

# テレビ番組をより楽しむための実時間ツイート選択システム

廣田 壮一郎<sup>†</sup>      笹野 遼平<sup>‡</sup>      高村 大也<sup>‡</sup>      奥村 学<sup>‡</sup>

<sup>†</sup>東京工業大学 総合理工学研究所, <sup>‡</sup>東京工業大学 精密工学研究所

hirota@lr.pi.titech.ac.jp, {sasano, takamura, oku}@pi.titech.ac.jp

## 1 はじめに

ソーシャルメディアの普及に伴い、テレビ番組の感想や意見をソーシャルメディアへとリアルタイムに投稿するユーザーが増えてきている。それと同時に、テレビ局側も番組制作において、関連する投稿を番組内で紹介し、それらの感想や意見を番組に反映させるといった取り組みを始めている。例えばNHKの「NEWS WEB」<sup>1</sup>では、代表的なソーシャルメディアの一つであるTwitter<sup>2</sup>に投稿された番組に関する投稿を募集し、視聴者の感想や意見を討論の進行に反映させている。しかし、「NEWS WEB」では放送時間の30分間に約5,000、投稿の多い時間帯には10秒間に100近い投稿が行われており、放送時間中に全てのツイートを確認することは大変手間のかかる作業である。

テレビ番組の内容に合わせて面白い投稿を自動的に選ぶようなシステムがあれば、インターネットとテレビ番組が組み合わせられた新しい視聴方法を提供する放送局が増えると考えられる。また、放送局にかぎらず、ユーザーもこのシステムを使うことで気軽にインターネットとテレビ番組を同時に楽しめるようになると考えられる。そこで、本研究では、テレビ番組と同時に見ることでより楽しめるようなツイートを自動選択する手法について述べる。

## 2 関連研究

大量にあるツイートの要約を作成する手法としてSharifiら[1]のものがある。Sharifiらはツイート集合に対してクエリを与え、クエリと一緒に咬かれている単語列パターンの頻度を計算した。クエリと連結している単語列の内、2回以上出現している単語列パターンを抽出することで要約を行った。

時間的な変化に着目したマイクロブログの要約を行った研究として、Takamuraら[2]のものがある。Taka-

muraらは単一トピックに関する投稿の中から他の投稿との被覆度が大きい投稿を選択し、時系列順に並べることによってトピックの要約を行っている。

また、イベントに対してリアルタイムに要約を行う手法としてNicholsら[3]やKuboら[4]の手法がある。イベントが起こればバーストと呼ばれる投稿数の増加が起きることに注目し、イベントの発生区間を特定する。各イベントごとの要約を行うことにより、トピックの終了を待つことなく要約を行うことが可能となる。

しかし、これらの研究ではツイート集合の要約を主目的としており、テレビ番組の内容に沿って人が面白いと感じるようなツイートを本研究とは、研究目的が異なっている。

人が面白いと感じるツイートを本研究の手法としてはStajnerら[5]の手法がある。Stajnerらは式(1)を用いて、 $g(S)$ を最大化するようなツイート集合 $S$ を選ぶことで人が面白いと感じるツイートの選択を行った:

$$g(S) = \lambda \sum_{m \in S} r(m) + (1 - \lambda) H_0(S). \quad (1)$$

ここで、 $S$ は選択したツイート集合であり、 $r(m)$ はツイート $m$ の面白さの度合いを計算している。 $r(m)$ の計算にはSupport Vector Regressionを使用しており、素性としては単語数やツイート数などを利用している。ここで、 $H_0(S)$ は選択したツイート集合の単語出現確率分布のエントロピーであり、選択ツイートの情報量が大きくなるようにしている。 $\lambda$ はツイートの面白さの度合いと、選択したツイート集合の情報量のどちらかを重視するかを決めるパラメータである。

この研究は、面白いツイートを静的にツイート集合の中から選択する手法を提案するものであるが、対して本研究では、テレビ番組の内容に合わせて人が面白いと感じるような投稿をリアルタイムに選択することを目的とする。その際、ツイートの面白さの学習と提示するツイートの学習を分けて行うことで、より良いツイートの選択を可能とする。

<sup>1</sup><http://www3.nhk.or.jp/news/newsweb/>

<sup>2</sup><https://twitter.com/>

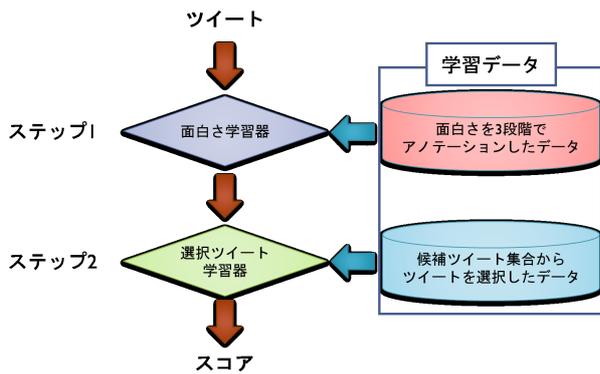


図 1: ツイートのスコア計算の概要

### 3 提案手法

本研究では番組放送中に呟かれたツイート系列を 10 秒毎に分割し、10 秒間に投稿されたツイート集合の中から 1 つを選択する。ツイートを選擇するにあたって大きく 2 つの計算を行う。

**ステップ 1:** 各ツイートの面白さ度合いを計算

**ステップ 2:** ステップ 1 のスコアを素性として使用し、既に選擇したツイートの内容を考慮して提示するツイートを選択

ツイートのスコア計算の概要を図 1 に示す。ステップ 1 とステップ 2 の計算の際には、別々のアノテーションデータを用意し学習を行う。学習には Support Vector Machine(SVM) を使用している。ここで、本研究ではテレビ番組の放送に合わせてリアルタイムにツイートを選擇することを目的としているため、ツイートを選擇する際には選擇時点よりも未来のツイートや番組の内容を知ることができない。

#### 3.1 ステップ 1: ツイートの面白さ度合いの計算

ツイートの面白さ度合いの学習を行うにあたり、表 1 に示す素性を使用する。

表 1 のうち、過去のツイート集合から作成される素性とは、過去 1 分間に投稿されたツイートを 10 秒毎に纏め、それぞれのツイート集合に対して計算されるスコアを素性に加えている。これにより投稿時間に関する影響を考慮することができ、Twitter における盛り上がりに対する被覆度合いなどを知ることができる。これは、盛り上がっている話題に関する情報を多く含むツイートは、面白いツイートとして選ばれる傾向があると考えたため使用している。また、テレビ番組の

表 1: 対象ツイートの面白さの学習に使用する素性  
ツイート単体から作成される素性 (Target)

単語数、感嘆符の数、疑問文か否か、URL の有無、一人称代名詞の有無、砕けた表現を含んでいるか、地名を含んでいるか、機能語の Bag-of-Words

候補ツイート集合から作成される素性 (Candidates)

ツイートに出現する単語の idf<sup>3</sup> の総和、余弦類似度が閾値を越えたツイート数

過去のツイート集合から作成される素性 (Past)

ツイートに出現する単語の idf<sup>4</sup> の総和、対象ツイートの対数尤度<sup>5</sup>、余弦類似度が閾値を越えたツイート数、隣接 10 秒の余弦類似度の差

テレビ番組の音声認識結果から作成される素性 (TV)

ツイートの単語のうち番組に出現した単語の数

音声認識結果から作成される素性は、テレビ番組の内容にどれくらい関係しているツイートかを知るために有効であると考えられるため使用している。

#### 3.2 ステップ 2: ツリート選擇のためのスコア計算

提示するツイートを決めるにあたり、表 2 に示す素性を使用する。

表 2: 提示するツイートの学習に使用する素性

ステップ 1 で計算した面白さ度合い (面白さ)

SVM での判別結果から得られる超平面からの距離

過去に選擇したツイートとの情報量関係 (情報量)

過去の選擇ツイート集合の単語出現確率分布の情報量、選擇ツイート 1 つ 1 つに対する余弦類似度

過去に選擇したツイートとの意味関係 (意味)

疑問が連続しているか、ポジネガの変化、テレビの内容に対する言及度合い

ステップ 1 で計算したツイートの面白さ度合いを素性に加えることにより、ツイートの面白さに影響を及ぼす要因を効率的に利用することができる。過去に選擇したツイートとの情報量関係と過去に選擇したツイートとの意味関係の素性は、自ツイートと過去 1 分間までに選擇されたツイート 1 つ 1 つの関係について計算を行っている。過去に選擇したツイートとの情報量関係を使用する理由は、如何に面白いツイートであっても既に選ばれたツイートと同じ内容のツイートは選

<sup>2</sup>候補ツイート集合から計算される。

<sup>3</sup>過去ツイート集合から計算される。

<sup>4</sup>対象の 10 秒間のツイート集合から作成した単語 1-gram を用いて計算した。

択の優先度が低いと考えられるためである。また、過去に選択したツイートとの意味関係は、既に選んだツイートとは異なる視点から投稿されたツイートが選択されやすい傾向があることを考慮するために使用する。

## 4 実験

### 4.1 アノテーションデータの作成

実際のテレビ番組の動画の録画と放送時間帯にそのテレビ番組を対象に眩かれたツイートの収集を行った。本研究では、20日分の「NEWS WEB」の番組と、公式ハッシュタグである「#nhk24」が付けられたツイートで放送時間中に投稿されたものを収集した。番組の内容を知る手がかりとするため、AmiVoice<sup>6</sup>という音声認識ソフトを利用し、出演者の発言内容の文字起こしを行った。

さらに、収集したツイートに対して、以下の2種類のアノテーションを行った。

#### 4.1.1 面白さ度合いのアノテーション

ツイートの面白さ度合いを学習するためのデータを作成した。アノテーションは3人のアノテータ (user1, user2, user3) によって行い、実験に使用するツイートに対してどのくらい面白いツイートかを3段階でタグ付けをした。具体的には面白いツイートを3、普通のツイートを2、面白くないツイートを1とタグ付けした。作成したアノテーションデータに対し、2人以上が面白いと判断したツイートを正例、2人以上が面白くないと判断したツイートを負例として学習に使用し、実験の際には5分割交差検定を用いて評価を行った。

#### 4.1.2 選択するツイートのアノテーション

提示するツイートを学習するためのデータを作成した。収集したツイートの投稿時間に従って10秒毎に纏め、10秒間に投稿されたツイートの中から1つのツイートを選択した。ここで提示するツイートのアノテーションには、過去に選択したツイートの内容を考慮し、テレビ番組の動画を視聴しながら選択している。アノテーションは4.1.1項と同様に3人のアノテータによって行った。収集した40,215ツイートから、1人あたり1,934ツイートを提示するツイートとしてタグ付けし、このデータを3人分用意した。実験の際には

20番組を4番組からなる5グループに分割し、5分割交差検定を用いて評価を行った。

### 4.2 ツイートの面白さ度合い計算の評価

実験を行うにあたって4.1.1項で作成したデータに対し、分類問題として精度、再現率、F値を比較することで各素性の評価を行った。ツイート単体から作成される素性 (Target)、候補となるツイート集合から作成される素性 (Candidates) を基本素性とし、過去のツイート集合から作成される素性 (Past)、テレビ番組の音声認識結果から作成される素性 (TV) を提案素性として比較実験を行った。

表 3: ツイートの面白さ度合い計算に使う素性の評価

	精度	再現率	F 値
Target	0.851	0.738	0.791
+Candidates	0.850	0.748	0.796
+Past	0.852	0.760	0.803
+TV	0.840	0.762	0.799
+Past+TV	0.849	0.770	0.807

結果は表3のとおりになり、提案素性を加えることで分類性能が向上することを確認した。ここで表3において、各モデルとその他のモデルのF値の差は有意<sup>7</sup>であることを確認した。精度に対して再現率が大きく向上していることから、素性を加えることで今まで取れていなかった面白いツイートを取ることができるようになったことがわかる。

### 4.3 提示ツイート選択の評価

#### 4.3.1 素性の評価

3.2節で述べた素性の評価を行う。4.1.2項で作成したアノテーションデータに対し、アノテータ毎に学習と評価を行った。これはツイートの面白さと比べ、選ばれるツイートのアノテータ毎の差異が大きいためである。評価にはシステムが選択したツイートの順位を用いた平均逆順位 (MRR) を使用した。本システムの評価実験の指標とした MRR は以下の式 (2) で計算される：

$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{rank(i)}. \quad (2)$$

<sup>6</sup><http://sp.advanced-media.co.jp/>

<sup>7</sup>Randomization Test(有意水準 5%)

MRR において、 $N$  は提示するツイートの計算を行った回数を表している。また  $rank(i)$  は、 $i$  番目のツイート選択時における 10 秒間に投稿されたツイートのスコアのうち、アノテータが選んだツイートの順位である。

表 4: 素性の評価 (MRR)

	user1	user2	user3	平均
ALL	0.340	0.326	0.341	0.336
-面白さ	0.329	0.310	0.329	0.323
-情報量	0.333	0.312	0.329	0.329
-意味	0.325	0.306	0.320	0.317

表 4 は素性の組み合わせを考慮して比較実験を行った結果である。3.2 節で述べた 3 種類の素性を全て足した素性 (ALL) に対して、ステップ 1 で計算した面白さ度合いの素性 (面白さ)、過去に選択したツイートとの情報量関係の素性 (情報量)、過去に選択したツイートとの意味関係の素性 (意味) のそれぞれを引いたときの MRR を比較することで評価を行っている。中でもツイートとの意味関係の素性を引いた際に大きく MRR の値が下がっている。これはアノテータが選んだツイートに、疑問を呈するツイートの次には疑問以外のツイートを選択したり、ポジティブなツイートの次にはネガティブなツイートを選択するなどの傾向があることから分かる。

#### 4.3.2 モデルの評価

2 つのステップで提示ツイートを選択する提案モデルの評価を行う。評価には 4.3.1 項と同様に MRR を用いた。比較モデルとして、1 ステップでツイートを選擇する 2 つのモデルを用意した。

比較モデル 1 は、ステップ 1 で用いた素性を利用し、4.1.1 項で作成したアノテーションデータから学習を行う。比較モデル 1 は面白いとアノテーションされたツイートを正例、面白くないとアノテーションされたツイートを負例とし、SVM で学習している。比較モデル 2 は、ステップ 2 で用いた素性からステップ 1 で計算した面白さ度合いの素性を除き、代わりにステップ 1 の素性を加えて学習を行っている。比較モデル 2 は 4.1.2 項で選擇されたツイートを正例、選擇されなかったツイートを負例として SVM で学習している。

結果は表 5 のとおりになり、本研究において提案モデルが最も高い値となることが分かった。ここで表 5 において、提案モデルと比較モデル 1 に対しては user2

表 5: モデルの評価 (MRR)

	user1	user2	user3	平均
提案モデル	0.340	0.326	0.341	0.336
比較モデル 1	0.319	0.314	0.307	0.314
比較モデル 2	0.329	0.310	0.330	0.323

を除いて、提案モデルと比較モデル 2 に対しては各ユーザにて、MRR の差は有意<sup>8</sup>であることを確認した。これは一度の学習でツイートを選択するのではなく、ツイートの面白さの学習と提示するツイートの学習を別々に行うことで、ツイートの面白さに関する情報を効率的に利用しつつツイートを選擇することができたからだと考えられる。

## 5 おわりに

本稿ではテレビ番組放送中に投稿されるツイートに対し、番組をより楽しめるようなツイートを自動的に選擇する手法を提案した。提案手法では、放送内容に合わせてツイートの選擇を行うため、テレビ番組出演者の発言を音声認識した結果を利用した。さらに、ツイートが投稿された時間も考慮することでリアルタイムなツイート選擇を実現している。また、ツイートの面白さの学習と提示するツイートの学習を別々に行うことで、テレビ番組をより楽しめるようなツイートを高い精度で選擇することができた。今後の課題として、より豊富なテレビ番組の内容と選擇されるツイートの関係を捉えることのできる素性の検討が考えられる。

## 参考文献

- [1] Beaux Sharifi, Mark-Anthony Hutton, and Jugal K. Kalita. Experiments in microblog summarization. In *Proc. of SocialCom'10*, pp. 49–56. 2010.
- [2] Hiroya Takamura, Hikaru Yokono, and Manabu Okumura. Summarizing a document stream. In *Proc. of ECIR'11*, pp. 177–188. 2011.
- [3] Jeffrey Nichols, Jalal Mahmud, and Clemens Drews. Summarizing sporting events using twitter. In *Proc. of IUI'12*, pp. 189–198. 2012.
- [4] Mitsumasa Kubo, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura. Generating live sports updates from twitter by finding good reporters. In *Proc. of WI'13*, pp. 527–534, 2013.
- [5] Tadej Štajner, Bart Thomee, Ana-Maria Popescu, Marco Pennacchiotti, and Alejandro Jaimes. Automatic selection of social media responses to news. In *Proc. of SIGKDD'13*, pp. 50–58, 2013.

<sup>8</sup>Randomization Test(有意水準 5%).