

SNSにおける利用者の感情伝播に関する研究

横山 直敬^{*1} 牧野 浩典^{*2}

Study of Propagating Emotion in Social Network Service

by

Naotaka YOKOYAMA^{*1} and Hironori MAKINO^{*2}

(received on Oct. 29, 2016 & accepted on Dec. 1, 2016)

Abstract

Twitter is a popular Social Network Service (SNS) where users post their own thoughts and daily events. Emotions of the users posting tweets can be extrapolated easily by the Natural Language Processing (NLP) using the WordNet Affect^{1,2)}. This paper is intended to devise an observation and visualization method of propagating emotions between the twitter users in the network, and to elucidate properties of the propagation with respect to each of the 8 primary bipolar emotions first proposed by Plutchik³⁾. We use the transfer entropy to detect the propagating emotion⁴⁾, and elucidate which emotion is predisposed to propagate widely beyond the community structures of the network or locally within the community structures.

Keywords: Natural language processing, Propagating Emotion, Transfer entropy, Complex Network, Plutchik 8 primary emotions

1. はじめに

Twitter は大勢のユーザを擁する世界規模のソーシャルネットワークサービス (SNS) の一つであり、多くの利用者が自身の日常生活や意見を発信している。ツイートにはユーザ自身の感情が反映されているだけでなく、タイムラインに現れる他のユーザのツイートが示す感情の影響が及ぼされている。結果として Twitter では利用者同士の感情の同期や伝播が発生する。本研究の目的は Twitter のユーザ同士の感情の伝播を観測する方法を考案し、伝わりやすい感情の存在を明らかにすることである。ツイートが示す感情を Plutchik の 8 つの基本感情に基づいた 8 次元ベクトルで表現し、Twitter ネットワークで発生する感情の伝播を捉える³⁾。これまで定性的に認知されることの多かった「感情」という抽象的な存在が伝達する過程を可視化し、Twitter ネットワークを伝播しやすい感情やコミュニティ内で局所的にのみ伝播しやすい感情を明らかにする。

Twitter から収集できる情報は投稿されたツイートだけに留まらない。ツイートには投稿日時ややり取りをしている相手の ID などが付随する。またユーザの設定によっては位置情報さえ付随している場合がある。こうした付加情報を利用する研究は国内外を通じ数多くあり、ネットワークの状態を可視化する技術と連携した様々な応用例が提示されている。また、自然言語処理 (NLP) の分野ではツイートなどの文章から書き手の感情を推定する研究が進められている。中でも本研究と重要な関係にあるものは田口らが取り組んだ「感情天気図」の研究、および内藤らによるイベント記事へのコメントを介した感情伝播の研究である^{5,6)}。

田口らの考案した感情天気図とはツイートに付随する日時と位置情報、そしてツイートが表現している感情概念を用いて日本地図上にツイートが投稿された箇所とそのツイートが含む感情概念を示したものである⁵⁾。田口らは日本国内における大衆の感情表現を定量化し、時空間的に可視化することを試みた。内藤らは感情語を含むコメントがイベントの印象に与える影響を調査した⁶⁾。報道機関が発信しているニュース記事のツイートと、その記事を引用しコメントを追加した一般ユーザによるツイートを収集、記事に対するコメントから感情概念を抽出し、コメントが表す感情を定義した。実験は被験者に対してニュース記事本文とそれに追加されたコメント部分を順番に提示するという手法で行われた。内藤らはイベント記事への印象に関するアンケート調査をおこない、感情語を含んだイベント記事に対する他のユーザからのコメント文が記事に対する印象の変化を著しく促進させること明らかにした。内藤らの研究はコメント文を介した感情の伝播が発生したことを定性的に捉えている。本研究での試みはこのような伝播を定量的にとらえることである。また、感情概念そのものを推定し可視化する田口らの研究を発展させ、感情概念の伝播に対する可視化をおこなう。

2. Twitter からのデータ収集

Twitter はマイクロブログと呼ばれる SNS の一種であり、ユーザは自身の出来事や雑記を短い文章で投稿できるほか、他のユーザと互いにコミュニケーションを取ることができる。Twitter においてユーザによる投稿は 3 種類に大別される。一つはユーザ自身による投稿という最も基本的なもので「ツイート」と呼ばれる。もう一つは、他のユーザの投稿への言及にあたる「メンション」であるが、一般的にはユーザ同士のやり取りを指す「リプライ」という用語で呼ばれることが多い。最後の一つは他のユーザ

^{*1} 工学研究科情報理工学専攻

^{*2} 情報理工学部情報科学科准教授

の投稿を引用し自身が再投稿することで、「リツイート」と呼ばれる。本稿では原則として、これら3種類の投稿を区別なくツイートと表現する。ユーザ個人のページ(タイムライン)に表示される他のユーザのツイートは自身でユーザ毎に選択でき、選択しているユーザを「フォロー」しているという。また自身をフォローしているユーザを「フォロワー」と呼ぶ。

Twitter はユーザが様々なデータを収集するための API とよばれるインターフェイスを提供している。API にはリアルタイムに情報を取得し続けることのできる Streaming API と特定のデータを収集する REST API の2種類がある。本研究では REST API のみを用いており、筆者の一人(横山)のフォロワー及びフォローしているユーザ163名を対象にデータ収集を行なっている。以下では収集するデータの種類および具体的な収集手順について説明する。

まず、フォロワー一覧を取得する GET followers/ids メソッドとフォローしているユーザの一覧を取得する GET friends/ids メソッドを使い、筆者のフォロワー及び筆者をフォローしているユーザのリストを作成する。Twitter ではユーザの識別に User Name, Screen Name, User ID という3種類のデータが用いられるが、本研究ではユーザ毎に固有で任意の変更ができない User ID を取得している。ユーザリストは具体的には Table 1 に示すようなフォーマットで記載されており、Source はフォローする側の User ID を、Target はフォローされる側の User ID を指している。

Table 1 ユーザ同士のフォロー関係を示すリストの例。User ID はユーザ固有のものであるため一部を伏せて記載している

Source	Target
1xx53xxx0	2xx767xxx0
1xx53xxx0	2xx26xxx6
1xx53xxx0	1xx298xxx3
2xx767xxx0	1xx53xxx0
2xx767xxx0	2xx26xxx6
2xx767xxx0	1xx298xxx3
2xx767xxx0	1xx427xxx4

続いて GET friendships/show メソッドを用いて、リストに記載されたユーザ同士がフォロー関係にあるか否かを示した情報を取得し、ユーザ同士のフォロー関係をノードとリンクからなるネットワークで示す (Fig.1)。このネットワークは有向グラフであるが Fig.1 の表示スケールではフォローの向きを示す矢印が確認できないため、Fig.2 にその一部を拡大したものを示す。

特定のユーザが投稿したツイート文を取得するには GET statuses/user_timeline メソッドを使う。ツイート文を用いて各ユーザの感情概念の分析をおこなう。

本研究で用いるデータセットは2015年12月6日から2016年1月5日の間に収集したものであり、この期間内の筆者のフォロワー及び筆者がフォローしているユーザのリストとそのツイート文で構成されている。収集したツイートにはリツイートも含まれているが、これは他のユーザの投稿であってもそれを引用する以上、そこにはリツイートを行なったユーザの感情による影響が介在していると考えたためである。

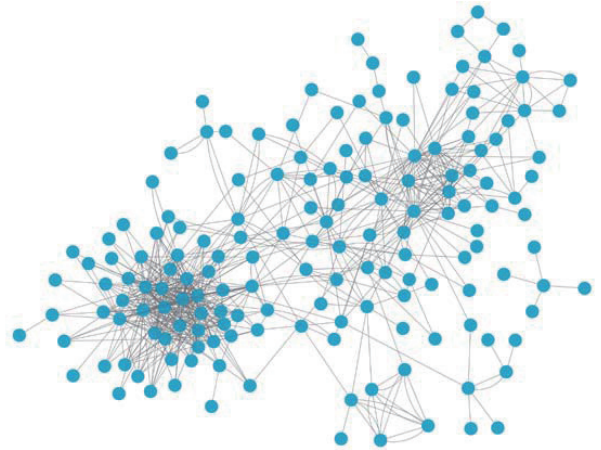


Fig.1 ユーザ同士のフォロー関係を表すネットワーク。各ノードは単一のユーザを表し、ユーザ同士の関係をリンクでつないでいる

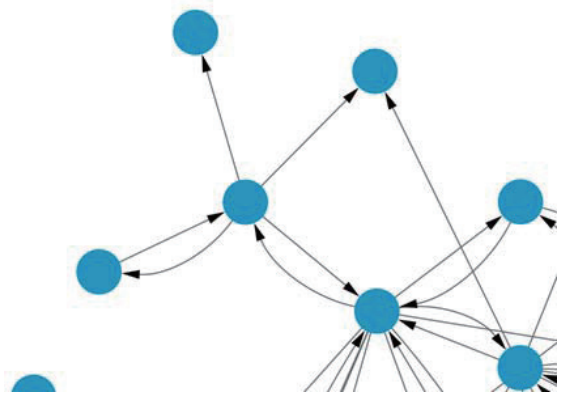


Fig.2 Fig.1の一部を拡大したもの

3. 感情分析と感情ベクトル

続いてツイートに含まれる投稿者の感情を推定する感情分析について解説する。感情分析はツイート文を構成する単語に対して感情概念を付与した感情語辞書を用いて行なう。本研究では感情語辞書に登録する単語の品詞を名詞、動詞、形容詞の3種類とし、日本語版 WordNet Affect の280の感情概念 synset を Plutchik の8つの基本感情に対応させることで感情語辞書を構築している。WordNet とはプリンストン大学の主導の下で運営されている概念辞書である^{1,2)}。WordNet に記述されている全ての単語は synset と呼ばれる同義語のグループに分類されており、synset 間には「上位語」や「下位語」といった概念上の関係がある。WordNet の synset に対して感情概念を付与したものが WordNet Affect である。Fig.3 に synset 間の関係の例を示す。同じ枠内にある単語は同一の synset に属することを表している。

WordNet には synset や単語を日本語に対応させた日本語版 WordNet があるが、WordNet Affect には日本語版が存在しない⁷⁾。そこで、本研究では WordNet Affect と日本語版 WordNet を用いた小澤による構築方法を採用し、日本語版 WordNet Affect

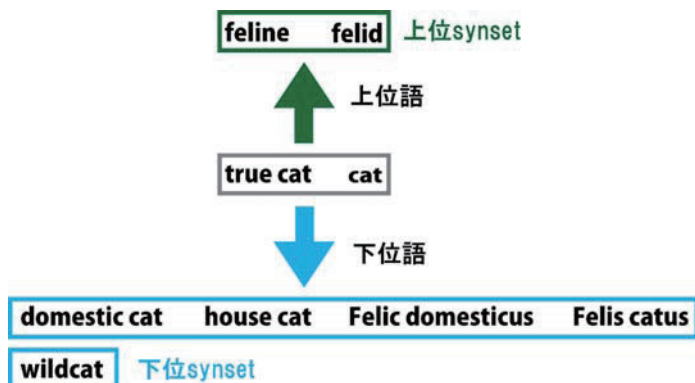


Fig.3 WordNet における synset 間の関係の例。同じ枠内にある単語は同一の synset に属することを表している。この事例では上位語に 1 つの synset、下位語に 2 つの synset が存在している

を得ている⁸⁾。我々が構築した日本語版 WordNet Affect の構成は synset が 798、単語数が 3378、感情概念が 280 となっている。辞書の中身は XML 言語で記述されており、具体的には Fig.4 に示すような品詞要素の下に感情概念 synset 要素、その下に単語要素を配置した構造になっている。

筆者らは日本語版 WordNet Affect 以外に感情表現辞典（中村明 東京堂出版 1993 年）⁹⁾ を用いた感情語辞書の構築にも取り組んでいる。ただし、感情表現辞典において各単語に割り当てられた感情概念は中村によって独自に選択されたものであり、客観的な裏付けが乏しい。日本語 WordNet Affect を使用して感情語辞書を構築したほうが感情概念を割り当てるメカニズムが明確である上に単語数において勝っており、ツイートからより多くの単語を感情語として抽出できる。

```
<adv-syn cat="confusion" caus-stat="stat" noun-synset="confusion.n.03"
offset="00294289" synset="confusedly.r.01"/>
-<adv-syn cat="despondency" caus-stat="stat" noun-synset="despondency.n.01"
offset="00301840" synset="despairingly.r.01"/>
<jpn-word lemma="真暗" pos="a" wordid="164403"/>
<jpn-word lemma="絶望的" pos="a" wordid="180547"/>
<jpn-word lemma="暗澹たる" pos="a" wordid="221837"/>
<jpn-word lemma="まっ暗" pos="a" wordid="234850"/>
<jpn-word lemma="救い難い" pos="a" wordid="238747"/>
<jpn-word lemma="真暗" pos="a" wordid="240847"/>
<jpn-word lemma="甲斐無し" pos="a" wordid="240966"/>
</adv-syn>
-<adv-syn cat="ambiguous-expectation" caus-stat="stat"
noun-synset="expectation.n.03" offset="00246077" synset="expectantly.r.01"/>
<jpn-word lemma="待ち受けて" pos="r" wordid="178408"/>
<jpn-word lemma="予期して" pos="r" wordid="205773"/>
<jpn-word lemma="期待して" pos="r" wordid="228892"/>
<jpn-word lemma="希望的" pos="a" wordid="230407"/>
<jpn-word lemma="楽観的" pos="a" wordid="235976"/>
</adv-syn>
-<adv-syn cat="daze" caus-stat="caus" noun-synset="daze.n.01"
offset="00082039" synset="dazzlingly.r.01"/>
<jpn-word lemma="すばらしい" pos="a" wordid="157495"/>
<jpn-word lemma="印象深い" pos="a" wordid="177204"/>
<jpn-word lemma="インプレッシブ" pos="a" wordid="196043"/>
<jpn-word lemma="インプレッシヴ" pos="a" wordid="223624"/>
<jpn-word lemma="空々" pos="a" wordid="228829"/>
<jpn-word lemma="印象的" pos="a" wordid="243479"/>
<jpn-word lemma="空空" pos="a" wordid="243594"/>
<jpn-word lemma="高入" pos="a" wordid="246841"/>
</adv-syn>
```

Fig.4 XML で記述された日本語版 WordNet Affect の一部

日本語版 WordNet Affect において個々の単語は単一ないしは複数の感情概念 synset にまたがって属し、単語が表す感情は感

情概念 synset により決定される。以下では日本語 WordNet に基づく感情語辞書の具体的な構築手法と、感情概念ベクトルの決定方法について述べる。

感情語辞書の構築は 3 つの段階からなる。まず 1 段階では 280 種類に細分化された日本語版 WordNet Affect の感情概念 synset を Plutchik の感情概念（8 基本感情 × 3 段階 + 8 応用感情の計 32 感情）へと変換することである。Plutchik の感情概念とは Plutchik が文献³⁾で提唱した Joy, Trust, Fear, Surprise, Sadness, Disgust, Anger, Anticipation の 8 つの基本感情をベースとする感情モデルであり、Fig.5 に示す Wheel of Emotion で表現される。各感情には 3 段階の強度があり、円の中心に向かうほど強度が高い。また各感情の間にはそれぞれの感情が合成された応用感情 optimism, love, submission, etc... が示されている。

感情概念 synset から Plutchik の感情概念への関連付けは自然言語処理ツール NLTK を用いた synset 間の類似度計算 path_similarity によって行なう¹⁰⁾。Fig.6 には類似度を計算する 2 つの感情概念 synset (synset A と synset B) の上位階層構造を概念的に表した。具体的には synset A を日本語版 WordNet Affect の感情概念 synset、synset B を関連付ける Plutchik の感情概念が属する synset とする。path_similarity では 2 つの synset の上位にある共通の synset を経由し、両 synset 間を最短経路で結ぶパスの長さを用いて類似度を計算している。synset 間を最短経路で結ぶパスの長さを P 、最長経路長で結ぶパスの階数を D とすると、類似度は自然対数を用いて $S = -\log(P/2D)$ で与えられる¹⁰⁾。synset と単語の類似度を求める場合は、単語が属する複数の synset の中で、synset 間の類似度が最も高いものを選び、それを synset と単語の間の類似度とする。この方法により類似度が最大値を示す Plutchik の感情概念 (32 単語) を 280 の感情概念 synset のそれぞれに関連付けている。

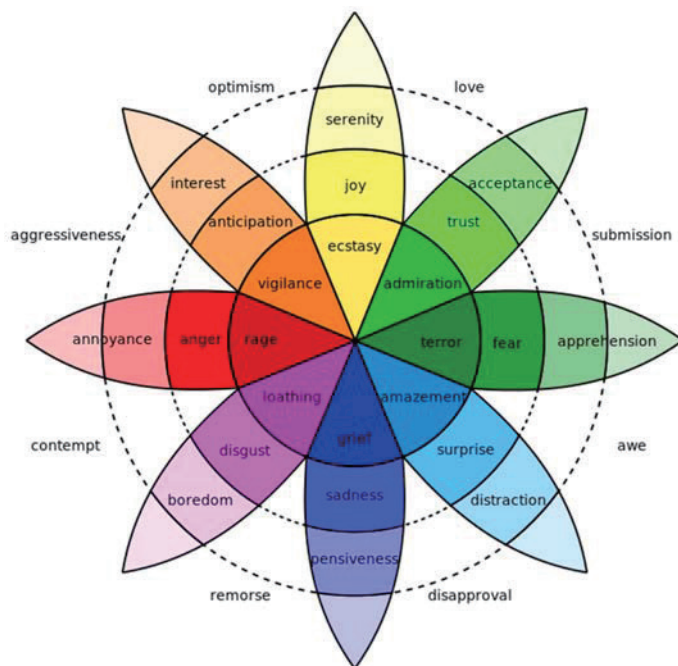


Fig.5 Plutchik の Wheel of Emotion。Wheel には 32 の感情概念（8 基本感情 × 3 段階 + 8 応用感情）が割り当てられている

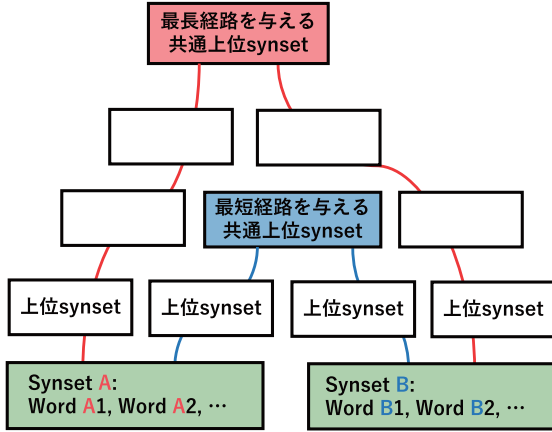


Fig.6 2つの synset(synset A と synset B) を隔てる上位階層構造の概念図。この図の事例では synset A と synset B の間を最長経路で結ぶ共通上位 synset までの階数が $D = 4$ 、最短経路で結ぶ経路長が $P = 4$ である

続いて感情概念 synset から感情概念ベクトルを求める手順について述べる。感情概念ベクトルは 8 次元ベクトル空間で表現され、Plutchik の基本感情の強弱に対応した成分を持つものとする。各成分は感情概念 synset に関連付けられた感情概念によって決定される。一つの感情概念が取り得る値は感情概念の強度が高い順に 1.25, 1.0, 0.75 とし、また感情概念 synset に応用感情が関連付けられている場合には、それを構成する 2つの感情概念に強度・中の 1.0 を割り当てている。日本語版 WordNet Affect において個々の単語は単一ないしは複数の感情概念 synset にまたがって属している。この場合、単語が属する複数の感情概念 synset（すなわち複数の synset A）がそれぞれ 1つの感情概念ベクトルを与えるので、それぞれの感情概念ベクトルの線形和をとることで、単語の感情概念ベクトルを単一に決定する。複数の単語のセットから単一の感情概念ベクトルを与える場合にも同様に線形和をとり、ベクトルは最後に正規化する。本研究では 1 日を基本単位とする感情概念ベクトルを構築する。一日分のツイートから単語セットを取り出し、線形和と正規化を経て、その日のユーザの感情を表す感情概念ベクトルを求める。

4. コミュニティ検出

現実世界に学校や企業あるいは地域などのコミュニティがあるように、Twitter にも現実の友人関係や共通した話題・趣味嗜好を持つ者が構成するコミュニティ構造がある。本節では Twitter からコミュニティ構造を抽出する手法を説明する。コミュニティ構造とはグループ内で密接にリンクし合っているが、グループ外とは疎な関係のリンクしか持たないノード集合のことである。こうした傾向の強弱を定量的に与える指標として一般にモジュラリティという量が用いられる¹⁸⁾。

コミュニティ構造を検出する研究は複雑ネットワークの研究よりも古く、早期のものとしては 1977 年の Zachary による大学の空手クラブのネットワークの解析が知られている¹⁵⁾。複雑ネットワークの分野では 2002 年の Girvan らによる枝の

媒介中心性に基づくコミュニティ検出方法の提案を契機に研究が盛んになった¹¹⁾。枝の媒介中心性とはネットワークを構成する枝（リンク）がどれだけ多くの最短経路上にあるかによって定義されるネットワークの特徴量である。Girvan らの研究以降では Pons らによるランダムウォークに基づくアルゴリズムや Newman による固有ベクトルを用いたアルゴリズム、Reichardt らによるスピンモデルを用いた焼きなまし法のアルゴリズムなど複数の分割アルゴリズムが考案されている^{12, 13, 14)}。これらアルゴリズムの分割精度を定量的に示すために Newman らが提案した指標がモジュラリティである¹⁸⁾。現在はこれがコミュニティ検出の精度を表す標準的な指標として広く用いられている。

モジュラリティとはリンクとノードで構成されているネットワークにおいて、ある何らかの方法によって実現したグループ分割 $\{G_1, G_2, G_3, \dots, G_K\}$ に対して「グループ内のノード同士がつながるリンクの割合 R 」から「リンクがランダムに配置された場合の R の期待値 $\bar{R}$ 」を引いた値として定義される¹⁶⁾。ネットワーク全体のノード数を N 、リンクの数を L 、各ノードのアドレスを $n = 1, 2, 3, \dots, N$ とし、ノード間のリンクの有無を 1 または 0 で表すネットワーク隣接行列を A_{nm} とするならば、隣接行列の成分の和は $\sum_{n=1}^N \sum_{m=1}^N A_{nm} = 2L$ となる。グループ G_k に属するノードとグループ G_l に属するノードがつながるリンク数の合計の、全リンク数に占める割合は

$$e_{kl} = \frac{1}{2L} \sum_{n \in G_k} \sum_{m \in G_l} A_{nm} \quad (1)$$

であり、リンクの割合 R と期待値 $\bar{R}$ は行列 $\mathbf{e}$ を用いて

$$R = \text{Tr}(\mathbf{e}) = \sum_{k=1}^K e_{kk}, \quad (2)$$

$$\bar{R} = \|\mathbf{e}^2\| = \sum_{k=1}^K \left(\sum_{l=1}^K e_{kl} \right)^2 \quad (3)$$

と表されるので、モジュラリティ M は

$$M = \text{Tr}(\mathbf{e}) - \|\mathbf{e}^2\| \quad (4)$$

と与えられる¹⁷⁾。 M の値は正にも負にもなりネットワークが強いコミュニティ構造を持つ場合には 1 に近い値を示し、コミュニティ構造を持たない場合には $M \sim 0$ を示す。一般に M が 0.3 程度より大きな値を示すならば、そのネットワークはコミュニティ構造を持つと判断される¹⁷⁾。Fortunato は複数考案されているコミュニティ分割アルゴリズムの効率を計算時間と検出精度（モジュラリティ）の 2つの観点から比較しており、モジュラリティの観点では Reichardt らの焼きなまし法アルゴリズムが最適であることを報告している¹⁹⁾。こうした背景を踏まえ、本研究では焼きなまし法に基づくアルゴリズムを用いて、第 2 節で得られた Twitter ネットワークのコミュニティ分割をおこなう¹⁴⁾。なお、統計解析言語 R 中のネットワーク解析ライブラリ igraph には Reichardt らの焼きなまし法に基づく分割アルゴリズムである spinglass.community 関数が用意されている。spinglass.community 関数はグループ分割数 K の最大値を引数としており、関数を用いる際は対象とするネットワークの大きさや構造を考慮しながら最適な最大値を代入する必要がある。

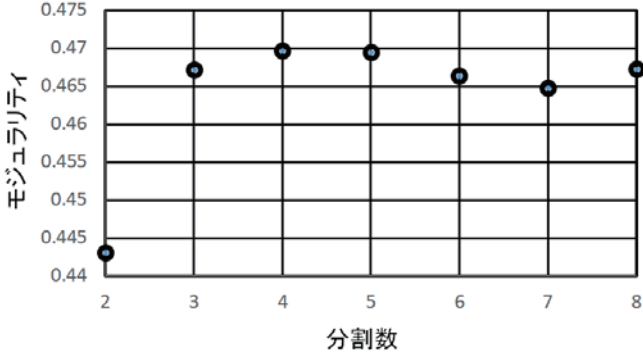
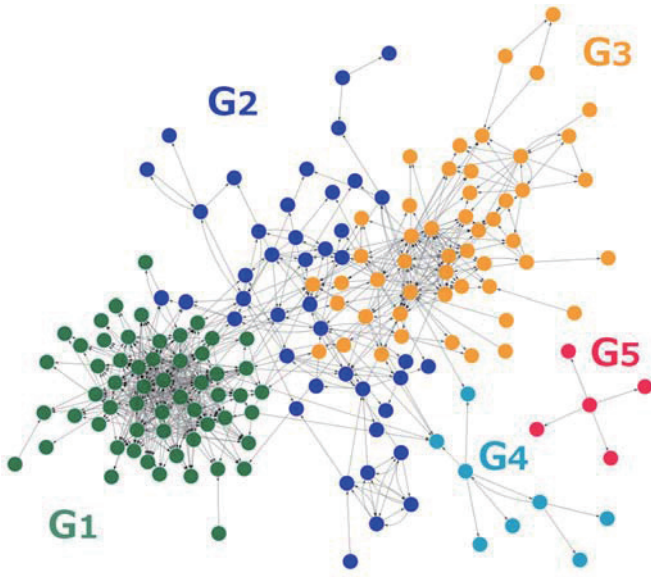

 Fig.7 分割数 K に対するモジュラリティ M の推移


Fig.8 Twitter ネットワークのコミュニティ分布

モジュラリティが最大値になるような分割数を理論的に決める有効な方法は現在までのところ存在しない。このため、本研究では試行的にモジュラリティを求め、最適な分割数を決めるよう工夫している。グループの最小構成ノードを3と仮定し、分割数を1ずつ増やしながら `spinglass.community` 関数を用いてネットワークを分割し、3ノードよりも小さなグループが出現するところで最大値を決定することにした。Fig.7にはツイッターネットワークに対して求めた分割数 K の値に対するモジュラリティ M の値の推移を示した。研究対象としたネットワークでは分割数が8を超えると3ノードよりも小さなグループが出現するため、分割数の最大値は8となっている。この計算例では分割数が4のところではモジュラリティが最大となり、グループ分割数の最適値を与えている。Fig.8には `spinglass.community` 関数によるコミュニティ分割の結果を示した。図中の G_5 (グループ5) はリンクを持たない孤立したグループのためモジュラリティへの寄与はなく、このネットワークの実質的な分割数は4である。

5. 移動エントロピー

移動エントロピーは2000年にSchreiberが提案した情報量の一種であり、2つの確率変数間の情報の流れを定式化したものである⁴⁾。確率変数間の相関のみを表す相互情報量とは異なり、情報の流れる方向までを含めた解析に適している。時刻 n における確率変数 X, Y の要素をそれぞれ x_n, y_n とするとき、 X に対する Y の影響を表す移動エントロピーは

$$TE_{Y \rightarrow X} = \sum_{x_{n+1}, x_n, y_n} p(x_{n+1}, x_n, y_n) \times \log \frac{p(x_{n+1}|x_n, y_n)}{p(x_{n+1}|x_n)} \quad (5)$$

で定義される。ここで、分布関数 $p(x_{n+1}, x_n, y_n)$ は x_{n+1}, x_n, y_n のそれぞれがある特定の値をとる同時確率を表し、 $p(x_{n+1}|x_n)$ は x_n がある値を取る条件において x_{n+1} がある特定の値を取る条件付確率をあらわしている。 $TE_{Y \rightarrow X}$ は非負の値であり、 Y の影響が X に及ぶとき $TE_{Y \rightarrow X} > 0$ 、 Y と X が互いに独立の時に $TE_{Y \rightarrow X} = 0$ を与える。この式は $p(x_{n+1}|x_n, y_n) = p(x_{n+1}, x_n, y_n)/p(x_n, y_n)$ と $p(x_{n+1}|x_n) = p(x_{n+1}, x_n)/p(x_n)$ の関係を用いて

$$TE_{Y \rightarrow X} = \sum_{x_{n+1}, x_n, y_n} p(x_{n+1}, x_n, y_n) \times \log \frac{p(x_{n+1}, x_n, y_n)p(x_n)}{p(x_n, y_n)p(x_{n+1}, x_n)} \quad (6)$$

と表される。 X の影響が Y に及ぶ逆方向のケースについても同様に

$$TE_{X \rightarrow Y} = \sum_{y_{n+1}, x_n, y_n} p(y_{n+1}, x_n, y_n) \times \log \frac{p(y_{n+1}, x_n, y_n)p(y_n)}{p(x_n, y_n)p(y_{n+1}, y_n)} \quad (7)$$

と表される。例として以下に示すような2進数で表される2ビットの時系列システム

システム X : 1110110100110011101101

システム Y : 1010101011011011011001

に対し移動エントロピーを求めると $TE_{Y \rightarrow X} = 0.044$ および $TE_{X \rightarrow Y} = 0.011$ が得られ、 $Y \rightarrow X$ の影響がその反対よりも大きいことがわかる。本研究ではシステム X とシステム Y をそれぞれ SNS 内のある特定のユーザとそのフォロワーであるとし、第3節で導入した感情概念ベクトルの8成分それぞれに対し移動エントロピーを求める。ただし、移動エントロピーは離散的な時系列データを扱う際に有効な方法であるため、そのままでは実数成分からなる感情概念ベクトルに適用することはできない。次節ではこの点を考慮し、実数データを離散データに変換する方法を考案し、移動エントロピーを適用可能にする。

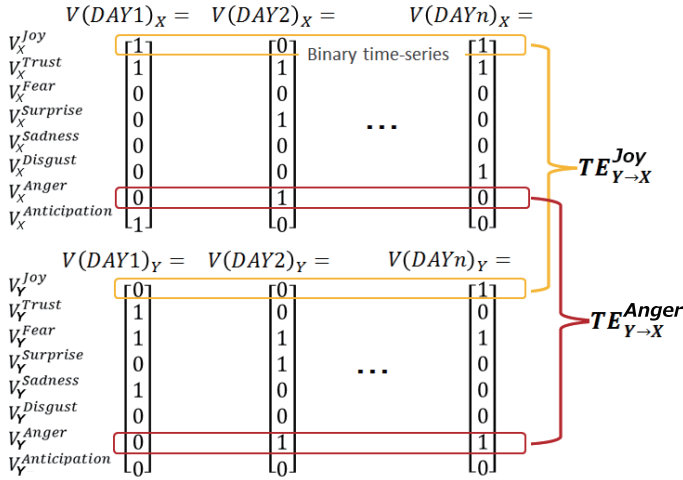


Fig.9 粗視化した感情表現ベクトル V_X, V_Y から移動エントロピー $TE_{Y \rightarrow X}$ を求める模式図

6. 感情概念の伝播

本研究は SNS におけるユーザ間の感情伝播を移動エントロピーを用いて定量的に捉えるとともに、ネットワーク内で感情が伝わる過程を可視化することで、伝播しやすい感情概念と伝播しにくい感情概念の存在を明らかにすることを目的としている。ユーザが投稿した 1 日分のツイートを形態素解析し、第 3 節で解説した感情語辞書を用いて感情概念を抽出、一日分のツイートから感情表現ベクトルを得る。ユーザ毎に 1 か月分に相当する 31 の感情概念ベクトルを時系列順に集め、感情概念ベクトルの 8 成分それぞれに対しユーザ・フォロワー間で移動エントロピーを求める。移動エントロピーを用いてユーザとフォロワー間の感情の伝播を可視化する。

まず、1 日分のツイートから作成した感情表現ベクトルの 8 成分のうち主要な 3 成分を 1、残り 5 成分を 0 に変換（粗視化）する。この変換により伝播源となるユーザの感情を支配的な 3 成分のみで表現するとともに、感情概念ベクトルを移動エントロピーの計算に必要な離散的な時系列データで表現する。3 成分で粗視化した明確な理由はないため、ここは 1 成分でも 2 成分でもよく、パラメータであるとする。続いて粗視化した感情表現ベクトルを 1 か月（31 日）分集め時系列順に並べることで、各ベクトル成分（感情概念）ごとに 31 ビットの時系列データ（2 進数列）を得る（Fig.9）。この 2 進数列を用いて、フォロー関係にあるユーザ間で 8 つある感情概念毎に第 5 節で解説した方法に沿って移動エントロピーを計算する。移動エントロピーには指向性があるので、順方向の $TE_{Y \rightarrow X}$ に加え逆方向の $TE_{X \rightarrow Y}$ についても計算する必要がある。得られた移動エントロピーの大小関係によりユーザ間を伝播する感情概念を捉え可視化をおこなう。

Fig.10 には感情概念ごとに移動エントロピーが最大値をとるノード（ユーザ）間のエッジを「伝播が起こっている」と判定し、色付きの太矢印で表示した。各ノードに対する配色は第 4 節のコミュニティ分割法により求めたコミュニティへの属性を表

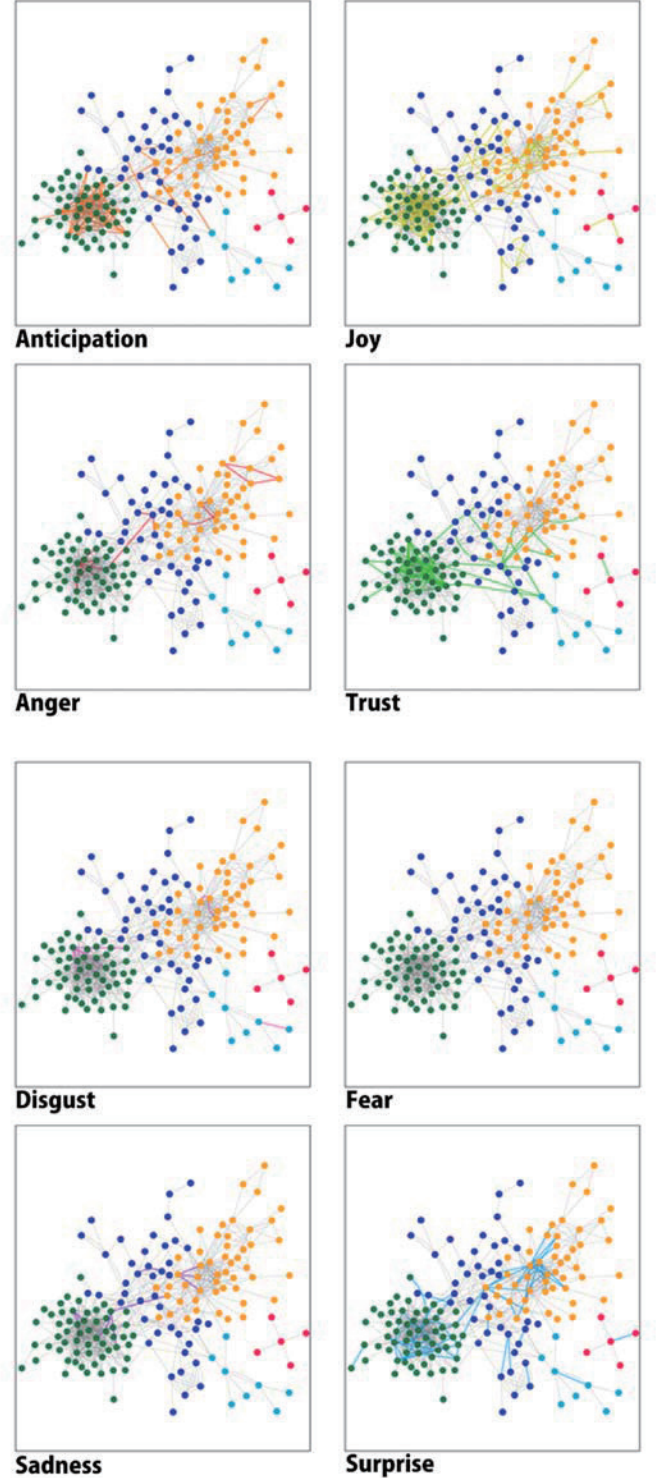


Fig.10

している。8種類の感情概念すべてに対し伝播のみられないリンクがあるが、これらは移動エントロピーの値が8成分全てに対し0の値をとるか、もしくはツイート収集期間で1度もツイートがなかった場合のどちらかに対応している。感情表現ベクトルの粗視化パラメータは3に設定しているが、我々のモデルではこのパラメータを1に選んでも2に選んでも伝播の起こっているエッジの空間パターンに大差はなく、結果に大きな差はみられない。

興味深いことにネットワーク全体では joy(喜び) と Surprise(驚き) が特に伝播しやすい傾向にあることがわかる。また、感情概念の中にはコミュニティ間を伝播しやすい性質を持つ成分と同一コミュニティ内に留まりやすい性質の成分があることが識別できる。コミュニティ分割を横断しネットワーク内を広く伝播している感情成分は Joy と Surprise であり、これら2成分はコミュニティ内でも活発な伝播がみられる。楽しい(面白い)あるいは驚き(素晴らしい)といった感情の伝播は共通の背景を必要とせずに伝播することができると考察できる。一方で局所的にコミュニティ内に留まりやすい伝播傾向を示す成分もあり、Trust と Anticipation にはこうした性質が顕著にみられる。信頼や期待といった感情は対象となる物事に対する共通の認識(話題や趣味嗜好といった価値観)を持っていなければ共感できないため、伝播はコミュニティ内に留まる傾向を示すと考察することができる。こうした局所的な感情の伝播は、その多くはコミュニティ内部で発生しているということがわかる。

7. まとめ・展望

本研究の目的は Twitter のユーザ同士の感情の伝播を観測・可視化する方法を提案するとともに、各感情概念にみられる伝播の特徴を明らかにすることである。ツイートが示す感情を Plutchik の8つの基本感情に基づいた8次元ベクトルで表現し、移動エントロピーを用いて Twitter ネットワーク上で感情の伝播を捉える方法を考案した。また、Reichardt らの焼きなまし法アルゴリズムを用いて Twitter ネットワークをコミュニティ分割し、コミュニティ内外を問わず広く伝播できる感情概念の存在とコミュニティ内に留まりやすい伝播傾向を示す感情概念の存在を明らかにした。

本研究で得られた感情概念の伝播に関する新たな知見には様々な利用価値が考えられるが、一つには情報の戦略的な配信であろう。joy(喜び) や Surprise(驚き) などの感情概念を普段から多く発するユーザはコミュニティ構造を越え「大きな声」で情報を配信することのできるエッジ構造を築いており、情報拡散や宣伝広告に利用できる高い付加価値がある。一方で Trust や Anticipation などの感情概念を普段から多く発するユーザは「ヒソヒソ声」で情報を配信することのできるエッジ構造を築いており、限定的な情報配信に利用できる高い付加価値がある。マーケティングの分野では情報をむやみに拡散させない戦略的な広告の配信が製品の価値を維持するために重要なことがしばしばあり、コミュニティ内での販売戦略に用いられれば、コミュニティ外での模造品の登場を回避することができるなど一定の成果を生み出すことが期待できる。

感情概念ベクトルの決定には一日分のツイートを用いているが、一日分のツイートをもってしてもユーザの感情概念が全成分で0になるケースが稀に起こり得る。ユーザからツイートが全く出ていない場合は感情概念ベクトルは0の値をとり問題はないが、ツイートが出ているにも関わらず、実際には感情が0であるケースが問題になる。この場合、全感情概念成分は小さい値ながらも0にならないため、感情概念ベクトルの正規化と粗視化という人為的な操作が、感情0の状態をかき消し、感情の発生を観測してしまうのである。しかし、感情伝播の有無は移動エントロピーを用いて決定しているため、人為的な感情の誘起が感情伝播に及ぼす影響は都合よく回避される。人為的な感情の誘起が感情伝播に及ぼす影響の大きさは、たかだかランダムな2進数列間で求めた移動エントロピーの大きさと同程度になり、2進数列が無限に長い極限では0に収束する。

Twitter ネットワーク内での感情伝播の研究における今後の課題は、Twitter ネットワークの分析と感情語辞書の改善という2種類に大別される。一つは Twitter ネットワーク上で感情を伝播させやすい傾向にあるユーザが持つ特徴を解明することである。実験結果からそうしたユーザの存在は明らかになったものの、共通する特徴を見出すことができなかった。媒介中心性や次数中心性といったノードが持つ性質に焦点を当てて分析する必要がある。

次に感情分析の精度向上に関する問題である。第6節で示したように感情ごとの伝播量には大きな差がある。感情語辞書に登録された感情概念別の単語数を調査したうえで単語数を充実させるなどの性能向上が必要となるだろう。また Twitter を対象とした感情分析では顔文字や新語、インターネット文化特有の表現の分析手法が大きな課題となる。例えば「笑い」を意味する「草生える」という新しい表現が生まれている。これに対して形態素解析を行なうと複数の形態素に分割されてしまう。こうした表現に対しては係り受けや文脈といった文章の構造を分析する構文解析を取り入れた感情分析の新たなアプローチ法を確立することが求められる。

人間の感情は言葉として表現され、言葉は変化し続ける。殊更、SNS において表現の変化は著しい。生まれ続ける新たな表現を収集し分析すると同時に、ネットワークそのものの性質を精確に捉えることができれば SNS におけるユーザの振る舞いについてより深い知見を得ることができよう。

謝辞

東京大学の鳥海不二夫先生には移動エントロピーの導入をすすめていただくなど貴重な助言を頂きました。小澤俊介氏には感情語辞書構築に関わるソースコードをご提供いただきました。心より感謝申し上げます。

参考文献

- 1) WordNet(最終閲覧 2016/10/1), Princeton University:
<http://wordnet.princeton.edu>

- 2) G.A. Miller, R.Beckwith, C. Fellbaum, D. Gross and K. Miller “Introduction to WordNet: An On-line Lexical Database”, Online Books Page (Revised in 1993)
- 3) Plutchik Robert, “Emotion: A psychoevolutionary synthesis” Harpercollins College Division, 1980., ISBN-13: 978-0060452353
- 4) T. Schreiber “Measuring Information Transfer ” Phys. Rev. Lett., **85**, pp.461, 2000.
- 5) 田口聖久, 三末和男, and 田中二郎. “大衆の豊かな感情の時間空間的な分布の可視化.” 情報処理学会研究報告. EC, エンタテインメントコンピューティング 2014.36 (2014): 1-8.
- 6) NAITO, KAZUHIRO, et al. “Twitter コメントに含まれる感情語がイベント印象に与える影響の評価.” 情報処理学会インタラクシオン (2012) pp.871-876.
- 7) Francis Bond, Timothy Baldwin, Richard Fothergill and Kiyotaka Uchimoto (2012) Japanese SemCor: A Sense-tagged Corpus of Japanese in The 6th International Conference of the Global WordNet Association (GWC-2012), Matsue.
- 8) 小澤俊介, skozawa’s blog, <http://skozawa.hatenablog.com/entry/2014/05/26/232429>
- 9) 中村明: 感情表現辞典, 東京堂出版, 1993, ISBN:9784490103394.
- 10) Natural Language Toolkit, <http://www.nltk.org/> (最終閲覧:2016/10/1)
- 11) Girvan, Michelle, and Mark EJ Newman “Community structure insocial and biological networks” Proceedings of the national academy of sciences 99.12 (2002): 7821-7826.
- 12) Pons, Pascal, and Matthieu Latapy. “Computing communities in large networks using random walks.” J. Graph Algorithms Appl. 10.2 (2006): 191-218.
- 13) Newman, Mark EJ. “Finding community structure in networks using the eigenvectors of matrices.” Phys. Rev. E 74.3 (2006): 036104.
- 14) Reichardt, Jrg, and Stefan Bornholdt. “Statistical mechanics of community detection.” Phys. Rev. E 74.1 (2006): 016110.
- 15) Zachary, Wayne W. “An information Flow Mode; for Conflict and Fission in Small Groups” Journal of anthropological research (1977): 452-473.
- 16) Newman, M. E. J.. “Fast algorithm for detecting community structure in networks” Phys. Rev. E **69**, 066133(2004)
- 17) 増田直紀 今野紀雄 「複雑ネットワーク 基礎から応用まで」 近代科学社 2010 年 ISBN-13: 978-4764903630
- 18) Newman, Mark EJ, and Michelle Girvan “Finding and evaluating community structure in networks” Phys. Rev. E 69.2 (2004): 026113.
- 19) Fortunato, Santo. “Community detection in graphs.” Phy. Rep. 486.3 (2010): 75-174.