IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 17 July 2024, accepted 7 August 2024, date of publication 16 August 2024, date of current version 25 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3444787

RESEARCH ARTICLE

Interactive Attention-Based Capsule Network for Click-Through Rate Prediction

SHENG XUE[1], CONGQING HE[2], ZHUXUAN HUA[1], SONGTIAN LI[1], GUANGWEI WANG[1], AND LIWEN CAO[3]

[1]Cloud Network Operation Technology Research Institute, Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China
[2]School of Computer Sciences, Universiti Sains Malaysia, Penang 11800, Malaysia
[3]Department of Economics and Information Management, Shanghai University of Finance and Economics Zhejiang College, Jinhua, Zhejiang 321000, China

Corresponding author: Liwen Cao (zjjhclw199374@163.com)

This paper work was fully supported by the Project Program of China Telecom (No.23HQJ3YF0035-001). The author would appreciate the anonymous reviewers and the editor for their valuable comments.

ABSTRACT With the continuous penetration of Internet applications in our lives, the ever-increasing data on clicking behavior has made online services a critical component of the economic sectors of internet companies over the past decade. This development trend has brought a large amount of information that reflects user needs but is relatively chaotic. Extracting user interests and needs from complex click behaviors is crucial for advancing online business development and precisely targeting product information The interactive attention-based capsules (IACaps) network is proposed in this paper to collate and analyze complex and changing click information for user behavior representation. Specifically, an interactive attention dynamic routing mechanism is proposed to mine the potential association information among different browsing behaviors, which facilitates the extraction and understanding of seemingly irrelevant information hidden in massive click data. To ensure the practicability of the proposed method, three different types of datasets were selected from Amazon Dataset for experiments, and the results of which shows the superior performance of the proposed method when compared with other models. Specifically, the reasonableness and effectiveness of the reported model are further proved by improvements of metrics obtained in the main experiments and ablation studies. Optimization of Hyper-parameters is also analyzed from the number of iterations, the number of capsules, and the dimension of capsules for better understanding of operating principles.

INDEX TERMS Interactive attention, capsules, click-through rate prediction.

I. INTRODUCTION

With the popularization of fourth/fifth generation mobile communication technology (4G/5G) in daily life and production, the average monthly data usage in mobile filed has multiplied with the swiftly shift of communication consumption from voice services to data services in recent years [1]. The rapid increase of data usage indicates the unstoppable trend of online consumption and the huger impact on all aspects of people's daily life with the further development and deeper permeability of Internet. Thus, online business are growing rapidly and have become an important part of economic sector in Internet companies in the past decade [2], [3], [4], [5], [6]. The rapid increase of data usage indicates the unstoppable trend of online consumption. Although capturing the real interests and demands from complex and changing user click behaviors is a critical node to push the development of online business as well as to reduce company cost and accurately push product information, most of valuable information and effective opportunities were hidden in the massive and related chaotic data [7], [8], [9], [10], [11], [12], [13]. Therefore, researches on how to improve the CTR prediction which reflects the effectiveness and attraction of marketing have become a hot direction [2], [14], [15], [16], [17].

In fact, a user's need may be triggered by different series of click behaviors, which means a reasonable and reliable CTR prediction must be based on the analysis of content-related groups, rather than a single, discrete click or nearby clicks.

The associate editor coordinating the review of this manuscript and approving it for publication was P. K. Gupta.

It is hard to directly link a outdoor barbecue demand with tin foil product or weather-related clicks individually [18], [19], but the possibility of outdoor barbecue demand is significantly increased when tin foil product clicks, weather-related clicks and kitchenware clicks are connected together for CTR. Effective information are hidden not only within each click or a same series clicks, but also among different types of clicks and even seemingly irrelevant clicks.

Basically, the thriving trend of online activities requires the great ability to handle the multiple and increasing behavior information in CTR prediction. Recurrent Neural Networks (RNNs) is a specialized neural network with feedback connection and allows the neural network to remember the past data when processing the next output [20]. The framework of RNNs captures the time dependence of behavior by modeling the historical behavior sequence of users, has been demonstrated to have strong advantages in recognition and classification, and are suitable to be applied in prediction tasks of massive data [21], [22], [23]. Meanwhile, the dependence of RNNs on time series leads to limitations in the face of complex and changeable user behavior data. With well robustness, high efficiency and automation advantages, RNNs are very friendly to non-expert users and have become competitive forecasting methods as being widely used in Autonomous engineering, flow prediction, consumer understanding, machinery [21], [24], [25], [26], [27].

Capsule Network was first proposed to be used in the field of computer vision, has been developed into one of the most efficient approaches for analyzing vast amounts of complex data [28], [29], [30]. It constructs the part-whole relationships with the encoded feature capsules to capture the important spatial hierarchies among different features. The importance of different behaviors and the correlation with different behaviors can be captured with high accuracy by transmitting user behavior sequence information among different levels of capsules [31], [32]. The part-whole spatial hierarchies avoids the loss of the possible correlation information in the mass data, as well as the validity of which is also confirmed in other fields [33], [34].

From the perspective of capturing valuable information hidden in different types of clicks and seemingly irrelevant clicks data, an interactive attention-based capsule network is proposed in this study to analyze the complex and changeable sequential user behavior data for click-through rate prediction. As a variant of RNNs, Gated Recurrent Unit (GRU) shows well performance on low-complexity sequences for retaining time series information. In detail, low-level capsule networks in this paper were obtained from user behavior sequence related items through GRU to avoid the loss of time series information while organizing user click behavior information [35]. Particularly, each low-level capsule network contains single or more related item vectors, ensuring the preservation of potential association information to the greatest extent possible. The spatial association structure was established through dynamic routing among diverse low-level capsule networks, which subsequently generates high-level capsule networks for ultimate predictions. Meanwhile, based on the public datasets, the reported model was evaluated from different perspective including the comparison with other algorithms in terms of AUC and other metrics, and the influence analysis of parameters and capsule dimensions on the model's function. Experimental results have demonstrated that the reported model exhibits exceptional performance with high effectiveness.

The major contributions of this research are summarized as follows:

(1) We propose a novel Interactive Attention-based Capsules Network (IACaps) for CTR prediction. This novel model effectively captures the complex and changeable sequential user behavior data, providing a more accurate and reliable CTR prediction.

(2) We propose an Interactive Attention Dynamic Routing Mechanism, which preserves the potential association information among different browsing behaviors to the greatest extent possible. It helps to extract and understand the seemingly irrelevant information hidden in massive click data.

(3) Our model has been thoroughly evaluated using three different types of datasets from the Amazon Dataset. The results show that the proposed model outperforms other existing models in terms of AUC and other metrics, demonstrating its effectiveness and superiority in CTR prediction tasks under different daily scenarios.

(4) A detailed analysis of hyper-parameter optimization has been conducted, including the number of iterations, the number of capsules, and the dimension of capsules. This analysis provides a better understanding of the operating principles of the IACaps model and contributes to its optimization and application in various practical scenarios.

II. RELATED TECHNOLOGIES

A. STATE-OF-THE-ART CLICK-THROUGH RATE PREDICTION BASED ON DEEP LEARNING

In 2022, Yang et al., categorized deep learning models as one of state-of-the-art CTR prediction models reported in online advertising, demonstrating the effectiveness and advance of deep learning in CTR [14]. As a main part of deep learning models, convolutional neural network represents advantages in finding local feature interactions and reducing the number of parameters through the shared weights mechanism. Lian et al. focused on the granularity of input features and integrated it with the mechanism of convolutional neural network (CNN) to generate higher-order feature interactions in 2020, which reducing the difficulty of deep neural network learning by optimizing the process of feature selection, and providing a typical study that emphasizes the pretreatment of input data, instead of optimizing the algorithm only [36]. In 2022, Damaševičius and Zailskaitė-Jakštė reported a deep learning-based framework for CTR Prediction [37]. The reported model considered sorting the scattered online interaction information with 3U (The user, usage, and usability) principal constituents before AI treatment. Actually, 3U constituents reflected the background, process and the involved roles, summarized all the sparse process information throughout the event into three principal elements. To some extent, this consideration solved the problem of typically sparse input functions, reduced the dependence of traditional prediction methods to high -level functions, and indicated the importance of the association among different item attributes. Xiao et al. observed that one user usually has multiple interests at a time, which appears the real result of CTR prediction should be the superposition of different probabilities, and final behavior changes were decided by dominant interests [38]. For analyzing the impact of this phenomenon in CTR prediction, the Deep Multi-Interest Network (DMIN) was developed by Xiao et al. in 2020. The proposed method designed Behavior Refiner Layer and Multi-Interest Extractor Layer to study latent multiple interests of users by capturing historical item representations and extracting multiple user interests. Experiment results show that the multiple-layer-based DMIN outperforms various state-of-the-art baselines in terms of CTR prediction. In addition, Liu et al. put forward the ideal of feature generation by convolutional neural network (FGCNN) which relies on feature generation and deep classifier for CTR prediction [39]. In detail, new features were generated by recombining the local patterns obtained through CNN, which also emphasized the importance of association relationship among different items in low dimension.

The mentioned related works indicting that the deep learning is useful and attractive in CTR prediction, and associating or analyzing features in the view of low granularity and dimension is the trend of this task.

B. CAPSULE NETWORK RELATED WORKS FOR PREDICTION

Capsule network architectures have been widely used in prediction tasks due to the well accuracy, robustness and interpretability. Zhu et al. demonstrated a hierarchical capsule prediction network to study the effectiveness of capsule networks on marketing campaign prediction tasks in 2022. Since it is difficult to parse and evaluate the effect of a single marketing campaign, Chu et al. extended the capsule neural networks through the disentangled feature capsule, event capsule, event cluster capsule, and final outcome capsule to reason the hierarchical relationships between the part and the whole in marketing campaigns effect prediction [40]. As the number of bite-sized videos grows exponentially, Han et al. designed an aspect-level sentiment capsule network to predict micro-video CTR with significant outperformance in 2021. Specifically, this work proposed pairing the aspect-level features of target micro-video with those of positive and negative feedback respectively to identify their sentiments and form the sentiment capsules. This operation helps improving the accuracy from the finer granularity with the background of massive data, which is similar to the method studied in this paper [41]. Meanwhile, a novel model which is named DemiNet (Dependency-Aware Multi-Interest Network) was reported by Wang et al. in 2021. The study relied on the consideration of various dependency types between item nodes and performed dependency-aware heterogeneous attention, improving the overall recommendation performance over several state-of-the-art baselines. This novel model further demonstrates the efficacy and interpretability benefits of inducing finer granularity analysis [42]. In 2020, Li et al. proposed an attentive capsule network (ACN) for CTR and conversion rate predictions in online adverting, aiming at capturing the diversity of user‘s interests from user behavior history adaptively. This method uses Transformers to achieve the interaction between different features, and captures multiple interests from user behavior history through capsule networks. In order to obtain the precise sequence representation related to advertisements, Li et al. further designed a modified dynamic routing algorithm integrating with the attention mechanism [43]. However, attention weights gained only by the dense layer may limit the intuitive interpretation of internal operation of the model, as well as hinder its further development and narrow the available range. Capsule network conception has also been extensively studied in other predicting direction. Liu et al. explored the value of capsule network in link prediction in 2023, Afshar et al. researched the potential of capsule network in malignancy prediction in 2021, and Mousa et al. developed TI-Capsule model based on capsule network for stock exchange prediction [44], [45], [46]. All above researches demonstrated that capsule network conception has unique advantages in prediction tasks, especially in processing massive data.

III. THE PROPOSED METHOD

In order to face the further development of Internet economy and better analyze the relationship between user behavior and target product, we proposed an interactive attention-based capsule network for CTR by aggregating advantages of deep learning in analyzing massive data and the well capacity of capsule network in item classification.

The interactive attention-based capsule network mainly consists of four layers assisted with a training process: an embedding layer, a feature extraction layer, an interactive attention-based capsule layer and an output layer. The relationship among different layers is clarified in Figure 1. Generally, a click action refers to a series of logical events, thus four feature categories were chosen to cover the whole logical click process, which are User Profile, User Historical Behavior, Target Item and Item Historical Behavior. User Profile and Target Item represent the detailed cause and result of a complete click process. From the subjective and objective perspectives respectively, User Historical Behavior and Item Historical Behavior concluded how the final result was achieved step by step from the original reason. Considering that the information of Item Historical Behavior and User Profile is relatively objective and stable, multi-head attention mechanism is introduced to capture the compatibility between current users and potential user groups of target products.

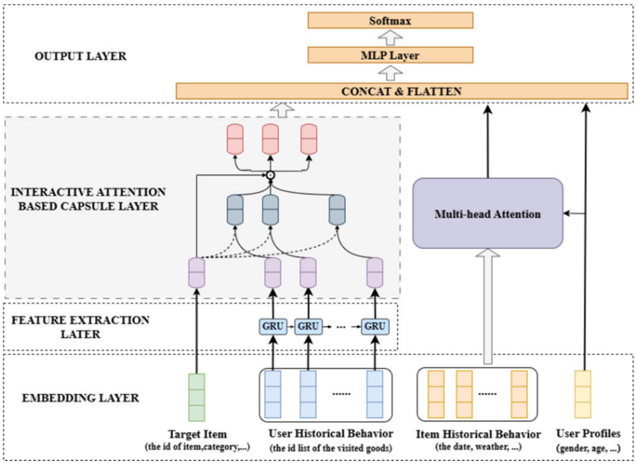

FIGURE 1. The architecture of the IACaps model.

A. EMBEDDING LAYER

There are several fields in each feature category to better cover functions of features in a logical click process. User Profile field contains gender, age and other user-related information. We removed the abnormal and invalid user information, and filled in the missing basic information before inputting the User Profile information into the embedding layer. User Historical Behavior field is actually the id list of goods the user visited, which reflects the interests and needs of users. All user behavior data were aggregated to form the sequence of User Historical Behavior. Target Item field includes kinds of id information, such as the item id, category id and any other id information that represents the target property. Meanwhile, the user's last click behavior is defined as a positive sample, and the randomly sampled item is defined as a negative sample in Target Item. Item Historical Behavior field aims at collecting objective information like the date, weather and so on. The function of embedding layer is transforming all data of four feature fields to embedding matrix with a usable format.

In detail, vectors with 18 dimensions were randomly initialized firstly, all the related ID, including the id of item, category in Target Item, and the id list of visited goods in User Historical Behavior, were mapped with these vectors in embedding layer subsequently to obtain the input of the feature extraction layer. This method can convert high-dimensional sparse data into low-dimensional dense vectors, reducing computational complexity as well as retaining the association information between IDs. In this paper, User Profile was transformed by $E_U \in \mathbb{R}^{N_u \times n_E}$, where N_u is the number of User Profile sparse features. While, the embedding matrix of User Historical Behavior was represented by $E_{hist} = \{x_1, x_2, \cdots, x_k\} \in \mathbb{R}^{n_E \times K}$, where K is the number of user historical behaviors and $x_t \in \mathbb{R}^{n_E}$ is the embedding of the t-th behavior with dimension n_E. Context and Target Item are similarly expressed as $E_C \in \mathbb{R}^{N_c \times n_E}$ and $E_I \in \mathbb{R}^{N_i \times n_E}$, where N_c and N_i are the number of Content and Target item sparse features, respectively. For the better comparison of the proposed method with others, the dimension n_E in this paper is set to 18, thus we randomly initialize a vector with 18 dimensions for each ID and update these vectors through gradient descent.

B. FEATURE EXTRACTION LAYER

In this paper, in order to capture the time dependency of user behavior and the better analysis of how past actions affect the future, GRU is introduced in feature extraction layer to encode the user's historical behavior sequence E_{hist} while retaining time series information of the historical behavior sequence. As a type of recurrent neural network, GRU enables to capture the temporal dependence in sequences and generate a comprehensive representation of user behaviors. Based on the current input x_t and the previous hidden state h_{t-1}, GRU updates the hidden state h_t at each time step t from the given historical behavior sequence $E_{hist} = \{x_1, x_2, \cdots, x_k\}$. The update and reset gates z_t and r_t are calculated as follows:

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right) \quad (1)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right) \quad (2)$$

Consequently, the candidate hidden state $\tilde{h}$ is then calculated as:

$$\tilde{h}_t = \tanh\left(W \cdot [r_t \odot h_t - 1, x_t]\right) \quad (3)$$

Finally, the hidden state h_t is updated as:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (4)$$

where σ is the sigmoid function, W_z, W_r, and W are weight matrices, $\odot$ denotes element-wise multiplication, and $[h_{t-1}, x_t]$ denotes the concatenation of h_{t-1} and x_t. The output of feature extraction layer serves as the representation of low-level capsules $H = \{h_1, h_2, \cdots, h_k\}$.

Referencing the previous work, different groups of clicks may be connected by some other key clicks for the clearer demand forecasting [42]. For instance, combined with the recent click history behavior of the kitchenware and the weather, a series of clicks on tin foil products actually increase the possibility of outdoor barbecue needs. The multi-head attention was adjusted and introduced in this paper to preserve the associated information for similar situation by representing potential multiple user groups for the target item.

$$MH^{(E_C)} = MULTI\_HEADATTENTION(E_C) \quad (5)$$

where E_{CU} was obtained through element-wise product and dot product interactions between the representation output of the Item Historical Behavior and the current User Profile.

C. INTERACTIVE ATTENTION-BASED CAPSULE LAYER

The interactive attention-based capsule layer is the core framework of the proposed model, which aggregates the representation of user behavior sequences. Then, representations of similar behaviors are aggregated into high-level capsules. Thus, each high-level capsule can represent one interest of the user. The hierarchical relationships between complex and changing user historical behaviors are captured on this layer with classification and aggregation. Furthermore, the introduction of Interactive Attention Dynamic Routing Mechanism in this layer enhances the ability of model to analyze the relationship between User Historical Behavior and Target Item.

Specifically, the transformation matrices were used to help low-level capsule i updating to high-level capsule j by generating the prediction vectors $\widehat{\mathrm{H}}_{\mathrm{j|i}}$ and $\widehat{\mathrm{E}}^{\mathrm{I}}_{\mathrm{j|i}}$ from low-level capsule i. Formally, corresponding transformations $\widehat{\mathrm{H}}_{\mathrm{j|i}}$ and $\widehat{\mathrm{E}}^{\mathrm{I}}_{\mathrm{j|i}}$ are computed as:

$$\widehat{\mathrm{H}}_{\mathrm{j|i}} = \mathrm{W}^{\mathrm{U}}_{\mathrm{j}}\mathrm{H}_{\mathrm{ij}} \tag{6}$$

$$\widehat{\mathrm{E}}^{\mathrm{I}}_{\mathrm{j|i}} = \mathrm{W}^{\mathrm{I}}_{\mathrm{j}}\mathrm{E}^{\mathrm{I}}_{\mathrm{ij}} \tag{7}$$

where W^{U} and W^{I} transformation matrices are sharing weights across the low-level capsules. The interactive attention dynamic routing is designed with nonlinear map following:

$$\left\{\widehat{\mathrm{H}}_{\mathrm{j|i}}, \widehat{\mathrm{E}}^{\mathrm{I}}_{\mathrm{j|i}}\right\}_{\mathrm{i}=1,\cdots,\mathrm{K},\mathrm{j}=1,\cdots,\mathrm{T}} \mapsto \{\mathrm{V}_{\mathrm{j}}\}_{\mathrm{j}=1}^{\mathrm{T}} \tag{8}$$

D. INTERACTIVE ATTENTION DYNAMIC ROUTING MECHANISM

The interactive attention dynamic routing is a key component of the capsule layer, as it is responsible for determining the coupling coefficients c_{ij} between the low-level and high-level capsules. The coupling coefficients are calculated based on the agreement between the prediction vectors $\widehat{\mathrm{H}}_{\mathrm{j|i}}$ and $\widehat{\mathrm{E}}^{\mathrm{I}}_{\mathrm{j|i}}$ of the low-level capsules and the output vectors of the high-level capsules.

The calculation process begins with initializing the coupling coefficients c according to Eq. (8) to associate User Historical Behavior with Target Item:

$$\mathrm{c}_{\mathrm{ij}} = \mathrm{softmax}\left(\widehat{\mathrm{H}}_{\mathrm{j|i}}, \widehat{\mathrm{E}}^{\mathrm{I}}_{\mathrm{j|i}}\right) \tag{9}$$

In addition, the mask method is applied to prevent low-level capsules which are meaningless (such as padding) from being updated to high-level capsules:

$$\mathrm{c}_{\mathrm{ij}} = \mathrm{MASK}\left(\mathrm{c}_{\mathrm{ij}}\right) \tag{10}$$

High-level capsule s_j is the weighted sum of all prediction vectors $\widehat{\mathrm{H}}_{\mathrm{j|i}}$ from low-level capsules:

$$\mathrm{s}_{\mathrm{j}} = \sum\nolimits_{\mathrm{i}} \mathrm{c}_{\mathrm{ij}}\widehat{\mathrm{H}}_{\mathrm{j|i}} \tag{11}$$

Following, the high-level capsules generated by user behavior, and item capsules are fused by interactive attention mechanism with equation:

$$\mathbf{s}_j = softmax(\widehat{\mathbf{E}}^I_{j|i}W^Q \cdot \left(\mathbf{s}_j W^K\right)^T) \cdot \mathbf{s}_j W^K \tag{12}$$

The output vector v_j of the high-level capsule is then computed as:

$$\mathrm{v}_j = \mathrm{s}_j W^O \tag{13}$$

And b_{ij} is updated by Eq. (13) after the routing was updated:

$$b_{ij} = b_{ij} + \widehat{\mathbf{H}}_{j|i} \cdot \mathbf{v}_j \tag{14}$$

In order to ensure that short vectors get shrunk to almost zero length and long vectors get shrunk to a length slightly below one, a non-linear "squash" function is introduced subsequently:

$$\mathbf{v}_j = \frac{\left\|\mathbf{v}_j\right\|^2}{1 + \left\|\mathbf{v}_j\right\|^2}\frac{\mathbf{v}_j}{\left\|\mathbf{v}_j\right\|} \tag{15}$$

The spatial association structure is established through dynamic routing among diverse low-level capsule networks to generates high-level capsule networks for ultimate predictions. The dynamic routing algorithm is shown in Algorithm 1.

Algorithm 1 Interactive Attention Dynamic Routing

1: INPUT: $\widehat{\mathrm{H}}_{\mathrm{j|i}}, \widehat{\mathrm{E}}^{\mathrm{I}}_{\mathrm{j|i}}$
2: OUTPUT: v_{j}
for all capsule i in layer l and capsule j in layer (l + 1):
$\mathrm{b}_{\mathrm{ij}} = \widehat{\mathrm{H}}_{\mathrm{j|i}} \cdot \widehat{\mathrm{E}}^{\mathrm{I}}_{\mathrm{j|i}}$
For r iterations do
for all capsule i in layer l: $\mathrm{c}_{\mathrm{i}} \leftarrow \mathrm{softmax}(\mathrm{b}_{\mathrm{i}})$
for all capsule i in layer l: $\mathrm{c}_{\mathrm{i}} \leftarrow \mathrm{MASK}\ (\mathrm{c}_{\mathrm{i}})$
for all capsule j in layer (l +1): $\mathrm{s}_{\mathrm{j}} \leftarrow \sum_{\mathrm{i}} \mathrm{c}_{\mathrm{ij}}\widehat{\mathrm{H}}_{\mathrm{j|i}}$
for all capsule j in layer (l +1): $\mathrm{s}_{\mathrm{j}} \leftarrow \mathrm{softmax}(\widehat{\mathrm{E}}^{\mathrm{I}}_{\mathrm{j|i}}\mathrm{W}^{\mathrm{Q}} \cdot (\mathrm{s}_{\mathrm{j}}\mathrm{W}^{\mathrm{K}})^{\mathrm{T}}) \cdot \mathrm{s}_{\mathrm{j}}\mathrm{W}^{\mathrm{K}}$
for all capsule j in layer (l +1): $\mathrm{v}_{\mathrm{j}} \leftarrow \mathrm{s}_{\mathrm{j}}\mathrm{W}^{\mathrm{O}}$
for all capsule i in layer l and capsule j in layer (l +1):
$\mathrm{b}_{\mathrm{ij}} = \mathrm{b}_{\mathrm{ij}} + \widehat{\mathrm{H}}_{\mathrm{j|i}} \cdot \mathrm{v}_{\mathrm{j}}$
end for
return squash $(\mathrm{v}_{\mathrm{j}})$

E. OUTPUT LAYER

The overall representation vector for CTR prediction is obtained by concatenating the output of the interactive attention-based capsule layer with User Embedding E_{U}, Target Item Embedding E_{I}, and Feature Extraction Embedding E_{CU}. Then two layers of multi-layer perceptron (MLP) with the Rectified Linear Unit (ReLU) activation function are used for the further learning of interaction. All final probability values for CTR prediction are generated in this layer after the modification by the Sigmoid function. The learning objective of this end-to-end trained network is calculating the weighted negative cross-entropy loss L:

$$\mathrm{L}(\mathrm{y}, \widehat{\mathrm{y}}) = -\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left[\mathrm{y}_{\mathrm{i}} \log(\widehat{\mathrm{y}}_{\mathrm{i}}) + (1 - \mathrm{y}_{\mathrm{i}}) \log(1 - \widehat{\mathrm{y}}_{\mathrm{i}})\right] \tag{16}$$

Specifically, N represents the total number of observations in datasets. $\widehat{y}$ are the predicted probability values calculated by the model, while y is the true label (0 or 1).

IV. EXPERIMENTS

In order to demonstrate the reliability and the universality of interactive attention-based capsule network in CTR prediction for complex and changing user behaviors, the proposed model is experimented and evaluated with Amazon open datasets obtained from real cases. Particularly, three different types of datasets as Beauty, Electronics and Sports (which represent three completely different life scenarios: daily life, entertainment, and outdoor respectively) are chosen to be tested for the further clarification of the reliability and the universality in different fields or categories. The performance of the model is compared with several baseline methods, and all experimental results are evaluated by AUC, F1 score, Accuracy, and LogLoss metrics to realize the comprehensive analysis of the proposed method

A. DATASET

The Amazon Dataset, which is composed of product reviews and metadata from Amazon, reflects the sufficient random and various, but real cases [48]. Three datasets obtained from Amazon Dataset: Beauty, Electronics and Sports, representing different scenarios and containing more than 250,000 users, 93,000 items, 2,000 categories and 1,066,000 samples, are used to verify the effectiveness of the proposed model in multiple interests scenarios, including daily, entertainment and outdoor fields. Furthermore, considering the brand numbers and selection habits, the type of the selected datasets have more categories and is updated more quickly when compared with other categories, which ensures the effectiveness and reliability of the proposed method verification process and conclusions. The datasets are randomly divided into three parts firstly as: training set (80%), test set (10%), and validation set (10%) for adjusting hyperparameters. The statistical information of the used data obtained from Amazon Dataset are shown in Table 1. All there datasets were randomly initialized to a vector with 18 dimensions and updated through gradient descent after classification, avoiding the loss of relevant information caused by data confusion.

TABLE 1. Statistics of the datasets.

DATASET	#USER	#ITEM	#CATEGORY	#SAMPLE
BEAUTY	22,363	12,101	221	352,278
ELECTRONICS	192,403	63,001	801	192,403
SPORTS	35,598	18,357	1,073	521,478

B. BASELINE MODELS

In order to better analyze the mechanism of the proposed model, ten typical and similar baseline models are selected to be compared with the proposed model from different perspectives, including:

- **SVD++** is a combined model that integrates both neighborhood and latent factor information.
- **Embedding & MLP** is the basic deep learning model for CTR prediction. It takes sum pooling operation to integrate behavior embeddings.
- **Wide & Deep** is composed of a deep model of id features and a linear model of wide features.
- **PNN** uses an embedding layer to learn distributed representations of categorical data, a product layer to capture interactive patterns between categories, and fully connected layers to explore high-order feature interactions.
- **DIN** introduces a local activation unit to adaptively learn the representation of user interests from historical behaviors based on candidate ads.
- **DIEN** uses interest extraction layer and interest evolution layer to capture user's dynamic interests.
- **DIEN_neg** is an extended version of DIEN. DIEN_neg denotes the DIEN method with negative sampling for an auxiliary loss.
- **DMIN** uses multi-head self-attention to refine the representation of user behavior sequence, and extract multiple user interests.
- **DUMN** uses an attention mechanism to measure item-to-item relevance and learn a unified user representation, and uses a relevance unit to measures user-to-user relevance based on similarities.
- **DRINK** introduces a transformer-based multi-representational item network that dynamically captures the characteristics of the item and decouples time information from item behavior.

C. EVALUATION METRICS

Metrics such as AUC, F1 score, Accuracy as well as LogLoss, are introduced in this section for model evaluation. AUC measures the area under the ROC curve to compare the quality of different models. As a classical index, F1 score is the harmonic mean of precision and recall, which feeds back the effectiveness of the tested model. Accuracy formula calculates the ratio of correct samples to the total, demonstrating the accuracy of the model. LogLoss is applied to assess the reliability of models. In addition, the lower value of LogLoss, the more successful of the corresponding model. Measurements in Evaluation Metrics section verified the comprehensive performance of IACaps by testing capabilities from all-round aspects with pertinent formulas or methods.

D. EXPERIMENTAL SETTINGS

All experiments in this paper were implemented on TensorFlow platform and trained on GPUs (GTX1070) with Adam optimizer. Moreover, all tests were repeated five times and their average values were reported as test results to avoid experimental randomness. Hyperparameters used during testing processes are tuned according to validation datasets, where learning rate is set to 0.001, batch size is adjusted to be 128 and the embedding size of features is modified to be 18. Since discussions of sequential models in this paper are mainly based on user and item behaviors, the hidden units of the multi-layer perceptron layer are fixed with [200, 80] and the truncation length of user behaviors is set as 20 for all models. For the proposed IACaps model, the related parameters are set appropriately as the hidden state of GRU to 36, number of capsules to 16, dimension of capsule to 32, routing to 3, the number of interactive attentions to 8. Other parameters consistent with that of baselines, ensuring the rationality of comparison.

V. RESULTS AND DISCUSSION

In this section, common metrics as AUC, F1, Acc. and LogLoss are introduced to demonstrate the superior performance of the proposed method. Furthermore, The main experiments and ablation studies are carried out to teat the reasonableness and effectiveness of the model. Optimization of Hyper-parameters is also analyzed from iterations and capsules aspects to help readers better understand the mechanism of the framework.

A. MAIN RESULTS AND DISCUSSION

Results of each method for different daily scenarios under its optimal hyper-parameters settings are shown in the below tables.

TABLE 2. Results of different methods based on Amazon Beauty dataset.

Amazon Beauty				
Methods	**AUC**	**F1**	**Acc.**	**LogLoss**
SVD+	0.5818	0.5039	0.5675	0.9946
Embedding & MLP	0.8076	0.7413	0.7273	0.5447
Wide & Deep	0.8115	0.7415	0.7278	0.5354
PNN	0.8095	0.7321	0.7282	0.5402
DIN	0.8233	0.7474	0.7411	0.5219
DIEN	0.8205	0.7443	0.7389	0.5251
DIEN_neg	0.8630	0.7890	0.7743	0.4625
DMIN	0.8799	0.7853	0.7882	0.4354
DUMN	0.8565	0.7749	0.7654	0.4854
DRINK	0.9079	0.8100	0.8151	0.3878
IACaps	0.9083	0.8203	0.8180	0.3813

Table 2 to 4 show the results of different methods for CTR task in daily domains as Beauty, Electronics and Sports. The metrics listed in tables including AUC, F1, Acc. and LogLoss are common metrics in model checking process. It is clear that the proposed IACaps has shown obvious advantages in terms of AUC, F1, Acc. and LogLoss metrics, demonstrating the superiority of IACaps and the effectiveness of the interactive attention-based capsule mechanism.

Specifically, except SVD+, all the mentioned methods are typical models constituted by Neural Network (NN), the significant improved performance of which compared with SVD++ indicates the necessity of NN. Embedding & MLP and Wide & Deep are classical NN-based discrete models but limited to dealing with discrete information. Comparing the average values of metrics of Embedding & MLP,

TABLE 3. Results of different methods based on Amazon Electronics dataset.

Amazon Electronics				
Methods	**AUC**	**F1**	**Acc.**	**LogLoss**
SVD+	0.6854	0.6403	0.6348	0.6914
Embedding & MLP	0.7528	0.7002	0.6844	0.5925
Wide & Deep	0.7655	0.7046	0.6942	0.5786
PNN	0.8552	0.7949	0.7874	0.5907
DIN	0.8893	0.8075	0.8066	0.3756
DIEN	0.8673	0.8078	0.7957	0.3789
DIEN_neg	0.9682	0.9222	0.9212	0.2148
DMIN	0.9667	0.9220	0.9209	0.2191
DUMN	0.9057	0.8849	0.8747	0.3383
DRINK	0.9636	0.9080	0.9076	0.2373
IACaps	0.9698	0.9292	0.9268	0.1992

TABLE 4. Results of different methods based on Amazon Sports dataset.

Amazon Sports				
Methods	**AUC**	**F1**	**Acc.**	**LogLoss**
SVD+	0.5818	0.5040	0.5675	0.9946
Embedding & MLP	0.7962	0.7300	0.7193	0.5563
Wide & Deep	0.8124	0.7357	0.7283	0.5340
PNN	0.8103	0.7276	0.7289	0.5426
DIN	0.8176	0.7415	0.7358	0.5281
DIEN	0.8164	0.7407	0.7356	0.5277
DIEN_neg	0.8586	0.7891	0.7713	0.4720
DMIN	0.8667	0.7758	0.7770	0.4581
DUMN	0.8564	0.7724	0.7651	0.4798
DRINK	0.9095	0.8102	0.8155	0.3842
IACaps	0.9091	0.8290	0.8225	0.3740

Wide & Deep and SVD+ methods with the which of the NN-based sequential models including PNN, DIN and other methods listed in tables, the average indicator values of AUC, F1 and Acc. are increased by 17.0%, 17.1%, 14.4% for Beauty domain, by 25.7%, 27.9%, 29.3% for Electronics domain, by 17.2%, 17.8%, 14.5% for Sports domain respectively. Meanwhile, the average values of LogLoss indicator decreased by 32.4%, 48.6%, 32.3% respectively. All these improvements demonstrated the reasonableness and effectiveness of NN model assisted by sequential information in CTR.

DIN and DIEN models are considered as the representation of single-interest user behavior, neglecting the multi-interest representation in user behavior sequences. In contrast, the proposed IACaps incorporates the feature aggregation characteristics of capsule networks, which automatically aggregates user behavior sequences to generate multi-interest representation. According to experiment results, IACaps increased the AUC by 5.2%, 0.2%, 5.9% in the analysis of Beauty, Electronic and Sports datasets when compared the proposed model with DIEN_neg model function. IACaps achieves the superior performance among all three daily datasets, indicating the introduced capsule units enable the further successful trace of user's latent multiple interests as well as the better modeling efficacy through the feature aggregation of capsule networks.

Moreover, DMIN is also a multi-interest representation model based on multi-head attention mechanism. Different from DMIN, IACaps model introduces an attention-based interactive capsule layer to construct a cross-attention mechanism between User Historical Behavior and the Target Item. It appears that IACaps obtains the superior performance to DMIN with 3.2%, 0.3%, 4.9% increase of AUC in terms of analyzing Beauty, Electronics and Sports datasets. Attributing to the cross-attention mechanism of IACaps, the proposed model effectively captured and facilitated the better representation of user multi-interests.

Besides, in this paper, the accuracy and reliability clarifications of IACaps are also proved by repetitive global training iterations. Fig. 2 visualizes the validation set performance of DRINK and IACaps based on the Amazon Electronics dataset. In terms of both AUC and LogLoss metrics, when the number of iterations exceeds 8000 times, IACaps consistently show significant superiorities to the which of DRINK. This persistent enhancement in performance further corroborates the efficacy of IACaps in effectively capturing the dynamics of sequential user behavior, improving the accuracy and reliability of CTR predictions.

B. ABLATION STUDIES

In order to prove the necessity and advantages of the combination between the proposed interactive attention and capsule mechanism, we designed two variants of IACaps in ablation studies, including IACaps_w_Caps and IACaps_w_Att. IACaps_w_Caps is designed to replace the proposed Interactive Attention-based Capsule layer with the original capsule layer, while IACaps_w_Att is designed to replace the Interactive Attention-based Capsule layer with an attention layer.

As shown in Table 5, the lack of interactive attention or capsule layer leads to the weakening of model performance as the generally value reduction of AUC, F1, Acc. metrics and the value increase of LogLoss indicator. Attributing to the cooperation between interactive attention mechanism and the introduced capsule units, the proposed Interactive attention-based capsule network has a stronger feature aggregation capability to aggregate inferior features.

C. OPTIMIZATION OF HYPER-PARAMETERS

The influence of hyperparameters on the Interactive Attention-based Capsule layer were also analyzed to further study the optimal configuration of the model. The exploration was mainly carried out from the three aspects, including the number of iterations r in Algorithm 1, the number of v_j capsules, and the dimension of the v_j capsules.

The performances of IACaps model with different numbers of routing iterations are shown in Fig.3 It is illustrated that when the routing algorithm is set as 3, values of AUC and LogLoss metrics reached maximum and minimum

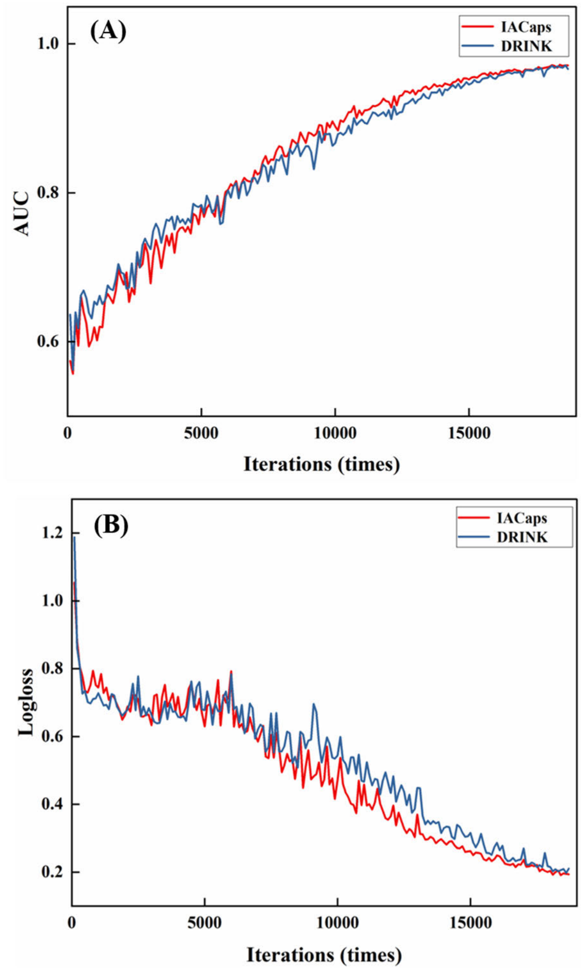

FIGURE 2. The value of AUC and LogLoss changing with the increase of iterations number.

TABLE 5. Results of different methods based on Amazon Sports dataset.

	Amazon Beauty			
Methods	**AUC**	**F1**	**Acc.**	**LogLoss**
IACaps	0.9698	0.9292	0.9268	0.1992
w/capsule	0.9674	0.9205	0.9198	0.2187
w/attention	0.9670	0.9184	0.9182	0.2182

respectively. In fact, the overfit generated by the excessive routing iterations might be the reason for this phenomenon.

For the better understanding of the affects caused by the capsule number, the different results were compared in the capsules number range of 8 to 20. As is shown in Fig. 4, the AUC indicator reaches the highest point with the lowest value of LogLoss indicator when the number of capsules is around 16. The phenomenon proved that the small number of capsules leads to the loss of effective information, while the large number of capsules leads to the accumulation of invalid information.

The dimension of capsule is the length of the output vector of a capsule as well as a critical parameter. High-dimension features enhance the express ability of multiple interests, but

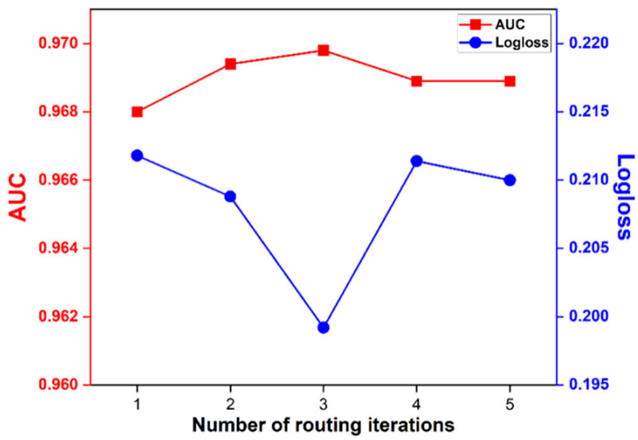

FIGURE 3. **The value of AUC and LogLoss changing with the increase of routing iterations number.**

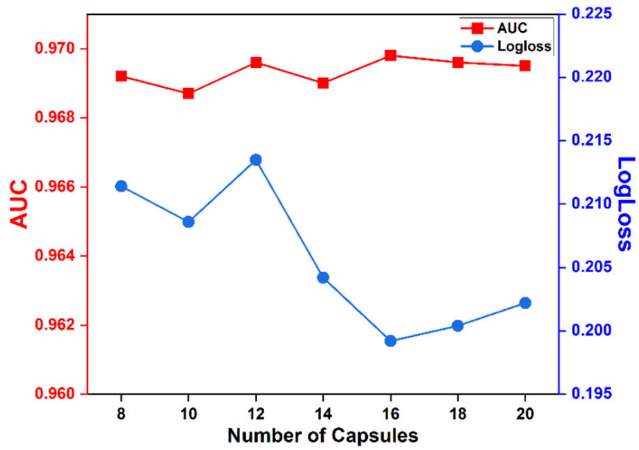

FIGURE 4. **The value of AUC and LogLoss changing with the increase of capsule number.**

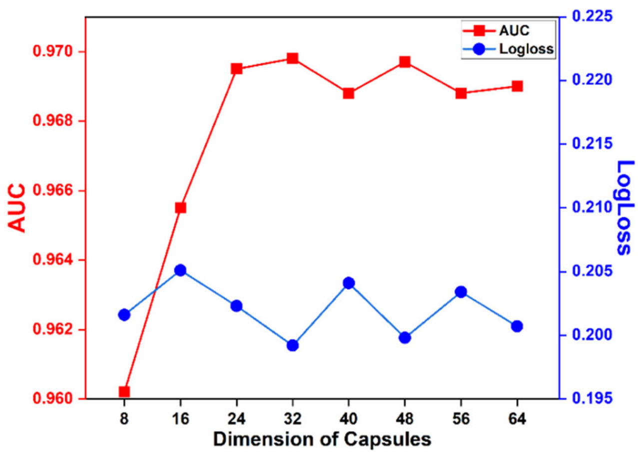

FIGURE 5. **The value of AUC and LogLoss changing with the increase of capsule dimension.**

may also produce potential noise and increase the complexity of the model. Hence, the dimensions of capsule were set to be 8, 16, 24, 32, 40, 48, 56, 64 for the further research. As shown in Fig. 5, the model appears the best performance when the size is 32. The large capsule dimension increased the amount of calculation and the time for training and reasoning, resulting in the risk of overfitting and performance degradation.

VI. CONCLUSION

In order to seize the opportunities that hidden in the thriving development of online economy, it is vital and effective to find a way that clearly predicts the real needs of customers, and accurately pushes product information and reduces company cost consequently. There is no doubt that any complete online transaction process must be realized through the click behavior on web pages, during the process of which, a large number of complex and changeable click behavior data that can reflect the real needs of users at the same time were generated. This study proposes an interactive attention-based capsule network for CTR prediction to capture the real demands of users from complex and changing click information.

Considering that effective information is hidden not only within each click or a same series clicks, but also among different types of clicks and even seemingly irrelevant clicks, the reported interactive attention-based capsule network improves the prediction accuracy by analyzing content-related groups to retain as much valid information as possible. In detail, after the preliminary correlation processing of discrete click information, the high-level capsules were obtained in interactive attention-based capsules layer through interactive attention dynamic routing mechanism which aggregates low-level capsules generated from feature extraction layer. During the level-up process of capsules in four layers framework, the easily ignored information hidden in massive click data and among different click groups were preserved by the cooperation between capsule unit and interactive attention dynamic routing mechanism.

In order to ensure the reliability and versatility of the model, three different types of data sets, including Beauty, Electronics, and Sports to represent different scenarios, were selected from Amazon Dataset to achieve comprehensive comparison experiments with ten baseline models in different aspects. The significant improvement of indicator values proves the effectiveness of the Interactive attention-based capsule network for CTR. Furthermore, the necessity of NN model assisted by sequential information, the effectiveness of the introduced low-level and high-level capsule units, as well as the importance of the cross-attention mechanism of IACaps, are verified by comparing and analyzing the results of different methods for CTR task in different daily scenarios as beauty, electronics and sports. Moreover, ablation studies indicate the contribution generated by the cooperation between interactive attention mechanism and the introduced capsule units. In order to better understand the principles of the model, the number of routing iterations, the number of capsules, and the dimension of capsules are limited to ranges of 1 to 5, 8 to 20, and 8 to 64 respectively while fixing other hyper-parameters on datasets, and study the changes in AUC and LogLoss metrics in condition that parameters are adjusted. The best model performance is obtained when the number of routing iterations is set to 3, the number of capsules is set to 16, and the dimension of capsules is set to 32.

In summary, benefiting from the introduction of capsules and interactive attention mechanism assistance, the interactive attention-based capsule network with sequential user behavior representing well performance in CTR predictions. This novel framework of the proposed model retaining the original information of each click event to take the hidden valuable information among different content click groups into considerations. All experimental results indicated that the proposed method is suitable for different daily scenes with high accuracy and stability, representing its huge potential in CTR prediction field.

REFERENCES

[1] N. Chen, J. Zhou, H. Guan, N. Xu, S. Xue, S. Li, and Z. Liu, "5G charging mechanism based on dynamic step size," *IEEE Access*, vol. 11, pp. 15069–15081, 2023.

[2] X. Wang, "A survey of online advertising click-through rate prediction models," in *Proc. IEEE Int. Conf. Inf. Technol., Big Data Artif. Intell. (ICIBA)*, vol. 1, Nov. 2020, pp. 516–521.

[3] H. Lee and C.-H. Cho, "Digital advertising: Present and future prospects," *Int. J. Advertising*, vol. 39, no. 3, pp. 332–341, Apr. 2020.

[4] B. Aslam and H. Karjaluoto, "Digital advertising around paid spaces, E-advertising industry's revenue engine: A review and research agenda," *Telematics Informat.*, vol. 34, no. 8, pp. 1650–1662, Dec. 2017.

[5] G. Chen, J. H. Cox, A. S. Uluagac, and J. A. Copeland, "In-depth survey of digital advertising technologies," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2124–2148, 3rd Quart., 2016.

[6] H. Choi, C. F. Mela, S. R. Balseiro, and A. Leary, "Online display advertising markets: A literature review and future directions," *Inf. Syst. Res.*, vol. 31, no. 2, pp. 556–575, Jun. 2020.

[7] G. Zhou, X. Zhu, C. Song, Y. Fan, H. Zhu, X. Ma, Y. Yan, J. Jin, H. Li, and K. Gai, "Deep interest network for click-through rate prediction," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 1059–1068.

[8] H. B. Mcmahan, G. Holt, D. Sculley, M. Young, D. Ebner, J. Grady, L. Nie, T. Phillips, E. Davydov, D. Golovin, S. Chikkerur, D. Liu, M. Wattenberg, A. M. Hrafnkelsson, T. Boulos, and J. Kubica, "Ad click prediction: A view from the trenches," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2013, pp. 1222–1230.

[9] Y. Kim, M. Kang, S. M. Choi, and Y. Sung, "To click or not to click? Investigating antecedents of advertisement clicking on Facebook," *Social Behav. Personality, Int. J.*, vol. 44, no. 4, pp. 657–667, May 2016.

[10] J. Gwizdka and D. Bilal, "Analysis of children's queries and click behavior on ranked results and their thought processes in Google search," in *Proc. Conf. Conf. Human Inf. Interact. Retr.*, Mar. 2017, pp. 377–380.

[11] S. Fosso Wamba, S. Akter, L. Trinchera, and M. De Bourmont, "Turning information quality into firm performance in the big data economy," *Manage. Decis.*, vol. 57, no. 8, pp. 1756–1783, Sep. 2019.

[12] D. Sen, M. Ozturk, and O. Vayvay, "An overview of big data for growth in SMEs," *Proc. Social Behav. Sci.*, vol. 235, pp. 159–167, Nov. 2016.

[13] S. Zillner, J. A. Gomez, A. G. Robles, T. Hahn, L. Le Bars, M. Petkovic, and E. Curry, "Data Economy 2.0: From big data value to AI value and a European data space," in *The Elements of Big Data Value: Foundations of the Research and Innovation Ecosystem*, vol. 2021. Cham, Switzerland: Springer, 2021, pp. 379–399.

[14] Y. Yang and P. Zhai, "Click-through rate prediction in online advertising: A literature review," *Inf. Process. Manage.*, vol. 59, no. 2, Mar. 2022, Art. no. 102853.

[15] A. Namin, M. L. Hamilton, and A. J. Rohm, "Impact of message design on banner advertising involvement and effectiveness: An empirical investigation," *J. Marketing Commun.*, vol. 26, no. 2, pp. 115–129, Feb. 2020.

[16] J. Chen, B. Sun, H. Li, H. Lu, and X.-S. Hua, "Deep CTR prediction in display advertising," in *Proc. 24th ACM Int. Conf. Multimedia*, Oct. 2016, pp. 811–820.

[17] Z. Gharibshah and X. Zhu, "User response prediction in online advertising," *ACM Comput. Surveys*, vol. 54, no. 3, pp. 1–43, Apr. 2022.

[18] F. Bienen, D. Kopljar, S. Geiger, N. Wagner, and K. A. Friedrich, "Investigation of CO_2 electrolysis on tin foil by electrochemical impedance spectroscopy," *ACS Sustain. Chem. Eng.*, vol. 8, no. 13, pp. 5192–5199, Apr. 2020.

[19] K. Eslamloo, V. Morshedi, M. Azodi, G. Ashouri, M. Ali, and F. Iqbal, "Effects of starvation and re-feeding on growth performance, feed utilization and body composition of Tinfoil barb (Barbonymus schwanenfeldii)," *World J. Fish Mar. Sci.*, vol. 4, no. 5, pp. 489–495, 2012.

[20] S. Das, A. Tariq, T. Santos, S. S. Kantareddy, and I. Banerjee, "Recurrent neural networks (RNNs): Architectures, training tricks, and introduction to influential research," in *Machine Learning for Brain Disorders*. Cham, Switzerland: Springer, 2023, pp. 117–138.

[21] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent neural networks for time series forecasting: Current status and future directions," *Int. J. Forecasting*, vol. 37, no. 1, pp. 388–427, Jan. 2021.

[22] W. Xiang, P. Musau, A. A. Wild, D. Manzanas Lopez, N. Hamilton, X. Yang, J. Rosenfeld, and T. T. Johnson, "Verification for machine learning, autonomy, and neural networks survey," 2018, *arXiv:1810.01989*.

[23] A. Alzahrani, P. Shamsi, M. Ferdowsi, and C. Dagli, "Solar irradiance forecasting using deep recurrent neural networks," in *Proc. IEEE 6th Int. Conf. Renew. Energy Res. Appl. (ICRERA)*, Nov. 2017, pp. 988–994.

[24] M. Mercangöz, A. Cortinovis, and S. Schönborn, "Autonomous process model identification using recurrent neural networks and hyperparameter optimization," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 11614–11619, 2020.

[25] L. Mudassar and Y. Byun, "Customer prediction on parking logs using recurrent neural network," in *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*. Cham, Switzerland: Springer, 2019.

[26] Y. Zhou, S. Mishra, J. Gligorijevic, T. Bhatia, and N. Bhamidipati, "Understanding consumer journey using attention based recurrent neural networks," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 3102–3111.

[27] A. Onan, "Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 5, pp. 2098–2117, May 2022.

[28] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–20.

[29] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.

[30] E. Xi, S. Bing, and Y. Jin, "Capsule network performance on complex data," 2017, *arXiv:1712.03480*.

[31] S. Yang, F. Lee, R. Miao, J. Cai, L. Chen, W. Yao, K. Kotani, and Q. Chen, "RS-CapsNet: An advanced capsule network," *IEEE Access*, vol. 8, pp. 85007–85018, 2020.

[32] M. K. Patrick, A. F. Adekoya, and A. A. Mighty, "Capsule networks—A survey," *J. King Saud Univ.-Comput. Inf. Sci.*, 2022, vol. 34, no. 1, pp. 1295–1310.

[33] Y. Wang, A. Sun, J. Han, Y. Liu, and X. Zhu, "Sentiment analysis by capsules," in *Proc. World Wide Web Conf. World Wide Web*, 2018, pp. 1165–1174.

[34] C. He, L. Peng, Y. Le, and J. He, "SECaps: A sequence enhanced capsule model for charge prediction," in *Proc. Int. Conf. Artif. Neural Netw.*, 2019, pp. 227–239.

[35] R. Cahuantzi, X. Chen, and S. Guttel, "A comparison of LSTM and GRU networks for learning symbolic sequences," in *Proc. Sci. Inf. Conf.*, 2023, pp. 771–785.

[36] Z. Lian and H. Ge, "FINET: Fine-grained feature interaction network for click-through rate prediction," in *Proc. 12th Int. Conf. Adv. Comput. Intell. (ICACI)*, Aug. 2020, pp. 334–339.

[37] R. Damasevicius and L. Zailskaite-Jakste, "Usability and security testing of online links: A framework for click-through rate prediction using deep learning," *Electronics*, vol. 11, no. 3, p. 400, Jan. 2022.

[38] Z. Xiao, L. Yang, W. Jiang, Y. Wei, Y. Hu, and H. Wang, "Deep multi-interest network for click-through rate prediction," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 2265–2268.

[39] B. Liu, R. Tang, Y. Chen, J. Yu, H. Guo, and Y. Zhang, "Feature generation by convolutional neural network for click-through rate prediction," in *Proc. World Wide Web Conf.*, May 2019, pp. 1119–1129.

[40] Z. Chu, H. Ding, G. Zeng, Y. Huang, T. Yan, Y. Kang, and S. Li, "Hierarchical capsule prediction network for marketing campaigns effect," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2022, pp. 3043–3051.

[41] Y. Han, P. Gu, W. Gao, G. Xu, and J. Wu, "Aspect-level sentiment capsule network for micro-video click-through rate prediction," *World Wide Web*, vol. 24, no. 4, pp. 1045–1064, Jul. 2021.

[42] Y. Wang, Q. Luo, Y. Ding, Y. Li, D. Wang, and H. Deng, "DemiNet: Dependency-aware multi-interest network with self-supervised graph learning for click-through rate prediction," 2021, *arXiv:2109.12512*.

[43] D. Li, B. Hu, Q. Chen, X. Wang, Q. Qi, L. Wang, and H. Liu, "Attentive capsule network for click-through rate and conversion rate prediction in online advertising," *Knowledge-Based Syst.*, vol. 211, Jan. 2021, Art. no. 106522.

[44] X. Liu, X. Li, G. Fiumara, and P. De Meo, "Link prediction approach combined graph neural network with capsule network," *Expert Syst. Appl.*, vol. 212, Feb. 2023, Art. no. 118737.

[45] P. Afshar, F. Naderkhani, A. Oikonomou, M. J. Rafiee, A. Mohammadi, and K. N. Plataniotis, "MIXCAPS: A capsule network-based mixture of experts for lung nodule malignancy prediction," *Pattern Recognit.*, vol. 116, Aug. 2021, Art. no. 107942.

[46] R. Mousa, S. Nazari, A. K. Abadi, R. Shoukhcheshm, M. N. Pirzadeh, and L. Safari, "TI-capsule: Capsule network for stock exchange prediction," 2021, *arXiv:2102.07718*.

[47] J. Zhang, F. Lin, C. Yang, and W. Wang, "Deep multi-representational item network for CTR prediction," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2022, pp. 2277–2281.

[48] J. McAuley, C. Targett, Q. Shi, and A. van den Hengel, "Image-based recommendations on styles and substitutes," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2015, pp. 43–52.

SHENG XUE received the B.S. degree in communication engineering from The North University of China, Taiyuan, China, in 2017, the M.S. degree in optical engineering from South China Normal University, Guangzhou, China, in 2020. He is currently an engineer in the Research Institute of China Telecom Co., Ltd... Guangzhou, China. He is the author of eight international journal papers. His research interests include analysis of cloud network operation business, optical sensors.

CONGQING HE received the B.S. degree from the School of Computer Science and Technology, Hefei Normal University, Hefei, China, in 2016, and the M.S. degree from the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China, in 2019. He is currently pursuing the Ph.D. degree with the School of Computer Science, Universiti Sains Malaysia, Penang, Malaysia. His research interests include natural language processing and legal intelligence.

ZHUXUAN HUA graduated from South China University of Technology in 2000 with a bachelor's degree in automation and a master's degree in computer technology in 2009.He is currently working in China Telecom Research Institute, mainly engaged in telecommunications business operations, telecommunications system research and development.

SONGTIAN LI received the B.S. degrees in computer science from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2017. He is working in China Telecom Research Institute, mainly engaged in telecommunications business operation, research and R&D of telecommunication systems.

GUANGWEI WANG received the M.Eng degree in Electronics and Communication Engineering form Beijing University of Posts and Telecommunications, Beijing, China, in 2022, His current research interests include Business & Operation Support System and Software Define Network.

LIWEN CAO was born in Jinhua, Zhejiang, P.R. China, in 1993. She received the Master degree from City University of HongKong. Now, she works in Department of Economics and Information Management, Shanghai University of Finance and Economics Zhejiang College, Her research interests include business data analysis and MIS.