The 2nd International Conference on Applied Mathematics, Modeling and Computer Simulation (AMMCS 2022)
August 13-14, 2022 Wuhan, China

ST-AGNN: Spatial-Temporal Attention Graph Neural Network for Pedestrian Trajectory Prediction

Yonghong LI, Jiayi CUI, Zhiqiang ZHAO, Laquan LI Paper ID: CS346
Chongqing University of Posts and Telecommunications, 400065, Chongqing, China.
E-mail: liyh@cqupt.edu.cn.

Abstract

Accurate and fast prediction of pedestrian trajectory is very important. However, the interaction between pedestrians is complex, pedestrians are affected not only by their own motion but also by the neighboring pedestrians. The Spatial-Temporal Attention Graph Convolutional Network (ST-AGNN) for pedestrian trajectory prediction is proposed in this paper. ST-AGNN model can extract the spatial interaction features of each time step and stack the spatial features to obtain the spatial-temporal features. It can predict the future trajectory by using Temporal Convolutional Network (TCN).

We evaluate the average displacement error (ADE) and final displacement error (FDE) of the proposed method on ETH and UCY datasets. The experimental results show that the proposed method is effective in accuracy and efficiency.

ST-AGNN Model

As shown in Figure 1, ST-AGNN consists of three main parts, including A-GNN, spatial-temporal fusion and TCN module. Finally, TCN is used to predict the future trajectory.

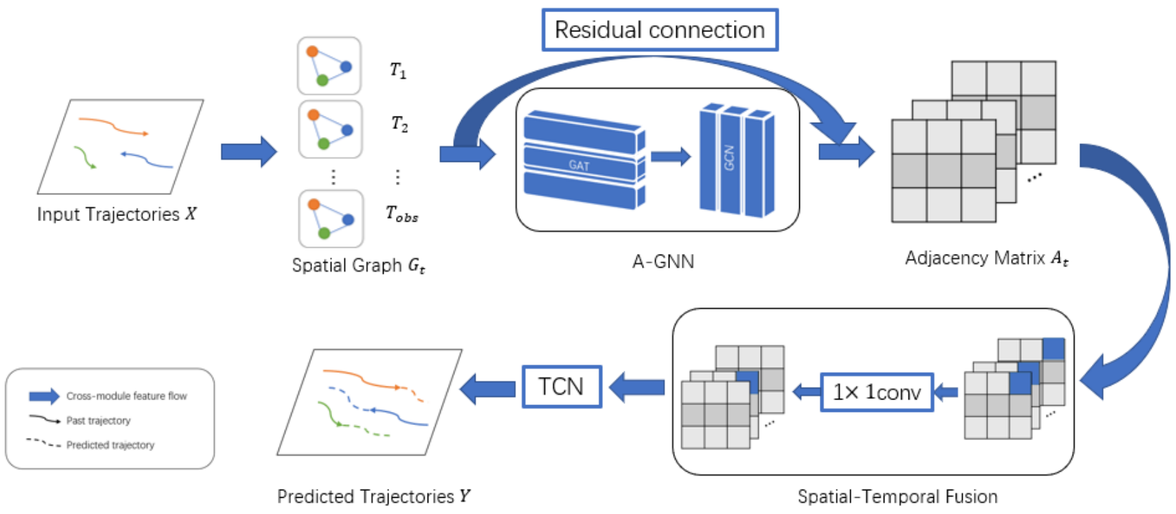

Figure 1. overall structure of ST-AGNN model.

1. Spatial Graph Representation

Firstly, we construct a spatial graph for the input trajectory. Let $t \in \{T_1, \dots, T_{obs}\}$ represent the set of nodes of N pedestrians in the scene and spatial graph $G_t = \{V_t, E_t\}$ represent the spatial interaction between pedestrians at time step t.

2. A-GNN Module

Then, the spatial interaction of pedestrians is encoded in the A-GNN module.

① GAT Module: In the GAT module, we get the features of each node in the graph and construct a weighted adjacency matrix $A_t \in \mathbb{R}^{N\times N}$.

$$\alpha_{ij} = \frac{\exp\left(LeakyRelu\left(e_{ij}\right)\right)}{\sum_{k\in N_i}\exp\left(LeakyRelu(e_{ik})\right)}$$

② GNN Module: For easy learning, we normalize the adjacency matrix, and then obtain the spatial graph features of each time step:

$$F_t = \sigma(\Lambda^{-\frac{1}{2}}\hat{A}_t\Lambda^{-\frac{1}{2}}V^{(l)}W)$$

3. Trajectory Prediction Module

In the spatial-temporal fusion module, the spatial interactions are stacked along the temporal channel. The spatial-temporal interaction characteristics are obtained by 1 × 1 convolution. TCN is used to predict the future trajectory.

Results

We **compare the performance** of our proposed model ST-AGCN against baseline methods. From an FDE perspective, the FDE of our method has a 15% lower error compared to the FDE of the Social-STGCNN. From ADE perspective, our method ADE is 0.45, the same as the ADE of the SR-LSTM model, but the FDE of our method is 25% lower than the FDE error of SR-LSTM, and our method does not use the scene image information. Therefore, the ST-AGNN model has excellent performance compared to other methods.

Compared with several other models, our ST-AGNN model's **inference time** is 0.0030 which is 0.001 lower than the Social-STGCNN. We have the same **number of parameters** as Social-STGCNN. Overall, our model reflects relatively good results in terms of inference speed and model size.

Conclusion

We proposed **ST-AGNN** model for pedestrian trajectory prediction, which can effectively capture the spatial and temporal social interaction between pedestrians in the scene.

- Firstly, a spatial graph of pedestrians at each time step is constructed and the spatial features of pedestrians at each time step in the A-GCN module is extracted.
- Then the spatial features of each time step are stacked and the spatial-temporal features are obtained.
- Finally, the temporal convolutional network is used to predict the future trajectory of the pedestrians.

The **main contributions** are below:

· The Attention Graph Neural Network (A-GNN) module is proposed, which is used to capture the spatial interaction features of pedestrians in the scene at each time step.

· Our model obtains excellent experimental results on ETH and UCY datasets, and the FDE result of our method is 25% lower than the SR-LSTM. The comparison results of our method in model size and inference speed are also very good.

Acknowledgements

This work was supported in part by the National Natural S cience Foundation of China (2020YFC2003502, 61876201,61 901074,72171031), the Natural Science Foundation Project of Chongqing (cstc2020jcymsxmX0649) and the Science and the Technology Research Program of Chongqing Municipal Education Commission (KJQN201900636)