

미세먼지 예측 성능 개선을 위한 시공간 트랜스포머 모델의 적용

김영광

(주)웨슬리퀘스트
(glory@wesleyquest.com)

김복주

우리에프아이에스(주)
(bjkim2004@gmail.com)

안성만

국민대학교
(sahn@kookmin.ac.kr)

미세먼지는 폐나 혈관에 침투해 각종 심장 질환이나 폐암 등의 호흡기 질환을 일으키는 것으로 보고되고 있다. 지하철은 일 평균 천만 명이 이용하는 교통수단으로, 깨끗하고 쾌적한 환경조성이 중요하나 지하철역을 통과하는 지하철의 운행 특성과 터널에 갇힌 미세먼지가 열차 풍으로 인해 지하철역사로 이동하는 등의 문제로 지하철역사의 미세먼지 오염도는 높은 것으로 나타나고 있다. 환경부와 서울시는 지하철역사 공기질 개선대책을 수립하여 다양한 미세먼지 저감 노력을 기울이고 있다. 스마트 공기질 관리 시스템은 공기질 데이터 수집 및 미세먼지 농도를 예측하여 공기질을 관리하는 시스템으로 미세먼지 농도 예측 모델이 중요한 구성 요소이다. 그동안 시계열 데이터 예측에 관한 다양한 연구가 진행되어왔지만, 지하철 역사의 미세먼지 농도 예측과 관련해서는 통계나 순환신경망 기반의 딥러닝 모델 연구에 국한되어 있다. 이에 본 연구에서는 시공간 트랜스포머를 포함한 4개의 트랜스포머 기반 모델을 제안한다. 서울시 지하철 역사의 대합실을 대상으로 한 시간 후의 미세먼지 농도 예측실험을 수행한 결과, 트랜스포머 기반 모델들의 성능이 기존의 ARIMA, LSTM, Seq2Seq 모델들에 비해 우수한 성능을 나타냄을 확인하였다. 트랜스포머 기반 모델 중에서는 시공간 트랜스포머의 성능이 가장 우수하였다. 데이터 기반의 예측을 통하여 운영되는 스마트 공기질 관리 시스템은 미세먼지 예측의 정확도가 향상될수록 더욱더 효과적이고 에너지 효율적으로 운영될 수 있다. 본 연구 결과는 스마트 공기질 관리 시스템의 효율적 운영에 기여할 수 있을 것으로 기대된다.

주제어 : 시계열 데이터 분석, 미세먼지 농도 예측, 어텐션 기법, 시공간 트랜스포머, 타임투벳 임베딩

논문접수일 : 2022년 3월 3일 논문수정일 : 2022년 3월 24일 게재확정일 : 2022년 3월 25일
원고유형 : 급행논문 교신저자 : 안성만

1. 서론

미세먼지에 대한 노출은 심장 및 폐 관련 질환 등을 발생시키거나 악화시킬 수 있으며, 미세먼지 농도가 높은 곳에서 오래 노출되는 경우 심혈관 질환, 호흡기 질환, 폐암 발생의 위험이 증가한다(질병관리청, 2020). 미세먼지가 사회적 이슈가 되면서 지하철과 지하철 역사의 공기질에 관한 관심도 커지고 있다. 제3차 지하철역사 공기

질 개선대책(2018~2022)(환경부, 2018)에 따르면 지하철은 일평균 천만 명이 이용하는 교통수단으로 깨끗한 환경이 중요하지만, 지하철 역사의 미세먼지 평균 오염도는 대규모점포, 지하도상가 등 21개 다중이용시설 중 실내주차장 다음으로 높은 오염도를 나타내고 있고, 지하철은 지하철역을 통과하는 운행 특성상 다른 대중교통 차량보다 미세먼지 농도가 높은 상태인 것으로 나타난다. 입자크기에 따라 직경 $10\mu\text{m}$ 이하인 먼지

를 미세먼지(PM10), 직경 $2.5\mu\text{m}$ 이하인 것을 초미세먼지(PM2.5)라고 하는데, 밀폐된 공간을 달리는 철도차량에서 발생하거나 지상으로부터 유입되는 미세먼지는 터널에 갇히게 되고 이렇게 터널에 갇힌 미세먼지는 지하철 차량 이동 시 열차 풍에 의한 피스톤 효과로 지하철 역사로 이동되어 역사의 공기를 오염시킨다(이현욱, 2020).

서울시는 2019년부터 2022년까지 총 7,958억원의 사업비를 투자하여 역사 환기시스템 개량, 스마트 공기질 관리 시스템 구축 등 지하역사 공기질 개선(서울특별시청, 2022)을 위해 노력하고 있다. 스마트 공기질 관리 시스템은 역사 내·외부 공기질 빅데이터를 수집 및 미세먼지 농도 예측을 통해 공조 환기 설비를 제어하여 공기질을 관리하는 시스템이다. 지하역사 공조 환기 설비 운영비는 연간 137억 원(2018년 기준)으로, 미세먼지 농도 예측을 통해 예측치가 높은 경우 설비를 사전 가동하고, 예측치가 낮은 경우 설비 운영을 줄이거나 중단하는 등 설비가동을 최적화함으로써 운영비 절감 및 유지보수 비용을 최소화할 수 있어 공조설비 운영비를 10% 이상(연간 13.7억) 절감할 수 있을 것으로 기대된다(권순박, 2018).

공기질 데이터는 시간의 흐름에 따라 관측되는 시계열 데이터로 공기질 예측을 위해서는 다양한 시계열 분석 기법들이 활용되고 있다. 데이터 마이닝 기법들 이전의 전통적인 시계열 분석은 지수평활법(Hyndman, Koehler, Ord, & Snyder, 2008)이나 ARIMA(Box, Jenkins, Reinsel, & Ljung, 2015)와 같은 통계 모델들이 활용되었다. ARIMA는 학계나 산업계에서 적용되고 있는 대표적인 시계열 분석 방법으로, 서울시의 공기질 예측연구에 활용되었다(오종민, 신현수, 신예슬, 정형철, 2017; 이정영, 이종현, 이영재, 김록호, 한진

석, 2007). 하지만 ARIMA는 데이터의 안정성(stationary) 등 이론적 요구사항을 고려하지 않는 경우 정확한 예측이 어렵고(Lara-Benítez, Carranza-García, & Riquelme, 2021), 데이터가 비선형적인 관계를 가지는 경우 정확도가 저하되며(Hamid & Sodoudi, 2016), 현대에서 생성되는 크고 다양한 형태를 가지는 시계열 자료에서 좋은 예측 성능을 발휘하지 못하는 단점(진세중, 조형준, 2020)이 있는 것으로 알려진다.

최근에는 인공지능 기술의 발전에 따라서 기계학습과 딥러닝 모델을 활용한 시계열 데이터 예측 연구가 활발히 이루어지고 있다. 인공신경망(Artificial Neural Networks)은 데이터 간의 비선형 관계를 모델링할 수 있는 특성 때문에 많은 경우 통계적 모형들보다 나은 성능을 보인다. 특히 데이터 분포에 대한 이론적 가정 없이 시간적 패턴을 자동으로 추출할 수 있어서 전처리에 대한 노력을 줄여준다는 장점이 있다(Lara-Benítez et al., 2021). 인공신경망은 각 입력데이터가 독립적이라고 가정한다. 하지만 문장이나 시계열 데이터와 같은 순차적 데이터를 입력데이터로 사용하는 경우, 현재 시점의 입력값은 과거의 입력값들의 영향을 받을 수밖에 없다(안성만, 정여진, 이재준, 양지현, 2017). 이러한 순차적 데이터를 모델링 하도록 설계된 순환신경망(Recurrent Neural Networks)은 입력과 출력 간의 관계뿐 아니라 입력데이터들 사이의 관계를 학습할 수 있는 특성 때문에 시계열 데이터 분석에서 활발하게 적용되었다. 하지만 전통적인 순환신경망은 기울기 소실 문제(Bengio, Simard, & Frasconi, 1994)로 인해 시계열 입력데이터의 길이가 증가할수록 장기 의존 관계를 학습하기 어려운 단점이 있다.

이러한 문제점을 해결하는 방법으로 제안된

LSTM (Hochreiter & Schmidhuber, 1997)이나 GRU(Cho et al., 2014b)는 기계번역, 음성인식, 이미지 캡셔닝(image captioning) 등 응용 분야에서 모델의 성능을 높였다. 이러한 순환신경망의 응용모델 중 입력데이터를 압축해서 고정된 크기의 문맥 벡터(context vector)를 생성하는 인코더와 이 벡터를 활용하여 출력 데이터를 생성하는 디코더로 구성된 시퀀스 투 시퀀스(sequence to sequence, Seq2Seq) 모델(Cho et al., 2014b; Sutskever, Vinyals, & Le, 2014)이 시계열 예측 분야에서도 우수한 성능을 보이면서, 시계열 예측에서 일반적으로 활용하는 모델이 되었다(T. Chen et al., 2018; Grigsby, Wang, & Qi, 2021). Seq2Seq 구조는 입력 시퀀스의 길이가 증가할수록 성능이 급격하게 감소하는 문제가 있다(Cho et al., 2014a). 이 문제를 해결하려는 방법으로 디코더에서 출력 결과를 예측하는 시점마다 인코더의 은닉상태(hidden state)를 다시 한번 참고하여 예측 결과와 연관이 있는 부분에 좀 더 집중하여 성능을 향상시킨 어텐션(attention) 기반의 인코더-디코더 네트워크(Bahdanau, Cho, & Bengio, 2014)가 제안되었다.

2017년 트랜스포머 모델(Vaswani et al., 2017)이 제안된 후로 트랜스포머 모델은 자연어 처리, 컴퓨터 비전, 오디오 처리와 같은 다양한 분야에서 큰 성과를 거두었다. 트랜스포머 모델은 시계열 예측 분야에도 적용되어 순환신경망 기반의 모델들보다 뛰어난 성능을 보이는 것으로 나타난다(Li et al., 2019; Wu, Green, Ben, & O'Banion, 2020). 하지만 트랜스포머 모델을 다변량 시계열 데이터에 적용하는 데는 추가적인 고려가 필요하다. 자연어 처리 분야에서는 입력 시

퀀스의 단일 토큰은 단일 개념(단어)을 표현하고 있지만, 다변량 시계열 데이터의 경우 매 시점의 입력데이터는 개별적인 의미를 가진 변수들로 구성된다. 일반적인 Seq2Seq 모델들은 매 시점의 다변수들을 단일한 토큰으로 취급하기 때문에 시점 간의 시간적(temporal) 어텐션만 학습하고 변수 간에 존재하는 공간적(spatial)¹⁾ 관계는 무시된다(Grigsby et al., 2021). 따라서 다변량 시계열 데이터에 트랜스포머 모델을 적용하기 위해서는 이 두 가지 관계를 모두 모델링할 필요가 있으며, 이에 관한 다양한 연구가 진행되고 있다(K. Chen et al., 2021; Grigsby et al., 2021; Padhi et al., 2021).

이에 본 연구에서는 지하철 역사 내 미세먼지 농도 예측을 위하여, 시공간 트랜스포머를 포함한 4개의 트랜스포머 기반의 모델을 제안한다. ARIMA, LSTM, 어텐션 기반의 Seq2Seq을 비교 모델로 활용하여, 서울교통공사에서 제공한 서울 지하철 역사의 미세먼지 농도 데이터, 기상청에서 제공하는 기온, 습도, 강수량, 풍향, 풍속 등 기상관측 데이터, 한국환경공단에서 제공하는 대기오염도를 이용하여 지하철 서울역, 동대문역사문화공원역, 시청역 대합실의 한 시간 후의 미세먼지 농도 예측실험을 하였다. 미세먼지가 공기 중에서 분산되는 시간지연과 한 시간 단위로 운영되는 환기 제어 시스템의 작동시간을 고려(권순박, 2017; Park et al., 2018)하여, 한 시간 후의 미세먼지 농도를 예측하였다. 본 연구는 기존의 통계와 순환신경망 기반의 모델 활용에 국한되어 있던 지하철 역사의 미세먼지 농도 예측에 트랜스포머 기반의 모델을 적용하여 우수한 성능을 확인한 첫 사례로 지하철 역사의 스마트 공

1) 본 논문에서 “공간”이라는 단어는 물리적 공간이 아닌 다중변수 간 존재하는 관계를 일컫는 의미로 사용하였다.

기질 관리 시스템과 연계되어 지하철 역사를 이용하는 시민들에게 쾌적한 환경을 제공하는 데 활용될 수 있을 것으로 기대한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 이론적 배경 및 선행연구를 소개하고, 3장에서는 제안 트랜스포머 모델의 구조를 소개한다. 4장에서는 논문에서 사용된 데이터와 실험 결과를 보고하고, 5장에서는 연구의 기여와 한계를 정리하고 후속 연구 방향을 제시한다.

2. 이론적 배경

2.1. ARIMA(Auto-Regressive Integrated Moving Average)

ARIMA는 대표적인 시계열 분석법으로 과거의 관측값과 오차를 사용해서 미래의 시계열 값을 예측하는 ARMA(Auto-regressive Moving Average) 모델을 일반화한 모델이다. ARMA 모델은 과거 관측값을 이용하여 미래값을 예측하는 AR(Auto-regressive, 자기 회귀) 모델과 과거 예측 오차를 기반으로 미래값을 예측하는 MA(Moving-average, 이동평균) 모델을 결합한 모델이다. 시계열의 과거 p 개 관측값과 q 개 오차를 이용하여 미래값을 예측하는 ARMA(p, q) 모델은 다음 식(1)으로 나타낼 수 있다.

$$y_t = c + \epsilon_t + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} \quad (1)$$

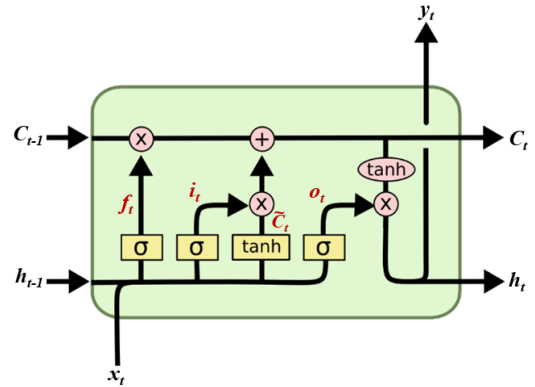
이 식에서 ϕ 는 자기 회귀 모형의 모수, θ 는 이동평균 모형의 모수, c 는 상수, ϵ 는 오차항이다.

ARMA 모델을 적용하기 위해서는 데이터가

안정적이어야 한다. ARIMA 모델은 다소 불안정한 특징을 보여도 적용할 수 있는데, 불안정한 데이터는 차분과정을 거쳐 안정적 시계열 데이터로 변환한다. 변환된 데이터는 자기 회귀와 이동평균 모형의 결합을 통하여 모델링된다. 데이터를 d 회 차분하고, 과거 p 개 관측값과 q 개 오차에 의해 예측되는 모델은 ARIMA(p, d, q)로 표현할 수 있다.

2.2. LSTM(Long Short-Term Memory)

순환신경망이 가지는 긴 입력 시퀀스에 대한 기울기 소실과 폭주 문제를 해결하는 방법으로 LSTM 구조가 제안되었다. LSTM은 입력 게이트(i_t), 망각 게이트(f_t), 출력 게이트(o_t)를 이용하여 정보의 흐름을 조절하여 기울기 소실과 폭주를 방지한다.

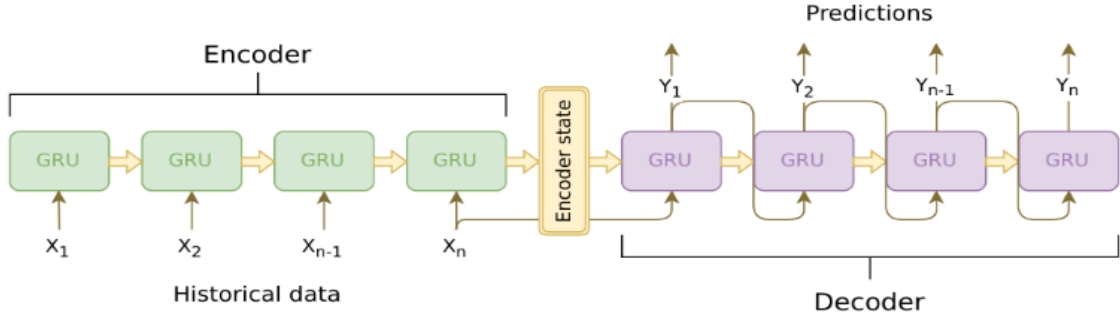


〈그림 1〉 Long Short-Term Memory Network(Olah, 2015)

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (4)$$



〈그림 2〉 Seq2Seq 구조(Suillin, 2018)

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(C_t) \quad (7)$$

망각 게이트는 이전 내부기억 노드 값(C_{t-1})의 정보를 얼마나 버릴 것인지 결정하고, 입력 게이트는 새로운 정보(x_t)를 얼마만큼 내부기억 노드 값(C_t)에 저장할 것인지를 결정한다. 현시점의 내부기억 노드 값(C_t)은 입력값(x_t)과 이전 시점의 은닉 노드 값(h_{t-1})의 선형결합으로 후보 값(\tilde{C}_t)을 구하고 이전 시점의 내부기억 노드 값(C_{t-1})을 더하여 계산한다. 마지막으로 현시점의 내부기억 노드 값을 얼마큼 출력할지가 출력 게이트에 의해 조절된 뒤 은닉 노드 값(h_t)이 출력된다. 위 식들에서 W_f, W_i, W_C, W_o 는 가중치를 포함한 행렬이고 $*$ 는 벡터의 성분별 곱이다.

2.3. Seq2Seq

Seq2Seq 구조는 기계번역과 같이 순차적 데이터를 입력으로 받아 또 다른 순차적 데이터를 출

력하는 모델로 <그림 2>와 같은 구조로 되어 있다. Seq2Seq 모델은 인코더, 문맥 벡터, 디코더의 세 가지 요소로 구성된다. 인코더와 디코더는 여러 개의 순환신경망 셀로 이루어지는데 일반적으로 성능향상을 위해서 LSTM이나 GRU 셀로 구성된다. 인코더의 각 셀은 입력데이터를 순차적으로 입력 받은 후 정보를 처리하여 은닉상태를 전파하고, 마지막 은닉층에서 모든 정보를 압축하여 문맥 벡터를 생성한다. 문맥 벡터에는 디코더의 예측을 돕기 위한 모든 입력데이터의 정보가 압축되어 있다. 디코더는 이 문맥 벡터를 첫 번째 은닉계층으로 입력받는다. 디코더의 각 순환신경망 셀은 이전 시점의 순환신경망 셀에서 생성하는 출력값과 은닉상태를 입력으로 받은 후 처리하여 자신의 은닉상태를 출력한다. 디코더 각 시점의 예측 출력값은 해당 시점의 은닉상태에 소프트맥스, 시그모이드, 선형함수 등을 적용하여 구한다.

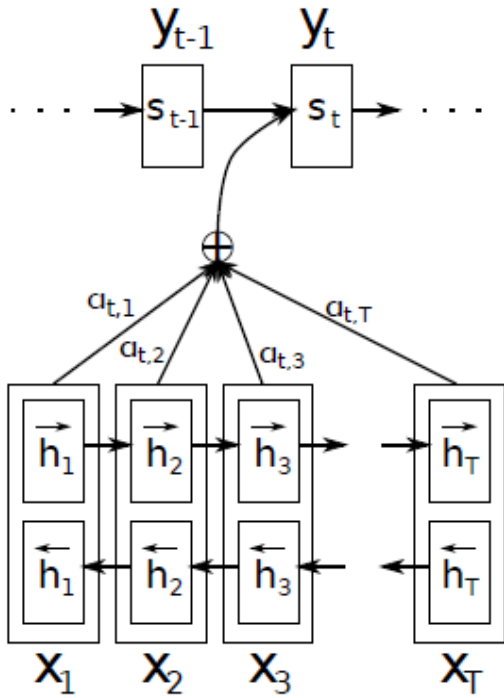
2.4. 어텐션 기법

입력 시퀀스의 길이가 길어지면 성능이 저하되는 장기 의존성 문제를 해결하기 위하여 제안

된 어텐션 기법(Bahdanau et al., 2014)은 디코더에서 출력을 생성할 때 인코더의 전체 입력 시퀀스를 다시 한번 참고하여 해당 시점에서 예측해야 할 출력과 연관이 있는 부분에 집중할 수 있도록 하는 기법이다. <그림 3>은 Bahdanau et al.(2014)이 제안한 어텐션 기법으로 y 는 디코더가 생성한 번역 결과, x 는 입력문장이다.

디코더는 조건부 확률을 학습함으로써 다음 단어를 예측하게 된다.

$$p(y_i | y_1, \dots, y_{i-1}, x) = g(y_{i-1}, s_i, c_i) \quad (8)$$



<그림 3> 어텐션 기법(Bahdanau et al., 2014)

식(8)에서 c_i 는 문맥 벡터이고, s_i 는 디코더의 i 시점에서의 은닉상태로 식(9)에 의해 계산되며,

f 는 일반적으로 LSTM이나 GRU를 활용한다. 문맥 벡터 c_i 는 인코더 은닉상태의 가중합으로 식(10)에 의하여 계산된다.

$$s_i = f(s_{i-1}, y_{i-1}, c_i), \quad (9)$$

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j \quad (10)$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})} \quad (11)$$

$$e_{ij} = a(s_{i-1}, h_j) \quad (12)$$

α_{ij} 는 각 출력이 어떤 입력 상태를 더 많이 참조하는지를 나타내는 어텐션 가중치로 식(11)의 소프트맥스 함수에 의하여 계산된다. 식(12)의 e_{ij} 는 이전 시점의 은닉상태인 s_{i-1} 와 인코더의 j 번째 은닉상태인 h_j 가 얼마나 유사한지를 나타내는 어텐션 점수다. 저자는 어텐션 점수를 다음 식(13)과 같이 계산하였다.

$$a(s_{i-1}, h_j) = v_a^T \tanh(W_a s_{i-1} + U_a h_j) \quad (13)$$

위의 식에서 v_a , W_a , U_a 는 학습 가능한 가중치 행렬이다. 어텐션 점수를 구하는 함수는 두 벡터의 유사도를 표현할 수 있으면 다른 방식이 활용될 수도 있다. Luong, Pham, & Manning(2015)은 일반(general attention) 기법과 내적(dot product) 기법을 제안하였고 Vaswani et al.(2017)은 조정된 내적(scaled dot product) 어텐션을 제안하였다. 임의의 벡터 x, y 가 주어졌을 때 각 기법의 어텐션 점수 함수는 <표 1>과 같다.

〈표 1〉 어텐션 기법들

이름	어텐션 점수 함수
일반(general)	$x^T W_a y$
내적(dot product)	$x^T y$
조정된 내적(scaled dot product)	$x^T y / \sqrt{\dim(x)}$

2.5. 시계열 데이터 예측 선행연구

본 선행연구에서는 공기질 데이터에 적용되는 시계열 데이터 예측과 관련한 연구를 살펴보고자 한다. 인공지능 기술이 인기를 얻기 이전에는 주로 통계적 기법들이 활용되었는데, Goyal, Chan, and Jaiswal(2006)은 다중회귀분석 모델, ARIMA 모델, 다중회귀분석과 ARIMA 혼합 모델을 적용하여 홍콩과 델리의 호흡 가능 부유 미립자(Respirable Suspended Particulate Matter, RSPM)의 일평균 농도를 예측하였다. 이정영 등(2007)은 단변량 ARIMA 모델을 이용하여 서울 지역 평균 오존농도를 예측하였다. Lee, Liu, Kim, Kim, and Yoo(2014)는 평행요인분석(parallel factor analysis)을 통한 공기질 해석 및 예측 시스템을 제안하고 학여울 지하철 역사를 대상으로 실험을 진행하였다. 오종민, 신현수, 신예슬과 정형철(2017)은 서울시 월별 평균 미세먼지 농도를 예측하는데, 회귀분석, ARIMA, 지수평활법 등을 적용하였다.

인공지능 기술의 발전과 더불어 다양한 인공지능 모델들을 활용한 시계열 데이터 예측이 활발히 이루어지고 있으며, 많은 경우 통계 모델들은 성능의 기준이 되는 기준모델로 제시되고 있다. 정철우와 김명석(2013)은 추세와 계절성 등 시계열 데이터의 특성에 따른 예측 모형의 성능을 비교하기 위하여 AR, SARIMA(seasonal

auto-regressive integrated moving average), 인공신경망과 GAM(generalized additive model) 모형을 비교·평가하였다. Park et al.(2017)은 외부 미세먼지 농도, 역사 운행 차량 횟수, 역사 환기량 정보(ventilation ratio)를 입력변수로 투입한 인공신경망 모델을 활용하여 서울시 6개 지하철 역사의 미세먼지 농도 예측연구를 수행하였다. Reddy, Yedavalli, Mohanty, and Nakhat(2018)은 베이징의 다양한 시점의 미세먼지 농도를 예측하는데 LSTM 기반 모델을 활용하여 서포트 벡터 머신(support vector machine) 보다 성능이 우수함을 보였고, Bui, Le, and Cha(2018)는 LSTM 인코더-디코더 구조를 활용하여 서울과 대구지역 미세먼지 농도를 예측하였다. Loy-Benitez, Vilela, Li, and Yoo(2019)는 미세먼지 예측 모델 실험을 위해 한국의 1개 지하철 역사를 선정하여 표준 순환신경망, LSTM, GRU 모델의 성능을 비교하였다.

또한, 다양한 알고리즘을 상호 결합한 연구들도 활발하게 수행되었는데 차진욱과 김장영(2018)은 인공신경망과 k-최근접 이웃 알고리즘을 결합하여 서울지역의 미세먼지 농도를 예측하는 실험을 하여 단일 모형보다 결합 모형의 성능이 더 뛰어남을 보였다. Soh, Chang, and Huang(2018)은 타이완과 베이징의 미세먼지 농도를 예측하기 위하여 LSTM, CNN, ANN을 결합하였는데, 대상 지역의 시계열 데이터 관계를 모델링하기 위해서 LSTM을, 대상 지역 주변의 지형정보와 공기질 변수 간의 정보를 추출하기 위해서 CNN을, 시공간적 유사도가 높은 지역들로부터 예측을 위한 특징들을 추출하기 위해서는 인공신경망을 활용한 통합 모델을 제안하였다. 홍성원(2020)은 시계열 데이터가 짧은 경우 예측력을 향상하는 방법으로 오토인코더를 통하

여 시계열 데이터 크기를 증가시킨 후 순환신경망 기반의 모델에 투입하는 방법을 제안하고 COVID-19 신규 확진자 수를 예측하였다.

트랜스포머 모델이 제안되어 다양한 분야에서 뛰어난 성과를 보인 후로는 시계열 예측 분야에서도 트랜스포머를 활용한 연구가 활발히 진행되고 있다. Li et al.(2019)은 교통량과 전기사용량을 예측하는 연구에서 ARIMA, 지수평활법, 순환신경망 기반 모델, 행렬분해 모델, 트랜스포머 모델을 비교하였고, Wu et al.(2020)은 유사독감 예측에 트랜스포머 모델을 적용하여 ARIMA, LSTM, 어텐션을 적용한 Seq2Seq 모델들보다 트랜스포머 모델의 성능이 우수함을 보였다. 김인경, 김대희와 이재구(2021)는 Temporal Fusion 트랜스포머(Lim, Arik, Loeff, & Pfister, 2019)를 주가와 베이징의 미세먼지 농도 예측에 적용하여 LSTM보다 더 성능이 우수함을 보였다. Temporal Fusion 트랜스포머는 알려진 미래 값(known future inputs), 정적 공변량(static covariates) 값과 시계열 데이터 등 다양한 유형의 입력값과 순환신경망, 셀프 어텐션층, 유의미한 특징 추출을 위한 변수 선택 층과 중요하지 않은 특징 억제층을 위한 개폐(gating) 층을 사용하는 모델이다.

트랜스포머 구조를 다변량 시계열 데이터 분석에 적용하기 위해서는 서로 다른 시점에서 관측된 변수들의 시간적 관계뿐만 아니라 서로 다른 변수 간의 공간적 관계를 고려할 필요가 있다. Grigsby et al.(2021)은 다변량 입력을 평탄화하여 각 입력 토큰들이 시점별 단일 변수값을 표현하도록 긴 시퀀스로 변환하고, 이를 활용하여 각 변수 간의 시간적, 공간적 어텐션을 학습할 수 있는 Spacetimeformer를 제안하였다. K. Chen et al.(2021)은 시간적, 공간적 종속성을 통합적으로

모델링하기 위하여 시간적 어텐션과 공간적 어텐션을 개별적으로 수행한 후 이를 곱셈연산으로 통합한 시공간 어텐션 지도(attention map)를 제안하였다.

모델 측면의 연구 외에도 시간 데이터를 효과적으로 표현하여 성능을 향상하기 위한 기법으로 Kazemi et al.(2019)이 제안한 Time2Vec이 있다. 저자는 Time2Vec 임베딩을 적용 모델과 관계없이 범용적으로 활용할 수 있도록 개발하고 다양한 모델을 대상으로 실험하여 성능이 향상됨을 보였다. 이 Time2Vec을 시계열 데이터의 임베딩으로 적용하는 연구들이 시도되고 있다. Grigsby et al.(2021)은 시계열 데이터 예측을 위한 트랜스포머 구조를 제안하면서 입력 임베딩 파이프라인의 주요 요소로 Time2Vec 기법을 활용하였고, Wass(2021)는 모바일 망 사용량 예측을 위한 모델을 개발하는 과정에서 포지셔널 인코딩(positional encoding) 대신 Time2Vec 임베딩 기법을 활용하여 트랜스포머 모델을 개발하였다.

3. 제안 모델

3.1. 제안 모델 개요

본 장에서는 지하철 역사의 미세먼지 농도 예측을 위해 본 연구에서 제안하는 트랜스포머 모델을 소개한다. 먼저 Vaswani et al.(2017)이 제안한 원조 트랜스포머 모델을 소개하고, 시간 변수를 효과적으로 활용하기 위해 Kazemi et al.(2019)이 제안한 Time2Vec 임베딩을 소개한다. 다음으로는 본 연구에서 제안하는 트랜스포머 모델들인 원조 트랜스포머의 구조를 적용한

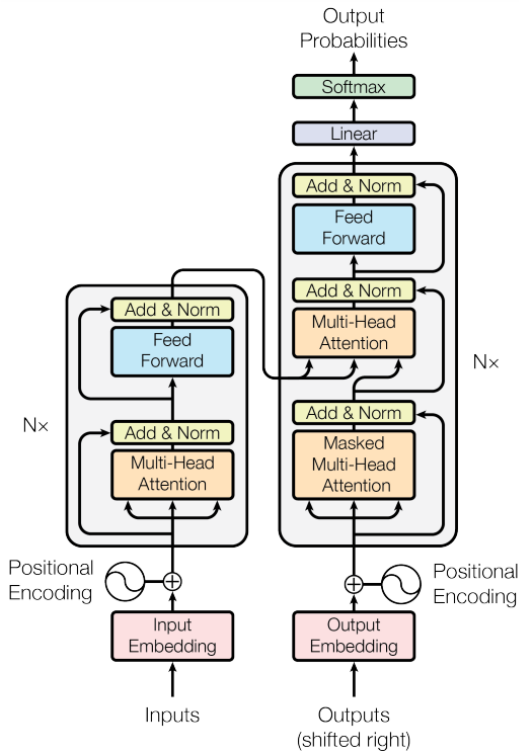
표준 트랜스포머(TFM), 포지셔널 인코딩 대신 Time2Vec을 적용한 표준 트랜스포머(TFM_t2v), 변수간의 시간적 공간적 관계를 모델링한 시공간 트랜스포머(ST-TFM), Time2Vec을 적용한 시공간 트랜스포머(ST-TFM_t2v)를 소개한다.

3.2. 원조 트랜스포머 모델

트랜스포머는 전형적인 인코더-디코더 구조로 <그림 4>와 같이 구성된다. 인코더는 여러 개 (Nx)의 인코더 층을 쌓아 구성하는데 각 인코더 층은 멀티 헤드 셀프 어텐션(multi-head self

attention)과 피드 포워드(feed forward) 두 개의 하부층으로 구성된다. 또한 각 하부층은 학습을 돕기 위한 잔차 연결과 층 정규화를 적용한다. 디코더도 여러 개(Nx)의 같은 층을 쌓아 구성하는데 각 디코더 층은 마스크 된 멀티 헤드 어텐션과 피드 포워드 두 개의 하부층과 인코더 출력과 디코더 입력 간의 어텐션을 학습하기 위한 멀티 헤드 어텐션층으로 구성된다.

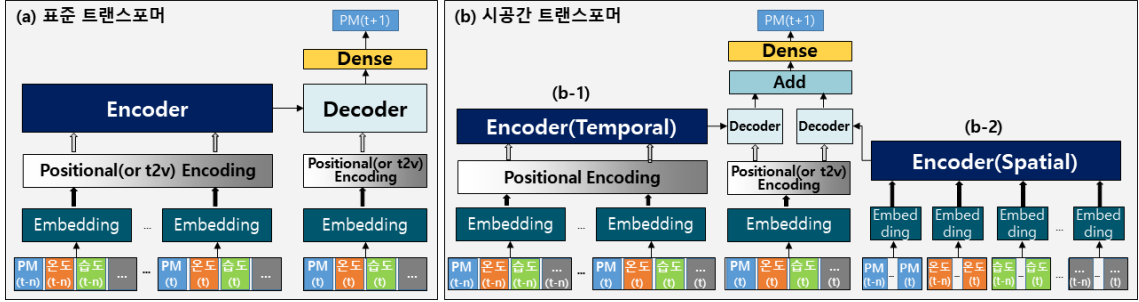
순환신경망은 입력을 순차적으로 받아서 처리하는 특성 때문에 시퀀스 내의 위치 정보를 가질 수 있었지만, 트랜스포머 모델은 입력을 한 번에 입력하기 때문에 입력 순서를 고려하지 못한다. 이를 보완하기 위해 포지셔널 인코딩을 사용하여 입력의 순서에 따라 생성한 위치 벡터값을 임베딩 벡터에 더하여 모델의 입력으로 사용한다. 트랜스포머의 모델의 핵심은 셀프 어텐션 기법으로, 기존의 Seq2Seq 구조와는 달리 어텐션을 수행할 은닉계층이 없는 한계를 극복하기 위하여(Quintana Valenzuela, 2021) 같은 입력 벡터들로부터 얻어진 쿼리(query)와 키(key) 벡터의 유사도를 구하여 키와 대응된 각각의 값(value) 벡터에 가중치를 적용하여 반영하는 방식을 도입하였다. 쿼리, 키, 값 벡터들은 입력 벡터의 선형 변환을 통하여 구해진다. 어텐션 가중치는 조정된 내적과 소프트맥스 정규화 연산을 통하여 구하게 된다. 이 어텐션 가중치를 값 벡터와 가중합하여 어텐션 값을 구하게 되며, 이는 다음의 수식(14)으로 표현할 수 있다.



<그림 4> 트랜스포머 모델 구조(Vaswani et al., 2017)

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (14)$$

Q, K, V는 각각 쿼리, 키, 값 행렬이며, $\sqrt{d_k}$



〈그림 5〉 (a) 표준 트랜스포머와 (b) 시공간 트랜스포머 구조

는 소프트맥스 함수의 기울기 소실 문제를 완화하기 위해 추가하였다. 트랜스포머 모델에서는 어텐션을 멀티 헤드 어텐션 기법을 사용하여 수행하는데, 이는 쿼리, 키, 값 행렬을 여러 개로 나누어 그 개수(헤드 수)만큼 나눠서 병렬적으로 어텐션을 수행한 후 각각의 값을 병합한다. 멀티 헤드 어텐션은 다양한 위치 정보에 집중할 수 있어 입력값들에 대한 다양한 관계를 파악할 수 있으며, 학습의 안정성에 기여하는 것으로 알려졌다(Liu, Liu, & Han, 2021).

3.3 Time 2 Vector 임베딩

순환신경망은 다양한 시퀀스 데이터를 처리하는데 훌륭한 결과를 보이지만 시간 변수를 효과적으로 활용하기 위해서는 해당 문제에 적합하도록 연구자가 피쳐(feature) 설계를 직접 수행하여 모델에 투입하기 때문에 도메인 전문성과 더불어 시간과 노력이 요구된다. Kazemi et al.(2019)은 이런 노력 없이 시간을 벡터로 표현할 수 있는 학습 가능한 Time2Vec 임베딩을 개발하였는데, 이는 모델과 관계없이 범용적으로 활용할 수 있는 기법이다.

Time2Vec 임베딩은 시간 데이터가 가지는 주기적(예: 계절에 따라 나타나는 기후의 주기적

패턴), 비주기적(예: 나이가 들면서 나타나는 질병) 패턴을 모두 모델링한다. 시간 스칼라값 τ 는 다음 수식(15)에 따라 $t2v(\tau)$ 표현으로 변환되는데, 비주기적 패턴은 다음식 (a)의 선형함수에 의하여 모델링 되고, 주기적 패턴은 (b)에 의하여 모델링 된다.

$$t2v(\tau)[i] = \begin{cases} \omega_i \tau + \phi_i, & \text{if } i = 0 \\ F(\omega_i \tau + \phi_i), & \text{if } 1 \leq i \leq k \end{cases} \quad (a) \quad (b)$$

(15)

$t2v(\tau)[i]$ 는 $t2v(\tau)$ 의 i 번째 원소이며, ω_i 와 ϕ_i 는 학습 가능한 모수이고, k 는 $t2v$ 의 차원이다. F 는 활성화 함수인데 저자는 다양한 활성화 함수를 실험하여 성능이 가장 좋은 사인함수를 선택하였다.

3.4. 제안 트랜스포머 모델 구조

본 논문에서 제안하는 트랜스포머 모델은 포지셔널 인코딩 적용 표준 트랜스포머(TFM), Time2Vec 적용 표준 트랜스포머(TFM_t2v), 포지셔널 인코딩 적용 시공간 트랜스포머(ST-TFM), Time2Vec 적용 시공간 트랜스포머(ST-TFM_t2v)의 4가지 모델이다. 표준 트랜스포머는 <그림 5>

의 (a)처럼 원조 트랜스포머의 구조를 따른다. 입력 시퀀스의 시간 순서를 고려하기 위한 방식을 달리하여, 포지셔널 인코딩 적용 표준 트랜스포머(TFM)와 Time2Vec 적용 표준 트랜스포머(TFM_t2v)로 구분하였다. 시공간 트랜스포머는 시간적 어텐션 외 변수 간에 존재하는 공간적 관계를 추가 모델링하기 위해서 K. Chen et al.(2018)의 제안을 참고하여 <그림 5>의 (b-2)처럼 입력 시퀀스의 시간 차원과 변수차원을 전치하여 입력 시퀀스를 구성 후 어텐션 기법을 적용하였다. <그림 5>의 (b-1)과 (b-2)의 인코더의 출력은 디코더 층으로 전달되고, 각 디코더의 출력값을 합한 후 Dense 층을 통하여 최종 예측값을 산출하였다. 시간적 관계를 모델링하는 경우에는 입력 시퀀스의 시간 순서를 고려하기 위해서 포지셔널 인코딩이나 Time2Vec을 적용했지만,

공간적 관계를 모델링하는 경우에는 입력 시퀀스가 시간의 순서와 무관하게 변경되기 때문에 포지셔널 인코딩이나 Time2Vec을 적용하지 않았다. 시공간 트랜스포머도, 포지셔널 인코딩 적용 시공간 트랜스포머(ST-TFM)와 Time2Vec 적용 시공간 트랜스포머(ST-TFM_t2v)를 구분하여 설계하였다.

4. 실험

4.1. 데이터

미세먼지 농도 예측실험 대상 지하철 역사는 서울역, 동대문역사문화공원역, 시청역 3개 역의 대합실이다. 예측 모델 활용 변수 중 PM10(미세

<표 2> 활용 데이터 목록

출처	변수	측정단위	데이터 설명
서울교통공사	미세먼지(PM10)	$\mu\text{g}/\text{m}^3$	• 동대문역사문화공원역, 서울역, 시청역 대합실의 공기질 측정 데이터 중 미세먼지 농도 • 5분 단위로 측정되는 데이터를 1시간 기준으로 평균
	열차 운행 횟수	회	• 측정 시간대의 해당 역 열차 운행 횟수
	승차 인원수	명	• 측정 시간대의 해당 역 이용객 승차 인원수를 합산
	하차 인원수	명	• 측정 시간대의 해당 역 이용객 하차 인원수를 합산
한국환경공단	대기오염도(PM10)	$\mu\text{g}/\text{m}^3$	• 측정 시간대의 역사 외부 미세먼지 농도(1시간 단위 제공)
기상청	기온	$^{\circ}\text{C}$	• 측정 시간대의 역사 외부 기온(1시간 단위 제공)
	상대습도	%	• 측정 시간대의 역사 외부 상대습도(1시간 단위 제공) • 상대습도 = (현재의 수증기량/현재 온도에서의 포화수증기량) x 100
	강수량	mm	• 측정 시간대의 역사 외부 강수량(1시간 단위 제공)
	풍향	각도($^{\circ}$)	• 측정 시간대의 역사 외부 풍향(1시간 단위 제공) • 북쪽을 0° 로 하고 시계방향으로 360° 의 각도로 나누어 풍향을 제공
	풍속	ms	• 측정 시간대의 역사 외부 풍속(1시간 단위 제공)
기타	월	-	• 미세먼지 측정 월(1~12)
	시간	-	• 미세먼지 측정 시간(1~24)
	요일	-	• 미세먼지 측정 요일(1~7, Label Encoding 변환)

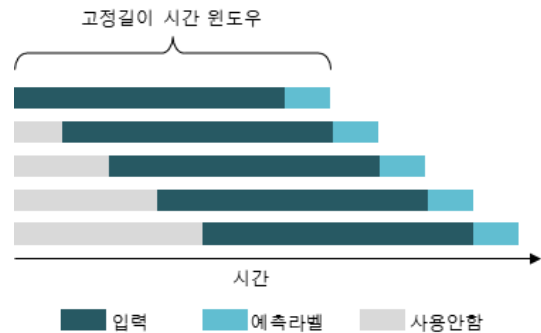
먼지), 열차 운행 횟수, 이용객 승하차 수 데이터는 서울교통공사에서, 역사 외부의 대기오염도는 한국환경공단, 기상관측 데이터는 기상청에서 관리하는 데이터를 수집하였다. 데이터는 2016년 1월부터 2018년 6월까지 3년 6개월간의 데이터를 수집하였으나, 미세먼지 측정 시스템 교체, 전원단절, 기기 점검 등에 의한 결측치를 고려, 데이터의 연속성 확보를 위하여 2016년 1월부터 12월의 데이터를 선별하였다. 데이터는 매시간별 데이터로 구성되어 있다. 기간 중 미세먼지(PM10)의 평균은 43.67, 표준편차는 23.16, 최솟값은 1.5, 최댓값은 394.32로 나타난다. 활용 데이터의 상세 목록은 <표 2>와 같다.

지도학습을 위한 데이터 세트를 구성하기 위하여, <그림 6>과 같은 고정길이 슬라이딩 윈도우(sliding window) 방식을 활용하여 입력과 출력 라벨을 구성하였는데, 과거 24시간의 데이터를 활용하여 1시간 후의 값을 예측하도록 하였다. 슬라이딩 윈도우 방식을 적용하기 전에 훈련데이터 변수별 최솟값과 최댓값을 활용하여 모든 데이터에 대한 최소최댓값 조정(min-max scaling)을 수행하였다. 훈련데이터는 2016년 1월부터 9월, 테스트데이터는 10월부터 12월까지의 데이터로 구성하였다.

4.2. 실험환경 구성

ARIMA 모델을 제외한 딥러닝 모델들은 훈련을 위한 미니 배치크기는 64로 설정하였으며, 학습 횟수는 조기 종료료를 적용하여 50 에포크(epoch)만큼 수행하였다. 또, 딥러닝 모형의 초기화 값 무작위성으로 인하여 동일 모델과 하이퍼파라미터를 사용해도 결과 성능의 차이가 발생하기 때문에, 모델별로 30회 반복 수행(1회당 50

에포크)하고 그 결과를 평균하였다. 실험환경은 Xeon 10 Core x 2, 5 GPU, SSD 960GB x 4, 64GB x 8 RAM 사양의 하드웨어에 Python 3.8과 딥러닝 프레임워크인 Tensorflow 2.7.0을 기반으로 구축하였다.



<그림 6> 고정길이 슬라이딩 윈도우 방식

4.2.1. ARIMA

ARIMA 모델은 미세먼지 데이터 외에도 추가적인 외생변수인 대기오염도, 기상관측데이터 등을 활용하기 위하여 다변량 시계열 데이터 분석 모델인 ARIMAX(Auto-regressive Integrated Moving Average with Exogenous Variables)를 활용하였다. 모델의 p(자기 회귀 차수), d(차분 차수), q(이동평균 차수)는 파이썬 pmdarima(Smith et al., 2017) 패키지를 활용하여 AIC(Akaike information criterion) 값을 최소화하는 단계적(stepwise) 탐색을 통하여 결정하였다. 최종적으로 서울역은 ARIMAX(3, 0, 0), 시청역은 ARIMAX(3, 0, 3), 동대문역사문화공원역은 ARIMAX(2, 0, 0)이 적용되었다.

〈표 3〉 모델별 성능평가 결과

모델	RMSE				MAE			
	전체 평균	서울역	동대문역사 문화공원역	시청역	전체 평균	서울역	동대문역사 문화공원역	시청역
ARIMA	14.7992	12.2264	14.0617	17.6022	11.1120	9.8514	10.1860	13.2987
LSTM	10.1044	7.6040	9.1238	12.8798	7.1229	5.5075	6.2021	9.6979
Seq2Seq	9.8536	7.5303	9.0261	12.3916	6.9809	5.4942	6.1698	9.3114
TFM_t2v	9.6489	7.3815	8.5939	12.2971	6.6044	5.2217	5.7192	8.9023
TFM	9.5473	7.5025	8.7472	11.8817	6.5185	5.2385	5.6810	8.6583
ST-TFM	9.3607	7.2619	8.6543	11.6481	6.3553	5.0509	5.5928	8.4448
ST-TFM_t2v	9.3591	6.8910	8.3071	12.1229	6.2640	4.7533	5.3468	8.7194

4.2.2. LSTM

LSTM 모델은 2개의 LSTM 층으로 구성되어 있고 마지막 층은 Dense 층을 연결하였다. LSTM 층은 24개 시점의 순차적 데이터를 입력으로 받아 처리하고, Dense 층은 두 번째 LSTM 층의 출력을 입력으로 받아서 한 시간 후의 미세먼지 농도 예측값을 출력하는 역할을 한다. LSTM 각 층은 각각 64, 32개의 유닛(units)을 사용하였으며, 규제를 위한 드롭아웃(dropout) 비율은 0.1을 적용하였다. 손실함수는 MSE(Mean Squared Error), 옵티마이저는 Adam을, 학습률은 0.01을 적용하였다.

4.2.3 Seq2Seq

Seq2Seq 모델은 인코더와 디코더로 구성되어 있다. 인코더는 2개의 LSTM 층으로 구성하여 24개 시점의 순차적 데이터를 입력으로 받아 부호화한 출력값과 마지막 시점의 은닉상태를 반환한다. 디코더도 인코더와 같이 2개의 LSTM 층으로 구성하였다. 인코더의 출력값을 입력으로 받고, 각 인코더 LSTM 층 마지막 시점의 은

닉상태가 디코더 LSTM 층의 첫 번째 은닉상태에 사용된다. Seq2Seq 모델에는 바다나우 어텐션 기법을 적용하여 디코더가 예측 시 인코더의 출력 시퀀스를 다시 한번 참조하도록 하였다. LSTM 층은 64개의 유닛을 사용하였으며, 규제를 위한 드롭아웃 비율은 0.1을 적용하였다. 손실함수는 MSE, 옵티마이저는 Adam을, 학습률은 0.01을 적용하였다.

4.2.4. 트랜스포머

트랜스포머 모델은 각 인코더와 디코더에서의 입력과 출력의 크기는 128, 피드 포워드 신경망 은닉층의 크기는 256, 병렬 어텐션을 수행하기 위한 헤드의 개수는 8, 인코더와 디코더 층의 개수는 1을 적용하였다. 규제를 위한 드롭아웃 비율은 0.1을, 손실함수는 MSE, 옵티마이저는 Adam을, 학습률은 학습률 스케줄러에 의해 조정되도록 하였다.

4.3 성능지표

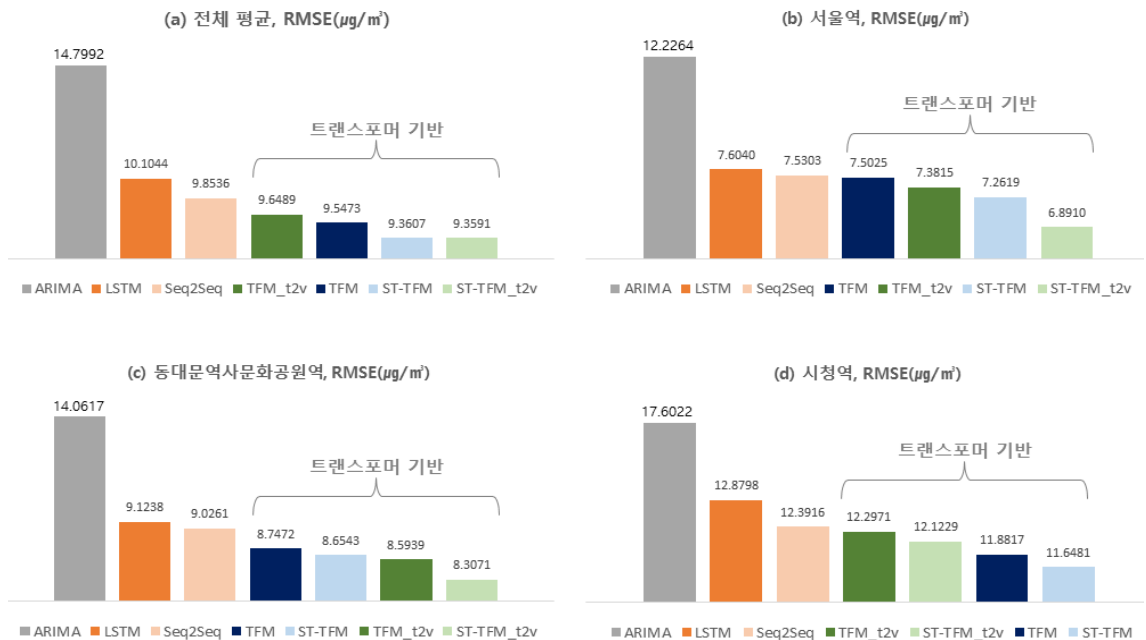
〈표 4〉 모델 간 성능(RMSE) 차이에 대한 통계적 검정 결과(p-value)

Paired t-test	ARIMA	LSTM	Seq2Seq	TFM_t2v	TFM	ST-TFM	ST-TFM_t2v
ARIMA		2.862e-16	2.862e-16	2.862e-16	2.862e-16	2.862e-16	2.862e-16
LSTM			0.03404	0.002586	5.027e-05	7.692e-06	2.862e-06
Seq2Seq				<u>0.06761</u>	0.005361	0.0001142	1.377e-05
TFM_t2v					<u>0.2085</u>	0.005244	0.005083
TFM						0.0446	0.02293
ST-TFM							<u>0.4924</u>
ST-TFM_t2v							

모델에 대한 성능을 평가하기 위한 성능지표는 RMSE(Root Mean Squared Error)와 MAE(Mean Absolute Error)를 사용하였다. RMSE와 MAE를 구하는 식은 각각 (16)과 (17)과 같으며, \hat{y}_i 와 y_i 는 각각 i번째 시점의 예측값과 실제값이고, n 은

측정값의 개수이다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (16)$$



〈그림 7〉 모델별 성능 비교

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (17)$$

4.4. 실험 결과

4.4.1. 전체 성능 비교

7개의 모델에 대한 RMSE와 MAE 성능 결과를 <표 3>에 수록하였고, <그림 7>의 그래프를 통하여 RMSE 기준으로 성능을 비교하였다. RMSE와 MAE 두 지표 모두 역사 전체 평균(a)을 보면 트랜스포머 기반 모델들의 성능이 가장 뛰어나고, 다음으로 Seq2Seq, LSTM, ARIMA 순으로 나타났다. 이는 서울역(b), 동대문역사문화공원역(c), 시청역(d) 개별 역에서도 모두 같은 결과를 보여준다. 트랜스포머 기반 모델들 간에는 3개 역사 전체평균에서는 시공간 트랜스포머

(ST-TFM, ST-TFM_t2v)들이 표준 트랜스포머(TFM, TFM_t2v)들을 앞서고 있으나, 일부 역에서는 임베딩 방식에 따라서 표준 트랜스포머가 시공간 트랜스포머보다 나은 성능을 보이기도 한다. 예를 들어, 시청역(d)의 경우에는 표준 트랜스포머(TFM)의 성능이 Time2Vec을 적용한 시공간 트랜스포머(ST-TFM_t2v)를 앞서고 있다. 이에 대해서는 4.4.3항에서 다시 살펴보도록 하겠다.

<표 4>는 역사 전체평균의 각 모델간 RMSE 차이에 대한 대응 표본 t-test 결과이다. Seq2Seq 모델과 TFM_t2v(p-value=0.06761)를 제외하면 본 논문에서 제안하는 트랜스포머 모델들의 성능이 비교모델인 ARIMA, LSTM, Seq2Seq 보다 유의수준 0.05 하에서 통계적으로 유의한 차이가 있다. 또한 시공간 트랜스포머 모델(ST-TFM, ST-TFM_t2v) 성능과 표준 트랜스포머 모델



<그림 8> 트랜스포머 기반 모델별 성능 비교

(TFM, TFM_t2v) 성능 간에도 통계적으로 유의한 차이가 있는 것으로 나타났다. 다만, 같은 구조의 트랜스포머 모델인 경우, 포지셔널 인코딩과 Time2Vec 임베딩 적용 모델인 TFM과 TFM_t2 간, ST-TFM과 ST-TFM_t2v 간의 성능은 통계적으로 유의한 차이를 보이지 않았다.

4.4.2. 트랜스포머 기반 모델들의 성능 비교

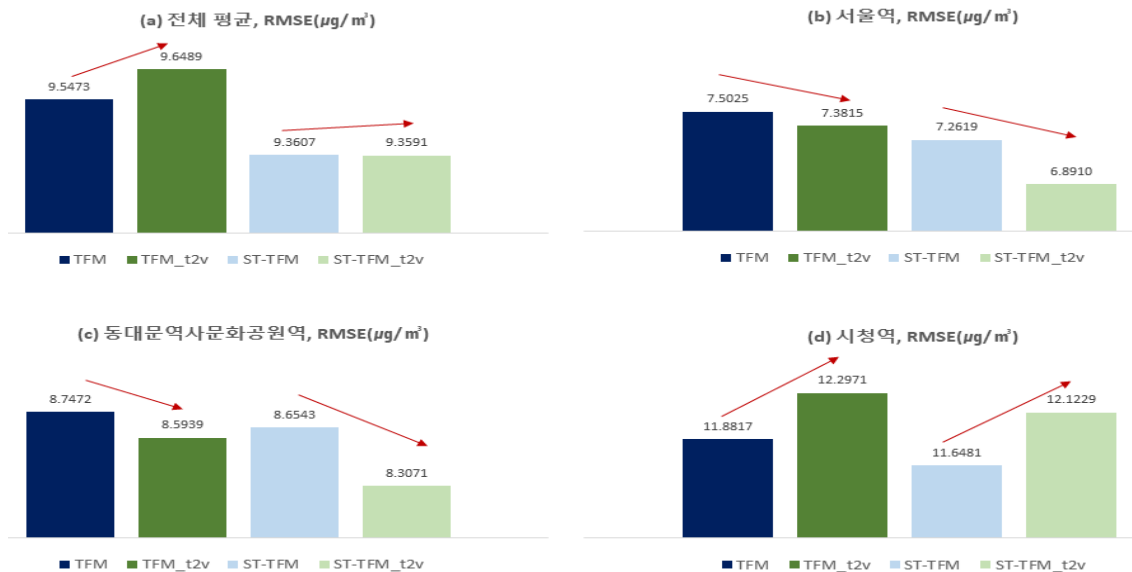
<그림 8>은 트랜스포머 기반 모델들의 성능을 비교한 것으로, 같은 임베딩을 적용하였을 경우, 3개 역사의 전체 평균(a)과 서울역(b), 동대문역사문화공원역(c), 시청역(d) 3개 역사 모두 시공간 트랜스포머 모델의 성능이 더 우수하게 나타났다. 포지셔널 인코딩을 적용한 모델인 TFM과 ST-TFM의 성능을 비교해 보면 모든 사례에서 ST-TFM의 성능이 우수하게 나타남을 확인할 수 있었다. Time2Vec을 적용한 모델인 TFM_t2v와

ST-TFM_t2v의 성능도 마찬가지로 ST-TFM_t2v의 성능이 모든 사례에서 우수하게 나타났다.

이러한 실험 결과는 본 연구에서 제안하는 시공간 트랜스포머가 변수 간에 존재하는 시간적, 공간적 관계를 잘 학습하는 것으로 해석할 수 있다. 매 시점에 모델에 투입되는 데이터(PM10, 외부 대기오염도, 온도, 습도 등)들은 개별적인 의미가 있는 변수로 구성되어 있는데, 표준 트랜스포머는 하나의 입력 시퀀스를 단일 개념으로 취급하기 때문에 변수 간 관계를 학습하는 데 한계가 있다. 이런 한계를 극복하기 위하여, 시공간 트랜스포머의 경우에는 <그림 5>의 (b-2)처럼 입력 시퀀스를 변환하여 모델에 투입함으로써 변수 간의 공간적 관계를 추가로 학습한다.

4.4.3. 포지셔널 인코딩과 Time2Vec 성능 비교

<그림 9>는 포지셔널 인코딩과 Time2Vec의



<그림 9> 포지셔널 인코딩과 Time2Vec 성능 비교

성능을 비교한 것으로 3개 역사 평균(a)과 시청역(d)을 보면 포지셔널 인코딩(TFM, ST-TFM) 적용 시 Time2Vec 적용 시(TFM_t2v, ST-TFM_t2v) 보다 성능이 우수한 것으로 나타났다. 하지만 서울역(b)과 동대문역사문화공원역(c)의 경우에는 Time2Vec 적용 시 성능이 더 우수한 것으로 나타났다. 포지셔널 인코딩과 Time2Vec 모두 입력 시퀀스의 상대적 위치를 반영하는 데 있어 일관성 있는 성능의 우위를 보이지는 않았으며, 적용 데이터의 특성에 따라 성능이 달라지는 것으로 보인다.

5. 결론

본 연구에서는 지하철 역사 내 미세먼지 농도 예측을 위한 트랜스포머 기반 모델을 제안하고, 서울역, 동대문역사문화공원역, 시청역 3개 지하철역 대합실에서 수집된 2016년 공기질 데이터와 지하철 운영자료, 역사 외부 대기오염도, 기상 데이터를 활용한 실증 연구를 진행하였다.

실험은 본 연구에서 제안하는 트랜스포머 기반 모델인 TFM, TFM_t2v, ST-TFM, ST-TFM_t2v와 비교모델인 ARIMA, LSTM, Seq2Seq를 사용하였다. 모델의 성능은 트랜스포머 기반 모델들, Seq2Seq, LSTM, ARIMA 순으로 나타났다. 트랜스포머 모델의 ARIMA, LSTM, Seq2Seq 모델 대비 우수한 성능은 Wu et al.(2020)의 유사 독감 예측실험에서도 동일하게 나타난다. 트랜스포머 기반 모델 중에서는 시공간 트랜스포머 모델이 표준 트랜스포머 모델들보다 높은 성능을 보였다. 이는 변수 간에 존재하는 시공간적 관계를 학습할 수 있도록 제안된 Spacetimeformer(Grigsby et al., 2021)와 Non-Auto regressive Spatial-Temporal Transformer(K. Chen et al.,



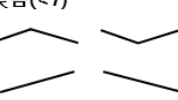
2021) 모두 시간적 관계만을 학습하는 표준 트랜스포머보다 예측 성능이 우수하다는 연구 결과를 보고한 주장을 뒷받침한다.

시계열 분석과 공기질 예측과 관련한 다양한 연구들이 진행되고 있지만, 지하철 역사의 미세먼지 농도 예측과 관련한 연구들은 통계나 순환 신경망 기반의 인공지능 모델 적용에 그치고 있다. 본 연구는 트랜스포머 기반의 모델들을 지하철 역사 미세먼지 농도 예측에 활용하여 기존의 모델들보다 우수한 성과를 확인한 첫 번째 사례라는 점에서 학술적 기여로 인정받을 수 있을 것이다. 환경부와 서울시는 지하철 역사의 미세먼지 저감을 위해서 많은 자원을 투입하고 있다. 미세먼지 농도 예측 모델의 성능이 개선될수록 스마트 공기질 관리 시스템은 더욱더 효과적이고 에너지 효율적으로 운영될 수 있다. 이러한 관점에서, 본 연구는 서울 지하철 역사의 실제 데이터를 사용하여 미세먼지 농도 예측의 성능 향상을 보였다는 점에서 실무적 측면의 기여로 인정받을 수 있을 것이다.

본 연구에서는 서울지역 3개 지하철 역사의 데이터를 사용하여 실험하였다. 따라서 본 논문에서 제안하는 모델의 성능을 수도권, 부산, 대구 광주, 대전 등 모든 지역의 지하철 역사로 확장하여 해석하는데 제약이 있다. 지역별 지하철 역사에 대한 데이터 수집이 가능하다면 본 모델의 지역별 확장 가능성을 더욱 정확하게 확인할 수 있을 것으로 기대한다. 또한, 본 연구에서 제안하는 모델은 한 시간 후의 미세먼지 농도만을 예측하는 단일스텝 모델이라는 한계가 있다. 다양한 시간대의 공기질 농도 예측에 적용하기 위해서는 모델 구조를 수정한 다중 스텝 모델의 개발이 필요하다.

지하철 역사 미세먼지 데이터는 특정 시점에

농도가 급격하게 변경되는 구간이 존재하며 이러한 시점에서 예측 성능이 떨어지는 것으로 나타난다. <그림 10>은 미세먼지 농도의 변동 폭 구간별 예측 모델의 성능을 비교한 것이다. 변동 폭이 ‘매우 높음’ 구간일 때 RMSE는 21.54로 전체평균 RMSE 9.24보다 2.3배 높은 수준으로 성능 저하가 큰 것으로 나타난다. 향후 변동 폭이 높은 지점에 대한 예측 성능 연구가 이루어진다면 모델 전체의 예측 성능을 향상할 수 있을 것으로 기대된다.

변동폭*	비율	구간별 RMSE		
매우 높음(20>) 	8%	21.54	21.54	9.24
중간(7~20) 	30%	9.10	7.38	
낮음(<7) 	62%	5.95		

*산식 : $|(T_n - T_{n-1}) - (T_{n+1} - T_n)|$

<그림 10> 미세먼지 농도 변동폭에 따른 예측 성능 차이

참고문헌(References)

[국내 문헌]

권순박. (2018). 공기청정기 실증 사례 분석-지하역사 스마트 공기질관리시스템 연구 사례. *Air Cleaning Technology*, 31(3), 39-46.

권순박, 강중구, 류승원, 남궁형규, 박세찬, 김민해, 김진호. (2017). 스마트 철도역사의 인공지능 기반 실내공기질 관리기술. *한국철도학회 학술발표대회논문집*, 578-580.

김인경, 김대희, 이재구. (2021). Temporal fusion transformer 모델을 활용한 다층 수평 시계열 데이터 분석. *한국정보처리학회 학술대회논문집*, 28(1), 479-482.

서울특별시청. (2022년, 1월 27일). 지하역사 공기질 개선(2022), Retrieved 2월 26일, 2022년, from <https://yesan.seoul.go.kr/wk/wkSelect.do?itemId=106477>

안성만, 정여진, 이재준, 양지현. (2017). 한국어 음소 단위 lstm 언어모델을 이용한 문장 생성. *지능정보연구*, 23(2), 71-88.

오종민, 신현수, 신예슬, 정형철. (2017). 시계열 분석을 활용한 서울시 미세먼지 예측. *Journal of the Korean Data Analysis Society (JKDAS)*, 19(5), 2457-2468.

이정영, 이종현, 이영재, 김록호, 한진석. (2007). ARIMA 모델을 이용한 서울지역 O₃ 오염도의 시계열 분석. *한국대기환경학회 학술대회논문집*, 1363-1366.

이현욱. (2020). 지하철 실내 공기 질 개선을 위한 철도차량 휠-레일 접촉 미세먼지 발생 연구 소개. *공업화학전망*, 23(4), 20-29.

정철우, 김명석. (2013). Comparison studies of hybrid and non-hybrid forecasting models for seasonal and trend time series data. *지능정보연구*, 19(1), 1-17.

진세중, 조형준. (2020). 머신러닝을 활용한 계절 시계열 예측. *Journal of the Korean Data Analysis Society*, 22(5), 1779-1791.

질병관리청. (2020년, 3월 23일). 미세먼지. Retrieved 2월 15일, 2022년, from <https://www.kdca.go.kr/contents.es?mid=a20304030300>

- 차진욱, 김장영. (2018). 미세먼지 수치 예측 모델 구현을 위한 데이터마이닝 알고리즘 개발. *한국정보통신학회논문지*, 22(4), 595-601.
- 환경부. (2018년, 8월 31일). 제3차 지하역사 공기질 개선대책. Retrieved 2월 15일, 2022년, from http://www.me.go.kr/home/web/policy_data/read.do?menuId=10276&seq=7188
- 홍성원. (2020). 오토인코더 기반 특징추출을 통한 국가별 covid-19 일일 신규 확진자 수 예측 모델링. *한국지능정보시스템학회 학술대회논문집*, 2020(6), 57-58.

[국외 문헌]

- Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv Preprint arXiv:1409.0473*.
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157-166.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.
- Bui, T., Le, V., & Cha, S. (2018). A deep learning approach for forecasting air pollution in south korea using LSTM. *arXiv Preprint arXiv:1804.07891*.
- Chen, K., Chen, G., Xu, D., Zhang, L., Huang, Y., & Knoll, A. (2021). NAST: Non-autoregressive spatial-temporal transformer for time series forecasting. *arXiv Preprint arXiv:2102.05624*.
- Chen, T., Yin, H., Chen, H., Wu, L., Wang, H., Zhou, X., & Li, X. (2018). Tada: Trend alignment with dual-attention multi-task recurrent neural networks for sales prediction. *Paper presented at the 2018 IEEE International Conference on Data Mining (ICDM)*, 49-58.
- Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014a). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv Preprint arXiv:1409.1259*.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014b). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv Preprint arXiv:1406.1078*.
- Goyal, P., Chan, A. T., & Jaiswal, N. (2006). Statistical models for the prediction of respirable suspended particulate matter in urban cities. *Atmospheric Environment*, 40(11), 2068-2077.
- Grigsby, J., Wang, Z., & Qi, Y. (2021). Long-range transformers for dynamic spatiotemporal forecasting. *arXiv Preprint arXiv:2109.12218*.
- Hamid, T. S., & Sodoudi, S. (2016). Statistical modeling approaches for PM10 prediction in urban areas; *A review of 21st-century studies*. *Atmosphere*, 7(2), 15.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
- Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach* Springer. Science & Business Media.
- Kazemi, S. M., Goel, R., Eghbali, S., Ramanan, J., Sahota, J., Thakur, S., . . . Brubaker, M. (2019). Time2Vec: Learning a vector representation of time. *arXiv Preprint arXiv:*

1907.05321

- Lara-Benítez, P., Carranza-García, M., & Riquelme, J. C. (2021). An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems*, 31(3), undefined.
- Lee, S., Liu, H., Kim, M., Kim, J. T., & Yoo, C. (2014). Online monitoring and interpretation of periodic diurnal and seasonal variations of indoor air pollutants in a subway station using parallel factor analysis (PARAFAC). *Energy and Buildings*, 68, 87-98.
- Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y., & Yan, X. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *arXiv Preprint arXiv:1907.00235*.
- Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2019). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748-1764.
- Liu, L., Liu, J., & Han, J. (2021). Multi-head or single-head? an empirical comparison for transformer training. *arXiv Preprint arXiv:2106.09650*.
- Loy-Benitez, J., Vilela, P., Li, Q., & Yoo, C. (2019). Sequential prediction of quantitative health risk assessment for the fine particulate matter in an underground facility using deep recurrent neural networks. *Ecotoxicology and Environmental Safety*, 169, 316-324.
- Luong, M., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *arXiv Preprint arXiv:1508.04025*.
- Martínez-Álvarez, F., Troncoso, A., Asencio-Cortés, G., & Riquelme, J. C. (2015). A survey on data mining techniques applied to electricity-related time series forecasting. *Energies*, 8(11), 13162-13193.
- Olah, C. (2015, August 27). Understanding LSTM Networks, Retrieved February 15, 2022, from <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Padhi, I., Schiff, Y., Melnyk, I., Rigotti, M., Mroueh, Y., Dognin, P., . . . Altman, E. (2021). Tabular transformers for modeling multivariate time series. *Paper presented at the ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3565-3569.
- Park, S., Kim, M., Kim, M., Namgung, H., Kim, K., Cho, K. H., & Kwon, S. (2018). Predicting PM10 concentration in seoul metropolitan subway stations using artificial neural network (ANN). *Journal of Hazardous Materials*, 341, 75-82.
- Quintana Valenzuela, D. (2021). *A study of deep learning techniques for sequence-based problems*. Master in Research in Informatics, Universitat Politècnica de Catalunya, Barcelona.
- Reddy, V., Yedavalli, P., Mohanty, S., & Nakhat, U. (2018). Deep air: Forecasting air pollution in beijing, china. *Environmental Science*, Retrieved 1 November, 2021, from https://www.ischool.berkeley.edu/sites/default/files/sproject_attachments/deep-air-forecasting_final.pdf
- Soh, P., Chang, J., & Huang, J. (2018). Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. *IEEE Access*, 6, 38186-38199.
- Smith, T. G. (2017). Pmdarima: ARIMA estimators

- for Python. 3 January, 2022, from <http://www.alkaline-ml.com/pmdarima>
- Suilin, A. (2018, May 9). How it works. Retrieved February 26, 2022, from https://github.com/Arturus/kaggle-web-traffic/blob/master/how_it_works.md
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30
- Wass, D. (2021). *Transformer learning for traffic prediction in mobile networks*. Degree Project in Computer Science and Engineering, KTH Royal Institute of Technology, Stockholm.
- Wu, N., Green, B., Ben, X., & O'Banion, S. (2020). Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv Preprint arXiv:2001.08317*.

Abstract

Application of spatiotemporal transformer model to improve prediction performance of particulate matter concentration

Youngkwang Kim* · Bokju Kim** · SungMahn Ahn***

It is reported that particulate matter(PM) penetrates the lungs and blood vessels and causes various heart diseases and respiratory diseases such as lung cancer. The subway is a means of transportation used by an average of 10 million people a day, and although it is important to create a clean and comfortable environment, the level of particulate matter pollution is shown to be high. It is because the subways run through an underground tunnel and the particulate matter trapped in the tunnel moves to the underground station due to the train wind. The Ministry of Environment and the Seoul Metropolitan Government are making various efforts to reduce PM concentration by establishing measures to improve air quality at underground stations. The smart air quality management system is a system that manages air quality in advance by collecting air quality data, analyzing and predicting the PM concentration. The prediction model of the PM concentration is an important component of this system. Various studies on time series data prediction are being conducted, but in relation to the PM prediction in subway stations, it is limited to statistical or recurrent neural network-based deep learning model researches. Therefore, in this study, we propose four transformer-based models including spatiotemporal transformers. As a result of performing PM concentration prediction experiments in the waiting rooms of subway stations in Seoul, it was confirmed that the performance of the transformer-based models was superior to that of the existing ARIMA, LSTM, and Seq2Seq models. Among the transformer-based models, the performance of the spatiotemporal transformers was the best. The smart air quality management system operated through data-based prediction becomes more effective and energy efficient as the accuracy of PM prediction improves. The results of this study are expected to contribute to the efficient operation of the smart air quality management system.

* Data Solution Business Department, WesleyQuest Co., Ltd.

** D&A Platform Department, Woori Finance Information System Co., Ltd.

*** Corresponding Author: SungMahn Ahn

School of Business Administration, Kookmin University

77 Jungrungro, Sungbukku, Seoul 02707, Korea

Tel: +82-2-910-4574, FAX: +82-2-910-4079, E-mail: sahn@kookmin.ac.kr

Key Words : Time series data analysis, PM concentration prediction, Attention mechanism, Spatiotemporal transformer, Time to Vector

Received : March 3, 2022 Revised : March 24, 2022 Accepted : March 25, 2022

Corresponding Author : SungMahn Ahn

저 자 소 개



김영광

현재 (주)웨슬리퀘스트 데이터솔루션 사업부에서 상무이사로 재직 중이다. 고려대학교 전산과학과에서 학사를, 뉴욕주립대학원에서 기술경영 석사를 취득 후 국민대학교 일반대학원 데이터사이언스학과에서 데이터사이언스전공 박사과정 중에 있다. 주요 관심분야는 비즈니스 애널리틱스, 시계열 데이터 분석, 추천시스템, AI application 등이다.



김복주

현재 우리에프아이에스(주)에서 AI팀 팀장으로 재직 중이다. 건국대학교 중문학과 에서 학사를 취득하고, 건국대학교에서 컴퓨터공학 석사를, 연세대학교에서 빅데이터 분석 석사 취득 후 국민대학교 일반대학원 데이터사이언스학과에서 데이터사이언스전공 박사과정 중에 있다. 주요 관심분야는 인공지능 자연어 처리, 인공지능 컴퓨터 비전, 강화학습 등이다.



안성만

현재 국민대학교 경영대학 경영학부교수로 재직하고 있다. 서울대학교 경영대학에서 학사를, George Mason University에서 정보기술학 박사학위를 취득하였다. Johns Hopkins University의 응용수학과에서 박사후과정을 보냈고 (주)쌍용정보통신에서 근무하였다. 주요 관심분야는 통계적 방법론, 모형선택, 데이터마이닝, 뉴럴네트워크 등이다.