

**科学研究費助成事業 研究成果報告書**

平成 27 年 6 月 26 日現在

機関番号：32682

研究種目：挑戦的萌芽研究

研究期間：2013～2014

課題番号：25540080

研究課題名(和文) 情報の曖昧化に関する研究

研究課題名(英文) Study on Information Clouding

研究代表者

中村 聡史 (Nakamura, Satoshi)

明治大学・総合数理学部・准教授

研究者番号：50415858

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：コンテンツを楽しみにしている人にとってネタバレ情報は忌むべきものである。本研究では、ウェブブラウザの拡張としてブラウジング中のネタバレを防止する手法や、Twitter上でのネタバレを防止するための手法を実現した。また、小説やマンガ、映画やドラマなどのネタバレに関するデータセットを構築してその分析を行いその特徴を明らかにするとともに、動画に重畳される形で提示されるコメントからのネタバレ除去に関する研究を行った。

研究成果の概要(英文)：Spoilers often spoil the pleasure of a user who is looking forward to enjoying the content. In this project, we proposed and implemented several systems to diminish spoilers on the Web and on the Twitter application. In addition, we generated the data set of spoilers about novels, comics, movies and dramas, and showed characteristics of spoilers by analyzing the data set. Furthermore, we proposed a method to detect spoilers from social comments to video clips on the video sharing web sites.

研究分野：ヒューマンコンピュータインタラクション

キーワード：情報曖昧化 ネットバレ Twitter ウェブ

## 1. 研究開始当初の背景

学校や仕事の都合でスポーツ番組をリアルタイムに視聴できず、録画し、家に帰ってから視聴することを楽しみにしている人にとって、そのスポーツの試合結果を示唆するようなネタバレ情報は避けたいものである。また、小説やマンガ、ドラマやアニメなどの展開に関するネタバレ情報もまた、そのコンテンツを楽しもうとしている人にとっては遮断したいと思っているものである(図1)。



図1 ネット情報に出会ってびっくり

このネタバレに関する問題は、コンテンツを全ての人と同じタイミングで楽しむことが出来ない事が理由であり、そのネタバレの情報元を遮断することである程度対処することは可能である。ここで、ネタバレの情報元をウェブ上に限ったとしても、ウェブ上の情報を全て遮断することは、生活などにおける情報収集および調査手段を大きく狭めてしまうものであり、容易ではない。また、TwitterやFacebookなどに代表されるソーシャルネットワークサービスはコミュニケーションツールとして重要な位置を占めており、こうしたサービスを多用しているユーザにとって、そのサービスを遮断することは他者とのコミュニケーションを取らないという事にも等しいものである。

ここで、情報を取り扱う研究としては、膨大な情報の中から適切な情報をユーザに提示する情報推薦技術や、情報量の多いコンテンツを短縮する情報要約技術、有害なコンテンツを遮断するフィルタリング技術など、様々な技術が進化している。しかし、先述のネタバレというユーザ視点に立った場合、上記の技術はユーザにとって十分なものではなく、余計なお世話である場合もある。

例えば、ユーザがとても好きなスポーツの試合を録画し、視聴を楽しみにしている場合、ユーザはその録画を視聴するまで試合結果に関する情報を積極的に遮断しようとするであろう。しかし、情報推薦技術は、そのユーザが興味を持つ情報を積極的に提示する

ため、スポーツの試合の結果をユーザに推薦し、ユーザの楽しみを低減させてしまう可能性がある。また、今後電子書籍が普及し、多くのコンテンツが流通するようになると、書籍の概要を作成するのに自動要約を用いることが考えられる。しかし、特に推理小説では、その要約を読むことで犯人の名前がわかってしまったり、ラストの展開がわかってしまったりしたら残念に思うことであろう。情報フィルタリング機能を備えたソフトウェアでは、多くの人にとって有害な情報を遮断してくれる。しかし、何がネタバレになるかは、個々人の興味や嗜好によって大きく異なり、事前に定義されたルールで除外できるものではない。

ネタバレについては、ユーザが興味を持っているからこそその情報の意味を曖昧化し、その核心に触れることを避ける事を可能にする技術を実現することが重要となる。これまでも、核心情報の及ぼす影響の調査や、Twitterにおける核心情報の遮断に関する調査研究などがあるが、どのようにして核心となる情報との接触を防ぐかといった研究は十分になされてはならず、研究として取り組む必要性がある。

## 2. 研究の目的

先述の通り、スポーツの試合結果や、テレビドラマの展開、推理小説の犯人の名前などのネタバレ情報(核心情報)は、そのコンテンツをこれから楽しもうと思っている人にとっては忌むべきものである。そこで本研究では、ユーザによって異なる、ユーザの見たくない・読みたくない情報を考慮し、自動的にネタバレ情報を曖昧化する情報曖昧化技術を実現する事を目的とする。この技術により、ユーザはどんな情報があるのかを把握しつつも、試合結果や犯人の名前といった直接的な情報との接触を避ける事を可能になると期待される。

## 3. 研究の方法

コンテンツを楽しみにしている人が、ウェブブラウジング中にニュースサイトやブログサイト、検索エンジンやまとめサイトなどにおいてネタバレ情報と出会ってしまうことを防ぐため、単純にネタバレ該当部分をなくしてしまう非表示手法、ネタバレ該当部分を塗りつぶす墨塗り手法、ネタバレ該当部分を任意の確率で反転させる結果反転手法、ネタバレ該当部分を増殖し結果を反転させたものを混ぜる木の葉を隠すなら森の中手法という4つの情報曖昧化手法(図2)をブラウザ拡張として開発する。また、ネタバレ防止手法を利用している時にユーザの視線がどのように動くのかということを明らかにするため、視線検出および可視化システムを構築する。

### Original

- Japan wins women's world cup 2011
- Japan 2-2 (PK: 3-1) USA
- Sweden 2-1 France

### Invisible

- 
- 
- Sweden 2-1 France

### Censored

- [REDACTED]
- Japan [REDACTED] USA
- Sweden 2-1 France

### Increasing

- Japan wins women's world cup 2011
- Japan loses women's world cup 2011
- Japan 2-4 (PK: 3-2) USA
- Japan 2-2 (PK: 2-4) USA
- Japan 2-2 (PK: 3-1) USA
- Sweden 2-1 France

### Inverse

- Japan wins women's world cup 2011
- Japan 2-2 (PK: 3-5) USA
- Sweden 2-1 France

図 2. 情報曖昧化手法を適用

Amazon などの EC 系サイトでは、小説やマンガ、映画やアニメの DVD などを買っており、そうした商品を購入する際に購入者のレビューを参考にすることが多々ある。こうしたレビュー記事においてネタバレ情報が提示されていることも多く、がっかりすることは珍しくない。そこで、こうしたレビュー記事からネタバレを推定する技術について取り組む。その第一歩目として、まずはレビュー記事に含まれるあらすじの判定について取り組む。次に、レビュー記事からのネタバレ除去に取り組むため、ネタバレはどのように記述されるのかということ明らかにするため、小説やマンガ、映画やアニメ、ゲームなどのコンテンツに対するネタバレデータセットを構築する。さらに、小説の中身の記述と、小説に対するネタバレがどのような関係にあるのか、どういったところでネタバレは記述されやすいのかといったことを明らかにするために、著作権が消滅している青空文庫を対象として、多面的なネタバレデータセットを構築する。

一方、Twitter や Facebook といったソーシャルネットワークサービス (SNS) 上では、多くの人が、自身が楽しんだコンテンツについて感想などの情報発信を行っている。こうした情報には、SNS 自体を遮断することで触れないようにすることが可能であるが、SNS はコミュニケーションツールとしても成り立っているため、SNS を遮断することは他者とのコミュニケーションも遮断することになってしまい難しい。そこで、こうした SNS 上でのネタバレを防止するため、Twitter アプリケーションのプラグインとして情報曖昧化を行う手法を開発する。ここでは、単純に

ネタバレ該当部分をなくしてしまう非表示手法、ネタバレ該当部分を塗りつぶす墨塗り手法、ネタバレ該当部分を任意の確率で反転させる結果反転手法、ネタバレ該当部分を増殖し結果を反転させたものを混ぜる木の葉を隠すなら森の中手法の 4 手法を実装し、その有用性について検討を行う。

ニコニコ動画などの動画の時間軸に対するコメントが投稿可能であり、動画とともに他者のコメントを楽しむことができるような動画共有サイトは、他者と感情を共有しながら視聴できるという点で成功を収めている。しかし、こうしたコメントにはしばしばネタバレ情報が含まれているものであり、公式的に配信されているアニメやドラマなどのネタバレを示され、がっかりしてしまった視聴者は少なくない。そこで、こうしたニコニコ動画上の動画コンテンツに対するコメントでのネタバレを防止するため、機械学習の手法を用いることによりコメントに含まれるネタバレを検知する仕組みを実現する。

## 4. 研究成果

まず、提案している 4 手法を Mozilla Firefox および Google Chrome という有名な 2 つのウェブブラウザのブラウザ拡張として実装し、ウェブ上で配布した。また、実装したシステムを長期的に運用テストすることでメリット・デメリット、さらにネタバレ判定の難しさについて明らかにすることが出来た。また、複数のユーザに体験してもらい、フィードバックをもらうことができた。さらに、この各種法についてその認知的負荷がどの程度あるのかということ調べるため、アイトラッキング装置を用いてユーザの詳細な分析を行うためのシステム開発を行い、ユーザの振る舞いに関するデータ収集を行うことができた。

次に、Amazon などの EC 系サイトからのネタバレを除去する一歩目として、あらすじを判定する仕組みについて研究を行った。ここでは特に、レビュー記事の文書構造やクラス情報などを利用することで、あらすじを分類することができた。さらに、あらすじを分類するだけでなく、ネタバレに取り組むために、青空文庫を対象としたデータセットを構築した。ここでは、6 人の実験協力者に 5 冊の本 (こころ、デパートの絞刑事、モルグ街の殺人、銀河鉄道の夜、赤髪連盟) を読んでもらい、ネタバレを書き下し、そのネタバレに度合いに応じたスコアを付けてもらった。このデータセットを利用することにより、人によってネタバレと思うレベルは異なること、ストーリーの後半にかけてネタバレは増加する傾向があること、人物名や地名などアイテム独自の単語がネタバレとなりやすいことなどを確かめることができた。

2 つ目のデータセットでは、複数名の協力者に任意のストーリー型コンテンツ (ドラマ、

アニメ、小説、漫画、ゲーム、映画を対象)について、巻数や話数を限定しながら、この情報に触れてしまうとコンテンツの楽しみが減退するというネタバレを簡潔に記述してもらった。この作業により500を超えるコンテンツに対するネタバレデータセットが構築され、どのような記述がネタバレとなる傾向があり、それを防ぐためにはどうしたらよいかということを検討可能となった。このデータセット構築により、ネタバレとしては「勝敗にまつわること」「生死にまつわること」「犯人にまつわること」「恋愛成就の成否にまつわること」「話の進度にまつわること(話が展開しないなど)」といったことがネタバレとして記述されやすいということがわかった。

一方、Twitterからのネタバレ防止を可能とするため、Windows OSおよびMac OSに対応しており広く利用されているJanetterというTwitterのクライアントソフトウェアのプラグインとして情報曖昧化手法を実装した。図3の左は墨塗り手法としてネタバレ防止を行っている様子で、図3の右は結果反転手法によりネタバレ防止を行っている様子である。図4はシステムの設定画面である。また、このシステムでは、辞書ベースでネタバレを防止するだけでなく、そのコンテンツに対応するハッシュタグ(あるジャンルに対するツイートをまとめるために、#soccer #japan #nhkなどのように発信者が付与するもの)を登録することで、そのハッシュタグを含むツイートから自動でネタバレと思しき語を推定し、その部分を情報曖昧化する、簡易的な手法も実現した。

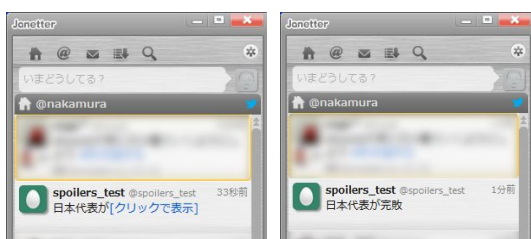


図3. Twitter上でのネタバレ防止

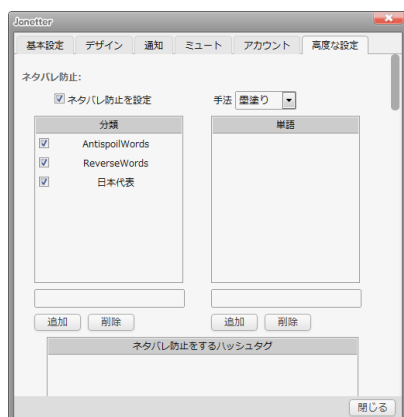


図4. 設定画面

さらに、ニコニコ動画などのように動画の時間軸に対してコメントを付与可能なシステム上でのネタバレを排除することを目的として、「名詞+は」「名詞+が」などのようなパターンやバーストする単語が別の場所で現れることを考慮するといったルールベースの手法と、教師付き学習に基づく手法によってネタバレコメントを自動検知する仕組みを実現し、ネタバレ検知の難しさについて明らかにした。

2年間ウェブ上におけるネタバレ防止に取り組んだことにより、ネタバレ防止に関する関連研究も増えてきており、研究分野を開拓することが出来た。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 0件)

〔学会発表〕(計 5件)

1. 田中駿, 廣田壮一郎, 高村大也, コメント機能付動画共有サービスにおけるネタバレ検知, 人工知能学会全国大会2015, 査読なし, 2015.
2. 前田恭佑, 土方嘉徳, 中村聡史, ストーリー文書内のネタバレの記述に関する基礎的調査, 第6回 ARG Web インテリジェンスとインタラクション研究会, 査読なし, 2015.
3. 前田恭佑, 土方嘉徳, 中村聡史, ストーリー文書内のネタバレの記述に関する基礎的調査, 第59回システム制御情報学会研究発表講演会(SCI'15), 査読なし, 2015.
4. 中村聡史, 川連一将, スポーツのネタバレを防止するTwitterクライアントの開発と諸検討, 第4回 ARG Web インテリジェンスとインタラクション研究会, 査読なし, 2014年5月25日.
5. 岩井秀成, 土方嘉徳, 西田正吾, レビュー文に対する文書構造と複数のクラス情報を考慮したあらすじ分類, WebDB Forum 2013 (第6回 Web とデータベースに関するフォーラム), 査読有り, 2013

〔図書〕(計 1件)

1. 人を幸せにする目からウロコ! 研究, 萩原一郎編著, 岩波ジュニア新書, 2013, pp.1-233 (p.1-21を中村聡史が担当).

〔産業財産権〕

出願状況(計 0件)

取得状況(計 0件)

〔その他〕

ホームページ等

ネタバレに関するデータセット  
<http://netabare.meguri.org/>

## 6. 研究組織

### (1) 研究代表者

中村 聡史 (NAKAMURA, Satoshi)  
明治大学・総合数理学部・准教授  
研究者番号：50415858

### (2) 研究分担者

小松 孝徳 (KOMATSU, Takanori)  
明治大学・総合数理学部・准教授  
研究者番号：30363716

### (3) 研究分担者

土方 嘉徳 (HIJIKATA, Yoshinori)  
大阪大学・基礎工学研究科・准教授  
研究者番号：10362641

### (4) 研究分担者

高村 大也 (TAKAMURA, Hiroya)  
東京工業大学・精密工学研究所・准教授  
研究者番号：80361773