ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20210311111839
Original scientific paper

Generator with Triangulation for Pedestrians Trajectory Prediction

Xiuhong MA, Haitao WANG*, Qiulin MA

Abstract: Pedestrian trajectory prediction is a basic task in computer vision field. The prosperity of artificial intelligence makes the automatic drive, human-robot interaction and surveillance video attract a great deal of attention. Generally, researchers always place emphasis on pedestrian trajectory. The focuses of pedestrian trajectory prediction task are motion pattern modelling and spatio-temporal interaction modelling in the current study. In our paper, we present a GAN-based framework to model pedestrian motion pattern. A Delaunay triangulation algorithm is applied to map the pedestrian interaction. From the perspective of space, both the position interaction and motion interaction of pedestrians can be considered. For example, the influence of the movement direction and motion potential energy of pedestrians on the surrounding pedestrians can be modelled.

Keywords: delaunay triangulation; generative adversarial nets; pedestrian trajectory prediction

1 INTRODUCTION

Pedestrian motion has played an important role in crowd monitoring scene analysis, robot cruise and traffic environment understanding from the original only Kalman filter [1] in the detection of correct path of robots to the joint prediction of pedestrian trajectory and robot motion path. That is to say, robots can improve the performance of their own path detection by learning the prediction task of pedestrian trajectory. Another example is the precise positioning of the target [2] with the prediction task. For the huge errors happening when the target turns its direction, this paper proposes an algorithm to predict the turning point of its motion path. This method can improve the approximation between depicted and real trajectory, and promote the localization and tracking accuracy of the whole system. In terms of the pedestrian trajectory prediction, it can learn the motion pattern and predict the interaction between pedestrians and the scene.

Pedestrians' trajectory prediction is getting more and more attention. In recent years, researchers are continuing to try various advanced methods. For examples: LSTM, GAN, GCN, GAT, etc. They all have achieved different degrees of achievement. Among them, the attention mechanism has achieved remarkable results. In this task, we usually focus on two aspects: one is motion pattern, the other is interaction. Motion patterns are used to model the physical changes of pedestrian trajectories, and interaction simulating the movement trend relation between pedestrians.

The purpose of pedestrian trajectory prediction is to forecast further location at the next t_{pre} time step, when the trajectories of t time steps are known. Generally, the dataset of the task is the surveillance video in crowd scenes. The number of pedestrians in the scene is variable, and the pedestrian motion pattern is different. The speed and direction of the individual will change with the neighbours or other factors in the scene. In Fig. 1, there are two crowd scenarios being shown, the upper one is dense crowd in *Zara* dataset, and the lower is sparse crowd in *Hotel* dataset.

Social-LSTM [3] uses LSTMs to learn the motion pattern by data driven. It also uses the social-pooling to simulate pedestrian interaction. This framework is one of the classical methods in trajectory prediction. It is proposed by Feifei-Li's team, and many studies have put forward the improvement on this framework.

In the pedestrian trajectory prediction task, motion pattern modelling is the focus of attention all the time. This paper uses a generative adversarial nets (GAN) [4] framework to learn pedestrian's motion pattern, because of the generate attributes of GAN. GAN network is an unsupervised learning method. It usually allows two neural networks to contest with each other in a game (in the form of a zero-sum game, where one agent's gain is another agent's loss). The GAN generator network randomly selects samples from the latent space as input, and its output results need to imitate the real samples in the training set as much as possible.

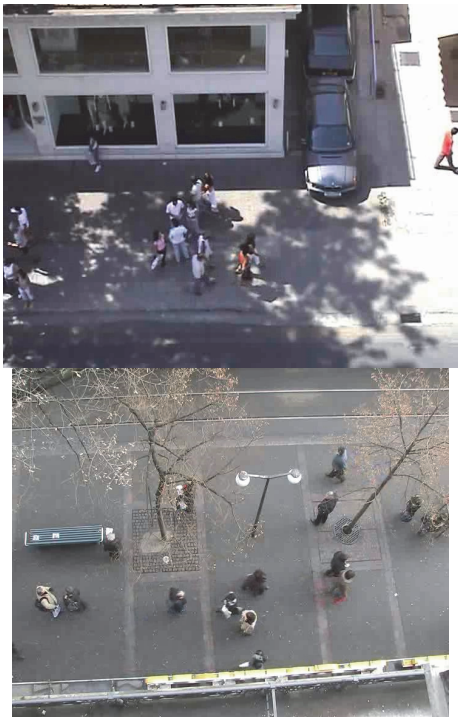

Figure 1 Two pictures of the crowd scenes in Zara and Hotel datasets

The input of the discriminator network is the real sample or the output of the generator network, and its purpose is to distinguish the output of the generator network from the real sample as much as possible. The

generator network should deceive the discriminator network as much as possible. The two networks confront each other and constantly adjust the parameters. The ultimate goal is to make the discriminator network unable to judge whether the output of the generator network is true.

However, GAN also has the shortcomings of unclear targets and poor controllability. In response to this problem, the researchers have proposed Condition GAN (CGAN) [5], the core of which is to integrate attribute information into the generator network and discriminator network. The attribute can be any label information. Therefore, this paper uses the GAN-based framework that is similar to CGAN, to establish a motion pattern learning model.

In the interaction modelling, triangulation topology is used to simulate the spatial relationship between pedestrians. This is because the positional relationship of pedestrians in a crowd scene changes all the time, and different pedestrians move at different velocities. For the target individual, it is not only the neighbours in the KNN or grid map relationship that have an important influence on its movement. There is a Delaunay triangulation in triangulation methods, which has uniqueness, neighbouring properties and non-intersect. That is, in Delaunay triangulation, a triangle is formed by the three closest points, and there is no intersection of two triangles. In the end, the triangle topological map formed is unique no matter which one starts. According to the characteristics of Delaunay triangulation, we use it to model the crowd relationship in the scene at each prediction time step.

In this paper, A GAN-based network similar to CGAN and Delaunay triangulation algorithm is used to model and predict pedestrians' trajectory.

2 RELATED WORKS

The tasks related to trajectory processing in computer vision include pedestrian detection and tracking, abnormal trajectory detection, trajectory prediction, and trajectory planning. Researchers have devoted a lot of effort to these tasks and achieved different results. Among them, pedestrian trajectory prediction is a relatively new type of research task. Because of the diversity of different application scenarios and targets in the scenarios, it has an intersection with areas such as autonomous driving and human-computer interaction.

In human-computer interaction projects, robots usually not only communicate with people, but they also need to share movement space with pedestrians. [6] thinks the ability to predict human actions can be useful for generating safe robot motions and action plan. However, knowing about what activity will be performed or the end location that a person is reaching or walking toward does not provide information about which specific portion of a shared human-robot workspace the human will occupy during the execution of that predicted action. We can leverage the additional information to ensure safe robot motion by enabling the robot to reason not only on the expected start and end locations, but on the entire expected human motion, [7] considers understanding the environment is a key requirement for any autonomous robot operation. This paper presents a probabilistic approach for general flow mapping, which can readily handle both of modelling the motion of objects (e.g., people) and flow of continuous media (e.g., air). Moreover, they present and compare two data imputation methods allowing building dense maps from sparsely distributed measurements. In addition, a first-person videos method [8] proposes a methodology for early recognition of activities from robot-centric videos (i.e., first-person videos) obtained from a robot's viewpoint during its interaction with humans. They introduce an algorithm to recognize human activities targeting the camera from streaming videos, enabling the robot to predict intended activities of the interacting person as early as possible and to take fast reactions to such activities (e.g., avoiding harmful events targeting itself before they actually occur).

In autonomous driving tasks, cars need to respond in a timely manner based on traffic and pedestrian conditions at intersections. That is to say, accurate traffic participant prediction is the prerequisite for collision avoidance of autonomous vehicles. In [9] paper, they propose to predict pedestrians using goal-directed planning. For this, they infer a mixture density function for possible destinations. They use these destinations as the goal states of a planning stage that performs motion prediction based on common behaviour patterns.

At present, the pedestrian trajectory prediction we are studying is only on people in a few fixed scenes. This type of task was proposed by Li Feifei's team and received attention. In this task, the researcher only pays attention to the pedestrian's own movement pattern, the influence of neighbours on him and the influence of several certain scenes on the choice of pedestrian trajectory.

In Social-LSTM [3], this paper presents a Social pooling layer with a grid map, which takes account of neighbours around the target. The authors use one LSTM for each trajectory and share the information between the LSTM through the Social pooling layer [10] utilizes Hamming distance to generate radial regions, and locate adjacent trajectories. However, this paper models a local neighbourhood when making the prediction, [11] proposes a novel GAN-based encoder-decoder framework for trajectory prediction capturing the multi-modality of the future prediction problem. They also propose a new pooling mechanism to model social interactions. The pooling mechanism captures "global" context, which uses relative positions to summarize all social information a person needs to make a decision. The above three methods are all proposed by Feifei-Li's team, and they provide researchers with a new research direction. Many methods proposed after these methods are improved on the basis of Social-LSTM or Social-GAN. For example, [12] is based on Social-LSTM and then takes into account the interaction with static (physical object) elements in the scene.

In the actual trajectory prediction task in a fixed scene, researchers consider the understanding of scene context on the basis of the spatio-temporal relationship models [13-17]. E.g. [16] develops a scene model, where the scene is divided into equal sized grid-cells which are further divided into sub-girds to provide more accurate spatial locations within the cell. Some researchers also use CNN to learn scene information, [17] blends a social attention mechanism with a physical attention that helps the model to learn where to look in a large scene and extract the most

salient parts of the images relevant to the path. However, scene context may cause the trajectory prediction task more suitable for specific scenes. Different from this, some methods use CNN to learn trajectory information for prediction directly. For example, [15] proposes a human trajectory prediction approach based on convolutional neural network (CNN). Unlike more recent LSTM-based models which attend sequentially to each frame, CNN-based model exploits the stack of convolutional layers to support increased parallelism and effective temporal representation. Similarly, [14] introduces a spatio-temporal convolutional neural network model for trajectory forecasting from visual sources.

3 GAN-BASED DELAUNAY TRIANGULATION MODEL

Assuming that the number of pedestrians in the scene at time step t is N_t, where the coordinate of pedestrian i is x_i^t, then the set of pedestrians at this time step is expressed as $X = \left\{x_i^t, x_2^t, ..., x_i^t, ..., x_{N_t}^t\right\}$. For our task, we give an observation sequence $\left\{x_i^1, x_i^2, ..., x_i^t\right\}$, and predict the next t_{pred} time steps $\left\{x_i^{t+1}, x_i^{t+2}, ..., x_i^{t+\text{pre}}\right\}$.

GAN is an emerging framework for estimating generative models through the confrontation process. GAN generally contains two models; a generative model G is used to fit the sample data distribution and a discriminative model D is used to estimate whether the input sample comes from the real training data or the generative model G. The generator maps the noise to the data space through the mapping function, and the output of the discriminator is a scalar, representing the probability that the data comes from the real training data instead of the G generated data. The optimization process of GAN is a "minimax two-player game" problem, and the value function is as follows:

Create equations with MathTypeEquation Editor (some examples are given below) (adjustments of sizes see Fig. 2).

$$\min_D \max_G V(D, G) = E_{x\sim pdata(x)}\left[\log D(x)\right] + \\ + E_{z\sim pdata(z)}\left[\log\left(1 - G(z)\right)\right] \tag{1}$$

This paper uses a network organization designed based on GAN to predict pedestrian trajectories, and map pedestrian spatial relationships through Delaunay triangulation. The network framework is shown in Fig. 2.

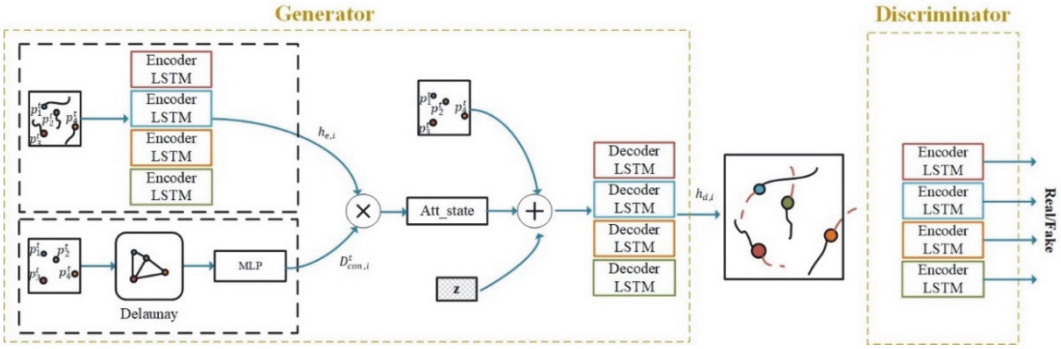

Figure 2 The pipeline of our framework

3.1 Delaunay Triangulation Graph

Triangulation is one of the key points of computational geometry research. The triangulation of a flat point set refers to the points concentrated in a finite plane, which are connected in a certain way to become a triangulation network that does not cross each other. At the same time, it is required that it is impossible to add new connecting lines on the graph after triangulation. Among all kinds of triangulation algorithms, Delaunay triangulation is the most widely used one.

Definition of triangulation: Assume that V is a finite point set on a two-dimensional real number field, edge e is a closed line segment composed of points in V as the endpoints, and E is a set of e. Then a triangulation $T = (V, E)$ of the point set V is a plane graph Gg, which satisfies the conditions: (1) Except for the endpoints, the edges in the plan view do not contain any points in the point set. (2) There are no intersecting edges. (3) All faces in the plan view are triangular faces, and the collection of all triangular faces is the convex hull of the scattered point se V.

Suppose an edge e in E (the two endpoints are a, b), e is called a Delaunay edge if it satisfies the following conditions: there is a circle passing through two points a and b, and the circle does not contain any other points in V. There are at most three points in the circle that are common, which is also called the empty circle feature. If a triangulation T of V contains only Delaunay edges, then the triangulation is called Delaunay triangulation DT. Delaunay triangulation graph is shown in Fig. 3. The red square indicates the target pedestrian, and the green circle indicates the neighbours.

In order to make Delaunay triangulation express pedestrian spatial relationship accurately, we design a distance feature (Eq. (2), Eq. (3)) to measure the interactions between observation and his/her neighbours, as follows:

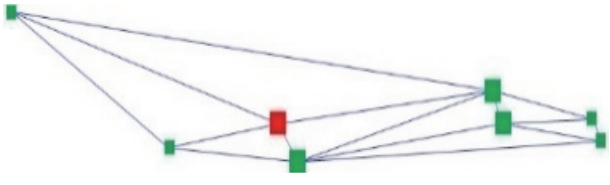

Figure 3 Schematic diagram of Delaunay triangulation between pedestrians at a certain time step

$$\overrightarrow{x_{ij}^t} = x_i^t - x_j^t, x_j^t \in DT\left(x_i^t\right) \tag{2}$$

$$D_{con,i}^t = \mathrm{MLP}\left(\frac{\overrightarrow{X_i}}{\left\|\overrightarrow{X_i}\right\|}\right) \tag{3}$$

where $DT\left(x_i^t\right)$ is the neighbors of $\left(x_i^t\right)$. $\overrightarrow{X_i}$ denotes the Delaunay mask matrix of $\overrightarrow{x_{ij}^t}$.

3.2 GAN-Based Framework

Trajectory data has time continuity, and LSTM is often used to learn the features of time series. This paper proposes to use an encoder-decoder model based on LSTM to learn the motion characteristics of pedestrian trajectories. As shown in Fig. 2, the generator of this framework consists of an encode-decoder module and a Delaunay triangulation graph.

Among them, the input of encoder-LSTM is the coordinates of observation sequence from time step 1 to current time step t, and the hidden state $h_{e,i}^t$ stores the motion pattern features, as the encoder-LSTM output.

$$\begin{aligned} &e_i^t = \phi\left(x_i^t; W_{e,i}\right), k \in (1, 2, ..., t), \\ &h_{e,i}^t = \mathrm{LSTM}_e\left(h_{e,i}^{t-1}, e_i^t, W_e\right) \end{aligned} \tag{4}$$

where $\phi(\cdot)$ is an embedding function with *ReLU* non-linearity, $W_{e,i}$ is the embedding weight. The encoder-LSTM weights W_e are shared between all people in a scene.

The encoder output is concatenated with the Delaunay mask $D_{con,i}^t$ of the output of Delaunay algorithm and noise z, together as the input of the decoder-LSTM.

$$\begin{aligned} &e_i^t = \phi\left(x_i^t; W_{\mathrm{d},i}\right), h_{\mathrm{d},i}^t = h_{e,i}^t, \\ &h_{e,i}^t = \mathrm{LSTM}_d\left(\left(h_{\mathrm{d},i}^{t-1} \otimes D_{con,i}^t\right), e_i^t, z, W_d\right) \end{aligned} \tag{5}$$

where $\phi(\cdot)$ is an embedding function with *ReLU* non-linearity, $W_{d,i}$ is the embedding weight. The decoder-LSTM weights are denoted by W_d.

The decoder-LSTM learns motion features and spatial features simultaneously, and outputs the predicted position $\bar{x}_i^{t+1}$ at the next time step $t + 1$.

$$\bar{x}_i^{t+1} = \gamma\left(h_{d,i}^t\right) \tag{6}$$

where γ is an MLP.

After experimental verification, the discriminator uses the encoder-LSTM module. Training data ground truth and the generated trajectory are respectively used as the input of the discriminator, and a score function is used to calculate the score of the predicted trajectory and ground truth. A *L2* function is applied as the loss function to calculate the generated loss. Another *L2* loss function is used to encourage the network to produce diverse samples. For each scene we generate k possible output predictions by randomly sampling z from $N(0, 1)$ and choosing the "best" prediction in *L2* sense as pure prediction.

$$L = \min_k \left\| Y_i - \bar{Y}_i^{(k)} \right\|_2 \tag{7}$$

where k is a hyperparameter.

4 EXPERIMENTS

We use the BIWI and UCY datasets which are the most popular datasets for pedestrian trajectory forecasting task. Both BIWI and UCY datasets are derived from high-angle cameras in natural scenes and contain variety of pedestrian motion patterns, such as going straight, turning toward a destination or avoiding pedestrians in the middle of the road.

4.1 Experiment Settings

The BIWI dataset [18] contains two outdoor scenarios of a long time period among which the eth-university shows an entrance of the school building, and the pedestrians in this scene all have clear destinations. The *Hotel* shows a street view in front of a *Hotel* building and the motion of pedestrians is complex in this scene. The UCY [19] is a crowd dataset by UCY Computer Graphics Lab. It includes three different scenarios: *Zara* Datasets, *Arxiepiskopi* and *University Students*, respectively. The data formats provided by two above datasets are not uniform. In order to compare them with other benchmark methods, we chose to keep the same with [20, 11] and transformed all data unit into meters.

Similar to the prior work, we report the prediction error with two metrics:

- *Average displacement error* (**ADE**): The mean Euclidean distance error over all estimated points of a trajectory and the true points.
- *Final Displacement error* (**FDE**): The distance between the predicted final destination and the true final destination at the end of the prediction period *T*.

Following [3], we use the leave-one-out approach, where four of the five scenes are used for training and validating, and the remaining one is used for testing. In these experiments, we set the sampling interval of videos in BIWI and UCY to 10 frames. In other words, the interval is 0.4 s at every time step. To be specific, the observed length of the trajectory is 8 time steps (3.2 s) and showing prediction results for 8 time steps (3.2 s).

4.2 Results Analysis

This paper uses 5 video datasets to conduct experiments, and the experimental results are quite different in different video datasets. As shown in Tab. 1, the experimental results of the *Eth* dataset have the largest gap with the other four datasets. In our analysis, this result is due to the fact that the *Eth* is a dataset with a large area

and a sparse population. Its population change law is quite different from the other four datasets. Therefore, the leave-one-out learning method has poor performance on the *Eth* dataset. On the other hand, due to *Eth*'s "large area and sparse population", the influence of the spatial relationship of its population is not obvious, and the prediction results of the learning model will also be inferior to other scenarios.

It can be seen from Tab. 1 that the prediction effects of *Zara*1 and *Univ* are worse than those of *Zara*2 and *Hotel*. Through analysis, it is found that *Zara*1 and *Zara*2 selected video clips with different traffic in the same scene. There are larger crowd density and more interaction between pedestrians in *Zara*2 than in *Zara*1, which also proves that pedestrian interaction can have an impact on pedestrian movement.

Table 1 Trajectory prediction results on five datasets

Methods	Metrics	*Eth*	Hotel	*Zara*1	*Zara*2	*Univ*
Ours	ADE	2.93	1.22	1.67	1.09	1.86
	FDE	9.81	4.97	9.08	9.03	12.13
Social-LSTM	ADE	0.73	0.49	0.27	0.33	0.41
	FDE	1.48	1.01	0.56	0.70	0.84
LSTM	ADE	0.70	0.55	0.25	0.31	0.36
	FDE	1.45	1.17	0.53	0.65	0.77

But another result appeared on the *Univ* dataset. The *Univ* scene has greater crowd density and more frequent pedestrian interactions, but it has not achieved the same prediction results as *Zara*2. Regarding this phenomenon, we believe that in addition to the intensive and frequent pedestrian movement in *Univ*, there are also a large number of small groups of pedestrians. There are some changes in these small groups (people come and go), but there are no changes (someone always stays still). The behaviour of these pedestrians is very difficult to judge in the prediction task. Small groups change in a small area, which affects the accuracy of prediction. Just like in the *Hotel* data set, its pedestrian density is greater than *Eth* and *Zara*1, but pedestrian interaction is not as complicated as *Univ*, and there is no formation and disappearance of complex small groups, so its prediction effect is the best.

As the baseline of pedestrian trajectory prediction, Social-LSTM reflects the datum line of this task research results. In contrast, our method has obvious disadvantages in the evaluation metric of FDE, which reflects the shortcomings of our method in the design and training, and also indicates that we should focus on solving the problem of inaccurate FDE in the future study.

We think that to solve the current bottleneck, we should add more learnable factors, more comprehensive understanding of the reasons for the formation of the trajectory and formulate it.

In summary, this paper believes that for the prediction task in a crowd scene, when only pedestrian motion patterns and interactions are considered, its prediction performance is affected by changes in the crowd in the scene. Complex crowd changes or too sparse crowds will cause the prediction results to be too little or to suffer from too much interference by the pedestrian interaction model.

5 CONCLUSION

This paper proposes a trajectory prediction method based on GAN, and Delaunay triangulation is used to model pedestrian interaction. The prediction performance is better than sparse scene or complex interaction when the crowd density is moderate and pedestrian interaction is simple. This shows that when considering the predictive influence factors, the external influence of pedestrians is to be taken into more comprehensive account in the current pedestrian trajectory prediction methods, for example the physical environment in the scene, the impact of small groups or high-density people on pedestrian motion.

In our future research, we will also learn the influencing factors of different pedestrian motions on the task of predicting pedestrian trajectories in different scenes. In addition to pedestrian motion pattern and interaction, additional considerations such as the characteristics of crowd, the characteristics of the flow of people, and the structure of the scene make the research of pedestrian trajectory prediction tasks more comprehensive.

Acknowledgements

This work is supported by the Fundamental Research Project of Humanities and Social Sciences in Colleges and Universities of Hebei Province (NO. GH171046).

We would like to show our deepest gratitude to the colleague, Ming Chen, who has provided us with valuable guidance in every stage of the writing of this paper. His support includes the following aspects: proper English language, grammar, punctuation and spelling.

6 REFERENCES

[1] Zolghadr, J. & Cai, Y. (2015). Locating a two-wheeled robot using extended Kalman filter, *Tehnički vjesnik*, *22*(6), 1481-1488. https://doi.org/10.17559/TV-20140531190647

[2] Qiu, Y. & Liu, C. (2014). Modelling and stimulation of target tracking and localization in wireless sensor network, *Tehnički vjesnik*, *21*(2), 233-238.

[3] Alexandre, A., Kratarth, G., Vignesh, R., Alexandre, R., Li, F. F., & Silvio, S. (2016). Social LSTM: Human Trajectory Prediction in Crowded Spaces. *IEEE Conference on Computer Vision and Pattern Recognition.* https://doi.org/10.1109/CVPR.2016.110

[4] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, Da., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *the International Conference on Neural Information Processing Systems*, *2*, 2673-2680.

[5] Mirza, M. & Osindero, Si. (2014). Conditional Generative Adversarial Nets, https://arxiv.org/abs/1411.1784

[6] Przemyslaw, A. L., Terrence, S., & Julie, A. S. (2017). A Survey of Methods for Safe Human-Robot Interaction. *Now Foundations and Trends*. https://doi.org/10.1561/2300000052

[7] Kucner, T. P., Magnusson, M., Schaffernicht, E., Bennetts, V. H., & Lilienthal, A. J. (2017). Enabling Flow Awareness for Mobile Robots in Partially Observable Environments. *IEEE Robotics and Automation Letters*, *2*(2), 1093-1100. https://doi.org/10.1109/LRA.2017.2660060

[8] Ryoo, M. S., Fuchs, T. J., Xia, L., Aggarwal, J. K., & Matthies, L. (2015). Robot-Centric Activity Prediction from First-Person Videos: What Will They Do to Me? *Proceedings of HRI*, 295-302. https://doi.org/10.1145/2696454.2696462

[9] Rehder, E., Wirth, F., Lauer, M., & Stiller, C. (2018). Pedestrian Prediction by Planning Using Deep Neural Networks. *2018 IEEE International Conference on Robotics and Automation* (*ICRA*), 5903-5908. https://doi.org/10.1109/ICRA.2018.8460203

[10] Alexandre, A., Vignesh, R., & Li, F. F. (2014). Socially-aware Large-scale Crowd Forecasting. *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). https://doi.org/10.1109/CVPR.2014.283

[11] Agrim, G., Justin, J., Li, F. F., Silvio, S., & Alexandre, A. (2018). Social GAN: Socially acceptable trajectories with Generative adversarial networks. *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). https://doi.org/10.1109/CVPR.2018.00240

[12] Sirin, H., Meiqing, W., He, W., & Siew, K. L. (2019). Situation-Aware Pedestrian Trajectory Prediction with Spatio-Temporal Attention Model. *24th Computer Vision Winter Workshop* (*CVWW*).

[13] Daksh, V. & Srinivasaraghavan, G. (2017). Human Trajectory Prediction using Spatially Aware Deep Attention Models. *31st Conference on Neural Information Processing Systems* (*NIPS*).

[14] Ehsan, P. & Christoph, H. L. (2018). Back to square one: Probabilistic Trajectory Forecasting without Bells and Whistles. *32nd Conference on Neural Information Processing Systems* (*NIPS*).

[15] Nikhil, N. & Morris, B. T. (2018). Convolutional Neural Network for Trajectory Prediction. *European Conference on Computer Vision-ECCV 2018 Workshops*. https://doi.org/10.1007/978-3-030-11015-4_16

[16] Huynh, M. & Gita, A. (2018). Scene-LSTM: A Model for Human Trajectory Prediction.

[17] Amir, S., Vineet, K., Ali, S., Noriaki, H., Hamid, R., & Silvio, S. (2019). SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints. *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). https://doi.org/10.1109/CVPR.2019.00144

[18] Stefano, P., Andreas, E., Konrad, S., Lcu, van G. (2009). You'll never walk alone: Modeling social behavior for multi-targer tracking. *IEEE 12th International Conference on Computer Vision* (*ICCV*). https://doi.org/10.1109/ICCV.2009.5459260

[19] Alon, L., Yiorgos, C., & Dani, L. (2007). Crowds by example. *Computer Graphics forum*, *26*(3), 655-664. https://doi.org/10.1111/j.1467-8659.2007.01089.x

[20] Pu, Z., Wanli, O., Pengfei, Z., Jianru, X. (2019). SR-LSTM: State Refinement for LSTM towards Pedestrian Trajectory Prediction. *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). https://doi.org/10.1109/CVPR.2019.01236

Contact information:

Xiuhong MA, PhD
Hebei University of Economics and Business
School of Management Science and Engineering, Hebei University of Economics and Business, Shijiazhuang, 050061, China
Email: maxiuhong@heuet.edu.cn

Haitao WANG, senior engineer
(Corresponding author)
Hebei University of Economics and Business
Modern Educational Technology Center, Hebei University of Economics and Business, Shijiazhuang, 050061, China
Email: wanght@heuet.edu.cn

Qiulin MA, PhD. student
Beijing Jiaotong University
Beijing Key Laboratory of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing, 100044, China
Email: 17112075@bjtu.edu.cn