Road Embedding using GNN to Estimate Walkability

Motivation and background

Walkability, it is a term for planning concepts best understood by the mixed-use of amenities in high-density neighbourhoods where people can access said amenities by foot.[1] It is based on the idea that a modern city street should not only focus on maximum vehicle throughput. Instead it should have a balance for a variety of uses, users and transportation modes and reduce the need for cars for travel. Walking also has the advantage that it is a much more predictable trip time than public transport or cars, where we have to allow for delays caused by poor service, congestion, and parking.[2]

A simple way to determine the walkability of a block, corridor or neighborhood is to count the number of people walking, lingering and engaging in optional activities within a space.[3] This process is a vast improvement upon pedestrian level of service (LOS) indicators, recommended within the Highway Capacity Manual.[4] However it may not translate well to non-Western locations where the idea of "optional" activities may be different.[5] In any case, the diversity of people, and especially the presence of children, seniors and people with disabilities, denotes the quality, completeness and health of a walkable space.[6] There are a couple of commercial walkability scores already existing like WalkScore and RateMyStreet. And they are using 2 different methods to calculate the walkability score which are distance only based and mix based. WalkScore is a walkability index based on the distance to amenities such as grocery stores, schools, parks, libraries, restaurants, and coffee shops.[7] Walk Score's algorithm awards maximum points to amenities within 5 minutes' walk (.25 mi), and a decay function assigns points for amenities up to 30 minutes away.[8] RateMyStreet is a webapp that combines crowdsourcing, Google Maps and a user rating system that allows users to rate their local street.

The existing methods that calculate walkability score between property and amenities are only considering the first order proximity which means that they are only considering the direct distance between property and amenities. As such, there is a need to develop a method that can capture more than first order proximity and try to involve more types of point of interests within a distance range.

Research questions and goals

We are planning to investigate how we can generate a road embedding that can capture both first order proximity and second order proximity, and the generated road embedding space ideally should diversify the types of points of interest. For example, a property on a road that is surrounded by five schools should have a lower walkability score compared to a property on the road that has two schools, two shops and two bars nearby, because the latter one is more convenient to residents.

Our goal is to provide a road embedding that can be used in other downstream tasks such as house price prediction. It should leverage the power of diversity of points of interest in the embedding space to provide more context to the downstream tasks.

Research methods

We are planning to use graph neural networks as the methodology to generate the road embedding. We have reviewed a few existing graph neural networks models like GCN, GAE and GraphSAGE [9, 10, 11]. And we picked GraphSAGE as our base model because the model itself provided an ability to do unsupervised learning without a target label which fits our purpose of generating the embedding. The reason we can not use GraphSAGE directly to generate the embedding is because by default, it is sampling by uniformed distribution so in the embedding space it will not consider the diversity of the types of points of interest. To be able to diverse the types of point of interests in the embedding space, we are proposing a new biassed random walk that will consider a variety types of point of interests during sampling, also at the same time, it will penalise on the neighbours that have repeated point of interests by select them as negative samples. The loss function then will try to push away the negative sampling nodes and the positive nodes in the embedding space so that the roads that have diverse types of points of interest will get embedded closer in the embedding space.

Plan

I have scheduled weekly meetings with my supervisor Kaiqi Zhao every tuesday. The goal of the weekly meeting is to discuss the research ideas, challenges and update progress to Kaiqi.

Week 1- 4 Think through topics and briefly read on the existing papers to find ideas. Align agreement between myself and supervisor.

Week 5 - 9 Data Preprocessing, find open source street data and construct into graphs. Read related papers to get more ideas.

Week 10 - 14 Implement GraphSAGE base model and see the result.

Week 15 - 28 Explore possibility of biassed walk and implementation. Compare the result with the base model.

Week 28 - 33 Wrap up and write the dissertation.

References

1. Dovey, Kim; Pafka, Elek (January 2020). "What is walkability? The urban DMA". *Urban Studies*. **57** (1): 93–108.
2. Robertson, Margaret (2014). Sustainability Principles and Practice. Routledge. pp. ppl: 208–222. ISBN 9780203768747.
3. Gehl, J. and Gemzoe, L. (1996). Public spaces and public life. Copenhagen: Danish Architectural Press

4. Transportation Research Board (2000). *Highway capacity manual: HCM2000*. Washington D.C.: National Research Council
5. Hutabarat Lo, R. (2009). "Walkability: what is it?", Journal of Urbanism Vol. 2, No. 2, pp 145-166.
6. Zehner, Ozzie (2012). *Green Illusions*. Lincoln and London: University of Nebraska Press. pp. 250–51, 265–66.
7. ceosforcities.org, Walking the Walk (2009)
8. ^ Walk Score Methdology
9. T. N. Kipf and M. Welling. SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS. In International Conference on Learning Representations (ICLR) 2017.
10. T. N. Kipf and M. Welling. Variational Graph Auto-Encoders. NIPS Workshop on Bayesian Deep Learning 2016.
11. W. L. Hamilton, R. Ying and J. Leskovec. Inductive Representation Learning on Large Graphs. *arXiv:1706.02216 [cs.SI]*, 2017.