

# Temporal Fusion Transformer 모델을 활용한 다층 수평 시계열 데이터 분석

김인경, 김대희, 이재구\*  
국민대학교 컴퓨터공학과  
\*jackoo@kookmin.ac.kr

## Multi-horizon Time Series Forecasting Using Temporal Fusion Transformer

Inkyung Kim, Daehee Kim, Jaekoo Lee\*  
Dep. of Computer Science, Kookmin University

### 요 약

시계열 형태의 데이터는 다양한 분야에서 수집되고 응용되기 때문에 정확한 시계열 예측은 많은 분야에서 운영 효율성을 높일 수 있는 중요한 분석 방법으로 고려된다. 그중 다층 수평 예측은 사용자에게 전반적인 시계열 데이터 경향성을 제공할 수 있다. 하지만 다양한 정보를 포함하는 시계열 데이터는 데이터에 내재한 이질성(heterogeneity)까지 포괄적으로 고려한 방법을 통해서만 정확한 예측을 할 수 있다. 하지만 지금까지 많은 시계열 분석 모델들이 데이터의 이질성을 반영하지 못했다. 이러한 한계를 보완하고자 우리는 Temporal Fusion Transformer 모델을 사용하여 실생활과 밀접한 관련이 있는 데이터에 적용하여 이질성을 고려한 향상된 예측을 수행하였다. 실제, 주식 데이터와 미세 먼지 데이터와 같은 실생활 시계열 데이터에 적용하였고 실험 결과 기존 모델보다 Mean Squared Error(MSE)가 0.3487 낮은 것을 확인하였다.

### 1. 서론

시계열 예측 문제는 과거의 관측에 기반해서 미래를 얼마나 정확하게 예측할 수 있는지에 대해서 다룬다. 시계열 데이터는 관측치가 시간적 순서를 가진 데이터를 의미한다. 시계열 데이터는 일반적으로 시간 순서대로 도착하여 축적된다. 많은 데이터가 시계열 형태를 가지고 있는 만큼 정확한 예측은 경제, 의료 등 다양한 분야에서 운영 효율성을 높일 수 있다. 다층 수평(multi-horizon) 예측은 시계열 데이터를 사용한 기계 학습에서 매우 중요한 문제 중 하나이다. 다층 수평 예측은 사용자들에게 전반적인 경향에 대한 짐작을 제공하기 때문에 사용자들이 미래에 대한 계획을 세울 수 있도록 도움이 될 수 있다.

하지만 시계열 데이터는 보통 다차원 형태를 이루면서 다양한 정보를 포함하기 때문에 전통적인 방법으로는 정확한 예측이 어렵다. 예를 들면 특정 가게에서 소비자들의 구매를 예측하려고 한다면 각 상품

에 대한 구매율과 함께 가게에 대한 정보, 소비자들의 특성 등 다양한 정보가 필요하다. 이렇게 이질성이 높은 데이터는 서로의 상호작용에 대한 정보를 매우 적게 제공하기 때문에 다층 수평 시계열 예측을 더욱 더 어렵게 만든다. 최근 깊은 신경망의 발전으로 순환 신경망 등을 사용한 방법[4][5][6][7]이 제안되었지만, 여전히 다양한 입력값을 활용한 예측에는 한계가 있다.

이러한 점을 개선하고자 본 논문에서는 Temporal Fusion Transformer (TFT)[1]를 주식 데이터와 미세 먼지 데이터에 적용하여 예측 결과를 확인하였으며, 정량적으로 Long Short-Term Memory (LSTM) 모델[2]과 성능을 비교하였다<sup>1</sup>. 실험 결과, TFT를 사용했을 때 가장 정확한 예측이 가능한 것을 확인하였다.

### 2. 관련 연구

수평(horizon) 예측은 일반적으로 한 단계 앞선 예

<sup>1</sup> 관련된 코드는 [https://github.com/IKKIM00/stock\\_and\\_pm2.5\\_prediction\\_using\\_TFT](https://github.com/IKKIM00/stock_and_pm2.5_prediction_using_TFT)에서 확인할 수 있습니다

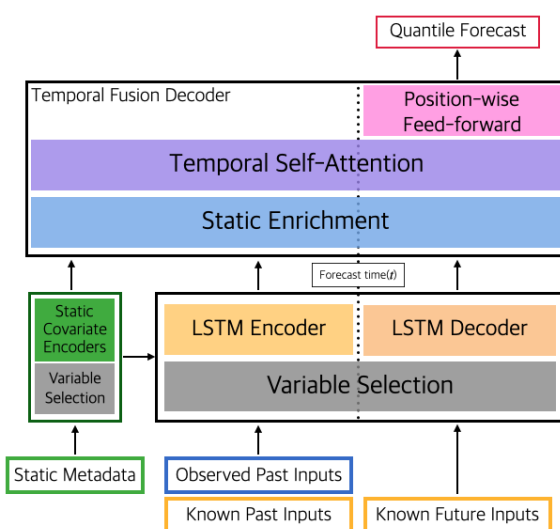
측에 중점을 둔다. 다층 수평 예측은 여러 미래 시간 단계에서 관심 변수를 예측하는 것을 의미한다. 시계열 예측은 단변량 (univariate) 분석과 다변량 (multivariate) 분석으로 나뉘어진다. 단변량 분석은 과거의 값을 통해서 미래의 관측값을 예측하는 접근을 의미한다. 반면 다변량 분석은 독립 변수로 사용하는 추가 시계열 데이터를 고려하여 단변량 분석을 확장한 것을 뜻한다.

깊은 신경망을 사용하기 이전의 시계열 예측 문제는 통계적 모델을 사용하여 단변량 분석을 진행하였다. 기존의 통계적 모델로는 simple moving average(SMA), exponential smoothing (ETS)나 autoregressive integrated moving average (ARIMA) 등이 있다. 하지만 위의 모델들은 상대적으로 단순하고, 정확한 다변량 분석에는 한계가 있다[3]. 최근 다층 수평 예측에는 깊은 신경망이 많이 사용되고 있으며 그 중 많은 모델들이 순환 신경망 구조를 채택하였다 [4][5][6][7]. 특히 최근에는 어텐션(attention) 기반의 방법을 사용한 구조들이 제안되었다[6][7]. 하지만 아직 많은 모델이 데이터의 다양한 형태인 이질성을 고려하지 못한다[4][5][7]. DSTP-RNN[6]의 경우 여러 종류의 입력 값을 고려하지만, 실제 값을 가진 데이터 집합만 사용해서 실험을 진행하였다.

하지만 시계열 데이터는 실제 값 이외에도 다양한 형태를 가지고 있어서 이를 모두 반영해야 정확한 예측이 가능하다. TFT 모델은 다양한 종류의 데이터와 여러 형태의 값들을 입력값으로 사용할 수 있다. 본 연구에서는 이전 연구들이 데이터의 이질성을 반영하지 못하는 한계점을 해결하기 위해서 TFT 모델을 사용하였다.

### 3. TFT를 활용한 시계열 데이터 예측

#### 3.1 Temporal Fusion Transformer(TFT) 모델



(그림 1) TFT 모델[1] 구조.

TFT 모델[1]은 어텐션 방법에 기반한 구조로 되어 있다. 다층 수평 예측을 위한 모델로 다양한 형태의

데이터를 입력 값으로 사용한다. 또한, 이전 모델들이 가진 한계점들을 보완하여 다양한 유형의 입력값을 고려하도록 설계되어 있다. 다층 수평 예측은 알려진 미랫값, 정적 공변량 값과 외인성의 시계열 데이터를 포함하는 복잡한 형태를 이루며 이 정보들은 상호 작용한다. 예를 들면 소매업 데이터[8]는 상품과 가게에 대한 메타데이터(metadata)를 포함하면서 일일 수준의 외인성 시간 변화에 따른 입력값을 포함한다. 알려진 미랫값으로는 정해진 값으로 볼 수 있는 달에서 며칠 인지, 세일 정보 등을 사용할 수 있다.

TFT 모델은 다양한 규모의 시간적 관계성을 파악하기 위해서 지역적 처리를 위한 순환 신경망 층과 장기 의존성을 위한 셀프-어텐션(self-attention) 층을 포함한다. 또한 유의미한 특징 추출을 위한 변수 선택 층과 중요하지 않은 특징을 억제하기 위한 개폐(gating) 층을 사용한다. 그래서 다양한 시나리오에서 좋은 성능을 보인다. 또한 분위수 손실 함수를 사용하여 결과의 q-분위수를 예측한다.

TFT 모델은 다음과 같은 입력 값들을 사용하여 모델을 구성한다. 먼저 주어진 시계열 데이터에서 고유한 존재  $i$ 가 존재한다고 가정한다. 예를 들면  $i$ 는 소매업에서는 가게를 의미한다.

$$\hat{y}_i(q, t, \tau) = f(q, \tau, y_{i,t-k:t}, z_{i,t-k:t}, x_{i,t-k:t+\tau}, s_i) \quad (1)$$

각 존재  $i$ 는 정적 공변량 값  $s_i$ 와 입력 값  $x_{i,t}$ , 스칼라 값  $y_{i,t}$  모두와 연관이 있다. 시간에 의존적인 입력 특징  $x_{i,t} = [z_{i,t}^T, x_{i,t}^T]$ 은 매 시간마다 측정되고 미리 알 수 없는 관측된 입력 값  $z_{i,t}$ 와 사전에 결정되는 알려진 입력 값  $x_{i,t}$ 로 나뉘어진다.  $\tau$ 는 분위수를 의미한다. 모델의 입력과 각각의 분위수 예측은 위의 수식(1)과 같이 이루어진다[1].

#### 3.2 데이터 설명

실험은 두 개의 공개 데이터를 사용하여 진행하였으며 실생활과 밀접한 관련이 있는 데이터로 선정하였다. 데이터는 Yahoo finance 에서 Carriage Services 기업[9]의 주식 데이터와 베이징의 미세먼지 데이터[10]를 사용하여 실험을 진행하였다.

주식 데이터는 기본적으로 예측이 어려운데, 이는 가격의 변동이 다양한 요인에 근거하고 복잡한 역학을 가지고 있기 때문이다. Yahoo finance 에서는 각 기업별로 시가, 고가, 저가, 종가, 거래량, 조정된 종가에 대한 값을 제공한다. 실험에는 시가, 고가, 저가, 종가와 거래량 값을 사용하여 진행하였으며 1996년 8월 9일부터 2021년 4월 8일까지의 일별 데이터를 사용했다.

베이징 미세먼지 데이터[10]에는 데이터의 측정 시점, 이슬점, 온도, 기압, 풍향, 풍속, 적설 시간, 누적 강우 시간, 미세먼지 농도에 대한 정보를 포함한다. 본 논문의 실험에서는 위의 모든 데이터를 사용해서 미세먼지 농도를 예측하였다.

### 3.3 전처리 방법

먼저 미세 먼지 데이터[10]의 경우 다음과 같이 데이터를 전처리를 진행하였다. 훈련 집합, 검증 집합, 테스트 집합을 각각 2010년 1월 1일~2012년 12월 31일, 2013년 1월 1일~2013년 12월 31일, 2014년 1월 1일~2014년 12월 31일로 설정했다. 미세 먼지 데이터[10]의 경우 시간 단위로 측정되었다. 그래서 예측 주기는 24시간의 이슬점, 온도, 기압, 풍향, 풍속, 적설 시간, 누적 강우 시간을 사용해서 다음 한시간의 미세 먼지 농도를 예측할 수 있도록 설정했다. TFT 모델에 적용하기 위한 입력값은 다음과 같다. 데이터가 측정된 날짜와 시간 값은 알려진 입력값으로 사용하였다. 정적 공변량 값으로는 데이터가 측정된 장소에 대한 정보를 사용했다. 관측된 입력 값으로는 이슬점, 온도, 기압, 풍향, 풍속, 적설 시간, 누적 강우 시간을 모두 사용하였다.

주식 데이터[9]의 경우 훈련 집합, 검증 집합, 실험 집합을 각각 1996년 8월 9일~2011년 12월 31일, 2012년 1월 1일~2016년 12월 31일, 2017년 1월 1일~2020년 4월 7일로 설정했다. 측정 주기가 하루 단위였기 때문에 예측 주기를 50일로 설정하여 다음 날 종가를 예측하도록 설정하였다. 입력값으로는 시가, 고가, 저가, 거래량 데이터를 사용했다. TFT 모델에 적용하기 위한 입력값은 다음과 같이 설정했다. 알려진 입력값으로는 주식의 데이터가 측정된 달을 사용하였다. 정적 공변량 값은 하나의 기업 데이터만 사용하였기 때문에 하나의 정수값으로 설정했다. 관측된 입력값으로는 고가, 저가와 거래량 값을 사용하였다.

## 4. 실험 결과

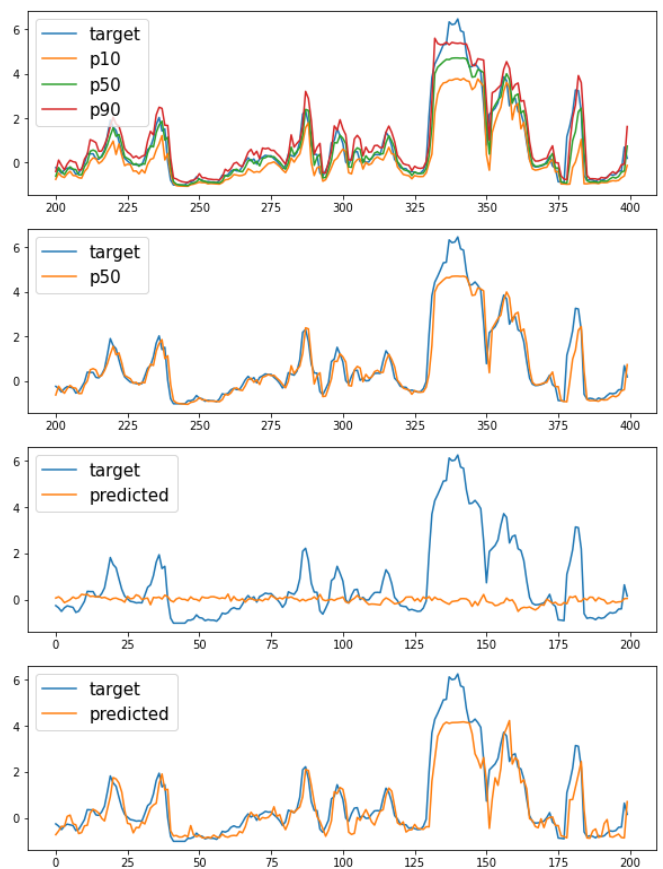
TFT 모델의 성능을 평가하기 위한 기준선 모델로는 단층 LSTM 모델과 다층 LSTM 모델을 사용했다. 실험 결과 모든 데이터 집합에 TFT 모델을 사용했을 때 가장 정확한 예측이 되는 것을 확인할 수 있다. TFT 모델의 경우 10<sup>th</sup>, 50<sup>th</sup>, 90<sup>th</sup>로 분위수 예측하도록 설정되어 있기 때문에 각각의 분위수에 대한 예측 값을 (그림 2)와 (그림 3)의 (a)에서 확인할 수 있다. 기준선 모델과 비교하기 위한 값으로는 50<sup>th</sup>에서의 예측 값을 사용할 수 있으며 명확한 비교를 위해 (그림 2)와 (그림 3)의 (b)에 TFT 모델의 50<sup>th</sup> 예측 결과를 출력했다.

<표 1> 각 데이터 집합의 모델 별 MSE 값.

		MSE 값		
		TFT[1]	단층 LSTM[2]	다층 LSTM[2]
데이터 종류	미세 먼지 데이터[10]	0.1310	1.0974	<b>0.1024</b>
	주식 데이터[9]	<b>0.0617</b>	0.4399	0.1402

먼저 미세 먼지 농도에 대한 예측 결과를 확인하면 (그림 2)의 (a)와 같이 TFT 모델을 사용한 실험 결과

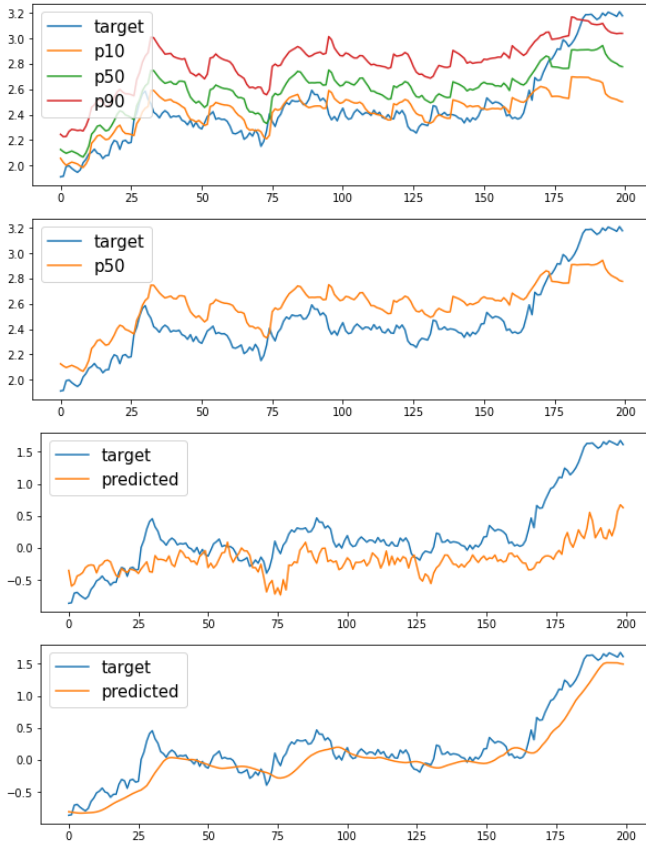
가 가장 정확한 것을 확인할 수 있다. 반면 (그림 2)의 (c)와 같이 단층 LSTM 모델을 사용했을 때 예측 결과가 가장 부정확한 것을 확인할 수 있다. 이러한 점을 통해서 단층 LSTM은 예측을 위한 모델의 수용력이 충분하지 못했다고 해석할 수 있다. (그림 2)의 (d)를 통해서 다층 LSTM을 사용했을 때 모델이 비교적 정확하게 미세 먼지 농도를 예측하는 것을 확인할 수 있다. 또한 <표 1>에서 확인할 수 있는 것처럼 다층 LSTM 모델을 사용했을 때 Mean Squared Error(MSE) 값이 TFT 모델에 비해 상대적으로 조금 더 낮은 것을 확인할 수 있다. 하지만 (그림 2)의 (b)와 (d)에서 확인할 수 있는 것처럼 TFT 모델이 상대적으로 더 정확하게 미세 먼지 농도를 예측하는 것을 확인할 수 있다.



(그림 2) 미세 먼지 데이터를 사용한 미세 먼지 농도 예측 결과. 순서대로 (a)TFT 모델의 예측 결과, (b)TFT 모델의 50<sup>th</sup> 예측 결과, (c)단층 LSTM 모델의 예측 결과, (d)다층 LSTM 모델의 예측 결과.

주식 데이터의 실험 결과는 아래와 같다. LSTM 모델을 사용한 실험 결과인 (그림 3)의 (c)와 (d)를 보면 TFT 모델을 사용했을 때의 결과인 (그림 3)의 (a)와 비교했을 때 예측 성능의 차이가 매우 크다는 것을 확인할 수 있다. 미세 먼지 농도 데이터 집합에 비해 주식 데이터 집합의 경우 LSTM 모델만을 사용했을 때 더 제대로 예측이 되지 않았다. 상대적으로 주식 데이터 집합의 입력 값의 종류가 3개로 더 적어서 미세 먼지 농도 데이터 셋에 비해 실험 성능이 더 안 좋

은 것으로 해석된다. (그림 3)의 (b)는 50<sup>th</sup> 분위 예측 결과를 보여주며 이는 (그림 3)의 (c),(d)와 비교했을 때 TFT 모델을 사용했을 때 예측의 결과가 훨씬 우수한 점을 보여준다. 또한 <표 1>의 결과를 통해서 TFT 모델의 MSE 값이 LSTM 모델에 비해 훨씬 낮은 것을 확인할 수 있다.



(그림 3) 주식 데이터를 사용한 예측 결과. 순서대로 (a)TFT 모델의 예측 결과, (b)TFT 모델의 50<sup>th</sup> 예측 결과, (c)단층 LSTM 모델의 예측 결과, (d)다층 LSTM 모델의 예측 결과.

## 5. 결론

실생활에서 사용되는 다수의 데이터 종류가 시계열 데이터라는 점에서 정확한 다층 시계열 분석은 여러 분야에서 운영 효율성을 높일 수 있게 도와준다. 하지만 시계열 데이터는 다양한 종류의 정보를 포함하기 때문에 이러한 데이터의 이질성을 모두 반영해서 예측해야 정확한 예측 결과를 얻을 수 있다. 그래서 다양한 형태의 모든 데이터를 반영해서 예측할 수 있는 TFT 모델을 사용해서 실생활과 밀접한 관련이 있는 데이터에 적용하여 예측 성능을 확인했다. 미세먼지 농도 데이터와 주식 데이터를 사용하여 실험을 진행하였고 실험 결과 두 데이터 집합 모두 TFT 모델을 사용했을 때의 성능이 가장 높은 것을 확인할 수 있었다. 이후 과정으로 더 다양한 데이터 집합에 TFT 모델을 적용하여 실험을 진행할 예정이다.

## 사사

이 성과는 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-01826, AI 기반 선도적 실전문제해결 연구인재 양성)

## 참고문헌

- [1] Lim, Bryan, et al., "Temporal fusion transformers for interpretable multi-horizon time series forecasting", arXiv preprint arXiv, 1912.09363, 2019.
- [2] Hochreiter, Sepp, et al., "Long short-term memory", Neural Computation, 9, 8, 1735-1780, 1997.
- [3] Hewamalage, Hansika, et al., "Recurrent neural networks for time series forecasting: Current status and future directions", International Journal of Forecasting, 37.1, 388-427, 2021.
- [4] Salinas, David, et al., "DeepAR: Probabilistic forecasting with autoregressive recurrent networks", International Journal of Forecasting, 36.3, 1181-1191, 2020.
- [5] Wen, Ruofen, et al., "A multi-horizon quantile recurrent forecaster", arXiv preprint arXiv, 1711.11053, 2017.
- [6] Liu, Yeqi, et al., "DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction", Expert Systems with Applications, 143, 113082, 2020.
- [7] Fan, Chenyou, et al., "Multi-horizon time series forecasting with temporal attention learning", Proceedings of the 25<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2527-2535, 2019.
- [8] C. Favorita, Corporacion favorita grocery sales forecasting competition, 2009.
- [9] Yahoo Finance, Carriage Services, Inc., 2021.
- [10] Liang, X., Zou, T., et al., "Assessing Beijing's PM 2.5 pollution: severity, weather impact, APEC and winter heating, Proceedings of Royal Society, 471, 2182, 20150257, 2015.