一般
投稿論文

SNSプライバシー保護とリスク管理の検討
―ソーシャルモニタリングツールの開発に向けて―

山下 晃弘[†1]　上村 卓史　川村 秀憲[†2]　鈴木 恵二[†2]

[†1]東京工業高等専門学校　[†2]北海道大学

家族や友人のみならず，就職活動やビジネスにおいてもSNSの利用が一般化している．一方で，炎上やネットストーカーに代表されるプライバシーリスクも顕在化してきた．スマホのアプリやサービスの選択肢が増える一方で，それらの情報源から第三者が個人情報を容易に推測可能な環境が整いつつある．特定のSNS上では安全でも，それらが結びつくことで生じるリスクは目視や手作業での把握や管理が難しい．筆者らは，多様化するSNSにおいて，利用者がプライバシーリスクを正確に把握し，適切な管理下で安全に利用可能な環境作りについて検討してきた．また独自にソーシャルモニタリングツールを開発して運用し，SNSに内在するさまざまなリスクについて知見を得てきた．本稿ではその取り組みを紹介し，今後さらに多様化が予想されるSNS利用形態とプライバシーリスク管理の課題について検討する．

1．はじめに

平成25年度情報通信白書[1]によれば，スマートフォン普及率は38.2%，SNS（Social Networking Service）利用率はFacebook 40.4%，Twitter 33.8%，Google+ 10.6%であった．「これまでにソーシャルメディアを利用したことはない」との回答は4.1%で，インターネット利用者の多くは何らかのSNSとかかわりを持っている．

今やSNSは家族や友人間だけではなく，就職活動やビジネスにも利用され，個人的な連絡はメールよりもSNSを利用するというユーザも少なくない．10代20代のディジタルネイティブ世代は一層抵抗なくWeb上でコミュニケーションを取り，SNSと無縁の利用者は若い世代ほど少数派といえるだろう．

一方で，Webの露出性や伝搬力，悪意を持つユーザへの認識や対処の甘さから，さまざまなリスクも顕在化してきた．「炎上」はその典型例である．いたずら心による投稿が批判の標的となり，実名，性別，年齢，住所，顔写真，家族や友人関係，所属企業や学校名等が瞬く間に調べ上げられてさらされる事案は，枚挙にいとまがない．炎上のように顕在化する事例だけではなく，ネットストーカーやなりすましによる嫌がらせ等の潜在的リスクも指摘され，無防備なSNS利用が原因で，気付いたときには手遅れといった事態が社会問題化している．

いったんネット上にさらされた情報を完全に抹消することは事実上不可能である．標的とされた個人は，進学，就職，結婚といった社会生活の根底を脅かされかねない．このようなリスクは，Webの匿名性やSNSの秘匿性の過信，スマホの普及で実生活とWebが密接に結びつくことも要因の1つと考えられる．

そこで筆者らは，SNS利用上のリスクを体系的に捉え，過去に顕在化した事例の詳細な分析と，利用者自身が適切に情報管理を行い，安心してSNSを利用するためのサポートツールの開発に取り組んできた．このようなサービス設計を行う際，単に現状の課題を研究要素と捉えて分析するだけではなく，多様に変化するSNSの利用形態をリアルタイムに捉え，利用者の利便性のみならず社会的課題を先回りして把握し，適切な対処法を検討するプラットフォームが必要である．我々はSNS利用をサポートする独自のシステムを，大学発ベンチャー企業である（株）調和技研と共同開発し，恒常的なサービスとして運用してきた．

本論文では，個人情報を含め，プライバシー侵害の標的となりやすい情報をプライバシー情報と定義する．また，プライバシー情報の意図しない第三者への流出を防ぐことをプライバシー保護と呼ぶ．その上で，SNS利用におけるプライバシー保護とリスク管理について，筆者らの取り組みから得られた知見や課題を述べる．また，筆者らが開発するソーシャルモニタリングツールについて報告する．

2. SNS の利用とプライバシー

2.1 SNS の利用実態と先行研究

Schrammel らは，ニューヨーク州ペース大学の 1 年生 200 名に対し，SNS のプライバシーに対するアンケート調査結果を報告している [2]．平均年齢は 18.6 歳であり，利用している SNS の内訳は Facebook（96.5%），MySpace（27%），Twitter（24.5%）である．報告によれば，プロフィールの公開設定について，非公開が 74.5%，公開が 14.4%，不明が 10.7% であったとしており，プライバシー公開設定を正確に把握していないユーザが 1 割以上存在している実態が明らかとなった．また，プライバシーポリシーや運営側での個人情報収集についても把握している学生は少数であると報告している．

一方で Johann らは，Facebook 上の個人情報の公開範囲について調査し，約 77% のアカウントで e-mail やインスタントメッセージを公開している実態を挙げ，ネットストーカー被害のリスクを指摘している [3]．また，人の性格特性を表現する 5 因子モデルを用いたオンラインコミュニティ上での情報開示パターンとの関連性については，強い相関は見られないとしている．つまり，実際の性格とオンライン上での情報開示の傾向の相関は低いとしている．

これらは欧米での調査であり，日本とは若干事情が異なるが，SNS の設定を初期設定のまま意識しないユーザは多く，また SNS リスクの意識の薄さと把握の困難さは日本でも同様に懸念される．

国内の先行研究として，鈴木ら [4] は，SNS 等 Web 上の公開情報に推論を適用し，ハイパーグラフモデルによって秘密情報抽出の可能性を議論している．課題意識は本研究と共通する部分であり，推論から秘密情報を取得する方法は興味深い．また，安藤ら [5] は，本研究と同様のモチベーションで複数の SNS 上のデータから個人特定の可能性について議論している．

SNS の個人属性に着目した研究として，Twitter のつぶやき履歴から所在地を推定する研究 [6],[7] や，年齢，性別等のグラフィック属性を推定する研究 [8],[9] 等が代表的である．またインターネット上の匿名性に言及した特集記事 [10] では，実名を使わないことと匿名性，仮名性に言及し，たとえ実名ではない ID やニックネームであっても，ある行動が同一人物によるものと推測できる（リンク可能性）場合は匿名性が失われる可能性があることに触れている．SNS 利用者が匿名と勘違いした状況において，意図しない領域で個人が同定され情報がリンクすれば，炎上やネットストーカーへのリスクは拡大する．

我々は，このような SNS 利用の実態とリスクの存在に対し，個人が適切に自身のプライバシーリスクを把握し，適切な管理と利用をサポートする環境作りについて検討を行ってきた．本論文ではその取り組みについて紹介する．

SNS リスクの状態を把握し
利用者自身が自らの発信や
設定を管理可能にする

2.2 SNS のデータ構造

まず，主要な SNS のデータ構造について整理しておく．日本国内で利用率の高い SNS のうち，特に上位である Twitter，Facebook，Google+，Instagram に絞り，そのデータ構造や API の調査を実施した．

筆者らは各 SNS のクローラを開発し，各 SNS の枠を越えて情報が結びつく可能性の定量的評価を実施している．各 SNS のデータ構造を調査し，蓄積されている情報を分析することで，まずはリスクの評価に必要な前提条件を整理した．なお，ここでは，プライバシー情報となりやすいプロフィール情報に絞って調査を行った．今後は，友人関係やユーザによる投稿情報についても議論する必要がある．

表 1 は各 SNS で蓄積されるプロフィールデータである．目的や利用シーンが異なるため，蓄積内容に差異があることが分かる．しかし，デモグラフィックな属性のほかに，自己紹介は 4 つの SNS すべてで項目が存在していた．また所属や場所の情報も人物特定の要因となる可能性が高い．

また，各 SNS で設定されるアカウント名も個人を特徴づける文字列である場合が多い．アカウント名は SNS 内で一意であるため，特徴的な文字列を設定する必要があり，同一または類似する文字列を複数の SNS で利用することが想定される．各 SNS のデータ構造には差異があるため，リンクした際に個人情報が露呈する可能性は高い．

複数の SNS 上の情報を注意深く観察すれば，ある SNS アカウントと別の SNS のアカウントが同一人物かを判断ができる可能性がある．本研究では第三者による情報の結びつきの可能性を検討し，潜在的なリスクとしてユーザに提示して適切な設定や管理を促す方法について検討している．

表 1　各 SNS に登録されたプロフィール情報のデータ構造

	Twitter	Facebook	Google+	Instagram
固有 ID	○	○	○	○
ユニークネーム	screen_name	username（URL）	―	username
氏名	名前	フルネーム， 姓，名，ミドルネーム	氏名（姓，名，敬称等）， ニックネーム， 表示用氏名	氏名
連絡先情報	―	メールアドレス	メールアドレスリスト	―
生年月日	―	生年月日 年齢域	生年月日 年齢域	―
性別	―	性別	性別	―
場所	場所	住んでいる場所 出身	現在の場所 住んだ場所のリスト	―
自己紹介	自己紹介	自己紹介	自己紹介 ユーザの「自慢」	自己紹介
所属・履歴	―	学歴 職歴	関係する組織リスト 略歴	―
プロフィール画像	プロフィール写真 カバー写真	プロフィール写真 カバー写真	プロフィール写真 カバー写真	プロフィール写真
URL	URL	URL	プロフィール URL URL リスト	URL
言語	言語設定	ロケール	言語	―
タイムゾーン	タイムゾーン		―	―
貨幣	―	貨幣	―	―
その他	アカウント作成 日時	政治観，宗教・信仰，好きなスポーツ選手，好きなスポーツチーム，尊敬する人物，アルバム，本，イベント，ゲーム，グループ，興味，映画，音楽，タグ付けされた写真，テレビ番組，ビデオ，家族	パートナー関係 （独身，既婚，交際中，など）	
つながり	非対称なフォロー・被フォロー関係	対称性のある友人関係	「タグ付け」による非対称な関係	非対称なフォロー・被フォロー関係

3. 炎上とプライバシー

SNSのリスクが顕在化する典型例が炎上である．一般的には，SNS上での不適切な発言が発端となり，その発言者の個人情報が何らかの方法で第三者によって暴かれ，Web上にさらされていく．そのプロセスには複数のユーザが関与している場合が多く，詳細に分析することで，SNS上のプライバシー情報の発見経路や，顕在化の知見を得ることができる．

我々は，SNSから発生した60件の炎上事例を詳細に調査し，その内容や炎上の広がりについて分析を行った[11]．**表2**は事例のうちいくつかをまとめたものである．炎上の発生件数は，2011年頃から発生し，一度2012年ごろにピークを迎えて減少した後，2013年度にはさらに多くの炎上事例が発生している．また，炎上の内容を調査した結果，①他人への誹謗中傷にあたる投稿（2件）②犯罪行為を暴露する投稿（交通違反，窃盗，器物損壊，未成年の飲酒・喫煙など34件），③職務の問題行為に関する投稿（業務に関するいたずら暴露，客に対する誹謗中傷，守秘義務違反など16件），④その他問題行為関する投稿（カンニング，他人のプライバシー侵害など7件），の4つのタイプに分類された．内訳としては，特に犯罪行為を暴露する投稿と，職場の問題行為に関する投稿が多かった．

一方で，炎上原因となる発言の投稿を発端としてそれが拡散し炎上に至るまでのプロセスを分析し，それぞれ時系列に，原因となる発言前の「正常な状態」，問題発言の「発見フェーズ」，複数の閲覧者による「拡散フェーズ」，「本人特定フェーズ」と定義・分類した．それぞれのフェーズごとに，それが炎上に至るのを防ぐための防止策についても検討を行い，対応づけた．**図1**は結果のまとめである．

本研究は，このような炎上事例の分析結果に基づいて，第三者がどの程度のプライバシー情報を取得可能であるかをあらかじめ定量的に評価し，利用者の適切な対策や投稿の制限などの管理をサポートするシステムの開発が最終的なゴールである．

4. SNS 上でのプライバシー管理

本研究で目指すシステムは，大きく2通りの利用局面を想定する．1つは，個人が自分自身のSNS利用を適切に管理するための機能である．複数のSNSを使い分けている場合や，長年SNSを利用している場合，全体として

表 2　近年の炎上事例のまとめ

発生時期	媒体	業種	概要
2011 年 1 月	Twitter	ホテル	本人が勤務するホテルのレストランに，有名サッカー選手と芸能人が来店したことをツイート．本人は過去にも同様のツイートを繰り返していた．ホテルの支配人が謝罪し，Web ページ上にお詫びを掲載.
2011 年 5 月	Twitter	スポーツ用品店	来店したスポーツ選手と同伴の女性を店員が Twitter で中傷し，炎上．店員は退職し，スポーツ用品店は選手と所属クラブに謝罪
2011 年 8 月	Twitter	ホテル	人気アイドルグループのメンバが勤務先のホテルに宿泊したとして, 部屋の写真などをインターネット上に流出.「やばいやばい www うちのホテルに□□□□くん泊まったんだか www これから泊まった部屋行ってくる www」
2011 年 9 月	Twitter	製薬業	製薬会社の社員が睡眠薬を不正に入手し, 飲み会で酒に混入したと発言．企業が謝罪のプレスリリース発表
2011 年 12 月	Twitter	大学生	某国立大学の学生が「iPod ゲット．忘れ物からパクった（笑)」と発言し，炎上
2011 年 12 月	Twitter	大学生	「無免で事故ったぁ？もはや開き直るぅ？だがしかしママが居て良かったぁ？けどママごめん (;__;) JAF 待ちー？」とツイートし炎上
2012 年 1 月	Twitter	大学生	「帰り俺が運転したけど相当危険だった？急ブレーキごめん！ やっぱ無免は運転しちゃだめだね (^_^;) 」とツイートし炎上
2012 年 2 月	Twitter	飲食業	「頭痛と吐き気がやばくて動けないんだけど疑われて無理やりシフト入れと言われたので肉鍋にゲロ吐いてきます．」とツイートし炎上
2012 年 3 月	Twitter	大学生	「誕生日祝ったり飲酒運転でちんさむしたり□□□家宅のみしに行って送ってもらってさっき帰宅ー」と飲酒運転＆未成年飲酒告白し炎上，アカウント閉鎖
2012 年 5 月	Twitter	病院	「今日すごい事実発覚！□□□□□□の選手のカルテ発見！住所も電話番号もわかるー … ！」とツイートし炎上
2012 年 6 月	Twitter	大学生	「□□大学に爆弾を仕掛けました．明日□□行ったら死にます．明日わ余裕で休校です.」と不適切な書き込みを行い無期停学処分に.
2012 年 12 月	Twitter	フリーター	「□□□□□□□で化粧品を万引きするミッション成功 !! 」とツイートし炎上.
2013 年 1 月	Twitter	専門学校生	「あ，そゆいえば今日暇だたけ□□□□のカルテ見てみた♪♪♪住所とか電話番号とか」と，医療専門学校生が日本代表 FW □□□□のカルテを研修先の病院で見たことを暴露．専門学校は学生を処分し，Web ページにお詫びを掲載

どのような情報が公開され，第三者が取得可能な状況にあるのかを把握することは困難である．また，軽い気持ちで危険な投稿をしてしまう可能性も否定できない．そこで必要と考えられる機能は，個人の公開設定の一覧化，プライバシー情報に結びつく可能性のある特徴語の表示，過去に投稿された NG ワード表示などである．

もう 1 つの利用局面はアルバイト従業員や会員などを多数抱える組織の管理者が，全体の利用状況を把握し，リスクを抑えると同時に，炎上などの事態にはいち早い対応が可能な仕組みを提供する機能である．具体的な機能としては，個別および全体の投稿量や公開設定の可視化，NG ワードの監視，危険な投稿に対する利用者への通知機能などである．この両方の利用局面において必要となる要素技術は重なる部分が多いが，機能的な観点から見ると, 普段の状況を監視するための「状態の可視化」, 炎上などのきっかけを防ぐための「不適切発言の防止」, 万が一炎上等の予兆が検知された場合の「拡散防止」の 3 つに分類できる．

前述のとおり，SNS 上で保持される個人情報はサービスごとに異なる．炎上事例の分析においても SNS 等の複数のネット上の情報源がリンクした際にプライバシーが露呈し，本人が意図しない情報がさらされる結果となる場合が多かった．第三者によって複数の SNS 上の情報が紐づけられる要因はいくつか考えられるが，筆者らの調査では，多くの場合，各 SNS 上に存在するアカウント名などの断片的な情報源から，同一人物と推測されることが原因であった．一方で，同姓同名や類似したアカウント名を持つまったく関係ないユーザの情報が同一人物として紐づけられ，誤った情報の拡散や炎上の可能性も十分に考えられる．そのような可能性は考慮しつつも，断片的な情報源が紐づけられる可能性を事前に評価し，リスク要因として認識することは重要である．

そこで本研究では，複数の SNS を使い分けているユーザについて，各 SNS 間の情報の関連性を調査し，第三者がそれらを結びつけ，利用者本人が意図しない個人情報が明らかになる可能性を検討した．本論文では特に利用者の多い Twitter（以下 TW）と Facebook（以下 FB）について，主として①TW 上で FB アカウントを明示しているユーザにおける両アカウント名の類似性，②TW 上で FB アカウントを明示していないユーザにおける FB アカウント発見の可能性，に関する実験を行った．次節ではその結果について報告する．実験にはあらかじめ開発したクローラによって収集した TW アカウントデータ 55,226 件を用いた．

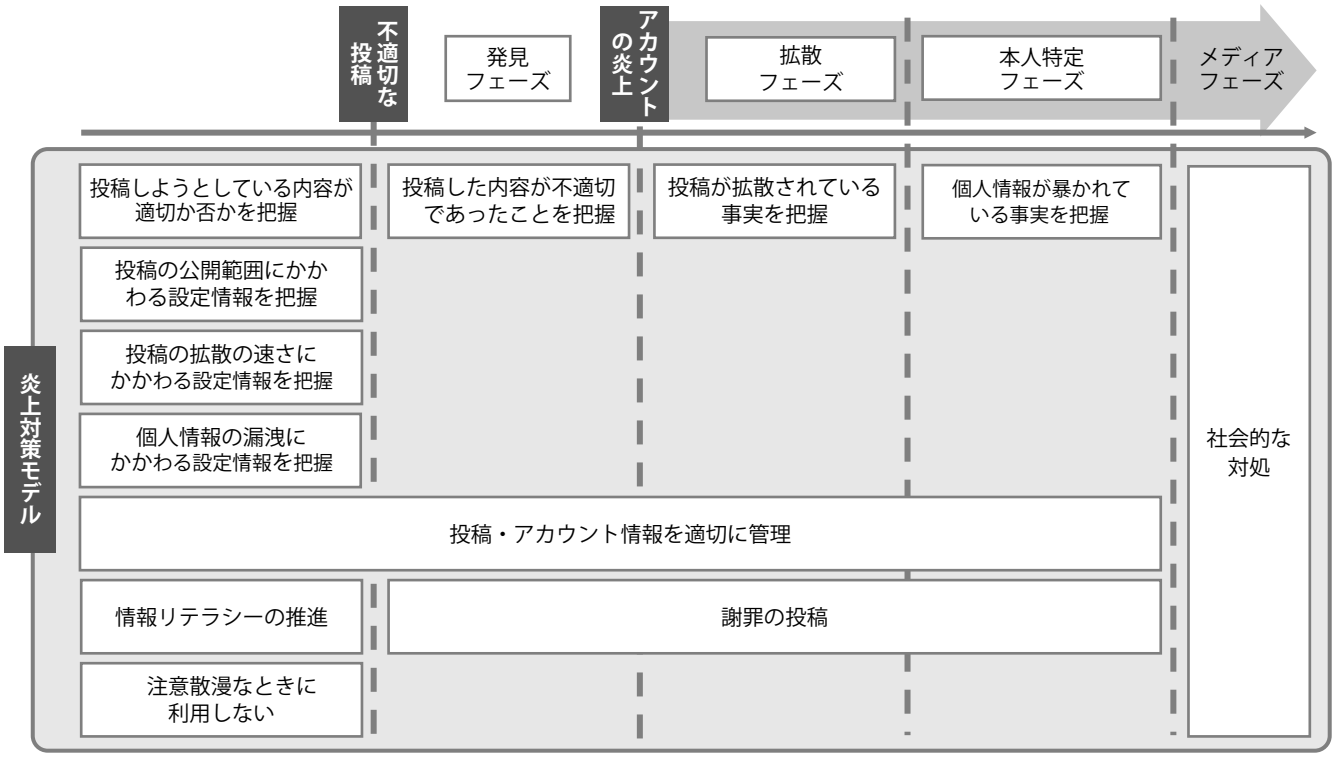

図1　炎上のプロセスとその対策モデル

4.1　TWアカウントからFBアカウント収集

入力となるTWアカウントに対応するFBアカウントを収集する際，(1) TWアカウント上で自分のFBアカウントを公開しているユーザ（FB明示），(2) TWアカウント名と同じFBアカウント名を使用しているユーザ（TWとFBで一致），(3) TWアカウントに対し，“_”を除去または“.”に置き換えることでFBアカウントと一致したユーザ（TWに変更を加えてFBと一致），というように場合分けし，TWアカウントに対しFBアカウント（ただしTWユーザ本人のものとは限らない）を収集した．ただし，FBページが見つかったものは除外した．この結果，全アカウント数55,226件中，それぞれ5,237，6,425，8,643アカウントが該当した．

4.2　TW上でFBアカウントを公開しているユーザにおける両アカウント名の類似性

TWアカウントには140文字以内の自己紹介文（description）があり，人となりや趣味等さまざまな項目が記載されている．自発的に自身のFBアカウントのURLを書くユーザがいる一方で，まったく別として運用しているにもかかわらずアカウント名や自己紹介文，投稿内容からFBページを特定され，予期せぬ情報の漏洩に発展してしまう場合も考えられる．

本節では，4.1節で分類したうちのFB明示（TWにおいてFBのURLを明示しているユーザ）について，両アカウント名の類似性を考察した．FBアカウントついては，http://www.facebook.com/xxxのようなURLを自己紹介文またはWebページURLから抽出した．アカウント名の類似性を評価するために，本研究では編集距離（レーベンシュタイン距離）$\mathrm{D}(c_1,c_2)$を導入する．編集距離$\mathrm{D}(c_1,c_2)$は，文字の挿入や削除，置換によって，文字列c_1を文字列c_2に変形するのに必要な手順の最小回数として与えられる．ただし，c_1とc_2の文字列の長さが異なる場合，単純な編集距離では直感と異なるため，本研究では，正規化編集距離

$$\mathrm{D}_{normalize}(c_1,c_2)=\frac{\mathrm{D}(c_1,c_2)}{\max(length(c_1),length(c_2))}$$

を用いた．

図2は，得られたFBアカウント名と，TWユーザ名およびスクリーンネーム（@で始まる文字列）を，正規化編集距離で比較した結果である．

評価は小文字に統一して計算を行った．また，TWでは区切りに“_”が用いられるが，FBでは“.”であるため，区切り文字を除去して計算した結果も合わせて掲載する．この結果から，区切り文字を残した場合でも全体の18.2%，除去した場合には55.4%が同一とみなせるアカウント名を利用しており，約半数のユーザはほぼ同一文字列をTWとFBのアカウント名に利用していることが明らかとなった．

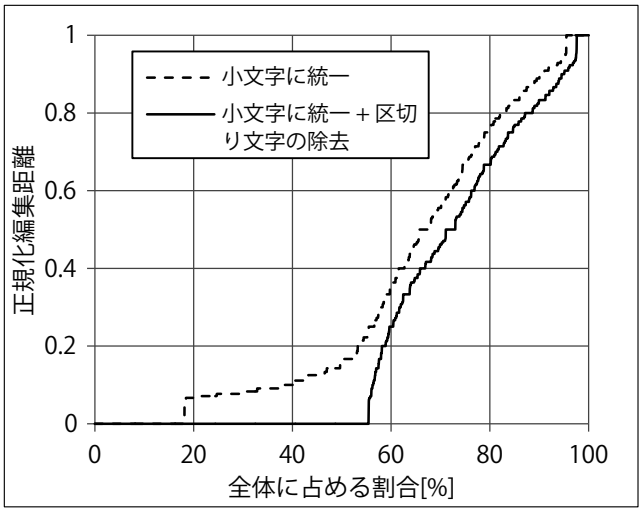

図 2　FB の利用を明示している TW ユーザにおけるアカウント名の類似性

この実験で対象としたユーザのように，自ら別の SNS のアカウント名を公開している場合は問題ではないが，FB は実名で利用し，TW ではハンドルネームで利用するといった使い分けをしているにもかかわらず，不用意に同一のアカウント名を利用した場合，個人情報漏洩のリスクが懸念される．

4.3　FB アカウント公開・発見状況に対する TW 利用状況

4.2 節の結果を踏まえて，TW では実名を公開しないものの，FB と同一または非常に類似したアカウント名を使用している場合のリスク調査を実施した．収集した TW アカウントに対し，(1) TW アカウント上で自分の FB アカウントを公開しているユーザ（FB 明示），(2) TW アカウント名と同じ FB 名を使用しているユーザ（TW と FB で一致），(3) TW アカウントに対し，“_” を除去または “.” に置き換えることで FB アカウントと一致するユーザ（TW に変更を加えて FB と一致），と場合分けを行い，TW のフォロワー（そのユーザの投稿を購読しているユーザ）数，フレンド（そのユーザがフォローしているユーザ）数，投稿数を比較した結果を図3，図4，図5に示す．

これらの図から，FB アカウントを明示しているユーザは，フォロワー数，フレンド数ともに多い傾向があることが分かる．一方で投稿数に関しては，FB アカウントが発見できなかった TW アカウントは投稿数が少ない傾向にあるが，それ以外の差異は見られなかった．FB を利用しているユーザの場合，TW と FB で似たような友人関係が存在すれば，その中の誰か一人の個人が特定されることで，連鎖的に自らのアカウントまで特定されてしまうリスクも考えられる．このようなユーザ同士の関係によるリスクについての考察は今後の課題である．

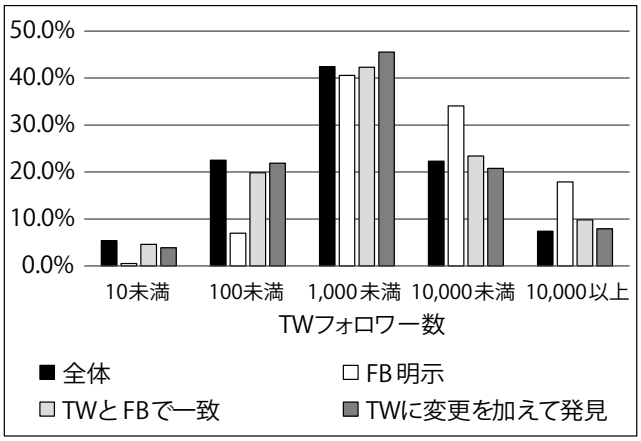

図 3　TW フォロワー数の差異

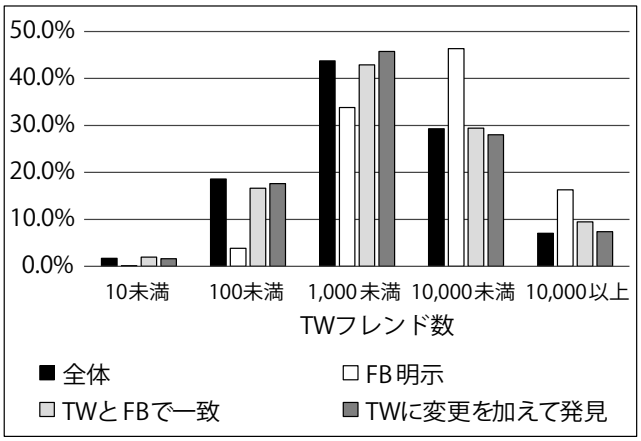

図 4　TW フレンド数の差異

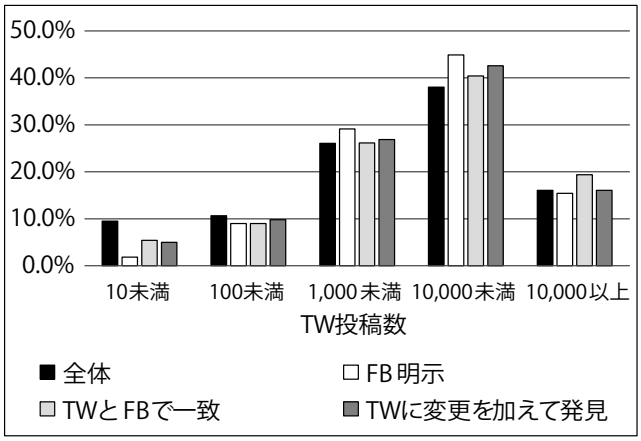

図 5　TW 投稿数の差異

4.4　TW からの FB アカウントの発見

前節までの検証において，TW と FB が一致するアカウントが確認できた際に，それが本当に同一人物であるかは評価していない．本節では，この問題について定性的な考察を加える．この実験では 4.1 節で示した 3 つの場合分けそれぞれから無作為に 100 件のアカウントを抽

出し，TWおよびFBの内容を目視による閲覧で判定を行った．このとき，

（ステップ1）アカウント名および氏名，TWの自己紹介，TW,とFBそれぞれのプロフィール画像，URL，基本データ等のプロフィール情報から同一人物または同一人物ではないと判定を試みる

（ステップ2）ステップ1では判定できず，TWとFBのタイムライン（本人の投稿等）を閲覧して同一人物と判定，または同一人物ではないと判定を試みる

（ステップ3）ステップ1, 2では判定できず，TWのフォロー・フォロワー，FBの友達関係を閲覧することで同一人物または同一人物ではないと判定を試みる

（ステップ1），（ステップ2），（ステップ3）いずれの方法でも判定不可

という場合分けを行った．**表3**はその結果である．

「TWとFBで一致」で22人，「TWに変更を加えてFBと一致」で17人は，同一人物と思われるFBアカウントを発見した．「TWに変更を加えてFBと一致」を加えると多少発見率は下がるが，簡単な操作でアカウントが同一人物候補として発見されてしまうことはリスクとなり得る．

また，本人のFBアカウントが発見され，かつTWは匿名性を確保して活動しているユーザのうち，「TWとFBで一致」から8人，「TWに変更を加えて発見」から3人は実名等の情報をFBから得ることができた．さらに，この中の4人はTWのアカウント設定でプロテクトをかけており，TWからは個人情報を公開しないように活動しているにもかかわらず，簡単な操作によりFB上での実名の情報が得られてしまうという，きわめてリスクの高い状態にあると判断できる．

5. ソーシャルモニタリングツール「Cikappo」の開発

「Cikappo」は，本研究の成果を踏まえ，筆者らが中心となり（株）調和技研と共同で開発したソーシャルモニタリングツールである[12]．本ツールの開発にあたり，第2章で述べたSNSのデータ構造の差異の分析や，第3章で述べた炎上事例の研究，および多数の企業ヒアリングを踏まえ，**表4**に示す機能の実装を行った．表4の各機能を実装するにあたっては，情報のリアルタイム性を重視し，不適切投稿の監視や情報の拡散監視については5分間隔でのチェックを実現する仕組みを構築した．また，第4章で述べた複数のSNSの断片的情報を結び付けて横断的にリスク評価する機能については現在実装を進めている．

現在は，主に炎上事例の舞台となることが多い，Twitterと2ちゃんねるの書き込みデータをバックグラウンドで常時監視し，ユーザが設定したNGワードや個人情報が発見された場合には，直ちにアラートを送信する機能実装している．また，単純な辞書マッチングでのNGワード検出だけではなく，学習アルゴリズムによる文脈解析によって危険と判断した書き込みにはアラートを送信する機能も実装している．炎上などのリスク要素が発見された場合は，対処までの時間を短縮することが非常に重要である．本システムでは，検出のリアルタイム性向上のために，SNSデータクローラと危険検出エンジンはAmazon EC2クラウドを用いて分散並列実装している．また，現在のSNS設定が適切に保たれているかどうかを診断する機能を実装しており，利用中にシステムが自動的にソーシャルリスクの状態を判断して利用者に知らせるようなサービスを提供する．

実装のプラットフォームとしては，Webアプリケーション版とiPhoneアプリ版の2種類を開発し，2013年より運用している．Webアプリケーション版は，主に大量のアルバイトや社員を抱える企業や団体を想定し，ソーシャルリスクからの社員の保護と社内管理を目的として開発したツールである．企業によって，社員の個人情報だけではなく，内部情報が外部に流出することを防ぐため

表3　TWアカウントからのFBアカウントの発見

「TWとFBで一致」	ステップ1	ステップ2	ステップ3
本人であると判定	17	5	0
本人でないと判定	60	5	0
不明と判定	23	13	13

「TWに変更を加えてFBと一致」	ステップ1	ステップ2	ステップ3
本人であると判定	16	1	0
本人でないと判定	72	2	1
不明と判定	12	9	8

表4　ソーシャルモニタリングツールの機能一覧

No	機能内容	カテゴリ
1	投稿数等の統計情報可視化	状態可視化
2	リツイートなどの一覧表示	
3	プロフィールや投稿内容が閲覧可能なアカウント数の提示	
4	投稿内容のNGワードチェック	不適切発言防止
5	機械学習による投稿内容の文脈解析	
6	不適切な内容投稿時のアラート	
7	投稿の拡散監視	拡散防止
8	拡散の異常検知とアラート	
9	2ちゃんねるへの転載検知	その他

に，一括管理でのNGワード登録機能と管理者による状況把握の機能を実装している．一方でiPhoneアプリは個人利用を想定し，SNSの利用状態の健全性や，個人情報が意図しない形で流出していないかをモニタリングするツールである．図6はiPhoneアプリの利用画面で，図7がWebアプリケーションの利用画面である．

本ツール構築に際し，多数のアルバイトを抱える企業や，顧客サービスを提供する企業等，多数の企業にヒアリングを実施し，プロトタイプの使用に関する評価を実施中である．現時点で得られた定性的な評価によれば，不適切な投稿に対するフィルタや，拡散状況の可視化に対する社会的なニーズは強く，特に炎上に至らない間での対処を行うためにもリアルタイム性は重要である．リアルタイム性を確保しつつ，使い勝手の良いシステムへの改善が今後の課題の1つである．一方で，本システムは2013年度より日本プロサッカーリーグ（Jリーグ）の選手にも実際に利用していただいている．Jリーグでは，SNSによるサポータとの積極的な交流を推進しており，不適切発言の防止は重要課題である．また，サッカーくじ等の関係から，スタメン情報や試合会場への移動が想定される発言も問題発言として検知したいというニーズをいただき，専用のカスタマイズを加えている．

図6　ソーシャルモニタリングツール「Cikappo」iPhone アプリの利用画面

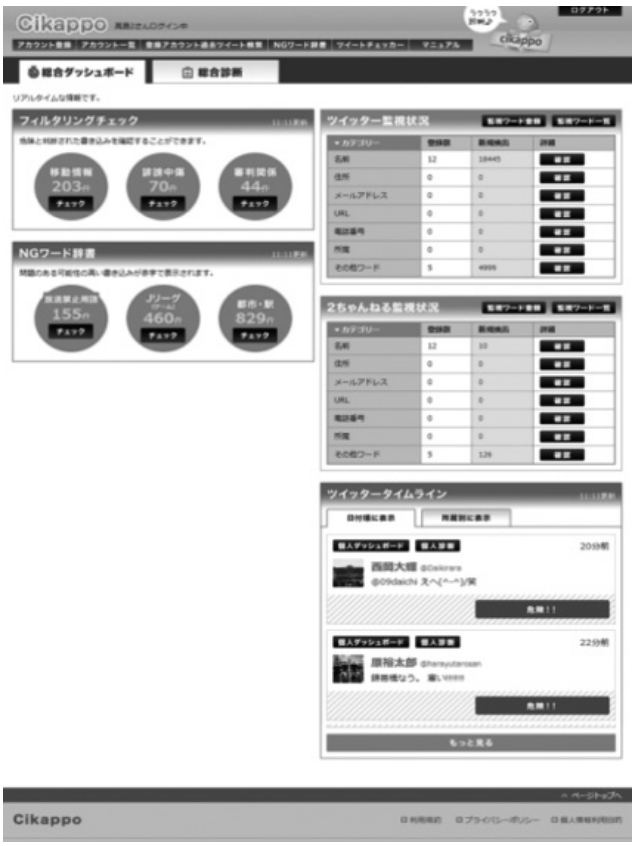

図7　ソーシャルモニタリングツール「Cikappo」Web アプリケーションの利用画面

図8はJリーグで本システムに登録されている全526名のアカウントを2シーズン利用した統計データであり，総ツイート数とワードフィルタでの検出数を示している．

途中でフィルタのカスタマイズ等多少の調整は行っているものの，全体としては総ツイート数に対してワードフィルタでの検出割合は減少傾向にあることが分かる．また，Jリーグ担当者からの意見として，次のコメントをいただいた．

- 炎上は頻繁には発生しないものの，選手によるツイートには危険なワードも含まれることがある．
- リテラシー強化のリアルなツールとして有用．このデータに基づいた勉強会も開催した．
- NGワードフィルタでは危険ではないワードも検出されるため運用場面では調整が必要．
- 炎上とは直接関係ないが，アカウントの「乗っ取り」を事前に察知できたことがあった．

このように，運用期間中において炎上は発生していな

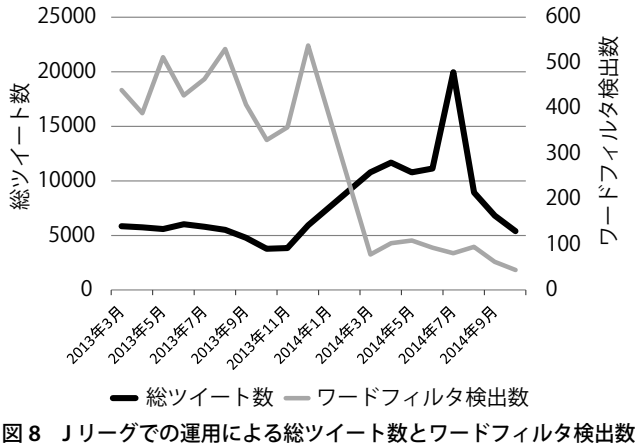

図8　Jリーグでの運用による総ツイート数とワードフィルタ検出数

いものの，リテラシシー強化のツールとして有用性があり，また有名人などに多い「アカウント乗っ取り」が本ツールによって早期に検出できたことは重要であると考えている．このような事例やヒアリング結果に基づいて今後より多方面で運用したいと考えている．

6. 終わりに

SNSは気軽な気分で友人知人とのコミュニケーションを楽しむツールであるがゆえ，つい軽い気持ちで個人的な発言をしがちである．また，特定の閉じたコミュニティに発信していると勘違いし，自己顕示欲や誇張によって，いたずら心で投稿した情報が炎上し，思わぬ結果を招く事態も少なくない．またそのように顕在化する事案の裏で，潜在的にはSNSを利用する大部分のユーザが，ネットストーカーなどの第三者によって想定を超えた情報漏洩を可能としているリスクを抱えていることが想像される．

スマートフォンの普及により，誰もが気軽に情報発信できる環境を提供した一方，アプリやサービスが乱立し，本人が意図せず位置や友人関係を露出してしまうリスクが拡大している．写真の投稿や買い物情報などはリアルな社会生活と密接にリンクし，それらが総合された場合，さまざまな個人情報を第三者が推測できる可能性を高めることは間違いない．

筆者らは，拡大するSNS利用の中で，本人が正しくその利用の実態を把握し，今現在の利用状況がどのような情報漏洩のリスクを抱えているのかを可視化する技術開発に取り組み，一部サービス運用を行ってきた．一方で，さらに利便性を高めて広範囲な管理を統合的に実現するためには，莫大な情報源からのデータクローニング技術や，リアルタイムなビッグデータ処理技術など，解決しなければならない課題も多い．筆者らの取り組みは，まずはそれらの基礎となるリスク管理モデルを確立し，総合的なサービス提供に向けた第一歩である．新しく魅力的な機能やサービスが日々生み出される一方で，多様化するSNSサービスを個人の立場から俯瞰し，安心安全を担保して適切なコミュニケーション環境を保守する仕組みを構築することも，我々IT研究者の重要な役目であると考えている．

謝辞　本研究の一部は，2013度総務省戦略的情報通信研究開発推進事業（SCOPE）（132101001，代表：山下晃弘）の支援を受けて実施した．

参考文献

1) 平成25年版情報通信白書，http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h25/pdf/25honpen.pdf
2) Lawler, J. P. et al.: A Surver of First-year College Student Perceptions of Privay in Social Networking, CCSC Eastern Conference, pp.36-41 (2011).
3) Schrammel, J. et al.: Personality Traits, Usage Patterns and Information Disclosure in Online Communities, HCI 2009, pp. 69-170 (2009).
4) 鈴木　遼 他：推論による情報漏えい防止のためのハイパーグラフモデル，電子情報通信学会論文誌，情報システム J95-D84, pp.812-824 (2012).
5) 安藤寿英 他：複数SNSサイトにおける発信情報分析による個人特定の可能性の検証，第11回情報科学技術フォーラム (2011).
6) Cheng, Z. et al.: You Are Where You Tweet: A Content-Based Approach to Geo-Locating Twitter Users, In CIKM '10 (2010).
7) Eisenstein, J. et al.: A Latent Variable Model for Geographic Lexical Variation, In EMNLP 2010 (2010).
8) Rao, D. et al.: Classifying Latent User Attributes in Twitter, In Proc. of the 2nd International Workshop on Search and Mining Usergenerated Contents, pp.37-44 (2010).
9) Burger, J. D. et al.: Discriminating Gender on Twitter, In EMNLP2011, pp.1301-1309 (2011).
10) 折田明子：ソーシャルメディアと匿名性，人工知能学会誌，Vol.27, No.1, pp.59-66 (2012).
11) 山形聖志，川村秀憲，鈴木恵二：SNS炎上の分析に基づく炎上対策方法の研究，人工知能学会研究会資料，SIG-KBS-B303, pp.13-18 (2014).
12) ソーシャルモニタリングツール，http://www.chowagiken.co.jp/wp-content/uploads/2013/11/leaflet_ver01.00.pdf

山下 晃弘（正会員）yamashita@tokyo-ct.ac.jp
2010年北海道大学大学院情報科学研究科博士後期課程修了．2011年（株）調和技研代表取締役を経て，2013年より東京工業高等専門学校情報工学科助教，現在に至る．知能システム，データマイニングの研究に従事．

上村 卓史（非会員）uemura.takashi@gmail.com
2011年北海道大学大学院情報科学研究科博士後期課程修了．同年ヤフー（株）入社．2012年（株）調和技研入社．2014年アクセンチュア（株）入社，現在に至る．

川村 秀憲（正会員）kawamura@complex.ist.hokudai.ac.jp
2000年北海道大学大学院工学研究科博士後期課程修了．同助手を経て，2006年同大学准教授，現在に至る．2007～2008年ミシガン大学客員研究員．マルチエージェントシステム，人工知能の研究に従事．

鈴木 恵二（正会員）suzuki@complex.ist.hokudai.ac.jp
1993年北海道大学大学院工学研究科博士後期課程修了．同大学助手，助教授を経て，2000年公立はこだて未来大学助教授．2004年同大教授．2008年北海道大学大学院情報科学研究科教授，現在に至る．マルチエージェントシステム，複雑系工学の研究に従事．

投稿受付：2014年5月7日
採録決定：2015年2月6日
編集担当：並木美太郎（東京農工大学）

本論文は特集「プライバシーフレンドリーシステム」への投稿論文です．