Toward Ubiquitous Interaction-Attentive and Extreme-Aware Crowd Activity Level Prediction

HUIQUN HUANG[†], University of Connecticut, USA
XI YANG[†], University of Connecticut, USA
SUINING HE[*†], University of Connecticut, USA
MAHAN TABATABAIE, University of Connecticut, USA

Accurate prediction of citywide crowd activity levels (CALs), *i.e.*, the numbers of participants of citywide crowd activities under different venue categories at certain time and locations, is essential for the city management, the personal service applications, and the entrepreneurs in commercial strategic planning. Existing studies have not thoroughly taken into account the complex spatial and temporal interactions among different categories of CALs and their extreme occurrences, leading to lowered adaptivity and accuracy of their models. To address above concerns, we have proposed `IE-CALP`, a novel spatio-temporal **I**nteractive attention-based and **E**xtreme-aware model for **C**rowd **A**ctivity **L**evel **P**rediction. The tasks of `IE-CALP` consist of *(a)* forecasting the spatial distributions of various CALs at different city regions (spatial CALs), and *(b)* predicting the number of participants per category of the CALs (categorical CALs). To realize above, we have designed a novel spatial CAL-POI interaction-attentive learning component in `IE-CALP` to model the spatial interactions across different CAL categories, as well as those among the spatial urban regions and CALs. In addition, `IE-CALP` incorporate the multi-level trends (*e.g.*, daily and weekly levels of temporal granularity) of CALs through a multi-level temporal feature learning component. Furthermore, to enhance the model adaptivity to extreme CALs (*e.g.*, during extreme urban events or weather conditions), we further take into account the *extreme value theory* and model the impacts of historical CALs upon the occurrences of extreme CALs. Extensive experiments upon a total of 738,715 CAL records and 246,660 POIs in New York City (NYC), Los Angeles (LA), and Tokyo have further validated the accuracy, adaptivity, and effectiveness of `IE-CALP`'s interaction-attentive and extreme-aware CAL predictions.

CCS Concepts: • **Information systems** → *Mobile information processing systems.*

Additional Key Words and Phrases: Crowd activity level, spatio-temporal interaction, points-of-interest (POI), extreme-aware prediction

ACM Reference Format:
Huiqun Huang, Xi Yang, Suining He, and Mahan Tabatabaie. 2024. Toward Ubiquitous Interaction-Attentive and Extreme-Aware Crowd Activity Level Prediction. *ACM Trans. Intell. Syst. Technol.* 0, 0, Article 0 (2024), 25 pages. https://doi.org/10.1145/1122445.1122456

[*]Corresponding author.
[†]Equal contribution.

Authors' addresses: Huiqun Huang, University of Connecticut, School of Computing, Storrs, CT, USA; Xi Yang, University of Connecticut, School of Computing, Storrs, CT, USA; Suining He, suining.he@uconn.edu, University of Connecticut, School of Computing, Storrs, CT, USA; Mahan Tabatabaie, University of Connecticut, School of Computing, Storrs, CT, USA.

ACM 2157-6904/2024/0-ART0
https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Accurate prediction of the citywide *crowd activity levels* (CALs), *i.e.*, the numbers of participants in different categories of activities (*e.g.*, dining or shopping venues) at certain time and locations, is essential for many smart city planning and mobility management services [1, 13, 42, 55, 61]. For instance, knowing the spatial and temporal trends of each category of citywide crowd activities in advance would enable venue recommendation [3] and assist decision making [3] of personal crowd activities, thus effectively reducing the commute overhead during the peak hours of certain activities.

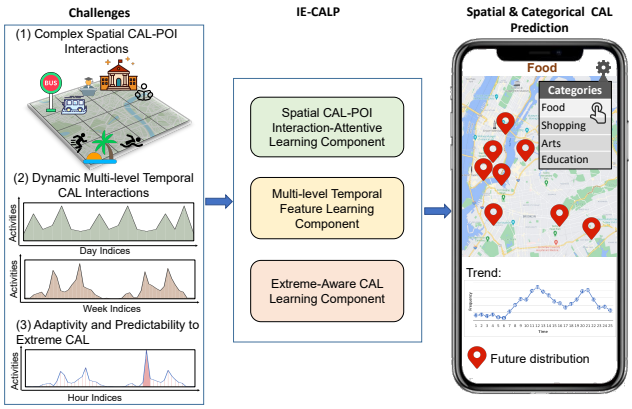


Fig. 1. Research challenges, motivations, and ubiquitous applications of citywide CAL prediction.

Despite the prior studies on the crowd activity prediction [22, 28, 40], as illustrated in Fig. 1, there remain the following three major technical challenges to be solved before a practical spatio-temporal CAL prediction system can be effectively deployed:

(1) **Complex Spatial CAL-POI Interactions:** The spatial distributions of various categories of crowd activities are highly correlated with different city regions, their respective points-of-interest (POIs), and their mutual spatial proximity and temporal correlations. On the other hand, the diverse functions of a city region can lead to co-occurrence of related crowd activities. How to extract the spatial *interactions* among different categories of crowd activities, as well as those among crowd activities and the region POIs, is essential for accurate forecasting of incoming crowd activities, which, however, remains largely unexplored.

(2) **Multi-level Temporal CAL Interactions:** Citywide CALs might demonstrate multiple short- and long-term recurrent patterns due to the diverse commute routines of the crowds, likely due to the related contexts and correlated categories of the location venues. For instance, the everyday crowd activities in both food (dining) and transportation categories in Tokyo might demonstrate similar peaks during the rush hours of morning and late afternoon, implying multi-level recurrent mobility patterns. How to capture and leverage the temporal dynamics of each category of crowd activities, and the temporal correlations among different categories of crowd activities is essential but challenging for the crowd activity prediction.

(3) **Adaptivity and Predictability to Extreme CAL:** In addition to the regular temporal dynamics, CALs may surge or drop within a short period due to impacts of various complex external factors, leading to rare and extreme occurrences of certain crowd activities that are much higher than majority of the historical records. Since these extreme crowd activities are relatively rare within the entire historical crowd activities, many existing deep learning

techniques (including long short-term memory [43, 51]) might only capture the overall trends of major crowd activities, but overlook these extreme crowd activities due to the *imbalanced* distributions of these extreme records, namely the problem of *extreme crowd activities prediction*. There exists a pressing need to actively and proactively learn and capture these extreme crowd activities within the prediction model, which, however, has not been thoroughly studied in the previous literature.

To address the above challenges concerning the ubiquitous and urban computing communities, we propose IE-CALP, a novel spatio-temporal **I**nteractive attention-based and **E**xtreme-aware model for **C**rowd **A**ctivity **L**evel **P**rediction. Specifically, our IE-CALP aims at *jointly* predicting *(a)* the spatial distributions of crowd activity levels at different categories, namely the *spatial CALs*, and *(b)* the aggregate number of participants with respect to each category of the crowd activities (including the extreme distributions), namely the *categorical CALs*. Towards the *joint* predictions of above CALs, we have made the following three major technical contributions:

(1) **Spatio-Temporal CAL-POI Graph Interaction-Attentive Learning:** To enhance the model learnability given Challenges *(1)* and *(2)*, we have formulated the geospatial interactions across the region POIs and CALs into *graph interaction*, and quantified the interactions across POIs and the regional crowd activities. Specifically, we formulate the city regions as nodes with the spatial proximity and temporal correlations as the edges, and design a novel *CAL-POI graph interaction learning* based on a multi-graph interactive attention mechanism. Then we further evaluate the spatial and temporal graph correlations among different categories of crowd activities.
(2) **Extreme-aware CAL Model Co-Design & Adaptive Learning:** To adapt to the impacts of the extreme distributions and the unbalanced frequencies of crowd activities as discussed in Challenge *(3)*, we have designed a novel extreme-aware CAL learning component with an adaptive loss function based on the *Extreme Value Theory* [9]. Through the novel co-designs with the graph interaction-attentive learning, our IE-CALP adaptively captures and integrates the temporal occurrence patterns of extreme CALs, enabling a ubiquitous extreme-aware CAL prediction framework.
(3) **Extensive Real-world Data Analysis & Experimental Studies:** We have conducted extensive data analytics and experimental studies on a total of 738,715 CAL records and 246,660 POIs from the cities of New York City (NYC), Los Angeles (LA), and Tokyo. Our extensive experimental studies have demonstrated that IE-CALP outperforms the other baseline approaches in terms of spatial and categorical CAL prediction accuracy (including extreme categorical CALs), with substantial error reduction by 24.12% on average.

• **System Framework:** We overview the system framework of IE-CALP in Fig. 2, which consists of three major technical designs.

(a) **Spatial CAL-POI Interaction-Attentive Learning Component**: In this module, IE-CALP takes in the spatial CALs, and captures the important regions for each category of crowd activities by evaluating the spatial correlations between city regions and the historical crowd activities. Specifically, IE-CALP models the city regions into graphs where each region is considered as nodes connected by the spatial, temporal, and POI similarity graphs. After capturing the spatial distributions of each category of crowd activities, and their interactions throughout the above graphs, IE-CALP jointly predicts the spatial CALs and categorical CALs.
(b) **Multi-level Temporal Feature Learning Component:** This component further leverages the gated recurrent unit (GRU) to model the recurrent and multi-level temporal patterns for

each category of CALs (*e.g.*, daily and weekly in our studies), and outputs another set of categorical CAL predictions.

(c) **Extreme-Aware CAL Learning Component:** To further capture the temporal characteristic of extreme CALs, IE-CALP models the recurrent occurrence probabilities of different categories of extreme CALs from the windows of historical CALs, quantifies the varying influences of historical extreme CALs upon the future CALs, and finally outputs the final categorical CAL predictions.

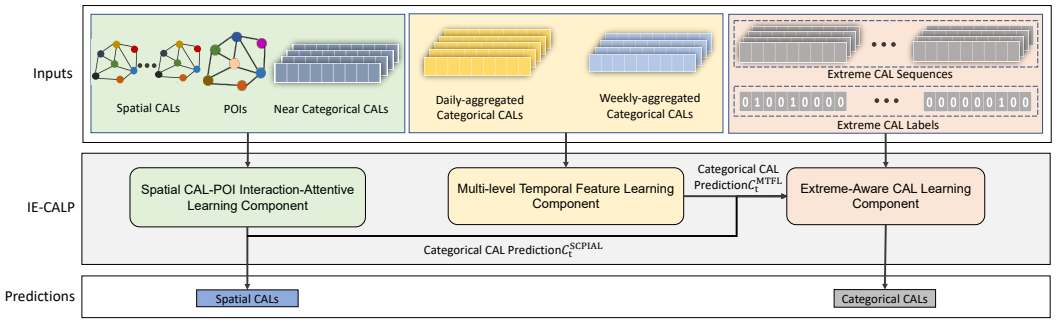


Fig. 2. Overview of the IE-CALP framework, with three categories of inputs and processes them with three novel modules.

• **Broader Societal Implications:** Our research studies and the developed IE-CALP as well as insights on the ubiquitous crowd activity levels (CALs) based on ubiquitous check-in analytics can pave new directions in ubiquitous/mobile/urban app designs and interactive user mobility analysis. While our studies here focus on crowd mobility data [15] such as social network check-ins, our core model and methodologies are general, and can be easily extended to other ubiquitous sensing modalities (such as cellular [40]), mobility platforms (say, bike/ride/vehicle sharing [34]), and urban event datasets [28, 56, 58] (such as anomalies or crimes). Our demonstrated joint predictions of both spatial distributions and extreme activities will benefit the city planners and many other urban computing practitioners in ubiquitous computing, urban applications [60] and system development, as illustrated in Fig. 1.

• **Paper Organization:** The rest of the paper is organized as follows. We review the related work in Sec. 2. Then we overview the data used in this study, importance concepts, and problem formulation in Sec. 3. We further detail the core designs of the interaction-attentive learning module in Sec. 4, and the extreme-aware learning module in Sec. 5. We present our experimental studies and results in Sec. 6, and finally conclude in Sec. 7.

2 RELATED WORK

We briefly overview the related work in the following two major categories.

• **Deep Learning For Ubiquitous Crowd Mobility Analytics:** Deep learning techniques have emerged as the effective tools for many big data applications, including crowd mobility analytics [4, 25, 52–54, 56]. Various spatio-temporal models [14, 32, 34, 35] have been studied for crowd mobility analytics, where graph neural networks have been considered in prediction of traffic volume [11, 16, 19, 44, 45, 59], bike flow [7, 24], and ride-hailing demand [18, 23]. Different from these studies [6, 20, 47], we have studied and utilized the graph attention neural network to model the important POI-to-CAL and CAL-to-CAL interactions for the ubiquitous CAL prediction, enabling an interaction-attentive learning mechanism.

Table 1. Crowd activity data from NYC, LA, and Tokyo.

City	Description	Total Records	Geographic Bounding Box Range
NYC	5 grouped categories, *i.e.*, *(i)* arts & entertainment & nightlife spot & outdoors & recreation, *(ii)* travel & transportation, *(iii)* food & shop & service, *(iv)* education & professional, and *(v)* residence.	227,428	[40.55085°*N*, 40.98833°*N*], [73.68382°*W*, 74.27476°*W*]
LA		10,329	[33.6099916°*N*, 34.1813999°*N*], [117.5323391°*E*, 118.4984708°*E*]
Tokyo		500,958	[35.51018469°*N*, 35.86715042°*N*], [139.4708776°*E*, 139.9125931°*E*]

Table 2. POI data from NYC, LA, and Tokyo.

City	Description	Total Records
NYC	9 categories, *i.e.*, commercial, cultural, education, health, recreational, religious, residential, social service, and transportation.	29,298
LA	6 categories, *i.e.*, arts & recreation, communication, education, health, social service, and transportation.	27,385
Tokyo	4 categories, *i.e.*, education, health, public transportation, and shop & food & service.	189,977

• **Extreme Detection and Prediction for Time Series:** Extreme values generally occur in various time-series datasets [9], including climate and network stream records. To mitigate the influence upon the time-series modeling and estimation, various statistical analysis and learning approaches have been proposed [29]. Lozano *et al.* [33] incorporated extreme value modeling for modeling extreme climate events. To detect the extreme values dynamically, Siffer *et al.* [39] and Na *et al.* [37] further proposed the *Extreme Value Theory* and *Local Outlier Factor* approaches, respectively. They studied how to detect outliers in data streams without manual determination of the thresholds and assumption of the underlying distributions. However, these approaches might only handle the univariate data streams. Manzoor *et al.* [36] further proposed a density-based ensemble outlier detector for high-dimensional feature-evolving streams. Su *et al.* [41] proposed a stochastic recurrent neural network for multivariate time series anomaly detection. However, these above studies could only detect the anomalies after the occurrences of the events rather than prediction, which might largely delay the response actions given certain anomalies in real-world applications. Besides extreme detection, how to forecast the extreme values has also attracted attention recently [12, 48, 50]. Deng *et al.* [10] considered the high-frequency and low frequency components within the time-series data to yield high prediction accuracy. Huang *et al.* [26] proposed predicting the anomalies of each region of a city by modeling the temporal dependency of anomaly occurrences of different regions. Inspired by the aforementioned studies [9, 12, 41], through integration of extreme value theory (EVT) with spatio-temporal CAL learning, we have proposed a novel model co-design which combines the learnability of novel interaction-attentive learning as well as the extreme-awareness of the EVT-based loss function. Our results show that `IE-CALP` models the extreme CALs and enhances the model prediction accuracy and adaptivity.

3 DATASETS, CONCEPTS, & PROBLEM DEFINITIONS

In this section, we present the details of datasets in Sec. 3.1, and introduce the important concepts of this study in Sec. 3.2. After that, we present the problem formulation in Sec. 3.3, and show the data analysis in Sec. 3.4.

3.1 Datasets Studied

In this study, we utilize the urban POI as well as the crowd activity level (CAL) data from cities of New York City (NYC), Los Angeles (LA), and Tokyo to model the citywide crowd activity

Table 3. Summary of the key symbols and their definitions.

Symbols	Definitions	Symbols	Definitions
M	Number of POI categories.	R	Number of regions.
T	Number of the time units within a day.	H	Hours of one time unit.
l	Time interval index of categorical CAL.	l'	time interval index of spatial CAL.
e	Each time interval in spatial CAL is $e \times H$ hours.	N	Number of both the categorical CAL and spatial CAL categories.
t'	The target time interval of spatial CAL.	t	The target time interval of categorical CAL.
L	Number of time intervals of near history data.	$\mathbf{C}_{l,n}$	Categorical CAL of category n in time interval l.
$\mathbf{C}_l$	Categorical CAL of N categories in time interval l.	$\mathbf{C}^{\texttt{Near}}$	Categorical CAL of N categories during the past L time intervals.
$\mathbf{C}^{\texttt{Daily}}$	P time intervals of categorical CAL which in the same time interval within a day as the target time interval t from the past P days.	$\mathbf{C}^{\texttt{Weekly}}$	Q time intervals of categorical CAL which in the same time interval within a day and the same day within a week as the target time interval t from the past Q weeks.
$\mathbf{S}$	The spatial CAL of each categories in all regions during the past L intervals.	$\mathbf{E}_k$	The T time intervals of extreme CAL sequence of the previous k^{th} day of the day of the target time step t.
$\mathbf{Q}_k$	The extreme label vector at time interval of $(t - k \times T)$.	$\mathbf{A}^{\texttt{Dis}}$	The region-to-region geographical distance adjacency matrix.
$\mathbf{A}^{\texttt{Tmp}}$	The region-to-region temporal CAL correlation adjacency matrix.	$\mathbf{A}^{\texttt{POI}}$	The region-to-region POI correlation adjacency matrix.
$\hat{\mathbf{C}}_t$	The categorical CAL prediction at time interval t.	$\hat{\boldsymbol{S}}_{t'}^{\texttt{SCPIAL}}$	The spatial CAL prediction at time interval t'.

participation. The details of the CAL and POI datasets are respectively summarized in Tables 1 and 2. We have further summarized the symbols and their definitions presented in this work in Table 3.

• *POI Data:* To model the spatial and temporal interactions with the CAL, we have harvested the Point of Interests (POIs) from the OpenStreetMap[1]. The POIs are classified into M distinct categories based on the function of the POIs (in our studies $M = 9$ for NYC, $M = 6$ for LA, and $M = 4$ for Tokyo).

• *Mobile CAL Data:* CAL venues reflect the types of crowd activities due to the urban region functions and the local venues. In this study, we consider a total of 5 CAL venue categories[2] for the three metropolitan cities: *(i)* arts & entertainment & nightlife spot & outdoors & recreation, *(ii)* travel & transportation, *(iii)* food & shop & service, *(iv)* education & professional, and *(v)* residential.

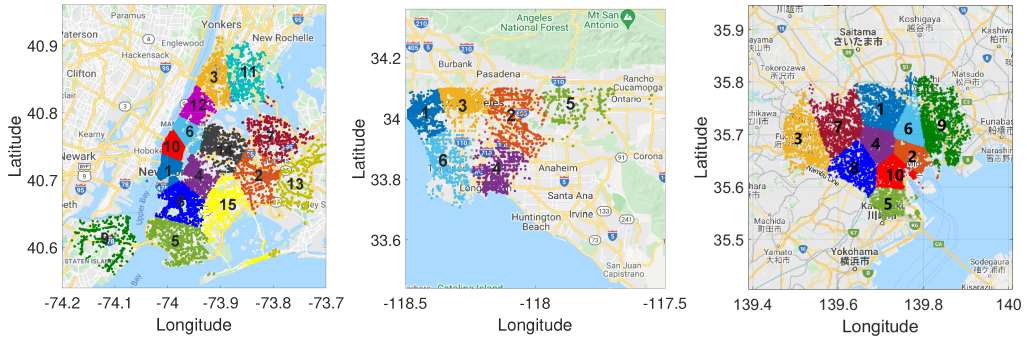
3.2 Important Concepts

Based on the above, we further present the important concepts in this study as follows.

Definition 3.1. Spatial Region Discretization. Conventional methods divide the city into rectangular distinct regions to model the citywide crowd mobility [57]. However, we notice that such partition method may results in the sparsity of data in some regions. To alleviate this problem and account for both the data density and geography distance of the regions, we partition the city

[1]https://www.openstreetmap.org

[2]https://developer.foursquare.com/docs/venues/categories



(a) NYC. (b) LA. (c) Tokyo.

Fig. 3. Spatial distributions and the corresponding regions of all the CALs of NYC (04/02/2012–02/16/2013), LA (12/01/2009–04/30/2010), and Tokyo (04/02/2012–02/16/2013).

into R disjoint *regions* by *k-means* clustering [21] using all the location data of the CALs within a city. We apply squared Euclidean distance to measure the spatial proximity across each pair of city regions. In this study, R is set as 15 for NYC, 6 for LA, and 10 for Tokyo due to different sizes of the cities studied. The partitioned regions are shown in the maps in Fig. 3, where we also provide the region indices within each city.

Definition 3.2. Time Discretization. For CAL computation, we discretize the time within a day equally into T time units. The length of each time unit is $H = 24/T$ hours (for instance, we set $T = 8$ and $H = 3$ in NYC). The categorical CALs and spatial CALs often occur with temporal patterns in different time ranges and scales. Therefore, to model the categorical CAL and spatial CAL predictions for ubiquitous CAL modeling, we set different lengths of time interval for predictions of the categorical CALs and spatial CALs.

Specifically, we have: *(i)* the length of the time interval for the categorical CALs prediction is one time unit, *i.e.*, H hours, and we denote such an time interval as l; *(ii)* the length of the time interval for spatial CALs prediction is e time units, *i.e.*, $e \cdot H$ hours, and we denote such an time interval as l'. In our current studies, for NYC, LA, and Tokyo, we respectively set (T, H, e) as (8, 3h, 8), (2, 12h, 2), and (24, 1h, 1) due to the sampling frequency variations from the three cities.

Definition 3.3. Categorical and Spatial CALs. We let the aggregate number of participants of CAL category $n \in \{1, \ldots, N\}$ within an interval l be the $\mathbf{C}_{l,n} \in \mathbb{R}^1$. Then we form the vector of categorical CAL as

$$\mathbf{C}_l = \{\mathbf{C}_{l,1}, \ldots, \mathbf{C}_{l,n}, \ldots, \mathbf{C}_{l,N}\} \in \mathbb{R}^N. \tag{1}$$

Based on above, we further denote

$$\mathbf{C}^{\mathsf{Near}} = \{\mathbf{C}_1, \ldots, \mathbf{C}_l, \ldots, \mathbf{C}_L\} \in \mathbb{R}^{L \times N}, \tag{2}$$

as the categorical CALs.

To evaluate the spatial distributions of different categories of CALs, we find the spatial occurrences of the N categories of CALs in all the R regions of the city in the past $L \in \mathbb{R}$ time intervals, *i.e.*, $\{t' - L, \ldots, t' - L + l', \ldots, t' - 1\}$, and denote the occurrences as $\boldsymbol{S} \in \mathbb{R}^{L \times R \times N}$.

Definition 3.4. Extreme CALs. To enable the extreme-aware CAL prediction, we further label the extreme CALs for the `IE-CALP`'s model learning. In particular, we denote the extreme CAL label $Q_{l,n} \in \mathbb{R}$ of the CAL category n in time interval l as an extreme value and label it as 1 if it is greater than θ ($\theta > 0$) percent of all the categorical CALs of the city, and 0 otherwise.

Given the labeled extreme CALs, we further construct the K historical extreme CAL sequence, and form the extreme CAL tuples, *i.e.*, $\{(\mathbf{E}_1, \mathbf{Q}_1), (\mathbf{E}_2, \mathbf{Q}_2), \ldots, (\mathbf{E}_K, \mathbf{Q}_K)\}$, as the model inputs, where

$$\mathbf{E}_k = [\mathbf{C}_{t-k\cdot T-T}, \ldots, \mathbf{C}_{t-k\cdot T-l}, \ldots, \mathbf{C}_{t-k\cdot T-1}] \in \mathbb{R}^{T\times N}, \tag{3}$$

denotes the T consecutive time intervals of categorical CALs on the previous k^{th} day ($k \in \{1, \ldots, K\}$) of the day of the target time step t, and

$$\mathbf{Q}_k = [Q_{t-k\cdot T,1} \ldots, Q_{t-k\cdot T,N}] \in \mathbb{R}^N, \tag{4}$$

denotes the k^{th} ($k \in \{1, \ldots, K\}$) historical extreme CAL label vectors at the time interval $(t - k \cdot T)$.

For instance, if we have $T = 8$, $H = 3$h, and $k = 2$ for the target time interval $t = [0\text{am}, 3\text{am}]$ on 09/05/2012, $\mathbf{E}_2$ is given by the sequence of the $T = 8$ categorical CALs from the interval $[0\text{am}, 3\text{am}]$, 09/02/2012 to $[9\text{pm}, 0\text{am}]$, 09/02/2012, and $\mathbf{Q}_2$ represents the extreme CAL labels of $[0\text{am}, 3\text{am}]$ on 09/03/2012. Through the historical extreme CAL sequences and label vectors, we can further enable `IE-CALP` to capture the occurrence patterns of the extreme CALs.

Definition 3.5. Daily-aggregated and Weekly-aggregated Categorical CALs. In order to conduct multi-level CAL learning to capture the daily and weekly patterns of the categorical CALs, we further construct the data vectors of P daily-aggregated categorical CALs from the past P days, denoted as

$$\mathbf{C}^{\texttt{Daily}} = \{\mathbf{C}_{t-P\cdot T}, \mathbf{C}_{t-(P-1)\cdot T}, \ldots, \mathbf{C}_{t-T}\} \in \mathbb{R}^{P\times N}, \tag{5}$$

from the past P days $\{t - P \cdot T, t - (P - 1) \cdot T, \ldots, t - T\}$. We also have Q weekly-aggregated categorical CALs from the past Q weeks, denoted as

$$\mathbf{C}^{\texttt{Weekly}} = \{\mathbf{C}_{t-7\cdot Q\cdot T}, \mathbf{C}_{t-7\cdot (Q-1)\cdot T}, \ldots, \mathbf{C}_{t-7\cdot T}\} \in \mathbb{R}^{Q\times N}, \tag{6}$$

from the time intervals $\{t - 7 \cdot Q \cdot T, t - 7 \cdot (Q - 1) \cdot T, \ldots, t - 7 \cdot T\}$.

• **CAL-POI Interaction Graph Formulation.** To evaluate the spatial distributions of different categories of CALs, we construct the graph $\mathbb{G}(\mathbf{V}, \boldsymbol{\mathcal{E}})$, where each node in $\mathbf{V} \in \mathbb{R}^R$ denotes each of the R regions in the city, and the weight of each edge in $\boldsymbol{\mathcal{E}}$ is based on the correlations across the city regions. We note that the designs of the effective encoded correlations among regions are important for the network parameter learning and accurate CAL prediction. We assign large weights to the edges between regions with similar CAL patterns.

To this end, we have designed the following three weighted adjacency matrices for $\boldsymbol{\mathcal{E}}$, as the *CAL-POI interaction graphs*, to further represent the following the spatial, temporal, and contextual interactions across CALs and POIs.

(1) **Spatial Region Distance:** To capture the correlations among the city regions that are geographically close to each other, we construct a distance adjacency matrix, denoted as $\boldsymbol{A}^{\texttt{Dis}} \in \mathbb{R}^{R\times R}$, where each element, $\boldsymbol{A}^{\texttt{Dis}}_{(i,j)}$, is given by the reverse of geo-distance (in km) between pairs of regions, $i \in \{1, 2, \ldots, R\}$ and $j \in \{1, 2, \ldots, R\}$. We set the diagonal elements as zeros, *i.e.*, $\boldsymbol{A}^{\texttt{Dis}}_{(i,i)} = 0$ for $i \in \{1, 2, \ldots, R\}$.

(2) **Temporal CAL Correlation:** As mentioned above, some regions may share similar temporal patterns of the CALs. With the historical aggregate occurrences of all CALs in different time intervals, we further construct the region-to-region temporal CAL correlation adjacency matrix to quantify the dynamic connectivities among regions. More specifically, we calculate aggregate occurrences of CALs over all historical intervals in each region. Then we find correlation adjacency matrix as $\boldsymbol{A}^{\texttt{Tmp}} \in \mathbb{R}^{R\times R}$. Each element $\boldsymbol{A}^{\texttt{Tmp}}_{(i,j)} \in \mathbb{R}$ denotes the Pearson coefficient [5] of the historical CALs between region i and region j. In particular, the total

number CALs in each time interval of the entire dataset in regions i and j is denoted as the vectors of $\mathbf{U}_i$ and $\mathbf{U}_j$. The Pearson coefficient between $\mathbf{U}_i$ and $\mathbf{U}_j$ is denoted as $\boldsymbol{A}^{\mathtt{Tmp}}_{(i,j)}$.

(3) **Region-to-Region POI Correlation:** The check-in patterns of a region may be dependent on the attractions of the POIs to and their potential interactions with the human visitations of the region. Therefore, we further construct another adjacency matrix $\boldsymbol{A}^{\mathtt{POI}} \in \mathbb{R}^{R\times R}$ to represent the *POI correlations* between regions where $\boldsymbol{A}^{\mathtt{POI}}_{(i,j)} \in \mathbb{R}$ denotes the cosine similarity of the POI distributions between regions i and j. In particular, we find the numbers of each of the M categories of POIs in regions i and j, and form the vectors of $\mathbf{B}_i \in \mathbb{R}^M$ and $\mathbf{B}_j \in \mathbb{R}^M$. Then we form the cosine similarity between $\mathbf{B}_i$ and $\mathbf{B}_j$ for $\boldsymbol{A}^{\mathtt{POI}}_{(i,j)}$. Similar to $\boldsymbol{A}^{\mathtt{Dis}}$ and $\boldsymbol{A}^{\mathtt{Tmp}}$, we set the diagonal elements $\boldsymbol{A}^{\mathtt{POI}}_{(i,i)} = 0$ for $i \in \{1, 2, \ldots, R\}$.

3.3 Problem Formulation

Based on the above concepts, we present the core problem formulation of IE-CALP. In particular, we take in the following as the inputs: (a) *spatial CALs and spatio-temporal CAL-POI interaction graphs:* including the spatial CALs $\boldsymbol{S} \in \mathbb{R}^{L\times R\times N}$ in the historical L time intervals, the region-to-region distance adjacency matrix $A^{\mathtt{Dis}} \in \mathbb{R}^{R\times R}$, the temporal CAL correlation adjacency matrix $A^{\mathtt{Tmp}} \in \mathbb{R}^{R\times R}$, and the POI similarity matrix $A^{\mathtt{POI}} \in \mathbb{R}^{R\times R}$; (b) *multi-level categorical CALs:* including the L historical time intervals of all categorical CALs, $\mathbf{C}^{\mathtt{Near}}$, the P time intervals of historical daily-aggregated categorical CALs, $\mathbf{C}^{\mathtt{Daily}}$, and the Q time intervals of historical weekly-aggregated categorical CALs, $\mathbf{C}^{\mathtt{Weekly}}$; and (c) *extreme categorical CALs:* the historical K-day extreme CAL sequence and label tuples, *i.e.*, $\{(\mathbf{E}_1, \mathbf{Q}_1), \ldots, (\mathbf{E}_K, \mathbf{Q}_K)\}$. Our IE-CALP aims at *jointly* predicting: *(i)* the spatial CALs $\hat{\boldsymbol{S}}_{t'} \in \mathbb{R}^{R\times N}$ and *(ii)* the categorical CALs $\hat{\mathbf{C}}_t \in \mathbb{R}^N$.

3.4 Spatial-Temporal CAL Data Analysis

We have further conducted large-scale spatial-temporal CAL data analysis to drive our model designs.

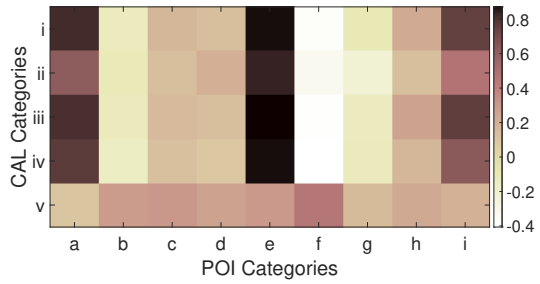
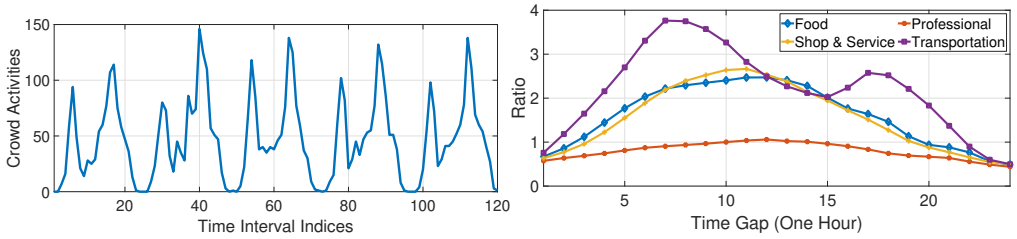


Fig. 4. The Pearson correlations among 5 categories of spatial CALs and 9 categories of POIs of NYC. The 5 categories of CALs are: (i) arts & entertainment & nightlife spot & outdoors & recreation, (ii) travel & transportation, (iii) food & shop & service, (iv) education & professional, and (v) residence. The 9 categories of POIs are: (a) commercial, (b) cultural, (c) education, (d) health, (e) recreational, (f) religious, (g) residential, (h) social service, and (i) transportation.

• **Spatial Interactions of CALs:** We show the spatial Pearson correlations between the 5 categories of CALs and 9 categories of POIs in 15 regions of NYC in Fig. 4. We can see that different interactions between POIs and CALs. For instance, the recreational POIs have an overall higher correlations with the CALs than other POI categories, while the *arts & entertainment* and *food & shop & service* CALs are more correlated with POIs than other CAL categories. We can learn from Fig. 4 that there are spatial interactions between the POI distributions and the CALs, and that the

spatial interactions vary across different categories. Such spatial interactions should be carefully characterized for accurate CAL predictions.

• **Temporal Multi-level Interactions and Extreme CALs:** Fig. 5a shows the temporal sequence of CALs for transportation category in Tokyo during 12/03/2012–12/07/2012, demonstrating the recurrent routines in the CALs. We can observe the temporal interactions between the daily activity patterns. Fig. 5b shows the average ratios of the CALs of food, professional, shop & service, and transportation categories in Tokyo. The horizontal axis represents the time gap between two intervals, ranging from 1 to 24 time steps, where the time step is 1 hour for the data of Tokyo. We can observe temporal interactions between the multiple categories of activities. Fig. 5b motivates our multi-level interaction learning designs, where daily and weekly trends are further captured for IE-CALP's accurate prediction. We further illustrate the extreme occurrences of the transportation activities in Fig. 6, from which we could observe several extreme occurrences (say, exceeding 200 records) during the rush hours. Since the extreme CALs are largely imbalanced in the historical records, we further design the Extreme-aware CAL Learning Component to enhance the model adaptivity and prediction accuracy.



(a) Trends. (b) Average ratio curves of crowd activities.

Fig. 5. The (a) original crowd activities of transportation category of Tokyo from 12/03/2012 to 12/07/2012, and the (b) average ratio curves of each two crowd activities in the same category and have the same time gaps from 1 time step to 24 time steps, using the crowd activity data of food, professional, shop & service and transportation categories of Tokyo.

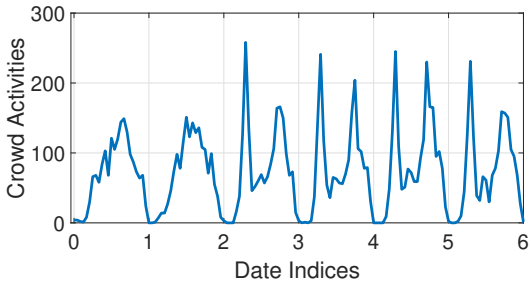


Fig. 6. Crowd activities from transportation category of Tokyo from 05/12/2012 to 05/18/2012.

4 INTERACTION-ATTENTIVE LEARNING DESIGNS

To realize the extreme-aware and interaction-attentive prediction, we have designed the core architecture of IE-CALP, as illustrated in Fig. 7, which consists of the three essential modules, *i.e.*, spatial CAL-POI interaction-attentive learning, multi-level temporal feature learning, and extreme-aware CAL learning. In this section, we first show the *Spatial CAL-POI Interaction-Attentive Learning* in Sec. 4.1 and the *Multi-level Temporal Feature Learning* in Sec. 4.2,

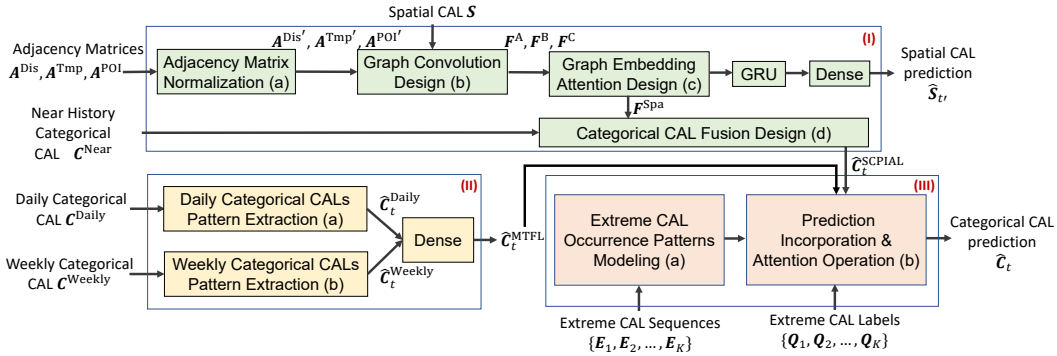


Fig. 7. The detailed design of IE-CALP framework. (I) The Spatial CAL-POI Interaction-Attentive Learning Component; (II) the Multi-level Temporal Feature Learning Component; and (III) the Extreme-aware CAL Learning Component.

4.1 Spatial CAL-POI Interaction-Attentive Learning Component

• **Motivations & Component Overview.** In this component, we leverage the regional POI-CAL interactions for both the spatial CALs (shown in Fig. 8) and categorical CALs (detailed in Fig. 9) prediction. In particular, we formulate three CAL-POI interaction graphs as discussed in Sec. 3.1, and then we predict the spatial CAL $\hat{\boldsymbol{S}}_{t'}$ via *(a)* adjacency matrix normalization, *(b)* graph convolution, and *(c)* graph embedding attention components. The spatial features $\boldsymbol{F}^{\text{Spa}}$ from the graph embedding attention component are fused with historical categorical CALs $\mathbf{C}^{\text{Near}}$ by *(d)* the categorical CAL fusion design to compute the categorical CAL prediction $\hat{\mathbf{C}}_t^{\text{SCPIAL}}$. Details of each component are presented as follows.

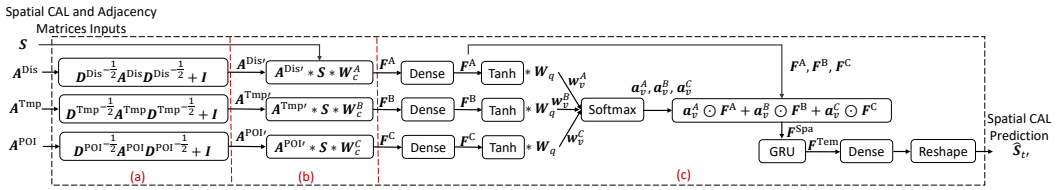


Fig. 8. Designs of Spatial CAL-POI Interaction-Attentive Learning Component in predicting the spatial CALs: (a) adjacency matrix normalization; (b) graph convolution; and (c) graph embedding attention.

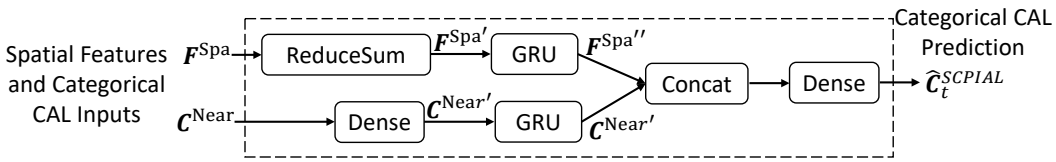


Fig. 9. The design of Spatial CAL-POI Interaction-Attentive Learning Component in predicting the categorical CALs.

• **Adjacency Matrix Normalization.** Given the three CAL-POI interaction graphs, we first normalize their adjacency matrix. Specifically, taking the distance adjacency matrix $\boldsymbol{A}^{\text{Dis}}$ as an example, we first normalize it by

$$\boldsymbol{A}^{\text{Dis}'} = \left(\boldsymbol{D}^{\text{Dis}}\right)^{-\frac{1}{2}} \boldsymbol{A}^{\text{Dis}} \left(\boldsymbol{D}^{\text{Dis}}\right)^{\frac{1}{2}} + \boldsymbol{I}, \tag{7}$$

where $\boldsymbol{D}^{\mathsf{Dis}} \in \mathbb{R}^{R\times R}$ is the degree matrix of the adjacency matrices of $\boldsymbol{A}^{\mathsf{Dis}}$ and $\boldsymbol{I} \in \mathbb{R}^{R\times R}$ is the identity matrix. Similarly, we normalize the $\boldsymbol{A}^{\mathsf{Tmp}}$ and $\boldsymbol{A}^{\mathsf{POI}}$, and obtain $\boldsymbol{A}^{\mathsf{Tmp}'}$ and $\boldsymbol{A}^{\mathsf{POI}'}$, respectively.

• **Graph Convolution Design.** To further capture the CAL-POI interactions, we have designed a graph convolutional network (GCN) [30], which takes in the normalized adjacency matrices and the spatial CALs, $\boldsymbol{S} = \{\boldsymbol{S}_{t'-L}, \ldots, \boldsymbol{S}_{t'-L+l'}, \ldots, \boldsymbol{S}_{t'-1}\}$, and generates three graph embeddings, *i.e.*, $\boldsymbol{F}^{\mathsf{A}}$, $\boldsymbol{F}^{\mathsf{B}}$, and $\boldsymbol{F}^{\mathsf{C}}$, respectively, for the three CAL-POI interaction graphs. Mathematically, the graph convolution is formulated by

$$\boldsymbol{F}^{\mathsf{A}}_{l'} = \boldsymbol{A}^{\mathsf{Dis}'}\boldsymbol{S}_{l'}\boldsymbol{W}^{\mathsf{A}}_{c}, \tag{8}$$

where $\boldsymbol{F}^{\mathsf{A}}_{l'} \in \mathbb{R}^{R\times L^{\mathsf{Emb}}}$ represents the resulting graph embedding of the spatial region distance graph in the time interval $l' \in \{t'-L, \ldots, t'-L+l', \ldots, t'-1\}$. Here we let $L^{\mathsf{Emb}} \in \mathbb{R}$ be the embedding length, $\boldsymbol{S}_{l'} \in \mathbb{R}^{R\times N}$ be the spatial CALs at time interval l, and $\boldsymbol{W}^{\mathsf{A}}_{c} \in \mathbb{R}^{N\times L^{\mathsf{Emb}}}$ be the trainable weight matrices. Similar to above, we obtain the graph embeddings, $\boldsymbol{F}^{\mathsf{B}}$ and $\boldsymbol{F}^{\mathsf{C}}$, for the temporal CAL correlation graph and region-to-region POI correlation graph, respectively.

• **Graph Embedding Attention Design.** Given the CAL-POI interaction graph embeddings, *i.e.*, $\boldsymbol{F}^{\mathsf{A}}$, $\boldsymbol{F}^{\mathsf{B}}$, and $\boldsymbol{F}^{\mathsf{C}}$, we have designed a graph embedding attention mechanism to differentiate the varying importance of different graph embeddings based on the occurrence of different CAL categories in each city region.

First, we apply a fully-connected layer whose number of units is set as L^{Emb} to further encode the spatial CALs, *i.e.,*

$$\boldsymbol{F}^{\mathsf{A}'} = \mathsf{Dense}\left(\boldsymbol{F}^{\mathsf{A}}\right). \tag{9}$$

We further transform the updated graph embedding $\boldsymbol{F}^{\mathsf{A}'} \in \mathbb{R}^{L\times R\times L^{\mathsf{Emb}}}$ through a nonlinear Tanh operation. Then, we apply one attention vector $\boldsymbol{W}_q \in \mathbb{R}^{L^{\mathsf{Emb}}\times 1}$ to obtain the embedding value $\boldsymbol{w}^{A}_{v} \in \mathbb{R}^{L\times R\times 1}$, *i.e.,*

$$\boldsymbol{w}^{A}_{v} = \mathsf{Tanh}\left(\boldsymbol{F}^{\mathsf{A}}\right)\boldsymbol{W}_q. \tag{10}$$

We integrate the same attention vector $\boldsymbol{W}_q$ for the other two graph embeddings, and obtain the updated embedding values, *i.e.*, $\boldsymbol{w}^{B}_{v}$ from $\boldsymbol{F}^{\mathsf{B}}$ and $\boldsymbol{w}^{C}_{v}$ from $\boldsymbol{F}^{\mathsf{C}}$, respectively.

We then have obtained the final attention scores, *i.e.*, $\boldsymbol{a}^{A}_{v}$, $\boldsymbol{a}^{B}_{v}$, $\boldsymbol{a}^{C}_{v} \in \mathbb{R}^{L\times R\times 1}$ by a Softmax function [7], *i.e.,*

$$\boldsymbol{a}^{A}_{v}, \boldsymbol{a}^{B}_{v}, \boldsymbol{a}^{C}_{v} = \mathsf{Softmax}\left(\boldsymbol{w}^{A}_{v}, \boldsymbol{w}^{B}_{v}, \boldsymbol{w}^{C}_{v}\right). \tag{11}$$

Given the attention scores, we weigh the graph embeddings and find the final spatial CAL features $\boldsymbol{F}^{\mathsf{Spa}} \in \mathbb{R}^{L\times R\times L^{\mathsf{Emb}}}$ by

$$\boldsymbol{F}^{\mathsf{Spa}} = \boldsymbol{a}^{A}_{v} \odot \boldsymbol{F}^{\mathsf{A}} + \boldsymbol{a}^{B}_{v} \odot \boldsymbol{F}^{\mathsf{B}} + \boldsymbol{a}^{C}_{v} \odot \boldsymbol{F}^{\mathsf{C}}, \tag{12}$$

where $\odot$ is the Hadamard (element-wise) product operator.

We then apply the gated recurrent unit (GRU) to capture the temporal features of the spatial CAL, *i.e.*,

$$\boldsymbol{F}^{\mathsf{Tem}} = \mathsf{GRU}\left(\boldsymbol{F}^{\mathsf{Spa}}\right), \tag{13}$$

where $\boldsymbol{F}^{\mathsf{Tem}} \in \mathbb{R}^{L^{\mathsf{Tem}}}$ represents the resulting temporal features with L^{Tem} being the size of the hidden state of the GRU.

Finally, to predict the spatial CALs in time interval t', $\hat{\boldsymbol{S}}_{t'} \in \mathbb{R}^{R\times N}$, we further apply another fully-connected layer on the temporal features $\boldsymbol{F}^{\mathsf{Tem}}$ with units of $R \times N$ followed by a reshape operation, *i.e.,*

$$\hat{\boldsymbol{S}}_{t'} = \mathsf{Reshape}\left(\mathsf{Dense}\left(\boldsymbol{F}^{\mathsf{Tem}}\right)\right). \tag{14}$$

• **Categorical CAL Fusion Design.** The spatial distribution of different categories of CALs are highly correlated with the future occurrences of the citywide categorical CALs. In this step, we utilize both the extracted spatial features $\boldsymbol{F}^{\mathsf{Spa}} \in \mathbb{R}^{L \times R \times L^{\mathsf{Emb}}}$ from the previous component and the N categories of categorical CALs $\mathbf{C}^{\mathsf{Near}}$ to predict the occurrences of the categorical CALs in the time interval t. The detailed designs of the processes are illustrated in Fig. 9.

First, to leverage the spatial features of the CALs, we sum up the extracted spatial features $\boldsymbol{F}^{\mathsf{Spa}}$ over all the city regions by

$$\boldsymbol{F}^{\mathsf{Spa}'} = \sum_{i}^{R} \boldsymbol{F}_i^{\mathsf{Spa}}, \tag{15}$$

to obtain the aggregate spatial features, *i.e.,* $\boldsymbol{F}^{\mathsf{Spa}'} \in \mathbb{R}^{L \times L^{\mathsf{Emb}}}$.

We then apply the GRU with the size of the hidden states as L^{Uni} to further extract the temporal dependencies across the spatial features, *i.e.,*

$$\boldsymbol{F}^{\mathsf{Spa}''} = \mathsf{GRU}(\boldsymbol{F}^{\mathsf{Spa}'}). \tag{16}$$

To leverage the historical categorical CALs, $\mathbf{C}^{\mathsf{Near}}$, for categorical CAL prediction, we capture the temporal features of categorical CALs by a fully-connected layer followed by the GRU with the size of the hidden states as L^{Uni}, *i.e.,*

$$\mathbf{C}^{\mathsf{Near}'} = \mathsf{GRU}\left(\mathsf{Dense}\left(\mathbf{C}^{\mathsf{Near}}\right)\right). \tag{17}$$

With both the spatial CAL features $\boldsymbol{F}^{\mathsf{Spa}''} \in \mathbb{R}^{L^{\mathsf{Uni}}}$ and the categorical CAL features $\mathbf{C}^{\mathsf{Near}'} \in \mathbb{R}^{L^{\mathsf{Uni}}}$, we further concatenate them together and utilize the fully-connected layer with units of N to obtain the predictions of the N categories of categorical CALs in time interval t, *i.e.,*

$$\hat{\mathbf{C}}_t^{\mathsf{SCPIAL}} = \mathsf{Dense}\left(\mathsf{Concat}\left(\boldsymbol{F}^{\mathsf{Spa}''}, \mathbf{C}^{\mathsf{Near}'}\right)\right), \tag{18}$$

where $\hat{\mathbf{C}}_t^{\mathsf{SCPIAL}} \in \mathbb{R}^N$ is the categorical CAL prediction in time interval t.

4.2 Multi-level Temporal Feature Learning Component

• **Motivations & Component Overview.** Crowd activities usually follow certain temporal patterns, which need to be captured for accurate CAL prediction. We note that the recurrent layers of gated recurrent unit (GRU) [8] is capable of capturing the shor-term and medium-term temporal dependencies with their memorization designs. However, GRU might still fail to recognize the long-term temporal characteristics. Previous studies [31, 46] leverage the periodic patterns among real-world datasets for time series prediction. Inspired by the idea that historical data in the same time period within a day/week/month/year as the predicted time step are highly correlated, we proposed the GRU-based Multi-level Temporal Feature Learning Component to further extract the daily and weekly long-term patterns of CAL.

As shown in Fig. 10, the Multi-level Temporal Feature Learning Component includes two identical sub-components to capture the daily and weekly patterns of CALs, respectively. Each sub-component is comprised of the GRU-based recurrent skip neural network. The outputs from the two sub-components are fed into the Dense operation to have the categorical CAL prediction in time interval t.

As discussed in Sec. 3.1, there are $T \in \mathbb{R}$ time units discretized for each day, and the length of each time unit is H hours ($H = 24/T$). Given categorical CALs (from N categories) $\mathbf{C}^{\mathsf{Daily}}$ from time intervals $\{t - P \cdot T, t - (P-1) \cdot T, \ldots, t - T\}$ (past P days), and weekly-aggregated categorical CAL $\mathbf{C}^{\mathsf{Weekly}}$ from time intervals $\{t - 7 \cdot Q \cdot T, t - 7 \cdot (Q-1) \cdot T, \ldots, t - 7 \cdot T\}$ (past Q weeks), we first capture the daily trend $\hat{\mathbf{C}}_t^{\mathsf{Daily}} \in \mathbb{R}^N$ and weekly trend $\hat{\mathbf{C}}_t^{\mathsf{Weekly}} \in \mathbb{R}^N$ of the categorical CALs

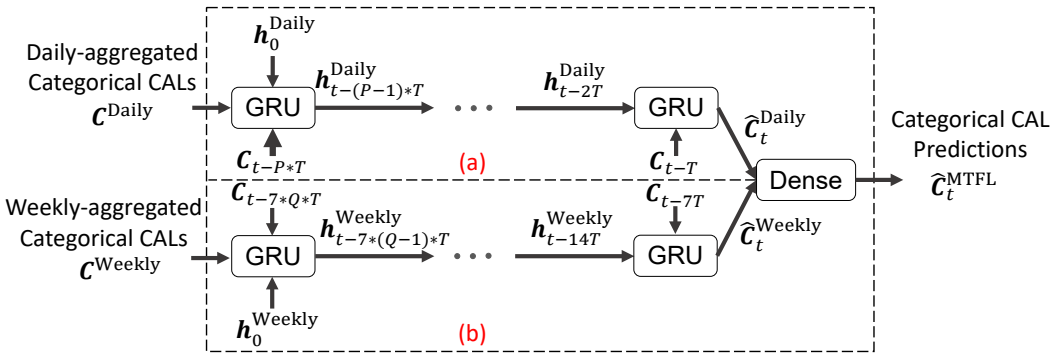


Fig. 10. Multi-level Temporal Feature Learning Component. (a) Daily pattern extraction; and (b) weekly pattern extraction

in time interval t. Then we take into account both of these two trends for the prediction of the N categories of CALs $\hat{\mathbf{C}}_t^{\mathtt{MTFL}} \in \mathbb{R}^N$ in time interval t via a fully connected layer (denoted as Dense).

Since we utilize the same structure to consider the daily and weekly patterns of CALs, we will focus on presenting the formulations of capturing the daily temporal patterns of CAL as an example.

(a) Daily Pattern Extraction. Having N categories of historical categorical CAL $\mathbf{C}^{\mathtt{Daily}} = \{\mathbf{C}_{t-P\cdot T}, \mathbf{C}_{t-(P-1)\cdot T}, \ldots, \mathbf{C}_{t-T}\} \in \mathbb{R}^{P\times N}$ in the past P days, we apply the GRU units to discover the daily temporal patterns of categorical CAL. The processes are formulated as, *i.e.,*

$$\boldsymbol{z}_p^{\mathtt{Daily}} = \sigma\left(\boldsymbol{W}^{\mathtt{dz}}\mathbf{C}_p + \boldsymbol{U}^{\mathtt{dz}}\boldsymbol{h}_{p-T}^{\mathtt{Daily}} + \boldsymbol{b}^{\mathtt{dz}}\right), \tag{19}$$

$$\boldsymbol{r}_p^{\mathtt{Daily}} = \sigma\left(\boldsymbol{W}^{\mathtt{dr}}\mathbf{C}_p + \boldsymbol{U}^{\mathtt{dr}}\boldsymbol{h}_{p-T}^{\mathtt{Daily}} + \boldsymbol{b}^{\mathtt{dr}}\right), \tag{20}$$

$$\hat{\boldsymbol{h}}_p^{\mathtt{Daily}} = \phi\left(\boldsymbol{W}^{\mathtt{dh}}\mathbf{C}_p + \boldsymbol{U}^{\mathtt{dh}}\left(\boldsymbol{r}_p^{\mathtt{Daily}} \odot \boldsymbol{h}_{p-T}^{\mathtt{Daily}}\right) + \boldsymbol{b}^{\mathtt{dh}}\right), \tag{21}$$

$$\boldsymbol{h}_p^{\mathtt{Daily}} = \boldsymbol{z}_p^{\mathtt{Daily}} \odot \boldsymbol{h}_{p-T}^{\mathtt{Daily}} + \left(1 - \boldsymbol{z}_p^{\mathtt{Daily}}\right) \odot \hat{\boldsymbol{h}}_p^{\mathtt{Daily}}, \tag{22}$$

where $\boldsymbol{z}_p^{\mathtt{Daily}} \in \mathbb{R}^N$, $\boldsymbol{r}_p^{\mathtt{Daily}} \in \mathbb{R}^N$, and $\hat{\boldsymbol{h}}_p^{\mathtt{Daily}} \in \mathbb{R}^N$ are the update gate, reset gate, and candidate hidden state of the GRU unit. $\boldsymbol{h}_p^{\mathtt{Daily}} \in \mathbb{R}^N$ is the hidden representation of the N categories of categorical CAL in time interval $p \in \{t-P\cdot T, t-(P-1)\cdot T, \ldots, t-T\}$. $\boldsymbol{W}^{\mathtt{dz}} \in \mathbb{R}^{N\times N}$, $\boldsymbol{U}^{\mathtt{dz}} \in \mathbb{R}^{N\times N}$, $\boldsymbol{b}^{\mathtt{dz}} \in \mathbb{R}^N$, $\boldsymbol{W}^{\mathtt{dr}} \in \mathbb{R}^{N\times N}$, $\boldsymbol{U}^{\mathtt{dr}} \in \mathbb{R}^{N\times N}$, $\boldsymbol{b}^{\mathtt{dr}} \in \mathbb{R}^N$, $\boldsymbol{W}^{\mathtt{dh}} \in \mathbb{R}^{N\times N}$, $\boldsymbol{U}^{\mathtt{dh}} \in \mathbb{R}^{N\times N}$, $\boldsymbol{r}_p^{\mathtt{Daily}} \in \mathbb{R}^N$, and $\boldsymbol{b}^{\mathtt{dh}} \in \mathbb{R}^N$ are parameters to learn. $\hat{\mathbf{C}}_p \in \mathbb{R}^N$ represents the N categories of categorical CAL in time interval p.

The final daily hidden representation $\boldsymbol{h}_t^{\mathtt{Daily}} \in \mathbb{R}^N$ of N categories of categorical CAL in time interval t is used as the daily trend prediction $\hat{\mathbf{C}}_t^{\mathtt{Daily}} \in \mathbb{R}^N$ of the categorical CALs.

(b) Weekly Pattern Extraction. The same operation is also utilized to capture the weekly trend of the categorical CALs. Given the Q weeks of N categories of categorical CAL $\mathbf{C}^{\mathtt{Weekly}} = \{\mathbf{C}_{t-7\cdot Q\cdot T}, \mathbf{C}_{t-7\cdot(Q-1)\cdot T}, \ldots, \mathbf{C}_{t-7\cdot T}\}$, we further use the same GRU-based recurrent skip neural network as in the *Daily Pattern Extraction* to predict the weekly trend of categorical CALs $\hat{\mathbf{C}}_t^{\mathtt{Weekly}} \in \mathbb{R}^N$ in the time interval t. Given both the $\hat{\mathbf{C}}_t^{\mathtt{Daily}}$ and $\hat{\mathbf{C}}_t^{\mathtt{Weekly}}$, a fully-connected layer is applied for weighting the three predictions and output the N categories of categorical CAL prediction $\hat{\mathbf{C}}_t^{\mathtt{MTFL}} \in \mathbb{R}^N$ at time interval t.

5 EXTREME-AWARE CAL LEARNING AND MODEL INTEGRATION

We first overview and present our motivations of the extreme-aware CAL learning component in Sec. 5.1, and then discuss the integration of the extreme value theory for the loss function designs in Sec. 5.2.

5.1 Extreme-aware CAL Learning Component

• **Motivations & Component Overview.** As shown in Sec. 3.4, the occurrences of CALs may follow their routines. However, some CALs will show extreme high/low records owing to some rare external factors. For instance, the outdoor recreation activities and nightlife activities might burst during festivals, whereas work activities drop significantly at the same time. The traditional deep learning models like LSTM/GRU can hardly learn the occurrence patterns of such an *imbalanced data distribution*, and fail to predict the extreme CALs.

To overcome the data imbalance issue in categorical CAL prediction, we propose a novel co-design called Extreme-aware CAL Learning Component, which integrates the Extreme Value Theory (EVT) [12], for adaptively modeling of the extreme categorical CALs. Specifically, our *Extreme-aware CAL Learning Component* accounts for capturing the extremely high occurrences of the CALs. As illustrated in Fig. 11, the Extreme-aware CAL Learning Component first models the probabilities and the temporal patterns of the historical occurrence of extreme CALs by applying the GRU on the K historical extreme CAL sequence and label pairs $\{(\mathbf{E}_1, \mathbf{Q}_1), (\mathbf{E}_2, \mathbf{Q}_2), \ldots, (\mathbf{E}_K, \mathbf{Q}_K)\}$. Then we utilize the attention operation to further differentiate the impacts of the occurrences of historical extreme CALs on the categorical CALs in time interval t.

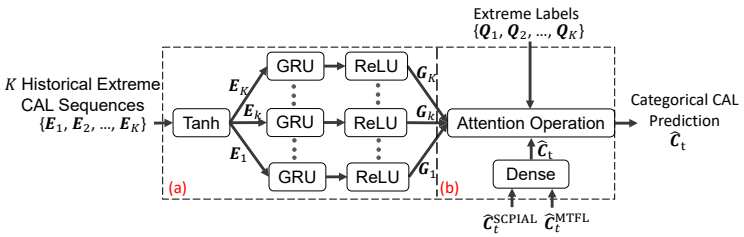


Fig. 11. Extreme-aware CAL Learning Component: *(a)* extreme CAL occurrence patterns modeling; and *(b)* prediction incorporation & attention operation.

(a) Extreme CAL Occurrence Patterns Modeling. To predict the occurrence of N categories of categorical CALs in time interval t, we first construct the K historical extreme CAL sequence and label pairs $\{(\mathbf{E}_1, \mathbf{Q}_1), (\mathbf{E}_2, \mathbf{Q}_2), \ldots, (\mathbf{E}_K, \mathbf{Q}_K)\}$. As defined in Def. 3.4, we note that $\mathbf{E}_k = (\mathbf{C}_{t-k\cdot T-T}, \ldots, \mathbf{C}_{t-k\cdot T-l}, \ldots, \mathbf{C}_{t-k\cdot T-1}) \in \mathbb{R}^{T\times N}$ denotes the k^{th} ($k \in \{1, \ldots, K\}$) historical extreme CAL sequence during time intervals $\{t-k\cdot T-T, \ldots, t-k\cdot T-l, \ldots, t-k\cdot T-1\}$, and $\mathbf{Q}_k \in \mathbb{R}^N$ denotes the $k^{\text{th}} \in \{1, \ldots, K\}$ ground truth historical extreme CAL labels of the N categories of categorical CALs at time interval $t-k\cdot T$.

To enable the occurrence pattern modeling of extreme high and normal CALs, we first classify the categorical CALs in each time interval into two classes. As discussed in Sec. 3.1, the categorical CAL in a time interval is labeled as one (extreme high) if its value is greater than θ percent of all the categorical CALs, and zero otherwise.

Given the extreme CAL sequence and label pairs, we first leverage a Tanh activation operation on the extreme CAL sequences $\{\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_K\}$, and then use K Gated Recurrent Units (GRUs) to each of the extreme CAL sequence $\mathbf{E}_k$, $k \in \{1, \ldots, K\}$. GRUs are running with a ReLU activation function. The last hidden state outputs of all the extreme CAL sequences from GRUs are denoted

as $\boldsymbol{G} \in \{\boldsymbol{G}_1, \boldsymbol{G}_2, \ldots, \boldsymbol{G}_K\}$, *i.e.*, the latent representation of the K historical extreme CAL sequences. $\boldsymbol{G}_k \in \{\boldsymbol{G}_{k,1}, \ldots, \boldsymbol{G}_{k,N}\}$ are the representations of each of the N categories of categorical CAL in extreme CAL sequence $\mathbf{E}_k$, where $\boldsymbol{G}_{k,n} \in \mathbb{R}$, $n \in \{1, 2, \ldots, N\}$.

The intuition behind utilizing the previous K historical extreme CAL sequence and label pairs to model the probabilities and the temporal patterns of extreme CALs is two-fold. *First*, the occurrences of CALs fall into different frequency ranges in different long time periods (different seasons, *etc.*). *Second*, the citywide extreme CALs of each category are usually influenced by the sudden change of some external factors. In this work, we focus on the short-term sudden frequency changes of CAL as extreme cases.

(b) Prediction Incorporation & Attention Operation. After finding the K latent representation vectors $\boldsymbol{G} \in \{\boldsymbol{G}_1, \ldots, \boldsymbol{G}_K\}$ of the K historical extreme CAL sequences $\{\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_K\}$, we further incorporate the N categories of categorical CAL prediction $\hat{\mathbf{C}}_t^{\texttt{SCPIAL}} \in \mathbb{R}^N$ from the Spatial CAL-POI Interaction-Attentive Learning Component and $\hat{\mathbf{C}}_t^{\texttt{MTFL}} \in \mathbb{R}^N$ from the Multi-level Temporal Feature Learning Component in time interval t for the final categorical CAL prediction in the time interval t.

We first produce the categorical CAL prediction $\hat{\mathbf{C}}_t \in \mathbb{R}^N$ in the time interval t through linear projection of $\hat{\mathbf{C}}_t^{\texttt{SCPIAL}}$ and $\hat{\mathbf{C}}_t^{\texttt{MTFL}}$ as follows,

$$\hat{\mathbf{C}}_t = \texttt{Dense}\left(\left[\hat{\mathbf{C}}_t^{\texttt{SCPIAL}}, \hat{\mathbf{C}}_t^{\texttt{MTFL}}\right]\right), \tag{23}$$

where $[\hat{\mathbf{C}}_t^{\texttt{SCPIAL}}, \hat{\mathbf{C}}_t^{\texttt{MTFL}}]$ represents the concatenation of $\hat{\mathbf{C}}_t^{\texttt{SCPIAL}}$ and $\hat{\mathbf{C}}_t^{\texttt{MTFL}}$.

To further leverage the temporal patterns of extreme CALs, we quantify the varying influences of the occurrences of historical extreme CALs by the Attention Operation [2], *i.e.,*

$$\boldsymbol{a}_k = \frac{\exp(c_j)}{\sum_{j=1}^{K} \exp(\mathbf{C}_j)}, \quad \text{and} \quad \mathbf{C}_j = \hat{\mathbf{C}}_t \odot \boldsymbol{G}_j, \tag{24}$$

where $\boldsymbol{a}_k \in \mathbb{R}^N$ represents the attention weight of the extreme CAL sequences $\{\mathbf{E}_k\}$. Furthermore, the overall influence $\hat{\boldsymbol{u}}_t = \{\hat{\boldsymbol{u}}_{t,1}, \hat{\boldsymbol{u}}_{t,2}, \ldots, \hat{\boldsymbol{u}}_{t,N}\} \in \mathbb{R}^N$ of extreme CAL sequences $\{\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_K\}$ on the categorical CAL prediction is evaluated by

$$\hat{\boldsymbol{u}}_t = \sum_{j=1}^{K} \exp(\boldsymbol{a}_j \odot \mathbf{Q}_j). \tag{25}$$

The final categorical CAL prediction which considers the spatio-temporal POI-CAL interactions, multi-level temporal patterns, and existence of extreme CAL is given by

$$\hat{\mathbf{C}}_t = \hat{\mathbf{C}}_t + \boldsymbol{W}^{\texttt{Ext}} \odot \hat{\boldsymbol{u}}_t, \tag{26}$$

where $\boldsymbol{W}^{\texttt{Ext}} \in \mathbb{R}^N$ is the model parameter to be trained.

5.2 Integrating Extreme Value Theory for Loss Function Designs

In this section, we further discuss how to integrate the extreme value theory for the loss function of `IE-CALP` in model training. For instance, Fig. 12 shows the transportation CALs of Tokyo (04/02/2012–02/16/2013), as well as the corresponding fitted truncated Gaussian distribution. We can learn from Fig. 12 that CALs with frequency higher than 150 follow a heavy-tailed distribution [38]. The occurrences of such activity frequencies located in the tail of the truncated Gaussian distribution make it hard to model the temporal dependencies of extreme CALs for conventional deep learning methods.

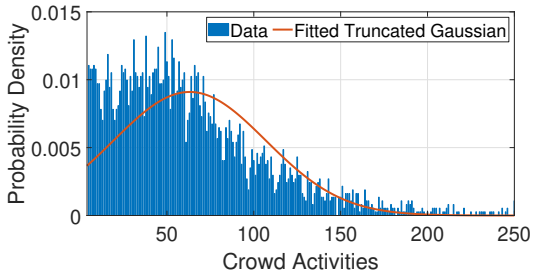


Fig. 12. Probability density function of CALs from transportation category in Tokyo and the fitted truncated Gaussian distribution (04/02/2012 to 02/16/2013).

Generalized Extreme Value Theory (GEVT) [17] takes an important step in modeling the distributions of heavy-tailed data (extreme CALs in our case). Formally, given I random variables $\{y_1, \ldots, y_I\}$, GEVT aims at modeling the distributions of the maximum data, which is formulated as

$$F(y) = \lim_{I \to \infty} P\{\max\{y_1, \ldots, y_I\} \leq y\}. \tag{27}$$

To make $F(y)$ non-degenerated to 0, we can further transform the maximum data distribution in Eq. (27) into $E(y)$, which is formulated by GEVT as follows,

$$E(y) = \begin{cases} \exp\left(-\left(1 - \frac{1}{\rho} y\right)\right)^{\rho}, & \text{if } \rho \neq 0,\ 1 - \frac{1}{\rho} y > 0; \\ \exp\left(-e^{-y}\right), & \rho = 0. \end{cases} \tag{28}$$

We further extended the GEVT to model the heavy-tailed distribution [49] by

$$1 - F(y) \approx (1 - F(\xi)) \left[1 - \log E\left(\frac{y - \xi}{f(\xi)}\right)\right], y > \xi, \tag{29}$$

where $\xi > 0$ represents a threshold parameter.

To take the extreme high CAL which follow the heavy-tailed distribution into consideration, we further proposed the Extreme Value Theory (EVT) based loss function as which based on Eq. (29) to minimize the prediction error of extreme CAL. Based on above, our model in IE-CALP takes in the following three perspectives to enable extreme-aware prediction of spatial and categorical CALs.

(a) Categorical CAL Prediction. Mean Squared Error (MSE) is the default loss function for many forecasting task. In this study, we also utilize the MSE as the loss function to formulate the prediction and ground truth values of categorical CAL, *i.e.,*

$$\text{MSE}^{\text{Cate}} = \frac{1}{N} \times \sum_{n=1}^{N} \left(\hat{\mathbf{C}}_{t,n} - \mathbf{C}_{t,n}\right)^2, \tag{30}$$

where $\hat{\mathbf{C}}_{t,n} \in \mathbb{R}$ and $\mathbf{C}_{t,n} \in \mathbb{R}$ are the prediction and ground-truth of categorical CALs of category n in time interval t.

(b) Spatial CAL Prediction. The spatial CAL prediction is also evaluated by MSE, *i.e.,*

$$\text{MSE}^{\text{Spa}} = \frac{1}{R \times N} \times \sum_{r=1}^{R} \sum_{n=1}^{N} \left(\hat{\boldsymbol{S}}_{t',r,n} - \boldsymbol{S}_{t',r,n}\right)^2, \tag{31}$$

where $\hat{\boldsymbol{S}}_{t',r,n}$ and $\boldsymbol{S}_{t',r,n}$ are the prediction and ground truth of the spatial CAL of category $n \in \{1, 2, \ldots, N\}$ in region $r \in \{1, 2, \ldots, R\}$ in time interval t'.

(c) Extreme CAL Sequence and Label Prediction. However, merely using the MSE may fail to consider the temporal distribution of extreme CAL. In order to improve the prediction accuracy of the categorical CAL, we proposed two additional loss functions which based on Eq. (29). We consider both the output latent representations $\boldsymbol{G} \in \{\boldsymbol{G}_1, \boldsymbol{G}_2, \ldots, \boldsymbol{G}_K\}$ of the K historical extreme CAL sequences $\{\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_K\}$, and the predicted extreme label $\hat{\boldsymbol{u}}_t = \{\hat{\boldsymbol{u}}_{t,1}, \hat{\boldsymbol{u}}_{t,2}, \ldots, \hat{\boldsymbol{u}}_{t,N}\} \in \mathbb{R}^N$ in the time interval t while training.

Specifically, based on Eq. (29), the temporal distribution of categorical CAL of category $n \in \{1, \ldots, N\}$ in time interval t is formulated as, *i.e.,*

$$1 - F\left(\hat{\mathbf{C}}_{t,n}\right) \approx \left(1 - P\left(\hat{u}_{t,n} = 1\right)\right) \log E\left(\frac{\hat{\mathbf{C}}_{t,n} - \xi_1}{f\left(\xi_1\right)}\right), \tag{32}$$

where $P\left(\hat{\boldsymbol{u}}_{t,n} = 1\right)$ represents the probability that $\hat{\mathbf{C}}_{t,n}$ is extreme CAL, $\boldsymbol{u}_{t,n} \in \{0, 1\}$ is the ground truth label of the categorical CAL $\hat{\mathbf{C}}_{t,n}$ of category n in time interval t, and $\xi_1 \in \mathbb{R}^+$ is the extreme threshold of the categorical CAL.

In this study, we treat the predicted extreme label $\hat{\boldsymbol{u}}_{t,n} \in [0, 1]$ of categorical CAL of category n in time interval t as the hard approximation for $\frac{\hat{\mathbf{C}}_{t,n} - \xi_1}{f(\xi_1)}$. Then we can formulate the temporal distribution of the extreme CAL by modifying the binary cross entropy. Taking the distribution of categorical CAL of category n in time interval t as example, we have

$$\begin{aligned} \mathsf{EUA}\left(\hat{\boldsymbol{u}}_{t,n}, \boldsymbol{u}_{t,n}\right) = &- \left(1 - P\left(\hat{\boldsymbol{u}}_{t,n} = 1\right)\right) \times \left[\log E\left(\boldsymbol{u}_{t,n}\right)\right] \hat{\boldsymbol{u}}_{t,n} \log\left(\boldsymbol{u}_{t,n}\right) - \left(1 - P\left(\hat{\boldsymbol{u}}_{t,n} = 0\right)\right) \times \\ &\left[\log E\left(1 - \boldsymbol{u}_{t,n}\right)\right] \left(1 - \hat{\boldsymbol{u}}_{t,n}\right) \log\left(1 - \boldsymbol{u}_{t,n}\right). \end{aligned} \tag{33}$$

The loss function of the output latent representations $\boldsymbol{G} \in \{\boldsymbol{G}_1, \boldsymbol{G}_2, \ldots, \boldsymbol{G}_K\}$ of the K historical extreme CAL sequences $\{\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_K\}$ is formulated as, *i.e.,*

$$\mathsf{EUA}^{Q} = \sum_{k=1}^{K} \sum_{n=1}^{N} \left(\mathsf{EUA}\left(\mathbf{Q}_{k,n}, \boldsymbol{G}_{k,n}\right)\right), \tag{34}$$

which aims at capturing the distribution differences of the predicted extreme CALs.

The loss function of both predicted and ground-truth extreme labels of the categorical CALs in time interval t is formulated as

$$\mathsf{EUA}^{u} = \sum_{n=1}^{N} \left(\mathsf{EUA}\left(\hat{u}_n, u_n\right)\right). \tag{35}$$

Given above Eqs. (30), (31), (34), and (35), the final loss function in this study is denoted as, *i.e.,*

$$\mathtt{Loss} = \eta_1 \times \mathsf{MSE}^{\mathtt{Cate}} + \eta_2 \times \mathsf{MSE}^{\mathtt{Spa}} + \eta_3 \times \mathsf{EUA}^{Q} + \eta_4 \times \mathsf{EUA}^{u}, \tag{36}$$

where η_1, η_2, η_3, and η_4 are the coefficients evaluating the importance of each prediction.

6 EXPERIMENTAL STUDIES

In this section, we first introduce the baseline approaches and the experimental settings in Sec. 6.1, and then we present the experimental results of this study in Sec. 6.2.

6.1 Baselines & Experimental Settings

In this study, we compare our proposed method IE-CALP with the following baselines or state-of-art algorithms. In particular, all of them are implemented into two versions to predict the spatial CAL and categorical CAL, respectively. In order to compare our proposed method IE-CALP with the baselines and state-of-art algorithms who require image-like data input, we further construct a

$H \times W$ heatmap based on the adjacency of the clustered regions of each city. For the heatmap of NYC, LA, and Tokyo, $H \times W$ are set as 6×5, 3×5, and 2×4, respectively. Specifically, we compare IE-CALP with the following baselines.

(1) *Recurrent Neural Networks* (RNN): In the prediction of spatial CAL, the aggregate spatial CAL $\boldsymbol{S}^{\text{Agg}} \in \mathbb{R}^{L \times H \times W}$ are flattened in each time interval before fed into RNN. The RNN is followed by a Dense layer to predict the spatial CAL of N categories in time interval t'. In the prediction of categorical CAL, L historical time intervals of categorical CAL of N categories are utilized to predict categories CAL in time interval t.
(2) *Gated Recurrent Unit* (GRU): We apply the same operation as RNN to predict both the categorical CAL and spatial CAL by GRU.
(3) *Long Short-Term Memory* (LSTM): To predict both the categorical CAL in time interval t and the spatial CAL in time interval t' via LSTM, we apply the same operation as RNN as mentioned above.
(4) *Convolutional LSTM Network* (ConvLSTM): In the predictions of both the categorical CAL in time interval t and the spatial CAL in time interval t', L historical time intervals of the aggregate spatial CAL $\boldsymbol{S}^{\text{Agg}} \in \mathbb{R}^{L \times H \times W}$ are fed into ConvLSTM. ConvLSTM is followed by a Dense layer.
(5) *Historical Average* (HA): In the prediction of categorical CAL, L historical time intervals of categorical CAL of N categories are fed into HA. In addition, we predict the spatial CAL in time interval t' of each of the N categories separately using the historical spatial CAL of each category.
(6) CHAT [27]: We also utilize a Dense layer as the last layer of Cross-Interaction Hierarchical Attention (CHAT) network to predict both the categorical CAL and the spatial CAL. In particular, L historical time intervals of the aggregate spatial CAL $\mathbf{C}^{\text{Agg}}$ are fed into CHAT.
(7) ST-ResNet [57]: The proposed model of ST-ResNet is adapted to take in the historical aggregate spatial CAL $\boldsymbol{S}^{\text{Agg}}$ to predict the categorical CAL and spatial CAL. In particular, the lengths of closeness, period, trend sequences in ST-ResNet are all set as L, P, and Q, respectively.
(8) ST-Norm [10]: which implements the Spatial and Temporal Normalization-based (ST-Norm) framework to predict the categorical CAl and spatial CAL separately.
(9) EVL [12]: which implements the Extreme Value Loss (EVL) to predict the categorical CAl and spatial CAL separately.

• **Parameters:** In this study, we evaluate our proposed method and the baselines with the crowd flow data, crowd activity data and Point of Interests (POIs) of NYC, Tokyo and LA. The data of NYC and Tokyo are both during 04/02/2012 – 02/01/2013. The data of LA are during 12/01/2009 – 04/30/2010. Unless otherwise stated, we use the following parameter settings by default.

For NYC, LA, and Tokyo, the last 20/30/20 days of each dataset are utilized for validation and testing, 10/15/10 days for validation and 10/15/10 days for testing, and the rest are used for training. The training epochs, batch size, and learning rate are set as: (500, 32, 0.0002), (500, 16, 0.0002), and (500, 16, 0.0002), respectively. In addition, the number of historical time intervals L, the number of days P, the number of weeks Q, the number of historical extreme CAL sequence and label pairs K, the number of CAL categories N are set as (3, 5, 4, 4, 5), (2, 3, 4, 4, 5), and (3, 6, 4, 4, 5), respectively, for NYC, LA, and Tokyo. The embedding length L^{Emb}, the GRU units L^{Uni}, and θ are set as (512, 128, 80), (64, 64, 80), and (256, 128, 80), respectively for the three cities.

• **Metrics:** The evaluation metrics in this study are included the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE), the Error Rate (ER), and the Mean Squared Logarithmic

Table 4. Prediction results and performance comparison on NYC.

Scheme	Categorical CAL Total (NYC)				Categorical CAL Extreme High (NYC)				Categorical CAL Normal (NYC)				Spatial CAL (NYC)			
	MAE	RMSE	ER	MSLE	MAE	RMSE	ER	MSLE	MAE	RMSE	ER	MSLE	MAE	RMSE	ER	MSLE
RNN	4.537	7.158	0.681	2.635	22.123	24.103	0.520	1.950	3.899	5.659	0.727	2.660	3.666	7.027	0.667	2.320
GRU	4.698	7.263	0.705	4.272	20.051	21.385	0.471	1.103	4.141	6.170	0.773	4.387	3.648	7.286	0.663	1.636
LSTM	5.216	8.828	0.783	3.936	30.084	31.717	0.707	8.441	4.314	6.654	0.805	3.772	3.696	7.700	0.672	1.812
ConvLSTM	6.013	10.432	0.903	3.706	40.436	40.673	0.950	15.391	4.765	7.264	0.889	3.283	3.425	6.642	0.623	1.456
HA	11.298	15.444	1.696	5.289	24.386	25.341	0.573	1.990	10.823	14.962	2.019	5.409	3.569	6.811	0.649	1.578
CHAT	5.098	8.561	0.765	2.235	33.141	33.480	0.778	4.491	4.081	5.941	0.761	2.153	3.434	5.995	0.624	1.543
ST-ResNet	5.969	9.723	0.896	4.505	30.692	32.368	0.721	8.094	5.072	7.744	0.946	4.375	3.607	6.265	0.656	2.253
ST-Norm	4.523	6.942	0.604	2.200	21.457	23.534	0.500	1.664	3.602	5.124	0.620	2.341	3.412	6.341	0.620	1.503
EVL	4.123	6.534	0.552	1.932	20.064	21.532	0.483	1.194	3.353	4.865	0.593	2.012	3.311	6.125	0.606	1.471
IE-CALP	**3.273**	**5.672**	**0.491**	**1.659**	**19.619**	**21.312**	**0.455**	**1.091**	**2.680**	**4.106**	**0.500**	**1.680**	**3.294**	**5.986**	**0.599**	**1.447**

Table 5. Prediction results and performance comparison on LA.

Scheme	Categorical CAL Total (LA)				Categorical CAL Extreme High (LA)				Categorical CAL Normal (LA)				Spatial CAL (LA)			
	MAE	RMSE	ER	MSLE	MAE	RMSE	ER	MSLE	MAE	RMSE	ER	MSLE	MAE	RMSE	ER	MSLE
RNN	7.342	12.714	0.862	5.502	34.615	35.095	1.000	26.434	**4.754**	7.753	**0.787**	3.516	1.299	2.131	0.700	0.940
GRU	5.472	7.525	0.642	3.201	6.882	8.421	0.199	0.133	5.338	7.434	0.883	3.492	1.433	2.509	0.772	1.272
LSTM	8.291	13.295	0.973	7.400	34.615	35.095	1.000	26.434	5.793	8.755	0.959	5.594	1.368	2.326	0.737	1.190
ConvLSTM	8.077	12.644	0.948	5.754	32.859	33.339	0.949	14.582	5.725	8.341	0.947	4.916	1.306	2.242	0.703	0.802
HA	6.741	10.222	0.791	1.492	10.369	11.417	0.300	0.179	6.397	10.101	1.058	1.617	**0.849**	**1.490**	**0.613**	**0.350**
CHAT	5.166	7.454	0.606	**1.387**	7.203	9.195	0.208	0.146	4.972	7.267	0.823	**1.504**	1.228	1.945	0.661	0.773
ST-ResNet	5.228	**6.799**	0.678	1.999	8.904	9.544	0.278	0.235	4.966	**6.560**	0.832	2.125	1.169	1.874	0.681	0.838
ST-Norm	5.210	7.572	0.619	1.591	6.942	8.623	0.221	0.156	5.152	7.321	0.835	2.521	1.274	2.094	0.699	0.798
EVL	5.012	7.242	0.596	1.521	6.364	8.075	0.184	0.113	4.989	7.142	0.814	1.663	1.253	1.983	0.687	0.703
IE-CALP	**4.989**	7.123	**0.586**	1.494	**6.193**	**7.850**	**0.179**	**0.096**	4.874	7.050	0.807	1.627	1.227	1.869	0.661	0.646

Table 6. Prediction results and performance comparison in Tokyo.

Scheme	Categorical CAL Total (Tokyo)				Categorical CAL Extreme High (Tokyo)				Categorical CAL Normal (Tokyo)				Spatial CAL (Tokyo)			
	MAE	RMSE	ER	MSLE	MAE	RMSE	ER	MSLE	MAE	RMSE	ER	MSLE	MAE	RMSE	ER	MSLE
RNN	4.390	8.753	0.401	1.425	15.824	20.609	0.276	0.313	2.398	**3.999**	0.840	1.618	0.728	1.712	0.665	0.412
GRU	4.161	8.609	0.380	1.164	**13.843**	18.800	0.241	0.230	2.475	5.045	0.866	1.326	0.744	2.002	0.679	0.402
LSTM	4.146	8.628	0.379	1.297	14.301	19.311	0.249	0.294	**2.377**	4.738	0.832	1.472	0.762	2.087	0.696	0.453
ConvLSTM	8.029	19.692	0.733	1.391	61.672	65.617	0.787	4.874	3.522	7.658	0.665	1.098	0.693	1.710	0.633	0.356
HA	9.746	17.792	0.890	4.666	32.261	37.930	0.412	0.723	7.854	14.910	1.484	4.997	0.785	1.983	1.083	0.574
CHAT	4.633	10.448	0.423	**0.743**	19.862	24.910	0.254	0.274	3.354	8.136	0.634	**0.783**	0.713	1.812	0.651	0.320
ST-ResNet	4.695	9.931	0.429	0.985	18.676	23.343	0.238	0.269	3.521	7.818	0.665	1.045	0.695	**1.645**	0.634	0.333
ST-Norm	4.012	8.477	0.369	0.931	14.965	20.043	0.263	0.210	3.132	7.421	0.583	1.044	0.701	1.709	0.625	0.365
EVL	3.732	8.342	0.348	0.902	14.325	18.932	0.203	0.162	2.899	6.954	0.521	0.981	0.681	1.692	0.603	0.332
IE-CALP	**3.564**	**8.212**	**0.325**	0.877	14.125	**18.756**	**0.180**	**0.121**	2.677	6.599	**0.506**	0.941	**0.641**	1.647	**0.585**	**0.327**

Error (MSLE), *i.e.*,

$$\text{MAE} = \frac{1}{N} \times \sum_{n=1}^{N} \left| \hat{\mathbf{C}}_{t,n} - \mathbf{C}_{t,n} \right|, \quad \text{ER} = \frac{\sum_{n=1}^{N} \left| \hat{\mathbf{C}}_{t,n} - \mathbf{C}_{t,n} \right|}{\sum_{n=1}^{N} \mathbf{C}_{t,n}}, \quad \text{RMSE} = \sqrt{\frac{1}{N} \times \sum_{n=1}^{N} \left(\hat{\mathbf{C}}_{t,n} - \mathbf{C}_{t,n} \right)^2},$$

$$\text{and} \quad \text{MSLE} = \frac{1}{N} \times \sum_{n=1}^{N} \left| \log_2 \left(\hat{\mathbf{C}}_{t,n} + 1 \right) - \log_2 \left(\mathbf{C}_{t,n} + 1 \right) \right|,$$

where $\hat{\mathbf{C}}_{t,n}$ and $\mathbf{C}_{t,n}$ denote the predicted and ground truth of categorical CAL of category n in time interval t. All the experiments are conducted upon the Google Colab[3] and a desktop of Intel i7-9700 CPU, NVIDIA GeForce RTX 2060 SUPER GPU, 16.0 GB RAM, and Windows 10. The proposed model is implemented in Python with Tensorflow-GPU-2.3.0.

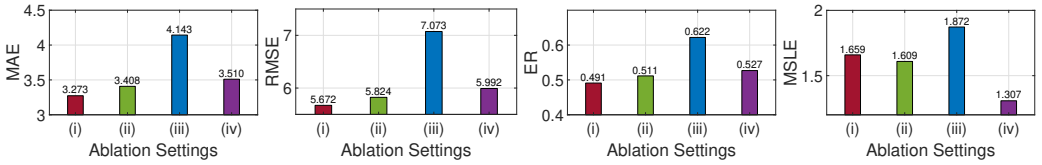
6.2 Evaluation Results

We present our experimental evaluation results as follows.

• **General Performance:** Tables 4, 5, and 6 show all the experimental results of predicting the categorical CAL and spatial CAL of NYC, LA, and Tokyo. Compared with other baselines and the state-of-the-arts, IE-CALP on average improves 24.12% in terms of all metrics. In particular, IE-CALP

[3]https://colab.research.google.com/notebooks/intro.ipynb#recent=true

improves 35.90% in the prediction of extreme CAL on average, thanks to the novel co-design of the CAL-POI graph interaction and extreme value theory.



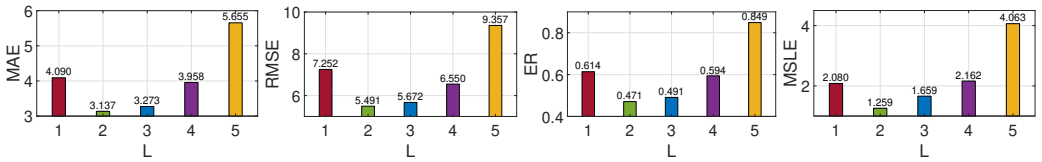
(a) MAE. (b) RMSE. (c) ER. (d) MSLE.

Fig. 13. Performance of categorical CAL prediction for NYC with (i) IE-CALP, (ii) IE-CALP without the Spatial CAL-POI Interaction-Attentive Learning Component, (iii) IE-CALP without the Multi-level Temporal Feature Learning Component, and (iv) IE-CALP without the Extreme-aware CAL Learning Component.

• **Ablation Studies:** To validate the component designs of our proposed model, we have conducted a thorough ablation study based on the dataset of NYC. Since removing some of the components in IE-CALP will disable the spatial CAL prediction, we conduct the ablation study just for categorical CAL prediction. Specifically, we compare the base design of (i) IE-CALP, with the variations of: (ii) IE-CALP without the spatial CAL prediction structure (detailed in Fig. 8) of the Spatial CAL-POI Interaction-Attentive Learning Component, (iii) IE-CALP without the Multi-level Temporal Feature Learning Component, and (iv) IE-CALP without the Extreme-aware CAL Learning Component. We can observe the most significant performance degradation in ablation setting (iii), which demonstrate the effectiveness and importance of Multi-level Temporal Feature Learning Component in capturing both the daily and weekly patterns of the categorical CAL.

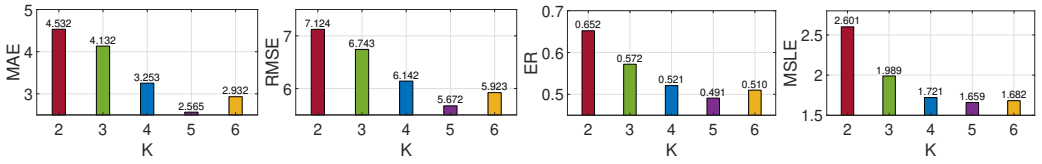
• **Sensitivity Studies:** In this section, we evaluate the influences of different parameter settings in IE-CALP on the categorical CAL and spatial CAL predictions. All the experiments in this section are running with the data of NYC. As shown in Fig. 14, the lengths of the near categorical CALs and spatial CALs in the Spatial CAL-POI Interaction-Attentive Learning Component plays an important role in the CAL prediction. We can see that with small and large L's, *i.e.,* small/large number of historical CALs for model input, IE-CALP achieves higher errors in all metrics. It is mainly because a small L may not bring enough information for modeling the CALs, while a large L may introduce noise and lead to inaccurate prediction.

To verify our selection of K in the construction of the historical extreme CAL sequence and label pairs, we implement our IE-CALP using K from 2 to 6 and show the results in Fig. 15. We can observe that the occurrences of extreme CAL are more related to the temporal patterns of the extreme CAL during the past 4 to 6 days. Small K may overlook the important CAL features. But large K may also increase the model and computational complexities and introduce noise within the data. Based on above, we consider $K = 5$ in our current studies.



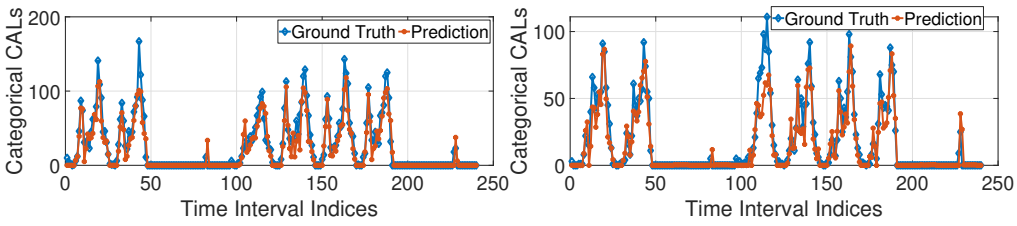
(a) MAE. (b) RMSE. (c) ER. (d) MSLE.

Fig. 14. Performance of categorical CAL prediction for NYC using L from 1 to 5.



(a) MAE. (b) RMSE. (c) ER. (d) MSLE.

Fig. 15. Performance of categorical CAL prediction for NYC using K from 2 to 6.



(a) Travel & transportation. (b) Food & shop & service.

Fig. 16. Categorical CAL predictions of categories of (a) travel & transportation; and (b) food & shop & service of Tokyo during 02/07/2013–02/16/2013.

• **Result Visualization:** Fig. 16 illustrates the categorical CAL prediction results of travel & transportation, and food & shop & service of Tokyo during 02/07/2013–02/16/2013. We can learn from the figures that IE-CALP works well in capturing the routines of categorical CALs as well as adapting to the extreme distributions. We also illustrate the attention scores (Eq. (11)) of the three CAL-POI interaction graphs for a selected region in NYC on 11/28/2012 in Fig. 17, demonstrating the spatial and temporal interactions across the three graphs. We can observe the relative importance of the graphs generated from the spatial region distance (dis), temporal CAL correlation (tmp), and region-to-region POI correlation (POI).

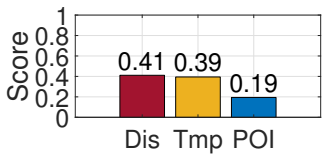
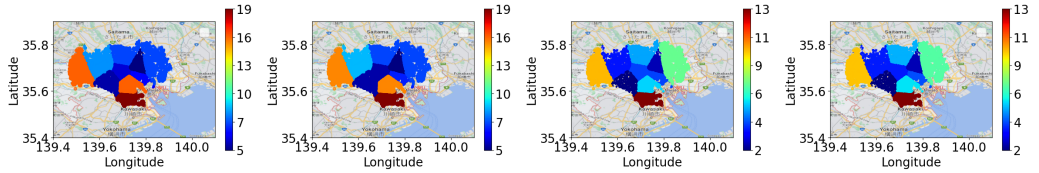


Fig. 17. Attention scores for the three interaction graphs.

Fig. 18 further illustrates the spatial CAL prediction of categories of travel & transportation, and food & shop & service of Tokyo of one selected time interval during 02/07/2013–02/16/2013. We can observe from the figure that, IE-CALP can effectively capture the spatial distributions of different categories of spatial CALs. This also validates our IE-CALP in modeling the region-to-region CAL-POI interactions.



(a) Ground truth. (b) Prediction. (c) Ground truth. (d) Prediction.

Fig. 18. Spatial CAL prediction of categories of travel & transportation ((a), (b)), and food & shop & service ((c), (d)) in one selected time interval during 02/07/2013–02/16/2013 using the data of Tokyo.

7 CONCLUSION

We have proposed IE-CALP, a spatio-temporal Interactive attention-based and Extreme-aware model for Crowd Activity Level Prediction. IE-CALP predicts both the categorical crowd activity levels (CALs) and spatial CALs by (a) capturing the spatial distribution of each categories of CAL by measuring the spatial interactions among different categories of POI and spatial CALs, and extract the spatial and temporal interactions among different categories of CALs; (b) modeling the daily and weekly temporal patterns of categorical CALs; and (c) integrating the temporal patterns of extreme categorical CAL by a novel Extreme-aware CAL Learning Component. We have also designed an adaptive loss function based on the Extreme Value Theory to capture and integrate the occurrence patterns of extreme CALs. Extensive experiments upon the crowd activity data and POIs of New York City (NYC), Los Angeles (LA), and Tokyo have further validated the accuracy and effectiveness of our proposed designs.

ACKNOWLEDGEMENT

We would like to thank the anonymous reviewers for their constructive comments. This project was supported, in part, by the National Science Foundation (NSF) under Grant 2303575. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] Maria Andersson, Fredrik Gustafsson, Louis St-Laurent, and Donald Prevost. 2013. Recognition of anomalous motion patterns in urban surveillance. *IEEE Journal of Selected Topics in Signal Processing* 7, 1 (2013), 102–110.
[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
[3] Jie Bao, Yu Zheng, and Mohamed F Mokbel. 2012. Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proc. of ACM SIGSPATIAL*. 199–208.
[4] Luca Bedogni, Shakila Khan Rumi, and Flora D. Salim. 2021. Modelling Memory for Individual Re-Identification in Decentralised Mobile Contact Tracing Applications. *Proc. ACM IMWUT* 5, 1, Article 4 (Mar 2021), 21 pages.
[5] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*. Springer, 1–4.
[6] Cláudio GS Capanema, Fabrício A Silva, Thais RMB Silva, and Antonio AF Loureiro. 2021. POI-RGNN: Using Recurrent and Graph Neural Networks to Predict the Category of the Next Point of Interest. In *Proc. of ACM PE-WASUN*. 49–56.
[7] Di Chai, Leye Wang, and Qiang Yang. 2018. Bike flow prediction with multi-graph convolutional networks. In *Proc. ACM SIGSPATIAL*. 397–400.
[8] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).
[9] Laurens De Haan, Ana Ferreira, and Ana Ferreira. 2006. *Extreme value theory: an introduction*. Vol. 21. Springer.
[10] Jinliang Deng, Xiusi Chen, Renhe Jiang, Xuan Song, and Ivor W Tsang. 2021. ST-Norm: Spatial and Temporal Normalization for Multi-variate Time Series Forecasting. In *Proc. of ACM SIGKDD*. 269–278.
[11] Zulong Diao, Xin Wang, Dafang Zhang, Yingru Liu, Kun Xie, and Shaoyao He. 2019. Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. In *Proc. AAAI*, Vol. 33. 890–897.
[12] Daizong Ding, Mi Zhang, Xudong Pan, Min Yang, and Xiangnan He. 2019. Modeling extreme events in time series prediction. In *Proc. ACM SIGKDD*. 1114–1122.
[13] Krittika D'Silva, Kasthuri Jayarajah, Anastasios Noulas, Cecilia Mascolo, and Archan Misra. 2018. The role of urban mobility in retail business survival. *Proc. ACM IMWUT* 2, 3 (2018), 1–22.
[14] Yali Fan, Zhen Tu, Yong Li, Xiang Chen, Hui Gao, Lin Zhang, Li Su, and Depeng Jin. 2019. Personalized Context-aware Collaborative Online Activity Prediction. *Proc. ACM IMWUT* 3, 4 (2019), 1–28.
[15] Zipei Fan, Xuan Song, Tianqi Xia, Renhe Jiang, Ryosuke Shibasaki, and Ritsu Sakuramachi. 2018. Online deep ensemble learning for predicting citywide human mobility. *Proc. of ACM IMWUT* 2, 3 (2018), 1–21.
[16] Zheng Fang, Qingqing Long, Guojie Song, and Kunqing Xie. 2021. Spatial-temporal graph ODE networks for traffic flow forecasting. In *Proc. ACM SIGKDD*. 364–373.
[17] Ronald Aylmer Fisher and Leonard Henry Caleb Tippett. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 24.

Cambridge University Press, 180–190.
[18] Xu Geng, Yaguang Li, Leye Wang, Lingyu Zhang, Qiang Yang, Jieping Ye, and Yan Liu. 2019. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In *Proc. AAAI*, Vol. 33. 3656–3663.
[19] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proc. of AAAI*, Vol. 33. 922–929.
[20] Mengyue Hang, Ian Pytlarz, and Jennifer Neville. 2018. Exploring student check-in behavior for improved point-of-interest prediction. In *Proc. of ACM SIGKDD*. 321–330.
[21] John A Hartigan and Manchek A Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series A (Applied Statistics)* 28, 1 (1979), 100–108.
[22] Suining He and Kang G Shin. 2019. Crowd-flow graph construction and identification with spatio-temporal signal feature fusion. In *IEEE INFOCOM*. IEEE, 757–765.
[23] Suining He and Kang G. Shin. 2019. Spatio-Temporal Adaptive Pricing for Balancing Mobility-on-Demand Networks. *ACM TIST* 10, 4, Article 39 (jul 2019), 28 pages.
[24] Suining He and Kang G Shin. 2020. Towards fine-grained flow forecasting: A graph attention approach for bike sharing systems. In *Proc. of The Web Conference 2020*. 88–98.
[25] Suining He, Bing Wang, Kang G. Shin, and Mahan Tabatabaie. 2022. Cross-zone and extreme-aware mobility learning of crowd interactions with built environments. In *Proc. ACM BuildSys*. 99–108.
[26] Chao Huang, Xian Wu, and Dong Wang. 2016. Crowdsourcing-based urban anomaly prediction system for smart cities. In *Proc. ACM CIKM*. 1969–1972.
[27] Chao Huang, Chuxu Zhang, Peng Dai, and Liefeng Bo. 2020. Cross-Interaction Hierarchical Attention Networks for Urban Anomaly Prediction. In *Proc. IJCAI*. 4359–4365.
[28] Huiqun Huang, Xi Yang, and Suining He. 2021. Multi-Head Spatio-Temporal Attention Mechanism for Urban Anomaly Event Prediction. *Proc. ACM IMWUT* 5, 3 (2021), 1–21.
[29] Renhe Jiang, Zekun Cai, Zhaonan Wang, Chuang Yang, Zipei Fan, Quanjun Chen, Xuan Song, and Ryosuke Shibasaki. 2022. Predicting Citywide Crowd Dynamics at Big Events: A Deep Learning System. *ACM Transactions on Intelligent Systems and Technology (TIST)* 13, 2, Article 21 (mar 2022), 24 pages.
[30] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
[31] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. 2018. Modeling long-and short-term temporal patterns with deep neural networks. In *Proc. ACM SIGIR*. 95–104.
[32] Ziqian Lin, Jie Feng, Ziyang Lu, Yong Li, and Depeng Jin. 2019. DeepSTN+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis. In *Proc. AAAI*, Vol. 33. 1020–1027.
[33] Aurelie C Lozano, Hongfei Li, Alexandru Niculescu-Mizil, Yan Liu, Claudia Perlich, Jonathan Hosking, and Naoki Abe. 2009. Spatial-temporal causal modeling for climate change attribution. In *Proc. ACM SIGKDD*. 587–596.
[34] Man Luo, Bowen Du, Konstantin Klemmer, Hongming Zhu, Hakan Ferhatosmanoglu, and Hongkai Wen. 2020. D3P: Data-driven demand prediction for fast expanding electric vehicle sharing systems. *Proc. ACM IMWUT* 4, 1 (2020), 1–21.
[35] Wenjun Lyu, Guang Wang, Yu Yang, and Desheng Zhang. 2022. Mover: Generalizability Verification of Human Mobility Models via Heterogeneous Use Cases. *Proc. ACM IMWUT* 5, 4, Article 171 (dec 2022), 21 pages.
[36] Emaad Manzoor, Hemank Lamba, and Leman Akoglu. 2018. xstream: Outlier detection in feature-evolving data streams. In *Proc. ACM SIGKDD*. 1963–1972.
[37] Gyoung S Na, Donghyun Kim, and Hwanjo Yu. 2018. Dilof: Effective and memory efficient local outlier detection in data streams. In *Proc. ACM SIGKDD*. 1993–2002.
[38] Tomasz Rolski, Hanspeter Schmidli, Volker Schmidt, and Jozef L Teugels. 2009. *Stochastic Processes for Insurance and Finance*. Vol. 505. John Wiley & Sons.
[39] Alban Siffer, Pierre-Alain Fouque, Alexandre Termier, and Christine Largouet. 2017. Anomaly detection in streams with extreme value theory. In *Proc. ACM SIGKDD*. 1067–1075.
[40] Yiwei Song, Dongzhe Jiang, Yunhuai Liu, Zhou Qin, Chang Tan, and Desheng Zhang. 2021. HERMAS: A Human Mobility Embedding Framework with Large-scale Cellular Signaling Data. *Proc. ACM IMWUT* 5, 3 (2021), 1–21.
[41] Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. 2019. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proc. ACM SIGKDD*. 2828–2837.
[42] Ke Sun, Tieyun Qian, Tong Chen, Yile Liang, Quoc Viet Hung Nguyen, and Hongzhi Yin. 2020. Where to go next: Modeling long-and short-term user preferences for point-of-interest recommendation. In *Proc. AAAI*, Vol. 34. 214–221.
[43] Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. 2012. LSTM neural networks for language modeling. In *Proc. ISCA*.
[44] Senzhang Wang, Meiyue Zhang, Hao Miao, Zhaohui Peng, and Philip S Yu. 2022. Multivariate Correlation-aware Spatio-temporal Graph Convolutional Networks for Multi-scale Traffic Prediction. *ACM TIST* 13, 3 (2022), 1–22.

[45] Xiaoyang Wang, Yao Ma, Yiqi Wang, Wei Jin, Xin Wang, Jiliang Tang, Caiyan Jia, and Jian Yu. 2020. Traffic flow prediction via spatial temporal graph neural network. In *Proc. WWW*. 1082–1092.
[46] Yuandong Wang, Hongzhi Yin, Hongxu Chen, Tianyu Wo, Jie Xu, and Kai Zheng. 2019. Origin-destination matrix prediction via graph convolution: a new perspective of passenger demand modeling. In *Proc. ACM SIGKDD*. 1227–1235.
[47] Zhaobo Wang, Yanmin Zhu, Qiaomei Zhang, Haobing Liu, Chunyang Wang, and Tong Liu. 2022. Graph-enhanced spatial-temporal network for next poi recommendation. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 16, 6 (2022), 1–21.
[48] Tyler Wilson, Andrew McDonald, Asadullah Hill Galib, Pang-Ning Tan, and Lifeng Luo. 2022. Beyond Point Prediction: Capturing Zero-Inflated & Heavy-Tailed Spatiotemporal Data with Deep Extreme Mixture Models. In *Proc. ACM SIGKDD*. 2020–2028.
[49] Rym Worms. 1998. Propriété asymptotique des excès additifs et valeurs extrêmes: le cas de la loi de Gumbel. *Comptes Rendus de l'Academie des Sciences Series I Mathematics* 5, 327 (1998), 509–514.
[50] Xian Wu, Yuxiao Dong, Chao Huang, Jian Xu, Dong Wang, and Nitesh V Chawla. 2017. UAPD: Predicting urban anomalies from spatial-temporal data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 622–638.
[51] Fengli Xu, Yong Li, and Shusheng Xu. 2020. Attentional Multi-graph Convolutional Network for Regional Economy Prediction with Open Migration Data. In *Proc. ACM SIGKDD*. 2225–2233.
[52] Xi Yang, Suining He, and Mahan Tabatabaie. 2023. Equity-Aware Cross-Graph Interactive Reinforcement Learning for Bike Station Network Expansion. In *Proc. ACM SIGSPATIAL*. Article 45, 12 pages.
[53] Xi Yang, Suining He, Mahan Tabatabaie, and Bing Wang. 2023. Towards Dynamic Crowd Mobility Learning and Meta Model Updates for A Smart Connected Campus. In *Proc. EWSN*. 126–137.
[54] Xi Yang, Suining He, Bing Wang, and Mahan Tabatabaie. 2022. Spatio-Temporal Graph Attention Embedding for Joint Crowd Flow and Transition Predictions: A Wi-Fi-based Mobility Case Study. *Proc. ACM IMWUT* 5, 4, Article 187 (dec 2022), 24 pages.
[55] Desheng Zhang, Tian He, and Fan Zhang. 2017. Real-Time Human Mobility Modeling with Multi-View Learning. *ACM Transactions on Intelligent Systems and Technology (TIST)* 9, 3, Article 22 (dec 2017), 25 pages.
[56] Huichu Zhang, Yu Zheng, and Yong Yu. 2018. Detecting urban anomalies using multiple spatio-temporal data sources. *Proc. of ACM IMWUT* 2, 1 (2018), 1–18.
[57] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proc. of AAAI*. AAAI Press, 1655–1661.
[58] Jason Shuo Zhang, Mike Gartrell, Richard Han, Qin Lv, and Shivakant Mishra. 2019. GEVR: an event venue recommendation system for groups of mobile users. *Proc. of ACM IMWUT* 3, 1 (2019), 1–25.
[59] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. 2020. GMAN: A graph multi-attention network for traffic prediction. In *Proc. AAAI*, Vol. 34. 1234–1241.
[60] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban Computing: Concepts, Methodologies, and Applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 3, Article 38 (sep 2014), 55 pages.
[61] Zhengyang Zhou, Yang Wang, Xike Xie, Lianliang Chen, and Hengchang Liu. 2020. RiskOracle: A Minute-Level Citywide Traffic Accident Forecasting Framework. In *Proc. AAAI*, Vol. 34. 1258–1265.