- Causal Reasoning from Meta-reinforcement Learning
  - 1. Introduction
    - 환경에서 인과 관계를 찾는 것은 Intelligent Agent 에게 큰 도전이다
    - Meta-RL 에서 인과 관계를 찾는다
    - 학습된 Agent 가 새로운 보상이 포함된 상황에서 인과 추론을 한다.
    - 주어진 데이터를 통해서 인과 추론을 하고, 이를 통해서 유의미한 상호관계를 선택한다.
    - And to And model-free-RL Approach 및 RL 환경에서 구조화된 exploration 에서 새로운 전략을 제공한다
    - Meta Learning Agent 는 환경과 Interaction 을 통한 Best Causal Structure 을 찾는것이 관심
    -
  - 2. Problem Specification and Approach
    - Reasoning 의 3 가지 실험 방법
      - Observation, Interventional, Counterfactual
      - Observation:
        - Passive Observation from env
        - Associative reasoning 을 추론
        - Cause-effect reasoning
      - Intervention
        - Agent 가 Action 을 하는데 Value 값을 세팅, 무슨 value 냐면 some variable 과 연속적인 다른 variable 을 Observation.
      - Counterfactual(반사실적)
        - Intervention 을 통해서 Causal structure 를 Agent 가 배우고, 이를 통해서 Counterfactual Question 을 답 할 수 있다.
    - 2.1 Causal Reasoning
      - *causal Bayesian networks* (CBNs)
        - irected acyclic graphical

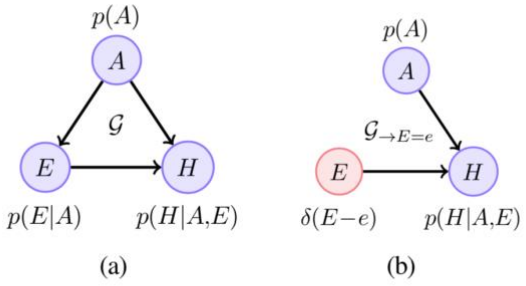


Figure 1. (a): A CBN $\mathcal{G}$ with a confounder for the effect of exercise (E) on heath (H) given by age (A). (b): Intervened CBN $\mathcal{G}_{\rightarrow E=e}$ resulting from modifying $\mathcal{G}$ by replacing $p(E|A)$ with a delta distribution $\delta(E-e)$ and leaving the remaining conditional distributions $p(H|E,A)$ and $p(A)$ unaltered.

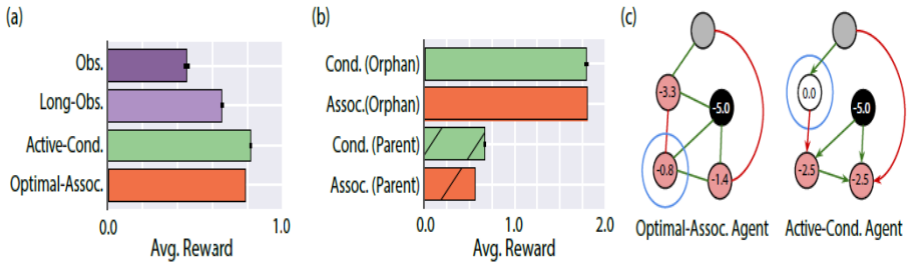
- Cause-effect Reasoning
  - Confounder 의 Graph 를 제거 하는 효과
- Counterfactual Reasoning
  - 예측된 질문"Does exercising improve cardiac health? " 질문에 답할 수 있다
  - 하지만, 일어나지 않은 일에 예측하는 질문에 답할 수 없다
  - 중간에 어떤 정보가 추가 되면 예측 할 수 있다(혈압이 높아서 신장박동이 빨라 졌다.) Cardiac Health 에 영향을 준다

○ 3. Task Setup and Agent Architecture

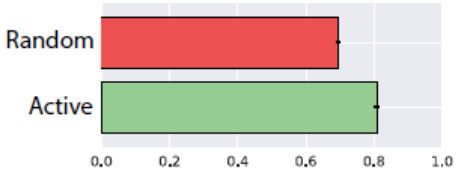
- Agent 가 different CBN G 와 상호 작용
- 각 Episode Information phase 와 Quiz phase 로 구성
  - Information Phase : First T-1 step 에서 G 로부터 Passively observing 상호작용함으로 써 Agent 가 information 을 획득
  - Quiz Phase: Final Step T 에서 Information Phase 에서 수집한 data 를 바탕으로 Agent 가 causal knowledge 를 탐색
- Causal Graph, observation, Action
  - Total Node N = 5
  - Each Node X_i = Gaussian Random Variable
  - Root node of G Always Hidden, 그래서 Agent 는 오직 oither 4node 만 볼 수 있음
  - 그리고 Quiz phase 에서 Intervention 는 Mt, Value of the node 는 V
    - Observation : O_t = [vt, mt]
  - Information Action:
    - Node 선택(N-1)
    - Information phase 에서 quiz action 을 하면 penalty rt=-10
  - Active vs Random Agent

- Information 단계
  - Random condition :Random choice node
  - Active condition: 학습을 바탕으로 선택
- 2 종류의 Learning
  - Inner loop learning : Quiz Phase 를 잘 수행하기 위해서 정보를 모으고 학습
  - Out loop learning: Information phase 의 observation 을 바탕으로 causal dependencies 를 discover 하기 위한 능력학습
- 학습:
  - LSTM base – A3C 학습

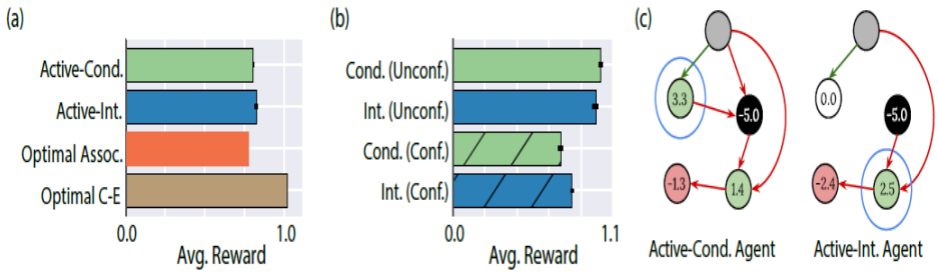
- 4. Experiments
  - Observation, Interventional, counterfactual
  - Performance measure 은 CBN 을 고정하고, Quiz phase 에서 reward 계산
  - 4 Agent 학습:
    - Observational, Long Observational, Active Conditional, Random Conditional, Optimal Associative Baseline
    - Observational Agent:
      - Longer -> *4 episode
    - Conditional Agent:
      - 특정 노드를 정했을 때의 조건부 확률 계산
        - Random 과 active 버전이 있음
    - Optimal Associative Baseline
      - Oracle baseline
      - 상관관계에서 얻을 수 있는 최대 보상을 받는다
  - Ex1. Result
    - 질문: 에이전트는 관측 데이터를 사용하여 원인 - 원인 추론을 수행하는 방법을 배우는가?



        - (C) 설명:
            - black node :intervened node
            - Green : Positive
            - Red: Negative
            - White: 0
            - Blue cycle : Agent select
        - (b) 중간에 intervened 했을 때 개입한 노드의 부모가 있다면 Cond 가 더 좋다. 하지만 없으면 비슷하다
    - 결론은: Cause-effect reasoning 을 한다
    - 질문: Agent 가 useful observation 을 선택하는 것을 배우냐?



- Ex2, Result(interventional setting)
    - Interventional data 를 Agent 가 받았을 때 경우
    - 질문: Agent 가 interventional data 를 받았을 때 cause-effect reasoning 을 잘 처리 할 수 있냐



    - (a) Active intervention Agent 가 좀더 좋다
    - (b) Confounder 이 있을 때 Intervention agent 가 더 좋다
    - © unobserved 된 Confounder 가 있을 때 Active intervention Agent 가 더 좋다
- Ex3, Result(Counterfactual setting)
    - 3 개 Counterfactual Agent:
        - Active, Random, Optimal


(a)
Active-Int.
Active-CF
Optimal C-E
Optimal CF
0.0
1.0
Avg. Reward
(b)
CF (Dist.)
Int. (Dist.)
CF (Deg.)
Int. (Deg.)
0.0
1.0
2.0
Avg. Reward
(c)
0.0
0.0
-1.7
-5.0
Passive-Int. Agent
0.6
0.2
-1.5
-5.0
Passive-CF Agent

-
-