

平成 29 年 6 月 25 日現在

機関番号：33403

研究種目：若手研究(B)

研究期間：2015～2016

課題番号：15K16543

研究課題名(和文) 携帯機器を対象としたサイバーパトロール支援に関する研究

研究課題名(英文) Research on cyber patrol support for mobile devices

研究代表者

安彦 智史 (ABIKO, SATOSHI)

仁愛大学・人間学部・講師

研究者番号：90560475

交付決定額(研究期間全体)：(直接経費) 2,200,000円

研究成果の概要(和文)：本研究では、携帯機器を対象としたアプリのConsumer Generated Media(CGM)から言語資源を取得し、未成年が巻き込まれる可能性がある危険度の高いアプリの解明を進めた。

スマホアプリ情報収集システムを構築し、2016年8月から11月までの4ヶ月間で47,385個のアプリと217,334件のレビューを収集した。本研究の提案手法を用いることにより、レビュー情報からの危険アプリを発見する手法の有用性の提示した。また、結果からスマホアプリのカテゴリと犯罪の発生場所の関連性の解明、レーティングの効果と問題点の解明に寄与した。

研究成果の概要(英文)：In this research, we acquired language resources from Consumer Generated Media (CGM) for mobile devices, and proceeded with elucidation of highly risky applications that may involve juveniles.

We built a smartphone application information collection system and collected 47,385 applications and 217,334 reviews over the 4 months from August 2016 to November 2016. From the analysis of the collected results, we present the usefulness of the method to discover dangerous applications from review information by using the proposed method of this research. Next, elucidate the relevance between the category of the smartphone application and the place of occurrence of the crime. Finally, I was able to achieve the result of elucidating the rating problem.

研究分野：情報学，子ども学，画像処理

キーワード：サイバーパトロール スマートフォン 有害情報 隠語 サイバー犯罪 レーティング CGM レビュー

1. 研究開始当初の背景

携帯機器の若年層への普及により、未成年を対象としたサイバー犯罪が急増し、その手口の多様化と複雑化が著しい。携帯機器の特徴である音声通話やカメラを用いた出会い系アプリなども出現し、ウェブマイニングやクローリングを中心とした従来の解析手法ではその全貌を明らかにすることが困難である。

2. 研究の目的

本研究では、携帯機器を対象としたアプリの Consumer Generated Media(CGM)から言語資源を取得し、未成年が巻き込まれる可能性がある危険度の高いアプリの解明を進める。これにより、携帯機器を用いたネット犯罪の全貌を明らかにし、サイバーパトロールの効率化を目指した。

3. 研究の方法

(1) 隠語概念化手法の開発とその評価

プライベートチャットアプリケーションの ID 交換を目的とした掲示板(ID 交換掲示板)において違法・有害な情報を含む書き込みが増加傾向にある。ID 交換掲示板では、多様な隠語表現を用いたやり取りが行われており意図的に崩された日本語が多く含まれるため、従来の手法では有害性評価を行うことが困難である。そこで本研究では、ID 交換掲示板における隠語表現を分類し、特に表層的な表記揺れが生じる環境下でも有害性判定を行える手法を検討した。まず、3 つの異なる ID 交換掲示板から 900 件の書き込みを収集し、記事に含まれる隠語の傾向と隠語が含まれる割合について調査した。次に、ID 交換掲示板で利用される隠語の傾向に応じた隠語の表記揺れを概念化する手法を開発し、最後に実証実験において、その有効性を検証した。

(2) CGM を利用した有害性アプリの発見手法の開発とその評価

携帯機器をアクセス手段とした事犯の多くは、直接インターネットを閲覧するのではなく、アプリケーション(以後：アプリ)を介して行われる。アプリを介したインターネット通信は、通常的手段ではアプリ外からアクセスすることができず、一度インストールしてしまうとフィルタリングサービスでは防げない場合がある。過去に検挙された事犯で利用されていたアプリとしては、児童の画像や卑猥な画像が原因で、児童ポルノや猥褻として検挙された「写真箱」、「写真カプセル」などの写真共有アプリや、チャットを介して出会い、性的な行為を行ったとして児童福祉法違反または青少年保護育成条例違反で検挙された「ぎゃるる」、「ひま部」などのチャットアプリなどがある。しかし、これらは氷山の一角であり、実際には数多くの犯罪が潜在的に発生していると考えられる。また、お絵かきリレーアプリが出会い系に利用されるなど、本来のアプリの利用目的とはかけ離れたアプリの使用の仕方青少年が危険な

事案に巻き込まれるなど、手口の多様化と複雑化が見られている。これらの現状をふまえ、本研究では、子どもを取りまくスマホアプリの実体把握に向けて、過去の検挙された事例や法律に関連の深いワードを起点にスマホアプリの情報を収集し、CGM(Consumer Generated Media)から危険性を調査する手法を提案する。

4. 研究成果

(1) 隠語概念化手法を用いたコーパスの構築

ID 交換掲示板に記載された書き込みに対して、5 つの処理を行うことにより発見した隠語を概念化し、コーパスと照らし合わせる手法を提案し、一定の有用性を示した。まず、隠語の概念化処理を行うにあたり、ID 交換掲示板の書き込みに含まれる隠語表現から、当て字として使われる「<」や「莓」、「千」などを登録した当て字辞書、表層的言い換えに利用されている「,」や「・」、「/」などの記号や感嘆符、文字を登録した表層的言い換え辞書を構築した。当て字変換処理では、入力データに当て字が含まれていた際に当て字辞書を用いて当て字をひらがなに変換する。また、数字については全角と半角が混在しているため、全て半角に変換する。次に、表層的言い換え変換処理では、入力データに表層的言い換え辞書に該当する記号が連続して使用された場合に記号を除去した書き込みデータと記号を1つのみ残した複数のデータを作成する。複数のデータを作成する理由については、表層的言い換えだけでなくマスクや短縮に対応するためである。「エッチ」という単語を例に挙げると、表層的言い換えでは「えっ、,,,,,,, ち」、マスクでは「エ チ」などのケースがある。この場合、単純に記号を除去するのみではマスクに対応することができない。そこで、書き込みの中に表層的言い換えが含まれていた場合、記号を全て除去した書き込みデータと記号を1つのみ残したデータを作成することで表層的言い換えとマスクのどちらの隠語表現であっても対応することが可能となる。そして、ひらがな化処理では入力された書き込みデータに含まれるすべての漢字と全角半角を含むカタカナをひらがな化する。ID 交換掲示板の記事では、漢字、ひらがなや半角カタカナ、全角カタカナや大文字小文字が入り混じって使用されることが多い。ここでは、MeCab のひらがな化機能を利用することで記事をひらがなに変換する。小文字変換機能ではひらがな化されたデータセットに含まれる小文字を全て大文字に統一を行う。最後に、エラーリストによる文字除去処理では、誤検出防止のためのエラーリストを作成し、特定の文字を除去した。援助交際を示唆する単語の中には、「えん」のように短い文字で構成されものが多い。そのため、「あまえんぼう」や「えんきより」などの単語で誤検出となる

可能性がある。エラーリストによる文字除去リストでは、このような誤検出となる単語を予め登録して除外した。尚、本提案手法における処理の順序について、記事内に当て字や表層的言い換えに該当する隠語表現が含まれていたと、正しくひらがな化処理を行うことができないため、当て字変換処理と表層的言い換え処理を最初に行う必要がある。そして、ひらがな化処理を行った後に小文字変換処理を行うことで、全ての漢字に含まれる小文字も大文字に変換することが可能である。本提案手法では、この5つの一連の処理を行うことによって、ID交換掲示板における表記揺れを含む書き込みの概念化を行った。有害性の評価では、概念化辞書を用いて表記揺れを解消する前と後のデータセットを用いて有害性評価における適合率、再現率、F値を比較し、有用性を示した。

(2) SPECを用いた隠語判定手法の精度向上手法の提案

(1)で構築した隠語概念化辞書を用いて有害判定を行うシステムをベースラインシステムとし、ベースラインシステムの精度を向上することを目的に、文パターン自動抽出手法 SPEC を用いた2項分類と有害性項目の詳細分類を行った。2項分類では、書き込み文章における有害・非有害の判定を行い、十分割交差検定により精度を確認した。そして、詳細分類では、SPECを用いた提案手法とSPECを用いないベースラインシステムを比較し、SPECを用いた提案手法の有用性を示した。

(3) スマホアプリ情報収集システムの構築

スマホアプリの実体調査を目的に、スマホアプリ情報収集システムを構築した。アプリ情報の収集手順として、まずキーワードを指定する。スマホアプリはGoogle Playだけを見ても200万件以上存在し、さらにそのレビュー情報となると膨大な数となる。そこで、本研究では、過去に事件にまで発展したアプリや取り締まり対象となる条例や法律から、「画像」、「写真」、「画像共有」、「動画共有」、「フリマ」、「出会い」、「チャット」、「通信ゲーム」、「通話」、「SNS」、「メル友」、「恋人」、「援交」の14個のキーワードを起点ワードとして定め、優先的にアプリの収集を行い、アプリ情報収集システムを2016年8月から11月までの4ヶ月間起動し、47,385個のアプリと217,334件のレビューを収集した。

(4) CGMを用いたスマホアプリの調査結果

スマホアプリ情報収集システムから得られたレビューを分析し、発見した知見についてまとめた。まず、本研究における危険アプリを青少年が犯罪に巻き込まれる可能性があるアプリと定義し、定義に該当するレビューや、レビューが含まれるアプリのレーティングやカテゴリを調査した。収集したレビュー情報から、危険性が含まれると判断した

レビューは589件あり、レビューが含まれていたアプリは329個であった。精査したアプリをGoogle Playに登録されているカテゴリに分類したものを図1に示す。

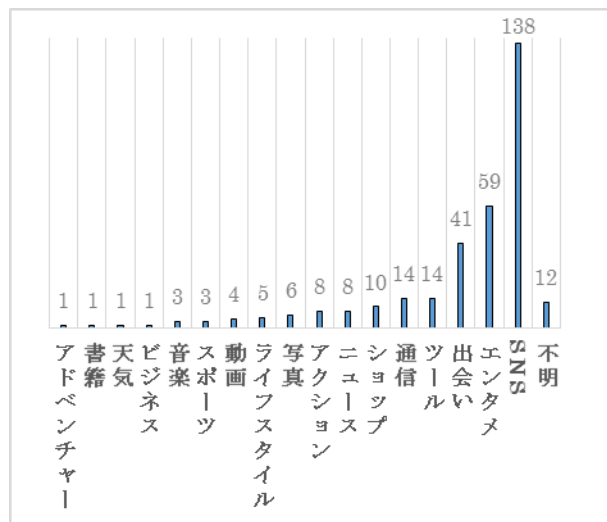


図1 危険アプリカテゴリ分類結果

により収集した329個のアプリ情報を対象に、各アプリのレーティングを調査した。各アプリにおけるレーティングの割合を図3に示す。分析の結果、レーティングを18歳以上としているアプリは102個であった。一方でレーティングを18歳未満としているアプリは227個あり、そのうち12歳以上は62個、3歳以上は132個であった。また、レーティングなしは22個であった。

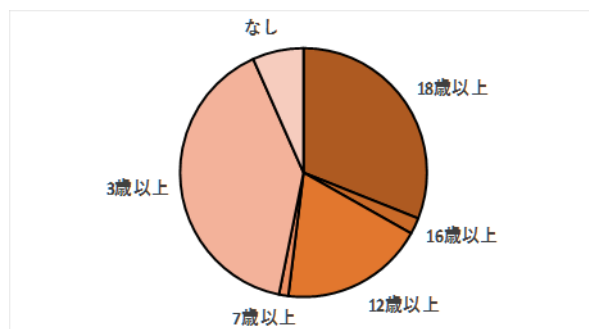


図2 危険アプリレーティングの割合

結果から、500件を超える被害レビューと329件の危険アプリの存在が明らかになった。危険アプリをGoogle Playのカテゴリ別に分類した結果では、SNSやエンタメ、出会いに分類されたアプリが多く検出された。また、「書籍」というカテゴリタグがついていても、アプリ名やアプリの内容が出会い系であるものも存在することが明らかになった。危険アプリのレーティングを調査した結果では、18歳未満が利用できるアプリでも多くの被害レビューが抽出された。18歳以上を対象とするアプリ102個については、ほとんどが出会い系もしくはチャットアプリであり、12歳以上対象アプリで

は、チャット、ゲームや掲示板が多数を占めた。「はじめに」で述べた通り、現在、事犯の多くはコミュニティサイトで起きており、その発生場所がインターネットでアクセスできる場所から、出会い系チャットアプリ等にとって変わられていることが本結果から明らかになった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表](計 1 件)

山田大, 安彦智史, 長谷川大, Michal Ptaszynski, 中村健二, 佐久田博司: ID 交換掲示板における書き込み有害性評価に向けた隠語概念化手法の提案, 言語処理学会第 22 回年次大会, pp. 49-50.

6. 研究組織

(1) 研究代表者

安彦 智史 (ABIKO Satoshi)

仁愛大学・人間学部コミュニケーション学
科・講師

研究者番号: 90560475

(4) 研究協力者

中村 健二 (NAKAMURA Kenji)

池辺 正典 (IKEBE Masanori)