

SNS 上の発生リスクとデータリンク構造

針尾 大嗣 摂南大学経営学部経営情報学科 准教授
吉見 憲二 佛教大学社会学部 准教授
上田 祥二 株式会社セールスフォース・ドットコム

1 研究目的と研究方法

1-1 研究の背景と目的

SNS の利用において子供達が直面する危険やシグナルの早期発見とその保護介入は、現代情報社会における重要な政策課題である。自立過程の不安定な自己に苦しむ成長期の子供達にとって、SNS は自己肯定的な他者との出会いを提供してくれる居心地のよい場となる。しかしながら、そこで共有される彼らの悩みや問題は、SNS により強められた共依存的な人間関係の内側で、より深刻な社会病理問題へと発展しやすい。実際に SNS での出会いを起因として、詐欺、違法薬物、人身売買、性暴力、殺人といった事件の被害者となった子供達は後を絶たない。特に Twitter を利用して売春や自殺相談の相手との出会いを求める子供達の探索的なコミュニケーション活動は活発化しており、これによる被害児童数は平成 27 年の 226 人から翌年には 446 人と約 2 倍にまで急増している。こうしたなかで近年、警察や学校を中心として危険な Twitter 投稿から子供達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、Twitter の投稿内容を表すハッシュタグ (#をつけた文字記号) 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、また 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、きわめて経験的な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、発生している問題の規模、それに関わる者達の特徴、繋がり(経路)、悩みの種別(問題テーマ)など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。

そこで本調査では、SNS 上でのコミュニケーション活動におけるリスクの発見とその全体像の理解及び被害防止の為の保護介入を支援する情報分析システムに有用なデータ分析モデルを検討する為、近年最も深刻な社会病理問題となっている「売買春」に焦点をあて、これらに関連する Twitter 上での問題投稿の傾向とそれに利用された一連のハッシュタグ及びそのつながりを表すデータリンク構造について調査する。

1-2 研究方法・進め方

本調査では、まずキーワード指定検索による Twitter 投稿(問題投稿)及びハッシュタグ収集の為の Web スクレイピングプログラムを制作する。指定検索するキーワードは「売買春」に関連した問題投稿に使用されている既知のハッシュタグであり、これらは研究代表者が関わる自治体・警察関連機関の担当者を対象としたヒアリング調査にて得られるものである。次に上記プログラムを用いて既知の問題ハッシュタグが使われた問題投稿を探索し、その投稿時に使われた他の未知の問題ハッシュタグをあわせて収集し、売買春に関わるハッシュタグと問題投稿全体の投稿傾向を調査する。さらに、これらのテキストデータを形態素解析により形態素化した後、調査の上で不要な文字記号等を除去するデータクリーニングを実行し、オントロジー理論等を参考としてデータ定義を行い構造化したのちデータベースを構築する。このデータベースを用いてネットワーク分析を行い、既知・未知のハッシュタグのつながりを表すデータリンク構造を明らかにする。これらをもとに発生リスクの全体像の把握を試み、以上の流れを自動化したリスク発見支援の情報分析システム開発に向けての課題を整理する。

2 調査環境 (Web スクレイピングプログラムの制作) について

本調査では、SNS への投稿を取得し、分析することが必要となる。SNS の中でも匿名性が高く、利用者の多い「Twitter」の投稿を分析の対象とする。したがって、Twitter 投稿を取得することが基本要件として容易に Twitter 投稿を取得できる手段の技術動向調査を行った。

調査した。調査には Twitter の投稿情報を取得するオープンソースのモジュールである「Node-Red」の「Twitter」ノードを用いて、投稿をモニターし投稿情報を取得してデータベース上で管理した。モニター対象は、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」、「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とした。投稿全体に関する傾向把握のための調査を、2018年5月2日から5月31日の30日間かけて行い、このうち問題投稿に使用される未知のハッシュタグを収集するための調査を、ゴールデンウィーク期間中の5月2日から8日の7日間に実施した。

3-2 調査概要

3-2-1 問題投稿の傾向

まず、売買春に関わる問題投稿に用いられている既知のハッシュタグ「援助」、「援」、「円」、「サポ」が使用された問題投稿の調査を行なった結果、総投稿数は120,019件、1日平均投稿数は約4,000件、投稿者総数は30,624名、1人あたりの平均投稿数は約3.9件であった。日別投稿では、曜日に若干の多寡はあるものの著しい増減は確認できなかった。また曜日別投稿では、週の半ば水曜日から木曜日にかけて投稿量が多いことが観察された。時間別投稿数では、6時から投稿活動が開始し23時頃をピークとして活発化していることが観察された。

3-2-2 問題投稿者の基本的な特徴

調査では、繰り返し問題投稿を行う投稿者の存在が確認されたため、投稿件数により区分し、件数区分ごとの投稿者数と投稿数をカウントし、図表3にまとめた。これによると、全投稿者の93.4%(28,609名)が行う投稿は一人あたり10回未満であり、全問題投稿のうち約半数(49.1%)分にあたる。一方で、残り半数(50.9%)の問題投稿は、わずか6.6%(2,015名)の投稿者により行われていることが判明した。

図表3. 投稿件数でみた傾向

区分	投稿者数	投稿者数累積	投稿者割合	投稿数	投稿数累積	投稿数割合
100件以上	64	64	0.2%	17,005	17,005	14.2%
90件以上100件未満	7	71	0.0%	660	17,665	0.5%
80件以上90件未満	9	80	0.0%	741	18,406	0.6%
70件以上80件未満	17	97	0.1%	1,294	19,700	1.1%
60件以上70件未満	37	134	0.1%	2,369	22,069	2.0%
50件以上60件未満	46	180	0.2%	2,507	24,576	2.1%
40件以上50件未満	64	244	0.2%	2,777	27,353	2.3%
30件以上40件未満	144	388	0.5%	4,800	32,153	4.0%
20件以上30件未満	638	1,026	2.1%	15,674	47,827	13.1%
10件以上20件未満	989	2,015	3.2%	13,211	61,038	11.0%
10件未満	28,609	30,624	93.4%	58,981	120,019	49.1%
合計	30,624		100.0%	120,019		100.0%

(1) 大規模投稿者の傾向

この6.6%(2,015名)の投稿者のうち、100件以上の問題投稿を行った大規模投稿者は全体の0.2%(64名)で、全体問題投稿の14.2%にあたる。また大規模投稿者は、活動対象地域を示唆するハッシュタグを多く(広範に都市を指定して)用いている。これら大量投稿者の情報を整理すると、図表4の通りとなった。ハッシュタグに関する調査結果は次節で取り上げる。

図表4. 大規模投稿の傾向

区分	投稿者数	投稿者数累積	投稿者割合	投稿数	投稿数累積	投稿数割合
500件以上	9	9	14.1%	6,466	6,466	38.0%
400件以上500件未満	2	11	3.1%	962	7,428	5.7%
300件以上400件未満	5	16	7.8%	1,725	9,153	10.1%
200件以上300件未満	15	31	23.4%	3,493	12,646	20.5%
100件以上200件未満	33	64	51.6%	4,359	17,005	25.6%
合計	64		100.0%	17,005		100.0%

(2) 小規模投稿者の傾向

全体の93.4% (28,609名)の投稿者が行う問題投稿は10件未満であり、このうちの約半数(15,894名)は問題投稿1件のみの小規模投稿者である。また小規模投稿者は、「デート援」「お小遣い」「家出」「泊めて」といった個人的支援を意味するハッシュタグが多様に用いられている。これらの小規模投稿者の情報を整理すると、図表5の通りとなった。

図表5. 小規模投稿者の傾向

区分	投稿者数	投稿者数累積	投稿者割合	投稿数	投稿数累積	投稿数割合
9件	245	245	0.9%	2,205	2,205	3.7%
8件	346	591	1.2%	2,768	4,973	4.7%
7件	465	1,056	1.6%	3,255	8,228	5.5%
6件	643	1,699	2.2%	3,858	12,086	6.5%
5件	1,023	2,722	3.6%	5,115	17,201	8.7%
4件	1,613	4,335	5.6%	6,452	23,653	10.9%
3件	2,674	7,009	9.3%	8,022	31,675	13.6%
2件	5,706	12,715	19.9%	11,412	43,087	19.3%
1件	15,894	28,609	55.6%	15,894	58,981	26.9%
合計	28,609		100.0%	58,981		100.0%

3-3 ハッシュタグに関する調査

3-3-1 問題投稿に使用されるハッシュタグの種類

次に、未知のハッシュタグを調査するために、ゴールデンウィーク期間中(5月2日～8日)の問題投稿を対象に分析を行った。この間、問題投稿を行った投稿者数は27,210名であり、彼らが投稿時に利用したハッシュタグの総数(グロス)は113,674であった。全体として265種類のハッシュタグが使用されていた。

3-3-2 問題投稿に使用されるハッシュタグ

図表6は、問題投稿に使用されていた既知・未知のハッシュタグ265種類を使用頻度順上位120位までをまとめたものである。上位3つには、既知のハッシュタグ「援」、「円」、「サポ」が入り、ついで未知のハッシュタグ「プチ」、「裏垢女子」となっている。1位「援」については、売買春を目的とする問題投稿の半数以上(付帯率53.8%)に利用されていることが明らかとなった。

図表6: ハッシュタグ265種類の付帯数と付帯率

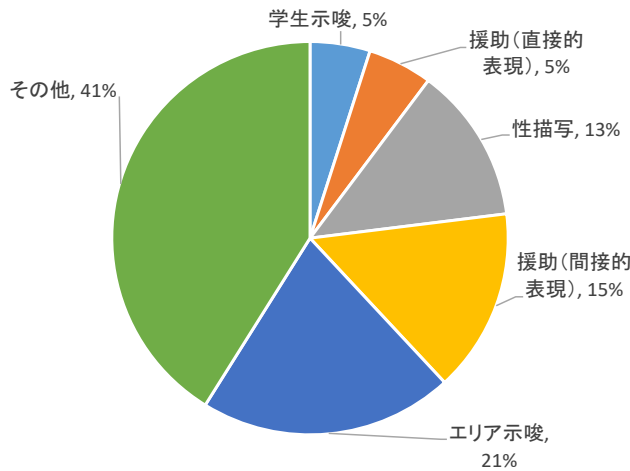
No.	付帯タグ(#)	付帯数	付帯率	No.	付帯タグ(#)	付帯数	付帯率	No.	付帯タグ(#)	付帯数	付帯率
1	援	5321	0.538	21	援募集	780	0.079	41	渋谷	438	0.044
2	円	3357	0.340	22	円募	772	0.078	42	裏垢男子と繋	415	0.042
3	サポ	3105	0.314	23	名古屋	758	0.077	43	札幌	398	0.040
4	プチ	2354	0.238	24	エロ	708	0.072	44	拡散希望	396	0.040
5	裏垢女子	2261	0.229	25	東京	694	0.070	45	募集	393	0.040
6	裏垢	1818	0.184	26	援助募集	664	0.067	46	援垢	388	0.039
7	エロ垢	1723	0.174	27	円募集	638	0.065	47	横浜	377	0.038
8	オフパコ	1702	0.172	28	割り切り	634	0.064	48	裏アカ女子	377	0.038
9	セフレ	1641	0.166	29	援助交際	622	0.063	49	パパ	366	0.037
10	えん	1629	0.165	30	援交募集	598	0.061	50	埼玉	355	0.036
11	援助	1553	0.157	31	泊めて	598	0.061	51	愛知	340	0.034
12	パパ活	1493	0.151	32	オナニー	570	0.058	52	難波	331	0.033
13	都内	1490	0.151	33	新宿	555	0.056	53	梅田	325	0.033
14	エログル	1403	0.142	34	サポ募集	534	0.054	54	処女	322	0.033
15	援交	1390	0.141	35	裏垢男子	521	0.053	55	メンヘラさんと	321	0.032
16	神待ち	1341	0.136	36	円光	496	0.050	56	依存したい	320	0.032
17	大阪	1065	0.108	37	JK	484	0.049	57	高収入	300	0.030
18	池袋	1025	0.104	38	援募	481	0.049	58	3P	299	0.030
19	家出	1013	0.103	39	オナ電	470	0.048	59	千葉	294	0.030
20	見せ合い	968	0.098	40	出会い	451	0.046	60	ぶち	289	0.029

No.	付帯タグ(#)	付帯数	付帯率	No.	付帯タグ(#)	付帯数	付帯率	No.	付帯タグ(#)	付帯数	付帯率
61	サポート	289	0.029	81	関東	194	0.020	101	一宮	149	0.015
62	jd	284	0.029	82	倉敷	189	0.019	102	巨乳	145	0.015
63	神奈川	276	0.028	83	家出少女	186	0.019	103	裏垢女子と繋	138	0.014
64	えっち	271	0.027	84	p活	183	0.019	104	エロ動画	136	0.014
65	SEX	269	0.027	85	エログループ	182	0.018	105	エン	134	0.014
66	裏アカ	258	0.026	86	関西	182	0.018	106	割	132	0.013
67	らぶりください	254	0.026	87	AV	178	0.018	107	ハメ撮り	130	0.013
68	裏アカ男子	241	0.024	88	複数	176	0.018	108	女子大生	126	0.013
69	JC	239	0.024	89	人妻	173	0.018	109	支援	125	0.013
70	バイト	236	0.024	90	ハッピーメール	172	0.017	110	尾張	121	0.012
71	京都	231	0.023	91	欲求不満	172	0.017	111	すすきの	120	0.012
72	岡山	231	0.023	92	なんば	171	0.017	112	モデル	116	0.012
73	LINE交換	218	0.022	93	通話	166	0.017	113	下着売ります	116	0.012
74	セフレ募集	215	0.022	94	副業	161	0.016	114	売子	115	0.012
75	オナ動画	212	0.021	95	デート援	158	0.016	115	円交	114	0.012
76	円交際	212	0.021	96	セックス	153	0.015	116	童顔	112	0.011
77	お助け	209	0.021	97	お小遣い	151	0.015	117	さぼ	109	0.011
78	滋賀	202	0.020	98	小岩	151	0.015	118	川崎	105	0.011
79	エッチ	201	0.020	99	アルバイト	150	0.015	119	プチ募集	103	0.010
80	福岡	201	0.020	100	デート	150	0.015	120	風俗	103	0.010

3-3-3 ハッシュタグの分類

図表7は、この265種類のハッシュタグを6つのカテゴリー「エリア示唆（エリアを示唆するもの）」「援助：間接的表現（援助交際を間接的に表したもの）」「援助：直接的表現（援助交際を直接的に表したもの）」「学生示唆（投稿者自身および投稿対象を学生ないしは未成年であることが明示されているもの）」「性描写」「その他」に分類し、まとめたものである。その割合は、「エリア示唆」のハッシュタグが最も多く21%、次いで「援助：間接的表現」が15%、「性描写」が13%、「援助：直接的表現」が5%、「学生示唆」が5%であった。

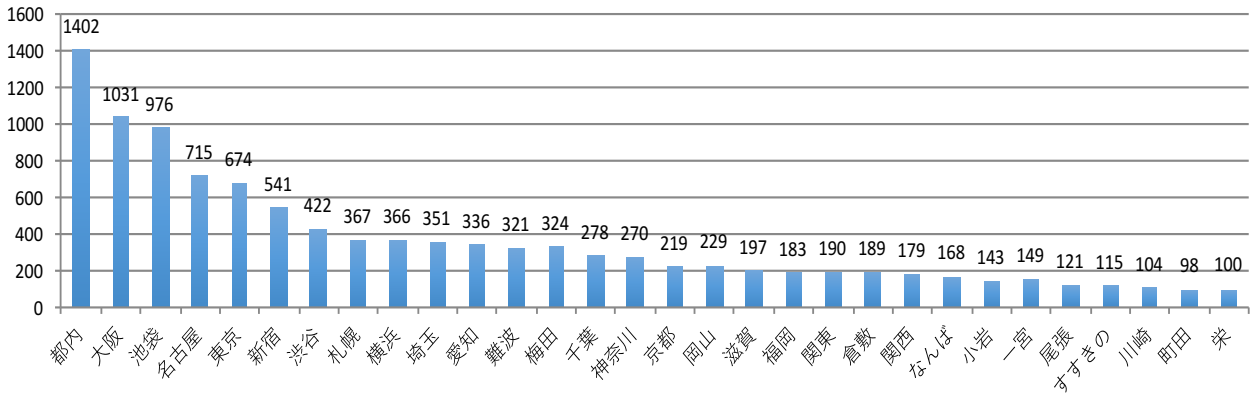
図表7. タグカテゴリーの種類と割合



3-3-4 “エリア” 関連ハッシュタグ

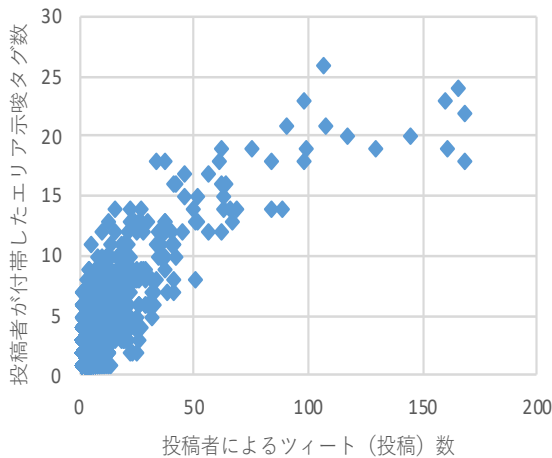
図表8は、「エリア示唆」のハッシュタグ33の付帯数順位をまとめたものである。問題投稿に最も使用されているエリアを示すハッシュタグは「都内」であり、次いで「大阪」、「池袋」であった。上位10エリアのうち7つのハッシュタグは関東圏内のエリアであることから、Twitterによる売買春活動は、関東圏の広域において盛んにすすめられていることがうかがえる。

図表 8. エリア関連ハッシュタグの付帯数順位



図表 9 は、投稿者による投稿数とその投稿に使用した「エリア示唆」ハッシュタグ数との相関図である。投稿数量が多いほど、その投稿に使用するエリア示唆のハッシュタグ数が多いことがわかる。売買春に関する投稿を大量に行っている投稿者の活動対象は広域であることがうかがえる。

図表 9. 投稿数量とエリア示唆タグ使用の関係
(相関係数=0.76)



4-2 調査結果

4-2-1 ハッシュタグ間の共起ネットワーク

共起ネットワーク分析の結果は、図表 10、11 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。これらの図表からは、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。こうした事実は過去の新聞報道等でも言及されており、改めてその事実が確認された。その他には地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。また、ハッシュタグ間の共起ネットワークは類似した意図で用いられているハッシュタグを把握することには有用であるものの、類似した意図で複数のハッシュタグが用いられている場合には、個々のハッシュタグ間の共起関係を過小評価してしまう懸念がある。

図 12.タグカテゴリー間の共起関係

/(Jaccard 係数=0.2)

図 13.タグカテゴリー間の共起関係

/(Jaccard 係数=0.1)

※図表 7 の区分「その他」に含まれるタグから「裏アカウント示唆」を取り出し追加した。

5 まとめ-売買春に関わる Twitter 上の発生リスクに関する考察と課題-

5-1 規模、投稿者像

調査期間の 30 日間における Twitter 上の売買春関連活動は、問題投稿（ツイート）数約 120,000 件、1 日平均約 4,000 件、投稿者数 約 30,600 名の規模で行われていた。このうち半数の問題投稿については、全投稿者の 7%の投稿者により行われたものであった。また、この 7%のなかに含まれる 64 名の投稿者については、調査期間の間に 100 件以上を超える大量の投稿を行っており、使用するエリア示唆のタグより活動対象も広範であった。この傾向から売買春の組織的活動を行う事業者である可能性が高いと考えられる。一方で、全投稿者のうち約 5 割については、調査期間中に 1 件の問題投稿を行ったのみであった。この 5 割の投稿者については、エリアを絞り、且つ「パパ活」「援」「家出」「泊めて」など具体的な支援を必要とする様子が見え、ハッシュタグを多様化する傾向が強く、多くが個人的関心により活動する個人である可能性が高いと考えられる。

5-2 経路（繋がり）、悩みの種別（問題テーマ）

Twitter 上の売買春関連ネットワークは、265 種類のハッシュタグで構成されていた。このうち「援」「円」「サポ」のハッシュタグは問題投稿の 3~5 割に含まれており、高確度で売買春活動者へたどり着くことができる。共起ネットワーク分析においても、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、「援」「円」「サポ」といった間接的に援助を求めるハッシュタグが中心的に用いられていることが明らかとなった。こうした間接的な援助を求める、いわゆる“隠語”を用いたハッシュタグと「家出」「泊めて」「神待ち」といったハッシュタグの共起関係は強く見られており、家出等で泊まる場所に困った青少年が Twitter を通して秘密裏に援助を呼びかける構図を読み取ることができる。加えて、こうした家出等に伴うハッシュタグ群と「パパ活」「デート援」「お小遣い」といった金銭的な支援を呼びかけるハッシュタグ群では異なる共起傾向を示しており、ハッシュタグの使用の差異が悩みの種別（問題テーマ）の差異を示唆していた。

タグカテゴリーに関する共起ネットワーク分析では、「援助：間接的表現」が他の多くのカテゴリーとの共起関係を有しており、隠語が中心的に用いられていることがこちらの結果でも支持されている。さらに、具体的な地名の入った「エリア示唆」は「援助：間接的表現」と強い共起関係を有しているため、この両者を含む投稿を抽出することで、より「売買春」の発生リスクが高い投稿を効率的に特定することが可能になる。なお、一連の調査結果より、多くの隠語がハッシュタグとして用いられていることが明らかとなっているが、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。そうした観点からも、ハッシュタグのデータリンク構造を通して全体像を把握することは有効なアプローチとなるものと考えられる。

5-3 リスク発見支援の情報分析システム開発に向けての課題

本調査では、Node-RED を用いて Twitter の問題投稿を監視し、特定のハッシュタグを抽出してデータベースに保存することを実現した。また、特定のハッシュタグの累積登場回数及び投稿者の投稿回数をカウントし、データベースに保存した。本件では、主として特定ハッシュタグが含まれるツイート（問題投稿）の投稿回数やそのツイートに含まれるハッシュタグの傾向を把握することが主たる目的であり、概ね目的は達成できた。今後はリスク発見支援のための機能を実装するための検討が必要である。

5-3-1 確度の高い Tweet の抽出

上述したとおり、問題投稿には投稿者の目的達成のために「エリア示唆」キーワードが含まれていることが多い。本件で把握できたそれらのキーワードをベースに、特定のハッシュタグに加えて、エリアキーワー

ドがハッシュタグまたは本文等に含まれているか確認して、データベースに登録することで、より確度の高いツイートを抽出することが可能になる。このエリアキーワードをデータベース化して、必要に応じてキーワードを追加できるようにし、このデータベースと照合してツイートを抽出する機能が必要であるとする。

5-3-2 ハッシュタグの追加

本件では「援助」、「援」、「円」、「サポ」のハッシュタグを初動監視対象とした。売買春のためのツイートの内容は日々変化することが推測される。今回は上述のキーワードを静的に監視することからはじめたが、変化やより有効なハッシュタグを監視対象とすることが早期のリスク発見には不可欠であるとする。したがって、問題投稿のツイート、且つ上述の静的キーワード以外の多頻度で用いられるハッシュタグを監視対象として動的に追加できる機能が必要となる。

5-3-3 投稿者の自動分類

投稿者としては、売り側（投稿内容から大半は女性と推察）と買い側（同様に大半は男性と推察）に大別される。ツイートのハッシュタグや本文のキーワードや助動詞や助詞、アイコン（画像）からそれらを自動的に推測し、分類することが、AI等を用いることによって技術的には可能である。この機能を開発して実装することで、リスクの種類（未成年の売春リスクや男性の買春リスクなど）を迅速に把握したうえで、対策を検討、実行に移すことが可能となる。

5-3-4 お気に入り登録分析

多数の買い側の投稿者のお気に入り登録を確認すると、自らが定期的に投稿を確認していると推察されるユーザが多数登録されている。その登録されているユーザはその内容から売り側であると考えられる。この買い側のお気に入り登録されているユーザを自動抽出し、データベースに登録して、その投稿内容を監視することが有効であると考えられる。また、お気に入り登録されている数が多いユーザは、リスクが高いことが推測されるため、お気に入り登録されている数もデータベースに格納し、監視の際に利用する機能が必要であるとする。

【参考文献】

- 1) 桂川泰典ほか（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
- 2) 警察庁（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」
- 3) 警察庁（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」
- 4) 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
- 5) 山本公平ほか(2006)「社会病理のリアリティ」学文社

〈発表資料〉

題名	掲載誌・学会名等	発表年月
問題投稿に付帯されるハッシュタグの特徴	社会情報学会全国大会にて発表	2018年8月
The pattern of victimization and social networking technology use	74 th American Society of Criminology Annual Meetingにて発表	2018年11月
問題投稿に付帯されるハッシュタグの特徴	情報処理学会第81回全国大会にて発表	2019年3月