



Learning Social Meta-knowledge for Nowcasting Human Mobility in Disaster

Renhe Jiang*
The University of Tokyo
jiangrh@csis.u-tokyo.ac.jp

Zhaonan Wang*
The University of Tokyo
znwang@csis.u-tokyo.ac.jp

Yudong Tao*
University of Miami
yxt128@miami.edu

Chuang Yang
The University of Tokyo
chuang.yang@csis.u-tokyo.ac.jp

Xuan Song[†]
The University of Tokyo
songxuan@csis.u-tokyo.ac.jp

Ryosuke Shibasaki
The University of Tokyo
shiba@csis.u-tokyo.ac.jp

Shu-Ching Chen[†]
University of
Missouri-Kansas City
s.chen@umkc.edu

Mei-Ling Shyu
University of
Missouri-Kansas City
shyum@umkc.edu

ABSTRACT

Human mobility nowcasting is a fundamental research problem for intelligent transportation planning, disaster responses and management, etc. In particular, human mobility under big disasters such as hurricanes and pandemics deviates from its daily routine to a large extent, which makes the task more challenging. Existing works mainly focus on traffic or crowd flow prediction in normal situations. To tackle this problem, in this study, disaster-related Twitter data is incorporated as a covariate to understand the public awareness and attention about the disaster events and thus perceive their impacts on the human mobility. Accordingly, we propose a Meta-knowledge-Memorizable Spatio-Temporal Network (MemeSTN), which leverages memory network and meta-learning to fuse social media and human mobility data. Extensive experiments over three real-world disasters including Japan 2019 typhoon season, Japan 2020 COVID-19 pandemic, and US 2019 hurricane season were conducted to illustrate the effectiveness of our proposed solution. Compared to the state-of-the-art spatio-temporal deep models and multivariate-time-series deep models, our model can achieve superior performance for nowcasting human mobility in disaster situations at both country level and state level.

CCS CONCEPTS

• Information systems → Spatial-temporal systems; • Applied computing → Sociology.

KEYWORDS

human mobility, twitter, disaster, hurricane, typhoon, COVID-19, spatiotemporal modeling, multivariate time series, meta-learning.

* Equal contribution.

[†] Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WWW '23, April 30–May 04, 2023, Austin, TX, USA

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9416-1/23/04...\$15.00
<https://doi.org/10.1145/3543507.3583991>

ACM Reference Format:

Renhe Jiang*, Zhaonan Wang*, Yudong Tao*, Chuang Yang, Xuan Song[†], Ryosuke Shibasaki, Shu-Ching Chen[†], and Mei-Ling Shyu. 2023. Learning Social Meta-knowledge for Nowcasting Human Mobility in Disaster. In *Proceedings of the ACM Web Conference 2023 (WWW '23)*, April 30–May 04, 2023, Austin, TX, USA. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3543507.3583991>

1 INTRODUCTION

In recent years, the world was hit by an increasing number of severe disasters, and tremendous amounts of human lives and properties are under the risks. For example, Coronavirus disease 2019 (COVID-19) has resulted in more than 6.5 million deaths as of October 2022, while the 2019 Pacific Typhoon season and 2019 Atlantic Hurricane season were both the costliest seasons on records. Against this background, nowcasting human mobility in disaster is critical for disaster management and response (*i.e.*, saving lives and reducing economic losses), which also aligns with the Sustainable Development Goals (SDGs) for promoting “Sustainable Cities and Communities” and “Good Health and Well-being”.

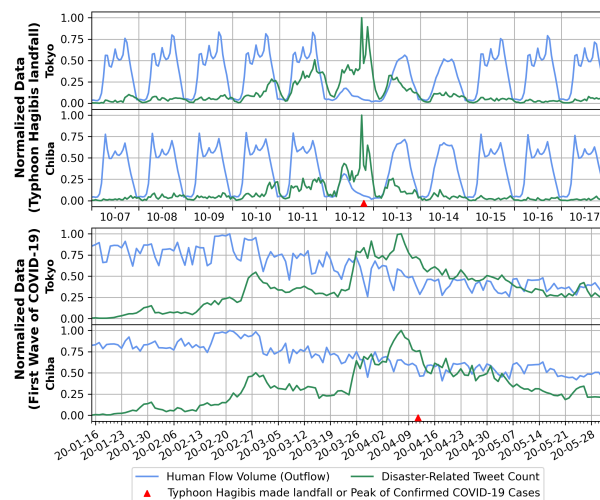


Figure 1: The normalized human outflow volume and disaster-related tweet count for two selected prefectures in Japan during the Typhoon Hagibis (upper, hourly aggregated) and the first wave of COVID-19 (lower, daily aggregated).

Each disaster event is distinct and almost unpredictable. Human mobility under disasters will deviate from its routine, which makes the human mobility nowcasting under disasters more challenging. By leveraging the the state-of-the-art deep learning technologies, enormous efforts [3, 6, 9, 20, 27, 31, 36, 39, 41, 43–45] have been made to forecast human (crowd/taxi/bike) flow or density. However, these models mainly focus on learning the periodic patterns of human mobility in normal situations, which makes it difficult to apply in anomalous disaster situations. Meanwhile, social media can provide valuable information to sense the collective sentiment under disasters. Previous study showed that geo-tagged social media data can be leveraged to precisely estimate the locations of earthquakes [26]. Residents' evacuation decisions have been modeled and predicted with their pre-disaster web search behaviors as well [38]. Inspired by these, we collected Twitter and mobility data from all 47 prefectures in Japan in 2019 Pacific Typhoon Season and during COVID-19 pandemic, where the human outflow volume and disaster-related tweet count in two selected prefectures have been depicted in Figure 1. It can be found that the "hills" of disaster-related tweet count co-occurred with the abnormal patterns in human mobility volume during the landfall of Typhoon Hagibis on October 12 and the prevalence of COVID-19 in March and April of 2020. In this sense, tweet count can be seen as a covariate to help the human mobility nowcasting in disasters.

Thus, in this study, we propose to incorporate the disaster-related social media data for human mobility nowcasting under disasters. We develop a novel model called Meta-knowledge-Memorizable Spatio-Temporal Network (MemeSTN) to learn the cross-modal correlations and inter-dependencies between the social media and mobility data. Specifically, we first extract the spatial meta-knowledge (SMK) and temporal meta-knowledge (TMK) from the social media data through one STN, then utilize the SMK and TMK to parameterize another STN for modeling the mobility data. Essentially, this "learning-to-learn" meta-learning procedure is to perform data fusion by "learning-social-to-learn-mobility". Furthermore, we utilize a learnable memory bank to store the socio-temporal prototypes (e.g., normal days, disaster coming) learned from the history, and let the current refer to the history via attention mechanism, through which the abnormality caused by the disaster can be quickly perceived. Our main contributions are summarized as follows:

- We propose to nowcast human mobility at country/state level under big disasters, which is a novel and significant topic for the field of web and society.
- The tweet count is utilized as an effective covariate for human mobility nowcasting during disasters, which guides the modeling of the heterogeneity in the responses to the disasters across different regions and the differences in human mobility under normal and abnormal scenarios.
- Meta-knowledge-Memorizable Spatio-Temporal Network (MemeSTN) is proposed for human mobility nowcasting. It leverages the memory network and meta-knowledge distillation to better learn the cross-modal dependencies between social media and mobility data.
- Three large-scale multimodal disaster datasets including Japan 2019 Typhoon, Japan 2020 COVID-19, and US 2019 Hurricane are developed with multi-month observations of

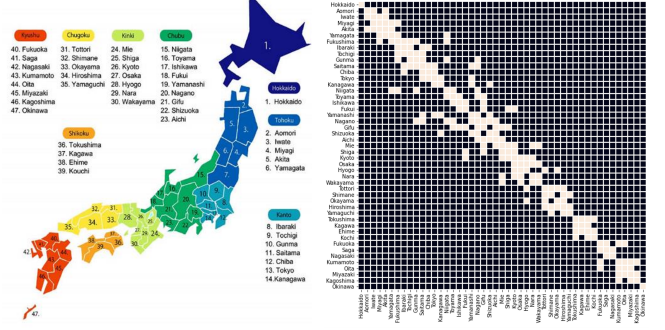


Figure 2: The 47 prefectures of Japan[33] and the binary adjacency matrix. White cell means two prefectures are adjacent.

tweet count and human mobility volume. Intensive experiments have shown the superior performance of our proposed model over the state-of-the-arts¹.

2 PROBLEM FORMULATION

Definition 1. (Region-Wise Human Mobility): Let $\{r_1, r_2, \dots, r_N\}$ be a spatial area consisting of N non-overlapping regions and $\{\tau_1, \tau_2, \dots, \tau_T\}$ be T consecutive and equally-divided timeslots (time intervals). We define different types of human mobility based on the original data source: (1) *GPS trajectory data*. Each trajectory in $\mathcal{J}$ is processed to a series of origin-destination (OD) pairs (essentially stay points) $[uid, (o_1, d_1), \dots, (o_i, d_i)]$ via trip segmentation, where $o.l$ and $d.l$ are the locations of the origin o and destination d , $o.t$ is the departure time leaving o , and $d.t$ is the arrival time reaching d . Then, the human mobility that takes region r as the origin or destination within timeslot τ is defined as follows:

$$\begin{aligned} x_\tau^{r(O)} &= |\{(o, d) \in \mathcal{J} \mid o.l \in r \wedge o.t \in \tau\}| \\ x_\tau^{r(D)} &= |\{(o, d) \in \mathcal{J} \mid d.l \in r \wedge d.t \in \tau\}| \end{aligned} \quad (1)$$

where $|\cdot|$ denotes the cardinality of a set. To be intuitive, we call $x_\tau^{r(D)}$ as human inflow (i.e., people who reach r in τ) and $x_\tau^{r(O)}$ as human outflow (i.e., people who leave r in τ). By aggregating all of the regions and timeslots, we can obtain an inflow or outflow tensor $X \in \mathbb{R}^{T \times N}$. (2) *POI visitation data*. The visitation record for each Point-Of-Interest (POI) denoted as p is structured as (id, l, t, v) , where id, l, t, v respectively represent POI ID, location, datetime, and people visit number. Then the human mobility based on POI visitation number is defined as:

$$x_\tau^{r(V)} = \sum p.v, \forall p, p.l \in r \wedge p.t \in \tau \quad (2)$$

Through this, the POI visit tensor $X \in \mathbb{R}^{T \times N}$ can be obtained.

Definition 2. (Spatial and Temporal Knowledge): Given the defined region set $\{r_1, r_2, \dots, r_N\}$, there are plenty of attributes to describe regional profiles (e.g. geospatial, demographic, socio-economic), which are in our context encapsulated in spatial knowledge, denoted by a matrix $K^{(S)} \in \mathbb{R}^{N \times f_S}$. And a topological graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ can be easily built, with region set $|\mathcal{V}| = N$ and edge set indicating spatial relations in regions. Figure 2 takes Japan as

¹Data and codes are available at <https://github.com/deepkashiwa20/MemeSTN.git>

an example by demonstrating adjacent relationship $\mathcal{A} \in \mathbb{R}^{N \times N}$ of its 47 prefectures. Similarly for the timeslot set $\{\tau_1, \tau_2, \dots, \tau_T\}$, we use the term temporal knowledge to represent the inherent properties of each timeslot (e.g. time-of-day, day-of-week, whether-holiday), denoted by matrix $K^{(T)} \in \mathbb{R}^{T \times f_T}$. f_S/f_T denote feature dimensions.

Definition 3. (Social Covariate): In this study, our motivation is leveraging social media to assist nowcasting human mobility in disaster/emergency situations. We have collected disaster-related geo-tagged tweets (tweetid, timestamp, location, text) and aggregate them into a tensor $S \in \mathbb{R}^{T \times N}$ for the given region and timeslot sets. This tensor is termed as social covariate as it shares same spatial and temporal granularity with the region-wise human mobility. Here we choose Twitter as the social data source because of its good accessibility and large volume.

Definition 4. (Nowcasting Human Mobility with Social Covariate): Given human mobility and social covariate from the past α timeslots, we aim to build a model $\mathcal{F}$ with learnable parameters θ to generate the inflow/outflow for N regions in the future β timeslots, denoted as follows.

$$[X_{T-\alpha+1}, \dots, X_T]; [S_{T-\alpha+1}, \dots, S_T] \xrightarrow[\theta]{\mathcal{F}} [\hat{X}_{T+1}, \dots, \hat{X}_{T+\beta}] \quad (3)$$

3 METHODOLOGY

In this section, we propose Meta-knowledge Memorizable Spatio-Temporal Network (**MemeSTN**), for social covariate-guided human mobility nowcasting.

3.1 Spatio-Temporal Network (STN)

Without loss of generality, we use Spatio-Temporal Network (STN) to denote a class of deep neural networks for spatio-temporal modeling [18]. As introduced in *Definition 2*, the underlying relationship in regions of irregular polygons can be modeled as an undirected graph, so we apply spectral graph convolution to handle the spatial dependency among regions as follows:

$$H = \sigma(X \star_A \Theta) \approx \sigma\left(\sum_{k=0}^K \tilde{A}^k X W_k\right) \quad (4)$$

where $\star_A$ denotes a graph convolution operation on graph $A \in \mathbb{R}^{N \times N}$ that can be approximated by Chebyshev polynomials to the order of K based on $\tilde{A}$ (normalized A) [5], W_k denotes the weights of graph convolution, and σ denotes a non-linear activation function. Here A is usually a static graph (e.g., binary adjacency matrix) pre-defined based on prior knowledge, such as regional adjacency, distance, functionality [20, 22, 39]. To avoid arbitrary choice and discover latent spatial dependency, we generate a self-learnable graph $\tilde{P}$ to replace $\tilde{A}$ in Eq. 4 utilizing a parameterized node embedding $E \in \mathbb{R}^{N \times e}$ [3, 36] with random initialization. This graph structure learning for $\tilde{P}$ is denoted by:

$$\tilde{P} = \text{softmax}(\text{relu}(EE^T)) \quad (5)$$

where the activation function relu regulates the product of node embedding E and its transpose to be non-negative and softmax further normalizes in a random walk fashion. Moreover, we exploit a recurrent structure (e.g., GRU) to model the short-term temporal dependency in nowcasting. By replacing matrix multiplications

with graph convolution operations (defined in Eq. 4), we build our spatio-temporal unit (STU), which is essentially graph convolutional recurrent unit [20] for simultaneous spatial and temporal modeling, denoted by:

$$\begin{cases} \mathbf{u}_\tau = \text{sigmoid}([X_\tau, H_{\tau-1}] \star_P \Theta_u + b_u) \\ \mathbf{r}_\tau = \text{sigmoid}([X_\tau, H_{\tau-1}] \star_P \Theta_r + b_r) \\ \mathbf{C}_\tau = \tanh([X_\tau, (\mathbf{r}_\tau \odot H_{\tau-1})] \star_P \Theta_C + b_C) \\ H_\tau = \mathbf{u}_\tau \odot H_{\tau-1} + (1 - \mathbf{u}_\tau) \odot \mathbf{C}_\tau \end{cases} \quad (6)$$

in which $\mathbf{u}$, $\mathbf{r}$ and $\mathbf{C}$ denote the update gate, reset gate and candidate cell of STU, respectively. Commonly, each parameter θ is a matrix with shape $\mathbb{R}^{(1+v) \times u}$. Projecting $[X_\tau, H_{\tau-1}] \in \mathbb{R}^{N \times (1+v)}$ with θ is practically equivalent to applying same parameter uniformly on N regions, which ignores the heterogeneity over space. Thereby, we adopt node-specific parameters $\Theta \in \mathbb{R}^{N \times (1+v) \times u}$ [3, 23] to handle diverse scales and patterns of human mobilities in different regions.

Given an observational sequence (i.e., α steps of human mobility), we use one STU layer as an encoder to extract hidden spatio-temporal representation $H_\tau \in \mathbb{R}^{N \times u}$ as follows:

$$H_\tau = \text{STU}^{(X)}(X_{T-\alpha+1}, \dots, X_{T-1}, X_T) \quad (7)$$

Then, instead of stepwise decoding H_τ with another STU layer, which leads to low inference efficiency and error accumulation, we implement a temporal deconvolutional (TDC) decoder to project H_τ for the multistep human inflow/outflow $\hat{X} \in \mathbb{R}^{N \times \beta}$ at one shot [3]. Let $*$ denote a 1D deconvolution operation with filter $g \in \mathbb{R}^{1 \times \beta}$ over horizon $\epsilon \in (1, \dots, \beta)$, we define the TDC operation as:

$$\hat{X}_{T+1}, \dots, \hat{X}_{T+\beta} = \text{TDC}(H_\tau) = H_\tau * g(\tau + \epsilon) \quad (8)$$

Combining STU encoder and TDC decoder gives the **vanilla STN**, which performs human mobility nowcasting (defined in Eq. 3) in an autoregressive manner without leveraging auxiliary social covariate S_τ . While concatenating S_τ to X_τ as dual-channel input for spatio-temporal encoder can be a straightforward solution, we find it is actually suboptimal since an shared encoder may not gain directly benefit from the raw input of social covariate. Instead, it can act as noise and hurt the performance (demonstrated later in Figure 4).

3.2 Meta-knowledge Distiller

To utilize the knowledge in social covariate to guide human mobility nowcasting in a more effective way, we propose a meta-knowledge distiller to learn two types of meta-knowledge, namely spatial meta-knowledge (SMK) and temporal meta-knowledge (TMK). Note that knowledge distillation commonly refers to the process of knowledge transferring from a large model to a small one. In our case, the teacher and student models do not differ in size, but in input source. Here we apply another STU layer (defined in Eq. 6) for social covariate input. The rationale is to let teacher model $\text{STU}^{(S)}$ to encode precursor to guide student model $\text{STU}^{(X)}$. Formally,

$$D_\tau = \text{STU}^{(S)}(S_{T-\alpha+1}, \dots, S_{T-1}, S_T) \quad (9)$$

3.2.1 Distilling Spatial Meta-knowledge. Given a set of regions $\{r_1, r_2, \dots, r_N\}$ and a group of attributes describing regional profiles (e.g., population), a matrix for spatial knowledge $K^{(S)} \in \mathbb{R}^{N \times f_S}$ can be built. In case where no prior knowledge is available, $K^{(S)}$ can

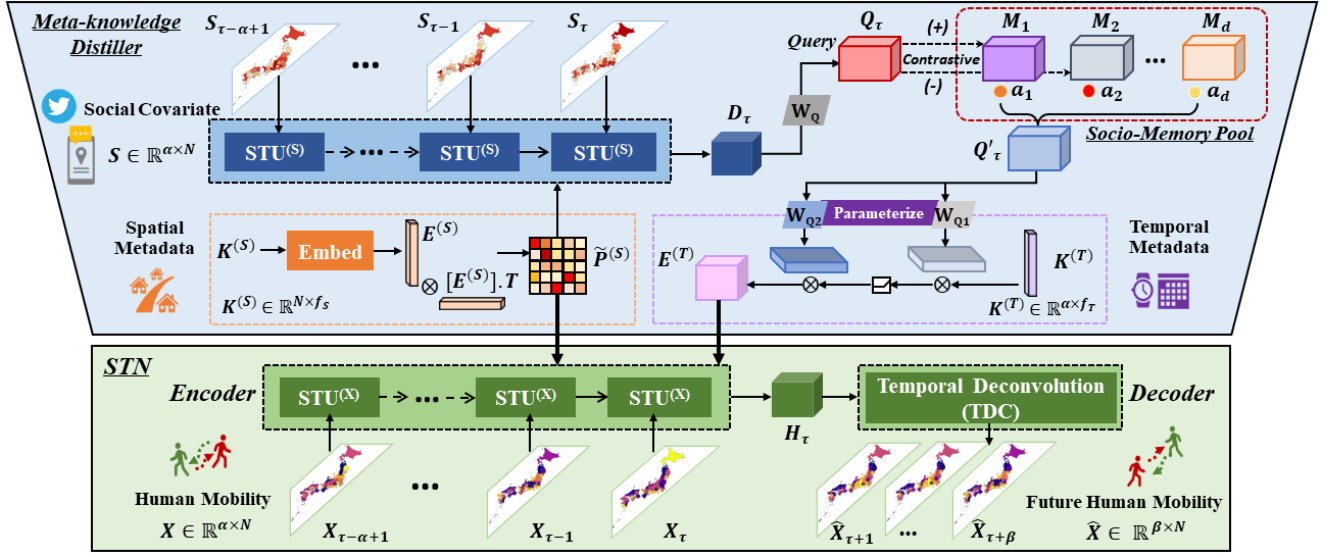


Figure 3: Meta-knowledge Memorizable Spatio-Temporal Network (MemeSTN): (1) Encoding Social Covariate and Distill Spatial Meta-knowledge; (2) Querying Socio-Memory Pool to Distill Temporal Meta-knowledge; (3) Encoding Human Mobility with Distilled Meta-knowledge and Decoding for Nowcasting.

be treated as a parameter matrix. Then, we embed $K^{(S)}$ and treat it as SMK for parameterizing spatial topology $\tilde{P}^{(S)}$ and node-specific parameters $\Theta^{(S)}$, formally:

$$\begin{cases} E^{(S)} = \sigma(K^{(S)}W_S + b_S) \\ \tilde{P}^{(S)} = \text{softmax}(\text{relu}(E^{(S)}E^{(S)T})) \\ \Theta^{(S)} = \tanh(E^{(S)}W_\Theta) \end{cases} \quad (10)$$

where $W_S \in \mathbb{R}^{f_s \times e_s}$ and $W_\Theta \in \mathbb{R}^{e \times (C+1) \cdot v \times u}$ denote trainable parameters. Distilled by updating $\Theta^{(S)}$ ($E^{(S)}$ -dependent) in (both) STU, SMK $E^{(S)} \in \mathbb{R}^{N \times e_s}$ encapsulates the high-level spatial knowledge of diverse social covariate patterns.

3.2.2 Distilling Temporal Meta-knowledge. Given a set of consecutive timeslots $\{\tau_1, \tau_2, \dots, \tau_T\}$ and a group of attributes describing each timeslot (e.g., time-of-day, day-of-week, whether-holiday), a matrix for temporal knowledge $K^{(T)} \in \mathbb{R}^{T \times f_T}$ can be built. Different from spatial knowledge, temporal knowledge is time-dependent, indicating TMK of a corresponding observational sequence should also be a function of time. We thereby utilize the extracted hidden social representation D_τ by teacher model $\text{STU}^{(S)}$ to distill TMK. Our motivation is to augment $K^{(T)}$, which essentially depicts temporal periodicity and regularity, by injecting anomaly awareness brought by social covariate. $K^{(T)}$ is commonly taken in by linear projection [41, 43, 45], denoted by:

$$E_\tau^{(T)} = \sigma(\sigma(K_\tau^{(T)}W_{T_1} + b_{T_1})W_{T_2} + b_{T_2}) \quad (11)$$

where $W_{T_1} \in \mathbb{R}^{f_T \times e_T}$ and $W_{T_2} \in \mathbb{R}^{e_T \times 1}$ denote two parameter matrices. Here we propose to reparameterize them by the representation extracted from the teacher model, denoted by:

$$Q_\tau = \bar{D}_\tau W_Q + b_Q \quad (12)$$

$$\begin{cases} W'_{T_1} = Q_\tau W_{Q_1} + b_{Q_1} \\ W'_{T_2} = Q_\tau W_{Q_2} + b_{Q_2} \end{cases} \quad (13)$$

where $\bar{D}_\tau \in \mathbb{R}^{N \times u}$ denotes the vectorized D_τ , and $W_Q \in \mathbb{R}^{N \times u \times d}$ denotes a parameter matrix for projecting D_τ to a query vector $Q_\tau \in \mathbb{R}^d$, which can be interpreted as a global social representation at the moment. $W_{Q_1} \in \mathbb{R}^{d \times f_T \times e_T}$ and $W_{Q_2} \in \mathbb{R}^{d \times e_T \times N \times 1}$ denote two reparameterizing matrices for deriving W'_{T_1} and W'_{T_2} to replace W_{T_1} and W_{T_2} in Eq. 11. We denote this TMK distillation process as $E_\tau^{(T)} = \Gamma(D_\tau, K^{(T)})$ for simplicity.

Given distilled SMK and TMK, we manage to improve vanilla STN to **Meta-knowledge STN (MeSTN)** by letting $E^{(S)}$ parameterize $\tilde{P}^{(S)}$, $\Theta^{(X)}$ of the student model $\text{STU}^{(X)}$ (in the same way illustrated in Eq. 10) and update input $X'_\tau \leftarrow [X_\tau, E_\tau^{(T)}]$ ($[\cdot]$ denotes a concatenation operation).

3.3 Meta-knowledge Memorizable Spatio-Temporal Network

Although utilizing spatial and temporal meta-knowledge, MeSTN directly applies momentary query Q_τ as the prototype of reparameterization, which does not fully exploit similar socio-temporal patterns in history (considering the multiwaves of typhoons and pandemic as illustrated in Figure 1). Therefore, we are further motivated to encourage the teacher model to refer to historical lessons by constructing a socio-memory pool for storing and guiding TMK distillation. To be specific, we parameterize a memory bank $M \in \mathbb{R}^{m \times d}$, which consists of m vectors and each vector represents a socio-temporal prototype (e.g., normal days, disaster coming). Whenever the teacher model finishes encoding, it makes a query to M using Q_τ (defined in Eq. 12) for searching and retrieving

a similar prototype, denoted by:

$$\begin{cases} a_i = \frac{e^{Q_\tau * M_i^T}}{\sum_{i=1}^m e^{Q_\tau * M_i^T}} \\ Q'_\tau = \sum_{i=1}^m a_i \cdot M_i \end{cases} \quad (14)$$

where a_i is a similarity measure or attention score [2, 30] between the query Q_τ and i -th item of the memory pool, and the reconstructed prototype Q'_τ can further replace Q_τ in Eq. 13 for memory-guided TMK distillation $E_\tau^{(T)} = \Gamma_{mem}(D_\tau, M, K^{(T)})$. In our case, we want the prototypes learnt by the socio-memory bank as diverse as possible, to cover both (but not limited to) normal and abnormal socio-temporal patterns. Motivated by a line of research on memory-augmented anomaly detection [8, 24], we put constraints on the learnable parameters to contrast difference of memory items. Particularly, we take advantage of the triplet loss to implement a memory contrastive constraint by treating query vector Q_τ as anchor, its most similar memory item M_p as positive sample, and the second similar memory item M_n as negative sample.

$$\mathcal{L}_{con} = \sum_{\tau=1}^T \max\{\|Q_\tau - M_p\|^2 - \|Q_\tau - M_n\|^2 + \rho, 0\} \quad (15)$$

where ρ denotes a margin between the positive and negative pairs. By applying this constraint, we encourage Q_τ to be near to M_p but distant from M_n to push memory items far away from each other, which essentially guides the memory bank to discriminate various socio-temporal patterns. Thus far, we have enhanced meta-knowledge STN to our proposed **Meta-knowledge Memorizable Spatio-Temporal Network (MemeSTN)**, as illustrated in Figure 3. The entire model is optimized by minimizing the weighted sum of prediction loss (*i.e.*, L1 Loss) and contrastive constraint:

$$\mathcal{L}_{task} = \sum_{\tau=1}^T \sum_{\epsilon=1}^{\beta} |\hat{X}_{\tau+\epsilon} - X_{\tau+\epsilon}| + \lambda \mathcal{L}_{con} \quad (16)$$

4 EXPERIMENT

4.1 Datasets

We take three real-world disasters as our experimental targets: *2019 Typhoon in Japan*, *2020 COVID19 in Japan*, and *2019 Hurricane in US*. The specifications of the multimodal datasets with different scales (country-level or state-level) are summarized in Table 1. Blog-

Table 1: Summary of Datasets

Dataset	Typhoon-JP	COVID-JP	Hurricane-US
Time	2019/7/1-2019/10/30 122 days by hour	2020/1/1-2021/2/28 425 days by hour	2019/7/1-2019/9/10 72 days by hour
Space	Japan 47 prefectures		Florida 67 counties
Mobility	Inflow and outflow volume		POI visit volume
Twitter	Disaster-related tweet number in each prefecture/county		

watcher GPS trajectory data is used as the Japan human mobility data, while SafeGraph POI visitation data is used as the US human mobility data. Geo-tagged tweets related to the disasters are collected through Twitter API V2. After preprocessing (*e.g.*, trip segmentation and spatio-temporal aggregation), human inflow and

outflow X in Typhoon case are respectively generated as a (2928, 47) tensor. Social covariate S is also a (2928, 47) tensor. Similarly, the inflow, outflow, and social covariate in COVID case are respectively a (10185, 47) tensor. The mobility and social tensor in Hurricane case are both (1728, 67). For each hour, we generate a 32-dimension vector as the temporal metadata, formed by hour-of-day (24), day-of-week (7), and is-holiday (1). Refer to Appendix for more details.

4.2 Settings

Nowcasting human mobility (*i.e.*, inflow, outflow, POI visit) in each of the disasters (*i.e.*, typhoon, COVID, hurricane) is conducted as five independent experiments. For each experiment, 60%, 20%, and 20% of the whole dataset are split for training, validation, and testing in chronological order. The testing period contains both normal days and disaster days. Adam optimizer with batch size = 64 and learning rate = 0.001 is used for model training. The models will be trained up to 200 epochs. If the validation error does not improve for 10 consecutive epochs, the training process will be early-stopped. α and β are both set to 6, which means the past 6-hour observations are used to nowcast the next 6 hours of human mobility volume. Our model and the baselines were implemented with Python 3.8.8 and PyTorch 1.9.1. The mobility data and the Twitter data were respectively normalized to range [-1, 1] with MinMaxScaler method provided by *scikit-learn*. The predicted values are rescaled back and the performance of the multi-step mobility prediction is evaluated using three metrics: *RMSE* (Root Mean Square Error), *MAE* (Mean Absolute Error), and *MAPE* (Mean Absolute Percentage Error). Two layers of GCRN with RNN hidden states = 64 and graph convolution kernel size = 2 are stacked together as the STN bone. 10 memory prototypes, each represented by an 8-dimension vector, form the social memory pool. The balancing factor λ for the two loss terms is tuned from 0 to 1.

4.3 Overall Evaluation

Baselines. We first implement two non-deep learning baselines. (1) **Historical Average (HA)**: it averages the historical values of the corresponding hours. (2) **CopyLastWeek**: it copies the corresponding inflow/outflow values from last week. The results are further corrected with a recent weekday or weekend if the target day or the “last week” day is a holiday and the other is not. Then we compare our model with the following deep learning models: (3) **STGCN** [39], (4) **ASTGCN** [9], (5) **DCRNN** [20], (6) **GW-Net** [36], (7) **LSTNet** [19], (8) **GMAN** [43], (9) **MTGNN** [35], (10) **AGCRN** [3], and (11) **STTN** [37]. For (3)~(6), we use the binary adjacency matrix in Definition 2 as the input graph, while (7)~(11) do not need the adjacency matrix as the input. To make a fair comparison with our model, we put the Twitter information as the second input channel for all baselines by following the similar strategy with [25]. Through this, LSTNet [19] can be seen as an improved version of [25], so we omit [25] from our baselines.

Overall Performance. In Table 2, we compare the overall performance between the baselines and the proposed model over the datasets, namely {Typhoon-JP, COVID-JP} × {Inflow, Outflow} and Hurricane-US (POI Visit). CopyLastWeek corrected for holidays works well and has relatively low MAEs and MAPEs as human mobility holds a strong weekly pattern in normal situation. Meanwhile, the graph-based deep learning models and the MTS models

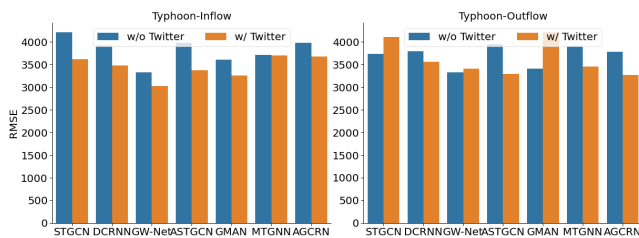
Table 2: Performance Comparison with Baselines on {Typhoon-JP, COVID-JP, Hurricane-US}

Model	Typhoon-JP						COVID-JP						Hurricane-US		
	Inflow			Outflow			Inflow			Outflow			POI Visit		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
HA	4061.2	1281.3	13.39%	4022.6	1289.6	13.44%	1986.3	818.1	14.19%	1963.6	820.7	14.27%	966.2	337.3	37.46%
CopyLastWeek	3980.5	947.3	10.81%	3981.8	949.0	10.77%	1531.3	536.1	10.12%	1522.6	536.3	10.12%	837.2	253.4	29.90%
STGCN[39]	3611.2	1308.3	16.43%	4098.3	1585.6	22.35%	1866.8	719.3	14.50%	1713.0	662.9	13.86%	1111.6	387.0	49.40%
ASTGCN[9]	3366.6	1065.1	15.18%	3288.1	1067.5	15.72%	1546.4	592.8	12.96%	1508.0	542.1	12.13%	883.8	278.3	32.12%
DCRNN[20]	3470.8	1081.8	12.66%	3551.3	1066.8	13.59%	1466.0	523.2	12.96%	1996.6	656.0	15.41%	864.4	291.0	37.46%
GW-Net[36]	<u>2856.4</u>	951.6	12.28%	<u>2922.3</u>	1007.7	13.68%	<u>1282.2</u>	<u>515.8</u>	13.51%	<u>1232.3</u>	530.6	14.75%	<u>786.6</u>	266.9	35.86%
LSTNet [19]	4635.9	1604.9	18.69%	4981.9	1672.8	21.50%	1744.1	644.1	16.24%	1545.0	637.0	14.64%	1059.3	354.5	39.39%
GMAN[43]	3248.8	1000.6	13.24%	4209.3	1188.1	17.37%	1367.1	543.9	10.34%	1431.3	602.0	11.17%	992.4	312.2	33.03%
MTGNN[35]	3698.4	1189.7	17.76%	3454.4	1152.3	17.61%	1616.5	597.8	14.05%	1329.7	<u>522.4</u>	12.32%	901.7	317.0	39.61%
AGCRN[3]	3675.7	1132.8	18.45%	3266.5	1033.6	16.07%	1601.2	603.1	15.96%	1532.8	592.8	15.60%	1023.7	346.4	41.09%
STTN[37]	3628.2	1207.5	18.25%	3431.0	1172.8	16.40%	1388.8	555.4	13.34%	1325.9	537.6	13.14%	1197.7	398.5	58.77%
MemeSTN	2470.5	805.8	9.62%	2491.0	822.1	9.87%	1125.4	438.8	9.68%	1079.6	424.0	9.60%	718.6	247.2	28.96%

Table 3: Performance Comparison with Model Variants on {Typhoon-JP, COVID-JP, Hurricane-US}

Model Variant	Typhoon-JP						COVID-JP						Hurricane-US		
	Inflow			Outflow			Inflow			Outflow			POI Visit		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
STN [Sec. 3.1]	3933.8	1283.2	18.86%	3257.0	1038.9	15.33%	1653.9	610.8	16.08%	1743.0	728.0	19.49%	1221.3	428.5	54.16%
MeSTN [Sec. 3.2]	2833.7	908.3	12.02%	2735.9	874.5	10.99%	1245.8	472.3	10.75%	1169.5	468.6	11.00%	770.9	271.5	30.25%
MemeSTN (w/ noise)	2991.8	1002.6	13.35%	2738.5	916.4	12.06%	<u>1234.4</u>	484.7	12.41%	1172.8	<u>451.2</u>	<u>10.24%</u>	788.8	279.9	31.92%
MemeSTN [Sec. 3.3]	2470.5	805.8	9.62%	2491.0	822.1	9.87%	1125.4	438.8	9.68%	1079.6	424.0	9.60%	718.6	247.2	28.96%

have their own advantages, among which GW-Net [36] based on Diffusion Graph Convolution and WaveNet generally achieved the second best performance. Our proposed model (MemeSTN) reaches the best performance in all metrics over all of the datasets. Besides, we compare the performances between using or without using Twitter information as the auxiliary input to the baselines in Figure 4, where we take RMSEs on Typhoon dataset as the representative. We can see that the performance gain could be achieved by some specific baselines over some specific datasets, but they were not consistent over all cases. This implies that (1) the social covariate cannot be straightforwardly learned by the state-of-the-art spatio-temporal models, and (2) it is necessary for us to explicitly extract and learn the knowledge from the social covariate by designing the corresponding modules. In terms of runtime, our trained model only takes 2 milliseconds to perform one prediction.

**Figure 4: Performances of Baseline Models with/without Social Covariate (Twitter Information).**

Ablation Study. Table 3 summarizes the performances of the variants over all the datasets. Specifically, (1) STN is just the STN bone of MemeSTN, i.e., the stacked two layers of GCRN that only takes the human mobility as the input and output; (2) MeSTN excludes the

memory module from MemeSTN and directly generate the social meta-knowledge and temporal meta-knowledge (TMK) through the social STN, which still takes both social covariate and human mobility as the input; (3) MemeSTN (w/ noise) uses the random floats with range $[-1, 1]$ as the noisy covariate to replace the social covariate (i.e., the Twitter data). Through Table 3, we can see: (1) compared with vanilla STN, the Meme module with social covariate (MemeSTN) gives us a huge performance gain (almost 40%); (2) memory-guided distillation mechanism in MemeSTN improves the overall performance of MeSTN by approximately 10%~12%; (3) compared with MemeSTN (w/ noise), we confirm that the social covariate did contain the essential information that can guide the human mobility nowcasting in disasters.

4.4 Case Study

Time Series. Typhoon Hagibis made landfall on the Greater Tokyo Area on October 12, 2019 at 8:00 pm. It is significant for us to verify whether our model can successfully adapt to such a sudden event from the normal situation. Thus, based on population and affected degrees, we select six representative prefectures, namely Tokyo, Chiba, Kanagawa, Osaka, Aichi, and Hokkaido, and plot the 1 hour ahead nowcasting results of the human outflow. The time-series charts in Figure 5 report the ground-truth and the prediction results from our model MemeSTN and three baseline models, namely CopyLastWeek, GW-Net [36], and GMAN [43]. These baselines all achieved relatively good overall performances as listed in Table 2. The baseline models could perform rather well before and after the disaster, namely the normal situation. However, they were quite struggling for the disaster period, especially CopyLastWeek and GMAN [43]. CopyLastWeek can't make a good "copy" for such an unprecedented disaster. Traditional regression models (e.g., panel

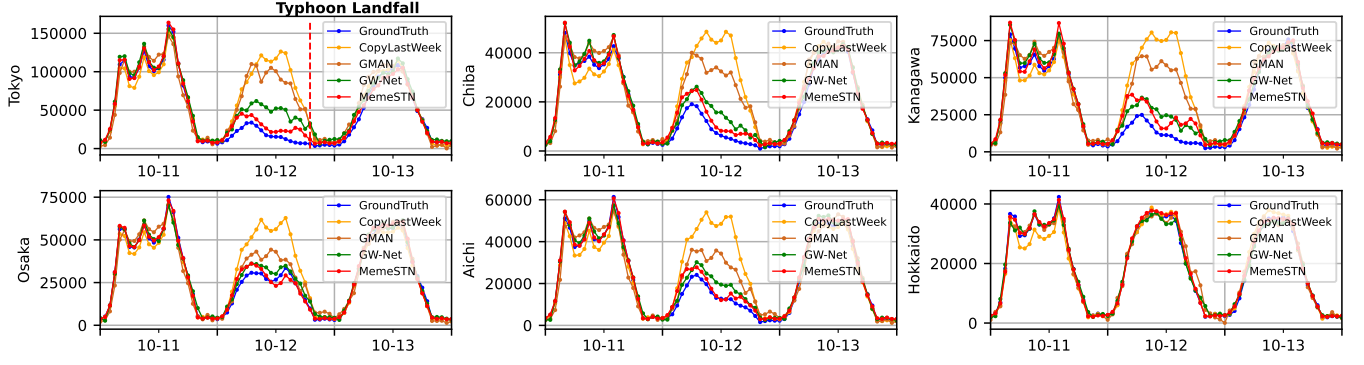


Figure 5: Case Study: 1-Hour Ahead Human Outflow Nowcasting during Typhoon Hagibis for Six Japan Prefectures.

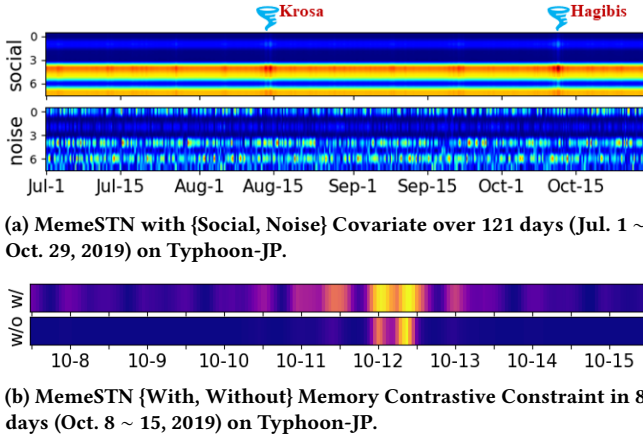


Figure 6: Query Weights of Selected Memory Items.

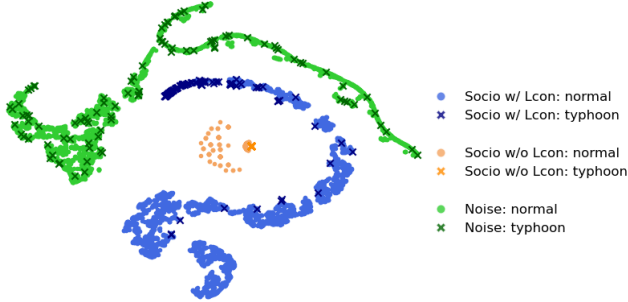


Figure 7: Visualization of Reconstructed Prototype (Q'_r) Embeddings of MemeSTN Variants with t-SNE on Typhoon-JP.

regression, mixed effects) are expected to have the same issue. GMAN [43] is a transformer-based deep model that requires the pre-calculated spatial and temporal embedding vectors as the input, corresponding to the position encoding module in Transformer. Such mechanism tends to make the model fit more to the periodicity (i.e., the regular time periods) and have less capability to fast adapt to the abnormality (i.e., the disaster time periods). On the other hand, GW-Net [36] designs an adaptive graph with learnable

parameters to gain the good adaptability to the current pattern, and thus it could achieve a relatively good performance in the disaster situation comparing to GMAN [43]. But still, MemeSTN could show a superior performance to GW-Net [36] on different prefectures. Note that Table 2 shows the general performance (most on normal situations) and Figure 5 focused more on the disaster date. By combining both, we can verify that our model can not only learn the normality but also hold good adaptability to the abnormality, which is the core contribution we would like to appeal for our study.

Latent Space. We further examine the latent space of MemeSTN in two steps. *Firstly*, Figure 6 illustrates the query weights of items in socio-memory pool through the typhoon season and under a selected typhoon. In Figure 6a, comparing two kinds of input signals, namely white noise and social covariate, we confirm the informativeness of the latter and the benefits by introducing this memory design, which not only regulates the input (certain memory items are queried frequently throughout the period) but memorize similar patterns in history for reacting to new ones (i.e. Typhoon Krosa and Hagibis). This capability is especially important when we are handling events with potential multiple strikes or waves (e.g. hurricane, pandemic). In Figure 6b, we verify the effectiveness of memory contrastive constraint. It is noticeable $\mathcal{L}_{con}$ boosts the power of socio-memory to echo about a specific event (i.e. typhoon Hagibis), before and after its occurrence. *Secondly*, Figure 7 shows 2D embeddings of reconstructed prototypes (Q'_r) produced by different MemeSTN variants. Apparently, white noise signal fails to help model distinguish typhoon (anomalous) cases from normal situations. In comparison, hour-level samples during typhoons are mostly clustered and differentiated from normal ones by social covariate input, which means we are using a good indicator for event recognition and prediction guiding. At the same time, the implementation with $\mathcal{L}_{con}$ tends to have a broader socio-temporal distribution than the one without it, which demonstrates the discriminative decomposing power brought by this constraint.

4.5 Hyperparameter Study

First, we conduct the hyperparameter study by varying the observation and prediction horizons among {3/3, 6/6, 9/9, 12/12}. The typhoon case is far more sensitive to these hyperparameters comparing with the COVID case, which may be ascribed to the stronger

fluctuation of the human mobility within a relatively short time window. Therefore, we plot the prediction errors (RMSE) for typhoon case at two specific horizons: 1st prediction horizon (1 hour ahead) shown in Figure 8-left and 3rd prediction horizon (3 hours ahead) shown in Figure 8-right. We can see $\alpha, \beta = 6, 6$ marks the lowest RMSEs for both 1 hour and 3 hours ahead prediction. Thus, they are chosen as the final hyperparameters so as to make nowcasting as long as possible while maintaining a high accuracy.

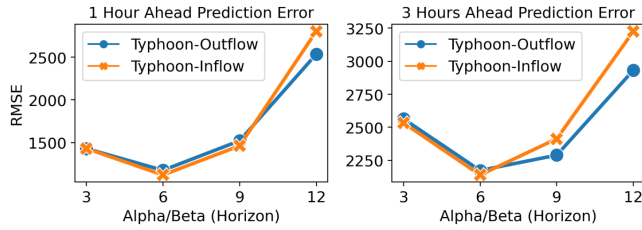


Figure 8: Hyperparameter Study on Observation/Prediction Horizons (α/β).

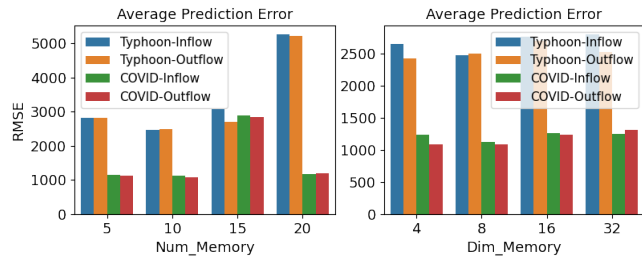


Figure 9: Hyperparameter Study on Number of Memory Prototypes and Dimension of Memory Prototype.

Second, we conduct the hyperparameter study on the number of memory prototypes and the dimension of each prototype in Memory-guided Meta-knowledge distiller (*Meme* module). As shown by Figure 9, 10 memory prototypes give us the best performance across all datasets among the options {5, 10, 15, 20}; when the dimension of each memory prototype varies among {4, 8, 16, 32}, 8 is generally the best across all datasets. Large number and dimension may be difficult to make the model converge, while small number and dimension may not be enough to cover all the underlying patterns. Therefore, $\{\text{Num_Memory}, \text{Dim_Memory}\} = \{10, 8\}$ is chosen as the final hyperparameter for *Meme* module.

5 RELATED WORK

Under major events or disasters, human mobility would vary from its routine and regular patterns dramatically. To address this, machine learning models to detect large deviations from the routine as anomalies have been developed based on tensor decomposition [34, 42], deep neural networks [1, 12, 21], Bayesian inference [11], and other techniques [40, 42]. A recent approach that efficiently explores the abnormal mobility patterns in a broader area and longer period was introduced by integrating the multi-head attention module in the deep learning model [13]. Wu et al. proposed an anomaly

prediction model based on Gibbs sampling and tensor decomposition to forecast future incidents based on the historical mobility data [34]. However, these models can only detect and predict the abnormal events and do not learn the patterns of human mobility under abnormal scenarios. To this end, an early approach [28] was developed to leverage additional disaster-related information such as locations and intensities of the events to simulate individual-level human trajectories after disasters. Furthermore, the severity of COVID-19 and local policy interventions were incorporated to train a conditional generative adversarial network to estimate the human mobilities under the COVID-19 pandemic [4]. Unfortunately, these methods can hardly be used for applications in disaster management and responses since the required additional disaster-related information are usually not available immediately after the disasters. To mitigate this issue, without using additional data, some deep learning models were proposed to predict the short-term crowd dynamics [14, 16], transportation demands [32], and individual trajectories [7, 15] under various major events. But nowcasting human mobility in disasters remains challenging for nationwide spatial scale and regions in different countries.

Web and social media technologies can provide essential information for situation understanding. For example, web search logs were found effective to predict residents' evacuation decisions in disaster [38]. In this paper, a different problem in this direction is tackled, i.e. nowcasting human mobility in disasters. Specifically, we exploit the integration of social co-variants based on timely social media data to improve the human mobility prediction. Recently, [25] proposed an LSTM model that integrates social media and evacuation flow volume data to forecast the traffic flow for specific highways under several major hurricanes in Florida. However, this naive LSTM-based model could not well capture the spatial correlations among regions or roads.

6 CONCLUSION

In this study, disaster-related Twitter data is incorporated to understand the public awareness and attention about the disaster events and thus perceive their impacts on the human mobility. Towards this, we propose a Meta-knowledge-Memorizable Spatio-Temporal Network (MemeSTN) that leverages memory network and meta-learning techniques to fuse the past social media and human mobility data and predict future human mobility under a disaster. Extensive experiments over Typhoon-JP, COVID-JP, and Hurricane-US datasets were conducted to verify the effectiveness of our model on different scales. In the future, we plan to improve the data preprocessing part by extracting semantic information from the social media data with advanced natural language processing techniques (e.g., sentiment analysis, topic modeling) and taking the accurate semantic information as the social covariate in our model. Meanwhile, we will try to validate our model on more disaster scenarios such as earthquake, flood, and blizzard. Also, we will investigate the possibility of applying our method to the origin-destination flow nowcasting task at nationwide scale [17].

ACKNOWLEDGMENTS

This work was supported by NSF Grant Number CNS-2125165 and JST SICORP Grant Number JPMJSC2104.

REFERENCES

- [1] Soto Anno, Kota Tsubouchi, and Masamichi Shimosaka. 2020. Supervised-CityProphet: Towards Accurate Anomalous Crowd Prediction. In *28th International Conference on Advances in Geographic Information Systems*. ACM, 175–178.
- [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] Lei Bai, Lina Yao, Can Li, Xianzhi Wang, and Can Wang. 2020. Adaptive Graph Convolutional Recurrent Network for Traffic Forecasting. In *34th Conference on Neural Information Processing Systems*.
- [4] Han Bao, Xun Zhou, Yingxue Zhang, Yanhua Li, and Yiqun Xie. 2020. Covid-gan: Estimating human mobility responses to covid-19 pandemic through spatio-temporal conditional generative adversarial networks. In *Proceedings of the 28th international conference on advances in geographic information systems*. 273–282.
- [5] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems*. 3844–3852.
- [6] 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 *The Thirty-Third AAAI Conference on Artificial Intelligence*. AAAI Press, 890–897.
- [7] Zipei Fan, Xuan Song, Tianqi Xia, Renhe Jiang, Ryosuke Shibasaki, and Ritsu Sakuramachi. 2018. Online deep ensemble learning for predicting citywide human mobility. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–21.
- [8] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. 2019. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1705–1714.
- [9] Shengnan Guo, Youfang Lin, Ning Feng, Chao Song, and Huaiyu Wan. 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 922–929.
- [10] Binxuan Huang and Kathleen M. Carley. 2019. A Large-Scale Empirical Study of Geotagging Behavior on Twitter. In *International Conference on Advances in Social Networks Analysis and Mining*. 365–373.
- [11] Chao Huang, Xian Wu, and Dong Wang. 2016. Crowdsourcing-based urban anomaly prediction system for smart cities. In *Proceedings of the 25th ACM international conference on information and knowledge management*. 1969–1972.
- [12] Chao Huang, Chuxu Zhang, Jiahu Zhao, Xian Wu, Dawei Yin, and Nitesh Chawla. 2019. Mist: A multiview and multimodal spatial-temporal learning framework for citywide abnormal event forecasting. In *The World Wide Web Conference*. 717–728.
- [13] Huiqun Huang, Xi Yang, and Suining He. 2021. Multi-Head Spatio-Temporal Attention Mechanism for Urban Anomaly Event Prediction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3 (2021), 104:1–104:21.
- [14] Renhe Jiang, Zekun Cai, Zhaonan Wang, Chuang Yang, Zipei Fan, Qunjun 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 (2022), 1–24.
- [15] Renhe Jiang, Xuan Song, Zipei Fan, Tianqi Xia, Qunjun Chen, Satoshi Miyazawa, and Ryosuke Shibasaki. 2018. DeepUrbanMomentum: An Online Deep-Learning System for Short-Term Urban Mobility Prediction. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. AAAI Press, 784–791.
- [16] Renhe Jiang, Xuan Song, Dou Huang, Xiaoya Song, Tianqi Xia, Zekun Cai, Zhaonan Wang, Kyoung-Sook Kim, and Ryosuke Shibasaki. 2019. DeepUrbanEvent: A System for Predicting Citywide Crowd Dynamics at Big Events. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2114–2122.
- [17] Renhe Jiang, Zhaonan Wang, Zekun Cai, Chuang Yang, Zipei Fan, Tianqi Xia, Go Matsuura, Hiroto Mizuseki, Xuan Song, and Ryosuke Shibasaki. 2021. Countrywide origin-destination matrix prediction and its application for covid-19. In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part IV 21*. Springer, 319–334.
- [18] Renhe Jiang, Du Yin, Zhaonan Wang, Yizhuo Wang, Jiewen Deng, Hangchen Liu, Zekun Cai, Jinliang Deng, Xuan Song, and Ryosuke Shibasaki. 2021. DL-Traffic: Survey and Benchmark of Deep Learning Models for Urban Traffic Prediction. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 4515–4525.
- [19] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. 2018. Modeling long- and short-term temporal patterns with deep neural networks. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 95–104.
- [20] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. 2018. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In *International Conference on Learning Representations*.
- [21] Ruiqiang Liu, Shuai Zhao, Bo Cheng, Hao Yang, Haina Tang, and Taoyu Li. 2020. Deep Spatio-Temporal Multiple Domain Fusion Network for Urban Anomalies Detection. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 905–914.
- [22] Zheyi Pan, Yuxuan Liang, Weifeng Wang, Yong Yu, Yu Zheng, and Junbo Zhang. 2019. Urban traffic prediction from spatio-temporal data using deep meta learning. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1720–1730.
- [23] Zheyi Pan, Zhaoyuan Wang, Weifeng Wang, Yong Yu, Junbo Zhang, and Yu Zheng. 2019. Matrix factorization for spatio-temporal neural networks with applications to urban flow prediction. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 2683–2691.
- [24] Hyunjong Park, Jongyoun Noh, and Bumsul Ham. 2020. Learning memory-guided normality for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14372–14381.
- [25] Kamol Chandra Roy, Samiul Hasan, Aron Culotta, and Naveen Eluru. 2021. Predicting traffic demand during hurricane evacuation using Real-time data from transportation systems and social media. *Transportation research part C: emerging technologies* 131 (2021), 103339.
- [26] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*. 851–860.
- [27] Samuel S. Sohn, Honglu Zhou, Seonghyeon Moon, Sejong Yoon, Vladimir Pavlovic, and Mubbasir Kapadia. 2020. Laying the Foundations of Deep Long-Term Crowd Flow Prediction. In *16th European Conference on Computer Vision (Lecture Notes in Computer Science, Vol. 12374)*. Springer, 711–728.
- [28] Xuan Song, Quanshi Zhang, Yoshihide Sekimoto, Ryosuke Shibasaki, Nicholas Jing Yuan, and Xing Xie. 2015. A simulator of human emergency mobility following disasters: Knowledge transfer from big disaster data. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- [29] Gábor J Székely and Maria L Rizzo. 2014. Partial distance correlation with methods for dissimilarities. *The Annals of Statistics* 42, 6 (2014), 2382–2412.
- [30] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.
- [31] 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 *Proceedings of The Web Conference 2020*. 1082–1092.
- [32] Zhaonan Wang, Renhe Jiang, Hao Xue, Flora D Salim, Xuan Song, and Ryosuke Shibasaki. 2022. Event-aware multimodal mobility nowcasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 4228–4236.
- [33] YABAI Writers. 2017. Japan 47 Prefectures. <http://yabai.com/p/2098>.
- [34] Xian Wu, Yuxiao Dong, Chao Huang, Jian Xu, Dong Wang, and Nitesh V. Chawla. 2017. UAPD: Predicting Urban Anomalies from Spatial-Temporal Data. In *European Conference on Machine Learning and Knowledge Discovery in Databases (Lecture Notes in Computer Science, Vol. 10535)*. Springer, 622–638.
- [35] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojuan Chang, and Chengqi Zhang. 2020. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 753–763.
- [36] Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, and Chengqi Zhang. 2019. Graph wavenet for deep spatial-temporal graph modeling. In *IJCAI*. 1907–1913.
- [37] Mingxing Xu, Wenrui Dai, Chunmiao Liu, Xing Gao, Weiyaolun Lin, Guo-Jun Qi, and Hongkai Xiong. 2020. Spatial-temporal transformer networks for traffic flow forecasting. *arXiv preprint arXiv:2001.02908* (2020).
- [38] Takahiro Yabe, Kota Tsubouchi, Toru Shimizu, Yoshihide Sekimoto, and Satish V Ukkusuri. 2019. Predicting Evacuation Decisions using Representations of Individuals' Pre-Disaster Web Search Behavior. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2707–2717.
- [39] Bing Yu, Haoteng Yin, and Zhanxing Zhu. 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 3634–3640.
- [40] Huichu Zhang, Yu Zheng, and Yong Yu. 2018. Detecting Urban Anomalies Using Multiple Spatio-Temporal Data Sources. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1 (2018), 54:1–54:18.
- [41] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. AAAI Press, 1655–1661.
- [42] Mingyang Zhang, Tong Li, Hongzhi Shi, Yong Li, Pan Hui, et al. 2019. A decomposition approach for urban anomaly detection across spatiotemporal data. In *IJCAI International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence.
- [43] Chuanpan Zheng, Xiaoliang Fan, Cheng Wang, and Jianzhong Qi. 2020. Gman: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 1234–1241.
- [44] Xian Zhou, Yanyan Shen, Yanmin Zhu, and Linpeng Huang. 2018. Predicting multi-step citywide passenger demands using attention-based neural networks.

In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 736–744.

- [45] Ali Zonoozi, Jung-jae Kim, Xiao-Li Li, and Gao Cong. 2018. Periodic-CRN: A Convolutional Recurrent Model for Crowd Density Prediction with Recurring Periodic Patterns.. In *IJCAI*. 3732–3738.

APPENDIX

A NOTATION TABLE

Table 4 describes all the important notations used in this paper.

Table 4: Description of Important Notations

Symbol	Description
N, T	Total number of regions and timeslots
$X_\tau, \hat{X}_\tau$	Real and predicted human mobility at time τ
S_τ	Social covariate tensor at time τ
$STU^{(X)}$	The STU encoder for human mobility
$STU^{(S)}$	The STU encoder for social covariate
TDC	The TDC decoder for human mobility
$\mathcal{A}$	Binary adjacency matrix of regions
$\tilde{P}$	Self-learnable adaptive graph
$W_c, W_s, \dots$	Weights in MemeSTN
$b_u, b_r, \dots$	Biases in MemeSTN
$\Theta_u, \Theta_r, \dots$	Node-specific parameters in MemeSTN
$K^{(S)}, K^{(T)}$	Input spatial and temporal metadata
$\tilde{P}^{(S)}, E^{(T)}$	Extracted SMK and TMK
M_i	The i -th prototype in memory pool

B EXPERIMENT

B.1 Datasets

Japan Human Mobility Data. The big human GPS trajectory data in Japan are obtained via the collaboration with Blogwatcher Inc. The data covers 5 million people in the 47 prefectures of Japan. Under user’s consent, the data is acquired from smartphone apps with the built-in module from Blogwatcher Inc, which contains six attributes: anonymized ID, timestamp, longitude, latitude, accuracy, and OS type. No personally identifiable information was collected. The raw data file involves around 180 GPS records per day per user and is roughly 1TB in the CSV format. The data covers around 9% of the total population of Japan. On average, 10 records of GPS data (either the origin or destination location) for each ID remain after data cleaning and trip segmentation. The linear regression between the population proportion and Census data for all the prefectures is $R^2 \geq 0.8$, which shows the good representativeness of our data. The human inflow and outflow tensor will be further generated according to *Definition 1*.

US Human Mobility Data. SafeGraph Point-of-Interest (POI) visitation data (<https://www.safegraph.com/academics>) records the hourly visit number for a large number of POIs in US from 2019 to 2020. The POI information includes state/county, longitude/latitude, location name, and the location type (category). Taking Miami metro area as an example, there are 40,964 POIs in total. We crop the data in Florida state during 2019 Atlantic hurricane season out and aggregate the total POI visits by county and hour.

Twitter Data. We collect the geo-tagged tweets related to the selected disasters in Table 1 through the full-archive search endpoint

in Twitter API V2 (<https://api.twitter.com/2/tweets/search/all>), and aggregate the total number of tweets for each prefecture. It should be noted that accurately distinguishing whether a tweet contains disaster-related information is a task of natural language processing (NLP), which is not the main concern of our study. Therefore, we simplify the procedure by using a set of pre-defined keywords to extract the relevant tweets. Using keywords *typhoon*, *strong wind*, *windproof*, and *big rain*, we collect a total 347,431 typhoon-related tweets from 89,324 users in Japan. Using keywords *COVID*, *infection*, *quarantine*, and *state of emergency*, we collect a total of 1,470,744 COVID-related tweets from 191,720 users in Japan. Using keywords *hurricane*, *strong wind*, *windproof*, *big rain*, and *storm*, we collect a total of 36,629 hurricane-related tweets from 15,908 users in Florida.

B.2 Settings

Metrics. Our evaluation metrics including *RMSE* (Root Mean Square Error), *MAE* (Mean Absolute Error), and *MAPE* (Mean Absolute Percentage Error) are defined as follows:

$$RMSE = \sqrt{\frac{1}{T_{test}} \sum_{\tau=1}^{T_{test}} \|\hat{Y}_\tau - Y_\tau\|^2}$$

$$MAE = \frac{1}{T_{test}} \sum_{\tau=1}^{T_{test}} |\hat{Y}_\tau - Y_\tau|$$

$$MAPE = \frac{1}{T_{test}} \sum_{\tau=1}^{T_{test}} \left| \frac{\hat{Y}_\tau - Y_\tau}{Y_\tau} \right|$$

where T_{test} is the total number of testing samples (timeslots), $Y_\tau \in \mathbb{R}^{\beta \times N}$ and $\hat{Y}_\tau \in \mathbb{R}^{\beta \times N}$ are the ground-truth tensor and predicted tensor, namely $[X_{\tau+1}, X_{\tau+2}, \dots, X_{\tau+\beta}]$ and $[\hat{X}_{\tau+1}, \hat{X}_{\tau+2}, \dots, \hat{X}_{\tau+\beta}]$. Zero values in ground-truth will be ignored when calculating MAPE.

Model Input/Output. Taking one batch of Typhoon-JP dataset as an example, the inputs to our model are: human mobility (in/outflow) tensor (64, 6, 47), social covariate tensor (64, 6, 47), and temporal knowledge tensor for both the past and the future 6 hours (64, 12, 32). The output human mobility is a tensor (64, 6, 47). In practice, we append an additional channel axis (64, 6, 47, 1) to the human mobility and social covariate tensor for the GCN operation.

Experiment Machine. Our machine is a GPU server with four *NVIDIA GeForce RTX 3090* graphics cards.

B.3 Social Covariate Effectiveness

As illustrated in Figure 1, the abnormal patterns of human mobility data under disasters co-occur with the “hills” of disaster-related tweet count. To further quantify the effectiveness of social covariate for human mobility nowcasting during the disaster events, we measure the partial distance correlation (PDisCorr) [29] between $[S_{\tau-\alpha+1}, \dots, S_{\tau-1}, S_\tau]$ and $[X_{\tau+1}, X_{\tau+2}, \dots, X_{\tau+\beta}]$ with respect to $[X_{\tau-\alpha+1}, \dots, X_{\tau-1}, X_\tau]$ and test whether its value equals to zero. This metric measures the correlation between past tweet count and future human mobility volume after removing the effects of past human mobility volume. To focus on the periods impacted by disaster events, a subset of our collected multimodal data, two days before and after typhoon events in 2019 Typhoon Season and three weeks before the peak of confirmed cases for each wave of

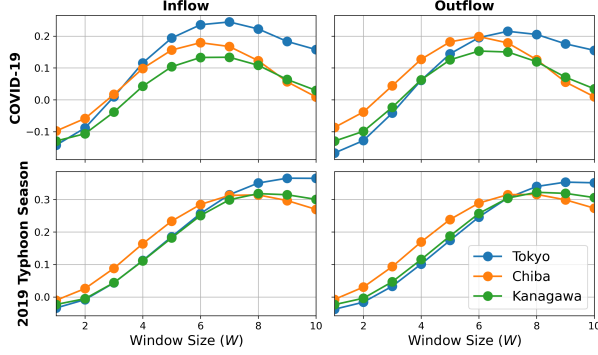


Figure 10: The partial distance correlations (PDisCorr) between the past tweet count and future human inflow/outflow volume with respect to the past human inflow/outflow volume at three selected prefectures in Japan. Non-zero partial distance correlation indicates the effectiveness of past disaster-related tweet count for human mobility nowcasting.

COVID-19 pandemic, is used for the analyses in this section. Figure 10 illustrates the PDisCorr metrics for both inflow and outflow volumes during COVID-19 and in 2019 Typhoon Season at various window sizes W , where we take $\alpha = \beta = W$ timeslots for both the past and future data to compute the PDisCorr. For all the scenarios when $5 \leq W \leq 8$, the values of PDisCorr do not equal to zero with $p < 0.01$, which suggests that disaster-related tweet count can help the human mobility nowcasting under disasters. Note that PDisCorr measures the correlation between the disaster-related tweet count and human mobility under disasters, not the causal relationship between them. Both disaster-related tweet count and human mobility under disasters are actually impacted by the disaster events. However, the complicated non-linear relationship between the social covariate and the human mobility remains unknown. To properly model their relationship and develop models for learning a better joint representation and more precise human mobility nowcasting under disasters, we propose MemeSTN based on spatio-temporal network and meta-knowledge learning techniques.

B.4 Evaluation on US Hurricane Dataset

We illustrate the normalized POI visit volume and hurricane-related tweet volume in three selected counties in Florida state in Figure 11. County Miami-Dade and Broward County together correspond to City of Miami, while Orange County covers Orlando City. These counties were impacted by hurricane Dorian on different days during September 1-3, 2019. We can clearly observe the decrease of POI visit and the increase of disaster tweets during this period. Moreover, we plot the time series charts in Figure 12 to show the prediction results on Hurricane dataset. Lastly, we do the t-SNE visualization for the reconstructed prototypes (Q'_t) produced by different MemeSTN variants (w/ socio or w/ noise) on Hurricane dataset in Figure 13. Apparently, socio signal helps our model better distinguish anomalous (hurricane) cases from normal situations, as hour-level samples during hurricane are more concentrated in the latent space of socio-input model and the hurricane samples scatter all over the latent space of noise-input model.

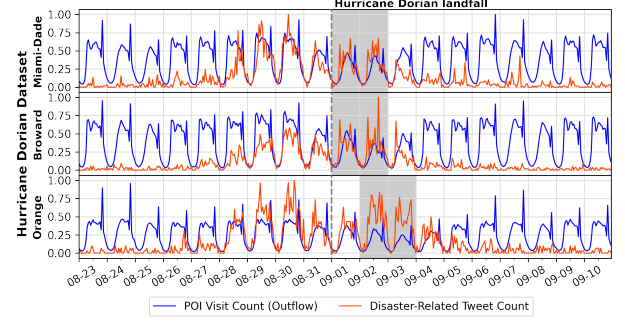


Figure 11: Illustration of POI Visit Number and Disaster-Related Tweet Number under Hurricane Dorian in US.

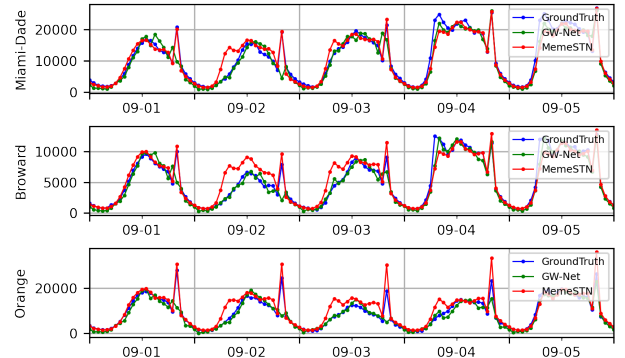


Figure 12: Case Study: 1-Hour Ahead POI Visit Nowcasting during Hurricane Dorian for Three Counties in Florida.

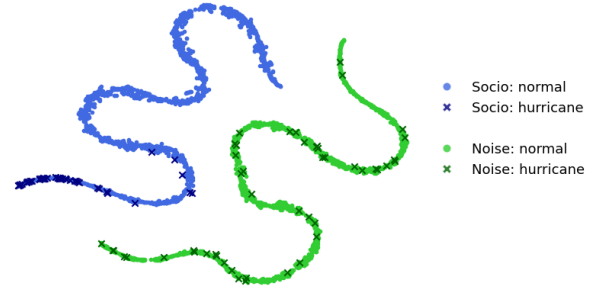


Figure 13: Visualization of Reconstructed Prototype (Q'_t) Embeddings of MemeSTN Variants with t-SNE on Hurricane-US

C DISCUSSION

We acknowledge that publicly accessible geo-tagged tweets only account for a very small proportion of the whole Twittersphere [10]. However, during the past severe disaster events, such as COVID-19 and Typhoon Hagibis, users in the impacted areas posted large amounts of disaster-related tweets. Thus, geo-tagged tweets can be effectively used as a good indicator to perceive people's awareness and attention of major disasters. For future disastrous events, the source of social covariate should be properly selected to reflect the real impact of disasters on users. Other geo-tagged social media data can also be used in the same way.