様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ９ 月 １６ 日現在

機関番号：１２６１２
研究種目：基盤研究(C)（一般）
研究期間：2011～2014
課題番号：２３５０００８２

研究課題名（和文）プライバシー保護のための個人情報の検知および照合技術の研究

研究課題名（英文）Development of technologies of detecting and linking personal information for privacy enhancement

研究代表者

吉浦　裕（Hiroshi, Yoshiura）

電気通信大学・情報理工学（系）研究科・教授

研究者番号：４０３６１８２８

交付決定額（研究期間全体）：（直接経費）　　3,900,000円

研究成果の概要（和文）：ソーシャルメディアを通じた個人情報の流出が問題になっている。そこで、メディアに投稿しようとする文章から個人情報の漏洩を検知する技術を開発し、11名の被験者の投稿文各1000件を用いた評価実験で、通勤・通学先及び職種情報の漏洩の約90%を検知することができた。一方、複数の個人情報の照合によるプライバシー侵害の問題が顕在化している。そこで、注目者の投稿文を本人の履歴書との照合により検知する技術を開発し、12名の被験者の投稿文各1000件と100人の背景ノイズ各1000件を用いた評価実験で、8名の被験者について、本人の投稿文と背景ノイズ100人の投稿文の中から、本人の投稿文を特定することができた。

研究成果の概要（英文）：Disclosure of personal information in social media is serious social problem. We developed a technology that detects revelation of personal information from natural language text to be posted in social medial. Evaluation using 1000x11 texts posted by 11 volunteers has shown that our developed technology detected 85% of revelations of volunteers' affiliations and job categories. On the other hand, privacy infringement by linking multiple fragments of personal information become a social concern.We developed a technology that detects texts in social media posted by a person of interest by linking the texts with a resume of the person. We have evaluated our developed technology using 1000x12 texts posted by 12 volunteers and 1000x100 texts posted by 100 background persons. The evaluation has shown that our developed technology worked successfully for 8 volunteers, i.e. it identified texts posted by the volunteer from texts posted this volunteer and 100 other persons.

研究分野：情報セキュリティ

キーワード：プライバシー　個人情報　ソーシャルネットワーク　情報セキュリティ

１．研究開始当初の背景

研究開始当初の 2011 頃には、個人情報のインターネットへの流出が社会問題として注目され始めていた。この個人情報の流出の原因はサービス業者の違法行為、ハッカーやウィルスの不正侵入など様々であったが、ソーシャルメディアからの流出が大きな問題になっていた。たとえば 2010 年の mixi ユーザ数は国内で 2000 万人、Twitter は世界で 1 億人であり、これらのユーザが自分自身や友人および周囲の人の個人情報を、日記やコメント、写真等のコンテンツの形で安易に開示していた。この不特定多数者の情報発信に伴うプライバシー問題は、自己表現やコミュニケーションといった人間の本質的欲求とメディア技術の発展に後押しされているため、今後益々顕在化していくと考えられた。

ソーシャルメディアからの個人情報の流出を防止する方法として、日記等の公開範囲の指定が実施されていたが、有効に機能していないことが社会科学者の分析により明らかになっていた。そのため、ソーシャルメディアからの個人情報の流出を防止する新たな技術的取り組みが求められていた。

一方、個人情報の流出経路は、単独のソーシャルメディアに留まらず、複数のメディア、業者の違法行為、不正侵入など様々である。特に、当時は、ユビキタス情報社会の実現が進んでおり、生活のあらゆる時間・空間で個人情報を提供するようになり、多数の断片的な個人情報の流出可能性が高まっていた。その結果、同一人物に関する複数の情報の照合、および、それを通じた個人の特定と広範囲の個人情報集積が、今後の高度情報社会の根本的なリスクとして顕在化してきていた。

特に、ソーシャルメディアは個人の特定につながる情報を大量に含むので、これと他の個人情報との照合は重大な脅威となりえた。たとえば、病院から流出した断片的な患者情報と、その患者が日頃開示しているソーシャルメディアコンテンツの照合によって、患者の氏名、住所、勤務先、病名などがセットとなってしまう可能性がある。この複数の個人情報の照合によるリスクに対応する技術的取り組みが求められていた。

この問題に対処するために、複数の情報が同一人物の情報であることを検知する Linking 技術が研究されていた。この技術により、複数の情報漏洩の組合せリスクを検証することが原理的には可能であった。しかし、当時の Linking 技術は、情報をベクトルやグラフの形式で表現した後、ベクトル間の距離計算やグラフの照合を行なっているので、ソーシャルメディアコンテンツのように単純に形式化できない情報は扱っていなかった。

２．研究の目的

上記の研究開発当初の背景を鑑み、以下の３つの研究目的を設定した。

（１）ソーシャルメディアを通じて開示しようとする投稿文から個人情報を自動検知する技術を確立する。これにより、ユーザへの警告、自動的な開示制御、さらにはプロバイダーのプライバシー保護への指針提供が可能になる。たとえば、コンテンツのどの部分からどのような個人情報が漏洩するかをユーザに示し警告することができると考えた。

（２）従来自動処理できなかった自然言語テキストの形式化と Linking を可能にする。これにより、ソーシャルメディアの投稿文同士あるいはソーシャルメディアの投稿文と他メディアのコンテンツの組合せリスクを健在化させ、明示することが可能になると考えた。

３．研究の方法

上記第１の目的（ソーシャルメディアを通じて開示しようとするコンテンツから個人情報を自動検知する技術の確立）を達成するために、以下の方法で研究した。

担当者らは、本研究期間の前に同目的のシステムを開発していた。このシステムでは、ソーシャルメディアのユーザが、漏洩したくない情報（たとえば通勤先）を NG ワードとしてシステム上で予め定義しておく。システムは、開示しようとする投稿文からの単語を抽出し、これらの単語の組合せをキーワードとして Web 検索を実行し、検索結果の中に NG ワードがどれだけ含まれるかによって情報漏洩を検知していた。しかし、このシステムでは、投稿文中の単語の組合せが多数であるため、Web 検索の回数が多く、検知の処理時間が 1 文平均 10 秒と大きかった。また、検知率が 70%程度、誤検知率が 30%程度であり、検知精度が不十分であった。さらに、漏洩したくない情報をユーザが NG ワードとして登録する必要があり、使い勝手が低かった。そこで、効率、精度、使い勝手を向上するため、以下の改良を行った。

(1-1)効率の向上

従来システムでは、形態素解析器として Mecab を用いていたが、その単語辞書が不十分であるため、「電気通信大学」のような複合語を「電気」、「通信」、「大学」のように複数の単語に分解してしまう。その結果、Web 検索のキーワードとなる単語の組合せ数が増加し、検索回数ひいては処理時間が大きかった。そこで、東京大学の提供する専門用語（キーワード）自動抽出システムおよび、はてなキーワード、Wikipedia の見出しを用いて複合語を認識し、効率化を図った。

また、従来システムでは、たとえば、ａ、ｂ、ｃの３つのキーワードを用いて Web 検索を実行する際に、abc，acb，bac など単語の全ての順列で検索していた。これを省略し、abc のみで検索した場合に、検知精度が低下するかを評価した結果,有意の精度低下は見られなかったため、一順列のみの検索とし、効率化を図った.

(1-2)精度の向上

個人情報の漏洩検知の目的から、誤検知率が増加しても、検知率を向上することが重要であるため、検知率に重点を置いた処理方式とした。また、誤検知率の増加については、ユーザに負担のないインターフェースを開発することで対処した。

(1-3)使い勝手の向上

従来システムでは、NG ワードとして、具体的な固有名詞等（たとえば通学先＝電気通信大学）を登録する必要があったが、これはユーザ毎に自分で登録する必要があり、使い勝手が低かった。そこで、通学先＝大学といった NG ワードの汎用的なメニューをシステム側で提供可能であるか検討した。このような汎用的な定義を用いて、個々のユーザの投稿文から当該ユーザの通学先の漏洩を検知できるかを評価した。

上記第２の目的（自然言語テキストの形式化と Linking）を達成するために以下の方法で研究した。

(2)企業等が社員や就職希望者の背景情報をソーシャルメディア上で調査することが、プライバシー侵害として欧米で問題になっている。この問題をリアルなプライバシーリスクの例として取り上げ、ソーシャルメディア上の多数の投稿文の中から、当該社員や就職希望者の投稿文を自動抽出できるかを検討した。提案技術は、多数の投稿文のうち当該社員や就職希望者の履歴書と最も良く照合する投稿文を本人の投稿文として検出する。本技術では、履歴書をベクトルで表現し、ベクトルの各要素を履歴書中の名詞の重みとした。一方、投稿文もベクトルで表現した。投稿文のベクトルの各要素は、履歴書の名詞の投稿文における重みとした。履歴書と投稿文の照合度合は、二つのベクトルのコサインによって定量化した。

履歴書ベクトルと投稿文ベクトルにおける名詞の重みとして、最初は TF-IDF（Term Frequency，Inverse Document Frequency）を用いた。ところが、履歴書の名詞は、本人の投稿文でも直接記載されることは殆どないことが判明した。TF-IDF は履歴書および投稿文における名詞の出現頻度に比例することから、履歴書ベクトルと投稿文ベクトルのコサインは、本人であっても小さな値となり、他人との区別ができないことが判明した。そこで、上記第１の目的達成のために開発した、Web 検索による単語間の間接的な繋がり検知手法を応用することにした。すなわち履歴書の名詞を NG ワードとみなし、投稿文における名詞の組合せから NG ワードへのつながりの強さを、投稿文ベクトルにおける名詞の重みとした。

４．研究成果

上記第１の目的に関して、以下の成果を得た。

(1-1)処理効率については、検索精度を維持しながら処理時間を平均で約 1/20 に短縮することができた。

(1-2)検知精度については、電気通信大学の学生 11 名を被験者とし、各 1000 件の投稿文について検知精度を評価した結果、平均して検知率が約 85%に向上した。提案技術が検知した情報漏洩のうち 10%は、人間が気付かなかった漏洩であり、提案技術により人間の能力を補うことが可能である。また、検知できなかった 15%の情報漏洩文を分析した結果、ソーシャルメディアの投稿文に特有の口語的表現および友人の間だけで通用する単語により、単語の自動認識に失敗していたことが判明した。そのため、今後の課題として、単語の認識に関して複数の可能性を試行する方式および検知に成功した投稿文から単語を学習する方式の必要性が明らかになった。一方、誤検知率は約 85%に増加した。誤検知について詳細に分析した結果、投稿文の約 5%についてシステムがアラートを発し、そのうちの 85%が不必要なアラートであったことが判明した。そのため、不必要なアラートの頻度は大きくない。そこで、ユーザにとって煩わしくないアラートの表示方法といったインターフェースの工夫によって対処可能である。

(1-3)使い勝手の向上については、通学先＝大学といった NG ワードの汎用的な定義を用いた場合、誤検知率が 95%に達することが判明した。その原因は、たとえば、通学先が電気通信大学であっても、投稿文から電気通信大学以外に東京大学などの有名大学が推定され、これを通学先情報の漏洩として検知することにある。そのため、今後の課題として、Web 検索結果に２つの大学名が同程度に含まれる場合、有名大学の推定度合を低くするといった方式の必要性が明らかになった。これは、単語の重み定義 TF-IDF（Term Frequency，Inverse Document Frequency）における Inverse Document Frequency に相当する要素を、個人情報の推定に取り入れることを意味する。

上記第２の目的に関して、以下の成果を得た。

(2)Twitter を例とし、履歴書との照合により本人のつぶやきを特定できるか評価した。企業の社員 2 人、電気通信大学の学生 10 人を被験者とし、背景ノイズとしてネット上からランダムに選定した 100 ユーザを用いた。各被験者について、本人のつぶやきと背景ノイズ 100 人のつぶやきの中から本人のつぶやきを特定できるかを評価した。

社員２人については、本人およびノイズ 100 人のつぶやきが各 100 件あれば、ほぼ 100%の精度で本人のつぶやきを特定することができた。一方、学生 10 人については、本人およびノイズ 100 人のつぶやきが各 100 件の場合は 50%以下の特定率であった。学生の特定率が低い原因は、履歴書の情報が少ない（たとえば職歴がない）ことであった。しかし、本人およびノイズ 100 人のつぶやきが各

1000 件になると、10 人中 6 人について本人のつぶやきを正しく特定することができ、3 人についても 101 人から 4 人まで絞り込むことができた。このことから、ソーシャルメディア上の投稿数が多くなるほど、個人が特定されやすくなることを定量的に示すことができた。

以上(1-1)，(1-2)および(2)の検討を通じて、ソーシャルメディアの投稿文におけるプライバシー情報の処理モデルおよび外部情報が個人特定に与える影響を明らかにした。本モデルは、プライバシー情報という曖昧性、個人性の高い情報を履歴書等のキーワード集によって簡潔に定義する仕組みと、投稿文におけるプライバシー情報の多様な表現と履歴書等のキーワードとを照合する仕組みから成る。Web 検索を用いてインターネット上の外部知識を利用することで、多様な表現とキーワードの間接的な照合を可能にする。

５．主な発表論文等

〔雑誌論文〕（計４件）

[1] Hoang-Quoc Nguyen-Son, Minh-Triet Tran, Hiroshi Yoshiura, Noboru Sonehara, Isao Echizen: Anonymizing Personal Text Messages Posted in Online Social Networks and Detecting Disclosures of Personal Information, IEICE Transactions on Information and Systems, Vol.E98-D, pp.78-88, 2015.

[2] 片岡春乃，奥野智孝，木村聡一，内海彰，吉浦裕：ソーシャルネットワークから注目者の発言を特定するシステムの提案と予備評価,日本セキュリティ・マネジメント学会誌，Vol.27，NO.3，pp. 13-28，2014.

[3] Akira Utsumi, Maki Sakamoto: Indirect categorization as a process of predicative metaphor comprehension, Metaphor and Sympobol, Vol.26, pp.299-313, 2011.

[4] 加藤慧，小宮山功一朗，瀬古敏智，一瀬友祐，河野耕平，中山心太，吉浦裕：コンテンツベースフィッシング検知手法大規模実例評価と改良，日本セキュリティマネジメント学会誌，Vol.25，No.2，pp.42-56，2011.

〔学会発表〕（計１５件）

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

[2] 吉浦裕：特定人物の投稿したつぶやきを多数のなかから見つけ出す技術，日本セキュリティ・マネジメント学会全国大会，東京，2014.6.21.

[3] 吉浦裕：匿名性維持可能性検討のための特定人物の投稿したつぶやきを多数の中から見つけ出す技術，日本セキュリティ・マネジメント学会 IT リスク学研究会，東京，2014.2.22.

[4] 吉浦裕：ソーシャルメディアのプライバシーと個人特定，電子情報通信学会マルチメディア情報ハイディング・エンリッチメント研究会（招待講演），仙台，2014.1.27.

[5] Hoang-Quoc Nguyen-Son, Anh-Tu Hoang, Minh-Triet Tran, Hiroshi Yoshiura, Noboru Sonehara and Isao Echizen: Anonymizing Temporal Phrases in Natural Language Text to be Posted on Social Networking Services, 12th International Workshop on Digital-forensics and Watermarking, LNCS8389, Auckland New Zealand, 2013.11.1-4.

[6] Tomotaka Okuno, Masatsugu Ichino, Isao Echizen, Hiroshi Yoshiura: Ineluctable Background Checking on Social Networks: Linking Job Seeker's Resume and Posts, 5th IEEE International Workshop on Security and Social Networking, San Diego, 2013.3.18.

[7] Hoang-Quoc Nguyen-Son, Minh-Triet Tran, Dung Tran, Hiroshi Yoshiura and Isao Echizen : Automatic Anonymous Fingerprinting of Text Posted on Social Networking Services, 11th International Workshop on Digital-forensics and Watermarking, LNCS7809, Shanghai, China, 2012.10.30-11.3.

[8] Midori Hirose, Tomoya Muraki, Akira Utsumi, Isao Echizen, Hiroshi Yoshiura : A Private Information Detector for Controlling Circulation of Private Information through Social Networks, Second International Workshop on Resilience and IT-Risk in Social Infrastructures, Prague, Czech, 2012.8.23-24.

[9] Hoang-Quoc Nguyen-Son, Quoc-Binh Nguyen, Minh-Triet Tran, Dinh-Thuc Nguyen, Hiroshi Yoshiura, and Isao Echizen: Automatic Anonymization of Natural Languages Texts Posted on Social Networking Services and Automatic Detection of Disclosure, 7th International Workshop on Frontiers in Availability, Reliability and Security, Prague, Czech, 2012.8.23-24.

[10] Tomotaka Okuno, Hiroshi Yoshiura: Identifying Anonymous Posts of Job Seekers, 21st USENIX Security Symposium, Bellevue, USA, 2012.8.8-10.
[11] Hoang-Quoc Nguyen-Son, Quoc-Binh Nguyen, Minh-Triet TRAN, Dinh-Thuc Nguyen, Hiroshi Yoshiura, Isao Echizen: New Approach to Anonymity of User Information on Social Networking Services, 6th International Symposium on Digital Forensics and Information Security, Vancouver, Canada, 2012.6.26-28.
[12] Akira Utsumi: Extending and evaluating a multiplication model for semantic composition in a distributional semantic model, 11th International Conference on Cognitive Modelling, Berlin, 2012.4.13-15.
[13] 村木友哉, 市野将嗣, 越前功, 吉浦裕：類似度に基づく匿名性定量化手法とその顔画像匿名化への応用, 第 29 回暗号と情報セキュリティシンポジウム, 金沢, 2012.1.30.
[14] 奥野智孝, 市野将嗣, 久保山哲二, 吉浦裕：ソーシャルメディア情報と履歴書情報の照合を通じた個人の言動の特定, 情報処理学会第 55 回 CSEC 研究会, 東京, 2011.12.5.
[15] 広瀬緑, 吉浦裕：学習を必要としない自然言語文からの個人情報検知技術, 情報処理学会第 55 回 CSEC 研究会, 東京, 2011.12.5.
[16] Tomotaka Okuno, Masatsugu Ichino, Tetsuji Kuboyama, and Hiroshi Yoshiura: Content-based De-anonymisation of Tweets, Seventh International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Dailin, China, 2011.10.14.

〔図書〕（計１件）

吉浦裕：セキュリティマネジメント学～理論と事例～, 第 2 章「工学的アプローチ」, 共立出版, 2011.

６．研究組織

(1)研究代表者

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

(2)研究分担者

内海彰（UTSUMI Akira）
電気通信大学・大学院情報理工学研究科・教授
研究者番号：３０２５１６６４