

テンソル分解を用いた避難移動軌跡データの分析

河井 悠佑[†] 石川 佳治[†] 杉浦 健人[†]

[†] 名古屋大学情報学研究科 〒464-8601 愛知県名古屋市千種区不老町

E-mail: [†]{kawai,sugiura}@db.is.i.nagoya-u.ac.jp, ishikawa@i.nagoya-u.ac.jp

あらまし 日本では非常に大きな災害が度々発生しており、災害対策に対する意識が高まっている。情報学からの試みとして災害時の避難シミュレーションが多数行われており、その活用が求められる。しかし、どのようにシミュレーションモデルに関する研究はよく行われている一方、シミュレーションデータの活用に関する研究は少ない。そこで本研究では、災害シミュレーションデータの分析技術の開発を目指して避難シミュレーションデータの分析に取り組む。避難シミュレーションデータが時間や空間を含む多次元データとして表せる点に着目してテンソル分解を用いた分析を行う。本論文では、避難シミュレーションデータをテンソルとして扱うための適切な設計について検討し、テンソル分解の結果を分析することで重要な災害や避難のパターンの獲得を目指す。

キーワード 避難シミュレーション, 移動軌跡データ, テンソル分解

1. はじめに

日本では、東日本大震災や熊本地震といった非常に大きな災害が度々発生しており、災害対策に対する意識が高まっている。政府では東日本大震災を踏まえて防災計画の見直しがなされており、従来の想定よりも大規模な災害に対するハード・ソフト両面の対策が求められる [18]。

災害対策として行われている方策は多数あるが、情報学からの試みとしては、大規模なシミュレーションによる被害や避難行動の予測が挙げられる。例えば、沖らの研究 [6, 9, 10] では、地震発生時における家屋の倒壊や火災の発生を考慮した人々の避難シミュレーションを行っている。こうしたシミュレーションで得られる被害状況を分析することで、事前に避難経路や防災計画の策定、都市のインフラ整備といった災害対策につなげられる。

しかし、シミュレーションデータの利活用に着目した技術開発に関する研究は少ない。避難シミュレーションのモデル化に関する研究 [6, 9, 10] はなされているものの、シミュレーション結果をどのように活用するかが課題となっている。そのため、多数のシミュレーションが試行されても、それらをうまく活用できていないのが現状である。

そこで本研究では、災害シミュレーションデータの分析技術の開発として、避難シミュレーションデータの分析に取り組む。避難シミュレーションは地震などの災害発生時に群衆が避難する様子をシミュレーションしたデータである。一人ひとりの各時刻における位置を持つ他、「建物に閉じ込められている」、「救助活動に参加している」などの状態に関する情報や、建物や道路の損傷状況の情報も持つ。このようなデータを分析することで、災害対策につながる有用な知見の発見を目指す。避難シミュレーションデータの特徴は、時間や空間を含む多次元データという点である。例えば、沖らの研究で行われている避難シミュレーションであれば、時空間に加え「ユーザ ID」と「状態」という 4 つの次元の情報を持つ。

そこで本研究では、多次元データの分析に用いられるテンソル分解を利用して避難シミュレーションの分析を行う。テンソル分解は、「テンソルの成分に関して少数のグループ（ランクと呼ぶ）に近似的に分割することが可能である」という低ランク性を仮定することによって、低ランクの行列（因子行列と呼ぶ）の積に分解する手法である。元のデータに含まれる特徴的な情報を残しつつ高次元のテンソルをより低次元のテンソルへ分解できるため、画像や音声におけるノイズ処理や特徴抽出 [11, 12] など様々な分野で応用されている。また、多次元で複雑な情報を総合的に考慮しながらグループ化ができるため、単純な分析とは異なる知見が得られることが期待されており、多数の属性に基づく購買履歴データにおけるアイテム推薦 [7, 8] にも用いられている。避難シミュレーションデータにおいては、重要な災害や避難のパターンの獲得が期待できる。

本論文の構成は以下の通りである。2. 章では、関連研究について述べる。3. 章では、テンソル分解の手法について説明する。4. 章では、今回使用する避難シミュレーションデータの説明と事前分析を行う。5. 章では、テンソル分解を用いた避難シミュレーションデータの分析を行う。6. 章では、本論文のまとめと今後の展望について述べる。

2. 関連研究

避難シミュレーションのモデル化に関する研究 [6, 9, 10] はなされているものの、生成されたデータの利活用に着目した技術開発に関する研究は少ない。

多次元で膨大なデータを分析をサポートする仕組みとして、配列形式のデータのサポートに特化した DBMS である配列指向 DBMS の活用が考えられる。多数の次元に伴ったシミュレーションデータに対して、配列指向 DBMS の一つである SciDB [2, 3, 15] を利用した研究がいくつか挙げられる。[4] では、避難シミュレーションデータを SciDB に格納して演算を適用することで、空間的に粒度を変えた管理を効率的に行っている。[5] では、血流シミュレーションデータに対する問合せに關

して、他の種類の DBMS と比較を行っている。これらの研究は、データベース管理や問合せが目的であり、データの詳細な分析はしていない。

実世界のデータを利用した大規模な災害時の行動分析に関して、いくつか研究されている [13,14]。これらの研究では、東日本大震災や福島原発事故の際のデータも含む、160 万人の GPS データを利用して、緊急時の行動のモデル化と予測が行われている。各人々にとっての重要な場所を定義して、その場所間での隠れマルコフモデルとしてモデル化を行い、交通ネットワークや都市グラフに沿った移動経路を計画する。そして、逆強化学習による最適化を行って、安全かつ速く目的地に到達できるよう訓練する。この研究は、実世界のデータを活用した点、避難行動のモデル化と予測に焦点を当てている点において、本研究とは異なる。

本研究と同様に、移動軌跡データに対してテンソル分解を適用した研究として [17] がある。タクシーの GPS データから、各時間帯の各道路の移動速度を集計する。しかし、実際に取得したデータだけでは、全ての道路の移動速度の情報が含まれていないため、テンソル分解によって成分の補完を行う。この研究はテンソルの成分を補完することを目的としており、本研究は分解後のテンソルを解釈することに主眼を置いている。

3. テンソル分解

本章では利用するテンソル分解の手法について述べる。本研究では、複数のテンソルを用いたテンソル分解を行う Non-negative Multiple Tensor Factorization (NMTF) [16] を用いる。この手法は、非負制約を維持しつつ CP 分解を行う Non-negative Tensor Factorization (NTF) [12] を、複数のテンソル分解へと拡張したものである。今回は三階のテンソルを例に説明を行う。また、因子行列からランクごとに復元を行うことによって、元のデータに合わせた定性的なランクの解釈が可能となる。

3.1 節で、本研究で適用する NMTF の説明を行う。3.2 節で、因子行列から元のテンソルへのランクごとの復元について述べる。

3.1 Non-negative Multiple Tensor Factorization

本研究で用いる NMTF について説明する。NMTF ではメインテンソルに補助テンソルをいくつか追加して同時にテンソル分解を行う。メインテンソルのスパース性を補うことによってテンソルの分解精度を上げたり、より多数の属性に基づいた多角的なデータ解釈を行うことが目的である。NMTF では、補助テンソルはメインテンソルと次元を一つ以上共有しており、同一の次元では同じ因子行列が生成されると仮定する。

このときの NMTF の概要を図 1 に示す。三階テンソル $\mathcal{X}$ をメインテンソルとし、補助テンソルとして三つの行列 $\mathbf{Y}^{(1)} = [\mathbf{y}_{il}^{(1)}] \in \mathbb{R}^{I \times L}$, $\mathbf{Y}^{(2)} = [\mathbf{y}_{jm}^{(2)}] \in \mathbb{R}^{J \times M}$, $\mathbf{Y}^{(3)} = [\mathbf{y}_{kn}^{(3)}] \in \mathbb{R}^{K \times N}$ (それぞれ補助行列 1, 補助行列 2, 補助行列 3 と呼ぶ) を追加する。 $\mathbf{Y}^{(1)}$ はメインテンソルの一番目の次元と、 $\mathbf{Y}^{(2)}$ は二番目の次元と、 $\mathbf{Y}^{(3)}$ は三番目の次元とそれぞれ共有している。ランク数 R で分解を行う。

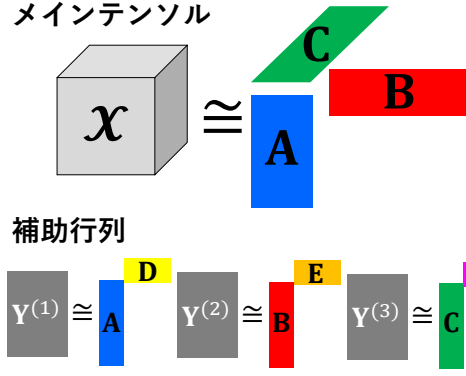


図 1: NMTF

メインテンソル $\mathcal{X}$ は三つの因子行列 $\mathbf{A} = [a_{ir}] \in \mathbb{R}^{I \times R}$, $\mathbf{B} = [b_{jr}] \in \mathbb{R}^{J \times R}$, $\mathbf{C} = [c_{kr}] \in \mathbb{R}^{K \times R}$ に因子分解される。分解により生成した因子行列から復元したテンソル $\hat{\mathcal{X}} = [\hat{x}_{ijk}] \in \mathbb{R}^{I \times J \times K}$ は、三つの因子行列の積として表現される。このテンソルの各要素 $\hat{x}_{ijk}$ は以下のように表される。

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

補助行列 $\mathbf{Y}^{(1)}$ は二つの因子行列 $\mathbf{A} = [a_{ir}] \in \mathbb{R}^{I \times R}$, $\mathbf{D} = [d_{lr}] \in \mathbb{R}^{L \times R}$ に分解される。同様に、補助行列 2 は $\mathbf{B} = [b_{jr}] \in \mathbb{R}^{J \times R}$, $\mathbf{E} = [e_{mr}] \in \mathbb{R}^{M \times R}$ に、補助行列 3 は $\mathbf{C} = [c_{kr}] \in \mathbb{R}^{K \times R}$, $\mathbf{F} = [f_{nr}] \in \mathbb{R}^{N \times R}$ にそれぞれ分解される。補助行列で生成される因子行列 $\mathbf{A}, \mathbf{B}, \mathbf{C}$ は、メインテンソルの分解によって生成された因子行列 $\mathbf{A}, \mathbf{B}, \mathbf{C}$ と同一である。因子行列から復元した補助行列 $\hat{\mathbf{Y}}^{(1)} = [\hat{y}_{il}^{(1)}]$, $\hat{\mathbf{Y}}^{(2)} = [\hat{y}_{jm}^{(2)}]$, $\hat{\mathbf{Y}}^{(3)} = [\hat{y}_{kn}^{(3)}]$ はそれぞれ以下のように求められる。

$$\hat{y}_{il}^{(1)} = \sum_{r=1}^R a_{ir} d_{lr} \quad (2)$$

$$\hat{y}_{jm}^{(2)} = \sum_{r=1}^R b_{jr} e_{mr} \quad (3)$$

$$\hat{y}_{kn}^{(3)} = \sum_{r=1}^R c_{kr} f_{nr} \quad (4)$$

因子行列 $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F}$ は、元のテンソル $\mathcal{X}, \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \mathbf{Y}^{(3)}$ と因子行列からの復元テンソル $\hat{\mathcal{X}}, \hat{\mathbf{Y}}^{(1)}, \hat{\mathbf{Y}}^{(2)}, \hat{\mathbf{Y}}^{(3)}$ との誤差 $\mathcal{D}$ が最小になるように求める。

誤差 $\mathcal{D}$ はテンソルの各要素同士の誤差 d の合計をとる。

$$\mathcal{D}(\mathcal{X}|\hat{\mathcal{X}}) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K d(x_{ijk}|\hat{x}_{ijk}) \quad (5)$$

$$\mathcal{D}(\mathbf{Y}^{(1)}|\hat{\mathbf{Y}}^{(1)}) = \sum_{i=1}^I \sum_{l=1}^L d(y_{il}^{(1)}|\hat{y}_{il}^{(1)}) \quad (6)$$

$$\mathcal{D}(\mathbf{Y}^{(2)}|\hat{\mathbf{Y}}^{(2)}) = \sum_{j=1}^J \sum_{m=1}^M d(y_{jm}^{(2)}|\hat{y}_{jm}^{(2)}) \quad (7)$$

$$\mathcal{D}(\mathbf{Y}^{(3)}|\hat{\mathbf{Y}}^{(3)}) = \sum_{k=1}^K \sum_{n=1}^N d(y_{kn}^{(3)}|\hat{y}_{kn}^{(3)}) \quad (8)$$

非負要素の誤差を評価する距離尺度 d には、ユークリッド距離や KL ダイバージェンスなどいくつかあるが、今回は一般化 KL ダイバージェンスを採用する。

$$d(p|q) = p(\log p - \log q) - (p - q) \quad (9)$$

NMTF ではメインテンソルと補助行列の誤差も重みを考慮する必要があるため、ハイパーパラメータ η を導入する。誤差は以下ようになる。

$$\begin{aligned} \mathcal{D}(\mathcal{X}, \mathbf{Y}^{(n)}, \eta^{(n)} | \hat{\mathcal{X}}, \hat{\mathbf{Y}}^{(n)}, \eta^{(n)}; n = 1, 2, 3) \\ = \mathcal{D}(\mathcal{X} | \hat{\mathcal{X}}) + \eta^{(1)} \mathcal{D}(\mathbf{Y}^{(1)} | \hat{\mathbf{Y}}^{(1)}) \\ + \eta^{(2)} \mathcal{D}(\mathbf{Y}^{(2)} | \hat{\mathbf{Y}}^{(2)}) + \eta^{(3)} \mathcal{D}(\mathbf{Y}^{(3)} | \hat{\mathbf{Y}}^{(3)}) \end{aligned} \quad (10)$$

目的関数は以下のように表される。

$$\begin{aligned} \underset{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F}}{\text{minimize}} \quad & \mathcal{D}(\mathcal{X}, \mathbf{Y}^{(n)}, \eta^{(n)} | \hat{\mathcal{X}}, \hat{\mathbf{Y}}^{(n)}, \eta^{(n)}; n = 1, 2, 3) \\ \text{subject to} \quad & \mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F} \geq 0 \end{aligned} \quad (11)$$

η の値が大きいほど、対応する補助行列を重視してテンソル分解を行っていることになる。

目的関数の導出については省略するが、誤差 $\mathcal{D}$ に対する各因子行列の偏微分が 0 になるように更新を行う。最終的な各因子行列の更新式は以下の通りである。

$$a_{ir}^{\text{new}} = a_{ir} \frac{\sum_{j=1}^J \sum_{k=1}^K \frac{x_{ijk}}{\hat{x}_{ijk}} b_{jr} c_{kr} + \eta^{(1)} \sum_{l=1}^L \frac{y_{il}^{(1)}}{\hat{y}_{il}^{(1)}} d_{lr}}{\sum_{j=1}^J \sum_{k=1}^K b_{jr} c_{kr} + \eta^{(1)} \sum_{l=1}^L d_{lr}} \quad (12)$$

$$b_{jr}^{\text{new}} = b_{jr} \frac{\sum_{i=1}^I \sum_{k=1}^K \frac{x_{ijk}}{\hat{x}_{ijk}} a_{ir} c_{kr} + \eta^{(2)} \sum_{m=1}^M \frac{y_{jm}^{(2)}}{\hat{y}_{jm}^{(2)}} e_{mr}}{\sum_{i=1}^I \sum_{k=1}^K a_{ir} c_{kr} + \eta^{(2)} \sum_{m=1}^M e_{mr}} \quad (13)$$

$$c_{kr}^{\text{new}} = c_{kr} \frac{\sum_{i=1}^I \sum_{j=1}^J \frac{x_{ijk}}{\hat{x}_{ijk}} a_{ir} b_{jr} + \eta^{(3)} \sum_{n=1}^N \frac{y_{kn}^{(3)}}{\hat{y}_{kn}^{(3)}} f_{nr}}{\sum_{i=1}^I \sum_{j=1}^J a_{ir} b_{jr} + \eta^{(3)} \sum_{n=1}^N f_{nr}} \quad (14)$$

$$d_{lr}^{\text{new}} = d_{lr} \frac{\eta^{(1)} \sum_{i=1}^I \frac{y_{il}^{(1)}}{\hat{y}_{il}^{(1)}} a_{ir}}{\eta^{(1)} \sum_{i=1}^I a_{ir}} \quad (15)$$

$$e_{mr}^{\text{new}} = e_{mr} \frac{\eta^{(2)} \sum_{j=1}^J \frac{y_{jm}^{(2)}}{\hat{y}_{jm}^{(2)}} b_{jr}}{\eta^{(2)} \sum_{j=1}^J b_{jr}} \quad (16)$$

$$f_{nr}^{\text{new}} = f_{nr} \frac{\eta^{(3)} \sum_{k=1}^K \frac{y_{kn}^{(3)}}{\hat{y}_{kn}^{(3)}} c_{kr}}{\eta^{(3)} \sum_{k=1}^K c_{kr}} \quad (17)$$

これを踏まえて、NMTF のアルゴリズム 1 を示す。元のメインテンソルと復元したメインテンソルとの誤差が収束するまで、因子行列の更新を繰り返す。

3.2 ランクごとの復元

推定して求めた因子行列から生成したテンソル $\hat{\mathcal{X}}$ は、元のテンソル $\mathcal{X}$ との誤差が大きいため、このままの状態では元のデータに基づいた分析は難しい。そこで、推定したテンソル $\hat{\mathcal{X}}$ を考慮しつつ、元のテンソル $\mathcal{X}$ に則した形になるよ

Algorithm 1 Estimation of factor matrices

Input: $\mathcal{X}, \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \mathbf{Y}^{(3)}, R, E$

Output: $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F}$

initialize ϵ

initialize $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{E}, \mathbf{F}$ with small random values

while $\epsilon \geq E$ **do**

 update $\mathbf{A} = [a_{ir}]$ by (12)

 update $\mathbf{B} = [b_{jr}]$ by (13)

 update $\mathbf{C} = [c_{kr}]$ by (14)

 update $\mathbf{D} = [d_{lr}]$ by (15)

 update $\mathbf{E} = [e_{mr}]$ by (16)

 update $\mathbf{F} = [f_{nr}]$ by (17)

 update $\hat{\mathcal{X}} = [\hat{x}_{ijk}]$ by (1)

$loss^{\text{new}} = \mathcal{D}(\mathcal{X} | \hat{\mathcal{X}})$

$\epsilon = loss^{\text{old}} - loss^{\text{new}}$

$loss^{\text{old}} = loss^{\text{new}}$

end while

うに、要素をランクごとに復元したテンソル $\hat{\mathcal{X}}$ を作成する。 $\hat{\mathcal{X}} = [\hat{x}_{ijk}] \in \mathbb{R}^{I \times J \times K \times R}$ は、元のテンソルの次元にランクを加えた四階のテンソルである。 $\hat{\mathcal{X}}$ の各要素 $\hat{x}_{ijk}$ は、ランクごとに復元した要素 $a_{ir} b_{jr} c_{kr}$ に関して、ランク方向の合計が元のテンソルの要素 x_{ijk} と等しくなるように正規化を行う。したがって、以下のような式で求められる。

$$\hat{x}_{ijk} = x_{ijk} \frac{a_{ir} b_{jr} c_{kr}}{\sum_{r=1}^R a_{ir} b_{jr} c_{kr}} \quad (18)$$

実際の分析では、NMTF により求められた因子行列だけでなく、 $\hat{\mathcal{X}}$ も用いて分析を行う。

4. 避難シミュレーションデータ

4.1 節で今回使用する避難シミュレーションデータの説明と、4.2 節で集計によるシミュレーション結果の事前分析を行う。

4.1 データセット

東京工業大学の大佛研究室では首都直下型地震を想定した広域避難シミュレーションが試行されており [6, 10]、本研究では大佛研究室より提供されたある避難シミュレーションデータを用いる。このシミュレーションでは、冬の 18 時に東京湾北部で震度 6 強の地震が発生したと想定されており、東京都の北千住地区の人々が避難する様子をシミュレーションしている。この地区は、古い木造家屋が多いため火災や倒壊の危険性が高いことに加えて、荒川と隅田川に囲まれているため広域避難が困難であり、災害において危険性が高い地域となっている。

このシミュレーションは、地震が発生して、いくつかの建物にて火災や倒壊が発生した状況において、広域避難場所への避難を目指すものとなっている。シミュレーション対象者は一定の領域内における建物内の滞留者と屋外の歩行者である。鉄道や自動車で移動中の人々や、領域外から流入する人々は考慮しない。建物の倒壊や道路の閉塞が起こっている中で、人々はそれぞれ行動を選択する。基本的に、行動可能な人々はまず近くの一時集合場所へ一時的に避難して、広域避難場所へ避難を行う。一方で、このシミュレーションでは避難を行うだけでなく、

属性	意味	備考
t	時刻	0.1 分刻みで 24 時間分
uid	人の ID	13,487 人分 (1 から連番)
x, y	X,Y 座標	平面直角座標系 IX 系
status	人の状態	表 2 参照

表 1: 人の避難データ

状態	意味
StayHome	発災時に自宅に滞在
StayOther	発災時に建物に滞在
Move	発災時に屋外で歩行
StayIttoki	一時集合場所に滞在
Safe	広域避難場所に滞在
Search	避難経路を探索中
HinanIttoki	一時集合場所に移動
HinanIttoki2	一時集合場所に到達
HinanKoiki	広域避難場所に移動
HinanKoiki2	広域避難場所に到達
Rescue	救助活動に参加中
Fire	消火活動に参加中
Konnan	街路内閉じ込め者
Death	死亡者

表 2: 人の状態

属性	意味	備考
t	時刻	表 1 と同様
uid	建物 ID	4,392 個
collapsed	倒壊状態	0:非倒壊/1:倒壊
stage	火災段階	0:未着火/1:延焼の恐れあり/ 2:着火/3:鎮火済

表 3: 建物の被害データ

属性	意味	備考
t	時刻	表 1 と同様
uid	道路 ID	2,720 本分
rubble	閉塞状態	0:非閉塞/1:閉塞

表 4: 道路の被害データ

救助活動や消火活動への参加をモデルに組み込んでおり、避難せずにそうした活動へ参加する人々も存在する。なお、本研究では提供された一回分のシミュレーションデータを用いるが、元の研究では救助活動の有無や炎上中の建物から逃げるタイミングを変えて複数回シミュレーションを行っており、その際の被害状況を検証している。

このシミュレーションデータには、発災から 24 時間分における 13,487 人の避難行動 (表 1) 及び、4,392 個の建物 (表 3) と 2,720 本の道路 (表 4) の物的被害の様子が記録されている。人の避難行動データは、発災からの各経過時間 (以降、時刻と表記) における座標と 14 種類の避難状態 (表 2) が記録された、状態付きの移動軌跡データである。建物と道路のデータには、火災や閉塞の状況が記録されている。

4.2 事前分析

この節では、シミュレーション結果の事前分析として、人の避難行動データや建物及び道路の被害データに関する簡単な集計を行う。状態の変化が大きく見られた発災から 10 時間後ま

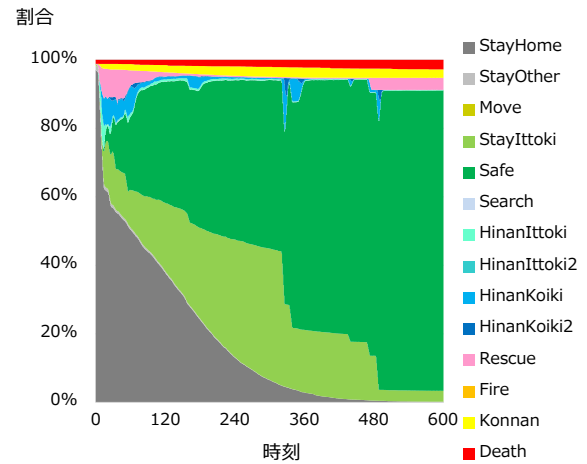


図 2: 10 時間分の各状態の割合

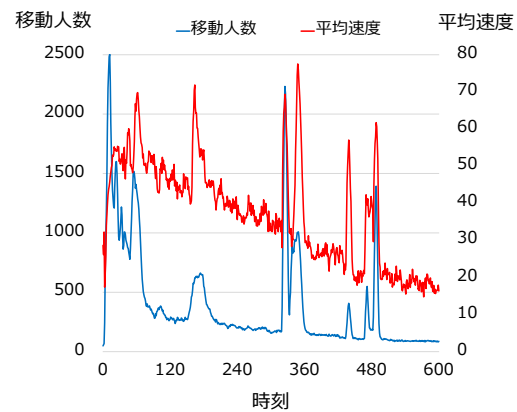
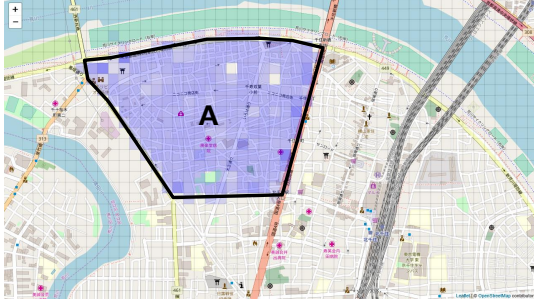


図 3: 10 時間分の移動人数と移動速度

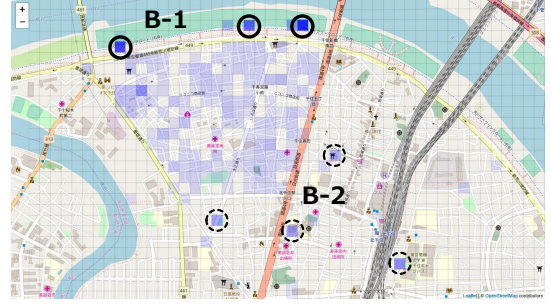
での避難行動データを分析対象とする。

10 時間分の各状態の割合を図 2 に示す。発災から時間が経過するにつれて、自宅に滞在している人々 (StayHome) が減っていき、広域避難場所への移動が完了 (Safe) した人々が増えていくことがわかる。その多くは、途中で一時集合場所に一時的に避難 (StayIttoki) を行っている。したがって、自宅に滞在している人々 (StayHome) の多くは、一時集合場所に移動した (StayIttoki) 後、広域避難場所へ避難 (Safe) することがわかる。5~6 時間後および 8 時間後付近において、避難完了者 (Safe) が大幅に増加しており、8 時間後以降は状態の変化はほとんど見られない。また、発災から 1 時間後付近は各状態とも大きな変化が見られる。その他の人々について、死亡者 (Death) は発災時に一定数発生し、緩やかに増加している。閉じ込め者 (Konnan) も緩やかに増加した後、減少傾向がみられることから、閉じ込め者の一部が死亡してしまったことが推測できる。救助活動を行っている人々 (Rescue) は、発災直後と大多数が避難が完了した後に一定数存在する。

次に、発災から 10 時間後までの各時刻において、移動した人々に関して集計した。一分ごとの位置座標に関して変化があった場合に移動状態にある人とみなし、その人々の人数 (以降、移動人数と表記) と一分間の移動距離の平均 (以降、平均移動速度と表記) を集計した。図 3 は移動人数 (青線) と平均



(a) 発災時



(b) 10 時間後

図 4: 各領域における人数

移動速度（赤線）を示したものである。横軸を時刻（分）として、上図の左縦軸を移動人数、右縦軸を移動速度（m/min）、下図の縦軸を全人数に対する各状態の割合とする。

各状態の変化が大きい時刻では移動人数や移動速度のピークが見られる。ピークを除くと、時間が経過するにつれて平均移動速度は低下する傾向がある。発災直後は移動速度はあまり大きくないものの、移動人数は多くなっている。移動人数は最大でも 2,500 人程度であり、全体の 19% 程度である。平均移動速度は最大でも 80(m/min) 程度であり、8 時間後以降は 20(m/min) まで低下する。多数の人々の避難が完了し、避難による移動が見られないことを考えると、救助活動 (Rescue) や閉じ込め者 (Konnan) が狭い範囲で移動していることが推測できる。

次に、人々の様子を領域ごとに可視化する。今回の避難シミュレーションは、実際の北千住地区に基づいて行われているため、地図上での可視化が有効である。地図上の可視化には、leaflet [1] を用いる。leaflet [1] は、ブラウザ上で Web 地図を表示することができる、オープンソースの JavaScript ライブラリである。地図上にマーカーや直線、ポリゴンなどのオブジェクトを設置して、分析した情報を示す。

発災時と 10 時間後における人々の滞在領域を図 4 に示す。50m × 50m のグリッドごとに人数を集計し、その大小を青色の濃淡で示した。発災時はラベル A で囲まれた領域内に人々が点在していることがわかる。また、避難完了した 10 時間後にはいくつかの場所に避難していることがわかる。大半の人は上部の 3 箇所（ラベル B-1）に避難しており、合わせると全体の 8 割を超える。一部の人は南東方面にある公園や学校（ラベル B-2）へ避難していることが確認できた。少数であるが、発災時の領域に残っている人が見られ、これらは救助活動 (Rescue) や閉じ込め者 (Konnan)、死者 (Death) に当たる。

5. 分析シナリオ

本章では、4.1 章で説明したデータセットから、実際にテンソルを作成して分析を行う。3. 章で述べた NMFT を行い、因子行列やランクごとに復元したテンソルに関して分析を行う。避難シミュレーションデータの分析としては様々なものが考えられるが、今回は人の移動軌跡データ（表 1）に焦点を当てて、人の状態や移動の様子における分析を例に挙げる。まず 5.1 節

テンソル	次元 (因子行列)	時刻T	領域G	状態S	方向D	人々U	建物B	道路L
		時刻T	領域G	状態S	方向D	人々U	建物B	道路L
メイン	$\mathcal{X}$							
補助	$\mathcal{Y}^{(1)}$							
	$\mathcal{Y}^{(2)}$							
	$\mathcal{Y}^{(3)}$							
	$\mathcal{Y}^{(4)}$							
	$\mathcal{Y}^{(5)}$							

図 5: 作成したテンソルと次元との対応

でテンソルの作成と、5.2 節で NMFT に必要なパラメータ η の設定を行う。5.3 節で NMFT とその結果の分析を行う。

5.1 テンソルの作成

作成したテンソルの一覧を表形式で図 5 に示す。表の行が作成するテンソル、列がテンソルの次元を表し、テンソルに対応する次元に色が塗ってある。また、テンソル分解によって対応する次元の因子行列が生成される。今回作成するテンソルは全部で 6 つであり、人の避難行動（表 1）から作成したテンソルが 4 つと、建物（表 3）及び道路（表 4）の物的被害から作成したテンソルが 1 つずつである。そのうち、時刻×領域×状態の三階テンソル $\mathcal{X}$ をメインテンソルとして、残りを補助テンソルとする。補助テンソルはあくまで分析に用いるテンソルの候補であり、後述する予備実験の結果により使用するか決定する。

各次元について説明する。時刻は分単位で 10 時間分としてサイズは $T = 601$ である。領域は 50m × 50m のグリッドを 1 単位としてサイズは $G = 3960$ である。状態は表 2 の通り扱い、サイズは $S = 14$ である。方向は 16 方位の移動方向を表し、サイズは $D = 16$ である。人々、建物、道路は元のデータの uid をそのまま使用し、サイズは $U = 13,487$, $B = 4,392$, $L = 2,720$ である。

作成したテンソルを以下に記す。

状態に関する 3 階テンソル $\mathcal{X} = [x_{tgs}^{(1)}] \in \mathbb{R}^{T \times G \times S}$

いつ、どこで、どのような状態の人が何人いたかという情報を持つ三階テンソルである。時刻、領域、状態ごとに人数を集計する。今回の分析ではこれをメインテンソルとする。

移動に関する 3 階テンソル $\mathcal{Y}^{(1)} = [y_{tgd}^{(1)}] \in \mathbb{R}^{T \times G \times D}$

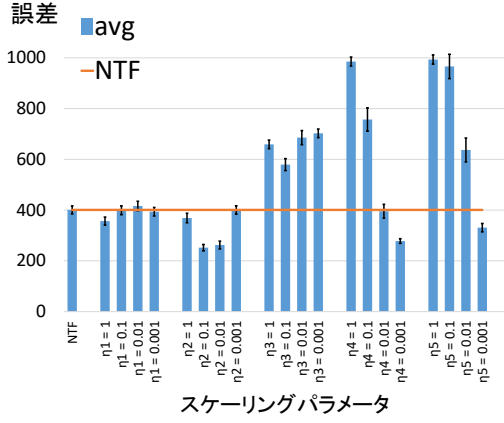


図 6: 各補助テンソルにおけるメインテンソルの分解精度

いつ、どこで、どのような方向に何人移動したかという情報を持つ三階テンソルである。時刻、領域、各移動方向ごとに人数を集計する。

ユーザの滞在場所に関する行列 $\mathbf{Y}^{(2)} = [y_{ug}^{(2)}] \in \mathbb{R}^{U \times G}$

誰が、どこにどのくらいの時間滞在していたかという情報を持つ行列である。各人々の各グリッド内での滞在時間を集計した。

ユーザの状態に関する行列 $\mathbf{Y}^{(3)} = [y_{us}^{(3)}] \in \mathbb{R}^{U \times S}$

誰が、どのような状態にどのくらいの時間なっているかという情報を持つ行列である。各人々の各状態になっている時間を集計した。

建物の被害状況に関する行列 $\mathbf{Y}^{(4)} = [y_{tb}^{(4)}] \in \mathbb{R}^{T \times B}$

いつ、どこの建物が、どのくらいの被害にあっているかという情報を持つ行列である。1分ごとの各建物の被害の様子を集計する。元のデータ(表3)のカテゴリを危険な状況に応じて分類し、0から1となるように被害の値を設定した。倒壊状況に関して倒壊を0.5、火災状況に関して着火済を0.5、延焼の危険性ありを0.25として、倒壊と火災状況を合算する。鎮火済は危険性は少ないと判断して未着火と同じ0とした。倒壊かつ着火済が最大の1をとることになる。

道路の被害状況に関する行列 $\mathbf{Y}^{(5)} = [y_{ul}^{(5)}] \in \mathbb{R}^{T \times L}$

いつ、どこの道路が、どのくらいの被害にあっているかという情報を持つ行列である。1分ごとの各道路の閉塞状況が1/0で表されている。

5.2 予備実験

メインテンソルの分解精度を上げるようにハイパーパラメータ η の値を調整する。全ての補助テンソルの組合せと η の大きさの候補は無数にあるため、メインテンソルと補助テンソル1つを用いて検証する。5つの補助テンソル $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \mathbf{Y}^{(3)}, \mathbf{Y}^{(4)}, \mathbf{Y}^{(5)}$ におけるスケーリングパラメータ $\eta_1, \eta_2, \eta_3, \eta_4, \eta_5$ について、それぞれ $\eta = 1, 0.1, 0.01, 0.001$ の4つを用意する。これにメインテンソル単独でテンソル分解を行う NTF を加えて合計 21 パターンのテンソル分解を行う。各パターンにつき 100 回実行し、メインテンソルと因子行列から復元したテンソルとの誤差の平均を評価の基準とする。誤差の尺度は一般化 KL ダイバージェンス(式9)の総和(式5)である。ランク数 R は5とした。

結果を図6に示す。横軸が補助テンソルのスケーリングパラ

メータ η 、縦軸が誤差の平均値であり、縦棒のグラフで表す。また、縦棒のエラーバーは標準誤差を表す。誤差が小さいほどメインテンソルの分解精度が良い。一番左の NTF がメインテンソル単独で分解した際の精度を表し、これを NMTF の基準としてオレンジ色の横線を引いて各補助テンソルにおける誤差と比較する。

補助テンソルとその η に応じて大きく異なる結果となった。 $\mathbf{Y}^{(2)}$ は全ての η において NTF の誤差以下の数値が出ており、メインテンソルに対して良い影響を与えるテンソルといえる。特に $\eta = 0.1, 0.01$ の場合は誤差が NTF の 60% 程度に減少しており、大幅に分解の精度が良好化した。 $\mathbf{Y}^{(1)}$ は $\eta = 1$ のとき NTF よりも精度が良く、それ以外の小さな値のとき NTF と同じくらいの結果になった。 $\mathbf{Y}^{(3)}$ は誤差及び標準誤差が他の二つより大きく、これの補助行列を加えたことでメインテンソルの分解が安定しなくなることがわかる。一番良い $\eta = 0.1$ でも NTF の 150% 程度の誤差があり、メインテンソルに対して悪い影響を与えていることが確認できる。建物と道路から作成した補助テンソル $\mathbf{Y}^{(4)}, \mathbf{Y}^{(5)}$ は、人の避難行動から作成した補助テンソル3つと比べて、 η の影響が大きいことが確認できた。 η が小さいほど精度が良く $\eta = 0.001$ のとき NTF の誤差を下回ったが、 $\eta = 1$ のときは非常に誤差が大きい。

このような結果になった要因として、各次元のスパース性が考えられる。状態の次元サイズは14と小さく、メインテンソルだけで十分密な情報を持っていたため、状態の次元を共有する $\mathbf{Y}^{(3)}$ では悪い影響を与えてしまった。一方で時刻や領域の次元のサイズは大きくスパースであったため、時刻や領域を持つ補助テンソルは精度が高くなったと考えられる。

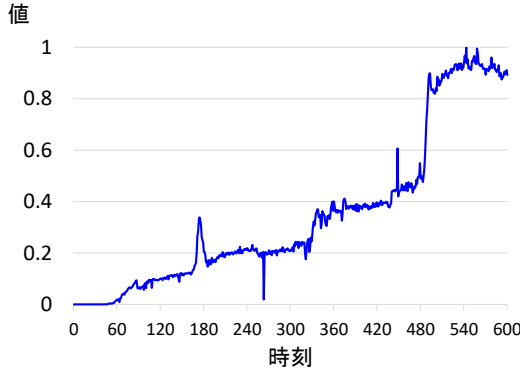
5.3 分析

NMTF によって生成された因子行列及びランクごとに復元したテンソルを用いて分析を行う。ランク数は $R = 10$ とし、5.2 節の予備実験の結果を踏まえて各パラメータは $\eta_1 = 1, \eta_2 = 0.1, \eta_3 = 0, \eta_4 = 0.001, \eta_5 = 0.001$ として NMTF を実行した。 $\mathbf{Y}^{(3)}$ はどの η でもメインテンソル単独の分解精度よりも悪いため使用しない。ここでは分析の一部として、5.3.1 節で rank1 に関する分析と、5.3.2 節で救助活動(Rescue)と閉じ込め者(Konnan)に関する分析について述べる。

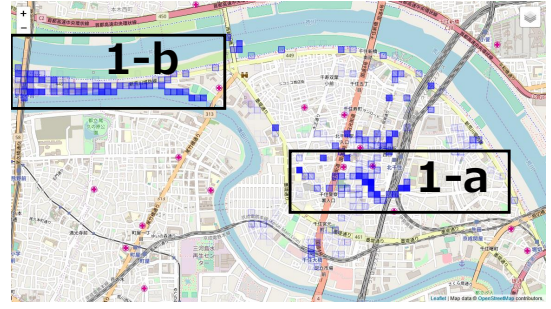
5.3.1 rank1

図7は時刻(図7a)と領域(図7b)の因子行列の様子を表したものである。時刻の因子行列は経過するにつれて増加傾向にあり、490分で大きく増加する。また、180分辺りと443分で一瞬の上昇が見られる。領域ではラベル1-aとラベル1-bで大きな値を持っている。

ラベル1-aは図4bにおけるラベルB-2の領域への避難に関わっていると推測できる。図8aはラベル1-aの領域において、時刻の因子行列でピークが見られた180分前後に滞在した人の位置を表したものである。連なった移動軌跡が確認でき、これはラベル1-aの値が大きい領域と一致する。つまり、ラベルB-2の領域への避難という一つの行動パターンが同じランクとして抽出できた。迂回した移動軌跡が見られることから、瓦礫等による閉塞が発生していることが推察できる。



(a) 時刻

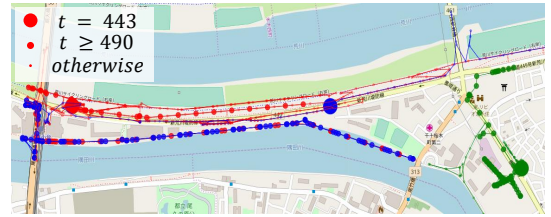


(b) 領域

図 7: rank1 の因子行列



(a) エリア 1-a の人の位置



(b) エリア 1-b を通った移動軌跡

図 8: rank1 の値が大きい領域

次にラベル 1-b の領域について見ていく．この領域を通った人はわずか 3 人であり，通常の避難とは大きく異なる行動をしている．その人らの移動軌跡を時刻の因子行列と対応させて図 8b に示す．3 人をそれぞれ赤色，青色，緑色の移動軌跡で示し，時刻のピークが見られた 443 分での位置を大きい円で，大きな上昇がみられた 490 分以降の位置を小さい円で強調した．それ以外の位置は小さい点で表す．青色の移動軌跡は 490 分以降に通った領域と広く重なっている上，ラベル 1-b の領域に始めた入った時刻と因子行列のピークである 443 分が一致しており，このユーザをピックアップしたランクといえる．つまり，単なる集計では得られにくい限られた行動パターンを強調できたといえる．1 人という限られた行動が強調された要因として，因子行列が割合に基づいた値であることが挙げられる．この領域を通った人が少ないため，割合的には大きな値となった．したがって，因子行列を分析することにより，母数としては少ない特徴的な行動パターンを検出しやすくなったといえる．

5.3.2 救助活動と閉じ込め者

ランクごとに復元したテンソル $\hat{\mathcal{X}}$ について，救助活動 (Rescue) と閉じ込め者 (Konnan) の時刻ごとの人数の推移を図 9 に示す図 2 により救助活動は発災直後と避難完了後に存在することは確認できていたが，図 9a ではそれらがはっきりランク分けされている．発災時は様々なランクに分かれているが，避難完了後は rank6 が全てを占めている．閉じ込め者 (図 9b) においても発災時は様々なランクに分かれているものの，最終的には rank8 という一つのランクが大部分を占めている．

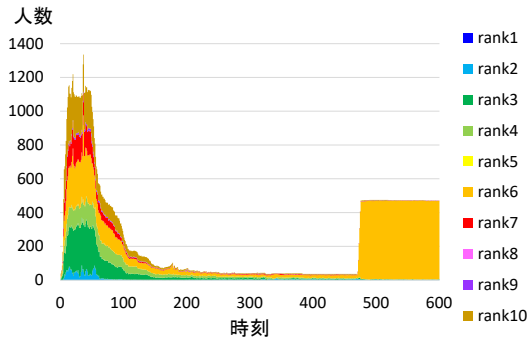
この結果を踏まえて救助活動と閉じ込め者の主要なランクである rank6 と rank8 について，10 時間後 (571 分～600 分) のランクごとに復元した移動の様子 $\hat{\mathcal{Y}}^{(1)}$ を図 10a と図 10b に

示す．閉じ込め者における rank8 の領域に四方八方に発散した移動の様子が確認できる．閉じ込められて狭い空間の中を動き回っていることが推察できる．rank6 についても各領域や移動の様子が閉じ込め者と類似していることが確認できる．rank6 は主に救助活動であるが，閉じ込め者を助けるために右往左往していることが推察できる．したがって，NMTF により複数のテンソルを同時に分解したことにより，時刻・領域・状態・移動を考慮してランク分けが行われた．

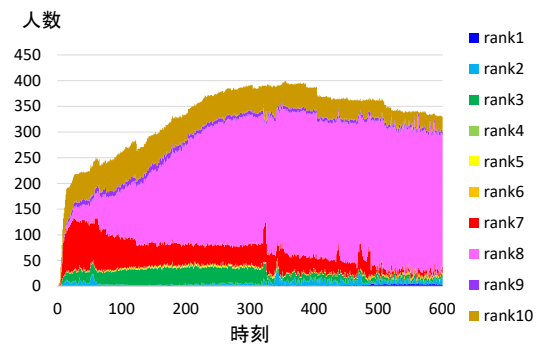
6. おわりに

本論文では，避難シミュレーションデータの分析を目的とし，テンソル分解によるアプローチを提案した．特に，移動軌跡データに焦点を当てて，避難における人の状態や移動の様子における分析を行った．避難シミュレーションデータを集計してテンソルを複数設計し，複数のテンソルを用いたテンソル分解の手法である NMTF を用いて分析を行った．テンソル分解によって生成された因子行列やランクごとに復元したテンソルを分析することによって，主要な避難行動や少数にしか見られない特徴的な行動パターンを獲得することができた．

今回は提供された一回分の避難シミュレーションデータに関して分析を行った．しかし，シミュレーションは複数回試行されていることが多いため，複数の結果の比較分析についても求められる．シミュレーション ID や設定パラメータを次元として設定してテンソルを設計することによって，テンソル分解による比較分析が行えると考えられる．多数のシミュレーションで共通して見られる結果や，少数のシミュレーションでしか見られない特徴の抽出が期待される．また，テンソル分解の精度向上のために，誤差尺度や最適化方法についても検討の余地がある．

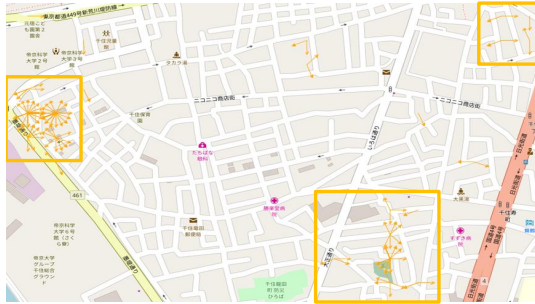


(a) 救助活動

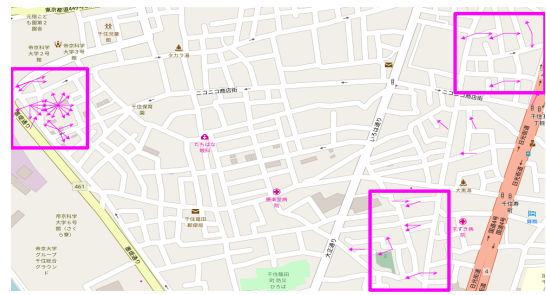


(b) 閉じ込め者

図 9: 各状態のランク別の人数



(a) rank6



(b) rank8

図 10: 10 時間後の移動の様子

謝 辞

避難シミュレーションデータの提供元である東京工業大学の
大佛研究室に感謝の意を表す。本研究の一部は、科研費
(16H01722), CREST「大規模・高分解能数値シミュレシ
ョンの連携とデータ同化による革新的地震・津波減災ビッグデ
ータ解析基盤の創出」による。

文 献

- [1] leaflet. <https://leafletjs.com>.
- [2] Paradigm4: Creators of SciDB a computational dbms. <http://www.paradigm4.com/>.
- [3] P. G. Brown. Overview of SciDB: Large scale array storage, processing and analysis. In *Proc. ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD 2010)*, pages 963–968, 2010.
- [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.*, 13(2):338–346, 2018.
- [5] H. Lustosa, F. Porto, P. Valduriez, and P. Blanco. Database system support of simulation data. *Proc. VLDB Endow.*, 9(13):1329–1340, 2016.
- [6] 沖 拓弥, 大佛 俊泰. 住民による救助活動を組み込んだ大地震時における木密地域の広域避難シミュレーション分析.
- [7] 松林達史, 幸島匡宏, 澤田宏. 非負値テンソル因子分解を用いた購買行動におけるブランド選択分析. *人工知能学会論文誌*, 30(6):713–720, 2015.
- [8] 松林達史, 幸島匡宏, 澤田宏. 複合データ分析技術と NTF(2・完) テンソルデータの因子分解技術と実応用例. *電子情報通信学会誌*, 99(7):691–698, jul 2016.
- [9] 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.*, 11(2):188–197, 2016.
- [10] T. Osaragi and T. Oki. Wide-area evacuation simulation incorporating rescue and firefighting by local residents. *J. Disaster Res.*, 12(2):296–310, 2017.
- [11] 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)*, pages 257–260, 2011.
- [12] A. Shashua and T. Hazan. Non-negative tensor factorization with applications to statistics and computer vision. In *Proc. Int'l Conf. on Machine Learning (ICML 2005)*, pages 792–799, 2005.
- [13] X. Song, Q. Zhang, Y. Sekimoto, and R. Shibasaki. Prediction of human emergency behavior and their mobility following large-scale disaster. In *Proc. ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2014)*, pages 5–14, 2014.
- [14] X. Song, Q. Zhang, Y. Sekimoto, R. Shibasaki, N. J. Yuan, and X. Xie. A simulator of human emergency mobility following disasters: Knowledge transfer from big disaster data. In *AAAI*, 2015.
- [15] M. Stonebraker, P. Brown, J. Becla, and D. Zhang. SciDB: A database management system for applications with complex analytics. *IEEE Computational Science & Engineering*, 15(3):54–62, 2013.
- [16] 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)*, pages 1199–1204, 2013.
- [17] Y. Wang, Y. Zheng, and Y. Xue. Travel time estimation of a path using sparse trajectories. In *Proc. ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2014)*, pages 25–34, 2014.
- [18] 中央防災会議. 東北地方太平洋沖地震を教訓とした地震・津波対策に関する専門調査会報告, Sep 2011.