

GNN을 사용한 도로 네트워크 데이터 분석 방법

오 병 우*

A Method for Analyzing Road Network Data using GNN

Byoung-Woo Oh*

이 연구는 국립금오공과대학교 대학 연구과제비로 지원되었음(2022~2024)

요 약

본 논문에서는 그래프 오토인코더(GAE) 기반의 그래프 신경망(GNN) 모델을 사용한 도로 네트워크 분석 방법을 제안한다. 기존의 중심성 기반 분석 방법은 도로 네트워크의 비선형적 특성이나 복잡한 관계를 충분히 반영하지 못하는 한계가 있다. 본 논문에서는 GAT와 GCN 두 가지 유형의 GNN 레이어를 적용한 GAE 모델을 구축하여 각각 학습시키고 분석을 수행하여 노드 및 엣지의 다양한 속성을 효과적으로 반영할 수 있도록 한다. 도로 네트워크는 방향성을 가진 그래프로 모델링하고, 교차로는 노드로 표현하며 도로는 엣지로 표현한다. 분석에는 노드에 연결된 엣지의 수와 기하 좌표를 노드 특성으로 포함하며, 엣지 특성으로는 도로의 속성인 길이, 차로 수, 최고 속도를 사용한다.

Abstract

This paper proposes a method for analyzing road network data using a Graph Neural Network(GNN) model based on a Graph Autoencoder(GAE). Existing centrality-based analysis methods have limitations in capturing the nonlinear characteristics and complex relationships of road networks. In this paper, GAE models are constructed using two types of GNN layers, Graph Attention Networks(GAT) and Graph Convolutional Networks(GCN), which are trained and analyzed separately to effectively integrate various node and edge attributes. The road network is modeled as a directed graph, where intersections are represented as nodes and roads as edges. The analysis considers the number of connected edges and geometric coordinates as node features, while edge features include road-specific attributes such as length, the number of lanes, and maximum speed.

Keywords

road, network, spatial data, GNN, GAE

* 국립금오공과대학교 컴퓨터공학과 교수
- ORCID: <https://orcid.org/0000-0003-4700-9961>

· Received: Dec. 27, 2024, Revised: Feb. 10, 2025, Accepted: Feb. 13, 2025
· Corresponding Author: Byoung-Woo Oh
Dept. of Computer Engineering
Kumoh National Institute of Technology, South Korea
Tel.: +82-54-478-7531, Email: bwoh@kumoh.ac.kr

I. 서론

도로는 과거로부터 현재까지 인류에게 있어서 매우 중요한 의미를 가진다. 특히, 자동차를 사용하여 이동하는 현대 사회에서는 도로 및 교통에 대한 중요성에 의해 다양한 연구가 수행되고 있다[1]-[4].

도로는 교차로와 도로 구간이 연결되어 있어서 그래프 구조로 표현하기 적합하다. 그래프는 현실 세계의 다양한 시스템을 모델링하기 위한 효과적인 도구로서 소셜 네트워크, 추천 시스템 등 광범위한 분야에서 활용되고 있다.

최근 딥러닝 기술의 발전과 함께, 그래프 신경망(GNN, Graph Neural Network)은 그래프 구조 정보를 효과적으로 학습할 수 있는 강력한 도구로 주목받고 있다[5]. 도로 및 교통과 관련된 분야에서도 GNN을 사용한 연구가 활발히 진행되고 있다[6]-[9].

특히, 그래프 오토인코더(GAE, Graph Auto-encoders)는 GNN을 활용하여 노드의 저차원 임베딩을 학습하고, 이를 통해 링크 예측을 수행하는 효과적인 방법으로 연구되고 있다[10].

본 논문에서는 도로에 해당하는 엣지(Edge) 속성 정보를 활용하여 GAE 기반 링크(Link) 예측을 통해 도로 네트워크를 분석하는 방법을 제안한다. 제안된 방법은 엣지 속성을 GAE 모델의 인코딩 과정에 통합함으로써, 노드 임베딩이 그래프의 구조적 정보뿐만 아니라 엣지의 특징도 반영하도록 한다.

본 논문의 구성은 다음과 같다. 2장에서는 데이터셋과 전처리 방법에 대해 설명하고, 3장에서는 제안된 GNN을 사용한 도로 네트워크 분석 방법을 상세히 설명한다. 4장에서는 도로 네트워크 분석 결과의 시각화에 대해 설명하고, 마지막으로 5장에서는 결론 및 향후 연구 방향을 기술한다.

II. 데이터셋

본 장에서는 도로 네트워크 데이터 분석을 위해 사용한 데이터셋을 전처리하는 방법을 설명한다.

2.1 전국 표준 노드 링크

전국 표준 노드 링크 체계는 교통 정보에 대한 일관성 있는 처리를 위해 국토교통부 국가교통정보센터에서 노드 ID 및 링크 ID를 표준화하여 교통망 DB를 구축한다. 이를 기반으로 도로 및 실시간 교통 정보를 제공할 수 있다.

노드는 교차로, 지하 차도 시작점 및 종료점, 터널 시작점 및 종료점 등과 같이 도로에서 차량의 속도 변화가 발생하는 곳을 표현한다. 링크는 도로, 지하 차도, 터널, 교량 등과 같이 노드와 노드를 연결한 선을 표현한다.

전국 표준 노드 링크 데이터는 국가교통정보센터에서 다운로드할 수 있으며, ESRI사의 Shapefile 포맷으로 배포된다. 공간 데이터의 기하 좌표를 저장하는 노드 shp 파일은 Point 타입이고, 링크 shp 파일은 Line(MultiLineString) 타입이다.

비공간 데이터를 저장하는 노드 dbf 파일은 고유한 10자리 숫자의 노드 ID(NODE_ID), 노드 유형(NODE_TYPE), 교차로 명칭(NODE_NAME) 등으로 구성된다. 링크 dbf 파일은 고유한 10자리 숫자의 링크 ID(LINK_ID), 시작(from) 노드 ID(F_NODE), 종료(to) 노드 ID(T_NODE)를 포함한다. 링크 dbf 파일에는 그 외에도 도로명, 도로 등급, 차로수, 최고제한 속도, 통과 제한 높이 등과 같이 유용한 정보도 포함하고 있다.

2.2 서울특별시 데이터 필터링

전국 표준 노드 링크 데이터는 전국의 모든 도로에 대한 대용량 데이터이므로 분석하기에 시간이 많이 소요되므로, 서울특별시 지역의 데이터를 필터링하여 데이터셋으로 사용한다.

노드 ID 및 링크 ID는 모두 10자리의 숫자로 저장되는데 앞의 3자리 숫자가 지역 코드이다. 지역 코드는 100번 종로구로부터, 101번 중구, 102번 용산구 순으로 124번 강동구까지 25개의 구에 대한 지역 코드가 서울특별시에 해당한다. 예를 들어, NODE_ID가 '1230044700'이면 송파구에 속한 교차로를 의미한다.

전국 표준 노드 링크 데이터로부터 서울특별시 데이터를 필터링하기 위해서는 GeoPandas를 사용한다.

노드 데이터 필터링 방법은 GeoPandas를 통해 읽은 노드 Shapefile로부터 모든 노드를 읽어 들인 후에, 각 노드의 NODE_ID 10자리 숫자의 처음 3자리 숫자가 100 이상이고 124 이하인 노드들만을 선별하여 Shapefile 포맷으로 저장한다.

링크 데이터 필터링은 노드 데이터 필터링과 유사하게 링크 Shapefile로부터 모든 링크를 읽어 들인 후에, 각 링크들의 LINK_ID 10자리 숫자의 처음 3자리 숫자가 100 이상이고, 124 이하인 노드들만을 선별하는 과정으로 시작한다. 도로에 해당하는 링크는 시작 노드와 종료 노드를 가지고 있으므로, 추가적으로 시작 노드와 종료 노드의 처음 3자리 숫자가 모두 100 이상이고 124 이하인 링크를 선별하여 Shapefile 포맷으로 저장한다.

필터링된 노드 데이터는 8,997개이고 링크 데이터는 23,530개이다. 본 논문에서는 이런 과정을 통해 필터링된 데이터셋을 사용하여 도로 네트워크 데이터 분석을 실행한다.

III. 도로 네트워크 데이터 분석

본 장에서는 표준 노드 링크 데이터를 필터링한 데이터셋을 기반으로 도로 네트워크 데이터를 분석하는 방법을 제안한다.

3.1 도로 데이터 로딩 및 그래프 생성

필터링하여 저장한 서울특별시 노드 링크 데이터도 Shapefile 포맷이므로 GeoPandas를 통해 로딩한다. 교차로에 해당하는 노드 데이터는 NODE_ID를 인덱스로 지정하고 점 기하 좌표를 nodes 변수에 저장한다. 도로에 해당하는 링크 데이터는 LINK_ID를 인덱스로 지정하고 라인 스트링 기하 좌표를 links 변수에 저장한다.

로딩된 노드 데이터 및 링크 데이터를 저장하고 있는 nodes 및 links를 사용하여 NetworkX를 사용하여 그래프를 구축한다. 표준 노드 링크 체계의 링크(도로)는 시작 노드(F_NODE)와 종료 노드(T_NODE)를 가지고 있으므로, 방향성이 있는 그래프인 DiGraph로 구축한다.

교차로인 노드를 그래프에 추가할 때는 고유 ID인 node_id 및 기하 점 좌표를 pos라는 이름의 애트리뷰트로 지정하여 인자로 전달한다. 일반적으로 그래프를 구축할 때 엣지를 추가하면 별도의 노드는 추가하지 않아도 자동으로 엣지의 시작 노드와 종료 노드가 추가된다. 그러나, 노드의 기하 좌표는 자동으로 저장되지 않으므로 별도로 pos라는 이름의 애트리뷰트에 기하 점 좌표를 저장하기 위하여 그래프에 노드를 추가하는 과정이 필요하다.

도로인 링크는 그래프에서 엣지로 추가하며 시작 노드인 u, 종료 노드인 v, 도로의 길이를 weight로 지정하고, 추가적인 데이터로 라인 스트링 타입의 기하 좌표들인 geometry, 도로가 몇 차로인지를 나타내는 lanes, 도로의 최고 속도인 max_spd 애트리뷰트를 인자로 전달한다.

3.2 그래프로부터 데이터 생성

생성된 NetworkX 도로 네트워크 그래프를 기반으로 PyTorch Geometric의 데이터 객체를 생성한다. 생성된 객체는 GNN 모델을 만들 때 입력으로 사용할 수 있는 형식이다. 그래프는 방향 그래프 G는 교차로에 해당하는 노드의 집합 V와 도로 구간에 해당하는 엣지의 집합 E로 표현된다.

PyTorch Geometric의 데이터 객체를 생성하기 위해서는 노드 특성, 엣지 인덱스, 엣지 특성이 필요하다. GNN에서 각 노드에 대한 정보를 저장하는 노드 특성, 엣지의 시작 및 종료 노드의 인덱스에 대한 정보를 저장하는 엣지 인덱스, 각 엣지에 대한 정보를 저장하는 엣지 특성은 모델이 학습하는데 필요한 중요한 정보이다.

교차로에 해당하는 각 노드 $v \in V$ 의 특성은 해당 교차로를 시작 또는 종료 노드로 사용하는 도로(엣지)의 개수의 합(degree)과 좌표 정보(x, y)로 구성된다. 각 노드의 도로 개수의 합과 좌표 값을 텐서로 변환한 후 PyTorch Geometric에서 요구하는 특성 벡터 형태로 변경하여 node_features 변수에 저장한다. 모든 노드 특성은 StandardScaler를 사용하여 정규화한다.

도로에 해당하는 각 엣지 $e \in E$ 의 인덱스는 해당 엣지를 연결하는 두 노드의 인덱스를 나타내는 텐서이다. 각 엣지의 시작 노드 u 및 종료 노드 v 에 해당하는 노드들의 인덱스를 리스트로 구성하고, PyTorch Geometric에서 사용할 수 있는 형태로 변환하여 `edge_index` 변수에 저장한다.

각 엣지 $e \in E$ 의 특성은 해당 도로의 차선 수 (lanes 변수), 길이(weight 변수), 최고 속도(max_spd 변수)로 구성된다. 각 엣지의 특성 값을 리스트로 만들고 PyTorch 텐서로 변환한 후 PyTorch Geometric에서 요구하는 특성 벡터 형태로 변경하여 `edge_features` 변수에 저장한다. 모든 엣지 특성은 StandardScaler를 사용하여 정규화한다.

최종적으로 `node_feature`, `edge_index`, `edge_feature`를 인자로 전달하여 PyTorch Geometric의 데이터 객체를 생성한다.

3.3 GNN 모델 생성 및 학습

본 논문에서는 도로 네트워크 분석을 위해서는 GAE(Graph Autoencoder) 기반의 GNN 모델을 사용한다. GAE는 노드 임베딩을 학습하는 인코더와 링크 존재 확률을 예측하는 디코더로 구성된다. 노드 임베딩은 모델이 노드를 이해할 수 있도록 해당 노드 특성이나 그래프에서의 역할을 반영하여 고차원 벡터로 변환한 것이다. 노드 임베딩을 통해 노드 간의 유사도 및 연관성을 수치적으로 계산할 수 있다.

GNN 모델 구조로는 GAT(Graph Attention Network) 및 GCN(Graph Convolutional Network) 두 가지 유형의 GNN 레이어를 구현하였으며 실행할 때 선택하여 적용한다[11][12]. 두 모델 모두 2개의 GNN 레이어를 가지는 GAE 구조를 사용한다.

GAT 기반의 GAE 인코더는 두 개의 GAT 레이어로 구성되며, 각 레이어는 어텐션 메커니즘을 사용하여 이웃 노드의 중요도를 학습한다. 어텐션 계수 a_{vu} 는 식 (1)과 같이 계산된다.

$$\begin{aligned} e_{vu} &= a^T \text{LeakyReLU}(W[h_v || h_u]) \\ a_{vu} &= \exp(e_{vu}) / \sum_{k \in N(v)} \exp(e_{vk}) \end{aligned} \quad (1)$$

GAT 레이어의 출력은 식 (2)와 같다. 엣지 특성

은 완전 연결 계층을 통하여 변환된 후 GAT 레이어의 출력에 더해진다.

$$h_v^{(l+1)} = \sigma(\sum_{u \in N(v)} a_{vu} h_u^{(l)} W^{(l)}) \quad (2)$$

GCN 기반의 GAE 인코더는 두 개의 GCN 레이어로 구성되며, 각 레이어는 식 (3)과 같은 메시지 전달 과정을 수행한다. 엣지 특성은 완전 연결 계층을 통해 변환되고 GCN 레이어의 출력에 더해진다.

$$h_v^{(l+1)} = \sigma(\sum_{u \in N(v)} (h_u^{(l)} \frac{W^{(l)}}{|N(v)|}) + b^{(l)}) \quad (3)$$

디코더는 식 (4)와 같이 내적을 사용하여 링크 존재 확률을 예측한다. 여기서 z_v 와 z_u 는 각각 노드 v 와 u 의 임베딩 벡터이다.

$$P(v,u) = \sigma(z_v^T z_u) \quad (4)$$

생성된 모델의 학습에는 Adam(Adaptive Moment Estimation) 최적화를 사용하며, 손실 함수로는 이진 교차 엔트로피(BCE, Binary Cross-Entropy)를 사용한다. 그림 1은 손실 값 그래프를 보여준다.

그림 1에서 보는 바와 같이 GCN을 사용했을 때 초기 손실 값은 1.585로 시작하여 GAT에 비해 상대적으로 빠른 수렴을 보이며 최종 손실 값 1.125로 일정하게 안정화된다. GAT는 초기 손실 값 1.517로 시작하여 상대적으로 오랜 시간 동안 학습이 진행되고 최종 손실 값 1.066으로 느리게 수렴하지만, 주의(Attention) 기법을 사용하여 더욱 복잡한 관계를 학습하는 데 강점이 있다.

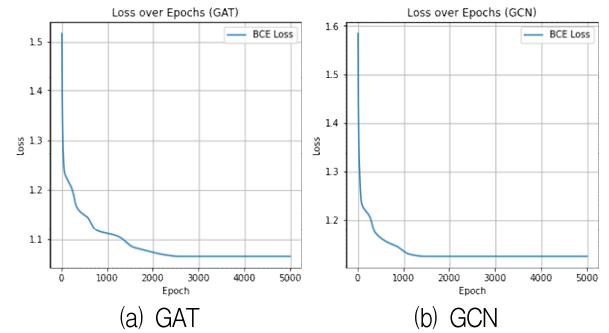


그림 1. 손실 값 그래프

Fig. 1. Loss value graph

3.4 엣지 예측

본 논문에서는 학습된 그래프 임베딩을 활용하여 그래프 내의 엣지 존재 확률을 예측하는 링크 예측 작업을 수행한다. 이는 그래프의 구조적 정보를 기반으로 엣지 연결의 중요도를 분석하고, 누락된 연결을 예측하거나, 미래의 연결 가능성을 예측할 수 있다. 앞에서 설명한 그래프 오토인코더(GAE) 모델을 통해 학습된 노드 임베딩을 활용하여 엣지 예측을 수행한다.

엣지 예측 과정은 다음과 같다. 먼저, 학습된 GAE 모델의 인코더를 통해 각 노드에 대한 임베딩 벡터를 생성한다. 인코더는 노드의 속성 정보와 그래프의 구조적 정보를 입력으로 받아, 저차원 공간에 노드를 표현하는 임베딩 벡터를 출력한다.

그 뒤에는 예측 대상 엣지 쌍을 정의한다. 엣지 예측은 두 노드 간 연결 존재 여부를 예측하는 이진 분류 문제로 볼 수 있다. 따라서 각 엣지 쌍에 대해 긍정 샘플과 부정 샘플을 구성한다. 긍정 샘플은 원본 그래프의 엣지를 사용하며, 부정 샘플은 부정 샘플링 기법을 통해 생성한다.

생성된 노드 임베딩과 엣지 샘플을 이용하여 엣지 존재 확률을 계산한다. 본 논문에서는 디코더를 통해 두 노드 임베딩의 유사도를 측정하여 엣지 존재 확률을 추정한다.

IV. 시각화

도로 네트워크 분석을 위한 GAE 모델은 GAT와 GCN 레이어 두 가지 유형으로 구현하였다. 도로에 해당하는 엣지 특성으로는 도로의 길이, 차로 수, 최고 속도를 사용하여 분석하였다.

그림 2는 교차로에 해당하는 노드 특성으로 방향이 있는 그래프에서의 엣지가 연결된 수의 총합을 사용하여 엣지 존재 확률을 예측하고, 확률 값을 평활화한 뒤에 확률이 높을수록 검은색에 가깝게 출력한 결과를 보여준다.

그림 3은 기존의 노드 특성에 추가적으로 노드의 기하 좌표를 포함하여 엣지 예측 분석을 수행한 결과를 보여준다.

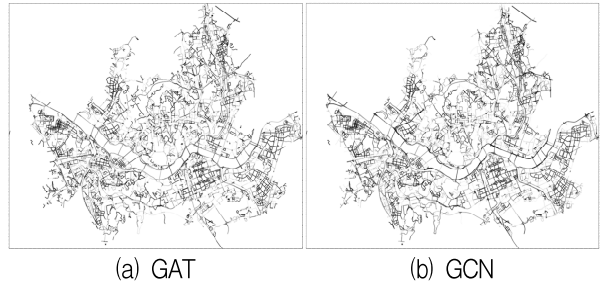


그림 2. 엣지 예측 결과
Fig. 2. Edge prediction result

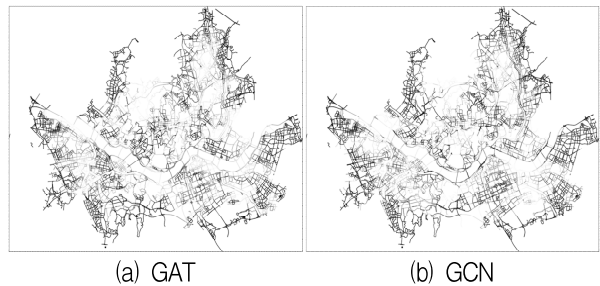


그림 3. 기하 좌표가 포함된 노드 특성
Fig. 3. Node features with geometric coordinates

노드 특성과 엣지 특성을 변경하면서 다양한 도로 네트워크를 분석하는 것이 가능하다.

V. 결론

본 논문에서는 전국 표준 노드 링크 데이터를 활용하여 도로 네트워크를 분석하기 위한 그래프 오토인코더(GAE) 기반의 그래프 신경망(GNN) 모델을 제안하였다. 도로 네트워크에서 도로에 해당하는 엣지가 가지는 속성 정보 중에서 도로의 길이, 차로 수, 최고 속도를 GAE 모델의 인코딩 과정에 통합함으로써, 노드 임베딩이 그래프의 구조적 정보뿐만 아니라 엣지의 특징까지 반영할 수 있었다.

기존의 Betweenness Centrality, Closeness Centrality 등과 같은 네트워크 중심성 기반 분석 방법에서는 교차로에 해당하는 노드의 좌표, 도로의 차로 수 등과 같은 도로 네트워크의 비선형적 특성이나 복잡한 관계를 반영하지 못하지만, 본 논문에서 제안한 방법은 노드 및 엣지의 다양한 속성 정보를 필요에 따라 반영하여 분석할 수 있음을 실험적으로 입증하였다.

GNN 모델로는 GAT 및 GCN 두 가지 유형의 레이어를 활용한 GAE 모델을 구현하였으며, 노드 특성으로 연결된 도로의 개수를 사용한 방법과 기하좌표를 포함한 방법을 사용하여 엣지 예측을 수행하고 그 결과를 시각화하였다.

향후 연구 방향으로는 추가적으로 제공되는 교통 정보, 사고 정보, 날씨 정보 등 더욱 다양한 특성을 활용하여 분석의 범위를 넓힐 수 있을 것이다.

References

- [1] T. Kim and J. Seo, "Traffic Accidents Risk Forecasting based on Deep Learning Models using Spatiotemporal Data Learning", *Journal of KIIT*, Vol. 22, No. 5, pp. 1-12, May 2024. <https://doi.org/10.14801/jkiit.2024.22.5.1>.
- [2] H. Park, "Deep Learning Model for Traffic Speed Prediction Considering Spatial-Temporal Correlation", *Journal of KIIT*, Vol. 20, No. 3, pp. 19-26, Mar. 2022. <http://doi.org/10.14801/jkiit.2022.20.3.19>.
- [3] S. Cho and J. Kim, "Development of a Multi-disciplinary Video Identification System for Autonomous Driving", *Journal of IIBC*, Vol. 24, No. 1, pp. 65-74, Feb. 2024. <https://doi.org/10.7236/JIIBC.2024.24.1.65>.
- [4] N. Bail, G. Kim, and J. Jang, "Design of Continuous Driving Enforcement System for FSORT-based Highway Passing Lane", *Journal of IIBC*, Vol. 24, No. 5, pp. 189-193, Oct. 2024. <https://doi.org/10.7236/JIIBC.2024.24.5.189>.
- [5] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model", *IEEE transactions on neural networks*, Vol. 20, No. 1, pp. 61-80, Dec. 2008. <https://doi.org/10.1109/TNN.2008.2005605>.
- [6] T. Ha, G. Lee, D. Kim, and S. Oh, "Road graphical neural networks for autonomous roundabout driving", In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Prague, Czech Republic, pp. 162-167, Sep. 2021. <https://doi.org/10.1109/IROS51168.2021.9636411>.
- [7] D. Jana, S. Malama, S. Narasimhan, and E. Taciroglu, "Edge-based graph neural network for ranking critical road segments in a network", *Plos one*, Vol. 18, No. 12, Dec. 2023. <https://doi.org/10.1371/journal.pone.0296045>.
- [8] Am. Sharma, As. Sharma, P. Nikashina, V. Gavrilenko, A. Tselykh, A. Bozhenyuk, M. Masud, and H. Meshref, "A graph neural network (GNN)-based approach for real-time estimation of traffic speed in sustainable smart cities", *Sustainability*, Vol. 15, No. 15, pp. 11893, Aug. 2023. <https://doi.org/10.3390/su151511893>.
- [9] G. H. Wesenberg and A. Ozaki, "TØIRoads: A Road Data Model Generation Tool", *Transactions on Graph Data and Knowledge*, Vol. 2, No. 2, pp. 6:1-6:12, Dec. 2024. <https://doi.org/10.4230/TGDK.2.2.6>.
- [10] T. N. Kipf and M. Welling, "Variational graph auto-encoders", *arXiv preprint arXiv.1611.07308*, Nov. 2016. <https://doi.org/10.48550/arXiv.1611.07308>.
- [11] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks", *Advances in neural information processing systems* 30, *arXiv preprint arXiv:1609.02907*, Sep. 2016. <https://doi.org/10.48550/arXiv.1609.02907>.
- [12] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks", *arXiv preprint arXiv:1710.10903*, Oct. 2017. <https://doi.org/10.48550/arXiv.1710.10903>.

저자소개

오 병 우 (Byoung-Woo Oh)



1993년 2월 : 건국대학교

전자계산학과(공학사)

1995년 2월 : 건국대학교

전자계산학과(공학석사)

1999년 2월 : 건국대학교

전자계산학과(공학박사)

1999년 6월 ~ 2004년 2월 :

한국전자통신연구원 텔레매틱스연구단 선임 연구원

2004년 3월 ~ 현재 : 국립금오공과대학교 컴퓨터공학과
교수

관심분야 : 공간 데이터베이스, GIS, 위치 기반 서비스