情報処理学会論文誌 Vol.58 No.12 1863–1874 (Dec. 2017)

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

橋本 英奈[1] 宮崎 夏美[1] 市野 将嗣[1] 久保山 哲二[2] 越前 功[3] 吉浦 裕[1,a)]

受付日 2017年3月8日, 採録日 2017年9月5日

概要：ソーシャルネットワークのプライバシリスクの明確化および悪用の抑止のために，匿名のソーシャルネットワークアカウントを組織が保有する履歴書と照合する手法を述べる．提案手法は機械学習を用い，履歴書に含まれる性別，趣味などの属性値ごとに，ソーシャルネットワークアカウントの投稿文が当該属性値を持つ人によって書かれたものかを判定する．この属性値ごとの識別器を組み合わせることにより，ソーシャルネットワークの投稿文が履歴書の当人によって書かれたものであるかを判定する．機械学習のための訓練データはソーシャルネットワーク上の他のアカウントから収集する．30 人の被験者のソーシャルネットワークアカウントと履歴書を用い，投稿文の特徴量を 2 種類，機械学習アルゴリズムを 5 種類，履歴書中の着目する属性群 3 セット，属性ごとのスコアの統合方法 2 種類により，提案手法を評価した．その結果，最良ケースにおいて，30 アカウント中 5 アカウントは本人の履歴書と正しく照合でき，14 アカウントは 30 人中 3 人に絞り込むことができ，19 アカウントは 6 人に絞り込むことができた．

キーワード：プライバシ，ソーシャルネットワーク，個人特定，機械学習

Linking Social Network and Resume by Using Machine Learning

Eina Hashimoto[1] Natsumi Miyazaki[1] Masatsugu Ichino[1]
Tetsuji Kuboyama[2] Isao Echizen[3] Hiroshi Yoshiura[1,a)]

Received: March 8, 2017, Accepted: September 5, 2017

Abstract: This paper describes a method that links anonymous accounts of social networks to resumes held by organizations. Using machine learning, the proposed method generates a classifier for each attribute value described in each resume, such as gender of female and hobby of dancing. It uses each classifier to judge posts in an account were written by a person who has such an attribute value. By combining scores from these resumes, the method judges the posts were written by a person of the resume. Training data for machine learning are collected from other accounts of the social network. The proposed method was evaluated by using 30 pairs of accounts and resumes with 2 kinds of sentence feature, 5 machine learning algorithms, 3 sets of resume attributes, and two methods of score fusion. In the best combination of parameters, the correct resumes were identified for 5 accounts, they were in 3 identified resumes for 14 accounts and in 6 identified resumes for 19 accounts.

Keywords: privacy, social network, identification, machine learning

1. はじめに

近年，ソーシャルネットワークサービス（以下，SNS）が広く普及し，人々のコミュニケーションを豊かにしている．SNS 上にはテキストや画像，位置情報など様々な情報が流通し，友人とのコミュニケーション手段だけでなく，企業によるマーケティング活用さらには，不正の内部告発

[1] 電気通信大学
University of Electro-Communications, Chofu, Tokyo 182–8585, Japan
[2] 学習院大学
Gakushuin University, Toshima, Tokyo 171–8588, Japan
[3] 国立情報学研究所
National Institute of Informatics, Chiyoda, Tokyo 101–8430, Japan
[a)] yoshiura@hc.uec.ac.jp

や政治活動など，利用目的も多岐にわたっている．その一方で，SNS 上にはユーザやその友人の個人に関する情報が漏洩し，プライバシ問題を引き起こしている．また組織の機密情報の漏えい，誹謗中傷などセキュリティの問題も引き起こしている．

これらのリスクの明確化やインシデントの抑止のために，SNS のアカウントが誰のものであるか，あるいは，SNS 上の投稿文の投稿者が誰であるかを特定する研究が行われている．このような SNS アカウントおよびコンテンツからの個人特定ができれば，内部告発者や政治活動者が特定されるリスクを明確化し，警鐘を鳴らすことおよび，機密情報の不正な漏洩者や誹謗中傷者の抑止にもつなげることができる．SNS アカウントおよびコンテンツからの個人特定の研究の例として，Almishari らは機械学習によって writing-style（文章の特徴）を学習し，異なる SNS アカウントを利用する同一ユーザを特定した [2]．また，Narayanan らも機械学習によって writing-style を学習し，同一ユーザの複数のブログを特定した [3]．しかし，これら先行研究の多くは SNS どうしまたは SNS と他のメディアの照合を行っており，個人の特定という意味では，間接的な攻撃手法であるといえる．たとえば，アカウント A とアカウント B が同一人物のものであると判明しても，その人物が誰であるかは分からない．

そこで本論文では，匿名の SNS アカウントと履歴書を照合する技術を提案する．SNS アカウントと履歴書の直接的な照合は困難であるため，提案技術では，機械学習によって，ある文章がある履歴書の人物によって書かれたものであるか，該当の度合いを算出するように識別器を構成する．この識別器を用いて，複数の履歴書の各々に対して，匿名アカウントの投稿文が当該履歴書の人物によって書かれたか該当の度合いを算出し，該当度合いが最大の履歴書の人物をアカウントの所有者と推定する．履歴書は個人を直接的に示すので，履歴書との照合により，SNS アカウントから個人を一意に特定することができる．また，企業や大学，公的機関などの多くの組織は履歴書あるいは履歴書に相当する情報を保持しているため，提案手法はプライバシ侵害のリスクの明確化および SNS の不正利用の抑止の両面で有効であると考える．以下，2 章では，本論文に関わる先行研究を述べる．3 章では提案方式を説明する．4 章はサンプルデータを説明し，5 章は予備評価，6 章は本評価を述べる．7 章は本論文の貢献，8 章は結論と今後の課題を述べる．

2. 先行研究

SNS のプライバシリスクを明確にするための研究が多数行われてきた．初期の研究では，ヒューリスティックアルゴリズムを用いたキーワードやグラフの照合によって，ユーザの個人情報を推定していた．たとえば，2007 年，Backstrom らは SNS 上の友人関係を表すソーシャルグラフから既知の人間関係を検知することで，ソーシャルグラフのノードの人物が誰であるかを特定した [4]．2008 年，Lam らはユーザの友人からのコメントをキーワード照合によって分析することで，72%のユーザのファーストネームを，また 30%のユーザのフルネームを正しく推定した [5]．2010 年頃からは機械学習などを用いた系統的な手法が提案されるようになった．2011 年，Mao らはナイーブベイズ分類器とサポートベクタマシン（以下，SVM）を用いることで，Twitter のつぶやきのうち旅行や病状などの個人情報を含むものを 76%の精度で，また飲酒中のつぶやきを 84%の精度で特定した [6]．Mao らの方式における訓練データは Twitter のつぶやきであり，個人情報を含む正例と含まない負例は人手によってラベル付けしていた．2012 年，Kótyuk らはアカウントのプロフィールに開示されていない年齢，性別，既婚・未婚などの情報を，プロフィールに開示された部分や，友人のアカウント情報，投稿文などから，ニューラルネットワークを用いることで推定した [7]．2014 年，Caliskan-Islam らはナイーブベイズ分類器とアダブーストを用いて，アカウントにおける個人情報の漏洩度合を 3 段階に分類した [8]．

最近の関連研究では，対象のアカウントや投稿を他のアカウントや投稿と照合することに焦点が置かれている場合が多い．2009 年，Narayanan らは複数の異なる SNS を利用する同一ユーザの Twitter と Flickr のアカウントを，サブグラフマッチングによって 12%の誤り率で照合した [9]．2010 年，Polakis らは SNS 上のユーザ名とユーザが使用しているメールアドレスの照合を行った [10]．その結果，Facebook から抽出されたユーザプロフィールのうち，43%が正しく照合された．2012 年，Goga らは位置情報，タイムスタンプ，writing-style などの特徴を解析することで，複数の異なる SNS（Yelp，Twitter，Flickr）を使用している同一ユーザを特定する方法を提案した [1]．同年，Narayanan らは SVM や線形判別分析など，複数の機械学習アルゴリズムを用いて，ブログから writing-style を学習することで，同一投稿者による複数のブログを特定した [3]．2014 年，Almishari らはナイーブベイズ分類器を用いて，Twitter アカウントの投稿文から writing-style を学習することで，同一人物の複数の Twitter アカウントを特定した [2]．アカウントや投稿を照合するこれらの手法のほとんどは，2 つのアカウントや投稿が同一人物のものであることを検知するだけで，その人物が誰かを特定することはなかった．なお，例外として，Goga らの手法 [1] は，個人を特定できるアカウントを前提とし，同じユーザの別アカウントを見出すことで，当該個人に関するより多くの情報の収集を可能にする．Goga らの手法は，ユーザが複数のアカウントを利用し，そのうちの 1 つのアカウントにおいて，個人が特定可能な場面を前提としている．一方，

提案手法は，攻撃者が，ユーザの履歴書を持っていることを前提とする．このように，Goga らの手法と提案手法は，異なる場面で有効である．

3. 機械学習を用いた匿名アカウントと履歴書の照合

3.1 機械学習利用における課題

提案手法で機械学習を利用する際に生じる課題の 1 つとして，訓練データの準備が困難であることがあげられる．1 章，2 章で述べたように，SNS アカウントあるいは投稿文を同一人物のアカウントあるいは投稿文に照合する手法は，ある人物のものであることが分かっているアカウントや投稿文から当該人物の writing-style を学習し，類似した writing-style を持つアカウントや投稿文を特定していた [2], [3]．これらの手法では，当該人物のブログやアカウントが入手できることを前提としていた．また，Kótyuk らの手法では，ユーザの年齢，性別，友達の数，既婚・未婚，使用言語などの属性情報が開示されているアカウントから，これらの属性情報間の関係を学習した．その学習結果を用いて，属性情報が一部のみ開示されたアカウントにおいて，非開示の情報を推定した [7]．Kótyuk らの手法における訓練データは開示されたアカウントの属性情報であるため，その収集は難しくはない．

しかし，訓練データの準備はつねに容易であるとは限らない．Mao らは，個人情報を含む投稿を正例，含まない投稿を負例とした訓練データを使用したが，正負のラベル付けは人手で実施されているため，準備に時間を要し，大規模な利用は困難である．また，新たな種類の個人情報（たとえば収入や住所，薬の使用など）を検知する場合には，そのつど人手によるレベル付けを行う必要がある．Caliskan-Islam らは，ソーシャルアウトソーシングによって研究者自身によるラベル付けを不要にした [8]．しかし，これは人手によるラベル付けの作業を不要化したわけではなく，作業者を研究者から外注労働者にシフトしただけであるため，これらの問題を解決したとはいえない．また，Hart らは訓練データにコーパスを使用した [11] が，コーパスの作成にも多くの時間と労力を必要とする．

提案手法のように匿名アカウントと履歴書を照合する場合，訓練データの準備はさらに困難である．なぜなら，匿名アカウントの投稿文は通常の文章であるため writing-style の学習が可能であるが，履歴書は通常の文章ではなく単語の羅列であるため，writing-style を持たない．したがって，writing-style を学習しても，それを用いて履歴書との照合を行うことはできない．また，そもそも，訓練データとして何を用いるか，何が正例あるいは負例であるかが不明確であるため，ラベル付けを外注労働者に依頼することもできない．そこで，次節に述べる方法によって，この問題の解決を試みた．

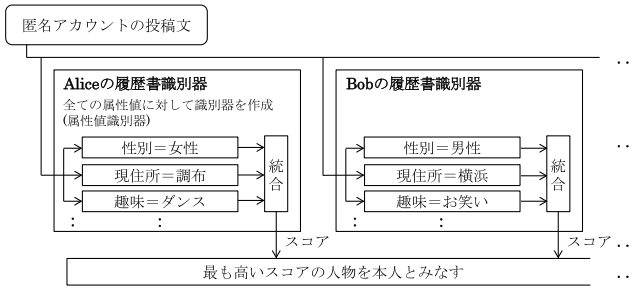


図 1 提案手法の概要
Fig. 1 Overview of proposed method.

3.2 提案方法

履歴書は，複数の属性とそれに対する属性値によって構成されている．たとえば，「性別 = 女性」，「現住所 = 東京都 A 市」，「帰省先 = 大阪府 B 市」，「所属 = 企業 C」，「学歴 = 1997 年 3 月 電気通信大学 卒業，1997 年 4 月 電気通信大学大学院 入学，2000 年 3 月 電気通信大学大学院 修了見込み」，「趣味 = ダンス」などである．提案手法では，まず履歴書から抽出した全属性値それぞれに対して，機械学習によって識別器を実装する．たとえば，匿名アカウントの人物の帰省先が大阪府 B 市であるかを識別する識別器，趣味がダンスであるかを識別する識別器，女性かどうかを識別する識別器[*1]などを実装する．以下，これらの識別器を属性値識別器と呼ぶ．

次に，履歴書に含まれるすべての属性値について，該当する属性値識別器を組み合わるせることによって，当該履歴書に対する識別器を構成する．以下，これらの識別器を履歴書識別器と呼ぶ．履歴書識別器は，匿名アカウントの人物が当該履歴書の人物であるか，すなわち匿名アカウントの人物が当該履歴書に記された属性値を有する人物であるかを特定する[*2]．履歴書識別器が出力するスコアは，当該履歴書から作成されたすべての属性値識別器のスコアを統合した値である．提案手法の手順をまとめると以下のとおりである（**図 1**）．

(1) 保有するすべての履歴書からすべての属性値を抽出し，各属性値に対して属性値識別器を作成する．
(2) 各履歴書から作成した属性値識別器を組み合わせることにより，履歴書識別器を構成する．
(3) 匿名アカウントから投稿文を入手する．
(4) 匿名アカウントから得られた投稿文を各々の履歴書識別器へ入力する．その結果，履歴書とアカウントの組合せごとに，それぞれの照合度合を示すスコアが出力される．
(5) 全履歴書識別器のうち，匿名アカウントに対して最も高いスコアを算出した履歴書識別器の人物を，匿名ア

[*1] 厳密には，当該アカウントの投稿文が女性によって書かれた文章であるかを識別する．
[*2] 厳密には，当該アカウントの投稿文が当該履歴書の属性を持つ人物によって書かれた文章であるかを識別する．

カウントの人物として推定する．

訓練データの収集はツイプロ（検索ワードをプロフィールに含むTwitterアカウントを検索するAPI）によって自動化し[12]，当該属性値をプロフィールに含むアカウントを訓練データの正例として用いた．また，プロフィールに当該属性値を含まないアカウントを訓練データの負例として用いた．

4. サンプルデータ

4.1 被験者から入手したデータ

著者らが所属する電気通信大学の学生30人のTwitterアカウントと履歴書を入手した．**表 1** は被験者の内訳である．履歴書情報は次の13項目を含んでいる．(1) 氏名，(2) 生年月日，(3) 性別，(4) 現住所（市・区・郡まで），(5) 帰省先住所（市・区・郡まで），(6) 学歴，(7) 職歴，(8) 交通手段，(9) 電車区間，(10) 得意科目，(11) 長所・特徴（自己PR），(12) 趣味（クラブ活動・サークルを含む），(13) 資格．これらの中で評価に用いた項目は，(3) 性別，(4) 現住所，(5) 帰省先住所，(6) 学歴，(10) 得意科目，(12) 趣味，(13) 資格の7種類である．ただし，一般的に学歴は複雑であるため，代表的な学歴情報として被験者が所属している学科を用いた．また，被験者のTwitterアカウントについて，1アカウントあたり2,167〜3,000件の投稿（つぶやき）を使用した．なお，1アカウントあたりの平均つぶやき数は2,771件であった．各被験者のつぶやきやプロフィールは，本人の情報が推測されないよう，日頃から被験者自身による個人情報の省略・変更や，偽の個人情報の使用などによる匿名化が施されている．

4.2 訓練データ

訓練データは，3.3節で述べたようにツイプロによって入手した．各属性値識別器のための訓練データの正例として，当該属性値をプロフィールに記載しているアカウントであって，つぶやき数が1,000以上であるものを30件検索し，それぞれのアカウントから最大3,000までのつぶやきを取得して使用した．たとえば，「趣味 = ダンス」の識別器を作成する場合，ツイプロに検索ワードとして「ダンス」を入力し，プロフィールに「ダンス」が記載されているアカウントを検索する．なお，当該属性値をプロフィールに記載し，つぶやきが1,000件以上であるアカウントが30件に満たない場合，検索結果が10件以上であれば正例として利用し，10件未満の場合には，当該属性値に対する識別器の学習を断念した．負例には，当該属性値をプロフィールに含まないアカウントであって，つぶやき数が1,000以上であるものをランダムに30件収集したものを使用した．

5. 予備実験

投稿文から抽出する特徴量，機械学習アルゴリズム，着目すべき属性，属性値識別器のスコアから履歴書識別器のスコアを算出するためのスコア統合方法を選定するために予備評価を行った．特徴量の候補としてbag-of-words（単語の出現頻度）とbinary（単語が文章中に出現したか否か）の2種類，機械学習アルゴリズムの候補として，基本的なアルゴリズムである線形SVM，非線形SVM，ロジスティック回帰，ナイーブベイズ，RandomForestの5種類，着目すべき属性の候補として性別，現住所，帰省先住所，学歴，得意科目，趣味，資格の7種類，スコア統合方法の候補として積，平均の2種類を評価した．

5.1 予備評価用データ

予備評価用データとして，30人中6人の被験者の履歴書とつぶやきを使用した．**表 2** は，予備評価用データの履歴書から抽出した属性と属性値を示している（被験者の現住所と帰省先は，プライバシの観点から匿名化している）．被験者6人の履歴書からは，7個の属性と46個の属性値が抽出されたが，6個の属性値については十分な正例数を取得できなかったため，識別器を作成しなかった．識別器を作成不可能であった属性値には，表2の中で取り消し線を引いている．資格については，被験者5以外は識別器を作成することができなかったため，この属性は評価に用いないことにした．以上から，予備評価では，特徴量2種類 × 機械学習アルゴリズム5種類 × 属性値40種類 = 400個の属性値識別器を作成し，被験者6人のつぶやきをテストデータとして用いることで，これらの識別器を評価した．

5.2 正規化

提案手法では，各属性値識別器が算出したスコアを統合することで照合を行う．このとき，属性値識別器によってスコアの値の範囲が異なることが問題になる．たとえば，属性値識別器Aの算出するスコアの値の範囲が，他の属性値識別器の算出するスコアの値の範囲より大幅に大きい場合には，属性値識別器Aのスコアが支配的となり，複数の

表 1 被験者の内訳
Table 1 Demography of subjects.

(a) 性別

男性	女性
20	10

(b) 年齢

20	21	22	23
10	14	5	1

(c) 学年

学部2年	学部3年	学部4年
4	21	5

(d) 学歴

総合情報学科	情報・通信工学科	知能機械工学科
24	2	4

表 2 予備評価に用いた属性と属性値
Table 2 Attributes and their values used for preliminary evaluations.

被験者 No.	性別	現住所	帰省先	学歴	得意科目	趣味	資格
1	F	神奈川県 A 市	埼玉県 E 市	総合情報 学科	プログラミング	お笑い芸人, オードリー, 眼鏡	~~普通自動車免許~~
2	F	神奈川県 A 市	長野県 F 群	総合情報 学科	独語	ピアノ, ~~ピアノの会~~	~~普通自動車免許~~
3	F	東京都 B 市	東京都 B 市	~~情報・通 信工学科~~	体育，音楽, 数学	合気道部, バスケット ボール,音楽，読書, お菓子作り	~~普通自動車免許~~
4	M	神奈川県 C 市	神奈川県 C 市	知能機械 工学科	~~加工学~~	ロボメカ, 工学研究部	~~普通自動車免許~~
5	M	東京都 D 市	北海道 G 市	総合情報 学科	文系科目	テニス, フットサル, テレビ, サッカー	~~普通自動車免許,~~ 証券外務員, FP
6	M	東京都 D 市	青森県 H 市	~~情報・通 信工学科~~	電子回路, 電磁気学, Web デザイン, プログラミング	野球部, 野球, ~~SNS を眺める~~	-

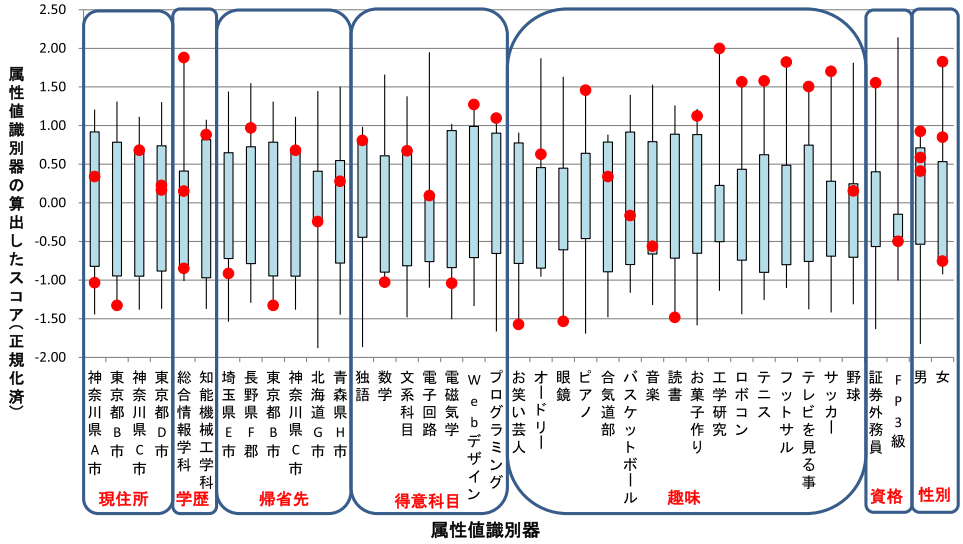


図 2 各属性値識別器のスコア分布（Bag-of-words と RandomForest を用いた場合）
Fig. 2 Scores from attribute-value classifiers with Bag-of words and RandomRorest.

属性値識別器のスコアを統合するメリットが得られない．そこで，各属性値識別器の算出するスコアの値の範囲がおおむね同じになるように，スコアの正規化を行った．

正規化の方法について説明する．属性値識別器ごとに，各アカウントに対して算出したスコアの平均値および標準偏差を計算する．そしてスコアから平均値を引き，その値を標準偏差で割ることにより正規化を行う．具体的には，M をアカウントの数，N を属性値識別器の数とする（予備実験では $M = 6$, $N = 40$）．x_{ij} は j 番目の属性値識別器が i 番目のアカウントに対して算出したスコア（$1 \leq i \leq M$, $1 \leq j \leq N$），$\overline{x_j}$ と σ_j は，j 番目の属性値識別器が算出した M 個のスコア x_{ij} の平均値と標準偏差とする．このとき，x_{ij} を正規化したスコア α_{ij} は，式 (1) のように計算することができる．

$$\alpha_{ij} = \frac{x_{ij} - \overline{x_j}}{\sigma_j} \tag{1}$$

5.3 結果

図 2 は，特徴量が bag-of-words，機械学習アルゴリズムが RandomForest である場合の属性値識別器のスコア分布である．横軸は属性値識別器，縦軸は属性値識別器が 6 個のアカウントに対して算出したスコアの正規化した値を示している．属性値識別器が算出したスコア（正規化済み）の分布は，箱ヒゲ図と点で表されている．箱ヒゲ図の箱は，

表 3 属性ごとにおける本人のアカウントの平均順位
Table 3 Averaged rank of corresponding persons for each attribute.

特徴	機械学習アルゴリズム	現住所	帰省先	学歴	得意科目	趣味	資格	性別	全属性平均	4つの属性平均	全属性－現住所
bag – of - words	線形 SVM	3.50	2.83	3.00	1.86	2.88	2.00	2.00	2.58	2.43	2.43
	非線形SVM	3.67	3.67	3.50	3.29	4.62	3.00	3.67	3.63	3.77	3.63
	ロジスティック回帰	3.33	3.00	3.00	2.14	2.69	2.00	2.00	2.59	2.46	2.47
	ナイーブベイズ	3.50	3.50	4.00	3.14	3.75	5.50	3.33	3.82	3.56	3.87
	RandomForest	3.83	3.83	2.75	2.86	2.75	3.00	2.67	3.10	2.76	2.98
	平均	3.57	3.37	3.25	2.66	3.34	3.10	2.73	3.14	2.99	3.07
binary	線形 SVM	3.17	3.67	2.75	3.14	4.06	2.00	4.00	3.26	3.49	3.27
	非線形SVM	3.60	4.10	4.00	3.00	3.10	4.50	3.6	3.70	3.43	3.72
	ロジスティック回帰	3.83	3.50	4.00	3.86	3.13	5.50	2.83	3.81	3.45	3.80
	ナイーブベイズ	3.50	3.50	4.00	3.14	3.75	5.50	3.33	3.82	3.56	3.87
	RandomForest	3.50	2.67	4.00	3.43	3.13	4.50	3.50	3.53	3.51	3.54
	平均	3.52	3.49	3.75	3.31	3.43	4.40	3.45	3.62	3.49	3.64

スコアの第 1 四分位点から第 3 四分位点まで（すなわち，スコアの 50%にあたる 3 個分相当）のばらつきを表し，箱の上下に伸びる 2 本の線は，スコアの上位 25%（1.5 個分相当）と下位 25%（1.5 個分相当）のばらつきを表している．分布にプロットされている点は，当該属性値を実際に有するアカウントのスコアである．したがって，この点の示す位置が高いほど，当該属性値識別器が正確であるといえる．たとえば，表の最左端は「現在地 = 神奈川県 A 市」の識別器から算出された 6 個のアカウントのスコア分布であり，分布中の 2 つの点は，実際に神奈川県 A 市に住んでいる 2 人の被験者のスコアを表している．

表 3 は，使用した特徴と機械学習アルゴリズムのそれぞれの組合せにおいて，当該属性値を実際に有するアカウントの 6 人中の順位を示している．たとえば，特徴を bag-of-words，機械学習アルゴリズムを RandomForest とした場合，現住所の属性値識別器において，実際にその住所に住んでいる被験者の平均順位は 3.83 位である．順位のとりうる値は 1 位～6 位であるため，期待値は 3.5 位であり，平均順位の値が小さいほど属性値識別器の精度が正確であると考えられる．表 3 より，特徴として binary よりも bag-of-words を用いた方が識別器の精度が高く，bag-of-words を用いた場合の属性ごとの順位に注目すると，学歴，得意科目，趣味，性別の 4 つの属性が効果的であることが見てとれる．

次に，以上の結果をふまえ，特徴として bag-of-words，機械学習アルゴリズムとして，線形 SVM，非線形 SVM，ロジスティック回帰，ナイーブベイズ，RandomForest の 5 種類，使用する属性として全属性，4 つの属性（学歴，得意科目，趣味，性別），全属性から現住所（予備評価で最も精度が悪かった属性）を除いた場合の 3 種類，スコアの統合方法として平均，積の 2 種類，以上の全通りの組合せ 30 ケースについて評価した．

表 4 RandomForest，全属性，平均を用いた場合の各履歴書識別器のスコア
Table 4 Scores of resume classifiers with RandomForest, all attributes and average.

アカウントNo.	履歴書No. 1	2	3	4	5	6
1	-1.8772	-1.8091	-1.7393	-0.6563	-0.5510	***-0.2724***
2	0.5327	***1.4159***	1.2572	0.4543	0.0738	-0.2863
3	-0.7332	***0.0897***	-0.5668	-1.3735	-1.6809	-1.5570
4	0.7635	0.4048	0.9986	***1.8512***	0.7651	1.7652
5	0.9798	0.4715	0.2185	-0.0616	***1.5058***	-0.1729
6	0.3344	-0.5728	-0.1682	-0.2141	-0.1128	***0.5234***

表 4 は，機械学習アルゴリズムとして RandomForest，使用する属性の組合せとして全属性，スコア統合方法に平均を用いた場合の，6 つの履歴書識別器から算出された実際のスコアである．太字イタリックで示しているスコアは，各行における最高得点（つまり，各アカウントに対して最も高いスコアを算出した履歴書識別器の値）である．対角線上のスコア（網掛けしているセル）が太字イタリックの場合は，アカウントと本人の履歴書が正しく照合されたことを示している．この場合，4 つのアカウントが正しく照合されている．

表 5 は，以上述べた 30 ケースについて，各被験者のアカウントが本人の履歴書に正しく照合された（アカウント所有者の履歴書識別器のスコアが 6 つの履歴書識別器のスコアの中で 1 位であった）アカウント数，2 番目に正しく照合された（スコアが 6 人中 2 位であった）アカウント数を表している．表 5 から，最も良い結果であった機械学習アルゴリズム，使用した属性，スコア統合方法の組合せは，

(a) ロジスティック回帰，全属性，平均
(b) ロジスティック回帰，4 つの属性，平均
(c) ロジスティック回帰，全属性—現住所，積
(d) RandomForest，全属性，平均
(e) RandomForest，4 つの属性，積
(f) RandomForest，全属性—現住所，平均

の 6 ケースであった．そこで，これらの 6 ケースについての本評価を行った．

表 5 本人の履歴書が 1 位，2 位になった数
Table 5 Number of times when corresponding resumes were ranked top and second.

機械学習アルゴリズム	使用した属性	統合方法	1位の数	2位の数
線形SVM	全ての属性	積	3	1
		平均	3	1
	4つの属性（学歴，得意科目，趣味，性別）	積	3	1
		平均	3	1
	全属性 - 現住所	積	2	2
		平均	4	0
非線形SVM	全ての属性	積	0	2
		平均	0	0
	4つの属性（学歴，得意科目，趣味，性別）	積	0	2
		平均	0	1
	全属性 - 現住所	積	0	1
		平均	0	1
ロジスティック回帰	全ての属性	積	4	0
		平均	***4***	***1***
	4つの属性（学歴，得意科目，趣味，性別）	積	4	0
		平均	***4***	***1***
	全属性 - 現住所	積	***4***	***1***
		平均	4	0
ナイーブベイズ	全ての属性	積	2	2
		平均	0	2
	4つの属性（学歴，得意科目，趣味，性別）	積	2	1
		平均	0	1
	全属性 - 現住所	積	3	1
		平均	0	2
RandomForest	全ての属性	積	2	1
		平均	***4***	***1***
	4つの属性（学歴，得意科目，趣味，性別）	積	***4***	***1***
		平均	3	1
	全属性 - 現住所	積	3	1
		平均	***4***	***1***

6. 本評価

6.1 結果

30 人の被験者の履歴書に含まれる属性値 119 個に対して，RandomForest とロジスティック回帰を用いて 238 個の属性値識別器を実装し，これらを用いて予備評価で有効であった 6 つのスコア統合ケースに対して本評価を行った．**図 3** は，ケース (e) における本評価の結果である．横軸は被験者のアカウント番号を表し，縦軸は各履歴書識別器によって算出された当該アカウントのスコア分布を箱ヒゲ図によって表している．また，表中の記号「●」，「▲」，「■」，「×」は当該アカウント本人の履歴書識別器から算出したスコアを表しており，「●」＝本人のスコアが 1 位である（アカウントが被験者の履歴書と正しく照合された）場合，「▲」＝本人のスコアが上位 10%（上位 3 位）以内の順位である場合，「■」＝本人のスコアが上位 20%（上位 6 位）以内の順位である場合，「×」＝それ以外の場合を示している．たとえば，被験者 1 のアカウントにおいて本人の履歴書識別器から算出したスコアは 30 人中 6 位以内の順位に相当する．

表 6 は，(a)～(f) それぞれのケースにおいて，本人の履歴書と正しく照合できたアカウントの数，上位 10%以内で照合できたアカウントの数，上位 20%以内で照合できたアカウントの数を示している．たとえば，ケース (e) では 5 件のアカウントが本人の履歴書と正しく照合され，14 件の

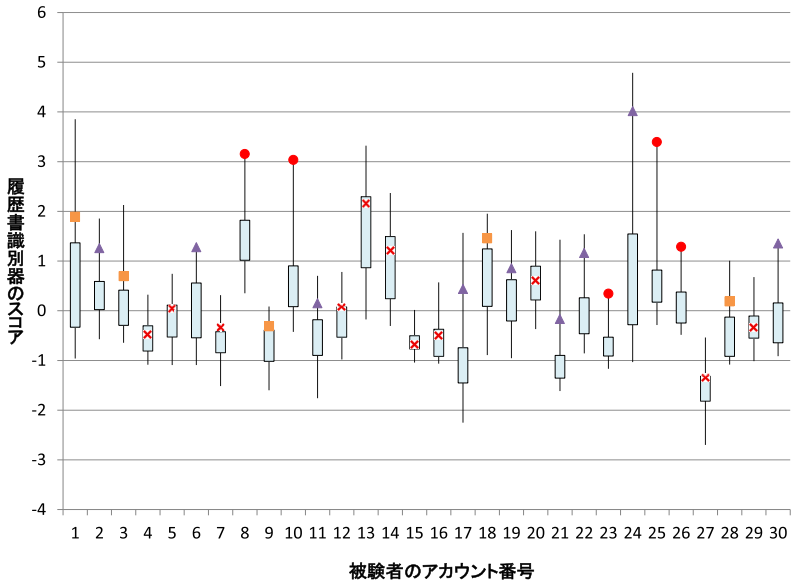
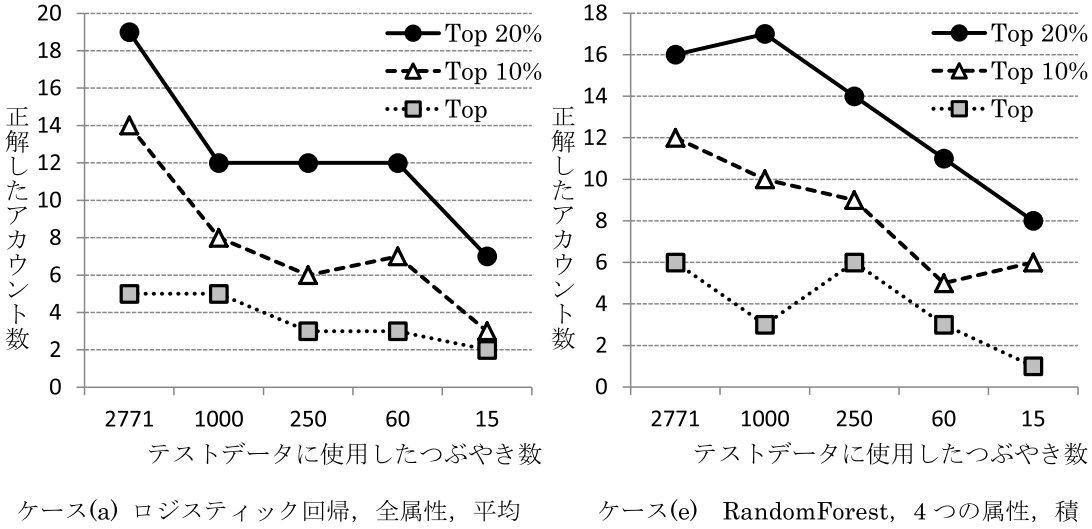


図 3 RandomForest，4 つの属性，積を用いた場合の各履歴書識別器のスコア分布
Fig. 3 Scores of resume classifiers with RandomForest, four attributes and product.

表 6 本人の履歴書が 1 位，上位 10%，上位 20%になったアカウントの数
Table 6 Number of accounts for which corresponding resumes are ranked top, in top 10%, and in top 20%.

ケース	機械学習アルゴリズム	使用した属性	統合方法	1位の数	上位10% (3位) 以内の数	上位20% (6位) 以内の数
(a)	ロジステッィク回帰	全属性	平均	6/30 (0.200)	12/30 (0.400)	16/30 (0.533)
(b)	ロジスティック回帰	4つの属性	平均	2/30 (0.067)	12/30 (0.400)	18/30 (0.600)
(c)	ロジスティック回帰	全属性 − 現住所	積	3/30 (0.100)	8/30 (0.267)	11/30 (0.367)
(d)	RandomForest	全属性	平均	3/30 (0.100)	10/30 (0.333)	15/30 (0.500)
(e)	RandomForest	4つの属性	積	5/30 (0.167)	14/30 (0.467)	19/30 (0.633)
(f)	RandomForest	全属性 − 現住所	平均	3/30 (0.100)	13/30 (0.433)	17/30 (0.567)



ケース(a) ロジスティック回帰，全属性，平均　　ケース(e) RandomForest，4 つの属性，積

図 4 少数のテストデータを用いた場合の照合精度の変化
Fig. 4 Accuracy of identification with small number of test data.

アカウント（5 件のアカウントを含む）が上位 10%以内，19 件のアカウントが上位 20%以内で照合されたことを示している．

図 4 は，(a)〜(f) のケースのうち，照合率が良かったケース (a) と (e) において，テストデータの数を減らした場合の性能の変化を折れ線グラフで示している．横軸は使用したテストデータ数，縦軸は照合できたアカウント数を示しており，●＝本人のアカウントが 1 位，▲＝上位 10%以内，■＝上位 20%以内で照合できたアカウント数である．これによると，多少の変動はあるものの，テストデータが少ないほど，照合精度が低下している．しかし，テストデータが 60 件までは，ランダム選択に比べて優位な精度で照合できている．

6.2 分析

本評価において，被験者 10 および被験者 25 のアカウントの照合率は全ケースで良好であった．理由として，これらのアカウントから投稿されたつぶやきには，得意科目や趣味といった，本人の履歴書の属性値に直接関連する単語が含まれていたことがあげられる．被験者 10 の履歴書には「得意科目＝微積分」が含まれており，当人のつぶやきにも「･･･「微積分の考え方」ですもんね．他の関数出てこないし」「そんなことより偏微分方程式やろうぜ！」など，得意科目に関連するフレーズが多く見られた．

次に，被験者 13，被験者 16，被験者 22 の場合は，ケース (e) では本人の履歴書がそれぞれ 10 位，12 位，3 位であるのに対し，ケース (a) では 5 位，5 位，1 位である．

一方で，被験者 6，被験者 21 の場合は，ケース (a) では 8 位，14 位であるのに対し，ケース (e) では 2 位，3 位であり，手法によって精度に大きな差が見られる．これにより，複数の手法から算出されたスコアを効果的に統合することで，履歴書識別器のさらなる精度向上が期待できると考えられる．

また，被験者 7 と被験者 14 の照合率はすべてのケース

で悪かった．被験者のつぶやきには，「おはようございます」「よるほー」などの挨拶や，1 単語のみで構成される発言（たとえば，「眠い」「帰ろ ···」など）が多く見られた．これらのつぶやきは本人の履歴書とまったく関係がなく，このような場合，履歴書から識別器を作成している本提案手法では照合が不可能である．被験者 14 は履歴書に「趣味＝音楽」と記載しており，つぶやき中にも音楽や歌手に関する単語が見られる．しかし，被験者 14 が興味を持っている楽曲および歌手は一般的に知名度が低く，訓練データとして収集した「趣味＝音楽」であるアカウント 30 件のつぶやきの中に，これらに関するつぶやきが含まれていなかった．このような場合，つぶやき中に現れる単語そのものではなく，単語から関連付けられるカテゴリを学習するなど，抽象化された手法をとる必要があると考えられる．

入手した 30 件の履歴書の全属性値の数は 169 個であったが，50 個の属性値に対しては十分な数の正例を取得できなかったため，本評価では 119 個の属性値識別器を用いた．しかし，今後 SNS のアカウントが増加し，作成不可能であった 50 個の属性値（たとえば，「趣味＝社会問題を考えること」，「資格＝ファイナンシャルプランナー 3 級」など）に対しても正例の収集が可能になれば，さらに照合精度を向上させることができると考える．

7. 本論文の貢献

7.1 新規性

3.1 節で述べたように，SNS の投稿文と履歴書は性質が異なるため，機械学習の適用方法が明らかではなかった．また，何を訓練データにするか明らかではなく，訓練データが明らかになったとしても，その収集が現実的には困難となる可能性があった．そのため，投稿文と履歴書という異種文書間の照合を，機械学習の問題に帰着させること，学習のための訓練データ（正例，負例）の収集を現実的に可能にすることが課題となった．

この課題に対して，提案手法では，履歴書が複数の属性値（たとえば性別＝女性，住所＝調布，趣味＝ダンスなど）の集合であることに着目した．そして，投稿文と履歴書の照合問題を，「投稿文が，履歴書内の全属性値を有する人物の書いた文章であるか」という一種の文書分類問題に帰着させた．この方針に沿って，図 1 に示したように，履歴書内の属性値ごとに，女性の投稿した文章の識別器，調布在住者の投稿した文章の識別器を設けた．その結果，訓練データは，女性と男性の投稿文，調布在住者と非在住者の投稿文となり，検索 API ツイプロによる自動収集が可能になった．履歴書内の全属性値についてスコアを算出し，その統合値を「投稿文が，履歴書内の全属性値を有する人物の書いた文章である」度合いとした．これをすべての履歴書について実施し，統合スコアの最も高い履歴書を投稿文の著者とした．このように，提案手法の新規性は，投稿文と履歴書という異種文書間の照合問題を，訓練データの収集が容易な文書分類問題に帰着させたことにある．

7.2 有用性

(1) 照合精度

5 章および 6 章では，提案手法の基本的な性質を評価したので，機械学習アルゴリズムは基本的なものを用い，属性値識別器のスコアを統合する方法も，単純な平均と積を用いた．しかし，より新しい機械学習アルゴリズムである GBDT（Gradient Boosting Decision Tree）を用いたところ，30 人中本人の履歴書と正しく照合されたアカウントが 11 人，上位 10%以内で照合できたアカウントが 16 人，20%以内が 23 人となった．なお，6 章の評価と同一のサンプルデータを用い，特徴量は bag-of-words，全属性のスコアを平均によって統合した．このケースのスコア分布を**図 5** に示す．「●」，「▲」，「■」，「×」の意味は図 3 の場合と同様である．

6 章の評価の最良ケースは，正しい照合が 5 人，10%以内が 14 人，20%以内が 19 人（あるいは 6 人，12 人，16 人）だったので，機械学習アルゴリズムを変更しただけで精度が大幅に向上した．機械学習アルゴリズムは急速な発展を続けているので，さらなる精度向上の可能性がある．またスコア統合方法についても，アダブーストなどの機械学習の手法を用いることで精度の向上が可能と考える．

(2) スケーラビリティ

本論文の評価では，30 件の履歴書に含まれる 119 の各属性値について，正例となる 10〜30 の Twitter アカウントと，負例となる 30 の Twitter アカウントを収集した．各アカウントのつぶやきを最大 3,000 件までダウンロードし，単語を抽出して特徴ベクトルを生成し，機械学習により識別器を生成した．これらの処理時間を**表 7** に示す．なお，測定に用いたハードウェアは，Intel Core i7，4 コア，3.4 GHz，16 GB メモリの PC1 台であり，ネットワークは 1 Gbps の有線 LAN である．

表 7 の 1 から 3 の処理時間は属性値の総数に比例するが，属性値の総数は，たかだか，履歴書の数に比例する（履歴書間で属性値の重複があるため，履歴書数が 2 倍になっても属性値数は 2 倍にはならない）．そこで，処理時間はたかだか履歴書の数に比例すると考え，表 7 の処理時間から推定すると，SNS 上で問題発言が見つかってから，3 日の間に，530 件の履歴書に対する識別器を生成することができる．また，上記の処理は属性値ごとに並列化できるので，複数台の PC の利用により，さらに多くの識別器を生成可能である．このように提案手法はある程度のスケーラビリティを有する．

なお，提案手法は，小規模な実施でも有用な場合がある．たとえば，内部告発をしたと思われる社員の候補者 50 人の履歴書について識別器を生成する場合もありうる．

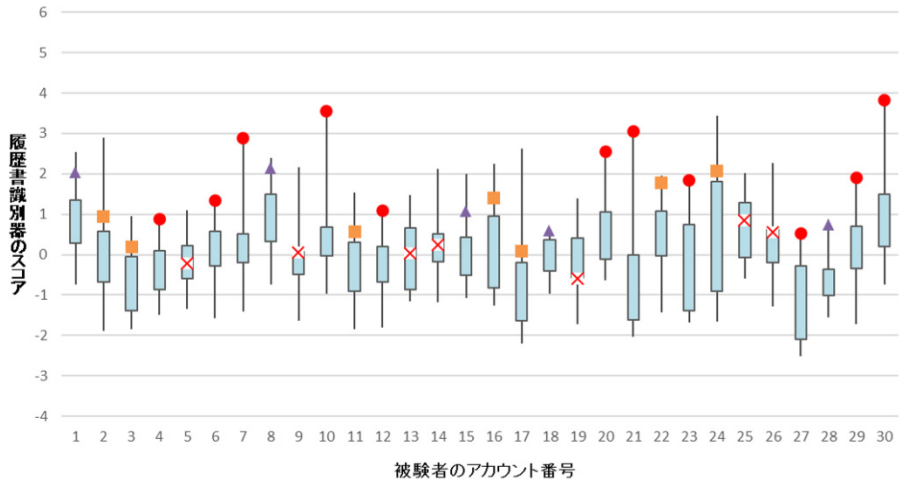


図 5 BDT，全属性，平均を用いた場合の各履歴書識別器のスコア分布
Fig. 5 Scores of resume classifiers with GBDT, all attributes, and average.

表 7 提案手法の処理時間の内訳
Table 7 Processing time of each part of proposed method.

No.	処理	時間
1	属性値毎に正例・負例のアカウントを収集し，各アカウントから最大 3000 件までのつぶやきをダウンロード	1 時間 39 分 25 秒
2	属性値識別器の訓練データとして，ダウンロードしたつぶやきから特徴ベクトルを生成（特徴量は bag-of-words）	1 時間 41 分 11 秒
3	特徴ベクトルを用いて，機械学習により属性値識別器を生成（アルゴリズムは GBDT）	41 分 40 秒
	合計	4 時間 2 分 16 秒

(3) 有用性

提案手法は，従来手法ではほとんど扱われなかった，より大きなプライバシリスク（SNS のアカウントと個人の直接照合）を明らかにすることができる．また，SNS の不正利用者の特定という別目的にも利用できる．

検知したい SNS アカウントが，履歴書の該当者のアカウントでない場合には，照合するべき履歴書が存在しないので，提案手法を適用できないという制約がある．しかし，企業や省庁，自治体，教育機関などの多くの組織は，構成員の履歴書を保有しているので，提案手法の仮定（SNS アカウントが履歴書の該当者のアカウントであること）は一定の現実性があると考える．たとえば，企業の社員による内部告発に対して，その告発の内容などから，社員であることが推定された場合には，告発のあったアカウントを社員の履歴書と照合することで，告発者を絞り込むことができる．

提案手法では，特定された履歴書の人物をただちに処罰するといった利用法は想定していない．提案手法によって人物を絞り込んだうえで，人手調査を行うことを想定している．そのため，提案手法の誤検知は，致命的な問題にはならないと考える．

このように，提案手法は，従来手法より大きなプライバシリスクを明らかにすることが可能で，一定のスケーラビリティを有し，精度についても今後の向上が期待できることから，一定の有用性を備えていると考える．

8. おわりに

SNS のプライバシリスクを示すために，匿名アカウントと企業などの組織が保有する履歴書を機械学習によって照合する手法を提案した．履歴書の属性値ごとに識別器を作成し，それぞれの識別器から出力されたスコアを統合することにより履歴書ごとの識別器を構成した．また，ツイプロ（Twitter アカウントのプロフィール検索ツール）によって，プロフィールに当該属性値を含むアカウントと含まないアカウントを自動的に収集し，それぞれを正例，負例とすることで訓練データの収集を容易化した．

被験者 30 人のアカウントと履歴書をサンプルとして提案方式を評価した．予備評価では，6 人のサンプルを用いて，投稿文から抽出する特徴量として bag-of-words と binary，機械学習アルゴリズムとして線形 SVM，非線形 SVM，ロジスティック回帰，ナイーブベイズ，RandomForest，着目すべき属性として性別，現住所，帰省先住所，学歴，得意科目，趣味，資格，属性ごとのスコアの等号方法として積と平均を評価した．その結果，bag-of-words，RandomForest，4 つの属性（学歴，得意科目，趣味，性別），平均によるスコア統合の組合せをはじめとする 6 通りのパラメータが効果的であった．本評価では，30 人の被験者の履歴書に含まれる属性値 119 個に対して 238 個の個の属性値識別器を実装し，これらを用いて予備評価で有効であった 6 つのパラメータについて評価を行った．その結果，最良のケースで 30 人中本人の履歴書と正しく照合されたアカウントが 5 件，本人の履歴書との照合結果が 10%以内のアカウント

が 14 件，20%以内のアカウントが 19 件であった．今後の課題としては以下があげられる．

(1) 本研究では機械学習アルゴリズムとして基本的なものを評価したが，ディープラーニングやディープフォレストなど，より新しい機械学習アルゴリズムを評価する．

(2) 単純な平均や積の代わりに，機械学習アルゴリズムによるスコア統合方式を検討する．履歴書識別器を構成する際には，何百もの属性値識別器から出力されたスコアを統合する必要があるため，アダブーストなどの手法により最適な統合を検討する．

(3) 本研究では，電気通信大学の学生 30 人を被験者として評価を行ったが，より多数の被験者，属性の異なる被検者による評価を行う．

(4) 図 1 の構成により，属性値識別器として writing-style に基づく識別技術（たとえば文献 [13]）を用いることが可能である．そこで writing-style の利用および bag-of-words などとの併用における照合精度を評価する．

謝辞 本研究の一部は，科学研究員補助金「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」の支援により推進しました．

参考文献

[1] Goga, O., Lei, H., et al.: On exploiting innocuous user activity for correlating accounts across social network sites, ICSI Technical Reports - University of Berkeley (2012). available from ⟨http://www.icsi.berkeley.edu/pubs/techreports/TR-12-008.pdf⟩(accessed 2016-02-20).

[2] Almishari, M., Kaafar, M., et al.: Stylometric Linkability of Tweets, *Proc. 13th Workshop on Privacy in the Electronic Society*, pp.205–208 (2014).

[3] Narayanan, A., Paskov, H., et al.: On the feasibility of Internet-scale author identification, *Proc. 33rd IEEE Symposium on Security and Privacy*, pp.300–314 (2012).

[4] Backstrom, R., Dwork, C. and Kleinberg, J.: Wherefore art thou R3579X? anonymized social networks, hidden patterns, and structural steganography, *Proc. 16th International World Wide Web Conference*, pp.181–190 (2007).

[5] Lam, I., Chen, K. and Chen, L.: Involuntary information leakage in social network services, *Proc. 3rd International Workshop on Security*, LNCS 5312, pp.167–183 (2008).

[6] Mao, H., Shuai, X. and Kapadia, A.: Loose Tweets: An analysis of privacy leaks on Twitter, *Proc. 10th ACM Workshop on Privacy in the Electronic Society* (2011).

[7] Kótyuk, G. and Buttyan, L.: A Machine learning based approach for predicting undisclosed attributes in social networks, *Proc. IEEE 4th International Work-shop on Security and Social Networking*, pp.361–366 (2012).

[8] Caliskan-Islam, A., Walsh, J. and Greenstadt, R.: Privacy detective: Detecting private information and collective privacy behavior in a large social network, *Proc. 13th Workshop on Privacy in the Electronic Society*, pp.35–46 (2014).

[9] Narayanan, A. and Shmatikov, V.: De-anonymizing social networks, *Proc. 30th IEEE Security & Privacy*, pp.173–187 (2009).

[10] Polakis, I., Kontaxis, G., et al.: Using social networks to harvest email addresses, *Proc. 9th ACM Workshop on Privacy in Electronic Society*, pp.11–20 (2010).

[11] Hart, M., Manadhata, P. and Johnson, R.: Text classification for data loss prevention, *Proc. 11th Privacy Enhancing Technologies Symposium*, pp.18–37 (2011).

[12] ツイプロ：Twitter プロフィール検索，入手先 ⟨http://twpro.jp/⟩（参照 2017-02-22）．

[13] Corney, M., Vel, O.D., Anderson, A. and Mohay, G.: Gender-preferential text mining of e-mail discourse, *Proc. 18th Annual Computer Security Applications Conference*, pp.282–289 (2002).

橋本 英奈

2015 年電気通信大学情報理工学部総合情報学科卒業．2017 年同大学大学院情報理工学研究科総合情報学専攻博士前期課程修了．現在，（株）NTT ドコモ勤務．

宮崎 夏美

2017 年電気通信大学情報理工学部総合情報学科卒業．同大学大学院情報理工学研究科情報学専攻博士前期課程在学中．

市野 将嗣 （正会員）

2003 年早稲田大学理工学部電子・情報通信学科卒業．2008 年同大学大学院理工学研究科博士課程修了．2007 年日本学術振興会特別研究員．2009 年早稲田大学大学院基幹理工学研究科研究助手．2010 年同大学メディアネットワークセンター助手．2011 年電気通信大学大学院情報理工学研究科助教．2016 年同大学院情報理工学研究科准教授．バイオメトリクス，ネットワークセキュリティに関する研究に従事．博士（工学）．電子情報通信学会会員．

久保山 哲二 （正会員）

1992 年九州大学工学部情報工学科卒業．1994 年同大学大学院システム情報科学研究科修士課程修了．1997 年より東京大学国際・産学共同研究センター助手，2007 年同助教．博士（工学）．2008 年学習院大学計算機センター准教授，2013 年より同教授．2009 年ベルン大学コンピュータ科学・応用数学研究所客員研究員（JSPS 特定国派遣・短期）．現在，離散データ構造を対象とした機械学習アルゴリズムの研究に従事．人工知能学会，電子情報通信学会各会員．

越前 功 （正会員）

1997 年東京工業大学大学院理工学研究科修士課程修了（応用物理学）．日立製作所システム開発研究所を経て，現在，国立情報学研究所所長補佐，同研究所情報社会相関研究系研究主幹・教授．総合研究大学院大学複合科学研究科情報学専攻教授（併任）．メディアセキュリティ，メディア情報処理の教育研究に従事．2010 年ドイツ・フライブルク大学客員教授，2011 年ドイツ・マルティン・ルター大学客員教授，2017 年より津田塾大学客員教授．情報セキュリティ文化賞（2016 年），ドコモ・モバイル・サイエンス賞（2014 年），情報処理学会論文賞（2014 年，2005 年），情報処理学会長尾真記念特別賞（2011 年）等受賞．Member, Information Forensics and Security Technical Committee, IEEE Signal Processing Society，電子情報通信学会フェロー．博士（工学）（東京工業大学）．

吉浦 裕 （正会員）

1981 年東京大学理学部情報科学科卒業．日立製作所を経て，2003 年より電気通信大学勤務．現在，同大学大学院情報理工学研究科教授．情報セキュリティ，プライバシ保護の研究に従事．博士（理学）．日立製作所社長技術賞（2000 年），情報処理学会論文賞（2005 年，2011 年），システム制御情報学会産業技術賞（2005 年），IEEE IIH-MSP best paper award（2006 年），日本セキュリティ・マネジメント学会論文賞（2010 年，2016 年），IFIP I3E best paper award（2016 年）等受賞．電子情報通信学会，日本セキュリティ・マネジメント学会，人工知能学会，システム制御情報学会，IEEE 各会員．本会フェロー．