IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 January 2024, accepted 4 April 2024, date of publication 17 April 2024, date of current version 29 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3390156

METHODS

Joint Urban Modeling With Graph Convolutional Networks and Crowdsourced Data: A Novel Approach

CHAO DENG[1], XUEXIA LIANG[1], XU YAN[2], YUHUA MO[1], SEN BAI[1], BIN LU[1], KAIDI CHEN[1], XIPENG LIU[1], AND ZHI CHEN[1]

[1]Internet Research Center, China Tobacco Guangxi Industrial Company Ltd., Nanning 530001, China
[2]College of Computer Science and Institute for Industrial Internet Research, Sichuan University, Chengdu 610065, China

Corresponding author: Yuhua Mo (moyuhua95@gmail.com)

ABSTRACT Graph Convolutional Networks (GCN) are a potent and adaptable tool for effectively processing and analyzing continuous spatial data. Despite the substantial potential of GCN in various domains, most existing spatial data prediction models are confined to defining weights solely based on distance. To overcome this limitation, this study proposes a novel approach to obtain the second-level embedding of Points of Interests (POIs) by employing Delaunay Triangulation (DT), Random Walk, and Skip-Gram model training. Subsequently, enhanced features are obtained through various aggregation strategies for regional embedding. The integrated grid data, including longitude and latitude coordinates, enhanced features, and target values, are then integrated. Finally, the GCN is utilized for training and fitting to achieve the final prediction target value. By considering the influence of weights on data prediction, this approach can more accurately reflect the distribution and relationships of data in the actual environment. Furthermore, we have experimentally validated the effectiveness of this approach, demonstrating that it significantly enhances the accuracy of spatial data prediction when compared to the original GCN model's approach.

INDEX TERMS Graph convolutional networks, POIs embedding, aggregation strategy.

I. INTRODUCTION

In smart cities, urban spatial data has become an indispensable data type. Its application occupying a significant position in various fields such as ecological construction [1], social networks [2], urban planning [3], and more. However, due to objective constraints, we often cannot obtain all the spatial data of interest. In the context of unavoidable constraints in data acquisition, the astute exploitation of extant spatial data assets for the purpose of precise predictive analytics has emerged as a pervasive and critical undertaking. The usual methods for completing this prediction task are regression or interpolation. Prediction tasks often serve as an important data source for other downstream tasks (such as urban planning or related analyses in the city), ensuring the accuracy of such generated data becomes particularly necessary. In order to enhance the accuracy of data diffusion and prediction, it is necessary to construct an effective spatial modeling method that can reveal the structure and characteristics of spatial data and accurately predict unknown data.

Traditional data differs from spatial data, which possesses unique characteristics. Spatial data is characterized by spatial dependence (spatial autocorrelation) and spatial heterogeneity (spatial structure) [4]. Spatial autocorrelation implies that each data point is associated with other data points, but the correlation between neighboring data points is usually much stronger than that between distant data points [5]. Spatial prediction tasks are largely built upon spatial autocorrelation and have developed methods such as regression and interpolation. In spatial modeling tasks, spatial interpolation is the most common approach. Interpolation is

The associate editor coordinating the review of this manuscript and approving it for publication was Jiju Poovvancheri.

an important method for handling missing data. However, it also has some limitations. Spatial interpolation often makes strong assumptions about the distribution of data and lacks the ability to handle data with different distributions simultaneously, making it less flexible. Furthermore, with the increasing popularity of machine learning algorithms and the significant similarity between spatial data and graph data, researchers have started to apply Graph Convolutional Networks (GCNs) to the processing and analysis of spatial data [6], [7]. Currently, the Kriging Convolutional Networks (KCN) model is considered a superior spatial interpolation method as it combines existing Kriging interpolation methods with GCN approaches [8].

Nevertheless, existing methods often overlook spatial heterogeneity within geographical data. Spatial heterogeneity refers to the phenomenon of differences or diversities that exist between different regions or locations in a geographical space. This heterogeneity can manifest as gradients, fluctuations, and non-uniform distributions. Notably, we have observed that spatial heterogeneity is relatively under considered in spatial modeling, potentially affecting the predictive capability of spatial data. Most existing spatial models utilize k-nearest neighbors (kNN) algorithms for composition training based on Euclidean distance. In data space, Euclidean distance is typically defined as the straight-line distance between two points. Based on this definition, a relatively small set of spatial points is often considered to have more similar features. However, this assumption does not correspond with real-life applications. Our living spaces are often obstructed rather than being flat, which means that using straight-line distance to measure similarity between spatial points is not accurate [9]. Therefore, our work aims to address this gap and enhance the precision of spatial prediction models.

To overcome the aforementioned limitations, we propose a regional feature aggregation module that leverages a substantial amount of crowdsourced Points of Interest (POIs) data to obtain high-dimensional features for each region. By utilizing these high-dimensional features, the performance of downstream tasks related to regional prediction can be significantly enhanced. However, during the implementation phase, we encountered two primary challenges: (1) In GCN, edge information in the graph is typically used to represent spatial proximity or physical links between nodes. Nevertheless, the k-nearest neighbor graph constructed based on Euclidean distance may not accurately reflect the actual proximity relationship between nodes. (2) Balancing spatial distance and regional functions to better capture the spatial structure within the overall region remains a challenge. To address these issues, this study proposes new weight definition methods comprising four specific strategies: (1) Utilizing the definition of regional function, which consists of a series of multi-dimensional matrices, each dimension representing a functional value in one region of multiple regions, with all functions summing up to 1. (2) Selecting crowdsourced POIs data from the city as the functional function for each region. (3) Employing contrastive learning based on the Word2Vec method to negatively sample random walks on the POIs graph, resulting in single-point embeddings for POIs. Subsequently, through aggregation strategies, single-point POIs embeddings in each region are combined into regional embeddings, which serve as the desired regional functional functions. (4) Constructing a corresponding network based on the distance between regions and the similarity between regions, and employing GCN methods for geographical modeling. The proposed method was evaluated using real data from Chinese cities to achieve the following objectives:

1. Balance spatial distance and regional functions to provide accurate geographic information support.
2. Enhance the model's performance in complex geographical environments for better application in real life.
3. Offer more precise geographic modeling compared to existing prediction models, address the gaps in current research, and offer a new perspective for spatial data analysis in smart cities.

II. RELATED WORK

A. SPATIAL DATA PREDICTION MODEL

Spatial data modeling has a rich history, with numerous classic regression or interpolation methods having been applied in the field of geographic modeling for prediction tasks. Kriging is the most representative and best-performing method across various scenarios. Recently, there has been an increase in research on the application of GCNs in spatial modeling tasks. Compared to traditional spatial modeling methods, GCNs are typically used for tasks that graphically represent spatial data, offering advantages in terms of flexibility and scalability. In GCNs, explicit graphs are constructed based on the geographical coordinates of spatial data points, and the weights of corresponding edges are defined. Information is passed through these edges, and the weights typically represent the correlation between two points, which can be naturally combined with the autocorrelation of spatial data. By setting multiple hidden layers, GCNs can capture more complex relationships between feature values and target values. Compared to traditional methods, GCNs are more flexible [10]. Among them, two models worth noting are KCN and Positional Encoder Graph Neural Networks (PE-GNN).

The KCN model [8] has been previously discussed in the previous text. It is a highly effective GCN that combines the inductive benefits of traditional Kriging interpolation with the complex fitting capabilities of GCN models, resulting in superior performance. Specifically, KCN employs Convolutional Neural Networks (CNNs) to extract and learn features from spatial data, generating representations that retain spatial context and features. Subsequently, based on these learned features, Kriging interpolation is utilized for spatial interpolation and prediction to estimate attribute values at unidentified locations. The KCN model's design takes into account the characteristics of knowledge graphs and their integration with GCNs, aiming to enhance the efficiency and accuracy of GCNs in handling complex network-structured data. By utilizing kernel functions to determine the weights of GCNs, KCN achieves favorable outcomes compared

to baseline models. PE-GNN [11] directly employs the distance reciprocal edge weight definition method in the GCN component, yielding positive results in their respective domains.

B. REGIONAL FUNCTIONALITY REPRESENTATION THROUGH EMBEDDING OF POIs

In the realm of urban data modeling, crowdsourced data such as POIs have garnered significant attention due to their ease of access. In the context of urban modeling tasks, approaches based on regions or grids often encounter challenges related to small sample sizes and low feature dimensionality. To facilitate end-to-end modeling of urban data, POIs emerge as a highly valuable data source. POIs are intrinsically linked to group behavior [12] and the socio-economic facets [13] of urban areas, making them one of the most widely utilized sources of crowdsourced data currently. Furthermore, POIs are more readily available compared to other data sources (e.g., group mobility data), which are typically restricted to specific regions and user groups. Consequently, despite the emergence of new types of crowdsourced data, POIs continue to be a valuable and easily accessible resource for enhancing regional features.

Early research primarily utilized POI frequency as regional features [14], [15]. This approach aimed to learn low-dimensional latent vector embeddings for POI classification (e.g., shopping malls and restaurants) based on spatial co-occurrence information and a sampling strategy. In recent years, the concept of learning POI and regional representations has emerged, highlighting the limitations of traditional feature engineering methods. Word2Vec, a word embedding technique introduced by Mikolov for generating word vectors [16]. Yao et al. were the first to employ Word2Vec for learning POI category embeddings, with the goal of capturing co-occurrence patterns between POI categories within POI strings [17]. Building upon this foundation, Yan et al. proposed the Place2Vec model, which also focuses on capturing category co-occurrence patterns but utilizes kNN as its core method [18]. Subsequently, Niu and Silva employed the Doc2Vec model, a variant of Word2Vec, to embed POIs and regions, resulting in functional land use classification [19]. Huang et al. proposed a semantically preserved POI embedding method in 2022, which applied random walks in the POI network and class semantics using manifold learning algorithms [20]. Recently, Yang et al. proposed a novel model that combines POIs, Place2vec and Latent Dirichlet Allocation (LDA) to significantly enhance the classification accuracy of urban functional areas [21].

After POIs are embedded, aggregation operations are usually required. Among these operations, average pooling is widely used. However, Huang et al. suggested that POIs may have varying importance levels during the aggregation process. To address this issue, they employed long short-term memory (LSTM) and attention mechanisms [20]. Nevertheless, due to the limitations of their supervised training approach, it may not be suitable for most tasks. Therefore, exploring effective POI embedding in an unsupervised environment remains a worthwhile endeavor.

III. METHODOLOGY

This study proposes a novel approach for urban area modeling utilizing GCN. The fundamental unit of this approach is an urban block, denoted as G. Each block comprises central coordinates C, discrete POIs points classified at first, second, and third levels, longitude and latitude coordinates, and a small amount of feature data X. Additionally, the target value Y that needs to be predicted through regression is also incorporated. In the context of GCN, central coordinates C are utilized to establish the geographical location relationship between grids. Based on these relationships, we select the k closest area blocks G_i to the region block and establish an edge between G_i and G to obtain the corresponding graph data for grid data. Correspondingly, the feature values X represent the feature values of the graph nodes, and the target value Y is employed to predict the nodes on the graph.

To address the potential issues of inadequate feature values and subpar prediction performance, we propose a two-layer embedding approach for cities. The first layer is the POIs embedding, which draws inspiration from Word2Vec. This embedding process involves utilizing the coordinates of POIs to construct a Delaunay Triangulation (DT) graph encompassing all POIs within a city. Subsequently, biased random walks are conducted on the generated graph. The resulting sequence is then trained through a Skip-Gram model, which incorporates a negative sampling procedure and maximizes hierarchical information, ultimately yielding the embeddings for all secondary categories of POIs [20].

After obtaining the embeddings of secondary categories of POIs, we employ an unsupervised approach to aggregate all POIs within urban area block G, resulting in the embedding E_G for that block. This strategy may involve multi-head attention mechanisms or average pooling, among other techniques. The obtained block embedding not only enhances the feature dimensionality of the block but also serves as a functional representation function, playing a crucial role in the GCN backbone prediction model. By combining E_G with the original features X, we obtain a higher-dimensional and more informative feature vector. Furthermore, for grid graph data, the edge weight is defined by the product of the similarity between the functional embeddings of two areas and their distance. This approach addresses the precise capture problem of spatial heterogeneity data mentioned earlier.

A. THE MAIN BODY OF GCN IN THE MODEL

Suppose there is a data point i, which includes a feature vector x_i and a label y_i. These feature vectors and labels are collectively represented by $\tilde{X}$ and $\tilde{Y}$ respectively. Subsequently, a graph G is constructed with $G = (V, E)$. This graph comprises: A set of data nodes $V = \{v_1, v_2, \ldots, v_n\}$ which symbolize the data point i; A set of edges $E = \{e_1, e_2, \ldots, e_m\}$ that are determined by the adjacency matrix

A of the graph. The formula for the standardized adjacency matrix $\bar{A}$ is presented below:

$$\bar{A} = D^{-\frac{1}{2}}(A + I)D^{-\frac{1}{2}} \tag{1}$$

In this formulation, D is a diagonal matrix defined as $D = \text{diag}(A1 + 1)$, which means the degree matrix plus one. The symbol I represents an identity matrix. A single GCN layer ℓ comprises $\bar{A}, \tilde{X}, \tilde{Y}$, with $\bar{A}$ and $\tilde{X}$ serving as inputs and the known label $\tilde{Y}$ being utilized as the target for fitting. Each GCN layer ℓ accepts an input $H_{\ell-1} \in \mathbb{R}^{n\times d_{\ell-1}}$ and produces an output matrix $H_\ell \in \mathbb{R}^{n\times d_\ell}$. Additionally, each layer is parameterized by a weight matrix W^ℓ, which has dimensions $d_{\ell-1} \times d_\ell$. Consequently, a series of L GCN layers form the final GCN model. Formally, the GCN is defined as follows:

$$H^\ell = \sigma\left(\bar{A}H^{\ell-1}W^\ell\right), \ell = 1, \ldots, L \tag{2}$$

Here, H^ℓ represents the output of the ℓ-th layer in the GCN. $H^{\ell-1}$ denotes the output of the previous layer $\ell - 1$. σ is a non-linear activation function. The initial output of the GCN, $\tilde{X}$, comprises all node feature vectors and can be expressed as $H^0 = \tilde{X}$. Ultimately, the final output of the GCN is $\hat{Y} = H^L$.

In addition, the GCN determines its training loss based on the available labels and predicts the unknown labels corresponding to the entries in $\hat{Y}$. To apply the GCN for predicting the labels of other unknown nodes, a graph can be constructed based on the data point positions (i.e., geographical coordinates), with each feature vector and its label attached to the corresponding graph node. For predicting unknown points, only the corresponding value from $\hat{Y}$ is required [8].

The GCN [10] is composed of multiple graph convolutional layers, akin to the perceptron. However, it also incorporates a neighborhood aggregation step based on spectral convolution. At time t, layer (ℓ) receives the adjacency matrix A_t and node embedding matrix $H_t^{(\ell)}$ as input and employs the weight matrix $W_t^{(\ell)}$ to update the node embedding matrix as output. It can be expressed as follows:

$$\begin{aligned} H_t^{(\ell+1)} &= \text{GCONV}(A_t, H_t^{(\ell)}, W_t^{(\ell)}) \\ &:= \sigma(\tilde{A}_t H_t^{(\ell)} W_t^{(\ell)}) \end{aligned} \tag{3}$$

The notation $\tilde{A}_t$ is employed to denote the normalized version of A_t, which is defined as follows (omitting the time index for clarity):

$$\tilde{A} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}, \tilde{A} = A + I, \tilde{D} = \text{diag}\left(\sum\nolimits_j \tilde{A}_{ij}\right) \tag{4}$$

where σ is an activation function, typically ReLU, which is suitable for all layers except the output layer. The initial embedding matrix is derived from node features, i.e. $H_t^{(0)} = X_t$. Assuming there are (L) layers of graph convolution. For the output layer, the function σ can be considered as an identity function, in this case, $H_t^{(L)}$ contains high-level representations of nodes transformed from the initial features; or it could be a SoftMax function for node classification, in which case $H_t^{(L)}$ contains predicted probabilities. The GNN updates information through the following message passing mechanism:

$$\begin{aligned} &h_v^{(\ell+1)} \\ &= \text{UPDATE}^{(\ell)}\left(h_v^{(\ell)}, \text{AGGREGATE}^{(\ell)}\left(h_u^{(\ell)} : u \in \mathcal{N}(v)\right)\right) \end{aligned} \tag{5}$$

where $h_v^{(\ell)}$ denotes the feature representation of node v at layer (ℓ), $\mathcal{N}(v)$ is the set of neighbor nodes of node v, $\text{UPDATE}^{(\ell)}$ and $\text{AGGREGATE}^{(\ell)}$ represent the update and aggregation functions respectively, operating at layer (ℓ).

B. THE EMBEDDING OF POIs AND REGIONS

The method primarily comprises a secondary category encoder for POIs and an area feature representation aggregator. The secondary categories are sufficiently informative without being excessively numerous. Firstly, we construct a POI network and devise a spatial explicit Random Walk strategy to sample the co-occurrence information of POIs. Subsequently, we capture the hierarchical structure of POIs categories [20]. The captured co-occurrence information and category semantic information are fed into a POI encoder ϕ, which generates POI embeddings by simultaneously optimizing the objective of Skip-Gram [16] and Laplacian Eigenmaps [22].

1) THE PROCEDURE OF EMBEDDING POIs

In the given area, POIs are interconnected using DT network, which possesses several advantageous characteristics that contribute to the generation of a compact and concise network. This facilitates the learning of spatial vector data embeddings [23]. Subsequently, a spatially explicit Random Walk sampling strategy is employed to capture the category co-occurrence patterns of POIs. The core of this sampling strategy lies in designing the transition probability between two points in the random walk. This biased transition probability between nodes (POIs) combines spatial distance decay, balances local and long-distance co-occurrence patterns between nodes, and distinguish between internal and cross-regional co-occurrence. The aforementioned approach is accomplished through the establishment of three distinct types of transition biases, whose precise mathematical formulations are detailed below:

The first is an inverse-distance transition bias α_d:

$$\alpha_d(p_2, x) = \log\left(\frac{(1 + D^{1.5})}{(1+d_{p_2,x}^{1.5})}\right) \tag{6}$$

The second transition bias α_b aims to equalize the discrepancy between local and long-range co-occurrence of POIs:

$$\alpha_b(p_2, x) = \begin{cases} \alpha_b^{loc}, & \text{if } \text{hop}_{p_2,x} = 0 \\ 1, & \text{if } \text{hop}_{p_2,x} = 1 \\ \alpha_b^{glob}, & \text{if } \text{hop}_{p_2,x} = 2 \end{cases} \tag{7}$$

The third transition bias α_r is defined as:

$$\alpha_r(p_2, x) = \begin{cases} 1, & \text{if} \{p_2, x\} \subseteq P_i \\ \alpha_r^{inter-region}, & \text{if} p_2 \in P_i \text{ and } x \in P_k \text{ and } r_i \neq r_k \end{cases} \tag{8}$$

The parameter α_d ensures that nodes with closer spatial distances are assigned higher probabilities. Here, D denotes the diagonal length of the minimum bounding rectangle encompassing all POIs within the study area, while d represents the spatial distance between two nodes, p_2 and x. $\text{hop}_{p,x}$ represents the minimum number of hops required to move from p (the previous node) to candidate node x. In theory, the value of $\alpha_r^{inter-region}$ should be less than 1, indicating that it is more likely to sample co-occurrence information within a region.

Finally, in space-explicit random walk, the unnormalized transition probability from the current node p_2 to each candidate node x is given by:

$$\text{tp}\left(p_{2,x}\right) = \alpha_d\left(p_{2,x}\right) \times \alpha_b\left(p_{2,x}\right) \times \alpha_r\left(p_{2,x}\right) \tag{9}$$

Using the aforementioned sampling strategy, we execute several random walks starting from each node in the network and generate multiple sequences of POIs, where each POI is represented by its second-level category c_i. For each sampled sequence, the first category is designated as the target category, while the remaining categories are considered context categories. To optimize the embedding of each POI second-level category, we employ a Skip-Gram neural network with negative sampling. This approach needs to minimize the objective function as follows:

$$\begin{aligned}\mathcal{L}_{co-occurrence} &= \sum_{c\in C^2}\sum_{c_q\in N_{R(c)}} \\ &\quad -\log\left(\frac{\exp\left(\boldsymbol{c}^T\boldsymbol{c}_q'\right)}{\sum_{c_n\in C^2}\exp\left(\boldsymbol{c}^T\boldsymbol{c}_n'\right)}\right) \\ &\approx \sum_{c\in C^2}\sum_{c_q\in N_{R(c)}} -\left(\log\left(\sigma\left(c^T c_q'\right)\right)\right. \\ &\quad \left.-\sum_{i=1}^{k}\log\left(\sigma\left(c^T c_{n_i}'\right)\right)\right)\end{aligned} \tag{10}$$

In this context, σ represents the sigmoid function. The symbol c denotes both the target embedding and the vector embedding of the second-level category, while c' represents the context embedding. $N_{R(c)}$ refers to the set of context categories captured by c during a random walk. The first line presents the original form of the Skip-Gram objective function, and the second line provides an approximation. To address the high computational cost, we opt for this approximation, where c_{n_i} signifies the category obtained through a negative sampling process (meaning that c_{n_i} does not co-occur with c) [20].

2) THE AGGREGATION YIELDS REGIONAL EMBEDDINGS

After obtaining the POIs category embeddings, it is crucial to aggregate them in a meaningful manner to generate regional embeddings. During this process, it is essential to consider the varying importance of each POI in defining its region, which typically necessitates multiple perspectives. Our objective is to learn regional representations that can serve multiple purposes, necessitating the simulation of this multifaceted impact. There are several methods for obtaining regional embeddings for all POIs points in a region. The more traditional approach involves using the average pooling of POIs embeddings, which can yield good results in unsupervised scenarios.

Additionally, the multi-head attention mechanism can also be employed for POIs aggregation, as developed by the *Transformer* [24], [25]. Specifically, this mechanism enables the attention operation to capture the relative relevance (importance levels) between two sets of entities, which in this case refers to the significance of each POI in defining its region from a singular perspective. The attention function maps the query vector Q to the output using key-value pairs K and V.

$$\text{Att}\left(Q, K, V\right) = \alpha\left(QK^T\right)V \tag{11}$$

In the (11), $Q \in \mathbb{R}^{n_v\times d_q}$, $K \in \mathbb{R}^{n_v\times d_k}$, $V \in \mathbb{R}^{n_v\times d_v}$. The scaling factor in the SoftMax function is denoted by α, which is calculated as $\alpha\left(o\right) = \text{softmax}(\frac{o}{\sqrt{d}})$. Initially, Q, K and V are projected onto h independent vectors with dimensions $d_k = d_q = d_v = \frac{d}{h}$, commonly referred to as multi-heads. Subsequently, an attention mechanism is applied to these h projections, resulting in a linear transformation and combination of all attention outputs:

$$\text{Multihead}\left(Q, K, V\right) = \text{concat}(O_1, O_2, \ldots, O_h)W^O \tag{12}$$

where the calculation of each O_j is achieved through the application of $\text{Att}(QW_j^Q, KW_j^K, VW_j^V)$, where $\left\{W_j^Q, W_j^K, W_j^V\right\}_{j=1}^{h}$ are learnable parameters. Furthermore, as per *set transformer* [25] an aggregation function $\text{AGG}_{\text{poi}-\text{region}}$ can be defined using the multi-head attention mechanism as follows:

$$\text{AGG}_{\text{poi}-\text{region}}\left(P_i\right) = H + \text{rFF}(H) \tag{13}$$

$$H = \vec{s}_i + \text{Multihead}\left(\vec{s}_i, P_i, P_i\right) \tag{14}$$

in which P_i is the embeddings of the POIs in a region r_i;$\vec{s}_i$, a randomly initialized and learnable seed vector, is used to compute the attention weights for the POI embeddings in the region r_i; The function rFF is a linear transformation that is subsequently followed by a ReLU activation function. This methodology produces the output of $AGG_{poi-region}$, which is the raw embedding $\vec{\boldsymbol{rw}}_{\boldsymbol{i}}$ for a region, representing the aggregating of embeddings of POIs within that region. The resulting $\vec{\boldsymbol{rw}}_{\boldsymbol{i}}$ captures varying levels of importance of POIs from multiple viewpoints, with each head representing a distinct perspective [26].

In this research, the Skip-Gram model is trained by maximizing the log-likelihood of the contextual embeddings of POIs. More specifically, the objective function for model training is the cross-entropy loss, which can be formulated as follows:

$$L\left(\theta\right) = -\sum_{(w,c)\in D}\log p(c|w;\theta) \tag{15}$$

where $L\left(\theta\right)$ denotes the loss function, D represents the set of all positive sample pairs (w, c), w is the target word, c is the context word and θ represents the model parameter.

C. DETAILS AND TRAINING OF MODEL

1) THE ARCHITECTURE OF MODEL

Our model employs a multi-layer GCN with the following specific structure:

Input layer: This layer receives the feature matrix X of nodes and the normalized adjacency matrix $\tilde{A}$.

Hidden layers: These layers consist of multiple GCN layers, where each layer applies the ReLU activation function and includes Batch Normalization to enhance training stability.

Output layer: The final GCN layer outputs the region embeddings, which are then passed through a SoftMax layer for multi-classification.

2) THE TRAINING PROCESS OF MODEL

To prevent overfitting, L2 regularization terms are incorporated into the loss function:

$$\mathcal{L}_{reg} = \mathcal{L}_{original} + \lambda \sum_{w} \|w\|^2 \tag{16}$$

where $\mathcal{L}_{original}$ is the original loss function, w represents the weight of the model, and λ is the regularization coefficient.

Our model is trained using the Adam optimizer with an initial learning rate of 0.001, a batch size of 32, and a total number of 100 epochs. To prevent overfitting, we also employ L2 regularization with a regularization parameter λ set to 0.0001. After each epoch, we assess the model's performance on the validation set and adjust the learning rate to avoid premature convergence. The enhancement in model performance is based on the reduction in loss on the validation set. If there is no further reduction in loss for 10 consecutive epochs, we halt training early. We list the algorithm as follows.

Algorithm 1 POIs Embedding

Input: POI dataset: $C_{POI}, FC_{POI}, SC_{POI}, TC_{POI}$
Output: Embedding of SC_{POI}
Function 1: DTG(d) → return graph data
Function 2: Rw(G) → return random walk sequence
Function 3: Skip-gram(s) → return embedding of SC_{POI}
G ← DTG(POI Dataset)
Si ← Rw(G)
for number of training steps(epochs) **do**
  for all s in Si **do**
  Y ← Skip-Gram(s)
  Compute loss
  end for
end for
return Skip-Gram(SC_{POI})

Algorithm 2 POIs-GNN Training

Input: Grid dataset: **C**, **X**, **Y**
POI second-level embedding PE
Output: Model predicted value **Y'**
Function: Aggregation-function (PE) → return grid embedding
Initialize model
Set hyper-parameter
for all grid in grid dataset
  for all POI in grid
    x′= aggregation (PE)
    x′= concat [x′; x]
  end for
  Construct a spatial graph with **C** using k-nearest neighbors
  Predict outcome **Y**′
  Compute loss (Y',Y)
  Update the parameters
end for
return Y'

IV. EXPERIMENTS

A. EXPERIMENTAL SETUP

In addition to the spatial modeling technology utilized in ecological construction and urban planning scenarios mentioned above, data generalization is crucial when delving into the factors influencing commodity consumption markets. For instance, the cigarette consumption market is influenced by numerous factors, encompassing not only substantial amounts of internal industry data such as the number of retailers and consumption rates but also external data sets that impact the market, including consumer demographics and holiday crowd flow. These data sets are often dynamic and subject to change, with variations in distribution occurring over time, geography, or other variables. Occasionally, this may result in missing data. Consequently, out-of-distribution generalization is essential for handling such data. In this study, we opted for a dataset pertinent to cigarette consumption for model evaluation purposes.

Datasets: We selected the cigarette dataset from Chongqing city for our analysis. During the initial data processing stage, we divided Chongqing into grid areas of 750 × 750 meters, resulting in a total of 5836 grid data points. Each grid data set encapsulates information pertaining to retail customers. This division was primarily influenced by the distance that consumers travel to purchase cigarette products, ensuring that adjacent grids exhibit both similarities and differences. The dataset for each grid (buss-id) comprises the following information:

1) **C**: This represents the grid coordinates selected from the center of the grid, which contain longitude and latitude information. As a result, it is a two-dimensional dataset with a total of 5836∗2 data points.
2) **X**: This represents the original data of the grid, including three dimensions of cigarette sales revenue, number of cigarette vendors, and POIs in the grid. Therefore, this dataset contains a total of $5836 * 3$ data points.
3) **Y**: This indicates the predicted target value of the model, reflecting the population in the grid. It only contains one type of information, thus it is a one-dimensional dataset with a total of 5836 data points.

In addition, a total of 612,078 POIs data were incorporated. Each POI record encompasses the geographical coordinates (latitude and longitude) of the POI, along with its first-level,

second-level, and third-level classifications. During the experiment, we partitioned the entire dataset into training, validation, and testing sets in a ratio of 6:2:2. This division was carried out for the purpose of training and evaluating the entire model.

Models: The calculation process of our proposed approach is detailed below. Firstly, urban grid data is selected, including longitude and latitude coordinates, feature values X, target values Y, and a large number of POIs. Secondly, DT is used to form a graph data of POIs points, where POIs are represented by their second-level category. Thirdly, a random walk process is conducted on the POIs graph to obtain corresponding sequences, which are then used to train the second-level category embedding of POIs through the Skip-Gram model. After obtaining the POIs embedding, it is aggregated in each grid to obtain the grid embedding, which is combined with the initial feature X to obtain enhanced feature. Finally, all data is integrated and the GCN is used for overall data training and fitting to obtain the final prediction results.

This study aims to investigate the utilization of POIs data embedding to enhance grid features and improve the performance of GCN models. To accomplish this, we have selected three fundamental GCN models, namely KCN, W-GCN, and F-GCN, as the research subjects. By comparing various embedding models and their strategies while employing the same dataset and evaluation criteria for comparison, we ensure the objectivity and credibility of the results. Finally, we will determine which method performs best in practical applications based on the experimental outcomes and provide valuable references for future research and enhancement.

KCN: The calculation process of the model is as follows. First, a dataset is selected, and the kNN algorithm is employed along with coordinates to generate edges and their corresponding weights. Second, the Radial Basis Function (RBF) kernel function is utilized to process the adjacency matrix. The RBF kernel function is a widely used technique that can map data from a low-dimensional space to a high-dimensional space, thereby making the data more linearly separable in the higher-dimensional space. Finally, the model is trained using the processed adjacency matrix and feature inputs [8]. In this study, the k-nearest neighbors of the processing point are selected, and their weights are defined according to (17):

$$A_{jk} = \exp\left(-\frac{1}{2\varphi^2} \parallel c_j - c_k \parallel_2^2\right), \forall j, k \in \beta_i \qquad (17)$$

where φ denotes the kernel length, which is a hyperparameter. The square term represents the square of the distance between two points. $\beta_i = \{i\} \cup \alpha_i$ signifies the set encompassing data point i and its neighbors, where α_i denotes the neighbor set of data point i in the training set. In this graph, when j and k are adjacent to each other, edge (j, k) possesses a greater weight, and vice versa.

W-GCN: The W-GCN is an optimization technique designed to enhance the original GCN model. Its fundamental principle involves employing a learnable weight matrix to adjust the embedding dimension. In the W-GCN framework, each layer's input is intricately connected to the weight matrix W. To process a data point, its k-nearest neighbors are identified, and the corresponding edge weights are established as the inverse of the distance between the two points. The weighted GCN can be mathematically expressed as (2) in the preceding text, which is based on the GCN.

F-GCN: F signifies the regional function. In this approach, identify the k points that are closest to the current processing point and assign edge weights based on the inverse of the distance between these points, multiplied by their cosine similarity in terms of feature vectors. The other calculation processes align with the principles of W-GCN.

B. EVALUATION METRICS

In this research, we employ the following evaluation metrics to gauge the precision of model predictions. These metrics, commonly referred to as evaluation coefficients, are primarily utilized to quantify the performance and efficacy of the model. By leveraging these evaluation metrics, we can gain a more comprehensive understanding of the predictive capability, fitting accuracy, and stability of the model.

(1) RMSE. The Root Mean Square Error is a widely adopted metric in regression tasks to assess the accuracy of predictive models on continuous data. It calculates the square root of the mean of the squared differences between predicted and actual values, providing an indication of the average deviation between these two sets of values. The smaller the RMSE value, the closer the predicted and actual values are, suggesting a stronger predictive ability of the model.

(2) MAE. The Mean Absolute Error is a commonly adopted evaluation metric in regression predictive models, which quantifies the discrepancy between predicted and actual values. In essence, MAE computes the average of the absolute deviations between predicted and actual outcomes. A lower value of this metric signifies a reduced difference between predicted and actual values, suggesting that the model exhibits superior predictive capabilities.

C. ENHANCING FEATURE REPRESENTATION THROUGH REGIONAL EMBEDDING TECHNIQUES

This study aims to enhance the accuracy of predictions on grid data in urban areas. To accomplish this objective, we conducted comparative experiments utilizing multiple GCN models and incorporated diverse embedding strategies for POIs data embedding within these models. Given that the feature dimension of the grid is relatively low and only approximately 5000 grids have data, while there are over 600,000 point of POI data points, organizing the vast number of POIs within the grid becomes a crucial aspect of data processing and feature extraction. In light of these considerations, we employ strategies outlined in Section III to enhance regional feature organization. Our specific aggregation strategies encompass attention mechanisms and embedding combined with pooling.

TABLE 1. The results of evaluation metrics.

Model	ChongQing	
	RMSE	MAE
KCN		
Raw Data	8425.30	4348.08
Embedding+Pooling	8232.03	4660.69
Embedding+Attention	8602.55	4132.10
W-GCN		
Raw Data	7706.36	3665.02
Embedding+Pooling	6725.75	3622.58
Embedding+Attention	6840.86	**3148.95**
F-GCN		
Raw Data	8981.16	4969.71
Embedding+Pooling	**6711.45**	3786.30
Embedding+Attention	8058.24	4224.85

Note: The best one is shown in bold.

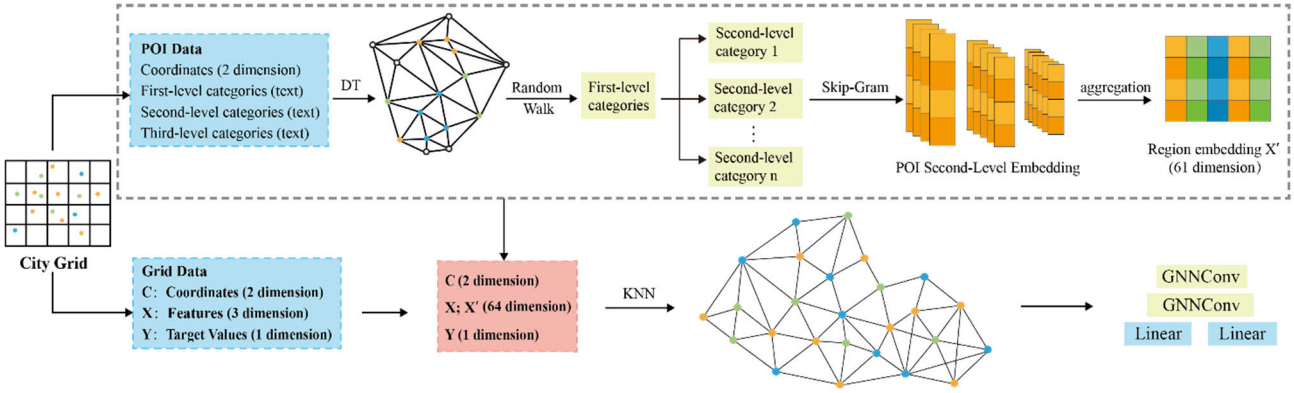

FIGURE 1. The primary framework of the methodology.

The original method utilized raw data for predictions. To investigate the optimal combination of embedding model and embedding strategy, we conducted comparative experiments by comparing this method with the results of POIs data embedding under different embedding strategies. The experimental findings are presented in Table 1. It can be observed that, in terms of the RMSE metric, the strategy of employing F-GCN model-based POIs embedding with pooling achieves the best performance, improving by 12.91% compared to the W-GCN model, which had the highest performance among the raw data. In terms of the MAE metric, the strategy of using W-GCN model-based POIs embedding with attention mechanism achieves the best performance, improving by 14.08% compared to the W-GCN model, which had the highest performance among the raw data. This indicates that our joint approach is effective.

D. PERFORMANCE CHANGES WITH DIFFERENT NUMBER OF NEIGHBORS

This study also investigates the impact of varying the number of neighbors on the performance of KCN, W-GCN, and F-GCN during graph construction. The results are presented in Fig. 2, which illustrates the variations in RMSE and MAE for these three models under both the original method and the POIs embedding methods with attention mechanism and pooling. The original method's W-GCN model consistently outperforms other models, as demonstrated in Fig. 2a, d, regardless of the value of k. Conversely, when employing pooling strategy with POIs embedding, the F-GCN model's performance becomes optimal as the value of k increases (Fig. 2b, e). Similarly, the W-GCN model consistently surpasses the other two models when using attention mechanism strategy with POIs embedding (Fig. 2c, f). This suggests that the F-GCN model is more suitable for employing the

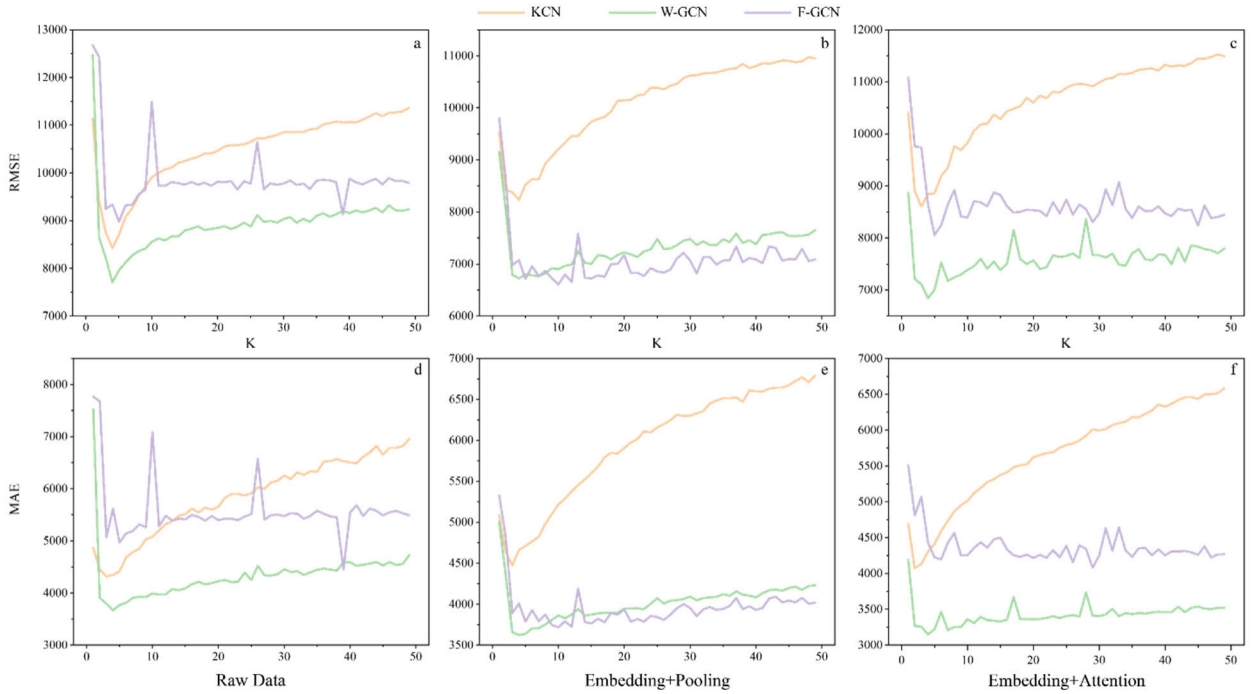

FIGURE 2. RMSE and MAE over k.

pooling strategy when embedding POIs data, while the W-GCN model is more appropriate for utilizing the attention mechanism strategy.

V. CONCLUSION

This study introduces a novel approach that leverages GCN and diverse POIs data embedding techniques to enhance urban regional features. We employ crowdsourced POIs data for feature embedding and explore various embedding models and strategies to optimize the predictive accuracy and computational efficiency of the model. In contrast to conventional models in this field, our experimental results suggest that our proposed approach demonstrates superior accuracy in spatial data prediction. This implies that our model can furnish dependable prediction outcomes in practical applications, which is of paramount importance for various domains such as urban planning and traffic management. This research provides robust support for addressing practical issues and serves as a valuable reference for further refining and advancing spatial data prediction methodologies.

REFERENCES

[1] J. T. Thorson, C. L. Barnes, S. T. Friedman, J. L. Morano, and M. C. Siple, "Spatially varying coefficients can improve parsimony and descriptive power for species distribution models," *Ecography*, vol. 2023, no. 5, p. 13, Apr. 2023.

[2] Y. Wang, H. Sun, Y. Zhao, W. Zhou, and S. Zhu, "A heterogeneous graph embedding framework for location-based social network analysis in smart cities," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2747–2755, Apr. 2020.

[3] H. Wu, Z. Gui, and Z. Yang, "Geospatial big data for urban planning and urban management," *Geo-spatial Inf. Sci.*, vol. 23, no. 4, pp. 273–274, Oct. 2020.

[4] L. Zhang, J. Cheng, and C. Jin, "Spatial interaction modeling of OD flow data: Comparing geographically weighted negative binomial regression (GWNBR) and OLS (GWOLSR)," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 5, p. 220, May 2019.

[5] N. Karasiak, J.-F. Dejoux, C. Monteil, and D. Sheeren, "Spatial dependence between training and test sets: Another pitfall of classification accuracy assessment in remote sensing," *Mach. Learn.*, vol. 111, no. 7, pp. 2715–2740, Jul. 2022.

[6] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[7] X. Zhang, Y. Ge, F. Ling, J. Chen, Y. Chen, and Y. Jia, "Graph convolutional networks-based super-resolution land cover mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7667–7681, 2021.

[8] G. Appleby, L. Liu, and L. P. Liu, "Kriging Convolutional Networks," presented at the AAAI, Honolulu, HI, USA, 2020.

[9] H. Zhu, X. Yang, B. Wang, and W.-C. Lee, "Range-based nearest neighbor queries with complex-shaped obstacles," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 5, pp. 963–977, May 2018.

[10] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," presented at the ICLR, 2017.

[11] K. Klemmer, N. Safir, and D. B. Neill, "Positional encoder graph neural networks for geographic data," presented at the Mach. Learn., 2021.

[12] Z. B. Sojahrood and M. Taleai, "A POI group recommendation method in location-based social networks based on user influence," *Expert Syst. Appl.*, vol. 171, Jun. 2021, Art. no. 114593.

[13] S. G. Wang, "Point of interest recommendation for social network using the Internet of Things and deep reinforcement learning," *Math. Biosciences Eng.*, vol. 20, no. 9, pp. 17428–17445, 2023.

[14] S. Jiang, A. Alves, F. Rodrigues, J. Ferreira, and F. C. Pereira, "Mining point-of-interest data from social networks for urban land use classification and disaggregation," *Comput., Environ. Urban Syst.*, vol. 53, pp. 36–46, Sep. 2015.

[15] Y.-S. Lu, W.-Y. Shih, H.-Y. Gau, K.-C. Chung, and J.-L. Huang, "On successive point-of-interest recommendation," *World Wide Web*, vol. 22, no. 3, pp. 1151–1173, May 2019.

[16] T. Mikolov, L. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 2, Dec. 2013, pp. 3111–3119.

[17] Y. Yao, X. Li, X. Liu, P. Liu, Z. Liang, J. Zhang, and K. Mai, "Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model," *Int. J. Geographical Inf. Sci.*, vol. 31, no. 4, pp. 825–848, 2017.
[18] B. Yan, K. Janowicz, G. Mai, and S. Gao, "From ITDL to Place2Vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts," presented at the 25th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst., Redondo Beach, CA, USA, 2017, doi: 10.1145/3139958.3140054.
[19] H. Niu and E. A. Silva, "Delineating urban functional use from points of interest data with neural network embedding: A case study in Greater London," *Comput., Environ. Urban Syst.*, vol. 88, Jul. 2021, Art. no. 101651.
[20] W. Huang, L. Cui, M. Chen, D. Zhang, and Y. Yao, "Estimating urban functional distributions with semantics preserved POI embedding," *Int. J. Geographical Inf. Sci.*, vol. 36, no. 10, pp. 1905–1930, Oct. 2022.
[21] X. Yang, Y. Yang, and X. Zheng, "Classifying urban functional zones by integrating POIs, place2vec, and LDA," *J. Urban Planning Develop.*, vol. 149, no. 4, Dec. 2023, Art. no. 04023034.
[22] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," presented at the 14th Int. Conf. Neural Inf. Process. Syst., Natural Synth., Vancouver, BC, Canada, 2001.
[23] X. Yan, T. Ai, M. Yang, and H. Yin, "A graph convolutional neural network for classification of building patterns using spatial vector data," *ISPRS J. Photogramm. Remote Sens.*, vol. 150, pp. 259–273, Apr. 2019.
[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," presented at the 31st Int. Conf. Neural Inf. Process. Syst., Long Beach, CA, USA, 2017.
[25] J. Lee, Y. Lee, J. Kim, A. R. Kosiorek, S. Choi, and Y. W. Teh, "Set transformer," in *Proc. Mach. Learn. Res.*, 2018, pp. 3744–3753.
[26] W. Huang, D. Zhang, G. Mai, X. Guo, and L. Cui, "Learning urban region representations with POIs and hierarchical graph infomax," *ISPRS J. Photogramm. Remote Sens.*, vol. 196, pp. 134–145, Feb. 2023.

CHAO DENG received the M.S. degree in internet computing from Edith Cowan University, Australia, in 2008, and the Ph.D. degree in agricultural information technology from China Agricultural University, China, in 2018. He is currently a Senior Engineer with China Tobacco Industry. His research interests include big data technology, artificial intelligence technology, and data visualization.

XUEXIA LIANG received the M.S. degree in computer science and technology from Guangxi University, China, in 2020. She is currently an Artificial Intelligence Researcher with China Tobacco Guangxi Industrial Company Ltd. She is an Engineer with China Tobacco Industry. Her research interests include data visualization, knowledge graphs, and artificial intelligence.

XU YAN is currently pursuing the bachelor's degree in computer science and technology with Sichuan University. His current research interests include GNN and knowledge graphs.

YUHUA MO received the M.S. degree in software engineering from the University of Electronic Science and Technology of China, in 2020. She is currently an Algorithm Engineer with China Tobacco Guangxi Industrial Company Ltd. Her current research interests include consumption data analysis, spatio-temporal data management, and social network knowledge discovery.

SEN BAI received the B.S. degree in agriculture from Henan Agricultural University, China, in 2001, and the M.S. degree in agricultural extension from Hunan Agricultural University, China, in 2013. He is currently an Agronomist with China Tobacco Industry. His main research interests include enterprise digital transformation, digitalization of agriculture, business intelligence, and application of big data technology.

BIN LU received the M.S. degree in software engineering from Guangxi University, China, in 2021. He is currently an Algorithm Engineer with China Tobacco Guangxi Industrial Company Ltd. His research interests include sentiment analysis, knowledge graphs, and machine learning.

KAIDI CHEN received the B.S. degree in software engineering from Fuzhou University, China, in 2017. She is currently a Data Analyst with China Tobacco Guangxi Industrial Company Ltd. Her current research interests include spatio-temporal data mining and sales forecasting.

XIPENG LIU received the M.S. degree in computer science and technology from South China Normal University, in 2021. He is currently a Data Analysis Engineer with China Tobacco Industry. His research interests include machine learning, data mining, and data analysis for consumers.

ZHI CHEN received the M.S. degree in computer technology from Dalian Minzu University, in 2020. He is currently a Data Analysis Engineer with China Tobacco Industry. His research interests include data governance, data mining, and data analysis for consumers.