

지능형 교통 시스템을 위한 Graph Neural Networks 기반 교통 속도 예측

Traffic Speed Prediction Based on Graph Neural Networks for Intelligent Transportation System

김 성 훈* · 박 종 혁** · 최 예 림***

* 주저자 : 서울여자대학교 데이터사이언스학과 석사과정

** 공저자 : 서울대학교 산업경영공학과 박사수료

*** 교신저자 : 서울여자대학교 데이터사이언스학과 교수

Sunghoon Kim* · Jonghyuk Park** · Yerim Choi***

* Department of Data Science, Seoul Women's University

** Department of Industrial Engineering, Seoul National University

*** Department of Data Science, Seoul Women's University

† Corresponding author : Yerim Choi, yerim.choi@swu.ac.kr

Vol.20 No.1(2021)

February, 2021

pp.70~85

pISSN 1738-0774

eISSN 2384-1729

<https://doi.org/10.12815/kits.2021.20.1.70>

Received 9 June 2020

Revised 6 July 2020

Accepted 5 January 2021

© 2021. The Korea Institute of Intelligent Transport Systems. All rights reserved.

요 약

최근 활발히 연구되는 딥러닝 방법론은 인공지능의 성능을 급속도로 향상시켰고, 이에 따라 다양한 산업 분야에서 딥러닝을 활용한 시스템이 제시되고 있다. 교통 시스템에서는 GNN을 활용한 공간-시간 그래프 모델링이 교통 속도 예측에 효과적인 것으로 밝혀졌지만, 이는 메모리 병목 현상을 유발하기 때문에 모델이 비효율적으로 학습된다는 단점이 있다. 따라서 본 연구에서는 그래프 분할 방법을 통해 도로 네트워크를 분할하여 메모리 병목 현상을 완화함과 동시에 우수한 성능을 달성하고자 한다. 제안 방법론을 검증하기 위해 인천시 UTIC 데이터 분석 결과를 바탕으로 Jensen-Shannon divergence를 사용하여 도로 속도 분포의 유사도를 측정하였다. 그리고 측정된 유사도를 바탕으로 스펙트럴 클러스터링을 수행하여 도로 네트워크를 군집화하였다. 성능 측정 결과, 도로 네트워크가 7개의 네트워크로 분할되었을 때 MAE 기준 5.52km/h의 오차로 비교 모델 대비 가장 우수한 정확도를 보임과 동시에 메모리 병목 현상 또한 완화되는 것을 확인할 수 있었다.

핵심어 : 지능형 교통 시스템, 교통 속도 예측, Graph Neural Network, Jensen-Shannon divergence, 스펙트럴 클러스터링

ABSTRACT

Deep learning methodology, which has been actively studied in recent years, has improved the performance of artificial intelligence. Accordingly, systems utilizing deep learning have been proposed in various industries. In traffic systems, spatio-temporal graph modeling using GNN was found to be effective in predicting traffic speed. Still, it has a disadvantage that the model is trained inefficiently due to the memory bottleneck. Therefore, in this study, the road network is clustered through the graph clustering algorithm to reduce memory bottlenecks and simultaneously achieve superior performance. In order to verify the proposed method, the similarity of road speed distribution was measured using Jensen-Shannon divergence based on the analysis result of Incheon UTIC data. Then, the road network was clustered by spectrum clustering based on the measured similarity. As a result of the experiments, it was found that when the road network was divided into seven networks, the memory bottleneck was alleviated while recording the best performance compared to

the baselines with MAE of 5.52km/h.

Key words : Intelligent transport systems, Traffic forecasting, Graph neural network, Jensen-Shannon divergence, Spectral clustering

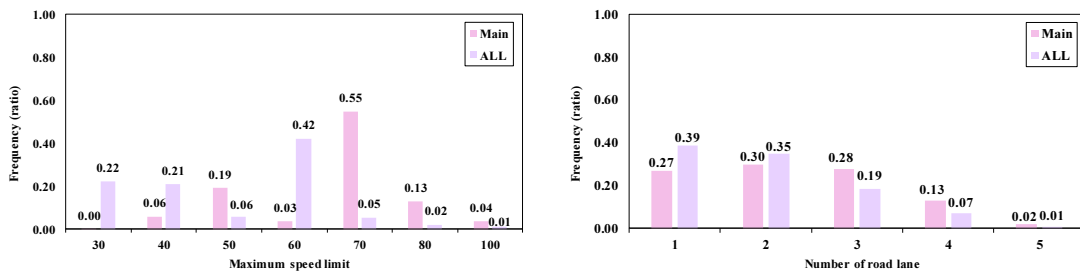
I. 서 론

국토교통부의 조사에 따르면 대한민국의 자동차 등록 대수는 1997년 1,000만 대를 돌파한 이래 2014년 2,012만대를 돌파하였다. 차량 수는 17년 동안 2배 이상 증가한 것으로, 이런 증가율이 유지된다면 2020년 경에는 2,500만대에 도달할 것으로 예측된다(K-indicator, 2020). 차량 수가 증가하면서 교통혼잡에 의해 발생하는 비용도 매년 지속적으로 증가하고 있다. 그 비용이 2017년 GDP 기준 3.4%에 달할 정도로 국가 경제활동에 큰 영향을 미치기 때문에 교통 혼잡을 완화하기 위한 지속적인 노력이 필요한 상황이다(K-indicator, 2020). 이에 따라 교통 혼잡을 줄이기 위하여 이용자 맞춤형 교통체계 구축, AI·자율주행·모빌리티 빅데이터 등 첨단 ICT 기술을 접목한 도로 이용의 효율성 증진과 지속적인 교통 수요 관리 정책이 요구되고 있다.

교통시스템 전체의 효율성을 극대화하기 위해 도로를 이용하는 차량의 속도를 예측하는 다양한 연구가 국내외에서 진행되고 있다. 특히 딥러닝을 기반으로 한 인공지능 방법론들의 발전은 인간이 인지하지 못하는 데이터의 특성을 반영할 수 있어 기존 방법론 대비 우수한 성능을 기록하고 있다. Fu et al.(2016)은 도로 속도 예측 시 시계열 특성을 반영하기 위한 RNN 기반의 모델을 제안하였고, Wang et al.(2016)은 물리적인 공간 정보를 반영하기 위해 CNN(convolutional neural network)기반의 모델을 설계하여 속도를 예측하였다. 이후 Yu et al.(2017)은 시계열 특성과 공간 정보 모두 반영하기 위해 RNN과 CNN을 결합한 SRCN(spatiotemporal recurrent convolutional network)을 제안하였다. 또한 시계열 특성과 공간 정보뿐만 아니라 예측하고자 하는 도로의 주변 도로 교통 상황까지 모델링하고자 GNN(graph neural network)기반의 모델이 활발하게 연구되기 시작 하였는데, Wu et al.(2019)이 제안한 Graph Wavenet은 특히 다른 기존 모델 대비 우수한 성능을 보였다.

하지만 대규모의 도로 네트워크를 그래프로 모델링 하는 것은 많은 계산량을 요구하는 단점이 있다. 이는 메모리 병목현상으로 이어져 학습의 비효율을 유발하는데, 기존 그래프 네트워크 관련 연구에서는 이를 해결하기 위해 METIS 알고리즘(George and Vipin, 1995)을 적용하여 대용량의 그래프를 하위 네트워크로 분할하고자 하였다(Mallick et al., 2019). 위 연구를 도로의 목적, 크기, 연결 관계 같은 특성과 함께 모델링에 반영할 수 있다면 더 좋은 성능과 더불어 메모리 병목 현상까지 완화할 수 있다.

복잡하게 연결된 도로 네트워크는 도로의 특성에 따라 하위 집단을 구분 할 수 있다. <Fig. 1>는 국토교통



<Fig. 1> Compare to main road and all road groups according to the attribute of traffic conditions

부의 지능형교통관리시스템에서 제공하는 표준노드링크 데이터의 인천지역의 도로들을 전체도로 집단과 도시교통 기초조사에서 선정한 주요 도로 집단으로 나누어 비교한 것이다. 왼쪽 그림은 도로의 최고제한속도를 기준으로 두 집단을 비교한 것으로, 전체도로 집단에서 5%만 차지하는 70km/h 도로가 주요 도로 집단에서는 55%에 해당하는 것을 알 수 있다. 반면 전체도로 집단에서 42%를 차지하는 60km/h 도로가 주요 도로 집단에서는 3% 밖에 되지 않는 것을 확인 할 수 있다. 오른쪽 그림은 차로 수를 기준으로 비교한 것으로 두 집단 모두 3차로 이하로 구성되어 있는 것을 확인할 수 있다. 그 중 주요 도로는 전체도로 대비 3차로의 비율이 높은 것으로 확인된다. 이를 통해 하위 집단은 도로들의 구성에 따라 다른 특성이 나타나는 것을 알 수 있다.

도로 네트워크는 속도분포의 유사성에 따라 하위 도로 네트워크로 분할될 수 있다. <Fig. 2>는 도시교통정보센터(urban traffic information center, UTIC)에서 제공하는 인천지역 도로들을 x축 속도, y축 빈도수(비율)인 히스토그램으로 나타낸 것이다. 네 개의 도로들은 오른쪽 꼬리분포와 왼쪽 꼬리분포 두 집합으로 나눌 수 있는데, 오른쪽 꼬리분포 집합은 왼쪽으로 치우쳐진 모양으로 해당하는 두 도로의 평균 속도가 높지 않다는 것을 알 수 있다. 반면, 왼쪽 꼬리분포 집합의 경우 오른쪽으로 치우쳐진 모양으로 왼쪽 꼬리분포 집합에 비해 평균 속도가 높은 것으로 확인된다.

본 연구에서는 위와 같은 도로의 특성을 반영하여 도로 네트워크를 하위도로 네트워크로 분할하고 도로의 교통 속도를 예측하였다. 대용량 도로 네트워크를 유사도로 네트워크로 분할하기 위해 Jensen-Shannon divergence(Endres and Schindelin, 2003)를 이용하여 도로 간 속도 분포의 유사도를 계산하고 스펙트럴 클러스터링(spectral clustering)을 이용하여 도로 네트워크를 군집화 하였다. 또한 시공간적 특성을 가진 그래프를 효율적으로 학습할 수 있도록 학습 공간에 투영하고 주변도로 교통상황뿐만 아니라 도로 간 숨겨진 종속성을 파악하여 속도 예측이 이루어지도록 아키텍처를 구성한 Graph Wavenet을 적용하였다.

본 논문의 구성은 다음과 같다. II장에서는 딥러닝 이론 및 딥러닝을 활용한 교통 속도 예측 관련 연구들에 대해 검토를 진행한다. III장에서는 제안된 모형의 이론적 배경을 설명하고, IV장에서 문제를 정의하고 제안된 모형을 설명한다. 마지막으로 V장에서는 모형을 평가하고 VI장에서 요약과 결론을 내린다.

II. 관련 연구 및 이론적 고찰

1. 딥러닝 알고리즘

1957년 Rosenblatt의 퍼셉트론을 시작으로 사람의 학습 방식을 모방한 ANN(artificial neural network)은 연구자들의 많은 관심에서 발전해왔다. 이후 Hinton et al.(2006)의 연구에서는 딥러닝이라는 명칭이 처음 소개되었다. 딥러닝은 기존 기계학습(machine learning)과 달리 사람의 구체적인 지시사항 없이 데이터만을 이용해 기계가 스스로 학습하여 기존 기계학습의 성능을 넘어서는 놀라운 성과를 보였다. 기존의 기계학습은 자연 데이터(raw data)를 그대로 처리하는 능력에 한계가 있어 도메인에 대한 전문 지식과 신중한 설계로 데이터의 패턴을 추출하였지만, 딥러닝은 학습 과정에서 데이터의 패턴을 스스로 추출해내는 모습을 보였다(Lecun et al., 2015). 딥러닝은 인공지능 커뮤니티에서 다년간 해결하려고 했던 문제를 풀어냄으로써 최신의 진보를 이루어 냈으며 현재 음성인식, 이미지 객체 인식, 객체 탐지, 자율주행차 등 매우 다양한 분야에서 뛰어난 성

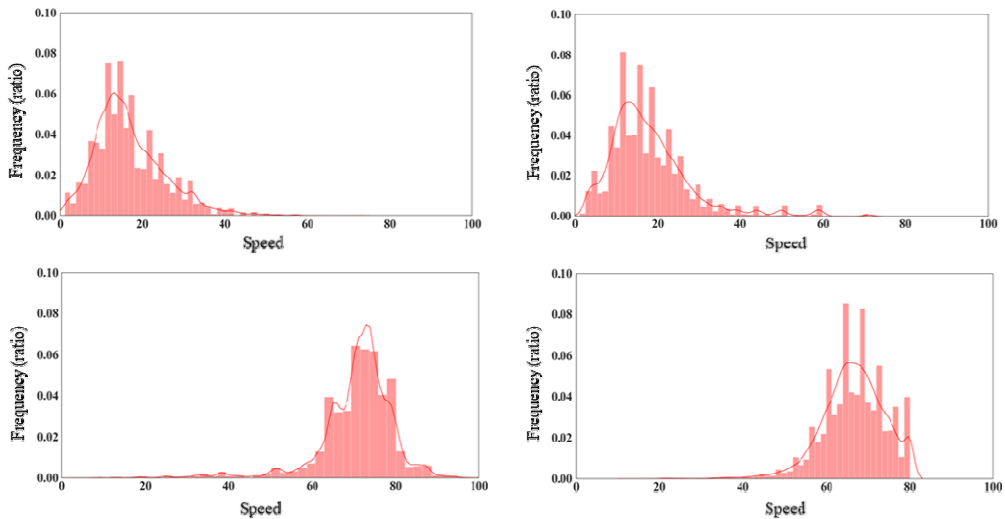
1) Jensen-Shannon divergence: 서로 다른 확률 분포 사이의 거리를 측정하는 척도

능으로 응용되고 있다.

2. 딥러닝 기반의 교통 속도 예측 관련 연구

<Table 1>은 딥러닝 기반의 교통 속도 예측 연구들을 요약한 표이다. 교통 속도 예측을 위한 딥러닝 모델은 Yisheng et al.(2014)가 설계한 DNN모델을 통해 처음 소개되었고 Kim et al.(2020)은 교통상황에 영향을 주는 요인들을 발굴하여 DNN모델의 예측 정확도를 향상시켰다. Fu et al.(2016)와 Zhao et al.(2017)은 시계열 특성을 모델에 반영하기 위해 RNN 기반의 모델을 사용하였다. 같은 시기에 Wang et al.(2016)와 Ma et al.(2017)은 도로 네트워크를 유클리드 공간의 무방향 그래프로 구성하고 CNN 기반의 모델을 적용하여 공간 정보를 이용한 교통 속도 예측을 진행하였다.

시계열 특성과 공간 정보를 동시에 반영하기 위해 Yu et al.(2017)은 CNN과 RNN 기반의 모델 앙상블(ensemble)을 시도하였고, Lv et al.(2018)은 과거 시계열 패턴과 날씨정보를 반영하도록 네트워크를 구성한 LCRNN(look-up convolution recurrent neural network)을 개발하였다. 또한 Kim et al.(2020)은 과거 시계열 패턴과 주변 도로의 교통상황을 입력 변수로 활용하여 LSTM(long short term memory)을 구축하였다. 하지만 교통 네트워크는 주변 도로의 교통 상황에 따라 임의성을 수반한 변동이 크기 때문에 교통 네트워크 내에서 고유한 토폴로지 정보가 손실되기 쉽다. 이를 방지하고자, GNN 기반의 모델을 통해 도로 네트워크를 모델링하는 연구 또한 많이 진행되었다. Yu et al.(2017)은 convolution filter의 가중치 공유(weight sharing)와 지역적 특징 학습(local feature learning)의 특성을 GNN에 적용한 GCN(graph convolution networks)을 통해 도로 네트워크 모델링을 진행하였다. Li et al.(2018)은 주변도로의 교통상황과 도로의 방향성을 모델링 할 수 있는 diffusion graph convolution layer와 RNN 기반의 모델인 GRU(gated recurrent unit)를 결합한 DCRNN(diffusion convolution recurrent neural network)을 제안하였고 향상된 예측 정확도를 보였다. Wu et al.(2019)의 Graph Wavenet은 diffusion graph convolution layer에 self adaptive adjacency matrix 알고리즘을 결합하여 모델이 숨겨진 도로 교통상황을 학습 할 수 있도록 유도하였고, 이는 다른 교통 속도 예측 모델들 대비 가장 뛰어난 성능을 보였



<Fig. 2> Speed distribution of four randomly selected roads

<Table 1> Summary of traffic forecasting base on deep learning

Dependency	Model		Reference
Temporal	DNN		Yisheng et al.(2014)
			Kim et al.(2020)
	RNN		Fu et al.(2016)
			Zhao et al.(2017)
Spatial	CNN		Ma et al.(2017)
			Wang et al.(2016)
Spatial and Temporal	LSTM		Kim et al.(2020)
	Ensemble	SRCN	Yu et al.(2017)
		LC-RNN	Lv et al.(2018)
	GNN	GCN	Yu et al.(2017)
		DCRNN	Li et al.(2018)
		GANN	Zhang et al.(2018)
		GraphWavenet	Wu et al.(2019)
		STGRAT	Park et al.(2019)

다. 이외에 Zhang et al.(2017)은 어텐션 메커니즘(attention mechansim) 기반의 GNN 모델을 설계하였고, Park et al.(2019)는 self-attention을 결합한 STGRAT(spatio-temporal graph attention)을 개발하여 도로 속도를 예측하는 등 다양한 알고리즘을 바탕으로 연구가 진행되고 있다.

GNN 기반의 모델이 좋은 성능을 보였음에도 불구하고 도로 네트워크가 크면 클수록 모델이 학습해야 하는 파라미터의 수가 증가하여 계산량이 늘어났다. 하지만 한정된 GPU 메모리에서는 대용량 도로 네트워크를 학습하는데 어려움이 있었다. 이를 해결하기 위해 Mallick et al.(2019)은 METIS 그래프 분할 알고리즘을 통해 대용량 도로 네트워크를 작은 크기로 분해하여 병렬적으로 학습하는 방식을 채택하였다. METIS 그래프 분할 알고리즘은 가능한 균등하고 빠르게 서버 네트워크로 분할하면서 분할된 네트워크 간 속성 또한 유사하도록 분해 하는 방식이다. 위 방법을 통해 Mallick et al.(2019)은 GPU 메모리 사용량을 감소시켜 대용량의 도로 네트워크를 효율적으로 학습할 수 있었다.

3. GCN 알고리즘

GNN은 데이터간의 상호작용으로 표현된 고차원의 그래프 데이터를 저차원의 벡터 공간으로 투영하여 학습하는 인공지능망이다. 그래프란 $G = (X, A)$ 로 정의되며 여기서 X 는 데이터의 속성을 나타내는 노드(꼭지점)의 집합이고 A 는 노드와 노드 사이의 관계를 나타내는 엣지(변)의 집합이다. 각 노드는 주변 이웃 노드로부터 정보를 전달받는 형식에 따라 다양한 방식으로 표현이 가능하다. Kipf and Welling(2016)의 GCN은 convolution 연산의 가중치 공유(weight sharing)와 지역적 특성 학습(local feature learning)기반으로 주변 이웃 노드의 정보를 전달하도록 고안된 모델이다. 즉 l 번째 계층 신경망에서 노드의 속성행렬(X)과 이웃 노드 간의 관계로 표현된 인접행렬(A)을 convolution 연산하여 $l+1$ 번째 노드행렬(X)을 생성하는 것이다. 이를 일반화하여 다음 식(1)과 같이 정의할 수 있다.

$$X^{(l+1)} = \tilde{A}X^{(l)}W^{(l)} \dots \dots \dots (1)$$

여기서,

$$\tilde{A} = A + I$$

A 는 노드와 노드 사이의 연결 관계를 수치로 표현한 인접행렬

I 은 단위행렬, X 는 속성행렬

W 은 입력값에 대한 가중치행렬

Li et al.(2018)은 인접한 도로에 위치한 노드들 중 중계 노드를 선정하여 중계 후보 노드들의 인접 정도 (k-hop)를 조정하고 노드와 노드 간의 방향성과 거리 기반으로 주변 이웃 노드의 정보를 전달하는 diffusion graph convolution layer를 제안하였다. Diffusion graph convolution layer는 중심 노드에서 k-hop 까지의 이웃 노드로의 연관성(P_f)과 k-hop 까지의 이웃 노드에서 중심 노드로의 연관성(P_b) 더하여 $l+1$ 번째 노드행렬(X)을 도출한다. 다음 식(2)을 통해 이를 확인할 수 있다.

$$X^{(l+1)} = \sum_{k=0}^K P_f^k X^{(l)} W_{k,1}^{(l)} + P_b^k X^{(l)} W_{k,2}^{(l)} \dots \dots \dots (2)$$

여기서,

$$P_f = \tilde{A} / \text{rowsum}(\tilde{A})$$

$$P_b = \tilde{A}^T / \text{rowsum}(\tilde{A}^T)$$

$W_{k,1}$ 은 순방향 입력값에 대한 가중치행렬

$W_{k,2}$ 은 역방향 입력값에 대한 가중치행렬

Wu et al.(2019)의 연구에서는 노드와 노드 사이 연관성에 대한 사전 지식이 없이 모델의 확률적 경사하강 법을 통한 학습으로 노드 간 종속성을 모델링하는 self adaptive adjacency matrix를 제안하였다. 이를 diffusion graph convolution layer와 합쳐 방향성에 따른 노드 간 종속성뿐만 아니라 모델 자체적으로 숨겨진 노드간 종속성을 파악하는 diffusion graph convolution layer + self adaptive adjacency matrix 알고리즘을 제안하였다. Self adaptive adjacency matrix는 초기 균일 분포 기반의 무작위 값으로 E_1 벡터와 E_2^T 벡터를 생성한다. 생성된 두 벡터를 내적하고 softmax 활성화함수를 통해 0과 1 사이 값으로 구성된 행렬(\tilde{A}_{apt})을 출력한다. 이렇게 계산된 self adaptive adjacency matrix와 diffusion graph convolution layer를 더하여 $l+1$ 번째 노드행렬(X)을 도출한다. 이후 역전파 시 self adaptive adjacency matrix의 두 벡터는 노드 간 종속성을 반영한 새로운 벡터로 학습된다. 제안된 알고리즘은 식(3)와 같다.

$$X^{(l+1)} = \sum_{k=0}^K P_f^k X^{(l)} W_{k,1}^{(l)} + P_b^k X^{(l)} W_{k,2}^{(l)} + \tilde{A}_{apt}^k X^{(l)} W_{k,3}^{(l)} \dots \dots \dots (3)$$

여기서,

$$\tilde{A}_{apt} = \text{softmax}(E_1 E_2^T)$$

$$\text{softmax}(z_j) = \frac{e^{z_j}}{\sum_{i=0}^n e^{z_i}} (i=0,1,\dots,n)$$

E_1 은 소스 노드 임베딩 벡터

E_2^T 은 타겟 노드 임베딩 벡터

$W_{k,3}$ 은 모델이 학습하게 되는 가중치

4. TCN(Temporal Convolution Network) 알고리즘

시계열 데이터는 각 시간 단계에 대한 입력 변수로 구성되며, 순차적 데이터 처리에 특화된 RNN 기반 모델을 통해 시계열 특징을 추출했다. 그러나 RNN 기반의 모델은 관련 정보와 그 정보를 사용하는 퍼셉트론 사이 거리가 멀 경우 역전파 시 기울기 사라짐 문제(vanishing gradient problem)가 발생하여 학습 능력이 크게 저하 된다. 이 문제를 극복하기 위해 TCN 기반의 모델 연구가 진행되었다.

TCN은 입력값에 대해 casual convolution filter를 취하여 출력값을 구하게 되는데 이는 convolution 연산을 현재 시점보다 이전 시간 단계의 영역으로만 제한을 둔 것이다. Casual convolution filter를 통한 연산은 수용 영역(receptive field)을 넓히기 위해 수많은 레이어를 쌓아야 하는데 순차적으로 데이터를 처리해야 하는 시계열 데이터에서는 많은 계산량으로 문제가 발생한다. 이를 보완하기 위해 Oord et al.(2016)는 <Fig. 3>과 같은 dilated convolution filter를 제안하였다. Dilated convolution filter는 dilation만큼의 casual convolution filter를 취하여 계산량은 줄이고 각 입력을 포괄하는 수용영역을 확보할 수 있도록 조정하였다. 이를 통해 효과적으로 레이어를 쌓아 시계열 특성을 얻을 수 있다. 1차원 시계열 벡터($x \in R^T$)에 dilated convolution filter ($filter: 0, 1, \dots, k-1 \in R$)를 연산하는 식(4)은 다음과 같다.

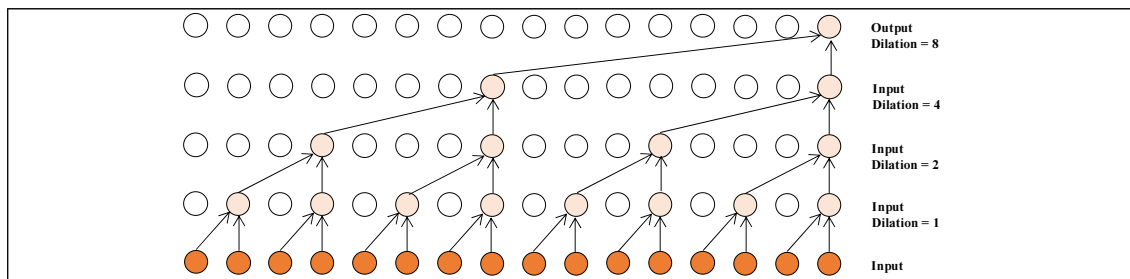
$$(x *_{d} f)(t) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-di} \dots \dots \dots (4)$$

여기서,
 $d = 2^v$, v 은 레이어의 층
 x 는 입력 벡터
 k 은 filter의 개수
 $s-di$ 은 이전 시점의 위치

또한 Oord et al.(2016)은 dilated convolution filter를 취한 입력값에 -1과 1 사이 값을 출력하는 tanh 활성화 함수와 0과 1 사이 값을 출력하는 sigmoid 활성화 함수(σ)를 \odot 하는 Gated TCN을 제안하였다. Gated TCN의 식(5)은 다음과 같다.

$$X^{(l+1)} = \tanh(W_f^{(l)} * X^{(l)}) \odot \sigma(W_g^{(l)} * X^{(l)}) \dots \dots \dots (5)$$

여기서,
 W_f, W_g 은 각 입력값에 대한 가중치
 $\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$
 $\sigma(z) = \frac{1}{1 + e^{-z}}$



<Fig. 3> Visualization of stack of dilated casual convolution(Oord et al., 2016)

III. 모형 구축

1. 문제 정의

딥러닝 모델은 최적 성능을 도출하기 위해 수많은 양의 데이터를 반복적으로 학습해야 한다. 빅데이터를 빠르고 효율적으로 학습하기 위해 일반적으로 많은 코어 수를 바탕으로 병렬적으로 수천 개의 연산이 가능한 GPU(graphics processing unit)를 활용한다. GPU가 한 번에 처리할 수 있는 데이터의 양은 내장되어있는 RAM의 용량에 따라 정해지며, 이를 초과할 시 메모리 병목현상이 나타난다. 본 연구에서는 11GB 메모리 용량을 가진 GPU인 RTX 2080 Ti를 사용하였으나, Graph Wavenet은 도로의 교통 속도뿐만 아니라 도로 네트워크 또한 모델의 추가 입력으로 넣어야 하기 때문에 11GB의 메모리 사용량을 초과한다. 따라서 도로 간 특성에 따라 도로 네트워크를 유사도로 네트워크로 분할하여 메모리 병목현상을 해결할 뿐만 아니라 각 유사도로 특성에 적합한 모델을 구축함으로써 모델의 성능까지 개선하고자 한다.

교통 속도 예측 모델이 이전 s 시점부터 t 시점까지의 도로 별 교통 속도(x_{t-s+1}, \dots, x_t)와 유사도로 네트워크($G(V, E)$)를 입력으로 사용하여 미래 시점의 도로 별 교통 속도를 예측하는 식(6)은 다음과 같다.

$$F([x_{t-s+1}, \dots, x_t]; G(V, E)) = [x_{t+1}, \dots, x_{t+s}] \dots \dots \dots (6)$$

여기서,
 x_t 은 t 시점에서의 속도
 s 은 window size

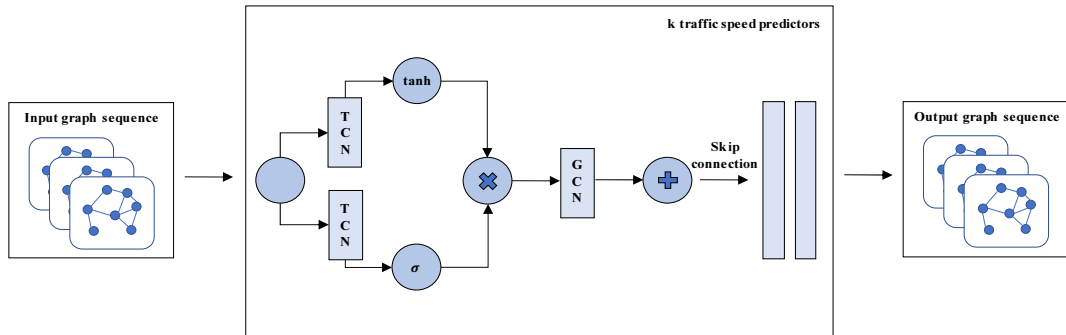
2. 제안 모형

본 연구에서는 도로 네트워크를 분할하기 위해 Jensen-Shannon divergence 기반의 spectral clustering을 진행한다. Spectral clustering의 진행 과정은 3단계로 이루어져 있으며 첫 번째 단계에서는 Jensen-Shannon divergence 알고리즘을 통해 두 도로 사이의 관계를 추출하고 이를 통해 유사도 네트워크를 생성한다. 두 번째 단계에서는 유사도로 네트워크를 사용하여 graph Laplacian matrix를 구성하고, 이를 고유 값 분해하여 고유 벡터를 통해 원본 데이터를 저차원으로 투영한다. 마지막으로 graph cut 알고리즘을 통해 도로 네트워크를 유사도로 네트워크로 분할한다.

<Fig. 4>는 분할된 유사도로 네트워크에 Graph Wavenet을 적용하는 과정을 도식화한 것이다. 먼저 도로 네트워크의 입력값이 서로 다른 활성화 함수로 구성된 Gated TCN으로 전달되어 시계열 특성이 추출된 후 diffusion graph convolution layer와 self adaptive adjacency matrix 특성을 가진 GCN으로 추출된 특성이 전달되어 각 유사도로 네트워크의 특성에 따른 시공간적 모델링이 진행된다. 이후 과정으로는 선형회귀 특성을 가진 완전연결 레이어(fully-connected layer)을 통해 예측값을 출력한다.

1) 유사 네트워크 생성

분석 도로들의 집합인 $V = \{v_1, v_2, \dots, v_n\}$ 에서 v_i 와 v_j 의 유사성에 대한 가중치를 $w_{ij}(w_{ij} \geq 0)$ 으로 표현한다. 각 도로쌍의 모든 유사성에 대한 가중치를 계산하고 이를 인접 행렬로 구성한 것이 $W = (w_{ij})_{i,j=1, \dots, n}$ 이다. 일반적으로 도로 네트워크를 인접 행렬로 구성할 때에는 단순히 도로가 연결되어 있는 경우 w_{ij} 를 1, 연결되어 있지 않을 경우 0으로 표현하지만 본 연구에서는 도로 간 속도 분포의 유사성



<Fig. 4> The architecture of the proposed method

에 따라 네트워크를 구성하고자 하였기 때문에 도로 간 속도 분포의 유사도로 w_{ij} 을 계산한다. 속도 분포의 유사도는 Jensen-Shannon divergence 알고리즘인 식(7)에서 계산할 수 있다.

$$JS(P \parallel Q) = \frac{1}{2} D_{KL}(P \parallel \frac{P+Q}{2}) + \frac{1}{2} D_{KL}(Q \parallel \frac{P+Q}{2}) \dots\dots\dots (7)$$

여기서,

$$D_{KL}(P \parallel Q) = \sum_{i=0} P_i * \log(\frac{P_i}{Q_i})$$

P_i, Q_i 은 각 도로 분포에서 i 번째 속도가 나올 확률

2) Graph Laplacian matrix 생성 및 고유 값 분해

도로들의 집합인 $v_i \in V$ 의 차수 d_i 은 각 도로에 연결된 도로의 개수를 의미한다. 차수 행렬 D 는 대각 원소가 d_1, d_2, \dots, d_n 으로 구성된 대각 행렬이다. 이를 통해 graph Laplacian matrix를 생성할 수 있으며 식(8)은 다음과 같다.

$$L = D - W \dots\dots\dots (8)$$

이렇게 구성된 graph Laplacian matrix는 불완전한 실수로 구성될 수 있어 차수에 정도에 따라 가중치를 부여하는 normalized graph Laplacian matrix의 식(9)로 변형할 수 있다.

$$L_{norm} = D^{1/2} L D^{1/2} = 1 - D^{1/2} W D^{1/2} = 1 - L' \dots\dots\dots (9)$$

Normalized graph Laplacian matrix은 고유 값 분해하여 k 개의 고유 값(eigenvalues)과 그에 대응하는 고유 벡터(eigenvectors)로 구분할 수 있으며, 이는 고차원으로 구성된 그래프를 고유 벡터를 통해 저차원 공간에 투영하여 보다 낮은 차원으로 본래의 그래프를 효과적으로 표현한 것이다.

3) Normalized cut(Ncut)을 통한 그래프 분할

그래프를 군집 간 분산을 최대화하면서 군집 내 분산을 최소화하며 분할하기 위해 normalized cut(Ncut)

알고리즘(10)을 적용한다. 즉 군집 간의 가중치를 최소화하는 분할된 그래프를 찾는 것이 Ncut의 목표로 목적식(11)은 다음과 같다.

$$Ncut(v_1, \dots, v_k) = \frac{1}{2} \sum_{i=1}^k \frac{W(\bar{v}_i, v_i)}{vol(v_i)} \dots \dots \dots (10)$$

여기서,

$$W(\bar{v}_i, v_j) = \sum_{i \in v_i, j \in v_j} w_{ij}$$

\bar{v}_i 은 v_i 에 대응하는 도로

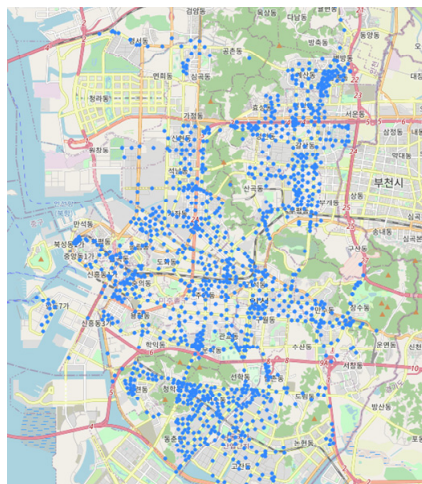
$v_i(A_i)$ 은 d_i

$$minimize \sum_{i=1}^k \frac{W(\bar{A}_i, A_i)}{vol(A_i)} \dots \dots \dots (11)$$

하지만 Ncut을 계산하는 문제는 각 단계에서 여러 가지의 가능성을 동시에 고려할 수 있어 다수개의 풀이 가 있는 NP-Hard 문제로 바뀐다. 이 문제를 해결하는 핵심 원리는 normalized graph Laplacian matrix을 통해 Ncut 목적식을 추적 최적화(trace optimization)로 변환하여 근사치를 찾도록 완화 시키는 방법이다. 따라서 완화된 목적식을 통해 Ncut으로 k개의 군집을 나눌 경우, normalized graph Laplacian matrix의 k번째 고유 벡터를 구하는 것과 같다.

3. 데이터 분석 구간 설정

UTIC은 전국 주요 도시에 교통정보센터, 수진제공장치(ToWay), CCTV, 도로전광판(VMS)등 기반 시설을 확충하고, 각 도시의 교통정보를 표준화하여 연계함으로써 전국 단위의 광역 교통정보를 수집하고 있다. 분석 데이터는 UTIC에서 제공하는 전국 주요 도시의 교통정보 중 인천광역시에 해당하는 교통정보로, 통행량이 많아 교통정보 수집 샘플을 충분히 확보할 수 있는 주요 도로로 설정하였다. 이런 주요 도로들이 <Fig. 5>에 점으로 표시되어 있다. 또한 주요 도로 구간에 인접해있는 상류 및 하류 구간도 분석 데이터로 활용하



<Fig. 5> UTIC dataset covering the main road of Incheon area, where traffic segments are plotted with colors

여 주요 도로로 유입되는 도로들의 영향권을 포함하도록 하였다. 따라서 데이터 분석 대상 구간은 2,783개 도로에 대한 2019년 4월부터 6월까지의 데이터이며 5분 단위로 측정된 도로의 교통 속도만을 이용하였다.

4. 훈련 데이터 구축

본 논문에서는 훈련 데이터를 구축하기 위해(Li et al., 2018)에서와 동일한 데이터 전처리 절차를 채택하였다. 직전 시간대 교통 속도는 분석 도로의 교통 속도에 가장 큰 영향을 미치는 요소이기 때문에 입력값을 예측 목표 시간 기준 이전 1시간으로 하였다. 따라서 2,783개의 도로의 26,141개 교통 속도 데이터 세트 중 70%는 모델 학습을 위한 훈련 데이터 세트로 활용하였고 10%는 모델 검증을 위한 검증 데이터 세트로 활용하였으며 20%는 모델 평가를 위한 평가 데이터 세트로 활용하였다. 한편 도로 간의 관계를 나타내는 인접 행렬은 가우스 커널로 투영된 도로 간 거리 따라 구성되는데, UTIC 데이터에서는 도로 간 거리를 제공하지 않아 표준노드링크 데이터에서 제공하는 노드 간의 거리를 통해 도로 간 거리의 근사치를 계산하였다. <Fig. 6>은 표준노드링크를 통해 도로 간 거리를 구하는 예시로 (a)와 (b) 도로 간의 거리를 구하기 위해 해당 도로들의 양 끝 노드에 해당하는 (A)와 (B), (B)와 (C)의 거리를 합한 후 절반의 값을 근사치로 사용하여 인접 행렬을 구성하였다.



<Fig. 6> Calculate of the distance from the node (A) to node (C) based on the provided distance of (a) and (b)

IV. 모형 평가

1. 비교모형 및 평가척도 설정

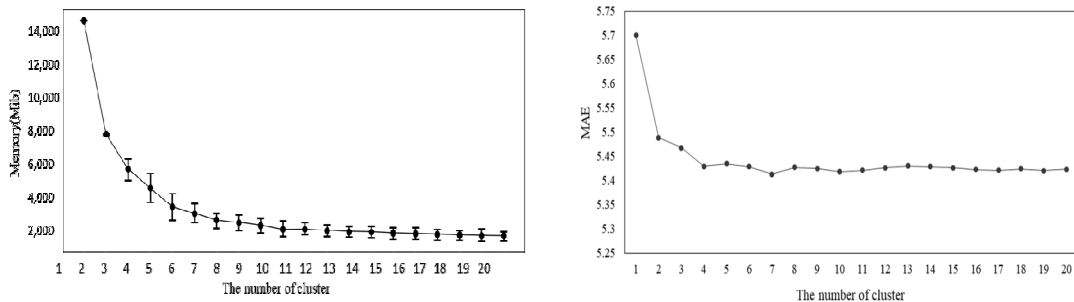
본 논문에서는 제안하는 교통 속도 예측 모델의 성능을 파악하기 위해 서로 다른 정보를 반영하는 교통 속도 예측 모델들과 비교하였다. 비교 모델로는 세 가지를 사용하였는데, 첫 번째로는 LSTM으로 시계열 데이터 모델링에서 시계열 특성을 파악하기 위해 가장 보편적으로 사용되는 RNN 기반의 모델이다. 두 번째로는 시계열 특성과 유클리드 공간상의 정보를 반영하기 위한 앙상블 모델로 1D convolution layer와 LSTM으로 네트워크를 구성한 CNN-LSTM을 활용하였다. 마지막으로 시계열 특성과 도로 간의 관계 정보인 도로 네트워크를 반영하기 위해 diffusion graph convolution layer와 RNN 기반의 모델인 GRU(gated recurrent unit)를

encoder-decoder 방식으로 결합한 DCRNN을 활용하였다. DCRNN은 본 연구에서 제안하는 모델인 Graph WaveNet과 비교하여 도로 네트워크를 모델링 한다는 공통점이 있지만 도로 네트워크를 반영하는 방식과 메모리 효율 측면에서의 차이가 있어 비교모형으로 설정하였다. 모델들 간 성능 평가 지표로는 회귀 예측의 대표적인 평가 지표인 MAE(mean absolute error), RMSE(root mean square error), MAPE(mean absolute percent error)을 활용하였다.

2. 최적 군집 수 설정

본 연구의 주요 목적은 메모리 병목 현상을 완화하는 것이다. 이를 위해 하위 네트워크의 최적 군집 수를 탐색하였다. 또한, 메모리 병목 현상을 정량적인 척도로 비교하기 위해 모델 학습 시 필요한 메모리량을 측정하였다. 여기서 메모리량이란 덤퍼닝 모델을 학습 시 GPU에 할당되는 메모리량을 의미하며, GPU에서 연산 가능한 메모리량을 초과하면 메모리 병목현상이 발생한다. <Fig. 7>에서 왼쪽 그래프는 군집 별 메모리량의 최대, 평균, 최소값을 비교한 것으로 도로 네트워크를 분해할수록 모델 학습에 필요로 하는 메모리량은 줄어드는 것을 확인할 수 있다. 이를 통해 모델 학습에 이용 가능한 메모리량이 늘어나고 모델 최적화를 수행하여 예측 성능과 훈련 속도를 향상시킬 수 있다.

<Fig. 7>에서 오른쪽 그래프는 Jensen-Shannon divergence 기반의 spectral clustering으로 생성된 군집의 개수에 따라 Graph WaveNet을 적용하고 성능 비교를 한 것이다. 앞서 제시한 평가 척도 중 RMSE를 이용해 각 군집마다의 성능을 구하고, 군집에 할당 된 도로 수에 비례하는 가중평균값 군집들의 대표 성능으로 설정하였다. 필요 메모리량과 MAE 지표를 종합적으로 고려한 결과, 학습에 필요한 메모리량은 2,128Mib, MAE값은 5.41인 군집의 수가 7개일 때 학습이 가장 효과적인 것으로 확인되었다. 이는 각 하위 네트워크의 고유한 시공간적 영향력을 모델링할 수 있어 정확한 예측이 가능해진 것으로 해석된다.



<Fig. 7> Performance of the proposed method and GPU memory that change as the number of cluster increases

3. 평가 환경 및 모형 평가

본 연구에서는 모형 평가를 위해 앞서 구축한 훈련 데이터로 비교 모형들을 학습하였다. 각 비교 모형들은 가용 가능한 메모리량에서 모델 최적화를 진행하였으며, 메모리 병목 현상이 나타날 경우 Jensen-Shannon divergence 기반의 spectral clustering을 적용하였다. 군집의 수는 위 실험을 통해 최적의 군집수로 판단되는 7로 설정하였다. 본 연구의 제안모델은 각 유사도로 네트워크에서 해당 모델을 최적화하기 위해 학습율이

<Table 2> Performance of the proposed method and the other comparison methods

	After 15 minute			After 30 minute			After 60 minute		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
LSTM	6.52	8.91	25.70%	6.86	9.20	27.32%	7.24	9.53	29.12%
CNN-LSTM	6.51	8.89	25.51%	6.85	9.18	27.12%	7.23	9.50	28.91%
DCRNN (clustering)	5.82	8.46	21.45%	6.13	8.74	22.81%	6.47	9.05	24.31%
Graph WaveNet	5.70	8.43	21.06%	5.95	8.66	22.28%	6.20	8.91	23.40%
Proposed method	5.41	7.83	20.33%	5.63	8.03	20.78%	5.81	8.11	21.82%

0.005인 Adam 최적화 알고리즘을 사용하였으며, 드롭아웃(dropout)의 비율은 0.5, TCN의 히든 레이어 개수를 128로 설정하였다.

<Table 2>는 본 연구에서 제안하는 모델과 비교 모델의 성능을 비교한 것이다. 결과 값이 작을수록 성능이 우수한 것으로 제안 모델이 비교 모델들보다 모든 시점, 모든 지표에서 가장 우수한 성능을 보였다. 구체적으로, 시계열 특성만을 반영한 LSTM, 시계열 특성과 유클리드 공간상의 정보를 반영한 CNN-LSTM, 주변 도로 관계 정보인 시공간 도로 네트워크를 반영하기 위한 DCRNN(clustering), Graph WaveNet, 제안모델 순으로 성능이 좋은 것을 확인할 수 있다. 특히 장기간의 예측에서 보다 뚜렷한 차이를 확인할 수 있는데, LSTM과 제안 모델 비교 시 제안모델의 MAE(60분 뒤 예측) 5.81km/h, RMSE(60분 뒤 예측) 8.11km/h, MAPE(60분 뒤 예측) 21.82%로 MAE 기준 약 1.4km/h까지 차이 나는 것을 확인할 수 있다. 이를 통해 교통 속도 예측에서는 도로들 간에 서로 영향을 주는 유기적인 연결고리가 있고, 이를 모델의 추가 변수로 반영하는 것이 전체적인 성능에 긍정적인 영향을 미친다는 것을 알 수 있었다.

또한 DCRNN은 메모리 병목 현상으로 제안 모델과 같은 군집으로 실험을 진행하였는데, 제안 모델과 비교 시 MAE 기준 최대 0.7까지 차이 나는 것을 확인할 수 있다. 이로부터 같은 GNN 기반의 모델일 경우에도 주변 노드 정보를 반영하는 알고리즘의 구성에 따라 정확도가 달라지는 것을 확인할 수 있었다. 한편, 도로 네트워크를 분할 한 제안 모델은 도로 네트워크를 분해하지 않은 Graph WaveNet에 비해 예측 정확도가 MAE 기준 최대 0.4km/h까지 개선되는 것을 확인할 수 있었는데, 이로부터 도로의 속도 분포에 따라 분할된 유사도로 네트워크에 고유한 시공간적 영향력을 모델링하는 것이 모델 성능에 효과적이라는 것을 확인할 수 있었다.

4. 군집 별 모형 평가 결과 분석

<Table 3>은 Jensen-Shannon divergence 기반의 spectral clustering로 분할 된 각 군집에서 본 연구에서 제안하는 모델과 비교 모델의 15분 뒤 예측 성능을 비교한 것이다. 비교 모델은 시간적 모델링, 공간적 모델링, 시공간적 모델링의 예측 정확도를 확인하기 위해 LSTM, CNN-LSTM, Graph WaveNet을 적용하였다. DCRNN은 Graph WaveNet에 비해 보다 더 많은 메모리량이 요구되고 예측 성능이 떨어지기에 비교 모델 군에서 제외하였다. 모든 군집에서 시공간적 모델링을 위한 Graph WaveNet이 LSTM과 CNN-LSTM에 비해 우수한 성능이 나오는 것을 확인할 수 있었고, 특히 군집 5를 확인해보면 MAE 기준 최대 2까지 차이 나는 것을 확인

<Table 3> Performance of the proposed method and the other comparison methods in the cluster

Cluster	Number of road	Model	MAE	RMSE	MAPE	Cluster	Number of road	Model	MAE	RMSE	MAPE
1	457	LSTM	8.69	10.99	28.83%	5	318	LSTM	4.46	6.41	33.93%
		CNN-LSTM	8.67	10.97	27.84%			CNN-LSTM	4.38	6.21	33.86%
		Graph WaveNet	8.01	9.89	26.31%			Graph WaveNet	2.45	4.89	17.01%
2	494	LSTM	7.83	10.11	36.41%	6	388	LSTM	4.17	5.72	21.11%
		CNN-LSTM	7.81	10.02	36.35%			CNN-LSTM	4.16	5.67	20.13%
		Graph WaveNet	6.98	9.16	30.69%			Graph WaveNet	3.16	4.95	15.76%
3	284	LSTM	7.46	10.38	17.61%	7	430	LSTM	5.36	7.33	19.17%
		CNN-LSTM	7.44	10.02	17.01%			CNN-LSTM	5.37	7.29	18.95%
		Graph WaveNet	6.55	9.84	15.69%			Graph WaveNet	4.19	6.44	14.71%
4	412	LSTM	6.92	9.16	18.69%						
		CNN-LSTM	6.88	9.15	18.67%						
		Graph WaveNet	6.15	8.64	16.91%						

할 수 있다. 이를 통해 주변도로의 영향력까지 반영한 시공간적인 모델링이 속도 예측에 유의미하다는 것을 알 수 있다.

한편 <Table 3>에서 군집 별로 예측 성능의 편차가 심하다는 것을 알 수 있다. 구체적으로 5번 군집은 318개의 도로들에서 Graph waveNet 기준 MAE 2.45km/h, RMSE 4.89km/h, MAPE 17.01%의 성능을 확인할 수 있는 반면 1번 군집은 457개의 도로들에서 Graph waveNet 기준 MAE 8.01km/h, RMSE 9.88km/h, MAPE 26.31%로 예측 성능의 차이를 확인할 수 있었다. 이는 도로 네트워크가 속도 분포에 유사성에 따라 분할됨에 따라 유사 도로들 고유 특성이 반영된 결과라고 판단되었다. 이를 검증하기 위해 <Fig. 8>에서는 차로 수와 최고제한속도로 도로의 특성별(column-wise) 군집의 할당 비율을 확인해보았다. 예측 오차가 비교적 큰 1,2,3번 군집들은 예측 성능이 작은 군집들 보다 차로 수가 많고 최고 제한 속도가 크다는 것을 확인할 수 있

Cluster	Number of lane					
	1	2	3	4	5	6
1	0.07	0.17	0.29	0.37	0.24	0.00
2	0.09	0.28	0.30	0.17	0.08	0.50
3	0.11	0.09	0.12	0.21	0.30	0.00
4	0.12	0.16	0.22	0.23	0.35	0.50
5	0.30	0.12	0.04	0.00	0.03	0.00
6	0.32	0.19	0.04	0.02	0.00	0.00
7	0.20	0.28	0.12	0.05	0.08	0.00

Cluster	Maximum speed limit					
	30	40	50	60	70	80
1	0.06	0.06	0.05	0.23	0.21	0.33
2	0.04	0.05	0.21	0.11	0.19	0.18
3	0.04	0.05	0.21	0.11	0.19	0.18
4	0.09	0.07	0.19	0.20	0.18	0.16
5	0.23	0.30	0.12	0.08	0.02	0.02
6	0.24	0.29	0.11	0.12	0.07	0.03
7	0.31	0.18	0.12	0.16	0.15	0.11

<Fig. 8> Ratio with column-wise to the attribute of traffic conditions

었다. 해당 군집들은 도로의 특성 상 통행량이 많아짐에 따라 교통 속도의 편차가 커질 것이고 이는 예측 오차를 계산하는데 영향을 미쳤을 것이라는 것을 알 수 있다. 이렇게 각 군집 별로 고유한 특성이 존재하고 이를 반영하여 모델링 할 수 있다면 보다 적합한 모델을 생성할 수 있다고 예상된다.

V. 결 론

본 논문에서는 도로 네트워크를 모델링하여 시계열 특성과 공간 정보를 반영할 수 있는 GNN 모델을 통해 도로의 교통 속도를 예측하는 연구를 진행하였다. 또한 인천지역의 도로 네트워크를 분석하여 도로 집합들 간 고유 특성을 확인하였고, 이를 반영한 교통 속도 예측 모델을 구축하였다. 구체적으로 개별 도로 고유 특성의 유사성에 따라 군집화를 수행하여 GNN 모델의 한계점인 메모리 병목 현상을 완화하였고 속도 분포의 유사성에 따라 네트워크를 분할하는 Jensen-Shannon divergence 기반의 spectral clustering을 제안하였다.

본 논문에서는 제안 방법론을 평가하기 위해 7개의 유사도로 네트워크로 분할하였고 서로 다른 시공간적인 정보를 반영하는 교통 속도 예측 모델들과 속도 예측 성능을 비교하였다. 그 결과 제안 방법론이 비교 모델들과 비교하여 가장 우수하였고, 유사도로 네트워크 별 성능 측정 결과에서 또한 제안 모델의 성능이 가장 우수한 것을 확인할 수 있었다. 이를 통해 제안모델이 메모리 병목 현상 문제를 완화함과 동시에 속도 예측 측면에서도 적합한 것을 확인할 수 있었다.

본 연구는 기상 정보, 돌발적인 상황 등 다양한 변수들의 영향력을 고려하지 않고 속도 변수만을 이용하였다는 단점이 있다. 이를 보완하기 위해 향후 연구에서는 하위 네트워크의 다양한 변수를 활용한 교통 속도 예측 연구도 같이 수행하고자 한다.

ACKNOWLEDGEMENTS

이 성과는 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No.2020R1C1C1009848)

REFERENCES

- Endres, D. M. and Schindelin, J. E.(2003), "A New Metric for Probability Distributions," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp.1858-1860.
- Fu, R., Zhang, Z. and Li, L.(2016), "Using LSTM and GRU Neural Network Methods for Traffic Flow Prediction," *Youth Academic Annual Conference of Chinese Association of Automation*, Wuhan, China, pp.324-328.
- George, K. and Vipin, K.(1998), "A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs," *SIAM Journal on scientific Computing*, vol. 20, no. 1, pp.359-392.
- Hinton, G. E., Osindero, S. and Teh, Y. W.(2006), "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation*, vol. 18, no. 7, pp.1527-1554.
- D. Kim, K. Hwang and Y. Yoon,(2019), "Prediction of Traffic Congestion in Seoul by Deep Neural

- Network,” *The Journal of the Korea Institute of Intelligent Transport Systems*, vol. 18, no. 4, pp.44-57.
- Y. Kim, J. Kim, Y. Han, J. Kim and J. Hwang,(2020), “Development of Traffic Speed Prediction Model Reflecting Spatio-Temporal Impact Based on Deep Neural Network,” *The Journal of the Korea Institute of Intelligent Transport Systems*, vol. 19, no. 1, pp.1-16.
- K-indicator(2020), <https://www.index.go.kr/>
- Kipf, T. N. and Welling, M.(2016), “Semi-Supervised Classification with Graph Convolutional Networks,” *arXiv preprint arXiv:1609.02907*.
- LeCun, Y., Bengio, Y. and Hinton, G.(2015), “Deep Learning,” *Nature*, vol. 521, no. 7553, pp.436-444.
- Li, Y., Yu, R., Shahabi, C. and Liu, Y.(2018), “Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting,” *The International Conference on Learning Representations*. Vancouver, Canada
- Lv, Z., Xu, J., Zheng, K., Yin, H., Zhao, P. and Zhou, X.(2018), “Lc-rnn: A Deep Learning Model for Traffic Speed Prediction,” *International Joint Conferences on Artificial Intelligence*, Stockholm, Sweden, pp.3470-3476.
- Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y. and Wang, Y.(2017), “Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction,” *Sensors*, vol. 4, no. 17, p.818.
- Mallick, T., Balaprakash, P., Rask, E. and Macfarlane, J.(2020), “Graph-Partitioning-Based Diffusion Convolution Recurrent Neural Network for Large-Scale Traffic Forecasting,” *Transportation Research Record*, vol. 2674, No. 9, pp.473-488.
- Oord, A. Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A. and Kavukcuoglu, K.(2016), “Wavenet: A Generative Model for Raw Audio,” *arXiv preprint arXiv:1609.03499*.
- C. Park, C. Lee, H. Bahng, K. Kim, S. Jin, S. Ko and J. Choo,(2019), “STGRAT: A Spatio-Temporal Graph Attention Network for Traffic Forecasting,” *arXiv preprint arXiv:1911.13181*.
- Wang, J., Gu, Q., Wu, J., Liu, G. and Xiong, Z.(2016), “Traffic Speed Prediction and Congestion Source Exploration: A Deep Learning Method,” *IEEE 16th International Conference on Data Mining*, Barcelona, Spain, pp.499-508.
- Wu, Z., Pan, S., Long, G., Jiang, J. and Zhang, C.(2019), “Graph Wavenet for Deep Spatial-Temporal Graph Modeling,” *International Joint Conference on Artificial Intelligence*, Macao, China.
- Lv, Y., Duan, Y., Kang, W., Li, Z. and Wang, F. Y.(2014), “Traffic Flow Prediction with Big Data: A Deep Learning Approach,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 2, no. 16, pp.865-873.
- Yu, B., Yin, H. and Zhu, Z.(2018), “Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting,” *International Joint Conference on Artificial Intelligence*, Stockholm, Sweden, pp.3634-3640.
- Yu, H., Wu, Z., Wang, S., Wang, Y. and Ma, X.(2017), “Spatiotemporal Recurrent Convolutional Networks for Traffic Prediction in Transportation Networks,” *Sensors*, vol. 7, no. 17, p.1501.
- Zhang, J., Shi, X., Xie, J., Ma, H., King, I. and Yeung, D. Y.(2018), “Gaan: Gated Attention Networks for Learning on Large and Spatiotemporal Graphs,” *Uncertainty in Artificial Intelligence*, California, USA.
- Zhao, Z., Chen, W., Wu, X., Chen, P. C. and Liu, J.(2017), “LSTM Network: A Deep Learning Approach for Short-Term Traffic Forecast,” *IET Intelligent Transport Systems*, vol. 2, no. 11, pp.68-75.