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.

The 9.0 magnitude Great East Japan Earthquake[1] occurred on 11 March 2011 off the east coast of Honshu, Japan's largest island. This was the most powerful recorded earthquake in Japan and one of the world's five most powerful earthquakes.[1] The human toll of this disaster was 15,881 deaths, 6,142 injured, and 2,668 missing persons.[2] As many as 128,801 buildings were damaged or destroyed.[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 Ukraine.[3]

The three separate events (earthquake, tsunami, and radioactive material release) created an unprecedented composite disaster that significantly impacted the people of Japan. In the wake of a disaster of this magnitude, there's an urgent need to develop an intelligent system able to objectively record people's movements following the event, analyze their behavioral 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

1541-1672/13/$31.00 © 2013 IEEE
Published by the IEEE Computer Society

Related Work on Human Mobility Patterns during Disasters

Recently, a number of studies on human mobility patterns during disasters have been proposed,[1,2] 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.[3] 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.[3–5]

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.[3] 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.

References

1. M. Moussaid et al., "Collective Information Processing and Pattern Formation in Swarms, Flocks, and Crowds," *Topics Cognitive Science*, vol. 1, 2009, pp. 469–497.
2. J. Hahm and J.H. Lee, "Human Errors in Evacuation Behavior During a Traumatic Emergency Using a Virtual Fire," *Cyberpsychology and Behavior*, vol. 12, 2009, pp. 98–98.
3. X. Lu, L. Bengtssona, and P. Holme, "Predictability of Population Displacement after the 2010 Haiti Earthquake," *Proc. Nat'l Academy Science USA*, vol. 109, no. 29, 2012, pp. 11576–11581.
4. C.M. Song et al., "Limits of Predictability in Human Mobility," *Science*, vol. 327, no. 5968, 2010, pp. 1018–1021.
5. M.C. Gonzalez, C.A. Hidalgo, and A.L. Barabasi, "Understanding Individual Human Mobility Patterns," *Nature*, vol. 453, no. 7196, 2008, pp. 779–782.

a simulation model that can be efficiently trained by using these discovered evacuation behaviors. This model helps us better understand human evacuation behaviors in general, and understand how those behaviors are impacted by various cities' states and social connections during disasters.

Moreover, based on the training model, DBAPRS can simulate or predict population mobility in various cities throughout Japan in an effort to inform future disaster relief and management. We believe that the data and results obtained by this system have enormous value and significance because they objectively and precisely reflect the behavior of people facing the huge, composite disasters. This will contribute to various research fields such as disaster prevention and management, civil engineering, intelligent transportation, urban management, and so on.

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 evacuations. 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

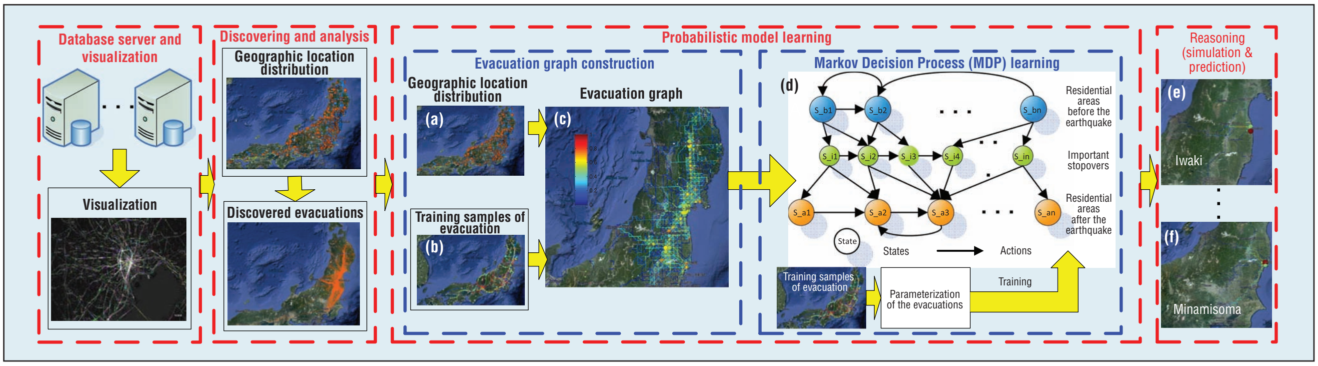


Figure 1. System overview. The Disaster Behavior Analysis and Probabilistic Reasoning System (DBAPRS) contains four modules: database server and visualization, discovery and analysis, learning, and probabilistic reasoning. DBAPRS uses (a) the geographic location distribution and (b) the training samples of people's evacuations to construct an evacuation graph. To learn, the simulation model has two stages: (c) evacuation graph construction and (d) Markov decision process learning. (e,f) Based on the trained probabilistic model, the system can automatically simulate or predict the population mobility in impacted cities.

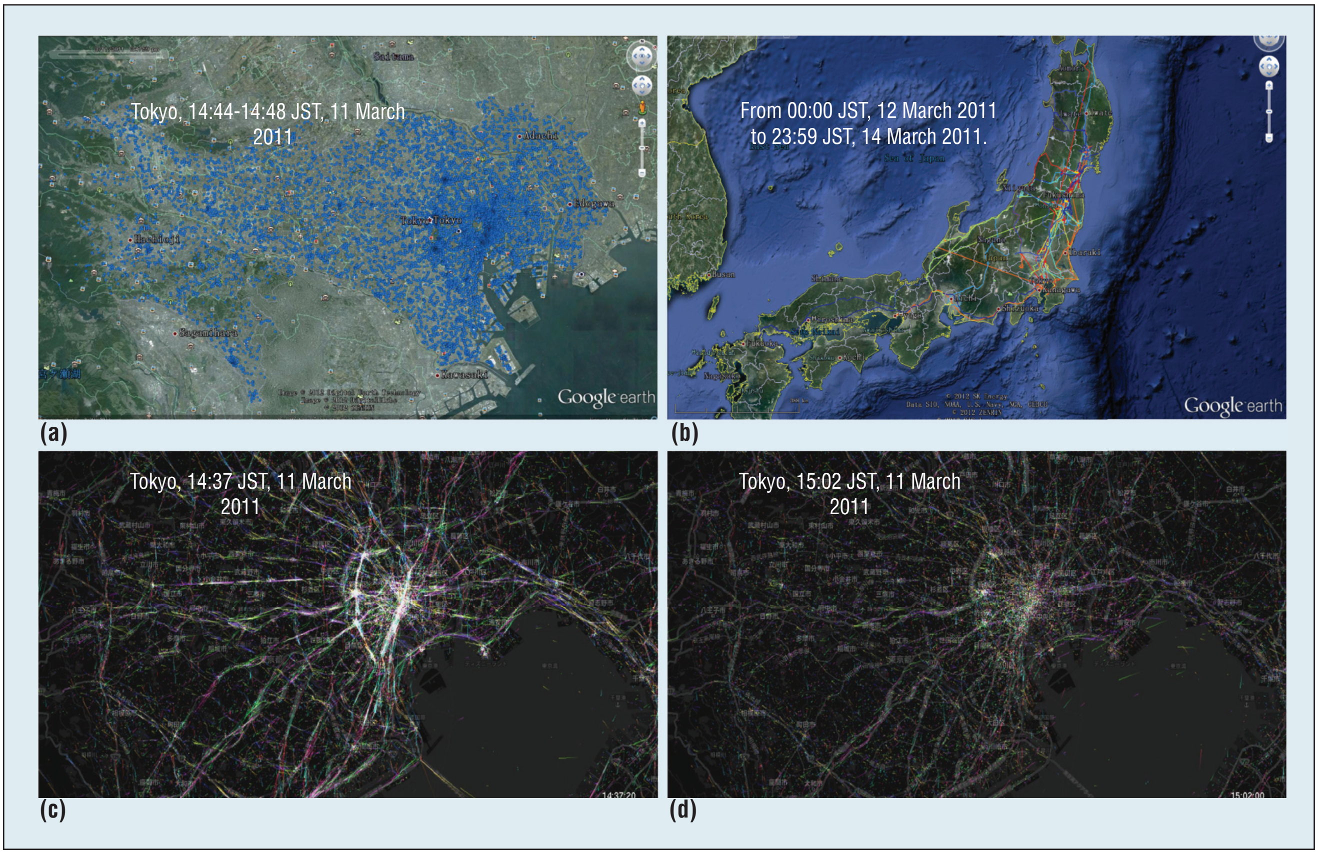
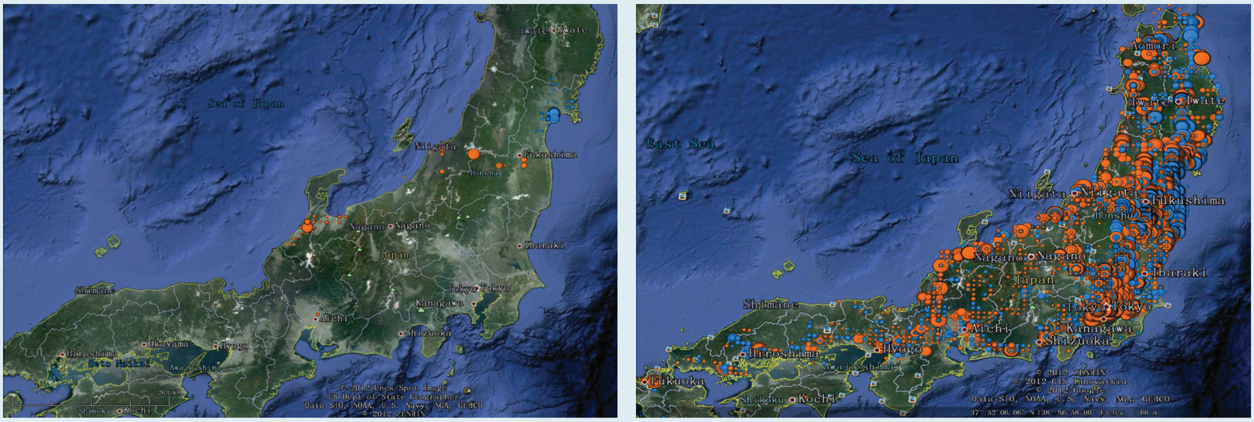


Figure 2. Visualization of the DBAPRS data in various styles. (a) The people's raw GPS recordings in Tokyo, from 14:44 to 14:48 JST, 11 March 2011 (the time when the earthquake occurred). (b) Some examples of the people's movements after the disasters. Different colors denote different persons. (c, d) Direction map and trajectories of all the people in the Greater Tokyo Area before and after the earthquake, respectively. Here, the colors denote people's travel directions. See http://shiba.iis.u-tokyo.ac.jp/song/?page_id=50 for the demonstrations.

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.



(a) (b)

Figure 3. Examples of people's geographic location distribution before and after the earthquake. The size of the circles indicates the probability of an individual person staying in a location at a specific time; larger circles indicate a higher probability that people stay or live there. Blue and orange circles indicate this distribution before and after the earthquake, respectively. The geographic location distribution of (a) a single person, and (b) multiple persons before and after the earthquake.

For each person, the geographic location history is a series of geographic positions including longitude, latitude, and time period. Let $X_k(t, T_{\text{period}}) = \{\mathbf{p}_k(t, d) : d \in T_{\text{period}}\}$ denote the geographic location history of person k in time $t(t \in T_{\text{time}})$ during period T_{period}, where $\mathbf{p}_k(t, d)$ is the geographic position in time t of day d. The geographic location distribution is represented by K bins distributed by

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

where $\sigma(\mathbf{p}_k(t, d))$ is a function that computes the bin index associated with the geographic location, and $\psi(n; \sigma(\mathbf{p}_k(t, d)), d \in T_{\text{period}})$ denotes the probability that person k will appear in location index n at time t during the period T_{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 α for before earthquake period $T^{\text{before}}_{\text{period}}$ and after $T^{\text{after}}_{\text{period}}$ as follows:

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

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)

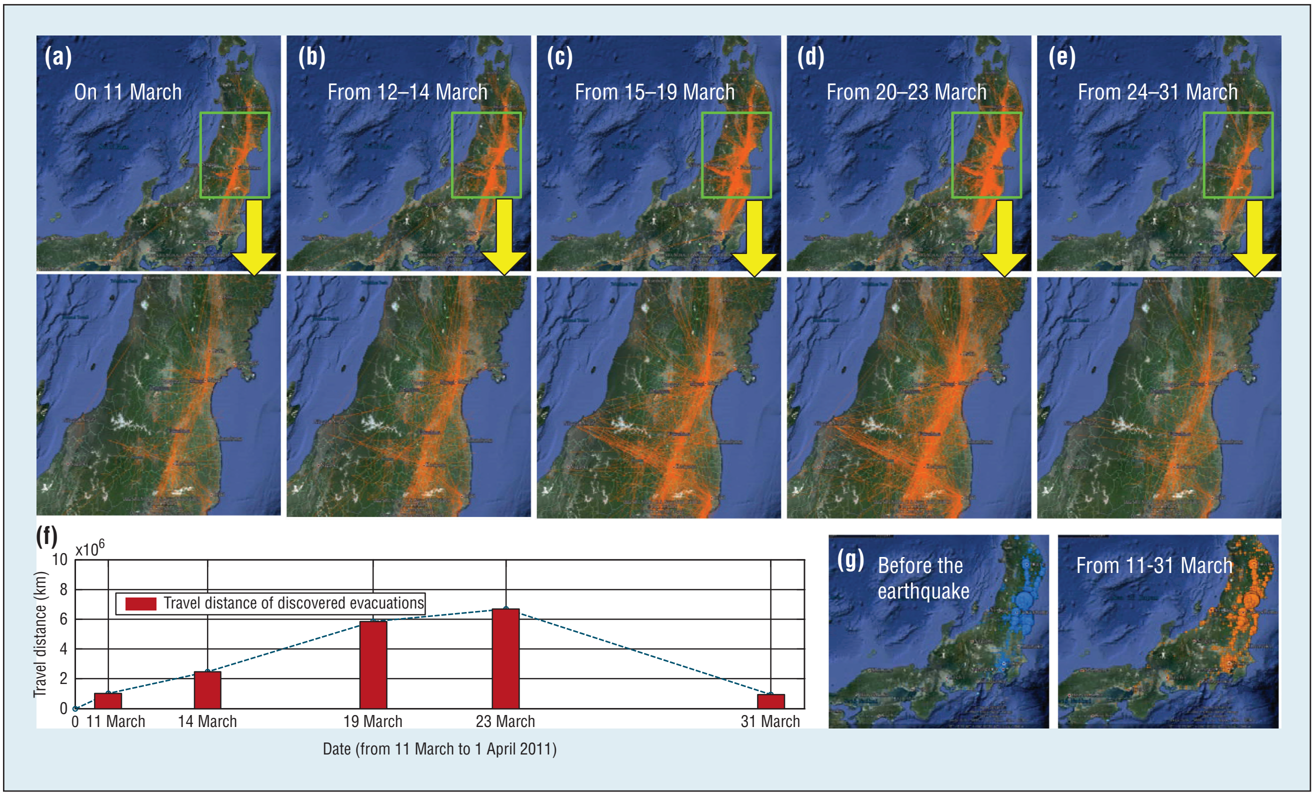


Figure 4. The discovered evacuations at different stages of the disasters in Fukushima, Miyagi, and Iwate prefectures. (a–e) The orange lines indicate the discovered evacuations that connect people's old and new primary residential areas. (f) The total travel distances of the discovered evacuations at different stages. (g) Residential populations before (blue) and after (orange) the earthquake. Larger circles indicate higher population densities. We analyzed the statistics in 10 × 10-km grids.

to construct the graph through collaborative learning.[4] 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.[5] Last, based on the trained probabilistic model, population mobility in various cities impacted by the disasters throughout the country is automatically simulated or predicted via probabilistic inference (see Figures 1e and 1f). More technical details are provided elsewhere.[6]

Probabilistic Reasoning

Based on the training model, we can simulate or predict people's evacuation or 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), $\zeta_{A\to B}$, the posterior probability of the destinations is computed by

$$P(\mathrm{dest}|\zeta_{A\to B}, \phi) \propto P(\zeta_{A\to B}|\mathrm{dest}, \phi)\, P(\mathrm{dest}), \quad (1)$$

where $P(\mathrm{dest})$ is the evacuation prior probability in a region A, and which we can compute by popular route inference[4] in the evacuation graph. $P(\zeta_{A\to B}|\mathrm{dest}, \phi)$ is the likelihood of observed evacuations, where ϕ 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[5] 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

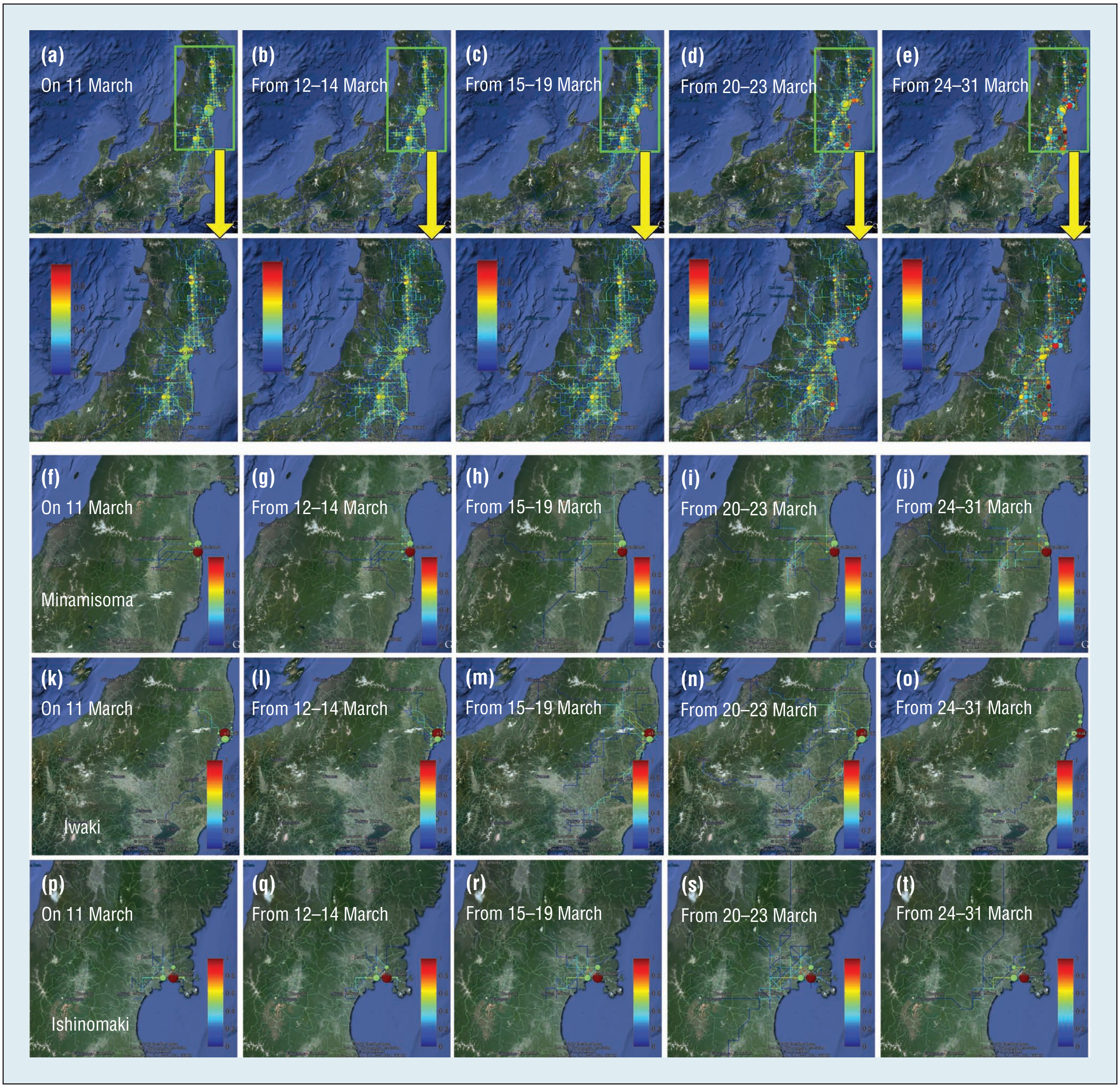


Figure 5. Evacuation graph and the simulation results. (a–e) The constructed evacuation graph for Fukushima, Miyagi, and Iwate prefectures at different stages of this disaster. Nodes denote important areas (for example, residential areas before and after the disaster, and some stopovers) where evacuation behaviors were observed; edges represent people's movements during the disaster (the edge value is normalized from 0 to 1). Circle size denotes node weights (nodes with very small weights aren't displayed in these figures). The edge color indicates the edge parameters. Circle color indicates the area type. Higher values represent areas from which people evacuated; lower values are areas where people sought refuge. This value is also normalized from 0 to 1. (f–t) Simulation results of people's evacuations at different stages following earthquake and nuclear accident in some major disaster areas, including (f–j) Minamisoma, (k–o) Iwaki, and (p–t) Ishinomaki. Given a specific area (red circle), the possible destinations and routes are simulated by the green circles. A green circle's size indicates the probability that large populations will evacuate to this area; larger circles indicate higher probabilities. Trajectories show the possible movements of these evacuations; and the color shows the probability normalized from 0 to 1.

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.

Table 1. Simulation accuracy.

Area (prefecture)	Simulation accuracy (%)
Minamisoma (Fukushima)	85.38
Futaba (Fukushima)	88.36
Iwaki (Fukushima)	83.29
Koriyama (Fukushima)	81.35
Ishinomaki (Miyagi)	84.67
Onagawa (Miyagi)	86.37
Kesennuma (Miyagi)	82.38
Wakabayashi Ward, Sendai (Miyagi)	77.35
Miyako (Iwate)	82.38
Kamashi (Iwate)	75.38
Hanamaki (Iwate)	81.59

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

THE AUTHORS

Xuan Song is a project assistant professor with the Center for Spatial Information Science, University of Tokyo. His research interests include artificial intelligence and its related research areas, including robotics and intelligent systems, computer vision, and data mining. Song has a PhD in signal and information processing from Peking University, China. Contact him at songxuan@csis.u-tokyo.ac.jp.

Quanshi Zhang is a PhD candidate at the Center for Spatial Information Science, University of Tokyo. His research interests include artificial intelligence, computer vision, and robotics, especially related to the construction of visual object knowledge bases, 3D point cloud processing, and knowledge mining. Zhang has an MS in spatial information science from the University of Tokyo. Contact him at zqs1022@csis.u-tokyo.ac.jp.

Yoshihide Sekimoto is a project associate professor at the Center for Spatial Information Science at the University of Tokyo. His research interests include large-scale data processing of moving objects and infrastructure data management. Sekimoto has a PhD in civil engineering from the University of Tokyo. He's a member of IEEE, ACM, GIS Association of Japan, Japan Society of Civil Engineers, and Japan Society of Traffic Engineers. Contact him at sekimoto@csis.u-tokyo.ac.jp.

Teerayut Horanont is a project researcher at Earth Observation Data and Information Fusion Research Initiative at the University of Tokyo. His research interests include human mobility analysis and large-scale data mining. Horanont has a PhD in civil engineering from the University of Tokyo. Contact him at teerayut@iis.u-tokyo.ac.jp.

Satoshi Ueyama is a project researcher at Earth Observation Data and Information Fusion Research Initiative at the University of Tokyo. His research interests include analyzing and visualizing data acquired from moving objects. Contact him at uym@csis.u-tokyo.ac.jp.

Ryosuke Shibasaki is a professor at the Center for Spatial Information Science at the University of Tokyo. His research interests include satellite and airborne remote sensing, tracking technologies, geospatial information gathering and integration among heterogeneous systems, and common service platforms for geospatial information. Shibasaki has a PhD in civil engineering from the University of Tokyo. Contact him at shiba@csis.u-tokyo.ac.jp.

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 DataCom for their support.

References

1. O. Norio et al., "The 2011 Eastern Japan Great Earthquake Disaster: Overview and Comments," *Int'l J. Disaster Risk Science*, vol. 2, no. 1, 2011, pp. 34–42.
2. Japanese National Police Agency, *Damage Situation and Police Countermeasures Associated with 2011 Tohoku District-off the Pacific Ocean Earthquake*, 2013; www.npa.go.jp/archive/keibi/biki/higaijokyo_e.pdf.
3. M. Holt, R.J. Campbell, and M.B. Nikitin, *Fukushima Nuclear Disaster*, Congressional Research Service Reports, 2012.
4. L. Wei, Y. Zheng, and W. Peng, "Constructing Popular Routes from Uncertain Trajectories," *Proc. ACM Special Interest Group Knowledge Discovery and Data Mining*, ACM, 2012, pp. 195–203.
5. B.D. Ziebart et al., "Maximum Entropy Inverse Reinforcement Learning," *Proc. AAAI*, AAAI, 2008, pp. 1433–1438.
6. X. Song et al., "Modeling and Probabilistic Reasoning of Population Evacuation during Large-Scale Disaster," *Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining*, ACM, 2013 pp. 1231–1239.