Summer Reading 6: DAG-Net: Double Attentive Graph Neural Network for Trajectory Forecasting

Social Robot Navigation Project @ Bot Intelligence Group

Paper:

Alessio Monti, Alessia Bertugli, Simone Calderara, Rita Cucchiara. DAG-Net: Double Attentive Graph Neural Network for Trajectory Forecasting: https://arxiv.org/abs/2005.12661v2

Summary:

Abstract

The authors propose a DAG-Net (Double Attentive Graphical Neural Network), a recurrent generative model that **considers both single agents' future goals and interactions between different agents.** The model uses a double attention-based GNN (graph neural network) to collect information about the mutual influences among different agents and to integrate it with data about agents' possible future objectives.

Introduction

Human trajectory forecasting is difficult because human motion is inherently multi-modal (when moving, people may follow several plausible trajectories, as there is a rich distribution of potential human behaviors).

Another challenge is social interaction. Interactions heavily impact on future trajectories (because people plan their paths by reading each other's future possible behaviors, each person's motion is influenced by the subjects around them).

The backbone of DAG-Net is integrated with a double Graph Neural Network (GNN)-based mechanism:

1. The first GNN defines the future objectives of each agent in a structured way, distilling each goal with proximity knowledge;
2. The second GNN models agents' interactions, filtering the hidden states of the recurrent network through neighborhood information.

Related Work

- Social Pooling
    - Fails to capture the global context, as it does not allow the model to consider the interactions between all the possible agents inside the scene in a computationally efficient manner
- Social GAN (Generative Adversarial Network)
- Social Ways
- STGAT

Method

DAG-Net leverages two graph attentive networks to model 2 different kinds of interactions:

1. the interactions between agents

    a. To effectively capture the coordination between the different subjects in the scene, DAG-Net shares goals information among agents relying on group interactions.
2. the relationships between future goals.
    a. To model the interactions between the different agents in the scene, the authors' model uses a graph based approach. Each agent is connected to the others as a node of a graph where edges weights are defined by a self attention mechanism.

Experiments

The authors' model is evaluated with respect to two error metrics on prediction results:

1. Average Displacement Error (ADE) to compute mean euclidean distance over all estimated points at each time step in the predicted positions and ground-truth positions (true positions)
2. Final Displacement Error (FDE) to compute the mean euclidean distance between the final predicted positions and the final ground-truth positions

The authors used *Stanford Drone Dataset* and *STATS SportVU NBA Datase*t.

- *Stanford Drone Dataset* is composed of a series of top-down videos recorded by a hovering drone in 8 different college campus scenes. This large scale dataset collects complex and crowded scenarios with various types of interacting targets: apart from classic pedestrians, we can also find bikes, skateboarders, cars, buses, and other vehicles, therefore the navigation inside such environments results particularly tough.
- *SportVU NBA Dataset* contains tracking positions from the 2016 NBA regular season on a span of over 1200 different games

Conclusion

The authors propose a novel architecture called DAG-Net, a double graph-based network that deals with both past interactions (between agents) and future goals through attentive mechanisms.

The authors' model shows impressive results also on long-term predictions.

Glossary: