스포츠 저널리즘의 변화와 AI의 활용

알고리즘이 사람의 손을 거치지 않고 기사 작성에 필요한 모든 역할을 자동으로 수행하는 형태의 언론 보도를 로봇 저널리즘(robot journalism)이라고 부른다.¹¹ 데이터 속에서 수학적 최적화 과정과 기계 학습 등을 통해 상대적으로 의미 있는 이야깃거리를 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.

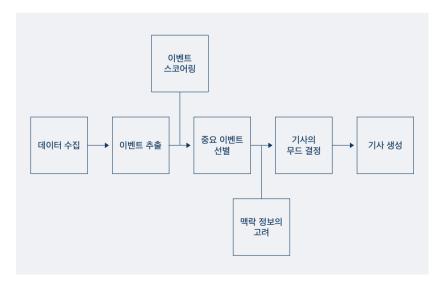
기사를 쓰는 알고리즘의 등장

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. '2 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. ¹³ 로봇 저널리즘 기술이 스포츠 분야에 처음 적용된 이유는 데이터가 풍부하게 축적돼 수학적인 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*4



동 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 *3 참고 | 프로야구 뉴스로봇, https:// www.facebook.com/kbaseballbot *4 특허 | 이준환 & 김동환(2017), 자 동 기사 생성을 위한 로봇 저널리 즘 시스템 및 방법(Robot Journalism Method and System for Automatic Article Generation), pg17, https:// patentimages.storage.googleapis. com/00/a2/84/80a56dbffbdf31/ KR20170058785A.pdf

*1 논문 | 김동환·이준환(2015), 〈로봇 저 널리즘: 알고리즘을 통한 스포츠 기사 자

글 l 김동환 dkim@lab2ai.com 학부와 석사에서 컴퓨터 공학과 HCl(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. In-Depth

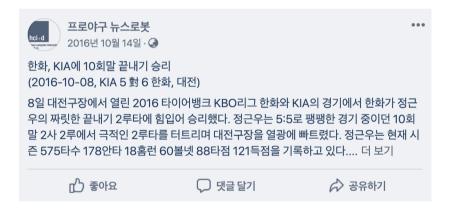
무엇이 기사가 되는가?

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 기록과특이사항 등 공시적이고 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*3



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. 5 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 자연스럽고 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.

94

^{*5} 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



●
●

</t