투빅스 15기 장아연

시계열 데이터, 특히 비정형 시계열 데이터에 대한 딥러닝 모델의 예측 성능이 뛰어나다고 합니다. 이에 따라 어떤 사람은 앞서 살펴본 시계열 모형 및 이론을 배울 필요가 없다고 주장할 수 있습니다. 이 주장에 대해 여러분의 찬성 및 반대 의견을 정리해 제출해주시기 바랍니다.

비정형 시계열 데이터에 대한 딥러닝 모델의 예측 성능이 우수하다고 하여 시계열 모형 및 이론을 배울 필요가 없다고 주장하는 것은 데이터를 분석하고 예측하는데 있어 그 근간인 통계의 중요성을 간과한 결과라고 생각한다.

통계학은 사회 현상을 수량적으로 관찰하고 분석하는 방법을 연구하는 학문이다. 특히 자연 현상과 달리 사회 현상에서 인간의 의도와 가치가 담겨 있어 가치 함축적이며 시공을 초월하는 보편성을 가진 동시에 구체적인 모습은 사회마다 다른 특수성을 보인다.

통계학에서는 데이터에 대한 이해를 바탕으로 하는 기술 통계학과 데이터에 대한 예측을 다루는 기술 통계학으로 구성되어 있다. 그리하여 데이터를 읽고 분포를 살피며 왜 이런 양상을 보이는지 이해하고 이상치들을 처리하며 전반적인 절차를 읽고 분석하는 과정이 중요시되며 통계학에서 크게 자리매김하고 있다. 딥러닝에서는 데이터에 맞게 최적화 하는 과정인 변수 튜닝과 인공 신경망 구성을 통해서 성능을 개선하는 과정이 중요시되고 있다. 딥러닝과 통계학의 공통된 목표인 사회현상을 관찰하고 분석하고 내포된 불확실성에서 생기는 위험을 줄이고 적절한 판단을 내리는 것에서는 두 과정이 필요하고 실제 연구 과정에서 완전히 분리하기 어려운 연관성을 지니기 때문에 어느 한 쪽이 뒤떨어진다고 그것이 사라질 것이라고 생각하는 것은 잘못된 관념이다.

통계 기법과 딥러닝은 모두 데이터를 사용해서 가설을 세우고 여러가지 방식을 사용해 그 가설을 입증하고 입증된 가설을 바탕으로 새로운 예측과 통찰을 이끌어낸다는 공통 목표를 가지고 있다. 이 두 분야는 가설을 입증하는 과정에서 사용하는 모델과 기술에서 차이점을 가지고 있다. 그러나 그 근간에는 통계학이 자리잡고 있다.

인공 신경망에서도 뉴런의 역할을 하는 노드를 연결하고 각 노드의 가중치를 부여해 학습하고 그 결과를 토대로 재조정하는 과정은 기존 확률과 통계학에서도 결과를 집계하고 그 확률을 재조정해 개선하는 형태로 서로 크게 다른 양상을 보이지 않는다.

더 나아가 실제 학습에 사용되는 데이터는 자연과 사회에서 추출한 일부의 사례일 뿐이다. 이를 다듬고 정제하여 비선형적인 상관성을 우수하게 표현해내는 딥러닝 일지라도 실제 세계를 완전히 학습하고 구현해내는 기술은 아니다. 통계학에서 시계열 이론과 모형을 이용하든 딥러닝을 이용해 그 모델을 예측하든 데이터를 통해서 얻은 모델은 현상의 일부만을 해석하거나 일부를 누락해 해석하는 경우가 빈번하다. 우리는 이 상황에서 편협하게 어느 한 쪽이 더 우수하다고 판단하기 보다는 이들이 가진 불확실성과 왜곡됨을 인지하고 모델을 보며 현명하게 예측하고 판단할 수 있는 안목을 기르는 것이 더 중요하다.