

시계열 예측을 고려한 속성 선택 딥러닝 모델

박광호, Lkhagvadorj Munkhdalai, 류근호*
 충북대학교 데이터베이스 바이오인포매틱스 연구실
 khblack, lhagii, khryu@dblab.chungbuk.ac.kr

Feature Selection Deep Learning Model considering Time Series Prediction

Kwang Ho Park, Lkhagvadorj Munkhdalai, Keun Ho Ryu*
 Database/Bioinformatics Laboratory, Chungbuk National University

요 약

최근 다양한 시계열 데이터의 분석이 딥러닝 방법을 통하여 수행되고 있다. 주로 RNN과 LSTM을 이용하여 많은 시계열 예측이 이루어지고 있다. 하지만 이러한 예측모델을 생성하는데 가장 중요한 것은 어떠한 변수를 얼마나 사용하는지가 중요하다. 이에 대하여, 본 연구에서는 3개의 신경망을 적용하여, 속성을 선택하는 Selection MLP, 속성에 가중치를 부여하는 Extraction MLP 그리고 예측을 진행하는 Prediction MLP로 이루어진 MLP-SEL 구조를 제안한다. 비교를 위하여 다른 순환 신경망에 대하여 시계열 데이터에 대한 예측을 진행하였으며, 그 결과 우리가 제안한 MLP-SEL 모델의 시계열 예측이 좋은 성능을 보였다.

1. 서론

시계열 분석은 질병 사건의 미래 발생 예측, 기후 변화 예측, 스마트 공장의 실시간 이상 탐지, 주가 예측 등과 같은 많은 분야에서 복잡한 문제 중 하나이다. [1-4] 또한 시계열 모델링 프로세스는 높은 변동성, 추세, 계절성 및 노이즈를 포함하는 고유 특성 때문에 매우 어려운 작업이다[5]. 특히 다변량 시계열에서 가장 관련성이 높은 변수와 적절한 지연 길이를 결정하는 것은 어려운 일이다.

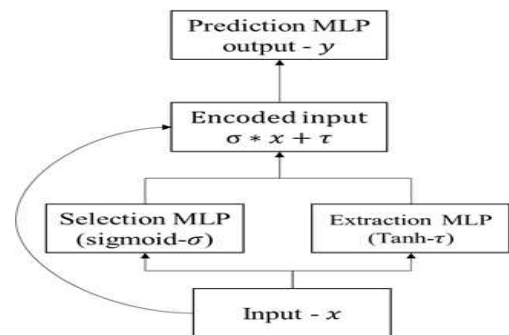
문헌에서, 상당한 수의 연구가 다변량 시계열 예측을 위한 변수 선택 접근법으로 여겨져 왔다.

최근, 반복 신경망(RNN)은 시계열 예측[6-9]에서 점점 더 인기를 얻고 있다. Tank [10, 11]는 희소 정규화를 통해 비선형 인과 변수를 식별하기 위해 구조화된 다중 계층 퍼셉트론(MLP) 또는 반복 신경망(RNN)을 적용하여 클래스 비선형 모델을 제안했다. Choi[12]는 주의 메커니즘을 이용한 변수의 해석 가능성뿐만 아니라 장기 메모리를 선택적으로 사용하는 RNN의 성능을 향상시킬 것을 제안했다. 또한 Attention 기반 신경망은 시간 단계에 걸쳐 숨겨진 상태의 일부를 선택하고, 다중 변수 시퀀스 데이터에 대한 반복 신경망의 해석 가능성을 제공하기 위

해 제안되었다[13, 14].

이 문제를 해결하기 위해, 우리는 입력 변수의 장기적 의존성을 포착하기 위해 게이트 반복 유닛 – GRU 네트워크의 게이트와 유사한 선택 및 추출 계층(MLP-SEL)을 소개한다 [15].

2. MLP-SEL 모델



(그림 1) Overview of the proposed deep learning architecture

그림 1은 세 가지 주요 MLP로 구성된 제안된 딥러닝 모델의 개요를 보여준다. Selection MLP라는 첫 번째 구성 요소는 Prediction MLP에 공급할 입력 변수를 결정하는 가중치를 생성하여 예측 목표

값을 계산한다. 두 번째 Extraction MLP는 항상 추출을 위해 가중치 입력 벡터를 업데이트하는 추가 벡터를 생성한다. 우리는 Selection MLP에서 초기 입력과 생성된 가중치 사이의 곱셈 계층과 스킵 연결을 사용한다. 또한, Extraction MLP에서 생성된 벡터를 가중 입력에 추가하여 목표값 예측을 위한 Prediction MLP를 공급하기 위해 인코딩된 입력 벡터를 얻는다. 추가 신경망의 과제는 동일 입력을 가져와서 다른 역할을 가진 벡터로 인코딩하는 것이다. 따라서, 우리는 분리된 신경망을 사용하여 그들을 훈련시킬 수 있다고 가정한다. Selection MLP의 역할은 Attention 계층의 역할과 유사하기에, Extraction 신경망의 입력에서 추가적인 유용한 정보를 처리할 수 있다.

2.1 Selection MLP

Selection MLP의 목표는 각 입력 변수의 영향을 제어하는 가중치를 생성하여 장기 종속성을 포착하는 것이다. 우리는 데이터 사전 처리 단계에서 과거 값만 입력으로 준비하기 때문에, 중요한 변수와 그 시차를 결정하는 것이 중요하다. 선택 MLP는 이 문제를 해결할 수 있다.

$Y_t = (y_1, y_2, \dots, y_T)$ 를 대상 변수를 나타내도록 합니다. 달력, 시차 및 이동 평균 변수와 같은 다른 변수를 사용한다고 가정하면 타임스탬프의 번째 입력은 다음과 같이 $T^{(i)} tuples(x_j^{(i)}) \in R^n$ 로 나타낼 수 있다. ($i = 1, \dots, T; j = 1, \dots, n$) 우리는 다음과 같이 선택 네트워크에서 가중치(v)를 찾기 위해 MLP를 사용한다.

$$\sigma = \text{sigmoid}(MLP_{\text{selection}}(x; \theta)) \quad (1)$$

여기서 θ 는 선택 네트워크의 가중치 매개 변수를 나타냅니다. 이 계층의 출력은 시그모이드(*sigmoid*) 활성화 함수를 이용한다. 따라서 선택 네트워크에서 생성된 벡터는 0과 1 사이로, 변수에 대한 상황에 따라 달라지는 중요도 가중치이다. 입력 변수 중 하나에 해당하는 가중치가 0에 가까우면 변수나 지연에 가중치를 곱한 값을 예측 출력에 영향을 줄 수 없습니다. 다시 말하면 장단기 메모리 (LSTM)의 망각 게이트와 유사하며, Selection MLP 네트워크는 어떤 정보를 잊어야 할 지를 입력 데이터에서 결정한다. 또한, 적응 가중치 사용의 중요성은 변화하는 환경에서 시간이 지남에 따라 일부 변수의 중요성이

변경될 수 있으며, 이때, 적응 가중치가 입력 변수를 신속하게 재조정하는 데 도움이 될 수 있다.

2.2 Extraction MLP

Selection MLP 네트워크를 사용하여 입력 데이터에서 중요한 정보를 검색하지만, 추가 새로운 정보를 생성하지는 않는다. 즉, 새로운 정보를 생성하기 위해 제안된 아키텍처에 Extraction MLP 네트워크를 추가한다. 우리는 Extraction MLP의 출력으로 쌍곡선 접선 활성화 함수(*tanh*)를 설정한다. 이 계층은 데이터가 0과 1의 범위 내에서 정규화되기 때문에 -1과 2 사이의 입력 변수를 인코딩할 수 있는 -1과 1 사이의 벡터를 생성한다.

$$\tau = \tanh(MLP_{\text{extraction}}(x; \omega)) \quad (2)$$

여기서 ω 는 추출 MLP의 가중치 파라미터를 나타낸다.

Extraction MLP 네트워크는 게이트 순환 유닛 (GRU) 네트워크의 업데이트 게이트와 유사한 기능을 하며, 가중 입력을 업데이트하기 위해 새로운 정보를 생성한다.

2.3 Encoding Layer

Encoding Layer는 Selection MLP에 의해 생성된 중요도 가중치, 업데이트 네트워크 및 입력 변수에서 벡터를 업데이트합니다. 우리는 다음과 같이 입력 변수를 인코딩하기 위해 스킵 연결을 사용한다.

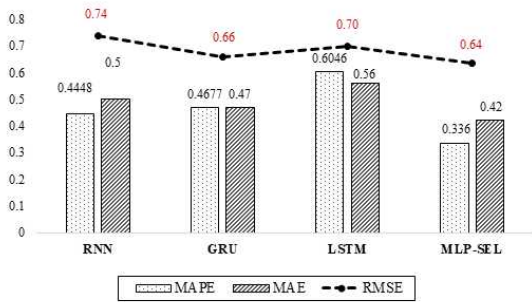
$$x' = \sigma \odot x + \tau \quad (3)$$

여기서 \odot 는 요소별 곱셈을 나타내고, $\sigma \in [0, 1]$ 은 중요 가중치를 나타내며, $\tau \in [0, 1]$ 은 업데이트 가중치를 나타낸다.

2. 실험 및 결과

제안된 MLP-SEL의 효과를 입증하기 위해, 우리는 모델이 시계열 데이터에서 다른 반복 신경망을 비교하도록 훈련 시킨다. 데이터는 De Vito et al., [16]에서 공개한 Air pollution 벤치마크 데이터를 사용하였고, 12개의 변수와 2004년 3월 10일부터 2005년 4월 4일까지의 시간당 기록된 데이터 9357개의 변수를 사용하였다. 훈련의 비교로는 RNN, GRU,

LSTM 알고리즘을 사용하였다. 아래 그림2에 각 기법에 대한 결과를 그래프로 나타내었다.



(그림 2) MLP-SEL 벤치마크 실험 결과

4. 결론

본 연구에서는 3개의 기능 레이어를 적용하여, 속성을 선택하는 Selection 레이어, 속성에 가중치를 부여하는 Extraction layer 그리고 예측을 진행하는 Prediction 레이어로 이루어진 MLP-SEL 구조를 제안한다. 비교를 위하여 다른 순환 신경망에 대하여 시계열 데이터에 대한 예측을 진행하였으며, 그 결과 우리가 제안한 MLP-SEL 모델의 시계열 예측이 좋은 성능을 보였다.

Acknowledgement

이 논문은 2020년도 미래창조과학부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (No.2020R1A2B5B02001717)

참고문헌

[1] Yang, Eunjoo, Hyun Park, Yeon Choi, Jusim Kim, Lkhagvadorj Munkhdalai, Ibrahim Musa, and Keun Ryu. "A simulation-based study on the comparison of statistical and time series forecasting methods for early detection of infectious disease outbreaks." *International journal of environmental research and public health* 15, no. 5 (2018): 966.

[2] O’Gorman, Paul A., and John G. Dwyer. "Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events." *Journal of Advances in Modeling Earth Systems* 10, no. 10 (2018): 2548–2563.

[3] Sonntag, Daniel, Sonja Zillner, Patrick van der Smagt, and Andras Lorincz. "Overview of the CPS for smart factories project: Deep learning, knowledge acquisition, anomaly detection and intelligent user interfaces." In *Industrial internet of things*, pp. 487–504. Springer, Cham, 2017.

[4] Munkhdalai, Lkhagvadorj, Tsendsuren Munkhdalai, Kwang Ho Park, Tsatsral Amarbayasgalan, Erdenebileg Erdenebaatar, Hyun Woo Park, and Keun Ho Ryu. "An End-to-End Adaptive Input Selection With Dynamic Weights for Forecasting Multivariate Time Series." *IEEE Access* 7 (2019): 99099–99114.

[5] Bowerman, Bruce L., and Richard T. O’Connell. *Time series and forecasting*. North Scituate, MA: Duxbury Press, 1979.

[6] L. Wang, Z. Cao, Y. Xia, and G. de Melo. "Morphological segmentation with window LSTM neural networks." *AAAI-16*, Phoenix, Arizona, USA, 2016.

[7] T. Guo, Z. Xu, X. Yao, H. Chen, K. Aberer, and K. Funaya. "Robust online time series prediction with recurrent neural networks." *DSAA’2016*, Montreal, Canada, 2016, pp. 816–825.

[8] T. Lin, T. Guo, and K. Aberer. *Hybrid neural networks for learning the trend in time series*. *IJCAI*, Vienna, Austria, 2017.

[9] T. Guo, and T. Lin. "Multi-variable LSTM neural network for autoregressive exogenous model." *arXiv preprint arXiv:1806.06384*, 2018.

[10] Tank, Alex, Ian Cover, Nicholas J. Foti, Ali Shojaie, and Emily B. Fox. "An interpretable and sparse neural network model for nonlinear granger causality discovery." *arXiv preprint arXiv:1711.08160* (2017).

[11] A. Tank, I. Covert, N. Foti, A. Shojaie, and E. Fox. "Neural granger causality for nonlinear time series." *arXiv preprint arXiv:1802.05842*, 2018.

[12] E. Choi, M.T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart. "Retain: An interpretable predictive model for healthcare using reverse time attention mechanism." *NIPS 2016*, Barcelona, Spain, 2016, pp. 3504–3512.

[13] Y. Qin, D. Song, H. Chen, W. Cheng, G.

Jiang, and G. Cottrell. "A dual-stage attention-based recurrent neural network for time series prediction." arXiv preprint arXiv:1704.02971, 2017.

[14] Y. G. Cinar, H. Mirisae, P. Goswami, E. Gaussier, A. Ait-Bachir, and V. Strijov. "Position-based content attention for time series forecasting with sequence-to-sequence rnns." ICONIP 2017, Guangzhou, China, 2017, pp. 533-544.

[15] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014, arXiv:1409.1259. [Online]. Available: <https://arxiv.org/abs/1409.1259>

[16] S. De Vito, M. Piga, L. Martinotto and G. Di Francia. "CO, NO₂ and NO_x urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization." Sens Actuators B Chem, vol. 143, no. 1, pp. 182-191, 2009.