

## オンラインチャットサービスにおける未成年者検出

## Detection of Minors in Online Chat Service

平野 雄一\*<sup>1</sup>  
Yuichi Hirano鳥海 不二夫\*<sup>1</sup>  
Fujio Toriumi高野 雅典\*<sup>2</sup>  
Masanori Takano和田 計也\*<sup>2</sup>  
Kazuya Wada福田 一郎\*<sup>2</sup>  
Ichiro Fukuda\*<sup>1</sup>東京大学  
The University of Tokto\*<sup>2</sup>株式会社サイバーエージェント  
CyberAgent Inc.

Online communication is essential not only for adults but also for such minors as junior high school and high school students. However, for minors, interaction with adults has some risks. In many cases of Internet-initiated sex crimes, a victim is a minor and a predator is a minor. Therefore, it is very important to estimate users' ages and to investigate communications between adults and minors. However, in many online service, users' ages is not known. In this study, we proposed a method to estimate ages of users from text data. Then, we showed that the number of communications between adults and minors is more than we initially envisioned.

## 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という点でも非常に便利なツールである。そのため、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おり、成年だけでなく、中学生、高校生を始めとした未成年者にとっても一般的に広まっている。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 [1] [2]。

このような問題に対して、鳥海らの研究 [3] では、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 [2] 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 歳以上の大人であったので、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、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も多く、正しい年齢が分からない場合も多い。従って、未成年者を性犯罪を始めとしたオンライン交流のリスクから保護するためには、まず、ユーザー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て出すこと、そしてその交流形態の特徴を見つけて出すことが重要である。そこで、本研究は、ユーザー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、他の世代同士の交流よりも危険度が高いことを示すことを目指す。

本研究では、年齢が取得可能な各ユーザーについて未成年者、大学生、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、他のユーザーを含めて年齢分類を行った。次に、求めたユーザーの年齢情報に加え、ユーザー間で交流することのできるチャット部屋と、各部屋の参加ユーザーの情報を元に各チャット部屋に参加しているユーザーの年齢分布を算出し、各チャット部屋を分布毎に分類した。そして、分類毎に他のユーザー

から話しかけられる頻度の違いを算出した。

## 2. データセットの構築

## 2.1 用いたサービス

本研究では、日本の 7gogo, Inc. によって提供されている 755(<https://7gogo.jp/>) というサービスにおけるユーザーによる「トーク」という名のチャット部屋への投稿の文章データと、「やじうまコメント」という名の「トーク」へのコメントを用いた。755 では、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」をすることにより、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、「トーク」がオープンであり、どのユーザーからも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」の設立者と他のユーザーが「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、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を「やじうまコメント」でした場合はサービスプロバイダから削除される。個人情報が流出することは、誘い出し等の事件に発展する可能性を生むので、このようなコメントの書かれた「トーク」は誘い出しに繋がる危険な「トーク」であると分かる。

## 2.2 用いたデータ

本研究では、以下の表 1 のデータを用いた。ただし、ユーザーについては「トーク」に 1 件以上文章を投稿したユーザーを表し、文章データを含み、部屋は「トーク」を表し、コメントは「やじうまコメント」を表す。

表 1: データ

名前	データ数	期間
ユーザー	212,709	2014/12/1 から 2015/3/27
部屋	481,473	2014/12/1 から 2015/1/26
コメント	10,000,000	2014/12/1 から 2015/2/7

## 2.3 年齢データの取得方法

このサービスではユーザーの年齢は公開されていないので、ユーザーの年齢を文章データから取得した。ユーザーの年齢を取得する方法の一つに Zhang ら [4] が用いたように、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が、Zhang らが対象とした twitter と異なり、755 においては、個人宛にメッセージを送ることが出来ない、オフラインの友人同士での利用が少ない、日本においては誕生日を祝う時に年齢を言わないことが多い、といった理由があるのでこの手法を用いることは出来ない。よって、本研究では、ユーザーの「私は n 歳です」という投稿によって、そのユーザーに 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ーのうち 5500 人について、日本国の修業年齢に基いて、

- ・ 18 歳未満を中高生
- ・ 18 歳以上 25 歳未満を大学生
- ・ 25 歳以上を大人

として、各ユーザーに対してラベルをつけた。結果として、表 2 の量のデータを得た。

表 2: 年齢分布

年齢層	中高生	大学生	大人	合計
人数	2344	2249	907	5500 人

## 3. 年齢分類の手法と結果

本研究では、Genkin ら [5] や、Sarar ら [6] と同様に、文章データから TF-IDF 行列を作成し、ロジスティック回帰を用いてラベルの付いたユーザーの年齢分類を行った。まず、各ユーザー  $i$  のトーク文章のデータから数字を除いた名詞のみを取り出した“bag-of-words”ベクトル  $N_i$  を作成した。それを元に、TF-IDF 行列  $X$  を求めた。本研究では、TF-IDF 行列  $X$  の各要素  $X_{i,j}$  は

$$TF_{i,j} = (N_i \text{ の中で単語 } j \text{ が出て来る回数}) \quad (1)$$

$$IDF_j = \log \frac{(\text{文章の数}) + 1}{(\text{単語 } j \text{ が出現した文章の数}) + 1} \quad (2)$$

として、

$$X_{i,j} = TF_{i,j} \times (1 + IDF_j) \quad (3)$$

と求めた。そして、求めた TF-IDF 行列を元に、各ユーザーの TF-IDF ベクトル  $v_i$  を

$$v_i = \frac{X_i}{\sqrt{\sum_j |X_{i,j}|^2}} \quad (4)$$

と求めた。次に、各ユーザーの TF-IDF ベクトル  $v_i$  と、年齢ラベルを元に多項ロジスティック回帰を用いて年齢を分類した。今回は、過学習を避けるため、L1 正則化、即ち Lasso 回帰 [5] を用いた。

5500 件のデータのうち、4950 件を訓練データ、550 件をテストデータに分割してスコアを出した。その結果、表 3 の値が得られ、特に中高生については高い精度で年齢を推定することが可能であることが明らかとなった。実際に、これらの値は表 4 に表した Zhang らの研究 [4] に比べても高い精度で推定出来ている。

表 3: 投稿文章からの年齢の分類

年齢層	精度	再現率	F 値	データ数
中高生	0.89	0.90	0.90	237
大学生	0.82	0.85	0.84	222
大人	0.81	0.70	0.75	91
平均	0.85	0.85	0.85	550

表 4: Zhang らの結果

年齢層	精度	再現率	F 値
平均	0.8069	0.8349	0.8138

## 4. 年齢グループ毎の交流

### 4.1 年齢分布を用いた部屋の分類

前章でユーザーの年齢をある程度分類できることが分かった。そこで、本章ではこの分類器を他のユーザーにも適用していき、チャット部屋毎に年齢分布を調べる。まず、チャット部屋毎に、その部屋に参加しているユーザーの年齢層を分類器から推定した。結果として、この分類器によって 212709 名のユーザーを分類し、各年齢層について、表 5 の人数を得ることが出来た。

表 5: 年齢分布

年齢層	中高生	大学生	大人
人数	71831	118321	21557

次に、年齢が推定できた 2 人以上のユーザーが参加した部屋について、表 6 のルールに従って、7 個のグループに分類をした。

このようなルールによって合計 44970 部屋を分類した結果、各グループの部屋の数、各グループが占める割合について、表 7 が得られた。

また、比較のために、次のようなシミュレーションを行った。まず、表 5 の各世代の人数の割合から、確率分布を導いた。次に、各部屋のユーザーの年齢をその確率分布に従って推定する。それによって生成された、部屋のユーザーの年齢分布と、表 6 のルールに従って部屋をグループに分ける。このシミュレーションによって表 8 が得られた。この結果と表 7 を比較すると、グループ 0,1,2,3,5 に関して統計的に有意な差が確認できた。

中高生と大学生、大学生と大人といった近い世代同士の交流は興味、関心の近さという観点や、年齢層の境界近くの人との誤分類という観点から自然である。一方で、大人と中高生が多い部屋は、オフラインの世界では友人になる可能性は低いため、オンライン特有の交流形態と考えられ、性犯罪の多い組み合わせなので、中高生保護のために特に注意が必要になってくる部屋である。このような観点から考えると、大人と中高生が多く、大学生が少ない部屋が比較実験よりも有意に多いことが確認できたことから、実際に、危険な交流は多く行われていることが分かる。

### 4.2 部屋のグループとコメントの関係

どのような部屋に話しかけるユーザーが多いのかを調べるために、コメントと部屋のグループの関係について調べた。

部屋のグループごとに 1 部屋あたりの平均コメント数について計算した結果、表 9 が得られた。

表 6: 部屋の分類

グループ名	特徴	中高生の割合	大学生の割合	大人の割合
0	中高生のみが多い	0.40 以上	0.30 未満	0.30 未満
1	大学生のみが多い	0.30 未満	0.40 以上	0.30 未満
2	大人のみが多い	0.30 未満	0.30 未満	0.40 以上
3	中高生と大学生が多い	0.40 以上	0.40 以上	0.20 未満
4	大学生と大人が多い	0.20 未満	0.40 以上	0.40 以上
5	大人と中高生が多い	0.40 以上	0.20 未満	0.40 以上
6	多い世代が存在しない	その他の部屋		

表 7: 部屋の割合

	特徴	部屋の数	割合
0	中高生のみが多い	7291	16.2%
1	大学生のみが多い	11439	25.4%
2	大人のみが多い	916	2.0%
3	中高生と大学生が多い	14486	32.2%
4	大学生と大人が多い	4261	9.5%
5	大人と中高生が多い	3044	6.8%
6	多い世代が存在しない	3533	7.9%

表 8: 理論値

	特徴	部屋の数	割合
0	中高生のみが多い	4839	10.8%
1	大学生のみが多い	14135	31.4%
2	大人のみが多い	460	1.0%
3	中高生と大学生が多い	15105	33.6%
4	大学生と大人が多い	4313	9.6%
5	大人と中高生が多い	2484	5.5%
6	多い世代が存在しない	3634	8.1%

表 9: グループとコメント数

グループ名	特徴	平均コメント数
0	中高生のみが多い	11.6 %
1	大学生のみが多い	13.3%
2	大人のみが多い	11.6 %
3	中高生と大学生が多い	14.9 %
4	大学生と大人が多い	20.1 %
5	大人と中高生が多い	21.1 %
6	多い世代が存在しない	4.9 %

結果として、平均コメント数はグループ 4、グループ 5 が多いことが分かった。中高生、大学生それぞれについて、その世代のみが多い部屋よりも、大人とその世代で交流している方が、片側 2.5% 検定で統計的に有意にコメント数が多い。即ち 1 部屋あたりで考えた時に、他者から話しかけられる数が高いことが発見できた。即ち、新たに交流を誘われる確率が高いことを示した。

## 5. 結論

本研究では、明示的な年齢情報がない 755 においてユーザーの年齢層を高い精度で判定することができた。さらに中高生を始めとした未成年者にとって、大人との交流は実際に危険な交流に巻き込まれる可能性が高いことを示した。実際に性犯罪に巻き込まれるのは未成年者が多いので、多くの SNS では年齢情報が分からないが、そのような中でもきちんと未成年者を検出することは、未成年者保護のためには重要である。その点で本研究は重要な意義があった。さらに、様々な世代との交流が可能なオンラインにおいて、誘い出しのリスクが高いことを示した。これらの成果は、オンラインチャットサービスにおいて、規制が必要であることを示唆している。

今後の研究の発展の方向性として、大人と中高生が多い部屋の特徴を詳しく調査し、どのような会話が行われているか等、交流の実態をさらに詳細に明らかにすること、未成年者がより安全により便利にオンラインサービスを使えるような規制を提案すること等が考えられる。

## 参考文献

- [1] B,Belsey. *Cyberbullying, An Emerging Threat to the 'Always On' Generation*. Recuperado el Vol 5, 2005
- [2] 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.
- [3] Fujio TORIUMI, Takafumi NAKANISHI,Mitsuteru TASHIRO,Kiyotake EGUCHI. *Encounters between Predators and Their Targets in Private Chat*.Journal of Transformation of Human Behavior under the Influence of Infosocionomics Society Vol1.2016.
- [4] Zhang,J;Hu,X;Zhang,Y;Liu,H. *Your Age Is No Secret: Inferring Microbloggers' Ages via Content and Interaction Analysis*.ICWSM'2016.
- [5] Genkin,A;Lewis,D,D;Madigan,D. *Sparse logistic regression for text categorization*.DIAMCS Working

---

Group on Monitoring Message Streams,Project Report.2005.

- [6] 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 ofACL'11,Portland.