CZECH TECHNICAL UNIVERSITY IN PRAGUE

Faculty of Electrical Engineering

Department of Cybernetics

BACHELOR THESIS

Anastasia Grigoryan

Predictive Temporal Models for Effective Social Distancing during Pandemic

Thesis supervisor: **Tomáš Krajník**

August, 2021

BACHELOR‘S THESIS ASSIGNMENT

CTU CZECH TECHNICAL UNIVERSITY IN PRAGUE

I. Personal and study details

Student's name:	**Grigoryan Anastasia**	Personal ID number:	**483742**
Faculty / Institute:	**Faculty of Electrical Engineering**		
Department / Institute:	**Department of Cybernetics**		
Study program:	**Open Informatics**		
Specialisation:	**Artificial Intelligence and Computer Science**		

II. Bachelor’s thesis details

Bachelor’s thesis title in English:

Predictive Temporal Models for Effective Social Distancing during Pandemic

Bachelor’s thesis title in Czech:

Prediktivní modely pro společenskou zdrženlivost během pandemií

Guidelines:

1) Research the impact of social distancing on the spread of pathogens in the population, e.g. [1].
2) Research the use of anti-covid mobile apps. (diagnostics, tracing, etc.), e.g. [2].
3) Research spatio-temporal models of human presence in mobile robotics, e.g., [3,4].
4) Propose a scenario to evaluate the impact of forecasting models on the reduction of individual infection risk.
5) Select or gather datasets relevant to the proposed scenario.
6) Design and implement a tool capable to automatically evaluate the performance of the individual methods.
7) Perform the experimental evaluation and discuss the feasibility of using objects to forecast people presence over time.

Bibliography / sources:

[1] Martín-Calvo D., et.al.: Effectiveness of social distancingstrategies for protecting a community from a pandemic with a data-driven contact network based on census and real-world mobility data.
[2] Drew D.A. et.al.: Rapid implementation of mobile technology for real-time epidemiology of COVID-19. Science, 2020.
[3] Vintr T. et al.: Spatio-temporal representation for long-term anticipation of human presence in service robotics. In ICRA 2019.
[4] Blaha, J. : Inferring models of people presence from environment structure. Thesis, CTU, 2020.

Name and workplace of bachelor’s thesis supervisor:

doc. Ing. Tomáš Krajník, Ph.D., Artificial Intelligence Center, FEE

Name and workplace of second bachelor’s thesis supervisor or consultant:

Date of bachelor’s thesis assignment: **07.01.2021** Deadline for bachelor thesis submission: **13.08.2021**

Assignment valid until: **30.09.2022**

doc. Ing. Tomáš Krajník, Ph.D. Supervisor’s signature	prof. Ing. Tomáš Svoboda, Ph.D. Head of department’s signature	prof. Mgr. Petr Páta, Ph.D. Dean’s signature

III. Assignment receipt

The student acknowledges that the bachelor’s thesis is an individual work. The student must produce her thesis without the assistance of others, with the exception of provided consultations. Within the bachelor’s thesis, the author must state the names of consultants and include a list of references.

.

Date of assignment receipt — Student’s signature

Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

In Prague on 13.08.2021 Grigoryan Anastasia

Acknowledgements

I am thankful to my supervisor Tomáš Krajník for his support, motivation and help during the hard time of distance education. Also i would like to thank my family for their faith in my strength and constant support throughout my studies. And I express my gratitude to my partner who was always by my side.

Abstract

As a result of the world-wide pandemic of 2020 caused by covid-19, social distancing is necessary to minimize virus spreading. The problem of predicting where and when events will happen with high probability is the main topic of this work.
This thesis tries to investigate methods, that allow to combine the public data and crowd sources to improve quality of information to the temporal modelling techniques. The goal of this thesis is development, implementation and experimental validation of predictive models, which enable to predict number of people in the given place at the given time. The main purpose of this model is to enable threatened people to plan their necessary trips, so they could minimize the amount of unnecessary social contacts.
The result shows that the recommendation system based on temporal models significaly reduces the risk for people who follow it.

Keywords: temporal models, social distancing, recommendation systems, predictive modeling.

Abstract

V důsledku celosvětové pandemie v roce 2020 způsobené virem COVID-19 je k minimalizaci šíření virů nezbytné sociální distancování. Problém predikce, kde a kdy se události stanou s vysokou pravděpodobností je hlavním tématem této práce.
Tato práce se snaží prozkoumat metody, které umožňují kombinovat veřejné údaje a zdroje davu za účelem zlepšení kvality informací v technikách dočasného modelování. Cílem této práce je vývoj implementace a experimentální validace prediktivních modelů, které umožňují predikovat počet lidí na daném místě v daném čase. Hlavním účelem tohoto modelu je umožnit ohroženým lidem naplánovat si nezbytné cesty tak, aby mohli minimalizovat množství zbytečných sociálních kontaktů.
Výsledek ukazuje, že systém doporučení založený na dočasných modelech významně snižuje riziko pro lidi, kteří ho dodržují.

Klíčová slova: časové modely, sociální distance, doporučovací systémy, prediktivní modelování.

Contents

1 Introduction **1**
1.1 Organisation 2

2 State of the art **3**
2.1 Types and usage of anti-covid mobile applications 3
2.2 Types of social restrictions 4
2.3 Crowdedness forecasting 5
2.4 Predictive modeling 6

3 System description **7**
3.1 Standart behaviour 7
3.2 Minimum recommendation strategy 7
3.3 l Additional amplitude model(AAM) and l Best amplitude model(BAM) . 8
3.3.1 Fourier transform 8
3.3.2 Fourier transform for parsed irregular data 9
3.3.3 Implementation 9
3.3.4 AAM and BAM 10
3.3.5 Implementation 11

4 Datasets **13**

5 Experiments **15**
5.1 Strategies 15
5.2 Performance of minimum algorithm 16
5.3 Performance of l AAM and l BAM 18
5.4 Comparison of methods by paired student's t-test 22

6 Conclusion **24**

List of Figures

1 Visualization of influence on the spread of the virus by self-distancing and teleworking [15] 5
2 Map with Index Living Mall and Big C Rangsit 14
3 Example of input sequence 15
4 Histograms for shops 7-11, Green canteen, big C and HOME RPO 16
5 Histograms for shops Makro, Puey Library, Robinson, Tesco Lotus 17
6 Histograms for shops Honda Dealer, Index Living, Khon Luang cluster . . 18
7 Result of AAM and BAM based on GPS dataset 19
8 Result of AAM and BAM based on GPS dataset 20
9 Result of AAM and BAM based on GPS dataset 20
10 Result of AAM and BAM based on observation dataset 21
11 Result of AAM and BAM based on observation dataset 21
12 Result of AAM and BAM based on observation dataset 22

1 Introduction

In the modern world, almost everyone uses the Internet every day for different purposes: watching movies, listening to music, reading the news, or shopping online. But sometimes person is puzzled by a huge selection of music, films and goods, because it is not always possible to decide what to choose for yourself this time. This is exactly what recommendation systems exist to make it easier for a person to make a choice.

In the past, scientists conducted experiments and observations and based on them, made conclusions about what would be better for a person. These were the first recommendation systems and predictive models, but with the development of technology, scientists do not have to manually calculate probabilities and give recommendations to people, because now computers are doing this.

A recommendation system is a set of algorithms that makes recommendations based on information that has been collected in the past. This tool tries to predict what goods or events a user would be interested in. It is widely used in various spheres of life : music and video services, social networks and e-commerce. It is equally important to understand what predictive models are. It is a mathematical process that tries to predict future events based on analyses of patterns that are likely to forecast future outcomes.

A fairly common example of a recommendation system can be considered, for example, the work of any platform for watching films. On the page of the selected movie, similar movies are almost always displayed along with the selected movie. Predictive models are also used in that sphere of life. For example, Store owners also use predictive models based on the data they have from their resources. This helps them improve sales, reduce advertising costs, and reduce customer churn. And the most important thing is the fact that they begin to understand their clients and their desires. This is one of the simplest examples of a recommendation system and predictive modeling used in the entertainment industry. But such systems are also used in those areas of our life that are critically important for humanity. And now, at a time when the world is gripped by the coronavirus pandemic and every day we are all at risk, predictive and recommendation systems can be used to mitigate that risk.

During the pandemic, it was revealed that social distancing is one of the most efficient methods to slow down spreading of this virus. And with the help of past data, about the number of people in the selected place, the predictive model can predict the further course of events and, based on it, give advice on what time it is best for the user to visit the place he needs with the minimum risk of being in the crowd and becoming infected.

And in this work, an analysis of several predictive methods will be carried out to recommend the most safe time to visit a particular place.

1.1 Organisation

The thesis is organised into five Sections, first being this Introduction. Second Section is “State of art”. In this section we discuss developments in the field of social distance, we will also analyze predictive temporal models, their work and new research in this area.

Third Section “System description” describes the complete principle of operation of those predictive algorithms that were used for all experiments in this paper.

Fourth Section “Datasets” shows data on which the experiments were carried out and on the basis of which the conclusion of this work was made. In this paper, predictions will be made based on data collected from real locations in Thailand. For the same locations, there will be two types of data, one of which was collected from GPS coordinates, and the other was obtained from accurate observation at each location. It is on the basis of this data that the analysis will be carried out.

Fifth section “Experiments” shows the process of working of algorithms with real data. This work will consider the operation of several algorithms based on the Fourier transformation, AAM (l additional amplitude model) and BAM (l best amplitude model), as well as the simplest algorithm for selecting the minimum and predicting the behavior of a standard person.

And the last, sixth section of this work is called "Conclusion" and it describes the result of all experiments. With real data at the end, predictive algorithms will be compared in order to understand which algorithm gives the most plausible results and which data can be best relied on.

2 State of the art

At the moment, the problem of the spread of coronavirus is still relevant and global and, of course, requires a solution. Social distancing, which can be done in a variety of ways, is a fairly effective way to reduce the spread of the virus.To make it easier for people to maintain social distance and fight the coronavirus, many different services and applications are being developed. Many of these applications are based on predictive models, but there are others that relate to another edge of the problem. This section will deal with exactly this: the variety of anti-covid applications, the effectiveness of social distance in the current environment, and the role of predictive models in all of this.

2.1 Types and usage of anti-covid mobile applications

Let's start by giving a classification to different types of anti-covid applications and briefly describe how it works.

1. Coronavirus Symptoms Apps

With symptom tracker app, person can record his symptoms almost anytime. It is a good way to control health every day without any problem. This type of application works by asking questions about your health condition and also about common symptoms of COVID-19. The entire history of a person's health records is saved in the application and can be shown to the doctor at any time. Examples of such applications are: Symptomate[1] and Data4Life[2] .

2. Coronavirus Contact Tracing Apps

This type of app is used by health departments to find people who may have been in contact with an infected person. Some could identify people who might have been exposed to the virus so they know how to isolate themselves and watch for symptoms. Some people wary of the fact that an app could track where they go and whom they meet, but information will be available only to health agencies. An example of such an application used in the Czech Republic is eRouška[3] .

3. Coronavirus Health Monitoring Apps

This type of application can be called a "doctor's assistant", because it helps doctors keep an eye on their patients' health. Combined with special sensors, health monitoring apps help doctors track the patient's health. Apps like that can collect data on important aspects of human health and ask questions about their symptoms, and then doctor collects this data from the application. Rave Mobile Safety offers a platform for that[4] .

4. Pulse oximeter apps

Patient who has a COVID-19 must ensure peripheral capillary oxygen saturation (SpO2) level above 95 percent, and anything below must require medical attention. And during the epidemic, it was difficult to buy oximeters, since they were sold out in most stores and pharmacies. However, at the moment there are some applications that offer the same functionality. These apps use the phone's flashlight and camera instead of infrared light in an oximeter to measure Spo2 levels. For example, CarePlix Vitals[5] shows fairly accurate results during each test.

5. Thermometer apps

Fever is considered one of the symptoms of coronavirus. Therefore, during the pandemic, many places require that people measure their own temperatures before entering. That is why it is important to have a device with you that could quickly and conveniently measure your temperature. Of course, a thermometer app can not be as precise as an actual thermometer, but Fever Tracker[6] comes very close to being reliable.

6. Coronavirus Prevalence Apps

Apps and dashboards can provide the latest information about the virus and also about health and safety resources. The World Health Organization (WHO) dashboard[7] follows the number of confirmed cases and deaths by world region. Unfortunately this service is not available for download, but it is conveniently arranged for viewing on a mobile device.

7. Coronavirus Research Apps

These applications were developed to help researchers study the spread of COVID-19. Users enter data about their health and possible symptoms, then this information is processed by researchers to find ways to prevent the spread of the disease. The COVID Control app[8] from Johns Hopkins University lets you enter your temperature every day.

8. Social Distancing Apps

Social distance is considered to be an effective way to prevent the spread of the virus. Special apps can help people to maintain social distance. There are lots of ways to use apps for this purpose. For example, cafes and restaurants can use them to avoid queues and crowds , so that customers will be informed about free table. A great example of such an application is FreMEn contra COVID[9].

9. Telemedicine Apps

What if the person suddenly felt unwell? Of course he can call for a taxi and go to the nearest hospital. But at the moment there is a great alternative for this. Such applications allow a person get immediate consultation from doctors using instant video chat and massage. At to present time, a big advantage of this method of consulting a doctor is that a person can avoid unnecessary social contacts. Good examples of such applications are MDLive[10] and Teladoc[11].

2.2 Types of social restrictions

Social distancing measures can come in many forms, but few can be identified:

1) Closure of educational institutions,

2) Self-distancing and distance work,
3) Closure of educational institutions with self-distancing and distance work,
4) Closure of public places (cafe's, museums etc.),
5) Closure of every place that is not necessary for everyday life of people,
6) All methods above within the metropolitan area.

Based on [15] we can say, that the most effective way for minimization of the virus spreading is social distancing and distance work.(see Figure 1)

Also World health organisation (WHO) in its list of recommendations to prevention also uses social distancing as a way for minimization of the virus spreading. [12]

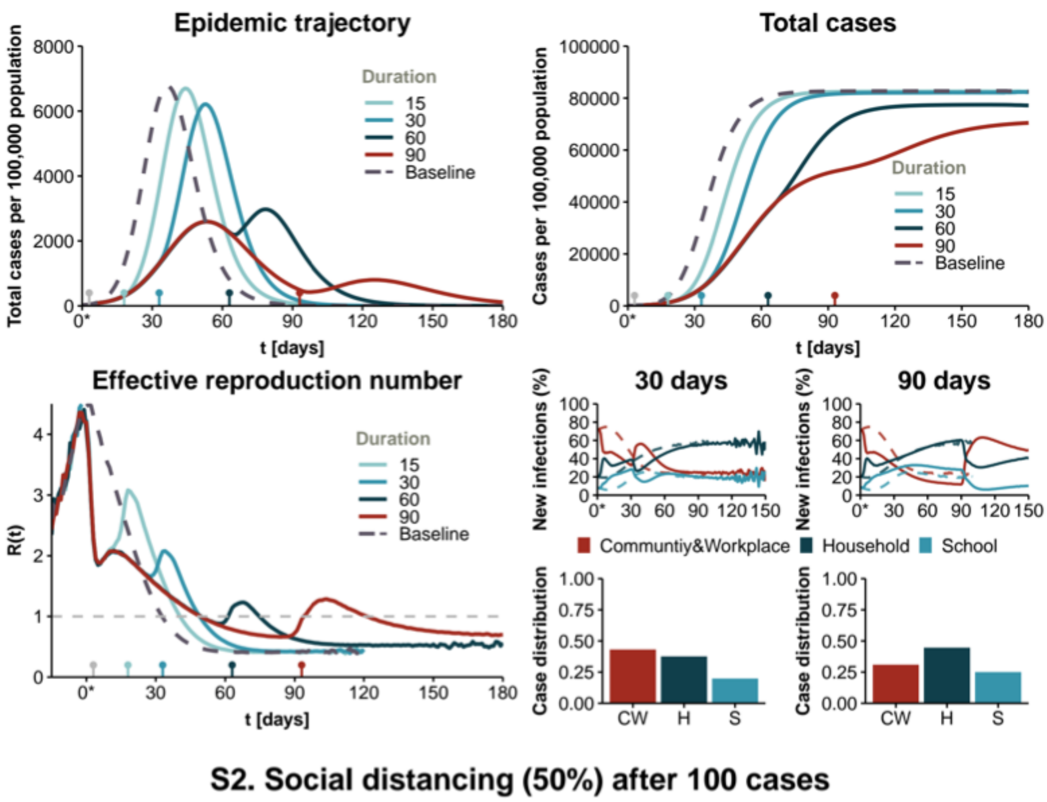

Figure 1: Visualization of influence on the spread of the virus by self-distancing and teleworking [15]

2.3 Crowdedness forecasting

If someone wants to follow recommendations and use social distancing efficiently, he needs to avoid visiting different places at peak times. This measure requires the ability to predict amount of people in this place, so one can plan his trip to the location he needs.

Nowadays a lot of researches indicate that forecasting crowdedness in public locations is possible. But for that we need to collect some data, which gives us an information about people's present at different locations. We can collect this information with using:

1) “Popular times” from Google Inc.

2) Facebook's Mobile Locator

These methods have deficiencies, such as long-term data collection, which is too slow for crowdedness dynamics. However, it's worth saying, that a major disadvantage of collecting data using Google is that The Global Positioning System (GPS) location determination provides not accurate information. An example would be a situation where the bus stop is located opposite the entrance to the store. All people standing at the bus stop will be identified by this system as people visiting the store, which is incorrect information.

But in [14] it is demonstrated, that previously collected data to forecast the number of people can be used for creating a schedule for safe trips.

2.4 Predictive modeling

As it was written above, there are applications to help people to maintain social distance. To analyze a place and choose the safest time to visit it, we can apply some of the simple methods. For example, we can divide locations, time, and both of that to understand the distribution of the population during the day. Based on these distributions, we can estimate the risk of infection for different segments of the population. However, effectively separating people in space and time is extremely hard. Different people still need access to the same essential locations, such as shops, offices and health care facilities. And that's why we use predictive modeling.

Predictive modeling is a statistical technique to predict future events. It works by analyzing past and current data and generating a model which helps to predict future outcomes. "Based on known past information, what is most likely to happen in the future?” – this is the main question that predictive models answer.

Social distancing with the help of predictive modeling can be done in 3 ways: 1. Predictive spatial models for distancing (separating people based on their location) . 2. Predictive temporal models for distancing (separating people based on time of visiting). Temporal distancing is an easy concept to grasp. 3. Predictive spatial-temporal models for distancing (separating people based on their location and time of visiting). Among the above, the most effective model is the predictive spatial-temporal model.

3 System description

The data collection app, ‘FreMEn explorer’ was released to the public in April 2020, and was primarily advertised in the Czech Republic, where the COVID-19 crisis started in mid March. Since the release of the app, until August 2020, people have provided around 4000 measurements from more than 60 locations. [14]

And now investigation involves methods to combine the public data and crowd sources to improve quality of information to the temporal modelling techniques.

3.1 Standart behaviour

For this I use two important formulas:

The number of people n(l) that a person would meet at a certain location l over one day is

$$n(l) = \int_{0h}^{24h} n(t,l)p(t|l)p(l)\mathrm{d}t \quad (1)$$

where n(l, t) is the number of people at a given location, l denotes the location occurring at the location l at time t, given that the person visits the location l that day. The p(l) is the probability of a person visiting the location that very day. In our evaluation, we assume that the visit of the location l is necessary and consider p(l) = 1. [14]

The p(t|l) depends on the fact if the person follows the recommendation of time visits or not. If yes, the p(t|l) is given by the output of our system. If not, we can assume that the shape of p(t|l) follows the density of people n(t,l) and obtain the p(t|l) through normalization as

$$p(t|l) = \frac{n(t,l)}{\int_{0h}^{24h} n(\phi,l)\phi} \quad (2)$$

The p(t|l) calculated by Equation 2 models a person with standard behaviour. [14]

3.2 Minimum recommendation strategy

I applied the ‘Minimum’ recommendation strategy to the forecasts obtained.

Algorithm 1 Algorithm Minimum

Input: Sequence of time and number of people for this time

Output: Time with minimum number of people

```
assume that minimum = 0
assume that recommended time = 0
for a, b ∈ dict[numberofpeople] do
    if a < b then
        minimum = a
    else
        minimum = b
    end if
end for
for c ∈ dict[time] do
    if dict[c] = minimum then
        recommended time = c
        break;
    end if
end for
return recommended time;
```

After that we can assume how many people a person will probably meet if he follows recommendation, for that we use formula (1). And how many people he will probably meet with his standart behaviour, for that we use formula (2).

3.3 l Additional amplitude model(AAM) and l Best amplitude model(BAM)

For working with spectral representation for temporal environment models we use idea proposed in [17] to use not a probabilities of the uncertainty of the environment states, but a probabilistic functions of time. For that reason we use Fourier transform.

3.3.1 Fourier transform

A Fourier transform (FT) is a mathematical transformation for decomposing functions that depend on space or time into functions that depends on spatial or temporal frequency. It is used in the field of signal processing. FT transforms a function of time into a function of frequency. Typically the Fourier transform of a function of time is a complex-valued function of frequency, whose absolute value represents the amount of that frequency present in the original function and argument represents phase shifts of the frequency components.

3.3.2 Fourier transform for parsed irregular data

Since the data we collect is not regular and has a lot of missing information, we cannot consider our data discrete. But Fourier transform is aimed at working precisely with discrete sequences of data.

This is why we can use the Fast Fourier Transform (FFT) algorithm, which calculates the frequency spectrum S (w) in a very efficient way. However, it has a significant disadvantage, since the FFT requires that the entire sequence of dates be fully restored. And the main disadvantage in our method can also be considered the fact that FFT requires constant, regular observations, which is quite difficult to implement. Thus, the predictive power of the method will become less and less accurate over time.

To solve this problem, a method was proposed in [17] for updating the space-time model based on irregular observations in stages. Every time the state s (t) is observed at time t, the above view is updated as [17]:

$$
\begin{aligned}
\gamma_0 &\leftarrow \frac{1}{(n+1)} \cdot (n \cdot \mu + s(t)) \\
\gamma_k &\leftarrow 1 \frac{1}{(n+1)} \cdot (n \cdot \gamma_k + s(t) \cdot e^{-j \cdot w \cdot w_k}) \\
n &\leftarrow n + 1
\end{aligned}
\tag{3}
$$

where n represents the number of observations.

3.3.3 Implementation

The implementation of the method described above was implemented in Python, the algorithm by which the complex-valued spectrum was calculated can be seen in Algorithm 2. In the further implementation of the AAM and BAM algorithms, we will use this particular algorithm instead of the standard Fourier transform.

Algorithm 2 Modified Fourier transform

Input:S: input signal
k: size of spectrum
time: set of timestamp corresponding to S
Output: Spect: complex-valued spectrum

$d \leftarrow 86400$ ▷ (Seconds in one hour)
$realParts[0] \leftarrow sum(S)$ ▷ (First element of spectrum is sum of all elements in S)
$complexPart[0] \leftarrow 0$
for i = 0 to size(time) **do**
  for j = 0 to k **do**
    $realParts[j] += S[i] * cos(2 * \pi * time[i] * j/d)$
    $imagParts[j] += S[i] * np.sin(2 * np.pi * time[i] * j/d)$
  end for
end for
$Spect = complex(realParts, imagParts)$ ▷ (Create complex numbers from real and imaginary parts of spectrum)

3.3.4 AAM and BAM

In each of the algorithms, not a standard Fourier transform will be used, but a modified one described by formula (3) and Algorithm 2. This is because our observations are not regular.

We will start with a description of the BAM algorithm, which works based on the choice of l maximum coefficients of the entire signal and on the basis of which it further reconstructs it. We start l BAM algorithm from usage of Fourier transform, after that we choose l coefficients with the highest amplitude. Further, for a smoother signal reconstruction, these coefficients will be used This method is called l Best Amplitude Model (BAM) [13] .However, the disadvantage of this method can be considered that it can not always recognize the magnitude of the input signal.

Using the technique ,which was borrowed from [16] , we modify the way of choosing l maximum coefficients. So with the help of several transformations in [13] a new method was obtained, which is called the Addition Amplitude Model (AAM). Now we do not immediately select the l maximal elements, but first obtain the spectrum using the Fourier transform. After that, from this spectrum, we select the frequency wk with the highest absolute value and subtract it from input signal. And until we obtain l desired coefficients this step iterates multiple times . In case we get a frequency that we have already encountered before, this absolute value is added to the absolute value of the frequency that we have encountered.

The result is stored as an l-sized set consisting of abs (k), arg (k), and k which describe the amplitudes, the phase shifts and frequencies of the spectral model.

The main purpose of AAM is to reconstruct the input signal and represent spatio-temporal signatures for each region. These goals are accomplished through the inverse Fourier transform.

These two algorithms are similar to each other, but the difference in the choice of the maximum l coefficients allows AAM to reconstruct the input signal more smoothly and determine the magnitude of the input signal with better accuracy.

3.3.5 Implementation

These methods were implemented in Python. Libraries such as numpy, geopy, scipy, pandas and matplotlib were used. These two algorithms AAM and BAM can be seen in Algorithm 2 and Algorithm 3.

Algorithm 3 Additional amplitude model(AAM)

Input:S: input sequence
total: maximum total frequency
Output: Rec: AAM reconstruction

$k \leftarrow 1$
//Create a collection of (abs(Wk), arg(Wk), Wk)
$wkAbs = []$
$wkArg = []$
$wkCollecton = []$
while $k < total$ **do**
    $k = k + 1$
    $Wk = FFT(S)[1] * 24/3600$ ▷ (Get the frequency with the highest amplitude)
    if $Wk \in wkCollecton$ **then**
        $idx = arg(Wk)$
        $wk_abs[idx] \leftarrow wkAbs[idx] + abs(Wk)$
        $wkArg = avg(wkArg)$
    else
        $wkCollecton \leftarrow Wk$
        $wkAbs[idx] \leftarrow abs(Wk)$
        $wkArg = arg(wk)$
        $S' \leftarrow abs(Wk) * cos($2∗π * Wk + arg(Wk)) ▷ (Create a cosine signal from k and substract)
        $Rec = S - S'$
    end if
end while

Algorithm 4 Best amplitude model(BAM)

Input:S: input sequence
k: size of spectrum
Output: Ret: BAM reconstruction

$d = 86400$ ▷ (Seconds in one hour)
$Sp = FFT(S)$
$max = maximum(Sp)$ ▷ (Find maximum element in Sp)
$maxIdx = argmax(Sp)$ ▷ (Find index of maximum element in Sp)
for i = 0 to size(Sp) **do**
    if $i = maxIdx$ **then**
        $MaxArray[i] = max$
    else
        $MaxArray[i] = 0$
    end if
end for
for t = 0 to size(Sp) **do**
    $v = 0$
    for j = 0 to k **do**
        v = v + real(MaxArray[j]) * np.cos(2 * np.pi * t * j / d) + imag(MaxArray[j]) * np.sin(2 * np.pi * t * j / d)
    end for
    $S' \leftarrow v$
end for
$Ret = S - S'$

4 Datasets

We use 2 sources of data for our experiments. Each dataset represents information about the number of people present in certain locations. The observations were carried out in two ways:

1. Using GPS coordinates (the number of people was determined at a certain point on the map, and later assigned to the place that lies closest to this point)

2. With accurate measurements at every location every hour

The measurements were taken in Thailand, Bangkok during couple of weeks. The collection of data took place through a mobile application that people used to send this data. Then all the data was entered into the tables. And then from these tables we take data for experiments.

Shops and offices were identified as the main location for observation. It is the following list of locations: 1. 7-eleven, Khlong Nueng, Khlong Luang District

2. Puey Library, Si Thammasat Rd, Khlong Nueng, Khlong Luang District

3. Green canteen, Khlong Nueng, Khlong Luang District

4. Khlong Luang Post Office ait, 58 Phahonyothin Rd, Khlong Nueng, Khlong Luang District

5. big C, 94 Phahonyothin Rd, Prachathipat, Thanyaburi District

6. ROBINSON, 94 Phahonyothin Rd, Prachathipat, Thanyaburi District

7. INDEX LIVING, 102 Phahonyothin Rd, Tambon Prachathipat, Thanyaburi District,

8. HOME PRO, 100 Phahonyothin Rd, Prachathipat, Thanyaburi District

9. Khon Luang cluster, Khlong Nueng, Khlong Luang District

10. Thai name place, 32/36, 13180 Khlong Nueng, Khlong Luang District

11. Tesco Lotus Khlong Luang, 101 Khlong Luang Rd, Khlong Nueng, Khlong Luang District

12. Makro, 39 Khlong Luang Rd, Khlong Song, Khlong Luang District

Each of the data collection methods are flawed and significantly affect the result. First of all, it is worth analyzing the disadvantage of the data collection method using GPS coordinates. For example, let's take two places that were being monitored: Index Living Mall and Big C Rangsit. As we can see in Figure 2, these two locations are very close to each other. This means that it can interfere with the further distribution of results. If we imagine that a person is standing in the place where the red dot is in Figure 1 then during the determination of his GPS coordinates, his location will be determined as the Living Index, because it is closer to him, but the person himself at this moment is inside Big C. This problem is not obvious and becomes clear when you start manually looking at the

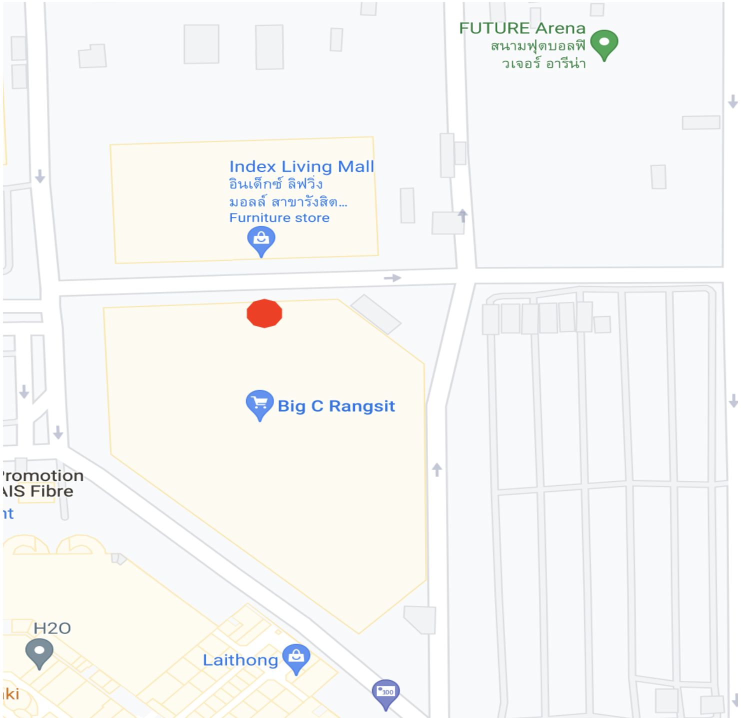

Figure 2: Map with Index Living Mall and Big C Rangsit

results of all observations. So in some locations, such a number of people can appear, which physically cannot be there. This means that this will affect the result of future predictions.

The disadvantage of collecting data using accurate observations is already more obvious, but it has less impact on the accuracy of the result when analyzing this data. The main disadvantage of this method is the difficulty in constantly monitoring all places and accurately tracking the number of people. Since every minute people move from one location to another, it is quite difficult to track the exact number of people at a certain time.

Due to the fact that each of these two methods of collecting data has its own disadvantages, we will consider them separately. All collected and filtered data about human presence are used as inputs for our models.

5 Experiments

The main task of the experimental evaluation is to evaluate the effectiveness of different predictive temporal algorithms to recommend the time of visiting a certain location based on past data, and then compare these methods to determine the most effective one.

There were two datasets for each location, the difference being mainly in the way they were collected. The first set was assembled based on GPS coordinates, while the other was based on accurate measurements. Each of these sets have been converted to a continuous time-dependent signal. An example of an input signal can be seen in the figure 3.

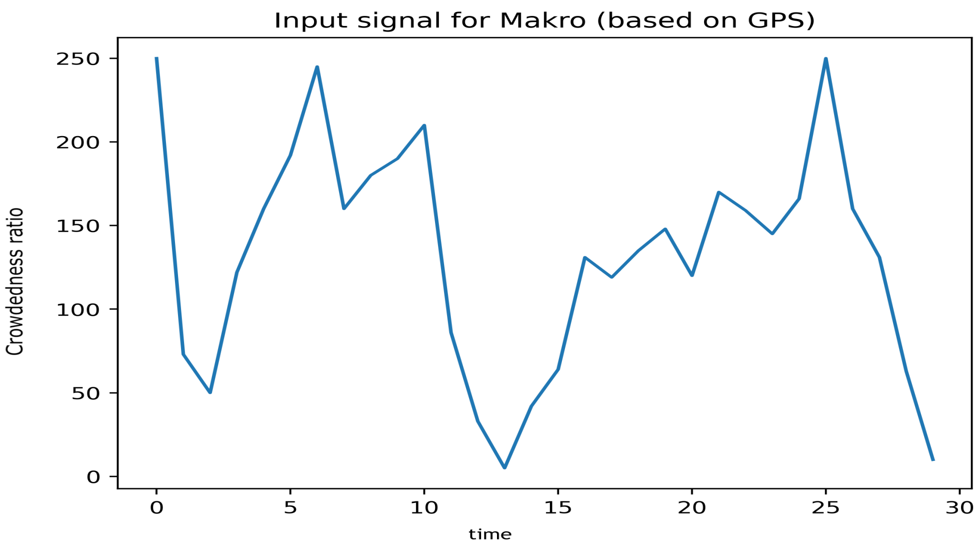

Figure 3: Example of input sequence

5.1 Strategies

In this section, a strategy is understood as how a person will plan his trips during the day. In our work, a person can be guided by three strategies:

1) **Standard** The person, who are guided by a standard strategy, draws up their plans by adapting to the standard behavior towards the population.

2) **Minimum** The person, who guided by recommended strategy, has some previous knowledge about amount of people in the location and by the simplest recommendation choose time for visiting.

3) **Recommended** The person, who guided by recommended strategy, has previous knowledge about amount of people in the location and tries to minimise the risk

5.2 Performance of minimum algorithm

Minimum algorithm was described in Algorithm 1. The principle of operation of this algorithm is quite simple. Based on past data, the algorithm recommends the hour of visit in which in the past there were the minimum number of people. For each place from our data training and testing sets were created for testing recommended time. Based on data in testing set, recommended time with minimum number of people was chosen. After that step, recommended time was used on training set. For all the places histograms were made (figure 4-6). On histograms there are testing and training sets for all shops in experiment.

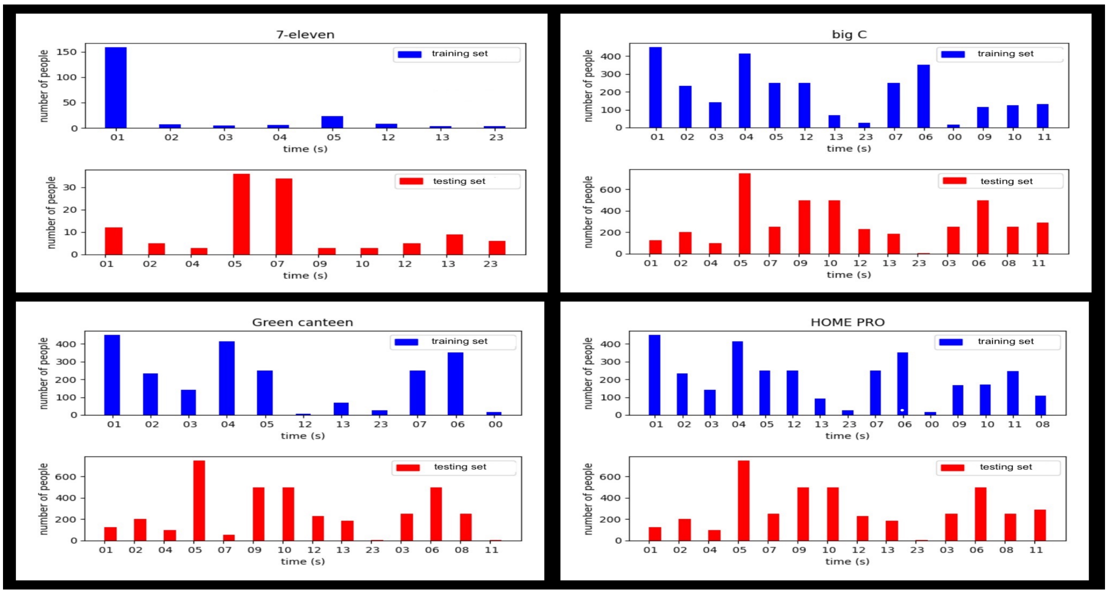

Figure 4: Histograms for shops 7-11, Green canteen, big C and HOME RPO

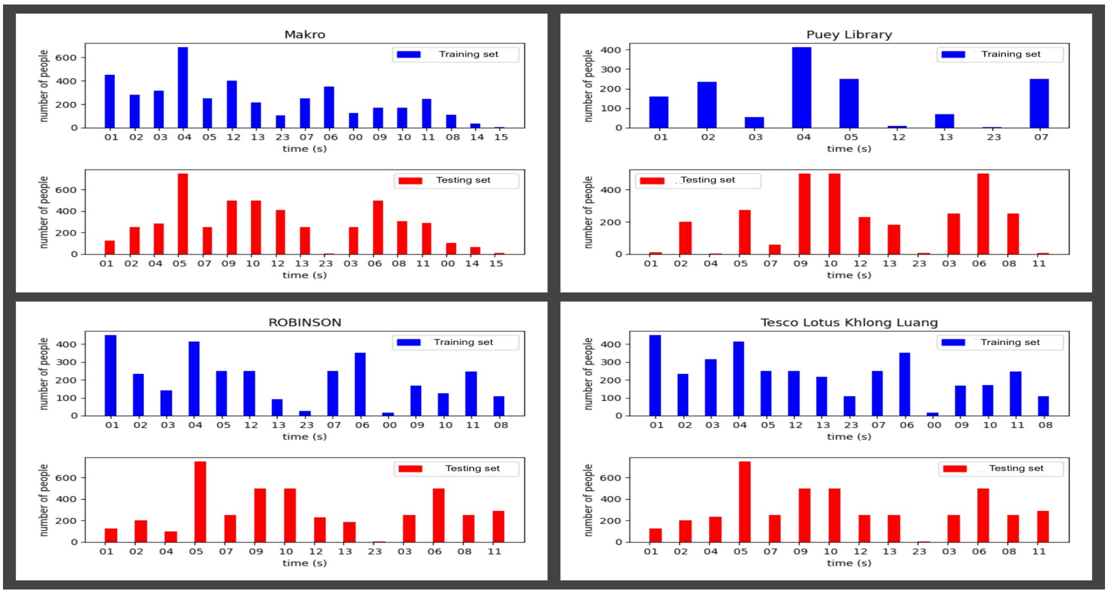

Figure 5: Histograms for shops Makro, Puey Library, Robinson, Tesco Lotus

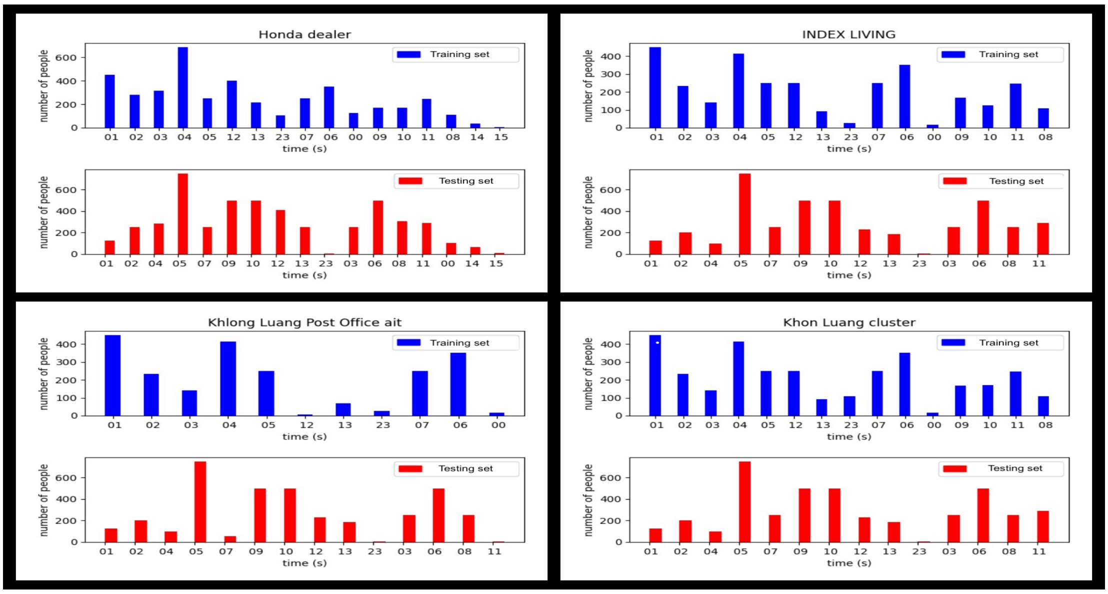

Figure 6: Histograms for shops Honda Dealer, Index Living, Khon Luang cluster

This algorithm is not predictive, since it does not make any assumptions about what may happen in the future, so it is difficult to call it sufficiently effective, however, such a recommender system may be considered as a working. The result of this algorithm can be seen in the table 1.

5.3 Performance of l AAM and l BAM

These algorithms are considered predictive and work on the basis of signal reconstruction. After the signal was generated, we applied the AAM and BAM algorithms to each of the input sequence. The output of each of these algorithms is a reconstructed signal. Further, for our experiments, we analyze each of the output sequence and find at what point the signal reaches its minimum. It is this minimum obtained on the time axis that we further consider the recommended time to visit the place for which this sequence was generated. A table with the recommended visiting hours for each place using different algorithms and sets of dates can be seen in the table 1.

Table 1: result of methods: recommended times to visit places

Method/place	7-11	Library	GC	bigC	Robinson	IndexLiving	HomePro	Market	Thai name	Tesco	Makro
AAM(GPS)	01	07	03	4	06	09	05	16	01	10	12
BAM(GPS)	04	07	07	12	06	09	12	13	01	01	02
AAM(observation)	08	11	07	17	16	19	09	13	-	14	10
BAM(observation)	08	11	08	13	16	19	09	07	-	15	10
Minimum algorithm	3	07	23	02	13	04	02	05	09	01	15

As a result of the operation of these algorithms, we received sequence that are quite similar to each other, which can be seen in the figures 7-12. This tells us about the similar operation of the two algorithms and that the reconstructed signal for both AAM and BAM looks similar.

We can also note that the result of the same algorithm for the same location for data collected using GPS and data collected using observations is quite different. This is the influence of the fact that different amounts of data were collected and their accuracy is also not high.

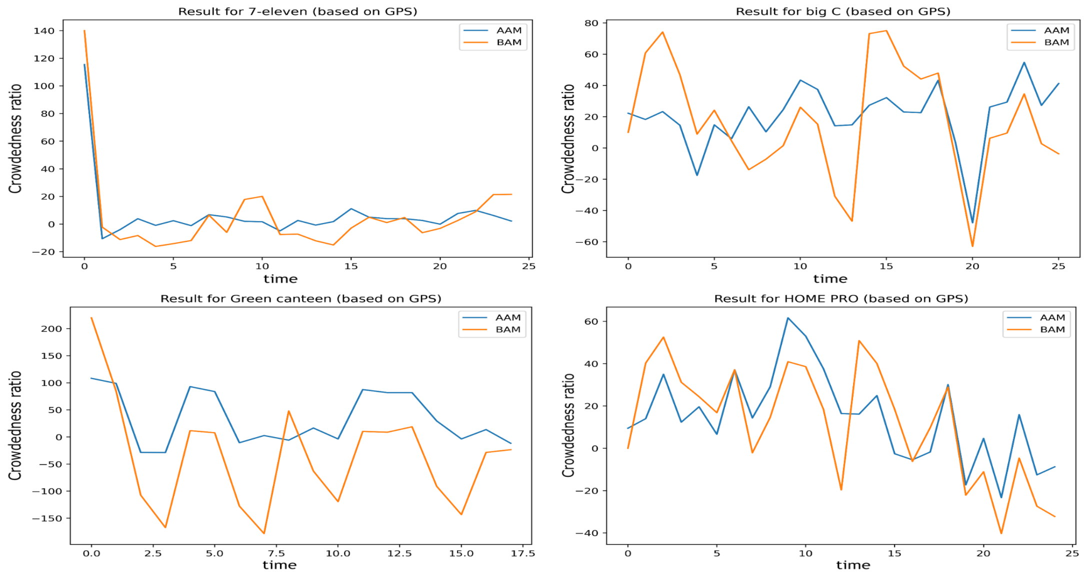

Figure 7: Result of AAM and BAM based on GPS dataset

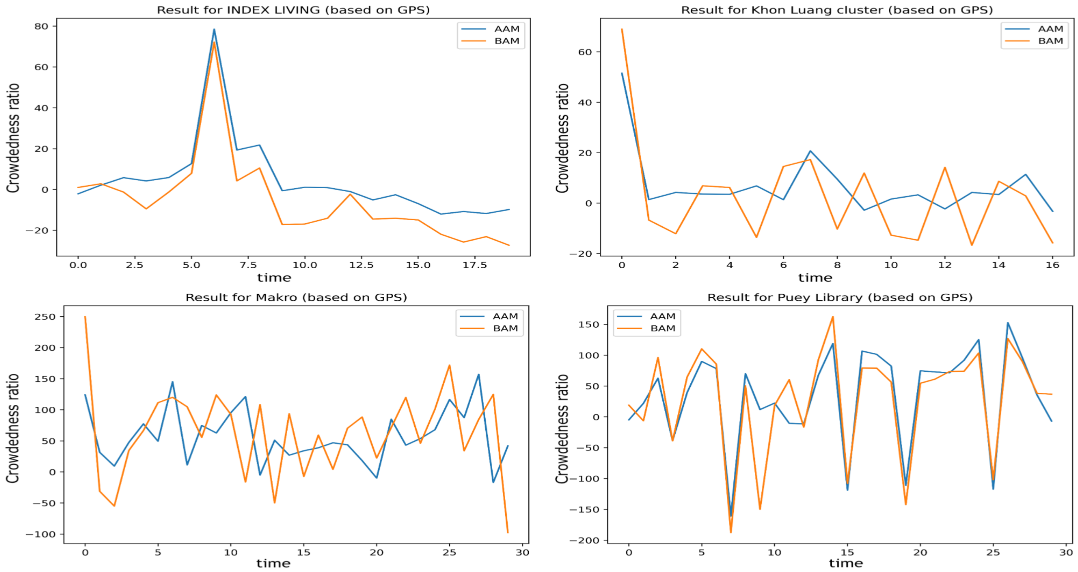

Figure 8: Result of AAM and BAM based on GPS dataset

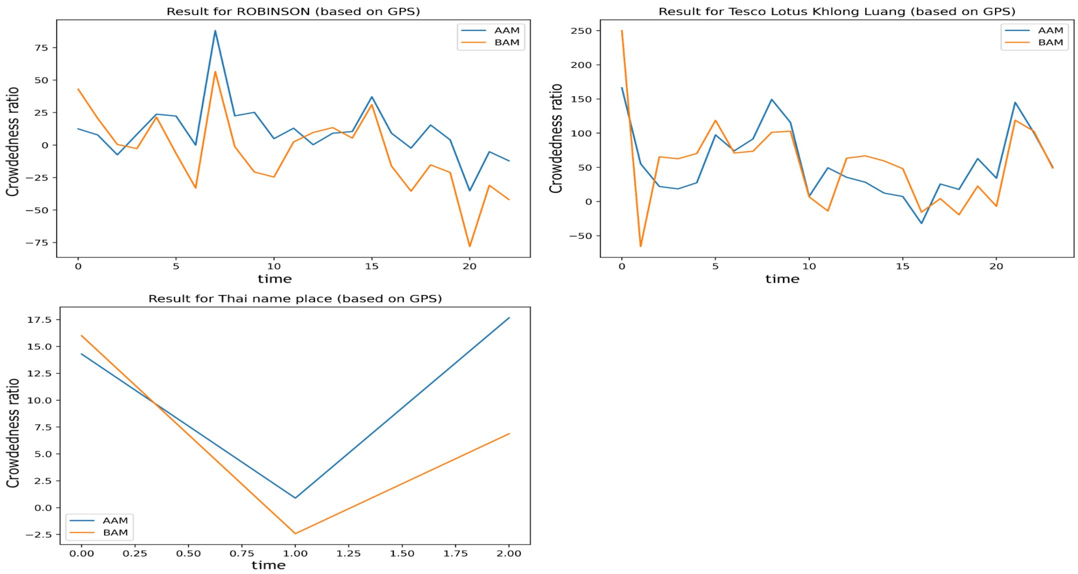

Figure 9: Result of AAM and BAM based on GPS dataset

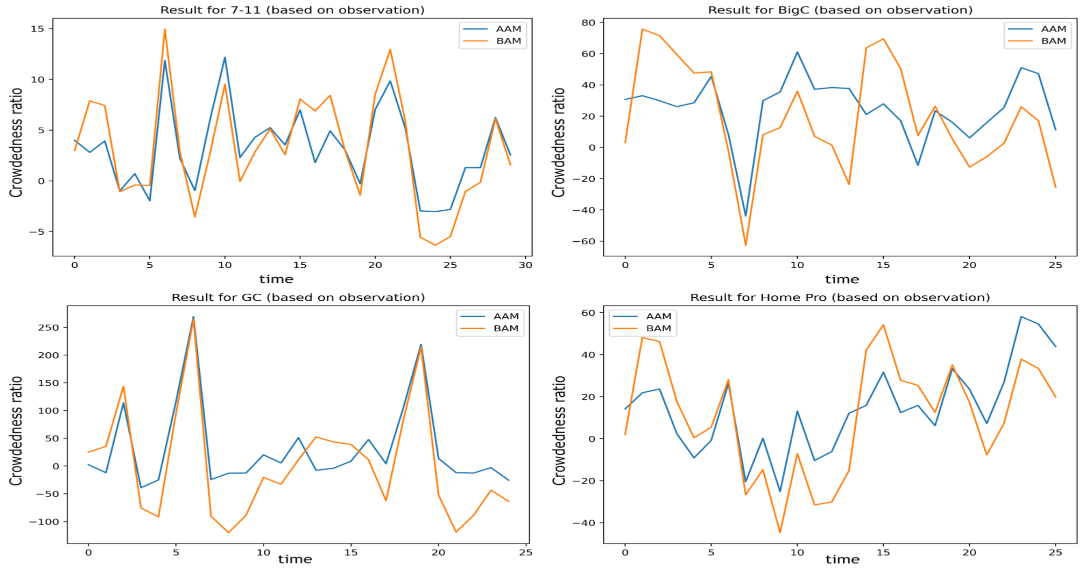

Figure 10: Result of AAM and BAM based on observation dataset

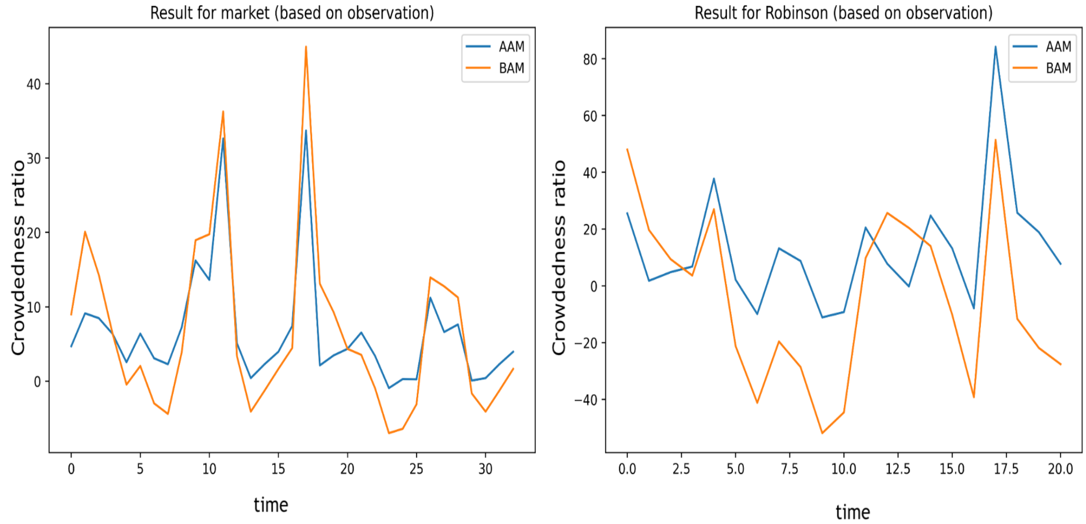

Figure 11: Result of AAM and BAM based on observation dataset

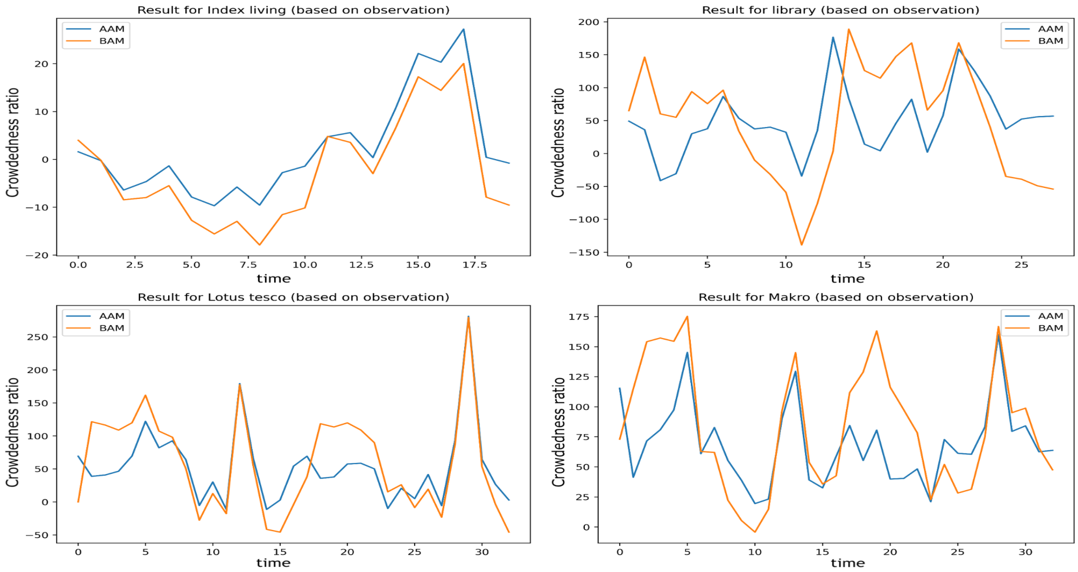

Figure 12: Result of AAM and BAM based on observation dataset

Table 2 shows us how many people a person will meet, depending on what he will be guided by when building his route. On this basis, we can conclude that it will be more efficient to use l AAM, but because it is predictive model we cannot completely rely on this result. That is why for each of the algorithms we will apply statistical hypothesis test, which is called Student's t-test.

Table 2: Number of people at the recommended time for each algorithm

Methods/place	7-11	Library	GC	bigC	Robinson	IndexLiving	HomePro	Market	Thai name	Tesco	Makro
l AAM(GPS)	4	57	118	72	61	6	48	23	16	139	160
l BAM(GPS)	3	57	140	100	61	6	25	14	16	1	119
l AAM(observation)	4	163	15	126	85	11	0	2	-	138	119
l BAM(observation)	4	163	81	74	85	11	0	38	-	130	119
Standart behaviour	13	122	251	146	115	8	105	3	0	139	282
Minimum algorithm	3	57	125	3	15	4	2	4	4	1	10

5.4 Comparison of methods by paired student's t-test

The paired t-test is also called dependent t-test shows us how significant the differences between two related units are. Typically, for this test we use the following two hypotheses:

1) Null hypothesis (H0): The mean difference between sample 1 and sample 2 is equal to 0.

2) Alternative hypothesis (H1): The mean difference between 1 and 2 objects is not equal to 0.

For paired t-test we use the formula below:

$$t = \frac{(m1 - m2) \cdot \sqrt{(n)}}{Sp} \tag{4}$$

where:

m1 and m2 - The average values of each of the sample sets,

n - The sample size,

Sp - The standart deviation of the differences of the paired data values.

The t score, which is calculated in formula above, is a ratio between the difference between two groups. The larger the t score, the more difference there is between groups. Every t-value has a p-value to go with it. A p-value is the probability that the results from sample data occurred by chance. Low p-values are good, because it indicates our data did not occur by chance. After all calculations, based on the probability obtained using the t-test, we can reject the null hypothesis or accept it.

For our experiments we say, that a p-value of minimum 0.05 is accepted to mean the data is valid.

The first step was a comparison between the results of the algorithms and the original data. Tests showed that AAM and BAM gave better results compared to the input sequence. Then, to carry out the test, we took the sequences that were the result of the work of the AAM and BAM for 11 places in which the data were collected using the GPS. The test showed that in 8 cases out of 11 the result will be the same, which means that the efficiency of their work in these 8 cases is approximately equal. However, in the remaining 3 cases, the test showed that the AAM algorithm is more efficient and gives the average sequence value less than the BAM, which means that a person who uses the recommendation given with the AAM method will meet fewer people on his way.

The next step was a comparison for the sequences that were the result of the work of AAM and BAM for 11 places in which data were collected using observations. The test showed that in 7 cases out of 10 the algorithms give, on average, a very similar result. But in the remaining 3 cases, it was found that AAM works more effectively.

6 Conclusion

The main purpose of this thesis was to study and test methods that allow to combine the public data and crowd sources to improve the quality of information to the temporal modeling techniques. Experiments were carried out on the real data provided to us. We have implemented two main models considered in this work. The first method was Additional Amplitude Model (AAM), the second method was Best Amplitude Model (BAM), both methods work on the principle of reconstruction of the input signal. Applying these two methods to real data, We were able to find an estimated time at which the least number of people would be in this place. After getting the result of these methods, we compared the effectiveness of each of them using a paired t-test. The results of the experiment comparing AAM and BAM showed us that despite the fact that both algorithms work with approximately the same efficiency, nevertheless, in several tests it was found that AAM gives more satisfactory results. So we can say that guided by this algorithm, a person will be able to better avoid unnecessary social contacts.

References

[1] https://symptomate.com.

[2] https://www.data4life.care/en/data4life-app/.

[3] https://erouska.cz.

[4] https://www.ravemobilesafety.com/covid-19-coronavirus-health-monitoring-app.

[5] https://www.ravemobilesafety.com/covid-19-coronavirus-health-monitoring-app.

[6] https://www.feevyapp.com.

[7] https://covid19.who.int.

[8] https://covidcontrol.jhu.edu.

[9] https://chronorobotics.fel.cvut.cz/cs/ai-vs-covid.

[10] https://mdlnext.mdlive.com.

[11] https://www.teladochealth.com.

[12] Ministry of health of the czech republic. https://www.mzcr.cz/category/uredni-deska/rozhodnuti-ministerstva-zdravotnictvi/aktualni-mimoradna-a-ochranna-opatreni-ke-covid-19/. *In review for Chaos, Solitons Fractals*, 2021.

[13] et.al. Ferdian Jovan. "a poisson-spectral model for modelling temporal patterns in human data observed by a robot". *IInternational Conference on Intelligent Robots and Systems (IROS)*, 2016.

[14] Tomáš Krajník. Crowd forecasting for social distancing during pandemics. *In review for Chaos, Solitons Fractals*, 4(4):11–14, 2020.

[15] et.al. Martín-Calvo D. "effectiveness of social distancing strategies for protecting a community from a pandemic with a data-driven contact network based on census and real-world mobility data.". *In review for Chaos, Solitons Fractals*, 2020.

[16] TM. Templeton. “time-series analysis of variable star data,”. *Journal of the American Association of Variable Star Observers (JAAVSO)*, 32(1):41–54, 2004.

[17] et.al. Tomáš Krajník, Jaime Pulido Fentanes. "fremen: Frequency map enhancement for long-term mobile robot autonomy in changing environments". *IEEE Transactions on Robotics*, 33(4):964–977, 2017.

Appendix

Contents of the enclosed CD

```
│ readme.txt ........ the file with CD contents description
├─ thesis.pdf ........ the thesis text in PDF format
└─ bp_en ........ the directory of source codes
   ├─ main.py ........ main file
   ├─ Aam.py ........ implementation of AAM algorithm
   └─ Bam.py ........ implementation of BAM algorithm
```