

시계열 모형 및 이론을 배울 필요가 없다?

시계열 데이터란 일정한 시간 동안 수집된 일련의 순차적으로 정해진 데이터 셋의 집합을 의미한다. 시계열 데이터는 시간에 관해 순서가 매겨져 있다는 점과, 연속한 관측치는 서로 상관관계를 갖고 있다.

시계열 모형이란 이러한 시계열이 갖고 있는 법칙성을 발견해 모형화한 것으로, 또 이 모형을 통해 미래의 값을 예측하는 것이다. 시계열 모형은 대표적으로 AR, MA, ARMA, ARIMA 가 있는데, AR는 자기상관 모형으로 이전 시점의 데이터가 다음 시점의 데이터에 영향을 준다고 가정한다.

MA는 이동평균을 시계열 모형으로 구성한 것으로, 시점 t 에서 얻게 될 $X(t)$ 의 평균값은 시점 $t-1$ 에서 발생한 error에 선형 변환을 한 것과 같다는 것을 뜻한다.

딥러닝 모델 중 순환 신경망 모델을 살펴보면, 모델은 메모리 셀이 존재하는데 이는 t 시점의 메모리 셀에는 $t-1$ 시점의 메모리 셀이 보낸 값, $t+1$ 시점의 메모리 셀에는 t 시점의 메모리 셀이 보낸 값이 저장되어 있다. 이를 다르게 보면 $\text{lag} = 1, 2, 3, \dots, t-1$ 로 설정하여 비교하는 것을 나타낸다. 또한 메모리 셀에도 가중치가 저장되어 있기 때문에 만약 1시점의 메모리셀 정보가 가중치를 통해 선형 변환을 하여 4시점의 메모리셀에 영향을 많이 주게 되어 loss값이 작게 되면 이는 AR모형의 개념과 비슷하게 된다. 즉, RNN의 모형 내부에서 시계열 모형과 비슷하게 동작하고 있다는 것을 말하게 된다.

시계열 모형들은 통계 이론을 기반으로 구축되어 설명을 할 수 있다. 하지만 딥러닝의 가장 큰 문제점은 설명이 불가능하다는 것이다. 반대로 시계열 데이터에 대하여 시계열 모형보다 딥러닝이 성능이 더욱 뛰어나다. 따라서 시계열 모형 및 이론에 대하여 배울 필요가 없다라기보다는 시계열 모형의 설명 가능한 점과 딥러닝의 뛰어난 성능을 서로 보완하며 발전해 나가야 한다고 생각한다.