Modeling and Probabilistic Reasoning of Population Evacuation During Large-scale Disaster

Xuan Song, Quanshi Zhang, Yoshihide Sekimoto, Teerayut Horanont,
Satoshi Ueyama and Ryosuke Shibasaki
Center for Spatial Information Science, The University of Tokyo, Japan
songxuan@csis.u-tokyo.ac.jp

ABSTRACT

The Great East Japan Earthquake and the Fukushima nuclear accident cause large human population movements and evacuations. Understanding and predicting these movements is critical for planning effective humanitarian relief, disaster management, and long-term societal reconstruction. In this paper, we construct a large human mobility database that stores and manages GPS records from mobile devices used by approximately 1.6 million people throughout Japan from 1 August 2010 to 31 July 2011. By mining this enormous set of Auto-GPS mobile sensor data, the short-term and long-term evacuation behaviors for individuals throughout Japan during this disaster are able to be automatically discovered. To better understand and simulate human mobility during the disasters, we develop a probabilistic model that is able to be effectively trained by the discovered evacuations via machine learning technique. Based on our training model, population mobility in various cities impacted by the disasters throughout the country is able to be automatically simulated or predicted. On the basis of the whole database, developed model, and experimental results, it is easy for us to find some new features or population mobility patterns after the recent severe earthquake, tsunami and release of radioactivity in Japan, which are likely to play a vital role in future disaster relief and management worldwide.

Categories and Subject Descriptors

H.4 [**Database Management**]: Data mining; H.4 [**Database Management**]: Spatial databases and GIS

Keywords

human mobility, disaster informatics, data mining

1. INTRODUCTION

The 9.0 magnitude Great East Japan Earthquake [1] occurred on 11 March 2011 off the east coast of Honshu, Japan's largest island. Since modern record keeping began in 1900,

this is considered the most powerful earthquake to have occurred in Japan and is one of the five most powerful historical earthquakes worldwide [1]. The Japanese National Police Agency confirmed 15,881 deaths [2], 6,142 injured [2], and 2,668 missing persons as well as 128,801 buildings damaged or destroyed across 18 prefectures [2]. The earthquake and the tsunami that followed severely damaged the Fukushima Daiichi nuclear power plant, causing the most extensive release of radioactivity since the 1986 Chernobyl accident in the former Ukraine [3]. The three separate events (earthquake, tsunami, and release of radioactive materials) created an unprecedented composite disaster that significantly impacted the people of Japan. In the wake of a disaster of this magnitude, *it is very important to record people's movements following the event to analyze their behavioral patterns and to develop simulation or predictive models for future disaster mitigation. The types of data, evacuation behavior pattern, and simulation model available following the Great East Japan Earthquake and the Fukushima Daiichi nuclear power plant meltdown are unique in human history, and are likely to play a vital role in future disaster relief and management worldwide.*

Therefore, in this paper, *we constructed a population mobility database that stored and managed daily GPS records from approximately 1.6 million individuals throughout Japan over one year.* By mining the whole database, we attempted to analyze and model evacuation behaviors of people during the Great East Japan Earthquake and ensuing Fukushima nuclear accident. Meanwhile, we developed a probabilistic inference model to effectively represent people's mobility patterns during the disaster to better understand human evacuation behaviors in general, and to understand how those behaviors were impacted by various city states under disasters. Moreover, based on our training model, *our objective was to simulate or predict population mobility in various cities throughout Japan in an effort to inform future disaster relief and management (as shown in Fig.1).*

The remainder of this paper is structured as follows: In the following section, we will introduce the evacuation behaviors discovery. Section 3 and 4 provide the details about evacuation behavior model learning and inference. Experimental results are presented in Section 5. Related work is briefly reviewed in Section 6, and the paper is finally summarized in Section 7.

2. EVACUATION BEHAVIORS DISCOVERY

The Great East Japan Earthquake and Fukushima nuclear accident cause large population movements or evacuations.
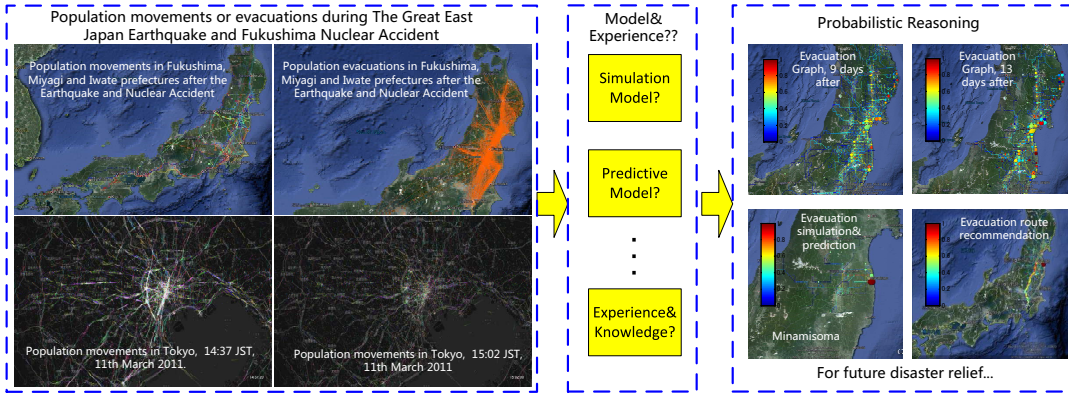
Figure 1: What kinds of experiences or model can we learn from the unprecedented composite disaster of Japan in 2011? The Great East Japan Earthquake and the Fukushima nuclear accident cause large human population movements and evacuations, and we constructed a human mobility database to record these disaster behaviors. Based on these data, can we learn some experiences, simulation or predictive models for future disaster relief and management worldwide?

Obviously, analysis of these movements and evacuations in short-term and long-term will play a vital role for planning effective humanitarian relief and disaster management in the future. In this section, we present the details relating to the discovery of people's evacuation behavior through enormous set of Auto-GPS mobile sensor data.

2.1 Geographic Location Distribution

For each person, the history geographic location are a series of geographic position with its longitude, latitude and time period. Let $X_k(t, T_{\text{period}}) = \{\mathbf{p}_k(t, d) : d \in T_{\text{period}}\}$ denotes the history geographic locations of person k in time t ($t \in T_{\text{time}}$) during period T_{period}, where $\mathbf{p}_k(t, d)$ is its geographic position in time t of day d, its geographic location distribution is represented by κ bins distribution by:

$$L_k(t, T_{\text{period}}) = \{\psi(n; \sigma(\mathbf{p}_k(t, d)), d \in T_{\text{period}})\}_{n=1,...,\kappa}, \quad (1)$$

where $\sigma(\mathbf{p}_k(t, d))$ is a function that computes the bin index associated with the geographic location, $\psi(n; \sigma(\mathbf{p}_k(t, d)), d \in T_{\text{period}})$ denotes the probability of person k will appear in location index n at time t during the period T_{period}. The example of this distribution is shown in Fig.2, and the size of the circles indicates the probability of staying location of an individual person at a specific time; larger circles indicate higher probability that persons stay or live there. Blue and orange circles indicate this distribution before and after the earthquake respectively. Based on this distribution, it is easy for us to find some regular and important places for each people, such as home, working area and etc.

2.2 Similarity Coefficient Computation

To discover the evacuations during the disasters, we can compute the similarity of geographic location distribution for individual person before and after the earthquake respectively, and if the similarity is small enough, this person's behavior is able to be seem as an evacuation (as shown in Fig.2). Here, we utilize the Jaccard Coefficient [4] to measure the similarity, and the coefficient α for before earthquake period $T_{\text{period}}^{\text{before}}$ and after earthquake one $T_{\text{period}}^{\text{after}}$ is able

to be calculated as follows:

$$\alpha_k(T_{\text{period}}^{\text{before}}, T_{\text{period}}^{\text{after}}) =$$
$$\frac{1}{\|T_{\text{time}}\|} \sum_{T_{\text{time}}} \frac{\sum_{n=1}^{\kappa} \min\{\psi(n, T_{\text{period}}^{\text{before}}), \psi(n, T_{\text{period}}^{\text{after}})\}}{\sum_{n=1}^{\kappa} \max\{\psi(n, T_{\text{period}}^{\text{before}}), \psi(n, T_{\text{period}}^{\text{after}})\}}$$
$$(2)$$

Therefore, based on this similarity, people's evacuations in short-term or long-term is able to be automatically discovered by mining the whole Auto-GPS mobile sensor database. Some examples of the discovered evacuations during the disaster are shown in Fig.3-A, and all these discovered evacuations form a sample dataset for further training or testing.

3. INFERENCE MODEL DEVELOPMENT AND LEARNING

To understand, simulate and predict human mobility during severe disasters, we need a concise model to effectively represent these discovered evacuations or movements. In this research, we develop a probabilistic inference model and utilize the discovered human evacuations to train its parameters via machine learning technique. We assume in this paper that our collected Auto-GPS mobile sensor data are representative of the general population movements during the disasters. The creation of this type of inference model is possible because social interactions, transportation networks, and political responses in some given cities (except for some highly destroyed cities) are typically stable through time, and large population movements (which are often influenced by these conditions) are likely to remain the same following different disasters.

To generate the probabilistic inference model, there are mainly two stages: evacuation graph construction and inference model learning (as shown in Fig.3 and Fig.4). Firstly, we construct the people's evacuation graph in some affected regions based on the geographic location distribution and the training samples of people's evacuations, and its node parameters and edge connections are able to be learned by *collaborative learning* [5]. Secondly, given the learned evacuation graph, we develop the reasoning model based on the Markov Decision Process (MDPs) [6], and utilize *Inverse*
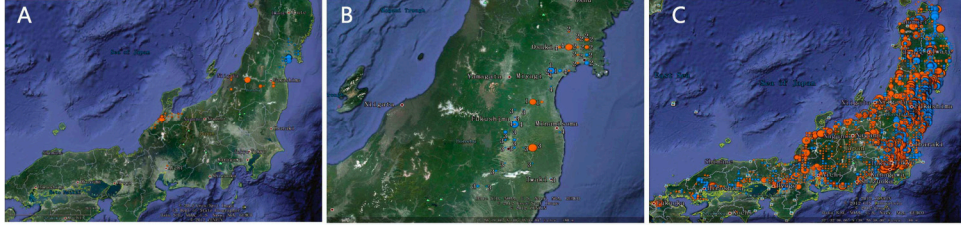
Figure 2: Examples of people's geographic location distribution before and after the earthquake. The size of the circles indicates the probability of staying location of an individual person at a specific time; larger circles indicate higher probability that persons stay or live there. Blue and orange circles indicate this distribution before and after the earthquake respectively. Fig.A shows geographic location distribution of a single person. Fig.B and Fig.C show the four persons case and multiple persons' case respectively.

Reinforcement Learning [7, 8] to train its parameters. In this section, we present the further details relating to this probabilistic inference model development and learning.

3.1 Evacuation Graph Construction

3.1.1 Region Construction

To construct the evacuation graph, we firstly need to discover connected geographical areas during disasters with the discovered evacuations (as shown in Fig.3). Here, we define people's evacuations as follows:

Definition 1. (Evacuation): Let $E_k = \{\mathbf{p}_k^b, \mathbf{p}_k^1, ...\mathbf{p}_k^n, \mathbf{p}_k^a\}$ denotes the evacuation of person k, where $\mathbf{p}_k^b$ and $\mathbf{p}_k^a$ are its major residential location before and after the earthquake respectively, and they are the highest probability location of person k based on its geographic location distribution. $\{\mathbf{p}_k^1, ...\mathbf{p}_k^n\}$ are the observed movements of person k during the disasters (as shown in Fig.3-A).

We divide the geographical range into disjoint cells by a given cell length l. The set of the cells is denoted as c, and the its grid ID is represented by (x, y). Thus, the specific position of the evacuation is able to be mapped into a cell, and overall evacuation is transformed into a sequence of cells as $\{c_{\mathbf{p}_k^b}, c_{\mathbf{p}_k^1}, ..., c_{\mathbf{p}_k^n}, c_{\mathbf{p}_k^a}\}$, where $c_{\mathbf{p}_k^n}$ represents the cell that $\mathbf{p}_k^n$ locates in. Given two cells $c(x, y)$ and $c'(x', y')$, the c and c' are called *spatially close* if $|x - x' \leq 1|$ and $|y - y' \leq 1|$.

To explore connected geographical regions, we need to define correlated relations among evacuations. Given a cell $c = c_{\mathbf{p}_k^n}$, for some $\mathbf{p}_k^n \in E_k$, we say that evacuation E_k traverses cell c , denoted $c \in E_k$. If E_k traverses from cell c to c', it is represented by $c \to c' \in E_k$, and the travel time is represented by $\lambda(c \to c', E_k)$. Thus, the spatio-temporally correlated relation (st-correlated) of evacuations is defined as:

Definition 2. (Spatio-temporally Correlated Relation): Given two evacuations E_i and E_j, and a temporal constraint θ, their sub-trajectories $tra_i' = \{\mathbf{p}_i^n, ...\mathbf{p}_i^{n'}\}$ and $tra_j' = \{\mathbf{p}_j^n, ...\mathbf{p}_j^{n'}\}$, where $tra_i' \subseteq E_i$, $tra_j' \subseteq E_j$, are correlated if
(1) $\exists \Delta t_1 \in \lambda(c_{\mathbf{p}_i^n} \to c'_{\mathbf{p}_i^{n'}}, E_i), \Delta t_2 \in \lambda(c_{\mathbf{p}_j^n} \to c'_{\mathbf{p}_j^{n'}}, E_j)$
$s.t. \frac{|\Delta t_1 - \Delta t_2|}{max\{\Delta t_1, \Delta t_2\}} \leq \theta$;
(2) one of the two rules is satisfied:
Rule1: $c_{\mathbf{p}_i^n}$ and $c_{\mathbf{p}_j^n}$ are spatially close, and $c'_{\mathbf{p}_i^{n'}} = c'_{\mathbf{p}_j^{n'}}$.
Rule2: $c_{\mathbf{p}_i^n} = c_{\mathbf{p}_j^n}$, and $c'_{\mathbf{p}_i^{n'}}$ and $c'_{\mathbf{p}_j^{n'}}$ are spatially close.

In addition, we define the connection support as follows:

Definition 3. (Connection support): Given a set of cells C, two specific cells $c, c' \in C$, where they are spatially close, the connection support of the cell pair (c, c') is defined as $|Q_1 \cup Q_2|$, and $Q_1 = \{(tra_i, tra_j)\}$, where its sub-trajectories tra_i' and tra_j' are st-correlated, and $c \to c'' \in tra_i'$, $c' \to c'' \in tra_j'$, for some $c'' \in C - \{c, c'\}$, and $Q_2 = \{(tra_i, tra_j)\}$ where its sub-trajectories tra_i' and tra_j' are st-correlated, and $c'' \to c \in tra_i'$, $c'' \to c' \in tra_j'$, for some $c'' \in C - \{c, c'\}$.

Hence, the neighbor cells and enclosed cells are able to be defined as:

Definition 4. (Neighbor cells): Given a set of cells C, two specific cells $c, c' \in C$, and a minimum connection support K, if the connection support of the cell pair (c, c') is greater than or equal to K, c and c' are neighbor cells, denoted as $neighbor(c, c')$.

Based on above definitions, the region as a connected geographical areas and the enclosed cell are defined as:

Definition 5. (Region): Given a set of cells C, it forms a region if for any two cells $c, c' \in C$, there exists a chain of cells $(c =)c_1 = c_2 = ... = c_n$, s.t. $neighbor(c_i, c_{i+1})$ for each $c_i \in C$ and $i \in [1, n]$.

Definition 6. (Enclosed cells): Given a set of cells C and a cell $c \in C$, the c is enclosed if there exists a region $r \subseteq C$, s.t. $c, c' \in r$, $\forall c' \in \{c' | c'$ and c are spatially close, $c' \in C\}$.

Hence, to construct the regions, we can generate all cell pairs from the set of cells C and then compute the connection support of each cell pair by checking other cells in C. We then verify whether the connection supports of cell pairs satisfy the given minimum connection support K to construct regions (as shown in Fig.3-B). Moreover, the computational cost is able to be reduced by building up a *grid index* [5] in which each cell has a unique grid ID, a value $grid(c)$, and a corresponding evacuations list, where $grid(c) = |\{E | c \in E, E \in D\}|$. Therefore, the overall algorithm of region construction is detailed in Algorithm 1 and Algorithm 2.

3.1.2 Edge Inference

Once the nodes in the evacuation graph are generated, we then need to infer edges and derive some edge information,
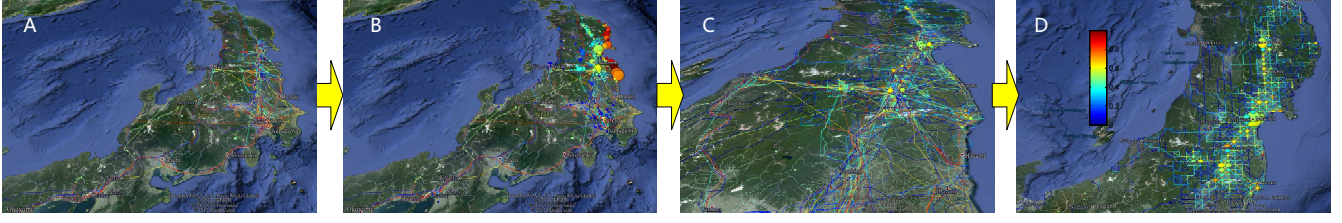
Figure 3: Evacuation graph construction. Given the discovered evacuations (Fig.A), we constructed some important regions as the nodes for the graph (Fig.B). Then, we utilized these evacuations traversing the regions to derive edge connections (Fig.C). The final evacuation graph was illustrated in Fig.D.

Algorithm 1: Region Construction

Input: The discovered evacuation dataset D, a set of cells C, a temporal constraint θ, and a minimum connection support K

Output: A set of regions R.

Initialization: $c' \leftarrow$ Sort cells in c in a decreasing order of $grid(c)$.

Do

1. $c \leftarrow$ Pop the cell from c';

2. **for each** E traversing c by the order stored in the grid index

    (a) $\tau(c) \leftarrow \{\mathbf{p}|c_{\mathbf{p}} = c, \mathbf{p} \in E\}$;

    (b) **for each** $\mathbf{p} \in E - \tau(c)$ and $c_{\mathbf{p}}$ is not enclosed

        i. **If** $c_{\mathbf{p}}$ is contained in some regions
          $r \leftarrow$ The region contains $c_{\mathbf{p}}$;

        ii. **Else**
          $r \leftarrow \phi$;

        iii. **If** $\mathbf{p}$ is before $\mathbf{p}'$ for all $\mathbf{p}' \in \tau(c)$
          $r \leftarrow CM(r, c_{\mathbf{p}}, c, \theta, K, Rule1)$;

        iv. **Else if** $\mathbf{p}$ is after $\mathbf{p}'$ for all $\mathbf{p}' \in \tau(c)$
          $r \leftarrow CM(r, c_{\mathbf{p}}, c, \theta, K, Rule2)$;

        v. **Else**
          $r \leftarrow CM(r, c_{\mathbf{p}}, c, \theta, K, Rule1)$;
          $r \leftarrow CM(r, c_{\mathbf{p}}, c, \theta, K, Rule2)$;

        vi. $R \leftarrow R \cup \{r\}$;

    (c) **end**

3. **end**

Until c' is empty or each cell in c is in some $r \in R$;
Return R;

Algorithm 2: Cell Merging (CM)

Input: A region r, a cell $c_{\mathbf{p}}$, a cell c, a temporal constraint θ, a minimum connection support K, and and indicator I.

Output: A region r.

Initialization: $c_{\mathbf{p}} = (x, y)$;

    1. **for each** cell $c' = (x', y') \in C - r$, where $|x' - x| \leq 1$ and $|y' - y| \leq 1$

        (a) **If** I is Rule1
          Verify whether $neigobor(c_{\mathbf{p}}, c')$ is held with given c by rule 1;

        (b) **Else**
          Verify whether $neigobor(c_{\mathbf{p}}, c')$ is held with given c by rule 2;

        (c) **If** $neigobor(c_{\mathbf{p}}, c')$

          i. **If** $r = \phi$
            $r \leftarrow \{c_{\mathbf{p}}\}$;

          ii. **If** c' is contained in some region
            $r \leftarrow$ The region contains c';

          iii. **Else**
            $c' \leftarrow \phi$

          iv. $r' \leftarrow CM(r', c', c, \theta, K, I)$;

          v. $r \leftarrow r \cup r'$

        (d) **end**

    2. **end**

Return r;

---

such as travel frequency, travel time and etc. In this research, the evacuation graph is a directed graph $G = (V, E)$, where V is a set of vertices and E is a set of edges. Each vertex v represents a geographical area, and the directed edge e indicates a transition relationship, including travel frequency and travel time.

Given the constructed regions R, and the discovered evacuation dataset D, we utilize these evacuations traversing the regions to derive edge connections and information within regions. For each evacuation traversing the region, we infer the shortest path between any two consecutive points of the evacuation by virtual bidirected edges in the region, and the travel time of each edge is estimated by the median of all the travel times of the edge. In addition, the travel frequency of each edge is able to be estimated by recording the number of traversing evacuations. Similarly, we can also generate edges between regions: if some evacuations traversers from

one region to another region, the edges is constructed between the two regions (as shown in Fig.3-C), and its edge information is estimated by the same methods as previous discussions.

3.2 Model Learning

Based on the constructed evacuation graph, the inference model is able to be developed by using the Markov Decision Precess (MDPs). MDPs provide a natural framework for representing sequential decision making, such as evacuating through various of geographical areas. In MDP theory, the agent takes as sequence of *actions* ($a \in A$), which transition between *states* ($s \in S$) and incur and action-based *cost* ($c(a) \in \Re$). The agent is trying to minimize the sum of costs while reaching some destination, and the sequence of action is called a *path* ζ. For MDPs, a set of *features* ($\mathbf{f}_a \in \Re$) characterize each action, and the cost of the action is a linear function of these features parameterized by a *cost weight* vector ($\boldsymbol{\phi} \in \Re$). Path feature, $\mathbf{f}_\zeta$ are the sum of the features of actions in the path: $\sum_{a \in \zeta} \mathbf{f}_a$. Thus, the cost
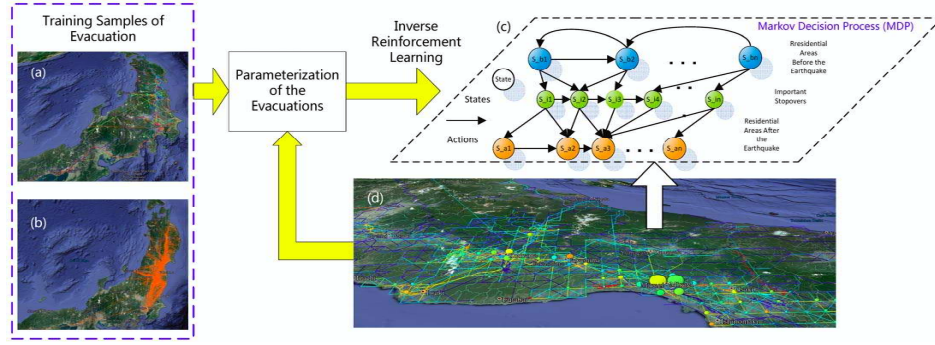
Figure 4: Inference model learning. Based on the constructed evacuation graph, the inference model was able to be developed by using the Markov Decision Precess (MDPs). The evacuation graph provided us a deterministic MDP, the geographical region (nodes) was able to be seem as *state*, the edge was the *action*, and the *path* was the parameterized evacuations by their path feature. We utilized the *Inverse Reinforcement Learning* to train the overall inference model.

weight applied to the path features is:

$$cost(\zeta|\phi) = \sum_{a \in \zeta} \phi^\top \mathbf{f}_a = \phi^\top \mathbf{f}_\zeta \qquad (3)$$

In our problem, the evacuation graph provide us a deterministic MDP, the geographical region (nodes) is able to be seem as *state*, the edge is the *action*, and the *path* is the people's evacuation (as shown in Fig.4). These evacuations are parameterized by their path feature $\mathbf{f}_\zeta$. For instance, a person's evacuation can be described by: evacuating through region A ($d = 25km$, $type = small$) to region B ($d = 55km$, $type = midle$), and finally stayed in region C ($d = 65km$, $type = large$) with route 1 ($frq = 0.53$, $time = 0.72$) ($A \rightarrow B$) an route 2 ($frq = 0.23$, $time = 0.52$)($B \rightarrow C$), where d is the distance from earthquake or nuclear power plant position, $type$ the city types, frq the travel frequency of the route, $time$ is the travel time of the route, and etc. Hence, we need to utilize all the discovered evacuations to train a MDPs model that is able to optimally demonstrate these people's behavior during the disasters. Obviously, this is an *Inverse Reinforcement Learning* problem. In this study, we utilize the *Maximum Entropy Inverse Reinforcement Learning* algorithm [7, 8] to train the overall simulation model.

Based on the *Maximum Entropy Principle*, the path distribution is able to be defined as:

$$P(\zeta|\phi) = \frac{e^{-cost(\zeta|\phi)}}{\sum_{path\zeta'} e^{-cost(\zeta'|\phi)}}. \qquad (4)$$

Hence, the cost weight vector ϕ from demonstrated behavior is learned by maximizing the entropy of the distribution over paths subject to the feature constraints from the discovered evacuations, and it implies that we maximize the likelihood of the observed data under the maximum entropy distribution as:

$$\phi^* = \arg\max_\phi L(\phi) = \sum_D \log P(\zeta_i|\phi) \qquad (5)$$

This function is convex for deterministic MDPs [8] and the optima can be obtained using gradient-based optimization methods. The gradient is the difference between expected empirical feature counts and the learner's expected feature

---

Algorithm 3: Expected Action Frequency Calculation

Input: Cost weight ϕ, initial state s_o, and goal state s_g.

Output: Expected action visitation frequencies $G_{a_{i,j}}$.

Backward Pass

1. Set $Z_{s_i} = 1$ for valid goal states, 0 otherwise;

2. Recursively compute for T iterations
$$Z_{a_{i,j}} = e^{-cost(a_{i,j}|\phi)} Z_{s:a_{i,j}}$$
$$Z_{s_i} = \sum_{a_{i,j} \text{ of } s_i} Z_{a_{i,j}} + 1$$

   Forward Pass

3. Set $Z'_{s_i} = 1$ for valid goal states, 0 otherwise;

4. Recursively compute for T iterations
$$Z'_{a_{i,j}} = e^{-cost(a_{i,j}|\phi)} Z'_{s_i}$$
$$Z'_{s_i} = \sum_{a_{i,j} \text{ to } s_i} Z'_{a_{j,i}} + 1$$

   Summing frequencies

5. $G_{a_{i,j}} = \frac{Z'_{s_i} e^{-cost(a_{i,j}|\phi)} Z_{s_i}}{Z_{s_{\text{initial}}}}$

---

counts, which can be expressed by:

$$\bigtriangledown L(\phi) = \widetilde{\mathbf{f}} - \sum_\zeta P(\zeta_i|\phi)\mathbf{f}_{\zeta_i} = \widetilde{\mathbf{f}} - \sum_a G_a\mathbf{f}_a, \qquad (6)$$

where G_a is the expected action visitation frequencies, and is able to be computed by enumerating all paths and probabilistically count the number of paths and times in each path the particular state is visited as Algorithm 3 and Eq.(7).

Algorithm 3 employs a more tractable approach by finding the probabilistic weight of all paths from the origin s_o to a specific action a, $Z'_a = \sum_{\zeta_{o \rightarrow a}} e^{-cost(\zeta)}$, all paths from the action a to the goal g, $Z_a = \sum_{\zeta_{a \rightarrow g}} e^{-cost(\zeta)}$, and all paths from the origin to the goal, $Z_o = Z'_g = \sum_{\zeta_{o \rightarrow g}} e^{-cost(\zeta)}$. Hence, the expected action visitation frequencies G_a is com-
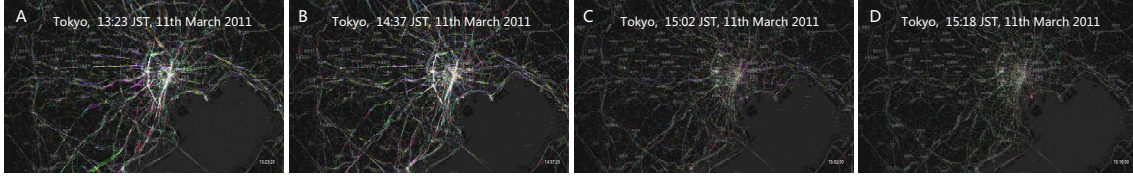
Figure 5: Visualization of population mobility during the earthquake. This figure shows population mobility of Greater Tokyo Area during the earthquake. The color denotes the directions of population mobility. Fig. A and B show population mobility before the earthquake, and Fig. C and D show the ones after the earthquake. We can clearly see that the whole transportation of Greater Tokyo Area was stopped suddenly during the earthquake (Fig C,D).

puted by:

$$G_a = \frac{Z_a Z'_a e^{-cost(a)}}{Z_o}. \qquad (7)$$

In summary, the inference model is able to be trained by finding the cost weight parameters with Eq.(5), Eq.(6), Algorithm 3 and Eq.(7) through the discovered evacuation dataset D. With this training model, the people's movement or evacuations during some future disasters is able to be easily simulated or predicted.

4. PROBABILISTIC INFERENCE AND REASONING

Based on the trained probabilistic inference model, people's evacuation or movements is able to be simulated or predicted for some similar disasters in the future. We utilize the Bayes's rule to perform this probabilistic inference: given the partial observed evacuations (such as some evacuations during first several hours after disasters), $\zeta_{A \to B}$, the posterior probability of the destinations is able to computed by:

$$P(\text{dest}|\zeta_{A \to B}, \phi) \propto P(\zeta_{A \to B}|\text{dest}, \phi)P(\text{dest}), \qquad (8)$$

where $P(\text{dest})$ is the evacuation prior in a region A, and it depends on the popular route inference [5] in the evacuation graph, and $P(\zeta_{A \to B}|\text{dest}, \phi)$ is likelihood, which is depended on:

$$P(\zeta_{A \to B}|\text{dest}, \phi) \propto \frac{\sum_{\zeta_{B \to dest}} e^{-cost(\zeta|\phi)}}{\sum_{\zeta_{A \to dest}} e^{-cost(\zeta|\phi)}}, \qquad (9)$$

and Eq.(9) is able to be easily inferred by taking the sums over paths from A to B to each possible destination using the forward pass of Algorithm 3.

Hence, the possible people's evacuation destination or route can be simulated or predicted by the *Maximum a Posteriori* (MAP) estimation of Eq.(8).

5. EXPERIMENTS AND RESULTS

5.1 Database and Visualization

The whole database contains GPS records of approximately 1.6 million anonymized users throughout Japan from 1 August 2010 to 31 July 2011, which contains approximately 9.2 billion GPS records, more than 600GB csv files. The visualization of human mobility in Greater Tokyo Area during the earthquake is shown in Fig.5. From this figure, we can clearly see that the urban transportation before the earthquake was very busy (Fig.5-A,B). In contrast, after the
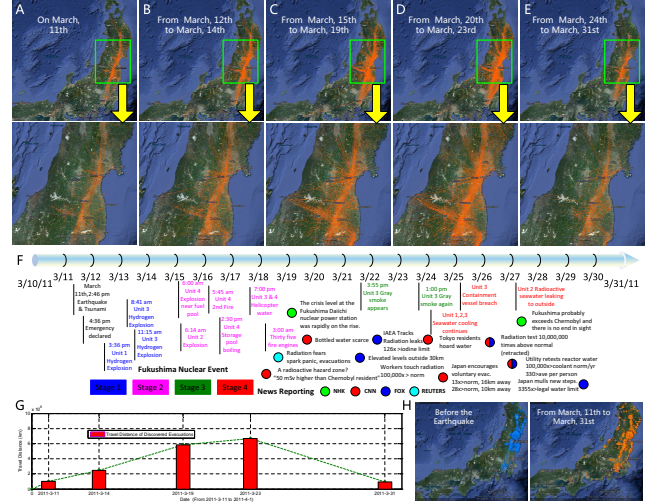

Figure 6: The discovered evacuations at different stages of the disasters in Fukushima, Miyagi and Iwate prefectures. (A-E) shows the discovered evacuations, the orange lines indicate the discovered evacuations that connect people's old and new primary residential areas. (F) shows some important government declarations and news reporting about this event. (G) shows the total travel distances of the discovered evacuations at different stages. (H) shows residential populations before (blue) and after (orange) the earthquake. Larger circles indicate higher population densities, and the statistics were analyzed in 10 km×10 km grids.

earthquake, the transportation network of whole Greater Tokyo Area was almost stalled (Fig.5-C,D).

5.2 Evacuations Discovery and Analysis

Generally speaking, most severe disasters cause large population movements and evacuations. But serious radioactive releases were a bit different from the traditional natural disaster [9], because its real affection was still uncertain to human society, and will last for a very long time. Facing such kind of unfamiliar disaster, people did not have direct feeling of its destructiveness, and their evacuations were usually influenced and caused by government declarations or various kinds of news reporting. To capture this influence, we collected government declarations [10, 11] as well as news reports from NHK in Japan [12] and other mainstream media from around the world [11] from 11 March to 31 March 2011. Based on this information, we divided the Fukushima nuclear event into four temporal stages and analyzed evac-

uation behaviors at these different sampling time periods. These stages included: (1) a period of early government declarations following the Fukushima nuclear event (12 March to 14 March 2011) before the significance of the event was fully known, (2) a period where the seriousness of the nuclear accident was more fully known (15 March to 19 March 2011), (3) a period where the nuclear event was beginning to be described through worldwide news reports (20 March to 23 March 2011), and finally (4) a period where the significance of the release of radioactive materials was more fully understood and described over worldwide news reports (24 March to 31 March 2011). The discovered evacuations during distinct sampling periods in Fukushima, Miyagi and Iwate prefectures are shown in Fig.6. Meanwhile, to get a overview of population distribution before and after the earthquake, we show (Fig.6-H) the residential distributions in Fukushima, Miyagi, and Iwate prefectures 161 days before the earthquake and 20 days after the earthquake.

On the other hand, to further analyze people's evacuations after this catastrophic disaster, we show (Fig.7) population change rate in Fukushima, Miyagi, and Iwate prefectures from short-term to long-term. As shown in Fig.7, we found that 12 days after the earthquake, the populations of Futaba (Fukushima), Onagawa (Miyagi), Kesnnuma (Miyagi), Minamisoma (Fukushima), Kamashi (Iwate), Ishinomaki (Miyagi) substantially decreased by 50% or more (Fig.7-E). In contrast, the populations of Hanamaki (Iwate) and Izumi Ward of Sendai (Miyagi) increased slightly (Fig.7-E), suggesting that people may have taken refuge in these areas. Twenty days after the earthquake, as the nuclear accident became more serious, the populations in Futaba (Fukushima) and Minamisoma (Fukushima) continued to decrease rapidly while populations of other disaster-impacted areas increased slightly where people may have chosen to return to their homes (Fig.7-E). Fifty days after the earthquake, the populations in most of the disaster areas, except for Futaba (Fukushima) and Minamisoma (Fukushima), continued to increase slightly (Fig.7-E). Eighty-one days after the earthquake, the populations in all disaster areas stabilized (Fig.7-E).

5.3 Probabilistic Reasoning

Based on our initial evaluation of movement behaviors following the composite disaster in Japan, we determined that movements were influenced by government declarations and news reports and that population movements differed over different stages of the Fukushima nuclear accident. Accordingly, we used discovered evacuations within respective sampling periods of the disaster stages to construct evacuation graph (Fig.8) for Fukushima, Miyagi, and Iwate prefectures in Japan. Based on this graph, we were able to identify critical evacuation routes as well as urban patterns and connections between different disaster areas (Fig.8A-E). In addition, according to the evacuation graph, we found that from 12 March to 23 March 2011, the graph structure was very similar despite with the different learned parameters (Fig.8B,8C,8D), suggesting that evacuation patterns in these areas share similar characteristics. For instance, people tended to evacuate from the most heavily impacted areas and sought shelter in large neighboring cities or government-appointed shelters.

Meanwhile, given any specific area or cities (with its parameters or states) in Fukushima, Miyagi, and Iwate pre-
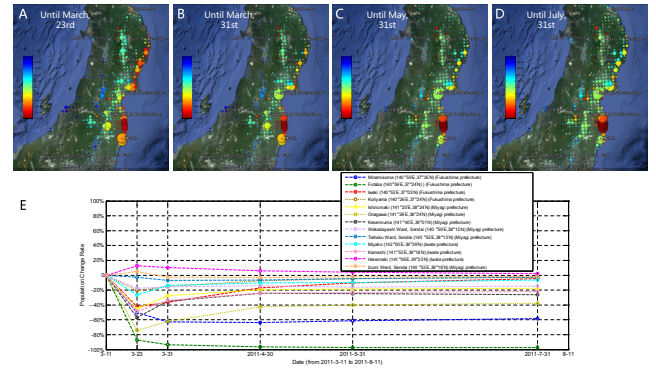

Figure 7: Rate of population change over different sampling periods in Fukushima, Miyagi, and Iwate prefectures in Japan (A-D) where the size of circles denotes the number of observed evacuations (larger circles indicate more observed evacuations) and the color of circles indicates the rate of population change in a specific area (ranging from -1 to 1 (-100% to 100%)). (E) shows the quantitative analysis of major disaster areas in these prefectures.

fectures, and some partially observed people's evacuation trajectories, our training model is able to automatically simulate or predict their population evacuations or mobility. To show the simulation results, we assumed that the same disaster (earthquake and release of radioactivity) occurred at the same places, and let the training model simulate population mobility. In other words, we can easily reproduce the mobility of population during the disasters. We show the simulation results (Fig.9) of some major disaster areas at different stages of this event, such as Minamisoma (Fukushima), Iwaki (Fukushima), Ishinomaki(Miyagi) and Wakabayashi Ward, Sendai (Miyagi).

On the other hand, based on the training model, we try to recommend some safe and fast evacuation routes if some similar disaster occur. Here, we assume that the high frequency routes during the diaster are usually the safe ones. Hence, we find the high frequency and fastest (low travel time) ones in the evacuation graph between origin and destination, and recommend them as evacuation routes during some possible disasters. Some selected results are shown in Fig.10.

5.4 Evaluation and Validation

To evaluate simulation results, we performed K-fold cross-validation. The discovered evacuation samples were randomly partitioned into three sub-samples: one sample was used as validation data while the other two were used as training data. The cross-validation process was then repeated three times, with each sub-one used exactly once as validation data. For each repetition, we computed the Jaccard similarity coefficient [4] between simulation results obtained by the training model and real evacuation distribution in testing samples for some major disaster areas (some high weight nodes) in Fukushima, Miyagi, and Iwate prefectures. For comparison, we evaluated the average simulation accuracy (Table 1) of the five models trained over both the five separate stages of the disaster (as discussed in the previous section) as well as over the full period of evacuations (11 March to 31 March 2011).
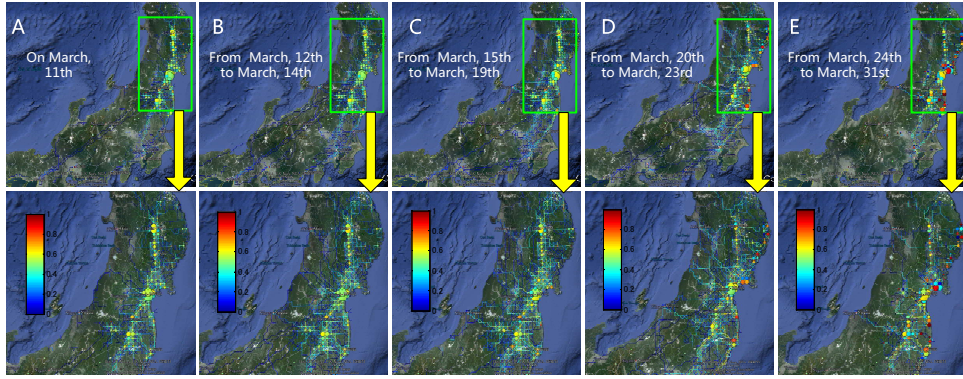
Figure 8: Constructed evacuation graph. (A-E) show the constructed evacuation graph for Fukushima, Miyagi and Iwate prefectures at different stages of this disaster. In this graph, nodes denote important areas (e.g. residential areas before and after the disaster, and some stopovers) where evacuation behaviors were observed, and edges represent people's movements during the disaster. The size of the circles denote the node weights. Here, it shows the number of evacuations observed in the corresponding areas. Please note that nodes with very small weights are not displayed in these figures. The edge color indicates the edge parameters. Here, it shows the travel frequency during the disaster; this value is normalized from 0 to 1. In addition, the circle's color indicates the type of area. Higher values represent areas from which people evacuate. In contrast, areas with lower values are areas where people seek refuge. This value is also normalized from 0 to 1.

Table 1: Simulation Accuracy

Areas	Accuracy	Accuracy (five stages)
Minamisoma	72.37%	85.38%
Futaba	73.62%	88.36%
Iwaki	68.71%	83.29%
Koriyama	63.38%	81.35%
Ishinomaki	78.75%	84.67%
Onagawa	81.72%	86.37%
Kesennuma	76.73%	82.38%
Wakabayashi Ward	63.77%	77.35%
Miyako	77.52%	82.38%
Kamashi	68.74%	75.38%
Hanamaki	73.78%	81.59%

According to this evaluation (Table 1), we found that the five-stage training model that considered the influence of government declarations and news reporting significantly improved simulation results, especially for the cities of Fukushima prefecture that were highly impacted by Fukushima nuclear accident. In addition, for most areas of the three prefectures, the simulation results of our training model reached above 80% accuracy.

6. RELATED WORK

Recently, a number of studies on human mobility patterns during disasters have been proposed [13, 14], mainly focusing on small-scale and short-term emergencies (e.g. crowd panics and fires). However, research on the dynamics of population movements on a national scale during large-scale disasters (e.g. earthquakes, tsunamis, and hurricanes) is very limited [9], likely the result of difficulties in collecting representative longitudinal data in places where infrastructure and social order have collapsed [15] and where study populations are moving across vast geographical areas [9]. In contrast, auto-mobile sensor data offer a new way to circumvent methodological problems of earlier research because they of-

fer high temporal and spatial resolution are instantaneously available, have no interview bias, and provide longitudinal data for very large populations [16, 17, 9, 15, 20, 21]. Meanwhile, human mobility or trajectory data mining [18, 19, 22] have become a very hot topic in various research fields.

7. CONCLUSION AND DISCUSSION

In this paper, we constructed a large population mobility database and have shown the population mobility patterns following the Great East Japan Earthquake and Fukushima nuclear accident. Meanwhile, we developed a general probabilistic model to simulate or predict population evacuations over complex geographic features in Japan in response to future disasters, and experimental evaluations demonstrated that the simulation or prediction of large population mobility in severe disasters were sometimes possible.

We note several limitations within our study. The population mobility database used was constructed from mobile devices and did not incorporate data from some representative portions of the population (i.e. people who did not own mobile devices or did not register for GPS service could not be incorporated into this study). Additionally, data were slightly biased towards younger age groups who were more likely to own GPS-based equipment than older age groups. However, we are confident that the data, which offers movement behaviors for the approximately 1.6 million people included in the database, are reflective of general movement patterns in the country following the composite disaster. A second limitation of our study was related to the difficulty in extrapolating movement patterns predicted by our simulation model for use in places outside of Japan or non-affected places by this disaster. Moreover, the actual performance of this model was difficult to be fully evaluated, except some similar events occur again in Japan.
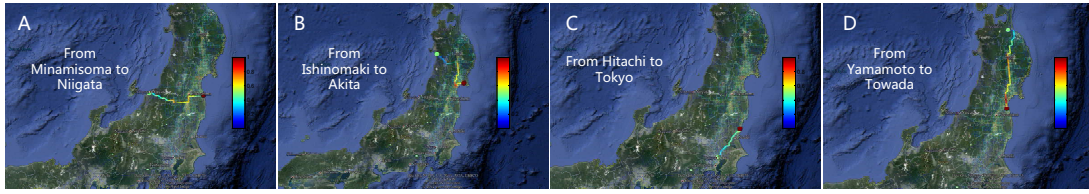
8. ACKNOWLEDGMENTS

Figure 10: Recommended evacuation routes. This figure shows some examples of the recommended evacuation routes between different origins and destinations, such as from Minamisoma to Niigata (Fig.A), Ishinomaki to Akita (Fig.B), Hitachi to Tokyo (Fig.C) and Yamamoto to Towada (Fig.D). The color denotes the route safety, and the warmer one means it will be safer.
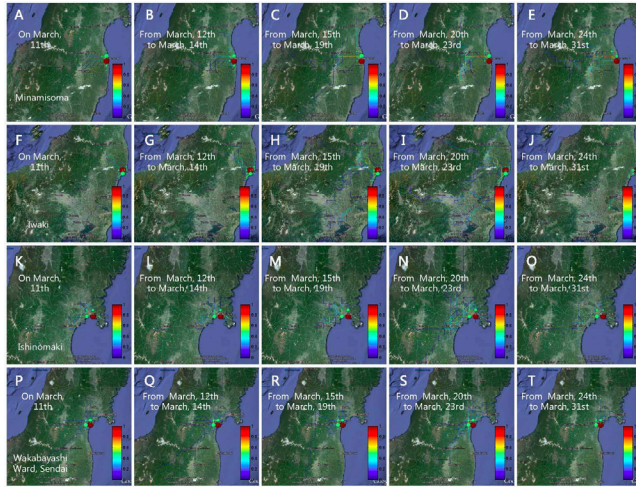

Figure 9: Simulation results. (A-T) show simulation results of people's evacuations at different stages following the Great East Japan Earthquake and Fukushima nuclear accident in some major disaster areas, including Minamisoma (A-E), Iwaki (F-J), Ishinomaki (K-O) and Wakabayashi ward, Sendai (P-T). Given a specific area (red circle), the possible destinations are able to be simulated by the green circles. The size of the green circle indicates the probability that large population will go there; larger circles indicate higher probabilities. Meanwhile, the trajectories show the possible movements of these evacuations, and the color shows the probability normalized from 0 to 1.

9. REFERENCES

[1] O. Norio, T. Ye, Y. Kajitani, P. Shi, and H. Tatano. The 2011 Eastern Japan Great Earthquake Disaster: Overview and Comments, *International Journal of Disaster Risk Science*, 2 (2011), pp. 34-42.

[2] Japanese National Police Agency. Damage Situation and Police Countermeasures associated with 2011 Tohoku district-off the Pacific Ocean Earthquake, http://www.npa.go.jp/archive, (2012).

[3] M. Holt, R. J. Campbell, and M. B. Nikitin. Fukushima Nuclear Disaster, Congressional Research Service Reports, (2012).

[4] P. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Addison Wesley, (2005).

[5] L. Wei, Y. Zheng, and W. Peng. Constructing Popular Routes from Uncertain Trajectories, *Proc. of ACM SIGKDD*, (2012), pp. 195-203.

[6] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley-Interscience, (1994).

[7] B. D. Ziebart, A. Maas, J.A. Bagnell, and A. K. Dey. Navigate Like a Cabbie: Probabilistic Reasoning from Observed Context-Aware Behavior, *Proc. of Ubicomp.*, (2008), pp. 322–331.

[8] B. D. Ziebart, A. Maas, J.A. Bagnell, and A. K. Dey. Maximum Entropy Inverse Reinforcement Learning, *Proc. of AAAI Conference on Artificial Intelligence (AAAI)*, (2008), pp. 1433-1438.

[9] X. Lu, L. Bengtssona, and P. Holme. Predictability of population displacement after the 2010 Haiti earthquake, *Proc. of the National Academy of Sciences of USA (PNAS)*, 109 (2012), pp. 11576-11581.

[10] Japanese Cabinet Secretariat, http://www.cas.go.jp/.

[11] R. Hoetzlein. Visual communication in Times of crisis: the Fukushima Nuclear Accident, *Leonardo Journal of Arts, Science and Technology*, 45 (2012), pp. 113-118.

[12] Japan Broadcasting Corporation (NHK), http://www.cas.go.jp/.

[13] M. Moussaid, S. Garnier, G. Theraulaz, and D. Helbing. Collective information processing and pattern formation in swarms, flocks, and crowds, *Top Cogn. Sci.*, 1 (2009), pp. 469-497.

[14] J. Hahm and JH. Lee. Human errors in evacuation behavior during a traumatic emergency using a virtual fire, *Cyberpsychol Behavior*, 12 (2009), pp. 98-98.

[15] JP. Bagrow, DS. Wang, and AL Barabasi. Collective response of human populations to large-scale emergencies, *Plos ONE*, 6 (2011).

[16] CM. Song, ZH. Qu, N. Blumm, and AL. Barabasi. Limits of predictability in human mobility, *Science*, 327 (2010), pp. 1018-1021.

[17] MC. Gonzalez, CA. Hidalgo, and AL. Barabasi. Understanding individual human mobility patterns, *Nature*, 453 (2008), pp. 779-782.

[18] Z. Chen, H.T. Shen, X. Zhou, Y. Zheng and X. Xie. Searching Trajectories by Locations - An Efficiency Study, *Proc. of ACM SIGMOD*, (2010), pp. 255-266.

[19] F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo and R. Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data, *The VLDB Journal*, 20 (2011), pp. 695-719.

[20] C. Song, T. Koren, P. Wang, and AL. Barabasi, Modelling the scaling properties of human mobility, *Nature Physics*, 6 (2010), pp. 818-823.

[21] X. Song, Q. Zhang, Y. Sekimoto, T. Horanont, S. Ueyama and R. Shibasaki, An Intelligent System for Large-scale Disaster Behavior Analysis and Reasoning, *IEEE Intelligent Systems*, 3 (2013), pp. 1-9.

[22] J. Yuan, Y. Zheng abd X. Xie. Discovering regions of different functions in a city using human mobility and POIs, *Proc. of ACM SIGKDD*, (2012), pp. 186-194.