Learning a Pedestrian Social Behavior Dictionary

Faith Johnson, Professor Kristin Dana
faith.johnson@rutgers.edu

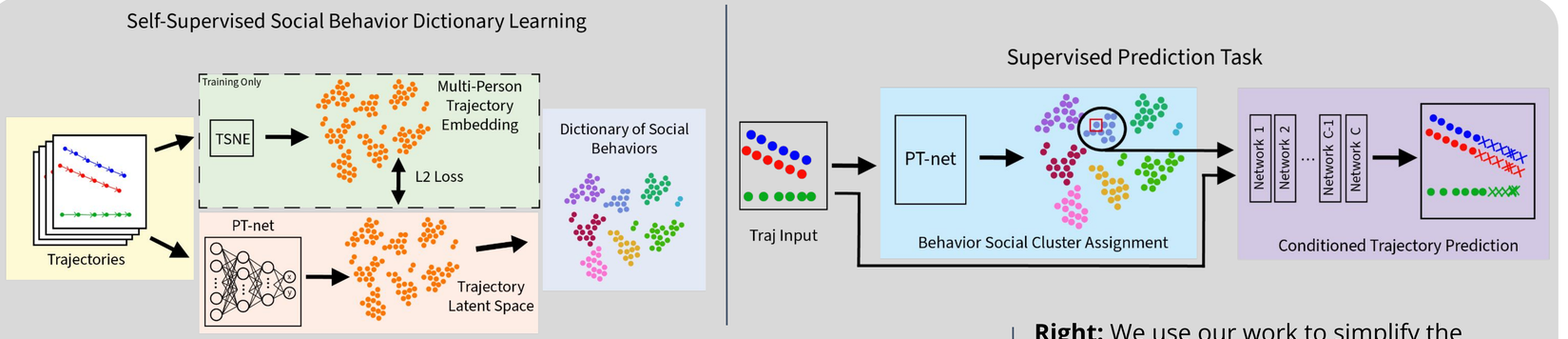
RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

Goal

- Create an **interpretable dictionary** that **clusters similar pedestrian social behaviors** based on historical positions in an **unsupervised** manner
- Use this dictionary to **simplify** downstream tasks
- **Answer questions** about social behavior and environmental characterization

Above: Given historical positions for a set of pedestrians, we identify the distinct, semantic social behavior they exhibit, and use it to answer key questions about the environment.

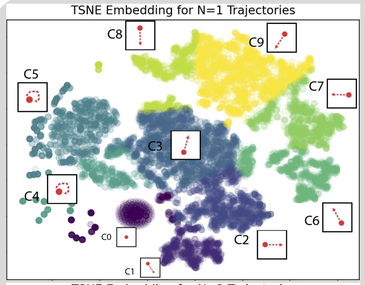
Methods



Left: To create our dictionary, we first process the raw trajectory data to get social features for each group of pedestrians. Then, we use t-SNE to create an embedding space from these features that clusters similar social behaviors together. We input those features into our network, PT-net, to learn a stable version of the t-SNE embedding. Finally, we k-means cluster the output of PT-net to get a set of distinct social behaviors for each group of pedestrians. These behaviors can be easily labeled with relatively few samples to give the embedding semantic meaning.

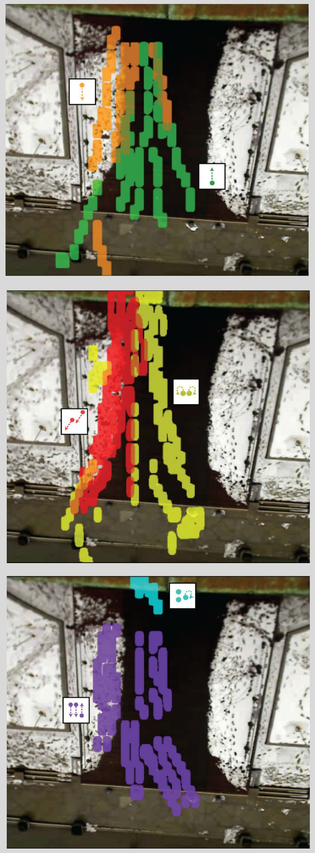
Right: We use our work to simplify the problem of pedestrian trajectory prediction. First, we use PT-net to find the social behavior cluster of a set of trajectories. Given that cluster, we choose a simple MLP network for the prediction from an ensemble, thereby deterministically conditioning our pedestrian trajectory prediction.

Results



We show the learned trajectory embeddings for groups of N=1, 2, & 3 pedestrians. Each cluster is a different color and accompanied by an illustration of its respective behavior.
These were hand labeled using relatively few samples.

There are multiple clusters that contain the same semantic behavior executed in a different direction or with a different pedestrian order due to the nature of the social trajectory features used as input.

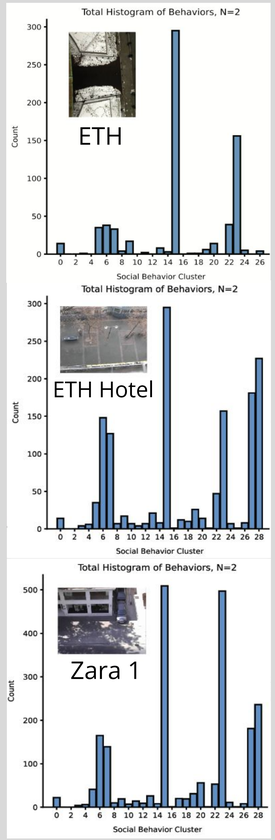


We also use PT-net to answer questions about the environment through social behavior heatmaps. Each colored region corresponds to a distinct semantic behavior (which may be a combination of clusters) from the social behavior dictionary.

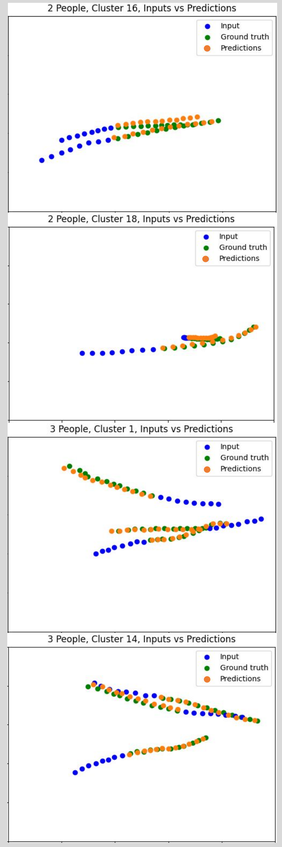
Top: People walking out of the building (orange) give right of way to people walking into the building (green).

Middle: Two people standing still (yellow) form a gap for two people exiting the building leader -follower (red), and generally try to stand out of the way.

Bottom: There is more space for movement away from the door where two people leave the building passing one person entering (purple). Near the door, there is more congestion as shown by three people stand still (blue).



The predicted pedestrian behavior histograms for N=2 people show that pedestrians in ETH (**top**) mainly exhibit horizontal leader-follower behaviors (C15,C23) allowing them to enter/exit the building. The other environments allow for more diversified movement like group congregating (C6,C7), walking side-by-side upwards (C28), and walking leader-follower diagonally downwards (C27). ETH Hotel (**middle**) is predisposed towards side-by-side vertical behavior because it has a train stop at the top of the frame, while UCY Zara1 (**bottom**) has more horizontal leader-follower or side-by-side behaviors because the sidewalk in front of the building is a more popular avenue than the alley at the edge of the frame.



We use our method to simplify the problem of pedestrian trajectory prediction. These example predictions come from four different social clusters. The input to the model is in blue, the ground truth is in green, and the prediction is in orange.

By using PT-net to deterministically condition our prediction, we drastically limit the state space of the problem, allowing very simple MLPs to achieve comparable results to the more complicated SOTA.