차량-인프라 협력 인지(V2X) 환경에서의 차량 궤적 예측 모델

이용선 [1], 함제석 [2*], 문진영 [3]
[1] 서울과학기술대학교 인공지능응용학과 학사과정
[2] 한국전자통신연구원 연구원
[3] 한국전자통신연구원 책임연구원

yongseon@seoultech.ac.kr, jsham@etri.re.kr, jymoon@etri.re.kr

Vehicle Trajectory Prediction Model in Vehicle-to-Infrastructure Cooperative Perception(V2X) Environments

Yongseon Lee[1], Je-Seok Ham[2*], Jinyoung Moon[2]
[1]Dept. of Applied Artificial Intelligence, Seoul National University of Science and Technology
[2]Electronics and Telecommunications Research Institute (ETRI)

요 약

기존 단일 차량 데이터 기반 도로주행은 차량의 시야 범위에 한정된 지각 능력을 가진다. 이로 인해 전체적인 주변 교통 상황을 충분히 반영하지 못하며, 결과적으로 미래 궤적 예측과 의사결정의 안정성에 한계가 존재한다. 이러한 문제를 해결하기 위해 인프라 기기로부터 수집된 데이터를 활용하여 지각 범위를 확장하려는 연구가 활발히 진행되고 있다. 본 논문에서는 차량과 인프라 기기에서 동시에 수집된 데이터를 포함하는 V2X-Seq Dataset 을 활용하여, 시점별로 수집된 차량 데이터의 시계열적 특성을 강화시킨 모델인 V2ITrajNet 을 제안한다. 실험 결과, V2ITrajNet 은 모든 평가 지표에서 벤치마크 모델 대비 성능 향상을 보였으며, 이를 통해 향후 복잡한 도로 환경에서도 인프라 기기와의 협력을 바탕으로 보다 안전하고 정확한 궤적 예측 및 의사결정을 내릴 수 있다.

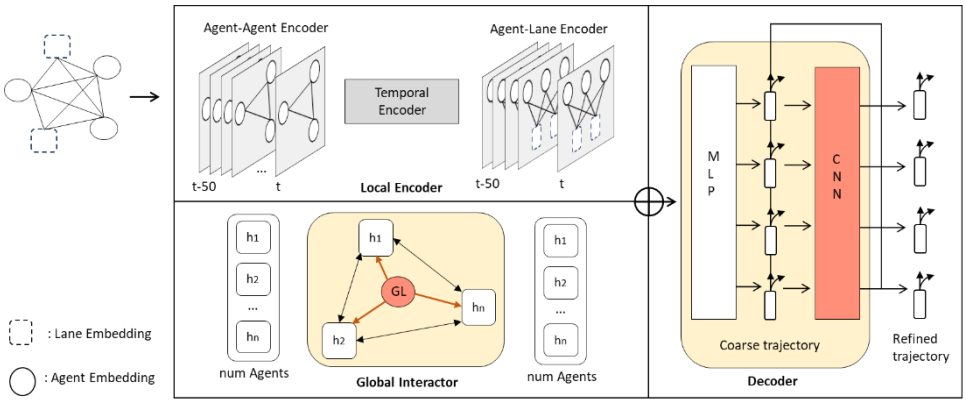
1. 서론

자율주행 차량의 미래 행동 지시에 있어 주변 차량들의 궤적 예측은 선행되어야 한다. 그러나 단일 차량 기반의 데이터는 복잡한 도심 환경에서 인식 공간이 크게 제한된다. 이 문제를 해결하기 위해 최근 차량 외부 인프라 기기의 데이터를 활용하는 Vehicle-to-Everything(V2X) 연구가 활발히 진행되고 있다.

인프라 기기 데이터의 활용은 자율주행 차량의 인식 범위를 확장하고 주행 안전성을 향상시킬 수 있음이 제시되어 왔다[1]. 본 연구를 뒷받침하기 위해 실제 자율주행 차량이 존재하는 도로 환경에서 수집된 대규모 시계열 V2X 데이터셋인 V2X-Seq[2]가 공개되었다. 인프라 기기와 차량으로부터 수집된 데이터가 모두 포함되어 있으며, 특히 V2X-Seq-TFD(Trajectory Forecasting Dataset)는 궤적 예측 과제를 지원한다.

본 논문은 V2X-Seq-TFD 데이터셋을 활용하여 기존 차량 궤적 예측 과제에서 성능을 입증한 HiVT 모델[3]을 기반으로, 새로운 모델인 V2ITrajNet 을 제안한다. V2ITrajNet 은 시점별 에이전트 주변에 존재하는 도로 환경과 주변 차량들의 관계를 효과적으로 학습하기 위해 Global Token 을 추가한다. 또한 기존 단일 디코더 기반의 궤적 예측 방식에서 확장하여, CNN 계열 보정 디코더를 도입함으로써 시계열적인 특성을 강화한다. 마지막으로 큰 오차에 대해서는 학습 안정성을 높이고 작은 오차에 대해서는 정밀한 예측을 유도하는 Smooth L1 Loss 를 반영한 Trajectory Loss 를 적용한다. 실험 결과, 제안하는 V2ITrajNet 은 기존 HiVT 모델 및 선행 연구에서 발표된 벤치마크 모델들과 비교하여 모든 평가 지표에서 우수한 성능을 보였다. 이는 시점별로 수집된 차량 데이터의 시계열적 특성을 강화하여 반영한 결과로 해석된다. 따라서 V2ITrajNet 은 자율주행 차량이 인프라 기기 데이터와 차량 데이

* 교신저자(Corresponding Author)



(그림 1) 제안하는 모델 V2ITrajNet 의 구조도

터를 효과적으로 통합·활용함으로써, 복잡한 도로 환경에서도 보다 안전하고 정밀한 궤적 예측 및 행동 지시를 가능하게 함을 입증하였다.

2. 관련 연구

차량 궤적 예측 과제에서 예측의 안정성을 높이기 위해, 하나의 궤적 대신 복수의 후보 궤적을 설정하는 다중 모달리티 기반 궤적 예측이 활발히 연구되어 왔으며, 이는 자율주행 분야의 핵심 과제로 자리 잡고 있다. 이러한 접근법의 대표적인 사례로, TNT (Target-driven Trajectory Prediction)[4]은 후보 지점(target)을 먼저 예측한 후, 해당 지점으로 이어지는 궤적을 생성하는 방식으로 다중 후보 궤적을 효과적으로 탐색한다. 이를 통해 장기 예측 시 다양한 주행 모드를 반영할 수 있으며, 지도 정보와 결합하여 복잡한 도로 구조에서도 안정적인 성능을 보인다. 또한, 본 논문에서 베이스라인으로 활용하는 HiVT 는 벡터 기반 트랜스포머 구조를 활용하여 다중 에이전트 간의 지역적•전역적 상호작용을 동시에 학습하는 모델이다.

차량과 인프라 데이터를 함께 이용하기 위해 V2INet[5]의 궤적 예측 모델은 멀티 뷰 정보를 결합하여 차량 단독 센서의 한계를 보완하였다. 그래프 기반의 접근법인 V2X-Graph[6]은 차량과 인프라에서 수집된 궤적을 그래프 노드로 표현하고, 이종 그래프 구조를 활용하여 멀티 뷰의 궤적 정보를 융합한다. 이를 통해 기존 단일 프레임 기반 방법이 포착하지 못했던 시계열적 맥락과 상호작용 정보를 반영할 수 있다. 마지막으로, HDGT(Heterogeneous Driving Graph Transformer)[7]는 보정 디코더를 추가함으로써 다중 디코더 구조로 확장하여, 다중 모달 궤적 예측에서 보다 세밀하고 안정적인 성능을 달성하였다.

3. 제안하는 모델

3.1. 데이터셋

본 연구에서는 V2X-Seq-TFD 데이터셋을 활용하여 모델 개발과 예측 성능 평가를 진행한다. V2X-Seq-TFD 데이터셋은 중국 베이징의 Yizhuang area 28 개 교차로에서 672 시간동안 수집한 데이터셋이다. 과거 인프라 기기로부터 수집된 차량 궤적 정보, 차량 기기로부터 수집된 차량 궤적 정보, 실시간 신호등 정보, 도로 Vector Map 을 포함하고 있다. 시점별 데이터 분포로는 80,000 개의 인프라 시점 데이터, 80,000 개의 차량 시점 데이터, 50,000 개의 협력 시점(Cooperative-view) 데이터로 구성되어 있다.

3.2. 베이스라인 모델

본 연구에서는 궤적 예측을 위한 베이스라인으로 HiVT(Hierarchical Vector Transformer) 모델을 채택하였다. HiVT 는 다중 에이전트 환경에서 발생하는 복잡한 상호작용을 효율적으로 학습하기 위해 제안된 모델로, 지역적 맥락 추출과 전역적 상호작용 모델링을 계층적으로 수행한다. 특히, 에이전트 중심의 지역적 표현을 기반으로 한 Local Encoder 와 지역 표현 간 메시지 패싱을 수행하는 Global Interaction Module 을 통해 에이전트 간의 국소적인 관계뿐만 아니라 장기적인 의존성까지 효과적으로 포착할 수 있다. 또한 HiVT 는 모든 에이전트의 궤적을 단일 순전파 과정에서 동시에 예측할 수 있기에 다중 에이전트 환경에 대응할 수 있는 장점을 지닌다. 이러한 장점으로 본 연구의 베이스라인을 HiVT 로 설정하고, 이를 개선하여 인프라 데이터를 활용한 협력적 차량 궤적 예측 성능을 탐구하였다.

3.3. 차량 궤적 예측 모델(V2ITrajNet)

본 연구에서 제안하는 V2ITrajNet 의 전체 아키텍처는 <그림 1>과 같다. 우선 데이터를 임베딩한 후 에이전트 간의 상호작용을 전역적으로 통합하기 위하여 Global Token 을 도입한다. Global Token 은 각 에이전트 임베딩에 결합되어 모델에 입력되며, 모든 에이전트 노드와 연결을 형성한다. 이를 통해 전체 장면의 정보를 요약하여 각 에이전트와 공유할 수 있다. 결과적으로 지역적 관계뿐만 아니라 전역적 맥락까지 반영할 수 있게 된다.

기존 HiVT 가 MLP 디코더를 통해 예측 궤적을 직접 산출하였다면, 본 연구에서는 Refine CNN 을 추가하여 궤적을 보정한다. 먼저 MLP 디코더에서 coarse trajectory(Y_{coarse})를 산출한다. Refine CNN 은 해당 궤적을 입력받아 시간 축 방향으로 convolution 연산을 수행하여 보정 오프셋 Δ 를 계산한다. 최종 출력은 다음

과 같이 정의된다.

$$Y_{refined} = Y_{coarse} + \Delta$$

이를 통해 Refine CNN 은 단일 디코더에서 발생할 수 있는 불연속적 궤적을 완화하고, 시계열적 연속성과 공간적 패턴을 반영한 정밀 궤적을 생성한다.

학습 과정에서는 Smooth L1 Loss 를 적용하여 궤적 예측의 안정성을 확보한다. Smooth L1 Loss 는 작은 오차 구간에서는 L2 Loss 처럼 정밀한 계산을 유도하며, 큰 오차 구간에서는 L1 Loss 처럼 완만한 학습을 진행한다.

$$SmoothL1(x) = \begin{cases} 0.5x^2 & (|x| < 1) \\ |x| - 0.5 & (|x| \geq 1) \end{cases},$$

이를 통해 모델은 이상치에 대한 강건성을 유지하면서도 세밀한 오차 보정을 수행할 수 있게 된다. 다중 모달 예측을 지원하기 위해 최대 6 개의 궤적 후보를 생성하며, 이들 중 최소 손실을 가지는 궤적을 선택하여 학습을 진행한다.

최종 학습 손실 Trajectory Loss 는 Smooth L1 Loss 와 기존 HiVT 에서 사용되던 Classification Loss 를 결합하여 다음과 같이 정의된다.

$$L_{Traj} = L_{cls} + \gamma L_{smoothL1}$$

여기서, γ 는 $L_{smoothL1}$의 비율을 의미하는 것으로 section 4.3 에서 변경 실험을 진행하였다. 결과적으로 V2ITrajNet 은 Global Token 을 통한 전역적 상호작용 강화, Refine CNN 을 통한 궤적 보정, 그리고 Trajectory Loss 를 이용한 학습 전략을 적용한다.

4. 실험 및 결과

4.1. 평가지표

본 연구에서 최소 평균 변위 오차(minADE), 최소 최종 변위 오차(minFDE), 예측 실패 비율(MR)을 성능 평가 지표로 사용한다. 실험에서는 다중 모달 궤적 예측을 위하여 6 개의 모드를 출력하며, 각 지표는 이들 후보 궤적 중 최적의 결과를 기준으로 계산된다.

minADE 는 모든 시간 단계에서 예측 궤적과 실제 궤적 간 평균 거리 오차의 최소값을 의미하여, 다음과 같이 정의된다.

$$\min ADE = \min_{k\in\{1,\dots,K\}} \frac{1}{H}\sum_{t=1}^{H} \left\| y_t^k - y_t \right\|_2$$

minFDE 는 최종 시점에서의 거리 오차 최소값을 의미하여, 수식은 다음과 같다.

$$\min FDE = \min_{k\in\{1,\dots,K\}} \left\| y_H^k - y_H \right\|_2$$

마지막으로 MR 은 최종 시점에서 예측 궤적의 종점이 실제 궤적으로부터 임계값 $\delta = 2m$ 이상 벗어나는 경우의 비율을 나타내며, 다음과 같이 표현된다.

$$MR = \frac{1}{N}\sum_{i=1}^{N} 1\left[\min_{k\in\{1,\dots,K\}} \left\| y_{H,i}^k - y_{H,i} \right\|_2 > \delta \right]$$

4.2. 정량적 평가

본 실험은 V2X-Seq-TFD 데이터셋을 사용하여 본 논문에서 제안하는 V2ITrajNet 모델과 베이스라인 모델 HiVT, 그리고 선행 연구에서 제안된 TNT, V2X-Graph, V2INet 모델의 성능을 비교하였다. <표 1>은 각 모델의 minADE, minFDE, MR 지표에 대한 결과를 제시하며, 각 지표는 값이 낮을수록 더 좋은 성능을 의미한다. 가장 높은 성능은 볼드체로, 두번째로 높은 성능은 밑줄을 그어 구분하였다.

<표 1> V2X-Seq-TFD 데이터셋에서
차량 궤적 예측 모델 성능 비교

Method	Cooperation	minADE↓	minFDE↓	MR↓
TNT [4]	Ego	8.45	17.93	0.77
TNT [4]	PP-VIC	7.38	15.27	0.72
HiVT [3]	Ego	1.34	2.16	0.31
HiVT [3]	PP-VIC	1.28	2.11	0.31
V2X-Graph [6]		<u>1.17</u>	2.03	<u>0.29</u>
V2INet [5]		1.19	<u>1.98</u>	**0.27**
V2ITrajNet(Ours)		**1.14**	**1.85**	**0.27**

실험 결과, 본 논문에서 제안하는 V2ITrajNet 은 minADE(1.14), minFDE(1.85), MR(0.27)에서 모두 가장 우수한 성능을 기록하였다. 이는 HiVT(1.28, 2.12, 0.31) 대비 각각 10% 이상의 성능 향상을 보인 것으로, 인프라 데이터와 차량 데이터를 통합한 본 연구의 접근법이 효과적임을 확인할 수 있다. 또한 V2X-Graph 와 비교하여서는 minADE 1.17 → 1.14 로, V2INet 와 비교하여서도 minFDE 1.98 → 1.85 와 같이 성능이 향상되어 벤치마크 모델들과 비교하였을 때, V2ITrajNet 은 가장 높은 성능(SOTA) 결과를 기록하며 안정적인 궤적 예측 능력을 입증하였다. 특히, minADE 와 minFDE 의 개선은 다중 모달 궤적 예측에서 보다 정확한 궤적 후보를 산출할 수 있음을 의미하며, 낮은 MR 값은 실제 궤적으로부터 크게 벗어나지 않는 예측을 더 자주 생성함을 보여준다. 이러한 결과는 V2ITrajNet 모델이 복잡한 도로 환경에서도 보다 안정적이고 정확한 예측 결과를 도출할 수 있음을 시사한다.

4.3. Trajectory Loss Ratio 변경 실험

본 실험에서는 Smooth L1 Loss 와 Classification Loss 의 결합 비율(Ratio)을 변화시키며 성능을 비교하였다. <표 2>는 Ratio 값에 따른 minADE, minFDE, MR 결과를 보여준다. 실험 결과, Ratio 가 0.4 와 0.5 일 때 가장 우수한 성능을 기록하였다. 특히 minADE 와 minFDE 가 각각 1.14, 1.85 로 최소값을 보였으며, MR 또한 0.27 로 안정적인 수준을 유지하였다.

반면, Ratio 값이 0.2 이하로 작아질 경우 Smooth L1 Loss 의 기여도가 낮아지면서 MR 이 증가하는 경향을 보였다. 이는 궤적의 세밀한 보정 효과가 충분히 반영되지 못한 결과로 해석된다. 반대로 Ratio 가 0.8 이상의 큰 값을 가질 경우 Smooth L1 Loss 가 지나치게 강조되어 학습이 불안정해지며 minFDE 와 MR 이 다시 상승하는 양상이 나타난다. 따라서 적절한 결합 비율을 설정하는 것이 성능 향상에 핵심적인 요소임을 확인할 수 있으며, 본 연구에서는 Ratio 를 0.4~0.5 범위로 설정하는 것이 가장 효과적임을 실험적으로 검증하였다.

<표 2> Trajectory Loss Ratio 변경 실험 결과

Ratio	minADE	minFDE	MR
0.1	1.20	1.99	0.30
0.2	1.16	1.89	0.28
0.3	1.15	1.88	0.30
0.4	**1.14**	**1.85**	**0.27**
0.5	**1.14**	**1.85**	**0.27**
0.6	1.14	1.87	0.27
0.7	1.18	1.91	0.29
0.8	1.15	1.86	0.28
0.9	1.16	1.89	0.28
1.0	1.14	1.87	0.27
1.1	1.17	1.90	0.28
1.2	1.17	1.88	0.27
1.3	1.15	1.88	0.28
1.4	1.17	1.89	0.28

5. 결론 및 향후 연구

본 연구에서는 V2X-Seq-TFD Dataset 을 활용하여 차량과 인프라 데이터를 통합한 궤적 예측 모델 V2ITrajNet 을 제안하였다. 제안한 모델은 Global Token 을 통해 전역적 상호작용을 강화하고, Refine CNN 을 통해 예측 궤적을 보정하였으며, Smooth L1 Loss 와 Cls Loss 를 결합한 학습 전략을 통해 안정성과 정확성을 동시에 확보하였다. 실험 결과, V2ITrajNet 은 기존 벤치마크 방법들 대비 minADE, minFDE, MR 지표에서 일관되게 향상된 성능을 보이며 제안한 접근법의 효과성을 입증하였다. 향후 연구에서는 보정 디코더 모듈을 더욱 고도화하여 복잡한 교통 상황에서도 안정적인 궤적 예측을 수행할 수 있도록 할 예정이다. 그 다음으로, 다양한 손실 함수 설계와 결합 비율 최적화에 대한 추가적인 탐구를 통해 성능 개선 가능성을 모색할 것이다. 마지막으로, 실제 도로 환경에서 수집한 데이터셋에 적용하여 제안한 모델의 실용성을 강화하는 방향으로 연구할 계획이다.

6. 사사

이 논문은 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(RS-2020-II200004, 장기 시각 메모리 네트워크 기반의 예지형 시각지능 핵심기술 개발).

참고문헌

[1] Eduardo Arnold, Mehrdad Dianati, Robert de Temple, and Saber Fallah, “Cooperative perception for 3d object detection in driving scenarios using infrastructure sensors”, *IEEE Transactions on Intelligent Transportation Systems*, 2020, 1, 2, 3

[2] Haibao Yu, Wenxian Yang, Hongzhi Ruan, Zhenwei Yang, Yingjuan Tang, Xu Gao, Xin Hao, Yifeng Shi, Yifeng Pan, Ning Sun, Juan Song, Jirui Yuan, Ping Luo, Zaiqing Nie, "V2X-Seq: A Large-Scale Sequential Dataset for Vehicle-Infrastructure Cooperative Perception and Forecasting," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, Canada, 2023, pp. 5486-5495.

[3] Zikang Zhou, Luyao Ye, Jianping Wang, Kui Wu, and Kejie Lu, “Hivt: Hierarchical vector transformer for multi-agent motion prediction”, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 8823–8833.

[4] Hang Zhao, Jiyang Gao, Tian Lan, Chen Sun, Ben Sapp, Balakrishnan Varadarajan, Yue Shen, Yi Shen, Yuning Chai, Cordelia Schmid, et al., "TNT: Target-driven trajectory prediction," *Proceedings of the Conference on Robot Learning (CoRL)*, London, UK, 2021, pp. 895–904.

[5] Xi Chen, Rahul Bhadani, and Larry Head, "Conformal Trajectory Prediction with Multi-View Data Integration in Cooperative Driving," arXiv preprint arXiv:2408.00374, 2024.

[6] Hongzhi Ruan, Haibao Yu, Wenxian Yang, Siqi Fan, and Zaiqing Nie, "Learning Cooperative Trajectory Representations for Motion Forecasting," *Annual Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, 2024.

[7] X. Jia, P. Wu, L. Chen, Y. Liu, H. Li, and J. Yan, "HDGT: Heterogeneous Driving Graph Transformer for Multi-Agent Trajectory Prediction via Scene Encoding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, pp. 13860-13875, 2023.