https://doi.org/10.20965/jdr.2019.p0521

Paper:

Analysis of Evacuation Trajectory Data Using Tensor Decomposition

Yusuke Kawai†, Yoshiharu Ishikawa, and Kento Sugiura

Graduate School of Informatics, Nagoya University
Furo-cho, Chikusa-ward, Nagoya 464-8601, Japan
†Corresponding author, E-mail: kawai@db.is.i.nagoya-u.ac.jp
[Received November 7, 2018; accepted March 2, 2019]

Owing to the advances in information technology and heightened awareness regarding disaster response, many evacuation simulations have been performed by researchers in recent years. It is necessary to develop suitable disaster prevention plans or evacuation plans using data generated by such simulations. However, it is difficult to understand the simulation results in their original form because of the detailed and voluminous data generated. In this study, we focus on tensor decomposition, which is employed for analyzing multi-dimensional data, in order to analyze the evacuation simulation data, which often consists of multiple dimensions such as time and space. Tensor decomposition is applied to the movement trajectory data generated in the evacuation simulation with the objective of acquiring important disaster or evacuation patterns.

Keywords: evacuation simulation, trajectory data, tensor decomposition

1. Introduction

The analysis of the damage status based on disaster simulations is expected to be an important disaster prevention measure in the area of disaster response. The analysis of the damage status can contribute to the determination of evacuation paths and disaster response plans, or the improvement of urban infrastructure. With the improved performance and increased storage capacity of computers in recent years, a vast amount of simulation data is being generated.

However, there have been relatively fewer studies on the development of technology related to the use of large datasets, and thus, although a large number of simulations have been conducted, their results have not been utilized effectively. Although there have been studies on the modeling of disaster simulations [1, 2], the effective usage of the simulation results remains an issue. It is necessary to understand the results and analyze the influencing factors of a simulation in order to obtain guidelines for new simulations or new findings. For instance, in evacuation simulations, it is necessary to identify the factors that hinder evacuation [3]. Various approaches [4–7] can be used to assist in such an analysis.

In this study, our objective is to develop a technology for analyzing disaster simulation data. Specifically, this paper is focused on the analysis of evacuation simulation data. Evacuation simulations are considered to have a high degree of uncertainty because the behavior patterns vary depending on the specific disaster conditions. Furthermore, the generated data are characterized by many attributes including temporal and spatial information. The classification of such highly uncertain results into characteristic patterns while taking into consideration the multiple attributes would be useful in the factor analysis of the simulation results.

In this study, we employ tensor decomposition as the analytical tool. In tensor decomposition, based on the assumption of low rank, i.e., that the tensor elements can be approximately decomposed into small groups (called ranks), the tensor is decomposed as a product of low-rank matrices (called factor matrices). As high-dimension tensors can be decomposed into low-dimension ones while retaining the characteristic information contained in the original data, tensor decomposition is applied in various fields such as noise processing and feature extraction of images or audio data [8, 9]. Furthermore, as high-dimension tensors can be used to comprehensively examine multi-dimensional, complex information and classify it into groups, it is expected to produce findings that are different from those based on simple piece-wise analyses and is used, for instance, in item recommendation based on purchase histories with multiple attributes [10]. In this study, we apply tensor decomposition to the movement trajectory data generated by an evacuation simulation in order to identify important patterns of disasters or evacuation.

This paper is structured as follows. Section 2 describes the dataset employed. Section 3 discusses the method of tensor decomposition. Section 4 describes the analysis used for the dataset. Section 5 presents the conclusion of our study and the future scope of study.

2. Evacuation Simulation Data

Section 2.1 describes the evacuation simulation data used in this study, and Section 2.2 presents an outline of the collated simulation results.

Table 1. Evacuation data.

Attribute	Meaning	Remarks
t	Time	In 0.1-min intervals over a 24-h period
uid	Person ID	13,487 people (serially numbered from 1)
x, y	X, Y coordinates	Japan Plane Rectangular Coordinate System IX
status	Status of person	See **Table 4**.

Table 2. Building damage data.

Attribute	Meaning	Remarks
t	Time	Same as **Table 1**
uid	Building ID	4,392 units
collapsed	Collapsed or not	0: Not collapsed; 1: Collapsed
stage	Fire stage	0: Fire not started; 1: Possibility of catching fire 2: Burning; 3: Extinguished

Table 3. Road damage data.

Attribute	Meaning	Remarks
t	Time	Same as **Table 1**
uid	Road ID	2,720 road sections
blocked	Blockage status	0: Unblocked; 1: Blocked
dens	Population density	Population density

Table 4. Person status.

Status	Meaning
StayHome	Present at home when earthquake occurs
StayOther	Present at some building when earthquake occurs
Move	Walking outdoors when earthquake occurs
StayIttoki	Staying at temporary site
Safe	Staying at wide-area evacuation center
Search	Searching for evacuation path
HinanIttoki	Moving toward temporary site
HinanIttoki2	Arrived at temporary site
HinanKoiki	Moving toward wide-area evacuation center
HinanKoiki2	Arrived at wide-area evacuation center
Rescue	Participating in rescue activities
Fire	Participating in fire-fighting activities
Konnan	Trapped in urban area
Death	Those killed

2.1. Dataset

The simulation data was provided by the Osaragi Laboratory, Tokyo Institute of Technology. A wide-area evacuation simulation [2] was conducted based on the Tokyo Inland Earthquake. The occurrence of a 6-plus seismic-intensity earthquake is simulated with its epicenter located in the northern part of Tokyo Bay, with the simulation focus being on the Kitasenju district of the Tokyo metropolitan area. This district has a high risk of fires and building collapse owing to its high concentration of old wooden buildings, and it is difficult to conduct a wide-area evacuation because the district is surrounded by the Arakawa and Sumidagawa Rivers, thus making it an extremely high-risk area at times of disaster.

In this simulation, it is assumed that, in a situation in which multiple fires have occurred in several buildings following the earthquake, the people in the area move toward wide-area evacuation centers. The people select actions amidst the occurrence of building collapses, road blockages, and fires spreading from building to building. Those able to move temporarily gather at temporary evacuation sites and then move on to the wide-area evacuation centers. In addition, depending on their circumstances, some people engage in rescue or fire-fighting activities.

The simulation data consists of the evacuation behavior (**Table 1**) of 13,487 people and damage status of 4,392 buildings (**Table 2**) and 2,720 roads (**Table 3**) during the 24-h period following the earthquake. The evacuation behavior data consists of the status-and-trajectory data including the individuals' geographic coordinates and their evacuation statuses (classified into 14 types; **Table 4**) at given times following the earthquake. The building and road data record show the status of fires and road blockages.

2.2. Collation of Results

In this section, we perform a simple collation of the evacuation behavior data in order to obtain an outline of

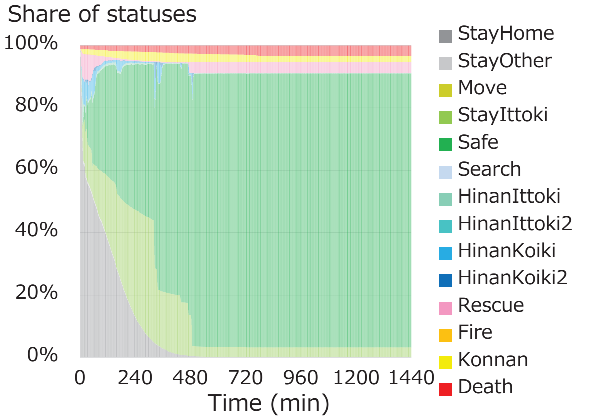


Fig. 1. Share of statuses during 24-h period following earthquake.

the simulation results. First, **Fig. 1** shows the collated results of the people's status over the 24 h following the earthquake. The abscissa and ordinate, respectively, represent the time (min) and the percentage of the number of people corresponding to various states.

It can be observed that, with the passage of time, the number of people present at home decreases, while the number of people who have arrive at the wide-area evacuation centers increases. On route, many of them temporarily evacuated to temporary sites. Thus, it can be observed that many of those staying home first move to the temporary sites and then move on to the wide-area evacuation centers. The number of those who completed the evacuation increases significantly approximately 5–6 h and 8 h after the occurrence of the disaster, and this situation remains mostly unchanged subsequent to 8 h after the disaster. Moreover, the statuses changes significantly approximately 1 h after the earthquake.

A certain number of people are killed at the time of the earthquake, following which this number increases gradually. The number of people trapped also increases gradually, then subsequently decreases, from which it can be speculated that some of them die during this time. Meanwhile, a certain number of people engage in rescue activities immediately after the earthquake and after the majority of people have been evacuated.

From the above observations, we see that the evacuation behavior data changes very little in the period that follows 10 h after the earthquake, and thus, we subject the data for up to 10 h to analysis. We thus collated the data of people who travelled since the earthquake occurred up to 10 h later. Those whose minute-by-minute position coordinates changed were considered as those who travelled, and we collated their number (hereafter "number of moving people") and their average travel distance per minute (hereafter "average travel speed"). The number of moving people (blue curve in upper graph) and the average travel speed (red curve in upper graph) are shown along with the percentages of statuses (lower graph) in **Fig. 2**. The abscissas represent the time (min), the left ordinate in the

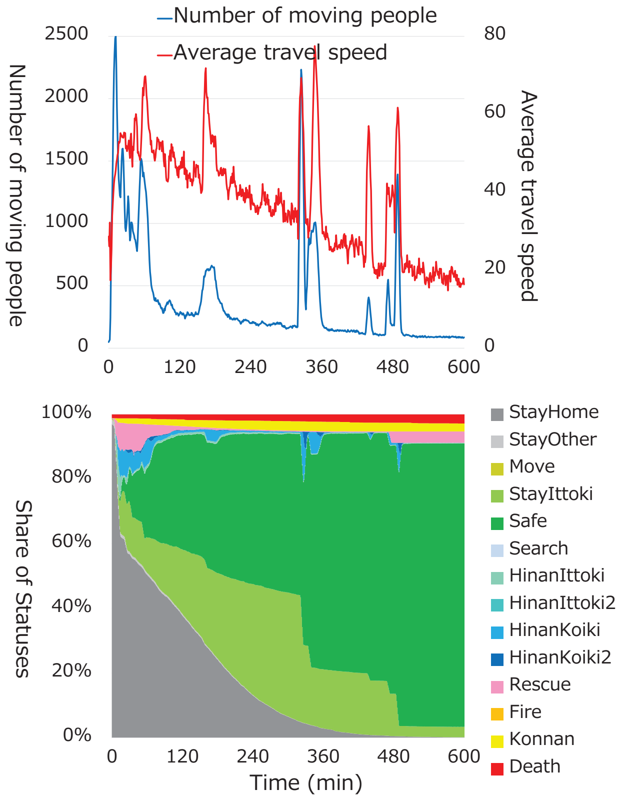
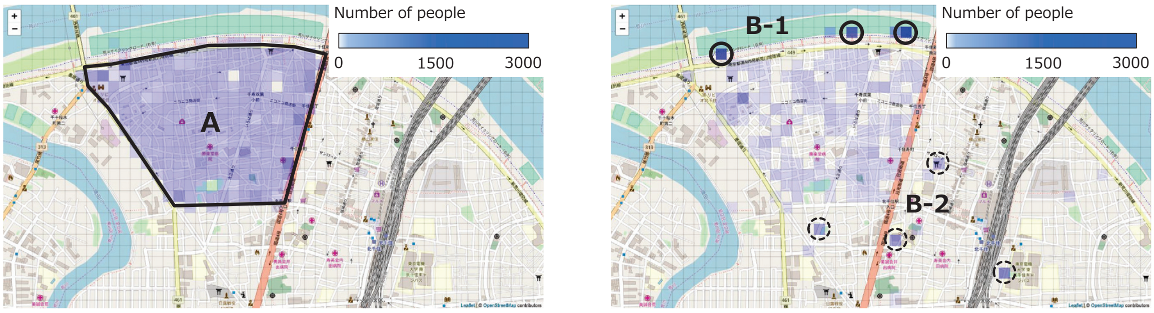


Fig. 2. Movement of people, travel speed, and share of statuses for the 10-h period.

upper graph the number of moving people, and the right ordinate the travel speed (m/min), while the ordinate of the lower graph represents the percentages of the statuses relative to the total number of people.

The times at which the statuses display considerable changes coincide with peaks in the number of moving people and travel speed. The average speed falls with the passage of time, except at the peaks. Although the travel speed is low immediately after the earthquake, the number of moving people is high. The maximum number of moving people is approximately 2,500, which is approximately 19% of the total number of people. The maximum travel speed is approximately 80 m/min, which decreases to 20 m/min subsequent to 8 h after the earthquake. Considering that travel for evacuation no longer takes place as the majority of people have completed evacuation, it can be speculated that this figure indicates the movements within restricted ranges of those engaged in rescue activities or those trapped in buildings.

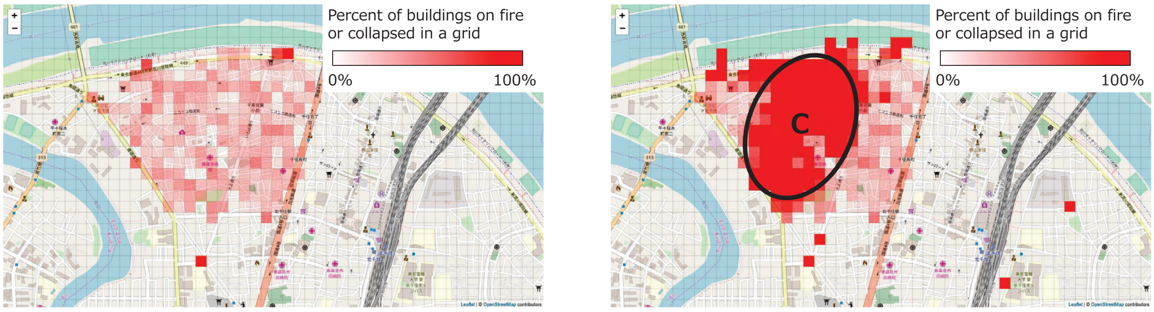
We next render a visual of the statuses of people, buildings, and roads. As the evacuation simulation is based on the actual location of Kitasenju district, visualization can be presented using a map. Leaflet [11] was used for visualization on the map. Leaflet is an open-source JavaScript library that can be used to display Web maps on a browser. It allows one to set up objects such as markers, lines, and polygons on the map. The objects can be set up as popups and embedded with information. Layers can also be



(a) At the time of earthquake occurrence

(b) 10 h after the earthquake occurrence

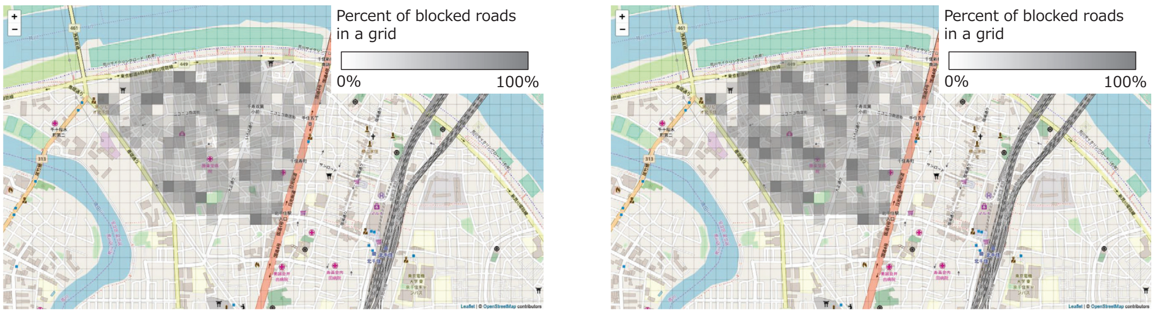
Fig. 3. Number of people in target zone and vicinity.



(a) At the time of earthquake occurrence

(b) 10 h after the earthquake occurrence

Fig. 4. Geographical locations of building fires and collapses in target zone and vicinity.



(a) At the time of earthquake occurrence

(b) 10 h after the earthquake occurrence

Fig. 5. Geographical locations of road blockages in target zone and vicinity.

introduced, where different layers can be switched to display different maps and objects.

The locations at which people are present at the time of the earthquake and 10 h later are shown in **Fig. 3**. The target area was divided into 50 m $\times$ 50 m grids, in each of which, the number of people was collated, the results of which are displayed using different shades of blue. It can be observed that people were scattered widely within the zone indicated by "A" at the time the occurrence of the disaster. It can also be observed that the people have evacuated to several sites 10 h later when evacuation was completed. The majority of the people evacuated to the three locations (labeled B-1) in the upper part of the map; they make up over 80% of the population considered. It can also be observed that some people evacuated to parks and schools (labeled B-2) located southeast of the target zone. Although low in number, there are also people still remaining in the target zone, who are considered to be engaged in rescue activities, trapped within collapsed buildings, or dead.

The statuses of buildings and roads at the time of the earthquake occurrence and 10 h later are shown in **Figs. 4** and **5**, respectively. The area was divided into grids as shown in **Fig. 3**, and the damage status of the buildings and roads are indicated by different shades of color. The grids with a high proportion of buildings that have caught fire or collapsed are indicated by red, while those containing many road blockages are indicated by grey. It can be observed that the damaged buildings and roads both increase with the passage of time. In particular, the buildings become extensively damaged, which indicates that fires have spread out centered in the zone labeled C.

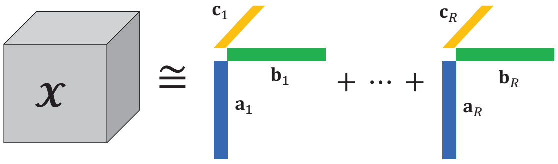


Fig. 6. CP decomposition.

Meanwhile, the roads tend to become blocked immediately after the earthquake, and the change in their damage status is low in comparison to the spreading of fires among buildings and the collapse of buildings.

3. Tensor Decomposition

In this section, we describe the method of Tensor decomposition applied to the dataset. Here, we decompose a third-order tensor into three low-rank matrices (called the factor matrices). We employ Non-negative Tensor Factorization (NTF) [8] in this study. In this method, CANDECOMP/PARAFAC (CP) decomposition (**Fig. 6**) is applied to a tensor while maintaining the non-negativity constraint. After the tensor has been decomposed, the elements are reconstructed from the factor matrices in conformance with the original tensor.

In Section 3.1, we describe the application of NTF to a third-order tensor. In Section 3.2, we describe how to reconstruct the elements of the tensor produced by decomposition.

3.1. Non-Negative Tensor Factorization

In NTF, a third-order tensor $\mathscr{X} = [x_{ijk}] \in \mathbb{R}^{I \times J \times K}$ is decomposed into three factor matrices of rank R, $A = [a_{ir}] \in \mathbb{R}^{I \times R}, B = [b_{jr}] \in \mathbb{R}^{J \times R}$ and $C = [c_{kr}] \in \mathbb{R}^{K \times R}$. The tensor produced from this decomposition $\hat{\mathscr{X}} = [\hat{x}_{ijk}] \in \mathbb{R}^{I \times J \times K}$ is expressed as the product of the three factor matrices. The elements $\hat{x}_{ijk}$ of the tensor are expressed as follows:

$$\hat{x}_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} \quad \ldots \ldots \ldots \ldots \quad (1)$$

A, B and C are determined in order to minimize the error between the original tensor $\mathscr{X}$ and the tensor after the decomposition of $\hat{\mathscr{X}}$. The sum of the errors of the corresponding elements of the tensors is used as the error.

$$\begin{aligned} \underset{A,B,C}{\text{minimize}} \quad & \mathscr{D}(\mathscr{X} | \hat{\mathscr{X}}) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} d(x_{ijk} | \hat{x}_{ijk}) \\ \text{subject to} \quad & A, B, C \geq 0 \end{aligned} \quad (2)$$

where the element-wise error $d(x_{ijk} | \hat{x}_{ijk})$ is the generalized Kullback-Leibler divergence:

$$d(p|q) = p(\log p - \log q) - (p - q) \quad \ldots \ldots \quad (3)$$

As it is difficult to simultaneously determine all of the variables, the local solution of a single variable is estimated while the two other variables are kept constant, and this procedure is repeated. In other words, the solution is updated in order to make the partial differential of the error $\mathscr{D}$ equal zero. The final update expressions of the factor matrices are as follows:

$$a_{ir}^{\text{new}} = a_{ir} \frac{\displaystyle\sum_{j=1}^{J} \sum_{k=1}^{K} \frac{x_{ijk}}{\hat{x}_{ijk}} b_{jr} c_{kr}}{\displaystyle\sum_{j=1}^{J} \sum_{k=1}^{K} b_{jr} c_{kr}} \quad \ldots \ldots \ldots \quad (4)$$

$$b_{jr}^{\text{new}} = b_{jr} \frac{\displaystyle\sum_{i=1}^{I} \sum_{k=1}^{K} \frac{x_{ijk}}{\hat{x}_{ijk}} a_{ir} c_{kr}}{\displaystyle\sum_{i=1}^{I} \sum_{k=1}^{K} a_{ir} c_{kr}} \quad \ldots \ldots \ldots \quad (5)$$

$$c_{kr}^{\text{new}} = c_{kr} \frac{\displaystyle\sum_{i=1}^{I} \sum_{j=1}^{J} \frac{x_{ijk}}{\hat{x}_{ijk}} a_{ir} b_{jr}}{\displaystyle\sum_{i=1}^{I} \sum_{j=1}^{J} a_{ir} b_{jr}} \quad \ldots \ldots \ldots \quad (6)$$

Algorithm 1, which is based on these expressions, is presented below. The factor matrices are updated until the error between the original tensor and the tensor produced from the estimated factor matrices converge.

Algorithm 1 Estimation of factor matrices

Input: $\mathscr{X}, R, E$
Output: A, B, C
  initialize ε
  initialize A, B, C with small random values
  while $\varepsilon \geq E$ **do**
    update $A = [a_{ir}]$ by Eq. (4)
    update $B = [b_{jr}]$ by Eq. (5)
    update $C = [c_{kr}]$ by Eq. (6)
    update $\hat{\mathscr{X}} = [\hat{x}_{ijk}]$ by Eq. (1)
    $loss^{\text{new}} = \mathscr{D}(\mathscr{X} | \hat{\mathscr{X}})$
    $\varepsilon = loss^{\text{old}} - loss^{\text{new}}$
    $loss^{\text{old}} := loss^{\text{new}}$
  end while

3.2. Reconstruction of Elements

As there is a large error between the tensor $\hat{\mathscr{X}}$ generated from the estimated A, B, and C, and the original tensor $\mathscr{X}$, without modification, $\hat{\mathscr{X}}$ cannot be compared with the original data. Therefore, by considering the estimated tensor $\hat{\mathscr{X}}$, a tensor $\bar{\mathscr{X}}$ is produced in which the elements are reconstructed such that they have the same form as the original tensor $\mathscr{X}$. $\bar{\mathscr{X}} = [\bar{x}_{ijkr}] \in \mathbb{R}^{I \times J \times K \times R}$ is a fourth-order tensor in which the rank is added to the dimensions of the original tensor. The elements $\bar{x}_{ijkr}$ of $\bar{\mathscr{X}}$ are normalized with respect to the elements a_{ir}, b_{jr}, and c_{kr}, which are reconstructed for each rank, such that their sum in the direction of the rank is equal to that of the elements x_{ijk} of the original tensor. Thus, we obtain the following expression.

$$\bar{x}_{ijkr} = x_{ijk} \frac{a_{ir} b_{jr} c_{kr}}{\sum_{r=1}^{R} a_{ir} b_{jr} c_{kr}} \quad (7)$$

The actual analysis is performed using $\bar{\mathscr{X}}$.

4. Analysis Scenario

In this section, we generate a tensor from the evacuation simulation data described in Section 2 and subject it to analysis. We perform the tensor decomposition described in Section 3 and conduct the analysis using the tensor $\bar{\mathscr{X}} = [\bar{x}_{ijkr}] \in \mathbb{R}^{I \times J \times K \times R}$, which is decomposed according to rank. While the evacuation simulation data can be subjected to various analyses, we focus on the person movement trajectory data (**Table 1**) as an example and analyze the statuses and movement of people.

Here, we conduct an exploratory analysis based on visualization. Visualization is a valid method for analyzing large-scale and detailed datasets as it allows one to understand the situation visually without directly referring to the numerical data. Visualization on the map is prepared using Leaflet [11] as mentioned in Section 2.

As the evacuation simulation data is closely tied to spatio-temporal patterns, we select the time and geographical location as essential dimensions in the tensor design. We also use a third dimension of the people's behavior and direction of movement, to produce two types of tensors – time $\times$ location $\times$ people's status and time $\times$ location $\times$ direction of movement – and decompose and analyze these tensors.

The method of determining the rank R is important in analyses that employ tensor decomposition. However, there is no standard for determining a suitable R, and it is determined depending on the characteristics of the tensor or the purpose of analysis [10, 12]. As there have been no previous works on the application of tensor decomposition to evacuation simulation data, we perform a coarse analysis to examine how the evacuation patterns are divided. The lowest dimensional sizes of the two types of tensors produced are 14 and 16, which are quite small. Thus, we use $R = 5$ in the present analysis.

4.1. Analysis of Status

4.1.1. Tensor Construction

We construct a third-order tensor $\mathscr{X} = [x_{ijk}] \in \mathbb{R}^{I \times J \times K}$ comprising the people's statuses obtained from the travel trajectory data. The dimensions are time, location, and person status, and the elements represent the corresponding numbers of people. In other words, the tensor contains information on how many people there are "when," "where," and "in what condition." The data are collated for time units of minutes and for geographical units of 50 m $\times$ 50 m grids. As the original data is recorded in 0.1-min intervals, we employ the average of ten position coordinates and the person statuses recorded at integer minutes. The sizes of the dimensions are $I = 601$, $J = 1770$ and $K = 14$, each consisting of non-negative integer-valued elements. There are 278,493 non-zero elements, which make up approximately 1.9% of the total elements.

4.1.2. Tensor Decomposition

We perform tensor decomposition for the rank $R = 5$. We observe the time, position, and person status for the five ranks and examine the results. From $\bar{\mathscr{X}}$, the time $\times$ rank matrix $\bar{A}$, location $\times$ rank matrix $\bar{B}$, and status $\times$ rank matrix $\bar{C}$ are obtained as follows:

$$\bar{A} = [\bar{a}_{ir}] = \sum_{j=1}^{J} \sum_{k=1}^{K} \bar{x}_{ijkr} \quad (8)$$

$$\bar{B} = [\bar{b}_{jr}] = \sum_{i=1}^{I} \sum_{k=1}^{K} \bar{x}_{ijkr} \quad (9)$$

$$\bar{C} = [\bar{c}_{kr}] = \sum_{i=1}^{I} \sum_{j=1}^{J} \bar{x}_{ijkr} \quad (10)$$

$\bar{C}$ is presented using a cumulative bar chart in **Fig. 7**. Rank 2 (orange) comprises a large portion of deaths and people engaged in rescue activities. It is also high in the number of people who have completed evacuation. Rank 3 (green) is high in the number of people who are trapped. Rank 4 (red) is high in the number of people at home before they evacuated. It is also relatively high in the number of people engaged in rescue activities. Ranks 1 (blue) and 5 (purple) make up a certain portion of those staying home, those at temporary sites, and those who have completed evacuation. As these ranks are absent among the deaths and those engaged in rescue activities, it can be speculated that these ranks include a high portion of the number of people who have evacuated safely.

Next, $\bar{A}$ is presented using a cumulative bar chart in **Fig. 8**. Rank 4 decreases with the passage of time in agreement with the observation made above that it is high in the number of people at home before evacuation. Rank 5 increases in time with the decrease in rank 4, which indicates that the people who were staying home are making a transition to completion of evacuation. Furthermore, rank 5 decreases significantly at approximately 330 min, which suggests that it corresponds to those who evacuate relatively early. It is then replaced by rank 3, which drastically increases at approximately 330 min. Rank 2 does not appear immediately after the earthquake, but increases gradually up to 10 h later. This is thought to correspond to the increase in deaths, which was observed in the original dataset.

If we take **Fig. 7** into consideration, those engaged in rescue activities immediately after the earthquake correspond to rank 4, while those engaged in rescue activities after completion of evacuation correspond to rank 2. This is confirmed by **Fig. 9(a)**, which shows only the number of people engaged in rescue activities according to rank at different times. Meanwhile, the deaths are also

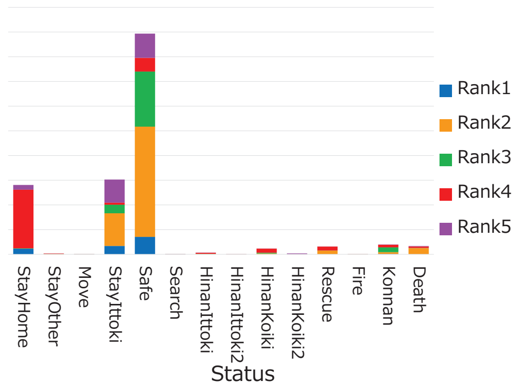


Fig. 7. Numbers of people in various statuses and their rank composition.

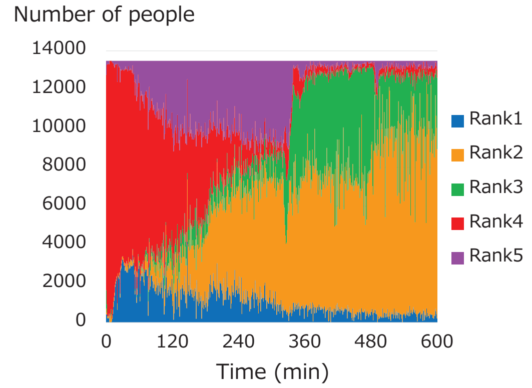
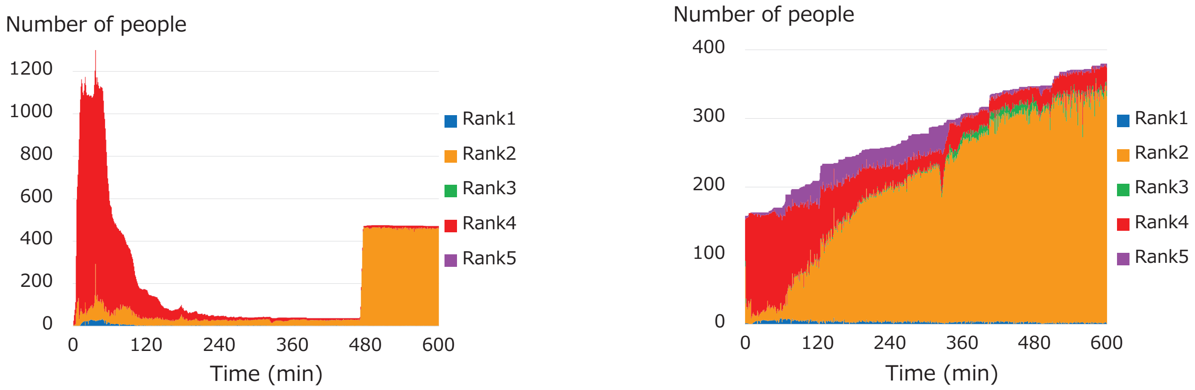


Fig. 8. Rank composition of people plotted against time.



(a) Rescue (b) Death

Fig. 9. Rank composition of people plotted against time.

represented by ranks 2 and 4, which correspond respectively with the deaths immediately after the earthquake and those that increase with the passage of time, as observed in **Fig. 9(b)**, which shows only the deaths according to rank at different times. As those who died should ideally have been rescued, rescue activities often take place at locations where deaths occur. Therefore, the finding that deaths and rescue volunteers are classified under the same ranks as the result of the tensor decomposition is a desirable outcome.

From the standpoint of simulation, deaths occurring immediately after the earthquake, as represented by rank 4, are closely connected to the locations of building collapse and fire occurrences and cannot be prevented by human actions immediately following an earthquake. Although it is important to undertake advance measures as those buildings are highly dangerous, it is difficult to deal with them with an evacuation simulation. Thus, in this case, it may be more relevant to focus on the increase in deaths as represented by rank 2.

While taking into consideration the above findings, we next observe the locations of people. **Fig. 10** shows the results of $\bar{B}$ on a map. Labels B-1 and B-2 are shown in **Fig. 3(b)**. Only the highest ranks for the grids are shown, where the darker shades indicate that the rank has a higher share. Those grids that contain very few people are omitted.

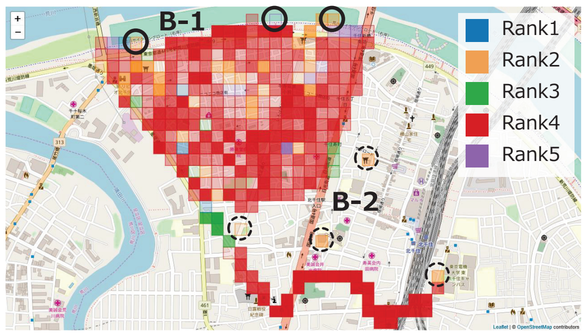


Fig. 10. Ranks of geographical locations.

It can be observed that rank 4 is dominant in a great majority of zones. Rank 4 is the dominant rank immediately after the earthquake and includes a large number of those staying home. As it decreases with the passage of time, we can speculate that people are moving from zones with a large share of rank 4 to other zones with a high proportion of some other rank. Among the latter zones, it is likely that the movement to zones of rank 5 occurs relatively early and that to zones 2 and 3 occurs relatively late. Rank 2 is found in those zones corresponding to labels B-1 and B-2, which indicates that the people evacuated to these wide-area evacuation centers. As this rank also cor-

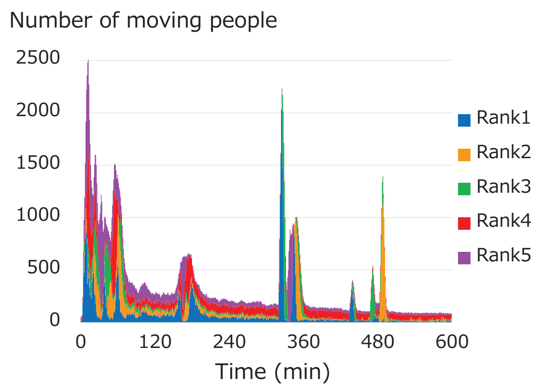


Fig. 11. Rank composition of people plotted against time.

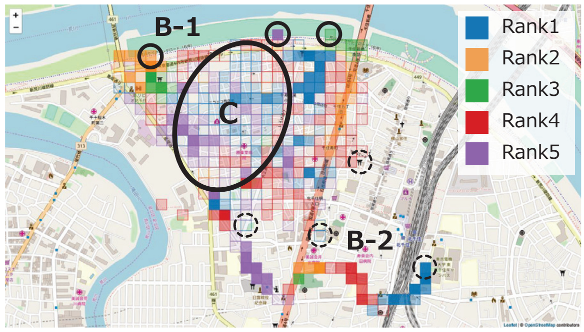


Fig. 12. Ranks of geographical locations.

responds to the increase in deaths and to those engaged in rescue activities after completing evacuation, we can state that the zones represented by rank 2 other than the evacuation centers require attention. Thus, by subjecting the dataset to tensor decomposition, it is possible to identify the characteristics of various locations without tracing the movement trajectories of individuals in detail.

4.2. Analysis of People's Movements

4.2.1. Tensor Construction

We construct a third-order tensor $\mathscr{X} = [x_{ijk}] \in \mathbb{R}^{I\times J\times K}$ on the movement of people from the movement trajectory data. We classify the direction in which a person moves at 1-min intervals at each time and each location into 16 directions and then collate the results. The time and locations are the same as in Section 4.1. The tensor has the dimensions of time, location, and direction of movement, with the corresponding numbers of people as the elements. In other words, the tensor contains the information of how many people moved in which direction, when, and where. The dimensions have the size $I = 601$, $J = 1770$ and $K = 16$, and there are $92,462$ nonzero elements, which represent approximately 0.5% of the total elements.

4.2.2. Tensor Decomposition

We perform tensor decomposition for rank $R = 5$. We observe the time, position, and movement direction according to the five ranks and examine the results. It should be noted that the tensor decomposition is independent of that described in the previous section, such that there are no correspondences with the previous ranks.

First, the time $\times$ rank matrix $\bar{A}$ is represented by a cumulative bar chart plotted against the elapsed time in **Fig. 11**. Rank 4 displays a characteristic quite different from the other ranks. The other ranks display significant increases when there are peaks in the people's movement. However, rank 4 comprises a large portion in the time zones that correspond to the steady movement of people. In particular, while the other ranks decrease after evacuation is nearly completed, rank 4 remains at the same level and thus comprises a large portion. We thus speculate that this rank includes the movement of people who do not belong to the major evacuating group. As mentioned in Section 2, such movements are those of people engaged in rescue activities and trapped people, as well as those evacuating to locations other than the major evacuation centers.

The other ranks are nearly absent after the evacuation has been completed and thus can be considered to be related to movements during evacuation. However, the ranks respectively display peaks at different times. Rank 5 is high in the relatively early time zones, while rank 3 is nearly absent during the steady-state time zones, it appears only during peaks. Rank 2 becomes dominant in the peaks occurring in the later time zones. Rank 1 displays peaks that are earlier than those of ranks 2 and 3.

While taking into consideration these observations, we examine the zones. **Fig. 12** shows the location $\times$ rank matrix $\bar{B}$ on a map. As in the previous section, only the highest ranks are shown, and the zones with very few people are omitted. The area labeled C and its vicinity mostly comprises ranks 1 and 5, which indicates that the people in these zones evacuated early. The majority of the areas wherein rank 3 dominates are found close to the zone labeled B-1, which comprises evacuation centers. Based on our earlier observation that rank 3 appears only during peaks of movement, this rank can be considered to be a major component of the movement of large numbers of people to evacuation centers. Rank 4 is observed in several zones along the major thoroughfare, and those located to the southeast represent movements to label B-2, which is a different evacuation site from the major evacuation centers.

A map representing $\bar{B}$ for the 30-min period after 9.5 h, when most of the evacuation has been completed, is shown in **Fig. 13**. It can be observed that the total number of people moving is low as compared to those in the other time zones, and that the areas in which movement occurs are limited. During this time period, rank 4 is found to be dominant in many areas. The movement of rank 4 during this time period is shown in **Fig. 14**. The zones labeled D, E, and F, which display many movements of rank 4, are magnified. The grids show the average damage status of buildings and roads as indicated by different shades

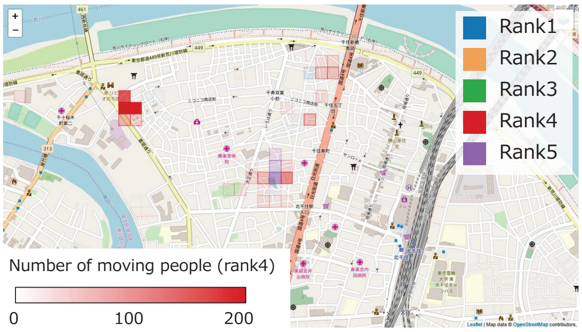


Fig. 13. Ranks of geographical locations after evacuation is complete.

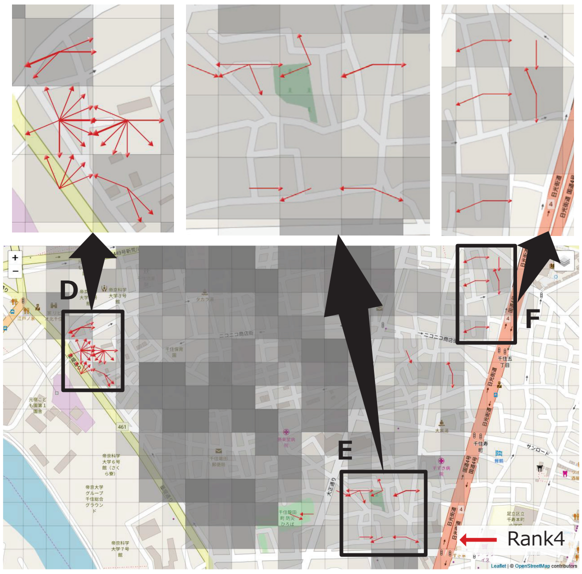


Fig. 14. Detailed movements of rank 4 after evacuation is complete.

of grey. Even though the evacuation has been completed, people continue to move, and the directions of movement within the individual zones tend to be dispersed in the north-south or east-west directions. Furthermore, these movements take place in locations that are somewhat distant from areas subjected to severe damage. Thus, it can be speculated that these zones comprise trapped survivors that still remain and wherein they and the rescue workers are moving to and fro within a narrow area. Therefore, this confirms the observation that rank 4 includes movements that are different from the major evacuation trajectories. These findings demonstrate that tensor decomposition of the people's movement facilitates the extraction of characteristic components such as the major evacuation movements and other specific actions.

5. Conclusion

In this paper, we proposed the use of tensor decomposition for analyzing evacuation simulation data. While specifically focusing on the movement trajectory data, we performed an analysis of the statuses of people during evacuation and the characteristics of their movement. The movement trajectory data were collated to construct tensors of the people's statuses and movement. NTF was applied to conduct an exploratory analysis based on visualization. The results showed that the data can be classified into ranks representing the major and minor evacuation movements, rescue activities, trapped victims, etc. In this manner, we were able to extract characteristic components regarding evacuation without tracing the movement trajectories of individuals in detail.

There remain several issues, however, with regard to the application of tensor decomposition. The method of assigning the appropriate rank is a major issue in the research of tensor-decomposition-based analysis [10, 12]. Although we selected $R = 5$, which is relatively low, to perform the tensor decomposition, the selection of a larger number may yield more detailed evacuation patterns. However, there exists the risk that the dataset will be classified into an excessive number of patterns and that a single behavior pattern will be divided into multiple ranks. It is thus necessary to select the rank in an exploratory manner according to the target "roughness of analysis" based on how fine a classification pattern one wishes to obtain. It is also necessary to examine the suitable metric of error convergence or the method of optimization to improve the accuracy of the tensor decomposition.

While we used NTF [8] for tensor decomposition in the present study, other approaches should be investigated as well. We constructed two tensors, one on the status of people and the other on their movements and decomposed them independently. If a tensor decomposition that takes into consideration both tensors can be performed, it would allow one to analyze the dataset by taking the people's statuses and movement into consideration simultaneously. Context Aware Tensor Decomposition (CATD) [13] and Non-negative Multiple Tensor Factorization (NMTF) [12] are two examples of decomposition based on multiple tensors. In these methods, the main tensor is augmented by auxiliary tensors or matrices to improve the decomposition accuracy or perform analyses using a greater number of dimensions. By suitably designing the tensors or matrices and applying these methods, we can expect to obtain characteristics that were not found in the present study.

In the present study, we subjected a single evacuation simulation trial to analysis. However, simulations are usually performed for several trials. This requires the comparative analysis of multiple results. Such a comparative analysis is possible by setting the simulation ID or setting parameters as the dimensions when designing the tensor. This should enable one to obtain results that are common to a high number of simulations or extract characteristics that are found in only a few of the simulations.

Acknowledgements

The authors would like to express their gratitude to Osaragi Laboratory, Tokyo Institute of Technology, for providing us with the evacuation simulation data. This study was partially funded under

CREST "Creation of Innovative Analysis Infrastructure for Earthquake and Tsunami Disaster Mitigation Big Data based on Large-Scale High-Resolution Numerical Simulations," Grant-in-Aid for Scientific Research (16H01722).

References:

[1] T. Oki and T. Osaragi, "Modeling Human Behavior of Local Residents in the Aftermath of a Large Earthquake – Wide-Area Evacuation, Rescue and Firefighting in Densely Built-Up Wooden Residential Areas," J. Disaster Res., Vol.11, No.2, pp. 188-197, 2016.

[2] T. Osaragi and T. Oki, "Wide-Area Evacuation Simulation Incorporating Rescue and Firefighting by Local Residents," J. Disaster Res., Vol.12, No.2, pp. 296-310, 2017.

[3] Ministry of Land, Infrastructure, Transport and Tourism, "Guidelines on Planning for Urban Development Preparing for Tsunami," Jun. 2013 (in Japanese).

[4] Y. Kawai, J. Zhao, K. Sugiura, Y. Ishikawa, and Y. Wakita, "An Analysis Technique of Evacuation Simulation Using an Array DBMS," J. Disaster Res., Vol.13, No.2, pp. 338-346, 2018.

[5] J. Zhao, Y. Ishikawa, Y. Wakita, , and K. Sugiura, "Difference Operators in Simulation Data Warehouses," J. Disaster Res., Vol.12, No.2, pp. 347-354, 2017.

[6] J. Zhao, K. Sugiura, Y. Wang, , and Y. Ishikawa, "Simulation Data Warehouse for Integration and Analysis of Disaster Information," J. Disaster Res., Vol.11, No.2, pp. 255-264, 2016.

[7] H. Lustosa, F. Porto, P. Valduriez, and P. Blanco, "Database System Support of Simulation Data," Proc. VLDB Endow., Vol.9, No.13, pp. 1329-1340, 2016.

[8] A. Shashua and T. Hazan, "Non-negative Tensor Factorization with Applications to Statistics and Computer Vision," Proc. Int'l Conf. on Machine Learning (ICML 2005), pp. 792-799, 2005.

[9] A. Ozerov, C. Févotte, R. Blouet, and J. Durrieu, "Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation," In Proc. Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP 2011), pp. 257-260, 2011.

[10] T. Matsubayashi, M. Kohjima, A. Hayashi, and H. Sawada, "Brand-choice analysis using non-negative tensor factorization," Transactions of the Japanese Society of Artificial Intelligence, Vol.30, No.6, pp. 713-720, 2015 (in Japanese).

[11] "leaflet," https://leafletjs.com [accessed March 14, 2019]

[12] K. Takeuchi, R. Tomioka, K. Ishiguro, A. Kimura, and H. Sawada, "Non-negative Multiple Tensor Factorization," In Proc. Int'l Conf. on Data Mining (ICDM 2013), pp. 1199-1204, 2013.

[13] Y. Wang, Y. Zheng, and Y. Xue, "Travel Time Estimation of a Path Using Sparse Trajectories," Proc. ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2014), pp. 25-34, 2014.

Name:
Yusuke Kawai

Affiliation:
Master Student, Graduate School of Informatics, Nagoya University

Address:
Furo-cho, Chikusa-ward, Nagoya 464-8601, Japan

Brief Career:
2017- Master Student, Graduate School of Informatics, Nagoya University

Selected Publications:
- "An Analysis Technique of Evacuation Simulation Using an Array DBMS," J. Disaster Res., Vol.13, No.2, pp. 338-346, 2018.

Academic Societies & Scientific Organizations:
- The Database Society of Japan (DBSJ)

Name:
Yoshiharu Ishikawa

Affiliation:
Professor, Graduate School of Informatics, Nagoya University

Address:
Furo-cho, Chikusa-ward, Nagoya 464-8601, Japan

Brief Career:
1994- Assistant Professor, Nara Institute of Science and Technology
1999- Assistant Professor, University of Tsukuba
2003- Associate Professor, University of Tsukuba
2006- Professor, Nagoya University

Selected Publications:
- "Probabilistic Range Querying over Gaussian Objects," IEICE Trans. on Information and Systems, Vol.E97-D, No.4, pp. 694-704, 2014.

Academic Societies & Scientific Organizations:
- Association for Computing Machinery (ACM)
- The Institute of Electrical and Electronics Engineers (IEEE) Computer Society
- Information Processing Society of Japan (IPSJ)
- The Institute of Electronics, Information and Telecommunication Engineers (IEICE)
- The Database Society of Japan (DBSJ)

Name:
Kento Sugiura

Affiliation:
Research Associate, Graduate School of Informatics, Nagoya University

Address:
Furo-cho, Chikusa-ward, Nagoya 464-8601, Japan

Brief Career:
2013- Master Student, Graduate School of Information Science, Nagoya University
2015- Ph.D. Candidate, Graduate School of Information Science, Nagoya University
2018- Research Associate, Graduate School of Informatics, Nagoya University

Selected Publications:
- "Grouping Methods for Pattern Matching over Probabilistic Data Streams," IEICE Trans. on Information and Systems, Vol.E100-D, No.4, pp. 718-729, 2017.

Academic Societies & Scientific Organizations:
- Association for Computing Machinery (ACM)
- Information Processing Society of Japan (IPSJ)
- The Database Society of Japan (DBSJ)