

Twitter 上で行われる議論要約のための、  
文脈を表現する指標構築のための検討Consideration for Constructing Measurement  
to Express Context between Messages on Discussion at Twitter.

与儀涼子<sup>†</sup> 當間愛晃<sup>‡</sup> 赤嶺有平<sup>‡</sup> 山田孝治<sup>‡</sup> 遠藤聡志<sup>‡</sup>  
Suzuko Yogi Naruaki Toma Yuhei Akamine Koji Yamada Satoshi Endo

## 1. はじめに

近年、リアルタイムで動画を配信できる Web サービスである USTREAM と、マイクロブログの Twitter 双方の普及により、USTREAM で中継されている内容についての議論を Twitter 上で行う、という状況が増えている。このとき、Twitter 上での議論を要約することは、議論の経緯を把握しやすくし、途中から参加する人にとって有益である。そういった議論要約のためには、ツイート同士がどのような関係にあるのか、また、各ツイートはどういった種類の発言であるかを特定する必要がある。先行研究である IDM[1] は、ツイートによる情報の伝播をモデル化する研究であり、メッセージ間に共通する語彙によって影響力を計算するため、語彙的な情報に依存するという特徴がある。本研究では、語の表層的な情報のみならず、ツイートの内容からそのツイートの種類を分類するための指標を提案し、その有効性について検討する。

## 2. Twitter 上における議論の要約

本論文では、Twitter 上で普段行われている同時性のない議論ではなく、「ある特定の場にメンバーが集合し、Twitter を付随的に使う」ような場を想定している。目標とする議論要約の形は、「このツイート A は、それ以前のツイート B に対する“提案”である」といった論理的な関係を導き出す事である。そのためには、

1. 各ツイートを、“意見”や“提案”や“質問”などに分類する
2. ツイート間の関係を得る

という工程が必要である。本研究では、1の「ツイートを種類ごとに分類する」ための指標を提案する。

## 3. ツイート分類のための指標モデル

ツイートを分類するために、各ツイートをベクトル化して機械学習させる。本研究では、語の表層的な情報のみならず、ツイートの内容からそのツイートの種類を分類するための指標を提案する。

<sup>†</sup>琉球大学大学院理工学研究科 情報工学専攻  
<sup>‡</sup>琉球大学工学部情報工学科

詞情報を利用した。各ツイートを形態素解析し、得た品詞情報を用いてベクトル化を行う。今回、以下の4種類のモデルを構築した。形態素解析には mecab[4] を、辞書として ipadic[5] を利用した。

## 3.1. 品詞モデル 1

品詞情報の集合を属性とし、出現回数を値としてベクトル化する、最も単純なモデル。品詞情報は、ipadic で定義されている品詞クラス 1 の情報とする。属性数は 13 個である。

## 3.2. 品詞モデル 2

品詞の詳細情報の集合を属性とし、出現回数を値としてベクトル化した。語の詳細情報として、ipadic に定義されている品詞クラス 1 と品詞クラス 2 の組み合わせを使った。属性数は 44 個である。

## 3.3. 品詞モデル 3

隣り合う品詞情報の組み合わせの集合を属性としたベクトルを生成し、出現回数を値とする。品詞モデル 1 を 2gram に拡張したものである。属性数は 111 個である。

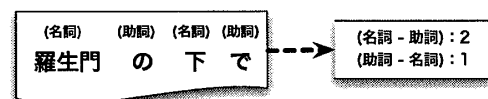


図 1: 品詞モデル 3 の例

## 3.4. 品詞モデル 4

隣り合う品詞の詳細情報の組み合わせ集合を属性としたベクトルを生成し、出現回数を値とする。品詞モデル 2 を 2gram に拡張したものである。属性数は 722 個である。

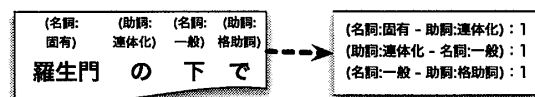


図 2: 品詞モデル 4 の例

## 4. 学習させるツイートのデータセット

本論文ではツイート分類の学習及びテストのために、2010年4月21日(水)に琉球大学工学部情報工学科

で行われた第一回情報工学科総会 [3] における専用ハッシュタグ #ieryukyu 内でのツイート 774 件をデータセットとして利用した。

#### 4.1. 教師データ

ツイートの分類は、各ツイートに“意見”や“提案”といったラベルをつけることで行った。ツイートは複数の文や意味を含むことがあるため、各ツイートが複数のラベルを同時に持ち得る、マルチラベル方式とする。全てのツイートについて、各ラベルが付く、または付かないことを学習させ、分類実験を行う。用意したのは以下の 10 種類のラベルである。

表 1: 各ラベルの意味

ラベル名	意味
aff	他のツイートに対する“賛同”であることを示す
ann	他のユーザ全員への“お知らせ”であることを示す
ans	あるツイートへの“回答”であることを示す
dem	何かに対する“要望”であることを示す
idea	アイデアを“提案”していることを示す
opin	何らかの“意見”であることを示す
ques	何かに対する“質問”であることを示す
sup	議論内容に関する“補足説明”など、事実を述べていることを示す
tsd	会場内にいるユーザが、現場で発表された発言や様子を要約して“実況”していることを示す。Twitter 用語で「tsudar」と呼ばれる種類のツイート
ets	その他、ハッシュタグのテストなどあまり議論内容とは関係がないツイートであることを示す

教師データは、全 774 件のツイートに対して手作業でラベルをつけることで作成した。

各ラベルの事例数を表 2 に示す。

**tsd ラベルについて** Twitter 上での議論において、tsd がついているツイートは、分類学習をさせるまでもなく特定することは容易である。そのため、tsd ラベルのみとなるツイート 105 件は学習データからは除外する事とした。

表 2: 各ラベルの事例数

ラベル名	事例数	ラベル名	事例数
aff	67	opin	384
ann	46	ques	73
ans	18	sup	113
dem	34	tsd	105
idea	78	etc	72

#### 4.2. 学習・テスト方法

ラベル付けされた 774 件のツイートのうち、tsd ラベルを除いた 669 件を分類学習させる。機械学習には、オープンソースのデータマイニングツールである Weka[2] を用いた。実験に使用した分類器は以下の 5 つである。

- J48: 決定木
- NB: ナイーブベイズ分類器
- NNge: k 近傍法
- SVMli: 線形カーネルの SVM
- SVMrb: RBF カーネルの SVM

実験は 10 分割交差検定で行う。

#### 5. 実験結果

今回学習するラベルは、正事例数が全事例数の 1 割にも満たないラベルも存在するため、単純な正解率では精度の判定ができない。よって、提案手法の精度評価のためのより妥当な基準として、F 値を利用する。

分類器によって A と分類された事例のうち、どれだけが実際に A であるかを適合率といい、A と分類されるべき全ての事例のうち、どれだけの事例を A と分類できたかを再現率という。F 値とは、再現率と適合率の調和平均である。

ラベルを付けるべきツイートに対する F 値を正事例の F 値、ラベルを付けるべきでないツイートに対する F 値を負事例の F 値、双方を平均した値を平均 F 値として、この 3 つを基準として精度を検証する。

##### 5.1. 各モデルの精度比較

各モデルの全ての分類器での、全てのラベルの判断精度を平均したグラフを図 3 に示す。

負事例の F 値についてはどのモデルもほとんど差はないが、正事例の F 値はモデル 4 が最大になっている。各モデルにおける事例を精査したところ、モデル 1 では、事例数の少ないラベルでは再現率 0% となってい

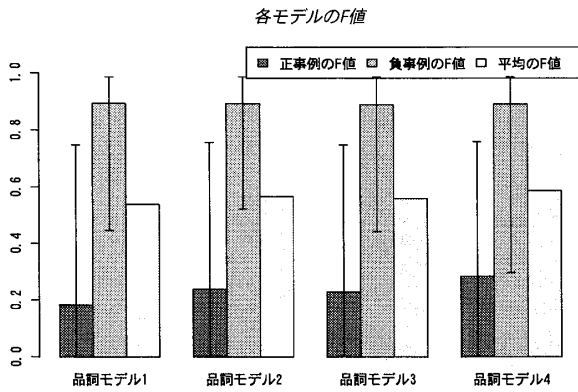


図 3: 各モデルの比較

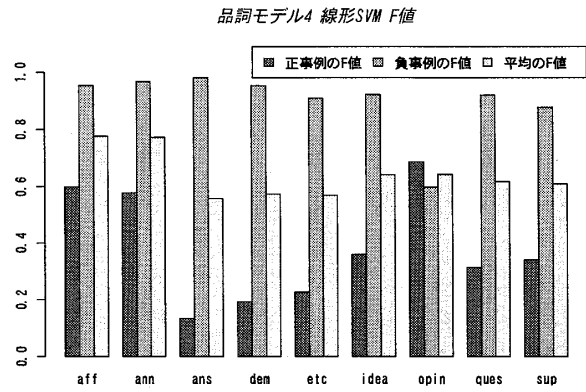


図 5: 各ラベルの比較

るものが多くあったが、モデル4ではそういった極端な判断は減っていた。情報量が多いモデルほど少ない正事例を拾いやすい傾向が観察された。

### 5.2. 各分類器の精度比較

各分類器の精度比較のため、モデル4における、全ラベルの平均精度比較グラフを図4に示す。

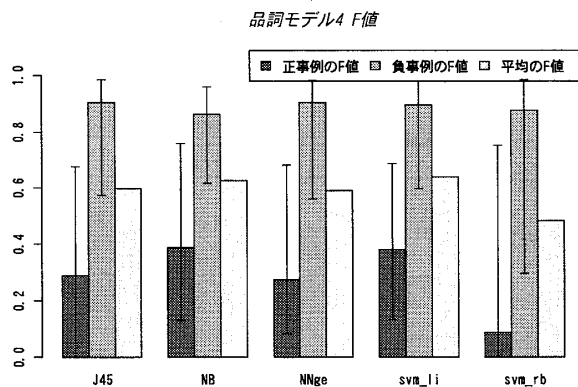


図 4: 各分類器の比較

全ての分類器で平均 F 値は 4 割を超えた。特に、線形 SVM(svm\_li) では 6 割を超え、今回最も良い結果となった。

### 5.3. 各ラベルの精度比較

各ラベルの精度比較のため、品詞モデル4で線形 SVM を用いた際の各ラベルの精度を図5に示す。

平均 F 値はだいたい 6 割を超えている。事例数の多い opin ラベルにおいて正事例の F 値が負事例より高くなっている。事例数の少ない ans や dem ラベルでは正事例の F 値が極端に低くなっている。

以降では、品詞モデル4で線形 SVM を用いた実験について論じる。

## 6. 各ラベルの傾向と対策

精度が高いラベルと低いラベルについて、原因や改善策を考案した。

### 6.1. ann

ann ラベルは、全ユーザーに対してお知らせをしているツイートに付加される。実験結果で平均 F 値が 7 割を超えた、精度の高いラベルである。

精度が高い理由として、リツイートによるツイート内容の拡散が考えられる。このラベルを持つツイートは、ツイート内容を更に広めようとするユーザーによってリツイートされる事が多い。そのリツイートにも同じ ann ラベルが付加されるため、よく似たベクトルを持つ事例がデータセット内に多く存在している事が、学習を容易にしていると考えられる。

ann ラベルの事例数 46 のうち、ann ラベルのツイートをリツイートしたものは 15 件あった。

### 6.2. ans

ans ラベルは、質問に対して回答しているツイートに付加される。実験結果で正事例の F 値が最も低かったラベルである。正事例を上手く拾えない理由として、このラベルの事例数が 18 件と少ないことが挙げられる。

また、別の理由として、リツイートによる混乱が考えられる。ans ラベルを持つツイートは、多くが ques ラベルのツイートをリツイートしたものである。よって両者はほとんど同じベクトルを持っており、片方が学習に使われもう一方がテストデータとして与えられたときに判断を誤る可能性が高い。これは ques ラベルの精度を下げる原因にもなっていると考えられる。

このことから、ツイートの RT 部分がどのラベルに属するかを判定することで、このラベルの精度を高めることができると予想される。また、対象ツイートが別のツイートに対するリツイートである場合、リツイート

元の持つラベルを対象ツイートにも付加する事で、ある程度混乱を低減できると考えられる。

### 6.3.opin

opin ラベルは、意見を述べるツイートに付加される。事例数が最も多く、実験での正事例の F 値が最も高いラベルである。このラベルを持つツイートは、ann ラベルや ans ラベルに見られるような顕著な特徴があまり見られないため、RT 処理などの前処理によって性能を改善することが難しいと考えられる。

意見のツイートは、post された時点での議論内容に關係する単語を含む事が多いため、ツイート中の語の表層格を利用したベクトルを作ることで改善可能だろう。また、TF-IDF などの語の局所性を利用したベクトルでも改善可能だと考えられる。

## 7. 改善策：RT 処理

いくつかのラベルの精度は、リツイートの影響を受けることがわかった。ここで、学習の際にリツイートを考慮した前処理を追加することで改善を試みる。

### 7.1.RT 元のラベルを追加

6.2 で述べた、「対象ツイートが別のツイートに対するリツイートである場合はリツイート元の持つラベルを対象ツイートに付加する」という前処理を追加した。

この処理を行った場合の、品詞モデル4で線形SVMを用いた際の各ラベルの精度グラフを図6に示す。

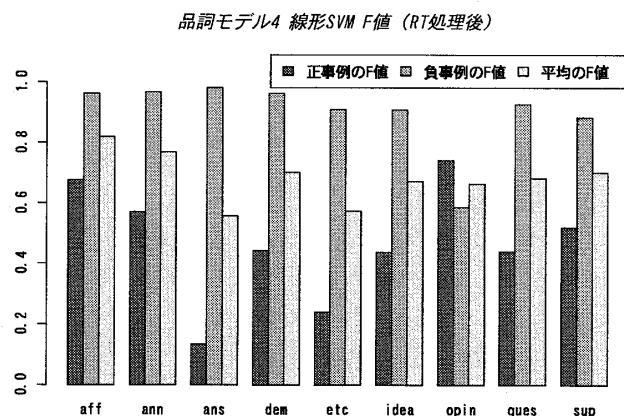


図6: 各ラベルの比較 (RT 処理後)

ほとんどのラベルにおいて、正事例の F 値と平均 F 値が改善されている。特に、dem ラベルや ques ラベルなど、他のユーザによってリツイートされやすいものの正事例の F 値が大きく改善されていることがわかる。aff,ann,dem,sup の4つのラベルで平均 F 値が7割を超えた。

## 7.2. 逐次学習の必要性

上記の方法で求めた精度は、リツイート部分が学習データに含まれている可能性がある、という前提の下での精度である。つまり、実利用する際にこの精度を出すためには、「現段階までのツイートを逐次学習させて更新する」という作業が必要になる。

## 8. おわりに

本研究では、Twitter 上での議論を要約するための一段階として、ツイートを機械学習によって分類することを目的とする4つのモデルを提案した。

ツイートを分類するための指標として、ツイート中の単語の品詞情報を用いた4つのモデルを構築し、実際の Twitter 上での議論を用いた分類実験を行いそれぞれの精度を検証した。

今後の課題として、このモデルを実装するための逐次学習方法の検討が挙げられる。また、分類実験において、語の局所性によって改善の余地があることを考察した。提案モデルを、TF-IDF などの局所性を判断する手法を用いて拡張することができるだろう。今回のモデルでは、RT 処理を除いて、ツイート間の関係性については考慮しなかったが、議論要約の生成のためにはツイート間の関係性の抽出が必要と考える。

最後に、7.1 で述べたように、今回の結果はリツイートによる補正を少なからず受けており、また、140 文字という限られた分量であることも、今回の実験結果に影響しているだろう。Twitter 固有の特徴に合わせたチューニングをすることで、より有用な指標を構築できるだろうと考える。

## 参考文献

- [1] 松村 真宏, “影響伝播モデル IDM の線形代数表現と Twitter 分析への応用”, 電子情報通信学会第二種研究会資料 (Web インテリジェンスとインタラクション), W12-2010-1 31, pp.73-78.
- [2] Weka  
<http://www.cs.waikato.ac.nz/ml/weka/>
- [3] 第一回情報工学科総会  
[http://www.u-ryukyu.ac.jp/tmp/ie\\_general\\_meeting20100423/index.html](http://www.u-ryukyu.ac.jp/tmp/ie_general_meeting20100423/index.html)
- [4] MeCab  
<http://mecab.sourceforge.net/>
- [5] ipadic  
<http://sourceforge.jp/projects/ipadic/>