# 多元情報のプロファイリングに基づく
# ソーシャルメディアからの個人の再特定

橋本 英奈
電気通信大学大学院 情報理工学研究科
博士（工学）の学位申請論文

2022 年 3 月

多元情報のプロファイリングに基づく

ソーシャルメディアからの個人の再特定

博士論文審査委員会

主査　　市野 将嗣 准教授

委員　　崎山 一男 教授

委員　　岩本 貢　教授

委員　　松本 光春 准教授

外部審査委員 京都橘大学　吉浦 裕　教授

# 和文要旨

ソーシャルメディアは重要なコミュニケーション基盤であるが，様々なプライバシー問題を引き起こしている．ソーシャルメディアのプライバシーを保護するために匿名化が多用されているが，匿名アカウントから個人が推定される再特定が問題となっている．そこで，本研究では，再特定の手法とその特定精度を明らかにすることで，ソーシャルメディアのプライバシーリスクを明らかにする．

匿名アカウントからの再特定の従来手法は，友人やフォロー・フォロワー等のリンク構造の照合に基づいて匿名アカウントを同一ユーザの実名アカウントに紐づける手法および，実名アカウントの投稿文から文章特徴を学習し，類似の文章の匿名アカウントに紐づける手法に分類できる．しかし，これらの手法は，再特定の対象となる匿名アカウント以外に同一ユーザの実名アカウントが開示されていることを前提としている．また，匿名アカウントのユーザの趣味はラグビー観戦であるといったヒント情報を入手しても，その情報を利用することができない．さらに，構造の照合に基づく手法は，匿名アカウントと実名アカウントの個数がほぼ等しく，両者が少なすぎないという条件を必要とする．以上の前提条件および問題点により，従来手法は実用性が低く，再特定のリスクを充分に示すことができなかった．一方，ソーシャルメディアアカウントのユーザの属性（性別，年代，趣味等）を推定するプロファイリングは再特定と補完関係にあるにも関わらず，プロファイリングと再特定は独立に研究されてきた．

先行研究の分析を踏まえ，本研究では，機械学習を用いた再特定においてプロファイリングを活用することで，従来とは異なる前提条件で動作し，従来手法とは補完関係となる手法を確立する．また，ヒント情報を有効活用可能とし，ソーシャルメディアアカウントの持つ投稿文，投稿画像，リンク等の様々なデータを利用することで，再特定精度の向上を可能とすることを課題とする．

提案法は，再特定候補者のプロファイルを入手することを前提とする．匿名アカウントの未知のユーザをプロファイリングし，未知ユーザのプロファイルと候補者のプロファイルを照合することで，未知ユーザと候補者が同一人物である確からしさを定量化する．すなわち，候補者のプロファイルに含まれる各属性値（たとえば趣味がピアノ演奏）について，匿名アカウントのユーザが当該属性値を有する確からしさを推定し，これらの属性値毎の確からしさを統合することで，匿名ユーザと候補者を紐づける．提案法は，匿名アカウントを別のアカウントと照合しないので，匿名アカウント以外に同一ユーザの実名アカウントを必要としない．また，再特定対象者に関するヒント情報（たとえばラグビー観戦が趣味）を属性値とみなすことで，提案手法の中で自然に利用することができる．

匿名アカウントのデータのうち投稿文を用い，再特定候補者のプロファイルとして履歴書を用いて，提案手法を実現した．匿名アカウントのユーザが候補者プロファイルの属性値を有するかのクラス分類器を機械学習により生成した．予備評価により，機械学習アルゴリズムには XGBoost，特徴量には Bag-of-words，スコア統合方法には平均を用いた組み合わせが最良であることを明らかにした上で，78 人の被験者データを用いた評価を

実施し，匿名アカウントから同一人物の履歴書を特定する問題，履歴書から同一人物の匿名アカウントを特定する問題，アカウントと同一人物の履歴書を 1 対 1 に紐づける問題において，各々48%，50%，55%の正解率を達成した．

次に，匿名アカウントのデータのうち投稿文中の地名を緯度経度に変換して移動履歴とした上で，現住所をプロファイリングした．上記の 78 人の被験者データを用いて住所をプロファイリングした結果，投稿文のみの正解率は 65%，移動履歴のみは 74%，投稿文と移動履歴のスコアレベル統合方式は 70%であり，移動履歴を用いることで，住所のプロファイリング精度の向上が可能である．さらに，移動履歴のみ，投稿文と移動履歴のスコアレベル統合の 2 方式によって現住所をプロファイリングし，そのスコアと現住所以外のスコアを統合することで，78 人のアカウントから個人を再特定した結果，投稿文のみを用いる方式に比べ，移動履歴を併用することで約 5%の正解率向上が見られた．

さらに，匿名アカウントのデータのうち投稿画像を用いることで，15 種類の趣味について，当該趣味を有することの推定精度を向上した．この趣味のプロファイリングと投稿文による他の属性値のプロファイリングとを統合することで，個人再特定の正解率を約 2%向上することができた．また，投稿文，移動履歴，投稿画像の全てを併用することで個人特定の精度をさらに向上できることを明らかにした．加えて，リンク情報を用いた性別および教育種別のプロファイリングが可能であることを評価実験により明らかにし，リンク情報の併用により，個人特定精度をさらに向上できる可能性を示した．

以上を通じて，プロファイリングに基づくソーシャルメディアからの個人の再特定手法を確立し，その再特定精度と拡張性を明らかにした．

## Abstract

Although social media are important communication infrastructures, they cause serious privacy problems. Although social media accounts are often anonymized to protect the users' privacy, there is a problem called "re-identification" in that the users of the anonymized accounts are identified. In this thesis, we clarify the privacy risk of social media by establishing an effective re-identification method and evaluating its precision.

Conventional methods of re-identifying anonymous social media accounts can be classified into two types. Methods in the first type map the set of anonymous accounts to the set of known accounts based on structural matching between the two social graphs represented by the two sets of accounts. Those in the second type map anonymous accounts to known accounts based on the similarity between writing styles of the anonymous and known accounts, where machine learning was used to measure the similarity. However, these two types of methods depend on the assumption that the attacker can use known accounts of the candidate of persons of re-identification. They cannot use hint information (e.g. the unknown user of an account having a hobbit of watching rugby games) to improve precision. The methods base on graph matching further depends on an assumption that the numbers of the anonymous and the known accounts are almost the same and they are not too small. Thus, conventional methos are not practical enough to clarify the privacy risk. On the other hand, although profiling social media accounts that estimates attributes (e.g. age, gender, and hobbits) of the account users could complement re-identifying the accounts, profiling and re-identification have been studied independently.

Based on the analysis of previous research, we aim to establish a new re-identification method that is based on an assumption different from the assumptions of the conventional methods and that can complement the conventional methods. We try to enable using hint information and enable using various kinds of data such as sentences, images, and links on social media accounts to improve re-identification precision.

Our proposed method assumes that the attacker knows candidates of persons to re-identify and knows their profiles. It profiles the unknown owners of anonymous accounts and compares the estimated profiles with the known candidate profiles to quantify degree of the unknown owner of each anonymous account being the same as each candidate person. The degree of the sameness is quantified by fusing degrees of the unknown owner having attribute values (e.g. having hobbit of playing piano) that the candidate profile has. The proposed method does not use

known social media accounts of the candidate persons. The method can make use of hint information (e.g. the unknown owner having hobbit of watching rugby games) by taking it as one of the attribute values. The proposed method thus solves the problems of the conventional methods.

We first implement the proposed method by using sentences posted on anonymous accounts to profile the unknown owners and using resumes of candidate persons as the candidate profiles. Machine learning is used to generate classifiers for quantifying the degrees of the unknown owners having the attribute values described in the candidate's resumes. XGboost for machine-learning algorithm, bag-of words for feature representation, and averaging for score fusion were chosen as the best combination through preliminary experimentation. In evaluation using accounts and resumes of 78 volunteers, precisions in linking anonymous accounts to resumes of the same persons, linking anonymous accounts and resumes in one-to-one mode, and linking resumes to accounts are 48%, 50%, and 55% respectively.

We next extended our method by using geographic information (i.e. latitude and longitude) extracted from place names in sentences posted on anonymous accounts. Addresses of the 78 volunteers were estimated by the geographic information extracted. Precisions of address estimation by using sentences (bag-of-words) only, the geographic information only, and both sentence and geographic information by concatenating the bag-of-words vector and the geographic vector are 65%, 74%, and 70%. Thus, using geographic information improved the precision. Degrees of the unknown owners having the estimated addresses were combined with degrees of the unknown owners having other attribute values, which had been estimated from the sentences, to re-identify the accounts. The experimentation clarified that using geographical information improves re-identification precision by 5%.

We also improve the precision of estimating an unknown owner having each of 15 hobbits by using images posted on the owner's account. Estimation of the owners having 15 hobbits is combined with other estimation based on sentences to improve the re-identification precision by 2%. Our experimentation clarified that using sentences, geographic information, and images all together could further improve the re-identification precision. Our experimentation also demonstrated that link information could be used to estimate gender and education type of the unknown owners and thus could be used to improve the re-identification precision.

Our research described in this thesis thus established a method of re-identifying anonymous social media accounts by using profiling that can solve the problems of the conventional methods as well as having shown its re-identification precision and extensibility.

# 目次

# 第1章　序論

## 1.1　ソーシャルメディアのプライバシー問題

人の関係を情報ネットワーク上で構築し，コメントやつぶやきなどのテキスト情報，画像，動画，位置情報，「いいね」などの評価情報を流通させるソーシャルメディアの利用および社会における役割が拡大している．例えば，世界最大規模のソーシャルネットワークサービス（以下，ソーシャルメディア）である Facebook の月間アクティブユーザ数（以下，MAU）は 23.7 億人，Twitter の MAU は 3.5 億人，写真掲載を中心とする Instagram の MAU は 10 億人となっている [1]．また，インスタントメッセンジャーアプリとして普及した LINE は，国内登録ユーザが 8600 万人を超え，日本の人口の 68% 以上をカバーしたと発表している [2]．また，2020 年より，招待制の音声配信ソーシャルメディアとしてサービスを開始した clubhouse は，世界登録ユーザ数が 200 万人となっている [3]．

サービスの拡大と共に，ソーシャルメディアの利用目的も広がっている．友人とのコミュニケーションや情報交換だけでなく，企業内コミュニケーション，就職活動や転職活動，マーケティングやブランド形成，さらには政治や民主化運動にも活用されている．また，新聞やテレビなどのマスメディアの利用が減少するに伴って，ソーシャルメディアがニュースの流通を担うようになってきた．たとえば，米国では，43.1%の人がソーシャルメディアからニュースを取得している [4]．

このように，ソーシャルメディアは重要なコミュニケーションインフラとなっているが，一方では，プライバシー情報や組織の機密情報の漏洩，誹謗中傷などの多くの問題を引き起こしている．また，近年では，ソーシャルメディア上のフェイクニュースにより，死亡事件が発生したり [5]，選挙の結果に影響を与えており [6]，大きな問題になっている．

本研究では，これらのソーシャルメディアの問題のうちプライバシー問題を取り上げる．ソーシャルメディアが原因となったプライバシー問題の事例としては， Twitter に旅行中であることを示唆する書き込みをしたユーザが空き巣の被害に遭ったという事件があった [7]．この事例ではユーザが居場所を絶え間なくソーシャルメディアに投稿していたり，位置情報付きの投稿を行っていたため，泥棒が犯行におよぶことができた．また，ソーシャルメディアの投稿からストーカーが住所や生活ぶりをさぐっていた事件 [8]も起きている．さらに，GPS 情報を付与した投稿から，自宅住所を割り出すサービス [9]や，外出先からの投稿から，自宅が不在状態であることを検知するサービス [10]もあり，英

国の空き巣の 8 割がソーシャルメディアを活用して犯行の事前準備をしているという調査もある [11].

ソーシャルメディア上のプライバシー情報を保護するために，サービスによっては，ユーザによる公開範囲の設定や，事業者による不適切な書き込みのチェックが行われている．しかし，公開範囲の設定はユーザの負担となり，情報発信のたびに適切な設定を行うことは難しい上，ソーシャルメディアの本来の目的である「コミュニケーションの楽しみ」を損なう．また，事業者によるチェックはコストや人手および通信の自由との関わりもあるため限界がある．

ユーザが自分の身元を隠し，匿名のアカウントを用いることでプライバシーを保護することも多く行われている．特に，日本人は匿名性を好むとされており，Twitter ではソーシャルメディアの利用の 76.5%が匿名アカウントによる [12]．また，ソーシャルメディアのデータ（投稿文や投稿写真，リンク構造）は学術研究等の目的で利用されるが，これらのデータの多くは名前等を除去し，匿名化した上で開示される [13]．しかし，匿名データと同一人物の実名データとの照合を通じて，匿名データの対象者を推定可能であることが多くの研究によって指摘されており，匿名アカウントや匿名投稿文からの個人の特定が懸念されている．

## 1.2　先行研究の概要

ソーシャルメディアのプライバシーに関する先行研究は，(1)ソーシャルメディアのプライバシー問題に関する社会調査，(2)プライバシー情報の漏洩防止対策，(3)ソーシャルメディアから個人情報を抽出する攻撃手法の 3 つに分類することができる．(1)の社会調査の研究では，ソーシャルメディアユーザのプライバシーに関する意識，プライバシー情報の開示状況，漏洩事例等の調査分析を行っている [14] [15] [16] [17] [18] [19] [20] [21] [22] [23]．(2)の漏洩防止対策の研究では，公開範囲の適切な設定方法や自動的な設定方法等を提案・評価している [24] [25] [26] [27] [28] [29] [30] [31]．(3)の攻撃手法の研究は，個人情報漏洩のリスクを示し，事業者やユーザに対策を促すことを目的としている [32] [33] [34] [35] [36] [37] [38] [39] [40]．なかでも，近年は(3)の攻撃手法の研究が盛んである．

以下では，攻撃手法の研究について概要を述べる．攻撃手法の研究は，主に，匿名投稿文や匿名アカウントから投稿者やアカウントユーザを推定する再特定の研究 [32] [33] [34] [35] [36] [37] [38] [39] [40]，投稿者やアカウントユーザの属性（年齢，性別，趣味，思想信条，能力など）を推定するプロファイリングの研究 [41] [42] [43] [44] [45] [46] [47]に分類することができる．

# 第１章　序論

匿名アカウントが再特定されると，匿名と思い安心して書いた情報（例えば，所属している企業の利害に関わることや，友人の信条に対する批判など）にも関らず，投稿者の身元が明らかにされるため，投稿者が企業から不利な扱いを受けたり，投稿者の人間関係が悪化する可能性がある．このように匿名アカウントが再特定されることで，安心して投稿したはずの投稿から，投稿者の実世界における情報や，人間関係などの個人情報が漏洩してしまう可能性がある．再特定の研究は，ソーシャルメディアにおけるこのようなプライバシーリスクを示すために行われている．再特定の研究では，ヒューリスティックな手法 [32] [33] [34]，構造の照合に基づく手法 [35] [36] [37]，機械学習に基づく手法 [34] [38] [39] [40] [35]が提案されている．ヒューリスティックな手法は，研究者の経験や勘に頼った手法であり，再特定研究の初期に提案された．構造に基づく手法は，アカウント間のリンク関係等をグラフ等によって表現し，実世界の人間関係と照合することで個人を特定する．機械学習に基づく手法は，匿名アカウントの投稿文等からユーザの特性を学習し，別の実名のアカウントのユーザの特性と照合することで，匿名アカウントと実名アカウントのユーザが同一であることを推定する．近年は，機械学習に基づく再特定の研究が多くなっている．

機械学習に基づく再特定の研究は，再特定の対象となる匿名のアカウントあるいは匿名投稿に対して，同一人物の別のアカウントや投稿文を特定する．典型的な手法は，匿名アカウントに投稿された文章からアカウントユーザの文章の特徴を学習し，複数の実名のアカウントの中から文章の特徴が最も近い実名アカウントを特定する．このように匿名アカウントを実名アカウントと照合することにより，匿名アカウントのユーザを再特定する．

しかし，機械学習に基づく従来の再特定手法には，以下の前提条件および問題点がある．

(1) 再特定の対象者が 2 つのアカウントを持ち，攻撃者は両方のアカウントのデータ（投稿文など）を入手可能であることが前提となる．しかし，この前提は常に満たされるわけではない．なお 2 つのアカウントが両方共匿名である場合には，同一人物であることを推定できたとしても，個人を特定することにはならない．

(2) 再特定の対象者に関してヒントとなる情報を入手できる場合がある．たとえば，候補人物はラグビー観戦が趣味であると分かっている場合がある．しかし従来手法はこの情報を活かすことができない．

攻撃手法の研究には，再特定以外に，投稿者やアカウントユーザの属性（年齢，性別など）を推定するプロファイリングがある．プロファイリングができれば，投稿者やアカウントユーザの再特定をしやすくなると考えられる．また，投稿者やアカウントユーザが推定できれば，その属性を推定しやすくなると考えられる．このように，再特定とプロファイリングは相補関係にあると考えられる．しかし，従来，再特定とプロファイリングは

独立に研究され，互いを有機的に利用することはなかった．

## 1.3　本研究の目的と位置付け

本研究の目的は，上述した従来手法とは異なる前提条件で動作する手法を確立することである．本研究で提案する手法は，図 1.1 に示すように，再特定の候補となる複数の人物毎のプロファイルを用意する．匿名アカウントのユーザである未知の人物をプロファイリングし，そのプロファイルを推定した上で，複数の候補者のうち最もプロファイルの近い人物を選定することで，匿名アカウントのユーザを特定する．このように，提案手法は，プロファイリングを活用して再特定を行う．

なお，図 1.1 のように，ソーシャルメディアのアカウントでは，ユーザのプロファイルを部分的に記載していることがある．たとえば，同じ趣味の人にアピールする目的で，趣味を記載する場合などがある．ソーシャルメディアでは，これらのアカウントに記載したプロファイルを慣例として「プロフィール」と呼んでいる．そこで，本論文では，ソーシャルメディアにおけるユーザ情報を「プロフィール」，現実世界における個人の属性を「プロファイル」と呼び，区別することにする．

提案手法は，再特定の候補となる人物のプロファイルが入手可能であることを前提とする．一方，従来手法は，匿名アカウントのユーザの持つ別アカウントのデータが入手可能であること前提としていた．このように，提案手法は従来手法とは異なる条件を前提とするので，補完関係にある．提案手法の位置付けを図 1.2 に示す．

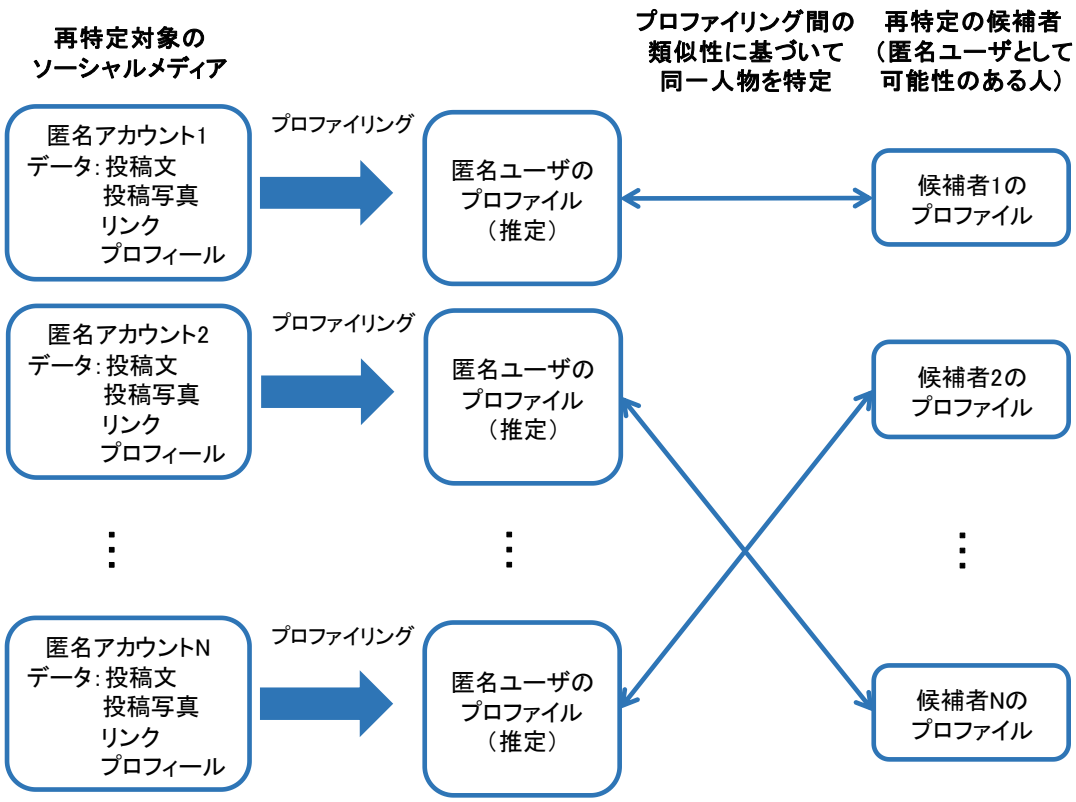


図 1.1　提案手法の基本方針

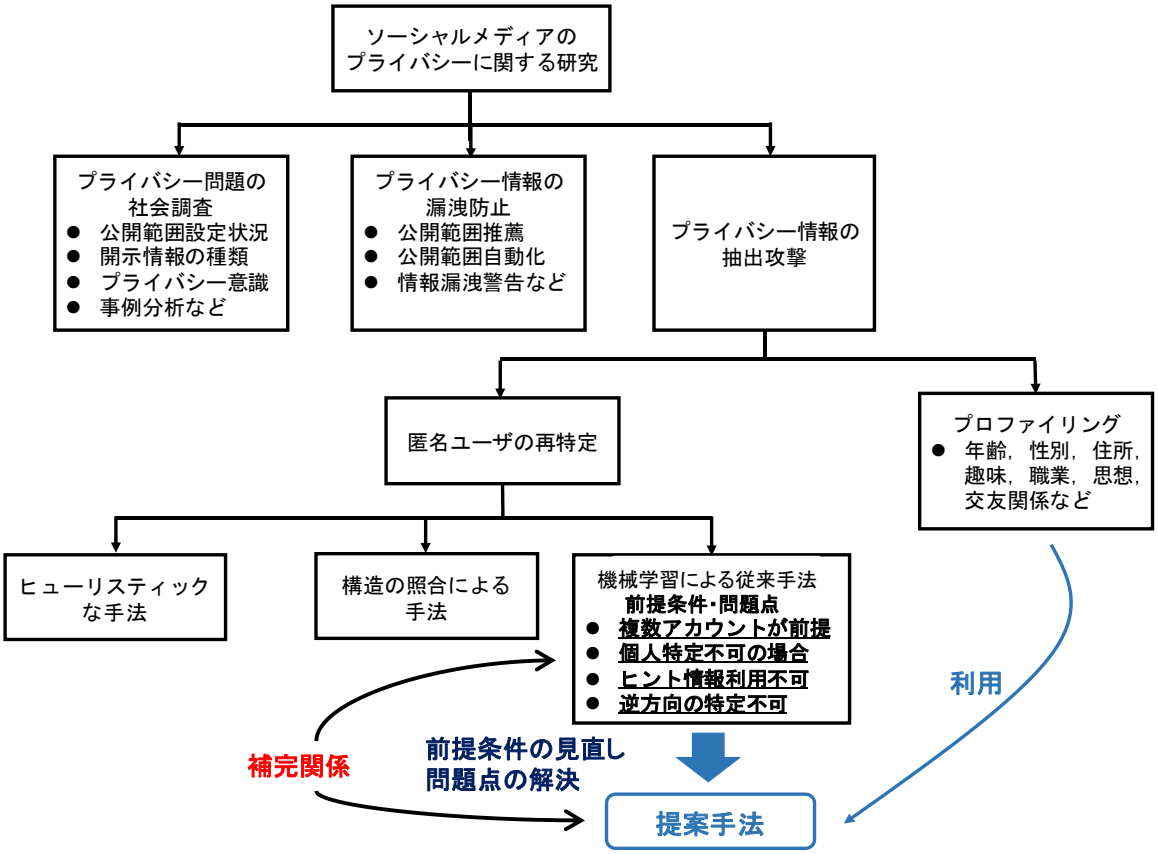


図 1.2　提案手法の位置付け

## 1.4　本研究の社会的位置付けと意義

再特定の研究は，再特定の危険性への警鐘や，個人情報漏洩防止対策の基礎として社会に活用されている．

再特定の危険性への警鐘については，再特定の研究成果に基づき，マスメディアが社会に警鐘を鳴らしている例がある．2018 年，英国のメディア Wired は，Twitter のメタデータを用いた再特定の研究 [76] を引用し，個人の身元特定に利用される危険性について警鐘を鳴らした [114]．この警鐘は Wired Japan を通じて日本でも広まった [115]．2018 年，NHK は，ソーシャルメディアの再特定の著者らの研究 [116] を引用し，就職活動時の企業によるプロファイリングの危険性について取り上げた [117]． 2019 年，BBC ジャパンは，カメラ映像からの個人の再特定の研究 [118] を引用し，ソーシャルメディアへの画像投稿の危険性について警鐘を鳴らした [119].

個人情報漏洩防止対策の基礎としては，再特定の研究成果に基づき，個人情報漏洩につながる語句の言い換えや，ユーザへの通知を行う技術の研究例がある．2017 年 He らは，2014 年に Gong らが提案したソーシャルネットワークによる個人特定の研究 [120] などを引用し，プライバシーと有用性を両立させるデータサニタイゼーション方法を提案した [28]．2012 年 Hoang らは，2005 年に Gross らが提案した匿名化されたユーザ名から実名を推定する研究 [14] などを引用し，場所や時間の粒度を拡大し上位語に置換する技術を提案した [121]．

本研究は，より高精度な再特定の可能性および，プロファイリングと連携した多面的なリスクを明らかにし，より効果的に社会に警鐘を鳴らし，漏洩防止技術につなげることを目的とする．

## 1.5　本論文の構成

本論文の 2 章では先行研究を分析し，その前提条件および問題点を明らかにする．3 章では，プロファイリングに基づく再特定手法を提案し，従来手法とは異なる前提条件で動作する点を明らかにし，問題点を解決できる見通しを示す．4 章では，プロファイリング対象のデータ（図 1.1 の左側）として投稿文を用い，照合先（図 1.1 の右側）として履歴書を用いた場合の提案手法の実現方法を述べ，実データを用いた有効性評価を行う．5 章では，プロファイリング対象のデータとして，投稿文に加え，投稿文から推定されるユーザの移動履歴を併用する場合の提案手法の実現方法と有効性評価を述べる（図 1.3）．6 章では，投稿文と投稿画像を併用する場合の提案手法の実現方法と有効性評価，および投稿文，移動履歴，投稿画像を全て用いる場合の実現方法と有効性評価を述べる．さらに，リ

ンク情報を併用する場合の予備実験について述べ，提案手法の拡張性を明らかにする．一方，提案手法の限界および従来手法との補完関係について述べる．7 章では，本研究の結論と今後の課題を述べる．

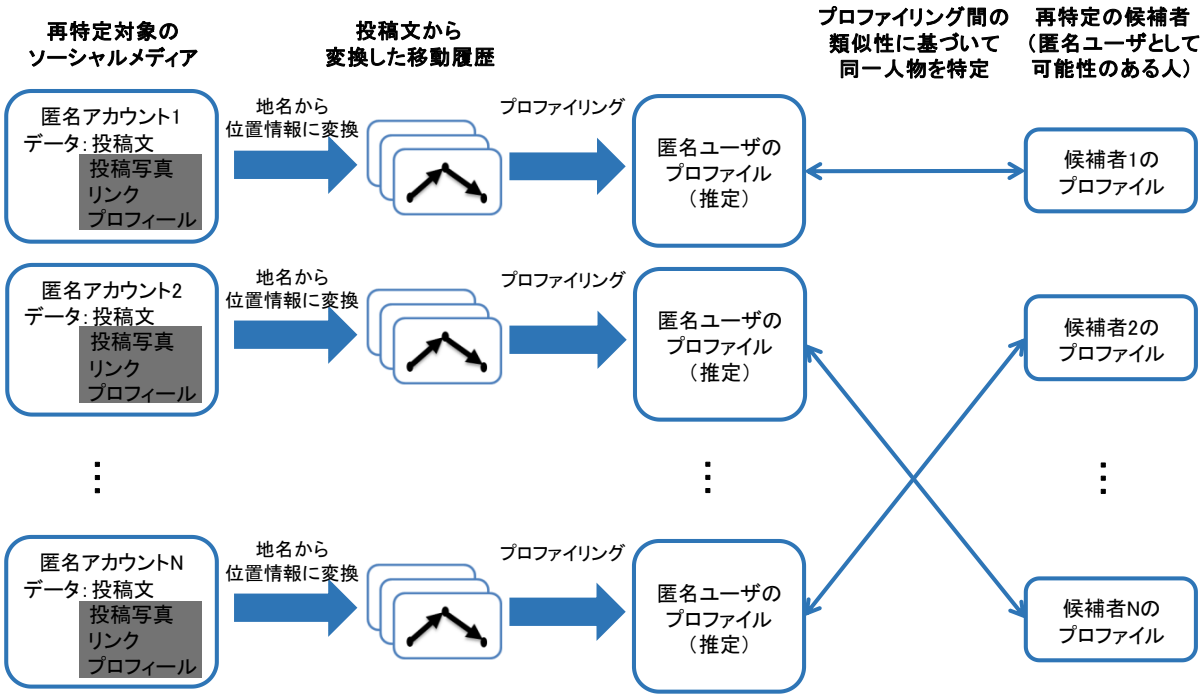


図 1.3　5 章における提案手法の基本方針

# 第2章　先行研究

## 2.1　はじめに

本章では，本研究に関わる先行研究を分析し，1 章で述べた先行研究の前提条件および問題点を明らかにすると共に，本研究の位置付けを明らかにする．まず，2.2 節でソーシャルメディアのプライバシーに関する研究を概観した後，2.3 節で近年の研究の中心であるプライバシーへの攻撃の研究を詳しく述べ，2.4 節で，本研究の関連研究にあたる機械学習を用いた再特定の研究について分析する．2.5 節では，ソーシャルメディア以外の分野における匿名データの再特定の研究を概観し，ソーシャルメディアの再特定の研究と同様の限界があることを明らかにする．

## 2.2　ソーシャルメディアのプライバシーに関する研究概要

ソーシャルメディアのプライバシーに関する先行研究は，(1)ソーシャルメディアのプライバシー問題に関する社会調査，(2)プライバシー情報の漏洩防止対策，(3)ソーシャルメディアから個人情報を抽出する攻撃手法の 3 つに分類することができる．本節では，このうち(1)と(2)について概観し，2.3 節，2.4 節にて(3)の攻撃研究について詳しく述べる．

(1)の社会調査では，ユーザのリテラシーに関して，ユーザによる公開範囲の設定（友達まで，全公開等）の状況調査 [14] [15] [16] [17] [18]，プライバシー保護機能の利便性および了解性の調査 [48] [19]，公開している情報の種類（本人，家族，友人の実名，誕生日，電話番号，現住所，出身地，政治観，友人の実名等） [14] [16] [49] [20]，プライベートな情報の割合 [50]などを調査している．また，ソーシャルメディアの機能に関わる問題として，リツイートや「友人の友人まで公開」，アプリケーションを介した想定外の情報拡散を調査している [51] [52] [53].

2015 年の総務省の調べによると，日本人の Twitter 実名利用率は 23.5%であり，その他のソーシャルメディアと比べると低いとされている [12]．2018 年，Shane らは，Facebook，Instagram，Twitter の利用者の傾向を性別，年齢，環境，プライバシーの懸念レベルの側面から調査した．調査によると，Facebook の利用者は，Instagram の利用者と比較すると自己開示レベルが低い一方，アカウント間の繋がりを重視している傾向が見られた．また，Twitter の利用者は自己開示レベルが最も高

く，異なる組織間における人脈を築き上げる傾向が見られた [21]．2017 年，竹地らは，プロファイリングの利用が，ターゲット広告配信や不正検知だけでなく，人事労務管理の分野にも広まりつつある現状に懸念を示している．プロファイリングの利用が，労務者にとってプライバシー問題や差別等の脅威になり得るとし，欧米諸国のようにプロファイリングに対する法的措置の実施が喫緊の課題であると訴えている [22].

組織がソーシャルメディアを利用して個人情報を収集する問題として，Background Checking とよばれる就職希望者を対象とした雇用前の身辺調査が問題視されている．Background Checking では，就職希望者が提出した履歴書などの情報を手がかりに，人事担当者が当該就職希望者のソーシャルメディアのアカウントを検索し，そこでの発言から，就職希望者の素行や友人関係，思想などを調査する．その結果次第では，選考対象から除外する企業もあるとされている [54]．ソーシャルメディアを用いた Background Checking は社会問題になっており，米国カリフォルニア州では，ソーシャルメディアのアカウントの開示を就職希望者に強要することを法律で制限している [55]．さらに，近年では，人工知能を用いて対象者の属性，思想信条，行動パターン，能力を推定するプロファイリングが大きな問題になっており，ソーシャルメディア上に開示された情報がプロファイリングに利用されることが指摘されている [23].

(2)のプライバシー情報の漏洩防止対策としては，ソーシャルメディア上の情報の適切な公開範囲の推薦 [56] [24]や自動設定 [57] [25]，投稿文の分析によるプライバシー漏洩の警告 [41] [26]，写真の投稿によるプライバシー漏洩の警告 [27]などが研究されている．2017 年，He らは，ソーシャルメディアのデータに対して，情報量を損なわないようにデータのプライバシーレベルを最大化するような最適化問題を定式化し，データのプライバシーと有用性を両立させるデータサニタイゼーション方法を提案した [28]．2018 年 5 月，EU 域内において，個人データの保護の強化を目的とする，EU 一般データ保護規則（GDPR：General Data Protection Regulation）が施行された [29]．2020 年，Google 社はこれを受け，2023 年を目途に，Chrome ブラウザにおいて，広告のためのサードパーティーCookie の利用を制限することを発表した [30]．近年，Facebook 等のソーシャルメディアを通じたオンライン広告配信が普及しているが，この規制により，広告主はユーザの Web 閲覧履歴に基づくレコメンド等が制限される．同年，Facebook は，GDPR やサードパーティーCookie の利用制限等を受けて，Facebook に広告を配信しているサードパーティーが収集した情報と，Facebook 上の個人情報との接続をユーザが拒否できる機能「Off-Facebook Activity」を提供している [31].

## 2.3　ソーシャルメディアへの攻撃の研究

攻撃の研究は，個人情報漏洩のリスクを示し，事業者やユーザに対策を促すことを目的としており，近年のソーシャルメディアのプライバシー研究の多くは攻撃研究である．匿名投稿文や匿名アカウントから投稿者やアカウントユーザを推定する再特定の研究[32] [33] [34] [35] [38] [39] [40] [58] [59] [60] [61] [62]と，投稿者やアカウントユーザの属性（年齢，性別，趣味，思想信条，能力など）を推定するプロファイリングの研究[41] [42] [43] [44] [45] [46] [63] [64] [65] [66]に分類することができる．

ソーシャルメディアのユーザの再特定の研究では，ヒューリスティックな手法 [32] [33] [34]，構造の照合に基づく手法 [58] [59] [60] [61] [62] [35]，機械学習に基づく手法 [34] [38] [39] [40] [35]が提案されている．ヒューリスティックな手法は，研究者の経験や勘を計算機により自動化した手法であり，再特定研究の初期に提案された．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 [32]．Polakis らはソーシャルメディア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 [33]．その結果，Facebook から抽出されたユーザプロフィールの内，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 を解析することで，複数の異なるソーシャルメディア（Yelp，Twitter，Flickr）を使用している同一ユーザのアカウントを特定する方法を提案したが，各々の特徴量（位置情報等）毎の同一性判定にはヒューリスティックな手法を用いている [34]．なお，3 つの特徴量毎の同一性判定の結果（類似度）を統合して，全体としての類似度を算出する部分には機械学習を用いている．

構造に基づく手法では，アカウント間のリンク関係等をグラフ等によって表現し，実世界の人間関係のグラフと照合することで個人を特定する．Backstrom らが 2007 年に提案した手法は，ソーシャルメディアのアカウントおよびアカウント間のリンクから成るグラフを，既知の人間関係のグラフと照合することで，アカウントのユーザを特定した [58]．2009 年，Narayanan らは，Twitter アカウントのリンクグラフと Flicker アカウントのリンクグラフの照合により，同一人物の Twitter アカウントと Flicker アカウントを特定した [59]．構造に基づく手法はいったん廃れたが，2015 年に，Niliadeh らは，2 つのソーシャルメディアのグラフをコミュニティ毎のサブグラフに分割して照合することで，誤ったノードやリンク（存在しないユーザやリンク）が混在していても正確な照合を行うことができる手法を提案した [60]．Gulyas らは，グラフ間の類似性を判定する新しい基準を導入することで，照合の精度を向上した [61] [62]．Lee らは，グラフ間の類似性の判定基準を機械学習によって最適化することで，照合の精度を向上した [35]．2 つの

グラフを照合する場合，各々のグラフから重要なサブグラフを抽出し，2 つのサブグラフを照合した上で，グラフ全体を照合することが多い．2019 年，Shao らは，既存手法の，事前の初期値設定を適切に行う必要がある点や，高い計算コストなどの課題を解消し，初期値の設定なしに高速に照合を行うことができる手法を提案した [36]．2018 年，Jiang らは，ソーシャルネットワークの構造の類似性と，ノードの属性の類似性を考慮した非匿名化方法を提案した [37]．構造の類似性では，異なるソーシャルネットワーク間の構造の特徴を照合し，属性の類似性では，各ノードの属性ベクトル（たとえば住所，専攻科目等）を照合する．

構造の照合に基づく再特定手法のうち，Backstrom らの手法は，攻撃者が特定対象者間の人間関係を知り，交友グラフを作成して，匿名アカウント間のグラフと照合することで，アカウントと人間を対応付ける [58]．そのため，Backstrom らの手法は実世界の人間を再特定できる．しかし，多くの場合，特定対象者の人間関係を知ることは困難である．Backstrom らの手法以外は，いずれも複数の匿名アカウント間のグラフを，別の複数のアカウント間のグラフと照合するので，再特定の対象者が複数のアカウントを開示していることが前提となる．また，照合先のアカウントが匿名である場合には，2 つの匿名アカウントが同一人物のものであることが判明するだけであり，その人物が誰であるかは特定できない．また，再特定対象の候補者について，ヒントとなる情報（たとえばラグビー観戦が趣味）を入手できたとしても，その情報を再特定に活かすことはできない．

さらに，2 つのグラフ間を精度よく照合するためには，グラフのノード数（すなわち対象となるアカウント数）がある程度多い必要がある．たとえば，図 2.1 のような 2 つの匿名アカウントのグラフと 3 つの実名アカウントのグラフを照合する場合，6 通りの照合が同じ確からしさで可能となるため，再特定は全くできない．2 つのグラフのノード数に大きな差がある場合には，小さいグラフは大きいグラフの様々なサブグラフと照合するため，再特定が困難となることが指摘されている [67]．

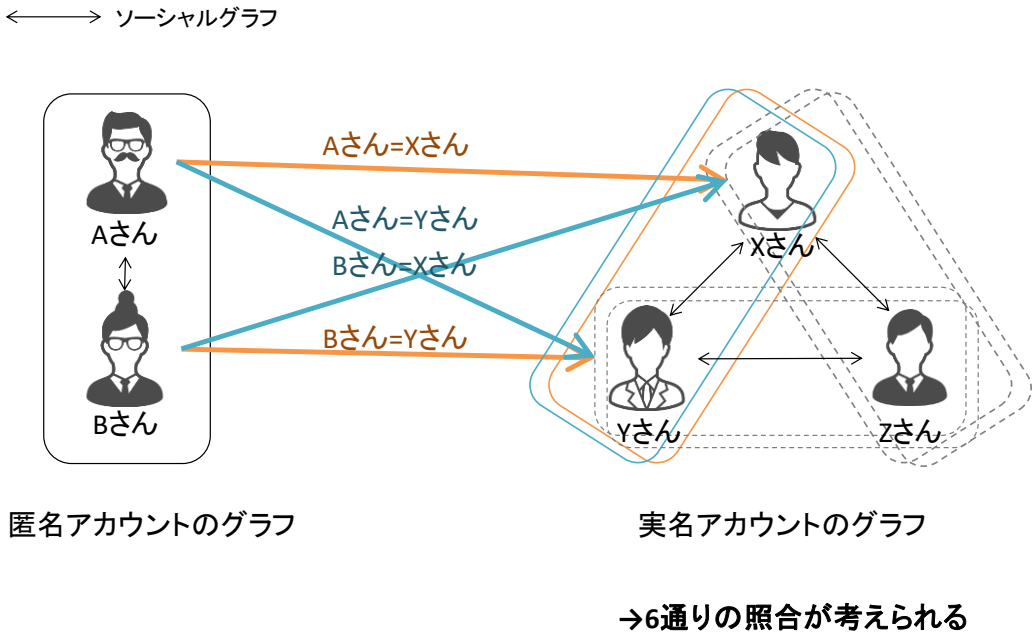


図 2.1　ノード数に差がある場合のグラフ間の照合

なお，Backstrom らの手法は，再特定の候補者間の人間関係を知ることが前提となっているが，その前提を満たすことは容易ではない．

以上の点から，構造に基づく再特定手法は利用できる条件が限定されているため実用性が低く，プライバシーリスクを示すには不充分である．

機械学習に基づく手法は，匿名アカウントの投稿文等からユーザの言語的特性を学習し，別の実名のアカウントのユーザの特性と照合することで，匿名アカウントと実名アカウントのユーザが同一であることを推定する．機械学習に基づく再特定の研究は，本研究の直接の関連研究であるため，2.4 節で詳しく分析する．

攻撃研究のもう一つの分野は，投稿者やアカウントユーザの属性（年齢，性別，趣味，思想信条，能力など）を推定するプロファイリングである．2011 年，Mao らはナイーブベイズ分類器とサポートベクターマシン（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 [41]．Mao 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 と Buttyan 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 [42]．この手法では，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．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 [43]．Hart らは既存のコーパスを使用することで，機械学習のための訓練データの準備の負荷を削減した [44]．Cheng らは，階層的確からしさモデルを用いて，Twitter のつぶやきの内容から，つぶやいた場所を推定した [45]．Burger らは，機械学習（Support Vector Machines，Naive Bayes，Balanced Winnow2）を用いて，Twitter のつぶやきから性別の推定を行った [46]．Pennacchiotti らは，Latent Dirichlet Allocation (LDA)を用いて，所属する政治的な党派，人種，好みのコーヒーブランド等の多様な属性を推定した [47]． 2016 年，Shigenaka らは，ソーシャルメディアの投稿画像から投稿者の性別を推定する手法として主流であった，画像から抽出される視覚的な特徴から推定する方法 [68]と，画像から抽出されるフィッシャーベクトル等の特徴量表現から推定する方法 [69]の 2 つのアプローチを統合し，既存手法よりも良好な精度で Twitter 画像の性別を識別した [63]．ユーザ属性の推測において，既存手法では友人のアカウント情報や，「いいね」などのユーザの行動記録のいずれかを利用していた [70] [71]．2018 年，Zhenqiang らは，友人情報と行動記録を統合した特徴を用いて，属性を推測する方法を提案した [64]．近年，テロ組織などの過激派ユーザが，一般ユーザと同様に，特別な制限なしでソーシャルメディアを通じた活動を実施している点が問題視されている．2016 年，Ferrara らは，機械学習によって過激派ユーザや，過激な内容のコンテンツ，過激派ユーザと一般ユーザとの関係性を検出する方法を提案した [65]．2018 年，Weerasinghe らは，Twitter のデータセットを用いて，精神疾患を抱える人の特徴的な言語パターンを分析し，機械学習によって精神疾患を抱える人物を推定する手法を提案した [66]．

## 2.4　機械学習を用いる再特定の研究

Goga らは，2.3 節で述べたように，複数のソーシャルメディア（Yelp，Twitter，Flickr）の中から同一ユーザのアカウントを特定した [34]．位置情報，タイムスタンプ，writing-style の 3 つの特徴量毎に，異なるソーシャルメディアのアカウント間の類似度をヒューリスティックな手法で算出する．2 値ロジスティック回帰を用いて，3 つの特徴量毎の類似度からアカウント間の総合的な類似度を算出する関数を学習し，この関数を用いた総合類似度が敷居値を超えた場合に，同一人物のアカウントと判定した．

Narayanan らは，2 つのブログアカウントの集合の中から，同一ユーザのアカウントのペアを特定した [38]．一方のブログ集合の各アカウントについて，ユーザの文章の特徴をナイーブベイズや線形判別分析などの複数のアルゴリズムを用いて学習した．特徴量としては，1 投稿当たりの単語数，長さ 1 から 20 までの単語の出現数等を用いた．また，

同一ユーザのアカウントの照合だけでなく，同一投稿者のブログの照合も同様の手法で行った．

Almishari らは，複数の匿名のつぶやき集合と，実名のアカウントの集合から，同一人物のつぶやき集合とアカウントのペアを特定した．ナイーブベイズアルゴリズムを用い，1 グラム，2 グラムの文字列を特徴量として，各アカウントのユーザの文章の特徴を学習し，この文章モデルを用いて，匿名のつぶやき集合とアカウントの類似度を算出した [39].

Overdorf らは，ブログ，Twitter の投稿，Reddit の投稿の 3 つの集合から，同一人物の投稿を特定した．上述した Narayanan らの手法を改良し，SVM とロジスティック回帰のアンサンブル学習を用いることで，異種のソーシャルメディア間の照合を可能にした [40].

Lee らは，グラフ照合と機械学習を組み合わせて，ソーシャルメディアの 2 つのアカウント集合から同一ユーザのアカウントを特定した．アカウントをノード，アカウント間の関係（友人リンク等）をリンクとして各々のアカウント集合をグラフ表現し，2 つのグラフを照合し，照合において対応するノードのペアを同一ユーザとした．グラフの照合には複数の可能性があるが，その中から最適な照合を選択するために，各ノードの周辺構造を特徴量として，部分構造間の類似度を算出する関数を SVM により学習し，部分構造間の類似度が最大になる照合を選択した [35].

既存手法では，異なるソーシャルメディア上の同一人物を照合する際に，関連する特徴（プロフィール，位置情報，リンク，文章特徴等）の抽出が必要であった [34] [72]．しかし，ソーシャルメディアのプラットフォーム毎に得られる特徴が異なる可能性があり，同一の特徴量に揃える必要があった．2018 年，Zhou らは， Deep Neural Network を用いることで，手間のかかる特徴量エンジニアリングを回避し，様々なソーシャルメディアに共通な特徴量を抽出する手法「DeepLink」を提案した [73].

2018 年，Wang らは，アカウント名や現住所等，テキスト形式で得られる属性値情報の文字列を連結し，これをユーザの特徴とすることで，ユーザ同士を照合する方法を提案した．既存手法では，プロフィールの情報を用いて同一ユーザを照合する場合，一つ一つの属性値を比較していた．Wang らの手法では，全属性を一つの文字列に集約し，N-gram 距離等の類似性尺度による文字列照合を行った結果，既存方法と比較して計算コストが削減され，照合精度も向上した [74].

2019 年，Li らは，ユーザ同士を 1 対 1 で紐付けるのではなく，ソーシャルネットワーク全体の分布とユーザの分布を写像変換によって照合する手法を提案した．写像変換関数の学習には，弱教師学習による Adversarial Learning を用いており，有意

な精度で照合が可能となった [75].

2018 年，Perez らは， KNN 等の機械学習手法を用いて，Twitter ユーザと，投稿者が不明なメタデータが付与された投稿を照合した．これにより，ユーザのプロファイル情報や「いいね」を押す等の行動情報，ユーザの位置情報を含むメタデータが，ソーシャルメディアにおけるユーザ特定の要素と成り得る可能性を提唱した [76].

以上の従来手法は，アカウントあるいは投稿文を同一人物の別のアカウントあるいは投稿文に照合している．また，Almishari らの手法は，投稿文を同一人物の別のアカウントに照合している．このことから，従来手法には下記の前提条件があることは明らかである．再特定の対象者は，再特定の対象となるアカウント（あるいは投稿文）以外に，他のアカウントを保有しており（あるいは他の投稿を行っており），攻撃者は両方のアカウントの投稿文（あるいは両方の投稿文）を入手可能であることが前提となるが，これが満たされるとは限らない．照合先となる別のアカウント（別の投稿文）が匿名である場合には，同一人物であることを推定できたとしても，個人を特定することにはならない．

また，再特定の対象者に関してヒントとなる情報を入手できる場合があるが，ヒント情報の活用は考慮していないという問題点がある．

## 2.5　他分野の再特定

匿名化された個人データからの個人の再特定は，ソーシャルメディア以外に，移動履歴 [77] [78] [79] [80] [81] [67]，Web 検索履歴 [82]，購買履歴 [83]，医療データ [84]に対しても研究されている．また， 汎用的な再特定手法の研究も存在する [85] [86]．このうち，移動履歴を対象とする手法 [81] [87] [84]，医療データに対する手法 [84]は機械学習を用いている．

Shokri ら [77]や Murakami [78]は，位置情報の欠損を含んだ実名移動履歴から，補完を行いながら，当該人物が 2 地点間を移動する確からしさを遷移行列の形式で表現し，匿名移動履歴が遷移行列にどれだけ整合しているかを定量化することで，匿名移動履歴の対象者を再特定した．その補完の方法として，Shokri らは Gibs サンプリング，Murakami は EM アルゴリズムを用いた．

Wang ら [81]の手法では，持ち主が同一の移動履歴のペアを教師データとして用いることで，同一人物の 2 つの移動履歴における地理的誤差と時間的誤差の分布をそれぞれ学習する．その後，2 つの分布を用いて実名移動履歴と匿名移動履歴の相違度を計算し，相違度が小さい場合に同一人物と判定する．

Feng ら [87]の手法では，recurrent network を用いた深層学習により，実名の移動履歴のモデルを学習し，このモデルに匿名移動履歴を入力することで匿名移動履歴と実名移

動履歴の類似度を算出し，匿名移動履歴を同一人物の実名移動履歴に対応付けている．

Santu ら [84]の手法では，ナイーブベイズ分類器を用いて，匿名の医療データを同一人物のソーシャルメディアデータに対応づけている．

以上に述べた手法では，匿名移動履歴を別の移動履歴に対応付け，匿名医療データをソーシャルメディアに対応付けている．そのため，ソーシャルメディアの場合と同様に，攻撃者は同一人物に関する複数のデータ（別の移動履歴，ソーシャルメディアデータ）を入手可能であることが前提となる．照合先の移動履歴やソーシャルメディアデータが匿名の場合には，照合に成功しても，個人の特定には至らない．また，再特定の対象者に関する付加的なヒント情報の活用は考慮しておらず，既知の人物の持つ匿名アカウントを特定する場合も考慮していない．以上のように，2.4 節で明らかにした従来方式の前提条件および問題点は，ソーシャルメディアの再特定に固有ではなく，再特定一般に共通している．したがって，これらの前提条件とは異なる前提で動作し，問題点を解決する手法を確立できれば，ソーシャルメディア以外のデータにも適用できる可能性がある．

## 2.6　まとめ

ソーシャルメディアのプライバシー対策については，社会調査，漏洩対策，攻撃の研究がおこなわれてきた．攻撃研究は，ヒューリスティックな手法，構造の照合に基づく手法，機械学習に基づく手法に分類できる．本論文の研究の直接の関連研究は，機械学習に基づいてソーシャルメディアのプライバシーを攻撃する研究であるが，これらの研究における攻撃手法は，いずれも，アカウントあるいは投稿を同一人物の別のアカウントあるいは投稿に照合している．そのため，再特定の対象者は，再特定の対象となるアカウント（あるいは投稿文）以外に，他のアカウント（投稿文）を開示しており攻撃者は両方のアカウント（投稿文）を入手可能であることが前提となる．また，照合先となる別のアカウント（別の投稿文）が匿名である場合には，同一人物であることを推定できたとしても，個人を特定することにはならない．さらに，再特定の対象者に関してヒントとなる情報を活用しておらず，既知の人物の持つ匿名アカウントを特定する攻撃を考慮していない．構造の照合に基づく手法においても，ほとんどの従来手法に同様の前提条件および問題がある．

# 第3章　プロファイリングに基づくソーシャルメディアの再特定

## 3.1　はじめに

本章では，プロファイリングを個人特定に利用することで，従来手法とは異なる前提条件で動作する手法を提案する．提案手法は，再特定対象の候補者のプロファイルまたはプロファイルを抽出可能なデータを，攻撃者が入手できることを前提とする．例えば，匿名アカウントの人物として，ある組織内の人物を特定したい場合，殆どの企業や学校，公共機関は，職員や学生の履歴書または相当情報を保有しているため，提案手法の前提を満たすことができる．

提案手法では，ソーシャルメディアアカウントのデータから，アカウント所有者の属性をプロファイリングし，推定したプロファイルと入手済のプロファイルとを照合する．プロファイルは実世界の個人と直結するので，候補者のプロファイルとの照合は実世界の個人特定となる．例えば，ある匿名アカウントが投稿した企業の内部告発の告発者を企業が特定する場合，匿名アカウントをプロファイリングした情報と，企業の保持している履歴書を比較することで，企業内のどの人物が告発者かを推定できる．

この手法を取ることで，同一人物が複数のソーシャルメディアアカウントを開示している必要はなく，従来手法の前提条件「複数のソーシャルメディアにアカウントを持つ人しか対象にならない.」,「照合の対象となるソーシャルメディアアカウントが公開されている必要がある.」を不要化できる．また，実世界の個人特定となるため，従来手法の前提条件に伴う問題点「照合先のソーシャルメディアアカウントが匿名の場合には，2つの匿名アカウントが同一人物のものであることが判明するだけで，実世界の個人が特定されない.」を解決できる．

## 3.2　匿名アカウントの再特定問題

匿名アカウントのユーザを再特定する問題を明確化する．

(1) 基本問題

多くの先行研究 [67] [78] [88]における再特定は図 3.1 のようにモデル化することができる．そこでは，$M$ 個の匿名アカウントの集合$X = \{x_i | 1 \leq i \leq M\}$と$N$個の別データの集合$Y = \{y_j | 1 \leq j \leq N\}$を想定する．ここで，各$y_j$を***サイドデータ***と呼ぶ．また，アカウント$x_i$とサイドデータ$d_j$が同一人物のものであることを$Same(i,\ j)$と表現することにする．このモデルにおける再特定とは，各$x_i (1 \leq i \leq M)$について，$Same(i,\ j)$となるサイドデータ$d_j$を特定することである．例えば，内部告発した匿名アカウントの発言者を特定する場合が基本問題に該当する．

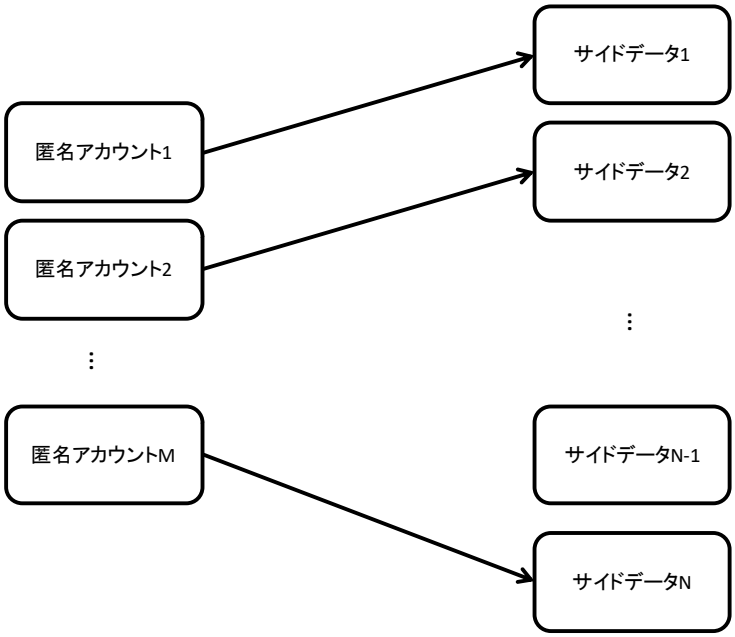


図 3.1 基本問題

(2) 逆問題

上記「(1) 基本問題」の状況において，各$y_j(1 \leq j \leq N)$について，$Same(i,\ j)$となる匿名アカウント$x_i$を特定する（図 3.2）．これは，既知の人物について，その人物が匿名で用いているアカウントを特定することを意味する．例えば，就職希望者が匿名で発言しているアカウントを特定する場合が逆問題に該当する．

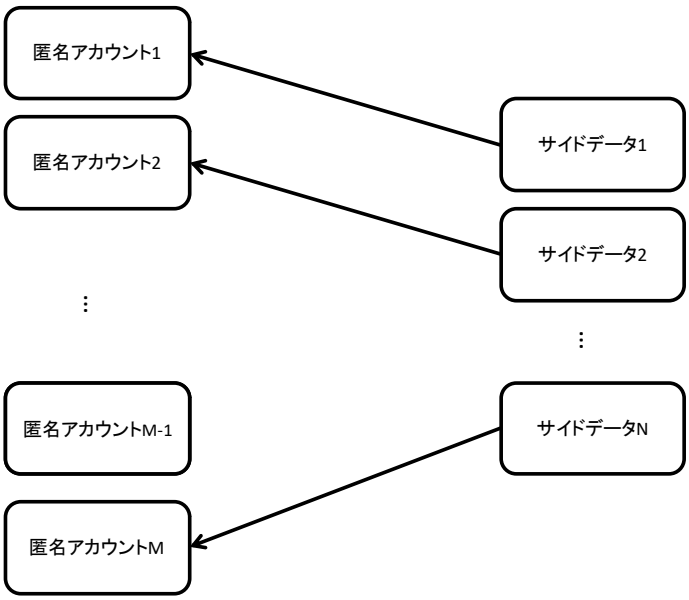


図 3.2 逆問題

(3) 1 対 1 問題

多くの先行研究（たとえば [38] [59] [77]）における再特定は，より強い前提を設けており，図 3.3 のようにモデル化することができる．そこでは，$M = N$かつ，関係$Same$は 1 対 1 である．このモデルにおける再特定とは，$X$の各要素$x_i$と$Y$の各要素$y_j$を$Same(i,\ j)$となるように 1 対 1 に対応付けることである．これは，基本問題あるいは逆問題において，アカウント数とデータ数が等しく，1 対 1 であるケースに該当する．本研究でも，先行研究にならって 1 対 1 問題を取り上げることにする．

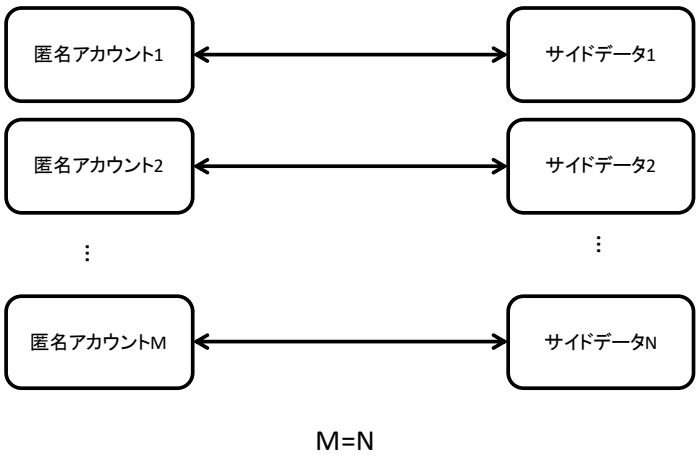


図 3.3 1 対 1 問題

本研究では上記の 3 つの問題を取り上げる．

## 3.3　課題と方針

本研究の目的は，従来手法とは異なる前提条件で動作する実用性の高い再特定手法の確立であるため，以下が課題となる．

(1) 再特定の対象となる匿名アカウント以外に，同一人物の他のアカウントが開示されていなくても，再特定を可能とする．
(2) 再特定対象者に関するヒント情報を入手した時に，アドホックな方法ではなく，提案法の枠組みの中で当該ヒント情報を利用し，再特定の精度を高めることができるようにする．

上記の課題を解決するための設計方針として，リンク構造の照合に基づくか，機械学習に基づくかの選択が考えられる．リンク構造の照合に基づく従来手法は，上記の課題を解決できないことに加え，再特定の対象となるアカウントおよび照合先となるアカウントが一定数以上存在し，両者のアカウント数に大きな差がないことを前提とするため，実用性に限界がある．これらのアカウント数に関する前提は，グラフの照合の本質的な問題点であるため回避が困難と考えられる．そこで，機械学習に基づいて提案法を検討することにした．

機械学習に基づく従来法は，再特定対象アカウントの投稿文を訓練データとして言語的特徴を学習するので，訓練データの準備（入手およびラベル付け）に関する問題は回避できる．しかし，一般に機械学習の利用にあたっては，訓練データの準備が実用上のボト

ルネックになることが多い．たとえば，2.3 節で述べたように，機械学習に基づくプロファイリングの研究では，訓練データの準備が問題になっている．そこで，本研究の第 3 の課題は下記の通りである．

(3) 機械学習のための訓練データの準備が，実用的な時間および労力の範囲内で可能となるようにする．

機械学習に基づく従来法は，投稿文のみに着目していた．しかし，ソーシャルメディアのアカウントには，投稿文以外に，投稿写真，友人やフォロー・フォロワーなどのリンク，ユーザプロフィールのデータが存在する．また，投稿文に付与された GPS 情報や投稿文の内容から，アカウントユーザの移動履歴を抽出することもできる．これらの情報を活かすことで，再特定の精度向上が期待できる．ソーシャルメディアの種類によっては，Instagram のように画像がメインで投稿されるものもあり，投稿文だけでは照合するための情報が不充分であることも考えられる．そこで，第 4 の課題は以下の通りである．

(4) 投稿文だけでなく，ソーシャルメディアアカウントの様々な種類のデータを活用して再特定の精度を向上することができる．

## 3.4 提案手法の基本方針

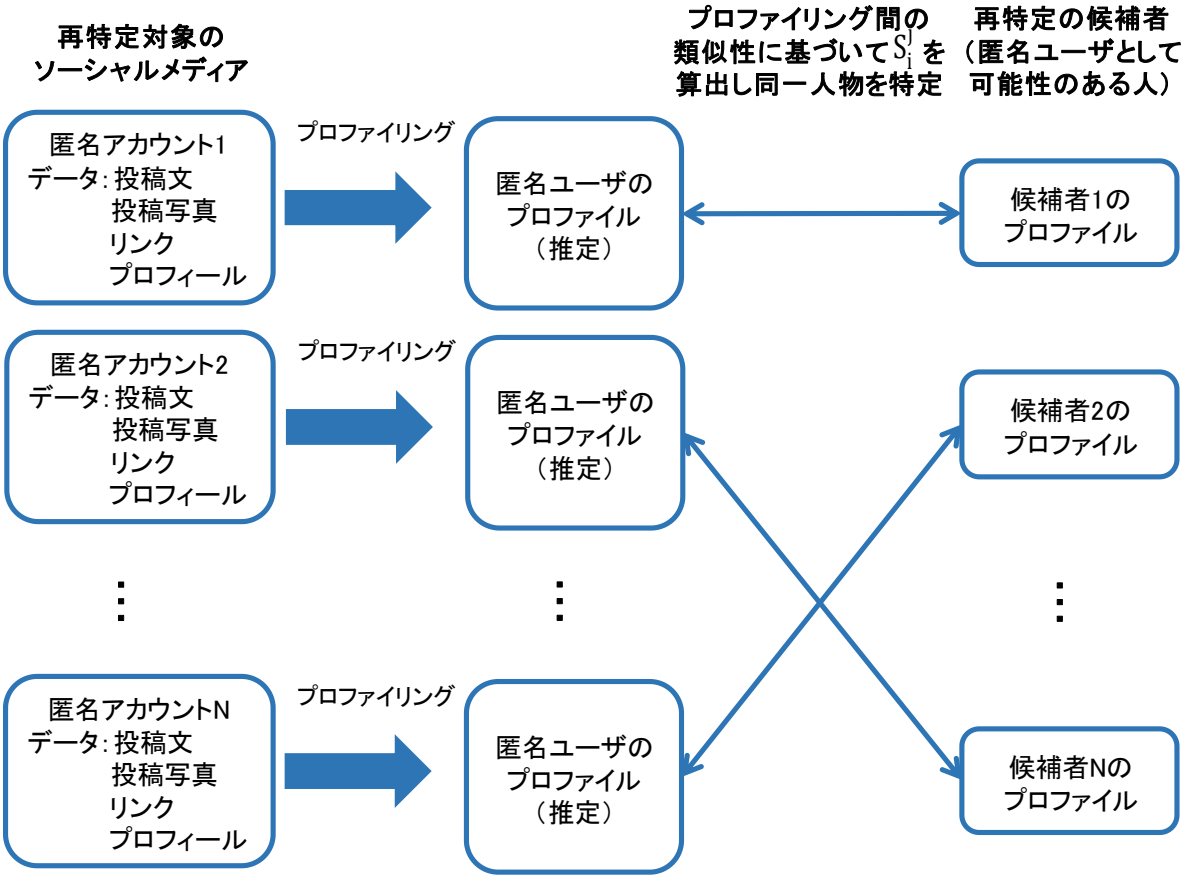


図 3.4　提案手法の基本方針

以上述べた再特定問題のモデル，課題，方針を踏まえ，下記の手法を提案する．

(1) 図 3.4 の右辺に示すように，攻撃者は，匿名アカウントのユーザの候補者集合を知っており，各候補者のプロファイル情報を入手できることを前提とする．

(2) 各候補者のプロファイルをサイドデータとする．候補者の人数を$N$とすると，$j$番目の候補者のプロファイルがサイドデータ$y_j$となる．

(3) 匿名アカウント$x_i$の人物をプロファイリングし，推定したプロファイルとサイドデータ$y_j$を比較することで，$Same(i, j)$であるかを判定する．すなわち，アカウント$x_i$とサイドデータ$y_j$が$Same(i, j)$を満たす確からしさを$S_i^j$とすると，プロファイリングを通じて，$S_i^j$を算出する．ここで，$S_i^j$が大きいほど$Same(i, j)$の可能性が高いとする．

(4) 機械学習を用いて$S_i^j$を算出する．

## 3.5　提案法の構成

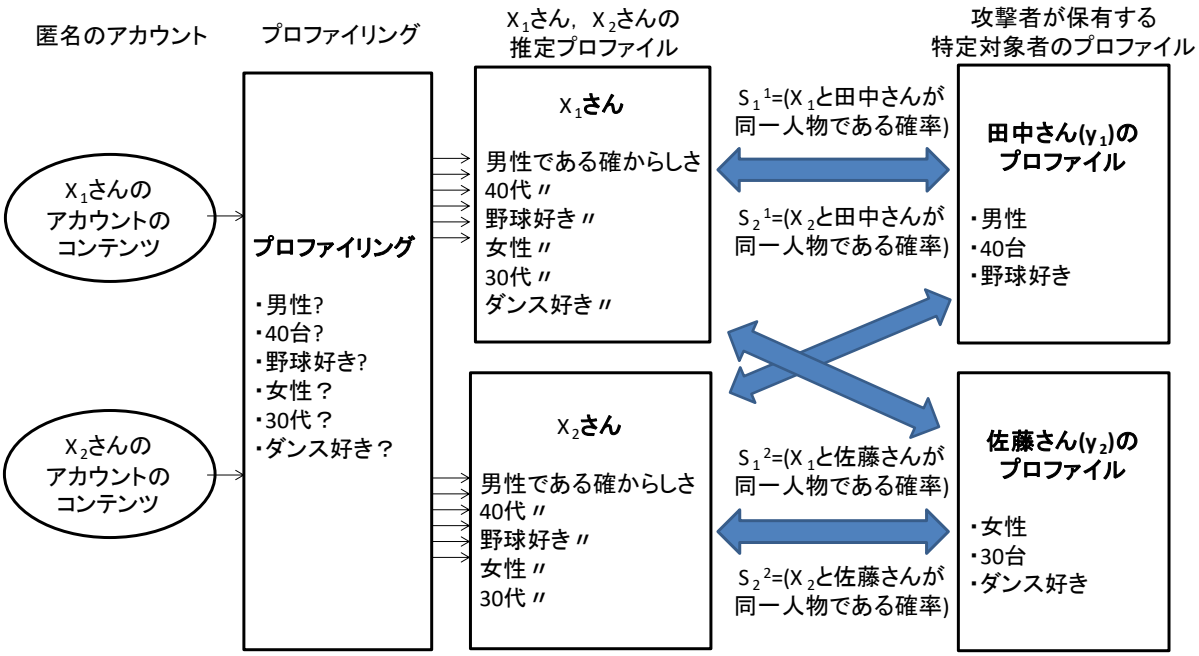


図 3.5　プロファイリングに基づく個人の特定方式

プロファイリングでは，各匿名アカウントの人物が，それぞれの候補者のプロファイルを構成している属性値（例えば，性別，年齢，趣味など）をどの程度有しているかをスコアとして算出し，これらの属性値毎のスコアから，各匿名アカウント$i$の人物が各候補者$j$である確からしさのスコア$S_i^j$を求める（図 3.5）．

匿名アカウントの人物が複数のプロファイル中の誰であるかを特定する場合，各匿名アカウント$i$から求めたスコア$S_i^j$を比較することで，どのプロファイルを持つ候補者$j$が当該匿名アカウントの人物$i$であるかを特定する．図 3.5 を例に説明する．匿名の人物$x_1$のアカウントが田中さん$y_1$，佐藤さん$y_2$のどちらのものであるかを特定する場合，$x_1$と田中さんが同一人物である確からしさ$S_1^1$と，$x_1$と佐藤さんが同一人物である確からしさ$S_1^2$を比較することで，$x_1$さんが二人のうちどちらの人物らしいかを特定する．

一方で，プロファイルが既知である人物の匿名アカウントを特定したい場合は，候補となる複数の匿名アカウントが当該人物である確からしさを求め，比較することで，どの匿名アカウントが当該人物であるかを特定する．図 3.5 を用いて，田中さんのプロファイルが既知であり，匿名アカウントの$x_1$と$x_2$が候補である場合を例に説明する．この場合，候補となる$x_1$，$x_2$それぞれについてプロファイリングを行い，それぞれのアカウントが田中さんである確からしさのスコア$S_1^1$，$S_2^1$を算出する．2 つのスコアを比較することで，$x_1$と$x_2$のうちどちらのアカウントが田中さんらしいかを特定する．

## 3.6　提案法の処理概要

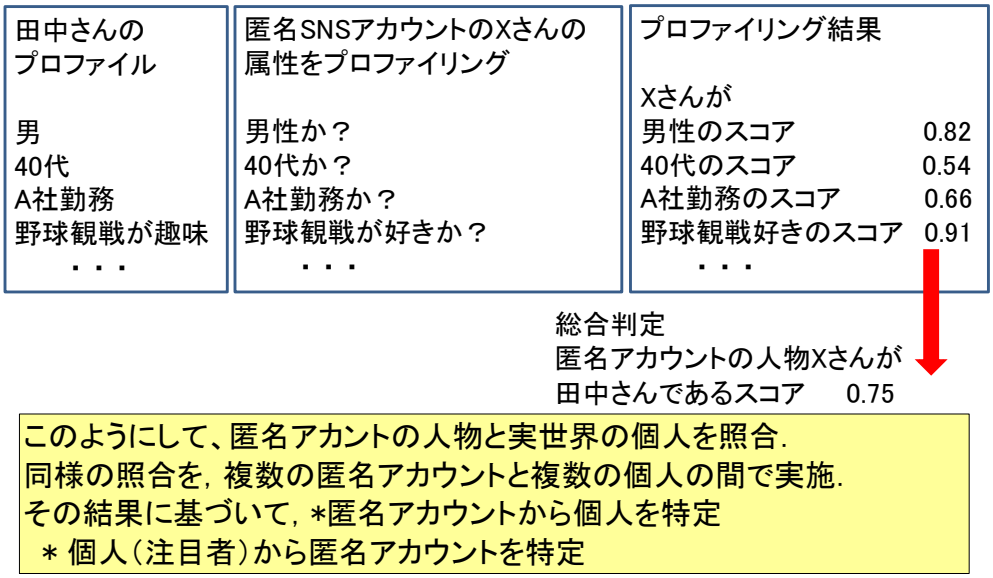


図 3.6　プロファイリングに基づく個人の特定方式

### 3.6.1　概要説明

最初に図3.6を用いて処理の例を説明した後，それを一般化し，アルゴリズムを説明する．例として，匿名アカウントの人物Xさんが，現実社会の田中さんであるかどうかをプロファイリングによって特定する場合の流れを説明する．田中さんのプロファイルは入手済みであるとし，属性値として性別＝男，年齢＝40代，勤務先＝A社，趣味＝野球が含まれていることとする．

まず，田中さんのプロファイルに記載されているそれぞれの属性値について，匿名アカウントのXさんがそれらの属性値を有する確からしさを機械学習によって算出する．この場合，性別が男であるかを識別するモデル，年代が40代であるかを識別するモデル，A社に勤務しているかを識別するモデル，野球観戦が好きかを識別するモデルを作成する．

次に，作成したそれぞれのモデルから，匿名アカウントのXさんが各属性値を有する確からしさのスコアを算出することでプロファイリングを行う．この場合，例えば，Xさんが男性である確からしさのスコアは0.82，40代であるスコアは0.54，A社勤務であるスコアは0.66，野球観戦好きであるスコアは0.91である．

最後に，各属性値に対して算出した確率を統合し，匿名アカウントのXさんが田中さんである統合スコア（ここでは0.75）を算出する．

これと同様のプロファイリングを，田中さん以外の人物のプロファイルでも行うことで，匿名アカウントのXさんが現実社会の誰であるかを推定することができる．また，田中さんがどの匿名アカウントであるかを特定したい場合は，その他の複数の匿名アカウン

トについて田中さんのプロファイルに基づいたプロファイリングを行うことで，どの匿名アカウントが最も田中さんらしいかを推定することができる．

### 3.6.2　アルゴリズム

以上の例を踏まえて，提案方式の一般的なアルゴリズムを説明する．

攻撃者が特定したい匿名アカウントの集合を$X$, 候補者の集合を$Y$とする．

匿名アカウント数を$M(=|X|)$ とする場合，匿名アカウントの集合 $X$に含まれる個々の匿名アカウントは$x_i(1 \le i \le M)$，$x_i \in X$である．また，アカウント$x_i$のデータを$d_i$とする．

候補者数を$N(=|Y|)$とする場合，候補者の集合 $Y$に含まれる個々の候補者は$y_j(1 \le j \le N)$, $y_j \in Y$ である．また，候補者$y_j$の持つ属性値の数を$L_j$とし，各属性値を$a_k^j$とする．候補者$y_j$のプロファイルを，$y_j$の持つ属性値の集合$profile^j = \{a_k^j(1 \le k \le L_j)\}$で表すことにする．

この時，攻撃者が見つけたい情報は，各匿名アカウント$x_i$に対する同一人物の候補者$y_j$，もしくは，各候補者$y_j$に対する同一人物の匿名アカウント$x_i$である．そこで，各アカウントの情報$d_i(1 \le i \le M)$を用いて，$x_i$が候補者$y_j$のプロファイル$profile^j(1 \le j \le N)$内に含まれる各属性値$a_k^j(1 \le k \le L_j)$を持つ確からしさのスコア$S_{i,k}^j$を算出する．これをまとめると，表 3.1 のように表すことができる．

表 3.1　$x_i(1 \le i \le M)$が$profile^j(1 \le j \le N)$に含まれる各属性値$a_k^j(1 \le k \le L_j)$を持つ確からしさのスコア

| | $a_1^1$ | … | $a_{L_1}^1$ | … | $a_1^j$ | … | $a_k^j$ | … | $a_{L_j}^j$ | … | $a_1^N$ | … | $a_{L_N}^N$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_1$ | $S_{1,1}^1$ | … | $S_{1,L_1}^1$ | … | $S_{1,1}^j$ | … | $S_{1,k}^j$ | … | $S_{1,L_j}^j$ | … | $S_{1,1}^N$ | … | $S_{1,L_N}^N$ |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| $x_i$ | $S_{i,1}^1$ | … | $S_{i,L_1}^1$ | … | $S_{i,1}^j$ | … | $S_{i,k}^j$ | … | $S_{i,L_j}^j$ | … | $S_{i,1}^N$ | … | $S_{i,L_N}^N$ |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| $x_M$ | $S_{M,1}^1$ | … | $S_{M,L_1}^1$ | … | $S_{M,1}^j$ | … | $S_{M,k}^j$ | … | $S_{M,L_j}^j$ | … | $S_{M,1}^N$ | … | $S_{M,L_N}^N$ |

全ての属性値$a_k^j$について$S_{i,k}^j$が算出された後に，$L_i$個の確率$S_{i,k}^j(1 \le k \le L_j)$を統合し，$x_i$と$y_j$が同一人物である確からしさのスコア$S_i^j$（統合スコア）を求める．$S_i^j$は，それぞれの$x_i$と$y_j$における$S_{i,k}^j$ を足し合わせた確率を，$y_j$のプロファイルに含まれる属性値数$L_i$で割った値，$S_i^j = \frac{S_{i,1}^j + S_{i,2}^j + \cdots + S_{i,Lj-1}^j + S_{i,Lj}^j}{L_j}$とする．以上を全ての$i$と$j$ のペアについて実行する．$x_i$

と$y_j$が同一人物である確率$S_i^j$をまとめると表3.2のように表すことができる．

表 3.2　$x_i$と$y_j$が同一人物である確率$S_i^j$

| | $y_1$ | $y_2$ | … | $y_{N-1}$ | $y_N$ |
|---|---|---|---|---|---|
| $x_1$ | $S_1^1$ | $S_1^2$ | … | $S_1^{N-1}$ | $S_1^N$ |
| $x_2$ | $S_2^1$ | $S_2^2$ | … | $S_2^{N-1}$ | $S_2^N$ |
| … | … | … | … | … | … |
| $x_{M-1}$ | $S_{M-1}^1$ | $S_{M-1}^2$ | … | $S_{M-1}^{N-1}$ | $S_{M-1}^N$ |
| $x_M$ | $S_M^1$ | $S_M^2$ | … | $S_M^{N-1}$ | $S_M^N$ |

3.1 節の基本問題の場合には，アカウントから個人を特定するので，各$i$行について，行方向で確率が最大となるプロファイル$j'$を，以下の式により特定する．

$$j' = Argmax_{1\le j\le N}(S_i^j)$$

3.1 節の逆問題の場合には，個人からアカウントを特定するので，各$j$列について，列方向で確率が最大となるアカウント$i'$を以下の式により特定する．

$$i' = Argmax_{1\le i\le M}(S_i^j)$$

3.1節の1対1問題の場合には，ハンガリアンアルゴリズムを用いることで，下記の式により$X$から$Y$ への全単射$\sigma'$を求めることにより，$M$ 個の$x_i$と$y_j$のペアを特定する．ここで，$\Sigma$は$X$から$Y$への全ての全単射の集合を表す．

$$\sigma' = Argmax_{x\in\Pi}\left(\sum_{1\le i\le M} S_i^{\Pi(i)}\right)$$

### 3.6.3　モデルの学習方式

次に，属性値モデルの学習方法について，上記の記号を用いて説明する．

$x_i$が属性値$a_k^j(1\le k\le L_j)$を持つ確率$S_{i,k}^j$ は，匿名のソーシャルメディアアカウント$x_i$のデータ$d_i$をモデル$F_k^j$に入力し算出されるものとする．$S_{i,k}^j = F_k^j(d_i)$と表すことができる．モデル$F_k^j$は機械学習を用いて学習するため，属性値$a_k^j$を持つ複数の公開アカウントと，属

性値$a_k^j$を持たない公開アカウントを収集し，これらのアカウントのデータを正例，負例としてモデル$F_k^j$を学習する．

学習データを公開データから容易に収集可能であれば，学習データの準備に関する課題（入手およびラベリングの時間，労力の軽減）を解決することができるが，その評価は，4 章以降の具体例において行う．

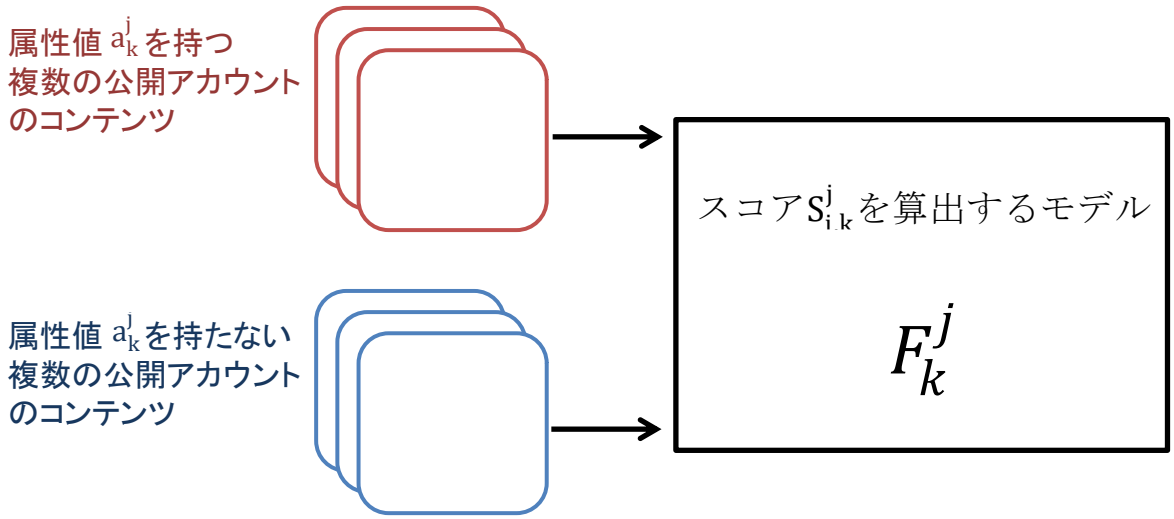


図 3.7　モデル$F_k^j$の学習方法

### 3.6.4　処理フロー

攻撃者が持っている情報（入力）：$X$と$Y$

見つけたい情報（出力）：各$x_i$に対して同一人物の$y_j$，各$y_j$に対して同一人物の$x_i$

① 属性値の洗い出し

$Y$の全ての候補者$y_j$のプロファイル$profile_j$に含まれるすべての属性値$a_k^j(1 \le j \le N, 1 \le k \le L_j)$を洗い出す．

② 学習データの収集

公開アカウントから正例，負例を収集する．各属性値$a_k^j$について，当該属性値を有する公開アカウントと有さない公開アカウントを収集し，それらのデータを属性値$a_k^j$の正例，負例とする．

③ 属性値モデル$F_k^j$の学習 $(1 \le j \le N, 1 \le k \le L_i)$

②で収集した正例，負例を用いて，属性値$a_k^j$を持つスコア確率$S_{i,k}^j$を算出するためのモデル$F_k^j$を学習する．

④ 各アカウントが各属性値を有するスコア$S_{i,k}^j$の算出 $(1 \le i \le M, 1 \le j \le N, 1 \le k \le L_i)$

$x_i$が各候補者$y_j$のプロファイル$profile^j(1 \le j \le N)$内に含まれる各属性値$a_k^j(1 \le k \le L_j)$を持つ確率$S_{i,k}^j$を算出する．

⑤ $S_i^j$の算出 $(1 \le i \le M, 1 \le j \le N)$

$x_i$と$y_j$が同一人物である確率$S_i^j$（統合スコア）を，$\frac{S_{i,1}^j+S_{i,2}^j+\cdots+S_{i,L_j-1}^j+S_{i,L_j}^j}{L_j}$により求める．

⑥ アカウント$i$とプロファイル$j$の対応付け

基本問題：$j' = Argmax_{1 \le j \le N}(S_i^j)$

逆問題：$i' = Argmax_{1 \le i \le M}(S_i^j)$

1 対 1 問題：$\sigma' = Argmax_{x \in \Pi}\left(\sum_{1 \le i \le M} S_i^{\Pi(i)}\right)$

## 3.7　まとめ

3 章では，匿名のソーシャルメディアアカウントをプロファイリングすることで，ユーザを再特定する方式を提案した．提案方式は，再特定候補者のプロファイルが入手可能であることを前提としている．匿名アカウントの人物をプロファイリングし，候補者のプロファイルを構成している属性値をどの程度有しているかをスコアとして算出する．匿名アカウントの人物の推定プロファイリングと候補者のプロファイルを照合し，匿名アカウントが候補者である確からしさをスコアとして算出する．

提案方式は，匿名アカウントとプロファイルを照合し，別のアカウントとは照合しないので，匿名アカウント以外に同一ユーザの実名アカウントを必要としない．また，再特定対象者に関するヒント情報（たとえばラグビー観戦が趣味）を属性値とみなすことで，提案手法の中で自然に利用することができるので，従来手法の問題点を解決できる．

# 第4章　投稿文と履歴書を用いた個人の再特定

## 4.1　はじめに

3 章では，ソーシャルメディアアカウントの人物をプロファイリングし，再特定候補者のプロファイルと照合する手法を提案した．この手法ではプロファイリングにより個人を照合するため，照合先の詳細なプロファイルが利用可能であれば，実世界の個人をより正確に特定できる．

本章では，照合先の詳細なプロファイルとして履歴書を，ソーシャルメディアアカウントのデータとして投稿文を利用する場合のシステム構成および評価結果を述べる（図 4.1）．履歴書は実世界の個人と直接対応しているため，提案手法におけるプロファイルとして履歴書を利用すれば，実世界の個人を一意に特定可能である．プロファイルに履歴書を用いる場合の提案手法の具体的な実現方法と，実データを用いた個人特定の精度を明らかにする．

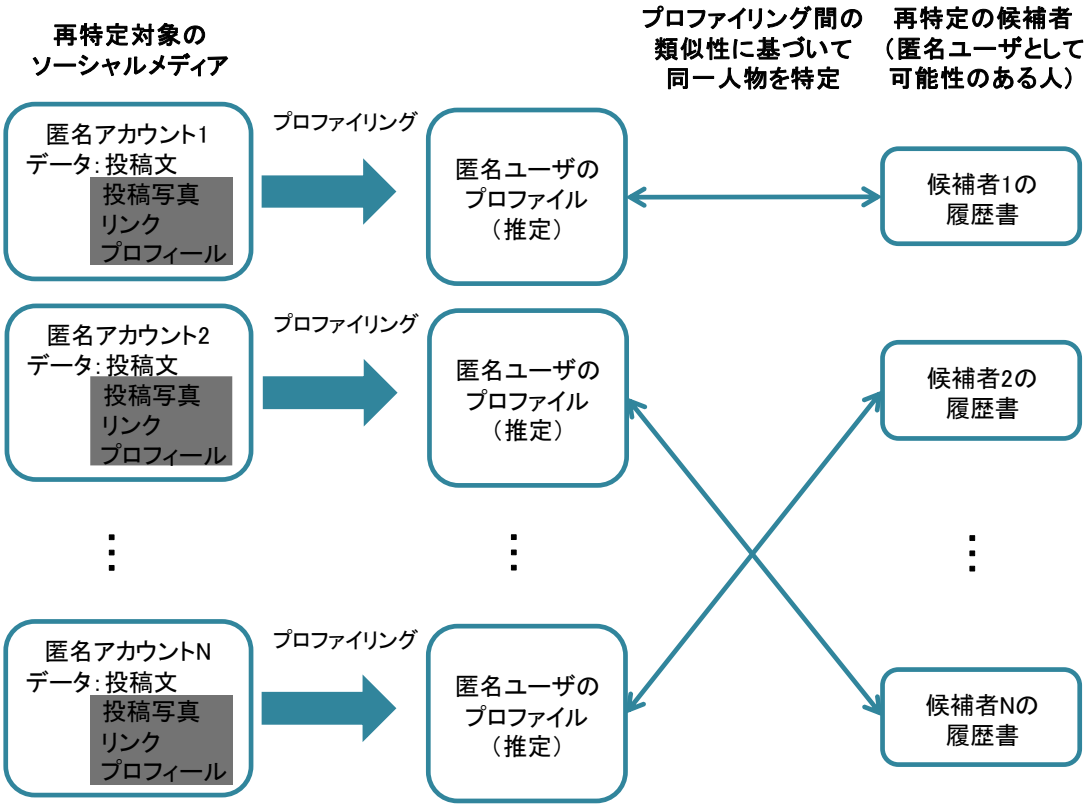


図 4.1 プロファイルとして履歴書を用いる場合のシステム構成

## 4.2 想定される応用と実用性

### 4.2.1　プライバシーへの応用

提案法では，ソーシャルメディアアカウントと履歴書を双方向に紐付ける．匿名のソーシャルメディアアカウントから，当該アカウントのユーザに対応する履歴書を特定する．一方，履歴書から，履歴書が表す人物のソーシャルメディアアカウントを特定する．これにより，ソーシャルメディアから個人が特定されるプライバシーリスクを明らかにする．

図 4.2 は，ソーシャルメディアアカウントから履歴書を特定する場合の例を示している．この例では，ある企業 A 社が，国のコロナ対策を批判する投稿文をソーシャルメディア上で発見し，同じアカウントの他の投稿文からアカウントユーザが A 社の社員であると推定している．このとき，A 社は本手法を用いることで，アカウントを社員の履歴書の一つに紐付け，投稿者が企業内の山田さんであると特定することができる．このようにして，A 社は社員の思想，信条にかかわるプライバシー情報を入手する．

図 4.3 は，履歴書からソーシャルメディアアカウントを特定する場合の例を示している．この例では，求職者である高橋さんが，履歴書を企業 B 社に提出した場合を想定している．企業は本手法を使用することによって，高橋さんの履歴書と，予め収集したいくつかのソーシャルメディアアカウントの候補を照合させることで，最も高橋さんのものらしいソーシャルメディアアカウントを特定し，高橋さんの交友関係や思想信条に関するプライバシー情報を入手する．その他にも，企業内での昇進候補者や対抗者の匿名ソーシャルメディアアカウントが特定されるリスクを明らかにすることもできる．

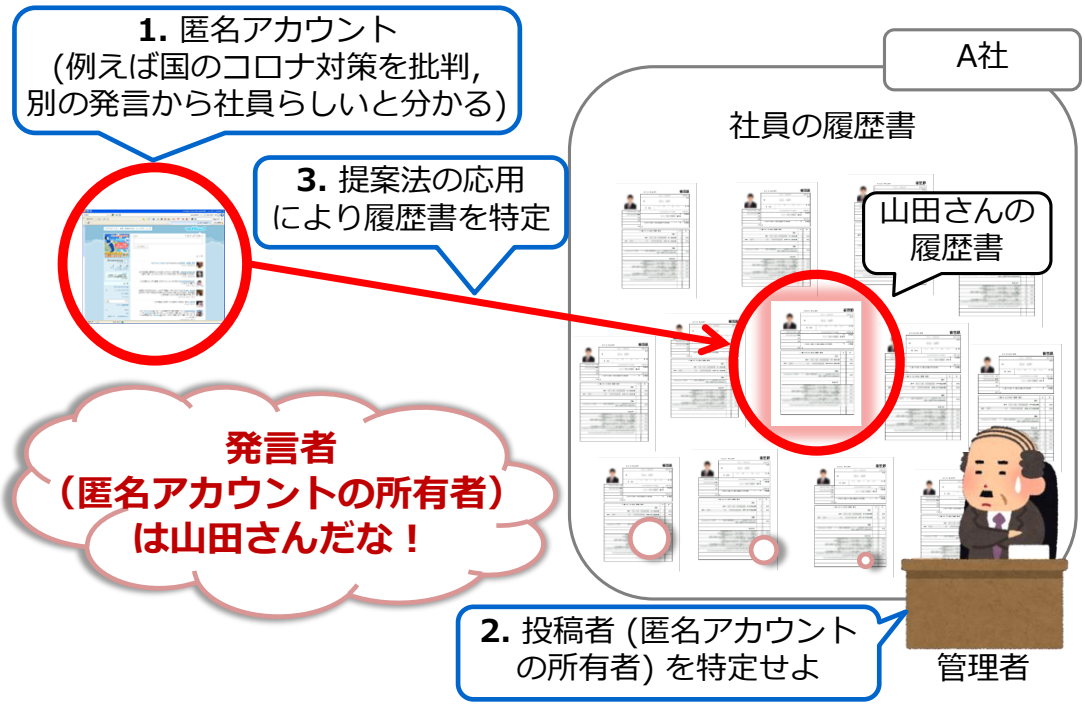


図 4.2　アカウントから履歴書を特定する場合の応用例（プライバシー）

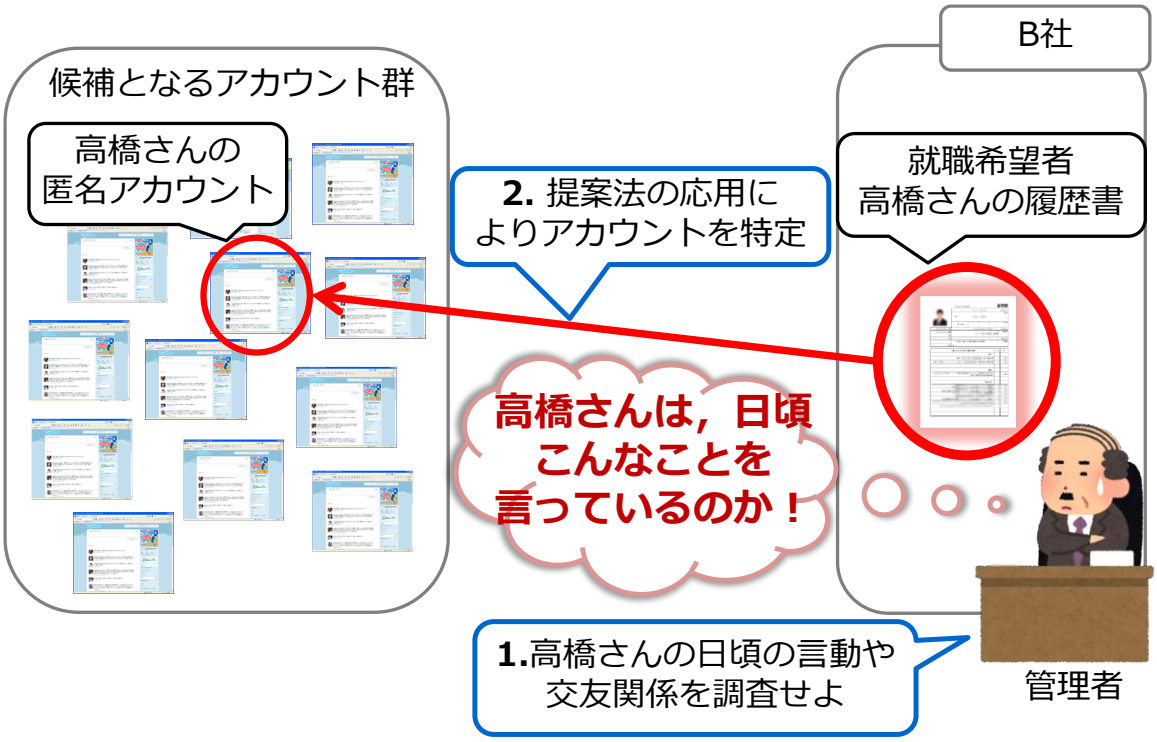


図 4.3　履歴書からアカウントを特定する場合の応用例（プライバシー）

### 4.2.2　セキュリティへの応用

本手法は，プライバシーのリスク明確化だけでなく，セキュリティ対策にも使用することが可能である．図 4.4 は，企業 C 社に勤めている山田さんが，社内の機密情報を本人が保有している匿名のアカウントに投稿し，漏洩させた場合を示している．この時，企業はこの機密情報を漏洩させたアカウントの人物が，どの社員であるかを特定するために本手法を使用する．本手法は，匿名アカウントを社内の全履歴書と照合し，漏洩させた社員が山田さんであることを特定する．このように，悪意を持って企業の秘密情報を漏洩した人物を特定することが可能であり，企業はその人物に対して相応の罰を与えるという選択ができる．また，そのような情報漏洩を抑止するために本手法を使用することもできる．その他にも，著作権侵害，いじめ，名誉棄損に関する投稿の抑止にも効果的である．

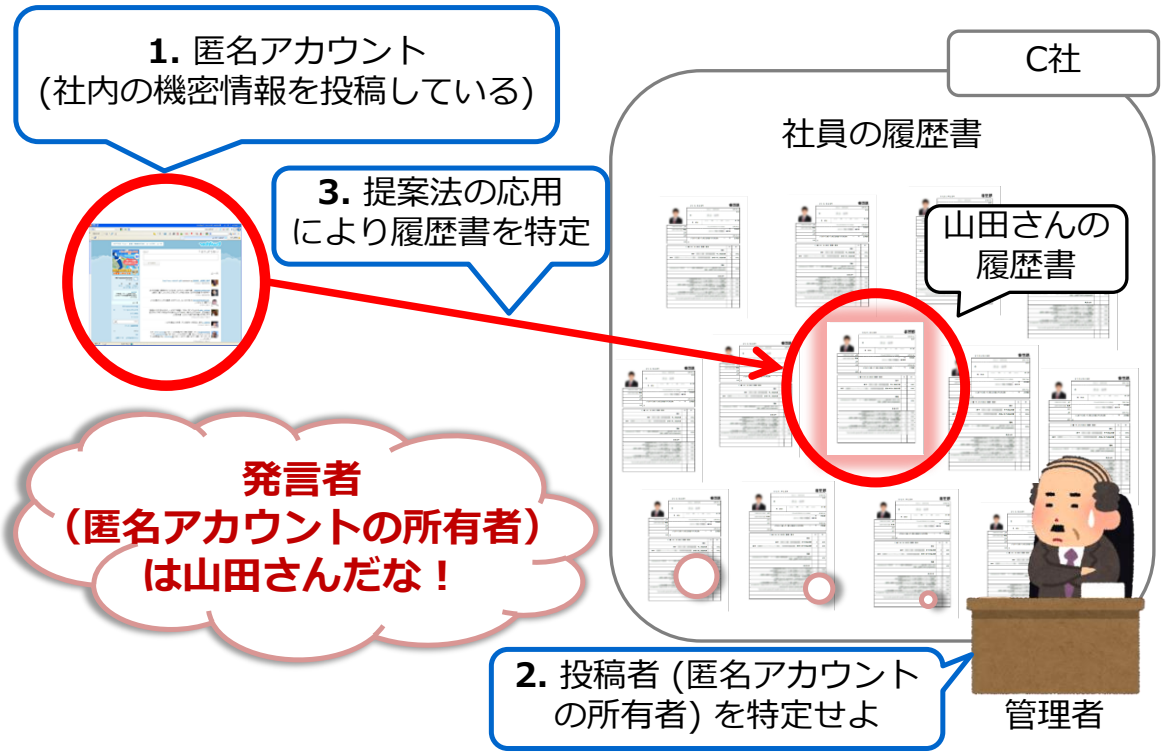


図 4.4　アカウントから履歴書を特定する場合の応用例（セキュリティ）

## 4.3 履歴書を用いた照合手法の構成

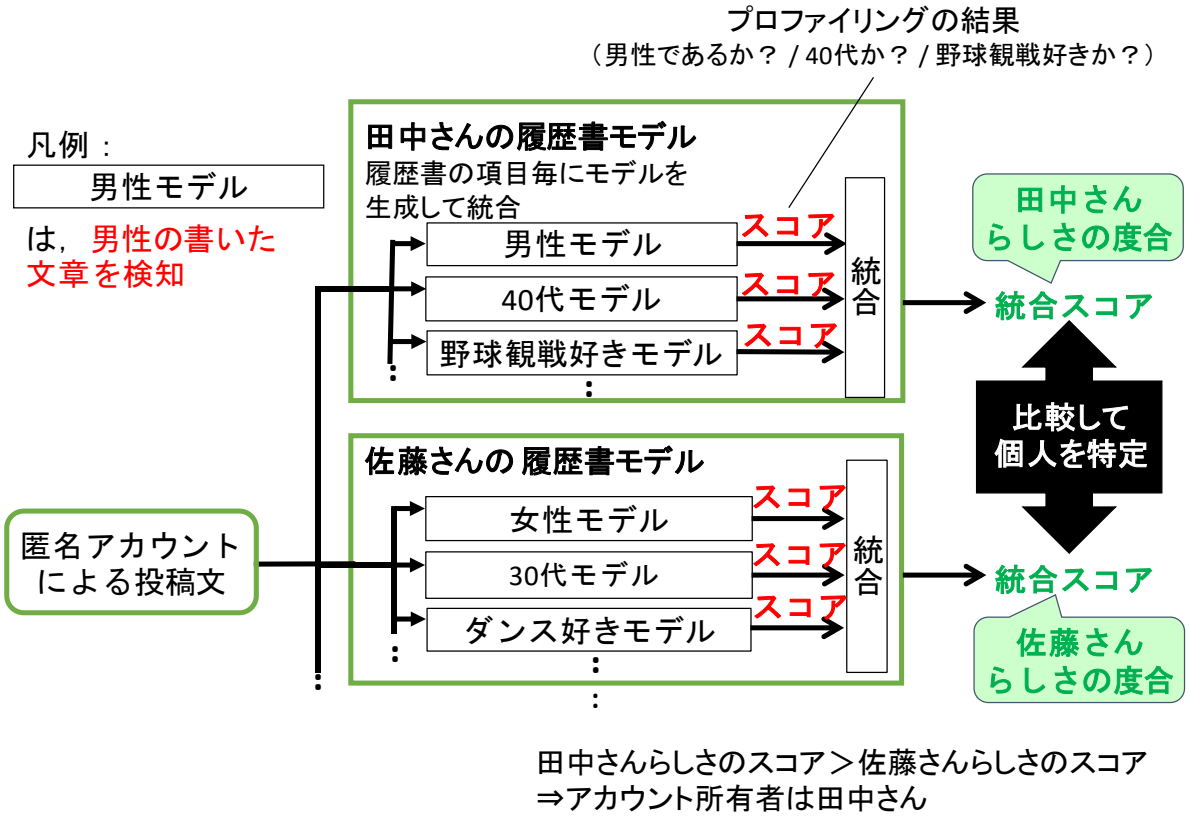


図 4.5　履歴書に基づく個人の再特定方式

プロファイルとして履歴書を，匿名アカウントのデータとして投稿文を用いた場合の提案法の構成を説明する．この構成では，特定対象者の履歴書毎に個人を識別するモデル（以下，履歴書モデル）を学習する（図 4.5）．履歴書モデルは，それぞれの人物の履歴書から作成する．すなわち，履歴書を構成している項目（例えば，性別，年齢，趣味など）毎に，人物がその属性値を有するかを識別するモデル（以下，属性値モデル）を学習し，それらのモデルを統合して履歴書モデルとする．

履歴書モデルは，各属性値モデルのスコアを統合し，匿名アカウントの人物が当該履歴書の人物らしいかを表す統合スコアを出力する．統合スコアを比較することによって，複数ある候補者の履歴書のうち，匿名アカウントがどの履歴書の人物であるかを推定する．

図 4.5 を用いて具体的に説明する．この場合，攻撃者は本人を特定したい匿名アカウントの候補の人物である田中さん，佐藤さんのそれぞれの履歴書に基づいて履歴書モデルを作成する．田中さんの履歴書モデルは，田中さんの履歴書に記載されている各属性値，性別=男性，年代=40 代，趣味=野球観戦のそれぞれについて識別する属性値モデルから構成される．同様に，佐藤さんの履歴書モデルも作成し，それぞれの履歴書モデルに匿名アカウントの投稿文を入力することで，各履歴書に対する統合スコアを得る．この統合スコアを田中さんと佐藤さんで比較することによって，匿名アカウントの人物がどちらの人物のものであるかを推定することができる．

一方で，プロファイルが既知である人物の匿名アカウントを特定したい場合は，プロファイルが既知である人物から作成した履歴書モデルに，候補となる複数の匿名アカウントの投稿文を入力し，匿名アカウントの統合スコアを比較することで，どのアカウントが当該人物であるかを特定する．

## 4.4　モデルの学習

### 4.4.1　属性値モデルの学習

属性値モデルは，ある人物がその属性値を有するかを識別する．属性値モデルの学習には，教師あり学習を用いる．例えば，属性値モデルとして，女性であるかどうかを識別するモデルを作成する場合，教師データの正例として女性の投稿文，負例として女性でない人物の投稿文を用いる．この時，教師データはユーザプロフィール情報に当該属性値を含む複数の公開アカウントから入手した投稿文を使用する．

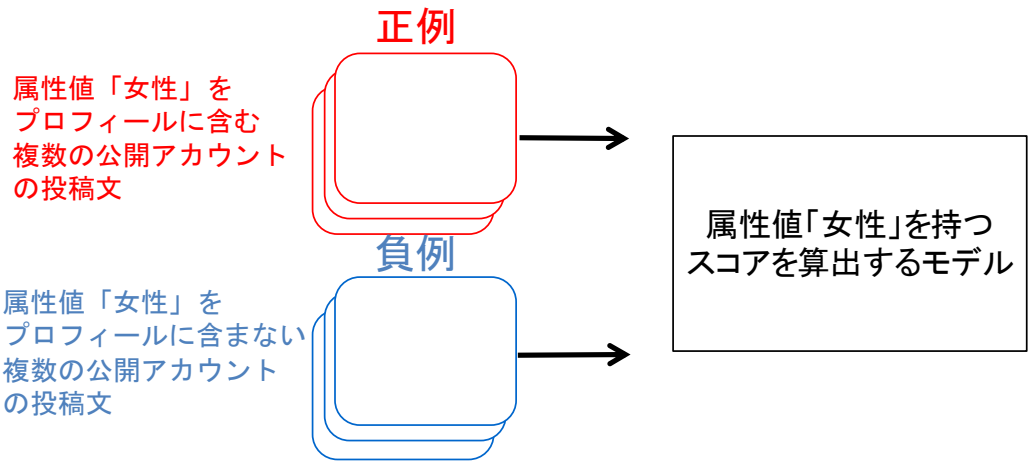


図 4.6　属性値モデルの学習方法

訓練データの収集はツイプロ（検索ワードをユーザ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 [89]，当該属性値をユーザ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，ユーザ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．この方法により，公開された不特定多数のアカウントから学習データを収集できるので，学習データの収集の難しさが解消され，人手によるラベル付けの手間を解決できる．

### 4.4.2　履歴書モデルの構成

各候補者の履歴書モデルは，履歴書に含まれる属性値に対応した属性値モデルから構成する．履歴書モデルの入力は匿名アカウントの投稿文であり，出力は匿名アカウントの人物と履歴書の人物が同一である確からしさのスコアである．履歴書モデルの出力は，各属性値モデルの出力を統合することで算出する．

## 4.5　予備評価

履歴書モデルを精度よく学習するためには，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，属性値モデルのスコアから履歴書モデルのスコアを算出するためのスコア統合方法を選定する必要がある．その選定の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．

### 4.5.1　データセット

著者らが所属する電気通信大学の学生 30 名の Twitter アカウントと履歴書を用いた．表 4.1 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，偽のユーザプロフィール情報の使用などにより匿名化が施されている．

表 4.1　被験者の内訳

(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) 学科

| 総合情報学科 | 情報・<br>通信工学科 | 知能機械工学科 |
|---|---|---|
| 24 | 2 | 4 |

訓練データは，4.4 節で述べたようにツイプロによって入手した．各属性値モデル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公開アカウントであり，かつ，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 件までのつぶやきを取得して使用した．例えば，「趣味=ダンス」のモデルを作成する場合，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モデルの学習を断念した．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り，かつ，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．

### 4.5.2　予備評価

**(1)** 小規模データセット

予備評価用データ 30 名分のうち 6 名の被験者の履歴書とつぶやきを使用して，最初に小規模な予備評価を行った．表 4.2 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，プライバシーの観点から匿名化している）．被験者 6 名の履歴書からは，7 種類の属性と 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，モデルを作成しなかった．モデルを作成不可能であった属性値には，表 4.2 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 以外はモデル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 個の属性値モデルを作成し，被験者 6 名のつぶやきをテストデータとして用いることで，これらのモデルを評価した．

表 4.2　予備評価に用いた属性と属性値

| 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,<br>オードリー,<br>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,<br>~~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,<br>音楽,<br>数学 | 合気道部,<br>バスケットボール,<br>音楽,<br>読書,<br>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,<br>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,<br>フットサル,<br>テレビ,<br>サッカー | ~~普通自動車免許~~,<br>証券外務員,<br>FP |
| 6 | M | 東京都　D市 | 青森県 H市 | ~~情報・通信工学科~~ | 電子回路,<br>電磁気学,<br>Webデザイン,<br>プログラミング | 野球部,<br>野球,<br>~~ソーシャルメディアを眺める~~ | – |

**(2)** 正規化

提案手法では，各属性値モデルが算出したスコアを統合することで照合を行う．この時，属性値モデル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，属性値モデル A の算出するスコアの値の範囲が，他の属性値モデル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，属性値モデル A のスコアが支配的となり，複数の属性値モデルのスコアを統合するメリットが得られない．そこで，各属性値モデルの算出するスコアの値の範囲が概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．属性値モデル毎に，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，$N$ を属性値モデルの数とする（予備実験では$M = 6$，$N = 40$）．$x_{ij}$は$j$番目の属性値モデルが$i$ 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$\sigma_j$は，$j$ 番目の属性値モデルが算出した$M$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$\alpha_{ij}$は，式(1)のように計算することができる．

$$\alpha_{ij} = \frac{x_{ij} - \overline{x_j}}{\sigma_j} \qquad (1)$$

**(3)** 結果と考察

図 4.7 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モデルのスコア分布である．横軸は属性値モデル，縦軸は属性値モデルが 6 個のアカウントに対して算出したスコアを正規化した値を示している．属性値モデルが算出したスコア（正規化済）の分布は，箱ヒゲ図と点で表されている．箱ヒゲ図の箱は，スコアの第一四分位点から第三四分位点まで（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．従って，この点の示す位置が高い程，当該属性値モデルが正確であると言える．例えば，表の最左端は「現在地=神奈川県 A 市」の属性値モデルから算出された 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 名の被験者のスコアを表している．

表 4.3 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み合わせにおいて，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．例えば，特徴を bag-of-words，機械学習アルゴリズムを RandomForest とした場合，現住所の属性値モデルにおいて，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．順位の取り得る値は 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モデルの精度が正確であると考えられる．表 4.3 より，特徴として binary よりも bag-of-words を用いた方が属性値モデルの精度が高く，bag-of-words を用いた場合の属性毎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，以上の結果を踏まえ，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，以上の全通りの組み合わせ 30 ケースについて評価した．

表 4.4 は，機械学習アルゴリズムとして RandomForest，使用する属性の組み合わせとして全属性，スコア統合方法に平均を用いた場合の，6 つの履歴書モデル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モデルの値）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．

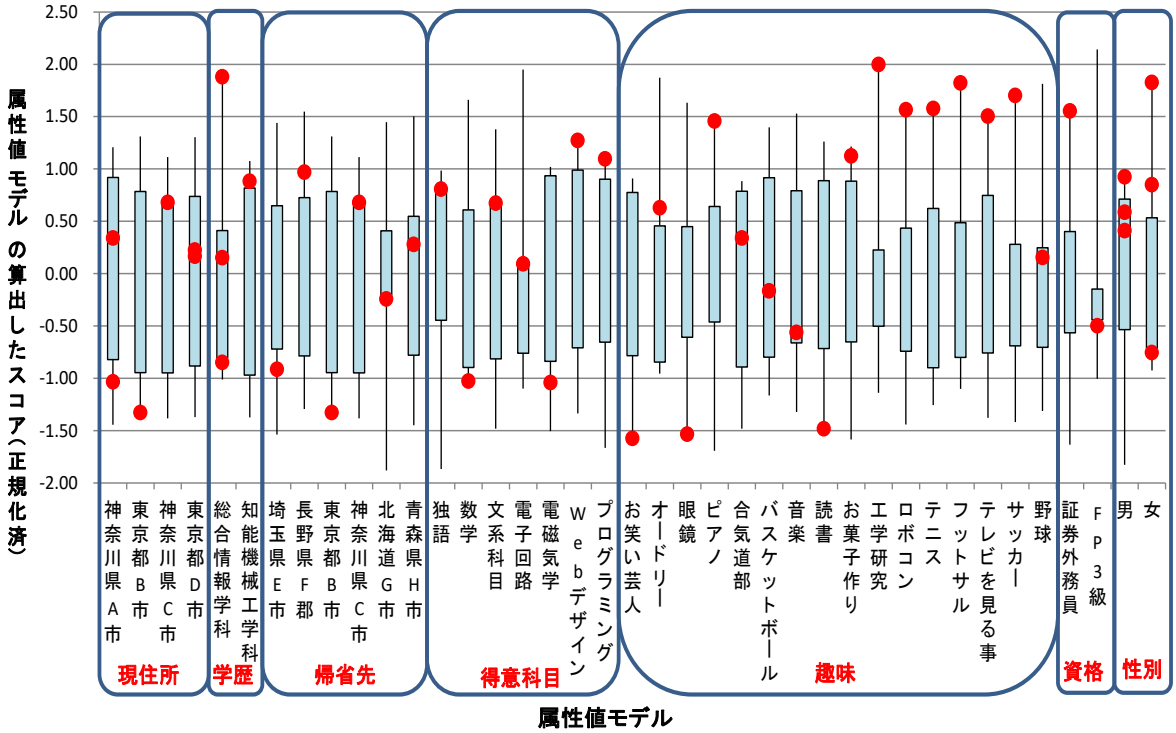


図 4.7　各属性値モデルのスコア分布 (Bag-of-words と RandomForest を用いた場合)

表 4.3　属性毎における本人のアカウントの平均順位

| 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 |

表 4.4　RandomForest，全属性，平均を用いた場合の各履歴書モデルのスコア

| | 履歴書No. | | | | | |
|---|---|---|---|---|---|---|
| アカウント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 4.5 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モデルのスコアが 6 つの履歴書モデルのスコアの中で 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 4.5 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み合わせは，

(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 ケースについて本評価を行った．

表 4.5　本人の履歴書が 1 位，2 位になった数

| 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*** |

## 4.6　本評価

### 4.6.1　再特定の精度

30 名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 個の属性値モデルを実装し，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 4.8 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，縦軸は各履歴書モデル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モデル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．例えば，被験者 1 のアカウントにおいて本人の履歴書モデル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 4.6 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．例えば，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．

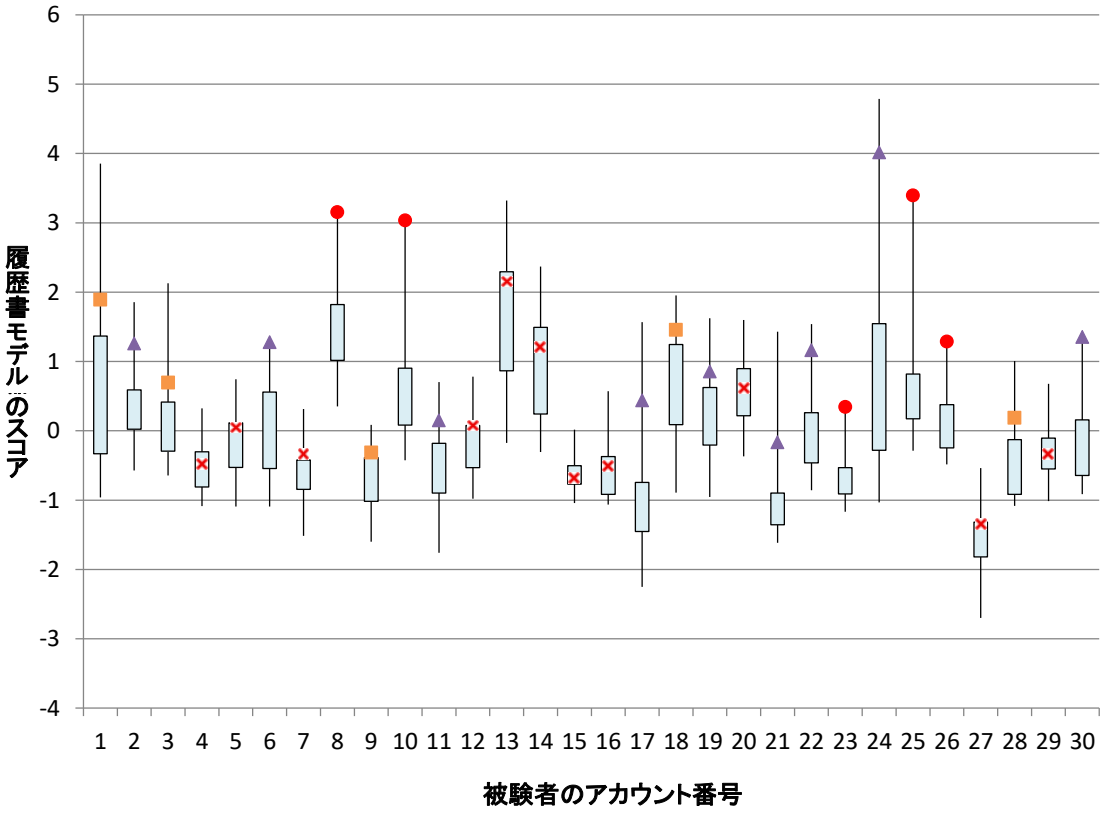


図 4.8　本人の履歴書が 1 位，2 位になった数
（RandomForest，4 つの属性，積の場合）

表 4.6　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

| 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) | ロジスティック回帰 | 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) |

本評価では，基本的な機械学習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 種類を採用したが，線形識別面を用いるロジスティック回帰と決定木を用いる RandomForest が最良であった．決定木を

用いるアルゴリズムは近年急激に進歩しており，XGBoost が高い精度を出している．そこで，さらに XGBoost を用いた評価を実施したところ，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，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．そこで，4.7 章の 2 次評価では機械学習アルゴリズムとして XGBoost，特徴量として bag-of-words，スコア統合方式として平均を用いた．

### 4.6.2　スケーラビリティの評価

本評価では，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，機械学習により属性値モデルを生成した．これらの処理時間を表 4.7 に示す．なお，測定に用いたハードウェアは，Intel Core i7，4 コア，3.4GHz，16GB メモリの PC1 台であり，ネットワークは 1Gbps の有線 LAN である．

表 4.7 の No.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，高々，候補者（すなわち履歴書）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，処理時間は高々候補者の数に比例すると考え，表 4.7 の処理時間から推定すると下記の式のように，ソーシャルメディア上で問題発言が見つかってから，3 日の間に，530 人の候補者に対する属性値モデルを生成することができる．

$$\frac{\frac{24\ \text{時}}{1\ \text{日}} \times 3\ \text{日}}{\frac{4.0378\ \text{時間}}{\text{人}}} \fallingdotseq 530\ \text{人}$$

上記の処理は属性値毎に並列化できるので，複数台の PC の利用により，さらに多くの属性値モデルを生成可能である．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属性値モデルを生成する場合もありうる．

表 4.7　提案手法の処理時間の内訳

| 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 3,000 件までのつぶやきをダウンロード | 1 時間 39 分 25 秒 |
| 2 | 属性値モデル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モデルを生成（アルゴリズムは GBDT） | 41 分 40 秒 |
| | 合計 | 4 時間 2 分 16 秒 |

### 4.6.3　分析

本評価において，被験者 10 及び被験者 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挙げられる．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，履歴書モデルの更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 の照合率は全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…」等）が多く見られた．これらのつぶやきは本人の履歴書と全く関係が無く，このような場合，履歴書から属性値モデル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及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取る必要があると考えられる．

## 4.7　2 次評価

### 4.7.1　データセット

本評価では，被験者が 30 名と少なく，また全員が電気通信大学の学生であるため一般性に乏しかった．そこで，本評価では，電気通信大学の学生 27 人，外部被験者 51 人から成る 78 人の被験者のデータを用いた．図 4.9 に示すように，78 人の被験者は予備評価の 27 人の被験者に比べると多様であり，一般性が向上している．78 人のツイート数は最小で 535 ツイート，最大で 3,222 ツイートであり，平均 2351.85 ツイートであった．

属性値モデルを学習するための教師データについては，本評価では正例と負例を各々 30 アカウントとしていたが，2 次評価では，より多くの正例および負例アカウントを収集した．正例および負例アカウントの使用数については，4.7.2 節で検討結果を述べる．

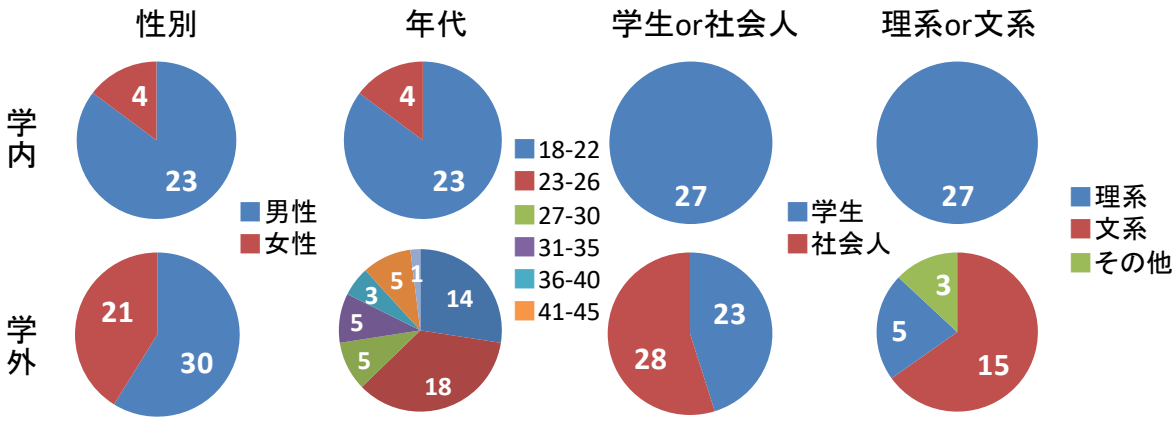


図 4.9　データセットの内訳

### 4.7.2　属性値もでとプロファイリング結果

被験者 78 人の履歴書からは，487 件の属性値が抽出された．最低でも 1 属性に対して 30 件以上の正例を収集できる属性値に絞った結果，最終的に 420 件の属性値モデルを構築した．

| | | 性別 | | 年代 | | 現住所 | 趣味 |
|---|---|---|---|---|---|---|---|
| | | 男性 | 女性 | 18-22 | 23-26 | 神奈川県F市 | プログラミング |
| アカウント番号 | 1 | **0.9995** | 0.0005 | **0.0714** | 0.4523 | 0.0047 | 0.5515 |
| | 2 | **0.9967** | 0.0033 | **0.5682** | 0.5839 | 0.1399 | 0.0119 |
| | 3 | **0.9770** | 0.0230 | **0.0224** | 0.7113 | 0.0771 | **0.0074** |
| | 4 | **0.9853** | 0.0147 | **0.3600** | 0.0505 | 0.0050 | **0.0163** |
| | 5 | **0.9788** | 0.0212 | **0.0173** | 0.0576 | 0.0242 | 0.7909 |
| | 6 | **0.9991** | 0.0009 | 0.0057 | **0.0654** | **0.9916** | **0.9251** |
| | 7 | **0.9961** | 0.0039 | **0.0345** | 0.9390 | 0.0402 | 0.9954 |
| | 8 | **0.9297** | 0.0703 | **0.3395** | 0.9688 | 0.0062 | 0.0057 |
| | 9 | **0.9970** | 0.0030 | **0.2699** | 0.1260 | 0.1358 | 0.0103 |
| | … | | | | | | |
| | 70 | 0.3936 | **0.6064** | 0.0118 | 0.1280 | 0.0927 | 0.0050 |
| | 71 | **0.9657** | 0.0343 | 0.0039 | 0.0296 | 0.0540 | 0.0079 |
| | 72 | 0.0405 | **0.9595** | 0.0073 | 0.2222 | 0.0651 | 0.0021 |
| | 73 | 0.9891 | **0.0109** | **0.1469** | 0.9790 | 0.0012 | **0.4646** |
| | 74 | 0.0152 | **0.9848** | 0.3671 | 0.8456 | 0.0011 | 0.0069 |
| | 75 | **0.9721** | 0.0279 | 0.0269 | **0.5871** | 0.1115 | 0.6344 |
| | 76 | **0.8931** | 0.1069 | 0.0416 | **0.9364** | 0.0067 | 0.2783 |
| | 77 | **0.9849** | 0.0151 | 0.5877 | 0.0225 | 0.1361 | 0.0008 |
| | 78 | 0.0046 | **0.9954** | 0.6591 | 0.8105 | 0.0129 | 0.0090 |


図 4.10　各アカウントにおけるプロファイリング結果のサンプル

図 4.10 はプロファイリング結果の一部を示す．各セルの数値は，当該行のアカウントが当該列の属性値を有する確からしさを表す．太文字の数値は，実際に当該アカウントの人物がその属性値を保有しているということを意味する．例えば，アカウント 6 の人物は，性別が男性，年代が 23-26，現住所が神奈川県 F 市，趣味がプログラミングである．この人物に対する各属性値モデルの結果は，男性である確からしさ，現住所が神奈川県 F 市である確からしさ，趣味がプログラミングである確からしさがそれぞれ 0.90 以上と高い精度で推定できている．一方で，年代が 23-26 である確からしさは 0.10 以下となっており，推定精度が低い．最終的に，これらの確からしさを統合することで，アカウント 6 の人物に対する本人らしさを表す履歴書スコアを算出する．

このようにして，プロファイリングではそれぞれのアカウントに対して，履歴書毎の属性値スコアを統合したスコアを算出し，これを統合したスコアに基づいて，当該アカウントが各履歴書とどの程度合致しているかを判定する．

表 4.8　属性毎の正解率

| カテゴリー | 項目数 | 正解率（%） |
|---|---|---|
| 性別 | 2 | 87.1 |
| 年代 | 7 | 59.3 |
| 通勤先・職種 | 26 | 60.4 |
| 大学・学部・学科・専攻 | 20 | 80.8 |
| 現住所 | 40 | 65.0 |
| 帰省先住所 | 57 | 75.9 |
| 特技・趣味<br>（部活・サークル含む） | 174 | 64.7 |
| 得意科目 | 30 | 62.8 |
| 資格 | 24 | 57.6 |

表 4.8 では，属性値を 9 カテゴリーに分類し，カテゴリー毎のプロファイリング精度を示している．最も精度良くプロファイリングできる属性は性別であった．続いて，大学・学部・学科・専攻，帰省先住所，現住所，特技・趣味，得意科目，通勤先・職種の順に精度良くプロファイリングできた．資格，年代のプロファイリング精度は低かった．

### 4.7.3　個人再特定の結果

(1) 基本問題

78 人の匿名アカウントをプロファイリングし照合した結果，表 4.9 に示すように，78 人中 38 人が本人の履歴書と正しく特定された．また，65 人が全体の上位 10%（7 位）以内，74 人を上位 20%（15 位）以内に絞り込むことができた．

(2) 逆問題

表 4.9 に示すように，被験者 78 人の履歴書と，被験者のものである 78 件アカウントを照合した結果，43 人の履歴書が本人のアカウントと正しく照合された．つまり，履歴書毎に計算される，各アカウントとの照合度合いを示すスコアについて，43 人の履歴書において，本人のアカウントとの照合スコアが 1 位であった．また，62 人が全体の上位 10%以内，73 人を上位 20%に絞り込むことができた．

(3) 1 対 1 問題

匿名アカウントと履歴書を 1 対 1 で照合した結果，39 人の履歴書とアカウントが正しく照合された．

表 4.9　78 人の被験者のアカウントと履歴書の照合結果

| | 基本問題 | 逆問題 | 1対1問題 |
|---|---|---|---|
| Top | 38(48.8%) | 43(55.1%) | 39(50.0%) |
| Top 10% | 65(83.8%) | 62(79.4%) | – |
| Top 20% | 74(94.8%) | 73(93.5%) | – |

### 4.7.4　使用する投稿数を減らした場合の評価

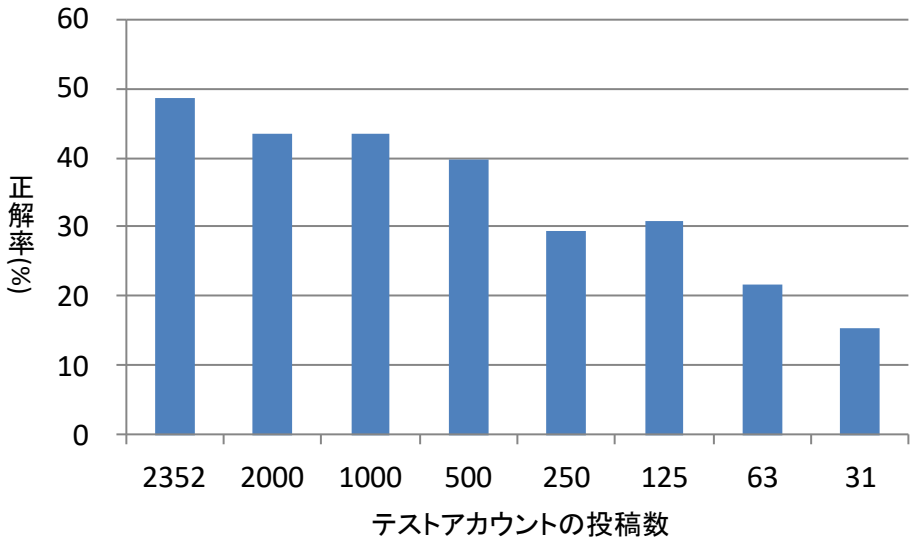


図 4.11 全 78 人の被験者アカウントに対して投稿文を削減した場合の正解率

今回，本評価のテストデータとして使用した 78 名の被験者のソーシャメディアへの投稿数は最大で 3,222 件，平均 2,351.85 件であった．しかし，本手法に対して，今回使用した被験者以外のアカウントをテストデータとして使用する場合，今回の被験者と同程度の投稿数が担保されるとは限らない．そのため，テストデータとして使用するアカウントの投稿数が少ない場合の本手法の精度について評価するため，被験者のアカウントの投稿数を削減した場合に，正しく被験者本人の履歴書を特定できた数（正解数）ついて評価した．

図 4.11 は，テストデータとして，各被験者アカウントの投稿数を削減していった場合に，正しく本人の履歴書と照合されたアカウントの数を示している．全アカウントに対して，最新の投稿文 2,352 件をテストデータとして使用する場合，全ての投稿文を使用した場合と比較して，正解数はそれほど減少しなかった．また，最新の 63 件の投稿文のみをテストデータとして使用した場合でも，78 名中 17 名（21%）が正しく照合された．これらの結果は，特定対象のアカウントについて，利用可能な投稿数が少ない場合でも，個

人特定に繋がる可能性を示している．

### 4.7.5　履歴書からアカウントへの照合

履歴書から本人のアカウントを特定したい場合，本人のアカウントの候補者は，保持している履歴書の数よりも多くなることが想定される．よって，本人のアカウントの候補者が多数存在する場合における，本手法の精度を評価するため，以下の通り実験を行った．

まず，被検者のアカウント以外に多数のノイズアカウントを用意し，その中から正しく本人のアカウントを特定できるかを確認した．具体的には，78 人の被験者のアカウントを N 個の不特定多数のアカウントの中に混ぜ込んだ後，78 の履歴書から（N+78）のアカウントへの照合を行い，被験者の 78 アカウントに正しく紐付けされる率を評価した．N 個のノイズアカウントは，不特定多数の公開された Twitter アカウントとし，その投稿文を用いた．

ランダムに収集された N=1,000 個のノイズアカウントと 78 名の被験者の履歴書をテストデータとして用いた場合，被験者 78 名中 9 名（12%）の履歴書が正しく照合された．また，59 名の履歴書の照合スコアが上位 100 位以内で照合された．100,000 件のノイズアカウントを追加した場合，12 名の履歴書（15%）が照合スコアの上位 100 位以内で照合された．これらの結果は，履歴書の人物のアカウントが絞り込めず，多数の候補の中から本人のアカウントを検知する場合でも，本人のアカウントが全体の 100 位以内に絞り込まれることで，これらの 100 件の候補アカウントのみを作業者が精査すれば，本人のアカウントを特定できる可能性を示している．

表 4.10　ノイズアカウントを追加した場合のアカウントと紐付いた履歴書の数

| ノイズアカウントの数（N） | 0 | 1,000 | 10,000 | 100,000 | 300,000 |
|---|---|---|---|---|---|
| Top | 16 (21%) | 9 (12%) | 5 (6%) | 4 (5%) | 1 (1%) |
| 100 位以内 | 78 (100%) | 59 (76%) | 20 (26%) | 12 (15%) | 7 (9%) |

## 4.8　提案方式の有用性

2 章で述べたように，再特定に関する従来手法はアカウント間の照合であるため，再特定対象のアカウントとその他の開示されているアカウントを用意する必要があった．また，再特定の対象者と同一人物のアカウントを推定できたとしても，実世界における個人を特定することにはならないという問題点があった．

提案手法では，アカウントと，履歴書から抽出されたアカウント保持者のプロファイルを照合させることによって，上記の課題を解決した．

ソーシャルメディア同士の照合ではなく，ソーシャルメディアと履歴書を照合するため，複数の公開アカウントの用意が不要である．また，履歴書は実在の人物を直接かつ一意に表すため，実世界の個人と直結する．よって，既知の履歴書との照合は実世界の個人特定となる．また，殆どの企業や学校，公共機関は，職員や学生の履歴書または相当情報を保有しているため，提案手法を適用することができる．

また，提案手法では，本人の情報として履歴書を用いることにより，匿名アカウントの人物特定だけでなく，従来手法では不可能であった既知の人物の匿名アカウントを特定する逆向きの照合にも対応した．ある実在する人物を特定したい場合，その人物の履歴書相当の情報を入手できれば，候補となる匿名アカウントをプロファイリングした情報に紐づけることで，どの匿名アカウントの人物が特定したい人物のアカウントであるかを推定することができる．

特定の対象者についてヒントとなる情報（たとえばワインが趣味）が得られた時に，従来手法ではこれを利用できなかった．提案手法では，ヒント情報を追加の属性値として，属性値モデルを生成することにより，個人特定の精度を向上することができる．すなわち，その情報に関して，公開されたソーシャルメディアの正例・負例（ワインが趣味のアカウントとそれ以外）を入手できれば，それらを用いて属性値モデルを生成し，個人特定に追加利用することができる．

提案法を用いてソーシャルメディアのプライバシーに関して警鐘を鳴らす場合の有用性を考察する．匿名アカウントから履歴書を通じて発言者個人が特定される基本問題と，個人から履歴書を通じて匿名アカウントが特定される逆問題について考える．

基本問題の場合は，特定対象の履歴書の母集団が小規模の場合でも現実的なリスクが存在する．例えば，ある企業の一つの課の職員が匿名アカウントで内部告発を行い，その告発内容から当該課の職員 10 名のうちの誰かが告発者であると推定できる場合には，10 名の履歴書の中から 1 名を特定することが現実的なリスクとなる．そのため，小規模でも提案法によって個人を特定できれば警鐘を鳴らす意味がある．

より有用な警鐘ツールとするためには，匿名アカウントと履歴書が同一人物である確からしさ（3.5 節の$S_i^j$が数学的な確率であることが望ましい．現在の提案法では，複数の候補者の中で最も同一人物らしい人を特定するのみであり，同一人物である確率は算出できない．そのため，たとえば匿名アカウントと履歴書が 1 つずつの場合には，実際には同一人物ではない場合も照合されてしまうため，提案法は有効ではない．$S_i^j$を確率として算出できれば，このような場合でも警鐘として有用である．

逆問題の場合は，特定候補となるアカウント集合を絞り込むことが難しい．たとえば，就職希望者の匿名アカウントの候補を絞り込む場合，ソーシャルメディアで企業名等のキーワード検索により抽出すると，数百から数千アカウントになる．よって，逆問題の場合

は多数のアカウントの中から履歴書を特定できることが望ましい．表 4.10 では，76%の履歴書について，本人のアカウントを 1,000 アカウントのうちの 100 アカウントに絞り込むことができた．100 アカウント程度であれば，人間が精査することで，本人のアカウントを特定できる可能性がある．表 4.10 の場合，100 アカウントへの絞り込みは，ノイズアカウント数が 10,000 の場合で 26%，100,000 の場合でも 15%であり，警鐘としての有用性はあると考える．今後，さらなる精度の向上が望まれる．

## 4.9　まとめ

本章では，匿名のソーシャルメディアアカウントを特定する手法として，照合先の詳細なプロファイルとして履歴書を，ソーシャルメディアアカウントのデータとしてソーシャルメディアへの投稿文を利用した場合の実現方法及び評価結果を述べた．

従来のソーシャルメディアアカウントの再特定手法では，公開されているソーシャルメディアアカウント間を照合していたので，再特定の対象者が 2 つ以上のアカウントを開示していないと利用できなかった．また照合先のアカウントが実名でないと個人を特定することができなかった．さらに，対象者に関してヒントとなる情報を入手しても利用することはできなかった．提案手法は，匿名アカウントから対象者のプロファイリングを行い，企業などが保持している履歴書と照合するので，実名アカウントを含む 2 つ以上のアカウントの開示がなくても個人を特定することができる．また，ヒント情報をプロファイルの一部とみなすことで，再特定精度の向上に利用することができる．

予備評価によって，機械学習アルゴリズム，特徴量，利用する属性値，スコア統合方法を選定した後，30 名の被験者データによる評価を行った．その結果，匿名アカウントから履歴書を特定する基本問題では 11 個のアカウントを本人の履歴書に正しく紐づけることができた．

さらに，78 名の被験者データを用いた 2 次評価を実施した結果，匿名アカウントから履歴書を特定する基本問題では 38 個のアカウントを本人の履歴書に正しく紐付けることができた．履歴書からアカウントを特定する逆問題では 43 個の履歴書を本人のアカウントに正しく紐付けることができた．アカウントと履歴書を 1 対 1 に紐付ける問題では，39 個のアカウントと履歴書を正しく紐付けることができた．

# 第5章　移動履歴を用いた再特定の精度向上

## 5.1　はじめに

本研究の提案法は，ソーシャルメディアアカウントの情報を利用し，匿名のアカウント所有者を機械学習によってプロファイリングすることで，個人を特定する．4 章では，アカウントの投稿文から得られた単語の頻度情報を特徴量としてプロファイリングを行う方式について述べた．しかし，投稿文の情報を直接的に用いた 4 章の方式では，現住所の推定の正解率は 65.0%と低かった．

投稿文には地名が含まれる．この地名を緯度経度の位置情報に変換し，投稿時刻と共に時系列データとすることで，アカウントユーザが何時何処にいたかという移動履歴の情報を得ることができる．この移動履歴情報は，現住所との関連が強いと予想される．そこで，本章では，投稿文から変換した移動履歴を用いてプロファイリングを行う手法および，その手法を組み込んだ個人再特定の精度向上について述べる．

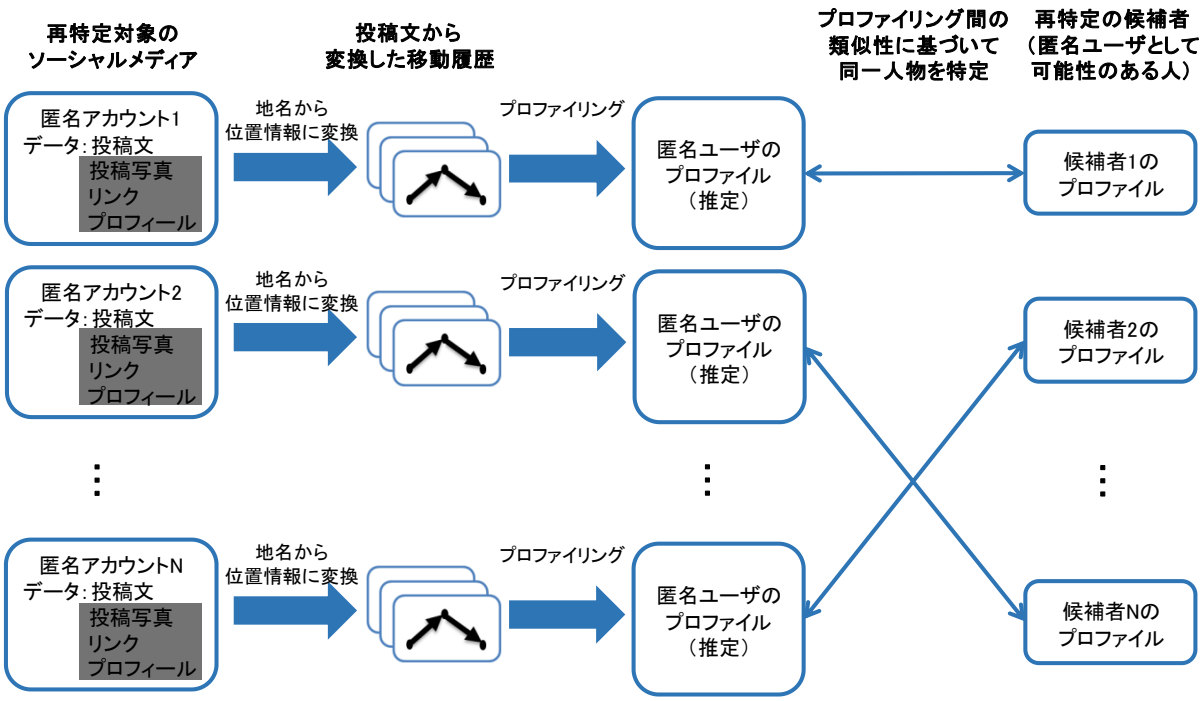


図 5.1 移動履歴を用いてプロファイリングを行う場合のシステム構成

## 5.2　位置情報を用いたプライバシーへの攻撃

本章では，移動履歴を用いてアカウントユーザの住所のプロファイリングと再特定を行

う．そこで，関連研究として，移動履歴を用いたプライバシーへの攻撃の研究をサーベイする．

移動履歴を用いたプライバシーへの攻撃には，個人の移動傾向の推定 [90] [91] [92]，移動先 [93] [94] [95]，移動手段 [96] [97] [98]，興味を持つ場所 [99]，現在行っている行為 [81]および移動人物の属性の推定 [100]，移動人物の再特定 [101] [102] [103] [104] 等がある．これらのうち本研究に関連の深い移動人物の属性推定と再特定について詳しく述べる．

人の属性の推定の例として，Matsuo らは，SVM によるドキュメント分類手法を応用し，研究所内の移動履歴からの年齢カテゴリー，仕事の種類および所属グループ，喫煙の有無，出勤形態等を推定した [100].

移動履歴からの再特定方式は，グラフの照合による方式 [67] [105]，場所間の遷移に基づく方式 [77] [80]，場所毎の訪問回数に基づく方式 [106] [78]，ノイズの位置の確率分布に基づく方式 [79]に分類できる．

グラフの照合による方式の例として，Srivatsa らは，複数の移動履歴間の接触関係のグラフと複数のソーシャルネットワークアカウント間のリンク関係のグラフを照合することにより，移動履歴と同一人物のアカウントを特定した [67]．Manousakas らは，移動履歴から，場所をノードとし場所間の遷移をエッジとするグラフを生成し，グラフ間の照合により，同一人物の複数の移動履歴を対応付けた [105].

場所間の遷移に基づく方式の例として，Shokri らは，再特定対象者の位置や移動に関する知識（自宅や勤務先等）から，場所間の遷移確率行列の形で，移動の傾向（移動プロファイル）を生成する．各移動履歴と各人物の移動プロファイルのペアについて，移動プロファイルを条件とし，移動履歴の条件付き生起確率を算出することで，移動履歴と移動プロファイルを対応付けた [77]． Gambs らは，移動履歴の 2 つの集合について，各々の移動履歴から場所間の遷移確率行列を生成し，行列間の類似度に基づいて，2 つの集合から同一人物の移動履歴を特定した [80].

場所毎の訪問回数に基づく方式の例として，Riederer らは，複数の位置情報サービスの利用履歴から得られた移動履歴の集合から，同じ人物の移動履歴を推定した [106]．自然な移動履歴では，場所毎の訪問回数がポアソン分布に従うとし，移動履歴のペアが同一人物に属すると仮定した場合と異なる人物に属すると仮定した場合の生起確率の比に基づいて，同じ人物の移動履歴を推定した．Murakami は，背景知識が少ない場合の移動プロファイルとして，場所間の遷移確率の行列よりも，場所毎の訪問確率のベクトルの方が有効であることを明らかにした [78].

最後に，ノイズの位置の確率分布に基づく方式の例として，Ma らは，ランダムノイズに沿って位置を変更した移動履歴を対象とし，元の移動履歴に対応付けた [79].

## 5.3　移動履歴を利用した現住所のプロファイリング

### 5.3.1　基本アイディア

4 章で述べた現住所推定では，属性値モデルの学習にソーシャルメディアへの投稿文を用いていた．例えば，アカウントの持ち主が神奈川県藤沢市在住かを推定する場合，実際に藤沢市に住んでいる人物のソーシャルメディアへの投稿文を正例，それ以外の地域に住んでいる人物の投稿文を負例として，藤沢市在住かを推定するモデルを学習する．そして，学習したモデルに現住所が未知である人物の投稿文を入力し，その人物が藤沢市在住であるかをテストする．しかし，その手法では本人の現住所の推定精度は正解率が 65%程度と低かった（表 4.8）．

現住所の推定精度を向上するために，本研究では以下の 2 点に着目した．1 つ目は，ソーシャルメディアの投稿文から移動履歴を抽出できる点である．松本らは，ソーシャルメディアの投稿文から地名を抽出し，投稿時刻と合わせて疑似移動履歴に変換した．さらに，この疑似移動履歴が WiFi 基地局から収集した同一人物の真の移動履歴に類似していることを明らかにしている [88]．2 つ目は，移動履歴から現住所が推定できることが従来研究によって明らかになっている点である [92]．

以上を踏まえて，匿名アカウントの人物の現住所を推定する新しい方式を以下のように提案する．

(1) ある住所に住んでいるかを推定するモデルを学習する

　はじめに，当該住所に住んでいる人物のアカウントの投稿文から疑似移動履歴を推定する．次に，当該住所に住んでいない人物のアカウントの投稿文から疑似移動履歴を推定する．これらを正例および負例として，現住所を推定するモデルを学習する．ソーシャルメディアの投稿文から真の移動履歴に類似した疑似移動履歴を抽出できることから，これらを正例，負例として，現住所の推定モデルを一定の精度で生成できると考えられる．

(2) 現住所が未知である人物がその住所に住んでいる確からしさを推定する

　投稿者の現住所が未知であるアカウントの投稿文を疑似移動履歴に変換後，これらをテストデータとして(1)で作成したモデルに入力し，その現住所の地名が指す場所にいる確からしさを求める．未知の人物のアカウントの投稿文から当該人物の真の移動履歴に類似した疑似移動履歴を抽出できることから，この疑似移動履歴をモデルに入力することで，現住所を推定できると考えられる．

次に，移動履歴を用いた現住所推定モデルを，全体の目標である匿名のソーシャルメデ

ィアアカウントの個人特定に組み込む方法を述べる．例えば，藤沢市在住かを推定する場合，藤沢市在住の人物の投稿文を移動履歴に変換したものを正例，それ以外の現住所の人物の移動履歴を負例としてモデルを作成する．同様に，匿名アカウントの移動履歴をモデルへ入力し，匿名アカウントの人物が藤沢市在住かをテストする．現住所以外の属性については，既存手法と同様に，投稿文からそれぞれの属性値推定モデルを作成し，匿名のソーシャルメディアアカウントの投稿文を入力として，各属性値の推定スコアを得る．最後に，投稿文から学習したモデルと，移動履歴から学習したモデルから得られたスコアを組み合わせて個人特定を行う．この際，移動履歴情報を用いて現住所推定の精度を向上することによって，個人特定精度の向上を図る．

### 5.3.2　処理方式

5.3.1 節で述べた，移動履歴を用いた現住所推定モデルの学習方法，学習したモデルを用いた匿名アカウントの所有者の現住所推定方法をソーシャルメディアアカウントの個人特定に組み込む方法について，詳細を述べる．

例として，ソーシャルメディアアカウントの人物の現住所が，神奈川県藤沢市であるかを判定する場合を説明する．

- Step 1：現住所推定モデルの学習

  - Step 1-1：投稿文の収集

正例として，現住所が神奈川県藤沢市である人物の投稿文，負例として，現住所が神奈川県藤沢市でない人物の投稿文を収集する．

なお，ソーシャルメディアアカウントの収集には投稿文からプロファイリングした 4 章の方式と同様にツイプロ [89]を，投稿文の収集には Twitter API [107]を用いる．

例えば正例の場合，「藤沢駅前で友達に会った.」「新江ノ島水族館に行ってきた.」「鵠沼海岸で海の家が始まった.」といったように，地元住民が良く訪れる場所名が投稿文に含まれる可能性が高い．一方，負例の場合「新宿で飲み会をした.」「伊豆までドライブ.」「みなとみらいに遊びに行った.」といったように，神奈川県藤沢市の住民があまり訪れない地名が投稿文に含まれる可能性が高い．

  - Step1-2：投稿文の分析

投稿文中に含まれる地名と投稿時間を抽出する．上記で挙げた正例の投稿文の例からは，それぞれ「藤沢駅」「新江ノ島水族館」「鵠沼海岸」が地名として抽出できる．また，負例の投稿文の例からは，「新宿」「伊豆」「みなとみらい」が地名として抽出できる．なお，

地名の抽出には MeCab [108]の形態素解析機能を用いる．

➢　Step1-3：疑似移動履歴への変換

上記で投稿文から抜き出した地名を座標に変換する．これを時系列順に並べ替え，時刻情報を付加したものを疑似移動履歴とする．なお，地名から座標への変換も，GeoNLP の API [109]を用いる．具体的には，投稿文を MeCab [108]で形態素解析した後，地名にあたる単語を GeoNLP [109]に通すことで，その地名の座標を取得する．

➢　Step1-4：特徴ベクトルの生成

疑似移動履歴から特徴量を抽出する．まず，作成した疑似移動履歴を訪問先が関東圏内であるものに限定する．この移動履歴を，関東圏を 1km メッシュに区切った際の，メッシュ毎の訪問回数として成形する．なお，地名を位置情報に変換する際に利用している GeoNLP [109]では，市区町村名が市役所，都道府名は県府庁等の位置に変換される．市区町村や都道府県は数 km の範囲に及ぶため，地名から変換した位置情報と，実際の訪問先の位置情報に誤差が生じる．この誤差による影響を緩和するため，疑似移動履歴のメッシュのカウントロジックでは，該当メッシュに 1，上下左右のメッシュに$\frac{1}{2}$，斜め四方のメッシュに$\frac{1}{2\sqrt{2}}$カウントを加える．これにより，各メッシュへの訪問回数を特徴量とする訓練データが作成される．正例（藤沢在住者）および負例（藤沢非在住者）の各アカウントから 1 つずつ特徴ベクトルが生成される．

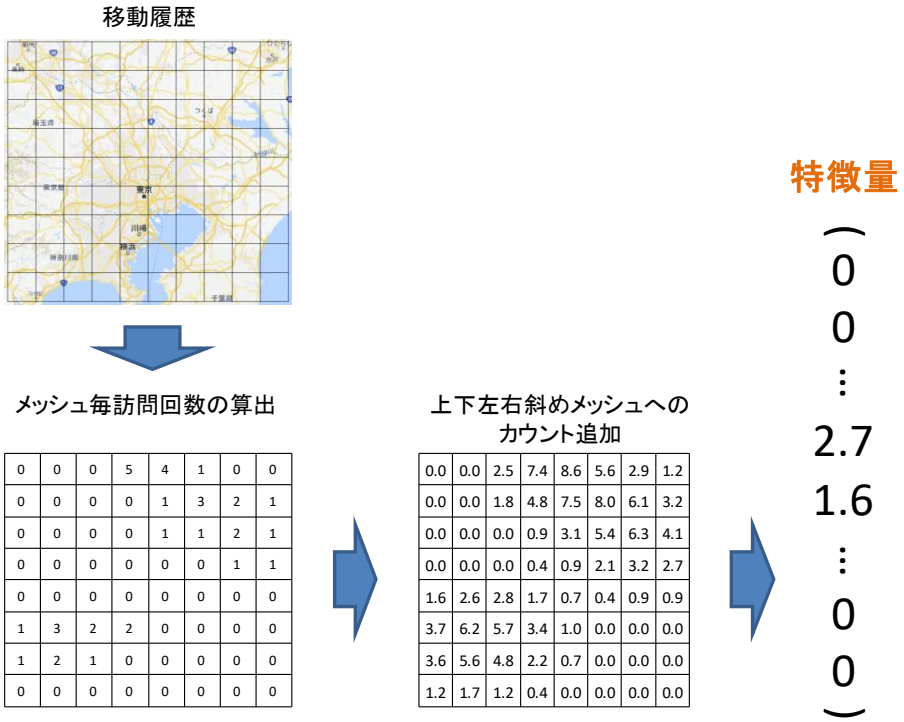


図 5.2　関東圏内のメッシュ

➢　Step1-5：モデルの学習

正例および負例の特徴ベクトルを用いて，機械学習アルゴリズムにより現住所推定モデルを学習する．

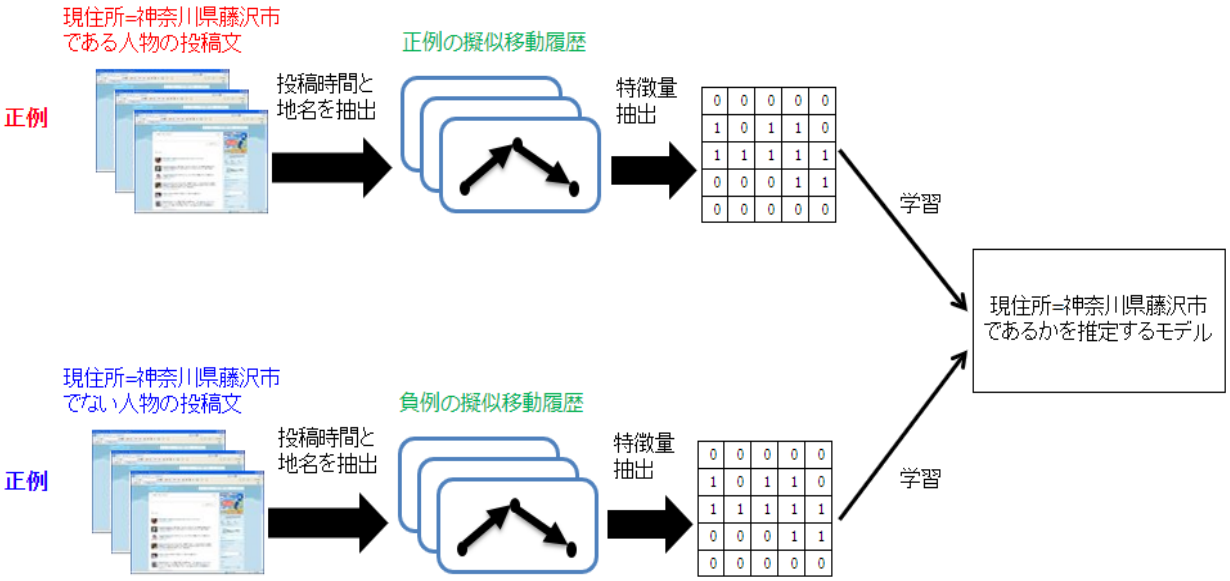


図 5.3　現住所が神奈川県藤沢市であるかを推定するモデルの学習方法

● Step 2：匿名アカウントの所有者の現住所推定

訓練データ作成課程の Step1-1～Step1-5 と同様に処理を行い，現住所が未知であるソーシャルメディアアカウントの人物の投稿文から作成した疑似移動履歴を用いて，現住所推定を行う．このテストデータを前述のモデルに入力することで，匿名アカウントの人物の現住所が神奈川県藤沢市である確からしさを得ることができる（図 5.4）．

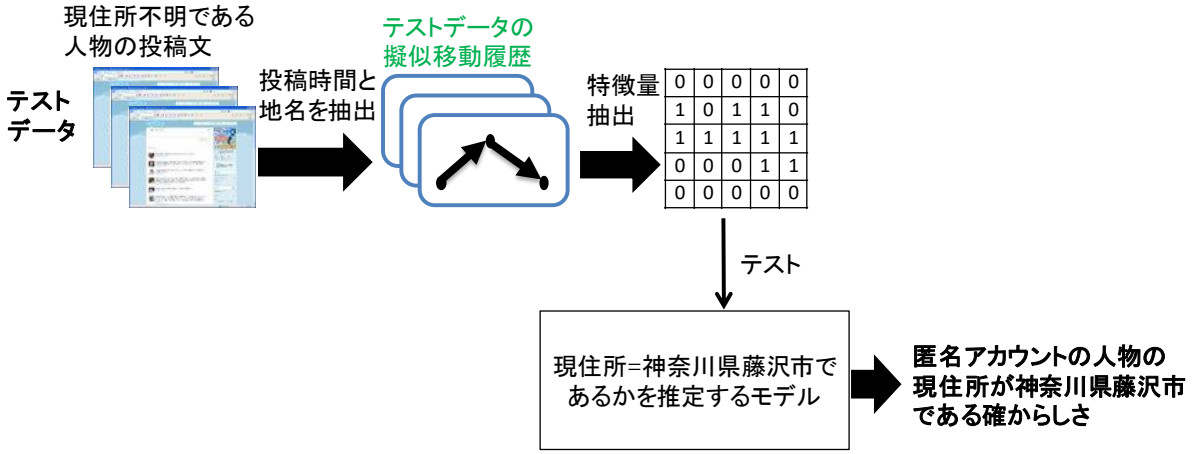


図 5.4　匿名アカウントの人物の現住所推定

候補となる現住所の各々について，神奈川県藤沢市の場合と同様の方法で推定モデルを学習する．各々の匿名アカウントから疑似移動履歴を抽出し，特徴ベクトルを生成して各現住所推定モデルに入力することで，匿名アカウントが，当該現住所エリアに居住している確からしさを得る．現住所以外の属性について，4 章で述べた方法により，各カウントが各属性値（性別，年代など）を有する確からしさを算出する．最後に，現住所に関する確からしさとそれ以外の属性値を有する確からしさを統合することによって，匿名アカウントの人物を特定する．具体的には，各属性値を有する確からしさから，匿名アカウントが履歴書に該当する確からしさ（統合スコア）を算出し，アカウント毎に統合スコアが最大となる履歴書を選択する（図 5.5）．

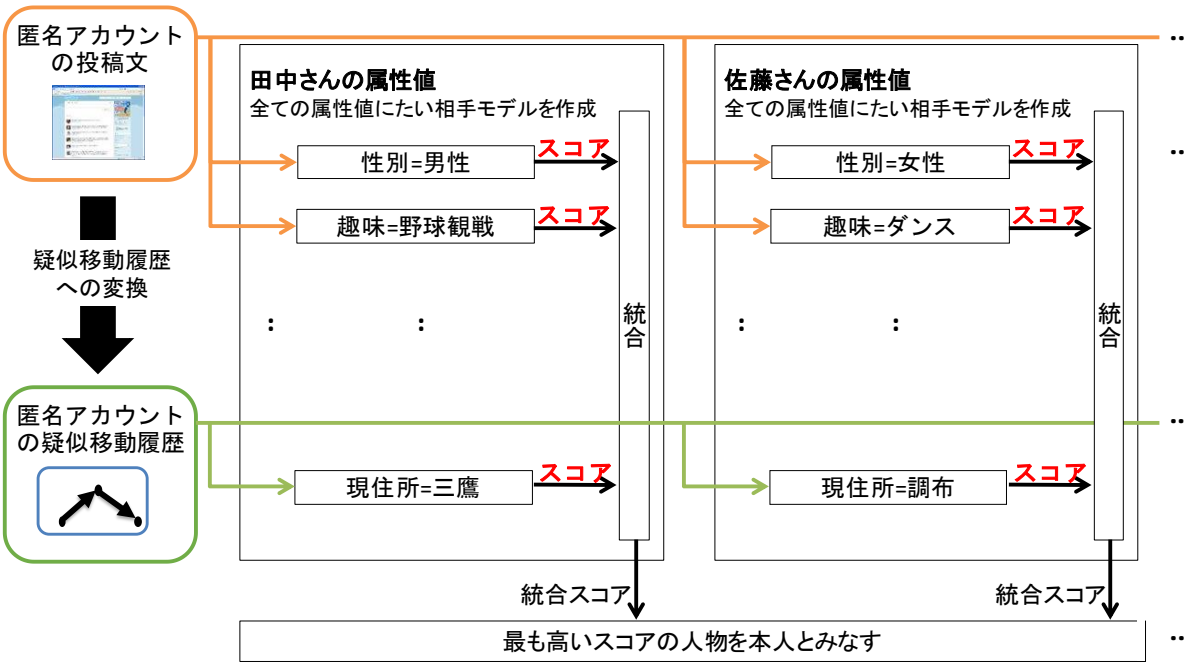


図 5.5　匿名アカウントの人物推定の概要

## 5.4　実装

以下の 4 つのプログラムにより提案手法を実装した．

(1) つぶやきを収集するプログラム．

言語は Python，行数は 350 行．ツールは Mecab を使用し，辞書には iPA 辞書と Wikipedia の見出し語を辞書化したものを使用した．

(2) ソーシャルメディアの投稿文を形態素解析し，モデルに入力するための特徴量に変換するプログラム．

言語は Python，行数は 515 行である．

(3) 地名を緯度経度に変換した後，既定されたメッシュ毎の移動頻度を集計するプログラム．

言語は Go 言語と Python，行数は計 1,411 行である．その内訳は，データ読み込み部分（位置情報の読み込み処理）112 行，設定ファイル読み込み部分（特徴量の範囲設定やメッシュの大きさ等の設定）364 行，データ書き出し部分（特徴量の出力処理 289 行，特徴量計算の関数が記述されている部分 349 行，特徴量の計算を実行する部分 297 行）である．

(4) 機械学習によって属性値モデルを学習し，アカウントの人物がその属性値を有するかを判別するプログラム．

プログラム言語は Python であり，行数は 612 行である．

実行環境は 3 台の PC であり，1 台目の PC は Linux Ubuntu 16.04.5 LTS（メモリサイズ：15GB / CPU コア数：4)，2 台目は Ubuntu 18.04.5 LTS（メモリサイズ：125GB / CPU コア数：24)，3 台目は Windows 10（メモリサイズ：16GB / CPU コア数：6）であった．

## 5.5　予備評価

疑似移動履歴を用いることが現住所推定に有効であるかを確認するための予備評価を行った．具体的には，二つの異なる現住所を識別するモデルの精度を評価した．取り上げた二つの現住所の内，一方の現住所を持つアカウントを正例，他方を負例とし，それぞれ N 個ずつ（計 2N 個）取得した．正例・負例からそれぞれ N-1 個（計 2N-2 個）のアカウントを訓練データとしてモデルを学習し，残りの 1 個（計 2 個）のアカウントをテストデータに用いてモデルの精度を評価した．この N 分割交差検定を繰り返すことで，テストデータの現住所を正しく推定できるか確認した．この実験を 4 章で述べた投稿文から現住所を推定する方式と，5.3 節で述べた投稿文を疑似移動履歴に変換して現住所を推定する方式で行った．また，これらの 2 つの方式の組合せを評価した．評価した方式は以下の 5 通りである．機械学習アルゴリズムとしては XGBoost，特徴量として bag-of-words を用いた（表 5.1)．

表 5.1　予備評価で比較する方式

| CASE. | 概要 | 特徴量 |
|---|---|---|
| CASE 1-1 | 投稿文から推定(4章の方式) | 投稿文<br>(以下では投稿文と略す) |
| CASE 1-2 | 疑似移動履歴から推定(5.3節の方式) | 疑似移動履歴 |
| CASE 1-3 | CASE1とCASE2の結果を平均したもの | 文&移by平均 |
| CASE 1-4 | CASE1とCASE2の結果を正規化後平均したもの | 文&移by正規化&平均 |
| CASE 1-5 | CASE1の特徴ベクトルとCASE2の特徴ベクトルを連接した特徴ベクトルを利用 | 文&移by連接 |

識別する現住所の組合せとして，「神奈川県横浜市/神奈川県藤沢市」，「神奈川県横浜市/茨木県結城市」，「東京都調布市/埼玉県さいたま市」，「東京都調布市/東京都足立区」「東京都北区/東京都足立区」「東京都北区/東京都大田区」「東京都足立区/東京都大田区」「埼玉県さいたま市/茨木県結城市」の計 8 通りを用いた．

### 5.5.1　予備評価データ

正例，負例として使用するツイートは，Twitter のプロフィール欄にその属性値に関する語句を書いているアカウントから収集した．このとき，1,000 件以上投稿しているユーザアカウントを採用し，1 アカウントあたり最新の 3,000 ツイートまでを収集した．識別する現住所の組合せに対して，最小で 21 アカウントずつ（埼玉県さいたま市 21 アカウント，茨木県結城市 21 アカウント），最大で 78 アカウントずつを用いた．各モデルに使用する正例・負例のアカウント数は，双方の内少ない方に合わせた．

### 5.5.2　結果

表 5.2　予備評価の結果例

| | CASE1-1 | | CASE1-2 | | CASE1-3 | | CASE1-4 | | CASE1-5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| テストに使用したペア | 神奈川県横浜市(label0) | 神奈川県藤沢市(label1) | 神奈川県横浜市(label0) | 神奈川県藤沢市(label1) | 神奈川県横浜市(label0) | 神奈川県藤沢市(label1) | 神奈川県横浜市(label0) | 神奈川県藤沢市(label1) | 神奈川県横浜市(label0) | 神奈川県藤沢市(label1) |
| ペア1 | 0.5873 | 0.9894 | 0.5288 | 0.2377 | 0.5581 | 0.6136 | -0.1956 | 0.5973 | 0.6741 | 0.9560 |
| ペア2 | 0.0347 | 0.2267 | 0.2319 | 0.1633 | 0.1333 | 0.1950 | -1.1584 | -0.7729 | 0.1179 | 0.0636 |
| ペア3 | 0.0027 | 0.6510 | 0.0122 | 0.9447 | 0.0074 | 0.7978 | -1.2053 | 0.7153 | 0.0040 | 0.8903 |
| ペア4 | 0.0139 | 0.9980 | 0.0101 | 0.9917 | 0.0120 | 0.9948 | -1.1918 | 1.1743 | 0.0051 | 0.9856 |
| ペア5 | 0.0368 | 0.7030 | 0.5739 | 0.8518 | 0.3053 | 0.7774 | -0.1752 | 0.2377 | 0.2007 | 0.7648 |
| ペア6 | 0.0175 | 0.9957 | 0.0549 | 0.8393 | 0.0362 | 0.9175 | -1.1521 | 1.0814 | 0.0074 | 0.9859 |
| ペア7 | 0.0127 | 0.8746 | 0.0129 | 0.5091 | 0.0128 | 0.6918 | -1.1917 | 0.8055 | 0.0086 | 0.8053 |
| ペア8 | 0.9959 | 0.0504 | 0.8958 | 0.7915 | 0.9459 | 0.4210 | 1.1054 | 0.0441 | 0.6852 | 0.9375 |
| ペア9 | 0.0138 | 0.9960 | 0.0108 | 0.5304 | 0.0123 | 0.7632 | -1.1964 | 0.6429 | 0.0057 | 0.9721 |
| ... | | ... | | ... | | ... | | ... | | ... |

表 5.2 は，予備評価結果の一部を示す．本表は，「神奈川県横浜市」のラベルを 0，「神奈川県藤沢市」のラベルを 1 としてモデルを学習した場合の，CASE1-1～CASE1-5 の評価方法の結果をまとめたものの一部である．なお，「神奈川県横浜市」，「神奈川県藤沢市」の現住所を持つアカウントとしてそれぞれ 54 個（計 108 個）を用いた．

予備評価では，テストデータとして，「神奈川県横浜市」と「神奈川県藤沢市」のアカウントから各 1 個ずつ，計 2 個のアカウントを選択し，これを 1 つのペアとした．ペアは，両現住所アカウントで重複しない組合せとし，全 54 ペア作成した．訓練データは，各ペアでテストデータとして用いる 2 個のアカウントを除いた，計 106 個のアカウントを用いた．

各行は，全 54 のペアの内，どのペアをテストデータに使用したかを示す．「神奈川県藤沢市」の列は，テストデータの現住所が神奈川県藤沢市である確からしさを示す．確からしさは，各サンプルの予測結果が，ラベル 0，ラベル 1 のどちらに近いかを示す数値である．つまり，テストデータの現住所が，予測結果が 0 に近いほど「神奈川県横浜市」，1 に近いほど「神奈川県藤沢市」である確からしさが高い．

例えば，CASE1-1 の 1 行目のケースでは，「神奈川県横浜市/神奈川県藤沢市」の全 54 のペアの内，1 つ目のペアをテストデータとして使用した場合の確からしさを示している．

この場合，「神奈川県藤沢市」が現住所であるサンプルの確からしさは 0.9894 であり，正解ラベル 1 に近い値が出力されている．しかし，「神奈川県横浜市」が現住所であるサンプルの予測確からしさは 0.5873 と横浜を示す 0 よりも藤沢を示す 1 に近い値が出力されており，正しく推定できていない．一方で，CASE1-1 の 3 行目のケースでは，「神奈川県横浜市」のサンプルは 0,「神奈川県藤沢市」のサンプルは 1 に近い値が出力されているため，正しく推定できていると言える．

表 5.3　予備評価結果のまとめ

| | | CASE1-1 投稿文 | | CASE1-2 疑似移動履歴 | | CASE1-3 文＆移by平均 | | CASE1-4 文＆移 by正規化＆平均 | | CASE1-5 文＆移by連節 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 現住所の組合せ | ペア数 | 正解数 | 割合 | 正解数 | 割合 | 正解数 | 割合 | 正解数 | 割合 | 正解数 | 割合 |
| 横浜/藤沢 | 54 | 36 | 0.67 | 27 | 0.50 | 42 | 0.78 | 40 | 0.74 | 37 | 0.69 |
| 横浜/茨城 | 21 | 15 | 0.71 | 19 | 0.90 | 18 | 0.86 | 18 | 0.86 | 18 | 0.86 |
| 調布/埼玉 | 47 | 35 | 0.74 | 41 | 0.87 | 43 | 0.91 | 42 | 0.89 | 41 | 0.87 |
| 調布/足立 | 47 | 39 | 0.83 | 39 | 0.83 | 44 | 0.94 | 44 | 0.94 | 43 | 0.91 |
| 北区/足立 | 57 | 32 | 0.56 | 35 | 0.61 | 37 | 0.65 | 39 | 0.68 | 41 | 0.72 |
| 北区/大田 | 78 | 58 | 0.74 | 54 | 0.69 | 72 | 0.92 | 70 | 0.90 | 73 | 0.94 |
| 足立/太田 | 57 | 40 | 0.70 | 41 | 0.72 | 50 | 0.88 | 49 | 0.86 | 52 | 0.91 |
| 埼玉/茨木 | 21 | 12 | 0.57 | 19 | 0.90 | 18 | 0.86 | 15 | 0.71 | 19 | 0.90 |
| 正解率の平均 | | | 0.69 | | 0.75 | | 0.85 | | 0.82 | | 0.85 |

表 5.3 に予備評価の結果をまとめる．予備評価の結果，各 CASE のうち最も正解率の平均が高かったものは，正解率 85%の CASE1-3（文&移 by 平均）と CASE1-5（文&移 by 連節）の二通りであった．また，CASE1-1（投稿文から推定）の結果が，他の 4 つの評価方式よりも正解率が悪いことが分かる．よって，現住所推定には疑似移動履歴の利用が有効であると言える．

## 5.6　本評価

### 5.6.1　現住所のプロファイリング

予備評価では，現住所の異なる 2 つのアカウントのペアを用いて，正しい住所を推定できるか評価し，疑似移動履歴を特徴量として利用することが現住所推定に有効であることを示した．本評価では，78 人の被験者の各現住所について，当該現住所に住む人物を正しく推定できるかを評価する．

評価した方式は，予備評価で評価した 5 つの方式のうち，CASE1-1（投稿文から推定），CASE1-2（疑似移動履歴から推定），精度が最も高かった 2 つの組合せ CASE1-3（文&移 by 平均）と CASE1-5（文&移 by 連節）である．

CASE1-1（投稿文から推定）と同様の方式として CASE2-1 を検討した．また，CASE1-2（疑似移動履歴から推定）に対して 2 種類の機械学習アルゴリズムと 2 種類の特徴量を組合せた方式として CASE2-2，CASE2-3，CASE2-4，CASE2-5 を検討した．機械学習アルゴリズムは，XGBoost に加え，4.5 節の評価で XGBoost に次ぐ精度であったロジスティック回帰を用いた．特徴ベクトルは，予備評価で用いた場所の訪問回数（Bag-of-words）に加え，訪問の有無（Binary）を用いた．なお，学習データの正例と負例の数の相違に対処する技術である Bagging を取り上げた．さらに，CASE1-3（CASE1-1 と CASE1-2 の平均）と CASE1-5（CASE-1-1 の特徴ベクトルと CASE1-2 の特徴ベクトルを連節した特徴ベクトルを利用）と同様の方式と，CASE1-3 に関して Min-Max scaling を適用した方式を採用した．合計で 8 通りの評価を行った（表 5.4）．

なお，Bagging はブートストラップサンプリングを繰り返して生成した多数の弱識別器を合成し，より判別精度の高い識別器を生成する手法である．正解と負例のサンプル数に偏りがある場合，Bagging が効果的に働く可能性がある．特徴ベクトルについて，Bag-of-words は加工していない基本的な値であるが，場所毎の訪問回数が大きく異なる場合，訪問回数の大きい場所だけが支配的になり訪問回数が少ないが特徴的である場所（他の人が行かない場所など）が無視されてしまう．この問題に対処するための標準的な手法として，Binary を用いた．8 つの方法を以下にまとめる．

表 5.4　本評価で比較する方式

| CASE . | 概要 | 特徴量 |
|---|---|---|
| CASE 2-1 | 投稿文から推定(4章の方式) | 投稿文<br>(以下では投稿文と略す) |
| CASE 2-2 | XGBoost+Baggingを用いた疑似移動履歴から推定 | Bag-of-words(移/Xgb/Bag) |
| CASE 2-3 | XGBoost+Baggingを用いた疑似移動履歴から推定 | Binary(移/Xgb/Bin) |
| CASE 2-4 | ロジスティック回帰+Baggingを用いた疑似移動履歴から推定 | Bag-of-words(移/Log/Bag) |
| CASE 2-5 | ロジスティック回帰+Baggingを用いた疑似移動履歴から推定 | Binary(移/Log/Bin) |
| CASE 2-6 | CASE2-1の特徴ベクトルとCASE2-2の特徴ベクトルを連節した特徴ベクトルを利用 | (文＆移by連節) |
| CASE 2-7 | CASE2-1の特徴ベクトルとCASE2-2の特徴ベクトルを連節し，特徴ベクトルをMin-Max scalingしたものを利用 | (文＆移by連節/Scale) |
| CASE 2-8 | CASE2-1とCASE2-2の平均 | (文＆移by平均) |

最初に，「神奈川県横浜市」「東京都北区」「神奈川県藤沢市」「東京都足立区」「埼玉県さいたま市」「東京都調布市」「茨城県結城市」「東京都大田区」「千葉県習志野市」「千葉県松戸市」の計 10 通りを用いて評価した．

CASE2-7 において，CASE2-1 の特徴ベクトルは投稿文の単語の出現頻度である．CASE2-2 の特徴ベクトルは緯度経度に基づいて地域をほぼ同じ大きさに区分けした各メッシュへの移動頻度である．よって，CASE2-1 と CASE2-2 の特徴ベクトルはそれぞれスケールが異なる．そこで，CASE2-1 の特徴ベクトルと CASE2-2 の特徴ベクトルを各々Min-Max scaling により正規化してサイズを合わせた後，特徴ベクトルを連節した．Min-Max scalling は，全ての値が 0 から 1 の間に位置するように元のデータを変換する正規化手法である．

表 5.5 に，CASE2-1 から CASE2-8 の評価結果を示す．表 5.5 における正解率は，実際に正例であるサンプルが正しく予測できている割合と，実際に負例であるサンプルが正しく予測できている割合を平均した数値である．つまり，正解率が高い方式が，匿名アカウントの現住所を正しく推定できていると言える．

表 5.5　各手法における正解率

※網掛けは，各現住所で最も正解率の高かった方式

| | 投稿文 | 移/Xgb/Bag | 移/Xgb/Bin | 移/Log/Bag | 移/Log/Bin | 文＆移 by連節 | 文＆移by 連節/Scale | 文＆移 by平均 |
|---|---|---|---|---|---|---|---|---|
| 神奈川県 横浜市 | 0.7589 | 0.811 | 0.8178 | 0.7178 | 0.7247 | 0.811 | 0.8562 | 0.8589 |
| 東京都 北区 | 0.7662 | 0.9026 | 0.8766 | 0.8117 | 0.3442 | 0.9026 | 0.8701 | 0.8247 |
| 神奈川県 藤沢市 | 0.9675 | 0.9545 | 0.961 | 0.4545 | 0.4481 | 0.9545 | 0.9481 | 0.961 |
| 東京都 足立区 | 0.7237 | 0.6776 | 0.4145 | 0.9408 | 0.8947 | 0.6776 | 0.4013 | 0.4934 |
| 埼玉県 さいたま市 | 0.9474 | 0.6842 | 0.6711 | 0.5921 | 0.6842 | 0.6842 | 0.625 | 0.7039 |
| 東京都 調布市 | 0.7587 | 0.769 | 0.6921 | 0.7548 | 0.7032 | 0.769 | 0.6563 | 0.7873 |
| 茨城県 結城市 | 0.4805 | 0.474 | 0.487 | 0.474 | 0.4935 | 0.474 | 0.4805 | 0.487 |
| 東京都 大田区 | 0.4342 | 0.9211 | 0.9211 | 0.875 | 0.9342 | 0.9211 | 0.9342 | 0.6908 |
| 千葉県 習志野市 | 0.8896 | 0.9286 | 0.9481 | 0.8506 | 0.9156 | 0.9286 | 0.9221 | 0.9156 |
| 千葉県 松戸市 | 0.9221 | 0.9351 | 0.9416 | 0.8701 | 0.9221 | 0.9351 | 0.9221 | 0.9416 |
| 平均 | 0.7649 | 0.8058 | 0.7731 | 0.7342 | 0.7064 | 0.8058 | 0.7616 | 0.7664 |

表 5.5 から，最も本人を特定できていると言えるのは，CASE2-2（移/Xgb/Bag）及び CASE2-6（文&移 by 連節）である．

次に，上記結果を元に，精度が良好であった CASE2-2，CASE2-6 と，比較対象として CASE2-1， CASE2-7，CASE2-8 の方式を全現住所に対して適用した結果を表 5.6 に示す．

表 5.6　全 46 現住所における各方式の結果

| 方式 | 投稿文 | 移/Xgb/Bag | 文&移by連節 | 文&移by連節/Scale | 文＆移by平均 |
|---|---|---|---|---|---|
| 当該方式の正解率が最良であった現住所の数 | 6/46 | 12/46 | 8/46 | 6/46 | 14/46 |
| 正解率の平均値 | 0.6454 | 0.7368 | 0.6477 | 0.6285 | 0.704 |

表 5.6 は，各方式を全現住所に適用した結果をまとめたものである．この表では，2 つの評価基準を設けている．1 つ目は，当該方式を用いた場合の正解率が，その他の方式を用いた場合と比べて最も高かった現住所の数を，方式毎に集計した数である．2 つ目は，各方式の正解率の全ての現住所における平均値である．その結果，移/Xgb/Bag，文&移 by 平均の方式が現住所推定に最も有効であることが分かった．以下では，これらの方式と，比較対象として投稿文方式を用いて実験を行った．

### 5.6.2　個人の再特定

4 章で述べた方式に，擬似移動履歴を用いた現住所推定方式を組み込んだ場合に，再特定精度がどの程度向上されるかを評価した．

具体的には，4.3 節で述べた通り，各アカウントについて被験者の履歴書モデルから統合スコアを算出し，統合スコアが最大となる履歴書を本人のものとして選択する．統合スコアは履歴書の各属性値についての確からしさを統合したものである．現住所以外の属性値のスコアは，4 章で述べた方式により算出する．現住所の属性値のスコアは，予備評価で有用であることが明らかになった移/Xgb/Bag，文&移 by 平均方式のスコアを用いた．

従来の方式と比較するため，以下の場合について再特定を行った．

1. 投稿文を特徴量として現住所モデルを作成する場合
2. 移/Xgb/Bag 方式によって現住所モデルを作成する場合
3. 文&移 by 平均方式によって現住所モデルを作成する場合

また，アカウントから本人の履歴書を特定する場合と，履歴書からアカウントを特定する場合，及びアカウントと履歴書を 1 対 1 で照合する場合を検討した．

表 5.7 に，移/Xgb/Bag 方式によって現住所モデルを作成した場合の，各履歴書モデルから算出された各アカウントの統合スコアを示す．

表 5.7　移/Xgb/Bag 方式で現住所モデルを用いた場合の各履歴書モデルの統合スコア

| アカウント番号 \ 履歴書番号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | … | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.8069 | 0.4762 | 0.4138 | 0.2912 | 0.4038 | 0.3145 | 0.0164 | 0.1748 | 0.2814 | 0.3029 | … | 0.1580 | −0.1551 | −0.1853 | −0.1496 | 0.3343 | 0.0228 | 0.4394 | 0.3372 | 0.2526 | −0.4010 |
| 2 | 0.2818 | 1.0161 | 0.3630 | −0.0026 | 0.6800 | 0.7451 | 0.5434 | 0.3592 | 0.6178 | 0.1275 | … | 0.0186 | −0.1489 | −0.1257 | −0.0647 | −0.2367 | 0.2640 | 0.2028 | 0.3100 | 0.5782 | −0.4522 |
| 3 | 0.2028 | −0.1102 | 1.2049 | −0.1411 | 0.3567 | 0.1870 | 0.1072 | 0.1773 | 0.0083 | 0.2753 | … | −0.0485 | 0.0154 | −0.1296 | −0.0991 | −0.4306 | 0.0281 | 0.1843 | −0.0708 | −0.2653 | −0.0900 |
| 4 | 0.1810 | 0.0584 | −0.0917 | 0.3825 | −0.1690 | −0.5800 | 0.0698 | −0.1604 | −0.3772 | −0.5942 | … | −0.1032 | −0.2094 | −0.0510 | −0.1333 | −0.1058 | −0.1327 | 0.3015 | −0.2522 | −0.3517 | −0.1092 |
| 5 | 0.0879 | −0.0643 | 0.3536 | 0.3717 | 0.7868 | 0.1909 | 0.1947 | −0.1224 | 0.4617 | 0.2002 | … | −0.1830 | −0.3456 | −0.1003 | −0.2426 | 0.3883 | 0.0738 | −0.2268 | 0.5412 | 0.3022 | −0.0956 |
| 6 | 0.1457 | −0.0970 | 0.3427 | 0.6739 | 0.1549 | 1.4828 | 0.6824 | 0.3865 | 0.2934 | 0.1871 | … | 0.0421 | −0.1048 | 0.0502 | −0.1094 | 0.1210 | −0.0116 | 0.0006 | 0.2150 | 0.0826 | −0.0028 |
| 7 | 0.4994 | 1.9056 | 1.2012 | 0.6282 | 1.5300 | 1.1020 | 2.0973 | 1.4372 | 0.7985 | 1.6716 | … | 1.2414 | 2.3648 | 1.2517 | 0.8788 | 0.9533 | 0.9595 | 0.7497 | 0.9746 | 1.1386 | 2.0165 |
| 8 | 0.3306 | 0.0576 | −0.0484 | −0.2237 | −0.0048 | 0.1780 | 0.0711 | 1.0724 | 0.1394 | 0.2742 | … | −0.2036 | −0.0910 | 0.2257 | 0.0706 | −0.3386 | 0.0140 | 0.0828 | 0.4860 | −0.0057 | −0.1624 |
| 9 | 0.2810 | 0.3088 | 0.1346 | −0.0822 | 1.0215 | 0.1012 | 0.0847 | 0.3217 | 2.1534 | −0.0146 | … | 0.3388 | 0.0009 | 0.7060 | 0.2037 | 0.0391 | 0.1850 | 0.4869 | 0.9049 | 0.3232 | 0.3815 |
| 10 | 0.0766 | 0.1658 | 0.6113 | −0.0961 | 0.3819 | 1.0215 | 0.7522 | 0.2408 | 0.4355 | 0.8213 | … | −0.3878 | −0.3064 | 0.1169 | 0.2720 | −0.0261 | 0.0958 | 0.7018 | 0.0395 | −0.3631 | 0.2663 |
| … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … | … |
| 70 | 0.0516 | −0.5164 | −0.0675 | −0.4807 | −0.5970 | −0.7563 | −0.6200 | −0.5667 | −0.1952 | −0.4403 | … | −0.1753 | 0.6546 | 0.3951 | 0.6120 | −0.1018 | 0.3473 | −0.5510 | 0.1167 | −0.3234 | 0.1562 |
| 71 | −0.1749 | −0.3737 | −0.4623 | −0.5861 | −0.1240 | −0.5835 | 0.1147 | −0.4316 | −0.5182 | −0.3713 | … | −0.1518 | −0.4077 | 0.7190 | 0.0069 | 0.1866 | −0.6577 | −0.2758 | −0.5048 | −0.2347 | −0.0587 |
| 72 | −0.0172 | −0.3460 | −0.4382 | 0.0645 | −0.4472 | −0.2670 | −0.7087 | −0.4511 | −0.5720 | −0.0910 | … | 0.0008 | 0.0321 | −0.1198 | 0.5321 | 0.0014 | 0.4715 | −0.1255 | −0.0305 | 0.0497 | 0.5279 |
| 73 | 0.1152 | −0.2490 | −0.0045 | 0.3409 | −0.0023 | 0.5053 | 0.4381 | −0.0320 | −0.1717 | 0.3467 | … | 0.5986 | −0.1208 | 0.3594 | 0.2269 | 1.2941 | 0.1077 | 0.2625 | 0.2702 | 0.1679 | −0.1745 |
| 74 | −0.2761 | −0.5243 | −0.5688 | −0.3582 | −0.0619 | −0.6250 | −0.6013 | −0.5590 | 0.0335 | −0.3346 | … | −0.1059 | 0.1292 | −0.2861 | 0.2140 | −0.0637 | 0.4806 | −0.3091 | −0.1097 | −0.2460 | 0.2345 |
| 75 | 0.8153 | 1.3620 | 0.4823 | 0.8477 | 0.7469 | 0.2065 | −0.1270 | 0.5074 | −0.0252 | 1.0623 | … | 1.1090 | 2.1698 | 0.8241 | 0.7838 | 0.4470 | 0.8289 | 1.4999 | 0.6096 | 1.0984 | 1.5864 |
| 76 | 0.1431 | 0.1827 | −0.0767 | −0.0595 | −0.2948 | 0.0835 | −0.4262 | −0.1678 | −0.2209 | −0.0054 | … | −0.1212 | 0.1099 | 0.0015 | 0.3741 | −0.2869 | 0.0294 | −0.0093 | 0.5198 | 0.5139 | −0.3859 |
| 77 | −0.1755 | 0.0441 | −0.0942 | 0.2256 | −0.2133 | −0.3666 | −0.0584 | −0.0893 | −0.2499 | −0.5100 | … | −0.1174 | −0.4979 | −0.5775 | −0.7503 | −0.5326 | −0.6668 | −0.4423 | −0.2869 | 0.2342 | −0.6774 |
| 78 | 0.0340 | −0.3597 | −0.1588 | 0.0129 | −0.2813 | −0.6283 | −0.4595 | −0.3422 | −0.2994 | −0.3534 | … | −0.0449 | 0.2175 | −0.1641 | 0.2875 | 0.1444 | 0.3691 | −0.1040 | −0.1789 | −0.5753 | 0.3094 |

アカウントから本人の履歴書を特定する場合（基本問題）の表の見方を説明する．例えば，匿名のソーシャルメディアアカウント 1 に対して本人の履歴書を特定する場合，表 5.7 の 1 行目の各列の値が，アカウント 1 に対する各履歴書の統合スコアとなっている．よって，1 行目の中で最も統合スコアの高い列に対応する履歴書が，アカウント 1 の人物の履歴書であると推定できる．

履歴書からアカウントを特定する場合（逆問題）の見方を説明する．例えば，本人のソーシャルメディアアカウントが未知である履歴書 1 から本人のアカウントを特定する場合，表 5.7 の 1 列目の各行の値が，同一の履歴書モデルから算出された，各アカウントの統合スコアとなっている．よって，1 列目の中で最もスコアの高い行に対応するアカウントが，履歴書 1 のアカウントであると推定できる．

アカウントと履歴書を 1 対 1 で照合する方法（1 対 1 問題）を説明する．1 対 1 の照合にはハンガリアンアルゴリズムを適用する．ハンガリアンアルゴリズムは，アカウントと履歴書の重複の無い全ての組み合わせに対して，スコアの合計値が最大となる組合せを出力する．表 5.7 の場合，対角線上の履歴書とアカウントのペア（(履歴書 1,アカウント 1），(履歴書 2,アカウント 2）,...,（履歴書 72,アカウント 72)）のスコアの合計値がその他の組合せの中で最大であれば，その履歴書とアカウントのペアを正しく推定出来ていると言える．

(1) 基本問題

アカウントから本人の履歴書を特定する場合の照合結果を以下の表 5.8 に示す．

表 5.8　被験者アカウントから本人の履歴書を推定した結果
※( )内は，特徴ベクトルに投稿文を使用した場合の特定人数との差分

| 特徴ベクトル | 1 位で特定できた人数 | 10 位以内で特定できた人数 |
|---|---|---|
| 投稿文 | 38 | 67 |
| 移/Xgb/Bag | 42(+4) | 71(+4) |
| 文&移 by 平均 | 43(+5) | 67(±0) |

現住所推定に投稿文のみを使用した場合，1 位として特定できたのは 38 人，10 位以内で特定できたのは 67 人であった．

一方で，現住所推定に移/Xgb/Bag 方式を適用した場合，1 位として特定できたのは 42 人，10 位以内で特定できたのは 71 人となり，投稿文を用いた場合に比べて照合できた人物が増えた．現住所推定に文&移 by 平均方式を用いた場合，1 位として特定できたのは 43 人，10 位以内で特定できたのは 67 人となり，投稿文を用いた場合に比べて 1 位で特定できた人数は 5 人増加したが，10 位以内で特定できた人数は変わらなかった．

(2) 逆問題

表 5.9　本人の履歴書から被験者アカウントを推定した結果
※( )内は，特徴ベクトルに投稿文を使用した場合の特定人数との差分

| 特徴ベクトル | 1 位以内で特定できた人数 | 10 位以内で特定できた人数 |
|---|---|---|
| 投稿文 | 43 | 67 |
| 移/Xgb/Bag | 46(+3) | 70(+3) |
| 文&移 by 平均 | 46(+3) | 70(+3) |

現住所推定に投稿文を使用した場合，1 位として特定できたのは 43 人，10 位以内で特定できたのは 67 人となった．

一方で，現住所推定に移/Xgb/Bag 方式及び文&移 by 平均方式を適用した場合，1 位として特定できたのは 46 人，10 位以内で特定できたのは 70 人となり，疑似移動履歴を用いた特定方式の方が，投稿文を用いた場合に比べて照合できた人物が増えた．

(3) 1 対 1 問題

表 5.10　被験者アカウントと本人の履歴書を 1 対 1 で照合した結果
※( )内は，特徴ベクトルに投稿文を使用した場合の特定人数との差分

| 特徴ベクトル | 1 位以内で特定できた人数 |
|---|---|
| 投稿文 | 39 |
| 移/Xgb/Bag | 50(+11) |
| 文&移 by 平均 | 43(+4) |

現住所推定に投稿文を使用した場合，1 対 1 で照合できたのは 39 人であった．

一方で，現住所推定に移/Xgb/Bag 方式を適用した場合，1 位として特定できたのは 50 人，文&移 by 平均方式を適用した場合 43 人となり，疑似移動履歴を用いた特定方式の方が，投稿文を用いた場合に比べて照合できた人物が増加した．これらの結果から，4 章で述べた方式に，擬似移動履歴を用いた現住所推定方式を組み込むことによって，本人推定精度が向上されることが明らかになった．

## 5.7　考察

5.5 節の予備評価で示した通り，現住所推定には，疑似移動履歴を特徴量とする場合と，投稿文から得られる単語の出現頻度と疑似移動履歴を連節した特徴ベクトル（以下では投稿文＋疑似移動履歴と示す.）を特徴量とする場合が有効であることが分かった．この理由について考察する．

表 5.8～5.10 から，投稿文，疑似移動履歴，投稿文＋疑似移動履歴の各特徴量に含まれる情報量を比較すると，投稿文＋疑似移動履歴の情報量は原理的には疑似移動履歴のみの情報量よりも多いと予想される．具体的に，どの部分で情報量に差が表れるかを考察する．例として，藤沢市在住のアカウントを取り上げる．

はじめに，投稿文から得られる情報に着目する．藤沢市在住のアカウントが「鵠沼に遊びに行った.」「とびっちょ（藤沢の有名店）でしらす丼食べた.」と投稿したとする．この場合，投稿文を形態素解析して得られる「鵠沼/に/遊び/に/行っ/た/」「/とびっちょ/で/しらす/丼/食べ/た/」の各形態素に対する出現頻度が特徴量として用いられる．鵠沼はご

当地の地名なので，藤沢在住の人でなければほとんど発言しないため，現住所の特定に有益な情報となる．また，地元の人に人気がある店やイベントは藤沢在住の人の投稿文に多く表れると考えられる．これらは地名ではないので，疑似移動履歴には表れないが，投稿文の特徴ベクトルには表れる（図 5.6）．

次に，疑似移動履歴から得られる情報に着目する．疑似移動履歴に含まれる情報は，地域を 1km 毎に区分けした際の各メッシュへの訪問回数である．疑似移動履歴のメッシュのカウントロジックでは，該当メッシュに 1，上下左右のメッシュに$\frac{1}{2}$，斜め四方のメッシュに$\frac{1}{2\sqrt{2}}$カウントを加える．そのため，「鵠沼」と「江の島」のように，1 ㎞より離れているが 2√2 ㎞以内の距離にある 2 つの地名が投稿文に現れた時，移動履歴に変換後の特徴ベクトルでは，各々の該当メッシュに 1 がカウントされるだけでなく，共通の周辺メッシュにカウントが重畳されるので，共通の特徴がベクトルに現れる．上記は疑似移動履歴の持つ緯度経度の情報によって可能となるので，投稿文の情報には現れない．

以上のように，投稿文から得られる単語の出現頻度と，疑似移動履歴から得られるメッシュの訪問回数は異なる情報を表しており，補完関係にある（図 5.6）．特徴量と疑似移動履歴を連節した特徴ベクトルは，上記で記した 2 つの特徴ベクトルの情報を合わせ持っており，最も情報が多い．

現住所のプロファイリング精度（表 5.6）および個人特定の精度（表 5.8，5.9，5.10）から，今回のデータセットについては，投稿文の固有の情報（ご当地の店やイベント）よりも移動履歴に固有の情報（地理的に近い場所の類似性）の方が，現住所および個人の推定に有効であったと考えられる．一方，原理的には，投稿文＋疑似移動履歴の情報は最も多いが，現住所のプロファイリングおよび個人特定の精度は，疑似移動履歴のみの精度と同等あるいは低かった．今回は，投稿文と疑似移動履歴の情報を統合する方法として，2 つの確からしさの平均および 2 つの特徴ベクトルの接続を用いたが，今後の課題として，より高度な統合手法を検討する必要がある．

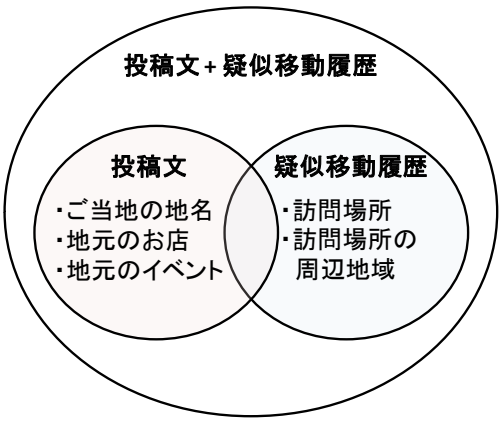


図 5.6　投稿文と疑似移動履歴の情報量

## 5.8　まとめ

4 章では，ソーシャルメディアの投稿文を直接用いて現住所を特定していたが，5 章では投稿文を直接用いるのではなく，投稿文から変換した疑似移動履歴を用いて，匿名アカウントの現住所推定方式を提案した．

78 人の被験者による現住所のプロファイリング評価では，疑似移動履歴を用いた正解率が，投稿文を用いたプロファイリングに比べて 9%上昇し，疑似移動履歴の利用が現住所推定に有用であることが明らかにした．

また， 4 章で述べた個人の再特定方式に擬似移動履歴を用いた現住所のプロファイリングを組み込むことにより，投稿文のみを用いる場合と比べて個人特定率が向上することが分かった．具体的には，アカウントから本人の履歴書を特定する場合，疑似移動履歴を用いることで，本人を 1 位で特定できた人数が 78 人中 38 人から 43 人に向上した．履歴書から本人のアカウントを特定する場合，本人を 1 位で特定できた人数が 78 人中 43 人から 46 人に向上した．また，履歴書とアカウントを 1 対 1 で照合する場合，本人を 1 位で特定できた人数が 78 人中 39 人から 50 人に向上した．これからの評価により，疑似移動履歴を利用する提案法が匿名の人物を特定するのに有益であることを明らかにした．

今後の課題としては，投稿文と疑似移動履歴の情報の統合方法を高度化することで，さらに精度を向上することがあげられる．

# 第6章　提案方式の拡張性と限界

## 6.1　まえがき

3 章では，匿名のソーシャルメディアアカウントについてプロファイリングによって個人を特定する提案手法の枠組み，4 章ではソーシャルメディアの情報のうち投稿文を利用した場合の提案手法の実現方法，5 章では疑似移動履歴を用いた場合の提案手法の実現方法について述べた．

本章では，これまでに述べた提案手法の発展形として，投稿画像を用いる場合，及びリンク情報を用いる場合の提案手法の実現可能性について検討し，その有効性を評価する．

また，ソーシャルメディアの多種多様なデータの活用による再特定精度の向上の可能性，及び提案手法の限界および従来手法との補完関係について考察する．

## 6.2　画像を用いたプロファイリングによる再特定

4 章および 5 章では，ソーシャルメディアアカウントの人物をプロファイリングする際に，アカウントが保有するデータのうちの投稿文，及び投稿文から抽出した疑似移動履歴（すなわち投稿文中の地名，投稿時刻，地名と座標の関係）を利用できることを示した．また，両者の情報を組み合わせることで，個人特定の精度を向上することができた．アカウントの保有するデータには，投稿文の他にも，投稿画像やリンク情報があるので，それらの情報を用いてソーシャルメディアカウントの人物をプロファイリングできると考えられる．

本節では，ソーシャルメディアアカウントの投稿画像を用いてプロファイリングを行い，個人を再特定する手法について検討する．

近年，instagram [110]のような，投稿文よりも画像の投稿が中心であるソーシャルメディアもあるため，投稿画像を用いてプロファイリングできることが望ましい．また，投稿文を十分に取得できる場合でも，投稿画像を併用してプロファイリングを行うことで，再特定の精度を向上できる可能性がある．

### 6.2.1　投稿画像を用いたプロファイリング手法

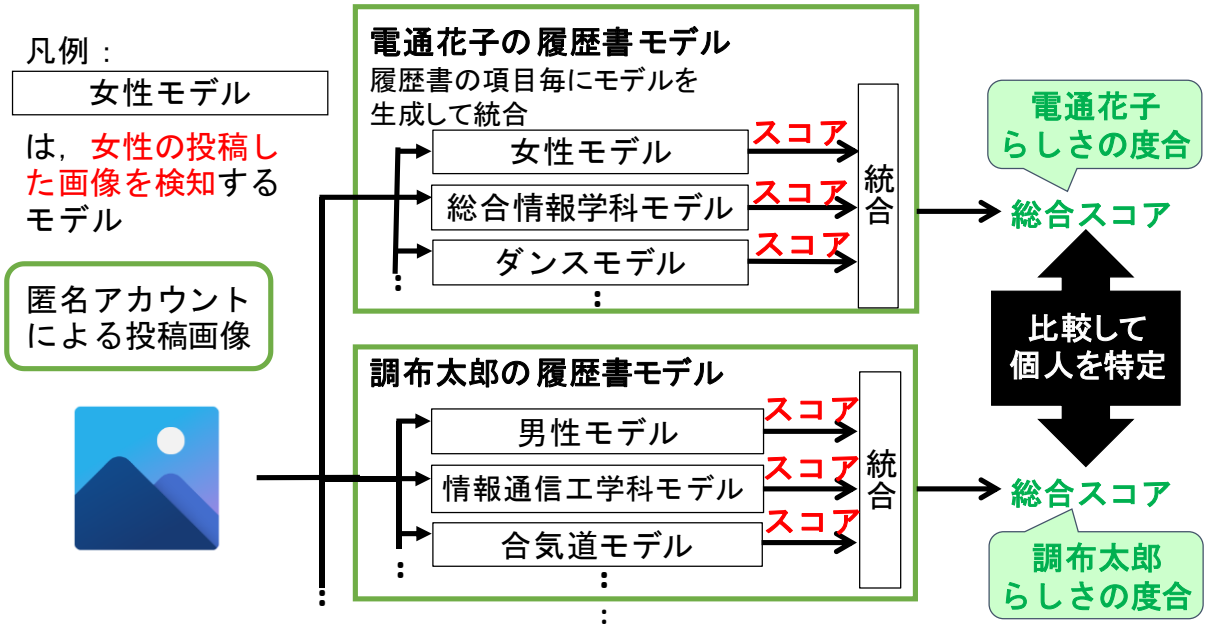


図 6.1　投稿画像を用いたプロファイリング

提案手法におけるプロファイルとして履歴書を，匿名アカウントのデータとして投稿画像を用いた場合のプロファイリング手法について，図 6.1 を用いて説明する．3 章で述べた手法と同様に，特定対象者の履歴書毎に個人を識別するモデル（以下，履歴書モデル）を学習する．履歴書モデルからは，匿名アカウントの人物が当該履歴書の人物らしいかを表す統合スコアが出力される．統合スコアは，属性値毎のモデルから算出されたスコアを統合処理した値を用いる．統合スコアを比較することによって，複数ある候補者の履歴書のうち，匿名アカウントがどの履歴書の人物であるかを推定することができる．

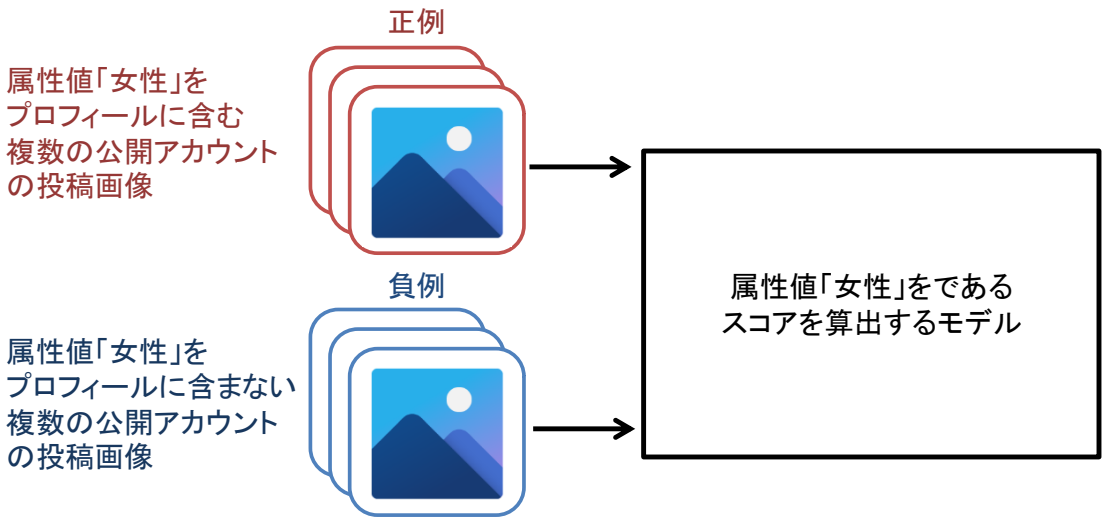


図 6.2　投稿画像を用いた属性値モデルの学習方法

投稿画像を用いてプロファイリングを行う場合，履歴書モデルを構成する属性値モデルは，その属性値を持つアカウントの投稿画像を正例として学習する（図 6.2）．例えば，属性値「女性」を持つかどうかを識別するモデルを学習する場合，正例としてプロフィールに女性であることが記載されているアカウントの投稿画像を，負例としてそれ以外のアカウントの投稿画像を用いる．

### 6.2.2　画像を用いたプロファイリングの実現例

画像を用いて投稿者をプロファイリングする上記の手法を丸山が実現しているので[111]，この実現方法を採用する．本節では，丸山の実現方法を説明する．

丸山の手法では，Google Cloud が提供する Vision API [112]を使用して，画像を description と呼ばれる単語集合に変換する（図 6.3）．Vision API は，事前に学習済みの機械学習モデルによって画像を認識するサービスである．Vision API に画像を入力すると，画像に含まれるオブジェクトが自動的に検出され，オブジェクトの属性が単語として出力される．丸山は，その中のラベル検出を使用した．

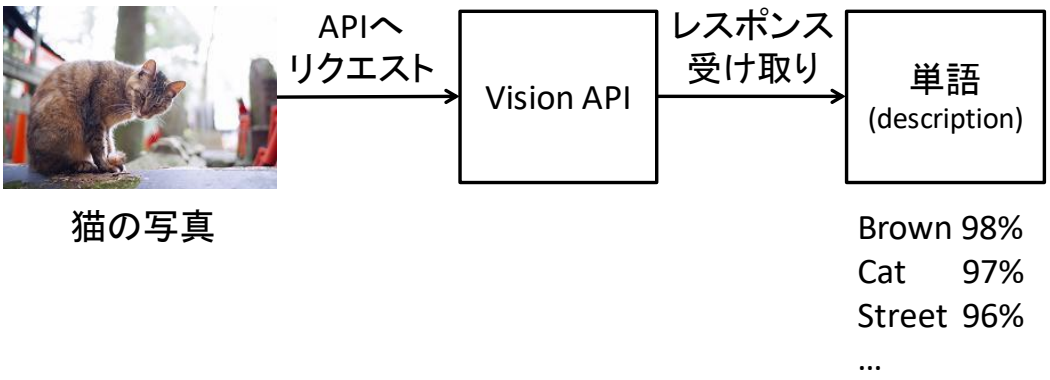


図 6.3 Vision API の使用

アカウントが保持する画像の各々に対して Vision API のラベル検出機能を適用して単語集合を抽出する．画像毎の単語集合の和集合を計算し，単語毎の出現頻度を特徴量として，モデルの学習および未知アカウントのプロファイリングを行う．

このように画像の集合を単語毎の出現頻度に変換することで，それ以降の処理については，4 章で述べた投稿文からのプロファイリングと共通化することができる（図 6.4）．

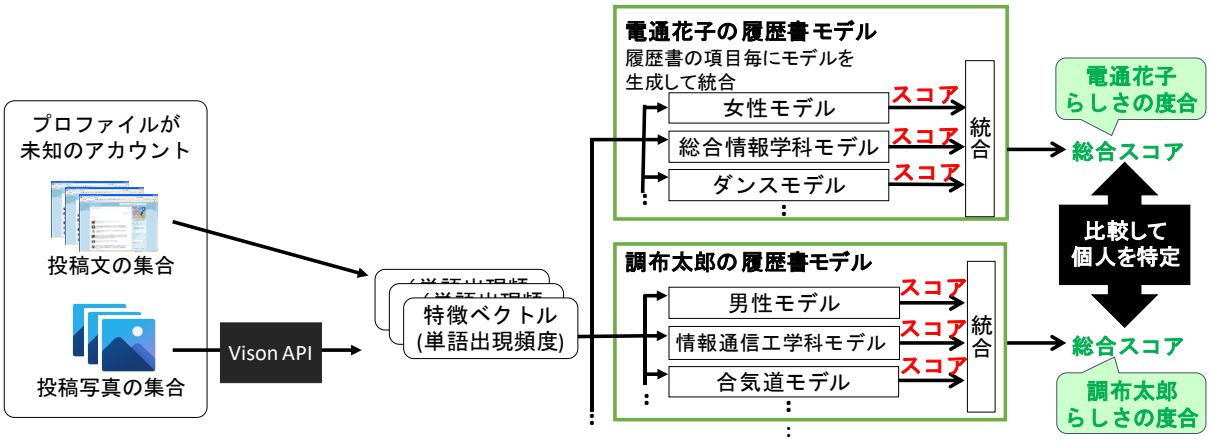


図 6.4　共通の枠組みによる投稿文と投稿画像を用いたプロファイリング

### 6.2.3　データセット

丸山は 4.7.1 節で述べた被験者 78 名のアカウントのうち投稿画像が入手可能な 51 名のアカウントを用いた．被験者の内訳は，電気通信大学に在籍する学生 21 人と，電気通信大学以外に在籍する学生や社会人等の外部の人間 30 人である． 51 名の履歴書と Twitter アカウントから入手した投稿画像をテストデータとした．被験者アカウントの平均画像投稿枚数は 261 枚，標準偏差は 221 枚である．

丸山は，プロファイリングする属性値として，表 6.1 に示す 15 種類の趣味に着目した．例えば趣味が「ギター」に該当する被験者は 3 人であり，学習データの正例は 82 件，負例も同数の 82 件のアカウントを 4 章の手法と同様に公開アカウントから収集した．「ギター」のモデル学習に使用したアカウントの平均投稿画像枚数は 359.10 枚であった．

これらの 15 種類の趣味について，51 人の被験者がその趣味を有するかを投稿画像からプロファイリングし，正解率を評価した．

表 6.1　データセットの内訳 [111]

| 趣味名 | 該当被験者数　学習データ件数 | 学習データ平均枚数 |
|---|---|---|
| ギター | 3 | 82 | 359.10 |
| バスケ | 2 | 31 | 220.48 |
| ピアノ | 1 | 74 | 250.96 |
| ベース | 2 | 81 | 202.07 |
| ラーメン | 1 | 65 | 330.77 |
| ラグビー | 1 | 62 | 152.21 |
| ロボット | 1 | 86 | 235.48 |
| 音楽鑑賞 | 4 | 74 | 265.61 |
| 将棋 | 2 | 80 | 249.24 |
| 食べ歩き | 3 | 81 | 376.54 |
| 登山 | 1 | 72 | 287.85 |
| 読書 | 4 | 69 | 271.59 |
| 野球 | 3 | 90 | 316.39 |
| 料理 | 5 | 49 | 234.33 |
| 手芸 | 2 | 177 | 437.29 |
| 平均値 | 2.33 | 78.2 | 279.33 |
| 標準偏差 | 1.25 | 30.2 | 71.30 |

### 6.2.4　画像を用いたプロファイリングの精度

6.2.3 節で示した趣味について，投稿画像を用いたプロファイリングを行い，正解率を求めた．その際，モデル作成時のパラメータチューニング方法として，チューニングなし，グリッドサーチ，ランダムサーチの 3 種類を検討した．また，データの正規化方法として，無編集（学習データの特徴量として bag-of-words をそのまま用いる），画像枚数（各被験者が保持している画像の枚数でベクトルを除算し正規化する），ノルム 1（特徴ベクトルが単位ベクトルとなるように正規化する），0-1（特徴ベクトルの最小値が 0，最大値が 1 となるように正規化する）の 4 種類を検討した．これらの組合せである 12 種類の評価を行った．その結果，モデル作成のパラメータはチューニングなし，データ正規化は画像枚数による正規化とした場合の正解率が最も高く，62.32%であった．

また，投稿画像から求めた各属性値モデルの確からしさと，投稿文から求めた各属性値モデルの確からしさの平均値を新たな確からしさとしてみなしてプロファイリングを行い，正解率を求めた．その結果，モデル作成はチューニングなし，正規化なしが最良で，その時の正解率は 67.08%であった．

一方，投稿文のみを用いて当該 15 種類の趣味をプロファイリングした場合の正解率の平均は 62.52%であった．そのため，投稿文と画像を組み合わせる方が，投稿文のみ使用する場合と比較して正解率が約 5%向上した．よって，プロファイリングにおいて，投稿文に加えて画像を用いることが有効であることがわかった．

### 6.2.5　再特定の精度評価

4 章では，投稿文のみを用いたプロファイリングに基づいて，アカウントと履歴書を照合し，アカウントから個人を特定した．本節では，上述した画像と投稿文を併用してプロファイリングを行う場合の，画像を入手できた 51 名の被験者に対する個人特定の精度を評価する．

上述した 15 種類の趣味に着目し，6.2.4 節で示したパラメータと正規化の最良の組合せ 2 ケース CASE1=「画像のみ パラメータチューニング=無し，正規化=画像枚数」，CASE2=「投稿文と画像の平均値 パラメータチューニング=無し，正規化=0-1」を用いて個人特定を行った．4.7.2 節と表 4.4 で示したように，投稿文を用いた個人特定では，アカウント毎に 420 種類の属性値を有する確からしさをプロファイリングによって求め，これらの被験者数×属性値数の確からしさの組合せに基づいて，アカウントと履歴書を照合し個人を特定していた．本節では，CASE1 と CASE2 のそれぞれについて，上述した 15 種類の趣味の属性値モデルの内，投稿文のみ用いた場合よりも正解率が向上した属性について，当該属性値を有する確からしさを CASE1 もしくは CASE2 で作成した属性値の値に置き換えた．その他の属性値に関しては，投稿文のみで作成した属性値モデルを用いて，4.6.3 節と同様に個人特定を行った．

(1) 基本問題（アカウントから本人の履歴書を特定する場合）

表 6.2　基本問題の再特定精度

| CASE. | 特徴量の作成方法 | 特徴量を置き換えた属性値 | 個人特定数<br>(個人特定率) |
|---|---|---|---|
| 0 | 従来手法(投稿文のみ) | - | 27/51(52.9%) |
| 1 | 画像のみ<br>・パラメータチューニング=無し<br>・正規化=画像枚数 | 料理, ラーメン, ギター, バスケ,<br>ロボット, 音楽鑑賞 | 28/51(54.9%) |
| 2 | 投稿文と画像の平均値<br>・パラメータチューニング=無し<br>・正規化=0-1 | 料理, ラグビー, ラーメン, ギター,<br>音楽鑑賞, バスケ, 読書 | 31/51(60.8%) |

表 6.2 の CASE.0 は，全特徴量を投稿文で作成した場合の個人特定数である．この場合，個人特定数は 27 人（52.9%）であった．CASE.1 は，画像のみを用いて作成した属性値モデルの内，投稿文を用いた手法よりも正解率が向上した料理，ラーメン，ギター，バスケ，ロボット，音楽鑑賞の属性値モデルの結果を，投稿文を用いた確からしさと置き換えた場合である．この場合，個人を特定できた人数は 28 人（54.9%）であった．

CASE.2 は，投稿文と画像でそれぞれ作成した属性値モデルの結果の平均値を確からしさとした際に，投稿文を用いた手法よりも正解率が向上したギター，音楽鑑賞，バスケ，読書の結果を，投稿文を用いた確からしさと置き換えた場合である．この場合，個人特定数が 31 人（60.8%）となり，従来手法の特定人数を上回った．

(2) 逆問題（履歴書から本人のアカウントを特定する場合）

表 6.3　逆問題の再特定精度

| CASE. | 特徴量の作成方法 | 個人特定数<br>(個人特定率) |
|---|---|---|
| 0 | 従来手法(投稿文のみ) | 31/51(60.8%) |
| 1 | 画像のみ<br>・パラメータチューニング=無し<br>・正規化=画像枚数 | 33/51(64.7%) |
| 2 | 投稿文と画像の平均値<br>・パラメータチューニング=無し<br>・正規化=0-1 | 31/51(60.8%) |

特徴量に画像を用いて，履歴書から本人のアカウントを照合した結果を表 6.3 に示す．CASE.0～CASE.2 は，(1) 基本問題の結果で説明したものと同様である．CASE.0 及び CASE.2 の個人特定数は 31 人（60.8%），CASE.1 の個人特定数は 33 人（64.7%）であった．

(3) 1 対 1 問題（アカウントと履歴書を 1 対 1 で照合する場合）

表 6.4　1 対 1 問題の再特定精度

| CASE. | 特徴量の作成方法 | 個人特定数<br>(個人特定率) |
|---|---|---|
| 0 | 従来手法(投稿文のみ) | 33/51(64.7%) |
| 1 | 画像のみ<br>・パラメータチューニング=無し<br>・正規化=画像枚数 | 35/51(68.6%) |
| 2 | 投稿文と画像の平均値<br>・パラメータチューニング=無し<br>・正規化=0-1 | 35/51(68.6%) |

特徴量に画像を用いて，アカウントと履歴書を 1 対 1 で照合した結果を表 6.4 に示す．

CASE.0～CASE.2 は，(1) 基本問題の結果で説明したものと同様である．CASE.0 の個人特定数は 33 人（64.7%），CASE.1 及び CASE.2 の個人特定数は 35 人（68.6%）であった．

基本問題，逆問題，1 対 1 問題の結果から，投稿文と画像を併用することにより，投稿文のみを用いる場合よりも個人特定の精度が向上する可能性を示した．

表 6.5　投稿文と疑似移動履歴と画像を組み合わせた個人特定の精度

| 被験者数 | | 特徴量の作成方法 | | 個人特定数(個人特定率) | | |
|---|---|---|---|---|---|---|
| | CASE. | 現住所 | 趣味 | 基本問題 | 逆問題 | 1対1問題 |
| 51人 | 0 | 投稿文 | 投稿文 | 29(56.8%) | 34(66.6%) | 31(60.7%) |
| | 1 | 文＆移by平均 | 投稿文 | 30(58.8%) | 34(66.6%) | 35(68.6%) |
| | 2 | 移/Xgb/Bag | 投稿文 | 29(56.8%) | 33(64.7%) | 37(72.5%) |
| | 3 | 投稿文 | 画像のみ | 29(56.8%) | 33(64.7%) | 35(68.6%) |
| | 4 | 投稿文 | 投稿文と画像の平均値 | 30(58.8%) | 34(66.6%) | 34(66.6%) |
| | 5 | 文＆移by平均 | 投稿文と画像の平均値 | 30(58.8%) | 35(68.6%) | 35(68.6%) |
| | 6 | 移/Xgb/Bag | 投稿文と画像の平均値 | 30(58.8%) | 34(66.6%) | 34(66.6%) |

現住所および趣味の特徴量として，4 章で検討した投稿文，5 章で検討した疑似移動履歴，本章で検討した画像を用いて，それぞれの特徴量の組合せで再特定を行った結果を表 6.5 に示す．

CASE.0 は，投稿文のみを用いて再特定を行った結果である．

CASE.1～CASE.4 は，現住所もしくは趣味のどちらかに対して投稿文を特徴量として用いて再特定を行った結果である．これらの CASE は，CASE.0 と比較して，基本問題，逆問題，1 対 1 問題のいずれかの問題で個人特定数が増加することはあっても，全ての問題で増加することはなかった．例えば CASE.2 では，1 対 1 問題においては個人特定数が 37 人となり特定人数が増加したが，逆問題においては 33 人と減少している．

CASE.5，CASE.6 は，現住所に疑似移動履歴を，趣味に投稿文と画像の平均値を特徴量として用いて再特定を行った結果である．CASE.5 における個人特定数は，基本問題で 30 人，逆問題で 35 人，1 対 1 問題で 35 人となり，全 CASE の中で個人特定数が最良となった．CASE.6 の個人特定数が CASE.5 よりも劣っているが，CASE.6 は現住所の特徴量に疑似移動履歴のみ利用しており，CASE.6 は疑似移動履歴と投稿文の両方を用いているため，CASE.5 よりも情報量が多くなったためと考えられる．

このことから，特徴量として投稿文のみ用いるよりも，疑似移動履歴や画像など，複数の情報を組み合わせた特徴量を用いる方が，個人特定精度が向上する可能性を示した．

今回検証していない特徴量の組合せを用いた個人特定精度の検証，およびソーシャル

メディアから得られるその他情報の特徴量としての活用は，今後の課題とする．

## 6.3　リンクを用いたプロファイリングによる再特定

ソーシャルメディアのデータは，これまでに言及した投稿文や投稿画像以外にも，友人との繋がりを示すリンク情報（図 6.5）を含んでいる．ソーシャルメディアのデータは，分析用に匿名化して提供される場合があるが，投稿文や投稿画像の内容はプライバシー情報を含む可能性が高いとして，リンク情報のみを提供する場合がある．例えば，スタンフォード大学では，ユーザ情報が匿名化されている Facebook のリンク情報を公開している [13]．このように，一般に公開されている匿名化されたリンク情報からプロファイリングができれば，プライバシーのリスクを示す上で有益である．また，匿名アカウントの情報として投稿文や投稿画像を収集できる場合でも，リンク情報も併用してプロファイリングすることで，再特定の精度をさらに向上できる可能性がある．本節では，リンク情報を用いたプロファイリング及び再特定手法について述べる．

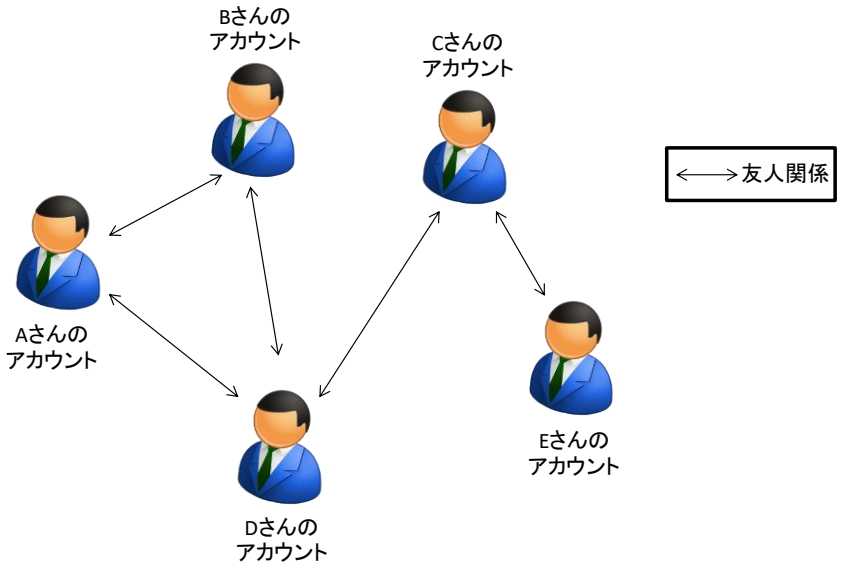


図 6.5　リンク情報のイメージ

### 6.3.1　リンク情報を用いたプロファイリング手法

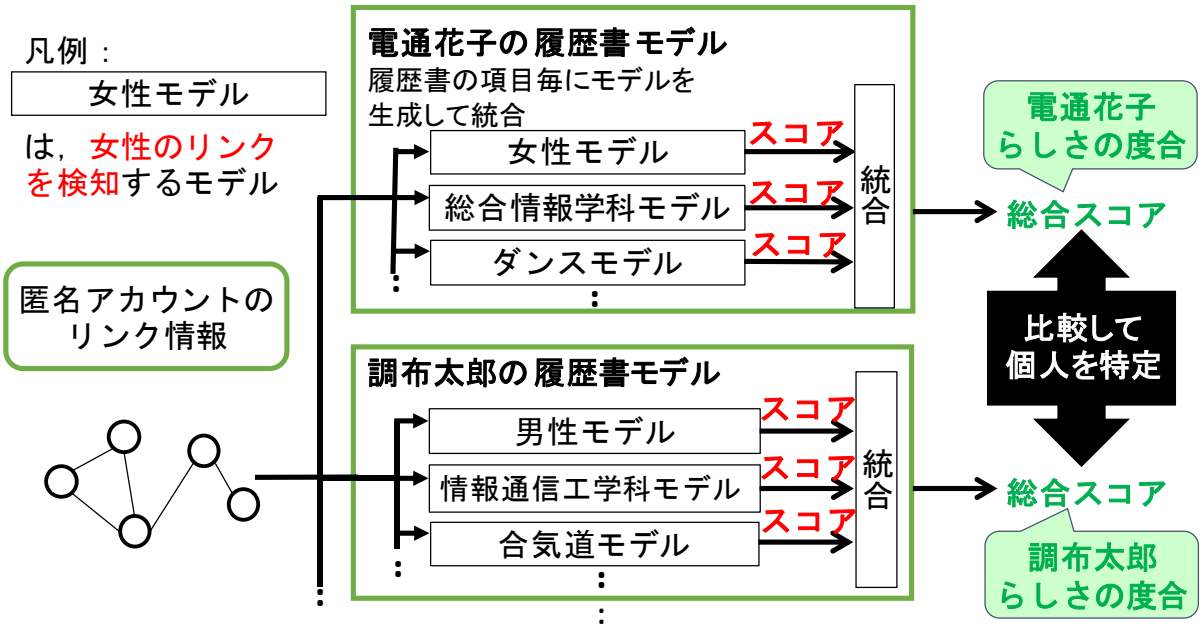


図 6.6　リンク情報を用いたプロファイリング

提案手法におけるプロファイルとして履歴書を，匿名アカウントのデータとしてリンク情報を用いた場合のプロファイリング手法について，図 6.6 を用いて説明する．リンクを用いたプロファイリングは，当該アカウントの人物の交友関係に基づいたプロファイリングとみなすことができる．4 章で述べた手法と同様に，特定対象者の履歴書毎に個人を識別するモデル（以下，履歴書モデル）を学習する．履歴書モデルからは，匿名アカウントの人物が当該履歴書の人物らしいかを表す統合スコアが出力される．統合スコアを比較することによって，複数ある候補者の履歴書のうち，匿名アカウントがどの履歴書の人物であるかを推定することができる．統合スコアは，各属性のモデルから算出されたスコアを統合処理した値を用いる．

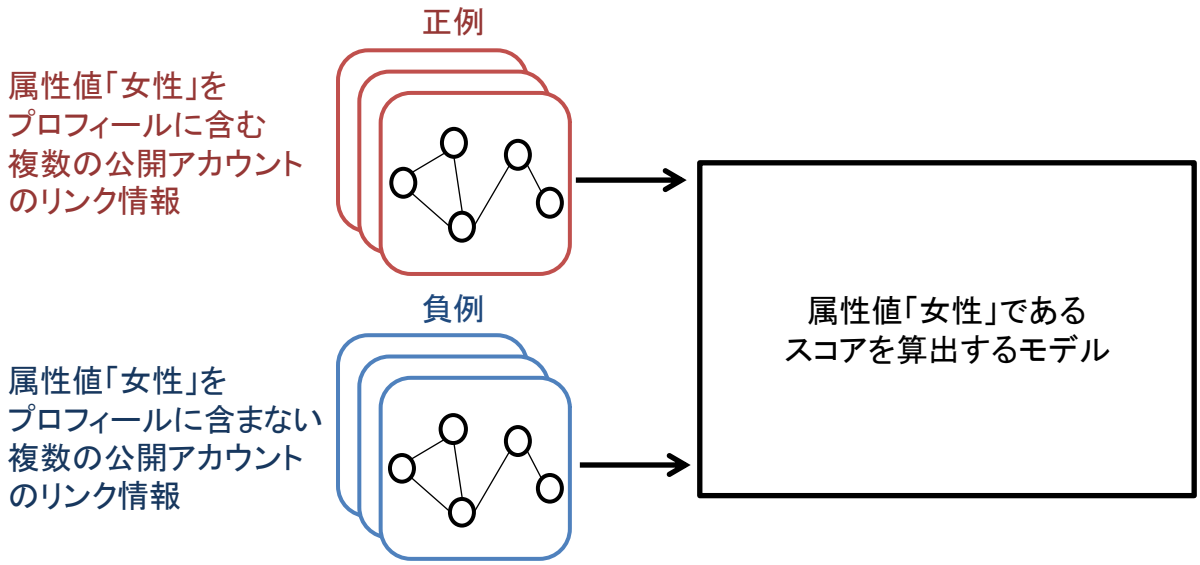


図 6.7　リンク情報を用いた属性値モデルの学習方法

履歴書モデルを構成する属性値モデルは，その属性値を持つアカウントのリンク情報を正例として用いて作成する（図 6.7）．例えば，属性値「女性」を持つかどうかを識別するモデルを作成する場合，正例としてプロフィールに女性であることが記載されているアカウントのリンク情報を，負例としてそれ以外のアカウントのリンク情報を用いる．これは，女性の交友関係と男性の交友関係をサンプルとして，女性の交友関係の特徴を学習することを意味している．

### 6.3.2　リンク情報を用いた場合のプロファイリングの実装

リンク情報を機械学習のモデルへの入力形式に変換するために，Node2vec [113]と呼ばれる Python ライブラリを利用した．Node2vec は，グラフ構造からベクトル空間上の特徴表現に変換する．例えば，Facebook において友達同士のユーザを線で結んでいくと，最終的には非常に大きな友人関係の繋がりを示すソーシャルグラフが出来る．Node2vec は，このようなソーシャルグラフについて，各アカウント（ソーシャルグラフのノード）の友人関係（エッジ）から各ノード周辺のグラフ構造を学習し，低次元ベクトルとして表現する（図 6.8）．

Node2vec を用いてリンク情報からベクトルを生成する手法について説明する．まず始めにグラフ構造からサンプリングを行う．従来のサンプリング方法は，木構造における Breadth-first-search（幅優先探索）のように，ルートノード周辺のノードを近傍とみなす Breadth-first-Sampling や，Depth-first-search（深さ優先探索）のように，ルートノードから末端の方向に探索を続け，その際に通ったノードを近傍とみなす Depth-first Sampling のような方法が取られていた． Node2vec では，この 2 つの手法をグラフ構造

の特性によって使い分けることで，元のデータ構造に適したサンプリングを行う．次に，skip-gram と呼ばれる手法を用いて，サンプリングされたデータからベクトルを生成する．

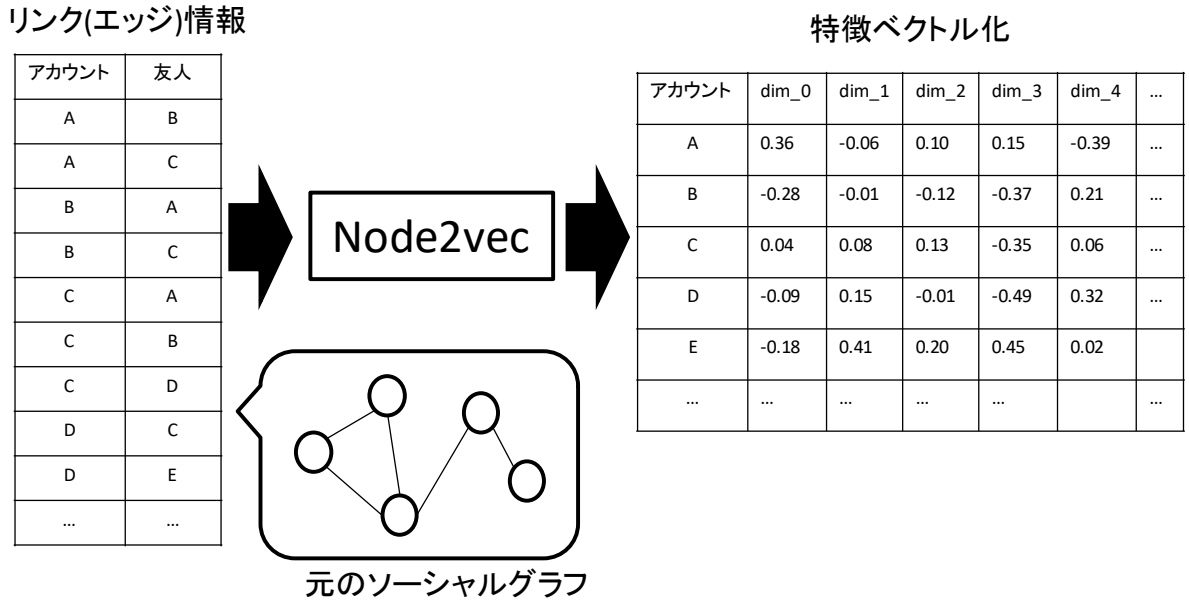


図 6.8　リンク情報の関係ベクトルへの変換

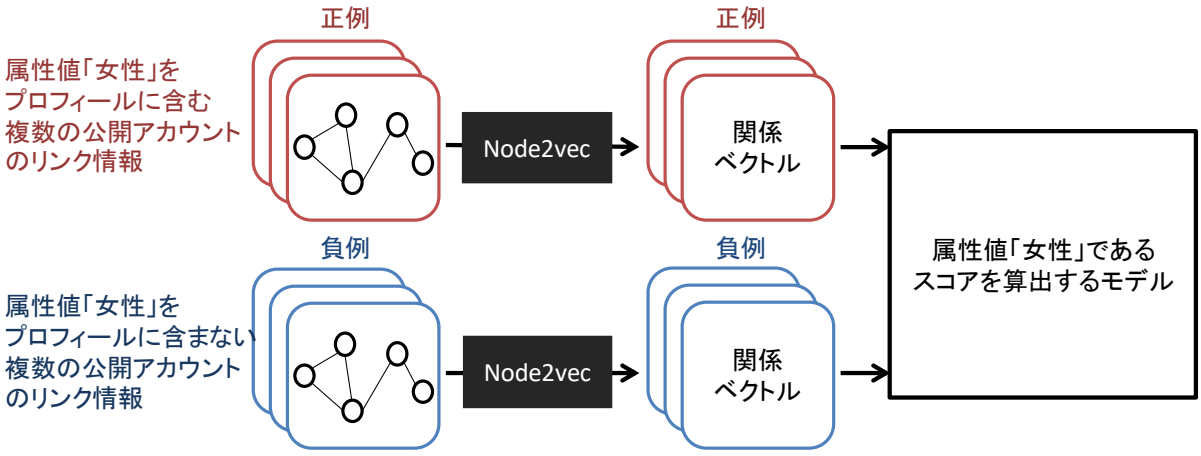


図 6.9　関係ベクトルに変換したリンク情報の学習方法

本手法では，アカウントが保持するリンク情報の各々に対して Node2vec [113]を適用し，リンク情報を低次元ベクトルに変換する．女性を識別するモデル作成を例に挙げて説明する（図 6.9）．正例のリンク情報を，Node2vec を用いて関係ベクトルに変換する．同様に負例のリンク情報も Node2vec を用いて関係ベクトルに変換する．得られたベクトルから女性を識別するモデルを学習する．

リンクの集合をベクトルに変換することで，それ以降の処理については，4 章で述べた投稿文からのプロファイリングと共通化することができている．

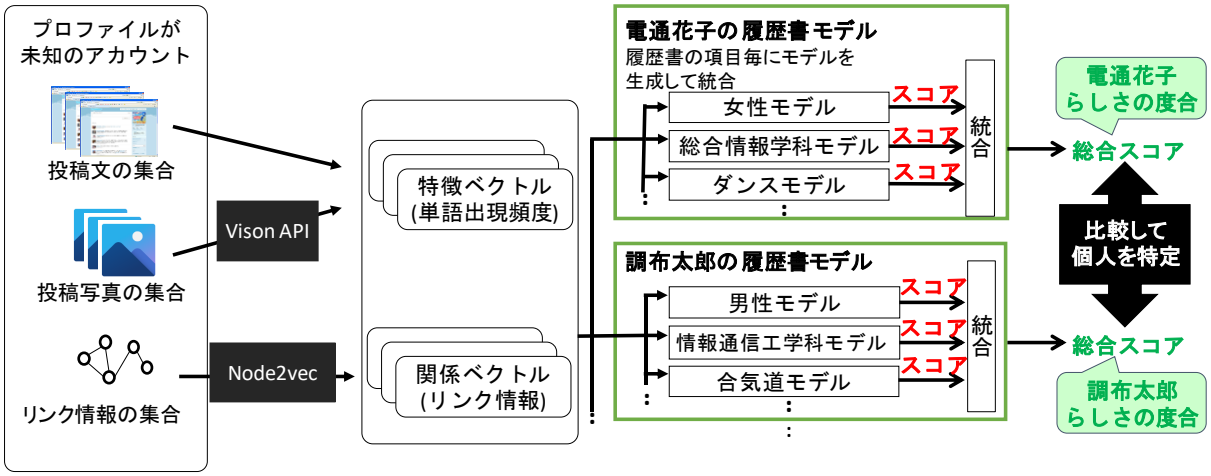


図 6.10　共通の枠組みによる投稿文，投稿画像，リンク情報を用いたプロファイリング

### 6.3.3　データセット

スタンフォード大学が公開している Facebook のリンク情報を用いた [13]．このデータは 4,039 名の被験者アカウント（ノード）から得られた 88,234 件のリンク情報を含んでおり，各アカウントには，匿名化された性別，教育タイプ，言語，仕事等の属性値情報が含まれている．

全アカウント名や属しているグループ名等は匿名化されている．また，属性は明示されているが，属性値は 0，1 などに数値化されており，その意味は明示されていない．例えば，性別の場合，属性値は 0 と 1 としているが，0 と 1 のどちらが男性でどちらが女性であるかは明示されていない．

### 6.3.4　リンク情報を用いたプロファイリングの精度評価

6.3.3 節で述べたデータセットを用い，リンク情報による性別と教育タイプのプロファイリング精度を評価した．これにより，図 6.3 の方法によって，リンク情報からのプロファイリングを通じた個人特定および，リンク情報からのプロファイリングと他の情報からのプロファイリングの併用による個人特定の可能性を示す．

なお，これまでは被験者の Twitter アカウントを用いた実験を行ってきたが，本節ではそれとは別の Facebook アカウントのデータセットを用いる．そのため，本節では，リンク情報によるプロファイリングについてのみ評価し，これまでの結果と組み合わせた上での再特定の精度評価は行わなかった．

性別では男性もしくは女性，教育タイプでは 4 種類の教育タイプのうち，どの教育を受けた経験があるかを推定する．使用したデータセットの各属性値の内訳を表 6.6 に示す．6.3.3 節で述べた通り，各属性値は 0，1 などの数値で表現されており，その数値の意味は明示されていない．性別は 0 か 1 で表現されており，どちらが男性であるかは分からない．また，教育タイプは，「フォーマル教育あり」，「フォーマル教育なしノンフォーマル教育あり」，「フォーマル教育なしノンフォーマル教育なしインフォーマル教育あり」，「教育無し」の 4 種類であり，0～3 に数値化されているが，どの数値がどの教育レベルを表しているかは明示されていない．

表 6.6　データセットの内訳

(a)性別の内訳[1]

| 性別 | アカウント数 |
|---|---|
| 0 | 2,504(62.0%) |
| 1 | 1,531(37.9%) |

(b)教育タイプの内訳[2]

| 教育タイプ | アカウント数 |
|---|---|
| 0 | 1,028(25.4%) |
| 1 | 2,806(69.4%) |
| 2 | 120(2.9%) |
| 3 | 88(2.1%) |

属性値モデルの作成に用いる機械学習アルゴリズムには，XGBoost とロジスティック回帰の 2 通りの方法を適用した．各モデルに入力する教師データは，Node2vec によってリンク情報を 128 次元に圧縮した特徴ベクトルを用いた．

### 6.3.4.1　リンク情報を用いた性別推定

教師データとテストデータは，全 4,035 件のサンプルアカウントを 5-分割交差検証することで作成した．この場合，テストデータの各ラベル数は不均衡となり，ラベル毎の正解率の比較が不平等となるため，以下の 3 つの方法でデータ数を補正した．

① 教師データとテストデータをアンダーサンプリング
　（データ数をラベル 1 の方に揃える）
② 教師データをオーバーサンプリングし，テストデータをアンダーサンプリング
　（データ数をラベル 0 の方に揃える）

[1]教育タイプのアカウント数を合計すると 4,042 件となり，全被験者数の 4,039 名を超えるが，これは，同一人物が複数の教育タイプに該当する場合があるためである．
[2]性別のアカウント数を合計すると 4,034 件となり，全被験者数の 4,039 名を下回るが，これは，性別情報を含まないアカウントが存在するためである．

③　教師データをバギングし，テストデータをアンダーサンプリング

表 6.7　性別モデルの正解率

| | | 平均正解率 | | 5分割交差検定時のデータ数 | |
|---|---|---|---|---|---|
| No. | 教師データの<br>サンプリング手法 | XGBoost | ロジスティック<br>回帰 | 教師データ | テストデータ<br>（全てアンダー<br>サンプリング） |
| 1 | アンダーサンプリング | 0.5532 | 0.5869 | 2448 | 612 |
| 2 | オーバーサンプリング | 0.5561 | 0.5816 | 4008 | 612 |
| 3 | バギング | 0.5542 | 0.5859 | 3228 | 612 |

それぞれの方法における正解率を表 6.7 に示す．①教師データとテストデータをアンダーサンプリングし，学習アルゴリズムにロジスティック回帰を使用した場合の正解率が最も良く， 0.5869 であった．

表 6.8　アンダーサンプリングとロジスティック回帰を使用した場合の混合行列

| | | 推定ラベル | |
|---|---|---|---|
| | | 0 | 1 |
| 正解ラベル | 0 | 170.4 | 135.8 |
| | 1 | 117.2 | 189.0 |

表 6.8 は，アンダーサンプリングとロジスティック回帰の組合せについて，5 分割交差検定で得られたそれぞれの混合行列の結果を平均した値を示している．混合行列は，実際のラベルと，予測されたラベルの内訳を行列としてまとめたものである．正しく推定できていれば，この表の対角線上に値が集中する．

未知である性別を推定する場合，ランダムな推定が正しい確からしさは 0.5 である．一方で，本手法による男女推定の正解率は 0.5 以上である．よって，リンク情報を用いて，有意の精度で性別がプロファイリング可能であると考えられる．

### 6.3.4.2　リンク情報を用いた教育タイプ推定結果

教育タイプの推定でも，5-分割交差検証によって教師データとテストデータを作成したのち，不均衡データを補正するため以下の 2 つの方法でデータ数を補正した．

①　教師データとテストデータをアンダーサンプリング
　（データ数をラベル 1 の方に揃える）

②　教師データをオーバーサンプリングし，テストデータをアンダーサンプリング

（データ数をラベル 0 の方に揃える）

③　教師データをバギングし，テストデータをアンダーサンプリング

表 6.9　教育タイプモデルの正解率

| | | 平均正解率 | | 5分割交差検定時のデータ数 | |
|---|---|---|---|---|---|
| No. | 教師データの<br>サンプリング手法 | XGBoost | ロジスティック<br>回帰 | 教師データ | テストデータ<br>（全てアンダー<br>サンプリング） |
| 1 | アンダーサンプリング | 0.3948 | 0.369 | 280 | 68 |
| 2 | オーバーサンプリング | 0.4123 | 0.4478 | 8980 | 68 |
| 3 | バギング | 0.4737 | 0.4062 | 3233 | 68 |

それぞれの方法における正解率を表 6.9 に示す．教師データをバギング，テストデータをアンダーサンプリングし，学習アルゴリズムに XGBoost を使用した場合の正解率が最も良く，0.4737 であった．

表 6.10　バギングと XGBoost を使用した場合の混合行列

| | | 推定ラベル | | | |
|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 |
| 正解ラベル | 0 | 7.2 | 5.2 | 3.0 | 2.2 |
| | 1 | 2.4 | 9.8 | 2.4 | 3.0 |
| | 2 | 2.8 | 4.0 | 8.0 | 2.8 |
| | 3 | 2.4 | 4.4 | 2.0 | 8.8 |

表 6.10 は，バギングと XGBoost の組合せについて，5 分割交差検定で得られたそれぞれの混合行列の結果を平均した値を示している．

未知である教育タイプを推定する場合，4 種類の教育タイプからランダム推定が正しい確からしさは 0.25 である．本手法による正解率は 0.25 を上回っている．よって，リンク情報を用いて，有意の精度で教育タイプのプロファイリング可能であると考えられる．

## 6.4　提案方式の拡張性について

4 章では，3 章の提案方式において，ソーシャルメディアのデータとして投稿文を用い，有意な精度で再特定が可能であることを示した．5 章では，同じ方式において，データとして，疑似移動履歴すなわち，投稿文中の地名，投稿時刻，および地名と座標の関係情報の組合せを用いることで，再特定の精度が向上できることを示した．6.2 節，6.3 節では，同じ方式において，データとして投稿画像，リンクを用いることで，さらなる精度向上の

可能性を示した．これ以外にも，投稿頻度や投稿時間帯などの情報を用いることも考えられる．以上のように，提案方式は，属性値毎のプロファイリングに基づいて再特定を行うので，ソーシャルメディアの多種多様な情報を活用して再特定精度を向上させることが可能であり，拡張性が高いと考えられる．

## 6.5　提案法の限界および従来法との補完関係

本論文では，匿名のソーシャルメディアのアカウントを個人のプロファイルに照合することで，匿名アカウントから個人を特定する方法を提案した．しかし，提案手法には以下の限界が存在する．

- 提案手法は，履歴書など特定したい人物のプロファイルを事前に入手済みであることを前提としている．そのため，プロファイルの情報を入手できない場合，提案手法を適用することはできない．
- アカウントの情報（テキスト，画像，リンク等）が本人のプロファイルを反映していない場合には，個人特定の精度が低い．例えば，アカウントユーザの趣味が音楽であっても，音楽についてほとんど発言しない場合には，音楽を含むプロファイルと照合できない．

上記の場合でも，従来手法と提案手法とを組み合わせることで個人特定が可能となる場合がある．提案手法と従来手法の組み合わせ方として以下のケースがある．

(1) 特定したい人物のプロファイルを入手できないが，再特定の対象となる匿名アカウント以外に，候補者の実名のアカウントを入手できる場合
(2) 再特定の対象となる匿名アカウントの情報が本人のプロファイルを反映していないが，候補者の実名のアカウントを入手できる場合
(3) 再特定の対象となる匿名アカウントの情報が本人のプロファイルを反映していないが，同一人物の別の匿名アカウントを入手できる場合

それぞれのケースについて，組み合わせ方法を述べる（表 6.11）．

表 6.11 提案手法と従来手法の組み合わせ方

| | | プロファイルを入手可能か？ | |
|---|---|---|---|
| 再特定対象アカウント以外のアカウントを入手可能か？ | アカウント情報がプロファイルを反映している | 可能 | 不可能 |
| 実名アカウントを入手可能 | 反映している | 従来手法<br>提案手法 | 従来手法 →(1) |
| | 反映していない | 従来手法 →(2) | 従来手法 →(1) |
| 匿名アカウントを入手可能 | 反映している | 提案手法 | - |
| | 反映していない | 従来手法＋提案手法 →(3) | - |
| 不可能 | 反映している | 提案手法 | - |
| | 反映していない | - | - |

まず，(1)と(2)の場合について述べる．従来手法は，異なるソーシャルメディアアカウント間の同一人物を特定する．そこで，特定したい人物のプロファイルが入手できない場合でも，候補者の実名アカウントを入手できれば，従来手法を用いて実名アカウントと特定対象のアカウントを紐付けることで，個人特定が可能となる．

(3)の場合，提案手法と従来手法を組み合わせることで個人特定を実現できる．まず，従来手法を用いて同一人物の匿名のソーシャルメディアアカウントのペア Ai，Aj を特定する．次に，提案手法を用いて Ai と Aj の両者からプロファイリングを行い，同一人物の 2 つのプロファイルを求める．それを合体したプロファイルと，照合先プロファイルを対応付けることで，Ai だけと対応付けるよりも，照合精度の向上が期待できる．この方法の具体化と有効性の検証は今後の課題としたい．

## 6.6 まとめ

本章では，これまでの発展形として，投稿文や疑似移動履歴の他に，投稿画像や友人関係を示すリンク情報を用いた個人再特定の実現性を検討した．

投稿画像による個人再特定は，Google Vision API [112]を用いて画像を単語に変換し，得られた単語から特徴ベクトルを作成して個人再特定の枠組みに組み込む方法を検討した．その 51 名の被験者のデータによる評価により投稿文のみ用いる場合の個人特定率が

56.9%であるのに対して，投稿画像も含めた場合は最良で 58.8%となり，投稿画像が個人特定精度の向上に貢献する可能性を示した．

リンク情報による個人再特定は，Node2Vec [113]を用いて友人関係の繋がりを示すソーシャルグラフを関係ベクトルに変換してプロファイリングを行い，個人再特定の枠組みに組み込む方法を検討した．FaceBook が提供しているソーシャルグラフを用いて個人特定を検証し，個人の性別と教育タイプのプロファイリングが優位の精度で可能であることを示した．

また，提案手法は，ソーシャルメディアのデータが本人のプロファイルを反映していない場合は有効でないという課題があるが，従来手法と組み合わせることでこの課題を解消できる可能性を提唱した．

# 第7章　結論

## 7.1　本研究のまとめ

ソーシャルメディアは重要なコミュニケーションインフラである一方，様々な社会問題を引き起こしている．本研究では，ソーシャルメディアの問題のうちプライバシー問題を取り上げた．なかでも，ソーシャルメディアの匿名アカウントから個人が特定されることによるプライバシーリスクの明確化を対象とし，プロファイリングを利用することで高い精度で個人を特定可能な手法を提案し，評価した．

本論文の第 1 章では，ソーシャルメディアの発展と社会における多様な役割を述べると共に，ソーシャルメディアが引き起こす様々な社会問題を明らかにした．ソーシャルメディアのプライバシーを保護するために，匿名性を利用することが多いが，匿名データから個人が特定される再特定の懸念がある．再特定によるプライバシーリスクを明らかにするために，再特定手法の研究が行われているが，従来の再特定手法は，匿名のソーシャルメディアアカウントを同一人物の実名アカウントと照合するので，対象者が複数のアカウントを利用し，そのうちの一つが実名であるという前提を必要とする．そのため，従来手法は実用性が低く，再特定のリスクを示すには不充分であった．また，対象者についてヒントとなる情報が得られても，従来手法はヒント情報を活かす枠組みがなかった．先行研究では，再特定以外に，ソーシャルメディアユーザの属性を推定するプロファイリングの手法が多く提案されている．プロファイリングは再特定に利用可能と考えられるが，従来はプロファイリングと再特定は独立に研究されてきた．そこで，本研究では，従来の再特定手法とは異なる前提条件で動作し，従来手法と補完関係になる手法を確立するために，再特定対象者のプロファイルを前提とし，匿名アカウントのユーザに対するプロファイリングを利用した再特定手法を検討する．

第 2 章では，先行研究をサーベイし，その問題点を明らかにした．ソーシャルメディアのプライバシーに関する研究は，プライバシー問題に関する社会調査，プライバシー情報の漏洩防止対策，ソーシャルメディアから個人情報を抽出する攻撃の 3 つに分類することができる．ソーシャルメディアへの攻撃の研究は，さらに，アカウントユーザの属性を推定するプロファイリングと，匿名アカウントや匿名投稿文から個人を特定する再特定の研究に分類できる．プロファイリング手法の多くは機械学習を用いている．再特定手法は，開発者の経験的知識を自動化したヒューリスティックな手法，友人やフォロー・フォロワー等のリンク構造の照合に基づく手法，機械学習に基づく手法の 3 種類に分類され，

そのうち後者の 2 種類は体系的な発展が可能である．

構造の照合に基づく手法は，複数の匿名アカウント間のリンクから成るグラフと，実名アカウント間のリンクから成るグラフを照合することで，同一人物のアカウントを検知し，匿名アカウントの対象者を再特定する．機械学習に基づく手法は，実名アカウントの投稿文の言語的特徴を学習し，匿名アカウントの投稿文の特徴と照合することで，匿名アカウントの対象者を再特定する．しかし，構造の照合に基づく手法と機械学習に基づく手法のいずれも，対象者が匿名アカウント以外に実名アカウントを公開していることを前提とするため，実用性が低く，プライバシーリスクを示すには不充分である．また，再特定の対象者についてヒント情報を入手できる場合でも，その情報を活かすことができない．構造に基づく手法には，再特定対象のアカウント数と照合先のアカウント数が少なすぎず，かつほぼ同数であるというさらなる制約がある．一方，プロファイリングと個人特定は補完関係にある．プロファイリングの結果を個人特定に利用可能であり，個人特定ができればプロファイリングを促進できると考えられるが，プロファイリングと個人特定の研究は従来独立に進められてきた．

第 3 章では，従来の再特定手法とは異なる前提条件で動作し，その問題点を解決するために，プロファイリングを利用したソーシャルメディアからの個人の再特定の手法を提案した．まず，匿名アカウントからの個人の再特定問題をモデル化し，匿名アカウントのユーザを複数の候補者の中から特定する基本問題，特定の対象者の匿名アカウントを複数のアカウントの中から特定する逆問題，複数の匿名アカウントと同数の候補者を 1 対 1 で対応付ける 1 対 1 問題に分類した．次に，提案手法の設計方針として機械学習の利用を採用し，課題として，従来手法の問題点の解決に加え，機械学習の訓練データの準備の容易化，投稿文だけでなく投稿写真やリンクなどのアカウントの持つ多様な情報の活用を挙げた．

問題のモデル化と設計方針，課題の設定に基づいて，プロファイリングに基づく再特定手法を提案した．提案法は，再特定候補者のプロファイルを前提として，匿名アカウントの未知のユーザの属性値をプロファイリングし，候補者のプロファイルと照合することで，匿名アカウントと候補者が同一人物である確からしさを定量化する．すなわち，候補者のプロファイルに含まれる各属性値について，匿名アカウントのユーザが当該属性値を有する確からしさをプロファイリングし，これらの属性値毎の確からしさを全属性について統合することで，匿名ユーザと候補者が同一人物である確からしさを定量化する．この定量値に基づいて，匿名アカウントと候補者を対応付ける．

提案法は，匿名アカウントを別のアカウントと照合しない．また，再特定対象者に関するヒント情報（ラグビー観戦が趣味）を属性値とみなし，その属性値について匿名アカ

ウントをプロファイリングすることで，ヒント情報を活かすことができる．プロファイリングのための訓練データの準備が容易であるか，投稿文だけでなく様々な情報からプロファイリングできるかは，第 4 章以降の具体例において評価する．

第 4 章では，匿名アカウントのデータのうち投稿文を用い，再特定候補者のプロファイルとして履歴書を用いて，提案手法を実現，評価した．履歴書中の各属性値について，匿名アカウントのユーザが当該属性値を有する確からしさを投稿文からプロファイリングし，これらの属性値毎の確からしさを履歴書の全属性について統合することで，匿名アカウントと履歴書の同一人物である確からしさを定量化する．この定量値に基づいて，複数の匿名アカウントと履歴書の間で同一人物のペアを推定する．本実現方式は，3 章の基本方式の特徴を受け継いでおり，従来法の問題点を解決できる．また，属性値識別モデルを学習するための訓練データを公開アカウントから自動収集できるので，データの準備を容易化できる．

予備評価によって，機械学習アルゴリズム，特徴量，利用する属性値，スコア統合方法を選定した後，30 人の被験者データによる評価を行った．その結果，機械学習アルゴリズムとして XGBoost，特徴量として Bag-of-words，属性値として全属性，スコア統合方法として平均を用いた場合が最良で， 37%（11 個）のアカウントを本人の履歴書に正しく紐づけることができた．

さらに，78 人の被験者データを用いた 2 次評価を実施し，匿名アカウントから履歴書を特定する基本問題では 49%の正解率，履歴書からアカウントを特定する逆問題では 55%，アカウントと履歴書を 1 対 1 に紐づける問題では 50%の正解率を達成した．

第 5 章では，匿名アカウントのデータのうち投稿文中の地名を用い，さらに，地名と緯度経度の関係についての情報を用いた．これらの情報から投稿者の移動履歴を推定し，住所をプロファイリングした．2 つの住所の一方を正しく推定できるかの予備評価では，4 章で述べた投稿文のみ用いる方式では正解率 69%，移動履歴を用いる方式は 75%，投稿文と移動履歴の特徴量レベル統合方式およびスコアレベル統合方式はいずれも 85%であった．さらに，4 章と同じ 78 人の被験者データを用いて現住所をプロファイリングした結果，投稿文のみの正解率は 65%，移動履歴のみは 74%，投稿文と疑似移動履歴の特徴レベル統合方式は 65%，投稿文と移動履歴のスコアレベル統合方式は 70%であった．これの結果により，移動履歴すなわち地名情報と地名―座標関係の情報を用いることで，現住所のプロファイリング精度の向上が可能であることを明らかにした．

さらに，投稿文のみ，移動履歴のみ，投稿文と移動履歴のスコアレベル統合の 3 方式によって現住所をプロファイリングし，そのスコアと現住所以外のスコアを統合すること

で，78 人のソーシャルメディアアカウントから個人を再特定する実験を行った．その結果，4 章で述べた投稿文のみを用いる方式に比べ，匿名アカウントから履歴書を特定する基本問題では 49%から 55%，履歴書からアカウントを特定する逆問題では 55%から 59%，アカウントと履歴書を 1 対 1 に紐づける問題では 50%から 64%に特定精度が向上することを明らかにした．

情報の多寡という観点では，投稿文と疑似移動履歴の両者を用いたプロファイリングが最も精度が高くなり，その結果，再特定精度も高くなると考えられる．しかし，今回の評価結果では，移動履歴のみ用いて現住所をプロファイリングする方式が，最も高精度であった．そのため，投稿文と移動履歴の両者の情報を活用するための統合方式の高度化が今後の課題となる．

第 6 章では，提案方式の拡張性と限界，従来方式との補完関係について述べた．4 章では，ソーシャルメディアアカウントの情報のうち投稿文を用いて提案方式を実現し，プロファイリングと再特定について有意の精度を明らかにした．5 章では，アカウントの情報のうち移動履歴を用いて，現住所のプロファイリング精度を向上し，投稿文の利用と組み合わせることで再特定の精度を向上した．本章では，アカウントの情報のうち投稿画像およびリンクの利用を検討した．投稿写真をキーワード集合に変換した上で趣味をプロファイリングし，一部の趣味については推定精度を向上した．投稿文の利用と併用することで，投稿文のみ用いた場合に比べて再特定の正解率を約 2%向上した．また，投稿文，移動履歴，投稿写真をすべて利用することで，さらに再特定精度を向上することができた．一方，アカウントの他のアカウントへのリンクを Node2Vec によってベクトル化した上で性別および教育種別のプロファイリングを試み，有意の精度を明らかにすることで，リンク情報を用いたさらなる再特定精度の向上の可能性を示した．以上の結果，3 章で提案した方法は，アカウントの様々な情報を用いて実現可能であり，複数種類の情報を組み合わせることで精度の向上が可能であることから充分な拡張性を有することを明らかにした．

提案方式は，攻撃者が再特定候補者のプロファイルを入手可能であること，ソーシャルメディアの情報がユーザのプロファイルを反映していることが前提となる．一方，従来方式は，再特定対象者が匿名アカウント以外に実名アカウントを有することが前提となる．両者の前提条件が異なるので，併用すれば，より広い条件で再特定が可能となる．さらに，両者の前提条件のいずれも満たさない場合でも，再特定対象者が複数の匿名アカウントを有し，複数のアカウントの情報を統合すればプロファイリングの精度を向上可能である場合には，再特定の可能性があることを明らかにした．

以上のように，本論文では，従来独立に進められてきたプロファイリングと再特定の

研究を統合し，匿名アカウントのユーザのプロファイリング結果を再特定候補者のプロファイルに照合する手法を提案した．ソーシャルメディアの情報のうち投稿文，移動履歴，投稿画像およびリンクを用いて提案方式の実現性を検討し，有意の再特定精度および拡張性を明らかにした．提案方式は，従来方式の前提条件を不要化すると共に，従来方式と組み合わせることで，より広い範囲の再特定を可能とする．

## 7.2 今後の課題

今後の課題として以下の研究に取り組みたい．

(1) 移動履歴のみを用いて現住所をプロファイリングした場合と移動履歴と投稿文の組合せによりプロファイリングした場合を比べると，組合せの精度は低かった．また，投稿画像と投稿文の組合せによって趣味の推定精度を向上することはできたが，再特定の精度はわずかであった．このように，アカウントの情報を広く併用しても精度向上につながらない場合がある．その原因を分析し，複数情報のより効果的な組合せ方法を検討する．

(2) リンク情報の利用については，まだ性別と教育レベルのプロファイリングに適用しただけであるため，より広い属性値に適用して有効性を確認すると共に，再特定精度への効果を実証する．

(3) より高精度な再特定の可能性を明らかにし，より効果的に社会に警鐘を鳴らし，漏洩防止技術につなげることを目的とする．具体的には，提案法を大規模実装した上で，メディアや展示会等で発表する．また，提案法を用いて，投稿文，写真，リンク等のソーシャルメディアアカウントのあらゆる情報からの再特定の可能性を検査し，ユーザへの通知および，言い換え，秘匿を行う技術を開発する．

# 謝辞

本研究を遂行し，学位論文としてまとめるにあたり，2020 年度までは主任指導教員として，その後も終始多大なるご指導とご教示を頂いた吉浦裕教授，2021 年度以降主任指導教員としてご指導いただいた市野将嗣准教授，副指導教員としてご指導とご教示を頂いた崎山一男教授に心より感謝の意を表します．また，博士論文の審査委員として，御指導いただいた岩本貢教授，松本光春准教授に深く感謝申し上げます．そして，吉浦研の研究室の皆様に深く感謝申し上げます．

最後に，社会人として働きながらの学位取得にご理解いただき激励くださった NTT ドコモの職場の皆様，そして，学位取得をいつも温かく応援してくれた夫，両親，家族の皆さまに深い感謝の意を表して謝辞といたします．

# 参考文献

[1]【最新版】2021 年 7 月更新．12 のソーシャルメディア最新動向データまとめ from <https://gaiax-socialmedialab.jp/post-30833/> (accessed 2021-07-24).

[2]LINE Business Guide_202101-06.pdf from <https://www.linebiz.com/sites/default/files/media/jp/download/LINE%20Business%20Guide_202101-06.pdf> (accessed 2021-07-24).

[3]How Many Users Does Clubhouse Have? 40+ Clubhouse Stats (2021) from <https://backlinko.com/clubhouse-users> (accessed 2021-07-24).

[4]諸外国の人達はどの媒体でニュースを目にしているのだろうか(2020 年公開版)（不破雷蔵） - 個人 - Yahoo!ニュース from <https://news.yahoo.co.jp/byline/fuwaraizo/20200321-00167866/> (accessed 2021-07-24).

[5]M. Delirrad and A. B. Mohammadi：New Methanol Poisoning Outbreaks in Iran Following COVID-19 Pandemic, Alcohol and Alcoholism Vol.55, No.4, pp.347-348 (2020)

[6]H. Allcott and M. Gentzkow：Social Media and Fake News in the 2016 Election, Alcohol and Alcoholism Vol.31, No.2, pp.211-36 (2017)

[7]How Many Users Does Clubhouse Have? SNS 投稿が発端の空き巣被害増　リアルタイム投稿にリスク｜NEWS ポストセブン - Part 2 from <https://www.news-postseven.com/archives/20190519_1371490.html/2> (accessed 2021-07-24).

[8]ツイッターへの書き込みでストーカー容疑？ 規制される行為の内容とは？ from <https://kanazawa.vbest.jp/columns/criminal/g_sex/1096/> (accessed 2021-07-24).

[9]「Exif 住所確認」 from <https://apps.apple.com/jp/app/exif-%E4%BD%8F%E6%89%80%E7%A2%BA%E8%AA%8D/id547108880> (accessed 2021-07-24).

[10]WeKnowYourHouse.com from <http://ww1.weknowyourhouse.com/> (accessed 2021-07-24).

[11]NHK「ニュースウオッチ 9」の空き巣と SNS について取材を受けました：新倉茂彦の情報セキュリティ AtoZ：オルタナティブ・ブログ from <https://blogs.itmedia.co.jp/niikura/2013/11/nhk9sns-2a27.html> (accessed 2021-07-24).

[12]総務省｜平成 27 年版 情報通信白書｜SNS の利用率 from <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/nc242220.html> (accessed 2021-07-24).

[13]SNAP: Network datasets: Social circles from <http://snap.stanford.edu/data/egonets-Facebook.html> (accessed 2021-07-24).

[14]R. Gross and A. Acquisti：Information Reve-lation and Privacy in Online Social Networks, Proc. ACM Workshop on Privacy in the Electronic Society, pp.71-80 (2005).

[15]K. Lewis, J. Kaufman and N. Christakis：The Taste for Privacy: An Analysis of College Student Privacy Settings in an Online Social Network, Journal of Computer-Mediated Communication Vol.14, No.1, pp.79-100 (2008)

[16]F. Viegas：Bloggers' Expectations of Privacy and Accountability: An Initial

Survey, Journal of Computer-Mediated Communication Vol.10, No.3 (2005)

[17]R. Dey, Z. Jelveh and K. Ross：Facebook users have become much more private: A large-scale study, Proc. IEEE International Conference on Pervasive Computing and Communications Workshops, pp.346-352 (2012).

[18]Y. Liu, K.P. Gummadi, B. Krishnamurthy, and A. Misolve：Analyzing facebook privacy settings: user expectations vs. reality, Proc. ACM SIGCOMM conference on Internet measurement conference, pp.61-70 (2011).

[19]インターネットサービス利用時の情報公開範囲の設定に注意！ from <http://www.ipa.go.jp/security/txt/2013/10outline.html> (accessed 2021-07-24).

[20]EA. Baatarjav, R. Dantu and S. Phithakkitnukoon：Privacy Management for Facebook, Proc. International Conference on Information Systems Security, Vol.5352, pp.273-286 (2008)

[21]C. Shane-Simpson, A. Manago, N. Gaggi, and K. Gillespie-Lynch: Why do college students prefer Facebook, Twitter, or Instagram? Site affordances, tensions between privacy and self-expression, and implications for social capital, Proc. Computers in Human Behavior, Vol.86, pp.276-288 (2018).

[22]竹地 潔: ビッグデータ時代におけるプロファイリングと労働者への脅威, Proc. 富山大学紀要．富大経済論集, Vol.63, No.1, pp.1-19 (2017).

[23]『プロファイリングの法的諸論点（試論）憲法の観点から』情報通信法学研究会平成29年度第2回 from <https://www.soumu.go.jp/main_content/000533322.pdf> (accessed 2021-07-24).

[24]町田 史門, 梶山 朋子, 嶋田 茂 and 越前 功: センシティブデータの漏洩検知による適応的な公開範囲設定システムのプロトタイプ実装, Proc. 電子情報通信学会技術研究報告, Vol.113, No.480, pp.51-56 (2014).

[25]L, Banks and SF, Wu: All friends are not created equal: An interaction intensity based approach to privacy in online social networks, Proc. International Conference on Computational Science and Engineering, Vol.4, pp.970-974 (2009).

[26]H Mao, X Shuai and A Kapadia: Loose tweets: an analysis of privacy leaks on twitter, Proc. 10th annual ACM workshop on Privacy in the electronic society, pp.1-12 (2011).

[27]曽根原 愛理, 白鷹 靖子, 小舘 亮之, 並河 大地, 南 裕也, and 下村 道夫: ペルソナ情報を用いた画像共有サービスにおける開示制御技術の提案, Proc. 電子情報通信学会技術研究報告, Vol.111, No.470, pp.97-102 (2012).

[28]Z. He, Z. Cai and J. Yu: Latent-data privacy preserving with customized data utility for social network data, Proc. IEEE Transactions on Vehicular Technology, Vol.67, No.1, pp.665-673 (2017).

[29]GDPR（General Data Protection Regulation：一般データ保護規則） from <https://www.ppc.go.jp/enforcement/infoprovision/laws/GDPR/> (accessed 2021-07-24).

[30]Chromium Blog: Building a more private web: A path towards making third party cookies obsolete from <https://blog.chromium.org/2020/01/building-more-private-web-path-towards.html> (accessed 2021-07-24).

[31]Off-Facebook Activity: Control your information from https://www.facebook.com/off-facebook-activity> (accessed 2021-07-24).

[32]IF. Lam, KT. Chen and LJ. Chen: Involuntary information leakage in social network services, Proc. International Workshop on Security, pp.167-183 (2008).

[33]I. Polakis, G. Kontaxis, et al.: Using social networks to harvest email addresses, Proc. 9th annual ACM workshop on Privacy in the electronic society, pp.11-20 (2010).

[34]O. Goga, H. Lei, S. Krishnan, G. Friedland, R. Sommer and R. Teixeira: On exploiting innocuous user activity for correlating accounts across social network sites, Proc. International Computer Science Institute Technical Reports(2012).

[35]WH. Lee, et al.: Blind De-anonymization Attacks using Social Networks, Proc. Workshop on Privacy in the Electronic Society, pp.1-4 (2017).

[36]Y. Shao, et al.: Fast de-anonymization of social networks with structural information., Proc. Workshop on Privacy in the Electronic Society, Vol.4, No.1, pp.76-92 (2019).

[37]H. Jiang, et al.: SA framework based de-anonymization of social networks, Proc. Procedia Computer Science, Vol.129, pp.358-363 (2018).

[38]A. Narayanan, et al.: On the feasibility of internet-scale author identification., Proc. IEEE Symposium on Security and Privacy, pp.300-314 (2012).

[39]M. Almishari, D. Kaafar, E. Oguz and G. Tsudik: Stylometric linkability of tweets, Proc. 13th Workshop on Privacy in the Electronic Society, pp.205-208 (2014).

[40]R. Overdorf and R. Greenstadt: Blogs, Twitter Feeds, and Reddit Comments: Cross-domain Authorship Attribution., Proc. Privacy Enhancing Technologies, Vol.3, pp.155-171 (2016).

[41]H. Mao, X. Shuai and A. Kapadia: Loose tweets: an analysis of privacy leaks on twitter, Proc. 10th annual ACM workshop on Privacy in the electronic society, pp.1-12 (2011).

[42]G. Kótyuk and L. Buttyán: A machine learning based approach for predicting undisclosed attributes in social networks, Proc. IEEE International Conference on Pervasive Computing and Communications Workshops., pp.361-366 (2012).

[43]A. Caliskan Islam, J. Walsh and R. Greenstadt: 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).

[44]M. Hart, P. Manadhata and R. Johnson: Text Classification for Data Loss Prevention, Proc. Privacy Enhancing Technologies, pp.18-37 (2011).

[45]Z. Cheng, J. Caverlee and K. Lee: You are where you tweet: a content-based approach to geo-locating twitter users., Proc. 19th ACM international conference on Information and knowledge management, pp.759-768 (2010).

[46]J. Burger, J. Henderson, et al.: Discriminating gender on Twitter, Proc. Conference on Empirical Methods in Natural Language Processing, pp.1301-1309 (2011).

[47]M. Pennacchiotti and A. Popescu: A machine learning approach to twitter user classification., Proc. International AAAI Conference on Web and Social Media, Vol.5, No.1, pp.281-288 (2011).

[48]S. Gurses, R. Rizk and O. Gunther: Privacy design in online social networks: Learning from privacy breaches and community feedback, Proc. International Conference on Information Systems (2008).

[49]A. Acquisti, R. Gross and F. Stutzman: Faces of Facebook: Privacy in the Age of Augmented Reality, Proc. BlackHat USA, (2011).

[50]インプレスジャパン: 『インターネット白書 2011』, インプレスコミュニケーションズ, pp.196 (2011).

[51]B. Meeder, J. Tam, P. Kelly, and L.F. Cranor: RT@ IWantPrivacy: Widespread violation of privacy settings in the Twitter social network., Proc. Web 2.0 Privacy and Security Workshop, Vol.2, No.1.2, pp.281-288 (2010).

[52]M. Häsel and LL. Iacono: Security in OpenSocial-Instrumented Social Networking Services, Proc. IFIP International Conference on Communications and Multimedia Security, pp.40-52 (2010).

[53]ガイアックス, 『Facebook ユーザーの時間・シチュエーションによる利用動向調査』を実施 from <http://gaiax-socialmedialab.jp/press/003> (accessed 2021-07-24).

[54]CareerBuilder.com, One-in-Five Employers Use Social Networking Sites to Reasearch Job Candidates, CareerBuilder.com Survey Finds from <http://www.careerbuilder.com/share/aboutus/pressreleasesdetail.aspx?id=pr459&sd=9%2F10%2F2008&ed=12%2F31%2F2008> (accessed 2021-07-24).

[55]California Background Check & Pre-Employment Laws from <https://crimcheck.net/california-background-check-laws/#:~:text=Regarding%20the%20use%20of%20social,applicant%20to%20access%20their%20social> (accessed 2021-07-24).

[56]町田 史門, 嶋田 茂 and 越前 功: SNS 上のプライバシーセンシティブ情報の漏洩検知に基づく公開範囲の設定方式, Proc. コンピュータセキュリティシンポジウム 2013 論文集, pp.566-573 (2013).

[57]Y. Zhu, Z. Hu, H. Wang, H. Hu and G. Ahn: A Collaborative Framework: for Privacy Protection in Online Social Networks, Proc. 6th International Conference on Collaborative Computing (2010).

[58]L. Backstrom, C. Dwork and J. Kleinberg: Wherefore Art Thou R3579X? Anonymized Social Networks, Hidden Patterns, and Structural Steganography, Proc. 16th international conference on World Wide Web, pp.181-190 (2007).

[59]A. Narayanan and V. Shmatikov: De-anonymizing Social Networks, Proc. 30th Symposium on Security and Privacy, pp.173-187 (2009).

[60]S. Nilizadeh, A. Kapadia and YY. Ahn: Community-Enhanced De-anonymization of Online Social Networks, Proc. ACM SIGSAC conference on computer and communications security, pp.537-548 (2014).

[61]GG. Gulyás and S. Imre: Measuring Importance of Seeding for Structural De-anonymization Attacks in Social Networks, Proc. IEEE International Conference on Pervasive Computing and Communication Workshops, pp.610-615 (2014).

[62]GG. Gulyás, B. Simon and S. Imre: An Efficient and Robust Social Network De-anonymization Attack, Proc. ACM on Workshop on Privacy in the Electronic Society, pp.1-11 (2016).

[63]R. Shigenaka, T. Yukihiro and K. Noriji: Content-Aware Multi-task Neural Networks for User Gender Inference Based on Social Media Images, Proc. IEEE International Symposium on Multimedia, pp.169-172 (2016).

[64]NZ. Gong and B. Liu: Attribute Inference Attacks in Online Social Networks, Proc. ACM Transactions on Privacy and Security, Vol.21, No.1, pp.1-30 (2018).

[65]E. Ferrara, WQ. Wang, O. Varol, A. Flammini and A. Galstyan: Predicting online extremism, content adopters, and interaction reciprocity, Proc. International conference on social informatics, pp.22-39 (2016).

[66]J. Weerasinghe, K. Morales and R. Greenstadt: Analyzing Machine Learning Models that Predict Mental Illnesses from Social Media Text, Proc. Privacy Enhancing Technologies Symposium (2018).

[67]M. Srivatsa and M. Hicks: Deanonymizing mobility Traces: Using Social Networks as a Side-Channel, Proc. ACM conference on Computer and communications security, pp.628-637 (2012).

[68]A. Krizhevsky, S. Ilya and EH. Geoffrey: Deanonymizing mobility Traces: ImageNet Classification with Deep Convolutional Neural Networks, Proc. 25th International Conference on Neural Information Processing Systems, Vol.1, pp.1097-1105 (2012).

[69]F. Perronnin, J. Sánchez and T. Mensink: Improving the Fisher Kernel for Large-Scale Image Classification, Proc. European conference on computer vision., pp.143-156 (2010).

[70]R. Dey, C. Tang, K. Ross and N. Saxena: Estimating age privacy leakage in online social networks, Proc. IEEE INFOCOM, pp.2836-2840 (2012).

[71]A. Chaabane, G. Acs and MA. Kaafar: You Are What You Like! Information Leakage Through Users' Interests, Proc. Network & Distributed System Security Symposium (2012).

[72]R. Zafarani and H. Liu: Connecting users across social media sites: a behavioral-modeling approach, Proc. 19th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.41-49 (2013).

[73]F. Zhou, L. Liu, K. Zhang, G. Trajcevski, J. Wu and T. Zhong: DeepLink: A Deep Learning Approach for User Identity Linkage, Proc. IEEE Conference on Computer Communications, pp.1313-1321 (2018).

[74]M. Wang, Q. Tan, X. Wang and J. Shi: De-anonymizing Social Networks User via Profile Similarity, Proc. IEEE Third International Conference on Data Science in Cyberspace, pp.889-895 (2018).

[75]C. Li, S. Wang, Y. Wang, P. Yu, Y. Liang, Y. Liu and Z. Li: Adversarial Learning for Weakly-Supervised Social Network Alignment, Proc. AAAI Conference on Artificial Intelligence, Vol.33, No.01, pp.996-1003 (2019).

[76]B. Perez, M. Musolesi and G. Stringhini: You are your Metadata: Identification and Obfuscation of Social Media Users using Metadata Information, Proc. Twelfth International AAAI Conference on Web and Social Media (2018).

[77]R. Shokri, G. Theodorakopoulos, JY. Le Boudec and JP. Hubaux: Quantifying location privacy, Proc. 32nd IEEE Symposium on Security and Privacy, pp.247-262 (2011).

[78]T. Murakami: A succinct model for re-identification of mobility traces based on small training data., Proc. 15th International Symposium on Information Theory and Its Applications, pp.164-168 (2018).

[79]CY. Ma, DK. Yau, NK. Yip and NS. Rao: Privacy vulnerability of published anonymous mobility traces, Proc. IEEE/ACM Transactions on Networking, Vol.21, No.3, pp.720-733 (2013).

[80]S. Gambs, M-O. Killijian and MN. Cortez: De-anonymization attack on geolocated data, Proc. 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, pp.789-797 (2013).

[81]Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang and H. Liu: E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures, Proc. 20th annual international conference on Mobile computing and networking, pp.617-628 (2014).

[82]J. Su, A. Shukla, S. Goel and A. Narayanan: De-anonymizing Web Browsing Data with Social Networks, Proc. 26th International Conference on World Wide Web, pp.1261-1269 (2017).

[83]T. Minkus and KW. Ross: I Know What You're Buying: Privacy Breaches on eBay, Proc. International Symposium on Privacy Enhancing Technologies Symposium, pp.164-183 (2014).

[84]SK. Karmaker Santu, V. Bindschadler, C. Zhai and CA. Gunter: NRF: A Naive Re-identification Framework, Proc. Workshop on Privacy in the Electronic Society, pp.121-132 (2018).

[85]G. Danezis and C. Troncoso: You cannot hide for long: De-anonymization of real-world dynamic behaviour, Proc. 12th ACM workshop on Workshop on privacy in the electronic society, pp.49-60 (2013).

[86]S. Ji, W. Li, M. Srivatsa and R. Beyah: Structural Data De-anonymization: Quantification, Practice, and Implications, Proc. ACM SIGSAC Conference on Computer and Communications Security, pp.1040-1053 (2014).

[87]Q. Lu, F. Chen and K. Hancock: On path anomaly detection in a large transportation network, Proc. Mathematics, Computer Science, Vol.33, No.6, pp.448-462 (2009).

[88]松本 瞬, 大岡 拓斗, 市野 将嗣 and 吉浦 裕: 対象者の人数と対象者間の関係に制約のない移動履歴とソーシャルネットワークの照合方式, Proc. 情報処理学会論文誌, Vol.61, No.12, pp.1814-1830 (2020).

[89]Twitter プロフィール検索 - ツイプロ from <https://twpro.jp/> (accessed 2021-07-24).

[90]LO. Alvares, V. Bogorny, JA. Fernandes de Macedo, B. Moelans and S. Spaccapietra: Dynamic modeling of trajectory patterns using data mining and reverse engineering, Proc. Tutorials, posters, panels and industrial contributions at the 26th international conference on Conceptual modeling, Vol.83, pp.149-154 (2007).

[91]LO. Alvares, V. Bogorny, B. Kuijpers, B. Moelans, JA. Fern, ED. Macedo and AT. Palma: Towards Semantic Trajectory Knowledge Discovery, Proc. Data Mining and Knowledge Discovery (2007).

[92]Z. Chen: Mining individual behavior pattern based on significant locations and spatial trajectories, Proc. IEEE International Conference on Pervasive Computing and Communications Workshops, pp.540-541 (2012).

[93]A. Monreale, F. Pinelli, R. Trasarti and F. Giannotti: WhereNext: a location predictor on trajectory pattern mining, Proc. 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.637-646 (2009).

[94]L. Chen, M. Lv, Q. Ye and G. Chen: A personal route prediction system based on trajectory data mining, Proc. Information Sciences, Vol.181, No.7, pp.1264-1284 (2011).

[95]AY. Xue, R. Zhang, Y. Zheng, X. Xie, J. Huang and Z. Xu: Destination Prediction by Sub-Trajectory Synthesis and Privacy Protection Against Such Predictio, Proc. 29th IEEE International Conference on Data Engineering, pp.254-265 (2013).

[96]S. Dodge, R. Weibel and E. Forootan: Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects, Proc. Computers, Environment and Urban Systems, Vol.33, No.6, pp.419-434 (2009).

[97]Y. Zheng, Y. Chen, Q. Li, X. Xie and WY. Ma: Understanding transportation

modes based on GPS data for web applications, Proc. ACM Transactions on the Web, Vol.4, No.1, pp.1-36 (2010).

[98]DJ. Patterson, L. Liao, D. Fox and H. Kautz: Inferring High-Level Behavior from Low-Level Sensors, Proc. International Conference on Ubiquitous Computing, pp.73-89 (2003).

[99]F. Giannotti, M. Nanni, F. Pinelli and D. Pedreschi: Trajectory pattern mining, Proc. 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.330-339 (2007).

[100]Y. Matsuo, N. Okazaki, K. Izumi, Y. Nakamura, T. Nishimura, K. Hasida and H. Nakashima: Inferring Long-term User Properties Based on Users' Location History., Proc. 20th International Joint Conference on Artificial Intelligence, pp.2159-2165 (2007).

[101]藤田 将成, 手塚 博久, 武藤 伸洋, 南 弘征 and 水田 正弘: GPS 移動履歴からの接触可能性キーワード抽出法と嗜好推定法の提案, Proc. 行動計量学, Vol.40, No.1, pp.3-15 (2013).

[102]Y. Zheng, L. Zhang, X. Xie and WY. Ma: Mining interesting locations and travel sequences from GPS trajectories, Proc. 18th international conference on World wide web, pp.791-800 (2009).

[103]相 尚寿: 観光研究への位置情報ビッグデータ展開の可能性, Proc. 観光科学研究 (2014).

[104]高柳 健司, 村尾 和哉, 望月 祐洋, and 西尾 信彦: 間欠的人流センシングにおける回遊状況推定, Proc. マルチメディア，分散協調とモバイルシンポジウム 2016 論文集, pp.234-241 (2016).

[105]D. Manousakas, et al.: Quantifying privacy loss of human mobility graph topology, Proc. 18th Privacy Enhancing Technologies Symposium, pp.5-21 (2018).

[106]C. Riederer, et al: Linking users across domains with location data: theory and validation, Proc. 25th International Conference on World Wide Web, pp.707-719 (2016).

[107]Twitter の API について from <https://help.twitter.com/ja/rules-and-policies/twitter-api> (accessed 2021-07-24).

[108]MeCab: Yet Another Part-of-Speech and Morphological Analyzer from <https://taku910.github.io/mecab/> (accessed 2021-07-24).

[109]GeoNLP - テキストを自動的に地図化する地名情報処理ソフトウェア from <https://geonlp.ex.nii.ac.jp/> (accessed 2021-07-24).

[110]instagram from <https://www.instagram.com/?hl=ja> (accessed 2021-07-24).

[111]丸山 翼, 市野 将嗣, 吉浦 裕: ソーシャルネットワークの投稿写真からの投稿者の属性推定とその効率向上, (2016).

[112]Cloud Vision のドキュメント | Cloud Vision API | Google Cloud from <https://cloud.google.com/vision/docs?hl=ja> (accessed 2021-07-24).

[113]node2vec from <https://snap.stanford.edu/node2vec/> (accessed 2021-07-24).

[114]Twitter’s vast metadata haul is a privacy nightmare for users from < https://www.wired.co.uk/article/twitter-metadata-user-privacy> (accessed 2021-07-24).

[115]Twitter のメタデータがあれば、個人を正確に特定できる：研究結果 from <

https://wired.jp/2018/08/19/twitter-metadata-user-privacy/> (accessed 2021-07-24).

[116]E. Hashimoto, M. Ichino, T. Kuboyama, I. Echizen and H. Yoshiura: Breaking Anonymity of Social Network Accounts by Using Coordinated and Extensible Classifiers based on Machine Learning, Proc. 15th IFIP Conference on e-Business, e-Services and e-Society, Lecture Notes in Computer Science 9844, pp.455-470 (2016)

[117]あなたもＡＩに選別される？｜サイカルジャーナル｜NHK NEWS WEB from < https://www3.nhk.or.jp/news/special/sci_cul/2018/06/story/special_180611/> (accessed 2021-07-24).

[118]星 周一郎: 街頭設置カメラの高機能化・生体認証機能と個人情報該当性：改正個人情報保護法と防犯カメラ条例の意義, 法学会雑誌, 57(2), pp.211-243 (2017).

[119]ストーカー、「瞳に映った景色」で女性の自宅を特定　日本 - BBC ニュース from < https://www.bbc.com/japanese/50010809> (accessed 2021-07-24).

[120]NZ. Gong, A. Talwalkar, L. Mackey and L. Huang: Joint Link Prediction and Attribute Inference Using a Social-Attribute Network, Proc. ACM Transactions on Intelligent Systems and Technology, Vol.5, Issue 2, No.27, pp.1-20 (2014).

[121]HQ. Nguyen-Son, QB. Nguyen, MT. Tran, DT. Nguyen, H. Yoshiura and I. Echizen: Automatic Anonymization of Natural Languages Texts Posted on Social Networking Services and Automatic Detection of Disclosure, Proc. 2012 Seventh International Conference on Availability, Reliability and Security, pp.358-364 (2012).

# 関連論文の印刷公表の方法および時期

- 学術論文

  全著者名：橋本 英奈 , 宮崎 夏美 , 市野 将嗣 , 久保山 哲二 , 越前 功 , 吉浦 裕

  論文題目：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

  印刷公表の方法および時期：情報処理学会論文誌, Vol. 58, No. 12, pp.1863-1874, 2017

  （3 章，4 章に関連）

- 国際学会発表論文

1. 全著者名：Eina Hashimoto, Masatsugu Ichino, Tetsuji Kuboyama, Isao Echizen,Hiroshi Yoshiura

   論文題目：Breaking Anonymity of Social Network Accounts by Using Coordinated and Extensible Classifiers based on Machine Learning

   印刷公表の方法および時期：The 15th IFIP Conference on e-Business, e-Services and e-Society, Lecture Notes in Computer Science 9844, pp.455-470, 2016 (最優秀論文賞受賞)

   （4 章に関連）

2. 全著者名：Eina Hashimoto, Masatsugu Ichino, and Hiroshi Yoshiura

   論文題目：Linking anonymous SNS accounts and real-world people through profiling

   印刷公表の方法および時期：Proceedings of International Workshop on Informatics (IWIN), pp. 127-134, 2019

   （3 章，4 章に関連）

3. 全著者名：Eina Hashimoto, Masatsugu Ichino, and Hiroshi Yoshiura

   論文題目：A Re-Identification Strategy Using Machine Learning that Exploits Better Side Data

   印刷公表の方法および時期：IEEE International Conference on Awareness Science and Technology (iCAST), pp.1-8, 2019

   （3 章，6 章に関連）

## 関連論文の印刷公表の方法および時期

- 表彰

1. 15th IFIP Conference on e-Business, e-Services and e-Society, Best paper award,
   “Breaking Anonymity of Social Network Accounts by Using Coordinated and Extensible Classifiers based on Machine Learning”, Eina Hashimoto, Masatsugu Ichino, Tetsuji Kuboyama, Isao Echizen,Hiroshi Yoshiura Sep, 2016

# その他の研究業績

- 特許

1. 特願 2019-037079
   L1 CSI_RSRP の取得に関する端末動作
   橋本英奈, 武田和晃, 川名昭博, 閔天楊
   2019.9.20（出願日）
   国際公開日 2021/3/25
   国際公開番号 WO2021/053825

- 対外発表

1. Telco AI Summit Europe 2020
   Eina Hashimoto
   2020 年 11 月

- 表彰

1. 2016 年度 電気通信大学 学生表彰