July 2018

A Study on Modelling Spatial-Temporal Human Mobility Patterns for Improving Personalized Weather Warning

YUE XU
University of Massachusetts Amherst

A STUDY ON MODELLING SPATIAL-TEMPORAL HUMAN MOBILITY PATTERNS FOR IMPROVING PERSONALIZED WEATHER WARNING

A Thesis Presented

by

YUE XU

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

May 2018

Department of Geosciences

A STUDY ON MODELLING SPATIAL-TEMPORAL HUMAN MOBILITY PATTERNS FOR IMPROVING PERSONALIZED WEATHER WARNING

A Thesis Presented

by

YUE XU

Approved as to style and content by:

____
Qian Yu, Chair

____
Brenda Philips, Member

____
Piper R. Gaubatz, Associate Department Head
Department of Geosciences

DEDICATION

For dear Naipo, in memoriam.

ACKNOWLEDGMENTS

This thesis is based, in part. upon work supported by the National Science Foundation under Grant AGS-1331572. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Thanks Qian Yu and Brenda Philips for supervising and mentoring.

Thanks to the whole UMASS Geosciences Department for the nice research environment.

Thanks to all my friends.

Thanks to my family’s support.

ABSTRACT

A STUDY ON MODELLING SPATIAL-TEMPORAL HUMAN MOBILITY PATTERNS FOR IMPROVING PERSONALIZED WEATHER WARNING
MAY 2018

YUE XU, B.S., NANJING NORMAL UNIVERSITY

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Qian Yu

Understanding human mobility patterns is important for severe weather warning since these patterns can help identify where people are in time and in space when flash floods, tornados, high winds and hurricanes are occurring or are predicted to occur. A GIS (Geographic Information Science) data model was proposed to describe the spatial-temporal human activity. Based on this model, a metric was designed to represent the spatial-temporal activity intensity of human mobility, and an index was generated to quantitatively describe the change in human activities. By analyzing high-resolution human mobility data, the paper verified that human daily mobility patterns could be clearly described with the proposed methods. This research was part of a National Science Foundation grant on next generation severe weather warning systems. Data was collected from a specialized mobile app for severe weather warning, called CASA Alerts, which is being used to analyze different aspects of human behavior in response to severe weather warnings. The data set for this research uses GPS location data from more than 300 APP users during a 14 month period (location was reported at 2 minutes interval, or at based on a 100m change in location). A targeted weather warning strategy was proposed as a result of this research, and future research questions were discussed.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Human mobility patterns are derived by collecting data on where people go, and how long they stay there by time of day as they go about their daily activities. Previous research demonstrates the existence and predictability of these patterns. A study of historical human trajectories conducted by González et al. shows a high degree of temporal and spatial regularity (González et al. 2008), which makes it possible to find the mobility pattern based on people's previous movement data. The study of Song et al (2010) answered the fundamental question "is human behavior predictable?" By measuring the entropy of individual's historical movement, they found 93% of the mobility could be predicted when time is incorporated. Besides, Song et al. (2010) found predicting human activity is largely independent of the size of the spatial area where people move around.

Understanding human mobility patterns is important for many fields, such as urban traffic planning (Guo et al. 2012), hazard management (Lindell and Hwang 2008), as well as crime prediction (Andresen and Malleson 2013a, Mohler et al. 2011). Studying human mobility for hazard management has become more significant these years because severe weather events occur more frequently due to global climate change and have caused losses of lives and property (Meehl et al. 2000). Sending an alert or evacuation guidance to people before or during these severe weather events can effectively reduce or avoid the losses. To better understand the human behavior related weather alert, we can divide the human activity into three stages. As Figure 1 shows, in pre-alert stage, people are engaged in their routine activities. After they receive the alert (in-alert stage), they begin to react to the alert and the weather event. After the alert stops (post-alert), they will resume routine activities or begin rescue.

Understanding people's mobility patterns in all three stages both spatially and temporally, can help optimize alert warning. By studying the pre-alert stage, we can send targeted and timely warning messages to a specified individual to give him/her a more clear suggestion or guidance on where and how to take protective action. Stakeholders can then develop hazard response strategies based on people's behavior in the in-alert and post-alert stages. For example, Becker et al. (2015) conducted research on people's behavior to floodwater. They analyzed why some people still entered floodwater area after the flood happened, and proposed suggestions on providing future public education. Similar works were also conducted by Basher (2006) which resulted in a new concept of a people-centered global early warning system that accounts for people's behavior.

Figure 1. Three Stages of Human Behavior Related to Weather Alert

Usually, severe weather warning systems send the same message to all people within a predicted hazard area. It does not consider an individual's or a specified group of individuals' situations (He et al. 2009). These general warning messages might cause confusion and uncertainty. Those people who were not really affected by hazard might become desensitized to warnings, while those affected people might think the message is too general to follow (Mileti and Peek 2000). To improve the accuracy of the warning messages, some hazard warning systems consider people's response to the alert message to make alerts more effective, but human behavior is one of the most complex aspects in these systems (Mileti and Sorensen 1990). As shown in Figure 2, public responses to warnings are monitored in the Response Subsystem and sent to the Management Subsystem, then can be used for adjusting warnings. However, the meaning of 'response' considered here is that how people interpret the broadcasted warning messages based on their perception. Thus, a better system which can actively monitor people's behavior is needed and human mobility patterns could be integrated into warning systems to generate personalized warning strategies. Specifically, to improve the warning system, the human mobility pattern in pre-alert stage should be considered in the Management Subsystem before generating warning messages, while the pattern in the in-alert and after-alert stages should be considered in the response subsystem.

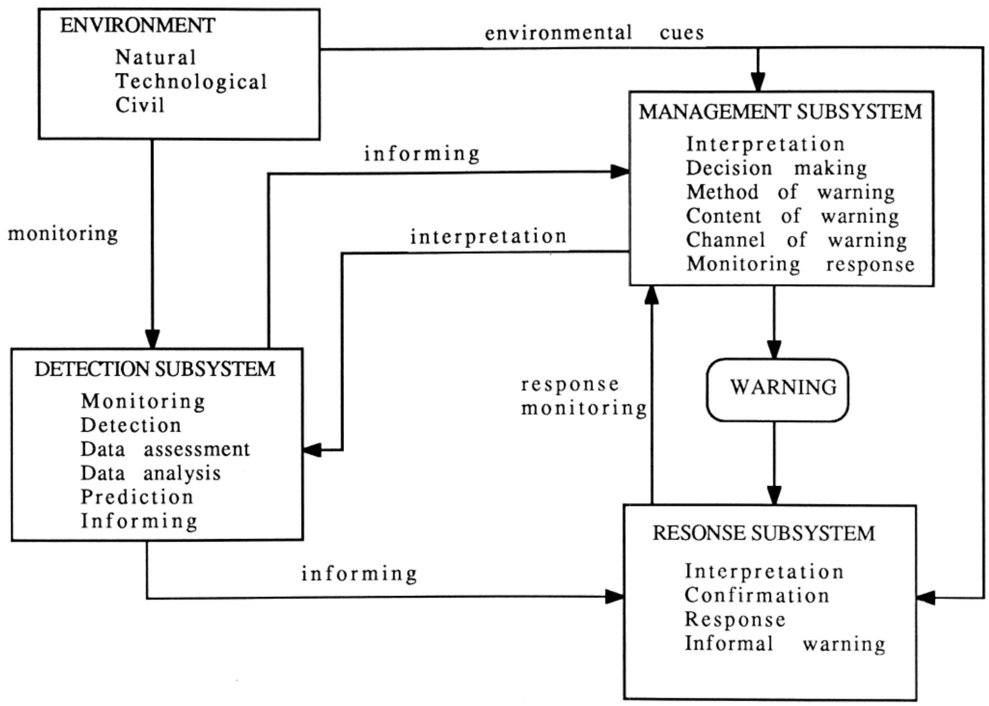


Figure 2. The General Components of An Integrated Warning System(Mileti and Sorensen 1990)

Four key challenges need to be considered for characterizing human mobility patterns. The first one is the data that represents human activities. Lack of high-quality movement tracking limits the spatial and temporal resolution of the derived human mobility pattern. For example, some previous researchers used the locations from twitter messages or mobile phone base stations as a proxy variable when the movement data is not available (Gao 2015, Wang et al. 2016). Analyses of Andresen and Malleson are based on street level crime data and monthly or seasonal level time interval analysis (Andresen and Malleson 2013a). 91% of American grown-ups possess a mobile phone, and numerous underserved populaces depend on mobile phones as their essential wellspring of data. (Bean *et al.* 2015). This penetration of cellphones makes it possible to study human behavior based on high resolution positioning data from mobile phones, if the big volume of the data can be processed efficiently.

The second challenge is the geographic boundary used as a unit to hold activities in order to conduct further analysis. Many researchers use clusters to represent activities within a certain spatial and time range. A popular way is to the use of changing geographic boundary of activity clusters. A cluster is generated by the activities of an individual sharing similar characters like time and distance. So the size and location of one cluster is unfixed and differs by each individual, which will cause overlap of locations and reduce the efficiency for further analysis. For example, someone used GPS track data of individual human activities to generate the cluster directly (Song et al. 2006, Etter et al. 2013). Thus, it will be very hard to model a group of people (citizens of a city) in a large area (the whole city) with clusters because generating every individual's activity clusters is low efficiency. It is also difficult to represent behaviors of grouped individuals with clusters.

The third challenge concerns describing the intensity of human activity within certain time and detecting human activity change. We still need to find a better way to detect human mobility changes, which is a crucial step on characterizing the pattern. Some previous research detected human activity changes by comparing the average duration difference of the activity points within the generated cluster (Huang et al. 2015). However, duration can only represent the time dimension of human activities. The spatial intensity of human activity is not considered. So, a new comparison method considering both spatial and temporal intensity is needed.

The last challenge is to quantitatively describe the difference among human activities. Generally, we use similar, increase and decrease to describe the variations of human activities qualitatively. Andresen (2009, 2016) provides an approach using Monte-Carlo resample method to generate a similarity index S which can quantitatively describe the similarity between human activities in two time stages. This similarity index is not enough to meet our purpose because it lacks the "increase" and "decrease" phases.

This paper proposes a method to quantitatively describe the spatial-temporal human mobility pattern, to improve the personalized weather warning. In chapter 2, we design a GIS (Geographic Information Science) model to describe the spatial-temporal movement pattern. A metric for describing the activity intensity was proposed based on this model. In Chapter 3, we define mobility pattern changes. Spatial intensity change and duration change were used to indicate the mobility pattern changes. An index was designed to indicate the magnitude of spatial-temporal pattern change. In chapter 4, based on the data collected by the 'CASA Alert mobile app, we calculated the index in different time intervals, and analyzed the relationship between the index and the daily mobility behaviors. By overlaying daily mobility behaviors with the distribution of housing units from demographic data from US Census Bureau, we conclude that the index can clearly indicate the daily pattern change. A warning process is proposed based on the proposed methods.

CHAPTER 2

DATA MODEL DESIGN

2.1 Data Description

This research uses high-resolution human mobility data collected by the "CASA Alert" platform developed by CASA (the Engineering Research Center for Collaborative Adaptive Sensing of the Atmosphere) at University of Massachusetts Amherst. As shown in Figure 3, multiple radars have been installed in the area around Dallas and Fort Worth, Texas, to monitor severe weather events. Weather reports and alerts are generated based on the radar data, and then sent to users through an APP called CASA ALERTS installed on their mobile devices. The APP also collects the users' location every 2 minutes in time interval and every 100 meters in spatial interval while using the app (250 m if the app is turned off). With the data collected within 14 months (from 30 June 2016 to 29 August 2017) from more than 300 users, we were able to analyze 5,529,335 high-quality GPS data points.

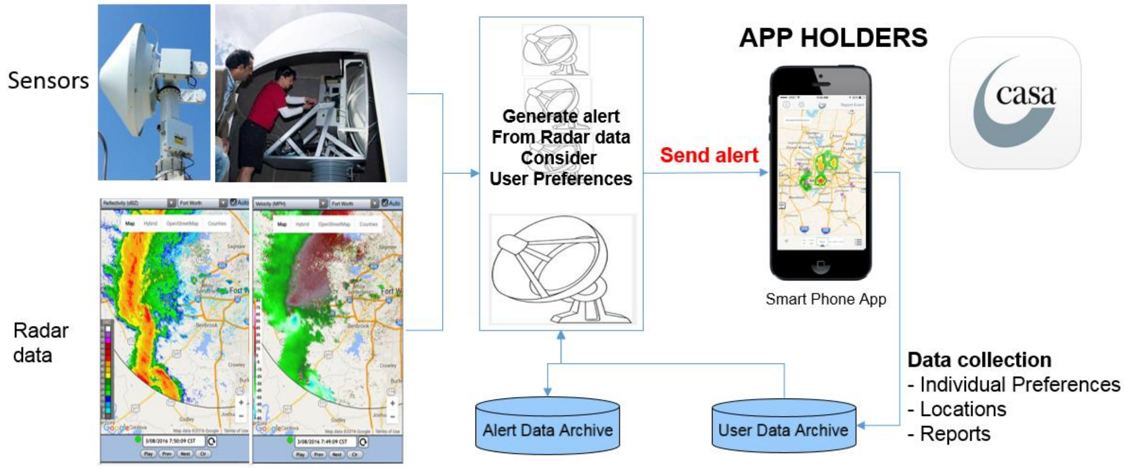


Figure 3. CASA Weather Alert System in Texas Developed by UMass CASA (Collaborative Adaptive Sensing of the Atmosphere) Engineering Research Center

2.2 Duration-Based Data Model Represent Human Activity

To model the human activity, we need first define its core components. Human activities are three dimensional because they are located in certain area (space, usually represented by longitude and latitude) and certain time interval (time) with a certain frequency (intensity). Figure 4 shows the trajectories of one person within two weeks, and the calculated occurrence density (activity intensity). We can see that this person mainly moves back and forth between two high activity intensity locations. This proves that time, space and intensity are the components needed to be considered in representing human activity.

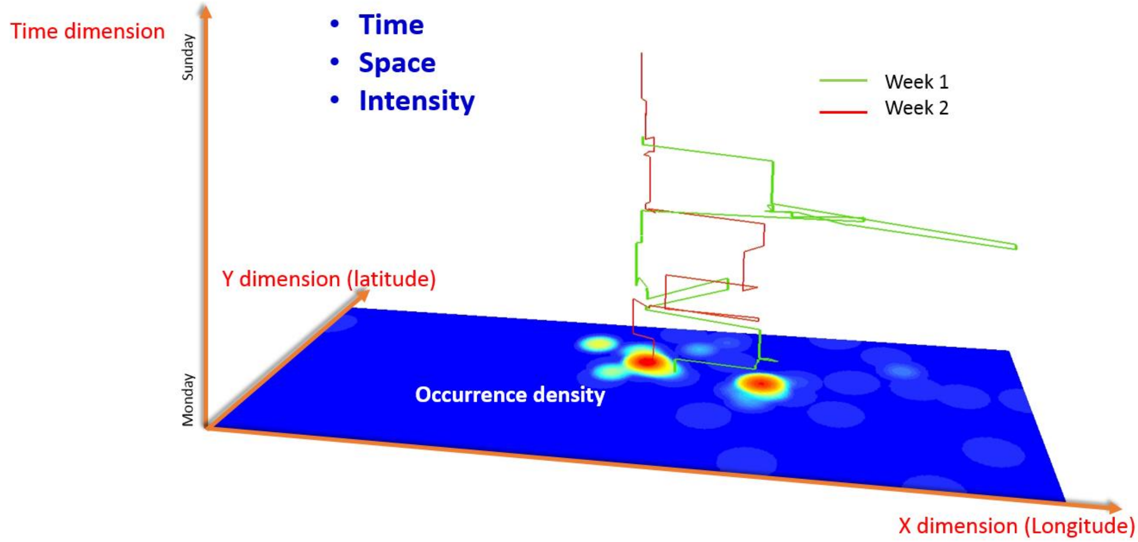


Figure 4. Three key Factor of the Human Activity

We propose a duration-based data model to represent the human activity based on above analysis. Figure 5 is the conceptual model. Human activity is composed of a series of single occurrences. A single occurrence has 'spatial location' and 'temporal location'. The Spatial Location can be defined by a pair of longitude and latitude. The 'temporal Location' is defined by the start-time and end-time, which we call a duration. Another attribute of human activity is the spatial-temporal magnitude, which can be described with spatial intensity and temporal intensity.

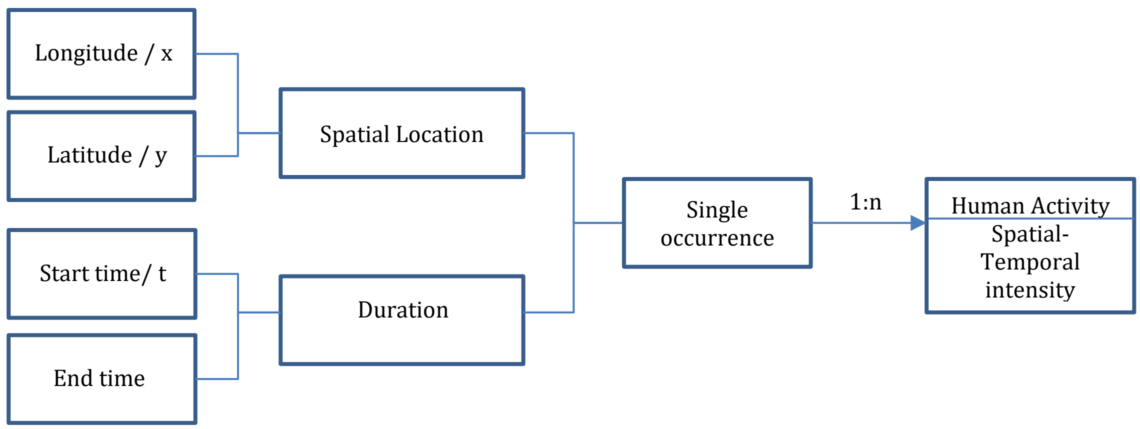


Figure 5. Conceptual Model Representing Human Activity

The raw data collected by the APP is structured as Table 1. The locations (longitude and latitude) of the user were collected every 2 minutes. If the user did not move during the data collection time interval, the two adjacent location points show in database will be the same. This caused redundancy of the computer memory and reduced the efficiency of calculation.

Table 1 Raw Data Structure from the CASA ALERTS

DataID	PersonID	latitude	longitude	Timestamp

We design the logical model based on the conceptual data model we proposed, as in Table 2.

Table 2 Proposed Logical Data Model

DataID	PersonID	latitude	longitude	Start-time	End-time	duration

We transformed the ‘timestamp’ in the raw structure into ‘duration’ with ‘start-time’ and ‘end-time’, which indicates the start and end time of the user’s single activity occurrence. When the person stops at the same point for some time length, we will know the ‘duration’ of his stop instead of the duplicated locations every 2 minutes. This will save the memory and improve the calculation speed.

We give the following definition of single activity occurrence:

Definition 1: A **single activity occurrence** is represented by a record in the **Logical Model**. The record is extracted from raw GPS Data dataset and transformed to the **new structure**.

2.3 Quantitatively Describing Activity Intensity

2.3.1 Define Temporal Activity Intensity by Considering Duration

We can use duration to describe the temporal activity intensity based on the designed conceptual model. As mentioned in section 2.2, the “CASA Alert” platform detects location change of a mobile device with 100 meters spatial resolution and update every 2 minutes. We can generate the duration of every record. If a user moves continuously in certain time period, we will get many short durations, which is considered high activity intensity during this time period. Figure 6 schematically shows the relationship between duration and time of single activities.

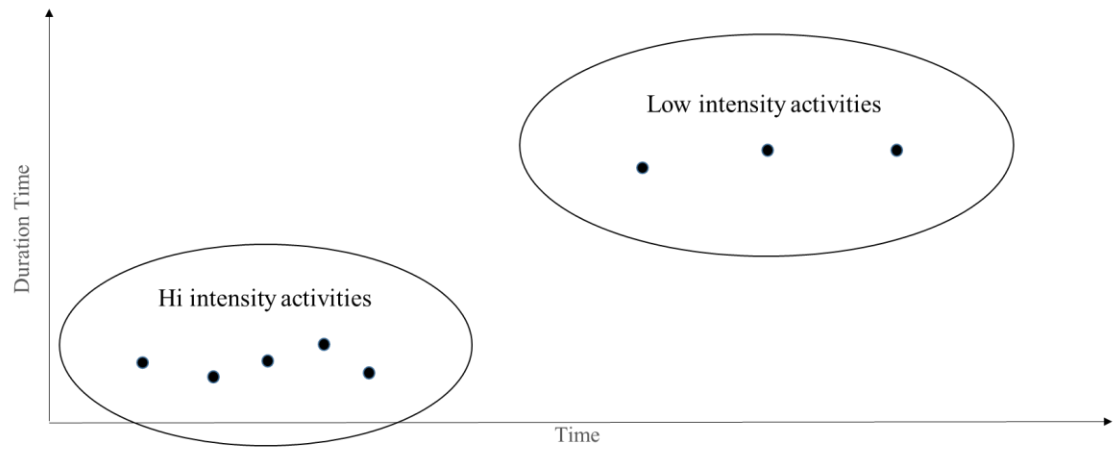
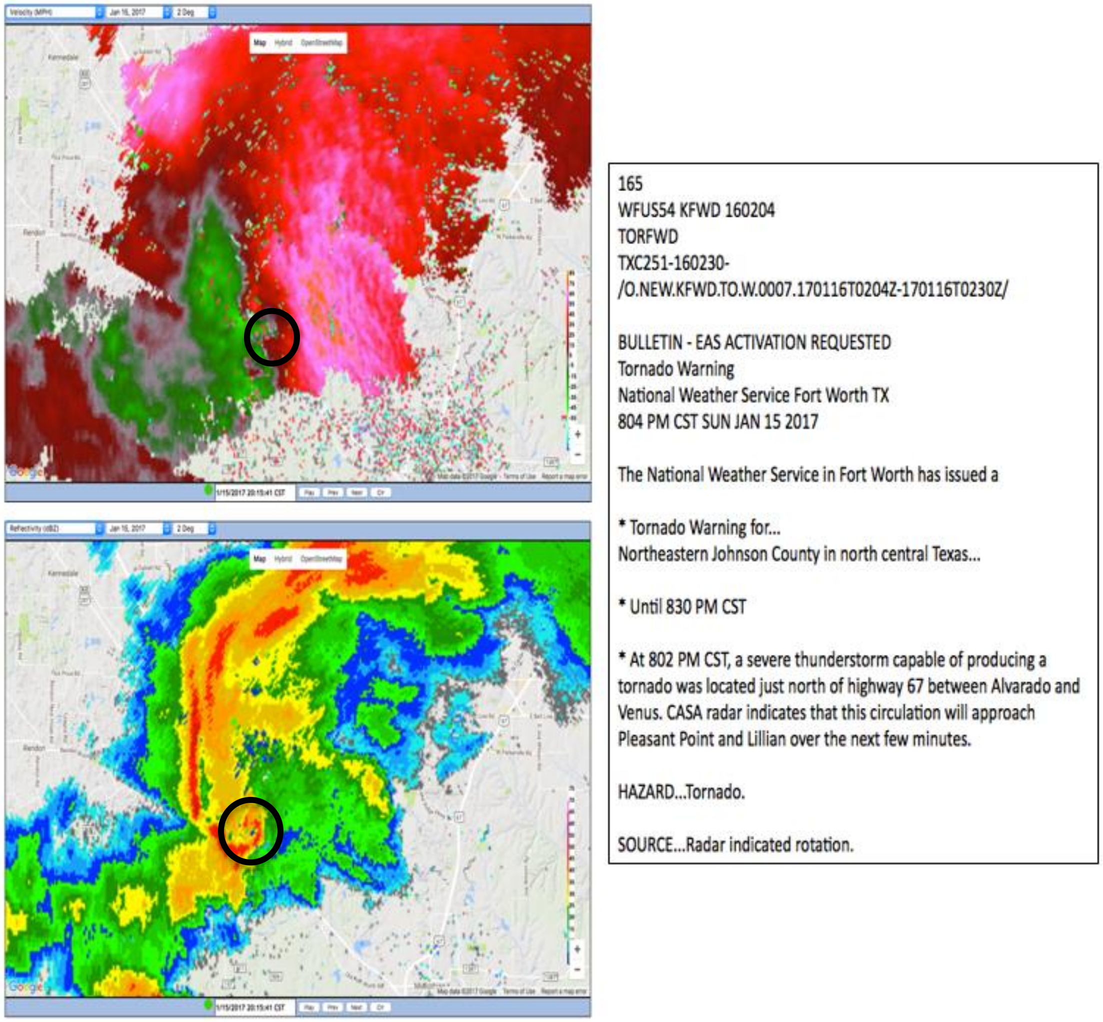


Figure 6. Relationship Between Duration and Start Time of Single Activities

We use a tornado event as a case study to further explain the proposed model. On January 15, 2017, a tornado (EF0) was detected by the CASA radar (Ridgeline Instruments, RXM 25) near Mansfield, TX. Several alerts were sent during 19:30 to 21:00 (as Figure 7 shows). There were 24 users of 'CASA Alert' around the area and received the alert message.



165
WFUS54 KFWD 160204
TORFWD
TXC251-160230-
/O.NEW.KFWD.TO.W.0007.170116T0204Z-170116T0230Z/

BULLETIN - EAS ACTIVATION REQUESTED
Tornado Warning
National Weather Service Fort Worth TX
804 PM CST SUN JAN 15 2017

The National Weather Service in Fort Worth has issued a

* Tornado Warning for...
Northeastern Johnson County in north central Texas...

* Until 830 PM CST

* At 802 PM CST, a severe thunderstorm capable of producing a tornado was located just north of highway 67 between Alvarado and Venus. CASA radar indicates that this circulation will approach Pleasant Point and Lillian over the next few minutes.

HAZARD...Tornado.

SOURCE...Radar indicated rotation.

Figure 7. Tornado Position (in black circle) and the Alert Message (Jan. 15, 2017)

With trajectories of these users, 130 durations of activities were generated. Figure 8 shows the activity start time versus duration length. We can see that, comparing to pre-alert period in which activities has long duration, more short duration activities occurred at post-alert period.

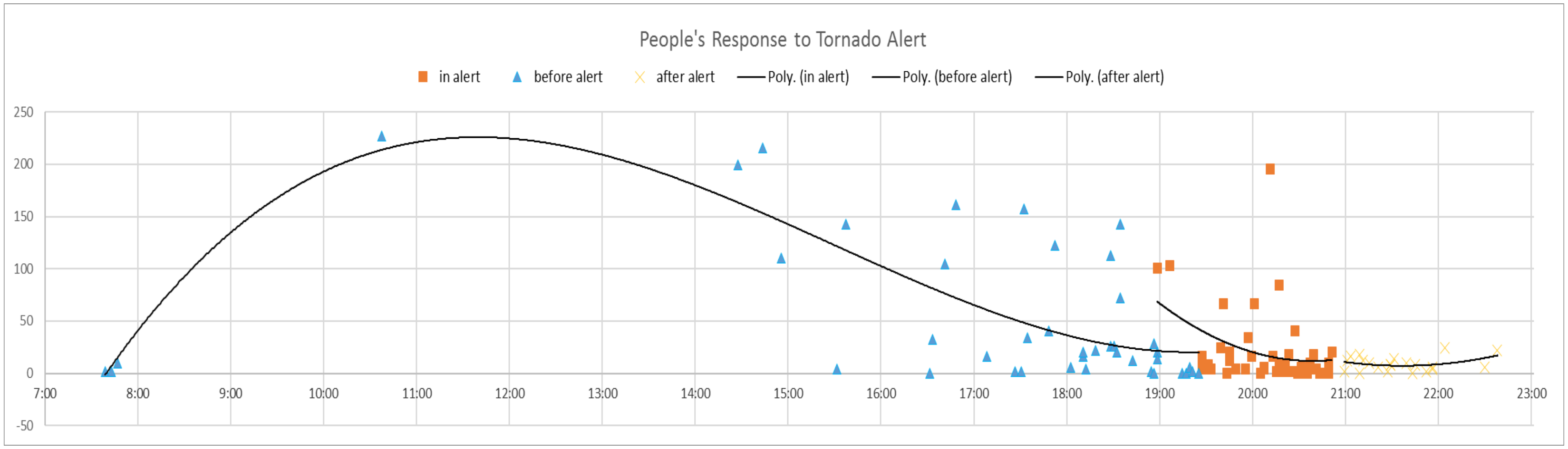


Figure 8. People's Response to Tornado Alert

Table 3 Movement Summary by Alert Stage

Alert Stage	Time	Activity frequency	Average stay duration
Pre-alert	7:30-19:30	0.08 move/min	50.04 min
In-alert	19:30-21:00	0.54 move/min	20.19 min
After-alert	21:00-23:00	0.22 move/min	9.74 min

Table 3 shows the activity frequencies and average stay duration along time. We can see from Figure 8 and Table 3 that before the alert, people were doing daily routine activities, so we got lowest move frequency and highest average duration. During in-alert period people move with the highest frequency but the average duration is lower, which means that those who were not in a safe place moved quickly to find a shelter and then stayed until the alert ended. At the time the alert starts, long durations mean people who had already been in safe place would not move until the warning ends. When the alert ended, the average duration decreased, and the move frequency is higher than pre-alert period. This indicated that the alert actually paused people's routine activities. Activities resumed in post-alert period.

2.3.2 Kernel Density Estimation for Point Mobility Data

Kernel density estimation is a method to quantitatively describe activity intensity. It can characterize spatial distribution of events and determine where events are more likely to occur in space (Thakali *et al.* 2015). The Kernel density estimator for **multivariate datasets** with **Kernel Function *K*** and **bandwidth h** is defined by:

$$\hat{f}(s) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{s-s_i}{h}\right) \quad (1)$$

It is a multivariate function.

The kernel function *K*(s) is a function, defined for **d-dimensional** s, satisfying:

$$\int_{R^d} K(s)ds = 1 \quad (2)$$

K is usually a radially symmetric uni-modal probability density function. There are many choices of kernel function. For example, ArcGIS uses the Epanechnikov kernel function described by Silverman (Silverman 1986). In my study, refer to the experiences in the Crime Process Modeling (Mohler *et al.* 2011), we use Gaussian kernel as K function

For spatial point p (x, y) we can use two dimensional kernel density estimator to calculate the spatial density. Here x and y represent the location of the point, usually are longitude and latitude respectively. In this case d=2, s=p (x, y). If we consider the time duration of each point, we can add duration as the weight value base on two-dimensional kernel density.

With above described method, we calculated a 2-dimensional kernel density based on data collected in the study area, as shown in Figure 9. The kernel density integrates spatial information with the duration, so we can visualize the spatial distribution and hotspot pattern. The density value can quantify the spatial intensity pattern of the points. Higher values in the northeastern and southwestern area of Figure 9 showed more activities happened there.

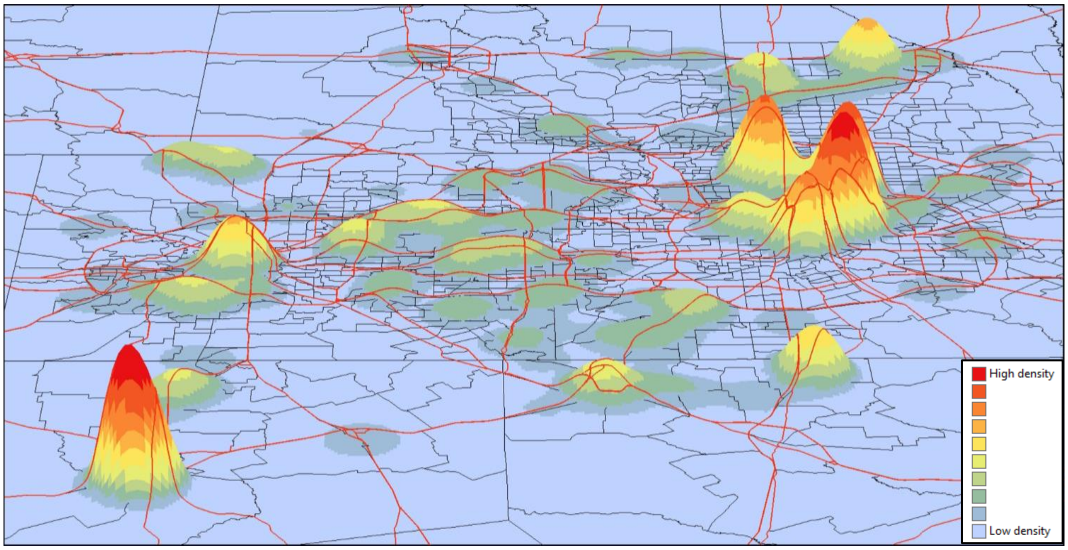


Figure 9. Visualization Example of Two Dimensional Kernel Density, K(x,y)

Alternatively, for spatial-temporal point s=p (t, x, y), we can use three-dimensional Kernel density to demonstrate the spatial-temporal pattern, where t represents the specific start time of an activity. In this case d is equal to 3. Figure 10 shows a visualization example of three-dimensional Kernel density estimation calculated from human activity points. We can see that people's spatial-temporal activity pattern represented by kernel density value changes as the time goes up along the vertical axes.

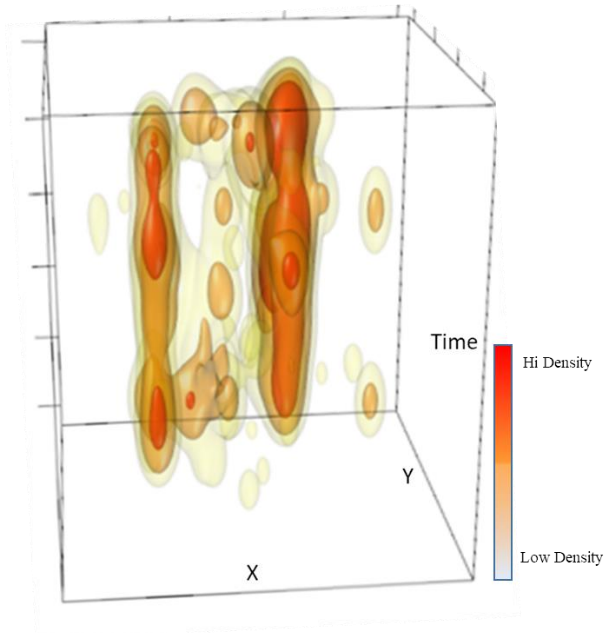


Figure 10. Visualization of Three Dimensional Kernel Density, K(x,y,t)

2.3.3 Describe Activeness of Human Spatial-Time Activities

We can use the points in the study area to generate an area-based kernel density. It is similar to the way of making the crime hotspot maps to visualize spatial-temporal crime patterns for near-repeat crime prediction (Mohler *et al.* 2011). The crime hotspot only flagged the highest density value area, while our proposed method can detect overall mobility pattern. Here we assign our estimation value to each area (polygon) we want to use. The estimation value in this thesis is called the Activeness value.

During time interval $[t_0, t_1]$, the input dataset is:

$$O = \{single\ activity\ occurrence\ (x_k, y_k, D_k);\ t_0\ \leq\ t_k < t_1\} \qquad (3)$$

Where, k is the total number of single activity occurrence, D_k is the duration. Each activity occurrence is an individual's behavior; more activity occurrences with short duration mean high activity intensity during this time interval.

For $\boldsymbol{O}$ in time interval [t_0, t_1],the Activeness is calculated according to:

$$f'(O(x,y,D)) = \sum_{1:k} K'(x_k, y_k; W_k) \quad (4)$$

Then, we can overlay activeness result to a polygon-based map by assigning average Activeness value on grid to each polygon.

Where in (4), K' is a two-dimensional (d=2) Kernel.

$$W_k = D_{max} - D_k, \quad (5)$$

is the weight function giving short duration activities more weight than long ones, $D_{max.}$ is the maximum duration value in $\boldsymbol{O}$. This activeness density will work better than regular kernel density when applying to describing activity intensity, as shown in Figure 11. The left figure is the regular kernel density generated with raw GPS data points of one person with (1). It shows the spatial mobility pattern of this person. The right figure is the activeness density calculated with (4). We can find mainly two 'nodes' in western and eastern part of the map with higher kernel density value, which means people stay longer and/or frequent around these two places. One node might be the location of residence and the other might be working place. Whereas, in the activeness figure, the connection between the two 'nodes' has higher activity intensity. This means the commuting pathway where this person travels between the two nodes is also detected. We generated a polygon-based kernel density map shown in Figure 12 by assigning the average density to each polygon. Here we used the census tracks as polygons. The scale of polygons can be adjusted according to the scale of analysis. For example. we can use city boundary or state boundary to make city level or state level analysis.

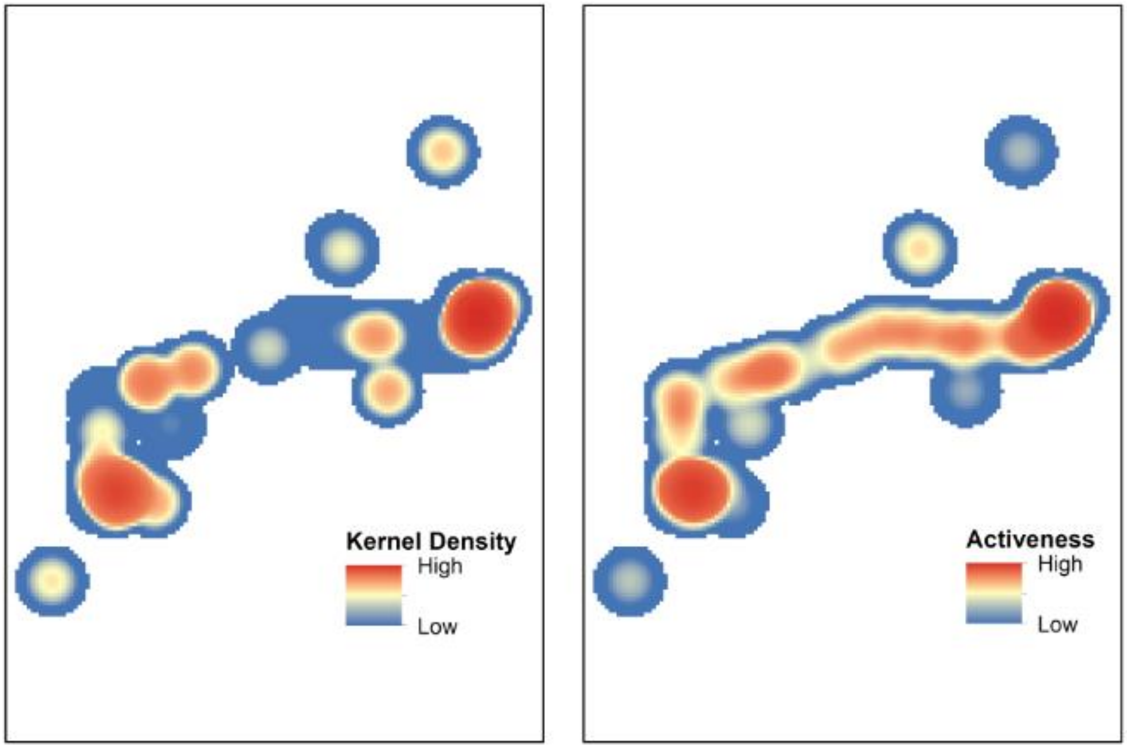


Figure 11. Visualization of Kernel Density and Activeness Estimation Result

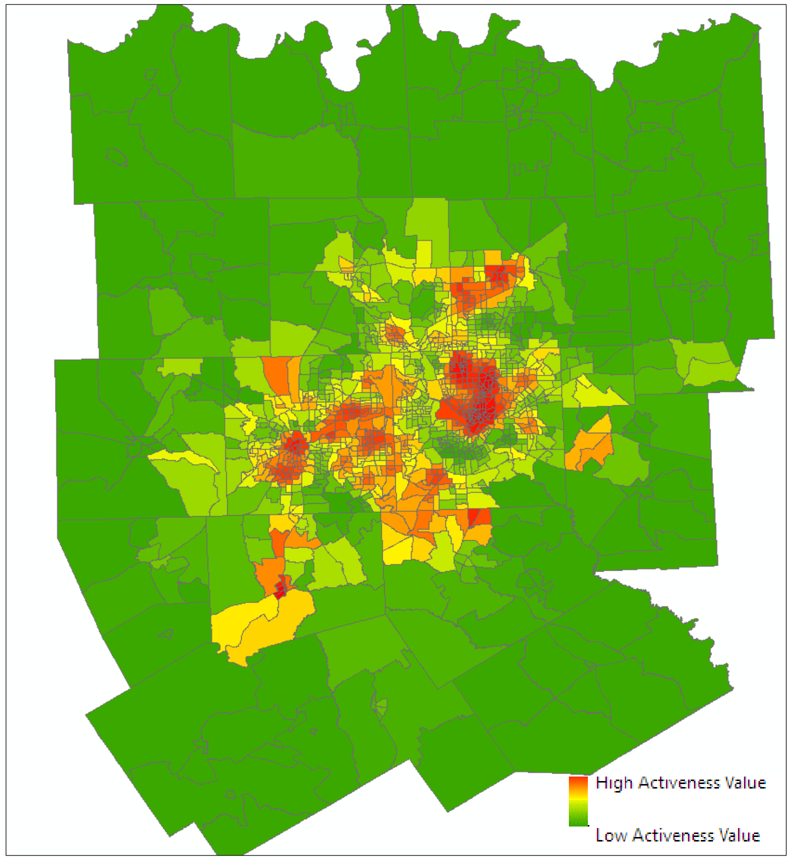


Figure 12. Visualization of Density Value by Census Tract

CHAPTER 3

PATTERN CHANGE DETECTION AND TEST

3.1 Define Mobility Change Based on Activeness

As we have known from section 2.3, higher activeness value is caused by more occurrences and shorter stay durations, which means more activities at that time interval in the area. Thus we can detect the spatial pattern change in each polygon by comparing the change of activeness value. If the activeness value of polygon P_1 at time t_1 is higher than that of same polygon at time t_2, we can say that there is more activity happened in t_1.

Definition 2: Within a specific area, changing of the activeness from one time interval to another time interval means the spatial intensity or temporal duration of human activities change.

Spatial intensity change or duration change can cause the change of activeness value. We explain the relationship between the changes of human activity pattern and the activeness value with Figure 13. At time 1, two people in area 1 are running to area 2, leaving eight short duration activity occurrences in area 1. At time 2, these same two people are moving slowly in area 2. One is moving around inside area 2. The other is moving back to area 1. They left four longer duration activity occurrences in area 2. Thus, for area 1, activeness decreases from time 1 to time2, due to spatial intensity decrease (8 occurrence → 0 occurrence). For area 2, activeness increases from time 1 to time 2, due to spatial intensity increase (0 occurrence → 4 occurrence). At time 3, one person starts running inside area 2, leaving four short duration activity occurrences. While, the other person is moving slowly in area 1, leaving two longer duration activity occurrences. Thus, for area

1: activeness increases from time 2 to time 3, due to spatial intensity increase (0 occurrence → 2 occurrence). For area 2, activeness increases from time 2 to time 3, due to duration decrease (4 long duration occurrence → 4 short duration occurrence).

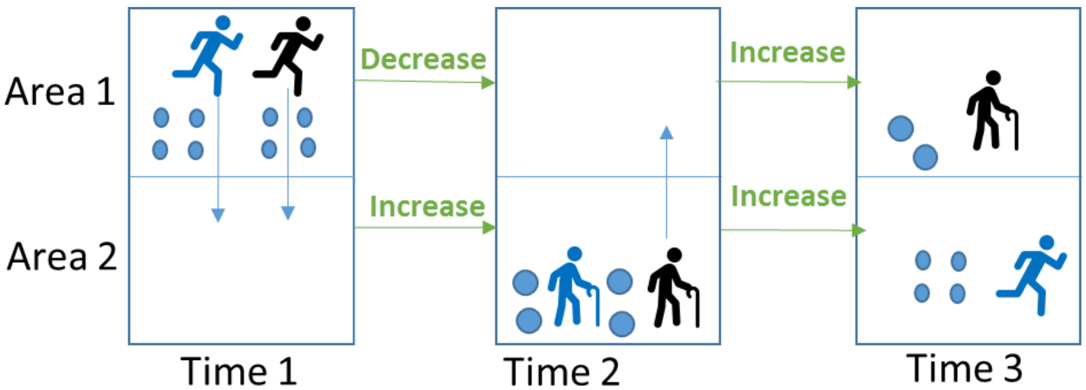


Figure 13. Schematically of Relationship between Changes of Activity Pattern and Activeness Value

3.2 Nonparametric Monte Carlo Approach for Detecting Spatial-Temporal Pattern Change

In this section we focus on comparing of spatial-temporal patterns in two different time intervals, which is represented by the activeness value. We proposed a modified statistical approach for the comprising, based on a nonparametric Monte Carlo approach given by a series of papers led by Andersen (Andresen 2009, Andresen and Linning 2012, 2012, Andresen and Malleson 2013b). They used this method to compare the similarity of two datasets based on polygons. We adopted and revised this similarity comparison method to detect our mobility change. Andresen created an index S to represent the similarity of two datasets. The index S is calculated as: $S = \frac{\sum_{i=1}^{N} s_i}{N}$, which summarizes the comparison result in the whole region of datasets by calculate percentage of similar spatial patterns in N polygons. S_i is signed to 1 if the spatial pattern of two data sets is considered similar for Polygon i, and 0 otherwise.

In this thesis, $X_{similar}$ index represents the overall spatial similarity of the study area as the S-index, but added the increase and decrease evaluation of activity intensity. Based on the criteria of Anderson's method, if S-index ⩾ 0.80, the spatial patterns will be considered similar. We used the same criteria for X. Otherwise, the pattern can be considered as change, either $X_{decrease}$ (denotes the proportion of polygons wich consider Activeness decrease) decrease or $X_{increase}$ for increase (denotes the proportion of polygons wich consider Activeness increase). Thus, we introduce the net increase $X_{increase}$ - $X_{decrease}$ to describe the relationship between increase and decrease. The extended activeness index A is defined as:

$$A = \begin{cases} X_{similar}, & X_{similar} \geq 0.8 \\ X_{increase} - X_{decrease}, & X_{similar} < 0.8 \end{cases}$$

Figure 14 shows the flow chart of calculating the extended index. We first resampled the data using the Monte Carlo approach and created a confidence interval of the Activeness value. Then we compare the confidence interval to the observed density from base point dataset. Finally, the overall index is calculated.

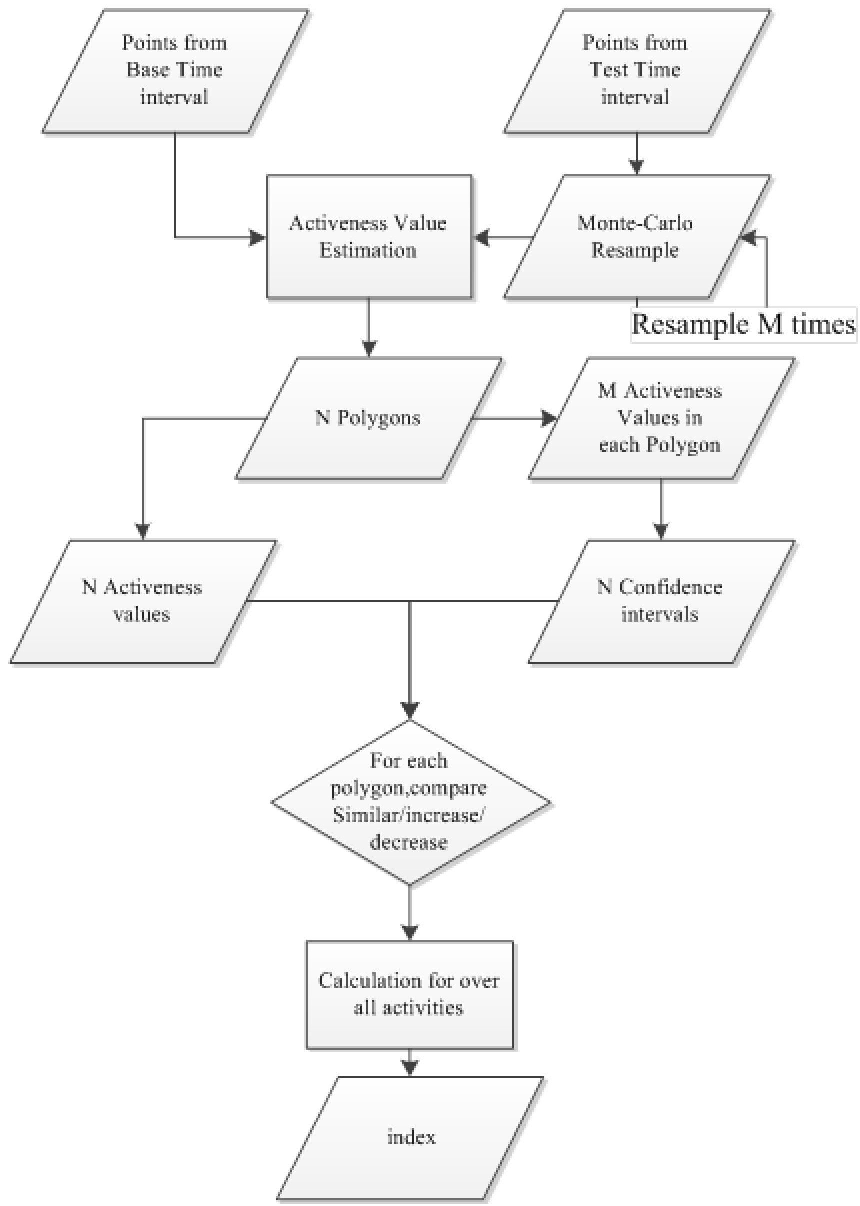


Figure 14. Flow Chart for Calculation

Table 4 Variables for Calculation

Variables	Explainations
Bp	Base point dataset at time interval t

Tp	Test points dataset at time interval t
P_i	polygons inside the area
Db_i	Density/Activeness values for Bp
Dt_{ij}	Density/Activeness values for Tp
T_i	Confidence interval for the test
X	result indexes indclding $X_{similar}$, $X_{increase}$, $X_{decrease}$
i	ranges from 1 to N
N	The number of polygon
j	ranges from 1 to M
M	the times of resample

.

Step 1: For comparing spatial pattern of two occurrence datasets of different time $s_1(x, y, t_1)$, $s_2(x, y, t_2)$, choose one dataset as base points Bp, and another dataset as test points Tp.

Step 2: Prepare geographic polygons P_1 to P_N in the study area as comparison unit

Step 3: Calculate the Activeness as in eq 4 for Bp, and for each polygon P_i , calculate mean Activeness value within the polygon as Db_i.

Step 4: Resample test points with replacement for M times, the total points for each sample is 0.85 of the original test points dataset. Also, for each sample and each polygon, calculate the mean Activeness/kernel density value, Dt_{ij}.

Step 5: For each P_i, calculate the 0.025 and 0.975 quantiles of Dt_i, save as confidence interval T_i.

Step 6: Compare base polygon's density D_{bi} with confidence interval Ti. The result fell into three categories. If D_{bi} falls inside Ti, this means two spatial patterns are similar; If D_{bi} falls outside Ti and is greater than the Base polygon's density, the activity intensity increases; Similarly, if confidence interval is lower than the Base polygon's density, the activity intensity is decreased.

Step 7: Count the number of polygons in each category of result. Divide the count number by N to get result total summaries: $X_{similar}$, $X_{increase}$, $X_{decrease}$.

By applying this method to our case study with different time interval from hourly to weekly, we got the similarity ranged from 0.0007 to 0.088. This means none of two human activity patterns are similar and it changes dramatically even by hourly.

CHAPTER 4

RESULT AND DISCUSSION

The method we proposed can describe both individual behavior and group overall behavior. We processed the data of over 300 APP users and got 32255 valid duration records. Activeness index in every one-hour interval of daily was calculated, and analysis was made based on the index.

4.1 Daily Spatial Pattern Analysis

To find the spatial-temporal regularities of daily human activities, we divided one day into time intervals with one hour for each interval and aggregate every day's records into each interval accordingly. Only weekday data was used because we believe people's behavior regularity will be different on weekends. We used one certain hour's records as base data to compare to the next hour's records.

Figure 15 shows the calculated result. As can be seen from the result, in working days (weekday), there are several activity increase intervals, from 06:00 am to 07:00 am, from 10:00 am to 11:00 am and from 18:00 pm to 03:00 am. There are several activity decrease intervals, from 03:00 am to 06:00 am, from 06:00 am to 09:00 am, and from 11:00 am to 17:00 pm. It is interesting that there is a slight decrease during 22:00 pm to 23:00 pm.

We think the increase during 06:00-07:00 am is the time interval for departure commuting, indicating people are waking up in the morning and departure for work. The big activity increase around 10:00 am is a combination result of business and commuting activities, indicating some people travel to different places for business after some preparation/meeting in office, and others are going to work or going home after finishing the morning work. The interval starting from 18:00 pm should be the time interval for leaving commuting, indicating people finished their work and started heading home or having non-work time activities. Most people fall into sleep during 03:00-05:00 am, and stay in office or home during 11:00 am-17:00 pm. The abnormal activities during 01:00-03:00 am is investigated later in section 4.2. The slight decrease during 22:00-23:00 pm might indicate that some people go to rest from where they are staying. In general, we can see that activities are decreased during working time and increased during non-working time.

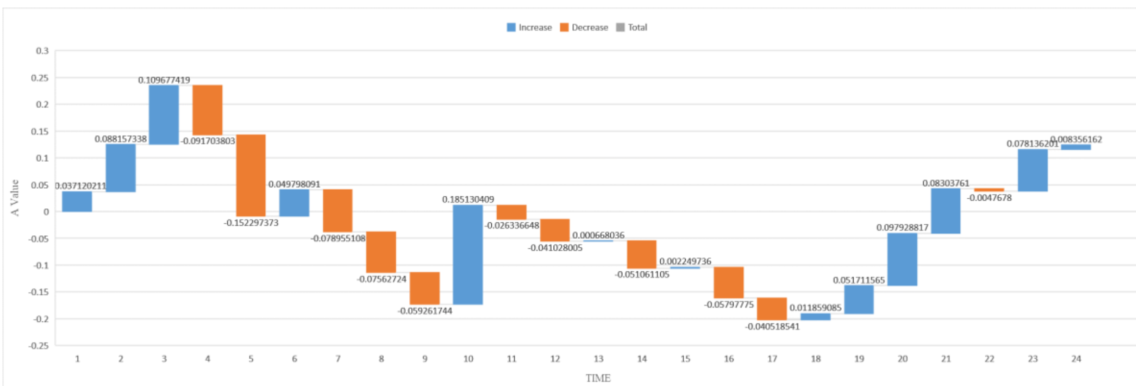


Figure 15. Index Result of Dataset Which Has Duration up to 72 Hours

We use Figure 16 to analyze and verify the temporal pattern changes. Figure 16 a and Figure 16 b show the activity changes at 10:00 am and 18:00 pm, the two commute time intervals we detected in Figure 15. The activities increased mainly around downtown, international airport and along main roads (the red lines in the map). At 10:00 am many activities happened in the northern part of the area, the university district. This justified the above deduction that some working activities happened in 10:00 am interval. Figure

16 c. and Figure 16 d. show the activity change comparison result at 06:00 am and 19:00 pm. We found that in 06:00 am people has an activity increase in the southern part of the city. Since this time interval was assumed by us as the departure commuting period in the morning, we can deduce that the southern part of this area is mainly residential zone. The pattern at 19:00 pm shows another high activity in the same area, indicating people are commuting home after work.

To further verify our deduction, we overlap demographic data with our maps. Figure 17 shows the number of total housing unit based on 5-year Estimate for census tracts from the American Community Survey in Texas State. We think the distribution of housing unit should not change a lot within recent years. Although the demographic data was collected in different year with our data, the distribution of housing unit quite well coincides with the area we found in Figure 16. We can see from Figure 17 that low housing unit in the central area are mainly the working area, the higher housing unit value distributing around the central are mainly residential area and university zone. The user of the APP might mainly live in the south part of the area.

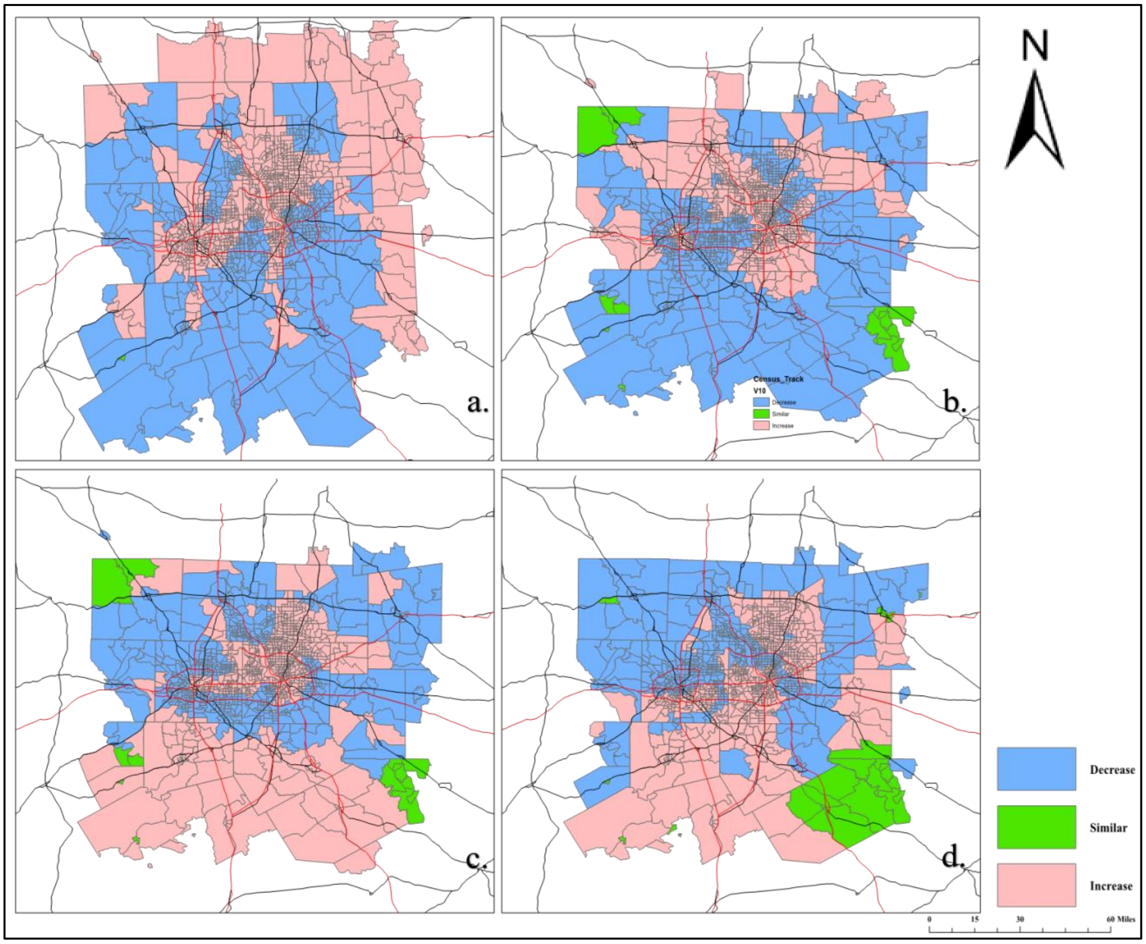


Figure 16. Activeness Similar/Decrease/Increase Based on Census Track, a: 10:00 am, b:18:00 pm, c: 06:00 am, d: 19:00 pm.

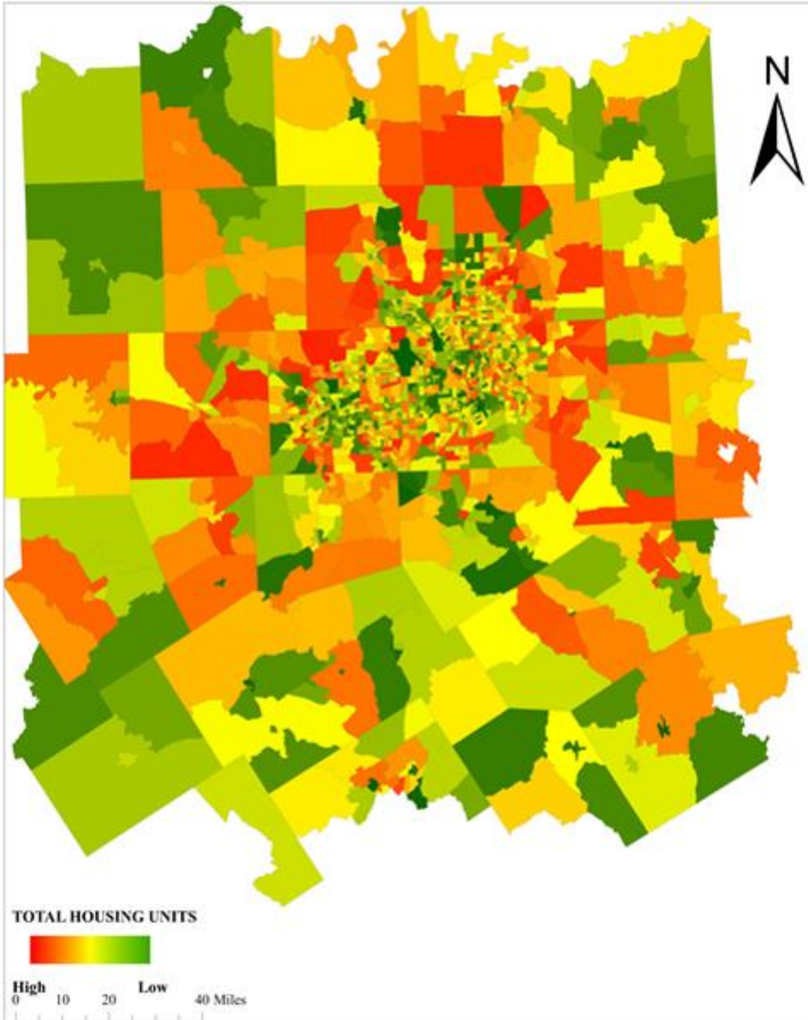


Figure 17. Total Residential Housing Unit by Census Track

4.2 The Abnormal Activities During 01:00-03:00 am

Furthermore, we try to further figure out why there is still some activeness index increase from 1:00 am to 3:00 am. We applied kernel density estimation to each individual person's GPS data, as shown in Figure 11 of section 2.3.3. Now we divide every person's historical footprint into several polygons according to the kernel density value (from high to low), as illustrated in Figure 18. It means people stay longer in high density value polygon than the low density value polygons. Figure19 shows the activity occurrence difference between top 20% polygon and all other polygons by hour of all individuals. We can see during midnight (1:00 – 3:00 am), activity occurred mainly in the top 20% kernel density polygon. This means people are mainly concentrated within top 20% polygons, so causes high intensity in 20% polygon with quite small space range. The locations of 20% polygons are very likely people's sleeping places at night. From the view of the whole region, the total activeness-index would show an overall increasing.

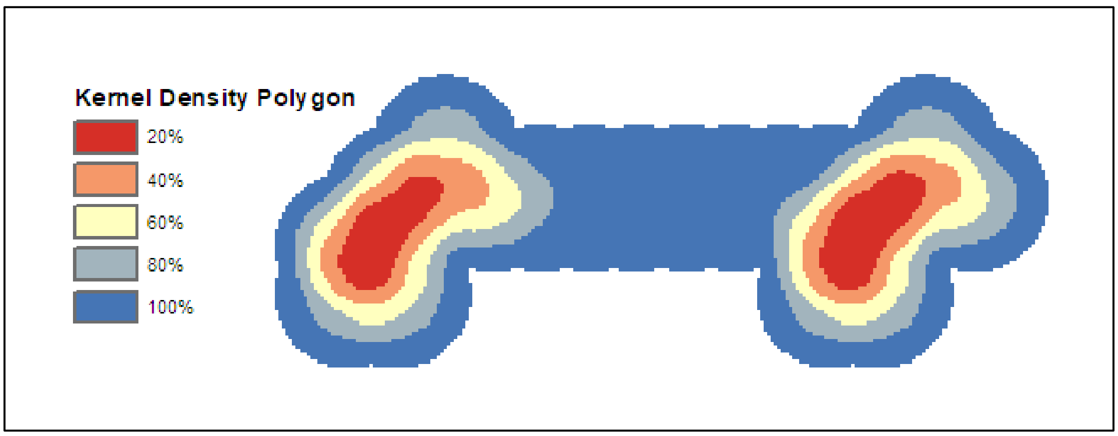


Figure 18. Divide Every Person's Historical Footprint into Several Polygons According to the Kernel Density Value (from high to low)

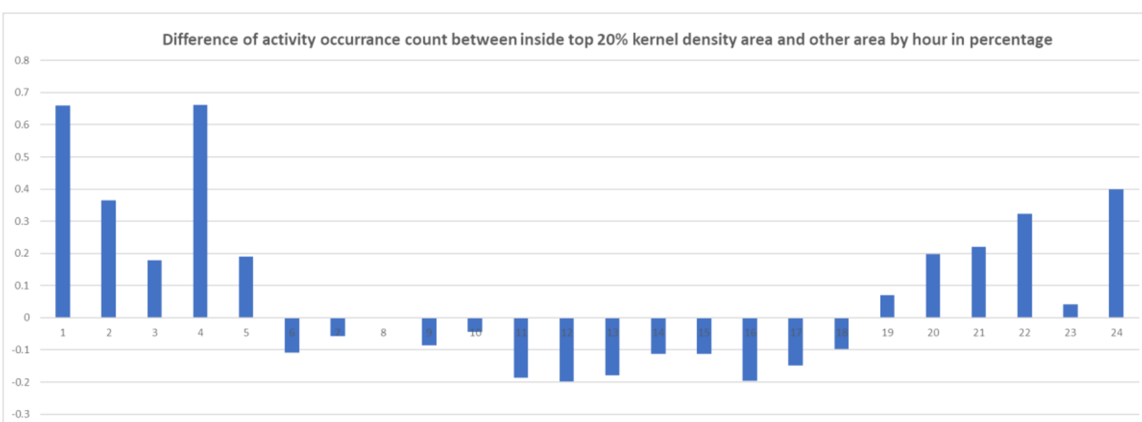


Figure 19. Difference of Activity Occurrence Between Top 20% Kernel Density Area and Other Areas

However, this conclusion is still not enough to explain the phenomenon of activity increase during 1:00-3:00 am. We further checked the activity occurrence of each day at 1:00 am to 3:00 am. The median value of the count of activity occurrence by day is 2, which mean people seldom move at midnight time. Surprisingly, we found that in 3/29/2017 during 1:00 am to 3:00 am, there are 157 activity occurrences, which is unusual and increase the total occurrence count during 1:00 am to 3:00 am by 19.2%, as Figure 20 shows. We thought this could be caused by error data. After further comparing with the system log records of CASA Alert, we found a better reason for that. In 3/29/2017 evening between 1:00 am and 3:00 am, major storms with 70-90 mph winds and tornadoes happened through our study area. People may have woken up and sheltered in their basements, or checked their properties. This late night episode further validates our method.

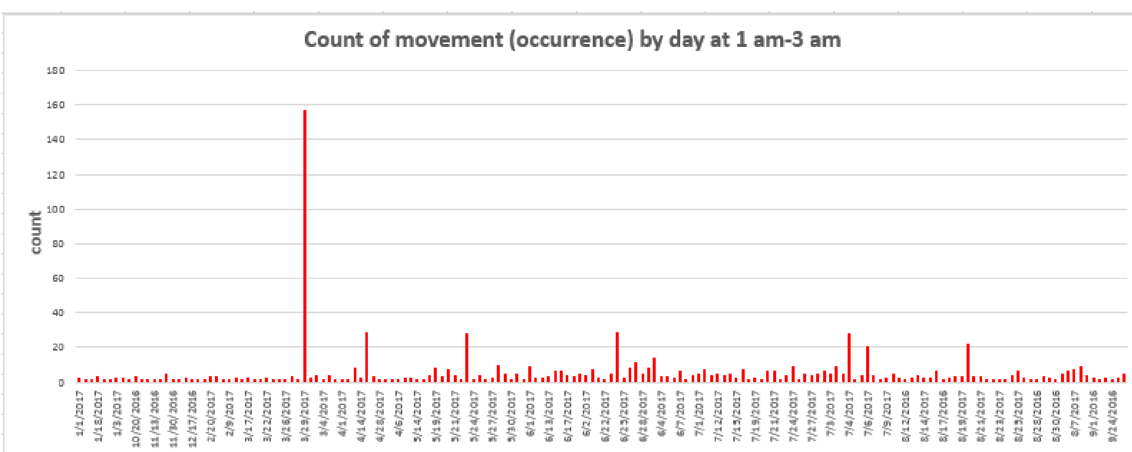


Figure 20. Abnormal Activity Occurrence During 1:00-3:00 Am, 3/29/2017

4.3 Eliminate the Noise Caused by Overlength Duration Sample Points

Another advantage of applying the inverse weight function defined in section 2.3.3 is that it can reduce the influence of ‘noise’ caused by the over-length durations. There are some over-length durations in the dataset been recorded when the APP user did not bring the mobile device with him/her or did not turn location service. These over-length durations will decrease the computational efficiency and accuracy.

Figure 21 and Figure 22 show the calculated results for subsets of data with different maximum duration length. Figure 15 is calculated using original dataset in which data points has a maximum duration of 72 hours, whereas Figure 21 and Figure 22 uses subset which duration less than 24 hours and 12 hours respectively. We can see that all of them have a similar overall trend as indicated in section 4.1.

The average duration time of data points hourly in each subset is shown in Figure 23. We can find in subset with long duration such as 24 and 72 hours, the trends become unclear due to the strong influence of long durations in each time interval. In the subsets with long durations less than 12 hours, the value goes low during 19:00 pm - 3:00 am because long duration has been filtered. Thus we think index works better than durations.

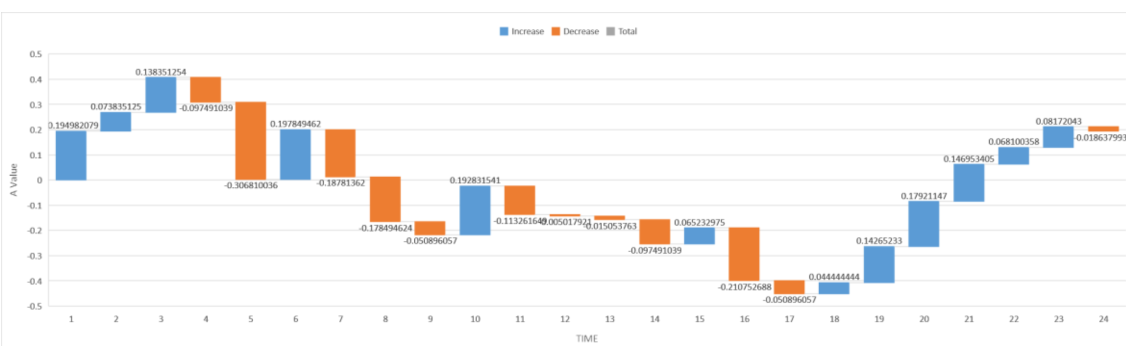


Figure 21. Index Result of Dataset Which Has Duration Less 24 Hours

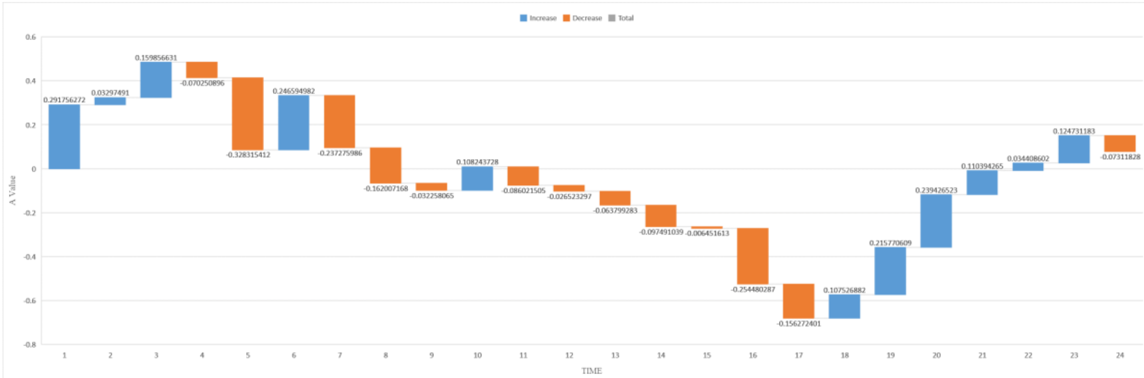


Figure 22. Index Result of Dataset Which Has Duration Less 12 Hours

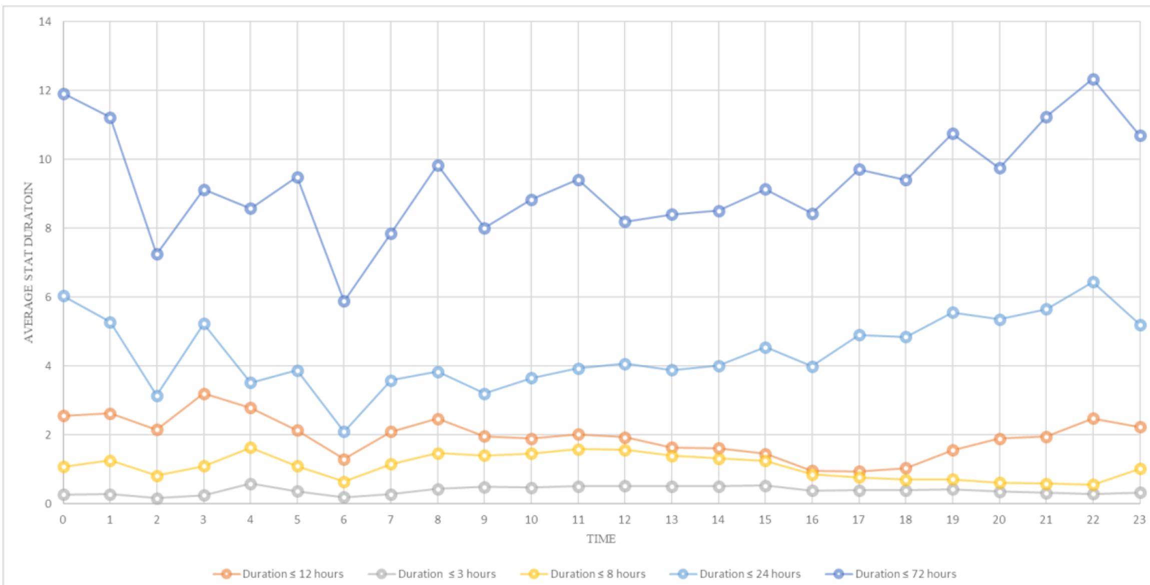


Figure 23. Average Stay Duration of Each Subset of Data Versus Time

4.4 Weather Warning Suggestions by Considering Hourly Human Activity Pattern

The approach can be used by weather alerting agencies. For every individual user or grouped user, we can collect their historical mobility data. By applying the method in this paper, their specified time interval and moving pattern can be found. With this spatial-temporal pattern, we can first generate people's historical footprint location polygon, then send the alert in an intentional way by comparing alert area polygon to user foot print polygon. For example, we can send weather alert about a potential tornado happening not just in the person's current location, but in his route to his place of work. The specific **algorithm** shows below:

Algorithm 1:

If (alert polygon location and alert time interval intersects user current location)

{

Send regular alert

}

If (alert polygon location and alert time interval intersects user historical foot print polygon)

{

Send regular Alert

Include intersected foot print location and time interval

}

4.5 Conclusion and Perspective Study

This paper proposed a GIS data model to describe the spatial-temporal pattern of human activity, a metric to represent the intensity of the activity, and an index to describe the change of the activity. By applying to high spatial-temporal resolution trajectories of CASA ALERTS users, we verified that human daily mobility pattern can be clearly described with the proposed methods. Especially, the method in this thesis is able to detect the influence of severe weather events to human mobility pattern. Targeted weather warning strategy can be designed based on the hourly human mobility pattern generated with this method. The method also shows high efficient by reducing data redundancy with a new data structure, and high robustness by reducing the influence of long duration data.

Future Research will focus on the following aspects:

First, deeper analysis based on more APP records and extended information of the users, including specified group mobility pattern based on profession and age of the APP users; pattern in longer time period considering seasons, holidays, etc. Transportation data, such as public transportation records, real-time traffic flow, will also be very helpful to conduct further analysis.

Second, research on how to predict the possible behavior of the users based on their routine mobility pattern.

Third, study the strategy of targeted alert during in-alert and post-alert periods based on our research methods, and its integration with the CASA Platform.

BIBLIOGRAPHY

Andresen, M.A., 2009. Testing for similarity in area-based spatial patterns: A nonparametric Monte Carlo approach. *Applied Geography*, 29 (3), 333–345.

Andresen, M.A., 2016. An area-based nonparametric spatial point pattern test: The test, its applications, and the future. *Methodological Innovations*, 9, 2059799116630659.

Andresen, M.A. and Linning, S.J., 2012. The (in)appropriateness of aggregating across crime types. *Applied Geography*, 35 (1–2), 275–282.

Andresen, M.A. and Malleson, N., 2013a. Crime seasonality and its variations across space. *Applied Geography*, 43, 25–35.

Andresen, M.A. and Malleson, N., 2013b. Crime seasonality and its variations across space. *Applied Geography*, 43, 25–35.

Basher, R., 2006. Global early warning systems for natural hazards: systematic and people-centred. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 364 (1845), 2167–2182.

Becker, J.S., Taylor, H.L., Doody, B.J., Wright, K.C., Gruntfest, E., and Webber, D., 2015. A Review of People's Behavior in and around Floodwater. Weather, Climate, and Society, 7 (4), 321–332.

Etter, V., Kafsi, M., Kazemi, E., Grossglauser, M., and Thiran, P., 2013. Where to go from here? mobility prediction from instantaneous information. Pervasive and Mobile Computing, 9 (6), 784–797.

Gao, S., 2015. Spatio-Temporal Analytics for Exploring Human Mobility Patterns and Urban Dynamics in the Mobile Age. Spatial Cognition & Computation, 15 (2), 86–114.

González, M.C., Hidalgo, C.A., and Barabási, A.-L., 2008. Understanding individual human mobility patterns. Nature, 453 (7196), 779–782.

Guo, D., Zhu, X., Jin, H., Gao, P., and Andris, C., 2012. Discovering Spatial Patterns in Origin-Destination Mobility Data: Discovering Spatial Patterns in Origin-Destination Mobility Data. Transactions in GIS, 16 (3), 411–429.

Huang, W., Li, S., Liu, X., and Ban, Y., 2015. Predicting human mobility with activity changes. International Journal of Geographical Information Science, 29 (9), 1569–1587.

Lindell, M.K. and Hwang, S.N., 2008. Households' Perceived Personal Risk and Responses in a Multihazard Environment. Risk Analysis, 28 (2), 539–556.

Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P., and Tita, G.E., 2011. Self-Exciting Point Process Modeling of Crime. Journal of the American Statistical Association, 106 (493), 100–108.

Renn, O., Jovanovic, A., and Schröter, R., 2011. Social unrest. OECD/IFP Project on "Future Global shocks.

Silverman, B.W., 1986. Density estimation for statistics and data analysis. CRC press.

Song, C., Qu, Z., Blumm, N., and Barabási, A.-L., 2010. Limits of predictability in human mobility. Science, 327 (5968), 1018–1021.

Song, L., Kotz, D., Jain, R., and He, X., 2006. Evaluating next-cell predictors with extensive Wi-Fi mobility data. IEEE Transactions on Mobile Computing, 5 (12), 1633–1649.

Wang, Z., Ye, X., and Tsou, M.-H., 2016. Spatial, temporal, and content analysis of Twitter for wildfire hazards. Natural Hazards, 83 (1), 523–540.

Mitchell, C. (2017, February 7). Psychology of warnings: Why do people ignore important weather alerts? Retrieved from https://www.accuweather.com/en/weather-news/psychology-of-warnings-why-do-people-ignore-weather-alerts/70000135

Bean, H., Sutton, J., Liu, B.F., Madden, S., Wood, M.M., and Mileti, D.S., 2015. The study of mobile public warning messages: A research review and agenda. Review of Communication, 15 (1), 60–80.

Mileti, D.S. and Peek, L., 2000. The social psychology of public response to warnings of a nuclear power plant accident. Journal of hazardous materials, 75 (2–3), 181–194.

Mileti, D.S. and Sorensen, J.H., 1990. Communication of emergency public warnings: A social science perspective and state-of-the-art assessment. Oak Ridge National Lab., TN (USA).

Thakali, L., Kwon, T.J., and Fu, L., 2015. Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. Journal of Modern Transportation, 23 (2), 93–106.