

자율주행 시뮬레이터에서 LLM 기반의 비정형 데이터 처리 방법

최현민, 문창주*

건국대학교

caleb5016@konkuk.ac.kr, *cjmoon@konkuk.ac.kr

"LLM-Based Unstructured Data Processing Methods in Autonomous Driving Simulators

Choi Hyun Min, Moon Chang Ju*

Konkuk Univ.

요약

본 논문은 자율주행 시스템에서 발생하는 비정형 데이터를 LLM(Large Language Model)을 활용해 처리하는 방법을 제안한다. CARLA 시뮬레이터에서 발생하는 로그 파일, 자연어 명령어, 교통 상황 등의 데이터를 분석하며 자율주행 시스템의 성능과 안전성을 향상시키는 방법을 탐구한다.

I. 서론

자율주행 기술은 전 세계적으로 교통 안전과 효율성을 향상시키기 위한 핵심 연구 분야로 자리 잡고 있다. 그러나 자율주행 시스템이 처리해야 하는 데이터의 상당 부분은 정형화된 데이터뿐만 아니라 비정형 데이터도 포함되며, 이는 기존의 전통적인 알고리즘으로는 처리하기 어려운 문제가 존재한다.

최근 대형 언어 모델(LLM, Large Language Model)은 자율주행 분야에서 그 가능성이 주목받고 있다. LLM은 자연어 처리(NLP)와 문맥 이해 능력을 통해 비정형 데이터를 해석하고 분석하는 데 있어 효과적이다. 이를 통해 자율주행 시스템에서 발생하는 비정형 데이터를 보다 효율적으로 분석하고 해석할 수 있다.

본 연구는 CARLA 시뮬레이터를 기반으로 자율주행 환경에서 발생하는 비정형 데이터를 LLM을 활용해 처리하는 방법을 제안한다. 이를 통해 자율주행 시스템의 성능과 안전성을 높이고자 한다. 특히, 시뮬레이터에서 생성되는 로그 파일 분석, 자연어 명령어 해석, 교통 상황 분석에 중점을 두어 자율주행 시스템의 성능과 안전성을 향상시키는 방법을 연구한다.

II. 관련연구

자율주행 시스템에서 LLM의 도입은 기존의 자율주행 기술의 문제를 해결하는 데 큰 잠재력을 가지고 있다. 특히, 비정형 데이터 처리에 대한 연구는 자율주행 시스템의 신뢰성과 안전성을 향상시키는 데 필수적인 부분이다. 본 연구와 관련된 주요 연구들은 자율주행 시뮬레이터 환경에서 LLM을 활용한 다양한 접근을 제시해왔다.

M. Kumar는 자연어로 작성된 교통 규칙을 LLM을 활용하여 MTL(Metric Temporal Logic)로 변환하는 자동화된 시스템을 제안했다 [1]. 이 연구는 LLM이 복잡한 논리와 시간 제약을 처리하는 데 뛰어난 성능을 발휘한다는 점을 강조하며 LLM을 활용한 비정형 데이터의 해석 가능성을 제시하였다.

K. Ananthajothi는 LLM을 자율주행 시스템에 통합하여 기존의 자율주행 알고리즘이 적용하기 어려운 상황을 해결하고, 인간처럼 상황을 이해

하고 대응할 수 있는 능력을 갖추도록 하는 방법을 제안했다[2]. 센서 데이터, 교통 신호 및 도로 표시기 등의 비정형 데이터를 해석하여, 인간과 같은 의사 결정 과정을 모방해 복잡한 주행 환경을 분석했다.

J. Zhang은 LLM을 사용해 자율주행 시뮬레이터에서 주요 시나리오를 생성하는 방법을 제시했다[3]. 이 연구는 LLM을 통해 자연어로 표현된 교통 상황을 안전성 테스트를 위한 시나리오로 변환하고, 이를 CARLA 시뮬레이터에서 실행 가능한 코드로 자동 변환하여 다양한 안전 중요 시나리오를 생성하는 것을 목표로 했다.

III. 연구 방법

이에 본 연구에서는 CARLA 시뮬레이터를 사용하여 자율주행 환경을 구축한다. CARLA 시뮬레이터는 실제 도로 상황과 유사한 환경을 제공하며, 다양한 주행 시나리오에서 발생하는 데이터를 수집할 수 있다. 이 환경에서 발생하는 비정형 데이터는 주로 로그 파일, 자연어 명령어, 교통 상황으로 구성된다.

시뮬레이터에서 생성되는 로그 파일은 자율주행 차량의 주행 중 발생하는 모든 정보를 기록한 데이터 파일로, 차량의 위치, 속도, 센서 입력값, 경로 정보, 주행 중 이벤트 등 다양한 요소를 포함하고 있다. 이 로그 파일들은 대부분 구조화되지 않은 형태로 방대한 양의 데이터를 포함하고 있어, 이를 효과적으로 분석하는 과정이 요구된다. LLM을 활용하여 로그 데이터를 요약하고, 주행 패턴과 센서 데이터 사이의 상관관계를 자동으로 분석하여 중요한 패턴을 식별할 수 있다.

또한 자율주행 시스템은 다양한 형태의 자연어 명령어를 처리하는 과정이 필요하다. "다음 교차로에서 좌회전해", "속도를 줄여", "주차공간을 찾아" 등의 명령어는 자율주행 차량이 주행 중에 실시간으로 받아들이고 실행해야 할 명령어들이다. 이러한 자연어 명령어는 정형화된 패턴 없이 사용되기 때문에 이를 이해하고 정확하게 실행하는 것이 중요하다. 본 연구에서는 LLM을 활용해 이러한 자연어 명령어를 실시간으로 해석하여 차량이 수행해야 할 행동을 정확하게 도출한다.

교통 상황은 자율주행 차량의 운행 중 중요한 의사결정에 필요한 데이터 중 하나이다. 이러한 교통 상황 데이터는 교차로의 신호 상태, 도로 상황,

주변 차량 및 보행자의 위치, 날씨 상태 등 다양한 요소로 구성되어 있으며, 대부분 구조화되지 않은 텍스트 또는 영상 데이터 형태로 존재한다. 이러한 비정형 데이터를 실시간으로 분석하여, 자율주행 시스템이 상황에 맞는 결정을 내리도록 도와줄 수 있다. 교차로에서 신호등의 상태와 주변 차량의 속도 및 위치를 분석하여, 자율주행 차량이 언제 멈추고 출발해야 하는지 결정하는 데 필요한 정보를 제공할 수 있다.

IV. 결론

본 연구는 LLM을 활용하여 자율주행 시스템에서 발생하는 비정형 데이터를 효과적으로 처리하는 방법을 제안하였다. CARLA 시뮬레이터와 같은 고도화된 자율주행 시뮬레이터에서의 로그 파일, 자연어 명령어, 교통 상황과 같은 비정형 데이터를 LLM을 통해 분석함으로써 다양한 비정형 데이터를 실시간으로 처리하고 분석할 수 있는 가능성을 확인하였다.

그러나 본 연구는 초기 연구로서 몇 가지 한계를 갖고 있다. 예를 들어, 실험 데이터의 규모와 다양성 측면에서 한계가 있었으며, 실제 도로 상황을 완전히 반영하지 못할 수 있다는 점에서 추가적인 연구가 필요하다. 또한, 자율주행 시스템의 실제 도로 주행 환경에서의 적용 가능성을 높이기 위해 실시간 처리 성능 및 대규모 데이터 처리 능력에 대한 연구가 필요하다.

결론적으로, LLM을 자율주행 시스템에 통합하여 비정형 데이터를 처리하는 것은 자율주행 기술의 발전에 중요한 기여를 할 수 있으며, 향후 자율주행 차량의 안전성과 성능을 높이기 위한 주요 도구로 활용될 수 있다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-학·석사연계ICT핵심인재양성 지원을 받아 수행된 연구임 (IITP-2024-RS-2020-II201834)

참 고 문 헌

- [1] M. Kumar, "TR2MTL: LLM based framework for Metric Temporal Logic Formalization of Traffic Rules," in 2024 IEEE Intelligent Vehicles Symposium (IV), Jeju Island, Korea, Republic of, Jun. 2024
- [2] K. Ananthajothi, "LLM's for Autonomous Driving: A New Way to Teach Machines to Drive," in 2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, India, Dec. 2023
- [3] J. Zhang, "ChatScene: Knowledge-Enabled Safety-Critical Scenario Generation for Autonomous Vehicles," in 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, Jun. 2024