Combining Reward and Rank Signals for Slate Recommendation

Imad Aouali, Sergey Ivanov, Mike Gartrell, David Rohde, Flavian Vasile,Victor Zaytsev, and Diego Legrand

Criteo AI Lab

CRITEO AI Lab

Abstract

We consider the problem of slate recommendation, where the recommender system presents a user with a collection of K items at once. If the user finds the recommended items appealing then the user may click and the recommender system receives some feedback. Two pieces of information are available: *was the slate clicked? (the reward)*, and *if the slate was clicked, which of its items was clicked? (the rank)*. In this paper, we formulate three Bayesian models that incorporate the *reward* signal (`Reward` model), the *rank* signal (`Rank` model), or both (`Full` model) for non-personalized slate recommendation. In our experiments, we analyze performance gains of the `Full` model and show that it achieves significantly lower error.

Introduction

In slate recommendation, historical data can be used to refine future recommendations by the use of two distinct signals:

- ***Reward signal.*** *was the slate clicked?*
- ***Rank signal.*** *if the slate was clicked, which of its items was clicked?*

Example. We consider a catalog of 3 items: ***phone, couscous***, and ***beer***. Ignoring order, there are 3 possible slates of size 2 that we can recommend: [***phone, couscous***], [***phone, beer***] or [***phone, couscous***]. An example of historical data is given in Table 1, where we display each of the 3 slates 700 times to the users. Here, slate [***couscous, beer***] is the best one. The most direct evidence for this is that it has the highest click through rate ($1-\frac{626}{700}\approx 0.11$), that is related to the *reward* signal. There is also indirect evidence using the *rank* signal that ***couscous*** is preferred to ***phone*** (29 clicks *vs.* 10), ***beer*** is preferred to ***phone*** (47 clicks *vs.* 9), and ***couscous*** is preferred to ***beer*** (46 clicks *vs.* 28). In aggregate, this ranking information also suggests that [***couscous, beer***] is the best slate. Thus, both signals shall contain important information of user preferences.

Slate	non-clicks	clicks on 1	clicks on 2
phone, couscous	661	10	29
phone, beer	644	9	47
couscous, beer	626	46	28

Table: Example of slate recommendation historical data.

Summary. We formulate three intuitive Bayesian models that use either the *reward* signal (`Reward` model), the *rank* signal (`Rank` model), or both (`Full` model). These algorithms learn from offline data similar to Table 1, and allow consistent estimation of the underlying reward model of slates. We demonstrate empirically that the `Full` model outperforms the other two approaches highlighting the benefits of combining *reward* and *rank* signals.

Bayesian Formulation

$$\Phi \sim \Gamma(1,1) \qquad \theta_i \sim \Gamma(1,1),\ i=1,\dots,N$$

Model Description	
`Full`	$nc, c_1, c_2 \mid I, \phi, \theta, a_1, a_2 \sim \text{Multinomial}\left(I, \frac{\phi}{\phi+\theta_{a_1}+\theta_{a_2}}, \frac{\theta_{a_1}}{\phi+\theta_{a_1}+\theta_{a_2}}, \frac{\theta_{a_2}}{\phi+\theta_{a_1}+\theta_{a_2}}\right)$
`Reward`	$nc, c \mid I, \phi, \theta, a_1, a_2 \sim \text{Multinomial}\left(I, \frac{\phi}{\phi+\theta_{a_1}+\theta_{a_2}}, \frac{\theta_{a_1}+\theta_{a_2}}{\phi+\theta_{a_1}+\theta_{a_2}}\right)$
`Rank`	$c_1, c_2 \mid I_c, \theta, a_1, a_2 \sim \text{Multinomial}\left(I_c, \frac{\theta_{a_1}}{\theta_{a_1}+\theta_{a_2}}, \frac{\theta_{a_2}}{\theta_{a_1}+\theta_{a_2}}\right)$

Table: Models formulation for slates of size 2.

with $[a_1, a_2]$ is the recommended slate. nc, c_1, c_2 denote, respectively, the number of non-clicks on $[a_1, a_2]$, the number of clicks on a_1 and the number of clicks on a_2. I is the number of impressions and I_c is the total number of clicks.

Learning

Maximum a posteriori & MCMC sampling. Having our data $\mathcal{D}$ of the form [**slate** a**, non-clicks on** a**, clicks on** a_1**, ..., clicks on** a_K]. Parameters ϕ and θ are inferred using both Maximum A Posteriori and Full Bayes principles where estimates $\hat{\theta}$ and $\hat{\phi}$ are obtained by maximizing the posterior $p(\theta, \phi|\mathcal{D})$. For violin plot (d), we generate samples $\tilde{\theta}_i$ from the posterior using MCMC methods in `Stan`.

Experimental setting

We generate n samples of user interactions with each slate, using a Multinomial distribution with known parameters $\phi = 100$ and θ containing values evenly spaced from 1 to 6. We then fit our models to generated data, and evaluate the ability of each model to estimate the true parameters of the generative process. Since all models estimate parameter θ, we use this parameter to evaluate the performance by computing:

$$L_1(p_{\hat{\theta}}, p_\theta) = \sum_{\text{all slates } a} \left| \frac{\hat{\theta}_{a_1}}{\sum_{j\in[K]}\hat{\theta}_{a_j}} - \frac{\theta_{a_1}}{\sum_{j\in[K]}\theta_{a_j}} \right| \tag{1}$$

Experimental results

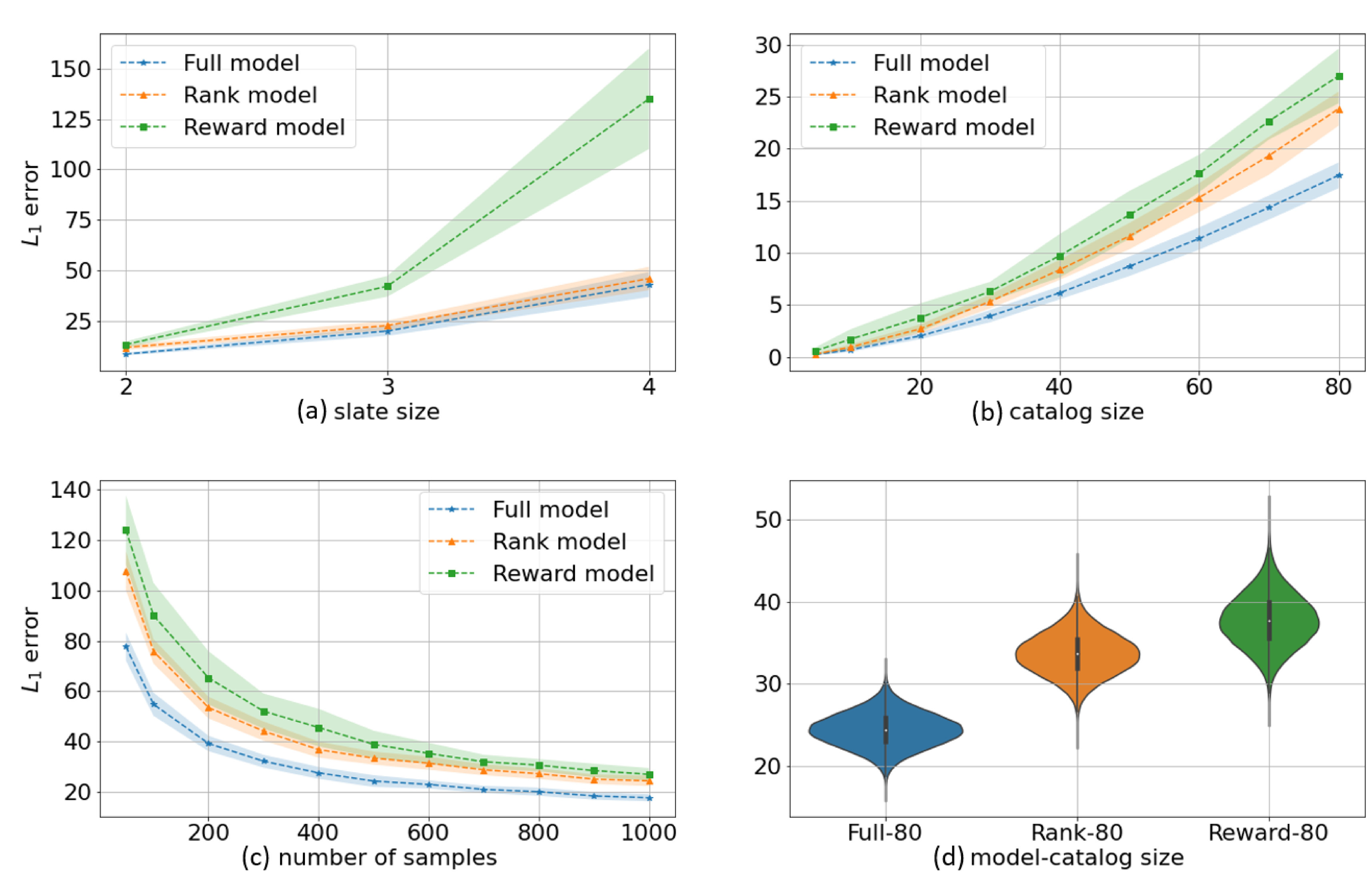

Figure: Figures (a, b ,c): L_1 error (Eq. 1) for varying slate size, catalog size, and number of samples. In each experiment, we run the models 50 times and average the results. Shaded areas represent uncertainty. Figure (d): Violin plot of L_1 errors distribution. Here, we generate samples $\tilde{\theta}_i$ from the posterior and calculate the L_1 distance (Eq. 1) between vectors p_θ and $p_{\tilde{\theta}_i}$ for all samples $\tilde{\theta}_i$. This results in a set of L_1 errors that we visualize using the violin plot.

conclusion

Combining both *reward* and *rank* signals can be beneficial in slate recommendation, especially as the catalog size and slate size grow.

Future work. Extend the `Full` model to personalized slate recommendation.