TYPE Original Research
PUBLISHED 08 October 2025
DOI 10.3389/fpsyg.2025.1641623

Predicting location emotions of users considering multidimensional spatio-temporal dependencies

Wei Jiang[1,2], Yiming Wang[1,2], Xiaoqing Song[1,2], Xinyue Zheng[1], Xiang Liu[1,2], Yi Long[2]*, Zuo Wang[1] and Ziran Wei[1]

[1]Anhui Normal University, Wuhu, China, [2]Nanjing Normal University Key Laboratory of Virtual Geographic Environment Ministry of Education, Nanjing, China

Check for updates

OPEN ACCESS

EDITED BY
Shuai Zhao,
Nanyang Technological University, Singapore

REVIEWED BY
Agustinus Bimo Gumelar,
Widya Mandala Catholic University Surabaya, Indonesia
Haobing Liu,
Ocean University of China, China

*CORRESPONDENCE
Yi Long
✉ yilong.njnu@hotmail.com

RECEIVED 05 June 2025
ACCEPTED 22 September 2025
PUBLISHED 08 October 2025

CITATION
Jiang W, Wang Y, Song X, Zheng X, Liu X, Long Y, Wang Z and Wei Z (2025) Predicting location emotions of users considering multidimensional spatio-temporal dependencies.
Front. Psychol. 16:1641623.
doi: 10.3389/fpsyg.2025.1641623

Emotion has significant spatio-temporal characteristics, and predicting the spatio-temporal changes in emotion is an important premise for monitoring the emotional state of urban residents. Most prediction methods focus on the prediction of emotion in time series without considering the spatial properties of emotion. Based on geotagged image data on the Weibo platform from Shanghai, a user location emotion prediction method that considers multidimensional spatio-temporal dependencies between different emotional states is proposed in this paper. The method introduces the HiSpatialCluster algorithm to identify the users' stay area. Then, the FaceReader algorithm is applied to determine the emotional quadrant of users from image data, and a graph embedding algorithm is employed to obtain the feature vector representing each stay area. Finally, an attention-based BiLSTM method is applied to construct the multidimensional spatio-temporal dependencies of emotion for prediction. Experiments on the Weibo dataset show that the prediction accuracy of location emotion reaches 75%, which is better than that of the single LSTM and CNN method. The results of this paper can not only deepen the understanding of the spatio-temporal variation patterns of emotion but also optimize location-based recommendation services.

KEYWORDS

location emotion, spatio-temporal emotion prediction, multidimensional spatio-temporal dependencies, attention-based BiLSTM, graph embedding

1 Introduction

Emotion determines how people view the past, present, and future (Zhu and Gao, 2015). It is the distinguishing characteristic between humans and animals. Emotion is a psychological and physiological condition caused by the interaction of feelings, thoughts, and actions. It is distinguished by diversity and variability in spatio-temporal processes. Location emotion prediction refers to predicting the next spatial location of individuals or groups and their corresponding emotional states. Location emotion prediction is an emerging direction in the field of emotional computing and an important premise for monitoring the emotional state of urban residents.

Existing location prediction research is mainly based on big location data. The majority of these studies calculate the next location of people by fully investigating the spatio-temporal change pattern. In the early stage, studies mostly utilized statistical probability models to build spatio-temporal frameworks and then predicted the next location. For example, Ying et al. (2011) introduced a clustering-based model to predict the next location of a user's movement. They analyze the geographical and semantic features of user trajectories to improve the prediction accuracy. Early methods have difficulty addressing spatio-temporal dependencies. Spatio-temporal dependencies refer to the relationship between data that are separated by different time periods and spatial distances. Deep learning methods provide effective approaches for constructing spatio-temporal dependencies. Some scholars have started to apply methods such as long short-term memory neural networks (LSTM) and conventional neural networks (CNN) to predict the next location. Zhang et al. (2019) proposed a CNN-LSTM network for location prediction; this method constructs spatio-temporal dependencies between trajectory data to predict the destination of the users. Existing studies show that a deep learning framework can effectively build spatio-temporal dependencies through time steps and gating mechanisms. Most existing prediction methods achieve a high level of accuracy. However, they still lack the consideration of the emotional attributes of users at the corresponding locations.

Emotion calculation is currently popular research in the field of computing. Most present emotion calculation approaches are based on machine learning. These methods can accurately extract the numerical features of the data and their contextual association. Yolcu et al. (2019) used multilayer convolutional neural networks to extract features of eyebrows, eyes and mouth regions in face images and achieved good prediction results on the RaFD dataset. Current methods can predict dynamic changes in emotional states in time series with a high accuracy. However, to our knowledge, there are no studies focusing on predicting emotion in different locations. Existing prediction methods have difficulty constructing multidimensional spatio-temporal dependencies, which is the key to location emotion prediction.

To address the above research problem, an attention-based BiLSTM method is proposed in this paper for predicting the target regions that users may visit and their corresponding emotional state. Compared with the existing methods, the proposed method achieves an accuracy of 75.21% in location emotion prediction. Our method can accurately and effectively predict location emotion. The results can not only deepen the understanding of the spatio-temporal change pattern of emotion but also provide an important theoretical basis for optimizing location-based personalized services.

2 Related works

2.1 Spatio-temporal location prediction

With the development of internet and mobile communication technology, the daily activities of urban residents have generated a large amount of spatio-temporal location data. Examples include social media data and cell phone data. The massive location data provide the possibility for prediction research. The existing location prediction methods can be broadly classified into shallow learning-based and machine learning-based prediction methods (Li et al., 2021).

Shallow learning methods mainly combine Markov chains (Douc et al., 2018) or Bayesian networks (Ben-Gal, 2008) to predict the location. Pavlovic et al. (2000) proposed a switching linear dynamical system model. The method is based on motion capture data to learn human motion characteristics for synthesis, classification and tracking of human motion to predict human trajectories. The results showed superiority over the traditional hidden Markov model (Eddy, 1996). To further improve the prediction accuracy, Kooij et al. (2014) developed a context-based dynamic Bayesian network model. The model integrates pedestrian situational awareness, situational criticality, and the spatial layout of the environment as potential states into the switching linear dynamical system. Thus, better prediction results than the switching linear dynamical system are obtained. Wang H. et al. (2019a) and Wang P. et al. (2019b) proposed a multi-order fusion Markov model to improve the algorithm. Based on the structural changes in the original trajectory, the model clusters the feature points to identify important locations and demonstrates superior predictive accuracy. However, the features extracted using shallow methods are insufficient. These methods also lack environmental semantic information and are complex to construct models. The final prediction results often deviate from the actual situation, and it is difficult to accurately predict spatio-temporal location.

In recent years, with the rapid development of machine learning, spatio-temporal location prediction methods based on deep learning have been rapidly developing. LSTM neural networks are often used to construct temporal dependencies to predict locations. Wang H. et al. (2019a) and Wang P. et al. (2019b) proposed a hybrid Markov model for location prediction. The model combines the Markov model and LSTM to capture the dependencies among location sequences to predict users' locations. With the emergence of the attention mechanism, Karatzoglou et al. (2018) explored the effect of the attention mechanism on the model performance. The model uses the attention mechanism to extend the LSTM to predict locations, and the results show superiority over the traditional single-layer LSTM model. Xue et al. (2018) proposed a hierarchical LSTM model consisting of three scales. The model captures person-, social- and scene-level information using three LSTMs and obtains better prediction results. In addition to LSTM methods, CNNs are also basic neural networks that extract spatial features for location prediction. Niu et al. (2019) proposed the L-CNN model to predict taxi-passenger location. The model utilizes CNN to extract spatial dimension features and shows good performance with regards to the RMSE values. To further improve the prediction accuracy, Fan et al. (2018) adopted both CNN and bidirectional LSTM networks to predict the next location of vehicles. The method models periodic patterns and dynamic features of vehicle trajectories and shows superiority over several existing methods. Compared with shallow learning methods, deep learning-based models can effectively extract spatio-temporal-dependent information and achieve better prediction results. Substantial progress has been made in the field of multimodal spatio-temporal prediction, where models have integrated various data types like text, images, and location to improve prediction accuracy (Ho and Hui Lim, 2022; Halder et al., 2021). Han et al. (2014) notably enhanced location prediction accuracy by combining geographic reference information in text. Zhou et al. (2024) integrated multimodal spatio-temporal data, including traffic, text, and Points of Interest (POI), and proposed a comprehensive spatio-temporal prediction framework that significantly improves existing traffic prediction methods. Transformer-based models have become increasingly popular in location prediction due to their capacity to capture complex long-range dependencies (Sang et al., 2024; Saxena and Cao, 2022). Yu et al. (2020) proposed a Transformer-based approach for modeling crowd interactions, effectively addressing the problem of trajectory prediction. Additionally, Hong et al. (2022) developed a transformer decoder-based neural network to predict a user's next location based on historical location, time, and travel mode.

However, existing models rarely take into account emotional information and lack in-depth exploration of spatio-temporal dependencies. The lack of spatio-temporal dependencies for emotion makes it difficult to predict spatio-temporal emotions. Therefore, capturing emotion dependencies is an urgent problem for current research.

2.2 Emotion calculation

With the rise of social media platforms such as Sina Weibo, Twitter, and Facebook, the daily online communication and sharing of urban residents has generated a large amount of social media data, which provides the possibility of emotion calculations. Existing emotion calculation methods broadly consist of two parts: emotion recognition and emotion prediction.

Traditional recognition methods mainly use manually designed features combined with classifiers for emotion recognition. Principal component analysis is one of the commonly used methods that can effectively extract the global features of the data and reduce the data dimensionality (Wold et al., 1987; Niu and Qiu, 2010). However, the recognition accuracy is limited, and it cannot make full use of the data information. For this reason, people started to extract features based on manually designed features for recognition. Such methods can improve recognition accuracy to a certain extent (Pantic and Stewart, 2007; Cootes et al., 1995). However, when key information is lost, the extracted features deviate, and the accuracy decreases. With the emergence of machine learning methods, people have started to apply them to extract features to improve the recognition accuracy. The optical flow method is often used to extract features, and Sun et al. (2019) proposed a multichannel deep spatio-temporal feature fusion neural network. This method uses optical flow to represent temporal features in static images, thus effectively recognizing face expressions in static images. Pan et al. (2019) extracted features from optical flow image sequences and designed a fully connected convergence layer for fusing different modal features. In addition to the optical flow method, an attention mechanism is also used to extract features. Sun et al. (2018) used CNN and a single-layer spatial attention mechanism to filter expression-related features. To further improve the accuracy, Mai et al. (2022) proposed a face expression recognition method based on a multiscale feature attention mechanism that uses two convolutional layers to extract information. The model introduces a channel attention mechanism to enhance the utilization of useful feature information. To construct the remote temporal association of features, LSTM is also used for emotion prediction. Liu and Wu (2020) proposed an emotion recognition model based on LSTM networks. The model further improves the accuracy of facial expression recognition by selecting multiple facial expression features for facial images.

Based on the correct recognition of emotion, classifiers can be combined or temporal associations of emotion can be constructed to predict future emotions. Nicolaou et al. (2012) proposed a novel output-associative relevance vector machine (OA-RVM) regression framework; the method employed a temporal window to learn temporal associations for emotion prediction in time series. Regarding text, word embedding are often used to extract semantic information from words and to analyze emotion. Thavareesan and Mahesan (2020) utilized Word2vec to represent each word and calculate their cosine similarity to classify different categories of emotions. BiLSTM is usually used to capture bidirectional long-time dependencies for emotion prediction. Li et al. (2020) used BiLSTM to extract sentence emotion and construct its contextual associations to predict text emotion. Zhang et al. (2020) introduced the whale optimization algorithm (WOA) to improve LSTM for the accurate prediction of public environmental emotions in time series. CNN is often employed to extract emotional features from images. Campos et al. (2017) studied the suitability of fine-tuning a CNN for visual feature extraction and used a linear layer to predict image emotion. Abdullah et al. (2018) proposed a coupled model of CNN and LSTM. It uses CNN to extract local features and LSTM to construct temporal correlations. The method achieves a higher prediction accuracy than CNN and LSTM.

Overall, these deep learning-based methods predict users' emotional states by extracting feature factors of the data and constructing their mapping relationships with emotions. However, these methods are not accurate enough to construct the temporal correlations of emotion. They also rarely consider the location information and have difficulty constructing the multidimensional spatio-temporal dependencies of emotions. As a result, these methods cannot accurately predict the user's spatio-temporal affective state.

3 Research methods and data sources

3.1 Experimental data and preprocessing

This study uses microblogs in Shanghai on the Weibo platform, and the original data include 748,790 microblogs with images posted by 267,737 users from 2017 to 2020. It is well known that Sina Weibo is the most popular microblog website in China. The platform has gained tremendous influence by significantly influencing the process of many real-world hot social events (Guan et al., 2014). The microblogs are collected based on the Weibo API, and each data sample includes the user's ID, coordinates, post time, and the URL of the image, as shown in Table 1. We obtained the image from the microblog based on the URL.

To create a reliable dataset, noise had to be removed. Microblog noise consists primarily of reposted microblogs, advertisements, and bot-posted microblogs. Geotagged microblogs cannot be reposted (Jiang et al., 2015). Therefore, geotagged microblogs have much less noise. During the preprocessing stage, we first obtained geotagged microblogs in Shanghai by filtering out coordinates. Most ads contain specific symbols, such as “【】.” Bot users often post duplicate messages for multiple days. We removed 42,167 microblogs containing these particular symbols. To avoid the sparsity of data on the experiment, we selected users who had posted at least 10 microblogs for 30 consecutive days as the experimental subjects based on existing social media research experience (Bao et al., 2021). Additionally, we strictly filtered the dataset to include only images containing human faces. The filtered dataset includes 206,881 microblogs from 11,996 users.

The spatial distribution of microblogs is shown in Figure 1. We discover that the spatial distribution of the microblogs is uneven.

TABLE 1 Microblog data.

Post time	y	x	ID	URL
Fri Jan 01 05:09:23 + 0800 2021	30.72**	121.35**	390***	http://wx1.sinaimg.cn/***.jpg
Fri Jan 01 03:16:56 + 0800 2021	30.70**	121.34**	189***	http://wx4.sinaimg.cn/***.jpg
Fri Jan 01 01:38:52 + 0800 2021	30.72**	121.33**	597***	http://wx1.sinaimg.cn/***.jpg
…	…	…	…	…
Fri Jan 01 01:11:50 + 0800 2021	30.72**	121.34**	293***	http://wx3.sinaimg.cn/***.jpg

**indicates hidden characters.

The urban centers concentrate the majority of data, whereas suburban locations post less data. Regarding the time dimension, Figure 2 illustrates microblog posting intervals, and clear temporal clustering can be observed. This research randomly chose six users and plotted their check-in time series distribution. There is a wide range of time between microblogs, ranging from a few days to several weeks.

3.2 Methodology

3.2.1 Overall framework

The flow chart of the location emotion prediction model is illustrated in Figure 3. The method is divided into four parts: stay area identification, user emotion measurement, emotional trajectory matrix construction, and emotional dependency relationship construction. Initially, the clustering algorithm HiSpatialCluster is applied to determine possible stay areas. The images posted by Weibo users are analyzed using the FaceReader algorithm and Russell model for emotion measurement. Then, stay areas are abstracted into graph vertices to construct an emotional interaction graph, and each user's trajectory matrix is calculated. Next, the original matrix is encoded according to the sequence. Time and space information are captured by a multi-head attention layer. BiLSTM layer outputs are concatenated with attention layer outputs to predict location and emotional quadrants.

3.2.2 Stay area identification

Considering the sparsity of data, we partition the users' staying area using the HiSpatialCluster clustering algorithm. HiSpatialCluster is an algorithm based on a fast search, density peak clustering and density connection filtering-based classification clustering. Adaptive clustering can be performed on massive point datasets with uneven spatial distributions of density. The algorithm divides check-in data into divisions of varying sizes based on data densities and balances the number of points inside each partition. Compared with existing clustering methods, such as DBSCAN (Ester et al., 1996) and K-Means (Bock, 2007), HiSpatialCluster has better robustness. It can effectively filter noise points, quickly cluster massive data, and quickly reach convergence. In this study, we adopt HiSpatialCluster for clustering large-scale geotagged social media data. Compared to DBSCAN and KMeans, HiSpatialCluster offers distinct advantages in handling datasets with uneven density and regional heterogeneity. Previous studies (Chen et al., 2018; Bao et al., 2021) have empirically demonstrated that HiSpatialCluster outperforms DBSCAN and KMeans in clustering large spatial point datasets, particularly those characterized by varying densities. Both studies emphasize that HiSpatialCluster not only improves the accuracy of delineating meaningful spatial regions but also maintains high computational efficiency when applied to large-scale datasets. Given the similarity of our dataset to those used in prior studies in terms of platform, scale, spatial distribution, and research objective, we adopted HiSpatialCluster to perform the clustering analysis in this study. After clustering the data and generating the clustered areas, the emotion attributes can be attached to the clustered areas.

3.2.3 User emotion measurement

In this paper, the Russell ring model (Russell, 1980) and FaceReader algorithm are combined to obtain the emotional quadrants and measure the emotion of expressions in images. Russell's ring model is an emotion classification method, as shown in Figure 4. It divides emotion into two dimensions: valence and intensity. The valence indicates the degree of pleasure of emotion, and the intensity indicates the level of emotion. As a result, we divided the emotions into four quadrants. The first quadrant is the high valence value-high intensity state, the second quadrant is the low valence value-high intensity state, the third quadrant is the low valence value-low intensity state, and the fourth quadrant is the high valence value-low intensity state. Thus, four significantly different emotional states can be distinguished corresponding to four emotional quadrants. All emotional emotions can be represented as a point in a two-dimensional coordinate system. Existing studies on emotion classification in images and text typically focus on differentiating between positive and negative polarity (Liu et al., 2023). Our approach extends this framework by employing a Russell ring model to segment emotions into four distinct quadrants, offering a more comprehensive and nuanced understanding of a user's emotional state. FaceReader can recognize the valence and intensity of facial expressions in images. It is a high-precision emotion recognition method based on a deep learning framework. FaceReader is proficient in accurately recognizing emotions based on facial expressions and effectively monitoring users' emotional states (Terzis et al., 2010). The Viola-Jones algorithm is applied to find the location of faces in images based on more than 500 key points of the face, and a deep neural network is utilized to determine the emotional state of the face. The output of FaceReader is the intensity and valence of seven-dimensional emotions (happy, sad, angry, surprised, fearful, disgusted, and neutral). Although FaceReader outputs seven discrete emotions, some of these expressions (e.g., angry and disgusted) share highly similar facial features, limiting the algorithm's ability to distinguish similar expressions. Directly using the seven categories as label may therefore propagate classification errors into subsequent analyses. To enhance prediction reliability and ensure clearer differentiation,

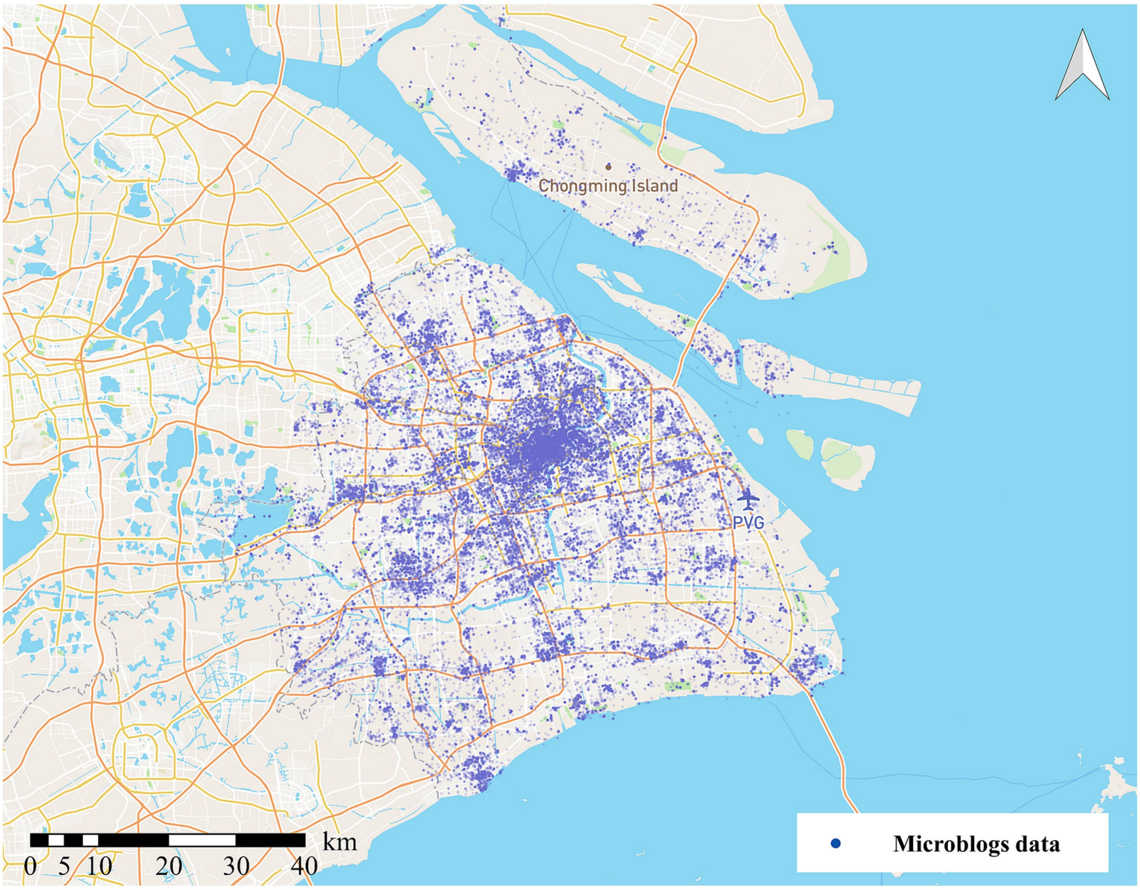

FIGURE 1
Microblog data distribution.

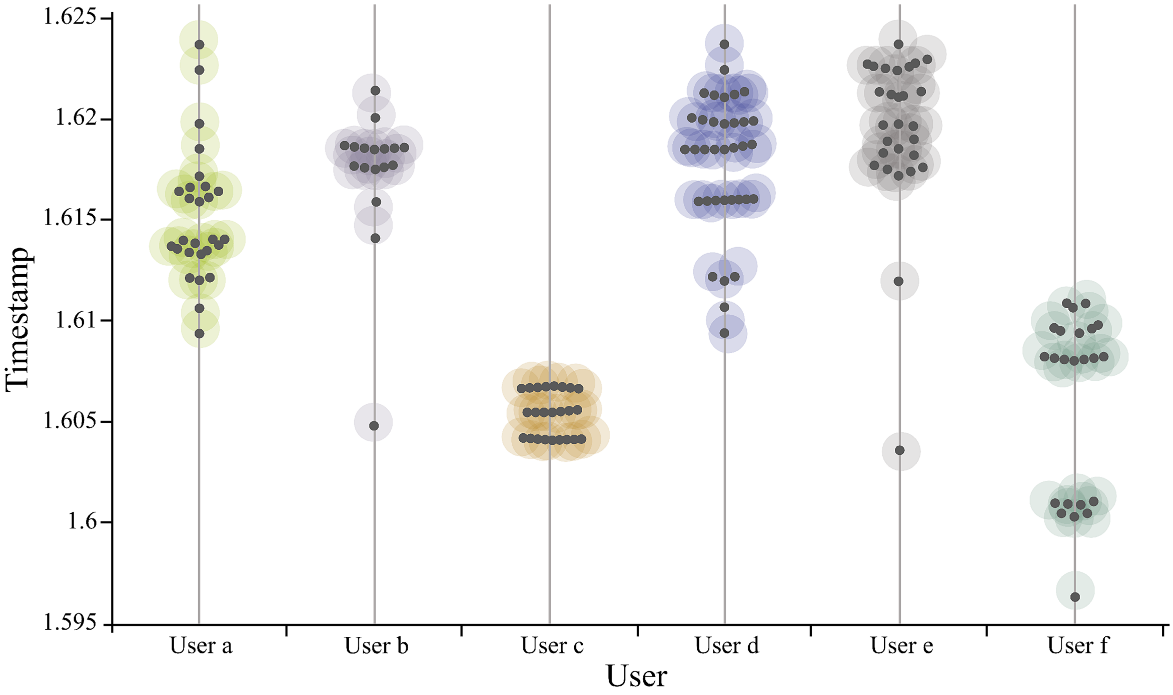

FIGURE 2
Check-in time distribution.

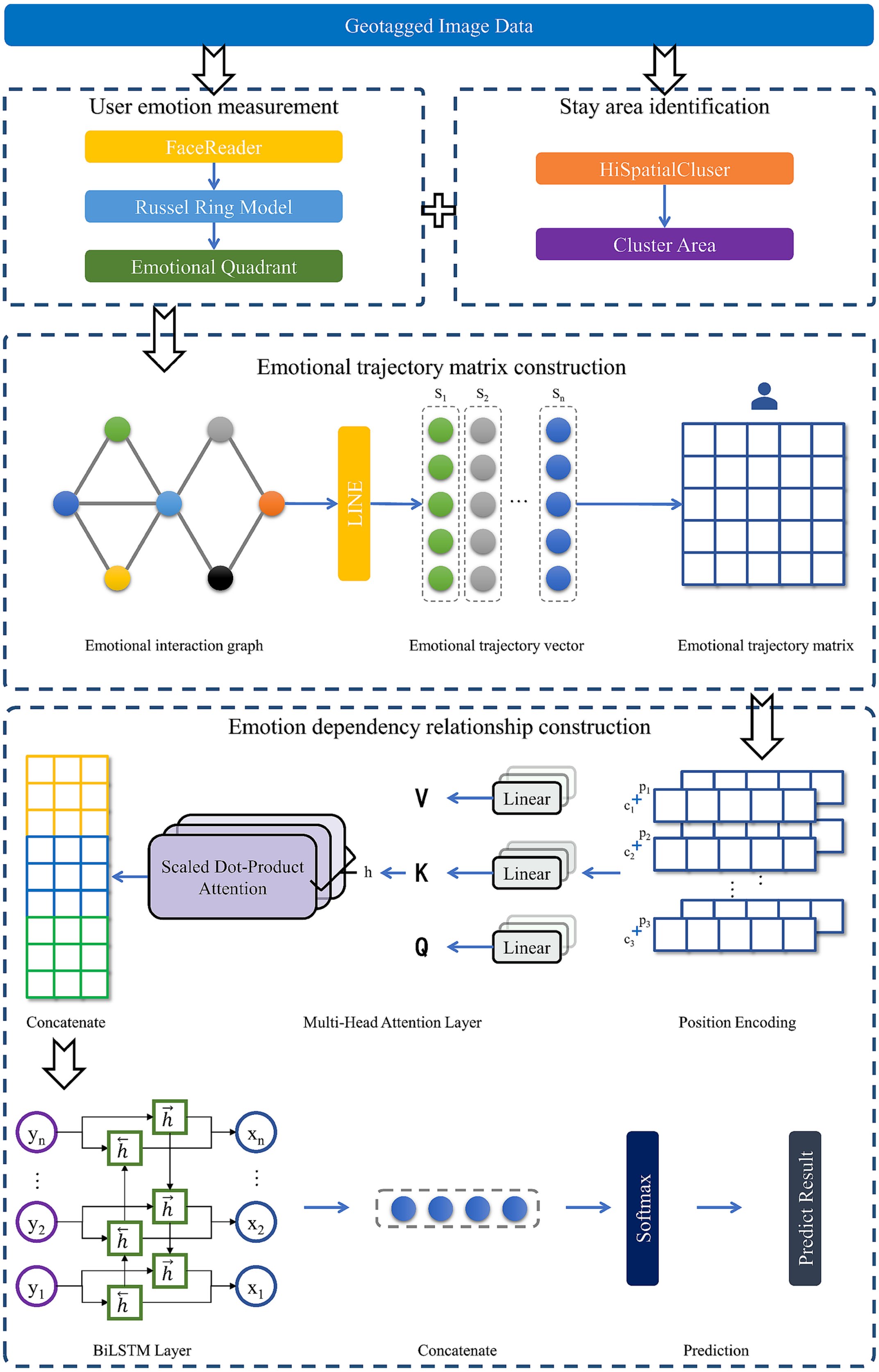

FIGURE 3
Flow chart of our proposed methodology.

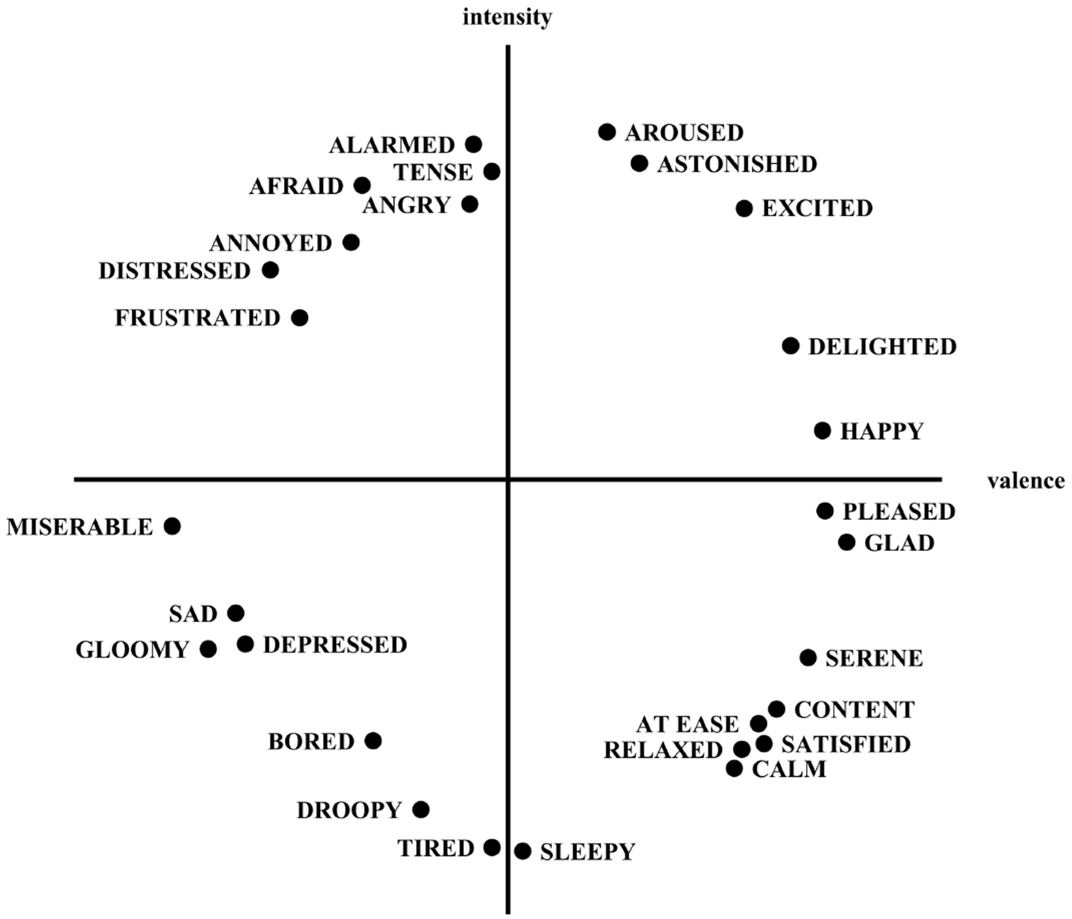

FIGURE 4
A circumplex model of emotions.

we reclassified the seven emotions into four emotion quadrants based on the Russell ring model (Cittadini et al., 2023).

FaceReader analyzes facial expressions and outputs emotion data that corresponds to a point on the Russell ring model. The coordinates of this point determine the corresponding valence (x_i) and intensity (y_i) of the emotion. For each image, the $Emotion_i$ is then represented using Equation 1:

$$Emotion_i = \left(x_i, y_i\right) \tag{1}$$

where x_i and y_i denote the valence and the intensity of image i, respectively.

For a Sina Weibo post containing n images, the $Emotion_i$ values for all images are aggregated, and the resulting composite emotion is classified into one of four emotional quadrants. These quadrants are defined as follows: the first quadrant corresponds to high valence and high intensity; the second quadrant corresponds to low valence and high intensity; the third quadrant corresponds to low valence and low intensity; and the fourth quadrant corresponds to high valence and low intensity. The dimensional representation of emotion, denoted as $\dim ension$, is a two-dimensional array derived from the aggregation of the $Emotion_i$ values, as indicated by Equation 2. The Dimension array's coordinates are then used to determine the corresponding emotional quadrant.

$$\dim ension = \sum_{i=1}^{n} Emotion_{\mathrm{i}} \tag{2}$$

The coordinates of the aggregated emotion ($\dim ension$) are mapped to the four quadrants of the Russell ring model, where each quadrant reflects a distinct combination of valence and intensity. The emotional quadrant allows us to identify the emotional state of a user in a particular spatio-temporal area. By acquiring the emotion quadrant of images, it can provide data support for the construction of emotional trajectory matrices.

3.2.4 Emotional trajectory matrix construction

The construction of the emotional trajectory matrix consists of two steps: the construction of the emotional graph and the calculation of the emotional trajectory vector.

Emotional interaction graph construction. We use the emotional interaction graph to quantify the travel preferences and the emotional state in each stay area, as shown in Figure 5. Based on the users' stay

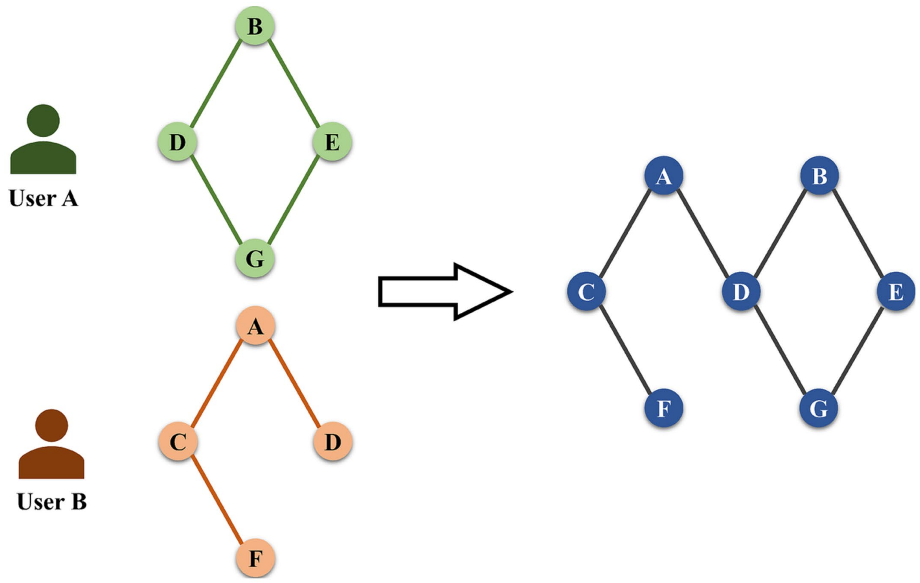

FIGURE 5
Emotional interaction graph construction.

area and the emotional quadrant, the emotional interaction graph is developed using graph theory. In addition, the graph quantifies the interaction relationship of the users' emotions between each stay area. There are *n* nodes in the emotional interaction graph, each of which contains two properties: the stay area and the emotion quadrant, as shown in Equation 3:

$$V=\left\{\left(C_1,E_1\right),\left(C_2,E_2\right),\left(C_3,E_3\right)\ldots\left(C_n,E_n\right)\right\} \tag{3}$$

where C_i denotes the stay area, and E_i represents the emotion quadrant. After constructing the emotional interaction graph, we use U_k to denote the k_{th} user trajectory as shown in Equation 4.

$$U_k=\left\{V_{1k},t_1,V_{2k},t_2\ldots V_{mk},t_m\right\}\left(t_1<t_2<t_3\ldots<t_m\right) \tag{4}$$

where V_k is the corresponding node in the emotional interaction graph, and t_m is the time of the m_{th} microblog posting.

In this paper, we define the weights of edges using the function g(·). The weights between two adjacent vertices V_i^k and V_j^k are computed as follows:

$$g\left(U^k_{V_i^k,V_j^k}\right)=\frac{1}{1+\mathrm{e}^{-\frac{Tmax}{\alpha\left(t_j-t_i\right)}}} \tag{5}$$

where α is used to regulate the weight of time intervals. Based on the experience of previous studies, α = 10 is set in this study as applicable to the dataset. T_{max} is the maximum time interval for the k_{th} user to post a microblog. If there are adjacent nodes V_i and V_j that are visited by multiple users, the weights corresponding to all users' sequences are first calculated separately using Equation 5. Then, Equation 6 is used to calculate the weights of the edges between V_i and V_j in the whole emotional interaction graph.

$$Edge=\left(V_i,V_j\right)=\Sigma_{k=1}^{s}g\left(U^k_{V_i^k,V_j^k}\right)if\exists V_i^k,V_j^k:V_i=V_i^k,V_j=V_j^k \tag{6}$$

where V is the set of vertices, and E_{VV} contains all weighted edges.

Emotional trajectory vector calculation. To transform the graph structure into vectors for model computation, this study constructs emotion trajectory vectors based on the LINE algorithm and builds the users' emotional trajectory matrix. LINE is an algorithm for graph embedding (Tang et al., 2015) that generates a low-dimensional vector representation of the graph's vertices. Liu et al. (2023) modeled and integrated user behavior and social influence by constructing a social behavior graph. Graph embedding is employed to transform social networks into learnable low-dimensional representations, thereby capturing latent social relationships. Similarly, our study develops emotion interaction graph to integrate users' spatial and emotional information. The LINE method is applied to generate embedding vectors that represent the spatio-temporal correlations of user emotions. Several approaches for graph embedding have been proposed, including DeepWalk (Perozzi et al., 2014), node2vec (Grover and Leskovec, 2016) and LINE. While DeepWalk and Node2Vec effectively capture local (first-order) neighborhood structures, their ability to represent higher-order proximities is limited, making it difficult to model indirect dependencies (Tang et al., 2015). In emotional trajectory analysis, such indirect relationships are essential, as emotional influence often propagates through multiple intermediaries rather than being confined to direct interactions. By explicitly preserving both first-order and second-order proximities, LINE provides a more comprehensive representation of influence patterns across the network. Compared with traditional graph embedding algorithms, the LINE algorithm captures both first- and second-order proximities. The ability to model second-order proximity enables LINE to represent latent, indirect connections between nodes. This provides a more nuanced representation of spatio-temporal emotional interactions, where immediate links may not fully explain the underlying dynamics. Meanwhile, LINE proposes an edge sampling algorithm to improve stability and is more suitable for embedding large-scale graphs. Figure 6 shows the LINE embedding process.

The vertex vectors generated from the LINE algorithm can effectively reflect the potential spatio-temporal correlation between the emotions of various stay areas. In the graph structure, highly correlated vertices correspond to larger edge weights. Even if two vertices are not directly connected, they may share the same vertices and thus can be considered to have similar states in the graph. This condition is similar to the second-order approximation of LINE. Therefore, LINE is used to process the spatio-temporal emotional interaction graph and construct the low-dimensional vector for each stay area C_i. In the graph, each stay area C_i plays two roles, its own and a specific context. There are two types of distributions that can be defined for these two kinds of roles: the conditional distribution $\hat{p}_i$ and the empirical distribution $\hat{p}_2$. The objective of LINE is to minimize the distance between the two distributions using Equation 7:

$$O_2 = -\sum_{i \in V} \lambda_i KL\left(p_i, \hat{p}_i\right) \tag{7}$$

KL(•, •) is the KL scatter between the two distributions. λ_i is the out-degree of C_i that represents the importance of the vertices in the whole graph. The low-dimensional vector representation $\vec{u_i}$ of each stay area C_i can be obtained by minimizing O_2.

After obtaining the vectors representing each stay area, the vectors are sequentially stitched in chronological order based on their historical temporal information. As shown in Figure 7, we obtain a matrix of users' spatio-temporal emotion trajectories as the input to the prediction model. The trajectory matrix of the k_{th} user can be expressed as $F_k = \left[\vec{u}_1^{\,k}, \vec{u}_2^{\,k}, \vec{u}_3^{\,k} \cdots \vec{u}_{m-1}^{\,k}\right]$.

3.2.5 Emotion dependency relationship construction

The construction of multidimensional spatio-temporal dependencies of emotions is divided into two parts: the encoding of location and the construction of dependencies. Building emotion spatio-temporal dependencies is an important part of extracting the interaction information from the trajectory matrix. By constructing the spatio-temporal dependencies of the emotion trajectory matrix, we can more accurately capture the dependencies of the matrix. Furthermore, it allows us to fully investigate the spatio-temporal emotion interaction information of users.

We encode the emotion trajectory vectors to capture the contextual information of the emotion trajectory vector. Then, we investigate spatio-temporal information from the emotion trajectory vectors. For each vector in the matrix, encoding vectors are generated based on the vector's position in the sequence and its dimensionality. Next, it is correspondingly added to the original emotion trajectory vectors. The position encoding is calculated using Equations 8 and 9.

$$PE_{(pos,2i)} = \sin\left(pos / 10000^{2i/d}\right) \tag{8}$$

$$PE_{(pos,2i+1)} = \cos\left(pos / 10000^{2i/d}\right) \tag{9}$$

where *pos* is the position of the emotion trajectory vector in the emotion trajectory matrix, and *d* is the dimension of the emotion vector.

After incorporating the contextual information, the multidimensional spatio-temporal dependencies of emotions are constructed based on the attention-based BiLSTM model. The main idea

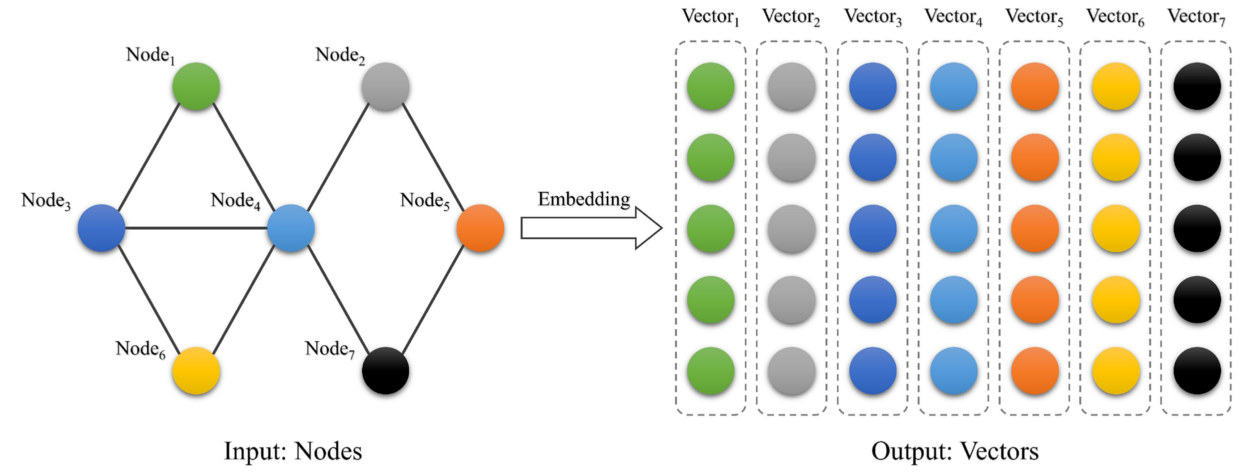

FIGURE 6
Graph embedding.

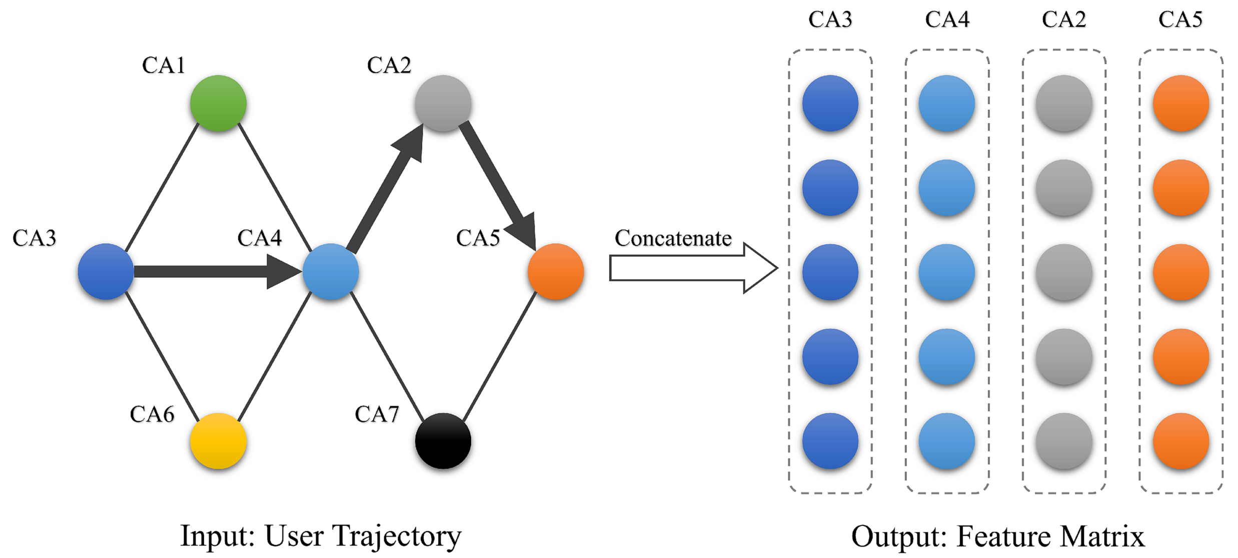

FIGURE 7
User emotion matrix.

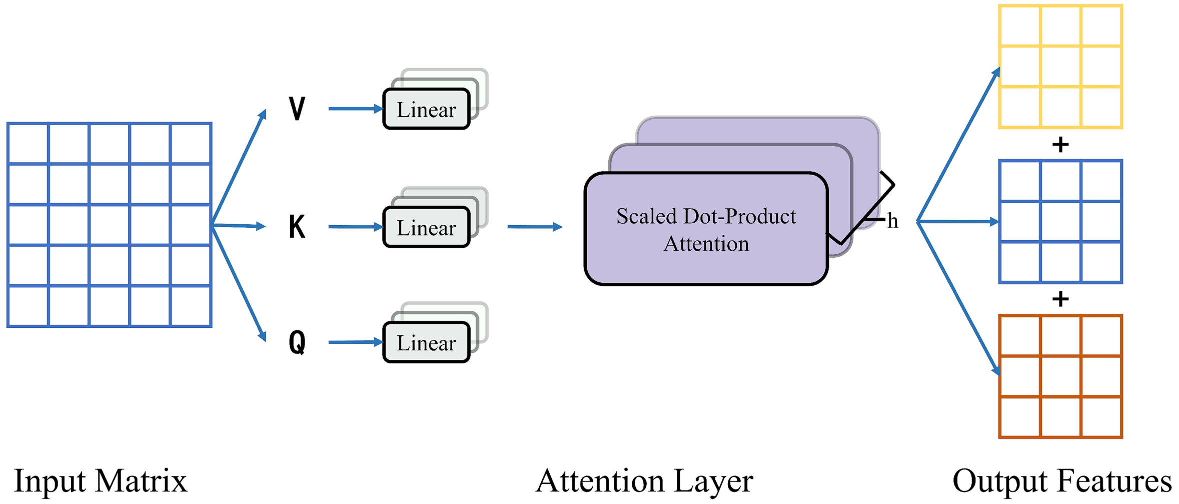

FIGURE 8
Multi-head attentional mechanism.

of the attention mechanism is to assign different weights to each part of the input and allocate varying degrees of attention according to the weight size during the decoding process. Unlike traditional RNNs, the attention mechanism allows the model to focus on the parts that contribute to prediction and ignore the irrelevant parts. It is possible to dynamically adjust the weight size of the model based on the input of the model to make more accurate judgments. To obtain multidimensional features, a multiheaded attention mechanism integrates multiple self-attentive mechanisms. As shown in Figure 8, the weights of each part are reassigned by softmax normalization after dynamically calculating the similarity between each vector. Finally, weighting and summing are performed to produce the new matrix. The initial three matrices K, Q, and V are scaled by the dot product between them and normalized by the softmax layer to extract the spatial and temporal dependence information of emotions. The attentional mechanism is calculated using Equation 10.

$$Attention\left(Q,K,V\right)=softmax\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \tag{10}$$

where Q is the query matrix, K is the key matrix, V is the value matrix, and d_k is the dimensionality of the key vectors.

A long short-term memory neural network is a kind of time-series recurrent neural network. Its network structure consists of one or more units with forgettable and memorable functions. Thus, LSTM is suitable for processing and predicting important events with very long intervals and delays in time series. It was proposed in 1997 to solve the problem

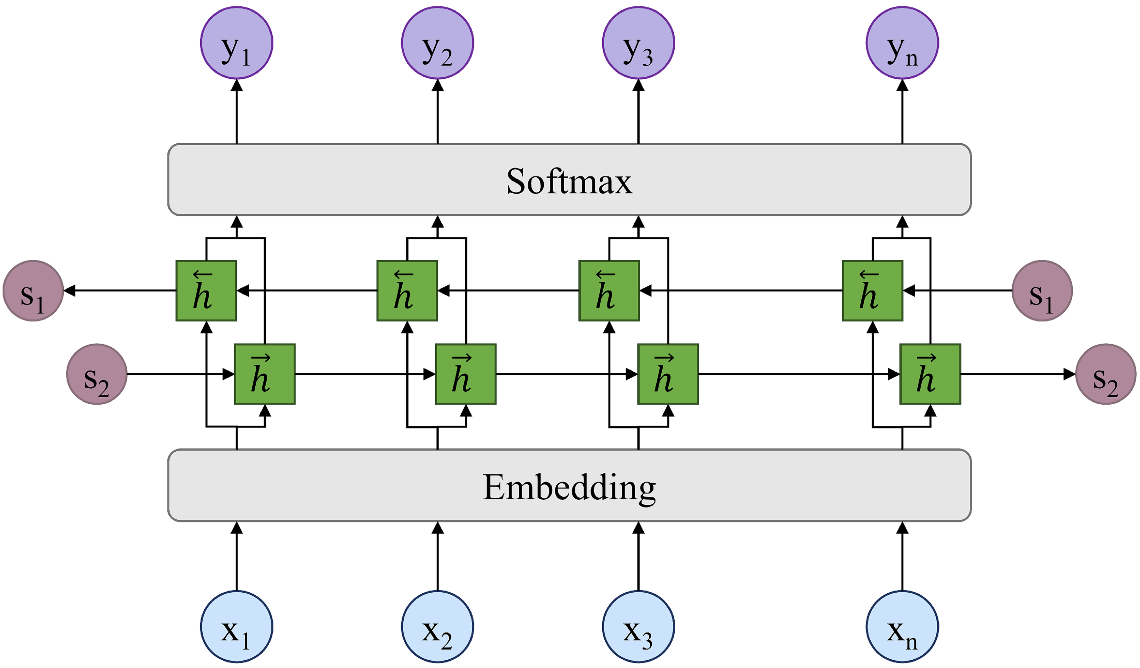

FIGURE 9
BiLSTM layer.

of weight disappearance in traditional RNN backpropagation over time. The important components include the forget gate, input gate, and output gate. These gates are responsible for deciding whether the current input is adopted, whether it is remembered in the long term, and whether the input in memory is output in the present. As shown in Figure 9, BiLSTM is a combination of a forward LSTM and a backward LSTM. As a result, it can capture both forward and backward timing information of time series, resulting in improved prediction results. Although the attention mechanism effectively captures global spatio-temporal dependencies, it provides limited explicit modeling of the local sequential continuity that characterizes emotion trajectories. To complement this, a BiLSTM layer is incorporated after the attention module. By processing the sequence bidirectionally, BiLSTM captures fine-grained temporal transitions and preserves contextual continuity. In this architecture, the attention mechanism supplies a global selective focus, while BiLSTM refines local temporal dynamics and reinforces long-range contextual information. Together, they yield a more comprehensive representation of spatio-temporal emotion interactions.

In the final step, we concatenate the feature vectors that represent spatio-temporal emotional dependencies. Next, concatenated vectors are inputted into a fully connected neural network containing a softmax layer to predict the location and corresponding emotional quadrants. The model predicts location and emotion jointly by employing shared layers.

4 Case study

4.1 Model evaluation

In this study, 80% of the data are taken at random from the sample set for training the model, while the remaining 20% are utilized to validate the model's correctness. The error is estimated using the cross-entropy loss function in Equation 11.

$$L = -\frac{1}{N}\sum_{i}\sum_{c=1}^{m} y_{ic}\log\left(p_{ic}\right) \tag{11}$$

where M denotes the number of categories; y_{ic} is the sign function that returns 1 if the true category of sample i is equal to c and 0 otherwise. p_{ic} is the predicted probability that the observed sample belongs to category c.

On this basis, we utilize k-fold cross-validation to divide the training set into k subsamples, maintaining one subsample for model validation and using the remaining k-1 samples for training. Cross-validation is performed k times, once for each subsample, and the results are then averaged k times to obtain a single estimate.

Training was performed on an Nvidia 4,060 GPU, requiring approximately 4 h for 256 epochs with a batch size of 64. The model contains approximately 470 K parameters, which is relatively lightweight compared to deep learning architectures typically used in this domain.

4.2 Analysis of the results

In this section, to present our prediction results in a more visual and detailed way, we selected one user and plotted his check-in locations and emotional status in different clustering areas. As shown in Figure 10, users' check-in locations are distributed in Areas A, B, C and D. Then, we will analyze the prediction accuracy of location emotions from three perspectives: clustering scale, emotional quadrant, and sequence length. By capturing and constructing the

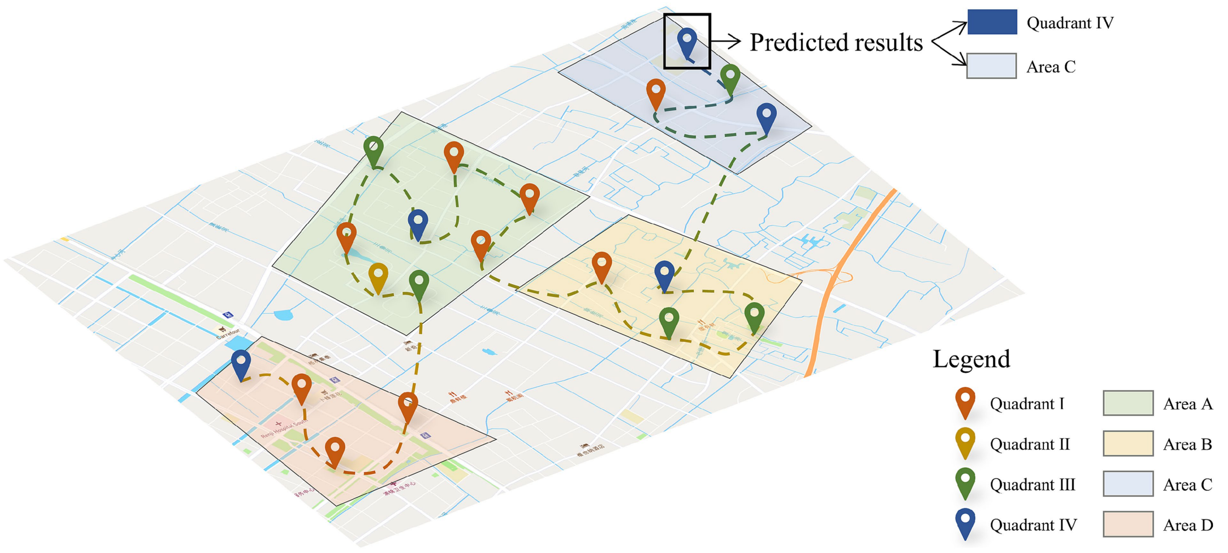

FIGURE 10
Spatial–temporal prediction of user emotions.

TABLE 2 Prediction performance comparison with baselines (accuracy).

Models	Number of clustered areas				
	100	200	300	400	500
ANN	0.6995	0.6563	0.6531	0.6392	0.6209
CNN	0.7381	0.6911	0.6647	0.6511	0.6492
BiLSTM	0.7337	0.6904	0.6773	0.6569	0.6636
ResNet	0.7176	0.6824	0.6619	0.6418	0.6314
ConvBiLSTM	0.7427	0.7161	0.6963	0.6915	0.6853
Attention-Based BiLSTM	**0.7521**	**0.7212**	**0.7208**	**0.7196**	**0.6949**

The bold values indicate the prediction performance of the model proposed in this study.

TABLE 3 Prediction performance comparison with baselines (precision).

Models	Number of clustered areas				
	100	200	300	400	500
ANN	0.6895	0.6469	0.6386	0.6261	0.6120
CNN	0.7166	0.6741	0.6635	0.6529	0.6343
BiLSTM	0.7298	0.6834	0.6644	0.6547	0.6419
ResNet	0.7074	0.6634	0.6502	0.6456	0.6228
ConvBiLSTM	0.7283	0.6952	0.6729	0.6635	0.6587
Attention-Based BiLSTM	**0.7423**	**0.7118**	**0.7072**	**0.6923**	**0.6859**

The bold values indicate the prediction performance of the model proposed in this study.

TABLE 4 Prediction performance comparison with baselines (recall).

Models	Number of clustered areas				
	100	200	300	400	500
ANN	0.6824	0.6403	0.6273	0.6156	0.6058
CNN	0.7092	0.6671	0.6463	0.6391	0.6277
BiLSTM	0.7181	0.6723	0.6573	0.6456	0.6316
ResNet	0.7049	0.6515	0.6363	0.6219	0.6116
ConvBiLSTM	0.7211	0.6884	0.6773	0.6663	0.6522
Attention-Based BiLSTM	**0.7312**	**0.7012**	**0.6978**	**0.6887**	**0.6756**

The bold values indicate the prediction performance of the model proposed in this study.

TABLE 5 Prediction performance comparison with baselines (F1 score).

Models	Number of clustered areas				
	100	200	300	400	500
ANN	0.6859	0.6436	0.6329	0.6208	0.6089
CNN	0.7129	0.6706	0.6548	0.6459	0.6310
BiLSTM	0.7239	0.6778	0.6608	0.6501	0.6367
ResNet	0.7061	0.6574	0.6432	0.6335	0.6171
ConvBiLSTM	0.7247	0.6918	0.6751	0.6649	0.6554
Attention-Based BiLSTM	**0.7367**	**0.7065**	**0.7025**	**0.6905**	**0.6807**

The bold values indicate the prediction performance of the model proposed in this study.

multidimensional spatio-temporal dependencies of users' check-in location and emotional status, the proposed model achieves high-precision prediction in location emotion and helps analyze the spatio-temporal pattern of users' emotional states.

The association between prediction accuracy, precision, recall and F1 score are shown in Tables 2–5. Compared with single models, our method has the advantage for all numbers of clustered areas. When the number of clustered areas is 100, the prediction accuracy of our proposed model is 75.21%. The accuracies of the ANN, CNN, BiLSTM, ResNet and ConvBiLSTM are 69.95, 73.81, 73.37, 71.76 and 74.27%, respectively. Our model accuracy has improved by 5.26, 1.40, 1.84, 3.45 and 0.94%. In addition to accuracy, our method also demonstrates

TABLE 6 Prediction accuracy for four emotional quadrants.

(a)					
Models	Number of clustered areas	100			
	Emotional quadrant	I	II	III	IV
ANN		0.7128	0.6628	0.6703	0.6998
CNN		0.7549	0.7035	0.7029	0.7404
BiLSTM		0.7514	0.7013	0.6992	0.7369
ResNet		0.7447	0.6883	0.6859	0.7271
ConvBiLSTM		0.7699	0.7176	0.7189	0.7551
Attention-Based BiLSTM		**0.7709**	**0.7154**	**0.7245**	**0.7683**

(b)					
Models	Number of clustered areas	200			
	Emotional quadrant	I	II	III	IV
ANN		0.7102	0.6527	0.6648	0.6988
CNN		0.7434	0.7014	0.6994	0.7498
BiLSTM		0.7475	0.7047	0.7027	0.7403
ResNet		0.7266	0.6818	0.6881	0.7272
ConvBiLSTM		0.7536	0.7015	0.7055	0.7489
Attention-Based BiLSTM		**0.7594**	**0.7065**	**0.7117**	**0.7519**

(c)					
Models	Number of clustered areas	300			
	Emotional quadrant	I	II	III	IV
ANN		0.6888	0.6365	0.6427	0.6907
CNN		0.7305	0.6795	0.6794	0.7276
BiLSTM		0.7288	0.6801	0.6883	0.7354
ResNet		0.7094	0.6557	0.6543	0.7059
ConvBiLSTM		0.7461	0.6815	0.6865	0.7339
Attention-Based BiLSTM		**0.7539**	**0.6895**	**0.6996**	**0.7422**

(d)					
Models	Number of clustered areas	400			
	Emotional quadrant	I	II	III	IV
ANN		0.6888	0.6365	0.6427	0.6907
CNN		0.7305	0.6795	0.6794	0.7276
BiLSTM		0.7288	0.6801	0.6883	0.7354
ResNet		0.7128	0.6596	0.6656	0.7072
ConvBiLSTM		0.7356	0.6816	0.6710	0.7387
Attention-Based BiLSTM		**0.7539**	**0.6895**	**0.6996**	**0.7422**

(Continued)

TABLE 6 (Continued)

(e)					
Models	Number of clustered areas	500			
	Emotional quadrant	I	II	III	IV
ANN		0.6888	0.6365	0.6427	0.6907
CNN		0.7305	0.6795	0.6794	0.7276
BiLSTM		0.7288	0.6801	0.6883	0.7354
ResNet		0.7176	0.6576	0.6571	0.7025
ConvBiLSTM		0.7422	0.6722	0.6859	0.7395
Attention-Based BiLSTM		**0.7539**	**0.6895**	**0.6996**	**0.7422**

The bold values indicate the prediction performance of the model proposed in this study.

consistent improvements in precision, recall and F1 score. Even when the number of clustered areas reaches 500, our method still has a prediction accuracy of 69.49%. In addition to accuracy, our method also demonstrates consistent improvements in precision, recall and F1 score.

In terms of emotional quadrants, we compared the location prediction accuracy of the model in the four emotion quadrants and for different clustering scales. The prediction results are shown in Tables 6A–E. The four quadrants here represent four emotional states with significant emotional differences. According to the Russell ring model, Quadrants I and IV represent positive emotional states, and Quadrants II and III represent negative emotional states. The prediction accuracy of our model is better than that of existing models in all emotional quadrants. In particular, our model has a significant prediction accuracy advantage with regards to positive emotional states (Quadrant I and Quadrant IV). The first quadrant (high valence-high intensity state) had the highest accuracy, followed by the fourth quadrant (high valence-low intensity state), then the third quadrant (low valence-low intensity state), and finally the second quadrant (low valence-high intensity state). This discrepancy is primarily attributable to the imbalanced distribution of emotions in online data. Online users tend to express positive opinions (Yang and Xu, 2023), resulting in a predominance of positive samples in our dataset. Specifically, Quadrants 1 and 4 (positive emotions) account for 73.4% of the data, while Quadrants 2 and 3 (negative emotions) together represent only 26.6%. Within this minority group, Quadrant 2 accounts for merely 12.7%. Such class imbalance restricts the model's exposure to negative samples during training, leading to insufficient learning of negative emotion features. Consequently, the model achieves higher prediction accuracy for Quadrants 1 and 4, while accuracy for Quadrant 2 is considerably lower.

The effect of the sequence length on the accuracy of emotion prediction is shown in Figure 11. In this study, the sequence length refers to the number of user check-in points, and sequence lengths range from 10 to 50. In the dataset employed in this study, the maximum length of a user check-in sequence is up to 51, thereby constraining the influence of sequence length on prediction accuracy to sequences of up to 50. The results of Figure 11 show that the prediction accuracy gradually increases with an increasing sequence length. The accuracy reaches 74.8% after the check-in sequence length reaches 50. The longer the sequence length is, the more

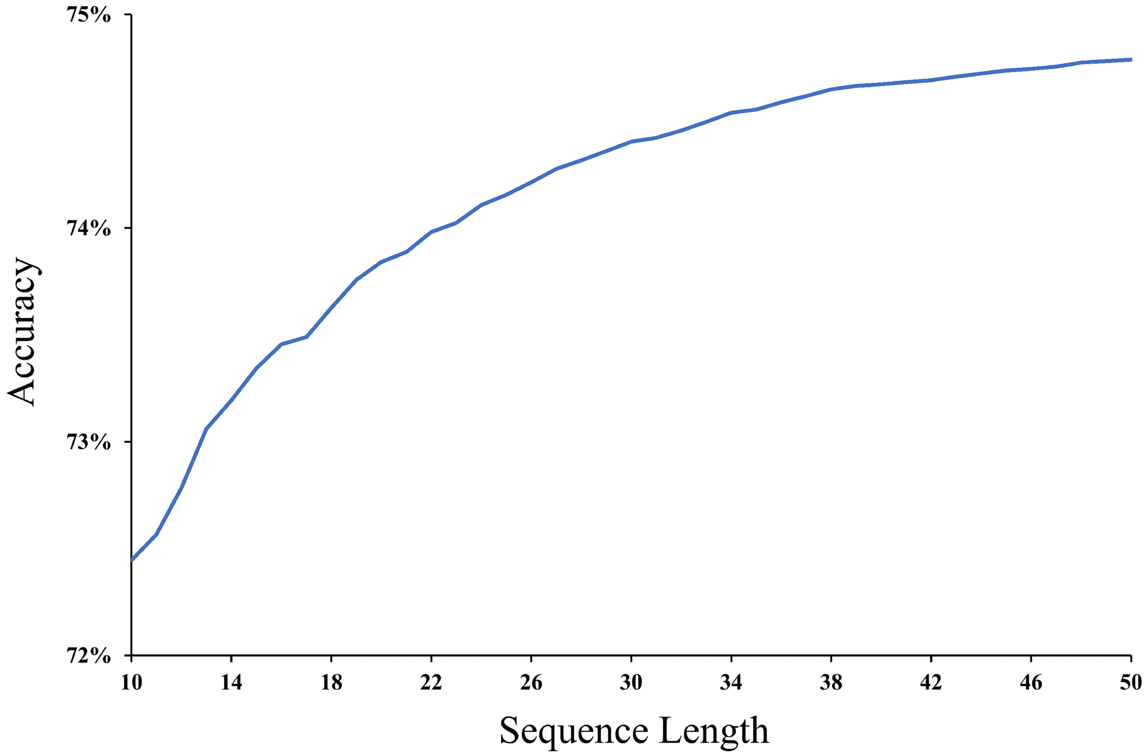

FIGURE 11
Prediction accuracy for different sequence lengths.

spatio-temporal emotional information users' emotion trajectory provides, which is the basis for constructing spatio-temporal emotional dependencies. By accurately constructing dependencies, our model achieves better prediction results.

5 Conclusion and discussion

Spatio-temporal emotion prediction is a major challenge for behavioral and psychological research. It is also important for promoting urban development, boosting the happiness of urban residents, and optimizing the public emotional experience. Existing prediction methods have difficulty effectively constructing multidimensional spatio-temporal emotion dependence. Therefore, human activities and emotional states are difficult to accurately predict. In this case, we propose a new prediction approach coupling an attention mechanism and BiLSTM to predict users' location emotions.

With the support of the geotagged image data, the proposed method achieves 75.21% accuracy in predicting location emotion. By coupling the attention mechanism with BiLSTM, multidimensional spatio-temporal emotion dependence can be generated. Based on this, the proposed method can not only differentiate four significantly different emotional states but also predict users' locations with a high precision and effectiveness. Existing research methods can only conduct emotion states or location alone. Compared with these methods, our approach can predict location and emotion at the same time and attach emotional attributes to location-based services. The results provide a novel approach for personalizing emotion-optimized services, enhancing service quality, and demonstrating humanistic care.

The approach in this paper has the following shortcomings. First, the data sources are limited, and more kinds of data, such as text, audio, and physiological electricity, can be introduced in the future. The integration of real-time prediction models will be explored, enabling adaptation to dynamic changes in user location and context. For example, a traffic prediction system could continuously update recommendations based on real-time data, including user location, traffic conditions, and weather patterns. Machine learning techniques, such as reinforcement learning algorithms, will also be employed to refine predictions and ensure more personalized and accurate suggestions for users. Additionally, given the privacy concerns associated with location-based services, differential privacy techniques will be investigated to add noise to location data, protecting user identities. Other privacy-enhancing methods, such as federated learning, will be explored to allow data processing on user devices, minimizing the need to share raw location data with centralized servers. Second, there is a lack of consideration of geographic environmental information. The influence of environmental information on the prediction will be further explored in the future. Environmental factors, such as weather conditions, public events, and surrounding infrastructure, play an important role in location-based predictions. We intend to collect fine-grained environmental data and integrate these elements as node attributes in the model. For instance, weather data (e.g., temperature, precipitation, humidity) could be included as time-dependent features at each node. Event schedules (e.g., concerts, sports events) could be encoded as categorical features associated with event-related nodes. These additions would enhance the model's ability to account for environmental factors and improve the accuracy of location-based predictions. Third, our model can only distinguish four categories of emotion, and it is difficult to calculate emotion attributes quantitatively. We will further improve the prediction methods, such as predicting the valence and intensity of emotions according to the Russel ring model. In terms of scalability, the dataset used in this work was constrained by the limitations of the Sina Weibo platform, which prevents further data collection[1]. We plan to extend this research by collecting additional data from other social media platforms, such as Twitter, to test the model's performance on a larger and more diverse dataset. Emotional expressions in online environments may diverge from those in offline contexts. Constrained by the limitations of online platforms and affected by expressive biases, online users are inclined to exaggerate their emotions (Caspi and Etgar, 2023). To mitigate this limitation, we plan to incorporate real-world offline data in future work. Through comparative analysis, we aim to correct the potential biases present in online datasets.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

WJ: Writing – original draft, Investigation, Visualization. YW: Data curation, Writing – original draft. XS: Conceptualization, Methodology, Writing – original draft, Formal analysis, Investigation. XZ: Investigation, Writing – original draft. XL: Validation, Visualization, Writing – review & editing. YL: Writing – review & editing. ZWa: Investigation, Writing – review & editing. ZWe: Validation, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China under Grant (42301490, 42171403) and the National Undergraduate Innovation and Entrepreneurship Training Program under Grant (202410370018).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2025.1641623/full#supplementary-material

[1] https://open.weibo.com/wiki/

References

Abdullah, M. , Hadzikadicy, M., and Shaikhz, S. (2018). SEDAT: sentiment and emotion detection in Arabic text using CNN-LSTM deep learning., in 2018 17th IEEE international conference on machine learning and applications (ICMLA), 835–840. doi: 10.1109/ICMLA.2018.00134

Bao, Y., Huang, Z., Li, L., Wang, Y., and Liu, Y. (2021). A bilstm-cnn model for predicting users' next locations based on geotagged social media. *Int. J. Geogr. Inf. Sci.* 35, 639–660. doi: 10.1080/13658816.2020.1808896

Ben-Gal, I. (2008). "Bayesian networks" in Encyclopedia of statistics in quality and reliability (Hoboken, New Jersey, USA: John Wiley & Sons, Ltd).

Bock, H.-H. (2007). "Clustering methods: a history of k-means algorithms" in Selected contributions in data analysis and classification. eds. P. Brito, G. Cucumel, P. Bertrand and F. de Carvalho (Berlin, Heidelberg: Springer), 161–172.

Campos, V., Jou, B., and Giró-i-Nieto, X. (2017). From pixels to sentiment: fine-tuning CNNs for visual sentiment prediction. *Image Vis. Comput.* 65, 15–22. doi: 10.1016/j.imavis.2017.01.011

Caspi, A., and Etgar, S. (2023). Exaggeration of emotional responses in online communication. *Computers in Human Behavior*, 146:107818. doi: 10.1016/j.chb.2023.107818

Chen, Y., Huang, Z., Pei, T., and Liu, Y. (2018). Hispatialcluster: a novel high-performance software tool for clustering massive spatial points. *Trans. GIS* 22, 1275–1298. doi: 10.1111/tgis.12463

Cittadini, R., Tamantini, C., Scotto di Luzio, F., Lauretti, C., Zollo, L., and Cordella, F. (2023). Affective state estimation based on Russell's model and physiological measurements. *Sci. Rep.* 13:9786. doi: 10.1038/s41598-023-36915-6

Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J. (1995). Active shape models-their training and application. *Comput. Vis. Image Underst.* 61, 38–59. doi: 10.1006/cviu.1995.1004

Douc, R., Moulines, E., Priouret, P., and Soulier, P. (2018). Markov Chains. Cham: Springer International Publishing.

Eddy, S. R. (1996). Hidden Markov models. *Curr. Opin. Struct. Biol.* 6, 361–365. doi: 10.1016/S0959-440X(96)80056-X

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996) A Density-based algorithm for discovering clusters in large spatial databases with noise., in Proceedings of the second international conference on knowledge discovery and data mining, (Portland, Oregon: AAAI Press), 226–231

Fan, X., Guo, L., Han, N., Wang, Y., Shi, J., and Yuan, Y.. (2018). A deep learning approach for next location prediction., in 2018 IEEE 22nd international conference on computer supported cooperative work in design ((CSCWD)), 69–74

Grover, A., and Leskovec, J. (2016). "node2vec: scalable feature learning for networks" in Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (New York, NY, USA: Association for Computing Machinery), 855–864.

Guan, W., Gao, H., Yang, M., Li, Y., Ma, H., Qian, W., et al. (2014). Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events. *Phys. A Stat. Mech. Appl.* 395, 340–351. doi: 10.1016/j.physa.2013.09.059

Halder, S., Lim, K. H., Chan, J., and Zhang, X. (2021). Transformer-based multi-task learning for queuing time aware next POI recommendation. In K. Karlapalem, H. Cheng, N. Ramakrishnan, R. K. Agrawal, P. K. Reddy and J. Srivastavaet al. (Eds.), Advances in Knowledge Discovery and Data Mining (pp. 510–523). Cham, Switzerland: Springer International Publishing

Han, B., Cook, P., and Baldwin, T. (2014). Text-based twitter user geolocation prediction. *J. Artif. Intell. Res.* 49, 451–500. doi: 10.1613/jair.4200

Ho, N. L., and Hui Lim, K. (2022). POIBERT: a transformer-based model for the tour recommendation problem. *IEEE Int. Conf. Big Data (Big Data)* 2022, 5925–5933. doi: 10.1109/BigData55660.2022.10020467

Hong, Y., Martin, H., and Raubal, M.. (2022). How do you go where? Improving next location prediction by learning travel mode information using transformers. Proceedings of the 30th International Conference on Advances in Geographic Information Systems, 1–10

Jiang, W., Wang, Y., Tsou, M.-H., and Fu, X. (2015). Using social media to detect outdoor air pollution and monitor air quality index (AQI): a geo-targeted spatiotemporal analysis framework with Sina Weibo (Chinese twitter). *PLoS One* 10:e0141185. doi: 10.1371/journal.pone.0141185

Karatzoglou, A., Jablonski, A., and Beigl, M. (2018). "A Seq2Seq learning approach for modeling semantic trajectories and predicting the next location" in Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems (New York, NY, USA: Association for Computing Machinery), 528–531.

Kooij, J. F. P., Schneider, N., Flohr, F., and Gavrila, D. M. (2014). Context-based pedestrian path prediction., in Computer Vision – ECCV 2014, eds. D. Fleet, T. Pajdla, B. Schiele and T. Tuytelaars (Cham: Springer International Publishing), 618–633

Li, D., Li, Y., and Wang, S. (2020). Interactive double states emotion cell model for textual dialogue emotion prediction. *Knowl.-Based Syst.* 189:105084. doi: 10.1016/j.knosys.2019.105084

Li, L., Zhou, B., Ren, W., and Lian, J. (2021). Review of pedestrian trajectory prediction methods. *Chin. J. Intell. Sci. Technol.* 3, 399–411. doi: 10.11959/j.issn.2096-6652.202140

Liu, H., Zhu, Y., Wang, C., Ding, J., Yu, J., and Tang, F. (2023). Incorporating heterogeneous user behaviors and social influences for predictive analysis. *IEEE Trans Big Data* 9, 716–732. doi: 10.1109/TBDATA.2022.3193028

Liu, J., and Wu, X. (2020). Real-time multimodal emotion recognition and emotion space labeling using LSTM networks. *Fudan Xuebao Ziran Kexue Ban* 59, 565–574.

Mai, S., Hu, H., Xu, J., and Xing, S. (2022). Multi-fusion residual memory network for multimodal human sentiment comprehension. *IEEE Trans. Affect. Comput.* 13, 320–334. doi: 10.1109/TAFFC.2020.3000510

Nicolaou, M. A., Gunes, H., and Pantic, M. (2012). Output-associative RVM regression for dimensional and continuous emotion prediction. *Image Vis. Comput.* 30, 186–196. doi: 10.1016/j.imavis.2011.12.005

Niu, K., Cheng, C., Chang, J., Zhang, H., and Zhou, T. (2019). Real-time taxi-passenger prediction with L-CNN. *IEEE Trans. Veh. Technol.* 68, 4122–4129. doi: 10.1109/TVT.2018.2880007

Niu, Z., and Qiu, X.. (2010). Facial expression recognition based on weighted principal component analysis and support vector machines., in 2010 3rd international conference on advanced computer theory and engineering (ICACTE), V3-174-V3-178

Pan, X., Ying, G., Chen, G., Li, H., and Li, W. (2019). A deep spatial and temporal aggregation framework for video-based facial expression recognition. *IEEE Access* 7, 48807–48815. doi: 10.1109/ACCESS.2019.2907271

Pantic, M., and Stewart, M. (2007). "Machine analysis of facial expressions" in Face Recognition (Vienna, Austria: I-Tech Education and Publishing), 377–416. doi: 10.5772/4847

Pavlovic, V., Rehg, J. M., and MacCormick, J. (2000). "Learning switching linear models of human motion," in Advances in neural information processing systems (Cambridge: MIT Press).

Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). "DeepWalk: online learning of social representations" in Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining (New York, NY, USA: Association for Computing Machinery), 701–710.

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714

Sang, H., Chen, W., Wang, H., and Wang, J. (2024). Mstcnn: multi-modal spatio-temporal convolutional neural network for pedestrian trajectory prediction. *Multimed. Tools Appl.* 83, 8533–8550. doi: 10.1007/s11042-023-15989-4

Saxena, D., and Cao, J. (2022). Multimodal spatio-temporal prediction with stochastic adversarial networks. *ACM Trans. Intell. Syst. Technol.* 13:18:1-18:23. doi: 10.1145/3458025

Sun, N., Li, Q., Huan, R., Liu, J., and Han, G. (2019). Deep spatial-temporal feature fusion for facial expression recognition in static images. *Pattern Recogn. Lett.* 119, 49–61. doi: 10.1016/j.patrec.2017.10.022

Sun, W., Zhao, H., and Jin, Z. (2018). A visual attention based ROI detection method for facial expression recognition. *Neurocomputing* 296, 12–22. doi: 10.1016/j.neucom.2018.03.034

Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., and Mei, Q. (2015). "LINE: large-scale information network embedding" in Proceedings of the 24th international conference on world wide web (Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee), 1067–1077.

Terzis, V., Moridis, C. N., and Economides, A. A.. (2010). Measuring instant emotions during a self-assessment test: the use of FaceReader. In Proceedings of the 7th international conference on methods and techniques in behavioral research, 1–4. doi: 10.1145/1931344.1931362

Thavareesan, S., and Mahesan, S.. (2020). Sentiment lexicon expansion using Word2vec and fastText for sentiment prediction in Tamil texts., in 2020 Moratuwa engineering research conference (MERCon), 272–276

Wang, P., Wang, H., Zhang, H., Lu, F., and Wu, S. (2019). A hybrid Markov and LSTM model for indoor location prediction. *IEEE Access* 7, 185928–185940. doi: 10.1109/ACCESS.2019.2961559

Wang, H., Yang, Z., and Shi, Y. (2019). Next location prediction based on an Adaboost-Markov model of mobile users. *Sensors* 19:1475. doi: 10.3390/s19061475

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometr. Intell. Lab. Syst.* 2, 37–52. doi: 10.1016/0169-7439(87)80084-9

Xue, H., Huynh, D. Q., and Reynolds, M.. (2018). SS-LSTM: a hierarchical LSTM model for pedestrian trajectory prediction., In 2018 IEEE winter conference on applications of computer vision (WACV), 1186–1194

Yang, Z., and Xu, W. (2023). Who post more negatively on social media? A large-scale sentiment analysis of Weibo users. *Curr. Psychol.* 42, 25270–25278. doi: 10.1007/s12144-022-03616-8

Ying, J. J.-C., Lee, W.-C., Weng, T.-C., and Tseng, V. S. (2011). "Semantic trajectory mining for location prediction" in Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems (New York, NY, USA: Association for Computing Machinery), 34–43.

Yolcu, G., Oztel, I., Kazan, S., Oz, C., Palaniappan, K., Lever, T. E., et al. (2019). Facial expression recognition for monitoring neurological disorders based on convolutional neural network. *Multimed. Tools Appl.* 78, 31581–31603. doi: 10.1007/s11042-019-07959-6

Yu, C., Ma, X., Ren, J., Zhao, H., and Yi, S. (2020). Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In A. Vedaldi, H. Bischof, T. Brox and J.-M. Frahm (Eds.), Computer Vision – ECCV 2020 (pp. 507–523). Cham, Switzerland: Springer International Publishing

Zhang, Q., Gao, T., Liu, X., and Zheng, Y. (2020). Public environment emotion prediction model using LSTM network. *Sustainability* 12:1665. doi: 10.3390/su12041665

Zhang, R., Guo, J., Jiang, H., Xie, P., and Wang, C.. (2019). Multi-task learning for location prediction with deep multi-model ensembles., In 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 1093–1100

Zhou, B., Liu, J., Cui, S., and Zhao, Y. (2024). A large-scale spatio-temporal multimodal fusion framework for traffic prediction. *Big Data Min. Anal.* 7, 621–636. doi: 10.26599/BDMA.2024.9020020

Zhu, H., and Gao, Q. (2015). Review on "emotional turn" and emotional geographies in recent western geography. *Geogr. Res.* 34, 1394–1406. doi: 10.11821/dlyj201507017