Lightweight Transformer-Based Event Embeddings with Log-Frequency Session Weighting for Purchase Prediction in E-Commerce Clickstream Data

Deniz Guzelcay[1(✉)], Sefa Çayır[1], Konuralp Bol[1], Halil İbrahim Simsek[2], Onur Cilgin[2], and Mehmet S. Aktas[1]

[1] Yildiz Technical University, Istanbul, Turkey
{deniz.guzelcay,sefa.cayir,konuralp.bol}@std.yildiz.edu.tr, aktas@yildiz.edu.tr
[2] Trendyol R and D Center, Istanbul, Turkey
{halilibrahim.simsek,onur.cilgin}@trendyol.com

Abstract. E-commerce platforms generate large volumes of clickstream data in the form of short user sessions. Predicting whether the next event in a session will be a purchase is a useful task for recommendation, marketing analytics, and real-time decision support. Many existing methods address this problem with complex sequential deep learning models, but such models often require substantial computational resources and large training datasets. This limits their practicality in resource-constrained settings. In this work, we study a lightweight alternative based on event embeddings and frequency-aware session representation. Each clickstream event token, such as a page view or user action, is encoded as a dense vector using a lightweight transformer-based embedding model. To reflect repeated behavior within a session, we apply a logarithmic frequency-weighting scheme so that recurring events contribute more strongly to the final session representation without dominating it. The weighted event embeddings are then aggregated into a fixed-length session vector and used as input to supervised classifiers for next-event purchase prediction. The approach is evaluated on both a public clickstream dataset and a large real-world e-commerce dataset. Experimental results show that frequency-weighted session representations perform better than unweighted representations across different embedding models and classifiers. The gains are more evident in highly imbalanced settings where purchase events are rare. These findings indicate that lightweight transformer-based embeddings combined with log-frequency weighting offer a practical approach for next-event purchase prediction in large-scale clickstream environments.

Keywords: Clickstream Data Analysis · Purchase Prediction · Event Embeddings · Session Representation · Lightweight Transformers · Frequency Weighted Aggregation

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16759, pp. 242–259, 2027.
https://doi.org/10.1007/978-3-032-30518-3_16

1 Introduction

E-commerce platforms now produce large amounts of interaction data as users browse, search, view products, and add items to their carts. These actions are typically stored as clickstream logs. When viewed over short periods, they form user sessions that can give clues about what the user intends to do next. This makes session data useful for several tasks, including recommendation, advertising, and behavior analysis. A particularly important question is whether a session will result in a purchase. Since purchases are directly related to revenue, predicting them has become a common problem in both data mining and recommender systems [1,2]. The present work approaches this problem from the angle of efficient session representation, looking for a balance between predictive accuracy and the computational cost of deployment in large-scale platforms.

Research on e-commerce has examined both purchase intent prediction and user behavior modeling. Early studies largely relied on statistical methods and manually designed behavioral features derived from clickstream logs, such as visit counts, session duration, and navigation depth [1]. These approaches were computationally efficient, but they often fell short in capturing the more complex patterns of user navigation. To improve predictive performance, later research began applying machine learning models to deeper feature sets [2]. Over time, representation learning methods and neural architectures also became more common for modeling the relationships among behavioral signals. Neural recommender models and factorization-based approaches, were used in click-through rate and conversion prediction and showed promising results [3,4]. At the same time, sequential modeling techniques, including recurrent neural networks and transformer-based architectures, were increasingly adopted to capture temporal dependencies in session-based recommendation and next-event prediction tasks [5].

These developments improved predictive performance, but they also raised practical concerns. Many deep sequential models deliver strong results, yet they are not always easy to use in real deployments. Training and deploying them often requires considerable computing power, enough labeled data, and fairly involved inference pipelines. In production environments, where fast and efficient predictions are essential, these requirements can become a real limitation. Their complexity can also make deployment and maintenance harder, especially on large commercial platforms. At the other end, lightweight session modeling methods are easier to work with, but they often oversimplify user behavior. When they rely on basic pooling or hand-crafted features, they may overlook the meaning of different event types and miss important behavioral signals in the interaction sequence [6,7]. As a result, there is a need for session representation methods that can do both: capture meaningful patterns in user behavior and still remain efficient and scalable in real-world settings. This need shapes the central question of the present study, namely how to build session representations that are informative enough for next-event purchase prediction while remaining computationally light. The remainder of this section reviews related embedding-based work and then introduces the proposed approach in more detail.

Several of the past studies have looked into using embedding techniques to model user behavior in clickstream data. For instance, researchers have used graph-based and natural language processing (NLP) inspired embedding methods such as Word2Vec, Node2Vec, and DeepWalk. These methods help represent user navigation sessions and analyze behavioral patterns in e-commerce platforms [8]. In later research, researchers also looked at embedding-based representations in system-level applications. These included predictive prefetching and smart data delivery in data-heavy client-server systems [9]. Several other studies have investigated the extent to which weighting methods can enhance embedding quality. One example is event frequency, inverse session frequency (EF-ISF) weighting, which helps in finding behavioral patterns [10] and aging-based temporal weighting strategies for improving session classification performance [11]. While these studies demonstrated the effectiveness of embedding techniques for representing clickstream data and modeling user behavior, they primarily focused on improving representation quality or clustering and classification performance through various embedding and weighting strategies. By contrast, our focus is on purchase prediction. To address it, we use a lightweight session representation that combines computationally efficient transformer-based event embeddings with log-frequency weighting to capture behavioral intensity within user sessions.

Within this context, the task we focus on is predicting whether the next event in a user session is a purchase, based on the navigation behavior observed up to that point. The central research problem is how to construct informative and efficient session-level representations from clickstream events that support this prediction task. More specifically, we examine whether repeated user actions within a session are associated with purchase intent and how this information can be incorporated into the session representation. To tackle this research problem, we need a way to represent the semantic features of event tokens. We also need to capture the intensity of behavior shown in repeated interactions during sessions.

Our approach to this challenge is to combine lightweight transformer-based event embeddings with a logarithmic frequency-weighting strategy at the session level. In the proposed method, each clickstream event token representing a page or user action is converted into a dense numerical vector. This transformation uses pretrained transformer-based embedding models, such as Sentence-BERT and related architectures [12–14]. Unlike large transformer models used in sequential recommendation tasks, the embedding models in this study are easy to run and fit for settings with limited computational resources. After event embeddings are generated, session-level representations are built by combining event vectors. This uses a logarithmic frequency-weighting method that boosts the impact of repeated events. It also limits excessive dominance with logarithmic scaling. In this way, the representation of the session captures both what the events represent and how strongly certain behaviors appear within the session.

To assess the approach, we run experiments on a public clickstream dataset and on a large-scale real-world e-commerce dataset. Event embeddings generated from lightweight transformer models are aggregated into session-level vec-

tors using both standard unweighted aggregation and the proposed log-frequency weighted aggregation strategy. These session vectors are then used as input features for supervised classification models, including Random Forest [15] and XGBoost [16]. To ensure reliable evaluation and prevent data leakage across user sessions, a stratified group cross-validation strategy is used during model training and testing. Model performance is evaluated using widely adopted classification metrics such as F1-score and accuracy, which are particularly suitable for evaluating classification tasks involving imbalanced datasets [17].

The main contributions of this study can be summarized as follows. First, we introduce a lightweight session representation framework that converts clickstream event tokens into numerical vectors using efficient transformer-based embedding models. Second, we propose a logarithmic frequency-weighting method that incorporates session-level event frequency information in creating session representations. Third, we demonstrate, through extensive experiments on both public and real-world datasets, that the frequency-weighted session representation significantly improves purchase prediction performance compared to standard unweighted aggregation methods. Finally, the results indicate that our approach offers an effective and efficient alternative to complex sequential models for purchase prediction in large-scale e-commerce settings.

The rest of this paper is organized as follows. Section 2 presents a review of the literature on session-based user behavior modeling, embedding-based clickstream representations, and the prediction of purchase intent. Section 3 introduces the proposed frequency-weighted session embedding framework, including the generation of event embeddings using lightweight transformer models and the log-frequency weighting method for creating session representations. Section 4 describes the experimental study, which includes the datasets, preprocessing steps, experimental setup, and evaluation methods. Additionally, this section discusses the experimental results, interpretation of the findings, and possible threats to validity. Finally, Sect. 5 concludes the paper and outlines several directions for future research.

2 Literature Review

Modeling user behavior from session-based interaction data has become a central research problem in recommender systems, clickstream analytics, and e-commerce intelligence. In many online platforms, user interactions occur in short browsing sessions without persistent user identifiers or long-term historical records. Predicting user intent from limited session-level behavior signals has gained a lot of interest in both research and business. This problem has been investigated through several modeling approaches, ranging from sequential neural architectures and embedding-based methods to graph-based models, frequency-aware session representations, and transformer-based sequential recommenders. Each of these approaches aims to capture different aspects of user behavior. They look at the order of events, the meaning behind their connections, and the different levels of importance of interactions during a session [7].

Early work on session-based recommendation mainly used recurrent neural networks to model the sequence of user interactions. A key contribution in this area was the GRU4Rec model introduced by Hidasi et al. [18]. It showed that gated recurrent neural networks can model time patterns in user sessions effectively. Later studies built on this idea and suggested various ways to improve sequential modeling. Li et al. [19], for example, proposed the Neural Attentive Recommendation Machine (NARM), which uses an attention mechanism to focus on the most informative items within a session. Quadrana et al. [20] later proposed hierarchical recurrent neural networks to capture patterns at both the session level and the user level. Later surveys highlighted the importance of sequential patterns in recommendation systems and provided a detailed classification of sequence-aware recommender algorithms [7]. These architectures work well for modeling user interactions over time. However, they often require a lot of training data and computing power.

To move beyond the limits of purely sequential models, several studies have explored embedding-based representation learning methods for modeling session data. Improvements in natural language processing inspire this approach. Sessions can be seen like sentences. Interaction events serve as words. In this context, embedding algorithms such as Word2Vec are often used to learn dense vector representations of items. These representations depend on how frequently the items appear together in sessions [21]. For instance, Greenstein-Messica and Rokach [22] proposed a session-based recommendation approach that learns item embeddings directly from session sequences. Similarly, graph embedding techniques such as DeepWalk [23] and Node2Vec [24] have been applied to interaction graphs constructed from session transitions. These approaches help show hidden similarities between items and uncover behavioral patterns across sessions. However, many embedding-based methods treat all events in a session equally. They do not consider that some interactions might be more important than others.

Several studies addressed this issue by adding weighting methods to session representations. Similar to TF-IDF in information retrieval, these methods assign more weight to important events and less to interactions that happen often but provide little information. In clickstream and behavioral modeling tasks, frequency-based weighting can also detect signals of behavioral intensity. This may indicate a stronger user interest or intent to buy. Event frequency-inverse session frequency is one example and has been explored for representing behavioral patterns and improving session classification performance [11]. The limitation is that these methods usually work as stand-alone statistical weighting schemes rather than as part of the embedding process itself. As a result, they can highlight important interactions, but they may still miss semantic relationships between different kinds of events.

More recent research has attempted to overcome these limitations by integrating attention mechanisms into sequential recommendation models. Attention-based architectures dynamically estimate the importance of events within a session and allow neural models to focus on interactions that are most predictive of user intent. For example, Li et al. [19] proposed an attention-

based recurrent model that learns both global session representations and local item importance weights. Similarly, self-attention mechanisms have been widely adopted in sequential recommendation models to capture contextual dependencies between interaction events [25]. These models improve representation learning by allowing the model to selectively focus on informative elements within the sequence. However, attention-based architectures typically rely on complex neural networks that require substantial computational resources for training and inference.

Graph-based approaches have also emerged as an effective solution for modeling session-based interactions. In graph-based session models, items are represented as nodes and transitions between items are modeled as edges in a directed graph structure. Graph neural networks can then be used to learn representations that capture complex relationships between interaction events. One of the most influential works in this area is SR-GNN proposed by Wu et al. [26], which models sessions as directed graphs and applies graph neural networks to capture item transition patterns. Graph-based approaches have demonstrated strong performance in session-based recommendation tasks because they explicitly model structural relationships between items. However, graph neural architectures often introduce additional computational complexity due to graph construction and message-passing operations.

More recently, transformer-based architectures have been introduced for modeling sequential user behavior in recommendation systems. These approaches leverage self-attention mechanisms to capture long-range dependencies between events in user sessions. For example, Sun et al. [5] proposed BERT4Rec, a transformer-based recommendation model that learns bidirectional representations of item sequences. Kang and McAuley [25] proposed SASRec, which applies self-attention to model sequential user behavior. In addition, industrial recommendation systems like Transformers4Rec have been created to connect natural language processing methods with large-scale recommendation processes [27]. While transformer-based architectures often achieve state-of-the-art performance, they generally require large training datasets and significant computational resources.

Several of our previous studies focused on embedding-based ways of modeling user navigation in clickstream environments. Earlier work examined graph-based and NLP-inspired methods, such as Word2Vec, Node2Vec, and DeepWalk, for representing user navigation sessions and analyzing behavioral patterns on e-commerce platforms. Subsequent studies examined embedding-based predictive prefetching strategies for intelligent data delivery mechanisms in data-intensive client-server systems. Other studies investigated the integration of this work by considering weighting mechanisms within embedding representations, including event frequency–inverse session frequency weighting and aging-based temporal weighting strategies for improving session classification [11]. Although these studies supported the use of embedding and weighting methods for user behavior modeling, they focused on representation learning and session classification rather than purchase prediction.

Overall, the current research shows that session-based behavioral modeling improves with sequential neural architectures, embedding-based representations, graph-based interaction modeling, and attention-based weighting methods. However, several challenges persist. Many leading methods depend on deep neural architectures that require a lot of computing power. This makes them unsuitable for real-time use in large-scale e-commerce settings. Additionally, embedding-based representations often do not clearly include signals of behavioral intensity, such as how often events happen within sessions. On the other hand, statistical weighting methods point out important interactions but overlook the semantic relationships between events. To tackle these issues, this study suggests a lightweight session representation framework. It combines transformer-based event embeddings with a logarithmic frequency-weighting method. This approach allows for the development of efficient and informative session representations for predicting purchases in large-scale clickstream environments.

3 Methodology

This section presents the proposed framework for constructing frequency-weighted session representations from clickstream event sequences for purchase intent prediction. The primary goal of the framework is to transform user navigation events into informative numerical representations that capture both the semantic properties of interaction events and the behavioral intensity reflected by repeated actions within a session. Instead of training computationally expensive sequential deep learning architectures such as recurrent neural networks or transformer-based sequence models, the proposed approach utilizes pretrained lightweight transformer-based embedding models to generate semantic representations of event tokens. These representations are then aggregated using a frequency-aware weighting strategy to construct session-level embeddings. As illustrated in Fig. 1, the proposed framework consists of three main stages:

1. Event representation learning using pretrained transformer-based embeddings
2. Frequency-based weighting of session events
3. Purchase prediction using supervised classification models

Event Token Embedding Generation: User navigation sessions consist of sequences of interaction events such as page visits, product views, searches, and cart interactions. In the proposed framework, each event type is treated as a token similar to words in natural language processing. This representation enables the application of embedding techniques to capture semantic relationships between interaction events. To generate numerical representations of event tokens, we employ pretrained lightweight transformer-based embedding models. Specifically, three Sentence-BERT (SBERT) based models are utilized in this study to balance computational efficiency and representation quality:

- *all-MiniLM-L6-v2* [28]
- *intfloat/e5-small-v2* [13]
- *BAAI/bge-small-en-v1.5* [14]

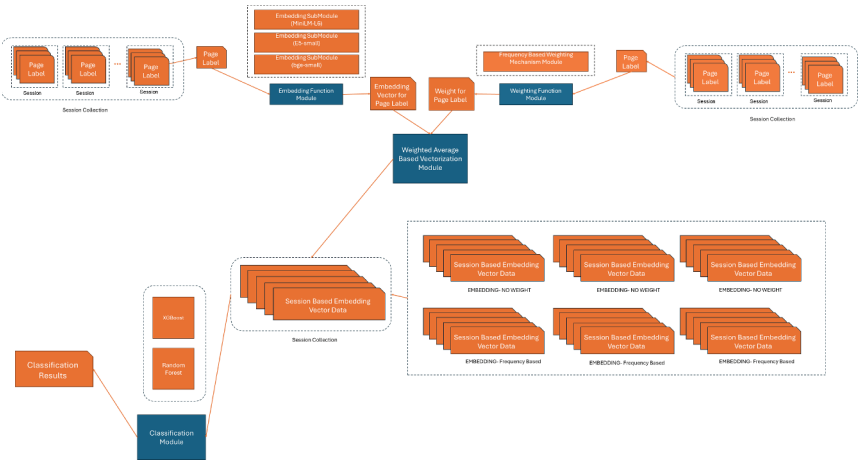

Fig. 1. Overview of the proposed framework.

Given an event token e_i, the embedding model $M(\cdot)$ generates a dense vector representation $v_i \in \mathbb{R}^D$ as follows:

$$v_i = M(e_i) \tag{1}$$

These embeddings position related events near each other in the vector space. This helps capture the similarities in interaction patterns.

Log-Frequency Weighting Mechanism: Embedding models are useful for capturing semantic relationships between event types, but standard aggregation methods like simple averaging still assume that every event contributes in the same way to the final session representation. In practice, this is often not the case. When users interact repeatedly with certain event types, those actions may carry stronger behavioral meaning or indicate higher intent. To reflect this more clearly, we incorporate a logarithmic frequency-weighting strategy into the session representation. Let $freq(e_i)$ denote the number of occurrences of event e_i within session S. The importance weight assigned to event e_i is defined as

$$w_i = \log(1 + freq(e_i)) \tag{2}$$

This logarithmic transformation increases the importance of repeated events while limiting the influence of events that occur very often.

Frequency-Weighted Session Embedding Construction: Once the event embeddings and frequency weights are available, the final session representation is formed through weighted aggregation of the event vectors. For a session $S = \{e_1, e_2, ..., e_n\}$, the session embedding V_S is computed as the normalized weighted sum of event embeddings:

$$V_S = \frac{\sum_{i=1}^{n} w_i \cdot v_i}{\sum_{i=1}^{n} w_i} \tag{3}$$

This formulation gives repeated events more influence on the final representation while keeping the values numerically stable.

Frequency-Weighted Session Embedding Algorithm: A summary of the proposed frequency-weighted procedure for constructing session embeddings is provided in Algorithm 1.

Algorithm 1. Frequency Weighted Session Embedding

Input: Session $S = \{e_1, e_2, ..., e_n\}$
Output: Session embedding vector V_S
Initialize $V_S \leftarrow 0$
Initialize $W \leftarrow 0$
for each event e_i in session S **do**
    $v_i \leftarrow M(e_i)$
    $w_i \leftarrow \log(1 + freq(e_i))$
    $V_S \leftarrow V_S + w_i \cdot v_i$
    $W \leftarrow W + w_i$
end for
$V_S \leftarrow V_S / W$
return V_S

The algorithm operates in two main stages. First, each event token in the session is transformed into a dense semantic vector using the selected embedding model $M(\cdot)$. These vectors capture semantic relationships between different interaction types in the clickstream. Second, a logarithmic frequency weighting is applied. This allows repeated events to contribute more to the session representation. However, it does not let very frequent events dominate the session. After that, the event vectors are combined using normalized weighted aggregation. This creates the final session embedding. It creates a clear representation that captures both semantic similarity and also the frequency of interactions. It also provides downstream machine learning models with a better basis for using session-level behavioral signals to predict purchases from large-scale clickstream datasets.

Purchase Prediction Framework: The session embeddings created in the earlier stage give short numerical descriptions of how users navigate. These representations then serve as an input feature for supervised learning algorithms meant to predict purchase intent. The framework transforms clickstream sessions of varying length into fixed-length vectors, which allows traditional machine learning methods to be used more efficiently without explicit sequence modeling. In this way, it provides a computationally simpler alternative to complex sequential deep learning models, while still retaining meaningful information about user navigation behavior.

Using frequency-weighted embeddings helps the framework retain behavioral cues that are often lost in traditional session modeling. Repeated interactions with certain event types may signal stronger intent or deeper interest in particular products or categories. When semantic event representations are combined

with frequency-based weighting, the resulting session vectors capture both the nature of the interactions and their intensity. This leads to a more informative feature space for supervised learning models. It allows classifiers to better differentiate between sessions that result in purchases and those that do not. As a result, the proposed framework makes it easier to predict purchase intent while keeping computational efficiency suitable for large-scale clickstream environments.

Computational Complexity Analysis: An important advantage of the proposed framework is its computational efficiency compared to sequential deep learning models that explicitly model event order within sessions. Let n denote the number of events in a session and d represent the embedding dimensionality. The generation of event embeddings requires a constant-time inference operation for each event using the pretrained embedding model. Therefore, the embedding stage has a computational complexity of $O(n)$ per session. The frequency-weighting step requires computing the frequency of each event within the session and applying the logarithmic transformation. Since these operations are performed once per event, this step also has a complexity of $O(n)$. The weighted aggregation of event embeddings requires a linear combination of n vectors of dimension d, resulting in a complexity of $O(n.d)$ per session. Consequently, the overall computational complexity of constructing a session embedding using the proposed framework is $O(n.d)$ which scales linearly with the number of events in a session. This linear complexity makes the proposed approach significantly more efficient than sequential neural architectures such as recurrent neural networks or transformer-based sequence models, which typically exhibit quadratic or higher computational costs with respect to sequence length. Therefore, the proposed framework provides a computationally efficient and scalable solution for large-scale clickstream analysis environments.

4 Experimental Study

This section presents the experimental evaluation of the proposed frequency-weighted session embedding framework. The objective of the experimental study is to investigate whether incorporating session-level event frequency information into embedding aggregation improves purchase prediction performance compared to conventional unweighted session representations. The experimental evaluation focuses on three main research questions:

- **RQ1:** Does frequency-weighted aggregation improve purchase prediction performance compared to standard unweighted embedding aggregation?
- **RQ2:** Is the proposed approach robust across different lightweight transformer embedding models?
- **RQ3:** How does the effectiveness of frequency weighting change under different class imbalance conditions?

These questions were investigated through experiments using both a public clickstream dataset and a large-scale proprietary e-commerce dataset.

Datasets: Two clickstream datasets with similar structural properties but different behavioral characteristics were used in the experiments. The first dataset contains user activity logs from a mobile coffee shop application published on Kaggle. Each record corresponds to a user session and includes information about the user identifier, page interactions, timestamp information, and other contextual attributes. A sample record from the dataset is shown in Table 1. This example illustrates how a single interaction event is represented using user-level contextual attributes.

Table 1. Sample User Data

User Information	Page Information	Day of the Week	Hour
0096610	002	03 (Wednesday)	23

Because user actions in sessions are recorded as sequential events, this dataset provides a suitable framework for session-based modeling. The dataset contains 276,566 sessions and 46 distinct event types performed by 96,200 users. The diversity of interaction types makes the dataset appropriate for evaluating embedding-based behavioral modeling approaches.

The second dataset consists of large-scale user clickstream data obtained from the real-life e-commerce platform. Each record contains a session identifier, a sequence duration array describing time intervals between interactions, and a sequence array representing the ordered list of user actions. A representative example of a session record from the real-life e-commerce website dataset is shown in Table 2. The table illustrates how a session is represented as a sequence of interaction events and corresponding temporal intervals.

Table 2. Sample from real-life e-commerce dataset

Field	Value
session_id	0000585b-f366-417b-aead-23d14b03a433
sequence_duration	0, 4, 1402, 2, 11, 4, 39, 3, 47, 3, 19, 4, 2, 3, 45, 4
sequence	PRODUCT, OTHER, PRODUCT, PRODUCT, PRODUCT, PRODUCT, PRODUCT, MAINPAGE, PRODUCT, OTHER, OTHER, SEARCH, SEARCH, SEARCH, SEARCH, SEARCH

Unlike the public dataset, the real-life e-commerce website dataset contains pre-labeled purchase outcomes, enabling direct supervised learning experiments without additional annotation. The statistical comparison of both datasets is presented in Table 3. As shown in the table, the datasets differ in several important ways, especially in terms of the severe class imbalance in the proprietary dataset and the longer average session length in the coffee shop dataset.

Table 3. Statistical Comparison of Datasets

Metric	Real-life e-commerce website	Coffee Shop
Total Sessions	547,715	276,566
Purchase Ratio (Balance)	1.08% (Highly Imbalanced)	54.38% (Balanced)
Vocabulary Size	37 Events	46 Events
Avg. Session Length	16.6 Events	156.6 Events

Using both balanced and highly imbalanced datasets allows us to test how robust and generalizable the proposed representation method is under different behavioral distributions.

Data Preprocessing: Clickstream datasets often include interactions that take place after a purchase. If we keep these events during training, the model may focus on the purchase itself instead of the behaviors that lead to it. To prevent this issue, we added a preprocessing step to end sessions right after the purchase event. For sessions with a purchase label, we removed all events that happened after the purchase. This step ensures the model only sees behaviors before the purchase and prevents data leakage during training.

Experimental Setup: The experimental prototype was built in Python 3.14 using popular data science libraries. We used *Pandas* and *NumPy* for data preprocessing and session manipulation. Event embeddings were created with pretrained models from the *SentenceTransformers* framework. We reduced embedding dimensionality to 64 using a dense projection layer built with *PyTorch*. Two widely used tree-based classification algorithms were used: Random Forest and XGBoost. These classifiers are reliable in high-dimensional feature spaces and perform well on tabular prediction tasks.

Cross-Validation Strategy: To get a reliable performance estimate, we used a 10-Fold Stratified Group Cross-Validation strategy with *StratifiedGroupKFold*. This evaluation strategy addresses two important issues in clickstream prediction: Maintaining class distribution across folds, Preventing data leakage across users. By grouping sessions by user identifier, interactions from the same user cannot appear in both training and testing partitions.

Evaluation Metrics: Two standard classification metrics were used: F1-score (Purchase class), Accuracy. Because purchase prediction represents an imbalanced classification task in real-world e-commerce environments, the F1-score for the purchase class was used as the primary evaluation metric.

Experimental Results: The experimental results obtained on the proprietary real-life e-commerce website dataset are summarized in Table 4. The table compares the performance of weighted and unweighted session representations across three embedding models and two classifiers.

To further highlight the effect of frequency weighting, Table 5 presents the relative improvement in F1-score obtained by the proposed method compared to the unweighted baseline.

Table 4. Performance on Proprietary(Real-life e-commerce website) Dataset

Model	Clf.	Method	F1-Score(Purchase)	Accuracy
MiniLM	RF	Unweighted	0.3759	0.9681
		Weighted	**0.6139**	**0.9924**
	XGB	Unweighted	0.3736	0.9673
		Weighted	**0.6828**	**0.9910**
E5-small	RF	Unweighted	0.3762	0.9681
		Weighted	**0.6100**	**0.9923**
	XGB	Unweighted	0.3732	0.9673
		Weighted	**0.6858**	**0.9911**
BGE-small	RF	Unweighted	0.3759	0.9681
		Weighted	**0.6142**	**0.9924**
	XGB	Unweighted	0.3728	0.9673
		Weighted	**0.6828**	**0.9910**

Table 5. Relative F1-Score Improvement on Proprietary Dataset

Model	Classifier	Improvement (%)
MiniLM	Random Forest	+63.31%
	XGBoost	**+82.76%**
E5-small	Random Forest	+62.15%
	XGBoost	**+83.76%**
BGE-small	Random Forest	+63.39%
	XGBoost	**+83.15%**

Similar experiments were conducted on the coffee shop dataset. The results are presented in Table 6.

Relative improvements for the balanced dataset are summarized in Table 7.

Discussion: The results suggest that frequency-aware weighting helps session embeddings perform better in purchase prediction. This pattern is particularly noticeable in the proprietary real-world e-commerce dataset, where purchase events are very rare. In that setting, repeated interactions with certain event types appear to signal stronger purchase intent. The logarithmic weighting scheme captures some of this intensity, which seems to help the classifier distinguish purchase sessions from non-purchase ones more effectively. The gain is smaller in the coffee shop dataset, which has a more balanced class distribution. Even so, the weighted version still performs better than the unweighted baseline across all embedding models and classifiers considered in the study.

Behavioral Interpretation: Beyond performance improvements, the experimental findings provide important insights into user behavior. When users interact multiple times during a session, it usually indicates they are more

Table 6. Performance on Coffee Shop Dataset

Model	Clf.	Method	F1-Score(Purchase)	Accuracy
MiniLM	RF	Unweighted	0.6487	0.6673
		Weighted	**0.7374**	**0.7359**
	XGB	Unweighted	0.6485	0.6672
		Weighted	**0.7605**	**0.7506**
E5-small	RF	Unweighted	0.6484	0.6671
		Weighted	**0.7369**	**0.7355**
	XGB	Unweighted	0.6489	0.6674
		Weighted	**0.7607**	**0.7511**
BGE-small	RF	Unweighted	0.6485	0.6672
		Weighted	**0.7366**	**0.7353**
	XGB	Unweighted	0.6484	0.6670
		Weighted	**0.7597**	**0.7502**

Table 7. Relative F1-Score Improvement on Coffee Shop Dataset

Model	Classifier	Improvement (%)
MiniLM	Random Forest	+13.67%
	XGBoost	**+17.27%**
E5-small	Random Forest	+13.65%
	XGBoost	**+17.23%**
BGE-small	Random Forest	+13.59%
	XGBoost	**+17.16%**

engaged with specific products or categories. For example, users might look at product pages several times, do multiple searches, or go back to certain categories before making a purchase. The suggested frequency-weighted representation captures this behavior by giving more importance to repeated events while using logarithmic scaling to avoid excessive influence. As a result, the created session embeddings reflect both the relationships among interaction events and the signals of behavioral intensity that indicate user intent.

Threats to Validity of the Results: Several potential threats to validity should be considered when interpreting the experimental results. First, although two datasets were used, both datasets originate from e-commerce or mobile application environments. Therefore, the generalizability of the proposed method to other domains of sequential user behavior remains an open question. Second, the experiments rely on pretrained embedding models without domain-specific fine-tuning. While this choice ensures computational efficiency, fine-tuned models may potentially achieve higher predictive performance. Third, the evaluation was conducted using offline experiments. Future work should evaluate the app-

roach in online production environments to assess its real-time performance and scalability.

5 Conclusion and Future Work

This study addressed the problem of predicting purchase intent from session-based clickstream data in e-commerce environments. Modeling user behavior from short interaction sessions remains a challenging task because clickstream sequences often contain limited contextual information while exhibiting complex behavioral patterns. Existing approaches frequently rely on computationally expensive sequential neural architectures, which may limit their applicability in large-scale production systems. To overcome these limitations, this work proposed a lightweight session representation framework that integrates transformer-based event embeddings with a logarithmic frequency-weighting mechanism.

Returning to the three research questions stated in Sect. 4, the experimental evidence can be summarized as follows. Regarding RQ1, the log-frequency weighted aggregation produced higher F1-scores than the unweighted baseline in every configuration tested in this study. As reported in Tables 5 and 7, the relative F1 improvements ranged from approximately 13% on the balanced coffee shop dataset to over 80% on the highly imbalanced proprietary dataset. Regarding RQ2, the gain was consistent across the three lightweight transformer embedding models considered (MiniLM, E5-small, and BGE-small); the differences between these embedding models were small relative to the gap between the weighted and unweighted variants. This suggests that the weighting scheme, rather than the choice of embedding model, is the main driver of the observed improvement. Regarding RQ3, the effect of frequency weighting was clearly larger on the imbalanced dataset, where purchase events are rare, than on the balanced dataset. A plausible interpretation is that, when positive examples are scarce, the additional signal carried by repeated intent-bearing events becomes more discriminative for the classifier.

The proposed framework constructs session-level representations by first transforming clickstream event tokens into dense semantic vectors using pretrained lightweight transformer embedding models. These event representations are then aggregated through a log-frequency weighting strategy that emphasizes repeated interactions within a session while preventing extreme dominance of highly frequent events. By integrating semantic event representations with frequency-aware behavioral signals, the proposed approach generates compact session embeddings that capture both interaction semantics and behavioral intensity without requiring explicit sequence modeling.

A detailed experimental study used both a publicly available clickstream dataset and a large real-world e-commerce dataset. The experiments compared the proposed frequency-weighted aggregation method with traditional unweighted embedding aggregation across several transformer embedding models and classification algorithms. The results consistently showed that adding

frequency-based weighting significantly improves purchase prediction performance. In particular, the proposed method achieved notable gains in F1-score and overall accuracy across different embedding models and classifiers. These improvements were especially significant in the proprietary e-commerce dataset, which had a serious class imbalance. This suggests that frequency-aware session representations effectively capture behavioral signals related to purchase intent.

Overall, the findings suggest that combining lightweight transformer-based event embeddings with frequency-aware weighting offers an effective and efficient alternative to complex sequential recommendation models. The proposed framework enables scalable modeling of user navigation behavior while preserving important patterns found in clickstream interactions. As a result, this approach delivers practical benefits for large-scale e-commerce platforms that require efficient and clear user behavior modeling.

Despite these promising results, several directions remain for future research. First, the current framework focuses exclusively on frequency-based weighting to capture behavioral intensity signals. Future work may investigate hybrid weighting strategies that jointly incorporate frequency, temporal recency, and contextual information to better model evolving user intent within sessions. Second, the present study formulates purchase prediction as a binary next-event classification task. Extending the framework to multi-class next-event prediction problems may enable modeling a broader range of user interaction outcomes, including product exploration, cart additions, and recommendation engagement.

Another promising direction is to enhance the quality of event representations through domain-adaptive embedding techniques. While this study uses pretrained lightweight transformer models, future work could investigate domain-specific fine-tuning strategies that adjust embedding models to the features of clickstream interaction data. Such adjustments could further improve the semantic representation of user navigation events and boost downstream predictive performance.

Acknowledgements. The authors would like to thank the Trendyol Research and Development Center for providing the industrial collaboration and data infrastructure that supported this study. This research paper used generative AI tools, including Google Gemini and ChatGPT, solely for improving language clarity, grammar, and overall readability of the manuscript. These tools were not used to generate scientific ideas, experimental designs, analyses, or research results. The authors retain full responsibility for the accuracy, originality, and integrity of the work.

References

1. Moe, W.W., Fader, P.S.: Dynamic conversion behavior at e-commerce sites. Manage. Sci. **50**(3), 326–335 (2004)
2. Sakar, C.O., Polat, S.O., Katircioglu, M., Kastro, Y.: Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and lstm recurrent neural networks. Neural Comput. Appl. **31**, 6893–6908 (2019)

3. Guo, H., Tang, R., Ye, Y., Li, Z., He, X.: Deepfm: a factorization-machine based neural network for ctr prediction, arXiv preprint arXiv:1703.04247 (2017)
4. Zhang, S., Yao, L., Sun, A., Tay, Y.: Deep learning based recommender system: a survey and new perspectives. ACM Comput. Surv. **52**(1), 1–38 (2019)
5. Sun, F., et al.: Bert4rec: sequential recommendation with bidirectional encoder representations from transformer. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 1441–1450 (Nov 2019)
6. Jannach, D., Ludewig, M., Lerche, L.: Session-based item recommendation in e-commerce: on short-term intents, reminders, trends and discounts. User Model. User-Adap. Inter. **27**(3), 351–392 (2017)
7. Quadrana, M., Cremonesi, P., Jannach, D.: Sequence-aware recommender systems. ACM Comput. Surv. **51**(4), 1–36 (2018)
8. Taşgetiren, N., Kılıç, T., Aktas, M.S.: On the use of graph embedding techniques for clustering user browsing navigational behaviours. Int. J. Web Grid Serv. **20**(4), 482–504 (2024)
9. Buyuktanir, T., Aktas, M.S.: On the use of embedding techniques for modeling user navigational behavior in intelligent prefetching strategies. Concurrency Comput. Pract. Exper. **37**(3), e8356 (2025)
10. Tasgetiren, N., Safak, I., Aktas, M.S.: An event frequency-inverse session frequency based weighting mechanism for data embedding methodologies. Softw. Pract. Exper. **55**(6), 983–1002 (2025)
11. Taşgetiren, N., cSafak, I., Aktaş, M.S.: Aging-based weighting for session classification in user behavior analysis, vol. 15886, pp. 16–34 (Jun. 2025)
12. Reimers, N., Gurevych, I.: Sentence-BERT: sentence embeddings using siamese BERT-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 3982–3992 (Nov. 2019)
13. Wang, L., et al.: Text embeddings by weakly-supervised contrastive pre-training, arXiv preprint arXiv:2212.03533 (Dec. 2022)
14. S. Xiao, Z. Liu, P. Zhang, N. Muennighoff, D. Lian, and J.-Y. Nie, "C-pack: Packed resources for general chinese embeddings," in *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Jul. 2024, pp. 641–649
15. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
16. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794
17. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Information Processing & Management **45**(4), 427–437 (2009)
18. Hidasi, B., Karatzoglou, A., Baltrunas, L., Tikk, D.: Session-based recommendations with recurrent neural networks, arXiv preprint arXiv:1511.06939 (Nov. 2015)
19. Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T., Ma, J.: Neural attentive session-based recommendation. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 1419–1428 (Nov 2017)
20. Quadrana, M., Karatzoglou, A., Hidasi, B., Cremonesi, P.: Personalizing session-based recommendations with hierarchical recurrent neural networks. In: Proceedings of the Eleventh ACM Conference on Recommender Systems, pp. 130–137 (Aug. 2017)
21. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781 (Jan. 2013)

22. Greenstein-Messica, A., Rokach, L., Friedman, M.: Session-based recommendations using item embedding. In: Proceedings of the 22nd International Conference on Intelligent User Interfaces. Association for Computing Machinery, pp. 629–633 (2017)
23. Perozzi, B., Al-Rfou, R., Skiena, S.: DeepWalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 701–710 (2014)
24. Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 855–864 (Aug. 2016)
25. Kang, W.-C., McAuley, J.: Self-attentive sequential recommendation. In: 2018 IEEE International Conference on Data Mining (ICDM), pp. 197–206. IEEE (Nov. 2018)
26. Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X., Tan, T.: Session-based recommendation with graph neural networks. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33(01), pp. 346–353 (2019)
27. de Souza Pereira Moreira, G., Rabhi, S., Lee, J.M., Ak, R., Oldridge, E.: Transformers4Rec: Bridging the gap between NLP and sequential/session-based recommendation. In: Proceedings of the 15th ACM Conference on Recommender Systems, pp. 143–153 (Sep. 2021)
28. Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., Zhou, M.: MiniLM: deep self-attention distillation for task-agnostic compression of pre-trained transformers. Adv. Neural. Inf. Process. Syst. **33**, 5776–5788 (2020)