Vol.2014-NL-215 No.6
2014/2/6

# キャプションテキスト感情の分析による プライバシー侵害シーン抽出

高田さとみ[†1] 周　子胤[†1] 高田 美樹[†1] 大本 茂史[†1] 岸本 拓也[†1] 奈良 育英[†1]
嶋田 茂[†1]

ウェアラブルコンピュータ装備のカメラによるプライバシー侵害を危惧する意見が高まりつつあり，その利便性が否定されかねない．そこで,我々は Google glass に関する Youtube 記事から機微な意見の多いプライバシー侵害動画記事を収集した．さらに，テキストに変換した動画の音声を利用し，動画記事に含まれる投稿者や撮影者の感情分析を行った．これにより，投稿者や撮影者の感情からプライバシー侵害のシーンを抽出している．

# Extraction accuracy of Privacy Invation Scene based on SNS caption emotion analysis

Satomi TAKADA[†1] Ziyin ZHOU[†1] Miki TAKATA[†1] Shigefumi OOMOTO[†1]
Takuya KISHIMOTO[†1] Ikuhide NARA[†1] Shigeru SHIMADA[†1]

Opinion to fear the invasion of privacy by the camera of the wearable computer equipped with is growing, its usability could be denied. Therefore, we have collected a privacy infringement Videos Recommended articles opinion sensitive from Youtube articles on Google glass. The more we , using the voice of the video has been converted to text, was emotion analysis of the photographer or reviewer included in the moving article. Accordingly, it is extracted scenes privacy from the emotion of the photographer or reviewer.

## 1. はじめに

### 1.1 研究背景

Twitter や Facebook をはじめとする SNS の利用傾向として，テキスト形式による投稿から，画像や動画等のマルチメディア形式の投稿が増加しており，今後もその傾向が強まることが予測される．Facebook で画像が含まれる投稿数は，2012 年には 1 日 3 億件に達し[1]，今後この割合は急速に増加する傾向にある．従来のスマートフォン搭載のカメラを用いて投稿するより，Google Glass に代表されるようなウェアラブルディバイスに装備されたカメラ（これを以降ウェアラブルカメラと略称する）を用いて投稿する方が簡便であると考えられる．このウェアラブルカメラは，簡易的にいつでもビデオ撮影を可能とする．この常時撮影特性を持ったウェアラブルカメラは，被撮影者に気がつかれず撮影することが可能である．ウェアラブルカメラで撮影され,投稿された動画記事がプライバシー侵害を引き起こすことが多くなる．即ち，このようなウェアラブルカメラによる投稿の増加は，プライバシー侵害リスクも連動して高まり，非常に大きな社会問題となることが危惧される．

また，SNS 上でプライバシー侵害が発生する投稿記事の多くが,ユーザーの不注意で投稿された後に発生している．しかしながら，投稿前にプライバシー侵害を防止する方式(Opt-in 方式)を適用することは現状では難しい．従って,従来 SNS を提供する各社は事後の対応策(Opt-out 方式)として，対応方針を示すプライバシーポリシーや投稿記事の公開範囲の推奨等を記述提示する対応を行っている.ところが，Opt-out 方式の記述提示では,全てのケースを網羅するための複雑で長文の記載が行われることが多く,全てを閲覧して理解するユーザーは少ない．対応策は各ユーザーの裁量にゆだねられる状況である．しかし,ウェアラブルカメラになると,スマートフォンと違って被撮影者にカメラを向けられている状況に気づかれないため,より深刻なプライバシー侵害記事が投稿される可能性が高い．ユーザーの裁量まかせとなる Opt-out 方式では手遅れになることもある．

実際に，ウェアラブルカメラ普及前の現在においても，投稿者の感情が高ぶるいわば興奮状態のときに判断が緩くなり不適切な内容となることもある．プライバシー侵害動画記事を見たことで閲覧者によって指摘を受ける事例も報告されている[2]．高級ホテルの飲食店でアルバイトをしていた大学生が,店に食事に来たサッカー選手と芸能人のカップルの様子に興奮し Twitter に投稿したことで芸能人のプライバシー漏洩となり投稿者への非難が殺到し炎上した事例がある[3]．

### 1.2 研究目的

本稿では，SNS へのマルチメディア投稿記事のプライバシー侵害の可能性が考えられるシーン検知に向けて，Youtube 動画記事の音声をテキストに変換した字幕（これを以降，キャプションとする.）を用いて投稿者や撮影者の

†1 首都大学東京 産業技術大学院大学
Tokyo Metropolitan University Advanced Institute of Industrial Technology.

ⓒ 2014 Information Processing Society of Japan

感情分析を行った．動画記事に対するコメントを解析し，プライバシー侵害を述べたコメントを抽出し,それらへの感情分析を行ったところ，閲覧者が感情的に記事に対してプライバシー侵害であることを示唆したプライバシー侵害動画記事の抽出には成功した．しかし，具体的なプライバシー侵害シーンの抽出までには至っていない．今回，キャプションを解析し投稿者や撮影者の感情を分析することで抽出されたシーンがプライバシー侵害シーンであるシーン検知率を算出した．これにより，投稿者や撮影者の感情とプライバシー侵害の親和性の高さを明らかし，感情を分析することでプライバシー侵害シーンの抽出をしていくこととする．

以降，2 章で SNS のプライバシーと感情について 3 章では SNS の構成について述べ，4 章でシーン抽出手法の提案を行い，5 章で評価を行う．

## 2. SNS におけるプライバシーと感情

### 2.1 プライバシー侵害の定義

一般に，プライバシー侵害の判定となるプライバシーの定義について既存の文献を調査すると，各国の法律として「個人情報保護法」がある[5] [6]．その内容は，個人情報を管理する立場からのルールが主体であり，個人情報を公開される側の個人の感情や価値観を反映するものではなく，プライバシー侵害判定のための定義とするには不十分である．一方，青柳は，プライバシー侵害を，「”個人識別情報”と”不可侵私的領域”が，本人の意思に反して結びつくことである」と定義している[7]．ここでは，”個人識別情報”とは，{氏名・性別・生年月日・住所等の個人を識別できる情報}，”不可侵私的領域”とは，{私生活上の領域，一般人の感受性を基準として公開を欲しないと思われる領域，非公知の領域，公開によって当該私人が現実に不快や不安の念を覚える領域}としている．しかし，どこまでが「私的生活」なのか，どこを「一般人の感受性」とするのか，異なる価値観を持つ多くのユーザーが集まる SNS においては，なお曖昧性が残る．

### 2.2 感情の定義

プライバシー侵害記事収集,抽出の研究を進めていくなかで SNS における感情モデルは閲覧者感情と当事者感情の 2 種に分類できることが判明した．投稿者や撮影者が動画撮影あるいは投稿したときに発する感情と投稿した記事を閲覧することで発する感情は必ずしも合致しない．当事者感情と閲覧者感情の定義を行うにあたり，SNS ユーザー分類を，小山ら[8]が定義した“投稿者 T:Tweeter”，S: Subscriber, ”指摘者 N: Notifier”の 3 種のうち 2 種を利用する．T:Tweeter は記事を投稿するユーザー，S: Subscriber はその記事を閲覧するユーザーである．尚，Youtube をはじめとする動画記事を扱う SNS では，T:Tweeter には撮影者と非撮影者も含める．

当事者感情は,T:Tweeter から発せられる感情として定義する．投稿者や撮影者が喜びや怒りで興奮している状態，非撮影者が怒りで興奮している状態が考えられる．

閲覧者感情は，S: Subscriber が記事を閲覧することで投稿記事に対して発せられる感情として定義する．T:Tweeter の記事が不適切である時に，怒りや悲しみなどの感情が現れている状態が考えられる．

### 2.3 関連研究

プライバシーと感情に関連する研究として，Yang Wang らによる Facebook における記事投稿後に起こる後悔の要因分析がある[9]．調査によれば，強い感情が現れている状態で Facebook に投稿するとプライバシーセンシティブ情報を漏洩させる可能性があり，それが投稿後に後悔する原因の一つとして挙げられている．該当論文はインタビュー形式による調査のみで，感情強度をトリガにしたプライバシーセンシティブ情報を漏洩させたことについて自動抽出する方式までには至っていない．

次に，Huina Mao らによる Twitter におけるプライバシー漏洩記事の抽出と原因分析がある[10]．休暇・アルコール過剰摂取・病気と，強い感情が発生しているシーンを位置情報やフレーズを根拠に 3 カテゴリに分類している．さらに，SVM（サポートベクトルマシン）や Naive Bayes(ナイーブベイズ)を利用し自動分類することにより記事を増やしている．しかし，なぜこれらのシーンに強い感情が出ているかの根拠が不十分である．

奈良ら[3]による SNS の感情分析をトリガーにしたウェアラブルカメラによるプライバシー侵害の要因分析がある．Google glass に関する Youtube 記事よりプライバシーを指摘する意見を抽出し，感情分析を行うことでプライバシー侵害要因を特定し，プライバシー侵害記事を収集する方式が提案されている．プライバシー侵害記事を閲覧し,記事に対する意見から閲覧者感情を利用した感情分析となっている．これにより，プライバシー侵害記事の収集は可能となったが，具体的なプライバシー侵害シーンの抽出までには至っていない．本稿は奈良ら[3]による研究の継続となる．

## 3. SNS 記事構成と収集方法

### 3.1 SNS 記事構成と定義

本稿で分析の対象とする SNS 記事は，製品発売前にもかかわらず注目を浴びているウェアラブルカメラである Google Glass に関するものである．また，分析対象とする SNS の選考にあたり，Twitter の収集方式はサンプリング形式であり，ある話題に関する一連のコメント全てを取集することが困難であるため対象としなかった．Facebook も，利用規約によりデータ収集に制限がある[11]．そこで，我々

は自動収集が可能であり，投稿ビデオに関するすべてのコメントを取得でき，関連する話題の一連の流れを分析対象にすることができる YouTube を選択した．

まず，YouTube への投稿ビデオから音声をテキスト化したキャプションに注目し，そのビデオの特性を明確にする．以上の情報は，下記の３つで構成され，これらは互いに関連付けて管理される（図１）．

1. ユーザー間での共有対象となるビデオ部
2. 音声認識にてテキストに出力されるキャプション部
3. 投稿ユーザーや閲覧者が動画に対する所感を入力するコメント部

図 1 YouTube のデータ構成

### 3.2 SNS 記事収集方法

分析対象とするデータは，過去 4 ヶ月（2013/06 - 2013/9）に蓄積された YouTube データアーカイブである．コンテンツを自動収集するため Google が提供している『YouTube Data API』を利用する[12]．YouTube にて検索キーワード『google+glass』でヒットするビデオは約 13,800,000 件となるが，同 API は，1 度で取得できるデータ件数は上位 1000 件までと制限があるため，一定間隔のポーリング収集を行い，クローリングを行った．また，上位 1000 件のソート順にも制約があり，指定可能な値は，relevance（関連度が高い順），published（公開日順），viewCount（再生回数順），rating（評価が高い順）の 4 種類であり，そのうち，published を選択することより常にその時点で最新の投稿内容を収集した．収集対象とするデータは，投稿されたビデオ属性データ（コメント数・カテゴリ・投稿者・概要等），ビデオに対するコメントデータ，YouTube で自動字幕起こしされたキャプションの３つとなる．また，収集対象とするビデオは，製品レビュー・ニュース・製品デモ等の Google Glass に関する話題の英語のビデオを対象にしているが，Google Glass で撮影したビデオそのもの（これを以降,スルーグラスとする）のみを実際の分析対象とする．

## 4. 当事者感情を利用した興奮状態シーン抽出の提案方式

### 4.1 抽出方式の概要

本手法は，テキスト解析による感情分析を行う．解析には 前章で提示したコメントデータ部とのキャプションを利用する．コメントデータ部は，奈良ら[3]によって提案された手法に基づいて解析しプライバシー侵害動画記事収集を行っている．動画記事からさらに具体的なシーンの抽出を行うにあたり，のキャプションを解析し当事者感情を分析した．それにより，興奮している状態のシーンを抽出する．尚，コメントデータ部を利用した閲覧者感情の分析目的はプライバシー侵害動画記事の収集である．それに対し，のキャプションを利用した当事者感情の分析目的はプライバシー侵害動画記事から興奮している状態のシーンを抽出することである．

### 4.2 抽出の処理フロー

T:Tweeter が興奮している状態が見られるシーン抽出のための処理フローは図 2 の通りとする．Step1 は奈良ら[3]により提案されたプライバシー侵害動画記事を抽出する方式で，これに興奮状態シーンを抽出するための方式である Step2 を追加した．

<u>Step1 コメント解析によるプライバシー侵害動画記事の抽出</u>

Profile of Mood States（これを以降, POMS とする）[13]を利用しコメントに対する感情抽出を行った．S: Subscriber が感情的な意見を残している動画記事が収集される．その中から,コメントに PS ワード[14]を含む動画記事を抽出することで，プライバシー侵害を指摘している動画記事が抽出される．

<u>Step2 キャプション解析による興奮状態のシーン抽出</u>

Step1 で抽出されたプライバシー侵害動画記事の全てがプライバシー侵害シーンとは考えにくい．そこで，属性データ内のキャプションより当事者感情を定量化した．キャプションは 4 秒単位で表示されており,全てのキャプションに感情分析行って感情の推移を把握する．その中で，T:Tweeter の興奮している状態が検知されたシーンのみ抽出することとする．

ⓒ 2014 Information Processing Society of Japan

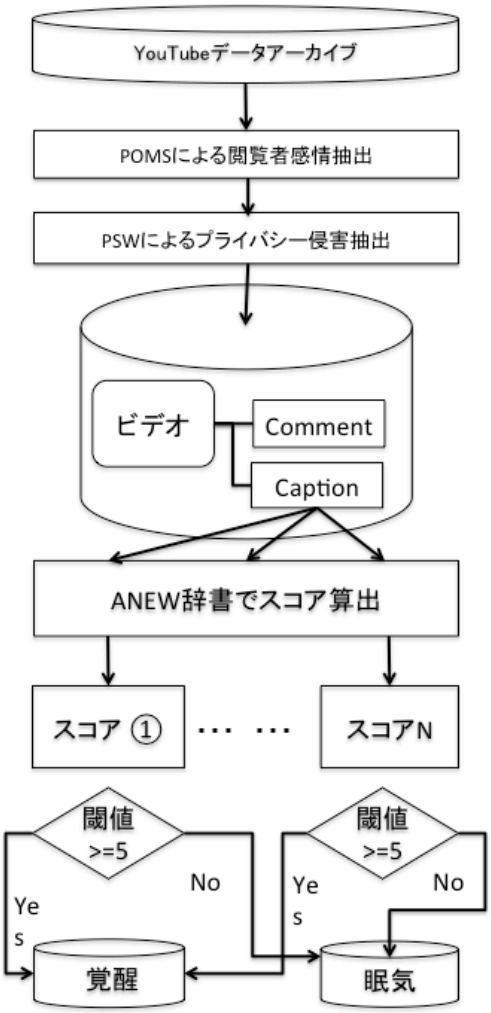


図 2 興奮状態のシーン抽出フロー

### 4.3 ANEW による感情抽出手法

奈良ら[]は POMS による感情分析の手法を適用したが，本稿では興奮している状態を抽出する必要がある.そこで，英語の標準的な感情表現をスコア化した ANEW(Affective Norms for English Words)[15]を採用した．ANEW は valence（誘発），arousal（覚醒），dominance（優位）の 3 つの次元で評点化された 1,034 語の単語セットである． valence（誘発）は幸福からスコアの低下に伴い不幸となり，肯定-否定の表現に適している. arousal（覚醒）は興奮からスコアの低下に伴い穏やかとなり，興奮している状態の表現に適している.

$$\mu_a = \frac{\sum_{i=1}^{n} \mu_{i,a}}{n}$$

当事者感情の抽出は，ANEW の arousal（覚醒）を参照し，算出した．算出方法は上記の通りで，抽出できた単語数と単語のスコアの平均値$\mu_\alpha$とした. α は arousal（覚醒）のスコア, n は ANEW で抽出できた単語数としている.

ANEW の標準的な閾値を利用した場合は，閾値 5.0 を基準に平均値$\mu_\alpha$1.0-5.0 未満は眠気･穏やかな状態となり, 5.0-9.0 は覚醒・興奮している状態となる．但し，本稿ではプライバシー侵害シーン抽出向上に伴い閾値を適宜変更することとする．キャプションから，T:Tweeter が興奮している状態のシーンを抽出する．抽出したシーンがプライバシー侵害シーンである確率を算出し，比較することで当事者感情が興奮状態であることとプライバシー侵害の親和性の高さを次章の評価実験にて明らかにする.

## 5. 評価実験

### 5.1 実験方法

プライバシー侵害と当事者感情が興奮状態であることの親和性の高さを明らかにするために，評価実験を行うこととする．今回は，当事者感情を分析することによるシーン抽出が目的のため，閲覧者感情の分析は評価対象外とする. Step1 にて収集した動画記事からアノテーターにより，予めプライバシー侵害の可能性があるシーンの選定を行った. これらを正解データとし，提案手法に基づいて抽出されたシーンがどれだけ正解データを網羅しているか検知率で把握する.ANEW にて覚醒の感情が示されたシーンのプライバシー侵害シーンの検知率と覚醒の対極に存在する眠気の感情が示されたシーンのプライバシー侵害シーンの検知率を比較することとする.

評価実験は 2 パターン行うとする.

1） 興奮状態シーンと眠気状態シーンの検知率比較
興奮状態として抽出されたシーンと眠気状態として抽出されたシーンそれぞれ検知率を算出し比較する.

2） ANEW の閾値変更による適合率と再現率の比較
本来，閾値 5.0 を基準に高スコアの場合は覚醒，低スコアの場合は眠気としている.今回抽出量増加のために，閾値を 3.0 から 6.0 の 1.0 単位で変更し適合率と再現率をそれぞれ算出し比較する．尚,適合率と再現率算出のための定義は以下の通りとなる.

True/True: 感情に興奮状態が見られ，プライバシー侵害シーンと判定した数

True/False: 感情に興奮状態が見られなかったが，プライバシー侵害と判定した数

False/True: 感情に興奮状態が見られても，プライバシー侵害シーンと判定しなかった数

False/False: 感情に興奮状態が見られなく，プライバシー侵害と判定しなかった数

### 5.2 実験結果

検知率の実験を行った結果はそれぞれ表 2,表 3 で示す. 表 2 では，興奮状態のシーンがプライバシー侵害シーンと

なるシーン検知率は眠気状態のシーンよりも高くなっている．これにより，当事者感情が興奮している状態であるシーンはプライバシー侵害となる可能性が高いといった
当事者感情の興奮状態とプライバシー侵害の親和性の高さは明らかとなった．

表 3 では．閾値を変更することで適合率，再現率からの判断により ANEW の標準閾値となる 5.0 が適切と考える．閾値を高くしてしまうと，プライバシー侵害の可能性があるシーンが抽出されない． また,閾値を低くしてしまうと，プライバシー侵害でないシーンまで抽出されることとなる．

表 2 感情分析によるプライバシー侵害シーン検知率

| 覚醒状態でのプライバシー侵害シーン検知率 | 眠気状態でのプライバシー侵害シーン検知率 |
|---|---|
| 43. 2% | 10. 8% |

表 3 ANEW の閾値変更による適合率と再現率

| ANEW 閾値 | 適合率 | 再現率 |
|---|---|---|
| 3.0～ | 54.0% | 52.6% |
| 4.0~ | 54.2% | 50.0% |
| 5.0~ | 57.1% | 53.3% |
| 6.0~ | 2.9% | 20.0% |

尚，プライバシー侵害動画記事の事例の一つとして，感情の推移からよみとれる感情の興奮状態とプライバシー侵害シーンの関係を図 3 に示す．グラフで示しているのは感情のスコア値となり，点線はプライバシー侵害シーンとして抽出されたキャプションとなる．
さらに，興奮状態とプライバシー侵害シーンが一致したときの Youtubu 画像を図 4 に示す．

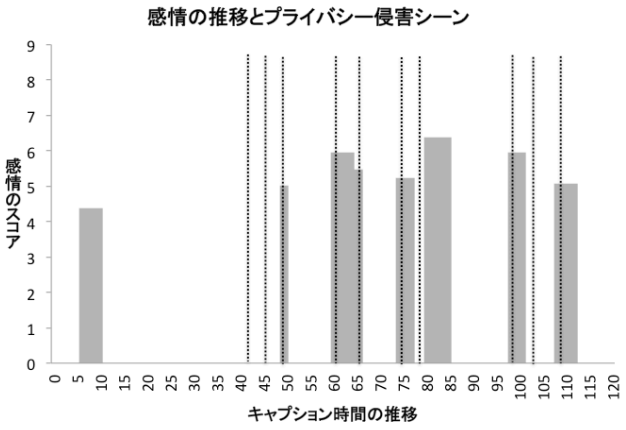


図 3 感情の推移とプライバシー侵害シーン

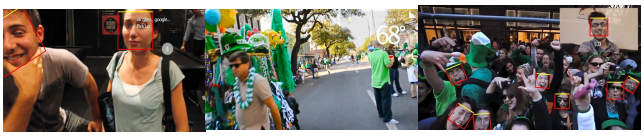

図 4 プライバシー侵害シーンとして抽出されたシーン

## 6. まとめ

本稿では，キャプションの感情分析した結果当事者感情が興奮状態であることとプライバシー侵害の親和性の高さを明らかにし,プライバシー侵害シーンの抽出することを目的とした．当事者感情が興奮している状態の時はプライバシー侵害が発生している可能性が高い．

今後は,ウェアラブルカメラを利用した画像や動画を SNS へアップロードする直前にプライバシー侵害を予測し、警告することで侵害を防止する方式（Opt-in 方式）にキャ T:Tweeter の感情というファクターを含めることを検討することとする．

## 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
2) 町田史門，小山貴之，宋洋，嶋田茂，越前功，デジタル私文書におけるプライバシーセンシティブ情報の漏洩検知に基づく公開範囲の設定方式の提案，電子情報通信学会 EMM 研究会技術報告,2013.
3) 奈良育英，高田さとみ，高田 美樹，大本 茂史，岸本 拓也，周 子胤，嶋田 茂, SNS の感情分析をトリガーにしたウェアラブルカメラによるプライバシー侵害の要因分析，電子情報通信学会 HCS 研究会技術報告,2013.
4) 「ネット炎上」を未然に防止する対策について，http://www.smbc-consulting.co.jp/upload/netpress_8717_.pdf
5) Data Protection Act 1998, http://www.legislation.gov.uk/ukpga/1998/29/contents
6) Federal Data Protection Act (BDSG), http://www.bfdi.bund.de/EN/DataProtectionActs/Artikel/BDSG_idFv01092009.pdf?__blob=publicationFile
7) 青柳武彦,日経新聞個人情報保護に行き過ぎ公益との調和必要プライバシー論議深めよ,2005.09.19
8) 小山貴之，宋洋,町田史門,嶋田茂，越前功， 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.
9) Y.Wang, G.Norcie, S.Komanduri, A.Acquisti, P.Leon, and L.Cranor, I regreted the minute I pressed share: A qualitative study of regrets on Facebook. SOUPS2011, 2011.
10) Huina Mao, XinShuai, ApuKapadia , Loose Tweets: An Analysis of Privacy Leaks on Twitter ,2011
11) Facebook 利用規約, https://www.facebook.com/legal/terms
12) YouTube Data API, https://developers.google.com/YouTube/
13) McNAIR et al. , Manual for the Profile of Mood States. San Diego, CA: Educational and Industrial Testing Service.
14) 高田さとみ,小山貴之,町田史門,宋洋,嶋田茂, 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.
15) M. M. Bradley and P. J. Lang. Affective Norms for English Words (ANEW): Stimuli, instruction manual, and affective ratings. (Te

ⓒ 2014 Information Processing Society of Japan