

SAMSUNG SDS

Realize your vision

Techtonic 2019

2019.11.14 (목)

삼성SDS 타워(잠실) B1F, 마젤란홀 / 파스칼홀

Partner



Foresee

Disrupt

Track 1

회로 설계 자동화를 위한 강화학습 적용기
제조현장 강화학습 적용 가이드

민찬호 프로 (AI선행연구Lab) / 삼성SDS

AGENDA

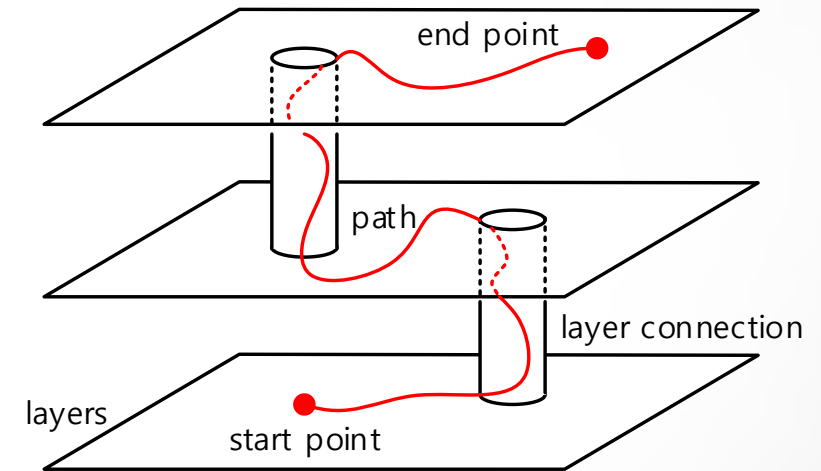
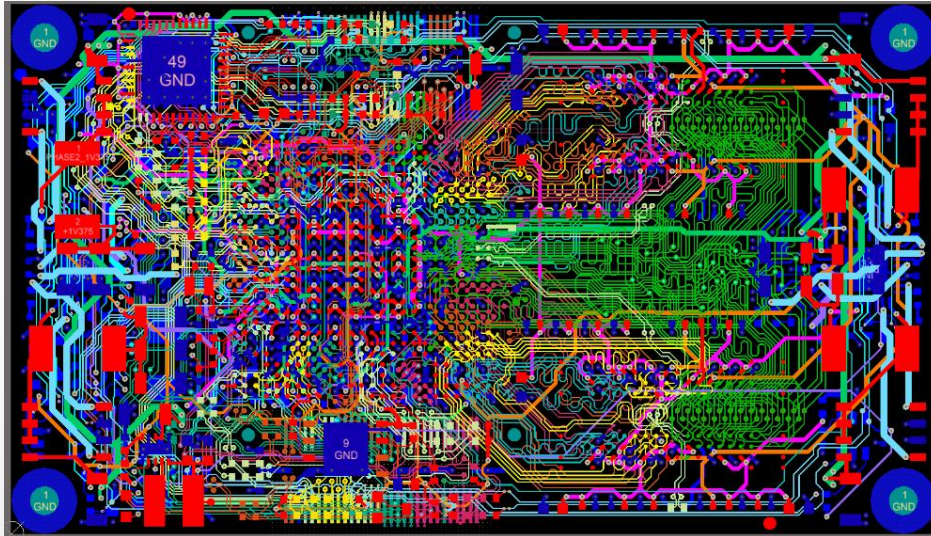
1. 시작하기에 앞서
2. 산업과 강화학습
3. 강화학습 적용기
4. 해결책과 결론

1

시작하기에 앞서

프로젝트 소개

반도체 회로(PCB) 설계 프로세스 자동화 기술

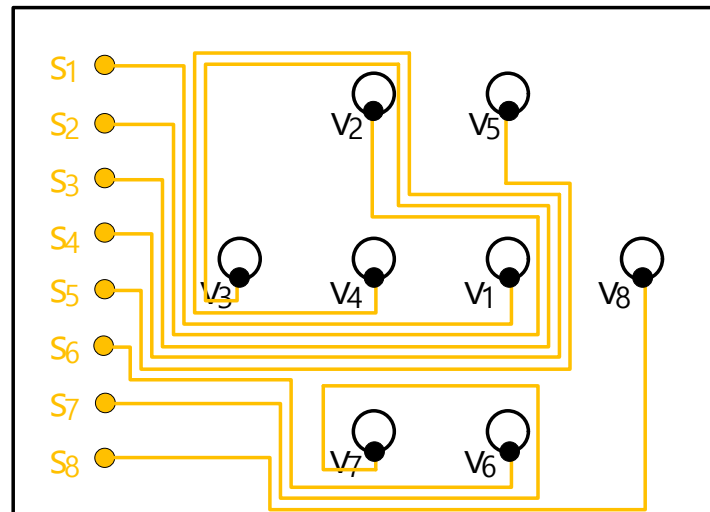
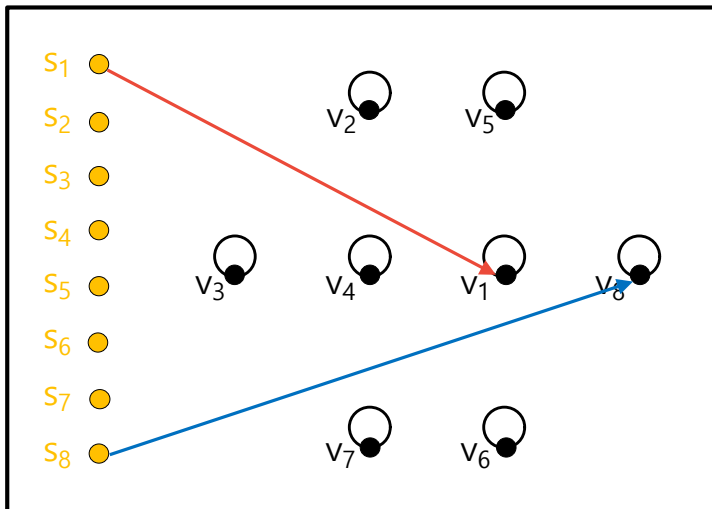


반도체 회로에서 Pin(start point)과 Ball(end point)을 잇는 회로를 설계

반도체 라우팅 프로젝트 특징

데이터가 적음

회로 설계 자동화



주요 특징

▶ 데이터가 적다

- 일반적인 딥러닝 기법 적용 불가능
- 현재 소수의 전문가들이 직접 "손"으로 설계

▶ 회로 간의 교차가 없어야 한다

- 회로 설계에선 회로 간의 간섭이 없어야 한다
- 교차가 생기면 설계도로서의 기능을 못한다

▶ 시작점과 끝점은 고정되어 있다

- 시작점과 끝점이 이어져야 한다
- 시작점과 끝점은 쌍으로 정해져 있다

강화학습의 기본 아이디어

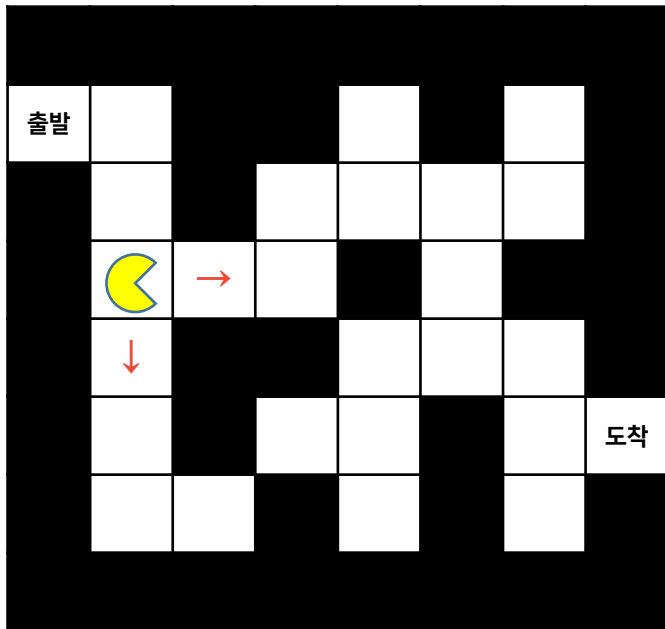
당근과 채찍



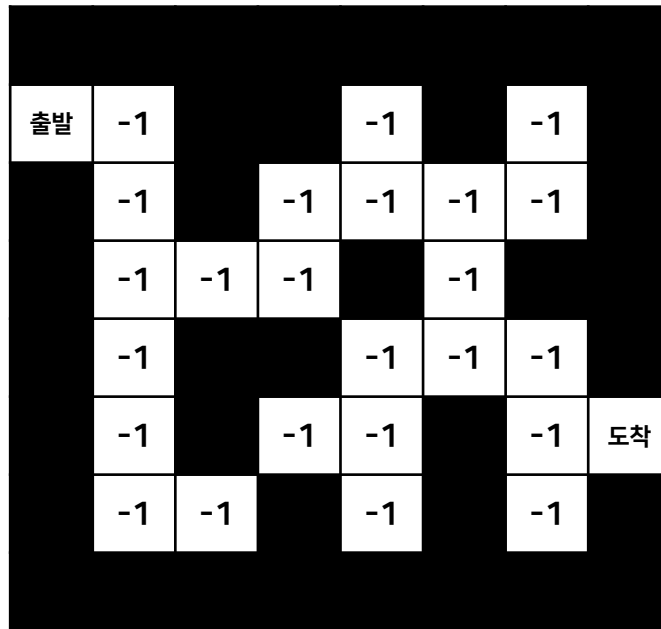
- ▶ 강화학습(RL):
잘하는 Agent에게 상
못하는 Agent에게 벌
- ▶ 실제 행동을 한 경험에서 배운다
데이터를 직접 생성할 수 있다
- ▶ 매우 일반적인 방법론으로 모든
문제에 적용 가능
- ▶ 기타 인공지능에 비해 성능이 떨어짐

강화학습의 구성요소

강화학습의 구성요소



Environment

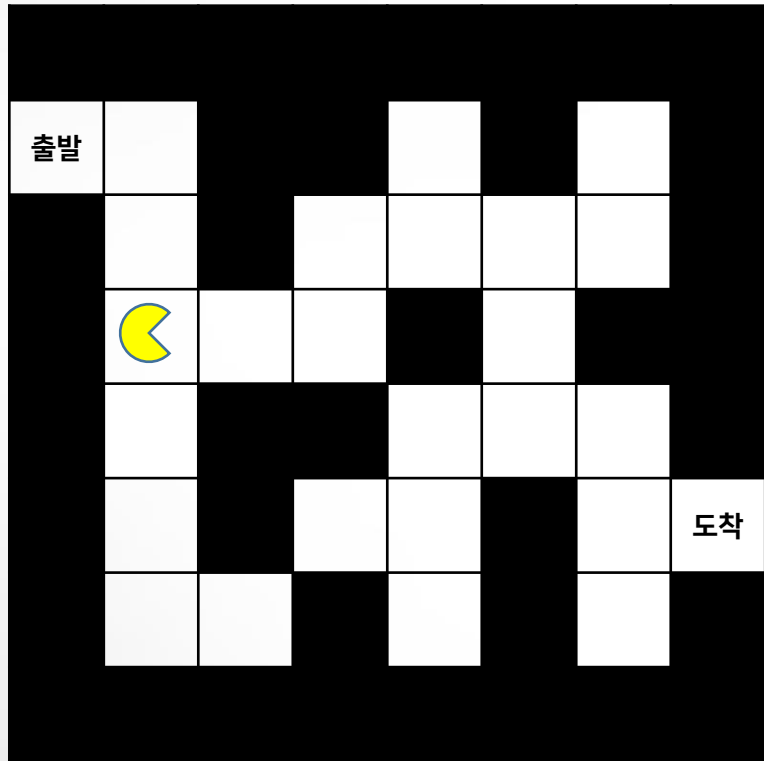


Reward

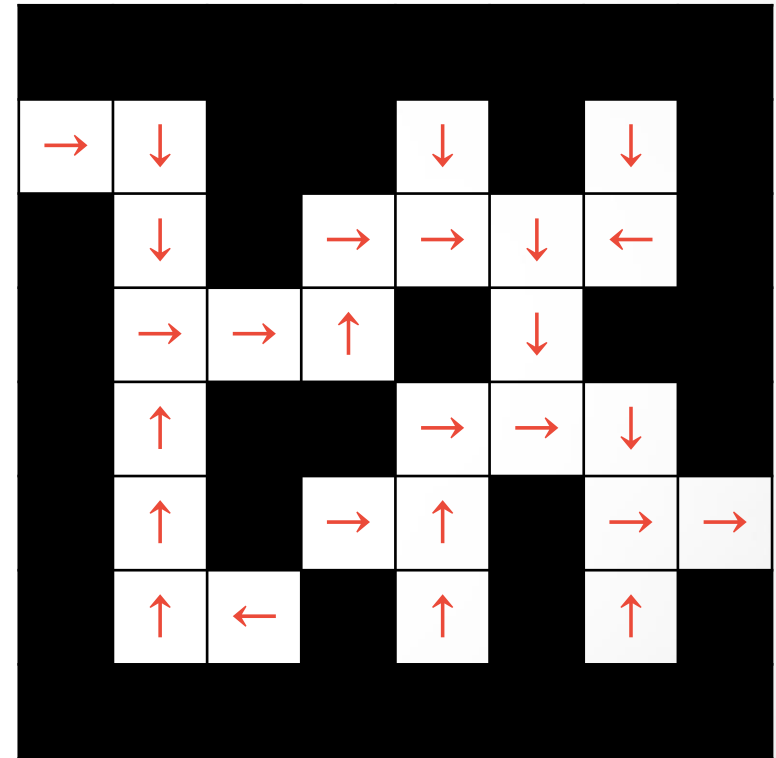
- 강화학습이 이루어지는 환경을 나타내는 Environment: 미로판
- Environment 안에서 행동하는 Agent: 미로판 안의 캐릭터
- Agent의 행동을 나타내는 Action: 캐릭터의 움직이는 방향
- Agent의 상태 State: 미로판 안에서 캐릭터의 위치
- Agent가 받는 보상 Reward: 미로에서 캐릭터가 받는 보상

강화학습의 종료

학습 완료의 정의



학습 완료



각 상태마다 최적의 행동을 도출

2

산업과 강화학습

산업과 강화학습

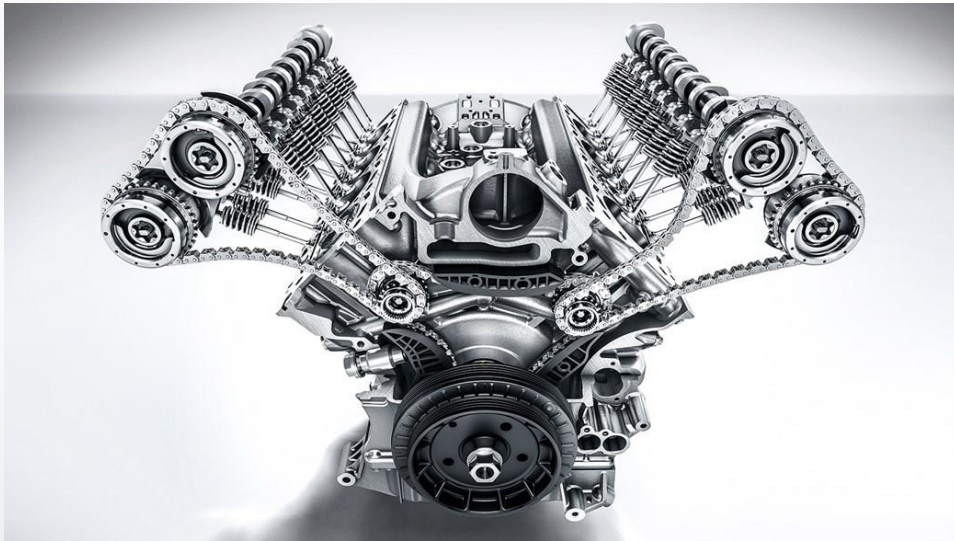
특정 상태에서 최적의 행동을 학습하는 일반적인 방법론



실제 산업체 적용에는 제약이 존재

강화학습의 역할

강화학습은 자동차의 엔진일 뿐이다



다른 부품도 갖추어야 기능을 한다

산업에서의 강화학습

강화학습 적용을 위한 구성 요소



문제를 나타내는 판
환경

- State와 Action을 이용하여 문제 표현
- 문제를 명확히 정의



데이터 생성 방향성
보상

- 탐색과 선택을 도울 수 있도록 보상 설정
- 학습 방향 가이드 역할



현재 상태를 해석
특징 추출법

- State를 feature로 표현
- Neural Network를 이용



학습의 주축이 되는
RL 알고리즘

- Policy Gradient
- Value Function

산업체에서의 강화학습

구성 요소를 모두 충족하기 어려움

Environment

Reward

Feature
Extraction

RL algorithm

- › 문제를 나타낼 수 있는 Environment 설정 자체가 힘들다
- › 데이터 탐색을 유도하는 Reward 설정에는 현업 전문가가 필요하다
- › 적절한 Reward를 통해서 Agent가 움직이더라도 현 상태를 분석하지 못한다면 학습이 불가능
- › 현재 논문으로 나오는 대부분의 RL알고리즘이 환경이나 reward 등이 확정된 후 평가된다
- › 위 4가지 요소가 하나라도 충족이 안 되면 학습이 불가능하다

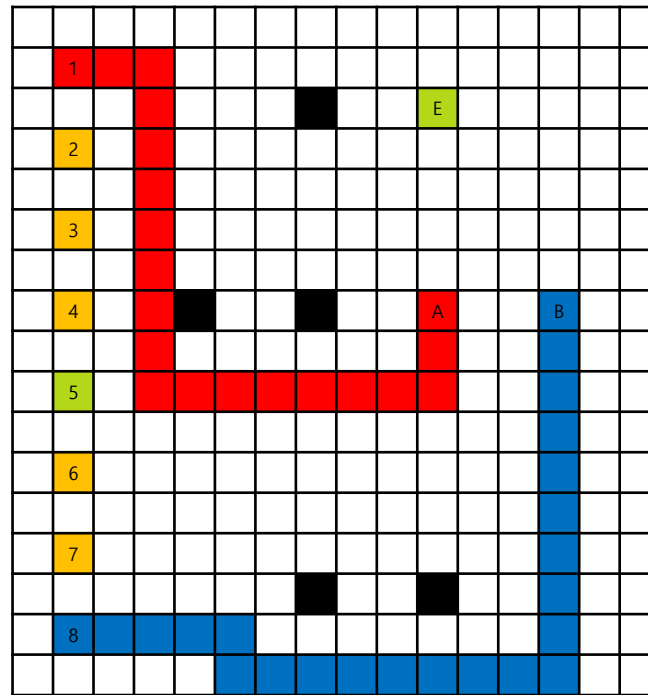
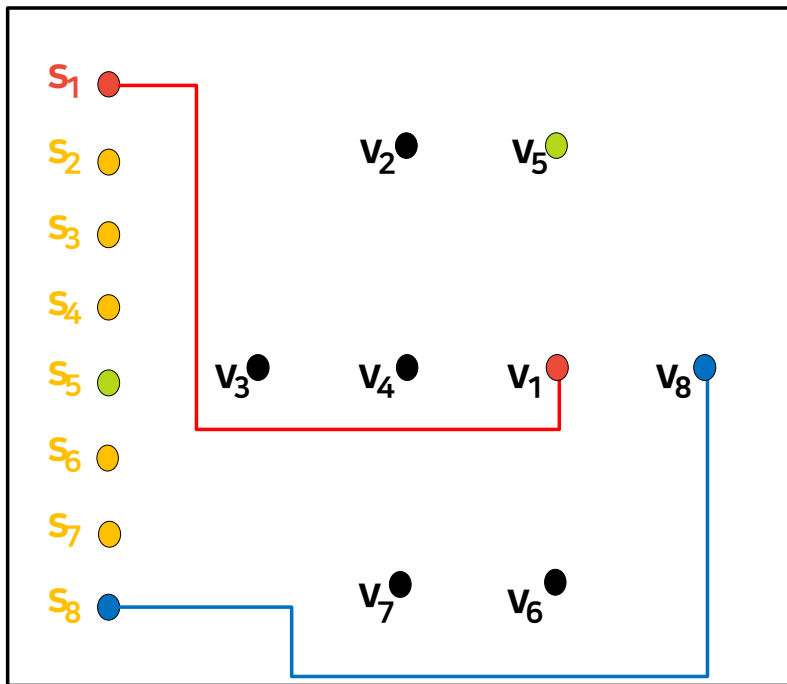
3

강화학습 적용기

미로로 보는 회로 설계

강화학습 적용 방법이 명확해 보였으나...

회로 설계 자동화



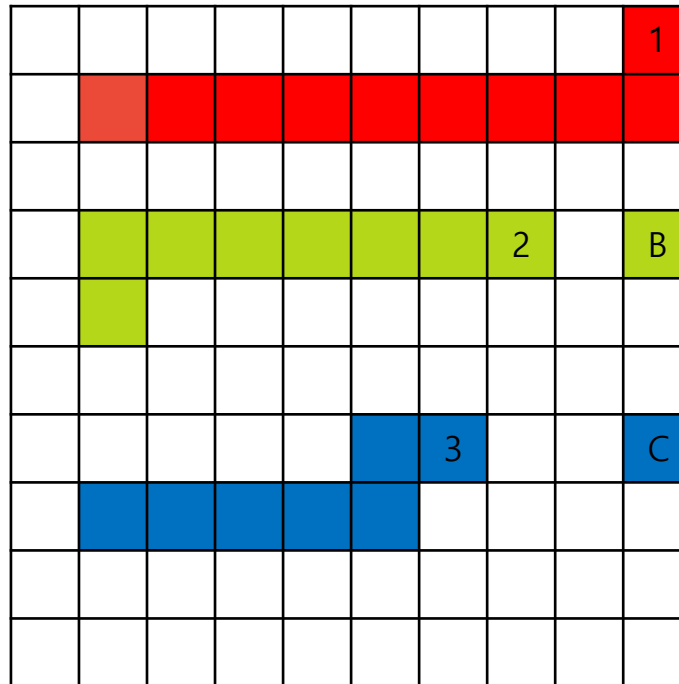
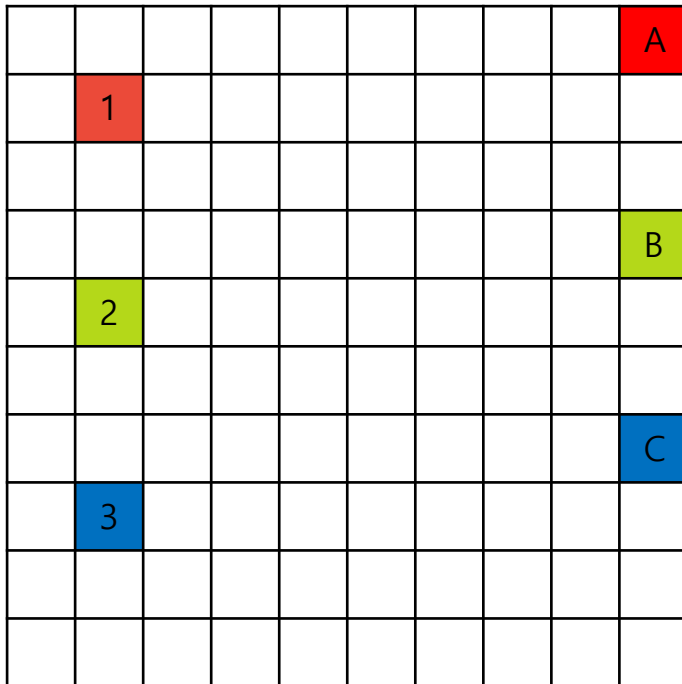
주요 특징

- ▶ 데이터가 적다
 - ▶ 강화학습을 이용해 데이터 생성
- ▶ 회로 간의 교차가 없어야 한다
 - ▶ 각 회로를 벽으로 인식하는 미로로 인식
- ▶ 시작점과 끝점은 고정되어 있다
 - ▶ 미로에서 출발점과 도착점이 고정되어 있다
- ▶ 회로 설계를 평가하는 알고리즘이 있다
 - ▶ 보상을 정해줄 수 있다

Environment 설정

Multi-agent를 활용한 미로 찾기 문제

회로가 벽인 미로로 표현

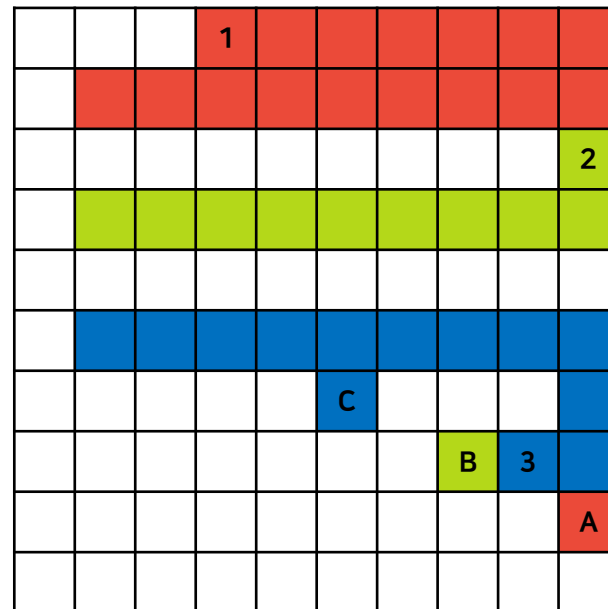
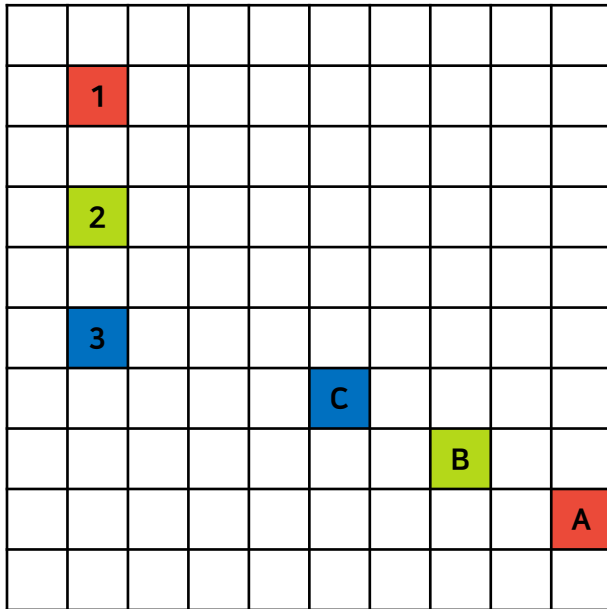


주요 특징

- 이미 지나간 길은 벽으로 인식
 - 이미 지나간 길은 벽으로 인식하기에 교차하지 않는다
- 판의 정보를 3d-Array로 표현
 - 각 위치마다 벽을 표시
 - Agent의 위치를 나타내는 별도의 Array 추가
- CNN Architecture를 활용
 - Feature Extraction을 위해 NN 활용

Reward 설정

학습 방향 가이드를 위한 Reward 설정

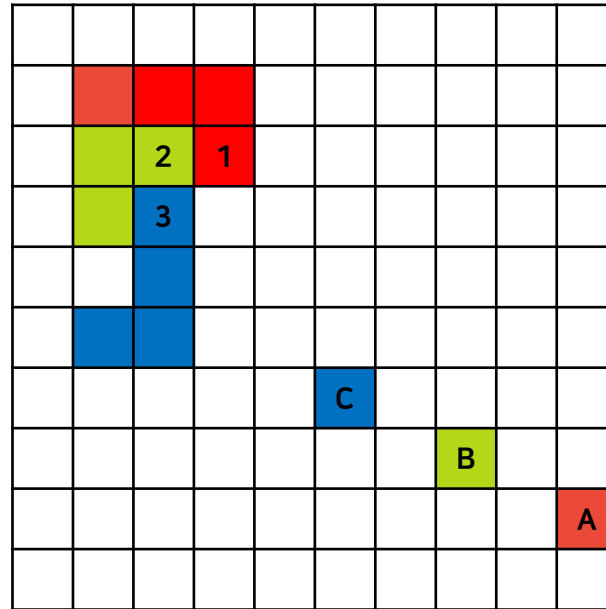
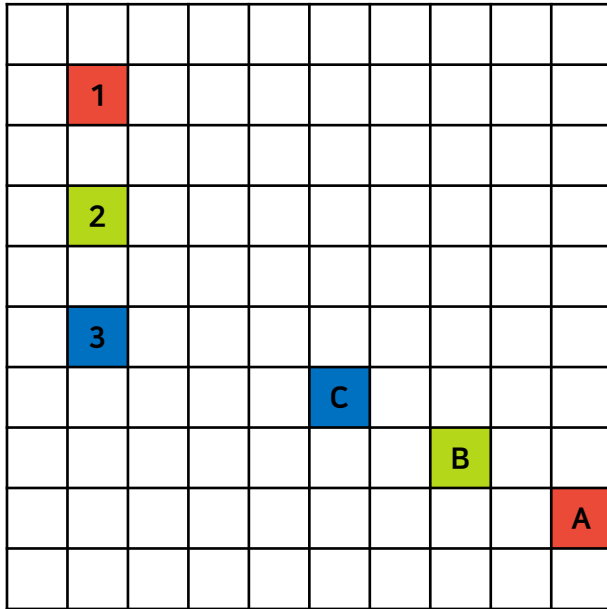


성공할 경우에만 Reward 1을 주는 경우

- 성공할 경우에만 Reward를 1로 주는 경우: 완전히 인간의 지식을 배제
- 가능한 회로를 오로지 무작위로 탐색하는 것으로 나타남
- Agent가 이동할 때마다 완전히 다른 Input으로 해석하기 때문
- 보다 더 나은 reward 설정이 필요

Reward 설정

학습 방향 가이드를 위한 Reward 설정

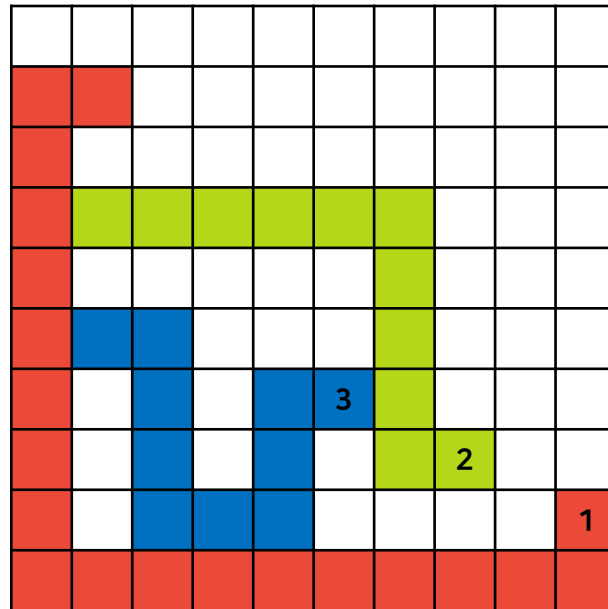
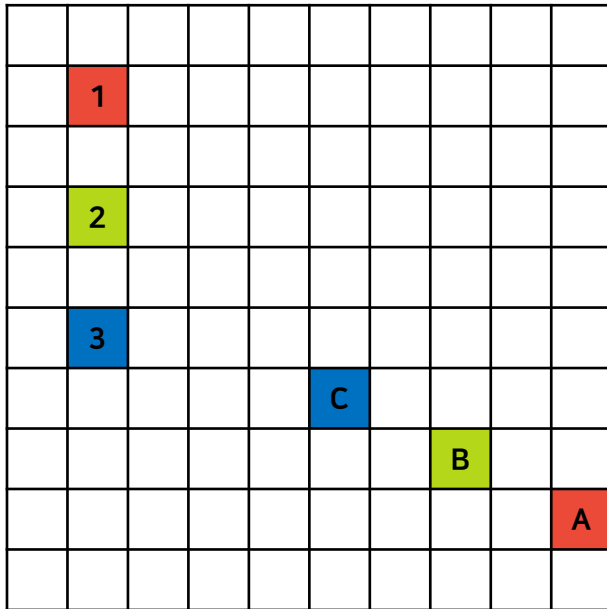


Reward를 매번 -1로 주는 경우

- ▶ 종료될 때까지 reward를 -1씩 주는 경우: 빠르게 종료가 되도록 학습을 유도
- ▶ 회로 탐색은 여전히 무작위: 가이드가 없기 때문
- ▶ 빠르게 종료를 하기 위해서 길을 막는 방법을 선택한다
- ▶ 보다 더 나은 Reward 설정이 필요

Reward 설정

학습 방향 가이드를 위한 Reward 설정

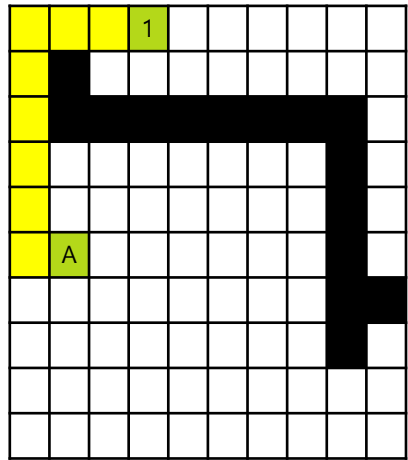


- ▶ 매번 Agent와 도착점까지의 거리에 비례해서 Penalty
- ▶ 회로 탐색할 때 가이드 역할: Agent들이 Penalty를 적게 받기 위해서 도착점으로 가려고 함
- ▶ 미로 풀기에 성공하는 경우 발생
- ▶ 그러나 학습의 수렴이 보장되지는 않음

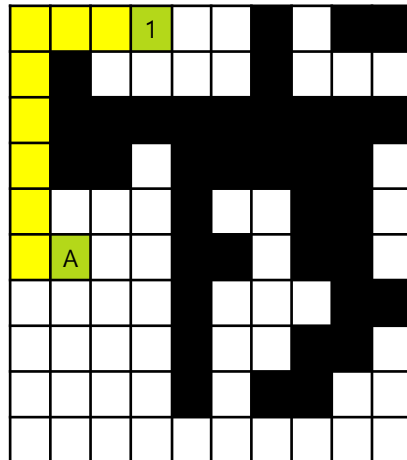
Reward 조정으로 가능한 데이터를 탐색

Feature Extraction

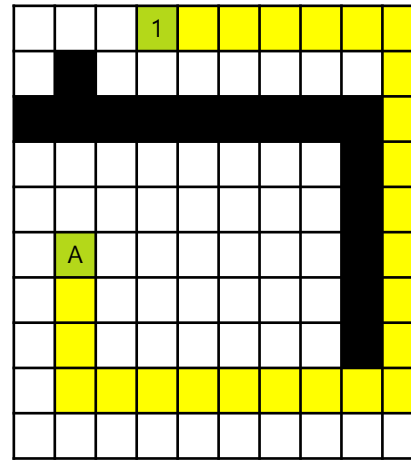
Feature extraction의 중요성



A



B



C

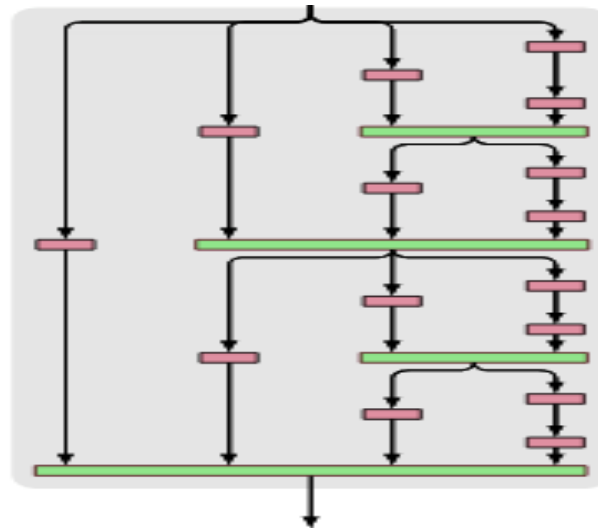
Reward가 정의되더라도 Feature Extraction이 필요

- ▶ 학습이 되지 않는 이유는 기하급수적으로 증가한 경우의 수 때문: 바둑의 경우의 수보다 더 많다
- ▶ 왼쪽 그림에서 A와 B는 장애물의 모양이 다르지만 가는 경로는 비슷하다
- ▶ 반대로 A와 C는 모양이 비슷하지만 가는 경로는 다르다
- ▶ 딥러닝을 활용한 Feature extraction이 필요하다

Feature Extraction

Feature extraction 최적화

0	1	0
0	2	0
0	0	0



Customized Convolution kernel과 Fractal Neural Network

- ▶ Customized Convolution Kernel을 활용하여 주변의 벽들을 인식한다
- ▶ Pooling Layer로 점차적으로 넓은 영역의 블록을 인식한다
- ▶ 각 pooling Layer들이 CNN을 적용한 결과를 concatenate 한 후 Dense Layer를 이용
- ▶ 총 5개의 Agent를 갖는 경우 성공

산업에서의 강화학습

강화학습 적용을 위한 구성 요소



문제를 나타내는 판
환경

- 회로판을 미로로 인식
- 점들을 잇는 것을 미로에서 길을 찾는 것으로 판단



데이터 생성 방향성
보상

- Agent의 탐색을 돕도록 Reward 설정



현재 상태를 해석
특징 추출

- Customized kernel을 활용
- Pooling layer를 이용한 Fractal Network 기반

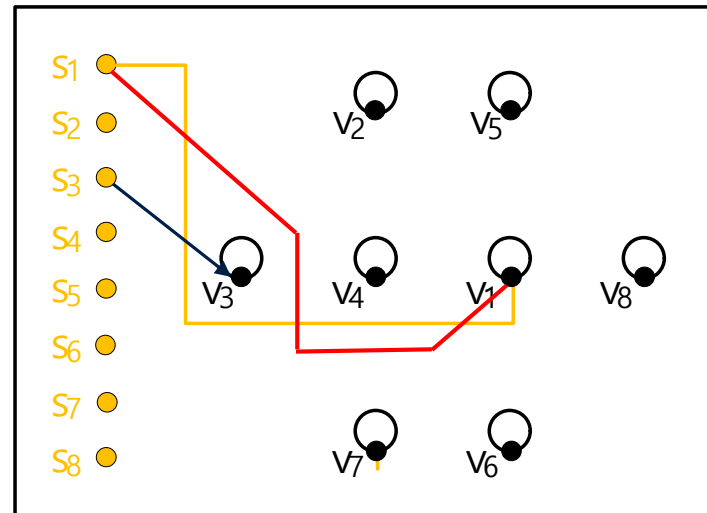
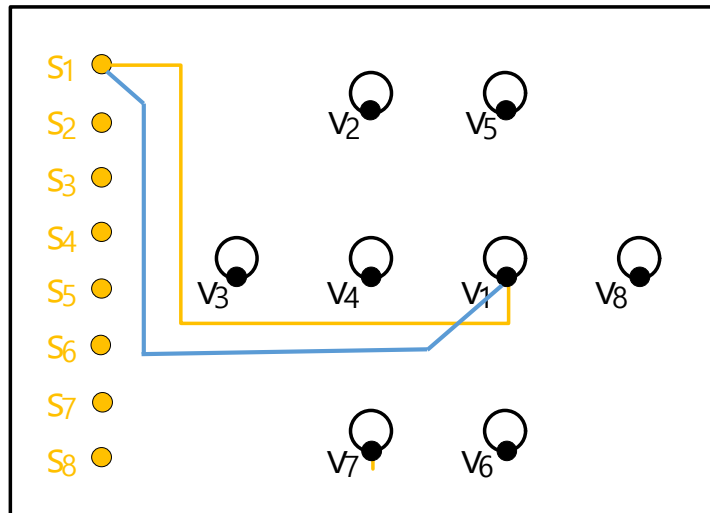


학습의 주축이 되는
RL 알고리즘

- Rainbow DQN 활용

새로운 Environment가 필요

Feature space를 최소화해야 한다

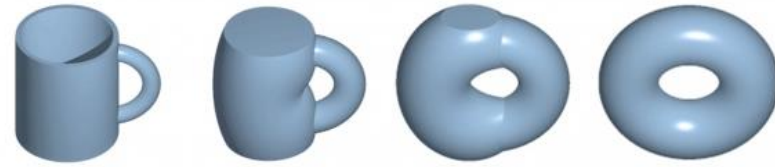


- ▶ 5개의 Agent에 대해서 성공을 했으나 실제 반도체는 많게는 200개까지의 Agent 필요
- ▶ 너무 많은 경우의 수: 판의 크기에 대해서 기하급수적으로 증가
- ▶ 불필요한 정보 과다: 길의 위치 데이터가 모두 필요한 것이 아니다
- ▶ 불필요한 정보를 필터할 데이터 구조가 필요

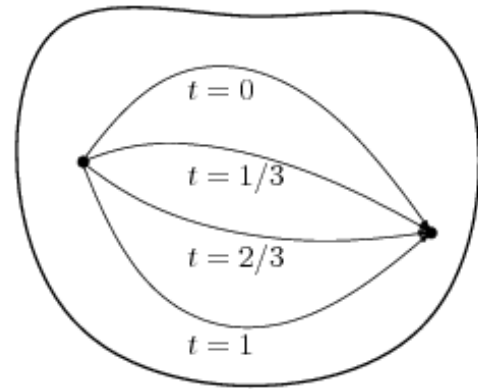
4

해결책과 결론

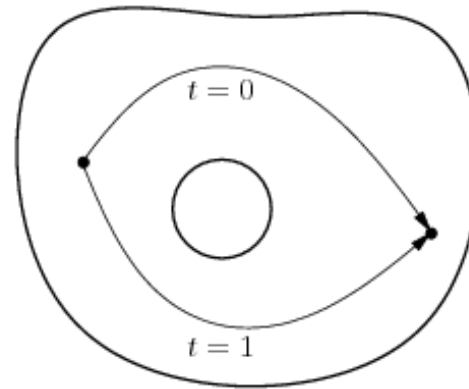
위상수학?



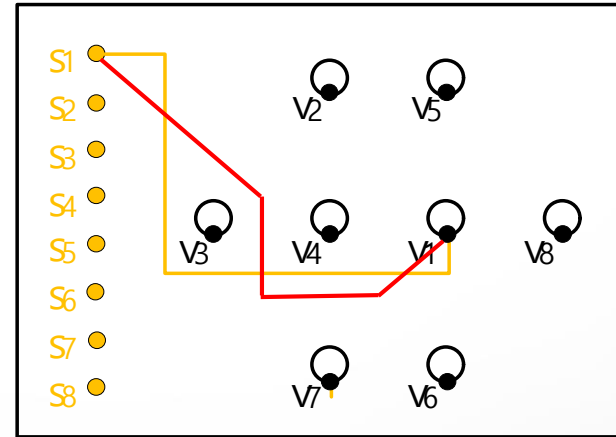
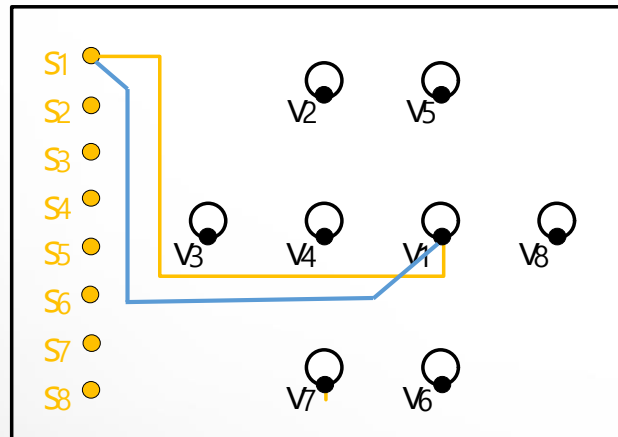
상대적인 위치만 고려하는 위상수학



(a)

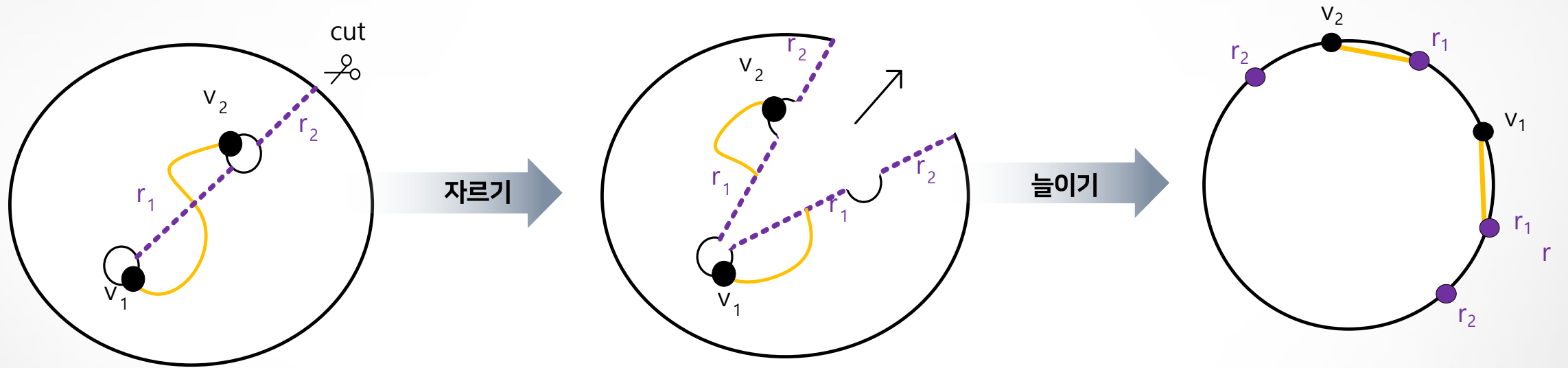


(b)



새로운 구조로 변형

회로를 위상수학으로 변형

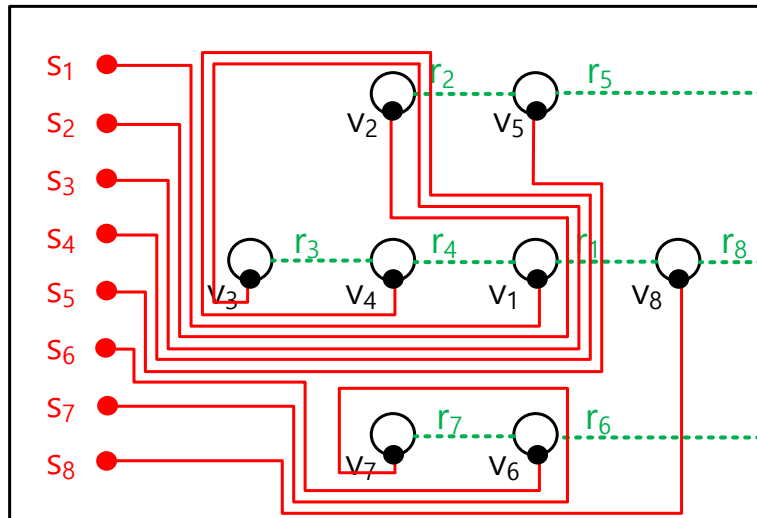


복잡한 길을 자르고 늘여서 원 위의 선분들로 변형

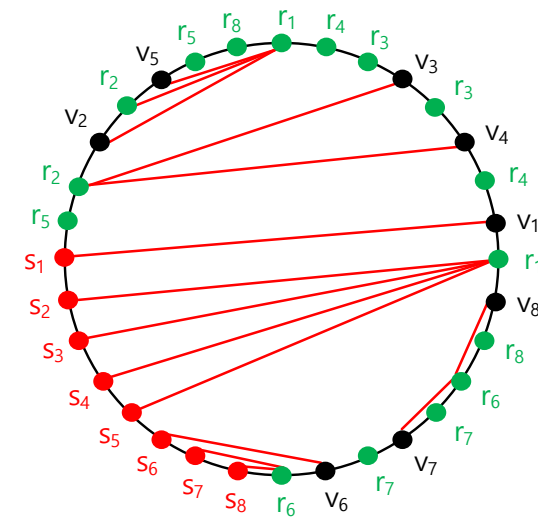
경우의 수 최소화

위상수학을 이용하여 환경을 극단적으로 단순화

Before



After

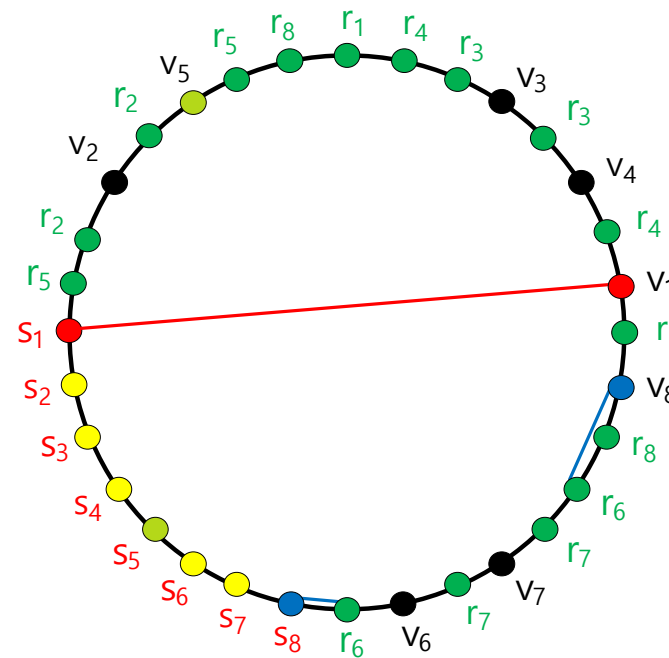
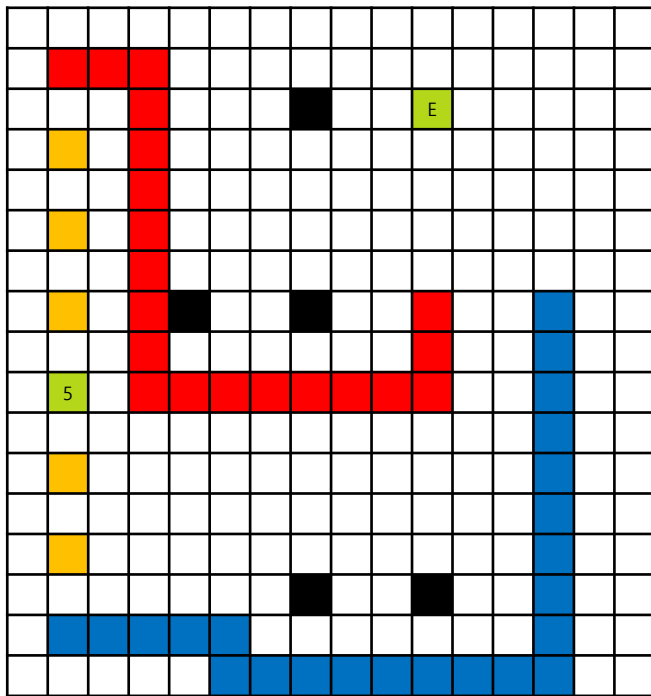


경우의 수를 극단적으로 축소

새로운 판

위상수학을 이용하여 판을 단순화

새로운 환경



주요 특징

- ▶ 새로운 판을 사용하여 경우의 수 축소
- ▶ 간단한 데이터 형식 덕분에 빠른 학습 가능
- ▶ 일반적인 라우팅 문제로 확장 가능
- ▶ 새로운 판을 사용하여 성공적 과제 마무리

결론

강화학습이 성공하려면...

- › Environment, Reward, Feature Extraction, RL algorithm 모두 필요하다
- › 위 네 가지 요소를 설정하기 위해서 개발자와 현업 전문가 간 협업이 필요하다
- › 개발자는 딥러닝에 대한 폭넓은 이해력과 구현력이 필요하다
- › 강화학습은 일반적인 딥러닝보다 더 많은 학습 데이터를 요구한다
- › 강화학습은 일반적인 supervised 러닝보다 성능이 떨어진다



Thank You



The graphic features the text 'Q & A' in a clean, sans-serif font. The 'Q' and 'A' are white, while the ampersand is a vibrant lime green. To the right of the text, two orange triangles point towards each other, forming a larger arrow shape that points into a large, dark blue circle on the right side of the image. The background is a dark blue gradient with a pattern of lighter blue circles on the left side.

Q & A

Partner Disrupt Foresee