統計数理研究所 75 周年記念事業（2019）

位置情報軌跡の匿名化技術
Anonymization Techniques for Trajectory Location Datasets

モデリング研究系 南 和宏（Kazuhiro Minami）

1. 位置情報の匿名化における課題

スマートフォンの普及に伴い，我々の位置情報の取得が容易になり，多くのユーザーの移動履歴は，交通情報の提供，都市設計といった社会サービス，また商圏分析等の商用ビジネスにも活用されている．その一方，位置情報から，個人の興味に関するプライバシーな情報が漏洩する危険性が懸念される．よって位置情報の安全な２次利用には，匿名化と呼ばれる個人の識別情報を取り除くデータ処理が不可欠である．一般に，位置情報の匿名化処理では，氏名等の個人の識別子を削除するだけでは不十分である．なぜなら目撃情報，名簿等の外部知識から特定の日時，場所の位置情報が識別され，その結果，その位置情報を含む軌跡全体が特定されるリスクが存在するからである．したがって，位置情報から k 未満のユーザーへの絞り込みを防ぐための k-匿名化処理（Sweeney, 2002）が必要となる．

しかし通常の k-匿名化の手法を位置情報軌跡に適用する場合，２つの課題が存在する．１つは，位置情報軌跡のような時系列データの場合，k-匿名化を実施するとデータの有用性が著しく劣化する問題である．長期の位置情報軌跡を匿名化する場合，k-匿名化の前提となる軌跡群へのグループ化が困難である．そのため，k-匿名化を実現するための一般化処理による情報損失は大きくなり，有益なデータ分析に堪えなくなる．２つめは，位置情報軌跡のデータ間に時空間の相関性が存在し，匿名化した位置情報から統計的推論により元の軌跡情報が復元される問題である．位置情報軌跡には，人の移動に関する物理的制約が反映し，車，電車といった交通手段により移動経路は限定される．また長期的な移動軌跡には通勤，病院への通院といった個人の生活習慣を反映した特徴的なパターンが現れる．そのような移動パターンに関する外部知識を用いると匿名化された位置情報から元の位置情報が復元される危険性がある．

近年，著者はこれらの課題を解決するための２つの匿名化技術に取り組んできた．１つは位置情報軌跡を複数のセグメントに分割する動的仮名交換手法（Tanjo et al., 2014）であり，ミックスゾーンと呼ばれる複数ユーザーの集積点でのランダムな仮名の再割当により移動先の不確定性を確保する手法である．もう１つは，ユーザーの移動パターンをマルコフ過程でモデル化し，隠れマルコフモデルにおける内部状態の推定問題として匿名化データの安全性の評価を行う手法（Minami, 2014）である．

2. ミックスゾーンにおける動的仮名割当

個人の行動パターンが顕著に現れる位置情報軌跡の場合，その中の幾つかの点に過ぎない外部知識を用いて軌跡全体の識別が可能であり，情報漏えいリスクが非常に高い．位置情報軌跡の開示リスクを局所するため，Mano et al.（2013）では位置情報軌跡に紐付けられる仮名を動的に更新し長期間の軌跡データを複数の軌跡セグメントに分割する方式を提案した．この仮名の更新は複数のユーザーが同一の時間，場所に存在する「ミックスゾーン」と呼ばれる領域でラ

ンダムな仮名交換の形式で実施し，ミックスゾーン前後の軌跡セグメント間の関連性を分断する．個人の位置情報はミックスゾーンを経由することで代替経路が増大するので，その不確定性に着目して位置情報軌跡の仮名化データに対するプライバシー指標を定式化した．ただし，攻撃者の外部知識と整合性を保持する代替経路の列挙には，ミックスゾーンを頂点，位置情報の軌跡セグメントを辺とするグラフにおける排他的辺素パス問題を解く必要がある．一般の排他的辺素パス問題は NP 困難であるため，(Tanjo et al., 2014)では排他的辺素パス問題を制約充足問題に変換する効率的な安全性評価手法を開発した．

3. 隠れマルコフモデルに基づく匿名化処理の安全性評価

仮名更新による位置情報軌跡の分割は位置情報軌跡の識別リスクを局所化する手法である．しかし位置情報が識別されると同じ軌跡セグメント内の位置情報は依然として漏洩してしまう．したがって分割して次元を削減した軌跡セグメント単位に k-匿名化を実施することが望ましい．ただし，位置情報軌跡には時空間の相関性が存在するので，通常の k-匿名化では不十分な場合が多い．

Minami (2014)では，ユーザーの移動パターンをマルコフ過程でモデル化し，匿名化データの安全性を隠れマルコフモデルにおける観測情報から内部状態の推定問題として定式化した．モデルにおける観測情報は匿名化データ，内部状態遷移は秘匿すべき元の位置情報に対応し，匿名化アルゴリズムは，内部状態から観測情報への確率的な変換を定義する記号出力行列として記述される．さらに匿名化データの安全性は，観測情報，記号出力行列，内部状態のマルコフ過程が与えられたときに正しく内部状態を推定する条件付き確率として定式化した．実データを用いた評価実験では，通常の k-匿名化処理では想定した安全性が確保できず，追加の秘匿処理の必要性を明らかにし，統計モデルに基づく安全性評価の有用性を実証的に示すことができた．

参 考 文 献

Mano, K., Minami, K. and Maruyama, H. (2013). Protecting Location Privacy with K-Confusing Paths Based on Dynamic Pseudonyms, *5th IEEE International Workshop on SEcurity and SOCial Networking*, March.

Minami, K. (2014). Preventing denial-of-request inference attacks in location-sharing services, *2014 Seventh International Conference on Mobile Computing and Ubiquitous Networking*, 50–55, January.

Sweeney, L. (2002). k-anonymity: a model for protecting privacy, *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, **10** (5), 557–570.

Tanjo, T., Minami, K., Mano, K. and Maruyama, H. (2014). Evaluating data utility of privacy-preserving pseudonymized location datasets, *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, **5** (3), 63–78, September.