

# 민감도 향상을 위한 트리거링

가짜연구소 인과추론팀  
온라인 통제 실험 연구자로 거듭나기

2024-05-14

황의림



Causal-Lab

## In Chapter 20...

방아쇠를 당기기 전에 목표물을 확실하게 식별하라  
- 톰 플린

### 트리거링이란

- 실험의 영향을 받지 않는 사용자가 생성한 노이즈를 필터링해서 통계적 검정력(민감도)를 개선하는 방법

조직의 실험 성숙도가 향상될수록, 더 많은 트리거 실험이 실행된다.

# *In Chapter 20...*

- *Examples of Triggering*
  - 의도적인 부분 노출
  - 조건부 노출
  - 적용 범위 증가
  - 머신러닝 모델에 대한 반사실 트리거링
  - 수치 예시
- *Trustworthy Triggering*
  - 최적 및 보수적 트리거링
  - 일반적인 함정들
  - 함정을 고려하여 실험효과 계산하기
- *Open Questions*

# Examples of Triggering

---

- 의도적인 부분 노출

특정 국가 사용자만을 대상으로 하는 **A/B** 테스트를 수행할 때,  
다른 국가의 사용자가 추가되면 노이즈 발생 + 통계적 검정력 감소

이미 수행 후 추가된 것을 알았을 때는 모든 사용자를 포함한 분석을 실시해야 함

# Examples of Triggering

- 조건부 노출

: 특정 페이지(결제)에 진입한 사용자만 트리거하기

- 결제 변경: 결제를 개시한 사용자만 트리거하기
- 공동 작업 변경: 공동 작업에 참여하는 사용자만 트리거하기
- 구독 취소 화면 변경: 이러한 변경이 표시된 사용자만 트리거하기
- 검색엔진 결과 페이지에 특정 답변(eg. 날씨)이 표시되는 방식 변경: 관련 검색어를 날린 사용자만 트리거하기

## < 예약 요청

혼치 않은 기회입니다.  
Terence님의 숙소는 보통 예약이 가득 차 있습니다.

예약 정보

날짜  
7월 7일~12일

수정

게스트  
게스트 1명

수정

결제 방식 선택하기

지금 ₩854,918 결제

○

요금 일부는 지금 결제, 나머지는 나중에 결제  
오늘은 ₩370,000, 2024년 6월 22일에는 ₩484,918의 금액을 결제하세요.  
추가 수수료는 없습니다. [상세 정보](#)

○

예약하려면 로그인 또는 회원 가입하세요

국가/지역  
한국 (+82)

전화번호

전화나 문자로 전화번호를 확인하겠습니다. 일반 문자 메시지 요금 및 데이터 요금이 부과됩니다. [개인정보 처리방침](#)

계속

또는


N

f

G

A

이메일로 로그인하기

힐링과 방갈로, 개울이 있는 자연의 집 - 양평 4  
개방  
종나무진 전제  
★ 4.95 (후기 475개) · ▼ 슈퍼호스트

요금 세부정보

₩148,000 x 5박

₩740,000

에어비엔비 서비스 수수료

₩114,918

총 합계 (KRW)

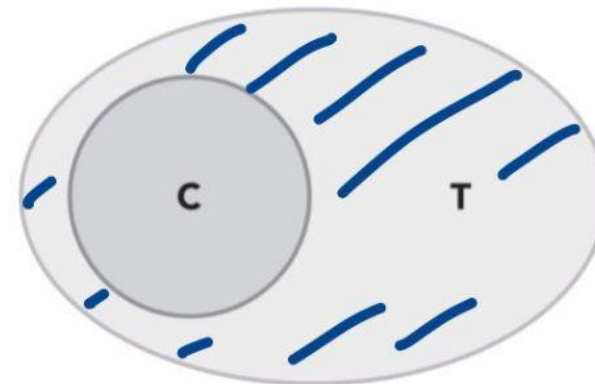
₩854,918

해외에서 결제가 처리되기 때문에 카드 발행사에서 추가 수수료를 부과할 수 있습니다.

# Examples of Triggering

- 적용범위 증가

실험군(T)를 “장바구니에 \$25이상 있어 무료배송 제의를 받는 유저”라고 하면,  
변경의 대상이 되는 \$25이상 \$35 미만 구매자에게  
변경사항이 적용되므로, 해당 유저에게만 트리거

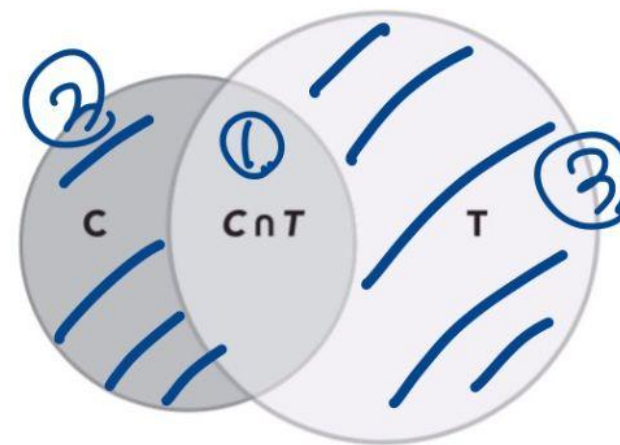


- 적용범위 변경

$T \cap C$ : 60일 이내 반품이력이 없고, 장바구니에 \$35 이상 있어 무료배송 제의를 받는 유저

$T \cap C^c$ : 60일 이내 반품이력이 있지만, 장바구니에 \$35이상 있어 무료 배송을 제의를 받는 유저

$T \cap C^c$ : 60일 이내 반품이력이 없고, 장바구니에 \$25이상 \$35미만이 있어 무료 배송 제의를 받는 유저



- 머신러닝 모델에 대한 반사실 트리거링(= 만약 그 기능이 도입되지 않았다면 어떤 일이 일어났을까?)
  - 대조군/실험군에서 구모델/신모델 모두 실행하여 결과를 저장
  - 각 집단에서 모델간 차이가 존재하는 유저들만 트리거

※ 각 집단에서 두 개의 모델이 모두 실행되어야 하므로 계산 비용이 증가함.

※ 두 모델을 병렬로 실행하지 않는 경우, 지연시간도 영향 받을 수 있음.

※ 모델 실행의 차이 등을 확인할 수 없음.

# Examples of Triggering

표준편차  $\sigma$ , 실험을 통해 변화를 보고자하는 크기(민감도 수준)  $\Delta$ ,  
95%의 신뢰수준, 80%의 검정력을 위한 최소 표본 크기를 구하려면,

$$n = \frac{16\sigma^2}{\Delta^2}$$

eg. 실험 기간 중 방문한 유저 중 5%가 실제 구매하는 사이트가 존재할 때,  
 $p = 0.05$ , 베르누이 시행으로  $\sigma^2 = 0.05*(1-0.05) = 0.0475$ , 구매 전환으로 5%의 변화를 보고자 할 때,  $\Delta = 0.05$   
즉, 최소  $16*0.0475/(0.05*0.05)^2 = \mathbf{121,600}$ 명의 사용자가 필요한 반면,

결제 페이지에 진입한 유저만 트리거한다고 하면,  
방문 유저 중 10%가 결제 페이지에 진입, 5%가 구매한다면  $p = 0.5$ 로  $\sigma^2 = 0.5*(1-0.5) = 0.25$   
즉,  $16*0.25/(0.5*0.05)^2 = \mathbf{64,000}$ 명



- 최적의 트리거 조건: 변수의 적용으로 변화가 생기는 유저만 분석
- 보수적인 트리거링

- 다중 실험군

대조군	실험군1	실험군1
1	0	0
0	1	0
0	0	1

- 사후 분석

결제 중 사용자 추천 모델이 반사실을 제대로 기록 못했다고 할 때,  
예시에서 두 모델간 차이가 있는 유저만을 트리거하기 어려운 상황이 된다.  
따라서 **“결제를 시작한 유저”**를 트리거 하는 게 효과적이다.

# Trustworthy Triggering - 꼭 수행해야 하는 두 가지 검사

- 샘플 비율 불일치(SRM)

- 전체 실험에 SRM이 없지만, 트리거된 분석이 SRM이 있는 경우 편향이 존재하는 것으로 판단됨.

- 보완 분석

- 트리거되지 않은 사용자에게 대해 AA 테스트를 수행
- 통계적으로 유의한 경우, 트리거 조건이 올바르지 않아 대상이 되지 않는 유저에게 영향을 미친다고 판단됨.

# Trustworthy Triggering

- 함정 1. 일반화하기 어려운 작은 세그먼트 실험
  - $n = \frac{\Delta_{\theta}}{M_{\omega C}} * \tau$ , ( $\tau$  = 전체 모집단 대비 트리거된 사용자 비율)
- 함정 2. 트리거된 사용자가 남은 실험기간 동안 제대로 트리거 되지 않음
- 반사실 기록의 성능 영향
  - A/A'/B 실험 수행
    - A, A'가 크게 다를 경우, 대조군에서 새로운 모델 관련 로그가 영향을 받고 있다는 경고가 났음.

# Trustworthy Triggering - 실험효과 계산하기

---

예시. 사용자 10%에 대해 수익이 3% 증가한 상황

1. 결제 프로세스가 변경되었고, 트리거된 사용자가 결제를 시작한 이용자인 경우
  - a. 수익이 결제로만 발생된다면 트리거된 수익 = 전체 수익 = 3%

# Trustworthy Triggering - 실험효과 계산하기



예시. 사용자 10%에 대해 수익이 3% 증가한 상황

## 2. 수익이 증가한 사용자가 평균 유저 매출의 하위 10%일 때

$w$ : 전체 사용자,  $\theta$ : 트리거된 사용자

$M_{wC}$ : 트리거되지 않은 대조군에 대한 지표 값

$N_{wC}$ : 트리거되지 않은 대조군 유저 수

$M_{wT}$ : 트리거되지 않은 실험군에 대한 지표 값

$N_{wT}$ : 트리거되지 않은 실험군 유저 수

$M_{\theta C}$ : 트리거된 대조군에 대한 지표 값  
 $= 0.1 * 0.1 * 0.03 = 0.0003 = 0.03\%$ 의 증가  
 $M_{\theta T}$ : 트리거된 실험군에 대한 지표 값

$N_{\theta C}$ : 트리거된 대조군 유저 수

$N_{\theta T}$ : 트리거된 실험군 유저 수

# Trustworthy Triggering - 실험효과 계산하기

예시. 사용자 10%에 대해 수익이 3% 증가한 상황

2. 수익이 증가한 사용자가 평균 유저 매출의 하위 10%일 때

$$\Delta_{\theta} = M_{\theta T} - M_{\theta C}$$

$$\delta_{\theta} = \Delta_{\theta} / M_{\theta C} = \text{트리거된 모집단에 대한 상대적 효과}$$

$$\tau = N_{\theta C} / N_{\omega C} = \text{트리거된 대조군 유저의 비율}$$

$$\frac{\Delta_{\theta} * N_{\theta C}}{M_{\omega C} * N_{\omega C}} = \frac{\Delta_{\theta}}{M_{\omega C}} * \tau = \delta_{\theta} * \tau$$

## 1. 트리거링 단위

- 전체 세션, 하루 종일, 혹은 실험 시작부터 어느 시점에 진행하는 것이 나을까?

## 2. 시간에 따른 지표 시각화

- 첫날 트리거 된 이후 접속하지 않는 사용자가 있다고 할 때, 어떻게 효과를 측정해야 할까
- 이런 문제를 야기하지 않기 위해서 어떻게 설계하면 좋을까

EOD.  
감사합니다.