Multitask Model with an Attention Mechanism for Sequentially Dependent Online User Behaviors to Enhance Audience Targeting

Marwa Hamdi El-Sherief, Mohamed Helmy Khafagy, Asmaa Hashem Sweidan
Faculty of Computers and Artificial Intelligence, Fayoum University, 63514 Egypt

***Abstract*—This paper proposes a multitask learning approach with an attention mechanism to predict audience behavior as sequential actions. The goal is to improve click-through and conversion rates by effectively targeting audience behavior. The proposed model introduces specific task sets designed to address the challenges specific to each prediction task. In particular, the first task, click prediction, suffers from data sparsity and a lack of prior knowledge, limiting its predictive power. To address this, a one-dimensional convolutional network (1D CNN) tower is used in the first task to learn local dependencies and temporal patterns of user activity. This design choice allows the model to better detect potential clicks, even without rich historical data. The task of conversion prediction is tackled by a fully connected convolution tower that selectively combines the corresponding features extracted from the first task using an Attention Mechanism, as well as the original shared embedding input data, enabling richer context for performing more accurate prediction. Experimental results show that the proposed multitask architecture significantly outperforms existing state-of-the-art models that do not consider tower architecture design to predict sequential online audience behavior.**

Keywords*—*Multitask learning; 1D convolution neural networks; attention mechanism; click through rate; conversion rate; audience behavioral targeting; audience behavior

I. Introduction

In Internet advertising, audience targeting delivers the right message to the right audience at the perfect time. It is a multifaceted process that depends on several factors and how they work together. Recently, researchers have become increasingly interested in using machine learning to improve audience targeting as the market's online advertising needs grow. Behavioral targeting (BT) [1], which considers the behavior of the online user, such as clicking links, visiting websites, submitting forms, sending messages, making purchases, etc., is one of the most efficient techniques for improving the target audience [2].

To reach the ultimate business goal of conversion, the user behavior in an advertising campaign takes a multistep path (impression $\rightarrow$ click $\rightarrow$ conversion). The terms“ impression”, “click”, and “conversion” denote the frequency of an online advertisement's display, clicks, and conversions, which include purchases that include the number of clicks that result in an action. The ratio of clicks on an online advertisement to the times the online ad is presented is called the click-through rate, or “CTR” [3]. In contrast, the conversion rate (CVR) is the ratio of the number of people who convert to the number of clicks at first; efforts were focused on developing different models to predict the click-through rate (CTR) [4], which gauges the effectiveness of the campaign. It is difficult to capture complex interactions in advertising systems using traditional machine learning classification methods, such as linear regression, support vector classification (SVC), and decision trees, particularly when high-dimensional data are present. Furthermore, a variety of deep learning architectures are employed in advertising systems to capture interactions between high-order features, including long-short-term memory (LSTM) and convolution neural networks (CNN) [5] and [6].

However, calculating the exact return on investment (ROI) has become crucial for marketers dealing with Internet advertising. Due to the complexity of user behavior, both the CTR and CVR predictions are required. The issue of limited data and delayed feedback is the main barrier to the prediction of the conversion rate [7]. Furthermore, clicks and conversion behaviors are sequential and depend on a multi-step conversion path. Recently, multitask models have used click behavior to address the problem of sparse data in conversion behavior. Deep learning researchers are actively investigating multitask learning (MTL) [8]. Through shared information, this learning paradigm seeks to learn several related tasks to enhance performance and generalization collaboratively. MTL has been used effectively by several academics in a variety of areas recently, including recommendation systems [9], computer vision, natural language processing, and reinforcement learning.

Multitask learning has been frequently applied to improve end-to-end conversion rates in the audience's multistep conversion path. According to task relations, multitask learning is divided into parallel, cascaded, and auxiliary tasks in multitask deep recommendation systems [9]. In this instance, a sequential dependency between tasks is termed a cascading task connection, where the prior task influences the computation of the current task. Numerous cascaded task relation models have been presented in user behavior sequence prediction to address the problems of overfitting and data sparsity in training sparse conversion behavior data. The authors in [10] and [11] proposed multigate mixture of experts (MMoE) and progression-layered extraction (PLE), respectively, taking into account CTR and CVR as non-sequential tasks. To address the problem of the sequential nature of the user behavior input data, [12] proposed the mixture of sequential experts (MoSE) to represent sequential behavior using long-short-term memory (LSTM) in cutting-edge multigate mixture of expert multitask models. However, this technique lacks information sharing between the top-level towers in the model, which holds valuable information from the previous tasks to model click-through rate (CTR) and conversion rate (CVR) prediction by two auxiliary tasks. Using an estimated conditional probability in the CVR task, the sequential dependence between user activities is taken into account in [13], [14], and [15].

To improve model performance, ESM2 [15] specifically masks conversion-related information during click prediction, therefore addressing the constraints of ESMM [13]. However, severe information loss might result from incorrect probability estimates. To overcome the difficulties of multitask learning with sequentially dependent tasks among multistep conversions, where prior tasks exchange valuable information with the subsequent task at the top of the tower [16], [17], and [18] proposed AITM, PIMM, and MNCM, respectively. Although many multitask learning models have been proposed for CTR and CVR prediction, most of them either overlook the sequential dependence across user behaviors or fail to share high-level task-specific information effectively among tasks. For example, models such as ESMM and MMoE don't explicitly model the action sequence of users or transfer informative signals from clicks to conversions using complicated mechanisms. Although recent models such as AITM, PIMM, and MNCM consider the sequential nature of user behavior, they still lack emphasis on the architectural design of task-specific towers. To the best of our knowledge, most of these models adopt simple feedforward networks as towers and overlook the potential benefits of using diverse tower structures. This is particularly crucial for the click prediction task, which suffers from a lack of prior knowledge and would greatly benefit from employing a more expressive architecture, such as a 1D convolutional network, to represent local temporal dependencies in user behavior. To address this gap, this paper proposes a novel multitask model that consists of a 1D CNN tower to improve the click prediction task and an attention-based mechanism to transfer beneficial information into the conversion prediction task. The proposed design addresses the challenges of data sparsity, sequential dependency, and task interaction of online user behavior prediction.

As a result, the following summarizes all the primary contributions in this paper:

1) Proposed a multitask model to improve audience behavioral targeting.
2) Shared embedding vector for all tasks.
3) Using 1D CNN improves the first task tower.
4) Use the attention mechanism to identify useful information from the information provided.
5) Processing of the Ali-CCP dataset.
6) Pre-processing of the Taobao dataset and feature engineering to be relevant to our work.
7) The data loader to access processed data that can be parallelized and shuffled.

Therefore, the remainder of this paper is structured as follows: Section II reviews the related work in multitask learning and user behavior modeling. Section III presents the proposed architecture in detail. Section IV explains the datasets, the preprocessing pipeline, and the experimental setup. Section V discusses the experimental results and performance evaluation. Finally, Section VI concludes the paper and suggests directions for future research.

II. Related Work

This section reviews recent studies on online audience behavior targeting using machine learning [19]. Previous research has shown that machine learning greatly improves the online audience-targeting process. Recent research uses multitask models to improve multistep user conversion path (click → conversion) returns [9]. According to task relation, models are classified into non-sequential tasks and sequential dependence tasks, as shown in Table I.

A. Non_Sequential Tasks

In studies [10], [11], and [12], tasks are non-sequential and are modeled separately. The multi-gate mixture of experts (MMoE) [10] uses network gates to bring together experts for various tasks, improving the model's capacity to capture complex task-specific patterns. The model AUCs are 0.6420 for purchases and 0.6047 for clicks.

The authors of study [11] presented a progressive layered extraction (PLE) model that clearly distinguishes between task-specific and task-shared experts to address the seesaw problem, which occurs when improving one task's performance may affect the performance of another task. The AUC models for click and conversion are, respectively, 0.6039 and 0.6417.

The authors of study [12] introduced MoSE. This model models user activity streams using long-short-term memory (LSTM) and offers a comprehensive sequential solution with a gated mixture of experts. After extensive testing with real-world G Suite user data, MoSE outperforms the production model with an AUC score of +4.8%.

In studies [10], [11], and [12], the mixture of experts shares expert modules across all tasks at the bottom of the multi-task model. However, the inability to exchange information between tasks in top towers, which contain richer and more useful information, may limit their ability to improve each other.

B. Sequential Dependence Tasks

Models are created to handle dependency-based actions, which means that when the first step is completed, the subsequent step could occur due to a sequential dependency.

The authors proposed the Entire Space Multitask Model (ESMM) in [13], which takes into account the sequential dependence between tasks and the probability of transfers for various activities to calculate the click and conversion rate (CTCVR) by calculating the product of (CVR) and (CTR). The model gets AUCs of 0.6022 and 0.6291 for click and conversion, respectively.

The authors of study [15] introduced ESM^2 to deliberately hide conversion-related information during click prediction to overcome the shortcomings of ESMM [13]. This method enhances model performance by using conditional probability rules based on user behavior graphs to convey fundamental information. However, it ignores richer representations in vector space, which results in a severe loss of information if any probability prediction is off.

In study [16], the authors proposed an AITM (Adaptive Information Transfer Multitask) to simulate multistep conversions and sequential reliance of the audience. To improve the performance of sequentially dependent multitask learning, the suggested adaptive information transmission (AIT) module combines the behavioral expectation calculator in the loss function to acquire knowledge of what and how much information to transmit for different conversion phases. With the Ali CCP dataset, AITM achieves AUCs of 0.6043 and 0.6525 for click and purchase, respectively. The authors of study [18] proposed the multilevel network cascades model (MNCM) with two adaptive information transfer modules, the task-level information transfer module (TITM) and the expert-level information transfer module (EITM), to address the information lacking in the first task. They achieve AUCs of 72.15, 87.16, 71.06, and 86.44 on click and buy, respectively, using the AliExpress-NL and AliExpress-US datasets. The authors used the prior information merged model (PIMM) to learn sequentially dependent tasks in [17]. The PIMM combines explicit premise information (i.e., probability of previous positive reinforcement) with latent representations (specialized knowledge) to strictly describe the logical dependency among tasks as learning several sequential dependence tasks under a curriculum-structured guidance. Using a soft sampling method, PIM randomly chooses the real label information to transfer to the downstream task during training, adhering to a curriculum paradigm that ranges from basic to advanced. They obtain AUCs of 0.6075 and 0.6561 for click and buy, respectively, using the Ali CCP dataset.

Most of the reviewed literature addressed the problem of targeting online audience behavior, and current models often fail to account for the dynamic nature of user behavior, which changes over time due to different factors, including evolving interests, outside influences, and seasonal trends. To bridge these gaps, the proposed approach in this paper uses a CNN tower for the click task to handle the absence of prior information in the first task. Furthermore, the attention method is utilized to convey pertinent information from the click task to manage sparse data in the conversion task.

III. The Proposed Model

In this paper, the methodology used to apply a multitask model with an attention mechanism [20] to anticipate the behavior of an online audience, which is intended to predict user clicks and conversion behaviors while accounting for the sequential dependency between these events; this means that the conversion event depends on the click event that occurred previously. Furthermore, as seen in Fig. 3, the shared embedding layer, the click tower, the conversion tower, the attention mechanism, and the click information extraction are the components of the proposed approach.

The approach consists of the following main phases:

A. Preprocessing of Data

The preparation of the data phase is crucial to the proposed model. To deal with various types of characteristics in the input data and to guarantee that the input data is consistent with the nature of sequentially dependent events. The preparation step consists of the following individual stages:

1) Data cleaning and transformation: The data cleaning and transformation process involved several key steps in preparing the dataset for modeling. First, missing values were identified and eliminated, as they did not significantly affect the performance of large datasets. Then, low-frequency characteristics, those with fewer than ten occurrences, were removed to enhance the conciseness and relevance of the input data. The numerical features were normalized to the same scale as the variables. Categorical characteristics were encoded to convert them to a numeric form so that they could be analyzed. In addition, timestamp fields were converted into a uniform date-time format to support temporal analysis. Finally, the data were ordered by user and date time to meet the sequential input data requirements.

2) Feature engineering: Generate additional features based on past user behavior.

3) Data splitting: divide the dataset into testing, validation, and training sets based on the number of days.

B. Data Loader

In the proposed model, the data loader plays a critical role in optimal data batching, shuffling, and loading. It begins by partitioning data and pulling associated feature names, and targets with sequential dependencies such as clicks and conversions. As an indication of process optimization, the data are batched according to a specified batch size. In addition, the data are randomized during training to improve the model's generalization capacity. Importantly, the data are dynamically loaded, which is extremely beneficial when dealing with a large dataset.

C. Shared Embedding Layer

During this stage, the embedding methods convert all input characteristics into low-dimensional dense vectors of a given size [21]. To create a common embedding module, all embedding vectors must be concatenated. By sharing the embedding layer between all tasks, the model can benefit from rich positive samples of previous tasks, which promotes information sharing and reduces the impact of the class imbalance conversion task. Sharing the embedding layer also contributes to reducing the total number of model parameters.

D. Click Task

This stage consists of two phases: click tower and click probability. Since the click tower, which represents the first task, has a significant impact on all subsequent tasks, as it suffers from a lack of useful information, since there are no previous tasks in this stage, this paper presents a new construction for the click tower, which addresses the representation of features using one-dimensional convolutional neural networks [22] as the tower. 1D Convolution Neural Networks are a unique kind of CNN in which the kernel analyzes one-dimensional input, such as sequential and temporal series. In CNN, the output size No is estimated as in Equation (1), the kernel size K refers to the size of the sliding kernel or the kernel filter, The number of kernels that slide before producing the output and the product points is known as the stride length S, and the size of the 0-Th frame that surrounds the input feature map N_n is known as padding P [23]. The proposed

TABLE I. Summary of Surveyed State-of-the-Art Multitask Models for Online user Behavior Prediction

Study	Model	Dataset(s)	Model Type	Tasks	Evaluation Metric(s)	Year
[13]	**ESMM**	Taobao's recommender system logs	DNN	CVR, CTR, CTCVR, SPP	Multi-AUC	2018
[10]	**MMOE**	Synthetic dataset, UCI Census	Neural network (8 hidden units)	CTR, CVR	AUC	2018
[11]	**PLE**	Census-income, synthetic data, Ali CCP	LSTM	CTR, CVR	AUC	2020
[12]	**MoSE**	Generated synthetic dataset	MLP	Task1, Task2	PR-AUC	2020
[16]	**AITM**	Ali CCP	MLP	CTR, Buy, Approval, Activation	AUC	2021
[18]	**MNCM**	AliExpress	MLP	CTR, CVR	AUC	2022
[17]	**PIMM**	Ali CCP	MLP	CTR, CVR	AUC	2023

model used multiple layers of a 1D convolution network to process the shared and embedded input characteristics, varying the stride and padding for every kernel size as specified by Eq. (1). Furthermore, weight normalization is handled by ReLU activation functions, and regularization is accomplished via the dropout approach. To pass through fully connected layers, the output of the 1D convolution layers is flattened, as shown in Fig. 1. Furthermore, the second phase is the click probability, which uses a sigmoid activation function to process the output data from the click tower after passing through a linear neural network to determine the output probability.

$$\mathbf{N_o} = \left(\frac{N_n - \mathrm{K} + 2 \times \mathrm{P}}{\mathrm{S}}\right) + 1 \tag{1}$$

E. Information Extraction

For the conversion task and other behaviors that follow, the information extraction phase is essential to improve the prediction process; learning is transferred between tasks through the attention mechanism, as shown in Fig. 2; the attention mechanism enhances the model's ability to recognize pertinent parameters required for the conversion task [24], and [25], which frequently involves sparse input. The attention mechanism that combines the information from the click information tower and the conversion tower, as shown in Fig. 3, where the feed-forward networks $Q(a)$, $K(a)$, and $V(a)$ represent the input vector to a single new vector representation, dot attention uses a dot product to determine how comparable queries $Q(a)$ and Key $K(a)$ are and scaled by dimension $\sqrt{d}$ in Eq. (2), and Eq. (3) uses a softmax function to determine attention weights w_a. Eq. (4) computes z_t, the weighted sum of the value vectors $V(a)$ using these attention weights w_a.

$$\mathbf{w_a} = \frac{Q(a) \cdot k(a)}{\sqrt{d}} \tag{2}$$

$$\mathbf{w_a} = \mathrm{softmax}(w_a) = \frac{\exp(w_a)}{\sum_a \exp(w_a)} \tag{3}$$

$$\mathbf{z_t} = \sum_a w_a \cdot V(a) \tag{4}$$

F. Conversion Task

This phase handles subsequent tasks using valuable knowledge obtained from the previous task, specifically from the click tower. It begins with the conversion tower, where a fully connected neural network is used to process input coming from the shared input embedding vector. Then, an information vector is generated using a feedforward neural network that models the output of the click tower. To enhance the learning process, an attention mechanism is applied to the output of the conversion tower, which is concatenated with the information vector of the previous task. In this step, the model can successfully utilize the knowledge acquired in the earlier stages. Lastly, the probability of conversion is found by feeding the attention output into a linear neural network, then applying the sigmoid activation function to get the final prediction.

G. Loss Calculation

The final loss function L_f in our suggested model is determined by calculating the loss of cross entropy L_c and the behavioral expectation calibrator L_{bc} combined with a constant weighting parameter α that regulates the behavioral force expectation calibrator as indicated in Eq. (7). The loss is determined using the binary cross-entropy loss (L_c) indicated in Eq. (5), where y is the label, $\hat{y}$ is the target value, T is the task, and N is the number of samples. In addition, a calibration of behavioral expectations (L_{bc}) indicated in Eq. (6) is added to ensure that the model result meets the actual production constraints, where the click task is expected to have a higher probability than conversion for the same user because click and conversion are sequentially dependent behaviors.

$$\mathbf{L_c}(y, \hat{y}) = -\frac{1}{N} \sum_T \sum_N \left(y \log(\hat{y}) + (1-y)\log(1-\hat{y})\right) \tag{5}$$

$$\mathbf{L_{bc}} = \frac{1}{N} \sum_T \sum_N \max(\hat{y}_{cl} - \hat{y}_{co}, 0) \tag{6}$$

$$\mathbf{L_f} = L_c + \alpha L_{bc} \tag{7}$$

H. Evaluation

The proposed approach adopts the AUC, or Area Under the ROC Curve, which is a commonly used evaluation metric in

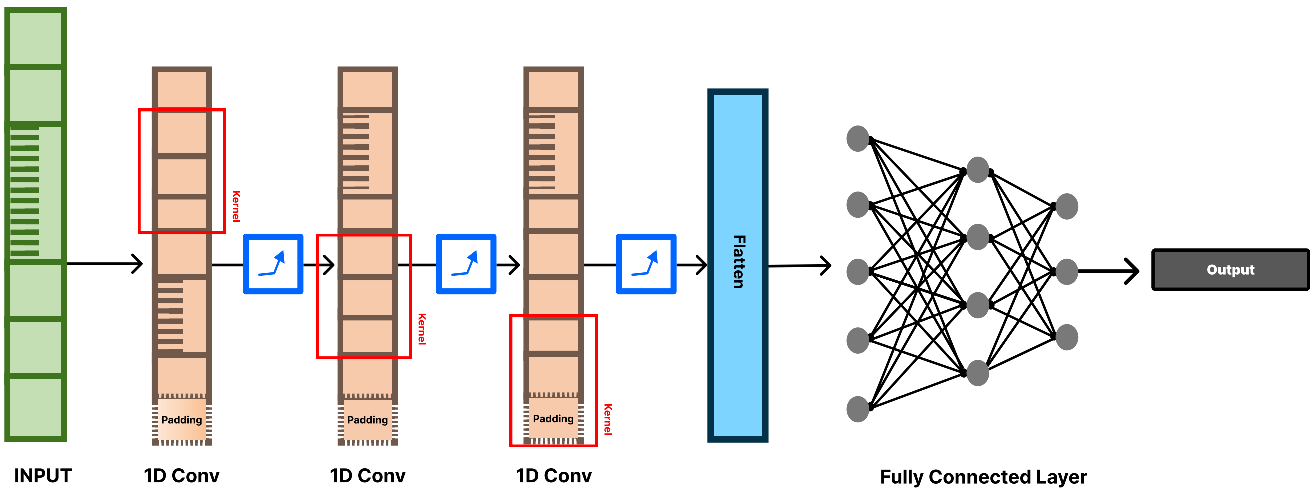

Fig. 1. 1D Convolution network.

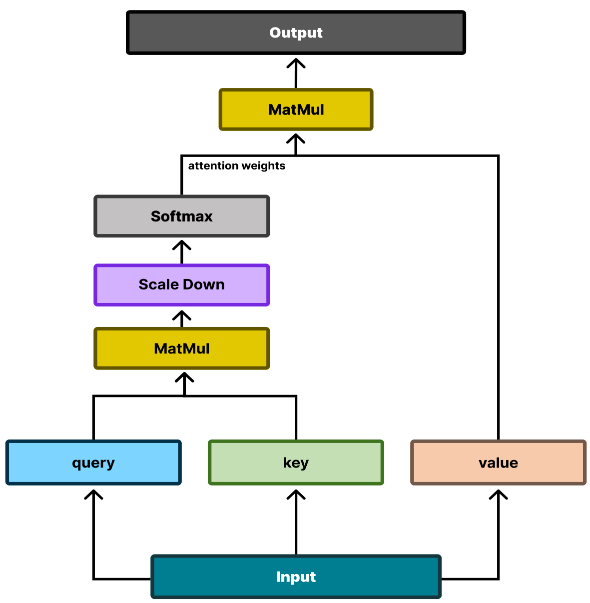

Fig. 2. Attention mechanism.

recommendation systems. Shows the probability that an item that has been clicked will rank higher than an item that has not been clicked.

IV. Results and Discussion

A. Datasets and Evaluation Metrics

Two public datasets were utilized in this study: Ali-CCP [13] and Taobao User Behavior for recommendation [26].

1) Ali-CCP Dataset [13]: This dataset is an actual traffic log from Taobao's recommendation system. The end-to-end user conversion process consists of several consecutive phases, starting from impression, moving to click, and eventually conversion (i.e., impression $\rightarrow$ click $\rightarrow$ conversion). Each observed impression is associated with a feature vector, denoted by x, representing both the user and item information. The label format is (x, y $\rightarrow$ z), where y and z are binary variables indicating whether a click (1) or no click (0), and a conversion (1) or no conversion (0), occurred, respectively. Here, as sequential tasks, the conversion can only happen if a click has occurred beforehand. The dataset exhibits significant data sparsity, as evidenced by the calculated click-through rate (CTR) and conversion rate (CVR), which are 3.89% and 0.54%, respectively, as illustrated in Fig. 4.

In the data preprocessing phase, the dataset is divided into subsets for training, validation, and testing. 50% of the data is allocated for training, 10% for validation, and the remaining 40% for testing. Low-frequency features—those with fewer than ten occurrences—are excluded, and categorical features are appropriately coded to prepare the data for input into the model. Analysis of the dataset shows that the most influential factor is the product categories the user has previously clicked on. This result demonstrates the importance of modeling user behavior history, as it has a significant impact on the prediction goal.

2) User behavior data from Taobao for recommendation dataset [26]: contains around 100 million user activities and serves as an essential resource for research and analysis, particularly in user behavior modeling. The dataset includes user behaviors gathered from one million randomly chosen Taobao users and records a wide range of online activities over nine days. As illustrated in Fig. 5, 846.9k users browse the items, while 6.20% add them to carts, 3.13% favorite them, and 2.27% proceed to purchase.

Regarding the preprocessing of the user behavior data from the Taobao recommendation dataset [26], the proposed approach narrowed the study to a nine-day observation period to check the fitness of the data and accelerate the processing time. It also determined the sequential dependency of the "buy"

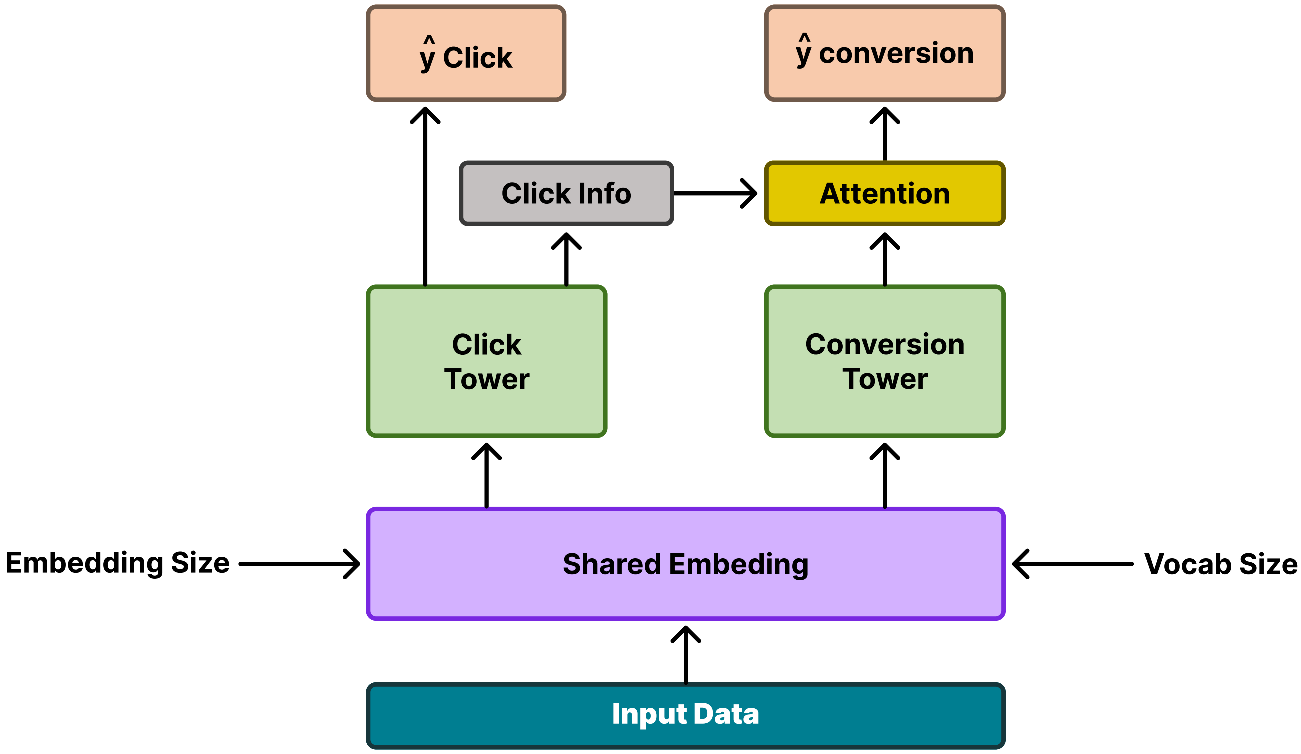

Fig. 3. Proposed model: Multitask model for sequential online user behaviors (click and conversion) with different tower architectures

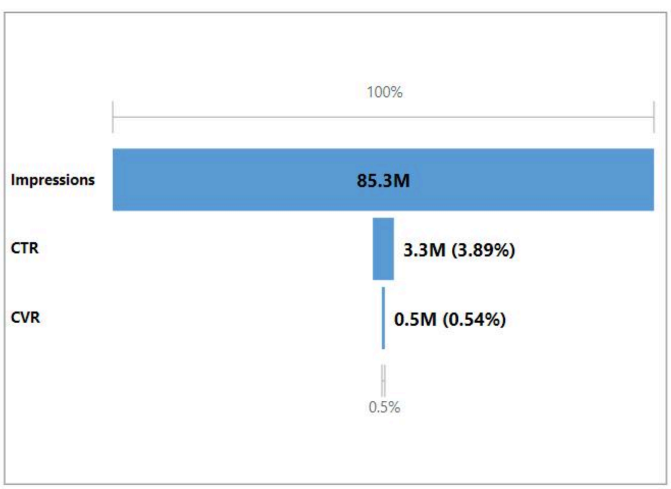

Fig. 4. Statistics of Ali_CCP dataset.

and “add-to-cart” activities, encoding the category feature ”behavior” to distinguish between different activities. Data cleaning followed, excluding non-conforming products from the desired browsing-to-purchase pattern. The data preparation to supply deep learning models was ordered chronologically by user and date. This form allows the model to learn time dependencies and improve the model’s ability to identify patterns and make consistent predictions. Furthermore, the presentation of the events in time-ordered sequence is consistent with the natural progression of user interactions and therefore improves the model’s ability to learn meaningful information. To feature engineer the user behavior data of the Taobao recommendation dataset [26], additional historical data was incorporated, including the user’s previous activities and the product categories on which the user previously clicked. This enhancement provides a contextual understanding of user behavior with more enrichment so that analysis can be deeper and more effective.

V. Experimental Analysis

This article’s studies were carried out using a PC equipped with a 16 GB GPU, 32 GB of RAM, and a 2.7 GHz Intel Core i7 CPU. The PyTorch Python libraries were used to develop our suggested strategy. Data are preprocessed using Apache Spark in Python (PySpark) to handle massive volumes of data in a distributed processing environment, especially to generate new historical characteristics that increase the size of the data. Two distinct open datasets were used to evaluate the suggested hypothesis. 4.10 GB of compressed train data and 4.68 GB of compressed test data make up the Ali CCP dataset. In comparison, the Taobao user behavior data used for the recommendation weighs 5 GB after compression.

The dictionary of vocabulary size in the embedding layer is set to the unique value of each input feature, and the embedding size is set to five. The proposed approach was experimented with different kernel sizes for the 1D Convolution

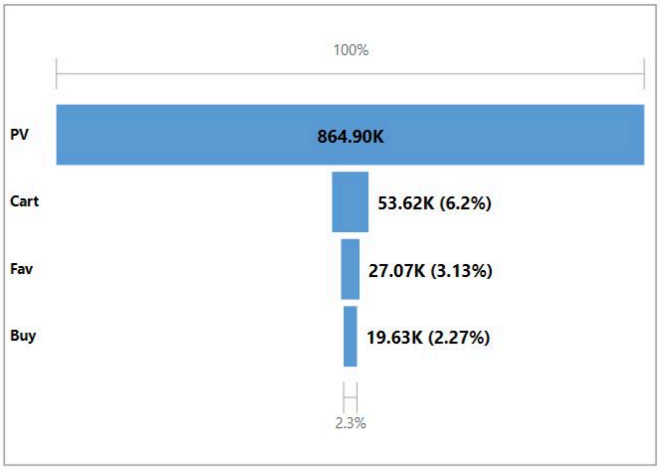

Fig. 5. Statistics of user behavior data from Taobao for recommendation dataset.

TABLE II. Evaluation Metrics for Clicks on Ali-CCP Public Dataset for Different Models

Model	AUC	Accuracy	Recall	Specificity
AITM Model	0.614	0.6	0.61	0.61
Proposed Model	0.661	0.64	0.66	0.66

tower to obtain the optimal AUC and adjusted the padding by Eq. (1). The input dimensions for the proposed approach are 128, 64, and 32 in the linear tower. The proposed approach was trained with a batch size of 1,000 samples over five epochs.

A. Comparative Evaluation of State-of-the-Art Models

This subsection compares the performance of the suggested model with the most advanced model, AITM. The performance of the proposed model was evaluated against the AITM model using two public datasets, Ali-CCP and Taobao, using evaluation metrics such as AUC, accuracy, recall, and specificity. ROC curves are presented to visually represent the performance of the models. Fig. 6 shows the ROC curve for the AITM model in the public dataset of the Taobao recommendation. Similarly, the ROC curves for the proposed model are shown in Fig. 7 for the Ali CCP public dataset and Fig. 8 for the Taobao recommendation public dataset. These curves highlight the comparative ability of each model to distinguish between classes.

1) Model performance on Ali CCP dataset: In terms of model performance on the Ali CCP dataset, the click task results show that the proposed model achieved an AUC of 0.661, which was better than the AITM model with an AUC of 0.614, as presented in Table II and Fig. 9. For the conversion task, the AUC of the proposed model increased further to 0.694 compared to 0.6320 for the AITM model, demonstrating the improved performance of the proposed model, as presented in Table III and Fig. 10.

2) Model performance on the taobao dataset: Concerning the performance of the model in the Taobao dataset, the results of the click task show that the proposed model obtained an AUC of 0.682, compared to that of the AITM model, which

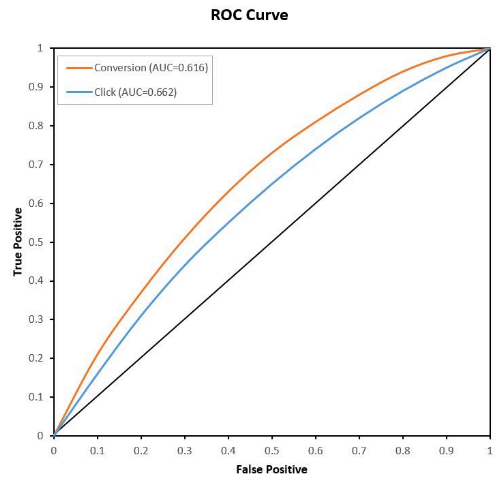

Fig. 6. ROC Curve for AITM model on the Taobao recommendation public dataset.

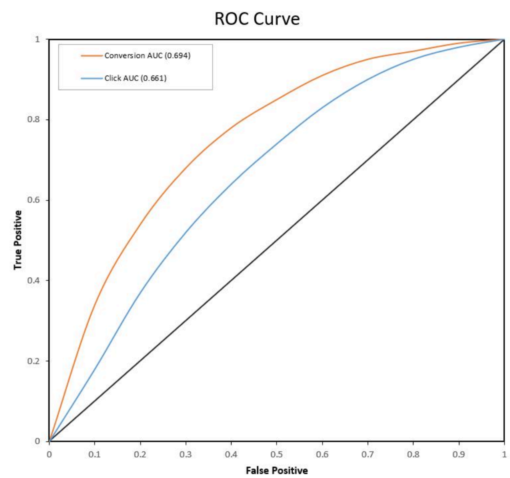

Fig. 7. ROC Curve for the proposed model on Ali-CCP public dataset.

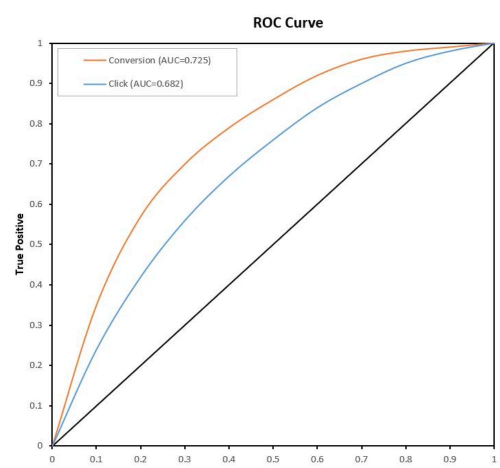

Fig. 8. ROC Curve for the proposed model on the Taobao recommendation public dataset.

TABLE III. EVALUATION METRICS FOR CONVERSION ON ALI-CCP PUBLIC DATASET FOR DIFFERENT MODELS

Model	AUC	Accuracy	Recall	Specificity
AITM Model	0.632	0.62	0.63	0.63
Proposed Model	0.694	0.68	0.69	0.69

TABLE IV. EVALUATION METRICS FOR CLICKS ON THE TAOBAO RECOMMENDATION PUBLIC DATASET FOR DIFFERENT MODELS

Model	AUC	Accuracy	Recall	Specificity
AITM Model	0.616	0.6	0.62	0.62
Proposed Model	0.682	0.65	0.68	0.68

TABLE V. EVALUATION METRICS FOR CONVERSION ON THE TAOBAO RECOMMENDATION PUBLIC DATASET FOR DIFFERENT MODELS

Model	AUC	Accuracy	Recall	Specificity
AITM Model	0.662	0.63	0.66	0.66
Proposed Model	0.725	0.7	0.73	0.73

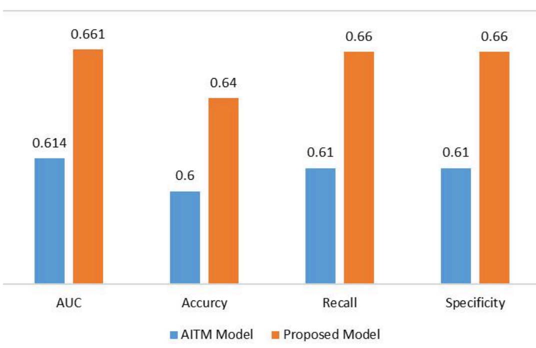

Fig. 9. Evaluation metrics for click on Ali-CCP public dataset for different models.

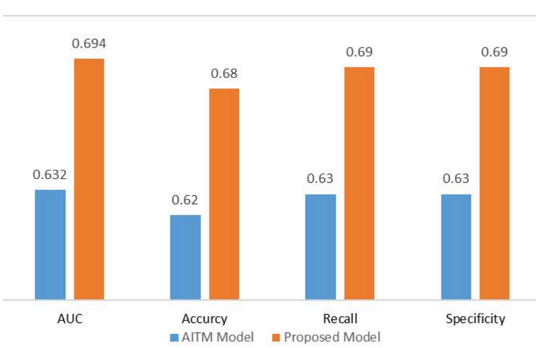

Fig. 10. Evaluation metrics for conversion on Ali-CCP public dataset for different models.

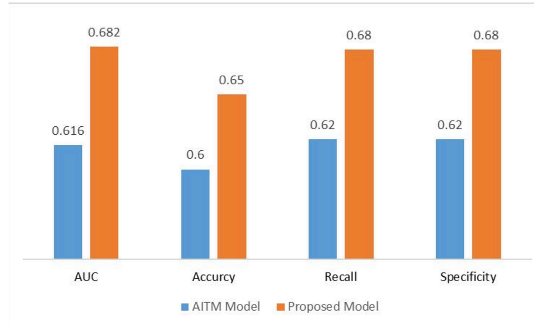

Fig. 11. Evaluation metrics for clicks on the Taobao recommendation public dataset for different models.

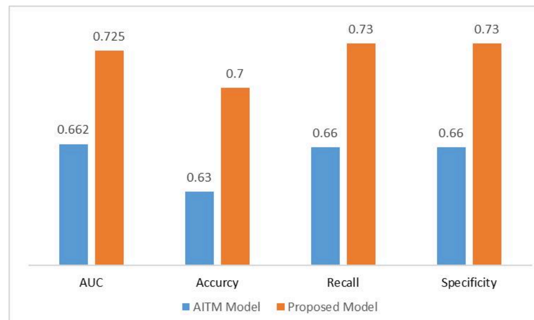

Fig. 12. Evaluation metrics for conversion on the Taobao recommendation public dataset for different models.

obtained an AUC of 0.616, as shown in Table IV and Fig. 11. As illustrated in Table V and Fig. 12, the differences were further noticeable for the conversion task, where the suggested model received an AUC of 0.725 compared to 0.662 for the AITM model.

Furthermore, the results suggest that the Taobao dataset may have slightly different structural properties than the Ali CCP dataset, resulting in marginally better model performance. This insight into the impact of the characteristics of the data set on model performance is valuable for future research.

VI. CONCLUSION

To improve audience targeting in online advertising, this research proposes a novel multitask model to evaluate sequential user behavior online, which could help with audience targeting. The method presents a unique multitasking model with several tower architectures specific to the task. For the click task (the first task), the proposed approach uses 1D convolutional neural networks without prior knowledge to improve audience targeting by focusing on those most likely to click. In the conversion task (the subsequent task), which utilizes information from the first task, a fully connected tower is used along with an attention mechanism. The suggested method outperforms another state-of-the-art multitask model (AITM) [16] in simulating user behavior with sequential online dependencies. Future studies may need to investigate more intricate tower architectures to overcome the restrictions above.

REFERENCES

[1] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen, "How much can behavioral targeting help online advertising?" in *proceedings of the 18th International Conference on World Wide Web*, 2009, pp. 261–270, https://doi.org/10.1145/1526709.1526745.

[2] J.-A. Choi and K. Lim, "Identifying machine learning techniques for classification of target advertising," *ICT Express*, vol. 6, no. 3, pp. 175–180, 2020, https://doi.org/10.1016/j.icte.2020.04.012.

[3] M. Richardson, E. Dominowska, and R. Ragno, "Predicting clicks: estimating the click-through rate for new ads," in *proceedings of the 16th International Conference on World Wide Web*, 2007, pp. 521–530, https://doi.org/10.1145/1242572.1242643.

[4] Y. Yang and P. Zhai, "Click-through rate prediction in online advertising: A literature review," *Information Processing & Management*, vol. 59, no. 2, p. 102853, 2022, https://doi.org/10.1016/j.ipm.2021.102853.

[5] P. P. Chan, X. Hu, L. Zhao, D. S. Yeung, D. Liu, and L. Xiao, "Convolutional neural networks based click-through rate prediction with multiple feature sequences." in *IJCAI*, 2018, pp. 2007–2013, https://doi.org/10.24963/ijcai.2018/277.

[6] K. Singh, A. Mahajan, and V. Mansotra, "1d-cnn based model for classification and analysis of network attacks," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, 2021. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2021.0121169

[7] M. Zhang, R. Yin, Z. Yang, Y. Wang, and K. Li, "Advances and challenges of multi-task learning method in recommender system: A survey," *arXiv preprint arXiv:2305.13843*, 2023, https://doi.org/10.48550/arXiv.2305.13843.

[8] Y. Zhang and Q. Yang, "An overview of multi-task learning," *National Science Review*, vol. 5, no. 1, pp. 30–43, 2018, https://doi.org/10.1093/nsr/nwx105.

[9] Y. Wang, H. T. Lam, Y. Wong, Z. Liu, X. Zhao, Y. Wang, B. Chen, H. Guo, and R. Tang, "Multi-task deep recommender systems: A survey," *arXiv preprint arXiv:2302.03525*, 2023, https://doi.org/10.48550/arXiv.2302.03525 .

[10] J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong, and E. H. Chi, "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts," in *proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 1930–1939, https://doi.org/10.1145/3219819.3220007.

[11] H. Tang, J. Liu, M. Zhao, and X. Gong, "Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations," in *proceedings of the 14th ACM Conference on Recommender Systems*, 2020, pp. 269–278, https://doi.org/10.1145/3383313.3412236.

[12] Z. Qin, Y. Cheng, Z. Zhao, Z. Chen, D. Metzler, and J. Qin, "Multitask mixture of sequential experts for user activity streams," in *proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3083–3091, https://doi.org/10.1145/3394486.3403359.

[13] X. Ma, L. Zhao, G. Huang, Z. Wang, Z. Hu, X. Zhu, and K. Gai, "Entire space multi-task model: An effective approach for estimating post-click conversion rate," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 1137–1140, https://doi.org/10.1145/3209978.3210104.

[14] L. Zhan, "Enterprise marketing decision: Advertising click-through rate prediction based on deep neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, 2023. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2023.0140988

[15] H. Wen, J. Zhang, Y. Wang, F. Lv, W. Bao, Q. Lin, and K. Yang, "Entire space multi-task modeling via post-click behavior decomposition for conversion rate prediction," in *proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 2377–2386, https://doi.org/10.1145/3397271.3401443.

[16] D. Xi, Z. Chen, P. Yan, Y. Zhang, Y. Zhu, F. Zhuang, and Y. Chen, "Modeling the sequential dependence among audience multi-step conversions with multi-task learning in targeted display advertising," in *proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 3745–3755, https://doi.org/10.1145/3447548.3467071.

[17] Y. Weng, X. Tang, L. Chen, and X. He, "Curriculum modeling the dependence among targets with multi-task learning for financial marketing," in *proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 1914–1918, https://doi.org/10.1145/3539618.3591969.

[18] H. Wu, "Mncm: multi-level network cascades model for multi-task learning," in *proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 4565–4569, https://doi.org/10.1145/3511808.3557644.

[19] Z. Shang and B. Ge, "Analysis of customer behavior characteristics and optimization of online advertising based on deep reinforcement learning," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 8, 2024. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2024.0150805

[20] S. Fang, X. Cai, Y. Xue, and W. Lu, "Adaptive residual attention recommendation model based on interest social influence," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 6, 2024. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2024.0150671

[21] F. Lyu, X. Tang, H. Zhu, H. Guo, Y. Zhang, R. Tang, and X. Liu, "Optembed: Learning optimal embedding table for click-through rate prediction," in *proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 1399–1409, https://doi.org/10.1145/3511808.3557411.

[22] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1d convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, p. 107398, 2021, https://doi.org/10.1016/j.ymssp.2020.107398.

[23] P. Purwono, A. Ma'arif, W. Rahmaniar, H. I. K. Fathurrahman, A. Z. K. Frisky, and Q. M. ul Haq, "Understanding of convolutional neural network (cnn): A review," *International Journal of Robotics and Control Systems*, vol. 2, no. 4, pp. 739–748, 2023, 10.31763/ijrcs.v2i4.888.

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017, https://doi.org/10.1111/cgf.14912.

[25] S. Liu, E. Johns, and A. J. Davison, "End-to-end multi-task learning with attention," in *proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1871–1880, https://doi.org/10.48550/arXiv.1803.10704 .

[26] H. Zhu, X. Li, P. Zhang, G. Li, J. He, H. Li, and K. Gai, "Learning tree-based deep model for recommender systems," in *proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 1079–1088, https://://10.1145/3219819.3219826.