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　　６　月　１７　日現在

機関番号：　１４４０１
研究種目：　挑戦的萌芽研究
研究期間：　2013～2014
課題番号：　２５５４０１４１

研究課題名（和文）匿名性を利用したネットいじめの数理モデル化と分析

研究課題名（英文）Cyber-bullying and Anonymity: Modeling and Analysis of Social Media Conversations

研究代表者
中野　賢（NAKANO, Tadashi）

大阪大学・生命機能研究科・招聘准教授

研究者番号：７０５７１１７３

交付決定額（研究期間全体）：（直接経費）　　2,800,000円

研究成果の概要（和文）：近年、ネットいじめが深刻な社会問題として取り上げられている。ネットいじめとは、パソコンや携帯電話、あるいは、インターネット上のサービスを利用して、加害者が被害者に対して、継続的に精神的苦痛を与えることと定義される。本研究では、ネットいじめの特徴を理解することを目的として、ソーシャルメディアサイトにおけるユーザ間のコミュニケーションデータを分析した。また、ネットいじめの被害者を推定する方法について考察した。

研究成果の概要（英文）：Cyber-bullying is a growing concern in today's information society. Cyber-bullying is a form of bullying that abuses technology to attack victims repeatedly over a long period of time. Cyber-bullies may use personal computers, cell phones or Internet services to negatively impact the mental health of victims. In this project, we collected a set of messages exchanged among users on a social networking web site, and analyzed the messages to characterize the features of cyber-bullying. In this project, we also discussed techniques to locate victims of cyber-bullying based on how users exchange messages.

研究分野：　情報ネットワーク

キーワード：　ネットいじめ　ソーシャルメディア　コミュニケーション

１．研究開始当初の背景

日本や欧米諸国においてネットいじめが深刻な社会問題として取り上げられている。ネットいじめとは、パソコンや携帯電話、あるいは、インターネット上のサービスを利用して、加害者が被害者に対して、継続的に精神的苦痛を与えることと定義される。近年の調査によると、10 代のインターネットユーザのうち 32%のユーザがネットいじめの被害経験をもつ。ネットいじめの被害者は、従来の学校現場等における対面のいじめの被害者と同様に、精神的苦痛を継続的に受けることによって鬱病等の感情障害に陥ることや、自殺や自殺未遂に追い込まれることが懸念されており、早急な対策が求められている。

２．研究の目的

ソーシャルメディアサイトがネットいじめのホットスポットになっているという報告がある。本研究では、ソーシャルメディアサイトにおけるユーザ間のコミュニケーションを分析し、ネットいじめの特徴を理解することを目的とする。また、ネットいじめを自動検出する新しい技術について考察する。

本研究では Springme におけるユーザ間のコミュニケーションを分析する。Springme は、質問と回答を基本としたユーザ間のコミュニケーションを支援するソーシャルメディアサイトである。Springme のユーザは匿名あるいはユーザ名を明らかにして質問を投稿し、他のユーザが回答する。類似する他のソーシャルメディアサイトと同様に、ユーザをフォローする機能やユーザのメッセージに対してスマイルする機能などが備わっている。2014 年 4 月の時点で約 400 万人のユーザが Springme を利用している。

Springme はユーザに匿名性を与えることによって自由な情報発信を保証している。その一方で、匿名性がネットいじめを助長していると考えられている。先行研究によると、攻撃的なメッセージの 8 割以上が匿名で投稿されていることなど、ユーザの匿名性と攻撃性に関係があることが指摘されている [M. J. Moore et al., “Anonymity and roles associated with aggressive posts in an online forum,” Computers in Human Behavior, vol. 28, no. 3, pp. 861-867, 2012]。また、Springme において投稿されたメッセージの 7~11%がネットいじめに関する内容を含んでいることから、高い頻度でネットいじめが発生している可能性が示されている [A。 Kontostathis et al., “Using Machine Learning to Detect Cyberbullying,” in Proc. International Conference on Machine Learning and Applications and Workshops, pp. 241-244, 2011]。

本研究では、ソーシャルメディアサイトの分析調査を行い、ネットいじめ発生原理の理解を深めることを目指す。ネットいじめに対する理解が深まれば、ネットいじめの発生を防止する方法やネットいじめを自動的に検出するシステムの開発に寄与できる。

３．研究の方法

まず、本研究の調査対象である Springme からユーザ間のコミュニケーションデータを取得するためのウェブクローラを開発し、データセットを構築する（１）。次に、構築したデータセットを利用して、メッセージの内容に基づき、ネットいじめの被害者と考えられるユーザを推定する（２）。更に、メッセージの内容を分析することなく、ネットいじめの被害者を推定できないか考察する（３）。以下に各々の詳細を述べる。

（１）ウェブクローラの開発とデータセットの構築

本研究で調査の対象とする Springme は、前述した通り、質問と回答を基本とするユーザ間のコミュニケーションを支援している。各ユーザは URL で識別されるページをもち、そのユーザがこれまでに投稿した質問や回答がそのページに公開される。

本研究で開発したウェブクローラは、ユーザをノード、質問・回答のペアをエッジとしたグラフに対して、指定したユーザを起点として幅優先探索で自動巡回する。その際、ユーザのプロファイルに記載されている情報、及び、質問・回答の履歴を取得し、以下の内容をもつレコードをデータベースに保存していく。

- #: レコードの識別子
- asker ID: 質問者のユーザ名
- answerer ID: 回答者のユーザ名
- question: 質問の内容（テキスト）
- answer: 回答の内容（テキスト）

なお、question や answer としては、テキストデータのみを取得し、静止画や動画データは取得しない。また、Springme の特徴として、質問者が匿名になる場合が多くあり、この場合は asker ID が Anonymous となる。

開発したウェブクローラは、各ユーザの質問および回答の各々について、あらかじめ指定した数を最大数としてデータを取得する。なお、Springme ではユーザがページをスクロールすることにより、過去の質問や回答を閲覧できるようになっているため、ウェブクローラが単純に HTTP の GET 要求を出すだけでは過去のデータを取得できない。本研究では、このような Web ページの収集、解析機能を備えている PhantomJS （Javascript ベースのライブラリ）を利用して、指定最大数のデータを取得できるウェブクローラを開発した。

（２）メッセージ内容に基づくネットいじめ被害者の推定

メッセージの内容に基づいて各ユーザが受けている被害の度合い（評価値）を算出し、ネットいじめの被害者を推定した。以下にその詳細を述べる。

まず、各レコードの question と answer の各々にネットいじめを示唆するような攻撃的表現が含まれるかどうかを調べ、yes（攻撃的である）あるいは no というフラグを付与した。このフラグを付与するために、インターネット上(www.noswearing.com) に公開されている bad word のリストを利用し、機械的なキーワードマッチングによって、question や answer に含まれる bad word の数をカウントし、1 つ以上の bad word が含まれる question や answer には yes というフラグを付与した。

次に、フラグ付けしたデータセットをもとに、各ユーザが受けている被害の度合い（評価値）を以下の式で定義し、算出した。

Ui（ユーザ i の評価値）=

$$w_1 m + w_2 f_m + w_3 n + w_4 f_n$$

ただし、この評価式に用いた変数は以下の通りである。

- m: ユーザ i に対する攻撃的メッセージの総数（``asker id = i"を満たすレコードのうち answer が攻撃的であるレコードの総数）
- f_m: ユーザ i に対するメッセージのうち攻撃的メッセージが占める割合（n の値を``asker id = i"を満たすレコードの総数で割った値）
- n: ユーザ i に対して攻撃的メッセージを 1 回以上投稿した（ユニーク）ユーザの総数
- f_n: ユーザ i に対してメッセージを投稿したユニークユーザのうち攻撃的メッセージを 1 回以上投稿した（ユニーク）ユーザの割合
- w_j（j = 1…4): 重み係数（0 以上の値）

攻撃的なメッセージを多く受取っているユーザや多くの（ユニークな）ユーザから攻撃的なメッセージを受取っているユーザは、U_i の値が大きくなり、ユーザ i はいじめの被害者であると推測される。

（3）グラフ理論的特徴量と評価値の相関分析

ユーザをノード、質問・回答のペアをエッジ（回答者から質問者への有向辺）としたグラフを作成し、各ノードの特徴量を算出する。更に、算出した特徴量からネットいじめの被害者を推定できないかを考える。このような推定方法は、メッセージの内容を分析する必要がないため、ユーザのプライバシーを保護できる。ネットいじめの被害者を検出する新しい技術の開発につながる可能性がある。

本研究ではノードの特徴量として以下を算出した。

- 入力次数: 他ノードから自ノードへの有向辺の本数（他ユーザから受取った回答の数）
- 出力次数: 自ノードから他ノードへの有向辺の本数（他ユーザの質問に回答した数）
- 次数: 入力次数と出力次数の和
- クラスタ係数: 自ノードと直接つながっている異なる 2 つノードの間に有向辺がある確率
- 近接中心性: 自ノードから他ノードへの平均距離の逆数（他ノードとの平均距離が小さいと、近接中心性が高くなる。）
- 中心媒介性: 自ノード以外の異なる 2 ノード間の最短経路上に自身が含まれている割合
- 離心率: 最も離れているノードとの距離
- ページランク: 周辺ノードの入出力次数を考慮したノードの重要度を表す特徴量

以上の特徴量が（2）で算出したユーザの評価値と相関があるかどうかを検証した。

4．研究成果

開発したウェブクローラを利用して、合計 14,568 ユーザの回答や質問の履歴を収集し、164,108 件のレコードを含むデータセットを構築した。構築したデータセットを分析した結果、攻撃的な質問は全体の 5.8%（9,575 件）を占め、攻撃的な回答は 5.3%（8,758 件）を占めることが分かった。更に、攻撃的な回答を一度でも受取った人の割合は 14.3%（2,086 人）であった。

また、攻撃的な質問に対して回答が攻撃的になる割合は 15.6%（1,496 件）であり、攻撃的でない質問に対して回答が攻撃的になる割合は 4.7%（7,262 件）であった。これより攻撃的な質問に対して回答が攻撃的になる傾向が確認できた。一方、攻撃的でない質問に対して回答が攻撃的になる件数は、攻撃的な質問に対して回答が攻撃的になる件数よりも 4.9 倍程度多いことが分かった。

（2）で導入した評価値の累積確率を図 1 に示す。評価値はユーザが受けている被害の度合いを表す。重み係数には任意の値を与えた。図 1 では、$w_1=w_3=w_4=1$ とし、w_2 = 10, 20, あるいは、30 としている。評価値の定義

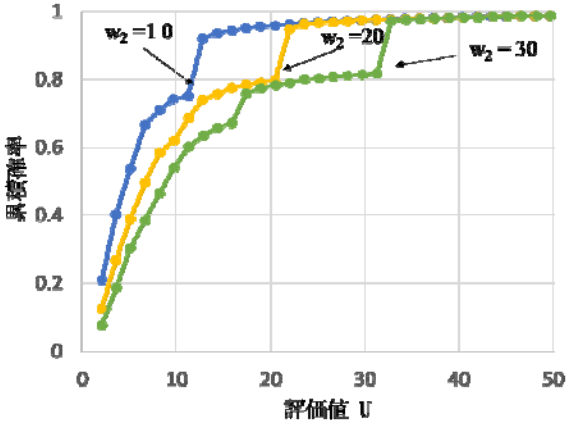

図 1 評価値の累積確率

より、重み係数が大きくなると評価値は大きくなる。図 1 より多数のユーザの評価値は小さく、ネットいじめの被害を受けていないと考える。例えば、w_2=10 の場合、評価値 10 以下のユーザが 74.2%を占める。一方、少数のユーザが非常に大きな評価値をもっていることが分かる。例えば、w_2=10 の場合、評価値 100 以上のユーザが 0.5%を占める。

本研究で取得したデータセットをもとに作成したグラフを図 2 に示す。このグラフにおいてノードはユーザを表し、エッジは質問者と回答者となるユーザを結んでいる。なお、エッジを質問と回答の頻度により重み付けしており、その重みを太さで表示している。ノード数は 14,568、エッジ数は 51,271、平均次数は 17.98、平均クラスタ係数は 0.0539 であった。

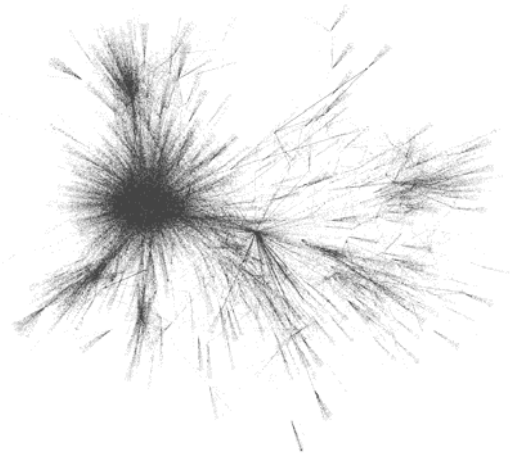

図 2 グラフ構造

（2）で導入したノードの評価値と（3）で述べたノードの特徴量の相関係数を表 1 にまとめた。入力次数、次数、および、ページランクとの相関係数は 0.74～0.80 であり、評価値と強い正の相関を示した。これに対して、中心媒介性との相関係数は 0.47、出力次数、近接中心性、および、離心率との相関係数は 0.2～0.3 であった。以上より、隣接ノードからの入力次数に着目したノードの特徴量はネットいじめの被害者を推定するのに有効であることが分かった。

表 1 ノードの特徴量と評価値の相関係数

入力次数	出力次数	次数	クラスタ係数
0.81	0.30	0.76	0.046

近接中心性	中心媒介性	離心率	ページランク
0.20	0.47	0.22	0.75

今後は、ネットいじめの存在とより強い相関がある指標やグラフ構造を特定するために研究を進めていく。従来のいじめと同様にネットいじめにも様々な役割をもつユーザが登場すると考えられる。例えば、被害者や加害者だけでなく、擁護者や傍観者が存在する。このようなユーザの役割を考慮して社会ネットワークを推測し、分析することによって、ネットいじめの発生を予測できるようになるかもしれない。例えば、社会的な優位性（例えば、多数のフォロワーをもつなど）が、攻撃的な発言を助長する可能性や、逆に擁護者や傍観者の存在がネットいじめを抑制する効果について仮説を立て、検証する。また、ネットいじめ特有の問題として、匿名性の影響も大きい。ネットいじめにおける被害者の周辺、あるいは、匿名ユーザと被害者間に見られる社会関係を理解できないか検討していく。

５．主な発表論文等

なし

６．研究組織

(1)研究代表者

中野 賢（NAKANO, Tadashi）
大阪大学・生命機能研究科・招聘准教授
研究者番号： 70571173

(2)研究分担者

なし

(3)連携研究者

なし