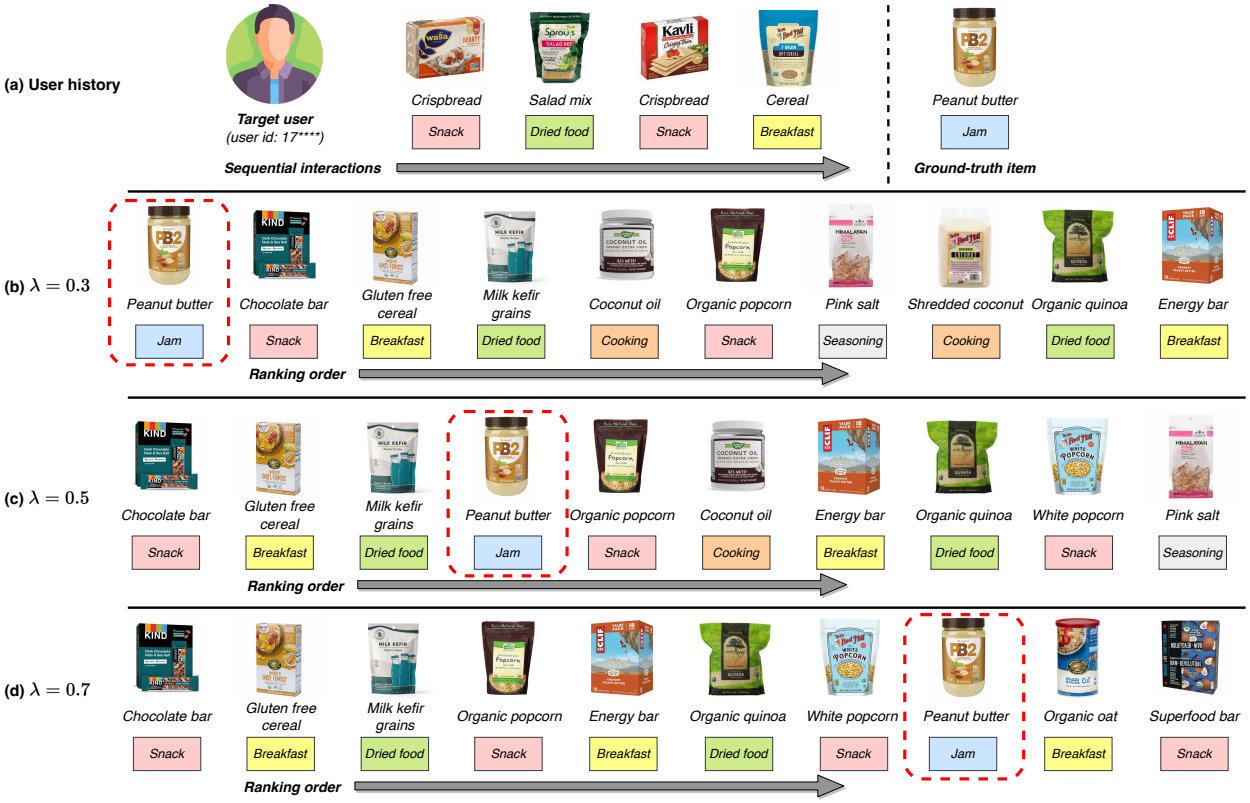

Figure 1: (a) A real user's sequential interactions and ground-truth next item in Grocery dataset. (b-d) The top-10 recommended items for the user by LeapRec with different λ: 0.3, 0.5, and 0.7. We mark the ground-truth item with a red box in each recommendation result. Colored boxes denote the categories. The values of $\mathcal{S}_{KL}$@10 are 0.5476, 0.3592, and 0.1084 respectively for $\lambda = 0.3$, $\lambda = 0.5$, and $\lambda = 0.7$. This case shows that LeapRec accurately recommends a relevant item from a non-interacted category even while improving calibration.

In Figure 1, we analyze a case to observe how LeapRec balances relevance and calibration when recommending items to a user who prefers a new category, on Grocery dataset. We observe how the user is provided with recommendations while varying the calibration level. Figure 1 (a) shows the user's history and the next item the user will interact with. Figures 1 (b-d) show the recommendation results of LeapRec for λ values 0.3, 0.5, and 0.7, respectively. In the results (b-d), LeapRec recommends the ground-truth item *Peanut butter* which is relevant to the user, despite its contrast to the user's category preference. As λ increases, we observe that *Peanut butter* is gradually lower ranked but is retained in the recommended list up to $\lambda = 0.7$. This case illustrates that LeapRec effectively reflects relevance even for an item that is from a category the user has not interacted with before.