도 역시 'Stream + MF'보다는 상대적으로 분리가 잘되는 경향을 보였으나, 여전히 상위 1000곡의 노래들은 같은 순위 안의 노래끼리 모이는 특성이 존재하였다.

[ 그림 5 ] Playlist + W2V를 통해 뽑아낸 음원 벡터의 시각화 - 발매 연도 관찰

Playlist + W2V: 발매 연도의 분포



사용자들의 플레이리스트 데이터를 가지고 Word2Vec 방법론을 통해 뽑아낸 음원 벡터(Playlist + W2V)의 경우 [그림 5]와 같이 발매 연도의 군집화가 'Stream + W2V'보다 더 잘되어 있는 경향을 보였으며, 장르 식별의 수준이나 상위 1000곡끼리 모여 있는 정도는 'Stream + W2V'와 비슷한 수준을 보였다.

지금까지 살펴본 3개의 음원 벡터 간 특징을 비교해보면 다음과 같이 정리해볼 수 있다.

'Stream + MF', 'Stream + W2V', 'Playlist + W2V' 3개 음원 벡터 간의 비교(상대적으로 어떤 음원 벡터에서 해당 특성이 잘 나타나는지를 부등호로 표현)

공통: 상위 1000곡은 상위 1000곡끼리 군집화된다.

장르 군집화 정도는?  
Stream + MF < Stream + W2V = Playlist + W2V

Popularity 정도는?  
Stream + MF > Stream + W2V = Playlist + W2V

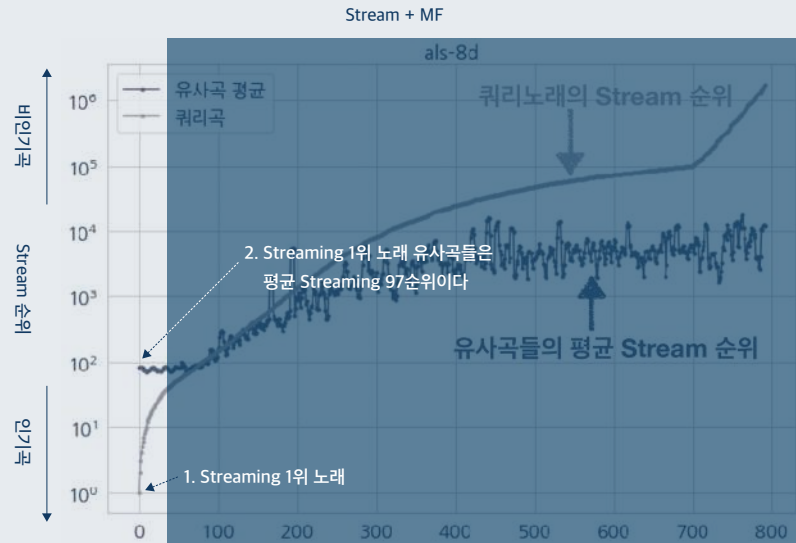
노래의 발매 연도 특성은?  
Stream + MF << Stream + W2V < Playlist + W2V

각 방법론을 통해 음원 벡터가 나오므로, 이것은 곧 벡터 간의 거리를 구하는 과정을 통해 쿼리 곡에 대해 유사곡 목록을 뽑아낼 수 있다는 것이다. 1000여 곡에 대해 유사곡 추천을 해보았을 때,

쿼리곡 장르와 유사곡으로 추천된 곡들의 장르가 같은 비율은 'Stream + MF'가 45.2%, 'Stream + W2V'가 60.1%, 'Playlist + W2V'가 59.9%로 조금씩 다른 비율을 보여주었다.

[ 그림 6 ] Stream + MF 음원 벡터로 유사곡을 추천한 경우 인기도 분포

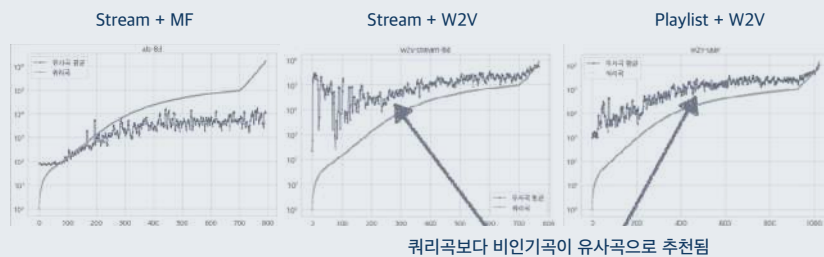
유사곡들의 Popularity 분포는?



또한 'Stream + MF' 음원 벡터의 경우는 쿼리곡이 스트리밍 순위로 몇 위의 곡이든 유사곡들의 평균 순위를 내보면 대부분 100위~1만 위 사이 곡들로 추천된다(그림 6) 참조).

[ 그림 7 ] 각 음원 벡터 방법으로 유사곡을 추천한 결과, 추천된 곡들의 평균 인기도 분포 비교

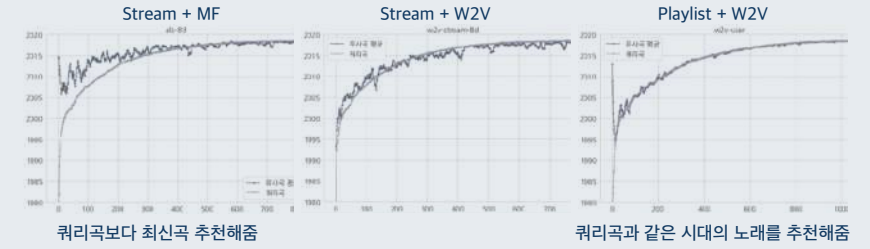
유사곡들의 Popularity 분포 Compare?



이에 반해 'Stream + W2V'와 'Playlist + W2V'의 경우에는 쿼리곡의 순위에 비해 유사곡으로 추천되는 곡들의 평균 순위가 더 낮은 경향을 보여주었다(그림 7) 참조).

[ 그림 8 ] 각 음원 벡터 방법으로 유사곡을 추천한 결과, 추천된 곡들의 평균 발매 연도 분포 비교

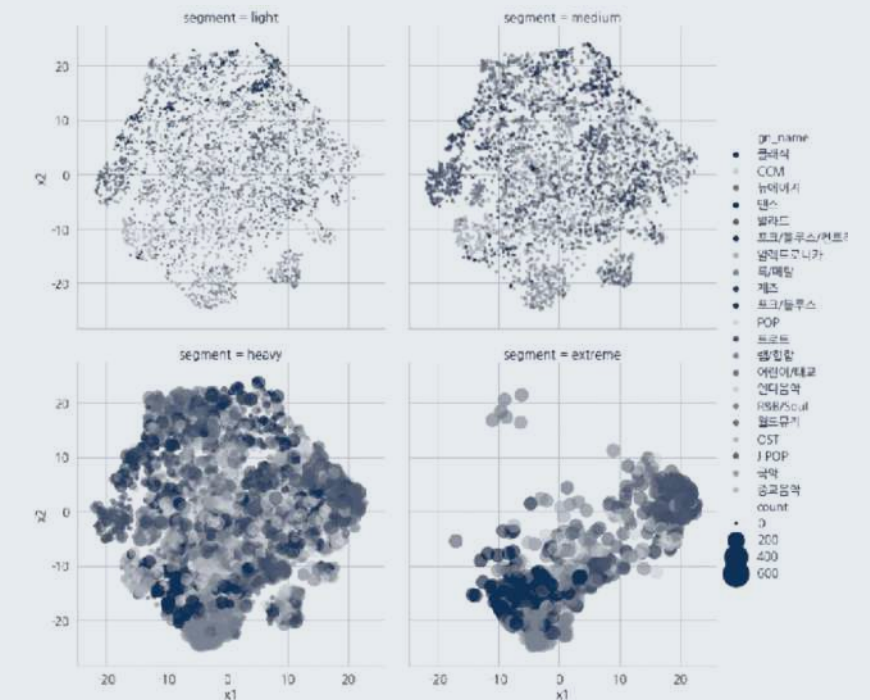
유사곡들의 발매 연도는?



또한 비슷한 방식으로 쿼리곡에 대해 유사곡으로 추천되는 곡들의 평균 발매 연도를 살펴본 경우, [그림 8]과 같이 'Stream + MF'는 쿼리곡보다 최신곡들이 추천되었다면, 'Playlist + W2V'는 쿼리곡과 같은 시대의 곡들이 유사곡으로 추천되었고, 'Stream + W2V'는 그 중간 정도 성격을 나타내었다.

[ 그림 9 ] 콘텐츠 기반 방법론(장르 분류 모델)으로 뽑아낸 음원 벡터의 시각화

Audio + Genre Classification DL



- Song Feature가 장르를 적절히 표현하는 것으로 보임
- 장르 간 글로벌한 특성까지 적절히 표현되어 보임
- Song Popularity와 상관없이 고르게 분포되어 있음

마지막으로 콘텐츠 기반 방법론으로 뽑아낸 음원 벡터의 경우 장르끼리 적절히 잘 군집화되는 특징을 보였으며, 접근 방법의 특색대로 인기도와 상관없이 고르게 분포되어 있는 모습을 보여 앞의 경우들과는 달리 상위 1000곡에 속하는 노래들도 음악적 특성에 의해 군집화되는 경향을 확인할 수 있었다(그림 9) 참조).

또한 장르별로 나누어 그림을 그려보면 OST, 인디음악 등 ‘음악적 특성으로 구분되기보다는 일종의 구분에 의한 장르’의 경우, 다른 장르들의 경우와는 달리 별도의 군집 특성 없이 전체적으로 흩어져서 분포되는 것을 확인할 수 있었으며, 발매 연도를 살펴봐도 딱히 상관성을 발견할 수는 없었다. 앞에서 살펴본 바와 같이 음원 벡터마다 다양한 특성을 갖고 있는데, 실제 음악 추천 환경(멜론, 카카오뮤직)에서 어떠한 음원 벡터가 좋은지는 내년 콘퍼런스 행사를 기약하며 세션이 마무리되었다.

‘눈으로 듣는 음악 추천 시스템’ 세션은 ‘음악’이라는 대상에 접근하는 방법에 따라 ‘유사하다’는 개념이 얼마나 다를 수 있는지를 보여주었다고 생각한다. 인기도에 따른 바이어스(popularity bias)를 피해 유사곡을 추천하려면 콘텐츠 기반으로 접근해볼 수 있다. 같은 스트리밍 데이터를 가지고도 음원 벡터를 만들어내는 방법에 따라 추천된 유사곡들의 평균 인기도나 발매 연도와 같은 특성이 달라졌다. 각 방법들의 장점을 청취 맥락이나 유저의 취향에 맞게 잘 살린다면 보다 다양한 유사곡 추천 성능을 보일 수 있을 것이다.



〈눈으로 듣는 음악 추천 시스템〉 브랜치로 연결되는 QR 코드입니다.

# 이미지로 이미지 검색하기

연예인들의 공항 패션은 항상 대중들로부터 많은 관심을 받는다. 편안함을 추구하며 일상적인 모습을 보여주는 사복 패션은 누구나 쉽게 따라 해볼 수 있는 스타일링으로 여겨지기 때문이다. 기사에 실린 사진 속 스타가 입을 옷에 대한 정보를 검색하려고 한다. 어떻게 검색할 수 있을까? '2018년 8월 18일, 설현, 공항 패션', 이런 식으로 연관된 키워드를 입력해서 검색해보니 다행히도 눈썰미 좋은 누군가가 상품의 브랜드를 소개한 글이 블로그에 올라와 있다. 하지만 텍스트로 이미지를 찾는 것은 한계가 있어 보인다. 어떤 경우에는 가능하지만 간혹 원하는 결과가 나오지 않을 때도 많아 이미지를 찾는 데 어려운 경우가 대부분이기 때문이다. 또한 텍스트로 검색해서 얻은 이미지들 중에서 원하는 결과를 찾는 데도 적지 않은 노력이 필요하다.

## 이미지 검색

이미지로 이미지를 검색하는 서비스는 쿼리(query)를 이미지로 받는다. 즉, 검색어를 입력하는 텍스트 기반의 단어 검색이 아닌 이미지를 이용하여 검색하는 방식이다. [그림 1]은 마이크로소프트(Microsoft)에서 제공하는 COCO(Common Objects in Context) 데이터 세트에 포함되어 있는 사진 중 하나이다. COCO는 디텍션(detection), 세그멘테이션(segmentation), 캡셔닝(captioning)을 위한 학습에 활용되는 데이터로, 아래 이미지는 다양한 동물들이 줄지어 있는 박물관 내부의 한 장면을 담고 있다. 사람마다 주목하는 대상이 달라 같은 그림이어도 이를 묘사하는 방식들이 다양하다. 그렇다면 과연 우리는 어떻게 원하는 이미지를 찾을 수 있을까?

[그림 1] 마이크로소프트의 COCO 데이터 세트(data set)\*1

id=55575



a long train of model animals on display to show how noah's arc was filled.  
a few people are standing near a large exhibit of animal statues.  
a large group of wild animals standing next to each other.  
a display of wild animals inside a building.  
a museum display of assorted animals like zebras and giraffes.

\*1 참고 | <http://cocodataset.org/#overview>

컨퍼런스 발표 | 이주영 michael.lee@kakaocorp.com

글 | 이가람 garam24@snu.ac.kr 서울대학교에서 수학과 통계학을 전공하고 최근 머신러닝, 딥러닝, 블록체인에 관심을 갖고 공부하며 그 접점을 찾으려 하고 있습니다. 궁금한 게 많은 문과적인 이과생이고 세상을 조금이라도 더 나은 곳으로 만드는 데 기여하길 바라며 노력하고 있습니다.

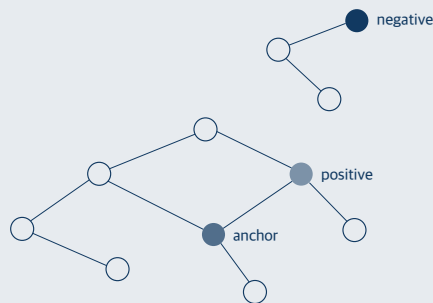
**방법 1) 텍스트로 검색한다**

문장 내에서 주어진 이미지를 가장 잘 표현하는 단어들을 추출한다. 문서를 검색하는 방식과 동일하게 추출된 단어를 사용해서 검색하면 해당 단어와 매핑(mapping)된 이미지를 결과로 얻을 수 있을 것이다. 하지만 이 경우 주어진 사진을 정확하게 묘사하는 문장이 필요하다는 문제점이 가지고 있다. SNS상에 별도의 내용 없이 사진만 올린 경우 검색이 불가능하다. 이를 위해서는 세상의 모든 대상을 인식하여 이미지를 분류하는 분류기가 필요하다. 또한 정제된 표현과 문장의 경우 사진과 직접 연관된 단어를 포함하고 있지만 단어와 사진이 매칭되지 않을 때는 학습에 어려움이 발생한다. 자신의 글을 더 많은 사람들에게 노출시키기 위한 목적으로 노이즈(noise)를 발생하는 단어를 사용했다면 이를 필터링(filtering)하는 과정이 추가적으로 필요하다.

**방법 2) 이미지로 검색하여 일정 거리 내에 있는 유사 이미지를 검색 결과로 제공한다**

각 이미지를 단어와 매핑하여 해당 단어와 일정한 거리에 있는 단어들에 대한 이미지를 검색 결과로 한 번에 노출시킨다. 이를 [그림 2]와 같이 방향성이 없는(undirected) 그래프로 표현할 수 있다. 각 노드(node)들은 이미지를 나타내며 에지(edge)는 각 노드가 연결된 관계를 표현한다. 사용자가 쿼리에 이미지를 입력하면 기준이 되는 노드인 앵커(anchor)로부터 특정 거리 내에 있는 노드들과 같은 그룹 안의 유사한 이미지가 검색 결과로 제공된다. 그래프 내의 노드 중에서 앵커로부터 연결되어 있는 노드는 포지티브(positive)로, 연결되어 있지 않은 노드는 네거티브(negative)로 라벨(label)을 붙인 후 샘플링(sampling)하여 이 세 노드의 관계를 통해 모델을 학습시킨다.

[ 그림 2 ] 이미지를 노드로 매핑한 네트워크 구조



모델은 앵커와 포지티브 노드 사이의 거리가 앵커와 네거티브 노드 사이의 거리보다 더 가까워지도록 학습된다. 이미지 데이터 자체가 매우 큰 경우 함수 F를 사용하여 정보를 압축한다. 또한 세상의 모든 이미지를 학습 데이터로 사용할 수 없기 때문에 새로운 데이터가 들어올 것을 감안하여 일종의 버퍼(buffer) 역할을 할 수 있게 마진(margin) m을 추가한다. 손실함수(loss function)로 트리플렛 함수(triplet function)를 사용하고 합성곱 신경망(convolutional neural network, CNN)을 통해 앵커와 연결된 포지티브 노드는 유사한 이미지로, 네거티브 노드는 다른 클래스의 이미지로 학습한다.

[ 그림 3 ] 학습을 위한 노드 간 거리 비교

$$\begin{aligned}
 & \text{Distance}(\text{Image}_{\text{Anchor}}, \text{Image}_{\text{Positive}}) < \text{Distance}(\text{Image}_{\text{Anchor}}, \text{Image}_{\text{Negative}}) \\
 & \quad \downarrow \\
 & \text{Distance}(F(\text{Image}_{\text{Anchor}}), F(\text{Image}_{\text{Positive}})) < \text{Distance}(F(\text{Image}_{\text{Anchor}}), F(\text{Image}_{\text{Negative}})) \\
 & \quad \downarrow \\
 & \text{Distance}(F(\text{Image}_{\text{Anchor}}), F(\text{Image}_{\text{Positive}})) + m < \text{Distance}(F(\text{Image}_{\text{Anchor}}), F(\text{Image}_{\text{Negative}})).
 \end{aligned}$$

**성능 비교**

다음 표는 스탠포드 온라인 프로덕트 데이터 세트(Stanford Online Product data set) 성능을 비교한 결과이다. 학습 데이터는 1만1300여 개의 클래스에 대해 6만 장의 이미지로 이루어져 있다. 성능 측정에 사용한 지표는 Recall@k로 ‘앵커로부터 k번째 가까운 노드까지의 거리 내에 포함되는 앵커의 이미지 개수’와 같은 클래스의 이미지를 갖는 노드 수의 비율을 나타낸다. k가 1인 경우 카카오(kakao)에서 개발한 모델의 성능이 논문에 발표된 가장 좋은 모델의 성능과 비슷하거나 약간 낮은 정도이고, k가 10 혹은 100일 때는 카카오 모델의 성능이 더 좋은 것을 확인하였다.

[ 표 1 ] 스탠포드 온라인 프로덕트 데이터 세트를 활용한 성능 비교<sup>\*2</sup>

Stanford Online Product data set [1]  
 120,053 images(train: valid=59,551:60,502)  
 22,634 classes(train: valid=11,318:11316)

	Recall@1	Recall@10	Recall@100
Triplet, Semi-Hard [1]	66.7	82.4	91.9
Lifted Struct [2]	62.5	80.8	91.9
Proxy NCA Loss [3]	73.7	-	-
Sampling Matters [4]	72.7	86.2	93.8
DAML [5]	68.4	83.5	92.3
kakao	73.3	86.2	93.9

\*2 참고 | [http://cvgl.stanford.edu/projects/lifted\\_struct](http://cvgl.stanford.edu/projects/lifted_struct)

### 서비스 응용

‘이미지로 이미지 검색하기’ 서비스는 카카오톡 대화방에서의 이미지 검색 기능(데모버전)에 활용되거나, 이미지를 통해 상품을 검색하는 쇼핑 서비스 그리고 유사 상품을 추천해주는 서비스 등에 응용되고 있다.

[ 그림 4 ] 유사 상품 추천 서비스

서비스 응용(1): 우린 비주얼이 되니까! 이미지 검색!



저 이미지를 뭐라고 검색해야 하지?  
애매했던 당신을 위해 이미지만 길게 누르면 검색이 되는 이미지 검색을 # 검색에서 지원합니다.



서비스 응용(2): 다음 쇼핑



<이미지로 이미지 검색하기>  
브랜치로 연결되는 QR 코드입  
니다.



# 딥러닝을 활용한 뉴스 메타 태깅

## 메타데이터(Meta-data)

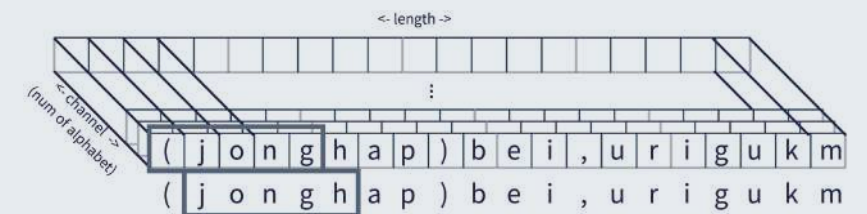
메타데이터는 뉴스 기사의 속성을 설명하기 위한 데이터로, 일반적인 정보들은 언론사가 카카오에 제공하기도 한다. 하지만 기사를 분석해야만 정의할 수 있는 정보들은 언론사에서 제공하지 않고, 제공할 수도 없는 경우가 있다. 카카오는 이런 경우에 해당되는 속성 태그를 생성하기 위해 직접 메타데이터를 만들어야 하는 번거로움을 겪는다. 사용자의 필요를 세밀하게 충족시키고 플랫폼에서 이를 활용하기 위해서는 뉴스의 메타데이터를 쉽게 파악할 필요가 있다. 카카오는 플랫폼을 통해 메타데이터를 소비할 수 있는 데이터의 흐름을 만들어 시스템화하고, 이를 '뉴스 메타 태깅 시스템'이라 이름 지었다.

## 기사의 형태 분류

카카오는 뉴스 기사를 서술된 내용을 기준으로 '사실 전달형 기사'와 '해설 묘사형 기사'로 나누어서 접근하였다. '사실 전달형 기사'는 "언제 누구에게 어떤 일이 있었다"를 육하원칙에 따라 서술한 기사이다. 반면 '해설 묘사형 기사'는 특정한 소재에 대해 맥락을 해설하거나, 현장을 상세하게 묘사하는 등의 심층 취재가 포함된 기사로 정의할 수 있다. 특히 '해설 묘사형 기사'는 '사실 전달형 기사'에 비해 상대적으로 포함된 정보가 다채롭거나 더 깊은 맥락을 제공하는 경우가 많아, 독자들이 깊이 읽어볼 만한 가치가 높은 특성을 가진다.

기사의 형태를 분류하기 위해 도입한 '합성곱 신경망(convolutional neural network, CNN)'은 6개 층의 '컨볼루션 레이어(convolutional layer)'와 3개 층의 '완전 연결 레이어(fully connected layer)'로 구성되어 있다. 텍스트의 형태소를 분석하거나 워드 임베딩(word embedding)을 하지 않고 그대로 분석할 수 있는 형태이면서 기사 안에서 사용되고 있는 괄호나 인용부호 등의 기호까지 의미 있게 반영하기 위하여 글자 단위로 학습을 진행하게 되었다. 또한 글자 단위에서 차원이 작은 상태로 처리하기 위해 한글과 한자를 모두 로마자 표기법(Romanization)으로 변환하여 학습 자료로 구성하였다.

[그림 1] 뉴스 기사 콘텐츠 분석을 위한 고려 사항들



- 형태소 분석, Word Embedding보다는 캐릭터(Character) 단위 처리
  - ▶ 기사 제목이나 본문의 특징적 기호 활용(예: [날씨], (종합), "나혼자")
- 한글, 한자 Romanization(로마자표기법 변환, 음역)
  - ▶ alphabet space 안에서 처리(예: 文정부, 靑 "소득성장가속")
- 문자와 단어의 순서 보존하여 기사의 문투 학습 ==> 1D Convolution

콘퍼런스 발표 | 김기도 olaf.kido@kakaocorp.com

글 | 김규형 civtale@gmail.com 고려대학교 대학원에서 전기전자공학을 공부하고 있으며 딥러닝 기반 자연어 처리에 관심이 많습니다. 성향 분석이나 이슈 검출을 자연어와 연결시켜 의미 있는 결과를 만들어내는 것을 좋아하고, 데이터가 많은 곳을 직접 찾아다니려는 성향이 강합니다. 명확하지 않은 분야의 분석 모델을 구축하는 것을 좋아합니다.

카카오는 뉴스 기사의 형태 분류 정보를 갖고 있지 않다. 딥러닝(deep learning)으로 학습을 하기 위해서는 정답 세트(set)가 필요하지만, 정답을 가지고 있지 않기 때문에 정답 세트를 어떻게 만들 것인지와 이에 대한 답을 찾아야 하는 문제가 발생한다. 카카오는 이러한 문제를 해결하기 위해 뉴스 기사를 파악한 결과값이 500자 내지 1200자면 사실 전달형 기사로 정의하고 3000자 이상의 기사는 해설 묘사형 기사로 정의하며, 뉴스 기사의 길이로 단위를 나누어 속성을 분류하였다. 잘못된 분류가 있다고 하더라도 오분류된 자료들에 특정한 패턴이 없다면 큰 단위에서 전체적인 학습을 하는 대세에는 영향을 주지 않는다. 따라서 자료의 양이 많은 트레이닝 세트(training set)는 자동으로 정답을 구성하고, 테스트 세트(test set)에 대해서는 전문가들이 직접 감수한 정답 세트로 검증하고 있다.

테스트에서 입력 차원의 크기는 중요한 정보들이 기사의 앞에 올 것이라는 가정을 세워서 제목과 본문을 합치고 로마자화하여 앞 600자로 구성하였다. 단어 차원의 크기는 알파벳 및 문장부호를 원핫 인코딩(one-hot encoding)으로 처리하여 그 크기를 정의하였다.

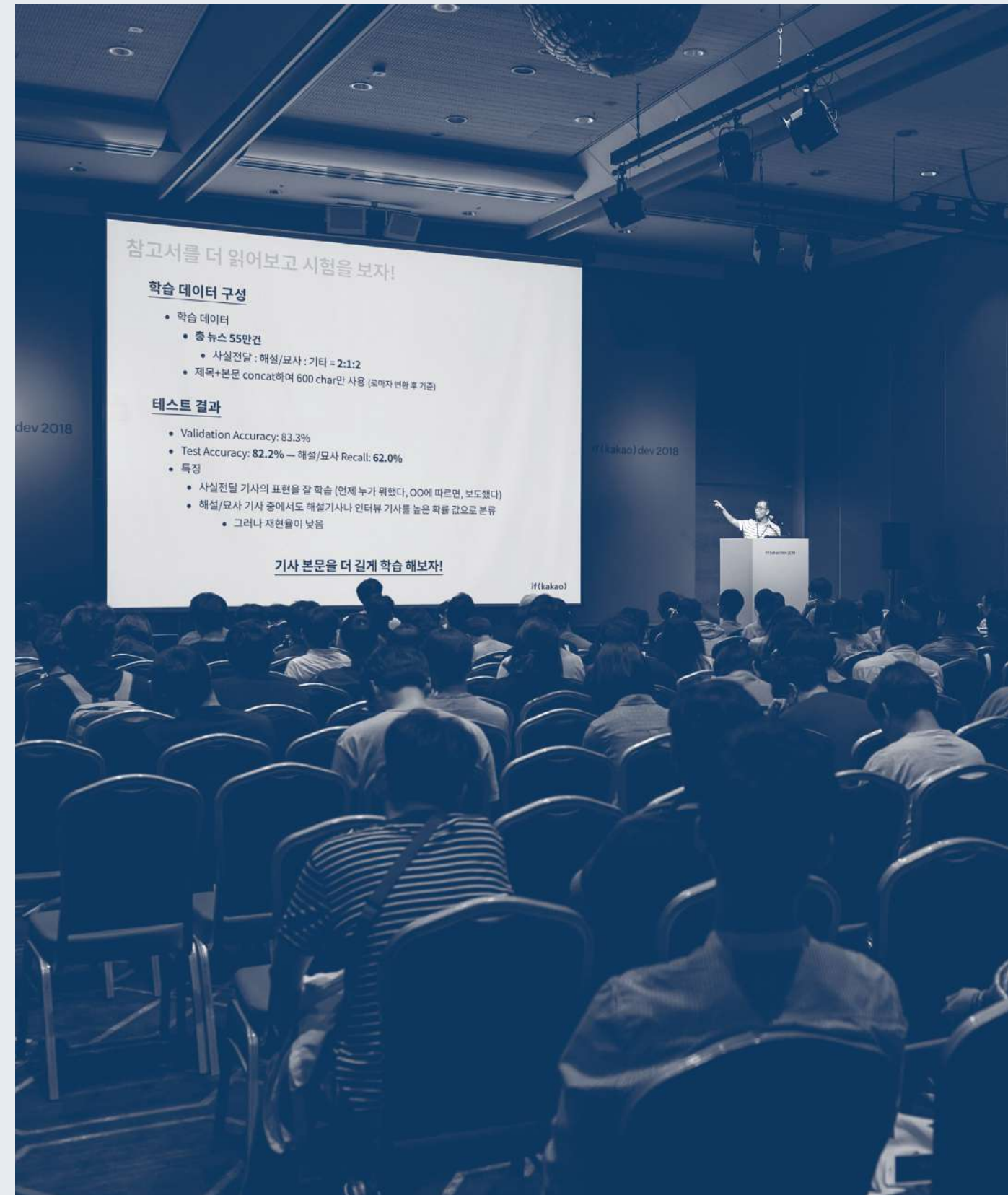
먼저 첫 번째 테스트는 사실, 해설, 기타 데이터 세트의 비율을 10:1:10으로 진행하였다. 정확도(accuracy)는 약 80%에 육박했지만, 해설형 기사의 데이터가 부족했기 때문에 해설 카테고리의 재현율(recall)은 43% 정도에 그쳤다. 데이터를 추가하고 2:1:2 비율의 데이터 세트를 구성하여 두 번째 테스트를 진행한 결과 정확도 82%, 재현율 62%의 수준을 보였다. 더 나아가 동일한 기사를 1014자를 사용하여 학습한 결과, 정확도가 81% 정도로 기존 대비 1% 떨어졌으나 해설 묘사형의 재현율은 75%로 상당히 향상되었다. 한편 연재나 기획 쪽 기사를 판별하는 능력은 향상되었으나, 오늘의 운세와 같은 종합 기사에서는 4000자가 넘어가는 경우도 있었기 때문에 학습 및 판별에서 혼동하는 양상이 발견되었다.

앙상블 모델(ensemble model)은 여러 개의 모델 결과를 조합해 더 좋은 결과를 구성하도록 하는 모델 구조를 말한다. 위의 테스트는 두 모델의 판정 결과로 나온 소프트맥스(softmax) 확률값을 더하여 카테고리를 분류하는 구조로 구성하였다. 전문가들은 위의 테스트 결과에 대해 상당히 좋은 평가를 내렸지만, 이와 비교할 대상이 없다는 것이 다소 아쉬움을 남긴다.

학습에서 원만한 분류가 이루어지지 못하는 예로 '외국어를 공부하는 코너 기사'를 들 수 있다. 해당 예에서는 외국어 문자가 삽입되어 있는 기사들을 적절히 판단할 수 있는 데이터가 부족하여 정확한 판정을 내릴 수 없는 문제가 나타난다. 이에 해당 데이터 자료에 대한 학습을 진행하여 결과물을 개선할 수 있는 방법이 필요하다.

### 뉴스 메타 태깅 시스템

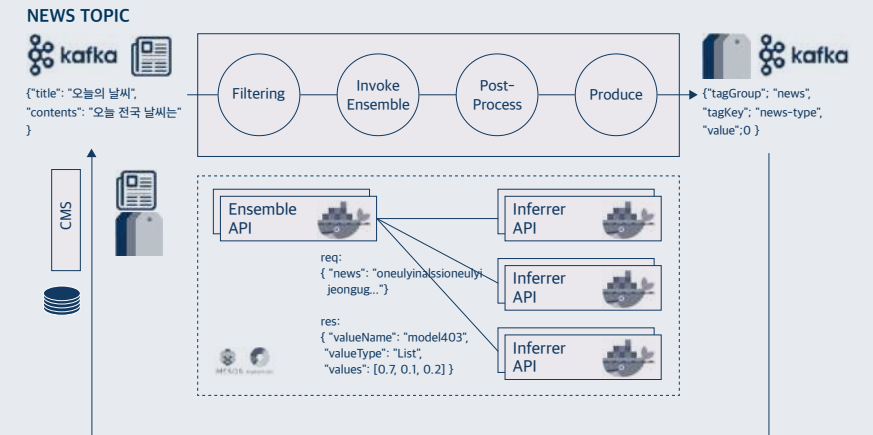
카카오에서는 태깅 시스템을 개발하면서 세 가지 설계 목표를 제시하였다. 첫째, 기사가 들어오는 대로 메타데이터를 빠르게 생성해야 하는 특성상 실시간(real-time)으로 작업이 진행되어야 한다. 둘째, 작업의 의존관계(dependency)가 없는 경우 여러 작업자들이 동시에 시스템에 접근해 메타데이터를 생성할 수 있도록 해야 한다. 이것을 비동기적 작업이라고 하며, 특정한 작업에서 문제가 발생했을 경우에도 나머지 작업 전체가 영향을 받지 않도록 하여 안정성을 높일 수 있다.





마지막으로 어떤 알고리즘이나 프레임워크(framework)를 사용하더라도 시스템이 작동하도록 설계해야 한다. 뉴스 본문을 분석하기 위해 사용하는 기계학습 및 딥러닝 기법에 제약을 두지 않기 위해서다.

[ 그림 2 ] 기사 형태 분류 시스템



카카오 뉴스의 태그 처리 시스템은 [그림 2]와 같은 절차를 거쳐 관리된다. 뉴스 데이터를 관리하는 카카오 CMS 플랫폼에서는 실시간으로 들어오는 뉴스를 뉴스 토픽(news topic)으로 전송한다. 뉴스 토픽에서 기사를 꺼내어 메타 태그를 생성할지 판단하고 태깅 서비스를 호출하는 일은 데이터 흐름 제어부가 담당한다. 생성하는 메타데이터의 종류가 많아질 수록 호출해야 할 태깅 서비스가 늘어나고 데이터 흐름 제어의 역할이 반복되므로 이 부분은 아파치 나이파이(Apache NiFi)를 사용하여 구현하였다. GUI 기반 환경에서 데이터 흐름을 구현할 수 있기 때문에 개발 시간을 단축할 수 있고, 데이터의 흐름을 메시지 단위로 추적하기 용이하기 때문이다.

태깅 서비스는 앙상블(ensemble) API와 인퍼러(inferer) API로 이루어져 있다. 앙상블 API는 인퍼러 API들의 결과 값을 종합하여 뉴스 기사에 대한 메타 태그값을 결정한다. 앞서 설명한 기사의 형태 분류의 예라면 기사가 사실 전달형 기사인지 해설 묘사형 기사인지가 메타 태그의 값이 된다. 이렇게 결정된 메타 태그는 태그 토픽(tags topic)으로 전송한다. 여기서 메타 태깅 과정을 끝낼 것 같지만 한가지 과정이 더 남아있다. 카카오 CMS 플랫폼은 태그 토픽에 저장된 메타 태그와 뉴스 본문 데이터를 병합하여 다시 뉴스 토픽으로 발행시킨다. 이와 같이 태그 데이터의 흐름을 구성하게 되면, 뉴스 토픽을 소비(consume)하는 것만으로도 뉴스 콘텐츠와 부가 생성된 메타 태그를 한 번에 받아들 수 있게 되어 소비자(consumer)들이 데이터를 여러 번 조회하지 않게 할 수 있다.

태깅 서비스를 구현하기 위해서 스파크 스트리밍(Spark Streaming)이나 스톰(Storm)같은 실시간 데이터 처리 프레임워크를 검토하기도 하였지만 특정 기술 세트에 대한 종속을 피하기 위해 필요한 역할에 따라 각기 독립적인 웹 API를 개발하였다. 개발된 API는 앙상블 API와 인퍼러 API로 구분되며, 그 설명은 다음과 같다.

### 앙상블 API(ensemble API)

앙상블 API는 말 그대로 학습한 모델들이 내어놓는 추론 값들을 종합하여 최종 결과 값을 반환하는 API이다. 앞서 설명한 기사의 형태 분류의 예라면 뉴스 기사를 로마자로 변환한 뒤, 600자 기준 판별 모델의 추론 값과 1014자 기준 판별 모델의 추론 값을 합하여 최종 결과 값을 내게 된다. 또한 인퍼러 API의 게이트웨이(gateway)로서 인증이나 로깅과 같은 공통 역할을 수행하기도 한다.

### 인퍼러 API(inferrer API)

인퍼러 API는 학습한 모델을 사용하여 데이터 추론을 하는 API로, 모델의 개수만큼 API를 개발한다. 앞서 설명한 기사의 형태 분류에서는 600자 기준 판별 모델과 1014자 기준 판별 모델이 각각의 인퍼러 API이다. 인퍼러 API는 최대한 단순하게 구현하는 것을 지향하여 다른 기능 없이 오직 모델의 추론 기능만을 구현한다. 모델이 재학습 되는 경우가 있으므로 버전 관리가 중요한데, 학습한 모델과 코드를 통합하여 관리하기 위해 도커라이징(dockerizing)을 하여 배포하고 버전을 관리한다.

뉴스에는 텍스트뿐 아니라 이미지, 차트, 동영상 등이 포함되어 있기 때문에, 향후 카카오톡에서는 텍스트뿐 아니라 이러한 멀티미디어 자료들을 분석할 수 있는 시스템을 개발할 예정이라고 한다.



〈딥러닝을 활용한 뉴스 메타  
태깅〉 브런치로 연결되는 QR  
코드입니다.

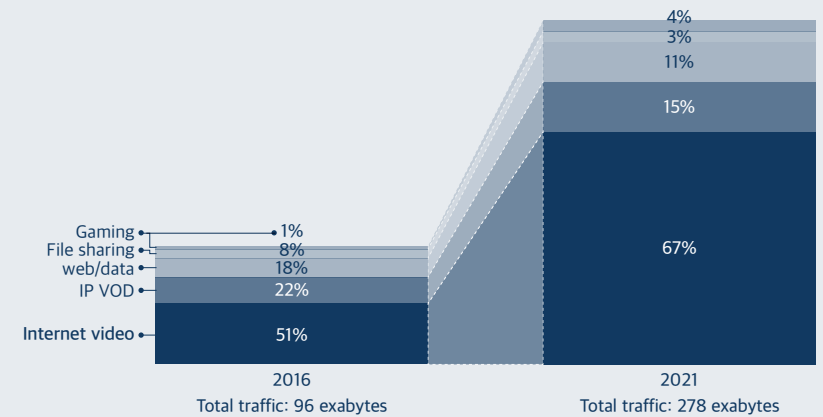
# 딥러닝을 이용한 실시간 인코딩 효율 최적화

동영상 분야에서 품질은 유지하면서 용량을 줄일 수 있는 방법을 딥러닝으로 개발한 내용을 다루었다. 단순히 동영상의 이미지를 학습시키는 것이 아니라, 인코딩할 때 동영상의 특성이 미치는 영향을 바탕으로 그 특성을 학습시키는 방법을 채택했다. 다시 말해 동영상을 인코딩할 때 알고리즘에서 사용되는 특징들을 분석해서 효율성을 높이는 것이다.

## 비트레이트(bitrate) 배분과 트래픽(traffic) 관리

하루에도 수많은 종류의 영상이 송출되고 있다. 카카오 TV 라이브 플랫폼도 그 영상을 받아서 인코딩하고, 스트리밍 서비스를 제공한다. 동영상 서비스를 제공할 때 가장 고민되는 부분은 무엇일까? 바로 트래픽이다. 트래픽은 비용과 직결되기 때문에 고품질 동영상을 무한정 서비스할 수 없으므로 트래픽과 화질을 고려하여 효율적으로 관리해야 한다. 전 세계적으로 동영상 트래픽이 폭발적으로 증가하는 추세이기 때문에 트래픽 관리는 더욱 중요한 문제가 되고 있다.

[그림 1] 글로벌 동영상 트래픽 증가 예측<sup>\*1</sup>



이 문제를 해결하기 위해 카카오에서는 먼저 압축 효율이 가장 좋은 최신 인코더를 도입해서 쓰는 접근 방식을 생각해보았다고 한다. 하지만 이 코덱(codec)을 적용하기 위해서는 현재의 인프라 시스템을 교체해야 하는데, 그러기에는 리소스 비용이 많이 필요하기도 하고, 실제로 적용해보면 압축 효율성이 예상보다 떨어진다. 그래서 카카오에서는 현재의 인프라로 트래픽 효율을 최적화하기 위해서 이 프로젝트를 시작하게 되었다.

동영상이 업로드되거나 라이브로 송출되면 가장 우선적으로 하는 일이 비디오 인코딩이다. 목적은 화질을 좋게 하면서 용량은 줄이는 것이다. 여러 가지 다양한 요소가 있겠지만, 직접적으로 맞닿아 있는 문제는 이것을 위해 실질적으로 비트레이트를 얼마나 할당하느냐의 문제이다. 사실 그동안의 경험으로 해상도별로 적정 비트레이트가 어느 정도 정해져 있다. 대부분의 서비스 업체에서 이런 방식으로 인코딩을 하고 있지만 이런 인코딩 방식은 효율적이지 않다. 움직임이 없는 동영상은 움직임이 많은 동영상에 비해 적은 비트레이트로 동일한 화질을 나타낼 수 있기 때문이다. 즉, 영상마다 적절한 비트레이트를 찾으면 트래픽을 효과적으로 사용할 수 있다는

콘퍼런스 발표 | 어욱형 jaden.y@kakaocorp.com

글 | 정소영 songr64@naver.com Swarm Intelligence를 이용한 Global Optimization Algorithm을 개발했고, 현재는 협업 필터링(Collaborative Filtering)을 바탕으로 추천시스템에서 Similarity Measure를 개발하는 연구를 진행 중입니다. 또 기계학습을 이용해 편미분 방정식을 푸는 연구를 진행하고 있습니다. 대학원 과정 동안 응용수학을 전공했고, 학부 때 조선공학을 전공했기 때문에 공학 기반의 수학 문제 해결에 관심이 많습니다.

\*1 참고 | Cisco



얘기가 된다. 그렇다면 적절한 비트레이트를 어떻게 찾을 수 있을까?

[그림 2] 구간별 동영상의 특성을 반영한 비트레이트 설정



단순하게 모든 필요 비트레이트와 해상도의 조합으로 인코딩을 해보면 된다. 화질 지수와 사이즈가 적절한 것을 찾으면 되는데, 이것은 많은 리소스가 필요해 현실적으로는 불가능하다. 그래서 카카오에서는 영상을 카테고리별로 분류하여 테스트했다. 애니메이션, 스포츠, 게임, 토크쇼를 분리해놓고 비트레이트를 다르게 설정해보았다. 하지만 테스트를 진행해보니 같은 영상이라도 구간별로 필요한 비트레이트가 다르다는 것을 알 수 있었다. 그래서 영상의 일정 구간마다 비트레이트에 영향을 줄 수 있는 요소를 분석하며 동영상을 더 세밀하게 분석했다.

인코딩은 어떤 패턴이 있는 것이 아니라 모든 요소마다 연관이 있기 때문에 수학적으로 풀어내기가 어렵고, 모호한 부분이 존재한다. 그래서 영상의 특성과 최적 파라미터(parameter)와의 관계를 딥러닝으로 분석해보기로 했다. 동영상 인코딩의 효율 최적화는 넷플릭스(Netflix), 유튜브(Youtube) 등에서도 적용하고 있는 방식이다. 넷플릭스는 3년 동안의 연구개발 끝에 타이틀 단위 인코딩(per-title encoding)을 통해 전체 트래픽의 30%를 감소시켰다. 유튜브의 경우 인공지능경망을 이용한 퀄리티 타겟(quality targets)을 개발 중이다. 이러한 시도들은 동영상에 고정적 비트레이트를 배분하는 것이 아니라 특징에 따라 유동적인 비트레이트를 배분하겠다는 시도이다.

**인코딩 효율 최적화를 위한 준비**

카카오에서는 어떤 패턴을 갖는 비디오의 특성들에 대해서 적절한 비트레이트의 설정값을 딥러닝으로 학습시키려 한다. 지도학습(supervised learning) 중 분류(classification) 방식이다. 비디오 특성 분석 데이터와 인코딩 파라미터가 입력(input) 데이터가 되고, 딥러닝 학습 시스템을 구동하여 비디오 특성 분류 프로그램이 출력(output)으로 나온다.

비디오 특성 분석 데이터란 영상 내 비트레이트에 영향을 주는 요소이다. 영상의 구조 단위는 크게 시퀀스(sequence), 그 안에 프레임(frame), 그 안에 다시 슬라이스(slice) 레벨로 나눌 수 있다. 슬라이스 내부는 매크로블록(Macro Block)으로 이루어져 있다. 매크로블록은 비디오 인코딩과 디코딩 연산의 최소 단위이며 주로 16x16픽셀이다. 경우에 따라서는 8x8, 4x4픽셀 등으로 더 작게 나누는 인코딩 프로세스가 수행되기도 한다.



[ 그림 3 ] 동영상 구조 단위 레벨 구분



매크로블록은 크게 세 가지 타입이 있다. 인트라 매크로블록(Intra Macro Block)은 하나의 프레임 내에서만 연관 관계를 찾아서 인코딩할 때 사용된다. 인터 매크로블록(Inter Macro Block)은 프레임 간 참조 관계를 통해서 인코딩을 할 때 사용된다. 스킵 매크로블록(Skip Macro Block)은 인터 매크로블록의 일종인데, 이전 프레임으로부터 비슷한 부분을 찾아서 데이터를 가져오는 점에서 비슷하지만, 현재의 프레임과 상관없이 이전의 프레임만을 가져와서 프레임을 구성한다는 점이 다르다. 인터 매크로블록의 경우 현재의 프레임과 이전 프레임의 상태를 분석해서 합친다.

[ 그림 4 ] 영상 속 하나의 프레임을 구성하는 요소



Type: I ( ), P ( ), Skip ( )

[그림 4]를 보면 하나의 프레임을 구성하는 요소를 확인할 수 있다. I, P, Skip(공백)으로 나누어져 있다. I와 P의 비율은 높고 Skip의 비율이 낮은 프레임이 있지만 I와 P의 비율은 낮고, Skip의 비율이 높은 프레임도 있다. 이렇게 다른 특성을 가진 동영상은 비트레이트 배분을 달리해야 한다.

비트레이트 배분에 영향을 주는 또 다른 요소가 있다. 움직임을 나타내는 요소, 모션 벡터(Motion Vector)이다. 현재 인코딩, 디코딩하려는 매크로블록이 이전 프레임의 어느 부분과 가장 일치하는지를 나타내는 요소로, 프레임 간 움직임을 추정할 수 있는 요소다. 다음 [그림 5]의 화살표가 바로 모션 벡터다. 모션 벡터량은 당연히 비트레이트 배분에 영향을 주게 된다.

[ 그림 5 ] 움직임을 나타내는 요소(Motion Vector)



최종적으로 영상을 분석할 때 움직임에 대한 정도를 나타내는 수치(움직임의 양, 강도), 프레임을 구성하는 블록의 타입별 분포, 프레임을 구성하는 블록의 파티션(partition) 사이즈 분포, 세그먼트 내의 복잡도 수치, 비트(bit) 배분율, 프레임 레이트(frame rate), 해상도, 그 외 영상과 관련된 전반적인 수치를 더하여 20여 가지 학습 데이터로 비디오 특성 분석 데이터를 구성했다.

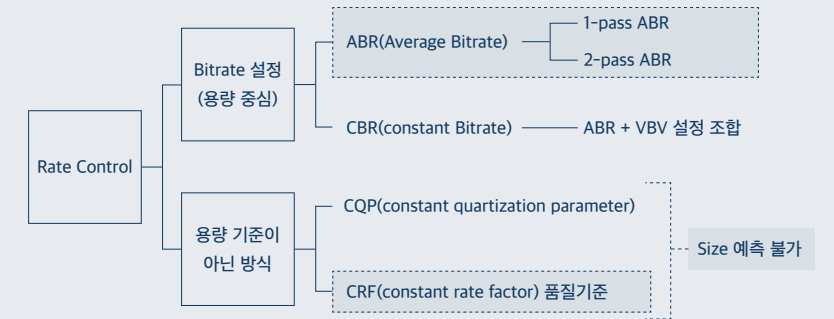
다음 인코딩 파라미터를 살펴보자. 파라미터는 두 가지로 구성했는데, CRF(Constant Rate Factor)와 Max rate를 택했다. CRF는 레이트 제어(rate control)의 한 종류이고, Max rate는 최대 비트레이트를 넘지 않도록 제한하는 수치이다.

레이트 제어는 용량을 고려하는지, 혹은 용량을 고려하지 않고 품질을 고려하는지를 기준으로 크게 두 가지 방식으로 나뉜다. 용량을 고려하는 가장 대표적인 방법으로는 ABR(Average Bitrate, 평균 비트레이트)이 가장 많이 쓰인다. ABR의 경우 미리 정해진 평균값을 중심으로 설정되는 비트레이트를 맞추는 방식이다. 레이트 제어 동작은 기본적으로 복잡도에 따라서 비트레이트를 배분하는데, ABR의 경우 복잡도를 미리 알 수 없는 것이 단점이다. 초기에 일정 비트레이트 배분을 해야 하지만 복잡도에 따라서 적극적으로 대응을 못 한다. 그래서 ABR 방식으로 2번 인코딩하는 2-Pass ABR도 존재하지만 충분히 효율적이지 못하다.

반면 CRF 방식은 용량의 제한 없이 품질을 맞추기 위해서 인코딩하는 방식이다. 현재 프레임의 복잡도에 따라서 비트 배분을 한다. 정해진 퀄리티 레벨(quality level)을 기반으로 인코딩을 하고 화질의 불균형도 발생하지 않는다. CRF는 덜 복잡한 부분은 비트 할당을 적게 하기 때문에 낭비되는 비트가 적을 수 있다.

하지만 CRF 대신 ABR이 많이 쓰이는 가장 큰 이유는 CRF의 인코딩 결과 사이즈를 예측할 수 없기 때문이다. 시스템을 운영하면서 트래픽을 예측할 수 없다는 것은 정말 치명적인 단점이다. 하지만 이러한 단점은 실험 데이터를 모아서 파악하면 극복할 수 있다. CRF 레벨은 0에서 51까지의 수치로 나뉘며, 레벨별로 비트레이트 설정 수치가 어느 정도 정해져 있다. 문제는 영상마다 CRF 레벨이 다른데, 복잡도가 비슷한 영상은 CRF 레벨과 비트레이트 곡선이 상당히 유사한 부분이 많다는 것이다. 카카오에서는 구간별로 동적인 비트레이트 설정을 할당하기 위해서 CRF를 선택했다. 한편 Max Rate의 경우 순간적으로 비트 배분이 높아지는 경우가 발생할 때를 대비하여 제한해야 하고, 이는 화질, 버퍼링 등에 상당한 영향을 미친다.

[ 그림 6 ] 레이트 제어(Rate Control) 분류 및 특성



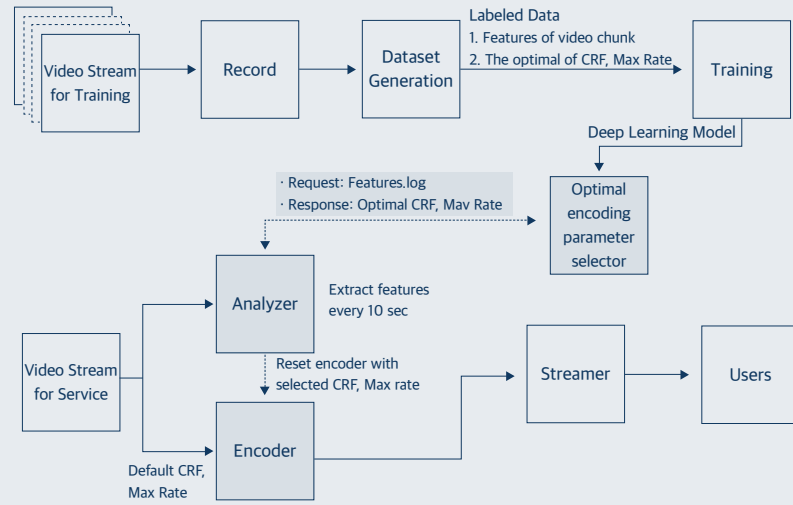
**인코딩 효율 최적화의 구현**

인코딩을 효율적으로 최적화하는 방법은 다음과 같다. 영상을 구간별로 잘라서 모든 CRF 레벨과 모든 Max rate 수치의 조합으로 인코딩을 실시한다. 그리고 품질 측정을 한다. 적절한 화질을 유지하면서 사이즈를 가장 작게 할 수 있는 파라미터를 고른다. 영상의 특징과 효율적인 파라미터를 매칭하고, 딥러닝을 이용하여 학습을 시킨다.

학습 결과로는 세그먼트 기준 총 30여만 개의 학습 데이터 세트를 생성해서 실험 정확도(test accuracy)를 계산했을 때 73% 수치가 나왔다. 100번 중에 73번은 학습시킨 의도대로 판별이 되었다는 뜻이다. 나머지 27번도 다른 값으로 판별되기는 하였지만, 원했던 값과 근사한 답을 주었고, 실제 서비스에 적용해도 무방한 수준이었다. 물론 알고리즘에 약간의 오류가 있어서 전혀 틀린 답을 뱉어낼 때도 있었지만 전체적으로 문제는 없는 수준이었다. 또 이전 설정값과 너무 큰 차이가 나면 이전 설정값을 사용해서 시스템의 안정성을 높였다. 이러한 방식을 적용하면 전체 트래픽의 약 20%를 줄일 수 있을 것으로 결과가 나왔다.



[ 그림 7 ] 시스템 아키텍처(System Architecture)




전체 시스템의 주요 아키텍처(architecture)는 다음과 같다(그림 7) 참고). 전체 시스템 아키텍처의 위쪽 레이어(layer)는 학습 시스템이고, 밑의 레이어는 기존에 갖고 있던 시스템이다. 애널리저(analyzer)와 최적의 인코딩 파라미터 셀렉터(optimal encoding parameter selector)를 붙여서 시스템을 구성하였다. 애널리저는 10초 간격으로 특징(feature)을 계속 뽑아내어(10초로 고정되어 있지 않음) 최적화된 파라미터 셀렉터에 파라미터를 질문한다. 그리고 파라미터를 받으면 해당 비트레이트의 배분에 맞게 인코딩을 한다.

이 시스템의 주요 목표는 데이터의 낭비를 줄이면서 퀄리티를 높여 데이터를 더 쓸 수 있게 하는 것이다. 고화질 서비스를 제공해야 할 때 비트레이트를 더 배분할 수 있도록 하고, 트래픽을 줄이면서 버퍼링 등의 문제를 예방해 스트리밍 서비스의 품질을 높이기 위해 노력했다. 또 원본 영상의 코덱 종류와 무관하게 동작하는 것을 목표로 했고, 지금은 H264로만 가정이 되어 시스템이 구성되어 있지만, 타깃 영상의 코덱 종류를 다양하게 지원하는 것도 염두에 두었다. 향후 H265 등의 고효율 압축 코덱도 서비스하려고 계획 중이다.

실제 일부 서비스에 적용해본 결과 화질은 -0.64% 정도로 눈으로 확인할 수 없을 만큼의 감소는 있었지만, 전체 용량은 약 20% 이상 절감 되었다. 조금 더 정적인 동영상에 적용을 해보니 화질은 겨우 1% 정도 나빠진 반면 용량은 거의 반으로 줄일 수 있었다.

이 프로젝트는 비디오의 특성과 비트레이트/화질과의 연관성을 입증하고, 비디오 특성을 분석한 데이터를 축적했다는 데 의의가 있다. 앞으로 정확도를 높이기 위해 더 많은 데이터를 확보해 분석할 수 있는 데이터를 추가하고, 모바일과 TV, 컴퓨터 등 기기별 최적화도 고려하며 좀 더 유연한 구조로 변경을 고려할 수 있도록 할 예정이다. 현재 카카오는 서비스 적용을 준비 중이며 연내 시스템 연동을 완료하는 것을 목표로 하고 있다.

 <딥러닝을 이용한 실시간 인코딩 효율 최적화> 브런치로 연결되는 QR 코드입니다.

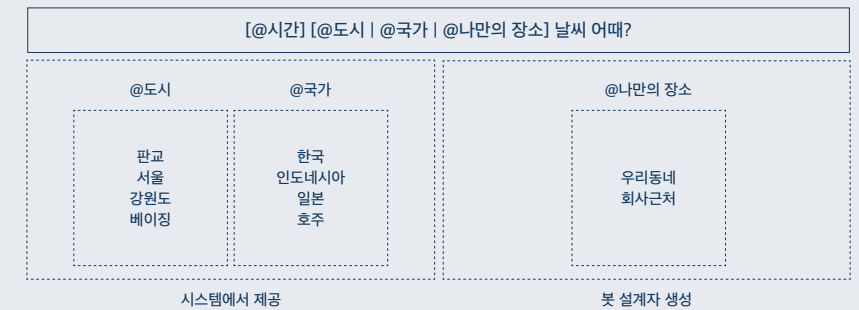
# 카카오 봇 플랫폼 소개

이번 세션에서는 봇 플랫폼(bot platform)을 활용한 봇 개발 방법과 기존 터치 기반 인터페이스가 아닌 자연어 처리(natural language processing, NLP) 방식의 인터페이스를 플랫폼에 어떻게 접목시킬 수 있는지, 그리고 다른 서비스와 어떻게 상호작용할 수 있는지를 소개하였다.

## 봇 만들기(Building a Bot)

봇(bot)이란 무엇일까? 봇은 AI 기술을 활용하여 질문에 대답하는 등 사용자와 상호작용하는 프로그램이다. 현재는 주로 챗봇(chatbot)의 형태로 고객센터 등에서 응대 역할을 수행하거나 개인 사용자의 비서처럼 활용되고 있다. 초창기 봇 개발 당시에는 각각의 봇을 하나하나 개발해야 하는 번거로움이 있었지만, 개발자들은 곧 이러한 봇들에 공통적인 구조가 있다는 것을 깨달았다. 이에 봇의 공통적인 부분을 묶어서 반복 작업 없이 봇을 만들 수 있는 봇 플랫폼이 개발되었다. 여기서는 봇의 공통적인 특성을 위주로 카카오 봇 플랫폼상에서 봇이 어떻게 개발되는지 살펴보자.

### [ 그림 1 ] 입력된 발화의 엔티티(entity) 분류



봇은 크게 사용자로부터 전달받은 메시지와 그에 상호작용하는 엔진, 그리고 엔진에 사용되는 데이터들로 구성되어 있다. 먼저 데이터 부분을 살펴보자. 각각의 원시 자료는 자연어 문장으로 이루어져 있으며, 각 문장에는 일정한 패턴이 존재한다. 이러한 패턴을 인식하고 학습하여 그 패턴에 따라 원시 자료를 분류하는 기능을 '패턴 분류기(pattern classifier)'라고 한다.

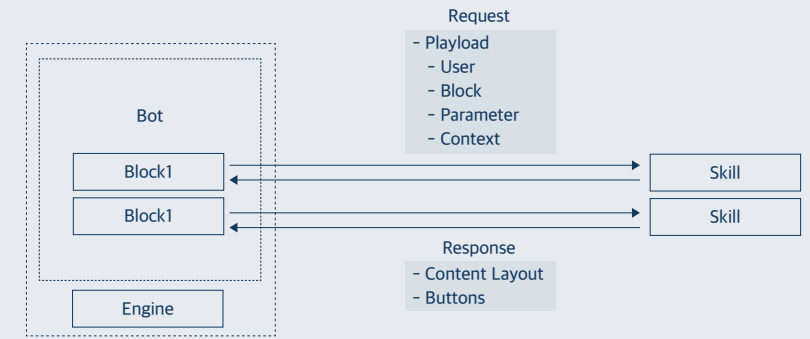
사용자가 입력한 원시 자료에는 핵심 질의(query)와 함께 질의를 처리하는 데 필요한 필수적인 매개변수(parameter)들과 부수적인 매개변수(optional parameter)들이 있을 것이다. 사용자가 그저 '날씨'라는 질의만 하고 날씨 봇이 이를 처리하는 경우를 예로 들어보자. 이 경우 사용자가 기본값(default)으로 지정된 '한 장소'와 '시간'이라는 필수적인 매개변수를 지칭하지 않았기 때문에 날씨 봇은 사용자의 물음에 답할 수 없게 된다. 이때 장소나 시간 등 같은 특성을 가진 매개변수들끼리 묶어서 단어 집합을 만든 것을 엔티티(entity)라고 한다. '장소'에는 서울, 경기, 부산 등이 들어갈 수 있고, '시간'에는 내일, 주말, 오후 등이 속할 수 있다. 이러한 엔티티는 장소 엔티티에 도시, 나라, 지역 엔티티를 두는 식으로 그 특성에 따라 서브 엔티티(sub entity)를 가지기도 한다.

콘퍼런스 발표 | 황지수 mario.h@kakaocorp.com

글 | 이형남 su8189@gmail.com 건국대학교에서 경제학과 통계학을 공부하고 있습니다. 머신러닝, 특히 딥러닝에 관심을 갖고 기술이 사람들에게 어떤 미래를 보여줄 수 있을지 꿈꾸며 학부 시절을 보내고 있습니다. 기술과 지식으로 쌓아올린 지혜는 삶을 풍요롭게 변화시킬 것이라 믿고 오늘도 배움의 길을 걷고 있으며, 기술과 문화예술 등 모든 분야의 흥미로운 생각을 사람들과 나누는 것을 좋아합니다.



[그림 2] 봇 플랫폼의 스킬 기능 모식도

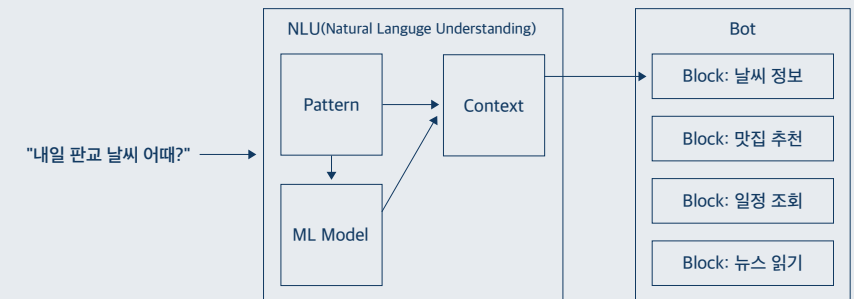


이러한 과정을 통해 매번 단어뭉치(word bag)를 생성할 필요 없이 기존에 정의된 엔티티들을 기반으로 봇을 만들 수 있게 되었다. 그러나 날씨를 알려주는 것과 같이 봇 플랫폼에서 기본적으로 제공하는 기능 외에 '사용자 맞춤 기능'에 대한 수요가 생기면서 스킬(skill)이 주목받게 되었다. 스킬은 봇 플랫폼상에서 봇을 개발할 때 필수적인 기능은 아니지만, 다른 앱과 상호작용하는 등 봇 내부적으로 구현된 기능 이상의 퍼포먼스를 내기 위해 사용할 필요가 있다.

**블록 분류(Block Classification)**

일반적으로 전화상담을 진행하는 고객센터에 전화를 걸면, 고객센터는 고객이 어떤 용무로 전화를 걸었는지 파악해야 한다. 그리하여 고객에게 여러 번 다이얼 누르기를 요청하며 고객이 전화를 건 목적을 파악하고자 할 것이다. 하지만 자연어 처리 방식으로 사용자와 상호작용하는 챗봇은 사용자가 자신의 목적을 스스로 분류하도록 시키는 것이 아닌, 사용자의 목적을 스스로 파악하여 한 번의 질의에도 답변을 되돌려준다. 그렇다면 이러한 일이 가능한 이유는 무엇일까?

[그림 3] 블록의 분류



가장 큰 차이점은 챗봇은 사람이 자연어로 입력한 정보를 그대로 받아들여 이해할 수 있기 때문이다. 봇은 자연어 이해(natural language understanding, NLU) 기능이 구현되어 있어서 충분한 정보가 담긴 질의를 받는다면 태그되어 있는 기능 블록에 연결되어 답변을 하게 된다.

봇이 받는 질의들을 구별하기 위해서 질의의 핵심을 구별해 미리 분류된 구분자를 붙여주는데 이를 태그(tag)라고 한다. 태그를 붙이게 되면 사용자가 날씨에 대해 묻든지, 기상정보에 대해 묻든지 의미상으로 같은 질문에 같은 답변을 할 수 있게 된다.

이 과정은 머신러닝(machine learning)을 통해 자연어 처리 과정을 거친다. 질의의 특징을 구별하기 위해 특성을 추출해내는 데 탁월한 성능을 보여주는 합성곱 신경망(convolutional neural network, CNN)을 활용한다. 머신러닝 분류기(machine learning classifier)에 의해 질의에서 특징이 추출되면, 질의 패턴 분류기의 결과와 종합하여 어떤 기능 블록에 해당하는 질문인지를 판단하고 기능 블록으로 질의를 전달한다.

기능 블록은 질의를 처리하기 위한 실제 기능이 구현된 부분이다. 자연어 이해 과정을 통해 사용자의 요구가 '서울'의 '내일' '날씨'가 궁금한 것이라면 날씨 정보를 구현해주는 날씨 기능 블록으로 연결된다. 서울의 지리 정보, 내일의 시간 정보에 따라 기상예보 정보를 알아내고 그 정보를 다시 자연어로 변환하여 사용자에게 답변해준다.

### 통합(Integration)

카카오 봇 플랫폼은 카카오톡 내에서 질의응답을 주고받는 것뿐 아니라, 타 앱과도 연동되어 있다. 챗봇을 통해 타 앱의 기능과 관련된 질의를 하게 되면 타 앱에서 기능이 동작하는 상호작용을 구현하는 것이다. 사용자가 멜론(Melon) 챗봇을 통해 '노래 틀어줘'라고 지시를 하면, 멜론 앱에서 노래를 재생하는 것이 가능하다는 것이다. 카카오는 타 앱과의 연동을 시작으로 다른 서드파티(3rd party)들과의 상호작용을 넘어서 사물인터넷(Internet of Things, IoT)과의 결합까지 플랫폼을 확장할 계획이라고 밝혔다.

오는 12월, 카카오는 '오픈 빌더(Open Builder)'라는 이름으로 봇 플랫폼의 챗봇 빌더를 공개 할 계획이다. 오픈 빌더가 공개되면 앞서 소개하였던 방식으로 일반 개발자들 또한 자유롭게 챗봇을 개발할 수 있다. 이미 공개되어 있는 다양한 봇 플랫폼도 있지만 많은 사람들이 국내 제일의 메신저에 기반한 오픈 빌더로 어떤 작업을 해낼 수 있을지 벌써부터 기대된다.



〈카카오 봇 플랫폼 소개〉  
브랜치로 연결되는 QR 코드  
입니다.

# In-Depth

Sports & AI

개황   디지털 사회로의 변화와 스포츠	56
경기   스포츠 경기력 향상을 위한 AI 활용 방안	64
분석   스포츠 경기 분석 전문가와 AI의 만남	70
판정   로봇 심판과 판정 알고리즘의 의미 있는 도전	78
도핑   스마트 도핑 검사의 도입과 발전 과제	84
보도   스포츠 저널리즘의 변화와 AI의 활용	92
챗봇   만능 스포츠봇의 등장과 발전 방향	98

# 디지털 사회로의 변화와 스포츠

사물과 통신, 지식과 산업이 융합하는 디지털 혁명의 거대한 물결이 사회 전체를 변화시키고 있다. 스포츠 분야라고 이러한 변화에서 예외일 수는 없다. 이번 글에서는 디지털 기술이 스포츠 산업에 미치는 영향과 향후 전개될 미래상에 대해 살펴보고자 한다.

스포츠의 본질은 무엇일까? 바로 이기고 싶다는 인간의 욕망이다. 그 대상은 경쟁자일 수도 있고, 나 자신일 수도 있다. 그런 점에서 볼 때 디지털 기술과 스포츠의 만남은 최고의 결합이라 할 수 있다. 어떻게 해야 이길 수 있는지를 알려주는 열쇠와 같은 존재가 디지털 기술이기 때문이다.

그 열쇠는 빅데이터의 모습을 하고 있거나 사물인터넷(IoT)이 될 수도 있다. 또한 어느 순간 인공지능(AI)으로 변신하거나 가상현실(VR)을 통해 해답을 알려주기도 한다. 삶의 모든 영역이 아날로그에서 디지털로 변해가고 있듯이, 스포츠 분야도 디지털이 일으키는 변화의 바람을 타고 새로운 시대로 진입하고 있다.

## 빅데이터와 스포츠

스포츠 산업과 디지털의 만남에서 가장 앞서나가는 기술은 빅데이터이다. 이기고 싶다는 인간의 욕망을 가장 충실히 수행하는 기술이라 여겨질 만큼, 스포츠에서 빅데이터를 활용하는 것은 이미 필수적인 과정으로 자리 잡았다.

과거 스포츠 경기에서 수집한 데이터를 모아 분석하면 유리한 전략을 도출해 경기에서 이길 확률이 높아진다는 빅데이터의 핵심이 이미 사실로 증명되었기 때문이다. 가장 대표적인 사례로 지난 2014년에 개최된 브라질 월드컵을 꼽을 수 있다. 당시 마이크로소프트(Microsoft)가 독자적으로 개발한 데이터 분석 예측 시스템이 16강전에서 벌어진 8경기의 결과를 모두 맞히며 화제가 되었다.<sup>1</sup> 이 예측 시스템에는 국가별 선수들의 경기력을 바탕으로 경기장의 지정학적 영향과 상대 국가와의 과거 전적, 경쟁 관계 여부 등 경기와 관련된 모든 데이터가 입력됐고, 마이크로소프트는 이를 복합적으로 분석하여 놀라운 결과를 만들어냈던 것이다.

빅데이터의 놀라운 능력은 여기서 그치지 않았다. 당시 우승국인 독일 역시 '매치인사이트(match insights)'라는 빅데이터 분석 시스템의 도움을 받았다. 이 시스템은 선수들의 몸에 부착한 센서로 데이터를 수집한 뒤 실시간으로 선수들의 기록 및 영상을 제공하여 코칭스태프가 빠르게 의사결정을 내리도록 하는 데 결정적인 도움을 줬다.<sup>2</sup>

## 사물인터넷과 스포츠

빅데이터와 사물인터넷의 관계는 실과 바늘 같다고 할 수 있다. 앞에서 예로 든 독일의 매치인사이트도 사물인터넷 기술을 응용한 '웨어러블(wearable)' 기술이 있었기에 빅데이터 분석이 가능했다. 매치인사이트는 훈련 중인 선수들의 무릎과 어깨에 부착된 4개의 이동식 센서를 통해 운동량을 파악할 수 있도록 설계되었다. 센서를 통해 선수들의 순간속도와 심박수 그리고 슈팅 동작 등에 대한 데이터가 실시간으로 수집되면 이를 분석하여 운동량을 파악하는 것이다.

<sup>1</sup> 참고 | <https://www.telegraph.co.uk/technology/microsoft/10959573/Germany-to-defeat-Argentina-in-World-Cup-final-predicts-Cortana.html>  
<sup>2</sup> 참고 | <https://www.sporttechie.com/sap-arms-reigning-world-cup-champs-with-tools-to-defend-title>

글 | 김준래 [stimes@naver.com](mailto:stimes@naver.com) 생명공학을 전공한 후 여러 연구기관에서 연구원으로 활동했습니다. 과학 전반에 대한 관심이 높아지면서 '과학문화 커뮤니케이션' 과정을 이수하였고, 현재 과학기술 전문 매체에서 객원기자로 활동하고 있습니다. 어렵게만 느껴지는 과학기술을 보다 쉽고 재미있게 전달하는 일에 관심이 많습니다.

[그림 1] 사물인터넷과 빅데이터의 융합으로 공의 움직임을 추적할 수 있는 아디다스 마이코치 시스템<sup>3</sup>



이와 유사한 사례로 아디다스(Adidas)의 마이코치(miCoach) 시스템을 들 수 있다. 매치인사이트와의 차이점이라면 센서를 부착하는 대상이 선수가 아닌 축구공이라는 점이다.<sup>4</sup> 마이코치 시스템은 공 안에 이동식 센서를 내장하여 선수가 공을 찰 때 공의 속도는 물론 회전수와 궤적 등을 종합적으로 분석할 수 있다. 따라서 코칭스태프는 선수가 공의 어느 부분을 강하게 찼는지, 또는 어떤 곡선을 그리며 공이 날아갔는지 등을 실시간으로 파악할 수 있다.

이 같은 웨어러블 기술은 스포츠 용품의 생산 방식에도 일대 혁신을 가져오고 있다. 얼마 전까지만 해도 스포츠 용품은 대량생산 방식이 주를 이루었다. 그러나 사물인터넷이 본격적으로 스포츠 용품에 적용되기 시작하면서 경기력 향상을 위한 '개인 맞춤' 형태나 '다품종 소량 생산' 형태로 변화하고 있다.

웨어러블 기술이 적용된 맞춤형 스포츠 용품을 소개할 때면 항상 거론되는 인물이 있다. 바로 메이저리그(MLB)에서 대표적 강타자로 활약하고 있는 LA 에인절스(Los Angeles Angels)의 마이크 트라웃(Mike Trout)이다. 트라웃은 연습할 때마다 스마트 배트(smart bat)라고 불리는 야구 방망이를 활용한다. 스포츠 용품 전문 기업 '제프(Zepp)'이 만든 이 스마트 배트는 야겟부분에 탑재된 센서를 통해 타자의 타구 속도와 궤적, 각도 등과 관련된 데이터를 수집한다. 그리고 수집된 데이터는 즉시 스마트폰 앱으로 전송되고, 선수는 이 데이터를 바탕으로 타격 자세를 교정할 수 있다.<sup>5</sup>

<sup>3</sup> 참고 | <https://www.thecoolist.com/adidas-micoach-smart-ball/adidas-micoach-smart-ball-fitness-tracker-soccer-ball-3>  
<sup>4</sup> 참고 | <https://www.swaggenmagazine.com/home/culture/gear-and-tech/adidas-micoach-smart-soccer-ball>  
<sup>5</sup> 참고 | [https://support.zepp.com/customer/portal/articles/2476168?b\\_id=12409](https://support.zepp.com/customer/portal/articles/2476168?b_id=12409)

### 인공지능과 스포츠

프로 바둑기사 이세돌 9단을 압도적으로 이기면서 전 세계 많은 사람들에게 충격을 주었던 인공지능 기술은 국내에서는 막연한 공포심의 대상으로 통한다. 하지만 기술 개발의 산실이라 할 수 있는 미국에서는 이미 인공지능이 주요 스포츠 분야에 많은 영향을 주고 있다.

그중에서도 다른 스포츠에 비해 유난히 데이터가 많이 축적되는 야구는 인공지능 기술이 가장 활발하게 적용되는 종목이라 할 수 있다. 예전에는 선수 기용부터 작전을 지시하는 순간까지의 모든 과정이 감독의 감에 의존하는 경우가 대부분이었다. 하지만 이제는 개인의 편견이 강하게 작용할 수 있는 감독의 판단보다는, 인공지능이 데이터를 분석하여 제시한 의견을 경기에 반영하는 사례가 점차 증가하고 있다.

대표적으로 메이저리그 구단 중 하나인 시카고 컵스(Chicago Cubs)가 활용한 키나트랙스(KinaTrax)라는 인공지능 시스템을 꼽을 수 있다. 키나트랙스는 선수들의 신체 정보를 실시간으로 추출하여 부상을 방지하고 선수들이 최상의 컨디션으로 경기를 치를 수 있도록 돕는 시스템이다.<sup>6</sup> 가령 선수의 성적이 부진할 때, 과거 성적이 가장 좋았던 시즌의 데이터와 현재의 데이터를 인공지능 시스템이 비교하여 문제점을 찾아낸다. 사람이 육안으로만 데이터를 비교하는 것과는 차원이 다른 결과를 도출해낼 수 있다.

### 가상현실과 스포츠

스포츠는 운동장이나 체육관에서만 즐길 수 있다는 편견을 깬 디지털 기술이 바로 가상현실(virtual reality, VR)이다. 사실 골프나 축구 같은 스포츠는 가상현실 기술이 등장하기 전에는 야외에서만 할 수 있었던 종목이었다. 그러나 이제는 가상현실이나 증강현실(augmented reality, AR) 기술을 통해 실내에서도 골프나 축구 같은 야외 스포츠를 충분히 즐길 수 있게 되었다. 특히 날씨나 미세먼지 등 외부 환경의 제약 없이 스포츠를 즐길 수 있다는 장점 때문에 훈련이나 건강관리 용도로 활용하는 사례가 늘어나고 있다.

[그림 2] 스트리버 랩(Strivr Labs)이 개발한 미식축구 훈련용 가상현실 시스템<sup>7</sup>



<sup>6</sup> 참고 | <https://www.sporttechie.com/imerit-cubs-kinatrax-biomechanics-pitching-hitting-cricket>  
<sup>7</sup> 참고 | <https://newsela.com/read/football-virtualreality>

가상현실 기술로 탁월한 훈련 효과를 거둔 스포츠 종목에는 미국의 국민 스포츠인 미식축구가 있다. 스트리버 랩(Strivr Labs)이라는 가상현실 전문 스타트업이 개발한 ‘훈련용 미식축구 게임’은 헤드셋을 통해 미식축구를 체험할 수 있는 프로그램이다.\*8 이 프로그램은 경기 전체를 조율하는 쿼터백의 시선으로 경기 내용을 파악할 수 있도록 설계되었다. 쿼터백의 시선으로 상대 선수들의 움직임을 한눈에 파악할 수 있기 때문에 이를 통해 선수들은 어떤 방식으로 상대팀을 공격할 수 있을지에 대한 전략을 세우기가 용이하다.

특히 가상현실의 장점을 충분히 살려 평소에는 접할 수 없었던 여러 방향에서 경기를 조망해볼 수 있다. 실제 경기 중에는 전후좌우로 130도 정도의 시야밖에 확보할 수 없지만, 헤드셋으로는 전후좌우의 시야뿐 아니라 위에서도 내려다볼 수 있기 때문이다. 또한 가상현실 속에서의 경기가 마무리되면 몇 번이고 헤드셋을 통해 경기 과정을 재현해볼 수 있다. 이 같은 기능 때문에 코칭스태프들은 경기 내용을 보다 복합적으로 파악할 수 있고, 경기 이후 선수들의 잘못된 점도 쉽게 고칠 수 있다.

이 같은 훈련용 가상현실 외에 최근 들어서는 건강관리를 위한 피트니스 분야에서도 가상현실을 적용한 서비스가 등장하여 주목을 끌고 있다. 얼마 전 유명 권투선수들과 시합을 하거나 그들의 연습 과정을 따라 하며 체중을 조절할 수 있는 복싱 프로그램이 선을 보여 주목을 끌었다.\*9

**디지털 영상과 스포츠**

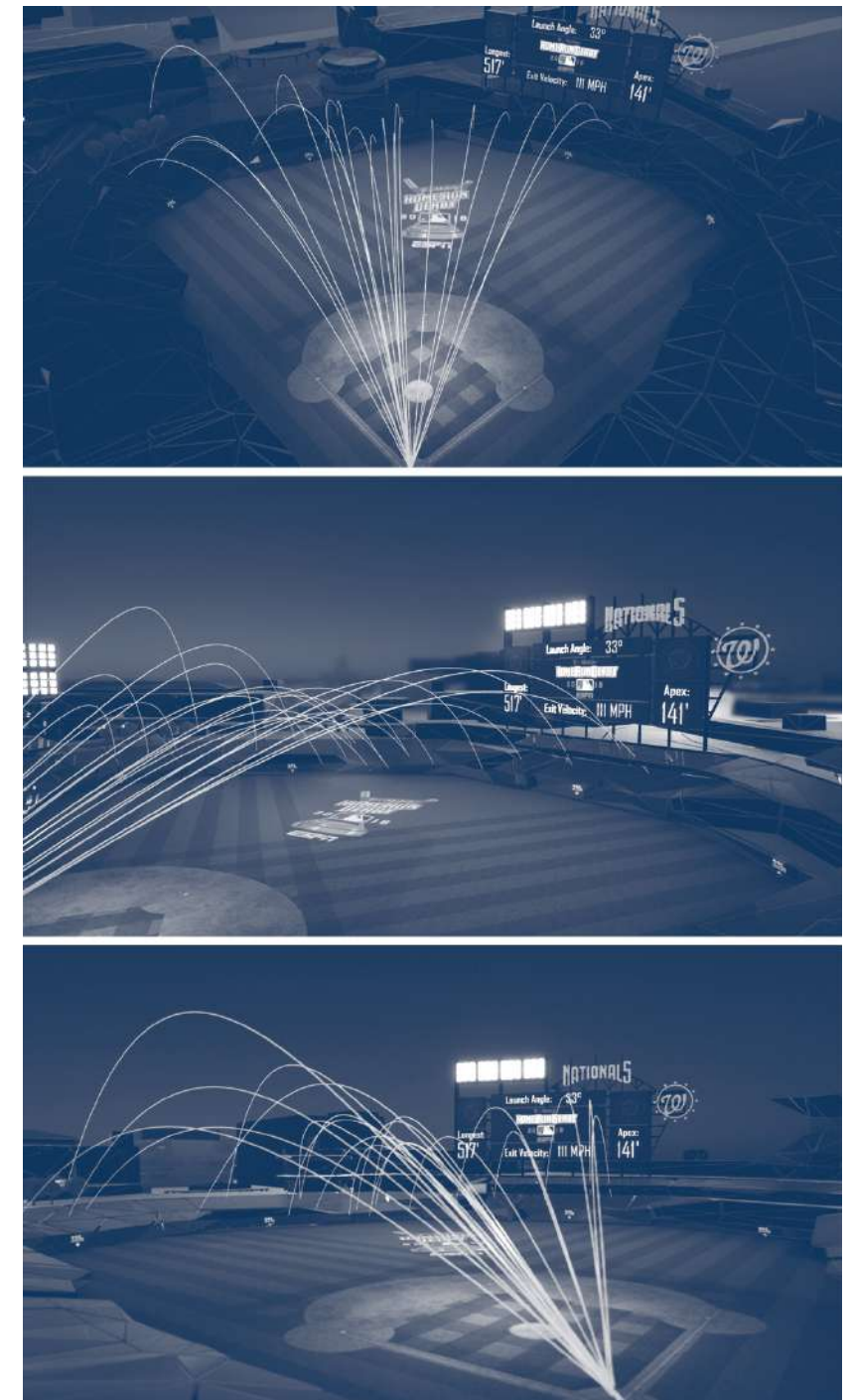
시합 중 심판의 착각이나 잘못으로 인해 판결이 잘못 내려지는 경우를 종종 볼 수 있다. 친선 경기나 비중이 크지 않은 시합의 경우 상대팀의 배려로 그냥 넘어가는 것이 대부분이지만, 월드컵이나 올림픽처럼 모든 선수들이 사활을 걸고 도전하는 시합에서의 오심은 도저히 묵과할 수 없는 사안이다.

이런 이유로 다양한 스포츠 종목에서 경기 중 일어나는 상황에 대해 비디오 분석과 판독을 하고 있다. 기록이나 득점을 중심으로 판정을 하는 육상이나 구기 종목은 물론이고, 심판들의 주관적 판단에 많이 의존하는 체조나 피겨스케이팅 등에서도 디지털 비디오를 통한 판독 과정이 도입되는 추세다. 디지털 비디오란 동영상과 동기화된 음성을 함께 디지털화하여 컴퓨터를 통해 볼 수 있도록 만든 데이터를 말한다. 실제로 우리는 지난 러시아 월드컵에서 디지털 비디오가 내리는 판정 시스템인 ‘VAR(video assistant referee)’의 위력을 실감한 바 있다.\*10

물론 디지털 영상이 스포츠 경기의 판결에만 사용되는 것은 아니다. TV나 인터넷 등으로 스포츠를 시청하는 시청자들을 위하여 새로운 디지털 기술이 적극적으로 활용되고 있다.

대표적으로 메이저리그에서 미사일 추적 기술 및 영상 처리 기술을 복합적으로 응용하여 개발한 스탯캐스트(statcast) 시스템을 들 수 있다.\*11 스탯캐스트는 공의 궤적뿐 아니라 주자의 반응속도, 타자의 달리기 속도 등을 실시간으로 중계 화면에 표시하여 야구를 보다 전문적으로 즐기기 원하는 시청자들을 사로잡고 있다. 이 시스템은 공을 추적할 수 있는 ‘레이더 기술’과 다각도에서 촬영이 가능한 ‘옵티컬 카메라 기술’을 결합한 것으로, 야구공을 포함해 필드 위 모든 선수들의 움직임을 정확하게 추적할 수 있다.

[그림 3] 야구 경기 중에 일어나는 모든 움직임을 추적할 수 있는 스탯캐스트\*11



\*8 참고 | <https://digit.hbs.org/submission/strivr-labs-vr-training-in-nfl/>  
 \*9 참고 | <https://vrscout.com/news/mayweather-boxing-fitness-vr-program-gym>  
 \*10 참고 | <https://www.engadget.com/2018/07/17/fifa-world-cup-2018-var-video-assistant-referee>  
 \*11 참고 | <https://www.sportsvideo.org/2018/07/16/live-from-mlb-all-star-espn-adds-new-dimensions-to-home-run-derby-with-4d-replay-3d-spray-charts>



### 스포츠의 미래상

인공지능이 발달하면 모든 스포츠 경기에서 심판이 사라질 수 있다는 예측도 나온다. 스포츠의 미래상이 자못 궁금해지는 지점이다. 이 같은 예측은 디지털 기술과 스포츠의 융합 없이는 도저히 불가능한 일이기 때문이다.

머지않은 미래에는 오로지 선수들만이 경기장에서 뛰고, 모든 판정은 디지털 기술에 의해 이루어지며, 관중들도 현장이 아닌 다양한 방식의 최첨단 기술을 통해 경기를 관람하는 시대가 펼쳐질지도 모른다.

하지만 스포츠는 스포츠만이 갖고 있는 숭고한 가치가 있다. 기적과도 같은 역전극이나 팬들의 열화와 같은 응원, 그리고 부상을 당했음에도 불구하고 팀을 위해 끝까지 경기에 임하는 선수 등이 그런 가치에 부응하는 모습들이다. 그런 점에서 볼 때 스포츠의 형태는 디지털 기술과의 융합을 통해 기존의 전통적인 방식과는 많이 달라질 수는 있어도 스포츠의 본질, 즉 이기고 싶다는 인간의 욕망이 사라지지 않는 한 스포츠는 영원히 존재할 것이다.\*12

\*12 참고 | <문화기술, 스포츠를 바꾸다>  
(2017), 《문화와 기술의 만남》 통권 52호,  
한국콘텐츠진흥원



<디지털 사회로의 변화와 스포츠> 브런치로 연결되는 QR 코드입니다.

# 스포츠 경기력 향상을 위한 AI 활용 방안

지난 2016년 구글(Google) 딥마인드(DeepMind) 인공지능(AI) 시스템인 알파고(AlphaGo)와 인간 대표 이세돌의 바둑 대결은 많은 사람들, 특히 스포츠 과학자들에게 AI의 가능성과 위협을 느끼게 한 사건으로 아직도 회자되고 있다.<sup>1</sup> 알파고 이전부터 FIFA 온라인 게임, 경기 분석(notational analysis) 분야에서 스포츠 AI 활용 방안에 대해 이미 논의되어 왔으나, 대부분의 스포츠 과학자들이 아직까지는 AI가 인간의 능력을 넘어서지 못할 것이라 예상했기 때문이다.

하지만 여전히 AI가 극복하지 못한 스포츠 영역들은 무수히 많다. 딥러닝과 같은 AI 기법은 신체적인 접촉이 경기 결과에 영향을 주지 않는 바둑과 같은 추상전략게임(abstract strategy game)에서는 엄청난 힘을 발휘하지만, 인간-인간, 인간-사물, 사물-사물의 신체적인 접촉이 경기 결과에 영향을 주는 축구, 농구, 배구, 야구, 컬링과 같은 반추상전략게임(semi-abstract strategy game)에서는 현재 알려진 대부분의 AI 기법들이 큰 힘을 발휘하지 못하고 있다.<sup>2</sup> 현재 기술 수준의 AI를 반추상전략게임에 적용하는 것이 매우 어렵다는 것을 알려주는 한 사례가 있는데, 바로 독일 축구대표팀이다.

독일 축구대표팀은 2014 FIFA월드컵(제20회 브라질 월드컵)에서 우승을 한 후, 2018 FIFA월드컵(제 21회 러시아 월드컵)에서는 조별 예선에서 우리나라 대표팀에 패배해 탈락했다. 2014 브라질 월드컵에서 우승한 후 언론과의 인터뷰에서 독일은 'SAP의 빅데이터 기술을 기반으로 선수들의 훈련을 체계적으로 관리했으며, 상대팀의 전략에 따라 맞춤형 전술을 사용한 것이 승리의 원동력'이었다고 말했다.<sup>3</sup> SAP의 기술이 퇴보한 것은 아닐 테니, 현재의 시로 스포츠 경기력을 향상시키는 것에 한계가 발생하는 원인은 무엇일까? 이 질문에 대한 답을 한다면, 크게 두 가지로 답할 수 있을 것이다. 첫째, AI 엔지니어와 스포츠 과학자들이 서로 다르게 정의하는 '경기력'이라는 용어는 스포츠에 AI를 활용하는 데에 장애 요인이 될 수 있다. 둘째, 유의미한 결과를 주면서 학습 가능한 수준의 데이터 세트(data set)의 부재이다. 위 두 가지 원인에 대해 상세히 다루고자 한다.

[그림 1] 독일 축구대표팀이 사용하는 SAP 스포츠 원(SAP Sports One) 솔루션<sup>4</sup>



<sup>1</sup> 참고 | <http://www.sisanews.co.kr/news/articleView.html?dxno=890>  
<sup>2</sup> 논문 | 박성건·이수원·윤형기(2017). <컬링 인공지능 적용을 위한 탐색적 연구> 《한국체육학회지》, 한국체육학회  
<sup>3</sup> 참고 | G. Clarke, "SAP: It was our Big Data software that won it for Germany", 2014-07-22, [https://www.theregister.co.uk/2014/07/22/germany\\_worldcup\\_sap\\_hana](https://www.theregister.co.uk/2014/07/22/germany_worldcup_sap_hana)  
<sup>4</sup> 참고 | [https://www.clarin.com/sociedad/tecnologica-secreta-alemania-ganar-mundial\\_0\\_r1TGtfsqPXg.html](https://www.clarin.com/sociedad/tecnologica-secreta-alemania-ganar-mundial_0_r1TGtfsqPXg.html)

글 | 박성건 [sgpark@sportsbon.com](mailto:sgpark@sportsbon.com) 배구선수이자 스포츠과학(운동생리학) 연구자로 활동하다 박사과정 중 컬링을 연구하면서 스포츠 인공지능 분야에 관심을 갖기 시작했습니다. 현재는 스타트업 '스포츠본주식회사' 대표로 활동하면서 스포츠 전문가 O2O 매칭 서비스를 운영 중에 있습니다. 인간 행동을 다학문적으로 접근하여 연구하는 것에 관심이 많고, 이를 배우고자 노력하고 있습니다.

글·감수 | 이수원 [swlee@ssu.ac.kr](mailto:swlee@ssu.ac.kr) 현재 송실대학교 IT대학 소프트웨어학부 및 정보통신소재융합학과(스포츠IT융합학 전공) 교수로 재직 중이며, 컬링 경기력 향상을 위한 승패예측모형 등 스포츠 인공지능 분야에서 연구를 수행하고 있습니다. 관심 분야로는 데이터 사이언스, 인공지능, 텍스트마이닝, 개인화 추천, 스포츠 IT 융합 연구입니다.

**팩트 체크: '경기력(performance)'**

AI를 활용하여 스포츠 경기력을 향상시키고자 한다면, 필자는 '경기력(performance)'이란 용어를 스포츠 과학자와 엔지니어 입장에서 바라본 후 재해석하여 연구할 것을 권장한다. 스포츠 과학자들이 정의하는 경기력은 '특정 종목을 효율적으로 수행하기 위해 필요한 생체역학 및 생리학적 기능(biodynamic & physiologic function), 심리적 요인(psychological factor), 종목에 대한 기술적 수준(technical level), 전략적 요인(strategy level)이 상호복합적으로 작용된 포괄적인 능력'<sup>5</sup>인 반면, 대부분의 엔지니어들이 생각하는 경기력은 '특정 종목을 수행할 때 필요한 인간의 기능을 수치화한 지표'이기 때문이다.

위 두 가지 정의를 곰곰이 생각해본다고 해도 독자들은 큰 차이점을 느끼지 못할 것이다. 하지만 스포츠 과학자와 엔지니어가 각각 정의하는 경기력 중 '과사용(overuse)'과 '부상(injury)'이란 변수를 어떻게 판단하는지 이해한다면 두 집단의 차이점을 명확히 알아낼 수 있다. 선수들이 경기 중 자신이 가진 근육의 힘보다 더 큰 힘을 사용하거나 상대방을 막기 위해 급격하게 방향 전환을 하는 일은 빈번하다. 이로 인해 선수들의 근육과 관절이 '과사용'되는 일이 발생하여 경기력이 저하되기도 하고, 상대방 또는 운동용품(공, 배트 등), 경기장 시설물에 부딪쳐 발생하는 '부상'으로 인해 경기력이 저하되기도 한다.<sup>6</sup> 이러한 상황을 스포츠 과학자는 '생체역학 및 생리학적 기능의 일부 제한'(마이너스, -)으로 생각하는 반면에, 엔지니어는 '운동수행의 종료'(곱하기 0, ×0)로 받아들인다.

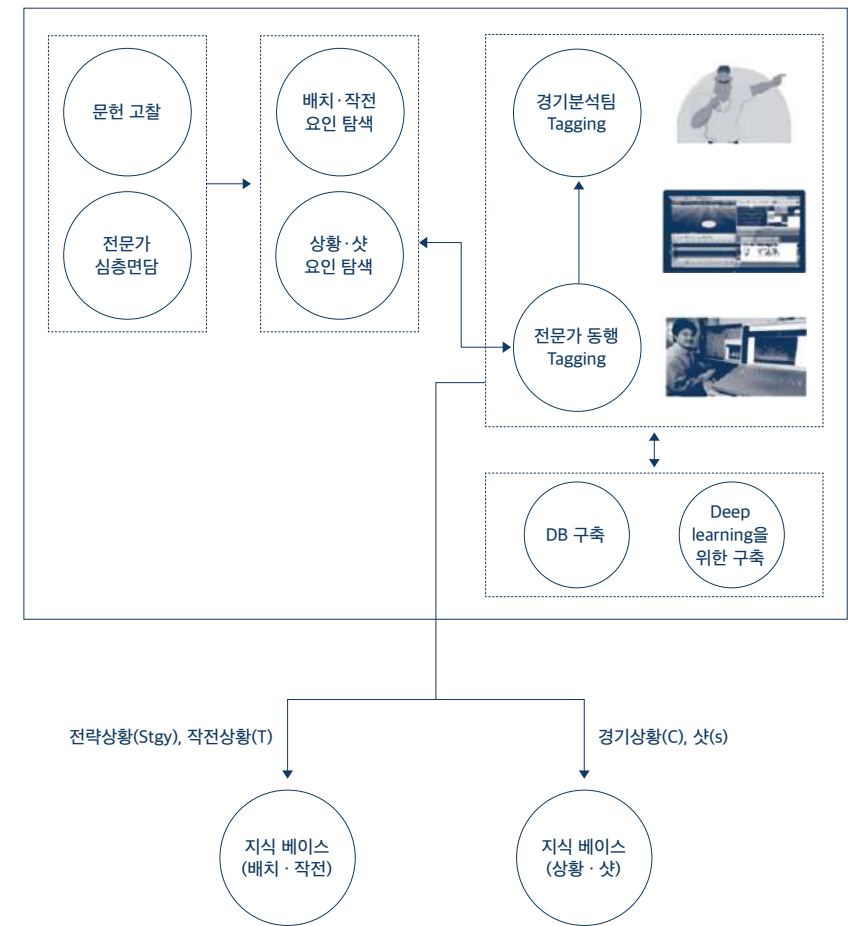
예를 들어보자. 2002년 한일 월드컵에서 우리나라 축구 국가대표팀의 황선홍 선수는 미국팀과의 경기 도중 상대 선수와 헤딩 경합 중 눈가가 찢어지는 부상을 당했다. 황선홍 선수는 심각한 부상에도 개의치 않고 붓대를 감는 응급처치를 받은 후 그라운드로 복귀한다.<sup>7</sup> 우리는 TV로 그런 선수의 모습을 보면서 '많이 다치지 않았어야 하는데...', '그래, 투혼이 필요해'라며 걱정을 했다. 되짚어보면 이런 걱정에는 '지금 이 순간을 참고 견뎌야 승리할 수 있어. 그러니까 아프더라도 뛰었으면 좋겠어'라는 속마음이 내포되어 있다. TV를 통해 경기를 지켜보든, 경기장에서 관람을 하든 지금 당장 선수가 그라운드에서 일어나서 경기에 승리하는 것을 보고 싶은 것이다. 선수 보호보다 승리가 우선인 것이 일반적인 사람들의 속마음이다. 하지만 인체 기능학적으로는 피부 조직이 찢어져서 출혈이 생길 경우, 우리 몸의 방어기전이 발동하여 적혈구와 백혈구 등이 서로 엉키면서 혈관 내부에 외부 공기가 들어오지 못하도록 그물망을 친다. 이 그물망에 재차 자극을 주게 되면 그물망은 다시 찢어지게 되고 이때 뇌는 선수에게 큰 고통을 느끼게 하여 외부와의 충동을 피하도록 경고한다.<sup>8</sup> 덧붙여 경기 중 선수가 출혈을 동반하는 부상이 발생될 경우, 체내 혈액량이 감소하게 되고 근육에 필요한 산소 공급이 저하되며, 이는 결국 근육경련(muscle cramps)으로 이어져 경기력이 저하되는 부수적인 문제가 발생할 수 있다. 이러한 상황에서도 대부분의 선수들은 고통을 참고 그라운드로 달려나가 승부를 펼치는데, 엔지니어들은 과연 이를 컴퓨터에 어떤 파라미터(parameter)로 표현하여 데이터값을 입력할 수 있을까? 아직까지 이 질문에 답할 수 있는 사람은 필자를 포함하여 아무도 없다.

\*5 논문 | M. Hughes, and I. Franks, "Notational Analysis of Sport: Systems for Better Coaching and Performance in Sport", In J Sports Sci Med, 2004  
 \*6 논문 | N. Maffulli, U. Longo, N. Gougoulas, D. Caine, and V. Denaro, "Sport injuries: a review of outcomes", British Medical Bulletin, 2011  
 \*7 참고 | 언론이 만들어 낸 '투혼' 프레임... 땅을 치는 선수들, <한국일보>, 2017.4.19 http://www.hankookilbo.com/News/Read/201704190486131017  
 \*8 논문 | C. Joo( 2011), <혈액응고 기전에 대한 고찰>, <J Korean Soc Neonatol>

**팩트 체크: 학습이 가능한 수준의 데이터 세트 존재 여부**

필자는 과거 미래창조과학부에서 추진한 '경기전략을 수립하고 경기 수행이 가능한 인공지능 컬링 로봇 기술 개발' 프로젝트를 수행하기 위해 다학제 연구팀을 구성한 경험이 있다. 불행인지 다행인지는 모르겠으나 최종적으로 과제 수주는 이루지 못했다. 그 외에도 다양한 정부 연구과제를 수행했을 때나 기업들이 데이터 분석을 의뢰할 때에도 필자는 공통적인 질문과 함께 학습을 내신 적이 많다. 더 나아가 스스로에게도 묻는다. 'AI(상세하게는 딥러닝(deep learning))를 적용할 만큼 의미 있는 데이터 세트는 존재하는가?' 결론은 '대부분 없다'이다. 결국 AI를 경기력 향상에 적용하고 구현하기 위해서는 학습이 가능한 수준의 데이터 세트를 '새로' 구축해야 한다.

**[그림 2] 필자가 과제 기획 시 참고한 '자료 수집을 통한 지식 획득 흐름'의 예시<sup>9</sup>**



앞서 '경기력'이란 정의부터 스포츠 과학자와 엔지니어 간에 차이점이 존재한다고 설명했다. 용어의 개념부터 차이가 존재하기 때문에 결론적으로 '학습이 가능한 수준의 데이터 세트는 없다'가 정확하다. 즉 AI를 활용하여 경기력 향상을 구현하기 위해서는 해당 스포츠 종목의 전문가와 스포츠 과학자, 엔지니어가 팀을 이루어 표준화된 편람(standardized terminology)을 기반으로 컴퓨터가 이해할 수 있는 수준의 데이터베이스(database, DB)를 구축하는 것이

\*9 참고 | <경기전략을 수립하고 경기 수행이 가능한 인공지능 컬링 로봇 기술 개발> 자료에서 일부 발췌

필요하다. 나아가, 스포츠 AI 개발을 위한 데이터베이스 구축 후 다양한 통계 방법을 이용하여 각 종목별 경기력 핵심 변인을 추출하는 과정이 필요하다. 진행 절차는 다음과 같은 순서로 할 것을 권장하며 이 글을 마무리하겠다.

[ 표 1 ] '자료 수집을 통한 지식 획득 흐름'에 대한 진행 순서

1단계	AI 개발 이해관계자들이 동일한 스포츠 경기 영상을 보고 경기 내용 기록하기
2단계	프로젝트에 참여한 사람들끼리 기록지를 바꿔가며 차이점 인식하기
3단계	AI 개발 이해관계자들이 함께 '경기력' 용어 재정의하기(경기력의 핵심 변수 선정)
4단계	표준화 용어 정의 검토하기(경기 기록자 중심으로)
5단계	경기력 결정 요인을 계량화할 수 있는 원시 데이터를 태깅하여 데이터베이스 구축하기
6단계	데이터베이스에 구축된 데이터 신뢰성 검증하기
7단계	경기에서 승리할 때 가장 중요한 경기력 결정 요인 찾기



〈스포츠 경기력 향상을 위한 AI 활용 방안〉 브런치로 연결되는 QR 코드입니다.

# 스포츠 경기 분석 전문가와 AI의 만남

필자는 어린 시절부터 운동을 했다. 운동선수로 6년간 활동했고, 대학에서 체육학을 전공했다. 지금도 스포츠를 자주 접하는 곳에서 생활하며 활동한다. 이런 필자에게 인공지능(AI)에 대한 글을 기고하는 기회가 주어졌을 때, 한치의 망설임도 없이 하겠다고 답을 할 수는 없었다. 왜냐하면 스포츠 경기 분석(performance analysis of sport) 분야에서의 인공지능은 여전히 진행형이며, 기술적인 부분에서 많은 전문가들이 존재하기 때문이다. 하지만 필자는 더 많은 사람들이 스포츠 경기 분석이란 분야에 대해 깊게 이해하고, 인공지능과 관련한 전문가들이 스포츠 전공자들과 협업하여 더욱 발전적인 미래를 만들어낼 수 있었으면 하는 바람을 가지고 있다. 이런 여러 가지 이유에서 용기를 내어 이 글을 기고한다.

## 스포츠 경기 분석

스포츠는 직접 참여함으로써 즐거움을 느낄 수도 있지만, 보는 것만으로도 즐거움을 얻을 수 있는 여가활동 중 하나이다. 스포츠에 대한 전반적인 내용을 모두 설명하기에는 많은 시간이 필요할 것이다. 하지만 스포츠를 업으로 삼고 생활하는 선수들과 지도자들은 그들이 하는 모든 활동 자체에 가치를 부여한다. 선수들 모두에게 스포츠란 그들의 경기력을 향상시키기 위한 활동의 장이자 목표이다.<sup>\*1</sup> 그런 이유에서 스포츠 경기 분석은 선수들의 경기력을 향상시키고, 더 나아가 선수들과 지도자들에게 차기 목표와 달성 방법을 설정하는 데 필요한 기초 자료를 제공해준다.<sup>\*2</sup> 스포츠 경기 분석은 국내로 들어오면서 여러 가지 의미로 해석되어 왔다.<sup>\*3</sup> 스포츠 경기 분석의 개념은 크게 두 가지로 접근할 수 있는데 스포츠 경기의 기록을 분석하는 의미로서의 스포츠 경기 분석(match analysis)과 스포츠 경기 중 관찰될 수 있는 내용을 중심으로 경기력(performance)을 분석하는 의미로서의 스포츠 경기 분석(performance analysis of sport)이다. 최근에는 스포츠에서 수집되는 자료를 바탕으로 역사적인 요인과 통계적인 요인을 병합하여 분석하는 스포츠 분석(sports analytics)<sup>\*4</sup>이란 개념으로 접근하고 있기도 하다. 무엇보다 어떠한 개념으로 스포츠 경기 분석을 접근하더라도 자료를 수집하는 과정에서 필요한 내용과 범위는 유사하다. 스포츠 경기 분석 관점에서는 이러한 자료 수집에 필요한 요인들을 분석인자(performance indicators)<sup>\*5</sup>라 부르며, 관계형 데이터베이스(database) 필드를 설정하는 개념과도 같다. 아래는 스포츠 경기 분석을 수행하기 위해 거쳐야 하는 단계들이다.

## 스포츠 경기 분석 단계

- (1) 자료 수집
- (2) 전처리
- (3) 자료 분석 및 통계 분석
- (4) 분석 자료의 시각화

스포츠 경기를 분석할 때 사용하는 여러 기법들은 분석의 목적과 목표에 맞게 적용하며, 경우에 따라서는 간단한 통계 처리부터 복잡한 통계를 접목한 분석기법도 활용되고 있다.

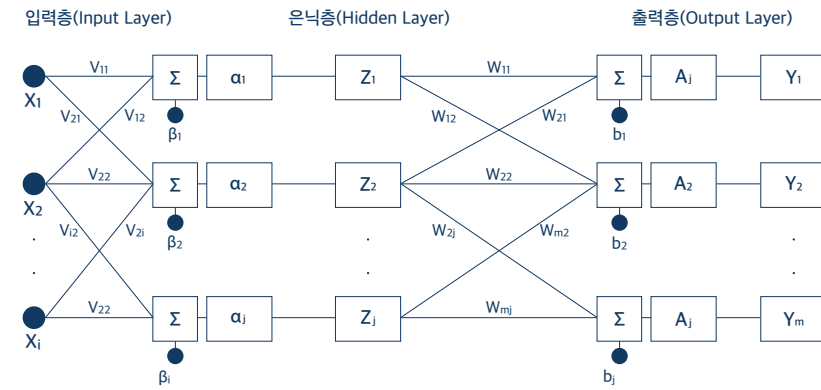
\*1 논문 | Fister, I., Fister, D., Deb, S., Mlakar, U., Brest, J., & Fister, I. (2018). Post hoc analysis of sport performance with differential evolution. *Neural Computing and Applications*, 1-10  
 \*2 논문 | McGarry, T., O'Donoghue, P., & Sampaio, J. (2013). *Routledge handbook of sports performance analysis*. (2013) Routledge Handbook of Sports Performance Analysis Xvii, New York, NY, US: Routledge/Taylor & Francis Group; US.  
 \*3 논문 | 최형준(2016). <축구 월드컵 경기 공식기록을 통한 데이터의 시각화>. 《한국체육측정평가학회》, pp.83-92  
 \*4 참고 | [https://en.wikipedia.org/wiki/Sports\\_analytics](https://en.wikipedia.org/wiki/Sports_analytics)  
 \*5 논문 | Hughes, M., Bartlett, R. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences*. 20, pp.739-754

글 | 최형준 [chj2812@dankook.ac.kr](mailto:chj2812@dankook.ac.kr) 체육측정평가를 전공하였으며, 세계스포츠경기분석학회 사무총장직을 수행하고 있습니다. 스포츠 경기 분석과 ICT 기술의 융합에 관심이 많아 스포츠 경기 분석 자료를 시각화하는 데 관심을 가지면서 AI와 접하게 되었습니다. 앞으로도 급속도로 발전하는 AI를 활용하여 스포츠 분야에 적용할 수 있도록 많은 콘텐츠를 개발하고 싶습니다.

**스포츠 경기 분석과 인공지능**

필자는 이 글에서 스포츠 경기 기록과 분석 자료를 토대로 어떠한 방식으로 인공지능을 활용하게 되었는지를 소개하고자 한다. 물론 기술적으로는 기존의 방식을 활용하는 수준이었지만, 스포츠 경기에서 얻을 수 있는 자료의 한계를 고려할 때 결코 쉬운 작업은 아니었다. 먼저 인공지능을 활용하여 2005년 영국 Wimbleton 테니스 대회의 경기 결과 예측에 관한 연구를 진행하였다. 2004년 127개 경기 중 학습이 불가능한 자료를 제외한 111개 경기 기록을 14개의 분석인자 값으로 입력한 후 학습을 진행하여, 2005년 경기 기록의 결과를 예측하는 방식으로 연구를 진행하였다.

[ 그림 1 ] 2005년 영국 Wimbleton 테니스 대회의 경기 결과 예측에 관한 연구에 사용된 인공신경망 모형\*6



[ 표 1 ] 경기 결과 예측 학습 과정에서 쓰인 매개변수의 설정값\*6

매개변수	설정값
목표 학습 횟수(epochs)	1,000
오차의 목표치(goal)	0
학습률(lr)	0.01
학습률 감소치(lr_dec)	0.7
학습률 증가치(lr_inc)	1.05
최대 학습 실패수(max_fail)	5
최대 오차 증가치(max_perf_inc)	1.04
관성상수(mc)	0.9
최소 경사 감소치(min_gad)	1e-006

연구결과를 보면 다양한 인공신경망 모델들은 결과를 도출하는 데 있어 매우 광범위한 결과를 가져올 수 있으며, 경기 기록을 토대로 결과를 도출하는 데도 이러한 특성을 고려해야 한다는 사실을 발견하였다. 또한 필자는 경기 기록만으로 예측에 필요한 인공지능을 학습시키는 것은 제한적이며, 이를 개선하기 위해서는 선수들의 특성을 고려한 자료 입력이 필요하다고 판단하였다. 필자는 선수들의 특성을 고려한 자료에 대해서 고민하게 되었다.

\*6 논문 | 최형준·김주학(2006), <인공 신경망(Artificial Neural Network)을 이용한 2005년도 영국 Wimbleton 테니스 대회의 경기 결과 예측에 관한 연구>, 《한국체육학회지》45(3), 한국체육학회, pp.459~467

따라서 전문적인 기자재를 활용하여 측정된 국가대표 선수들의 체격 및 체력 측정 자료를 인공지능에게 학습시킨다면, ‘인공지능이 선수의 체격 및 체력 데이터를 바탕으로 최적의 스포츠 종목을 알아낼 수 있지 않을까?’라는 고민을 하게 되었다.

스포츠정책과학원에서 수집하고 있는 국가대표 선수들의 체격 및 체력 자료를 이용하여 인공신경망 종류의 하나인 자기구성지도(Self-organizing Map 혹은 Kohen’s feature map)에 적용해보았다. 연구 대상의 체격 및 체력검사 내용은 [표 2]와 같았으며, 연구 대상은 근대 5종, 농구를 비롯한 19개 종목에서 288명의 선수들 자료를 사용하였다.

[ 표 2 ] 자기구성지도의 입력층에 입력되었던 연구 대상자의 체격 및 체력검사 내용\*7

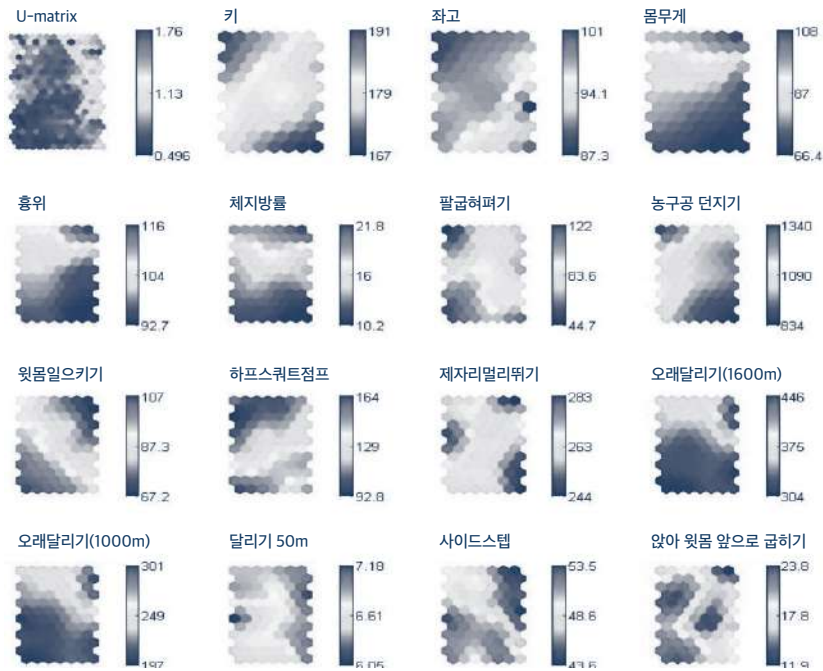
매개변수	요인	검사
체격	길이	신장
	길이	좌고
	무게	체중
	둘레	흉위
	신체 조성	체지방률
체력	상지근지구력	팔굽혀펴기
	상지순발력	농구공 던지기
	복근지구력	윗몸일으키기
	하지근지구력	하프스쿼트점프
	하지순발력	제자리멀리뛰기
	심폐지구력	오래달리기-걸기(1600m)
	심폐지구력	오래달리기(1000m)
	스피드	50m 달리기
	민첩성	사이드스텝 테스트
	유연성	앉아 윗몸 앞으로 굽히기

학습된 자기구성지도에 투영된 각 분석 인자들의 특성은 [그림 2]와 같다. 자기구성지도의 특성상 같은 셀에 있는 분석 인자들의 자료 범위가 모두 포함됨으로써, 학습된 자기구성지도에 새로운 자료가 투영될 때 각 셀에 포함되어 있는 분석 인자들의 특성을 반영하게 된다.

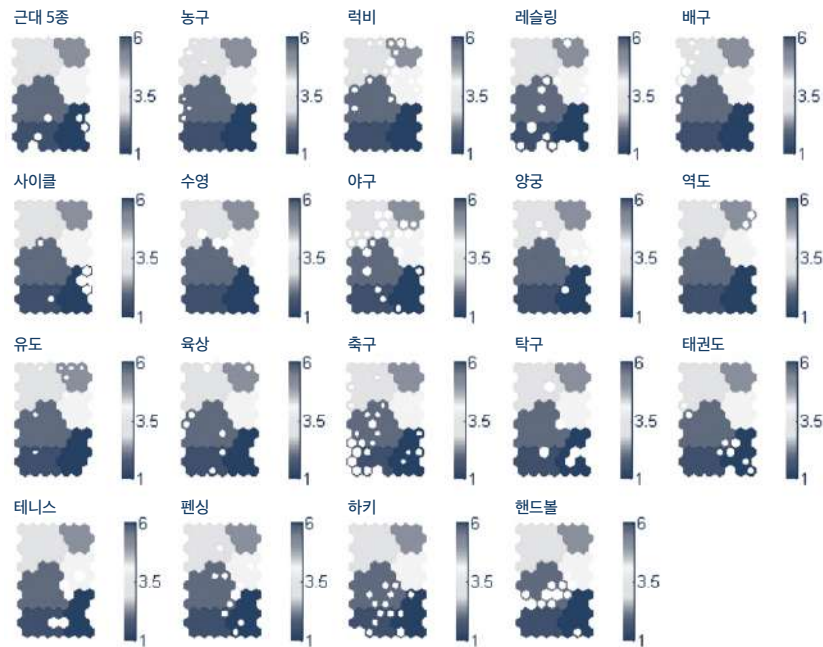
이러한 자기구성지도에 투영된 모든 분석 인자들의 특성을 토대로 군집분석(cluster analysis)을 실시하여 6개 군집으로 축약하였으며, 종목별 선수의 체격 및 체력의 특성을 반영하여 [그림 3]과 같은 결과를 도출하였다.

\*7 논문 | 최형준·정연성·고병구(2009), <한국 남자 국가대표 선수의 체격, 체력검사 결과와 자기구성지도(Self-Organized Map)를 이용한 스포츠 종목의 군집분석-사례연구를 중심으로>, 《체육과학연구》20(3), 한국스포츠정책과학원, pp.486~498

[ 그림 2 ] 자기구성지도에 투영된 분석인자의 자료 특성 시각화<sup>7</sup>



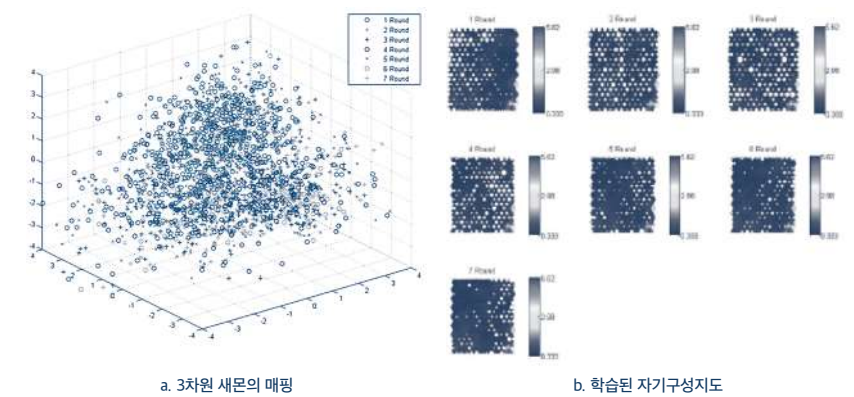
[ 그림 3 ] 6개 군집으로 나타난 자기구성지도에 투영된 스포츠 종목별 선수들의 원자료<sup>7</sup>



결과적으로 자기구성지도에 6개의 군집으로 구분된 종목별 스포츠 선수들의 체격 및 체력 자료는 대부분의 종목에서 가능하였다. 하지만 모든 종목에서 뚜렷한 분류 기준을 밝혀낼 수는 없었다. 그 이유는 포지션별로 선수를 구분하거나 체급이 존재하여 선수들의 체격 및 체력 특성에 차이를 보이는 종목이 존재하기 때문이다. 그러나 보다 광범위한 사실은 기존의 통계분석을 이용해 각 분석 인자에 대한 범위 내에서 분석한 방법에 비해서 인공지능의 포괄적인 분석을 적용할 수 있다는 점과 모든 자료 처리 과정을 시각화하여 살펴볼 수 있다는 점이다.

필자는 스포츠에서 생성되는 자료의 공공성과 인공지능의 자료 처리 및 시각화 기술에 대한 장점을 살려 스포츠 공식기록에서 도출되는 자료의 시각화에 관심을 갖게 되었다. 2005년부터 2010년까지 테니스 4대 그랜드슬램대회(US Open, Austrian Open, Wimbledon Open, French Open)에서 생성되는 자료의 공공성을 토대로 테니스 선수의 경기 기록을 시각화하는 연구를 진행하였다. 필자가 관심 있는 부분은 다차원적 자료(multi-dimensional data)의 시각화인데, 기존의 차원 축소 방법인 요인분석이나 판별분석으로는 자료의 다양한 특성을 시각화하는 데 한계가 있다고 판단하였다. 이에 다차원적 자료를 2차원 혹은 3차원으로 축소시킬 수 있는 새몬의 매핑(Sammon's mapping)과 자기구성지도를 이용하여 2005년부터 2010년까지 테니스 4대 그랜드슬램대회 2060세트를 시각화하였다. 이때 1라운드부터 7라운드까지의 경기를 토너먼트 형식으로 진행하는 테니스 경기의 특성을 반영하여 7라운드까지 진출한 선수의 경기 기록과 1라운드에서 경기를 마무리한 선수의 특성 간에는 분명한 차이가 있을 것이라 가정하였고, 이에 대한 비교를 중점적으로 조명하였다.

[ 그림 4 ] 인공지능 기법을 이용한 남자 테니스 선수의 경기 기록 자료<sup>8</sup>



연구결과를 살펴보면 새몬의 매핑은 자료의 수가 증가함에 따라서 시각화 진행 후 결과를 해석하는 데 어려움을 보였다. 반면, 학습된 자기구성지도를 통해 나타난 분석 결과는 자기구성지도의 각 셀에서 도출할 수 있는 분석 인자의 범위에 대한 시각화가 가능하다는 점에서 우수성을 보였다. 비록 스포츠 경기 기록 자료를 시각화하는 방법에 대한 간단한 비교였지만, 인공지능을 활용하여 스포츠 경기 기록을 시각화하고 자료를 처리하는 과정에서 도출되는 자료를 토대로 2차 가공이 가능하다는 점에서 의미가 있다고 하겠다.

<sup>7</sup> 논문 | 최형준·권민혁(2012), <인공지능(AI: Artificial Intelligence) 기법을 이용한 남자 테니스 선수 경기기록의 시각화>. 《한국체육측정평가학회지》 제14권 제3호, pp.47-56

**마치며**

인공지능은 산업 분야 전반에 걸쳐 다양하게 적용되고 있고, 그 활용 범위 또한 날이 갈수록 확대되고 있다. 스포츠 경기 분석 분야도 예외는 아니다. 특히 스포츠 경기 기록 분야에서의 활용은 인공지능의 진가를 발휘하기 좋은 주제이다. 최근에는 모든 스포츠 종목에서 경기 기록을 시각화하고 활용하고자 하는 노력이 계속되고 있다. 옵타(Opta)와 SAP 같은 기업들은 이미 스포츠 경기 기록 분야에서 새로운 시장을 창출해내고 있으며, 이를 스포츠 현장에 적용하는 사례도 다양하게 나타나고 있다.

하지만 이와 같은 아웃풋(output)을 내기 위해서는 사람이 판단해야 할 내용도 다수 포함되며, 사람이 판단하는 내용을 보완하기 위해서는 여러 가지 작업이 동시에 필요하다. 스포츠 경기 분석은 목표를 설정하는 단계부터 사람이 많은 작업을 진행한다. 스포츠 전문가와 인공지능 기술의 만남은 앞으로 스포츠 현장을 혁신적으로 뒤바꿔 놓을 것이며, 이를 위해 여러 가지 시도가 계속되어야 한다.



〈스포츠 경기 분석 전문가와 AI의 만남〉 브랜치로 연결되는 QR 코드입니다.

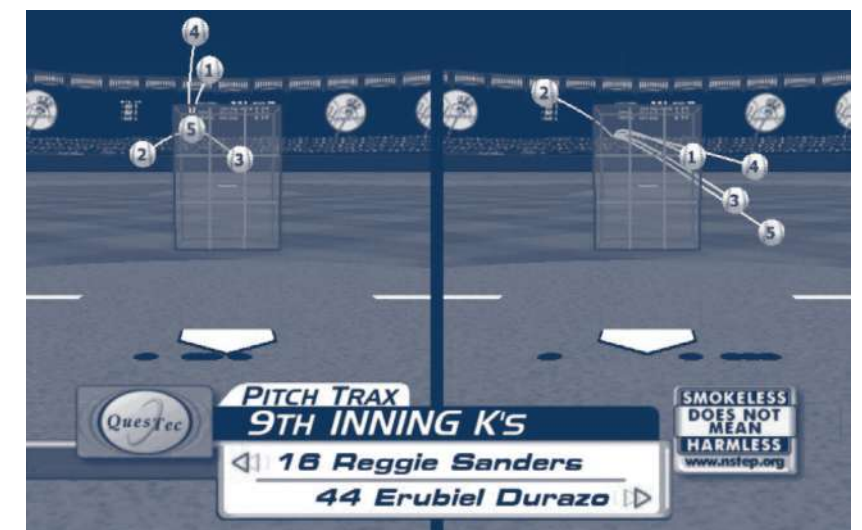


# 로봇 심판과 판정 알고리즘의 의미 있는 도전

2003년 6월 2일, 애리조나 다이아몬드백스(Arizona Diamondbacks) 소속의 투수 커트 실링(Curt Schilling)은 메이저리그 베이스볼 커미셔너(Major League Baseball Commissioner) 사무국으로부터 벌금 1500달러를 선고 받았다. 5월 24일 샌디에이고(San Diego)와의 경기에서 홈구장에 설치된 카메라 한 대를 부숴버렸기 때문이다. 카메라는 퀘스텍(QuesTec) 시스템의 일부로 사무국이 심판의 스트라이크 판정을 평가하기 위해 도입한 최신 장비였다. 실링은 해당 경기의 주심으로부터 "기계의 측정치에 맞춰 스트라이크존(strike zone)을 바꾸고 있다"는 말을 들은 뒤 화를 참지 못했던 것이다.

당시 메이저리그에서는 존(zone) 판정에 대한 불만이 고조되고 있었다. 높은 공은 판정이 지나치게 박하고, 바깥쪽으로 먼 공은 스트라이크 판정이 잦았기 때문이다. 이에 사무국은 심판 판정의 정확도를 평가하겠다는 계획 하에 카메라로 공의 궤적을 촬영해 홈플레이트(home plate) 통과 위치를 분석하는 장비를 도입했다. 그러나 이것은 격렬한 논란을 촉발시켰다.

[그림 1] 야구공의 실시간 위치 및 움직임 추적이 가능한 퀘스텍 시스템<sup>1)</sup>



'야구가 순수성을 잃었다'는 개탄이 일었고, 심판 노조의 반발도 거셌다. 일부 투수들은 퀘스텍 시스템 때문에 심판 판정이 너무 박해졌다고 불평했다. 김병현의 동료이자 앞 시즌에서 23승을 거뒀던 실링의 '난폭한' 의사 표현도 그 연장선상이었던 것이다.

반대로 '기계가 스트라이크 판정을 해야 한다'는 주장도 있었다. 하지만 당시의 기술로는 불가능했다. 측정의 정확도에 대해 논하기 전에, '실시간 처리'조차 불가능했던 것이다. 경기가 끝난 뒤에야 데이터를 처리하여 통계자료를 만들 수 있었던 것이다.

2007년에는 PITCH f/x(투구추적시스템)가 도입됐다. 스포츠 방송용 FX 솔루션을 만들던 스포츠 비전(Sports Vision)이 스트라이크존에서의 볼의 궤적을 정확하게 분석해내기 위한 목적으로 개발했던 시스템이다. 3대의 고해상도 광학카메라로 공의 움직임을 촬영해서 3차원 좌표를 계산한다. 스트라이크존의 통과 위치뿐 아니라 릴리즈 포인트(release point), 피치 무브먼트(pitch movement) 정보까지 실시간으로 수집할 수 있다.

글 | 신동윤 dyonshin@gmail.com 데이터는 신비로운 마법도, 절대적 진리도 아닙니다. 그 대신 "당신 야구 얼마나 해봤는데?"라고 묻지도 않습니다. 그래서 편견 없는 소통의 언어이며 협력의 플랫폼이 될 수 있습니다. 20년쯤 게임 산업 분야에서 일했고 거의 비슷한 시간 동안 한양대학교에서 게임디자인과 관련한 강의를 했습니다. 지금은 타구/투구 추적 데이터에 기반한 분석과 스포츠 미디어 개발을 하고 있습니다.

<sup>1</sup>참고 | <http://baseball-in-play.com/333>

2015년 메이저리그 공식 플랫폼으로 채택된 스탯캐스트(StatCast)에는 더 발전된 기술이 사용된다. 투수가 던진 150km/h의 속구는 홈플레이트까지 0.405초 만에 도달하는데 광학카메라 기반 추적 시스템은 30FPS일 때 12컷, 60FPS일 때 24컷의 이미지로 투구 궤적 전체를 유추한다. 반면 도플러레이더(Doppler radar) 기반의 스탯캐스트는 초당 2만 개의 데이터를 수집한다.

스트라이크존 통과 위치를 실시간으로 측정하는 것은 충분히 가능해졌다. 사용하는 기술에 따라 다소 차이가 있지만 0.2초를 넘기지 않는다. 원한다면 공이 타자에게 도착하기 전에 알아낼 수도 있다. 투수의 손끝을 떠난 공이 0.1초쯤 비행을 하면 추적레이더는 공의 속도와 회전을 읽어낸다. 통제할 수 없는 변수는 갑자기 부는 바람인데, 야구공의 질량과 크기를 고려해본다면 그 영향은 미미하다. 기압, 습도, 온도, 질량 그리고 마찰력과 같은 다양한 요인을 고려해 설정해놓은 물리 시뮬레이션에 넣으면 예상 도착 지점의 오차는 거의 없다. 몇백만 원짜리 컴퓨팅 파워로도 공이 실제 도착하는 시간보다 더 빠르게 예상 도착 지점을 계산해낸다. 정확도도 충분하다. 평가 방법에 따라 이견은 있지만 오차범위가 1인치라면 정확도는 95% 혹은 99% 이상이다.

그렇다면 스트라이크, 볼 판정에 한해서 로봇 심판 도입은 현실적으로 가능했을까? 아직은 오작동과 같이 여전히 해결해야 할 문제가 남아 있다. 어떤 공에 대해 완전히 영동한 측정값을 내놓거나 아예 인식을 못할 수 있다. 물론 이는 극히 희박한 경우이다. 또한 연습구를 구분하지 못하거나 2루수나 중견수의 홈송구를 투구로 인식하는 문제 역시 여전히 남아 있다.

물론 사람은 더 자주 실수한다. 그러나 한두 개 빠진 공을 '인간 심판'이 잘못 판정해도 경기는 문제없이 진행된다. 이것은 인간이 실수하는 방식이다. 로봇 심판은 '인식 실패'를 할 경우 판정을 내리지 않고 멈춘다. 브룩스 베이스볼(Brooks Baseball)의 운영자 댄 브룩스(Dan Brooks)는 이렇게 표현한다. "심판도 공을 놓치지 않나요? 물론이죠. 하지만 공을 놓쳐도 투수가 공을 던졌다는 사실만큼은 알고 있죠."

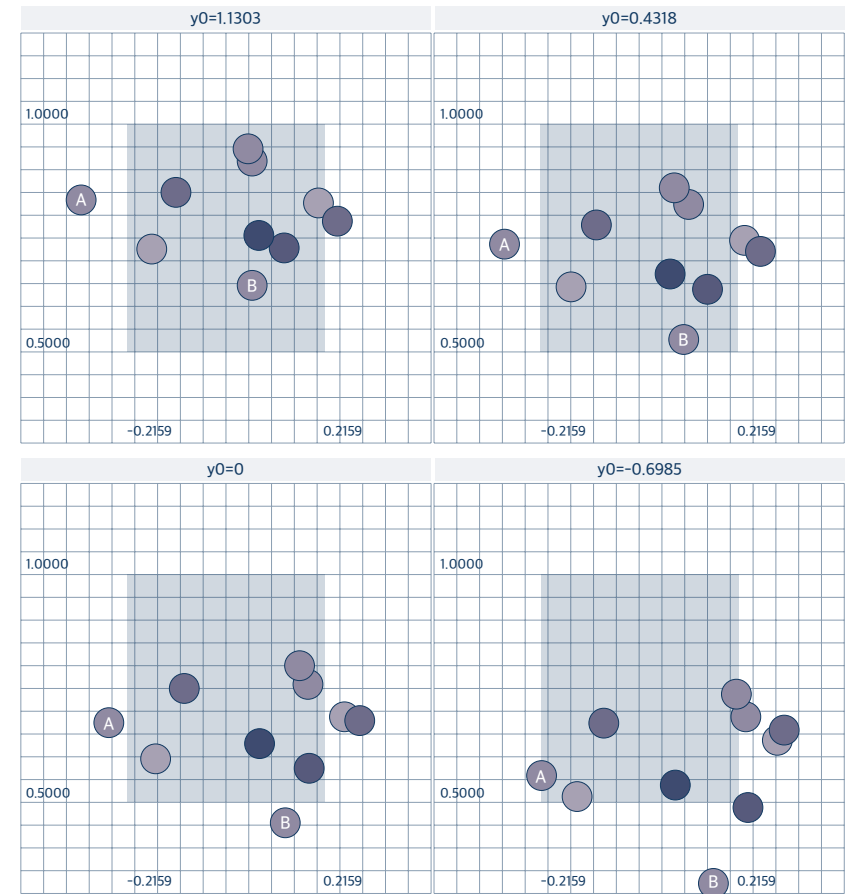
스트라이크존의 세로가 가변적인 것도 어려움이다. 야구 규칙은 '유니폼의 어깨 윗부분과 바지 윗부분 중간의 수평선을 상한선으로 하고 무릎 아랫부분을 하한선'으로 정의한다. 또 '스트라이크존은 타자의 스탠스에 따라 결정된다'고 한다. 지금의 추적 시스템은 타자의 신장 정보를 이용한 알고리즘으로 존의 위아래를 판단하기도 하지만 타격폼이 타자마다 다르다. 결국 사람이 눈으로 보면서 값을 보정해야 한다. 물론 이런 것들은 기술적 난제일 뿐 현실의 로봇 심판 도입에 치명적 장애는 아니다. 완전한 무인 자동화 시스템이 아니라 사람 오퍼레이터(operator)의 도움을 받으면 되기 때문이다. 진짜 문제는 다른 것에 있다.

한번 생각해보자. 중계 화면에 표시되는 그래픽은 평면이다. 하지만 스트라이크존은 '홈플레이트 상공'에 해당하는 입체적인 공간이다. 걸치고 나가는 공과 걸치고 들어오는 공이 있다. 물론 더 진보된 기술을 사용하고 컴퓨팅 파워를 늘리면 실시간으로 궤적 전체를 추적할 수도 있다. 그렇다면 존 앞부분에 1cm 정도 걸치면 스트라이크일까? 혹은 5cm 이상은 걸쳐야 스트라이크일까?

로봇 심판은 공의 위치를 거의 정확히 알아낼 수 있다. 하지만 그 위치가 스트라이크인지 아닌지 판단하는 것은 좀 다른 문제다. 여기서 규범적이고 철학적인 문제가 출현한다. 바로 스트라이크가 무엇인지 명확하지 않다는 것이다.

다음은 5월 29일 사직구장에서 진행된 KBO(Korea Baseball Organization) 리그에서 롯데 투수 브룩스 레일리(Brooks Raley)가 던진 공의 스트라이크 투구 로케이션(location) 모습이다.

[그림 2] 측정 위치에 따른 Pitch Location 차이



\* 우타자 타석. 좌상-우상-좌하-우하 순서로 배터리박스 앞 - 홈플레이트 앞 - 홈플레이트 끝 - 배터리박스 끝에서 측정한 위치다. 회색 존의 가로폭은 홈플레이트 너비와 같다.

투수가 던진 공은 배터리박스(batter's box)를 지나며 우리가 생각하는 것보다 궤적이 훨씬 많이 변화한다. A는 커브 공이다. 좌투수가 던지는 공은 우타자 기준 먼 곳에서부터 가까운 쪽으로 휘어져 들어온다. 이 공은 배터리박스 라인의 앞부분에는 한참 못 미치지만 홈플레이트 뒤쪽의 존에는 닿게 된다. '걸치고 들어온' 스트라이크다. 역시 커브 공인 B는 배터리박스 앞에서 존 한가운데를 통과하지만 배터리박스 끝에서는 거의 타자 발목 근처까지 32cm만큼 낙하한다. 홈플레이트 앞부분과 뒷부분을 통과하는 동안에도 8cm 낙하한다. 이런 공이 포수 미트(catcher's mitt) 위치에 오게 되면 거의 땅바닥에 닿게 되고 스트라이크를 판정하는 데 논란거리가 된다.

야구 규칙에 어떤 공을 스트라이크로 봐야 하는지 명시되어 있기는 하다. 하지만 존에 걸치는 공에 대해서는 명확하지 않다. 홈플레이트 좌우로 걸치는 경우는 [그림 2] 형식으로 걸치고 나가거나 걸치고 들어오면 스트라이크라고 해석한다. 하지만 어느 정도 수준으로 걸쳐야 스트라이크인지는 명확하지 않다. 위아래로 걸치는 공에 대해서는 언급조차 없다. 슬로 커브(slow curve)라면 홈플레이트 상공을 통과하는 도중에 15cm 이상 높이 차이가 생길 수 있다. 프로 레벨의 투수가 작성하고 던진다면 이보다 훨씬 더 큰 낙차를 만들어 걸칠 수도 있다. 타자의 시선은 투수 쪽을 향하기 때문에 머리 위로 지나간 공이 스트라이크 판정을 받았다고 느낄 수 있다.

KBO 리그 데이터 일부를 샘플링해보면 히팅 중 31%는 홈플레이트 앞쪽 면보다 투수에 더 가까운 쪽에서 걸처진다. 홈플레이트 뒷부분 위치보다 포수한테 더 가깝게 걸처지는 타구는 전체에서 13%에 불과하다. 그러니 이렇게 걸치고 들어오는 공을 타자가 치는 것은 아주 어렵다. 대부분의 심판은 이런 공은 '볼'로 판정한다. 우리가 아는 야구에서 이런 공은 스트라이크가 아니다. 일종의 불문(관습)법이다.

야구 규칙은 다른 어떤 스포츠보다 방대하고 세밀하다. 그럼에도 가장 기본이 되는 스트라이크 판정에 대해서는 어느 정도 성문법이 아닌 불문(관습)법을 따르고 있다. 사람이 판정하는 게임에서는 전혀 문제될 것이 없다. 하지만 로봇 심판에게는 난처한 일이다. 요구사항에 남겨진 빈칸은 프로그래머에게는 악몽이다.

머신러닝은 도전할 가치가 있는 해법이다. 실제 경기에서 수집된 투구 궤적과 판정 결과를 학습시키면 로봇 심판은 불문(관습)법을 상당한 수준까지 익히게 될 것이다. 특징(feature)이 매우 명확한 지도학습 모델이다. 데이터가 충분히 쌓일 때까지 기다리면 된다. 하지만 이 학습에서 로봇 심판은 인간 심판의 오류와 편향도 함께 배우게 된다. 예를 들어 인간 심판의 경우 대부분 3볼이 되면 존이 넓어지고 2스트라이크가 되면 존이 좁아진다. 심리학자들은 이를 '부작위 편향'으로 설명한다. 로봇 심판도 비슷한 성향을 가지게 될 것이다. 기계학습 기반의 판정 시스템은 확률에 기반한다. 그렇다면 오심도 생긴다. 하지만 로봇 심판이 이처럼 인간 심판과 비슷하게 작동한다면 로봇 심판이 왜 필요했던 것일까?

지금의 기술은 공의 위치를 꽤 잘 인식할 수 있지만 어디서부터 어디까지를 스트라이크로 판정해야 할지는 여전히 남아 있는 숙제이다. 이런 이유로 인공지능 또는 판정 알고리즘의 도전은 이제부터 진짜 시작이다. 통계 분석에 활용한다면 미미한 예외적 상황이나 오차는 버리면 그만이다. 하지만 공 하나에 승패가 갈리는 경기장이라면 이것은 골치 아픈 일이다. 연구 성과로만 봤을 때 98%의 정확도로 작동하는 판단 시스템은 연구성과로는 충분할지 몰라도 한 경기 내 250개 판정 중 5개가 오심이라면 야구팬들 사이에서 '조작 시비'가 생긴다.

실험과 도전은 물론 진행 중이다. 미국 마이너리그 일부 경기에서 투구추적시스템을 이용한 로봇 심판을 시험 운영한 적도 있다. 로봇 심판에 대한 가장 현실적인 아이디어는 심판이 고글 형태의 웨어러블 디바이스(wearable device)를 착용하고 자신의 눈으로 본 것과 투구추적시스템으로부터 제공받은 정보를 종합해서 판단하는 하이브리드 방식이다.

만약 이것보다 더 멀리 가고 싶다면 다음부터는 기술이 아닌 다른 종류의 문제가 된다. 우리는 100년 이상 야구란 게임을 즐겨왔고, 10조 원 규모의 거대한 비즈니스가 되었지만, 굳이 하지 않아도 되었던 질문인 "야구에서 스트라이크란 과연 무엇일까?"에 답해야 한다. 지금 시점에서 뜬금없는 질문이기는 하다. 하지만 모든 스포츠와 기술 사회가 본래 그런 게 아닐까? 테크놀로지가 의미 있는 변화를 이끌어내려 할 때 가장 중요하고 어려운 일은 '기술적 해법을 찾는 것에 앞서 문제 자체를 제대로 정의'하는 일이다.



〈로봇심판과 판정 알고리즘의 의미 있는 도전〉 브런치로 연결되는 QR 코드입니다.

# 스마트 도핑 검사의 도입과 발전 과제

최근 국내의 스포츠의 핵심 키워드를 뽑으라고 하면 바로 도핑(Doping)이다. 특히, 러시아의 도핑 파문은 2018 평창 동계올림픽과 동계패럴림픽(장애인올림픽)의 큰 이슈였다. 러시아는 2014 소치 동계올림픽 당시 자국 선수들의 샘플을 조직적으로 조작한 정황이 드러나며, 세계도핑방지기구(World Anti-Doping Agency, WADA)와 국제육상경기연맹(International Association of Athletics Federations, IAAF) 등 여러 단체로부터 아직까지 제재를 받고 있다. 이런 가운데 평창 동계올림픽과 패럴림픽이 막을 내린 직후인 지난 3월 21일부터 23일까지(현지시간) 3일간 '올림픽의 수도' 스위스 로잔에서 900여 명의 관계자가 참석한 가운데 '깨끗한 스포츠의 미래'를 주제로 제14회 세계도핑방지기구 심포지엄이 열렸다. 소치 동계올림픽 도핑 파문 이후 전 세계의 스포츠 도핑 흐름을 한눈에 볼 수 있는 자리였다. 이 중에서 가장 눈에 띄는 대목은 인공지능(AI)을 활용한 스마트한 도핑 검사 예고였다.

세계도핑방지기구는 인공지능을 활용해 수상한 선수들을 확인하고 검사 대상을 더욱 효율적으로 파악할 수 있기를 기대하고 있다. 또한, 지속적으로 수집하고 있는 방대한 양의 데이터를 분석하면 더욱 정확한 조사 대상을 선정해 적절한 도핑 검사를 받을 수 있을 것으로 예측한다.

## 도핑이란?

도핑이란 선수가 하나 또는 다수의 도핑방지규정을 위반하는 경우를 말하며, 선수 및 선수 관계자는 무엇이 도핑방지규정 위반 행위인지 그리고 금지목록에 포함된 약물과 방법은 무엇인지 알아야 할 책임과 의무를 가진다. 대표적인 도핑 규정 위반으로는 스포츠에서 운동능력 향상을 위해 금지된 약물을 복용하는 것이 있다.

[표 1] 도핑 규정 위반 유형<sup>1)</sup>

연번	규정 위반 내용
금지약물의 검출	선수의 소변 또는 혈액 샘플에서 금지약물, 대사물질(metabolites), 표시물(markers)이 검출될 경우
금지약물/방법의 사용	금지된 도핑 약물 및 도핑 방법의 사용 또는 사용 시도가 적발된 경우
샘플 채취 거부	사전 공지가 있는 후에 타당한 이유 없이 소변 또는 혈액 샘플 채취를 거부하거나 불응할 경우
소재지 정보 미제출 & 검사 불참	8개월 내에 선수의 소재지 정보 미제출 또는 검사 불응 횟수가 총 3회가 될 경우 도핑방지규정 위반으로 간주
도핑 관리 과정의 특정 부분에 대한 부정행위 및 시도	부적절한 목적이나 방법으로 도핑 검사 과정을 방해하거나 결과를 조작할 경우
금지약물의 소지	금지약물을 소지할 경우
금지약물의 부정거래 및 시도	금지약물 또는 방법의 부정거래 및 거래를 시도할 경우
금지약물의 투여	선수에게 금지약물 투여 또는 투여를 시도할 경우
공모	도핑방지규정의 위반 및 위반 시도에 대해 협조, 조장, 교사, 은폐 행위를 할 경우
금지된 연루	그밖에 선수 및 선수 관계자가 금지된 행위에 연루될 경우

<sup>1)</sup> 참고 | 《국제스포츠지식》 2017-7호  
<http://sport-strategy.org/wp-content/uploads/2017/12/%EC%A7%91%EC%A4%91%EC%A1%B0%EA%B0%90-2017-7%ED%98%B81.pdf>

글 | 박주희 joohee.park@sport-strategy.org 이화여자대학교 체육과학대학에서 학사, 석사를 마치고, 경희대학교에서 스포츠 의과학 박사학위를 취득했습니다. 국내 1호 국제 도핑검사관으로 국내뿐 아니라 올림픽, 아시안게임을 포함한 다양한 국제대회 도핑관리 경험을 가지고 있습니다. OCA(아시아올림픽평의회), WADA(세계도핑방지기구)에서 활약했으며, 현재 국제크리쉬연맹 의무-반도핑위원장으로 활동 중입니다. (재)국제스포츠전략위원회 사무국장으로 근무하며 빠르게 변화하는 국제 스포츠의 흐름에 발빠르게 대처하고자 노력하는 국제 반도핑 전문가입니다.

근육강화제, 심장흥분제 등 선수의 경기력 향상에 직접적인 영향을 주는 PEDs(Performance Enhancing Drugs)부터 이뇨제(소변 배출을 촉진시켜 약물이 신체에서 빨리 빠져나가도록 유도)와 은폐제, 장비 도핑까지 다양한 방법들이 불법적으로 경기력을 향상시키기 위해 사용되고 있다. 도핑은 공정성이라는 스포츠의 핵심 가치를 훼손하고, 이로 인해 동료 선수들에게 피해를 끼치기 때문에 전 세계적으로 엄격하게 금지하고 있다. 이를 위해 모든 스포츠 대회들에서 도핑 검사가 요구되고 있으며, 언제 어디서나 선수들의 혈액과 소변 샘플에 금지약물 성분이 있는지 다양한 분석과 검사가 이루어지고 있다.

**세계도핑방지기구(World Anti-Doping Agency, WADA)<sup>\*2</sup>**

세계도핑방지기구는 1999년 스위스 사법(Swiss Private Law)에 의해 설립되었으며, 현재 캐나다 몬트리올에 본부를 두고 있다. 국제올림픽위원회(International Olympic Committee, IOC)가 공식적으로 인정하는 독립 조직으로서 ‘전 세계 모든 선수들이 도핑 없는 환경에서 경쟁하는 것(A world where all athletes can compete in a doping-free sporting environment)’을 비전으로 두고 도핑 방지 커뮤니티(Anti-Doping Community)와 함께 도핑과의 전쟁에 앞장서고 있다.

세계도핑방지기구는 처음 설립되었을 당시 검사에만 집중하였으나, 효과가 충분치 않다고 판단하여 현재는 도핑 방지에 관한 연구 및 교육 시행, 도핑 방지 역량(capacities)의 개발, 세계도핑방지규정(World Anti-Doping Code) 준수 감시, 금지약물 목록 제정 등 도핑 방지와 관련된 모든 영역에서 활동을 하고 있다. 또한 세계도핑방지기구는 매년 엄격한 심사를 통해 실험실 인증을 실시하며, 인증된 실험실에서 승인된 방법으로만 도핑 분석을 시행할 수 있다. 세계도핑방지기구는 IOC와 각국 정부의 재원으로 예산을 편성하여 활동하고 있으며, 스포츠 대표와 정부 대표들로 구성된 이사회 승인을 거쳐 각종 정책들을 시행하고 있다. 대한민국은 1999년 세계도핑방지기구 창설 당시부터 2006년까지 이사국 회원으로 활동하였으며, 2013년에 재선출되어 2019년까지 활동할 예정이다.

[ 표 2 ] 도핑 방지 커뮤니티(Anti-Doping Community)<sup>\*3</sup>

연번	구분	역할
01	WADA	도핑 방지의 최고 권위 조직이며, 모든 도핑 방지 활동은 세계도핑방지규정(World Anti-Doping Code)에 의거함
02	IOC, IPC, IFs	· IOC, IPC(국제패럴림픽위원회): 올림픽과 패럴림픽에서의 검사, 결과 관리, 위반 제재에 대한 책임을 가짐 · IFs(국제연맹): 각 IF 주관 경기에서의 검사 및 경기 외 검사 실시, 교육 프로그램 시행, 도핑방지 규정 위반 제재 역할을 담당함
03	Governments	도핑 검사가 가능하도록 하는 역할, 국가의 검사 프로그램 지원, 도핑에 연루된 자에 대한 재정적 지원 보류(withholding), 불법 물질의 제조 및 매매에 대한 제재, 도핑방지 교육 및 연구에 대한 재정을 지원함

\*2 참고 | <https://www.wada-ama.org>  
 \*3 참고 | 《국제스포츠지식》 2018-1호 <http://sport-strategy.org/wp-content/uploads/2018/05/%EC%A7%91%EC%A4%91%EC%A1%B0%EA%B0%90-2018-1%ED%98%B8.pdf>

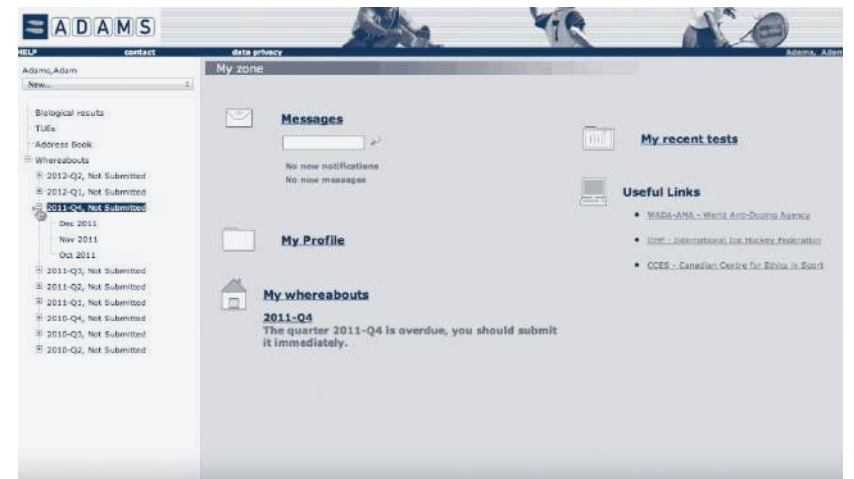
04	NOCs, NPCs, NFs	NOC와 NPC가 도핑 방지 코드를 준수하도록 IOC와 IPC 각각 관련 규정을 제정해야 하며, IF 규정은 NF의 '세계도핑방지규정 준수'를 반드시 포함해야 함
05	NADOs, RADOs	자국의 선수들과 자국에서의 경기 또는 훈련에 참가하는 외국 선수들에 대한 경기, 경기 외 검사의 실시, 도핑방지규정의 위반에 대한 판결, 도핑 방지 교육 등에 대한 책임을 가짐
06	Athletes	도핑방지규정의 모든 조항은 국제 및 국내 수준의 모든 선수들에게 적용되어야 함
07	CAS	스포츠 관련 논쟁들의 중재와 조정을 위한 독립 조직으로, WADA는 도핑방지규정의 효력이 미치는 범위 내에서 도핑과 관련된 사안들에 대해 CAS에 제소할 권리를 가짐
08	LABs	도핑 검사 실험실들은 반드시 국제표준(International Standard)에 의거하여 WADA의 인가를 받아야 함
09	ITA	· 러시아 도핑 파문 이후 IOC 집행위원회의 승인을 받아 설립된 독립적인 국제검사기구 · 스포츠 단체, 국가 이의 단체로부터 독립적으로 운영되는 기구이며 IF나 MEO(Major Event Organization)에게 도핑 검사 서비스를 제공함

**ADAMS, 그리고 WADA 페이퍼리스(Paperless) 정책**

세계도핑방지기구는 모든 도핑 커뮤니티 관계자들이 도핑방지규정을 준수하고, 정보들을 효율적으로 활용할 수 있도록 지난 2005년 ADAMS(Anti-Doping Administration & Management System, 도핑방지행정관리시스템)를 개발하여 도입하였다.

ADAMS는 도핑 관리 절차들을 하나의 시스템에 담고 있으며, 관계자들은 자신들에게 부여된 권한을 가지고 필요한 정보를 언제든지 열람하여 확인할 수 있다. 현재 176개 국가들과 대부분의 스포츠 단체들이 사용하고 있는 ADAMS는 다음과 같은 기능들을 가지고 있다.

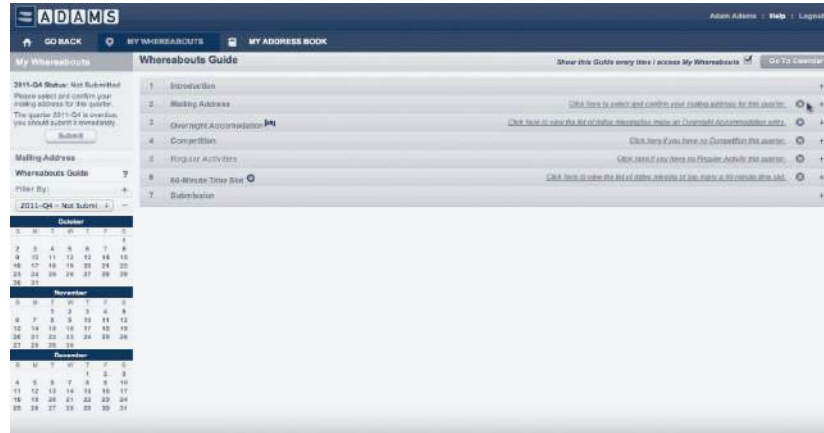
[ 그림 1 ] ADAMS 화면<sup>\*2</sup>



### 1) 선수 소재지(Athlete Whereabouts)

선수들은 불시에 도핑 검사의 대상이 된다. 따라서 ADAMS에 자신의 소재지를 정확하게 기입해야 하는 의무를 가지고 있다. 도핑 방지 기구들은 ADAMS를 통해 선수의 소재지를 간편하게 확인할 수 있으며, 이를 기반으로 불시에 방문 검사를 진행할 수 있다. 소재지 정보를 기입하지 않거나 허위로 기입한 선수는 제재의 대상이 된다.

[ 그림 2 ] ADAMS 선수 소재지 등록 시스템\*4



### 2) 치료 목적 사용 면책(Therapeutic Use Exemptions, TUE) 관리

선수는 치료 목적으로 허가받은 약물을 복용하는 것은 가능하다. 선수는 ADAMS에 TUE를 신청하여 허가를 받을 수 있고, ADAMS는 이와 관련된 모든 관계자들에게 해당 정보를 제공한다. 모든 절차는 한 번의 신청으로 완료될 수 있을 정도로 간편하며, 선수는 ADAMS를 통해 TUE 증명서를 발급받을 수도 있다.

### 3) 도핑 검사 계획 & 결과 관리(Test Planning & Results Management)

도핑 관리 관계자들은 ADAMS를 통해 도핑 검사 계획과 일정 등을 조율할 수 있으며 결과 관리 또한 가능하다. ADAMS로 인해 검사기관과 실험실 등의 협력이 가능하고 중복 검사를 방지할 수 있어 효율적인 도핑 관리가 이루어질 수 있다.

### 4) 실험실 결과 관리 모듈(Lab Results Module)

세계도핑방지기구의 인증을 받은 실험실들은 ADAMS를 통해 선수들의 샘플 분석 결과를 세계도핑방지기구를 포함한 관련 단체들에게 전달한다. 이를 통해 단체들은 간편하게 정확한 결과를 받을 수 있으며, 세계도핑방지기구는 도핑 관리 과정에서의 투명성을 확보할 수 있다.

또한, 세계도핑방지기구는 도핑 관리에 요구되는 문서들을 시스템화하여 간편화시키는 이른바 페이퍼리스(Paperless) 운영체제로 전환하고 있다. 도핑 수법이 더욱 교묘해지고 복잡해지면서

도핑 검사 항목 또한 증가하였고, 선수 한 명당 방대한 양의 자료가 생성되었다. 이에 따라 최신 기기들을 활용하여 종이 없이 간편하게 시스템에 등록하고 관리하도록 권장하고 있다. 세계도핑방지기구는 페이퍼리스 도입으로 인해 방대한 분석 자료들의 시스템화를 추진하고, 도핑 방지 커뮤니티의 모든 관계자들이 더욱 효율적으로 정보들을 활용할 수 있도록 지원하고 있다.

### 한계

하지만 전 세계에서 도핑 관리를 효과적으로 하기 위해 ADAMS 시스템을 도입한 이후로도 여러 문제점들이 제기되었다. ADAMS 내에 등록되는 정보의 양은 지속적으로 증가하고 있으나 세계도핑방지기구의 정보부(Information & Intelligence Team)가 실제로 조사할 수 있는 양은 한정되어 있다. ADAMS로 인해 절차는 간편화되었지만 이로 인해 더욱 많은 정보와 자료들이 생성되었고, 이들의 진위 여부를 판단하기 위해서는 많은 인력과 재정이 소비되어 효과적인 도핑 방지 활동이 이루어지지 않는다는 것이다. 이에 대한 가장 큰 이유는 세계도핑방지기구의 인력과 재정 부족이다. 세계도핑방지기구는 자체적인 수익사업이 없고 IOC와 각 정부들로부터 받는 기여금으로 예산을 편성하기 때문에, ADAMS에 등록된 방대한 양의 정보들을 분석할 자원이 부족하다.

또 다른 문제로 ADAMS 시스템 내에 허위 정보를 기입했을 때 절차상의 문제만 없다면 이를 쉽게 적발할 수 없다는 문제점이 드러났다. 실제로 '2014 소치 동계올림픽' 당시 러시아의 조직적인 도핑 샘플 조작은 당시 시스템상으로는 별 다른 문제가 나타나지 않았지만, 독립 감시단(Independent Observers)의 조사와 내부고발자의 증언으로 발각되었다는 점은 전 세계에 충격으로 다가왔다. 이와 더불어 몇몇 국가들과 선수들 사이에서 아직도 행해지는 도핑의 실태는 페이퍼리스 시스템 구축만으로는 도핑을 효율적으로 적발하기 힘들다는 점을 보여주고 있다. 이에 따라 세계도핑방지기구 내외부에서는 ADAMS에 등록된 수많은 정보들을 더욱 효율적으로 모니터링할 방법의 필요성이 제기되었다.

### 보안책 - 인공지능

세계도핑방지기구는 올해 3월 이러한 문제점에 대한 해결책 중 하나로 인공지능 시스템을 도입하자고 주장하였다. 도핑 방법이 더욱 교묘해지면서 적발하기가 더 어렵게 발전하였기 때문에 도핑방지 방법도 이에 따라 발전해야 한다는 것이다.

세계도핑방지기구 사무국장 올리비에 니글리(Olivier Niggi)가 발표한 내용에 따르면, 현재 세계도핑방지기구의 가장 큰 고민은 방대한 양의 선수 데이터를 제시시간 안에 분석하지 못한다는 것이다. 그런데 인공지능을 활용하면 더욱 빠르게 도핑이 의심되는 선수를 찾아내 추가 조사를 진행할 수 있다. 물론 인공지능이 도핑 의혹 선수에 대한 유무죄 판단을 내릴 수는 없지만, 인공지능을 활용한 새로운 시스템이나 알고리즘을 개발한다면, 데이터를 활용해

\*4 참고 | KADA 홈페이지, <https://www.kada-ad.or.kr/page/100e01>

세계도핑방지기구의 검사 및 조사를 더욱 효율적으로 진행할 수 있다. 특히 인공지능을 활용하면 의심되는 선수들에 대한 집중 검사가 이뤄질 수 있을 것이고, 방대한 데이터도 더욱 효율적이고 정확하게 분석하여 스마트한 도핑 검사가 기대된다. 올리비에 니글리 사무국장은 가까운 미래에 인공지능을 시범 운용할 것이며, 이에 기반을 둔 시스템을 구축할 것이라고 발표하였다.

물론 인공지능 시스템이 도입되기 위해서는 많은 개발과 재정이 필요하다. 보다 효율적인 시스템을 구축하기 위해서는 이에 요구되는 예산을 확보하는 것이 중요하다. 세계도핑방지기구는 현재 각국 정부들과 IOC로부터 자금을 조달함으로써 예산을 추가로 확보하여 도핑 방지 프로그램 개선에 활용한다는 계획을 가지고 있으며, 앞으로 인공지능을 활용한 스마트하고 효율적인 도핑 관리로 발전할 전망이다.

더불어 올 초 핀란드도핑방지위원회 주관으로 핀란드 헬싱키에서 세계도핑방지기구를 비롯한 각국 도핑 방지 관련 인사들이 모여 네트워크 및 정보수집 강화와 관련된 논의를 진행했다. 11개 국가도핑방지위원회(National Anti-Doping Organizations, NADO)와 2개 국제경기종목연맹(International Federation, IF), 인터폴(Interpol), 세계도핑방지기구 정보수사(Intelligence and Investigations, I&I) 부서 대표들이 참석한 본 회의에는 최신의 정보수사 혁신 기술과 세계도핑방지기구 집행위 및 이사회에 보고될 도핑방지 정보수사 네트워크(Anti-Doping Intelligence and Investigations Network, ADIIN) 운영계획 등이 논의되었다. 도핑방지 정보수사 네트워크는 경험, 전문성, 경영 및 운영 데이터를 안전하게 공유할 수 있는 환경을 제공하는 네트워크라고 한다.

2018 평창 동계올림픽의 성공을 기점으로 명실상부 국제 스포츠 강국으로 자리매김한 대한민국 역시 급변하는 전 세계 스포츠의 흐름에 발맞추며 도핑 방지와 관련해서도 다양한 시도와 준비가 필요한 시점이다.

\*참고문헌

1. <https://inews.co.uk/sport/wada-artificial-intelligence-doping-cheats>
2. <https://www.insidethegames.biz/articles/1063094/exclusive-wada-considering-name-change-and-broader-anti-drugs-focus-to-generate-more-funding>



〈스마트 도핑검사의 도입과 발전 과제〉 브런치로 연결되는 QR 코드입니다.

# 스포츠 저널리즘의 변화와 AI의 활용

글 | 김동환 [dkim@lab2ai.com](mailto:dkim@lab2ai.com) 학부와 석사에서 컴퓨터 공학과 HCI(Human-Computer Interaction)를 전공한 후 서울대학교 언론정보학과에서 로봇 저널리즘 연구로 박사학위를 취득했습니다. 현재는 랩투아이(Lab2AI)에서 공동 대표이사를 맡아 스포츠 분야에서의 인공지능 솔루션 및 서비스를 개발하고 있습니다.

알고리즘이 사람의 손을 거치지 않고 기사 작성에 필요한 모든 역할을 자동으로 수행하는 형태의 언론 보도를 로봇 저널리즘(robot journalism)이라고 부른다.<sup>\*1</sup> 데이터 속에서 수학적 최적화 과정과 기계 학습 등을 통해 상대적으로 의미 있는 이야기거리를 찾아내고 자연스러운 문장으로 만드는 이 기술은 현재 국내외의 다양한 언론매체에서 적극적으로 활용되고 있다. 이 글을 통해 스포츠 저널리즘 분야에서 알고리즘 솔루션이 동작하는 방식과 적용 사례, 그리고 가능성에 대해 살펴보고자 한다.

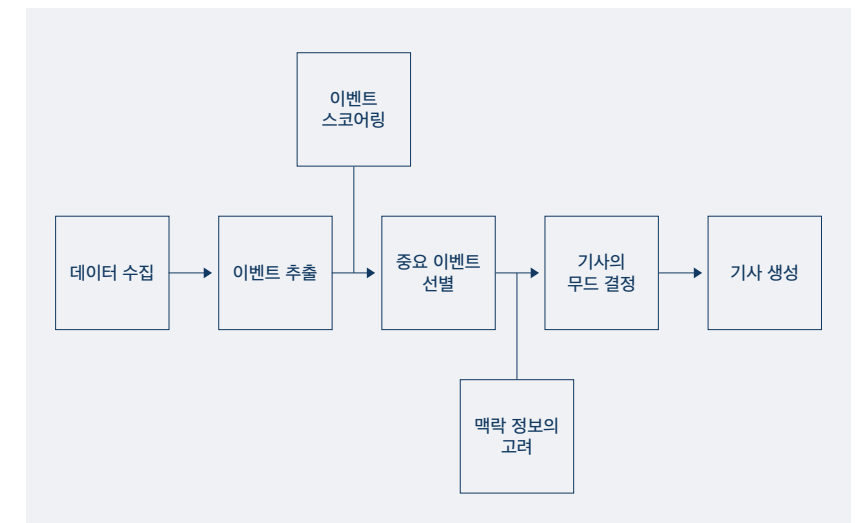
## 기사를 쓰는 알고리즘의 등장

2014년 3월 17일 오전 6시 25분, 미국의 메이저 신문사인 로스앤젤레스 타임스(Los Angeles Times)에 LA 웨스트우드(Westwood)에서 8km 떨어진 곳에서 발생한 진도 4.7의 지진 기사가 실렸다.<sup>\*2</sup> 이 기사가 모두의 관심을 끈 이유는 기사 말미에 '이 정보는 미국 지질조사국(United States Geological Survey, USGS)의 지진 경보 서비스를 통해 왔고, 이 포스트는 기자가 만든 알고리즘에 의해 생성되었다'라는 문구 때문이었다.

기자이자 프로그래머인 켄 슈웬키(Ken Schwencke)는 지진 경보 이메일을 받을 경우 지진의 강도에 따라 뉴스 가치를 판단해 기사를 자동으로 작성하는 알고리즘을 만들었고, 미국 캘리포니아 남서부에서 발생한 이 지진을 다룬 기사로 인해 알고리즘이 기사 작성에 관계된 모든 역할을 자동으로 수행하는 언론 보도인 로봇 저널리즘에 대한 관심이 폭발적으로 증가했다.

서울대학교 이준환 교수와 필자는 지난 2015년 3월 프로야구 기사를 작성하는 로봇 저널리즘 솔루션을 국내 최초로 개발하며 관련 연구의 시작을 알렸다.<sup>\*3</sup> 로봇 저널리즘 기술이 스포츠 분야에 처음 적용된 이유는 데이터가 풍부하게 축적돼 수학적 연산과 분석 모델을 구축하기가 용이하기 때문이었다. [그림 1]은 로봇 저널리즘 솔루션의 단계별 동작 방식이다.

[그림 1] 자동 기사 생성을 위한 로봇 저널리즘 시스템 및 방법<sup>\*4</sup>



\*1 논문 | 김동환·이준환(2015), <로봇 저널리즘: 알고리즘을 통한 스포츠 기사 자동 생성에 관한 연구>, 《한국언론학보》 59(5), pp.64-95  
 \*2 참고 | [http://www.slate.com/blogs/future\\_tense/2014/03/17/quakebot\\_los\\_angeles\\_times\\_robot\\_journalist\\_writes\\_article\\_on\\_la\\_earthquake.html](http://www.slate.com/blogs/future_tense/2014/03/17/quakebot_los_angeles_times_robot_journalist_writes_article_on_la_earthquake.html)  
 \*3 참고 | 프로야구 뉴스로봇. <https://www.facebook.com/kbaseballbot>  
 \*4 특허 | 이준환 & 김동환(2017), 자동 기사 생성을 위한 로봇 저널리즘 시스템 및 방법(Robot Journalism Method and System for Automatic Article Generation), pg17. <https://patentimages.storage.googleapis.com/00/a2/84/80a56dbffbf31/KR20170058785A.pdf>



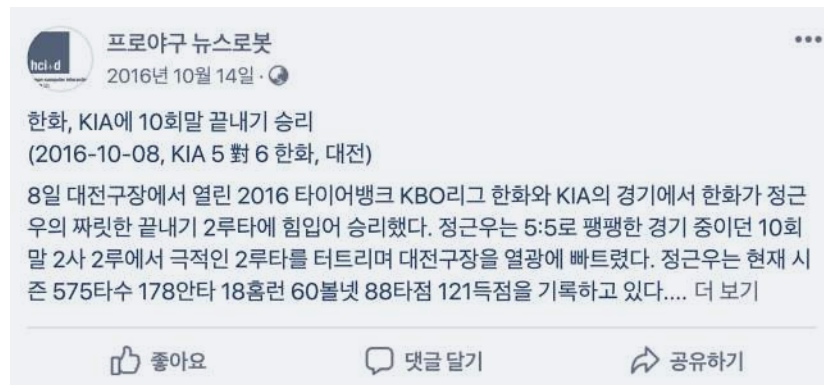
## 무엇이 기사가 되는가?

로봇 저널리즘 솔루션의 동작 방식을 간단히 살펴보면, 일차적으로 수집된 데이터 속에서 알고리즘이 의미 있는 정보를 찾아낸다. 예를 들어 프로야구에서는 볼카운트, 타격 결과, 구종과 구질 등 경기에서 발생하는 실시간 데이터는 물론이고 타자와 투수의 상대 전적, 선수의 통산 기록과 특이사항 등 공식적이고 통시적인 데이터를 모두 고려해 이벤트(사건의 단위)를 추출할 수 있다.

여기서 쓰이는 알고리즘은 이런 이벤트 중에서 ‘무엇이 기사가 될 수 있는가’에 대한 가치판단을 스스로 하도록 설계되었다. 기존 기록의 갱신, 일반 범주에서 벗어나는 기록, 특정 이벤트의 연속 발생 등 뉴스가 될 만한 사건을 찾기 위해 수학적 최적화(numerical optimization) 과정과 기계 학습(machine learning)을 통한 스코어(score) 보정이 반복적으로 수행된다.

중요한 이벤트를 선별한 후에는 일관된 표현을 통해 사건을 전달할 수 있도록 글의 관점을 정하는 과정이 뒤따른다. 연장 끝에 끝내기 안타로 경기를 승리한 경우 ‘짜릿한’ 안타가 ‘극적으로’ 터진 상황을 전달하고, 결승타를 기록한 선수와 팀의 승리를 보다 극적으로 표현한다면 [그림 2]와 같이 더 자연스러운 기사를 작성할 수 있다.

[그림 2] 로봇 저널리즘을 통해 작성한 기사 예시<sup>3</sup>



컬럼비아대학교(Columbia University) 토센터(Tow Center)의 보고서에 의하면

AP통신, 포브스(Forbes), 뉴욕타임스(New York Times), 프로퍼블리카(ProPublica),

워싱턴포스트(Washington Post) 등 해외의 선도적인 언론매체들이 속도 경쟁에서 우위를 점하고 단순한 일과를 대체하는 비용 절감 측면에서 로봇 저널리즘 기술을 직접 개발해 뉴스 기사 작성에 적극적으로 나서고 있다.<sup>5</sup> 최근에는 파이낸셜뉴스의 로봇 기자 lamFNBOT이나 연합뉴스의 사커봇 등 국내 언론에서도 이런 솔루션을 직접 개발하거나 도입하는 사례가 늘어나는 추세이다.

## KBO 퓨처스리그 케이봇(KBOT) 로봇 기자

스타 플레이어들이 대거 출전하는 KBO 리그에 비해 2군 구장에서 벌어지는 KBO 퓨처스리그 경기는 미디어의 외면 속에 상대적으로 주목받지 못한다. 이에 KBO는 퓨처스리그의 저변 확대와 활성화를 위해 인공지능 솔루션인 케이봇(KBOT)을 도입해 2018년 7월부터 퓨처스리그 기사를 자동으로 발행하기 시작했다.

랩투아이(Lab2AI)는 지금까지 축적된 로봇 저널리즘 개발 경험과 기술을 활용해 케이봇을 개발하며 크게 다음의 두 가지 사항을 중점적으로 고려했다.

첫 번째 목표는 다양한 경기의 양상에 따라 중요한 상황을 판단할 수 있는 자동화된 알고리즘 솔루션의 개발이었다. 이를 위해 경기 흐름과 주요 기록에 따라 데이터의 상대적인 중요도를 판별하는 복합적인 가중치 매트릭스(matrix)를 만들어 활용했다. 독자들이 주로 관심을 가지는 득점 장면, 홈런, 투수의 승/패/세이브 등과 같이 야구 기사에서 자주 등장하는 장면들과 연속 기록 혹은 진기록의 발생 여부와 같이 상대적으로 중요한 사건을 구분해 수학적 최적화 과정에 대입했다. 이러 이유로 케이봇은 매 경기에서 주목할 만한 사건을 놓치지 않고 찾아내고 있다.

두 번째 목표는 1년에 576번 발행되는 기사 속에서 같은 문구가 반복적으로 등장하지 않도록 자연스럽게 다양하게 문장을 작성하는 자연어 생성 기술을 구현하는 일이었다. 케이봇은 사람이 작성한 문장을 기계적으로 학습하는 개체명 인식 알고리즘과 구문을 분석하는 자연어 처리 알고리즘이 적용되었다. 분석된 문장은 알고리즘을 통한 생성에 활용될 수 있게 템플릿화되고, 전산화된 경기 기록에 따라 상황에 최적화된 템플릿을 선택하는 모델을 구축했다.

케이봇이 본격적으로 KBO 퓨처스리그 기사를 작성하면서부터 그동안 주목받지 못했던 퓨처스리그 경기가 지금은 KBO 홈페이지와 포털사이트, 각종 SNS에 연재되고 있고 사람들의 관심 역시 증가하고 있다. 매일 벌어지는 퓨처스리그 경기를 기자가 직접 취재한다면 지금보다 2배의 인력이 필요하겠지만, 알고리즘은 이 과정을 효율적이고 효과적으로 대체하며 성공적으로 퓨처스리그의 미디어화를 이끌어내고 있다.

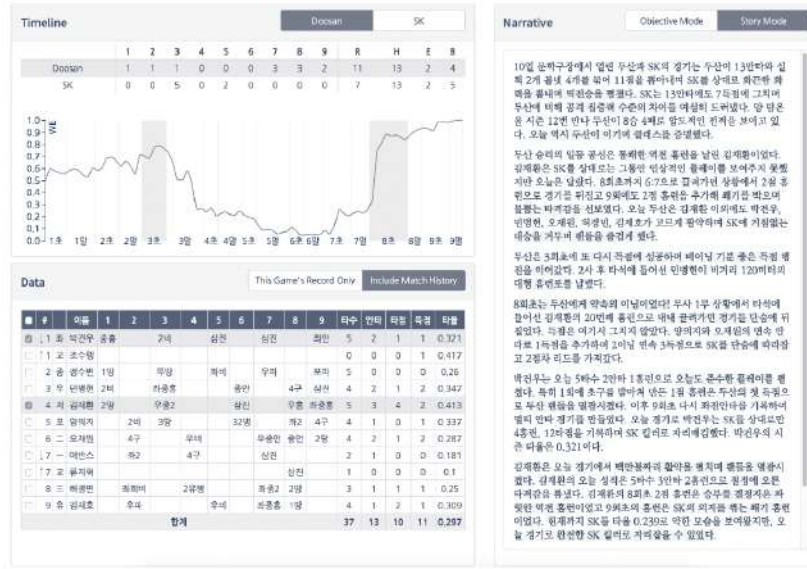
## 새로운 가능성의 발견, 뉴스의 개인화(personalization)

필자는 연구진과 함께 2017년 한 해 동안 6만여 건 이상의 사회인 야구 경기를 기사로 작성하는 프로젝트를 진행한 바 있다. 사회인 야구 경기는 매주 1천여 번 벌어지는데, 이 경기를 직접 취재하며 글을 쓰는 기자는 없다. 경기 수를 감당하기도 힘들뿐더러 관심을 갖는 독자층도 존재하지 않아 기사를 발행할 수도 없기 때문이다. 하지만 사회인 야구 리그에 참여 중인 선수 본인에게는 자신과 팀의 활약이 담긴 기사는 높은 가치를 지닌다.

이와 같이 전통적인 매스미디어(mass media)의 개념에서는 불가능하지만 주기적으로 생성되는 데이터를 해석해 이야기를 만들어내는 알고리즘의 동작 방식은 모두가 보는 하나의 기사가 아니라 개인을 위한 맞춤형 기사를 만드는 일에도 상당히 적합하다.

<sup>5</sup> 논문 | Andreas Graefe. (2016). Guide to Automated Journalism, Columbia University Academic Commons, <https://doi.org/10.7916/D80G3XDJ>

[ 그림 3 ] 독자가 직접 원하는 내용을 선택해 기사를 만드는 인터랙티브 뉴스 서비스 예시\*6



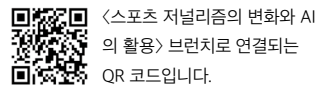
[그림 3]은 독자가 직접 인터페이스(interface)를 조작해 개인이 관심을 가진 팀과 경기 상황, 선수를 선택해 기사를 만드는 인터랙티브 뉴스(interactive news) 서비스의 예시 화면이다.\*6 실험 설문조사를 진행한 결과, 이와 같은 방식으로 기사를 제공했을 때 더 재미있고 흥미로운 기사로 느껴진다는 참가자들의 반응이 있었다.

**마치며**

알고리즘이 반드시 인간과 같은 방식으로 기사를 작성할 필요는 없다. 초기 인공지능 기자의 개발 방향은 인간을 재매개(remediation)하는 ‘기자처럼 행동하는’ 로봇 기자의 개발이었지만, 신문에 실리던 상보 기사 형태를 답습하는 방식보다 개인화된 인터페이스를 통한 맞춤형 기사의 생성이나 학습에 기반한 질의응답형 챗봇 및 AI 스피커와의 인터랙션(interaction)도 다변화된 매체 환경에 적합한 좋은 대안이 될 수 있다.

이처럼 빠르게 변화 중인 저널리즘의 지형에서 인공지능 솔루션은 보다 다양한 차원에서 스포츠 저널리즘에 깊이를 더해줄 수 있을 것으로 기대한다.

\*6 논문 | Kim, D., & Lee, J. (2018). Designing an Algorithm-Driven Text Generation System for Personalized and Interactive News Reading. International Journal of Human-Computer Interaction, pp.1-14



# 만능 스포츠봇의 등장과 발전 방향

프리미어리그인 영국 프로축구나 프로야구를 좋아하는 사용자들 사이에 스포츠봇이 화제다. 카카오톡에서 '프리미어리그봇', '프로야구봇' 플러스친구를 친구로 추가하면 자신이 좋아하는 팀의 각종 경기 일정과 기록, 하이라이트 영상 등을 손쉽게 찾아볼 수 있다. 알림 설정을 해주면 깜박 잊어버리기 쉬운 내가 응원하는 팀의 경기날 라인업이나 생중계, 경기 결과를 챗봇이 알려준다. 또한, 팬들끼리 서로 메시지를 주고받을 수 있어 같은 팀을 응원하는 팬들과 훈훈한 응원의 메시지를 보내거나 라이벌팀 팬들과 묘한 신경전도 벌일 수 있는 잔재미도 있다.

프리미어리그봇은 1월, 프로야구봇은 3월에 오픈하였고, 처음에는 팬 카페나 커뮤니티를 중심으로 입소문을 타며 사용자가 늘어났다. 현재, 프로야구봇의 경우 8만 명이 넘는 사용자가 하루에 42만 회 이상 서비스를 사용하고 있다.

[그림 1] 프로야구봇이 제공하는 기능<sup>1)</sup>

1. 간편하게 경기정보를 받아볼 수 있어요. 경기 일정, 결과, 기록, 순위, 하이라이트 등 경기관련 정보를 볼 수 있어요.

2. 선수의 시즌 및 경기별 기록을 제공해요. 공금한 선수의 이름과 타이틀 기록을 보여주세요.

3. 팬들과 메시지를 주고받아요. 팬들과 메시지를 주고 받을 수 있고, 메시지에 좋아요, 싫어요 반응을 보낼 수 있어요.

4. 알림을 기준으로 설정해요. 알림을 설정하는 정보에 따르면 알림을 받을 수 있고, 알림을 받지 않는 정보도 선택할 수 있어요.

5. '정보수령은 '실정' 막힐때는 '말을 잘 입력하세요'를 보여주세요. '정보수령은 '실정' 막힐때는 '말을 잘 입력하세요'를 보여주세요.

프로야구봇의 경우 사용자의 90% 이상이 경기 알림을 받고 있고, 매일 30%가 넘는 사용자가 자발적으로 챗봇 대화방을 찾아와 서비스를 사용한다. 서비스 초반의 반짝 효과에 그치지 않고 오픈 초기부터 지금까지 서비스 사용 지표는 날로 상승하고 있다. 여러 사용자들로부터 “좋은 서비스를 만들어주어 고맙다”라는 황송한 피드백을 받고 있으나, 반대로 이렇게 적극적으로 사용해주셔서 오히려 감사할 따름이다.

[그림 2] 프로야구봇 사용자 피드백 중

“ 날랜에게 '프로야구 시즌 시작했어?'라고 하자, 이미 알고 보고 있었으며 기록이 프로야구봇 이런 특별한 서비스가 있다며 자랑을, 인벤프로이만.”	“ 프로야구봇 초마다. 날짜를 넣으면 지-나-간 날짜 하이라이트도 볼 수 있는데 그 훨씬 규칙으로 3분야구도 지원하면 금상첨화겠습!”
“ 요즘 제일 실례는 카톡 ㅋㅋㅋㅋ #프로야구봇 #야구”	“ 세상에 이렇게 좋은걸 지금 말았더니... 네이버 스포츠 가기도 귀찮았는데 인벤 카톡에서 다 해결되는구만! 좋아좋아!”
“ 카카오톡에 프로야구봇 신기하네 ㅋㅋㅋㅋ 이제 카톡하면서 야구 볼 수 있네!”	“ 카톡 프로야구봇 아저 신기하다... 으윳! 되게 편하네 ㅋㅋㅋㅋㅋㅋㅋㅋ”
“ 나 오늘 놀러가서 프로야구봇 처음 써봤는데 *판하고 좋더라... 나만 아저 오늘 알았나봐...!!”	“ 프로야구봇 추가해놓으니깐 편하네... 없어서 척척하...”

### 스포츠봇 탄생의 배경

사실대로 말하자면, 필자가 소속된 미래미디어파트에게 주어진 미션은 스포츠봇을 만드는 것이 아니라, 카카오톡의 챗봇 플랫폼인 카카오톡 오픈빌더(open builder)를 사용해서 레퍼런스로 삼을 만한 서비스를 만드는 것이었다. 2016년 페이스북(Facebook)에서 메신저 API(application programming interface)를 공개한 이후 수많은 챗봇 서비스가 개발되었고, 그중 몇몇 챗봇이 인기를 얻었지만 누구나 쉽게 떠올릴 수 있는 레퍼런스가 될 만한 챗봇은 거의 없었다. 처음 한두

글 | 유승원 cree.yoo@kakaocorp.com 10년 넘게 미디어 관련 서비스 개발을 하고 있습니다. 수많은 채널과 콘텐츠 속에서 사용자에게 가치 있는 콘텐츠를 더 잘 그러모아 사회의 다양한 면을 제대로 조명하고, 더 좋은 세상을 만드는 데 일조하는 서비스를 만들어가고 싶습니다.

<sup>1)</sup> 참고 | <https://brunch.co.kr/@openbuilder/1>

달은 어떤 붓을 어떤 콘셉트로 만들 것인가에 대해 고민하였다. 우리는 단순히 기존의 서비스를 챗봇으로 전환하는 것이 아니라, 왜 챗봇이어야만 하는가에 대한 명확한 근거가 필요했다.

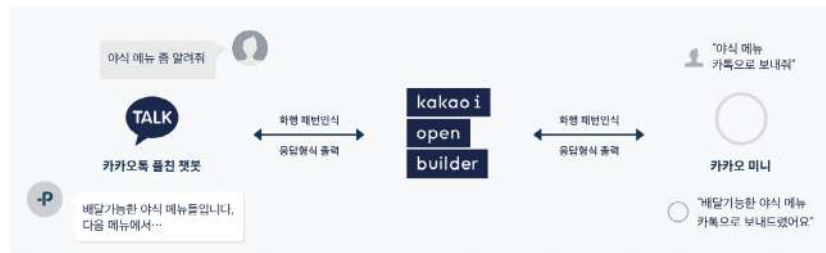
우리가 주목한 점은 챗봇은 하위 뎁스(depth)로 제공되는 정보를 대상으로 해야 한다는 것이다. 사람들이 '리버풀 경기 일정'을 확인하기 위해서는 포털사이트에서 검색을 하거나 스포츠 섹션 혹은 리그의 공식 홈페이지나 기록 페이지를 찾아보며 경기 일정을 확인해야 한다. 수십 자의 타이핑, 혹은 스크롤과 클릭 등 사용자의 시간과 동작을 줄이는 것이 '편리함'이라고 한다면 이러한 정보를 한 번에 제공할 수 있게 하는 것이 챗봇에 적합한 서비스일 것이다.

그리고 챗봇 플랫폼으로서의 강점인 알림 서비스가 적합하게 쓰일 수 있어야 한다. 푸시 알림은 유용한 마케팅 도구이며 사용자에게 즉시 정보를 전달할 수 있다는 강점이 있지만 잦은 알림은 도리어 서비스 이탈률을 높게 된다. 반면 사람들은 메신저 알림에는 상대적으로 관대한 측면이 있다. 많은 토론 끝에 우리는 스포츠 도메인(domain)을 타깃으로 삼았다. 챗봇 론칭 시점이 시즌 중이라는 점과 팬층이 비교적 젊은 유저(user) 기반이지만 헤비 유저층(heavy user)이 존재하는 '프리미어리그'를 첫 번째로 출시했다.

스포츠봇의 구조

스포츠봇은 카카오 AI를 통합 관리하는 오픈 빌더 플랫폼 위에서 동작한다. 사용자가 특으로 메시지를 전송하면 오픈 빌더는 자연어 처리 및 패턴 인식 등을 통해 붓의 어떤 기능을 실행해야 하는지 판단한다. 이때 오픈 빌더는 붓에 필요한 매개변수인 날짜, 시간, 선수 이름, 기록 종류 등을 인식하고 적절한 형태로 가공하여 붓에 전달한다.

[ 그림 3 ] 카카오 i 오픈 빌더<sup>2</sup>



오픈 빌더는 챗봇과 카카오톡에 공용으로 사용되는 플랫폼이며, 큰 틀에서는 특을 통해 텍스트상으로 명령을 하고 결과를 받는지(챗봇), 스피커를 통해 음성으로 명령을 내리고 결과를 받는지(인공지능 스피커/음성 챗봇)의 차이가 있다고 할 수 있다. 실제로 카카오톡에서 서비스되는 스포츠 콘텐츠도 스포츠봇과 같은 붓 에이전트(bot agent)에서 동작하고 있다.

에이전트가 오픈 빌더에게 데이터를 전달받으면, 카카오 스포츠 서비스에서 사용하는 다양한 API로부터 데이터를 수집한 후 사용자에게 카카오톡의 말풍선 형태로 결과를 가공하여 응답한다. 붓 에이전트는 시나리오에 따라 동시에 여러 API를 호출할 수도 있고, 한 번 호출한

\*2 참고 | <https://i.kakao.com/openbuilder/docs>

API 결과에 따라 다른 API를 선택적으로 호출하기도 한다. 스포츠봇 에이전트 서버는 스프링 프레임워크(Spring Framework)를 사용하는데, 스프링에서 제공하는 리액터(Reactor)의 퍼블리셔(Publisher) 구현체인 Flux/Mono와 관련 연산자(Operator)는 여러 API로부터 돌려받는 데이터들을 비동기적으로 다루면서도 다양한 시나리오에 따라 재조합하거나 변환하는 데 높은 개발 효율성을 제공했다.

[ 그림 4 ] 스포츠봇 시스템 구조



스포츠 경기 알림은 불규칙하게 발생할 수 있다. 프로야구 경기 종료는 정해진 시간이 없고, 경기 취소와 같이 돌발적인 상황에 대한 알림도 존재한다. 이를 주기적으로 확인하기 위해 아파치 에어플로(Apache Airflow) 기반의 스케줄러(scheduler)가 별도로 존재하며, 동시다발적으로 수만에서 수십만 번의 알림이 발생할 수 있기 때문에 메시지 큐(message queue)와 아파치 스톰(Apache Storm) 클러스터(cluster) 구조의 알림 발생기를 사용한다.

챗봇이 다른 서비스 플랫폼과 다른 점은?

챗봇 서비스는 챗봇 플랫폼 위에서 동작하기 때문에 서비스 개발에 편리한 점이 많다. 오픈 빌더는 사용자와의 인터페이스(interface)를 '스마트'하게 관리해주며 다양한 부가 기능(시나리오 관리, 매개변수 관리, 이벤트 관리 등)을 제공한다. 챗봇 서비스 개발자는 온전히 어떤 서비스를 어떤 맥락으로 사용자에게 제공할 것인지에 집중하면 되기 때문에 서비스 개발 사이클(cycle)이 웹, 혹은 앱 개발에 비해 단축될 수 있다.

한편 스포츠봇은 다양한 종류의 스포츠 데이터를 메뉴 계층의 구별 없이 사용자의 직접적인 호출에 따라 제공할 수 있다. 사용자는 우리가 생각했던 수준 이상으로 많은 데이터를 원했고, 서비스의 디테일한 부분까지 사용했다. 우리가 웹이나 앱에서 서비스를 기획할 때 기획 의도나 가독성, 개발 효율성 등을 고려해서 누락시키는 데이터들도 챗봇에서는 모두 흠뻑려 놓을 수 있으며, 사용자는 언제나 비슷한 정도의 수고만으로 자신이 원하는 데이터에 접근하게 된다. 많은 사용자들이 프로야구 1군 콜업(call up) 정보나 축구의 팀 이적 정보를 원했고, 우리는 가지고 있는 데이터가 부족함에 아쉬움을 느껴야 했다. 또한, 쉽게 다양한 기능을 추가할 수 있음에도

챗봇의 기반 앱이라고 할 수 있는 카카오톡을 더 무겁게 만든다거나 사용자가 원하지 않는 부담을 주지 않는다는 점도 챗봇의 주요 장점이라 할 수 있다.

[ 그림 5 ] 프로야구봇 서비스 화면



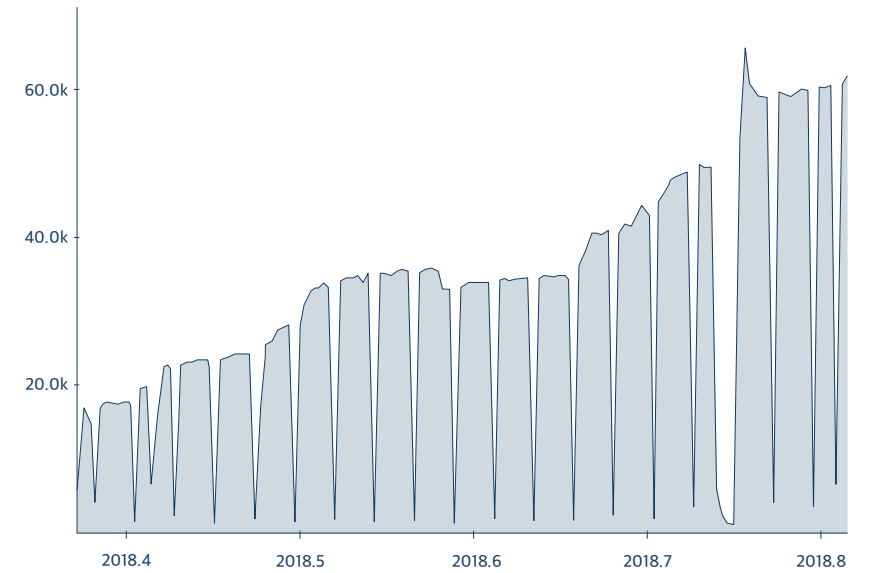
UX(user experience) 측면에서 챗봇은 일반 플랫폼과는 다른 고민을 하게 한다. 먼저 챗봇에서 표현할 수 있는 내용이 카카오톡에서만 가능한 것으로 제한되어 있다. 아직까지는 흔히 사용하는 표도, 애니메이션도, 다양한 형태의 이미지도, 심지어 한 번에 표현할 수 있는 글자 수까지도 제한적이다. 이러한 한계를 조금이라도 극복하기 위해 스포츠봇의 경우는 텍스트와 이모지(emoji)를 최대한 조합하여 결과물을 표현했다.

또 다른 한 가지는 서비스의 흐름이 상하 스크롤로 정해져 있다는 것이다. 챗봇에서 웹뷰(Webview)를 띄우거나 다른 앱으로 전환시킬 수 있겠지만, 챗봇을 벗어나지 않는 범위로 한정한다면 챗봇에서는 화면이 전환되거나, 탭이 있거나, 팝업이 뜨지 않는다. 서비스를 기획할 때 이를 소홀히 생각한다면, 한 번에 너무 많은 데이터를 배치하면서 사용자가 현재 위치를 잃어버리거나 불필요한 스크롤을 하게 되는 현상을 겪을 수도 있다.

스포츠봇의 매우 인상적인 지표 중 하나는 알림 사용자 비율이 90%가 넘는다는 점이다. 챗봇이 새로운 채널로서 자리하기 위해서는 푸시 알림이 매우 중요했지만, 사용자가 플랫폼에 익숙해지기 전 챗봇 알림에 귀찮음을 느끼고 서비스를 이탈하는 것을 막기 위해 처음부터 보수적인 접근법을 사용했다. 특히 한국시간으로 밤/새벽에 경기가 열리는 프리미어리그는 사용자들이 가장 중요하게 생각하는 라인업 정보만 알림으로 제공했는데, 오픈 한 달 후 알림 사용자가 전체 사용자의 84%에 달하는 긍정적인 신호를 얻었다. 이것을 기반으로 프로야구봇은 사용자들이 원하는 알림 종류를 라인업, 중계 시작, 경기 기록, 하이라이트, 경기 취소 등으로 세분화해 제공했다. 프로야구는 상대적으로 프리미어리그에 비해 경기시간대가 이르지만, 거의 매일 있기에 알림이 더 잦을 수밖에 없음에도 하나 이상 알림을 받는 사용자 비율이 90%가 넘는다는 것은 매우 고무적인 현상이다. 이것으로 미루어 알림의 횟수보다는 사용자가 원하는 정보를 원하는 시점에 전달하는 것이 가장 중요하다는 것을 알 수 있다. 여기에 한 단계 더

나아가 프로야구봇은 '특종계'라는 새로운 개념의 서비스를 제공하는데, 특종계는 경기 상황을 아웃카운트(out count)별로 요약하여 실시간으로 사용자에게 알려준다. 대략 2시간 30분 동안 50번이 넘는 알림을 받게 되는데, 예상보다 많은 사람들이 이 서비스를 이용하는 것을 알 수 있었다.

[ 그림 6 ] 프로야구봇 순방문자



[ 그림 7 ] 프로야구봇에서 알림을 하나 이상 받는 비율



### 스포츠봇은 앞으로 어떻게 발전할 것인가?

AI 기술의 발전에 따라 사람들의 말을 더욱 잘 이해하고, 드러나지 않은 의도까지 파악하는 스마트한 봇의 등장을 생각할 수 있다. 챗봇끼리 서로 연동하여 사용자가 야구를 보는 동안 치킨을 주문해주고, 다양한 봇의 기능을 하나로 묶은 비서봇으로 발전할 수도 있다. 단톡방에 스포츠봇을 초대해서 여러 사람들과 선수나 경기에 대한 상황을 분석하고 토론할지도 모른다. 그러나 가장 중요한 본질은 편리함을 끝까지, 집요하게 추구해야 한다는 점이다. 사용자들이 스포츠봇을 선호하는 이유는 스포츠봇이 '똑똑해서', '스마트해서'가 아니라 '편해서', '편리해서' 좋아한다는 사실을 주지하자.



〈만능 스포츠봇의 등장과 발전  
방향〉 브랜치로 연결되는 QR  
코드입니다.

# Tech Insider

Apache S2Graph 기반 머신러닝 모델 환경 구축 <small>AI in Kakao</small>	108
딥러닝이 탐구하지 못한 언어와 5가지 태스크	116
2018 ICML을 통해 살펴보는 AI 연구 동향	126
2018 CVPR 논문 동향 및 주요 연구 소개	138
카카오 크루들의 커피 주문 방식	146

# Apache S2Graph 기반 머신러닝 모델 환경 구축

서비스에 머신러닝(machine learning)을 적용하기 위해서는 데이터 수집, 데이터 전처리, 모델 생성, 서비스에 API 제공 등 각각의 쉽지 않은 일들을 잘 처리해야 합니다.

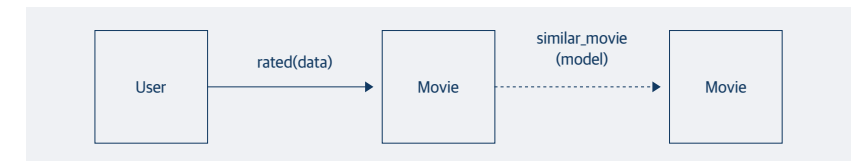
지금까지 좋은 모델을 생성하기 위해서는 어떤 알고리즘을 활용해야 하는지에 대해서는 많이 공유되었습니다. 하지만 모델을 만드는 것만큼 중요하다고 여겨지는, 만들어진 모델을 서비스에 잘 적용하고 쉽게 실험해볼 수 있는 환경을 구축하는 문제에 대한 공유는 많지 않았습니다.

이 글에서는 누구나 쉽게 따라 해볼 수 있도록 실제 데이터가 아닌, 모두에게 공개된 MovieLens(무비렌즈) 데이터 세트를 사용하여 영화 추천이라는 친숙한 도메인 예제를 통해 그래프 데이터베이스를 활용한 머신러닝 모델(machine learning model)과 데이터를 하나로 추상화하는 과정을 살펴보도록 하겠습니다. 이 글의 모든 내용은 깃허브(GitHub)의 MovieLens 데이터 튜토리얼<sup>1</sup>을 통해 직접 실행해볼 수 있습니다.

## The abstraction(개요)

MovieLens 데이터 세트를 S2Graph의 기반 모델인 프로퍼티 그래프(Property Graph) 모델로 모델링해보면 다음과 같습니다.

[ 그림 1 ] MovieLens 데이터 세트의 프로퍼티 그래프 모델링



그리고 지금부터는 MovieLens 데이터 세트의 Vertex, Edge의 표현 방식을 자세하게 살펴보도록 하겠습니다.

### 1) Vertex: 데이터에 존재하는 노드(node)들을 표현합니다.

(1.1) Movie: 사용자들이 MovieLens 데이터 세트에서 평점을 부여한 영화들

#### (1.1) 예시

movielid, title, genres

- 1, Toy Story(1995), Adventure|Animation|Children|Comedy|Fantasy
- 2, Jumanji(1995), Adventure|Children|Fantasy
- 3, Grumpier Old Men(1995), Comedy|Romance

...

글 | 윤도영 shon.0@kakaocorp.com 카카오에서 데이터 엔지니어로 일하고 있고, 데이터 파이프 라인과 머신러닝에 관심이 많은 개발자입니다. 현재 Apache S2Graph(incubating) 프로젝트를 리드하고 있고, 카카오의 그래프 DB팀에서 일하고 있습니다.

<sup>1</sup>참고 | <https://github.com/apache/incubator-s2graph/tree/master/example/movielens>



(1.2) User: 영화에 평점을 부여한 사용자들

```
(1.2) 예시
userId
1
2
3
...
```

2) Edge: Vertex와 Vertex 간의 관계를 표현합니다.

(2.1) Rated: 사용자들이 영화에 부여한 평점

```
(2.1) 예시
userId, movied, rating, timestamp
1, 31, 2.5, 1260759144
1, 1029, 3.0, 1260759179
1, 1061, 3.0, 1260759182
...
```

(2.2) similar\_movie: 비슷한 영화들 간의 관계를 표현

이 글에서 가장 중요한 부분으로 ALS 알고리즘을 통해 만들어진 머신러닝 모델이 저장될 관계를 표현합니다. 다른 관계(Edge)들은 MovieLens 데이터 세트에 존재하는 데이터를 저장소에 저장하고, 저장된 원래의 데이터(raw data)를 조회하는 방식이지만, similar\_movie는 사용자들의 평점 기록을 통해 ALS 알고리즘이 생성한 모델의 예측 결과를 조회하는 방식이라는 차이가 있습니다.

**The technologies stack: 카카오 활용 오픈소스**

다음은 실제 카카오에서 활용되는 오픈소스들입니다. 카카오에는 사용자가 검색한 검색어, 사용자가 본 콘텐츠, 사용자가 맺은 친구 관계 등 다양한 종류의 데이터들이 존재합니다. 많은 경우에 이렇게 모인 데이터들을 다음과 같은 기술 스택(technologies stack)으로 저장하여 모델을 생성하고, 생성된 모델을 서비스에 적용하고 있습니다.

**Apache S2Graph<sup>2</sup>**

카카오에서 개발하고 Apache 소프트웨어 재단에 오픈소스로 제공한 그래프 데이터베이스로, 이 튜토리얼에서는 MovieLens 데이터 세트 전체를 저장하는 저장소로 사용됩니다. S2Graph는 S2GraphQL이라는 REST 인터페이스(REpresentational State Transfer Interface)를 통해 저장된 원래의 데이터뿐 아니라 머신러닝 모델도 조회하는 기능을 제공합니다.

<sup>2</sup> 참고 | <https://s2graph.apache.org>

**Apache Spark<sup>3</sup>**

Apache Spark(아파치 스파크)를 활용해 MovieLens 데이터 세트를 가공하고, MLLib이라는 머신러닝 라이브러리(machine learning library)의 ALS 알고리즘을 활용하여 모델을 생성합니다. 현재까지는 하나의 라이브러리로 Apache Spark을 활용했지만, 다른 머신러닝 라이브러리들도 충분히 활용 가능합니다.

**Annoy4s<sup>4</sup>**

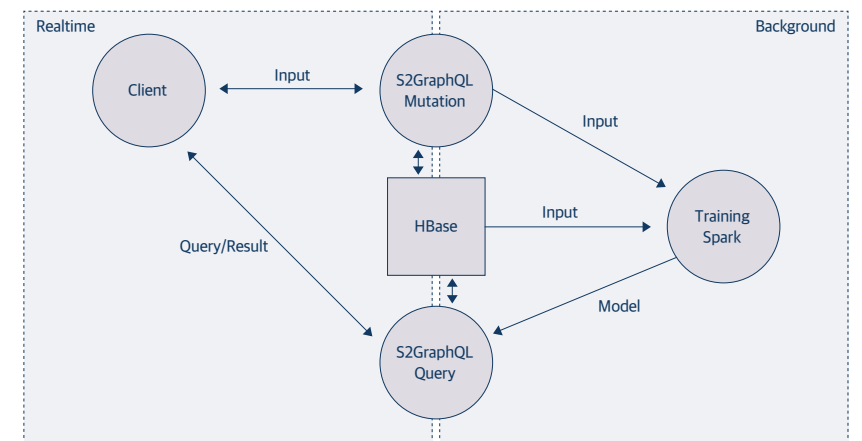
Spark MLLib의 ALS 알고리즘을 통해 만든 모델을 그대로 사용하지 않고, annoy<sup>5</sup>라는 라이브러리를 통해 approximate nearest neighbors(근사 근접 이웃)를 찾는 인덱스(index)를 생성하고, 이 인덱스를 최종적으로 저장소에 저장하여 사용자들이 좋아할 만한 영화를 찾아주는 태스크(task)에 활용합니다.

**The architecture(구조도)**

**전체 프로세스 설명**

[그림2]와 같은 형태로 모델링되는 MovieLens 데이터 세트를 가지고 간단하게 영화 추천 서비스를 구성하는 과정을 설명해보도록 하겠습니다.

[ 그림 2 ] Apache S2Graph의 구조도<sup>6</sup>



이 글에서는 협업 필터링(collaborative filtering)의 한 종류인 ALS 알고리즘을 이용해 사용자들이 매긴 영화 평점 데이터(rated edge) 입력값(input)을 토대로 비슷한 영화를 알려주는 모델을 구축하고, 이를 기반으로 사용자의 반응을 분석하여 선호할 만한 영화들을 추천해주는 평점 기반의 추천 서비스를 소개하려고 합니다. 실제 튜토리얼 정보는 MovieLens 데이터 세트의 디스크립션(description)<sup>7</sup> 부분을 참고하시면 됩니다.

<sup>3</sup> 참고 | <https://spark.apache.org>  
<sup>4</sup> 참고 | <https://github.com/annoy4s/annoy4s>  
<sup>5</sup> 참고 | <https://github.com/spotify/annoy>  
<sup>6</sup> 참고 | <https://github.com/apache/incubator-s2graph/tree/master/example/movielens#the-architecture>  
<sup>7</sup> 참고 | <https://github.com/apache/incubator-s2graph/tree/master/example/movielens#description>

(1) Schema 생성

먼저 사용자의 데이터를 그래프로 모델링하는 과정을 가장 먼저 고민해야 합니다. 대부분 직관적으로 그림을 그리고 관계를 연결하다 보면 자연스럽게 모델링이 되는데, 이 부분이 그래프 데이터베이스의 가장 큰 장점입니다. 실제 데이터에서 존재하는 Vertex를 schema로 생성하고, 이 Vertex들 간의 관계도를 구축하고 나면 데이터를 수집·저장할 준비가 됩니다.

(2) Data import

Schema를 생성한 뒤에는 S2Graph의 API를 활용하여 데이터를 저장소에 저장합니다. S2Graph는 실시간으로 발생하는 사용자 데이터를 뛰어난 성능으로 저장할 수 있는 저장소로, 카카오 내에서 널리 활용하고 있습니다. 특히, S2Graph는 대량으로 작업이 가능한 벌크로드(bulkload) 기능을 제공함으로써 100억 건 이상 되는 대량의 데이터도 한 번에 저장할 수 있게 해줍니다. 이를 통해 물리적으로 분리되어 있는 여러 데이터베이스의 대량의 데이터를 한곳에 그래프 형태로 저장할 수 있습니다. 이 글의 MovieLens 데이터 세트에서는 평점(rated) 데이터가 가장 큰 데이터로 이를 벌크로드 기능을 통해 저장하면 아래와 같은 기본적인 조회가 가능해집니다.

[ 그림 3 ] 1번 사용자가 평점을 매긴 영화 10편(왼쪽), 제목에 Toy를 포함한 영화 5편(오른쪽)

```

**1번 사용자가 평점을 매긴 영화 10편**
query {
  movielens {
    User(id: 1) {
      rated(limit: 10) {
        Movie {
          title
        }
      }
    }
  }
}

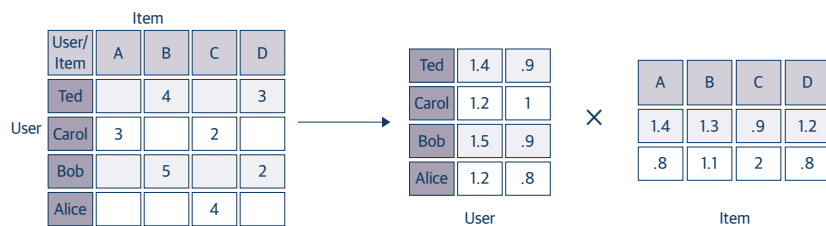
**제목이 Toy를 포함한 영화 5편**
query {
  movielens {
    Movie(search: "title: *Toy*", limit: 5) {
      title
    }
  }
}
    
```

(3) ALS 알고리즘을 활용한 비슷한 영화 모델 생성

모델을 생성하기 위해서는 ALS 알고리즘을 이용하여 사용자들의 평점 데이터를 행렬분해(Matrix Factorization)합니다. 우리의 목표는 비슷한 영화를 찾는 것이기에 행렬분해 결과 중 항목(item)에 해당하는 행렬(matrix)만 annoy 알고리즘의 입력값(input)으로 사용합니다.

다음으로 annoy 알고리즘을 활용하여 근사 근접 이웃 탐색(approximate nearest neighbor search)이 가능한 annoy 지수(index)를 생성합니다.

[ 그림 4 ] 추천 데이터 계산\*8



\*8 참고 | <https://mapr.com/ebooks/spark/08-recommendation-engine-spark.html>

(4) 생성 모델의 schema 등록

위에서 생성한 annoy 인덱스로 S2GraphQL의 update schema API를 통해 schema를 업데이트 합니다. 업데이트가 완료된 후 similar\_movie 관계에 대해 조회를 해보면, 저장된 원래 데이터를 조회하는 대신 annoy 인덱스를 활용해서 3번에서 만든 모델에 기반한 비슷한 영화 목록을 추천해 주도록 시스템이 변경된 것을 볼 수 있습니다.

(5) GraphQL을 통해 데이터 조회 및 API 제공

[그림 5]처럼 2에서 저장된 원 데이터와 4에서 등록된 모델 데이터를 합쳐 데이터를 조회할 수 있습니다. 특히 눈여겨볼 부분은 similar\_movie를 조회하는 부분인데, 데이터를 조회하는 사용자는 해당 데이터가 직접 저장한 원래의 데이터인지, 모델이 다이나믹하게 만들어내는 데이터인지 알지 못하고 사용하게 됩니다.

보통 일반 그래프 데이터베이스는 사용자가 미리 관계를 저장해놓은 데이터를 조회하는 기능만을 지원합니다. 그러나 앞서처럼 모델 서빙(model serving)을 지원하는 S2Graph는 미리 준비한 관련 데이터가 없어도 GraphQL을 통한 완벽한(seamless) 인터페이스로 실시간(dynamic) 유사도 계산과 직관적인 조회를 지원합니다.

[ 그림 5 ] 1번 사용자가 평점을 부여한 10편 영화와 비슷한 5편의 영화(왼쪽), 코미디 장르의 성격을 가지고 제목에 '1995'를 포함하는 5편의 영화(오른쪽)

```

**1번 사용자가 평점을 매긴 영화 10편과 비슷한 영화 5편(item based cf)**
query {
  movielens {
    User(id: 1) {
      rated(limit: 10) {
        Movie {
          title
          similar_movie(limit: 5) {
            Movie {
              title
            }
          }
        }
      }
    }
  }
}

**Comedy가 포함된 장르를 가지고 있고, 제목에 1995를 포함하는 영화들과 비슷한 영화들 5편**
query {
  movielens {
    Movie(search: "genres: *Comedy* AND title: *1995*", limit: 5) {
      title
      genres
      similar_movie(limit: 5) {
        Movie {
          title
          genres
        }
      }
    }
  }
}
    
```

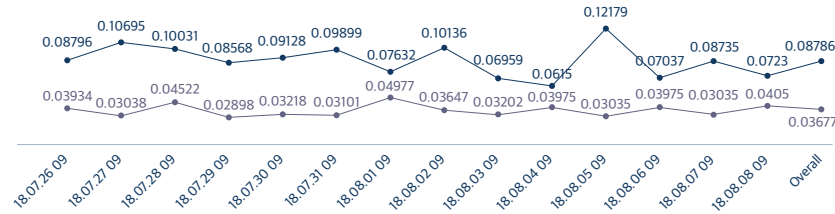
(6) API를 활용한 A/B 실험, 모델 개선

사용자가 좋아할 만한 영화들을 찾는 태스크(task)는 여러 가지 방식(서비스 로직(logic), 알고리즘)으로 시도해볼 수 있는데, 이 시도들 가운데 실제로 사용자들의 반응률이 가장 좋은 방식을 데이터에 기반해 결정할 수 있도록 S2Graph에서는 A/B 테스트 환경을 자체적으로 제공합니다.

인기 있는 상위 10편의 영화를 모든 사용자에게 보여주는 기존 로직 A가 있다고 가정합니다. 그리고 새롭게 ALS 알고리즘을 통해 만든 로직 B를 제공하려고 합니다. 두 개의 로직 중 실제 사용자들의 선호가 높은 방법을 판단하기 위해 5번처럼 로직 B 쿼리(query)를

S2Graph 어드민(admin)에 등록합니다. 쿼리를 등록한 이후에는 각각의 로직에 전체 트래픽 중 몇 %를 할당할지 정할 수 있습니다. 또한 A/B 각각의 쿼리들은 자신의 고유한 ID를 발급받습니다. 실제 사용자가 추천 결과에 노출, 클릭 등의 반응을 하면, 해당 고유 ID의 피드백(feedback)을 클라이언트(client)로부터 S2Graph 서버로 전달받습니다. 이런 과정을 거쳐 다음과 같이 어떤 로직이 효율적이고 좋은 반응을 얻었는지 확인 가능합니다.

[ 그림 6 ] A/B 두 가지 로직을 동시에 실험하여 각 로직별로 기존 대비 효율을 확인하는 예

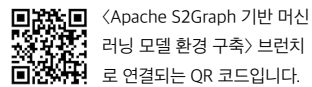


**마치며**

카카오에서는 앞서 설명한 Apache S2Graph라는 그래프 데이터베이스를 활용하여 많은 부문에서 데이터 기반 의사결정과 액션이라는 목표를 달성해오고 있습니다. 단순히 추천이라는 애플리케이션 외에도 검색, 광고, 사용자 타게팅 등 카카오 서비스 내 여러 분야에서 활용되고 있는 Apache S2Graph가 궁금하면 깃허브의 내용을 참조해 주시길 바랍니다.\*9

머신러닝 기술을 카카오 서비스에 적용하는 데이터 파이프라인(data pipeline)을 개발, 운영하면서 몇 가지 느낀 점이 있습니다. 간단한 알고리즘·모델이라도 서비스에 쉽고 빠르게 적용하고, 기존보다 끊임없이 효율을 개선하는 과정이 중요합니다. 그리고 무엇보다 데이터와 모델을 서비스에 쉽게 적용하고 테스트해볼 수 있는 환경을 갖추는 것이 더 나은 서비스를 제공하기 위해 필수적임을 말씀드리며 글을 마무리하겠습니다.

\*9 참고 | <https://github.com/apache/incubator-s2graph>



# 딥러닝이 탐구하지 못한 언어와 5가지 태스크

**글·정리 | 이수경 samantha.lee@kakaobrain.com** 2016년 3월 알파고와 이세돌 9단이 펼치는 세기의 대결을 두 눈으로 목도한 이후 AI를 제대로 공부해봐야겠다고 결심했습니다. AI의 본진이자 연구소인 카카오브레인으로 걸어들어온 이유입니다. 인공지능 기술과 이로 인해 바뀔 미래 사회를 다루는 글을 통해 사람들과 소통하고 싶습니다.

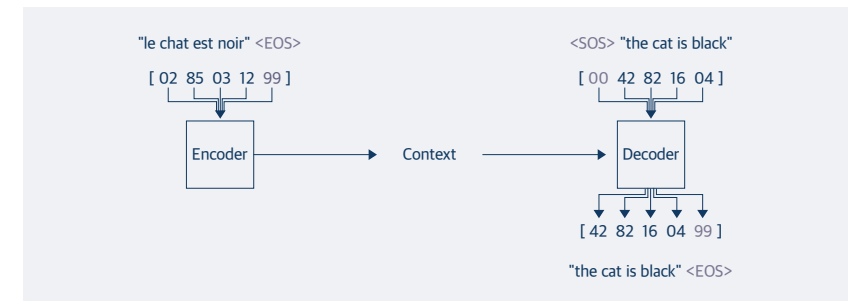
**발표·감수 | 박규병 kyubyong.park@kakaobrain.com** 카카오브레인에서 언어를 재료로 한 딥러닝 연구를 하고 있습니다. 언어와 교육, 생명정보, 환경 등에 관심이 있습니다. 기술을 중시하지만, 기술이 아닌 사람이 중심이 되어야 한다고 믿습니다. <https://github.com/kyubyong>에서 그동안의 연구 활동과 사진을 보실 수 있습니다.

그간 딥러닝(deep learning)은 기계 번역(machine translation), 음성 인식(speech recognition), 음성 합성(speech synthesis) 등 여러 응용 분야에서 기술적으로 큰 발전을 이루었다. 그러나 딥러닝으로 충분히 해결하지 못한 언어 관련 과제들은 여전히 많다. 대화 시스템(dialog systems)이나 챗봇(chatbot)이 대표적인 예다. 상업적인 가치가 큰 것으로 알려진 챗봇의 경우 소비자의 만족을 끌어낼 정도의 기술 수준에는 여전히 도달하지 못했다. 그 외 딥러닝이 아직 충분히 해결하지 못한 언어 과제에는 무엇이 있을까?

## 첫 번째, 비지도 기계 번역(unsupervised machine translation)

NMT(neural machine translation)의 첫 번째 모델인 seq2seq(sequence to sequence) 네트워크<sup>1</sup>는 지도 학습(supervised learning)<sup>2</sup> 방식의 기계 번역에 큰 혁신을 가져왔다. 문장 전체의 모든 정보, 즉 단어 의미와 단어 순서, 문장 구조, 단어 간의 의존 관계 등 번역에 필요한 모든 정보(문장 벡터)를 토대로 문맥을 이해하는 자연스러운 번역 결과를 냈기 때문이다.<sup>3</sup>

[그림 1] 인공신경망 기반의 기계 번역 구조<sup>4</sup>



[그림 1]에서처럼 'le chat est noir'를 입력하면 'the cat is black'이라는 번역 문장을 도출하는 원리는 다음과 같다. 먼저 단위정보(단어(word) 또는 토큰(token)) 시퀀스<sup>5</sup>를 인코더(encoder)에 입력한다. 인코더는 이를 분석해 고정 길이의 벡터 표현(vector representation)을 추정한다.

디코더(decoder)는 이 벡터를 활용해 또 다른 단위정보의 시퀀스를 생성한다.<sup>6</sup>

양질의 대규모 병렬 말뭉치(large parallel corpus)가 필요한 지도 학습에서는 데이터 양이 충분치 않으면 학습효과가 거의 없다.<sup>7</sup> 이처럼 병렬 말뭉치 데이터가 부족한 상황에서는 어떻게 신경망을 훈련해야 할까? 라벨링(labeling)된 데이터와 그렇지 않은 데이터 모두를 훈련에 사용하는 준지도 학습(semi-supervised learning)<sup>8</sup> 방법이 제안됐으나 큰 성과는 없었다.

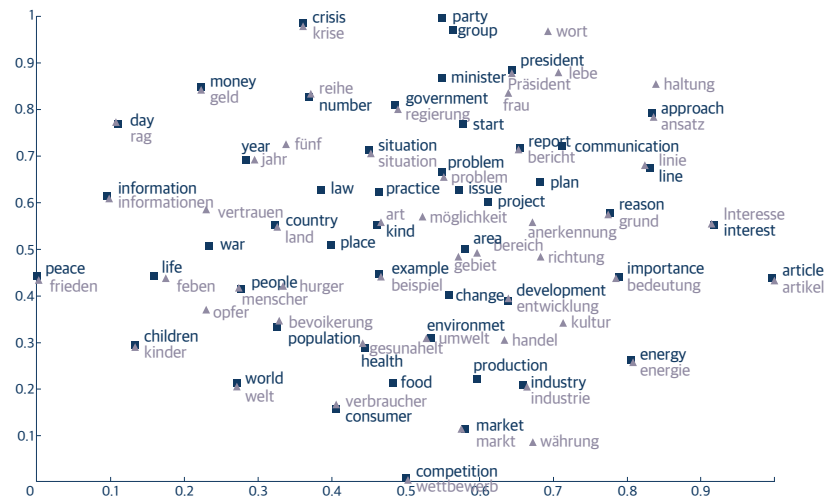
최근에는 비지도 학습 교차언어 임베딩(unsupervised cross-lingual embeddings)이 가시적인 성과를 내고 있다.<sup>9</sup> 훈련 과정은 다음과 같다.

<sup>1</sup> 참고 | <https://datascience.stackexchange.com/questions/25411/what-exactly-is-the-input-of-decoder-in-autoencoder-setup>  
<sup>2</sup> 설명 | 기존 모델이 가지고 있는 입력과 출력 사이 길이 차이가 생기는 문제를 고정된 상태 벡터를 출력으로 내놓게 함으로써 해결함. 물론 가장 단순한 형태의 seq2seq 모델은 만족할 만한 수준의 성능을 내지 못할. 신경망(neural networks)이 가진 잠재력을 확인한 정도라 해도 무방함(신경망 번역 모델의 진화 과정 <https://brunch.co.kr/@kako-it/155>)  
<sup>3</sup> 설명 | 지도 학습: 정답이 라벨링(labeling)된 데이터로 모델을 학습시키는 방식  
<sup>4</sup> 참고 | 일상생활 속으로 들어온 기계 번역. <http://kiss.kstudy.com/public/public2-article.asp?key=50905527>  
<sup>5</sup> 설명 | 시퀀스: 연속된 입력 데이터. 문장(sentences)은 문법과 같은 규칙에 의거한 일련(sequential)의 단어 집합이라는 점에서 시퀀스의 예를 가운데 하나라고 볼 수 있음  
<sup>6</sup> 참고 | Sequence to Sequence 기술을 이용한 음성 합성 기술개발 동향. [http://www.riss.kr/search/detail/DetailView.do?p\\_mat\\_type=1a0202e37d52c72d&control\\_no=2ab1d078bd0d406d4884a65323211ff0](http://www.riss.kr/search/detail/DetailView.do?p_mat_type=1a0202e37d52c72d&control_no=2ab1d078bd0d406d4884a65323211ff0)  
<sup>7</sup> 설명 | 머신러닝 성능에 영향을 미치는 요소를 중요도순으로 나열하면 양질의 데이터>좋은 특징(feature)>효율적인 알고리즘이라고 할 수 있음  
<sup>8</sup> 설명 | 라벨링 데이터를 얻는 일은 사람의 손을 거쳐 일일이 만들어야 하는 만큼 큰 비용을 수반함. 이에 연구자들은 대규모 비(非)라벨링 데이터에 대해 적은 양의 라벨링 데이터를 활용할 경우 학습 정확도가 높아짐을 확인함  
<sup>9</sup> 참고 | Learning bilingual word embedding

- (1) 먼저 언어 A와 언어 B에 대한 단일 언어 말뭉치(monolingual corpora)를 이용해 각각의 단어 임베딩(word embedding)을 생성한다.
- (2) 가장 쉬운 단어들만 모아 대역(bilingual) 쌍 샘플 목록을 구축한다.
- (3) 대역 말뭉치를 공유 임베딩 공간(shared embedding space)에 매핑(mapping)하는 변환 네트워크(transformation network)를 학습시킨다. 그러면 나머지 다른 단어들도 자동으로 비슷한 위치에 매핑된다.

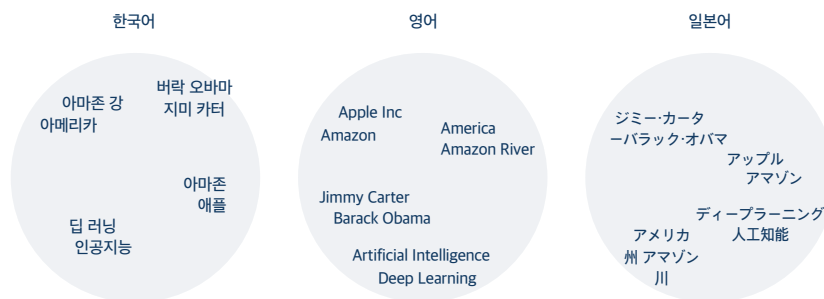
이 방식은 일부 언어 쌍에 대한 지도 학습 기반 교차언어 임베딩보다 성능이 좋은 것으로 알려졌다.<sup>\*10</sup>

[ 그림 2 ] 영어와 독일어 말뭉치를 공유 공간에 매핑한 모습<sup>\*11</sup>



하지만 이런 접근 방식에는 근본적인 한계가 있다. 워드 임베딩 그 자체는 비결정론적(non-deterministic)이다. 알고리즘 학습에 활용하는 말뭉치 종류와 정렬 방식, 어떤 알고리즘을 활용하는지에 따라 임베딩 결과는 천차만별로 달라질 수 있다는 의미다. 이로 인해 같은 의미를 나타내는 단어 A(언어 a)와 단어 B(언어 b)가 있으면 각 언어라는 맥락에서 봤을 때 다른 단어 분포(word distribution)를 가질 확률도 배제할 수 없다.

[ 그림 3 ] 같은 의미를 뜻하는 단어가 한국어, 영어, 일본어에서 다른 단어 분포를 보인다.<sup>\*12</sup>



아울러 word2vec은 특정 문맥이나 문장 내에서 단어가 갖추는 특정한 의미를 이해하지 못한다. word2vec은 문장 내 한 단어와 같이 출현하는 다른 단어를 관련 단어로 학습하는 것일 뿐 단어의

의미를 학습하지는 않기 때문이다.<sup>\*13</sup> 예를 들어, ‘민수는 지난주 수술을 받고 다리(의미 1)를 고쳤다’와 ‘서울 한강에 있는 다리(의미 2)는 총 28개다’라는 문장을 보자. 여기서 학습 모델은 각 문장에 등장하는 ‘다리’의 의미 차이를 이해하지 못한다. 단지 철자의 다름을 통해 서로 다른 단어를 인식할 뿐이다.

**문장별 ‘다리’의 의미 차이**

- (의미 1) 사람이나 동물의 몸통 아래 붙어 있는 신체의 부분. 서고 걷고 뛰는 일 따위를 맡아 한다.
- (의미 2) 물체의 아래쪽에 붙어서 그 물체를 받치거나 직접 땅에 닿지 않게 하거나 높이 있도록 버티어 놓은 부분.

어휘집(seed lexicon)이 되는 적은 양의 병렬 말뭉치를 초기 학습에 이용한다는 점에서 온전한 비지도학습이라고 보기 힘든 어려운 측면도 있다. 따라서 한 언어에 대한 대규모 단일 언어 말뭉치(monolingual corpora)만을 가지고도 언어 간 번역을 가능케 하는 기술 연구가 필요하다.

**단어 임베딩(word embedding)**

단어, 문장 등의 의미를 나타내기 위한 분산 표상(distributed representation)으로, 저차원의 벡터 공간으로 표현된다. 대표적인 임베딩 학습 방법으로는 word2vec이 있다.

**word2vec<sup>\*14</sup>**

같은 맥락에 있는 단어는 가까운 의미를 가진다는 전제에서 출발한다. 한 단어와 함께 출현하는 다른 단어를 관련 단어로 인공신경망에 학습시킨다. 문서 내 연관 단어들은 가까운 곳에 출현할 가능성이 높아 학습을 반복하는 과정에서 주변 단어가 비슷한 두 단어는 가까운 벡터 공간에 놓이게 된다.

**두 번째, 맥락 인지 기계 번역(context-aware machine translation)**

전형적인 기계 번역은 문장 단위로 번역된다. 그러나 이런 접근 방식은 이상적이지 않을 수도 있다. 예를 들어, [예시 1]처럼 ‘the bank’가 포함된 a 문장은 ‘저기 은행(the bank)을 봐라’로 번역될 수 있다. 그러나 b 문장처럼 문맥에 따라서는 ‘둑(the bank)’으로 해석될 수 있지만 상용 번역기는 ‘저기 은행을 보세요. 남자들이 낚시하고 있어요’로 번역한다. 문장 단위 번역이 문맥을 고려하지 못하고 있음을 보여주는 단적인 예이다.

**[ 예시 1 ]**

- Look at the bank over there.
- Look at the bank over there. Men are fishing from it.

문맥을 고려하지 않은 기계 번역이 가진 문제점을 보여주는 또 다른 예시를 보자([예시 2]).

**[ 예시 2 ]**

He took another full piece and chewed it. “It is a strong full-blooded fish,” he thought. “I was lucky to get him instead of dolphin. Dolphin is too sweet. This is hardly sweet at all and all the strength is still in it.” - From the Old Band and the Sea

(구글 번역 결과) 그는 또 다른 조각을 가져다가 씹었다. “그것은 강한 전혈 물고기입니다.” 그는 생각했다. “나는 돌고래 대신에 그를 잡아서 운이 좋았다. 돌고래는 너무 달다. 이것은 전혀 달콤하지 않으며 모든 힘이 여전히 남아 있습니다.”

\*10 참고 | <https://github.com/kweonwoo/papers/issues/100>  
 \*11 참고 | <http://ruder.io/cross-lingual-embeddings>  
 \*12 참고 | <http://youngl.org/wp-content/uploads/2017/12/위키와-다국어-검색-이만재-박사.pdf>

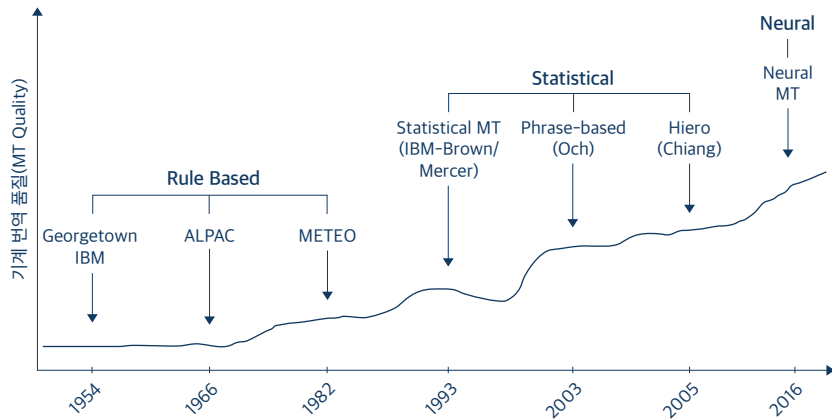
\*13 설명 | 이런 문제를 보완하고자 sense2vec과 같은 모델이 개발되기도 함  
 \*14 참고 | Word2Vec을 이용한 반복적 접근방식의 그래프 기반 단어 중의성 해소, <http://www.dbpia.co.kr/Journal/ArticleDetail/NODE06646570#>

문맥상 'him'은 '그(사람)'가 아니라 '물고기(fish)'를 지칭한다. 따라서 '나는 돌고래 대신에 그 녀석(물고기를) 잡았다'라는 번역이 조금 더 자연스럽다. 그러나 번역기는 문맥을 고려하지 않은 결과를 도출했다. 아울러 톤이 일관적이지 않은 현상(~했다, ~했습니다)도 확인할 수 있다. 그렇기 때문에 자연스러운 결과물을 얻기 위해서는 문법성(grammatical gender)<sup>\*15</sup>, 단수/복수(number or plurality), 경어체(honorifics), 생략(ellipsis)<sup>\*16</sup>과 같은 언어학적 특징(linguistic features)을 고려해야 하는 이유다.

이런 언어학적 특징을 잘 포착하기 위해서는 문맥 파악이 선행되어야 한다. 그렇다면 문맥은 어떻게 정의해볼 수 있을까? 그 방식에 따른 여러 가지 방법론을 생각해볼 수 있다. 한 논문에서는 메모리 네트워크(memory network)를 제안했다. 보통은 번역하려는 문장 바로 앞에 나오는 문장을 함께 학습하면, 실제로 이전보다 번역 성능이 높아짐을 확인할 수 있었다. 하지만 현재 문장을 번역하는 데 필요한 모든 정보가 바로 앞 문장에만 있지 않다는 사실은 직관적으로 알 수 있는 부분이다. 이에 따라 앞서 등장한 모든 정보를 기억하는 외부저장장치(external memory)를 두는 방안도 고려해볼 수 있다. 다만 이 방법은 모델 복잡도가 높아지는 경향이 짙다.

기계 번역의 역사

[ 그림 4 ] 기계 번역의 역사 © 스탠퍼드대학교<sup>\*17</sup>



규칙 기반 기계 번역(rule-based machine translation, RBMT)

제2차 세계대전이 끝나고 냉전이 시작된 1950년대, 미국 국방성은 러시아어를 자동으로 번역하는 연구를 시작했다. 이때부터 1970년대에서 1980년대까지는 사람이 실제로 번역하는 프로세스를 본떠서 만든 기계 번역이 고안됐다. 형태소 분석(morphological analysis), 구문 분석(synthetic analysis), 의미 합성(semantic composition)의 과정을 거쳐 중간언어(interlingual)를 만든 다음 역의 과정을 거쳐 중간언어를 타깃 언어의 단어로 맵핑(mapping)하는 식이다. 그러나 안타깝게도 이 방식으로는 완벽에 가까운 기계 번역을 제대로 해낼 수 없음이 드러났다.

통계적 기계 번역(statistical machine translation, SMT)

1988년 IBM은 NLP(natural language processing) 팀을 아무리 크게 키워도 도저히 희망이 안 보인다고 말하기까지 했다. '언어학자 한 명을 해고할 때마다 번역의 정확도가 높아진다'는 프레데릭 제리넥(Frederick Jelinek)의 말이 우스갯소리처럼 떠돌 정도였다. 그 이유는 언어학자들이 설계한 규칙 기반 시스템보다 확률에 기반한 '바보 같은' 접근 방식의 결과가 더 좋았기 때문이다. 이로써 1990년대를 전후로 SMT 시대가 열렸다.

신경망 기계 번역(neural machine translation, NMT)

인공신경망에 기반한 기계 번역의 기본 구조는 다음과 같다. 먼저 입력 문장을 문장 벡터 형태로 변환, 이를 기반으로 출력 언어의 문장을 생성하는 방식으로 번역을 수행한다. 입력 문장을 문장 벡터로 변환하는 인코더와 출력 문장을 생성해내는 디코더는 인공신경망으로 구성돼 있다. 데이터를 통해 인코더와 디코더의 파라미터(parameter) 값이 조정된다.

<sup>\*15</sup> 설명 | 단어에 남성과 여성, 중성 같은 성(性)을 부여하는 것  
<sup>\*16</sup> 설명 | 동일한 언어 표현 중 생략이 일어나기도 함. 예제 문장은 다음과 같음. All the children had an ice-cream today. Eva chose strawberry (ice cream), Arthur had orange (ice cream) and William (had orange ice cream) too. (담화분석) - 현대 영어학 총서9, pp.97-98  
<sup>\*17</sup> 참고 | <https://nlp.stanford.edu/projects/nmt/Luong-Cho-Manning-NMT-ACL2016-v4.pdf>

세 번째, 맞춤법/문법 검사기(grammar checker)

문법이나 맞춤법, 띄어쓰기 등의 오류(error)는 콘텐츠에 대한 신뢰도를 떨어뜨린다. 글을 쓸 때 심적 부담감을 일으키는 요인이다. 이에 사람들은 맞춤법/문법 검사기를 이용해 오류를 최소화하는 데 힘쓴다. 맞춤법/문법 검사기는 다음과 같은 오류를 찾아내 바로잡는다.

- (1) 맞춤법/문체(spelling/style) Is this orngal? → Is this original?
- (2) 문법(grammar) I have an apples. → I have an apple.
- (3) 연어(collocation)<sup>\*18</sup> I like powerful coffee. → I like strong coffee.
- (4) 어조(tone)/태(voice)<sup>\*19</sup>/격식(formality)

이런 오류를 찾아내기 위해 취했던 전통적인 방법은 다음과 같이 다섯 가지로 볼 수 있다.

- (1) 지식 베이스(knowledge base)에 존재하는 지식과 비교·대조한다. 매칭되는 결과가 없다면 사용자 입력값에 오류가 있다고 인지한다. 대표적으로 철자 오류는 어휘 사전에 존재하지 않으므로 단순한 형태소 분석을 통해서도 그 오류의 검출 및 교정이 비교적 간단하다고 볼 수 있다.
- (2) 대량의 말뭉치에 기반한 통계치를 바탕으로 주어진 문장의 구문을 분석하는 통계(statistics) 기반 분석 방법이 있다. 예를 들어, 사람들이 자주 틀리는 표현을 구문 분석에 활용하는 방식이다.
- (3) 몇 번의 추가 편집으로 단어 A를 단어 B로 만들 수 있는지를 측정하는 편집 거리(edit distance, Levenshtein distance)를 이용하기도 한다. 철자 오류와 편집 거리가 작은 단어를 추출하고, 이를 철자 오류가 발생한 단어와 치환하는 식이다.
- (4) 주어진 문맥상에서 다음에 나올 적절한 글자, 단어 또는 문장을 예측하는 언어 모델(language model)을 만드는 방법도 있다. 예를 들어, '요금 결제'에서 '결제'는 (1)에서 등록된 올바른 단어이지만 함께 쓰인 '요금'이라는 문맥을 고려해 '결제'가 되도록 바꾼다.<sup>\*20</sup>
- (5) 전문가가 만든 규칙(human-generated rules)을 통해 문법적 오류를 수정하는 방법도 있다. 이를 이용하면 언어에 대한 상세한 구조를 정확하게 파악할 수 있다. 규칙이 많을수록 성능을 높일 수 있는 이유다.

지금까지는 대체로 규칙이 더 나은 성능을 보여왔다. 오늘날 가장 많이 사용되는 한국어 맞춤법/문법 검사기 또한 규칙에 기반을 둔 문맥에 의존해 철자 오류를 교정한다. 문제는 전문가의 지식에 기반하여 규칙을 만드는 만큼, 막대한 비용과 시간이 소요된다는 점이다. 아울러 발생 빈도가 높거나 정형화된 오류는 규칙으로 찾아낼 확률이 높으나, 그렇지 않은 유형은 규칙으로도 교정하지 못할 확률이 높다. 딥러닝이 더 나은 성능을 낼 것으로 기대되는 이유다.

맞춤법/문법 검사에 대한 딥러닝의 접근 방식은 크게 두 가지로 나누어볼 수 있다. 하나는 최대우도추정(maximum likelihood estimation, MLE)<sup>\*21</sup>을 이용한 seq2seq이다. 바르게 입력되지 않은 문장 X(source)를 입력받아 바르게 입력된 문장 Y(target)를 반환할 확률을 최대로 하는 모델 파라미터를 학습한다. 다른 하나는 강화 학습(reinforcement learning)이다. 강화 학습은 누적 보상을 최대화해 소프트웨어 에이전트(AI)가 특정 목표에 도달하기 위해 취해야 할 조치를 결정하도록 가르친다. 감독 학습과 달리 강화 학습에서는 라벨링된 입력값과 결과값이라는 데이터 쌍을 필요로 하지 않는다. 인공지능은 자신만의 방식을 찾아 명시적인 지시를 받지 않고도 성능을 개선한다. 여기에서는 성능을 미세조정(fine tuning)할 때 seq2seq를 활용하는 경우가 많다.

그러나 이런 방법론은 규칙에 기반을 둔 전통적인 접근 방식만큼 좋은 성과를 내지 못한다. 먼저 올바른 문장으로만 구성된 데이터 세트를 학습한 딥러닝 알고리즘은 단순히 정답을 내놓을 줄만 안다. 이는 틀린 문장을 입력받아도 오류로 인식하지 못한다는 의미다. '틀린 문장, 교정된

<sup>\*18</sup> 설명 | 특정한 뜻을 나타낼 때 흔히 함께 쓰이는 단어 간의 결합을 뜻함  
<sup>\*19</sup> 설명 | 동작을 바라보는 관점 차이로 발생하는 동사의 표현 형식을 일컫는 말  
<sup>\*20</sup> 참고 | 어절 N-gram을 이용한 문맥 의존 철자 오류 교정 <http://dx.doi.org/10.5626/JOK.2014.41.12.1081>  
<sup>\*21</sup> 참고 | 주어진 샘플 x에 대해 우도(likelihood)를 가장 크게 해주는 모수를 찾는 방법 <https://datascienceschool.net/view-notebook/864a2cc43df44531be32e3fa48769501>

문장'으로 구성된 병렬 말뭉치를 만들더라도, 만족할 만한 성능을 보이는 알고리즘을 만드는 데 충분한 규모를 확보할 수 있을지에 대해서도 미지수다. 합성 데이터 생성(synthetic data generation)을 통해 랜덤하게 틀린 문장을 생성하는 방안을 고려해볼 수는 있으나, 괄목할 만한 성과를 내는 전략으로는 충분하지 못할 것으로 보인다.

**네 번째, 한정된 자원을 활용한 음성합성(speech synthesis with limited resources)**

음성합성(speech synthesis)은 주어진 텍스트를 정확하고 자연스럽게 읽는 음성 데이터를 생성하는 기술을 의미한다. 주어진 텍스트에 대응하는 음성을 출력한다는 점에서 TTS(text to speech)라고 표현하기도 한다.

1990년대 초 규칙 기반 음성합성 기술에 이어, 2000년대에는 대용량 음성 데이터 기반 음편 조합(unit selection) 방식이 주목을 받았다. 전통적인 신호처리 방식 중 대용량 음성 데이터 기반인<sup>22</sup> 이 방식은 짧은 단위의 음편을 저장한 후 텍스트에 해당하는 음편을 연결해 합성음을 출력한다. 이 방식은 실제 사람의 음성을 녹음한 음편이기에 원음에 가까운 고품질 합성음을 만들 수 있다는 장점이 있는 한편, 음편의 연결 경계가 부자연스럽고 주어진 문장에 항상 똑같이 말하는 문제 등의 단점을 지닌다. 이러한 한계를 극복하고자 통계적 파라미터 방식이 고안됐으며, 이후 은닉 마코브 모델(hidden markov model, HMM)과 심층 신경망(deep neural network, DNN)을 기반으로 한 모델로 발전하였다.<sup>23</sup> 현재 웨이브넷(WaveNet), 타코트론(Tacotron), 딥보이스(Deep Voice) 등의 LSTM(long short term memory)을 이용한 seq2seq 네트워크가 주로 활용되고 있다.<sup>24</sup>

오늘날의 음성합성 시스템은 크게 3단계를 거친다. 첫 번째 단계에서는 문장 데이터에서 특징 정보를 분석한다. 두 번째 단계에서는 문장 내 특징 정보로부터 음성 파라미터를 추출한다. 마지막으로 세 번째 단계에서는 음성 파라미터에서 음성을 재구성(vocoder)한다. 문자 단위로 입력된 텍스트로부터 음성을 합칠 수 있도록 하나의 모듈로 이루어진 E2E(end to end) 시스템이라고도 볼 수 있다. 기존 방식과 달리 각 모듈에 대한 전문적인 지식이 필요하지 않아 진입장벽이 낮고, 각 모듈에서 손실(loss)이 누적되는 문제를 해결할 수 있다.<sup>25</sup>

현재의 음성합성 기술은 인간과 거의 비슷한 음성을 구사할 수 있는 기술 수준까지 도달했다. 특히 지난 2017년 12월 구글(Google)이 발표한 타코트론2(Tacotron2)는 현존하는 최고 음질의 합성음을 만들어내고 있다. 하지만 여기에는 선행조건이 따른다. 원하는 수준의 목소리 품질을 얻기 위해서는 잡음(noise)이 없는 깨끗한 음질을 다량 확보해야 한다는 점이다. 적어도 한 사람의 목소리로 구성된 20시간 이상 분량의 녹음 데이터가 필요하다. 그러나 현실에서는 이런 데이터를 쉽게 얻을 수 없고, 얻을 수 있더라도 스튜디오 녹음 및 편집 비용이 많이 든다. 기존에 보유하고 있는 데이터 세트는 크기 자체가 부족해 배우지 않은 발음과 억양이 생길 수밖에 없다. 따라서 어느 정도 잡음이 있거나 적은 분량의 음성 데이터로도 깨끗한 목소리를 생성할 수 있는 강인한(robust) 인공지능을 만드는 연구가 후행되어야 할 것이다. 현재 1분 정도의 샘플 데이터만으로는 고품질의 음성 합성 엔진 개발은 현실적으로는 불가능에 가깝다. 다만 해당 목소리를 보유한 사람의 목소리인 척 흉내를 내는 모델 정도는 기대해볼 수 있다.

**다섯 번째, 제주어 재활성화(jejeuo revitalization)**

지구상에는 6000여 개의 언어가 존재한다. 이중 인류의 절반 규모인 50억 명 정도가 20개의 상위권 언어를 구사한다. 그리고 상위권 언어를 제외하면 한 언어를 구사하는 인구가 1만 명도 되지 않는다. 소수가 사용하는 언어의 비중이 절대적으로 높다는 의미다. 소수 언어를 사용할 사람이 점차 없어지면 해당 언어는 지구상에서 사라진다. 학계에서 세계 언어의 50%가 10년 내 소멸할 것으로 전망하는 배경이다. 이는 6000여 개의 언어를 매개로 향유하던 무형자산이 사멸될 수 있음을 시사한다.<sup>26\*27</sup>

제주어 또한 소멸 위기에 처한 언어 중 하나다. 이에 유네스코(UNESCO)는 지난 2010년 제주어를 소멸 위기 언어 5단계 중 사멸 직전인 4단계(아주 심각한 위기에 처한 언어(critically endangered language))로 분류했다. 제주어를 사용하는 인구가 급격하게 감소하고 있기 때문이다.

**유네스코에서 소멸 위기에 놓인 언어를 판단하는 9가지 기준<sup>28</sup>**

- (1) 세대 간의 언어 전승
- (2) 화자의 절대 수
- (3) 전체 인구 대비 해당 언어 구사자 비율
- (4) 언어 사용 분야의 변화
- (5) 새로운 영역, 매체에(해당 언어를 사용하여) 반응하기
- (6) 언어 교육과 읽고-쓰기 자료 사용 여부
- (7) (해당 언어에 대한) 공식적인 지위 부여/사용을 포함한 정부와 해당 기관의 언어에 대한 태도와 정책
- (8) 자신의 언어에 대한 해당 그 지역 사회 구성원의 태도
- (9) (해당 언어) 자료의 양과 질

**[ 표 1 ] 세대 간 언어의 계승 정도**

소멸 위기 정도	층위(단계)	언어의 세대 간 계승
안전	-	모든 세대가 해당 언어를 사용하고 있음. 세대 간 계승이 계속됨
취약	1	대부분의 아이가 해당 언어를 사용하고 있음. 그러나 이 언어는 한정된 영역에서 사용됨(보기: 집)
위험	2	아이들이 해당 언어를 집에서 모국어로 배우지 않음. 특히 부모 세대와 그 이전 세대가 사용함
매우 위험	3	조부모들이 해당 언어를 사용함. 부모 세대는 이 언어를 이해할 수 있지만, 그들끼리 또는 아이들과 이 언어를 사용하지 않음
소멸 직전	4	조부모 이상의 세대가 해당 언어의 가장 젊은 사용자임. 그들은 단지 부분적으로만 드물게 이 언어를 사용함
소멸	5	해당 언어의 사용자가 더는 존재하지 않음

© 유네스코 한국위원회

대한민국 표준어가 있음에도 제주어의 소멸을 막는 방안을 검토해야 하는 이유는 무엇일까? 제주어가 언어학적 관점에서 연구 가치가 높기 때문이다. 제주어는 훈민정음의 아래아를 사용하는 것은 물론, 표준어에 없는 때가림소(시제) ‘ㅇ, ㄴ’을 활용한 제주어만의 문법을 지니고 있다.<sup>29</sup> 이에 제주도는 제주어를 보전하고자 지난 2007년 ‘제주어 보전 및 육성’ 조례안을 제정, 제주어를 활성화(revitalization)할 다양한 방법을 모색하고 있다. 녹음, 구술 자료집 편찬, 제주어 교육, 연구 및 사전 편찬과 같은 활동이 바로 여기에 해당한다. 그러나 이는 보존(preservation)이나 유지(sustaining)를 위해 필요한 언어 자료화(documentation)에 더 가깝다고 볼 수 있다.

<sup>\*22</sup> 참고 | 한국어 TTS 시스템을 위한 엔드투엔드(end to end) 합성 방식 연구, <https://doi.org/10.13064/KSS5.2018.10.1.039>  
<sup>\*23</sup> 참고 | 지식 기반 프랑스어 발음 열 생성 시스템(A knowledge-based pronunciation generation system for French) <https://doi.org/10.13064/KSS5.2018.10.1.049>  
<sup>\*24</sup> 참고 | Sequence to Sequence 기술을 이용한 음성합성 기술개발 동향  
<sup>\*25</sup> 참고 | 한국어 TTS 시스템을 위한 엔드투엔드 합성 방식 연구

<sup>\*26</sup> 참고 | 니컬러스 에번스 저음, 김기혁, 호정은 옮김, 《아무도 모르는 사이에 죽다》, 글항아리  
<sup>\*27</sup> 참고 | 소멸위기 언어 보존 사례 분석을 통한 제주어 보전 방안, <http://www.jst.re.kr/report.do?curPage=4>  
<sup>\*28</sup> 참고 | <http://www.unesco.org/new/en/culture/themes/indangered-languages/faq-on-endangered-languages/>  
<sup>\*29</sup> 참고 | 시한부 생명 제주어 향후 10년에 달했다, [http://news.chosun.com/site/data/html\\_dir/2012/07/15/2012071500489.html](http://news.chosun.com/site/data/html_dir/2012/07/15/2012071500489.html)

그렇다면 어떻게 하면 사람들이 제주어를 적극적으로 사용하게 만들 수 있을까? 딥러닝을 활용한 표준어-제주어 기계 번역과 제주어 TTS가 가장 현실적인 대안이 될 수 있을 것이다. 물론 이 과정은 쉽지만은 않을 것으로 예상된다. 먼저 '표준어 문장-제주어 문장'이라는 대역 말뭉치 데이터가 충분하지 않다. 제주어로만 작성된 자료가 충분하다고 하더라도, 이를 표준어로 대역하지 못한 자료가 수두룩하다. 그나마 제주특별자치도가 내놓은 《제주어구술자료집(2017)》을 활용해보는 방안을 고려해볼 수 있다. 여기서 추출한 대역 말뭉치 중 글자 수가 5개 이상인 문장 쌍이 대략 6만 3000개가 된다. 상용 번역 서비스(영한/일한/중한 번역)에 활용된 데이터와 비교하면 그 양이 턱없이 모자라다. 추가로 데이터 세트를 확보하고, 이를 학습에 용이한 형태로 정제하는 데 막대한 시간과 비용이 드는 점도 장벽이다. 아울러 큰 틀에서 보면 표준어 어휘를 제주어 어휘로 일대일로 대체하는 작업인 만큼, 딥러닝보다는 형태소 분석 방식이 더 적합하다는 가능성도 완전히 배제할 수 없다.



〈딥러닝이 탐구하지 못한 언어와 5가지 태스크〉 브랜치로 연결되는 QR 코드입니다.



# 2018 ICML을 통해 살펴보는 AI 연구 동향

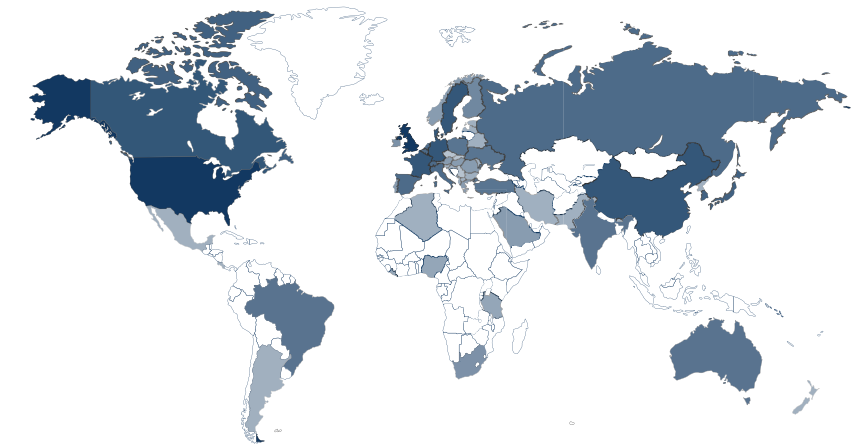
글 | 박찬연 chan.y.park@kc-ml2.com / chan@morulabs.com 대학에서 물리학을 전공하였고, 졸업하기 전 2년여간 웹 프로그래머와 시스템 프로그래머로 일했습니다. 이후 대학원에 진학하여 이론물리학, 그중에서도 끈이론을 전공하였고, 2017년까지는 물리학과 기하학 그리고 프로그래밍에 대한 관심과 경험을 바탕으로 한 연구를 계속했습니다. 그러던 중 머신러닝에서 데이터를 바탕으로 수학적 모델을 세우고 이를 프로그래밍으로 구현하는 흥미로운 일이 많음을 알게 되어 딥러닝 연구에 뛰어들기로 결심합니다. 2018년부터 반도체 제조 솔루션 기업 KC의 Machine Learning Lab(ML2)의 principal research scientist와 스타트업 모루랩스(Moru Labs)의 CTO를 맡아서 여러분들의 도움을 받으며 많은 것을 배우고 있습니다.

ICML(International Conference on Machine Learning)은 머신러닝(machine learning)과 딥러닝(deep learning) 분야에서 NIPS(Neural Information Processing Systems), ICLR(International Conference on Learning Representations)과 함께 전 세계 주요 학회 중 하나입니다. 2018년 ICML은 스웨덴 스톡홀름(Stockholm)에서 7월 10일부터 15일까지 개최되었습니다.

올해 ICML에는 5000여 명이 참가하고 2500편 이상의 논문이 제출되는 등 역대 최대 규모를 기록했습니다. 이는 지난 ICML과 비교하여도 큰 폭으로 증가한 숫자이며, 머신러닝 및 딥러닝에 대한 많은 연구와 관심이 반영된 것으로 볼 수 있습니다. 폭발적으로 늘어난 제출 논문들로 논문 심사 과정에서 논문을 제출하신 분들과 심사하시는 분들 모두 많은 어려움을 겪었다는 이야기도 여러 곳에서 들었습니다.

[그림 1] 연도별, 국가별 ICML 참가자 수<sup>\*1</sup>

5000 registered participants (2300 in 2007, 3200 in 2016, 1600 in 2015)



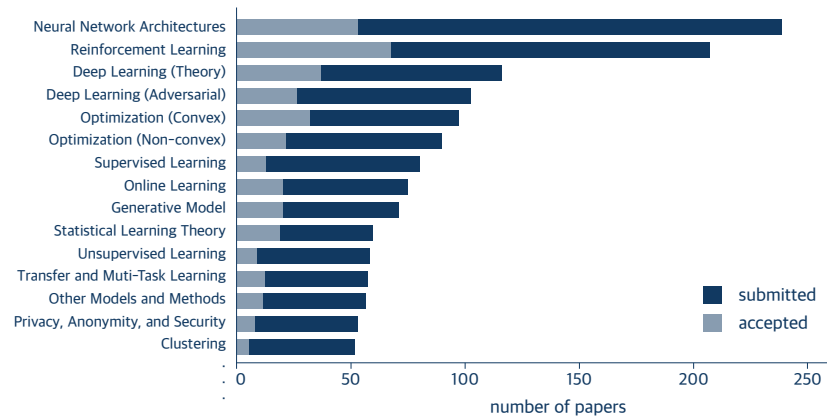
[그림 2] 연도별 제출 및 통과된 논문 수 추이<sup>\*1</sup>



\*1 참고 | F. Bach, Opening remarks, ICML 2018

[그림 3]과 같이 제출된 논문을 주제별로 나누어보았을 때, 가장 많이 제출된 분야는 신경망 구조(neural network architecture)이지만 가장 많이 채택된 분야는 강화 학습(reinforcement learning)이라는 점이 흥미롭습니다. 개인적인 추측으로는 새로운 구조(architecture)를 제안하는 연구는 그것이 모델에 대한 이해를 돕거나, 기존의 모델과 비교할 때 명확하게 향상된 성능을 보여주지 않는다면 좋은 연구성과로 받아들여질 수 없다는 점이 그 이유가 아닐까 생각합니다. 또 다른 흥미로운 점은 딥러닝의 이론적인 이해에 대한 연구가 세 번째로 많이 다루어지고 채택된 연구주제라는 것입니다. 머신러닝과 딥러닝을 연구하는 분들께서 현재의 성과에 안주하지 않고 근본적인 딥러닝의 이해에 대해 계속 고민하고, 이를 통해 더욱 큰 발전을 바라고 계신다는 생각이 들었습니다. 마찬가지로 머신러닝의 핵심이라고 할 수 있는 최적화(optimization)도 여전히 중요한 주제로 다루어지고 있다는 점도 눈길을 끌었습니다.

[ 그림 3 ] 주제별 제출 및 통과된 논문 수<sup>\*1</sup>



이번 ICML에서는 구두 발표(oral presentation)와 워크숍(workshop)의 규모도 굉장하였습니다. 학회 첫날인 7월 10일에는 9개의 튜토리얼(tutorial)이 3개의 동시 세션(parallel session)으로 나누어 진행되었고, 11일부터 13일까지 3일간의 학회 메인 기간 동안에는 매일 200개가 넘는 논문 발표가 10개의 세션에서 동시에 운영되었습니다. 또한 3일간 67개의 워크숍도 개최되면서 상황리에 진행되었습니다. 상황이 이렇다 보니 매일 어떤 세션에 참여할지 결정하는 것도 큰 일이었고, 어떤 결정을 내리던 많은 내용들을 놓칠 수밖에 없어 안타깝기도 하였습니다. 또한 현재의 연구 상황과 앞으로의 연구 방향에 대한 의견을 나누며 연구자분들과 교류하는 데 많은 시간을 할애하다 보니 더 많은 발표에 참석하지 못한 아쉬움도 남았습니다.

올해 ICML의 규모를 고려하여 이 글에서는 모든 내용을 담기보다는, 이번 학회에서 인상 깊었던 발표 내용들을 간단히 정리한 후 이어서 제가 평소 관심을 가지고 있는 연구 주제 중심으로 정리해보고자 합니다. 다행히도 꽤 많은 발표 영상들이 학회 중에 생중계되거나 녹화되어 인터넷에 올라왔습니다. 조만간 모든 발표가 녹화된 영상이 올라온다고 하니, ICML 웹사이트에서 관심 분야의 발표 내용을 확인해보시면 좋을 것 같습니다.<sup>\*2</sup> 그리고 이번 ICML에서 가장 많이 다루어진 주제인 강화 학습 위주로 학회 내용을 정리한 데이빗 아벨(David Abel)의 글<sup>\*3</sup>도 살펴보기길 추천해드립니다.

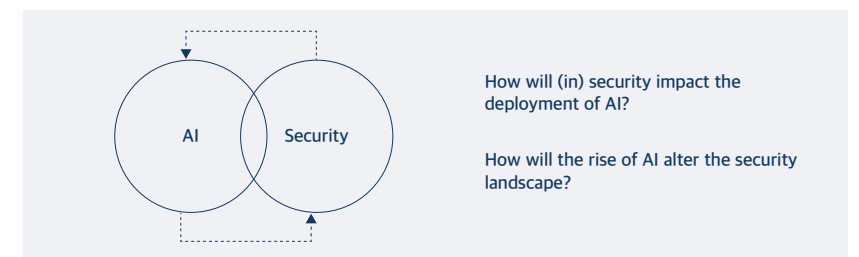
\*2 참고 | <https://icml.cc/Conferences/2018/Schedule>  
 \*3 참고 | [https://david-abel.github.io/blog/posts/misc/icml\\_2018.pdf](https://david-abel.github.io/blog/posts/misc/icml_2018.pdf)

### 머신러닝과 보안

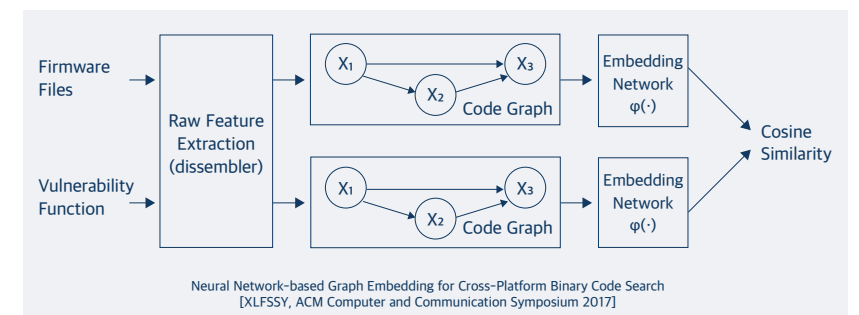
정보기술 기반 산업에서 보안은 언제나 중요하게 생각해야 할 부분입니다. 개인정보 및 사업정보 등 대량의 데이터를 기반으로 하는 데이터 사이언스(data science)와 머신러닝이 더욱 보편화됨에 따라 앞으로 새로운 보안 문제와 대책이 중요하게 다루어질 것으로 생각됩니다. 머신러닝 커뮤니티에서도 이러한 생각을 강조하고 서로 공유하고 있다는 것을 이번 학회를 통해 다시 한 번 느낄 수 있었습니다.

먼저 기조연설 중 하나였던 다운 송(Dawn Song)의 'AI and Security: Lessons, Challenges and Future Directions(인공지능과 보안: 레슨, 도전 과제 및 향후 방향)'에서는 다양한 예제를 통해 새로운 머신러닝 기술이 기존의 보안 문제에 어떤 도움을 줄 수 있으며, 또한 머신러닝 기술의 발달과 함께 어떠한 새로운 보안 문제가 등장할 수 있는지에 대해서 다루었습니다. 딥러닝을 이용한 사물인터넷(IoT) 디바이스의 취약성 발견과 시각 인식 시스템을 적대적으로 공격하는 기법인 적대적 공격(adversarial attack) 등이 그 예입니다. 제시된 문제점 중 가장 마음에 와 닿았던 것은, 만일 머신러닝 모델들의 학습 능력이 데이터를 모두 기억할 수 있을 정도로 뛰어나다면, 머신러닝 모델의 입력값을 적절히 조절하여 그에 해당하는 출력값을 통해 학습된 데이터를 볼 수 있을까라는 것이었습니다. 머신러닝 모델, 그중에서도 특히 딥러닝 모델들이 데이터를 어떻게 처리하고 저장하는지에 대한 이해는 아직 부족하고 이와 관련된 연구도 아직 활발하게 진행 중입니다. 특히 딥러닝 모델들의 경우 산술적으로는 모든 데이터를 기억할 수 있을 정도로 큰 학습 능력을 가지고 있다는 것과 몇몇 모델들의 경우에는 실제로 기억에 의한 학습이 많은 부분을 차지한다는 것을 고려해봤을 때, 의료 정보나 신용정보와 같은 민감한 개인정보를 다루는 머신러닝 모델의 보안성은 매우 중요한 주제입니다.

[ 그림 4 ] AI와 보안<sup>\*4</sup>



[ 그림 5 ] 딥러닝을 이용한 사물인터넷 디바이스의 취약성 찾기<sup>\*4</sup>



\*4 참고 | AI and Security: Lessons, Challenges and Future Directions, D. Song, ICML 2018

[ 그림 6 ] 딥러닝을 적대적으로 이용한 사례<sup>4</sup>



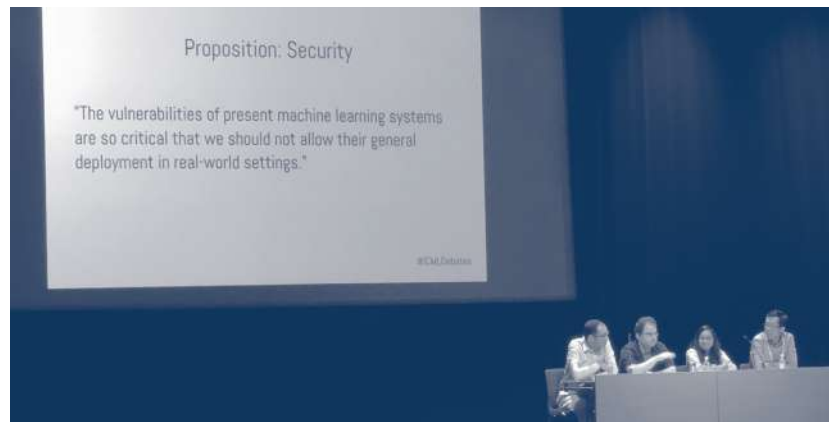
이번 ICML의 최우수 논문 중 하나<sup>5</sup>도 보안과 관련된 것이었습니다. 특히 논문 내용 외에 발표 내용 중 “보안 문제를 방어하는 연구를 하고 싶다면 먼저 최신 보안 방어 시스템을 공격하는 연구를 해봐야 한다”는, 다시 말해 해커의 자세를 가져야 한다는 이야기가 가장 인상 깊었습니다. 또한 딥러닝 모델 해커들이 방어 방법에 대한 연구논문을 모두 읽고 이해한 후 그 취약성을 공격할 것이라는 가정 하에 보안 방어 연구를 진행해야 한다는 지적도 설득력 있게 다가왔습니다.

**보안 연구자가 가져야 할 마음가짐<sup>6</sup>**

The threat model MUST assume the attacker has read the paper and knows the defender is using those techniques to defend.

토요일에 열린 워크숍 중 하나<sup>7</sup>에서는 머신러닝과 관련된 여러 주제들에 관해 토론을 하는 자리가 마련되었습니다. 그중 보안과 관련해서 주된 논점 중 하나는 우리가 한정된 연구자원을 첨단 머신러닝 기술 개발에 투입하는 것이 옳은가, 아니면 연구의 시작부터 보안을 강화하는 데에 투입해야 하는가에 대한 것이었습니다. 물론 한 번의 토론으로 결론이 날 수 있는 주제가 아니라 머신러닝 커뮤니티가 지속적으로 함께 고민해보아야 할 문제입니다만, 이렇게 모두가 문제의식을 공유하고 그에 대해 의견을 주고받고 있다는 점이 머신러닝 커뮤니티가 건강하다는 증거가 아닐까 생각되었습니다.

[ 그림 7 ] 머신러닝과 보안에 관한 토론<sup>8</sup>



<sup>5</sup> 논문 | Athalye, A., Carlini, N., Wagner, D. (2018). Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:274-283  
<sup>6</sup> 참고 | Carlini, N., Best Paper Session 1, ICML 2018  
<sup>7</sup> 참고 | <https://www.machinelearningdebates.com>  
<sup>8</sup> 참고 | <https://twitter.com/LaflammeSteph>

**공정한 머신러닝**

머신러닝으로 학습된 모델은 데이터가 반영하는 현재 사회의 편견(bias)까지도 배우기 때문에 그에 따르는 사회적 영향을 고려해야 합니다. 이는 딥러닝의 혁신적인 발전으로 머신러닝 모델이 많은 데이터를 학습하는 것이 쉬워졌을 때부터 중요하게 생각되어 왔습니다.

한편 딥러닝이 주목받기 이전부터 사회과학 분야의 연구자들은 적극적으로 데이터 기반의 머신러닝 방법론을 받아들였습니다. 예를 들어 과거에는 공공 데이터로서 공개된 많은 정부 문서들이 사회과학 연구자들에게는 아주 귀중한 자료임에도 양이 워낙 방대하여 그 자료들을 연구에 활용하는 데 어려움을 겪었습니다. 그러나 이제는 LDA(latent dirichlet allocation)와 같은 자연어 처리 모델이 개발된 덕분에 연구 주제와 관련된 문서를 찾아내기가 훨씬 수월해졌습니다. 이와 같이 머신러닝을 활용한 다양하고 흥미로운 사회과학 연구에 대해서는 2016년 SciPy(Scientific Python Conference)의 기조 연설 중 하나였던 한나 월러크(Hanna Wallach)의 ‘Machine Learning for Social Science(사회과학 부문의 머신러닝)’<sup>9</sup>에서 살펴볼 수 있습니다. 또한 계량경제학이나 미시경제학을 연구하는 많은 분들도 사회과학과 수학, 그리고 데이터 사이언스가 만나는 지점에서 다양한 활동을 하면서 머신러닝과 딥러닝의 발전으로 많은 혜택을 보게 되었습니다. 그리고 2008년 미국 대통령 선거에서 데이터를 기반으로 한 선거전이 화두로 떠오르면서<sup>10</sup> 정치학에서도 머신러닝과 데이터 사이언스의 중요성은 더욱 커지고 있습니다.

머신러닝 발전이 사회에 어떠한 영향을 미치며, 특히 그중에서 좋지 않은 영향을 줄이는 방법에 대한 주제는 머신러닝 커뮤니티 내에서만 연구하기에는 어려운 주제입니다. 따라서 전통적으로 사회과학의 여러 주제에 대해 깊은 관심과 고민을 가지고 있던 연구 커뮤니티와의 협력이 중요합니다. 이번 ICML의 또 다른 최우수 논문 중 하나<sup>11</sup>도 정량적으로 머신러닝의 공정함에 대해서 생각해 보는 연구였습니다. ‘Fairness, Accountability, and Transparency in Machine Learning(머신러닝의 공정성, 책임성 및 투명성)’이라는 주제로 워크숍도 개최되어 앞으로 사회과학과 머신러닝의 조화롭고 생산적인 협력을 기대할 수 있게 되었습니다.

**빅데이터와 베이저안 추론**

GPGPU(General-purpose computing on graphics processing units)의 발전과 그에 따른 딥러닝의 르네상스가 찾아오기 전까지 머신러닝의 최전선은 베이저안 추론(Bayesian inference)이었다고 해도 무방합니다. 베이저안 추론 기반의 머신러닝은 상대적으로 적은 수의 데이터를 가지고도 해당 분야의 전문 지식까지 포함하는 좋은 모델을 만들 수 있다는 장점이 있지만, 많은 데이터를 효율적으로 처리하기 어렵다는 단점도 가지고 있습니다. 하지만 그 장점들 때문에 여전히 외행성 발견, 중력파 검출과 같은 과학 연구와 비행기 잔해 탐색, 교통량 분석, 무담보 소액대출(microcredit), 야생동물 보호, 제트엔진 분석 등 산업의 많은 분야에서 중요하게 활용되고 있습니다([그림 8] 참고). 이번 학회의 튜토리얼 중 하나인 브로데릭(Broderick)의 ‘Variational Bayes and Beyond: Bayesian Inference for Big Data’가 바로 이러한 베이저안 추론을 빅데이터 영역에 활용하는 방안을 연구한 것이었습니다. 특히 발표의 주된 내용은

<sup>9</sup> 참고 | <https://youtu.be/oqfKz-PP9FU>  
<sup>10</sup> 참고 | <https://fivethirtyeight.com/tag/2008-presidential-election>  
<sup>11</sup> 논문 | Liu, L.T., Dean, S., Rolf, E., Simchowitz, M. & Hardt, M. (2018). Delayed Impact of Fair Machine Learning. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:3150-3158

근사적인 베이지안 추론 대신, 효과적인 데이터 샘플링을 통해 베이지안 추론을 하는 방법론에 대한 것이었습니다.

[ 그림 8 ] 베이지안 추론의 다양한 활용<sup>\*12</sup>

**Bayesian inference**

- Analysis goals: Point estimates, coherent uncertainties
- Interpretable, complex, modular; expert information



기조연설 발표자 중 한 분인 웰링(Welling)도 베이지안 추론을 딥러닝에 적용하는 것과 관련해 많은 연구를 진행하신 바 있습니다. 기조연설의 주제는 'Intelligence per Kilowatt-hour'로, 현재의 딥러닝 연구 결과가 실제로 활용되기 위해서는 지금보다 에너지를 훨씬 적게 사용하는 방법론이 필요하다는, 그러한 방법론의 예로 베이지안 추론을 이용하여 모델의 크기를 줄이는 연구에 대한 내용을 다루었습니다.

**딥러닝의 이론적 이해**

기존의 결정론적(deterministic)인 컴퓨터 알고리즘이나 자료구조(data structure)가 최근의 딥러닝 모델들과 다른 점 중 하나는 딥러닝 모델들의 작동에 대한 이론적인 이해가 부족하다는 것입니다. 다시 말하면 어떤 딥러닝 모델의 학습 능력이 뛰어난지, 어느 정도의 학습 능력을 가지고 있으며 어떻게 학습하지 않은 영역으로 일반화를 할 수 있는지, 어떤 구조를 가져야 주어진 데이터로 학습이 가능한지를 이해하지 못하고 있다는 것입니다.

그러한 질문에 대한 이론적인 이해를 위해서는 단순하지만 명확하게 수학적으로 정의될 수 있는 문제를 찾아서 정량적으로 질문에 대한 답을 구하려는 연구가 필요합니다. 이번 ICML에서 발표된 연구 내용 중 그 예를 찾아보자면<sup>\*13</sup>, 두 가지 분류(binary classification)를 위해 학습된 신경망(neural network)이 모든 가까운 최소값들(local minima)에서 학습오차(training error)가 '0'이 되기 위한 조건에 대한 것입니다. 연구 결과로 나온 이를 충족시키는 조건은 1) 활성화 함수(activation function)가 증가하는 순볼록(strictly convex) 형태를 가지며, 2) 신경망은 1개의 숨겨진 층(hidden layer)을 가지거나 또는 쇼트컷트 커넥션(shortcut connection)을 가지는 다층

네트워크(multi-layered network)이어야 하고, 3) 손실 함수(loss function)가 힌지 로스(hinge loss)를 매끄럽게 만든 형태여야 한다는 것입니다. 하지만 최근에 널리 사용되는 깊은 신경망(deep neural network)의 경우 위의 조건을 만족하는 경우는 없습니다. 이는 간단명료한 결론을 내릴 수 있는 문제를 만들 경우 그 대상에 포함되는 모델들이 toy model이 되는 경우가 많다는 현재의 딥러닝의 이론적인 이해 한계를 보여주고 있습니다.

또 다른 예<sup>\*14</sup>는 'tropical geometry'의 풍부한 기하학적인 구조를 활용해 깊은 신경망을 이해하려는 연구로, ReLU(rectified linear unit) 활성화를 가지고 가중치(weight parameter)들이 정수(integer)로 주어지는 피드포워드 망(feed-forward network)과 'tropical rational map'의 상관관계를 보여줬습니다. 하지만 위의 조건을 만족시키는 깊은 신경망 역시 실제로 활용되고 있는 모델들과 비교했을 때 매우 단순한 경우에 국한됩니다.

이번 ICML의 튜토리얼 중 하나인 산지브 아로라(Sanjeev Arora)의 'Toward Theoretical Understanding of Deep Learning(딥러닝의 이론적 이해를 위해)'에서도 현재의 딥러닝에 대한 이론적인 이해가 앞서 말한 것과 같이 매우 제한적이라는 것을 이야기하고 있습니다. 물론 위와 같은 연구들이 향후에 중요한 이론적인 연구 결과가 나올 수 있는 원천이 될 수 있음을 인지하고 있어야 합니다. 동시에 튜토리얼에서는 단순히 발표된 모델의 구조를 이해하려고 하는 연구에 그치지 않고, 더 나아가 주어진 데이터를 효율적으로 학습할 수 있는 모델의 구조와 이를 가능하게 하는 최적화 방법을 먼저 제시할 수 있도록 노력해야 한다는 점이 강조되었습니다. 마무리 즈음에는 라그랑지안(Lagrangian)이나 해밀토니안(Hamiltonian)을 이용한 변분법(calculus of variation)의 사용과 같이 물리학에서 비롯된 직관(insight)의 역할이 중요할 수 있다는 점이 언급되어서, 물리학을 전공한 사람의 입장에서 반갑기도 했습니다.

**딥러닝의 표현 학습**

딥러닝이 지금과 같은 성공을 거둘 수 있었던 이유는, 연구자가 직접 특징(feature)을 선택하거나 생성해야 했던 이전 머신러닝 모델의 학습과는 다르게 역전파(backpropagation)를 이용해 모델이 직접 특징을 생성하고 연구자는 모델의 구조(architecture)와 학습목표(objective function) 설계에 집중할 수 있는 환경이 만들어졌기 때문입니다. 따라서 이렇게 학습된 특징이 사람이 이해할 수 있도록 의미를 가지는 표현(representation)과 어떻게 대응될 수 있는지에 대한 연구도 활발히 진행되고 있습니다.

지금까지 CNN(convolutional neural network)과 같은 딥러닝 모델은 유클리드 공간(Euclidean space)에 규칙적으로 배열된 데이터를 학습하는 데에 큰 성공을 거뒀습니다. 하지만 중요한 자료구조 중 하나인 그래프나 3차원 공간에 정의된 2차원 곡면과 같은 비유클리드 공간에 정의된 데이터의 학습, 또는 이러한 공간을 딥러닝 표현(representation learning) 학습에 사용하기 위해서는 비유클리드 기하학을 이용하여 이전의 모델을 확장하는 연구가 필요합니다. 이러한 주제를 다루는 연구를 아울러서 기하학적 딥러닝(geometric deep learning)이라고 부르며<sup>\*15</sup>, 이번 ICML에서는 기하학적인 딥러닝 연구 중 계층구조(hierarchical structure)를

\*12 참고 | T. Broderick, Variational Bayes and Beyond: Bayesian Inference for Big Data, ICML 2018. (http://www.tamarabroderick.com/files/broderick\_tutorial\_2018\_icml\_part\_i.pdf)  
 \*13 논문 | LIANG, S., Sun, R., Li, Y. & Srikant, R. (2018). Understanding the Loss Surface of Neural Networks for Binary Classification. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:2835-2843

\*14 논문 | Zhang, L., Naitzat, G. & Lim, L. (2018). Tropical Geometry of Deep Neural Networks. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:5824-5832  
 \*15 논문 | Bronstein, M. M., Bruna, J., Lecun, Y., Szlam, A., & Vandergheynst, P. (2017). Geometric Deep Learning: Going beyond Euclidean data. IEEE Signal Processing Magazine, 34(4), 18-42. doi:10.1109/msp.2017.2693418

가지는 데이터를 쌍곡기하학(hyperbolic geometry)에 임베딩(embedding)하는 것과 관련된 논문 두 편이 발표되었습니다.

그중 하나는 쌍곡기하학을 이용한 딥러닝의 첫 연구<sup>\*16</sup>에 이은 후속 연구<sup>\*17</sup>로, 쌍곡선 공간(hyperbolic space)의 푸앵카레 모델(Poincare model)을 사용하여 이루어진 이전의 모델링 방식을 로렌츠 모델(Lorentz model)을 이용하는 방식으로 변경하여 더욱 효율적인 리만 최적화(Riemannian optimisation)를 이끌어내는 것을 제안했습니다.

다른 하나<sup>\*18</sup>는 쌍곡선 기하학에서의 함축 콘(entailment cone)을 닫힌 형태(closed form)로 정의하고 이를 이용하여 방향성 비순환 그래프(directed acyclic graph)의 효율적인 임베딩을 보여준 연구였습니다. 또한 그래프를 다루기 위해 사용되는 딥러닝 모델 중 하나인 그래프 합성곱 네트워크(graph convolutional network)를 개선하기 위한 연구<sup>\*19</sup><sup>\*20</sup>와 실제 데이터와 유사한 그래프를 생성하기 위한 심층 생성 모델(deep generative model)에 대한 연구<sup>\*21</sup><sup>\*22</sup>도 흥미로웠습니다.

기하학의 아이디어를 활용한 표현학습의 또 다른 연구로는 모델이 특징의 대칭성을 유지하고 학습하도록 해서 더 적은 수의 특징으로도 더 많은 정보를 학습할 수 있도록 하는 방법이 있습니다. 여기에서 핵심이 되는 개념인 등변(equivariance)은 수학적으로는 다음과 같이 정의됩니다.

$$\varphi(Tg\mathbf{x})=T'\mathbf{g}\varphi(\mathbf{x})$$

여기에서  $\varphi$ 는 깊은 신경망 전체 혹은 일부에 해당하는 비선형함수(non-linear function),  $\mathbf{x}$ 는 그 함수의 입력값(input),  $\mathbf{g}$ 는 입력 공간(input space)의 대칭 그룹(symmetry group)인  $G$ 의 원소, 그리고  $T\mathbf{g}$ 는 그에 해당하는 변형(transformation)입니다. 위와 같은 등변식이 최대한 유지되도록 하면서 모델을 설계하려는 연구<sup>\*23</sup><sup>\*24</sup>가 다양하게 진행되고 있으며, 한편으로는 이론적인 이해를 위해 미분기하학(differential geometry)에서 널리 사용되는 principal G-bundle을 이용한 연구<sup>\*25</sup>도 진행되고 있다는 사실 또한 이번 ICML의 ‘Towards learning with limited labels: Equivariance, Invariance, and Beyond’ 워크숍에서도 알 수 있었습니다. 아직은 이러한 연구들이 시작 단계에 불과하지만 앞으로의 발전이 기대되는 연구들이라고 생각합니다.

생성 모델(generative model)의 경우 잠재적 공간(latent space)에서의 표현을 이해하는 것이 중요한데, 그러기 위해서는 표현의 얽힘(entanglement) 문제를 해결하는 것이 필요합니다. 이번 ICML에서는 이를 위해 새로운 학습방법<sup>\*26</sup> 및 모델의 구조<sup>\*27</sup>, 그리고 새로운 학습목표<sup>\*28</sup>를 제시한 연구 결과들이 발표되었습니다. 또한 ‘Theoretical Foundations and Applications of Deep Generative Models’이라는 주제로 열린 워크숍에서도 현재 진행 중인 관련 연구들이 대거 소개되었습니다.

<sup>\*16</sup> 논문 | Nickel, M., Kiela, D., Poincare Embeddings for Learning Hierarchical Representations, NIPS 2017

<sup>\*17</sup> 논문 | Nickel, M. & Kiela, D., (2018). Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:3779-3788

<sup>\*18</sup> 논문 | Ganea, O., Becigneul, G. & Hofmann, T., (2018). Hyperbolic Entailment Cones for Learning Hierarchical Embeddings. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:1646-1655

<sup>\*19</sup> 논문 | Chen, J., Zhu, J. & Song, L., (2018). Stochastic Training of Graph Convolutional Networks with Variance Reduction. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:942-950

<sup>\*20</sup> 논문 | Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K. & Jegelka, S., (2018). Representation Learning on Graphs with Jumping Knowledge Networks. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:5453-5462

<sup>\*21</sup> 논문 | Bojchevski, A., Shchur, O., Zugner, D. & Gunnemann, S., (2018). NetGAN: Generating Graphs via Random Walks. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:610-619

<sup>\*22</sup> 논문 | You, J., Ying, R., Ren, X., Hamilton, W. & Leskovec, J., (2018). GraphRNN: Generating Realistic Graphs with Deep Auto-regressive Models. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:5708-5717

<sup>\*23</sup> 논문 | T.S. Cohen, M. Welling, Group Equivariant Convolutional Networks. Proceedings of the International Conference on Machine Learning (ICML), 2016

<sup>\*24</sup> 논문 | Sabour, S., Frosst, N., Hinton, G.E. (2017). Dynamic Routing Between Capsules. NIPS 2017

<sup>\*25</sup> 논문 | T.S. Cohen, M. Geiger, M. Weiler, The Quite General Theory of Equivariant Convolutional Networks (forthcoming), 2018

<sup>\*26</sup> 논문 | Karnitsas, K., Castro, D., Folgoc, L.L., Walker, I., Tanno, R., Rueckert, D., Glocker, B., Criminisi, A. & Nori, A. (2018). Semi-Supervised Learning via Compact Latent Space Clustering. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:2459-2468

<sup>\*27</sup> 논문 | Parascandolo, G., Kilbertus, N., Rojas-Carulla, M. & Scholkopf, B., (2018). Learning Independent Causal Mechanisms. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:4036-4044

<sup>\*28</sup> 논문 | Kim, H. & Mnih, A., (2018). Disentangling by Factorising. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:2649-2658

## 경험적인 해법을 머신러닝으로 대체하기

경험적인 해법(heuristic)은 여행하는 판매원 문제(travelling salesperson problem)와 같이 정확한 답을 찾기가 어려운 문제의 답을 현실적인 제약조건 안에서 찾기 위해 컴퓨터 알고리즘과 자료구조에서 자주 사용되는 방법입니다. 큰 의미에서 보면 알고리즘과 자료구조의 설계 자체도 주어진 문제를 해결하기 위해 연구자들이 이전까지의 경험과 문제에 대한 직관을 바탕으로 이끌어낸 경험적인 해법이라고 할 수 있을 것입니다.

따라서 머신러닝의 발전과 함께 이전의 경험적인 해법을 학습된 모델로 대체하려고 하는 시도는 자연스러운 연구 방향이라고 할 수 있습니다. 이는 알고리즘이나 자료구조를 설계할 때 문제를 해결하기 위한 세부적인 논리적 설계를 고민하기보다는 어떠한 학습 목표를 설정하는 것이 좋은지를 고민할 수 있게 함으로써, 경험적 해법을 이끌어내기 위해 필요하게 마련인 시행착오(trial and error)에 투입되는 시간과 노력을 줄이거나 효율적으로 활용할 수 있기 때문입니다.

이러한 연구 방향 중 흥미로웠던 결과는 구글(Google)이 이끌었던 두 개의 연구였는데, 그중 하나는 강화 학습을 통해 딥러닝 모델의 산출 그래프(computational graph)를 여러 GPU node에 배분하는 것을 최적화하여 전체 실행 시간을 단축하는 것이었습니다.<sup>\*29</sup> 또한 파일 시스템(file system)과 데이터베이스에서 널리 사용되는 B트리(B-tree) 구조와 같은 색인구조(index structure)를 학습된 모델로 대체하는 연구 결과도 흥미로웠습니다.<sup>\*30</sup> 특히 두 번째 연구의 경우, 기존의 색인구조와 학습된 색인구조(learned index structure)의 장단점을 비교하고, 이 둘을 함께 사용하는 구조를 제안함으로써 기존의 자료구조를 학습된 모델을 이용하여 점진적으로 발전시킬 수 있는 방법을 제시했다는 점이 인상적입니다.

지난 6월 26일, 구글의 제프 딘(Jeff Dean)이 구글 서울 캠퍼스에 방문했을 때 위의 두 연구들에 대한 질문을 할 수 있는 기회가 있었습니다. 저의 첫 번째 질문은 “경험적인 해법을 학습 모델로 대체할 수 있는 또 다른 연구 방향으로 어떤 것이 있을까?”였는데, “현재의 컴퓨터 시스템의 소프트웨어와 하드웨어에 다수의 경험적인 해법들이 포함되어 있기 때문에 이를 학습된 모델로 대체하여 성능을 향상시킬 만한 연구 방향 또한 많이 있을 것이고, 그중에서 캐시 정책(cache policy)과 관련된 연구가 흥미로울 것”이라는 답을 들었습니다. 단일 CPU 코어(CPU core)의 클럭(clock)을 증가시키는 것이 현실적으로 어려워진 지금의 상황에서 시스템의 성능을 크게 좌우하는 메모리 계층 구조(memory hierarchy)를 개선하는 것은 지연(latency) 및 OS 보안 커널(OS kernel security)과 같이 여러 가지 어려움이 예상되더라도 확실히 큰 성능 향상을 기대할 수 있는 연구 방향입니다. 이번 ICML에서는 구글에서 진행 중인 관련 연구가 발표되었는데<sup>\*31</sup>, 캐시 예측(cache prediction)을 N-gram 문제로 보고 LSTM(long short-term memory)을 이용하여 캐시 미스 주소(cache miss address)의 차이(delta)를 입력 데이터로 모델을 학습하는 구조를 제안했습니다. 이 연구는 발표자 스스로도 말했던 것처럼 해답보다는 질문을 많이 얻게 된 연구였습니다. 하지만 한편으로 이 연구는 딥러닝과 하드웨어의 결합이라는 새로운 방향을 제시한 흥미로운 연구라고 생각합니다.

머신러닝과 하드웨어를 접목하는 연구는 2001년에 신경망(neural network)과 의사결정 나무(decision tree)를 이용한 분기 예측(branch prediction)에 대한 연구<sup>\*32</sup>와 2008년에 진행됐던 강화 학습을 이용하여 DRAM 메모리 제어기(DRAM memory controller)를 개선하는

<sup>\*29</sup> 논문 | Mirhoseini, A., Pham, H., Le, Q.V., Steiner, B., Larsen, R., Zhou, Y., Kumar, N., Norouzi, M., Bengio, S. & Dean, J., (2017). Device Placement Optimization with Reinforcement Learning. Proceedings of the 34th International Conference on Machine Learning, in PMLR 70:2430-2439

<sup>\*30</sup> 논문 | Kraska, T., Beutel, A., Chi, E. H., Dean, J., Polyzotis, N. (2018). The Case for Learned Index Structures. In Proceedings of the 2018 International Conference on Management of Data (SIGMOD '18). ACM, New York, NY, USA, 489-504. DOI: <https://doi.org/10.1145/3183713.3196909>

<sup>\*31</sup> 논문 | Hashemi, M., Swersky, K., Smith, J., Ayers, G., Litz, H., Chang, J., Kozyrakis, C. & Ranganathan, P., (2018). Learning Memory Access Patterns. Proceedings of the 35th International Conference on Machine Learning, in PMLR 80:1919-1928

<sup>\*32</sup> 논문 | Jimenez, D. A., Lin, C. Dynamic branch prediction with perceptrons. In HPCA, pp. 197-206, 2001

연구<sup>\*33</sup>에서 볼 수 있는 것처럼 최근 딥러닝이 발전되기 전부터 활발하게 진행되고 있었습니다. 딥러닝 모델의 발전이 하드웨어 분야에도 적용된다면 이전의 연구에서 한발 더 나아갈 수 있는 계기가 될 것입니다.

이렇게 머신러닝과 컴퓨터 하드웨어를 함께 고민하는 연구에 대해 최근 많은 관심이 모이고 있습니다. 구글에서 발표한 TPU(tensor processing unit)와 같은 특정 도메인 아키텍처(domain-specific architecture)도 이러한 연구 중 하나라고 할 수 있습니다. 제프 딘에게 이와 관련하여 “TPU처럼 머신러닝 모델을 위한 또 다른 하드웨어 개발 방향은 어떤 것이 있을까?”라는 두 번째 질문을 했을 때, “TPU와 같이 범용성과 특수성을 모두 갖고 있는 또 다른 응용 분야를 찾기가 당장은 어려워 보이나, 데이터 센터에서 사용되고 있는 TPU와 같은 선형대수가속기(linear algebra accelerator)를 저전력 종단 장치(edge device)에서 사용할 수 있도록 하는 것을 생각해볼 수 있을 것”이라는 답을 들을 수 있었는데, 이 역시 최근 구글에서 발표한 에지 TPU(Edge TPU)<sup>\*34</sup>를 염두에 둔 발언이 아니었을까 생각합니다.

머신러닝을 이용해 경험적인 해법을 개선하는 방법 중 하나로, 경험적인 해법을 차선책(sub-optimal expert)으로 생각하고 이를 뛰어넘는 주체(agent)를 학습하는 모방 학습(imitation learning)을 고려해볼 수 있습니다. 이번 ICML의 한 튜토리얼<sup>\*35</sup>에서 다뤄진 모방학습의 다양한 연구와 그 응용은 강화 학습과 상호 보완적인 역할을 통해 경험적인 해법을 향상시킬 수 있는 머신러닝 모델의 개발에 대한 기대를 높였습니다.

## 마치며

학회에 참석하면 즐거운 일 중 하나는 비슷한 관심사를 가진 새로운 분들을 만나볼 수 있다는 것입니다. 하지만 큰 규모의 학회장에서는 아는 사람을 찾기도 쉽지 않은데요, 이번 ICML에서는 학회에서 제공한 편리한 모바일 앱을 통해 다양한 모임방이 열려서 그러한 어려움을 덜 수 있었습니다. 이번 ICML에 참석한 물리학 전공자 혹은 물리학에 관심이 많은 머신러닝 연구자들은 모바일 앱을 이용하여 의견도 교환하고 함께 점심 모임을 가지기도 하였습니다.

제가 참석했던 2016년도 ICML과 비교해볼 때, 올해의 ICML은 새로운 돌파구를 여는 연구 성과보다는 머신러닝 분야의 성숙된 분위기를 반영하는 데에 더욱 중점을 둔 것같이 여겨졌습니다. 올해 주요 학회 중 하나인 NIPS도 이와 같은 모습을 이어갈지, 아니면 깜짝 놀랄 만한 새로운 연구 결과가 공개될지 사뭇 기대가 됩니다.

\*33 논문 | Ipek, E., Mutlu, O., Martinez, J. F., Caruana, R., Self-Optimizing Memory Controllers: A Reinforcement Learning Approach, 2008 International Symposium on Computer Architecture, Beijing, 2008, pp.39-50. doi: 10.1109/ISCA.2008.21

\*34 참고 | <https://cloud.google.com/edge-tpu>

\*35 참고 | Imitation Learning, Y. Yue and H. M. Le, ICML 2018



〈2018 ICML을 통해 살펴보는 AI 연구 동향〉 브런치로 연결되는 QR 코드입니다.

# 2018 CVPR 논문 동향 및 주요 연구 소개

'컴퓨터 비전 및 패턴 인식 학회(IEEE Conference on Computer Vision and Pattern Recognition, CVPR)'는 4차 산업혁명 및 인공지능에 대한 관심에 힘입어 급격하게 성장하고 있는 컴퓨터 비전 분야의 명성 있는(Top-tier) 학술대회이다. 특히 자율주행이나 의료영상 분석 등에 대한 연구들이 활발하게 발표되면서 더욱 큰 관심을 모으고 있다. 등록 인원은 매년 20% 이상 증가하고 있고, 투고되는 논문 수 또한 필자가 마지막으로 참석했던 2015년에 비해 1.5배 이상 증가하여 올해는 3309편이 제출되는 등 CVPR의 인기가 매우 뜨겁다.<sup>\*1</sup>

학회의 덩치가 커진 만큼, CVPR를 충분히 즐기기 위해서는 미리 준비해야 할 부분들이 상당히 많아졌다. 많은 수의 논문이 발표되기 때문에 관심 있는 논문들을 미리 체크하는 데에도 꽤 많은 시간을 들여야 했고, 함께 개최되는 워크숍도 무려 48개나 되었기 때문에 어떤 워크숍에 참석할지 결정하는 것도 쉽지 않았다. CVPR의 엄청난 인기를 실감하듯, 올해는 점심시간조차 아껴가면서 포스터 세션(poster session)에 참석해야 하는 강행군이었다. 기업체들의 전시 부스는 전보다 커졌고, 인재 영입도 치열하게 이루어져 전쟁터와 다름이 없었다.

최근 몇 년 사이 한국에서 투고하여 발표되는 논문 숫자가 많아지면서, 올해는 미국, 중국에 이어 한국이 세 번째로 참가자가 많은 국가<sup>\*\*</sup>였다. 이러한 열기를 반영하듯 올해도 CVPR에 참석하는 한국인들의 페이스북(Facebook) 그룹 채팅방이 생겨났고, 비행기에서 우연히 친구들을 만나기도 했다. 첫날 학회장에서는 등록 데스크의 줄이 학회장을 둘러 약 300m 이상 늘어지는 진풍경이 벌어지기도 했다.

[ 그림 1 ] 학회장 풍경



\*1 참고 | <http://cvpr2018.thecvf.com/files/CVPR%202018%20Opening%202.pdf>

\*\*2 참고 | [https://www.cc.gatech.edu/~parikh/citizenofcvpr/static/slides/brown\\_welcome\\_%E5%A4%A7%E5%AE%B6.pdf](https://www.cc.gatech.edu/~parikh/citizenofcvpr/static/slides/brown_welcome_%E5%A4%A7%E5%AE%B6.pdf)

글 | 황순민 [jjang9hsm@gmail.com](mailto:jjang9hsm@gmail.com) KAIST에서 박사과정을 밟고 있는 공돌이. 로봇틱스 및 컴퓨터비전 연구실 권인소 교수님의 지도를 받으며 자율주행을 위한 인지 기술에 대해서 연구하고 있습니다. 이 외에도 여러가지 재미있는 컴퓨터 비전 문제에 폭넓은 관심을 가지고 있습니다. 연구를 위한 연구보다 정말 쓰일 수 있는 연구를 지향하며, 언젠가 세상을 바꾸는 데 조금이라도 도움이 되는 엔지니어가 되고 싶다는 꿈을 꿉니다.

**연구 경향**

이번 CVPR에서 발표된 논문들의 제목을 분석해보면 ‘학습(Learning)’이라는 키워드가 돋보인다. 즉 이제는 많은 컴퓨터 비전 문제들이 학습 기반으로 해결책을 찾아가고 있다는 것을 의미한다. 또한 물체 검출(object detection)이나 의미론적 분할(semantic segmentation), 거리/자세 추정(depth/pose estimation), 초해상도 영상 변환(super-resolution), 주석 생성(captioning) 등 다양한 문제들에 대한 연구가 이루어지고 있었다.

학습기반으로, 특히 딥러닝(deep learning)을 이용하여 이러한 컴퓨터 비전 문제를 해결하기 위해서는 많은 데이터가 필요하다. 이에 따라 교사학습(supervised learning)에 필요한 정답 데이터에 대한 부담으로 인하여, 비교사학습(unsupervised, self-supervised)이나 약한교사학습(weakly-supervised) 등이 활발하게 연구되고 있다. 또한 약 2, 3년 전부터는 최신 게임 제작에 활용되는 컴퓨터 그래픽 엔진을 활용하여 실제와 같은 영상을 합성(synthesis)하고 이를 인공지능 학습에 활용하려는 시도들이 활발해지고 있다. 그러나 이런 접근은 합성 데이터와 실제 데이터의 미묘한 차이에서 오는 도메인 이동(domain shift) 문제를 야기하였고, 이를 해결하기 위한 도메인 적응(domain adaptation) 또는 전이 학습(transfer learning)과 같은 연구도 많아지고 있다.

논문 제목에서는 잘 나타나지 않았지만, 최근 많이 연구되고 있는 방향으로는 (1) 기존에 인간이 경험적으로 디자인한 모델을 데이터 주도적으로 학습된 인공지능 모델로 대체하는 것(deep learning approach)과 (2) 여러 종류의 입력 데이터를 함께 사용하는 것(multimodal learning), (3) 불안정한 정답 데이터를 활용해 학습하는 것(weakly-/partially supervised learning)이 있다. 이와 관련된 흥미로운 연구 몇 편을 소개한다.

**인상 깊었던 연구 소개**

**1) [Intel] 어둠 속에서 보는 법 배우기(Learning to See in the Dark<sup>\*3</sup>)**

이 연구에서는 몇 단계로 구성된 기존의 카메라 영상 처리 파이프라인(pipeline)을 인공지능망으로 대체하기 위한 방법을 제안한다. 기존 방법으로는 조도가 매우 낮은 환경에서 빠른 속도로 촬영하는 경우 아주 어둡거나 잡음이 매우 심한(noisy) 영상을 얻을 수밖에 없었는데, 이 연구에서 제안하는 인공지능망 기반의 파이프라인을 이용하면 낮은 조명 환경에서도 고품질의 영상을 획득할 수 있음을 보였다. 이 연구는 고 명암비 영상(high dynamic range image, HDR)이나 움직이는 물체 등에 대한 고려가 없는 등 몇 가지 한계점이 있지만, 인공지능망을 영상 획득 과정의 초기 단계부터 활용하면 고품질의 영상을 얻는 데 매우 유용할 수 있음을 보여주는 중요한 연구이다.

\*3 참고 | Chen, C. et al. (2018). Learning to See in the Dark, CVPR.

[그림 2] CVPR 2018 발표 논문 제목에서의 단어 빈출 현황<sup>\*4</sup>



※범례: 글자의 크기와 빈출 정도는 비례함.

[그림 3] 인공지능망 기반의 파이프라인을 이용한 영상 처리<sup>\*3</sup>

**1) 기존의 카메라 영상 처리 파이프라인과 제안된 방법의 결과 비교**

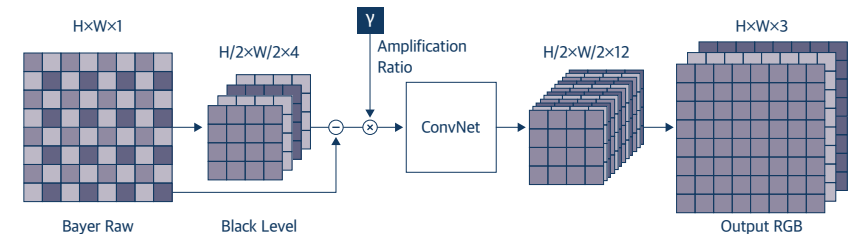


(a) Camera output with ISO 8,000 (b) Camera output with ISO 409,600 (c) Our result from the raw data of (a)

**2) 카메라 내부에 활용된 전통적인 방법**



**3) 논문에서 제안한 인공지능망을 이용한 영상 처리 파이프라인**



\*4 참고 | https://worditout.com



2) [MIT] 무선 신호를 이용하여 벽으로 가려진 사람의 자세 추정(Through-Wall Human Pose Estimation Using Radio Signals<sup>5</sup>)

이 연구는 무선 신호(radio frequency signal)로부터 사람의 자세를 추정하는 인공지능경망을 소개하고 있다. 무선 신호는 사람이 직관적으로 이해할 수 없기 때문에 교사학습(supervised learning)을 위한 정답을 직접 들어줄 수가 없다. 이 연구에서는 영상으로부터 사람의 자세를 추정하는 최신 모델(Open Pose)<sup>6</sup>을 교사 네트워크(teacher network)로 활용하여, 동기화된 무선 신호로부터 같은 사람의 자세를 추정하도록 하는 교차 감독(cross-modal supervision)을 통해서 학생 네트워크(student network)를 학습한다. 학습 이후에는 무선 신호만을 사용한 사람의 자세를 추정하는 것이 가능하데, 사용된 무선 신호는 벽을 가로지르기 때문에 벽 뒤에 있는 가려진 사람의 자세 또한 추정할 수 있게 된다.

3) [Facebook AI Research] 모든 것을 분할하는 법 배우기(Learning to Segment Everything<sup>7</sup>)

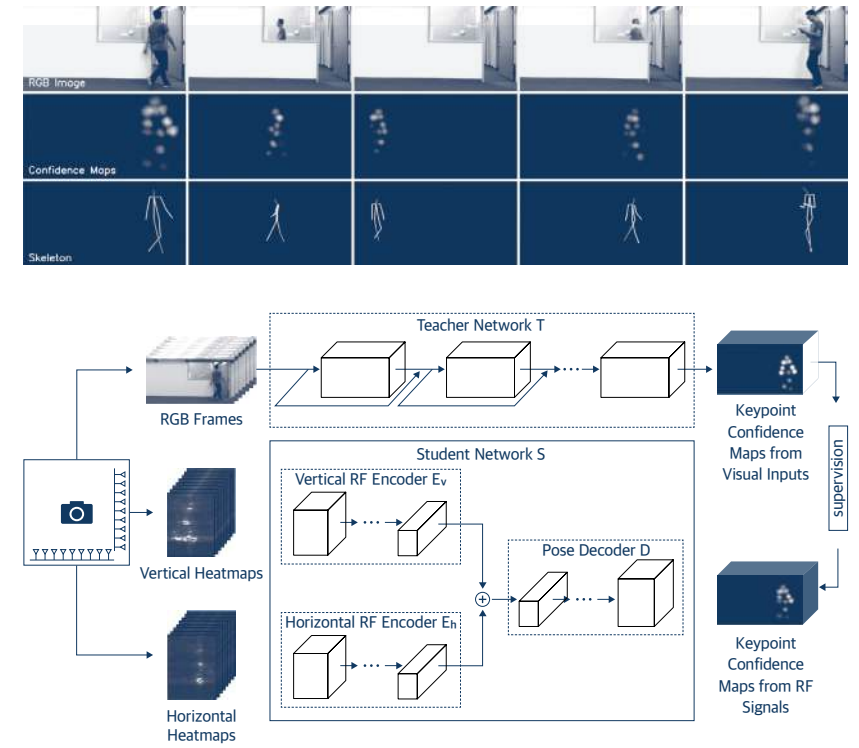
이 연구에서는 물체 분할(object instance segmentation) 문제를 학습하기 위한 데이터가 부분적으로 주어진 경우(partially supervised training)를 다루고 있다. 물체 분할을 위해서 활용되는 유명한 모델인 Mask R-CNN<sup>8</sup>의 학습을 위해서는 영상 내 물체의 위치 정보를 나타내는 상자 정보(bounding box) 및 픽셀(pixel) 단위의 마스크 정보(instance mask)가 모두 필요하다. 이 연구에서는 물체의 상자 정보와 마스크 정보가 모두 있는 카테고리 A의 데이터를 이용하여 상자 가중치(box weights)로부터 마스크 가중치(mask weights)를 추정하는 가중치 전달 함수(weight transfer function)를 학습한다. 이를 이용하여 마스크 정보가 없는 카테고리 B에 대해서도 마스크 정보를 추정해낼 수 있다. 많은 카테고리를 다루기 위해서 학습에 필요한 마스크 정보를 모두 제공하는 것은 막대한 비용이 들기 때문에, 완전한 지도 없이 모델을 학습하기 위한 방법을 고안하는 것이 매우 중요하다는 점에서 이 연구는 큰 의미가 있다.

조언에서 주언으로 워크숍

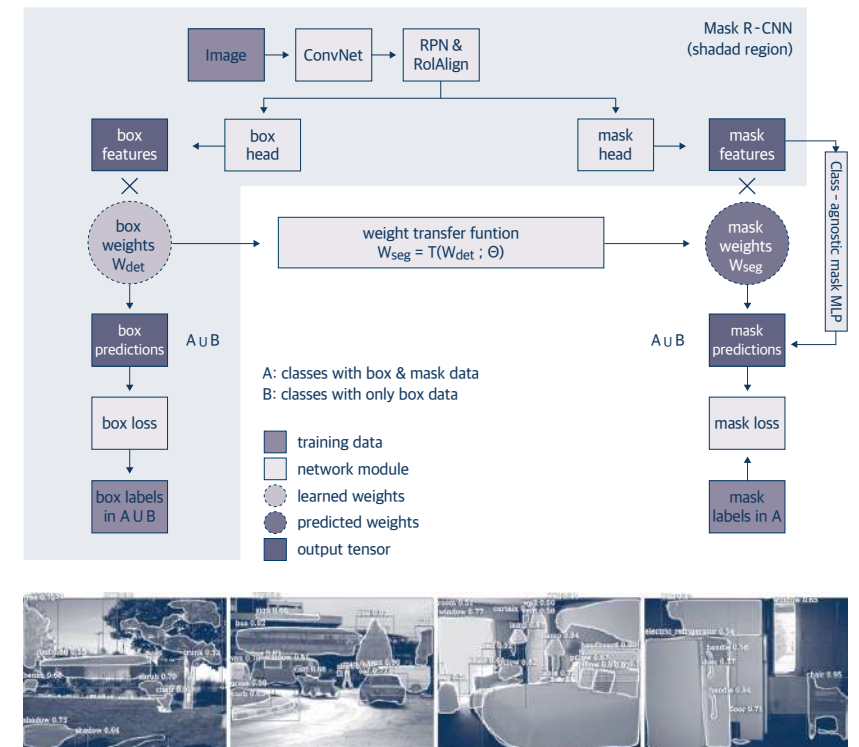
올해는 무려 48개의 워크숍이 함께 개최되면서 CVPR에 참석한 수많은 연구자들의 발길을 잡았다. 과거의 워크숍들이 초청 강연 위주로 진행되어 상대적으로 많은 관심을 받을 수 없었던 것과 달리 최근 CVPR과 함께 개최되는 워크숍들은 미래지향적이고 구체적인 문제를 다양한 방식으로 다루으로써 큰 인기를 끌고 있다. 특히 요즘 워크숍들은 자체적으로 새로운 데이터 세트를 제공하고 경진대회를 여는 경우도 많아 다양함을 더한다. 이번 CVPR에서는 메인 학회 일정의 첫 순서로 기계학습 분야에서의 경연대회 역할과 워크숍에서 진행되는 몇몇 경연대회를 소개하는 특별 세션을 가질 정도로, 이제는 워크숍/경연대회가 학회의 주언으로 발돋움하고 있다.

구체적으로 '미세 분류 (Fine-grained Visual Categorization)<sup>9</sup>', '행동 인식 (ActivityNet Large Scale Action Recognition Challenge)<sup>10</sup>', '비디오 물체 분할 (DAVIS Challenge on Video Object Segmentation)<sup>11</sup>', '자율주행을 위한 강인한 시각 인지 경연대회(Robust Vision Challenge)<sup>12</sup>' 등 다양한 문제에 대한 경연대회가 진행되었다. 분야의 특성상 주어진 문제에 대해

[그림 4] 단일 객체에 대한 테스트 예제(위)와 RF-Pose 모델 구조(아래)<sup>5</sup>



[그림 5] 제안하는 방법의 구조(위)와 결과 예시(아래)<sup>7</sup>



※ 초록색 상자는 학습 시 마스크 정보가 활용된 경우, 빨간색 상자는 상자 정보만으로 학습된 경우(실제 이미지는 브러시를 참고)

<sup>5</sup> 참고 | Zhao, M. et al. (2018). Through-Wall Human Pose Estimation Using Radio Signals, CVPR  
<sup>6</sup> 참고 | Cao, Z. et al. (2017). Realtime multi-person 2D pose estimation using part affinity fields, CVPR  
<sup>7</sup> 참고 | Hu, R. et al. (2018). Learning to Segment Every Thing, CVPR  
<sup>8</sup> 참고 | He, K. et al. (2017). Mask R-CNN, ICCV  
<sup>9</sup> 참고 | https://sites.google.com/view/fgvc5/home  
<sup>10</sup> 참고 | http://activity-net.org/challenges/2018  
<sup>11</sup> 참고 | https://davischallenge.org/challenge2018  
<sup>12</sup> 참고 | http://www.robustvision.net

얼마나 높은 성능을 달성하는가에 대한 연구는 연구자들뿐만 아니라 기업에서도 관심이 많기 때문에, 다양한 기업들이 각 워크숍/경연대회를 후원하고 많은 사람들이 도전하는 선순환 구조가 만들어지고 있다. 이런 다양한 벤치마크, 경연대회가 진행되면서 해당 문제에 대한 최신 성능이 매우 빠르게 발전하고 있기 때문에, 같은 문제를 고민하는 연구자들은 해당 워크숍을 눈여겨보는 것이 좋을 듯하다.

### 네트워킹(networking)

전시 부스들도 해마다 늘어나고 있는데, 최근에는 기업들이 더욱 공격적으로 인재 영입에 나서고 있는 것을 느낄 수 있었다. 몇몇 회사들은 학회장에 별도의 회의실을 마련하여 현장에서 엔지니어와 인터뷰를 진행하는 경우도 있었고, 공식적으로 진행되는 부스 전시 외에 비공개 기술 강연 등을 통해 기술력을 과시하기도 했다. 비공개 기술 강연이 있는 회사의 경우 전시 부스에서 강연에 초대되는 기회를 얻을 수도 있었다.

저녁에는 구글(Google), 애플(Apple), 페이스북(Facebook), 아마존(Amazon) 등의 북미 회사들뿐 아니라 텐센트(Tencent), 화웨이(Huawei), 바이두(Baidu) 같은 중국 기업들도 네트워킹 파티를 열었다. 많은 연구자들이 파티에 참석해서 음료와 식사를 즐기며 사교활동을 할 수 있었는데, 이런 서양식 파티가 익숙하지 않은 필자는 꽤나 어색함을 느끼기도 했다. 이런 어색함을 이겨내고 잘 어울린다면 학회에서 인맥을 넓히는 데 도움이 될 수도 있을 것이다.

### 마치며

OpenAccess를 지향하는 arXiv<sup>\*13</sup> 덕분에 원한다면 얼마든지 공부할 수 있는 세상이다. 그러나 이런 세계적인 학회에 직접 참석해서 그 분위기를 느끼고, 적어도 그곳에 있는 며칠만큼은 연구 생각만으로 가득 차는 경험은 온라인에서 쉽게 할 수 없다. 현장에서 공부한 다양한 연구 내용들을 바탕으로, 함께하는 동료들 또는 학회에 참석한 연구자들과의 교류가 새로운 아이디어로 이어지는 경험을 할 수 있다면 연구자로서 매우 큰 동기부여가 될 것이다. 앞으로도 더 많은 한국인들이 이런 세계 무대를 적극적으로 경험하고 자극받아 4차 산업혁명 시대를 선도할 수 있기를 희망해본다.

\*13 참고 | <https://arxiv.org/list/cs.CV/pastweek>



〈2018 CVPR 논문 동향 및 주요 연구 소개〉 브런치로 연결되는 QR 코드입니다.

# 카카오 크루들의 커피 주문 방식

글 | 최은필 reyna.c@kakaocorp.com IT 업계의 에반젤리스트(evangelist)를 자칭하며 빠르게 변화하는 시대에 그 누구보다 부지런히 움직이며, 좋은 것들을 나누며 살고 있습니다. 재즈 힙합(Jazz Hip-hop)을 즐겨 들으며, 언젠가 추리소설을 쓰겠다는 원대하지만 소박한 꿈을 가지고 있습니다. 전자공학을 전공했지만 실험보다는 글쓰기를 더 좋아하며, 오랜 시간 '전략, 기술, 기획'이라는 단어들과 함께하였습니다. 전 세계 수많은 사람들의 마음을 움직여 보고 싶다는 커다란 꿈을 가지고, 하루하루 즐겁고 재미있게 살고 있습니다.

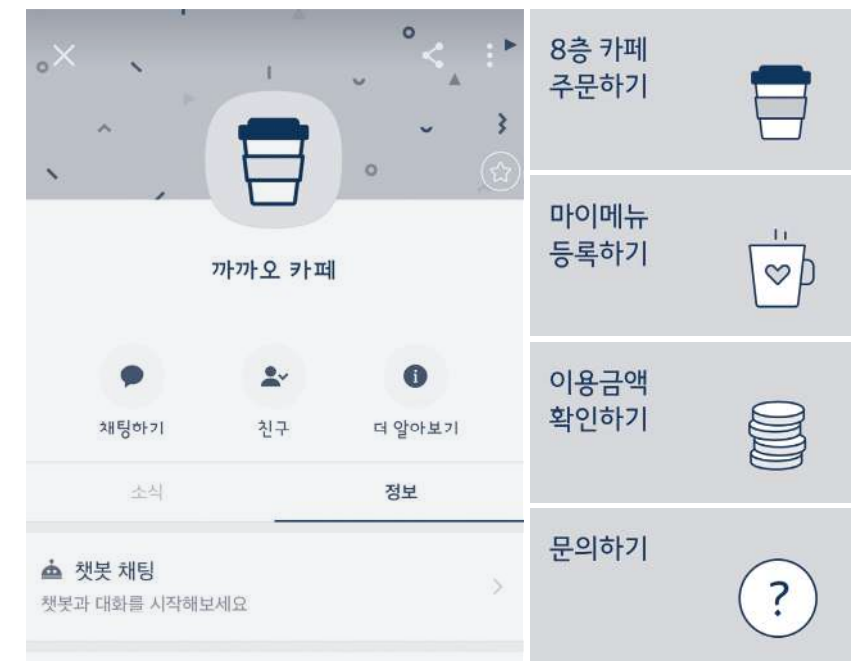
## 까까오 카페

“카페에 줄을 서서 기다리는 시간을 줄일 수 있을까?”  
 “주문을 편하고 쉽게 할 수 있는 방법이 있을까?”  
 “챗봇이 해결할 수 있지 않을까?”

카카오 사내 카페에서는 이제 더이상 길게 줄을 서서 음료를 주문하고 기다리는 데 시간을 보내지 않아도 됩니다. 바로 사내에서 사용하고 있는 '까까오 카페' 챗봇 덕분입니다. 출근 시간과 점심시간 등 개인 시간에 맞춰서, 사무실이나 혹은 카카오톡을 사용할 수 있는 어디에서나 카카오 사내 카페에 주문을 넣을 수 있습니다. 과연 '카카오 카페'도 아닌 '까까오 카페'의 정체는 무엇일까요? 누구나 사용할 수 있는지, 일반 카페에서도 사용할 수 있는지, 또 주문과 결제는 어떻게 하는 것일까요? <카카오 시리포트> 리뉴얼 호의 마지막 시간으로 카카오 크루(Krew)들의 카페 이용 방식과 스타일을 알아보았습니다.

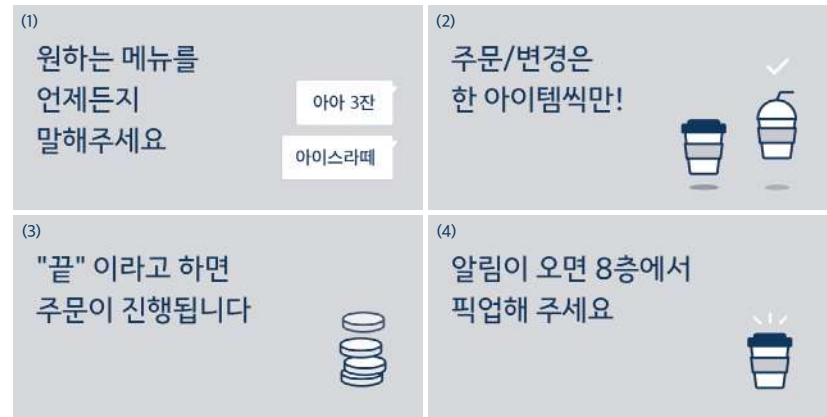
카페 챗봇은 과연 누가, 왜, 어떻게 만들게 되었을까요? 소문에 의하면 카페챗 탄생이 개발자 특유의 '귀차니즘'에서 시작된 것이라고 하는데 과연 어떻게 된 일일까요? 사람들이 붐비는 아침 출근 시간대와 점심시간에 카페까지 걸어가서 주문하고, 길게는 15~20분까지 기다려야 하는 경우가 많습니다. 이런 광경은 카카오 사내 카페뿐만이 아니라, 전 세계 모든 카페에서 흔히 볼 수 있는 광경입니다. 특히 아침시간이나 바쁜 일정에는 카페에서 음료를 기다리는 시간조차 아까운 경우도 있고, 주문하고 기다리는 일이 굉장히 귀찮게 느껴지기도 합니다. 이러한 상황이 너무나도 불편하고 귀찮게 느껴졌던 카카오 개발자 크루들, 해결 방법을 찾아보자는 취지에서 '까까오 카페'라는 챗봇 개발 프로젝트가 시작되었다고 합니다.

[그림 1] '까까오 카페' 카페챗의 메뉴 화면



챗봇의 원리와 사용방식을 간단하게 설명드리면 다음과 같습니다. 카카오톡을 이용해 '까까오 카페'라는 챗봇을 실행하고, 챗봇을 이용해 간단하게 주문을 입력하고 내역을 확정하면 즉시 카카오 사내 카페로 주문이 통보됩니다. 카페에서는 챗봇을 통해 들어온 주문에 따라 음료를 제조하고, 준비가 완료되면 주문자에게 음료가 준비되었다고 알림을 보냅니다. 카페 사용자가 직접 대면으로 주문하고, 확인하고 결제까지 해야 하는 시스템이 아닌, 카카오톡의 챗봇을 통해 편리하게 주문하고 픽업하는 방식입니다.

[ 그림 2 ] 까까오 카페 챗봇을 이용한 주문 방법

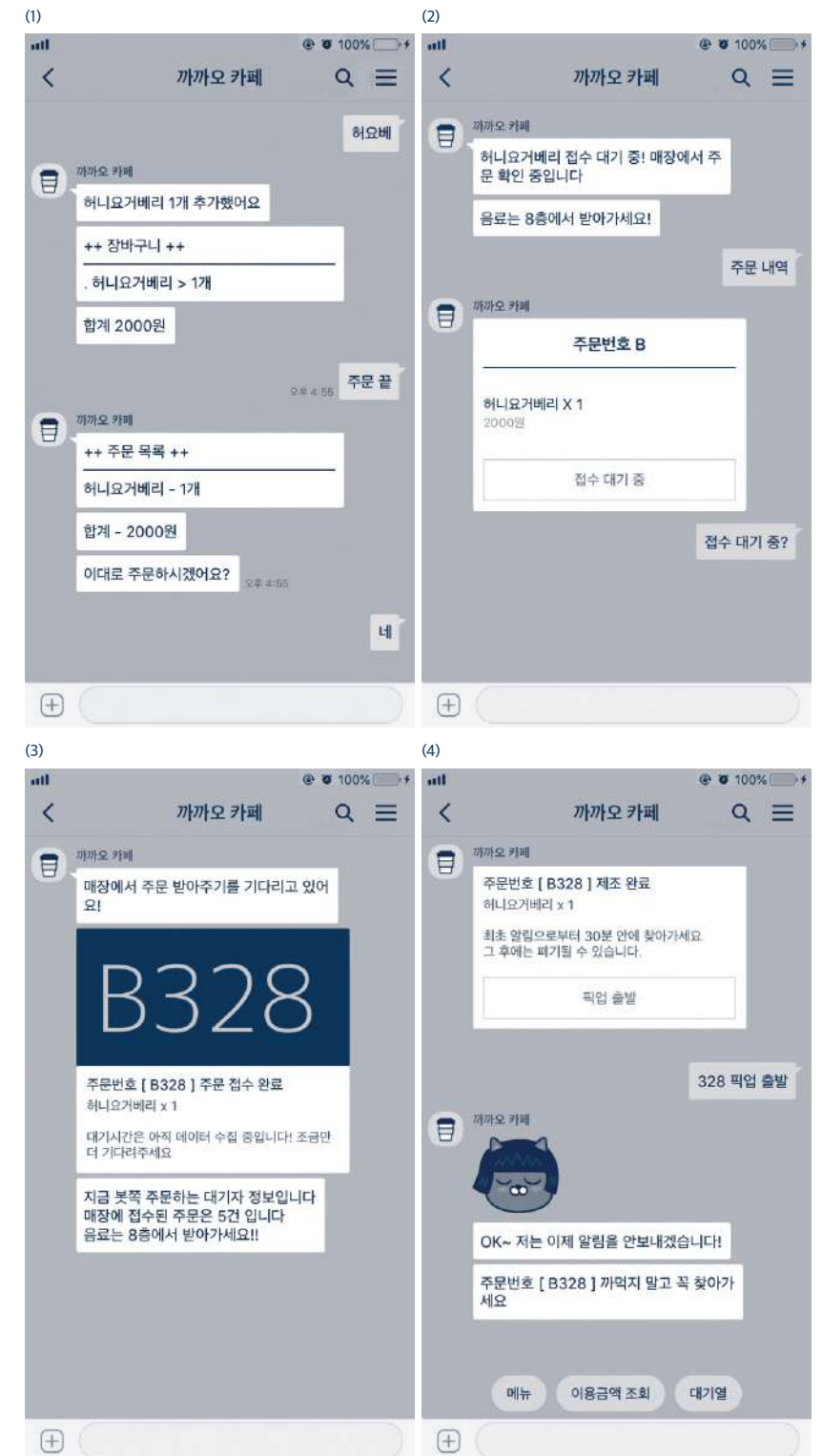


그렇다면 '까까오 카페' 챗봇은 과연 사용자들에게만 편리할까요? 서비스 측면에서 살펴본다면 주문자가 직접 주문을 입력하고 결제가 진행되는 셀프 오더(self-order) 시스템이기 때문에 카페 이용자와 매장 운영자 양쪽 모두에게 편리한 시스템이라고 할 수 있습니다.

현재 스타벅스(Starbucks)의 사이렌 오더(siren order)와 같은 원격 주문 시스템도 있지만, 이 방식은 별도의 앱을 실행시켜야 하고 앱을 부팅하는 데 시간이 오래 걸린다는 불편함이 있습니다. 하지만 '까까오 카페' 챗봇은 카카오톡에서 바로 주문이 가능하다는 장점이 있는데, 실제 카페봇을 사용한 주문을 살펴보도록 하겠습니다.

카페봇은 상당히 똑똑한 주문 처리 능력을 가지고 있습니다. 아아(아이스 아메리카노), 뜨아 or 따아(따뜻한 아메리카노) 같이 단축어를 말해도 정확한 주문 명령어를 입력합니다. 이것을 가능하게 만든 기술이 바로 카카오의 자연어 처리 기술입니다. 특히 대면 주문이 아닌 카페봇 주문에서는 자연어 처리 기술이 상당한 강점을 가지는데, 카카오 사내 카페에 있는 메뉴인 '허니요거베리' 같은 메뉴들도 '허요베'로 간단하게 축약해서 주문할 수 있습니다.

[ 그림 3 ] 카페봇을 이용한 실제 주문 화면



### 까까오 카페봇의 뒷이야기

카페봇을 탄생시킨 이들은 과연 어떤 분들일까요? 실제 카페봇 개발에 참여했던 두 분은 서로 다른 조직에서 일을 하고 있는데, 게임 동호회에서 논의된 ‘귀차니즘’이 이 모든 일의 시작이었다고 합니다. 오버워치 동호회에서 시작된 인연이 카카오 카페의 이용을 획기적으로 변화시킨 출발점이 되었던 거죠.

‘까까오 카페’의 소소하지만 거대한 프로젝트는 ‘카페까지 걸어가서 주문하고 기다리는 것이 너무 귀찮다’라는 것에서 시작되었습니다. 처음에는 개인적인 질문에서 시작된 프로젝트가 사내 카페 리뉴얼 일정과 맞물리면서, ‘카카오 사내 카페 전면 개편’이라는 전사 프로젝트로 발전하게 된 것입니다. 결국 작은 아이디어에서 시작된 물음이 전사적으로 발전하며 카카오 전체 크루들의 시간과 비용을 줄이는 데 엄청난 기여를 하게 된 것입니다.

하지만 항상 모든 성공 스토리에 훈훈한 얘기만 있지는 않겠지요? 가볍고 소소하게 시작했던 카페봇 개발 과정은 생각보다 순탄치 않았습니다. 개발 과정에 있었던 몇 가지 에피소드를 알아보았습니다. 사실 카페봇 개발 초기에는 디자인이 지금처럼 다듬어진 형태는 아니었습니다. 그저 재미로 개발자분들께서 동그라미 하나, 직선과 곡선 하나씩 그려놓은 수준이었습니다. 하지만 카페봇의 설계가 어느 정도 마무리되고 카페봇 대표 이미지와 사투를 벌이던 개발자분들을 옆에서 지켜보던 디자이너께서, 그런 모습에 더는 참지 못하고 순식간에 똑딱 카페봇 이미지를 그려냅니다. 그렇게 나오게 된 카페봇 디자인이 현재 사용되고 있는 버전인데, 카페봇에게 멋진 옷을 선사한, 훌륭한 일을 하신 거죠.

사내의 카페 리노베이션 일정과 맞물리면서, 개인적인 흥미와 불편함을 없애자는 취지에서 재미로 시작했던 프로젝트에 ‘데드라인’이라는 것이 생겼네요. 카페봇 개발에 필요했던 것은 개발자분들의 손과 컴퓨터 그리고 POS(point of sales)기와 영수증 인쇄기였습니다. 특히 개발에 가장 어려움이 많았던 것이 POS기였다는데, 같이 협업했던 POS기 개발 담당자분들이 계속 변경되면서 커뮤니케이션에 많은 어려움을 겪었다고 합니다. 그러다 실제 서비스가 오픈되기 이틀 전 저녁(7월의 어느 주말 저녁이라고 합니다) 책상 위에 놓여 있던 영수증 인쇄기에서 구매 영수증이 인쇄되는 역사적인 순간을 맞이하게 됩니다. 지면으로 전달하는 것이라 당시의 감동과 희열을 그대로 전해드리지는 못하지만, 영수증 인쇄가 이렇게 어렵고 감동적인 것인지 처음 느꼈다고 합니다.

카페봇을 기획하고 방법을 고민하고, 실제 사내 카페에 적용하기 위해 힘써주신 담당자 분들께(개발: 이철민, 전현우/ 도입·활용: 최솔) 몇 가지 질문을 드려봤습니다.

[ 그림 4 ] 카페봇 개발 테스트 영수증



Q. 카페봇 개발 과정에 대한 소감을 말씀해주세요.

A. 지금은 지난날을 회상하면서 웃으며 말할 수 있지만, 최초 영수증이 인쇄되기 전까지 수많은 시행착오들과 카페 오픈 일정을 맞추기 위해 아꼈었던 순간도 있었습니다. 아쉬운 점이라고 하면 아직 담아내지 못한 서비스가 있습니다. ‘오늘은 제가 쏘니다’와 같은 골든벨 기능, ‘커피 사주세요’와 같은 조르기 등의 기능인데 이 기능은 향후에 반영될 수 있지 않을까 생각합니다.

Q. 카페봇 이름이 왜 ‘까까오 카페’인가요?

A. 사실 처음에 재미 삼아 시작한 일이었고 비공식 프로젝트였기 때문에 ‘카카오’라는 이름을 공식적으로 쓸 수 없었습니다. 그래서 재미있고 귀여운 어감으로 ‘까까오 카페’라고 이름 붙인 거죠.

Q. 영수증 인쇄 테스트를 하시면서 아꼈었던 순간이 있었다고 들었는데 어떤 사건이었나요?

A. 끊임없이 챗봇과 POS기, 영수증 인쇄기 연동 테스트를 진행했었습니다. 테스트 과정이 생각보다 순탄치 않았습니다. 카카오 사내 카페에서는 사원증 태깅으로 카페 구매 내역이 결제되는데, 테스트 과정에서 재미 삼아 밀크티 1656잔, 카페라떼 3500잔, 자몽에이드 5000잔을 테스트했던 적이 있습니다. 합계 가격만 4680만 원 정도가 되는데, 영수증 인쇄는 성공했지만 카페에 실제 주문으로 입력되어 결제된 것이 아닐까라는 생각에 순간 등에 식은땀이 흐르기도 했습니다. 사내에서는 ‘만수르’가 나타났다면 많은 분들이 구경하러 몰려오기도 했지요.

Q. 카페봇의 가장 큰 성공 요인은 무엇일까요?

A. 가장 큰 특징은 실제 사용하는 사람들의 필요에 의해서 모두가 필요로 하고 도움이 되는 방향으로 만들어진 것이고, 카페를 사용하는 이용자와 카페를 운영하는 사람들 모두가 필요로 하는 것을 붓으로 구현한 것이라고 생각합니다. 아직 주문하기 명령어와 UI/UX 등의 개선과 보완이 필요하지만 이는 계속해서 발전할 것이라고 생각합니다.

Q. 현재 붓은 1세대로 생각됩니다. 향후 2세대, 3세대 붓은 어떤 모습으로 변화할까요? 생각하시기에 미래의 붓은 어떻게 정의할 수 있을까요?

A. 제 생각에는 자비스(영화 <아이언맨>에 나오는 인공지능)의 모습을 가지지 않을까 생각합니다. 저는 플랫폼 개발자로서, 플랫폼을 잘 이해하고 있는 사람이 붓을 만들었다는 것은 결국 기술개발 순환의 좋은 사례를 보여준 것이라 생각합니다.

### 카카오의 미래? 챗봇의 미래!

카카오 카페봇은 별도의 결제 프로세스와 장치가 필요 없고, 앱을 따로 설치할 필요도 없으며 긴 시간 줄을 서서 기다려야 하는 번거로움도 없습니다. 서비스를 이용하는 사용자와 서비스를 제공하는 운영자 모두에게 편리한 서비스라는 점이 가장 큰 장점이자 매력 포인트라고 생각합니다. 처음 프로젝트의 시작도 비효율적인 것을 바꿀 수 있을까?라는 질문에서 시작하였고, 그렇다면 서비스를 사용하는 모두에게 편리한 모습을 만들자는 취지에서 프로젝트가 전개되었습니다. 다른 누군가에 의해 서비스가 기획되고 개발된 것이 아닌, 지금 우리가 느끼는 불편함과 필요에서 거대한 여정이 시작된 것입니다. 그리고 '서비스가 어떻게 쓰이고, 누구를 위해 만들어져야 할까?'를 부단히 고민한 끝에 모두가 좋아하고 편리해질 수 있는 결과가 나올 수 있지 않았을까 생각합니다.

카카오 카페봇은 계속해서 진화하고 새로운 언어들을 배워가고 있습니다. 새로운 메뉴들과 다양한 주문 명령어를 이해하고 실행하는 것인데, 이것의 핵심이 바로 자연어 처리 기술입니다. 그리고 앞으로 다양한 기능들이 추가되고, 진화된 모습을 만날 수 있지 않을까 생각합니다. 현재 카카오 사내 카페에서 대면 주문 방식과 챗봇으로 주문하는 비대면 주문의 비율이 6:4 정도 된다고 합니다. 아마 앞으로는 지금보다 더 새로운 방식의 카페 이용과 주문 & 픽업 등의 모습이 펼쳐질 것으로 예상됩니다.



〈카카오 크루들의 커피 주문 방식〉 브랜치로 연결되는 QR 코드입니다.

#### COVER

〈카카오 시리포트〉의 표지에선 시와 관련된 의미 있는 코드들을 소개하고 있습니다

#### Vol.14 코드 | 박규병 [kyubyong.park@kakaobrain.com](mailto:kyubyong.park@kakaobrain.com)

카카오브레인의 박규병 연구원이 진행한 제주어 프로젝트(파일럿) 소스 코드 중 일부입니다. 제주어 화자의 구술을 받아쓰고 이를 표준한국어로 옮긴 병렬말뭉치인 《제주어구술자료집》을 학습 데이터로 삼았습니다. 최신 기계번역 모델 중 가장 뛰어난 성능을 보인 Transformer를 모델로 하였습니다. ([https://github.com/Kyubyong/jejueo/blob/master/make\\_parallel.py](https://github.com/Kyubyong/jejueo/blob/master/make_parallel.py))

## AI CODE vol.14

```
from glob import glob
import re
import codecs
from tqdm import tqdm
import xlrd
from jamo import h2j # syllable to jamo
from itertools import chain
from urllib.request import urlopen

def refine(text):
    text = re.sub("[.+]?", "", text)
    text = re.sub("(.+?)", "", text)
    return text

def load_hypua2jamo():
    url = "https://raw.githubusercontent.com/mete0r/hypua2jamo/master/data/hypua2jamocomposed.txt"
    lines = urlopen(url).read().decode("utf8").splitlines()

    hypua2jamo = dict()
    for line in lines:
        if line.startswith("U+"):
            try:
                hypua, jamo = re.split(" => ", line)
            except ValueError:
                print(line) # U+F86A =>
                continue
            hypua = chr(int(hypua[2:], 16))
            jamo = jamo.split("%")[0] # trim comments
            jamo = "".join(chr(int(j[2:], 16)) for j in jamo.split())
            hypua2jamo[hypua] = jamo
    return hypua2jamo

...
```