

# ドリフトベクトルモデルによる 非対称な非類似度に基づく多次元尺度構成法について

2020SS059 逢坂駿

指導教員：塩濱敬之

## 1 はじめに

MDS（多次元尺度構成法）は物事の距離や類似度などのデータを使って、多変量データを可視化する手法である。この時に用いられるのは対称な距離や類似度、誤差を無くしたデータを使用している。しかし、実際のデータ解析においては非対称な距離や類似度のデータも多い [2]。本研究では非対称 MDS（非対称多次元尺度構成法）を使用して非対称なデータの可視化を行う。

## 2 ドリフトベクトルモデル

非対称な距離行列の例として、以下の表 1 に用意した愛知県の 7 つの市町村間の市役所（町役場）を車で移動する際にかかる予想所要時間のデータ（2023 年 6 月 12 日 7 時 20 分）を用いる。

表 1 2023 年 6 月 12 日 7 時 20 分

(着) \ (発)	名古屋市	一宮市	常滑市	豊橋市	岡崎市	豊田市	設楽町
名古屋市	0	65	95	148	115	55	142
一宮市	51	0	121	180	140	118	169
常滑市	89	113	0	125	93	100	158
豊橋市	143	182	124	0	72	95	75
岡崎市	102	140	90	75	0	47	77
豊田市	61	120	105	104	52	0	77
設楽町	149	189	172	91	78	77	0

このような非対称な非類似度（距離）に対する多次元尺度構成法のモデルにドリフトベクトルモデルがある [1]。ドリフトベクトルモデルでは、2 次元配置の座標を表す点の配置とその点の変化の方向と強さを示す矢印（ドリフトベクトル）を図に同時に表示することができる。これらのドリフトベクトルの長さや方向が非対称行列の行の値に対応している。ドリフトベクトルの 2 次元空間における長さや方向は以下のように計算する [5]。

1. 非対称な近接行列  $P = [p_{ij}]$  に対して対称行列  $A = [a_{ij}]$  と非対称成分からなる行列  $B = [b_{ij}]$  を計算する。

$$a_{ij} = (p_{ij} + p_{ji})/2, \quad b_{ij} = (p_{ij} - p_{ji})/2$$

2. 行列  $A$  に対して、非計量 MDS を適用する。
3. MDS によって得られた  $x_i$  と  $x_j$  ( $i \neq j$ ) に対して、ベクトル  $y_{ij} = x_j - x_i$  を計算する。
4.  $y_{ij}$  を単位長に正規化する。  $y_{ij}^* = y_{ij}/(y_{ij}^T y_{ij})^{1/2}$ 。
5.  $y_{ij}^*$  に近接行列の非対称成分の要素  $b_{ij}$  をかける。  $z_{ij} = b_{ij} \times y_{ij}^*$ 。
6. 地点  $i$  に対して  $z_{ij}$  の平均を求める。  $w_i = n^{-1} \sum_j z_{ij}$ 。

7.  $w_i$  の長さ  $\|w_i\| = \sqrt{w_{i1}^2 + w_{i2}^2}$  を計算する。

8.  $y$  軸に対する角度  $\alpha_i = \arccos\{w_i u^T / \sqrt{w_i \cdot w_i^T}\}$  を計算する。ただし、 $w_i = (w_{i1}, w_{i2})$ ,  $u = (0, 1)$  である。

表 1 の愛知県主要市町村間の予想所要時間のデータに対して、ドリフトベクトルモデルのアルゴリズムのステップ 1 を適用すると、非対称な非類似度行列は次のように分解される。

$$A = \begin{bmatrix} 0 & 58 & 92 & 145.5 & 108.5 & 58 & 145.5 \\ 58 & 0 & 117 & 181 & 140 & 119 & 179 \\ 92 & 117 & 0 & 124.5 & 91.5 & 102.5 & 165 \\ 145.5 & 181 & 124.5 & 0 & 73.5 & 99.5 & 83 \\ 108.5 & 140 & 91.5 & 73.5 & 0 & 49.5 & 77.5 \\ 58 & 119 & 102.5 & 99.5 & 49.5 & 0 & 77 \\ 145.5 & 179 & 165 & 83 & 77.5 & 77 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 7 & 3 & 2.5 & 6.5 & -3 & -3.5 \\ -7 & 0 & 4 & -1 & 0 & -1 & -10 \\ -3 & -4 & 0 & 0.5 & 1.5 & -2.5 & -7 \\ -2.5 & 1 & -0.5 & 0 & -1.5 & -4.5 & -8 \\ -6.5 & 0 & -1.5 & 1.5 & 0 & -2.5 & -0.5 \\ 3 & 1 & 2.5 & 4.5 & 2.5 & 0 & 0 \\ 3.5 & 10 & 7 & 8 & 0.5 & 0 & 0 \end{bmatrix}$$

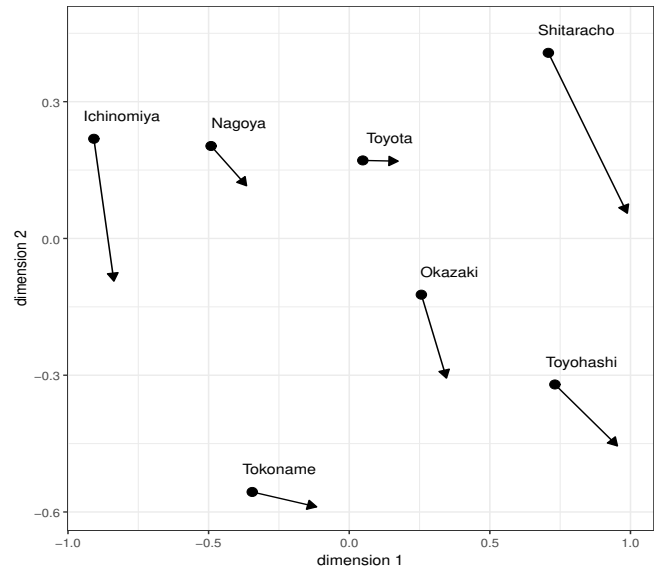


図 1 ドリフトベクトルモデルによる愛知県主要市町村の 2 次元配置

図 1 は、愛知県主要市町村のドリフトベクトルモデルによる 2 次元座標の配置をプロットしたものである。図の点は、対称な距離行列  $A$  に基づく MDS による座標で、ドリフトベクトルモデルによって得られた矢印の長さや角度が非対称性を表している。

### 3 提案手法

ドリフトベクトルモデルのステップ 2 において得られた座標のベクトルを  $\mathbf{x}_i^{(A)}$ , またドリフトベクトルモデルによる座標を  $\mathbf{x}_i^{(B)}$  と表す. (非) 計量 MDS に対して, 当てはまりの尺度としてストレス値が一般によく利用されている. ストレス値とは, 2 地点  $i, j$  間の本来の距離  $d_{ij}(\mathbf{X})$  と, MDS によって表現された距離  $\hat{d}_{ij}$  の平均 2 乗誤差の平方根であり, 値が小さいほど当てはまりの良い尺度として解釈される.

$$\text{Stress} = \sqrt{\frac{\sum_{i < j} (\hat{d}_{ij} - d_{ij}(\mathbf{X}))^2}{\sum_{i < j} d_{ij}^2(\mathbf{X})}}$$

ドリフトベクトルの長さの決定について, 様々な提案手法がある [4]. 本研究では, 地点  $i, j$  間の距離として, ドリフトベクトルモデルによって得られる対称な非類似度による座標  $\mathbf{x}_i^{(A)}$  とドリフトベクトルモデルによる座標  $\mathbf{x}_j^{(B)}$  を用いて以下のような距離を定義する.

$$\hat{d}_{ij} = \begin{cases} \|\mathbf{x}_i^{(A)} - \mathbf{x}_j^{(B)}\| & i < j \\ \|\mathbf{x}_j^{(A)} - \mathbf{x}_i^{(B)}\| & i > j \end{cases}$$

この距離に基づいて, ドリフトベクトルモデルのストレス値を次のように定義する.

$$\text{Stress}^{(P)} = \sqrt{\frac{\sum_{i,j=1}^n (\hat{d}_{ij} - d_{ij}(\mathbf{X}))^2}{\sum_{i,j=1}^n d_{ij}^2(\mathbf{X})}}$$

ドリフトベクトルモデルのアルゴリズムのステップ 7 において, ドリフトベクトルの長さはプロットする際に分析者が自由に調整できる. そのスケールを  $\rho$  と表すと, 点  $\mathbf{x}_i^{(B)}(\rho) = (x_{i1}^{(B)}(\rho), x_{i2}^{(B)}(\rho))$  は点  $\mathbf{x}_i^{(A)}(\rho) = (x_{i1}^{(A)} + \rho \|w_i\| \cos \alpha_i, x_{i2}^{(A)} + \rho \|w_i\| \sin \alpha_i)$  と表すことができる. 最適な  $\rho$  はストレス値  $\text{Stress}^{(P)}(\rho)$  を最小にするように選択すれば良い.

$$\hat{\rho}^{(opt)} = \underset{\rho \in \mathbb{R}}{\text{argmin}} \text{Stress}^{(P)}(\rho)$$

提案手法によって, 推定された先程のデータの  $\hat{\rho}$  の値は  $\hat{\rho} = -0.14$  であった. その時のドリフトベクトルによる距離行列は次のようになった.

$$\hat{d}_{ij}(\hat{\rho}^{opt}) = \begin{bmatrix} 0 & 0.42 & 0.77 & 1.32 & 0.81 & 0.53 & 1.21 \\ 0.41 & 0 & 0.95 & 1.72 & 1.21 & 0.95 & 1.62 \\ 0.78 & 0.97 & 0 & 1.09 & 0.74 & 0.82 & 1.43 \\ 1.34 & 1.73 & 1.11 & 0 & 0.52 & 0.85 & 0.74 \\ 0.82 & 1.22 & 0.75 & 0.50 & 0 & 0.36 & 0.70 \\ 0.55 & 0.96 & 0.83 & 0.83 & 0.35 & 0 & 0.70 \\ 1.22 & 1.63 & 1.43 & 0.72 & 0.69 & 0.70 & 0 \end{bmatrix}$$

またそのときのストレス値は

$$\text{Stress}(\hat{\rho}^{opt}) = 0.01160394$$

となった.

### 4 データ解析

10月1日(日)の愛知県の7つの市町村間の市役所(町役場)を車で移動する際にかかる予想所要時間のデータをドリフトベクトルモデルを使って比較し, データ解析を行う.

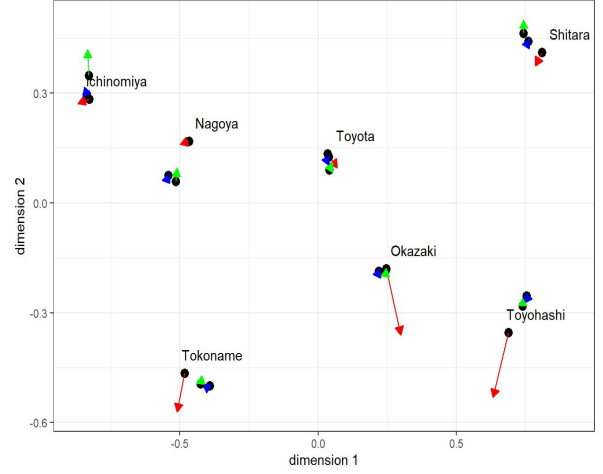


図2 10月1日7時から9時

図2は10月1日(日)の7時から9時までのデータをドリフトベクトルモデルによってプロットしたものである. それぞれ矢印は赤色が7時, 青色が8時, 緑色が9時を表している. 7時では, 南方向の地点に向かう時間の方が長くなる傾向がみられる. 8時では, 北方向の地点に向かう時間の方が長くなる傾向がみられる. 9時では, 北方向の地点に向かう時間の方が長くなる傾向がみられる.

### 5 まとめ

今回は非対称な非類似度のデータを多次元尺度構成法で推定するモデルとして, ドリフトベクトルモデルを用いた手法について検討した.

### 参考文献

- [1] Borg, I., Groenen, P. (2005), Modern multidimensional scaling. Theory and applications. Second Edition, Springer-Verlag, New York.
- [2] Chino N. (1978), A graphical technique for representing the asymmetric relationship between N objects, Behaviormetrika, no. 5, 23-40.
- [3] 永田靖., 棟近雅彦. (2001), 多変量解析法入門. サイエンス社, 東京.
- [4] Okada A., and Imaizumi T. (2022), Determining the length of drift vector I and II, Proceedings of the 50th Annual Meeting of the Behaviormetric Society.
- [5] Zaborski, A. (2016), Switch preference analysis by the drift vectors method.