

Twitter 上で共感を生み出すツイートの性質に関する考察

Consideration on Characteristics of Sympathy-arousing Tweets on Twitter

大川 陽聡^{1*} 高間 康史¹

Hisato Ohkawa¹, Yasufumi Takama¹

¹ 首都大学東京

¹Tokyo Metropolitan University

Abstract: Twitter 上では様々なコミュニケーションが発生しており、その情報伝播特性や震災時に果たした役割など、多様な観点から分析がなされている。我々は、不特定の相手向けになされたツイート（つぶやき）に対し多数の人が共感を抱くケースに着目し、その発生メカニズムの解明を目的としている。本稿では、ツイート自体が持つ特徴に着目して機械学習を適用することにより、共感を生み出すツイートの性質について考察する。

1. はじめに

Web は情報の共有やコミュニケーションの場として広く活用されており、特に近年のスマートフォンの発展・普及によりその傾向はさらに強まっている。Web を活用するためのサービスとして、ブログ・SNS・ミニブログ等が広く普及しているが、各サービスにおいてやり取りされる情報やコミュニケーションに対して多様な観点から分析がなされている。特にミニブログサービスの代表的な例である Twitter¹については、平成 23 年 3 月の東日本大震災時に被災地や避難所の情報共有や被災者への励ましなどに活用されたことが知られている。ここで注目すべきは、役立つ情報の共有だけでなく、他者とつながるためのメディアとして Twitter が利用されている点である。特に、Twitter や Face book のようなソーシャルメディアでは、他者とつながる事自体を目的とした利用の割合が高いと考える。この時重要なのは、役立つ情報を含んでいるか否かより、読み手の共感を得ることが重要であると考えられる。

我々は、不特定の相手向けになされたツイート（つぶやき）に対し多数の人が共感を抱くケースに着目し、そのメカニズムを解明することを目指し研究を進めている。本稿では、ツイート自体が持つ特徴に

着目して機械学習を適用することにより、共感を生み出すツイートを識別する分類器を生成し、共感を発生させるツイートが持つ性質について考察する。

2. 関連研究

2.1. Twitter 上における情報伝播

Twitter 上では、ユーザは他のユーザをフォロー (follow) し、また他のユーザからフォローされることにより、フォロー・フォロワーというつながりが発生する。このつながりをフォローネットワークと呼び、ユーザはフォローネットワーク上で主に以下の 6 つの機能を用いてコミュニケーションを取り合い、情報共有を行うことができる。

1. ツイート (tweet)
2. 返信 (リプライ, reply)
3. 公式リツイート (RT, official retweet)
4. 非公式リツイート
(非公式 RT, unofficial retweet)
5. 言及 (mention)
6. お気に入り (favorite)

ツイート、リプライ、リツイート、言及は短文を投稿する機能であり、お気に入りは他者の投稿内容を保存しておく機能である。

ツイートは、ユーザが 140 文字以内の短文を投稿することの総称であり、特定の相手へ向けたものではないものが大部分である。リプライは、文頭に「@ユーザ名」をつけた投稿であり、特定のユーザを対象とした投稿として用いられる。リツイートは、他者の投稿を引用し、紹介や拡散を行う機能である。

¹ <http://twitter.com>

*連絡先：首都大学東京大学院システムデザイン研究科
情報通信システム学域
〒191-0065 東京都日野市旭が丘 6-6
E-mail: ohakawa-hisato@sd.tmu.ac.jp

公式リツイートは他者の投稿を改変することなくそのまま引用するが、非公式リツイートは他者の投稿を自身の投稿の一部として用いることで、自身の意見を加えた投稿をすることができる。言及は、文頭以外の文中に「@ユーザ名」を含む投稿であり、特定のユーザに関するツイートであるという意味合いを持つ。

Twitter については、今までに以下の様な研究がなされている。

1. フォローネットワークによる情報伝播に関する研究[1]
2. Twitter ユーザの特性に関する研究[2]
3. リツイートに関する研究[3]
4. Twitter 外部からの影響に関する研究[4]
5. Twitter 上での特定の話題に関する研究[5]
6. ツイートの内容に関する研究[6]

情報伝播に関する研究としては、フォローネットワーク上での情報伝播についての研究である震災による情報伝播ネットワークの変化[1]などがなされている。Twitter ユーザの特性に関する研究としては、Twitter 上でフォローすべきユーザの推薦手法に関して、コミュニケーションに着目した Twitter フォローユーザ推薦[2]に関する研究がなされている。リツイートに関する研究としては、非公式リツイートに含まれる訂正コメントに着目した、リツイート構造を用いたデマ拡散防止支援手法[3]が提案されている。また Twitter 外部からの影響に関する研究として、Twitter 上のテキストが含む URL の傾向を把握する目的で、リンクを含むつぶやきに注目した Twitter の分析[4]が行われている。Twitter 上での特定の話題に関する研究としては、Twitter 上で注目されているキーワードに関する情報を抽出し提示するシステムの構築を目的として、Twitter におけるつぶやきの関連性を考慮した改良相関ルール抽出に関する研究[5]がされている。ツイート自体の内容に関する研究としては、任意のツイートとツイートへのリプライを対象とし、リプライの内容を分析することでツイートが持つ感情表現を推定する研究[6]がなされている。

これらの研究では、Twitter 上でのツイートの伝わり方や、リツイート、リプライなどを介した二次的な情報を中心に研究が行われているが、一時的な情報であるツイートそのものが持つ性質についての分析は行われていない。

2.2. 共感とソーシャルメディア

共感という概念に関する研究は、ソーシャルメディア分野のみでなく、経営学の分野でも行われてい

る。相原らは、ビジネスを行う特定の集団における経営イノベーションの向上を目的とし、その達成のために共感を用いた経営イノベーションと共感ネットワークについて研究している[7]。この研究では、「意思の疎通を図る＝親密になる」というプロセスを踏まえて組織内でのコミュニケーションが深まっていく段階における一部分として共感を定義しており、共感が発生するための条件には一定のコミュニケーションが必要であると指摘している。

また、ソーシャルメディア分野で用いられている共感の事例として、「SIPS」が挙げられる。「SIPS」とは、ソーシャルメディアが普及した状態で考えられる消費者の消費行動プロセスの考え方であり、共感 (Sympathize) → 確認 (Identify) → 参加 (Participate) → 共有・拡散 (Share&Spread) というプロセスのことである。これは、Twitter を「情報を共有し、新たな情報に出会える場」と認識した際の考え方であり、消費行動はユーザへの共感から始まるという仮定の基に研究が行われている。

これらの研究では、ソーシャルメディアやコミュニケーションにおける共感の重要性が指摘されているが、どのような行動が共感を得ることができるかについては議論されていない。

3. 共感を生み出すツイートの識別手法

本稿では、ツイートの共感発生メカニズムの解明を目的として、共感を呼ぶツイートの性質について考察する。分析手順として、分析対象とするツイートを収集し、データセットを構築する。収集したツイートについて、ツイートが共感を生み出すか否かに関係すると考えられる属性を定義し、特徴ベクトルを求める。最後に、特徴ベクトル集合に対して機械学習を適用し、学習結果からツイートの性質について考察する。本稿では Twitter 上におけるツイートに対する共感を「そのツイートを不特定多数のユーザが閲覧したときに、投稿した背景が想像でき、それに同感できる」ことであると定義する。ツイートを収集して特徴ベクトルを求め、機械学習によって共感を生み出すツイートとそれ以外を分類することで、ツイートの持つどの性質がユーザに共感を呼び起こさせるのかを考察する。

3.1. ツイートの収集と共感のラベル付け

ツイートの収集には Twitter Streaming API と Twitter REST API を用いる。また、本研究の対象であるツイ

ートは「不特定多数のユーザが閲覧するツイート」であるため、著名人のユーザ（投稿者）に対して収集時点における最新のツイート、および非公式リツイートを含めて一定の件数分取得する。取得した各ツイートを閲覧し、ツイートに対する共感が発生しているか否かを手作業により判別しラベル付けする。本稿では、第一著者の判断によりラベル付けを行った。以下に、共感を生み出すとラベル付けされたツイート、それ以外とラベル付けされたツイートの例を表 1 に示す。共感を生み出すと判定したツイートに関しては、テレビ番組の録画に失敗しているという悲しい状況から思わずツイートをしたくなるという経験がある人は比較的多いと考えられる。また、共感を生み出さないと判定したツイートは、宣伝がメインであり、共感する人は少ないと考える。

表 1 ツイートのラベル付けの例

ラベル	ツイート本文
共感を生み出す	録画失敗！！絶望だ。。。なんて最悪な日だ。。。
共感を生み出さない	息子が初めてしゃべる言葉を「パパ～バカ～」を目指してます。「●●のピンネタ LIVE」■/■（水）開場 18：30／開演 19：00【会場】▲▲【チケット】前売 1500 円／当日 1800 円（全席自由）P コード：××-×× チケット発売中！

3.2. 属性の定義と抽出方法

本稿で用いる属性を表 2 に示す。これらの属性は共感に関係すると考えられるもののうち、Twitter 上で使用可能なサービスやツイート内のテキストから取得可能であることを考慮して決定している。表 2 に抽出した属性名と属性値の形式および取りうる値を示す。fav, rt はユーザに依存する属性、それ以外はツイートに依存する属性である。fav および rt は、それぞれ対象ツイートが持つお気に入り登録数と公式リツイートをされた回数である。これらは、Twitter Streaming API のパラメータを参照することによって取得可能である。お気に入り登録やリツイートなどはユーザに閲覧されることにより発生するため、共感を生み出す前段階である閲覧の有無に関連すると想定している。term は対象ツイートと対象ツイートの直前に投稿されたツイートの投稿間隔であり、単位は分以下を切り捨てた時間である。これらは、Twitter Streaming API のパラメータから参照した値から計算して求める。頻繁に出現するツイートとた

まにしか出現しないツイートの間に共感発生の有無に関して差異があるのではないかと考える。この属性を定義している。characters は対象ツイートの文字数であり、twitter の仕様上 1 から 140 の値を取りうる。テキストの文字数はツイートの第一印象に関連すると考える。unofficial は対象ツイートが非公式リツイートをしているか否かであり、「RT @」をテキスト内に含む場合に yes、それ以外は no の値をとる。gladness, anger, sadness, pleasure はそれぞれ喜怒哀楽の感情表現がテキスト内に含まれているか否かであり、判定には感情語辞書を作成し用いた。ツイートに含まれる感情表現によって、閲覧したユーザへの影響があるのではないかと考える。これら agreement はツイートが肯定的か否定的か中立かということを示しており、thinking はツイートが賛成的か反対的中立かということを示している。これらはそれぞれ賛成語辞書、肯定語辞書を作成して判断している。ツイートの持つ印象がユーザになんらかの影響を与えるとの考えに基づき定義している。

表 2 属性名と属性値の形式

属性名	属性値の形式
fav	real 型, 0～
rt	real 型, 0～
term	real 型, 0～
characters	real 型, 1～140
unofficial	{yes, no}
gladness	{yes, no}
anger	{yes, no}
sadness	{yes, no}
pleasure	{yes, no}
agreement	{agree, disagree, neutral}
thinking	{positive, negative, neutral}

3.3. 機械学習

表 2 に示す各ツイートの属性値を説明属性、共感発生の有無を目的属性と定めてデータセットを作成し、機械学習を行う。本稿の目的に適した学習器が不明であるため、ここでは、対象となるデータセットに対して複数の学習器を使用し分類学習を行った結果について考察する。交差検証法を用い、適合率 (Precision)、再現率 (Recall)、正答率 (Correctly Classified Instances) により性能を評価する。各指標の定義を式 (1) (2) (3) にそれぞれ示す。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\begin{aligned} \text{Correctly Classified Instances} \\ = \frac{TP + TN}{TP + FP + FN + TN} \end{aligned} \quad (3)$$

ここで、TP, FP, FN, TN は以下を意味する。

- TP : 実際の共感発生は有で、有と判定した数
- FP : 実際の共感発生は無で、有と判定した数
- FN : 実際の共感発生は有で、無と判定した数
- TN : 実際の共感発生は無で、無と判定した数

共感を生み出すツイートの数は少ないことが想定され、実際に本稿で収集したデータでも 4.1 節で示すように共感を生み出すツイートの割合は低くなっている。このようなデータの特性および本稿の目的を考えると、共感を生み出すツイートを取りこぼすこと、すなわち FN が大きくなることを避けるべきである。そこで分類学習においては、FN へのペナルティが大きくなるようにコストを設定して学習を行う。

4. 実験結果

4.1. データセット

今回の実験で収集するツイートは「不特定多数のユーザが閲覧可能であるもの」を対象とするため、著名人の Twitter ユーザによるツイート、非公式リツイート、および言及を対象としている。著名人ユーザは一般人ユーザに比べてフォロワー数が非常に多い傾向があり、またユーザ層を特定しないため不特定多数のユーザが閲覧していることが多い。著名人ユーザは芸能人・企業人・学者・作家・スポーツ選手等様々な種類のタイプが挙げられるが、ユーザの絶対数が多く、ツイートの頻度が高めなタイプである芸能人、芸人を対象とした。また、リプライは特定の相手を対象としたつぶやきであること、公式リツイートは対象ユーザが第一次情報源でない投稿であることから、本実験においては除外している。

データセットは前述の芸能人、芸人の著名人ユーザの中から 5 名を選択し、各ユーザのツイートを収集し作成した。ツイート収集の手法として、各ユーザのツイート、非公式リツイート、言及であるツイートを 2012 年 9 月 29 日において最新のツイートから時系列順に数えて 50 件を取得した。作成したデータセットの概要を表に示す。

表 3 データセットの概要

ユーザ名	タイプ	フォロワー数	フォロワー数	共感発生数
庄司智春	芸人	133	12,611	7
陣内智則	芸人	174	144,809	18
有吉弘行	芸人	204	1,699,163	13
ROLA	タレント	133	1,048,947	9
ベッキー	タレント	0	826,753	14

4.2. 機械学習結果の考察

用意したデータセットに対して、データマイニングツール weka²を用いて機械学習を行った。精度の推定には交差検証法を用い、表 2 に示す属性のデータ型が利用可能で、かつコスト設定が可能なものについて学習を行った。本稿で用意した 5 つのデータセットについて全般的に性能がよかった 6 種類の手法 : SGD, SMO, MultilayerPerceptron, LWL, AdaBoostM1, LMT について表 4~表 9 にそれぞれ示す。ここで、性能評価は再現率を重視し、多くのデータセットに対して再現率が 7/10 以上でかつ適合率が 1/3 以上となるものを選択している。後者の条件は、あまりにも多数の負例を誤検出してしまうものを除くためである。また、コストは FP に対するコストを 1 とした場合の FN に対するコストを表す。

実験結果より、共感を生み出すツイートが持つ性質について考察する。陣内智則氏のデータセットを除き、ツイートの文字数 (characters) がある値以下の時に共感を生み出すと判定する学習結果が多く見られた。ツイートの長さが冗長であればあるほどユーザがテキストの全てを閲覧することがなくなってしまい、共感の発生を妨げる要因になっているのではないかと考える。また、要点がまとまっている短めの記事の方が共感を発生させやすいと考える。一人だけこのような傾向が見られなかった陣内智則氏のデータセットでは、ツイート中の文字数が増える顔文字を使用する頻度が多かったが、顔文字は感情を表現しやすく閲覧者に伝わりやすい傾向があるため他とは異なる結果になったと考える。

一方、agreement や thinking など、ツイート自体が持つ意見の傾向は共感の発生に影響を及ぼしている想定していたが、学習結果からはそのような傾向は読み取れなかった。この点については、より大きなデータセットでさらに検討をすすめる必要があると考える。gladness, anger, sadness, pleasure の感情表現に関する属性については、特に悲しみの感情を

² <http://www.cs.waikato.ac.nz/ml/weka/>

表すツイートが他の感情表現を表すツイートより共感を生み出すツイートの属性となる場合が多く見られた。悲しい記憶は人間の心に残りやすく、そのためツイートを投稿したユーザの状況をイメージしやすくなり、共感が発生する可能性が高くなるためと考える。

表 4 SGD の評価値

ユーザ名	コスト	精度	再現率	正解率
庄司智春	1.0or5.0	0.50	0.86	0.86
陣内智則	8.0	0.74	0.94	0.86
有吉弘行	17	0.63	0.92	0.84
ROLA	4.0	0.58	0.78	0.86
ベッキー	13or20	0.56	1.0	0.78

表 5 SMO の評価値

ユーザ名	コスト	精度	再現率	正解率
庄司智春	25	0.33	0.86	0.90
陣内智則	4.0	0.69	1.0	0.84
有吉弘行	2.0	0.71	0.77	0.86
ROLA	7.0	0.86	0.78	0.94
ベッキー	8.0	0.48	0.93	0.70

表 6 MultilayerPerceptron の評価値

ユーザ名	コスト	精度	再現率	正解率
庄司智春	9.0or10	0.45	0.71	0.84
陣内智則	5.0	0.84	0.89	0.90
有吉弘行	1.0	0.77	0.77	0.88
ROLA	7.0	0.58	0.78	0.86
ベッキー	20	0.65	0.79	0.84

表 7 LWL の評価値

ユーザ名	コスト	精度	再現率	正解率
庄司智春	18~20	0.33	0.86	0.74
陣内智則	1.0	0.67	1.0	0.82
有吉弘行	5.0~10	0.55	0.92	0.78
ROLA	20	0.33	0.67	0.70
ベッキー	8.0	0.48	0.86	0.70

表 8 AdaBoostM1 の評価値

ユーザ名	コスト	精度	再現率	正解率
庄司智春	15	0.25	0.57	0.70
陣内智則	4.0~6.0	0.69	1.0	0.84
有吉弘行	15	0.57	0.92	0.80
ROLA	11	0.64	0.78	0.92
ベッキー	11	0.52	0.86	0.74

表 9 LMT の評価値

ユーザ名	コスト	精度	再現率	正解率
庄司智春	18	0.24	0.57	0.68
陣内智則	5.0	0.68	0.94	0.82
有吉弘行	5.0	0.67	0.92	0.86
ROLA	4.0	0.80	0.89	0.94
ベッキー	7.0	0.52	0.86	0.74

5. おわりに

本稿では、Twitter において不特定多数のユーザがツイートを閲覧したときにツイートに対して共感を生み出すメカニズムについて調査するため、共感を生み出すツイートを識別する分類器を学習し、得られた分類規則の考察を行った。分析結果より、ツイート文字数や悲しみの感情が共感の発生に関係ある可能性のある結果が得られた。本稿で用いたデータセットはユーザ数、ユーザあたりのツイート数も小規模であるため、今後はより大規模なデータセットを用意して分析を行う予定である。また、今回のデータセットでは第一著者がラベル付けを行ったが、複数の評価者による判定でラベル付けを行う必要があると考えている。さらに、ツイートを閲覧する側の状況や性質に関する分析も必要であると考えている。Twitter ユーザは、ツイートを閲覧した際にリプライやリツイート等の様々なサービスを用いてアクションを起こすことができる。これらのアクションが共感を生み出すツイートに対してどう起こされるのか今後検討をしていく。

参考文献

- [1] 臼井 翔平, 鳥海 不二夫, 石井 健一郎, 間瀬 健二: 震災による情報伝播ネットワークの変化, 人工知能学会全国大会, 1C3-OS-12-3, 2012
- [2] 北村 太一, 小川祐樹, 諏訪博彦, 太田敏澄: コミュニケーションに着目した Twitter フォローユーザ推薦, 人工知能学会全国大会, 3E1-R-6-5, 2012
- [3] 中原 英美, 富永 一成, 牛尼 剛聡: リツイート構造を用いたデマ拡散防止支援手法, DEIM Forum 2012, F2-3, 2012
- [4] 吉田 光男, 乾 孝司, 山本 幹雄: リンクを含むつぶやきに着目した Twitter の分析, DEIM Forum 2010, A5-1, 2010
- [5] 鈴木 啓太, 新美 礼: Twitter におけるつぶやきの関連性を考慮した改良相関ルール抽出による話題抽出, 言語処理学会 第 17 回年次大会 発表論文集, pp.468-471, 2011

- [6] 堀宮 ありさ, 坂野 遼平, 佐藤 晴彦, 小山 聡, 栗原 正: Twitter における発話者へのリプライを用いたユーザー感情推定手法, DEIM Forum 2012, F2-1, 2012
- [7] 相原 憲一, 倉島 由佳子: 経営イノベーションと共感ネットワーク, 国際 P2M 学会 研究発表大会, pp. 151-160, 2010