

# 수학자는 어떻게 인공지능을 발전시키는가?

제가 머신러닝(machine learning)이란 용어를 처음 접한 건 2013년도 겨울로 기억합니다. 당시 저는 SNS상으로 종종 인사드리던 고등학교 선배들과 운동 모임을 정기적으로 가졌었는데, 민현석 박사는 그 중 한 분이었습니다. 영상처리와 패턴인식 이론을 전혀 모르는 수학과 대학원생이었던 저는 민현석 박사님 덕분에 이 분야의 발전 가능성을 미리 접할 수 있게 되었습니다. 제가 학계에서 벗어나 산업계로 첫걸음을 내딛게 된 데에는 민현석 박사님의 조언이 매우 큰 영향을 주었습니다.

※ 편집자 주: 물체에도 필자의 의도가 담겨 있다고 판단해서 존칭 표현을 그대로 담았습니다.

## 딥러닝과 수학의 연결 고리

그러던 어느 날 민현석 박사님이 합성곱 연산(convolution operation)의 수학적인 직관에 대해 물어 보셨습니다. 사실 합성곱 연산은 신호처리 분야 뿐만 아니라 편미분방정식, 확률론 등 현대수학의 해석학(analysis) 분야에서 절대적인 위치를 차지하고 있습니다. 제가 박사과정 때 주로 연구했던 분야도 조화해석학(harmonic analysis)과 확률론이었습니다. 두 이론은 합성곱 연산과 푸리에 분석(Fourier analysis) 또는 웨이블릿 분석(wavelet analysis) 사이의 유기적인 관계를 이용해 시그널의 확률적인 패턴을 분석합니다. 민현석 박사님은 장시간에 걸쳐 딥러닝의 등장 배경과 CNN(convolutional neural networks)의 구조를 설명해 주셨고, 2012년도부터 이미지넷(ImageNet)<sup>1</sup> 등의 대회를 통해서 본격적으로 알려지게 된 딥러닝의 성능 발전에 대해서도 알려 주셨습니다.

한 동안 저는 딥러닝 분야에 수학자가 크게 기여할 수 있는 일이 많지 않다고 생각했습니다. 데이터 기반(data-driven)이면서 모델 비의존적(model-free)인 방법론을 추구하는 학문에서는 수학적인 접근보다 오히려 데이터를 잘 축적하기 위한 파이프라인(pipeline) 설계가 더 중요하다고 생각했기 때문입니다. 저의 이러한 방어적(?) 태도에 대해 민현석 박사님은 다음과 같이 답변했습니다.

“지금으로선 나도 잘 모르겠다. 하지만 딥러닝의 가치가 커질수록 수학이 중요해질 날이 분명 올 것이다.”

## 딥러닝 엔지니어가 수학을 꼭 전공해야 하나?

딥러닝은 수학적으로 어려운 구조는 아닙니다. 행렬 연산과 비선형 함수(activation function)를 여러 층으로 합성해 쌓는 것이 기본 구조입니다. 1차 미분만 계산하는 역전파(back propagation) 알고리즘은 학부 대학생 정도면 쉽게 구현할 수 있고, 고등학생도 어렵지 않게 이해할 수 있습니다. 게다가 자동미분(automatic differentiation) 라이브러리가 포함된 프레임워크가 오픈소스화 되면서부터 구현하는 것 또한 손쉬운 일이 되었습니다. 텐서플로우코리아(TensorFlow Korea)나 파이토치(PyTorch) 한국 사용자 그룹 같은 SNS 그룹에는 본인이 직접 만든 알고리즘을 깃허브(GitHub)로 공개하는 중학생도 더러 있을 정도입니다. 이런 경향 때문인지 “수학을 전공하는 것이 딥러닝 엔지니어에게 도움이 될까요?”라는 질문이 SNS 상의 커뮤니티에 올라오곤 합니다. 갑론을박이 오고 가는 가운데 전문적인 수학 지식이 왜 필요한가에 대해 대중적인 공감을 얻은 의견은 아직까지 없습니다. 오히려 딥러닝 논문의 수식들은 현학적인 장벽처럼 느껴지고 코드로

아이디어를 이해하는 것이 더 효율적이라는 의견도 있습니다.

저 역시 딥러닝을 목적으로 수학을 깊게 공부하는 것은 시간 낭비라고 생각합니다. 2~3년의 시간을 투자하여 수학과와 전공 수업을 공부하기에는 그 범위가 너무 넓고, 수학 외에도 공부해야 할 더 중요한 지식들이 산더미이기 때문입니다. 게다가 딥러닝의 철학은 기존의 수학적 모델링의 성과를 과감히 포기하고 효율적 연산과 데이터 의존적인 엔드-투-엔드(end-to-end)를 추구하는 점에 있습니다. 그래서 저는 모든 딥러닝 연구자가 수학을 깊게 공부해야 한다는 의견에는 선뜻 동의하기 어렵습니다.

그렇다면 반대 논리로 수학 전공자가 딥러닝 연구에 기여할 가능성은 낮을까요? 그렇지는 않다고 생각합니다. 모두가 수학을 전공할 필요는 없지만 수학에 능숙한 연구자는 과학, 공학 연구에 언제나 필요하기 마련인데 가령 주어진 문제에 내재된 제약들을 다른 수학적 조건으로 전환하는 방법을 제시할 수 있습니다. 작년 초에 많이 회자되었던 마틴 아조브스키(Martin Arjovsky)의 베셔슈타인(Wasserstein) GAN(Generative Adversarial Network) 논문이 그 중 하나의 사례로 볼 수 있습니다. 마틴은 기존의 머신러닝 논문에서 주로 사용하던 쿨백-라이블러 발산(Kullback-Leibler divergence) 대신 베셔슈타인 거리함수<sup>2</sup>를 GAN에 적용하는 아이디어를 수학적인 타당성과 함께 제시했습니다. 그 이후로는 연구자들이 베셔슈타인 거리함수를 쿨백-라이블러 발산을 대신하여 새롭게 시도할만한 손실 함수(loss function)로 인식하기 시작했고, 최근엔 오토 인코더(auto-encoder)<sup>3</sup>나 강화 학습(reinforcement learning)<sup>4</sup> 연구에도 도입되기 시작했습니다.

주어진 문제에 알맞은 정규화(regularization) 방법을 제시하는 것도 수학을 전공한 연구자가 현업에서 풀어야 할 과제입니다. 대개 정규화는 과적합(overfitting)을 막고 일반화(generalization) 성능에 도움을 주기 위한 도구로 인식됩니다. 그러나 정규화 방법들을 어떤 문제에도 동일하게 적용할 수 있는 건 아닙니다. 예를 들어 데이터가 상대적으로 적거나 노이즈가 적은 상황에선 모델의 복잡도를 제한할 필요가 있으므로 패러미터(parameter)의 크기를 제한하는 가중치 감소(weight decay) 기법이 제 역할을 하는 경우가 많습니다. 반면, 데이터에 노이즈가 크게 내재된 경우에는 확률적 정규화 기법(stochastic regularization technique)<sup>5</sup>들을 사용하는 베이지안 신경망(Bayesian neural network)이 우수한 성능을 낼 때가 있습니다. 요컨대 정규화 기법은 주어진 학습 문제의 최적해 공간(optimal solution space)을 특정하는 역할로 해석할 수 있고, 최적해 공간이 적절하게 상정되었을 때 일반화 성능이 개선되는 것입니다. 이처럼 “어떤 해를 찾을 것인가?”라는 질문에 대해 데이터 집합의 특징을 고려하여 적절한 선행적 지식(prior knowledge)을

글 | 임성빈 leo.brain@kakaobrain.com

수식과 증명을 사랑하던 수학자가 로보틱스와 인공지능의 매력에 빠지게 되었습니다. 카카오브레인에서 수학과 엔지니어 사이의 가교 역할을 담당하는 수리과학자(Mathematical Scientist)로 활동하고 있습니다. 안전하고, 신뢰도 높은 인공지능을 개발해 우리 사회를 보다 윤택하게 만들고 싶습니다.

반영하는 모델을 찾는 것은 수학적 근거를 필요로 합니다.

또한 수학은 지금까지는 시도되지 않은 종류의 데이터에 접근할 수 있는 방법론을 제시할 수 있습니다. 후자의 대표적인 예시는 최근에 주목을 받고 있는 지오메트릭 딥러닝(geometric deep learning, GDL)<sup>6</sup>입니다. 지오메트릭 딥러닝은 2017년 NIPS(Neural Information Processing Systems) 튜토리얼(tutorial)에 소개 될 정도로 그 중요성이 새롭게 인식되는 분야입니다. 지오메트릭 딥러닝은 데이터를 그래프나 다양체(manifold) 같은 비유클리드(non-Euclid) 공간 상의 대상인 경우로 상정하고, 그 위에서의 딥러닝 네트워크를 학습하는 방법을 다룹니다. 지오메트릭 딥러닝 분야에 대한 상세한 내용은 다음 기회에 다루도록 하겠습니다.

### 수학자들의 도전을 기다리는 인공지능 이론

지금까지는 수학 전공 연구자들이 딥러닝 연구에 기여할 수 있는 내용들에 대해 살펴보았습니다. 그렇다면 학계의 수학자들은 딥러닝을 어떻게 바라보고 있을까요? 종종 “딥러닝은 블랙박스라 수학적으로 이해할 수 없지 않나요?”라는 질문을 듣게 되는 경우가 있지만, 역사적으로 보면 수학자들은 오래전부터 신경망 연구를 시도하였고 어느 정도 성공을 거둔 영역도 있습니다.

[그림 1] 찰스 페퍼먼(Charles Fefferman)<sup>7</sup>



조화해석학은 물리학과 신호처리 이론의 중요한 근간이 되는 학문입니다. 딥러닝 연구도 이 조화해석학을 통해 많은 성질들이 규명되었는데 찰스 페퍼먼(Charles Fefferman, [그림 1])이 그 대표적인 수학자라고 할 수 있습니다. 페퍼먼은 현대 수학의 다방면에서 활약한 천재 수학자인데 신경망 연구 뿐만 아니라 머신러닝 전반에 걸쳐 어려운 문제들을 풀어내고 있습니다. 그 중 하나가 다양체 가설(manifold hypothesis) 문제입니다. 이는 고차원의 데이터를 차원을 축소하여 저차원의 데이터로 표현하는 다양체의 존재성에 대한 가설입니다. 다양체 가설은 머신러닝

연구자들이 대부분의 이미지나 텍스트 데이터에서 참으로 간주합니다. 여러 실험 결과들이 이를 뒷받침하고 있지만<sup>8</sup>, 이를 수학적으로 증명하는 것은 어려운 일입니다. 페퍼먼은 이 다양체 가설이 수학적으로 검증 가능하다는 것을 2016년 Journal of AMS에 게재된 ‘Testing the Manifold Hypothesis’ 논문에서 증명하였습니다.

[그림 2] 스테판 말렛(Stephane Mallat)(좌)<sup>9</sup>과 조안 브루나(Joan Bruna)(우)<sup>10</sup>



스테판 말렛(Stephane Mallat, [그림 2])과 조안 브루나(Joan Bruna, [그림 2])는 2012년 ‘Invariant Scattering Convolution Networks’<sup>11</sup>라는 논문에서 웨이블릿 필터(wavelet filter)를 가진 CNN 구조가 가지는 수학적 성질들을 밝혀낸 적이 있습니다. 구체적으로 ‘CNN은 차원의 저주(curse of dimension)를 극복하고 이미지 데이터에 내재된 불변량(invariant)을 찾아낸다’라는 추측을 수학적으로 참인 명제로 증명해 낸 것입니다. 이 논문의 뒤를 이어서 토마스 와이토스키(Thomas Wiatowski)와 헬멧 볼스키(Helmut Bolcskei)는 일반적인 필터들로 이루어진 CNN도 ‘Invariant Scattering Convolution Networks’ 논문에서 말한 CNN 성질과 같은 성질을 만족함을 증명했습니다.<sup>12</sup> 이 외에도 ‘Mathematics of Deep Learning’<sup>13</sup> 논문을 보시면 딥러닝 연구에 기여한 수학적인 결과들을 개괄적으로 보실 수 있습니다.

앞에서 소개한 수학자들의 연구 결과들도 놀랍지만 인공지능 분야에서의 수학 연구들은 많은 부분에서 현재 진행형입니다. 이 분야에서 수학이 필요한 이유는 재현 가능성(reproducibility)의 보장, 이론적인 예측과 실험 결과의 일치가 필요 조건인 과학 이론을 정당화하는 것이 바로 수학의 역할이기 때문입니다. 물론 ‘모든 과학 이론은 수학적 증명 없이는 이론이 될 수 없다’는 환원주의적인 주장이 항상 옳은 것은 아닙니다. 그러나 인공지능 이론이 그 범주 밖에 있다고 생각하지 않습니다. 다비드 힐베르트(David Hilbert, 1862-1943)의 제자 중에 리하르트 쿠란트(Richard Courant, 1888-1972)라는 수학자가 있었습니다. 쿠란트는 물리학과 현대 응용수학의 발전에 수많은 기여를 했던 사람인데, 생전에 다음과 같은 말을 남긴 적이 있습니다.

“실험적 증거(empirical evidence)는 결코 수학적 존재성(existence)을 의미하지 않으며, 엄밀한 증명에 대한 수학자들의 요구를 쓸데없는 것으로 여기게 하지도 않는다. 현상에 대한 수학적 설명이 의미를 가지게 만드는 건 오로지 수학적인 증명을 통해 가능한 것이다.”<sup>14</sup>

2017년도 여름에 열린 제주 머신러닝 캠프에서 뉴욕대학교(New York University, NYU)의 조경현 교수님이 “딥러닝이 정말 블랙박스인가?”라고 질문을 던진 적이 있습니다. 현재는 딥러닝을 포함해 인공지능 분야의 연구가 실험적인 결과들에 의해 대부분 진행이 되고 있습니다만, 언젠가 상당수의 명제들이 수학적으로 증명된 명제들로 논증되는 시점이 올 것이라 생각합니다. 그리고 그렇게 이루어진 인공지능 연구가 우리의 삶을 더 윤택하게 만들 수 있다고 저는 확신합니다. 어쩌면 오랜 시간이 필요한 작업이 될 수 있을 것입니다. 그러나 힐베르트의 명언을 인용하면 우리는 알아야만 하며 언젠가는 알게 될 것입니다.

### 글을 맺으면서

예전에 모 대학의 수학과 교수님께서 저에게 “어떤 수학 전공자를 인공지능 업계에서 선호합니까?”라고 물어 보신 적이 있습니다. 그 분께서는 “어떤 전공이 중요한지에 대한 답을 기대하셨지만, 저는 “특정 전공보다 소통 능력이 더 중요합니다”라고 답한 바 있습니다.

저는 수학자들의 연구가 비약적인 성공을 거둔 후에도 공학을 포함한 비전공자들의 직관적인 이해를 돕거나 성능 향상에 도움이 되려면 수학자들의 더 많은 노력이 필요하다고 생각합니다. 왜냐하면 용어가 지나치게 전문적인데다 수학자들의 논증이 실제적인 예제로 설명하기가 어려울 정도로 비전공자의 시각에선 추상적이기 때문입니다. 사실 수학이 추상화를 추구하게 된 원인은 지난 2세기 동안 과학의 발전에 성공적으로 공헌했던 것이 큼니다. 이는 학문적 가치관으로서의 옳은 방향이었지만, 공학적 연구에 도움이 되려면 이제는 소통의 관점에서 접근해야 합니다. 저는 최근에 카카오브레인에 합류한 최성준 박사님과 로보틱스 분야에서의 머신러닝 활용을 연구하고 있습니다. 막연하게 떠오른 아이디어를 수학적 언어로 토론하고 코드로 구현하여 실험합니다. 그렇게 구체화되는 연구 결과들을 보면서 매일 큰 보람을 느끼고 있습니다. 그 과정에서 공학자와 수학자 간의 협업에 있어 위에서 말한 소통이 매우 중요하다는 점을 깨닫고 있습니다.

제가 생각하는 STEM<sup>15</sup>업계에서의 성공적인 조직 문화 요건들 중 빼놓을 수 없는 것이 ‘소통’ 과 ‘다양성’<sup>16</sup>입니다. 소통과 다양성이 왜 필요할까요? ‘다양성과 복잡성(Diversity and Complexity)’의 저자인 스콧 페이지(Scott Page)는 복잡한

과학적 문제의 해결은 단순히 한 분야의 전문성에 의존하기 보다는 기존의 접근 방식과 새로운 방식을 접목하고 비틀어보는 ‘발상의 전환’에서 나온다고 설명합니다.<sup>17</sup> 복잡한 문제일수록 변수와 구조가 다차원적이기 때문입니다. 새로운 발상이 통찰력 있는 아이디어가 되려면 문제를 바라보는 다양한 시각들이 얼마나 서로 소통하고 적용되는지에 따라 결정됩니다. 페이지 교수는 이러한 소통과 다양성의 메커니즘을 머신러닝 알고리즘 중 하나인 랜덤 포레스트(random forest)<sup>18</sup>에 비유했습니다. 복잡계 문제를 푸는 알고리즘과 팀 조직 구성이 비슷한 면모를 가진다는 점이 흥미롭습니다.

인공지능 연구는 아직 미완성이며 그 가운데 해결해야 할 많은 난제들이 쌓여 있습니다. 앞으로 다양한 수학 전공자들이 인공지능 연구 분야에 뛰어들어 기술 발전에 공헌할 날이 오기를 기대해 봅니다.

<sup>1</sup> 설명 | 2010년부터 매년 개최되는 대회로 물체 인식, 물체 검출, 동영상 물체 검출, 장면 분류, 영상 분할 등 인공지능 및 딥러닝 관련 문제를 해결하는 경연대회로 구글, 마이크로소프트, 퀄컴 등 세계적 IT 기업이 출전 하는 대회이다. <sup>2</sup> 참고 | 두 거리함수에 대한 구체적인 비교는 Wasserstein GAN 논문 및 링크 자료 (<https://www.slideshare.net/ssuser7e10e4/wasserstein-gan-i>) <sup>3</sup> 참고 | Wasserstein Auto-Encoders (<https://arxiv.org/abs/1711.01558>) <sup>4</sup> 참고 | On Wasserstein Reinforcement Learning and the Fokker-Planck equation (<https://arxiv.org/abs/1712.07185>) <sup>5</sup> 설명 | 페라미터에 사전(prior) 확률분포를 부여해서 노이즈의 성능감소를 감쇄하는 페라미터를 학습하는 기법. 드롭아웃(dropout)이 대표적인 방법이다. <sup>6</sup> 참고 | <http://geometricdeeplearning.com/> <sup>7</sup> 참고 | <https://tecnico.ulisboa.pt/en/events/symposium-interpolation-and-approximation-in-several-variables/> <sup>8</sup> 논문 | Ian Goodfellow et al. Deep Learning, p.161 <sup>9</sup> 참고 | <https://www.di.ens.fr/~mallat/> <sup>10</sup> 참고 | <https://cims.nyu.edu/~bruna/> <sup>11</sup> 논문 | Invariant Scattering Convolution Networks (<https://arxiv.org/abs/1203.1513>) <sup>12</sup> 논문 | A Mathematical Theory of Deep Convolutional Neural Networks for Feature Extraction (<https://arxiv.org/abs/1512.06293>) <sup>13</sup> 논문 | Mathematics of Deep Learning (<https://arxiv.org/abs/1712.04741>) <sup>14</sup> 원문 | Empirical evidence can never establish mathematical existence nor can the mathematician's demand for existence be dismissed by the physicist as useless rigor. Only a mathematical existence proof can ensure that the mathematical description of a physical phenomenon is meaningful. <sup>15</sup> 설명 | Science, Technology, Engineering, Mathematics의 약자 <sup>16</sup> 참고 | 보통 다양성(diversity)은 성(gender), 인종(race), 사회경제(socioeconomic)적인 배경 등의 문화적 차이를 의미하지만 본 글에서는 전공 별 다양성에 초점을 두었습니다. <sup>17</sup> 참고 | <https://steemit.com/writing/@sungyu1223/the-diversity-bonus-how-great-teams-pay-off-in-the-knowledge-economy-our-compelling-interests> <sup>18</sup> 설명 | 랜덤 포레스트는 부트스트랩(bootstrap)시킨 훈련데이터에 랜덤하게 선택된 서로 다른 변수들을 포함하는 나무 모형(tree model)으로 적합시키는 앙상블(ensemble) 기법으로, 성능을 높이면서 동시에 과적합을 방지합니다.