A HYBRID AGENT-CENTRIC AND SCENE-CENTRIC APPROACH FOR MULTI-AGENT TRAJECTORY PREDICTION

Tang Yi, German O.V.
Department of Information Technologies in Automated Systems,
Belarussian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus
E-mail: tangyijcb@163.com, ovgerman@tut.by

Accurately predicting the future trajectories of agents in autonomous driving is crucial for safe navigation and decision-making. Traditional trajectory prediction models have limitations when dealing with complex multi-agent interactions. In this paper, we propose a hybrid approach that leverages the strengths of both agent-centric and scene-centric models by using agent-centric normalization for dynamic agents and a scene-centric framework for static map elements.

Introduction

The task of trajectory prediction is fundamental to the safe operation of autonomous vehicles (AVs). AVs must accurately predict the future positions of surrounding dynamic agents, such as vehicles and pedestrians, while also understanding the static context, such as road boundaries and traffic signals. Traditional approaches can be broadly categorized as instance-centric or agent-centric.

Agent-centric models normalize the world around each agent individually. The advantage of this method is that it simplifies the prediction problem for each agent by reducing relative motion and viewpoint complexity. However, this approach can introduce computational overhead, as each agent requires a separate normalization process.

In contrast, scene-centric models operate in a global reference frame. All agents and static elements (such as lanes and traffic signs) are modeled together in the same coordinate system. This reduces computational complexity but can lead to inaccuracies when dealing with interactions, especially when agents are moving in different directions or at varying speeds[1].

In this paper, we propose a hybrid agent-centric and scene-centric approach to overcome these limitations. By combining the strengths of both paradigms, our method enables accurate trajectory prediction in complex multi-agent environments while maintaining computational efficiency.

I. Problem statement

The task of multi-agent trajectory prediction involves predicting the future positions of each agent based on their past trajectories and environmental context. Mathematically, let $X^i = \{\mathbf{x}_t^i\}_{t=1}^T$ represent the observed trajectory of agent i, where $\{\mathbf{x}_t^i\}$ denotes the position of agent i at time t. The goal is to predict the future trajectory $\hat{X}^i = \{\hat{\mathbf{x}}_t^i\}_{t=T+1}^{T+H}$, where H represents the prediction horizon.

In existing approaches, agent-centric models apply normalization for each agent, which involves recalculating the positions of surrounding agents and static elements relative to the target agent. This often leads to redundant computations, especially when dealing with multiple agents. Scene-centric models, while computationally efficient, struggle with relative viewpoint shifts among agents, leading to lower prediction accuracy in complex multi-agent scenarios.

II. Description of existing methods

Agent-centric methods normalize the environment by transforming the past and surrounding trajectories of all agents relative to a specific target agent. In this approach, each agent's motion is analyzed in a local coordinate system that is centered on the agent itself. This allows for precise modeling of the agent's behavior and movement patterns, as its trajectory is examined relative to its own motion. The transformation into a local frame helps to eliminate the effects of external global references, making the analysis more focused on the agent's perspective[2]. However, while agent-centric methods can ensure a highly accurate understanding of each individual agent's motion, they come with a significant computational cost. For each agent, the system must continuously recalculate the positions of all other agents and static elements relative to the target agent. This process, while effective for a single or a small number of agents, becomes computationally expensive as the number of agents increases, leading to inefficient scaling in multi-agent scenarios. In complex environments with many interacting agents, the need to independently transform and normalize the trajectories for each agent can quickly overwhelm the system, limiting its scalability and overall efficiency.

On the other hand, the scene-centric approach models all agents and static elements within a global reference frame without recalibrating each agent. This approach is computationally efficient because it avoids repeating the normalization process for each agent's local environment. While this approach simplifies the modeling process and reduces the computational burden, there are trade-offs in terms of accuracy, especially when dealing with multi-agent interactions. Scene-centered models can be difficult to dynamically adjust to each agent's unique perspective, including differences in viewpoint, direction of motion, and velocity. This lack of flexibility can lead to reduced accuracy when predicting interactions between agents, as the global framework may not be able to accurately capture the relative motion between agents. In addition, the scene-centric approach has a significant drawback in the form of weak generalization of the global scene. Such models are prone to overfitting static elements in a particular scene and the global environment during training. When the scene in the training data changes, the prediction accuracy of the model decreases significantly. For example, subtle changes in lane layouts, traffic signs, or road structures may lead to inaccurate predictions of a scene-centered model in a new scene because it lacks adaptability to changes in the scene[3].

III. Method description

To address the limitations of existing approaches, we propose a hybrid agent-centric and scene-centric approach for multi-agent trajectory prediction. This method uses instance-centric normalization for dynamic agents and scene-centric representation for static elements like roadways, lane segments, and traffic signs.

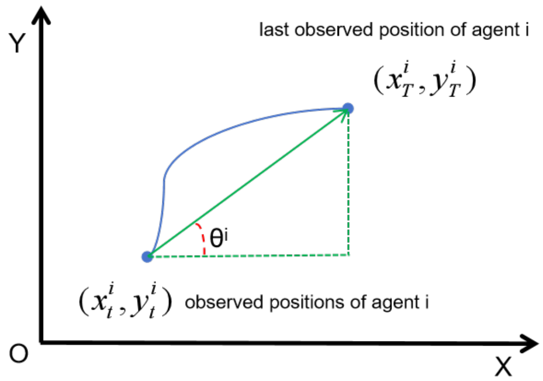

Рис. 1 – Displacement vector and heading angle

For each dynamic agent i, we transform the trajectory into a local reference frame based on the agent's last position and heading. Figure 1 shows the displacement vector and heading angle. The trajectory transformation is as follows:

$$\tilde{\mathbf{x}}_t^i = R^i(\mathbf{x}_t^i - \mathbf{x}_T^i) \tag{1}$$

where $\mathbf{x}_T^i$ is the agent's current position at time T, and R^i denotes the rotation matrix that aligns the agent's heading direction with the x-axis:

$$R^i = \begin{bmatrix} \cos(\theta^i) & -\sin(\theta^i) \\ \sin(\theta^i) & \cos(\theta^i) \end{bmatrix} \tag{2}$$

where $\theta^i = \arctan(\mathbf{x}_t^i - \mathbf{x}_T^i)$ being the agent's heading angle derived from its past trajectory. This instance-centric normalization ensures that the agent's future trajectory prediction is robust to variations in heading direction, allowing the model to focus on predicting future movement patterns relative to the agent's current motion.

Static elements such as lane lines, road boundaries, traffic signs and traffic signals are kept represented in a global coordinate system without depending on specific dynamic intelligences. These static elements are characterized by a fixed position that does not change over time, so keeping them in the global reference system avoids the repetitive computation of coordinate transformations for different intelligences each time. Road geometries such as lane lines are described by multiple sampling points and represented under the global coordinate system, which provides consistent reference information for all intelligences[4]. For example, a lane segment j can be represented by its centroid C_j and orientation θ_j, which is derived from the displacement vector between its endpoints:

$$\theta_j = \arctan 2(y_2^j - y_1^j, x_2^j - x_1^j) \tag{3}$$

This approach eliminates redundant calculations by keeping static elements in a global frame, while dynamic agents are modeled in their local reference frames. Especially in multi-intelligent body scenarios, the trajectory prediction of dynamic intelligences can be directly based on the interaction of these global static elements, which improves the computational efficiency, especially in complex scenarios containing a large number of static elements.

IV. Conclusion

We proposed a novel hybrid approach that combines the strengths of agent-centric and scene-centric trajectory prediction methods. By applying instance-centric normalization for dynamic agents and maintaining scene-centric representation for static elements, our method achieves both high prediction accuracy and computational efficiency. This hybrid approach is particularly effective in complex, multi-agent environments, where both dynamic and static elements play a crucial role in trajectory prediction. Future work will explore integrating more sophisticated interaction modeling between agents and extending the method to larger-scale real-world scenarios.

1. Liang, M., Yang, B., Hu, R., & Urtasun, R. (2020). Learning Lane Graph Representations for Motion Forecasting. Arxiv preprint arXiv:2007.13732.
2. Zhang, L., Li, P., Liu, S., & Shen, S. (2024). SIMPL: A Simple and Efficient Multi-Agent Motion Prediction Baseline for Autonomous Driving. Arxiv preprint arXiv:2402.02519.
3. Bharilya, V., & Kumar, N. (2023). Machine Learning for Autonomous Vehicle's Trajectory Prediction: A Comprehensive Survey. Arxiv preprint arXiv:2307.07527.
4. Huang, Y., Du, J., Yang, Z., Zhou, Z., & Chen, H. (2022). A Survey on Trajectory-Prediction Methods for Autonomous Driving. IEEE Transactions on Intelligent Vehicles, 7(3), 652-667.