

# Save, Revisit, Retain: A Scalable Framework for Enhancing User Retention in Large-Scale Recommender Systems

Weijie Jiang, Armando Ordorica, Jaewon Yang,  
Olafur Gudmundsson, Yucheng Tu\*, Huizhong Duan

Pinterest Inc.

weijiejiang,aordorica,jaewonyang,ogudmundsson,ytu,hduan@pinterest.com

## Abstract

User retention is a critical objective for online platforms like Pinterest, as it strengthens user loyalty and drives growth through repeated engagement. A key indicator of retention is revisitation, i.e., when users return to view previously saved content, a behavior often sparked by personalized recommendations and user satisfaction. However, modeling and optimizing revisitation poses significant challenges. One core difficulty is accurate attribution: it is often unclear which specific user actions or content exposures trigger a revisit, since many confounding factors (e.g., content quality, user interface, notifications, or even changing user intent) can influence return behavior. Additionally, the scale and timing of revisitations introduce further complexity; users may revisit content days or even weeks after their initial interaction, requiring the system to maintain and associate extensive historical records across millions of users and sessions. These complexities render existing methods insufficient for robustly capturing and optimizing long-term revisitation.

To address these gaps, we introduce a novel, lightweight, and interpretable framework for modeling revisitation behavior and optimizing long-term user retention in Pinterest’s search-based recommendation context. By defining a surrogate attribution process that links saves to subsequent revisitations, we reduce noise in the causal relationship between user actions and return visits. Our scalable event aggregation pipeline enables large-scale analysis of user revisitation patterns and enhances the ranking system’s ability to surface items with high retention value. Deployed on Pinterest’s Related Pins surface to serve 500+ million users, the framework led to a significant lift of 0.1% in active users without additional computational costs. Our data analysis reveals novel insights, such as the impact of content topics on revisitation rates; for example, users are more likely to revisit aesthetically pleasing topics.

## Introduction

User retention is an important business objective for online platforms such as Pinterest, as it fosters stronger relationships with users, builds loyalty, and ultimately drives both growth and revenue. A key behavior indicating strong user retention is revisitation (Zhang et al. 2021), which occurs when a user returns to view specific content that they

previously accessed and saved. This behavior often signals user satisfaction, typically driven by personalized content and recommendations. Therefore, identifying and leveraging previously saved content that encourages users to return can serve as a valuable signal for fostering long-term user retention.

Modeling and optimizing revisitation behavior presents several challenges. The first is attribution: it is difficult to determine what exactly triggers a user to revisit. Many confounding factors can influence revisitation, such as content quality, user experience, and platform notifications or reminders (Adar, Teevan, and Dumais 2008; Zhang et al. 2021). For instance, if a user returns with a completely different intent than in their previous session, it would be inappropriate to attribute that revisit to their earlier activities. The second challenge concerns scale. Since revisitations can occur within a day, a week, or even a month (Adar, Teevan, and Dumais 2009), tracking them effectively requires maintaining extensive records of user activity over indefinite periods. This creates significant data processing and scaling challenges (Jin et al. 2017). To date, there has been little concrete research on whether it is reasonable to disregard revisitations after a certain period (e.g., after X days).

Because of these challenges, many existing methods prioritize immediate engagement metrics (e.g., clicks, saves) over longer-term retention (Gugnani and Misra 2020; Wang et al. 2020; Dang et al. 2025; Gong et al. 2025; Jia et al. 2025). In some cases, effective modeling also increases the time users spend (Guo et al. 2022). Reinforcement learning (RL) offers a framework for optimizing long-term rewards. Although RL methods have improved intra-session engagement (Afsar, Crump, and Far 2022; Cai et al. 2023; Liu et al. 2024), applying RL to optimize longer-term outcomes such as revisitation (events that may occur days after exposure) remains highly challenging. Existing RL approaches often lack the scalability and robustness to handle multi-day sequences that may encompass thousands of recommended items. To the best of our knowledge, there are currently no RL-based approaches that, at Pinterest’s scale, explicitly target long-term revisitation while jointly addressing attribution and scalability.

In this study, we propose a novel lightweight and interpretable framework for modeling user revisitation behavior and optimizing for long-term user retention in the

\*Work done at Pinterest.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

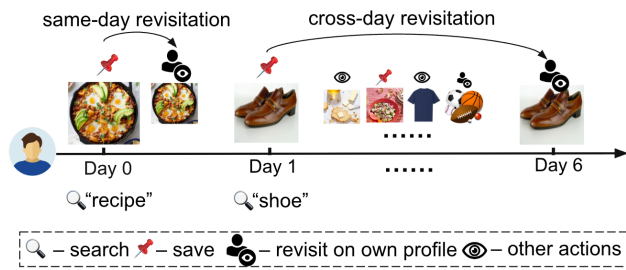


Figure 1: Revisitation Attribution Illustration

recommendation context of search. Research shows that well-chosen user behaviors can serve as effective surrogate rewards for predicting and optimizing long-term engagement in recommender systems (Wang et al. 2022). Therefore, to address the first challenge of attribution, we establish a surrogate for the causal effect of saving an item on subsequent revisitation events to reduce ambiguity and noise in revisitation signals. Figure 1 illustrated two types of revisitation attribution process.

- *Same day revisitation:* A user searched for “recipe” on day 0 and saved a particular recipe item to their profile. Later that day, they revisited the saved recipe item in their profile. In this case, we attribute revisitation to the saved recipe item. We do not attribute other items the user viewed on other days to the saved recipe item.
- *Cross-day revisitation:* The user saved a shoe item via search on day 1 and viewed some other items in subsequent days. We do not attribute revisitation to the shoe item until the user revisited it in own profile on day 6.

To leverage the constructed revisitation attributions and address the second challenge of scale, we first conduct a comprehensive analysis of user revisitation patterns and then translate the findings into a scalable data pipeline. The pipeline aggregates event data over a seven-day window across sessions and presentation surfaces by linking search-surface save events to profile-surface revisit events. This aggregation allows us to capture the most important revisitation signals for saved items, which serve as an auxiliary prediction task for the ranking model. In this way, the recommender system prioritizes items with high predicted probabilities of saving and revisit in the top positions, thereby driving downstream return visits and promoting long-term retention.

This framework was shown to be effective in significantly increasing active users by 0.1% through both extensive offline experiments and a large-scale online A/B test with 24 million users on Pinterest’s Related Pins surface (Liu et al. 2017) over two months. In summary, our contributions include:

- **Large-scale analysis:** We provide the first large-scale analysis of user revisitation patterns on online platform with hundreds of millions of users.
- **Methodology & metrics improvements:** We propose a novel, lightweight, and scalable framework for modeling user revisitation behavior without additional cost, which

is demonstrated to improve long-term user retention via a large-scale online experiment of 24 millions of users.

- **Interpretability & insights:** The proposed framework offers greater interpretability than the SOTA methods, and it is able to show the reasoning behind the metrics improvement and the long-term and short-term revisitation patterns on topics.
- **Deployment:** The proposed method has been deployed on Pinterest’s Related Pins surface (Liu et al. 2017), efficiently serving 500+ millions of users without incurring additional computational costs..

## Related Work

User retention is critical objective for online platforms as sustained engagement typically drives more growth and revenue than short-term interactions. However, typical recommender systems in online platforms mainly focused on immediate engagement metrics such as click-through rates but struggled to capture the temporal dynamics essential for user retention. The first challenge is that there are lots of noise in the collected data of user behaviors when modeling retention and it is difficult to find the factors that contribute to user retention (Sun et al. 2025). Ding et al. (2023) designed a contrastive multi-instance learning framework to explore the rationale and improve the interpretability of user retention, and they argue that recommender systems should rank different items for a user according to the user-item retention scores. The second challenge is the supervised information of user retention has been expected to be much sparser than immediate explicit feedback, such as click and save. Generative Flow Networks (Liu et al. 2024) have been proposed to model retention by treating it as a probabilistic flow over user sessions, which aims to address the challenges of sparse and delayed retention signals. DT4Rec (Zhao et al. 2023) leveraged the Decision Transformer to optimize long-term user retention by modeling sequences of interactions. This approach incorporates an autodiscretized reward prompt that preserves partial order relationships between rewards, enhancing retention modeling over time.

Other workstream such as customer purchase behavior analysis in industries like fashion have empirically validated that diverse product recommendations not only increase purchase rates but also positively influence long-term retention (Kwon, Han, and Han 2020). Fashion is one of the most popular topic on Pinterest. However, our past experiments when intentionally improving recommendation diversity hurt engagement metrics like click and save by more than 1% and we observed the tradeoff tension between relevance and diversity.

Wang et al. (2022) demonstrated that well-chosen user behaviors can serve as effective surrogate rewards for predicting and optimizing long-term engagement in recommender systems. Through deep analyses on the revisitation behaviors of users on Pinterest, we observed that user revisitation behaviors are less sparse than some other user behaviors such as long click and share. Therefore, we establish our surrogate as the causal relationship between a saved item and its associated revisitation events. Although Zhang et al.



Figure 2: Illustration of Save, Revisitation Impression and Revisitation Grid-click on Pinterest

(2021) designed a separate model for modeling the probability of click  $\rightarrow$  revisit and used it together with the ranking model that predicts click rate for predicting retention, deploying an additional model is very costly and increases serving latency. In this work, we propose a revisitation modeling framework that jointly models users' save and revisitation actions within the existing multi-task recommender, adding no cost or latency relative to the current system.

### Revisitation Behavior Analysis

Users engage in a variety of behaviors on online platforms, such as browsing, clicking on grids, liking, saving, sharing, and commenting. They can also return to the platform for many reasons, from upper-funnel activities like aesthetically-pleasing content to completing lower-funnel projects like DIY and Crafts, or performing outbound clicks on previously saved outfits. Among these behaviors, revisitation stands out as a special type of return behavior, where a user goes back to content they had previously saved, an action that arguably indicates user satisfaction.

In this work, we leverage the common “save” action on online platforms to build an attribution between saved content and subsequent revisitation behaviors. Specifically, when a user saves an item, it is added to their personal profile (such as a collection or board). Users can then revisit their saved items in two primary ways:

- **Impression-based revisitation:** Scroll or navigate through the saved content without deeper interaction on their personal profile (such as collection or board).
- **Grid-click-based revisitation:** Tap on the saved content to access more information on their personal profile (such as collection or board).

Figure 2 shows an example of a user who saved a Pin of a ramen recipe on Pinterest, had a revisitation impression on the Pin on their own board several days later, and then a revisitation grid-click to zoom in the Pin for more details.

We conducted extensive analyses on these two revisitation behaviors and observed the following patterns:

1. **More users tend to have revisitation impressions rather than revisitation grid-clicks, and the percentage of users revisiting decays (Figure 3a).** Suppose a user saved a Pin on day 0, 14.6% of the users tend to have a revisitation impression on the saved Pin on day 0

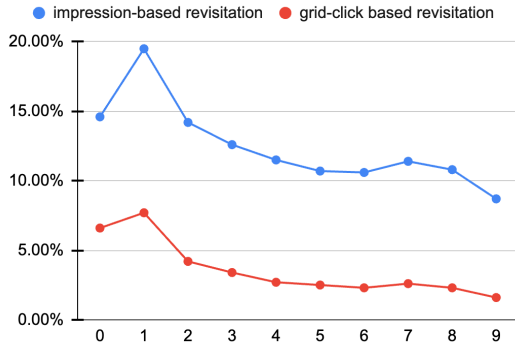
immediately and 19.5% of the users on day 1, following a significant decay till 8.7% on day 9. The percentage of users have a revisitation grid-click starts at 6.6% at day 0 and peaked at day 1 of 7.7%, and decays to only 1.6% day 9.

2. **The volume of revisitation grid-clicks decays dramatically in the first three days (Figure 3b).** Only 4.7% of the saved Pins get a revisitation grid-click on day 0 and the percentage dropped by half on day 1 followed by another half on day 3. The percentage dropped more steadily until it reaches 0.5% by day 5 and 0.3% by day 9.
3. **The sooner they revisit, the higher the expected number of days active in the next 1-month period.** Figure 4 shows that: (1) users who revisited by day  $t$  tend to have more days active in the following 1-month period compared to users who did not revisit by day  $t$  (distribution less left skewed); (2) users who revisited sooner tend to have more days active in the following 1-month period. It is worth mentioning that the fact that they revisit soon does not imply that they will be more active. Instead, it could also be true that users who are more active tend to come back sooner.
4. **Revisitation grid-click drives deeper engagement than just revisitation impression.** We have one group of users with only revisitation impressions and another group of users with revisitation grid-click. We computed the difference in their active days between the two groups in the 1-month period after the initial revisitation by Day 0 to Day 6 and estimate their incremental change within the 1-month period. Figure 5 shows that **Grid-click-based revisitation** has a higher correlation with the change in the 1-month activity than pure Impression-based revisitation. Unsurprisingly, this suggests that more “intentional” revisitation (i.e. via a grid click over just an impression) yields superior retention outcomes.

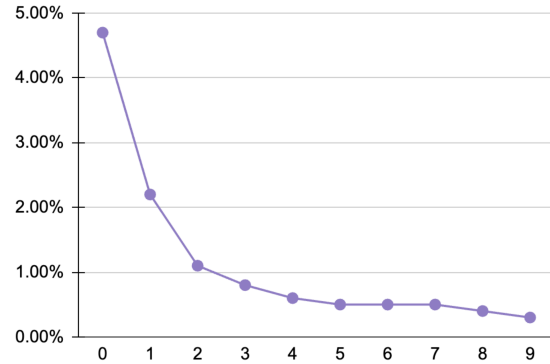
### Revisitation Modeling in Multi-task Learning Recommender System

We proposed revisitation modeling on the Related Pins surface on Pinterest (Liu et al. 2017) (Figure 6), a search-like surface which serves around 50% of the traffic on the entire platform. Users enter the Related Pins surface when they grid-click a Pin on any other upper-stream surface (such as Homefeed or Search). There are two main parts of the Related Pins surface: (1) A closer view of the Pin (query Pin) they just grid-clicked on from an upper-stream surface where they can read the title and descriptions of the Pin, and save the Pin to their own profile so they can revisit it any time later (Figure 6 left). (2) When user scrolls down, they will see a grid view of recommended Pins under title “More like this” which are related to the query Pin above (Figure 6 right). Users can engage with the the recommended Pins, a.k.a. candidate Pins with the following main actions:

- **Impression:** Scroll or navigate through a Pin without deeper engagement.



(a) Percentage of Users Revisiting Per Day



(b) Percentage of Grid-click based Revisitation Volume Per Day

Figure 3: Revisitation analysis (users & volume) - Day 0 is when a Pin was saved

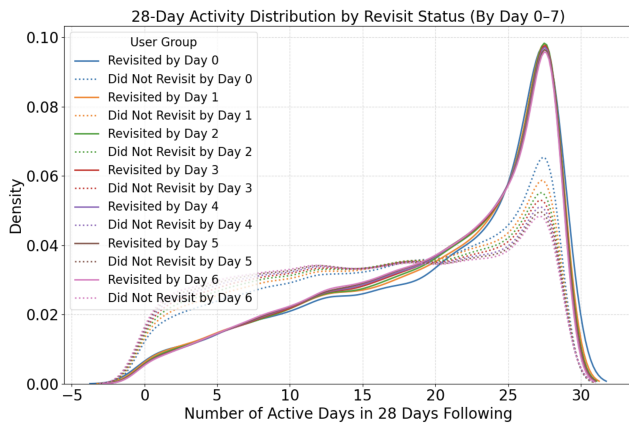


Figure 4: 28-day activity distribution by revisit status (revisit by day 0-6)

- **Grid-click:** A single tap on the Pin that makes it appear larger. When a grid-click happens, user will enter another Related Pin feed where the query Pin becomes the previously tapped Pin and new recommendations will be generated for that Pin.
- **Save (a.k.a. Repin):** Click on the “Save” button to save the Pin to user’s own profile so they can revisit it any time later.
- **Click:** Click on the “Visit” button to view more information about the Pin from a third-party website (if there is the Visit button for that Pin).
- **Long click:** User stay on the third-party website for longer than 35s.

### Multi-task Learning Recommender System

The goal of a multi-task recommender system is to jointly optimize for multiple related tasks by leveraging shared representations and correlations among those tasks (Zhang and Yang 2021). Figure 7 (left) shows the main architecture of the multi-task learning ranking model for recommendation

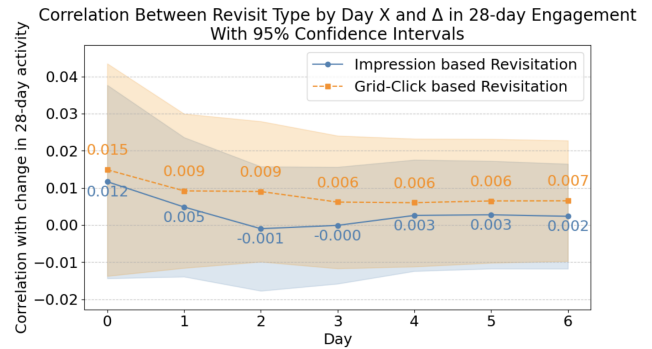


Figure 5: Correlation Between Revisit Type by Day X and Δ in 28-day Engagement with 95% Confidence Intervals

on Related Pins. The input layer includes query Pin features, candidate Pin features, user features, and a transformer-based user sequence module that encodes user historical actions. The summarization layer includes some feature crossing manipulation between query Pin features and candidate Pin features and dimension reductions, where the output embedding is fed to a DCNv2 module (Wang et al. 2021) to learn deeper feature interactions. A final MLP layer aggregates the output embeddings of DCNv2 into a single embedding and feed it to an MMoE module (Ma et al. 2018) for multi-task learning. Each action is binary so the loss function is the sum of binary cross entropy loss (Equation (1)),

$$\text{loss}_1 = \sum_{i \in \mathbb{L}} w_i \cdot [c_i \log p(c_i | q, \theta) + (1 - c_i) \log p(1 - c_i | q, \theta)] \quad (1)$$

where  $\mathbb{L} = \{\text{grid-click}, \text{repin}, \text{click}, \text{longclick}\}$ ,  $q$  refers to query Pin,  $c$  refers to candidate Pin,  $\theta$  denotes model parameters,  $c_i$  is binary label, where 1 indicating action  $i$  is positive and 0 means negative, and  $w_i$  refers to the loss weight for task  $i$ .

The recommender system requires an aggregated score based on the prediction of each task to rank all the candi-

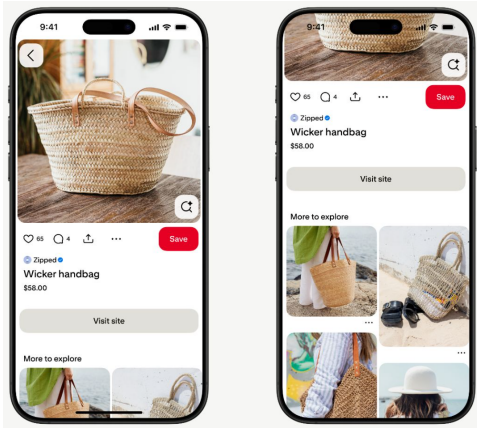


Figure 6: Related Pins Surface on Pinterest

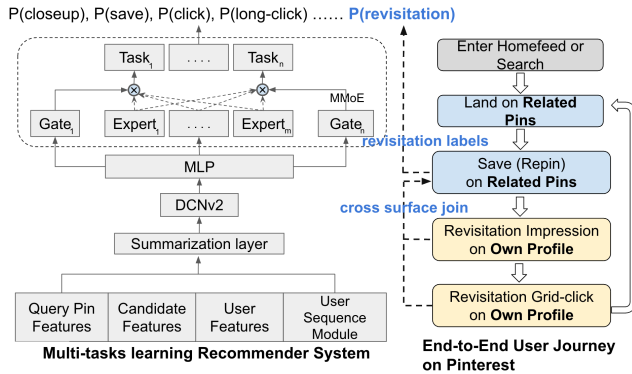


Figure 7: Revisitation Data Generation and Modeling in Multi-task Learning Recommender System

date Pins. For each task, a utility weight  $u_i$  representing the importance of the action based on business need is tuned and assigned to  $p(c_i|q)$  and all the tasks are aggregated to calculate the final score for ranking the candidate Pins (Equation (2)).

$$\text{score} = \sum_{i \in \mathcal{L}} u_i \cdot p(c_i|q, \theta) \quad (2)$$

## Revisitation Label Design

*Goal.* We seek a revisitation signal that balances label density (coverage) and intent (precision) as the best proxy for retention.

*Interaction type.* Impression-based revisits provide high coverage but include incidental scroll noise; grid-click revisits are rarer but indicate deliberate return intent. Section “Revisitation Behavior Analysis” shows grid-click revisits correlate more strongly with downstream activity than impressions.

*Temporal window.* In section “Revisitation Behavior Analysis”, we observed that revisits concentrate shortly after saves (sharp decay within the first week). In addition, the sooner revisitation happens, the more likely it is for users to be more active within the following 1-month period. There-

fore, we utilized the revisitation behaviors that happened 0-6 days after repin as revisitation labels as they are more recent and take up most of the revisitation actions. We surmised that driving more revisitation behaviors on the saved Pin in the following week could lead to an increase in active days by users and potentially improve user retention. In addition, based on our finding that revisitation grid-clicks take less than half of the volume of revisitation impressions but they tend to drive deeper engagement than revisitation impressions, we utilized revisitation grid-clicks that happened 0-6 days after repin while only use same-day revisitation impressions as repin. The final revisitation labels include:

- **Same-day Revisitation Impression (1dRevImpre):** User saved a Pin and has a revisitation impression of the saved Pin on the same day. The volume of same-day revisitation impressions consists of 25.8% of the repin volume in a day.
- **Same-day Revisitation Grid-click (1dRevGrid):** User saved a Pin and has a revisitation grid-click of the saved Pin on the same day. The volume of same-day revisitation grid-clicks takes up 4.7% of the repin volume in a day.
- **7-day Revisitation Grid-click (7dRevGrid):** User saved a Pin and has a revisitation grid-click of the saved Pin in the following 0-6 days. The volume of 7-day revisitation grid-click consists of 9% of the repin volume in a day in total as we see a quick decay of revisitation grid-click after repin on day 0 in Figure 3b.

We merge the three types of labels as the final revisitation label, and add a revisitation head in the multi-tasks learning ranking model for predicting the probability of a candidate Pin being saved and revisited in the future using the constructed revisitation label. Therefore, the loss function with the additional revisitation head is shown in Equation (3), where  $RP$  indicates *Repin* and  $RV$  stands for *Revisitation*.

$$\text{loss}_2 = \text{loss}_1 + w_{RP\&RV} \cdot [c_{RP\&RV} \log p(c_{RP\&RV}|q, \theta) + (1 - c_{RP\&RV}) \log p(1 - c_{RP\&RV}|q, \theta)] \quad (3)$$

The final ranking score is shown in Equation (4), and we tune the utility weight  $u_{RP\&RV}$  for the new task.

$$\text{score} = \sum_{i \in \mathcal{L}} u_i \cdot p(c_i|q, \theta) + u_{RP\&RV} \cdot p(c_{RP\&RV}|q, \theta) \quad (4)$$

## Revisitation Features Design

The current recommender system on the Related Pins surface takes various Pin features as input, including content features, such as text embeddings and visual embeddings (Zhai et al. 2019), and graph based embeddings, such as *graphsage* embeddings (Hamilton, Ying, and Leskovec 2017), *pinnersage* embeddings (Pancha et al. 2022), *omnisage* embeddings (Badrinath et al. 2025), and engagement features like *Navboost* (Kislyuk et al. 2015). One of the most important engagement features that have been adopted on all the surfaces is called Pin perf, counting the historical number of actions on a specific Pin. For each positive action (including grid-click, save, click, and long click), the Pin perf

feature counts the how many times a Pin has been engaged with that action on Pinterest in the past 90 days.

In this work, we designed a set of revisitation Pin perf features to evaluate the popularity of the saved Pins receiving revisitations in the past 7 days, 30 days, and 90 days. Corresponding to the three types of revisitation labels, we designed three types of revisitation Pin perf features, which we refer as Related Pins-triggered revisitation Pin perf features since we only count +1 if the revisited Pin is saved on the Related Pins surface by the same user. This design is for enhancing the model learning correlations between a Pin’s revisitation popularity in history (features) and its current revisitation performance (labels). We aggregated the features on the 7-day (update every day), 30-day (update every 3 days), and 90-day (update every week) basis for the sake of higher feature coverage.

- Same-day Revisit Impression Pin Perf:  $\sim 60\%$  coverage
- Same-day Revisit Grid-click Pin Perf:  $\sim 49\%$  coverage
- 7-day Revisit Grid-click Pin Perf:  $\sim 51\%$  coverage

In addition, to capture the general popularity of a Pin being revisited, we designed a set of overall revisitation Pin perf features where we do not require the repin happened on the Related Pin surface and do not define the time gap between repin and revisitation.

- Overall Revisit Impression Pin Perf:  $\sim 80\%$  coverage
- Overall Revisit Grid-click Pin Perf:  $\sim 70\%$  coverage

For each Pin perf feature, we include a count of “number of actions” and a count of “unique users” to reflect volume and popularity, respectively. Volume alone could be noise to the system since the Pin revisitation count could be heavily driven by a single enthusiastic user.

### Cross-surface Revisitation Data Pipeline

Figure 7 (right) illustrates the end-to-end user journey on Pinterest. When user lands on Related Pins surface from other surface (such as Homefeed and Search) and grid-clicks a Pin they like, they can save (repin) the Pin to their own profile. Later on, they can go back to their own profile and browse the Pin (revisitation impression); they can also tap on the saved Pin to have a bigger and deeper view of the Pin (revisitation grid-click) where they will re-land on Related Pins with the saved Pin as query Pin and browse the recommended Pins that are to it. Users’ actions on all the candidate Pins in the grid view on the Related Pins surface (staging 2 and 3 with blue color) are logged to a daily generated dataset for training the ranking model for recommendations on the Related Pins surface. However, users’ revisitation actions on their own profiles were not included in the training data. In order to further track users’ revisitation behaviors on their saved Pins, we did a cross-surface join of the existing training data of user actions on the Related Pins surface and data of user actions on their own profile on user ID and Pin ID to map users’ revisitation actions in their own profile to their saved Pins on related Pins surface.

Figure 8 illustrates the revisitation data generation and the final training data generation pipeline.

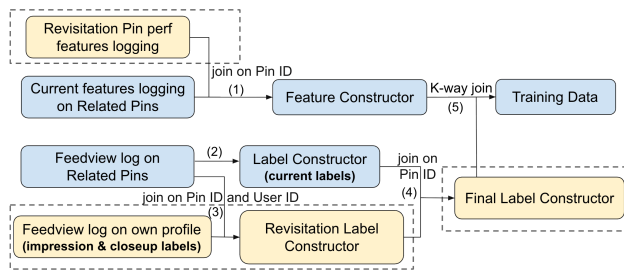


Figure 8: Revisitation Data Generation and Training Data Generation Pipeline

1. Log the revisitation Pin perf features, join them with the other features, and convert raw features log entries into TabularML.
2. Extracts all the user actions on the Related Pins surface from Feedview Log into labels.
3. Cross-surface and cross-session (day) join of users’ save actions on the Related Pins surface and users’ revisitation impressions and revisitation grid-clicks actions on their own profile (join key: User ID and Pin ID, condition:  $time_{save} < time_{revisit} < time_{save} + 7(days)$ ).
4. Join the revisitation labels with the repin labels and on User ID and Pin ID and add a new column for the revisitation label indicating which saved Pin has been revisited in the final label constructor.
5. Join feature TabularML and the final labels (with revisitation label added) on API request id and candidate Pin id.

Step (1) (2) and (3) can be conducted in parallel, followed by step (4) and step (5) sequentially. This solution of mapping user revisitation labels to training data is scalable to any other surface on Pinterest and other platforms where users can save content and revisit the saved content.

### Offline Experiment

We trained the multi-task learning ranking model on the Related Pins surface on Pinterest illustrated in Figure 7 with the designed revisitation labels and revisitation Pin perf features. We used 27 days of data for training, the last day of training data for calibration and the following 3 days of data for evaluation. The size of the entire training data is around 6.6 billion and the size of the evaluation data is around 700 million. We train the model for 1 epoch. We also tuned the utility weight  $u_{RP\&RV}$  for the added revisitation task in Equation (4) and experiment results showed that the model achieved the best gain without tradeoff when  $u_{RP\&RV} = 1.27 * u_{Repin}$ .

Table 1 first shows the offline evaluation results. For ranking metrics per task (head), we observed revisitation head achieved 20%-40% lift and repin head also achieved 0.05%-0.03% lift compared to the current ranking model without revisitation modeling. Because all the revisited Pins were saved Pins (repins), the model was able to prioritize ranking the Pins that may receive more revisitations in the top positions.

Metric	Repin	Revisit
NDCG@3 Lift (%)	<b>0.1279</b>	<b>40.15</b>
MAP@3 Lift (%)	<b>0.06453</b>	<b>43.44</b>
Recip_Rank@3 Lift (%)	-0.02666	<b>27.62</b>
Recall@3 Lift (%)	<b>0.2705</b>	<b>35.4</b>
Pariwise Accuracy Lift (%)	<b>0.10</b>	<b>21.62</b>
Hits@3 Lift (%)	<b>0.59</b>	<b>0.65</b>

Note: Bold numbers indicate metrics gain.

Table 1: Offline Evaluation Results Lift (%)

In addition, we used *Hits@3* to evaluate the overall ranking performance based on the final ranking score defined in Equation (4). For example, we incremented *Hits@3* (*Repin*) if the user saved any of the top 3 recommendations. We tuned the utility weight for the revisitation head in Equation (4) based on *Hits@3* metrics. By modeling revisitation in the ranking model, the model improved the *Hits@3* metrics on repin by 0.59% and revisitation by 0.65% significantly without trade-offs on other heads. Empirical evidence at Pinterest showed that *Hits@3* usually lines up the most with the online A/B experiment among all of the metrics.

## Online A/B Experiment

We conducted a large-scale online A/B experiment from April 29, 2025 to June 26, 2025 on Pinterest Related Pins surface (Liu et al. 2017). The number of users in each group was around 12 million. The experiment ran for 2 months from which we reported the accumulated results.

### Revisitation Online Metrics

Table 2 first shows the three types of revisitation metrics corresponding to the proposed revisitation labels, i.e., 1dRevImpre, 1dRevGrid, and 7dRevGrid. Compared to the control group, the treatment group achieved 0.95% - 1.42% lift on the revisitation grid-clicks metrics. The revisitation grid-clicks metrics gains are higher than the repin gain, which indicating revisitation grid-click rate gain overall after removing the impact of repin gain on revisitation metrics. Notably, the treatment group achieved higher revisitation ratio in terms of both volume and propensity, indicating that the revisitation modeling in the treatment group was driving larger proportion of revisitations from the saved Pins. Compared to the same-day revisitation grid-click metrics, the treatment group achieved higher lift on the 7-day revisitation grid-clicks metrics, which demonstrated strong accumulative effect of revisitation modeling in driving users' longer term revisitation behaviors on the platform.

We also observed that the gain on the same-day revisitation impression is lower than the grid-click based revisitation metrics. This aligns with our previous finding that revisitation grid-clicks drives deeper engagement than just Impression-based revisitation.

Metric Name	Actions Per User (Volume)	Unique Users (Propensity)
1dRevImpre	0.65**	0.93**
1dRevGrid	0.95**	1.17**
7dRevGrid	1.18**	1.42**
Repin	0.94*	0.64***
Sessions $\geq$ 5min	0.41 ***	-0.03
Web Or API Requests	0.35 ***	0.00
Time Spent (RP)	0.53 ***	0.02
Time Spent (OP)	0.68**	0.13
Total Time Spent	0.39 ***	0.01
Active Users	0.10**	0.08**

RP: Related Pins surface, OP: own profile

\*\*\*:  $p < 0.01$ , \*\*:  $p < 0.05$ , \*:  $p < 0.1$

Table 2: Online Experiment Results Lift (%)

### Engagement Metrics and User Retention Metrics

Table 2 also shows the engagement metrics lift on the Related Pins surface between the control group and the treatment group with revisitation modeling. Aligned with the offline metrics, the treatment group achieved better metrics of 0.94% volume gain and 0.64% propensity gain on repin. Table 2 showed that the treatment group with revisitation modeling is driving +0.41% user sessions that are longer than 5 minutes and +0.35% Web or API requests overall on Pinterest.

We also observed from Table 2 that the users in the treatment group spent longer time on the Related Pins surface (+0.53%) as well as their own profile (+0.68%) where revisitations happen, and overall in App (+0.39%) than users in the control group. Given that our model only directly affected the recommendations on the Related Pins surface, the gains on site-wide retention metrics were considered as significantly notable. All these metrics gains contributed to 0.1% volume gain and 0.08% propensity gain on active users.

### Visualization

Based on the online A/B experiment results, we further conducted multiple revisitation visualizations on the 16 main topics of Pins on Pinterest.

**Interpretability** By incorporating revisitation modeling into the multi-task recommendation system (Equation (3)), the model is prioritizing recommending Pins that achieved more saves and revisitations. Pins that received more revisitations historically will tend to receive more repins by the proposed revisitation modeling framework. We calculated the average probability of repin and revisitation  $P(C_{RP\&RV}|q, \theta)$  on each topic for the control group, and the repin volume lift (%) per user of treatment group over control group. Figure 9 shows that the topics that have the highest  $P(C_{RP\&RV}|q, \theta)$  (such as *Beauty*, *Architecture*, and *Entertainment*) generated higher repin volume lift than other topics in the treatment group, while the topics that have the

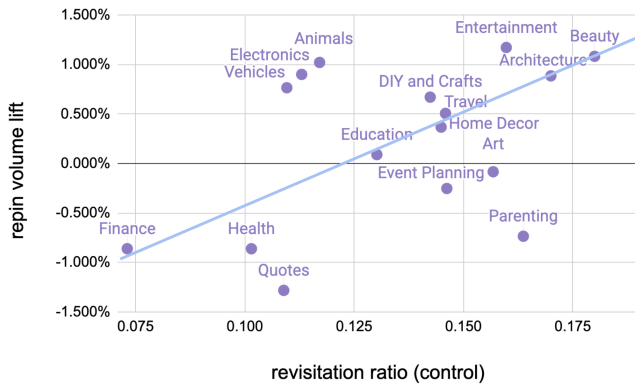


Figure 9: Interpretability (x-axis: revisitatio relative ratio of control group (absolute value hidden while order preserved), y-axis: repin volume lift %)

Topic	repin rate	overall revisit rate	revisit grid-click rate	1-day revisit grid-click rate	3-6 day revisit grid-click rate
DIY and Crafts	Green	Red	Red	Red	Red
Parenting	Green	Red	Red	Red	Red
Entertainment	Green	Red	Red	Red	Red
Health	Green	Red	Red	Red	Red
Home Decor	Green	Red	Red	Red	Red
Beauty	Green	Green	Green	Green	Green
Education	Green	Red	Red	Red	Red
Quotes	Green	Red	Red	Red	Red
Architecture	Green	Green	Green	Green	Green
Travel	Green	Green	Green	Green	Green
Event Planning	Green	Green	Green	Green	Green
Animals	Green	Red	Red	Red	Red
Art	Green	Green	Green	Green	Green
Vehicles	Green	Red	Red	Red	Red
Finance	Green	Red	Red	Red	Red
Electronics	Green	Green	Green	Green	Green

Figure 10: Heatmap of Repin Rate v.s. Revisitation Rate by Topic (Green represents higher rate and red vice versa)

lowest  $P(C_{RP\&RV}|q, \theta)$  (e.g., *Finance*, *Health*, and *Quotes*) incurred more negative repin lift.

**Repin Rate v.s. Revisitation Rate** Figure 10 uses a heatmap to illustrate the average repin rate, average overall revisitation rate, and revisitation grid-click rates for each topic. We observed that topics such as *DIY and Crafts*, *Parenting*, and *Health* which requires more real-life practices received higher repin rates but lower revisitation rates. On the other hands, *Beauty*, *Architecture*, *Travel*, *Event Planning*, *Art*, *Electronics* received lower repin rates but high revisitation rates.

**Long-term v.s. Short-term Revisitation** In Figure 3b, we observed that the volume of revisitation grid-clicks decays dramatically in the first three days. 8% of the saved Pins received a revisitation grid-click after 0-2 days while the ratio dropped to only 1.9% in the next four days (day 3 to day 6). In Table 3, we calculated  $revisit_{3-6day}$ , the ratio of 3-6 day revisitation grid-click volume to the 0-6 day revisitation grid-click volume for each topic, where a higher ratio indicates that the topic tends to be revisited 3-6 days after saving (longer term revisitation interests). The longer term revisita-

Topic	ratio	Topic	ratio
Event Planning	0.1881	Entertainment	0.1639
Health	0.1836	Animals	0.1633
Home Decor	0.1780	Education	0.1602
DIY and Crafts	0.1733	Art	0.1572
Quotes	0.1725	Architecture	0.1495
Beauty	0.1666	Vehicles	0.1334
Parenting	0.1659	Electronics	0.1079
Travel	0.1654	Finance	0.1056

Table 3: Long-term v.s. Short-term Revisitation by Topic

tion interests are *Event Planning*, *Health*, *Home Decor*, and *DIY and Crafts*, which are innately related to long-term user interests. The topics which users have the most short-term revisitation interest are *Finance*, *Electronics*, *Vehicles*, *Architecture*, and *Art*.

## Conclusion

This work introduces a revisitation modeling framework in a multi-task learning recommender system, using Pinterest as a test bed. By analyzing user save (repin) actions and their cross-session, cross-date revisitation behaviors, we establish causal links between saved Pins and user retention. Our approach encompasses comprehensive behavior analysis, label and feature design for revisitation, and demonstrates scalability to other online platforms. Differing from traditional recommender systems that rely on in-session actions, our model leverages cross-session, cross-date, and cross-surface revisitation data, aiming to enhance long-term user retention and core engagement metrics, as validated through both of-line and large-scale online experiments.

While our results show the framework effectively drives engagement, revisitation, and retention metrics, we identified a bottleneck: highly active users contribute disproportionately by repinning more than less active users, leading to limited revisitation label data from the latter group. To address this, future work could broaden revisitation actions beyond explicit repin events to include implicit signals, such as users interacting with Pins on similar topics across consecutive sessions. Overall, our work offers a novel, scalable framework for estimating and optimizing user revisitation and long-term retention.

## Acknowledgments

We would like to acknowledge our cross-team colleagues **Jing Zhang** and **Yinuo Liu** for their invaluable contributions to the productionization of the revisitation labels and revisitation features, respectively.

We are grateful to **Haoran Guo** and **Liyao Lu** for their excellent guidance in shipping this work to production. We also thank **Shivin Thukral**, **Tianyu Feng**, **Wendy Shao**, **Jinfeng Rao**, **Lu Liu**, **Jinyu Xie**, **Matt Chun**, **Michael Chau**, **Karim Wahba**, and **Rajat Raina** for their generous support throughout this work.

## References

- Adar, E.; Teevan, J.; and Dumais, S. T. 2008. Large scale analysis of web revisitation patterns. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 1197–1206.
- Adar, E.; Teevan, J.; and Dumais, S. T. 2009. Resonance on the web: web dynamics and revisitation patterns. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 1381–1390.
- Afsar, M. M.; Crump, T.; and Far, B. 2022. Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys*, 55(7): 1–38.
- Badrinath, A.; Yang, A.; Rajesh, K.; Agarwal, P.; Yang, J.; Chen, H.; Xu, J.; and Rosenberg, C. 2025. OmniSage: Large Scale, Multi-Entity Heterogeneous Graph Representation Learning. *arXiv preprint arXiv:2504.17811*.
- Cai, Q.; Liu, S.; Wang, X.; Zuo, T.; Xie, W.; Yang, B.; Zheng, D.; Jiang, P.; and Gai, K. 2023. Reinforcing user retention in a billion scale short video recommender system. In *Companion Proceedings of the ACM Web Conference 2023*, 421–426.
- Dang, Y.; Zhang, J.; Liu, Y.; Yang, E.; Liang, Y.; Guo, G.; Zhao, J.; and Wang, X. 2025. Augmenting Sequential Recommendation with Balanced Relevance and Diversity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 11563–11571.
- Ding, R.; Xie, R.; Hao, X.; Yang, X.; Ge, K.; Zhang, X.; Zhou, J.; and Lin, L. 2023. Interpretable user retention modeling in recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, 702–708.
- Gong, S.; Liu, Y.; Dang, Y.; Guo, G.; Zhao, J.; and Wang, X. 2025. Multiple Purchase Chains with Negative Transfer Elimination for Multi-Behavior Recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 11717–11725.
- Gugnani, A.; and Misra, H. 2020. Implicit skills extraction using document embedding and its use in job recommendation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 13286–13293.
- Guo, X.; Wang, S.; Zhao, H.; Diao, S.; Chen, J.; Ding, Z.; He, Z.; Lu, J.; Xiao, Y.; Long, B.; et al. 2022. Intelligent online selling point extraction for e-commerce recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 12360–12368.
- Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30.
- Jia, J.; Wang, Y.; Li, Y.; Chen, H.; Bai, X.; Liu, Z.; Liang, J.; Chen, Q.; Li, H.; Jiang, P.; et al. 2025. LEARN: Knowledge Adaptation from Large Language Model to Recommendation for Practical Industrial Application. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 11861–11869.
- Jin, L.; Feng, L.; Liu, G.; and Wang, C. 2017. Personal web revisitation by context and content keywords with relevance feedback. *IEEE Transactions on Knowledge and Data Engineering*, 29(7): 1508–1521.
- Kislyuk, D.; Liu, Y.; Liu, D.; Tzeng, E.; and Jing, Y. 2015. Human curation and convnets: Powering item-to-item recommendations on pinterest. *arXiv preprint arXiv:1511.04003*.
- Kwon, H.; Han, J.; and Han, K. 2020. ART (Attractive Recommendation Tailor) How the Diversity of Product Recommendations Affects Customer Purchase Preference in Fashion Industry? In *Proceedings of the 29th ACM international conference on information & knowledge management*, 2573–2580.
- Liu, D. C.; Rogers, S.; Shiau, R.; Kislyuk, D.; Ma, K. C.; Zhong, Z.; Liu, J.; and Jing, Y. 2017. Related pins at pinterest: The evolution of a real-world recommender system. In *Proceedings of the 26th international conference on world wide web companion*, 583–592.
- Liu, Z.; Liu, S.; Yang, B.; Xue, Z.; Cai, Q.; Zhao, X.; Zhang, Z.; Hu, L.; Li, H.; and Jiang, P. 2024. Modeling User Retention through Generative Flow Networks. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 5497–5508.
- Ma, J.; Zhao, Z.; Yi, X.; Chen, J.; Hong, L.; and Chi, E. H. 2018. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 1930–1939.
- Pancha, N.; Zhai, A.; Leskovec, J.; and Rosenberg, C. 2022. Pinnerformer: Sequence modeling for user representation at pinterest. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 3702–3712.
- Sun, R.; Kong, R.; Milton, A.; Kluver, D.; Paterson, I.; and Konstan, J. A. 2025. Why They Come And Go: A Case Study of Productive Flyby Users and Their Rating Integrity Challenge in Movie Recommenders. In *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval*, 1–11.
- Wang, C.; Kim, L.; Bang, G.; Singh, H.; Kociuba, R.; Pomerville, S.; and Liu, X. 2020. Discovery news: a generic framework for financial news recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 13390–13395.
- Wang, R.; Shivanna, R.; Cheng, D.; Jain, S.; Lin, D.; Hong, L.; and Chi, E. 2021. Dcn v2: Improved deep & cross network and practical lessons for web-scale learning to rank systems. In *Proceedings of the web conference 2021*, 1785–1797.
- Wang, Y.; Sharma, M.; Xu, C.; Badam, S.; Sun, Q.; Richardson, L.; Chung, L.; Chi, E. H.; and Chen, M. 2022. Surrogate for long-term user experience in recommender systems. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 4100–4109.
- Zhai, A.; Wu, H.-Y.; Tzeng, E.; Park, D. H.; and Rosenberg, C. 2019. Learning a unified embedding for visual search at pinterest. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2412–2420.
- Zhang, Y.; Wang, D.; Li, Q.; Shen, Y.; Liu, Z.; Zeng, X.; Zhang, Z.; Gu, J.; and Wong, D. F. 2021. User Retention:

A Causal Approach with Triple Task Modeling. In *IJCAI*, 3399–3405.

Zhang, Y.; and Yang, Q. 2021. A survey on multi-task learning. *IEEE transactions on knowledge and data engineering*, 34(12): 5586–5609.

Zhao, K.; Zou, L.; Zhao, X.; Wang, M.; and Yin, D. 2023. User retention-oriented recommendation with decision transformer. In *Proceedings of the ACM Web Conference 2023*, 1141–1149.