



책, 영화, 음식, 뉴스...

너의 취향을 추천할게

글 이정아 기자



뉴스
협업 필터링 기술과
인공신경망으로 관심 뉴스 추천

영화
협업 필터링 기술로 개인
맞춤형 추천

‘피피피카부, 피피피카부, 흥이 난 여우 그런 나라구~’. 요즘 가장 좋아하는 노래가 끝잠을 깨웠다. 잠을 깨우는 알람으로 이만한 게 없다. “아리아, 다음 노래 틀어줘”라고 말하자 인공지능(AI) 스피커가 유료 음원 제공 사이트를 연결해 다른 노래로 바꾼다. ‘내 혈관 속 DNA가 말해줘~ 내가 찾아 헤매던 너라는 걸~’. 버스에 앉아 스마트폰으로 평소 이용하는 소셜네트워크서비스(SNS)에 접속한다. 친구 추천에 ‘최지원’ ‘권예슬’ 이름이 뜬다. SNS는 이들이 내 동료라는 걸 안다. SNS 하단에는 ‘빅센 히어로즈’ 야구잡바를 사라는 광고가 뜬다. SNS는 내 ‘최애(가장 사랑하는다는 뜻)’ 팀도 이미 알고 있다. 사무실에 도착해 인터넷 포털 사이트를 검색한다. 과학뉴스가 쏟아진다. 과학기자라는 내 직업에 맞춘 뉴스 추천 서비스다. 내가 좋아하는 의류 브랜드 세일 정보도 뜬다. 한번이라도 검색한 브랜드는 놓치는 법이 없다. 점심 약속을 앞두고 맛집을 검색하자 비슷한 종류의 음식점이 같이 뜬다. 추천한 곳 중 하나를 선택하면 예약 끝. 내가 하루 종일 이용한 추천 서비스는 총 30개. 말하지 않아도 내 취향을 먼저 알고 추천해주는 너란 서비스, 이전 너 없이 못할 것 같다.

쇼핑
인공신경망으로
내가 좋아할 만한
상품 추천

도서
최다 검색

추천의 기술 1

마트와 서점은 그룹별 최다 검색

음악과 동영상에서부터 맛집, 이동 경로, 심지어 친구까지 추천받는 세상이다. 이는 내가 은연중에 남긴 ‘사이버 발자국’ 때문이다. 본인 동의를 얻은 개인정보(성별, 연령, 거주 지역 등) 나뉘어 검색, 구매 이력, 인터넷 서핑 경로 등 온라인에 남겨진 개인의 흔적은 모두 데이터로 변환된다. 이른바 빅데이터다.

빅데이터를 이용한 추천 기법 가운데 가장 간단한 것은 ‘최다 검색’을 기준으로 삼는 것이다. 가장 많이 재생한 음악, 가장 많이 검색한 음식점 등 다수가 원하는 정보를 최우선으로 추천하는 것이다. 건국대 경영학과에서 빅데이터를 연구하는 함유근 교수는 “최다 검색 결과는 모든 사람의 평균 수치”라면서 “구매로 이어질 확률이 랜덤 추천보다 높다”고 설명했다.

우선 성향이 비슷한 사람끼리 그룹으로 묶어, 그룹별로 상품을 추천할 수 있다. 이를테면 남성, 여성처럼 성별에 따라 그룹을 묶거나, 연령, 자주 찾는 품목 등 공통적인 특징이나 관심사를 이용해 그룹을 묶는다. 카드회사나 포털 사이트, 쇼핑몰 애플리케이션(앱) 등이 주로 이 방식을 쓴다.

이 경우 그룹 내 구성원이 새롭게 검색하거나 구매한 제품을 다른 구성원에게 추천할 수 있다. 그룹에 묶인 사람들은 관심사가 비슷하다고 전제하는 만큼 한 구성원이 제품을 구입하면 다른 구성원도 그 제품에 구매 의사를 보일 가능성이 높다. 예를 들어 쇼핑몰 앱에서 니트를 자주 검색하거나 이미 구입한 사람들을 한 그룹으로 묶고 이 그룹의 일부가 털장갑을 샀다면, 나머지 구성원에게도 털장갑을 광고창에 띄우는 식이다.

교보문고 등 온라인 서점이나 이마트 등 대형 할인점은 이런 방식으로 개인 맞춤형 상품을 추천한다. 이마트의 경우 2017년 10월부터 구매 이력을 모은 빅데이터를 활용해 고객들에게 상품을 추천하고 있다.

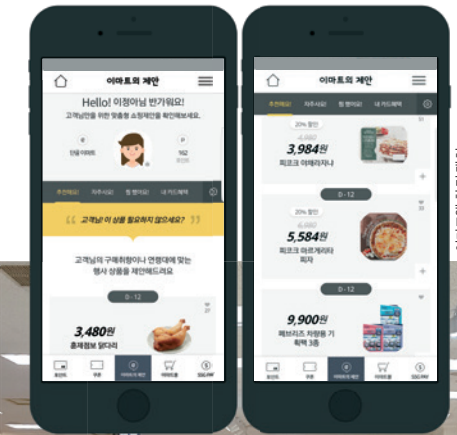
이마트 관계자는 “고객이 지금까지 이마트에서 구매한 상품이 어떤 카테고리에 속하는지 분류하는 방식으로 구매 이력을 분석한다”며 “연령대가 동일하거나 자주 구매하는 제품의 카테고리가 동일한 사람끼리 그룹으로 묶어 이들의 선호도에 맞춰 행사 상품을 추천한다”고 밝혔다.

함 교수는 “고객이 속한 계층의 과거 데이터를 활용함으로써, 최다 검색보다는 개인에 초점을 맞춰 관심 가질 만한 상품을 추천하는 기법”이라고 말했다.



인터넷 교보문고 화면 캡처

과거에는 가장 많이 구매한 품목을 순위별로 추천했다(사진). 현재 교보문고에서는 이런 방식 외에도 구매 이력에 따라 고객의 관심사에 맞게 도서를 추천하고 있다.



이마트에서는 고객의 구매 이력이나 연령대에 따라 행사 상품을 추천하는 ‘이마트의 제안’을 운영하고 있다.

이마트 영남권 화면 캡처



이정아

아마존과 넷플릭스는 협업 필터링 기술에서 시작

빅데이터를 단순히 그룹별로 분류하는 것만으로는 추천에 한계가 있다. 연령, 나이, 거주 지역도 비슷하고, 자주 검색하는 제품의 카테고리 또한 같아서 동일한 그룹에 묶이는 두 사람이 디자인이나 상품 가치를 따지는 기준에서는 다를 수 있다. 개인의 취향을 저격해 구매로 이어지게 유도하기 위해서는 결국 개인에게 초점을 맞출 수밖에 없다.

아마존은 2000년대 초 세계 최초로 협업 필터링 기술을 이용해 개인 맞춤형 상품 추천 엔진을 개발했다. 고객들이 상품 정보를 읽기 위해 상품명을 클릭한 횟수, '좋아요'를 누른 횟수, 상품 구매 후 남긴 평가 등 다양한 종류의 빅데이터를 수집한 뒤 머신러닝 시스템(A9)이 이들을 사용자나 아이템 기반으로 분류한다.

최근에는 인공지능망 기술을 적용해, 제품 사이의 상관관계와 유사도를 측정해 한 제품을 구매한 사람에게 유사한 제품을 추천한다. 이 방식은 새로 가입해 구매 이력이 전혀 없는 고객이라도 상품을 쉽게 추천할 수 있다는 장점이 있다. 미국 미네소타대 컴퓨터과학과 조셉 콘스탄과 존 리들 교수는 2012년 9월, 과학기술 분야 잡지인 '국제전기전자기술자협회 스펙트럼(IEEE Spectrum)'에 쓴 칼럼에서 "아마존은 머신러닝 추천

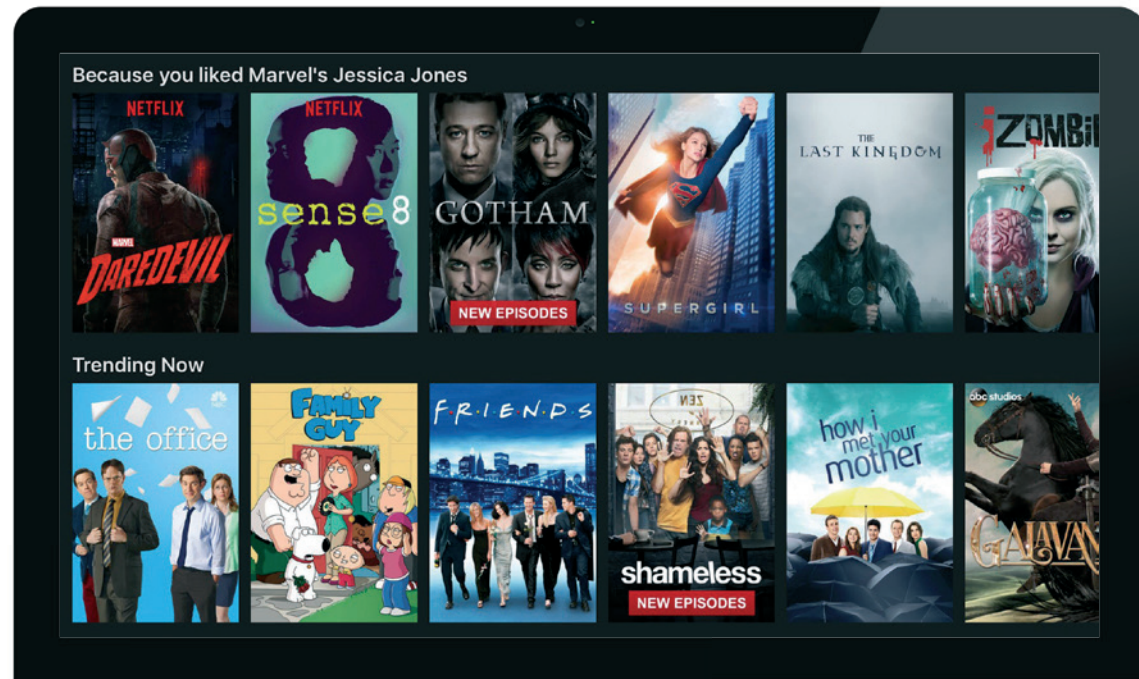
엔진으로 매출이 10~30% 정도 늘었다"고 밝혔다.

넷플릭스도 협업 필터링 기술을 이용한 개인 맞춤형 서비스로 엄청난 성공을 거둔 것으로 평가받고 있다. 넷플릭스는 영화를 지역이나 장르, 제작자, 배경, 줄거리 등 8만 여개로 세분화했다. 사용자의 개별 취향과 매우 유사하게 추천해줄 수 있다. 넷플릭스 전체 사용자의 70% 이상이 추천받은 작품을 본다는 조사 결과도 있다.

국내에서도 협업 필터링 기술을 이용한 추천 서비스가 많아졌다. 롯데카드사는 협업 필터링 기술을 이용해 고객의 카드 결제 장소를 수집하고 개인별 소비 패턴 유사도를 자동으로 계산하는 머신러닝 알고리즘을 개발했다. 고객이 최근에 어떤 매장에 갔는지, 연간 카드 결제금액은 얼마나 되는지 등을 분석한 뒤 그 고객이 향후 어떤 매장에 방문할지 예상해 쿠폰과 가까운 매장 정보를 보낸다.

김상욱 한양대 컴퓨터소프트웨어학부 교수는 "현재 주변에서 흔히 볼 수 있는 추천 서비스는 대부분 빅데이터를 활용해 파악한 사용자의 취향을 추천한다"며 "도서나 영화 등 어떤 상품을 판매하느냐에 따라 업체마다 최적의 알고리즘을 개발해 사용하고 있다"고 말했다.

넷플릭스는 영화를 지역과 장르, 배경, 줄거리 등 8만 여개로 세분화해 사용자의 개별 취향에 따라 영화를 추천하고 있다. 협업 필터링 기술을 적용한 덕분이다.



Netflix 화면 캡처

네이버는 딥러닝 알고리즘을 이용해 뉴스 등 콘텐츠나 상품을 개인 맞춤형으로 추천하는 서비스를 제공 중이다.

- ① '에어스'는 사용자가 평소 관심을 갖고 있는 분야의 뉴스와, 관심사가 같은 사람들의 패턴에 따라 다른 분야의 뉴스를 추천한다.
- ② '에이아이템즈'는 구매 이력뿐만 아니라 사용자가 평소에 읽었던 뉴스나 블로그, 카페 게시물, 검색 이력 등 콘텐츠를 종합적으로 고려해 취향을 저격한 상품을 추천한다.

네이버는 인공지능망 활용

인터넷 포털 네이버는 2016년 기사와 동영상, 웹툰 등 콘텐츠를 추천하는 딥러닝 기반 기술인 '에어스(AiRS)'를 국내 최초로 개발했다. 에어스 기술의 핵심은 두 가지다. 관심사가 비슷한 사용자들을 네트워킹하는 머신러닝 기법(협업 필터링 기술)과 개인의 텍스트 콘텐츠 소비 유형을 분석해 예측하는 인공지능망 기술(RNN-사용자의 패턴을 인식하는 인공지능)이다.

최재호 네이버 에어스팀 리더는 "내가 읽은 기사와 관련이 있는 다른 분야의 기사를 직접 찾으려면 수분 걸리지만, 에어스 기술은 수초 만에 추천해준다"며 "2016년 2월부터 뉴스에 에어스 기술을 적용한 뒤 1인당 기사 소비량이 30~40% 가량 증가했고, 사람들마다 읽는 기사의 분야도 다양해졌다"고 설명했다.

네이버는 제휴 언론사만 380개가 넘고, 하루에 제공하는 기사는 3만 건이 넘는다. 기사마다 일일이 사실 여부를 판별하기가 현실적으로 불가능하다. 이전에는 '클릭 수'나 '추천 수' 기준으로 모든 사람들에게 기사를 추천했다. 그래서 일부 언론사들이 자극적인 제목으로 클릭을 유도하는 등 인위적으로 추천수를 높일 수 있었다.

에어스팀은 스스로 '좋은 기사를 찾아내는 딥러닝 알고리즘(CNN)을 개발했다. 최리더는 "알파고가 프로 바둑기사의 기보를 외웠듯이, CNN은 좋은 기사를 수만 개 학습했다"며 "기사의 질까지 판단하는 능력을 확보한 게 핵심"이라고 말했다.

CNN은 클릭 수에 연연하지 않는다. 오히려 사람들이 오랫동안 머물면서 찬찬히 정독한 기사에 높은 점수를 주도록 돼



네이버 화면 캡처

있다. 사람들이 한 기사에 머무른 시간 데이터를 모아 학습한 뒤, 새로운 기사에 대해 사람들이 얼마나 머무르지 예측한다. 만약 CNN이 예측한 시간보다 사람들이 오래 머물렀다면 좋은 기사로 판단한다.

하지만 아직까지 풀어야 할 숙제가 남아 있다. 속보로 뜬 기사는 노출 시간이 짧아 사실상 데이터가 쌓이지 않고, 1보와 2보 등 종합기사가 뜨기 전 단편적으로 나오는 뉴스는 내용이 조금씩 덧붙여지거나 수정돼 동일한 기사로 인지되기 힘들다. 에어스는 헤드라인과 속도처럼 누구나 봐야 하는 기사는 관심분야가 아니더라도 바로 노출되도록 만들었다.

네이버 쇼핑에서도 추천의 기술이 진화하고 있다. 네이버는 2016년 11월 딥러닝 기반 상품 추천 시스템인 '에이아이템즈(AiTEMS)'를 개발해 패션 분야에 시범 적용했다. 그 결과 서비스를 시작한 지 2주 만에 상품 클릭수가 평균 6.4배 늘어났다. 사용자들이 네이버에서 검색한 단어, 블로그와 카페, 뉴스에서 읽은 정보 등 다양한 데이터를 AI가 스스로 학습해 개인별로 취향을 분류한 덕분이다.

네이버에서 사용자 상품 추천 서비스를 담당하고 있는 이정태 리더는 "에이아이템즈를 통해 지금까지 인지도가 낮았던 쇼핑물이나 상품도 추천되는 확률이 높아졌다"며 "조만간 내가 좋아하는 연예인의 패션을 간파해 비슷한 스타일의 상품을 알아서 추천하는 딥러닝 알고리즘을 개발할 예정"이라고 밝혔다. 65