Making Real-Time Predictions of People's Irregular Movement in a Metropolitan Scale under Disaster Situations

Takahiro Yabe, Yoshihide Sekimoto, Hiroshi Kanasugi, Takehiro Kashiyama

Abstract

In this paper, we present our method of accurately predicting people's movement under disaster situations in a metropolitan scale. Recently many studies have suggested methods on predicting the daily movement of the people from various datasets. However because of the lack of accumulated data and irregular behaviors of people in emergency, people movement prediction under disaster situations has been challenging. Our method combines multi agent simulation which uses heterogeneous characteristics of people in the behavioral model, and real-time observation data. We verified the effectiveness of our method by experimenting it on the Great East Japan Earthquake, and also showed that the hourly update of the optimum parameters of the model was effective in keeping a high prediction accuracy. Also, the number of people bound for each destination matched the survey results carried out by different research, supporting the accuracy of the method.

T. Yabe (Corresponding author), Y. Sekimoto, T. Kashiyama, H. Kanasugi
Department of Engineering, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, JP
Email: yabe0505@iis.u-tokyo.ac.jp

Y. Sekimoto
Email: sekimoto@iis.u-tokyo.ac.jp

T. Kashiyama
Email: ksym@iis.u-tokyo.ac.jp

H. Kanasugi
Email: yok@iis.u-tokyo.ac.jp

1. Introduction

Many countries and areas suffer from devastating natural disasters every year, and disaster prevention is one of the most important tasks for policy makers. Damage could become significant especially in high dense urban areas which heavily rely on public transportation, since many people would not be able to return home located in the outskirts if they become paralyzed. Excessive concentration would cause secondary disasters, such as the lateness of rescue activities and inefficient distribution of emergency supplies. These phenomena were actually observed in metropolitan Tokyo during the Great East Japan Earthquake, where over 5.5 million people were left in city centers far from their homes with nowhere to stay overnight [Ito *et al.*, Hiroi *et al.*]. Many metropolitan areas mainly in Southeast Asia such as Manila, Jakarta, and Dhaka are facing rapid population increase as well as high natural disaster risks. There is a huge demand for the real-time prediction of people movement in a metropolitan scale during disaster situations in order to avoid the expected consequences.

Recently, aggregated real-time population data are being provided by companies like Zenrin Data Com and NTT, which are aggregations of GPS data or Call Detail Record Data (CDR) from mobile phones of users who have agreed to give their location data as sample data to mobile phone companies. These data are aggregated in square meshes in which both lengths are 250 meters (Figure 1). Figure 1 shows the density in central Tokyo area. The darkness of red shows the highness of density.

Although these aggregated population data can be used to detect the real-time density of all areas, it does not provide information on the movement of each of the people, which is preferred in further application of people movement data.

Many research have introduced methods to predict daily individual movement, using various datasets such as Call Detail Records (CDR) of mobile

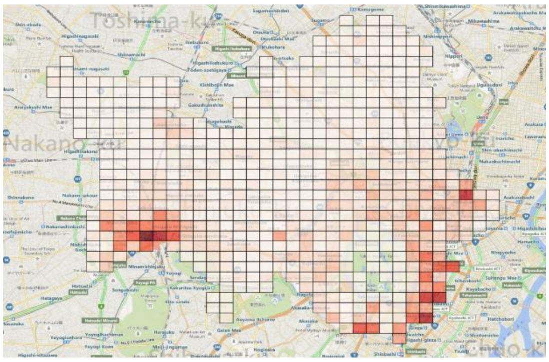

Figure 1 Example of Density Map Data

phones [Hasegawa *et al.*, Calabrese *et al.*, Kanasugi *et al.*], disaggregated GPS data [Ashbrook *et al.*, Herrera *et al.*, Song *et al.*, Horanont *et al.*], and surveys of daily people activities such as Person Trip Surveys (PTS). [Sekimoto *et al.*] predicted the movement of people in metropolitan Tokyo and metropolitan Hanoi using the PTS which is a survey that asks a sample of around 10% about their daily movement patterns. The PTS also includes personal information additional to daily movement such as age, sex, address information, and transportation methods, which are very useful in many ways.

While many researches present methods on the prediction of daily individual movement, the movement of people in disaster situations is difficult because of the lack of data in these situations. Many researches have used multi agent simulations and various constraints to predict the movement of people in that situation. However, most of these researches focus on small scale areas and specific situations [Levi *et al.*, D'Orazio *et al.*]. Since the problem of being stranded in urban areas is an urgent and serious problem in Japan, some researches in Japan have focused on large scale simulation [Osaragi]. While multi agent simulations can predict people movement in disaster situations, the accuracy of the results is not guaranteed since they do not take real-time observed data into account.

Some research attempt to assimilate real-time observation data in order to improve the accuracy of the prediction [Chen *et al.*, Madey *et al.*]. However, these work only focus on the locations of the agents and do not consider their personal characteristics or the reasoning of their movement.

In this paper, we present a method to predict people's irregular movement in a large scale area under disaster situations by combining multi agent simulations using personal characteristics obtained from surveys and real-time observation data. This way, we can accurately predict irregular people movement while also being able to give descriptions on why people are actually showing that movement.

2. Used Datasets

3 datasets are used in our method, all of which are purchasable and are useable without legal regulations. Based on the PTS provided by the Japanese government, we create the location data of all the people in the time of the occurrence of the disaster, since we can apply the additional information such as the personal characteristics including age, sex, home location, trip purpose, and transportation methods, which are included in the survey in our simulation. Also for the simulation, we use the road data and building maps

provided by Zenrin Data Com ltd. As the observation data during the disaster situations, we use Density Data provided also by Zenrin Data Com. Density Data is mesh-aggregated population data made from sample GPS disaggregated data, and is provided every hour. We can assume that Density Data will also be provided in disaster times because it was provided during the Great East Japan Earthquake.

3. Methods

Figure 2 shows the whole flow of our method. It is constructed from 4 steps. In the first step, we create the primary location data of all the persons in the area at the time of the occurrence of the disaster. This must be accurate, since in the second step, we create many scenario candidates of the people's movement 1 time step later with our multi agent simulation. In the third step, we compare the real-time observed population distribution data with all the scenario candidates we made to evaluate which set of parameters generated the scenario closest to the reality. After calculating the optimum parameter set based on the likelihood with the real-time population distribution, in the fourth step we predict the future people movement using that optimum parameter set. We repeat the second step to the fourth step every time real-time data of population distribution is obtained, and we update the optimum parameter set each time. This way, we can continuously predict the future movement based on the real-time situation.

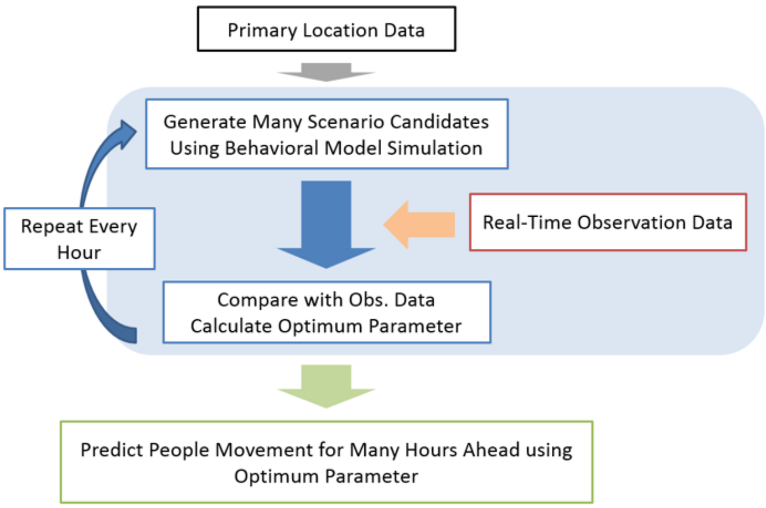

Figure 2 Diagram of Method

We must also note that we will predict the people movement under the situation that all public transportation have stopped from the damage of the disaster, as we have observed in the Great East Japan Earthquake and many other natural disasters. The following sections focus on each step more closely.

3.1 Creating Primary Location Data

The main goal of this step is to create an accurate location data at the time of the occurrence of the disaster for every person with personal characteristics from the PTS. We will use the personal characteristics included in the PTS in our simulation model to increase the reality of it. The PTS has several demerits, which are a) the lack of updates, and b) the vagueness of the location information. Firstly, the problem with the lack of updates is that the magnification factor (MF) changes as time passes, and since the last PTS in Tokyo was held in 2008, there will be some difference from the current population distribution. To cover this demerit, we modified the MF for each person so that the total population in each PTS zone matches the observed population data. Secondly, the PTS only gives location data by zone codes, where each zone is large enough to contain over ten 250 m-meshes. In order to raise the accuracy of people distribution when compared with DD in 250m-mesh aggregated data, we distributed each person in each zone to meshes in that zone corresponding to the ratio of population of all the meshes in the zone in the observed data. By these two methods, we obtained an accurate location data of all the people in the area at the time of the occurrence of the disaster.

3.2 Making Scenario Candidates

In order to be able to investigate ad predict a scenario close to the reality which is difficult to predict because of the irregular, disastrous situations, we take an approach that generates many scenario candidates and expect that some among them closely reflect the reality. Therefore, to be able to generate many scenario candidates in a real-time speed, time must be reduced as much as possible, and to realize that, there must be only a few parameters, which are also very effective, in the simulation model.

From the primary location data, we move the agents using multi agent simulation based on a behavioral model shown on figure 3, until the next time step.

The behavioral model is consisted of 4 parameters describing the possible actions of people during and after a disaster. The simulator describes the slowing down of people movement due to traffic congestion by applying the queueing theory. Also, we have to note that, since we are simulating in a situation where all the public transportations have stopped, all the people who were riding on them will start moving by foot from the time of the occurrence of the disaster. People who were riding their own vehicles will stay on them.

The 4 parameters are shown in Figure 3, and they represent the probability of going home directly, going to the station, wandering around, and going home after staying at the station, respectively. The action "wandering around" represents actions such as going to nearby stores to buy food, moving to nearby buildings for shelter etc.

Regarding the decision of whether to go home directly by foot or not, previous research has focused on the relation of the decision with the age of the individual and also the distance to his/her home [Osaragi]. Therefore in our method, we multiple a coefficient on parameter 1 to get the revised probability of going home, which considers the age and distance to home of the individual. Based on previous work, we set the coefficient to the value plotted on figure 4.

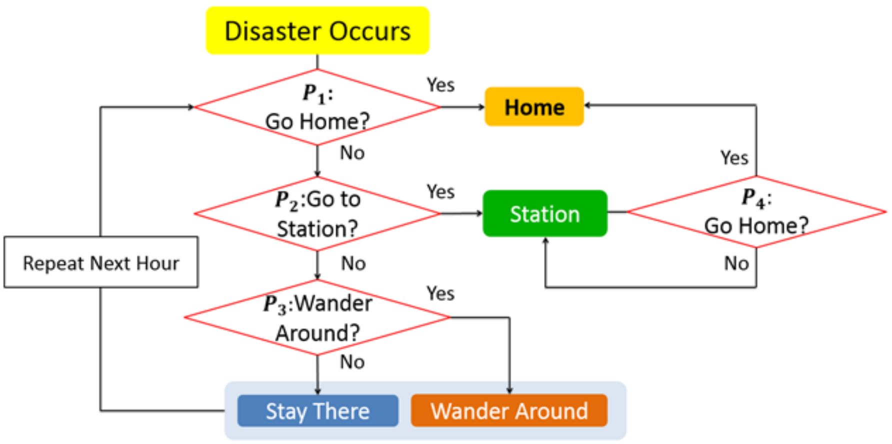

Figure 3 Behavioral Model

From figure 4, we can see that in all age groups, people located over 20 kilometers from their homes will not try to go home directly by foot.

For each parameter, we assign discrete values also shown as examples in table 1, and we generate different scenario candidates by every combination of the parameter values. For instance, if all parameters have 4 discrete values, $4^4 = 256$ scenario candidates will be generated.

As previous work [Wako *et al.*] mentions with actual data as proof, people's movement differs greatly from that of daily activities. Also it is also suggested that people movement under disaster situations is predictable since most people make movements that are rare but taken sometime before in their lives, such as going to their relatives' or friends' homes [Lu *et al.*]. By generating many scenario candidates and covering a wide range of movements after disaster in our method, we have a high possibility of generating a scenario that reflects the reality well, which is difficult to guess right with only one scenario.

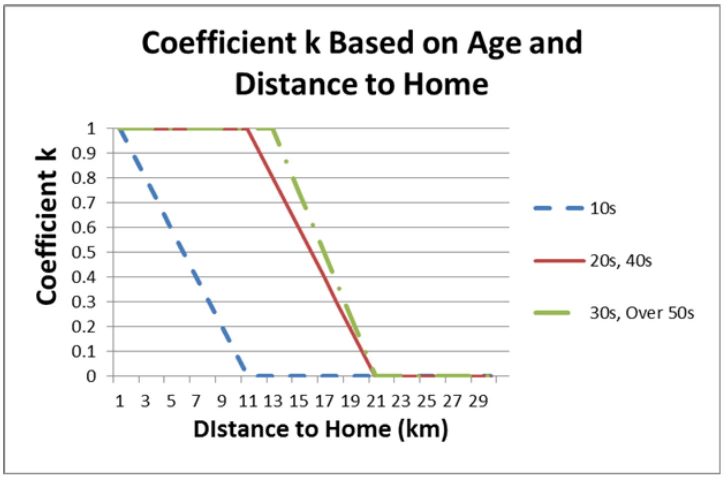

Figure 4 Coefficient k Based on Age and Distance to Home

3.3 Comparison with Real Time Observed Data

In the next step, we compare the scenario candidates we made with the real-time observed data. Any type of dataset can be used as the real-time observed data, but in this method we assume the usage of "Density Map Data" (DMD) provided by Zenrin Data Com co. which is a set of real-time population distribution data of all the areas in Japan. DMD is produced from GPS data of mobile phone users who have agreed in providing their location information as samples. In DMD, the population in each 250 meter-square-mesh is shown, so we aggregate the population in each of the scenario candidates into the same meshes for comparison.

To evaluate the error of each scenario candidate from the real-time observed data, we calculate the root mean squared error for each scenario candidate by the formula (i),

$$RMSE = \sqrt{\frac{\sum_i (V_i - V_i')^2}{n}} \tag{i}$$

where n and i represent the number of meshes and the mesh number respectively, and V_i, V_i' represent the simulated population and observed population in mesh i respectively.

Using the RMSE values, we calculate the likelihood from formula (ii).

$$L_i = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(RMSE)^2}{2\sigma^2}\right) \qquad \text{(ii)}$$

where L_i is the likelihood for the scenario candidate i, σ is 25% of the average population per mesh in that area.

Then, to earn the optimum parameter, we calculate the weighted average of the parameters in all of the scenario candidates using the likelihood, from formula (iii).

$$(p_1', p_2', p_3', p_4')^T = \begin{pmatrix} p_1^1 & p_1^2 & \cdots & p_1^n \\ p_2^1 & p_2^2 & \cdots & p_2^n \\ p_3^1 & p_3^2 & \cdots & p_3^n \\ p_4^1 & p_4^2 & \cdots & p_4^n \end{pmatrix} * (L_1', L_2', L_3' \ldots, L_n')^T \qquad \text{(iii)}$$

This way, even though our parameters are discrete values, our optimum parameters are continuous values.

3.4 Predicting with Optimum Parameter

Finally, in the fourth step, we run the simulation until several hours ahead using the optimum parameter set to predict the future people movement during the disaster situation. We repeat steps two to four, every time observation data is obtained to reflect the real-time movement of the people.

We evaluated our predictions by comparing them to the observed DMD of the same day and calculating their correlation coefficients.

4. Experiment

To verify our method, we carried out an experiment to simulate the people movement in the Great East Japan Earthquake which occurred on 14:47 of 11th March, 2011. The earthquake stopped all public transportation until midnight of the 11th of March in Metropolitan Tokyo, and over 5.5 million people could not return home. Although direct damage from the earthquake was barely done in Tokyo, confusion and secondary damage caused by congestion and crowdedness had a huge impact.

Figure 4 shows the 2 areas of study: 3 central wards in metropolitan Tokyo (Shinjuku, Chiyoda, Bunkyo wards) in blue, and Fujisawa City in Kanagawa prefecture in green, which is located around 50 kilometers from central Tokyo.

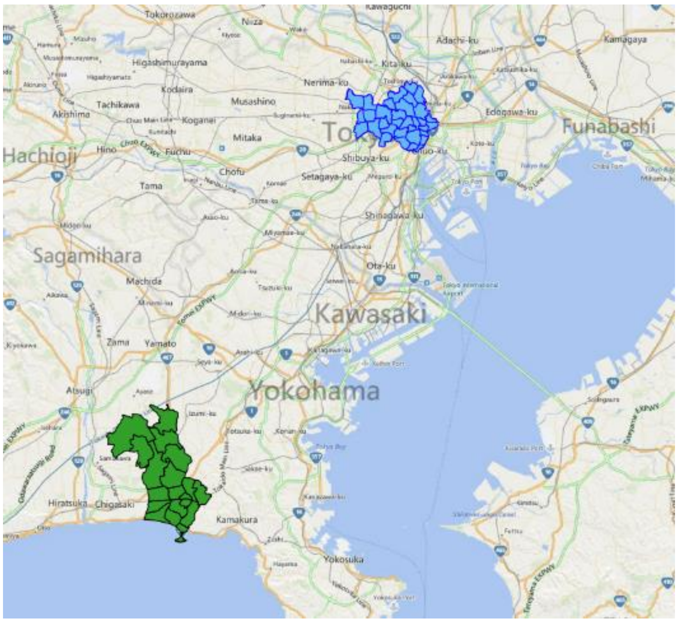

Figure 5 Two Areas of Experiment

We carried out experiments on these 2 areas to see the difference of people movement depending on the size and population density of the city. For the experiment, we used the 2008 Tokyo Metropolitan PTS, building data map and DMD provided by Zenrin Data Com co., and the Digital Road Map provided by DRM.

We were able to produce the primary location data for each of the areas with a very high accuracy, with a correlation coefficient of 0.996 and 0.997 in central Tokyo and Fujisawa respectively when evaluated in 250 meter-square-meshes against the DMD of ZDC.

Central Tokyo area contains 3 million people at daytime, and 92% of them are located at their “goal” location of the day, for example workplace or school, and only 16.5% of them are at home. This is reasonable since many people live in the outskirts of Tokyo and most of them are coming to central Tokyo for work at daytime.

On the other hand, Fujisawa city contains 0.4 million people and 46% of them are at home. The significant difference between the 2 areas, are very convincing, considering their locational and functional characteristics.

For the parameters, we set 4 discrete values {0, 0.2, 0.4, and 0.6} for each parameter. Also, we assumed that the real-time observation data can be obtained once in every hour. We simulated the people movement until 8 hours after the disaster, since after 8 hours, trains start to move in Tokyo, which we do not take in to account in this model. Therefore, we update the optimum parameter every hour in this experiment.

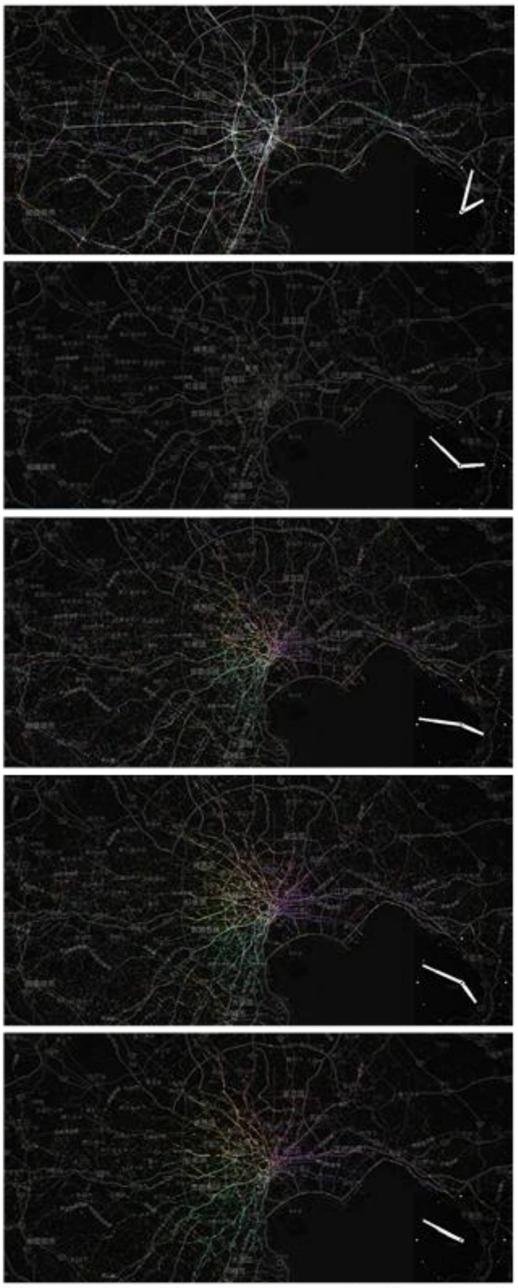

Figure 6 Movement of People in Metropolitan Tokyo after Earthquake

5. Results

Figure 6 shows an animated version of the result of the prediction in metropolitan Tokyo area. We can see that before the earthquake (1st frame, 14:05), the brightness is high meaning that the movement is very active in all areas. However, when the earthquake hit Japan (2nd frame, 14:47), we can see all the movements stop, darkening the whole metropolitan Tokyo, due to the confusion and shock. After 1 hour (3rd frame, 16:47), although slow in speed, people gradually start to move mainly in the central parts of Tokyo. More people start moving as time passes (4th frame, 17:47), and at night (5th frame, 22:47), we can clearly see that people are spreading from central Tokyo to the outskirts where their homes are. These movements represent the people who decided not to stay in central Tokyo and walk home instead.

Figure 7 shows the accuracy of prediction of each optimum parameter until 8 hours. For example, the blue line represents the accuracy of prediction if we keep on using the optimum parameter calculated after the 1st hour. We can see that, by updating the optimum parameters every hour, the accuracy of prediction is increasing in both areas. Therefore, we can conclude that the repetition of steps 2 to 4 are effective in this method. However, the prediction accuracy decreases as time passes even when we update the optimum parameters. This is because as time passes, more and more people take movements which are not covered by the behavioral model in our simulation. A good example for this is going home and then going to school to pick up their kids. We have not included additional trips after arriving at home, which is common in an emergency situation.

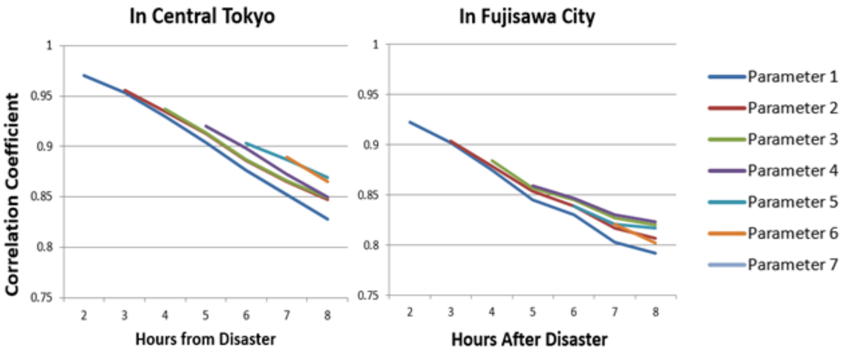

Figure 7 Accuracy of Prediction of all Optimum Parameter Sets

To relatively evaluate our method, we compared the results with 2 other methods, one of which is the daily people movement, and the other, the movement when we used the Poisson distribution to determine the parameters in the same behavioral model we used. The Poisson distribution is used in previous work [Osaragi] as parameter values in disaster scenario simulations. Figure 8 shows the comparison of the correlation coefficient and RMSE values of the 3 methods. We can see that compared to the other 2 methods, our proposed method has the highest correlation with the actual observed data, and the RMSE is also the smallest among them. By comparing the result of our method to the method where parameters were set according to the Poisson distribution, we can conclude that generating many scenario candidates and calculating the optimum parameter is effective in catching real-time movement especially in a situation where prediction is difficult.

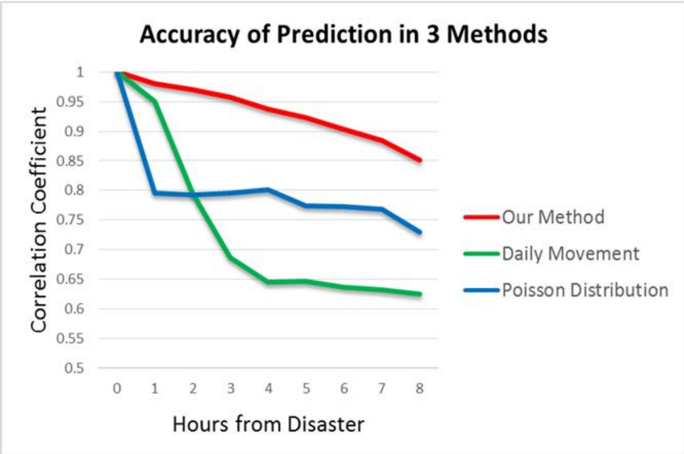

Figure 8 Accuracy of Prediction in 3 Methods

Figure 9 shows the number of people who moved to each destination in each hour in our optimum scenarios in central Tokyo area. We can see that as time passes, the number of people who head home directly by foot decreases, while the number of people heading to the nearby stations conversely increase. This result accurately reflects the peoples' actions on the day of the disaster, where many people tried to return home when it was early and bright, but hesitated to do so as it got darker. This result is also accurate compared to past research based on a survey, which mentions that

	To Home		To Station		Wandered Around	
Time	Number of People	%	Number of People	%	Number of People	%
15:47	428950	13.00	31950	0.97	132700	4.02
16:47	304400	9.22	45450	1.38	105200	3.19
17:47	261650	7.93	54300	1.65	91350	2.77
18:47	218300	6.62	69400	2.10	91400	2.77
19:47	204150	6.19	66950	2.03	100700	3.05
20:47	176900	5.36	71800	2.18	114250	3.46
21:47	160500	4.86	75000	2.27	123300	3.74
22:47	139650	4.23	83300	2.52	128850	3.90

Figure 9 Behaviors of People in Simulation of Central Tokyo

a total of 50% of the people started to head home by 20:30, where 48% of them did in our prediction.

Looking at the number of people wandering around, we can also see that the number of people increases right after the disaster occurs, and at nighttime. This is explainable because it can be assumed that right after the disaster many people evacuate to nearby buildings, and at night people would go shopping for food and commodities to spend the night away from home. By using a behavioral model in our simulation, we were able to analyze and give descriptions to the peoples' movement. Also, this allowed us to evaluate qualitatively that our method accurately simulated the movement of the people on the day of the Great East Japan Earthquake.

6. Discussion

In this paper, we proposed a method to predict people movement under disaster situations by combining multi agent simulation based on a behavioral model with real-time observation data. Our behavioral model considered personal characteristics when determining actions of individuals, which added persuasiveness to our model and output. In order to predict people movement which is difficult compared to daily movement, we generated many scenario candidates and calculated the optimum parameter set based on the comparison of each candidate with the observed real-time data. This allowed us to cover a wide range and variety of movement of the people, which is needed when predicting an uncertain situation like post disaster times.

Based on the experiment on the Great East Japan Earthquake, we verified the effectiveness of our method compared to other methods, and also confirmed the effectiveness of updating the parameter every time observation data is obtained. Also, by looking at the number of people by destination and comparing it to the results of a survey, we confirmed the accuracy of the method qualitatively.

However, there are some work left for the future. Firstly, the behavioral model we used in our method does not include trips after arriving at home in order to keep the model simple and keep the number of parameters less. As we have seen in the results, the prediction accuracy lowers as time passes, and there is a need to consider actions after arriving home especially as more people start arriving home.

Also, when predicting the people movement of hours ahead, in this method we used the optimum parameter set calculated each hour, and did not consider the optimum parameter sets calculated before when updating every hour. Although we were able to predict with high accuracy in our current method, applying a new method for updating optimum parameters using not only the newest optimum parameter but also the ones before would be an interesting topic.

7. Conclusion and Future Work

In the Great East Japan Earthquake, Tokyo experienced great congestion and confusion mainly caused by the stoppage of public transportation. With the rapid of growth of urban areas in many countries, and continuous if not increasing risk of damage caused by natural disasters, the need of disaster preparation technologies is rising. An accurate prediction of people movement in disaster situations would give decision makers quantitative grounds for making decisions and policies rather than making assumptions or experience based judgments.

In this paper, we introduced a method to accurately predict people movement in disaster situations using real-time observation data. Through an experiment, we confirmed the effectiveness of the method in an earthquake situation, and future work would include expanding the application of the method to other disasters such as typhoons and snowstorms. Also, due to the damage from the disaster, some data might become fragmented. A method to deal with partial observation data would be needed to make this method a more practical one.

Regarding the issues raised in section 6, future work would also involve the increase of options of movement for especially people who have once arrived at home. Also, considering a new algorithm for updating optimum parameters would be an interesting topic for future research.

ACKNOWLEDGEMENTS

We would like to appreciate Zenrin Data Com for providing us with the Density Map Data of the day of the earthquake.

REFERENCES

Ashbrook, Daniel, and Thad Starner. (2003) "Using GPS to learn significant locations and predict movement across multiple users." Personal and Ubiquitous Computing 7.5: 275-286.

Calabrese, Francesco, et al. (2011) "Estimating origin-destination flows using mobile phone location data." IEEE Pervasive Computing : 36-44.

Chen, Zhai, Madey (2011) “Dynamic Adaptive Disaster Simulation: Developing a Predictive Model of Emergency Behavior Using Cell Phone and GIS Data”, Proceedings of the 2011 Workshop on Agent-Directed Simulation

Herrera, Juan C., et al. (2010) "Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment." Transportation Research Part C: Emerging Technologies 18.4: 568-583.

D’Orazio, L. Spalazzi, E. Quagliarini, G. Bernardini, (2014) “Agent-based model for earthquake pedestrians’ evacuation in urban outdoor scenarios”: Behavioral patterns definition and evacuation paths choiceSafety Science, 62, pp. 450–465

Hiroi, U., et al. (2011) "Questionnaire Survey Concerning Stranded Commuters in Metropolitan Area in the Great East Japan Earthquake." The Annals of Institute of Social Safety Science 15: 343-353.

Horanont, Teerayut, et al. (2013) "Weather effects on the patterns of people's everyday activities: a study using GPS traces of mobile phone users." PloS one8.12: e81153.

Ito, Aono, Oomori (2013) “Study on Peoples movements in Metropolitan Tokyo Based on Surveys” Journal of The City Planning Institute of Japan

Kanasugi, Hiroshi, et al. (2013) "Spatiotemporal route estimation consistent with human mobility using cellular network data." Pervasive Computing and Communications Workshops (PERCOM Workshops), 2013 IEEE International Conference on. IEEE.

Levi, Lion, et al. (1998) "Hospital disaster management simulation system." Prehospital and disaster medicine 13.01: 22-27.

Madey, et al. (2006) "WIPER: A multi-agent system for emergency response." Proceedings of the 3rd International ISCRAM Conference.

Osaragi, Toshihiro. (2013) "Modeling a spatiotemporal distribution of stranded people returning home on foot in the aftermath of a large-scale earthquake." Natural hazards 68.3: 1385-1398.

Ryo Wako, Yoshihide Sekimoto, Hiroshi Kanasugi, Ryosuke Shibasaki (2014) "Analysis of People's Route and Destination Choice in Evacuation Using GPS Log Data", Journal of Japan Society of Civil Engineers

Sekimoto, Y., Shibasaki, R., Kanasugi, H., Usui, T., & Shimazaki, Y. (2011). Pflow: Reconstructing people flow recycling large-scale social survey data.IEEE Pervasive Computing, 10(4), 0027-35.

Song, Xuan, et al. (2013) "Modeling and probabilistic reasoning of population evacuation during large-scale disaster." Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013.Song

Xin Lu, Bengtsson, Holme, (2011) "Predictability of population displacement after the 2010 Haiti earthquake", PNAS vol. 109, no.29 p.11581

Yoko Hasegawa, Yoshihide Sekimoto, Takehiro Kashiyama, Hiroshi Kanasugi, (2014) "Transportation Melting Pot Dhaka: Road-link Based Traffic Volume Estimation from Sparse CDR Data", The 1st International Conference on IoT in Urban Space (Urb-IoT 2014), Rome