Journal of KIIT. Vol. 21, No. 7, pp. 39-48, Jul. 31, 2023. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2023.21.7.39

STPA-RL: 강화학습을 이용한 STPA에서 손실 시나리오 분석

장지영*, 권령구**, 권기현***

STPA-RL: Analyzing Loss Scenarios in STPA with Reinforcement Learning

Jiyoung Chang*, 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

시스템의 위험 분석 기법인 STPA(System Theoretic Process Analysis)에서는 안전하지 않은 제어 행동(Unsafe Control Action)으로 야기된 위험의 원인과 결과 관계를 설명하는 손실 시나리오 식별이 필수적이다. 지금까지는 전문가의 수작업, 주관적, 비체계적인 방법으로 손실 시나리오를 식별하였다. 이에 본 논문에서는 강화학습(RL, Reinforcement Learning)을 결합하여, 손실 시나리오를 자동으로 도출하는 STPA-RL을 제안한다. 이를 위하여 STPA 분석 결과를 바탕으로 안전한 제어 행동을 강화 학습하는 환경을 모델링하고, 이어서 위험 상태에 도달하는 시스템 상태 전이 과정을 탐색한다. 산업 공정 시스템을 사례연구로 실험한 결과, 약 400개의 손실 시나리오를 생성할 수 있었다. 그 결과, 높은 빈도를 갖는 위험을 파악할 수 있었을 뿐만 아니라 위험에 이르는 상태 변화 과정을 시각화하여 손실 시나리오 이해를 높일 수 있었다.

Abstract

In System Theoretical Process Analysis (STPA), a hazard analysis technique for systems, it is essential to identify loss scenarios that describe the cause and effect relationships of hazards caused by unsafe control actions. Until now, loss scenarios have been identified by experts using manual, subjective, and unsystematic methods. In this paper, we propose STPA-RL, which combines reinforcement learning (RL) to automatically derive loss scenarios. For this, we model an environment where safe control actions are reinforced based on STPA analysis results, and then explore the system state transition process that leads to a hazard state. Experimenting with an industrial process system as a case study, we were able to generate about 400 loss scenarios. As a result, we were able to not only identify high-frequency hazards, but also visualize the state transition process leading to the hazard to improve the understanding of loss scenarios.

Keywords

loss scenario, STPA, reinforcement learning, hazard analysis, process industry system

* 경기대학교 SW안전보안학과 석사과정
- ORCID: https://orcid.org/0000-0002-8605-8805
** 경기대학교 컴퓨터과학과 박사과정(교신저자)
- ORCID: https://orcid.org/0000-0002-4942-247X
*** 경기대학교 컴퓨터과학과 교수
- ORCID: https://orcid.org/0000-0002-8221-4939

· Received: Jun. 25, 2023, Revised: Jul. 19, 2023, Accepted: Jul. 22, 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

시스템 수준의 위험 분석을 위해 FTA(Fault Tree Analysis)[1], FMEA(Failure Mode Effect Analysis)[2] 같은 전통적인 기법이 널리 사용되고 있다. 이들 기법은 시스템 구성 요소의 개별적인 고장이 전체 시스템의 고장을 유발한다는 신뢰성 이론에 기초한다. 한편, 본 논문에서 활용할 STPA(System Theoretic Process Analysis)는 개별 요소 고장이 아닌 구성 요소 간의 잘못된 상호작용이 위험을 일으킨다는 시스템 이론에 기반을 둔 위험 분석 방법론이다[3].

STPA는 컨트롤 스트럭처(Control structure)를 이용하여 위험을 유발할 수 있는 안전하지 않은 제어 행동(UCA, Unsafe Control Action)의 발생원인 식별이 가능하다. 또한, 시스템의 정상적인 운영 중에 발생할 수 있는 잠재적인 위험 상황을 의미하는 손실 시나리오를 식별하고 예방하는 것은 중요하지만, 동시에 매우 어려운 작업이다. 이는 분석가의 주관적인 판단에 의존할 수 있으며 복잡한 시스템의 경우에는 손실 시나리오 분석이 쉽지 않다[4]. 뿐만 아니라, 잠재적인 위험 요소를 놓치거나 중요하지 않은 요소를 과도하게 강조하는 문제점이 있다[5].

기존에는 STPA에 모델 체킹(Model checking) 기법을 연계하여 시스템 모델을 구축하고 동작의 정확성을 검증하는 연구가 존재한다[6]. 다만 모델 체킹은 상태 공간의 전수 검사라는 특성 때문에 대규모 시스템에서는 상태 공간이 기하급수적으로 증가하는 상태 폭발 문제를 가지고 있다. 이러한 점을 보완하기 위해, 본 논문에서는 STPA에 강화학습(RL, Reinforcement Learning)을 결합한 STPA-RL을 제안하여 시스템의 위험과 원인을 명확히 하는 손실 시나리오를 확보하고자 한다. 강화 학습은 에이전트가 환경과 상호 작용하여 보상을 최대화하는 행동을 학습하는 기계학습의 한 분야로, 복잡한 상태 공간에서 손실 시나리오를 효과적으로 식별하고, 위험을 학습하여 예방하는 능력 향상에 기여할 것이다.

STPA-RL은 첫째, 시스템을 STPA로 분석한 내용과 시스템에 영향을 미치는 제어 관계를 중심으로 강화학습 환경을 모델링한다. 둘째, 컨트롤 알고리즘을 학습하여 안전한 행동을 수행하도록 보상을 최대화하는 강화학습을 진행한다. 셋째, 학습 모델 평가 단계에서, 상태 변화 경로 도출을 통해 손실 시나리오를 식별한다. 넷째, 발견한 설계 오류 정보를 이용하여 모델을 수정하며 시스템 설계 시뮬레이션에 활용한다. 결과적으로, STPA-RL은 위험에 도달하게 되는 상태 변화의 경로를 추적함으로써 세 가지의 분석 방법인 위험 빈도 분석, 잠재적인 손실 시나리오 제시, 환경 변화 패턴 파악에 기여한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 설명하고, 3장에서는 시스템의 위험 분석을 위해 STPA-RL을 제안한다. 4장에서는 산업 공정 시스템을 이용한 사례 연구의 과정을 보인다. 이어서 5장에서는 사례 연구의 결과 분석을 수행하고, 6장에서는 결론 및 향후 연구를 기술한다.

Ⅱ. 관련 연구

2.1 STPA

STPA는 신뢰성 이론을 기반으로 하며 시스템 또는 구성요소들 간 제어 문제에서 위험이 발생함을 전제로, 제어 관계 중 위험을 유발할 수 있는 부적절한 제어를 식별하는 방식으로 위험을 분석한다[7].

STPA를 활용한 위험 분석은 다음의 네 단계로 수행된다[8]. 분석 대상과 관련한 손실 및 위험을 정의하고, 분석의 목적에 따라 대상의 범위를 결정하는 1단계, 컨트롤 스트럭처를 도식화하는 2단계, UCA를 도출하는 3단계 그리고 원인 요소들을 도출하고 손실 시나리오를 정의하는 4단계로 이루어진다. 손실 시나리오는 다음 두 가지를 설명하기 위해 작성된다[9]. 첫째, 잘못된 피드백, 부적절한 요구사항, 설계 오류, 구성요소 고장 및 기타 요인이 UCA를 초래하고 궁극적으로 손실을 초래하는 방법이다. 둘째, 안전한 컨트롤 액션(CA, Control Action)이 제공되지만 제대로 지켜지지 않거나 실행되지 않아 손실을 초래하는 방법이다.

다만, STPA만으로는 손실 시나리오의 생성 시 수동 판단 및 분석에 의존하여 임의성이 초래될 수 있으며, 포괄성과 정확성이 보장되기 어려우므로 후속 안전 요구 사항 작성 시 구체성이 약화될 수 있다. 따라서, 본 논문에서는 STPA의 손실 시나리오를 생성하는 과정에 있어 컴퓨터의 계산 및 처리

기능을 활용하여 현재 시스템의 제어 관계의 적절성을 파악하려 한다[10]. 이를 위해 시스템을 STPA로 분석한 1~3단계의 산출물을 이용하며, 시스템 설계에 직접적인 도움이 되도록 강화학습과 연계하여 STPA의 4단계를 발전시키고자 한다.

2.2 강화학습

강화학습은 기계 학습의 한 분야로, 에이전트라고 불리는 학습 주체가 환경과 상호작용하며 행동을 선택하여 보상을 최대화하는 방법을 학습하는 알고리즘이다[11]. 강화학습은 지도학습과 달리, 명시적인 정답이 주어지지 않고 시간에 따라 행동의 결과로부터 학습한다. 다양한 도메인에서 적용 가능하며 예측이나 분류와 같은 문제보다는 실제 시스템에서의 의사결정과 제어에 적합하다.

강화학습의 핵심 개념은 마르코프 결정 과정(MDP, Markov Decision Process)이다[12]. MDP는 환경을 시간적으로 분리된 상태(State), 액션(Action), 보상(Reward), 상태전이확률(Transition probability)의 요소로 모델링한다. 에이전트는 현재 상태를 관측하고 선택 가능한 액션 중에서 특정 액션을 선택하여 환경에 영향을 준다. 그리고 환경은 에이전트의 액션에 대한 보상을 제공하고, 상태전이확률에 따라 새로운 상태로 전이된다. 행동 가치 함수(Q-value function)은 식 (1)과 같으며 상태-액션 쌍의 기댓값의 가치를 나타낸다[13]. 여기서 R은 보상, γ는 할인 계수(Discount factor), s는 상태, a는 액션이다.

$$Q(s,a) = E[R_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) \mid s_t = s, a_t = a] \quad (1)$$

에이전트는 학습 알고리즘을 사용하여 최적의 정책(Policy)를 학습한다. 정책은 주어진 상태에서 선택할 행동을 결정하는 규칙을 나타내며, 보상을 최대화하는 정책을 찾는 것이 강화학습의 목표이다[14]. 정책 업데이트를 위해 목적 함수를 최적화하는 방식으로 학습하는 PPO(Proximal Policy Optimization) 알고리즘의 목적 함수는 식 (2)와 같으며 정책의 성능을 최대화하면서 이전 정책과의 차이를 제어한다[15]. 여기서 $r_t(\theta)$는 이전 정책과 새로운 정책간의 상대적인 비율을 나타내며, t는 시간 단계를, θ는 정책의 파라미터를 나타내는 변수이다. A_t는 어드밴티지 함수로 현재 상태-액션 쌍의 가치를 나타낸다. 이를 통해 얼마나 좋은 액션을 선택했는지 평가한다. $clip$ 함수는 정책 업데이트 과정에서 정책 간의 차이 값의 범위를 제한하며 ϵ는 범위 제한을 조절하는 파라미터이다.

$$L(\theta) = E[min(r_t(\theta)A_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t)] \quad (2)$$

본 논문에서는 이러한 강화학습의 수식을 기반으로 시스템의 위험 상태 탐지와 예방을 수행한다. 시스템의 강화학습 환경을 구축하여 환경과 제어 액션의 상호작용을 통해 시스템 환경에서의 동작을 모델링한다. 위험 상태를 탐지하며 이를 회피하는 액션을 수행할 정책을 학습한다. 이를 통해 시스템은 위험 상태로의 경로를 식별하고, 안전하지 않은 액션을 강화학습에서 배제하며 위험을 예방하도록 한다.

III. 제안한 STPA-RL 알고리즘

이 장은 STPA에 강화학습을 결합한 STPA-RL을 제안한다. 주 목적은 STPA 네 번째 단계의 손실 시나리오를 학습의 결과물로 도출하는 것이다. STPA-RL은 절차는 그림 1과 같다.

3.1 시스템 환경 모델링 단계

3.1.1 상태 구성을 위한 프로세스 모델 정의

STPA에서 프로세스 모델(Process model)은 컨트롤러가 CA를 제공하기 위해 필요한 시스템 상태 정보를 갖고 있다. 이를 활용하여 환경의 상태를 정의한다. 프로세스 모델은 시스템의 동작 방식, 입력 및 출력 변수, 동작 제약 조건 등을 정의한다. 다만, 시스템 학습 환경을 구성하기 위한 프로세스 모델이 충분히 정의되지 않았다면 UCA를 도출하는 과정에서 도출한 특정 상황 또는 조건인 컨텍스트(Context)로부터 프로세스 모델을 추가로 정의하고, 기존 STPA 분석 결과를 보완한다.

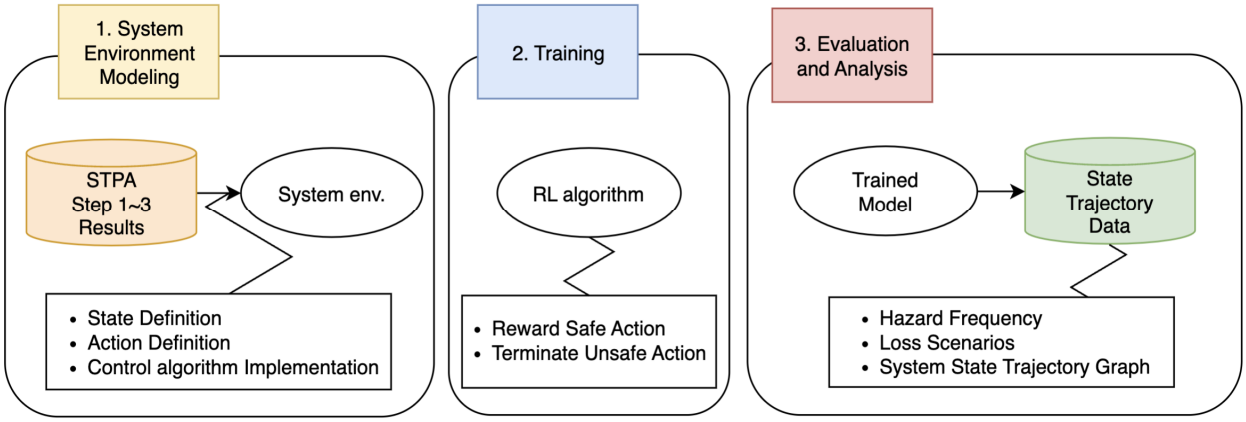

그림 1. STPA-RL 절차
Fig. 1. Procedure of STPA-RL

3.1.2 액션 구성을 위한 컨트롤 액션 정의

CA는 시스템의 동작을 조정하거나 변화시켜 시스템을 안전한 상태로 유지하도록 하는 역할을 한다. 따라서, CA를 선정한 후 이를 강화학습 환경의 액션으로 사용한다. 액션은 시스템의 제어 변수 또는 조작 명령어로 정의될 수 있으며, 강화학습 알고리즘을 통해 최적의 액션을 학습하게 된다.

3.1.3 컨트롤 알고리즘 구성 및 구현

컨트롤 알고리즘(Control algorithm)은 CA를 결정하는 규칙 또는 정책을 구성하고 구현하는 단계이다. 컨트롤 알고리즘을 활용하여 “왜 UCA가 발생하는가?”의 질문에 집중하여 손실 시나리오를 도출할 수 있는 환경을 구현한다. 또한, 각 액션에 대해 제어되는 상태와 조건을 설정하여 정확히 동작하도록 한다.

3.1.4 위험 재구분과 우선순위 부여

위험(Hazard)은 시스템의 위험 요소로 정의되며, 재구분과 우선순위 부여 과정을 거쳐야 한다. UCA를 위험 상태에 도달하였다는 조건으로 활용한다. 또한, 상태 조건에 따라 위험의 심각성을 판단하여 우선순위를 부여하고, 세분화하여 컨트롤 알고리즘에 적용한다.

3.1.5 환경 상태 변화 과정 설정

실제 시스템에서 발생할 수 있는 다양한 상황들을 고려하여 시스템 환경의 상태 변화를 설정한다. 이는 기본적인 동작뿐만 아니라 변칙적인 동작이나 예외 상황 등을 포함하여 시스템 전이를 표현한다. 상태의 변화 과정은 실제 시스템의 동작에 따라 강화학습 모델로 시뮬레이션 기능을 제공하도록 한다.

3.2 시스템 강화 학습 단계

3.2.1 학습 알고리즘 선택 및 강화 학습

구현한 환경에 알맞은 강화학습 알고리즘을 선택한다. 선택한 학습 알고리즘을 이용하여 시스템을 강화 학습한다. 알고리즘은 강화학습 환경의 상태를 관찰하고, CA를 선택하여 시스템을 안전한 동작으로 유도한다. 학습은 일련의 에피소드를 통해 진행되며 성능 개선을 위해 반복적으로 학습을 수행한다. 상태 전이 데이터 흐름이 분석에 용이한 적절한 길이로 학습된 모델을 생성하는 것이 목표이며, 평균 보상 값을 통해 성능을 확인한다.

3.2.2 상태 조건 기반으로 안전한 액션 시 보상

컨텍스트와 상태 조건을 고려하여 안전한 액션이 수행되었을 때에는 보상을 부여한다. 이는 강화학습이 안전한 액션을 장려하고 보상을 최대화하는 방향으로 학습하도록 도와준다.

3.2.3 상태 조건 기반으로 위험한 액션 시 위험

컨텍스트와 상태 조건을 고려하여 UCA가 수행되었을 때에는 위험에 도달하였다 설정하고 에피소드 종료를 판단한다. 이는 강화학습이 위험한 액션을 피하도록 학습하고, 안전한 액션 선택을 강화하는 방향으로 학습하도록 도와준다.

3.3 모델 평가 및 분석 단계

3.3.1 환경 상태 변화 흐름 저장

학습된 모델의 평가 과정에서 발생하는 환경 상태 변화의 추적하기 위해, 경로 추적(Trace) 목록을 생성하고 저장한다. 현재 상태에서 안전한 액션을 수행할 경우, 해당 상태와 수행한 액션을 추적 목록의 “path”에 연결하여 저장한다. 위험한 액션을 수행할 경우, 추적 목록의 “H{num}”에 해당하는 위험에 연결하여 저장한다. {num}에는 위험의 번호를 입력한다. 또한, 각 에피소드가 종료될 때마다 현재까지의 경로 길이도 함께 저장한다.

3.3.2 손실 시나리오 도출 및 시스템 위험 분석

저장한 환경 상태 변화 경로 추적 목록을 기반으로 손실 시나리오를 도출하고 시스템을 위험, 손실 시나리오, 상태 흐름 세 가지 방법으로 분석한다.

IV. 사례 연구

4.1 산업 공정 시스템

시스템 설계 시 제안한 STPA-RL의 적용 과정을 살펴보기 위해 산업 공정 시스템을 이용한다. 가스 생산 시설의 공정 산업 공장의 설비에는 생산 유정, 유동 라인, 집열 시스템 및 냉각기, 분리기, 압축기가 포함되어 있다. [16]에서는 가스, 탄화수소 응축액 및 물과 같은 서로 단계를 분리하는 분리기(Separator) 플랜트를 STPA로 분석하였다. 탄화수소 응축액(Condensate)와 물의 수위 계면 수준에 따라 제어 하에 각각 응축액 안정화 장치와 물 탱크로 보내진다. 응축액 수준이 저수준이면 응축액 출구 제어 밸브를 닫아 응축액 출구에서 나오는 흐름을 중단하며 수위가 저수준에 도달하면 출구 제어 밸브를 닫아 출구로부터 흐름을 중단한다. 그리고 응축액 수준이 고수준이면 분리기의 압력은 높아진다.

그림 2는 STPA-RL을 적용할 시스템의 컨트롤 스트럭처이다. 기본 공정 제어(BPC, Basic Process Control)는 공정 산업에서 운영자(Operator)를 대신하여 시스템을 제어하는 자동화 시스템이다[17]. 이는 일련의 제어 알고리즘과 센서로부터의 분리기의 상태, 압력, 수위 등 입력 신호를 기반으로 분리기와 연결된 응축액 밸브와 생산된 물의 흐름을 제어하는 워터 밸브를 개폐하며 공정을 제어한다.

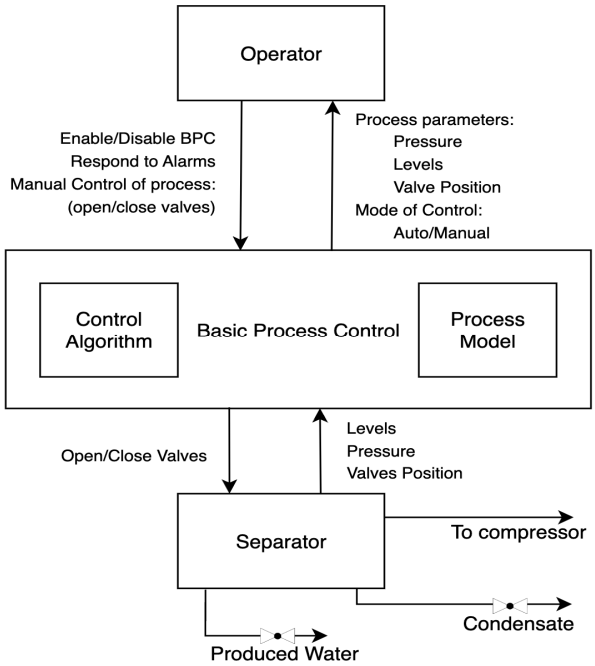

그림 2. 산업 공정 시스템의 컨트롤 스트럭처
Fig. 2. Control structure of process industry system

4.2 시스템 모델링을 통한 환경 구현

산업공정 시스템의 제어 문제를 강화학습 환경으로 모델링한다. 정의한 프로세스 모델은 컨트롤 스트럭처와 UCA의 정보를 기반으로 강화학습 환경의 상태로 구성한다. 표 1에서는 상태를 정의한다. 상태 0,1 과 2는 각 분리기의 압력, 수위, 응축액 수준으로 저수준(0), 보통 수준(1), 고수준(2)의 값을 갖는다. 상태 3와 4는 워터 밸브와 응축액 밸브의 상

태로 열림(0), 닫힘(1) 값을 갖는다. 상태 5는 운영 모드로, BPC의 자동제어 모드(0)과 운영자의 매뉴얼 모드(1)의 값을 갖는다. 상태 6는 현재 공정의 상태로 비정상(0), 정상(1)의 값을 갖는다. 실제 시스템에 존재하는 제어 변수들을 제공하여 시뮬레이션이 가능하도록 상태를 설정한다.

표 1. 산업 공정 시스템의 상태 정의
Table 1. State definition of process industry system

num	State name	num	State value
0	Separator pressure	0	low
		1	normal
		2	high
1	Water level	0	low
		1	normal
		2	high
2	Condensate level	0	low
		1	normal
		2	high
3	Produced water valve position	0	opened
		1	closed
4	Condensate valve position	0	opened
		1	closed
5	Operation mode	0	auto (BPC)
		1	manual (Operator)
6	Process condition	0	out of normal
		1	normal

표 2는 액션을 정의한다. 액션 0부터 3은 두 가지 밸브를 열고 닫는다. 4부터 7은 운영자 수행하는 액션으로 BPC를 활성화하거나 무력화하며 위험 상황을 알리는 경보에 응답하거나 응답하지 않는다.

표 2. 산업 공정 시스템의 액션 정의
Table 2. Action definition of process industry system

value	Control action for action
0	water valve open
1	condensate valve open
2	water valve close
3	condensate valve close
4	operator enable BPC
5	operator disable BPC
6	operator respond to alarm
7	operator not respond to alarm

기존 STPA 결과에서는 UCA 16가지가 모두 H1과 H2의 위험에 영향을 끼칠 수 있다, 정의하고 있지만[16], 시스템의 동작 흐름을 위해 정의한 프로세스 모델, CA를 컨트롤 알고리즘에 적용하여 구현하기 위해서 위험을 표 3과 같이 5가지로 상세하게 분류한다.

표 3. 산업공정시스템의 위험 목록
Table 3. Hazard List of process industry system

Hazard	Hazard description
H1,H2	Potentially release hydrocarbon from the process and cause hydrocarbon enter to the water stream
H3	Causing high pressure in the separator and condensate enter to gas stream
H4	Causing gas enter to condensate stream
H5	Process is out of normal condition

4.3 강화 학습 수행

시스템의 최적 제어를 이해하고 안전성을 고려한 의사를 결정할 수 있도록 본 논문에서는 Stable Baselines3의 파이썬 라이브러리의 PPO 알고리즘을 이용하여 학습시킨다[18]. PPO 알고리즘은 안정적인 정책 업데이트를 위해 제약 조건을 설정한다. PPO는 이전 정책과 새로운 정책의 차이를 관리하여 너무 큰 변화를 방지하고 학습의 안정성을 보장한다[19]. 이를 통해 에이전트는 시스템 환경과의 상호작용을 반복하며 점진적으로 보상을 최적화하는 정책을 학습한다.

V. 사례 결과 및 분석

이 장에서는 시스템 강화학습 모델을 활용하여 시스템의 평가 결과 분석을 3가지 방식으로 수행한다. 첫째, 위험 분석은 시스템에서 해당 위험이 발생하는 정도를 분석하여 비교 제시한다. 둘째, STPA 4단계의 산출물인 손실 시나리오를 도출한다. 셋째, 시스템 상태 흐름 시나리오를 이용하여 시스템의 환경 변화 패턴을 파악한다.

5.1 위험 분석

위험 분석은 시스템의 잠재적 위험을 식별하고 분석하여 시스템에서 발생할 수 있는 피해 정도를 평가하는 과정이다. 각 위험에 대해 빈도를 비교하여 관리자나 운영자에게 정보를 제공한다.

표 4는 학습 과정의 에피소드 수를 다르게 하여 학습 모델 7가지를 생성한 것을 각 1000개의 에피소드 수로 평가를 진행하여 평균 보상 값을 비교한 결과이다. M1 모델의 경우 1000번의 학습 에피소드 수로 1.92의 보상 값을 얻었지만, M4는 10000번으로 19.6의 보상을 얻어, 10배 이상의 성능 차이를 보였다. 학습 40000번의 M7은 536.3의 평균 보상 값으로, M4에 비해 약 27배의 보상 크기 차이를 보였다. 따라서, 학습 에피소드의 수가 많을수록 최적의 학습을 통해 안전한 액션을 선택하여 진행함을 파악할 수 있다.

표 4. 학습 에피소드 수 별 모델의 평가 평균 보상 값
Table 4. Model evaluation reward by train episode

Model	Train episode	Evaluation episode	Average reward
M1	1000	1000	1.92
M2	5000	1000	7.73
M3	8000	1000	12.1
M4	10000	1000	19.6
M5	15000	1000	57.8
M6	20000	1000	78.8
M7	40000	1000	536.3

그림 3은 각 모델 별 위험(H1, H3, H4, H5)의 발생 빈도를 비교한다. 위험의 빈도는 H1이 평균적으로 0.393의 값을 가지며 가장 많이 나타나고 H3이 0.046으로 상대적으로 적게 나타나는 것을 파악한다. 이는 H1 위험의 발생 확률과 해당 위험의 중요성을 나타내며, 안전 제약사항 수립 시 H1에 관한 시스템 수준의 요건을 강화할 필요가 있다는 것이다.

본 논문에서는 손실 시나리오와 시스템 상태 흐름의 분석에 용이성을 위해 평균 보상 값 19.6를 갖는 M4 모델을 이용한다. 이는 평균적으로 19 정도 길이의 상태 변화 흐름을 뽑아낼 수 있는 것이다. M4 모델의 평가 결과로 적정 수준 길이의 위험까지의 상태 경로 데이터를 도출하여 분석한다.

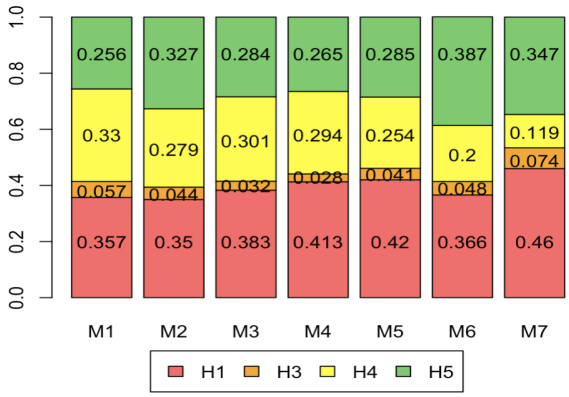

그림 3. 모델 별 위험 발생 빈도 비교
Fig. 3. Comparison of hazard frequency by model

5.2 손실 시나리오

위험까지의 상태 경로 데이터를 이용하여 손실 시나리오를 확보한다. 이렇게 도출된 손실 시나리오들은 시스템은 시스템의 특정 조건이나 상황에서 위험에 빠지게 되는 경로를 보여준다. M4 모델에서는 약 400개의 손실 시나리오를 발견하였다. 표 5는 위험에 도달하게 된 시나리오의 상태를 n번째 상태라 하여, n-1번에서 n번으로 전이된 상태-액션 쌍의 값을 나타낸다. LS는 Loss Scenario을 줄인 표현이다.

표 5. 손실 시나리오의 상태-액션 쌍
Table 5. State-action pair of loss scenarios

LS	State(n-1)	Safe action	State(n)	Unsafe action
LS16	[0,2,0,1,1,1,1]	4	[0,2,0,1,1,0,1]	2
LS25	[0,1,0,0,0,0,0]	3	[1,1,1,0,1,0,0]	1
LS119	[1,2,2,1,1,0,0]	5	[2,2,2,1,1,1,0]	3
LS244	[1,2,0,1,1,1,0]	6	[0,2,0,1,1,1,1]	5
LS396	[0,2,0,1,0,0,0]	0	[0,2,0,0,0,0,0]	1

다음은 손실 시나리오 목록 중 각각 다른 위험에 속하는 LS16, LS119, LS244와 LS396에 대해서 기술한다. LS16은 H1에서 0.02의 빈도로, LS119는 H3에서 0.09의 빈도로, LS244는 H5에서 0.01의 빈도로 발생, LS396은 H4에서 0.04 빈도로 발견되었다. 안전한 시스템 동작을 위해 강화학습을 수행했음에도 불구하고 발생한 위험들이다. 이는 시스템의 복잡성과 다양한 상호작용으로 예기치 않은 결과가 발생할 수 있다는 것을 시사한다.

이를 통해 우리는 안전성을 향상시키기 위해 추가적인 조치와 개선이 필요함을 인식한다.

LS16. 운영 모드가 BPC의 자동 제어 모드로 설정되었을 때, 수위가 높고 워터 밸브가 닫혔으나 BPC가 비정상 작동하여 워터 밸브를 닫았다. 결과적으로, 탄화수소가 워터 스트림으로 유입된다. [H1, H2]

LS119. 분리기의 압력이 보통에서 고압으로 변하고 응축액 수준이 높을 때, 운영자가 응축액 밸브를 닫았다. 결과적으로, 분리기 내 압력이 높아지고 응축액이 가스 스트림으로 유입된다. [H3]

LS244. 공정 상태가 정상 범위를 벗어났을 때, 운영자가 경보에 대응하여 상태를 정상으로 돌렸지만, 운영자 모드가 수동으로 변경되었을 때, 다시 BPC를 비활성화하였다. 결과적으로, 공정이 정상 범위를 벗어난다. [H5]

LS396. 워터 밸브가 닫혀 있을 때, BPC가 수위를 맞추기 위해 작동되었으나, 응축액 수준이 낮을 때 BPC가 응축액 밸브를 열었다. 결과적으로, 가스가 응축액 스트림으로 유입된다. [H4]

5.3 시스템 상태 흐름

시스템 설계 내의 문제를 해결하는 방법으로, 시스템의 상태 변화를 추적하여 비정상적인 상황을 감지하고, 조치를 통해 시스템의 안전성을 유지하고자 한다. 손실 시나리오 LS25는 응축액의 수준이 보통일 때, BPC가 응축액 밸브를 열어 H4 위험에 도달한 것에 해당한다. 이는 현재 시스템의 컨트롤 알고리즘이 잘못된 것을 확인한 시나리오이다. 왜냐하면 H4는 응축액이 고수준이고 밸브를 열지 않을 때 발생하는 위험으로 보통 수준에서는 열어두어도 되기 때문이다. 파악한 설계 오류를 설계에 대한 변경으로 이용한다. 이와 같은 방식으로 상태 조건과 제어 관계에 문제가 있음을 확인하고, 문제를 해결하기 위해 시스템 환경 모델링을 변경하는 시뮬레이션 기능을 제공한다.

LS25. BPC가 응축액 밸브를 닫았을 때, 응축액 수준이 보통으로 돌아갔으며 BPC가 응축액 밸브를 열었다. 결과적으로, 가스가 응축액 스트림으로 유입된다. [H4]

그림 4는 상태 전이 패턴을 파악하고 잘못된 설계를 개선하기 위해 상태-액션 쌍은 노드(Node)로, 전이는 간선(Edge)으로 하여 그래프로 표현한 것이다. LS25를 도출하게 된 상태의 경로를 나타내며, 초기 상태에서부터 안전하지 않은 액션으로 H4에 도달하게 되는 흐름을 통해 시스템 동작의 다양한 특성을 이해하고자 한다. 여기서 n-1번 노드는 65번이고 n번 노드는 H4이다. 그래프의 전이 흐름을 살펴보면 스스로를 반복적으로 전이하는 노드로는 15, 22, 28, 45, 46과 54가 있다. 이는 해당 상태에서 액션을 수행했을 때, 상태에 변화가 생기지 않아 최적의 액션으로 다시 선택하여 반복적으로 나타나는 것이다. 경로 A(40 - 34 - 35 - 36 - 37 - 38 - 39 - 40)나 경로 B(53 - 49 - 50 - 51 - 52 - 53)의 경우에는 40과 53의 노드가 경로 반복의 중심이 되어 나타나는 관계를 예측할 수 있다. 발견한 관계 정보를 이용하여 시스템의 현재 제어 관계에 문제의 유무를 상세히 파악한다.

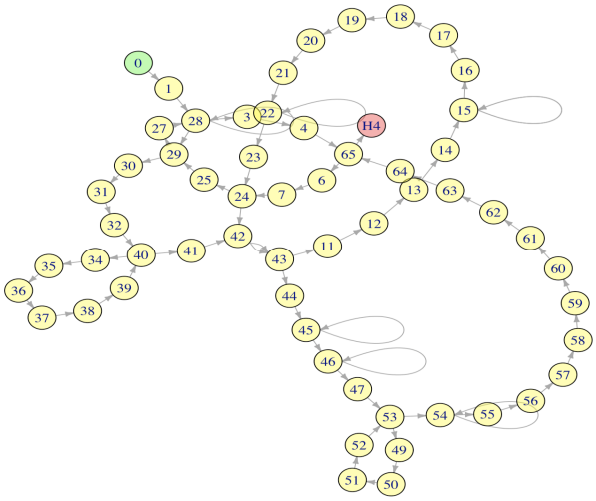

그림 4. LS25: H4에 도달하는 상태 흐름 시나리오
Fig. 4. LS25: State trajectory scenario to H4

VI. 결론 및 향후 과제

본 논문에서는 STPA 네 번째 단계의 손실 시나리오를 식별하기 위한 과정에 강화학습을 결합한 STPA-RL을 제안하였다. 그리고 이를 적용한 사례 연구로, 산업 공정 시스템의 상태 변화 흐름 데이터를 이용하여 손실 시나리오를 도출하고 3가지 분석 방법을 상세히 제시하였다.

제안한 STPA-RL은 시스템을 STPA 산출물을 통해 강화학습 환경을 모델링한다. 이어서 위험 상황을 감지하고, 안전한 액션을 내리는 방법을 강화학습 시켜 학습 모델을 생성한다. 해당 모델의 평가 과정에서는 위험 상태까지의 여러 경로를 추적하여 손실 시나리오를 산출하고, 상태 변화 흐름에서 위험에 관한 원인과 결과 관계를 규명한다. 사례 연구로 삼았던 산업 공정 시스템에서는 학습 모델 M4의 기준으로 손실 시나리오는 총 400개에 도달하였다. H1 위험의 빈도가 41%로 가장 자주 발생하였으므로, 해당 위험 관련한 안전 조치의 구현을 가장 고려해야 한다. 또한, 상태 전이 그래프를 그려 반복적인 경로와 상태-액션 쌍을 구별해내어, 제어 관계의 형태를 시각적으로도 확인하였다. 이러한 분석의 과정으로 현재 시스템 모델링의 위험에 대한 허점들을 발견하여 적절한 안전 제약사항으로 추가하거나, 시스템의 기존 STPA 결과를 수정할 수 있도록 시스템의 설계 시뮬레이션에 유리한 방법을 제안하였다.

강화학습은 다양한 상태와 액션의 조합으로 이루어진 복잡한 시스템에서 손실 시나리오 확보에 유용함을 본 연구를 통해 확인하였다. 다만, 강화학습 모델은 주어진 환경에서 학습된 상태에서의 동작을 최적화하는 방향으로 학습하기 때문에, 학습에 사용된 환경과 실제 운영 환경의 차이로 인해 학습된 모델이 일반화되지 않아 예기치 않은 동작이 나타날 수 있다는 문제가 존재한다. 따라서, 향후 과제로서 학습된 모델이 충분히 검증 되도록, STPA 산출물들을 강화학습 시스템 환경 모델링에 최대한 정확하게 반영할 수 있는 구체적인 방안을 개발하여 제시할 예정이다.

References

[1] E. Ruijters and M. Stoelinga, "Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools", Computer Science Review, Vol. 15-16, pp. 29-62, Feb. 2015. https://doi.org/10.1016/j.cosrev.2015.03.001.

[2] D. H. Stamatis, "Failure mode and effect analysis: FMEA from theory to execution", Quality Press, 2003.

[3] N. G. Leveson, "Engineering a safer world: Systems thinking applied to safety", The MIT Press, 2016.

[4] V. A. Yunusov, S. A. Demin, and A. A. Elenev, "The study of statistical features of the evolution of complex physical systems using adaptive machine learning methods", Journal of Physics: Conference Series, Vol. 2270, 2022. https://doi.org/10.1088/1742-6596/2270/1/012042.

[5] H. Li, J. Ki, J. Pimentel, G. Gruska, R. Xu, and F. Xu, "Complete Safety Analysis of Known and Unknown Scenarios in Autonomous Vehicles Based on STPA Loss Scenarios", SAE Technical Paper, No. 2022-01-7023, Jun. 2022. https://doi.org/10.4271/2022-01-7023.

[6] L. Dakwat and E. Villani, "System safety assessment based on STPA and model checking", Safety Science, Vol. 109, pp. 130-143, Nov. 2018. https://doi.org/10.1016/j.ssci.2018.05.009.

[7] J. Thomas, "Extending and Automating a Systems-Theoretic Hazard Analysis for Requirements Generation and Analysis", Ph. D. dissertation, Massachusetts Institute of Technology, USA, 2013.

[8] Telecommunication Technology Association, "New Approach to Hazard Analysis: Guide of Hazard Analysis using STPA", 2018.

[9] N. G. Leveson and J. Thomas, "STPA Handbook", Massachusetts Institute of Technology, Mar. 2018.

[10] J. Chang, R. Kwon, and G. Kwon, "Machine

Learning-based STPA for Platform Screen Door", In Proc. of Advanced Engineering and ICT-Convergence, Vol. 6, no. 1, pp. 38-42, Feb. 2023.

[11] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction", The MIT Press, 2018.

[12] H. Guo, Q. Cai, Y. Zhang, Z. Yang, and Z. Wang, "Provably efficient offline reinforcement learning for partially observable Markov Decision Processes", International Conference on Machine Learning, pp. 8016-8038, PMLR, 2022.

[13] J. Chang and G. Kwon, "Analysis of Hyper-parameters in Solving Sokoban using Q-learning", The Journal of KIIT, Vol. 20, No. 11, pp. 65-72, Nov. 2022. https://doi.org/10.14801/jkiit.2022.20.11.65.

[14] Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe, and j. Morimoto, "Deep learning, reinforcement learning, and world models", Neural Networks, Vol. 152, pp. 267-275, Aug. 2022. https://doi.org/10.1016/j.neunet.2022.03.037.

[15] G. Zhan, X. Zhang, Z. Li, L. Xu, E. Zhou, and X. Yang, "Multiple-UAV reinforcement learning algorithm based on improved PPO in Ray framework", Drones, Vol. 6, No. 7, pp. 166, Jul. 2022. https://doi.org/10.3390/drones6070166.

[16] A. Yousefi and M. R. Hernande, "Using a system theory method (STAMP) for hazard analysis in process industry", Journal of Loss Prevention in the Process Industries, Vol. 61, pp. 305-324, Sep. 2019. https://doi.org/10.1016/j.jlp.2019.06.014.

[17] M. R. Hernande and A. Yousefi, "A systematic approach methodology to measure process and plant safety", Chemical Engineering Transactions, Vol. 90, pp. 259-264, May 2022. https://doi.org/10.3303/CET2290044.

[18] Stable Baselines3, https://pypi.org/project/stable-baselines3/. [accessed: Jun. 24, 2023]

[19] S. Han and T. Liang, "Reinforcement-learning-based vibration control for a vehicle semi-active suspension system via the PPO approach", Applied Sciences, Vol. 12, No. 6, pp. 3078, Mar. 2022. https://doi.org/10.3390/app12063078.

저자소개

장 지 영 (Jiyoung Chang)

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

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