IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 1, 2019, accepted August 10, 2019, date of publication August 13, 2019, date of current version August 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2935121

Customer Purchase Intent Prediction Under Online Multi-Channel Promotion: A Feature-Combined Deep Learning Framework

CHEN LING[1,2], TAO ZHANG[1,3], AND YUAN CHEN[1,3]
[1]School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China
[2]School of Medical Instruments, Shanghai University of Health Science, Shanghai 201318, China
[3]Shanghai Key Laboratory of Financial Information Technology, Shanghai University of Finance and Economics, Shanghai 200433, China

Corresponding authors: Tao Zhang (taozhang@mail.shufe.edu.cn) and Yuan Chen (chen.yuan@shufe.edu.cn)

This work was supported in part by the Humanities and Social Sciences Planning Fund of Ministry of Education of China under Grant 19YJA630116, in part by the Shanghai Natural Science Foundation under Grant 19ZR1417200, and in part by the Fundamental Research Funds for the Central Universities and the National Natural Science Foundation of China under Grant 71371115.

ABSTRACT The micro-level customer purchase intent in promotions is crucial for the overall purchase conversion rate of promotions. In the context of joint promotions on multiple online channels, customers can access and compare prices and services by navigating between channels to aid their purchase decision. The interactions between customer and promotion channels offer another angle to predict their purchase intent during promotions. In this paper, we propose a feature-combined deep learning framework, in which a full-connected long short-term networks (FC-LSTM) is used for modeling the interactions between customers and promotion channels, as well as the nonlinear sequence correlations and cumulative effects between customer's browsing behavior. To improve the performance of the prediction, the framework incorporates other features of customer profile including purchase history and demographics, integrating them into an end-to-end framework. We apply our method in a real prediction task for online multichannel promotion for concert tickets. Extensive experiments show that the proposed approach exhibits overall good performance compared with state-of-the-art methods on standard metrics such as precision, recall, f-measure, area under curve (AUC), and lift.

INDEX TERMS Multiple channels promotion, ticketing service, purchase intent, FC-LSTM, feature combination.

I. INTRODUCTION

Today, Internet technological advances are transforming the marketing in the ticketing service industry for entertainment products such as concerts, sports, and movies [1]. Authorized online ticket distributors such as Gewara, Damai, Ticketmaster, StubHub, are booming in the market [2], providing customers with more choice for ticketing service (Figure 1). The marketing of ticket service industry also takes advantage of online channels to conduct sales promotions while customers can have immediate access to this promotional information and compare prices and services between different online channels to aid their purchasing decisions [3]. As a key determinant of a company's revenue [4], the overall purchase conversion rate in promotions are concerned by marketers. Although most research demonstrates a macro-level increase in purchase conversion rate in the presence of promotions [5], still, some potential customers may not complete the purchase for various reasons such as missing the promotion information. Personalized marketing strategy is expected to target potential customers for improving their purchase conversion rate in promotions [6]. Hence, the prediction of micro-level customer purchase intent is a premise for conducting personalized marketing and plays a fundamental role in improving the overall purchase conversion rate [7].

During promotions, customers purchase intent prediction can be formulated as a binary classification task [8], where customers are classified into two groups according to whether they have purchase intents or not. Many research efforts have sought to research customer purchase intent

The associate editor coordinating the review of this article and approving it for publication was Alicia Fornés.

in promotions [9], and demonstrate that customer purchase intent closely relates to customer profile such as purchase history [10], [11], and demographics [12]–[14]. In addition, customer web browsing (clickstream) behavior is also regarded as a kind of important feature [15]. Traditionally, customer browsing behavior within a particular e-commerce website is emphasized for analyzing their purchase conversion rate [12]. Emerging studies have consistently pointed out that customers' browsing behaviors within the online environment can reveal the antecedent of customers' purchase outcomes [16]. Based on these features, a stream of machine learning methods have been applied in the prediction [14], [17], and obtain good prediction results.

While in the online B2C e-commerce market consists of multiple homogeneous distribution channels such as the ticketing service industry (Figure 1), the situation is significantly different. Given the popularity of online promotions through multiple distribution channels ("multi-channel promotion"), it is necessary for marketers to pay attention to customers' online browsing behavior between distribution channels. In this context, individual customer's clickstream data across channels is a key predictor to purchase intent [9], because customers often navigate between different online channels to search and compare prices and services [5]. Consequently, only using those carefully hand-crafted features within a particular e-commerce website and traditional machine learning methods cannot fully model the correlations and dynamics between customer's browsing behavior in the context of multi-channel promotions.

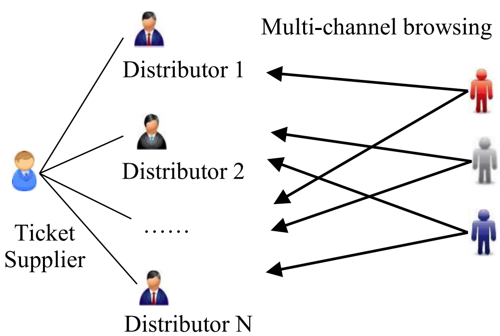


FIGURE 1. The multiple channels market structure of e-ticketing service.

In fact, a customer's purchase decision under promotions is a dynamic and evolving process [7]. In the case of promotions in multiple online channels, customers often compare prices and services through multiple channels before purchase [18]. customers' response behavior usually takes the form of searching, browsing, comparing sales information between channels, and placing orders [18]. Sequence correlation and cumulative effects [19] exist within a customer's browsing behavior on distribution channels [20], [21], where modeling these correlations and cumulative effects is a challenging task. In fact, in promotions, customer experience a series of psychological phases which play an important role as a predictor for the prediction of their purchase intent. For instance, Russell [22] proposed the AIDA model (Figure 2), which describes customers' experiencing Attention, Interest, Desire, and Action as a series of steps taken before making a purchase decision with a promotion or advertisement. AIDA was previously a popular framework for clearly separating customers' promotion-response behaviors into several stages, but it has been criticized for assuming that these stages are linear and sequential, some research [23], [24] believe that nonlinear effects exist between the four stages of the AIDA model (Figure 2). Moreover, although the AIDA model provides us with a framework to analyze customer intent with promotion, yet, what the AIDA model describes is an implicit process and discrete phases which are very difficult to quantify. In summary, the implicit nature, the assumption of discrete phase and the linear effects of the AIDA model hinder the direct application of the AIDA model in predicting customer purchase intent in promotions, because it cannot model the nonlinear evolving process of customer purchase decision.

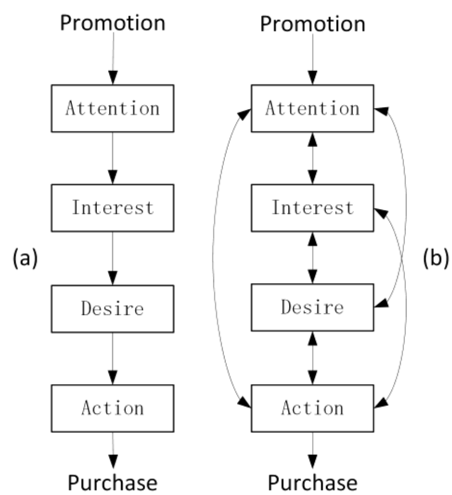


FIGURE 2. Original AIDA model with linear effects (a) in a promotion and improved AIDA model with nonlinear effects (b) in a promotion.

In summary, the traditional features and approaches cannot fully model the nonlinear sequence correlations, and the cumulative effects of customer browsing behavior relating to purchase in online multi-channel promotions. In this context, the interactions between customers and online multiple channels (Figure 3) can be viewed as an evolving signal of purchase intent, which provides us another angle to research customer purchase intent. According to literature, the interactions between customers and channels have never been addressed. Our solution is to model the dynamics of purchase conversion process using customer browsing sequence between promotion channels.

In order to predict micro-level customer purchase intent under online multiple channels promotions for the e-ticket of entertainment service products, we propose a feature-combined deep learning framework, in which, a full-connected long short-term networks (FC-LSTM) is used for

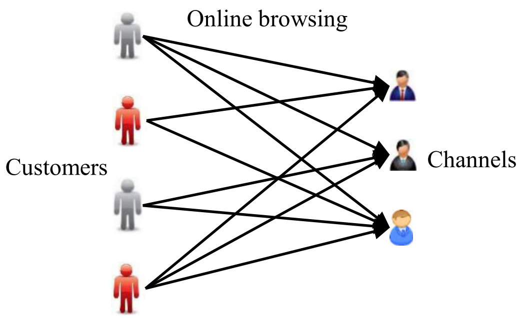


FIGURE 3. Interactions between customers and promotion channels.

modeling the interactions between customers and promotion channels, as well as the nonlinear correlations and cumulative effects between customer's browsing behavior. In order to improve the prediction performance, the framework incorporate information of customer profile such as purchase history and demographics, integrate them into an end-to-end framework.

The main contributions of this paper are three-fold: (1) The micro-level customer purchase intent prediction in sales promotion for the e-ticket of entertainment service products in multiple homogeneous distribution channels has never been discussed according to our knowledge. (2) we propose a full-connected long short-term networks (FC-LSTM) for modeling the interactions between customers and promotion channels, as well as modeling the nonlinear correlations, and cumulative effects between customer's browsing behavior. (3) Then, we propose a feature-combined deep learning framework, incorporating the FC-LSTM and customer profile information including purchase history and demographics. In our framework, multi-source and heterogeneous features are combined in an end-to-end model, providing high performance and practical framework for other similar problem in marketing. Finally, we apply our framework in an online sales promotion for e-ticket of symphony concert conducted by a famous concert hall in China. Extensive experiments demonstrate that our proposed feature-combined deep learning approach outperforms other state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 summarizes existing work relating to promotion effects, customers' purchase process, and prediction of purchase intention prediction. Section 3 demonstrates the feature extracting and feature representation. Section 4 presents the construction of our feature-combined deep learning framework. Section 5 conducts extensive experiments and analysis. Section 6 gives our conclusions and summary and identifies limitations that call for future works.

II. RELATED WORK

A. EFFECTS OF PROMOTIONS

In marketing, promotion refers to any type of marketing communication utilized to inform or persuade the target customer of the relative merits of a product, service, brand, or issue. The purpose of promotion is to increase attention, create interest, generate sales, or create brand loyalty [25]. Promotional channels include offline channels [26], online channels [27], dual channels [28], and multi-channel [5]. Multi-channels consist of both homogeneous and heterogeneous channels although our research uses it to refer solely to online homogeneous channels. Internet technological advances enable promotion to take place within online channels [29].

Many studies have investigated the effects of promotion in various contexts. For instance, Breugelmans and Campo [5] discussed the cross-channel effects of price promotions in the multi-channel grocery retail sector. A consistent finding in the promotion studies is the increase in purchase conversion rate after the promotion [30]–[32], which implies a positive response from customers' purchasing in many realms such as airline tickets [33], automotive industry [34], internet marketing [35].

Although most existing research discussed the overall effect of promotion in many industries, yet personal level purchase intent plays a more fundamental role in the overall purchase conversion rate [7]. However, little research focuses on the personal-level response of customer which are crucial for personalized marketing. Research into personal level purchasing behavior confirms that consumers move through a series of psychological stages when making purchase decisions [20]. According to the AIDA model [22], these stages are specified as *Attention*, *Interest*, *Desire*, and*Action* in sequence. The AIDA model is commonly used to analyze customers' purchase behavior in promotions and has been widely applied in marketing and advertising [36]. Kojima *et al.* [20] proposed that different promotional advertisements should be adopted in different AIDA stages to promote potential customers' psychological transformations for automobile purchases. Mohammadi *et al.* [36] found that it is necessary to vary the promotional strategies in the sports industry according to the different AIDA stages of customers. However, the purchase process actually does not strictly follow AIDA's step-by-step progression. Customers often skip stages or take them in different orders. Different customers also have different decision processes when making purchases [23]. In fact, customers' online purchase decision making is dynamic and highly flexible, regularly adding, skipping, or reordering the AIDA steps [37]. Therefore, the AIDA model does not always describe the process within multiple online channels in a promotion. The process frequently diverges from the main path and repeats stages [24].

Others have proposed improvements to the AIDA model. For instance, Barry and Howard [38] showed that there should be a *Conviction* stage before the *Desire* stage. Lavidge and Steiner [39] proposed that customers go through six psychological stages instead of four stages before purchase, which results in more complicated purchase behavior. Furthermore, Wijaya [40] proposed a nine-stage process consisting of several cognitive and behavior stages to reflect the complex, interactive nature of customer psychological

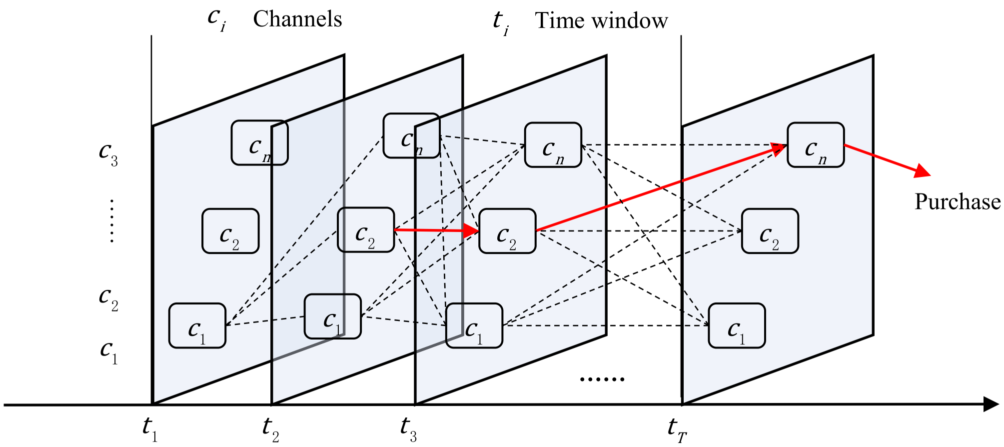


FIGURE 4. The dynamic process of customers' response (browsing in multiple channels) to online multiple channels promotions, which represents the interactions between customer and channels.

transformation. Yoon *et al.* [41] suggested that consumers process promotional information via cognitive (thinking) and affective (feeling) processes simultaneously. These empirical studies demonstrate that the evolving psychological process in the presence of promotion is a nonlinear and complex interaction mechanism.

Note that all these response models to sales promotions consist of several discrete and implicit phases which are very difficult to quantify, preventing the direct application of AIDA in the prediction of customer purchase intent. But the AIDA models provide us a kind of framework for analyzing customer's personal response to sales promotions.

B. CUSTOMER PURCHASE INTENT PREDICTION

The prediction of customers' purchase intent can be formulated as a binary classification task [8]. Several machine learning methods are used for the prediction task, including hidden Markov model [42], logistic regression [43], support vector machine [17], random forest [13], and gradient boosting decision tree [10]. These researchers have used customer profile as features, including purchase history [10], [11], and demographics [12]. Purchase history includes data such as average purchase interval, purchase frequency, date of last purchase, average tickets per order, and average ticket price. Demographics include information such as gender, age, education, and location [12]–[14]. These features have been proven to be effective indicators of future purchase intent [12]. Customer clickstream data is another kind of important features but it is frequently underestimated in previous studies [21]. Traditionally, customer browsing behavior within a particular E-commerce website is emphasized for analyzing their purchase conversion rate [12], [44]. For example, Customer's web browsing behavior, such as depth of browsing path, browsing time, and the interactions between customers and commodities [12], [45] have been discussed.

However, in the context of online multichannel promotion, customer access promotion information, and compare prices and services to facilitate their decision [9], [12], [19] between multiple channels before the eventual purchase. Padmanabhan *et al.* [9] showed that a customer's clickstream data across channels is a key predictor to purchase intent. The main challenge in this context is how to model the interaction between customer and channels.

Note that customer purchase is a dynamic and evolutionary process with nonlinear sequence correlation and cumulative effects [19], [21]. Deep learning methods can automatically extract complex feature representation from raw data through deep architectures to improve classification performance [46], [47]. Among deep learning methods, recurrent neural network (RNN) can extract sequence correlations within time series data [48]. The long and short-term memory neural network (LSTM) is an improved version of the RNNs that can be used in sequence prediction, including time sequence as well as other sequences. LSTM is suitable for analyzing long term and short term sequentially-structured data [49] and predict important events with relatively long intervals and delays in time series data [50], avoiding the gradient explosion and disappearance in traditional RNN [49]. Information fusion is now a hotpot emphasized in many areas [51]. the combination of multi-source and heterogeneous data is expected to improve the accuracy of purchase prediction [10]–[12].

Based on the above analysis, and enlightened by Wide & Deep architecture [52], we propose a feature-combined full-connected long and short-term memory neural network framework (FCD), incorporating the FC-LSTM and customer profile features in an end-to-end deep learning architecture for our prediction task.

III. FEATURE EXTRACTION

A. MULTI-CHANNEL BROWSING

As presented in Figure 4, to model the interactions between customer and promotion channels as well as a customer's response to a promotion, we propose procedures to extract features of customers' browsing across multiple channels in the presence of sales promotions.

Multiple studies have indicated that frequent customer visits to online channels imply strong interests or purchase intent in the products or services [11], [12], [14], [21]. We adopt a similar position and use the frequency of customer visits to the promotion channels (page view, PV) in a particular time window as the browsing feature. The PV simply calculates the frequency of browsing on channels during the promotion, ignoring the details of browsing behavior.

In order to extract features from browsing sequence in promotions that occurred across those multiple channels, we conduct feature extraction procedure from URLs recorded in weblogs provided by all the distribution channels. Taking all of the preceding into account, our overall process of feature extraction is as follows.

1) STEP 1. PRODUCT IDENTIFICATION

The components of an URL provide not only the online channel name but also clues about the products themselves, as shown in Table 1. However, most online channels sell a variety of kinds of products, and there are many "noise" URLs among the URLs obtained from customer browsing data from online channels. Hence, to target URLs for products, we first manually collect those URLs which contain product information from online channels as a candidate URL set, then we screen out all related URLs from weblogs by search and sift out URLs matching with any URLs in the candidate URL set.

2) STEP 2. CHANNELS IDENTIFICATION

According to our framework, customer research into product details and prices occurs within online channels. Thus, we extract the keyword indicating the channel name from each URL. For example, considering the promotions of symphony concert tickets, which the promotions are conducted in multiple online channels. The channels identified from URLs are listed in Table 1, the keywords representing online channels are *SYMPHONY*, *DAMAI*, *GEWARA*, etc.

3) STEP 3. FEATURE REPRESENTATION ACROSS CHANNELS

We use X_m^b to denote the sales channels for products, where m represents each online channel, and b represents browsing behavior in the channel. Each customer's multi-channel browsing data are expanded in chronological order to construct a two-dimensional matrix. We introduce a one-day time window in the time dimension to calculate the number of page view (PV) in the channel within a day. We then map the behavior data to a time series matrix denoted by X^b in (1).

$$X^b = \begin{bmatrix} X_1^b \\ X_1^b \\ \cdots \\ X_M^b \end{bmatrix} = \begin{pmatrix} x_{1,1}^b & x_{1,2}^b & \ldots x_{1,T}^b \\ x_{2,1}^b & x_{2,1}^b & \ldots x_{2,T}^b \\ \vdots & \ddots & \vdots \\ x_{M,1}^b & \ldots & \ldots x_{M,T}^b \end{pmatrix}. \quad (1)$$

where $x_{m,t}^b$ represent the browsing frequency (TF) in channel m in day t as shown in (2), where $t \in \{1, 2, \ldots, T\}$ represents the time step, T is the total number of time intervals, and $m \in \{1, 2, \ldots, M\}$ represents each online sales channel.

$$x_{m,t}^b = \sum_t TF_m, \quad m \in \{1, 2, ., M\}. \quad (2)$$

By the procedure of feature extraction, we get a multivariate time series (MTS) X^b.

Definition 1: (Formulation of customer purchase prediction based on browsing between multiple channels).

Suppose we observe a dynamical system over page view (PV) between multiple channels represented by a $X \times 1$ Vector X_m which consists of I channels. The variables inside the vector showing distinction across date and across channels, Thus, we will get a sequence of vector $X_{m1}^b, X_{m2}^b, \ldots, X_{mt}^b$, denoting by $X_{M\times T}$.

The temporal sequence classification problem is to predict the purchase outcome in the future given the previous T observations. With the aforementioned definition of the explanatory variables, we can formulate the customer's purchase intention prediction as formula 1.

$$\tilde{Y}_{(t+k)} = \underset{W,b}{argmax}\, p(Y_{(t+k)}|X_{M(t+1)}^b, X_{M(t+2)}^b, \ldots, X_{M(t+k-1)}^b). \quad (3)$$

W and b in (3) represent the weights and the biases of the prediction model. We note that the temporal sequence prediction problem is a multi-step input and one-step output time series prediction problem because the prediction target of our task is a binary variable which belongs to {0, 1}.

B. PURCHASE HISTORY AND DEMOGRAPHICS

The purchase history, as an important feature category, reflects customer preferences and their intention to purchase in the future [11], [13]. Purchase history features past purchase frequency [53], past spending amount and quantity [54], and time elapsed between purchases [55]. In line with these studies, we extract purchase history features relating to the frequency, cost, last date, and specific channels of actual purchases. Formally, we represent this feature as $(X_1^p, X_2^p, ., X_l^p)$, where p denotes feature category of purchase history, and l indicates the specific characteristic (e.g., average price or purchase frequency).

Customer demographics have also been widely used to predict purchase intention [12], [13] using characteristics such as gender, age, and postal code [14]. Following these studies, we extract demographic features including gender, age, postal code, education level, and occupation, which we represent as $(X_1^d, X_2^d, \ldots, X_j^d)$, where d denotes feature category of demographics, and j represents the particular characteristic (e.g., age or gender).

Note that the features of products often influence their sales, but in the context of promotion, these features do not play a key role in determining customer's response to promotion. Therefore, they are not considered in this paper.

IV. METHODOLOGY

In this section, we will describe the proposed feature-combined deep learning framework, which has three input

TABLE 1. Obtaining data from URLs.

Channel	URL	Concert	Keywords
Gewara	http://www.*gewara*.com/drama/365650899	The Misty of Europa (III): Resurrection of Mahler	*GEWARA*
Damai	https://piao.*damai*.cn/140719.html	The Misty of Europa (III): Resurrection of Mahler	*DAMAI*
Symphony	http://www.xymphony.com/item-index-id-885.html	The Misty of Europa (III): Resurrection of Mahler	*SYMPHONY*

branches, incorporating multi-source and heterogeneous data in an end-to-end neural network architecture. First, a full connection long short-term networks (FC-LSTM) is used for modeling the interactions between customers and promotion channels, as well as the nonlinear sequence correlations and cumulative effects between customer's browsing behavior. We use LSTM to capture the time-delayed information in the customer browsing data.

A. FULL-CONNECTED LONG SHORT-TERM MEMORY NETWORKS

LSTM adopts a memory unit to replace the hidden nodes in the original RNN. The memory unit consists of memory cells, forget gates, input gates, and output gates. Memory cells store history information, recording and updating history information through a state parameter, and the three gate structures determine the choice of information through a sigmoid function. Each LSTM cell has three gates to control whether the current information is ignored or passed to the next cell [50]. Every time a new input comes, its information will be accumulated to the cell if the input gate i_t is activated. Also, the past cell status c_{t-1} could be "forgotten" in this process if the forget gate f_t is on. Whether the latest cell output c_t will be propagated to the final state h_t is further controlled by the output gate o_t [50]. Figure 5 depicts a typical LSTM.

The input of the LSTM is the vector sequence with T time steps and denoted as $x_t = (x_1, x_2, \ldots, x_T)$. The output of the LSTM is the vector sequence denoted as $h_t = (h_1, h_2, \ldots, h_T)$, which stores the information from the previous time step and passes it to the next time step. The state update of the LSTM can be given by (4) - (8).

$$i_t = \sigma(w_{xi}x_t + w_{hi}h_{t-1} + w_{ci} \circ c_{t-1} + b_i). \tag{4}$$

$$f_t = \sigma(w_{xf}x_t + w_{hf}h_{t-1} + w_{cf} \circ c_{t-1} + b_f). \tag{5}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh(w_{xc}x_t + w_{hc}h_{t-1} + b_c). \tag{6}$$

$$o_t = \sigma(w_{xo}x_t + w_{ho}h_{t-1} + w_{co}c_t + b_o). \tag{7}$$

$$h_t = o_t \circ \tanh(c_t). \tag{8}$$

In (4)-(8), the operator σ is the point pair sigmoid function given in (8); the operator $\circ$ is the point pair product (Hadamard Product); tanh is a non-linear activation function given in (9); $u_i, u_f, u_c, u_o, w_i, w_f, w_c, w_o$ are the weight matrices of the implied state; and b_i, b_f, b_c, b_o represent the biases of the corresponding weights.

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{9}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{1 + e^{-x}}. \tag{10}$$

Depending on the structure of input data and output data, LSTM can be designed in different forms. For instance, Ke, Zheng, Yang, and Chen [56] provide an LSTM architecture for multi-source data fusion for forecasting taxi passenger demand. We use their architecture in our own approach.

To model the correlations of browsing behavior between promotion channels, we adopt a full-connected architecture (Figure 6). Each node in the input layer is connected to all the nodes in the hidden layer, consequently, the global correlations among channels can be caught.

As can be seen from Figure 6, in each time step t, the input to the input layer consists of two parts, one is the X_t, the other is the state of the hidden layer H_{t-1}. While at the last time step T, we get the output O_t. The reason to adopt an FC-LSTM is to catch the global correlations and sequence correlations between the browsing behavior on channels simultaneously.

B. FEATURE-COMBINED DEEP LEARNING (FCD) FRAMEWORK

Second, to improve the performance of the prediction, the framework incorporates other features of customer profile such as purchase history and demographics, integrating them into an end-to-end framework. In our study, the prediction of customer purchase intention in online multiple channels promotion is a binary classification problem. Since we combine customer features from different sources, the input data of the classification problem involve multiple heterogeneous-structure features and multiple time steps while the output data contain multiple variables and a single time step. Accordingly, we design our proposed FCD to use data structures for multi-source heterogeneous-structure input and a single output. Specifically, the proposed FCD method consists of an input layer, the LSTM layer, a merge layer, a full connection layer, and a sigmoid layer, as presented in Figure 3. The FCD framework consists of three branches (network channels), which handle online browsing behavior, purchase history and demographics separately.

First, there are three parallel branches in the input layer (Figure 7). Branch 1 deals with customer browsing data

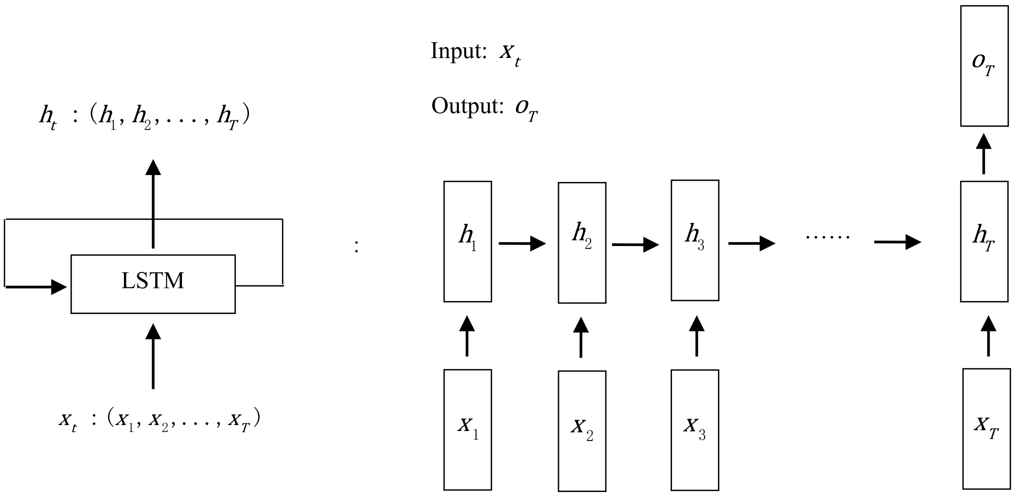


FIGURE 5. Exploded time-sequence view (right) of the original LSTM (left).

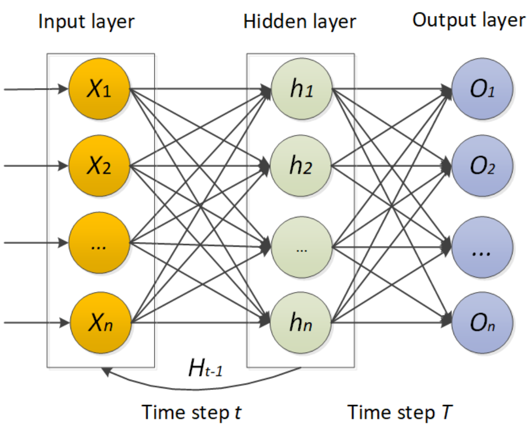


FIGURE 6. The full-connected architecture of the FC-LSTM.

during the promotion period. This information is sent to the LSTM layer for extracting the complex correlations within the browsing patterns across channels. Branch 2 and 3 handle purchase history and demographics. A dense layer is to map the input variables into a new feature space, where the input features are transformed into a new feature representation that is favorable for classification.

Second, the merge layer combines the three categories of features together and then maps them to the full-connection layer. There are several alternative methods for combining features. The series combination method [57] is one of the most adopted methods and has been confirmed as effective in many scenarios. Our FCD method incorporates Peng and Zhang's work [57] to combine browsing features produced by the LSTM and denoted by $X^o = (x_1^o, x_2^o, \ldots, x_T^o)$ with the purchase history and demographic features. We then obtain the combined features, denoted by the vector X^f, with

$$X^f = X^d \& X^p \& X^o. \tag{11}$$

$$X^{dense} = \sigma(WX^f + b). \tag{12}$$

In (11) and (12), W is the weight matrix, X^{dense} represents the output of the full connection layer, and $\widehat{Y}$ is the output vector representing the expected probability for each class.

Third, we employ another full connection layer (dense layer) with the sigmoid function σ commonly used for binary classification tasks, mapping X^{dense} to a vector $\widehat{Y}$, which uses the same label space as Y, given in (13). We then obtain the predicted probability distribution of the label.

$$\widehat{Y} = \sigma(WX^{dense} + b). \tag{13}$$

In addition, we adopt a cross-entropy loss function to measure the distance between two probability distributions, as given in (14) and (15). In our case, the two probability distributions refer to the real distribution of label classes of an instance and the predicted probability distribution of the label classes of an instance.

$$\min_{\omega,b} L_{H(Y,\widehat{Y})} = H(P_Y \parallel P_{\widehat{Y}}). \tag{14}$$

$$H(P_Y \parallel P_{\widehat{Y}}) = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\widehat{y}_i) + (1-y_i)\log(1-\widehat{y}_i)]. \tag{15}$$

In (14) and (15), $P_{\widehat{Y}}$ is the probability distribution of the two classes in the dataset calculated by the FCD method, and P_Y is the real distribution of the two classes in a dataset. More specifically, $\widehat{y}_i$ is the probability that an instance label is predicted to be 1, while $1 - \widehat{y}_i$ is the probability that the instance label is predicted to be 0.

V. EXPERIMENTS AND RESULTS

A. SETTINGS AND DATA

In our experiments, we use data collected from a large and famous concert hall in China. Established in the 1870s, its

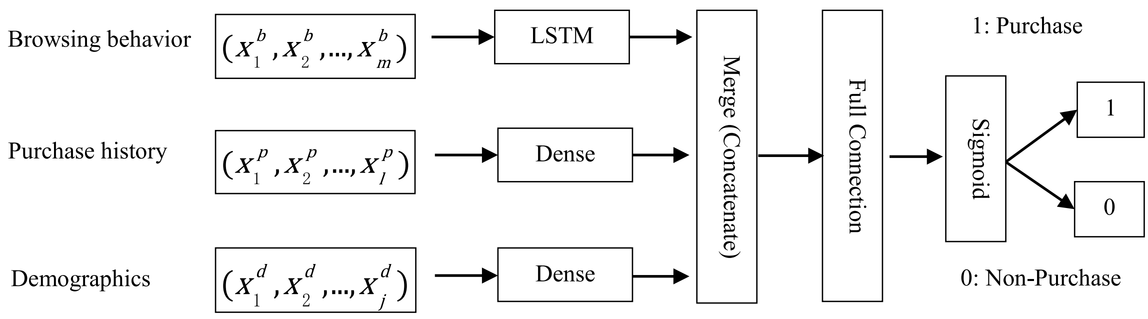


FIGURE 7. The proposed FCD framework.

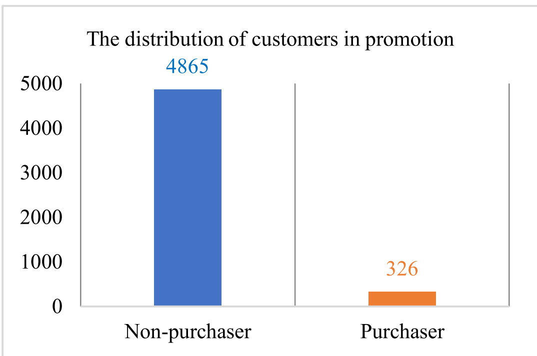


FIGURE 8. The distribution of customers response in the promotion.

main business is the operation of symphony concerts. Most of its concert tickets are sold through online channels such as *Gawara* and *Damai*, both of which are well-known in China. Online channels now account for the majority of concert ticket sales. [58]. In recent years, this concert hall has suffered from a lower attendance rate of under 60% on average. To encourage potential customers to purchase concert tickets, the hall decided to carry out sales promotions to stimulate demand. At the end of 2017, they launched a half-year promotion in multiple online channels to offer discounts for concert tickets. This promotional campaign provides us a unique dataset to observe customer behavior in response to the promotions in these online channels and to predict their purchase intentions.

During the promotional period, customers bought 7877 tickets for 25 concerts through more than 6 online channels, with sales totaling 1,086,300 Chinese yuan (about 158,273 US dollars). A total of 5191 people browsed these online channels, and 326 people purchased the tickets (Figure 8). Thus, we divide all of the customers in our dataset into two subsets: the positive set containing customers who purchased tickets and the negative set with customers who did not purchase tickets. We derive 11,156,965 URLs from the browsing data and obtain their purchase history and demographic details.

B. EXPERIMENT DESIGN

Based on the dataset, we design three groups of experiments with two objectives. The first objective is to compare the prediction performance of our proposed FCD method with alternative methods, with and without feature combination. The other objective is to evaluate the prediction performance of our proposed FCD and to compare it with other benchmark methods.

First, *Experiment_G1* is to examine the fitting and convergence performance of FCD during the training process. We conduct experiments to evaluate the performance of FCD in term of Loss and ACC curves. We perform other experiments to test the performance of FCD with different parameter settings, including dropout, learning rate, and L1 and L2 regularization, in order to find the best parameter settings.

Second, *Experiment_G2* is to verify the effectiveness of the feature combination technique. In term of a variety of feature combinations, we conduct experiments to compare the performance of our proposed FCD method with some benchmark methods already used in customer purchase prediction such as LR, SVM, RF, GBDT. Moreover, we further test the performance of some state-of-the-art methods widely used in Click Through Rate (CTR) prediction and recommendation, including Factorization Machines (FM) [59], and Field-aware Factorization Machines (FFM) [60]. We evaluate their performance by using the common metrics of precision, recall, f-measure, ROC (Receiver Operating Characteristic), and AUC (Area Under Curve). Among these metrics, we calculate precision, recall, and f-measure according to (16) - (18), where TP represents the number of true positives, FP represents the number of false positives, and FN represents the number of false negatives.

$$precision = \frac{TP}{TP + FP}. \tag{16}$$

$$recall = \frac{TP}{TP + FN}. \tag{17}$$

$$F - measure = \frac{(1 + \beta^2) \times recall \times precision}{\beta^2 \times (recall + precision)}. \tag{18}$$

The precision refers to the number of instances correctly identified over the number of instances identified, and recall refers to the number of instances that are correctly identified over the number of instances that should be identified. The F-measure is a tradeoff of precision and recall measuring the effectiveness of classification in terms of a ratio of the weighted importance on both recall and precision as

TABLE 2. Description of feature variables.

Datasets	Features	Description	Variable Type
Purchase history X^p	*sum of expenditures*	The total consumption in yuan	Continuous
	sum of tickets	The total number of tickets brought by customers	Continuous
	sum of orders	The consumption frequency	Continuous
	recency	The timing of the most recent consumption	Continuous
	lead time	The average purchase lead time prior to a concert	Continuous
Demographics X^d	*gender*	Customer's gender (male or female)	Discrete
	age	Customer's age	Continuous
	income(approximation)	Customer's income approximated by the average GDP of the customer's location	Continuous
	housing	The average house price in customers' location	Continuous
	education	Customer's education level (high school, Bachelor's degree, Master's degree, Ph.D., etc.)	Discrete
	job	Customer's occupation (e.g., lawyer, private business owner, freelancer, etc.)	Discrete
Online browsing X^b	*PV on channels*	Page view of each channel each day during the promotion	Integer

TABLE 3. Experimental parameters.

Methods	Br	Parameters
Logistic Regression	LR	max_iter=100
Support Vector Machine	SVM	C= 1.0, kernel= rbf
Random Forest	RF	n_estimators=300, max_depth=24
Gradient Boosting DT	GBDT	n_estimators=300, max_depth=24
Factorization Machines	FM	k=12, learning_rate=0.0005, batch_size=64
Field-aware Factorization Machines	FFM	k=16, learning_rate=0.0002, batch_size=64

determined by a coefficient β set manually. The ROC and AUC measures are widely used in classification tasks. The ROC curve summarizes classifier performance over a range of tradeoffs between the true positive rate (TPR) and false positive rate (FPR) as defined in (19) and (20). AUC is a traditional performance metric describing the overall performance of a classifier.

$$TPR = \frac{TP}{TP + FN}. \tag{19}$$

$$FPR = \frac{FP}{FP + TN}. \tag{20}$$

Finally, *Experiment_G3* is to test the performance of the FCD method. In the field of marketing, the lift is a commonly used indicator to reflect the return on investment (ROI) and has been widely applied to test the effectiveness of a prediction model [61]. To calculate lift, customer instances are first sorted in descending order based on the purchase intention scores obtained by the prediction model. Then, the lift is computed as the ratio of the percentage of the correctly classified purchasers (the minority class) in the top 10% ranked cases ($\beta_{10\%}$) and the percentage of actual purchasers in the whole data set as given in (21). For example, a top-decile lift value of 3 indicates that the classifier identifies thrice as many purchasers in the top 10% ranked customer group as a random classifier would do.

$$top - decilelift = \beta_{10}\%/\beta_0. \tag{21}$$

Since improving the ROI is the essential objective of the promotion campaign in our context, we design this group of experiments to assess the performance of FCD by comparing it with logistic regression, support vector machine, random forest, and gradient boosting decision tree methods according to the lift metric given by (21). (21) and (22) refer to the same concept, while the form of (22) emphasizes the specific calculation details.

$$lift = \frac{TP/(TP + FP)}{(TP + FN)/(TP + TN + FP + FN)}. \tag{22}$$

We implement our experiments using Python 3.5, along with the machine learning packages *Keras* and *scikit-learn*. We also apply the Synthetic Minority Over-sampling Technique (SMOTE) [62] and Random Under-Sampling technique (RUS) to balance the dataset. Table 3 shows the parameters for the FCD method and the other alternative methods in our experiments.

C. RESULTS AND DISCUSSION

For *Experiment_G1*, Figures 9, 10, and 11 present the fitting and convergence performance of the FCD model regarding the loss curve for different parameter settings during the training process. First, as shown in Figure 9, for the three dropout rates, when the dropout rate is 0.2, the FCD reach its best perform best with approximately 180 epochs.

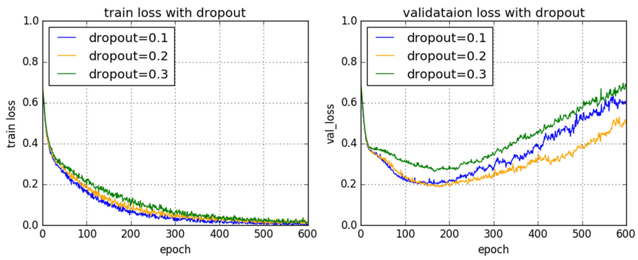


FIGURE 9. Loss curves of FCD with different dropout rates.

Figure 10 presents the loss curves of FCD for the three learning rates. The FCD model performed best with the learning rate set to 0.002 and the number of training epochs set to around 90, which indicates a fast convergence rate. A smaller learning rate often gets a local minimum while misses the global optimum in this case.

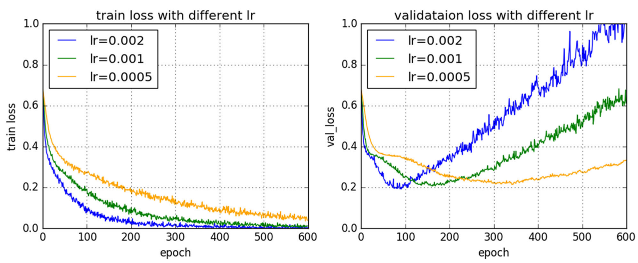


FIGURE 10. Loss curves of FCD for different learning rates.

Figure 11 shows the loss curves of the FCD with three L1 and L2 regularization values used to avoid the overfitting problem. The FCD model performs best with the L1 and L2 regularization values both set to 0.01.

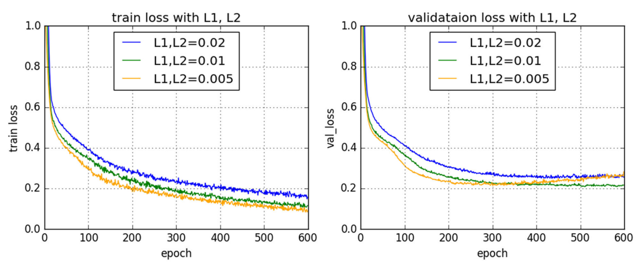


FIGURE 11. Loss curves of FCD with different L1 and L2 regularization rates.

Then, we conduct experiments to examine the performance of FCD regarding the ACC and loss curves with three parameter sets. Figure 12 shows FCD performance when the dropout rate is 0.2, and Figure 13 presents its performance when the dropout rate is 0.2 with a learning rate of 0.001. According to Figure 14, the FCD method performed best with the parameters combination of dropout rate set to 0.2, the learning rate set to 0.002, and the regularization values both set to 0.01.

Experiment_G1 shows that, the FCD model is easy to train, with good fitting and convergence performance. Moreover,

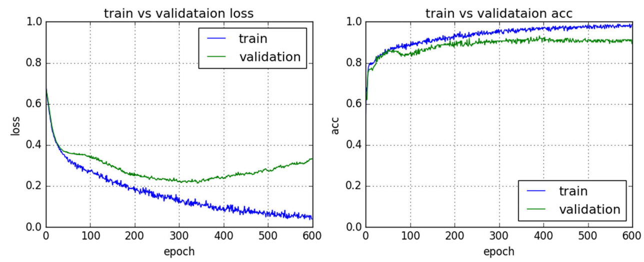


FIGURE 12. ACC curves and loss curves of FCD (dropout = 0.2).

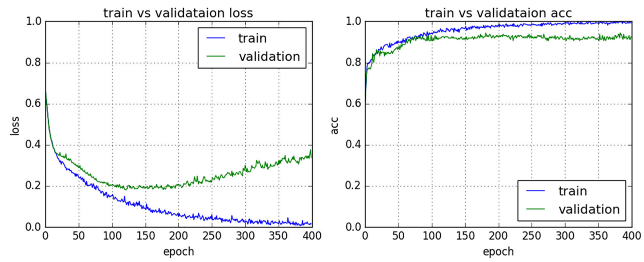


FIGURE 13. ACC curves and loss curves of FCD with dropout (0.2) and learning rate (0.001).

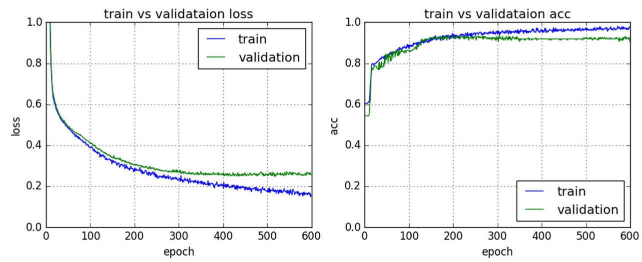


FIGURE 14. ACC curves and loss curves of FCD with dropout (0.2), learning rate (0.002) and regularization (L1 = 0.1, L2 = 0.1).

we can obtain the best parameters settings for the FCD. In the following experiments, we will directly apply the best parameters settings obtained from *Experiment_G1*.

The results of *Experiment_G2* are presented in Table 4, Table 5, and Figure 15. Table 4 shows that the best performance regarding the precision, the f-measure, and the AUC for our FCD algorithm are obtained using the combination of all the features (browsing behavior, demographics, purchasing history, expressed as B+D+P). Table 5 shows the performance comparison of FCD with other benchmark methods (LR, SVM, GBDT, RF, FM and FFM) using feature combination, the proposed FCD method outperforms other methods significantly regarding the precision, the f-measure, and the AUC.

Figure 15 presents the ROC curve results for all of the tested methods using feature combination. As shown in Figure 15, the FCD method had a significant advantage over the other methods. From the results of *Experiment_G2*, we conclude that the combination of multi-source features improves prediction performance across the board, with the FCD method offering the best overall performance.

TABLE 4. FCD Performance with different feature combinations.

	Precision	Recall	F-measure	AUC
B	0.9264 (0.0526)	0.8496 (0.0491)	0.9105 (0.0403)	0.9016 (0.0454)
B+D	0.9538 (0.0348)	0.9510 (0.0373)	0.9612 (0.0522)	0.9596 (0.0306)
B+P	0.8459 (0.0412)	0.9601 (0.0390)	0.9131 (0.0379)	0.9200 (0.0411)
B+D+P	0.9778 (0.0268)	0.9448 (0.0213)	0.9687 (0.0232)	0.9650 (0.0253)

D: Demographics; P: Purchase history; B: Browsing in multi-channel, (*): * means standard deviation

TABLE 5. Performance comparison of FCD with other benchmark models.

	Precision	Recall	F-measure	AUC
LR	0.8714 (0.0441)	0.6993 (0.0454)	0.7759 (0.0448)	0.8843 (0.0327)
SVM	0.8610 (0.0513)	0.5490 (0.0632)	0.6705 (0.0593)	0.8717 (0.0274)
GBDT	0.8327 (0.0251)	0.7668 (0.0453)	0.7984 (0.0352)	0.8942 (0.0237)
RF	0.9012 (0.0406)	0.7821 (0.0401)	0.8374 (0.0404)	0.9124 (0.0232)
FM	0.8577 (0.0460)	0.7382 (0.0380)	0.7919 (0.0246)	0.8846 (0.0141)
FFM	0.8711 (0.0504)	0.7228 (0.0318)	0.7890 (0.0288)	0.8999 (0.0080)
FCD	0.9778 (0.0268)	0.9448 (0.0213)	0.9687 (0.0232)	0.9650 (0.0253)

(*): * means standard deviation

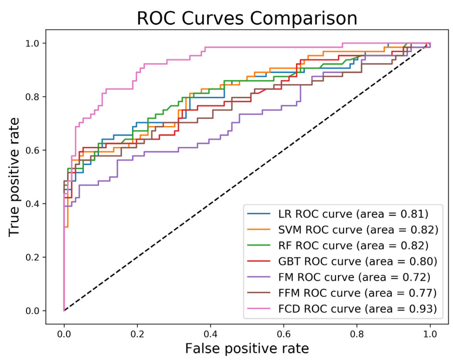


FIGURE 15. Comparison of FCD with other methods regarding ROC curves in a random test.

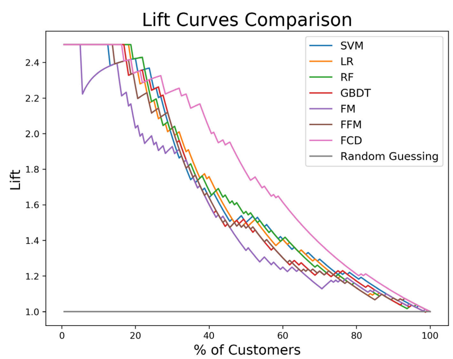


FIGURE 16. Comparison of FCD with other methods regarding lift curves in a random test.

Figure 16 presents the results of *Experiment_G3*. Within the lift curves of all of the tested methods, in the first 20% of the depth (depth is the proportion of the number of customers targeted to the total quantity of customers), FCD, RF and GBDT are in the first level. But, then FCD has an absolute vantage over other methods from 20% to 100% regarding the depth. Moreover, the FCD lift curve is extremely robust compared to the other methods, As shown in Figure 16, there is no fluctuation in the lift curve of FCD, while some algorithms have volatile lift curves more or less. The FCD lift curve has a significant advantage over other methods, which means that when you plan to target potential customers in a promotion campaign, the return on investment (ROI) of FCD is always better than other state-of-the-art methods.

In summary, the experiment results show that feature combination improves prediction performance. In the context of online multiple channels promotion, our proposed FCD method exhibited excellent performance, making it promising for predicting customer purchase intent in practice. Furthermore, our FCD method is robust overall, and easy to train, making it more effective as a prediction method.

D. INSIGHT INTO THE FEATURES

The limitation of deep learning lies in its interpretability. because we cannot interpret the learning process, as well as fully understand the contributions of the prediction features. Therefore, we utilize a data visualization approach to see how some of the prediction features work.

1) VISUALIZATION OF CUSTOMER BROWSING BEHAVIOR BETWEEN CHANNELS

In order to get an insight into customers' browsing patterns in multiple online channels, we apply data visualization during our analysis. We randomly select two customers and then draw a heatmap with normalized data (23) according to the customers' page view (PV) in multiple channels in each time step. Customer 1 is a positive instance who made a purchase in the promotion while customer 2 is a negative instance who

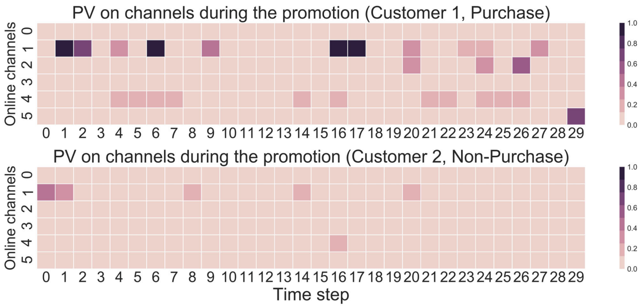


FIGURE 17. **Comparison of two customers' page view (PV) on multiple channels in each time window.**

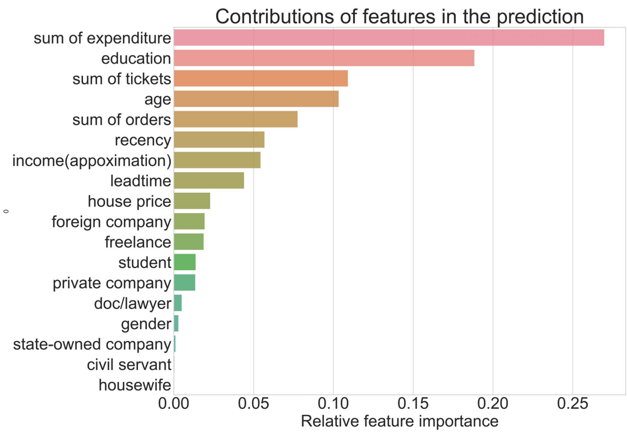


FIGURE 18. **Relative contributions of features to the customer's purchase intent in the promotion.**

did not make a purchase.

$$x = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})}. \tag{23}$$

Figure. 1 presents the heatmap of two customers' browsing frequency in multiple channels in each time step. Obviously, the two customers had different browsing patterns. Customer 1 has more PV than customer 2. This can partially reveal the different outcomes (purchase or not) in the sales promotion regarding the two types of customers.

2) VISUALIZATION OF FEATURE IMPORTANCE OF CUSTOMER PROFILE

In order to get an insight into the contributions of features extracted from the customer profile, we calculate the relative importance of each feature in the customer profile in the process of construction of classification trees for prediction. The variable importance is measured by the decline of the Gini coefficient, which describes the purity of features. The higher the purity of the set, the more important of the variable. The calculation of the Gini coefficient is shown in (24).

$$Gini(D) = 1 - \sum_{i=1}^{n} p_i^2. \tag{24}$$

In Figure 18, the horizontal axis reflects the features' importance and the longitudinal axis reflects the name of the features. As can be observed from the graph, the sum of expenditure derived from customers' purchase histories is the most important influencing factor, followed by education, the sum of tickets and age, other features are relatively trivial. These findings can considerably help guide future marketing activities.

VI. CONCLUSION

In this paper, we have proposed a feature-combined deep learning framework for predicting customers' purchase intention in promotion under multiple online channels. Our framework is based on the AIDA model and extracts customer browsing behavior across channels. We have demonstrated how combining this browsing behavior with purchase history and demographic information was expected to improve prediction performance but raised methodological challenges as well. Therefore, we have proposed an FCD using the LSTM to model sequence correlation within customer browsing data and process the multiple sources and heterogeneous data structures of feature combination. Extensive experiments showed that for all of the tested methods, the combination of all features had the best prediction performance, with our proposed FCD outperforming the other machine learning methods according to precision, recall, f1, AUC, and lift. Our proposed approach, although made in the context of concert tickets promotion, can be applied to other similar promotion problems.

Our study contributes to and extends the research into promotion. First, little research had been made into online multiple channels promotions and their effects on the personal-level response of customer especially when applying the state-of-the-art machine learning approaches to predict purchase intention in the presence of promotion. Our study is widely applicable to the entertainment service products such as symphony concert, cinema, and sports game. Second, prior studies have not combined browsing data with purchase history and demographic information in predicting customer purchase intention. To our knowledge, our study is the first to combine these feature categories together for prediction. Third, our proposed feature-combined deep learning approach uses the LSTM architecture, which can be modified to solve prediction problems in other contexts with similarly heterogeneous-structured features. In addition, our study also has important implications for firms in understanding customer behavior and for providing a basis for further personalized marketing strategies.

Our study suggests several avenues for future researchers to explore, partly due to the limitations on our own research. Although we have utilized the sequence correlation in customer's browsing behavior in our prediction task, yet, other correlations may exist in customer's browsing behavior. Thus, further research can take other correlations into consideration. In addition, there are some other methods which can deal with multi-source data fusion, especially for heterogeneous data, which need to be discussed in future research. Another limitation is that our research only focuses on the entertainment service products such as concerts, sports or movies, products of other industry need to be discussed in our future research.

REFERENCES

[1] M. Arce-Urriza, J. Cebollada, and M. F. Tarira, "The effect of price promotions on consumer shopping behavior across online and offline channels: Differences between frequent and non-frequent shoppers," *Inf. Syst. e-Bus. Manage.*, vol. 15, no. 1, pp. 69–87, 2016. doi: 10.1007/s10257-016-0310-2.

[2] J. Drayer and N. T. Martin, "Establishing legitimacy in the secondary ticket market: A case study of an NFL market," *Sport Manage. Rev.*, vol. 13, no. 1, pp. 39–49, 2010. doi: 10.1016/j.smr.2009.09.003.

[3] E. de Haan, T. Wiesel, and K. Pauwels, "The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework," *Int. J. Res. Marketing*, vol. 33, no. 3, pp. 491–507, 2016. doi: 10.1016/j.ijresmar.2015.12.001.

[4] N. Gudigantala, P. Bicen, and M. Eom, "An examination of antecedents of conversion rates of e-commerce retailers," *Manage. Res. Rev.*, vol. 39, no. 1, pp. 82–114, 2016. doi: 10.1108/mrr-05-2014-0112.

[5] E. Breugelmans and K. Campo, "Cross-channel effects of price promotions: An empirical analysis of the multi-channel grocery retail sector," *J. Retailing*, vol. 92, no. 3, pp. 333–351, 2016. doi: 10.1016/j.jretai.2016.02.003.

[6] J. Zhang and M. Wedel, "The effectiveness of customized promotions in online and offline stores," *J. Marketing Res.*, vol. 46, no. 2, pp. 190–206, 2009. doi: 10.1509/jmkr.46.2.190.

[7] O. Mokryn, V. Bogina, and T. Kuflik, "Will this session end with a purchase? Inferring current purchase intent of anonymous visitors," *Electron. Commerce Res. Appl.*, vol. 34, Mar./Apr. 2019, Art. no. 100836. doi: 10.1016/j.elerap.2019.100836.

[8] C. Sismeiro and R. E. Bucklin, "Modeling purchase behavior at an e-commerce Web site: A task-completion approach," *J. Marketing Res.*, vol. 41, no. 3, pp. 306–323, 2004. doi: 10.1509/jmkr.41.3.306.35985.

[9] B. Padmanabhan, Z. Zheng, and S. O. Kimbrough, "Personalization from incomplete data: What you don't know can hurt," in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, 2001, pp. 154–163. doi: 10.1145/502512.502535.

[10] A. Martínez, C. Schmuck, S. Pereverzyev, C. Pirker, and M. Haltmeier, "A machine learning framework for customer purchase prediction in the non-contractual setting," *Eur. J. Oper. Res.*, to be published. doi: 10.1016/j.ejor.2018.04.034.

[11] Y.-H. Tao, T.-P. Hong, W.-Y. Lin, and W.-Y. Chiu, "A practical extension of Web usage mining with intentional browsing data toward usage," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3937–3945, 2009. doi: 10.1016/j.eswa.2008.02.058.

[12] D. Van den Poel and W. Buckinx, "Predicting online-purchasing behaviour," *Eur. J. Oper. Res.*, vol. 166, no. 2, pp. 557–575, 2005. doi: 10.1016/j.ejor.2004.04.022.

[13] A. Vanderveld, A. Pandey, A. Han, and R. Parekh, "An engagement-based customer lifetime value system for e-commerce," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, 2016, pp. 293–302. doi: 10.1145/2939672.2939693.

[14] E. Y. Kim and Y.-K. Kim, "Predicting online purchase intentions for clothing products," *Eur. J. Marketing*, vol. 38, no. 7, pp. 883–897, 2004. doi: 10.1108/03090560410539302.

[15] Q. Su and L. Chen, "A method for discovering clusters of e-commerce interest patterns using click-stream data," *Electron. Commerce Res. Appl.*, vol. 14, no. 1, pp. 1–13, 2015. doi: 10.1016/j.elerap.2014.10.002.

[16] H. Wang, Q. Wei, and G. Chen, "From clicking to consideration: A business intelligence approach to estimating consumers' consideration probabilities," *Decis. Support Syst.*, vol. 56, pp. 397–405, Dec. 2013. doi: 10.1016/j.dss.2012.10.052.

[17] Y. Tsuboi, A. Jatowt, and K. Tanaka, "Product purchase prediction based on time series data analysis in social media," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*, Singapore, Dec. 2015, pp. 219–224. doi: 10.1109/wi-iat.2015.170.

[18] K. L. Ailawadi, J. P. Beauchamp, N. Donthu, D. K. Gauri, and V. Shankar, "Communication and promotion decisions in retailing: A review and directions for future research," *J. Retailing*, vol. 85, no. 1, pp. 42–55, 2009. doi: 10.1016/j.jretai.2008.11.002.

[19] W. W. Moe and P. S. Fader, "Capturing evolving visit behavior in clickstream data," *J. Interact. Marketing*, vol. 18, no. 1, pp. 5–19, 2004. doi: 10.1002/dir.10074.

[20] T. Kojima, T. Kimura, M. Yamaji, and K. Amasaka, "Proposal and development of the direct mail method PMCI-DM for effectively attracting customers," *Int. J. Manage. Inf. Syst.*, vol. 14, no. 5, pp. 15–22, 2010. doi: 10.19030/ijmis.v14i5.9.

[21] W. W. Moe and P. S. Fader, "Dynamic conversion behavior at e-commerce sites," *Manage. Sci.*, vol. 50, no. 3, pp. 326–335, 2004. doi: 10.1287/mnsc.1040.0153.

[22] C. P. Russell, *How to Write a Sales-Making Letter*. London, U.K.: Printers' Ink, 1921, pp. 49–56.

[23] S. Karimi, K. N. Papamichail, and C. P. Holland, "The effect of prior knowledge and decision-making style on the online purchase decision-making process: A typology of consumer shopping behaviour," *Decis. Support Syst.*, vol. 77, pp. 137–147, Sep. 2015. doi: 10.1016/j.dss.2015.06.004.

[24] K. N. Papamichail and I. Robertson, "Supporting distributed decision processes using an evolution model," *Oper. Res.*, vol. 8, no. 3, pp. 279–297, 2008. doi: 10.1007/s12351-008-0019-1.

[25] H. J. van Heerde and S. A. Neslin, "Sales promotion models," in *Handbook of Marketing Decision Models*, vol. 254, B. Wierenga and R. van der Lans, Eds. Boston, MA, USA: Springer, 2017, pp. 13–77. doi: 10.1007/978-3-319-56941-3_2.

[26] P. Zhang, Y. He, and C. Shi, "Retailer's channel structure choice: Online channel, offline channel, or dual channels?" *Int. J. Prod. Econ.*, vol. 191, pp. 37–50, Sep. 2017. doi: 10.1016/j.ijpe.2017.05.013.

[27] O. F. Koch and A. Benlian, "Promotional tactics for online viral marketing campaigns: How scarcity and personalization affect seed stage referrals," *J. Interact. Marketing*, vol. 32, pp. 37–52, Nov. 2015. doi: 10.1016/j.intmar.2015.09.005.

[28] J. Chen, L. Liang, D.-Q. Yao, and S. Sun, "Price and quality decisions in dual-channel supply chains," *Eur. J. Oper. Res.*, vol. 259, no. 3, pp. 935–948, 2017. doi: 10.1016/j.ejor.2016.11.016.

[29] K. Du, "The impact of multi-channel and multi-product strategies on firms' risk-return performance," *Decis. Support Syst.*, vol. 109, pp. 27–38, May 2018. doi: 10.1016/j.dss.2018.01.009.

[30] C. C. H. Chan, C.-B. Cheng, and W.-C. Hsien, "Pricing and promotion strategies of an online shop based on customer segmentation and multiple objective decision making," *Expert Syst. Appl.*, vol. 38, no. 12, pp. 14585–14591, 2011. doi: 10.1016/j.eswa.2011.05.024.

[31] H. J. van Heerde, P. S. H. Leeflang, and D. R. Wittink, "Decomposing the sales promotion bump with store data," *Marketing Sci.*, vol. 23, no. 3, pp. 317–334, 2004. doi: 10.1287/mksc.1040.0061.

[32] J. Saunders, D. Jobber, and V. Mitchell, "The optimum prepaid monetary incentives for mail surveys," *J. Oper. Res. Soc.*, vol. 57, no. 10, pp. 1224–1230, 2017. doi: 10.1057/palgrave.jors.2602053.

[33] E. Crespo-Almendros and S. Del Barrio-García, "Online airline ticket purchasing: Influence of online sales promotion type and Internet experience," *J. Air Transp. Manage.*, vol. 53, pp. 23–34, Jun. 2016. doi: 10.1016/j.jairtraman.2016.01.004.

[34] M. Chaudhuri, R. J. Calantone, C. M. Voorhees, and S. Cockrell, "Disentangling the effects of promotion mix on new product sales: An examination of disaggregated drivers and the moderating effect of product class," *J. Bus. Res.*, vol. 90, pp. 286–294, Sep. 2018. doi: 10.1016/j.jbusres.2018.05.020.

[35] G. I. Doukidis, K. Pramatari, and G. Lekakos, "OR and the management of electronic services," *Eur. J. Oper. Res.*, vol. 187, no. 3, pp. 1296–1309, 2008. doi: 10.1016/j.ejor.2006.09.014.

[36] S. Mohammadi, N. Esmaeily, and N. Salehi, "Prioritization of promotion tools based on AIDA model by Analytic Hierarchy process in production sector of sport industry," *Arch. Appl. Sci. Res.*, vol. 4, no. 4, pp. 1670–1675, 2012.

[37] C. Dorn, T. Burkhart, D. Werth, and S. Dustdar, "Self-adjusting recommendations for people-driven ad-hoc processes," in *Proc. 8th Int. Conf. Bus. Process Manage.*, vol. 6336, Hoboken, NJ, USA, 2010, pp. 327–342. doi: 10.1007/978-3-642-15618-2_23.

[38] T. E. Barry and D. J. Howard, "A review and critique of the hierarchy of effects in advertising," *Int. J. Advertising*, vol. 9, no. 2, pp. 121–135, 2015. doi: 10.1080/02650487.1990.11107138.

[39] R. J. Lavidge and G. A. Steiner, "A model for predictive measurements of advertising effectiveness," *J. Marketing*, vol. 25, no. 6, pp. 59–62, 1961. doi: 10.2307/1248516.

[40] B. S. Wijaya, "The development of hierarchy of effects model in advertising," *Int. Res. J. Bus. Stud.*, vol. 5, no. 1, pp. 73–85, 2012. doi: 10.21632/irjbs.5.1.73-85.

[41] K. Yoon, R. N. Laczniak, D. D. Muehling, and B. B. Reece, "A revised model of advertising processing: Extending the dual mediation hypothesis," *J. Current Issues Res. Advertising*, vol. 17, no. 2, pp. 53–67, 1995. doi: 10.1080/10641734.1995.10505032.

[42] J. Zhu, J. Hong, and J. G. Hughes, "Using Markov chains for link prediction in adaptive Web sites," in *Proc. 1st Int. Conf. Soft Issues Design, Develop., Oper. Comput. Syst.*, vol. 2311, D. Bustard, W. Liu, and R. Sterritt, Eds. 2002, pp. 60–73. doi: 10.1007/3-540-44566-8_51.

[43] J. Qiu, Z. Lin, and Y. Li, "Predicting customer purchase behavior in the e-commerce context," *Electron. Commerce Res.*, vol. 15, no. 4, pp. 427–452, 2015. doi: 10.1007/s10660-015-9191-6.

[44] C. O. Sakar, S. O. Polat, M. Katircioglu, and Y. Kastro, "Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks," *Neural Comput. Appl.*, vol. 12, pp. 1–16, 2018. doi: 10.1007/s00521-018-3523-0.

[45] C. H. Park, "Online purchase paths and conversion dynamics across multiple Websites," *J. Retailing*, vol. 93, no. 3, pp. 253–265, Sep. 2017. doi: 10.1016/j.jretai.2017.04.001.

[46] M. Mittal, L. M. Goyal, S. Kaur, I. Kaur, A. Verma, and D. J. Hemanth, "Deep learning based enhanced tumor segmentation approach for MR brain images," *Appl. Soft Comput.*, vol. 78, pp. 346–354, May 2019. doi: 10.1016/j.asoc.2019.02.036.

[47] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. J. Amaratunga, "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting," *Appl. Soft Comput.*, vol. 54, pp. 246–255, May 2017. doi: 10.1016/j.asoc.2017.01.015.

[48] A. M. Grachev, D. I. Ignatov, and A. V. Savchenko, "Compression of recurrent neural networks for efficient language modeling," *Appl. Soft Comput.*, vol. 79, pp. 354–362, Jun. 2019. doi: 10.1016/j.asoc.2019.03.057.

[49] H. Deng, L. Zhang, and X. Shu, "Feature memory-based deep recurrent neural network for language modeling," *Appl. Soft Comput.*, vol. 68, pp. 432–446, Jul. 2018. doi: 10.1016/j.asoc.2018.03.040.

[50] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

[51] A. K. Raz, C. R. Kenley, and D. A. DeLaurentis, "A system-of-systems perspective for information fusion system design and evaluation," *Inf. Fusion*, vol. 35, pp. 148–165, May 2017. doi: 10.1016/j.inffus.2016.10.002.

[52] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, R. Anil, Z. Haque, L. Hong, V. Jain, X. Liu, and H. Shah, "Wide & deep learning for recommender systems," in *Proc. 1st Workshop Deep Learn. Recommender Syst.*, Boston, MA, USA, 2016, pp. 7–10. doi: 10.1145/2988450.2988454.

[53] K. N. Lemon, T. B. White, and R. S. Winer, "Dynamic customer relationship management: Incorporating future considerations into the service retention decision," *J. Marketing*, vol. 66, no. 1, pp. 1–14, 2002. doi: 10.1509/jmkg.66.1.1.18447.

[54] B. Baesens, S. Viaene, D. Van den Poel, J. Vanthienen, and G. Dedene, "Bayesian neural network learning for repeat purchase modelling in direct marketing," *Eur. J. Oper. Res.*, vol. 138, pp. 191–211, Apr. 2002. doi: 10.1016/s0377-2217(01)00129-1.

[55] C. Wu and H.-L. Chen, "Counting your customers: Compounding customer's in-store decisions, interpurchase time and repurchasing behavior," *Eur. J. Oper. Res.*, vol. 127, no. 1, pp. 109–119, 2000. doi: 10.1016/s0377-2217(99)00326-4.

[56] J. Ke, H. Zheng, H. Yang, and X. Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 591–608, Dec. 2017. doi: 10.1016/j.trc.2017.10.016.

[57] P. Yan and Z. Dao-Qiang, "Semi-supervised canonical correlation analysis algorithm," *J. Softw.*, vol. 19, no. 11, pp. 2822–2832, 2008. doi: 10.3724/SP.J.1001.2008.02822.

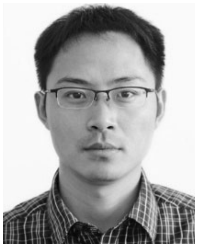
[58] E. S. Soegoto and A. Marica, "Influence of Internet marketing on concert ticket purchasing," in *Proc. IOP Conf. Ser., Mater. Sci. Eng.*, vol. 407, 2018, Art. no. 012015. doi: 10.1088/1757-899x/407/1/012015.

[59] S. Rendle, "Factorization machines," in *Proc. IEEE Int. Conf. Data Mining*, Washington, DC, USA, Dec. 2010, pp. 995–1000. doi: 10.1109/ICDM.2010.127.

[60] Y. Juan, Y. Zhuang, W.-S. Chin, and C.-J. Lin, "Field-aware factorization machines for CTR prediction," in *Proc. 10th ACM Conf. Recommender Syst.*, Boston, MA, USA, 2016, pp. 43–50. doi: 10.1145/2959100.2959134.

[61] B. Zhu, B. Baesens, and S. K. L. M. vanden Broucke, "An empirical comparison of techniques for the class imbalance problem in churn prediction," *Inf. Sci.*, vol. 408, pp. 84–99, Oct. 2017. doi: 10.1016/j.ins.2017.04.015.

[62] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002. doi: 10.1613/jair.953.



CHEN LING was born in Anhui, China, in 1980. He received the B.S. degree in science from Shaanxi Normal University, in 2003, and the M.S. degree in science from East China Normal University, in 2008. He is currently pursuing the Ph.D. degree in management science and engineering with the Shanghai University of Finance and Economics. He is also a Lecturer with the Shanghai University of Medicine and Health Sciences. He has authored more than four journal papers. His current research interests include business intelligence, big data, artificial intelligence, and computer vision.

TAO ZHANG was born in Henan, China, in 1970. He received the B.S. and M.S. degrees in engineering from Northeastern University, Shenyang, Liaoning, and the Ph.D. degree in engineering from Northeastern University, in 2000. He has been a Professor with the Shanghai University of Finance and Economics, since 2004. His current research interests include ERP, supply chain management, big data, and system engineering.

YUAN CHEN received the M.S and Ph.D. degrees in management from Northeastern University, Shenyang, China, in 2003 and 2008, respectively. Since 2012, she has been an Associate Professor with the Electronic Commerce Department, Shanghai University of Finance and Economics, Shanghai, China. She has published more than 20 papers in journals and conferences in information systems and management science. Her expertise is UGC and customers' online behavior.