

《 _____ 확률과 통계 실생활 탐구 사례 보고서 》

하이에듀

주제 날씨예측과 조건부확률의 관계 ✓전공적합성 ✓학업역량

관련 도서



가이드	<p>서론 주제 선정 이유 [자료1,2]</p> <p>날씨 예측을 하는 것은 단순히 우산을 다음날 챙겨야 할까 말까의 고민을 위해서가 아니라고 생각합니다. 날씨 예측의 실패가 우산을 미처 챙기지 못해 편의점에서 사지 않아도 될 우산을 하나 더 사는 정도의 불상사에서 그치지 않기 때문입니다. 작게는 휴가계획 크게는 우박, 태풍과 같은 자연 재해를 미리 알아 인명피해를 최소화할 수 있으며, 농촌에서 큰 비가 오지는 않을지 가뭄이 들지는 않을지 날씨와 밀접한 연관을 갖고 있는 직업군에게 중요한 정보가 되기 때문입니다.</p> <p>하지만 우리나라의 기상청 예보가 자꾸 틀리고 뉴스와 기사에서는 기상청을 탓하는 것을 보며 근본적으로 날씨 예측은 어떻게 하는 것인지 궁금해졌고, 그것에 슈퍼컴퓨터가 이용된다는 것을 알고 나서는 슈퍼컴퓨터를 통해 연산을 함에도 불구하고 왜 예측이 틀리는 것인지 궁금하게 되었고 이를 보완하기 위해 AI 기술을 접목시키면 더 나은 예측을 할 수 있지 않을까 생각되어 탐구 주제로 선정하게 되었습니다.</p>
	<p>본론 AI에서 활용되는 수학적 업적 확률과 통계 이론 [자료3,4]</p> <p>예보를 하려면 위성과 레이더 관측 자료가 필요하고, 슈퍼컴퓨터의 수치 모델링이 필요합니다. 한국은 처음 기상예보에서 슈퍼컴퓨터가 도입된 2000년부터 일본의 수치예보 모델을 사용하다 기존 모델이 틀릴 때가 많아 2010년부터 영국 기상청과 협약을 맺고 영국 수치예보를 활용했습니다. 하지만 영국 모델역시 영국과 한국의 지형적 특성의 차이로 맞지 않는 경우가 생겨 2020년 운영을 목표로 한국역시 '한국형수치예보모델' 개발사업을 18년부터 추진하였습니다.</p> <p>확률과 통계 수업 시간 때, 조건부확률은 “어떤 사건 A가 일어 났을 때 다른 사건 B가 발생한 확률”이라고 배웠습니다. $P(B A)$라고 표기하는 이것은 사건 A와</p>

사건 B가 동시에 발생하는 $P(A\&B)$ 와는 다릅니다.

예를 들어, 두 개의 주머니 A와 B에 각각 A에는 빨간색 공 7개 파란색 공 3개, B에는 빨간색 공 5개 파란색 공 5개가 들어있다고 가정합니다. 어떤 주머니에서 뽑은 것인지 모른 상태로 공을 하나 선택했을 때 그 공이 빨간 공이었다면 주머니 A에서 나온 빨간색 공일 확률은 $P(A|\text{빨간색 공}) = P(A) P(A\&\text{빨간색 공}) / P(\text{빨간색 공})$ 이 될 것입니다.

이제 빨간공을 데이터라고 생각하면 주위에서 관찰할 수 있는 사건들로부터 특정 사건이 발생했을 때 그것이 일어나게 된 동기 또는 과정을 A 또는 B라고 할 수 있습니다. 규칙기반 시스템에서 “A이면 B이다. B이면 C이다” 와 같은 삼단 논법을 사용하면 “A이면 C이다”라는 결론을 만들어 낼 수 있습니다. 하지만 현실 상황에서는 좀 더 다양한 인과관계와 더 복잡한 상황이 있기 때문에 명확한 규칙을 만들어 내기 어려워 통계적인 방법을 쓰는데 이 때 조건부 확률이 사용 됩니다.

강수의 재료인 수증기가 상승 운동이 없더라도 중하층 대기에 수증기가 풍부하다면 호우가 발생할 수 있는 잠재 확률이 있다는 것, 여름철은 기온이 높으므로 수증기를 함유할 수 있는 그릇이 크다는 것, 이때 대기가 불안정해지면 적란운의 형태까지 발달할 수 있고 천둥과 번개가 동반될 수 있다는 것. 등등 조건부 확률을 포함하여 이러한 방정식을 종합하여 수치모델이 생성됩니다.

기상청의 예보 분석팀 예보국의 자료에 의하면 수치모델을 통해 여름철 기압계의 특성은 아래와 같이 설명한다고 합니다.

1. 여름철에는 하층 대기로 수증기 공급이 원활해지면서 강한 대류 운동을 일으킬 수 있으므로 하층 수증기량에 대한 분석이 중요하다.
2. 수치모델은 북태평양 고기압의 일변화에 대한 진동폭이 크다. 21시~03시에는 강수대를 실황보다 남쪽으로, 03~09시에는 북쪽으로 모의하는 경향이 있다.
3. 수치모델은 500hPa 고도장의 급격한 변화를 따라가지 못한다.
4. 500hPa 북태평양 고기압보다 200hPa 티베트 고기압에 대한 수치모델의 예측성이 높다. 따라서 폭염에 대한 중기예보 시에는 200hPa 일기도를 활용한 다.
5. 하층제트 북단에서 집중호우가 발생하며, 수치모델은 야간에 강화되는 하층 제트를 과소모의 하는 경향이 있다.

자료 : https://www.kma.go.kr/down/e-learning/hands/hands_43.pdf

이러한 데이터를 바탕으로 예보관이 기상예측을 한다고 합니다. 이제 AI 기술을 접목하면 이제 예보관이 아니라 AI가 통계를 기반으로 확률 모델과 확률분포를 통해 더 많은 데이터를, 더 빠른 속도로 예측할 수 있습니다.

결론 **의미와 향후 연구 방향 & 학생의 생각과 주장으로 마무리**

각종 데이터와 슈퍼컴퓨터의 능력, 예측 모델 모두 중요하지만 결국 이 모든 것을 해석하는 것은 결국 사람이라는 것을 알게 되었습니다. 계명대 환경학부 김해동 교수는 매일경제와의 인터뷰에서 “기상 예측은 방정식, 슈퍼컴퓨터 외에도 ‘숙련된 예보관의 경험’이 플러스 알파로 필요”하다고 설명했습니다.

슈퍼컴퓨터의 한계를 AI와 양자컴퓨터를 활용하여 극복하기 위해 최근 중국과 유럽에서 연구가 활발히 진행되고 있다는 기사가 근래 발표되고 있지만 예측에 활용하는 방안 역시도 다른 나라의 모델을 사용하면 일본과 영국 모델을 사용하여 실패했을 때와 같은 일이 발생할 것이라 생각합니다.

데이터를 바탕으로 해석하는 것이 사람이었듯, 인공지능 역시 데이터를 바탕으로 날씨를 예측하게 될 텐데 결국에는 우리나라에서 우리나라에 맞는 모델링 설계가 이뤄져야 올바른 데이터를 바탕으로 인공지능 역시 옳은 예측을 할 수 있을 것 같습니다.

자료1 AI와 확률의 관계성

○ AI는 확률 속에서 움직인다

AI 비서라 불리는 시리나 빅스비, 언어 치료를 돕는 AI 로봇 뮤지오 등 대화형 AI는 음성으로 데이터를 인식합니다. 이들을 통틀어 음성 인식 AI라고 부릅니다. 이처럼 어떤 데이터를 인식하느냐에 따라 AI를 구분하기도 합니다. 이미지 데이터를 처리하면 이미지 인식 AI, 움직이는 영상을 인식해서 처리하면 영상 분석 AI로 분류됩니다. 글자를 인식할 수도 있어서 사진과 글자가 섞인 문서를 넣으면 글자만 골라내는 AI도 등장했습니다.

[SEP] AI가 어떤 서비스를 제공하는 그 서비스를 가능하게 하려면 데이터를 학습해야 합니다. 그런데 아무리 많은 데이터를 입력한다고 해도 현실에서 나오는 데이터보다는 양이 적을 수밖에 없습니다. 그래서 AI가 관측한 데이터는 불확실성(오차)을 갖습니다.

[SEP] 수학에서는 완전하지 않은 데이터로 세상을 설명하기 위해 모형(모델)을 만듭니다. AI에 쓰이는 학습 알고리즘도 수학자들이 만든 모형입니다. 완전하지 않은 데이터와 수학자가 도입한 여러 가정 등으로 이 모형 역시 불확실성을 가질 수밖에 없습니다. 결국 데이터와 모형이 가진 불확실성 때문에 AI가 내놓는 답 역시 100% 확실하다고 장담할 수 없습니다.

[SEP] 수학에서 불확실성을 최소화해 정답에 가장 가까운 답을 내놓도록 돕는 개념이 확률입니다. 다양한 변수를 고려해 최적의 결과를 줄 수 있는 기대치를 계산하는 겁니다. 확률에서는 어떤 사건이 일어날 수 있는 모든 경우의 수를 가진 '전체 공간(모집단)'과 일부 경우의 수만 포함하는 '표본 공간'이 있습니다. 이런 의미에서 AI가 학습한 데이터의 집합은 현실에서는 전체 공간 중 일부분만 포함된 표본 공간이라고 볼 수 있습니다. AI가 데이터를 인식하는 단계에서부터 확률의 개념이 쓰이는 것입니다.

[SEP] 이런 불확실성을 고려하기 위해 확률 모형을 세워야 합니다. 이 과정에서 관심있는 사건을 풀기 위해 AI가 관측한 데이터를 확률변수 X 로 표시합니다. 데이터에 따라 변하는 값입니다. 또 각각의 확률 모형을 설정하기 위한 변수를 Y 라고 하면 AI가 알고리즘에 적용된 확률 모형을 학습한다는 것은 데이터 X 를 가장 잘 설명하는 Y 를 찾는 과정이라고 할 수 있습니다.

[SEP] 확률론을 연구하는 학자를 '빈도주의자'와 '베이즈주의자'로 구분합니다. 빈도주의자들에게 확률은 특정 사건을 시행할 때 A 가 나타날 경우의 수를 이 사건이 일어나는 모든 경우의 수로 나눠 구하는 것을 의미합니다. 예를 들어 주사위를 던지면 1부터 6까지 여섯 가지 경우의 수가 나오는데 이때 1이 나올 경우의 수는 한 가지뿐입니다. 결국 주사위를 1번 던질 때 1이 나올 확률은 6분의 1이라고 할 수 있어, 빈도적 확률 개념을 적용한 대화형 AI라면 어떤 상황을 입력해도 주사위를 던져 1이 나올 확률은 6분의 1이라고 대답할 것입니다.

[SEP] 반면에 베이즈주의자들은 이와 다르게 봅니다. 무한히 주사위를 던졌을 때 1이 나올 확률

이 6분의 1인 것은 분명합니다. 그런데 어떤 사람이 주사위를 10번 던졌는데 경험적으로 1이 7번 나왔다고 주장할 수 있습니다. 이 사람에게 1이 나올 확률은 10분의 7인 것입니다. 만약 주사위를 10번 던진 사람과 20번 던진 사람이 있다면, 이 두 사람은 실제로 경험한 결과에 따라 1이 나올 확률을 다르게 느낍니다. 교과서에서는 이를 사건 A가 일어났을 때 사건 B가 일어날 확률이라는 의미를 가진 조건부확률의 개념으로 설명하고 있습니다. [Step]조건부확률은 베イズ 정리나 베イズ 확률로 부르기도 합니다. 이를 적용하면 사건에 영향을 미칠 수 있는 상황을 최대한 고려할 수 있습니다. 여러 상황을 가정한 AI의 알고리즘일 수록 연산 과정이 복잡해질 수밖에 없습니다. 수학자나 컴퓨터 과학자는 빈도적 확률 개념과 베イズ 확률 개념을 이용해 AI의 기계 학습 알고리즘을 설계합니다.


AI 작동 과정에서 확률의 역할 TOP3

AI가 외부에서 들어온 데이터를 인식하는 것부터 시작해 문제를 탐색하고, 이를 통해 답안을 추론해내기까지 거의 모든 과정에서 다양한 확률 개념이 적용되고 있어!

1

데이터 인식


음성이나 이미지처럼 외부에서 들어오는 데이터가 가진 불확실성(오차)을 확인해 데이터 자체의 신뢰도를 계산해야 해.



2

문제 영역 탐색

데이터를 처리할 때는 전체 공간에서 일부만 뽑아서 확인하는 확률의 표본공간 개념으로 탐색 공간의 범위를 좁혀야해. 답을 찾는 데 걸리는 시간을 줄일 수 있거든!



3

답안 생성하기

주어진 데이터와 탐색 범위 내에서 최적의 답안을 만들려면 사전에 유력한 조건을 고려해 판단하는 '조건부확률' 개념이 쓰일 수 있어.



출처 : <https://www.dongascience.com/news.php?idx=46883>

자료2 날씨와 AI

AI 와 날씨의 역사

NCAR, NOAA(National Oceanic and Atmospheric Administration), 미국 기상청과 같은 조직에서 근무하는 날씨 및 기후 예측 과학자 들에게 AI 기술은 낯설지 않습니다 . 예측을 개선하고 , 더 정확하 고 넓은 범위의 날씨 , 기후뿐 아니라 해양과 우주의 날씨 정보를 제공하기 위해 대규모 컴퓨팅 성능과 AI 를 활용해왔습니다 .

일일 날씨 예보와 극한의 날씨 경고의 정확도는 스마트 알고리즘 과 슈퍼컴퓨터에 달려있습니다 . 많은 양의 데이터를 관리 , 분석 , 시각화해야 하는 일기 예보에서는 수년간 AI 를 활용해왔습니다 . 20 년 전 기상 데이터를 수집하여 자동화된 정확한 예측을 제공 하기 위해 DICAST(Dynamic Integrated foreCAST) 시스템이 등장했을 때 AI 도 그 일부였습니다 . 그리고 주요 민간 일기 예보 서비스 기 업에서 이 시스템을 사용했습니다 .

Hewlett Packard Enterprise 의 지구과학 부문 관리자 Ilene Carpenter 는 이렇게 말합니다 . “오랫동안 일기 예보에 AI 를 사용했지만 지 금은 엄청난 양의 데이터와 GPU 성능 으로 기계 학습이 개선되어 다시 부상하고 있습니다 . 일기 예보 센터에서 처음 슈퍼컴퓨터를 사용하기 시작했습니다 . 이제 슈퍼컴퓨터에서 물리적 모델링과 AI 및 데이터 중심

접근법을 함께 활용하여 예측 개선을 지원합니다 .”

각 기술은 단독으로 작업 수행이 불가능하기 때문에 서로 의존합 니다 . 날씨 및 기후 예 측 분야의 상황은 전 세계 다른 산업 분야에 서 일어나는 일과 비슷합니다 . 독립적인 업 계 분석가 Jeff Kagan 은 “이러한 기술의 발전으로 날씨 와 기후뿐 아니라 모든 활동에서 자연스러운 조합이 생겨나고 있 다” 고 말합니다 . “물론 AI 가 새로운 것은 아니지만 최 근 급격히 성장하고 , HPC 도 새로운 수준으로 발전했습니다 . 이러한 기술을 함께 활용 할 수 있다면 새로운 방식으로 새로운 패러다임이 등장 할 것입니다 . 기술의 발전으로 모 든 산업 분야에서 변화가 일어날 것입니다 . 우리는 새로운 세계에 첫발을 들여놓았을 뿐 입니다 .

다양한 기술을 함께 활용 전 세계 위성 , 지표면 , 바다에 위치한 날씨 센서가 날씨 및 기 후 데이터를 끊임없이 제공합니다 . 사람뿐 아니라 기존의 컴퓨터 시 스템도 데이터를 분 석하고 패턴을 검색하기엔 양이 매우 많습니 다 . 정보에서 의미를 찾지 못한다면 쓸모가 없기 때문에 문제가 됩니다 . Gagne 은 “데이터가 너무 많아 예보관들이 정보를 파악하고 결 정 을 내리는 데 어려움을 겪는다” 고 말합니다 . “이미 너무 많은 데 이터와 너무 많은 정보 잡음이 있어서 사람이 그 안에 있는 패턴 을 파악하기 위해 고군분투하는 시점은 오 래 전에 지났습니다 . 아주 효과적인 시각화 분석 툴을 개발해야 합니다 .” 과학자들이 AI 시스템 , ML, 신경망 , DL(딥 러닝) 을 활용하여 작업 중입니다 . 패턴 인식 기능이 관 련 작업에 적합하게 조정되었기 때 문입니다 . 시스템이 많은 양의 데이터를 처리하여 번 개나 토네이도 를 동반하는 태풍을 발견하는 방법을 학습합니다 . 심각한 태풍이 나 눈보 라로 이어질 가능성이 있는 패턴을 인지할 수 있습니다 . 이 러한 패턴 인식은 날씨 및 기 후 데이터 세트를 통해 작동합니다 . 하지만 많은 양을 어떻게 관리할 수 있을까요 ? 바로 이 부분에서 HPC 가 도움이 됩니다 . NERSC(National Energy Research Scientific Computing Center) 데이 터 및 분석 서비스 팀의 그룹 리더인 Prabhat 은 이렇게 말합 니다 . “기후 변화와 관련된 일반적인 논의는 전 세계 연간 지표면 온도 상승 , 해수면 상 승과 같이 과도하게 간소화된 수치를 중심으로 합니다 . 하지만 기후 변화가 본인의 삶에 어떤 영향을 미칠지 알 고 싶어 하는 사람이 늘고 있습니다 . 더 세분화된 수준으로 정밀 하게 처리하는 시뮬레이션이 필요합니다 . 일례로 대규모 데이터 세트가 생성되어 메인주 에 어떤 영향을 주는지 또는 내가 사는 도시에 어떤 영향을 주는지를 물어보려면 정밀한 분석을 수행해 야 합니다 . AI 는 더욱 정확한 시뮬레이션과 정밀한 분석으로 우 리 작업 에 도움이 됩니다 .” 몇 년간 NERSC 의 주력 슈퍼컴퓨터는 Cori 라고 알려진 레거시 Cray XC401 이었습니다 . 622,336 개의 인텔 ® 프로세서 코어가 탑 재되어 이론상 초당 30,000 조 개의 계산 수행이 가능합니다 .

출처 : <https://www.hpe.com/psnow/doc/a50007052kop>

날씨는 어떻게 예측하는 걸까?

기상 예보 과정은 △관측자료 수집 △수치 모델링 △예보관 분석 △예보 순으로 진행된다. 첫 번째 단계인 관측자료 수집은 예보에 있어서 가장 기초적인 단계다. 이 단계에선 기상 요소인 △강수량 △기압 △습도 △풍향 등을 바탕으로 대기과 해양의 상태를 △땅 △바다 △우주 △하늘에서 입체적으로 관측한다. 윤기환<기상청 대변인실> 통보관은 “기상전문인

공위성인 천리안 2A호로 우리나라 주변 기상 상황을 2분 간격으로 찍어서 현재 날씨 상태를 알 수 있다”며 “기상레이더를 통해선 강수량을, 윈드 프로파일러를 통해선 상층의 바람 상태를 알아낸다”고 말했다. 이처럼 현재 기상청에선 다양한 관측기구를 통해 기상 상태를 면밀히 관측하려고 노력 중이다.

기상청의 예보가 빗나가는 원인은?

최근 기상청의 기상 예보가 자주 빗나가면서 기상청에 대한 불신이 커졌고 그 결과 많은 국민이 노르웨이 등 다른 나라 기상청을 통해 우리나라 기상 예보를 보는 웃지 못할 일이 벌어지고 있다. 기상청의 예보가 빗나가는 원인은 무엇일까.

첫째는 기상청의 미흡한 기술력 때문이다. 기상청은 예보 정확도를 높이기 위해 지난 4월, 앞서 언급된 수치예보모델의 한 종류인 한국형수치예보모델을 도입했다. 기존에는 영국모델을 차용했지만 우리나라의 지형 및 기상 특성이 영국과 다른 점을 고려해 더 정밀한 예측값을 만들어 내고자 독자 기술로 수치예보모델을 만든 것이다. 하지만 최근 기상 예보가 빗나가는 일이 잦아지며 기상청의 예보 정확도가 나아졌는지에 대한 의문이 제기됐다. 실제로 기상청이 지난 5월 ‘올여름 기상 전망’에서 예측한 기온과 강수량이 모두 빗나갔다. 지난 10년간 기상청은 한국형수치예보모델을 개발하기 위해 약 780억 원을 들었다. 막대한 금액을 들여 우리나라 기후에 최적화된 수치예보모델을 개발했으나 제대로 효과를 보지 못하면서 기상청은 국민의 비판으로부터 자유롭지 못하게 됐다.

예보 정확성을 높이기 위한 해결책은 ?

기후변화가 빠르고 다양하게 나타나면서 기상 예측에 영향을 주는 변수가 많아지고 있다. 그렇기에 기상청은 국민의 생명과 재산 보호를 위해 날씨 예측의 기술력을 높여야만 하는 과제에 당면했다. 이에 관해 윤 통보관은 “한국형수치예보모델은 최근에 도입됐기 때문에 앞으로 자료를 연구하고 발전시킨다면 예보의 정확성이 개선될 것”이라면서도 “날씨는 자연현상이기에 현재의 기술로써 오차가 발생하는 것이 불가피하다”고 밝혔다. 이어 윤 통보관은 “현재 날씨 상태를 자세히 표현할 수 있는 기술력을 갖춰 나가는 것도 중요하다”고 덧붙였다.

기상청은 2000년부터 기상업무 전용의 슈퍼컴퓨터 1호기 도입을 시작으로 현재는 슈퍼컴퓨터 4호기까지 도입했고 내년 말까지 슈퍼컴퓨터 5호기 구축을 계획 중이다. 하지만 이것만으로는 부족하다는 것이 기상청의 입장이다. 윤 통보관은 “슈퍼컴퓨터는 방대한 양의 관측 자료를 계산하기에 슈퍼컴퓨터의 데이터가 기하급수적으로 늘어난다”며 “데이터가 계속 쌓이기 때문에 더 많은 슈퍼컴퓨터의 보급이 필요하다”고 밝혔다. 이어 윤 통보관은 “보급뿐만 아니라 방대한 양의 데이터를 순식간에 처리하기 위해선 슈퍼컴퓨터의 성능 개발도 지속적으로 병행돼야 한다”고 덧붙였다.

출처 : <https://www.hynews.ac.kr/news/articleView.html?idxno=10420>

문제는 일본 기상청과 한국 기상청 예보 내용이 크게 달랐다는 점이다. 태풍이 24일 전라 지역을 거쳐 수도권 남쪽으로 이동할 것이라고 일본 기상청이 예보한 시점은 23일 오전 6

시다. 24일 실제 이동경로와 거의 일치한다. 반면 한국 기상청은 23일 오전 9시 태풍 상륙 위치를 전북 군산으로 바꾼 후 늦은 오후가 돼서야 또다시 전남 서해안으로 이동경로를 변경했다. 양국 기상청의 태풍 진로 예측 오차는 2017년 일본이 24시간 기준 82km로 한국의 93km보다 낮았다. 태풍 반경이 수백 km인 것을 감안하면 수십 km 오차는 분석에 큰 영향을 미치지 않는다는 시각도 있다.

기상청 관계자는 "일본은 4면이 바다이며 태풍 등 자연재해 피해가 극심해 분석 노하우가 있는 인력이 많다"며 "한국은 차차 발전해가는 수준이라 일본에 미치지 못하는 것"이라고 시인했다. 김해동 계명대 환경학부 교수는 "기상 예측은 방정식, 슈퍼컴퓨터 외에도 '숙련된 예보관의 경험'이라는 플러스 알파가 필요하다"고 지적했다. 이론상이 아닌 실제 환경에서 실수도 하면서 '실전 경험'을 쌓은 예보관들이 많이 나오는 것이 중요하다는 것이다. 손병주 서울대 지구환경과학부 교수는 "우리나라는 항공기를 태풍 상층부에서 띄워 관측하거나 레이더를 통한 관측 등이 대만 일본 등에 비해 부족한 것은 사실"이라며 "아무리 좋은 수치모델을 쓰더라도 투입되는 관측자료가 부족하면 예측력이 떨어질 수밖에 없다"고 관측자료 확보를 주문했다.

출처 : <https://n.news.naver.com/mnews/article/009/0004206925?sid=102>

자료3 빅테크 AI 모델

AI 도입해 변수 분석 속도와 예측 정확도 크게 개선

인공지능(AI)을 활용해 기상예측의 정확도를 높이기 위한 도전이 이어지고 있다.

기상예측은 온도, 강수량, 기압, 바람, 습도, 일사량 등 수많은 날씨 변수를 조합한 고차 방정식을 슈퍼컴퓨터로 풀어낸 결과를 토대로 사람의 판단을 추가하는 형태로 이루어지고 있다.

그러나 슈퍼컴퓨터도 수많은 기상 변수를 한번에 하나씩 차례로 분석하기 때문에 시간이 많이 걸리고, 지구 온난화에 따른 기상 이변으로 새로운 변수가 속속 등장하는 바람에 정확한 예측이 어려운 실정이다. 일기예보가 자주 틀리거나 오차를 보이는 이유다.

이같은 난점으로 인해 신뢰할 만한 예보 기간은 10년에 하루 정도 늘어나고 있다. 덕분에 1980년에는 3일 예보만 정확했다면 지금은 7일 예보까지는 어느정도 정확한 편이다.

여기에 AI가 추가되면서 정확도가 더 높아지기 시작했다는 보도가 이어지고 있다. AI가 복잡한 변수를 계산하는 시간을 대폭 줄이면서 정확도를 높이는데 기여하고 있다는 내용이다. 특히 기계학습(ML)이 도입되면서 기존 데이터가 없어도 예측이 가능한 모델이 속속 등장해 관심을 모으고 있다.

워싱턴포스트와 MIT테크놀로지리뷰 등 최근 외신에 따르면 엔비디아를 비롯한 기술 대기업들이 기상예측 모델에 AI를 도입하면서 변수를 분석하는 속도와 예측 정확도를 크게 개선해 나가고 있다.

출처 : <https://www.aitimes.com/news/articleView.html?idxno=152286>

대부분 국가에서는 기상 예측 모델을 컴퓨터로 구동합니다. 수학적 방법론에 따라 대기과

해양 상태를 가상으로 구현해봅니다. 휴대전화 날씨 앱들은 다양한 기상 데이터를 받아 미래를 예상합니다. 러스 슈마허 콜로라도 주립대학교 기후학자는 이런 예측 도구들은 과학자들이 수십 년에 걸쳐 이룩한 성과라고 평가합니다. 그는 "이 도구들은 기후 예측에 큰 발전을 가져왔다"며 "더 많은 데이터와 모델 개선, 발전된 컴퓨터로 예측은 점점 더 정확해지고 있다"고 강조했습니다.

시에는 "인공지능이 수학적 방법론으로 날씨를 예측하는 업무에서 컴퓨터 시뮬레이션 모델보다 빠르다"고 주장했습니다. 그에 따르면 기존 모델은 토네이도나 우박처럼 갑작스러운 기상 위험을 파악하기 어렵습니다. 또 AI 시스템보다 더 많은 자원이 필요하다는 단점도 있습니다.

인공지능인 판구웨더를 훈련하기 위해 전 세계 기상 자료 39년 치를 입력했습니다. 과거 사용한 예측 기술보다 1만 배 더 빠르고 정확도도 뒤떨어지지 않았습니다. 내장된 3D 모델도 도움이 되었습니다. 이전에 개발된 AI 예보와 달리 다양한 기압 레벨에서 대기 상태를 기록해 세밀한 기후 예측이 가능했습니다. 판구웨더는 5~7일 앞 미래 날씨를 추정할 수 있었습니다.

강수량을 알지 못한다는 단점도 있습니다. 그로 인해 토네이도 같은 기상이변은 예측하지 못합니다. 하지만 네이처에 발표된 두 번째 나우캐스트넷을 적용한다면 간극을 메울 수 있을지도 모릅니다.

나우캐스트넷은 강우 현상을 상세히 계산합니다. 미국과 중국에서 레이더로 관측한 자료와 인공지능 신경망 기술로 미국 동부와 중부에서 257 제곱킬로미터 공간의 강수량을 3시간 전부터 알 수 있습니다. 중국 기상학자 62명이 참여한 일기예보 연구에서도 폭우를 71% 정확히 맞춰 1위를 차지했습니다.

에이미 맥거번 미국 국립과학재단 인공지능 연구소 소장은 "이런 생성형 AI 모델은 유망하다"고 말했습니다. 동시에 이런 기술들로 일기예보를 하려면 좀 더 개선이 필요하다고 밝혔습니다. 그는 물리학에 기반한 수학적 계산 절차가 부족하다고 지적합니다. 대기에서 이동하는 수분, 공기, 열을 고려하면 더 정교한 예보를 내릴 수 있습니다. 그는 "이 논문들은 아직 이론 증명 단계이다"며 "극단적인 날씨를 예상하기 위해 물리 법칙을 적용하지 않았다"고 강조합니다. 또한 AI 기술이기에 가지는 편향성도 걱정했습니다. 인공지능은 입력하는 데이터 질에 따라 역량이 달라집니다. 품질이 낮은 자료나 특정 지역에 쏠린 정보로 학습하면 예측 능력이 지역에 따라 한정될 수 있습니다.

AI가 예술부터 의학까지 다양한 영역으로 뻗어나가면서 기상학에도 닿았습니다. 현재 기술은 당장 실용화하기까지 난관이 많지만, 미래에는 다를 수 있습니다. 맥거번은 "5~10년이 지나면 AI 모델들의 역량에 놀라게 될 것이다"고 말했습니다.

출처 : <https://www.popsci.co.kr/news/articleView.html?idxno=20991>

자료4 수식과 그래프 활용 자료

내용 : 책에서 발췌한 수학적, 확률적 지식을 포함하고 있으며 이를 교과 내용과 관련 지어 구성하였는가

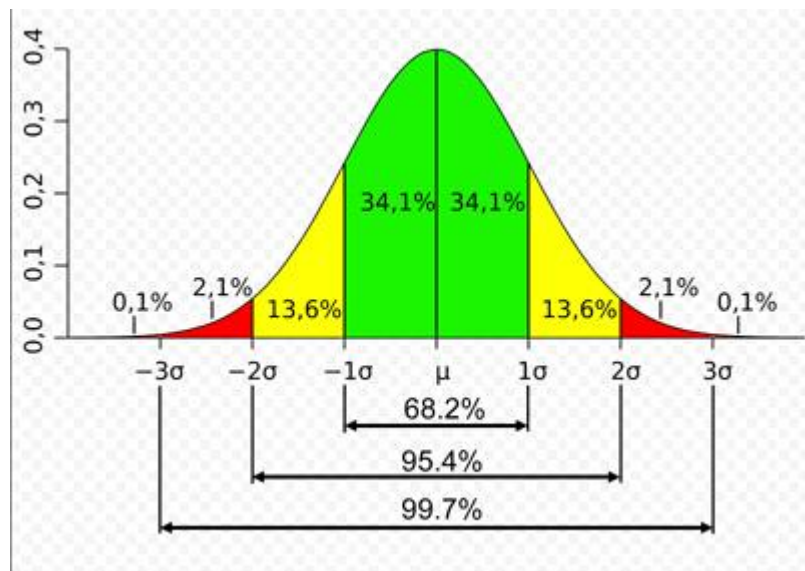
내용 : 실생활 맥락 안에서 책의 내용을 재해석 하였는가

참고 :

<https://magazine.contenta.co/2015/12/%EC%9D%B8%EA%B3%B5%EC%A7%80%EB%8A%A5-%EC%A1%B0%EA%B1%B4%EB%B6%80%ED%99%95%EB%A5%A0/>

▶ 해당 내용은 조건부 확률이 인공지능 관점에서 어떻게 해석되고 있는지를 설명하고 있습니다. 학생의 진로희망학과인 '컴퓨터공학'/'소프트웨어학과'와의 연결을 통해 실생활(진로에서의 적용)으로 확장하여 책의 내용을 재해석한 모습을 보여주면 좋겠습니다.

표현 : 확률을 나타내는 자료 (다이어그램, 그래프, 표)를 포함하는가



참고 자료 :

https://artificialnetworkforstarters.readthedocs.io/en/latest/_post/chap6.html

▶ 해당 자료를 읽고, 그 중 학생이 이해한 부분의 그래프 자료를 활용하는 것을 추천해 드립니다.