位置情報の活用と
プライバシー保護

統計数理研究所
南　和宏

統計数理研究所 公開講演会
「空から眺める未来都市：空間ビッグデータと統計数理」
2017年11月7日

講演内容

- 位置情報の利活用とプライバシー保護
  - 活用事例
  - プライバシーとの関連
- 位置情報の匿名化の課題
  - 超多次元データ
  - 時空間の相関性
- 位置情報の匿名化手法
  - 動的仮名交換による移動軌跡の分割
  - 隠れマルコフモデルによる安全性評価

位置情報の活用と
プライバシー保護

位置情報の様々な利活用

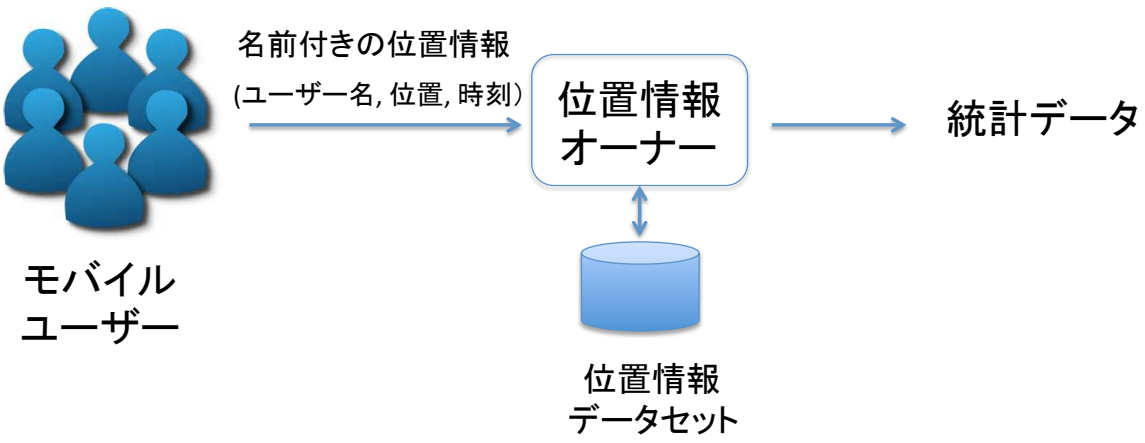
- 交通状況のリアルタイムのモニタリング
- 動的人口分布
- 商圏分析
- 災害リスクアセスメント

位置情報の形態

- GPS位置座標
- 携帯基地局
- 無線LAN アクセスポイント
- ポイントカードによる購買（店舗情報）
- 乗車履歴

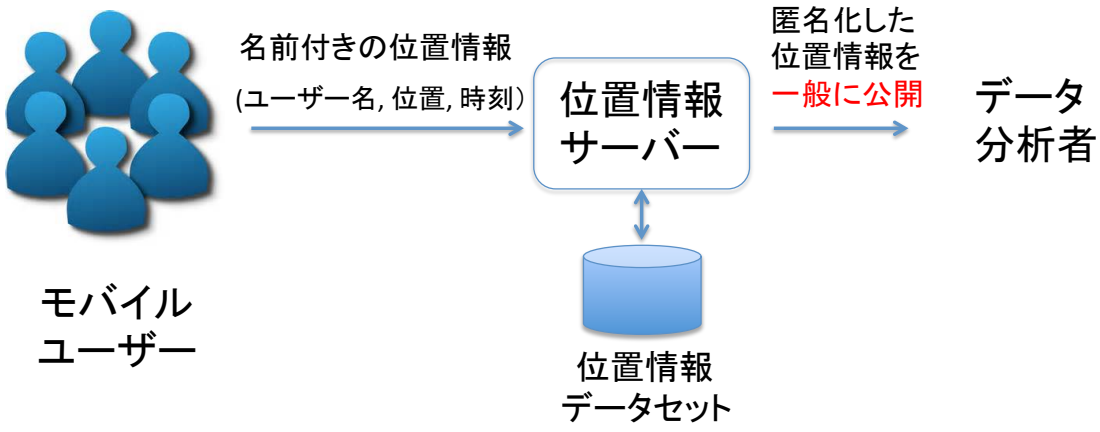
現在の位置情報の利活用

- 位置情報は一部の企業に囲い込まれていて、統計情報のみ提供



匿名化データ提供による2次利用

- 位置情報には様々な分析ニーズが存在
- 個人情報を削除した匿名データで第3者に提供



位置情報には
プライバシー漏洩の懸念

病院 ⇨ 病気

カフェ ⇨ 休憩のとり過ぎ
秘密の会議

位置情報の匿名化の課題

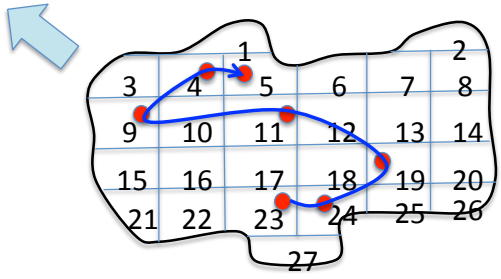
位置情報軌跡データの例

- （位置情報ID, タイムスタンプ)のシーケンス

2016年5月15日

氏名	8:00	8:30	9:00	9:30	10:00	10:30	11:00	
伊藤	1	5	4	8	12	15	9	
佐藤	10	15	24	14	21	20	19	・・・・・・・・・
鈴木	3	8	6	6	7	10	15	
高橋	23	24	19	11	9	4	5	

GPS座標
から変換



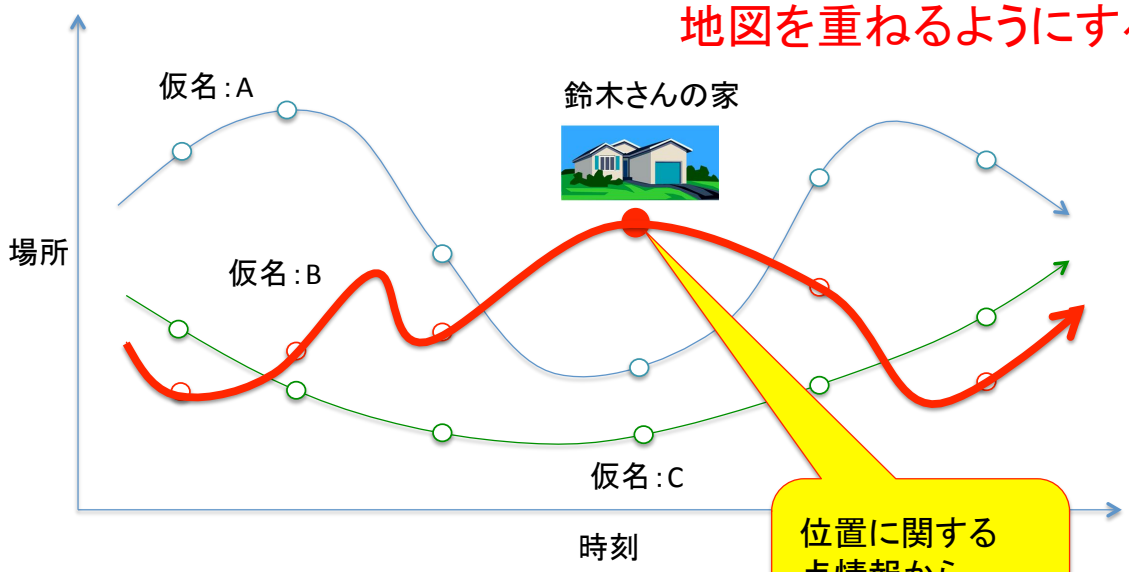
仮名化で十分か？

2016年5月15日

仮名	8:00	8:30	9:00	9:30	10:00	10:30	11:00	
A	1	5	4	8	12	15	9	
B	10	15	24	14	21	20	19	・・・・・・・・・・
C	3	8	6	6	7	10	15	
D	23	24	19	11	9	4	5	

位置情報に関する（断片的な）外部知識を得るのは比較的簡単

地図を重ねるようにする



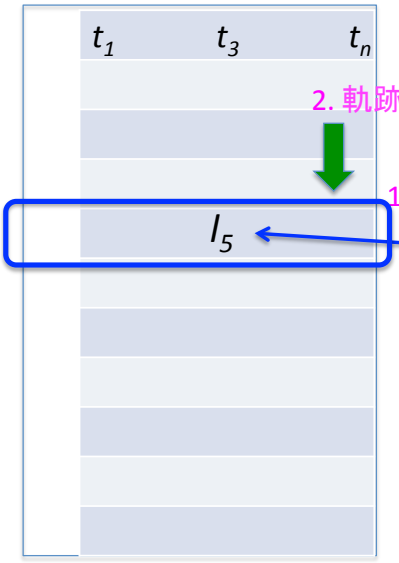
- 自宅、オフィスの住所
- 物理的な目撃

仮名化のリスクに関する従来研究

- 65人のドライバーのGPSデータから、85%の住所を特定 [Hoh06]
- 午前3時以前の最後の位置の平均を住所と推定（誤差60m) [Krumm07]

位置情報のk-匿名化

匿名化した位置情報軌跡



2. 軌跡全体の情報を入手

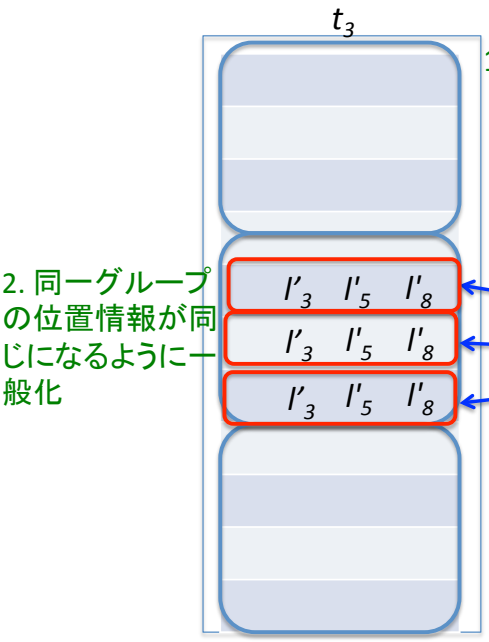
1. レコードのリンケージ

標的ユーザーに関する部分的知識

	t_1	t_2	t_3	t_4
Bob				
Tom			l_5	
Ken	l_7			

位置情報のk-匿名化

匿名化した位置情報軌跡

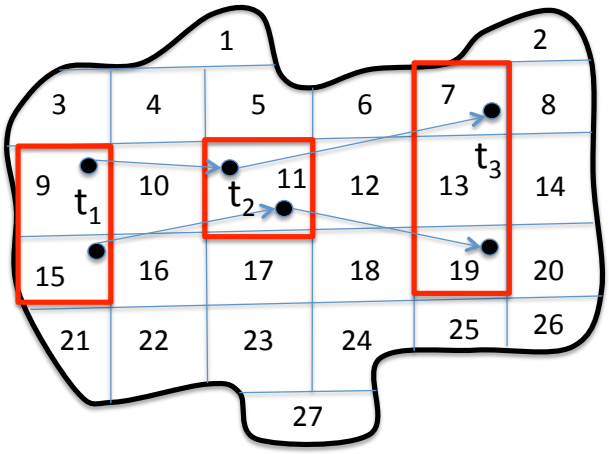


標的ユーザーに関する部分的知識

	t_1	t_2	t_3	t_4
Bob				
Tom			l_5	
Ken	l_7			

2-匿名化の例

- k人が入るように領域を拡大



元データ

ユーザー	時刻 t_1	時刻 t_2	時刻 t_3
田中	{9}	{11}	{7}
鈴木	{15}	{11}	{19}

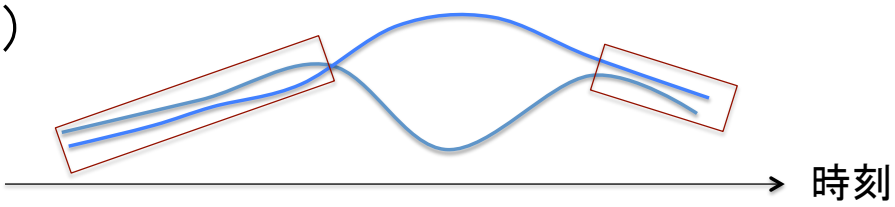
一般化処理

2匿名化データ

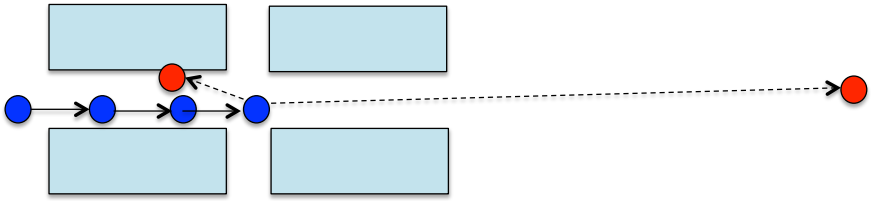
仮名	時刻 t_1	時刻 t_2	時刻 t_3
A	{9, 15}	{11}	{7, 13, 19}
B	{9, 15}	{11}	{7, 13, 19}

位置情報軌跡の匿名化の課題

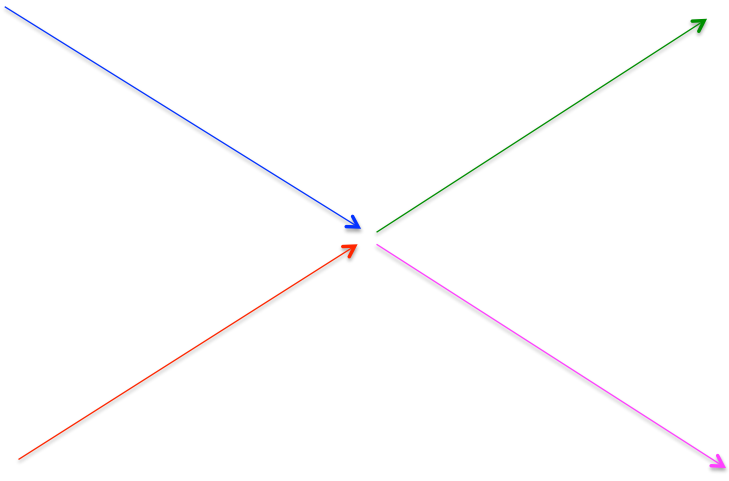
1. 軌跡が長くなるに従い、類似する複数の軌跡を見つけることは困難になる（次元の呪い）



2. アルゴリズム知識、地図情報、時空間の相関性による元軌跡の復元攻撃への対策

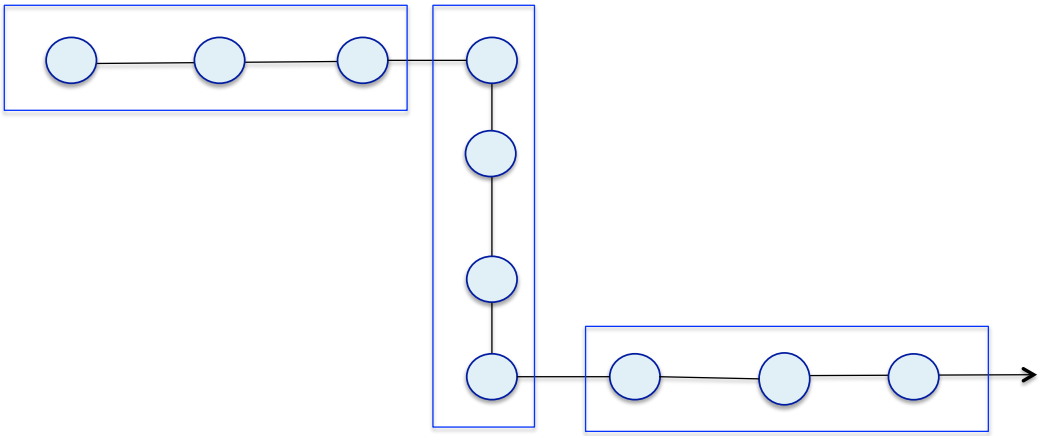


動的仮名交換による位置情報軌跡の分割



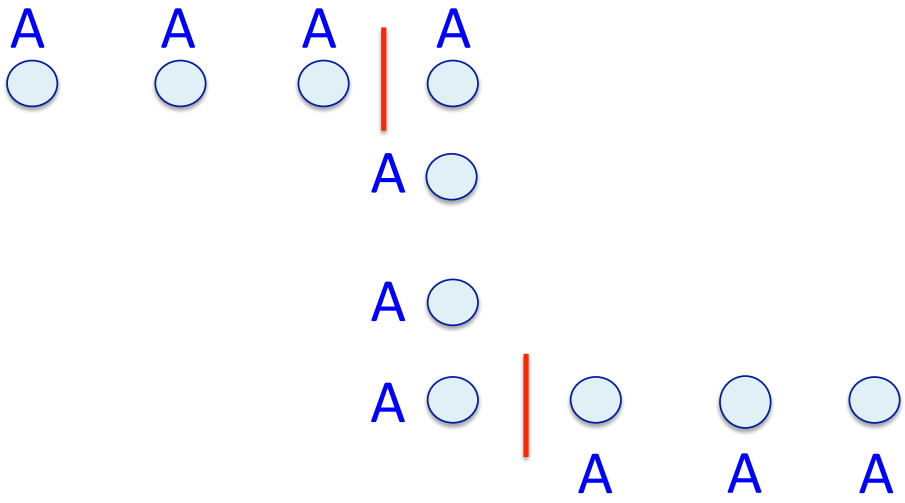
軌跡分割のアイデア

- 軌跡を分割すれば、類似する他の軌跡とグループ化しやすい



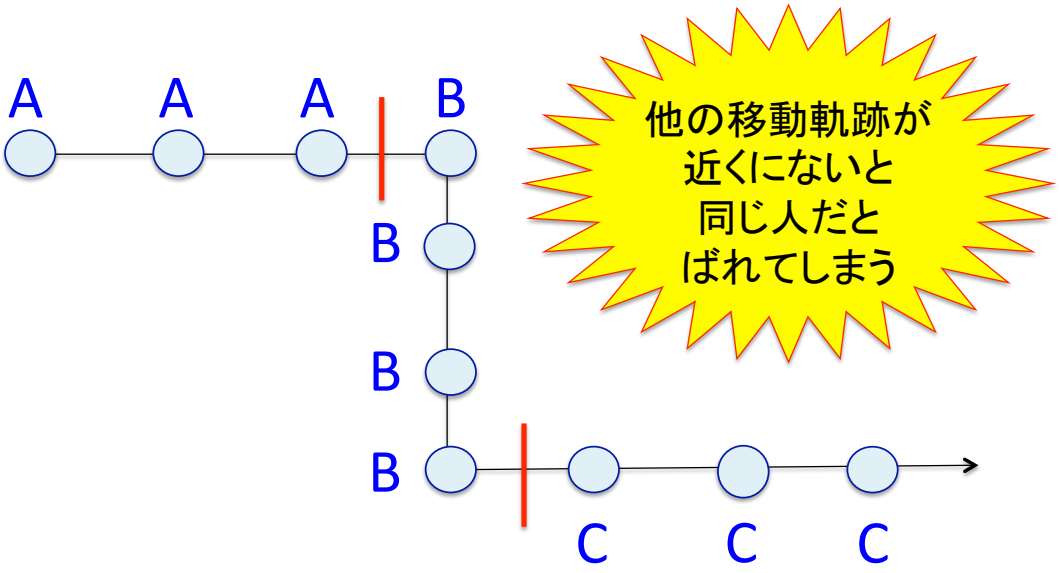
軌跡分割のアイデア

- 軌跡の分割は、仮名に置換に相当する

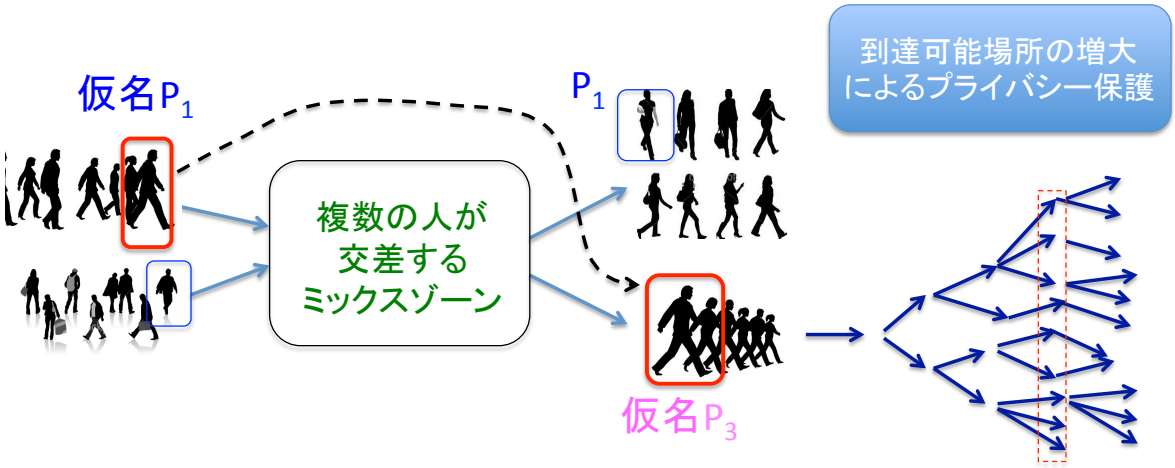


軌跡分割のアイデア

- 軌跡の分割は、仮名に置換に相当する



ミックスゾーンにおける仮名交換

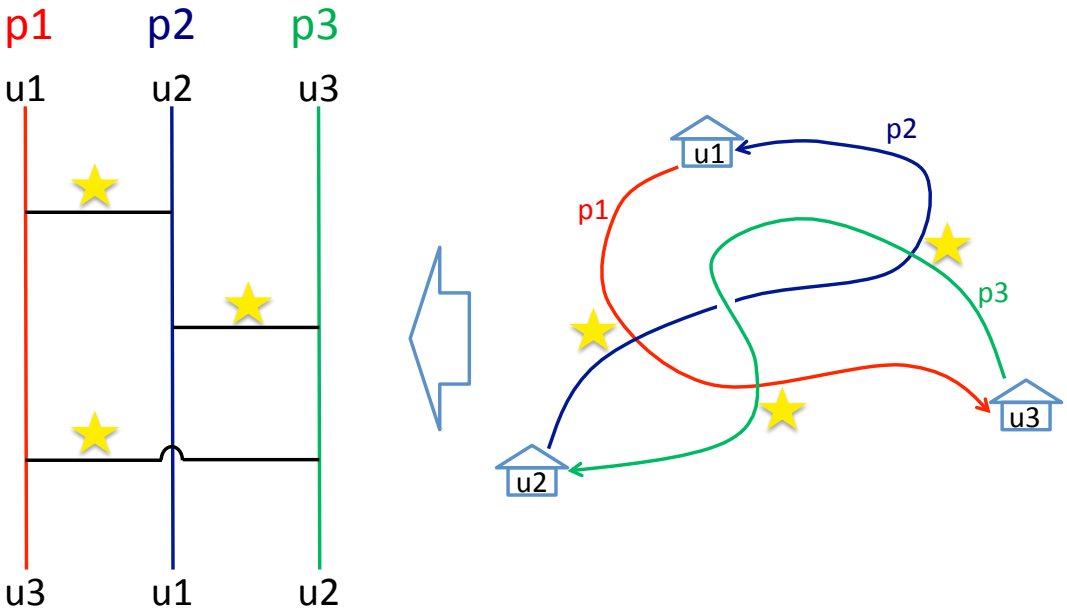


外部知識（自宅情報）を利用されると可能経路は大幅に減ってしまう

つまり同じ場所に戻ってくる
経路しか可能経路であり得ない

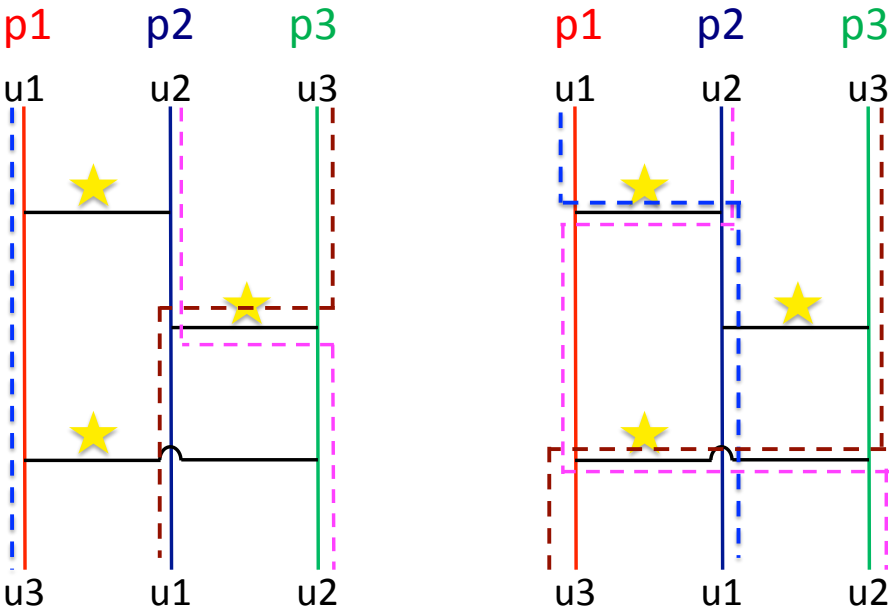
大抵の人は自宅を朝出発し、夜帰宅する。

安全性はアミダモデルで評価

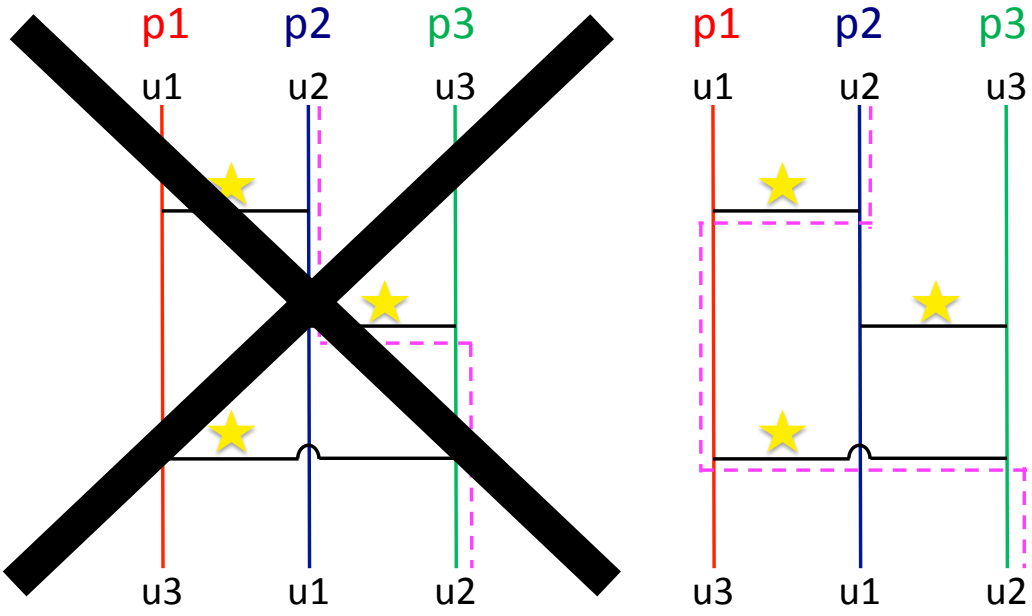


“横棒が選択できるアミダ”

Q: u2 の可能な経路は？



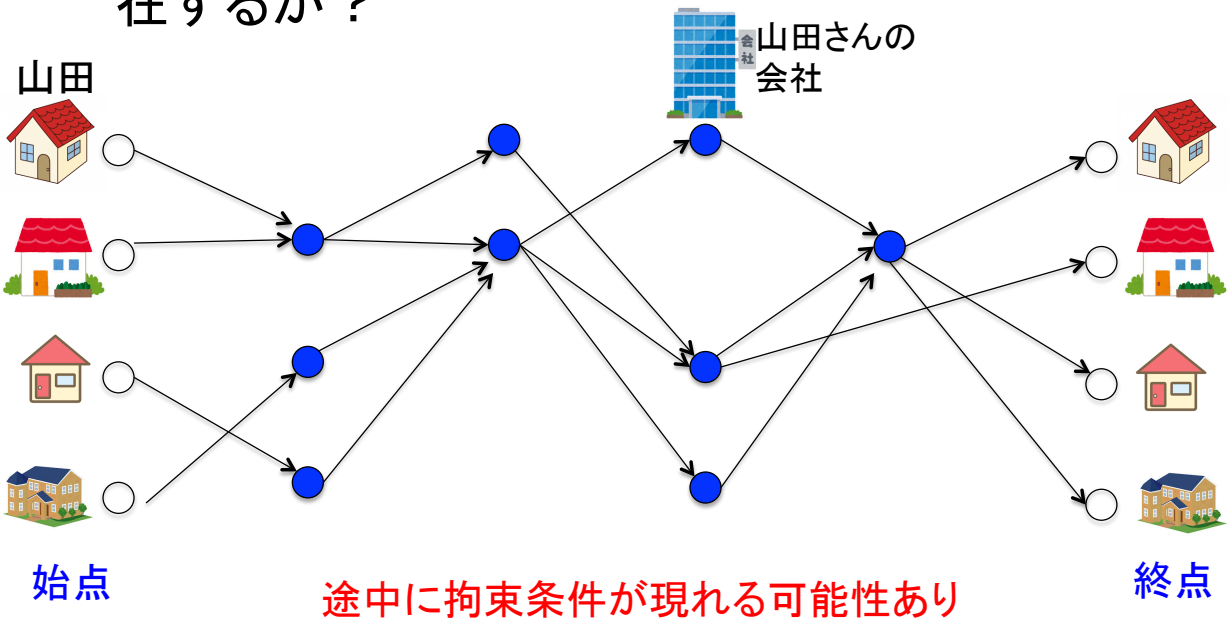
Q: u2 の可能な経路は？



排他的辺素パスを数え上げる必要がある

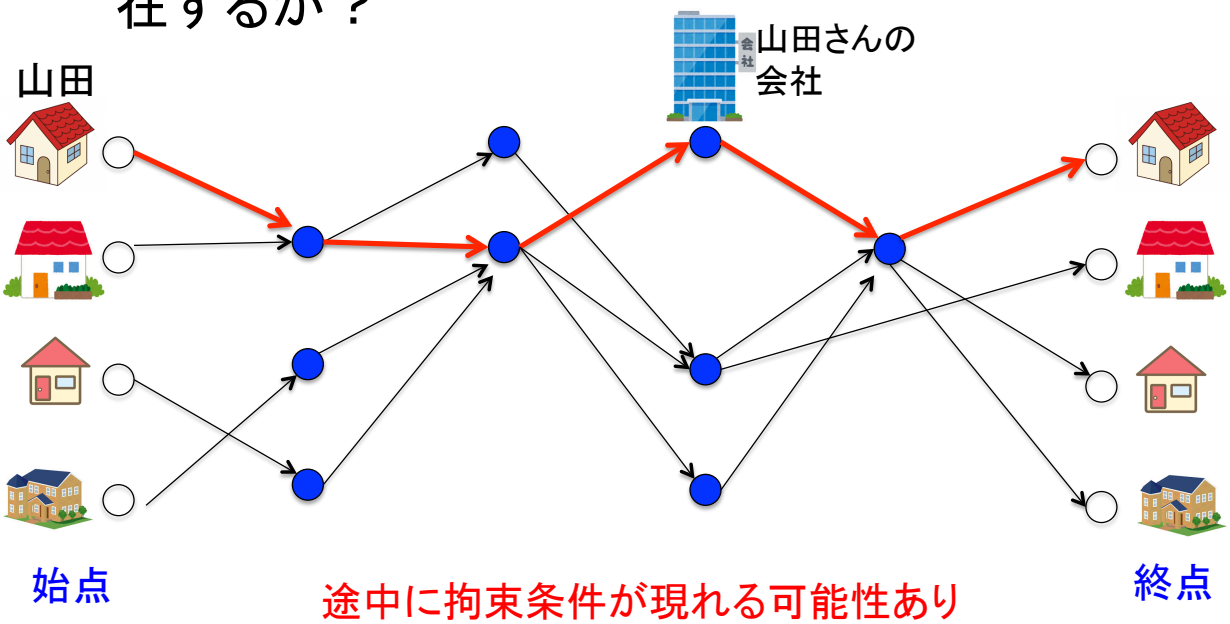
グラフ理論の排他的辺素パス問題

- 辺を共有しない*n*組の排他的パスがいくつ存在するか？



グラフ理論の排他的辺素パス問題

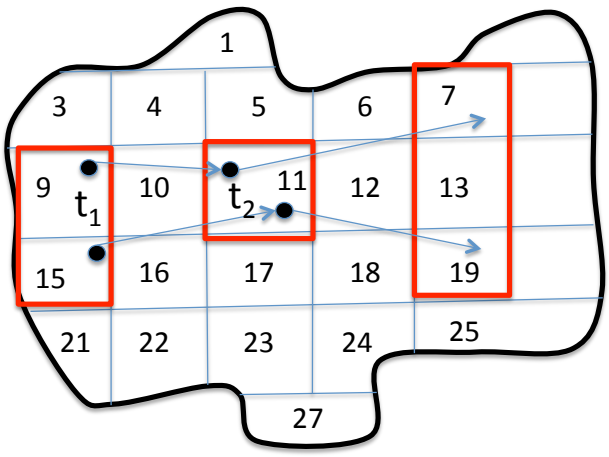
- 辺を共有しない*n*組の排他的パスがいくつ存在するか？



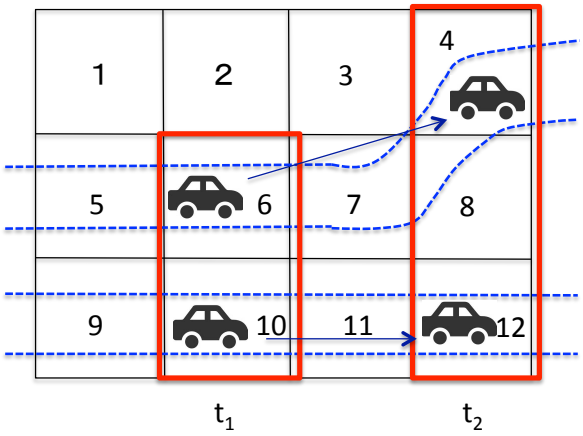
統計モデルに基づく
統計データの安全性評価

位置情報の*k*-匿名化

1. k個の類似する軌跡をグループ化
2. 各時刻ごとにk個の位置情報を囲む最小の矩形に位置情報を一般化



道路情報を用いた推論



元データ

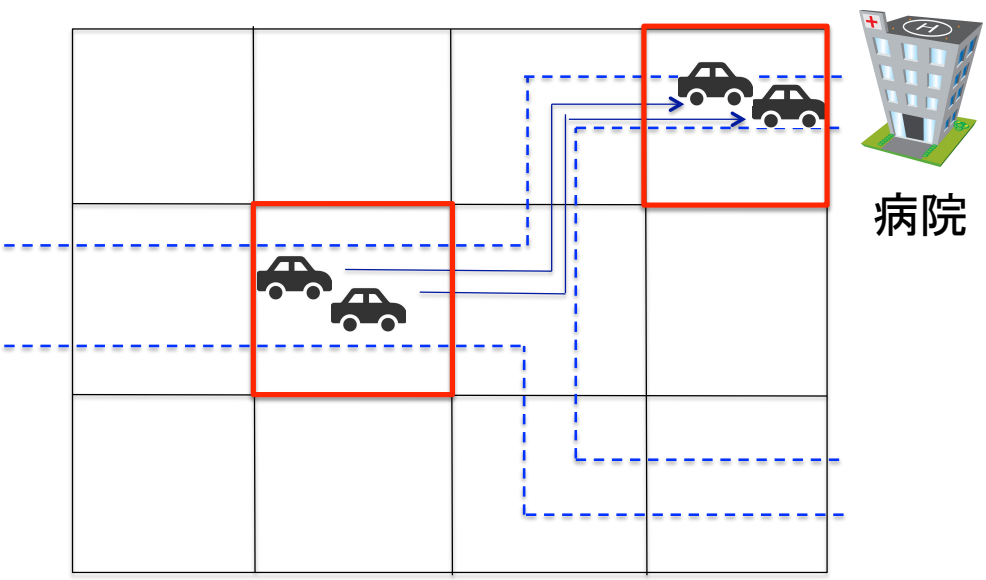
ユーザー	時刻 t_1	時刻 t_2
田中	6	4
鈴木	10	12

一般化処理

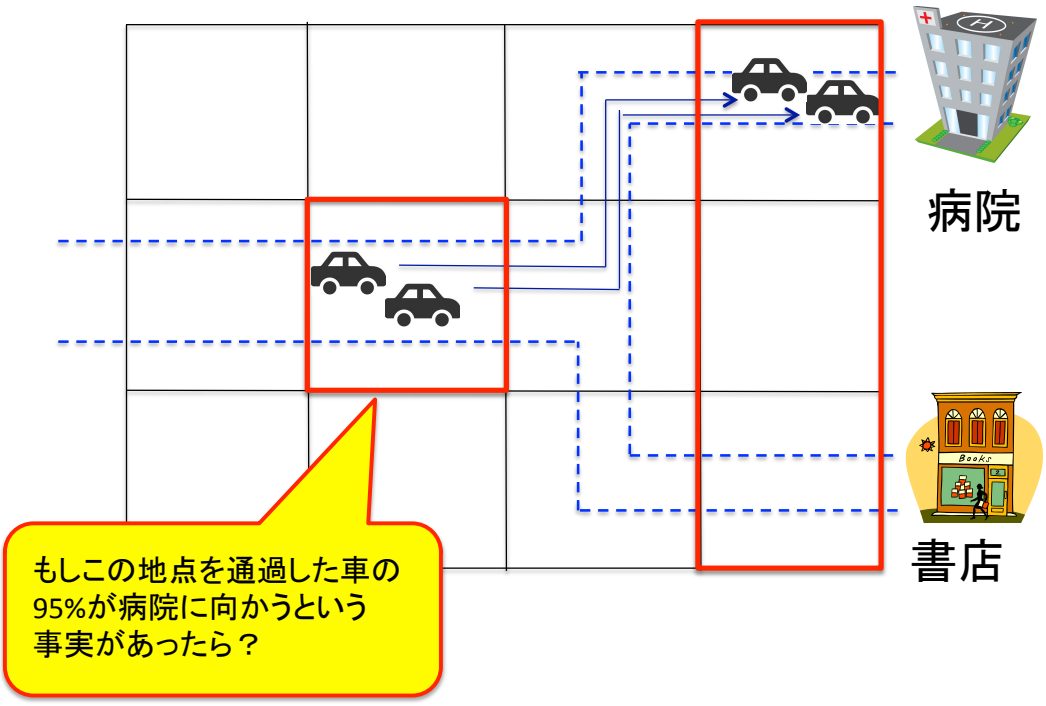
2匿名化データ

仮名	時刻 t_1	時刻 t_2
A	{6, 10}	{4, 8, 12}
B	{6, 10}	{4, 8, 12}

道路が分岐していれば大丈夫か？



病院の訪問を隠すためには領域を広げる必要がある



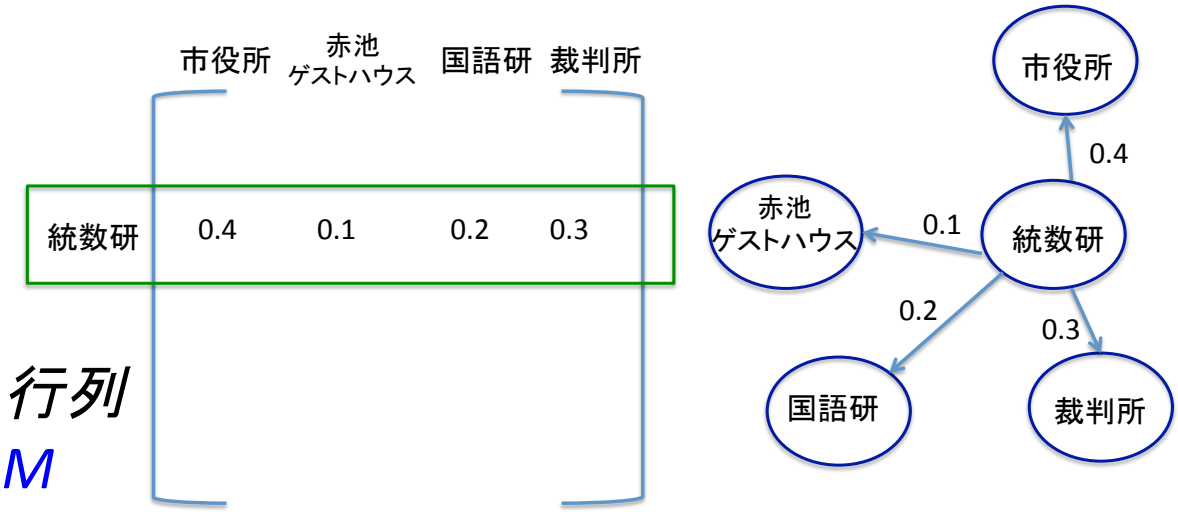
「位置A➔Bへの移動」の推論が成立すると匿名化が破綻

- 地図情報、移動パターンの知識等さまざまな外部知識が存在
  - ルールの記述による個別対応は困難

過去の移動履歴を学習して、様々な移動パターンを統一的に表現するマルコフチェーンで表現

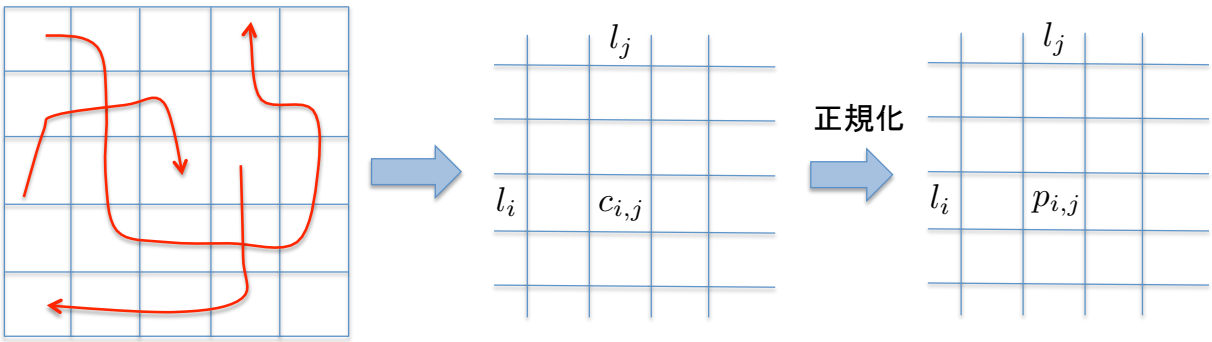
マルコフ連鎖を用いて予測する攻撃者を想定

- 外部知識：ユーザーの過去の全ての移動履歴
- 各位置を状態とするマルコフモデルを構築



マルコフチェーンの生成

- 過去の移動軌跡を学習データとして使用



過去の移動軌跡

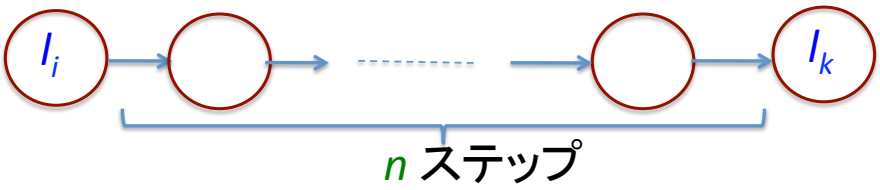
状態遷移の回数を数える

Markov chain
P^u

行列Mで推論攻撃 (l_i, l_k, n)のリスク評価

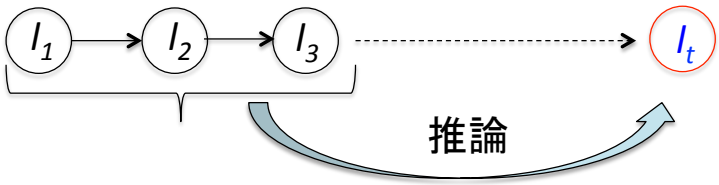
- 現在位置 l_i からnステップ後にl_k に移動する確率は閾値 t 未満であるべき

$$M_{i,k}^{(n)} < t$$

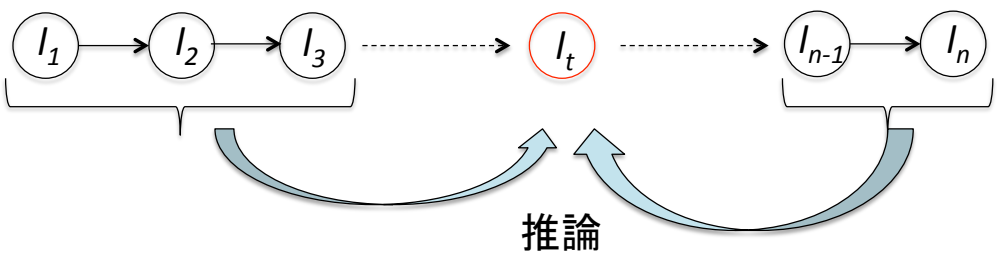


過去を振り返って推論する可能性もある

- 未来の移動を推論

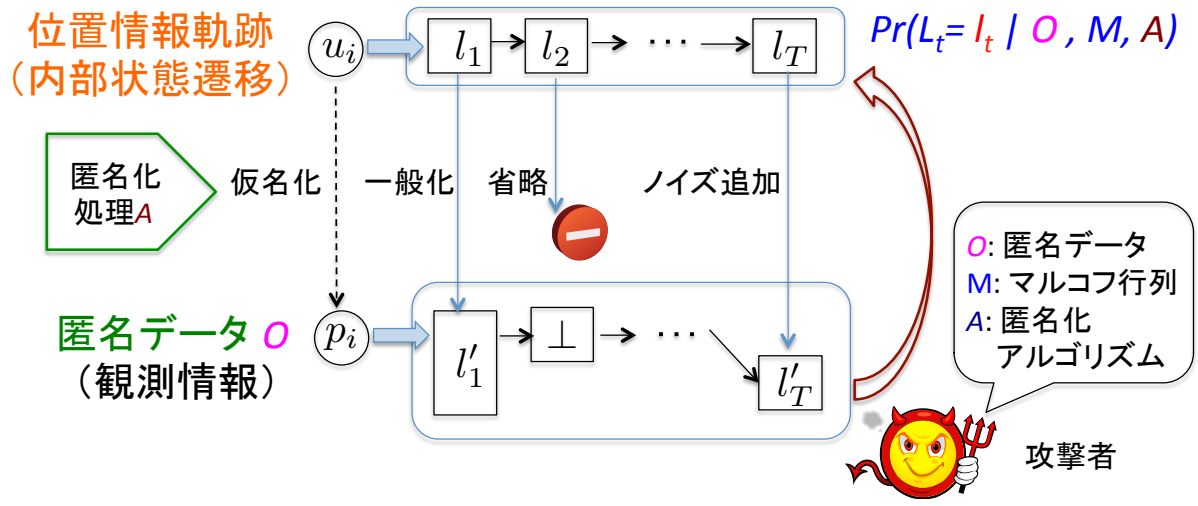


- 過去の歯抜けの部分を推論



隠れマルコフモデルにおける内部状態の推論問題

- 安全性は、マルコフチェーン*M*, 匿名化アルゴリズム*A*, 観測情報*O*を与えられて内部状態 l_t を推測する条件付き確率として定式化



まとめ

- 位置情報活用のポテンシャルは高く、匿名化データでの流通が望まれる
- ただし、位置情報の匿名化には多次元性、時空間の相関性の2つの課題があり、軌跡の分割と状態空間モデルに基づく軌跡セグメントの匿名化が必要
- 匿名化データの安全性は相対的なものであり、プライバシー保護とデータ効用のバランスが重要