

O-015

プロフサイト内の人間関係解析によるネットパトロール支援に関する研究 Net Patrol Support by Analyzing Human Relations in Profile Site

荒川 裕子†
Yuko ARAKAWA

中村 健二‡
Kenji NAKAMURA

小柳 滋‡
Shigeru OYANAGI

1 はじめに

近年携帯電話利用者が低年齢化しており、中高生がモバイルインターネットに端を発した犯罪に巻き込まれるケースが頻発している。モバイルインターネットには様々なサービスがあり、中でも、中高生を中心としたプロフ遊びが盛んにおこなわれている。また、プロフに起因した問題 [1] も懸念されており、非行逸脱行為の誘発やネットいじめなど様々な問題 [2] が顕在化している。

現在、これらの問題への対策として、教育現場ではネットパトロールを実施している。ネットパトロールとは、インターネット上に存在するプロフィールサイト（以下、「プロフ」と略記）や掲示板、SNS (Social Network Service) などを見回り、問題行動を予防する活動である。しかし、ネットパトロールは一般的に人手での作業が中心であり、インターネット上の情報を網羅的に監視するには多大な負担がかかる。

そこで、本研究ではネットパトロール対象者、特に非行逸脱行為を行っているユーザ（以下、「非行者」と略記）の特徴を解析し、非行者を効率的に発見する手法を考察する。

2 研究の位置づけ

著者らは、ネットパトロール支援に関する研究プロジェクトを立ち上げ、インターネット上を自動的に探索し、継続的に監視するシステムの開発に取り組んでいる。その研究では、非行者を探索する方法として、非行に関するキーワードを登録した非行逸脱行為判定辞書を用いて、非行者を自動的に判別する処理を行っている。図1にシステムの全体図を示す。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からプロフユーザを網羅的に収集し、データベースに蓄積する。次に、個人領域クローラが、蓄積されたプロフユーザ全てに対して、プロフページやリンク情報を収集し、各プロフユーザを中心としたネットワークをそれぞれ抽出する。そして、プロフページを用いて非行逸脱行為判定機能で非行者を判別する。最後に、人間関係解析機能にて人間関係を解析し、継続的にクロールする際の優先順位を決定する。そして、優先順位に応じて個人領域クローラが継続的にユーザ情報をクロールする。

本研究では、継続的にクロールする際の優先順位決定に利用可能な指標について調査する。クロールの優先順位を決定することにより、非行者の効率的な探索が可能となり、ネットパトロールの支援につながると考えられる。

3 研究概要

本研究では非行者を効率的に発見するために、継続して監視すべきユーザを判別するための指標について調査する。本研究では非行者とのコミュニケーションが多いユーザは継続して監視すべきであると考え、コミュニケーションの中心人物を検出する手法について

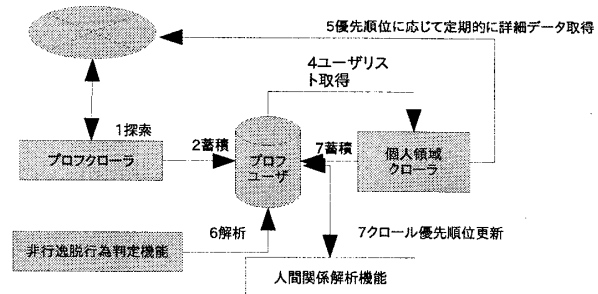


図1: システム構成図

検討する。そして、実験にて中心的人物と非行者との関係を解析し、非行者を効率的に探索する手法を考察する。中心人物の検出には、非行者を含むネットワークの中心性を用いる。

本研究では、中心性の算出において一般的に用いられる4つの指標（次数中心性、距離中心性、媒介中心性、固有ベクトル中心性）[3]により中心人物の抽出を行い、どの指標が適切であるかを考察する。固有ベクトル中心性にはページランク [4] のアルゴリズムを用いた。

4 実験

4.1 実験内容

本実験では、楽天社が運営する「前略プロフィール」[5]を対象とし、前述のシステムにおいて個人領域クローラを用いて、非行者を中心に収集した人間関係ネットワーク（以下、「全体Network」）を用いる。ネットワークの規模はノード数1,696ノード（一般ノード1,584、非行者ノード112）である。

本実験では、非行者とコミュニケーションが多いノードと中心性の高いノードの関係を調査する。実験は以下の手順で行った。

1. 各ノードの隣接ノードを調べ、非行者ノードの含まれる割合 w を算出
2. 全体Networkから非行者ノード以外のノードを省いた非行者Networkを抽出
3. 非行者Networkと全体Networkでそれぞれ中心性を計測
4. 中心性の高い上位 k 個のノードと w の大きい上位30個のノードに共通して含まれるノード数を算出
5. 中心性の順位と w の大きいものからつけた順位の相関を調べる

手順1において、非行者ノードへのリンクの数を見た場合、非行者ノードへのリンクが多くても対象ノードの持つリンクの絶対数が大きければ、非行者とのコミュニケーションは浅いと考えられる。そのため、本研究では隣接ノード中に含まれる非行者ノードの割合が高いノードが非行者とコミュニケーションが多いとし、非

† 立命館大学大学院理工学研究科, Graduated School of Science and Engineering, Ritsumeikan University

‡ 立命館大学, Ritsumeikan University

表1: 上位30ノードに共通するノード

	全体 Network	非行者 Network
出次数	4	28
入次数	14	24
距離	12	29
媒介性	6	23
固有ベクトル	14	24

行者ノードの割合 w を用いる。割合 w は、下の式で算出する。

$$w = \frac{\text{隣接ノード} \cap \text{非行者ノード}}{\text{隣接ノード}} \quad (1)$$

手順2において、一般ノードが含まれる全体 Network では、一般ノードと多く関係するノードの中心性が高くなると考えられる。そのため、一般ノードとの比較のために、今回は非行者 Network を抽出する。

手順4において、中心性が高いノードは非行者とコミュニケーションが多いかを検証するために、中心性の高いノードと割合 w が大きいノードの順位相関を計測した。相関の指標にはスピアマンの順位相関係数を用いる。

4.2 実験結果と考察

(1) 割合 w の上位ノードの分析

割合 w を算出した結果、上位10ノードのうち非行者ノードは8ノード、上位30位のうち非行者ノードは15ノードであった。割合 w が大きい上位30ノードのうち、非行者ではないと判定された15ノードの「ブログ」や「りある」を目視で確認したところ、15ノードのうち8ノードが喫煙や飲酒などの非行行動を確認することができた。また残りの7ノードについて、非行行動の書き込みはみられなかったが、髪色が極端に明るいなど、主観的に非行に走りそうであると判断できるノードが4ノードであった。このことから周囲に非行者が集まっているノードは非行者である確率が高いと考えられる。

これらのノードは、プロフィールページに非行に関する語句が書かれておらず、検出できなかったと考えられる。今回の結果から、非行者判定の際に、非行逸脱行為判定機能と、人間関係解析を併用すると、より検出の精度が上がると考えられる。

(2) 中心性と割合 w の関係の分析

中心性の上位30ノードと割合 w の上位30ノードに共通するノードの出現数を表1に示す。全体 Network では共通するノードの最大値が14であり、半数以下である。それに対し、非行者 Network では全ての中心性で、20以上のノードが共通する結果となった。割合 w と中心性順位の相関を表2に示す。全体 Network と非行者 Network での結果を比較すると、どの中心性においても非行者 Network の方が高い相関が得られている。また非行者 Network での結果の中でも固有ベクトルの結果が最も相関が高く、媒介性の結果は最も相関が低くなっている。

表1の結果より、ネットワークの中心的なユーザはより周囲に非行者が集まりやすいということがわかった。この傾向は、非行者 Network を対象に解析した際に顕著に見られる。このことから、クローリングの優先順位決定において中心性を用いることは妥当であると考えられる。また、表2の結果より、非行者 Network で中心性を計測した結果の中でも、固有ベクトルで計

表2: 割合の高さとの相関

	全体 Network	非行者 Network
出次数	-0.74	0.26
入次数	-0.32	0.45
距離	-0.73	0.26
媒介性	-0.74	0.05
固有ベクトル	0.33	0.63

測した結果は最も相関が高いことがわかる。このことから、中心性の指標には固有ベクトルで中心性を用いることが適切であると考えられる。

5 おわりに

本研究では、ネットパトロールの支援を目的として、効率的に非行者を発見するために、継続して監視すべきユーザの決定に役立つ指標を調査した。著者らは、非行者とコミュニケーションが多いユーザは継続して監視すべきであると考え、コミュニケーションの中心人物の特徴を調査した。

実験の結果、非行者は周囲に非行者が集まりやすく、その中でも、固有ベクトル中心性の高いノードの周囲には非行者が集まることが分かった。さらに、現在の内容解析による非行者の検出手法に、人間関係解析を併用することで、非行者ノードの検出精度が向上する可能性を示し、人間関係解析の有効性を示した。

今後は「前略プロフィール」のサービスの1つである、24時間日記「りある」を解析し、人物間の関係強度を考慮した人間関係解析を行う。そして、日々の行動から非行行為に巻き込まれる危険性の高いユーザを判定する手法について検討する。

謝辞

本研究の一部は、平成21~22年度独立行政法人科学技術振興機構社会技術研究開発センター研究開発プログラム「犯罪からの子供の安全」(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」)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] 下田博次: 学校裏サイト, 東洋経済進報社, 2008
- [2] 下田博次: 子供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, 産学官連携ジャーナル, 科学技術振興機構, Vol.5, No.8, pp.27-29, 2009
- [3] 篠田孝祐, 松尾豊, 中島秀之: 複数の中心性尺度によるネットワーク生成モデル, 知能と情報, 日本知能情報ファジィ学会, Vol.20, No.6, pp.1470-1479, 2005
- [4] Page, L., Brin, S., Motwani, R. and Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. Computer Science Department Technical Report, Stanford University, 1998.
- [5] 前略プロフィール: <http://pr.cgiboy.com/>