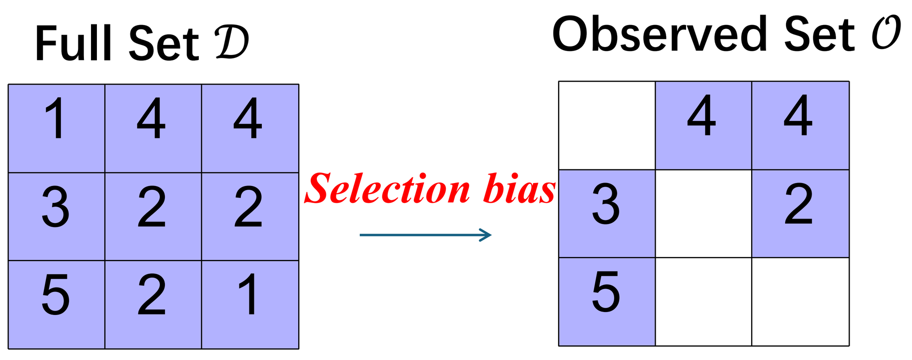
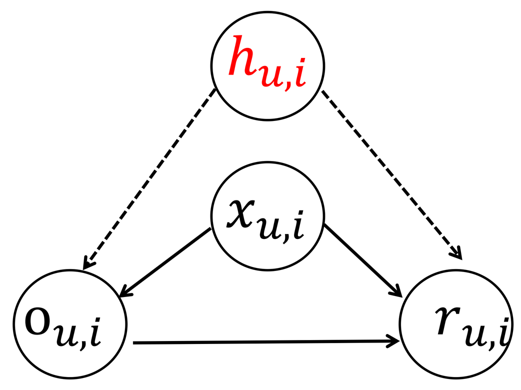
Temporally-Aware Integration of Randomized Controlled Trials for Mitigating Hidden Confounding in Recommendations

Yuchao Zhang[1], Tianyu Xia[2]

[1]Beijing University of Chemical Technology,[2]Peking University

Problem Setup





Selection bias happens in *explicit feedback data* as users are free to choose items to rate, so that the observed ratings are not a representative of all ratings. *Hidden confounding* occurs when *unobserved factors* jointly affect both item exposure and user feedback, causing biased recommendation results.

- $\mathcal{U} = \{u_1, u_2, ..., u_m\}$ is the set of m users. $\mathcal{I} = \{i_1, i_2, ..., i_n\}$ is the set of n items.
- $\mathcal{D} = \mathcal{U} \times \mathcal{I}$ is the set of all user-item pairs. $x_{u,i}$ is the feature of user u and item i.
- $\mathbf{R}^* \in \mathbb{R}^{m\times n}$ is the true rating matrix, $\hat{\mathbf{R}} \in \mathbb{R}^{m\times n}$ is the rating prediction matrix for $\mathbf{R}^*$. $r^*_{u,i}$ is the true rating of user u on item i, while $r_{u,i}$ is the observed rating.
- $o_{u,i}$ is an indicator. $o_{u,i} = 1$ means the rating of user u for item i is observed.
- True propensity score: $p_{u,i} = \mathbb{P}(o_{u,i} = 1|x_{u,i})$ or $p_{u,i} = \mathbb{P}(o_{u,i} = 1|x_{u,i}, h_{u,i})$ with hidden confounder $h_{u,i}$, and is estimated by a propensity model $\pi(x_{u,i};\psi)$.

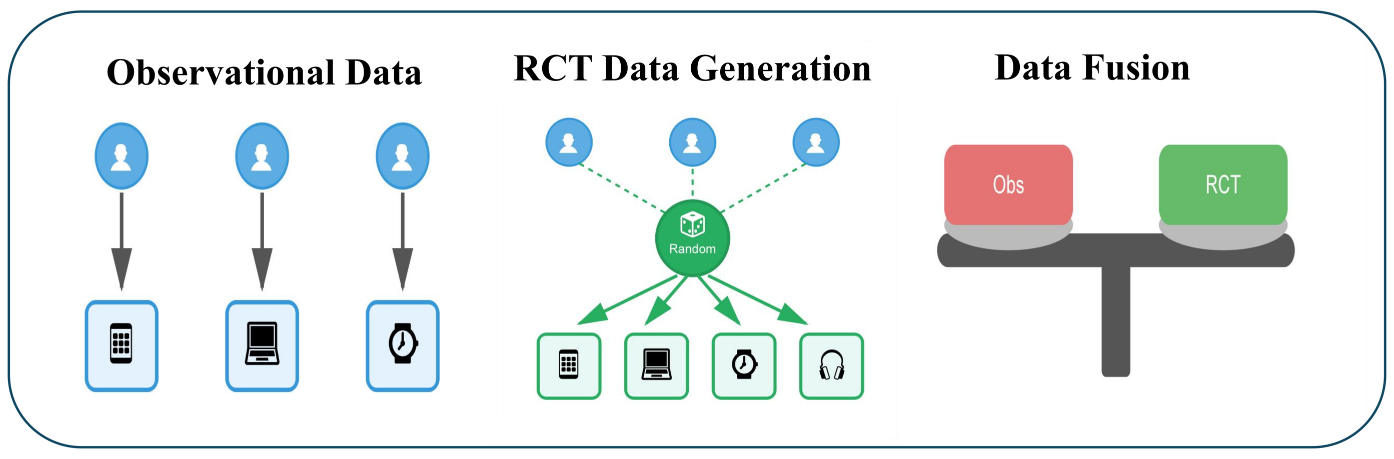
If all the ratings are observed, we can train the prediction model directly by minimizing the following ideal loss:

$$\mathcal{L}_{\text{ideal}}(\theta) = \frac{1}{|\mathcal{D}|}\sum_{(u,i)\in\mathcal{D}} e_{u,i}$$

where $e_{u,i} = \mathcal{L}(\hat{r}_{u,i}, r_{u,i})$ is the prediction error.

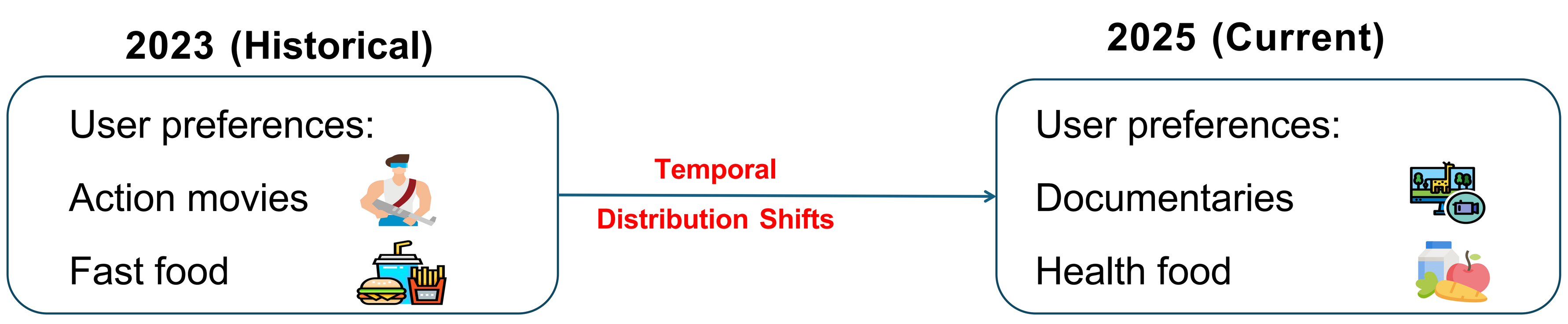
- Despite many methods have been proposed for achieving unbiased learning to under data MNAR, **previous methods don't consider use Randomized controlled trial (RCT) data to deal with unobserved confounder.**
- **RCT Data Fusion**: Auxiliary data can be used to correct propensity estimation errors induced by unobserved confounders:

$$\mathcal{L}_{\text{RCT}}(\theta) = \frac{1}{|\mathcal{D}_{\text{RCT}}|}\sum_{(u,i)\in\mathcal{D}_{\text{RCT}}} e_{u,i}$$



Motivation

- RCT-based methods rely on sufficient RCT data, but collecting such data is costly and results in scarcity.
- Practitioners merge historical and current RCT data, but temporal shifts lead to distribution mismatch.
- Naive merging can actually introduce bias. The historical RCT loses its "unbiased" property and introduces hidden confounding.



Method

- To align historical RCT data with current RCT data

Temporal Alignment Condition:

$$\frac{1}{|\mathcal{D}^{his}_{\text{RCT}}|}\sum_{(u,i)\in\mathcal{D}^{his}_{\text{RCT}}} w^{temp}_{u,i}\, e_{u,i} = \frac{1}{|\mathcal{D}^{cur}_{\text{RCT}}|}\sum_{(u,i)\in\mathcal{D}^{cur}_{\text{RCT}}} e_{u,i}$$

$\mathbf{w^{temp}_{u,i} = g(x_{u,i};\xi)}$ is temporal correction weight for alignment.

- The above alignment condition naturally provides a super-vision signal for learning g(ξ), formulated as

$$\mathcal{L}_{align}(\xi) = \left[\frac{1}{|\mathcal{D}^{his}_{\text{RCT}}|}\sum_{(u,i)\in\mathcal{D}^{his}_{\text{RCT}}} w^{temp}_{u,i}\, e_{u,i} - \frac{1}{|\mathcal{D}^{cur}_{\text{RCT}}|}\sum_{(u,i)\in\mathcal{D}^{cur}_{\text{RCT}}} e_{u,i}\right]^2$$

- Combining this with an entropy-based regularization term, the overall training objective for g(ξ) is given by:

$$\mathcal{L}_{temp}(\xi) = -\sum_{(u,i)\in\mathcal{D}^{his}_{\text{RCT}}} w^{temp}_{u,i}\log\left(w^{temp}_{u,i}\right) + \lambda\mathcal{L}_{align}(\xi)$$

- To Correct propensity scores for unobserved confounders

$\hat{p}_{u,i}$ deviates from $p_{u,i} \triangleq \mathbb{P}(o_{u,i} = 1|x_{u,i}, h_{u,i})$, to correct this discrepancy, we introduces balancing weights $w_{u,i} = w(x_{u,i};\phi)$ such that the reweighted inverse propensity score satisfies :

$$\frac{w_{u,i}}{\hat{p}_{u,i}} = \frac{1}{p_{u,i}}$$

Hidden Confounding Balancing Condition:

$$\frac{1}{|\mathcal{D}_{\text{OBS}}|}\sum_{(u,i)\in\mathcal{D}_{\text{OBS}}} w_{u,i}\frac{e_{u,i}}{\hat{p}_{u,i}}$$

$$= \frac{1}{|\mathcal{D}^{cur}_{\text{RCT}}|}\sum_{(u,i)\in\mathcal{D}^{cur}_{\text{RCT}}} e_{u,i} + \beta\left(\frac{1}{|\mathcal{D}^{his}_{\text{RCT}}|}\sum_{(u,i)\in\mathcal{D}^{his}_{\text{RCT}}} w^{temp}_{u,i}\, e_{u,i}\right)$$

- Balancing loss is defined as:

$\mathcal{L}_{bal}(\phi)$

$$= \left[\frac{1}{|\mathcal{D}_{\text{OBS}}|}\sum_{(u,i)\in\mathcal{D}_{\text{OBS}}} w_{u,i}\frac{e_{u,i}}{\hat{p}_{u,i}} - \frac{1}{|\mathcal{D}^{cur}_{\text{RCT}}|}\sum_{(u,i)\in\mathcal{D}^{cur}_{\text{RCT}}} e_{u,i} - \beta\left(\frac{1}{|\mathcal{D}^{his}_{\text{RCT}}|}\sum_{(u,i)\in\mathcal{D}^{his}_{\text{RCT}}} w^{temp}_{u,i}\, e_{u,i}\right)\right]^2$$

- Overall objective:

$$\mathcal{L}_w(\phi) = -\sum_{(u,i)\in\mathcal{D}_{\text{OBS}}} w_{u,i}\log\left(w_{u,i}\right) + \alpha\mathcal{L}_{bal}(\phi)$$

Reweighting with $\frac{w_{u,i}}{\hat{p}_{u,i}}$ approximates inverse true propensity weighting $\frac{1}{p_{u,i}}$, thereby guiding the prediction model toward unbiased learning. The resulting temporally enhanced objective is formulated as follows:

$$\mathcal{L}_{TE-Bal-IPS}(\theta) = \frac{1}{|\mathcal{D}_{OBS}|}\sum_{(u,i)\in\mathcal{D}_{OBS}} w_{u,i}\frac{e_{u,i}}{\hat{p}_{u,i}}.$$

Learning Algorithm of TE-Bal-IPS

Algorithm 1 Learning Algorithm of TE-Bal-IPS Estimator

Require: Three datasets $\mathcal{D}_{\text{OBS}}$, $\mathcal{D}^{\text{his}}_{\text{RCT}}$ and $\mathcal{D}^{\text{cur}}_{\text{RCT}}$, prediction model $f(\theta)$, propensity model $\pi(\psi)$, temporal alignment model $g(\xi)$, balancing weight model $w(\phi)$.

```
while stopping criteria is not satisfied do
    for steps for training propensity model do
        Sample a batch from D_OBS
        Update ψ by descending along gradient ∇_ψ L_ce(ψ)
    end for
    for steps for training temporal alignment model do
        Sample a batch from D^his_RCT and D^cur_RCT
        Update ξ by descending along gradient ∇_ξ L_temp(ξ)
    end for
    for steps for training balancing weight model do
        Sample a batch from D_OBS, D^his_RCT and D^cur_RCT
        Update φ by descending along gradient ∇_φ L_w(φ)
    end for
    for steps for training prediction model do
        Sample a batch from D_OBS
        Update θ by descending along gradient ∇_θ L_TE−Bal−IPS(θ)
    end for
end while
```

Ensure: θ

Experimental Results

Performance comparison in terms of AUC, NDCG@5 (N@5), and NDCG@10 (N@10). The best results are bolded.

Method	Music AUC	Music N@5	Music N@10	Coat AUC	Coat N@5	Coat N@10
CausE	0.731	0.551	0.656	0.761	0.500	0.605
KD-Label	0.740	0.580	0.680	0.750	0.504	0.610
MF (biased)	0.727	0.550	0.655	0.747	0.500	0.606
MF (uniform)	0.573	0.449	0.591	0.579	0.358	0.482
MF (combine)	0.730	0.554	0.659	0.750	0.503	0.611
IPS	0.723	0.549	0.656	0.760	0.509	0.613
Bal-IPS	0.727	0.564	0.668	0.771	0.521	0.628
Tem-Bal-IPS	**0.731**	**0.579**	**0.677**	**0.774**	**0.558**	**0.655**

S42208