RETAIN 모델 설명

핵심

**1. attention weight 사용**

**2. 두 set의 weight 사용**

- visit level attention(시간)

- variable level attention(변수)

🡺 두 layer에서 나온 rnn hidden layer에 적절한 비선형 함수(기본적으론 softmax)를 씌워 예측에 얼마나 영향을 주었는지에 대한 측도로 사용될 수 있는 weight를 뽑아냄.

🡺 뽑아낸 alpha, beta는 서로 곱해진 뒤에, 입력 벡터와 element wise multiplication 함 ( 선형결합). 이 결과가 최종 예측에 사용됨.

**해석방법**

우리 모델에서 window를 7로 설정하였기 때문에 하루 뒤 예측을 위해 이전 7일의 데이터가 사용됨.

bidirection RNN모델에 7일씩 window가 이동되며 입력이 되기 때문에 7 \* 7의 형태가 기본적으로 생성된다고 말할 수 있음.

따라서 시점 layer의 weight alpha는 7\*7\*1

변수 layer의 weight beta는 변수가 각 7개이기 때문에 7\*7\*7 의 형태로 반환됨

이를 결합하여 하나의 기여도로 나타내기 위해 딥러닝 weight인 W와 beta를 먼저 element wise multiplication한 matrix에 alpha matrix곱 연산을 함

이러면 최종적으로 7\*7\*7의 기여도 matrix를 얻을 수 있게 됨( 각 변수별 시점별 기여도 7일치)

대본

다음으로 사용되는 모델은 RETAIN입니다. RETAIN 모델은 기존의 black box형태로 해석이 불가능한 RNN 기반 모델에서 해석력을 보존하기 위해 변형된 모델입니다. 가장 큰 특징은 시점과 변수 보존 해석력 보존을 위해 attention layer가 각각 alpha, beta로 분리되어 학습된다는 점입니다.

attention layer는 무엇일까요, attention layer는 앞서 언급했던 black box 형태인 RNN hidden layer에 비선형함수 softmax값을 씌워 나온 weight을 예측값 계산에 활용하는 매커니즘입니다. 여기서 alpha로 표기된 attention weight가 해석에 사용됩니다.

해석은 다음과 같습니다. 시간 축을 기준으로 변수별로 예측에 어느정도 기여했는지 알 수 있습니다.

해석에 사용되는 기여도 matrix 계산은 다음과 같습니다.

앞서 언급했듯이 모델에서 window를 7로 설정하였기 때문에 하루 뒤 예측을 위해 이전 7일의 데이터가 사용됩니다. 때문에 7 \* 7의 형태가 기본적으로 생성됩니다.

변수가 없고 시점 하나인 alpha layer의 weight는 7\*7\*1, 변수가 7개인 beta layer는 7\*7\*7의 형태로 반환됩니다.

이를 결합하여 하나의 기여도로 나타내기 위해 딥러닝 weight인 W와 beta를 먼저 element wise multiplication한 matrix에 alpha matrix곱 연산을 하면 최종적으로 7\*7\*7의 기여도 matrix를 얻을 수 있게 됩니다.

이렇게 만들어진 matrix의 총계를 내 다음과 같은 해석을 할 수 있었습니다.

예측기여도의가 클수록 진학 색깔을 가진다고 해석하면

예측력이 높았던 동 들에 대해 주거지구는 업무지구보다 상대적으로 날씨의 영향력이 컸던 것을 알 수있었습니다. 특히 마실거리의 경우 다른 카테고리보다 날씨의 영향력이 컸습니다.

예측기여도를 예측에 사용된 날짜별로 총계를 내어 확인해본 결과, 전주 동요일예측이 좋지 않은 다산동의 경우에는 하루전의 판매량이 기여도가 높았고, 비교적 판매량이 예측에 기여하는 정도가 적을 때 기온과 같은 타 변수에 영향력도 확인할 수 있었습니다.