NeurIPS 2023

Estimating Propensity for Causality-based Recommendation without Exposure Data

Zhongzhou Liu, Yuan Fang, Min Wu

2025.07.15

서정호

DGIST DATA SCIENCE LAB

DGIST 대구경북과학기술원 Daegu Gyeongbuk Institute of Science & Technology

Contents

Abstract & 1. Introduction

a. Correlation-based RS

1. Classical paradigm

b. Causality-based RS

2. Causality-based RS w/ observed [Exposure Data] or [Propensity Scores]

3. Causality-based RS w/ observed [Exposure Data] but w/o [Propensity Scores]

4. Causality-based RS w/o observed [Exposure Data] and [Propensity Scores]

Year

Abstract & 1. Introduction

b. Causality-based RS

- Cause: Exposure/Recommendation to item
- Result: User-Item Interaction to item [Causal Effect]
  - Click
  - Purchase

Abstract & 1. Introduction

RS의 4가지 paradigm

1. Classical paradigm
   - [Limitation] Ignore the Causal Impact behind recommendation
2. Causality-based RS w/ observed [Exposure Data] or [Propensity Scores]
   - [Problem] Exposure Data를 얻을 수 없음.
3. Causality-based RS w/ observed [Exposure Data] but w/o [Propensity Scores]
   - Estimate propensity scores
   - Limitations
     1. 대부분의 SOTA[state-of-the-art] methods: Propensity estimator을 훈련시키기 위해 Exposure Data가 필요.
     2. Prior knowledge를 propensity estimator에 통합 실패 → Not robust estimation
4. Causality-based RS w/o observed [Exposure Data] and [Propensity Scores]
   - **Only interaction data** 사용.
   - (propensity, item popularity) pairwise relationship을 prior knowledge로 통합
     → more Robust Propensity Estimation
   - model에 영향을 주는 factors에 대한 분석 제공.

Interaction Probability → Recommend

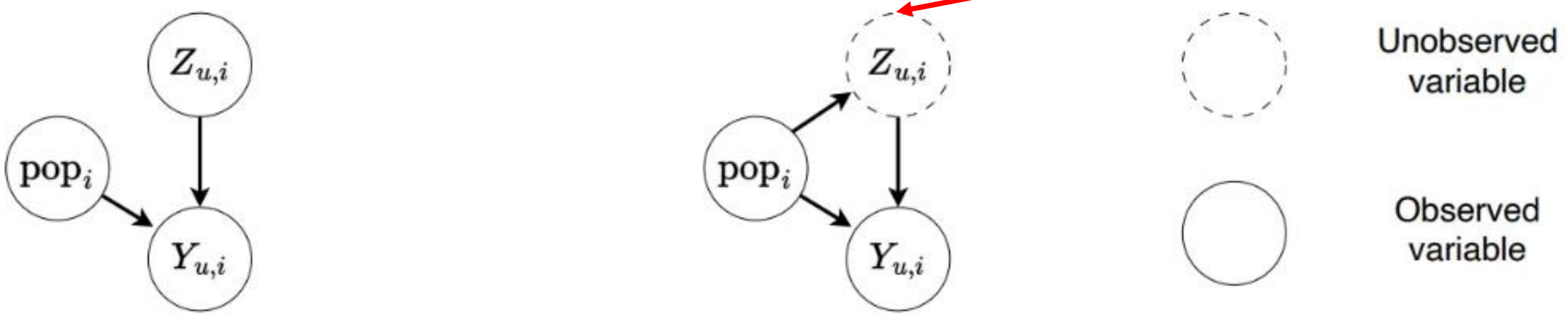
technical or privacy constraints

This Paper's Model: PROPCARE

Abstract & 1. Introduction

Estimation 후 추정치 사용.

Previous Propensity Estimation Vs PROPCARE



(a) Previous framework where exposure is observable

(b) Our framework PropCare where exposure is unobservable and popularity is incorporated as prior

Figure 1: Causal diagrams under different frameworks. pop_i is the popularity (prior) of item i. $Y_{u,i}$ indicates if user u interacts with item i. $Z_{u,i}$ indicates if item i is exposed to user u.

- PROPCARE 장점.
  1. Propensity or Exposure Data가 전혀 필요하지 않음.
  2. Prior Information을 Robust Estimation에 포함시켜 사용.
- Paper Contributions
  1. Previous Causality-based RS: propensity score and/or exposure data are often unavailable but required for model training or inference.
     → 이러한 문제 해결.
  2. (propensity, item popularity) pairwise relationship을 prior knowledge로 통합 → more Robust Propensity Estimation
  3. model에 영향을 주는 factors에 대한 분석 제공.
  4. PROPCARE의 효과를 quantitative and qualitative 결과를 통해 검증.

2. Related Work

Causal effect estimation in recommendation

- Typical RS: positive feedback or interactions(clicks, purchases)가 성공적인지를 고려
  - → Recommendations로 인한 causal effect를 최적화하는 것이 더 가치있다.
- Causal Effect를 구하는 것의 challenges
  - Online A/B tests: exposure strategies를 비교할 수 있음. But 비싸고 selection bias에 취약.
  - → 해결책: Causal Effect Estimators.

1. Naive Estimator
2. Inverse Propensity Score (IPS) estimator
   - propensity score := probability of exposure
   - Non-parametric approach
3. CausCF
   - Parametric models 사용해서 outcomes prediction.
4. doubly robust estimator
   - parametric model + non-parametric IPS estimator → bias and variance 감소.

→ **문제점: training stage 단계에서 propensity scores and/or exposure data가 필요함.**

2. Related Work

Propensity estimation in recommendation

- Existing Causal Effect Estimation approaches
  - **문제점 (1): training stage 단계에서 propensity scores and/or exposure data가 필요함.**
  - **문제점 (2): Missing-Not-At-Random (MNAR) 문제가 발생.** KDD'10 Training and testing of recommender systems on data missing not at random
- Some methods: Estimate propensity in a **heuristic way [경험적인 방법]**
  1. Using item popularity
  2. Using other side Information. e.g., items participating in promotional campaigns

  → **문제점: Personalization 부족 & Noisy results.**
- Interaction models를 이용하는 approaches == click models
  - propensity scores, relevance, interactions과 관계있음.
  - **문제점: 추가적인 constraints 없이는 interaction model 단독으로는 optimize하기 어려움. [Sect. 4.1.]**
- Matrix factorization, Linear regression, Dual learning, Double robust learning: propensity scores 학습 가능.
  - Exposure data를 training labels or known variables로 가정
  - **→ 문제점: Exposure data가 필요**

2. Related Work

Propensity estimation in recommendation

- Existing Causal Effect Estimation approaches
  - **문제점 (1): training stage 단계에서 propensity scores and/or exposure data가 필요함.**
  - 문제점 (2): Missing-Not-At-Random (MNAR) 문제가 발생. KDD'10 Training and testing of recommender systems on data missing not at random
- Some methods: Estimate propensity in a **heuristic way [경험적인 방법]**
  1. Using item popularity
  2. Using other side Information. e.g., items participating in promotional campaigns

  → **문제점: Personalization 부족 & Noisy results.**
- Interaction models를 이용하는 approaches == click models
  - propensity scores, relevance, interactions과 관계있음.
  - **문제점: 추가적인 constraints 없이는 interaction model 단독으로는 optimize하기 어려움.** [Sect. 4.1.]
- Matrix factorization, Linear regression, Dual learning, Double robust learning: propensity scores 학습 가능.
  - Exposure data를 training labels or known variables로 가정
  - **→ 문제점: Exposure data가 필요**

DGIST 대구경북과학기술원 Daegu Gyeongbuk Institute of Science & Technology

2. Related Work

KDD'10 Training and testing of recommender systems on data missing not at random

MSAR Problem

Table 1: Simplistic Example for ratings missing not at random (MNAR): test data where users rated only what they liked or knew.

		users							
		horror fans				romance lovers			
	h	5		5	5				
m	o	5	5						
o	r		5		5				
v	.	5		5	5				
i	r					5	5		5
e	o						5	5	5
s	m					5		5	
	.						5	5	5

- Romance movies를 좋아하는 사람은 romance movies에만 최고 점수 5점을 주고, horror movies는 평가하지 않음.(누락)
  1. Missing (누락): user가 각자 좋아하거나 알고 있는 movies만 평가
  2. test sets에 MNAR 문제가 발생.
  3. 항상 user - item(movie)의 pair가 본인이 좋아하는 영화만 pair로 생성.
  4. (user, item)이 항상 최고 점수 5점만 발생, 항상 5점을 예측하는 쓸모 없는 추천 시스템 탄생
- [해결책] (user, item)에서 ratings(평가)이 observed or missing item 모두에 대하여 prediction/ranking을 해야 함.

3. Preliminaries

Data notations

- Typical recommendation dataset
  - $D = \{(Y_{u,i})\}$: collection of observed **training** user-item **interaction data**
- Interactions between users-items.: Purchases or Clicks [Result]
  - $Y_{u,i} \in \{0, 1\}$: observed interaction
    - $Y_{u,i} = 1$: user u interacted to item i
  - User $u \in \{1, 2, 3, ..., U\}$
  - Item $i \in \{1, 2, 3, ..., I\}$
- Unobservable indicator variable for Exposure [Cause]
  - $Z_{u,i} \in \{0, 1\}$
    - $Z_{u,i} = 1$: item i is exposed/recommended to user u
- Propensity score := probability of exposure
  - $p_{u,i} := P\left(Z_{u,i} = 1\right)$

3. Preliminaries

Causal effect modeling

- Potential outcomes for different exposure statuses: $Z_{u,i} = 0 \ or \ 1$
  - 윗점자: Exposure 여부. [Cause]
  - 값: Interaction 여부. [Result]
  - $Y^0_{u,i} \in \{0,1\}$: **Exposure** ❌일 때의 interaction
  - $Y^1_{u,i} \in \{0,1\}$: **Exposure** ⭕일 때의 interaction
  - **문제점: real-world에서는 특정 (u,i)에 대하여 $Y^0_{u,i}$ 또는 $Y^1_{u,i}$만 관측할 수 있음. [Counterfactual Nature]**
- Counterfactual Model
  - Causal Effect: $\tau_{u,i} := Y^1_{u,i} - Y^0_{u,i} \in \{-1,0,1\}$ means **Exposure(Recommendation) → Interaction Relationship**
    - $\tau_{u,i} = 1$: **recommending item i to user u** ⇒ user-item interaction 증가.
      - 3가지 주체(Users, Sellers, Platforms) 모두 positive causal effects를 결과로 하는 recommendation으로부터 이득을 얻음.

      💡 **RS의 주체: User, Seller, Platform** 3가지
    - $\tau_{u,i} = -1$: **recommending item i to user u** ⇒ user-item interaction 감소.
    - $\tau_{u,i} = 0$: **recommending or not** ⇒ user-item interaction에 영향을 끼칠 수 없음.

DGIST 대구경북과학기술원 Daegu Gyeongbuk Institute of Science & Technology

DGIST DATA SCIENCE LAB

3. Preliminaries

Causal effect estimation

- Causal effect $\tau_{u,i}$: [Counterfactual Nature] 때문에 observed data로부터 direct하게 계산 불가능. ⇒ Estimation 필요.
  - $Y_{u,i}^1, Y_{u,i}^0$을 동시에 observation할 수 없다.
- CausCF or Doubly robust estimator
  - direct parametric models라서 potential outcomes의 prediction error에 민감
    - → 고품질 labeled Exposure Data 필요. (for parametric models)
    - → this paper의 setup이 아님.
- **This paper**: IPS estimator ∈ Non-parametric approach 사용.
  - [**Appendix B**]

$$\hat{Y}_{u,i}^1 = \frac{Z_{u,i}Y_{u,i}}{p_{u,i}}, \quad \hat{Y}_{u,i}^0 = \frac{(1-Z_{u,i})\,Y_{u,i}}{1-p_{u,i}} : \text{Unbiased (IPS) Estimator}$$

$$\text{Since} \quad Y_{u,i} = Z_{u,i}Y_{u,i}^1 + (1-Z_{u,i})\,Y_{u,i}^0 \quad (a.10)$$

$$\mathbb{E}[Z_{u,i}] = 1 \cdot p_{u,i} + 0 \cdot (1-p_{u,i}) = p_{u,i}$$

$$\hat{\tau_{u,i}}^{IPS} = \hat{Y}_{u,i}^1 - \hat{Y}_{u,i}^0 = \frac{Z_{u,i}Y_{u,i}}{p_{u,i}} - \frac{(1-Z_{u,i})\,Y_{u,i}}{1-p_{u,i}} : \text{Also Unbiased Estimator} \quad (1)$$

3. Preliminaries

Causal effect estimation

- Causal effect $\tau_{u,i}$: [Counterfactual Nature] 때문에 observed data로부터 direct하게 계산 불가능. ⇒ Estimation 필요.
  - $Y_{u,i}^{1}, Y_{u,i}^{0}$을 동시에 observation할 수 없다.
- CausCF or Doubly robust estimator
  - direct parametric models라서 potential outcomes의 prediction error에 민감
    - → 고품질 labeled Exposure Data 필요. (for parametric models)
    - → this paper의 setup이 아님.
- **This paper**: IPS estimator ∈ Non-parametric approach 사용.
  - [**Appendix B**]

$$\hat{Y}_{u,i}^{1} = \frac{Z_{u,i}Y_{u,i}}{p_{u,i}}, \quad \hat{Y}_{u,i}^{0} = \frac{(1-Z_{u,i})\,Y_{u,i}}{1-p_{u,i}}: \text{Unbiased (IPS) Estimator}$$

$$\text{Since} \quad Y_{u,i} = Z_{u,i}Y_{u,i}^{1} + (1-Z_{u,i})\,Y_{u,i}^{0} \quad (a.10)$$

$$\mathbb{E}\left[Z_{u,i}\right] = 1 \cdot p_{u,i} + 0 \cdot (1-p_{u,i}) = p_{u,i}$$

$$\hat{\tau_{u,i}}^{IPS} = \hat{Y}_{u,i}^{1} - \hat{Y}_{u,i}^{0} = \frac{Z_{u,i}Y_{u,i}}{p_{u,i}} - \frac{(1-Z_{u,i})\,Y_{u,i}}{1-p_{u,i}}: \text{Also Unbiased Estimator} \quad (1)$$

3. Preliminaries

💡 IPS estimator에서 더 발전한 점이 무엇인가?

문제점: training stage 단계에서 propensity scores and/or exposure data가 필요함.

→ [해결]: 추정치 사용하여 대체하였음. [Section 4.5]

- $p_{u,i} \rightarrow \hat{p_{u,i}}$
- $Z_{u,i} \rightarrow \hat{Z_{u,i}}$

Interaction model

- Assumption: Relationship betw. interactions, propensity and relevance

$$y_{u,i} = p_{u,i} \times r_{u,i} \quad (2)$$

- $y_{u,i} = P(Y_{u,i} = 1)$: **Probability of interaction** betw. user u and item i [**Result: Interaction Prob.**]
- $p_{u,i} := P(Z_{u,i} = 1)$: Propensity score := **probability** of exposure [**Cause: Exposure Prob.**]
- $r_{u,i}$: **Probability** that **item i** is **relevant** to **user u** [Relevant Prob.]

💡 Assumption: Interaction Prob. = Exposure Prob. * Relevant Prob.

4. Proposed Approach: PROPCARE

4.1. Naive propensity estimator

- **Setup: only interaction data** are observable
- Objective: Estimate propensity scores and exposure
  1. Main focus: Estimation of propensity scores $\hat{p_{u,i}}$
  2. → Propensity로부터 corresponding Exposure은 쉽게 sampling 가능. using threshold. [Section 4.3]
- **Naive Loss Function** for the interaction model $y_{u,i} = p_{u,i} r_{u,i}$

$$\mathcal{L}_{\text{naive}} = -Y_{u,i} \times \log f_p(\mathbf{x}_{u,i}; \Theta_p) f_r(\mathbf{x}_{u,i}; \Theta_r) - (1 - Y_{u,i}) \times \log(1 - f_p(\mathbf{x}_{u,i}; \Theta_p) f_r(\mathbf{x}_{u,i}; \Theta_r)) \quad (3)$$

- Interaction model: $y_{u,i} = p_{u,i} r_{u,i}$
  - $\hat{p_{u,i}} = f_p\left(\mathbf{x_{u,i}}; \Theta_p\right)$
  - $\hat{r_{u,i}} = f_r\left(\mathbf{x_{u,i}}; \Theta_r\right)$
  - $\Rightarrow \hat{y_{u,i}} = \hat{p_{u,i}} \times \hat{r_{u,i}}$
  - $\mathcal{L}_{\text{naive}} = -Y_{u,i} \times \log \hat{y_{u,i}} - (1 - Y_{u,i}) \times \log(1 - \hat{y_{u,i}})$
    : **Binary Cross-Entropy Loss ftn.** [label*log (prob.)] ⇒ **Point-wise manner**

4. Proposed Approach: PROPCARE

4.1. Naive propensity estimator

- $\mathbf{x}_{u,i} = f_e\left(u, i; \Theta_e\right)$: Joint user-item embedding output
  - f_e: learnable embedding function
  - $\mathbf{x}_{u,i}$: f_p, f_r의 Input으로 사용.
- f_p: learnable propensity function → Estimated propensity score $\hat{p_{u,i}}$
- f_r: learnable relevance function → Estimated relevance probability $\hat{r_{u,i}}$

- Exposure Prob., Relevant Prob. 모두 구할 수 없음 → Estimation 해서 사용. with learnable function
- [MLP] 구조
  - [Input] (u,i) → embedding vector 변환 → MLP로 $\mathbf{x}_{u,i} = f_e\left(u, i; \Theta_e\right)$ [Output] 출력.
    Θ_e: Weights of MLP ftn. f_e
  - [Input] $\mathbf{x}_{u,i}$ → MLP로 $\hat{p_{u,i}} = f_p\left(\mathbf{x_{u,i}}; \Theta_p\right), \hat{r_{u,i}} = f_r\left(\mathbf{x_{u,i}}; \Theta_r\right)$ [Output] 출력.
    Θ_*: Weights of MLP ftn. f_*
- MLP: [Input] (u,i) → $\hat{p_{u,i}} = f_p\left(\mathbf{x_{u,i}}; \Theta_p\right), \hat{r_{u,i}} = f_r\left(\mathbf{x_{u,i}}; \Theta_r\right)$ [Output]
  - Exposure Prob., Relevant Prob. Estimation.

4. Proposed Approach: PROPCARE

4.1. Naive propensity estimator

- Each learnable function f_*: parameter을 가짐 → Θ_*: parameter set of f_* → MLP로 parameter 학습됨.
- 문제점: $\hat{y_{u,i}} = f_p\left(\mathbf{x_{u,i}}; \Theta_p\right) \times f_r\left(\mathbf{x_{u,i}}; \Theta_r\right)$: f_p, f_r이 곱해진 form
  - → $\mathcal{L}_{\text{naive}}$로부터 f_p or f_r을 학습시킬 수 없음. $\hat{y_{u,i}} = \hat{p_{u,i}} \times \hat{r_{u,i}}$가 학습됨.
  - → Propensity Score 추정 불가능. $\hat{p_{u,i}} = f_p\left(\mathbf{x_{u,i}}; \Theta_p\right)$을 구할 수 없음.

4. Proposed Approach: PROPCARE

4.2. Incorporating prior knowledge

- Naive Loss Function의 해결법: Prior Knowledge로 f_p or f_r을 constrain 하는 것.
- *Observation: more popular items will have a higher chance to be exposed* [Popularity → Exposure Prob.]
  - pop_i: Popularity of item i

$$pop_i := \frac{\sum_{u=1}^{U} Y_{u,i}}{\sum_{j=i}^{I} \sum_{u=1}^{U} Y_{u,j}}$$

    - 모든 (user, item) pairs 의 interactions #중, item i에 대하여 얼마나 많은 user가 interaction을 하였는가?
    - **Popularity of item i ⇒ 전체 interaction 중 item i에 대한 비중.**
  - 직관적이지만 popularity-exposure 관계를 설명하기에 적절하지 못함.
  - → **반례**: 높은 interaction 확률을 가진 item ⇒ 높은 노출 chance를 가지는 경향이 있음.
    - [Interaction Prob. → Exposure Prob.]
    - [문제점] Popularity, Interaction Prob. → Exposure Prob.
    - Causal Effect: Exposure → Interaction
      - Thus, Popularity factor을 소거해야 함.
  - **[해결책]** Popularity를 propensity/exposure estimation에 대한 Prior Knowledge로 통합시킴.
    - → Interaction Prob. control. → **Assumption 1 (Pairwise Relationship on Popularity and Propensity)**

4. Proposed Approach: PROPCARE

4.2. Incorporating prior knowledge

Assumption 1 (Pairwise Relationship on Popularity and Propensity) == Prior Knowledge

1. Popularity
2. Interaction Prob. → fixed. [control]
3. Exposure Prob. = Propensity

Popularity → Exposure Prob. 관계만을 보기 위함.

∴ Prior Knowledge: 인기가 많은 item이 노출이 많이 된다. [Propensity 크다]

- user: u
- pair of items: (i, j)
- $pop_i > pop_j,\ y_{u,i} \approx y_{u,j} \ \Rightarrow\ p_{u,i} > p_{u,j}$
  - **정확히는,** $y_{u,i} \approx y_{u,j} \ \Rightarrow\ (pop_i > pop_j \Leftrightarrow\ p_{u,i} > p_{u,j})$
    : Popularity, Propensity의 대소 관계가 같다. if Interaction Prob. fixed.

4. Proposed Approach: PROPCARE

4.2. Incorporating prior knowledge

▼ Empirical validation of Assumption 1

Assumption 1을 만족시키는 3가지 dataset(DH_original, DH_personalized and ML)에서의 실험적 검증. by calculating

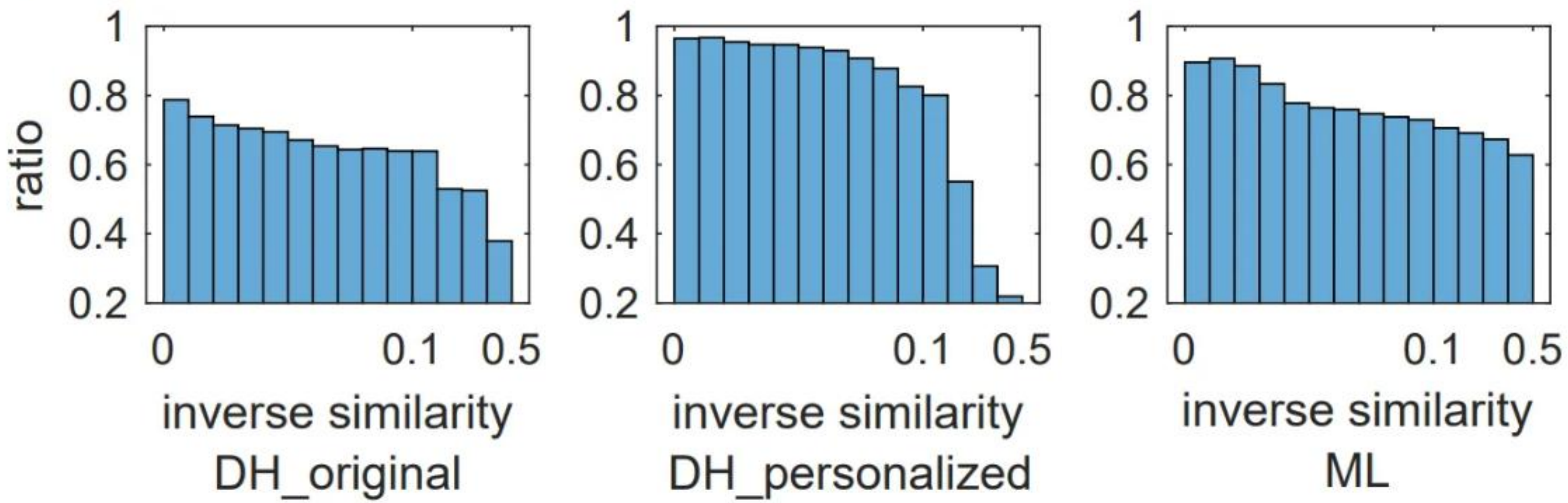


Figure 2

- x-axis: bins [intervals]: **Inverse Similarity** in **Interaction Prob.** $|y_{u,j} - y_{u,i}|$
  - **ps. Inverse Similarity == Distance.**
- y-axis: ratio_b: 각 bin에 대하여 Assumption 1을 만족시키는 items의 확률

💡 $(|y_{u,j} - y_{u,i}| \downarrow \Rightarrow \ \mathrm{ratio}_b \uparrow) \ \text{means} \quad [y_{u,i} \approx y_{u,j} \ \Rightarrow \ (pop_i > pop_j \Leftrightarrow \ p_{u,i} > p_{u,j})]$

즉, Assumption 1이 실험적으로 검증 끝.

4. Proposed Approach: PROPCARE

4.2. Incorporating prior knowledge

ratio_b 계산.

user u에 대하여 **특정** bin b에 존재하는 item pairs 중 $pop_i > pop_j \Rightarrow p_{u,i} > p_{u,j}$을 만족시키는 pairs의 개수. [확률]

$$ratio_b = \frac{1}{U}\sum_{u=1}^{U} \frac{\#\text{ item pairs } (i,j) \text{ for user } u \text{ in bin } b \text{ s.t. } (p_{u,j} - p_{u,i})(\text{pop}_j - \text{pop}_i) > 0}{\#\text{ item pairs } (i,j) \text{ sampled for user } u \text{ in bin } b} \quad (4)$$

4. Proposed Approach: PROPCARE

4.2. Incorporating prior knowledge

Integrating prior knowledge

- $\mathcal{L}_{\text{naive}}$에서 $\hat{y_{u,i}} = \hat{p_{u,i}} \times \hat{r_{u,i}}$가 학습됨 → f_p or f_r을 **학습시킬 수 있도록,** 분리된 Loss function 사용.
- $pop_i > pop_j, y_{u,i} \approx y_{u,j}$일 때, $f_p\left(\mathbf{x_{u,i}}\right) > f_p\left(\mathbf{x_{u,j}}\right)$ 여야 함. From. Assumption 1 [Prior Knowledge]

$$\text{loss} = -\log\left[\sigma\left(f_p\left(\mathbf{x_{u,i}}\right) - f_p\left(\mathbf{x_{u,j}}\right)\right)\right] \in (0, \infty) \qquad (5)$$

  - $f_p\left(\mathbf{x_{u,i}}\right) - f_p\left(\mathbf{x_{u,j}}\right) \uparrow \Rightarrow \text{loss} \downarrow$: 따라서 loss를 minimize하려는 방향이 우리의 Prior Knowledge에 잘 부합한다.
  - σ: Sigmoid function.
  - loss: pair-wise 방식으로 popularity를 잘 사용했음.
- Above loss의 Advantages
  1. f_p, f_r 분리
  2. Interaction data으로 계산이 가능한 item popularity만 사용. ← $pop_i := \frac{\sum_{u=1}^{U} Y_{u,i}}{\sum_{j=i}^{I}\sum_{u=1}^{U} Y_{u,j}}$
  3. $pop_i \neq pop_j$이므로 예측값 $\hat{p_{u,i}} \approx 0 \text{ or } 1$ 방지. [Section 4.5의 Remark 2]
- Final Popularity-loss function: Popularity → Exposure Prob. $\hat{p_{u,i}}$ [Assumption 1]

$$\mathcal{L}_{\text{pop}} = -\kappa_{u,i,j}\log\left[\sigma(\text{sgn}_{i,j}\cdot(f_p(\mathbf{x}_{u,i}) - f_p(\mathbf{x}_{u,j}))) + \sigma(\text{sgn}_{i,j}\cdot(f_r(\mathbf{x}_{u,j}) - f_r(\mathbf{x}_{u,i})))\right] \qquad (6)$$

$$\arg\min_{\Theta_{\text{pop}}} \mathcal{L}_{\text{pop}} = \hat{\Theta_{\text{pop}}} \rightarrow \hat{p_{u,i}}, \hat{r_{u,i}}$$

4. Proposed Approach: PROPCARE

4.2. Incorporating prior knowledge

$$\text{loss} = -\log\left[\sigma\left(f_p\left(\mathbf{x_{u,i}}\right) - f_p\left(\mathbf{x_{u,j}}\right)\right)\right] \in (0, \infty) \quad (5)$$

$$\mathcal{L}_{\text{pop}} = -\kappa_{u,i,j} \log\left[\sigma(\text{sgn}_{i,j} \cdot (f_p(\mathbf{x}_{u,i}) - f_p(\mathbf{x}_{u,j}))) + \sigma(\text{sgn}_{i,j} \cdot (f_r(\mathbf{x}_{u,j}) - f_r(\mathbf{x}_{u,i})))\right] \quad (6)$$

$$\arg\min_{\Theta_{\text{pop}}} \mathcal{L}_{\text{pop}} = \hat{\Theta}_{\text{pop}} \rightarrow \hat{p_{u,i}}, \hat{r_{u,i}}$$

For a Fixed $y_{u,i}$,

- $\hat{p_{u,i}} = f_p\left(\mathbf{x_{u,i}}\right)$
- $\hat{r_{u,i}} = f_r\left(\mathbf{x_{u,i}}\right)$

$$\mathcal{L}_{\text{pop}} = -\kappa_{u,i,j} \times \log\left[\sigma(\text{sgn}_{i,j} \cdot (\hat{p_{u,i}} - \hat{p_{u,j}})) + \sigma(\text{sgn}_{i,j} \cdot (\hat{r_{u,j}} - \hat{r_{u,i}}))\right] \quad (6')$$

- Popularity → Exposure Prob. → Relevance Prob.: 위 loss function은 결국 popularity에 의존 [$\mathcal{L}_{\text{pop}}$]

1. $sgn_{i,j} = \text{sign}\left(pop_i - pop_j\right) \in \{1, -1\}$: $pop_i > pop_j, pop_i < pop_j$ 모두 고려.
2. $\kappa_{u,i,j} = e^{\eta(y_{u,i} - y_{u,j})^2}, \eta < 0$: weighting function. $|y_{u,i} - y_{u,j}| \downarrow \Rightarrow \kappa_{u,i,j} \uparrow$
   - η: learnable parameter
   - Assumption 1의 조건에 부합할수록 loss의 가중치를 크게 만듦.: $y_{u,i} \approx y_{u,j}$ 조건 고려.
3. Interaction model: $y_{u,i} = p_{u,i} \times r_{u,i}$ ⇒ for a fixed $y_{u,i}$, $p_{u,i} \uparrow \Rightarrow r_{u,j} \downarrow$
   - f_p 만 고려하지 않고, f_r까지 고려. → model training을 더 향상시킴.
   - 뒷 부분 relevance term이 j - i인 이유. $p_{u,i} \uparrow \Rightarrow r_{u,j} \downarrow$

4. Proposed Approach: PROPCARE

4.3. Propensity learning

1. $\mathcal{L}_{\text{naive}}$: $\hat{y_{u,i}} = \hat{p_{u,i}} \times \hat{r_{u,i}}$**가 학습됨** → Interaction model $y_{u,i} = p_{u,i} \times r_{u,i}$ **최적화**
2. $\mathcal{L}_{\text{pop}}$: $\hat{p_{u,i}}, \hat{r_{u,i}}$**가 학습됨**. [Pairwise-loss] → Popularity를 Propensity learning을 위한 Prior information으로 사용.

   → 통합시킨 total loss function. $\mathcal{L}_{\text{total}}$ 사용하자.
3. Propensity score $\hat{p_{u,i}}$ Regularization for [$\hat{p_{u,i}} \approx 0 \text{ or } 1$ 방지].

   → Regularization Term: $\mu \cdot \mathrm{KLD}\left(Q \| \mathrm{Beta}(\alpha, \beta)\right)$, Regularization parameter: μ

$$\mathcal{L}_{\text{total}} = \sum_{u,i,j} \left(\mathcal{L}_{\text{naive}} + \lambda \cdot \mathcal{L}_{\text{pop}}\right) + \mu \cdot \mathrm{KLD}\left(Q \| \mathrm{Beta}(\alpha, \beta)\right) \quad (7)$$

$$\arg\min_{\Theta_{\text{total}}} \mathcal{L}_{\text{total}} = \hat{\Theta}_{\text{total}} \rightarrow \hat{p_{u,i}}, \hat{r_{u,i}}$$

4. Proposed Approach: PROPCARE

$$\mathcal{L}_{\text{total}} = \sum_{u,i,j} (\mathcal{L}_{\text{naive}} + \lambda \cdot \mathcal{L}_{\text{pop}}) + \mu \cdot \text{KLD}\,(Q \| \text{Beta}(\alpha, \beta)) \quad (7)$$

$$\arg\min_{\Theta_{\text{total}}} \mathcal{L}_{\text{total}} = \hat{\Theta_{\text{total}}} \;\rightarrow\; \hat{p_{u,i}}, \hat{r_{u,i}}$$

4.3. Propensity learning

- 소수의 인기 아이템만 노출확률이 크기 때문에, propensity scores [**ground-truth**]는 **long-tailed distribution**.
  - → 마찬가지로 long-tailed인 **Beta distribution**을 propensity scores의 regularization에 사용.
  - Q: Empirical distribution of all estimated propensity scores $\hat{p_{u,i}}$
  - α, β: parameters which are selected to simulate a long-tailed shape.
  - $\text{KLD}\,(\cdot\|\cdot)$: **Kullback-Leibler Divergence betw. two distributions. → 작을수록 예측 분포와 실제 분포가 비슷함.**
- λ, μ: trade-off hyper-parameters [weighting term.]
  - λ : $\mathcal{L}_{\text{naive}}, \mathcal{L}_{\text{pop}}$ 조율.
  - μ: Regularization 조율.
- Estimated propensity score $\hat{p_{u,i}} \rightarrow \hat{Z_{u,i}}$ 예측.
  - $\hat{Z_{u,i}} = 1 \quad if \;\; \text{Norm}\,(\hat{p_{u,i}}) \geq \epsilon$
  - $\hat{Z_{u,i}} = 0 \quad \text{otherwise}$
  - ϵ: threshold hyper-parameter
  - Norm: Normalization function such as Z-score normalization

DGIST DATA SCIENCE LAB

DGIST 대구경북과학기술원 Daegu Gyeongbuk Institute of Science & Technology

4. Proposed Approach: PROPCARE

4.3. Propensity learning

- [Input] $D = \{(Y_{u,i})\} \rightarrow \mathcal{L}_{\text{total}} \rightarrow \hat{p_{u,i}}, \hat{r_{u,i}} \rightarrow \hat{Z_{u,i}}$ [Output]
- $\hat{r_{u,i}}$도 함께 구할 수 있지만, 실제 PROPCARE framework에는 사용되지 않음.

4. Proposed Approach: PROPCARE

4.4. Causality-based recommendation

- DLCE: Debiased Learning for the Causal Effect
  - SOTA [state-of-the-art] Causality-based Recommender w/ IPS estimator
  - Input: Interaction $Y_{u,i}$, Exposure $Z_{u,i}$, Propensity $p_{u,i}$
  - Output: Ranking Score $\hat{s_{u,i}}$ for each user-item (u,i) pair
    - User u를 위한 Recommendation Ranking을 정할 때 사용되는, each item의 Ranking Score.
  - Given $(u, i, j) \quad s.t.\ i \neq j$, the **DLCE loss function**

$$\mathcal{L}_{\text{DLCE}} = \frac{Z_{u,i} Y_{u,i}}{\max(p_{u,i}, \chi^1)} \log\left(1 + e^{-\omega(\hat{s_{u,i}} - \hat{s_{u,j}})}\right) + \frac{(1 - Z_{u,i}) Y_{u,i}}{\max(1 - p_{u,i}, \chi^0)} \log\left(1 + e^{\omega(\hat{s}_{u,i} - \hat{s}_{u,j})}\right) \quad (8)$$

$$\mathcal{L}_{\text{DLCE}} = \frac{Y_{u,i}}{\max(p_{u,i}, \chi^1)} \log\left(1 + e^{-\omega(\hat{s}_{u,i} - \hat{s}_{u,j})}\right) \times \mathbb{I}(Z_{u,i} = 1) + \frac{Y_{u,i}}{\max(1 - p_{u,i}, \chi^0)} \log\left(1 + e^{\omega(\hat{s}_{u,i} - \hat{s}_{u,j})}\right) \times \mathbb{I}(Z_{u,i} = 0) \quad (8)$$

$$\boxed{\hat{s_{u,i}} = f_s(u, i, \Theta_s), \quad \arg\min_{\Theta_s} \mathcal{L}_{\text{DLCE}} = \hat{\Theta}_s \rightarrow \hat{s_{u,i}}}$$

- χ^1, χ^0, ω: hyper-parameters

DGIST DATA SCIENCE LAB

4. Proposed Approach: PROPCARE

4.4. Causality-based recommendation

- [This paper] Ground truth 대신 추정치 사용.
  - $p_{u,i} \rightarrow \hat{p_{u,i}}$
  - $Z_{u,i} \rightarrow \hat{Z_{u,i}}$

$$\mathcal{L}_{\text{PC-DLCE}} = \frac{\hat{Z_{u,i}} Y_{u,i}}{\max(\hat{p_{u,i}}, \chi^1)} \log\left(1 + e^{-\omega(\hat{s_{u,i}} - \hat{s_{u,j}})}\right) + \frac{(1 - \hat{Z_{u,i}}) Y_{u,i}}{\max(1 - \hat{p_{u,i}}, \chi^0)} \log\left(1 + e^{\omega(\hat{s_{u,i}} - \hat{s_{u,j}})}\right) \qquad (8')$$

$$\hat{s_{u,i}} = f_s(u, i, \Theta_s), \quad \arg\min_{\Theta_s} \mathcal{L}_{\text{PC-DLCE}} = \hat{\Theta}_s \rightarrow \hat{s_{u,i}}$$

  - PC: PROPCARE

DLCE: [Input] $Y_{u,i}, \hat{p_{u,i}}, \hat{Z_{u,i}} \rightarrow \hat{s_{u,i}}$ [Output]

4. Proposed Approach: PROPCARE

4.4. Causality-based recommendation

PROPCARE Structure

- [Input] $D = \{(Y_{u,i})\} \rightarrow \hat{s_{u,i}}$ [Output]
- User u에 대하여, $\{\hat{s_{u,1}}, \hat{s_{u,2}}, \ldots, \hat{s_{u,I}}\}$를 sorting하여 가장 큰 상위 K개 선발.
  - → Top-K Recommendation Lists. [Final Output]

4. Proposed Approach: PROPCARE

4.5. Theoretical property

- 기존 Causal Effect의 IPS estimator: Unbiased Estimator.

$$\hat{\tau_{u,i}}^{IPS} = \frac{Z_{u,i}Y_{u,i}}{p_{u,i}} - \frac{(1 - Z_{u,i})\, Y_{u,i}}{1 - p_{u,i}} : \text{Unbiased Estimator} \quad (1)$$

- [This paper] Ground truth 대신 추정치 사용. → IPS estimator: Biased Estimator.
    - $p_{u,i} \rightarrow \hat{p_{u,i}}$
    - $Z_{u,i} \rightarrow \hat{Z_{u,i}}$

$$\hat{\tau_{u,i}}^{PC-IPS} = \frac{\hat{Z_{u,i}}Y_{u,i}}{\hat{p_{u,i}}} - \frac{\left(1 - \hat{Z_{u,i}}\right) Y_{u,i}}{1 - \hat{p_{u,i}}} : \text{Biased Estimator} \quad (1')$$

4. Proposed Approach: PROPCARE

4.5. Theoretical property

Proposition 1

$$\text{bias}\left(\hat{\tau_{u,i}}^{PC-IPS}\right) = \left(\frac{p_{u,i} + \mathbb{E}\left[\hat{Z_{u,i}} - Z_{u,i}\right]}{\hat{p_{u,i}}} - 1\right) Y_{u,i}^{1} - \left(\frac{1 - p_{u,i} - \mathbb{E}\left[\hat{Z_{u,i}} - Z_{u,i}\right]}{1 - \hat{p}_{u,i}} - 1\right) Y_{u,i}^{0} \quad (9)$$

Remark 1

$\text{bias}\left(\hat{\tau_{u,i}}^{PC-IPS}\right)$의 major factors:

$$\frac{p_{u,i}}{\hat{p_{u,i}}},\ \frac{1-p_{u,i}}{1-\hat{p_{u,i}}},\ \mathbb{E}\left[\hat{Z_{u,i}} - Z_{u,i}\right] \quad \rightarrow \quad \left(\hat{p_{u,i}} = p_{u,i},\ \hat{Z_{u,i}} = Z_{u,i} \Rightarrow \text{bias}\left(\hat{\tau_{u,i}}^{PC-IPS}\right) = 0\right)$$

DGIST DATA SCIENCE LAB

DGIST 대구경북과학기술원 Daegu Gyeongbuk Institute of Science & Technology

4. Proposed Approach: PROPCARE

4.5. Theoretical property

Remark 2

$$\hat{p_{u,i}} \approx 0 \text{ or } 1 \;\Rightarrow\; \text{bias}\left(\hat{\tau_{u,i}}^{PC-IPS}\right) \approx \pm\infty$$

- Exposure variable: $Z_{u,i} \in \{0,1\}$ is Binary variable. → F1 score같은 binary classification metrics 사용.
  - $\mathbb{E}\left[\hat{Z_{u,i}} - Z_{u,i}\right] \rightarrow \text{bias}\left(\hat{\tau_{u,i}}^{PC-IPS}\right)$이므로 $\hat{Z_{u,i}}$를 잘 추정할수록 bias가 작아짐.
- Propensity: $p_{u,i} := P\left(Z_{u,i} = 1\right)$ is Continuous variable. → KLD, Kendall's Tau같은 metrics 사용. [Section 5.2]
  - Remark 2에서 $\hat{p_{u,i}} \not\approx 0 \text{ or } 1$이어야 함. → Eq. (7)처럼 Regularization 사용하는 것이 좋다.

5. Experiment

5.1. Experiment setup

Datasets

- 3가지 standard Causality-based Recommendation Benchmarks
  : **DH_original, DH_personalized, MovieLens 100K (ML 100K)**
- **DH_original, DH_personalized** ∈ DunnHumby dataset
  - purchase and ptomotion logs @ 오프라인 소매점, 93주 기간동안.
  - DH_original: 주간 전단지 → Exposure → ground-truth Propensity Scores
  - DH_personalized: Simulation → ground-truth Propensity Scores
- **ML 100K**
  - Users' Rating on movies
  - Simulated Propensity Scores ← ratings & user behaviors
- PROPCARE: ground-truth propensity scores → Model Output Evaluation에만 사용.
  - Note: training 단계에서 ground-truth values 사용 ❌

5. Experiment

5.1. Experiment setup

Baselines

- PROPCARE vs Baselines [other methods]
- Propensity Estimators
  - **Ground-truth values: propensity $p_{u,i}$, exposure $Z_{u,i}$ → input of DLCE on training**
    1. **Ground-truth**: datasets → Propensity Score & Exposure values
  - **Estimate propensity $\hat{p_{u,i}}$ → Derive exposire $\hat{Z_{u,i}}$ → input of DLCE on training**
    2. **Random**: Propensity Scores $\in (0, 1)$ randomly
    3. Item Popularity (**POP**): Propensity Scores = Normalization of POP $\in (0, 1)$
    4. **CJBPR**: Propensity → Relevance → Propensity → Relevance → ... point-wise optimization
    5. **EM**: Expectation-Maximization algorithm → Propensity Scores point-wise learning

Parameter settings

- Validation data → Tuning hyper-parameters
  - PROPCARE: Use the trade-off hyper-parameters as
    - $\lambda = 10$
    - $\mu = 0.4$
  - Other settings: Appendix C.2.

DGIST DATA SCIENCE LAB

5. Experiment

5.1. Experiment setup

Evaluation metrics

- **Performance** of Causality-based Recommendation → Evaluation metrics [Appendix C.3.]
  1. **CP@10, CP@100**: Causal effect-based Precision (CP)
  2. **CDCG**: Causal effect-based Discounted Cumulative Gain (CDCG)

$$\mathrm{CP@K} = \frac{1}{U}\sum_{u=1}^{U}\sum_{i=1}^{I}\frac{\mathbf{1}(\mathrm{rank}_u(\hat{s}_{u,i}) \leq K)\tau_{u,i}}{K} \qquad (a.11)$$

$$\mathrm{CDCG} = \frac{1}{U}\sum_{u=1}^{U}\sum_{i=1}^{I}\frac{\tau_{u,i}}{\log_2(1+\mathrm{rank}_u(\hat{s}_{u,i}))} \qquad (a.12)$$

5. Experiment

5.2. Results and discussions

Table 2: Performance comparison on downstream causality-based recommendation.

Methods	DH_original			DH_personalized			ML		
	CP@10↑	CP@100↑	CDCG↑	CP@10↑	CP@100↑	CDCG↑	CP@10↑	CP@100↑	CDCG↑
Ground-truth	.0658±.001	.0215±.001	1.068±.000	.1304±.001	.0445±.001	1.469±.003	.2471±.001	.1887±.000	16.29±.006
Random	.0154±.001	.0071±.002	.7390±.004	.0479±.004	.0107±.005	.8316±.039	<u>.0124</u>±.002	<u>.0135</u>±.005	<u>13.16</u>±.076
POP	.0200±.000	<u>.0113</u>±.000	<u>.7877</u>±.001	.0457±.000	.0096±.001	.8491±.002	-.142±.001	-.092±.001	11.43±.005
CJBPR	<u>.0263</u>±.001	.0087±.001	.7769±.002	<u>.0564</u>±.008	.0106±.005	.8528±.032	-.410±.002	-.187±.001	9.953±.006
EM	.0118±.001	.0067±.001	.7247±.001	.0507±.002	<u>.0121</u>±.001	<u>.8779</u>±.003	-.437±.002	-.194±.002	10.21±.011
PROPCARE	**.0351**±.002	**.0156**±.001	**.9268**±.005	**.1270**±.001	**.0381**±.000	**1.426**±.001	**.0182**±.002	**.0337**±.002	**13.80**±.011

Results are reported as the average of 5 runs (mean±std). Except Ground-truth, best results are bolded and runners-up are underlined.

- **Performance** of Causality-based Recommendation [Evaluation metrics 비교.]
- Ground-truth: performance 가장 좋다. [Evaluation metrics 가장 큼.]
  - 실제 propensity, exposure values를 DLCE에 그대로 사용하기 때문.
    - → But real-world에서는 사용 불가.
  - PROPCARE: 가장 Ground-truth 값에 가까움. 특히 DH_personalized에서는 큰 차이 ❌
  - PROPCARE > CJBPR, EM
    - → Pairwise method by Assumption 1.이 좋다.

DGIST DATA SCIENCE LAB

5. Experiment

5.2. Results and discussions

Table 3: Performance comparison on propensity score (KLD, Tau) and exposure (F1 score) estimation.

Methods	DH_original			DH_personalized			ML		
	KLD↓	Tau↑	F1 score↑	KLD↓	Tau↑	F1 score↑	KLD↓	Tau↑	F1 score↑
Random	.5141±.001	.0002±.000	.4524±.013	3.008±.002	.0001±.000	.4463±.021	.0363±.002	.0002±.000	.4511±.022
POP	.5430±.000	**.4726**±.000	.2851±.000	4.728±.000	**.6646**±.000	.2772±.000	.0615±.000	**.4979**±.000	.5050±.000
CJBPR	.3987±.008	.3279±.011	.2853±.005	2.650±.022	.6477±.013	.2825±.005	.0230±.006	.4956±.045	**.5189**±.020
EM	.6380±.002	.0834±.000	.4974±.001	2.385±.001	.0934±.002	.4954±.009	.0517±.001	.1321±.002	.3653±.005
PROPCARE	**.3851**±.023	.3331±.065	**.5846**±.006	**1.732**±.038	.4706±.072	**.6059**±.017	**.0204**±.005	.3889±.034	.4847±.020

Results are styled in the same way as in Tab. 2.

- Propensity, Exposure Estimation Accuracy
- POP: Baselines 중에서, Kendall's Tau 기준 가장 좋다.
  - But Table 2를 보면 POP의 causality metrics는 좋지 못함.
    - KLD 값이 큰 것으로 보아, propensity score의 distribution 예측이 잘 되지 않았기 때문. → ill-fit propensity distribution.
  - F1 score가 작음 → Exposure estimation도 잘 되지 못하였음.
- PROPCARE: F1 score, KLD에서 효과가 좋음 & Table 2에서 causality metrics도 좋은 성능을 의미.
  - Tau scores가 다른 baselines보다 약간 나쁘지만, 나머지 두 지표가 좋음.
    - → Propensity Score, Exposure 둘 모두 estimation 잘됨. → Causal Performance 좋음.

5. Experiment

5.2. Results and discussions

Ablation study

$$\mathcal{L}_{\text{pop}} = -\kappa_{u,i,j} \log \left[\sigma(\text{sgn}_{i,j} \cdot (f_p(\mathbf{x}_{u,i}) - f_p(\mathbf{x}_{u,j}))) + \sigma(\text{sgn}_{i,j} \cdot (f_r(\mathbf{x}_{u,j}) - f_r(\mathbf{x}_{u,i}))) \right] \quad (6)$$

💡 $\mathcal{L}_{\text{pop}}$의 term를 변화시켰을 때, PROPCARE에 비하여 성능이 많이 떨어짐 → 변화된 term이 중요한 term이다.

- Derive 5 variants
  1. NO_P: *Removing the constraint on estimated* $\hat{p_{u,i}}$ by deleting the term with $f_p(\mathbf{x}_{u,i}) - f_p(\mathbf{x}_{u,j})$
  2. NO_R: *Removing the constraint on estimated* $\hat{r_{u,i}}$ by deleting the term with $f_r(\mathbf{x}_{u,j}) - f_r(\mathbf{x}_{u,i})$
  3. NO_P_R: *Removing* $\mathcal{L}_{\text{pop}}$ entirely from the overall loss $\mathcal{L}_{\text{total}}$ to eliminate Assumption 1 altogether
  4. NEG: Reversing Assumption 1 by *replacing* $\text{Sgn}_{i,j}$ *with* $-\text{Sgn}_{i,j}$ to assume that more popular items have smaller propensity scores
     - *Removing the condition* $(pop_i > pop_j \Leftrightarrow p_{u,i} > p_{u,j})$
  5. $\kappa = 1$: Setting all $\kappa_{u,i,j} = 1$ → equal weighting of all training triplets.
     - *Removing the condition* $y_{u,i} \approx y_{u,j}$

5. Experiment

5.2. Results and discussions

Ablation study

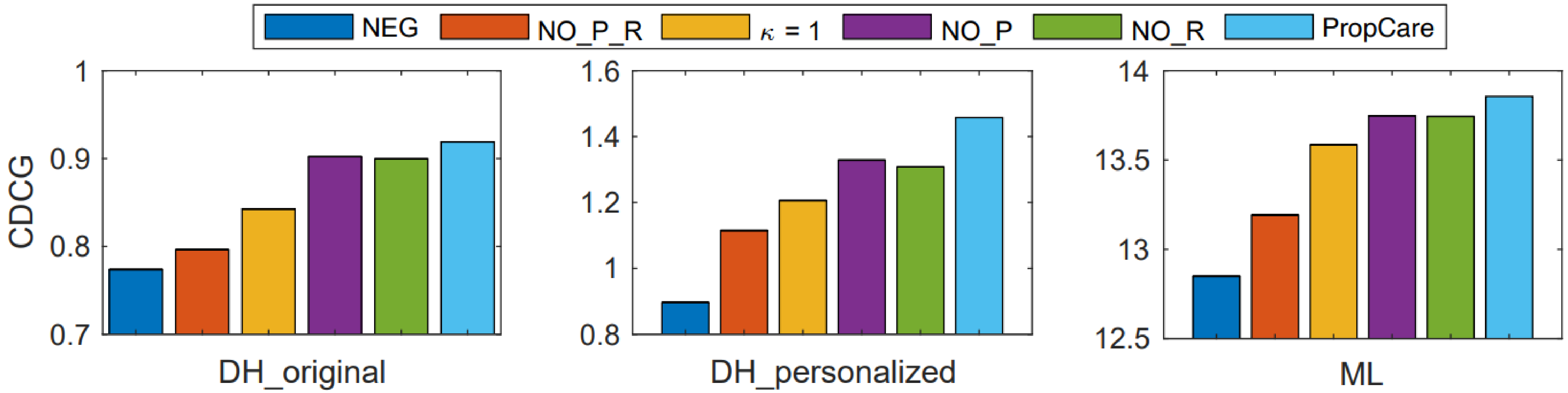


Figure 3: Ablation study on PROPCARE.

- x-axis: dataset
- y-axis: performance
- PROPCARE: best performance
- NEG: worst performance → Assumption 1 is most important.

Importance:

1. $(pop_i > pop_j \Leftrightarrow p_{u,i} > p_{u,j})$
2. $\mathcal{L}_{pop}$
3. $y_{u,i} \approx y_{u,j}$
4. $\hat{p_{u,i}}$
5. $\hat{r_{u,i}}$

5. Experiment

5.2. Results and discussions

Effect of regularization

$$\mathcal{L}_{\text{total}} = \sum_{u,i,j} (\mathcal{L}_{\text{naive}} + \lambda \cdot \mathcal{L}_{\text{pop}}) + \mu \cdot \text{KLD}\left(Q \| \text{Beta}(\alpha, \beta)\right) \quad (7)$$

$$\arg\min_{\Theta_{\text{total}}} \mathcal{L}_{\text{total}} = \hat{\Theta_{\text{total}}} \rightarrow \hat{p_{u,i}}, \hat{r_{u,i}}$$

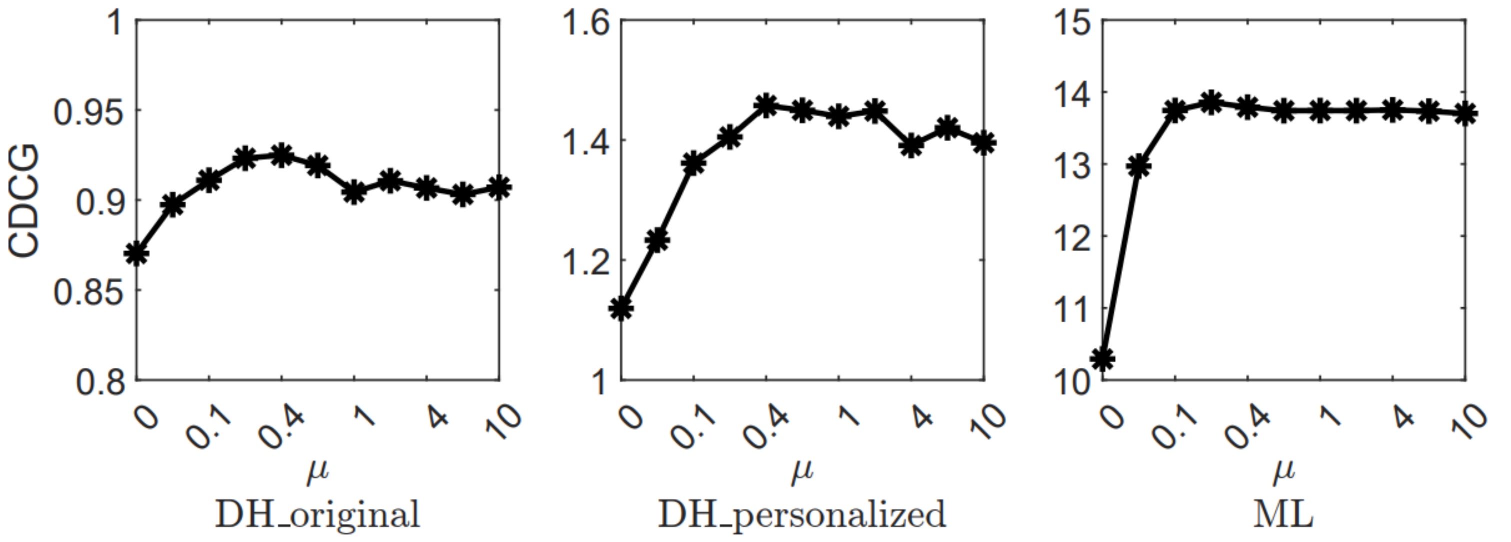


Figure 4: Effect of regularization term.

In this paper, Used 0.4

- $\mu \approx 0 \;\Rightarrow\;$ performance CDCG $\downarrow$
- $\mu \uparrow \Rightarrow\;$ performance CDCG $\uparrow$, $\quad \mu_{\text{peak}} \in [0.2, 0.8]$

5. Experiment

5.2. Results and discussions

Factors influencing causality-based recommendation

- 방법 1: ground-truth propensity or exposure values에 Noises Injection [각각 (b), (a)]

> Estimation Accuracy Vs Causality-based Recommendation Performance: **Important Factors**
>
> Propensity Score - Estimation Accuracy Vs Exposure - Estimation Accuracy

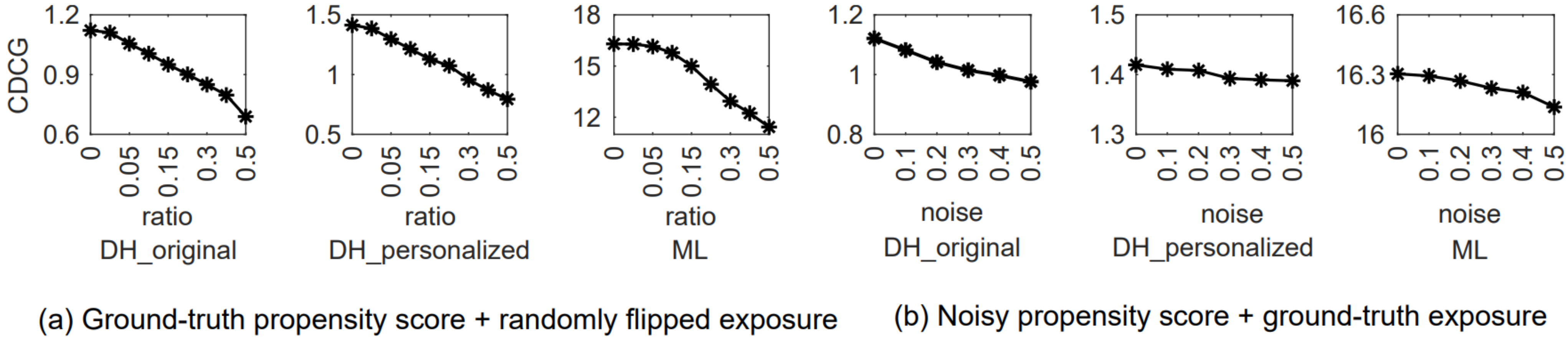


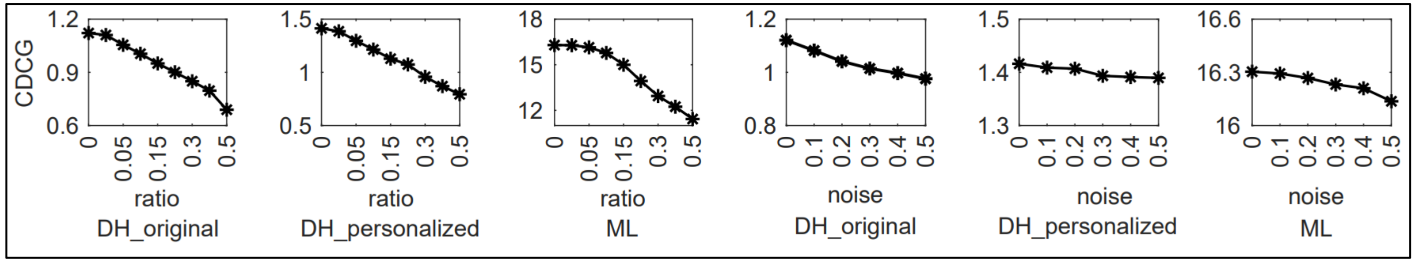
Figure 5: Analysis of factors that influence causality-based recommendation performance.

5. Experiment

5.2. Results and discussions

Factors influencing causality-based recommendation

- (a) Ground-truth propensity scores $p_{u,i}$를 DLCE training에 사용하면서, $Z_{u,i}$를 $0 \leftrightarrow 1$로 일부분 randomly flip.
  [$Z_{u,i}$ 오염.]
  - x-axis: Flip ratio
  - y-axis: CDCG performance
  - 오염 비중이 커질 수록 성능 급격히 하락.
    - → Causality-based Recommendation: Exposure의 Estimation에 매우 민감.



💡 **Exposure - Estimation Accuracy가 더 중요하다.**

- (b) Ground-truth exposure $Z_{u,i}$를 DLCE training에 사용하면서, Add Gaussian Noises to the propensity scores.
  [$p_{u,i}$ 오염.]
  - x-axis: Variance of Noises
  - y-axis: CDCG performance
  - 오염 비중이 커질 수록 성능 적당히 하락.
    - → Causality-based Recommendation: Propensity Score의 Estimation에 적당히 민감.

5. Experiment

5.2. Results and discussions

Factors influencing causality-based recommendation

- 방법 2: Correlation betw. Estimation Accuracy & Recommendation Performance

💡 Estimation Accuracy Vs Causality-based Recommendation Performance : **Correlation**

  - Dataset: Only DH_original dataset

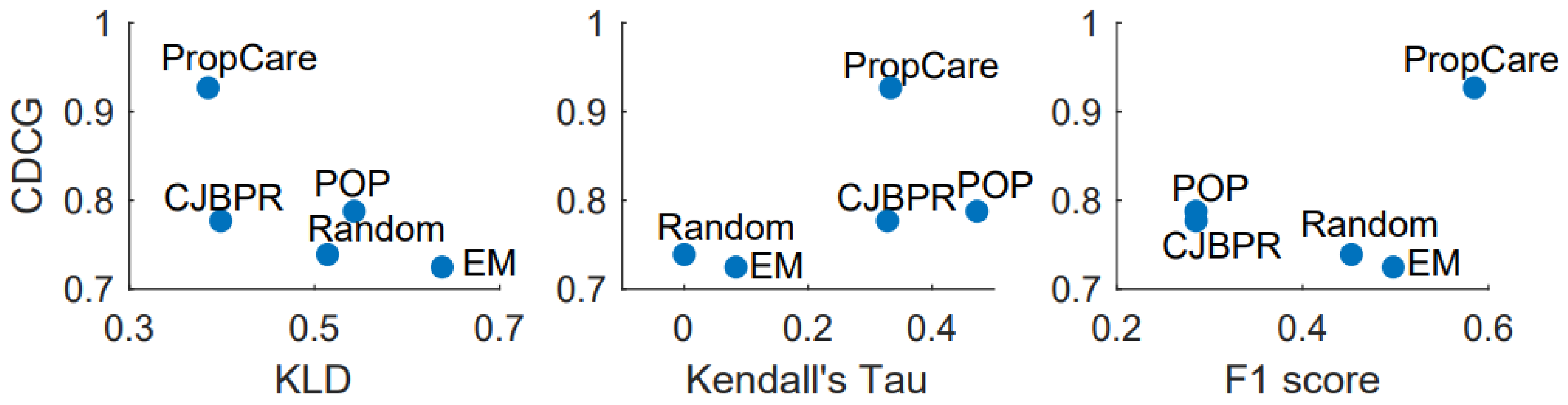


Figure 6: Correlation analysis on factors influencing recommendation.

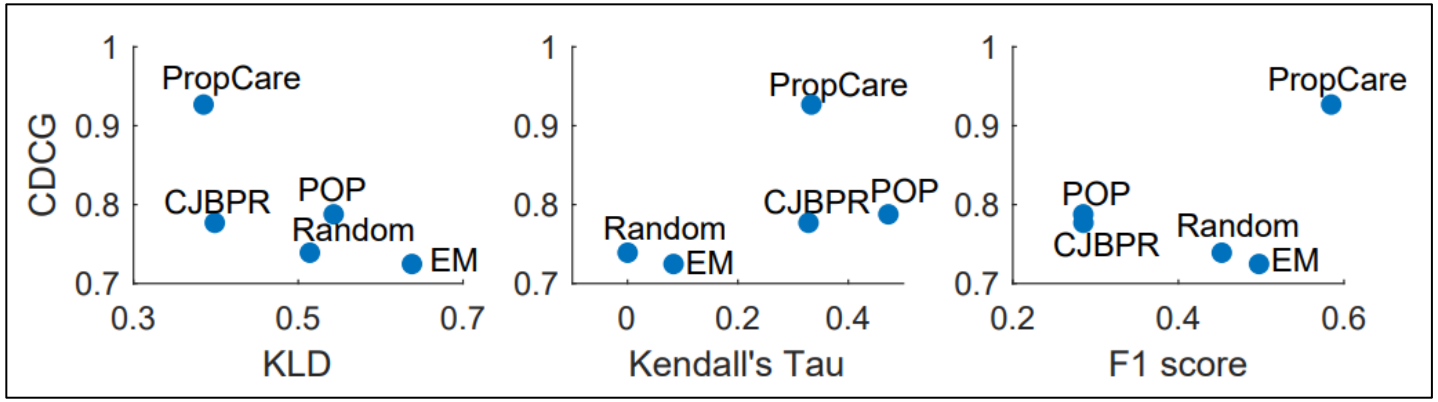
DGIST 대구경북과학기술원 Daegu Gyeongbuk Institute of Science & Technology

DGIST DATA SCIENCE LAB

5. Experiment

5.2. Results and discussions

Factors influencing causality-based recommendation



- x-axis: Estimation Accuracy
  - KLD, Kendall's Tau: Propensity Scores
  - F1 score: Exposure
- y-axis: CDCG performance [Recommendation Performance]

1. $\text{KLD} \downarrow \Rightarrow \text{CDCG} \uparrow$
   : Estimated Propensity Scores $\hat{p_{u,i}}$ 분포가 long-tailed인 beta 분포와 비슷할수록 Performance good.
2. $\text{Kendall's Tau} \uparrow \Rightarrow \text{CDCG} \uparrow$
3. $\text{F1 score} \uparrow \Rightarrow \text{CDCG} \uparrow$

5. Experiment

5.3. Case study

PROPCARE: Ranking-based Recommendation

Table 4: Case study of an anonymous user.

Ground-truth	POP	CJBPR	PROPCARE
garlic bread (1176)	bananas (1310)	‡ **fluid milk** (1169)	‡ **infant soy** (1232)
‡ **cleansing wipes** (737)	toilet tissue (742)	bananas (1310)	‡ **fluid milk** (1169)
‡ **fluid milk** (1169)	‡ **fluid milk** (1169)	cereal (1090)	bananas (1310)
‡ **primal** (807)	white bread (675)	**strawberries** (834)	pure juice (1277)
alkaline batteries (754)	¶ tortilla chips (634)	margarine tubs/bowls (1245)	coffee creamers (1169)

Each column represents the recommendation list output by DLCE trained with the estimated propensity and exposure by the corresponding baseline. The purchased items are highlighted in bold. Items with positive causal effect ($\tau_{u,i} = 1$) and negative causal effect ($\tau_{u,i} = -1$) are marked by ‡ and ¶, respectively, and unmarked items have zero causal effect ($\tau_{u,i} = 0$). Numbers in brackets are the popularity ranks in the training set.

5. Experiment

5.3. Case study

- Top-5 Recommend items [User ID 2308, DH_personalized dataset]

1. **Ground-truth: DLCE가 ranking list를 효과적으로 생성하였음.**
   - Most items가 Positive Causal Effect.
     - $\tau_{u,i} := Y^1_{u,i} - Y^0_{u,i} = 1$
     - → **Recommending item i to user u** ⇒ user-item interaction[click or purchase] 증가.
   - All items in positive causal effect: 모두 purchased 되었음. → Goal of Causality-based Recommendation 성공.
2. **CJBPR, PROPCARE: Purchased items의 Causal Effects는 각기 다르다.**
   - **CJBPR - strawberries:** causal effect ❌
     - **Recommending or not ⇒ user-item interaction[Purchasing]에 영향을 끼칠 수 없음.**
   - **PROPCARE - infant soy:** Positive causal effect
     - **Recommending item i to user u ⇒ user-item interaction[click or purchase] 증가.**
3. **POP: Negative Causal Effect를 가진 item(tortilla chips)도 recommend를 한다.**
   - **→ POP: 좋은 method가 아님.**

DGIST DATA SCIENCE LAB

DGIST 대구경북과학기술원 Daegu Gyeongbuk Institute of Science & Technology

6. Conclusion

- PROPCARE: w/o ground-truth of propensity and exposure data
- Observation of (propensity scores, item popularity) → Key Assumption → Prior Information → Causality-based RS
- Factors for bias in estimated causal effects
- Empirical studies: PROPCARE > Baselines [other methods]
- Future research suggestion:
  1. Direct exposure estimation w/o propensity scores [i.e, w/o propensity estimation]
  2. Parametric causal effect estimators [IPS estimator: Non-parametric approach]