

大規模移動履歴の特徴と匿名加工方法

疋田敏朗 * †

Toshiro HIKITA

1 はじめに

近年、GPSを始めとする測位デバイスが携帯機器の機能の一部として搭載されるようになり、さらに携帯網の発展により携帯機器が測位した位置情報をセンター上のサーバに送信し、蓄積することで位置履歴情報を生成することが現実的になっている。

本稿では位置情報独特の課題について説明をした上で、利用履歴や乗降履歴の長さが長くなるにつれ、高くなるという履歴特有の本人特定リスクを検討し、安全な移動履歴について検討を行う。さらに大規模な移動履歴である「人の流れプロジェクト」の移動履歴から乗降履歴を生成し、移動履歴の80%以上が二点間の往復移動であること、また二点間の移動を除くほぼすべての履歴が一意であることを示す。

なお、本研究は東京大学空間情報科学研究センターの「人の流れプロジェクト」との共同研究である。

2 大規模移動履歴を利用したり リスク検討

「履歴が固有のために個人が特定される場合」について検討をすることとする。そこで実際の移動履歴を元にどの程度の履歴長であれば、個人が特定をされるのかについて実際のデータを元に検討することとする。

今回は東京大学空間情報科学研究センターの「人の流れプロジェクト」[1]の「2008年東京都市圏 人の流れデータセット（空間配分版）」を利用することとした。この実験データは東京都市圏交通計画協議会（東京地区）が収集したパーソントリップ調査によるデータを元になっている。

鉄道を利用した移動について1日のデータを履歴長とその数で整理すると表1に示す。

表 1: 移動履歴数と履歴長 (東京地区)

履歴長	履歴数	一意履歴数	個人特定率
1	8000	5679	71.0
2	154997	64837	41.8
3	17396	16766	96.4
4	6257	6185	98.8
5	1132	1132	100
6	370	370	100
7	79	79	100
8	37	37	100
9	5	5	100
10	1	1	100
11	1	1	100

この場合に一意である履歴は95092件であり全履歴の50.5%が他の履歴と重複しない一意のものであることがわかる。

すなわち乗車カードやIoTによる移動履歴単体をそのまま利用した場合は時刻情報を削ったとしても日単位で50%の移動履歴は単一であって、他の履歴と区分されるため、ユーザを外部観察できていれば個人を特定することができることになる。

また2区間の履歴を除くと個人特定率は高まる3区間以上に限定すれば特定率は96%以上となる。

この特性を踏まえて移動履歴の匿名加工方法について論じる

参考文献

- [1] Y. Sekimoto, R. Shibasaki, H. Kanasugi, T. Usui and Y. Shimazaki: "Pflow: Reconstructing people flow recycling large-scale social survey data", IEEE Pervasive Computing, **10**, 4, pp. 0027–35 (2011).

*東京大学大学院情報理工学系研究科 電子情報学専攻
toshi@yamagula.ic.i.u-tokyo.ac.jp

†トヨタ IT 開発センター hikita@jp.toyota-itc.com