

Activity pattern similarity:
a multidimensional sequence
alignment method

Chang-Hyeon Joh, Theo Arentze,
Frank Hofman, Harry Timmermans

東京大学大学院羽藤研究室

M1 松村草也

2008/10/30

研究の背景

- アクティビティパターンの類似性
 - 活動目的 / 活動地点 / 開始時刻 / 終了時刻 / 継続時間 / 交通手段 といった指標による判断
- 研究目的
 - 類似性指標はアクティビティパターンの基礎分析において重要な役割を持つ
 - アクティビティモデルの精度検証においても同様.

問題設定

ソースパターン

$$s = s[s_1, \dots, s_k, \dots, s_K]$$

$$s_k = s_k[s_{k0}, \dots, s_{ki}, \dots, s_{km}]$$

ターゲットパターン

$$g = g[g_1, \dots, g_k, \dots, g_K]$$

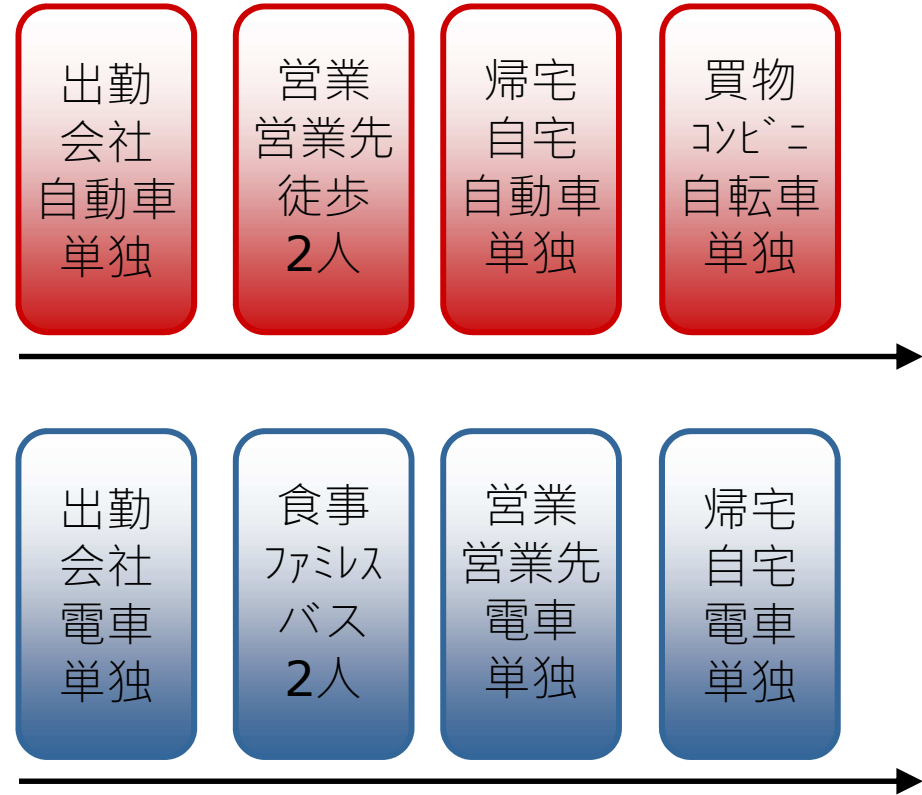
$$g_k = g_k[g_{k0}, \dots, g_{kj}, \dots, g_{kn}]$$

k:attributeの数

i:ソースパターンのステップ数

j:ターゲットパターンのステップ数

$$i=0, \dots, m; j=0, \dots, n; k=1, \dots, K; s_{ki} \in A_k, g_{kj} \in A_k$$



attribute : 活動属性・
活動目的 / 活動場所 / 交通手段 / 同行者
など

SAM (Sequence Alignment Method)

- 4つの操作： () 内はスコア
 - Deletion:削除 (1)
 - Insertion:挿入 (1)
 - Substitution:置換 (2)
 - Identity:確認 (0)
- 異なる2つの活動属性シーケンスを等しい形にする・スコア合計が最小の操作セットを求める。

操作例

s[ACB] → s[ABC] にかかる操作コストは？

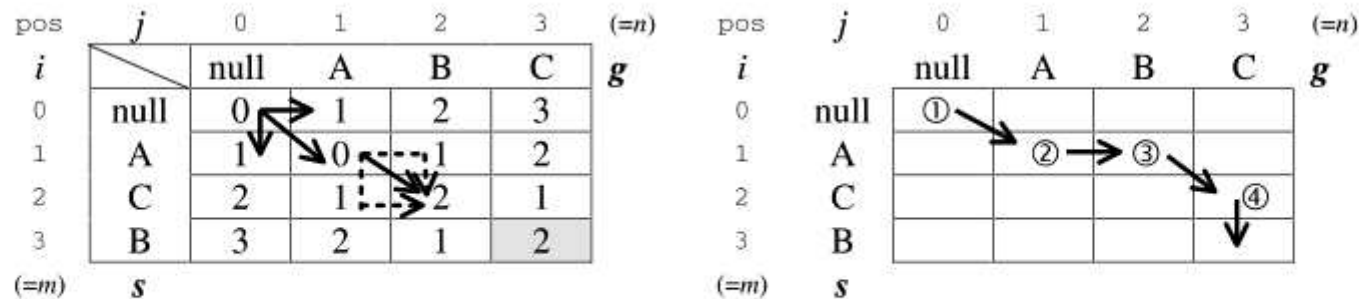


Fig. 1. An example of comparison table and its optimum trajectories. *Note:* The null elements initialize the comparison. The operation weights of deletion, insertion, substitution and identity are given 1, 1, 2 and 0, respectively.

	null	A	B	C
null	ACB	AACB	ABACB	ABCACB
A	CB	ACB	ABCB	ABCCB
C	B	AB	ABB	ABCB
B	null	A	AB	ABC

右方向：挿入
 下方向：削除
 右下方向：確認or置換

統合手法の提案

問題点

複数のattributesを考慮したモデルへの拡張・
従来手法では単純に重ね合わせ (UDSAM)
→ 個々のattributeについてsequenceの類似性のoperation costを計算し、
足し合わせると、attribute間の相互作用を考慮できない。

- 「Segment」の提案
- β : attributeの重み
- w : 個々のoperationのcost

操作コストの定義：

$$C = \max(\beta_1, \beta_2) \times w_d$$

$$(\beta_1 + \beta_2) \times w_d - \max(\beta_1, \beta_2) \times w_d = \min(\beta_1, \beta_2) \times w_d$$

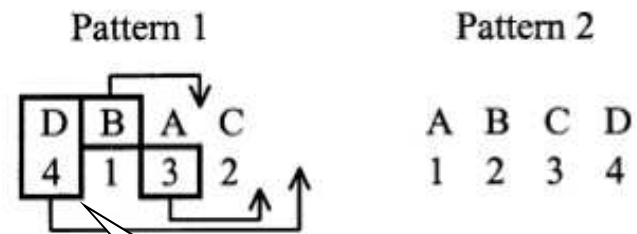


Fig. 2. A comparison of two-dimensional patterns.

複数のOperationを
Segmentとして
まとめて考える。

1次元オペレーションセット (Uni-dimensional operation set)

1. operationの内容
2. operationの位置
3. operationが適用されたattribute

によって操作を記述する。

$$O_{k,l} = \{ p \mid p = \underbrace{d(i, k)}_{\text{deletion}} \vee \underbrace{i(j, k)}_{\text{insertion}} \vee \underbrace{s(i, j, k)}_{\text{substitution}} \}$$

- $d(i,k)$: s_k の*i*番目の要素を削除
- $i(j,k)$: g_k の*j*番目の要素を s_k の*j*番目の要素へ挿入
- $s(i,j,k)$: g_k の*j*番目の要素と s_k の*i*番目の要素を置換

$$i \in \{1, \dots, m\}, \quad j \in \{1, \dots, n\}, \quad 1 \leq k \leq K$$

ex. 左図の操作セット

$$O_{2,1} = \{i(1,2), i(3,2), d(2,2), d(3,2), i(5,2)\}$$

$$O_{2,2} = \{i(1,2), i(2,2), d(3,2), d(3,2), i(5,2)\}$$

→異なる経路が*l*通りある。

pos	<i>j</i>	0	1	2	3	4	5	(=n)
<i>i</i>		null	A	B	C	D	E	<i>g</i>
0	null	0	1	2	3	4	5	
1	B	1	2	3	4	5		
2	E	2	3	4	5			
3	A	3	2	3	4	5		
4	D	4	3	4	5			
(=m)	<i>s</i>							

Fig. 3. Two different trajectories of the same operation set.

OT-MDSAM:

Optimum Trajectories-based MultiDimensional Sequence Alignment Method

多次元オペレーションセット

$$T = \prod_{k=1}^K T_k$$

T:attributeごとの最小操作量経路の組み合わせ数・
最小操作量を満たす経路が複数存在する時・
それらの集合を多次元オペレーションセットと呼ぶ・

$$O^t = \{p \mid p = d(i, \mathbf{k}) \vee i(j, \mathbf{k}) \vee s(i, j, \mathbf{k})\}$$

たとえば、右のような例を考える・
3つのattribute・

$$O_{1,l'}^t = \{d(2,1), i(4,1), s(5,6,1)\},$$

$$O_{2,l}^t = \{d(3,2), i(4,2), s(5,5,2)\},$$

$$O_{3,l''}^t = \{d(2,3), i(4,3), s(5,5,3)\}$$



$$\underline{O^t = \{d(2, \{1,3\}), d(3, \{2\}), i(4, \{1,2,3\}), s(5,5, \{2,3\}), s(5,6, \{1\})\}},$$

多次元類似性

attributeの組み合わせごとにコストを計算した合計.

$$C^t = \sum_{p \in O_t} c_p$$

$$\text{where } c_p = \begin{cases} w_d \beta_k^{\max} & \text{if } p = d(i, \mathbf{k}) \\ w_i \beta_k^{\max} & \text{if } p = i(j, \mathbf{k}) \\ w_s \beta_k^{\max} & \text{if } p = s(i, j, \mathbf{k}) \end{cases}$$

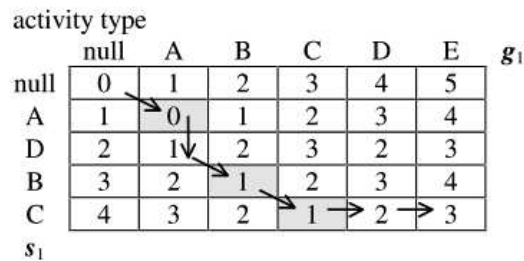
多次元類似性指標を以下のように定義する.

$$C^{t*} = \min[C^1, \dots, C^t, \dots, C^T]$$

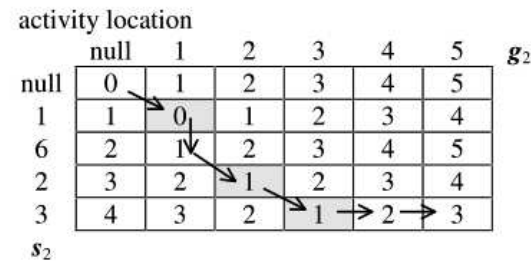
計算実行例

Source pattern				Target pattern					
A	D	B	C	A	B	C	D	E	→ activity-type sequence
1	6	2	3	1	2	3	4	5	→ activity-location sequence
a	b	c	f	a	b	c	d	e	→ transport-mode sequence
α	δ	φ	γ	α	φ	γ	δ	ε	→ accompanying-person sequence

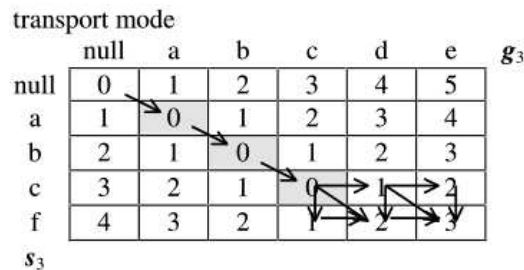
活動目的
Tk=1



活動場所
Tk=1



交通手段
Tk=5



同行者
Tk=1

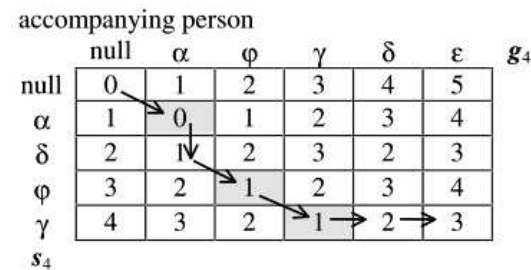


Fig. 4. Optimum trajectories of each attribute sequence.

計算実行例

$$O_{1,1} = \{d(2, 1), i(4, 1), i(5, 1)\}, \quad O_{1,2} = \{d(2, 2), i(4, 2), i(5, 2)\},$$
$$O_{1,3} = \{i(4, 3), i(5, 3), d(4, 3)\}, \quad O_{1,4} = \{d(2, 4), i(4, 4), i(5, 4)\},$$

$$O_{2,1} = \{d(2, 1), i(4, 1), i(5, 1)\}, \quad O_{2,2} = \{d(2, 2), i(4, 2), i(5, 2)\},$$
$$O_{2,3} = \{i(4, 3), s(4, 5, 3)\}, \quad O_{2,4} = \{d(2, 4), i(4, 4), i(5, 4)\},$$

$$O_{3,1} = \{d(2, 1), i(4, 1), i(5, 1)\}, \quad O_{3,2} = \{d(2, 2), i(4, 2), i(5, 2)\},$$
$$O_{3,3} = \{i(4, 3), i(4, 3), d(5, 3)\}, \quad O_{3,4} = \{d(2, 4), i(4, 4), i(5, 4)\},$$

$$O_{4,1} = \{d(2, 1), i(4, 1), i(5, 1)\}, \quad O_{4,2} = \{d(2, 2), i(4, 2), i(5, 2)\},$$
$$O_{4,3} = \{s(4, 4, 3), i(5, 3)\}, \quad O_{4,4} = \{d(2, 4), i(4, 4), i(5, 4)\},$$

$$O_{5,1} = \{d(2, 1), i(4, 1), i(5, 1)\}, \quad O_{5,2} = \{d(2, 2), i(4, 2), i(5, 2)\},$$
$$O_{5,3} = \{i(4, 3), i(4, 3), d(5, 3)\}, \quad O_{5,4} = \{d(2, 4), i(4, 4), i(5, 4)\}.$$

計算実行例

$$O_1 = \{d(2, \{1, 2, 4\}), d(4, \{3\}), i(4, \{1, 2, 3, 4\}), i(5, \{1, 2, 3, 4\})\},$$

$$O_2 = \{d(2, \{1, 2, 4\}), i(4, \{1, 2, 3, 4\}), i(5, \{1, 2, 4\}), s(4, 5, \{3\})\},$$

$$O_3 = \{d(2, \{1, 2, 4\}), d(4, \{3\}), i(4, \{1, 2, 3, 4\}), i(5, \{1, 2, 3, 4\})\},$$

$$O_4 = \{d(2, \{1, 2, 4\}), i(4, \{1, 2, 4\}), i(5, \{1, 2, 3, 4\}), s(4, 4, \{3\})\},$$

$$O_5 = \{d(2, \{1, 2, 4\}), d(4, \{3\}), i(4, \{1, 2, 3, 4\}), i(5, \{1, 2, 3, 4\})\}.$$

$$C^1 = 2 \times 1 + 1 \times 1 + 2 \times 1 + 2 \times 1 = 7,$$

$$C^2 = 2 \times 1 + 2 \times 1 + 2 \times 1 + 1 \times 2 = 8,$$

$$C^3 = 2 \times 1 + 1 \times 1 + 2 \times 1 + 2 \times 1 = 7,$$

$$C^4 = 2 \times 1 + 2 \times 1 + 2 \times 1 + 1 \times 2 = 8,$$

$$C^5 = 2 \times 1 + 1 \times 1 + 2 \times 1 + 2 \times 1 = 7.$$

$$C^* = \min[C^1, C^2, C^3, C^4, C^5] = 7.$$

手法の有効性検証

- OT-MDSAMとUDSAMの比較
 - 従来型の単純加算手法に比べて、提案手法を用いることで活動属性間の相互作用について、より有効な知見が得られていることを示す。
- 提案手法の計算可能性
 - データ構造を縮約したことによる計算速度の向上について検証を行う。

検証データ概要

- アクティビティダイアリーデータ (97年オランダ・2都市)
- 連続した2日間についてアンケート
- 2974パターン中の71パターン, $(71 \times 71 - 71) / 2 = 2485$ ペアを対象とした。
- 活動属性は活動目的, 活動場所, 交通手段の3種類。
 - 活動目的: 25種類 + 自宅
 - 活動場所: 32施設
 - 交通手段: 4種類 (車, 徒歩, 自転車, 助手席の)



類似性検証結果

Table 1
Pairwise comparison results

	Number of comparisons	Mean	Minimum	Maximum	Standard deviation
Sum of UDSAMs	2485	27.7336	2	73	12.2389
OT-MDSAM	2485	13.1038	2	32	5.1867

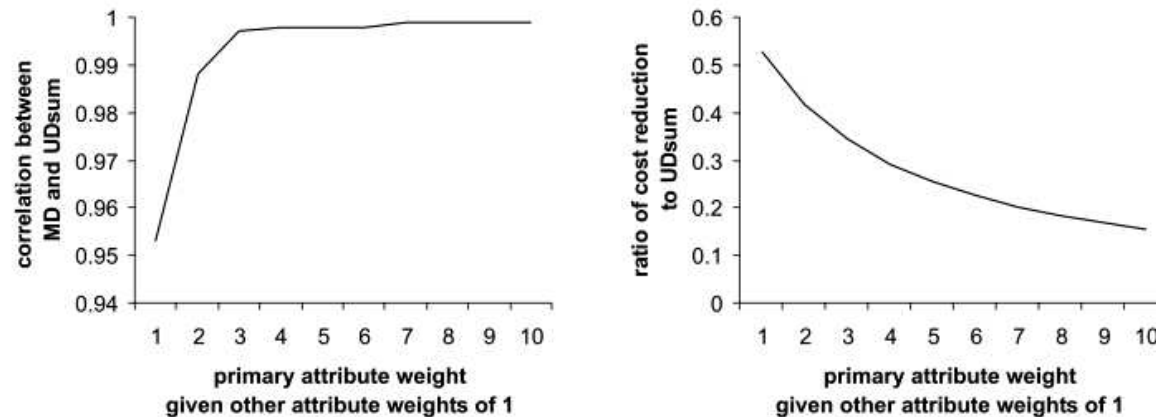


Fig. 6. Correlation between sum of UDSAM costs and OT-MDSAM cost with varying weights of activity-type attribute.

- 2485ペアを対象に類似性コストを2つの方法で計算し、結果を比較。
- 提案手法が算出した平均コストが従来手法に比べて14.63ポイント低下。
- OT-MDSAMのプライマリ活動属性の重みを3以上の値に設定すると、UDSAMの結果との相関係数がほぼ1となる（左図）。同時にUDSAMの結果に対する平均コストの「減少率」が下がる（右図）。

活動属性間比較

Table 2
Correlation between sum of UDSAMs and two-dimensional OT-MDSAM

Attributes included in the activity patterns	Pearson's r	Spearman's r
Activity type, activity location	0.977	0.977
Activity type, transport mode	0.970	0.973
Activity location, transport mode	0.967	0.969

検証では、活動目的は行き先よりも交通手段のほうが相関関係にある、という結果が得られている。

交通手段と移動先の関係のほうが、移動先と活動内容に比べ、固定されているといえる。

計算速度に関する検証

Table 3
Performance of OT-MDSAM versus ExMDSAM

	MD cost		Number of MD operation sets		Computing time ^a
	< ^b	= ^c	Mean	Standard deviation	
ExMDSAM	1	2484	853105.1	16425997.0	62 h 47 min 32 s
OT-MDSAM			667.7	3551.8	40 s

^a Computations based on a 350 MHz Pentium II processor.

^b The number of cases where OT-MDSAM costs are bigger than ExMDSAM costs.

^c The number of cases where OT-MDSAM costs are the same as ExMDSAM costs.

従来手法に比べ、格段に計算速度が上がっている。

Segmentを用いることによるデータの縮約の効果が現れているといえる。

まとめ

- シーケンシャル情報，活動属性間の相互依存性について説明を与える指標として，活動パターン間の類似性を提案した．
- 特に，多次元での活動パターン比較を可能にした．複数要素をSegmentとしてまとめることを提案した．