9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Foot Traffic Prediction Using Graph Neural Networks

Keywords: Foot Traffic Prediction, Graph Neural Networks, Human Mobility, Business Site Selection, Customer Demographics

Extended Abstract

Accurately estimating the potential foot traffic for a new business is a crucial task since it directly impacts a business's ability to generate revenue. Therefore, selecting the right location for a new business is critical as it can significantly impact its future financial well-being. However, this is a challenging task given the complex and dynamic nature of human mobility patterns. The site selection process requires careful consideration of factors such as accessibility, demographics, and competition to ensure a business attracts its target customer base and maximizes its earning potential.

Classical spatial methods, such as Huff gravity model[1], make use of a business's attractiveness to customers (e.g., store area) and its distance to customer location to model the probability of a customer visiting a particular business location. However, such methods fail to capture the socio-economic drivers behind human mobility. Recent work focuses on data-driven approaches, in which location based datasets (e.g., mobility) are used to formulate predictive models. For instance, Karamshuk et al. [2] predict the popularity of business locations in different areas based on well-established spatial interaction and mobility features, extracted from a social media platform, to explain visit patterns to target areas. Furthermore, formalizing the foot traffic prediction task with a graph representation has drawn attention due to their ability to capture the complex interactions between nodes and their neighbors. For example, Yeghikyan et al. [3] represent car flows between grids of a city given their socio-economic features as an origin-destination (OD) matrix and use a graph neural network to extract location embeddings based on node neighborhood aggregations. In this study, we leverage large-scale longitudinal mobility data and frame the relationship between neighborhoods and businesses as a directed bipartite network. We predict the visit flux from each neighborhood to a particular business with the help of Graph Neural Networks (GNNs) in a classification setting.

To achieve this we use *SafeGraph Weekly Visit Patterns*[1] dataset, which provides weekly aggregated visits from Census Block Groups (CBG) to businesses also referred to as point-of-interests (POI), collected from mobile phone of users who opted to share their location data through certain applications. In order to obtain a more balanced distribution, we aggregate the visits from CBG level to census tract level that covers larger area and contain more residents. Then we construct a bipartite graph between census tracts and POIs, in which edge weights store the yearly aggregated number of visits from a census tract to a certain POI. For each census tract, we compile a multitude of socio-economic features using 2014-2018 5-years American Community Survey (ACS) data[2], and assign them as node features in addition to its geo-location (i.e., centroid coordinates), POI count and the diversity of POI categories in the area. POIs are represented with their business category embeddings, geo-location, store area in square meters, and the socio-economic features of their home census tract. We assign edge labels based on the quartiles of aggregated number of visits in each business category.

GraphSAGE[4] is an inductive GNN framework, in which a node's local neighborhood is taken into account to generate its embeddings, instead of considering all the nodes in the graph.

[1] `https://docs.safegraph.com/docs/weekly-patterns`
[2] `https://data.census.gov/`

To predict the potential foot traffic, we use a GNN architecture that relies on GraphSAGE census tract and POI embeddings. We use two stacked GraphSAGE layers to extract the node embeddings in a heterogeneous graph setting. Resulting census tract and POI embeddings are concatenated to be fed to three linear layers with *leakyReLU* activation functions in between to obtain the final logit values for each class label.

We evaluate the proposed method in two different urban settings with differing population densities namely: Monroe County, NY (192 census tracts, 6K POIs and 384K edges) and Greater Boston Area (914 census tracts, 33K POIs and 2.8M edges) in year 2018. To account for the sampling bias in the mobility data, we apply post-stratification on the number of visits based on the ratio of observed mobile devices in each time step and the actual population of a census tract. We obtain the train and test sets by randomly splitting the edges of the bipartite graph and train the model in 3000 epochs with *Adam* optimizer and calculate loss with *Cross Entropy Loss*. During the training phase, we also add *Dropout* layers between linear layers to regularize the model. The model yields significant scores of **0.66** and **0.64** AUROC on test sets for Monroe County and Greater Boston Area, respectively.

Such results indicate that our model is suitable for analyzing the total customer flux and demographics of a new business's customer base. We show that the inductive learning of GraphSAGE layers enables our model to generalize embeddings of neighborhood and business node embeddings. As a future research direction, we are investigating new GNN architectures that predict not only the number of visitors a new business would receive but also its impact on the visitor distribution for existing businesses.

References

[1] David L. Huff. A Probabilistic Analysis of Shopping Center Trade Areas. *Land Economics*, 39(1):81–90, 1963.

[2] Dmytro Karamshuk, Anastasios Noulas, Salvatore Scellato, Vincenzo Nicosia, and Cecilia Mascolo. Geo-spotting: mining online location-based services for optimal retail store placement. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 793–801, 2013.

[3] Gevorg Yeghikyan, Felix L. Opolka, Mirco Nanni, Bruno Lepri, and Pietro Liò. Learning Mobility Flows from Urban Features with Spatial Interaction Models and Neural Networks. In *2020 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 57–64, 2020.

[4] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

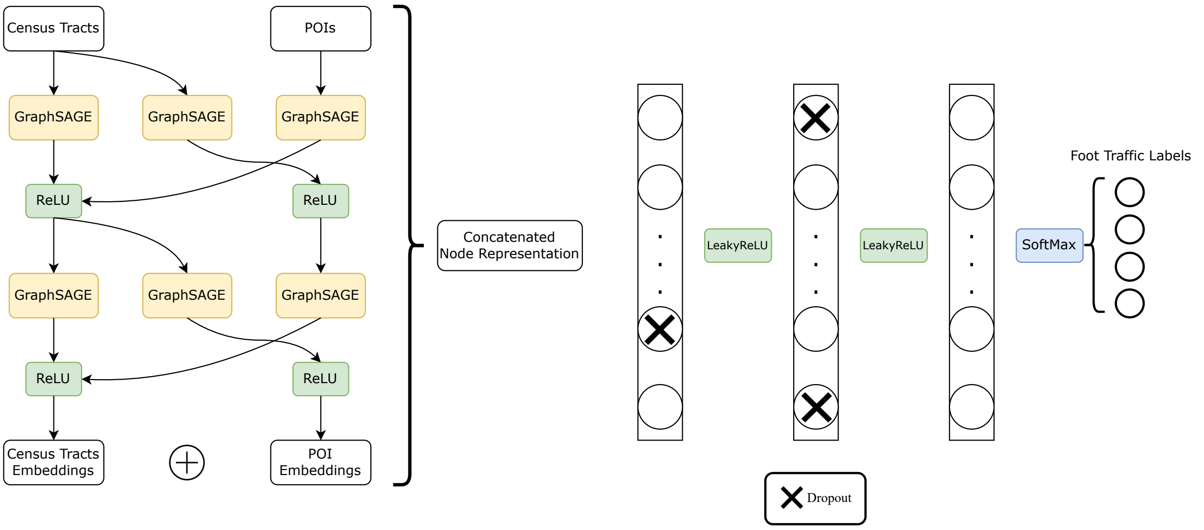

Figure 1: Overview of the graph neural network architecture used to predict the potential foot traffic. GraphSAGE layers learn the embeddings of census tract and POI nodes in the bipartite network. Resulting embeddings are concatenated and fed to stacked linear layers with dropouts in between to perform the final prediction of foot traffic label.