STPA를 활용한 학습 기반 End-to-End 자율주행차 위험분석 사례 연구

김한결, 서준호*, 장은지**

한국정보통신기술협회

hgk0426@tta.or.kr, *jhseo@tta.or.kr, **eun6759@tta.or.kr

A Case Study on the Risk Analysis of Learning-based End-to-End Autonomous Vehicle Using STPA

Kim Han Gyul, Seo Jun Ho*, Jang Eun Ji**

Telecommunications Technology Association

요 약

본 논문은 현재 새롭게 대두되고 있는 학습 기반 End-to-End 자율주행차에서 개발 완료 이전에 위험을 사전에 파악하기 위한 위험분석 방법으로 STPA의 유효성을 확인한다. 이를 위해, 자율주행 개발기업의 도움을 받아 기존 자율주행차에 학습 기반 End-to-End 시스템이 적용되는 경우를 가정하여 위험분석 사례 연구를 수행하였다. 사례 연구 결과 STPA가 자율주행차와 같은 복잡한 시스템의 위험을 사전에 파악하고 안전 요구사항 도출에 효과적임을 확인하였다.

Ⅰ. 서 론

최근 자율주행차는 인지(Perception), 계획(Planning), 제어(Control) 등 자율주행을 위해 필요한 주요 기능을 개별적으로 개발하여 차량에 통합하는 모듈식 패러다임을 활용하고 있다[1]. 기존 모듈식 자율주행차에서 주행 계획을 수립하는 계획 모듈을 개발하기 위한 일반적인 접근 방법은 정교하게 설계된 규칙을 활용하는 것이다. 그러나 규칙 기반 접근 방법은 주행 중에 발생하는 새로운 상황에 대응하기에 부족하고, 다양한 상황 대응을 위해선 규칙의 계산 복잡성이 급격하게 증가다는 한계를 지니고 있어 자율주행차가 높은 수준의 자율성에 도달하는데 장벽이 된다. 이에 대한 대안으로 센서에서 입력된 원천 데이터로부터 바로 액추에이터로 제어 신호를 출력하는 학습(Learning) 기반 End-to-End 자율주행 시스템에 관한 연구와 관심이 증가하고 있다[2].

그럼에도 현재 자율주행차에 학습 기반 End-to-End 시스템을 적용하기 어려운 이유는 학습 기반 시스템의 특성상 일반화(Generalization) 문제가 존재하기 때문이다[3]. 학습 기반 시스템은 특정 환경에서 잘 동작한다고 해서 모든 환경에서 잘 동작하는 것을 보장하지 않는다. 학습 기반 시스템은 특정 입력에 대해서 예상하지 못한 방식으로 동작할 수 있는 위험을 내포하고 있다. 이는 치명적인 사고로 이어질 수 있기에 높은 수준의 안전성과 신뢰성을 요구하는 자율주행 시스템에 적용하기 위해서는 이에 대한 해결이 선행되어야 한다.

본 논문에서는 학습 기반 End-to-End 자율주행 시스템에서 새로이 발생할 수 있는 위험과 그에 따른 요구사항을 파악하기 위해 기존 규칙 기반 자율주행 시스템이 학습 기반 End-to-End 자율주행 시스템으로 대체되는 경우를 가정하여 변경되는 영역에 초점을 맞추어 위험분석을 수행하였다. 본론에서는 수행된 위험분석 과정에 관해 설명하고, 위험분석 결과로부터 도출된 안전 요구사항을 제공한다. 마지막으로 결론에서는 앞선 위험분석에 대한 시사점과 위험분석 결과와 관련하여 수행된 후속 연구 내용을 소개한다.

Ⅱ. 본론

End-to-End 자율주행 시스템은 각 모듈의 출력을 최적화한 이후에 후속 모듈로 전달하는 모듈식 자율주행 시스템과 달리 자율주행과 같은 복잡한 과업을 하나의 모듈로 해결하려는 접근 방법이다. End-to-End 접근 방법에는 필연적으로 딥러닝과 같은 학습 기술이 사용되는데, 인지, 예측, 계획 등 구성요소 간 특징 표현을 전파하여 공유하며 역전파를 통해 전체 손실을 최소화하는 방향으로 학습된다[1]. 이를 도식화하면 아래 그림과 같다.

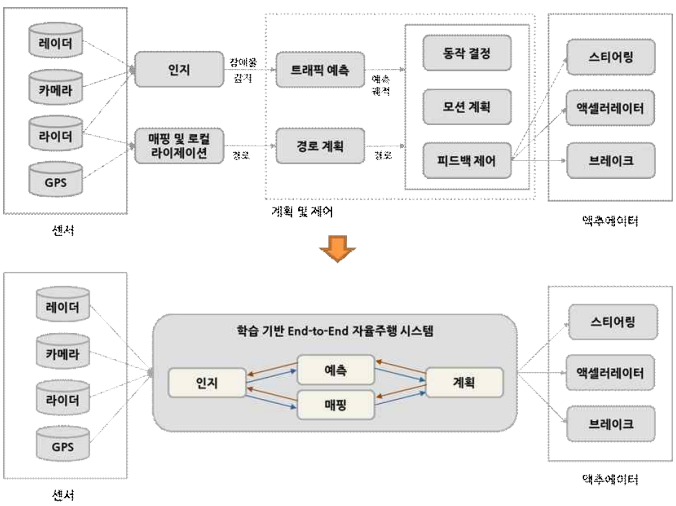

그림 1. 규칙 기반 모듈식 자율주행 시스템 대비
학습 기반 End-to-End 자율주행 시스템 구성 변화

학습 기반 End-to-End 자율주행차에서 새로이 발생하는 위험들은 위의 시스템의 구성 변화에 기인한다. 본 논문에서는 학습 기반 End-to-End 자율주행차의 위험분석을 수행하기 위해 STPA(System Theoretic Process Analysis)를 활용하였다[4]. STPA는 시스템 이론에 기반한

위험분석 기법으로, 시스템 구성요소의 고장만이 아니라 구성요소의 상호작용으로 발생하는 사고를 분석하는 것에 장점이 있다. STPA는 1) 사고 정의, 2) 제어구조 정의, 3) 위험 제어 식별, 4) 사고 시나리오 도출의 4단계로 구성된다. STPA는 위험이 잘못된 제어로 인해 발생한다는 점에서 착안하여 제어와 피드백으로 추상화된 시스템 모델(제어구조)을 구성하고, 제어구조에서 위험 제어를 식별하고 이에 대한 사고 시나리오들을 체계적으로 도출한다. 학습 기반 End-to-End 자율주행차에서 새로 발생하는 사고 위험을 분석한 STPA 결과는 다음과 같다.

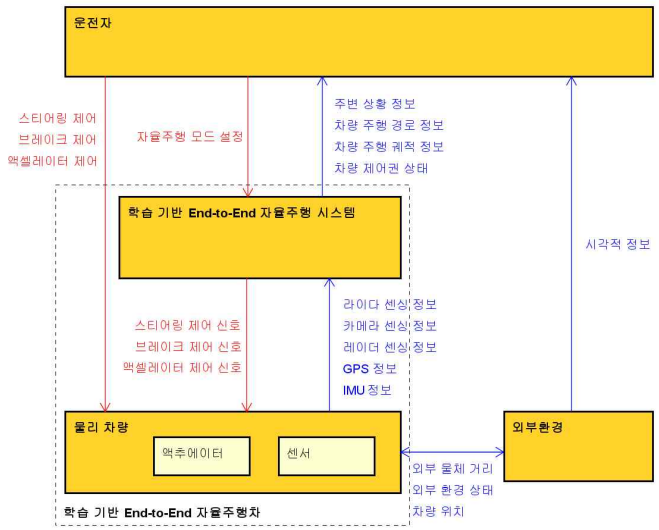

그림 2. 학습 기반 End-to-End 자율주행차의 제어구조

표 1. STPA 분석 결과 도출된 사고 시나리오 유형

사고 시나리오 유형	건수	비율
피드백 정보 훼손/손실	18	22.78%
상황 인지 오류	24	30.38%
제어 결정 오류	27	34.18%
제어 충돌	5	6.33%
환경 불확실성	5	6.33%
합계	79	100.00%

위 분석 결과는 기존 자율주행차에서 발생하는 사고 시나리오는 제외하고 학습 기반 End-to-End 시스템이 적용되어 새로 발생하는 위험 원인에 초점을 맞추어 분석하였다. 이에 따라 사고 시나리오는 학습 기반 End-to-End 자율주행 시스템이 잘못된 결정(제어 결정 오류, 34.18%)을 하거나 운전자나 시스템이 정보를 잘못 이해하여(상황 인지 오류, 30.38%) 주로 발생한다. STPA 결과로 도출된 대표적인 사고 시나리오 사례는 다음과 같다.

- 사례 1: 학습 기반 End-to-End 자율주행 시스템이 다중 작업 학습(Multi-task Learning) 최적화에 실패로 인해 브레이크 제어 신호를 너무 늦게 전달하여 사고가 발생함(제어 결정 오류)
- 사례 2: 학습 기반 End-to-End 자율주행 시스템이 희귀 케이스(Rare Case) 예측 실패로 인해 액셀레이터 제어 신호를 잘못 전달하여 사고가 발생함(제어 결정 오류)
- 사례 3: 운전자가 학습 기반 End-to-End 자율주행 시스템이 제공하는 정보를 잘못 이해하여 스티어링을 제어하지 않아 사고가 발생함(상황 인지 오류)
- 사례 4: 학습 기반 End-to-End 자율주행 시스템이 주변 환경 모델링(World Model) 실패로 인해 브레이크 제어 신호를 잘못 전달하여 사고가 발생함(상황 인지 오류)

STPA 분석 결과 도출된 사고 시나리오들을 살펴보면, 자율주행차에 학습 기반 End-to-End 자율주행 시스템을 적용하는 경우에 발생하는 새로운 사고들은 대부분 불확실성(Uncertainty)과 관련이 있다. 이는 학습 기반 End-to-End 자율주행 시스템의 두 가지 특성으로부터 기인한다. 하나는 데이터를 학습하여 동작하는 학습 기반 시스템 내포하고 있는 확률 기반 추론의 불확실성이다. 학습 기반 시스템의 추론 결과는 확률을 기반으로 판단하기에 결과의 재현성이 떨어질 수 있고, 새로운 입력에 대해서는 성능이 급격하게 떨어지는 일반화 문제가 발생할 수 있다. 다른 하나는 학습 기반 End-to-End 시스템의 내부의 상태를 파악하기 어렵다는 점이다. 학습 기반 시스템의 경우 기본적으로 동작 방식을 파악하기 어려운 특성이 있는데 End-to-End 시스템은 복잡한 작업들을 통합하여 수행하기에 내부 상태나 의사결정 과정을 더욱 파악하기 어려워진다. 이러한 블랙박스 특성은 운전자에게 시스템의 의사결정 과정과 이유를 이해하기 어렵게 만들어 운전자 의사결정의 불확실성도 증가시킨다. 이에 불확실성을 주요 위험 요소로 파악하고 안전 요구사항을 도출하였으며 결과는 다음과 같다.

- **안전요구사항1**: 학습 기반 End-to-End 자율주행 시스템 제어 결정의 불확실성을 저감시킬 수 있는 수단을 마련한다.
- **안전요구사항2**: 운전자가 학습 기반 End-to-End 자율주행 시스템의 결정을 이해할 수 있도록 돕는 수단을 마련한다.

Ⅲ. 결론

본 연구에서는 학습 기반 End-to-End 자율주행 시스템이 자율주행차에 적용되는 경우 발생할 수 있는 위험을 파악하기 위해 STPA를 통해 위험분석을 수행하였다. 이를 통해, STPA가 자율주행차와 같은 복잡한 시스템에서 일부 구성요소가 변경될 때 발생할 수 있는 새로운 위험을 효과적으로 파악하고 이에 대한 안전 요구사항을 도출할 수 있음을 알 수 있었다. 다만 STPA는 정성적인 분석 방법이기에 위험분석을 진행하기 위해서 도메인 전문가와 관련 문헌에 대한 도움이 많이 필요하다는 한계점이 일부 존재한다. 또한 위험분석 결과 도출된 안전 요구사항들의 유효성에 관한 확인이 필요하다는 점이다. 이는 자율주행 개발기업과의 후속 연구를 통해 확인할 수 있었는데, 기본 End-to-End 자율주행 모델과 불확실성 저감 수단(Regularization & Uncertainty Quantification)이 적용된 End-to-End 자율주행 모델을 비교하는 시뮬레이션 연구가 수행되었다.

참 고 문 헌

[1] Chen, L., Wu, P., Chitta, K., Jaeger, B., Geiger, A., & Li, H. (2023). End-to-end autonomous driving: Challenges and frontiers. arXiv preprint arXiv:2306.16927.

[2] Le Mero, L., Yi, D., Dianati, M., & Mouzakitis, A. (2022). A survey on imitation learning techniques for end-to-end autonomous vehicles. IEEE Transactions on Intelligent Transportation Systems, 23(9), 14128-14147.

[3] Brunke, L., Greeff, M., Hall, A. W., Yuan, Z., Zhou, S., Panerati, J., & Schoellig, A. P. (2022). Safe learning in robotics: From learning-based control to safe reinforcement learning. Annual Review of Control, Robotics, and Autonomous Systems, 5, 411-444.

[4] N. G. Leveson and J. P. Thomas, "STPA Handbook," Cambridge, MA, USA: MIT Press, 2018, (http://psas.scripts.mit.edu/home/materials/)