

字幕テキストの利用による マイクロブログからのテレビ番組に言及したメッセージ検出手法

山本祐輔[†] 及川孝徳[‡] 山名早人^{§,¶}

[†] 早稲田大学基幹理工学部 〒169-8555 東京都新宿区大久保 3-4-1

[‡] 早稲田大学大学院基幹理工学研究科 〒169-8555 東京都新宿区大久保 3-4-1

[§] 早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1

[¶] 国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: {y.yamamoto,t_oikawa,yamana}@yama.info.waseda.ac.jp

あらましマイクロブログのひとつである Twitter には、放送中のテレビ番組に関して、多くのユーザから感想や意見がリアルタイムに投稿されている。Twitter に投稿される番組に関する感想や意見は、他人の考えや番組内容、視聴者が番組において着目している事柄や登場人物の理解に有効であるが、一部のユーザのみの意見ではなく網羅的に意見を検出することが重要である。本論文では、Twitter からテレビ番組に言及している投稿を検出する手法を提案する。提案手法では、テレビ番組の字幕テキストから番組の特徴となるキーワードを抽出し、番組放送時間帯に Twitter に投稿されるメッセージに番組のキーワードが含まれているかをみることにより、テレビ番組に言及している投稿を検出する。実験の結果、提案手法は平均して適合率 82%を保ちながら 54%の再現率を達成した。さらに、Twitter 特有のハッシュタグの情報を提案手法と併用することで、84%の適合率と 68%の再現率を得ることができた。

キーワード: マイクロブログ, Twitter, テレビ

Detecting Tweets related to TV Programs by using Closed Caption Texts

Yusuke YAMAMOTO[†] Takanori OIKAWA[‡] Hayato YAMANA^{§,¶}

[†] School of Fundamental Science and Engineering, Waseda University 3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

[‡] Graduate School of Fundamental Science and Engineering, Waseda University 3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

[§] Science and Engineering, Waseda University 3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555, Japan

[¶] National Institute of Informatics 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, 101-8430, Japan

E-mail: {y.yamamoto,t_oikawa,yamana}@yama.info.waseda.ac.jp

1. はじめに

近年、Twitter をはじめとするマイクロブログが盛んに利用されている。マイクロブログとは、情報発信を行うことができるブログの性質とリアルタイムにコミュニケーションがとれるチャットの性質をあわせ持つサービスである。マイクロブログへの投稿は短文で良く、一投稿に必要な時間が短いため、マイクロブログには現在行っていることや考えていることがリアルタイムで投稿される傾向が強い。Twitter はマイクロブログサービスで最も有名なものの 1 つであり、140 文字以内でメッセージを投稿することができる。Twitter は 2006 年にアメリカで開始され、2008 年に日本語化されたサービスであり、2010 年 10 月の時点で日本から twitter.com へのアクセスだけでも月間 1700 万回にのぼるとされている [1]。Twitter に投稿されるメッセージ

は多岐にわたり、ユーザ自身に関連すること以外にも、ユーザが興味を持ったテレビ番組やニュースサイト等、他のメディアに関することや、企業の情報発信の場としても利用されている。

マイクロブログの発達に伴い、2007 年以降、Twitter を対象とした研究が多数行われている。Twitter 自体を対象とした研究 [2][3] では、ユーザの傾向や影響力のあるユーザの発見などについて研究がされてきた。また、Twitter にはユーザの身の回りに起こっていることが書き込まれる傾向があるため、Twitter と突然発生するイベントとを関連付けた研究もなされている [4]。さらに、Twitter には、テレビやニュースサイト等の影響力の大きなメディアから発信された情報に関するメッセージが多く投稿される。Twitter と他のメディアとして代表的なテレビ番組との関連性についても研究がされている [5][6][7][8]。

Twitterには、放送中のテレビ番組に関して、多くのユーザから感想や意見がリアルタイムに投稿されている。テレビ番組を視聴しながら意見や感想を書き込むという行為はマイクロブログから始まったわけではなく、以前からWebページにおいて行われている。例えば、掲示板サイトには、多くの視聴者が集まり意見の発信や交換が行われている。このような行為が行われる背景として、他人の意見を知る、番組の補足情報を提供しあうといった理由のほか、大勢でテレビ番組を見ているような一体感が得られることが挙げられる。これらのテレビ番組に言及している投稿は、他人の考えや番組内容、視聴者が番組において着目している事柄や人物の理解に有効であり、番組の感想や意見が書き込まれる掲示板サイトは、有用な情報を抽出する研究にも使用されている。

Twitterにおいても、テレビ番組の視聴者がリアルタイムに書き込んだ番組の感想や意見を検出する、小林らの研究[10]が行われている。小林らの研究[10]では、番組冒頭時にテレビ番組が始まったことを象徴する投稿が多くなることを利用して、番組開始直後に投稿を行ったユーザに番組を見ているか否かのラベルを付加する。そして、ラベルを付加したユーザが番組放送時間帯に投稿したメッセージ中の単語の出現頻度により、投稿の検出器を作成している。また、ツイテレ[11]というサービスでは、主要なテレビ局で放送されている番組についての投稿を見ることができる。しかし、小林らの手法では番組冒頭に検出対象とする番組について投稿を行うユーザが多く存在していなければ、番組に言及した投稿を検出することができない等の問題がある。また、ツイテレでは、テレビ局のハッシュタグを付けたユーザのメッセージしか検出できないという問題がある。ハッシュタグとは、特定のトピックについて投稿する際にメッセージに付加することができるタグ情報である。

そこで本論文では、テレビの字幕テキストを用いることで、テレビ番組について言及されたTwitterへの投稿を網羅的に検出する手法を提案する。投稿検出の対象とするユーザを限定しないことで、一部のユーザのみの感想や意見ではなく、より多くのユーザの網羅的な意見や感想を知ることができる。字幕テキストとは、テレビ番組の字幕放送において表示されるテキストを指す。字幕テキストはテレビ番組の出演者の発言等を文字情報として含んでいるため、番組に関連する語が含まれている。本手法では、字幕テキストからテレビ番組に関連のある語を抽出し、番組放送時間帯にTwitterに投稿されるメッセージに番組のキーワードが含まれているかをチェックすることにより、テレビ番組に関して言及した投稿の網羅的な検出を試みる。

本論文では以下の構成をとる。第2節で関連研究について述べる。第3節で提案手法についての説明をし、第4節で実験と評価を行い、第5節でまとめを述べる。

2. 関連研究

本節では、関連研究について述べる。以下、(1)テレビ番組についてリアルタイムに書き込まれるコメントを対象とした研究、(2)Twitterと他のメディアとの関連性についての研究、(3)Twitterからテレビ番組について言及している投稿を検出する研究の3つの研究について述べる。

(1)テレビ番組についてリアルタイムに書き込まれるコメントを対象とした研究として、宮森ら[9]は、インターネット掲示板にリアルタイムに書き込まれるテレビ番組に関するコメントを利用することにより、番組の盛り上がり場면을視聴者の視点で抽出し、番組のビューを生成する研究を行った。

(2)Twitterと他のメディアとの関連性についての研究について述べる。

Shammaら[5]は、テレビで放送されたアメリカ大統領選挙の候補者同士のディベートを題材に、ライブメディアイベントにおけるマイクロブログの使い方について研究を行った。

Akiokaら[6]は、日本におけるTwitterユーザの特徴ならびに、出版物やテレビ番組などのメディアがTwitterのコミュニティに与える影響について研究を行った。

澤井ら[7]は、Twitterに投稿されるメッセージを利用して、協調フィルタリングによりテレビ番組を推薦する研究を行った。

加藤ら[8]は、Twitterに投稿されるメッセージから出現頻度が急に高くなった固有名詞を抽出し、テレビ番組で放送される話題を追跡する研究を行った。

(3)Twitterからテレビ番組について言及している投稿を検出する研究について説明する。小林ら[10]は、投稿されるメッセージ中の単語の出現頻度により、テレビ番組に言及した投稿を検出する研究を行った。まず、番組冒頭にテレビ番組開始を象徴する投稿が行われることを利用し、番組冒頭に投稿を行ったユーザにテレビ番組を見ているか否かのラベルを付加する。具体的には、テレビ番組開始直後に投稿されるメッセージから出現頻度の高い単語により特徴ベクトルを作成する。この特徴ベクトルをSVMにより識別し、投稿を行ったユーザにテレビを見ているか否かのラベルを付加する。次に、ラベルを付加したユーザの、番組放送時間帯のその後の投稿から出現頻度の高い単語により特徴ベクトルを作成する。そして、この特徴ベクトルにより学習したSVMを検出器とし、番組放送時間

帯に投稿されるメッセージからテレビ番組に関連している投稿を検出する。

小林らの手法では、テレビ番組に関わらず汎用的なイベントについての投稿検出に適用可能であるが、番組冒頭で検出対象とする番組について投稿を行うユーザが多く存在していなければ番組に言及した投稿を検出することができない。また、投稿を検出する基準が番組冒頭に投稿を行ったユーザにより決定されてしまうため、番組冒頭に投稿を行ったユーザが着目しない事柄について投稿を行ったユーザのメッセージは検出することが難しい。

また、テレビ局のハッシュタグで番組に言及した投稿を検出する場合、ハッシュタグを使用する一部のユーザの投稿しか検出することができない。

本研究では、テレビ番組に関して言及した投稿の網羅的な検出を試みる。

3. 提案手法

本節では提案手法について述べる。まず提案手法の流れを図1に示す。

提案手法では、まず過去に放送されたテレビ番組の字幕テキストを形態素解析し、現れる名詞の idf 値を求めておく(図1中の①)。その後、感想や意見検出の対象となる番組を選択する。選択したテレビ番組の字幕テキストを形態素解析し、名詞を抽出した後、名詞の idf 値が閾値以下となる名詞をテレビ番組に関係のない一般名詞として取り除く。残った名詞をテレビ番組に関連する名詞として採用する(図1中の②右)。そして、番組放送時間帯に Twitter に投稿されるメッセージに、採用したテレビ番組に関連する名詞が含まれているかをみることにより、テレビ番組の感想や意見メッセージを検出する。また、テレビ番組に言及しているメッセージと判断されたメッセージを投稿したユーザはテレビ番組を見ているユーザとし、その後の番組放送時間帯に Twitter に投稿されるメッセージも番組に関するものである可能性があるため検出する(図1中の②左)。

提案手法の詳細について 3.1 節でテレビ番組の特徴語抽出について述べ、3.2 節で字幕テキストに現れる名詞の idf 値算出について説明する。3.3 節でテレビ番組に言及したメッセージを投稿したユーザの番組終了までの投稿追跡について述べる。

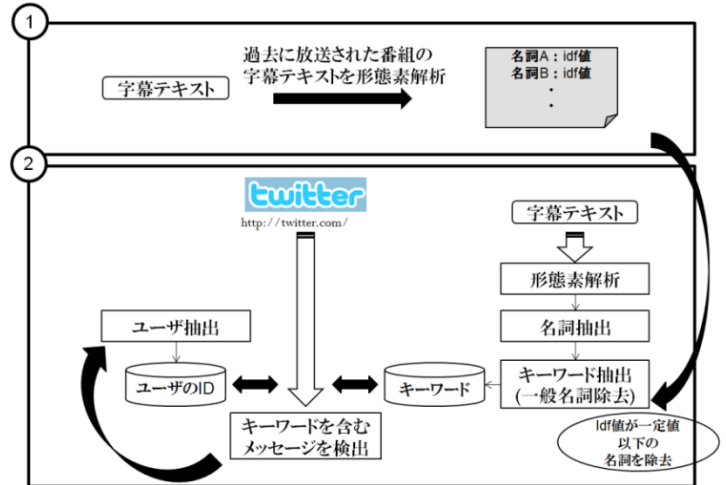


図1 提案手法の流れ

3.1. テレビ番組の特徴語抽出

テレビ番組の特徴となる名詞を、そのテレビ番組の字幕テキスト中から取得する手順について述べる。

テレビ番組の特徴語として名詞を採用する理由は、テレビ番組の感想や意見として有用なものが、「すごい！」など単なる感情表現ではなく、テレビ番組に登場する人物や事柄についてどのようにとらえたかを表しているメッセージだからである。

次に、テレビ番組の特徴となる名詞の抽出手法について説明する。番組に言及したメッセージを検出する対象としたテレビ番組の字幕テキストを形態素解析し、名詞を抽出する。そして、抽出した名詞のうち 3.2 節で説明する idf 値が閾値以下の名詞を、番組に関連しない一般的な名詞として削除する。そして、残った名詞をテレビ番組の特徴となる名詞として採用する。

3.2. 字幕テキストに現れる名詞の idf 値算出

名詞の idf 値は 3.1 節で述べたようにテレビ番組の特徴となる名詞を抽出する際に、一般名詞を除去するために使用する。

テレビ番組の字幕テキストに現れる名詞の idf 値算出方法について説明する。字幕テキストが存在するテレビ番組の総数を P 、字幕テキストに名詞 $term_i$ が現れるテレビ番組の数を p_i としたとき、名詞 $term_i$ の idf 値 idf_i を以下の式(1)で定義する。

$$idf_i = \log \frac{P}{p_i} \quad (1)$$

3.3. テレビ番組に言及したメッセージを投稿したユーザの番組終了までの投稿追跡

テレビ番組に言及しているメッセージとして 3.2 節

で検出された名詞を含んだメッセージの投稿者はテレビ番組を見ているユーザである。これらのユーザはその後テレビ番組の感想や意見を投稿する可能性が高い。そこで、3.2 節で検出された名詞を含むメッセージの投稿者が、番組終了までに Twitter へ投稿する全てのメッセージを、テレビ番組に言及しているメッセージとして採用する。

4. 評価実験

本節では、提案したテレビ番組に言及しているメッセージを検出する手法の精度および再現率の評価実験と結果について説明する。4.1 節で使用データについて説明し、4.2 節で評価方法を説明する。4.3 節で一般名詞除去における idf 値の閾値決定について説明し、4.4 節で評価結果について述べる。

4.1. 使用データ

実験に用いる字幕テキストは、2005 年 1 月から 2010 年 11 月までの、関東で放送された 7 局のテレビ番組における字幕テキストである。Twitter のデータは 2009 年 12 月から 2010 年 12 月まで Twitter API により収集したデータを使用した。実験に使用したデータを表 1 にまとめる。

まず、2005 年 1 月から 2010 年 11 月までの字幕テキストに現れる名詞の idf 値を算出した。この期間中に字幕テキストが存在するテレビ番組の総数は 226,426 番組であった。

そして、番組に言及しているメッセージを抽出する対象とするテレビ番組を選択し、そのテレビ番組の字幕テキストと番組放送時間帯に Twitter に投稿されたメッセージのデータを使用して、4.2 節で説明する実験を行った。

表 1 実験に使用したデータ

	収集期間	データ量
字幕テキスト	2005 年 1 月から 2010 年 11 月	226,426 番組
Twitter のデータ	2009 年 12 月から 2010 年 12 月	208 日分

4.2. 評価方法

実験対象として、ドラマ、ドキュメンタリー、バラエティー、アニメのジャンルにおいて各々 1 番組を用いた。また、各番組の放送時間帯に API で取得できた Twitter に投稿されたメッセージから、あらかじめ番組に言及しているメッセージを手手により抽出し、正解セットを作成した。なお正解セットは、投稿されたメッセージのみに着目し、(1)テレビ局およびテレビ番組のハッシュタグが付けられている投稿、(2)番組の内容

に即している投稿、(3)出演者やキャスト等の番組に登場する人物についての投稿の 3 つの基準により作成した。そして、実験対象とした番組の字幕テキスト、番組時間帯に Twitter に投稿されたメッセージ、および作成した正解セットを用いて、以下の 3 つの手法により投稿検出の適合率と再現率を測定した。

1. テレビ局とテレビ番組のハッシュタグが付加されているメッセージを検出する手法
2. 提案手法
3. 提案手法およびテレビ局とテレビ番組のハッシュタグが付加されているメッセージを検出する手法

実験で使用した番組と Twitter のデータを表 2 に示す。

表 2 実験に使用した番組と Twitter のデータ

	番組 A (ドラマ)	番組 B (ドキュメンタリー)	番組 C (バラエティー)	番組 D (アニメ)
番組の放送時間	45 分間	30 分間	30 分間	30 分間
番組放送時間帯に API で取得した Twitter に投稿されたメッセージ数	8,972	4,423	6,214	3,888
番組に言及しているメッセージ数	389	97	135	120

4.3. 一般名詞除去における idf 値の閾値決定

本節では、一般名詞として除去する名詞の idf 値の閾値決定のための事前実験について説明する。一般的に閾値を低く設定し過ぎると一般的な名詞を除去することができず、番組に言及していない誤ったメッセージを検出する可能性が高まる。また、閾値を高く設定し過ぎると番組の特徴となるべき名詞をキーワードにできないことがあり、番組に言及しているメッセージの網羅的な検出ができなくなる。

算出した名詞の idf 値のヒストグラムを図 2 に示す。図 2 より、字幕テキスト中の名詞の idf 値の分布は、idf 値が 10 以上の範囲に集中していることがわかる。これは、字幕テキストに頻出する名詞は多くないということと、字幕テキストに頻りに現れない、番組の特徴的な語となる単語は多くあるということを表している。

一般名詞除去における idf 値の閾値を決定するため

に使用する番組を示す。各番組の放送時間帯に API で取得できた Twitter に投稿されたメッセージから、テレビ局および番組のハッシュタグが付けられているメッセージを正解セットとした。

表 3 idf 値の閾値決定に使用した番組と Twitter のデータ

	番組 1	番組 2	番組 3	番組 4
番組の放送時間	45 分	49 分	45 分	60 分
番組放送時間帯に API で取得した Twitter に投稿されたメッセージ数	10,842	9,099	10,409	7,519
テレビ局および番組のハッシュタグが付けられているメッセージ数	121	173	122	52

一般名詞除去における idf 値の閾値の決定方法について述べる。以下、一般名詞除去における idf 値の閾値を $Threshold_{idf}$ と定義する。まず、 $Threshold_{idf}$ を変化させながら、表 3 に示した 4 番組について、提案手法による番組に言及したメッセージ検出の適合率と再現率を求める。適合率は検出したメッセージのうち、正解セットに含まれているメッセージの割合とする。また、再現率を正解セットのうち、提案手法により検出できたメッセージの割合とする。そして、各番組について式(2)で定義される F -measure を求める。そして、4 番組の F -measure の平均が最大となる idf 値を、最終的な $Threshold_{idf}$ とする。

$$F - measure = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (2)$$

まず、 $Threshold_{idf}$ を 0 から 12 まで 1 刻みで変化させ、4 番組の F -measure の平均が最大となる idf 値の範囲を限定する。その後、制限した範囲内で $Threshold_{idf}$ を 0.1 刻みで変化させ 4 番組の F -measure の平均が最大となる idf 値を求めることにする。 $Threshold_{idf}$ を 0 から 12 まで 1 刻みで変化させたときの 4 番組の F -measure の平均を図 3 に示す。図 3 より、4 番組の F -measure の平均が最大となる idf 値の範囲は 5.0 から 7.0 の間に絞り込まれる。次に、5.0 から 7.0 の範囲で $Threshold_{idf}$ を 0.1 刻みで変化させたときの 4 番組の F -measure の平均を図 4 に示す。図 4 より、4 番組の F -measure の平均が最大となる idf 値の閾値は 6.0 となることがわかる。よって、最終的に採用する $Threshold_{idf}$ を 6.0 と定める。

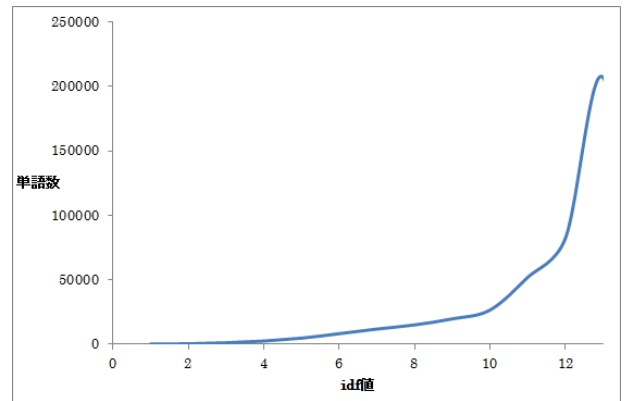


図 2 idf 値のヒストグラム

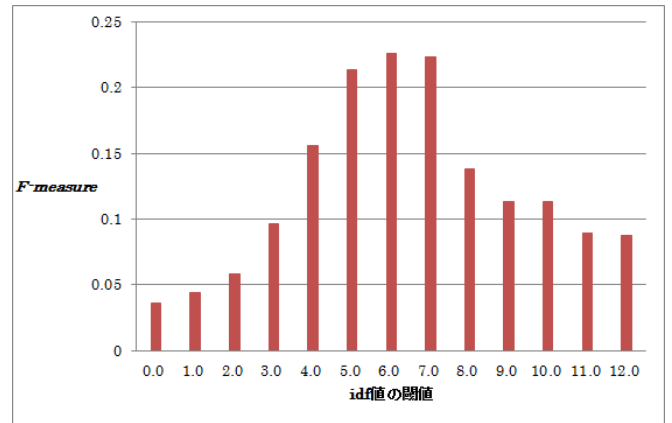


図 3 4 番組の F -measure の平均値

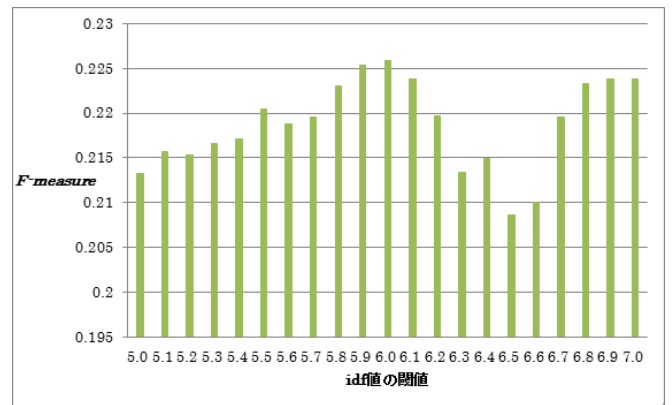


図 4 4 番組の F -measure の平均値

4.4. 評価結果

各手法により検出したメッセージのうち、正解セットに含まれているメッセージの割合を適合率とする。また、正解セットのうち、各手法により検出できたメッセージの割合を再現率とする。評価実験の結果得られた適合率を図 5 に、再現率を図 6 に、F 値をに示す。実験の結果、提案手法では平均で 82% の適合率を保ち、

テレビ局とテレビ番組に関連するハッシュタグで検出するより最低でも 1.3 倍の再現率を達成できた。また、提案手法とテレビ局およびテレビ番組に関連するハッシュタグを組み合わせることにより、ハッシュタグのみで検出する手法に比べ最低でも 1.9 倍の再現率を達成できている。よって、字幕テキストを利用することで、Twitter に投稿されたメッセージからテレビ番組に言及したメッセージを検出することにおいて、再現率の向上に有効であるとわかる。

提案手法により検出できたテレビ番組に言及したメッセージの例を示す。例えば、ドキュメンタリイでは、スマートフォンについて述べられており、以下のようなメッセージが検出された。

- 日本のメーカーはわかってないな。ワンセグや高画質カメラより、いかに使える。楽しいアプリがあるかだよ。
- iPhone って言うてからスマートフォンと言い換えるあたりが NHK っぽい。
- 親父がスマートフォン特集のテレビ見てるんだがまず携帯電話の仕組み事態理解できない人に理解するのは無理だと思います。

また、ドラマでは大河ドラマで実験を行い、次のような投稿が検出できた。

- 容堂いい！「こやつは上士じゃない」武市（南朋）の行く末、涙さそう。
- 松陰さん捕まったがじゃ。無謀と言われても自分を貫く姿勢、かっこいいぞ。
- 熱いなあ。吉田松陰

提案手法では、番組に言及している投稿のうち、検出例で示したような番組のキーワードとなるような名詞を含む投稿は検出できている。

しかし、提案手法では『しかし周りが良ければ良いほど福山の存在感が薄れていく畏』のようなドラマのキャスト自体について言及している投稿は番組の字幕テキストに現れないため検出することができない。提案手法で検出できない投稿は以下の 4 つに分類することができる。

1. ドラマのキャスト等、字幕テキストに現れない情報についての投稿
2. 番組のキーワードが含まれていない投稿
3. 番組のキーワードが誤字になっている投稿
4. 番組に言及している投稿であるが、含まれている名詞の idf 値が閾値以下である投稿

また、番組を録画する旨を表しているメッセージ等を誤検出してしまう。提案手法で誤検出してしまう投稿は以下の 3 つに分類することができる。

1. 番組を録画しているまたは裏番組を見ていることを示す投稿

2. 番組のキーワードを含んでいるが、番組に言及しているわけではない投稿
3. 提案手法により番組を見ているとして追跡しているユーザが、番組に言及している投稿以外の内容を投稿している

このような投稿は字幕テキストから得られる情報のみでは検出することが難しく、異なる手法が必要である。

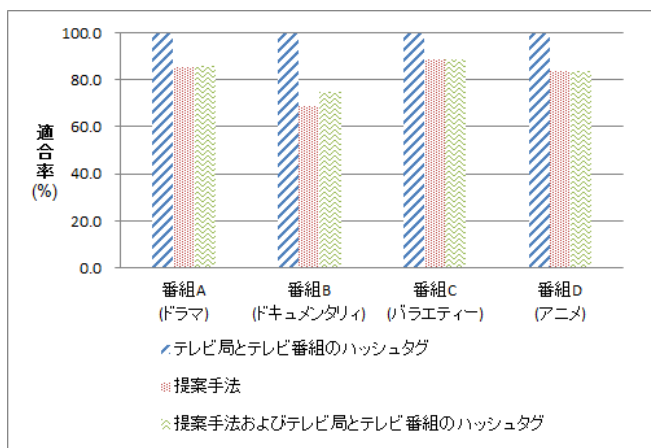


図 5 実験結果における適合率

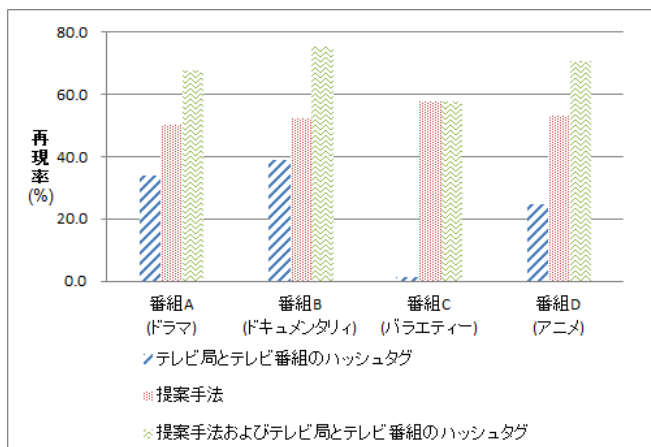


図 6 実験結果における再現率

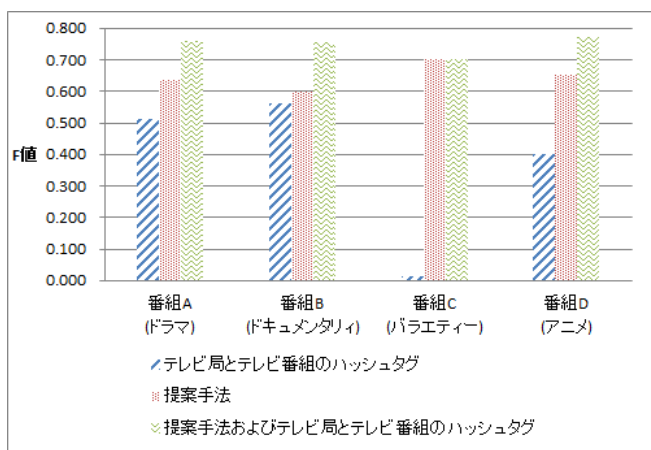


図 7 実験結果における F 値

5. まとめ

本稿では、字幕テキストの利用によるマイクロブログからのテレビ番組に言及した投稿の検出手法を提案した。実験の結果、提案手法は平均で適合率 82%を保ちながら平均して 54%の再現率を達成し、また、提案手法とテレビ局およびテレビ番組に関連したハッシュタグを組み合わせるにより平均して 68%の再現率を達成した。よって、字幕テキストはテレビ番組に言及したメッセージ検出の再現率向上に有効であるとわかった。

今後の課題として、字幕テキストに現れないがテレビ番組の重要な情報に言及している投稿等を検出し、再現率の向上を図ることである。また、提案手法を利用した、テレビ番組の Twitter における視聴率の概算等の影響力の調査を考えている。

謝辞

本研究の一部は、文部科学省・次世代 IT 基盤構築のための研究開発「多メディア Web 解析基盤の構築及び社会分析ソフトウェアの開発」の補助によるものである。なお、解析用データ提供において国立情報学研究所の佐藤真一様にお世話になりました。ここに感謝いたします。

参 考 文 献

- [1] 日本における Twitter.com の月間利用者数 (2010.12.5 アクセス)
https://www.google.com/adplanner/?pli=1#siteSearch?identifier=twitter.com&geo=JP&trait_type=1&lp=false.
- [2] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng: “Why We Twitter: Understanding Microblogging Usage and Communities”, Proc.of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp.56-65, 2007.
- [3] JianshuWeng, Ee-Peng Lim, Jing Jiang, and Qi He: “TwitterRank: finding topic-sensitive influential twitterers”, Proc.of the 3rd ACM international conference on Web search and data mining, pp.261-270, 2010.
- [4] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo: “Earthquake shakes Twitter users: real-time event detection by social sensors”, Proc.of the 19th international conference on World Wide Web , pp.851-860, 2010.
- [5] David A. Shamma, Lyndon Kennedy, and Elizabeth F. Churchill: “Tweet the Debates: Understanding Community Annotation of Uncollected Sources”, Proc.of the 1st SIGMM workshop on Social media, pp.3-10, 2009.
- [6] SayakaAkioka, Norikazu Kato, Yoichi Muraoka, and HayatoYamana: “Cross-media Impact on Twitter in Japan”, Proc.of the 2nd international workshop on Search and mininguser-generated contents, pp.111-118, 2010.
- [7] 澤井 里枝, 有安 香子, 藤沢 寛, 金次 保明: “SNS を利用した協調フィルタリングによる番組推薦手法”, 情報処理学会研究報告, Vol.2010-DBS-151 No.43, 2010.
- [8] 加藤 慶一, 秋岡 明香, 村岡 洋一, 山名 早人: “ミニブログにおける注目語抽出手法の提案と注目語を用いたメディア間での話題追跡”, 情報処理学会研究報告, Vol.2010-DBS-151 No.22, 2010.
- [9] 宮森 恒, 中村 聡史, 田中 克己: “番組実況チャットに基づく視聴者視点を利用した放送番組のビュー生成”, 電子情報通信学会第 16 回データ工学ワークショップ, 4B-i9, 2005.
- [10] 小林 尊志, 野田 雅文, 出口 大輔, 高橋 友和, 井手 一郎, 村瀬 洋: “Twitter における実況書き込み検出手法の検討”, 映像情報メディア学会技術報告, Vol. 34, No.25, pp.129-130, 2010.
- [11] ツイテレ <http://twtv.jp/>.