社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

# ［招待講演］複数用途からなる交通ICカードデータの再識別リスク分析 (from AINA 2018)

伊藤聡志† 原田玲央† 菊池浩明††

† 明治大学大学院, 〒164–8525 東京都中野区中野 4-21-1
†† 明治大学, 〒164–8525 東京都中野区中野 4-21-1
E-mail: †{mmhm,kikn}@meiji.ac.jp

**あらまし**　匿名加工とはデータセットから個人が特定されないように加工する技術であり，匿名加工されたデータから個人を特定しようと試みる攻撃を再識別という．多くの匿名加工の研究は，購買履歴や乗降履歴など，単一用途のデータに注目したものである．しかしながら，攻撃者は複数用途のデータを用いて個人の再識別を試みることが想定されるため，複数用途のデータを組合せたときのリスクを調査する必要がある．我々は交通 IC カードデータが複数の用途からなっていることに注目し，各用途の危険度やそれらを組み合わせたときの再識別リスクをエントロピーを用いて評価する．

**キーワード**　匿名加工，再識別，プライバシーリスク評価

# [Invited Talk] Risk of Re-identification from Payment Card Histories in Multiple Domains (from AINA 2018)

Satoshi ITO†, Reo HARADA†, and Hiroaki KIKUCHI††

† Meiji University Graduate School, 4–21–1 Nakano, Nakano-ku, Tokyo, 164–8525 Japan
†† Meiji University, 4–21–1 Nakano, Nakano-ku, Tokyo, 164–8525 Japan
E-mail: †{mmhm,kikn}@meiji.ac.jp

**Abstract**　De-identification is the process of modifying a data set to prevent the identification of individual people from the data. However, most studies consider only the De-identification of data from a single domain. No study has been made on the risk of re-identification from combined data sets involving more than one domain. This paper proposes an evaluation of the risk of re-identification from payment card histories in multiple domains by entropy.

**Key words**　De-identification, Re-identification, Privacy Risk Evaluation

## 1. 概　　要

匿名加工とはデータセットから個人が特定されないように加工する技術であり，匿名加工されたデータから個人を特定しようと試みる攻撃を再識別という．2015 年の個人情報保護法改正により「匿名加工情報」が新たに定義され，それに伴い 2015 年から匿名加工・再識別コンテスト PWSCUP [1] が開催されている．

企業は購買履歴データのようなビッグデータを利活用する際に，そのデータのプライバシーリスク評価と匿名加工を行ったのち，匿名加工されたデータを総合的に評価する必要がある．匿名加工されたデータの評価は有用性と安全性の面からされることが多く，匿名加工データを評価する指標が数多く提案されている．[2] [3] [4] [5]

しかしながら，ほとんどの研究が単一用途のみのデータに注目をしている．例えば [4] や [5] では移動履歴データの，[6] では購買履歴データの再識別リスク評価の研究が行われているが，2 つ以上の用途を組み合わせたデータの再識別リスク評価は多くない．その理由の一つとして，個人情報データを他のデータと組み合わせる同意が得られないことや，公開データが組み合わせることを制限されていることなどがあげられる．さらに，全く異なる特徴を持つデータセットを組み合わせたデータのための数学的モデルを定式化することも容易ではない．

複数履歴からなるデータセットの再識別リスクを明らかにするために，移動履歴データと購買履歴データの相関を考える必要がある．おそらくこれらの履歴データは完全に独立ではなく，何かしらの相関があることが予想される．しかしながら，2 データ間の相関を数学的にモデル化する最良の数学的性質はい

まだ知られていないため，我々は複数用途の履歴（移動履歴と購買履歴）が組み合わさったデータの再識別リスクを評価する手法を提案する．我々の手法は購買履歴データが与えられた時の移動履歴データの再識別リスクを定量化し，結合されたデータの匿名性レベルを効率的に測ることができる．

本研究では交通 IC カード Suica [7] の履歴データを実験に用いる．Suica には鉄道の乗降履歴だけでなく，購買履歴やチャージ履歴等のデータも保存されており，我々はこのデータから，同一顧客の移動履歴データと購買履歴データの両方を得ることができる．我々は 2 つの異なる用途の履歴間の相関をエントロピーを用いて求め，相互情報量によって顧客の再識別リスクを定量化する．

我々は 31 人の顧客から実際の交通 IC カードデータを収集し，そのデータを用いて提案手法の評価を行った．データの読み取りには，Android のアプリケーション「IC カードリーダー by マネーフォ ワード」[8] を用いた．結合されたデータのリスクを測り，移動履歴と購買履歴の間の相関を評価する実験を行った結果，交通 IC カード内の履歴データには相関があることと，複数用途のデータを組み合わせることにより再識別リスクは大きく上がることが判明した．

我々の主な貢献は以下の通りである．

（1） 我々は複数用途の履歴を含むデータの再識別リスク評価のための，エントロピーを用いた指標を提案した．

（2） 我々は 31 人の顧客から収集した交通 IC カードデータのリスク評価実験を行った．

（3） 我々は匿名加工データの有用性と安全性を評価するプラットフォームを開発し，評価結果を報告した．

## 文　献


[1] H. Kikuchi, T. Yamaguchi, K. Hamada, Y. Yamaoka, H. Oguri and J. Sakuma, “What is the Best Anonymization Method? - a Study from the Data Anonymization Competition Pwscup 2015”, Data Privacy Management Security Assurance (DPM2016), LNCS 9963, pp. 230 - 237, 2016.

[2] Josep Domingo-Ferrer, Sara Ricci and Jordi Soria-Comas, “Disclosure Risk Assessment via Record Linkage by a Maximum-Knowledge Attacker”, 2015 Thirteenth Annual Conference on Privacy, Security and Trust (PST), *IEEE*, 2015.

[3] Koot, M. R., Mandjes, M., van’t Noordende, G., and de Laat, C., “Efficient probabilistic estimation of quasi-identifier uniqueness”, In Proceedings of ICT OPEN 2011, 14-15, pp. 119-126, 2011.

[4] A Monreale, R Trasarti, D Pedreschi, C Renso and V Bogorny, “*C*-safety: a framework for the anonymization of semantic trajectories”, Transactions on Data Privacy, Vol. 4 (2), pp. 73-101, 2011.

[5] A. Basu, A. Monreale, R. Trasarti, J. C. Corena, F. Giannotti, D. Pedreschi, S. Kiyomoto, Y. Miyake and T. Yanagihara, “A risk model for privacy in trajectory data”, Journal of Trust Management, 2:9, 2015.

[6] Daqing Chen, Sai Liang Sain, and Kun Guo, “Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining,” Journal of Database Marketing and Customer Strategy Management, Vol. 19, No. 3, pp. 197–208, 2012.

[7] EAST JAPAN RAILWAY COMPANY, `http://www.jreast.co.jp/e/`, June 24, 2017.

[8] Money Forward, `http://corp.moneyforward.com/`, June 24, 2017.