

未成年女性のネットリスク分析

Analysis of Minor Women's Online Risk

平野 雄一 *¹
Yuichi Hirano

鳥海 不二夫 *¹
Fujio Toriumi

高野 雅典 *²
Masanori Takano

和田 計也 *²
Kazuya Wada

福田 一郎 *²
Ichiro Fukuda

*¹ 東京大学
The University of Tokyo

*² 株式会社サイバーエージェント
CyberAgent Inc.

With the spread of smartphones and SNS sites, online communication has become indispensable for social life. However, the risk of online communication for minors such as sexual crime and cyberbullying is increasing. In particular, many minor women suffer sexual crime through the Internet. In this research, we analyzed the communication among sex and age group using online chat service data. As a result, we showed that what kind of communication has high risk for minor women.

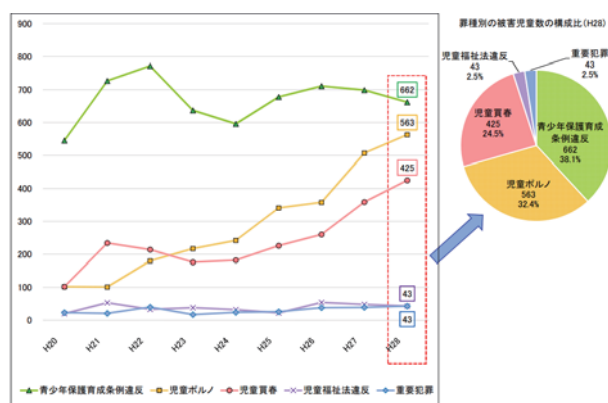


図1 被害児童数の推移 2(警察庁資料 [3] より抜粋)

1. イントロダクション

現在、世界中で普及しているインターネットは、情報を簡単に取得できるだけでなく、世界中の誰とでも簡単に会話できるといった利便性があることから、コミュニケーション手段としての重要度が高くなっている。メールやSNS、チャットアプリ等を始めたインターネットを通じたコミュニケーションは社会生活上、必要不可欠なものとなってきている。

この傾向は中学生、高校生を初めとした未成年者にとっても同様のことがいえる。総務省「平成 28 年情報通信メディアの利用時間と情報行動に関する調査報告書」[2]によると日本国で代表的なコミュニケーションツールの一つである LINE の 10 代における普及率は 79.3% であり、多くの未成年者がインターネットをコミュニケーションの場を利用して利用しているといえる。

しかしながら、オンラインでのコミュニケーションには、不用意な発言から炎上やネットいじめ [1] に発展したり未成年者が悪意を持った成人に誘い出され性犯罪の事件に発展したりする等、未成年者にとってのリスクも含んでいる。警察庁「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]によると、図 1 にあるように、オンラインでのコミュニケーションによって被害に会う未成年者の数は性犯罪を中心として年々増えている。

このような問題に対して、Toriumi らの研究 [5] では、コミュ

ニケーションペアに着目し送受信した文章の数や文字量等の情報を元に、ユーザーの年齢を含めずに潜在的な被害者と加害者の候補の抽出を行った。しかしながら、Wolak らの研究 [4] によると、オンラインで誘い出しが行われ性犯罪に至ったケースの多くが、未成年女性と 26 歳以上の成人男性であったので、潜在的な被害者や加害者特定のためには年齢や性別を考慮に入れることは重要である。一方で、ユーザーが偽りの年齢でサービスに登録している場合や、年齢を登録する必要のないサービスである場合等、運営会社側にとって正しい年齢が分からない場合も多い。

未成年女性にとって成人男性の交流全てが危険というわけではなく、オンラインでの自由な交流により多様な価値観に触れることができる等、彼女らの成長に繋がる部分も多いと考えられる。従って、彼らのマッチングのうち、リスクの高い交流と安全な交流を分離することが重要である。

本研究では、まずユーザーの年齢を推定した上で未成年女性と成人男性の交流を検出し、次に、彼らの交流形態がどのようになっているのかを求め、さらに、どういったユーザー同士のマッチングがリスクが高いのかについて発見することを目指す。

2. データ

2.1 サービス

本研究では株式会社 7gogo によって提供されている「755」*¹を用いた。「755」は、ユーザーが設立した「トーク」に、参加したユーザーが投稿したり、参加していないユーザーでも「やじうまコメント」と呼ばれるコメントを付与することにより、各ユーザーが交流することのできるサービスである。「755」の特徴は芸能人のユーザーが多いことと、「トーク」がオープンであり、どのユーザーからも投稿内容を見ることができることである。「755」においては一般的に、はじめに「トーク」の設立者や他のユーザーが「やじうまコメント」を通じて仲良くなり、その後そのユーザーが「トーク」に招待され、参加し、他の参加者たちと交流する。即ち、「トーク」での投稿は仲の良いユーザー同士の交流であることが多いと推察できる。

また、LINE や Twitter を始めた他のサービスの ID やメールアドレスなど個人情報や、公序良俗に反する内容等を含む投稿を「やじうまコメント」として投稿した場合は運営会社

*¹ <https://7gogo.jp/>

表 1 利用データ

名前	データ量	備考
ユーザー	191,448	投稿内容を含む
トーク	153,903	トークに対するやじうまコメントを含む

表 2 年齢分布

属性	未成年男性	未成年女性	成人男性	成人女性	合計
人数	126	193	241	319	879

表 3 年齢性別推定結果

label	適合率	再現率	F 値	データ数
未成年男性	0.65	0.22	0.33	126
未成年女性	0.67	0.82	0.74	319
成人男性	0.69	0.81	0.74	241
成人女性	0.64	0.54	0.59	193
平均	0.67	0.67	0.65	879

表 4 年齢性別推定分布

属性	未成年男性	未成年女性	成人男性	成人女性
人数	3,203	14,260	18,863	6,340

から削除される。本研究では、運営会社から削除されたコメントを「NG コメント」と定義し、NG コメントを投稿したユーザーを「NG ユーザー」と定義する。

2.2 使用データ

本研究では 2015 年 2 月 1 日から 2015 年 2 月 7 日までの 1 週間におけるユーザーによるチャット部屋（トーク）と、「トーク」へのコメント（やじうまコメント）への投稿文章データを用いた。但し、次節の求める年齢性別推定で用いたデータは学習データ量の確保のため、2014 年 12 月 1 日から 2015 年 3 月 31 日までのデータを用いた。

「755」ではトークにサービス上参加していなくてもコメントを通じて交流しているユーザーは多くいること、またサービス上参加していても殆ど活動していないユーザーもいるが、本研究でトークに参加したユーザーの定義は、トークに対して投稿した、もしくはコメントを行ったユーザーであるとする。

2.3 年齢性別推定

本研究ではトークに対して 20 件以上投稿したユーザーの年齢性別ラベルを、Hirano ら [6] と同様の手法で次のように取得した。

- 1. 年齢性別正解ラベルを投稿文章から取得
- 2. 正解データから TF-IDF とロジスティック回帰を用い [7]、年齢性別推定分類器を学習
- 3. 分類機から全ユーザーの年齢性別ラベルを推定

本研究で用いた正解データ数は表 2 に、10 分割交差検証による精度は表 3 に、推定したユーザーの年齢性別ラベルは表 4 に示した。

なお、本研究では未成年者の定義は 18 歳以下、成人は 19 歳以上とした。

3. 未成年女性のネットリスク

本章は、多数対多数のコミュニケーションの中で、どのような特徴を持つ交流で未成年者が高いリスクに晒されているかを解

表 5 投稿文章からの NG ユーザー推定

	適合率	再現率	F 値	データ数
一般ユーザー	0.75	0.76	0.76	3151
NG ユーザー	0.76	0.75	0.75	3151
平均	0.76	0.76	0.76	6302

表 6 NG ユーザーの特徴語 (上位)

順位	単語
1	LINE
2	カカオ
3	寝
4	ライン
5	寝る

明することを目的とする。まず、本研究におけるネットリスクを定義し、そのネットリスクについての考察を行った。続いて、現状のサービスではどういったコミュニケーションにおいてリスクが高くなるか分析した。最後に、潜在的にリスクの高いコミュニケーションに着目し、その言語的特徴について分析した。

3.1 ネットリスクの定義

本研究では、期間内に「1 件以上 NG コメントを投稿したことのあるユーザー」との交流をネットリスクであると定義する。

3.2 NG ユーザーの性質

まず初めに、NG ユーザーがどのような発言が多いか、単語ベースでの分析を行った。NG ユーザーと一般ユーザーそれぞれ 3,151 名ずつについて、年齢性別推定と同様に、投稿文章を TF-IDF からベクトル化し、ロジスティック回帰を用いて、10 分割交差検証で学習し、各評価指標を求めた。結果として表 5 を得た。

特徴語の分析から表 6 を得た。ここで、「LINE」や「ライン」は LINE 株式会社の LINE*2、「カカオ」はカカオ社のカカオトーク*3 というサービスを指していると考えられる。いずれのサービスも、利用者の非常に多いチャットアプリケーションであり、特に LINE は 10 代の 79.3% が利用している [2]。これらのアプリケーションではチャットが主要な機能であり、会話内容を第三者に閲覧されること無く利用できる。実際にオフラインで出会う場合、第三者に見られない場所で住所や待ち合わせの場所等の情報を交換すると考えられるため、これらのアプリケーションへの招待は誘い出しの第一歩であると考えられる。本結果から、それらの NG ユーザーには誘い出しを目的としたユーザーが多いと考えられる。さらに、図 1 にあるように、児童買春や児童ポルノといった誘い出しと関連性の深い要因が近年増大していることから、彼らと未成年者のマッチングの危険性が非常に高いことが分かる。

次に、NG ユーザーの年齢性別と、各年齢性別層に占める NG ユーザーの割合を求めた。分類結果を表 7、表 8 に示した。NG ユーザーの数については、成人男性が、1 ユーザーあたりの NG ユーザーの割合については未成年男性、成人男性が多いという結果になった。これらの結果は、本サービスにおいて、成人男性を中心として注意を払う必要があることを意味している。

*2 <https://line.me/ja/>
*3 <http://www.kakaotalk.jp/>

表 7 NG ユーザー年齢性別推定分布

属性	未成年男性	未成年女性	成人男性	成人女性
人数	352	583	2,013	203

表 8 1 ユーザーあたり NG ユーザー年齢性別割合

属性	未成年男性	未成年女性	成人男性	成人女性
人数	11.0%	4.09%	10.7%	3.20%

表 9 部屋の分類

グループ名	未成年女性の割合	成人男性の割合	トーク数
グループ 0	>0.5	>0	736
グループ 1	> 0	>0.5	841
グループ 2	> 0	>0	7,933
グループ 3	>0.5	0	4,295
グループ 4	0	>0.5	2,631
グループ 5	その他	その他	9,101

表 10 各グループでの NG コメントが書かれたトークの割合

	NG トーク数	NG トーク割合
グループ 0	9	1.22%
グループ 1	36	4.28%
グループ 2	24	0.30%
グループ 3	26	0.61%
グループ 4	16	0.61%
グループ 5	14	0.15%

3.3 トークの分割

トークにおいて、成人男性と未成年女性がどのような交流が行われているかを理解することを目的に、未成年女性と成人男性が参加しているかどうか、未成年女性か成人男性どちらかが参加人数において過半数を占めているかどうかという 2 つの基準を用いて 2 人以上が参加しているトークを 6 つに分割し、結果を表 9 に示した。例えば、グループ 1 は未成年女性、成人男性が参加していて、かつ成人男性が参加人数の過半数を占めているトークを表していて、グループ 4 は成人が参加しているが、未成年女性が参加していないことを示している。

表 9 の結果を元に、どのグループのリスクが高いかについて直接的リスク、潜在的リスクという 2 つの観点から調べた。

3.4 直接的リスクの分析

まず 1 つ目は、実際にリスクに遭遇した未成年女性がどのグループに多いかを調査することを目的とした。そのために、各グループにどれくらい直接 NG コメントが書かれたかを調べ、表 10 を得た。表 8 において、例えば、グループ 0 では NG コメントが投稿されたトークが 9 件であり、これはグループ 0 の全 736 個トークのうち、1.22% を占めていることを表している。NG コメントは個人情報や社会的問題のあるものを初めとした、実際に運営会社に削除されたコメントであることから、NG コメントがされたトークに参加していた未成年女性は実際に危険に巻き込まれたことを意味している。表 10 から、グループ 1 が量、割合両方の観点からリスクが高いことを示している。このことは、成人が多いトークに参加した未成年者は危険な交流に巻き込まれる可能性が高いことを示唆している。

表 11 各グループでの NG ユーザーの参加割合

グループ	未成年男性	未成年女性	成人男性	成人女性
0	0.00%	2.30%	1.63%	0.40%
1	3.45%	4.99%	10.34%	1.78%
2	0.19%	0.41%	0.50%	0.08%
3	0.02%	0.65%	0.00%	0.02%
4	0.19%	0.00%	1.36%	0.04%
5	0.07%	0.03%	0.06%	0.08%

3.5 間接的リスクの分析

次に、2 つ目は、潜在的に危険に巻き込まれる未成年者がどこに多いかということ調査することを目的とした。そのために、期間内に一度でも NG コメントを投稿したことのある NG ユーザーがどのグループに多いかを調査し、表 11 を得た。表 11 は各グループのトークの中で、どの年齢性別層の NG ユーザーが含まれていたかを表しており、例えば、グループ 1 のトークのうち、未成年女性の NG ユーザーが参加していたトークが 4.99% 存在し、成人男性の NG ユーザーが参加していたトークは 10.34% 存在することを表している。NG ユーザーは実際に NG コメントを投稿したことのあるユーザーであることから、ユーザーの性質的に今後も機会があればそのような投稿をする可能性の高いユーザーであると考えられるため、未成年女性にとって NG ユーザーとの交流は現時点で被害がなくても潜在的なリスクの高い交流であると言える。

表 11 から、どの NG 年齢層においてもグループ 1 が最も高い値となる事がわかった。特に、未成年女性が含まれているかという点で条件が異なるグループ 4 との比較から、成人男性のうち、危険なユーザーは参加する相手として、成人男性同士で交流しているトークよりも、未成年女性が参加しているトークの方を意図して選択して参加している可能性が高いということが分かる。従って、成人男性が多く参加しているトークに参加している未成年者は危険に巻き込まれる確率が潜在的にも高いということが分かる。

また、グループ 0 との比較により、成人男性は参加するものの、未成年女性が中心のトークよりも成人男性が中心のトークの方が、未成年女性自身から不適切な投稿をするユーザーが多いと考えられることが分かる。グループ 1 において、多くの NG 成人男性が存在していること未成年男性や成人女性においても高い値となることと合わせて考察すると、この原因には不適切な投稿が多い空間では不適切な投稿をする心理的障壁が小さくなるから、不適切な投稿を促すユーザーに誘導されてしまったから等といった可能性が考えられる。

3.6 考察

以上 2 つの分析から、成人男性が多く参加しているグループ 1 が実際に被害に合っている未成年女性が多い、危険性が高いトークであることがわかり、さらに潜在的にも高いリスクを抱えているトークであるということが分かる。実際に被害があったかどうかだけでなく、潜在的リスクの高い交流に注意を払うことも重要であると考えられる。

3.7 リスクの高いトークルームの特徴

前節まででグループ 1 がリスクの高いトークであるということが示された。本節ではグループ 1 の中でもどのようなトークにリスクが高いかを調べるために、NG ユーザーが 1 人以上参加したグループ 1 のトーク (NG トーク) と、1 人も参加していないトーク (一般トーク) の特徴語を調べた。それぞれ 90 トークずつについて、TF-IDF とロジスティック回帰を用いて分類

表 12 NG トーク推定

	適合率	再現率	F 値	データ数
一般トーク	0.71	0.77	0.74	90
NG トーク	0.75	0.69	0.72	90
平均	0.73	0.73	0.73	180

表 13 NG トークの特徴語 (上位)

順位	単語
1	?
2	し
3	!
4	です
5	こんばんは
6	～
7	(=ω)
8	ライン
9	たい
10	LINE

した。分類結果を示した表 12 から、直接 NG コメントが投稿されたとは限らない NG トークについても単語的な特徴によって判別することが可能であることが分かる。

また、回帰係数上位の単語を示した表 13 から、NG ユーザーと同様に「LINE」や「ライン」といった単語が上位に来ることが分かる。加えて、「?」が最上位であることから、NG トークは一般のトークに対して疑問文が多いということが推察できる。疑問文は一般的に、「好きな食べ物は何か?」、「どのアイドルが好きですか?」といった形で、相手から情報を引き出すために用いられるため、疑問文が多いチャット環境の方が各参加者の年齢や趣味といった個人の性質情報、LINEID 等の個人情報が流出しやすいと考えられる。即ち、「LINE」等や疑問文の係数が高いことから、NG トークは言語的特徴という観点から考えても未成年保護という観点からは好ましくない環境と言える。

これらの分析結果から、多数の大人男性と少数の未成年女性が参加しているトークの中でも NG ユーザーが参加しているかどうかによって、トークの単語的特徴に違いが見受けられることが分かり、NG トークは未成年女性にとって言語的に好ましくない環境であり、前節で考察したように直接的リスクだけでなく間接的リスクにも注意が必要ということが示された。

4. 結論

本研究は多数対多数のコミュニケーションについて直目し、どういったチャット空間が未成年女性にとってリスクが高いのかを分析した。本研究によって、年齢性別層の組み合わせの違いによって、ネットリスクの差異があることが示された。特に、多数の成人男性と少数の未成年女性が交流するチャット空間のリスクが高いことが示された。

さらに、多数の成人男性と少数の未成年女性が交流するチャット空間の中でも言語的分析によって危険な発言をしたことのあるユーザーが参加しているチャット空間について分析し、そのチャット空間が未成年女性にとって危険であることを示した。

これらの結果から、本研究は特に多対多のコミュニケーションにおける未成年女性のネットリスク理解に意義が大きく、ネットリスクの軽減に繋がる結果を得ることが出来たと言える。

今後の研究の方向性としては、次のようなものが考えられる。

まず、時系列的特徴に着目し、どういった環境が危険なユーザーのトークへの参加を促しているのかについての分析が考えられる。続いて、他サービスとの比較である。他の多数体多数のコミュニケーションを行うことのできるサービスを分析し、本研究の結果と比較することにより、

参考文献

- [1] B,Belsey. *Cyberbullying, An Emerging Threat to the 'Always On' Generation*.Recuperado el Vol 5, 2005
- [2] 総務省平成 28 年情報通信メディアの利用時間と情報行動に関する調査報告書.http://www.soumu.go.jp/main_content/000357570.pdf.22.11.2017 取得
- [3] 警察庁平成 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 https://www.npa.go.jp/cyber/statics/h28/h28_community_shiryou.pdf22.11.2017 取得
- [4] J,Wolak,;D, Finkelhor; and K.,Mitchell. *Internet-initiated sex crimes against minors: Implications for prevention based on findings from a national*” Journal of Adolescent Health, vol35, no 5, pp424–e11. 2004.
- [5] Fujio,TORIUMI;Takafumi,NAKANISHI; Mitsuteru,TASHIRO;Kiyotake,EFUCHI. *Encounters between Predators and Their Targets in Private Chat*.Journal of Transformation of Human Behavior under the Influence of Infosocionomics Society Vol1.2016.
- [6] Yuichi,Hirano;Fujio,Toriumi;Mitsuteru,TASHIRO; Kiyotaka,EGUCHI. *Finding Minors Faced with Online Risk*.Journal of Transformation of Human Behavior under the Influence of Infosocionomics Society Vol1.2018.
- [7] Genkin,A;Lewis,D,D;Madigan,D. *Sparse logistic regression for text categorization*.DIAMCS Working Group on Monitoring Message Streams,Project Report.2005.