

Spatiotemporal Prediction of Foot Traffic

Samiul Islam
George Mason University
Fairfax, Virginia, USA
sislam22@gmu.edu

Dhruv Gandhi
George Mason University
Fairfax, Virginia, USA
dgandhi2@gmu.edu

Justin Elarde
George Mason University
Fairfax, Virginia, USA
jelarde.com

Taylor Anderson
George Mason University
Fairfax, Virginia, USA
tander6@gmu.edu

Amira Roess
George Mason University
Fairfax, Virginia, USA
aroess@gmu.edu

Timothy F. Leslie
George Mason University
Fairfax, Virginia, USA
tleslie@gmu.edu

Hamdi Kavak
George Mason University
Fairfax, Virginia, USA
hkavak@gmu.edu

Andreas Züfle
George Mason University
Fairfax, Virginia, USA
azufle@gmu.edu

ABSTRACT

Foot traffic is a business term to describe the number of customers that enter a point of interest (POI). This work aims to predict future foot traffic: the number of people from each census block group (CBG) that will visit each POI of a study region with potential applications in marketing and advertising. Existing techniques for spatiotemporal prediction of foot traffic use location-based social network data that suffer from sparsity, capturing only a handful of visits per day. This study utilizes highly granular foot traffic data from SafeGraph, a data company that collects mobility data regarding hundreds of millions of visits per day in the United States alone. Using this data, we explore solutions to predict weekly foot traffic data at the POI level. We propose a collaborative filtering approach using tensor factorization on the (POIs $\times$ CBGs $\times$ Weeks) data tensor. This approach provides us with a de-noised estimation of visits in previous weeks for all POI-CBG pairs. Using this tensor, we explore various time series prediction models: weekly rolling average, weighted weekly rolling average, univariate linear regression, polynomial regression, and long short-term memory (LSTM) recurrent neural networks. Our results show that of all the prediction models, the collaborative filtering step consistently improves prediction results. We also found that a simple weighted average consistently performed better than the more sophisticated approaches. Given this abundance of foot traffic data, this result shows that we can improve the spatiotemporal prediction of foot traffic data by harnessing collaborative filtering.

CCS CONCEPTS

• Information systems $\rightarrow$ Geographic information systems.

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.

LocalRec'21, November 2–5, 2021, Beijing, China

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9100-9/21/11...\$15.00

<https://doi.org/10.1145/3486183.3490997>

KEYWORDS

Spatiotemporal prediction, foot traffic, location prediction, next POI prediction, human mobility, visitor prediction

ACM Reference Format:

Samiul Islam, Dhruv Gandhi, Justin Elarde, Taylor Anderson, Amira Roess, Timothy F. Leslie, Hamdi Kavak, and Andreas Züfle. 2021. Spatiotemporal Prediction of Foot Traffic. In *5th ACM SIGSPATIAL International Workshop on Location-Based Recommendations, Geosocial Networks, and Geoadvertising (LocalRec'21)*, November 2–5, 2021, Beijing, China. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3486183.3490997>

1 INTRODUCTION

Spatiotemporal prediction, also known as spatiotemporal forecasting, is the task of predicting the future value of variables across both space and time. Applications of spatiotemporal prediction include road-traffic (flow) prediction to predict future states of a traffic network [15, 26, 27, 45], the prediction of future COVID-19 cases and deaths [25, 29, 34], and the prediction of bike-sharing systems to understand supply and demand at bike-rental stations [4, 18, 41], to name a few.

This study focuses on the spatiotemporal prediction of human foot traffic, that is, visits of individuals to public points of interest (POIs), which has applications in local marketing, geoadvertising, and location recommendation. For example, users who are predicted to visit a sports stadium may be interested in sports merchandise. Similarly, users who are predicted to visit bookstores regularly may be interested in discount vouchers for themselves and their friends.

Existing solutions for spatiotemporal prediction of foot traffic leverage location-based social network (LBSN) data or check-in data. LBSN data, as detailed in Section 2, is sparse, having only a few thousand check-ins per day distributed across the world. Such sparse data makes it challenging to model, learn, and predict human foot traffic patterns [12].

In order to overcome limitations associated with data sparsity, we use large-scale foot traffic data from SafeGraph¹, a data company that shares anonymized foot traffic data that has been collected through numerous smartphone applications. This data captures

¹For detailed information, see <https://docs.safegraph.com/docs/weekly-patterns>.

trips from census block groups (CBGs) to specific POIs. CBGs are defined as spatial neighborhoods that contain between 600 and 3,000 people. "A point of interest is a specific physical location which someone may find interesting. Restaurants, retail stores, and grocery stores are all examples of points of interest" [30]. To enhance privacy, SafeGraph excludes CBG information if fewer than two devices visited an establishment in a month [31].

In this study, we tackle the problem of predicting foot traffic one week into the future, for which data is not available. Towards achieving this goal, this paper is organized as follows. Section 2 surveys related work for spatiotemporal prediction of foot traffic data using check-ins in LBSN data. Then, Section 3 describes foot traffic data, a tensor representation of the data, and formally defines the problem of spatiotemporal foot traffic prediction. A collaborative filtering-based approach using tensor factorization to reduce noise in the data is presented in Section 4. Based on this noise-reduce tensor, a range of prediction methods are described in Section 5. Our experimental evaluation in Section 6 presents a case study using Fairfax County, VA, USA, which has a population of 1.1 million people, providing a benchmark for spatiotemporal foot traffic prediction using state-of-the-art solutions.

2 RELATED WORK

A large body of research established by the location-based social network (LBSN) research community proposes the use of LBSN data for spatiotemporal prediction of foot traffic, where visits to POIs are captured as check-ins. The task of spatiotemporal prediction using check-in data is often referred to as **location prediction** [1, 5, 19, 21, 35, 42]. These predictions can be used for **location recommendation**, which leverages check-ins of users and their ratings in the user-location network to recommend new locations to users (e.g., [10, 20, 24, 33, 38, 40, 43] for a survey see [2]).

While there has been a plethora of work for spatiotemporal prediction of foot traffic in LBSN research, all existing work suffers from a common shortcoming. LBSN data sets are trivially small [12]. The most commonly used LBSN dataset is the Gowalla dataset [22]. While this dataset captures more than 36 million check-ins, these are distributed over 20 months, averaging 60,000 check-ins per day, distributed globally. After removing users with less than 15 check-ins and removing locations with less than ten visitors, more than half of the visitors in the data are eliminated [23]. Thus, the Gowalla dataset, publicly available at the Stanford Network Analysis Project (SNAP) [7], only has 6 million check-ins total and only about 10,000 check-ins per day worldwide. That translates into only a dozen of check-ins per major city per day.

Consequently, it has been observed in [16] that the lower bound of predictability of the human spatiotemporal behavior, defined in [16], is as low as 27% using LBSN data. [16] concludes that "[r]esearchers working with LBSN data sets are often confronted by themselves or others with doubts regarding the quality or the potential of their data sets." and that "it is reasonable to be skeptical."

3 PROBLEM DEFINITION

Foot traffic data obtained from SafeGraph [31] provides us with weekly data on the number of visits to each POI grouped by CBG.

	CBG 1	CBG 2	CBG 3	...
POI 1	5	0	0	...
POI 2	0	7	0	...
POI 3	1	0	2	...
.	.	.	.	...
.	.	.	.	...
.	.	.	.	...

Figure 1: A visual depiction of the tensor

This section formally defines such foot traffic data and the problem of spatiotemporal prediction of foot traffic data.

Definition 3.1 (Foot Traffic Observation). Let $\mathcal{P} = \{P_1, \dots, P_{|\mathcal{P}|}\}$ be a set of POIs, let $\mathcal{R} = \{R_1, \dots, R_{|\mathcal{R}|}\}$ be a set of spatial regions such as census block groups, and let $\mathcal{T} = \{T_1, \dots, T_{|\mathcal{T}|}\}$ be a set of time intervals. A foot traffic observation $(p \in \mathcal{P}, r \in \mathcal{R}, t \in \mathcal{T})$ denotes an observed visit of a user from region r at POI p during time interval t .

We define a foot traffic database as a collection of foot traffic observations.

Definition 3.2 (Foot Traffic Database). A foot traffic database $\mathcal{D} = \langle o_1, \dots, o_N \rangle$ is a collection of foot traffic observations where each $o_i \in \mathcal{D}$ is a foot traffic observation $o_i = (p_i \in \mathcal{P}, r_i \in \mathcal{R}, t_i \in \mathcal{T})$.

We note that $\mathcal{D}$ is a collection (or multiset), as it may contain duplicates. That is, more than one individual from the same region may be observed at the same POI during the same time interval. A foot traffic database can be represented by three mode foot traffic tensor formally defined as follows:

Definition 3.3 (Spatiotemporal Foot Traffic Tensor). Let $\mathcal{D}$ be a foot traffic database, then $X \in \mathbb{N}^{|\mathcal{P}| \times |\mathcal{R}| \times |\mathcal{T}|}$ is a spatiotemporal foot traffic tensor, where

$$(X)_{i,j,k} = |\{o \in \mathcal{D} | o = (P_i, R_j, T_k)\}|$$

denotes the number of visits at POI P_i coming from region R_j during time interval T_k .

An example of a spatiotemporal foot traffic tensor is shown in Figure 1. Given a foot traffic database $\mathcal{D}$, we define the problem of spatiotemporal prediction of foot traffic as the problem of predicting the number of observations, for each POI and for each region, for the next time interval $T_{|\mathcal{T}|+1}$. We formally define:

Definition 3.4 (Spatiotemporal Prediction of Foot Traffic). Let $\mathcal{D}$ be a foot traffic database that captures observed visits at POIs $\mathcal{P}$ from regions $\mathcal{R}$ during time intervals $\mathcal{T} = \{T_1, \dots, T_{|\mathcal{T}|}\}$. The task of spatiotemporal prediction of foot traffic is to predict the number of foot traffic observations at times $T_{|\mathcal{T}|+1}$ for each POI $p \in \mathcal{P}$ and each region $r \in \mathcal{R}$.

In the following, we propose solutions to the problem of spatiotemporal prediction of foot traffic.

4 COLLABORATIVE FILTERING

While SafeGraph data is denser and more comprehensive compared to traditional LSBN data, we observed a different set of challenges when applied collaborative filtering using tensors. The first challenge is related to tensor sparsity. Furthermore, there are additional challenges associated with the heterogeneity of the data. We describe both challenges in detail here.

The first challenge of data sparsity is that, in practice, most cells of a spatiotemporal foot traffic tensor with a value of zero are a function of small POIs that are rarely frequented by faraway regions. For example, it is rare that people will travel hundreds of kilometers from their home region to visit a fast-food restaurant chain (e.g., “McDonald’s”). This leads to a very sparse spatiotemporal foot traffic tensor T . Yet, a simple solution of only considering local foot traffic, thus ignoring POI-region pairs beyond a certain distance will yield a loss in information. For example, for some POIs, such as a specific brand of POI that is not found everywhere (e.g., the Swedish furniture store “Ikea”), people may travel from far away regions. The second problem of heterogeneity is that the data varies substantially across each tensor mode, stemming from the inherent properties of foot traffic data. First, POIs may have very different visitor patterns. Some POIs may be small businesses with only dozens of visitors per day, whereas other POIs may be large stores with thousands of visitors per day. Second, different CBGs may have very different visiting patterns due to their location (as people tend to be likely to visit nearby POIs) and socioeconomic factors. For example, some areas may have a higher median per-capita income and thus, may be able to visit POIs that people from other regions do not visit. And third, visiting patterns may vary over time, as some POIs may vary seasonally (for example, POIs selling ice cream may have more foot traffic in summer), some POIs may close temporarily, and there may be global trends due to the closure of certain types of POIs due to pandemic prevention policies.

To capture the differences between POIs, spatial regions, and time intervals in spatiotemporal foot traffic data, we propose to use a tensor factorization approach to decompose the raw spatiotemporal foot traffic tensor into a product of factor matrices that capture the latent features of different POIs, spatial regions, and time intervals. Capturing similarities (and dissimilarities) between POIs, spatial regions, and time intervals, we can retain only the latent features that explain the most variance to create an approximated spatiotemporal foot traffic tensor. Using this tensor, we can estimate the expected number of visitors at each POI, from each spatial region, at each time interval, informed by visitor numbers at similar POIs, from similar regions, and at similar time intervals.

Such an approach of estimating values based on information about similar objects is also known as collaborative filtering, frequently used in recommendation systems. Here, we use Tucker tensor factorization [36] to factorize the spatiotemporal foot traffic tensor X into 1) a matrix that captures K_1 latent features of each POI, 2) a matrix that captures K_2 latent features of each spatial region, and 3) a matrix that captures K_3 latent features for each time interval, such that the product $\hat{X}$ of these matrices (and a core tensor that captures the singular values of X to capture the relative weight of each latent feature) approximates the raw data tensor X . This approach is formalized in the following.

Definition 4.1 (Tucker Tensor Decomposition). Let $X \in \mathbb{N}^{|\mathcal{P}| \times |\mathcal{R}| \times |\mathcal{T}|}$ be a spatiotemporal foot traffic tensor. Tucker tensor decomposition factorizes X having

$$X \sim G \times_1 U_1 \times_2 U_2 \times_3 U_3,$$

where $G \in \mathbb{R}^{K_1 \times K_2 \times K_3}$ is called the core tensor and $U_1 \in \mathbb{R}^{|\mathcal{P}| \times K_1}$, $U_2 \in \mathbb{R}^{|\mathcal{R}| \times K_2}$, $U_3 \in \mathbb{R}^{|\mathcal{T}| \times K_3}$ are factor matrices. The operator $\times_i$ denotes the mode- i tensor-matrix product, and K_1, K_2, K_3 are parameters that denote the degree of approximation of each factor.

Intuitively speaking, Tucker decomposition factorizes the spatiotemporal foot traffic tensor X into three matrices. The first matrix U_1 describes each of the $|\mathcal{P}|$ POIs with K_1 latent features; the second matrix U_2 describes each of the $|\mathcal{R}|$ spatial regions with K_2 latent features, and the third matrix U_3 describes each of the $|\mathcal{T}|$ times intervals with K_3 latent features. These factor matrices are similar to the factor matrices of (truncated) singular value decomposition (SVD) [14] and eigenvalue decomposition [9], with the main difference of having three-factor matrices instead of two. The core tensor G contains the 1-mode, 2-mode and 3-mode singular values of X , and can be seen as equivalent to the diagonal eigenvalue matrix in eigenvalue decomposition (but for three-mode tensors instead of two-mode matrices).

Unlike SVD and eigenvalue decomposition, where the number of latent features per mode is a single parameter, Tucker decomposition allows us to specify the number of latent features (or topics) for each mode. Thus, we may choose the parameter K_1 that specifies the number of latent features representing each POI to be larger than the number K_2 of latent features of spatial regions and K_3 of time intervals.

By expanding the factorized representation back into a tensor, we obtain an approximation of the data matrix $\hat{X}$ in which the K_1, K_2 , and K_3 most important (explaining most variance) latent features of POIs, spatial regions, and time intervals (respectively) are retained, but less important components are omitted.

Definition 4.2 (Approximated Spatiotemporal Foot Traffic Tensor). Let $X \sim G \times_1 U_1 \times_2 U_2 \times_3 U_3$ be the tensor factorization of spatiotemporal foot traffic tensor X . We define

$$\hat{X} = G \times_1 U_1 \times_2 U_2 \times_3 U_3$$

as the approximated spatiotemporal foot traffic tensor of X .

Intuitively, the resulting data tensor $\hat{X}$ will contain values $\hat{X}_{i,j,k}$ that approximate the number of visitors at POI p_i from spatial region r_j during time interval t_k by leveraging collaborative filtering with similar POIs, similar spatial regions, and similar time intervals while unexplained signal (noise) is removed. Our rationale is that this noise-reduced representation of the original spatiotemporal foot traffic tensor X will improve spatiotemporal foot traffic prediction.

Based on the approximated spatiotemporal foot traffic tensor $\hat{X}$ of X we explore solutions for predicting future foot traffic. In the following Section 5. By experimenting on the different periods and different subset of the data, we then empirically show that the approximated representation (which leverages collaborative filtering through tensor factorization) consistently improves predictions as shown in Section 6.

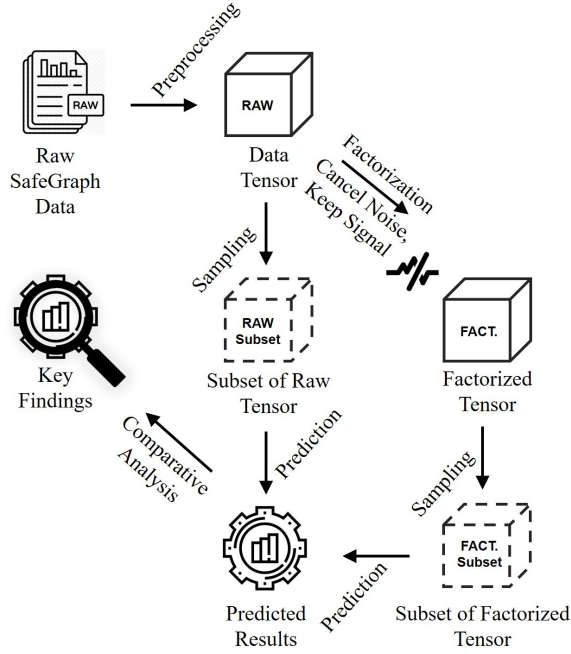


Figure 2: Spatiotemporal Foot Traffic Prediction Workflow

5 SPATIOTEMPORAL FOOT TRAFFIC PREDICTION

From the raw data that has been provided by SafeGraph [31] to the final prediction, a workflow consists of multiple steps is presented in Figure 2. First, the raw data obtained from SafeGraph is preprocessed into the raw spatiotemporal foot traffic tensor. This step groups all the raw data records by home CBG of the visiting user, the POI visited by the user, and the week of this visit. The resulting three-mode tensor having dimensions $\mathbb{N}|\mathcal{P}| \times |\mathcal{R}| \times |\mathcal{T}|$ is split into training and test data. A subset is sampled (and removed) from the raw data tensor and called “RAW Subset” in Figure 2. This sample is later used to evaluate the results. By default, we sample 10% of the raw data tensor for evaluation. The remaining data, up to week T , which includes all of the raw data tensors without the sampled evaluation data, is then factorized as described in Section 4 yielding a factorized tensor. This tensor is then used for predicting the spatiotemporal foot traffic at time $T + 1$ as described in the following. Then, the predicted values at time $T + 1$ (using only data up to the time T) is compared against the ground truth spatiotemporal foot traffic data. Evaluation metric and corresponding results are described in Section 6.

The following sections (Section 5.1-5.5) explain the prediction approaches to predict the number of visits at POI p_i from region r_j at time $T_{|\mathcal{T}|+1}$.

5.1 Previous Value

To predict the number of visits $X_{i,j,T+1}$ at POI p_i from region r_j at time $T_{|\mathcal{T}|+1}$, we first explore a naive baseline that simply predicts the number of visits at the previous time interval, that is:

$$X_{i,j,T+1}^{Prev} = \hat{X}_{i,j,T}$$

Intuitively, this value corresponds to the number of visits that the collaborative filtering would expect if next week is identical to the previous week.

5.2 Average Sliding Window

Instead of simply using the previous value to predict a future value, a sliding window approach looks at a window of the previous k observed time intervals and uses the mean of these observations as a prediction. Formally, to predict the number of visits at POI p_i from region r_j at time $T_{|\mathcal{T}|+1}$, this naive baseline simply predicts the number of visits at the previous time interval, that is:

$$X_{i,j,T+1}^{Slide} = \frac{1}{k} \sum_{\Delta=0}^{k-1} \hat{X}_{i,j,T-\Delta}$$

5.3 Weighted Sliding Window

Like subsection 5.2, here we use a sliding window approach but with different weights for each of the historical values. Foot traffic observations that are most recent will get the highest weight and k th oldest observations will get the least weight. We use a linearly decreasing weight to reduce the importance of older observations, formally:

$$X_{i,j,T+1}^{wSlide} = \sum_{\Delta=0}^{k-1} \frac{k-\Delta}{\frac{1}{2}k(k+1)} \hat{X}_{i,j,T-\Delta}$$

5.4 Temporal Univariate Linear Regression

We also employ a linear regression model, which fits a linear model to the previous k foot traffic observations of a given (POI, spatial region) pair. The intuition is that this approach is better able to capture a linear trend, such as increasing (or decreasing) foot traffic. Modeling the foot traffic $X_{i,j,T}$ as a function of time T , and using ordinary least squares fitting to minimize the squared model error, we obtain the slope β of a linear regression as:

$$\beta = \frac{\sum_{\Delta=0}^{k-1} (\hat{X}_{i,j,T-\Delta} - \overline{\hat{X}_{i,j,T-\Delta}})(T - \Delta - \bar{T})}{\sum_{\Delta=0}^{k-1} (\hat{X}_{i,j,T-\Delta} - \overline{\hat{X}_{i,j,T-\Delta}})^2},$$

where $\overline{\hat{X}_{i,j,T-\Delta}} = \frac{1}{k} \sum_{\Delta=0}^{k-1} \hat{X}_{i,j,T-\Delta}$ is the average foot traffic in the last k time intervals, and $\bar{T} = T - \frac{1}{2}k$ is the average time interval among the last k time intervals. We obtain the intercept of this linear regression as:

$$\alpha = \overline{\hat{X}_{i,j,T-\Delta}} - \beta \cdot \bar{T}.$$

This linear model allows us to predict the foot traffic at time $T + 1$ as:

$$X_{i,j,T+1}^{reg} = \alpha + \beta \cdot (T + 1)$$

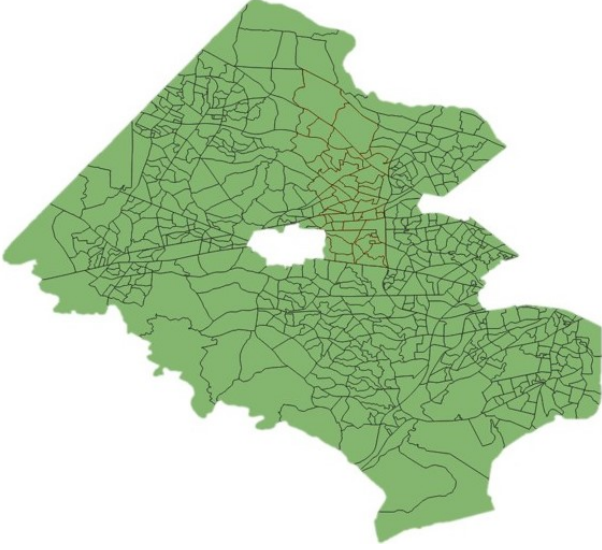


Figure 3: CBGs in Fairfax County

5.5 Long Short-Term Memory Recurrent Neural Network (LSTM)

Our problem resembles most of the classical time series forecasting problems; since Long Short-Term Memory (LSTM) has a track record of being successful in similar cases [6], [32], [3], we tried it too. The LSTM model is a type of RNN (Recurrent Neural Network) model, meaning, it reads each time step of an input sequence one at a time. The LSTM has an internal memory to keep, forget, or ignore previous data points based on a probabilistic model and to accumulate an internal state as it reads across the steps of the input data sequence. We tried to predict foot traffic from a spatial region to a POI and since our data is comprised of such observations (how many people visited a POI from a CBG), we have used a univariate LSTM model to do the prediction. We focus on keeping 75 percent of our data for training the model and the rest for testing.

6 EXPERIMENTAL EVALUATION: CASE STUDY OF FAIRFAX COUNTY, VA, USA

Our experimental discussion starts by describing our dataset which we obtained from SafeGraph [31] for our study region of Fairfax County, Virginia, USA, in Section 6.1. Following that, we summarize the prediction algorithms we use in Section 6.2 and describe our used evaluation metric in Section 6.3. Our experimental results for three subsets of the data are presented in Section 6.4.

6.1 Dataset Description

We use foot traffic data obtained from SafeGraph [31], which records the weekly foot traffic between points of interest (POIs) and census block groups (CBGs). This dataset is made available at no cost for academic institutions and has been used in the past to understand changes in mobility due to COVID-19 [8, 11] and to inform COVID-19 simulation models with fine-grained human mobility [13, 28]. Thus, in our experimental evaluation, the set of regions $\mathcal{R}$ corresponds to CBGs. We note that, this data does not report foot traffic from CBGs having fewer than two individuals in the data to protect

their privacy. Since SafeGraph data captures visits to millions of POIs and the United States has more than 200,000 CBGs, we focus our experiments on the study area of Fairfax County, Virginia, which has a population of 1.146 million people [37]. We obtain weekly foot traffic data that captures trips between 649 CBGs and 15,197 POIs in this region from Jan 2018 to Jun 2021, giving us 1.7 billion observations to be used in our analysis. Thus, for our experiments, we have a set of $|\mathcal{P}| = 15,197$ POIs, a set of $|\mathcal{R}| = 649$ CBGs, and a total of $|\mathcal{T}| = 182$ weeks. The resulting $15197 \times 649 \times 182$ tensor has 1.7 billion cells, most of which are zero. Figure 3 presents a map of Fairfax County and the CBGs that belong to it. Note that Fairfax County does not include Fairfax City, explaining the donut-like shape to the region. The large volume of data produces a relatively large data tensor (refer to Section 4 for more information about the tensor) which requires a significant amount of computational run time to investigate the wide range of possible predictive algorithms.

6.2 Spatiotemporal Foot Traffic Prediction Algorithms

We evaluated five algorithms (as described in Section 5) for spatiotemporal prediction of foot traffic one week into the future:

- Previous Value (Prev) (see Section 5.1)
- Four weeks average sliding window (AVG) (see Section 5.2)
- Four weeks weighted sliding window (WSW) (see Section 5.3)
- Linear regression (LReg) (see Section 5.4) and a polynomial regression of Degree 2 (PReg).
- Long Short-Term Memory (LSTM) (see Section 5.5)

6.3 Evaluation Metric

To evaluate our spatiotemporal foot traffic prediction models, we measure the one-week prediction accuracy in terms of Root Mean Squared Error (RMSE) over all POI-CBG pairs, thus we measure:

$$RMSE(X^M) = \sqrt{\frac{\sum_{i \in \mathcal{P}, j \in \mathcal{R}} (X_{i,j,T+1}^M - X_{i,j,T+1})^2}{|\mathcal{P}| \cdot |\mathcal{R}|}},$$

where $M \in \{Prev, Slide, wSlide, Reg, LSTM\}$ is the evaluated model (see Section 5 for details on each model), $X_{i,j,T+1}^M$ is the predicted foot traffic at time $T + 1$ for POI i at CBG j , and $X_{i,j,T+1}$ is the corresponding ground truth foot traffic.

6.4 Experimental Results

In the following, we apply the five prediction approaches (see section 6.2) to predict foot traffic in Fairfax County for each of the 178 weeks from Feb 2018 to Jun 2021. We exclude the first four weeks in Jan 2018 from prediction as these weeks do not have sufficient prior data to make prediction. We note that for each week $T + 1$ to be predicted, we use a separate factorized tensor that only includes data up to time T . This is to avoid overfitting by allowing the collaborative filtering to extract information of the future to be predicted. Due to the large number of POI-CBG-Week pairs to be predicted, we evaluate our spatiotemporal foot traffic prediction algorithms on three subsets of data, including:

- **Single CBG:** We select a single randomly selected CBG and 500 randomly selected POIs. The purpose of this experiment, described in Section 6.4.1, is to evaluate how well we can predict the spatiotemporal foot traffic of a single CBG;



Figure 4: Predicting Foot Traffic from a Randomly Selected CBG to 500 Randomly Selected POIs

- **Clustered CBGs:** We select a cluster of 50 CBGs and all 3052 POIs within these CBGs. Figure 5 shows the location of the CBGs and POIs considered in this subset. Using this subset, Section 6.4.2 shows how well we can predict the foot traffic in a whole area having approximately 100,000 people.
- **Most popular POIs:** This dataset includes all CBGs of Fairfax County (c.f. Figure 3) but considers only the most visited 1670 POIs. Section 6.4.3 shows the spatiotemporal foot traffic prediction results for this data set.

For each of these dataset subsets, we evaluate all five proposed algorithms (note, however that there are two variants of univariate regression) on both the raw spatiotemporal foot traffic data tensor X and the approximated tensor $\hat{X}$ after factorization. For the tensor factorization, we used $K_1 = 500$, $K_2 = 50$, and $K_3 = 20$ as the approximation degree (rank) of the factor matrices, which correspond to the number of latent features to describe POIs, CBGs, and time intervals, respectively.

6.4.1 Prediction Results for a Single CBG. Figure 4 shows the prediction results for the Single CBG foot traffic data subset. First, we observe that the regression models and the LSTM model yield a much larger error than the simple models. This is likely because these approaches overfit outlier observations. As the number of visitors from the selected CBG to a POI fluctuates between weeks, the assumption of a linear (or even second-degree polynomial) relationship between time and number of visits yields to potential over or underestimation. This is evident in the case of having outliers that the models may overfit. For example, if a POI had a sequence of (20,3,2,3) visitors from the selected CBG, then a linear relationship will predict a negative number of visitors next week (at time $T + 1$) by overfitting to the single outlier (and giving residuals a squared week using simple least-squares minimization).

We also observe that the patterns learned by the LSTM do not improve the prediction of foot traffic, which is again contributed to overfitting. While the LSTM considers all weeks at time $t > T + 1$, it may try to reproduce individual patterns observed in the past. For example, if a pattern of (3,2,3,20) was observed in the past, and in the three weeks before time $T + 1$ we observe a pattern of (3,2,3), then the LSTM is likely to predict a very high number of visitors at time $T + 1$. In summary, the problem is that we do not observe a sufficiently large number of weeks for each (POI, CBG) pair for the LSTM to converge into a stable model.

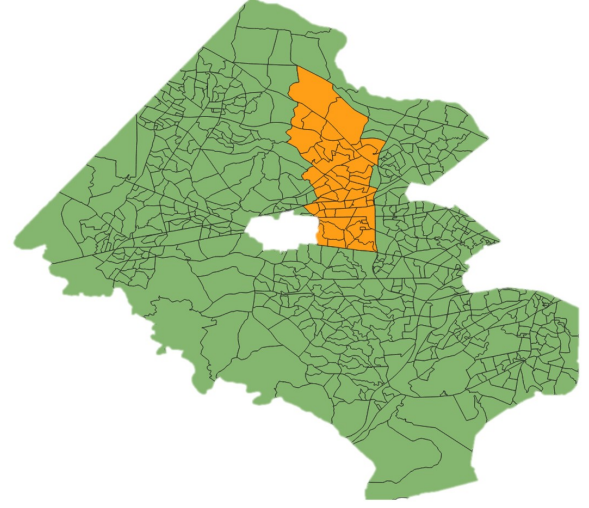


Figure 5: Cluster of CBGs in Fairfax County.

We further observe that the three naive models, which simply average previous observations, produce better results. Much of the variance that we observe in weekly foot traffic is a function of random noise rather than a trend. We observe that the approach that simply uses the previous observation has a much higher RMSE than the approach using the average of the previous four weekly observations. What is interesting is that the weighted approach, which gives more recent weeks a higher weight, does not improve prediction, which again indicates that, for most POIs, the weekly variance cannot be explained by trends.

What is interesting is that for all approaches, the use of the approximate matrix $\hat{X}$ derived from the factorized model yields better predictions. This shows that the collaborative filtering approach does indeed help prediction. We note that among all our experiments using this data sample, the approach of using a simple average of the last weekly observations on the factorized matrix $\hat{X}$ yields the lowest RMSE.

6.4.2 Prediction Results on Clustered CBGs. Figure 6 shows the prediction results for the clustered group of CBGs shown in Figure 5. In addition to comparing spatiotemporal foot traffic prediction results on the raw data tensor and the factorized model, we additionally show results 1) using a non-negative tensor factorization (which constrains the factorized matrix $\hat{X}$ to hold non-negative values) and 2) a simple improvement of the factorized model of Section 4 where negative elements of $\hat{X}$ are simply set to zero. For this set of experiments, we have omitted the results for the LSTM model which has consistently shown worse results than the other approaches, and we only show results of the linear regression model for the raw and factorized model due to the long run time of fitting $3025 \cdot 50$ linear models for each of the 178 weeks (we exclude the first four of 182 weeks from the evaluation since there is insufficient previous data).

Again, we observe that the approach using a simple average of the previous four weeks yields the highest prediction accuracy (the lowest RMSE). Interestingly, the non-negative factorization approach does not improve the prediction results. We explain this by the additional optimization constraint (of disallowing negative



Figure 6: Predicting Foot Traffic from the orange-colored CBGs of Figure 5 to 3052 POIs of that same location

values) which may yield a bias in the approximated spatiotemporal foot traffic data tensor. We also note that the simple solution of simply truncating negative values in the (original, not non-negative) matrix factorization yields a slight improvement in accuracy. This is expected, as truncating negative approximations to zero is guaranteed to decrease the RMSE (i.e., ground truth observations are guaranteed to be non-negative).

6.4.3 Prediction Results using popular POIs. Figure 7 shows our prediction results using all CBGs of Fairfax County and a subset of POI data that includes 1670 of the most visited POIs only. We note that this experiment covers the entire geographical location of Fairfax County and minimizes the number of zeros (flat lines) of (POI,CBG) pairs.

Again, we omit the results of the regression and LSTM solutions due to their high prediction error. For these highly visited POIs, we observe similar prediction errors for the three naive approaches of simply using the foot traffic of the previous week, using the average of the previous four weeks, and using the weighted average of the previous four weeks. We note that the approach that uses the approximated value estimated for the previous week using tensor factorization yields the lowest error. We also note that simply using the average of the previous week without the tensor factorization yields a substantially higher error. This error is reduced by averaging over previous weeks.

We further note that for this experiment, using 1670 POIs, 649 CBGs, and 178 weeks our models took hours to produce the resulting 193 million predictions. Since all models use an analytics solution to compute results, their run-times are similar, such that we did not evaluate. We note that we were not able to run the regression and LSTM algorithms due to run-time issues.

In summary, our experiments using three subsets of spatiotemporal foot traffic data obtained from SafeGraph shows that simple time series prediction approaches are able to best predict the number of visitors between (POI, CBG) pairs. We also show that the tensor factorization approach is able to reduce noise in the data to improve prediction results for all prediction algorithms consistently. However, we note that this study only applied straightforward solutions for spatiotemporal prediction and that more sophisticated solutions may promise better results.

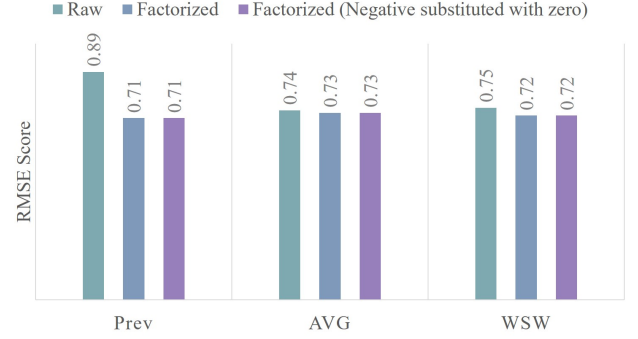


Figure 7: Predicting Foot Traffic from all the CBGs of figure 3 to the Most Visited 1670 POIs of the Fairfax County

7 DISCUSSION, CONCLUSIONS, AND FUTURE WORK

An efficient and stable framework to predict foot traffic has a plethora of applications in the fields of business, marketing, and advertisement. Due to these applications, a wide range of existing works has studied the problem of spatiotemporal prediction of check-ins using location-based social network data. However, prior research suffered from a lack of available data. Before the emergence of large sets of foot traffic data like those provided by SafeGraph, the main bottleneck of researching and developing such a system was the scarcity of data. With the current availability of data, this limitation is reduced.

The goal of this workshop paper is to organize such large-scale foot traffic data into a data representation that can be used to provide first benchmark results for spatiotemporal foot traffic prediction. Therefore, we propose simple time series prediction solutions to predict the foot traffic from a given region to a given point of interest at a future time interval $T + 1$, based on observations at previous time intervals. In addition to evaluating simple prediction approaches, we observed that foot traffic data is very noisy. To reduce the noise and find the signal in the data, we employed a tensor factorization approach. This approach reduces the data using latent features of POIs, spatial regions, and time intervals. Our experiments show that factorization consistently improves spatiotemporal foot traffic prediction results in all our evaluated cases. Still, we observe that the foot traffic for a specific POI from a specific spatial region is highly volatile, and our first attempts of modeling the trend in this data using univariate regression and using long short-term memory neural networks (LSTMs) do not offer improved predictions. A further challenge that we observed is run-time due to predicting each combination of point of interest, spatial region, and time interval individual using a separate model.

For our future work, we aim to explore two additional research directions. The first direction will focus on computing only a subset of spatiotemporal foot traffic predictions for selected (POI, CBG) pairs. This yields (for a given time $T+1$) a sparse matrix of prediction having most (non-sampled) pairs missing. We want to explore using a matrix factorization approach to estimate the missing (POI, CBG) pairs. This approach may allow us to obtain accurate foot traffic estimations for all (POI, CBG) pairs while only having to evaluate a small sample of individual models. The second direction

that we want to employ is to apply more advanced solutions for the spatiotemporal prediction that have been recently proposed for data-rich domains such as traffic prediction [39, 44], urban flow prediction [27] and bike-rental prediction [17] which extract spatial and temporal features from locations to predict future spatiotemporal observations. The main challenge working with foot traffic data is that the number of locations, which is usually in the tens to hundreds for these existing applications (traffic measurement locations, metro station, bike rental stations), may now exceed tens of thousands (points of interests). While this workshop paper only presents benchmark solutions, we hope that our work will spur interest in this new research direction of spatiotemporal prediction of foot traffic.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation as follows: 1) Grant No. 2109647 titled "Data-Driven Modeling to Improve Understanding of Human Behavior, Mobility, and Disease Spread" and 2) Grant No. 2030685 titled "An Ensemble Approach to Combine Predictions from COVID-19 Simulations". This work is also supported by George Mason University Office of the Provost and Executive Vice President's 2021 Summer Team Impact Grant.

REFERENCES

- [1] R. Assam and T. Seidl. Check-in location prediction using wavelets and conditional random fields. In *ICDM*, pages 713–718. IEEE, 2014.
- [2] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel. Recommendations in location-based social networks: a survey. *Geoinformatica*, 19(3):525–565, 2015.
- [3] J. Cao, Z. Li, and J. Li. Financial time series forecasting model based on ceemdan and lstm. *Physica A: Statistical Mechanics and its Applications*, 519:127–139, 2019.
- [4] D. Chai, L. Wang, and Q. Yang. Bike flow prediction with multi-graph convolutional networks. In *Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems*, pages 397–400, 2018.
- [5] C. Cheng, H. Yang, M. R. Lyu, and I. King. Where you like to go next: Successive point-of-interest recommendation. In *IJCAI*, volume 13, pages 2605–2611, 2013.
- [6] V. K. R. Chimmula and L. Zhang. Time series forecasting of covid-19 transmission in canada using lstm networks. *Chaos, Solitons & Fractals*, 135:109864, 2020.
- [7] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *KDD*, pages 1082–1090. ACM, 2011.
- [8] J. Elarde, J.-S. Kim, H. Kavak, A. Züfle, and T. Anderson. Change of human mobility during covid-19: A united states case study. *arXiv preprint arXiv:2109.09022*, 2021.
- [9] J. N. Franklin. *Matrix theory*. Courier Corporation, 2012.
- [10] Q. Gao, G. Trajcevski, F. Zhou, K. Zhang, T. Zhong, and F. Zhang. Deeptrip: Adversarially understanding human mobility for trip recommendation. In *ACM SIGSPATIAL*, pages 444–447, 2019.
- [11] S. Gao, J. Rao, Y. Kang, Y. Liang, and J. Kruse. Mapping county-level mobility pattern changes in the united states in response to covid-19. *SIGSpatial Special*, 12(1):16–26, 2020.
- [12] J.-S. Kim, H. Jin, H. Kavak, O. C. Rouly, A. Crooks, D. Pfoser, C. Wenk, and A. Züfle. Location-based social network data generation based on patterns of life. In *2020 21st IEEE International Conference on Mobile Data Management (MDM)*, pages 158–167. IEEE, 2020.
- [13] J.-S. Kim, H. Kavak, C. O. Rouly, H. Jin, A. Crooks, D. Pfoser, C. Wenk, and A. Züfle. Location-based social simulation for prescriptive analytics of disease spread. *SIGSPATIAL Special*, 12(1):53–61, 2020.
- [14] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, aug 2009.
- [15] H.-P. Kriegel, M. Renz, M. Schubert, and A. Zuefle. Statistical density prediction in traffic networks. In *Proceedings of the 2008 SIAM International Conference on Data Mining*, pages 692–703. SIAM, 2008.
- [16] M. Li, R. Westerholt, H. Fan, and A. Zipf. Assessing spatiotemporal predictability of lbsn: a case study of three foursquare datasets. *Geoinformatica*, pages 1–21, 2016.
- [17] Y. Li and Y. Zheng. Citywide bike usage prediction in a bike-sharing system. *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1079–1091, 2019.
- [18] Y. Li, Y. Zheng, H. Zhang, and L. Chen. Traffic prediction in a bike-sharing system. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–10, 2015.
- [19] D. Lian, V. W. Zheng, and X. Xie. Collaborative filtering meets next check-in location prediction. In *WWW*, pages 231–232. ACM, 2013.
- [20] B. Liu, Y. Fu, Z. Yao, and H. Xiong. Learning geographical preferences for point-of-interest recommendation. In *ACM SIGKDD*, pages 1043–1051, 2013.
- [21] Q. Liu, S. Wu, L. Wang, and T. Tan. Predicting the next location: A recurrent model with spatial and temporal contexts. In *AAAI*, pages 194–200, 2016.
- [22] X. Liu, Y. Liu, K. Aberer, and C. Miao. Personalized point-of-interest recommendation by mining users' preference transition. In *CIKM*, CIKM '13, pages 733–738, New York, NY, USA, 2013. ACM.
- [23] X. Liu, Y. Liu, K. Aberer, and C. Miao. Personalized point-of-interest recommendation by mining users' preference transition. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 733–738, 2013.
- [24] Y. Liu, T.-A. N. Pham, G. Cong, and Q. Yuan. An experimental evaluation of point-of-interest recommendation in location-based social networks. *Proc. VLDB Endowment*, 10(10):1010–1021, 2017.
- [25] C. J. Lynch and R. Gore. Short-range forecasting of covid-19 during early onset at county, health district, and state geographic levels using seven methods: Comparative forecasting study. *Journal of medical Internet research*, 23(3):e24925, 2021.
- [26] W. Min and L. Wynter. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, 19(4):606–616, 2011.
- [27] Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang. 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*, pages 1720–1730, 2019.
- [28] J. Pesavento, A. Chen, R. Yu, J.-S. Kim, H. Kavak, T. Anderson, and A. Züfle. Data-driven mobility models for covid-19 simulation. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Advances in Resilient and Intelligent Cities*, pages 29–38, 2020.
- [29] F. Petropoulos and S. Makridakis. Forecasting the novel coronavirus covid-19. *PLoS one*, 15(3):e0231236, 2020.
- [30] SafeGraph Inc. GUIDE TO POINTS OF INTEREST DATA: POI DATA FAQ (<https://www.safegraph.com/points-of-interest-poi-data-guide>).
- [31] SafeGraph Inc. Weekly Patterns data (<https://docs.safegraph.com/docs/weekly-patterns>).
- [32] A. Sagheer and M. Kotb. Time series forecasting of petroleum production using deep lstm recurrent networks. *Neurocomputing*, 323:203–213, 2019.
- [33] M. Sarwat, J. J. Levandoski, A. Eldawy, and M. F. Mokbel. Lars*: An efficient and scalable location-aware recommender system. *TKDE*, 26(6):1384–1399, 2014.
- [34] R. K. Singh, M. Rani, A. S. Bhagavathula, R. Sah, A. J. Rodriguez-Morales, H. Kalita, C. Nanda, S. Sharma, Y. D. Sharma, A. A. Rabaan, et al. Prediction of the covid-19 pandemic for the top 15 affected countries: advanced autoregressive integrated moving average (arima) model. *JMIR public health and surveillance*, 6(2):e19115, 2020.
- [35] Y. Su, X. Li, W. Tang, J. Xiang, and Y. He. Next check-in location prediction via footprints and friendship on location-based social networks. In *MDM*, pages 251–256. IEEE, 2018.
- [36] L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.
- [37] United States Census Bureau. United States Census. <https://www.census.gov/>, Accessed Sept 10, 2021.
- [38] H. Wang, M. Terrovitis, and N. Mamoulis. Location recommendation in location-based social networks using user check-in data. In *ACM SIGSPATIAL*, pages 374–383, 2013.
- [39] J. Xu, D. Deng, U. Demiryurek, C. Shahabi, and M. Van der Schaar. Mining the situation: Spatiotemporal traffic prediction with big data. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):702–715, 2015.
- [40] G. Yang and A. Züfle. Spatio-temporal site recommendation. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 1173–1178. IEEE, 2016.
- [41] Z. Yang, J. Hu, Y. Shu, P. Cheng, J. Chen, and T. Moscibroda. Mobility modeling and prediction in bike-sharing systems. In *Proceedings of the 14th annual international conference on mobile systems, applications, and services*, pages 165–178, 2016.
- [42] J. Ye, Z. Zhu, and H. Cheng. What's your next move: User activity prediction in location-based social networks. In *SIAM Data Mining (SDM)*, pages 171–179. SIAM, 2013.
- [43] M. Ye, P. Yin, and W.-C. Lee. Location recommendation for location-based social networks. In *ACM SIGSPATIAL*, pages 458–461, 2010.
- [44] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma. Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks. *Sensors*, 17(7):1501, 2017.
- [45] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li. T-gcn: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9):3848–3858, 2019.