

## SNSにおける集合気分のダイナミクス

## Dynamics of Collective Mood in Online Social Networks

笹原 和俊 \*1

Kazutoshi Sasahara

\*1 名古屋大学 大学院情報科学研究科

Graduate School of Information Science, Nagoya University

Social media such as Twitter often exhibit collective phenomena associating with real-life events, during which emotional expressions about the events spread over the user networks. To quantify collective mood in online networks, we proposed “emoticon networks,” in which nodes denote emoticons and Japanese adjective words in tweets, and directed links denote information flows among them. Using Twitter data in 2011, we constructed the emoticon networks to examine the dynamics of collective mood before and after the Japan earthquake. We found that the emoticon network before the earthquake had the one-directional chains of positive emoticons and adjectives; in contrast, that after the earthquake had a strange loop among positive and negative ones as well as many isolated nodes.

## 1. はじめに

ソーシャルメディアでは、しばしば実世界の事象と連動した集合現象が生じ [Sasahara 13], その事象に関する人々の感情情報が伝搬することが報告されている [Bollen 11a, Dodds 11]. また、そのような感情情報を利用してして株価を予測するという応用研究もなされている [Bollen 11b]. これらの先行研究は、ツイートに含まれる感情と関連した言語表現やキーワードの出現頻度を主に解析に利用している。しかし日本語の場合、自然言語処理も難しさもあって、既存の手法だけでは感情分析の精度が出ないというのが現状である。

一方、オンライン・コミュニケーションには言語だけでなく種々の顔文字が使用され、テキストに独特の感情情報を付加しているという事実がある [Park 13]. 特に日本語ではその傾向が顕著である。例えば、英語では笑顔を表現するときに :-) やこれをアレンジした記号列が使われるが、それほどバリエーションは多くない。一方、日本語の場合、笑顔の表現は (^\_^), (^\_^), \ (^\_^) /, (=^\_^=) など枚挙にいとまがない。このような顔文字の情報をうまく利用して、既存の自然言語処理や感情分析を補おうという研究が近年出てきている。しかし、顔文字は言語表現と比べて圧倒的に出現頻度が低いことや使用形態もまだ明らかでないことが多いため課題が多い。

そこで本稿では、感情に関係する顔文字と言語表現（形容詞）の相互作用に着目し、ソーシャルメディアにおける集合気分を定量化するためのツールとして「顔文字ネットワーク」(Emoticon Networks) を提案する。表出された顔文字と言語表現の1つ1つがユーザーの感情と1対1対応するとは限らないし、そもそもそれは誰にも（時に本人にさえも）わからない。しかし、それらを大規模に集めて分析することで、マクロな「気分」のダイナミクスを捉えられる可能性がある。

## 2. 方法

## 2.1 データの取得

ツイートの収集は、Twitter REST API\*1 を用いてスノーボール・サンプリングによって行い、2011年4月から約1年間をかけて、約40万人のユーザーから約5億ツイートのデータ（ユーザータイムライン）を取得した [Sasahara 13]. 各データには、ツイートのテキストの他に、ユーザープロフィール、投稿時間や位置情報（利用が許可されている場合）などのメタデータが含まれている。

## 2.2 顔文字ネットワークの構成

顔文字ネットワークとその構成手順を以下に述べる。顔文字ネットワークとは、感情に関わる複数の顔文字とそれらに関連する形容詞をノードとし、これらのノード間の情報の流れをネットワークとして表現したものである。

まず、調べたい顔文字とそれに関連する形容詞を選定する。顔文字の候補は Wikipedia\*2 や日本の顔文字\*3 などのサイトを参考にした。顔文字や形容詞の中にはツイートで多用されるものもあれば、ほとんど使われないものもあるため、予めこれらの出現頻度を確認しておく必要がある。あまりにも使用頻度が小さいものを用いた場合、これらの間の情報の流れは正しく推定されない。

次に、注目する期間の全ツイートを解析して、予め選定した顔文字と形容詞のすべて種類の出現時間と頻度を記録する。形態素解析には MeCab\*4 と NAIST-jdic\*5 を用いた。頻度時系列の各要素は、1時間ごとの頻度の総計を顔文字や形容詞を含む1日の総ツイート数で除した相対頻度を用いた。

こうして得られた形容詞と顔文字の頻度時系列を、順列エントロピー [Bandt 02] の考え方にしたがって、値の大きさの順番のみ着目して離散化する。例えば、頻度時系列が  $X = \{4, 7, 9, 10, 6, 11, 3\}$  の順番で観測されたとする。データを観測する窓のサイズを2として大きさを比較していくと、4と7は

\*1 <https://dev.twitter.com/docs/api/>\*2 [http://en.wikipedia.org/wiki/List\\_of\\_emoticons](http://en.wikipedia.org/wiki/List_of_emoticons)\*3 <http://japaneseemoticons.net>\*4 <http://mecab.googlecode.com/>\*5 <http://sourceforge.jp/projects/naist-jdic/>

12, 7 と 9 は 12 という具合にマップすることができ、もとの頻度時系列は {12, 12, 12, 21, 12, 21} と離散化することができる。

最後に、このようにして離散化した顔文字と形容詞の頻度時系列の間の移動エンロピー (TE) [Schreiber 00] を計算する。いくつかの先行研究において、ソーシャルデータにおいても移動エンロピーの計測が有効であることが示されている [Ver Steeg 12, Oka 13]。ただし、ここではサロゲートデータ (Y') を用いて見かけの移動エンロピー (TE) を計算し、それを TE から減じて補正した有効移動エンロピー (ET) [Marschinski 02] を用いた。頻度時系列 Y を 1000 回ランダムシャッフリングすることでサロゲートデータ Y' を得た。時系列 Y から X への移動エンロピー  $TE_{Y \rightarrow X}$  と有効移動エンロピー  $ET_{Y \rightarrow X}$  はそれぞれ次のように定義される。

$$ET_{Y \rightarrow X} = TE_{Y \rightarrow X} - \overline{TE}_{Y' \rightarrow X},$$

$$TE_{Y \rightarrow X} = \sum_{x_{n+1}, x_n, y_n} p(x_{n+1}, x_n, y_n) \log_2 \frac{p(x_{n+1}|x_n, y_n)}{p(x_{n+1}|x_n)}.$$

選定した顔文字と形容詞のそれぞれの頻度時系列を用いて有効移動エンロピーを総当たりで計算し、顔文字または形容詞をノード、有効移動エンロピーを有向リンクとして顔文字ネットワークを構成する。ここでは、 $TE_{X \rightarrow Y}$  と  $TE_{Y \rightarrow X}$  を比較して大きい方をリンクとして採用した。

このような手続きに基づいて顔文字ネットワークを構成し、非言語的な感情表現 (顔文字) と言語的な感情表現 (形容詞) の間の相互作用を可視化する。

### 3. 結果

最初に、記号による感情表現の統計を示し、どのような顔文字 (や漢字) がツイートに頻出するのかを確認する。そのデータをもとに 5 種類の顔文字と形容詞をそれぞれ選定し、東日本大震災前後のツイートを用いてそれぞれの頻度時系列を求め、それらから顔文字ネットワークを構成する。

#### 3.1 感情の記号表現

普通のツイートにおいて、感情を表現するのにどのような記号的表現が用いられているのかを知るために、取得した全ツイートを形態素解析し、顔文字や (笑) などの漢字による感情表現などを含めた計 528 種類の頻度順位を調べた。表 1 によると、10 位以内のうち 1 位から 9 位までの全てがポジティブな感情に関するものである。ちなみに 20 位までを見ても、そのうち 16 個がやはりポジティブな感情に関するものだった。したがって、通常のツイートではポジティブな顔文字を表出する方へバイアスがかかっていることが示唆される。先行研究において、英語の書物やニュースなどに用いられる表現では、ネガティブよりポジティブな単語の方が使用頻度が高いことが知られており [Kloumann 12]、ツイートにおける顔文字の使用もこれと同様の性質を示している。

#### 3.2 顔文字と形容詞の頻度時系列

先ほどの頻度順位において比較的頻出する 5 種類の顔文字 (T.T, ^\_^;, (^o^), (^o^)/, ^o^ ) とそれらに関係しそうな形容詞 (すごい, やばい, 怖い, 楽しい, ひどい) を選び、東日本大震災前後におけるそれらの頻度時系列を調べた。

図 1 は 2011 年 2 月から 4 月までの上記顔文字の頻度時系列である。これを見ると、震災当日は T.T を除く 4 つの顔文字の

表 1: 記号による感情表現

順位	感情の記号表現	相対頻度	感情
1	(笑)	0.159	ポジティブ
2	(^o^)	0.104	ポジティブ
3	^_^	0.068	ポジティブ
4	(^o^)/	0.039	ポジティブ
5	^^;	0.039	ややポジティブ
6	( (^o^), (^o^)/ )	0.034	ポジティブ
7	\ (^o^)/	0.034	ポジティブ
8	^_^;	0.033	ややポジティブ
9	(^O^)	0.033	ポジティブ
10	orz	0.030	ネガティブ

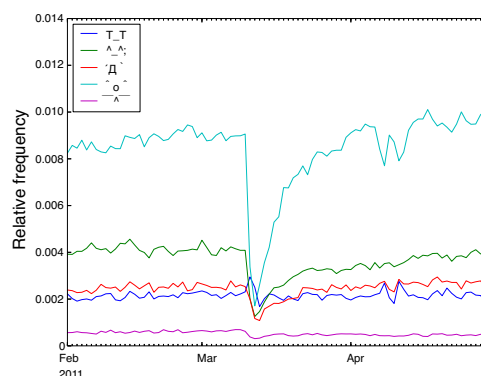


図 1: 震災前後における顔文字の頻度時系列。

出現頻度が減少していることがわかる。特に、(^o^) の出現頻度は急激に減少し、もとのレベルに回復するのに 1 ヶ月ぐらいかかっていることがわかる。これは、震災後 1 ヶ月程度は、多くの人が (^o^) を表出できない (あるいは意図的にそうしない) 心的状態にあり、人々の気分が落ち込んでいたことを示唆している。

図 2 は上記形容詞の頻度時系列である。震災当日は、「楽しい」を除く全ての形容詞の出現頻度が急増した。特に、「怖い」はその後も震災以前よりも高いレベルを示し、スパイク的な増大は余震や原発事故などと関連していると思われる。ここに示した頻度時系列には集団レベルの気分の変遷がある程度反映されていると考えられる。

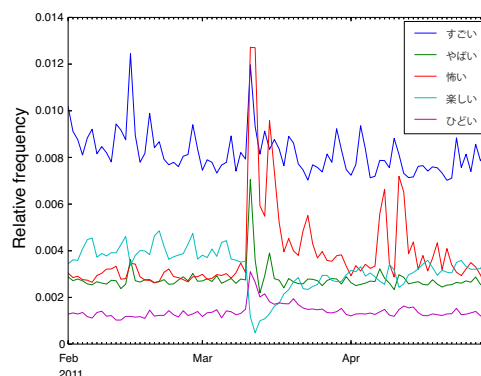


図 2: 震災前後における形容詞の頻度時系列。

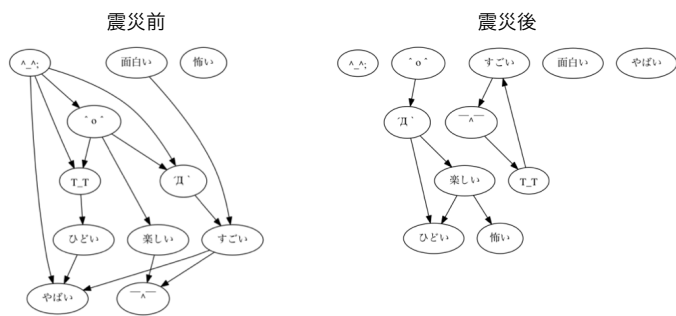


図 3: 震災前後の顔文字ネットワーク。

### 3.3 顔文字ネットワークの例

これまで見てきた 5 種類の顔文字と形容詞の頻度時系列のうち、震災前後一週間のデータを使って顔文字ネットワークを構成した。頻度時系列を離散化する際に使う窓のサイズは 2 とした。図 3 の左が震災前、右が震災後の顔文字ネットワークである。

震災前は、顔文字と形容詞が階層的かつ一方向的に結合されたネットワークになっている。例えば、 $\wedge \wedge \rightarrow \wedge \circ \wedge \rightarrow$ 「楽しい」というポジティブな情報の流れや、 $T.T \rightarrow$ 「ひどい」 $\rightarrow$ 「やばい」というネガティブな情報の流れも確認できる。また、「やばい」は使用される文脈によってポジティブにもネガティブにもなりうるため、「すごい」からも「ひどい」からも情報の流れがあることや、「怖い」は他の顔文字や形容詞との相互作用がないこともわかる。

一方、震災後は、「すごい」 $\rightarrow \wedge \rightarrow T.T \rightarrow$ 「すごい」のような解釈不明の閉ループや、「楽しい」が本来つながらず「ひどい」や「怖い」とつながるといった、通常とは異なる情報の流れが生じている。また、相互作用をしない孤立した顔文字や形容詞も多い。

このような顔文字ネットワークの結合構造の違いは、頻度時系列のみからはわからない震災前後の集合気分の変遷に関する非自明な相互作用を反映していると考えられる。

## 4. まとめ

本稿では、ツイートに含まれる感情に関わる顔文字や形容詞の頻度時系列を求め、有効移動エントロピーを用いて顔文字ネットワークを構成する方法を提案した。そして、顔文字の頻度の統計と頻度時系列の性質を確認した後、東日本大震災前後のデータを用いて顔文字ネットワークを構成した。その結果、震災前は顔文字と形容詞の間に一方向的な結合があるネットワークになったのに対して、震災後はそれとは異なる結合やループをもつネットワークになった。この違いは震災前後の集合気分の変遷と関連していると考えられる。

今後、この方法の有効性を、様々な事象に関するツイートを対象として様々なパラメータ設定を用いて確認する。特に、ノードに使用する顔文字や形容詞の選定や頻度時系列を離散化する場合の窓の大きさは結果に大きく影響するため、様々な条件で検討する必要がある。また、顔文字と形容詞のツイート内での共起関係を表すネットワークを構成し、顔文字ネットワークとの構造特性の違いを比較する。

## 謝辞

本研究は堀科学芸術振興財団の研究助成を受けて行われました。

## 参考文献

[Bandt 02] Bandt, C. and Pompe, B.: Permutation Entropy: A Natural Complexity Measure for Time Series, *Physical Review Letters*, Vol. 88, No. 17, p. 174102 (2002)

[Bollen 11a] Bollen, J., Gonçalves, B., Ruan, G., and Mao, H.: Happiness is assortative in online social networks, *Artificial Life*, Vol. 17, No. 3, pp. 237–251 (2011)

[Bollen 11b] Bollen, J., Mao, H., and Zeng, X.: Twitter Mood Predicts the Stock Market, *Journal of Computational Science*, Vol. 2(1), pp. 1–8 (2011)

[Dodds 11] Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., and Danforth, C. M.: Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter, *PLoS ONE*, Vol. 6, No. 12, p. e26752 (2011)

[Kloumann 12] Kloumann, I. M., Danforth, C. M., Harris, K. D., Bliss, C. A., and Dodds, P. S.: Positivity of the English Language, *PLoS ONE*, Vol. 7, No. 1, p. e29484 (2012)

[Marschinski 02] Marschinski, R. and Kantz, H.: Analysing the Information Flow between Financial Time Series, *The European Physical Journal B*, Vol. 30, No. 2, pp. 275–281 (2002)

[Oka 13] Oka, M. and Ikegami, T.: Exploring Default Mode and Information Flow on the Web, *PLoS ONE*, Vol. 8, No. 4, p. e60398 (2013)

[Park 13] Park, J., Barash, V., Fink, C., and Cha, M.: Emoticon Style: Interpreting Differences in Emoticons Across Cultures, *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media* (2013)

[Sasahara 13] Sasahara, K., Hirata, Y., Toyoda, M., Kitsueregawa, M., and Aihara, K.: Quantifying Collective Attention from Tweet Stream., *PLoS ONE*, Vol. 8, No. 4, p. e61823 (2013)

[Schreiber 00] Schreiber, T.: Measuring Information Transfer, *Physical Review Letters*, Vol. 85, pp. 461–464 (2000)

[Ver Steeg 12] Ver Steeg, G. and Galstyan, A.: Information Transfer in Social media, *WWW '12 Proceedings of the 21st international conference on World Wide Web* (2012)