

산림 분야, AI 활용 확대



월간 산림 2020년 2월호

구자춘 한국농촌경제연구원 연구위원



2020년 12월 9일과 16일, Mnet에서 ‘다시 한번’이라는 프로그램이 방영되어 큰 화제가 됐다. 작고한 터틀맨과 김현식의 영상과 음성이 AI 기술로 등장했기 때문이다. 김현식은 박진영의 ‘너의 뒤에서’를, 터틀맨은 가호의 ‘시작’을 불렀다. 두 가수가 한 번도 불러보지 못한, 아니 부를 수 없었던 노래를 부른 것이다. 김현식과 터틀맨은 1990년과 2008년에 세상을 떠났고, ‘너의 뒤에서’와 ‘시작’은 1994년과 2020년에 세상에 나왔기 때문이다. 영상 속 청중의 반응은 비슷했다. “와, 저 일이 진짜 가능하네” 였다. 그리웠던 가수가 진짜 노래를 부르는 것 같아 보는 내내 가슴이 뭉클했다. 2016년 3월 큰 이슈가 되었던 이세돌과 알파고의 바둑 대결도 어느덧 5년이 되어 간다. 그 사이 AI는 분야를 가리지 않고 우리 곁에 더 가까이 다가온 것이 분명하다. AI를 산림과 임업에 더 현명하게 사용할 방법은 무엇이 있을까?

AI를 가능하게 하는 것들

AI(Artificial Intelligence)는 스스로 판단하여 문제를 해결하고 학습하는 능력을 인공으로 만든 것이다. 김현식의 노래를 AI의 결과라 하는 이유는 기계가 학습을 통해 그가 단 한 번도 부르지 않은 노래를 부르게 하는 데 성공했기 때문이다. 많은 청중이 노래 중에 눈물을 흘리고, 노래 끝에 절로 손뼉을 친 것은 그만큼 똑같았다는 이야기니까.

이것이 가능했던 이유를 좀 더 따져보자. AI 기술은 데이터, 알고리즘, 검증과 상상력을 통해 실현된다. 과거 방송에 남아 있는 그의 목소리와 숨소리가 데이터이다. 이 재료를 쓰기 좋게 가공하여 박스에 담아 기계를 ‘학습’ 시킨다. 머신 러닝(machine learning)이라 부르는 단계다. 학습 방법은 사람이 가르치는지 여부에 따라 나뉜다. 전자는 하나씩 묻고, 맞았는지 틀렸는지 알려주는 방법으로서 ‘지도 학습’이라 한다. 횟수가 반복될수록 식별과 예측력이 높아진다. 사람이 개입하지 않는 방법에는 두 가지가 있다. 흩뿌려진 데이터의 분포, 구조, 패턴을 알아내 비슷한 것끼리 모으는 ‘비지도 학습’과 우리가 어떤 기술을 익히듯

성공과 실패를 반복하면서 가장 적합한 경로를 찾아내는 ‘강화 학습’ 이 있다.

이런 학습 방식을 ‘알고리즘’ 이라 하는데, 이 단어의 사전적 의미는 ‘어떤 문제를 해결하기 위해 명령들로 구성된 순서화된 절차’ 이다. 인간 뇌의 뉴런 구조를 따라 만든 알고리즘을 ‘인공신경망’ 이라 하고, 이를 사람 뇌처럼 여러 겹으로 쌓은 알고리즘을 ‘딥 러닝(deep learning)’ 이라 한다. 즉, 머신 러닝이 인공신경망을 통해 딥 러닝으로 발전한 것이며, 이를 통해 머신 러닝이 맞닥뜨린 여러 장벽과 한계를 뛰어넘을 수 있었다.

이제 머리 아픈 용어 설명은 그만하겠다. AI에서 재료가 많을수록, 학습의 횟수가 많을수록, 미리 선택한 알고리즘이 적절할수록, 정확한 답을 빠르게 찾게 된다는 것만 기억하자. 이런 과정을 거치고 거쳐 사람의 뇌가 가수의 목소리를 기억해 흉내내듯, 기계가 그의 목소리로 “어, 제, 는, 비, 가, 내, 렷, 어” 라는 음절은 물론 음절 간 빈 공간을 채워 곡으로 만들어 낸 것이다.

산림 분야 AI, 어디까지 왔나?

AI를 활용한 흥미로운 사례는 차고 넘친다. 인터넷 검색엔진에 ‘인공지능’ 을 입력하면 수많은 페이지 앞에 서게 된다. 인공지능 비서, 자율주행 자동차, 스마트 공장과 농장, 의사 로봇, 판사 로봇 등이 대표적 사례다. 이런 것까지 있었나 싶었을 정도로 특이한 것도 많다. 반려동물의 습관을 교정해 주는 로봇도 있고, 그림을 그리고 노래를 작곡하는 로봇이 있는가 하면, 개별 돼지의 ‘꿀꿀’ 울음소리로 배고픔의 정도를 인식해 먹이를 자동으로 공급하는 시스템도 있다. AI를 통한 변화, 즉 효율성 증진 정도는 이전의 기술로 절대 달성하지 못했을 것이다. 이러한 과급력의 크기 때문에 많은 과학자가 AI를 국가나 지구 차원의 사회와 경제에 근본적인 변화를 이끌 수 있는 기술을 뜻하는 범용 기술(general purpose technology)의 유력 후보로 주저치 않고 꼽는다. 참고로 이전의 범용 기술은 인쇄술, 증기기관, 전기, 컴퓨터, 인터넷이었다. AI, 어느 분야를 막론하고 지금 지혜롭게 써야 하는 도구임은 분명하다.

이제 산림 AI로 눈을 돌려보자. 산림청은 K-뉴딜에 대응하는 K-포레스트를 지난해 7월에 발표했다. 여기에 ‘AI’ 란 단어가 곳곳에 들어가 있다. ‘AI 기술을 접목한 산불단계별 의사결정 지원 개발’ 과 ‘드론 3차원 수치지표 표면모델 기술을 활용한 산림병해충 예찰’ 등이 대표적이다. AI 기술이 실용화된다면, 산불 진화와 예찰률을 획기적으로 높일 것으로

기대된다. 2021년에 예정된 산림 분야 연구개발 사업에서도 AI라는 단어를 쉽게 확인할 수 있다. 초실감 산림 환경을 만드는 연구, 돌발 홍수 발생 시, 산림 휴양 시설 이용자의 최적 피난 경로를 찾는 연구, 효과적 산불 진화를 위한 초기 예측 및 대응 시스템 개발 연구에 AI가 활용될 예정이다.

한편 작년 10월, 국립수목원은 AI를 활용한 단풍 예측지도를 발표했다. 지난 10년간 축적된 데이터를 학습시켜 전국 주요 명산을 중심으로 단풍의 ‘절정’ 시기를 예측한 것이다. 2009년부터 2020년까지 당단풍나무의 꽃 피고 잎 나는 현장 관측 자료를 썼고, 학습 방법으로는 ‘랜덤 포레스트(random forest)’를 활용했다. 랜덤 포레스트는 모양만 보면 숲(forest)보다 나무(tree)에 더 가깝다. 이것은 입력 변수로 목פות값을 예측하는 ‘의사결정나무’를 여러 개 겹쳐 해를 찾는 방식이다. 직접 계산해 본 오차 일수 평균은 더하고 뺐을 때 6.3일 수준이다. 데이터양이 많아지고 학습량이 늘어날수록, 알고리즘이 개선될수록 오차 일수는 더 줄어들 것이다. 그 오차가 하루 정도까지 준다면, 헛걸음하는 사람 수도 줄 것이다. 반대로 그날 그곳은 북새통을 이루겠지만 말이다.

다른 나라에서도 산림 분야 AI 활용이 본격화되고 있는데 우리나라처럼 병해충, 산불 등 재난에 효과적으로 대응하기 위해 주로 쓰이고 있다. 우리나라와 비슷한 사례이므로 이에 대한 상세 설명은 생략하고, 산업에 응용되어 부가가치를 높이는 사례 하나만 소개코자 한다. 이탈리아의 마이크로텍(Microtec) 사의 사례다. X선 CT로 원목 내부를 촬영해 결함을 알아내고, 그에 맞는 가공 방식을 선택하는 방법이다. 조금의 손실도 없이 부가가치를 최대로 하기 위해 고안된, 간단하지만 직관적이고 확실한 방법이 아닌가?

필요한 건, 상상력과 데이터

미국의 컴퓨터 공학자 래리 테슬러는 “(인공)지능은 아직 실현되지 않은 어떤 것을 의미한다”고 말했다. AI의 3가지 요소인 데이터, 알고리즘, 결핍과 상상력을 조합하는 경우의 수는 무궁무진하다. 그 어떤 것 중 하나가 산림과 임업의 발전을 획기적으로 도모할 수 있을 것이다.

그 어떤 것을 찾기 위해 가장 필요한 것은 3가지 요소 중 갈증에서 비롯된 상상력이라 생각한다. 누군가 김현식의 노래를 간절히 듣고 싶었기에, 들을 수도 있지 않을까 하는 호기심 덕분에 작년 말 텔레비전에서 그의 목소리를 다시 들을 수 있었기 때문이다. AI 분야의

선두를 달리는 미국과 중국이 벤처를 적극적으로 지원하는 것은 창업자의 상상력을 자극하기 위해서다. 대만 출신 전산학자 리카이푸가 그의 저서 『AI 슈퍼파워』에서 AI 초강대국의 조건 중 하나로 ‘포기를 모르는 전문가’를 꼽은 것도 이 때문일 것이다. 그는 풍부한 데이터, 잘 훈련된 AI 과학자, 정부 지원을 나머지 3개 조건으로 꼽았다.

AI 상상력 분출을 위한 공간 마련이 필요하다. 미국 코넬대학교가 운영하는 arXiv.org라는 논문 저장소를 참고할 만하다. 이 논문 저장소는 1991년 8월에 문을 연 인터넷 사이트로 수학, 물리학, 천문학, 전기공학, 컴퓨터과학, 생물학, 통계학, 수학 금융, 경제학 분야의 ‘심사받지 않는 논문’을 올리는 공간이다. 정제되지 않은 아이디어가 넘치는 집단지성의 장이다. 한 달에 업로드되는 논문의 수가 1만 건 이상, AI 관련 논문의 수는 2021년 1월 5일 기준으로 5592개에 달한다. 가장 최근에 올려진 코로나19 논문의 제목이 「백신 공급 관리를 위한 블록체인 플랫폼」이다. 어떨까? 지금 꼭 필요한 질문이고, 꽤 참신하지 않은가? 이 같은 열린 공간에서 우리가 간절히 찾는 그 어떤 것이 나올 수 있을 것이다. 지금껏 새로운 아이디어를 발굴하기 위해 늘 써왔던 공모전에서 좀 벗어나 보는 건 어떨까?

사실 AI 구현을 위한 알고리즘 장벽은 꽤 낮아졌다. 필자 같은 데이터 사이언스 분야의 범인도 중간 정도 성능의 컴퓨터로 파이선이나 R로 딥 러닝과 머신 러닝을 따라 해볼 수 있는 책이 속속 나오고 있으니 말이다. 이제 남은 건 풍부한 양질의 데이터다. 충분한 데이터 없이는 기계를 학습시킬 수 없기 때문이다. 이 시점에서 산림 부문의 데이터 실태를 냉정하고 꼼꼼하게 따져볼 필요가 있다. 기계를 학습시킬 만한 재료가 될 수 있는지 말이다. 참고로 2021년 1월 현재, 공공데이터포털에 등록된 산림 분야 공공 데이터는 파일 데이터 67건, 오픈 API 29개, 표준 데이터 세트 122개다.

그 어떤 것을 기다리며

지금까지 산림 분야에서 AI가 제대로 활용되길 바라는 마음으로 AI의 의미, 산림 부문의 AI 현황, 산림 분야 AI 활성화를 위한 몇 가지 조건을 차례로 살펴봤다. 아무쪼록 그 누군가의 포기를 모르는 열정으로 산림과 임업의 고질적 문제를 해결하고, 두세 발자국 더 도약할 수 있도록 하는 그 어떤 것이 ‘K_FOREST_AI(가칭)’라는 이름으로 등장하길 기대해 본다. 누군가의 상상력을 자극하고 분출시키는 공간 마련과 빅데이터 수집을 위한 정부의 과감한 지원을 청하면서 글을 마친다.