Journal of KIIT. Vol. 21, No. 8, pp. 165-174, Aug. 31, 2023. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2023.21.8.165

강화 학습 모델의 안전성 평가 방법: 자율 주행 사례 연구

조수희*, 권령구**, 권기현***

Safety Evaluation for Reinforcement Learning Model: Case Study of Autonomous Driving

Suhee Jo*, Ryeonggu Kwon**, and Gihwon Kwon***

이 논문은 과학기술정보통신부 및 정보통신기술진흥센터의
고안전 SW 개발을 위한 안전 분석 및 검증 도구 기술 개발 사업의 연구 결과로 수행되었음(No. 2021-0-00122)

요 약

강화 학습은 의료, 자율 주행 자동차 등 다양한 분야에서 사용된다. 그러나 강화 학습은 에이전트의 의사 결정 과정에 대해 이해하기 어렵고, 지도학습처럼 출력된 결과에 대한 정확성을 측정할 수 없다. 그렇기 때문에, 강화 학습 시스템의 안전성을 평가하는 연구가 매우 필요하다. 따라서, 본 논문에서는 행동 주도 개발 기법과 위험원 분석을 결합하여 강화 학습 모델의 안전성을 평가하는 방법을 제안한다. 자율 주행 시뮬레이션 환경에 대해 제안한 방법을 적용한 결과, 강화 학습 모델의 평균 보상이 유사하더라도 안전 제약 조건을 위반한 빈도의 편차가 나타나는 것을 관찰할 수 있었다. 안전성 평가로 도출된 결과는 다양한 측면에 걸친 분석과 반복적인 개선을 통해 모델의 안전성 향상에 기여할 수 있을 것으로 기대된다.

Abstract

Reinforcement Learning(RL) finds application in diverse fields such as healthcare and autonomous driving. However, due to its inherent complexity in agent decision-making and the inability to measure outcomes with the same precision as in supervised learning, there is a pressing need for research into assessing the safety of RL systems. Therefore, this paper introduces a method for evaluating the safety of RL models by combining behavior-driven development techniques with risk analysis. When applied to a driving simulation environment, the proposed approach reveals variations in the frequency of safety constraint violations among RL models, even when their average rewards are similar. The outcomes of safety assessment are anticipated to contribute to the improvement of model safety through thorough analysis and iterative enhancements across various aspects.

Keywords
reinforcement learning, safety, hazard analysis, behavior driven development

* 경기대학교 SW안전보안학과 석사과정
- ORCID: https://orcid.org/0000-0001-8009-9829
** 경기대학교 컴퓨터과학과 박사과정(교신저자)
- ORCID: https://orcid.org/0000-0002-4942-247X
*** 경기대학교 컴퓨터공학부 교수
- ORCID: https://orcid.org/0000-0002-8221-4939

· Received: Jun. 25, 2023, Revised: Aug. 17, 2023, Accepted: Aug. 20, 2023
· Corresponding Author: Ryeonggu Kwon
Dept. of Computer Science, Kyonggi University, 154-42,
Gwanggyosan-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, South Korea
Tel.: +82-31-249-9666, Email: rkkwon@kyonggi.ac.kr

Ⅰ. 서 론

강화 학습은 기계 학습의 한 분야로서, 환경과의 상호작용에서 발생하는 시행착오를 통해 학습하는 분야이다[1]. 기계 학습은 의료[2], 자율 주행 자동차[3], 항공 우주[4] 등을 포함한 다양한 영역에서 사용된다. 하지만 기계 학습은 의사 결정 과정이 블랙박스 특성을 가져서, 전문가조차 이해하는 데 어려움을 초래하여 시스템의 신뢰성과 안전성에 대한 우려가 제기되고 있다[5][6]. 따라서 예상치 못한 사고 및 상황에 대한 분석이 어렵기 때문에, 기계 학습을 안전이 중요한 시스템에 적용하는 것은 어려운 일이다[7]. 특히, 에이전트의 시행착오를 통해 학습하는 강화 학습은 예측 가능하고 불규칙하지 않아야 하는 안전성의 특징과는 거리가 멀다[8].

한편, 주어진 데이터 세트를 기반으로 사전의 정의된 출력을 매핑하도록 학습하는 지도 학습은 정확도(Accuracy), 정밀도(Precision), 재현율(Recall) 등을 기반으로 예측된 출력의 정확성을 측정할 수 있다[9]. 반면, 강화 학습은 환경과 상호 작용하여 시간이 지남에 따라 누적 보상을 최대화하는 정책을 학습하기 때문에 평균 보상, 수렴 속도 등을 기반으로 에이전트의 성능을 평가한다[10]. 따라서, 어떤 모델의 성능이 제일 좋은지 비교적 쉽게 판단할 수 있는 지도 학습과 다르게, 강화 학습은 모델의 성능에 대한 통찰력을 제공할 수는 있지만 그것이 정확성이나 안전성을 나타내는 것은 아니다.

이에 따라 강화 학습 모델의 안전성을 종합적으로 평가하고 개선을 하기 위한 노력이 요구된다. 본 연구에서는 강화 학습 모델의 안전성을 평가하기 위한 기준을 세우기 위해 위험원 분석(Hazard analysis)[11]과 행동 주도 개발(BDD, Behavior Driven Development)[12]을 결합한 방법을 제안한다. 위험 분석 기법을 통해 강화 학습 모델의 안전성 기준을 세우고 BDD를 이용하여 환경에 대한 에이전트의 행동을 분석한다. 먼저, 위험 분석 기법을 통해 목표한 강화 학습 모델과 관련된 잠재적 위험을 식별하고 분석하여 안전 제약 조건을 추출한다. 그 다음, 추출한 안전 제약 조건을 기반으로 BDD 시나리오를 작성하여 테스트 케이스를 생성한다. 마지막으로, 학습된 모델에 대해 사전에 작성한 테스트 케이스를 적용하여 결과를 분석한다. 이를 통해, 학습된 강화 학습 모델에 대한 검증을 수행하여 모델의 신뢰성을 확보하는 데 기여하는 것이 목표이다.

사례 연구로 가상 주행 시뮬레이션 환경을 사용한다. 이 환경에서 사용된 에이전트(자율 주행 차량)에 대해 제안한 방법을 적용하여 안전성을 평가하는 과정을 보여준다. 일반적으로 자율 주행 차량은 ISO 26262와 같은 표준을 기반으로 안전성을 평가하는데[13], 본 연구에서는 이를 기반으로 수행된 위험 분석을 적용하여 에이전트의 안전성을 분석하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 제안하는 방법을 이해하기 위한 배경 지식 및 관련 연구에 대해 소개한다. 3장에서는 제안하고자 하는 방법에 관해 서술하고, 4장에서는 제안한 방법을 바탕으로 사례 연구를 진행하여 결과를 분석한다. 마지막으로 5장에서는 본 연구의 기여점, 한계 및 향후 연구에 대해 기술한다.

Ⅱ. 배경 지식 및 관련 연구

2.1 강화 학습 모델의 안전성

강화 학습 모델의 안전성을 위해 다양한 연구가 진행되고 있다. 그중에는 안전이 시스템의 운영뿐만 아니라 학습 과정에서도 중요하다고 보는 관점이 있다[14]. 이 관점은 주로 에이전트의 행동에 제약을 두어 안전한 탐험(Safe exploration)을 하도록 유도하는 방식으로 접근한다[15]. 이를 통해 에이전트가 안전 제약 조건을 위반하지 않는 작업을 학습할 수 있도록 한다[16]. 이와 관련하여 [17]-[19]는 제약 조건이 존재할 때 에이전트가 학습 도중 위반하지 않도록 하는 방법을 제안한다.

이러한 방법은 에이전트의 행동을 제한하여 안전성을 강화하는 것을 목표로 하지만, 원치 않는 동작을 방지하기 위해 탐험을 제한한다면 시스템의 안전성은 증가할 수 있지만 전체적인 학습 성능이 제한될 수 있다는 것에 주의하여야 한다[8]. 이와 달리, 본 논문에서는 모델의 학습이 완료된 후에 해당 모델의 안전성을 평가하는 것을 목표로 한다.

2.2 위험원 분석

위험원 분석은 시스템 또는 기술과 관련된 잠재적 위험을 식별 및 분석, 평가하는 데 사용되는 프로세스이다[11]. 위험원 분석을 수행하기 위해 STPA(System Theoretic Process Analysis)[20], FMEA(Failure Mode Effectiveness Analysis)[21], HARA(Hazard Analysis and Risk Assessment)[22] 등 다양한 기법이 개발되었다. 위험원 분석의 주요 산출물 중 하나는 안전 제약 조건(Safety constraints)을 식별하는 것이다. 안전 제약 조건은 시스템의 안전 요구 사항 준수 여부를 평가하는 데 중요한 역할을 한다. 이는 시스템이 사전에 정의한 안전 요구사항의 경계 내에서 작동하도록 보장하여 위험을 완화하고 안전 및 신뢰성을 향상시키는 데 도움을 준다.

기계 학습에 대해 위험원 분석을 적용한 사례는 다음과 같다. Hodge[23]은 훈련된 방식을 통해 안전성에 대한 확신을 얻는 것뿐만 아니라 학습된 모델 자체의 충분성에 대한 증거를 생성해야 한다고 보았다. 이를 위해 FFA(Functional Failure Analysis)[24]를 사용하여 심층 강화 학습을 통해 개발된 드론 내비게이션의 안전성을 분석한다. Qi[25]는 안전이 중요한 시스템에 기계 학습이 적용되는 추세에 따라 시스템 기반 위험 분석인 STPA를 확장하여 DeepSTPA를 제안하였다. Henriksson[26]은 HARA의 사용이 권장되는 ISO 26262에 대해 자동차에서 기계 학습의 개발을 허용하도록 조정해야하는지에 대해 논한다.

2.3 행동 주도 개발

BDD는 다양한 이해 관계자들 간의 협업과 커뮤니케이션을 강조하는 애자일 소프트웨어 개발 방법론으로, 개발자, 테스터, 비즈니스 담당자 등이 함께 시스템의 동작을 정의하고 소프트웨어의 동작을 원하는 결과와 일치시키는 것을 목표로 한다[27]. 이를 위해 BDD는 “Given[Context], When[Event], Then[Outcome]” 형식의 구문을 사용하여 시스템의 동작을 일종의 시나리오로 작성한다. 표 1은 BDD 시나리오의 구성에 대해 설명한다.

표 1. BDD 시나리오 구성
Table 1. Structure of BDD scenario

	Content
Given	set up the initial state of the system
When	describe the action or event being performed
Then	specify the expected outcome or behavior

Wang et al.[28]은 안전성 검증을 위해 STPA와 BDD를 결합한 방식을 제안하였다. 위험원 분석을 통해 추출되는 안전 제약 조건은 다음과 같이 BDD 형식으로 작성하여 표현할 수 있다.

- Given: 위험이 발생할 수 있는 환경에서
- When: 어떤 행동이나 입력이 들어올 때
- Then: 어떠한 사고나 출력이 발생하지 않아야 한다

본 연구에서는 이런 방식으로 위험원 분석을 통해 도출된 안전 제약 조건을 BDD 형식으로 재작성하여 활용한다. BDD 방식을 사용한 이유는 두 가지 측면에서 설명할 수 있다. 먼저, BDD는 특히 행동 테스트에 집중하는 방법론[29]으로, 에이전트의 행동에 따라 상황이 변하는 강화 학습의 맥락과 유사하다. 두 번째로는 자연어를 기반으로 작성된 시나리오는 모든 사람들이 쉽게 이해가 가능하기 때문에 코드에 대한 이해 없이도 참여할 수 있는 장점이 있다[30]. 따라서, 기존의 개발자뿐만 아니라 도메인 전문가나 안전 엔지니어와 같은 다양한 분야의 전문가들도 테스트 결과를 분석하여 강화 학습 모델의 안전성을 종합적으로 평가할 수 있을 것으로 기대된다.

III. 강화 학습 모델의 안전성 평가

본 논문에서는 위험원 분석과 BDD 방법론을 결합하여 학습된 강화 학습 모델의 안전성을 평가하기 위한 프로세스를 제안한다. 3장은 제안하고자 하는 평가 프로세스의 개요와 각 단계의 역할에 대해 소개한다. 그리고 제안한 프로세스를 사례 연구인 가상 주행 시뮬레이션 환경을 사용하여 적용하는 과정에 대해 보여준다.

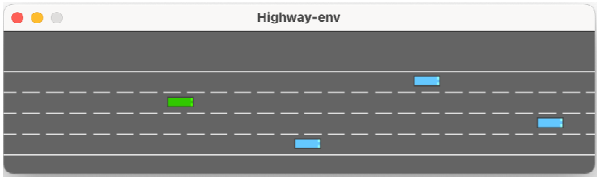


그림 1. 자율 주행 환경
Fig. 1. Environment of autonomous driving

그림 1은 자율 주행 에이전트의 성능을 평가하기 위해 설계된 Highway 가상 시뮬레이션 환경이다 [31]. 이 환경에서 에이전트는 주변 차량과의 충돌을 피해 최대 속도로 주행하는 것을 목표로 한다. 본 연구에서는 이 환경을 활용하여 에이전트의 행동에 대한 안전성을 조사한다. 그림 1은 에이전트(초록색 자동차)가 4차선 고속도로의 환경에서 주행하는 모습을 보여준다. 에이전트는 차선을 변경하거나 가속 및 감속하는 행동을 할 수 있다.

3.1 평가 프로세스

그림 2는 본 연구에서 제안하는 강화 학습 모델의 안전성을 평가하기 위한 전반적인 프로세스를 보여준다. 개발하고자 하는 강화 학습의 기능으로부터 안전성 분석, 강화 학습 두 가지 과정으로 나눌 수 있다.

안전성 분석은 강화 학습 모델로부터 잠재적인 위험을 식별하고 최종적으로는 안전성을 평가하기 위한 테스트 케이스를 작성하는 것이 목적으로 그림 2의 왼쪽 라인에서 해당 과정을 볼 수 있다. 안전성 분석의 단계는 다음과 같다.

Step 1. 위험원 분석을 통한 안전 제약 조건 식별
Step 2. BDD 기반 안전 시나리오 작성
Step 3. 테스트를 위한 코드 작성

강화 학습은 사이클의 반복을 통한 모델의 안전성 향상을 목적으로 한다. 그림 2의 오른쪽 라인에서 해당 과정을 볼 수 있다. 강화 학습의 단계는 다음과 같다.

Step 1. 모델링
Step 2. 모델 학습
Step 3. 모델 테스트

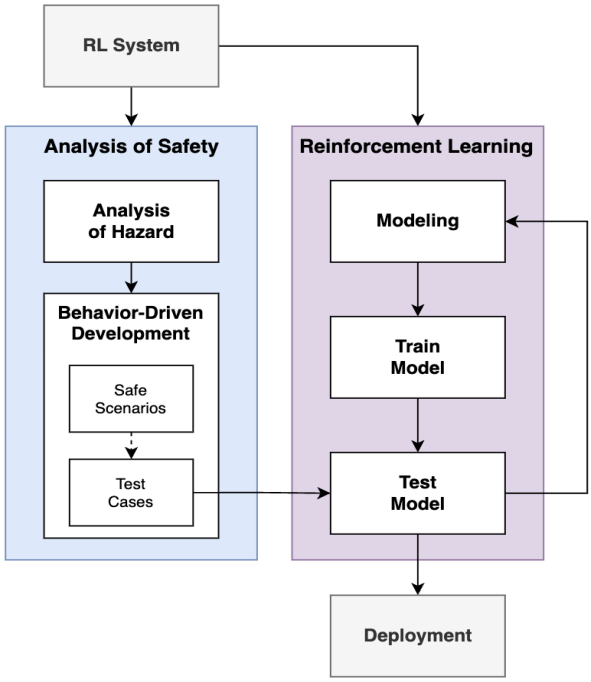


그림 2. 강화 학습 모델의 평가 프로세스
Fig. 2. Evaluation process of RL model

3.2 안전성 분석

Step 1. 위험원 분석을 통한 안전 제약 조건 식별

잠재적인 위험을 식별하기 위해 FMEA, STPA와 같은 다양한 위험원 분석 기법을 사용하여 강화 학습 시스템에 대한 안전 제약 조건을 식별한다. 식별된 안전 제약 조건은 강화 학습 모델이 안전하게 동작하는지 판단하기 위한 기준으로서 사용된다.

Highway에 대한 위험원 분석을 수행하여 안전 제약 조건을 식별하기 위해 [32]에서 무인 자동차에 대해 ISO 26262 기반으로 HARA를 수행한 결과를 인용하였다. 이를 통해, 가상 시뮬레이션 환경에 적용하기 위해 식별된 안전 제약 조건은 표 2와 같다.

표 2. Highway에 대한 안전 제약 조건
Table 2. Safety constraints for Highway

ID	Content
SC-1	Do not collide with the vehicle when controlling speed.
SC-2	Do not collide with the vehicle during lane change control.
SC-3	Do not accelerate when the distance from the previous vehicle is close.

Step 2. BDD 기반 안전 시나리오 작성

그 다음, 안전 제약 조건을 “Given, When, Then”의 구문을 사용하는 BDD 시나리오로 변환한다. BDD 시나리오의 구성은 어떤 환경(Given)에서 어떤 행동(When)을 할 때 발생할 결과(Then)에 대해 설명하기 때문에 특정 환경에 따른 에이전트의 행동과 그 결과에 대해 평가 및 분석할 수 있다. 이 구성에 따르면 결과(Then)에 대한 위반은 곧 안전 제약 조건을 충족하지 못했다는 의미로 해석할 수 있다. 표 3은 앞서 식별된 안전 제약 조건에 대해 BDD 시나리오로 변환한 결과이다.

[32]에서 속도나 거리에 대한 구체적인 수치를 제공하지는 않기 때문에, 표 3에 대해 “Given”에서 다른 차량을 근처로 인식하는 기준은 에이전트를 기준으로 앞뒤는 점선 약 3칸, 차선은 위아래 하나의 차선 기준이라고 가정하였다.

표 3. 안전 제약사항에 대한 BDD 시나리오
Table 3. BDD scenarios for safety constraints

ID	Context	Content
SC-1	Given	there’s another vehicle in front or behind.
	When	when controlling the speed of the vehicle.
	Then	it must be prevented from colliding.
SC-2	Given	there’s another vehicle in the mearby lane.
	When	when controlling the lane change of the vehicle.
	Then	it must be prevented from colliding.
SC-3	Given	there’s another vehicle in front it.
	When	when controlling the speed of the vehicle.
	Then	it must not accelerate.

Step 3. 테스트를 위한 코드 작성

변환된 BDD 시나리오들은 강화 학습 모델의 안전성을 평가하기 위한 테스트 케이스로써 사용된다. 테스트를 수행하기 위한 코드를 작성할 때, BDD 시나리오의 구성을 토대로 함수를 구성한다. 예를 들어, SC-1의 대한 의사 코드는 그림 3과 같다.

```
// given('there'a another vehicle in front of begind')
func other_vehicles_in_front_or_begind():
    if near_the_same_lane() == true:
        return true
    else return fasle

// when('when controlling the speed of the vehicle')
func speed_controlling_vehicle():
    if acceleration_vehicle() == true \
            or deceleration_vehicle() == true:
        return true
    else return false

// then('it must be prevented from colliding')
func not_crash_between_vehicles():
    if not crash_vehicle():
        return true
    else return false
```

그림 3. SC-1 기반 BDD 시나리오에 대한 의사 코드
Fig. 3. Pseudocode for SC-1 based on BDD scenario

3.3 강화 학습

Step 1. 모델링

모델링 단계에서는 에이전트, 환경, 파라미터 설정 등 훈련 전의 모든 준비 과정을 포함한다. 표 4는 본 연구에 사용된 환경에 사용된 보상 구조를 보여준다. 기본적으로 최대 속도로 주행하거나 가장 오른쪽 차선을 통해 주행할 때 보상을 받으며, 다른 차량과 충돌할 때는 페널티를 받도록 구성된다.

표 4. Highway 보상 구조
Table 4. Reward structure of Highway

Reward	Content
collision_reward (default: -1)	The reward received when colliding with a vehicle.
right_lane_reward (default: 0.1)	The reward received when driving on the right-most lanes, linearly mapped to zero for other lanes.
high_speed_reward (default: 0.4)	The reward received when driving at full speed, linearly mapped to zero for lower speeds.
lane_change_reward (default: 0)	The reward received at each lane change action.

본 연구에서는 고속도로 시뮬레이션 환경에서 PPO(Proximal Policy Optimization) 알고리즘을 사용하여 강화 학습을 수행한다. PPO는 누적 보상을 최대화하기 위해 순차적으로 조치를 취하는 환경에서 정책을 최적화하도록 설계된 강화 학습 알고리즘이다[33]. PPO 알고리즘의 목적 함수는 식 (1)과 같다. θ는 정책 파라미터, $r(\theta)$는 주어진 행동에 대해 새로운 정책과 이전 정책의 확률 비율을 나타낸다. A는 이전 정책과 비교하여 얼마나 더 나은지를 나타내는 어드밴티지 함수이고 ϵ는 클리핑 범위를 나타내는 작은 상수이다.

$$L(\theta) = E[min(r(\theta)A, clip(r(\theta), 1-\epsilon, 1+\epsilon)A)] \quad (1)$$

Step 2. 모델 학습

앞서 수행한 모델링을 기반으로 학습을 시작한다. 학습이 완료된 강화 학습 모델은 안전성 평가를 위해 테스트 단계로 넘어간다.

Step 3. 모델 테스트

마지막 테스트 단계에서는 학습이 완료된 강화 학습 모델에 대해 BDD 시나리오로부터 변환된 테스트 케이스를 사용하여 모델의 안전성을 평가한다. 이러한 테스트 케이스를 통해 사전에 도출한 안전 제약 조건을 준수하는지 확인한다. 분석된 결과는 강화 학습 모델의 개선을 위한 정보로 활용될 수 있다. 예를 들어, SC-1에 대한 위반을 판단하는 로직은 그림 4와 같다.

테스트가 완료되면 위반 사항을 개선하기 위해 모델링 단계로 돌아간다. 목표한 기준을 충족할 때까지 이 과정을 반복하여 모델의 안전성을 향상시킨다.

```
// Given
if other_vehicles_in_front_or_behind() == true:
    // When
    if speed_controlling_vehicle() == true:
        // Then
        if not_crash_between_vehivles() != true:
            // SC-1 violation
```

그림 4. SC-1 위반에 대한 의사 코드
Fig. 4. Pseudocode for SC-1 violation

Ⅳ. 실험 결과

4.1 실험 모델

본 연구에서는 같은 환경에 대한 여러 가지 모델의 안전성을 비교하기 위해 보상 구조나 내부 알고리즘은 변경하지 않고 하이퍼 파라미터만 조정하여 10가지의 모델을 생성하였다. 표 5는 본 연구를 위해 조정한 모델의 하이퍼 파라미터를 보여준다.

표 5. 하이퍼 파라미터 설정
Table 5. Configuration of hyper-parameters

Model	Total timesteps	Learning rate
1	10,000	0.0001
2	10,000	0.0002
3	10,000	0.0003
4	10,000	0.0004
5	10,000	0.0005
6	20,000	0.0001
7	20,000	0.0002
8	20,000	0.0003
9	20,000	0.0004
10	20,000	0.0005

4.2 결과 비교 분석

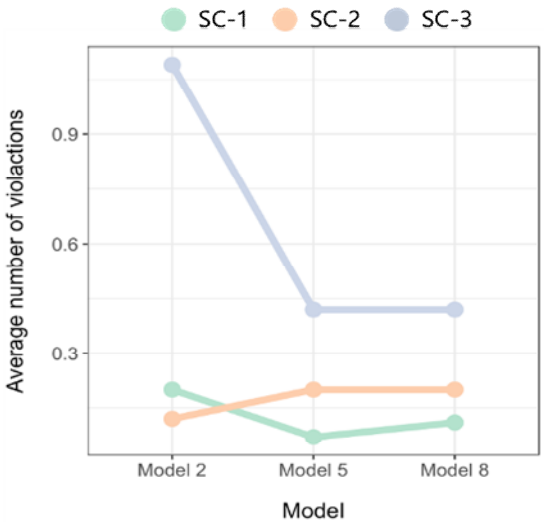
학습이 완료된 모델에 대해 테스트 코드를 적용하여 각 100번의 에피소드를 통해 결과를 집계하였다. 표 6은 안전 제약 조건을 위반한 합계와 평균 보상을 보여준다.

표 6. 주행 시뮬레이션 안전성 평가 결과
Table. 6. Driving simulation safety assessment results

model	SC-1	SC-2	SC-3	mean rewards
1	51	1	223	19.03
2	20	12	109	23.82
3	33	21	75	18.40
4	26	30	73	17.57
5	7	20	42	23.39
6	51	5	179	17.11
7	37	4	172	18.22
8	11	20	42	23.69
9	34	0	173	20.80
10	26	13	117	21.87

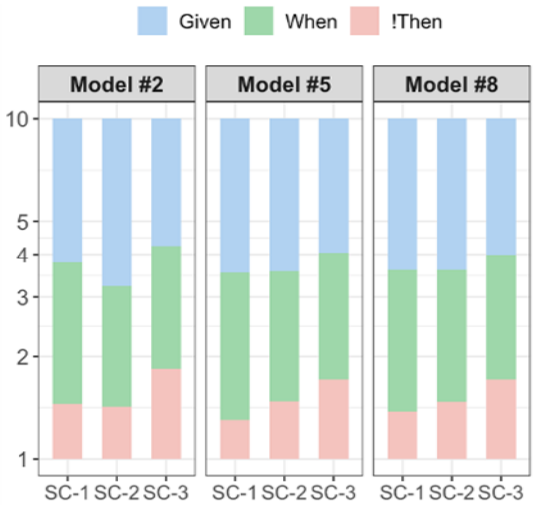
이 중 평균 보상이 가장 높은 모델 2, 5, 8의 평균 보상이 각각 23.82, 23.39, 23.69의 유사한 수치를 보였으나 안전 제약 조건을 위반한 횟수는 서로 차이가 나는 것을 관찰할 수 있었다. 그림 5는 모델 2, 5, 8의 안전 제약 조건 위반 정도를 서로 비교하기 위해 작성한 그래프이다.

그림 5-(a)는 안전 제약 조건을 위반한 평균을 보여준다. 모델에 대한 SC-1과 SC-2의 위반은 비교적 큰 차이가 없지만 SC-3는 model 2에서 비교적 높은 수치로 위반하였다는 결과를 관찰할 수 있다.



(a) 안전 제약 조건 위반 평균

(a) Average number of violations for each safety constraint



(b) “Given”, “When”, “!Then”의 발생 비율

(b) Percentage of occurrence for “Given”, “When”, and “!Then”

그림 5. 모델 2, 5, 8의 안전성 비교

Fig. 5. Safety comparison of models 2, 5, and 8

그림 5-(b)는 각 안전 제약 조건마다 BDD 시나리오를 만족했던 횟수를 비율을 통해 보여준다. 이때 “!Then”은 안전 제약 조건을 위반하였다는 의미로 사용한다. “Given”과 “When”의 발생 비율에 비해 “!Then”은 작은 수치이기 때문에 가시성을 확보하기 위해 비율에 로그 스케일을 적용하였다. 이를 통해, 각 모델에 대해 위험이 발생할 수 있었던 상황과 안전이 지켜지지 못한 상황의 비율을 비교하는 것이 가능하다.

모델 2는 평균 보상이 가장 높았던 모델이다. 하지만, 모델 5, 8과 비교했을 때, 안전성 평가 결과 특히 SC-3 부분이 위험이 발생할 수 있었던 상황이 가장 많이 발생했고 평균 위반 횟수도 가장 높았다. SC-2의 위반은 가장 낮은 수치를 기록했지만 SC-1의 위반은 가장 높은 수치를 기록했다는 것을 관찰할 수 있다.

Ⅴ. 결론 및 향후 과제

본 논문에서는 강화 학습 모델의 안전성 평가를 위한 프로세스를 위해 위험원 분석과 BDD 방법론을 결합하였다. 이 프로세스를 통해 강화 학습 시스템의 잠재적인 위험을 식별하고 안전 제약 조건을 추출한 후, 이를 자연어 기반의 BDD 시나리오로 변환하여 안전성 평가에 활용하였다.

사례 연구로 가상 주행 시뮬레이션 환경을 사용하였고, 이 환경으로부터 제안한 프로세스를 적용하여 사용하는 과정을 보였다. 하이퍼 파라미터를 변경하여 10가지의 모델을 생성하였고, 학습이 완료된 모델에 대해 안전성 평가를 수행하였다. 이를 통해, 각 모델의 평균 보상이 비슷하더라도 안전 제약 조건에 대한 위반은 서로 차이가 나는 것을 관찰할 수 있었다. 따라서 안전이 중요한 시스템에 강화 학습 모델을 적용할 때는 평균 보상뿐만 아니라 안전성 또한 평가하여 다양한 측면에서 분석해야 할 것이다.

본 연구는 크게 두 가지 관점에서 의의를 가진다. 첫 번째, 위험원 분석을 통해 도출된 안전 제약 조건을 사용한다. 이는 강화 학습 시스템의 잠재된 위험에 대해 분석하기 때문에 안전성 측면에서의 신뢰성을 높일 수 있을 것으로 보인다.

두 번째, 자연어를 기반으로 작성된 BDD 시나리오와 테스트 결과는 개발자가 아닌 이해관계자들이 보더라도 쉽게 이해가 가능하기 때문에, 원활한 의사소통과 협업을 통해 강화 학습 시스템의 안전성에 대해 종합적으로 평가할 수 있을 것으로 기대된다.

한편, 본 연구에서 사용한 가상 시뮬레이션 환경과 임의로 수정한 안전 제약 조건은 현실 세계에서의 다양한 상황과 위험 요인들을 완벽하게 대변하지는 않는다. 따라서 추후에는 실제 환경과 현실적인 안전 제약 조건을 사용하여 제안한 프로세스의 적용 가능성을 검토하고자 한다.

References

[1] Z.-H. Zhou, "Machine learning", Springer Nature, 2021.

[2] S. K. Zhou, H. N. Le, K. Luu, H. V. Nguyen, and N. Ayache, "Deep reinforcement learning in medical imaging: A literature review", Medical image analysis, Vol. 73, Oct. 2021. https://doi.org/10.1016/j.media.2021.102193.

[3] A. Norouzi, H. Heidarifar, H. Borhan, M. Shahbakhti, and C. R. Koch, "Integrating machine learning and model predictive control for automotive applications: A review and future directions", Engineering Applications of Artificial Intelligence, Vol. 120, Apr. 2023. https://doi.org/10.1016/j.engappai.2023.105878.

[4] S. Chinchanikar and A. A. Shaikh, "A review on machine learning, big data analytics, and design for additive manufacturing for aerospace applications", Journal of Materials Engineering and Performance, Vol. 31, pp. 6112–6130, Jul. 2022. https://doi.org/10.1007/s11665-022-07125-4.

[5] Q. Bi, K. E. Goodman, J. Kaminsky, and J. Lessler, "What is machine learning? A primer for the epidemiologist", American journal of epidemiology, Vol. 188, No. 12, pp. 2222-2239, Dec. 2019. https://doi.org/10.1093/aje/kwz189.

[6] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics", Electronics, Vol. 8, No. 8, p. 832, Jul. 2019. https://doi.org/10.3390/electronics8080832.

[7] M. Cummings, "Rethinking the Maturity of Artificial Intelligence in Safety-Critical Settings", AI Magazine, Vol. 42, No. 1, pp. 6-15, Apr. 2021.

[8] P. V. Wesel and A. E. Goodloe, "Challenges in the Verification of Reinforcement Learning Algorithms", NASA Technical Reports Server, Jun. 2017.

[9] B. Liu, "Supervised Learning", Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, pp. 63-132, Jan. 2011. https://doi.org/10.1007/978-3-642-19460-3_3.

[10] M. Wiering and M. Otterlo, "Reinforcement Learning", Adaptation, learning, and optimization, Vol. 12, No. 3, p. 729, 2012. https://doi.org/10.1007/978-3-642-27645-3.

[11] A. Clifton and I. Ericson, "Hazard analysis techniques for system safety", J. Wiley.

[12] M. Soeken, R. Wille, and R. Drechsler, "Assisted behavior driven development using natural language processing", Objects, Models, Components, Patterns: 50th International Conference, TOOLS 2012, Vol. 7304, pp. 269-287, May 2012. https://doi.org/10.1007/978-3-642-30561-0_19.

[13] R. Mariani, "An overview of autonomous vehicles safety", 2018 IEEE International Reliability Physics Symposium (IRPS), Burlingame, CA, USA, Mar. 2018. https://doi.org/10.1109/IRPS.2018.8353618.

[14] G. Dulac-Arnold, D. Mankowitz, and T. Hester, "Challenges of real-world reinforcement learning", arXiv preprint arXiv:1904.12901, Apr. 2019. https://doi.org/10.48550/arXiv.1904.12901.

[15] J. Garcıa and F. Fernández, "A comprehensive survey on safe reinforcement learning", Journal of Machine Learning Research, Vol. 16, No. 1, pp. 1437-1480, 2015.

[16] G. Dulac-Arnold, et al., "Challenges of real-world reinforcement learning: definitions, benchmarks and analysis", Machine Learning, Vol. 110, No. 9, pp. 2419-2468, Apr. 2021. https://doi.org/10.1007/s10994-021-05961-4.

[17] G. Dalal, K. Dvijotham, M. Vecerik, T. Hester, C. Paduraru, and Y. Tassa, "Safe exploration in continuous action spaces", arXiv preprint arXiv:1801.08757, Jan. 2018. https://doi.org/10.48550/arXiv.1801.08757.

[18] J. Achiam, D. Held, A. Tamar, and P. Abbeel, "Constrained Policy Optimization", in International conference on machine learning, PMLR, Vol. 70, pp. 22-31, 2017.

[19] J. Huang, J. Chen, L. Zhao, T. Qin, N. Jiang, and T.-Y. Liu, "Towards deployment-efficient reinforcement learning: Lower bound and optimality", arXiv preprint arXiv:2202.06450, Aug. 2022. https://doi.org/10.48550/arXiv.2202.06450.

[20] R. Sadeghi and F. Goerlandt, "A proposed validation framework for the system theoretic process analysis (STPA) technique", Safety science, Vol. 162, pp. 106080, Jun. 2023. https://doi.org/10.1016/j.ssci.2023.106080.

[21] M. Ben-Daya, "Failure mode and effect analysis", Handbook of maintenance management and engineering, Springer, pp. 75-90, 2009. https://doi.org/10.1007/978-1-84882-472-0_4.

[22] K. Beckers, D. Holling, I. Côté, and D. Hatebur, "A structured hazard analysis and risk assessment method for automotive systems—A descriptive study", Reliability Engineering & System Safety, Vol. 158, pp. 185-195, Feb. 2017. https://doi.org/10.1016/j.ress.2016.09.004.

[23] V. J. Hodge, R. Hawkins, and R. Alexander, "Deep reinforcement learning for drone navigation using sensor data", Neural Computing and Applications, Vol. 33, pp. 2015-2033, Jun. 2021. https://doi.org/10.1007/s00521-020-05097-x.

[24] D. J. Pumfrey, "The principled design of computer system safety analyses", phdthesis, University of York, Sep. 1999.

[25] Y. Qi, Y. Dong, S. Khastgir, P. Jennings, X. Zhao, and X. Huang, "STPA for learning-enabled systems : a survey and a new practice", In: 26th IEEE International Conference on Intelligent Transportation Systems ITSC 2023, Bilbao, Bizkaia, Spain, pp. 24-28, Sep. 2023.

[26] J. Henriksson, M. Borg, and C. Englund, "Automotive safety and machine learning: Initial results from a study on how to adapt the ISO 26262 safety standard", in Proceedings of the 1st International Workshop on Software Engineering for AI in Autonomous Systems, pp. 47-49, May 2018. https://doi.org/10.1145/3194085.3194090.

[27] C. Solis and X. Wang, "A study of the characteristics of behaviour driven development", in 2011 37th EUROMICRO conference on software engineering and advanced applications, Oulu, Finland, pp. 383-387, Aug. 2011. https://doi.org/10.1109/SEAA.2011.76.

[28] Y. Wang and S. Wagner, "Combining stpa and bdd for safety analysis and verification in agile development: a controlled experiment", in Agile Processes in Software Engineering and Extreme Programming: 19th International Conference, XP 2018, pp. 37-53, May 2018. https://doi.org/10.1007/978-3-319-91602-6_3.

[29] M. Wynne, A. Hellesoy, and S. Tooke, "The cucumber book: behaviour-driven development for testers and developers", Pragmatic Bookshelf, 2017.

[30] R. K. Lenka, S. Kumar, and S. Mamgain, "Behavior driven development: Tools and challenges", 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), pp. 1032-1037, Oct. 2018. https://doi.org/10.1109/ICACCCN.2018.8748595.

[31] E. Leurent, An Environment for Autonomous Driving Decision-Making. GitHub, 2018. https://github.com/eleurent/highway-env [accessed: Jul. 03, 2023]

[32] T. Stolte, G. Bagschik, A. Reschka, and M. Maurer, "Hazard analysis and risk assessment for an automated unmanned protective vehicle", in 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, pp. 1848-1855, Jun. 2017. https://doi.org/10.1109/IVS.2017.7995974.

저자소개

조 수 희 (Suhee Jo)

2023년 2월 : 경기대학교 컴퓨터공학부(공학사)
2023년 2월 ~ 현재 : 경기대학교 SW안전보안학과 석사과정
관심분야 : 소프트웨어 공학, 소프트웨어 안전성, 기계 학습

권 령 구 (Ryeonggu Kwon)

2011년 2월 : 경기대학교 컴퓨터과학과(이학사)
2013년 2월 : 경기대학교 컴퓨터과학과(이학석사)
2013년 3월 ~ 현재 : 경기대학교 컴퓨터과학과 박사과정
관심분야 : 기계 학습, 정형 합성

권 기 현 (Gihwon Kwon)

1985년 2월 : 경기대학교 전자계산학과(이학사)
1987년 8월 : 중앙대학교 전자계산학과(이학석사)
1991년 2월 : 중앙대학교 전자계산학과(공학박사)
1991년 2월 ~ 현재 : 경기대학교 컴퓨터공학부 교수
1999년 ~ 2000년 : 미국 CMU 전산학과 방문교수
2006년 ~ 2007년 : 미국 CMU 전산학과 방문교수
2014년 ~ 2016년 : 한국정보과학회 소프트웨어공학 소사이어티 회장
2021년 ~ 현재 : 경기대학교 SW중심대학 사업단장
2022년 ~ 현재 : 경기대학교 소프트웨어경영대학 학장
관심분야 : 소프트웨어 공학, 소프트웨어 안전성, 정형 검증 및 정형 합성