Efficiently Maintaining Next Basket Recommendations under Additions and Deletions of Baskets and Items

Benjamin Longxiang Wang  
University of Amsterdam & ABN AMRO Bank  
wlongxiang1119@gmail.com

Sebastian Schelter  
AIRLab, University of Amsterdam  
s.schelter@uva.nl

ABSTRACT
Recommender systems play an important role in helping people find information and make decisions in today’s increasingly digitalized societies. However, the wide adoption of such machine learning applications also causes concerns in terms of data privacy. These concerns are addressed by the recent “General Data Protection Regulation” (GDPR) in Europe, which requires companies to delete personal user data upon request, when users enforce their “right to be forgotten”. Many researchers argue that this deletion obligation does not only apply to the data stored in primary data stores such as relational databases, but also requires an update of machine learning models whose trainingset included the personal data to delete. As a consequence, the academic community started to investigate how to unlearn user data from trained machine learning models in an efficient and timely manner.

We explore this direction in the context of a sequential recommendation task called Next Basket Recommendation (NBR), where the goal is to recommend a set of items based on a user’s purchase history. We design efficient algorithms for incrementally and decrementally updating a state-of-the-art next basket recommendation model in response to additions and deletions of user baskets and items. Furthermore, we discuss an efficient, data-parallel implementation of our method in the Spark Structured Streaming system.

We evaluate our implementation on a variety of real-world datasets, where we investigate the impact of our update techniques on several ranking metrics and measure the time to perform model updates. Our results show that our method provides constant update time efficiency with respect to an additional user basket in the incremental case, and linear efficiency in the decremental case where we delete existing baskets. With modest computational resources, we are able to update models with a latency of around 0.2 milliseconds regardless of the history size in the incremental case, and less than one millisecond in the decremental case.

1 INTRODUCTION
Recommender systems (RS) are ubiquitous in our digitalized societies with applications in e-commerce, social networks and streaming services [32]. However, the wide adoption of such machine learning applications has been causing concerns in terms of fairness, accountability, privacy, and transparency in the general public and research community alike [5, 30, 38]. It is widely recognised that we need to tackle such responsibility issues at the intersection of machine learning and data management [19, 27, 29, 31].

The right to be forgotten. In this work, we focus on one specific aspect of privacy stipulated by the General Data Protection Regulation (GDPR), namely, “the right to be forgotten” [7], which states that the “data subject shall have the right to […] the erasure of personal data concerning him or her without undue delay […] where the data subject withdraws consent”. When users can enforce their right to be forgotten, companies must delete their personal user data. Many researchers argue that this deletion obligation does not only apply to the data stored in primary data stores such as relational databases, but also requires an update of machine learning models whose trainingset included the personal data to delete. Legal scholars even point out the continual use of AI systems trained on deleted data could be considered illegal under certain interpretations [34]. In some cases, the retained lossy representations in ML models can even be restored by adversarial parties [26].

Towards timely and efficient unlearning. A naive way to realize data deletion is to retrain the model from scratch on the remaining dataset, leaving out the user data to delete. However, this approach is typically expensive in terms of time and computational resources. Furthermore, the GDPR mentions companies should erasure data “without undue delay” [7] but does not directly specify when deletions should be fully conducted. In practice, data deletion is still a lengthy process. For example, Google could take two to six months to fully delete someone’s data 1; for Facebook, this period is 90 days 2. We argue that it is an open and important academic question to determine how fast data can be erased from recommendation models, and that we should design algorithms and systems to empower users to exercise their right to be forgotten as timely as possible, instead of making them wait for several months.

Efficiently maintaining a TIFU-kNN model for next basket recommendation. In this paper, we explore the direction of efficiently updating a recommendation model in the context of a sequential recommendation task called “Next Basket Recommendation” (NBR, Section 2.1), where the goal is to recommend a set of items based on a user’s purchase history. We focus our efforts on TIFU-kNN (Temporal Item Frequency-based User-kNN) [13], a nearest neighbor-based state-of-the-art next basket recommendation model (Section 2.2), which outperforms neural networks in the NBR task, despite its simplicity. We design efficient algorithms for incrementally and decrementally updating a state-of-the-art next basket recommendation model in response to additions and deletions of user baskets and items. Provided the training data has remained unchanged, the TIFU-kNN model can be updated in an efficient manner.

1https://policies.google.com/technologies/retention
2https://www.facebook.com/help/250563911970368
deletions of user baskets and items.

We formalize the problem of maintaining a TIFU-kNN model under additions and deletions of baskets and items in Section 3. Next, we design efficient online algorithms that extend TIFU-kNN to learn incrementally as new data arrives (incremental updates Section 4.2) and to forget efficiently as deletion requests come in (decremental updates Section 4.3). We implement our approach in a data-parallel manner in the Apache Spark Structured Streaming system [37] (Section 5). Furthermore, we conduct extensive experiments in Section 6 where we analyze the recommendation performance and update efficiency of our approach, and point out its limitations for an extreme number of deletions in Section 6.3. In summary, this paper provides the following contributions.

- We introduce the problem of incremental/decremental learning for next basket recommendation (Section 3).
- We design efficient online algorithms for incremental and decremental updates of the state-of-the-art NBR model TIFU-kNN (Section 4).
- We discuss a data-parallel implementation of our approach in the Spark Structured Streaming system (Section 5).
- We conduct experiments on three real-world datasets, where we evaluate our approach in terms of predictive performance and update time efficiency. We find that we are able to update the model with constant time efficiency in the incremental case and linear time efficiency in the decremental case with millisecond latency (Section 6).

2 BACKGROUND

2.1 Next Basket Recommendation

We focus on a specific recommendation task called Next Basket Recommendation (NBR), where we aim to recommend a set of items to users based on their historical purchases [25]. NBR is of great interest to the e-commerce and retail industry, where we want to recommend a set of items to fill a user’s shopping basket. A basket \( b \) is a set of items, i.e., \( b = \{i_1, i_2, \ldots, i_j, \ldots, i_h\} \), where \( i_j \in I \), and where \( I \) denotes the universe of all items. For a given user, \( b \) has access to a sequence of \( n \) historical baskets (in increasing chronological order, such that more recent items are at the tail) denoted as \( \mathcal{H} = [b_1, b_2, \ldots, b_i, \ldots, b_n] \), where \( b_1 \in I \). The goal of NBR is then to find a model which takes the historical baskets \( \mathcal{H} \) as input and predicts the next basket \( b_{n+1} \) as recommendation.

2.2 TIFU-kNN

Our work centers around adapting the state-of-the-art next basket recommendation model TIFU-kNN [13]. This method builds on the user-based collaborative filtering approach [17]. A target user is first modeled by an embedding vector, and then recommendations are made based on a weighted combination of the aggregated representations of similar users and the target user’s own vector embedding. TIFU-kNN calculates user vector representations during the training phase leveraging repeated purchase patterns [1, 2] commonly observed in e-commerce, under the assumption that recent purchases have more predictive power than older ones.

User vector representation. A user vector is calculated to represent a user’s past purchase patterns, with a time-decayed window to assign a higher weight to recent items. However, a single time decayed weight is not flexible enough to model another property of temporal dynamics: consecutive steps have small changes while steps far from each other have large changes. Therefore, TIFU-kNN applies a hierarchical structure where user baskets are first divided into groups, baskets within a group are aggregated with time decayed weights to obtain a group vector, and afterwards, these group vectors are aggregated again with time decayed weights to obtain a final user vector. In detail, the algorithm applies the following steps:

Step 1: Multi-hot encoding. Without loss of generality, we denote the historical baskets of our target user as \( \mathcal{H} = [b_1, \ldots, b_i, \ldots, b_n] \) where \( b_i \in I \) and \( |\mathcal{H}| = n \). Each basket \( b_i \) is converted into a multi-hot encoded vector whose dimension is equal to \( |I| \) with ones in dimensions for items present in the basket, and zeros in the remaining positions. For example, suppose we have a corpus containing four items, i.e., \( I = \{1, 2, 3, 4\} \), and a user’s history with two baskets \( \mathcal{H} = [(1, 4), (1, 2, 3)] \). The corresponding multi-hot vector would then look as follows.

\[
[(1, 4), (1, 2, 3)] \xrightarrow{\text{multi-hot encoding}} \left\{ [1, 0, 0, 1], [1, 1, 1, 0] \right\}
\]

Step 2: Group vector generation. To model the user’s behavior shift over time, TIFU-kNN utilizes a hierarchical grouping mechanism to capture the dynamics. Given a fixed group size \( m \) as a hyperparameter, the number of groups is denoted as \( k = \left\lceil \frac{n}{m} \right\rceil \) where \( \lceil \cdot \rceil \) is the ceiling function. The group size dictates the number of baskets in a group. Given a temporal decay rate \( 0 < \tau_b \leq 1 \) and baskets partitioned into groups of equal length (except for the last group which contains the remaining baskets in case \( n \) is not fully divisible by \( m \)), a group division example looks as follows.

\[
\mathcal{H} = [b_1, \ldots, b_m, b_{m+1}, \ldots, b_{2m}, \ldots, b_{n-1}, b_n]
\]

Each group is represented by a group vector \( v_g \) which is calculated by a time-decayed weighted average, where the \( i \)-th vector within a group is multiplied by a time-decayed weight \( r_i^{m-1} \), where \( r_i \) is a multi-hot encoded vector. Note that the indicator for the group index is left out for brevity.

\[
v_g = \frac{r_1^{m-1}v_{b_1} + r_2^{m-2}v_{b_2} + \cdots + v_{b_m}}{m} = \sum_{i=1}^{m} r_i^{m-1}v_{b_i}
\]

Step 3: User vector generation. Similar to how a group vector is calculated from basket vectors, TIFU-kNN calculates the final target user vector \( v_u \) based on a time-decayed weighted average of the group vectors \( \mathcal{H} = [v_{g_1}, v_{g_2}, \ldots, v_{g_k}] \). Given the group decay rate \( 0 < \tau_g \leq 1 \), we have

\[
v_u = \frac{r_1^{k-1}v_{g_1} + r_2^{k-2}v_{g_2} + \cdots + v_{g_k}}{k} = \sum_{i=1}^{k} r_i^{k-1}v_{g_i}
\]

Personalized Collaborative Filtering Recommendation. After obtaining the vector representations of all users, TIFU-kNN generates recommendations inspired by collaborative filtering (CF) [17]. Given a predefined distance measure, it finds the top-k most similar
users for each target user. The final prediction combines two parts: (i) the personalized component, e.g., the user vector representation denoted by \( \mathbf{u}_{\text{target}} \), corresponding to the user’s temporal repeated purchase pattern; (ii) the collaborative filtering component, which is calculated as the mean of the nearest neighbors’ vector representations, denoted by \( \mathbf{u}_{\text{neighbors}} \).

The final prediction vector is calculated via a linear combination \( \mathbf{p} = \alpha \mathbf{u}_{\text{target}} + (1 - \alpha) \mathbf{u}_{\text{neighbors}} \) of both components, where \( \alpha \) is a hyper-parameter used to adjust the weight of each component. Note that the numeric values in the prediction vector \( \mathbf{p} \) indicate the strength of preference of the user towards a certain item \( \mathbf{p} \), and the highest scored items are typically leveraged for recommendation.

3 PROBLEM STATEMENT

As outlined in the introduction, we focus on two aspects of the maintenance of TIFU-kNN models: (i) how to incrementally learn user representations in response to additional baskets (incremental updates); (ii) how to efficiently delete user data from the learned user representations (decremental updates).

Let \( \mathcal{D} = \{ \mathcal{H}_1, \mathcal{H}_2, \cdots, \mathcal{H}_i, \cdots, \mathcal{H}_m \} \) denote a dataset containing historical baskets for \( M \) users, and the corresponding user vectors are \( \mathcal{U} = \{ \mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_i, \cdots, \mathbf{u}_m \} \) where \( \mathbf{u}_i \) is a real-valued vector of size \( |I| \). Our goal is to maintain the user vector of a TIFU-kNN model. As each user vector is modeled independently (e.g., only calculated from the baskets of the particular user), we consider only a single user in the following.

Constant time incremental updates. Given the historical baskets \( \mathcal{H} = [b_1, b_2, \cdots, b_i, \cdots, b_n] \) and a newly arriving basket \( b_{n+1} \) for user \( \mathbf{u} \), we can always recalculate the corresponding user vector from the complete history. For incremental updates, we want to improve the asymptotic time complexity from \( O(|\mathcal{H}|) \) for retraining from scratch to a constant effort \( O(1) \) per additional basket. Given the history \( \mathcal{H} \) and the current user vector \( \mathbf{u}_n \), we are interested in an efficient approach \( f_{\text{incr}}(\mathbf{u}_n, b_{n+1}) \) to update the current user vector \( \mathbf{u}_n \) to \( \mathbf{u}_n' \) by incorporating \( b_{n+1} \).

Efficient decremental updates. In the incremental case, given the current user vector \( \mathbf{u}_n \) and a deletion request \( r \) (identifying a basket or item to be deleted), we aim to efficiently delete \( r \) from the corresponding user vector \( \mathbf{u}_n \). We want to improve the asymptotic time complexity from \( O(|\mathcal{H}|) \) for retraining from scratch to \( O(|\mathcal{H}| - p) \) where \( p \) is the position of the basket where a deletion is applied. Formally, we look for a decremental update function \( f_{\text{decr}}(\mathcal{H}, \mathbf{u}_n, r) \) which takes three inputs: the historical baskets \( \mathcal{H} \), the existing user vector \( \mathbf{u}_n \) and the deletion request \( r \), and outputs an updated user vector \( \mathbf{u}_n' \). We want \( f_{\text{decr}} \) to satisfy \( \forall \mathbf{u}_n, \forall r : f_{\text{decr}}(\mathcal{H}, \mathbf{u}_n, r) = \mathbf{u}_n' \), e.g., we want the result of a decremental update to be the same as computing the user vector after deletion \( \mathbf{u}_n' \) by running TIFU-kNN from scratch on \( \mathcal{H} \setminus r \).

4 APPROACH

As mentioned in Section 3, we compare our approach to the baseline approach of retraining the whole model from scratch whenever there is an incremental or decremental update. This baseline has an asymptotic complexity of \( O(|\mathcal{H}|) \) per update, be it incremental or decremental. This baseline has several disadvantages: (i) the (re)training phase is expensive in terms of time and resources, especially if it has to be executed repeatedly for a small number of data deletion requests (ii) the user has to potentially wait long for their data to be actually deleted from the model if the retraining is only executed periodically.

To overcome these challenges, we design online algorithms which update the model with low latency in response to additional data or deletion requests concerning baskets or items. We first discuss general rules for maintaining the decaying average of a series in Section 4.1 and then apply them to derive our approach to incremental updates in Section 4.2 and decremental updates in Section 4.3 under different scenarios.

4.1 Maintaining the Decaying Average of a Series

Calculating decaying averages of vector series is at the heart of TIFU-kNN (Section 2.2) as it is the core computation for both the group vector (Equation 1) and user vector (Equation 2). Without loss of generality, we discuss this problem with a series of real numbers. Let \( S = \{x_1, \cdots, x_n\} \) denote a series of real numbers, where the next item is \( x_{n+1} \). Let \( \bar{x}_n \) denote the decaying average of the first \( n \) numbers. Given \( 0 < r \leq 1 \), the decaying average of the series is \( \bar{x}_n = \frac{\sum_{i=1}^{n} r^{n-i} x_i}{n} \). We examine how the decaying average changes in response to incremental and decremental updates. Note that it is straightforward to extend our approach to a series of vectors of real numbers by replacing the enclosed elements by vectors of the same length, which essentially leads to the group vector (Equation 1) and user vector (Equation 2) definitions of TIFU-kNN. We now look at three scenarios: incremental updates, decremental updates, and in-place updates.

Incrementally adding a number. With a newly arriving item \( x_{n+1} \), we update the corresponding decaying average as follows.

\[
\bar{x}_{n+1} = \frac{1}{n+1} \sum_{i=1}^{n+1} r^{n+1-i} x_i
\]

\[
= \frac{r^n x_1 + r^{n-1} x_2 + \cdots + r x_n + x_{n+1}}{n+1}
\]

\[
= \frac{r n \bar{x}_n + x_{n+1}}{n+1} \tag{3}
\]

With this update rule, we can efficiently update the decaying average in constant time using only the current decaying average \( \bar{x}_n \), the count of items \( n \), the additional item \( x_{n+1} \) and the decay rate \( r \).

Decrementally deleting a number. If an existing number \( x_i \) needs to be deleted from the series, we would like to find an efficient way to update the decaying average without resorting to a full recalculation. Before deletion, the decaying average is:

\[
\bar{x}_n = \frac{r^{n-1} x_1 + \cdots + r^{n-(i-1)} x_{i-1} + r^{n-i} x_i + r^{n-(i+1)} x_{i+1} + \cdots + x_n}{n}
\]

After deletion, the updated decaying average becomes:

\[
\bar{x}_{n-1}' = \frac{r^{n-2} x_1 + \cdots + r^{n-(i-1)} x_{i-1} + r^{n-i} x_{i+1} + \cdots + x_n}{n-1}
\]

\[
\bar{x}_{n-1} = \frac{r^{n-2} x_1 + \cdots + r^{n-(i-1)} x_{i-1} + r^{n-i} x_i + r^{n-(i+1)} x_{i+1} + \cdots + x_n}{n-1}
\]
We rearrange the terms such that we express $\bar{x}_{n-1}'$ using $\bar{x}_n$:

$$\bar{x}_{n-1}' = \frac{n\bar{x}_n + \{r^{n-1}(x_{n+1} - x_1) + \cdots + r(x_n - x_{n-1}) - x_n\}}{(n-1)r}$$

$$= \frac{n\bar{x}_n + \mathcal{D}(\{x_1, \cdots, x_n\})\mathcal{R}(r, n-1)}{(n-1)r}$$

(4)

where $\mathcal{D}(.)$ denotes a vector of the first order difference of the slice of the series after the deleted item: $\mathcal{D}(\{x_1, \cdots, x_n\}) = \{x_{i+1} - x_i, \cdots, x_{n-1} - x_n\}^T$ and $\mathcal{R}(r, n-1) = \{r^{n-1}, \cdots, r, 1\}^T$. We discuss a few observations:

- The decremental update rule (Equation 4) is mathematically symmetrical to its incremental counterpart (Equation 3), as in both cases we update by incorporating the previous value and the change introduced by the additional or removed item. This view provides a unified perspective to deal with incremental and decremental updates.

- To perform a deletion, it is sufficient to access a slice starting from the deleted item. In the worst case when the deleted item is the very first item $x_1$, the slice is equivalent to the full series; in the best case when the deleted item is the very last item, we only need to access a single item. Compared to a full recalculation where (re)access to the full history is needed, our update rule reduces the amortized number of items impacted by a deletion from $|S|$ to $|S|/2$, assuming a uniform distribution of deletions.

- Lastly, the term $\mathcal{D}(.)^T\mathcal{R}(.)$ is a dot product and can be efficiently computed by vectorization, which further improves the efficiency of decremental updates.

**Updating an existing number.** Suppose $x_i$ is updated to a new value $x_i'$, then the new decaying average becomes

$$\bar{x}_n' = \frac{r^{n-1}x_1 + \cdots + r^{n-1}x_i' + \cdots + x_n}{n}$$

$$= \frac{r^{n-1}x_1 + \cdots + r^{n-1}x_i' + \cdots + x_n + r^{n-1}x_i - r^{n-1}x_i}{n}$$

$$= \frac{n\bar{x}_n + r^{n-1}(x_i' - x_i)}{n}$$

(5)

Note that for an in-place update, we only need to update the previous decaying average with a term dependent on the difference introduced by the updated item.

The approaches to update decaying averages for the incremental, decremental, and in-place scenarios serve as a basis for the online algorithms for TIFU-kNN developed in the following subsections.

### 4.2 Incrementally Adding New Baskets

Given an existing user vector $\vec{u}_u$ and an additional basket $\vec{b}_{m+1}$, we discuss an algorithm to update the user vector efficiently based on the rules discussed in Section 4.1. Note that it is sufficient just to study the problem for a single user as all user vectors are computed independently. Due to the grouping mechanism in TIFU-kNN (Section 2), adding a basket to the history will either lead to 1) a new group with a single basket (if the last existing group is fully occupied) or 2) an update of the last existing group with one new basket appended. This is best illustrated with an example. Consider a user with four baskets, group size $m = 2$, hence number of groups $k = \left[\frac{4}{2}\right] = 2$ which leads to two groups with each group containing two baskets. Now we add two new baskets one by one. The group composition evolves as follows.

A group with a single basket is created after adding a new basket to a fully occupied group, and afterwards, the group is updated with the second additional basket. Our update mechanisms differentiate these two scenarios:

**Scenario 1: Adding a group with a single basket**. When transitioning from $\mathcal{H}$ to $\mathcal{H}'$ (Equation 6), we add an additional group vector (with a single basket) to update the user vector while the existing group vectors remain unchanged. Recall that we work with a group size $m$, a number of baskets $n$ and a number of groups $k = \left[\frac{n}{m}\right]$. Since the new group only contains the newly added basket, the new group vector is equivalent to the new basket vector $\vec{g}_1(k+1) = \vec{b}_{m+n+1}$. Given the additional group vector $\vec{g}_k(k+1)$, the existing user vector $\vec{u}_u$ as defined in Equation 2, and the group decay rate $r_g$, we obtain the updated user vector $\vec{u}_u'$ by applying the incremental rule derived in Equation 3.

$$\vec{u}_u' = \frac{r_g\vec{u}_u + r_g^{-1}\vec{u}_u + \cdots + r_g\vec{u}_gk + \vec{u}_g(k+1)}{k + 1}$$

(7)

Note that we can efficiently update the user vector in constant time regardless of the size of the user’s history with access to the current user vector $\vec{u}_u$ and the number of groups $k$.

**Scenario 2: Updating an existing group.** When transitioning from $\mathcal{H}'$ to $\mathcal{H}''$ (Equation 6), we have $r < m$ where $r$ is the number of baskets in the last group and $m$ is the group size. In this scenario, the last group vector $\vec{g}_k$ needs to be updated to take the added basket into account. Given the number of existing baskets in the last group $r$, the basket decay rate $r_b$, the additional basket denoted as $\vec{b}_{r+1}$, we update the group vector $\vec{g}_k$ by applying the incremental rule derived in Equation 3.

$$\vec{g}_k' = \frac{r_b\vec{g}_b + r_b^{-1}\vec{g}_b + \cdots + r_b\vec{g}_b + \vec{g}_b(r+1)}{r + 1}$$

(8)

Analogous to the previous scenario, we only need access to the current group vector $\vec{g}_k$ and the number of baskets in the last group $r$ in order to obtain the updated last group vector efficiently in constant time. Finally, with the updated group vector $\vec{g}_k'$, the target user vector after addition $\vec{u}_u'$ is recomputed with access to the last user vector $\vec{u}_u$, the last group vector $\vec{g}_k$ and the number
of groups $k$ by applying the in-place update rule in Equation 5.

$$
\forall u = r_g^{k-1}v_g + r_g^{k-2}v_g + \ldots + r_gv_g(k-1) + v'_g \\
= v_u + \frac{v'_g - v_gk}{k} \tag{9}
$$

### 4.3 Decrementally Removing Existing Baskets and Items

First, we discuss how a user vector will change if a user requests to delete a specific basket from her basket history. In practical applications, this could happen when a customer on an e-commerce website requests to delete a shopping session for privacy reasons or when a customer made a return. It is sufficient just to look at one basket because we can handle the removal of multiple baskets by applying our algorithm repeatedly. Subsequently, we discuss how to remove a single item from a basket instead of the basket as a whole, which corresponds to a scenario where a customer requests to remove certain sensitive items from a previous shopping cart. Again, we can handle these cases via the update rules discussed in Section 4.1.

**Varying group size.** Before diving into the algorithmic details, we first discuss a relaxation of the original algorithm to which we refer as varying group size. Given a user’s historical baskets, if we were to maintain a fixed group size, removing a basket at any location would cause cascading changes in terms of group composition for any subsequent groups following the deleted basket. For users with a large number of baskets, it quickly becomes impractical to regroup and recalculate all group vectors, especially when the basket under deletion is at the beginning. As a consequence, we propose an alternative to avoid expensive regrouping while still maintaining the temporal patterns with recency effects. We allow for a varying group size, and thereby restrict the regrouping impact to the enclosing group of the to-be-deleted basket. We demonstrate how the group composition changes under a varying group size as follows.

$$
\begin{align*}
\forall g = [v_{g1}, v_{g2}, v_{g3}, v_{g4}, v_{g5}, v_{g6}, v_{g7}, v_{gb}] &\iff v_{gb} \\
\forall g = [v_{gb}, v_{g2}, v_{g3}, v_{g4}, v_{g5}, v_{g6}, v_{g7}, v_{gb}] &\iff v_{gb} \\
\forall g = [v_{gb}, v_{g2}, v_{g3}, v_{g4}, v_{g5}, v_{g6}, v_{g7}, v_{gb}] &\iff v_{gb}
\end{align*}
$$

where we show that after removing baskets for $v_{g3}$ and $v_{g6}$, the impacted groups are limited to minimal to $v'_{g2}$ and $v'_{g3}$ respectively. We differentiate three different deletion scenarios:

**Scenario 1: Deleting a basket from an existing group.** If the enclosing group contains more than one basket, deleting a basket from it leads to an updated group vector. Given the to-be-deleted basket vector denoted as $v_{bh}$, the current group vector $v_g$, the basket decay rate $r_b$ and the number of baskets in the enclosing group $\tau$, we can apply the decremental rule (Equation 4) to obtain the updated group vector,

$$
\forall g = r_bv_g + D([v_{bh}, \ldots, v_{bt}])^T R(r_b, \tau - i)
$$

where we have $D([v_{bh}, \ldots, v_{bt}]) = [v_{bh} - v_{bt}, \ldots, v_{bh} - v_{bt}, \ldots, v_{bt}]^T$ and $R(r_b, \tau - i) = [r_b^{\tau-i}, \ldots, r_b]^T$. Similar to the observations for the decremental rule (Equation 4), we highlight one important property of Equation 10: we only need to access a slice starting from the deleted basket instead of the full history to perform a deletion. Compared to a full retraining, we reduce the amortized number of baskets impacted by a deletion from $|H|$ to $|H|/2$ (assuming a uniform distribution of deletes). Finally, following the in-place update rule (Equation 5), we compute the updated user vector $u'_u$ from the updated group vector.

$$
\forall u = u + \frac{k}{k} (v'_g - v_gk) \tag{11}
$$

**Scenario 2: Deleting a single-basket group.** We consider the scenario where a group with a single basket is removed, and the corresponding group vector $v_g$ will vanish, so we only need to update the user vector $u$. Given the number of groups $k$, the vanished group vector $v_g$, the group decay rate $r_g$, the current user vector $u_v$, we update the user vector by applying the decremental rule (Equation 4) again, where we have $D([v_{gb}, \ldots, v_{gb}]) = [v_g - v_g, \ldots, v_g - v_g, \ldots, v_g]^T$ and $R(r_g, k-i) = [r_g^{k-i}, \ldots, r_g]^T$.

$$
\forall u = \frac{k}{k} (v'_u - v_u) \tag{12}
$$

**Scenario 3: Removing a single item from a basket.** Instead of requesting to delete a whole basket, one can also delete an item within a basket, and update the corresponding user vector. Deleting an item for a target basket vector $v_{bi}$ will either 1) result in an updated basket vector by changing the corresponding item dimension from one to zero in the basket vector or 2) cause the basket vector to vanish if it only contained the item to remove. In the first case, given the updated basket vector $v'_{bi}$, the current group vector $v_g$, the number of baskets in the group $r$ and decay rate $r_g$, we update the group vector by applying the in-place update rule (Equation 5) again.

$$
\forall g = v_g + \frac{r_g^{i} (v'_{bi} - v_{bi})}{\tau} \tag{13}
$$

Now that we have $v'_g$, we can apply Equation 11 to obtain the updated user vector. In the second case where the target basket vanishes, we fall back to the previous scenario 1 or 2 depending on if the vanished basket also causes a vanished group.

### 5 IMPLEMENTATION

In the following, we discuss a data-parallel implementation of our approach in a popular stream processing system. We briefly show the system overview, then discuss how to leverage the parallel nature of TIFU-kNN, and how to combine incremental and decremental updates to fulfill low-latency requirements. Our implementation is conducted in Scala in Apache Spark’s Structured Streaming abstraction [37]. Spark Structured Streaming allows us to express a streaming computation as a state update based on data changes.
The Spark engine takes care of running it incrementally and continuously, and of updating the result as new streaming data continues to arrive.

**System overview.** The system (Figure 1) consists of two stages: 1) the training phase where user baskets and deletion requests are processed and user vectors are produced or updated; 2) the prediction phase where the target user vector and its neighborhood vectors are combined to make recommendations.

**Parallelism.** Since each user vector is calculated independently, we can leverage this fact to achieve parallelism on the user level, which in turn makes our implementation scale to a large number of users for practical usage. We leverage the mapGroupsWithState API in Apache Spark for this purpose, because this operation allows us to apply user-defined code on grouped datasets to update user-defined state. We treat each user’s basket stream as a dataset grouped by a unique user identifier, and the user vectors are stored in a state store.

**Joint incremental and decremental state updates.** The incremental algorithm and the decremental algorithm in the streaming engine are jointly responsible for updating the user vectors while the former processes incoming user data and the latter handles deletion requests. We show a simplified version in Algorithm 1 where the \( f_{\text{incr}} \) (Section 4.2) updates the user vector state without accessing the history; the \( f_{\text{dec}} \) procedure (Section 4.3) receives the deletion request, the current state and the user’s historical baskets and returns the updated user vector state. In our implementation, the users’ historical baskets are stored in memory indexed with user identifiers for fast retrieval. Our code is available under an open source license at https://github.com/0xeeff/amnesiac_recsys.

**Datasets.** The state-of-the-art performance of TIFU-kNN has already been shown by the original authors in [13]. We only adapt this algorithm and show that our incremental and decremental algorithms do not introduce major performance regressions while at the same time providing for efficient incremental learning and data deletion capabilities. We base our experiments on the publicly available datasets (TaFeng, Instacart and ValuedShopper) and the tuned hyper-parameters (listed in Table 1) used for experimentation in the original TIFU-kNN paper [13]. The basic statistics for the datasets are shown in Table 1. These datasets contain timestamped transactions in which items bought in the same order are tampered transactions in which items bought in the same order are conducted randomly and delete 10% of baskets for those users. We select these users at random and delete 10% of baskets for those users.

**Experimental setup.** For each user, the last basket is held out as a test basket for evaluation. We report the average over all users for the metrics. The baseline, incremental and decremental experiments are implemented according to Section 2, Section 4.2 and Section 4.3 respectively. Recall that the baseline triggers a full retraining for any incremental or decremental updates. For the decremental experiment, as it is difficult to estimate a realistic number of deletion requests, we opt for the setup according to [28]: approximately 1 out of 10,000 users might issue deletion requests in industrial scenarios. Due to the limited number of users in our datasets (Table 1), we opt for a more extreme estimate of 1 out of 1,000 users, which is a more difficult experimental setting. We select these users at random and delete 10% of baskets for those users.

**Results and discussion.** As shown in Table 2, the incremental updates produce exactly the same results and therefore performance.
We expect this behavior, as the incremental updates compute the
we only need to access the state and the newly added item or
we are working with a single user and a vocabulary of one single
Basket additions. The baseline algorithm needs to access the full
history size. In this experiment, we follow the setup where we
sequentially add baskets and record the update time along the
way. As shown in Figure 2a, our incremental algorithm achieves
constant update time efficiency, i.e., the time it takes to retrain does
not depend on the size of history, whereas the baseline approach
suffers from a linear increase in update time as the number of
baskets grows.

Basket deletions. The update time for a deletion depends on the
location of the deleted basket. The more recent (closer to the end)
it is, the less time it takes to retrain the model. To better illustrate
this, we investigate three situations: 1) removing baskets one by
one starting from the end; 2) removing baskets one by one starting
from the very beginning; 3) removing baskets one by one from
any random location. In situation (1), the expected update time
should be close to constant. As we remove baskets from the end of
the historical baskets, the number of subsequent baskets it needs
to access to update the user vector is constant. As we can see in
Figure 2b, the experiment result validates our assumption though
it shows a slight downward trend, which we attribute to regularly
triggered maintenance operations in Spark. In situation (2), the
expected update time is approximately linear, as we need to access
the whole history every time a deletion occurs from the begging
of the history. In situation (3), which is a more realistic scenario,
the update time lies in between the previous two situations. In
summary, our approach is on par with the baseline for worst-case
scenarios and consistently outperforms it in other cases.

6.3 Limitations of Repeated Deletions
In the following, we discuss a limitation of our approach, based
on a potential numerical instability issue that could occur under
extreme conditions (a large number of repeated deletions for a
specific user). We theoretically analyze the numerical properties
of our decremental algorithm, and empirically a potential issue in an
extreme case scenario.

We can rewrite the user vector update rule (Equation 12) as
\( u' = au + C \) where \( a = \frac{k}{(k-1)r_q} > \frac{1}{r_q} > 1 \). Suppose the initial
user vector is \( u_0 = \hat{u}_0 + \epsilon \), where we use the hat notion \( \hat{u}_0 \)
to represent the true value, and the \( \epsilon \) denotes the initial error. After \( n \)
steps of deletions, the accumulated error becomes \( |u_n - \hat{u}_n| = \epsilon a^n \).
Since \( a > 1 \), the decremental algorithm is numerically unstable.
This conclusion implies that there is a limit in terms of how many
deletions we can perform for a given user before the error becomes
unbearable.

Experimental setup. In this experiment, we first build up the user
vector through incremental updates, then we continuously conduct
decremental updates during which we calculate the true value of

Table 1: Dataset basic statistics and hyper-parameters (from left to right: group size \( m \), decay rates \( r_p \) and \( r_q \), the number of
neighbors \( k \), the combining weight of the target user vector and neighborhood component \( \alpha \)).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#users</th>
<th>#items</th>
<th>#baskets</th>
<th>avg basket size</th>
<th>avg #baskets/user</th>
<th>hyper-parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>TaFeng</td>
<td>13949</td>
<td>11997</td>
<td>79423</td>
<td>6.2</td>
<td>5.7</td>
<td>[7, 0.9, 0.7, 300, 0.7]</td>
</tr>
<tr>
<td>Instacart</td>
<td>19935</td>
<td>7999</td>
<td>158933</td>
<td>8.9</td>
<td>8.0</td>
<td>[3, 0.9, 0.7, 900, 0.9]</td>
</tr>
<tr>
<td>ValuedShopper</td>
<td>10000</td>
<td>7874</td>
<td>568573</td>
<td>9.1</td>
<td>56.9</td>
<td>[7, 1.0, 0.6, 300, 0.7]</td>
</tr>
</tbody>
</table>

Table 2: Predictive performance under incremental and decremental updates compared to the baseline approach of retraining
the model from scratch. Our online algorithms provide equal predictive performance.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>metric</th>
<th>baseline</th>
<th>incr.</th>
<th>decr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TaFeng</td>
<td>Recall@10</td>
<td>0.1298</td>
<td>0.1298</td>
<td>0.1297</td>
</tr>
<tr>
<td></td>
<td>NDCG@10</td>
<td>0.0847</td>
<td>0.0847</td>
<td>0.0847</td>
</tr>
<tr>
<td></td>
<td>Recall@20</td>
<td>0.1859</td>
<td>0.1859</td>
<td>0.1856</td>
</tr>
<tr>
<td></td>
<td>NDCG@20</td>
<td>0.0816</td>
<td>0.0816</td>
<td>0.0817</td>
</tr>
<tr>
<td>Instacart</td>
<td>Recall@10</td>
<td>0.1709</td>
<td>0.1709</td>
<td>0.1710</td>
</tr>
<tr>
<td></td>
<td>NDCG@10</td>
<td>0.5789</td>
<td>0.5789</td>
<td>0.5788</td>
</tr>
<tr>
<td></td>
<td>Recall@20</td>
<td>0.2291</td>
<td>0.2291</td>
<td>0.2291</td>
</tr>
<tr>
<td></td>
<td>NDCG@20</td>
<td>0.5037</td>
<td>0.5037</td>
<td>0.5035</td>
</tr>
<tr>
<td>ValuedShopper</td>
<td>Recall@10</td>
<td>0.1436</td>
<td>0.1436</td>
<td>0.1437</td>
</tr>
<tr>
<td></td>
<td>NDCG@10</td>
<td>0.4058</td>
<td>0.4058</td>
<td>0.4060</td>
</tr>
<tr>
<td></td>
<td>Recall@20</td>
<td>0.2133</td>
<td>0.2133</td>
<td>0.2131</td>
</tr>
<tr>
<td></td>
<td>NDCG@20</td>
<td>0.3465</td>
<td>0.3465</td>
<td>0.3465</td>
</tr>
</tbody>
</table>
the user vector using the baseline algorithm. The hyper-parameters are fixed: $m = 2$, $r_g = 0.7$, $r_p = 0.9$.

**Results and discussion.** As shown in Figure 2c, the error grows exponentially with the number of decremental updates, which is confirmed by the error fitting curve. While this issue might be theoretically worrying, we argue that its impact in practice may be negligible, as the overall number of deletions is typically low, and these deletions will be interleaved with incremental updates. The incremental algorithm is a numerically stable algorithm that helps deduce the accumulated error. In our experiment, to incur 1% relative error, it takes 180 continuous deletions, which is arguably an extreme case.

### 7 RELATED WORK

**Decremental learning.** Contrary to conventional machine learning tasks, decremental learning attempts to remove the impact of training data from the trained model efficiently. This line of research dates back to some early work on incremental and decremental versions of Support Vector Machines [6, 15], while it has recently gained traction due to the rising concerns for privacy in ML models. Ginart et al. [8] study data deletion mechanisms for variations of k-means clustering. Schelter et al. [26] propose decremental learning methods for a selection of non-iterative ML models such as item-based collaborative filtering and KNN. Recent work from Guo et al. [10] studies this problem for neural networks and shows promising theoretical results albeit apparent scalability bottlenecks due to the requirement of reversing Hessian matrix. Bourtoule et al. [3] propose a framework in which training samples are divided into shards and trained in silo to obtain a combination of weaker learners, and later on, combined to form a stronger model. When there is a deletion request concerning specific user data, it’s sufficient to update the corresponding impacted weaker learner.

**Deep learning for NBR.** Deep learning [4, 9, 18, 22, 23] gains momentum in RS given its potential in learning powerful latent representations of users/items [39]. Its ability to model sequential patterns is especially interesting for NBR. Many new methods based on recurrent neural networks and attention mechanisms [33] have been proposed [11, 12, 20, 21, 24, 35, 36]. However, questions for the effectiveness of deep learning in the RS domain, especially next basket recommendation/sequential recommendation, remain open. Other studies show that simple neighborhood-based methods outperform complex neural models easily [13, 14, 16].

### 8 CONCLUSION

In this work, we focus on efficient data deletion mechanisms for recommender systems in light of privacy protection and introduce the problem of decremental learning to Next Basket Recommendation. We then present efficient and effective algorithms capable of learning and forgetting with millisecond-level latency. Our approach is evaluated on real-world datasets and demonstrates its effectiveness in terms of ranking metrics (recall and NDCG). Moreover, it also proves to have desirable update time efficiency as shown in our experiments. In addition to that, we provide a fully parallel implementation in Apache Spark to facilitate practical use cases. For theoretical completeness, we also study the limitations under extreme conditions, that is, the decremental algorithm could suffer from numerical instability for a large number of continuous deletions. For future work, we intend to investigate: 1) how to incorporate incremental and decremental algorithms for other more complex recommendation scenarios; 2) numerical stability properties for other ML models under similar settings. In the end, we hope for more advances in this area to realize a future of truly “amnesiac” recommender systems.

**Acknowledgments.** This work was supported by Ahold Delhaize. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.
REFERENCES


