Sam Park

Summer Reading 1:
Social LSTM: Human Trajectory Prediction in Crowded Spaces

Social Robot Navigation Project @ Bot Intelligence Group

Paper:
Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, Silvio Savarese. Social LSTM: Human Trajectory Prediction in Crowded Spaces: https://cvgl.stanford.edu/papers/CVPR16_Social_LSTM.pdf

Sam Park

Summary:

Abstract

Autonomous vehicles navigation is to predict the future trajectory of people based on their past positions. Human trajectory prediction can be seen as a sequence generation/prediction task. The authors suggested using a Social LSTM (long short-term memory) model that is able to jointly predict their future paths from general human movements (common sense rules, social conventions, etc.).

LSTM networks have shown success in learning and generalizing the properties of isolated sequences like handwriting and speech. The authors developed an LSTM model for human trajectory prediction, and it is called Social-LSTM.

Introduction

Modeling common sense rules and social conventions that humans follow while walking in a crowded public space is valuable but challenging.

1. A model that considers static environments (sidewalks, grass areas, etc.) predicts the trajectory of pedestrians in future instants more accurately than when it did not.
2. Some works proposed that human-human interactions (social forces) can increase robustness and accuracy in multi-target tracking problems.

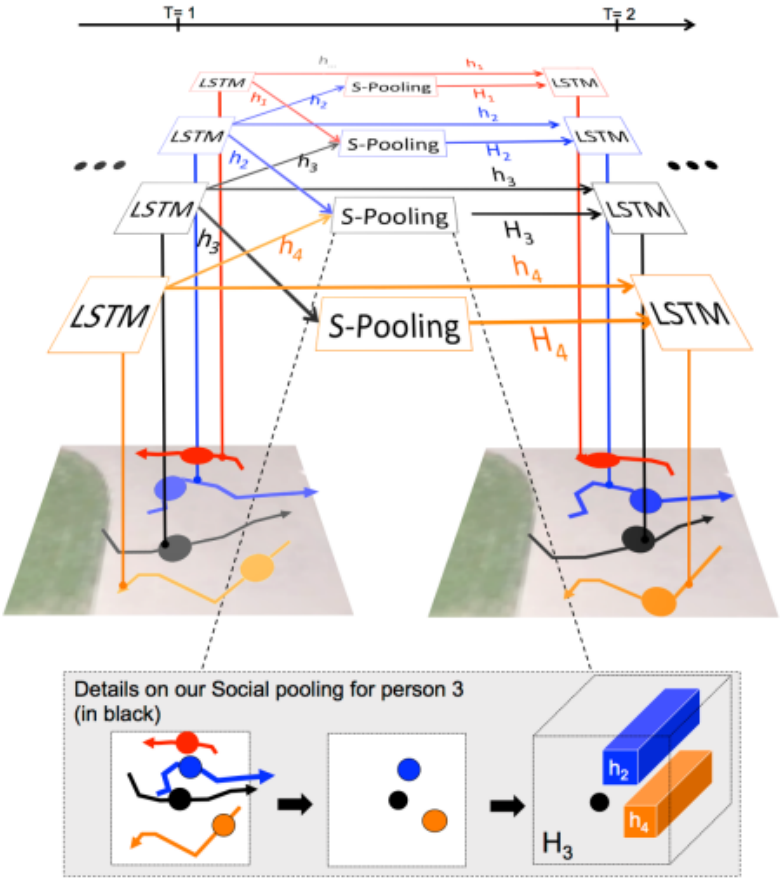
Problem: These works are usually hand-crafted and are limited to simple interactions.
Proposal: Social LSTM is a more generic data-driven model that can automatically learn general interactions (common sense rules, social conventions, etc.) without any additional annotations to learn.

Related Work

1. Human-human interactions
    - a pedestrian motion model with attractive and repulsive forces (traditional social forces)
    - Similar approaches have been mostly hand-crafted.
2. Activity Forecasting
    - Learns motion patterns via clustering trajectories.
    - Works have been mostly restricted to using static scene information to predict human motion and activity.
3. RNN (Recurrent Neural Networks) models for sequence prediction
    - proven to be very successful for sequence prediction tasks.
    - proven to be effective for tasks with densely connected data.
    - capable of learning the dependencies between spatially correlated data (e.g. image pixels).

Authors' Model

- Accounts for behaviors of people within a large group during human trajectory prediction.
- Obtains spatial coordinates (x^i_t, y^i_t) of all people (i for each person) from a scene at different time instants t from time t = 1 to t = T_observe, and then predicts their positions for the next time instants from time t = T_observe+1 to t = T_predict.
- One separate LSTM is used for each person in a scene to predict its future positions.
- Since separate LSTM for each person does not capture interactions of people in a neighborhood, the authors connected separate LSTMs through a Social pooling layer.



- The hidden state h^i_t (**I think:** factors from neighbors that change a person to alter its path) of the LSTM at time t captures the latent representation of the ith person in the scene at that instant. Then, this representation is shared with neighbors.

Experiments

- Used 2 publicly available human trajectory datasets, ETH and UCY.
- Calculated prediction error based on 3 different metrics: MSE (mean squared error), final displacement error, and MSE at non-linear regions of a trajectory.
- At 0.4 fps, the authors observed 8 frames of a scene for 3.2 seconds to predict the next 12 frames equivalent to 4.8 seconds.

Conclusion

- The authors' Social LSTM model outperformed other methods on 2 public human trajectory datasets, ETH and UCY.
- Future work will be extending the model to multi-class settings to account for various objects (e.g. bicycles, skateboards, etc.) and including local static-scene images to jointly model human-human and human-space interactions.

Sam Park

Glossary:

LSTM (Long Short-Term Memory)

- Long short-term memory is an artificial neural network used in the fields of artificial intelligence and deep learning. Unlike standard feedforward neural networks, LSTM has feedback connections. Such a recurrent neural network can process not only single data points, but also entire sequences of data.

MTT (Multi-Target Tracking)

RNN (Recurrent Neural Networks)

MSE (Mean Squared Error)