

コミュニティサイトにおける危険ユーザ発見手法の検討

深尾 和寿† 中村 健二‡ 小柳 滋†

† 立命館大学情報理工学部

‡ 大阪経済大学情報社会学部

1 はじめに

近年、携帯電話の利用者が低年齢化しており、携帯電話は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。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)のサービスの中でも、自身の趣味や学校、友人関係などの情報を公開するプロフィールサービス(以下、「プロフ」と略記)は、中学生や高校生の間で広く利用されている。しかし、プロフなどのモバイルインターネット上のサービスを介したトラブルも多く発生しており、子供たちが危険にさらされている[1]。実際に、プロフや携帯ゲームサイトなどのコミュニケーション機能がある非出会い系サイトを利用して、児童買春や青少年保護育成条例違反などの事件に巻き込まれた被害児童数は2008年に792名、2009年に1,136名、2010年に1,239名と増加傾向にあり、1つの社会問題となっている。そこで、本研究では任意のユーザが出会い目的で利用している可能性の度合いを算出する手法を開発し利用者に提示することで、コミュニケーションの初期段階で利用者が任意のユーザと直接連絡することの危険性を理解し、事件に発展する可能性が低減されることを目標としている。

2 既存手法

危険な関係に発展する可能性のある出会い目的の書き込みを有害な情報と考えた場合、関連研究として、インターネット上の情報に含まれる特徴に基づき、有害な情報であるかどうかを自動的に判別するための手法があげられる。

2.1 教師あり学習による分類手法

教師あり学習による分類手法[2]では、教師データから有害情報を示す特徴を学習し、その学習した特徴に基づき未知の情報が有害情報である度合いを判別器により算出する。これらの研究では、有害情報を判別するための特徴データとして、形態素を用いる手法[3]、形態素の組み合わせを用いる手法[4]などがある。しかし、危険な関係に関連する出会い目的の危険な書き込みを判定対象とした場合、「書き込み内容から一様に危険かど

うかを判断することは困難」と「継続的に利用可能な教師データを構築することが困難」という2つの課題がある。具体例として、高校生の同性同士が「遊びに行こう」と書いた場合と、女子高生と40代男性が「遊びに行こう」と書いた場合、同様の有害度となる。しかし、一般的に後者の方が危険であると判断できる。そこで著者らは、人と人の属性関係を用いて有害度を判定する研究[1-3]に取り組んでいる。本研究はそれらの取り組みの一つである。

3 提案手法

既存手法では、テキスト情報に基づきユーザの危険性を判断していたが、書き込み内容のみでは精度の高い判定をすることは難しい。そのため、本研究では、ユーザの個人属性とそのユーザを取り巻く人間関係を考慮した危険性の度合いを算出する手法を提案する。本手法の処理フローを図1に示す。図に示す通り3つの機能より構成される。

3.1 プロフページ情報収集機能

コミュニティサイトからユーザの各個人属性、及び人間関係を収集する。収集する個人属性は氏名、URL、年齢、性別、居住地域などとする。

3.2 プロフユーザの属性推定機能

収集したユーザの個人属性を解析して、既存研究[2]に倣い年代と性別を推定する。

3.3 危険なユーザの推定機能

表1に示す各ユーザの個人属性、性別、年齢、人間関係を用いてユーザベクトルを生成する。その後、生成したユーザベクトルを用いて次に示す2つの処理を行う。

3.3.1 教師データ抽出処理

同年代の一般的なユーザの個人属性・人間関係を One-Class-SVM[5]を用いてモデル化した後、それに基づき偏りのあるユーザを抽出する。そして、抽出したユーザを「出会い目的の可能性が高いユーザ」と「その他の目的であるユーザ」に分類し、最後に、分類結果を特徴づける特性を決定し、人間関係の特性をモデル化する。

3.3.2 クラス分類処理

[教師データ抽出処理]で得られたモデルに基づき、出会い目的の可能性の高いユーザを抽出する。そして抽出

Examination of the Dangerous User Discovery Technique in Community site

†Kazutoshi FUKAO ‡Kenji NAKAMURA †Shigeru OYANAGI

†College of Information Science and Engineering, Ritsumeikan University

‡Faculty of Information Technology and Social Science Osaka University of Economics

表 1: ベクトル指標表

属性種別	ベクトル指標
ユーザの基本属性	年齢, 性別
人間関係 (性別)	友人の男女比率
人間関係 (年代+性別)	小学生, 中学生, 高校生, 大学生以上, で定義された各年代毎の男女比率
人間関係 (居住距離)	0-50km, 50-100km, 100-150km, 150-200km, 200km 以上, で定義された各居住距離毎の比率

表 2: 実験結果

実験回数 (回)	1	2
テストデータ数 (件)	500	500
有害ユーザ検出数 (人)	22	19
有害ユーザ数 (人)	9	9
判別不可ユーザ数 (人)	4	3
正答率 [判別不可ユーザを含む](%)	55	64

4.3 実験結果と考察

実験結果を表 2 に示す. 結果として, 判別に有効だと考えられるベクトルには, [異性の友人比率の高さ], [女子高校生の友人比率の高さ], [居住距離が離れている友達比率の高さ] などが挙げられた. また検出されたユーザの中には, なりすましユーザと思われるユーザやサイトからペナルティを受けているユーザも発見できた. しかし, 中にはプロフィールなど目視で確認できる情報が何もないユーザも発見され, 有害であるかの判別は困難な例も見受けられた.

5 おわりに

本稿ではコミュニティサイトにおける危険ユーザ発見手法を提案した. 実験結果より, なりすましと考えられるユーザや実際にペナルティを受けているユーザの発見に成功した. 今後は, ネットいじめ問題などの取り組みに取り組みでいきたい.

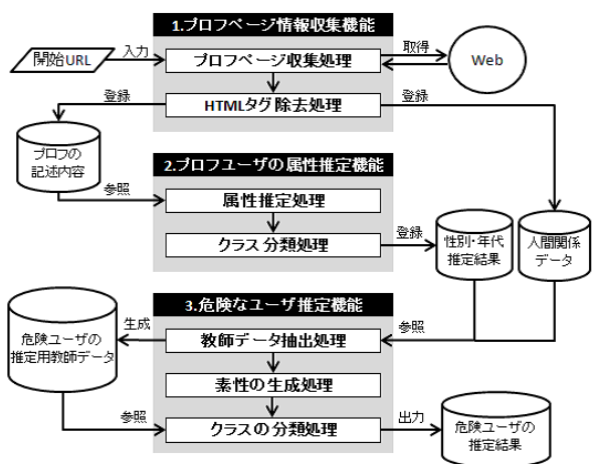


図 1: 研究の処理フロー

結果を解析し, 判別精度を目視で確認する. また, これによって得られた「出会い目的の可能性が高いユーザ」に基づき再びモデル化する作業を繰り返すことで判別精度の向上を目指す.

4 評価実験

4.1 実験内容

本実験では, プロフユーザの属性推定結果を用いて, 出会い目的の危険なユーザを推定可能な判別器を構築し, その判別器を用いて危険なユーザを判定し, その際の判定精度の評価, 繰り返し実験による精度向上, 判別に有効であるベクトル指標の特定を目的としている.

4.2 実験環境

本実験では検証サイトを「大手携帯ゲームサイト」とし, 検証対象は男子高校生と定義する. また, [一般的な高校生] のモデル作成には目視により 500 件のデータを収集し用いる. そして有害ユーザ判別器の精度検証にはランダムに 500 件を抽出して用いる. 尚, それぞれのモデル作成には One-Class-SVM を用いる.

参考文献

- [1] 中村健二: 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, 社会技術研究開発事業研究開発プログラム「犯罪からの子どもの安全」平成 24 年度研究開発実施終了報告書 (2012).
- [2] 井上健治, 中村健二, 小柳滋: 著者属性の推定結果を用いたプロフの出会い目的の書き込み検出のための教師データ自動構築手法, 情報処理学会論文誌: 情報処理学会, Vol.4, No.3, pp.53-69 (2011).
- [3] 荒川裕子, 中村健二, 小柳滋: 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, 第 9 回情報科学技術フォーラム講演論文集, No.4, pp.517-518 (2010).
- [4] 菊池琢弥, 内海彰: 語の共起情報に基づく有害サイトフィルタリング手法, 第 9 回情報科学技術フォーラム講演論文集, 情報処理学会, 電子学会, No.2, pp.1-6 (2010).
- [5] Chih-Chung Chang, Chih-Jen Lin: LIBSVM – A Library for Support Vector Machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>