SNS 動画像投稿記事のトピック抽出とそのカテゴリー化によるプライバシー侵害理由の推定方式

尾崎敏司，輪島幸治，隅岡隆之，村上陽子，
邓 超頴，嶋田茂

産業技術大学院大学 〒140-0011 東京都品川区東大井 1-10-40

{ a1305so, a1328kw, a1311ts, a1325ym, a1333ct,shimada-shigeru}@aiit.ac.jp,

概要 近年，Google Glass 等のウェアラブルデバイスの出現と共に，あらゆるところで動画像の撮影可能な環境が整いつつある．この状況から，被撮影者の動画像が無断で SNS へ投稿されるリスクが高まっており，意図しないプライバシー侵害を検知することが求められている．そこで本研究では，動画投稿時にプライバシー侵害判定を行い，侵害理由を提示するサービスのために， Google Glass に関する YouTube 記事のコメントとキャプションを対象として，LDA によるトピック抽出とそのトピックの結合やカテゴリー化による，プライバシー侵害理由の推定方式を検討する．

キーワード LDA トピック抽出 理由推定 プライバシー侵害

1 導入

1.1 背景

本近年，Twitter や Facebook 等の SNS では写真や動画による投稿が増加している[1]. これはデスクトップ PC 等の静的環境から，スマートフォン等のモバイル環境への移行により，写真や動画による投稿がより容易な環境が形成されているためと考えられる．この傾向は今後も継続される上，最近では，各種センサに通信機能を装備して着用可能なウェアラブルデバイスが市場に出始めており，FuelBand に代表されるような着用者の動作や脈拍等の身体状況の計測データをクラウド上に送信するものの外に[2], Google Glass に代表されるようなカメラとヘッドアップディスプレーを備え，撮影した写真や動画をクラウド上へ送信する機能を備えるものもある[3].特にこの Google Glass に装備されているカメラ(以下ウェアラブルカメラ)は，着用したまま常時撮影を行うことが可能で，撮影されていることを周囲の人間が気付き難くなっている．そのため，ウェアラブルカメラを着用したユーザーが，撮影した写真やビデオを SNS へ不用意に投稿することにより，そこに映り込まれた他者のプライバシーを意図せず侵害してしまう可能性があるなど，プライバシー侵害の面での問題点が指摘されている． このため，米国におけるレストラン等の一部の施設において，GoogleGlass の利用が禁止されるなど，新しいウェラブルディバイスの普及が阻止されるといった社会問題が発生している[4].

1.2 関連研究

このような状況を打開するため，SNS への写真やビデオの投稿から発生するプライバシー侵害を事前に検知して，個人情報の漏洩を保護するようなオプトイン形式のプライバシー保護方式が必要となる．既に，Anna Squicciarini らは，SNS へ投稿された写真の画像分析からその写真特性を抽出し，クラスター化して，写真の公開範囲をそのクラス毎に変える方式を提案している[5]. 一方，大本らは，GoogleGlass に関する YouTube 記事を対象に，"privacy" に関連するビデオ記事の抽出から，プライバシー侵害シーンを絞り込み，そのシーンを形成する画像やコメント及びキャプションから教師データを生成し，機械学習によりプライバシー侵害を自動判定する方式を開発している[6]. これらの研究はいずれもプライバシー侵害の発生を未然に防ぐ意味でのオプトイン方式となっているが，そのプライバシー検知からユーザーへの警告として提示されるメッセージは，単純にプライバシー侵害の有無を示す程度の指摘であるため，その指摘が理解できないユーザーにとっては，煩わしい警告に捉えられ，無視されるような事態に陥ることが危惧される．

1.3 研究目的

本研究は，この問題に対応するものであり，プライバシー侵害が検知された場合に，単にその侵害有無の警告だけではなく，そのプライバシー侵害理由を提示することにより，ユーザーがプライバシー侵害を回避する行動

をとり易くすることを狙うものである. 前年度ではプライバシー関連語(Privacy Sensitive Words :PSW)[7] によりプライバシー関連記事を抽出した後, 人手によりプライバシー侵害記事を抽出して, その限定したサンプルの画像・キャプション・コメント・位置・時間等の特徴ベクトル化したものを教師データとして, SVM によりプライバシー侵害の有無を検知している. しかし, この方式では. 単にプライバシー侵害の有無だけが判定され, 侵害の原因となる話題(トピック)が把握されていない. そこで, この侵害検知に加えて, 侵害となる記事に含まれるトピックを抽出して, それらのトピックの関連から, 侵害理由を推定する方式を提案する.

2 SNS 投稿に起因するプライバシー侵害とその分析対象

2.1 関連研究

既に町田らによる前研究[8]で触れているように, モバイルディバイス装備のウェアラブルカメラにより撮影された写真やビデオを YouTube 等の SNS へ投稿して公開する場合には, 各種のプライバシー侵害を引き起こすことが考えられる. その場合の類型として, ①他者侵害型, ②自己漏洩型, ③侵害反論型, ④侵害指摘型, ⑤間接侵害型 の5つの代表的なものがあることが纏められている[9]. この中で, GoogleGlass 等のウェアラブルカメラの場合には, 自分自身を映す(いわゆる自撮)構成にはなっていないので, ②の類型は当てはまらない. 一方この類型を, ウェアラブルカメラを使用するユーザーの観点で分類すると, ①はウェアラブルカメラを装備したユーザー自身の視点(Through Glass とも言う)からみた主観的なプライバシー侵害の捉え方にあたり, ③④⑤はウェアラブルカメラを使用しているユーザー第三者的に客観的なプライバシー侵害の捉え方にあたる.

2.2 分析対象とする SNS 記事

本研究で, これらの類型に属するプライバシー侵害を分析する SNS としては, 前年度の研究に引き続き, YouTube のアーカイブ(前年の 7 ヶ月分(2013/06 - 2013/12)に蓄積された YouTube 記事)とする. この YouTube のアーカイブに記録されている記事は, 次の4つのメディアで構成される(図 1 参照)

(1) ユーザー間での共有対象となるビデオ
(2) ビデオの音声が約 4 秒ごとにテキスト化されたキャプション
(3) 投稿ユーザーや閲覧者がビデオに対する所感を入力するコメント
(4) 動画タイトルや公開範囲などのメタデータ

今回の研究では4つのメディアのうちコメントとキャプションを対象に分析を行う. コメントは動画を閲覧した閲覧者(第三者)からなされるものであり, 動画の閲覧に際し閲覧者が想起した内容が含まれている. ただし, 動画とまったく関係のない話題が話される可能性があるため, 動画との結びつきが弱くなることがある. また, 動画内の特定のシーンにコメントを結びつけるのが難しい.

一方, キャプションは動画内の音声であるため, 動画の内容との結びつきが強く, また, 動画を分割した場面(シーン)に対応したキャプションというものを考えることができる. 特にウエラブルカメラにおいては, 撮影者の音声が主になり, 撮影者の感情が反映される(表 1 参照).

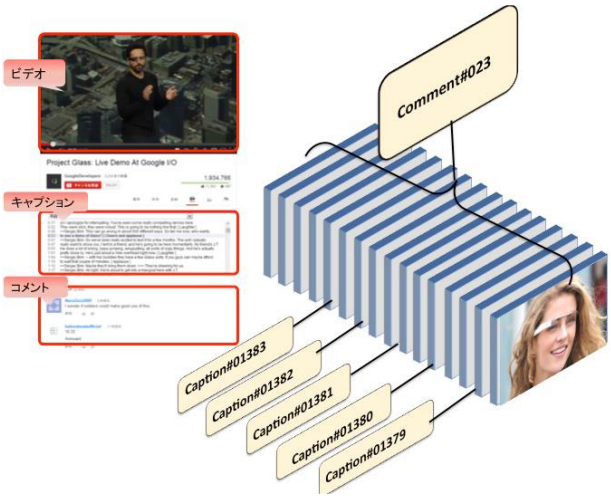

図 1 YouTube 記事の構造

表 1 コメントとキャプションの動画との関連性の違い

	内容	動画との結びつき	動画内の各シーンとの対応付け
コメント	所感	弱い	難しい
キャプション	動画の音声	強い	可能

2.3 分析対象の選定

コメントとキャプションの動画との結びつきの違いを考え, コメントとキャプションに対してそれぞれ異なるデータの選定方法を用いている

コメント: 動画と関係なくプライバシー侵害に関する議論が行われている可能性があるため, 侵害しているかしていないかに関わりなく, プライバシーに関連の高いコメントが含まれている可能性が高いデータの抽出を行う. そのため, Google Glass に関する YouTube 記事から, より多種の PSW が含まれているものを抽出し分析の対象とした. コメント内で行われている動画への指摘や議論の中にプライバシー侵害に関連している情報が含まれていることを期待する.

キャプション: シーンとの対応付けが可能であることを利用して, 前年度までの研究において実際にプライバシー侵害を犯している可能性が高いとされるシーン(侵害シーン)の含まれている動画データを対象とする. 侵害シーンの音声情報にそこで行われているプライバシー

侵害に関連している情報が含まれていることを期待する．なお，客観的な評価をコメントから，主観的な認識からの評価をキャプションから，行うことを想定している．

3 プライバシー理由の推定方式

3.1 コメントによるプライバシー侵害理由の推定フロー

2 章で述べたとおり，前年度までの方式では侵害の原因となる話題(トピック)を把握することができないので，プライバシー侵害理由を推定するための別の手法が必要である．本研究では，プライバシーに関連性の高い記事やプライバシー侵害となる記事に含まれるトピックを抽出して，それらのトピックの関連から，侵害理由を推定する．

PSW で抽出した記事を対象にトピック抽出による解析を行う．抽出された記事数を M とすると，全コメントの集合 C は，

$$C = \{c_j \mid 1 \leq j \leq M\} \ldots(式\ 1)$$

と定められる．ここで，C は抽出した YouTube 記事のコメントすべてを表す文書集合であり，c_j は個々の YouTube 記事それぞれのコメントを表す文書集合である．

文書集合 C に対して，トピック数 N でトピック抽出を行い，これを上位トピック$U = \{u_i \mid 1 \leq i \leq N\}$…(式 2) と定める．文書集合 W に，トピック数 N でトピックを抽出する処理を $f_T(N, W)$ とすると，上位トピック U は

$$U = \bigcup_{1 \leq i \leq N} u_i = f_T(N, C) \quad \ldots(式\ 3)$$

と記述できる．

また，同様に，文書集合 c_j に対して，トピック数 n でトピック抽出を行い，個々の記事のトピックである d_j を求める．この d_j の和集合を下位トピック$D = \{d_{jk} \mid 1 \leq j \leq M, 1 \leq k \leq n\}$…(式 4) として定める．上位トピックと同様に，$f_T(N, W)$ を用いて記述すると

$$d_j = \bigcup_{1 \leq k \leq n} d_{jk} = f_T(n, c_j) \quad \ldots(式\ 5)$$

$$D = \bigcup_{1 \leq j \leq M} d_j \quad \ldots(式\ 6)$$

となる．

次に，この上位トピック $u_i \in U$ と下位トピック $d_{jk} \in D$ の組に対して dice 係数を計算する．求めた dice 係数が ω 以上になるものを上位トピック u_i と下位トピック d_{jk} が結合したと定める．

dice 係数を求める処理を$f_D(u_i, d_{jk})$…(式 7) と記述すると，上位トピック u_i に dice 係数が ω 以上で接続された下位トピックの集合 $D(u_i)$ は下記のように定めることができる．

$$D(u_i) = \{d_{jk} \in D \mid f_D(u_i, d_{jk}) \geq \omega\} \ldots(式\ 8)$$

つまり，u_i との dice 係数が ω 以上となる d_{jk} の集合を $D(u_i)$とする．本研究では，この u_i がプライバシーに関連していると判断できる場合に，$D(u_i)$がどの程度プライバシーに関連しているか人手で評価を行い．さらに，この $D(u_i)$にプライバシー侵害の理由に関連する可能性のある単語が存在しているか確認する．

まず，YouTube アーカイブからコメントを取得した．その後，そのコメントに対して PSW が 10 種以上含まれるように動画の抽出を行った(図 2 step1 参照)．これにより 39 本の記事が抽出された．つまり，記事数 M=39 となる．これらの各動画のコメント c_j (1≦ j ≦39)に対してトピック数 n = 10 としてトピック抽出を行い下位トピック d_j を求めた．同時にすべての動画のコメントをあわせた文書集合 C に対してもトピック数 N = 20 としてトピック抽出を行い上位トピック U を求めた(図 2 step2 参照)．次に，上位トピック u_i と下位トピック d_{jk} の間で dice 係数を計算し，これが $\omega \geqq 0.4$ となった際に，その上位トピックに下位トピックが結合されたとした(図 2: step3 参照)．この条件を満たす d_{jk} の集合が $D(u_j)$ にあたる．最後に，プライバシーとの関連の判定を U に対して人手により行った(図 2: step4 参照)．次節より各ステップの詳細を述べる．

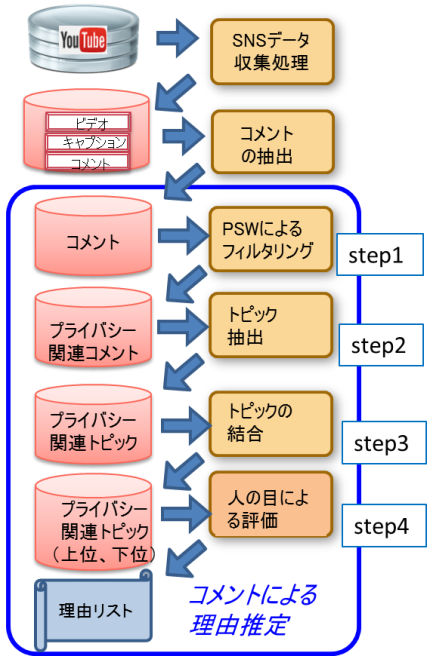

図 2 コメントによる理由推定フロー

3.1.1PSW によるフィルタリング

まず，プライバシー関連のコメントが投稿されている動画を抽出するため，プライバシーに関連性の高い単語の辞書である PSW を用いて，動画の抽出を行った．

今回の研究においてはこの PSW の作成に，英語版 Wikipedia のアブストラクトを文書群とした．これを WordNet により抽象化したのち”privacy”と bigram で共起している単語の出現頻度を計算し，上位 20 単語に ”privacy” を加えた 21 単語を PSW として用いている（表 2 参照）．

今回は，コメントにこの PSW が少なくとも 10 種類含まれるという条件で動画の抽出を行った．その結果 39 本の動画が抽出された．

表 2 PSW リスト

Privacy	Differential	Equivalent
Preserving	Internet	Information
Protection	Concern	Online
Surveillance	Freedom	Electronic
Invasion	Consumer	Data
Protect	Security	User
Commissioner	Advocate	Financial

3.1.2コメントのトピック抽出

抽出された 39 本の動画のコメントに対して Latent Dirichlet Allocation (LDA) (Blei et al. 2003)[10] を使用して以下のようにトピックの抽出をおこなった．

上位トピック U :抽出された M = 39 本すべてのコメントをあわせた文書集合 C に対してトピック数 N = 20 となるように，トピック抽出を行った．つまり，$U = f_T(20,C)$ を求めた．

下位トピック D :抽出された M = 39 本 ，それぞれの動画のコメント c_j に対して n = 10 個のトピックを抽出させそれぞれ $d_j = f_T(10, c_j)$ を求めた．

つまり，下位トピックは，$D = \cup d_j\ (1 \leqq j \leqq 39)$となる．

3.1.3各動画のトピックの結合

先のトピック抽出により，390 個のトピックをもった下位トピックの集合 D が生成された．この 390 個のトピックを，20 個のトピックをもつ上位トピック U と結びつける．これには井上らの提案しているトピックのマージ手法を参考にし，dice 係数により上位トピック u_i と下位トピック d_{jk} の結合を行った[11]．この際，本実験では上位トピック u_i と下位トピック d_{jk} との dice 係数が 0.4 以上になる下位トピックをその上位トピックに結合されるとした．

3.1.4評価

まず，20 の上位トピックについて内容を確認し，タグ付けを行い，プライバシーに関連のあるものと関連のないもの，どちらともとれるものの，3 種に分類した．

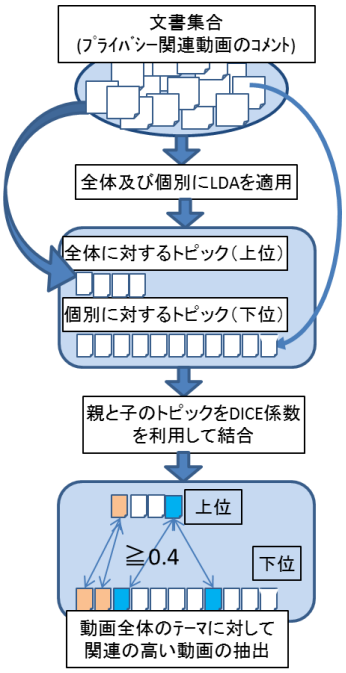

図 3 トピックの結合の概念図

また，20 の上位トピックのうち，下位トピックと結合されたのは表 3 の 6 つであった．このうちタグ:「公共でのプライバシー」，「撮影への畏怖」の 2 トピックをプライバシーに関連しており，これらをプライバシーと関連していないトピックである，タグ:「運転と警察」「動画一般」の 2 つを比較することで，プライバシーに関連のある上位トピックを用いて，プライバシー関連の下位トピックを得ることが可能かを評価する．

そのために，各上位トピックに結合された下位トピックがプライバシーに関連するかどうかをそれぞれ確認し，割合を計算した(表 3 参照)．例えば，タグ:「撮影への畏怖」に結合された下位トピックは 13 個存在するが，そのうち，プライバシーに関連していると人間が判断可能なトピックが 6 個，関連していないと判断可能なトピックが 3 個，どちらにもとりうるトピックが 4 個存在していた．この場合，プライバシーに関連しているトピックの含有率は 46.2%となる．表 3 をみると「公共でのプライバシー」が 33%，「撮影への畏怖」が 46% の割合でプライバシー

表 3 下位トピックと結合の行われた上位トピック

タグ	Topic 数	単語	内容	プライバシーとの関連
公共でのプライバシ	6	glasses already people privacy wear public security cameras cops would	グーグルグラスのプライバシーやセキュリティの話題	あり
撮影行為への畏怖	13	google glass glasses take awesome video record say picture scary	グーグルグラスによる写真撮影への畏怖の話題	あり
運転と警察	8	police good thing buy would use one drive probably think	運転中の行動と警察の話題	なし
動画一般	34	movie like see people would without face something think could	動画一般の話題	なし
人間と法律	1	going people let law head like feel hud think laws	人間と法律の話題	なし
不明	3	like look love bad people fucking shit make life looks	不明？	どちらにも

関連のトピックが含まれているのに対して，下位トピックはどちらも 25%以下であった．このことより，上位トピックにおいてプライバシーに関連していると下位トピックもプライバシーに関連がある可能性が高いことがわかる．

表 4 プライバシー関連トピックの含有率(タグ比較)

		下位トピック		
		関連しない	どちらともとれる	関連する
上位トピック	公共でのプライバシ	50%	16.6%	33.3%
	撮影行為への畏怖	23.1%	30.8%	46.2%
	運転と警察	62.5%	12.5%	25%
	動画一般	56%	20.6%	23.5%

また,プライバシーに関連する上位トピックと関連しない下位トピックをそれぞれ合算して比較した場合を示す(表 5 参照).

表 5 プライバシー関連トピックの含有率(合算)

		下位トピック		
		関連しない	どちらともとれる	関連
上位トピック	プライバシー関連	23.8%	19.0%	57.1%
	関連しない	44.4%	27.8%	27.8%

表 5 より，上位トピックがプライバシーに関連している場合は関連しないものと比べて，約 2 倍程度の割合でプライバシーに関連の下位トピックが含まれることがわかる．

また，プライバシーに関係している上位トピックに結合される下位トピックの単語リストが表 6 である．これをみると，出現回数が多い単語は google, glass, people, would, like, video などの抽出条件に依存した単語や一般的な単語となっており，理由の作成には利用できない．一方で，出現回数の少ない単語においてプライバシー侵害理由に関連しそうな単語がいくつか見られた．例えば，身体部位を表す face や 親族や弱者に関連のあると思われる kid, girl, 犯罪に関連のあると思われる police, robber, cops, 人名である john, 金銭に関係のある pay などがあげられる．これらの単語はプライバシー侵害の理由に相当している可能性があると考えられる．

3.2 キャプションによるプライバシー侵害理由の推定フロー

上述のコメントのトピック抽出により，理由に該当しうる単語の検討を行うことができた．しかし，これは第三者の議論によるもので，撮影者の主観的な情報は含まれていない．そこで撮影者の発言を記録しているキャプションからも理由推定に利用できる情報が抽出できないか検討した．まず，前年度の研究においてプライバシー侵害だと判定されたシーンを含む動画を抽出し(図 4 step1 参照)，次に，その侵害シーンに対応するキャプションについてトピック抽出を行った(図 4 step2 参照) ．

3.2.1各動画の侵害シーンの抽出

前年度の大本らの研究[6]により，プライバシー侵害シーンが含まれていると判定された 22 本の動画の内，キャプションの付いている 12 本の動画を選定した．

その動画の内容を確認してそれぞれ複数のシーンに分割した．プライバシー侵害だと判定されたシーン数は 51 シーンであった(図 4 step1). この 51 の侵害シーンから，トピック数 3 でトピック抽出を行った(図 4 step2 参照).

表 6 プライバシー関連上位トピックに結合された下位トピックに現れる単語リストと出現回数

単語	数	単語	数	単語	数
google	19	watch	2	kids	1
glass	16	wearing	2	life	1
glasses	15	300	1	make	1
people	15	9000	1	makes	1
video	12	apple	1	mark	1
would	12	around	1	myopia	1
like	11	assault	1	need	1
get	5	attention	1	pay	1
one	4	bar	1	police	1
take	4	better	1	power	1
think	4	buy	1	probably	1
wear	4	camera	1	real	1
lol	3	cool	1	recording	1
look	3	cops	1	right	1
already	2	could	1	robber	1
funny	2	driving	1	say	1
picture	2	even	1	shit	1
privacy	2	face	1	stupid	1
really	2	girl	1	talk	1
record	2	going	1	thank	1
time	2	imagine	1	vegeta	1
use	2	john	1	want	1
				works	1

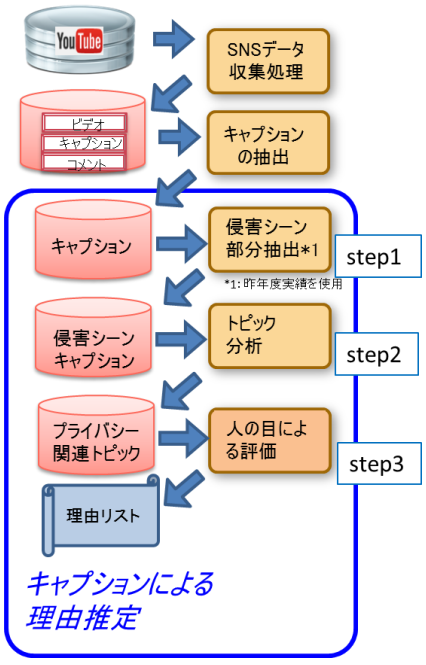

図 4 キャプションによる理由推定フロー

そのトピックを人手によりプライバシーに関連しているか評価した(図 4　step3 参照).

3.2.2評価

実際に抽出されたトピックの単語を確認すると下記のような問題があることがわかった.

1. 単一シーンでは十分な文書量が確保できない

たとえば, 動画 ID 5R1snVxGNVs の侵害シーンの 1 つは約 30 秒程度の長さのあるシーンであるが, 下記のトピックが作成された.

Topic1 (gay tonight money line buy really sweet hey tickets man)
Topic2 (tickets man hey really buy line sweet money gay tonight)
Topic3 (tonight gay sweet money line buy really hey man tickets)

これらのトピック中の単語は, 順序が異なるが同一のものである.　そのシーン中に含まれているキャプションを確認すると下記のみであった

I'm oh man really hey there and line sweet money to buy tickets for was your gay no tonight

つまり, このシーン分割によるアプローチでは, トピック抽出に必要な十分な文書量が確保できていない場合があることが確認できた.

2. 感嘆表現, 繰り返し, 簡潔な表現などが頻出する口語である.

下記のように ah,oh 等の感嘆表現や, 繰り返しなども キャプションは記録するため, 抽出の妨げになることがあることがわかった.

glasses open face to with what I'm no no no open just face QR shit %ah movie has I know that okay

以上より, 特に文書量の問題によりキャプションから侵害理由を求めることは難しいことがわかった.

4 結論

YouTube のコメントの PSW による抽出とトピックの結合により, コメントの文書集合からプライバシーに関連づいているトピックを取得することができた. これはコメント内で議論が発生し, プライバシー侵害に対する客観的な指摘が含まれている可能性があるためと考えられる. また, 抽出されたトピックを結合し評価を行うことで, プライバシーに関連のある下位トピックを効率よく収集できる可能性があることがわかった. さらに, その下位トピックには, プライバシー侵害の理由となる可能性のある単語も含まれていることがわかった.

一方, キャプションへのトピック抽出は, プライバシー侵害の理由リストの作成には向かないということがわかった. これは 1 シーンに話される単語数の問題が主であるが, 口語表現がトピック抽出の妨げになっている.

以上より, YouTube のコメントに対して PSW による動画抽出とトピックの結合を行うことがプライバシー侵害の理由リストを作成する手段の一つとなりうることがわかった.

5 展望と課題

Wikipedia のアブストラクトが各項目を短く説明した文書であるため今回の PSW は抽象的な単語を多く含んでいた. 文章集合を変更するなどし, より具体的なプライバシー関連辞書を作成できれば, 理由リストの作成のためのデータを効果的に収集することができると考えられる. また, 今回人手により評価した部分に関して客観的な指標を作成し, このステップの自動化を行いたい.

参考文献

[1] BUSINESS INSIDER , ” Facebook Users Are Uploading 350 Million New Photos Each Day ”, http://www.businessinsider.com/facebook-350-million-photos-each-day-2013-9 (2013/12/29)

[2] NIKE+ FUELBAND http://www.nike.com/jp/ja_jp/c/nikeplus-fuelband

[3] Google Glass Explorer Program https://www.google.com/glass/start/

[4] CNET, “Privacy officials from 6 countries reques etails on Google Glass”, http://news.cnet.com/8301-1023_3-57589973-93/privacy(2013/12/30)

[5] Anna Squicciarini, Smitha Sundareswaran, Dan Lin, ” A3P:Adaptive Policy Prediction for Shared Images over Popular Content Sharing Sites”, ACM New York, 2011

[6] 大本 茂史, 岸本 拓也, 髙田 美樹ほか "ウェアラブルカメラを利用した SNS 記事投稿によるプライバシー侵害を保護する方式の提案", DEIM Forum 2014

[7] 高田さとみ, 小山貴之, 町田史門ほか "SNS 画像投稿時に発生するプライバシー侵害の要因分析", 電子情報通信学会 EMM 研究会技術報告, 2012

[8] 町田史門, 小山貴之, 宋洋, 高田さとみほか "SNS 写真投稿に起因するプライバシー侵害の類型化とその保護策", 電子情報通信学会 EMM 研究会技術報告, 2012

[9] 高田さとみ, 周子胤, 髙田美樹ほか SNS 画像投稿時のプライバシー侵害予知サービスの提案, DEIM Forum 2013 F8-4

[10] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. Journal of Machine Learning Research, Vol. 3,pp. 993–1022, 2003

[11] 井上祐輔, 小池大地, 宇津呂武仁ほか "複数の粒度での LDA 適用結果におけるトピック集約", 言語処理学会, 第 20 回年次大会, 2014