2I1-2

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

Detecting Story Spoilers for Commentable Video Hosting Service

田中駿 Shun Tanaka　廣田壮一郎 Soichiro Hirota　高村大也 Hiroya Takamura

東京工業大学
Tokyo Institute of Technology

We apply a rule-based method and a machine-learning-based method to story spoiler detection in commentable video hosting services. In both methods, we used textual patterns and burstness of words as well as post time. The result suggests that the simultaneous use of the textual patterns and the burstness of words is effective under the strict setting where no comments of the target movie are available as training data, and that the unigram features and the post time features show better performance compared with other features under the lenient setting where some comments of the target movie are available as training data.

1. はじめに

ニコニコ動画 *1 などの動画共有サービスには，ユーザがコメントを投稿できる機能がある．コメントは動画中の各時点に紐付けられており，その場面に関する様々な感想や意見などが記述されている．コメントは動画を再生する際に表示され，視聴者がより動画を楽しめるという効果を生む．しかしその一方，動画中のある時点でそれより後の時点の内容を記述してしまう，いわゆるネタバレが問題になっている．例えばミステリにおいて，事件の犯人やトリックについて記述してしまうことにより，視聴者が推理を楽しめなくなるなどの事態が発生することがある．このような背景のもと，ネタバレを含むコメントを自動的に検知する技術の開発が望まれている．

そのために本稿では，ネタバレの検知においてコメントのテキスト部分がどのように使えるかについて考える．特に，コメント内の単語が，動画の内容に関する重要な情報を含むか否かの推測が検知においてどのように役に立つかについて調査する．動画に対するコメントにおけるネタバレ検知は，レビューにおけるネタバレ検知と関連した研究課題であるが，いくつか本質的な差異がある．まず，動画に対するコメントは一般的に非常に短いという特徴がある．また，動画から得られる情報をユーザが共有しているという前提があるので文脈がテキストに明示されていないことも特徴的である．これらの特徴により，動画に対するネタバレコメントの検知は非常に挑戦的な研究課題となっている．それゆえ，どのような状況においてどのような情報が検知に効果的であるかについての知見は，ネタバレ検知システムの構築において非常に有用になると考えられる．本稿では，コメントのテキストに加え，単語のバーストや投稿時刻 (コメントが紐づけられた動画中の位置) の情報について調査する．

2. 関連研究及び要素技術

2.1 関連研究

岩井らは，書籍などのレビューにおけるネタバレを検知するために，あらすじを表す文を分類する手法を提案した [岩井ら 13]．Guo らは，映画のレビューを，ネタバレを含むものとそうでないものに分類する手法を提案した [Guo 10]．いずれもレビューを対象としており，動画に対するコメントは短くかつテキストからわかる文脈情報が僅かであるという点で，課題の質が異なっている．Boyd-Graber らも映画のレビューを対象にしており，各文をネタバレを含むものとそうでないものに分類する手法を提案した [Boyd-Graber 13]．彼らは同じ映画に関するデータが訓練データと評価データに入らないような実験設定に従うことで，より一般的な状況で実験を行っている．しかし，データにおけるネタバレとそうでないものが同数となるように設定しており，この点においてはやや人工的な設定となっている．

中村らはスポーツ中継に関するネタバレツイートを検知する手法を提案した [中村ら 13]．彼らの用いたキーワードマッチングによる手法は，スポーツ中継のように使用語彙の多様性が低いと思われるドメインでは有効であるが，一般の動画において高い性能を期待することは難しい．Jeon らは，テレビ番組に関するネタバレツイートを検知する手法を提案した [Jeon 13]．彼らは，固有表現，頻出動詞，URL の有無，主な時制を素性とした Support Vector Machines (SVMs) を用いた．あるリアリティショーの 1 シーズン (12 エピソード) 分のツイートをデータとして用い，3 分割の交差検定で評価している．同じテレビ番組のデータが訓練に使える設定を考えている点，およびリアリティショーという特定のドメインに特化しているという点が特徴的である．

2.2 要素技術

ここでは，本稿で提案する手法において要素技術として用いる，藤木らによるバースト単語判定方法について述べる [藤木ら 04]．この判定方法においては，ある単語の発生時刻の系列が与えられたとき，その各発生間隔 t を用いてバースト判定を行う．より具体的には，λ を動画内の発生間隔の平均として，定常状態の指数分布 ($\lambda e^{-\lambda t}$) とバースト状態の指数分布 ($s\lambda e^{-s\lambda t}$, $s(>1)$ はパラメータ) を考える．そして，これらの分布が t に与える確率値 p に対して，$-\log p$ を状態の持つコストとし，さらに定常状態からバースト状態に遷移するときに遷移コスト τ が必要となるとする．その上で，系列全体で課されるコストの総和が小さくなるように Viterbi アルゴリズムを用いて各時点での状態を判定する．

連絡先: 田中駿，**shun@lr.pi.titech.ac.jp**

*1 http://www.nicovideo.jp

3. データ構築および分析

実験データの動画はニコニコ動画より API を用いて取得した．5 つの動画について，それぞれ投稿から 24 時間経った時点での最新 1000 件のコメント (old)，投稿から一ヶ月経った時点での最新 1000 件のコメント (new) を取得した．

表 1 に実験データの統計値を示す [*2]．各コメントについて

表 1: 実験データ

タイトル	長さ (分:秒)	ネタバレ数 /全コメント数
進撃の巨人　第 1 話「二千年後の君へ」	24:16	26/1980
シュタインズ・ゲート 第 1 話「始まりと終わりのプロローグ -Turning Point-」	23:47	90/1949
劇場版「空の境界」第一話「伽藍の洞 I」	23:46	91/1901
TV アニメ「Fate/stay night [Unlimited Blade Works]」＃ 00 プロローグ	47:49	123/1871
ダンガンロンパ　＃ 1「ようこそ絶望学園」	24:33	47/1958
合計	–	465/9659

3 人のアノテータにネタバレか否かを判定してもらい，多数決によって最終的なラベルを決定した．動画の長さは，最終コメントの投稿時刻 [*3] により決定している．

表からわかるように，全コメントにおけるネタバレコメントの割合は動画によって異なるが，平均的には約 5%に過ぎず，ネタバレでない方に大きく偏ったデータであることがわかる．また，投稿時刻に対してコメント数がどのように変化するか，およびネタバレコメント数がどのように変化するかを度数分布で表現し，図 1 に示す．1 分単位で度数を算出している．図からわかるように，全体のコメント数とネタバレコメント数は同じ形の曲線で表現されてない．すなわち，ネタバレは，通常のコメントと異なる要因によって生じていることが示唆される．

個々のコメントについて分析すると，テキストに現れている手がかりが非常に少なく，高精度の検出には高度な背景知識が必要であることわかる．例えば,『劇場版「空の境界」』において「黒桐！」というコメントがネタバレとされているが，表層的な手がかりは少なく，動画の内容を理解していなければこのコメントをネタバレと判定することは困難である．また，例えば「犯人はヤス。」[*4] のように，表層的にはネタバレと思われるが実はそうではなく，ユーザの作るコミュニティ内でのみ理解しうる特殊な意味を持ったコメントなどもある．

*2　各動画には以下の URL よりアクセス可能である：
http://www.nicovideo.jp/watch/1365403220
http://www.nicovideo.jp/watch/1302085709
http://www.nicovideo.jp/watch/1372908375
http://www.nicovideo.jp/watch/1412240325
http://www.nicovideo.jp/watch/1373013567

*3　本稿では，コメントが紐づけられた動画の再生時間中の位置を投稿時刻とよぶ．

*4　http://dic.nicovideo.jp/a/犯人はヤス

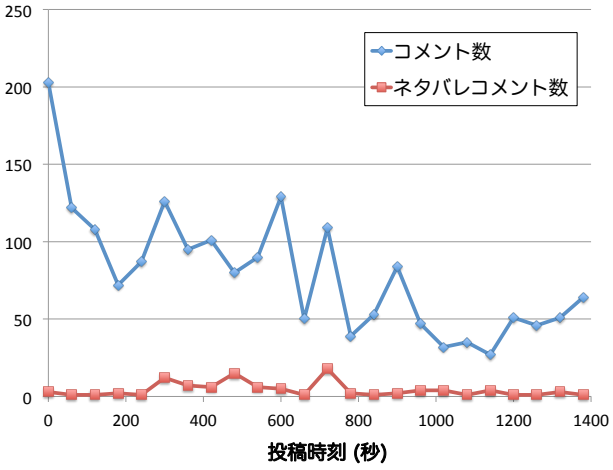

図 1:『劇場版「空の境界」』におけるコメント数の度数分布

4. ネタバレ検知手法

テキスト情報を主に用いたネタバレ検知手法を考える．ここでは，訓練データを必要としない単純なルールベース手法と，機械学習に基づく手法を考える．いずれも，類似した情報を用いている．まず，ネタバレは「犯人は [人名]」のような形式で出現しやすいという想定の下，そのようなテキストパターンの出現をルールベース手法および機械学習手法の両方で用いる．また，コメント内の単語が重要な情報を含んでいるかを捉えるために，単語のバーストを用いる．単語のバーストの計算には，対象動画のコメントをすべて利用する．テレビ番組などと異なり，動画共有サービスではオンデマンドで動画が再生できる．よって，ある時点でのコメントを分類するにあたり，それよりあとの時点のコメントも利用できるという設定が現実的である．これに対し，テレビ番組などに対するツイートにおけるネタバレ検知では，あとの時点のツイートを使えないとする設定が現実的であり，両者は問題の性質が大きく異なることに注意されたい．

4.1 ルールベース手法

次の 3 種類のルールベース手法を考える．

1. **RULE+ptrn**:“[名詞]+は” あるいは “[名詞]+が” というテキストパターンを含むコメントはすべてネタバレとする．
2. **RULE+brst**:以降の時点でバーストする単語を含むコメントをネタバレとする．
3. **RULE+ptrn+brst**:“[名詞]+は” あるいは “[名詞]+が” というテキストパターンを含み，以降の時点でバーストする単語を含むコメントをネタバレとする．

4.2 教師付き学習に基づく手法

以下の 3 種類の素性を用いた対数線形モデルを考える．

1. **1grm**:単語 1 グラム
2. **time**:投稿時刻．各動画を時間に従い 20 分割したうちのいずれの時間帯に，対象コメントの投稿時刻が対応するか

3. **brst**:対象コメントが，その投稿時刻以降でバーストしている単語を含むか

ネタバレのクラスとそうでないクラスにおいて，サイズが後者に大きく偏っていることが 3 節より事前にわかっている．この問題に対応するために，単純に対数線形モデルの分類結果を用いるのでなく，対数線形モデルはコメントが与えられたときのネタバレクラスの条件付き確率を出力するので，この確率がある閾値 th を上回った場合にネタバレであると判定する．

5. 実験

5.1 実験設定

実験データには 5 つの動画があるので，このうち 3 つを訓練データ，1 つを開発データ，残りの 1 つを評価データとする．5 つの動画をこのように分けるすべての分け方について実験を行い評価指標の値を算出し，その平均値を出した．評価指標としては F 値を用いた．本設定では，mean average precision なども評価指標の候補となる．しかし，ルールベース手法は事例をランキングすることができないので，mean average precision などのランキングに基づく指標を算出できない．結果として mean average precision ではルールベース手法と機械学習手法を比較することができなくなるので，ここでは F 値を用いた．

対数線形モデルの実装としては，liblinear[*5] を用いた．開発データを用いて，正則化パラメータ C (1, 10, 100, 1000, 10000)，および対数線形モデルの閾値 th (0.50, 0.40, 0.30, 0.20, 0.10, 0.09, 0.08, ⋯, 0.01) を調整した．対数線形モデルの出力する確率が閾値 th を上回った場合に，そのコメントはネタバレであると判定する．また，バースト単語判定のパラメータについては，$s = 10$，$\tau = 1$ とした．

加えて，同一動画のコメントが訓練データと評価データに入っている場合の評価を行う．すなわち，各動画について何らかの方法で訓練データを作成できると仮定した場合の評価である．具体的には，各動画を 3 節に記述した old と new におおよそ二等分し，それぞれを訓練データと評価データとして用いて実験した．ここでは最適な C の値を用い [*6]，様々な閾値 th についての F 値を算出してグラフで示すことにする．

5.2 実験結果

表 2 に実験結果を示す．RULE がルールベース手法，CLS が機械学習分類器に基づく手法であり，それぞれ brst，ptrn，1grm などで用いるルールや素性を記述している．また，すべてネタバレと判定した場合をベースラインとしている．まず，いずれの手法も F 値が非常に低く，本研究課題の困難さを示している．単純なパターンに基づくルールベース手法 (RULE+ptrn) と 1 グラムに基づく機械学習手法 (CLS+1grm) を比較すると，ルールベース手法の方が良い．これは，1 グラムにより多くの素性を導入しても，現状のデータ量では単純なパターン以上の手がかりが得られていないことを示している．また，ルールベース手法と機械学習手法のいずれにおいても，テキスト情報と単語バーストを同時に用いることで，僅かながら性能向上があり，その効果が示されている．また表では省略したが，投稿時刻に関する素性 (time) の有効性は示されず，むしろ F 値は減少した．これは，各動画によってネタバレが発生しやすい時間帯は異なっていることを示している．

表 2: 実験結果

手法	F 値 (%)
ベースライン	7.5
RULE+brst	9.0
RULE+ptrn	15.3
RULE+ptrn+brst	15.7
CLS+brst	9.0
CLS+1grm	14.7
CLS+1grm+brst	15.4

同一動画のコメントが訓練データと評価データに入っている場合の機械学習手法の結果を，図 2 に示す．ルールベース手法で最も高かった F 値は，RULE+ptrn+brst の 22.1 である．訓練データ量は増加していないにも関わらず，表 2 の数値と比較すると，図 2 の数値は大きく向上していることがわかる．各動画のコメントに人手でラベル付けをするのはコスト面から現実的でないが，ユーザがコメントに付与するスコア [*7] などを有効利用することで，性能向上が期待できる．また，投稿時刻の素性 (time) を追加することで F 値が向上していることから，動画ごとにネタバレが発生しやすい時間帯が存在することがわかる．

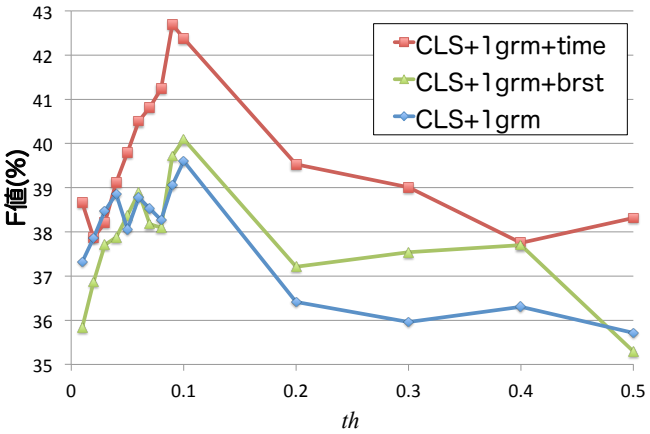

図 2: 同一動画のコメントが訓練および評価データに入っている場合の F 値．横軸は閾値 th

実際はネタバレであるが正しくネタバレと判定できなかった例を挙げる．例えば，女性と思われるキャラクタが映像で映っている際に，「こいつ男」などのようにまだ明らかにされていない設定を明かすコメントがあった．これを正しく処理するためには，映像の情報とコメントの情報が矛盾していることを認識する必要があり，非常に難しい例である．また，ストーリーが盛り上がる直前に「くるぞー」などのように，まもなく盛り上がることを明かしてしまうコメントがあった．これは，このコメントが盛り上がりを示唆していること，さらに実際にその直後に盛り上がりがあり，このコメントがこの盛り上がりを参照していることを認識する必要がある．

*5 http://www.csie.ntu.edu.tw/~cjlin/liblinear/
*6 実際の応用では，開発データを用いて C の値を決定する．
*7 ニコニコ動画では実際にコメントにスコアを付ける機能が存在する．このスコアの信頼性や有用性については今後の研究課題とする．

6. おわりに

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に，ルールベース手法と機械学習手法を適用した．いずれの手法でも，テキストパターンと単語のバーストを用いており，両者を同時に用いることの有効性を示唆する実験結果が得られた．また，同一動画のコメントが訓練データに入っている場合は，分類結果が大幅に改善すること，また投稿時刻の情報が有用であることがわかった．

今後の課題としては，まず重要情報の認識が考えられる．本稿では重要情報の認識のために単語のバーストを用いたが，結果からわかるように十分ではない．そのために，動画の映像や音声から情報を抽出する，関連する Wikipedia からストーリーを捉えるなど，テキスト以外のリソースや外部リソースを利用していく必要がある．また，ユーザがコメントに付与するスコアを利用して擬似的な訓練データを作成する方法も有望である．さらに，例えば “犯人” や “黒幕” などのようにネタバレに使用されやすい単語などを事前に収集し，素性として利用するなどの方法も考えられる．また，現時点ではデータが小さいことが実験結果に影響を与えている可能性があり，データ量を増やして実験する必要もある．

謝辞

本研究は JSPS 科研費 25540080 の助成を受けたものです．

参考文献

[Boyd-Graber 13] Boyd-Graber, J., Glasgow, K., and Zajac, J. S.: Spoiler Alert: Machine Learning Approaches to Detect Social Media Posts with Revelatory Information, in *Proceedings of the 76th Annual Meeting of the American Society for Information Science and Technology (ASIST)* (2013)

[藤木ら 04] 藤木稔明, 南野朋之, 鈴木泰裕, 奥村学: document stream における burst の発見, 情報処理学会研究報告 SIGNL-160, pp. 85–92 (2004)

[Guo 10] Guo, S. and Ramakrishnan, N.: Finding the storyteller: automatic spoiler tagging using linguistic cues, in *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 412–420 (2010)

[岩井ら 13] 岩井秀成, 池田郁, 土方嘉徳, 西田正吾: レビュー文を対象としたあらすじ分類手法の提案, 電子情報通信学会論文誌 D, J96-D, no.5, pp. 1222–1234 (2013)

[Jeon 13] Jeon, S., Kim, S., and Yu, H.: Don’t Be Spoiled by Your Friends: Spoiler Detection in TV Program Tweets, in *Proceedings of the 7th International Conference on Weblogs and Social Media (ICWSM)*, pp. 681–684 (2013)

[中村ら 13] 中村聡史, 小松孝徳: スポーツの勝敗にまつわるネタバレ防止手法の検討, 情報処理学会論文誌, vol.54, no.4, pp. 1402–1412 (2013)