

# 포털 검색량 데이터 활용 명절 기간 교통 예측 시스템

신효민<sup>1</sup>, 서동혁<sup>2</sup>, 김상욱<sup>3\*</sup>

<sup>1</sup> 한양대학교 인공지능학과 석사과정

<sup>2</sup> 한양대학교 컴퓨터 소프트웨어학과 박사과정

<sup>3</sup> 한양대학교 컴퓨터 소프트웨어학과 교수

mehyom@hanyang.ac.kr, hyuk125@hanyang.ac.kr, wook@hanyang.ac.kr

## A Traffic Prediction System for Holiday Periods Using Portal Search Data

Hyomin Shin<sup>1</sup>, Dong-Hyuk Seo<sup>2</sup>, Sang-Wook Kim<sup>3\*</sup>

<sup>1</sup>Dept. of Artificial Intelligence, Hanyang University

<sup>2</sup>Dept. of Computer Science, Hanyang University

<sup>3</sup>Dept. of Computer Science, Hanyang University

### 요약

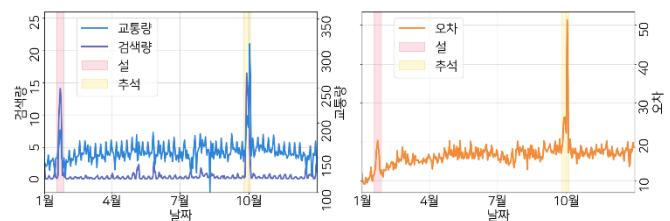
설, 추석 등으로 대표되는 명절 기간에는 평소의 교통 패턴과는 매우 다른 교통 패턴이 관찰된다. 이러한 교통 패턴은 과거 교통량의 패턴을 학습하여 미래 교통량을 예측하는 기준 교통 예측 시스템의 예측을 방해한다. 이처럼 데이터의 한계로 인해, 오로지 과거 교통량 데이터에 의존하는 기준 연구들은 명절 기간에 교통 예측 정확도가 크게 저하되는 문제를 겪고 있다. 이에 본 연구에서는 포털 사이트에서 제공하는 여러 검색어에 대한 포털 검색량 데이터를 추가 정보로 활용하여 정확도를 개선하는 딥 러닝 기반 교통 예측 시스템을 제안한다. 경부고속도로의 8년간 명절 기간 데이터를 대상으로 실시한 실험 결과, 제안 방안은 우리나라 경부고속도로 데이터에 대해 최신 교통 예측 시스템과 비교하여 3%에 가까운 개선된 교통 예측 정확도를 나타냈다. 이를 통해 본 연구는 포털 검색량 데이터를 활용하는 접근법이 명절과 같이 교통 상황이 급변하는 경우에 대한 예측 정확도를 효과적으로 개선할 수 있음을 보였다는 점에서 큰 의미가 있다..

### 1. 서론

교통 예측은 효율적으로 최단 이동 경로를 제공하기 위한 필수적인 기술이자 도시 계획, 물류, 대중교통 운영 등 다양한 분야에서 효율성과 편의성을 제공하기 위한 핵심 기술이다. 이에 따라 교통 예측과 관련한 다양한 연구가 진행되어 왔으며 [1-16], 일반적으로 (1) 도로 구간 간의 공간적 상호작용과 (2) 시간의 흐름에 따른 시간적 상호작용을 모델링하는 것에 초점을 둔다. [1-12]. 이를 위해 최근에는 딥 러닝 기반 접근 방식 [1-10]이 주류를 이루고 있으며, 특히 공간적 상호작용 모델링을 위한 그래프 신경망(Graph Neural Network, 이하 GNN) [17, 18]과 시간적 상호작용 모델링을 위한 순환 신경망(Recurrent Neural Network, 이하 RNN) [19, 20]을 활용한 방법들이 주목받고 있다.

그러나, 이러한 대부분의 연구 결과는 [1-16] 명절 기간의 교통 예측에 대해 평상시보다 상대적으로 낮은 예측 정확도를 보인다. 이는 귀성·귀경, 성묘, 여행 등 다양한 목적의 대규모 인구 이동으로 인해 명절에

는 평소와 전혀 다른 양상의 교통 흐름이 나타나기 때문이다. 특히 인구가 수도권에 집중된 국가의 경우, 단기간 내 수도권과 비수도권 지역 간 양방향 교통량이 급증하여 특정 시간대나 경로에 쓸리는 현상이 발생한다. 이처럼 평소와 다른 이동 양상은 기존 교통 예측 시스템이 전제로 삼는 일반적인 시간적·공간적 패턴을 크게 벗어나며, 결과적으로 높은 예측 오차를 초래한다.



(그림 1) 경부고속도로의 일별 교통량(파랑), 검색어 ‘고속도로 교통 정보’의 일별 검색량(보라), 교통 예측 MAE 오차(주황)의 연간 추세.

이러한 문제점을 보다 명확히 확인하기 위해, 본 연구에서는 한국의 주요 도로 중 하나인 경부고속도로를 대상으로 실제 교통 흐름과 예측 정확도를 비교 분석하였다. 그림 1에는 경부고속도로에서 수집된 실제 교통량의 시간적 변화가 파란색 선으로, 최신 교통 예측 시스템인 MegaCRN [3]이 동일한 기간 동안 도출한 예측 결과의 오차가 주황색 선으로 나타나 있다. 그림에서 확인할 수 있듯, 설(빨간색 박스)과 추석(노란색 박스) 기간에는 실제 교통량이 평상시와는 매우 다른 양상을 보인다. 이로 인해 해당 기간 동안 교통 예측 시스템의 성능이 눈에 띄게 저하되는 현상이 나타나며, 명절이라는 특수 상황이 기존 방법들의 예측 능력을 크게 약화시킨다는 점을 확인할 수 있다.

본 연구는 이러한 예측 정확도 저하 문제를 개선하기 위한 방안으로, 포털에서 수집한 검색량 데이터를 교통 예측의 추가 정보로 활용하는 접근 방식을 제안한다. 명절 기간 이전에는 많은 운전자들이 ‘고속도로 교통 상황’과 같은 검색어를 검색하는 경향이 있으며, 이로 인해 특정 검색어의 검색량이 단기간 내 급격히 증가하는 현상이 나타난다. 그림 1에 제시된 보라색 선은 ‘고속도로 교통 상황’이라는 검색어의 연간 검색량 변화를 나타낸 것으로, 설과 추석 기간을 전후로 검색량이 급증하는 양상을 확인할 수 있다. 그림에서는 보이지 않았지만, 이와 유사하게 ‘주차’, ‘시외버스’ 등 교통 관련 검색어들 또한 명절 기간에 검색량이 크게 증가하는 경향을 보였다. 이러한 정보는 명절 기간 동안의 교통 수요를 간접적으로 반영할 수 있는 유용한 데이터로서 활용될 수 있다.

본 논문은 다음의 두 가지 공헌을 가진다: (1) 검색량 정보를 교통 예측에 활용하는 방법을 제안하고, (2) 국내 경부고속도로의 교통량 데이터를 기반으로 한 실험을 통해 기존 8 개의 최신 교통 예측 시스템들과 비교함으로써, 제안한 방법의 효과를 검증한다.

## 2. 관련 연구

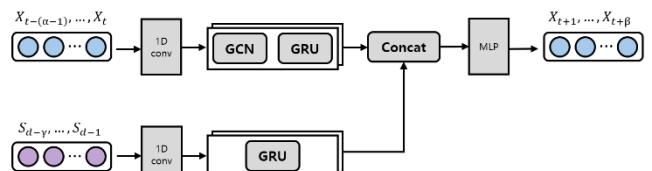
최근 교통 예측 분야에서는 시계열 데이터의 예측 정확도를 향상시키기 위한 다양한 딥 러닝 기반 접근 방식이 활발히 연구되고 있다. 초기에는 ARIMA [11], VAR [21] 과 같은 통계적 모델이 주로 활용되었으나, 복잡한 비선형 패턴을 자동으로 학습할 수 있는 딥 러닝 기법의 등장 이후 RNN 및 GNN 기반 모델이 주요한 연구 방향으로 자리잡고 있다. RNN 계열 모델은 시간 흐름에 따른 패턴을 효과적으로 학습할 수 있는 구조를 갖추고 있으며, 특히 LSTM(Long Short-Term Memory) [20] 과 GRU(Gated Recurrent Unit) [19]와 같은 구조는 교통량과 같은 시계열 데이터의 장기적인 추세와 주기성을 포착하는 데 강점을 가진다. 이러한 특성 덕분에 RNN 기반 모델은 다양한 교통 예측 문제에서 우수한 성능을 보여주고 있다.

한편, 도로 네트워크와 같은 공간적 구조를 반영하기 위해 GNN(Graph Neural Network)이 도입되었으며,

이는 도로 구간 간의 복잡한 공간적 상관관계를 효과적으로 학습할 수 있다. GNN 기반의 대표적인 모델인 Graph Convolutional Network(GCN) [17]는 그래프 구조 내에서 주변 노드의 정보를 합성(convolution)함으로써 도로 간의 공간적 특징을 효과적으로 추출한다. 이와 같은 RNN 및 GNN 기반 모델들은 교통 예측 분야에서 지속적으로 연구가 이루어지고 있다.

## 3. 제안 방안

본 연구는 교통량에 영향을 미칠 가능성성이 높은 포털 검색량 데이터를 추가 정보로 활용하여, 도로 구간 단위의 미래 교통량을 보다 정확하게 예측하는 딥 러닝 기반의 교통 예측 시스템을 제안한다. 기존의 교통 예측 시스템은 과거 교통량 데이터에 주로 의존해왔으며, 이는 외부 환경 변화에 따른 불규칙한 교통 패턴을 반영하는 데에 한계가 있다. 본 연구는 이러한 한계를 보완하기 위해, 사람들의 관심도를 포함하는 포털 검색량 데이터를 추가 정보로 활용하는 방법을 제안한다. 구체적으로는, 교통량 데이터에 대해서 도로 간 영향을 포착하는 GCN [17] 기반의 그래프 학습, GRU [19] 기반의 시계열 학습, 그리고 검색량 데이터에 대한 GRU [19] 기반의 시계열 학습을 통합한 구조를 구성하였다. 전체적인 제안 방안의 오버뷰는 그림 2에 제시되어 있다.



(그림 2) 제안 방안의 오버뷰.

제안 방안의 학습 및 검증을 위해, 설과 추석 같은 대규모 인구 이동이 발생하는 명절 기간에 집중하여 데이터를 수집하였다. 실험에서 5 분 단위의 교통량 데이터는 과거  $\alpha$  개의 시점인  $X_{t-(\alpha-1)}, \dots, X_t$ 을 입력으로 활용하며, 1 일 단위인 검색량 데이터는 날짜  $d$ 를 기준으로 과거  $\gamma$  일간의 검색량  $S_{d-\gamma}, \dots, S_{d-1}$ 을 입력으로 사용한다. 여기서  $d$ 는 시점  $t$ 가 속한 날짜를 의미한다.

이러한 입력 데이터는 다음과 같은 관계를 갖는 예측 구조를 따른다:

$$X_{t-(\alpha-1)}, \dots, X_t, S_{d-\gamma}, \dots, S_{d-1} \Rightarrow X_{t+1}, \dots, X_{t+\beta}$$

여기서  $X_i \in R^N$ 은 시점  $i$ 에서의  $N$  개 도로 구간에 대한 교통량을 의미하며,  $S_j \in R^M$ 은 날짜  $j$ 에서의  $M$  개 검색어에 대한 검색량이다. 출력은 향후  $\beta$  개 시점에 대한 교통량 시계열  $X_{t+1}, \dots, X_{t+\beta}$ 로 구성된다.

제안 방안은 교통량 시계열과 검색량 시계열이라는 서로 다른 두 종류의 정보를 결합하여 향후 도로 구간별 교통량을 예측하는 구조로 구성된다. 교통량

데이터는 도로 간 상호작용과 시간적 흐름을 동시에 반영해야 하므로, 공간 정보와 시간 정보 학습이 결합된 형태로 구성된다. 공간적 상호작용과 시간적 흐름을 동시에 고려할 수 있도록, (1) 먼저 각 도로 구간을 하나의 노드로 간주하여 그래프를 구성하고, 이를 기반으로 GCN [17]을 적용하여 공간적 특징을 학습한다. (2) 이어서, 개별 도로 구간의 시간에 따른 교통량 변화를 효과적으로 포착하기 위해 GRU [19]을 적용하여 시간적 특징을 학습한다. 반면 검색량 데이터는 공간적 관계를 전제하지 않기 때문에, GRU [19]만으로 처리된다. 이를 통해 얻은 검색어 별 임베딩은 교통량 시계열만으로는 포착하기 어려운 명절 기간 교통패턴을 예측하는 데에 도움을 준다.

마지막으로, 도로 구간별로 학습된 교통량 임베딩과 전체 검색어 임베딩을 순차(concatenate)하여 하나의 임베딩으로 만들며, 이를 완전 연결층(Fully Connected Layer)에 입력하여 각 도로 구간의 향후 교통량을 출력한다. 모델은 예측값과 실제 교통량 간의 평균 절대 오차(Mean Absolute Error, MAE)를 최소화하도록 종단 간(end-to-end) 방식으로 학습되며, 그 손실 함수는 다음과 같다:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

제안하는 구조는 단순히 교통량 데이터에만 의존하던 기준 접근과 달리, 사용자 관심도의 변화를 포함하는 검색량 데이터를 함께 활용함으로써 예측의 정확도와 실효성을 높일 수 있다는 점에서 차별성을 갖는다.

## 4. 실험

### 4-1. 데이터 집합

<표 1> 데이터 통계

데이터집합	데이터 수	시점 수	시간 단위
교통량 데이터	86	50688	5 분 단위
검색량 데이터	44	176	1 일 단위

본 연구에서는 설 및 추석과 같은 명절 기간 동안 교통량이 급증하는 현상을 예측하기 위해, 경부고속도로 전 구간의 교통량 데이터와 검색량 데이터를 함께 활용하였다. 교통량 데이터는 고속도로 공공데이터 포털에서 수집하였다. 구체적으로는 서울과 부산을 연결하는 경부고속도로 구간을 대상으로 하며, 2016년부터 2023년까지 총 8년간의 명절(설 및 추석) 기간에 해당하는 16 개의 구간을 수집하였다. 각 명절 절에 대해 명절 당일을 기준으로 앞뒤 5 일씩 총 11 일간의 5 분 단위 교통량 데이터를 확보하였으며, 전체 시계열 길이는 50,688 시점에 달한다. 교통 예측에는 최신 교통 예측 시스템[1-10]과 동일하게 과거 12 step을 입력으로 하여 미래 12 step을 예측하는 방식으로 설정하였다.

검색량 데이터는 대한민국에서 가장 많이 사용되는 검색 포털인 네이버(Naver)의 데이터 랩 검색어 트렌

드에서 수집하였으며, 다음의 5 가지 주제에 따라 총 44 개의 검색어를 선정하였다:

- 연휴 관련: 설, 추석, 명절, 연휴, 귀성, 귀경
- 고속도로 정보 관련: 경부고속도로, 고속도로 교통 정보, 고속도로 교통 상황, 교통 체증, 도로 공사, 사고, 실시간 교통 정보, 단속
- 여행 관련: 동물원, 예매, 한국민속촌, 어린이 대공원, 여행, 나들이, 졸음쉼터, 주유소, 주차, 맛집, 카페, 휴게소 등
- 대중교통 관련: 고속버스, 시외버스, 버스, 터미널
- 지역 관련: 서울, 부산

일반적으로 포털 검색은 실제 운전보다 이전에 이루어지기 때문에, 검색량 데이터는 교통량보다 더 먼 과거를 볼 필요가 있다. 우리는 검색량 데이터를 얼마나 많이 보아야 하는지를 실험을 통해 검증하였으며 (4.2 절 참조), 이에 따라 검색량 입력 범위  $\gamma$  는 3(3 일)으로 설정하였다. 검색량 데이터는 1 일 단위로 수집되어 총 176 시점으로 구성된다.

### 4-2. 실험 결과

<표 2> 제안 방안과 최신 교통 예측 방법들의 성능 비교

	MAE	MAPE	RMSE
STGCN [10]	31.1673	0.6398	47.3093
DCRNN [7]	14.9635	0.142	23.1079
Graph WaveNet [9]	<b>14.5716</b>	0.1312	22.6512
AGCRN [1]	15.2	0.1544	23.97
MTGNN [2]	14.3144	<b>0.1419</b>	22.6364
GTS [8]	15.0504	0.138	<b>22.6157</b>
MegaCRN [3]	16.0002	0.1425	25.1757
HimNet [6]	18.6134	0.2145	38.9832
제안 방안	<b>14.2924</b>	<b>0.1278</b>	<b>22.5584</b>
개선	0.15%	2.59%	0.25%

<표 3> 검색량 데이터 사용 기간( $\gamma$ )에 따른 성능 비교

사용 기간 ( $\gamma$ )	MAE	MAPE	RMSE
1 일 전까지	14.3152	0.1285	22.602
2 일 전까지	14.3088	0.1292	22.5943
3 일 전까지	<b>12.2924</b>	<b>0.1278</b>	<b>22.5584</b>
4 일 전까지	<b>14.2951</b>	<b>0.1280</b>	22.5710
5 일 전까지	14.3047	0.1283	<b>22.5635</b>

우리는 제안하는 방법의 효과를 검증하기 위해 다양한 최신 교통 예측 시스템들과 비교 실험을 수행하였다. <표 2>는 기존 방법들과 제안 방안의 예측 성능을 MAE, MAPE, RMSE 세 가지 지표로 비교한 결과이다. 실험 결과, 제안 방안은 모든 지표에 대해 가장 우수한 성능을 보였다. 이는 검색량 데이터의 활용이 교통 예측 정확도 개선에 도움이 됨을 보여준다.

또한, <표 3>은 교통 예측 목표 시점을 기준으로 과거 검색량 데이터를 활용한 기간( $\gamma$  일)에 따른 교통 예측 정확도를 비교한 결과이다. 실험 결과,  $\gamma=1$  부터  $\gamma=3$  까지는 그 정확도가 점차 증가하였으나,  $\gamma=4$  부터는 다시 정확도가 감소하는 경향을 보였다. 이는 적절한 검색량 데이터 활용 기간을 설정하는 것이 중요하다는 점을 보여준다. 이러한 실험 결과는 교통 예

측 시스템에 검색량과 같은 추가 정보를 결합하는 것 이 실질적인 성능 개선으로 이어질 수 있음을 효과적으로 보여준다.

## 5. 결론

본 연구에서는 명절과 같이 교통 패턴이 급격히 변화하는 상황에서, 기존 교통량 데이터만으로는 미래의 교통량을 예측하기 어려운 문제를 개선하기 위해 포털 검색량 데이터를 추가 정보로 활용하는 교통 예측 시스템을 제안하였다. 제안 방안은 기존 교통 예측 시스템들에 비해 높은 교통 예측 성능을 보였으며, 이는 실제 교통량 변화를 파악하는 데에 검색량 데이터가 유용한 정보를 제공함을 시사한다. 이와 같이 본 연구는 교통 예측에 외부 데이터를 결합함으로써, 명절과 같은 특수 상황에서도 예측 정확도를 개선할 수 있음을 입증하며, 향후 도시 교통 관리 및 관련 응용 분야에서의 활용 가능성을 제시한다.

## 사사

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원과 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.RS-2022-00155586, 실세계의 다양한 다운스트림 태스크를 위한 고성능 빅 하이퍼그래프 마이닝 플랫폼 개발(SW 스타랩), No.RS-2020-II201373, 인공지능대학원지원(한양대학교))

## 참고문헌

- [1] BAI, Lei, et al., Adaptive graph convolutional recurrent network for traffic forecasting, *Advances in Neural Information Processing Systems*, 33, 17804-17815, 2020.
- [2] GAO, Jianliang, et al., MTGNN: multi-Task Graph Neural Network based few-shot learning for disease similarity measurement, *Methods*, 198, 88-95, 2022.
- [3] JIANG, Renhe, et al., Spatio-temporal meta-graph learning for traffic forecasting, *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, 8078-8086.
- [4] SEO, Dong-hyuk, et al., Empowering Traffic Speed Prediction with Auxiliary Feature-Aided Dependency Learning, *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024, 4031-4035.
- [5] SONG, Junho, et al., St-gat: A spatio-temporal graph attention network for accurate traffic speed prediction, *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, 4500-4504.
- [6] DONG, Zheng, et al., Heterogeneity-informed meta-parameter learning for spatiotemporal time series forecasting, *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, 631-641.
- [7] LI, Yaguang, et al., Diffusion convolutional recurrent neural network: Data-driven traffic forecasting, *arXiv preprint arXiv:1707.01926*, 2017.
- [8] SHANG, Chao; CHEN, Jie; BI, Jinbo, Discrete graph structure learning for forecasting multiple time series, *arXiv preprint arXiv:2101.06861*, 2021.
- [9] WU, Zonghan, et al., Graph wavenet for deep spatial-temporal graph modeling, *arXiv preprint arXiv:1906.00121*, 2019.
- [10] HAN, Haoyu, et al., STGCN: a spatial-temporal aware graph learning method for POI recommendation, *2020 IEEE International Conference on Data Mining (ICDM)*, 2020, 1052-1057.
- [11] KUMAR, S. Vasantha; VANAJAKSHI, Lelitha, Short-term traffic flow prediction using seasonal ARIMA model with limited input data, *European Transport Research Review*, 7, 1-9, 2015.
- [12] STATHOPOULOS, Anthony; KARLAFTIS, Matthew G., A multivariate state space approach for urban traffic flow modeling and prediction, *Transportation Research Part C: Emerging Technologies*, 11, 2, 121-135, 2003.
- [13] CUI, Yue; XIE, Jiandong; ZHENG, Kai, Historical inertia: A neglected but powerful baseline for long sequence time-series forecasting, *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, 2965-2969.
- [14] KIM, Namhyuk, et al., Context-aware Traffic Flow Forecasting in New Roads, *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, 4133-4137.
- [15] KIM, Namhyuk, et al., APOTS: A model for adversarial prediction of traffic speed, *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 2022, 3353-3359.
- [16] SHAO, Zehui, et al., Exploring progress in multivariate time series forecasting: Comprehensive benchmarking and heterogeneity analysis, *IEEE Transactions on Knowledge and Data Engineering*, 2024.
- [17] Kipf, Thomas N., Welling, Max, Semi-supervised classification with graph convolutional networks, *arXiv preprint arXiv:1609.02907*, 2016.
- [18] Veličković, Petar, Graph attention networks, *arXiv preprint arXiv:1710.10903*, 2017.
- [19] CHO, Kyunghyun, et al., Learning phrase representations using RNN encoder-decoder for statistical machine translation, *arXiv preprint arXiv:1406.1078*, 2014.
- [20] HOCHREITER, Sepp; SCHMIDHUBER, Jürgen, Long short-term memory, *Neural Computation*, 9, 8, 1735-1780, 1997.
- [21] SIMS, Christopher A., Macroeconomics and Reality, *Econometrica*, 48, 1, 1-48, 1980.