Intelligent System for Human Behavior Analysis and Reasoning Following Large-Scale Disasters

Xuan Song, Quanshi Zhang, Yoshihide Sekimoto, Teerayut Horanont, Satoshi Ueyama, and Ryosuke Shibasaki, University of Tokyo

By mining “big GPS records” of 1.6 million users, an intelligent system automatically discovers, analyzes, and simulates population evacuations during the Great East Japan Earthquake and the Fukushima Daiichi nuclear accident. By mining this large dataset of spatially referenced mobile sensor data, DBAPRS can automatically discover and analyze evacuation behavior patterns and simulate or predict human mobility for future disaster mitigation. The types of data, evacuation behavior patterns, and simulation models available following the Japanese earthquake and nuclear power plant meltdown are unique in human history, and are likely to play a vital role in future disaster relief and management worldwide.

In this article, we introduce the novel Disaster Behavior Analysis and Probabilistic Reasoning System (DBAPRS) to analyze and simulate people’s evacuation behaviors during the Great East Japan Earthquake and the Fukushima nuclear accident. (For others’ work in this area, see the related sidebar.) DBAPRS is an intelligent system that stores and manages daily GPS records from approximately 1.6 million individuals throughout Japan over a one-year period (from 1 August 2010 to 31 July 2011). By mining this large dataset of spatially referenced mobile sensor data, DBAPRS can automatically discover and analyze evacuation behaviors of people during the disasters. Meanwhile, DBAPRS constructs...
Recently, a number of studies on human mobility patterns during disasters have been proposed, mainly focusing on small-scale and short-term emergencies (for example, crowd panics and fires). However, research on the dynamics of population movements on a national scale during large-scale disasters (such as earthquakes, tsunamis, and hurricanes) is limited, most likely due to difficulties in collecting representative longitudinal data in places where infrastructure and social order have collapsed and where study populations are moving across vast geographical areas. Recently, smartphones and PDAs, which are typically equipped with GPS sensors, have become ubiquitous in daily life. Mobile sensor data from these devices offer a new way to circumvent methodological problems of earlier research, because they offer high temporal and spatial resolution, have no interview bias, and provide longitudinal data for large populations.

Xin Lu and his colleagues collected data from 1.9 million mobile users in Haiti to analyze population displacement after the 2010 Haitian earthquake, which was the first study to analyze large-scale human-mobility patterns after a severe disaster. They concluded that people’s evacuation patterns following the natural disaster were highly correlated with their daily movements prior to the event. However, human-mobility patterns after the Great East Japan Earthquake were different from those following the 2010 Haitian earthquake for two reasons. Large population movements in Japan were caused by the Fukushima Daiichi nuclear accident in addition to the major earthquake and tsunami. Compared to more common natural disasters (for example, earthquakes and tsunamis alone), serious releases of radioactivity have been historically rare. Facing the most extensive release of radioactivity since 1986, human-mobility patterns were expected to differ from previous ones. In addition, the land area of Japan is much larger than that of Haiti, and evacuation behaviors in Japan were therefore more geographically complicated. Therefore, in the study discussed in the main article, we try to develop a system to analyze and simulate population mobility in large-scale disasters of Japan.

**Related Work on Human Mobility Patterns during Disasters**

**References**


**Overall System**

Figure 1 illustrates the DBAPRS architecture, which contains four modules: database server and visualization, discovery and analysis, learning, and probabilistic reasoning. The database server and visualization module stores and manages the GPS data for all people being tracked. The discovery and analysis module analyzes people’s behaviors during the disaster, and automatically discovers long- or short-term population evaucations. The learning module uses the discovered evacuation behaviors to build a probabilistic model. The probabilistic reasoning module simulates or predicts population mobility or evacuations in various cities impacted by possible disasters throughout Japan.

**Database Server and Visualization**

The database server of DBAPRS stores and manages GPS records of approximately 1.6 million anonymized users throughout Japan from 1 August 2010 to 31 July 2011; it now contains approximately 9.2 billion GPS records and more than 600 Gbytes of comma-separated value (CSV) files. The database server and visualization module preprocesses these data, providing indexing, retrieval, editing, and visualization services for users. Moreover, the visualization module visualizes these data in various styles, such as raw GPS records, travel trajectories, travel directions, and so on (see Figure 2), and some parts of the visualization module are based on Google Earth.

**Discovery and Analysis Module**

Generally speaking, most severe disasters cause large population movements or evacuations. Obviously, analysis of these short- and long-term evacuation behaviors will play a vital role in future disaster relief and management worldwide. Hence, we...
designed the discovery and analysis module to discover human evacuation behaviors during the Great East Japan Earthquake and the Fukushima Daiichi nuclear accident. To achieve this task, DBAPRS used several months of data collected before the earthquake (1 October 2010 to 11 March 2011) to compute geographic location information for individual people (see Figure 3). By measuring the similarity of this distribution before and after the earthquake, we discovered people’s evacuation behaviors at different sampling periods.
For each person, the geographic location history is a series of geographic positions including longitude, latitude, and time period. Let $X_k(t, T_{\text{period}}) = \{p_k(t, d) : d \in T_{\text{period}}\}$ denote the geographic location history of person $k$ in time $t(t \in T_{\text{time}})$ during period $T_{\text{period}}$, where $p_k(t, d)$ is the geographic position in time $t$ of day $d$. The geographic location distribution is represented by $K$ bins distributed by

$$L_k(t, T_{\text{period}}) = \{\psi(n; \sigma(p_k(t, d)), d \in T_{\text{period}})\}_{n = 1, \ldots, K},$$

where $\sigma(p_k(t, d))$ is a function that computes the bin index associated with the geographic location, and $\psi(n; \sigma(p_k(t, d)), d \in T_{\text{period}})$ denotes the probability that person $k$ will appear in location index $n$ at time $t$ during the period $T_{\text{period}}$. Figure 3 shows an example of this distribution. Based on the analysis of this distribution, DBAPRS can also find some regular and important places for each person, such as home, working areas, and so on.

To discover the evacuation routes and locations during the disasters, DBAPRS computes the similarity of geographic location distribution for an individual before and after the earthquake, and if the similarity is small enough, this person’s behavior is classified as an evacuation (as shown in Figures 3 and 4). Here, we use a Jaccard coefficient to measure this similarity due to its efficiency, and we calculate the coefficient $\alpha$ for before earthquake period $T_{\text{period}}$ and after $T_{\text{period}}$ as follows:

$$\alpha_k \left( T_{\text{before}, T_{\text{period}}}, T_{\text{after}, T_{\text{period}}} \right) = \frac{1}{T_{\text{time}}} \sum_{T_{\text{time}}} \sum_{n=1}^{K} \max \left( \psi(n, T_{\text{before}, T_{\text{period}}}), \psi(n, T_{\text{after}, T_{\text{period}}}) \right).$$

Therefore, based on this similarity, the system can automatically discover people’s evacuation routes and locations across the entire country in the short- or long-term by mining the auto-GPS mobile sensor database.

**Learning Module**

To effectively understand, simulate, and predict human mobility during severe disasters, DBAPRS builds a probabilistic model and uses the discovered human evacuation behaviors (movement trajectories during the disasters, as shown in Figure 1a) to train its parameters via the machine learning technique (see Figure 1b). In this research, we assume that the stored auto-GPS mobile sensor data in DBAPRS are representative of the general population’s movements during disasters. The creation of this type of predictive 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 disasters.

To learn the simulation model, DBAPRS has two stages: evacuation graph construction (as shown in Figure 1c) and Markov decision process (MDP) learning (as shown in Figure 1d). First, DBAPRS constructs the people’s evacuation graph in some affected regions, revealing and modeling population movements or evacuations between different disaster areas. Because most public transportation systems were usually not available after the large-scale disasters occurred, we couldn’t use transportation networks for building it. Hence, DBAPRS uses the geographic location distribution (see Figure 1a) and the training samples of people’s evacuations (see Figure 1b).
to construct the graph through collaborative learning. In this graph, the nodes usually denote some disaster areas (cities or regions) affected by the disasters, and the edges indicate some important evacuation routes among these disaster areas (see Figures 5a–5e). Second, given the learned evacuation graph, DBAPRS builds the simulation model based on the MDP. Here, the evacuation graph provides a deterministic MDP, the geographical region (nodes) is considered a state, the edge is the action, and the path is the people’s evacuation trajectory (as in Figure 1d). These evacuation trajectories can be parameterized by their path feature (for example frequency of used routes, travel time, and so on). Hence, DBAPRS uses these parameterized evacuation trajectories to train the parameters of MDP via inverse reinforcement learning.

**Probabilistic Reasoning**

Based on the training model, we can simulate or predict people’s evacuation movements for similar future disasters. DBAPRS uses the Bayes’ rule to perform the probabilistic inference: given the partial observed evacuations (such as some evacuations during the first several hours or days after disasters), the posterior probability of the destinations is computed by

\[
P(\text{dest}|\zeta_{A\rightarrow B}, \phi) \propto P(\zeta_{A\rightarrow B}|\text{dest}, \phi) P(\text{dest}),
\]

where \(P(\text{dest})\) is the evacuation prior probability in a region A, and which we can compute by popular route inference in the evacuation graph. \(P(\zeta_{A\rightarrow B}|\text{dest}, \phi)\) is the likelihood of observed evacuations, where \(\phi\) is the learned parameters of the MDP model and is computed by taking the sums over paths from region A to region B to each possible destination using the forward-pass algorithm in the learned MDP model.

Hence, we can simulate and predict possible evacuation routes and destinations by the maximum a posteriori estimation of Equation 1.

**Experimental Results**

Figure 4 shows evacuations that DBAPRS discovered at different periods of the disasters in Fukushima, Miyagi, and Iwate prefectures. From these results, we found distinct patterns of human movement in each of the periods following the earthquake and nuclear disaster. Some evacuations began...
prior to the first time period, as people responded to the earthquake and tsunami themselves (see Figure 4a).

During the first period following the nuclear accident, we believe that most people in the Fukushima, Miyagi, and Iwate prefectures didn’t understand the serious nature of the release of nuclear materials despite
government declarations of an emergency situation in those prefectures. Hence, only small numbers of people began to evacuate over short distances at this time (see Figures 4b).

Over the second period, when people began to better understand the accident’s seriousness, the number of evacuations and travel distances substantially increased (see Figure 4c). The number of evacuations and travel distances continued to increase over time period three as news reports began to describe the nuclear event (see Figure 4d).

Finally, during the fourth period, the number of evacuations and travel distances began to decrease (see Figure 4e) despite the fact that the mainstream worldwide media dubbed the Fukushima Daiichi nuclear accident as one of the most serious accidents in human history. We believe this was because most people who had previously evacuated to safe places in days prior chose to return home or had found a safe place to stay. Figure 4f graphs the data from these images. Meanwhile, to get an overview of population distribution before and after the earthquake, DBAPRS presents the residential distributions in Fukushima, Miyagi, and Iwate prefectures 161 days before and 20 days after the earthquake (see Figure 4g).

Figures 5a through 5e show the constructed evacuation graph for the Fukushima, Miyagi, and Iwate prefectures and corresponding simulation results of some major disaster areas at different stages of this event. According to these results, we found that certain patterns of human movement in each of the prefectures were linked. Because Minamisoma and Iwaki were both heavily impacted by the earthquake, tsunami, and nuclear accident, population mobility in these locations shared similar characteristics (see Figures 5f through 5o). From 11 March to 14 March 2011, evacuations were heavily concentrated in nearby cities or shelters as people lost their homes due to the earthquake and tsunami. From 15 March to 31 March 2011, the Fukushima Daiichi nuclear accident became more serious, and the range of evacuations became substantially larger from those prefectures. However, patterns of evacuation from Ishinomaki (where the impact of the composite disaster was less severe), were different (see Figures 5p through 5t).

From 11 March to 31 March 2011, evacuations were concentrated in specific regions and didn’t change through time, although some people in Ishinomaki extended their evacuation range between 20 March and 23 March 2011 (see Figure 5s). We think that people might have stayed in place because the earthquake and tsunami had seriously destroyed two regions; also, due to the grief people felt at losing houses or relatives, many seemed to not care about the radioactive releases, even though they weren’t far from the Fukushima Daiichi Nuclear Power Plant. These people just wanted to find a safe place to stay.

To evaluate simulation results, DBAPRS performed K-fold cross-validation. Here, we used evacuation trajectory samples from Fukushima, Miyagi, and Iwate prefectures during 11 March to 31 March 2011. DBAPRS randomly partitioned these samples into three subsamples: 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 subsample used exactly once as validation data. For each repetition, DBAPRS computed the Jaccard similarity coefficient between simulation results obtained by the training model and real evacuation distribution in testing samples for some major disaster areas (high-weight nodes in Figure 5’s evacuation graph). Here, we used the similarity coefficient as simulation accuracy, which Table 1 shows. From this evaluation, we see that the accuracies for individual cities or regions ranged from 75.38 to 88.36 percent, with the majority higher than 80 percent.

In this study, we’ve demonstrated that the accurate simulation or prediction of large population mobility in severe disasters is possible. Further, on the basis of DBAPRS results, we found that in regions instantaneously impacted by the earthquake and tsunami, large numbers of people sought immediate refuge in nearby cities or government shelters. However, in regions more impacted by the release of nuclear materials, evacuation patterns were highly influenced by government declarations and news reports. Evacuations became substantially more extended and disorderly as people

<table>
<thead>
<tr>
<th>Area (prefecture)</th>
<th>Simulation accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minamisoma (Fukushima)</td>
<td>85.38</td>
</tr>
<tr>
<td>Futaba (Fukushima)</td>
<td>88.36</td>
</tr>
<tr>
<td>Iwaki (Fukushima)</td>
<td>83.29</td>
</tr>
<tr>
<td>Koriyama (Fukushima)</td>
<td>81.35</td>
</tr>
<tr>
<td>Ishinomaki (Miyagi)</td>
<td>84.67</td>
</tr>
<tr>
<td>Onagawa (Miyagi)</td>
<td>86.37</td>
</tr>
<tr>
<td>Kesennuma (Miyagi)</td>
<td>82.38</td>
</tr>
<tr>
<td>Wakabayashi Ward, Sendai</td>
<td>77.35</td>
</tr>
<tr>
<td>Miyako (Iwate)</td>
<td>82.38</td>
</tr>
<tr>
<td>Kamashi (Iwate)</td>
<td>75.38</td>
</tr>
<tr>
<td>Hanamaki (Iwate)</td>
<td>81.59</td>
</tr>
</tbody>
</table>
became informed of the significance of the radioactive materials being released as a result of the disaster.

We note several limitations within this system and our study. The dataset of population movements used was constructed from mobile devices and didn’t incorporate data from some representative portions of the population (that is, people who didn’t own mobile devices or didn’t register for GPS service couldn’t be incorporated into this study). Additionally, data were slightly biased toward younger age groups who were more likely to own GPS-based equipment than older age groups. However, we’re confident that the data, which offers movement behaviors for the approximately 1.6 million people included in the dataset, are reflective of general movement patterns in the country following the composite disaster. A second limitation of this system was related to the difficulty in extrapolating movement patterns predicted by DBAPRS for use in places outside of Japan in places not affected by this disaster. Actually, the prediction or simulation is available only for some highly affected cities or regions in the east of Japan. Moreover, the actual performance of DBAPRS was sometimes difficult to fully evaluate. Further study is needed during similar events, even though we hope that similar events never occur.

For future work, this research can be extended and improved. Obviously, people’s evacuation patterns are complicated and influenced by various factors (such as media coverage, city characteristics, and so on). Fortunately, the simulation model of DBAPRS is a general model, and we can easily add these types of information for consideration. Hence, we need to study more factors that influence human mobility and thereby develop a more accurate simulation model.

Acknowledgments
This work was partially supported by a grant of Japan’s Ministry of Land, Infrastructure, Transport, and Tourism, and by the Data Integration and Analysis System/Green Network of Excellence (DIAS/ GRENE) project of Japan’s Ministry of Education, Culture, Sports, Science, and Technology. We especially thank Zenrin Data-Com for their support.

References