Summer Reading 7: PreTraM: Self-Supervised Pre-training via Connecting Trajectory and Map

Social Robot Navigation Project @ Bot Intelligence Group

Paper:

PreTraM: Self-Supervised Pre-training via Connecting Trajectory and Map
Chenfeng Xu[1]*, Tian Li[3]*, Chen Tang[1], Lingfeng Sun[1], Kurt Keutzer[1],
Masayoshi Tomizuka[1], Alireza Fathi[2], and Wei Zhan[1]

Summary:

Abstract

The scarcity of trajectory data inhibits the data-hungry deep-learning models from learning good representations. While mature representation learning methods exist in computer vision and natural language processing, these pre-training methods require large-scale data.

To work around the scarcity of trajectory data, the authors resort to another data modality closely related to trajectories - HD maps.

The authors propose PreTraM. PreTraM is a self-supervised pre-training model that connects trajectories and maps for trajectory forecasting.

Introduction

Data-driven supervised learning dominated the current deep learning models for trajectory forecasting. However, both the collection and the annotation of trajectory data in supervised learning are extremely difficult and costly. This complexity limits the scale of the data.

The scarcity of trajectory data prohibits the models from learning good trajectory representation, which restrains their performance when trained with such a small amount of data.

The authors propose PreTraM. PreTraM jointly pre-trains the trajectory and map encoders of a trajectory prediction model in 2 ways:

1. Trajectory-Map Contrastive Learning (TMCL)
    - Trajectories and maps are projected to a shared embedding space with cross-modal contrastive learning.
    - Trajectories are contrasted with corresponding map patches to enforce the model to capture their relationship.
    - The TMCL objective teaches the model to encode the relationship between maps and trajectories into the representation. By capturing the relationship, the trajectory embedding contains the information of the underlying map conditioned on the input trajectory, which implies the geometric and routing information of the future trajectories for the predicto
2. Map Contrastive Learning (MCL)
    - Map representation is enhanced by contrastive learning on large quantities of HD maps.
    - The authors train a stronger map encoder with contrastive learning on large quantities of trajectory-decoupled map patches, which outnumber the agent-centric ones by 782x.

- MCL is conducted on map encoder using large batch size on trajectory-decoupled map patches, where there are not necessarily agent trajectories.
- Through MCL, at each training iteration, the authors randomly crop N map patches from a random subset of HD-Maps in the dataset.

PreTraM is a synergy of TMCL and MCL.
TMCL benefits trajectory representation by understanding trajectory-map relationship.
MCL further enhances that understanding by improving map representation.

Background

Contrastive Learning

- Contrastive learning is a powerful method for self-supervised representation learning.
- Contrastive Learning pulls the semantically-close neighbors together and push away non-neighbors.

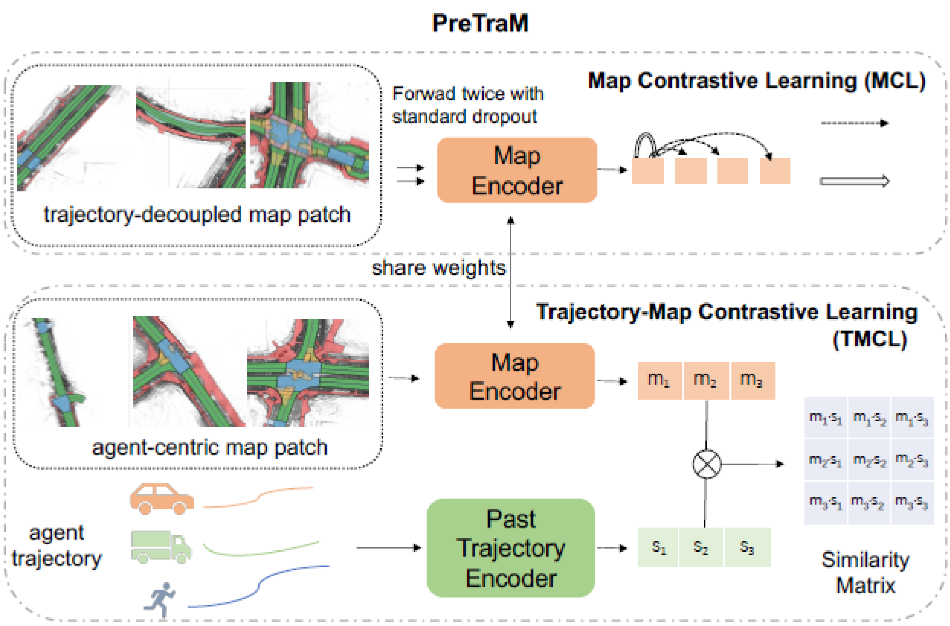

Fig. 2: Top: Map Contrastive Learning (MCL). On the contrary to agent-centric map patches, trajectory-decoupled ones do not necessarily contain agent trajectories. During training, we randomly crop those patches from the whole map around positions on the road. Bottom: Trajectory-Map Contrastive Learning (TMCL).

Method

The authors propose a novel self-supervised pre-training scheme by connecting trajectory and map (PreTraM) to enhance the trajectory and map representations when there is small-scale trajectory data, but large-scale map data.

The authors jointly pre-train a trajectory encoder and a map encoder to obtain good trajectory representation by encoding the trajectory-map relationship into the representation.

After pre-training, the authors load the pre-trained weights and then fine tune the pre-trained model under the prediction objective with the same training schedules as the original models.

Experiments

The authors used the nuScenes dataset.

- nuScenes is a recent large-scale autonomous driving dataset collected from Boston and Singapore. It consists of 1000 driving scenes with each scene annotated at 2Hz.
- nuScenes provides 11 semantic classes.

The authors performed experiments with Pre-TraM on two models, AgentFormer and Trajectron++. Both of them are CVAE models including a past trajectory encoder, a map encoder, a future trajectory encoder, and a future trajectory decoder.

Pre-traing: Pre-training is applied to the past trajectory encoder and map encoder.

1. To train TMCL, the authors pair the historical trajectories of last 2s and map patches of context size 100100.
2. The authors randomly rotate the trajectories and maps simultaneously for data augmentation.
3. For MCL, the authors collect the trajectory-decoupled map patches dynamically at training. For each instance in the mini-batch, the authors crop 120 map patches centered at random positions along the road in the HD-map.

The authors use the common trajectory prediction metrics:

- Average Displacement Error (ADE)
- Final Displacement Error (FDE).

The authors also leverage the metrics including Kernel Density Estimate-based Negative Log Likelihood (KDE NLL) and Boundary Violation Rate.

- KDE NLL measures the NLL of the ground truth trajectory under a distribution created by fitting a kernel density estimate on trajectory samples, which shows the likelihood of the ground truth trajectory given the sampled trajectory predictions.
- Boundary Violation Rate is the ratio of the predicted trajectories that hit road boundaries.

Conclusion

The authors design Trajectory-Map Contrastive Learning (TMCL) to help models capture the relationship between agents and the surrounding HD-map, and Map Contrastive Learning (MCL) to enhance map representation via a large number of augmented map patches that are not associated with the agents. With PreTraM, the authors enhance the prediction performance of Trajectron++ and AgentFormer on FDE-10 on the challenging nuScenes dataset.

Glossary: