

## ユーザの反応を利用したネタツイート自動分類手法

林田 宗一郎<sup>†</sup> 牛尼 剛聡<sup>‡</sup>

<sup>†</sup>九州大学芸術工学部 〒815-8540 福岡市南区塩原 4-9-1

<sup>‡</sup>九州大学大学院芸術工学研究院 〒815-8540 福岡県福岡市南区塩原 4-9-1

E-mail: <sup>†</sup>1DS10185K@kyushu-u.ac.jp <sup>‡</sup>ushiama@desighn.kyushu-u.ac.jp

**あらまし** 近年, SNS は爆発的に普及し, 情報源として重要な役割を担っている. SNS には様々な種類が存在するが, 代表的な SNS の一つに Twitter がある. 一般的に, Twitter のユーザは, 日常的に膨大な数の記事(ツイート)を閲覧している. そして, 閲覧したツイートの中には, ユーザにとって価値があり, 後から見直したいツイートが存在する場合がある. それらを管理するために, ユーザはツイートを「お気に入り」に登録することが多い. 一般に, ユーザは, 「お気に入り」に登録されたツイートをフォルダに分類したり, タグを付与したりすることはなく, 全てのツイートを「お気に入り」用のフォルダに一括して格納するだけである. そうしたことから, ユーザがお気に入りに入れたツイートを後で読み返そうとしても, 自分が読み返したいツイートを見つけることが困難な場合がある. この問題を解決するために, ツイートを自動的に分類することが期待されている. 分類には, 様々な基準が考えられるが, ユーザにとってのツイートの価値や目的に基づいた分類が重要である. しかし, そのような分類においては, ツイート本文に出現する単語等の言語的特徴に一貫性がなく, 従来一般的なテキスト分類法での分類は困難である. 本論文では, そのような分類の代表例として, 読み手を笑わせることを目的とした「ネタツイート」を例題として, ツイートに対するユーザ反応を利用することでネタツイートを自動的に分類する手法を提案する. 具体的には「引用ツイート」, 「ツイートへのリプライ(コメント)」, 「リツイート直後のツイート」によるユーザの反応を利用して, ネタツイートの自動分類を行う.

**キーワード** Twitter, ユーザ反応, テキスト分類

### 1. はじめに

現在, SNS(ソーシャルネットワークサービス)は多くのユーザに利用されている. Twitter は代表的な SNS の一つである. Twitter では, 興味を持つ投稿する可能性があるユーザをフォローすることにより, それらのユーザが投稿した記事(ツイート)が自分のタイムラインに配信される. 一般に, タイムラインには様々な種類のツイートが配信され, それらの中には後で読み返したいと思うツイートが含まれる場合がある. そうしたツイートを保存するために, 「お気に入り」が利用されることが多い.

「お気に入り」は Twitter が提供する機能の一つであり, ユーザは後で読み返しや利用を行う可能性があるツイートを「お気に入り」に保存できる. 「お気に入り」に保存したツイートは後から確認できる. しかし, 「お気に入り」に登録したツイートを, ユーザやハッシュタグ別に分類する機能は提供されていないため, お気に入りに登録したツイートの数が増えると, 読み返したいツイートの探索に時間を要する. そうしたことから, お気に入りを効果的に管理することを目的としたウェブサービスが提案されている. そうしたウェブサービスの代表例である「ふぁぼろぐ」[1]等を利用することで, 「お気に入り」に保存したツイートを投稿日時や宛先別に表示可能である. また, キーワード検索も可能である. しかし, 「ふぁぼろぐ」を利用したとしても, 読み返したいと思うツイートの投稿日時や

キーワードが分からなければ, 目的のツイートの発見は困難である. そうしたことから, 現在の「お気に入り」は, ユーザが, 読み返したいツイートを好きな時に読むために利用するにも関わらず, 目的のツイートを発見するためのユーザに対する負担が大きい.

この問題を解決する手段として, お気に入りしたツイートのカテゴリ毎に分類することが考えられる. 分類する際の基準には様々なものが考えられるが, 「役に立つツイート」, 「笑えるツイート」等, ツイートに対するユーザの価値や目的に基づいて分類できれば, 目的のツイートを探し出しやすくなる. しかし, 情報の粒度が小さく, 大量の情報を処理する必要がある Twitter において, ユーザが手動で個々のツイートを分類することは, ユーザに対するコストが大きく現実的でない. そこで, ツイートに含まれるテキスト情報に基づいて, ツイートを自動的に分類することが考えられる. ツイートに含まれる単語と分類したいカテゴリの依存関係が強い場合は, 従来テキスト分類を利用すれば, 効果的な分類が可能であると考えられる. 例えば, 「映画」, 「スポーツ」等といったカテゴリにおいては, それらのカテゴリに出現する可能性がある単語の出現傾向に特徴があるため, 従来一般的なテキスト分類手法を利用した自動的な分類が有効に機能すると考えられる. しかし, 「笑える」, 「役に立つ」といった, ユーザにとっての価値や目的に基づいたカテゴリに自動的に分類

することを考える場合には、ツイートに含まれる単語とカテゴリの依存関係が弱いために、従来の自動分類は有効でない。

我々は、ツイート本文のテキスト情報を利用したテキスト分類では効果的な分類が困難である対象を、自動的に分類可能とする手法を開発することが重要であると考え、本研究では、ユーザの目的や価値に基づいてツイートを自動的に分類する手法を開発することを目的とする。本研究では、自動分類が困難であると考えられるユーザにとっての価値や目的に応じたツイートのカテゴリの代表例として、「**ネタツイート**」を例題に利用する。ネタツイートとは、読み手を笑わせることを目的としたツイートである。ネタツイートのテーマは多種多様であり、様々なジャンルにまたがっている。したがって、ネタツイート自体には特定のジャンルに強く依存した単語の出現傾向があるわけではない。

上記の問題を解決するために、本論文では、ツイートの本文の特徴ではなく、そのツイートに対する反応の特徴を利用した分類手法を提案する。例えば、ネタツイートの本文自体には共通の言語的な特徴は存在しないが、ネタツイートに対する反応に対しては、それを読んだユーザが「笑った」というような共通した内容を表現するための言語的な特徴があると期待できる。我々は、ネタツイートに対する反応が持つこの特徴を利用すれば複数のツイートの中からネタツイートを分類できると考えた。本論文では、このアプローチに基づいてネタツイートを自動的に分類する手法を提案する。

2節では関連研究について述べる。3節では、本研究で利用する Twitter の機能を述べる。4節では、ユーザ反応を抽出するアプローチとその具象的な方法を述べる。5節では、実験とその結果、考察について述べる。6節では、まとめと今後の課題を述べる。

## 2. 関連研究

田中ら[2]は、Twitter におけるツイートの分類手法を提案している。ここでは、広く一般の人にとって有用であるかどうかという観点から、ツイートを分類する。具体的には、ツイートの以下の2点に基づいて分類を行っている。

- ツイートを投稿したユーザが、一般への情報発信を指向しているユーザか、主にコミュニティ内のコミュニケーションのためにtwitterを使用しているユーザか。
- ツイートの内容が一般への情報発信を指向しているか。

さらに、ユーザが広く一般に向けた情報発信を指向しているかどうかの判定を、そのユーザへのフォロー関係を、特定のコミュニティ内のフォロー関係とそうでないものに分類することによって行う手法と、そのユーザのツイートが「お気に入り」に登録される頻度を用いる手法を提案し比較している。また論文内で、Twitter上に発信されるツイートは、情報の質に関して大きく分けて以下の4つの属性を持つことを定義し

ている。

1. 情報発信性: ツイートが他ユーザに対して発信している意識があるかどうか
  2. リアルタイム性: そのツイートをリアルタイムで見ることに有益であるかどうか
  3. 社会性: 広く一般に有用であるかどうか
  4. 有用性: 他ユーザにとって有用であるかどうか
- また、ユーザの属性として以下の様なものを挙げている。

1. 利用目的: ユーザが何を目的として Twitter を利用しているか
2. 有名度: 現実や web 上でのそのユーザの人気
3. 嗜好: ユーザの嗜好、あるツイートを有用と感じるかどうか
4. 情報収集意識: Twitter 上で情報を収集する意識が高いか低い
5. 情報発信意識: ユーザがフォロワーに対して情報を発信する意識が高いか低い

他にも、Twitter に関する研究では、上記のようなツイート属性やユーザ属性を考慮したクラスタリングや、フォロー関係に基づくユーザ分析が行われている。

伊藤ら[4]は、対象ツイートに対し、関連ツイートの投稿時間と投稿したユーザの分布に基づくツイート分類手法を提案している。西田ら[5]は、Twitterにおけるツイート情報の圧縮されやすさに着目し、新しいツイートを、着目する話題に関するツイートの集合(話題モデル)と、それ以外のツイートの集合(比較モデル)の両方を基にして圧縮。データ圧縮を利用することで、形態素解析の結果に依存せず、新語、口語、俗語が多く含まれるツイートを精度良く分類する手法を提案している。

本研究では上記のような、ツイート属性やユーザ属性を考慮したクラスタリングや、フォロー関係に基づくユーザ分析行わず、ツイート本文とツイートに対する反応のみを利用し、ネタツイートの分類手法を提案する。

## 3. ツイートに対するユーザ反応の種類

本研究では Twitter 上に投稿され、ユーザによってお気に入りに入れられたツイート集合を対象に、ネタツイートを自動的に分類する手法を開発することを目的とする。そのために、ツイートに対する「ユーザ反応」(ツイートに対する反応ツイート)を利用する。ユーザ反応を表すツイートは、「引用ツイート」、「リプライ・メンション」、「リツイート」の3種類に分類できる。以下にこれらのツイートの特徴を述べる。

### (1) 引用ツイート

引用ツイートとは、元のツイート本文を引用し、自身のツイートを付け加えて投稿されたツイートである。他ユーザのツイートとそれに対する自身のツイート(意見)を単一のツイートとして投稿に出来るため、自分が興味をもったツイートとそれに対する反応の関係を解りやすく

提示できる。この形式では、情報が集約され、ツイートの情報的な価値が高くなると考えられる。

## (2) リプライ

リプライとは、特定ユーザ個人へ向けたツイートである。ツイートの先頭部分に@とユーザ ID (例:@taro) を付けることで、そのユーザ ID を持つユーザ宛てのツイートを投稿できる。リプライは、対象となるツイートに対し、ツイートの投稿者に直接コメントを送るようにツイートするために利用される。リプライと類似した役割のツイートとしてメンションと呼ばれるものがある。メンションでは@とユーザ ID が、文頭ではなくツイート本文のどこかに入っている。リプライは送信元と送信先、そしてその2人をフォローしているユーザのタイムラインに表示されるのに対し、メンションツイートはフォロワー全員のタイムラインに表示される。より多くのユーザに見られるため、情報発信性・情報拡散性はメンションツイートの方が高い。

## (3) リツイート

リツイートとは Twitter の提供する機能の一つであり、ユーザのタイムラインに表示された他人のツイートを自分が発信者としてフォロワーに送付する機能である。一般的な Twitter クライアントでは、表示されたツイートの下に、リツイート用のボタンが存在し、それを押すことでそのツイートをリツイートできる。リツイートによって、ユーザは自身のフォロワーと、自身が興味をもったツイートを共有できる。また Twitter ではリツイート直後に、そのツイートに対する自身の意見を投稿するという行動が多く見られる。そのようなツイートはユーザの反応を表していると考えられる。

## 4. ユーザの反応を利用したネタツイート自動分類手法

本研究はツイート集合の要素を、ネタツイートとそれ以外のツイートに自動分類することを目的とする。ネタツイートを自動分類するにあたって以下の問題がある。

1. ネタツイートは読み手を笑わせることを目的としているが、ツイートが笑わせることを目的としているか否かの判断を行うには人間と同程度の高度な能力を必要とする。ツイートが笑わせることを目的としているか否かをコンピュータが自動的に判断することは困難である。
2. ツイート本文は 140 字以内であり、非常に短いため、既存のテキスト分類手法では効果的に行うことは困難であると予想される。また、ネタツイート本文には、共通している部分に一貫性がないので、単語の出現傾向が似ている文書同士は似ているとする従来の一般的なテキスト分類法を利用した自動分類は困難である。

上記の問題を解決するために、本手法では、ツイートに

対するユーザ反応を利用して、ネタツイートか否かを判断するアプローチを採用する。このアプローチの概要を図 1 に示す。このアプローチでは、ネタツイートの付いている Twitter ユーザの反応パターンと、判別対象とするツイート(ツイート A とする)に付いているユーザの反応が似ているのならば、ツイート A はネタツイートである可能性が高いと考える。

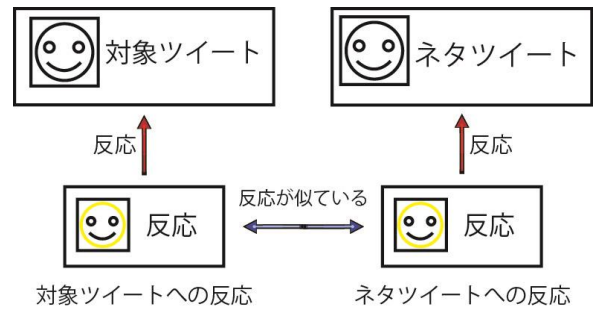


図 1 ユーザ反応を利用したネタツイート分類のアプローチ

そこで、以下の図 2 に示すネタツイート判定手法を提案する。この手法では、あらかじめ収集したツイートと反応ツイートの集合に対して、ネタツイートか否かの判定を行い学習データとする。この学習データを利用して、機械学習により分類器を生成する。そして、分類対象となるツイートに対して、システムは自動的に反応ツイートを収集し、得られた分類器を利用して、それらがネタツイートか否かを判定する。

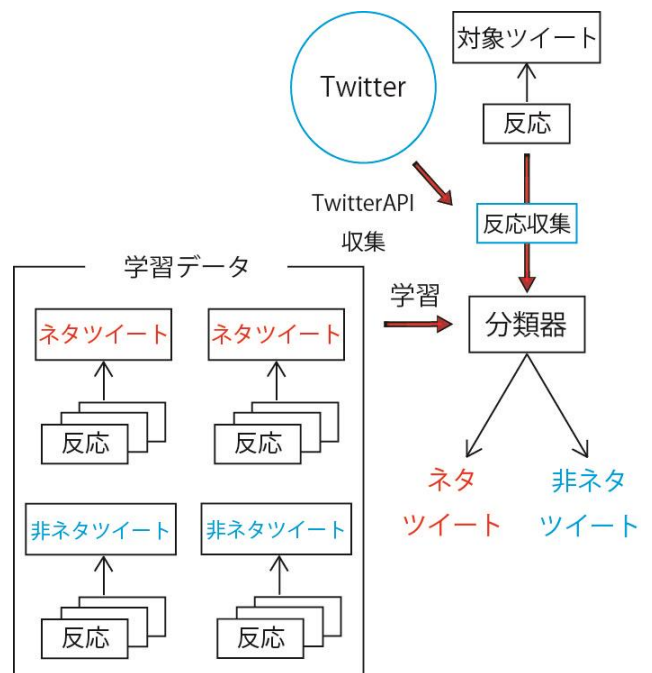


図 2 ネタツイート判定手法

提案手法を説明するために、まずツイートに対するユーザ反応を抽出する手法について述べ、次に抽出した反応から

ネタツイートを分類する手法を述べる。

#### 4.1. 反応ツイートの抽出

Twitter におけるユーザ反応ツイートの抽出のために、「リプライ・メンションツイート抽出」、「引用ツイート抽出」、「リツイート直後のツイート抽出」の三つの手法を用いる。以下にそれぞれの手法について説明する。なお、本研究では Twitter API を利用して、反応ツイートを収集する。

Twitter 上に存在するツイートには、それぞれステータスが割り振られている。ステータスを構成する代表的な属性としては、ユーザ ID を表す screen\_name, ユーザ名を表す name, ツイート ID を表す id\_str, アイコンを表す profile\_image\_url, ツイート本文を表す text, 投稿時間を表す created\_at, ツイートがリプライである場合に振られるリプライ ID を表す in\_reply\_to\_status\_id\_str 等がある。これらの属性を利用してユーザ反応の抽出を行う。

##### 4.1.1 リプライ・メンションツイート抽出

引用ツイートは、元ツイート本文の一部を引用し、自身の意見ツイートを付け加えているという特徴がある。この特徴を利用して引用ツイートを抽出する。通常、ツイートの全文をクエリとして Twitter 内で検索すると、そのツイート 1 件のみが検索結果として返されるのが一般的である。しかし、引用ツイートは、対象とするツイートの本文を引用するため、ツイート全文を引用している場合、引用文をクエリとして検索することにより、対象ツイートと複数の引用ツイートが検索結果として返される。返された引用ツイートから、引用文を除外すれば、意見部分が抽出できる。

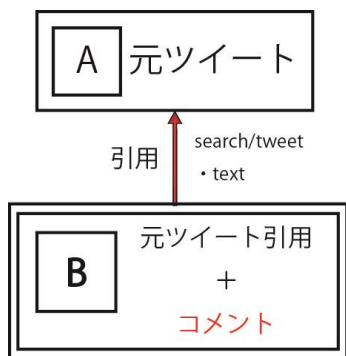


図 3 引用ツイート抽出

##### 4.1.2 引用ツイート抽出

リプライとメンションツイートは、ユーザが特定のユーザへ向けて投稿されたツイートである。ツイート本文に、「@送信対象のユーザの ID」を含めることで、そのユーザへ向けたツイートを投稿できる。これを抽出する方法を以下に述べる。

リプライとメンションツイートは、本文中に必ずユーザ ID を

含む。Twitter 内でユーザ ID をクエリとして本文を検索すれば、そのユーザへ送られたリプライとメンションツイートが表示される。次に、検索されたリプライが、どのツイートへの反応として投稿されたものかを判断する。そのために、ツイートのステータス内の in\_reply\_to\_status\_id\_str 属性を利用する。

ツイートは in\_reply\_to\_status\_id\_str (以下 reply\_id) と呼ばれるステータスを持っている。通常、各ツイートにはそれぞれツイート ID である id\_str が与えられているが、reply\_id はそれとは別に与えられる「あるツイートに対するリプライを示す ID」である。あるツイートに対するリプライを送った場合、そのリプライは id\_str の他に、ツイート A の id\_str と同じ値を持つ reply\_id を与えられる。したがって、元ツイートを投稿したユーザ ID を本文中に含み、かつ元ツイートの id\_str と一致する reply\_id を持つツイートを探し出せばよい。

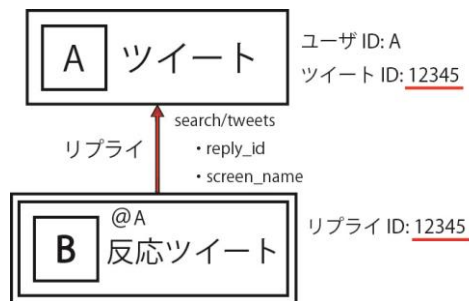


図 4 リプライ・メンション抽出

##### 4.1.3 リツイート直後のツイートの抽出

Twitter におけるユーザは、リツイートした直後に、それに対する自身の反応を示すツイートを投稿することが多い。そこで、対象とするツイートをリツイートしたユーザのタイムラインを取得し、リツイートした直後のツイートを取得する。現在の Twitter API では、特定のツイートに対するリツイートを自動的に取得する機能は提供していない。しかし、特定のツイートに対して、そのツイートをリツイートしたユーザは取得することが可能である。そこで、指定したツイートをリツイートしたユーザー一覧を取得する Twitter API の search/retweets を使い、指定ツイートをリツイートした最新ユーザ 100 件 (API で取得できる最大件数) を取得する。さらに特定のユーザタイムラインを取得する statuses/user\_timeline を用いて最新ユーザ上位から順にタイムラインを取得し、リツイートされた指定ツイートの直後のツイートを抽出する。

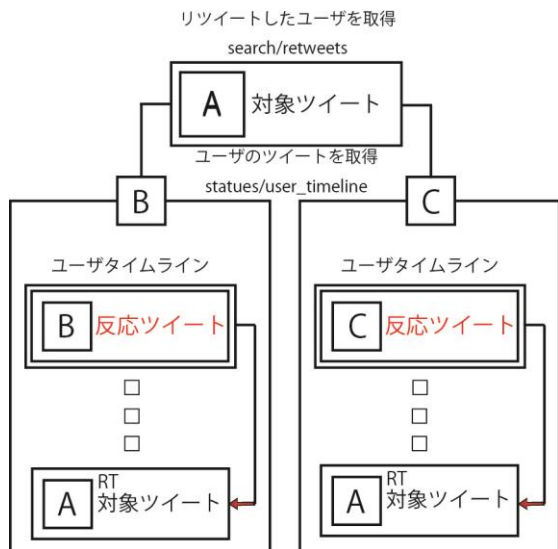


図 5 リツイート直後ツイート抽出

## 4.2. 反応パターンを利用したツイートの特徴のモデル化

4.1.で挙げた 3 つの方法で抽出した、ツイートに対するユーザ反応ツイートを利用して、ツイートの特徴のモデル化を行う手法について述べる。

本手法では、ツイートの特徴をモデル化するために、ツイート本文に含まれるテキストの特徴ではなく、ツイートに対する反応ツイートのテキストの特徴を利用する。

次に、反応ツイートの特徴のモデル化について述べる。一般に文書分類のためには、本文を形態素解析し、出現する単語の出現頻度に基づいて TF-IDF 手法などを利用した重み付けにより、文書の特徴をモデル化する場合が多い。しかし、単語に基づくテキストの特徴は、テキストのジャンルや話題に強く依存すると考えられる。ユーザの嗜好の解析などのような場合には、テキストのジャンルや話題に強く依存する特徴量が適切であるが、本研究のように特定の話題に依存しないためには適切でないと考えられる。そうしたことから、本研究では、N-gram を利用した特徴量を利用する。

N-gram とは文章を N 文字で区切り、連続する N 文字の出現パターンを利用する方法である。例えば、「アンパンと蒸しパンだ」という文章に 2-gram を適用すると、

アン・ンパ・パン・ンと・と蒸・蒸し・しパ・パン・ンだ  
となる。

我々は、反応ツイート  $r$  の特徴ベクトル  $rv_r$  を以下のように定義する。

$$rv_r = \begin{pmatrix} freq(r, p_1) \\ freq(r, p_2) \\ \vdots \end{pmatrix}$$

ここで、 $freq(r, p)$  は、反応ツイート  $r$  における 2-gram パターン  $p$  の出現頻度を表す。

次に、反応ツイートの特徴を利用して、対象ツイートの特徴をモデル化する。いま、対象ツイート  $t$  が有する反応ツイート集合が  $\{r_1, r_2, \dots, r_n\}$  であるとするとき、 $t$  の特徴ベクトル  $tv_t$  を以下のように定義する。

$$tv_t = \sum_{i=1}^n rv_{r_i}$$

我々は、予備実験として 100 件のネタツイートとその反応ツイートを収集し、収集した反応ツイートの本文を 2-gram で分解しデータベースに保存し、かつ空白や記号などのノイズを取り除いた。その結果、2-gram のパターンとして 2 万件が取得できた。これらの上位は、「ww」「わる」「!!」「面白」「って」などのパターンとなり、直感的には良好な結果が得られる可能性が大きいと考えられた。

## 4.3. ネタツイート判定

4.2.で述べたツイート反応に基づいてモデル化されたツイートの特徴に基づいてネタツイートを判定する手法を述べる。

ネタツイートの判定には、機械学習を利用する。サンプルとするツイートを収集しそれらに対して、人手によりネタツイートとそれ以外のツイートに分類する。そして、それぞれのツイートに対して、上記に示したやり方で反応ツイートを収集し、特徴ベクトルを構成し、それらを教師データとして分類器を学習させる。

ツイートがネタツイートか否かを判別する際には、システムは、入力されたツイートに対して自動的に反応ツイートを収集して、対象ツイートに関する特徴ベクトルを構成する。そして、生成された特徴ベクトルを、上記で学習された分類器で分類する。機械学習による訓練が可能な分類器として、SVM、ナイーブベイズ、ニューラルネット等、様々な手法が提案されている。本研究では、現在広く利用されている SVM を利用する。

## 5. 実験

提案手法のネタツイート分類精度を確かめるために実験を行った。ランダムに選出したツイートを対象とし、ツイート全 100 件のデータセットと全 200 件のデータセット、これら 2 組のデータセットを使用する。分類器に学習させるため、手動でデータセットの全ツイートをネタツイートとそれ以外に分類する。この時、ネタツイートに分類したツイートを実験における正解とする。手動で分類した後、データセットのツイート全てに対し 4.1.で述べた手法で反応ツイートの抽出を行い、4.2.で述べた手法でシステムによるネタツイート判定を行う。判定された結果から、人が手動でネタツイートに分類を行った正解と、提案手法による分類の正解率、適合率、再現率を測る。



なお、今回の実験では、分類には、libsvm を R から利用する e1071 パッケージを利用した。なお、今回は学習データとテストデータは同一のデータを利用した。

### 5.1. 実験結果

ツイート 100 件のデータセットを使用しネタツイート判定を行った結果を示す。

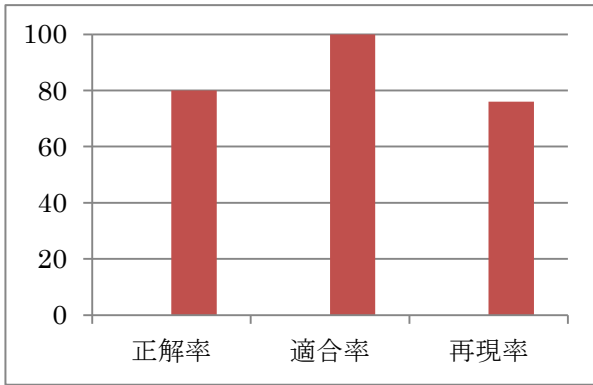


図 6 ツイート 100 件の判定結果

100 件のツイートの内、人が選択した正解は、ネタツイートが 66 件、それ以外のツイートが 34 件である。提案手法では、ネタツイートが 86 件、それ以外が 34 件と判定された。これらと比較し、正解率は 80%、適合率 100%、再現率 76%という結果となった。

次に、ツイート 200 件のデータセットを使用しネタツイート判定を行った結果を示す。

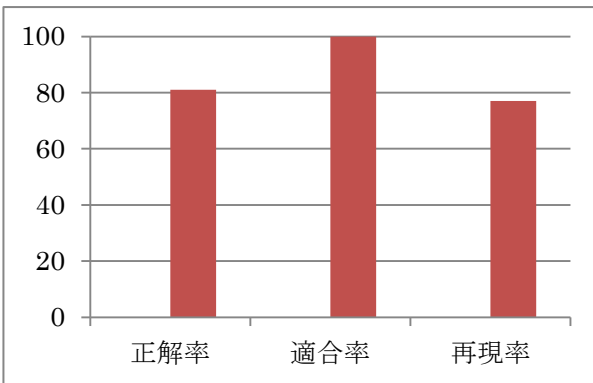


図 ツイート 200 件分類結果

200 件のツイートの内、人が選択した正解は、ネタツイートが 128 件、それ以外のツイートが 72 件である。提案手法では、ネタツイートが 163 件、それ以外が 37 件と判定された。これらと比較し、正解率は 81%、適合率 100%、再現率 77%という結果となった。

### 5.2. 考察

実験結果から、提案手法によるネタツイート分類の考察を

述べる。ツイート 100 件のデータセットにおいて、81%と非常に高い正解率を得ることができた。200 件のデータセットにおいても同様に、有意な結果が得られた。今回の実験では、訓練データとテストデータが同一としたため、精度が高いのは当然であると考えられるが、対象ツイートに付いているユーザ反応に基づいて特徴ベクトルを形成し、機械学習によって判定を行うことは、ネタツイート分類において有用である可能性が高いと考えられる。今後、データセットの量を増やし、交差検証法などにより、提案手法の有効性をより正確に評価する予定である。

### 6. まとめ

本論文では、「お気に入り」に登録されたツイートに対して、ユーザにとっての価値や目的に基づいた自動的な分類を行うことを目的に、ユーザの反応を利用したツイート分類手法を提案した。例題としてネタツイートを対象を限定して研究を行った。一般的に、ネタツイートには、様々なジャンルのツイートが含まれているので、ネタツイートをジャンルごとに分類することは実現可能であると考えられるが、ネタツイートとそれ以外のツイートに分類することは、ツイート本文のテキストを利用する一般的なテキスト分類手法では非常に困難である。そこで本論文では、ツイートに対する反応に注目し、反応ツイートのテキストの特徴を利用して、ツイートの特徴をモデル化した。本論文では、「引用ツイート」、「リプライ・メンションツイート」、「リツイート直後ツイート」の 3 つをツイートに対する反応とし、これらを抽出するための方法を提案した。そして、反応ツイートの本文の、n-gram の出現頻度に基づいて、対象ツイートの特徴をモデル化する。分類は、代表的な期外学習手法である SVM を利用した。提案手法の有効性を検証するために、を使用した実験を行った。実験の結果、提案手法は訓練データとテストデータが同一である場合には、高い精度で分類を行うことができた。今後、テストデータの数を多くして、交差検定法等の一般的な評価方法に基づいて提案手法の評価を行う予定である。

提案手法は、ユーザ反応を利用してネタツイート判定をする、というアプローチであるゆえにユーザ反応が全くないネタツイートに対しては提案手法が適用できない。この問題は自身でリツイート直後に反応パターンに沿ったツイートを投稿することで解決できる。また、本研究ではツイート本文とその反応に着目したが、ユーザのフォロー関係やプロフィールは考慮していない。自身やネタツイートを投稿しているユーザのユーザ属性、Twitter 利用指向やフォロー関係によるユーザクラスタリング、ツイート属性クラスタリングを行うことで、より精度の高い自動分類が可能になる。

### 参考文献

- [1] ふぁぼろぐ, <http://favolog.org/>
- [2] 田中淳史, 田島敬史, “Twitter のツイートに関する

分類手法の提案”，DEIM Forum 2010，2010  
<http://db-event.jpn.org/deim2010/proceedings/files/A5-4.pdf>

- [3] 白木原 渉, 大石 拓也, “Twitter における流行語先取り発言者の検出システムの開発”, 情報処理学会研究報告, 情報学基礎研究会報告 2010-IFAT-99(2), 1-8, 2010-07-28
- [4] 佐藤 勇也, 浅野 泰仁, “投稿日時とユーザの広がりに基づくツイート分類手法”, DEIM Forum 2011, 2011  
<http://db-event.jpn.org/deim2011/proceedings/pdf/a10-5.pdf>
- [5] 西田 京介, 坂野 遼平, “データ圧縮による Twitter のツイート話題分類”, DEIM Forum 2011, 2011  
<http://db-event.jpn.org/deim2011/proceedings/pdf/a1-6.pdf>
- [6] 田中 淳史, 田島 敬史, “Twitter のフォロー関係のユーザの意図に基づく分類”, DEIM Forum 2011, 2011  
<http://db-event.jpn.org/deim2011/proceedings/pdf/f5-1.pdf>
- [7] 松尾 潤, 川村 秀憲, “Twitter のコメント分析による広告の推薦”, 情報処理学会研究報告. ICS, [知能と複雑系]2011-ICS-162(4), 1-6, 2011-03-01
- [8] 山下 澄枝, 川喜田 佑介, “Twitter のための単語の出現頻度を用いたツイート有益度推定(「コミュニティ上での知識の共有・再利用」及び一般)”, 電子情報通信学会技術研究報告. AI, 人工知能と知識処理 110(301), 7-11, 2010-11-12
- [9] 橋本 泰一, 村上 浩司, “文書クラスタリングによるトピック抽出および課題発見”, 社会技術研究論文集 Vol. 5(2008), page216-226
- [10] 平博 順, 向内 隆文, “Support Vector Machine によるテキスト分類”, 情報処理学会研究報告. 自然言語処理研究会報告 98(99), 173-180, 1998-11-05
- [11] 藤川 智英, 鍛冶 伸裕, “マイクロブログ上の流言に対するユーザの態度の分類(テーマセッション, 大規模マルチメディアデータを対象とした次世代検索およびマイニング)”, 電子情報通信学会技術研究報告. DE, データ工学 111(76), 55-60, 2011-05-30
- [12] 天谷 祐介, 荒木 健治, “テキストの面白さ評価によるユーモアの認識”, ファジィシステムシンポジウム講演論文集 28, 233-238, 2012-09-12
- [13] Zi Yang, Jingyi Guo, “Understanding Retweeting Behaviors in Social Network”, CIKM '10 Proceedings of the 19th ACM international conference on Information and knowledge management, pages 1633-1636