Initial project proposal --Reveal urban dynamics by using deep neural net spatial interaction models

Problem background

Machine learning methods are increasingly used to help solve geospatial issues and many advancements have been made by incorporating state-of-the-art machine learning methods(e.g., deep neural networks). However, obtaining higher predictive accuracy or better classifications does not typically increase our understanding of geospatial processes. Furthermore, the black-box nature of many machine learning techniques can limit geographic knowledge building. As a result,explaining the outcome of machine learning algorithms is a barrier that must be overcome in order to better understand, model, and anticipate the dynamic and diverse activities by cities.

On the other side, spatial interaction models (SI models) are used to facilitate the explanation and prediction of human and economic interaction over geographic space. However, it is beyond the capability of current computing resources to simulate the same with a real city and also abstraction is also not feasible. There are always diversity in geographical contexts that lead to a bias when applying any simple model. So neural net spatial interaction models were introduced to increase the capability of SI models for nonlinear process. The missing parts in explainable machine learning and machine learning implementation in SI models are the motivation of this class project.

Literature

Spatial interaction (SI) can be defined as the movement of individuals, commodities, capital, and information over (geographic) space resulting from a decision process (Fotheringham & O'Kelly, 1989). SI therefore encompasses research into migration, shopping, recreation, commodity and capital flows, communication, transportation networks, and commuting, as well as animal mobility, spatial dependence (e.g., plants competing for sunlight, water, and soil nutrients), and even some biophysical/environmental phenomena.

A recent trend in SI modeling is the use of non-parametric techniques, which means either that no parameters need to be estimated or there are no underlying distributional assumptions, or both. The primary focus of these models is predicting SI and includes neural network SI models and "universal models."

Neural networks (NN) are a computing framework that draw on an analogy to neurons in the brain (Miller, 2009). Given sufficient input data and an appropriate NN structure, it is possible to approximate or "learn" any data-generating process with few a priori assumptions about the data. Openshaw (1993) was the first to propose the use of NNs to model SI flows, which are often noisy, non-linear, and may vary from place to place. Consequently, NN SI models have been used in various contexts, such as predicting telecommunication traffic, journey-to-work trips, and commodity flows (Fischer, 2002; Mozolin et al., 2000) and generally boast higher predictive capabilities than legacy gravity-based SI models.

Task

Task for this project can be concluded as a prototype of deep neural net spatial interaction module which would be compatible with the standards of PySAL Spatial Interaction module, as well as the calibration of the new models. In addition, a pipeline to reveal quantitative spatial-temporal characteristics of urban dynamics will be established through implemented SI models.

Potential sources of data

As the IT technology develops quickly, location information is more accessible and available for urban geography study. For this project, we are anticipating data sources with as little bias to crowd movement as possible so that to be a good representative. Public transit data in metropolitan area is such a good candidate under our expectations. First, there will be more data types available and second, bigger aggregated data amount is available to avoid the cases that data count is too sparse to study spatiotemporal dynamics.

New York City (NYC) is the most populous city in the United States which results in a rich migration flow data generated everyday. Thanks to the government of NYC, multiple data sources of public transportation like Metro, Bus, Taxi and bike sharing data are available for research and other purposes. For the initial data exploration, Taxi trip and Citi Bike trip data are two target datasets fits requirements for calibrating SI models. Both datasets have fields of start and end datetime and location as well as individual’s id. Original data is in .csv file and is easy to be processed by Geopandas Python module.

We collect May, 2019 data for example. First inspection into the datasets shows that there are more than 28 million bike trips and more than 136 million taxi trips recorded in one month period. These rich datasets still contain limitations of data bias of population. We have no idea of if the population distribution of bike or taxi riders is random ‘enough’.

Python tools

Python spatial analysis library (PySAL) has contained spatial interaction modeling (SpInt) module (Oshan, 2016) which can provide calibration for legacy gravity model. Python module for neural network SI models (NNSI) is not released in public domain based on our current knowledge. We are planning to write NNSI code by exploiting prevalent deep learning framework PyTorch and elaborate the outcome to wrap it in the PySAL package.

Envisioned challenges

First challenge is to implement the neural network code which few appeared in the existing literature. Second, the explanation for the neural net, no matter good or bad result, is uncertain in how to proceed.

Exploratory data analysis

o Some simple plots

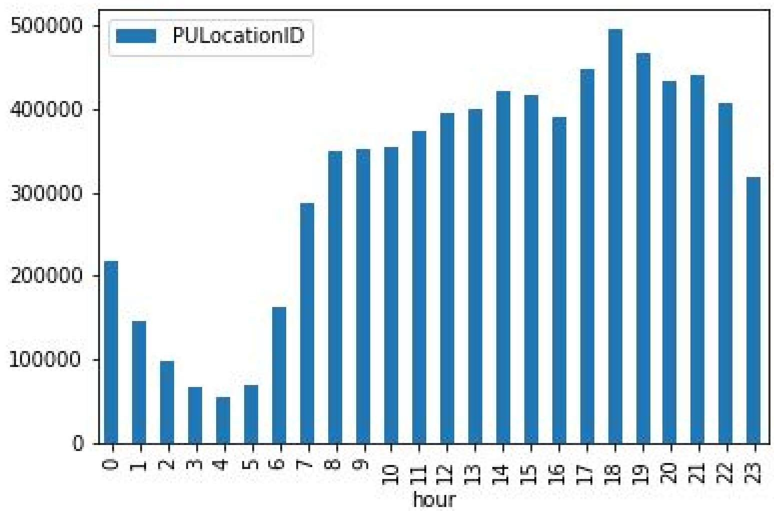

Figure 1 Hourly trip counts

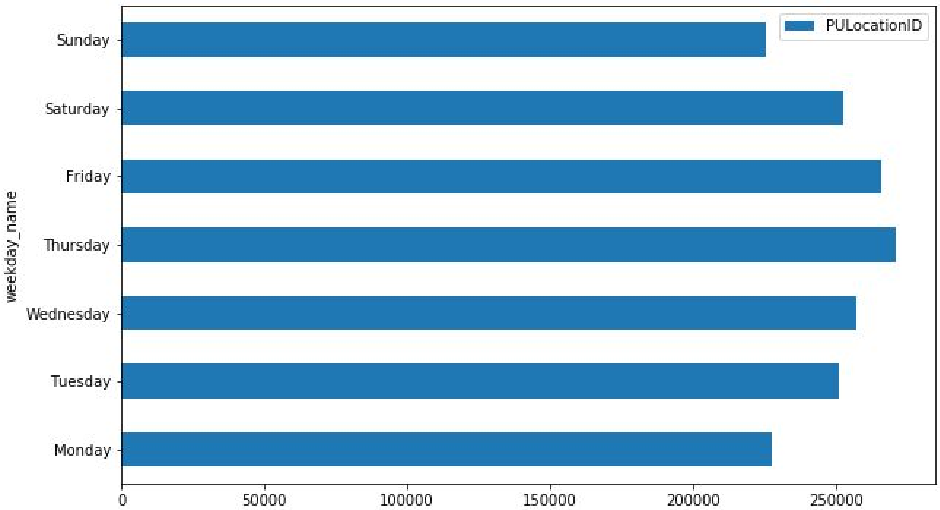

Figure 2 Trip counts by weekday

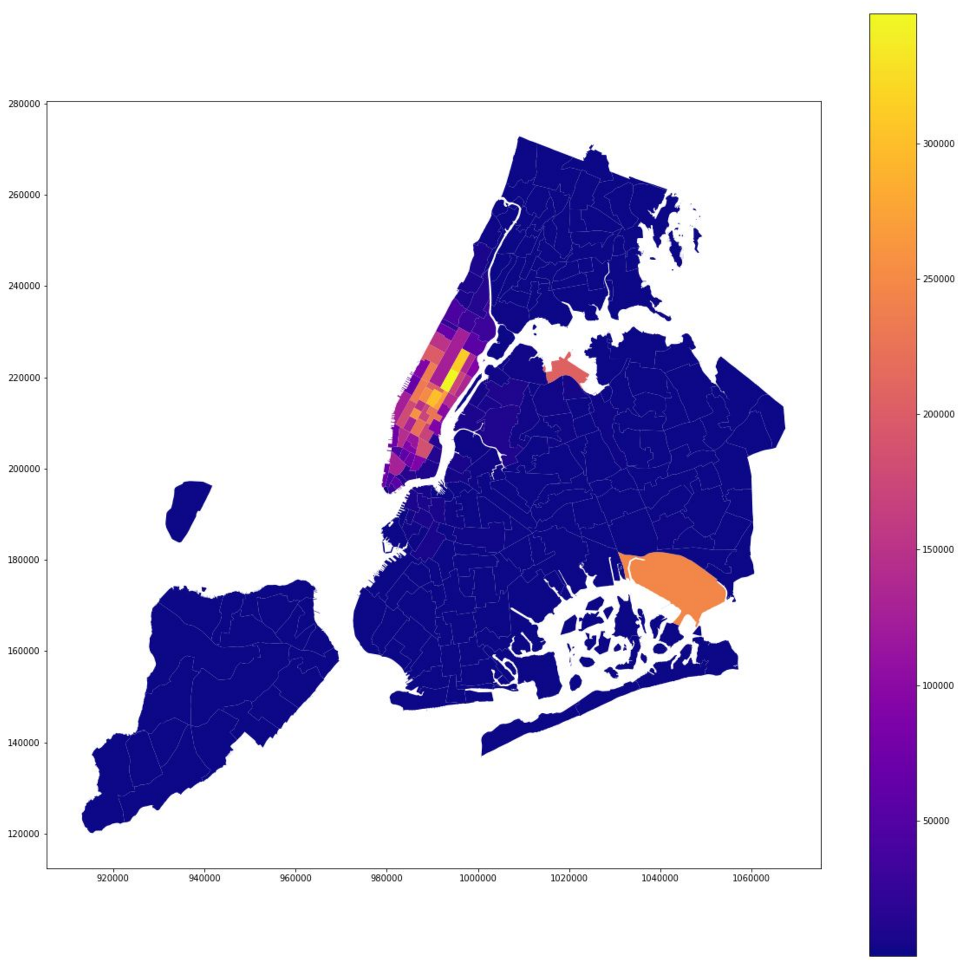

Figure 3 Spatial distribution of all pick-up locations in May, 2019

References:

Fischer, M. M. (2002). Learning in neural spatial interaction models: a statistical perspective. Journal of Geographical Systems, 4(3), 287-299.

Fotheringham, A. S., & O'Kelly, M. E. (1989). Spatial interaction models: Formulations and applications. Dordrecht: Kluwer Academic Publishers.

Miller, H. J. (2009). Geocomputation. In Fotheringham, A. S. and Rogerson, P.A. (Eds.) The SAGE Handbook of Spatial Analysis, (pp. 398-418). SAGE Publications, Ltd.

Mozolin, M., Thill, J. C., & Lynn Usery, E. (2000). Trip distribution forecasting with multilayer perceptron neural networks: A critical evaluation. Transportation Research Part B: Methodological, 34(1), 53-73.

Openshaw, S. (1993). Modelling spatial interaction using a neural net. In M. M. Fischer, & P. Nijkamp (Eds.) Geographic Information Systems, Spatial Modelling and Policy Evaluation, (pp. 147-164). Berlin, Heidelberg: Springer Berlin Heidelberg.

Oshan, T. (2016). A primer for working with the Spatial Interaction modeling (SpInt) module in the python spatial analysis library (PySAL). REGION. 3, 2 (Dec. 2016), R11-R23.