
原著論文

計量テキスト分析を用いた問題投稿に付帯されるハッシュタグの傾向把握 —サイバー空間における売買春活動の検知システムの開発に向けて—⁽¹⁾

Understanding the Trend of Hashtags Attached to Problematic Posting Using Quantitative Text Analysis: Toward the Development of a Detection System for Prostitution Activities in Cyberspace

キーワード：

SNS, 問題投稿, ハッシュタグ, 売買春活動, 計量テキスト分析

keyword：

SNS, Problematic Posting, Hashtag, Prostitution Activities, Quantitative Text Analysis

成蹊大学 吉見憲二

Seikei University Kenji YOSHIMI

セールスフォース・ドットコム 上田祥二

salesforce.com, Inc Shoji UEDA

摂南大学 針尾大嗣

Setsunan University Daiji HARIO

要約

スマートフォンやソーシャルメディアの普及を背景に、SNS等を通じた児童の被害が増加しており、青少年を取り巻く情報環境は大きなリスクと隣り合わせになっている。本研究ではその中でも近年深刻

原稿受付：2019年9月10日

掲載決定：2020年2月3日

な社会病理となっている「売買春」に着目し、Twitterにおける問題投稿に付帯されるハッシュタグの特徴について計量テキスト分析の手法を用いて分析した。

結果より、登場頻度の高いハッシュタグについて「エリア示唆」「援助（間接的表現）」「援助（直接的表現）」「裏アカウント示唆」「学生示唆」「性描写」「その他」の7つに分類し、その中でも地域名を明示する「エリア示唆」と売買春を間接的に示唆する「援助（間接的表現）」の共起関係が強いことが明らかとなった。こうした知見はサイバー補導における効率的な事前検知につながることを期待できる。

Abstract

Due to the widespread popularity of smartphones and social media to people specially to youths, the risk and danger in people's life is increasing. This research focused on 'prostitution' which has become a serious social pathology problem in recent years, analyzed the use of hashtags attached to problematic postings on Twitter using quantitative text analysis as one factor contributing to the issue. From the results, it was possible to classify seven groups with hashtags of high frequency appearance. Among them, it became clear that the co-occurrence relation between "area names" and "prostitution (indirect expression)" is especially strong. Such findings can be expected to lead to effective preliminary detection in cyber patrol.

1 はじめに

スマートフォンの普及を背景に、ソーシャルメディアの利用が一般化、低年齢化していくことに伴い、子どもたちが直面するリスクも増大してきている。警察庁(2017, 2018)では、出会い系サイト規制法施行以降、出会い系サイトに起因する事犯の被害児童数が減少し、過去最少を記録している一方で、SNS等における被害児童数は増加傾向が継続し、過去最多となっていることが指摘されている(図1)。また、被害児童のSNSへのアクセス手段は約9割がスマートフォン経由であること、Twitterにおける被害児童数は平成27年から平成28年にかけてほぼ倍増しており、平成29年には全被害児童数の1,813人のうちTwitterによる被害が3分の1以上の695人に達していることも調査から示されている。このまま有効な対策が打たれなければ被害数はさらに増大していくことが予想されるため、事業者や行政の連携によって有効な対策が行われる必要がある。



図1 SNS等に起因する事犯の被害児童数の推移
(出典)警察庁(2018)

Twitterにおける犯罪被害という点では、座間9遺体事件が記憶に新しい。これはTwitterで自殺願望を示していた複数の被害者が言葉巧みに加害者に呼び出され、犯罪被害にあってしまったという事件である。以下の記事では被害者がTwitterの投稿内容を表すハッシュタグ(＃をつけた文字記号)を用いて自殺願望を示し、見知らぬ

第三者とダイレクト・メッセージを用いてコミュニケーションをとっていたことが描写されている。

一連の事件の発端となった被害者の女性(23)は自身のTwitter上で「#自殺募集」というハッシュタグを使い、「だれか一緒に死んでくれる方」を募集していた。

女性は、その書き込みをみたA容疑者とメッセージのやりとりをし、事件に巻き込まれたとみられる。

女性の失踪後、女性の兄は同じアカウントから、こうつぶやいている。

「自殺募集をしたところ、多くの人がこのアカウントにDMで連絡をしていました」

「トントン拍子で話は進み、その4時間後には駅で合流してしまったみたいです」

(出典)BuzzFeedNEWS⁽²⁾

この事件を受けて、Twitter社の日本法人では利用規約を変更し、「自殺、自傷行為をほのめかず投稿を発見した場合は助長や扇動を禁じます」という一文を加えたうえで、違反者の投稿の削除やアカウント凍結を行う方針を示している⁽³⁾。事件を受けて対策が行われること自体は望ましいものの、児童の犯罪被害を増大させているITサービスが殺人事件にまでつながってしまっている事実は看過できるものではない。特に、売買春に関連した投稿では座間の事件と同様にハッシュタグを用いた探索的なコミュニケーションが行われているため、同様の手法で児童が被害にあってしまう懸念がある⁽⁴⁾。

こうしたなかで近年、警察や学校を中心として危険なTwitter投稿から子どもたちを守るための見守り活動が始まっている。これらの見守り活動では、ハッシュタグの検索機能を利用して危険性の高い問題投稿を個々に発見する探索的な調査を行なっている。ところがTwitterの仕様上、構造化されていないハッシュタグは統制されておらず、Twitter利用者によって各々生成されている

ため、見守り活動においては、子供達が使用する膨大なハッシュタグから問題投稿に用いられるであろうハッシュタグを類推し検索するという手探りの作業が行われている。そのため、スポット検索による部分的なものとなり、発生リスクの全体像を捉えるまでには至れないのが現状である。

本研究では、こうした問題意識から青少年を取り巻くリスクの中でも近年最も深刻な社会病理となっている「売買春」に焦点をあて、これらに関連するTwitterの問題投稿とそれに利用された一連のハッシュタグについて調査を行った。各章の概要は以下の通りである。第2章では青少年のソーシャルメディア利用傾向に関する先行研究を踏まえて、フィルタリングのような従来型の対応が難しい理由について検討する。併せて、ソーシャルメディア上の問題投稿に対するアプローチとしてサイバー補導を取り上げ、その課題について確認する。第3章では独自に収集したTwitterの問題投稿に関するデータから付帯ハッシュタグの傾向と共起関係について、計量テキスト分析の手法を用いて分析する。第4章では全体のまとめを行い、サイバー補導に資するようなシステムについて考察する。

2 先行研究

2.1 青少年のソーシャルメディア利用傾向に関する研究

前述のように青少年を取り巻く情報環境は大きなリスクと隣り合わせになっている。桂川ほか(2012)は全国の私立中高生5766名への調査を行い、ソーシャルメディア利用者のうち15%にソーシャルメディアを通じた対面経験があること、ソーシャルメディアを「なんとなく」利用している層では対面失敗経験が高まることを明らかにしている。ここでの「対面者」や「対面失敗経験」は表現に幅があるため、即座に問題視されるべきものとは限らないものの、ソーシャルメディ

アの普及によって面識のない人間と接するハードルが下がっていることは憂慮すべき事態である。

北村ほか(2016)では10代後半層のTwitter利用者への動機調査から、「会ったことのない者も多く含む相互フォローの関係を持つ多数のフォロワーに対して、10代後半層は『知って欲しい』という理由で投稿していることが多いと考えられる」と結論付けている。併せて、「10代後半層に意識されている他者とは、フォロワーの誰かではあるものの、さほど強く具体化(表象)されないもので、しかしながら実体はフォロワーの一部をなす親しい者(オンライン、オフラインの関係を問わない)であると推測される」として、オフラインでの面識の有無がソーシャルメディア上のコミュニケーションでは重視されていない現状に言及している。

こうした青少年のメディア利用に関する危惧に対しては、従来フィルタリングサービスの利用がその対応策として期待されていた。「平成29年度青少年のインターネット利用環境実態調査」ではスマートフォンを利用している青少年の保護者の実施している取組として、「フィルタリングを使っている」が44.0%、「子供のネット利用状況を把握している」が36.1%と上位であることを報告している。他方で、齋藤・新垣(2011)による「平成21年度青少年のインターネット利用環境実態調査」のローデータを用いた分析では、中学生・高校生と学年が上がるにつれてフィルタリングに対する知識と実際の使用状況の相関関係が低くなり、明確な関係性があるとは言えない結果となっている。知識は知識として修得しているものの、その知識が行動規範に至っていないという指摘は、桂川ほか(2012)における結論とも整合的である。

さらに、齋藤・新垣(2018)では「平成21年度青少年のインターネット利用環境実態調査」と「平成25年度青少年のインターネット利用環境実態調査」の比較分析を行い、啓発教育の実践が

統計的に有意に増加している一方で、中学生を除いて他の学齢期においては青少年の啓発教育の経験と実際のフィルタリングの利用率に有意な関係が示されなかったことを指摘している。特に、高校生では小中学生と比べて啓発教育を受けた機会数の平均値は高い数値を示しているものの、そうした啓発教育の経験は実際のフィルタリング利用に有意に結びついてはいない。

加えて、児童の被害が出会い系サイトからSNS等の普段から利用しているコミュニティサイトに移っていることもフィルタリングによる解決を難しくしている。齋藤・新垣(2011)ではこうした現状に対し、青少年の発達段階を考慮した協働学習を実施すること、齋藤・新垣(2018)では携帯デバイスの購入前に啓発教育を提供する社会的な仕組みを構築すること、青少年の発達段階に応じたフィルタリング設定を普及していくことを提案している。こうした提案自体は効果的なアプローチとして賛同できるものの、Twitterを代表としたソーシャルメディアを通じた児童の被害が右肩上がりに増え続けていることを踏まえれば、より即効性のある対策も同時に求められている。

2.2 サイバー補導に関する研究

サイバー補導とは「インターネット上で援助交際を求めるなどの不適切な書き込みをした未成年者に対し、注意や指導をすること(出典:デジタル大辞泉)」であり、2013年10月より全国の警察で実施されている。「サイバー補導の現状と今後の取組について」(2015年3月12日広報資料)では試行期間である2013年4月から2014年末までに18歳未満の児童597人が補導され、その6割以上に非行・補導歴がなかったことを報告している。ただし、本研究の分析対象としている援助交際(売買春)目的の児童は下着売買目的の児童と比べ、非行・補導歴をもつ割合は多くなっている。さらに、補導された児童の9割以上がスマートフォンとSNS等のコミュニティサイトを利用している。

近年ではさまざまな地域でサイバー補導が増加していることが報道されており、警察庁(2018)におけるSNS等での被害児童数の増加やその背景とも一致した傾向を見せている。

矢作(2014)は全国12か所の都道府県警察本部少年課を対象とした聞き取り調査から、サイバー補導の課題として「効率よく検索できる仕組み」を挙げている。そもそもサイバー補導における交信作業はさまざまな制約から「不適切な書き込みを発見し、現場での接触にこぎつけられるのは、ほんの数%程度である」という難しさがある。その中で、「ネット上にあふれる交流サイトの掲示板の中から、不適切な書き込みを発見する作業は、容易ではない。変化を続ける掲示板等のサイトの情報を、取りまとめてくれる人材と、情報共有できる仕組みが欲しい」というのが現場からの声となっている。警察に代わって対象者との交信作業を行うことは現実的ではないが、サイバー補導の対象となる書き込みを「効率よく検索できる仕組み」については情報システムによる解決が期待できるものである。

2.3 海外の先行研究

海外でも売買春とソーシャルメディアの関連を論じた先行研究は増えてきている。Musni-Tagaytay et al.(2016)はFacebookがフィリピンにおいて売買春の仲介に用いられていることを指摘し、売買春を示す“Walk”や若い売春婦を意味する“Candy”, すぐに対応が可能であることを示す“Xpress”といったSexual Slang(隠語)が用いられていることを明らかにしている。Cheung et al.(2016)は香港におけるcompensated dating(援助交際)に関与した少女へのインタビュー調査を通して、ソーシャルメディアを含めた情報通信技術が用いられていることに言及している。Irawan et al.(2017)はエスノグラフィーの手法を用いて実際に売買春のコミュニティに参加し、Facebookをプロモーションに用いた上で

その後のコミュニケーションをメッセージングサービスで行っている実態を指摘している。このように隠語の利用やソーシャルメディアとメッセージングサービスの併用など日本における問題事例との類似点が見られるものの、海外の先行研究においては定性的な研究が中心となっている。そのため、本研究のようにソーシャルメディア上の問題投稿を抽出し、計量テキスト分析を用いるようなアプローチを採用しているものはまだ多くはないものと考えられる。

2.4 先行研究の課題

青少年のインターネット利用に関する問題では、フィルタリングやリテラシー教育の実態についての先行研究は蓄積されているものの、青少年を取り巻く問題投稿に着眼した研究は海外事例を含めてまだ多くはない。特に、ハッシュタグを利用した情報探索については実際に被害者が存在する事件で行われているものの、その実態は明らかにされていない。そのため、警察や学校における見守り活動やサイバー補導の効率的な実施の阻害要因となっている。

実際に、援助交際を示唆する問題投稿において隠語が中心に用いられていることは過去の新聞報道等でも言及されている⁽⁵⁾。こうした隠語の組み合わせは流動的なものであるため、最新の利用傾向について経験則的に手探りで把握し続けることは容易ではない。そのため、問題投稿に付帯されるハッシュタグの種類や関係性を把握することができれば、青少年に有害な蓋然性の高い投稿を特定することがより容易になることが期待される。加えて、ハッシュタグに用いられやすい単語の傾向を把握することで、青少年を取り巻くリスクについて定量的に把握することが可能となる。

2.5 研究目的

上記のような社会背景や先行研究を踏まえて、本研究ではTwitter上の投稿内容の分析から、売

買春に関する問題投稿の蓋然性が高いハッシュタグの傾向を明らかにすることを目的とする。Twitterに着目する理由としては、現実に被害児童数が増えていること、青少年にとって面識のない他者との交流を促進する利用者数の多いソーシャルメディアであることが挙げられる。また、ハッシュタグに関しては、若年層の情報探索において中心的に用いられていることや後述するように売買春に関連した問題投稿において付帯されることが多いことから、取り上げる。さらに、売買春に着目する理由としては、対面接触を伴うため、より犯罪に巻き込まれるリスクが高いことが挙げられる。

第3章では、実際に独自に収集したTwitter投稿内容のデータから、計量テキスト分析の手法を用いて売買春に関する問題投稿に付帯されるハッシュタグの特徴について分析する。特に、サイバー補導に資する検知情報システムを考えるために、対面接触をする上で必要となるだろう地域名と隠語のいずれもが同時に記されている投稿が売買春に関する蓋然性が高いものであるという仮説を設定し、既知のハッシュタグから共起関係を通じて隠語を特定することを試みる。

3 調査・分析

3.1 調査概要

本調査では、特に売買春に関するTwitterの問題投稿の実態把握を目的として、投稿の状況及び特徴を調査した。収集対象は、事前の試行調査及び自治体・警察関係者へのヒアリング調査を参考として、援助交際を示唆するTwitterの問題投稿に付帯されることが多い「援助」、「援」、「円」、「サポ」の4つのハッシュタグのいずれかを含んだ投稿（オリジナルツイートとリツイート）とした⁽⁶⁾。投稿全体に関する傾向把握のためのデータ収集を、2018年5月2日から5月31日の30日間かけて実施した。

なお、本研究における問題投稿は前述の通り売買春に関連したものとして広く定義しており、「援助」、「援」、「サポ」についてはそのほとんどが業者によるものも含めて売買春に関連した文脈で用いられていることを確認している。ただし、「円」については「日本円」を意味する投稿として用いられている投稿も多かったため、売買春とは明確に関係ないと考えられる投稿については目視の範囲で分析対象から除外した。

3.2 調査環境

Twitter投稿はTwitter APIを活用して取得することが一般的であるが、技術動向調査の結果からTwitter投稿を容易に取得できるモジュールが利用可能な開発ツール「Node-RED」が有効であることが判明し、それを利用した。なお、Node-REDは英国IBMのHursley研究所によって開発されたオープンソースの開発環境であり、主にはIoTデバイスからのデータ取得等を行うアプリケーション開発向けのものである。

まず、CentOS(Linux関連)のサーバにNode-REDやDBのMySQLなどをインストールしてシステム環境基盤を構築した。次にNode-REDをもとに、標準的なノード及びJavascript言語を用いて自由にプログラムすることができるfunctionノードを多用してTwitter投稿を取得し、必要なデータをDBに格納するプログラムを作成した。

Twitter APIで取得できる投稿データはJSON形式であり、そのJSONには単に投稿内容だけでなく、投稿者やハッシュタグ、投稿日時など様々な関連情報が存在している。そのJSONデータから研究に必要となる項目のデータを取得し、DBに格納するようにした。特に、「投稿者」と「ハッシュタグ」については、問題投稿に関する有用な分析の対象となるため、次の通りとした。問題投稿者の累積投稿回数が容易に把握できるように、投稿のJSONデータから投稿者情報を取得し、投稿者テーブルに該当のデータの有無を確認し、データ

が「無」の場合はテーブルにデータを新規追加し、投稿回数の項目を1とし、データが「有」の場合は既存の投稿者レコードの回数を1プラスするようにレコードを更新した。また、ハッシュタグについては、1回の投稿で複数のハッシュタグが付与されていることもあるため、ハッシュタグテーブルに、該当データの有無の確認と必要に応じたデータの新規追加、或いはデータの更新をハッシュタグ別を実施した。図2は、Twitter APIからデータを取得し、DBにデータを格納するイメージを示している。問題投稿と推察されるハッシュタグをこのシステムに登録することで投稿を監視し、自動的に問題投稿をDBに登録することを可能とした。

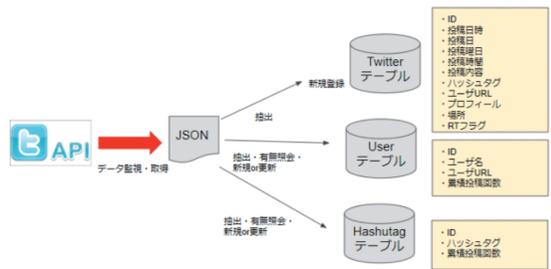


図2 データ取得のイメージ

3.3 調査・分析手法

ハッシュタグ間の関係性に着目した研究として、村井(2012)は東日本大震災後のTwitterにおける震災関連ハッシュタグの計量的分析を行い、話題のピークや中心的な話題の変化について検証している。当該の研究では、ハッシュタグの共起関係について、共起ネットワーク分析を用いて可視化しているため、ハッシュタグの利用傾向を調査目的とする本研究でも同様の手法を採用した。

本研究では、該当するハッシュタグの共起関係について、計量テキスト分析ソフトウェアKH Coderを用いて共起ネットワーク分析を行った。共起ネットワーク分析は「抽出語またはコードを用いて、出現パターンの似通ったものを線で結ん

だ図, すなわち共起関係を線 (edge) で表したネットワーク図を描く」(KH Coder公式サイト; 樋口, 2014) ものであり, 描写の基準として出現数と Jaccard 係数を用いた。Jaccard 係数は類似性の指標であり, 単語 X と単語 Y において, $|X \cap Y| / |X \cup Y|$ の式によって求められる。単語間の共起関係を表すものとして, 計量テキスト分析の手法で広く使用されている。共起ネットワーク分析では描写の明確な基準は存在しないが, 視認性を重視して 60 語程度を目安とすることが多い⁽⁷⁾。ただし, Jaccard 係数の上位から 60 語を基準とした場合には共起関係の基準の変化による差異を把握しにくいことから, 本研究では Jaccard 係数 0.1 と 0.2 を共起ネットワーク描写の基準として採用し, それぞれの結果を比較する⁽⁸⁾。また, 出現数については全体の 0.5% 以上の登場数とし, 600 以上を基準として採用した。これは後述する地域名の頻出上位 20 語を含むものとして設定している。

3.4 結果概要

期間中に取得した総投稿数は 120,020 件で, 日別の投稿数は図 3 に示している。毎日 3500 件以上もの前記 4 ハッシュタグのいずれかを含む投稿が観察され, そのうち RT (リツイート) は 1000

件から 1500 件程度であった。曜日ごとの変動はあったものの, 一貫した傾向は見られなかった。

投稿者に関しては, 期間中に 30,578 アカウントが該当した。各アカウントの投稿数の中央値は 1 であり, 半数以上は期間中 1 回のみでの投稿であった。なお, 投稿数の最大値は 787 であったが, 投稿数上位のほとんどのアカウントは後日調べたところ凍結されており, 投稿数 300 以上のアカウントはすべて凍結されていた。

ただし, 特定のハッシュタグによる抽出であるため, これらの投稿や投稿者がすべて売買春を目的としたものではない点には留意が必要である。また, 売買春に関するハッシュタグが付帯されているものの, 内容が全く別のもの (わいせつ動画の宣伝, 別の出会い系サイト・サービスの紹介等) も少なくなかった。以下では, こうした単純なキーワードからの検索の限界を踏まえて, 付帯ハッシュタグの傾向分析から, 地域名と隠語が同時に記されている投稿がサイバー補導の対象となるような売買春に関する蓋然性が高いものであるという仮説を検証する。

3.5 出現数上位のハッシュタグ

登場したハッシュタグは 5,045 種類となった。

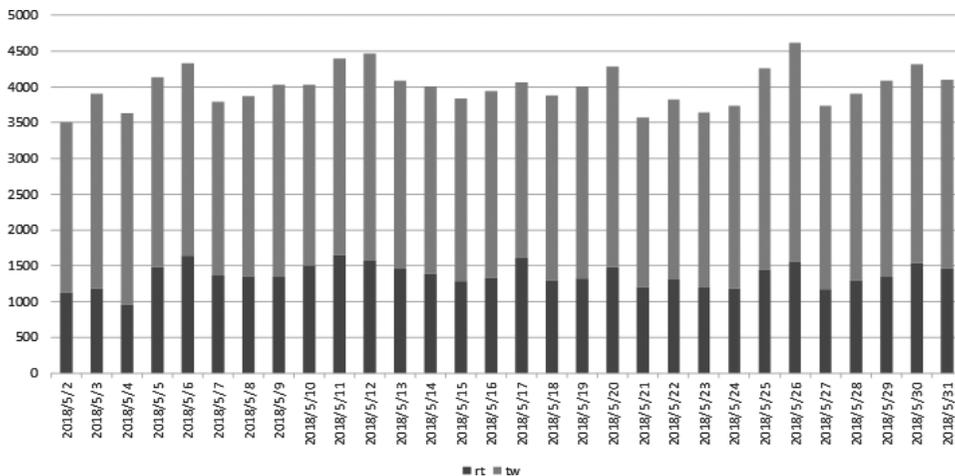


図 3 日別の投稿数と RT 数

今回はハッシュタグを分析対象としているため、以下ではこの5,045種類のハッシュタグを強制抽出の対象として登録し、分析を行った。出現数上位のハッシュタグは、「援」「サポ」「円」「オフパコ」「援助」「裏垢女子」「家出」となり、いずれも1割以上の投稿に当該ハッシュタグが含まれていた(表1)。「パパ活」や「神待ち」といった通常のコミュニケーションでは用いないようなハッシュタグの利用も見られた。

また、「都内」「池袋」「大阪」といった地域名が出現数上位に見られている。地域名のみに着目した出現数上位のハッシュタグは表2の通りである。付帯ハッシュタグ数の平均値は4.8個、中央

値は4個であり、複数のハッシュタグが同時に用いられる傾向にあった。10個以上のハッシュタグが用いられている投稿も少なくなかったが、こうした極端にハッシュタグが多い投稿の詳細を確認すると、業者によるものと疑われる投稿も散見された。なお、ハッシュタグではないものの、本文に「DM(ダイレクト・メッセージ)」という単語を含む投稿は28,122件で全体の23.4%に達していた。

ただし、単純集計では出現数上位のハッシュタグは把握できるものの、複数同時に用いられていることが予想されるハッシュタグの関係性については十分に検討できない。そのため、全投稿(N

表1 出現数上位のハッシュタグ(全体)

No.	付帯タグ(#)	付帯数	付帯率
1	援	70703	58.9%
2	サポ	41685	34.7%
3	円	37995	31.7%
4	オフパコ	25574	21.3%
5	裏垢女子	15674	13.1%
6	家出	13643	11.4%
7	援助	12300	10.2%
8	プチ	12713	10.6%
9	裏垢	11704	9.8%
10	神待ち	11600	9.7%
11	エロ垢	10343	8.6%
12	セフレ	9660	8.0%
13	えん	9574	8.0%
14	パパ活	9567	8.0%
15	都内	9501	7.9%
16	援交	6492	5.4%
17	泊めて	6790	5.7%
18	池袋	5979	5.0%
19	見せ合い	5611	4.7%
20	エログル	4802	4.0%

表2 出現数上位のハッシュタグ(地域名)

No.	付帯タグ(#)	付帯数	付帯率
1	都内	9501	7.9%
2	池袋	5979	5.0%
3	大阪	4582	3.8%
4	東京	4024	3.4%
5	名古屋	3569	3.0%
6	新宿	2493	2.1%
7	渋谷	1964	1.6%
8	横浜	1933	1.6%
9	札幌	1802	1.5%
10	愛知	1504	1.3%
11	難波	1196	1.0%
12	千葉	1164	1.0%
13	梅田	1148	1.0%
14	滋賀	1081	0.9%
15	埼玉	1063	0.9%
16	神奈川	1045	0.9%
17	岡山	956	0.8%
18	福岡	936	0.8%
19	京都	857	0.7%
20	倉敷	754	0.6%

= 120, 020) を対象とした共起ネットワーク分析を次に行った。

3.6 共起ネットワーク分析

共起ネットワーク分析の結果は、図4と図5の通りである。共起ネットワーク分析は前述の通り描写の基準に応じて結果が変わるため、Jaccard係数の異なる結果を示している。図4と図5からは、援助交際を示唆する「援」「円」「サポ」といったハッシュタグが強い共起関係にあることが明らかとなっている。このことは複数の隠語が同時に用いられていることで、その意味を補強していることを示している。その他には地域名同士の関係性の強さや特定のハッシュタグ間の共起頻度の多さも見られている。ただし、後者については業者アカウントが繰り返し同様の表現を用いていることから、結果として共起頻度が高まっているものも含まれている。

さらに、類似した意図に関してさまざまなハッシュタグが用いられる場合には、個々のハッシュタグの共起関係を過小評価してしまう可能性がある。そこで、分析対象となったハッシュタグ（期間中出現数600以上）に関して、表3のように同様の意図を持って用いられているハッシュタグごとにカテゴリを分けて類型化を行った。類型化に当たっては著者ら複数人の主観をもとに分類を行っているが、特に見て援助交際を意味するもの（「割り切り」含む⁽⁹⁾）や援助を直接呼びかけるものを直接的表現とし、売買春に関する隠語の知識がないと意味をつかめないものを間接的表現とした。ただし、「パパ活」のように当初の意味を離れて売買春に関連した用語として認知されつつあるものもあるため、直接的表現と間接的表現を厳密に分類することは難しい⁽¹⁰⁾。

類型化の結果、「エリア示唆」「援助（間接的表現）」「援助（直接的表現）」「裏アカウント示唆」「学生示唆」「性描写」「その他」の7つの分類が見られた。図6ではそれぞれの類型と全投稿に占

める比率を示している。「援助（間接的表現）」は全投稿の9割程度を占めており、地域名等を含んだ「エリア示唆」も全投稿の3割以上に達している。「サポ」や「神待ち」に代表される隠語はある程度的前提知識がないと売買春に関連した用語であると見抜くことは難しい。そのため、「援助（直接的表現）」がなく、「援助（間接的表現）」が含まれる投稿はより発見がしづらいものと考えられる。

最後に、類型化したハッシュタグ間の共起関係について図7と図8に示している。ここでは「援助（間接的表現）」を中心に「エリア示唆」「裏アカウント示唆」「援助（直接的表現）」「性描写」といった共起関係がはっきりと示されている。特に、地域名を明示している「エリア示唆」とJaccard係数0.2以上の共起関係を有しているカテゴリが「援助（間接的表現）」のみであるという事実は、地域名と隠語が同時に組み合わせられている投稿が売買春に関する蓋然性が高いものであるという仮説を裏付けるものである。

その証左として、図9の最低1つの地域名のハッシュタグを含む投稿を対象とした共起ネットワーク分析では、「援助（直接的表現）」がどのカテゴリともJaccard係数0.2以上の共起関係を有していない一方で、「エリア示唆」と「援助（間接的表現）」の共起関係は0.96にも達している（ $N=34,828$ ）。このことは最低1つの地域名のハッシュタグを含む投稿のほとんどが隠語に近いハッシュタグと共起関係にあることを意味している。この結果からも本研究の仮説は支持される。

なお、本研究と同様に計量テキスト分析の手法を用いて問題となる表現の特定を行っているものに後藤（2018）があり、クラスタ分析や対応分析（コレスポンデンス分析）といったさまざまな分析が行われている。しかしながら、前述のように本研究では分析対象である売買春に関連したハッシュタグに同様の意図で複数の異なるハッシュタグが用いられていることから、個々のハッシュタグ間の関係性を捉える分析手法を用いるこ

表3 頻出ハッシュタグの分類内訳

カテゴリ	代表的なハッシュタグ
エリア示唆	都内, 池袋, 大阪, 東京, 名古屋, 新宿, 渋谷, 横浜, 札幌, 愛知, 難波, 千葉, 梅田, 滋賀
援助 (間接的表現)	サボ, 神待ち, パン活, 募集, パン, サポート, パン活初心者, パン活募集, エン, p活
援助 (直接的表現)	援交, 割り切り, 援助交際, 援交募集, 援助, 援助募集, 泊めて, お小遣い, お助け
裏アカウント示唆	裏垢女子, 裏垢, エロ垢, 裏垢男子, 裏垢男子と繋がりたい, 援垢, 裏アカ女子, 裏アカ男子
学生示唆	JK, jd, JC, 女子大生
性描写	オフパコ, エロ, えっち, エッチ, セフレ
その他	エログル, 出会い, 依存したい, LINE 交換

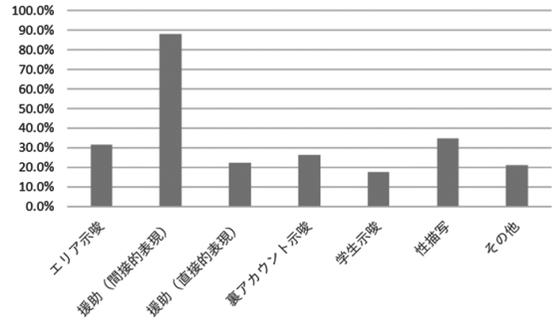


図6 頻出ハッシュタグの分類と全投稿に占める比率

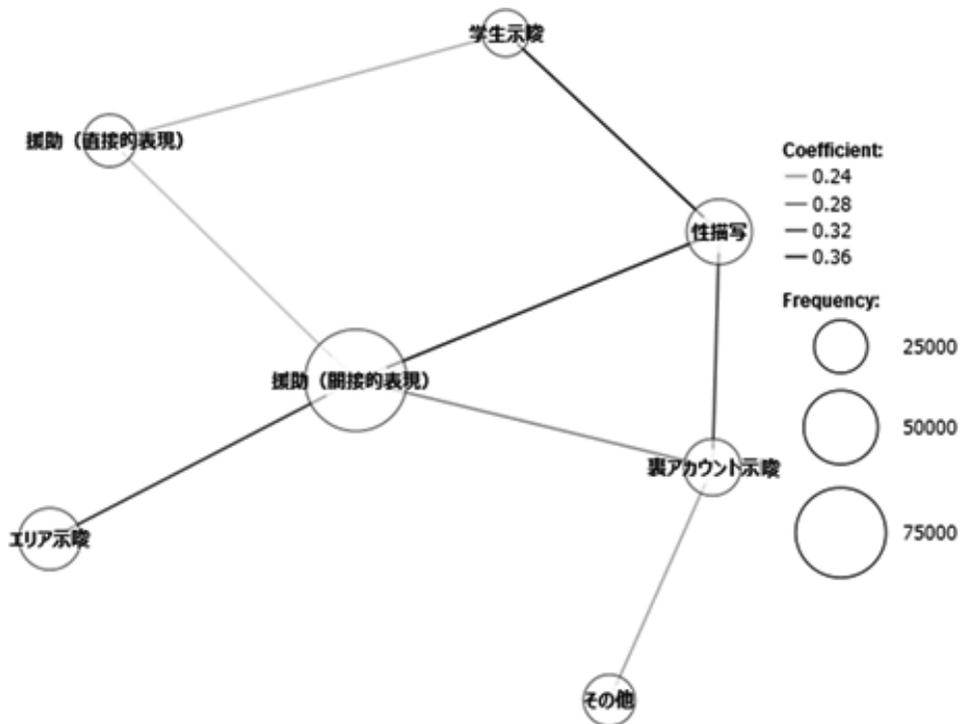


図7 類型化したハッシュタグ間の共起関係 (Jaccard係数=0.2)

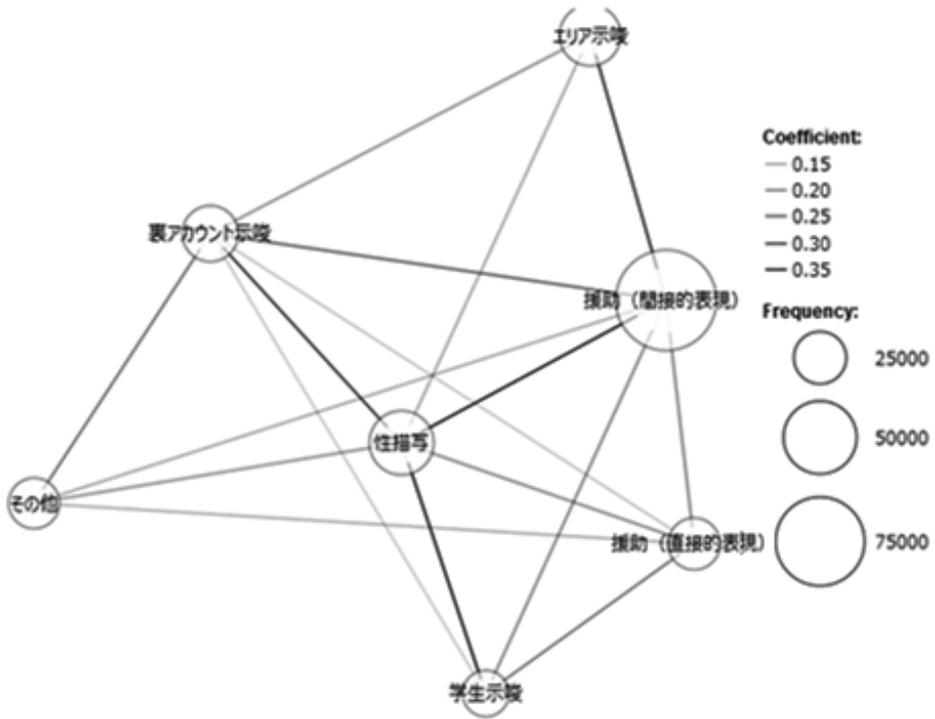


図8 類型化したハッシュタグ間の共起関係 (Jaccard係数=0.1)

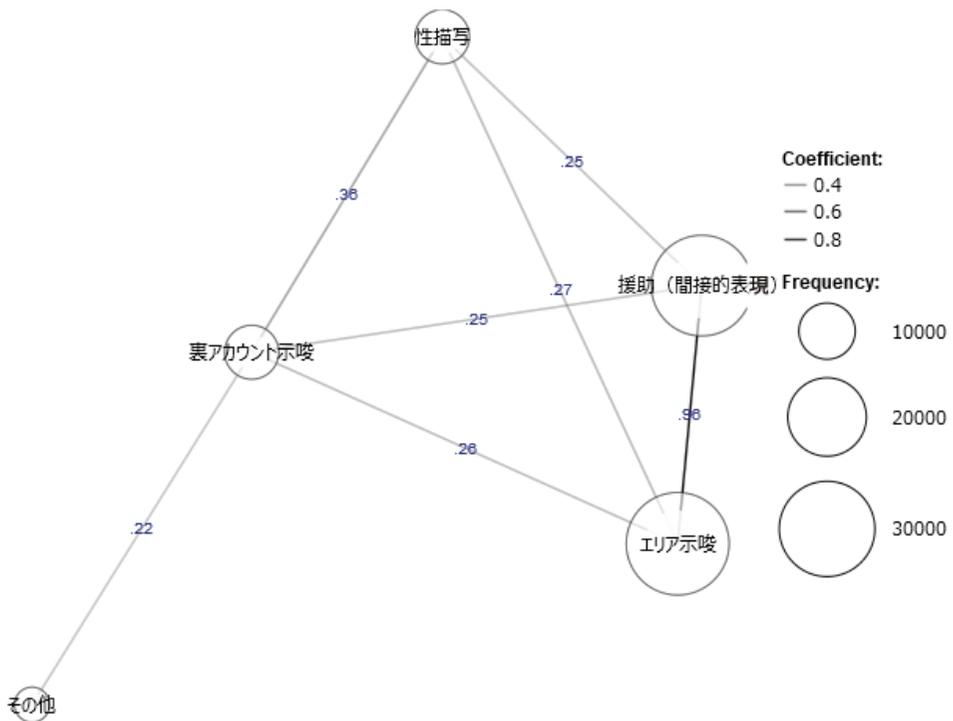


図9 地域名を含む投稿のみの共起関係 (Jaccard係数=0.2)

とは難しい。そのため、頻出のハッシュタグを人の目によって分類し、カテゴリ化した上での共起ネットワークの結果を示すに留めている。

3.7 考察

一連の調査結果より、4つのハッシュタグのみを収集条件とただけでも毎日数千件の問題投稿が発生しており、Twitterという青少年が日常的に利用するツールにおいてそれらを目にする可能性があることが明らかとなった。これは警察庁(2018)においてTwitterでの被害児童数が近年急増している事実とも整合的である。座間の事件を受けて自殺や自傷に対する利用規約が改正されるなど対策が行われているものの、「売買春」という別の社会病理問題に対しても深刻な状況であることが確認された。

さらに、「裏アカウント」であることを明示しているアカウントが多いため、仮に不適切な投稿の削除やアカウントの凍結といった措置を取ったとしても、すぐに新しいアカウントが作成されるといういたちごっこになってしまうことも想像に難くない。実際に、期間中の投稿数が多いアカウントに対する凍結措置は行われているものの、半数以上のアカウントでは期間中に1回しか投稿されていない。このことは、相互フォロー関係による継続的なコミュニケーションではなく、ハッシュタグを中心とした即時的なコミュニケーションがTwitter上では志向されていることを示唆している。

ハッシュタグに関する分析では、直接的に「売買春」を表現している「援助交際」や「援交」といったハッシュタグよりも、隠語を示す「円」「円幕」「サポ」や身元を特定させないような「裏垢」「裏垢女子」といったハッシュタグが頻繁に用いられていることが明らかとなった。こうした複数のハッシュタグが用いられる背景要因については本研究の直接の研究対象ではないが、Musni-Tagaytay et al.(2016)が指摘するように警察等

の監視から逃れるために隠語でのコミュニケーションを志向している可能性は高い。さらに、ハッシュタグの利用傾向が相手を見極めるためのある種のシグナルとして機能していることも考えられる。このことはサービス提供企業による監視やサイバー補導等の対策が強化されるに伴い、隠語でのコミュニケーションのニーズが高まることを示唆している。同時に、付帯ハッシュタグから隠語を示すハッシュタグの内容を自然と学習してしまう懸念があるため、特定の単語のみ禁止ワードに設定することが効果的な対策であるかは慎重に議論する必要があるだろう。加えて、「円」のような一般的にも用いられる隠語が増えることで、売買春が疑われる投稿の探索が難しくなることも考えられる。

また、地域名を示すハッシュタグの登場頻度が多いことも特徴的な傾向であった。これはコミュニケーションが仮想空間に留まらず実世界でも展開されている危険性を示すものである。併せて、多くの投稿の本文中にDM(ダイレクト・メッセージ)の記載があることも、こうした見方を補強している。本研究の手法では不適切なハッシュタグを用いた問題投稿の検出はできるものの、DMを通したやり取りが行われてしまえば第三者から把握することは困難である。これは前述の座間の事件において実際に行なわれていたことであり、同様の事件が自殺願望に関わる投稿だけでなく、「売買春」に関わる投稿において発生する可能性も否定できない⁽¹¹⁾。

他方で、売買春に関連した投稿において隠語が使われていることは以前から知られていることであったが、問題投稿に含まれる既知のハッシュタグの共起関係から隠語を含めた類似した意味合いを持つハッシュタグを特定し、その分類ごとの共起関係から蓋然性の高い問題投稿を効率的に探索できる可能性が示されたことは本研究の大きな貢献と言える。本研究のこのような提案手法を用いれば、仮に新しい隠語が登場したとしても、既知

のハッシュタグとの共起関係から新規に流行する隠語をいち早く特定することが可能となる。それにより、これまで経験則的に行われてきた問題投稿の探索作業を効率化することができる。さらに、地域名と売買春の蓋然性の高い隠語との組み合わせを検出できれば効率的なサイバー補導や警告⁽¹²⁾につなげることが期待できる。

4 まとめ

本研究では、「売買春」に関連するTwitterの問題投稿とそれに利用されたハッシュタグについて調査・分析を行った。テキストマイニングを用いた付帯ハッシュタグの分析から、地域名や隠語を含めた複数のハッシュタグが同時に利用されている傾向があり、こうしたハッシュタグを特定することで問題投稿の事前検知に有効であることが示された。調査した範囲だけでも実際に毎日数千件の問題投稿が発生していることから、事前検知による問題のある投稿の削除やアカウントの凍結は1つの対策として考えられる。

ただし、事前検知による削除や凍結の対応が隠語を用いたコミュニケーションを促進してしまう懸念も同時に示唆されている。加えて、ハッシュタグのような公開されている投稿内容については把握ができるものの、DMを通じたコミュニケーションが行われると外部から把握することは困難となる。本研究ではこうした点については検討できていないが、本質的な解決のためには看過できない問題である。

しかしながら、Twitterがきっかけとなり青少年が見知らぬ第三者と知り合い犯罪被害を受けてしまう事案が増えている中で、事前検知からそもそもその問題投稿を特定することは必要なプロセスであり、本研究の手法が活用できる余地がある。特に、サイバー補導の課題として「効率よく検索できる仕組み」が希求されていることから、本研究の提案手法をシステム化することで、その一

助となることが期待できる。問題投稿に含まれる既知のハッシュタグの共起関係から隠語を含めた類似した意味合いを持つハッシュタグを特定し、その分類ごとの共起関係から蓋然性の高い問題投稿を効率的に探索するという手法は自殺教唆等、他の社会病理問題にも応用可能であるため、事前検知からの効果的な対策を含めて引き続き検討していきたい。

謝辞

本研究は公益財団法人電気通信普及財団の2017年度研究助成（「SNS上の発生リスクとデータリンク構造（代表：針尾大嗣）」）を受けたものである。

本論文の執筆にあたり、3名の査読者、及び、編集委員の先生より大変多くの有益なご指摘をいただいた。この場を借りて厚く御礼申し上げたい。

注

- (1) 本稿の内容は第38回情報通信学会大会における報告内容「SNSにおける問題投稿の内容と付帯ハッシュタグの特徴」、及び、2018年社会情報学会（SSI）学会大会における報告内容「問題投稿に付帯されるハッシュタグの特徴—売買春を事例として—」を加筆・発展させたものである。有益なコメントをいただいた先生方にこの場を借りて御礼申し上げる。
- (2) 「きっかけはTwitterだった。9遺体事件、被害者が使った「#自殺募集」を見かけたら」(BuzzFeedNEWS)
<https://www.buzzfeed.com/jp/kotahatachi/how-to-save>
 なお、元の記事では容疑者名が明記されているものの、ここでは伏せている。
- (3) 「Twitterルール」
<https://help.twitter.com/ja/rules-and-policies/twitter-rules>より。なお、当該ルー

ルには児童の性的搾取を禁止する旨も明記されている。

- (4) 警察庁の調査では児童ポルノ及び児童買春事犯が増加傾向にあること、被疑者と会った理由では、「金品目的」及び「性的関係目的」に関連する理由が4割強を占めることが指摘されている。なお、座間の事件を受けて「青少年ネット利用環境整備協議会」への情報提供を行い、緊急提言に協力するなど、当該事件を児童被害の観点から深刻に受け止めている様子が窺える。
- (5) 「サイバー補導163人最多 福岡県警・昨年 援交「小遣い稼ぎ」少女6割、非行歴なし（西日本新聞2017年9月13日）」
(<https://www.nishinippon.co.jp/feature/attention/article/379458/>), 「SNS「サイバー補導」急増 少女が援助交際目的投稿（岐阜新聞2018年8月3日）」
(<https://www.gifu-np.co.jp/news/20180803/20180803-62856.html>) 等で隠語を用いたコミュニケーションが行われていることに言及されている。
- (6) 売買春に関連する投稿内容について予備調査を行い、援助交際を示唆する投稿に含まれているハッシュタグとして選定した。そのため、「パパ活」や「ママ活」といったハッシュタグについては初期の抽出条件からは除外している。しかしながら、近年では売買春に近い文脈で「パパ活」「ママ活」という用語が用いられることも多いため、その影響については本研究の結果では過小評価している懸念がある。
- (7) KH Coderの初期設定ではJaccard係数の上位60語を表示するようになっている。
- (8) 例えば、柳瀬（2012）ではクラスター分析を行った関連語のJaccard係数について0.075～0.336という数値を示しているため、0.1と0.2はこの範囲に収まるものとし

て捉えることができる。

- (9) 「割り切り」は売買春から派生した俗語であるが、認知度が高いため直接的表現として分類した。
- (10) 「パパ活」という言葉を作った交際クラブ関係者、「恋愛化するパパ活」に戸惑う」（ダイヤモンド・オンライン）
<https://diamond.jp/articles/-/206457>
上記の記事では、「パパ活」という言葉がもともと「交際クラブ・デートクラブ」と呼ばれている業態において使われているものであり、売買春を意図するものでなかったことが指摘されている。
- (11) 実際に座間の事件の後もTwitterのDMを用いたやり取りから殺人事件や未成年者の誘拐事件が発生している。
- (12) 「パパ活ツイートに直接返信で警告 効果抜群、全国警察に広がる」（毎日新聞）
<https://mainichi.jp/articles/20191105/k00/00m/040/349000c>
上記の記事の通り、最近ではサイバー補導に加えて、問題投稿に直接返信することで成果を上げている。ただし、問題投稿の探索にかかる負担は変わらないため、効率的な検知へのニーズは依然として大きい。

参考文献

- 桂川泰典・渡邊由里子・小野僚子・佐藤佳奈・関泰子・坂野葵・福田安奈・平田夏鈴・蔵田三沙代（2012）「中高生のソーシャルメディア利用実態（2011年）：ソーシャルメディアにおける“出会い”とその失敗の分析を中心に」『情報文化学会誌』第19巻第1号，pp.16-24.
- 北村智・佐々木裕一・河井大介（2016）『ツイッターの心理学:情報環境と利用者行動』誠信書房
- 後藤和史（2018）「教職員のわいせつ行為のニュース記事のテキストマイニングによる分析（2）年代コホートと時季との関連」『瀬木学園紀要』

- 12号,pp.3-13.
- 警察庁 (2015) 「サイバー補導の現状と今後の取組について」
- (2017) 「平成28年におけるコミュニティサイト等に起因する事犯の現状と対策について」
- (2018) 「平成29年におけるSNS等に起因する被害児童の現状と対策について」
- 齋藤長行・新垣円 (2011) 「青少年のインターネット利用における規範意識を育てるための協働学習についての研究」『情報文化学会誌』第18巻第2号, pp.60-67.
- (2018) 「啓発教育受講経験とフィルタリング利用との分析による携帯電話安全利用に向けた啓発教育政策のあり方の検討」『情報文化学会誌』第6巻2号, pp.15-30.
- 内閣府 (2018) 「平成29年度青少年のインターネット利用環境実態調査」
- 樋口耕一 (2014) 『社会調査のための計量テキスト分析—内容分析の継承と発展を目指して』ナカニシヤ出版
- 村井源 (2012) 「東日本大震災後のTwitter利用傾向 -震災関連ハッシュタグの計量的分析-」『情報知識学会誌』22巻2号, pp.97-106.
- 矢作由美子 (2014) 「サイバー補導の現状と課題：不良行為少年の補導活動の拡大化」『早稲田大学社会安全政策研究所紀要』第7号, pp.113-140.
- 柳瀬公 (2012) 「計量テキスト分析によるメディア・フレームの探索的検討：「放射性セシウム汚染牛問題」の新聞記事を通して」『社会情報学』第1巻2号, pp.61-76.
- Cheung, C.K., Jia, X., Li, J.C.M., Lee, T.Y. (2016) Engaging adolescent girls in transactional sex through compensated dating, *Journal of Adolescence*, Volume 52, pp.124-134.
- Irawan, F., Gultom, I., Amsani, H. (2017) Social Media As A Means For College Students Prostitution In Medan City, *Proceedings of the 2nd International Conference on Social and Political Development (ICOSOP 2017)*. <<https://www.atlantis-press.com/proceedings/icosop-17/25892095>> Accessed 2019, December 1
- Musni-Tagaytay, M., Viloan-Sagarino, E., Ancheta, P.B., Macrohon, C.N., Olarte, A.M. (2016) Walkers & handlers: Exploring the dynamics of Facebook mediated prostitution, *International Journal of Applied Business and Economic Research*, Volume 14, Issue 2, pp.1385-1404.