

非負値テンソル因子分解を用いた R3 冬の札幌圏大雪における断面交通量分析

Section Traffic Data Analysis of Heavy Snowfall in Sapporo Area in the Winter of 2022 using Non-negative Tensor Factorization

北海道大学工学部環境社会工学科 ○学生員 橋本晟矢 (Seiya Hashimoto)
北海道大学大学院工学研究院 正員 杉浦聡志 (Satoshi Sugiura)

1. はじめに

令和3年度の冬、札幌市は市の想定を大幅に上回る大雪に見舞われた。一連の大雪により市内では公共交通機関の運休、除排雪の大幅な遅延、道路機能の低下による渋滞などが発生し、市民生活に重大な影響が及んだ。¹⁾

当時、大雪により道路供給性能が低下していたことは明らかであるが、空間的・時間的に見た影響の程度などといった当時の道路の具体的な状態は不明瞭であり、これを明らかにすることは今後の雪害対策に基礎的な知見をもたらすと考えられる。本稿では JARTIC が提供する断面交通量オープンデータ²⁾を対象として雪害による影響の空間的・時間的分布の特徴を分析することを目的とする。具体的には、多次元のデータの特性抽出を可能とする非負値テンソル因子分解を用いて断面交通量を日付、時間帯、観測地点別でそれぞれ特性を分析する。

テンソル分解や行列分解を用いて交通関連情報を特性分析した例はいくらか報告されている。原田ら³⁾は、携帯電話位置情報データにスパース非負値行列因子分解を適用することにより COVID-19 の影響による都道府県間を跨ぐ行動の変化の度合いを確認したとともに、都道府県間を移動する旅行パターンを抽出し各パターンの影響を確認した。また、もとの情報と非負値行列因子分解した行列による再現行列との残差から通常の行動変化パターンでは説明できない COVID-19 によると考えられる行動変化があることを確認した。桑野ら⁴⁾は、バス経路検索サービスのログデータに非負値テンソル因子分解を適用し、バス利用者の行動の特徴を抽出するとともにテンソル分解が交通特性分析に適用できることを明らかにした。

本稿では札幌市内の JARTIC の断面交通量データに非負値テンソル因子分解を適用して得られた特徴行列から断面交通量の分解された因子の特徴を確認する。次に、一度分解したテンソルを再度合成し、もとのテンソルとの残差を分析することによって一連の大雪で、いつどこで異常が発生し、その程度はどれほどであったかを明らかにする。

2. 分析手法

2.1 非負値テンソル因子分解の概要

テンソル分解とは、3 階以上のテンソルで表現されたデータの各要素軸の特徴を抽出することが可能な分析手法である。テンソルとは多次元の配列のことを指し、0 階のテンソルはスカラー、1 階のテンソルはベクトル、2 階

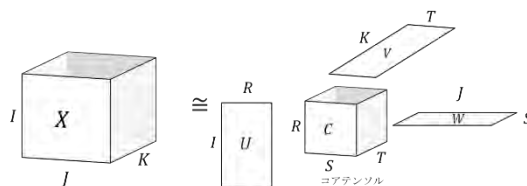


図-1 タッカー分解の概念図

のテンソルは行列である。分解によって分析対象のテンソルはコアテンソルと特徴行列に分解され、特徴行列に各要素軸が持っている特徴が抽出されることになる。また、元のテンソルと分解した結果から復元したテンソルの差をとった残差のテンソルには因子分解において説明できない特殊な変化の情報が含まれている。したがって、残差を分析することによって通常時とは異なる挙動を補足することができる。本研究ではこの性質を利用し、残差のテンソルを分析することによって大雪の際の異常な現象を捉えることを目指す。

テンソル分解の手法としては主に CP 分解とタッカー分解の2つが挙げられる。本研究では非負値タッカー分解を用いる。タッカー分解は大きさが $I \times J \times K$ である3階のテンソル X を例とすると、分解における入出力の概念図は図-1 のように表現することができ、次式のように定式化される。

$$X \cong C \times U^T \times V^T \times W^T \quad (1)$$

ここで、 C はコアテンソル、 U, V, W は特徴行列である。また、 R, S, T は各要素軸の次元を圧縮する際の次元数であり、因子数と呼ばれる。因子数は分析者が任意に設定する。因子数は、機能的にはある要素軸の特徴を分解する基底の数を決定する役割を担っている。

次に、実際に分解を行う流れについて簡単に述べる。式(1)におけるパラメータは C, U, V, W であり、式(2)で示される2乗誤差を最小にするようにパラメータを決定する。

$$\min_{C,U,V,W} \|X - C \times U^T \times V^T \times W^T\|_F^2 \quad (2)$$

ここで、 $\|\cdot\|_F^2$ はフロベニウスノルムを表す。本研究で用いるのは非負値タッカー分解であるため、これに加えてコアテンソルと特徴行列の全成分が非負であるという制約が加わる。

なお、本研究では python のライブラリ Tensorly により非負値タッカー分解を実装している。詳細な計算方法やアルゴリズムについては文末の参考文献⁵⁾を参照された

い。

2.2 使用するデータの概要

断面交通量データは JARTIC が提供する断面交通量オープンデータ²⁾を使用する。分析の期間は2020年11月1日から2021年3月31日、2021年11月1日から2022年3月31日のうち欠測の多かった2021年2月27日を除いた計301日間とし、時間帯は6時～9時台、10時～13時台、14時～17時台、18時～21時台の4つとする。分析期間を2020年11月以降としているのは、それ以前の冬期間が新型コロナウイルスの流行前であり、分析対象である令和3年度の冬季と状況が大きく異なることが予想されるためである。分析の対象とする範囲は定山溪地域を除いた札幌市内全域とし、範囲内の観測地点のうち、各日付の各時間帯の全てで少なくとも1回以上断面交通量が計測されている計1048地点を用いる。したがって、分析対象となるデータは日付(301日間)×時間帯(4時間帯)×観測地点(1048地点)の3要素を持つ3次元のテンソルとなる。このテンソルに対し、各要素の因子数をそれぞれ(日付, 時間帯, 観測地点) = (6, 4, 5)として非負値テンソル因子分解を適用する。

3. 非負値テンソル因子分解

本項では非負値テンソル因子分解により得られた各要素の特徴行列と、元のデータと分解した行列から復元したテンソルとの残差について分析を行う。

3.1 日付に関する特徴行列

ここでは抽出された6つの因子のうち、特徴的であった第1, 2, 4因子の3因子について述べる。

(1) 第1, 2因子

第1, 第2因子の日別推移を図-2に示す。ここで、縦軸は因子の成分の大きさ、横軸は日付を表しており、図

中のピンク色の縦棒は土日祝日を表している。図-2から、第1因子と第2因子は双方とも周期的な変動をしていることがわかり、第1因子は土日祝日以外の部分、第2因子は土日祝日や年末年始といった長期休暇の部分で成分が大きくなっていることが確認できる。このことから第1因子は平日における特徴を示しており、第2因子は土日祝日や長期休暇などにおける特徴を示していることがわかる。

(2) 第4因子

第4因子の成分の日別推移を図-3に示す。縦軸と横軸は第1, 2因子の場合と同様である。図-3から、第4因子は2021年の積雪期間と比較すると、2022年の積雪期間において大きな値をとっていることが確認できる。したがって、第4因子は2022年の冬の積雪期間の断面交通量の挙動がこの冬に特有なものであることを示唆しており、2021年の冬の積雪期間とは異なる状態であったことが示唆されていると考えられる。

3.2 時間帯に関する特徴行列

得られた時間帯に関する特徴行列を表-1に示す。表-1から、第1因子は4つの時間帯全てにおいて平均的な値をとっており、各時間帯に共通した特徴を示している因子といえる。第2因子は6時から9時台の時間帯において最も成分の値が大きく、18時から21時台に小さな値をとっている。したがって、朝の通勤の挙動を示してい

表-1 時間帯に関する特徴行列

	第1因子	第2因子	第3因子	第4因子
6～9時台	0.463816	0.77138	0.211309	0.000179
10～13時台	0.485486	0.412188	0.174536	1.97478
14～17時台	0.548043	0.366538	0.080103	1.88899
18～21時台	0.395765	0.004908	0.84846	0.403015

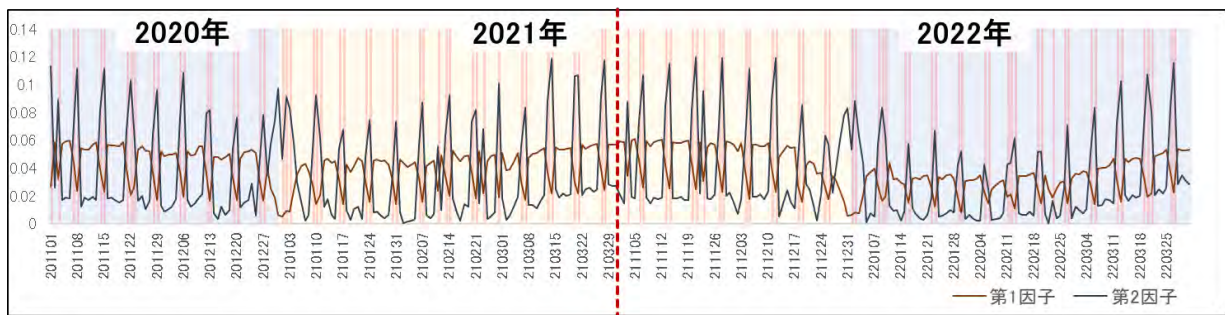


図-2 第1因子、第2因子の日別推移

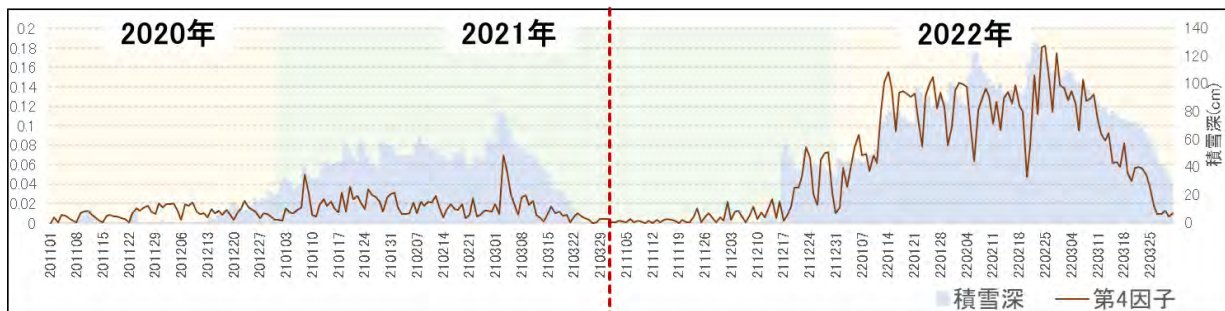


図-3 第4因子の日別推移

ると考えられる。一方で第3因子は18時～21時台の時間帯において成分の値が卓越しており、帰宅時間帯の行動を表す因子であると解釈できる。第4因子は10時～18時台の2つの時間帯において非常に大きな成分の値となっていることがわかる。このことからこの因子は日中の勤務時間帯の行動や買い出しなどといった日常生活上での行動を抽出していると考えられる。このように時間帯に関する特徴行列は時間帯による断面交通量の特徴を因子として表現できているといえる。

3.3 観測地点に関する特徴行列

観測地点に関する特徴行列で抽出された5つの因子のうち、ここでは空間的な特徴として解釈できる第3～5

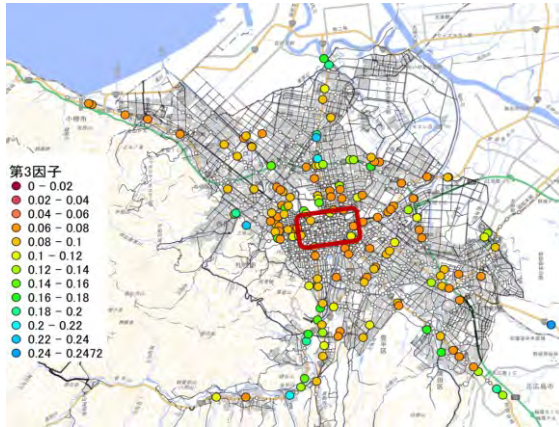


図-4 観測地点の第3因子

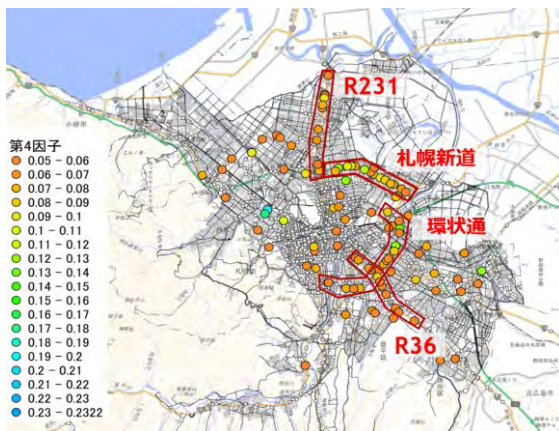


図-5 観測地点の第4因子

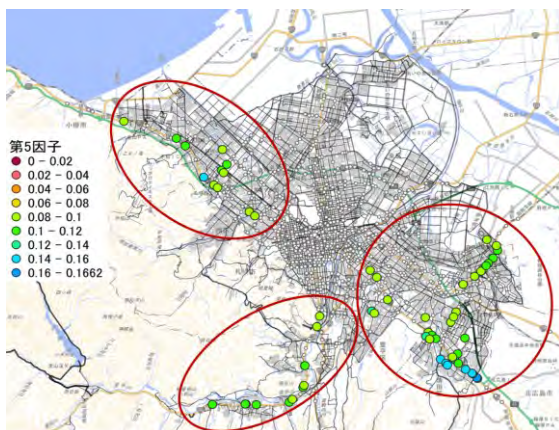


図-6 観測地点の第5因子

因子について述べる。

(1) 第3因子

第3因子の因子負荷量を色により表現し、観測地点の空間的配置と整合させたものを図-4に示す。図中では成分の値が0.06を上回った地点のみ抽出している。図-4により、札幌市内中心部は0.06を上回る地点がなく、市内中心部から放射状に伸びる主要幹線に成分の値が大きい部分が集中していることがわかる。このことから第3因

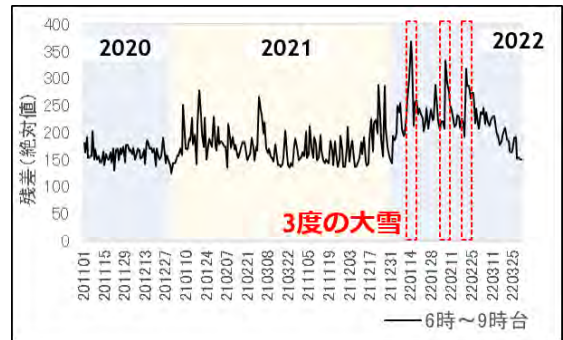


図-7 6時～9時台の残差の日別推移

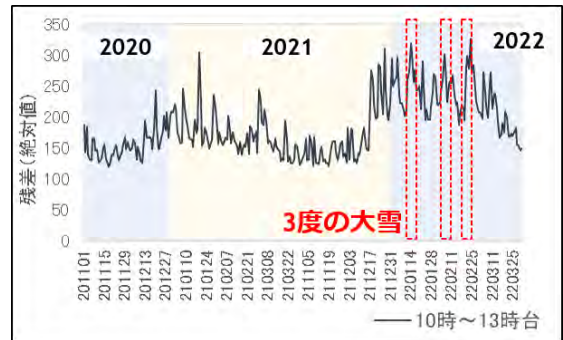


図-8 10時～13時台の残差の日別推移

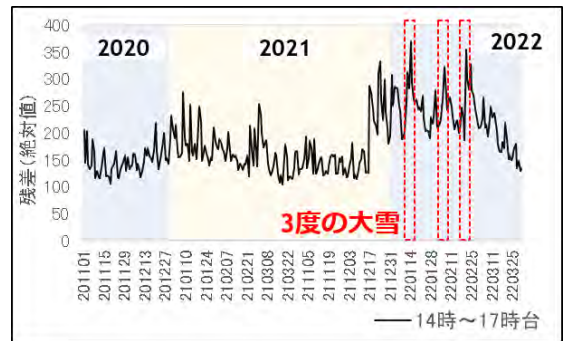


図-9 14時～17時台の残差の日別推移

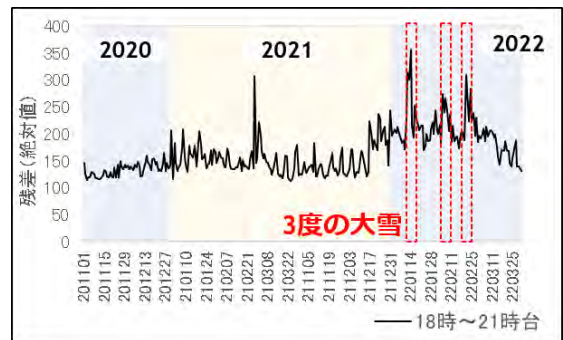


図-10 18時～21時台の残差の日別推移

子は都心部に入出入りする部分や都心部付近を移動する部分を主に示していると解釈できる。

(2) 第4因子

第4因子の因子負荷量を色により表現し、観測地点の空間的配置と整合させたものを図-5に示す。図中では成分の値が0.05を上回った地点のみ抽出している。図-5により、環状通や札幌新道といった市内の環状線や国道36号線、国道231号線などにおいて成分の値が大きくなっていることがわかる。第3因子を示す図-4と比較すると空間的に外側に広がっているようにも見て取ることができ、市内全体を動き回るような部分を示していると解釈することができる。

(3) 第5因子

第5因子の因子負荷量を色により表現し、観測地点の空間的配置と整合させたものを図-6に示す。図中では成分の値が0.08を上回った地点のみ抽出している。図-6により、図中左上の小樽方面に向かう箇所や右下の千歳・江別方面に向かう箇所、左下の定山溪方面に向かう箇所などで成分の値が大きな部分が集中していることがわかる。このことから第5因子は札幌とその周辺都市を繋ぐような郊外部を主に示していると解釈できる。

3.4 残差の分析

全観測地点で平均した交通量について、もとの観測値テンソルと分解により得られたコアテンソルと特徴行列の積をとった再現交通量との残差を分析する。この残差はすなわち分解された因子では表現できない特異な値を示す。残差の日別推移を時間帯別に示したものを、図-7～10に示す。縦軸は残差の絶対値の平均値を、横軸は日付を表している。図-7～10から、札幌市内で55cmの24時間降雪量を記録した12/17を境として4つの全ての時間帯において残差が積雪期間を通して大きな値が継続していることが確認できる。とくに、図-8～9に示される10時～17時台の時間帯については特にその傾向が顕著に現れているといえる。また、図-7～10に示した2022年冬期に共通して3回現れる残差が大きな部分は、札幌市の「令和3年度の大雪対応に係る検証と今後の対策」¹⁾において「特筆する3度の大雪」とされている2022年1月12～14日、2022年2月6日、2022年2月21日～23日の期間とその直後に一致していることがわかる。この結果は、2022年の冬の札幌市内では因子分解で説明できない、つまり大雪による異常な状態が冬の期間を通して継続していたことを示唆しているものと考えられる。

3.5 擬似逆行列による因子成分ごとの残差分析

残差の分析ではコアテンソルとすべての特徴行列の積をとった再現交通量テンソルを利用した。ここでは、要素軸の情報を特徴行列の擬似逆行列を用いて圧縮し、因子ごとに分析する。観測地点の情報を圧縮することを例として手順を示す。式(3)で表される残差テンソル E' を考える。

$$E' = Y - C \times U^T \times V^T \quad (3)$$

ここで、 Y は $Y = X \times \theta$ で得られるテンソルである。 θ はデータテンソル X に乗ずることで注目する要素軸(=観測地点)の情報をもとの特性を失わないよう圧縮する行列であり、式(4)の性質を満たす。

$$X \times \theta \cong C \times U^T \times V^T \times W^T \times \theta \quad (4)$$

これを満たす行列 θ は観測地点の特徴行列の擬似逆行列であり、 Y, E' はそれぞれ X, E の1048地点の観測地点の情報を5つの因子に圧縮したテンソルとなる。

この操作で得られた残差テンソル E' を分析することによって、各因子の特徴に基づいた残差の分析が可能になることが予想される。

4. おわりに

本稿では、令和3年度に生じた札幌の雪害における交通状態を分析するために、断面交通量データに非負値テンソル因子分解を適用し、その特性分析を行った。特徴行列の分解結果からは日付別の行列においては平日・休日での差異や2022年冬に特有の特徴がある点、時間帯別の行列からは時間帯によって異なる断面交通量の特徴がある点、観測地点別の行列からは地域によって特徴が異なる点をそれぞれ確認した。一方で残差の分析については、現時点では雪害発生時の残差量を把握するにとどまっておらず、空間的な特性を詳細に把握するには至っていない。擬似逆行列の因子成分ごとの残差分析の結果は講演時に報告する。

参考文献

- 1) 札幌市ホームページ：令和3年度の大雪対応に係る検証と今後の対策
https://www.city.sapporo.jp/kensetsu/yuki/documents/r3ooyuki_houkokusho.pdf (2022/12/05 閲覧)
- 2) JARTIC：各種情報の提供（オープンデータ）
<https://www.jartic.or.jp/service/opendata/> (2022/12/05 閲覧)
- 3) 原田 魁成, 山口 裕通, 寒河江 雅彦：スパース非負値行列因子分解を用いた COVID-19 流行期の県間旅行行動の変容分析, 土木学会論文集 D3 (土木計画学), Vol. 77, No. 2, 160 - 173, 2021.
- 4) 桑野 将司, 木下 礼央, 福山 敬, 谷本 圭志, 菅原一孔, 川村 尚生：非負値テンソル因子分解を用いた検索ログデータからのバス利用特性の分析, 土木学会論文集 D3 (土木計画学), Vol.73, No.5 (土木計画学研究・論文集第34巻), I_569-I_578, 2017.
- 5) Yong-Deok Kim and Seungjin Choi, "Non-negative tucker decomposition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1-8, 2007