

ネットパトロール支援システムの研究開発

中 村 健 二

要旨

近年、携帯電話利用者の低年齢化が進み、モバイルインターネットから端を発した問題が様々な局面で子供の安全を脅かしている。そのため、各教育機関では、生徒に対してモバイルインターネットの危険性を伝える活動や、ネットパトロールを通して自校の生徒の活動を継続的に見守る活動を実施している。しかし、現在のネットパトロールは、検索エンジンを利用した人手での活動が中心であるため、自校の生徒のみを見つけ出し、その生徒の活動を継続的に監視することは困難である。そのため、本研究では、モバイルインターネットの中でも、中高生の利用率が高いプロフサービスを起点として、そのサービスから広がる子供たちの行動を継続的に監視するネットパトロール支援システムを開発する。本システムでは、プロフサイトの記述内容と個人間の関連度から生徒と各学校との関連性を求めることで、特定の学校に関係の深い生徒を監視することが可能である。

1. は じ め に

近年、携帯電話の利用者が低年齢化しており、携帯電話が子供たちの生活に欠かせないコミュニケーションツールとなりつつある。2009年5月に文部科学省が実施した「子どもの携帯電話等の利用に関する調査」¹⁾によると、小学校6年生の24.7%、中学2年生の45.9%、高校2年生の95.9%が携帯電話を所有している状況である。携帯電話には、様々な機能があり、通話やメール、ゲームやインターネットなど、子供に対して魅力的な道具となっている。こうした機能により、携帯電話は、保護者の監視の目をすり抜け、友人や他人とつながることができるダイレクトコミュニケーションの特性を持ったパーソナルメディアであり、子供たちが危険にさらされている状況である。特に、携帯電話のインターネット機能（以下、「モバイルインターネット」と略記）は、子供たちの人間関係を学校内だけにとどまらず、格段に広げている状況である。ここで、モバイルインターネットの利用目的¹⁾を確認すると、高校2年生では、「メールのやりとりをする90.5%」「音楽などのダウンロードをする70.2%」「他人のプロフやブログなどを見る63.3%」「他人のプロフやブログなどに書き込みをする52.6%」となっており、約半数以上がプロフやブログのチェックをしていることが分かる。また、モバイルインターネットを利用した自身の情報発信も盛んであり、「自分のプロフを公開する40.1%」「自分のブログを公開する35.5%」と、約4割がプロフやブログを利用して自身の情報を公開している状況¹⁾となっている。しかしながら、携帯電話の利用者の低年齢化に伴い、携帯電話やモバイルインターネットに関する様々な問題²⁾⁻⁷⁾が顕在化している。実際に、2008年7月の東京都教育委員会の「子供の

インターネット・携帯電話利用についての実態調査⁸⁾では、高校生の29.2%がモバイルインターネット利用において何らかのトラブルに巻き込まれた経験があると回答している。プロフに関するトラブルの顕著な例として、2008年10月にさいたま市立中学3年の女子生徒が自身のプロフにおいてネットいじめに巻き込まれ自殺した事件や2008年12月に東京都足立区に住む私立高3年の男子生徒がプロフの書き込み内容をめぐり同級生を包丁で刺す事件などが発生している。このように、プロフを中心として、ネットいじめや非行・逸脱行為が発生しており、ひとつの社会問題となっている。これらの原因は、プロフのメディア特性とそのプロフを利用する子供の利用方法に起因すると考えられる。下田は、著書⁹⁾の中でプロフを利用した問題点について次のように述べている。

- 子供が無自覚に個人情報を流してしまう（その結果、自分だけでなく友達の情報も悪用される）
- 子供が意図的に、たとえばいじめ目的で、友達の個人情報を流したり、友達になり済まして発信するなど悪質な使い方がされやすい
- 危ない大人を呼び込んでしまう（インターネットナンパ被害）
- 出会い系サイトや十八禁サイトなど、インターネット風俗の有害情報につながってしまう
- 刺激的発信の材料を自分で作ろうとして、非行・逸脱行為に走るようになる（たばこを吸ったり酒を飲んだりしている写真を発信するなど）
- 自分のプロフの人気を上げたくて、刺激的な発信をはじめてしまう（裸の写真やデマなど有害情報）

このように、モバイルインターネットの中でも、プロフには様々な問題が潜在しており、プロフ内の危険から子供を保護するためには、プロフにおける子供たちの行動を継続的に監視し、問題がある子供を的確に指導することが重要な課題となっている。そのため、本研究では、子供たちがモバイルインターネット上で公開している情報を解析し、各学校の教員が自身の生徒の行動を見守ることが可能なシステムを開発する。このことにより、非行・逸脱行為を起こしそうな生徒や危険な大人とつながっている生徒を事前に把握でき、事件を未然に防ぐことができると考えられる。

2. 関連する取組

現在、モバイルインターネットなどにおける子供の安全に関連する取組として、有害情報フィルタリングやネットパトロールに関する取組が実施されている。有害情報フィルタリングに関する取組では、2009年4月の「青少年ネット規制法（青少年が安全に安心してインターネットを利用できる環境の整備等に関する法律）」の施行に伴い、各携帯電話会社は有害なサイトをフィルタリングするサービスの提供を開始している。また、従来のフィルタリングサービスでは、フィルタリングの形態によっては、コミュニティサイトやショッ

ピングサイトなども一様に有害と判断されアクセスできない状況となり、利便性が損なわれるという問題が発生する。そのため、モバイルインターネットのコンテンツの健全性を審査するモバイルコンテンツ審査・運用監視機構（EMA: Content Evaluation and Monitoring Association）¹⁰⁾ が設立され、健全なモバイルサイトを認定する活動が行われている。その結果、健全と認定されたモバイルサイトは、携帯電話会社のフィルタリング対象から除外される動きがなされており、フィルタリングソフトの利便性が向上している状況である。しかし、フィルタリングソフトの利用率は低く、2009年1月時点において、高校生の69.3%は利用していない¹¹⁾ という状況であり、喫緊の問題には対処できていないのが現状である。さらに、今回対象とする問題はプロフなどのコミュニティサイトが起点となっており、対象サイトにて取り交わされる人と人とのコミュニケーションから問題へと発展する事例である。そのため、コンテンツの内容を審査するもしくは、コンテンツの内容に応じてフィルタリングするフィルタリングソフトでは、根本の問題を解決することは困難な状況である。

一方、ネットパトロールに関する取組では、（A）教育委員会や学校関係者による自主的な取り組み^{12)~15)}、（B）ガイアックス¹⁶⁾のスクールガーディアンやピットクルー¹⁷⁾の監視などの民間企業を中心とした取り組みと（C）非営利団体やボランティア団体による活動¹⁸⁾がある。（A）の取り組みでは、学校裏サイトやプロフの監視業務を専門とする職員を雇い、人手でインターネットから情報を収集する活動を行っている。しかし、人手で検索しているため、網羅的に大量の情報を収集することが困難である。（B）の取り組みでは、基本的には人手での監視を中心としているが、過去のノウハウを利用した専門の検索エンジンを用いているため、学校裏サイトやプロフなどを中心として効率よく情報収集することができる状況である。しかし、依頼された学校に対して情報を検索し、その時点での解析を行い報告するため、対象となる学校が制限されるという問題や継続的なネットパトロールには費用がかかるという問題がある。（C）の取り組みの中でも、文献¹⁸⁾では、市民インストラクター支援システム（CISS: Civil Instructor Support System）¹⁸⁾を用いて、ネットの遊び場の情報、トラブル事例・対応策をデータベース化し、子供の見守り・指導をする人材の育成、活動支援している。この活動の中では、非行・逸脱行為の有無や、その前兆となる生徒の情報を見つけ出し、その生徒について継続的に見守っている。しかし、市民インストラクターによる監視活動であるため、（A）の取り組みと同様、監視範囲に限界がある。

これらの取組から、安全なモバイルインターネットを実現するためには、ネットパトロールの強化が重要であると考えられる。しかし、現在実施されているネットパトロールは、人手によるものが一般的であり、インターネット上の情報を網羅的に監視することは困難である。また、本システムの構築時において教育関係者にヒアリングした結果、個人を中心として広がる人間関係が、子供の非行・逸脱行為もしくは、事件につながる危険性があることが分かった。そのため、子供の安全を確保するためには、「非行・逸脱行為を起こした生徒のその後の行動」、「非行・逸脱行為を起こした生徒と友人関係にある生徒の行動」

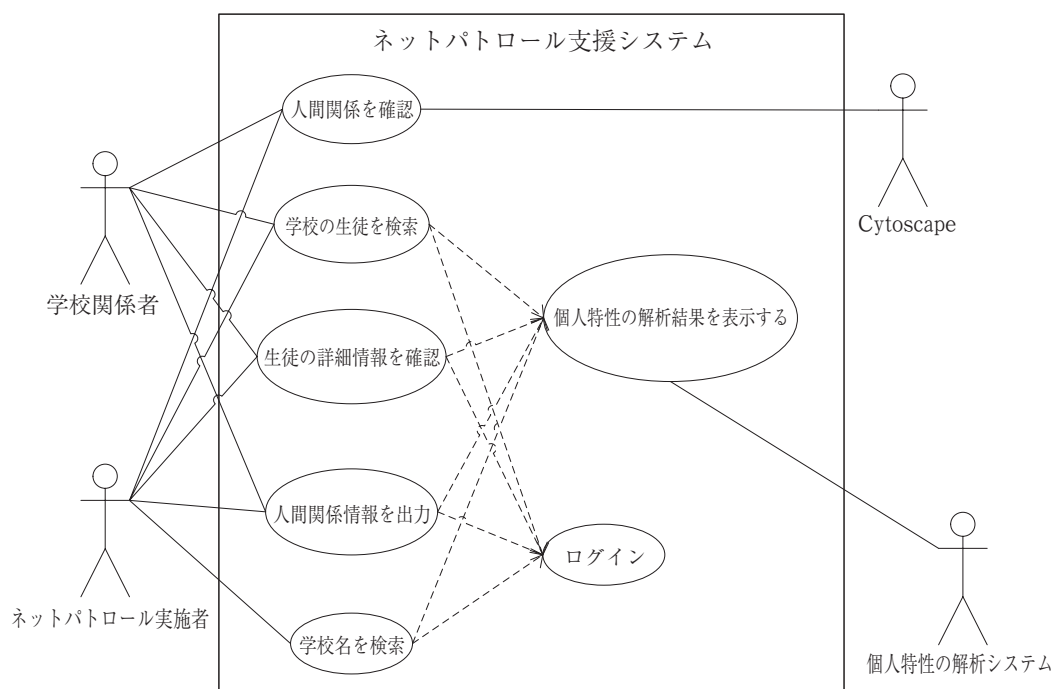


図1 ネットパトロール支援システムのユースケース図

や「見知らぬ人との交流の有無」などの情報を定期的に継続して監視することが必要であると考えられる。しかし、有人の監視では、継続的に広範囲の生徒を確認することが困難である。そのため、本研究では、ネットパトロールの支援を目的として、プロフサイトを中心に広がるモバイルインターネット上での子供のネット遊びを継続的に監視し、人の繋がりや非行・逸脱行為などの解析結果を提供するシステムを開発する。

3. ネットパトロール支援システムの構成

3.1 ネットパトロール支援システムの概要

ネットパトロール支援システムは、モバイルインターネットにおける子供の非行・逸脱行為を自動的に監視し、その結果をネットパトロール実施者および学校関係者に提供するシステムである。ネットパトロール支援システムのユースケースを図1に示す。

本研究では、学校関係者やネットパトロール実施者が自校の生徒のインターネット上での危険な行動を監視し、危険な行動を起こしている場合は、適切に指導可能な情報の提供を目指している。そのため、ネットパトロール支援システムでは、プロフに記述されている内容から所属する学校を類推し、個人を特定した上で定期的に監視する仕組みを実現する。そして、各個人の所属学校の類推結果を用いて、特定の学校の生徒を検索する機能や生徒の人間関係を確認可能な機能を提供する。このことにより、学校の教師が把握できていないインターネットからつながった関係や表層化していない各生徒の非行・逸脱行為を

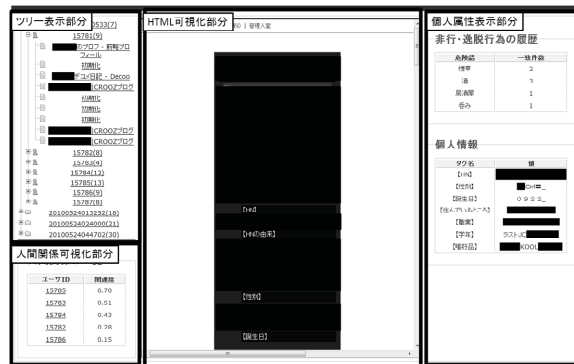


図2 個人情報ブラウザ画面

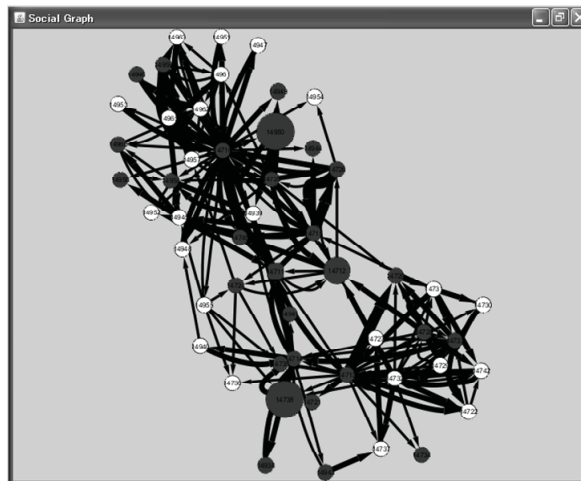


図3 ソーシャルグラフの例

準リアルタイムに把握できる。

さらに、人間関係を確認するためのソーシャルグラフを自動的に出力する機能を実装することで、特定の個人から広がる人間関係を容易に俯瞰する仕組みを提供する。このことにより、「非行・逸脱行為を起こした生徒と友人関係にある生徒の行動」や「見知らぬ人との交流の有無」を容易に確認できる。

3.2 ネットパトロール支援システムの画面遷移

ネットパトロール支援システムは、学校名を検索キーとして生徒の一覧を検索し、検索結果から生徒を選択することで、その生徒についての日々の行動や人間関係を確認できる。生徒の詳細情報を確認する機能として、個人情報のクローリング結果を確認する個人情報ブラウザ画面（図2）と生徒の親密な人間関係の情報をソーシャルグラフとして出力し、Cytoscape¹⁹⁾にて可視化する機能（図3）がある。本稿では、個人情報保護のため図2に

表1 ソーシャルグラフの属性情報

出力種類	属性	説 明
ノード	形状	□：出力する際の中心となる生徒
		○：出力する際の中心となる生徒に関係する人物
	ノードの大きさ	非行・逸脱傾向の強度（非行・逸脱用語辞書との一致件数）
	色	赤色：非行・逸脱行為 あり
		白色：非行・逸脱行為 なし
	表示名	名前と学年に関わる情報を出力
	個人属性	個人情報の抽出結果を出力
	プロフ URL	実際のプロフサイトへ接続する URL を表示
エッジ	グループ化 ID	関連度に基づきグループ化した結果を出力
	矢印の向き	関係の方向性
	エッジの太さ	関連度（メッセージの通信回数）
	ニックネーム	A→Bのエッジが記述されている場合、AさんがBさんと呼ぶ場合のニックネームを出力

は墨塗りを施し、図3はIDのみの表示としており、実際のシステム画面とは一部異なるものを表示している。

図2の個人情報ブラウザ画面では、個人領域をクローリングし、その際に解析した結果を確認することができる。個人情報ブラウザ画面は、ツリー表示部分、HTML可視化部分、個人属性表示部分と人間関係可視化部分の4つで構成される。ツリー表示部分では、その個人について解析した日付と、その際に得られたページの情報を一覧で表示する。ツリービューに表示されているノードを選択することで、そのページの情報がHTML可視化部分に表示される。このことにより、継続的に解析した結果を時系列で確認することが可能となる。HTML可視化部分では、ツリービューで選択したノードに一致するHTMLを表示する。個人属性表示部分では、非行・逸脱行為の履歴、プロフから取得した個人情報を表示する。非行・逸脱行為の履歴では、クロールした際に、非行・逸脱用語辞書の内容と一致した記述が含まれていた場合にその情報を出力する。このことにより、飲酒や喫煙などの非行・逸脱行為の有無やネットいじめなどへの関連を概観することが可能となる。プロフから取得した個人情報の表示では、プロフページに記述している個人に関する情報を自動的に取得した結果を表示する。人間関係可視化部分では、その生徒に直接関係する人を関係の深い順に人間関係の一覧を表示する。このことにより、その生徒が「見知らぬ大人と密に連絡を取り合っている」や「非行・逸脱行為を行っている生徒と友人関係にある」など、学校や保護者が把握できていない関係を把握可能にする。

図3の人間関係の可視化機能では、個人情報の解析結果をCytoscapeにて可視化する。Cytoscapeとは、複雑なネットワークを可視化、解析するためのプラットフォームであり、

Cytoscape の機能を使いこなすことで、利用者のニーズに合わせた結果を柔軟に表現できる。しかし、ネットパトロール支援システムは、ネットパトロール実施者や学校関係者を対象としているため、簡易な操作で生徒の情報を把握できる必要がある。そのため、本システムでは、Cytoscape 用に出力する情報に対して、解析結果を属性として付与することで、可読性の高いソーシャルグラフを出力する。ソーシャルグラフに出力する属性を表 1 に示す。

3.3 ネットパトロール支援システム構築時の課題と対応方法

3.3.1 ネットパトロール支援システム構築時の課題

本研究では、学校関係者が実現場にて利用可能なネットパトロール支援システムを構築するために、次に示す 2 つの課題に対応する必要がある。

● 個人領域の特定問題

生徒指導へつなげるためのネットパトロールでは、生徒一人一人の日々の行動を各個人単位に分類して監視する必要がある。そのため、本研究では、プロフを起点としてつながる様々な Web ページの中で、そのプロフユーザが関係するりあるやブログなどのサービスは、URL 構造が類似する特性に着目し、そのユーザが関係する領域のみを抽出する手法を実現する。

個人領域とは、プロフからリンクするページの中で、そのユーザが関連する情報であり、具体的には、そのユーザが公開するホーム、ブログ、りあるやゲストブック（以下、ゲストと略記）のサービスを利用して公開した Web ページ群である。これらのサービスは、プロフも含めて様々な企業が提供しており、各企業のプロフサービスに特化しない汎用的な手法を考案する必要がある。

● 生徒情報の検索に関する問題

本研究では、ネットパトロールの結果を学校の教育現場へ反映するため、生徒一人一人の個人情報および個人領域の記述内容に基づき、そのプロフユーザが通っている学校を推定する。しかし、実際のプロフサイトを確認したところ、個人情報として学校名（もしくは学校名の略称）を記述しているユーザと学校に関する情報を記述していないユーザがまばらに存在することが分かった。プロフに学校の情報を記述していないユーザは、学校名による検索で一致しないため、ネットパトロールの対象から外れる可能性がある。そのため、本研究では、学校に関する情報を記述していないユーザであったとしても、ユーザ間の関連度に基づきプロフユーザをクラスタリングすることで、学校とユーザ間の関連度を算出する手法を用いる。このことにより、生徒情報の検索に関する問題に対応したネットパトロール支援システムを構築する。

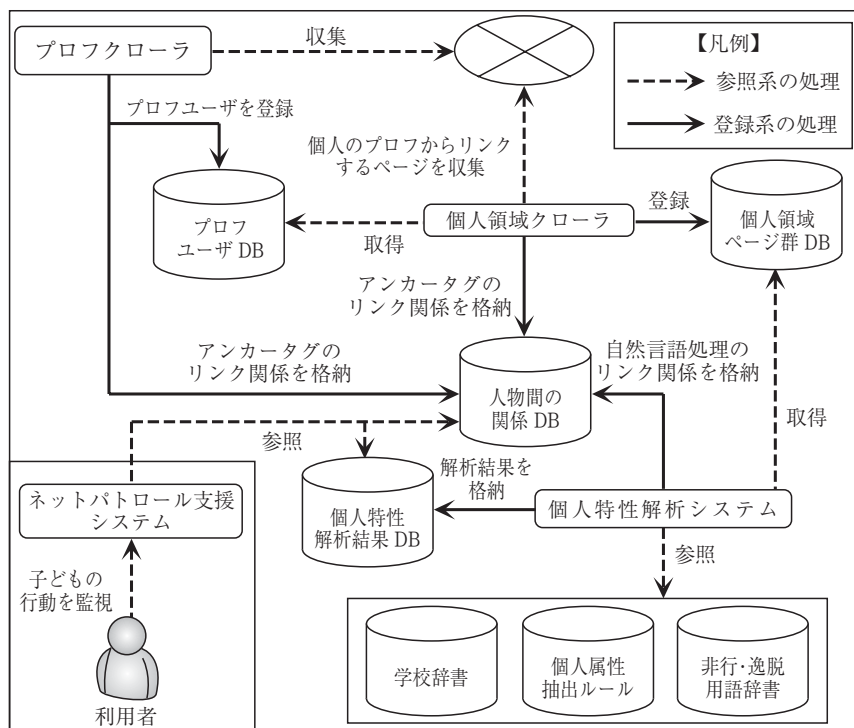


図4 システムの構成

4. システム構成

4.1 全体構成

本研究は、ネットパトロール支援システムとそのシステムで提供する情報を収集・解析する3つの内部システムで構成される。内部システムは、インターネット上から情報収集を行うプロフクローラと個人領域クローラ、その収集結果を解析する個人特性解析システムで構成される。本研究のシステム構成を図4に示す。

本システムでは、まず、プロフクローラがインターネット上のプロフサービスを解析し、プロフサイトに登録しているユーザを網羅的に収集して、プロフユーザ DB に登録する。次に、プロフクローラが収集したプロフユーザの情報を起点として、個人領域クローラが、各プロフユーザに関係する情報を収集し、個人領域ページ群 DB に登録する。そして、収集された個人情報ページ群 DB を解析し、そのユーザの人間関係およびユーザの個人情報を解析して、それぞれを人物間の関係 DB および個人特性解析結果 DB に登録する。ここで、本システムでは、これらの情報収集と解析処理を一定期間毎に実行することで、持続的に子供の非行・逸脱行為を監視する。そして、ネットパトロール実施者は、ネットパトロール支援システムを用いて、それらの解析結果および人物間の関係を確認し、子供の非行・逸脱行為の監視を行う。それぞれのシステムの役割を次に述べる。

4.2 プロフクローラ

プロフクローラは、プロフサービスに登録されているユーザをプロフページのリンク構造に基づき網羅的に収集する。本処理では、各プロフサービスとサービス毎のユーザ特定の URL フィルタを入力として、プロフサービスと同一ドメインページを網羅的に収集する。そして、収集されたページの URL およびそのページのリンク先 URL を解析し、ユーザ特定の URL フィルタと一致した場合は、その URL をプロフのトップページとしてプロフユーザ DB に登録する。このことにより、当該プロフサービスに登録されているユーザを網羅的に収集することが可能となる。

4.3 個人領域クローラ

個人領域クローラは、プロフクローラが収集したプロフユーザ DB に格納されている各プロフユーザのトップページを起点として、リンクするページを収集・解析し個人領域を特定する。リンクするページを収集する際に、単純にリンク構造のみをたどると広告ページやプロフランキングページなどの不要なページを収集し、個人の特性や個人から広がる人間関係を正しく判断できないという問題が発生する。そのため、本研究では URL の構造に着目し個人領域を特定する手法を実装することで、起点となるプロフユーザに関連するページのみを収集する。個人領域の特定手法については、第 5 章にて論じる。

4.4 個人特性解析システム

個人特性解析システムでは、プロフクローラおよび個人領域クローラが収集した情報を解析し、ネットパトロール支援システムで利用する情報をデータベースに格納する。個人特性解析システムでは、「個人領域に含まれる個人情報抽出」、「人と人の関係情報の抽出」と「所属学校の推定」の 3 つの解析を行う。個人領域に含まれる個人情報の抽出処理では、個人属性抽出ルールを用いて、個人領域に含まれる情報を個人の属性として取得する。また、非行・逸脱用語辞書を用いて、個人領域に含まれる非行・逸脱用語が含まれるかどうかを単語一致で判定する。人と人の関係情報の抽出処理では、個人領域間のリンク関係を抽出する。りあるやブログなどの日記サイトでは、その日の出来事を紹介する際に友人の名前が頻出するが、その名前に対してリンクを貼り付けることは非常にまれである。そのため、通常のリンク関係のみを対象とした場合、結果として非常に親密な関係であるが、プロフ間ではリンクがつながっていないもしくは、リンク数が 1 件と疎な関係と認識される可能性がある。そのため、本研究では、通常のリンク関係に加えて自然言語処理でのリンク関係²⁰⁾を用いて人と人との関係グラフを構築する。所属学校の推定処理では、個人情報の解析によって得られた情報から、学校名称および住所等の情報を利用して所属学校の推定を行う。このとき、一人のユーザから取得した場合には、学校名に関する記述がない場合に所属学校の推定を行うことができないという問題が発生する。そのため、本研究では、人と人の関係情報に基づきクラスタリングを行い、クラスタリングされたユーザの情報に基づき所属学校の推定を行う。ここで、所属学校を一意に特定するのではなく、そ

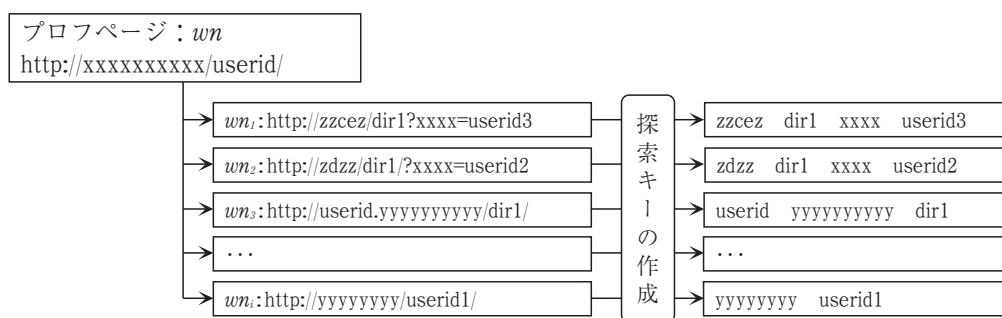


図5 探索キーの生成処理

の学校の生徒であると判定されたものを所属学校の候補として取得する。

これらの解析により得られた結果から学校とユーザの関連度を算出する。そして、関連度に基づいてネットパトロール支援システムの検索結果順位を決定する。学校とユーザの関連度算出手法については、第6章にて論じる。

5. 個人領域の特定手法

5.1 個人領域の特定の流れ

個人領域の特定処理では、各ユーザのプロフページからリンクするページの中で、そのユーザに関連するページのみを収集する。個人領域の特定処理では、プロフページからリンクしている他のサービス（ホーム、りある、ゲスブやブログなど）を個人領域としてそのプロフユーザに関連付ける。個人領域の特定では、プロフページからリンクする他のサービスのURLは、そのサービスにおけるユーザ領域（ex:http://domain/userid/）を示しており、ユーザ領域から派生するページは、そのユーザ領域の配下もしくは同階層に格納される点に着目し、個人領域の収集を行う。

5.2 本研究の手法

個人領域の特定は、「探索キーの生成処理」、「探索対象の判別処理」と「個人領域の判定処理」3つのステップで実現する。

（1）探索キーの生成処理

探索キーの生成処理（図5）では、プロフページからリンクするURLをもとに探索の有無を判定するための探索キーを生成する。まず、プロフクローラにて収集した任意のプロフページ wn を取得する。ここで、プロフユーザの集合を N 、各プロフユーザを n とする。次に、取得した wn にリンクする1階層目リンクのURLを取得する。取得したURLの集合を $(wn_1, wn_2, wn_3, \dots, wn_i)$ とする。最後に、URLを解析し、ドメイン、ディレクトリ名、クエリ引数や値の集合を wn_i の探索キー $Element(wn_i)$ として生成する。

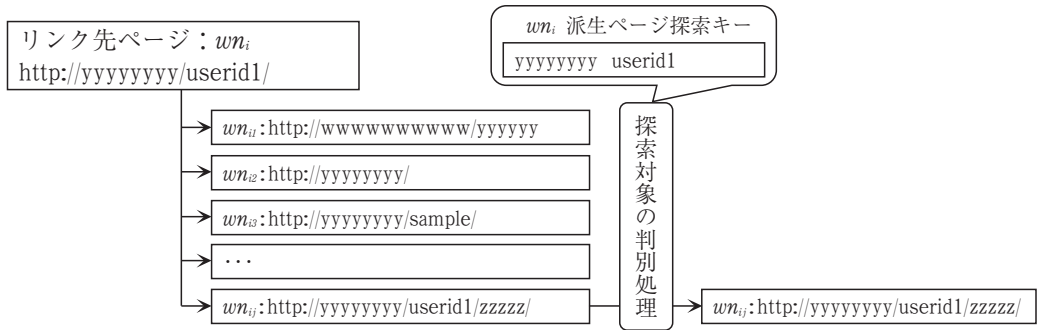


図6 探索対象の判別処理

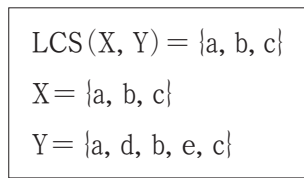


図7 LCSの例

(2) 探索対象の判別処理

探索対象の判別処理（図6）では、プロフページからリンクするページのURLを取得し、探索キーとの一致率に基づき探索対象かどうかを判別する。まず、探索元となる wn_i を取得し、そのページから派生リンクを取得する。ここで、 wn_i から取得したURLの集合を $WN_{ij} = (wn_{i1}, wn_{i2}, \dots, wn_{ij})$ とする。次に、式(1)にて $Ext(wn_i, wn_{ij})$ の値を求める。

$$Ext(wn_i, wn_{ij}) = \frac{(count(LCS(Element(wn_i), Element(wn_{ij}))))}{(count(Element(wn_i)))} \quad (1)$$

式(1)において、 $LCS(X, Y)$ は、2つの集合 X, Y 間の Longest Common Subsequence (LCS: 最長共通部分列) を算出する関数である。 $LCS(X, Y)$ の出力例は図7に示す通りとなる。そして、 $Ext(wn_i, wn_{ij})$ の結果が1となるURLを次階層探索候補URL集合として抽出する。ここで、ユーザ特定ルールもしくは、事前に作成した広告サイトなどのURLを登録したブラックリストであるスパムリストと一致する場合は、無条件で探索対象外と判別する。最後に、次階層探索候補URL集合の各要素を探索元URLとして、次階層探索候補URLを取得できないもしくは、探索制限回数の上限に至るまで繰り返し探索を行う。

(3) 個人領域の判定処理

個人領域の判定処理（図8）では、探索対象の判別処理にて抽出したURLを個人領域に含めるかどうかの判定を行う。まず、 wn_i から派生して収集したURLの集合 WN_{ij} を取得する。次に、要素数 $Count(WN_{ij})$ が閾値 α 以上となる WN_{ij} を抽出し、それらの集

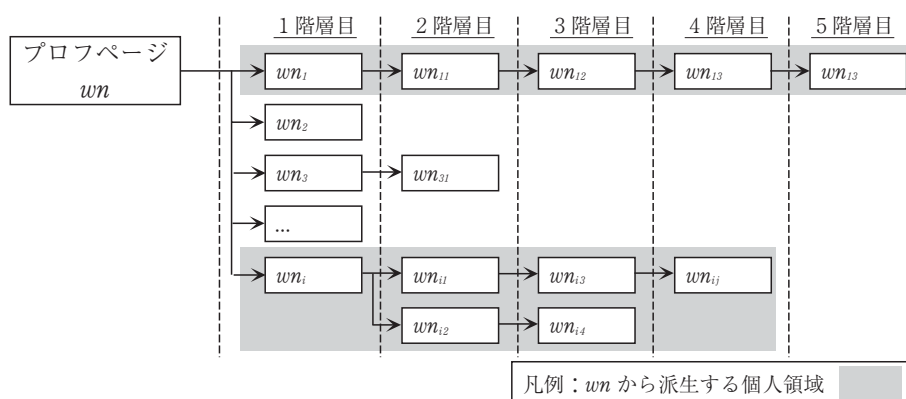


図8 個人領域の判定処理

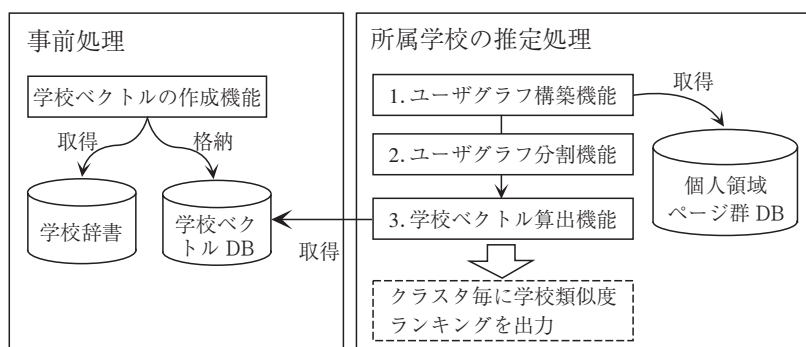


図9 学校とユーザの関連度算出の流れ

合をプロフユーザ n の個人領域 PD_n として定義する。

6. 学校とユーザの関連度算出手法

6.1 学校とユーザの関連度算出の流れ

本章では、プロフクロウで収集したプロフユーザと個人領域クロウで収集したプロフユーザごとの個人領域を解析し、学校との関連度を算出する学校とユーザの関連度算出手法について解説する。学校とユーザの関連度算出手法の流れを図9に示す。

本処理では、まず、事前に学校ベクトル作成機能により国土交通省の国土数値情報（公共施設データ）から学校ベクトルを作成する。次に、ユーザグラフ構築機能によりプロフクロウで収集したプロフユーザを解析してユーザグラフを構築する。そして、ユーザグラフ分割機能によりプロフユーザ間の関連度からグラフを分割することでプロフユーザのクラスタを作成する。最後に、学校ベクトル算出機能によりクラスタごとにクラスタを代表するクラスタベクトルを作成し、学校ベクトル作成機能で作成した学校ベクトルと比較することによりクラスタと各学校の間で類似度を算出する。このことにより、学校に關す

る個人情報を記述していないユーザであったとしても、学校に関連する生徒の一覧を取得できると考えられる。

6.2 Web ページクラスタリング手法の選定

Web ページをクラスタリングする関連研究では、文書の特徴を解析してクラスタリングする手法とネットワーク構造の特徴を解析してクラスタリングする手法が提案されている。それぞれの手法の特徴を述べ、本研究で採用する手法を決定する。

文書の特徴を解析してクラスタリングする手法には、教師有りクラスタリングによる手法と教師無しクラスタリングによる手法がある。教師有りクラスタリングの手法では、SVM (Support Vector Machine)²¹⁾ に基づく方法や Naïve Bayes Classifier²²⁾ を用いた方法などが提案されている。これらの手法では、あらかじめ教師データを入力し分類するクラスタごとの特徴を学習させ、分類器を作成する。そして、この分類器に新たな文書を入力することで、入力文書の所属するクラスタを判別する。しかし、プロフィールは、記述内容の更新頻度が高く記述内容のトレンドも短期間で変化するため、分類器の更新を頻繁に行う必要があるという問題点がある。また、隠語や略語が多用されるため、出現単語の特徴に基づく辞書的な手法は適用が難しい。一方、教師無しクラスタリングに基づく手法では、SOM (Self-Organizing Map)²³⁾ による方法²⁴⁾ や k-means²⁵⁾ による方法^{26), 27)} などが提案されている。これらの手法は、分類器の更新は必要無いが、プロフィールには簡潔に記述された短い情報が掲載されるため、クラスタリングに用いる特徴量が十分に取得できないという問題がある。また、プロフィールは、現実の人間関係を基にクラスタを形成しており、プロフィール間のみでなく、現実の人間関係を考慮する必要であると考えられる。

ネットワーク構造を解析してクラスタリングする手法では、HITS アルゴリズム^{28), 29)} に基づく手法³⁰⁾ や MCL (Markov Clustering Algorithm)³¹⁾ に基づく手法³²⁾ などが提案されている。これらの手法では、Web ページの内容ではなくリンク構造やリンクの強さに基づき Web ページをクラスタリングする。

本研究で対象とするデータは、Web ページ間の関係から人間関係を意味するネットワークを生成している。本研究では、人間関係の強度（親密度）が高ければ同様のクラスタに分類されるようにネットワークを分割することが目的であるため、MCL を採用する。

6.3 本研究の手法

本提案手法では、学校ベクトルとユーザ間の類似度を算出するために、学校ベクトル作成処理、ユーザグラフ構築処理、ユーザグラフ分割処理と類似度算出処理の4つの処理を行う。

6.3.1 学校ベクトル算出処理

本処理では、学校ごとに住所と学校名から単語を抽出し、その単語ごとの TF/IDF 値³³⁾ から学校ベクトルを作成する。まず、国土交通省の国土数値情報（公共施設データ）から小

学校、中学校、高等学校、大学の集合 $S=(s_1, s_2, \dots, s_k)$ を抽出する。次に、学校ごとに学校名と住所から語句の集合 $SWk=(swk_1, swk_2, \dots, swk_l)$ とその個数 $SWCk=(swck_1, swck_2, \dots, swck_l)$ を抽出する。最後に、対象学校における語句の特徴量 $TFIDF(sw_{il})$ を式(2)-(4)にて算出する。ここで、学校毎の特徴ベクトルを $SWWk=(swwk_1, swwk_2, \dots, swwk_l)$ とする。式(4)において、 df は、任意の語句が含まれる学校の数を目指す。

$$TFIDF(sw_{il}) = TF(sw_{il}) * IDF(sw_{il}) \quad (2)$$

$$TF(sw_{il}) = \frac{swck_l}{\left(\sum_{m=1}^{|SWk|} swck_m \right)} \quad (3)$$

$$IDF(sw_{il}) = \frac{\log(|S|)}{(df(sw_{il}))} \quad (4)$$

6.3.2 ユーザグラフ構築処理

ユーザグラフ構築処理は、次に示す2つの処理で実現する。

(1) グラフの構築

本ステップでは、プロフクローラが収集したプロフユーザをノード、プロフユーザ間の繋がりをエッジとしたユーザグラフを構築する。まず、プロフユーザDBから x 人のプロフユーザ集合 N を取得する。ここで、 N に含まれる任意のプロフユーザ n のプロフページ wn 、プロフユーザの個人領域 PDn もユーザと関連付けて取得する。次に、各プロフユーザの個人領域 PDn に含まれる Web ページ集合を解析する。このとき、任意のユーザの個人領域間でハイパーリンクが存在する場合は、それらのユーザ間で関連付けを行い、ユーザグラフ G を構築する。ユーザグラフ構築の手順を次に示す。

Step. 1: ノード集合 $Node=\varphi$ とエッジ集合 $Edge=\varphi$ を作成

Step. 2: 全 n をノード $Node$ に追加

Step. 3: $node_p$ に対応する PD_p からハイパーリンクが張られている PD_q に対応する $node_q$ を取得し、 $node_p$ から $node_q$ へのエッジ $edge_{pq}$ を $Edge$ に追加

Step. 4: Step. 3 を $p=0$ から $p=x$ まで繰り返し実行

(2) 関連強度の算出

本ステップでは、ユーザ間のハイパーリンク関連強度と自然言語リンク関連強度の調和平均からユーザ間の関連強度を求め、ユーザグラフのエッジの重みとする。まず、任意のユーザ p と q 間において PDp から PDq へのハイパーリンクの数を $countlink(p, q)$ とし p と q 間のハイパーリンク関連強度とする。次に、 PDp から PDq へのハイパーリンクに設定されたテキストを PDp から PDq へのニックネームとして PDp から抽出する。そして、 PDq へのニックネームが、 PDp 中に出現する回数を自然言語リンク関連強度

	node1	■	■	■	■	nodex
node1	LW(1,1)	■	■	■	■	LW(1,x)
■	■	■				■
■	■		■			■
■	■			■		■
■	■				■	■
nodex	LW(x,1)	■	■	■	■	LW(x,x)

図10 ユーザ間の関係行列

$countnlplink(p, q)$ として算出する。最後に、 p と q 間の関連強度 $LW(p, q)$ を式(5)により算出し、ユーザー p と q 間のエッジの重みとする。式(5)において、 R は E 尺度の重みであり、ハイパーリンク関連強度と自然言語関連強度のどちらを重視するかを示す値である。

$$LW(p, q) = \frac{(1 + R^2)}{\frac{R^2}{countlink(p, q)} + \frac{1}{countnlplink(p, q)}} \quad (5)$$

(3) ユーザグラフ分割処理

本処理では、既存手法³²⁾に倣い、ユーザグラフ構築処理によって構築したグラフ G を MCL にて分割する。グラフ G を MCL を用いて分割する処理の流れを次に示す。まず、行列を $Node$ 、値を $Node$ 間のエッジの重みとした x 次正方行列 M_G (図10) を作成する。

この時、エッジが存在しない $Node$ 間のエッジの重みは 0 とする。次に、MCL のフローに基づき関係行列の値が収束するまで計算を行う。MCL のフローを次に示す。

Step. 1: グラフ G の全てのノードにノード自身へのエッジを追加

Step. 2: $M = M_G$

Step. 3: $M = M^2$

Step. 4: $M = \Gamma_r M$

Step. 5: $M^2 \neq M$ の間 Step. 3, 4 を繰り返し実行

MCLでは、expansion と inflation の2つの演算を交互に実行することで、グラフを分割

する。行列 M に対する expansion 演算は、 $M=M^2$ で与えられる。また、inflation 演算 $M=\Gamma_r M$ は、式(6)で定義される。なお、 r は inflation パラメータと呼ばれ $r>1$ となる。

$$(\Gamma_r M)_{pq} = \frac{(M_{pq})^r}{\sum_{t=1}^x (M_{tq})^r} \quad (6)$$

そして、行列 M に対応するグラフを H とし、 H の連結成分を1つのクラスタとして、 y プロフユーザのクラスタ集合 $C=(c_1, c_2, \dots, c_y)$ を生成する。

(4) 類似度算出処理

本処理では、ユーザグラフ分割処理によって取得したユーザのクラスタ Cy ごとにクラスタベクトルを作成し、学校ベクトルと比較することでクラスタに所属するユーザと学校間の類似度を算出する。まず、クラスタ内の全ユーザのプロフページから語句の集合 $CWy=(cwy_1, cwy_2, \dots, cwy_z)$ を抽出する。次に、語句の重みを $TF(cwy_z)$ としたクラスタベクトル $CWWy=(cwwy_1, cwwy_2, \dots, cwwy_z)$ を作成する。そして、類似度を算出するために、 $CWWy$ と $SWWk$ に含まれる語句を次元として、同一次元数のベクトルを作成する。ここで、作成されたベクトルの次元数を o とする。また、それぞれのベクトル間を比較した際に一方に語句が存在しない場合はその要素の値を0として補完する。最後に、クラスタベクトルと学校ベクトル算出処理で作成した学校ベクトルから式(7)に示すベクトル空間モデルのコサイン尺度により類似度 $Sim(CWy, SWk)$ を算出する。

$$Sim(CWy, SWk) = \frac{\sum_{h=1}^o (cwwk_h \times swwy_h)}{\sqrt{\sum_{h=1}^o (cwwy_h)^2} \times \sqrt{\sum_{h=1}^o (swwk_h)^2}} \quad (7)$$

7. ネットパトロール支援システムの評価実験

7.1 実験目的

本章では、ネットパトロール支援システムを構築する際の「個人領域の特定問題」と「生徒情報の検索に関する問題」の2つの課題について解消できているかを評価する。その後、本システムで解析したプロフから広がる人間関係をソーシャルグラフとして出力し、本研究で構築したネットパトロール支援システムの有用性を確認する。

7.2 個人領域の特定に関する評価実験

7.2.1 実験内容

個人領域の特定に関する評価実験では、本研究で提案した個人領域の特定方法にて、ネットパトロール支援システムの構築時の一つの課題である「個人領域の特定問題」が解消されているかどうかを評価する。本実験では、4つのプロフサービスからランダムで抽出したユーザ30名を対象として、個人領域クローラを実行し、収集したURLについて評価した。実験の評価には、クローラ対象となる120名のユーザが利用するサービスを人手で調

表2 個人領域の特定手法の評価実験用データ

プロフサービス名	プロフサービス名サービス利用者数					
	ホームペ	りある	ゲスブ	ブログ	アルバム	リンク
前略プロフィール	7	28	27	9	5	4
魔法のプロフ	1	0	0	3	0	0
ヤッピープロフ	2	9	14	2	4	0
プロフプロフ	4	5	2	0	0	0

表3 個人領域の特定手法の実験結果

プロフサービス名	サービスの抽出結果（抽出件数/サービス利用者数）						合計 (割合)	不要 URL 数
	ホームペ	りある	ゲスブ	ブログ	アルバム	リンク		
前略プロフィール	7/7	28/28	26/27	9/9	4/5	4/4	78/80 (0.98)	1
魔法のプロフ	1/1	0	0	3/3	0	0	4/4 (1.00)	1
ヤッピープロフ	2/2	9/9	14/14	2/2	4/4	0	31/31 (1.00)	2
プロフプロフ	4/4	5/5	2/2	0	0	0	10/11 (0.91)	0

査し、それらのサービスを正解データとしてシステムの出力結果と比較した。また、抽出精度の評価において、対象サービスの URL を取得できていることに加えて、広告などの不要なリンクが正確に除去されていることも併せて確認する。本実験で利用するテストデータを表2に示す。表2では、プロフからリンクされるサービスとしてホームペ、りある、ゲスブ、ブログ、アルバムとリンクを採用した。

7.2.2 実験結果と考察

個人領域の特定に関する評価実験の結果を表3に示す。表3を確認すると、次に示す2つの内容が実現できていることが分かる。

- 特定の学校の生徒を高精度に取得可能

表3の結果から、プロフユーザが利用するサービスの URL を9割以上の確率で取得できていることが分かる。このことから、個人が利用するサービスの URL を網羅的に収集できていることが分かる。

- 不要な URL を除去した効率的な収集が可能

表3の不要 URL 数の結果から、ユーザ領域に含まれていた不要 URL 数が非常に少ない

ことが分かる。この結果は、プロフィールに表示される様々な広告リンクやアフィリエイトサイトへのリンクを誤抽出することがなく、クロール対象の個人に関するページのみを効率よく収集できていることが分かる。このことにより、他サイトを経由したユーザ間の関係を誤抽出する問題が発生する確率を低下させ、高精度なユーザ間の関係を抽出することが可能となる。

これらの結果から、個人領域クローラは、クロール対象となるプロフィールユーザに関連するサービスの領域のみを抽出可能であることが分かる。このことにより、ネットパトロールを行う上で重要となる次の3つの機能を実現することができる。

- プロフィールユーザが利用するサービスの変化を判断可能

個人領域クローラでは、日々更新される広告リンクやアフィリエイトサイトへのリンクを選別し、個人に関するサービスのみを抽出することが可能である。そのため、個人領域に含まれるページのURLを監視することで、プロフィールユーザが利用するサービスの変化を迅速に発見することが可能である。このことにより、更新されたユーザを重点的に調査することが可能となり、効率の良いネットパトロールを実現できる。

- タイムスタンプとともに個人領域を確保可能

個人領域クローラでは、個人領域を正確に特定して情報を収集することが可能である。そのため、クロールした日付とクロール結果を対にして保存することで、個人領域内での情報の変化を効率的に把握することができる。特に、ゲスプでは、追加・更新や削除が日々繰り返して行われている。それに対して、本システムを利用することで、クロール時点の情報の解析結果が確認可能となり、過去にさかのぼって状況を確認することができる。

- 新たな人間関係の変化を時系列で監視可能

本システムでは、個人領域を特定した後に、個人特性解析システムにて個人領域間の関係からそのプロフィールユーザの交友関係を解析する。この関係の広がりを継続的に監視することにより、新たな人間関係の変化を把握することが可能となる。そのため、ネットパトロールにて、新たに追加された人間関係を重点的に確認することで、問題の早期発見を支援することが可能となる。

7.3 生徒情報の検索精度に関する評価実験

7.3.1 実験内容

生徒情報の検索精度に関する評価実験では、本研究で提案した学校とユーザの関連度算出手法にて算出した結果を用いて生徒情報の検索を行うことで、ネットパトロール支援システムの構築時の一つの課題である「生徒情報の検索に関する問題」が解消されているかどうかを評価する。本実験では、事前に準備したテストデータセット内から特定の学校の生徒を正確に抽出できるかを評価する。

本実験で採用するテストデータは、次に示す4つの手順で作成した。

表4 テストデータセットに含まれる生徒の学校名の一覧

学校名	数	学校名	数	学校名	数
都立城東高校	29	都立葛飾総合高等学校	2	都立江東商業高等学校	1
都立橘高等学校	11	都立向丘高等学校	2	都立荒川工業高等学校	1
都立葛飾野高等学校	8	都立小岩高等学校	2	都立三田高等学校	1
都立江戸川高等学校	6	都立深川高等学校	2	都立青山高等学校	1
都立芝商業高等学校	5	都立南葛飾高等学校	2	都立浅草高等学校	1
都立青井高等学校	5	都立墨田川高等学校	2	都立足立工業高等学校	1
都立日比谷高等学校	5	東京成徳学園	2	都立足立東高等学校	1
関東第一高等学校	4	京華女子高等学校	1	都立大崎高等学校	1
都立上野高等学校	4	駒場学園高等学校	1	都立日本橋高等学校	1
新渡戸文化短期大学	3	県立城東高校	1	都立飛鳥高等学校	1
都立篠崎高等学校	3	県立城南高等学校	1	都立豊島高等学校	1
都立足立高等学校	3	三郷北高等学校	1	東京農業大学短期大学部	1
都立足立西高等学校	3	修徳高等学校	1	八尾高等学校	1
鹿島学園高等学校	2	昭和学院高等学校	1	遊学館高等学校	1
都立一橋高等学校	2	村田女子高等学校	1		
都立葛飾商業高等学校	2	都立戸山高等学校	1		

Step. 1：前略プロフィール内の「城東高校」のキーワードを含むプロフユーザを無作為に10名取得する。テストデータの作成において「城東高校」を採用した理由は、城東高校の略称で呼ばれる高校として、「徳島県立城東高等学校」、「福岡工業大学附属城東高等学校」、「岡山県立岡山城東高等学校」と「東京都立城東高等学校」の4校が存在し、それらの高校を区別して正確に検索できるかを評価するためである。また、プロフサービスとして前略プロフィールを採用した理由は、個人領域の特定に関する評価実験から利用者が多様なサービスを利用しており、多くのプロフユーザ間の関連が取得できると考えたためである。

Step. 2：取得した10名を起点として個人情報クロウラを実行し、各ユーザの個人領域を取得する。個人領域の取得時に、各プロフユーザの個人領域内にプロフユーザへのリンクが含まれる場合は、そのユーザもテストデータセットとして追加する。ここで、追加するユーザの深さは、起点ユーザから3階層以内とする。その結果、得られた204名を本実験のテストデータセットとする。

Step. 3：個人領域の解析を行い204名のユーザ間の関連度を算出する。

Step. 4：検索精度を評価するため、個人領域に含まれるページを人手で解析し、204名の所属学校を推定する。人手で解析する際には、プロフに記述されている単語以外に、ホームページ、リンクページのタイトル、グープのコメントや画像内の文字列を参照した。テストデータセットに含まれる生徒の学校名の一覧を表4に示す。

表4を確認すると、テストデータセットを作成するための起点となった「城東高校」に

表5 生徒情報の検索精度に関する評価実験の結果

	通 常						不明ユーザ削除後の精度				
採用順位	1	10	20	30	40	50	1	10	20	30	40
結果件数	16	22	22	65	79	84	14	19	19	48	61
正解件数	13	17	17	25	26	27	13	17	17	25	26
クラスタ数 (全体46)	3	5	5	10	11	12	3	5	5	10	11
適合率	0.81	0.77	0.77	0.39	0.33	0.32	0.93	0.90	0.90	0.52	0.43
再現率	0.45	0.59	0.59	0.86	0.90	0.93	0.45	0.59	0.59	0.86	0.90
F 値	0.58	0.67	0.67	0.53	0.48	0.48	0.60	0.71	0.71	0.65	0.58

関連するユーザが30名（都立城東高等学校29名，県立城東高等学校1名）含まれていることが分かる。

本実験では，提案手法を用いて表4に示すテストデータから最も件数の多い「都立城東高校」に関する生徒のみを抽出する場合の検索精度について評価する。検索精度の評価には，一般的に検索エンジンの精度評価に利用されるF尺度を採用した。

7.3.2 実験結果と考察

生徒情報の検索精度に関する評価実験の結果を表5に示す。

表5では，各クラスタの学校所属ベクトルの採用順位に応じて，F値を算出した結果を表している。生徒情報の検索精度に関する評価実験の結果を確認すると，次に示す5つの内容が実現できていることが分かる。

- 特定の学校の生徒を高精度に取得可能

表5を確認すると，採用順位が20位以上の場合は，不明のユーザを含めた場合は適合率が0.77（検索結果22件中17名が正解），不明を抜いた場合は0.90（検索結果19件中17名が正解）となり，非常に高い適合率を実現できていることが分かる。これは，クラスタ内に特定の学校の生徒が多く含まれている場合に，クラスタの学校への所属確率が高くなることを表している。

- 特定の学校の生徒を網羅的に取得可能

表5を確認すると，採用順位が30位以上で再現率が0.86（正解データ29名中25名を抽出），40位以上で0.90（正解データ29名中26名を抽出），50位以上で0.93（正解データ29名中27名を抽出）となり，非常に高い再現率を実現できていることが分かる。これは，クラスタ内に他の学校の生徒が多く含まれているが，その学校の生徒も含まれていることを表している。実際に，検索にかかった12件のクラスタの内容を確認したところ，いずれのクラス

表6 城東高校の生徒29名の情報取得元

収集元	プロフィール	ホームページ	画像	ゲスブ
件数	24	3	1	1

タにおいても正解データが含まれていることが分かった。

- 略称が同様に正式な学校名が異なる学校を除外可能

本実験で利用したテストデータには、表4の通り略称が同様となる「都立城東高等学校」と「県立城東高等学校」の学校の生徒が含まれている。「県立城東高等学校」の生徒が含まれているクラスタについて所属学校の推定順位を確認すると、「県立城東高等学校」が3位となっており、「都立城東高等学校」は50位以内に含まれていない状況であった。これは、「県立城東高等学校」の生徒と関係の深いユーザのプロフページに「徳島＝城南」と記述されていたためであると考えられる。

- プロフに記載されていないユーザを判別可能

本実験では、各ユーザの学校名を手で推定する際に、プロフの記述内容から判別できない場合は、個人領域に含まれる他のサービスに記載されている内容を用いて判断している。城東高校の生徒29名のデータ収集元を表6に示す。表6からプロフ以外に、ホームページやプロフのTOP画面、ゲスブから学校名を推定していることが分かる。

本実験の評価結果である表5において、採用順位50位にした際に検索漏れとなった2名について、取得元を確認したところ、プロフから情報を取得したユーザであった。そのため、本手法では、プロフに記載されていないユーザに対しても、同様のクラスタに分類された他ユーザの情報から所属学校を判定し、検索結果として抽出可能であることが分かる。

- 学校名の表現の多様性に対応可能

学校名の表現方法の多様性を確認するため、本実験で正解データとした城東高校生29名の学校名の表現方法を集計（表7）した。表7の結果を確認すると、城東高校の表現方法は8種類あり十人十色であることが分かる。これらの生徒に対して、本手法では、同様のクラスタに分類された他ユーザの情報から所属学校を判定し、検索結果として抽出可能であることが分かる。

これらの結果から、学校とユーザの関連度算出手法にて算出した結果に基づく検索方法は、特定の学校の生徒を効率的に検索できていることが分かる。本システムで採用した検索方法では、利用用途に応じて学校ベクトルの採用順位を変更することで、ネットパトロールを支援することが可能である。具体的には、飲酒や喫煙などの非行・逸脱行為を行っている自校の生徒を検索する場合は、対象を絞り込むために自校の採用順位を上位に設定する。また、他校の生徒との繋がりや危険な関係が含まれていないかを確認する場合には、自校の生徒と関連する可能性のある生徒を検索するために採用順位を下位に設定する。このように、パトロールの目的に応じて採用順位を指定可能とすることで、関連する生徒を

表7 城東高校の表現方法

記述名称	件数
城東高校	9
都立城東高校	8
城東生	4
城東	4
(都) 城東高校	1
Johtoh	1
JOHTOH high school	1
都立 J 高校	1
城東高校	9

効率よく見守ることが可能となると考えられる。

7.4 ソーシャルグラフの評価実験

7.4.1 実験内容

ソーシャルグラフの評価実験では、プロフサービスから女子中学生、女子高生を示す隠語を入力し、その結果得られたプロフユーザを対象にソーシャルグラフの構築実験を行った。本実験により、プロフから広がるネット上での人間関係を容易に把握することが可能であり、ネットパトロールの支援に役立つと考えられる。

7.4.2 実験結果と考察

ソーシャルグラフの構築実験の評価では、特徴的な結果が見られたユーザに焦点を当てて考察する。女子高校生と女子中学生の各ユーザから得られた3階層のソーシャルグラフをそれぞれ図11、図12に示す。ここで、本実験で表示するソーシャルグラフでは、実際のネットパトロール支援システムとは異なり、各ノードのラベルにシステムIDを採用する。ソーシャルグラフ図11と図12を確認すると、次に示す3つの内容が確認できる。

● 学外の大人との関係を把握可能

図11のAの領域では、起点ユーザ14724と14273がリンク関係にあり、14273を中心として、4つのグループに分類されていることが分かる。14273の情報を解析すると1988年生まれの男のプロフであることがあり、現在22歳であることが分かった。14273は、14724以外にも、14275、14276、14277とも連絡を取っており、特に14277と深い関係にあることが分かった。14273と関連する4名のうち3人は高校生であり、ゲスブの内容を確認すると出会い目的で高校生向けに連絡を取っていることが分かった。

● 非行・逸脱行為を行っている友人との関係を把握可能

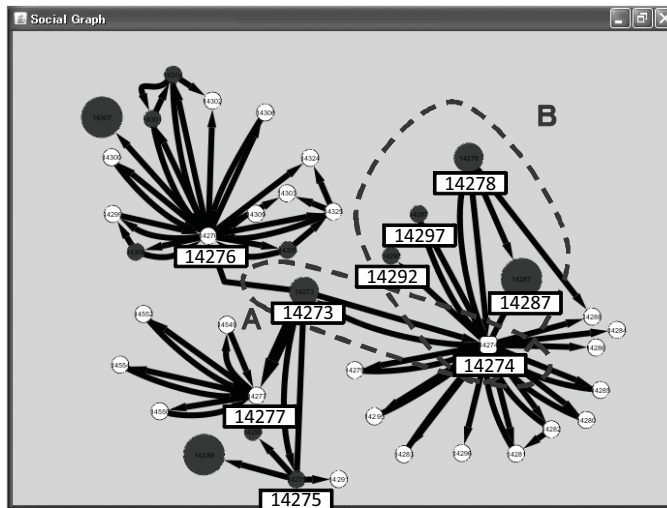


図11 女子高校生のソーシャルグラフ

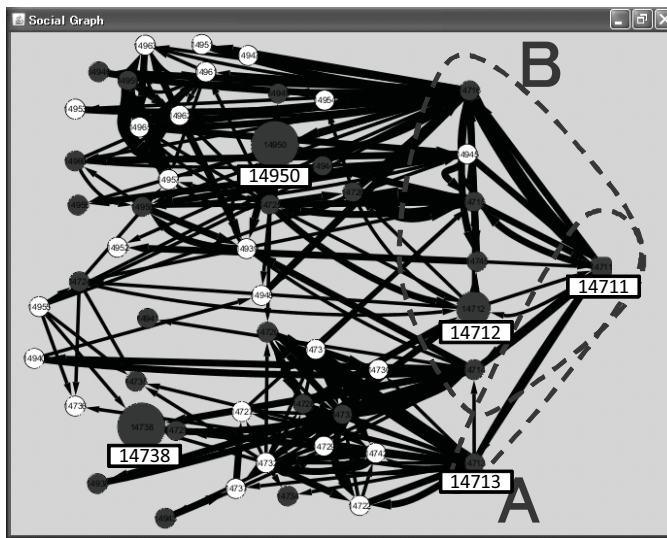


図12 女子中学生のソーシャルグラフ

図11のBの領域では、起点ユーザ14274の友人関係において、14292、14297、14278と14287が非行・逸脱行為を行っていることが分かる。特に、14278と14287は、ノードサイズが大きくなっていることからわかるように、非行・逸脱用語辞書との一致する件数が多くなっている。

また、図12は、女子中学生（14711）から構築したユーザ関係であるが、非行・逸脱行為を行っているユーザが全体の約6割に達していることが分かる。特に、14712、14950、14738は、ノードのサイズが大きくなっており、非行・逸脱用語辞書との一致する件数が多くなっていることが分かる。実際に、それらのユーザのプロフを確認すると、飲酒や喫

煙行為を行っていることが分かった。

- 現実社会の友人関係を把握可能

図12のAの領域では、起点ノード14711と14713の関係を表している。この関係は、恋愛関係にあり、エッジの太さが比較的大きくなっていることが分かる。また、図12のBの領域では、地元の友人関係が集まっており、共有する時間が長いことから、関連度が高くなっていることが分かる。

このように、本システムを利用することにより、起点ユーザ14724と14711から広がるプロフ上の関係を自動的に解析し、出会い目的などの「危険な人間関係」、「非行・逸脱行為を行っている友人との人間関係」や「現実社会の友人関係」を確認することが可能となる。このことにより、りあるやゲスプなどの様々なサービスを横断的に確認する必要がなく、解析対象のユーザの行動や人間関係を俯瞰でき、ネットパトロールの効率化に寄与できると考えられる。

8. お わ り に

本研究では、モバイルインターネットの中でも、中高生の利用率が高いプロフサービスを起点として、そのサービスから広がる子供たちの行動を継続的に監視するネットパトロール支援システムを開発した。本システムでは、プロフサービスに登録しているユーザをURLの特性に基づき一意に識別し、そのユーザ単位で非行・逸脱行為の有無や個々人の人間関係を解析する仕組みを実現している。さらに、ネットパトロールでは、事件の未然防止や生活指導を行うために、実際の現場と関連付ける情報が必要不可欠となることから、ユーザと学校間の関連度に基づく検索機能を提供し、自校の生徒を効率よく検索する仕組みを実現している。このことにより、現在、人手で行っているネットパトロール作業を効率化させることが可能となり、より精度の高いネットパトロールが可能になると考えられる。

今後は、本研究に関連した内容として、次に示す3つの内容に取り組む予定である。

- 実現場への適用とフィードバックによるシステムの改善

本システムは、現状の教育現場におけるニーズを踏まえた上で開発したシステムである。しかし、その際にヒアリングしたニーズは、人手でネットパトロールを行っている際の課題点であり、その課題を解決した後にも様々な課題や問題が浮き彫りになると想定される。そのため、今後は、多くの教育現場へ適用した後に、どのような改善が必要かのヒアリングを行い、その結果をフィードバックしたより有用なシステムへ改善する予定である。

- 違法・有害情報の判別精度の向上

本システムでは、非行・逸脱行為の判別において、辞書との一致による判定を行っている。そのため、違法・有害情報の有無の判定において、フィルタリングの技術を適用し、違法・有害情報の判別精度を向上する予定である。

● 有害な人間関係の推定

本システムで構築したソーシャルグラフは、解析対象のユーザと関係が深い人物を出力する。しかし、出力したユーザ間の詳細な関係は、人手で解析する必要がある。そのため、出会い系目的の関係やネットいじめの関係を把握するためには、ユーザの個人領域に含まれるページを確認する必要がある非効率である。そのため、今後は、記述内容やネットワークの構造に基づき人間関係の詳細を推定する研究を行う予定である。

謝辞

本研究の一部は、平成21～22年度 JST RISTEX「犯罪からの子どもの安全」研究開発領域研究開発プログラム「犯罪からの子どもの安全」(研究課題「子どもの犯罪に関わる電子掲示板記事の収集・監視手法の検討」)、平成20～24年度私立大学戦略的研究基盤形成支援事業(研究課題「セキュアライフ創出のための安全知循環ネットワークに関する研究」)から助成を受け、その成果を公表するものである。

参考文献

- 1) 文部科学省：子どもの携帯電話等の利用に関する調査（2009）
- 2) 下田博次：思春期メディアとしての携帯電話の利用問題，日本社会情報学会学会誌，日本社会情報学会，Vol.20, No.1, pp.11-25（2008）
- 3) 藤川大祐：学校・家庭でできるメディアリテラシー教育(4) 見せかけだけの優しい大人-児童ポルノ・児童売春とネット，児童心理，金子書房，Vol.63, No.5, pp.127-133（2009）
- 4) 小針誠：学校裏サイトにおける「ネットいじめ」の構造と対策，児童心理，金子書房，Vol.63, No.13, pp.94-99（2009）
- 5) 今一生：「ブログ」「プロフ」の中でのケンカとトラブル，児童心理，金子書房，Vol.63, No.13, pp.110-114（2009）
- 6) 石川一：高校生向けケータイ依存予防教育プログラムの開発的研究，生徒指導学研究，日本生徒指導学会，Vol.7, pp.142-152（2008）
- 7) 群馬大学社会情報学研究センター：青少年の携帯電話等の利用問題に関する全国指定都市・都道府県教育委員会調査結果（2010）
- 8) 東京都教育委員会：子供のインターネット・携帯電話利用についての実態調査（2008）
- 9) 下田博次：学校裏サイト，東洋経済新報社（2008）
- 10) モバイルコンテンツ審査・運用監視機構：<http://www.ema.or.jp/>
- 11) 総務省：平成20年度電子通信サービスモニターに対する第2回アンケート調査結果〔報告書〕（2009）
- 12) 札幌市教育委員会：平成22年度教育委員会実施プラン（2010）
- 13) 宮崎県学校教育改革推進協議会：平成20年度宮崎県学校教育改革推進協議会〔協議のまとめ〕（2008）
- 14) 北海道教育委員会：教育行政執行方針（2010）
- 15) 和歌山県教育委員会：平成21年度教育委員会事務の点検及び評価報告書（2009）
- 16) ビットクルー：<http://pit-crew.co.jp/>
- 17) ガイアックス：<http://www.gaiax.co.jp/>

- 18) 下田博次：子どものインターネット利用問題解決のための社会システム開発——地域，地方からの解決を——，産学官連携ジャーナル，科学技術振興機構，Vol.5, No.8, pp.29-31 (2009)
- 19) Cytoscape：http://www.cytoscape.org/
- 20) 池辺正典，田中成典，古田均，中村健二，小林建太：Web リンク構造解析と自然言語処理による組織関係の抽出についての研究，情報処理学会論文誌，情報処理学会，Vol.47, No.6, pp.1687-1695 (2006)
- 21) Cortes, C. and Vapnik, V.: Support Vector Networks, Machine Learning, Springer Netherlands, Vol.20, pp.273-297 (1995)
- 22) Bishop, T.: Pattern Recognition and Machine Learning, Springer (2006)
- 23) Kohonen, T.: Self-Organizing Maps, Springer Series in Information Sciences 30, Springer (2000)
- 24) Mahalakshmi, B. and Duraiswamy, K.: Self-organizing Map for Document Categorization Using Latent Semantic Analysis, International Conference on Innovative Computing technologies, IEEE, pp.1-6 (2010)
- 25) MacQueen, J.: Some Methods for Classification and Analysis of Multivariate Observations, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Vol.1, pp.281-297 (1967)
- 26) Gao, Z. and Jin, N.: Detecting Community Structure in Complex Networks Based on K-means Clustering and Data Field Theory, Control and Decision Conference 2008, IEEE, pp.4411-4416 (2008)
- 27) Dai, W., Jiao, C. and He, T.: Research of K-means Clustering Method Based on Parallel Genetic Algorithm, Intelligent Information Hiding and Multimedia Signal Processing 2007, IEEE, Vol.2, pp.158-161 (2007)
- 28) Kleinberg, J.: Authoritative Sources in a Hyperlinked Environment, Journal of the ACM, ACM, Vol.46, No.5, pp.604-632 (1998)
- 29) Gibson, D., Kleinberg, J. and Raghavan, P.: Inferring Web Communities from Link Topology, Proceedings of 9th ACM Conference on Hypertext and Hypermedia, ACM, pp.225-234 (1998)
- 30) Kumar, R., Raghavan, P., Rajagopalan, S., and Tomkins, A.: Trawling the Web for Emerging Cyber-communities, Proceedings of 8th International World Wide Web Conference, Computer Networks, pp.403-415 (1999)
- 31) Enright, J., Dongen, S. and Ouzounis, C.: An Efficient Algorithm for Large-scale Detection of Protein Families, Nucleic Acids Research, Oxford University, Vol.30, No.7, pp.1575-1584 (2002)
- 32) 加藤一民，松尾啓志：Markov Cluster Algorithm を用いた Web コミュニティ群の発見手法，自然言語処理研究会研究報告，情報処理学会，Vol.2005, No.22, pp.87-93 (2005)
- 33) Salton, G. and McGill, M.: Introduction to Modern Information Retrieval, McGraw-Hill Book Company (1984)