

観測データに基づく人の移動履歴の推定

梅谷 俊治 (大阪大学), 熊野徹 (JFE スチール), 蓮池 隆 (大阪大学), 森田 浩 (大阪大学)

概要

防災, 交通, マーケティングなどの分野で, 個人の移動履歴を収集・加工したデータベースを利用して時々刻々と変化する人の流れを解析する研究や調査が盛んに行われている. しかし, これらのデータベースは個人のプライバシー情報を多く含むため, 企業や個人などの幅広い層にデータベースを完全な形で公開することは困難である. また, 収集した個人の移動履歴データの欠損や不備によりデータベースの構築が困難な場合も少なくない. 本論文では, 地域内の各地点に配置されたセンサから観測される各時刻の交通量を入力データとし, 時空間に拡張された大規模なネットワーク上における整数計画問題を解くことで, 限られた観測データから個人の移動履歴を推定する手法を提案する.

1 はじめに

防災, 交通, マーケティングなどの分野で, 個人の移動履歴を収集・加工したデータベースを利用して時々刻々と変化する人の流れを解析する研究や調査が近年盛んに行われている [8, 9]. 特に最近では, Suica などの交通系 IC カードや携帯端末など個人の移動履歴データを取得する手段が整備されるようになり, データベースの充実にともない潜在的な需要は増加の一途をたどっている [1, 7]. しかし, これらのデータベースは個人のプライバシー情報を多く含むために, 企業や個人など幅広い層にデータベースを完全な形で公開することは困難である. また, 個人の移動履歴データの欠損や不備によりデータベースの構築が困難な場合も少なくない. このような状況下で, 限られた観測データから個人の移動履歴を高い精度で推定する手法は, 現代社会のサービスを支える重要な基盤になると期待できる.

本論文では, 地域内の各地点に配置されたセンサから観測される各時刻の交通量を入力データとし, 人の移動経路の選択に関する適当なモデルの下で, 時空間に拡張された大規模なネットワーク上における整数計画問題を解くことで, 各人の移動履歴を推定する手法を提案する (図 1).

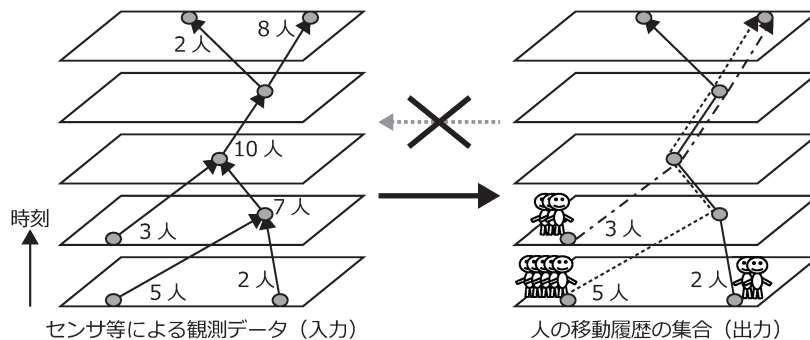


図 1: 観測データに基づく人の移動履歴の推定

センサから得られる部分的な観測データから観測対象の移動経路を推定する問題は移動体追跡問題 (multi-target tracking problem) と呼ばれ, 例えば, レーダにより取得した観測データから飛

行物体の識別と航路を推定する [4, 5], 監視カメラにより取得した動画データから人物の識別と移動経路を推定する [3, 6] など, 多くの分野に応用を持つ基本的な問題である [10]. しかし, 大規模な都市圏において数万~数百万人の移動履歴の推定は今までにない大規模な移動体追跡問題であり, 本論文ではこの問題をネットワーク上における整数計画問題に定式化することで実用に耐える効率的な解法を実現する.

2 問題の定式化

本論文では, 地域内の交通網を表すネットワークを時間軸方向に拡張した時空間ネットワーク (time space network) を用いて各地点に配置されたセンサから観測される各時刻の交通量を枝のフロー量で表すことで問題を定式化する (図 2).

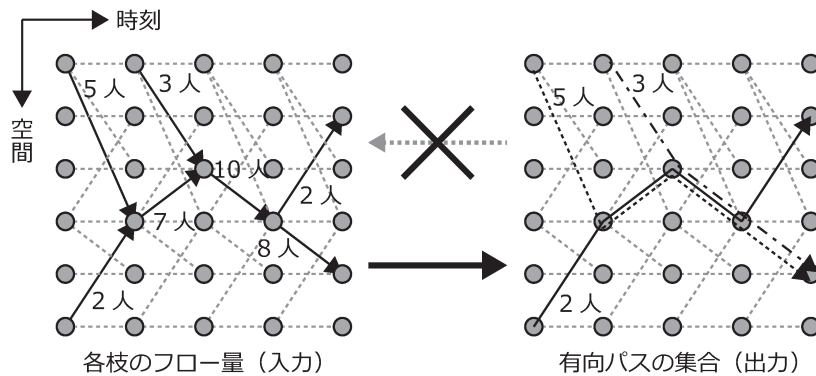


図 2: 時空間ネットワークによる移動体追跡問題の定式化

地点 $i \in N$, 時刻 $s \in T$ を表す時空間ネットワーク上を点 (i, s) とする. 地域外から点 (i, s) への流入者数 $e_{(i,s)}$, 点 (j, t) から地域外への流出者数 $l_{(j,t)}$, 点 (i, s) から点 (j, t) への移動者数 $f_{(i,s),(j,t)}$ が観測データとして与えられる. ただし, 時空間ネットワーク上では時刻 s と時刻 $s+1$ の点同士を結ぶ枝のみが存在する. 個人の移動履歴は時空間ネットワーク上での単純パスで表現され, 本論文ではこれを経路と呼ぶ. 可能な経路候補全体の集合を P , 始点 (i, s) と終点 (j, t) を繋ぐ経路候補の集合を $P_{(i,s),(j,t)}$ と表す. 各経路候補 $p \in P$ を利用する人数を変数 x_p で表すと時空間ネットワーク上の各枝のフロー量の推定値を計算できる. これと観測データから得られる各枝のフロー量の実測値との誤差を最小化する問題は以下の通りに定式化できる.

- $i, j \in N$: 地点
- $s, t \in T$: 時刻
- P : 可能な経路候補全体の集合
- $P_{(i,s),(j,t)}$: (i, s) を始点, (j, t) を終点とする経路候補全体の集合
- $e_{(i,s)}$: (i, s) における地域外からの流入者数
- $l_{(j,t)}$: (j, t) における地域外への流出者数
- $f_{(i,s),(j,t)}$: (i, s) から (j, t) への移動者数
- $a_{(i,s),(j,t)}^p$: 枝 $((i, s), (j, t))$ が経路 p に含まれていれば 1, そうでなければ 0

$$\begin{aligned}
& \text{minimize} && \sum_{(i,s) \in (N,T)} \sum_{(j,t) \in (N,T)} |y_{(i,s),(j,t)}| + \sum_{(i,s) \in (N,T)} |g_{(i,s)}| + \sum_{(j,t) \in (N,T)} |h_{(j,t)}| \\
& \text{subject to} && \sum_{(j,t) \in (N,T)} \sum_{p \in P_{(i,s),(j,t)}} x_p + g_{(i,s)} = e_{(i,s)}, \quad (i,s) \in (N,T), \\
& && \sum_{(i,s) \in (N,T)} \sum_{p \in P_{(i,s),(j,t)}} x_p + h_{(j,t)} = l_{(j,t)}, \quad (j,t) \in (N,T), \\
& && \sum_{p \in P} a_{(i,s),(j,t)}^p x_p + y_{(i,s),(j,t)} = f_{(i,s),(j,t)}, \quad (i,s), (j,t) \in (N,T), \\
& && x_p \in \mathbb{Z}_+, \quad p \in P.
\end{aligned} \tag{1}$$

3 経路候補集合の生成

前述の定式化には、(1) 可能な全ての経路候補 $p \in P$ に対して変数 x_p を用意するため問題の規模が大きくなり過ぎる、(2) 最適解が多数存在しておよそ人が取り得ないような経路を出力する可能性が高い問題点がある。そこで、東京大学空間情報科学センターが各都市圏のパーソントリップ調査をもとに作成した研究期間向けの人の流れデータ [11] を用いて、時空間ネットワーク上の始点と終点の各組に対して実際に利用された経路の種類を集計した。表 1 は平成 8 年 10 月 1 日の高知都市圏、平成 14 年 10 月 1 日の仙台都市圏、平成 13 年 10 月 1 日の中京都市圏において k 種類の経路が利用された始点と終点の組の数を表している。表 1 より、始点と終点のほとんどの組に対して実際に人が利用する経路は数種類しかないことが確認できる。

表 1: 始点と終点の各組に対して実際に利用された経路の種類

	総移動者数	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k \geq 7$
高知都市圏	24,001	21,834	1,733	333	84	14	2	1
仙台都市圏	84,606	79,130	4,335	873	195	51	18	4
中京都市圏	205,170	174,232	22,742	6,463	1,483	222	21	7

提案手法では、人の流れデータ [11] より実際に利用された経路の集合 $\bar{P} (\subseteq P)$ を取り出した上で、各経路 $p \in \bar{P}$ の出発時刻を前後にずらして得られる経路候補からなる集合 $\tilde{P} (\supset \bar{P})$ を生成する。可能な全ての経路 $p \in P$ ではなく生成した経路候補 $p \in \tilde{P}$ のみを考慮した定式化を考えることで、問題規模の縮小と人が取り得る経路の出力を実現する。さらに、観測データから得られる各枝のフロー量を用いることで経路候補を絞り込むことが可能である。時空間ネットワーク上においてフロー量が 0 の枝を通る経路候補が実際に利用されたとは考えにくい。そこで、(1) 観測データから得られるフロー量が正の枝を少なくとも 1 本以上含む経路候補の集合 $\tilde{P}' \subset \tilde{P}$ 、(2) 観測データから得られるフロー量が正の枝のみを含む経路候補の集合 $\tilde{P}'' \subset \tilde{P}$ のみを考慮する 2 通りの方法で経路候補の絞り込みを実現する。

4 数値実験

人の流れデータ [11] を用いて、平成 8 年 10 月 1 日の高知都市圏、平成 14 年 10 月 1 日の仙台都市圏、平成 13 年 10 月 1 日の中京都市圏の 6~9 時を対象に各人の移動履歴を推定する数値実験を実施した。各都市圏に対して単位領域を長さ 1km の正方形、単位時間を 10 分とする時空間ネットワークを生成し、6~9 時の観測データ (時空間ネットワークの各枝のフロー量) と同日の人の移動履歴 (0~24 時) から生成した経路候補の集合 \bar{P} を入力データとして与えた。(0) 人の移動履歴から生成した経路候補の集合 \tilde{P} 、(1) フロー量が正の枝を少なくとも 1 本以上含む経路候補の集合

\tilde{P}' , (2) フロー量が正の枝のみを含む経路候補の集合 \tilde{P}'' で定式化した問題例の概要を表 2 に示す. 総移動人数は 6~9 時の時間帯に実際に都市圏内を移動した延べ人数を表す. “#cst.” と “#var.” はそれぞれ定式化した際の制約式と変数の数を表す.

表 2: 数値実験に用いた問題例の概要

	総移動者数	地点数	定式化 (0)		定式化 (1)		定式化 (2)	
			#cst.	#var.	#cst.	#var.	#cst.	#var.
高知都市圏	13,181	1,405	217,097	405,912	98,435	117,827	32,112	30,447
仙台都市圏	31,521	2,729	632,235	1,391,251	297,454	397,500	91,483	90,949
中京都市圏	85,315	8,068	2,262,174	3,738,404	881,346	712,858	302,860	205,861

数値実験では定式化 (2) の各問題例に対して整数計画ソルバー CPLEX12.5.1 を適用した. Intel Xeon E5640(2.7GHz) の CPU と 32GB のメモリを搭載した PC 上で計算を実行したところ, いずれの問題例でも約 30 秒で目的関数値が 0 となる最適解が得られた. 提案手法の推定精度を以下に定義する再現性 R と適合性 P で評価した. これは, 情報検索の分野で用いられる評価指標で再現性は推定漏れの少なさを適合性は推定ノイズの少なさを表す [2].

$$R = \frac{\sum_{p \in P} \min\{\bar{x}_p, x_p\}}{\sum_{p \in P} \bar{x}_p} \quad (2)$$

$$P = \frac{\sum_{p \in P} \min\{\bar{x}_p, x_p\}}{\sum_{p \in P} x_p}. \quad (3)$$

ここで, $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{|P|})$ は各経路候補 $p \in P$ を実際に利用した人数を表す. 表 3 に定式化 (2) の各問題例に対する推定結果の評価を示す. いずれの問題例でも時空間ネットワーク上の大半の経路を正しく推定できたことが分かる. 全ての問題例で再現率と適合率の値が等しくなるのは得られた解 x が $\sum_{p \in P} \bar{x}_p = \sum_{p \in P} x_p$ を満たしているためである. 高知都市圏, 仙台都市圏, 中京都市圏の問題例において正の値を取る変数は 12,758 個, 31,006 個, 85,112 個であり, 1 人しか利用していない経路が非常に多いことが分かる. また, 他の時間帯に対する数値実験でも同程度の再現率, 適合率の解が得られることを確かめた.

表 3: 定式化 (2) の各問題例に対する推定結果の評価

	再現率 R	適合率 P
高知都市圏	0.859	0.859
仙台都市圏	0.899	0.899
中京都市圏	0.791	0.791

これまでの数値実験では時間帯や都市圏に関わらず高い精度で人の移動履歴を推定できることを確かめた. しかし, 実際に都市圏内の全ての地点にセンサを配置することは困難であり, 一部の地点のみセンサが配置されると考えるのが自然である. そこで, 図 3 に示すように交通量の多い地点のみセンサが配置された状況で人の移動履歴を推定した. 表 4 に仙台都市圏において交通量の多い地点のみセンサを配置された状況で人の移動履歴を推定した結果を示す. 定式化 (2) ではフロー量が正の枝のみを含む経路候補の集合 \tilde{P}'' を用いるため, 一部の地点のみセンサが配置された状況では推定に必要な経路候補が不足し推定精度が著しく低下する.

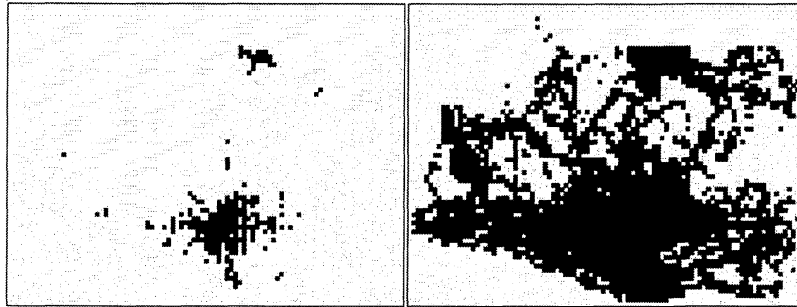


図 3: 仙台都市圏におけるセンサの配置例 (左:10%, 右:100%)

表 4: 交通量の多い地点のみセンサを配置された状況での推定結果 (仙台都市圏)

	100%	95%	90%	80%	60%	40%	20%	10%	5%	2%
再現率 R	0.899	0.484	0.440	0.394	0.322	0.281	0.230	0.194	0.173	0.108
適合率 P	0.899	0.480	0.444	0.420	0.362	0.330	0.309	0.312	0.342	0.295

5 おわりに

本論文では、地域内の各地点に配置されたセンサから観測される各時刻の交通量を入力データとし、人の移動経路の選択に関する適当なモデルの下で、時空間に拡張された大規模なネットワーク上における整数計画問題を解くことで、各人の移動履歴を推定する手法を提案した。高知都市圏、仙台都市圏、中京都市圏の人の流れデータを用いて数値実験を実施し、地域内の全ての地点にセンサが配置された状況では高い精度で人の移動履歴が推定できることを確かめた。一方で、地域内の一部の地点のみにセンサが配置された状況では推定に必要な経路候補が不足するため推定精度が著しく低下する問題点も明らかになった。

参考文献

- [1] M. C. González, C. A. Hidalgo and A. Barabási, Understanding individual human mobility patterns, *Nature*, **453** (2008), 779-782.
- [2] C. D. Manning and P. Raghavan and H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, 2008.
- [3] P. Nillius, J. Sullivan and S. Carlsson, Multi-target tracking-linking identities using Bayesian network inference, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, 2187-2194, 2006.
- [4] A. B. Poore and N. Rijavec, A Lagrangian relaxation algorithm for multidimensional assignment problems arising from multitarget tracking, *SIAM Journal on Optimization*, **3** (1993), 544-563.
- [5] A. B. Poore, Multidimensional assignment formulation of data association problems arising from multitarget and multisensor tracking, *Computational Optimization and Applications*, **3** (1994), 27-57.

- [6] K. Okuma, A. Taleghani, N. D. Freitas, J. J. Little and D. G. Lowe, A boosted particle filter: Multitarget detection and tracking, *Computer Vision-ECCV 2004*, 28-39, Springer, Berlin, 2004.
- [7] C. Ratti, D. Frenchman, R. M. Pulselli and S. Williams, MobileLandscapes: using location data from cell phones for urban analysis, *Environment and Planning B: Planning and Design*, **35** (2006), 727-748.
- [8] Y. Sekimoto, R. Shibasaki, H. Kanasugi and T. Usui, PFlow: Reconstructing people flow recycling large-scale social survey data, *IEEE Pervasive Computing*, **10** (2011), 27-35.
- [9] Y. Sekimoto, A. Watanabe, T. Nakamura, H. Kanasugi, T. Usui, Combination of spatio-temporal correction methods using traffic survey data for reconstruction of people flow, *Pervasive and Mobile Computing*, **9** (2013), 629-642.
- [10] J. Vermaak, S. J. Godsill and P. Pérez, Monte Carlo filtering for multi-target tracking and data association, *IEEE Transactions on Aerospace and Electronic Systems*, **41** (2005), 309-332.
- [11] 東京大学空間情報科学センター, 人の流れプロジェクト, <http://pflow.csis.u-tokyo.ac.jp/>