Taxi Trajectory Prediction via Swin-Transformer

Whitegive: Zepu Wang, Xincheng Zhu, Zhiyu Lei, Yifei Sun

March 13, 2023

Project Statement

Precise trajectory prediction of vehicles is a challenging and important task for modern traffic systems. Taxis are one of the most popular travel tools in large cities, and accurately predicate their destinations can benefit a lot of location-based services and applications. For instance, it is advantageous for early warning of potential public emergencies, accurate location-aware advertisements for business companies, improving operations and enhancing the customer experience for taxi companies, assessing the future motion of surrounding vehicles for Autonomous cars, and so forth.

However, most traditional approaches like RNNs and MLP which are used by Brebisson et al[1] or Markov model which adopts in [2][3], take the trajectory prediction as a single-dimension prediction problem. These methods always have difficulties in capturing complex nonlinear spatial-temporal correlations, which means it's hard to predict trajectories that are highly related to real road situations like corners or windings.

To incorporate the spatial information contained within a trajectory, it is common for researchers to transform the trajectory into a two-dimensional grid [5]. This technique is demonstrated in figure 1, wherein the trajectories are projected onto a two-dimensional image (figure 1, Left). The utilization of a grid is a popular approach when analyzing trajectories across varying scales, as it permits the division of the map into cells, with spatial points belonging to the same cell aggregated into a single unit. Figure 1, right displays a micro-spatial scale with a dense grid, which records trajectories with enhanced granularity. By converting trajectories into images, multiple computer vision techniques, including convolutional neural networks (CNNs), can be implemented to extract meaningful patterns and features from the data.

With the advent of Vision Transformer, the Swin Transformer has demonstrated impressive performance in various computer vision tasks. However, due to its novelty, it has not been applied to the problem of trajectory prediction. This project aims to explore the feasibility of replacing convolutional neural networks

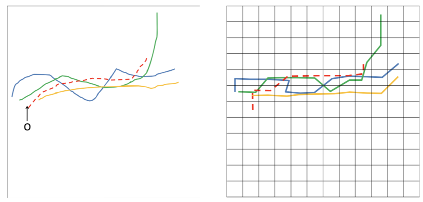

Figure 1: visualization of turning trajectory into 2-d grids

with the Swin Transformer (or its components) to test its applicability to this task.

Dataset

In this project, the ECML-PKDD competition's real taxi trajectory dataset is employed, consisting of data from 442 taxis operating in Porto city over a complete year, from 2013-07-01 to 2014-06-30. The dataset comprises 1.7 million complete observations, with each observation representing a full taxi trajectory. For the purposes of this study, only one attribute is utilized - the POLYLINE - which is a sequence of GPS positions that provide latitude and longitude coordinates measured every 15 seconds. An example of the data is provided as follows: [[-8.618643, 41.141412], [-8.618499, 41.141376], [-8.620326, 41.14251], [-8.622153, 41.143815], [-8.623953, 41.144373], [-8.62668, 41.144778], [-8.627373, 41.144697], [-8.630226, 41.14521], [-8.632746, 41.14692], [-8.631738, 41.148225], [-8.629938, 41.150385], [-8.62911, 41.151213], [-8.629128, 41.15124], [-8.628786, 41.152203], [-8.628687, 41.152374], [-8.628759, 41.152518], [-8.630838, 41.15268], [-8.632323, 41.153022], [-8.631144, 41.154489], [-8.630829, 41.154507], [-8.630829, 41.154516], [-8.630829, 41.154498], [-8.630838, 41.154489]]. Each trajectory contains 23 pairs of coordinates, with the first 22 pairs representing the trajectory, and the last pair representing the destination.

Given the computational demands of the task, the goal is to use as much data as possible for training. However, due to the computational efficiency, we might not able to do it. Hence, we want o try at least 50000 data samples will be used for training purposes.

Methods

For this project, we chose RNNs as our baseline since trajectory prediction falls under the category of time series prediction. Additionally, we will employ CNNs as another baseline to predict the trajectory.

In this stage, our dataset is pretty simple; hence, we do not think data hypothesis and inductive biases/priors are important in this proposal. Hence, we would like to explain more about how we are going to process the data. Every input will be an image that shows the trajectory of each taxi. The output will be a pair of coordinates that represent the final destination. Therefore, it is a supervised learning problem. First, we will define each image's size according to all the cabs' activity ranges. Next, each image will be divided into M*M grids, and M is set to be variable according to each taxi's trajectory range. If the cab passes through the grid, we set the value not to be 0. the early stages of the trajectory will be assigned a relatively low value. the last stages of the trajectory will be assigned a relatively high value. Hence, we can store the time information in the image.

Moreover, we will test Swin Transformer for this project. However, unlike traditional images who have 3 channels, trajectory image only has one channel. At the same time, many resolutions of trajectory images are set to be 0. Hence, we suppose that the traditional Swin Transformer might be too complex. Hence, in this project, we will work on a simplified Swin Transformer to tackle the problem.

Innovation

First, our innovation lies in treating a time series problem as an image processing challenge. we accomplish this by representing each trajectory data as an image and dividing each observation into an MxM grid. The combination of using a time-series prediction approach and image processing approach enables us to capture spatial attributes more effectively than only using time-series prediction methods.

Second, the Swin Transformer represents a novel technique in the field of computer vision. To the best of our knowledge, this study marks the first attempt by researchers to employ the Swin Transformer in the problem of trajectory prediction.

Acknowledgement

Thanks Chatgpt for helping us write this proposal.

References

[1] de Brébisson, Alexandre, et al. "Artificial Neural Networks Applied to Taxi Destination Prediction." ArXiv:1508.00021 [Cs], 21 Sept. 2015, arxiv.org/abs/1508.00021. Accessed 13 Mar. 2023.

[2] Xue, Andy Yuan, et al. “Destination Prediction by Sub-Trajectory Synthesis and Privacy Protection against Such Prediction.” IEEE Xplore, 1 Apr. 2013, ieeexplore.ieee.org/document/6544830. Accessed 13 Mar. 2023.

[3] Li, Xiang, et al. “T-DesP: Destination Prediction Based on Big Trajectory Data.” IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 8, 1 Aug. 2016, pp. 2344–2354, ieeexplore.ieee.org/document/7434640, https://doi.org/10.1109/TITS.2016.2518685. Accessed 13 Mar. 2023.

[4] https://www.kaggle.com/c/pkdd-15-predict-taxi-service-trajectory-i [5] Lv, Jianming, et al. ”T-CONV: A convolutional neural network for multi-scale taxi trajectory prediction.” 2018 IEEE international conference on big data and smart computing (bigcomp). IEEE, 2018.