

未成年者のネットリスク軽減に向けたソーシャルメディアの分析

東京大学大学院工学系研究科准教授 鳥 海 不二夫

TORIUMI Fujio

東京大学大学院工学系研究科 平 野 雄 一

HIRANO Yuichi

情報法制研究所 江 口 清 貴

EGUCHI Kiyotaka

- I 緒 言
- II デ ー タ
- III 年齢の推定
- IV 年齢層ごとのトーク分析
- V 年齢層によるトークペア分析
- VI 考 察
- VII 結 論
- VIII 謝 辞

I 緒 言

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という点だけでなく、オフラインで実際に会わなくても他人と簡単に交流できることから、コミュニケーションの基盤として利用されている。携帯電話やスマートフォンの普及とともに、メールやSNS、チャットアプリ等を始めとし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た。

その利用は中学生、高校生を始めとした未成年者にとっても日常的なコミュニケーションツールとして広まっている。「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[1]」によれば、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に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に発展したり未成年者が悪意を持った大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[2][3]。このような問題に対して、Toriumi らは[4]、送受信した文章の数や文字量等の情報を元に、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、この研究ではユーザの年齢情報については考慮していない。Wolak らの研究[3]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、未成年者と 26 歳以上の大人であることが分かっており、潜在的な被害者や加害者特定のためには年齢を考慮に入れることは重要である。

一方で、多くのネットサービスでは年齢を登録する必要がないため、ユーザが偽りの年齢でサービスに登録していることもあるため、正しい年齢が分からない場合が多い。従って、未成年者を性犯罪（誘い出し、児童ポルノ等）やネットいじめといったネットコミュニケーションリスクから保護するためには、まず、ユーザの年齢を推定した上で、大人と未成年者のオンラインでの交流を見つけ出すこと、そしてその交流形態の特徴を見つけ出すことが重要である。そこで、本研究は、ユーザの年齢情報が無い場合においても、年齢を取得した上で、未成年者と大人の交流を検出し、他の世代とのコミュニケーションがどのように行われているのかを分析する。

本論文では、日本で利用されているコミュニケーションツールの一つである LINE プレイを対象に分析を行った。まず年齢が取得可能な各ユーザについて未成年者、学生、大人のいずれかのラベルと投稿文章データが含まれた教師データを作成し、教師あり学習から年齢分類器を作成して、他のユーザを含めて年齢分類を行う。次に、求めたユーザの年齢情報に加え、各年代のユーザのコミ

ユニケーションツールの利用状況、および各年代間のコミュニケーションがどのように行われているのかを分析する。特に、利用時間やコミュニケーション量に着目し、分析対象となるコミュニケーションツールが未成年者にとってどの程度の潜在的リスクをはらんでいるのかを検討する。

II データ

1 利用データ

本研究では株式会社 LINE によって運営されているサービス、LINE プレイ¹⁾のデータを用いた。このサービスではユーザはアバターを作り、アバターの服を着せ替え、友達と交流し、各種ゲームを楽しむことができるサービスである。

このサービスにおいては、一般的にユーザはまず初めにスクエアという名前の多人数が参加する交流空間に入り、興味ある人を見つけ、ささやきトークという私的チャットを開始する(図1参照)。

本研究では、ささやきトークについて、2015年2月1日～2015年4月6日までの期間において、75,863 ユーザ 10,836,377 発言を対象に分析を行った。



図1 LINE プレイとささやきトーク

2 ささやきトーク分析

ささやきトークは一对一のチャットである。ささやきトークを行ったペアは 237,607 あり、二人

のユーザが行った一連ささやきトークについて基本統計分析を行った。

ささやきトークは図2のような流れで行われる。ささやきトークは話しかけた側と話しかけられた側がいるため、それぞれに発信側、受信側と呼ぶ。ここで、発信側はささやきトークを開始したユーザであり、受信側はささやきトークを受けた側である。たとえば、以下のような対話があったとする。

ユーザ A：こんにちは♪
ユーザ B：こんにちは！

このとき、ユーザ A が発信側、ユーザ B が受信側となる。受信側が全く返信を行わないこともあるが、その場合でもトークは成立したものととする。

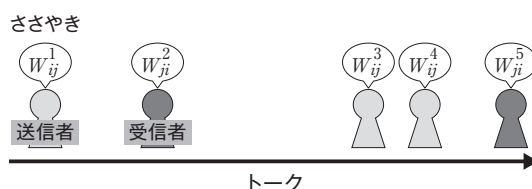


図2 ささやきトーク

片方のユーザが発言を行ってから、もう一方のユーザが発言を行うまでの時間を返信時間と呼び、一連のトークが開始してから最後の発言が行われるまでの時間をトーク時間と呼ぶ。図3に、各ユーザが行ったささやきトーク数の分布を示す。横軸がトーク回数であり、縦軸がその回数トークを行ったユーザ数である。ただし、トークにはトークを開始する送信側とトークを受け取る受信側があるため、それぞれについて両対数グラフにプロットした。これより、各ユーザが行うささやきトーク数はべき分布となることが分かる。すなわち、非常に大量のトークを行うユーザがいる一方で、多くのユーザはほとんどトークを行っていないといえる。事実、トークを送信した回数が一回以下

1) <http://lp.play.line.me/>

のユーザが全体の 70.5%を占め、トークを受信した回数が一回以下のユーザも 58.9%となっており、多くのユーザが高々 1 回の送受信しか行っていない。

次に、図 4 にささやきトークの利用時間を示す。横軸に時間を、縦軸にその時間に行われたささやきトークの一日あたりの平均値を示す。もっとも盛んにささやきトークが利用されているのは深夜帯であり、一分当たり 150 回以上のささやきが投稿されている。Twitter などのコミュニケーションツールでは、午前中や昼休みの時間帯にもそれなりに投稿があることが良く知られているが、チャットという一対一でリアルタイムに応答が求められるコミュニケーションツールの性質のためか、本データでは 19 時～深夜 3 時程度にピークがある以外は、主なピークは見つからなかった。

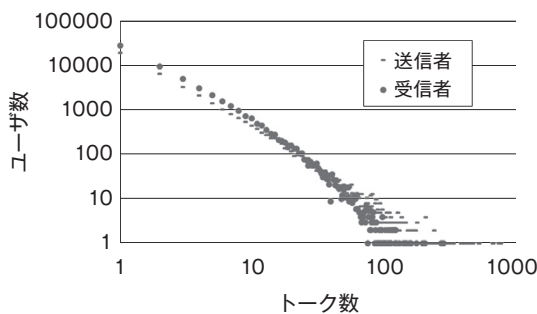


図 3 ささやきトーク回数

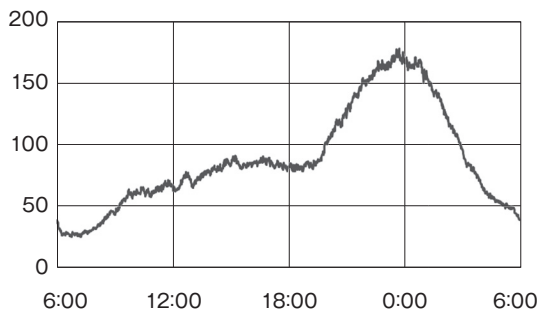


図 4 ささやきトークの利用時間

Ⅲ 年齢の推定

1 正解データの生成

LINE プレイにおいて、ユーザの年齢は公開されていないため、ユーザの年齢を文章データから推定した。

ユーザの年齢を取得する方法の一つに Zhang ら[5]が用いた、友人から送られた「happy y-th birthday」というフレーズから取得する方法がある。しかしながら、日本においては誕生日を祝う時に「〇〇歳の誕生日おめでとう」等と言わないことが多いため、この手法をそのまま適用するのは難しい。

そこで、本研究ではユーザ自身が「私は n 歳です」などという投稿によって、そのユーザに単に n 歳とラベルを付けるという手法を採用した。この手法によって 12 歳から 50 歳までと分かったユーザのうち 3,900 人について、日本国の修業年齢に基づいて、

- ・ 18 歳未満：未成年
- ・ 18 ～ 24 歳：学生
- ・ 25 歳以上：大人

として、各ユーザに対してラベル付けを行った。なお、学生ラベルは年齢層から想起される身分を名前として付けたものであり、必ず学生であることを保証するものではないことに注意されたい。

2 SVM による年齢推定の精度評価

LINE プレイのユーザ全てが自らの年齢を申告しているとは限らない。そこで、ユーザの対話内容から年齢を推定する。たとえば、未成年であれば「宿題」や「部活」という言葉が出てくる可能性が高いと推測され、大人であれば「残業」や「子供」といった話題が多く出ると考えられる。

そこで、各ユーザが対話で用いた単語に基づいたユーザの年齢推定を行った。

まず、各ユーザの対話データに基づいて、対話に出てくる単語に基づいてユーザの特徴ベクトルを生成する。この際、ユーザに特徴的な単語を重視するため、TF-IDF 法を用いた。

tf-idf は、tf (Term Frequency: 単語の出現頻度) と idf (Inverse Document Frequency: 逆文書頻度) の二つ

の指標に基づいて以下の様に算出する。

$$tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}}$$

$$idf_i = \log \frac{|D|}{|\{d: d \ni t_i\}|}$$

$$tfidf_{ij} = tf_{ij} \cdot idf_i$$

これによって、単に単語の出現頻度ではなく、当該ユーザの発言で特有の単語には高い tfidf 値が付けられ、ユーザの特徴とみなすことができる。

次に、各ユーザが持つ単語 tfidf 値をユーザの特徴量として、2 クラス分類として代表的な手法である Support Vector Machine (SVM) [6] を用いて年齢推定を行った。なお、SVM のカーネルには多項式カーネルを用いた。

本手法の精度を確認するため、年齢が既知のユーザ 3,900 人のうち 3,510 人を学習用に使い、のこり 390 人で判別を行い正しく判別されているかどうかを確認した。

結果を表 1 に示す。適合率は各年齢層と判別されたユーザが実際にその年齢層だった割合を、再現率は実際にその年齢層だったユーザのうち正しく判別された割合を示している。また、F 値は適合率と再現率の調和平均であり、

$$F = \frac{2 \cdot \text{適合率} \cdot \text{再現率}}{\text{適合率} + \text{再現率}}$$

によって算出される尺度である。

この結果、本手法によって未成年 (18 歳未満) と大人 (25 歳以上) について 80% 以上の高い適合率で判別可能であることが明らかとなった。

表 1 年齢推定の精度

	適合率	再現率	F 値	データ数
18 歳未満 (未成年)	0.84	0.93	0.88	212
18 ~ 24 歳 (学生)	0.60	0.41	0.49	63
25 歳以上 (大人)	0.80	0.78	0.79	115
平 均	0.79	0.80	0.79	390

3 年齢推定結果

得られた SVM 分類機によって全ユーザについて年齢を求めた。その結果得られた年齢分布を表

2 に示す。

表 2 年齢推定結果

	人 数	人数比
18 歳未満 (未成年)	37,933	0.500
18 ~ 24 歳 (学生)	7,569	0.100
25 歳以上 (大人)	30,361	0.400

年齢推定の結果、未成年、大人が大半を占め、18 ~ 24 歳の学生数は多くないと判断された。ただし、学生の判別精度は高くはないことから、実際にはこの年代のユーザ数はもう少し多い可能性がある。しかしながら、本研究の主な目的は未成年者と大人とのコミュニケーションからリスクを判断することにあるため、大きな問題とはならないと考えられる。

IV 年齢層ごとのトーク分析

1 年齢別によるトーク数およびトーク長

各年齢層のユーザがどの程度 LINE プレイのささやきトークを行っていたのか、その平均を表 3 に示す。各年齢層のユーザが平均何回トークを開始したかを送信数、何回トークを受けたかを受信数としてカウントした。また、送受信数をトータルトーク数とした。

この結果より、未成年 < 学生 < 大人の順でトーク参加数は増加していることが分かる。トークを開始するか他人に話しかけられるかを見ると、未成年者は自ら開始するよりも受けることのほうが多い。一方で、18 歳以上のユーザは話しかける回数のほうが多い。

表 3 年齢層別トーク数

	トーク開始数 (送信数)	被トーク数 (受信数)	トータル
18 歳未満 (未成年)	1.72	1.83	3.55
18 ~ 24 歳 (学生)	2.84	2.72	5.56
25 歳以上 (大人)	4.87	4.72	9.59

2 年齢別利用時間分析

各年代に推定されたユーザがどのくらいの時間帯にLINE プレイを利用しているかを発言時間から求めた。その結果を図5に示す。横軸に発言が行われた時間を、縦軸に割合を示す。すなわち、未成年者であれば全発言のうちの8%程度が21時台に行われていることになる。これより、年代によって利用時間のピークが異なることが明らかとなった。未成年者は21時ごろに最も利用が多く、一方大人は0時付近がピークとなっている。

この結果は、未成年者の利用は比較的早い時間で終わることを意味しており、ある程度健全な使い方をしていることを示唆している。その一方で、深夜までLINE プレイを利用している未成年者がいることも示されている。

そこで、次に深夜0時以降に一度でもささやきを行ったユーザが各年代でどの程度いたのか、その割合を求めるため、一日を0-6時、6-12時、12-18時、18-24時の四区間に分け、それぞれの区間で一度でも発言したことのあるユーザの割合を年齢別に算出した。その結果を図6に示す。これより、未成年者は12-24時の利用率が高く、0-6時の深夜帯の利用率は低いことが明らかとなった。大人や学生は0-6時の利用率も比較的高く、最も利用率が低いのは6-12時であることと対称的である。

以上の結果から、大人と未成年ではLINE プレイにおけるささやきトークの利用時間帯のピークは、ずれているといえる。

このようなずれによって、大人と未成年者がトークを行う可能性が同年代よりも低くなると予想される。一方で、深夜帯など同年齢層のユーザが少ない時間帯にささやきトークを行う未成年者にとっては、コミュニケーション対象が大人である可能性が高いことを意味し、その意味ではコミュニケーションにリスクが増加する可能性もある。

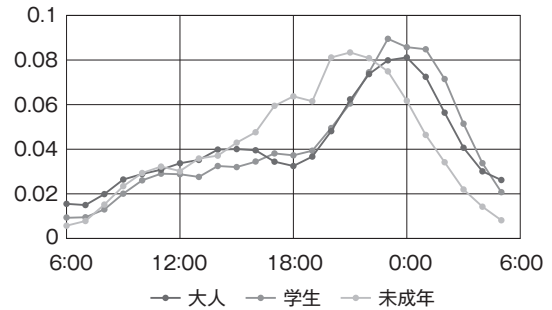


図5 年代別利用時間

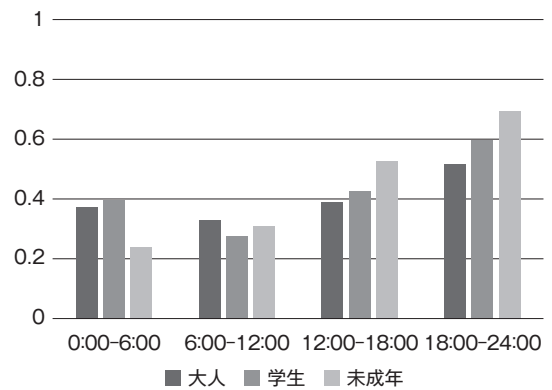


図6 時間帯別利用者割合

3 年齢層とユーザクラス

次に、各年齢層のユーザが、Toriumi らのユーザ分類結果[4]において得られたクラスタのうちどのクラスタに所属しているかを求めた。なお、各クラスタの主な特徴は表4に示すとおりである。

その結果を図7に示す。横軸は[4]にて定義されたクラスタを示し、縦軸は各年齢層のユーザが当該クラスタに所属する割合を示している。ただし、トークを一回しか行っていないユーザはコミュニケーションに積極的ではないユーザとして、その他ユーザとしている。ここで、Cluster0-3は主にトークを持ちかけられることが多い「受信クラスタ」、Cluster5-8はトークを開始することが多い「発信クラスタ」と定義されている[4]。

この結果から、未成年者はCluster0, 6, 7に所属することが少なく、大人はCluster1, 2, 7が多いことが分かる。このうちCluster7は利用がトークの発信、受信ともに数が多く、発言数も多い

ことから活発な利用者ということができるため、そのような利用者は未成年者には少ないといえる。特に、発言数が多い上位3クラスタ(0, 6, 7)に関しても、未成年者が所属する割合が特に小さい。したがって、未成年者の中にはささやきトークを積極的に使うようなユーザは相対的に少なかったということが分かる。

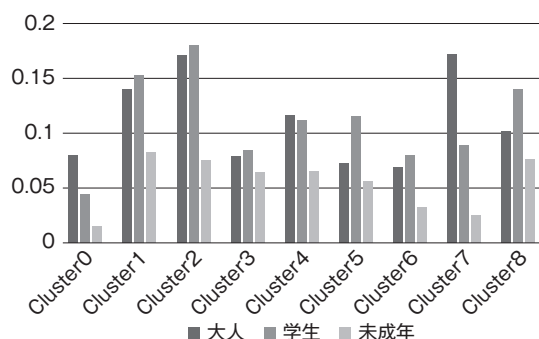


図7 年齢層別クラスタ所属割合

ただし、これはもともとの各クラスタに所属するユーザの割合を考慮していない。そこで、大人、学生、未成年の人数から、各クラスタに所属する年齢層別ユーザ数の期待値からの乖離率を求めた。ここで、期待値とはランダムにユーザを選んだときに、当該クラスタに所属する各年齢層のユーザ数である。乖離率は、期待値通りであれば1.0となり、期待値より多ければ1.0以上、期待値より少なければ1.0以下の値となる。その結果を図8に示す。これより、未成年者はCluster0, 7に所属する確率が期待値よりも有意に低い(有意水準1%)ことが明らかとなった。逆にCluster3, 5, 8に所属する確率が期待値よりも有意(有意水準1%)に高い。これらのクラスタは発言数が少ない3クラスタである。すなわち、未成年者はトークを行うとしても、その発言数は少ないといえる。

表4 各クラスタの主な特徴

	送信数	受信数	発言数	被発言数
Cluster 0	2.7	13	344.7	379.3
Cluster 1	1.0	2.5	42.5	46.8
Cluster 2	0.0	4.2	58.9	70.6
Cluster 3	0.0	2.0	15.2	19.0
Cluster 4	3.0	3.4	99.5	102.8
Cluster 5	3.0	0.0	39.9	34.3
Cluster 6	9.1	2.9	238.5	224.6
Cluster 7	13.4	9.4	604.1	604.3
Cluster 8	2.0	1.0	36.3	34.7

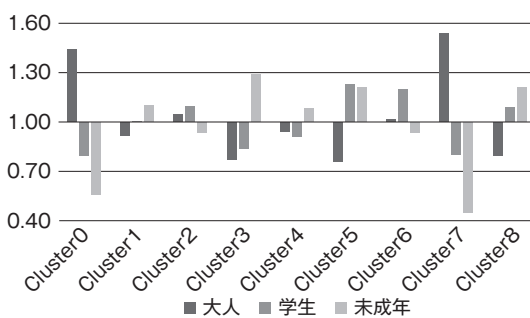


図8 クラスタ所属数の期待値からの乖離率

V 年齢層によるトークペア分析

1 年齢層別トークペア生成率

本研究の目的は、未成年者を誘い出す、あるいは個人情報を得るなどの目的を持った大人と未成年者との接触を防ぐことにある。そこで、どのような年齢層のペア間でささやきトークが行われているのかを分析する。本節では、各トークペアにおいて、どちらがトークを始めたかという点に着目する。

まず、表5に、各年齢層から各年齢層に対してのトークが何回行われたかを示す。例えば、送信者が未成年で受信者が大人、すなわち未成年が大人に話しかけた回数が22,619回、大人が未成年に話しかけた回数が26,873回である。

これより、未成年者同士、大人同士のトークが多いことが分かる。特に、大人同士のトークは未成年同士のトークの3倍近くあり、未成年者のほ

うが大人よりもユーザが多いことを考慮すると、トークを主に使っているのは25歳以上の大人であり、大人同士の会話に利用しているといえる。

そこで、次に未成年者、学生、大人の数から算出されるトークペア期待値との、実際のトーク数との比較を行った。ここで、トークペア期待値とはささやきトークを行った75,863ユーザからランダムに2ユーザを選んだ時にできる各年齢層ペアの確率である。例えば、未成年者と未成年者によるトークが行われるトークペア期待値は、

$$E_{mm} = \frac{N_m}{N} \cdot \frac{N_m}{N} \cdot T = 58,668$$

となる。ただし、 E_{mm} は未成年同士のトークペア期待値、 N_m は未成年者数、 N は全ユーザ数、 T は総トーク数である。

実際のトークペア数とトークペア期待値の比を表6に示す。比が1.0であれば、期待値と同じ回数だけのトークが行われたことを意味する。

これより、学生と大人、学生同士のトークはほぼ期待値通りであるが、未成年が加わるトークの回数は期待値の50～65%程度であることが分かる。ここから、未成年者がトークを行う割合は1%の有意水準で少ないことが明らかとなった。一方で、大人同士のトークの発生確率は期待値の3倍程度と大きい。

以上より、ささやきトークにおいて主にコミュニケーションを行っているのは大人であり、大人と未成年者のコミュニケーションは期待値より少ないことから、LINEプレイは未成年者にとって危険なコミュニケーションツールとは言えないことが分かる。

表5 トークペアの年齢層

		受信者		
		未成年	学生	大人
送信者	未成年	35,155	6,641	22,619
	学生	7,555	2,758	11,096
	大人	26,873	11,191	109,475

表6 トークペアの期待値との比

		受信者		
		未成年	学生	大人
送信者	未成年	0.599	0.567	0.482
	学生	0.645	1.181	1.184
	大人	0.572	1.194	2.913

2 トーク開始時間によるトークペアの年齢分布の変化

次に、時間によって未成年者がトークを行う対象が変化するかどうかを確認した。図9に、未成年者が対象となる各時間に開始されたトークの相手の割合を示す。横軸が時間であり、縦軸がその対象となる年齢層の割合である。ここから、未成年者が最も多くささやきトークを利用している16～22時の時間帯は比較的大人とのトークは少ない。しかし、それ以外の時間帯では大人とペアになる可能性は高い。

未成年者と大人とでは、ささやきトークを利用する時間帯のピークにずれがあることは分かっている。本分析からそれぞれのピークに利用しているユーザは同年齢層のユーザとコミュニケーションをとる可能性が高いが、ピークを外れた時間帯に利用するユーザは潜在的リスクが増加することが示された。

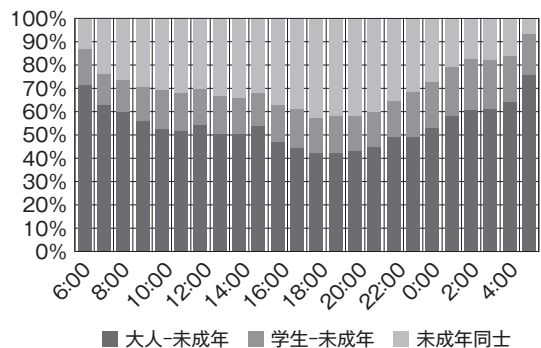


図9 時間別未成年者のトーク相手の割合

VI 考 察

1 年齢の推定

未成年者のネット利用リスクを評価するため、LINE 株式会社が提供するコミュニケーションツール LINE プレイのプライベートチャット機能「ささやきトーク」について分析を行った。当該サービスでは、ユーザの年齢が登録されていないため、まずプライベートチャットの利用者の年齢推定を、各ユーザの発言に基づく SVM による判別によって行った(Ⅲ章)。チャット中に年齢を明かしているユーザのデータを教師データとして推定手法の評価を行ったところ、約 80% の精度でユーザを 18 歳未満 (未成年)、18 ～ 25 歳未満 (学生)、25 歳以上 (大人) にクラス分類することができると明らかとなった。

また、各年齢層に分類した各ユーザの利用時間を確認したところ (Ⅳ 2 節)、未成年者の利用時間が夕方から 22 時程度にピークを迎え、提案手法によっておおむね正しくユーザの分類ができていることが示唆された。

2 未成年者の利用リスクの評価

推定された年齢に基づいて、ユーザを未成年者、学生、大人に分類しささやきトークの分析を行った。

その結果、未成年者のささやきトーク利用者は多いものの、積極的な利用者数は多くないことが明らかとなった。プライベートチャットは他人の目が届かない空間であることから、そこでのコミュニケーションが少ないことはリスクが低いことを意味する (Ⅳ 1 節)。

特に、ささやきトークの利用者は多いということは、少なくとも一度はささやきトークを利用していることを意味するにも拘わらず、トータルの利用回数は少ないことから、ささやきトークの存在は知ったうえで利用していないということの意味する。他のサービスとの比較をしなければ確実なことは言えないが、少なくともプライベートチャットというリスクを含むコミュニケーション手段の利用率が未成年者において低いことは、リスクマネジメントの観点から適切であるといえる。

次に、ユーザごとにどのような利用方法をしているかをクラスタリングの手法を用いて分析した。その結果、未成年者の多くがプライベートチャットを開始したとしても長くは続けず、かつその相手の数も少ないクラスに所属していることが明らかとなった。

Toriumi らの研究[4]によって話しかけられやすいユーザ群「受信クラスタ (表 4 における Cluster0-3)」が存在することが明らかとなっていたが、その中でも発言数が多い、すなわちトークに応じているクラスタ (Cluster0) に所属する未成年者の割合は期待値と比較して大幅に低く、話しかけられても相手にしない未成年者が多いと言えよう (Ⅳ 3 節)。

さらに、未成年者と大人がペアとなったトークそのものも期待値よりも低い確率でしか生じておらず、その意味でも未成年者が危険度の高いコミュニケーションをしているとは言えない。特に、大人同士のペアが期待値の 3 倍近く生じていることは、本サービスが主として大人同士のコミュニケーションに利用されていることを示唆している (Ⅴ 1 節)。

ただし、時間別にみると深夜になると未成年者と大人のペアによるトークの割合が増加していることから、時間帯によっては注意が必要であるといえるだろう (Ⅴ 2 節)。

VII 結 論

本研究では、明示的な年齢情報がない LINE プレイにおいてユーザの年齢層の推定を行った。その結果、およそ 80% という高い精度で年齢層の推定が正しく行えることが確認できた。また、推定結果を用いてプライベートチャットの利用について年齢層別の特徴を分析することで、未成年者の利用、また年齢層によるペアのでき方やその利用状況について統計的な特徴を分析した。その結果、LINE プレイにおけるプライベートチャット機能であるささやきトークは、未成年者の潜在リスクが高いとは言えず、特に危険なコミュニケーションツールであるとは言えないことが明らかとなった。

一方で、未成年者と大人とのマッチングが防が

れているわけではないため、そのようなコミュニケーションの中に潜在的なリスクを持つコミュニケーションがどの程度あるのか、より詳細な分析を行う必要がある。さらに、リスクを持つコミュニケーションをどのような手段によって防ぐかを検討し、未成年者にとってより安全なコミュニケーションツールを実現するための手段を明らかにする必要がある。

VIII 謝 辞

本研究の遂行に当たり、データを提供していただいた LINE 株式会社に感謝する。

【参考文献】

- [1]. 情報通信政策研究所. 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 (2016)
- [2]. B. Belsey. Cyberbullying, An Emerging Threat to the 'Always On' Generation. Recuperado el Vol. 5, 2005.
- [3]. J. Wolak; D. Finkelhor; K. Mitchell. Internet-initiated sex crimes against minors: Implications for prevention based on findings from a national, Journal of Adolescent Health, vol. 35, no.5, pp. 424-e11. 2004.
- [4]. Fujio TORIUMI, Takafumi NAKANISHI, Mitsuteru TASHIRO, Kiyotaka EGUCHI. Encounters between Predators and Their Targets in Private Chat. Journal of Transformation of Human Behavior under the Influence of Infosociconomics Society Vol.1. 2016.
- [5]. Zhang, J. ; Hu, X. ; Zhang, Y. ; Liu, H. Your Age Is No Secret: Inferring Microbloggers' Ages via Content and Interaction Analysis. ICWSM, 2016.
- [6]. Cortes, C. and Vapnik, V. Support-vector networks, Machine learning, Vol. 20, No. 3, pp. 273-297 (1995).
- [7]. Genkin, A. ; Lewis, D.; Madigan, D. Sparse logistic regression for text categorization. DIAMCS Working Group on Monitoring Message Streams, Project Report, 2005.
- [8]. Sara, R.; Kathleen, M. Age Prediction in Blogs: A Study of Style, Content, and Online Behavior in Pre- and Post-Social Media Generations. In Proceedings of ACL'11, Portland.