Asia Pacific Journal of Information Systems
Vol. 33 No. 3 (September 2023), 737-767

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2023.33.3.737

Predicting Session Conversion on E-commerce: A Deep Learning-based Multimodal Fusion Approach

Minsu Kim[a], Woosik Shin[b], SeongBeom Kim[c], Hee-Woong Kim[d,*]

[a] *Data Scientist, LG UPlus Corporation, Korea*
[b] *PhD Candidate, Graduate School of Information, Yonsei University, Korea*
[c] *PhD Candidate, Graduate School of Information, Yonsei University, Korea*
[d] *Professor, Graduate School of Information, Yonsei University, Korea*

ABSTRACT

With the availability of big customer data and advances in machine learning techniques, the prediction of customer behavior at the session-level has attracted considerable attention from marketing practitioners and scholars. This study aims to predict customer purchase conversion at the session-level by employing customer profile, transaction, and clickstream data. For this purpose, we develop a multimodal deep learning fusion model with dynamic and static features (i.e., DS-fusion). Specifically, we base page views within focal visist and recency, frequency, monetary value, and clumpiness (RFMC) for dynamic and static features, respectively, to comprehensively capture customer characteristics for buying behaviors. Our model with deep learning architectures combines these features for conversion prediction. We validate the proposed model using real-world e-commerce data. The experimental results reveal that our model outperforms unimodal classifiers with each feature and the classical machine learning models with dynamic and static features, including random forest and logistic regression. In this regard, this study sheds light on the promise of the machine learning approach with the complementary method for different modalities in predicting customer behaviors.

Keywords: Purchase Conversion, Multimodal Fusion, Clickstream Data, Rfmc, Deep Learning

Ⅰ. Introduction

The proportion of online sales to total retail sales increased to 21.3% in 2020 from 10.7% in 2015 (Digital Commerce 360, 2021). Despite this substantial growth in the e-commerce market and increased traffic, online retailers still face a challenge in converting customers' visits to purchases (i.e., visit-to-purchase conversion). According to a survey, the average e-commerce visit-to-purchase conversion

*Corresponding Author. E-mail: kimhw@yonsei.ac.kr

rate is 2.27%, with conversion rates by industry ranging from 1% to 4% (Ogonowski, 2021). Online retailers have sought to implement diverse marketing strategies and decision support systems to improve their conversion rates. These efforts include e-coupon targeting, advertising campaigns, and product recommendation systems. Nevertheless, most retailers have seen little improvement in their conversion rates.

To address this low conversion rate issue, a thorough understanding of online customers' behavior has been of great importance to online retailers. Unlike the high tendency in offline customers' visits to result in purchases, most customer visits to e-commerce platforms end up browsing to search and compare products and opinions on them but only a small portion of the online customer visits result in making purchases. Accurate predictions of purchase behavior are fundamental to arriving at strategies to reverse these low visit-to-purchase conversion rates and in turn boost the performance of online retailers. If online retailers could predict the purchase likelihood of customers during each visit, they could target those customers rated most likely to make purchases with individualized marketing strategies, thereby reducing marketing costs and ultimately enhancing profitability. Customized marketing strategies based on precise predictions of customers' purchase likelihood could convert customers' browsing to purchases within a given visit, significantly influencing the revenue of online retail platforms. Therefore, advances in real-time predictions of purchase behavior have meaningful practical implications for online retailers.

Because of the growing availability of big customer data, scholars have attempted to identify diverse user behavior on online platform and predict customers' behavior and especially their purchase behavior on e-commerce platform (Chaudhuri et al., 2021; Koehn et al., 2020; Mokryn et al., 2019). Existing literature on customer purchase behavior could be divided into two streams of literature. The first focuses on the dynamics of browsing patterns during a site visit (i.e., session[1]) to predict purchases (Baumann et al., 2018; Bucklin and Sismeiro, 2009). Customers' browsing patterns encompass every aspect of their behavior in interactions with e-commerce sites that are captured by clickstream data. Previous literature in this stream predicts session-level purchases and next-page visits by tracking customers' dynamic browsing behavior based on this data (Koehn et al., 2020; Wu et al., 2015). The second stream of literature has examined customers' historical characteristics on online platforms. Compared with the other stream's concentration on dynamic browsing behavior during a site visit, this stream has focused on static customer-platform interaction[2] (i.e., customers' past interactions with the online retailer) by aggregating past visits and transactions. For example, Park and Park (2016) examined the patterns of customers' visits to predict purchase conversion. Existing works in this stream have investigated the impacts of customers' demographics on purchase intention (Law and Ng, 2016). Further, research in this stream has examined the patterns of customers' visits and purchases to predict a customer lifetime value (CLV) based on recency, frequency, and monetary value (RFM) (Fader et al., 2005).

Despite the vast of literature on online purchase behavior, these previous works have limitations. First, there is a paucity of research on purchase behavior prediction that encompasses customers' character-

1) In the online platform context, a session indicates a single visit to the platform.

2) We regard customers' demographics and customers' past visits and purchases as static customer-platform interaction because those are obtained as fixed values based on historical information before the focal visit, compared with dynamic browsing behavior during the focal visit on the online platform.

istics and the dynamic platform engagement aspects of e-commerce. Most previous studies on session-level purchase prediction have used clickstream data in focusing on the dynamics of browsing features (e.g., Baumann et al., 2018; Toth et al., 2017). However, customers' buying decisions depend not only on their current dynamic behavior but also on their past interaction patterns with the platform (Park and Park, 2016). To comprehensively understand a customer's purchase behavior within a given visit, both dynamic browsing patterns during the focal visit and the customer's past interactions should be considered simultaneously.

Second, although some extant studies considered dynamic platform engagement and customers' characteristics for session-level purchase behavior prediction (Chaudhuri et al., 2021), those works disregarded the dynamic browsing patterns of customers by aggregating them at the session level. This approach also has a limitation in that it is not applicable to a real-time purchase predictive method. Real-time predictions of purchase likelihood are crucial to enable online retailers to immediately implement their customized marketing strategies. To accomplish this, it is needed to properly process both dynamic platform engagement and static customer features that have different modalities. Specifically, the static attributes of customers are captured as a snapshot at the beginning of a given session, but dynamic browsing patterns within the focal session should be processed sequentially at the page-view level. Because the static and dynamic platform engagement features of customers are composed of different data sources and representations but might contain complementary information (Koehn et al., 2020), we could expect that a multimodal fusion approach would improve prediction results by taking both features into account. However, few works in this realm have adopted multimodal approaches that combine different modalities into a joint representation to predict customers' purchase behavior.

To address this gap, we undertook to develop a deep learning-based multimodal fusion method to predict session-level purchase behavior by comprehensively using both dynamic platform engagement and customers' static attributes. Specifically, we extracted the dynamic platform engagement features from clickstream data and the static features of customers from their profiles, visits, and transaction data. Then, we used a multimodal fusion approach with deep learning architectures to reflect different modalities and to improve predictive performance, including long short-term memory (LSTM), and multilayer perceptron (MLP).

We validated the performance of our model by using real-world e-commerce data over six months. Our results have important implications from both the academic and practical perspectives. Academically, we are among the first to propose a method for predicting customers' purchase behavior based on a deep learning-based multimodal fusion approach that combines both dynamic platform engagement behavior and the static attributes of customers. We also extend the literature on predicting customers' purchase behavior by adapting the recency, frequency, monetary value, and clumpiness (RFMC) measures that have been used to estimate customers' value. From a practical perspective, online retailers can use our approach that predicts the purchase likelihood of customers in real time. Those businesses with low conversion rates will find our method especially helpful because our model can help them improve their real-time marketing efficiency by sorting customers in a given session into those most likely and less likely to make a purchase, thereby improving sales and overall business performance.

Ⅱ. Conceptual Background

2.1. Dynamic Platform Engagement: Clickstream Data

Dynamic platform engagement is defined as every browsing behavior of customers on online platform sites. This behavior contains information about awareness, interest, and consideration of the customer's decision funnel. Clickstream data is the main source of dynamic platform engagement data that captures user's interactions with online platforms (Bucklin and Sismeiro, 2003). Previous literature employed clickstream data to predict a customer's next visit (Bogina and Kuflik, 2017), purchase conversion (Park and Park, 2016; Zhu et al., 2019), and product choice (Iwanaga et al., 2016). In addition, this data is useful for real-time prediction of customers' behavior in e-commerce (Bucklin and Sismeiro, 2009). Given that a clear behavioral difference exists between with- and without-purchase sessions on e-commerce (Baumann et al., 2018; Lu et al., 2005), customers' browsing patterns should be considered in predicting their purchases. Processing clickstream data for dynamic engagement features to predict customers' behavior is broadly categorized into two methods: clipping at every click and sequence classification.

First, the clipping at every click approach is to label each page view for prediction (VanderMeer et al., 2000). This approach is suitable for a standard classification problem using supervised machine learning because it processes page views into aggregated metrics with a tabular format (Koehn et al., 2020). Specifically, each page view of sessions is labeled as a data point and labeled page views within a session are concatenated and transformed into a two-dimensional matrix format as feature vectors. Although the clipping at every click approach is useful to extract features that reflect the connectivity of page views in a session, this method cannot capture chronological order of page views within a session.

The sequence classification approach uses a sequence of pageviews as a single instance with a single label. This method uses each page view as a feature to predict the session outcome (Koehn et al., 2020). Consequently, this approach can preserve the temporal order of page views within a session, thus making this approach more suitable for deep learning models such as recurrent neural network (RNN)-based architectures. These architectures sequentially process input features whereas classical machine learning models do not consider the temporal order of input features. Thus, recent studies have used RNN-based architectures for conversion classification based on clickstream data (e.g., Bogina and Kuflik, 2017; Sheil et al., 2018; Toth et al., 2017; Wu et al., 2015). This method also can incorporate diverse page-level features, including page content, time spent on a page, and the number of page views (Koehn et al., 2020). However, the predictive performance of the sequence classification approach suffers because of the overfitting and noisiness inherent in the complex structure of clickstream data, which consists of tremendous number of routes and different sizes of session sequences (Bigon et al., 2019).

In addition to these uses of clickstream data, our approach incorporates the use of the static attributes of customers to better understand the purchase decision-making processes of customers by supplementing the information on their browsing patterns with their historical and demographic features.

2.2. Customers' Static Features

Unlike dynamic engagement features, customers' static attributes are predetermined based on previous

interactions with online retailers. That is, these static features encompass demographics and patterns of past transactions and visits, which remain unchanged while users are browsing pages in e-commerce. These features may not be directly related to real-time (session-level) predictions. Nevertheless, they have attracted considerable attention because these attributes have significant effects on customers' behavior and loyalty (i.e., purchase behavior and CLV). Demographic characteristics, including age and gender, have been regarded as key factors for predicting customers' purchases and loyalty (Kim and Kim, 2004; Larivière and Van den Poel, 2005; Ndubisi, 2006; Sorce et al., 2005).

Customers' historical factors derived from data on their transactions and visits have been widely used to predict their value and behavior. Recency, frequency, and monetary value have been commonly used to estimate CLV. Moreover, RFM measures have been applied to predict customers' purchase and spending (Wei et al., 2010). Van den Poel and Buckinx (2005) found that historical purchase factors based on RFM measures are significant for online purchasing behavior. Furthermore, Zhang et al. (2015) proposed the RFMC framework to estimate CLV. Unlike traditional RFM, clumpiness captures bingeable customer activities by measuring the temporal intervals of customers' visits and purchases (Zhang et al., 2013). Previous works found a significant association of visit- and purchases-based clumpiness and a firm's marketing performance in terms of customer churn and firm marketing performance (Zhang et al., 2015). Because customers' static features contain meaningful information related to their past behavior and loyalty, we incorporated them into our study as part of a comprehensive approach to predict session-level purchase behavior.

2.3. Online Purchase Prediction

In this section, we provide a systematic review of previous literature closely related to our research. We review recent studies on predicting purchase behavior at session level that employs dynamic or static features. As discussed in previous sections, we focus on detailed modeling approaches of each study including whether dynamic or static features are used, which types of approach are employed to process clickstream data (dynamic platform engagement features) and types of static features are used (customers' static features), and which algorithms (models) are used. Moreover, we include applicability of models in real time scenarios. <Table 1> presents the summaries of reviewed literature.

<Table 1> shows that most previous studies employ a unimodal feature type (i.e., dynamic platform engagement feature or customers' static feature) to predict purchase behavior. First, given distinguished browsing patterns between sessions with purchases and without purchases (Baumann et al., 2018), recent studies have attempted to use clickstream data to predict purchase behavior. In terms of processing clickstream data, both clipping at every click and sequence classification approaches are widely employed. As the clipping at every click approach transforms page views within a session into tabular form, studies using such approach used traditional machine learning approaches including logistic regression, naïve regression, random forest and XGBoost (Baumann et al., 2018; Yeo et al., 2020). For example, Baumann et al. (2018) processed page views of session-level clickstream data as nodes in graph theory to extract graph metrics within sessions involving the structure, distance, and centrality of page views. Then, they used logistic regression to examine how these predictors affect customers' pur-

<Table 1> Reviews of Related Literature

Reference	Research Modeling Approach				
	Dynamic Platform Engagement Features (Clickstream data)	Customers' Static Features	Algorithms	Dependent Variable	Real-time Prediction
Baumann et al. (2018)	Clipping at every click	-	LR, RF, XGB	Purchase conversion	-
Yeo et al. (2020)	Clipping at every click	-	Naïve Regression	Purchase conversion	-
Bigon et al. (2019)	Sequence classification	-	LSTM	Purchase conversion	✓
Koehn et al. (2020)	Sequence classification	-	LSTM, GRU	E-coupon redemption	✓
Chaudhuri et al. (2021)	Clipping at every click	Customer attributes (frequency, age, gender, recency, and tenure)	DT, RF, SVM, ANN, DNN	Purchase conversion	-
Esmeli et al. (2022)	Clipping at every click (Aggregation)	Purchase frequency, visit frequency, dates of visit and device used	RF, DT, Bagging, DNN	Purchase conversion	-
Rahim et al. (2021)	-	RFM	DT, SVM, MLP	Repurchase behavior	-
Our Study	Sequence classification	Visit and purchase-level RFMC and demographics (age and gender)	Multimodal fusion Approach (LSTM + MLP)	Purchase conversion	✓

Note: Logistic regression (LR), random forest (RF), XGBoost (XGB), decision tree (DT), support vector machine (SVM), artificial neural network (ANN) and Deep neural network (DNN)

chase behavior. However, such approach has limitations in a real-time prediction scenario. To model temporal dependencies of page views, recent studies also adopt the sequence classification approach to process clickstream data. <Table 1> reports that corresponding studies employed RNN-based architectures including LSTM and GRU (Bigon et al., 2019; Koehn et al., 2020) as those algorithms are capable of modeling temporal dependencies of page views by sequentially processing input features. For example, Koehn et al. (2020) employ LSTM and GRU algorithms to predict shopping behavior (i.e., order values) by using the sequence classification approach for clickstream data. However, those studies using dynamic engagement features based on clickstream data have limitations that the prediction performance is low due to possible noisiness and do not consider customers' static features.

Next, with great interest in customer lifetime value in the domain of marketing, corresponding studies have employed various customer static features to predict purchase behavior including age, gender, tenure and RFM. Rahim et al. (2021) apply purchase-level RFM features to predict repurchase behavior. They used traditional machine learning models such as decision tree and support vector machine and deep

learning for tabular data such as MLP and simple DNN. Similarly, previous studies using both dynamic engagement features and customers' static features employ various customers' attributes including frequency, recency, tenure, age, gender, device used and date of visits to predict purchase behavior at session level (Chaudhuri et al., 2021; Esmeli et al., 2022).

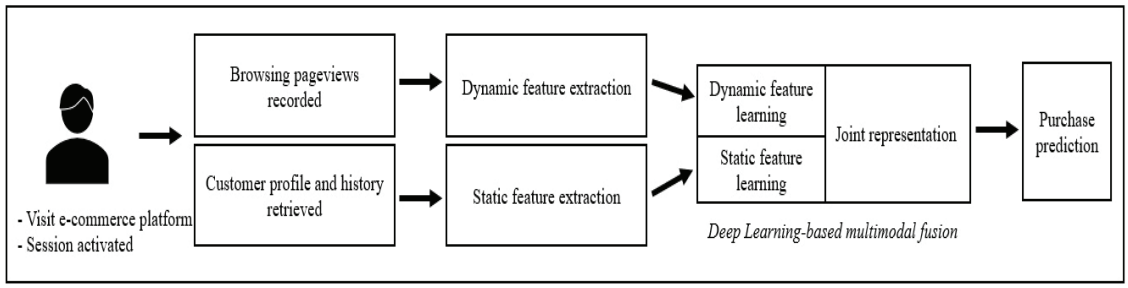
Last, a couple of studies employ both dynamic and static features to predict purchase behavior (Chaudhuri et al., 2021; Esmeli et al., 2022). However, those studies employ the clipping at every click approach to transform sequential page views into session-level feature vectors comparable to customers' static feature vectors at session level. However, this approach has limitations in that it cannot model sequential dependencies of page views within sessions.

Taken together, despite importance of both dynamic platform engagement and customers' static features in predicting purchase behavior at session level, only a few prior studies employ both modalities. Even those studies employing both feature types have limitations that they do not comprehensively encompass customers' static features and model chronological orders of page views in terms of dynamic engagement features. In comparison with previous literature, our study employs the sequence classification approach to extract dynamic features that reflect temporal dependencies at page-view level. In terms of customers' static features, we comprehensively extract customers' static features from transaction history data including visit and purchase-level RFMC, age and gender. Furthermore, to model different modalities of both features, we employ a multimodal fusion approach. This study is among the first to employ a multimodal fusion approach in the context of online purchase prediction. We present details of multimodal fusion approaches in the next section.

2.4. Multimodal Fusion Approach

A multimodal fusion approach aims to integrate data from different modalities into one representation for analysis tasks (Hu et al., 2019; Poria et al., 2017). By capturing complementary information, the multimodal fusion method can generate synergistic effects to increase the accuracy of overall predictions compared with the performance of single models of each modality (Zhang et al., 2020b). The multimodal feature fusion approach has been widely used for data with complementary information such as static and dynamic features (e.g., Liu et al., 2019; Wang et al., 2004). For example, texts and image data are applied for sentiment analysis and fake news detection (Zhang et al., 2020a; Yuan et al., 2021), and dynamic and static API call information is used to detect malware in cyberspace security (Han et al., 2019). Because dynamic platform engagement and customers' static features on e-commerce platforms indicate different modalities, the multimodal fusion approach is suitable for capturing the probabilistic correlations and is expected to have better predictive performance.

Methods for multimodal fusion can be classified into three types—late, early, and intermediate. Late fusion (or decision level fusion) is the method in which the features of each modality are analyzed independently, and the results of each analysis are combined to compute a final decision (Zhang et al., 2020a). Second is the early fusion approach. This method combines different modalities of raw data ahead of feature learning process for each modality (Dong et al., 2014; Glodek et al., 2013). Third, intermediate fusion (or feature level fusion) is the method that combines the early and late fusion strategies to avoid problems from each approach (Poria et al., 2015). This intermediate approach has advantages over early and late fusion because it can learn features within



<Figure 1> Research Overview

each modality and better capture correlations between features from different modalities (Gao et al., 2020).

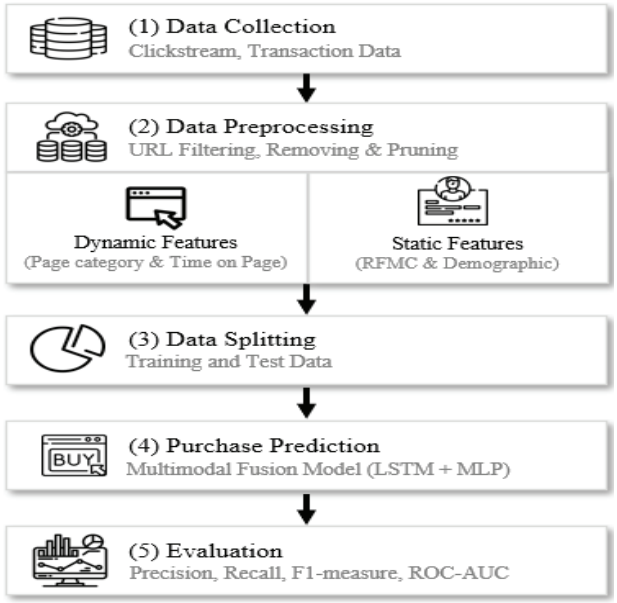
In terms of dealing with multimodal data, early fusion is limited in timing the synchronization of different modalities, and late fusion ignores correlations of features at the input level (Zhang et al., 2020b). In this regard, we adopt the intermediate fusion method as the best approach for combining different modalities of static and dynamic customer information captured from e-commerce data to predict customers' purchase behavior. Specifically, we used different neural network architectures to learn features from each modality at an early stage. We then use a joint representation derived from the features at the early stage for a purchase behavior prediction task. <Figure 1> shows an overview of the research framework in this study.

Ⅲ. Research Context and Data

For our study, we obtained data from K-Mall, a large online retailer in South Korea.[3] At the end of 2020, K-Mall had a monthly average of approximately 260,000 active users. The company sells numerous products, ranging from apparel and cosmetics to groceries, through multiple channels such as PCs, mobile apps, and mobile sites. We obtained clickstream and transaction data from the focal retailer. Specifically, our data set includes clickstream data with pageviews of customers' session information (e.g., user ID, session ID, browser, page URL, and time spent on each page) and transaction data with customers' demographics for 6-month period. The pages in the clickstream data consist of 16 major categories based on the page URLs provided by the company (see <Appendix A>). <Figure 2> presents flowchart of our analytical procedure for predicting purchase behavior using clickstream and transaction data.

We preprocessed the data as follows. First, we deleted incomplete sessions and those of only one page view; we regard these as "bouncers" uninterested in the website and having no intention of making a purchase. Second, we removed users whose channel was not a mobile browser, which consists of lower than 10% of the entire sessions because online customers' behavior may differ depending on the context (e.g., mobile or PC) (Wagner et al., 2020). Third, with-purchase sessions were pruned of confirmation-of-purchase pages (i.e., order page) because the goal of our model was to predict customers' purchase behavior at the session level. As a result, the cleansed clickstream data consisted of 217,063 page views with 47,732 unique sessions of 16,067

3) K-Mall is a pseudonym used for reasons of confidentiality.



<Figure 2> Flowchart of Analytical Procedure

distinct active customers. Of these cleansed sessions, 32,604 were of confirmed purchases and no purchase was observed in the rest 15,128 sessions. We then split the preprocessed data set into training and test set. The training and data sets consist of 33,412 and 14,320 sessions, respectively.

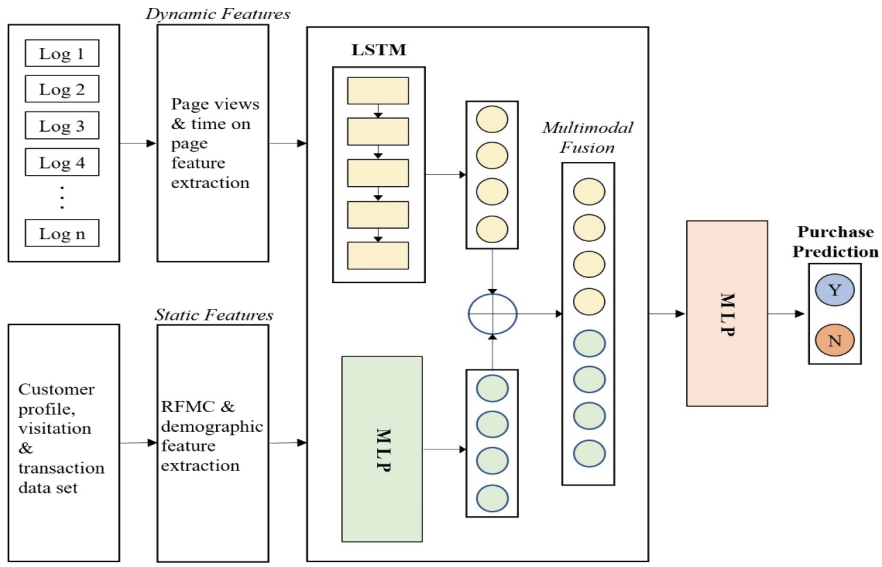
Ⅳ. Methodology

Dynamic and static feature extractions and multimodal fusion model in the proposed method are presented in <Figure 3>. We first processed data from clickstream, customers' profiles, transactions, and visits to extract customers' dynamic and static features. These features were then fed into the deep learning architectures with the multimodal fusion approach for a binary purchase classification.

4.1. Feature Extraction

The features for predicting customers' purchases in e-commerce are divided into two extraction processes according to customers' e-commerce behavior — static and dynamic feature extraction. When a customer lands on an e-commerce page, the customers' static features are retrieved based on visit and transaction records and demographics. Dynamic platform engagement features are sequentially captured by page views within single session at the e-commerce site. In summary, we processed data on transactions and visits to extract static features and clickstream data for dynamic features.

To capture customers' static features, we adopted the RFM analysis and the clumpiness measure in previous works (Fader et al., 2005; Zhang et al., 2013; Zhang et al., 2015). We computed purchase-based RFMC and visit-based frequency and clumpiness, and then coded those measures at a session level.



<Figure 3> Proposed Prediction Framework

Specifically, we defined the purchase-level recency (*Purchase-R*) as the number of days since the last purchase. *Purchase-R* thus represents the number of days without a purchase. Visit- and purchase-level frequencies (*Visit-F* and *Purchase-F*) were measured by the number of visits or transactions during the past eight weeks. We defined monetary value (*Monetary value*) as the total amount of transaction values for the past eight weeks. Finally, we captured visit-level clumpiness and purchase-level clumpiness (*Visit-C* and *Purchase-C*), which we defined as the extent to which visiting or buying intervals are regularly distributed (see details in <Appendix C>). Both clumpiness features were measured from 0 to 1, in which, if a customer had high clumpiness, then the customer had binge visiting and buying patterns. <Table 2> indicates the description of RFMC features and their measurement formula used in this study. We also used demographic information to extract customers' ages and gender. In sum, we obtained eight customer static features for our predictive modeling. The descriptive statistics of those features are presented in <Table 3>.

To extract dynamic platform engagement features, we used the sequence classification method to process clickstream data. We first prune sessions into length of 15 given that 95.6% of sessions are comprised of less than and equal to 15 pageviews. Then, we zero-pad sessions which have less than 15 pageviews. Pruning is suitable for real-time purchase prediction with incomplete sessions (Koehn et al., 2020). Each row of clickstream data that consists of a session ID, page category, and time-on-page corresponds to a single page view in a session. By using the cleansed clickstream data, we extracted entire page views of each session as well as time-on-pages. Notably, we observed there is a difference between with- and without-purchase sessions in that the former has more page views and is more prone to repeat product detail pages (see <Table B1>, <Appendix B>). This is in line with a previous finding that reoccurring product page visits are important predictors of purchase behavior (Baumann et al., 2018). Capturing such patterns of purchase sessions may have sig-

<Table 2> Description and Formula of RFMC Measurement

Features	Description	Formula of Measurement
Purchase-R	Number of days since the last purchase	$T_p - T_{p-1}$
Visit-F	Number of sessions (visits) during the past 8 weeks	$\sum F_v$
Purchase-F	Number of transactions during the past 8 weeks	$\sum F_p$
Monetary value	Total amount of transaction value during the past 8 weeks	$\sum M_p$
Visit-C	Variation in visiting intervals	$1 + \frac{\sum_{v=1}^{n+1} \log(x_v) * x_v}{\log(n+1)}$
Purchase-C	Variation in buying intervals	$1 + \frac{\sum_{p=1}^{n+1} \log(x_p) * x_p}{\log(n+1)}$

Note: $\mathbf{T_P}$ indicates dates of purchase session in session p; $\boldsymbol{F_V}$ and $\mathbf{F_P}$ denote the number of visits and purchases during the past eight weeks, respectively; $\mathbf{M_P}$ represents the total amount of purchase during the past eight weeks. For clumpiness measures, n denotes the number of visits and purchases, respectively, and $\mathbf{X_V}$ and $\mathbf{X_P}$ indicate the interevent times (IETs) of visits and purchases (see measurement details in <Appendix C>).

<Table 3> Summary Statistics of Customers' Static Features

Pageview Orders		#1	#2	#3	#4	#5	#6	⋯	#15
S1:	Page category	2	1	2	0	0	0	⋯	0
	Time on page	7	5	12	0	0	0	⋯	0
S2:	Page category	1	4	4	4	4	12	⋯	0
	Time on page	10	6	8	20	15	5	⋯	0

<Table 4> Examples of Pageviews and Time on Page Within Sessions

Features	Mean	SD
Purchase-R	35.536	28.772
Visit-F	14.213	34.019
Purchase-F	1.023	4.894
Visit-C	0.426	0.495
Purchase-C	0.534	0.495
Monetary value	31.105	162.073
Age	34.9	1.408
Gender (Female)	0.87	-

Note: The currency for Monetary value is in U.S. dollars (1 dollar = 1,100 Korean Won).

nificant predictive power. These dynamic engagement features and customer static features were incorporated into inputs for our proposed model. <Table 4> shows the examples of chronological orders of pageviews for dynamic feature learning.

4.2. Proposed Model

To predict session-level purchase behavior, we proposed a model based on a multimodal fusion with deep learning architectures. The proposed model aims to fuse dynamic platform engagement and customer static features into a joint representation for session-level purchase prediction. By using an intermediate fusion approach, our multimodal fusion model consisted of four steps—input data preprocessing, individual feature learning, fusion, and classification. In the input data preprocessing step, we prepared two modalities of dynamic and static features to feed into separate neural network architectures. For dynamic features, we used the sequence inputs of page views and time-on-page from clickstream data. We processed the page-view vector

sequence as $s = \{(p_1, t_1), \ldots, (p_l, t_l), \ldots, (p_L, t_L)\}$, where p_l and t_l denote the l-th page category and time-on-page, and L stands for session length. These session vectors were prepared for inputs of LSTM architecture that effectively analyzes the temporal dynamic dependencies of sequences (Hochreiter and Schmidhuber, 1997). For static features, we used eight features derived from demographics and RMFC measures by using customers' profiles, transactions, and visit data. These features were processed as the standard vector formats, $x = [x_1, x_2, \ldots, x_8]$, to be fed into MLP, which has been commonly used for tabular data.

In the individual feature learning step, input data preprocessed in the previous step was separately fed into the neural network architectures to learn different features, including MLP and LSTM. The MLP architecture for static feature learning was composed of three hidden layers with the rectified linear unit (ReLU) as the activation function:

$$\mathbf{R}^{(m)} = ReLU\left(W^{(m)} \cdot \mathbf{R}^{(m-1)} + b^{(m)}\right). \quad (1)$$

In (1), $\mathbf{R}^{(m)}$ represents the hidden state of the m-th layer; $\mathbf{R}^{(0)}$ as the input is the static feature matrix, $\mathbf{X}$; $W^{(m)}$ denotes the weight matrix of m-th layer; and $b^{(m)}$ is the bias vector. The output from the MLP architecture was used for static feature representation. In the MLP structures, we used multiple hidden layers instead of one hidden layer with many neurons because they capture nonlinear relationships better in the data set (Saide et al., 2015).

For dynamic feature learning, we used LSTM architecture. The architecture for the dynamic feature is composed of two layers of the LSTM. The LSTM consists of a memory cell and several gates, including an input gate, a forget gate, an output gate, and a hidden state (see LSTM cell details in <Appendix D>), which can capture long-range dependencies by controlling the proportion of information from a previous state and the input to the memory cell (Hochreiter and Schmidhuber, 1997). The session vectors, S, were sequentially processed as inputs for the LSTM layer. The output from the LSTM was used for the dynamic feature representation:

$$\mathbf{H}^{(n)} = LSTM(\{\langle p_1, t_1\rangle, \ldots, \langle p_l, t_l\rangle, \ldots, \langle p_L, t_L\rangle\}). \quad (2)$$

In (2), ***LSTM*** denotes the LSTM network at the n-th layer and $\langle p_l, t_l\rangle$ represents input vectors at timestep l.

In the fusion step, the static feature representation at the m-th layer, $\mathbf{R}^{(m)}$, and the dynamic feature representation at the n-th layer, $\mathbf{H}^{(n)}$, were concatenated to build a joint representation:

$$\mathbf{J} = (\mathbf{R}^{(m)} \parallel \mathbf{H}^{(n)}). \quad (3)$$

In (3), "$\parallel$" indicates a concatenation operator. The joint representation, $\mathbf{J}$, generated from two modalities in the fusion stage was finally fed into the MLP to produce the session-level classification result (i.e., purchase and nonpurchase). Our model in the classification step consists of two layers of fully connected neural networks. The classifier layer is as follows:

$$\hat{y} = \sigma(W_d \cdot \mathbf{J}_d + \mathbf{b}_d). \quad (4)$$

In (4), $\mathbf{J}_d$ represents the hidden layer output of d-th layer in the fusion stage. W_d represents the weight matrix and b_d is the bias vector at d-th layer. Our model adopted the negative log likelihood as the loss function to be minimized.

In the following section, we report the performance evaluation of the proposed model and compare it with baseline models to validate our model. To devel-

INPUT for Dynamic Features: Sequential pageviews of a customer at session **S** such that $s_l \in \mathbf{S}$ and s_l = (page category p_l, time on page t_l), (l =1,2,3…,15) and S ≠Ø
INPUT for Static Features: Compute visit and purchase-level RFMC of a customer at session visit and concatenate with demographic features as static features $x_i \in \mathbf{X}$
x_i = [Purchase-R, Visit-F, Purchase-F, Monetary value, Visit-C, Purchase-C, gender, age]
OUPUT: Binary outcome y for purchase at a session
- Dynamic feature learning step
  for S_i in **S**:
    DynamicOutput ← LSTM($[s_1, \dots, s_l, \dots, s_L]$):
    $i_l = \sigma(W_i \cdot [s_l, h_{l-1}] + b_i)$, where $i_l =$ *input gate at page timestamp* l; $s_l =$ *input vector at* l; $h_{l-1} =$ *hidden state vector at* $l-1$; $W =$ *weight matrixes*; $b =$ *bias vectors*; *and* $\sigma =$ *sigmoid function*
    $f_l = \sigma(W_f \cdot [s_l, h_{l-1}] + b_f)$, where $f_l = forget\ gate\ at\ l$
    $o_l = \sigma(W_o \cdot [s_l, h_{l-1}] + b_o)$, where $o_l = output\ gate\ at\ l$
    $\tilde{c}_l = tanh(W_c \cdot [s_l, h_{l-1}] + b_c)$, where $\tilde{c}_l = update\ gate\ at\ l$
    $c_l = c_{l-1} \circ f_l + \tilde{c}_l \circ i_l$, where $c_l = state\ gate\ at\ l$
    $h_l = o_l \circ \tanh(c_l)$, where $h_l = hidden\ state\ vector\ at\ l$
- Static feature learning step
  for x_i in X:
    StaticOutput← MLP(x_i):
    $R^{(m)} = ReLU(W^{(m)} \cdot R^{(m-1)} + b^{(m)})$, where $R^{(m)} =$ *hidden state vector at layer* m; $R^{(0)} = input\ vector\ x_i$; $ReLU =$ *Rectified linear unit activation function*; $W^{(m)} = weight\ matrices\ at\ layer\ m$; $W^{(m)} = weight\ matrices\ at\ layer\ m$; *and* $b^{(m)} =$ *bias vectors at layer* m
- Fusion step
  Concatenate *DynamicOutput* and *StaticOutput*
  $J = (DynamicOutput \parallel StaticOutput)$, *where* $\parallel$ *is a concatenation operator*
  $\hat{y} = logistic(ReLU(W \cdot J + b))$, where $logistic(x) = \frac{1}{1+e^{-(\beta \cdot x)}}$;
  $W =$ *weight matrix*; $\beta =$ *weight vector*; *and* $b =$ *bias vector*
Update β *and* W *to minimize loss:* $L = -log\ (y - \hat{y})$

<Figure 4> Pseudocode of Proposed Model

op the proposed models, we used Keras 2.8.0, a neural networks library. Keras library acts as an interface for the TensorFlow library 2.8.2, an open-source platform for machine learning developed by Google Brain. Our proposed model is summarized in <Figure 4>.

4.3. Evaluations

We used diverse metrics for classification tasks to evaluate predictive performance. First, we used precision, recall, and the F-measure.

Given that our data consists of a myriad of sessions without purchase, these evaluation metrics are relevant and widely used for imbalanced data (He and Ma, 2013). Second, the area under the receiver operating characteristic (ROC) curve (AUC) was used to compare the performance of the models (Bradley, 1997). Compared with the first performance measures, the AUC allows comparison of the aggregate measures of performance at various threshold settings. Third, we applied cumulative gains and lift charts (Jamal and Bucklin, 2006), which are popular business value measurements (Ling and Li, 1998). These assessment approaches have been widely used in the literature on purchase behavior prediction (e.g., Baumann et al., 2018; Park and Park, 2016).

$$precision = \frac{TP}{TP+FP}. \quad (6)$$

$$recall = \frac{TP}{TP+FN}. \quad (7)$$

$$F\text{-}measure = 2 * \frac{precision * recall}{precision + recall}. \quad (8)$$

<Table 5> Network Structure and Hyper-Parameter Setting

Layer	Hyperparameters	Search Space	Selected Values
MLP in Individual Feature Learning	Number of hidden layers	1-3	3
	Number of neurons	32,64,128	32
LSTM in Individual Feature Learning LSTM	Number of hidden layers	1-3	2
	Number of neurons	32,64,128	64
Fusion Layer	Number of hidden layers	-	1
	Number of neurons	32,64,128	32
	Optimizer	Adam, RMSProp	Adam
	Batch size	32,64	32
	Activation Function in hidden layers	ReLU, Tanh	ReLU
	Activation Function in output layer	-	SoftMax

Specifically, the cumulative gains chart indicates the percentage of purchase cases by targeting a certain percentage of the population, which is ordered according to a model's estimated purchase probability. Based on the cumulative gains chart, the lift measure is defined as the ratio of the cumulative gains from a model to those from a random sample. In this regard, those measurements are directly related to the profitability of marketing practice.

Ⅴ. Results

5.1. Comparison with Baseline Models

Our model uses LSTM and MLP to base its multimodal feature fusion on dynamic and static features. We considered six baseline models using dynamic or static features to compare the performance of the proposed model. First, we used bi-directional LSTM (BiLSTM) for dynamic feature learning in multimodal fusion model (see architecture of BiLSTM in <Appendix D>). Recent studies in various prediction contexts have shown improvement in performance by using BiLSTM architecture (Siami-Namini et al., 2019; Yang and Wang, 2022). We do not use BiLSTM approach in the main proposed model as it is not applicable in a real-time prediction task, which is our main interest of purchase prediction model development. This approach employs future browsing information (i.e., page views after the current page view) to predict the outcome of focal session but future page views are not available in live scenario (Koehn et al., 2020). In this regard, we used this approach as an advanced approach to compare prediction performances with our proposed model.

Second, we used LSTM and BiLSTM classifiers using dynamic platform engagement features. Third, we used an MLP classifier with customer static features. If either of these unimodal baseline models were to outperform our proposed model, then we would not have to proceed to build a complex multimodal fusion model with dynamic and static features, which incur higher computation costs.

Third, we considered an alternative approach that used traditional machine learning algorithms. In comparative testing of the *clipping at every click* method that processes clickstream data into aggregated

<Table 6> Performances of Different Hyper-Parameters

Parameter Settings		Performance		
Layers for Dynamic Feature Learning	Layers for Static Feature Learning	Precision	Recall	F-Measure
LSTM	MLP			
1	1	0.863	0.973	0.915
	2	0.867	0.99	0.924
	3	0.879	0.978	0.926
2	1	0.864	0.979	0.918
	2	0.874	0.984	0.926
	3	**0.884**	**0.974**	**0.927**
3	1	0.861	0.993	0.922
	2	0.878	0.977	0.925
	3	0.879	0.977	0.925
BiLSTM	MLP			
2	1	0.878	0.968	0.921
	2	0.888	0.964	0.925
	3	**0.883**	**0.978**	**0.928**
3	1	0.868	0.983	0.922
	2	0.886	0.968	0.925
	3	0.887	0.959	0.922

features, we used a random forest classifier to predict purchases. This approach has been used for purchase classification tasks (e.g., Baumann et al., 2019; Moe and Fader, 2004). Following the previous studies, we used duration of session, number of clicks by page categories, age, gender, RFMC features for predicting purchase behavior. Finally, we used a logistic regression classifier with historical factors to evaluate its predictive powers, including customers' visits and purchase-level recency and frequencies. Earlier works in this realm indicated that historical factors are a strong classifier (Van den Poel and Buckinx, 2005).

5.2. Model Performance

We first examined different configurations of our proposed model architecture. The selected values of hyperparameter settings are present in <Table 5>. The prediction performance metrics of 15 different configurations (changing layers for dynamic and static feature learning) based on the proposed model is presented in <Table 6>. In addition to LSTM approach, we report BiLSTM approach for dynamic feature learning in the proposed model. As a result, we observe that the model with two LSTM layers for dynamic feature learning and three MLP layers perform the best although overall performance of other configurations also performed better than baseline models.

<Table 7> summarizes the precision, recall, and F-measure of our model and the baseline models. We can see that multimodal fusion models (i.e., LSTM+MLP and BiLSTM+MLP) outperformed unimodal classifiers in precision and in the F-measure. The results reveal the importance of considering both dynamic and static features to predict customers' be-

<Table 7> Comparison of Model Performance

Model	Precision	Recall	F-Measure	AU-ROC
MLP	0.836	0.99	0.908	0.84
Random forest	0.777	0.833	0.804	0.69
Logistic regression	0.667	0.99	0.799	0.72
LSTM	0.762	0.929	0.837	0.73
BiLSTM	0.766	0.932	0.841	0.74
DS-Fusion (LSTM+MLP)	**0.884**	**0.974**	**0.927**	**0.91**
BiLSTM + MLP	0.883	0.978	0.928	0.91

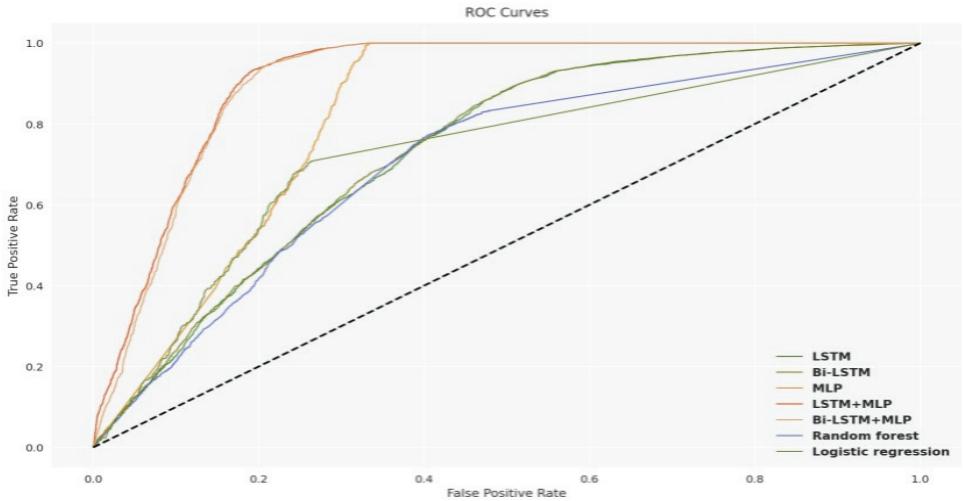
havior through direct comparisons of our model with unimodal classifiers such as MLP, LSTM and BiLSTM using static or dynamic features. The DS-Fusion model had a higher F-measure (0.927) than either of these baseline models (0.908, 0.837 and 0.841, respectively).

The results indicate that the baseline models with the MLP and logistic regression had higher recall rates than our model. However, these baseline models were much less precise, returning many false positives. Such a shortcoming by these baseline models would be a serious problem in target marketing through erroneously selecting as potential buyers online visitors with low intention to purchase. In addition, the results reveal that the F-measure of the model with the static feature (MLP) is higher than that of the logistic regression model (i.e., 0.908 vs. 0.799). And the F-measure of the model with the dynamic features (LSTM) is higher than that of the random forest classifier that uses the *clipping per every click* approach (i.e., 0.837 vs. 0.799).

We also compared the performance of our model with the baselines by using ROC curves that consider various classification threshold settings. The ROC curves were plotted with a true positive rate on the x-axis against the false positive rate on the y-axis, depending on the classification threshold. <Figure 5> presents comparisons of these ROC curves. The curve of our model approaches closer to the top left of the graph than any of the baseline models. In addition, the AUC in <Table 7> confirms that the multimodal fusion models performed better (i.e., 0.91) than the baseline models by at least 7%.

For further validation of our model, we conducted a t-test using 10-fold cross-validation (Wang and Xu, 2018). Specifically, we obtained 10 F-measure scores for each model from the cross-validation. Then, we conducted a t-test to compare the F-measure of our model with those of the baseline models. The results indicate that the F-measure of the DS-Fusion model is significantly higher than that of traditional machine learning models and unimodal classifiers (i.e., MLP, LSTM and BiLSTM). For instance, the difference between the F-measure of our model (mean = 0.935) and the best baseline model— the MLP model with static features (mean = 0.92; $t = 3.203$, $p < 0.05$)—is statistically significant. However, in terms of multimodal fusion models (i.e., LSTM+MLP and BiLSTM+MLP), the results of t-tests show that the performance difference between two models is not statistically significant ($t = -0.2365$, $p = 0.816$). This implies that the prediction performances of both models are not different but comparable.

For real time prediction, the model should be fast enough to produce prediction likelihood to be effective in marketing strategies. <Table 8> reveals summary statistics of prediction runtimes of the proposed



<Figure 5> Comparison of ROC-Curves of Models

model and baseline models. Specifically, two models using multimodal fusion approaches (i.e., LSTM + MLP and Bi-LSTM + MLP) are relatively slower than the unimodal models (i.e., LSTM, BiLSTM and MLP). However, the average runtime of the proposed model (LSTM + MLP) indicates 44.826 milliseconds, sufficiently fast to be employed for real-time prediction.

Furthermore, we assessed effects of different feature sets on prediction performance by designing eight scenarios involving the main feature sets. For static features, we used baseline features (gender and age), visit-level features (*Visit-F* and *Visit-C*), and purchase-level RFMC features (*Purchase-R, Purchase-F, Monetary value,* and *Purchase-C*). For dynamic features, we used different sizes of dynamic input features by pruning page length (i.e., 3 to 15 pageviews and time on page) as follows:

<Table 8> Prediction Runtimes of Models

Model	Mean of Runtimes	S.D. of Runtimes
MLP	39.340	15.664
LSTM	41.145	18.524
BiLSTM	41.417	19.601
DS-Fusion (LSTM+MLP)	**44.826**	**20.189**
BiLSTM+MLP	46.544	24.345

Note: Runtime is in milliseconds

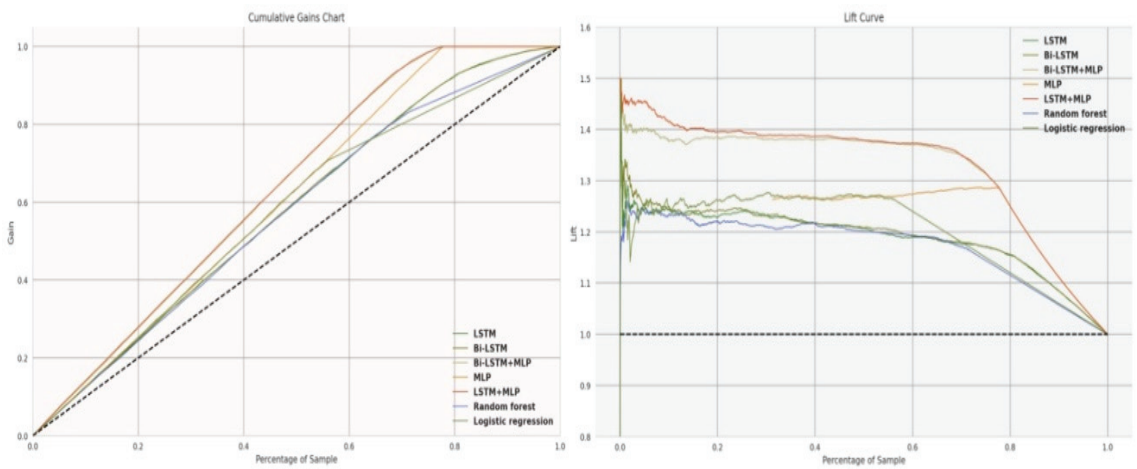
- *Scenario A*: Dynamic features (up to 15 pageviews and time on page) + gender + age
- *Scenario B*: Dynamic features (up to 15 pageviews and time on page) + gender + age + 2 visit-level RFMC features
- *Scenario C*: Dynamic features (up to 15 pageviews and time on page) + gender + age + 4 purchase-level RFMC features
- *Scenario D* (Full model): Dynamic features (up to 15 pageviews and time on page) + entire static features (8 static features)
- *Scenario E*: Dynamic features (up to 12 pageviews and time on page) + entire static features
- *Scenario F*: Dynamic features (up to 9 pageviews and time on page) + entire static features
- *Scenario G*: Dynamic features (up to 6 pageviews and time on page) + entire static features
- *Scenario H*: Dynamic features (up to 3 pageviews and time on page) + entire static features

To assess the performance of each scenario, we employed the proposed multimodal fusion model. <Table 9> presents the prediction performance of each scenario. It shows that Scenario D using entire dynamic and static features outperforms the other scenarios. In terms of static features, we can observe that both visit and purchase-level RFMC features (Scenario B and C) significantly improve prediction performances compared to Scenario A based on baseline static features (i.e., gender and age). When it comes to dynamic features, we can see slight but gradual improvement in prediction performances as more pageviews are employed for dynamic feature learning (Scenario D-H).

<Table 9> Comparisons of Impacts of Feature Combinations

Scenarios	Precision	Recall	F-Measure
Scenario A	0.766	0.928	0.839
Scenario B	0.856	0.943	0.898
Scenario C	0.886	0.96	0.922
Scenario D	**0.884**	**0.974**	**0.927**
Scenario E	0.869	0.992	0.926
Scenario F	0.871	0.972	0.919
Scenario G	0.867	0.967	0.914
Scenario H	0.866	0.961	0.911

In addition to prediction performances, we evaluate the models in terms of business value. <Figure 6> illustrates the cumulative gains and lift curves, respectively. In <Figure 6>, the cumulative gains chart shows the ratio of the with-purchase sessions if the marketing initiative targets a certain percentage of sessions. The chart indicates that our model and the baseline models are all significant improvements over the results with random targets, which are represented as a diagonal line. For example, our model allows us to identify 69% of potential buying sessions by targeting 50% of the samples, which indicates that our model improves 38% ((0.69 – 0.5) / 0.5), compared with the random sample. Moreover, our model (DS-Fusion) has an obviously steeper curve than those of the baseline models. The lift chart in <Figure 6> indicates that the lift index of our model exceeds that of the baseline models until approximately 78% of the samples are targeted. These results indicate that our model outperforms the baseline models at identifying potential buyers.



<Figure 6> Comparison of Cumulative Gains and Lift Index curves of Models

Ⅵ. Discussion

6.1. Discussion of Findings

Because of the extremely low purchase conversion rate among online customers, a thorough understanding of online purchase behavior is of great importance to online retailers. They would especially benefit from the capability to accurately predict the purchase behavior of customers during a given session. Thus equipped, marketing managers then could target these susceptible customers with customized and targeted marketing, thus facilitating visit-to-purchase conversions. In addition, adopting advanced machine learning algorithms has become imperative because of their capability to extract insightful information from the abstract and complex representations of data, leading to high predictive performance (Najafabadi et al., 2015). Motivated by such practical concerns and a gap in the previous literature, we used dynamic platform engagement and customers' static features to develop a deep learning-based multimodal fusion model to predict session-level purchase behavior.

Our experimental results show our multimodal model outperforms unimodal classifiers. This notion implies that dynamic platform engagement and customers' static attributes play complementary roles in explaining customers' buying behavior, resulting in improved predictive power (Chaudhuri et al., 2021). Our results show that the deep learning-based multimodal fusion approach can appropriately capture and learn representations of different modalities.

Although recent studies have shown improvement in performance by using BiLSTM over LSTM architecture (Siami-Namini et al., 2019), our results show that the performance of the proposed model using LSTM for dynamic feature learning is comparable with that of other multimodal fusion model using BiLSTM. In addition, given that BiLSTM approach is less applicable in a real-time prediction task, we can suggest that LSTM approach is more suitable in real-time purchase prediction task. Lastly, our results show that the deep learning models such as LSTM, BiLSTM and MLP have higher predictive power than the traditional machine learning models such as random forest and logistic regression. These results further support previous studies that have shown deep learning-based models have greater predictive power for large data sets than conventional machine learning models (Chaudhuri et al., 2021; Lee and Choeh, 2014). This is because the deep neural network structure better accommodates learning nonlinear relationships.

6.2. Implications for Research and Practice

This study has several important implications for academics and practice. First, from an academic perspective, our study is among the first to propose a deep learning-based multimodal fusion method in the context of online purchase prediction. By using actual e-commerce data, our multimodal fusion method empirically demonstrated its predictive superiority in comparisons with unimodal models and conventional machine learning models. These comparisons supported arguments advanced in previous studies for the importance of multimodal fusion approaches to enhance overall predictive performance because these methods can efficiently learn different modalities of data (Rastgoo et al., 2019; Zhang et al., 2020b). However, literature on purchase predictions of online customers has still struggled and rarely attempted to combine different modalities of customers' features. In this regard, our proposed model using the multimodal fusion method may pave the way for predictions of online purchase behavior by effectively combining two different modalities in

online customer data sets (i.e., clickstream data and transaction data).

Second, our study contributes to the literature on predicting customers' behavior by incorporating CLV in marketing domain into computational methods. We use the RFMC analysis to predict session-level purchase behavior. The RFMC analysis has been widely applied to customer segmentation and CLV estimation (Chen et al., 2012; Zhang et al., 2015). Few studies have adopted RFMC measures to enhance predictive performance in purchase prediction context (Moro et al., 2015). This study has demonstrated that those RFMC-derived features have a high predictive power for customers' purchase behavior by extracting features of customer historical interactions with e-commerce platforms. Unlike dynamic platform engagement features, features derived from the RFMC analysis can be easily extracted from transaction and visit data. Hence, we can argue that this study paves new ways for adapting RFMC analysis to predict customers' diverse behavior.

Third, the usage of advanced deep learning and machine learning techniques in our study reflects the increased interest in artificial intelligence and big data applications in diverse contexts, especially in online retail and customers' behavior (Herhausen et al., 2020; Rust, 2020). We provide further evidence for the relevancy of applying deep learning techniques for large data sets by comparing the predictive performance of deep learning algorithms with those of machine-learning algorithms. Future research in this domain could consider these deep learning methods to enhance predictive power, compared to conventional alternatives.

This study has important practical implications. First, as our model outperformed diverse baseline models and produced reliable predictive performance, this method can be applied to the actual e-commerce market. Using our method, online retailers can predict customer purchase behavior in real time. Based on computed purchase likelihood at session level, e-commerce platforms can conduct personalized target marketing initiatives in real time. This allows marketing managers at online retailers to reduce unnecessary costs and to maximize their revenues by better sorting customers into targets.In addition, our comprehensive use of customer transaction data for static feature learning shows significant improvement in prediction performance. With increased privacy concerns of customers, e-commerce confronts with difficulties in collecting personal information. In this regard, our predictive results can provide guidelines for collecting data to be effectively applied for marketing strategies.

6.3. Limitations and Future Research

Despite the significant contributions of this study, this study has some limitations, which offer valuable opportunities for future research. First, we employed clickstream and transaction data of customers from a large e-commerce in South Korea. Therefore, future research could be conducted using sizable real-world data sets from different countries to further generalize our results. As customer behavior might differ depending on their cultural backgrounds and demographics, the predictive performances might not be generalized. In addition, future studies could investigate the fusion of different sources of customer data, such as images, text, and videos, in order to improve the accuracy and comprehensiveness of predictive models. This is especially relevant given the growing trend of using diverse data sources in IS research (Sun et al., 2022). Second, regarding our study design, we focus on browsing patterns of the focal session in predicting purchase behavior.

However, future research could expand upon this by also considering browsing patterns from previous sessions and using dynamic feature learning techniques, such as concatenating adjacent sessions. Additionally, while we utilized LSTM, BiLSTM, and MLP for multimodal feature fusion, future research could explore alternative network architectures to improve predictive performance. One possible option is to apply dynamic features to a deep transformer architecture, which utilizes self-attention mechanisms to capture complex dynamics in sequential data. Last, our proposed model is limited in terms of interpretability of prediction results, which may impede practical application in business settings. Thus, we urge future research to consider developing explainable models using surrogate explanatory models. These models are expected to provide deeper domain understanding and enable proactive decision-making based on the model's predictions (Choi et al., 2022).

Ⅶ. Conclusion

In this study, we developed a deep learning-based multimodal fusion model (DS-Fusion model) for predicting purchase behavior. We first extracted dynamic platform engagement and customers' static attributes based on the data generated through customers' browsing patterns and their historical interactions with an e-commerce platform. We then proposed the multimodal fusion method with deep learning architectures to learn individual modalities and obtain joint feature representations, thereby predicting a session-level purchase behavior. The results of the experiment show our deep learning-based multimodal fusion model outperformed diverse baseline models that were based on deep learning and on conventional machine-learning algorithms. Marketing managers and platform designers in online retail platforms could apply our predictive model to improvise personalized marketing initiatives such as real-time e-coupons and to improve product recommendation engines that reflect customers' purchase likelihood in each session. These applications will enhance overall operating efficiency for online retailers without disruptive changes to their systems.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (No. 2022R1F1A1073133).

<References>

[1] Baumann, A., Haupt, J., Gebert, F., and Lessmann, S. (2018). Changing perspectives: Using graph metrics to predict purchase probabilities. *Expert Systems with Applications*, *94*, 137-148. https://doi.org/10.1016/j.eswa.2017.10.046

[2] Baumann, A., Haupt, J., Gebert, F., and Lessmann, S. (2019). The price of privacy. *Business & Information Systems Engineering*, *61*(4), 413-431. https://doi.org/10.1007/s12599-018-0528-2

[3] Bigon, L., Cassani, G., Greco, C., Lacasa, L., Pavoni, M., Polonioli, A., and Tagliabue, J. (2019). Prediction is very hard, especially about conversion. Predicting user purchases from clickstream data in fashion e-commerce. *arXiv preprint arXiv:1907.00400*.

[4] Bogina, V., and Kuflik, T. (2017). Incorporating dwell time in session-based recommendations with recurrent neural networks. In *CEUR Workshop Proceedings*, (Vol. 1922 pp. 57-59).

[5] Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145-1159. https://doi.org/10.1016/S0031-3203(96)00142-2

[6] Bucklin, R. E., and Sismeiro, C. (2003). A model of web site browsing behavior estimated on clickstream data. *Journal of Marketing Research*, *40*(3), 249-267. https://doi.org/10.1509/jmkr.40.3.249.19241

[7] Bucklin, R. E., and Sismeiro, C. (2009). Click here for internet insight: Advances in clickstream data analysis in marketing. *Journal of Interactive Marketing*, *23*(1), 35-48. https://doi.org/10.1016/j.intmar.2008.10.004

[8] Chaudhuri, N., Gupta, G., Vamsi, V., and Bose, I. (2021). On the platform but will they buy? Predicting customers' purchase behavior using deep learning. *Decision Support Systems*, *149*, 113622. https://doi.org/10.1016/j.dss.2021.113622

[9] Chen, D., Sain, S. L., and Guo, K. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. *Journal of Database Marketing & Customer Strategy Management*, *19*(3), 197-208. https://doi.org/10.1057/dbm.2012.17

[10] Choi, H., Kim, D., Kim, J., Kim, J., and Kang, P. (2022). Explainable anomaly detection framework for predictive maintenance in manufacturing systems. *Applied Soft Computing*, *125*, 109147.

[11] Digital Commerce 360. (2021). US ecommerce grows 44.0% in 2020. Retrieved from https://www.digitalcommerce360.com/article/us-ecommerce-sales/#:~text=Online%20spending%20represented%2021.3%25%20of,to%20Digital%20Commerce%20360%20estimates.&text=Online's%20share%20of%20total%20retail,2019%20and%2014.3%25%20in%202018

[12] Dong, Y., Gao, S., Tao, K., Liu, J., and Wang, H. (2014). Performance evaluation of early and late fusion methods for generic semantics indexing. *Pattern Analysis and Applications*, *17*(1), 37-50. https://doi.org/10.1007/s10044-013-0336-8

[13] Esmeli, R., Bader-El-Den, M., and Abdullahi, H. (2022). An analyses of the effect of using contextual and loyalty features on early purchase prediction of shoppers in e-commerce domain. *Journal of Business Research*, *147*, 420-434. https://doi.org/10.1016/j.jbusres.2022.04.012

[14] Fader, P. S., Hardie, B. G., and Lee, K. L. (2005). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, *42*(4), 415-430. https://doi.org/10.1509/jmkr.2005.42.4.415

[15] Gao, J., Li, P., Chen, Z., and Zhang, J. (2020). A survey on deep learning for multimodal data fusion. *Neural Comput*, *32*(5), 829-864. https://doi.org/10.1162/neco_a_01273

[16] Glodek, M., Reuter, S., Schels, M., Dietmayer, K., and Schwenker, F. (2013). Kalman filter based classifier fusion for affective state recognition. In Z. H. Zhou, F. Roli, and J. Kittler (Eds.), *Multiple Classifier Systems*. Berlin, Heidelberg.

[17] Graves, A., and Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, *18*(5), 602-610. https://doi.org/10.1016/j.neunet.2005.06.042

[18] Han, W., Xue, J., Wang, Y., Huang, L., Kong, Z., and Mao, L. (2019). MalDAE: Detecting and explaining malware based on correlation and fusion of static and dynamic characteristics. *Computers & Security*, *83*, 208-233. https://doi.org/10.1016/j.cose.2019.02.007

[19] He, H., and Ma, Y. (2013). *Imbalanced Learning: Foundations, Algorithms, and Applications*. Wiley-IEEE Press.

[20] Herhausen, D., Miočević, D., Morgan, R. E., and Kleijnen, M. H. P. (2020). The digital marketing capabilities gap. *Industrial Marketing Management*, *90*, 276-290. https://doi.org/10.1016/j.indmarman.2020.07.022

[21] Hochreiter, S., and Schmidhuber, J. (1997). Long

short-term memory. *Neural Computation*, *9*(8), 1735-1780. https://doi. org/10.1162/neco.1997.9.8.1735

[22] Hu, D., Wang, C., Nie, F., and Li, X. (2019). Dense multimodal fusion for hierarchically joint representation. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton, UK.

[23] Iwanaga, J., Nishimura, N., Sukegawa, N., and Takano, Y. (2016). Estimating product-choice probabilities from recency and frequency of page views. *Knowledge-Based Systems*, *99*, 157-167. https://doi.org/10.1016/j.knosys.2016.02.006

[24] Jamal, Z., and Bucklin, R. E. (2006). Improving the diagnosis and prediction of customer churn: A heterogeneous hazard modeling approach. *Journal of Interactive Marketing*, *20*(3-4), 16-29. https://doi.org/10.1002/dir.20064

[25] Kim, E. Y., and Kim, Y. K. (2004). Predicting online purchase intentions for clothing products. *European Journal of Marketing, 38*(7), 883-897. https://doi.org/10.1108/03090560410539302

[26] Koehn, D., Lessmann, S., and Schaal, M. (2020). Predicting online shopping behaviour from clickstream data using deep learning. *Expert Systems with Applications*, *150*, 113342. https://doi.org/10.1016/j.eswa.2020.113342

[27] Larivière, B., and Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, *29*(2), 472-484. https://doi.org/10.1016/j.eswa.2005.04.043

[28] Law, M., and Ng, M. (2016). Age and gender differences: Understanding mature online users with the online purchase intention model. *Journal of Global Scholars of Marketing Science*, *26*(3), 248-269. https://doi. org/10.1080/21639159.2016.1174540

[29] Lee, S., and Choeh, J. Y. (2014). Predicting the helpfulness of online reviews using multilayer perceptron neural networks. *Expert Systems with Applications*, *41*(6), 3041-3046. https://doi.org/10.1016/j.eswa.2013.10.034

[30] Ling, C. X., and Li, C. (1998). Data mining for direct marketing: problems and solutions. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*. New York, NY.

[31] Liu, A., Tan, Z., Li, X., Wan, J., Escalera, S., Guo, G., and Li, S. Z. (2019). Static and dynamic fusion for multi-modal cross-ethnicity face anti-spoofing. *arXiv preprint arXiv:1912.02340*.

[32] Lu, L., Dunham, M., and Meng, Y. (2005). Mining significant usage patterns from clickstream data. In *International Workshop on Knowledge Discovery on the Web*. Berlin, Heidelberg.

[33] Moe, W. W., and Fader, P. S. (2004). Dynamic Conversion Behavior at E-Commerce Sites. *Management Science*, *50*(3), 326-335. https://doi.org/10.1287/mnsc.1040.0153

[34] Mokryn, O., Bogina, V., and Kuflik, T. (2019). Will this session end with a purchase? Inferring current purchase intent of anonymous visitors. *Electronic Commerce Research and Applications*, *34*, 100836. https://doi.org/10.1016/j.elerap.2019.100836

[35] Moro, S., Cortez, P., and Rita, P. (2015). Using customer lifetime value and neural networks to improve the prediction of bank deposit subscription in telemarketing campaigns. *Neural Computing and Applications*, *26*(1), 131-139. https://doi.org/10.1007/s00521-014-1703-0

[36] Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., and Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*(1), 1. https://doi.org/10.1186/s40537-014-0007-7

[37] Ndubisi, N. O. (2006). Effect of gender on customer loyalty: A relationship marketing approach. *Marketing Intelligence & Planning, 24*(1), 48-61. https://doi.org/10.1108/02634500610641552

[38] Ogonowski, P. (2021). Ecommerce Conversion Rate Statistics. Retrieved from https://www.growcode.com/blog/ecommerce-conversion-rate

[39] Park, C. H., and Park, Y.-H. (2016). Investigating purchase conversion by uncovering online visit patterns. *Marketing Science*, *35*(6), 894-914.

https://doi.org/10.1287/mksc.2016.0990

[40] Poria, S., Cambria, E., Bajpai, R., and Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, *37*, 98-125. https://doi.org/10.1016/j.inffus.2017.02.003

[41] Poria, S., Cambria, E., and Gelbukh, A. (2015). Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal.

[42] Rahim, M. A., Mushafiq, M., Khan, S., and Arain, Z. A. (2021). RFM-based repurchase behavior for customer classification and segmentation. *Journal of Retailing and Consumer Services*, *61*, 102566. https://doi.org/10.1016/j.jretconser.2021.102566

[43] Rastgoo, M. N., Nakisa, B., Maire, F., Rakotonirainy, A., and Chandran, V. (2019). Automatic driver stress level classification using multimodal deep learning. *Expert Systems with Applications*, *138*, 112793. https://doi.org/10.1016/j.eswa.2019.07.010

[44] Rust, R. T. (2020). The future of marketing. *International Journal of Research in Marketing*, *37*(1), 15-26. https://doi.org/10.1016/j.ijresmar.2019.08.002

[45] Saide, C., Lengelle, R., Honeine, P., Richard, C., and Achkar, R. (2015). Nonlinear adaptive filtering using kernel-based algorithms with dictionary adaptation [10.1002/acs.2548]. *International Journal of Adaptive Control and Signal Processing*, *29*(11), 1391-1410. https://doi.org/10.1002/acs.2548

[46] Schuster, M., and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, *45*(11), 2673-2681. https://doi.org/10.1109/78.6500 93

[47] Sheil, H., Rana, O., and Reilly, R. (2018). Predicting purchasing intent: Automatic feature learning using recurrent neural networks. *arXiv preprint arXiv: 1807.08207*.

[48] Siami-Namini, S., Tavakoli, N., and Namin, A. S. (2019, 9-12 Dec. 2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. In *2019 IEEE International Conference on Big Data (Big Data)*. https://doi.org/10.1109/BigData47090.2019.9005997

[49] Sorce, P., Perotti, V., and Widrick, S. (2005). Attitude and age differences in online buying. *International Journal of Retail & Distribution Management*, *33*(2), 122-132. https://doi.org/10.1108/09590550510581458

[50] Sun, C., Adamopoulos, P., Ghose, A., and Luo, X. (2022). Predicting stages in omnichannel path to purchase: A deep learning model. *Information Systems Research*, *33*(2), 429-445.

[51] Toth, A., Tan, L., Di Fabbrizio, G., and Datta, A. (2017). Predicting shopping behavior with mixture of RNNs. In *Proceedings of SIGIR 2017 eCom*. Tokyo, Japan.

[52] Van den Poel, D., and Buckinx, W. (2005). Predicting online-purchasing behaviour. *European Journal of Operational Research*, *166*(2), 557-575. https://doi.org/10.1016/j.ejor.2004.04.022

[53] VanderMeer, D., Dutta, K., Datta, A., Ramamritham, K., and Navanthe, S. B. (2000). Enabling scalable online personalization on the web. In *Proceedings of the 2nd ACM conference on Electronic commerce*. New York, NY.

[54] Wagner, G., Schramm-Klein, H., and Steinmann, S. (2020). Online retailing across e-channels and e-channel touchpoints: Empirical studies of consumer behavior in the multichannel e-commerce environment. *Journal of Business Research*, *107*, 256-270. https://doi.org/10.1016/j.jbusres.2018.10.048

[55] Wang, L., Ning, H., Tan, T., and Hu, W. (2004). Fusion of static and dynamic body biometrics for gait recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, *14*(2), 149-158. https://doi.org/10.1109/TCSVT.2003.821972

[56] Wang, Y., and Xu, W. (2018). Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. *Decision Support Systems*, *105*, 87-95. https://doi.org/10.1016/j.dss.2017.11.001

[57] Wei, J. T., Lin, S. Y., and Wu, H. H. (2010). A review of the application of RFM model. *African Journal of Business Management*, *4*(19), 4199-4206.

https://doi.org/10.5897/AJBM.9000026

[58] Wu, Z., Tan, B. H., Duan, R., Liu, Y., and Mong Goh, R. S. (2015). Neural modeling of buying behaviour for e-commerce from clicking patterns. In *Proceedings of the 2015 International ACM Recommender Systems Challenge* (pp. 1-4). https://doi.org/10.1145/2813448.2813521

[59] Yang, M., and Wang, J. (2022). Adaptability of financial time series prediction based on BiLSTM. In *Procedia Computer Science*, *199*, 18-25. https://doi.org/10.1016/j.procs.2022.01.003

[60] Yeo, J., Hwang, S. W., S, K., Koh, E., and Lipka, N. (2020). Conversion Prediction from Clickstream: Modeling Market Prediction and Customer Predictability. *IEEE Transactions on Knowledge and Data Engineering*, *32*(2), 246-259. https://doi.org/10.1109/TKDE.2018.2884467

[61] Yuan, H., Zheng, J., Ye, Q., Qian, Y., and Zhang, Y. (2021). Improving fake news detection with domain-adversarial and graph-attention neural network. *Decision Support Systems*, 113633. https://doi.org/10.1016/j.dss.2021.113633

[62] Zhang, K., Geng, Y., Zhao, J., Liu, J., and Li, W. (2020a). Sentiment Analysis of Social Media via Multimodal Feature Fusion. *Symmetry*, *12*(12). https://doi.org/10.3390/sym12122010

[63] Zhang, W., Yu, J., Hu, H., Hu, H., and Qin, Z. (2020b). Multimodal feature fusion by relational reasoning and attention for visual question answering. *Information Fusion*, *55*, 116-126. https://doi.org/10.1016/j.inffus.2019.08.009

[64] Zhang, Y., Bradlow, E. T., and Small, D. S. (2013). New measures of clumpiness for incidence data. *Journal of Applied Statistics*, *40*(11), 2533-2548. https://doi.org/10.1080/02664763.2013.818627

[65] Zhang, Y., Bradlow, E. T., and Small, D. S. (2015). Predicting customer value using clumpiness: From RFM to RFMC. *Marketing Science*, *34*(2), 195-208. https://doi.org/10.1287/mksc.2014.0873

[66] Zhu, G., Wu, Z., Wang, Y., Cao, S., and Cao, J. (2019). Online purchase decisions for tourism e-commerce. *Electronic Commerce Research and Applications*, *38*, 100887. https://doi.org/10.1016/j.elerap.2019.100887

<Appendix A> URL category

<Table A1> Page URL and Categories

Category Index	Page URL	Page Category
1	/main/initMain.action	*Main*
2	/search/search.action?kwd=	*Search*
3	/shop/initShopBest100.action	*Best shop*
4	/goods/initGoodsDetail.action?goods_no=	*Product detail*
5	/shop/initPlanShop.action?disp_ctg_no=	*Plan shop*
6	/shop/initShopLuckyDeal.action	*Deal*
7	/shop/initEkidsShop.action	*Kids shop*
8	/shop/initPlanShopMain.action	*Exhibition shop*
9	/dispctg/initDispCtg.action?disp_ctg_no=	*Display*
10	/dispctg/searchBrandIndexBaseInfo.action	*Brand search*
11	/dispctg/initBrandShop.action?brand_no=	*Brand shop*
12	/cart/initCart.action	*Cart*
13	/order/initOrder.action?cart_no=	*Order*
14	/mypage/initMypageMain.action	*My page*
15	/mypage/initMyPointList.action	*My page point*
16	/mypage/initMyCouponList.action	*My page coupon*

<Appendix B> Page View Patterns

<Table B1> Comparison of Patterns between Sessions without Purchase and with Purchase

Session	Rank	Page View Pattern	Percentile
Without Purchase	1	Product detail → Product detail	27%
	2	Exhibition shop → My page	13%
	3	Product detail → Product detail → Product detail	5.2%
	4	My page → Exhibition shop	4.4%
	5	Main → Plan shop	3.8%
With Purchase	1	Product detail → Product detail	17%
	2	Product detail → Product detail → Product detail → Product detail	11%
	3	Product detail → Product detail → Product detail → Product detail → Product detail → Product detail	6.9%
	4	Display → Display	5.4%
	5	Product detail → Product detail → Product detail → Product detail → Product detail → Product detail → Product detail → Product detail	4.4%

<Appendix C> Visit- and Purchase-Based Clumpiness Measure

Following RFMC measures of previous works (Zhang et al., 2013; Zhang et al., 2015), we measure the clumpiness of visit and purchase. We first process customer transaction and visit data to daily incidence data. Second, we compute the inter-event times (IETs) of visits and purchases, respectively:

$$x_v = \begin{cases} t_1, & if\ v = 1, \\ t_v - t_{v-1}, & if\ v = 2, 3, \ldots, n, \\ N + 1 - t_v, & if\ v = n + 1. \end{cases} \quad (9)$$

In (9), t_i denotes the occurrence time of ith event and x_v indicates the IETs of customer visits. N is the total time intervals. Then, to control for observation window size, we rescale the inter-events times, v_i, by diving it by N+1. Last, we compute visit-based clumpiness as follows:

$$Visit\text{-}C = 1 + \frac{\sum_{v=1}^{n+1} \log(x_v) * x_v}{\log(n+1)}. \quad (10)$$

Similar to the visit-based clumpiness measure, we compute purchase-based clumpiness by replacing the inter-event times of visits to the those of purchases as follows:

$$Purchase\text{-}C = 1 + \frac{\sum_{p=1}^{n+1} \log(x_p) * x_p}{\log(n+1)}. \quad (11)$$

In (11), x_p indicates the IETs of purchases.

<Appendix D> LSTM and BiLSTM Network Architecture

LSTM network as an extension of RNN was introduced to address the problem of long-term dependency (Hochreiter and Schmidhuber, 1997). LSTM cell is comprised of input, forget and output gates, and cell state. Specifically, those gates are computed using the input vector and previous hidden state vector, and the hidden state is computed using cell state and output gate as follows:

$$\boldsymbol{i_t} = \boldsymbol{\sigma}(\boldsymbol{W_i} \cdot [\boldsymbol{x_t}, \boldsymbol{h_{t-1}}] + \boldsymbol{b_i}). \quad (12)$$

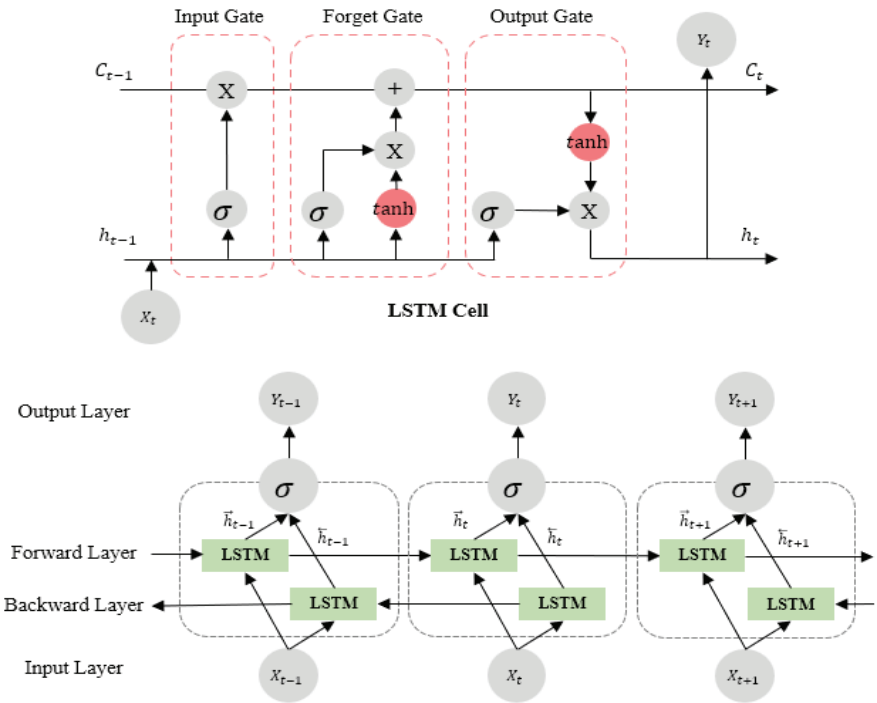
$$\boldsymbol{f_t} = \boldsymbol{\sigma}(\boldsymbol{W_f} \cdot [\boldsymbol{x_t}, \boldsymbol{h_{t-1}}] + \boldsymbol{b_f}). \quad (13)$$

$$\boldsymbol{o_t} = \boldsymbol{\sigma}(\boldsymbol{W_o} \cdot [\boldsymbol{x_t}, \boldsymbol{h_{t-1}}] + \boldsymbol{b_o}). \quad (14)$$

$$\boldsymbol{c_t} = \boldsymbol{tanh}(\boldsymbol{W_c} \cdot [\boldsymbol{x_t}, \boldsymbol{h_{t-1}}] + \boldsymbol{b_c}). \quad (15)$$

$$\boldsymbol{h_t} = \boldsymbol{o_t} \circ \mathbf{tanh}(\boldsymbol{c_t}). \quad (16)$$

In (12)-(16), x_t and h_t indicate input and hidden state vectors, respectively; i_t denotes the input gate at timestep *t*; f_t is the forget gate at timestep *t*; o_t is the output gate at timestep *t*; c_t represents the cell state at timestep *t*; and W and b with different subscriptions are the weight matrix and bias vector. Given that the LSTM network is suitable to handle sequential data, we employ the LSTM to process the dynamic pageview patterns in the model.



<Figure D1> Single LSTM Cell and BiLSTM Architecture Example

<Appendix D> LSTM and BiLSTM Network Architecture (Cont.)

To overcome the limitations of LSTM network that outputs of LSTM learn based previous timesteps but cannot learn future ones. LSTM network was modified into Bi-directional LSTM (BiLSTM) by combining forward and backward LSTM networks (Graves and Schmidhuber, 2005; Schuster and Paliwal, 1997). Based on LSTM network, BiLSTM network employs both previous and future context in output layer. For example, an input sequence $S = (s_1, s_2, \dots, s_n)$ in BiLSTM is processed in forward direction, $\overrightarrow{h_t}$= ($\overrightarrow{h_1}$,$\overrightarrow{h_2}$, ..., $\overrightarrow{h_n}$) and then in backward directions, $\overleftarrow{h_t}$=($\overleftarrow{h_1}$,$\overleftarrow{h_2}$, ...,$\overleftarrow{h_n}$). The output y_t is obtained by both $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$. <Figure D1> displays the single cell of LSTM and BiLSTM architecture.

◆ About the Authors ◆

Minsu Kim

Minsu Kim is a data scientist. After finishing his master's degree in the area of Information Systems at Yonsei University, he has been applying data science methods for marketing at LG UPlus, one of three biggest telecom companies in Korea. His research focuses on social media marketing, electronic commerce, and deep learning.

Woosik Shin

Woosik Shin is a PhD candidate of the Graduate School of Information at Yonsei University. His research interests focus on societal impact of IS, information security and privacy, and business value of IT. He also employs applied econometrics, experiments, and machine learning methods to explore the research topics.

SeongBeom Kim

SeongBeom Kim is a PhD candidate of the Graduate School of Information at Yonsei University. His research interests focus on deep learning, recommender systems and personalized marketing. Specifically, he is interested in topics related to customer behavior on e-commerce.

Hee-Woong Kim

Hee-Woong Kim is a Professor of the Graduate School of Information at Yonsei University. Before joining Yonsei University, he was a faculty member in the Department of Information Systems and Analytics at the National University of Singapore (NUS). He has served as Associate Editor for MIS Quarterly and the Editorial Boards of the Journal of the Association for Information Systems and IEEE Transactions on Engineering.

Submitted: January 13, 2023; 1st Revision: May 10, 2023; Accepted: July 10, 2023