User Profiling for Personalized Service Recommendation with Dual High-order Feature Learning

Guobing Zou[1], Liangrui Wu[1], Shengxiang Hu[1*], Song Yang[1], Chenyang Zhou[1], Yanglan Gan[2], Bofeng Zhang[3*]
[1] School of Computer Engineering and Science, Shanghai University, Shanghai, China
[2] School of Computer Science and Technology, Donghua University, Shanghai, China
[3] School of Computer and Information Engineering, Shanghai Polytechnic University, Shanghai, China
{gbzou, lr_w, shengxianghu, yangsong, chenyangzhou}@shu.edu.cn, ylgan@dhu.edu.cn, bfzhang@sspu.edu.cn

***Abstract*—With the surge in Web service users, user profiling has become increasingly prominent in personalized service recommender system. Graph Neural Networks (GNNs) has emerged as a key technology for user feature extraction. However, these methods mostly focus on modeling pairwise interaction relationships by type and overlook the high-order interaction relationships and deep semantic correlations. Moreover, GNNs' limited receptive fields restrict their capacity to capture user high-order features effectively. To address these issues, we propose a novel framework for advanced user profiling named Heterogeneous Interaction Graph Transformer (HIGT). Firstly, HIGT constructs a weighted heterogeneous interaction graph from historical user-service interactions, using edge types for interaction modes and weights for their frequency. Secondly, it uses a Transformer to extract high-order semantic attribute correlations and enhance global understanding through self-attention, while proposing a structure-enhanced attention mechanism to incorporate the graph structure into the Transformer architecture for extracting high-order interaction features of users. This dual high-order feature learning method provides deeper insight into users' preferences for Web services. Extensive experiments on two real-world e-commerce service datasets reveal that HIGT brings a significant performance boost compared with competing models for user profiling.**

***Index Terms*—Service Recommendation, Service User Profiling, High-order Feature, Graph Transformer, User-Service Heterogeneous Interaction Graph**

I. Introduction

With the rapid evolution of information technology and network infrastructure, Web services now play a pivotal role in attracting global users and accumulating large amounts of data. Personalized service recommendation techniques, which efficiently extract pivotal information from massive user data, optimize service resource allocation, and enhance user experience, have found widespread application. However, as the volume and complexity of data generated by various services increase, effectively conducting user analysis and accurately identifying user preferences have become critical prerequisites and challenges for realizing personalized service recommendation.

* Corresponding author.

User profiling is commonly employed as a crucial element for personalized recommendation of Web services on e-commerce and social media [1]–[4], describing and categorizing users from extensive data. In general, primary data sources used for user profiling include user content data, which covers what users create; behavioral data, showing how users interact with the services; and relational data, mapping out users' connections. The intricate nature of these data makes model development challenging. To tackle this, many efforts focus on developing models that can efficiently handle these complex data types while keeping their design simple.

User profiling models typically predict user attributes such as age [5]–[10], gender [6]–[10], and geolocation [11], [12], commonly viewed as classification tasks. Existing research can be broadly divided into single-source methods and fusion-based methods. Traditional research paradigms [13]–[27] often narrow their focus only to a single data source, leading to limited user representations. Recent shifts towards fusion methods integrate multiple data sources, overcoming this limitation. However, hand-crafted-based fusion strategies [28]–[31] face scalability challenges. Bridging this gap, automated fusion models using deep learning, particularly deep graph learning, offer a solution by modeling complex interrelations in a graph structure, where users and services are nodes, and their interactions are edges. Subsequently, employing Graph Neural Networks (GNNs) for feature propagation, researchers predict user attributes with high accuracy. Such deep graph learning models [7]–[9], [12], [32] have shown superior performance in user profiling tasks compared to traditional methods, highlighting their potential in this field.

Despite the effectiveness of graph-based user profiling methods, we argue that three critical issues have not yet been fully considered. Firstly, from a modeling perspective, current methods tend to focus on the direct interactions between entities, often neglecting the more complex high-order interactions. These high-order interactions involve multi-hop associations between user-to-user or user-to-service. Research in network theory and complex systems underscores the importance of high-order relationships for a deep understanding of complex

2836-3868/24/$31.00 ©2024 IEEE
DOI 10.1109/ICWS62655.2024.00049

networks [33], [34]. At the same time, these methods also ignore the high-order semantic correlations of entities, thereby limiting a comprehensive understanding of user behavior patterns. Secondly, from a technical perspective, GNNs are commonly used as the technique to extract user features along interaction graphs due to their effectiveness in modeling entities and their complex relationships. However, they struggle to capture high-order user features due to their limited receptive fields. Additionally, these approaches typically consider only the type of interaction, neglecting the significance of interaction frequency, which can directly indicate a user's engagement with specific services, thus serving as a more precise indicator for understanding user preferences.

Towards the above limitations, we introduce a novel framework for user profiling, named Heterogeneous Interaction Graph Transformer (HIGT), which adaptively extract dual high-order user features. Specifically, differing from traditional GNNs-based methods, HIGT utilizes a self-attention mechanism in the Transformer architecture to learn global semantic attribute correlations and a path-augmented structural encoding strategy to capture high-order interactions between user-to-user and user-to-service. Then, by employing a structure-enhanced attention mechanism to aggregates attribute and structural high-order features, a comprehensive understanding of user preference is ensured. A classifier then leverages these refined representations for accurate user attribute prediction. Finally, this end-to-end training process allows the model to automatically learn the relative importance of various user or services for target user within the weighted user-service heterogeneous interaction graph. In summary, the main contributions of this paper are summarized as follows:

- We introduce a novel framework named Heterogeneous Interaction Graph Transformer (HIGT), which innovatively optimizes and enhances the conventional Transformer architecture to extract user features from the modeled weighted user-service heterogeneous interaction graph, thereby refining user profiling for personalized service recommendations.
- In the feature extraction process of HIGT, we design an innovative dual high-order feature learning approach to refine user representation by integrating both semantic and structural insights from the graph. It employs a self-attention mechanism to identify high-order semantic attribute correlations, and a path-augmented encoding strategy to capture high-order interactions along the graph's topology.
- Extensive experiments are conducted on two real-world e-commerce Web services datasets, demonstrating that our method receives significant advancements in user profiling.

The remainder of this paper is organized as follows. Section II formulates the research problem. Section III illustrates the overall framework of HIGT and elaborates the approach of user profiling. Section IV shows the experimental results and analyses. Section V reviews the related work. Finally, we conclude the paper and discuss the future work in Section VI.

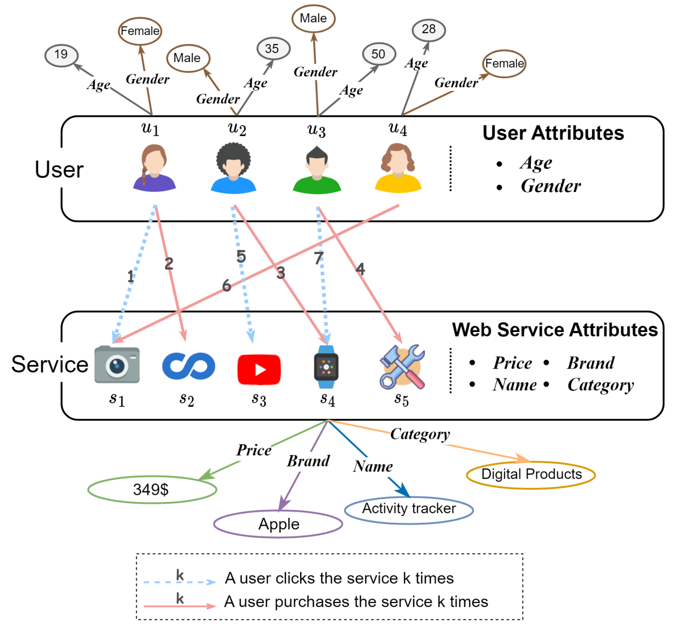


Fig. 1. User-Service heterogeneous interaction graph for user profiling in the service recommendation system.

II. Problem Formulation

In service recommender system, interactions among entities are typically modeled using a heterogeneous interaction graph. These entities, rich in attribute, are mainly categorized as 'users' and 'services'. We define these entities as follows:

Definition 1 (Service User). *Service users mainly refer to individuals who have interacted with one or more Web services. Let* $\mathcal{U}$ *be a user set, where* $\mathcal{U} = \{u_1, u_2, ..., u_m\}$*. Each* $u \in \mathcal{U}$ *is characterized by a set of attributes* $\mathcal{A}_u$*.*

Definition 2 (Web Service). *For user profiling, the Web service primarily pertains to the suite of services offered by e-commerce platforms or social media. Let* $\mathcal{S} = \{s_1, s_2, ..., s_n\}$ *denote a set of Web services. For each service* $s \in \mathcal{S}$*, it can be described by an attribute set* $\mathcal{A}_s$*.*

Next, we will explore how these entities interact with each other in a complex network structure, known as the "user-service heterogeneous interaction graph". The following is a definition of this graph.

Definition 3 (User-Service Heterogeneous Interaction Graph). *The User-Service Heterogeneous Interaction Graph can be defined as* $\mathcal{G} = \{\mathcal{U}, \mathcal{I}, \mathcal{A}, \mathcal{E}, \mathcal{W}, \mathcal{T}\}$*. This graph* $\mathcal{G}$ *consists of two types of nodes: user nodes* $\mathcal{U}$ *and service nodes* $\mathcal{S}$*, each with respective attributes* $\mathcal{A}_u$ *and* $\mathcal{A}_s$*. The interactions between users* $\mathcal{U}$ *and services* $\mathcal{S}$ *are represented as edges in the set* $\mathcal{E}$*. Each edge, denoted by* (u, s, w_{us}, t_{us})*, connects a user* $u \in \mathcal{U}$ *with a service* $s \in \mathcal{S}$*, where* $w_{us} \in \mathcal{W}$ *denotes the frequency of interactions and* $t \in \mathcal{T}$ *signifies the type of interaction. This graph models the intricate relationships between users and services, forming the basis for user profiling.*

Fig.1 presents a toy example of user-service interaction heterogeneous graph in the service recommendation system. It

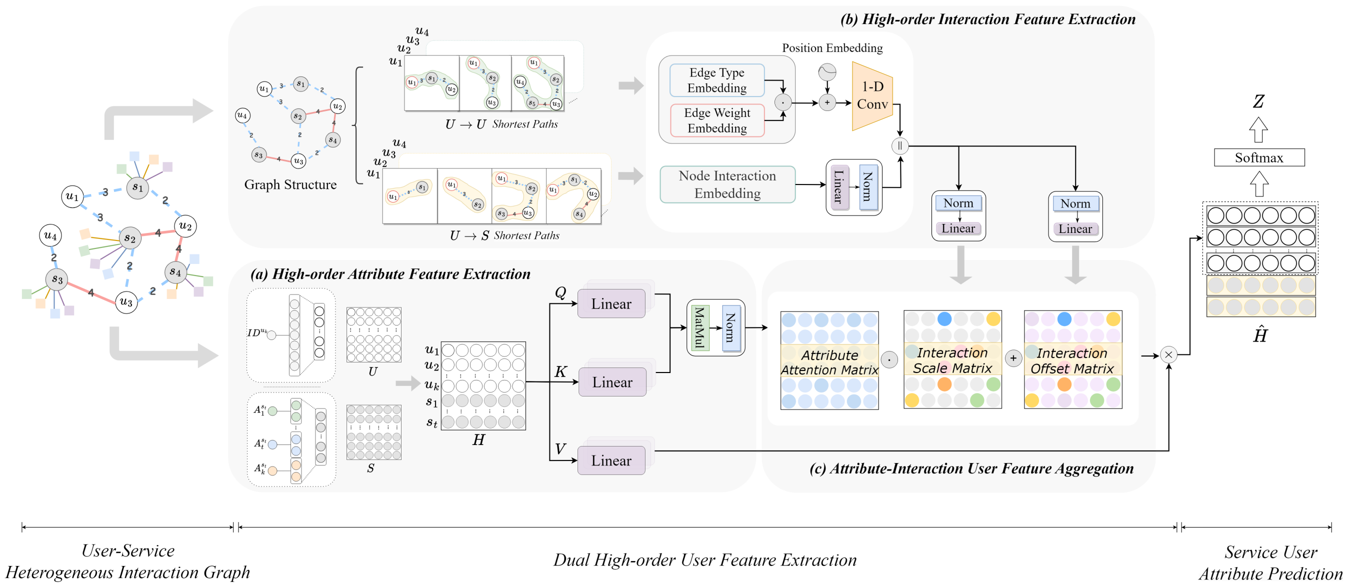


Fig. 2. The overall framework of HIGT for service user profiling.

shows users $\mathcal{U}$ with attributes such as age and gender, services $\mathcal{S}$ with attributes including price, brand, and category, and the weighted edges $\mathcal{E}$ representing the frequency of user clicks and purchases.

Definition 4 (Service User Profiling). *Given the constructed user-service heterogeneous interaction graph $\mathcal{G}$, service user profiling is defined as $T_{up} = \langle \mathcal{G}, M, \hat{\mathcal{A}} \rangle$. Here, $M(\cdot|W)$ denotes the predictive model, equipped with parameters W, which is applied to the graph $\mathcal{G}$. The set $\hat{\mathcal{A}}$ includes the inferred user attributes. The process of profiling is thus described as $\hat{\mathcal{A}} = M(\mathcal{G}|W)$, which translates interactions within the graph into individual user attributes in service recommender system.*

III. Approach

This section introduces the details of the proposed HIGT for user profiling. As shown in Fig.2, it consists of three core components: (1) The Graph Modeling Stage, which lays the foundation by construct a weighted user-service heterogeneous interaction graph. (2)The Dual High-order Feature Extraction Stage, which employs a refined self-attention mechanism within the Transformer architecture to extract high-order attribute features of users, and a structure-enhanced attention mechanism to embed the graph's topology into the Transformer architecture for extracting high-order interaction features of users; (3)The User Attribute Prediction Stage, which accepts the refined user representations and predicts specific user attributes.

A. User-Service Heterogeneous Interaction Graph Modeling

In this section, we elaborate on the construction of the User-Service Heterogeneous Interaction Graph. This graph effectively illustrates the complex interactions between users and services through diverse connection types, interaction strengths, and the specific attributes of services, providing a comprehensive view of user behaviors.

Before constructing the graph, we ensure the consistency and reliability of the raw data extracted from Web services. It consists of two primary steps:

a) Interaction Type Prioritization: Given the variety of user-service interaction modes, we introduce a hierarchy to sequence interaction types. This approach posits that certain interactions, like purchases, naturally encompass others, such as clicks. Thus, our strategy retains only the highest priority interaction in cases of multiple interactions between a user and a service.

b) Core User/Service Filtering: To balance the graph's representativeness and scale, we adopt a filtering technique to include core services interacted with by a minimum of k users and core users who have interacted with at least t different services. By doing so, we sharpen the graph's focus on capturing crucial interaction patterns from historical user behaviors.

Following data pre-processing, we construct the User-Service Heterogeneous Interaction Graph through the following steps: First, we initiate **Node Creation**, where every entity within Web services is represented as a node. Each user node, denoted by u, not only signify individual users but also encompass attribute information A_u that is the label for the user profiling task. Each service node, denoted by s, are detailed with attributes $\mathcal{A}_s$, aiding in the user profiling process. Following this, in the **Edge Establishment** phase, the interactions between users and services are depicted as weighted edges to reflect the strength of these interactions. Notably, these edges are undirected, it implies that there is no specific direction between users and services.

Next, the user-service heterogeneous interaction graph serves as a basic input to our proposed model.

B. Dual High-Order User Feature Extraction

In this section, we optimize and enhance Transformer architectures to user feature learning inspired by their recent success in graph analysis [35], and focusing on the effective integration of semantic user/service attribute and structural interaction information within the graph.

*1) **High-order Attribute Feature Extraction**:* The first step is node embedding, transforming users and services into semantic vectors in a d-dimensional space, $\mathbb{R}^d$, through embedding functions that convert discrete attributes into continuous vectors, a process also known as attribute encoding.

For users, embeddings are initialized based on their unique identifiers, formalized as $\phi_U(u) : (u, ID) \mapsto \mathbf{h_u} \in \mathbb{R}^d$. Specifically, $\phi_U(u)$ represents the function implemented by the Multi-Layer Perceptron (MLP), mapping a user's identifier to an embedding vector $\mathbf{h_u}$. For services, the embedding process is more intricate due to the diversity of their attributes. Textual attributes, such as names, which are semantically significant, are encoded using pretrained Large Language Model (LLM) framework, with the function $\phi_A(t) : (t) \mapsto \mathbf{a_t} \in \mathbb{R}^d$. Here, $\phi_A(t)$ signifies the process of calling the OpenAI's embedding API. This step transforms the rich textual attribute t into structured, low-dimensional vectors $\mathbf{a_t}$ that capture the essence of each service's identity. Additionally, for other categorical attributes, an embedding function $\phi_A(c) : (c) \mapsto \mathbf{a_c} \in \mathbb{R}^d$ maps each category attribute c to a point in the same latent space via an MLP, yielding the embedding vector $\mathbf{a_c}$.This ensures consistency with the text-based embeddings. Thus, the embedding of a service, $\phi_S(s) : (s, a_{t_1}, ..., a_{t_m}, a_{c_1}, ..., a_{c_n}) \mapsto \mathbf{h_s} \in \mathbb{R}^d$, integrates its textual and categorical attributes, formalized as:

$$\mathbf{h_s} = \sum_{i=1}^{m} \mathbf{a_{t_i}} + \sum_{j=1}^{n} \mathbf{a_{c_j}} \tag{1}$$

Following initialization, user embeddings form $H_u \in \mathbb{R}^{|U| \times d}$ and service embeddings form $H_s \in \mathbb{R}^{|S| \times d}$. These are merged into a unified matrix $H \in \mathbb{R}^{(|U|+|S|) \times d}$. This process encodes nodes with semantic information, enriching the context for analyzing user interactions and uncovering implicit preferences and behavioral patterns, which reflect user profiles.

To capture this depth of context over longer distances and further enhance our understanding of user behaviors, we have utilized a self-attention mechanism within the Transformer architecture to learn and extract the global semantic similarity within the network. This approach efficiently identifies and emphasizes entities most semantically relevant to the target user, whether they be other users or services. This aligns with the idea that user profiles can be reflected not only through their personal interaction preferences, but also from user groups with similar behavior patterns.

This means that the input matrix H is transformed by three distinct projection matrices $W_Q \in \mathbb{R}^{d \times d_K}$, $W_K \in \mathbb{R}^{d \times d_K}$, and $W_V \in \mathbb{R}^{d \times d_V}$ to obtain the query matrix Q, key matrix K, and value matrix V, respectively, which are then used within the self- attention mechanism to compute the semantic attention matrix $A \in \mathbb{R}^{(|U|+|S|) \times (|U|+|S|)}$. Consequently, the computation of pairwise semantic association in the user-service heterogeneous interaction graph is as follows:

$$Q = HW_Q \tag{2}$$

$$K = HW_K \tag{3}$$

$$V = HW_V \tag{4}$$

$$A = \frac{Q\,(K)^T}{\sqrt{d_K}} \tag{5}$$

where $a_{ij} \in A$ indicates the strength of semantic association of node v_i and v_j, and each node belonging to the set $U \cup S$. To illustrate with simplicity, we focus on the single-head self-attention framework and operate under the assumption that $d_K = d_V = d$. Furthermore, generalizing to the multi-head attention is a standard and uncomplicated process.

*2) **High-order Interaction Feature Extraction**:* The feature extraction process based on attribute adeptly discerns the semantic influences exerted by various entities on the target user, facilitating the mining of the user's potential high-order features. Meanwhile, the behavior patterns of users are also an important aspect in understanding the user's attributes. Through the analysis of users' direct interactions, we can gain initial insights into their explicit preferences. However, relying solely on this direct information may not completely reflect the users' complex preferences.

Further considering that users may exhibit similar behavior patterns, such as similar purchasing histories, we can explore high-order connections between users through these shared behavior patterns.

Specifically, the fact that similar purchasing behaviors among users imply similar attributes leads us to deduce high-order connectivity between users from shared service interactions along multi-hop paths. Formally, a user u is connected to another user u' at the L-th order if there exists a path $u \xrightarrow{e_{us_1}} s_1 \xrightarrow{e_{u_1 s_1}} u_1 \xrightarrow{e_{u_1 s_2}} s_2 \xrightarrow{e_{u_2 s_2}} \cdots \xrightarrow{e_{u_{L-1} s_L}} s_L \xrightarrow{e_{u' s_L}} u'$, where $u_i \in \mathcal{U}$, $s_i \in \mathcal{S}$, and $e_{u_s s_i}$ represents the edge between user and service in $\mathcal{E}$. Similarly, high-order connectivity between a user and a service is established based on the concept of User-User connectivity.

Under the definition of high-order connectivity based on shared interaction patterns, it is obvious that the topology of the user-service heterogeneous interaction graph intuitively maps complex interaction patterns, revealing the direct and indirect connections between users and services. To this end, we propose a path-augmented structure encoding strategy, which works by constructing multiple paths from the target user to different entities (users and services), and focuses on the k-order shortest paths reachable. By strategically encoding edges and node pairs in paths, the target user feature reflected by interaction patterns are finely learned.

a) Edge Attributes Encoding: Within graph's topology, the shortest path between two nodes can effectively reflect their structural correlation. This aspect becomes even more pertinent in heterogeneous graphs, where it is essential to consider not just the length of the shortest path, but also the types and weights of the edges along this path. For instance, in the user-service heterogeneous interaction graph depicted in Fig. 2, the shortest path connecting users u_1 and u_4 can be represented as $u_1 \overset{e_{u_1 s_2}}{\to} s_2 \overset{e_{u_3 s_2}}{\to} u_3 \overset{e_{u_3 s_3}}{\to} s_3 \overset{e_{u_4 s_3}}{\to} u_4$. This particular path not only encompasses two distinct types of edges, each symbolizing different user behaviors, but also incorporates three varying weights, each corresponding to the frequency of actions within the specific type of interaction. Thus, in order to learn topology dependencies between node pairs, we construct a shortest path matrix $E_{(u_i v_j)} \in \mathbb{R}^{2\times spd(u_i,v_j)}$ for each node pair $< u_i, v_j >$, where v_j could represent either a user $u_j$or a service s_j. The matrix:

$$E_{(u_i v_j)} = [\mathbf{p_{et}}; \mathbf{p_{ew}}] \tag{6}$$

where u_i and v_j represent the starting and ending node of the shortest path, respectively and $spd(u_i, v_j)$ aims at calculating the length of the shortest path. Each dimension of the vector $\mathbf{p_{et}}$ represents the type of edge on the shortest path, while each dimension in the second row $\mathbf{p_{ew}}$ represents the corresponding edge weight.

Subsequently, each type and weight of edges is encoded through a trainable embedding vector with real value, which is processed by two distinct full-connected embedding layers, respectively. To achieve this, we define a set of predefined edge types $\mathcal{T}$ and edge weights $\mathcal{W}$, along with corresponding embedding functions $f_t : \mathcal{T} \to \mathbb{R}^{d_t}$ and $f_w : \mathcal{W} \to \mathbb{R}^{d_t}$. These functions map the edge types and weights to a vector space of d_t dimensions of real value. Given a sequence of edges $(e_1, e_2, \ldots, e_{spd})$ on the shortest path, where each edge e_i has an edge type $t_i \in \mathcal{T}$ and an edge weight $w_i \in \mathcal{W}$, we can obtain two matrices:

$$X_{et} = [f_t(t_1); f_t(t_2); \ldots; f_t(t_{spd})] \tag{7}$$

$$X_{ew} = [f_w(t_1); f_w(t_2); \ldots; f_w(t_{spd})] \tag{8}$$

where X_{et} denotes the embedding matrix for edge types and X_{ew} indicates the embedding matrix for edge weights. These matrices are combined using the Hadamard product, defined as follows:

$$X_e = X_{et} \odot X_{ew} \tag{9}$$

where $X_e \in \mathbb{R}^{spd(u_i,v_j)\times d_t}$ represents the feature matrix for the shortest path from node u_i to v_j. Each row in this matrix corresponds to the feature vector of a specific edge in the shortest path.

b) Edge Positional Encoding: In addition, following the standard architecture of the Transformer, we incorporate a crucial step of identifying the position of each edge on the shortest path. Positional encoding is crucial because, without positional information, the model might treat all edges as if they were interchangeable, potentially leading to incorrect conclusions about the graph structure. We achieve this by augmenting each row of X_e with a positional encoding, calculated as follows:

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{2i/d_t}}\right) \tag{10}$$

$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{2i/d_t}}\right) \tag{11}$$

$$X'_e = X_e + PE \tag{12}$$

where pos refers to the index of an edge within the specified path and i represents the column index in a row vector. We then apply a one-dimensional convolutional layer to the matrix X'_e , effectively compressing its column dimensions. Consequently, this operation yields the final path feature $\mathbf{x_{(ij)e}} \in \mathbb{R}^{d_t}$, which comprehensively encodes the edge information along the shortest path.

c) Node Interaction Encoding: In the process of extracting high-order features of users based on graph topology, the role of nodes is as significant as that of edges. Specifically, the degree of a node is an essential metric that indicates its importance within the graph. However, focusing solely on individual nodes may overlook complex relationships, especially in heterogeneous graphs. Thus, we extend this concept from individual nodes to pairs of nodes to mine the relative importance of different entities. For node pairs $< u_i, v_j >$, we construct a real-valued vector $\mathbf{x_{(ij)n}}$ to quantify the correlation between neighboring nodes on the shortest paths from node u_i to v_j, which is formulated as follows:

$$\mathbf{x_{(ij)n}} = \begin{cases} [c_{u_i u_{i+1}}, c_{u_{i+1} u_{i+2}}, \ldots, c_{u_{j-1} u_j}] & ,\text{for} < u_i, u_j > \\ [\frac{w_{u_i i_1}}{N(u_i)}, \frac{w_{u_{i+1} i_1}}{N(u_{i+1})}, \ldots, \frac{w_{u_{j-1} i_j}}{N(u_{j-1})}] & ,\text{for} < u_i, i_j > \end{cases} \tag{13}$$

For user node pairs $< u_i, u_j >$, $\mathbf{x_{(ij)n}} \in \mathbb{R}^{\frac{spd(u_i,u_j)}{2}}$ quantifies the number of shared interaction services between adjacent user nodes along the shortest path, with each dimension indicates the count of shared interaction services between the user u_k and u_{k+1}.

For user-service node pairs $< u_i, s_j >$, $\mathbf{x_{(ij)n}} \in \mathbb{R}^{spd(u_i,s_j)}$, each dimension of this vector reflects the relative frequency of interaction between the user and a specific service. Where $w_{u_m s_n}$ represents the interaction intensity between user u_m and service s_n, and $N(u_m)$ represents the total frequency of interactions the user u_m has with all services they directly interact with.

The lengths of $\mathbf{x_{(ij)n}}$ are respectively half the shortest path for user-user pairs and equal to the full shortest path distance for user-service pairs. Subsequently, this vector is projected through a linear transformation and normalization layer, resulting in a higher-dimensional representation $\mathbf{x_{(\hat{ij})n}} \in \mathbb{R}^{d_t}$, which effectively captures the interactions of node pairs on the shortest path.

*3) **Attribute-Interaction User Feature Aggregation**:* To this point, we have completed the extraction of two key aspects of user features: the extraction of high-order semantic features based on attribute information and the extraction of high-order interaction features based on the graph structure. In order to build a more comprehensive and in-depth user representation, we integrate these two types of features through an effective fusion strategy. This is achieved through a structure-enhanced attention mechanism that builds upon the foundation of the Transformer's self-attention mechanism. It adjusts the standard attention coefficients a_{ij} in accordance with topological characteristics. Specifically, the final user interaction feature based on graph structure $\mathbf{x_{ij}} \in \mathbb{R}^{2d_t}$ are computed first by concatenating the edge feature vector $\mathbf{x_{(ij)e}} \in \mathbb{R}^{d_t}$ and the node interaction feature vector $\mathbf{x_{(\hat{ij})n}} \in \mathbb{R}^{d_t}$ along the shortest path between the nodes pairs between u_i and v_j. Subsequently, the concatenated vector is subjected to a linear transformation and normalization to yield the final feature representation:

$$\mathbf{x_{ij}} = \text{Normalize}\left(W_t(\mathbf{x_{(ij)e}} \parallel \mathbf{x_{(\hat{ij})n}}) + b_t\right) \tag{14}$$

where $W_t \in \mathbb{R}^{2d_t \times 2d_t}$ represents a learnable weight matrix, $b_t \in \mathbb{R}^{2d_t}$ denotes a bias vector, both of which are parameters to be optimized during training. The resulting structural features x_{ij} informs the adjustment of attention coefficients through an affine transformation as follows:

$$\alpha_{ij} = \mathbf{w_s}^T \cdot \mathbf{x_{ij}} \tag{15}$$

$$\beta_{ij} = \mathbf{w_b}^T \cdot \mathbf{x_{ij}} \tag{16}$$

$$\hat{a}_{ij} = \alpha_{ij} a_{ij} + \beta_{ij} \tag{17}$$

$$\hat{H} = \text{Normalize}(\hat{A})V \tag{18}$$

where w_s and w_b are the trainable vectors. α_{ij} and β_{ij} represent the structural scale and structural offset coefficients, respectively; $\hat{a}_{ij}$ is the refined attention coefficient for nodes v_i and v_j. The matrix $\hat{H} \in \mathbb{R}^{(|U|+|S|)\times d'}$ indicates the final output and embed all the users, capturing the comprehensive semantic correction and topological dependency.

C. User Profiling and Model Training

The prediction layer predicts the labels of user attributes based on the learned user representations. We first select the top $|U|$ rows corresponding to users from the matrix $\hat{H}$, and then linearly projected to a lower-dimensional space, resulting in the matrix $Z \in \mathbb{R}^{|U| \times F_c}$, where F_c is the number of attribute categories for classification. It can be formalized as:

$$Z = \hat{H}_U W_h + b_h \tag{19}$$

$$\hat{Y} = \textit{softmax}(Z) \tag{20}$$

where W_h is the projection matrix that reduces the dimension to F_c, and b_h is a bias vector. A softmax operation is then applied row-wise function on the matrix Z to yield a probability distribution $\hat{Y}$ over the labels for each user.

Upon obtaining the probability distribution over the labels for each user, the model employs the cross-entropy loss function for effective multi-class classification. The loss function is defined as:

$$\mathcal{L}(\theta) = -\frac{1}{|U|} \sum_{i \in U} \sum_{j \in F_c} y_{ij} \log(\hat{y}_{ij}) + \lambda \|\theta\|_2^2 \tag{21}$$

here, $\mathcal{L}(\theta)$ represents the loss function dependent on all the model parameters θ. The term $-\frac{1}{|U|}\sum_{i \in U}\sum_{j \in F_c} y_{ij} \log(\hat{y}_{ij})$ quantifies the average cross-entropy loss across all users, where y_{ij} is the ground truth label of user i for the j-th category within a particular attribute, and $\hat{y}_{ij} \in \hat{Y}$ is the corresponding predicted probability. The regularization term $\lambda\|\theta\|_2^2$ mitigates overfitting by penalizing large parameter values, with λ being a pre-defined regularization coefficient.

We optimize the model parameters θ using the mini-batch AdamW [36] algorithm. This algorithm is a widely used optimizer, incorporating weight decay regularization to enhance performance and reduce overfitting.

TABLE I
STATISTICS OF LABEL IN ALIBABA-DATASET.

Gender		**Age**				
Male	Female	1	2	3	4	5
32765	26601	3989	10681	19157	18332	7207

TABLE II
STATISTICS OF LABEL IN JD-DATASET.

Gender		**Age**			
Male	Female	< 26	26-35	36-55	> 55
21769	13881	6204	26336	2482	628

IV. Experiments

A. Experimental Datasets and Setup

1) Datasets Description: We conduct comprehensive experiments using two large-scale public service datasets from Alibaba and JD.com, the leading e-commerce platforms in China. These datasets, identified as the Alibaba-dataset[1] and the JD-dataset[2], consist of authentic user behavior and profile data, crucial for reflecting real-world e-commerce interactions. In both datasets, the heterogeneous graphs are extracted with multiple relationships among users and services. The Alibaba-dataset, encompassing data from users on Taobao, includes a rich history of advertising displays and user behavior logs. It details user-service interactions like "purchase," "click," "add to shopping cart," and "favorite," alongside product attributes such as categoryID and brandID, as well as user attributes including gender, age, consumption level, and city tier, etc. In the JD-dataset, users and services have 'click' and 'purchase' interactions, with service attributes extending to category, brand, name, and price. Notably, service names are provided in Chinese. Additionally, it includes user attributes like age and gender. Aligning with previous works [7]–[9], we use the user's gender and age as the label for our user profiling task. Labels of users' age in Alibaba-dataset have been pre-classified in the original dataset. Additionally, both datasets undergo a filtering process to refine their structure for enhanced efficacy in model training and evaluation. Detailed descriptions of the labels within two datasets are presented in Table I and Table II.

[1]https://tianchi.aliyun.com/dataset/dataDetail?dataId=56
[2]https://github.com/guyulongcs/IJCAI2019_HGAT

TABLE III
OVERALL PERFORMANCE COMPARISON OF VARIOUS MODELS ON JD-DATASET AND ALIBABA-DATASET.

	Task	***Age***				***Gender***			
	Model	Accuracy	Macro-P	Macro-R	Macro-F1	Accuracy	Macro-P	Macro-R	Macro-F1
JD	LR	62.82	0.24	0.25	31.44	50.12	0.50	0.50	48.10
	SVM	63.52	0.25	0.25	32.27	50.51	0.50	0.50	49.92
	GCN	62.67	0.25	0.25	31.86	53.29	0.50	0.51	51.36
	GAT	63.71	0.26	0.26	32.24	51.82	0.49	0.50	49.39
	HGCN	66.29	0.33	0.34	34.65	52.47	0.52	0.51	51.48
	HGAT	66.71	0.32	0.34	34.83	52.91	0.52	0.52	51.52
	RHGN	68.37	0.33	0.35	36.91	53.48	0.53	0.52	51.91
	HIGT	**72.24**	**0.44**	**0.46**	**41.24**	**56.61**	**0.56**	**0.55**	**54.95**
Alibaba	LR	22.90	0.18	0.20	18.07	55.93	0.51	0.51	49.16
	SVM	23.08	0.19	0.18	20.31	57.82	0.51	0.52	52.59
	GCN	24.14	0.20	0.18	20.71	58.40	0.53	0.52	53.48
	GAT	24.92	0.21	0.20	21.28	57.38	0.52	0.51	52.22
	HGCN	23.25	0.20	0.19	20.73	59.35	0.54	0.53	54.01
	HGAT	23.94	0.20	0.21	20.98	58.96	0.53	0.54	53.89
	RHGN	43.71	0.28	0.29	29.29	60.82	0.59	0.58	59.46
	HIGT	**46.36**	**0.30**	**0.32**	**30.73**	**61.67**	**0.61**	**0.62**	**60.83**

2) Experimental Setup: To evaluate the effective of our proposed HIGT, we conduct a series of comprehensive experiments. These experiments are performed using a workstation furnished with two NVIDIA GTX 4090 GPUs, an Intel(R) Xeon(R) Gold 6130 processor clocked at 2.60 GHz, and 1024GB of RAM. The HIGT's components are implemented using Python 3.8.16 and Pytorch 1.13.0.

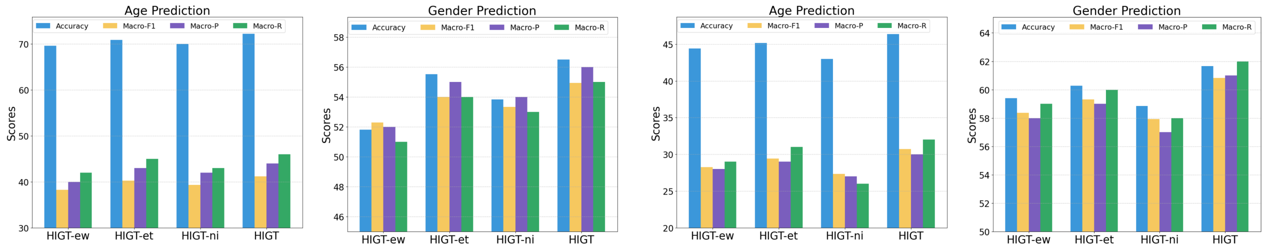
In our experiments, following prior works [7], [8], we randomly divide the labeled users into training, validation, and test sets in a 75:12.5:12.5 ratio. During the training phase, we engage all embeddings and labels assigned to users in the training set. For validation and testing phases, we evaluate our model using the labels of users in the validation and test sets respectively. To facilitate calculation, the embedding dimensions for both users and services are fixed at 128, i.e., $d = 2d_t = d' = 128$. Considering both computational resources and the scale of the dataset, in the process of extraction of high-order user interaction features by calculating the shortest paths between node pairs, we focus on up to third-order neighbors of the user node, i.e., $spd(u_i, v_j) \leq 3$, employing Dijkstra's algorithm for shortest path computations.

We adopt a grid-search strategy to find the optimal parameter combination for the model. The number of Transformer encoder layers is searched within {1,2,3,4,5}; the number of heads in multi-head attention is searched within {1,2,4,8,16}. We employ the early stopping technique for training the model, achieving optimization within 50 epochs, and set the weight decay to 0.1. For gender prediction, the learning rate, dropout rate, and mini-batch size are set to 0.005, 0.6, and 64, respectively. For age prediction, they are set to 0.1, 0.2, and 32, respectively. In both the path-augmented structural encoding module and the attribute-based semantic encoding module, we use GELU [37] as our activation function. All baseline implementations follow their respective original papers.

B. Competing Methods and Evaluation Metrics

To evaluate the performance of the proposed HIGT, we compared it with classical and state-of-the-art graph-based methods for user profiling task. The details of these comparative methods are outlined below.

- **LR** [38]: It is widely used for its efficiency and interpretability, often serving as an initial benchmark for user profiling.
- **SVM** [39]: It excels in classification tasks by constructing a hyperplane in a high-dimensional space that maximizes the margin between different classes of nodes.
- **GCN** [40]: It is specifically designed for graph-structured data and stands out in node classification tasks through its semi-supervised learning algorithm.
- **GAT** [41]: It refines the process of node classification through an attention-based mechanism that assigns varying levels of significance to the nodes in a neighborhood. While GCN and GAT serve as representative and strong baselines for numerous tasks centered around homogeneous graphs, they do not inherently account for multiple types of relations and entities.
- **HGCN** [7]: It extends traditional graph convolutional networks to better handle heterogeneous graphs for semi-supervised user profiling. It utilizes the graph convolutional operation to effectively capture complex user features from diverse types of nodes and edges.



(a) JD-dataset

(b) Alibaba-dataset

Fig. 3. Performance of HIGT and variants on two datasets.

- **HGAT** [7]: It is a variant of HGCN, employs an attention mechanism and focuses on dynamically allocating significance to various node interactions.
- **RHGN** [8]: It is a state-of-the-art method in the field of user profiling on heterogeneous graphs. It integrates a relation-aware strategy, specifically designed to handle the intricate dynamics between diverse node types.

User profiling is essentially a classification task. In our experiments, which involve the classification tasks of gender prediction and age prediction, we employ four key metrics to measure the accuracy of user attributes prediction among the competing approaches: Accuracy, Macro-Precision, Macro-Recall, and Macro-F1 Score [42]. These metrics are widely used in classification and user profiling problems.

C. Experimental Results and Analyses

Comparison with Competing Methods. In this study, we evaluate the performance of our HIGT model against baseline models on the JD-dataset and Alibaba-dataset, specifically focusing on gender and age prediction tasks. The experimental outcomes, as illustrated in Tables III, clearly demonstrate the superiority of our HIGT model over baseline models. It can be observed from the results that traditional machine learning methods like LR and SVM show limited performance on both datasets. This performance gap is attributed to their limited capacity to incorporate the rich network structure and complex node interrelations into the predictive modeling process, relying solely on initial node attributes for user representation. Graph neural network-based methods, such as GCN and GAT, outperform traditional methods, with GAT showing better results because of its attention mechanism, which accounts for varying node importances. Additionally, models specifically tailored for heterogeneous graphs, including HGAT, HGCN, and RHGN, nearly exhibit superior performance in age and gender prediction tasks compared to techniques intended for homogeneous graphs. This highlights the critical role of heterogeneous graphs in deriving diverse semantic representations from varied data sources, offering a richer contextual understanding of users.

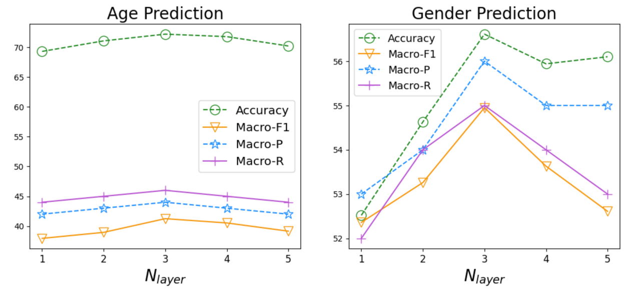
Most notably, HIGT not only achieves the highest accuracy across both tasks and datasets but also excels in Macro-Precision, Macro-Recall, and Macro-F1 score metrics. Specifically, HIGT's accuracy and Macro-F1 score for gender prediction are elevated by 5.85% and 5.86% on the JD-dataset and by 1.40% and 2.30% on the Alibaba-dataset, respectively. For age prediction, these metrics increase by 5.66% and 11.73% on the JD-dataset, and by 6.06% and 4.92% on the Alibaba-dataset, respectively. Our model exhibits superior performance on the JD dataset compared to its performance on the Alibaba-dataset. The variations could be attributed to the differing intensity and nature of high-order interactions within these datasets. One possible explanation is that the high-order interactions between users and other entities are not as obvious in the Alibaba-dataset as in the JD-dataset, which weakens the functionality of the path-enhanced structure encoding module and limits its capacity to leverage these shared interaction patterns. Another possible reason could be that in the Alibaba-dataset, the number of services with few interactions surpasses the number of services that are frequently interacted with. the JD-dataset demonstrates a more uniform distribution in the intensity of user interactions.

Although the HIGT model performed best on both tasks for both datasets, it is important to note that all models generally performed poorly on the age prediction task. Age prediction, unlike the binary classification of gender, is a multi-class challenge with subtle indicators that are often intertwined with various demographic and behavioral factors, making it slightly challenging for models to treat it as a classification task only.

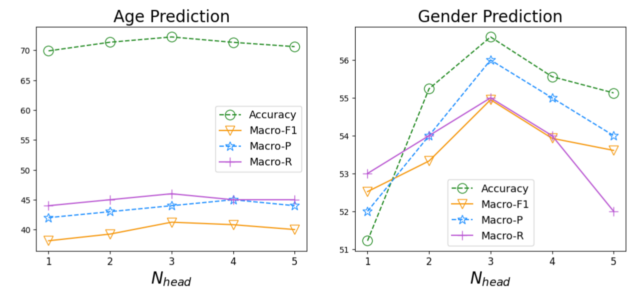
In general, HIGT achieves the best results primarily for the following reasons. Firstly, HIGT utilizes heterogeneous information from different data sources and meticulously reconstructs the real interactions between users and services using a user-service heterogeneous interaction graph, thereby gaining a deeper understanding of user behavior. Furthermore, HIGT introduces a dual self-attention mechanism to aggregate structural and attribute information to extract high-order features of the target users, thereby obtaining high-quality user representations.

Ablation Study. To evaluate the effectiveness of the path-augmented structural encoding module within HIGT, we conduct experiments with modified versions of the full HIGT model, each lacking different components of the module. The performance of these variants was evaluated in comparison to the original HIGT model, with the findings depicted in Figure 3. These variants include:

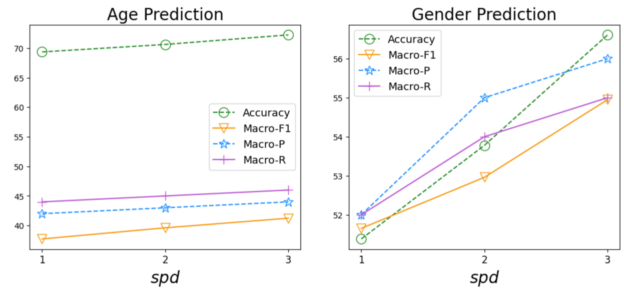
- HIGT-ew: It is a variant of HIGT, but it only uses the edge type encoding and the node interaction encoding in



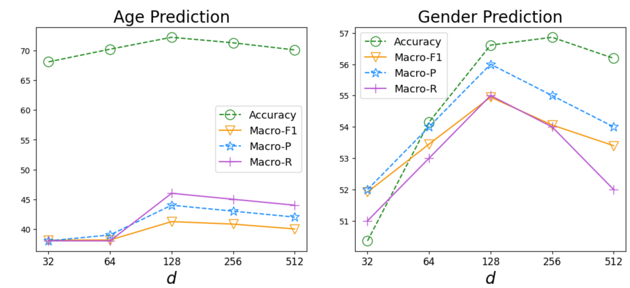
(a)



(b)



(c)



(d)

Fig. 4. Parameter sensitivity of HIGT on JD-dataset.

the path-augmented structural encoding module(i.e., no edge weight encoding module). The modified structural encoding module will be fused with the original global attribute-based semantic encoding module to extract user features, ensuring a focused adaptation while maintaining the integrity of semantic analysis. Comparing its performance with the standard HIGT highlights the role of edge weight information in understanding user interactions and preferences.

- HIGT-et: It is a variant of HIGT, but it only uses the edge weight encoding and the node interaction encoding in the path-augmented structural encoding module(i.e., no edge type encoding module). This comparison can reveal the importance of considering edge type information for the model's accuracy and efficiency.
- HIGT-ni: It is a variant of HIGT, but it only uses the edge type encoding and the node weight encoding in the path-augmented structural encoding module(i.e., no node interaction encoding module). This setup provides insights into the significance of quantifying interactions between neighbors along the shortest paths.

From the results, it is clear that HIGT outperforms other model structures in both datasets, underscoring the efficacy of its comprehensive path-augmented structural encoding module. The comparison with its variants—HIGT-ew lacking edge weight encoding, HIGT-et without edge type encoding, and HIGT-ni missing node interaction encoding—emphasizes the significance of each component in the model. Specifically, edge weight encoding enhances the model's understanding of interaction strengths, making it a more critical factor than edge type in our model, as it provides a deeper insight into the intensity and significance of user interactions. This finding suggests that the quantitative aspect of user interactions plays a more pivotal role than the qualitative aspect in understanding user behavior. Furthermore, edge type encoding offers insights into the nature of these interactions, making the model's predictions more relevant to actual user behaviors. Additionally, the node interaction encoding capitalizes on graph topology to infer user-user similarities based on shared neighbors and user-service similarities from the relative importance of services to different users, effectively using high-order connectivity to augment the understanding of user behavior. This approach implies that users with similar behaviors are likely to share similar attributes, thereby enhancing the model's accuracy in user profiling.

D. Performance Impact of Parameters

In this section, we explore how four crucial parameters affect our model's effectiveness: the number of encoding layers N_{layer}, the number of attention heads N_{head}, the shortest path distance spd, and the embedding dimension d. We display the results from the JD-dataset in Fig.4.

Based on Fig.4 (a), we can observe that there is a positive correlation between the model's performance and the number of layers in the encoding layer within a certain range. A larger N_{layer} is beneficial for enhancing the model's performance, but it incurs additional computational costs and increases the risk of overfitting. In our experimental scenario, N_{layer} is set to 3 to achieve a balance between computational resource utilization and model performance. Furthermore, it can be seen from Fig.4 (b) that the multi-head self-attention mechanism in the Transformer works well in our task, as the model's performance improves with the increase of N_{head}. However, too many heads can decrease performance. Therefore, we set $N_{head} = 4$. For the shortest path distance spd, from Fig.4 (c), due to limited computational resources, we only consider the maximum distance of 3. Within this range, age prediction remains mostly stable, while gender prediction shows marked improvement, suggesting that gender prediction benefits from a broader structural information captured by a higher spd. The analysis of Fig.4 (d) shows that the performance metrics improve as the embedding dimension d increases for both age and gender prediction, peaking at $d = 128$, except for Accuracy. The final choice of 128 is guided by the peak Macro-F1 score at this dimension, which balances precision and recall effectively.

V. RELATED WORK

User profiling, defined as the process of labeling user, has attracted widespread attention and research from both academia and industry over the past decades. Initially, user profiling predominantly served as an intermediary step in service recommendation system tasks. For instance, He et al. [43] employed matrix factorization (MF) strategies for user analysis based on user ratings or click behaviors. With the rapid development of the Internet, the significance of user profiling has increasingly come to the forefront, leading researchers to focus on it as a distinct field of study. In this context, the methods used for extracting user features have evolved significantly. Existing research on these methodologies can be broadly divided into two categories: single-source based methods and fusion-based methods.

A. Single-source based User Profiling

Existing methods for user profiling typically first extract features from text, behavior, relationships, etc., and then apply machine learning or deep learning techniques to infer user attributes.

Considering mining user characteristics from texts, Flekova et al. [5] employed linear and nonlinear machine learning regression methods to explore the relationship between stylistic and syntactic features and users' age and income; Zamal et al. [44] employed SVM and GBDT models using the homophily principle to derive features from Twitter profiles and friends' posts to enhance user characteristics, improving the performance of attributes with moderate to high combinatoriality. Beyond basic user attributes, Preoţiuc-Pietro et al. [45] developed models for user-level race and ethnicity predictions based on Twitter text, using logistic regression to predict membership in the four largest racial and ethnic groups.

Besides text, user behavior also reflects attributes to some extent, as indicated by [6], which inferred users' age and gender based on patterns in daily mobile communication; Wang et al. [46] proposed a novel Structured Neural Embedding (SNE) model to automatically learn representations from user purchase data, thus predicting multiple demographic attributes simultaneously.

Additionally, scholars have focused on identifying user dependency relationships across social networks to help infer user attributes. Ding et al. [47] devised different strategies to calculate relational weights between users, applying graph-based semi-supervised learning algorithms for attribute prediction; Mislove et al. [48] discovered that users with common attributes are more likely to become friends, leading to a novel community detection method that can infer the attributes of the remaining users with only 20% of them providing their attributes.

B. Fusion-based User Profiling

In recent years, scholars have placed greater emphasis on how to model heterogeneous information from multiple user data sources automatically to learn user representations, overcoming the limitations of traditional methods that often involve single data types and relationships and require manual feature design.

Specifically, the model proposed by Miura et al. [11] integrates text, metadata, and user network representations with an attention mechanism to predict user geolocation. Similarly, the deep learning approach in [10] is capable of extracting and fusing information from different modalities, utilizing shared representations across these modalities to infer age, gender, and personality traits of social media users. However, these methods still face challenges with manual feature design or fusion techniques. Consequently, some scholars have begun to apply the advantageous properties of graphs, which are adept at handling unstructured data, to the combination of multi-typed user data related to user attributes. For instance, HGAT [7] learns representations for each entity through the constructed heterogeneous graph and used an attention mechanism to differentiate the importance of each neighboring entity, utilizing unsupervised information and limited user labels to build predictors for users' age and gender. RHGN [8] further emphasizes the distinctions among different types of interactions, employing a relation-aware heterogeneous graph approach for user analysis. IHNN [9] adopts the idea of heterogeneous attention mechanisms and designs hypergraph convolutional operations to mine interactions beyond pairwise relations among users from multiple views, thereby enhancing the performance of user analysis.

However, these methods all overlook the differences within the same interaction type, and more importantly, the high-order interactions between different entities in heterogeneous graphs are also ignored, which may limit the generalizability of these methods in different application scenarios.

VI. CONCLUSION AND FUTURE WORK

In this paper, we leverage graph structure to model users' historical interactions on web services, aiming to enhance user profiling for personalized service recommendation. For this purpose, we propose an innovative framework named Heterogeneous Interaction Graph Transformer (HIGT) to extract user dual high-order feature. By constructing a weighted user-service heterogeneous interaction graph, we employ a self-attention mechanism and path-augmented encoding strategy within Transformer architecture, efficiently capturing global semantics feature and high-order interactions features. Experiments on real-world large service datasets reveal that this dual strategy significantly enhances the model's understanding of user behaviors and improves the accuracy of user profiling, outperforming the most advanced baseline methods. Furthermore, we validate the effectiveness of each component in HIGT.

In the future, we plan to further explore how to apply the HIGT framework to dynamic heterogeneous graphs effectively. Additionally, we will also investigate how to further optimize the computational efficiency and scalability of the model to accommodate larger datasets.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (No. 62272290, 62172088) and Shanghai Natural Science Foundation (No. 21ZR1400400).

REFERENCES

[1] L. Wu, J. Li, P. Sun, R. Hong, Y. Ge, and M. Wang, "Diffnet++: A neural influence and interest diffusion network for social recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 10, pp. 4753–4766, 2020.

[2] Z. Lu, S. J. Pan, Y. Li, J. Jiang, and Q. Yang, "Collaborative evolution for user profiling in recommender systems." in *International Joint Conference on Artificial Intelligence (IJCAI)*, vol. 30, 2016, pp. 3804–3810.

[3] L. Liao, X. He, H. Zhang, and T.-S. Chua, "Attributed social network embedding," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 12, pp. 2257–2270, 2018.

[4] H. Fang, X. Li, and J. Zhang, "Integrating social influence modeling and user modeling for trust prediction in signed networks," *Artificial Intelligence*, vol. 302, p. 103628, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000437022100179X

[5] L. Flekova, D. Preoţiuc-Pietro, and L. Ungar, "Exploring stylistic variation with age and income on twitter," in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016, pp. 313–319.

[6] Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla, "Inferring user demographics and social strategies in mobile social networks," in *International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, 2014, pp. 15–24.

[7] W. Chen, Y. Gu, Z. Ren, X. He, H. Xie, T. Guo, D. Yin, and Y. Zhang, "Semi-supervised User Profiling with Heterogeneous Graph Attention Networks," in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2019, pp. 2116–2122.

[8] Q. Yan, Y. Zhang, Q. Liu, S. Wu, and L. Wang, "Relation-aware Heterogeneous Graph for User Profiling," in *International Conference on Information and Knowledge Management (CIKM)*, 2021, pp. 3573–3577.

[9] S. Yan, T. Zhao, and J. Deng, "Interaction-aware Hypergraph Neural Networks for User Profiling," in *International Conference on Data Science and Advanced Analytics (DSAA)*, 2022, pp. 1–10.

[10] G. Farnadi, J. Tang, M. De Cock, and M.-F. Moens, "User profiling through deep multimodal fusion," in *ACM International Conference on Web Search and Data Mining (WSDM)*, 2018, pp. 171–179.

[11] Y. Miura, M. Taniguchi, T. Taniguchi, and T. Ohkuma, "Unifying Text, Metadata, and User Network Representations with a Neural Network for Geolocation Prediction," in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2017, pp. 1260–1272.

[12] A. Rahimi, T. Cohn, and T. Baldwin, "Semi-supervised User Geolocation via Graph Convolutional Networks," in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018, pp. 2009–2019.

[13] L. Flekova, D. Preoţiuc-Pietro, and L. Ungar, "Exploring stylistic variation with age and income on twitter," in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2016, pp. 313–319.

[14] F. A. Zamal, W. Liu, and D. Ruths, "Homophily and Latent Attribute Inference: Inferring Latent Attributes of Twitter Users from Neighbors," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 6, no. 1, pp. 387–390, 2012, number: 1. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14340

[15] D. Preoţiuc-Pietro and L. Ungar, "User-Level Race and Ethnicity Predictors from Twitter Text," in *International Conference on Computational Linguistics (ICCI)*, vol. Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 1534–1545.

[16] S. Rosenthal and K. McKeown, "Age Prediction in Blogs: A Study of Style, Content, and Online Behavior in Pre- and Post-Social Media Generations," in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2011, pp. 763–772.

[17] V. Thakur and A. Tickoo, "Text2Gender: A Deep Learning Architecture for Analysis of Blogger's Age and Gender," 2023, publisher: arXiv Version Number: 1. [Online]. Available: https://arxiv.org/abs/2305.08633

[18] R. G. Guimaraes, R. L. Rosa, D. De Gaetano, D. Z. Rodriguez, and G. Bressan, "Age Groups Classification in Social Network Using Deep Learning," *IEEE Access*, vol. 5, pp. 10 805–10 816, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7932459/

[19] Y. Gu, Z. Ding, S. Wang, and D. Yin, "Hierarchical User Profiling for E-commerce Recommender Systems," in *International Conference on Web Search and Data Mining (WSDM)*, 2020, pp. 223–231.

[20] Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla, "Inferring user demographics and social strategies in mobile social networks," in *International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, 2014, pp. 15–24.

[21] P. Wang, J. Guo, Y. Lan, J. Xu, and X. Cheng, "Your Cart tells You: Inferring Demographic Attributes from Purchase Data," in *International Conference on Web Search and Data Mining (WSDM)*, 2016, pp. 173–182.

[22] C. Wu, F. Wu, J. Liu, S. He, Y. Huang, and X. Xie, "Neural Demographic Prediction using Search Query," in *International Conference on Web Search and Data Mining (WSDM)*, 2019, pp. 654–662.

[23] M. Diao, Z. Zhang, S. Su, S. Gao, and H. Cao, "UPON: User Profile Transferring across Networks," in *International Conference on Information and Knowledge Management (CIKM)*. Association for Computing Machinery, 2020, pp. 265–274.

[24] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: inferring user profiles in online social networks," in *International Conference on Web Search and Data Mining (WSDM)*. Association for Computing Machinery, 2010, pp. 251–260.

[25] A. Mahmoudi, M. R. Yaakub, and A. A. Bakar, "The Relationship between Online Social Network Ties and User Attributes," *ACM Transactions on Knowledge Discovery from Data*, vol. 13, no. 3, pp. 26:1–26:15, May 2019.

[26] S. Bartunov, A. Korshunov, S.-T. Park, W. Ryu, and H. Lee, "Joint Link-Attribute User Identity Resolution in Online Social Networks."

[27] Y. Ding, S. Yan, Y. Zhang, W. Dai, and L. Dong, "Predicting the attributes of social network users using a graph-based machine learning method," *Computer Communications*, vol. 73, pp. 3–11, Jan. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140366415002455

[28] J. Wang, S. Li, M. Jiang, H. Wu, and G. Zhou, "Cross-media User Profiling with Joint Textual and Social User Embedding," in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018, pp. 1410–1420.

[29] Y. Gu, Z. Ding, S. Wang, and D. Yin, "Hierarchical User Profiling for E-commerce Recommender Systems," in *International Conference on Web Search and Data Mining (WSDM)*, 2020, pp. 223–231.

[30] G. Farnadi, J. Tang, M. De Cock, and M.-F. Moens, "User Profiling through Deep Multimodal Fusion," in *International Conference on Web Search and Data Mining (WSDM)*, 2018, pp. 171–179.

[31] I. M. Al-Zuabi, A. Jafar, and K. Aljoumaa, "Predicting customer's gender and age depending on mobile phone data," *Journal of Big Data*, vol. 6, no. 1, p. 18, Feb. 2019. [Online]. Available: https://doi.org/10.1186/s40537-019-0180-9

[32] M. Diao, Z. Zhang, S. Su, S. Gao, and H. Cao, "UPON: User Profile Transferring across Networks," in *International Conference on Information and Knowledge Management (CIKM)*. Association for Computing Machinery, 2020, pp. 265–274.

[33] G. Bianconi, *Higher-Order Networks*, ser. Elements in Structure and Dynamics of Complex Networks. Cambridge University Press, 2021. [Online]. Available: https://books.google.com.tw/books?id=1-5UEAAAQBAJ

[34] A. R. Benson, D. F. Gleich, and D. J. Higham, "Higher-order network analysis takes off, fueled by classical ideas and new data," *arXiv preprint arXiv:2103.05031*, 2021.

[35] C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, Y. Shen, and T.-Y. Liu, "Do transformers really perform badly for graph representation?" *Advances in Neural Information Processing Systems*, vol. 34, pp. 28 877–28 888, 2021.

[36] I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in *International Conference on Learning Representations (ICLR)*, 2019.
[37] D. Hendrycks and K. Gimpel, “Gaussian error linear units (gelus),” *arXiv preprint arXiv:1606.08415*, 2016.
[38] S. Rosenthal and K. McKeown, “Age prediction in blogs: A study of style, content, and online behavior in pre-and post-social media generations,” in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2011, pp. 763–772.
[39] F. Al Zamal, W. Liu, and D. Ruths, “Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors,” in *International AAAI Conference on Web and Social Media (ICWSM)*, vol. 6, no. 1, 2012, pp. 387–390.
[40] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016.
[41] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph attention networks,” vol. abs/1710.10903, 2017.
[42] C. Wu, F. Wu, J. Liu, S. He, Y. Huang, and X. Xie, “Neural demographic prediction using search query,” in *ACM International Conference on Web Search and Data Mining (WSDM)*, 2019, pp. 654–662.
[43] X. He, H. Zhang, M.-Y. Kan, and T.-S. Chua, “Fast matrix factorization for online recommendation with implicit feedback,” in *International ACM SIGIR conference on Research and Development in Information Retrieval (SIGIR)*, 2016, pp. 549–558.
[44] F. Al Zamal, W. Liu, and D. Ruths, “Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors,” in *International AAAI Conference on Web and Social Media (ICWSM)*, vol. 6, no. 1, 2012, pp. 387–390.
[45] D. Preoţiuc-Pietro and L. Ungar, “User-level race and ethnicity predictors from twitter text,” in *Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018, pp. 1534–1545.
[46] P. Wang, J. Guo, Y. Lan, J. Xu, and X. Cheng, “Your cart tells you: Inferring demographic attributes from purchase data,” in *ACM International Conference on Web Search and Data Mining (WSDM)*, 2016, pp. 173–182.
[47] Y. Ding, S. Yan, Y. Zhang, W. Dai, and L. Dong, “Predicting the attributes of social network users using a graph-based machine learning method,” *Computer Communications*, vol. 73, pp. 3–11, 2016.
[48] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, “You are who you know: inferring user profiles in online social networks,” in *ACM International Conference on Web Search and Data Mining (WSDM)*, 2010, pp. 251–260.