様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　５　月　２９　日現在

機関番号：　１２６１２
研究種目：　基盤研究(C)（一般）
研究期間：　2014～2017
課題番号：　２６３３０１５３

研究課題名（和文）匿名化技術への体系的な個人特定攻撃および防御手法の研究

研究課題名（英文）Systematic studies on de-anonymization attacks and their countermeasures

研究代表者
吉浦　裕（Yoshiura, Hiroshi）

電気通信大学・大学院情報理工学研究科・教授

研究者番号：４０３６１８２８

交付決定額（研究期間全体）：（直接経費）　　3,700,000円

研究成果の概要（和文）：ネットビジネスの発展やIoT環境の実現に伴い，個人の様々な活動に関するデータを集約し，データマイニングを通じて活用することが始まっているが，個人情報の利用に伴うプライバシー保護が問題になっている．そこで，データ中の個人情報を匿名化する技術が期待されてるが，その安全性の検討は不十分であった．本研究では，匿名化技術の安全性の分析を目的として，匿名化されたデータから個人を特定する攻撃手法を検討した．また，その検討結果に基づいて，攻撃に耐える匿名化の強度を検討した．SNSの投稿文，およびWiFi基地局から収集した移動履歴を事例データとし，機械学習を用いた個人特定技術の開発と評価と行った．

研究成果の概要（英文）：While data about various personal activities had been collected and used through data mining, such use of personal data had caused privacy problem. Anonymization techniques had been expected to solve the privacy problem but security of these techniques had not been sufficiently analyzed. To clarify the security of anonymization techniques, we studied de-anonymization attacks that enable re-identifying persons from anonymized data. We also studied level of anonymization that stands against the de-anonymization attacks. We developed our de-anonymization techniques based on machine learning by using real data of SNS posts and of location histories collected through WiFi base stations.

研究分野：　情報セキュリティ

キーワード：　プライバシー保護　個人情報保護　情報セキュリティ　匿名化　個人特定

１．研究開始当初の背景

本研究は 2014 年度に開始した．当時は，ビッグデータの利活用が始まった時期であり，センシングネットワーク，オンラインビジネス，ソーシャルネットワーク（SNS），携帯電話や WiFi 通信システムなど多種多様な経路を通じて，個人や組織の行動に関するデータが収集されるようになっていた．これらのデータをマイニングし，営利および公共活動に活用することの期待が高まる一方，データに含まれる個人情報の扱いが問題になっていた．そのような状況下で，個人情報を含むデータの活用とプライバシー保護を両立する手段として，匿名化が期待されていた．

匿名化は，情報と個人との結びつきを弱くする技術である．特に，情報の削除や修正により、当該情報が誰に関する情報であるか推定困難にすることを主たる目的とする．匿名化の手法としては、識別情報の除去や符号化、ノイズの重畳、複数の個人情報の識別不能化（k-匿名化およびその改良手法）等が提案されていた。しかし，匿名性の保証範囲，すなわち、どのような条件下でどこまで匿名性が維持できるかは明らかでなかった．例外として，差分プライバシーに基づくノイズ重畳は，匿名性の範囲を数学的に保証することができたが，ノイズによる情報の劣化が著しいため，実用性が低かった．

匿名性の保証範囲が不明確である原因は，匿名性を破る攻撃の研究とそれに対する防御の研究が共に少なく、匿名化技術の安全性が充分に分析されていない点にあった．

匿名性を破る攻撃の研究、すなわち匿名情報から個人を特定する研究は、2000 年頃に、匿名情報と実名情報の照合の観点から検討され、一定の成果をあげていた。しかし、ノイズ重畳により匿名化された数値情報とノイズのない数値情報との距離を算出し、距離の近いペアを対応付ける方法が主流であり、2000 年以降に提案された k-匿名化等の重要な匿名化手法を扱っていなかった。また、単純な数値ではない情報、たとえば移動履歴やソーシャルネットワークの開示情報など近年重要になった個人情報を扱っていなかった。

2000 年以降は、個別の具体例において個人が特定できることを示す事例研究が多い[2,3]。たとえば、Srivatsa らは、2013 年に，複数人の移動履歴から推定される人間関係と、ソーシャルネットワークから推定される人間関係の照合により個人の特定が可能であることを示した。しかし、これらの研究は、場当たり的な事例研究に留まっていた。

一方，人工知能，特に様々な機械学習技術が成熟すると共に，深層学習技術の研究が活発になり，機械学習をプライバシー保護に適用する試みが始まっていた．

２．研究の目的

上述した場当たり的な個人特定（たとえば Srivatsa）を分析すると、匿名情報と実名情報の照合を非数値情報に対して行っていることが分かる。ここで，照合とは，同じ人物の 2 つの情報の対を見出すことであり，照合を通じて匿名情報の対象者が特定される．また，複数の匿名情報を照合することで，対象者に関する多面的な情報の入手が可能となり，対象者の特定につながる．人間が匿名情報から個人を想起する際にも、記憶内の実名あるいは別の匿名の情報と照合していると考えられるため、匿名情報と実名情報（あるいは他の匿名情報）の照合は、個人特定の汎用的なモデルになると考えられる。

そこで、本研究の第 1 の目的は、匿名情報を含む情報間の照合を、非数値情報や新しい匿名化法にも適用できるように拡張し、個人特定攻撃の体系的な手法を確立することである。

照合には距離の近さ（類似度）の算出が必要であるが、非数値の匿名情報と実名情報に対して、同一人物の情報が最も類似となるような類似度の定義を人手で考案することは容易ではない。そこで、第 2 の目的は、非数値の匿名情報と実名情報に対して、機械学習による類似度定義の学習を実現し、個人特定の精度を向上することである。

３．研究の方法

匿名情報を含む情報間の照合について，当時，2 つのアプローチがあった．一つは，前述した Srivatsa の手法のように，N 人分の情報の集合が 2 組存在する時に，集合の要素間を対応付ける手法である（図 1）．対応付けのために，N 人の間の関係をグラフ等の形で表現し，個々の集合を構造化することで，2 つの構造間のマッチングを行う．Srivatsa の手法の場合は，N 人の間の友人関係を用いて集合をグラフ化し，グラフ間のマッチングを行った．

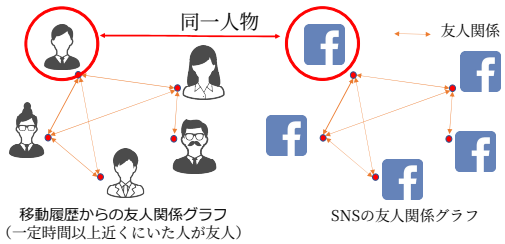

図１　構造マッチングによる照合例

もう一つの方法は，攻撃の対象となる 1 人分の匿名情報と，照合の候補となる N 人分の情報を前提とし，攻撃対象情報と各候補情報の類似度関数（あるいは距離関数）を機械学習等によって実現する（図 2）．N 人の情報のうち匿名情報に最も類似した情報を同一人物の情報と推定する．

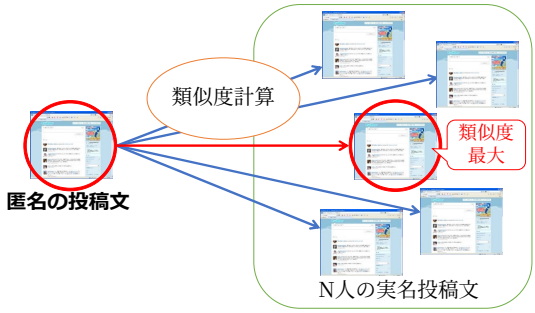

図 2　類似度による照合例

上記のうち第 1 の方法は，互いに関係のある N 人分の情報を 2 組入手する必要があり，一般性に欠けるため，本研究では第 2 の方法を採用し，類似度関数を機械学習で実現することにした．

機械学習を用いた照合の先行研究は，主に SNS の投稿文を対象としていた．N 人分の実名の投稿文から各投稿者の文章の傾向を，N クラス分類器の形で学習する．匿名の投稿文を N クラス分類器に入力し，一つのクラスを推定し，そのクラスに対応する人物を匿名投稿文の投稿者であると判定する．しかし，先行研究の手法は，攻撃対象の匿名情報と照合対象の情報がいずれも自然言語の文章であることを前提としており，大きな制約があった．

そこで，この制約を打破するために，本研究では，以下の 2 つの方針を採った．

(1)異種情報間の照合のために，異種情報間の類似度関数を機械学習によって実現する（図 3）．その第 1 段階として，先行研究で取り上げられていた SNS の投稿文を匿名情報とし，N 個の照合対象情報として，N 人の履歴書を取り上げた．履歴書は，企業や公共機関等の殆どの組織が保有している．また，学校も学生や生徒の履歴書に相当する情報を保有している．そこで SNS 上の内部告発，情報漏洩，いじめ発言等の投稿者を，履歴書との照合を通じて特定するといった現実的な応用が考えられる．

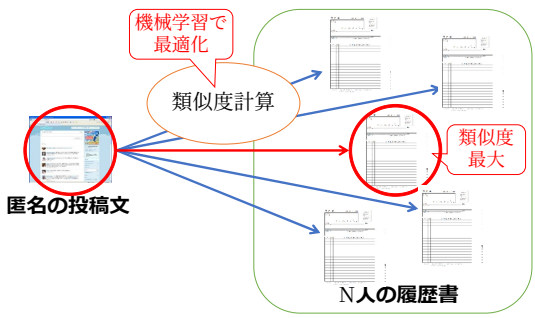

図 3　SNS と履歴書の照合

(2)移動履歴を匿名情報とし，SNS の投稿文をと照合する（図 4）．本研究を開始する前年（2013 年）に，携帯電話利用者の移動履歴（基地局を通じて取得した場所と時刻の列）を活用するシステム（モバイル空間統計）が我が国ではじめて実用化されており，移動履歴の匿名性には大きな関心が集まっていた．また，移動履歴は自然言語文とは性質が大きく異なるため，自然言語文以外の照合技術を確立する上で適切な対象である．

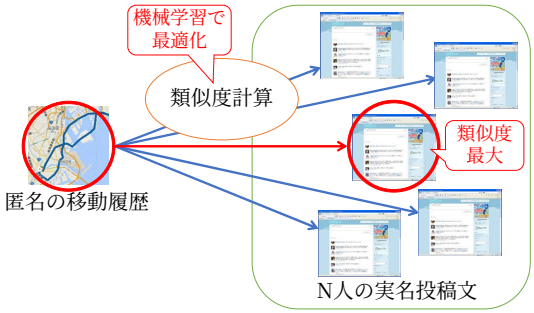

図 4　移動履歴と SNS の照合

(3)移動履歴から対象者の属性（年齢，性別，国籍）を推定することにした．これは，照合による個人の特定ではないが，匿名情報から対象者に関する隠された情報を推定することに相当し，推定した情報が個人の特定につながる．

4．研究成果

(1)SNS 投稿文と履歴書の照合

先行研究では，N 人の実名投稿文の傾向を N クラス分類器で学習し，匿名投稿文の傾向に基づいて分類器がクラス(すなわち人）を推定していた．しかし，これは学習データと匿名データが両方共自然言語文であることを前提としている．本研究では，履歴書が実名投稿文に相当するが，履歴書は文章ではなく，キーワードの列である．たとえば，（性別・女性），（年齢・24 才），（所属・電気通信大学情報学専攻在学）といった形式になっている．そのため，文章の傾向を直接学習するというアプローチは不可能であった．

そこで，履歴書を，（属性・属性値）の集合とみなし，当該履歴書に含まれる（属性・属性値）毎に，ネット上の投稿文を用いて 2 クラス分類器を学習する．たとえば，（性別・女性）については，男性であるかの分類器を，ネット上の男性の投稿文と女性の投稿文から学習する．また，電気通信大学情報学専攻在学であるかの分類器を，ネット上の当該専攻学生の投稿文とそれ以外の投稿文から学習する．そして，これらの（属性・属性値）毎の分類器の組合せにより，履歴書の人物の識別器を構成する．たとえば，（性別・女性），（所属・電気通信大学総合情報学科），（趣味・テニス）の 3 つの 2 クラス分類器全てに該当する投稿文は，この 3 つの（属性・属性値）から成る履歴の人物が投稿したと判定する（図 5）．

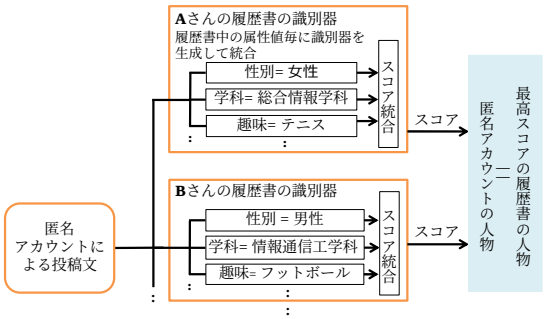

図 5　SNS と履歴書の照合方式

機械学習アルゴリズムとして，SVM（線形），ナイーブベイズ，ロジスティック回帰，SVM（ガウシアンカーネル），ランダムフォレスト，XG ブーストの 6 種類，特徴量として単語の出現頻度と出現有無の 2 種類，（属性・属性値）毎の出力を統合する方法として平均と積の 2 種類を用いて，上記の提案手法を実装した．30 人の被験者を用いた評価実験を行った結果，XG ブースト・出現頻度・平均の組合せが最良で，11 人が照合に成功，16 人が上位 10%に絞込み可能であった．

(2)移動履歴と SNS 投稿文の照合

移動履歴として，WiFi 基地局から収集した緯度・経度・収集時間から成る時系列情報を取り上げた．一方，SNS の投稿文から地名を検知し，緯度・経度と投稿時刻によって，時系列情報に変換した．そして，この 2 つの時系列情報を照合した．

SNS から生成した時系列情報には，ノイズが大量に含まれる．たとえば，「ローマに旅行したい」のように，実際には行っていない場所が含まれる．また，「昨日新宿で買い物をした」のように，実際にその場所にいた時刻と投稿時刻が異なる場合がある．さらに，自然言語処理ツールが人名，会社名，地名を混同することが多い．また，遠隔地に移動したが，その地名を投稿しなかった場合には，同一人物の WiFi 移動履歴と SNS 移動履歴であっても，距離が遠くなる．WiFi 移動履歴は 1 日に 100 データ程度であるが，SNS からの移動履歴は 1 日に 1 データ程度であり，データの頻度が 2 桁異なる．以上により，この照合は容易ではなかった．

そこで，機械学習の検討に先立って，以下のアドホックな手法を検討した（図 6）．

- 信頼性の高い WiFi 移動履歴を基準とし，WiFi 移動履歴の各地点から最も近い SNS 移動履歴の地点をペアリングする．全ペアの距離の平均値を当該 WiFi 移動履歴と SNS 移動履歴の距離とし，距離の最も近い SNS 移動履歴を WiFi 移動履歴と同一人物の情報とする．これにより，SNS 移動履歴中のノイズが距離に反映されないようにする．
- 上記の WiFi 移動履歴と SNS 移動履歴のペアのうち遠い方から一定割合を無視することで，遠隔地に移動したが，その地名を投稿しなかったケースに対応する．
- 場所と時間の両方が一致する場合には，本人の可能性が高いので，距離が小さいとみなす．
- 時間が近いが場所が遠い場合には，他人の可能性が高いので，距離が大きいとみなす．

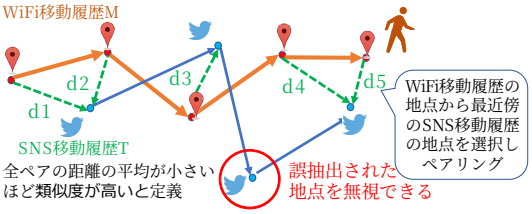

図 6　移動履歴と SNS の照合方式

上記のアドホックな手法を実装し，60 人の被験者データを用いて，3 交差検定を行った結果，20 人中平均 9.5 人の照合に成功した．

このアドホックな手法の知見を踏まえて，機械学習による照合手法を設計し，機械学習アルゴリズム XG ブーストを用いて実装した．現在は，実装した機械学習システムの評価実験を行っている．

(3)移動履歴からの個人の属性の推定

WiFi 通信事業者との契約を経て，訪日外国人観光客 160000 人の移動履歴を入手し，移動の傾向から個人の属性を推定する手法を設計，実装した．機械学習アルゴリズムとして，SVM（線形），ロジスティック回帰，SVM（ガウシアンカーネル），ランダムフォレスト，XG ブースト法を用いた評価の結果，人数の多い 8 か国のなかから 45%の精度で国籍を特定，10 代から 60 代までの 6 年代のなかから 25%の精度で年代を特定することができた．性別の推定は困難であった．これは，訪日旅行において家族旅行やペア旅行など男女が一緒に行動する場合が多いからであると予想される．現在，推定が困難である原因を分析中である．

5．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 5 件）

[1] Takayasu Yamaguchi, Hiroshi Yoshiura: Maintaining Information in Differential Privacy by Using Insensitive Relationships between Personal Attributes, International Journal of Informatics Society, Refereed, Vol.10, No.1, 2018（掲載決定）

[2] 橋本英奈，宮崎夏美，市野将嗣，久保山哲二，越前功，吉浦裕：機械学習を

用いたソーシャルネットワークと履歴書の照合方式の提案，情報処理学会論文誌，査読有，Vol.58，No.12，pp.1863-1874，2017.
https://ipsj.ixsq.nii.ac.jp/ej/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=185087&item_no=1&page_id=13&block_id=8

[3] 小川陽平，市野将嗣，久保山哲二，吉浦裕：ソーシャルネットワークの発言者を特定するシステムの提案と予備評価，日本セキュリティ・マネジメント学会誌，査読有，Vol.30，No.2，pp.3-19，2016.

[4] Hoang-Quoc Nguyen-Son, Minh-Triet Tran, Hiroshi Yoshiura, Noboru Sonehara, Isao Echizen: Anonymizing Personal Text Messages Posted in Online Social Networks and Detecting Disclosures of Personal Information, IEICE Transactions on Information and Systems, Refereed, Vol.E98-D, Issue 1, pp.78-88, 2015.
DOI: https://doi.org/10.1587/transinf.2014MUP0016

[5] 村木友哉，内田貴之，市野将嗣，越前功，吉浦裕：安全性と有用性のトレードオフが制御可能な顔画像の匿名化方式，日本セキュリティ・マネジメント学会誌，査読有，Vol.28，No.3，pp.3-16，2015.

〔学会発表〕（計１４件）

[1] Yoichi Midorikawa, Masatsugu Ichino, Hideki Yoshii, Hiroshi Yoshiura: Estimating Nationality from Behavioral Data, 2018 International Conference on Business and Information - Winter Session -, Okinawa, 2018.1.25.

[2] Fumito Nakazawa, Masatsugu Ichino, Hideki Yoshii, Hiroshi Yoshiura: Utility of Anonymized Location Histories - Experiments Using Large-Scale Real-World Data -, 2018 International Conference on Business and Information - Winter Session -, Okinawa, 2018.1.24.

[3] Takayasu Yamaguchi, Hiroshi Yoshiura: A Strategy for Mitigating Information Loss in Anonymization by Using Nature of Personal Data, 2017 International Workshop on Informatics, Zagreb, Croatia, 2017.9.4.

[4] Takayasu Yamaguchi, Hiroshi Yoshiura: Privacy-Preserving Recommender System Based on Mining Large-Scale Data Distributed among Organizations, 2017 International Conference on Business and Information, Hiroshima, 2017.7.4.

[5] 佐々木千慧，浅見航太朗，吉井英樹，久保山哲二，越前功，吉浦裕：移動履歴とソーシャルメディアの照合による個人の特定，暗号と情報セキュリティシンポジウム 2017，沖縄，2017.1.27.

[6] 山口高康，寺田雅之，吉浦裕：差分プライバシに基づく一括開示と対話開示のデータ有用性の評価 -多属性に関する考察-，コンピュータセキュリティシンポジウム 2016，秋田，2016.10.13.

[7] Eina Hashimoto, Masatsugu Ichino, Tetsuji Kuboyama, Isao Echizen,Hiroshi Yoshiura: Breaking Anonymity of Social Network Accounts by Using Coordinated and Extensible Classifiers based on Machine Learning, 15th IFIP Conference on e-Business, e-Services and e- Society, Swansea, UK, 2016.9.12. (Best Paper Award 受賞)

[8] 山口高康，寺田雅之，吉浦裕：差分プライバシに基づく一括開示と対話開示のデータ有用性の評価，第 74 回コンピュータセキュリティ研究会，山口，2016.7.7.

[9] Kilho Shin, Tetsuji Kuboyama, Takako Hashimoto, Dave Shepard: Super-CWC and super-LCC: Super fast feature selection algorithms, IEEE International Conference on Big Data, Santa Clara, US, 2015.10.

[10] Hoang-Quoc Nguyen-Son, Minh-Triet Tran, Hiroshi Yoshiura, Noboru Sonehara, Isao Echizen: A Rule-Based Approach for Detecting Location Leaks of Short Text Messages, Workshop on Privacy by Transparency in Data-Centric Services, Poznan, Poland, 2015.6.26.

[11] Yohei Ogawa, Eina Hashimoto, Masatsugu Ichino, Isao Echizen, Hiroshi Yoshiura: De-Anonymising Social Network Posts by Linking with Resume, Workshop on Privacy by Transparency in Data-Centric Services, Poznan, Poland, 2015.6.26.

[12] Hoang-Quoc Nguyen-Son, Minh-Triet Tran, Hiroshi Yoshiura, Noboru Sonehara, Isao Echizen: A System for Anonymizing Temporal Phrases of Message Posted in Online Social Networks and for Detecting Disclosure, 4th International Workshop on Resilience and IT-Risk in Social Infrastructures, Fribourg, Switzerland, 2014.9.8.

[13] Haruno Kataoka, Yohei Ogawa, Isao Echizen, Tetsuji Kuboyama, Hiroshi Yoshiura: Effects of External

Information on Anonymity and Role of Transparency with Example of Social Network De-anonymisation, 4th International Workshop on Resilience and IT-Risk in Social Infrastructures, Fribourg, Switzerland, 2014.9.8.
[14] 内田貴之, 村木友哉, 市野将嗣, 越前功, 吉浦裕: 安全性と有用性のトレードオフが調整可能な顔画像の匿名化, 日本セキュリティ・マネジメント学会全国大会, 八王子, 2014.6.21.

６．研究組織

(1)研究代表者
吉浦　裕（YOSHIURA, Hiroshi）
電気通信大学・大学院情報理工学研究科・教授
研究者番号：４０３６１８２８

(2)研究分担者
久保山　哲二（KUBOYAMA, Tetsuji）
学習院大学・計算機センター・教授
研究者番号：８０３０２６６０