

A Report of Trajectory Prediction based on Deep Learning

Chenxiao Tian
tian01@ads.uni-passau.de
University of Passau

Yashu Wang
wang52@ads.uni-passau.de
University of Passau

firstname lastname
mail@example.org

firstname lastname
mail@example.org

ABSTRACT

Pedestrians trajectories prediction has become a popular topic in recent years. Many approaches can be used to accomplish the prediction, including deterministic and non-deterministic ways. In this report we conduct research and discussion about how to use the a non-deterministic method i.e., deep learning to solve the prediction problem. The algorithm we are going to use is the combination of SocialGAN and InfoGAN. First, we train a model by the adopted algorithms. Second, during the research, we propose our research questions and try to answer it in the report.

PHASES ASSIGNMENT

- **Phase1** Chenxiao Tian
- **Phase2** Zubair Ahmed
- **Phase3** Yashu Wang
- **Phase4** Rita Akhmetova

INTRODUCTION

Systems and electronic devices that use pedestrian trajectories prediction can be found everywhere in our daily life, e.g., service robots, automatic drive and city planning etc. Hence, it is useful and important to make predictions about pedestrians movement. Many researches have proposed miscellaneous approaches that tackle this problem. Helbing and Molnar [6] propose the Social Force model. Yi [8] introduces the factor of stationary group to the modeling of pedestrians trajectories with an energy map. The aforementioned ways are deterministic ways for prediction, they can not utilize the valuable information in the trajectories data.

Over the last few years, following the widespread usage of machine learning and deep learning, researchers use various neural networks to tackle the trajectories prediction problem. Zhou et al. [9] build a linear dynamic system, applying Expectation Maximization (EM) algorithm to estimate parameters, to learn motion patterns in crowded scenes. Althé [2] proposes a method that predicts the trajectory on the highway using Long Short-Term Memory (LSTM). Alahi et al. [1] give a sequence model based on LSTM as well as a social pooling that aggregates the human-human interaction in a scene.

However, these approaches mentioned previously learn only the pattern of human motion from data. Predicting human trajectory is a complex task. This is because both internal and external stimuli, such as intentions and other directly or indirectly observable influences, can affect human motion, as mentioned in the survey [7]. In

addition to the location, which is usually recorded in the dataset, many factors that are not explicitly recorded in the dataset, such as speed, direction, or even not recorded, such as route and human intent. Recent researches have shown that Generative Adversarial Network (GAN) can better capture these uncertainties with latent space and thus naturally preserve multimodality. Gupta et al. [5] used GAN and a Pooling Module to predict socially acceptable trajectories and found that certain directions in the latent space are related to direction and velocity. What is more, The study of Amirian et al. [3] has shown that InfoGAN, an information-theoretic extension to the Generative Adversarial Network [4], partly improves the performance on commonly used datasets that have the largest variance in the prediction distribution, while still leaving some room for improvement.

Even though these researches give various effective models that fulfill the prediction task and attempt to encompass hidden aspects that influence the trajectory, they have not disentangled these factors in the latent space. If we know the factors that affect pedestrians' trajectory and apply these factors in specific scenarios. We can obtain better performance of prediction on various distributed datasets and to mitigate the limitations of the observed data. Therefore, we decide to consider the hidden factors behind different datasets.

In this study, we focus on what factors we can obtain that influence human trajectories and try to develop a model that can be controlled by these factors. we assume that different datasets have different static environments and so the data in a dataset share some specific common features. We consider three factors: obstacles (obstacles information such as the presence of static obstacles and the coordinates), maps (geometry and topology), and semantics (environment semantics such as no-go-zones, crosswalks, sidewalks, or traffic lights) in static environments, which are denoted by the survey [7]. We propose to develop a conditional generation model that is controlled by factor c to have different static environments. We demonstrate that human movement is influenced by these three factors we consider in a static environment, and so with inputting different factors in static environments, our model can achieve better performance on different datasets.

1 PROBLEM STATEMENT

In this paper, our goal is to develop a controllable generative model to predict pedestrian trajectories. Consider the problem of predicting the future trajectory of each pedestrian. Let (x_i^t, y_i^t) denote the

position of the i pedestrian at time t , and a sequence of coordinates $[(x_i^t, y_i^t), (x_i^{t+1}, y_i^{t+1}), \dots, (x_i^{t+n}, y_i^{t+n})]$ denote the trajectory of pedestrians from time t to $t+n$.

Given the observed trajectory of n_{obs} steps $X_i^t = [(x_i^t, y_i^t), (x_i^{t+1}, y_i^{t+1}), \dots, (x_i^{t+n_{obs}}, y_i^{t+n_{obs}})]$, with certain condition c and random variable z , we want to fit a function to generate the prediction of trajectory for the next n_{pred} steps $Y_i^t = [(x_i^{t+n_{obs}+1}, y_i^{t+n_{obs}+1}), (x_i^{t+n_{obs}+2}, y_i^{t+n_{obs}+2}), \dots, (x_i^{t+n_{obs}+n_{pred}}, y_i^{t+n_{obs}+n_{pred}})]$. That is

$$Y_i^t = f(X_i^t | c, z)$$

The prediction Y_i^t is conditioned on the vector c , where consist of (c_1, c_2, c_3) . So we can control the factors of obstacles, maps, and semantics respectively. These factors might vary over time.

REFERENCES

- [1] Alexandre, Kratharth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. 2016. Social LSTM: Human Trajectory Prediction in Crowded Spaces. (2016), 961–971. <https://doi.org/10.1109/CVPR.2016.110>
- [2] Florent Althé and Arnaud de La Fortelle. 2017. An LSTM network for highway trajectory prediction. (2017), 353–359. <https://doi.org/10.1109/ITSC.2017.8317913>
- [3] Javad Amirian, Jean-Bernard Hayet, and Julien Pettre. 2019. Social Ways: Learning Multi-Modal Distributions of Pedestrian Trajectories With GANs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
- [4] Xi Chen, Yan Duan, Rein Houthooft, John Schulman, Ilya Sutskever, and Pieter Abbeel. 2016. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*. 2180–2188.
- [5] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. 2018. Social GAN: Socially Acceptable Trajectories With Generative Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [6] D. Helbing and P. Molnar. 1995. Social force model for pedestrian dynamics. *Physical review E* (1995), 4282–4286. <https://doi.org/10.1103/PhysRevE.51.4282>
- [7] Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M Kitani, Dariu M Gavrila, and Kai O Arras. 2020. Human motion trajectory prediction: A survey. *The International Journal of Robotics Research* 39, 8 (2020), 895–935.
- [8] Shuai Yi, Hongsheng Li, and Xiaogang Wang. 2015. Understanding pedestrian behaviors from stationary crowd groups. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3488–3496. <https://doi.org/10.1109/CVPR.2015.7298971>
- [9] Bolei Zhou, Xiaoou Tang, and Xiaogang Wang. 2015. Learning Collective Crowd Behaviors with Dynamic Pedestrian-Agents. 111, 1 (2015). <https://doi.org/10.1007/s11263-014-0735-3>