

# テンソル分解を用いた避難移動軌跡データの分析

## Analysis of Evacuation Trajectory Data Using Tensor Decomposition

河井悠佑<sup>†</sup>  
Yusuke Kawai

石川佳治<sup>†</sup>  
Yoshiharu Ishikawa

杉浦健人<sup>†</sup>  
Kento Sugiura

### 1. はじめに

災害対策において、災害シミュレーションによる被害状況の想定は、重要な防災手段として期待されている。被害状況を想定することで、避難経路や防災計画の策定、都市のインフラ整備につながれる。近年では、計算機の性能向上及びストレージの容量拡大から、シミュレーションデータが多数生成されている。

しかし、この膨大なデータを利活用に着目した技術開発に関する研究は少ない。災害シミュレーションのモデル化に関する研究 [4] がなされている一方で、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。災害シミュレーションデータは、時間や空間に基づいた大規模で詳細なデータが、多数生成されていることが特徴である。そのため、多次元で膨大なデータに対する分析技術が求められる。

そこで本研究は、分析手法としてテンソル分解の利用を検討する。テンソル分解は、高次元のテンソルから、より低次元のテンソルに分解し、元の特徴を保ったまま必要な次元以外を削減する手法である。そのため、災害シミュレーションデータのような多次元で膨大な情報から、重要な災害や避難のパターンの獲得が期待できる。この手法は、購買履歴データにおけるアイテム推薦 [5] や、画像 [1] や音声 [2] におけるノイズ処理や特徴抽出など、様々な分野で応用されている。

2 章では、今回使用するデータセットについて述べる。3 章では、テンソル分解の先行研究として、補助行列を用いたテンソル分解について述べる。4 章では、データセットに関する分析シナリオについて述べる。5 章では、本稿のまとめと今後について述べる。

### 2. データセット

今回使用するデータセットの説明を行う。対象とするシミュレーションは、首都直下型地震を想定した広域避難シミュレーション [4] である。東京都北千住を対象として、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 2) と 2,720 本の道路 (表 3) の物的被害の様子をシミュレーションしたものである。人の避難行動データは、各時刻において座標と 14 種類の避難状態 (表 4) 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は、火災や閉塞の状況が記録されている。

### 3. 補助行列を用いたテンソル分解

移動軌跡データにテンソル分解を適用した先行研究 [3] について述べる。この研究では、補助行列 (context matrix) を追加した CATD (Context-Aware Tensor Decomposition) が用いられる。タクシーの移動軌

表 1: 移動軌跡

属性	意味
t	時刻
uid	人の ID
x, y	X, Y 座標
status	状態 (表 4)

表 2: 建物

属性	意味
t	時刻
uid	建物 ID
collapsed	倒壊状態
stage	火災段階

表 3: 道路

属性	意味
t	時刻
uid	道路 ID
rubble	閉塞状態
dens	人口密度

表 4: 人の状態区分

状態	意味
Home	発災時に自宅に滞在
Other	発災時に建物に滞在
Move	発災時に屋外で歩行
SIItoki	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
HIItoki	一時集合場所に移動
HIItoki2	一時集合場所に到達
HKoiki	広域避難場所に移動
HKoiki2	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

跡データを対象として、各道路の移動速度を推定する。作成するテンソルは道路セグメント、ドライバー、時間帯の三つを次元とし、成分を移動速度とする三次テンソル  $\mathcal{A} \in \mathbb{R}^{N \times M \times L}$  である。しかし、各ドライバーが通過する道路には限りがあるため、移動軌跡データだけではごく僅かな成分しか求められない。そこで、テンソルを因子行列に分解し、その因子行列から元のテンソルを復元することによって、未知の成分を補完する。しかし、非常に疎なテンソルでは補完精度が悪いため、補助行列を追加してメインテンソルの分解精度を向上させる。

三次のメインテンソル  $\mathcal{A}$  は、次元サイズの小さいコアテンソル  $S \in \mathbb{R}^{d_R \times d_U \times d_T}$  と、道路セグメント、ドライバー、時間帯に関する因子行列  $R \in \mathbb{R}^{N \times d_R}, U \in \mathbb{R}^{M \times d_U}, T \in \mathbb{R}^{L \times d_T}$  へと分解する。補助行列はメインテンソルと次元を共有した密な行列を作成する必要がある。この研究では、空間グリッドごとの車両数を集計した行列  $X$  と、道路セグメントの特徴に関する行列  $Y$  の二つを作成する。  $X$  は時間帯  $T \in \mathbb{R}^{L \times d_T}$  と空間グリッド  $G \in \mathbb{R}^{d_T \times P}$  の二つの因子行列に分解する。  $Y$  は道路セグメント  $R \in \mathbb{R}^{N \times d_R}$  と道路特徴  $F \in \mathbb{R}^{d_R \times Q}$  の二つの因子行列に分解する。  $X$  は時間帯の因子行列  $T$  を、  $Y$  は道路セグメントの因子行列  $R$  を、メインテンソルの因子行列とそれぞれ共有し、  $T, R$  の分解精度を向上させる。最終的な CATD の目的関数は式 (1) である。

<sup>†</sup>名古屋大学大学院情報学研究科  
Graduate School of Informatics, Nagoya University

$$\begin{aligned}
& \mathcal{L}(S, R, U, T, F, G) \\
&= \frac{1}{2} \|A - S \times_R R \times_U U \times_T T\|^2 \\
&+ \frac{\lambda_1}{2} \|X - TG\|^2 + \frac{\lambda_2}{2} \|Y - RF\|^2 \\
&+ \frac{\lambda_3}{2} (\|S\|^2 + \|R\|^2 + \|U\|^2 + \|T\|^2 + \|F\|^2 + \|G\|^2).
\end{aligned} \tag{1}$$

第1項はメインテンソルの分解による損失, 第2項, 第3項は補助行列の分解による損失, 第4項は過学習を防ぐための正則化項をそれぞれ表す.  $\lambda_1, \lambda_2, \lambda_3$  は, 分解の損失と正則化のバランスを制御するハイパーパラメータである. この目的関数について, 各変数に関して勾配法を適用することによって, 局所的最適解に収束するまで変数の更新を繰り返す. そして推定された  $S, R, U, T$  から, 補完したメインテンソル  $A_{rec} = S \times_R R \times_U U \times_T T$  を求める.

#### 4. 分析シナリオ

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検討する. 避難シミュレーションデータの特徴として以下が挙げられる.

- 時空間に基づいた大規模で詳細なデータ
- 多数のシミュレーション

この二点の特徴に基づいて分析シナリオを検討する.

##### 4.1. CATD による高精度な補完推定

計算機の性能向上やストレージの容量拡大により, 多数の属性を持った多量のデータが生成されることが多い. また, 実世界におけるシミュレーションであるため, 時間や空間に基づいたデータであることが多い. そのため, 多次元なテンソルを複数個作成でき, 複数のテンソルや行列を用いたテンソル分解である CATD の利用が期待できる. そこで, 3章で述べたような高精度なテンソルの補完を検討する.

避難シミュレーションを評価する一つの指標として, 状態別の人数が挙げられる. 避難が完了した人数や, ケガ等により避難が困難な人数, 死亡者数の大小は, 避難リスクの評価において重要である. そこで, 移動軌跡データから, 各時間帯, 各領域, 各状態における人数を集計した三次のテンソルを作成する. しかし, 人がいない地域の状態別人数は集計できず, 避難リスクの評価が難しい. この未知の成分をテンソル分解による補完推定により解決する. なお, 元データは二次元の空間で記録されているが, 今回は空間をグリッド状に分割した領域を分析の単位とする.

これに加えて時間帯や領域, 状態に関する補助行列を作成する. 例としていくつか列挙する.

**建物や道路の被害状況** 各時間帯ごとに各領域に存在する建物や道路の被害状況を集計する. 領域や時間帯の関連性を表す.

**状態間の遷移** 人の状態に関する遷移数を集計する. 人の状態間の関連性を表す.

**領域間の距離** 領域間の距離を計算する. 領域間の関連性を表す.

**状態別の時間** 各ユーザが各状態になった時間を計算する. 状態間の関連性や頻度を表す.

このような補助行列を追加することによって, メインテンソルの補完精度の向上を目指す.

#### 4.2. 複数のシミュレーション結果の特徴抽出

シミュレーションはシナリオやパラメータを定めて多数試行されることが多い. そのため, 複数のシミュレーションにわたる統合的な分析や比較が求められる. 異なるシナリオや初期条件においても共通で見られる結果は, そのシミュレーションにおいて保持しておくべき重要な情報である可能性が高い. 加えて, ほとんどのシミュレーション見られない情報も, 特異なパターンとして着目すべきである. 一方で, シミュレーションはランダム性を持って試行されるため, ノイズとなる結果が生成される場合がある. 複数のシミュレーションにおける, 重要な情報の保持や特異なパターン検出, ノイズの除去のために, テンソル分解の利用を検討する.

シミュレーション ID を次元として設定することによって, 複数のシミュレーションにわたるテンソルを作成する. 例としてシミュレーションにおける避難時間について考える. シミュレーション ID, 領域, 避難時間の階級を次元として, 発災時に滞在した領域ごとに, 避難時間の度数を成分とした三次テンソルを作成する. テンソル分解から求められた因子行列から, ノイズとなるシミュレーション結果の除去や, 多数のシミュレーションに共通する結果の強調が期待できる.

#### 5. おわりに

本稿では, テンソル分解による避難シミュレーションデータの分析アプローチを提案した. 今後はテンソル分解を実装して, メインテンソルの設計や補助行列の選択の有用性について検討する. また, 実際にテンソル分解後の結果について考察を行い, 避難シミュレーションの評価を行う.

#### 謝辞

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ー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.

#### 参考文献

- [1] S. Amnon and H. Tamir. Non-negative tensor factorization with applications to statistics and computer vision. In *ICML*, pp. 792–799, 2005.
- [2] A. Ozerov, C. Fvotte, R. Blouet, and J. L. Durrieu. Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation. In *ICASSP*, pp. 257–260, 2011.
- [3] Y. Wang, Y. Zheng, and Y. Xue. Travel time estimation of a path using sparse trajectories. In *SIGKDD*, pp. 25–34, 2014.
- [4] 沖拓弥, 大佛俊泰. 大地震時の地域住民による救助活動と逃げ遅れについて (都市計画). 日本建築学会関東支部研究報告集, Vol. 85, pp. 425–428, 2015.
- [5] 松林達史, 幸島匡宏, 林亜紀, 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, Vol. 30, No. 6, pp. 713–720, 2015.