

大衆の豊かな感情の時空間的な分布の可視化

田口聖久^{†1} 三末和男^{†2} 田中二郎^{†2}

大勢の人々が抱く豊かな感情の時空間的な変化の把握を支援することを目的に、我々は感情をテーマにした主題地図による可視化システムを開発した。感情の空間的広がりを可視化する従来の研究には肯定・否定の評判分析の結果を可視化したものがある。しかし、本来感情はより豊かに、より細かく分類できるものである。本研究では、その細かく分類された感情の主題地図を作成するための設計指針を2つのプロトタイプ製作によって導いた。例えば、細かく分類する感情の個数や、地理的広がりを提供する的方法などである。それら指針をもとに、豊かな感情の主題地図を設計し、計算機による描画システムを開発した。

Visualization for Spatiotemporal Distribution of People's Rich Emotions

KIYOHISA TAGUCHI^{†1} KAZUO MISUE^{†2} JIRO TANAKA^{†2}

To grasp spatiotemporal changes of rich emotions for a large number of people, we developed a visualization system with thematic maps of the emotions. Several conventional studies visualizing the spatiality of emotions as a result of sentiment analysis exist. In many of them, the emotions are classified into positive ones or negative ones. However, primary emotions can be classified more richly and more precisely. By creating two prototypes, we obtained design criteria for visualizing the multidimensional emotions and their spatiotemporal distribution (e.g. the number of emotions to be classified, how to visualize the spatiality of emotions etc.). Based on these criteria, we designed the thematic maps of rich emotions and we developed a system of drawing using a computer.

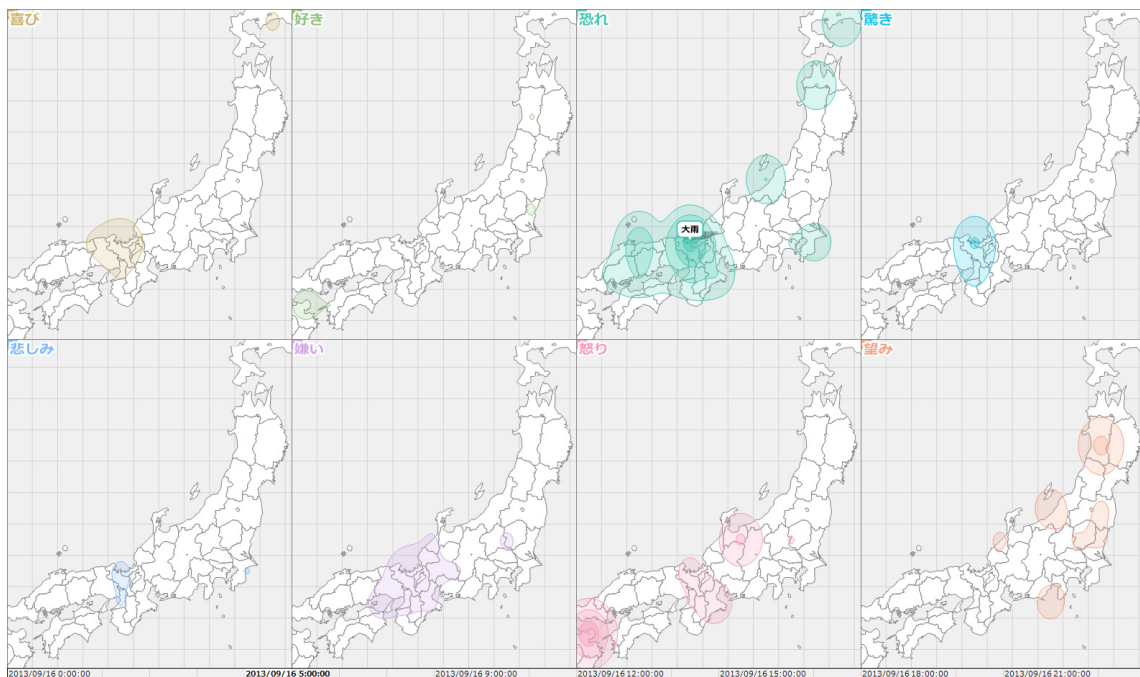


図1 「感情天気図」の可視化例 (2013年9月16日5時台の様子)

Figure 1 Visualization example of "Emotional Weather Map" (2013, Sept. 16th, 5:00 to 6:00).

1. はじめに

ソーシャルメディアの登場によって、個人がより容易に情報を発信できる環境が整いつつある。ソーシャルメディア

^{†1} 筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻
Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba

^{†2} 筑波大学システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba

は生活との親和性が高く、今感じている自分の意見や気持ちを気軽に投稿する場となっている。マーケティングへの応用や社会的指標の廉価な獲得を期待して、そのような投稿集合をもとに、製品に対する評判や生活満足度などを判定し、その地理的広がりを主題地図(主題図)によって表現する研究がある。主題地図とは、地形や地名などの地理的要素を中心のテーマとして表現した一般地図(一般図)に対して、例えば人口や交通量などの地理的でない要素を

中心に表現した地図のことであり、感情の主題地図を作成したこれら研究の多くは感情を快・不快、肯定・否定、または喜び・悲しみといった1次元で捉えるものであった。しかし、本来人間が抱く感情はより豊かであって多次元で表現されるものである。その豊かな感情が表現できれば、より身近でより具体的な知見が獲得できると考える。

そこで本研究は、大衆の豊かな感情の時空間的分布の把握を支援することを目的とする。そのために、多次元に分類した感情をテーマとした主題地図「感情天気図」の設計と開発を行った。この設計には2つのプロトタイプを製作することで得られた指針を採用している。また、実データを適用した感情天気図を観察することで、具体的な知見が得られるかどうかを調査した。図1は日本時間2013年9月16日の1日分のデータを適用した例である。5時台の空間的分布を表したもので、京都府の辺りで「恐れ」の感情の高まりが分かる。以降、本研究での時刻は日本時間を表す。

2. 関連研究

2.1 感情の分類方法と抽出方法

ある事象に対する快・不快の感情はそれぞれを肯定・否定、総じては評判と呼ばれる。感情を1次元に分類する評判に対して、感情を多次元に分類する試みもある。Izardは顔の表情を分類することで、基本となる感情を10個挙げている[1]。また、それを引き継ぎながら、Plutchikが基本感情を8つ挙げた[2]。その他にも多くの分類方法がある。

また、さまざまな情報源から感情を抽出する方法が感情分析として研究されている。自然言語からの感情抽出として、那須川らは製品に対する文章からその評判を判断する技術を開発した[3]。多次元の感情抽出方法としては、Izardの分類方法を用いた Neviarouskayaら[4]や菅原[5]による自然言語処理がある。また他にも、音声や顔の表情、生体信号などを情報源とした感情抽出が試みられている。

本研究は、感情を詳細に分類し抽出する技術よりも、むしろこれら技術によって抽出された感情情報の加工、また、その時空間的な分布の視覚的な提示に焦点を合わせている。抽出が適切に行われたとしても、複雑に分類された感情の時空間的变化やその相互作用を意味合いまで含めて解釈するのは計算機では未だに難しいという問題がある。本研究ではそれを適切な視覚的表現の設計をして、人間の認識能力を補助することで解決する。

2.2 大衆感情の主題地図の作成

Haoらは視覚的な評判分析として、評判の主題地図の作成を試みた[6]。ソーシャルメディアの1つであるTwitter[a]である製品に対する発言と地理情報を含む投稿を収集し、その発言が肯定的・中立的・否定的であるかを判定する。続いてそれぞれを緑色・灰色・赤色に対応付ける。視覚的提示として、地理情報に対応する地図上の地点にその色で

a) <http://twitter.com/>

点を描画する主題地図と、その評判の色で塗ったキーワードの文字列を地図上に描画する主題地図を持つ。Shookらはオンライン百科事典であるWikipediaの英語版[b]の記事から、地理情報と喜び・悲しみの感情情報を抽出し、地理的分布の視覚的提示を行った[7]。喜びを青、悲しみを赤に、その度合いを青から白を経て赤のグラデーションに対応付けて主題地図を描いた。また、Schwartzらはアメリカの郡単位での生活満足度を表す主題地図を作成し、分析結果の提示を行った[8]。地理情報を含むTwitterの投稿の文章などから生活満足度を算出し、その値が高いことを緑に、低いことを赤に、その度合いを緑から黄緑を経て赤のグラデーションに対応付けている。また、張らはウェブで公開されている新聞社の記事を対象に、Plutchikの感情分類方法を用いて感情を4次元で分類し、感情分析を行った結果を視覚提示するシステムを構築した[9]。記事に対応する新聞社の所在地付近に吹き出しを描画し、その中にラインチャートを描画することで感情の時間変化を提示している。

ここでは3つの問題点がある。1つ目の問題点は感情の分類方法である。人間が抱く感情は1次元で表される単純なものだけではない。本研究では、把握や分析ができる感情の分類の数を検討した。2つ目の問題点は感情の種類提示の方法である。本研究では多次元に分類した感情について、人間が把握したり分析を進めたりするのに易しい設計をした。3つ目は感情以外の情報の提示とその方法である。感情の分類のみを示すだけでは、そこに意味を見いだすのは難しい。そこで本研究は感情の時空間的な分布を分かりやすく提示することでより具体的な知見を得られることを示した。

3. プロトタイプ1

人々の感情とその分布を捉えるため、人・感情・時刻に注目した分析を行うことを考えた。そこで、素朴な座標系表現を持つプロトタイプ1の製作を行った。

3.1 可視化の流れ

3.1.1 生データの収集

対象メディアにはTwitterを選んだ。特に、投稿者が自身の言葉を含むだろうと考える通常の投稿(ツイート)・非公式リツイート[c]・メンション・リプライ[d]を対象とする。また、これらをまとめて投稿と呼ぶこととする。

投稿を収集する人々の対象には、筆者らが所属する研究室のメンバ67人で構成されるコミュニティを選んだ。具体的には、Twitterで上記メンバのリストを作成し、REST APIによってそのリストに含まれる者による投稿を収集する。生データは「投稿者・投稿時刻・投稿文章」からなる。

b) <http://en.wikipedia.org/>

c) 投稿者がもとの文章をコピーして自分の投稿とする引用型の投稿。もとの文章の前後に投稿者が加筆することもある。

d) あるツイートに対する言及のうち、文頭に誰か投稿者を表す記述のあるものはリプライ、そうでないものはメンションと呼ばれる。

3.1.2 提示データへの変換

プロトタイプ 1 での感情の分類は、概念辞書である WordNet の日本語版[10]を用いて構築する。WordNet は同義語の集合である概念の群とそれら同士の関係情報を持っている。この関係情報のうち、今回は下位関係と属性関係を用いる。下位関係とは当該概念が対象概念を包含する関係、属性関係は当該概念の属性を表す際に対象概念を用いる関係である。感情の種類を表す概念を「感情」を所属語に含む概念の下位関係にある 45 種類概念群とした。これらの概念それぞれで、その概念とその概念の下位・属性関係にある全ての概念に所属する語をその概念を表現する感情語とする。こうして感情の種類とそれを表す感情語の集合である感情語辞書を作成した。

感情の抽出には簡単な自然言語処理を行う。まず、投稿文章を形態素解析エンジン MeCab[e]によって単語に分け、活用があるものは基本形に修正したのち、感情語辞書との文字列照合を行い、その投稿が含む感情を導く。1 つの投稿で複数個の異なる感情概念を含む可能性がある。ただし、その感情の強さについては判定しない。また、慣用句や否定・疑問・皮肉といった句や文、文章単位での特徴についても考慮しない。

この段階で、生データは「投稿者・投稿時刻・感情の種類」からなる提示データへと変換される。

3.1.3 視覚的表現への対応付け

投稿者は、縦軸の位置に対応付ける。投稿者の並び順に特別な意味はない。

投稿時刻は横軸の位置に対応付ける。具体的には左端が最も古いもの、そこから右にいくにつれて新しいものに対応する。

感情の種類は HSV 色空間での H (色相) に対応させる。色空間とは色を立方的に定義する方法であり、HSV 色空間は色を H (色相)・S (彩度)・V (明度) で定義するものである。色に対応付けた理由は、他の表現より視覚的重なりを少なくでき、概観を捉えやすくなるためである。また、色相に対応付けた理由は、量を表現するには拙いがカテゴリの表現に適しているためである。具体的に感情は赤 (H=0, S=100, V=100) を基準として間隔を感情概念の個数である 45 個に等分した色相へと対応付ける。色相の対応付け順に特別な意味はない。また、感情を含まないものは半透明の白色とする。前述の位置に、全て同じ大きさの矩形を対応した色で塗りつぶす。ただし複数の感情を持つ場合は、その数で矩形を縦方向に等しく分割し、それぞれの色で塗る。

3.2 可視化例

図 2 にプロトタイプ 1 の可視化結果の一例を示す。対象は、2011 年 9 月 20 日午前 0 時から同年 10 月 7 日午前 0 時までの 17 日間の投稿約 5,000 個である。そのうちいずれか

の感情を含むものは約 500 個である。投稿者ごとに白の横線で、0 時を基準に 1 日ごとに白の縦線で区切りを入れた。投稿が比較的多くてもそのうち感情を含むものが少ない者がいることや、逆に投稿数は少なくとも感情を含む割合が高い者がいることが分かる。

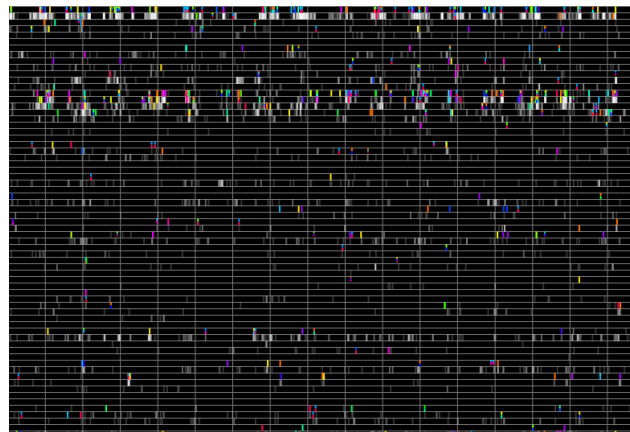


図 2 プロトタイプ 1 の可視化例

Figure 2 Visualization example of Prototype 1.

3.3 考察

対象メディアとして Twitter を選んだことは良かったと考える。67 人もの投稿者それぞれの意見・考え・気持ちがそう感じたときに吐露されているように思う。

続いて、感情の分類方法とその視覚的表現への対応付けについて述べる。作成した感情語辞書を用いることで、感情自体はある程度抽出できていると考えるが、それがどの感情かを把握するのはこの分類あるいはこの表現では難しかった。つまり、感情を分類する数が多すぎると認知的にも視覚的にも混雑さが高まり分析に支障が出る。しかし、分類数が少なすぎると人間の感情を説明するには不十分である。したがって、肯否よりも詳細な分類はしつつも、視覚的提示をした場合には明快さが求められるため、感情の分類を人間が把握するのに自然な数に、また視覚的に識別できる表現にすることが必要である。

4. プロトタイプ 2

プロトタイプ 1 で得られた知見を踏まえながら、豊かな感情の時空間的分布を捉えるため、それを地図上に展開するプロトタイプ 2 の作成に取り組んだ。

4.1 可視化の流れ

4.1.1 生データの収集

プロトタイプ 2 でも Twitter を対象メディアとする。ただし Twitter で公開されているものを全体として、そのうち無作為抽出された 1% のうち、さらに日本の地理情報を持ち、日本語である投稿とした。具体的には、日本の地域に限定する Streaming API によって得られる投稿を収集する。生データは「投稿文章・投稿時刻・緯度・経度」からなる。

e) <http://mecab.googlecode.com/>

4.1.2 提示データへの変換

プロトタイプ 2 での感情の分類は、Plutchik が提唱する基本感情 8 つを採用する。Plutchik は、全ての感情は基本感情「Joy・Acceptance・Fear・Surprise・Sadness・Disgust・Anger・Anticipation」が組み合わせられたものだと説明している。ここで和訳を「喜び・好き・恐れ・驚き・悲しみ・嫌い・怒り・望み」とする。「喜び」と対極をなす感情として「悲しみ」、また「喜び」に近い感情として「好き・望み」があり、全体として循環性があるなど、Plutchik は感情同士の関係構造についても言及している。基本感情の分布を把握することで、大きなまとまりとしての感情を捉えることができると考え、構造化されていなかったプロトタイプ 1 の感情 45 種類から、この感情 8 種類に改めた。

感情の抽出には簡単な自然言語処理を行う。プロトタイプ 2 では感情語辞書を簡易的に作成した(表 1)。プロトタイプ 1 と同様に、投稿文章を形態素解析したのち、感情語辞書との文字列照合を行い、どの感情を含むか判定する。

表 1 感情語辞書の一部
 Table 1 Part of the emotional dictionary.

| 感情の種類 | 感情語 |
|-------|-----------------------------|
| 喜び | 喜ぶ, 幸せ, 嬉しい, 愉快, 笑, |
| 好き | 愛, 恋しい, 可愛い, おめでとう, |
| 恐れ | 怖い, 怯える, 恐怖, 心配, ホラー, |
| 驚き | 驚く, 焦る, 驚愕, 仰天, ビックリ, |
| 悲しみ | 悲しい, 惜しい, 泣く, 不憫, 絶望, |
| 嫌い | 嫌, 憎む, 悩む, 苦しい, 軽蔑, |
| 怒り | 怒る, 叱る, 激怒, 我慢, むかつく, |
| 望み | 欲しい, 楽しみ, 希望, わくわく, |

ここで、可視化の前に、実データの感情別・時間別・空間別の分布を見てみる。ここで紹介するのは 2013 年 6 月 14 日午前 0 時から同年 7 月 14 日午前 0 時までの 30 日間のものである。

まず、感情別の投稿数の分布を表 2 に示す。「喜び」の感情を含むものが他の感情と比べて高いことが分かった。その理由として、この 30 日間「喜ばしい」ことが多かったとも考えられるが、次のような 3 つの要因からバイアスがかかっている可能性がある。まずは、生理的な要因である。人間の抱くあらゆる感情が抽出できたとして、それぞれ分類された感情の個数が一様とは限らない。続いてはメディアの要因である。感情は言葉や表情、手紙やソーシャルメディアなど、メディアを介して人に伝わる。実際にある感情を抱くのと、その感情がそれぞれのメディアによって表出されて観測可能になることにはギャップがあり、その違いによって特定の感情の抽出がしづらくなる可能性がある。最後に、感情抽出方法の要因である。今回の感情抽出の方法に限らず、現状の感情抽出方法では、各感情の抽出精度

を等しくすることは難しい。これら 3 つの要因によるバイアスを感情的バイアスと呼ぶこととする。絶対数の時空間的な分布以外にも、抽出できた個々の感情としての特徴量の時空間的な分布を把握したい必要があるはずである。したがって、この感情的バイアスを解消する方法も検討する。

表 2 感情別の投稿数の分布
 Table 2 Emotional distribution of tweets.

| | 個数 | 総数に対する割合 |
|--------------|-----------|----------|
| 投稿総数 | 6,800,080 | - |
| いずれかの感情を含むもの | 1,876,058 | 27.59 % |
| 「喜び」を含むもの | 1,067,192 | 15.69 % |
| 「好き」を含むもの | 597,812 | 8.79 % |
| 「恐れ」を含むもの | 45,215 | 0.66 % |
| 「驚き」を含むもの | 25,490 | 0.37 % |
| 「悲しみ」を含むもの | 127,592 | 1.88 % |
| 「嫌い」を含むもの | 83,927 | 1.23 % |
| 「怒り」を含むもの | 33,930 | 0.50 % |
| 「望み」を含むもの | 90,492 | 1.33 % |

続いて、時間別の投稿数の分布を図 3 に示す。図 3 は横軸が 1 時間ごとの時間帯、縦軸が感情を含むか否かを問わずに集積した投稿数を示すヒストグラムである。投稿数は、朝方には少なく、夜中に多いことが分かった。

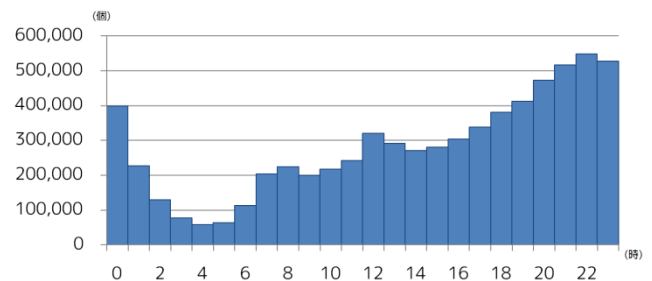


図 3 1 時間単位の投稿数の分布
 Figure 3 Temporal distribution of tweets.

最後に、空間別の投稿数の分布を図 4 に示す。図 4 は緯度経度 1 度ずつの領域に分け、その領域に該当する感情を含むか否かを問わずに集積した投稿数について、最大のもを基準に、その何%になるかを領域を塗りつぶすことで表現したものである。最大となるのは東京都心を含む領域で、1,830,080 個に相当する。都市部を含む領域は投稿数が多いことが分かった。図 4 の地形や県境などのデータはクリエイティブコモンズライセンス (CC BY-NC 2.1 JP) のもの [f] を利用しており、以降の主題地図にもこれを用いる。

f) <http://www.kabipan.com/geography/whitemap/>

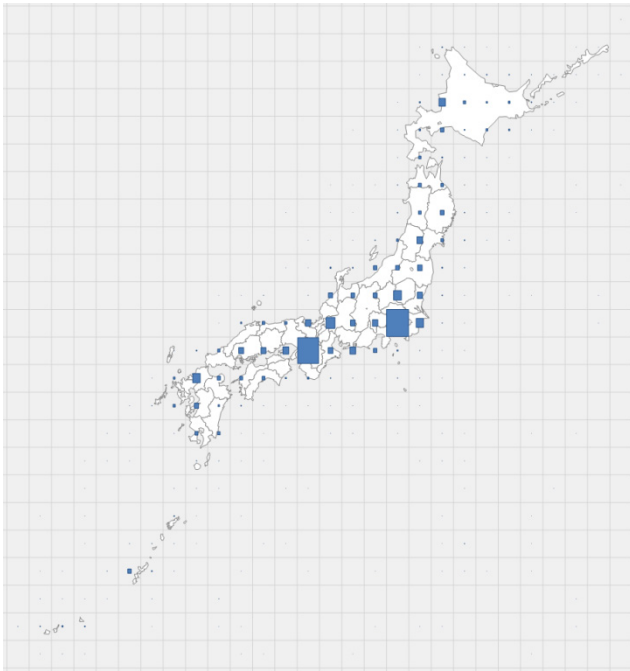


図 4 緯度経度 1 度の矩形単位の投稿数の分布

Figure 4 Spatial distribution of tweets.

投稿数が多ければ、抽出される感情の絶対数も大きくなることは想像に難くない。このバイアスを、時間別のは時間的バイアス、空間別のは空間的バイアスと呼ぶこととする。

以上を踏まえた上で、データセット内での、特定の時間帯・地理的領域で、特定の感情がどの程度特徴的であるかを表す感情スコアを算出する。

まず、計算を簡単にするため、また、視覚的に単純なものにするためのデータ変換を行う。全ての投稿について、(感情の種類・時間帯・地理的領域)で分割し、それぞれの個数を数え上げる。最大のもを基準として 0 から 1 までの値に正規化することで、基本となる感情スコアを導く。これは、前述した感情的・時間的・空間的バイアスがかかった値である。

続いて、バイアスの解消方法について述べる。まず、それぞれがかかっているバイアスの強さを算出する。

感情的バイアスの強さの求め方は次のとおりである。まず感情別の投稿を数え上げ、その最大値を基準として、それがその何%になっているか算出し、その逆数が強さとなる。

時間的バイアスの強さの求め方は次のとおりである。感情を含むか否かを問わず、時間帯別に投稿を数え上げ、その最大値を基準として、それがその何%になっているか算出し、その逆数が強さとなる。

空間的バイアスの強さの求め方は次のとおりである。感情を含むか否かを問わず、地理的領域別に投稿を数え上げ、その最大値を基準として、それがその何%になっているか算出し、その逆数が強さとなる。ここで、最大となる領域

の投稿数に対してその領域内の投稿数の割合が一定数未満となる領域を標本数が少なすぎる結果であるとしてバイアスの強さを 0 と改める。今回は、時間帯を 1 時間ごと、地理的領域を緯度経度 1 度の矩形ごとに分割する場合には、経験的に 1% 未満とした。

このように、それぞれバイアスがかかっている強さと投稿数が少なすぎる領域を算出したのち、解消したいバイアスを選択し、その強さと前述した基準となる感情スコアを乗算し、その最大値を基準に 0 から 1 までの値に再び正規化することで、バイアスが解消された感情スコアを導く。

最終的に、「(感情の種類・時間帯・地理的領域)の組とそれに対応する感情スコア」からなる提示するデータができあがる。

4.1.3 視覚的対応付け

時間帯は、閲覧者が操作によって切り替えることに対応付ける。

地理的領域は、地図上にそのまま描画することに対応付ける。ただし、1 つの感情に対し 1 枚の地図を描く。

感情の種類は、色と地図の位置で表す。具体的な配色を図 5 に示す。これは Plutchik の本[11]の表紙などで用いられている対応付けであるが、彼が感情と色との対応付けを行ったわけではない。しかし、顔を真っ赤にして怒る、ブルーな気分になるなど経験的に納得できる部分もある。また、Plutchik は感情と具体的な色の対応付けこそしていないが、感情の構造と色相環の構造とが似ていることには言及している。感情と色との対応付けの一部を基準に、残りを色相環に倣って配色すると、図 5 のようになるのは適当である。また、横長のディスプレイを想定し、画面を縦 2 つ横 4 つの 8 つに分け、そこに感情別の主題地図を描く。順序は、左上から右に、喜び・好き・恐れ・驚き、下段に移って右下から右に、悲しみ・嫌い・怒り・望みの順である。こうすると上下で対になる感情が、左右に似た感情が配置されることになる。



図 5 プロトタイプ 2 における感情と色の対応付け

Figure 5 Coloration in Prototype 2.

感情スコアの高さは、対象の地理的領域を塗りつぶす面積によって表す。

4.2 可視化例

図 6 にプロトタイプ 2 の可視化結果の一例を示す。日 2013 年 6 月 14 日午前 0 時から翌日午前 0 時までの 24 時間分について表示したものである。この例では時間帯の単位を午前 0 時を基準に 24 時間、地理的領域の単位を緯度 0 度経度 0 度を基準に 1 度ごとに区切った矩形とする。全てのデータが 1 つの単位時間内にあるために、時間的バイアスは考慮する必要が無い。

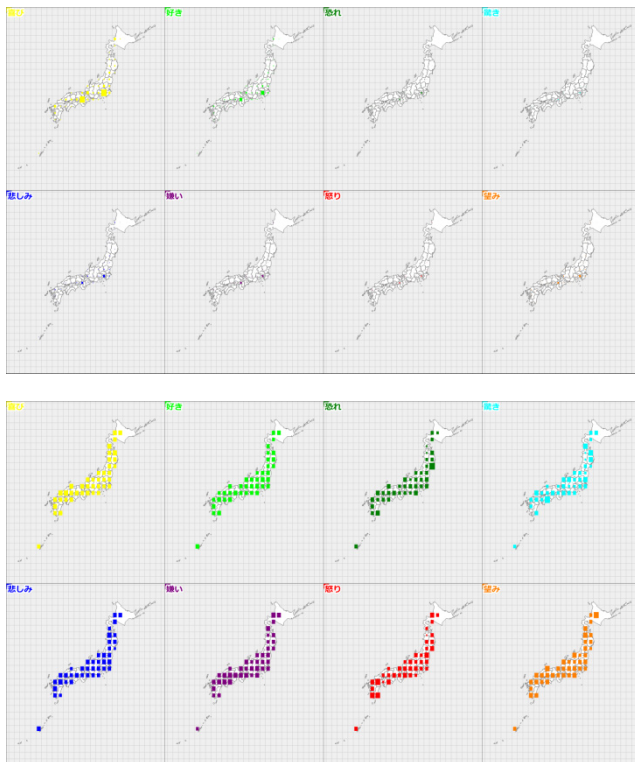


図 6 プロトタイプ 2 の可視化例

Figure 6 Visualization examples of Prototype 2.

図 6 上は全てのバイアスがかかったもので、観測できた感情を含む投稿数の絶対数を示しており、例えば「喜び」の投稿 1 つと「悲しみ」の投稿 1 つが等価である。このとき、黄色に対応する「喜び」の感情スコアが高く、また広く分布していることが分かる。また、それぞれの感情について都市部のスコアが高いことは分かる。図 6 下は感情的・空間的バイアスを解消したもので、感情別に、データセット内での割合としての空間的分布の高さを示している。「恐れ」の分布について宮城県の辺りが高いスコアであることが分かる。

4.3 考察

まず、バイアスの存在について考察する。プロトタイプ 2 製作の過程で感情的・時間的・空間的バイアスがある可能性に気づき、それらの解消をしたデータと解消しないデータとを可視化し比較した。すると、感情スコアの高さやその広がりについて得られる知見に大きな差があったので、

バイアスの解消には意味があると考えられる。ただし、それぞれバイアスを解消するのかもしれないかは閲覧者が感情をどう捉えるか、何を知りたいかと関係するため、それに応じて設定する必要がある。

続いて、感情の分類方法とその視覚的表現への対応付けについて述べる。45 種類から 8 種類に減らしたため、感情の種類と対応付ける色も 8 つになり、視覚的・認知的に捉えやすくなったことで、その意味の把握ができるようになった。しかし、配色には問題があり、用いた色の知覚的な明度が異なることで、量の差を読み取りづらくなっている。そこで、感情の種類と対応付けた色について、知覚的な明度をそろえる必要がある。

5. 感情天気図

2 つのプロトタイプを製作して得た知見をもとに、感情の時空間的分布を把握するための感情の主題地図「感情天気図」の設計および開発をした。

5.1 要件

プロトタイプ製作を通じて、感情天気図に求められる要件を整理した。

まず、多次元の感情を表現することである。豊かな感情を把握することで、1 次元の感情を表現する場合より具体的な知見獲得を狙う。

続いて、次の 3 点が把握できることである。

- ある時間、ある場所において、特定の感情の特徴がどの程度（数 / 割合）なのか
- 特定の感情が狭域へ集中しているのか、あるいは広域へ分布しているのか
- 特定の感情が短時間で発生しすぐに消滅したのか、あるいは長時間にわたって滞留したのか

さらに、喜びの共有や人々の悲しみを知ることなどは、データ分析のための情報可視化技術の知識の少ない一般の人々に求められるものであると考える。そこで感情天気図は、特別な知識を必要としない、できるだけ簡素で分かりやすい表現であることも重要だと考える。

5.2 可視化の流れ

5.2.1 生データの収集と提示データへの変換

生データの収集と提示データへの変換はプロトタイプ 2 と同じ手順で行う。ただし、各バイアスについて、解消するかどうかを GUI で切り替えることができるようにした。

5.2.2 視覚的対応付け

時間帯は、操作によって切り替えることで対応付ける。

地理的領域は、地図上にそのまま描画することで対応付ける。ただし、感情 1 つに対して 1 枚の地図を描く。

感情の種類は、色と地図の位置で表す。プロトタイプ 2 では HSV 色空間の色相に対応付けたが、感情天気図では L*a*b*色空間を用いた色相に改める（図 7）。L*a*b*色空間は、L*（明度）、a*（正が赤、負が緑の色味）、b*（正が

黄, 負が青の色味) の3つで色を定義するものである。これは、心理的な4原色に則っており、2色を表す座標の距離が実際の知覚的距離と等しく対応付けられている。これを採用した理由は、Plutchik が定義した感情同士の関係を知覚的により近く表現できるようにするため、また、知覚的な明度が等しい色を採用することで、量の表現をできるだけ歪ませないためである。さらに、これはプロトタイプ2で用いた色相とよく対応している。具体的には、喜び・好き・恐れ...・望みを0・1・2...・7としてnで表すとすると、次の式で定義される色に対応付けた。

$$L^* = 74, a^* = 40 \times \sin\left(-\frac{2\pi}{8} \times n\right), b^* = 40 \times \cos\left(\frac{2\pi}{8} \times n\right)$$

$L^*a^*b^*$ 色空間は、値によっては一般的なディスプレイでは正しく表示できない色領域を持つ。それを避けつつ「喜び」を黄色で表すことを基準にして、等しい明度で色差が最大となる値を選んだ。



図7 感情天気図における感情と色の対応付け
 Figure 7 Coloration in Emotional Weather Map.

感情スコアの高さは、対象領域を塗りつぶす面積に対応付けると他の領域とのスコアの差がはっきりと分かるが、周囲の領域との接続がはなれ地理的な広がりが見づらい(図8左)。そこで、2次元メタボールが与える影響力を段階的に描画することによってスコアの高さを段階的に表示する(図8右)。スコアと領域が近いものを緩くつなげることによって、感情スコアの高さの地理的広がりを大まかに把握できることを狙った表現である。

描画までの具体的な手順について説明する。はじめに、地図と重なる影響力場に、地理的領域1つについて1つのメタボールを追加していく。影響範囲は全てのメタボールに共通で1.5度の円とする。最大影響力を該当する感情スコアの高さとし、距離が遠いほど受ける影響力を小さくした。続いて、このように与えた影響力場に対して、それぞれ1/5・2/5・3/5・4/5の4段階の閾値で境界を作ると、影響力が1/5~2/5・2/5~3/5・3/5~4/5・4/5~となる領域が導かれる。最後に、それぞれを不透明度1/5・2/5・3/5・4/5で塗りつぶし、境界線は不透明度100%の色で描画することで閲覧者に提示する主題地図が得られる。

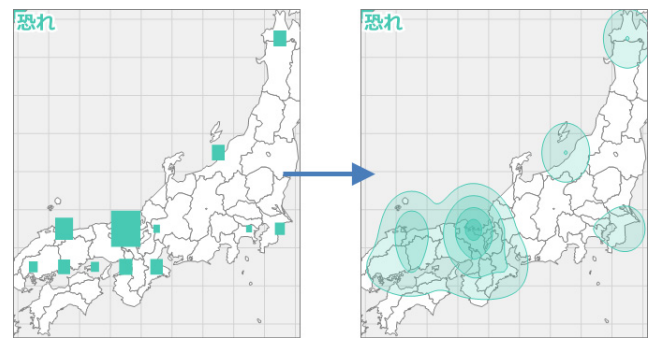


図8 感情スコアの提示

Figure 8 How to visualize emotional scores.

5.3 補助的機能

実際の利用において、前述の視覚提示のみでは不十分な部分がある。ここでは、それを補うために導入した機能について紹介する。

5.3.1 トピック表示

感情の時間的・空間的分布が把握できれば、その原因が知りたいだろう。それをトピックとして、今回は試験として簡易的に示すことにする。

メタボールレンダリングの結果4/5以上を示す領域で、ハッシュタグ[[g](#)]と感情語でない名詞の集合のうち、頻度が最も高いもの1つをその領域のトピックと定義し、吹き出しでその領域を代表する場所に重畳表示する(図9)。

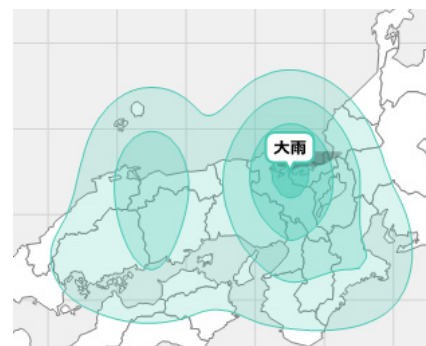


図9 トピックの表示

Figure 9 How to visualize topics.

5.3.2 投稿の詳細ウィンドウ

右クリックした場所が該当する(感情の種類・時間帯・地理的領域)の組に対応する投稿集合を新しいウィンドウに表示する。この機能によって、実際に投稿された文章についても確認できるようにした。

6. 適応例

感情天気図の有用性を示すために適用例を示す。今回は試用として、感情が誘発されそうな出来事があった期間の実データを選んで適用した。

g) Twitter において、投稿文章中で話題を明確に示すための特別な表記。

ここで紹介する実データは2013年8月4日午前0時から翌日5日午前0時までの24時間分である。投稿総数は217,661個であった。この日は、宮城県沖と震源地とする地震があった。生データから提示データへの変換は、時間帯の単位を午前0時を基準に1時間、地理的領域の単位を緯度0度経度0度を基準に緯度経度1度ずつの矩形とする。

感情別に、投稿総数に対する割合としての感情の高まりを把握したいとする。この場合、感情的・時間的・空間的バイアス3つ全てを解消した値を把握することが対応する。

図10は上が12時台、下が13時台を表した感情天気図である。上図から、12時台では「恐れ」と「驚き」の感情が東北地方、特に震源地に近い宮城県あたりで高いことが分かる。またこの2つの図から、「恐れ」の感情が2時間にわたって高く広く分布しているのに比べて、「驚き」の感情の高まりは1時間限りでおさまったことが見てとれる。また、上図の「恐れ」の主題地図上には「津波」のトピック表示が見られる。これによって、「恐れ」の主な原因は地震によって引き起こされる津波であると推測できる。

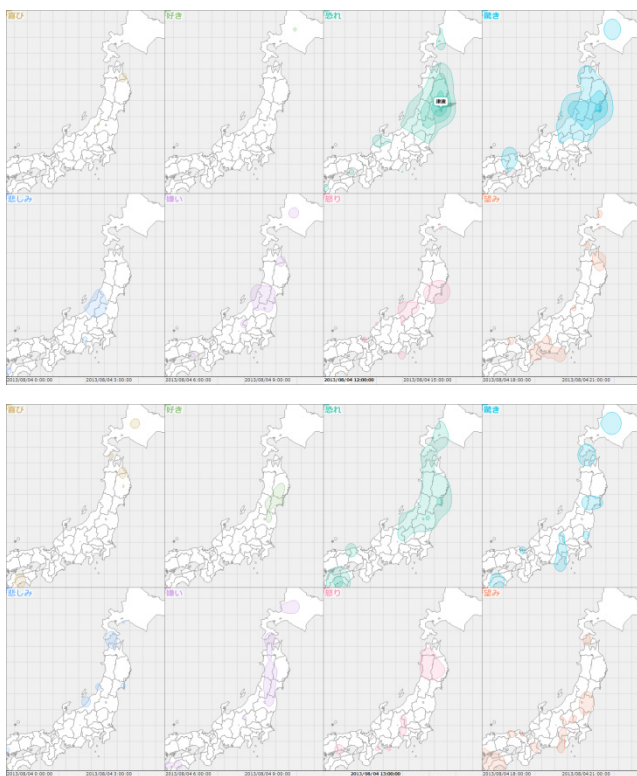


図10 感情天気図の適用例

Figure 10 Examples of Emotional Weather Map.

ただし、「津波」のトピック表示のある部分の投稿の詳細を見てみると、地震情報を知らせるボットの投稿も多いことが分かった。人が抱く感情を扱いたいとき、今回のようにTwitterでそれを集める場合には、それが個人による投稿であるかどうかを判断する技術が必要である。

7. おわりに

本論文では、大衆が抱く豊かな感情の時空間的分布の把握を支援するために、2つのプロトタイプ製作で得られた指針をもとにした「感情天気図」の設計を行った。また、実際に感情天気図を自動的に描画するシステムを開発し、実データの適用で具体的知見が得られることを確認した。

提案した方法によるバイアスを解消した感情スコアは、あくまでデータセット内での特徴量である。したがって、その投稿数が多い「喜び」の感情は、「喜ばしい」ことが少し増えただけではその事実がなかなか観測できない。また、感情の時空間的变化がある部分のみからなるデータセットではその変化を読みとれない。「喜び」の投稿数がいつも高めになることを踏まえた新たな普遍状態を定義し、その差を表現するバイアス解消方法の開発が課題である。また、8つの感情とそのスコアを1枚の地図でも混雑せずに表現する方法や、よりよい時間変化の提示方法の開発も望まれる。

また、本研究はデータ変換と視覚的表現への対応付けに焦点を合わせており、感情の抽出方法は非常に簡易的な自然言語処理を用いている。展望として、より精度の高い感情抽出方法によるデータの適用が望まれる。また、本来感情の高まりは誰でもいつでもどこでも発生あるいは表出するものである。将来的にそれらを精細に抽出する方法が開発され、またその情報を共有する基盤が構築されることによって、真に「人が抱く感情を表した世界の感情天気図」ができあがることに期待する。

参考文献

- 1) Izard, C. E.: The Face of Emotion, Appleton-Century-Crofts, 1971.
- 2) Plutchik, R.: Emotion: A Psychoevolutionary Synthesis, Harper & Row, 1980.
- 3) Nasukawa, T. and Yi, J.: Sentiment Analysis: Capturing Favorability Using Natural Language Processing, K-CAP'03, pp.70-77, 2003.
- 4) Neviarouskaya, A. et al.: Textual Affect Sensing for Sociable and Expressive Online Communication, ACHI'07, pp.220-231, 2007.
- 5) 菅原 久嗣, Neviarouskaya, A., 石塚 満: 日本語テキストからの感情抽出, The 23rd Annual Conference of the Japanese Society for Artificial Intelligence, 2009.
- 6) Hao, M. C., Rohrdantz, C., Janetzko, H., et al.: Visual sentiment analysis of customer feedback streams using geo-temporal term associations, Information Visualization Volume 12, pp.273-290, 2013.
- 7) Shook, E., Leetaru, K., Cao, G., et al.: Happy or not: Generating topic-based emotional heatmaps for Culturomics using CyberGIS, 2012 IEEE 8th International Conference on E-Science, pp.1-6, 2012.
- 8) Schwartz, H. A., Eichstaedt, J., Lucas, R. et al.: Characterizing Geographic Variation in Well-Being using Tweets, ICWSM 2013, 2013.
- 9) 張 建偉, 河合 由起子, 熊本 忠彦, 田中 克己: 地域性に基づく発信者の観差を可視化するセンチメントマップシステムの提案, 情報処理学会論文誌データベース Vol. 3 No. 1 (TOD 45), pp. 38-48, 2010.
- 10) Isahara, H., et al.: Development of Japanese WordNet, Language Resources and Evaluation Conference 2008, Marrakech, 2008.
- 11) Robert Plutchik: Emotions and Life: Perspectives from Psychology, Biology, and Evolution, American Psychological Association, 2002.