
Causal Inference: Question of Why?

DMQA Open Seminar

Changhyun Kim

Department of Industrial and Management Engineering

Korea University

April 07, 2023





❖ 김창현 (Changhyun Kim)

- 고려대학교 산업경영공학과 석사과정 (2021.09 ~)
Data Mining & Quality Analytics Lab (김성범 교수님)

❖ Research Interest

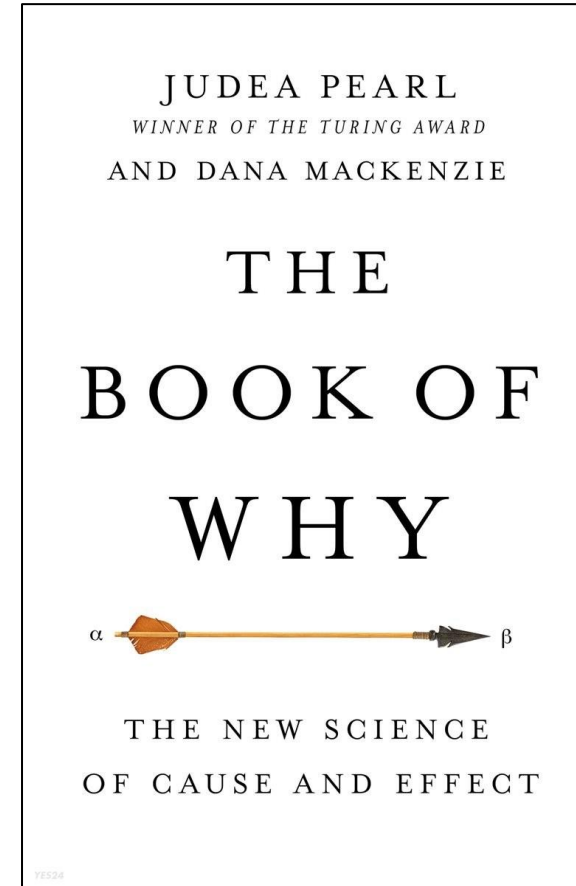
- Forecasting models
- Application of deep learning and machine learning algorithms

❖ Contact

- E-mail: charlie17@korea.ac.kr

목차

- 인과추론(Causal Inference)이란?
- 상관관계(Correlation)와 인과관계(Causality)
- 인과관계 분석 방법
 1. Randomization
 2. Causal Graphical Models
 3. Potential Outcome
- 인과관계 분석 도구
- 결론



인과추론(Causal Inference)이란?

창현

What is Causality?



Causality refers to the relationship between an event (the cause) and a second event (the effect), where the second event is understood as a consequence of the first. In other words, causality is the relationship between cause and effect, where the cause is the reason why something happens, and the effect is the result of that cause.

Causality is a fundamental concept in many fields, including philosophy, physics, psychology, and economics. It is often used to explain and predict the behavior of systems, to test theories, and to establish the basis for scientific knowledge. In order to establish causality, researchers must use methods that allow them to rule out alternative explanations for the relationship between two events, such as spurious correlations or coincidence.



창현

What is causal inference



Causal inference is the process of drawing a conclusion about causality based on the evidence available, which may include observations, experiments, and statistical analysis. In other words, it is the process of making inferences or conclusions about the causal relationship between two variables, based on data or other types of evidence.

Causal inference is a critical aspect of scientific inquiry, as it allows us to establish whether there is a causal relationship between variables of interest, and to understand how changes in one variable may affect another. This is important in a wide range of fields, including medicine, epidemiology, economics, and public policy, among others.

However, establishing causal inference is not always straightforward, as there may be confounding variables or other factors that can affect the relationship between the variables of interest. To address these challenges, researchers use various statistical and experimental methods, including randomized controlled trials, natural experiments, and observational studies with rigorous adjustment for confounding variables.

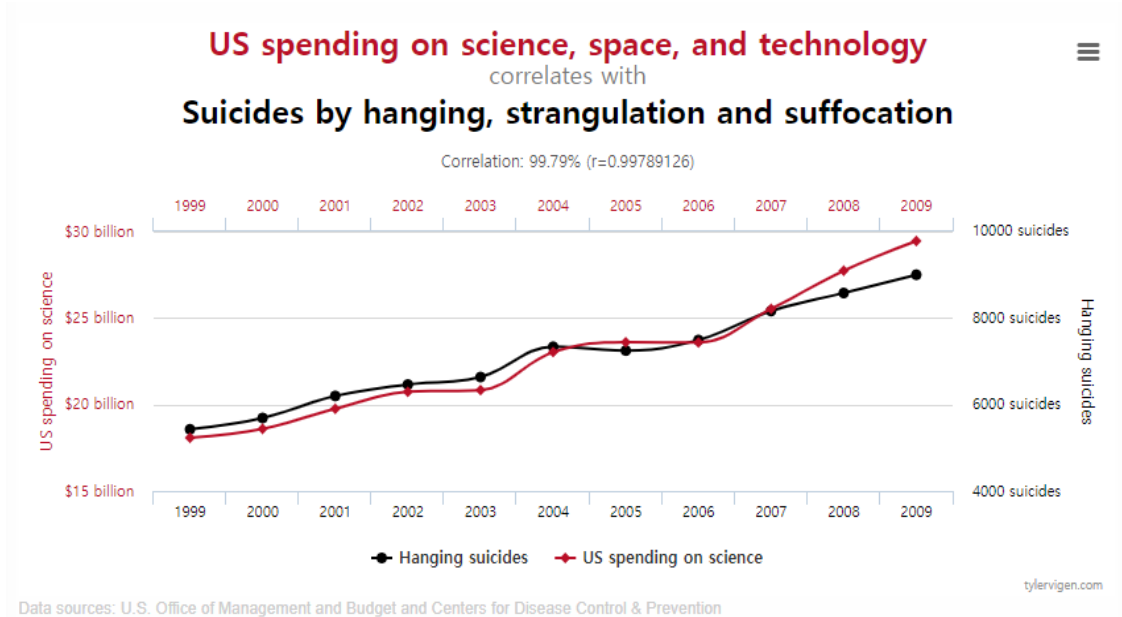
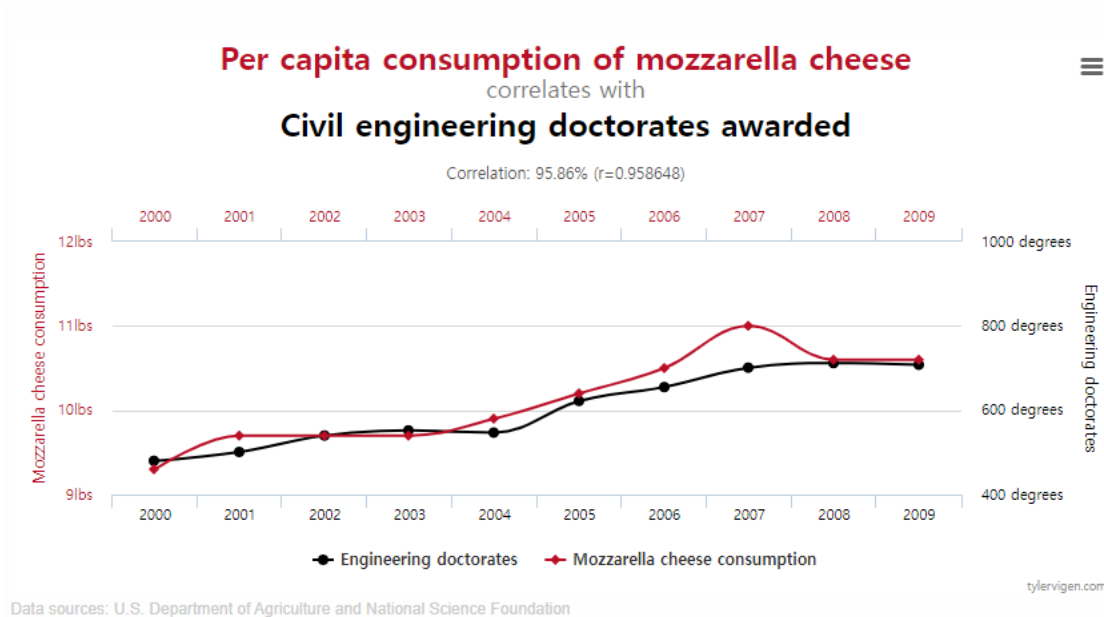


인과추론은 어떤 사건이 다른 사건의 원인이 되었다고 추론하는 것
(원인과 결과 사이에 인과관계가 있다 추론!)

상관관계(Correlation)와 인과관계(Causality)

상관관계는 인과관계를 의미하지 않음

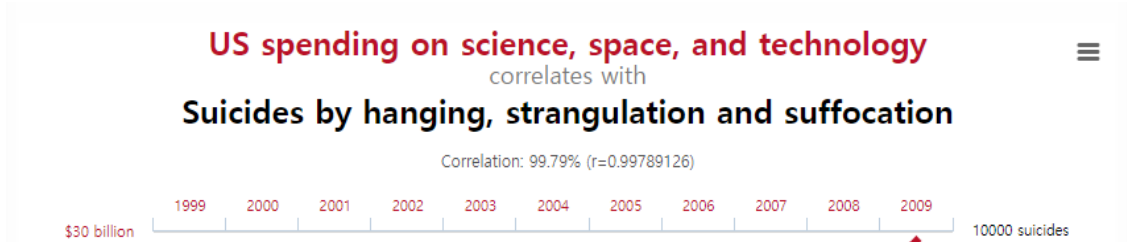
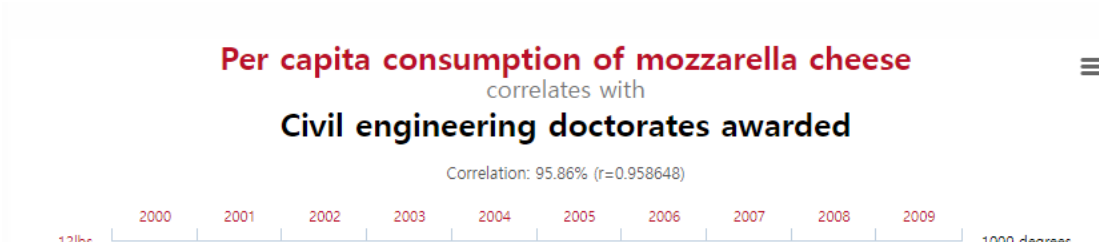
- 상관관계: 통계적 변수와 다른 통계적 변수들이 공변(covariance)하는 관계 → **예측** 목적
- 인과관계: 선행하는 한 변수가 후행하는 다른 변수의 원인이 되고 있다고 믿어지는 관계 → **원인 설명** 목적



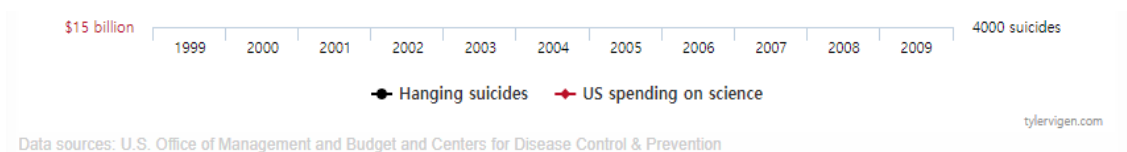
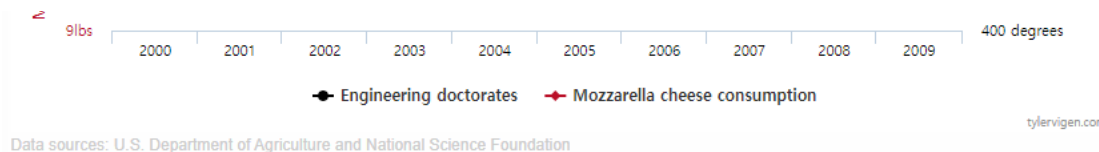
상관관계(Correlation)와 인과관계(Causality)

상관관계는 인과관계를 의미하지 않음

- 상관관계: 통계적 변수와 다른 통계적 변수들이 공변(covariance)하는 관계 → **예측** 목적
- 인과관계: 선행하는 한 변수가 후행하는 다른 변수의 원인이 되고 있다고 믿어지는 관계 → **원인 설명** 목적



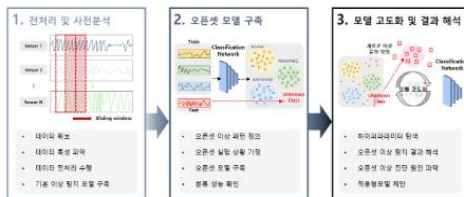
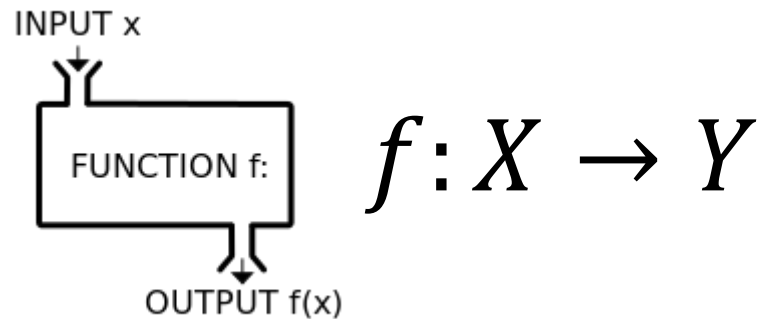
상관관계 $\neq$ 인과관계



AI 상관관계를 학습하는데 용이함

AI는 데이터의 상관관계 패턴을 학습함

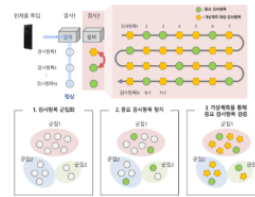
- Input(X)와 Output(Y)간 상관관계를 학습하고 이에 기반하여 예측, 분류 등 task 수행
- 외부 환경에 따른 데이터의 변화나 패턴 변화가 적은 분야에서는 성능이 뛰어남 (NLP, Vision, Collaboration Filtering)



Machine Learning Algorithms

알려지지 않은 이상 패턴 탐지를 위한 오픈셋 분류 AI 기술 개발(삼성전자)

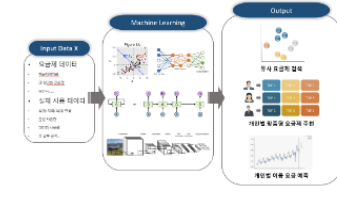
분류



Machine Learning Algorithms

휴대폰 검사 효율화를 위한 예측 모델 연구(삼성전자)

예측



Machine Learning Algorithms

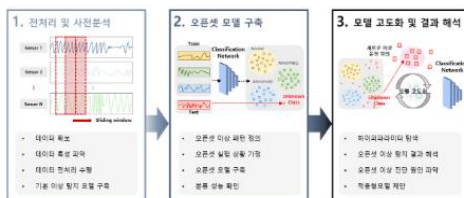
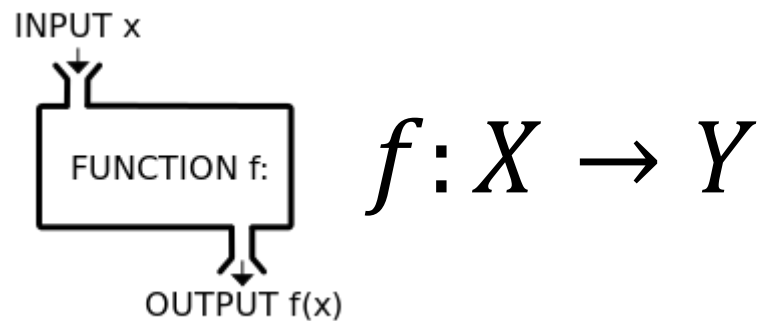
이동통신 요금제 추천 빅데이터 분석 솔루션(다음소프트)

추천

AI 상관관계를 학습하는데 용이함

AI는 데이터의 상관관계 패턴을 학습함

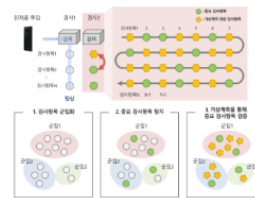
- Input(X)와 Output(Y)간 상관관계를 학습하고 이에 기반하여 예측, 분류 등 task 수행
- 외부 환경에 따른 데이터의 변화나 패턴 변화가 적은 분야에서는 성능이 뛰어남 (NLP, Vision, Collaboration Filtering)
- 그러나 왜? 분류, 예측, 추천의 결과가 나왔는지에 대해서는 **궁극적으로 설명하지 못함**



Machine Learning Algorithms

알려지지 않은 이상 패턴 탐지를 위한 오픈셋 분류 AI 기술 개발(삼성전자)

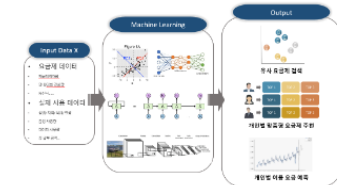
분류



Machine Learning Algorithms

휴대폰 검사 효율화를 위한 예측 모델 연구(삼성전자)

예측



Machine Learning Algorithms

이동통신 요금제 추천 빅데이터 분석 솔루션(다음소프트)

추천

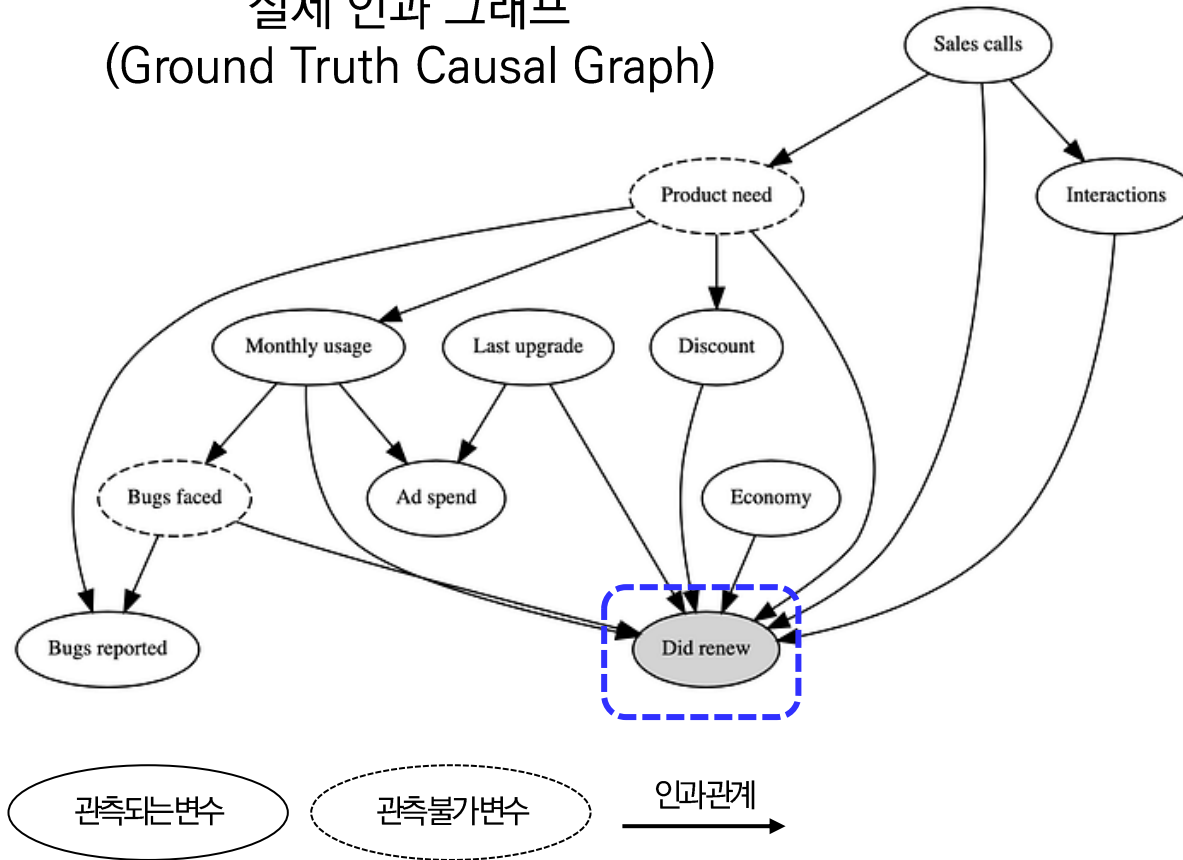
XAI로도 인과관계는 알 수 없음

XGBoost

$$f: X \rightarrow Y$$

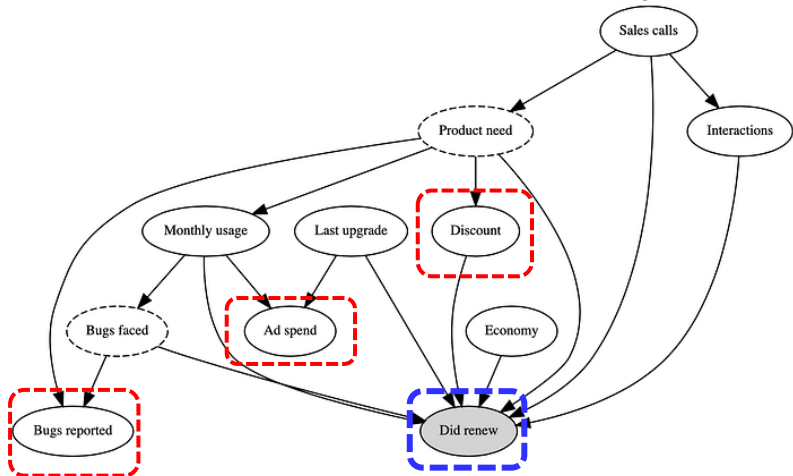
Sales Calls
Interactions
Economy
Last Upgrade
Discount
Monthly Usage
Ad Spend
Bugs Reported
Did Renew
(구독갱신)

실제 인과 그래프
(Ground Truth Causal Graph)



XAI로도 인과관계는 알 수 없음

실제 인과 그래프
(Ground Truth Causal Graph)



XGBoost

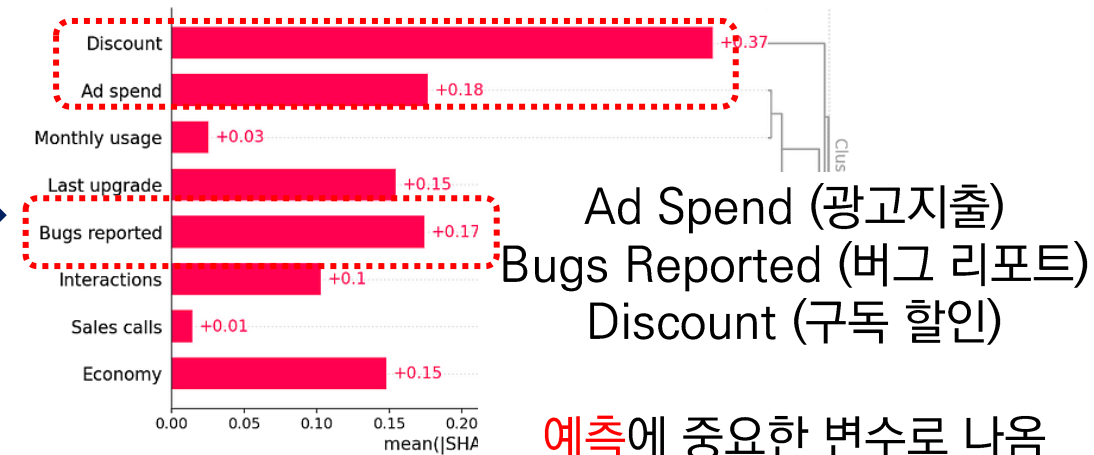
$$f: X \rightarrow Y$$

Sales Calls
Interactions
Economy
Last Upgrade
Discount
Monthly Usage
Ad Spend
Bugs Reported

Did Renew
(구독갱신)



SHAP 기반 변수 중요도

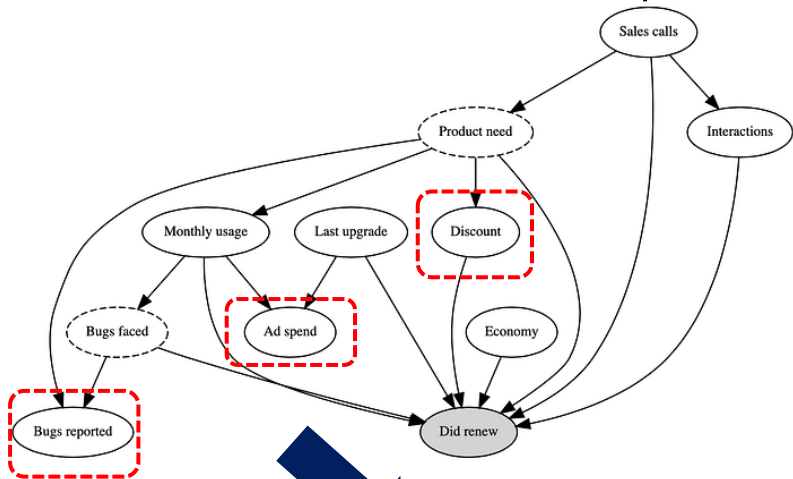


Ad Spend (광고지출)
Bugs Reported (버그 리포트)
Discount (구독 할인)

예측에 중요한 변수로 나옴
(높은 상관관계)

XAI로도 인과관계는 알 수 없음

실제 인과 그래프
(Ground Truth Causal Graph)

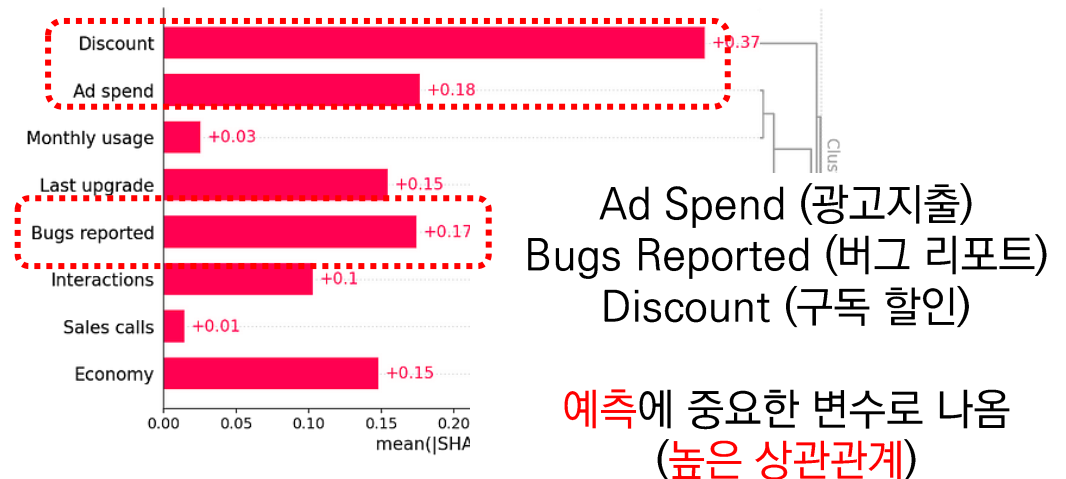


XGBoost
 $f: X \rightarrow Y$

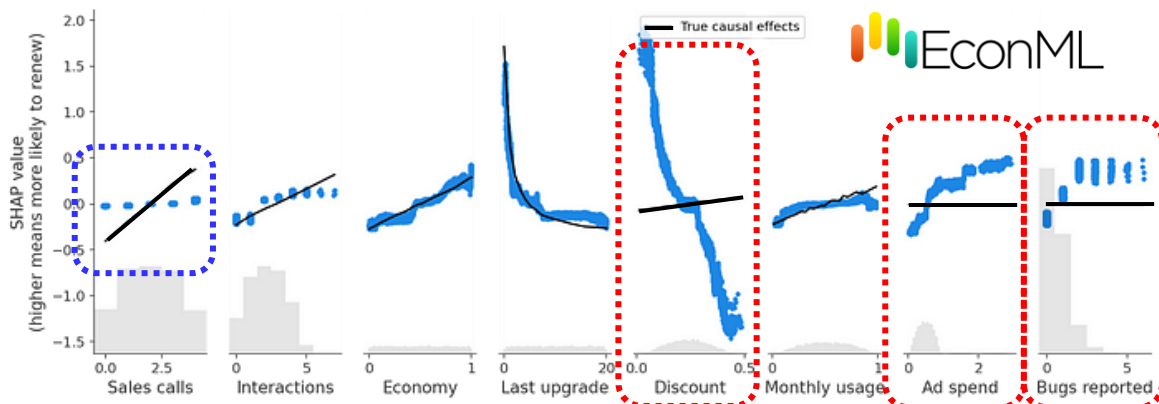
Sales Calls
Interactions
Economy
Last Upgrade
Discount
Monthly Usage
Ad Spend
Bugs Reported

Did Renew
(구독갱신)

SHAP 기반 변수 중요도



SHAP 변수 중요도 vs 실제 인과효과 (True Causal Effects)



Ad Spend(광고 지출),
Bugs Reported(버그 리포트)
구독 갱신과 **인과 관계 없음**
Discount(구독 할인)
음의 상관관계였지만 **양의 인과관계**

Sales Calls(영업전화)
상관관계는 없지만 양의 인과관계

그래서 인과추론 어디에 사용하는데?

예측은 됐고
그래서 구독갱신
왜 올랐는데?

After

구독갱신 예측엔
Ad Spend,
Bugs Reported,
Discount 변수가
큰 상관관계를 보여줬습니다.

Data
Analyst

Sales Calls.

CEO

Roll Series ¹		WIN %		RECORD	
Roll Series ²	Roll Series ³ Roll Series ⁴ = Win %	EXPECTED	ACTUAL	EXPECTED	ACTUAL
OAK	88V ¹ 78V ² 6	.6525	.6296	106-56	102-60
2001	88V ¹ 69V ² 8	.6145	—	99-63	0-0
OAK	88V ¹ 69V ² 8	.6145	—	99-63	0-0
PROTECTION	88V ¹ 69V ² 8	.6145	—	99-63	0-0
SEA	91V ¹ 85V ² 8	.6861	.7160	111-51	—
2001	91V ¹ 85V ² 8	.6861	.7160	111-51	—
	71V ¹ 50V ² 0	.6047	—	99-63	0-0
	71V ¹ 50V ² 0	.6047	—	99-63	0-0
	69V ¹ 88V ² 8	.5597	.5804	91-71	—
	69V ¹ 88V ² 8	.5597	.5804	91-71	—
	69V ¹ 88V ² 8	.6178	—	100-62	—

인과관계 분석 방법

인과관계 분석 방법은 크게 3가지로 나뉨

- Randomization (랜덤화 추출)
- Causal Graphical Models (인과 그래프 모형)
- Potential Outcome (잠재적 결과)

1. Randomization

- Randomized Clinical Trial (A/B Test)
- Reinforcement Learning Based
 - Multi-Armed Bandit (MAB)

2. Causal Graphical Models

- Directed Acyclic Graphs (DAG)
 - Greedy 알고리즘
 - Epsilon Greedy 알고리즘
- Upper Confidence Bound 알고리즘

3. Potential Outcome

- Machine Learning Based
 - Meta Learners (T, S, X)
- Neural Net Based
 - TARNet
 - CFRNet

인과관계 분석 방법1

Randomization (랜덤화 추출): Randomized Clinical Test

- Randomized Clinical Test (무작위 임상 시험) → 현실에서는 A/B test 라고 불림
- 측정하고자 하는 변수 이외에는 모든 것들을 최대한 고정하고 확인하고 싶은 변수만 변경하면서 실험하는 방법 → 실험 기반
- 무작위 배정으로 평균적인 효과를 측정할 수 있음

Product: A

Product A: Standard box art



Product: B

Product B: Upside-down box art



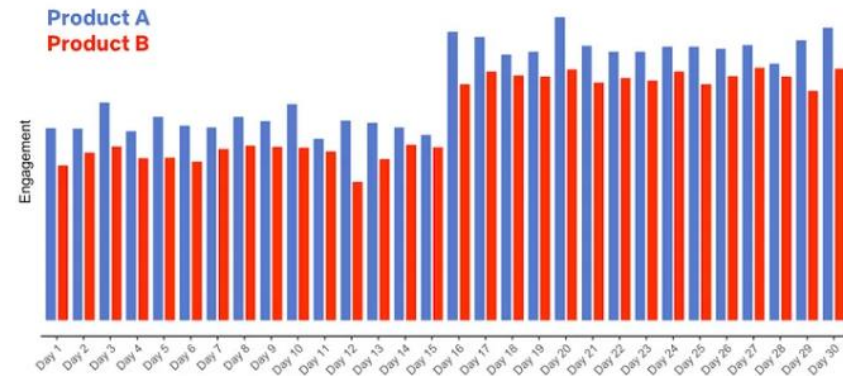
서로 다른 화면 (뒤집혀 있음)

Netflix Members



무작위 배정

Compare member behavior



평균적인 효과 측정

인과관계 분석 방법1

Randomization (랜덤화 추출): Randomized Clinical Test

Q: Product B가 A보다 좋은 방법인가???

	Product A	Product B
시행 수	1000명	1100번
클릭 수	20명	30명
클릭률(표본비율)	0.02	0.027

두 집단의 분포는 이항 분포를 따름 (클릭을 했냐 vs 안 했냐)

$$n_1 = 1000, \quad n_2 = 1100$$
$$p_1 = \frac{20}{1000}, \quad p_2 = \frac{30}{1100}$$

귀무가설: $p_1 - p_2 = 0$

대립가설: $p_1 - p_2 < 0$ or $p_1 - p_2 > 0$



두 집단의 모비율 차이에 대한 가설 검정 필요 !!

$$\tilde{p}_1 - \tilde{p}_2 \sim N(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}) \quad \text{by 중심극한 정리}$$

$$Z \text{ 검정 통계량} = \frac{\bar{X} - \mu}{SE} = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

$$\tilde{p} = \frac{m_1 + m_2}{p_1 + p_2} = \frac{20 + 30}{1000 + 1100} = 0.024 \quad \text{by 합동표본비율}$$

$$Z \text{ 검정 통계량} = \frac{(\bar{p}_1 - \bar{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}} = \frac{(0.02 - 0.027) - (0)}{\sqrt{\frac{0.024(0.0976)}{1000} + \frac{0.024(0.0976)}{1100}}} = -1.05$$

이때의 p-value는 0.1469

A: Product B가 Product A 보다 낮다고 판단할 수 없음

인과관계 분석 방법1

Randomization (랜덤화 추출): Multi-Armed Bandit (MAB)

- A/B 테스트는 탐색-수확 Tradeoff 가 존재 → MAB은 강화학습 기반으로 탐색과 활용을 최적화하여 수익률을 극대화!
- 탐색(Exploration) 관점

모든 Bandit Machine 계속 당겨보자 (A가 더 좋아 보이는데 혹시 모르니 B도 테스트) → 예상대로 A가 좋았다 → 비용 발생!

- 활용(Exploitation) 관점

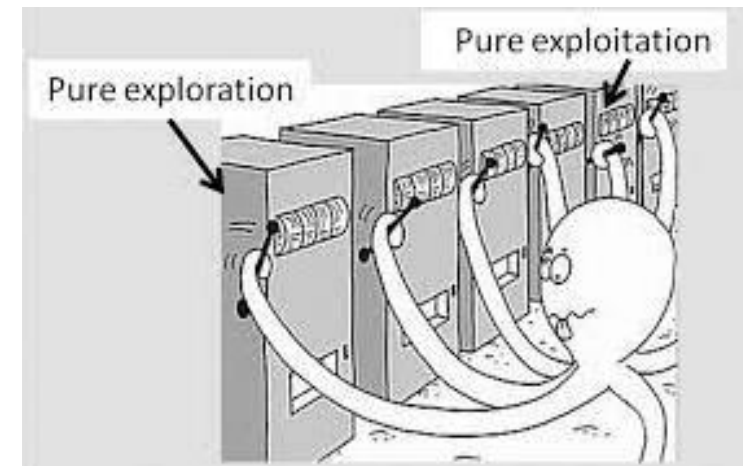
모든 Bandit Machine을 한 두 번만 당겨보고 가장 수익률이 높은 머신을 선택 → 신뢰성 문제 → 만약 테스트를 더 오래하면?

카지노의 Bandit Machine은
당첨 확률이 모두 다름

Bandit Machine



Multi-Armed Bandit (MAB)



인과관계 분석 방법1

Randomization (랜덤화 추출): Multi-Armed Bandit (MAB)

- Greedy 알고리즘

한 번씩 해보고 Reward가 가장 좋은 알고리즘 선택 $A_t = \operatorname{argmax}_a Q_t(a)$
(현재 시점 t 까지 기대 보상 $Q_t(a)$ 를 최대화 하는 Action a 선택)

- Epsilon Greedy 알고리즘

$1-\epsilon$ 의 확률로는 Greedy 알고리즘 $A_t = \operatorname{argmax}_a Q_t(a)$, ϵ 의 확률로 랜덤하게 선택

- Upper Confidence Bound (UCB) 알고리즘

Epsilon Greedy는 최적값과 멀어지게 탐색을 과도하게 하는 문제 존재

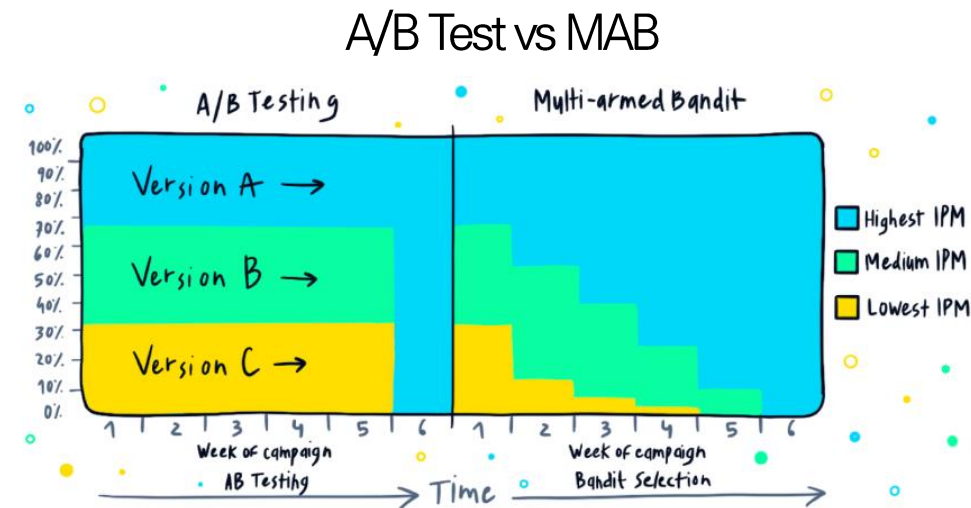
→ 일종의 신뢰 구간을 설정 $N_t(a)$ = 현재 시점까지 행동 a 를 한 횟수, c = Hyperparameter

$A_t = \operatorname{argmax}_a [Q_t(a) + c \sqrt{\frac{\ln t}{N_t(a)}}]$ → 분모에 $N_t(a)$ 를 넣어서 탐험하지 않은 행동에 가중치 부여

행동(Action) = MAB에서 선택된 제품 (ex. A안, B안) = 시점 t 의 행동 = A_t

보상(Reward) = 한 번의 행동에 따른 수치화된 결과 (ex. 클릭, 구매) = R_t

가치(Value) = 행동으로 인한 기대 보상 = 시점 t 에 추정된 가치 = $Q_t(a)$



인과관계 분석 방법2

Causal Graphical Models (인과 그래프 모형): Directed Acyclic Graphs (DAG)

- Randomization은 평균적인 인과 효과만을 확인할 수 있음 (어떤 사람들에게 더 효과적인지 세부적인 정보 확인 어려움)
- 인과 그래프 모형은 인과 관계를 Directed Acyclic Graphs (DAG) 라는 그래프로 표현하는 방식! → 명확한 인과관계 파악 가능

Causal Graphical Models은 Bayesian Network라고 불림

변수들의 결합 확률 분포를 DAG로 나타낸 모형으로, 딥러닝과 달리 조건부 확률이 투명하게 공개되므로 “White-Box Model”

Bayesian Network $B = \langle G, \Theta \rangle \rightarrow$ 두 개의 요소로 구성되어 있음

$G = \text{Directed Acyclic Graphs (DAG)}$

$\Theta = \text{조건부 확률 (conditional probability)의 집합}$

} 이게 뭔데?!

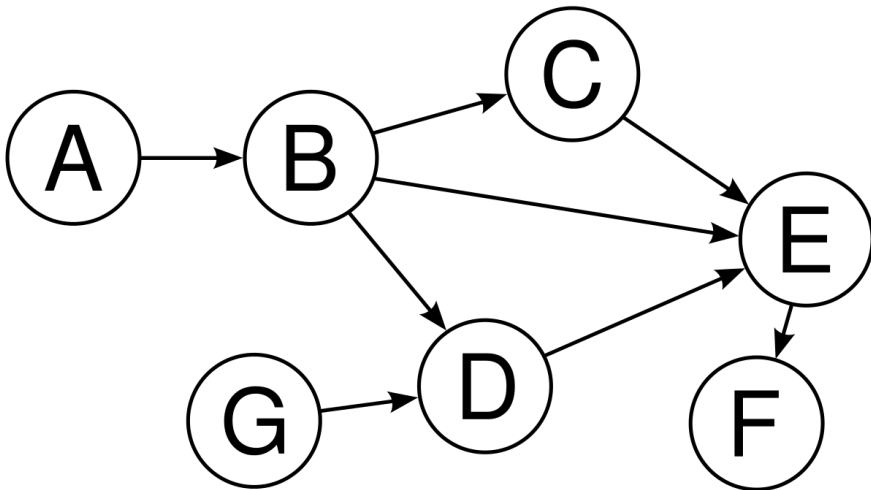
인과관계 분석 방법2

Causal Graphical Models (인과 그래프 모형): Directed Acyclic Graphs (DAG)

- $G = \text{Directed Acyclic Graphs (DAG)}$, $\Theta =$ 조건부 확률 (*conditional probability*)의 집합
- Causal Graphical Models = Bayesian Network = $B = \langle G, \Theta \rangle \rightarrow$ 두 개의 요소로 구성되어 있음
- Bayesian Network는 여러 인과 관계를 조건부 확률로 표현

Directed Acyclic Graphs (DAG)

방향은(Directed) 있고 순환성은 없는(Acyclic) 그래프를 의미함



$Node = \{A, B, C, D, E, F, G\}$

$Edge = \rightarrow$

ex. $A \rightarrow B$: A가 B의 원인

A: 부모(parent), B: 자식(child)

인과관계 분석 방법2

Conditional Probability (Y가 y로 주어졌을 때 X의 확률)

$$P(X|Y = y) = \frac{P(X, Y = y)}{P(Y = y)} = \frac{P(Y|X)P(X)}{\sum_{x \in X} P(Y|X = x)P(X = x)}$$

Causal Graphical Models (인과 그래프 모형): Directed Acyclic Graphs (DAG)

- $G = \text{Directed Acyclic Graphs (DAG)}$, $\Theta = \text{조건부 확률 (conditional probability)의 집합}$
- Causal Graphical Models = Bayesian Network = $B = \langle G, \Theta \rangle \rightarrow$ 두 개의 요소로 구성되어 있음
- Bayesian Network는 여러 인과 관계를 조건부 확률로 표현

$\Theta =$ 변수 X_i 의 모든 가능한 값 x_i 에 대해 $\theta_{x_i|\Pi_{X_i}} = P_B(x_i|\Pi_{X_i})$ 의 확률들의 집합

* Π_{X_i} = DAG에서 표현된 X_i 의 직계 부모의 집합

Bayesian Network B 는 다음의 결합 확률 분포 표현 가능

$$P_B(X_1, \dots, X_n) = \prod_{i=1}^n P_B(X_i|\Pi_{X_i}) = \prod_{i=1}^n \theta_{X_i|\Pi_{X_i}}$$

예시

Directed Acyclic Graphs (DAG)

$$X_1 \rightarrow X_2 \rightarrow X_3$$

X_1 의 직계 부모 없음, X_2 의 직계 부모는 X_1 , X_3 의 직계 부모는 X_2

첫번째 확률 $\theta_{X_1|\Pi_{X_1}} = P_B(X_1)$

두 번째 확률 $\theta_{X_2|\Pi_{X_2}} = P_B(X_2|X_1)$

세 번째 확률 $\theta_{X_3|\Pi_{X_3}} = P_B(X_3|X_2)$

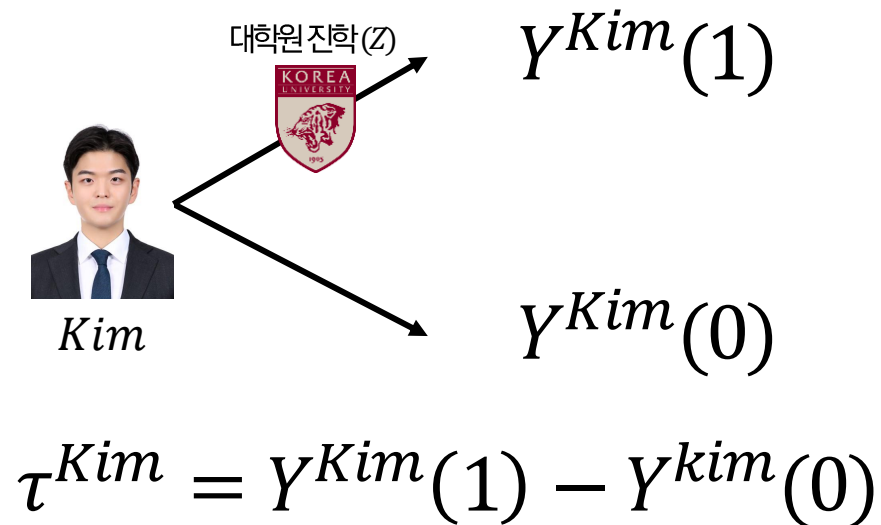
$$\begin{aligned} \text{DAG의 결합 확률 분포} &= P_B(X_1, X_2, X_3) \\ &= P_B(X_1)P_B(X_2|X_1)P_B(X_3|X_1, X_2) \\ &= P_B(X_1)P_B(X_2|X_1)P_B(X_3|X_2) \end{aligned}$$

인과관계 분석 방법3

Potential Outcome (잠재적 결과):

- 인과관계는 모든 결과를 관측할 수 없는 근본적인 한계가 존재 (현실에서는 시간을 돌릴 수 없음)
- 관측되지 않은 “가상의 결과”를 고려해서 처리 효과를 계산하는 방식 → “이렇게 했다면 어땠을까?”
- Treatment Effect (처리 효과) $\tau = Y(1) - Y(0)$

대학원 진학과 계약연봉에 대한 인과 효과 (이론)



하지만 현실에서는 진학했거나 진학하지 않은 경우만 관측 가능함!

대학원에 진학했다면, $Y^{kim} = Y^{Kim}(1)$

대학원에 진학하지 않았다면 어땠을까? $Y^{Kim}(0)$

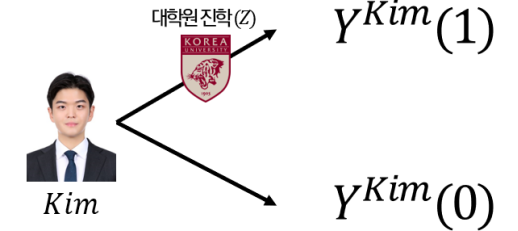
인과추론은 Counterfactual (반사실)을 추정하여 인과효과를 추정하는 기술!

인과관계 분석 방법3

Potential Outcome (잠재적 결과):

- 개인에 대한 효과를 측정하는 대신 모든 사람의 효과를 비교하고 평균을 구하자
- Average Treatment Effect (ATE)

대학원 진학과 계약연봉에 대한 인과 효과 (이론)



$$\tau^{Kim} = Y^{Kim}(1) - Y^{Kim}(0)$$

성별	나이	처리(z_i)	$Y_i(0)$	$Y_i(1)$
남	29	1	?	6000
남	31	0	6000	?
여	25	1	?	7000
여	27	0	4000	?



$Y_i(0)$	$Y_i(1)$
5500	6000
6000	6300
6700	7000
4000	4500

처리 효과(τ_i)
$Y_1(1) - Y_1(0)$
$Y_2(1) - Y_2(0)$
$Y_3(1) - Y_3(0)$
$Y_4(1) - Y_4(0)$

⋮

빈 칸을 그냥 머신러닝으로 예측해서 채워 넣자!

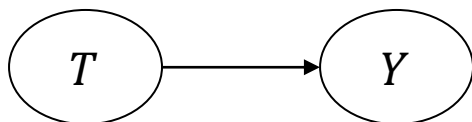
평균적인 효과 측정 가능 (ATE)

$$\tau_{ATE} = E[Y(1)] - E[Y(0)]$$

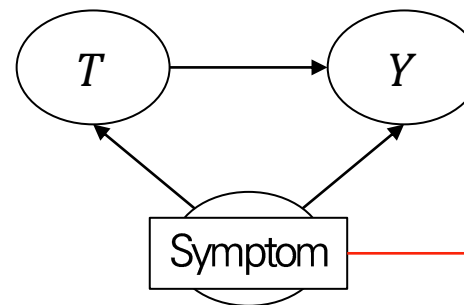
인과관계 분석 방법3

Potential Outcome (잠재적 결과):

- Simpson's Paradox: 각 변수를 고려하지 않고 전체 통계 결과를 유추하다 발생하는 오류
- Confounder(교란 변수)는 처리(Treatment)와 결과(Y)에 모두 영향을 주는 변수임
- Confounder(교란 변수)를 고려했을 때 처리효과는?



Death Rate	Mild Symptom	Severe Symptom	Total
Treatment A	15% (195/1300)	30% (30/100)	<u>16.1%</u> (225/1400)
Treatment B	<u>10%</u> (10/100)	<u>20%</u> (100/500)	18.3% (110/600)



Confounder (교란변수 Z)!

Death Rate	Mild Symptom	Severe Symptom	Total	Causal Effect
Treatment A	15% (195/1300)	30% (30/100)	<u>16.1%</u> (225/1400)	19.5%
Treatment B	<u>10%</u> (10/100)	<u>20%</u> (100/500)	18.3% (110/600)	<u>13.0%</u>

$$\sum_c P(Y|T, S)P(S)$$

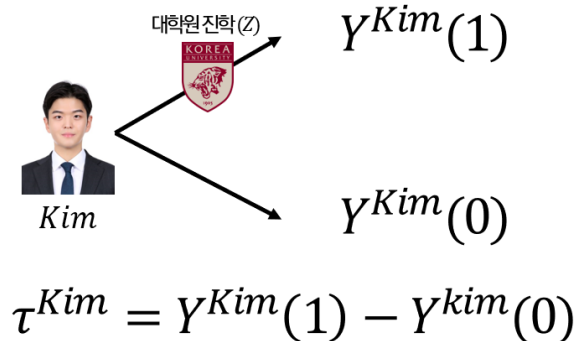
$$P(Y|T, S = mild) \quad P(Y|T, S = severe) \quad P(Y|T)$$

인과관계 분석 방법3

Potential Outcome (잠재적 결과):

- Confounder(교란 변수, Z)를 고려했을 때 처리효과는 $\rightarrow$ Conditional Average Treatment Effect (CATE)

대학원 진학과 계약연봉에 대한 인과 효과 (이론)



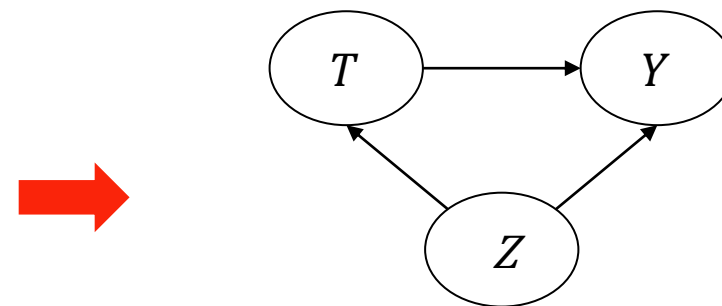
$$\tau_i = Y_i(1) - Y_i(0)$$

개별 수준 처리효과는 관측 불가

처리 효과(τ_i)
$Y_1(1) - Y_1(0)$
$Y_2(1) - Y_2(0)$
$Y_3(1) - Y_3(0)$
$Y_4(1) - Y_4(0)$

$$\tau_{ATE} = E[Y(1)] - E[Y(0)]$$

ATE: 교란변수 관측 불가



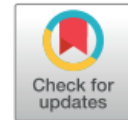
$$\tau_{CATE} = \frac{1}{n} \sum_{i=1}^n \{E[Y_i(1)|Z_i] - E[Y_i(0)|Z_i]\}$$

CATE: 교란변수까지 관측 가능

인과관계 분석 방법3

Potential Outcome (잠재적 결과): Machine Learning Based

- Metalearners for estimating heterogeneous treatment effects using machine learning (607회 인용)
- Machine Learning을 기반으로 CATE를 추정하는 방법
- Meta Learners (T-Learner, S-Learner, X-Learner)



Metalearners for estimating heterogeneous treatment effects using machine learning

Sören R. Künnel^{a,1}, Jasjeet S. Sekhon^{a,b}, Peter J. Bickel^a, and Bin Yu^{a,c,1}

^aDepartment of Statistics, University of California, Berkeley, CA 94720; ^bDepartment of Political Science, University of California, Berkeley, CA 94720; and ^cDepartment of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720

Contributed by Bin Yu, December 18, 2018 (sent for review March 16, 2018; reviewed by Jake Bowers and Dylan Small)

인과관계 분석 방법3

Potential Outcome (잠재적 결과): Machine Learning Based

- T (Two Model) – Learner
- 1. 처리를 받지 않은 개체와 처리를 받은 개체를 각각 다른 모형 M 을 사용해서 학습
- 2. 이들의 차이로 CATE 추정

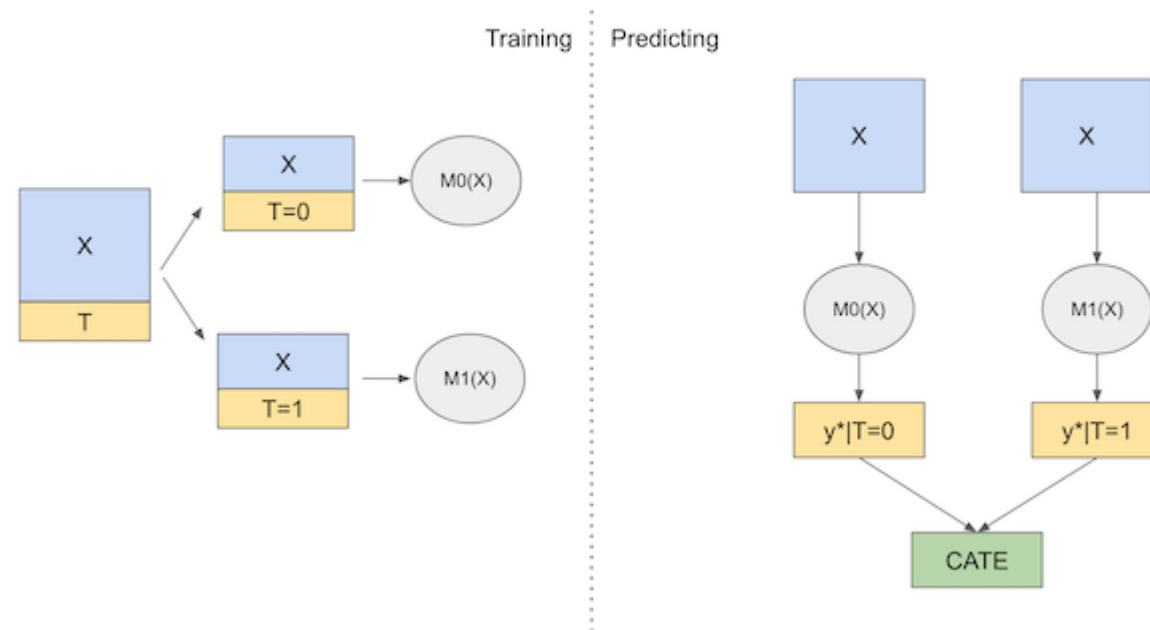
Algorithm 1 T-learner

```
1: procedure T-LEARNER( $X, Y, W$ )  
2:    $\hat{\mu}_0 = M_0(Y^0 \sim X^0)$   
3:    $\hat{\mu}_1 = M_1(Y^1 \sim X^1)$   
  
4:    $\hat{\tau}(x) = \hat{\mu}_1(x) - \hat{\mu}_0(x)$ 
```

M_0 and M_1 are here some, possibly different, machine-learning/regression algorithms.

(X^0, Y^0) : 처리받지 않은 변수와 반응값, (X^1, Y^1) : 처리 받은 변수와 반응값
 $\hat{\mu}_0: E[Y(0)|X = x]$ 의 추정값, $\hat{\mu}_1: E[Y(1)|X = x]$ 의 추정값

$$CATE = \hat{T}(x) = \hat{\mu}_1(x) - \hat{\mu}_0(x)$$



인과관계 분석 방법3

Potential Outcome (잠재적 결과): Machine Learning Based

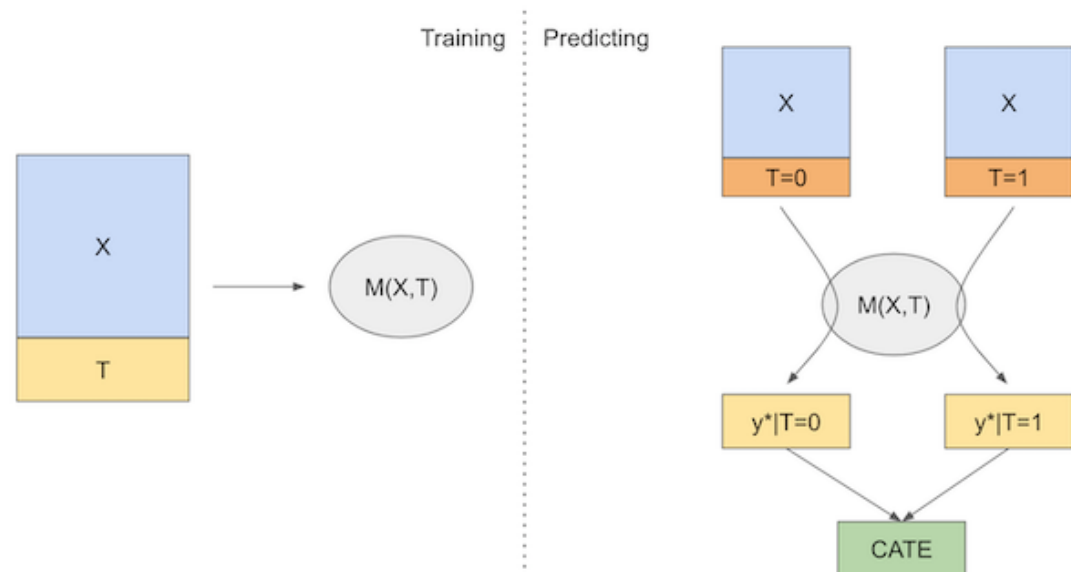
- S (Single Model) – Learner
- 1. Y 를 반응변수로 두고 처리 여부를 같이 설명 변수로 설정하여 모형 M 학습
- 2. $W = 1$ 일 때의 추정값에서 $W = 0$ 일 때의 추정값의 차이로 CATE 추정

Algorithm 2 S-learner

```
1: procedure S-LEARNER( $X, Y, W$ )  
2:    $\hat{\mu} = M(Y \sim (X, W))$   
3:    $\hat{\tau}(x) = \hat{\mu}(x, 1) - \hat{\mu}(x, 0)$ 
```

$M(Y \sim (X, W))$ is the notation for estimating $(x, w) \mapsto \mathbb{E}[Y|X = x, W = w]$ while treating W as a 0,1-valued feature.

$$CATE = \hat{T}(x) = \hat{\mu}(x, 1) - \hat{\mu}(x, 0)$$



인과관계 분석 방법3

Potential Outcome (잠재적 결과): Machine Learning Based

- X (Cross) – Learner
- 1. T-Learner 처럼 처리를 받은 개체와 처리를 받지 않은 개체를 나누어 $\hat{\mu}_0, \hat{\mu}_1$ 계산 → First Stage
- 2. 이 모델의 예측값을 다시 Counterfactual (반사실)로 사용하여 X와 차이를 학습 → CATE 0, CATE 1
- 3. 마지막으로 CATE에 가중 평균 $g(x)$ 을 곱하여 CATE Final 계산
- 4. T-Learner가 가지고 있는 대상군, 대조군 사이의 데이터 양의 차이를 해결함

Algorithm 3 X-learner

1: **procedure** X-LEARNER(X, Y, W, g)

2: $\hat{\mu}_0 = M_1(Y^0 \sim X^0)$

3: $\hat{\mu}_1 = M_2(Y^1 \sim X^1)$

4: $\tilde{D}_i^1 = Y_i^1 - \hat{\mu}_0(X_i^1)$

5: $\tilde{D}_i^0 = \hat{\mu}_1(X_i^0) - Y_i^0$

6: $\hat{\tau}_1 = M_3(\tilde{D}^1 \sim X^1)$

7: $\hat{\tau}_0 = M_4(\tilde{D}^0 \sim X^0)$

8: $\hat{\tau}(x) = g(x)\hat{\tau}_1(x) + (1 - g(x))\hat{\tau}_0(x)$

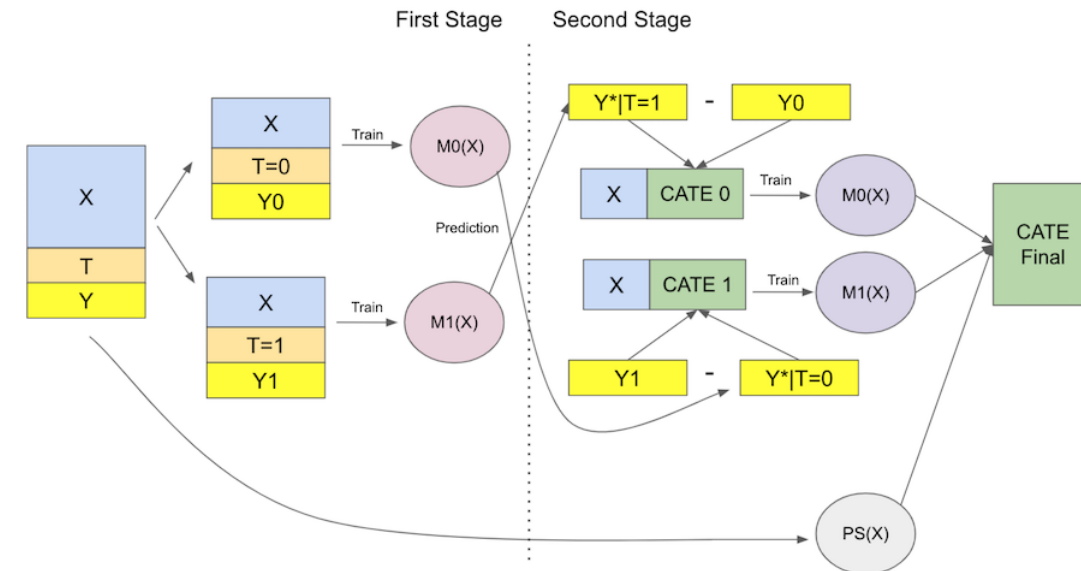
$g(x) \in [0, 1]$ is a weighting function that is chosen to minimize the variance of $\hat{\tau}(x)$. It is sometimes possible to estimate $\text{Cov}(\tau_0(x), \tau_1(x))$, and compute the best g based on this estimate. However, we have made good experiences by choosing g to be an estimate of the propensity score.

▷ Estimate response function

▷ Compute imputed treatment effects

▷ Estimate CATE in two ways

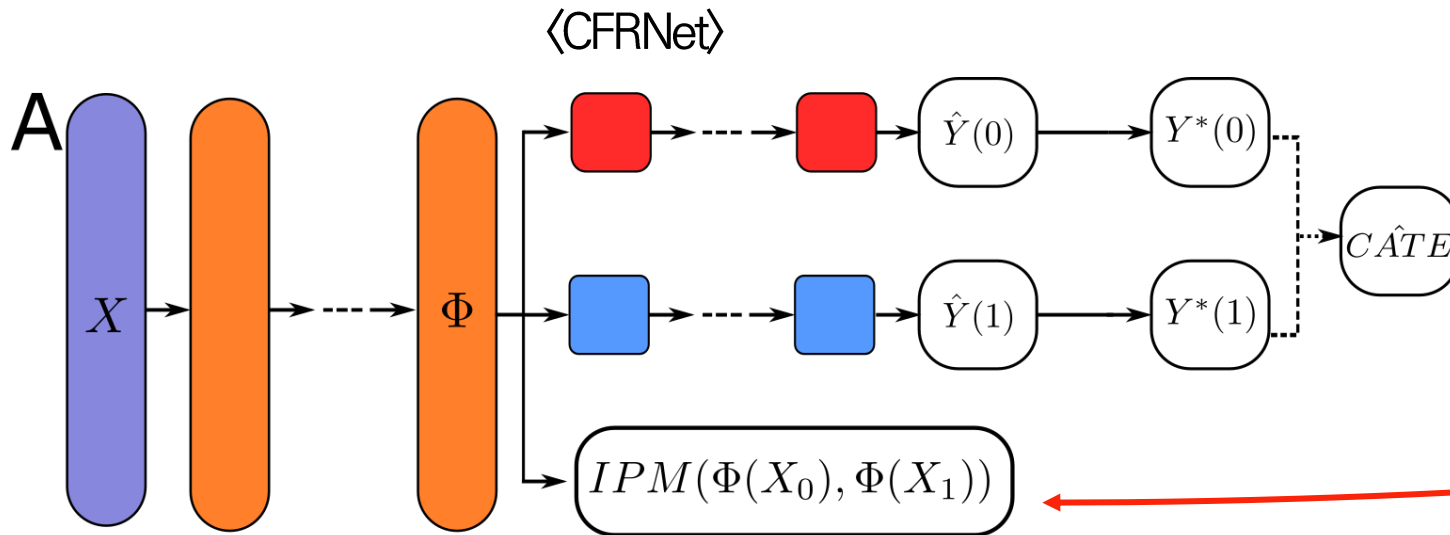
▷ Average the estimates



인과관계 분석 방법3

Potential Outcome (잠재적 결과): Neural Net Based

- Estimating individual treatment effect: generalization bounds and algorithms (PMLR 2017, 690회 인용)
- TARNet: Treatment-Agnostic Representation Network (기본 모델)
- 2개의 Head가 representation layer를 공유하며 학습되고 control/treatment 집단의 결과를 예측함
- CFRNet는 TARNet에 Regularization (IPM) 을 추가하여 Treatment Group과 Control Group의 차이를 최소화



We add an IPM as a loss in TARNET to more explicitly encourage the representation layers to balance the covariate distribution of the treated group, and the covariate distribution of the control group

인과관계 분석 도구

Potential Outcome (잠재적 결과): Neural Net Based

- CausalML(Uber): <https://github.com/uber/causalml>
- EconML(Microsoft): <https://github.com/py-why/EconML>



- Tree-based algorithms
 - Uplift tree/random forests on KL divergence, Euclidean Distance, and Chi-Square [2]
 - Uplift tree/random forests on Contextual Treatment Selection [3]
 - Causal Tree [4] - Work-in-progress
- Meta-learner algorithms
 - S-learner [5]
 - T-learner [5]
 - X-learner [5]
 - R-learner [6]
 - Doubly Robust (DR) learner [7]
 - TMLE learner [8]
- Instrumental variables algorithms
 - 2-Stage Least Squares (2SLS)
 - Doubly Robust (DR) IV [9]
- Neural-network-based algorithms
 - CEVAE [10]
 - DragonNet [11] - with `causalml[tf]` installation (see [Installation](#))



Estimation Methods

- ▶ Double Machine Learning (aka Rlearner) (click to expand)
- ▶ Dynamic Double Machine Learning (click to expand)
- ▶ Causal Forests (click to expand)
- ▶ Orthogonal Random Forests (click to expand)
- ▶ Meta-Learners (click to expand)
- ▶ Doubly Robust Learners (click to expand)
- ▶ Double Machine Learning with Instrumental Variables (click to expand)
- ▶ Doubly Robust Machine Learning with Instrumental Variables (click to expand)
- ▶ Deep Instrumental Variables (click to expand)

See the [References](#) section for more details.

Interpretability

- ▶ Tree Interpreter of the CATE model (click to expand)
- ▶ Policy Interpreter of the CATE model (click to expand)
- ▶ SHAP values for the CATE model (click to expand)

Conclusion

- AI 모델은 데이터의 패턴을 파악하여 상관관계를 학습하는 task (예측, 분류)에는 뛰어난 성능을 보여줌
- 하지만 왜 그런 결과가 나왔는지 원인을 밝히는 데에는 많은 아직 한계가 존재함
- 이를 위해서는 상관 관계가 아닌 인과 관계를 추론할 줄 알아야 함
- 최근에는 ML/DL을 활용하여 인과를 추론하는 연구가 활발하게 진행되고 있음

Thank you

본 세미나 내용에 대한 문의사항이 있으시면
아래의 이메일 주소로 연락주시길 바랍니다.

E-mail: charlie17@korea.ac.kr

Reference

- Künzel, Sören R., et al. "Metalearners for estimating heterogeneous treatment effects using machine learning." Proceedings of the national academy of sciences 116.10 (2019): 4156–4165.
- Shalit, Uri, Fredrik D. Johansson, and David Sontag. "Estimating individual treatment effect: generalization bounds and algorithms." International Conference on Machine Learning. PMLR, 2017.
- Shi, Claudia, David Blei, and Victor Veitch. "Adapting neural networks for the estimation of treatment effects." Advances in neural information processing systems 32 (2019).
- Be Careful When Interpreting Predictive Models in Search of Causal Insights, Towards Data Science
- Netflix Tech Blog (<https://netflixtechblog.com/what-is-an-a-b-test-b08cc1b57962>)
- 인과추론 기술과 마케팅 활용 (<https://www.youtube.com/watch?v=VwECmgje5A>)
- Causal Inference : Primer (https://www.slideshare.net/lumiamitie/causal-inference-primer-20190601?from_action=save)
- Korea Summer Session on Causal Inference 2021 (<https://sites.google.com/view/causal-inference2021>)
- 프로젝트 데이터 분석 커뮤니티 (<https://playinpap.github.io/mab/>)
- Be Geeky Blog (<https://assaeunji.github.io/bayesian/2020-04-10-causal/>)
- ZZAEBOK's BLOG (https://zzaebok.github.io/ab_test/ab-testing/)