

筑波大学大学院 博士課程

システム情報工学研究科 修士論文

大衆感情の時空間的分布の 可視化に関する研究

田口 聖久

修士 (工学)

(コンピュータサイエンス専攻)

指導教員 三末和男

2014年3月

概要

現在まで、大勢の人々の感情を集計した情報がマーケティングなどで活用されてきた。本研究では、その大勢の人々の感情について、その時空間的変化の把握を支援することを目的とする。その目的を達成するために、感情をテーマにした主題地図を設計し、計算機による自動生成を試みた。

従来の感情の可視化を目指した研究には、肯定・否定の評判分析を地図上に可視化したものはある。しかし、人間の感情はより豊かなものだと考え、本研究ではその豊かさを大衆の発言を 8 つの感情に分類することで表現した。それらに分類された発言の絶対数だけでなく、表出された、また、観測された特徴量としての感情スコアの算出方法を提案した。また、その感情の時空間的分布の識別に配慮した可視化設計をした。

それらの結果として、その変化や相互作用、およびその意味合いの解釈をする手助けになることを示した。特に、特徴量の算出方法を適用するかどうかで得られる知見が異なることが明らかになり、また、視覚的に誤解を与えることや視覚的に複雑化することを避けた可視化設計によって情報可視化技術に詳しくない人でも容易に大衆感情の時空間的変化の把握ができるようになった。

目次

第1章	はじめに.....	1
1.1	感情情報と大衆感情.....	1
1.2	人の地理情報.....	1
1.3	情報可視化技術としての主題地図.....	1
1.4	研究の着想.....	2
1.5	本研究の目的.....	2
1.6	研究の流れ.....	3
1.7	本論文の構成.....	3
第2章	関連研究.....	5
2.1	主題地図の作成.....	5
2.2	感情の分類方法と抽出方法.....	5
2.3	感情情報の可視化.....	6
第3章	可視化の手順.....	8
第4章	予備実験1.....	10
4.1	予備実験1の目標.....	10
4.2	可視化の手順.....	10
4.3	可視化例.....	14
4.4	考察.....	15
第5章	予備実験2.....	17
5.1	予備実験2の目標.....	17
5.2	可視化の手順.....	17
5.3	可視化例.....	26
5.4	考察.....	32
第6章	感情天気図.....	33
6.1	3つの分析方法.....	33
6.1.1	時間分布の分析.....	35
6.1.2	空間分布の分析.....	36
6.1.3	時空間分布の分析.....	38
6.2	補助的機能.....	39
6.2.1	感情天気記号の提示.....	39
6.2.2	トピックの提示.....	39

6.2.3	投稿の詳細ウィンドウ	40
6.3	実装.....	41
第7章	適用例.....	42
7.1	前処理.....	42
7.2	事例 1:地震.....	42
7.2.1	生データと提示データへの加工.....	42
7.2.2	可視化結果の観察と得られた知見.....	42
7.3	事例 2:高校野球大会.....	51
7.3.1	生データと提示データへの加工.....	51
7.3.2	可視化結果の観察と得られた知見.....	51
第8章	まとめ.....	57
謝辞	59
参考文献	60

図目次

図 1 「感情天気図」を適用した例.....	4
図 2 情報可視化の 4 つの手順.....	8
図 3 WordNet による感情の分類.....	12
図 4 予備実験 1 の可視化の図例.....	13
図 5 予備実験 1 での感情と色との対応付け.....	14
図 6 予備実験 1 の可視化例.....	15
図 7 対象となる地理的な範囲.....	18
図 8 Plutchik による基本感情の環.....	19
図 9 1 時間単位の投稿数の分布.....	21
図 10 1 日単位の投稿数の分布.....	21
図 11 緯度経度 1 度の矩形領域を単位とする投稿数の分布.....	22
図 12 予備実験 2 における感情と色の対応付け.....	25
図 13 全てのバイアスがかかったもの（予備実験 2）.....	27
図 14 感情的バイアスのみを解消したもの（予備実験 2）.....	28
図 15 空間的バイアスのみを解消したもの（予備実験 2）.....	29
図 16 全てのバイアスを解消したもの（予備実験 2）.....	30
図 17 感情的バイアスのみを解消したもの（2013 年 6 月 14 日午前 4 時 / 予備実験 2）	31
図 18 全てのバイアスを解消したもの（2013 年 6 月 14 日午前 4 時 / 予備実験 2）...31	31
図 19 感情天気図における感情と色の対応付け.....	34
図 20 L*a*b*色空間において一般的ディスプレイが表示可能な色領域.....	35
図 21 ヒストグラムの例.....	35
図 22 感情スコアの表示.....	36
図 23 影響力と距離の関係.....	37
図 24 感情天気記号の表示.....	39
図 25 トピックの表示.....	40
図 26 投稿の詳細ウィンドウ.....	41
図 27 事例 1 の全てのバイアスがかかった時間分布ビュー.....	44
図 28 事例 1 の全てのバイアスがかかった空間分布ビュー.....	45
図 29 事例 1 の全てのバイアスがかかった時空間分布ビュー（午後 0 時台）.....	46
図 30 事例 1 の全てのバイアスを解消した時間分布ビュー.....	47

図 31 事例 1 の全てのバイアスを解消した空間分布ビュー	48
図 32 事例 1 の全てのバイアスを解消した時空間分布ビュー（午後 0 時台）	49
図 33 事例 1 の全てのバイアスを解消した時空間分布ビュー（午後 1 時台）	50
図 34 事例 2 の時間分布ビュー	52
図 35 事例 2 の投稿の詳細ウィンドウ	53
図 36 事例 2 の空間分布ビュー	54
図 37 事例 2 の時空間分布ビュー（午後 2 時台）	55
図 38 事例 2 の時空間分布ビュー（午後 11 時台）	56

表目次

表 1	予備実験 1 の感情語辞書.....	12
表 2	予備実験 2 の感情語辞書.....	19
表 3	実データの感情の分類の様子.....	20
表 4	予備実験 2 における感情と色の対応付け.....	25
表 5	感情天気図における感情と色の対応付け.....	34

第1章 はじめに

1.1 感情情報と大衆感情

感情情報とは、ある人が抱く気持ちを取り出したものである。また、大衆感情とは、大勢の人々の感情情報を総合したもののことである。大衆感情を知ることは、さまざまな分野で利用されている。

そのひとつに、意思決定やコミュニケーションなどの行動のきっかけになることが挙げられる。例えば大勢の人が困っている状況を知ることができれば、その悩みを解決してあげたいと思うだろう。2011年3月、震災によって、福島県を始めとして多くの地域の人々が大きな被害にあい、各地の人々から物資や義援金がおくられた。この支援行動は、物資が不足している事実のもとより、それで困っている人がいることこそが誘発させた行動であると考えられる。

また、感情分析という分野で、人々の感情を捉えようとする研究が行われており、主にマーケティングに応用できるとされている。例えばあるホラー映画の市場評価として、それが狙った恐怖感を与えることができたかを検証することができる。また、世論の調査として、大災害のショックから生じた自粛ムードがどのタイミングを回復したと見て広報活動を再開させるのかを決定する指標となりうる。

1.2 人の地理情報

人の地理情報とは、ある人がどこにいるのか、という情報である。私は、ある人物がある場所にいる、ということ自体に重要な意味があると考えられる。地理情報は、その人物の性質に深い関わりがある。どの場所にもそれぞれの文化があり、住人としてならばその文化そのものを、訪問者ならば自分が持つ文化との差異をその場で表現するだろう。

1.3 情報可視化技術としての主題地図

大規模なデータの中の特異点やデータ同士の相関を見つけ出すといったタスクは、計算機が得意とするものである。一方で、その意味合いや相互作用を考察するのは、人間のほうが得意なことである。そこで、計算機と人間とをつなぐ技術が必要になるが、その1つに情報可視化が挙げられる。情報可視化は、人間にとっては複雑な情報を、計算機が的確で理解しやすい視覚的表現に置き換えて提示することで、人間がその意味合いを把握す

るのを支援する技術である。

主題地図の作成は、情報可視化の1つと言える。主題地図とは、地形や地名など地理学的情報だけを記した地図ではなく、たとえば気象や交通情報など、あるテーマを重点的に表した地図のことである。主題地図は、そのテーマについての空間的分布を把握するのに役立っている。

1.4 研究の着想

世界中の人々がどんな気持ちでいるのかを知りたいというのが初めの着想であった。そのために、感情が世界中をどのように分布しているのかを把握する主題地図「感情天気図」が作成できないかと考えた。どのようなところの人々が、どのような気持ちでいるのかを知ること、喜びを一緒に共有したり、悲しんでいるところを励ましたり、怒っているところを落ち着かせたりするはたらきかけのきっかけができる。

感情の主題地図を作成する研究はいくつかあるが、それらの多くは感情を肯定・否定に分類するものである[1], [2], [3]。人間はより豊かな感情を持っているものであり、その把握がしたいと考える。詳細に分類した感情の主題地図を作成する研究[4]もあるが、それが提供するの主要な地点での時間的分布のみで、地理的な広がり概観を提供しない。感情それぞれの地理的な分布を提示することで、その存在がより具体的により身近に感じられると考える。

このような感情の主題地図に求められている要件は、次の3点であると考えられる。

1. ある時間帯、ある場所において、特定の感情が特徴に値するかどうか、またその特徴がどの程度なのかを把握できる
2. 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を把握できる
3. 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を把握できる

また、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。

1.5 本研究の目的

本研究は、大衆感情の時空間的変化の把握を支援することを目的とする。すなわち、ある人がある場所、ある時点でどんな感情を抱いたのか、という情報を多く集め、それが大

きなまとまりとして空間的にどう分布しているのか、その分布が時間経過によってどう変化していくのかという把握を容易にできるようにすることを目指す。そのために、感情をテーマとした主題地図の設計とその主題地図の計算機による自動生成を試みた。本研究は、感情を詳細に分類し抽出する技術よりも、むしろそ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。

1.6 研究の流れ

本節では、筆者がとった研究の進め方について説明する。本研究は初めから感情の主題地図の作成にとりかかったのではなく、人々の気持ちを把握する方法を模索しながら進めていった。

まず始めに、筆者が所属する研究室のメンバの気持ちの変化を把握することを目標に、予備実験を行った。ここでは感情の分類および抽出方法とその視覚的提示の検討をした。結果として、感情を分類した種類が多すぎると視覚的にも認知的にも分析に支障が出るという知見を得た。次に、日本全体の気持ちの変化を把握することを目標に、2番目の予備実験を行った。ここでは感情の分類の見直しと、データ変換によって得られる知見の違いを考察した。2つの予備実験で得た設計指針を踏まえながら、最終的に図1のような大衆感情の時空間的分布を示す「感情天気図」を設計および開発した。図1は2013年9月16日の1日分のデータに対して、5時台の空間的分布を表したものであり、京都府の辺りで「恐れ」の感情の高まりが分かる。このように、研究として最後に、「感情天気図」に実データを適用し、具体的な知見を得られることを確認した。

1.7 本論文の構成

第2章で、本研究に関連のある研究や既存のサービスについて示す。それ以降行方情報可視化の説明をより分かりやすくするために、次の第3章で情報可視化の手順について紹介する。第4章で1番目の予備実験について、第5章で2番目の予備実験について示す。続く第6章では、2つの予備実験で得られた知見を踏まえて設計した感情の主題地図である「感情天気図」について示す。第7章では実データを適用した「感情天気図」について示す。最後に第8章で本研究のまとめをする。

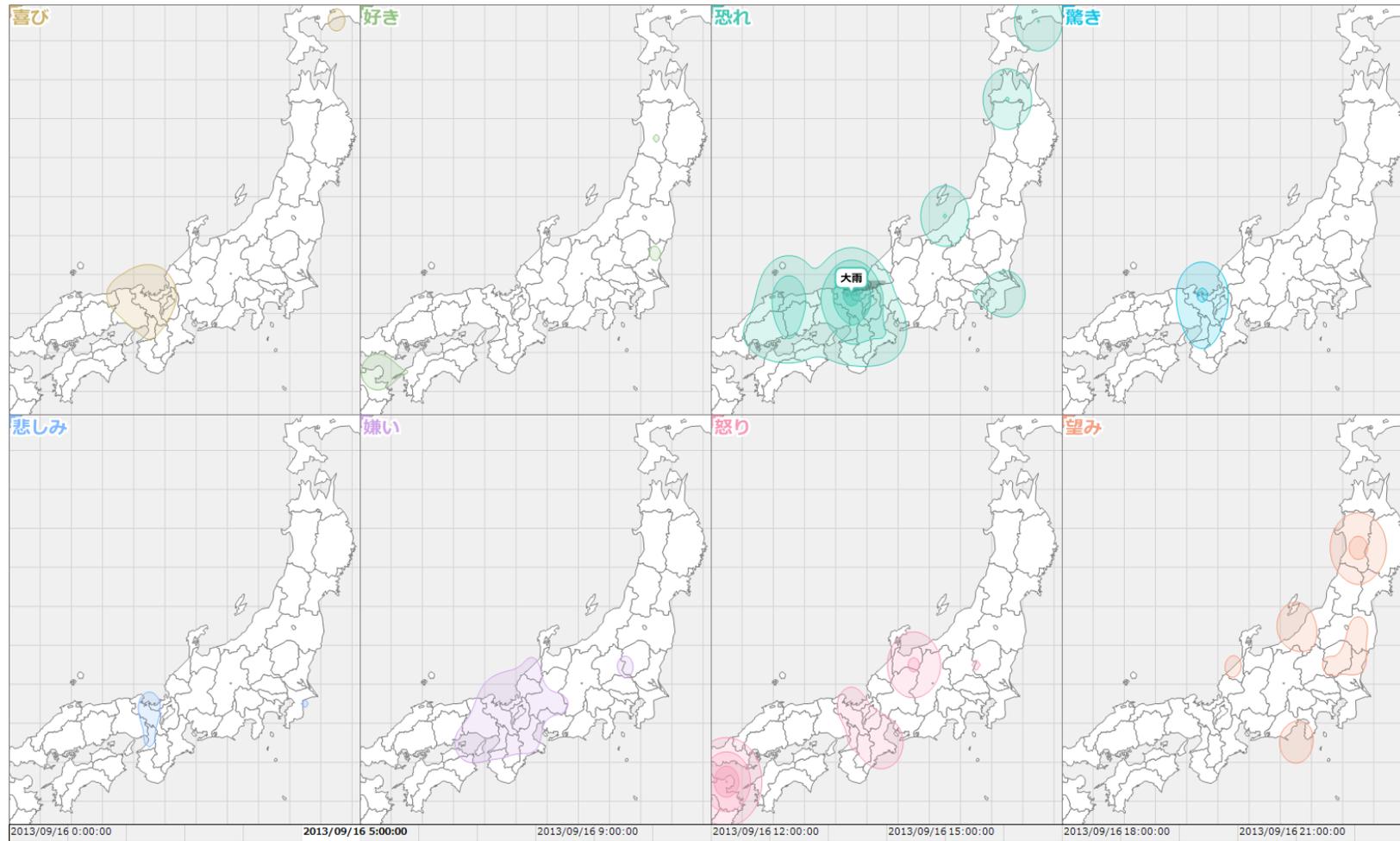


図1 「感情天気図」を適用した例

第2章 関連研究

本章では、本研究に関連する研究や既存サービスについて述べる。本研究はソーシャルメディアの発言から感情を抽出し、そのデータを地図上に展開する主題地図を作成手法を構築した。関連するものとしては、主題地図を作成するもの、感情を分類する、あるいはその分類に基づいて感情を抽出するもの、感情情報を可視化するものが挙げられる。

2.1 主題地図の作成

大衆から発信された情報をもとにして主題地図を作成するサービスがいくつかある。高橋らは、ソーシャルメディアの1つである Twitter[5]上での花粉症の症状を訴える発言を収集して、花粉症被害の主題地図を作成した[6]。ニフティ株式会社より実際にそれを「みんなの花粉症なう！β」[7]という Web サービスとして提供した。また、株式会社ウェザーニューズによる「SAKULiVE」[8]では、桜の様子を投稿をもとにして、桜の開花情報の主題地図を作成している。

本研究でも、大衆が発信する情報をもとにして主題地図を作成するが、テーマとして抽象的で、分類の種類もさまざまある「感情」というものを扱う部分に難しさがある。

2.2 感情の分類方法と抽出方法

前述したとおり、感情にはさまざまな定義や分類の方法が存在する。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。しかし人間はより豊かな感情を持ち合わせており、さらに詳しい分類をする試みも多くある。言語学では Davitz が、同義語辞書から感情語辞書を作成しそれを分類する形で、感情を 12 種類のまとまりに分類した[9]。Izard は顔の表情を分類する形で、基本となる感情を 10 個挙げている[10]。また、それを引き継ぎながら、感情の心理進化説の立場から、Plutchik が基本感情を 8 つ挙げた[11]。

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の研究として行われている。那須川らは、製品に対するある文章が肯定的か否定的かという評判を判断する自然言語処理の技術を開発した[12]。さらに詳細な分類での感情抽出方法として、Neviarouskaya らによるオンライン上での会話テキストを対象にした自然言語処理がある[13]。そこでの感情は、Izard の分類を用いて分類される。この研究をもとに、菅原も日本語を対象に自然言語処理による感情抽出を試みた[14]。またほかにも、音声[15]や顔の表情[16]、生体信号[17]

など、それぞれをもとにする感情抽出が試みられている。どれも、それぞれその分野にあった、あるいは分析の目的に応じた感情の分類方法を採用している。

本研究は、感情の分類方法およびその抽出方法の発展を目的としていない。ただ、視覚的に分析をする際に有利となる感情の分類については検討している。その分類を採用しさえすれば、抽出方法は問わない。問題となるのは、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現状である。本研究ではそれを、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。

2.3 感情情報の可視化

Yi らは、大規模テキスト分析のプラットフォームである **WebFountain** を対象に、評判分析とその結果の視覚提示を行った[18]。複数の製品とそれらの評価情報を、肯定か否定かの評判を識別し、製品カテゴリを縦軸、肯定や否定の多さを横軸にした、積み重ねバーチャートを用いて提示している。また、Mishne らの **MoodViews**[19]は、**LiveJournal** というブログサービスを対象として感情の視覚提示を行った例の1つである。**LiveJournal** では、投稿時に投稿者が「**Mood** 情報」を付加することができる。その **Mood** 情報はデフォルトで 132 種類用意されており、また投稿者自身で記入することもできる。**MoodViews** はこの情報を収集し、感情 1 種類ごとに時間変化のラインチャートを提示することで投稿者集合の感情変化を分析、追跡、提示するツールである。さらに、青島らは **Twitter** での投稿を対象に、**Plutchik** の感情の分類方法を用いて 1 ピクセルに 1 つの投稿を対応させる視覚提示を行い、60 万画素からなる画面において 60 万の感情を概観できるとした[20]。バーチャートによってそれぞれの感情の集積量を示す分析補助のためのビューも持ち合わせている。

感情情報の分析の結果の提示のために主題地図を作成する研究がある。**Hao** らは視覚的な評判分析を試みた[1]。まず、ある **Twitter** 上の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、否定的、あるいは中立的な意見かという肯否判定を行う。続いて、肯定・中立・否定をそれぞれ緑色・灰色・赤色に対応付ける。最終的に、分析者への視覚的提示として、地理情報に対応する地図上の地点にその色で点を描画するビューと、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ビューを持つ。**Shook** らは、英語版の **Wikipedia** の記事から、地理情報と喜びと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[2]。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いる。また、**Schwartz** らは、アメリカの郡単位での生活満足度の主題地図を作成した[3]。地理情報を含む **Twitter** の投稿の文章から生

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次元とし、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[4]。地図上に吹き出しを描画し、さらにその上にラインチャートを描画することで提示している。

ここで問題なのは、感情の分類方法の選択とその視覚表現の適切さである。人間が抱く感情は肯定・否定といった単純なものだけではない。一方で感情を130種類以上にも分類し、それらに対応した独立しているチャートをまたいで、視覚的に分析を進めるのは困難である。本研究ではある程度詳細な感情の分類を採用しながら、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することで、感情の時空間的分布がより分かりやすく、また容易に深い知見が得られることを期待する。また、識別可能な分類数にするとしても、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。そこで本研究はその時空間的な分布を提示することでより具体的な知見を獲得できることを示す。

第3章 可視化の手順

本研究では感情情報の可視化によって、その把握の支援を行う。本章では、これ以降で行う情報可視化の説明を分かりやすくするために、情報可視化の手順とそこで生じる技術的課題について説明する。

Cardらによって情報可視化の手順が3つに整理されている[21]が、本研究における可視化の手順としては、これに「生データの収集」を加える形で、最終的な閲覧者への表示までに4つの手順を踏む(図2)。第1手順は生データの収集、第2手順は提示データへの変換、第3手順は視覚的表現への対応付け、第4手順は表示領域への適用である。

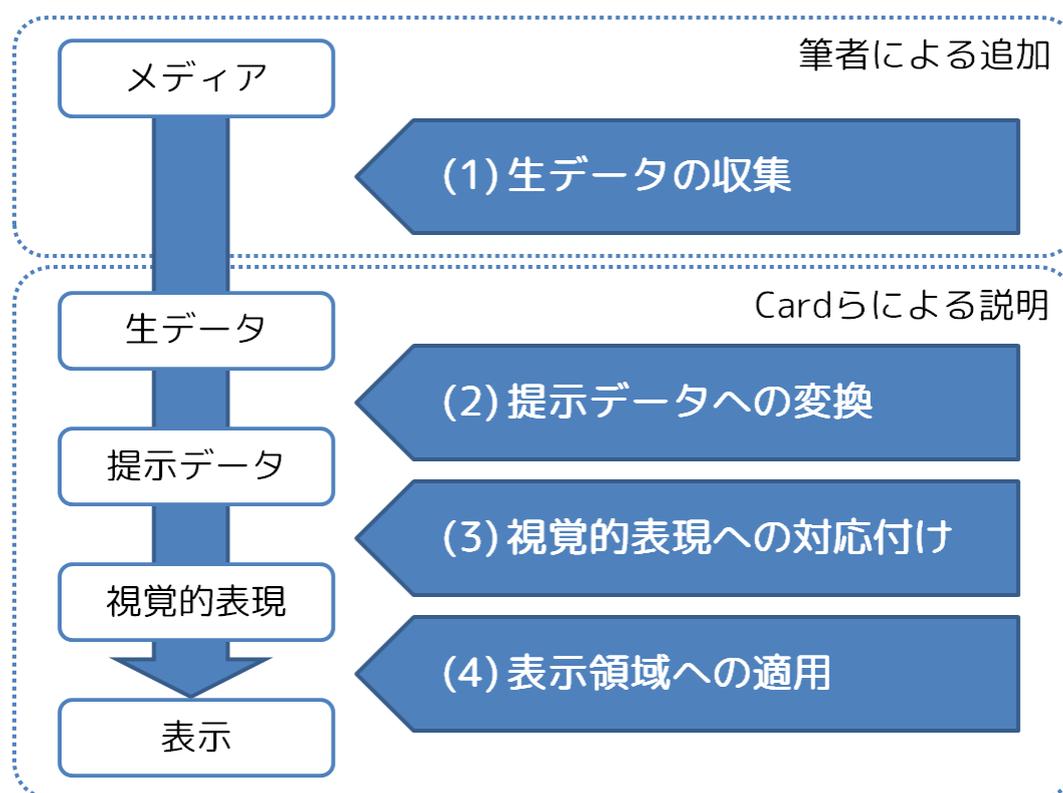


図2 情報可視化の4つの手順

(1) 生データの収集

初めは、Card らの手順に私が追加した、何かしらのメディアから生データを収集する段階である。例えば降水量の様子を把握のために、各観測施設から実際の降水量の情報を収集する段階である。

本研究が目指す大衆感情を把握するためであれば、どのような方法で大勢の人々の情報を集めるのかに当たる。大衆の気持ちを知る手段には、面接調査や電話調査を行う、世論を代表するものとしての新聞記事やソーシャルメディアへの個人による投稿を収集するなどある。これを決めることは、どのような人々の情報を収集するのかに関係することになる。

(2) 提示データへの変換

続いては、生データを提示するためのデータへ変換する段階である。生データから分析の対象となる要素を改めて構築したり、または目的にあった形に変換したりする必要がある。例えば都道府県別の降水量の様子を把握するために、各都道府県ごとに生データの平均を算出する段階である。

本研究においては、感情をどう扱うかを決定する必要がある。前述のとおり、感情の分類方法はさまざまある。さらに、感情の抽出方法についても留意すべきである。

(3) 視覚的表現への対応付け

次の手順では、データと視覚的表現との対応付けを行う。この対応付けが適切になされないと、情報を見逃させたり、見誤らせたりしてしまうことがある。

本研究においては、特に感情の種類を視覚的表現への対応付けに注意を払った。

(4) 表示領域への適用

最後の手順では、第3手順でした対応付けに沿って、実際に表示する領域に描画する。これら手順の結果として、閲覧者への視覚的提示が得られる。

第4章 予備実験 1

4.1 予備実験 1 の目標

予備実験 1 では、手始めとして、対象とするメディアの検討や、感情に関するデータ構造とその視覚的表現に対する設計指針となる知見を獲得することを目標とする。

4.2 可視化の手順

第 3 章で述べた手順に沿って、予備実験 1 での可視化の手順を述べる。

(1) 生データの収集

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発信する場となっている。そこで本研究は、生データを収集するメディアにソーシャルメディアの 1 つである Twitter を選んだ。また、Twitter 側から一般の人々の投稿を収集する仕組みが提供されており、データの取得が容易であることも Twitter を選択した理由の 1 つである。

Twitter での活動にもさまざまな種類がある。例えば、ツイートと呼ばれる基本的な短文の投稿に始まり、そのツイートに対する言及や他の投稿者とのやり取り（メンション、リプライ*）、引用する形の投稿（リツイート）、気に入った投稿へのマーキング（お気に入り）などが挙げられる。リツイートには、Twitter が提供する機能で投稿したもの（公式リツイート）、投稿者がもとの文章をコピーして投稿するもの（非公式リツイート）とがある。本研究では、投稿者が自身の言葉を投稿に含むだろうと考える通常のツイート・非公式リツイート・メンション・リプライを対象とする。また、これらをまとめて投稿と呼ぶこととする。

また、それら投稿を収集する人々の対象には、私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。ほとんどが日本語話者で、日本語で投稿している。また、収集することについて、メンバに知らせていない。

具体的な手順について説明する。まず、前述のメンバで構成されるリストを Twitter が提供している機能によって作成する。続いて、Twitter が提供する REST API のうち GET lists/statuses のメソッドを用いて、1 時間ごとにそのリストに含まれる者による新しい投稿

* あるツイートに対する言及のうちでも、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。

を収集する。より厳密に言えば、このメソッドで取得する対象の投稿はリストに含まれる投稿者の通常のツイート（・非公式リツイート・メンション）とリストに含まれる投稿者同士のリプライであり、リストに含まれる者が投稿したリプライであってもリストに含まれていない者へのリプライは対象ではない。

最終的に、投稿者・投稿文章・投稿時刻からなる生データを収集する。

(2) 提示データへの変換

感情情報の可視化をする研究では、感情を肯定・否定に分類する方法を採用するものが多いが、人間はより豊かな感情を持っている。本研究では、その感情の豊かさを扱いたい。そこで、その豊かな感情どのように分類するのか、また、それに基づいてどのように感情を抽出して提示データへ変換するのかについて検討した。基本的に、投稿された文章を自然言語処理することで導く。

予備実験 1 での感情の分類は、概念辞書である WordNet の日本語版[22]を用いて構築する。図 3 に予備実験 1 で扱ううちの WordNet の構造の例を示す。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。この関係情報はさまざまあるが、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の概念を包含する関係、属性関係とは当該概念の属性を表す際に対象の概念を用いる関係である。総体としての「感情」を表す概念の直下の下位関係にある概念 45 種類を感情の種類を表す概念（感情概念）と定義する。すなわち、感情が 45 種類に分類されるものである。また、前述の感情概念とその下位関係あるいは属性関係にある全ての概念に所属する語 1,382 語を感情語とする。こうして感情概念とそれらを表す感情語の集合の組みである感情語辞書が作成できる。表 1 に作成した感情語辞書の一部を示す。

以降、より具体的な説明をする。まず、感情概念を定義することを考える。総体としての「感情」を表す概念を「感情」を所属語に含む概念とすると 5 つ存在する（WordNet の ID として 00026192-n, 05916739-n, 07480068-n, 07481951-n, 07513035-n で表されるもの）。その概念と下位関係にある概念を個々の感情を表す概念として採用したい。このとき実際には下位関係にある概念は 56 個あるが、ID が 05916739-n の概念は 00026192-n の概念に対して下位関係にあり、重複するためにこれを除く。続いて、それぞれの概念を表す言葉を集めたい。それにはその概念に所属する語のほかに、その概念に対して下位関係・属性関係にある概念に所属する語も含める。その際、再帰的に下位関係・属性関係にある概念に所属する語も含める。また、感情語には漢字・ひらがな・カタカナで表現されるもののみを採用する。こうすると「感情」の下位関係にある概念のうち、属する語が存在しない概念が 10 個存在するためにこれらは感情概念から除く。こうすることで感情概念として 45 種類を定義し、また、それぞれの概念を表す感情語の集合が得られる。

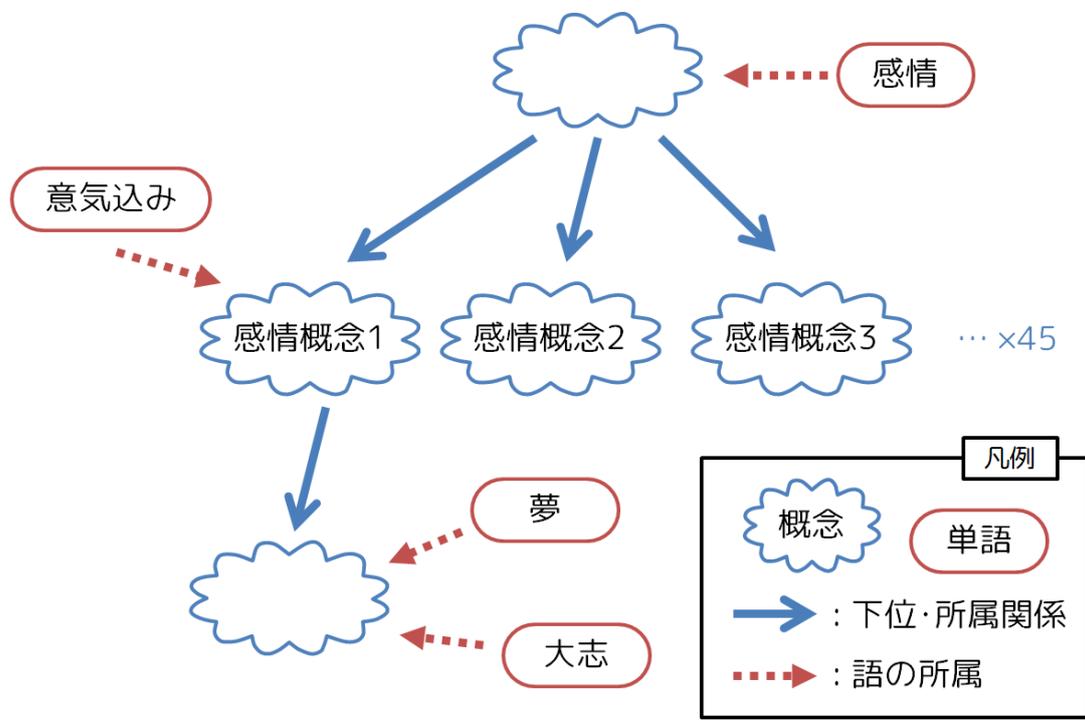


図3 WordNetによる感情の分類

表1 予備実験1の感情語辞書

感情概念	感情語
感情概念1	意気込み、夢、大志、希望、意欲、憧れ、……
感情概念2	嬉しい、幸せ、幸福、御機嫌、ハッピー、……
感情概念3	嫌悪、軽蔑、憎しみ、憎悪、反感、……
:	:
感情概念45	陽気

続いて、予備実験1での感情の抽出方法について述べる。前述の感情語辞書を用いて、投稿文章からそれぞれの感情を含むかどうか識別する。まず、文章をオープンソースの形態素解析エンジンである MeCab[23]によって単語に分け、活用があるものは基本形に修正する。そして、感情語とその基本形とを文字列照合する。文字列が等しい場合は辞書から感情概念を導き、その投稿はその感情を含むとする。ここでは、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こと、その感情の強さについては判定しないこと、慣用語のほか、否定や疑問、皮肉といった句や文、文章単位での特徴については考慮しない。

最終的な提示データは、どの投稿者が、どの時刻に、どのような感情の投稿をしたかとなる。投稿者と投稿時刻については、投稿に付加されているメタデータを用いる。

(3) 視覚的表現への対応付け

図4は、予備実験1での提示データと視覚的表現との対応付けの図例である。

投稿者は縦軸の位置で表す。投稿者の並び順に特別な意味はない。

時刻は横軸の位置で表す。具体的には左端から右端へ、古いものから新しいものへと対応している。

感情を表現するには、投稿者と投稿時刻からなる位置を歪ませないため、また視覚的重なりをなるべく避けるために、HSV色空間でのH（色相）に割り当てることにする。色空間とは色を立法的に定義する方法であり、HSV色空間は色をH（色相：赤・緑・黄色といった色の種類で0~360の値を取る）・S（彩度：色の鮮やかさで、0~100%の値を取る）・V（明度：色の明るさで0~100%の値を取る）で定義するものである。色相には循環性があり、量の表現には乏しいが、カテゴリの表現に適しているために割り当てた。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する（図5）。このとき、どの色についてもS・Vは最大の100%とする。色相の割り当て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にして表現した。前述の位置に、全て同じ大きさの矩形で、対応した色相で塗る。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。

(4) 表示領域への適用

図4のとおり、背景には黒を選んだ。投稿者は白の横一線で区切りを付けて表す。

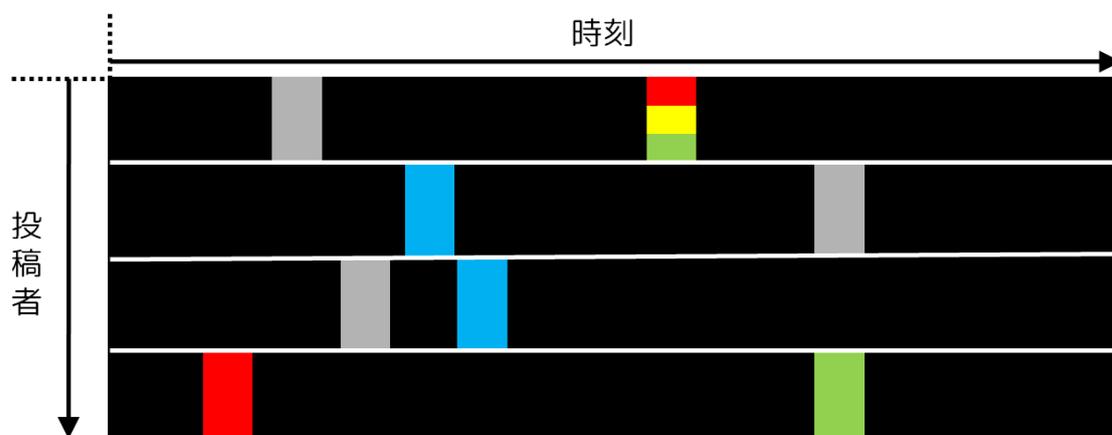


図4 予備実験1の可視化の図例

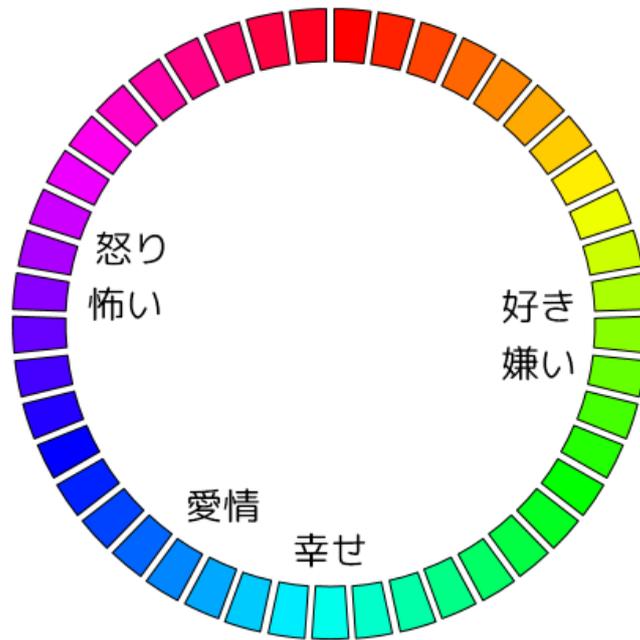


図5 予備実験1での感情と色との対応付け

4.3 可視化例

図6に予備実験1の可視化結果の一例を示す。図6で示されているのは、私の研究室に所属するメンバ67人の投稿で、日本時間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、合計約5,000個である。そのうち約500個が感情語にマッチした。0時を基準に1日ごとに白い縦線で区切りを入れた。

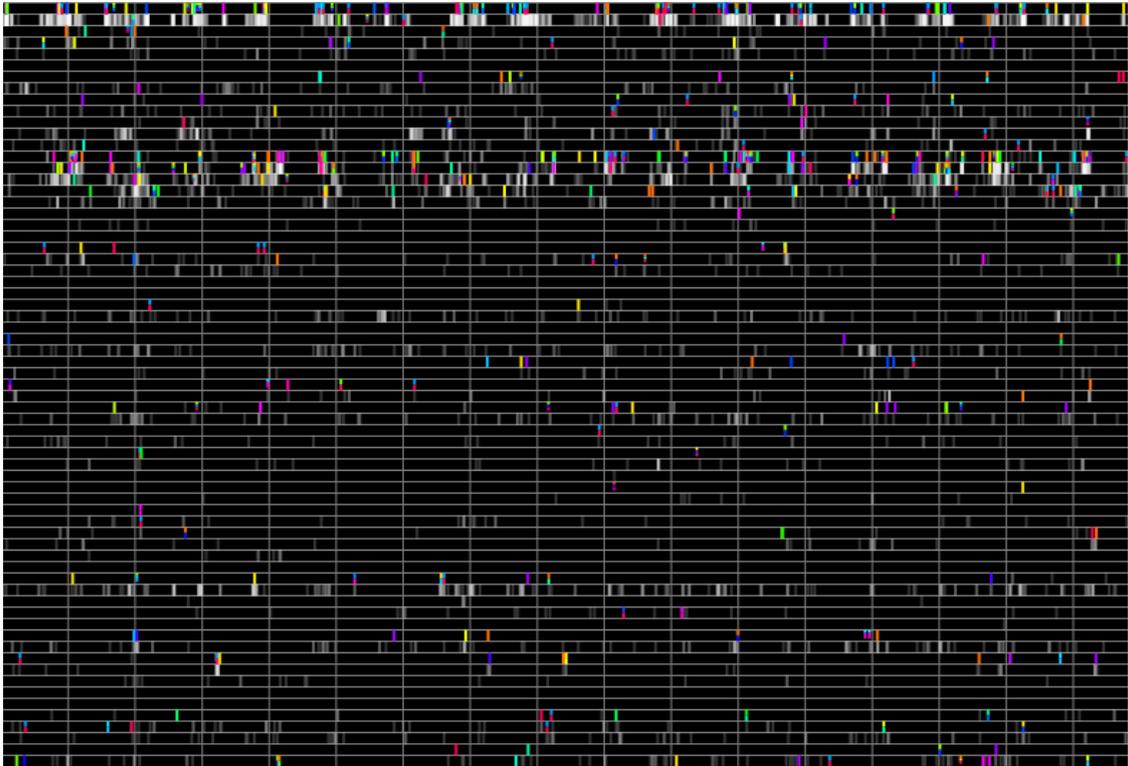


図 6 予備実験 1 の可視化例

感情に関係する部分で言えば、投稿が比較的多くてもそのうち感情語を含むものが少ない者がいることや、逆に投稿自体が少なくても感情語を含む割合が高い者がいることなどが分かる。

それ以外には、全体的に夕方から深夜にかけての投稿が多いことや、よく投稿をする者とそうでない者の差が大きいことが分かる。

4.4 考察

まず対象メディアとして Twitter を選んだことについて、有益だったと考える。67 人も投稿者全員一人ずつに随時そのときの気持ちをインタビューなどしてまわることは現実的でなく、仮に気持ちを聞いたとしても違和感のあるものになりがちだろう。Twitter ではそれぞれの意見・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。作成した感情の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。しかしながら、それがどのような感情なのかを把握するに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、分類する数が少なすぎでは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、感情の分類を人間にとって自然な数に、また視覚的に把握できる表現にすることが必要である。

また、感情語を含む投稿同士の時間間隔は大きく、その時点限りであることが多かった。また、他の投稿者との感情の同調などが期待されたがこの表現では見つけられなかった。この問題に対する原因として、前述の視覚的表現への対応付けが不適當であったこともあるが、対象コミュニティが小さく、投稿が少ないことも考えられる。そのため、投稿を時間に対して密にする必要がある。

第5章 予備実験 2

5.1 予備実験 2 の目標

予備実験 1 で得られた知見を踏まえながら、感情の主題地図を作成するにあたっての技術的課題を洗い出すことを目的とする。

5.2 可視化の手順

第 3 章で述べた手順に沿って、予備実験 2 での可視化の手順を述べる。

(1) 生データの収集

予備実験 2 でも、Twitter を対象メディアとして分析することを試みる。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・日本語のテキストを含む、通常のツイート（・非公式リツイート・メンション・リプライ）とした。日本のものに絞ったのは、筆者が日本語話者であって実際の分析を進めることができるから、また、感情抽出に日本語の自然言語処理を行うからである。SemioCast 社によれば、2012 年 6 月の 10 億 5800 万の公開されているツイートのうち、0.77% ほどが位置情報を付加しており、また、10.6% ほどが日本語によるもので、英語に次いで 2 番目に多い[24]という。

予備実験 2 における具体的な生データの収集方法について述べる。Twitter が提供する Streaming API のうち、POST statuses/filter メソッドを用いる。このメソッドでは対象の地域を限定するパラメータを渡すことができる。今回は前述のとおり日本のものを集めたいので、南は緯度 23.45 度から北は 46.50 度まで、西は経度 122.22 度から東は 150.27 度までをパラメータとして渡した。さらに外国の部分を除くため、得られた投稿に付与されている経緯座標のメタデータを用いて、

$$\text{緯度} \leq \text{経度} - 95.00$$

を満たすもののみを対象とした。つまり、図 7 に表現された領域の地理情報をメタデータに含む投稿が対象となる。また、言語や時刻についても、メタデータを用いた。こうすると、1 日でおおよそ 200,000~300,000 個の投稿がある。

最終的に、投稿文章・緯度・経度・投稿時刻からなる生データを収集する。

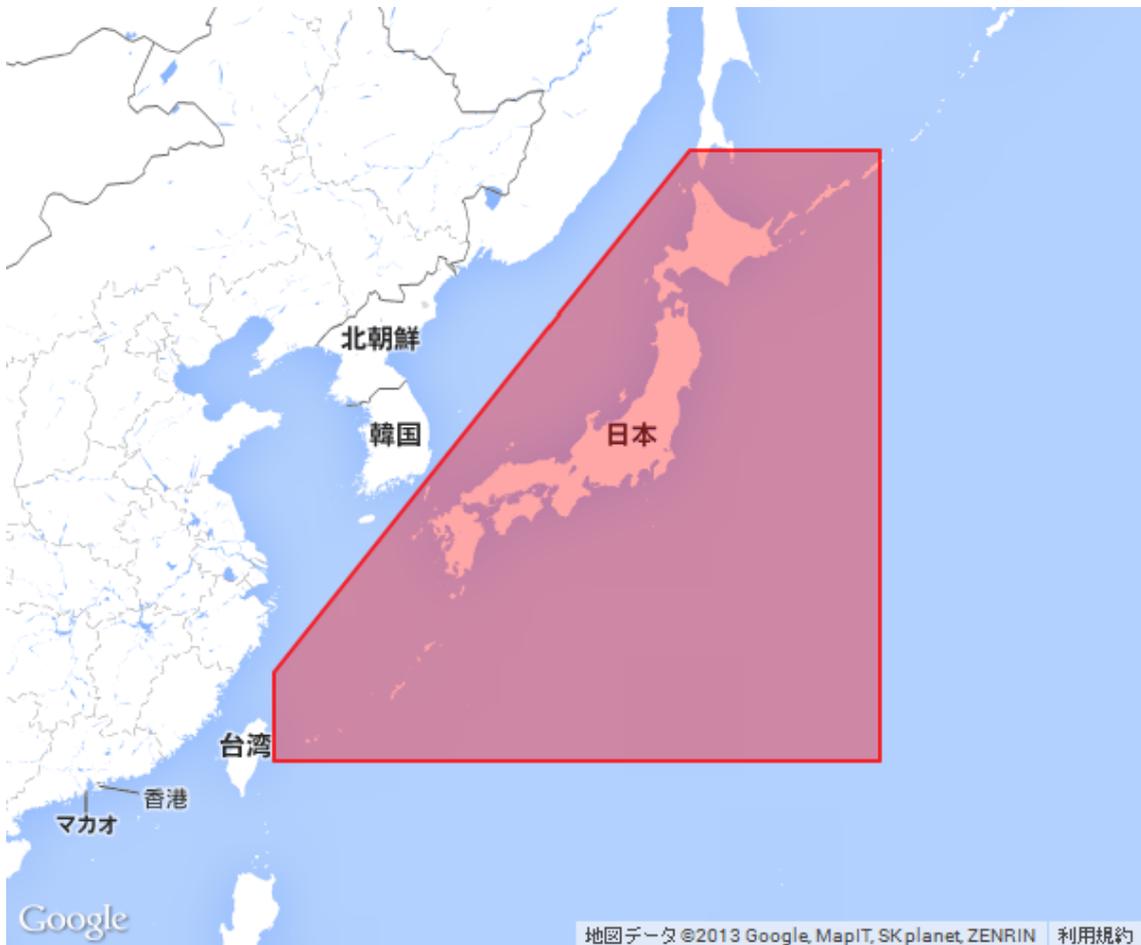


図7 対象となる地理的な範囲

(2) 提示データへの変換

次に、感情の分類方法について考える。予備実験1では、感情を45種類にも分類したために、その把握が難しいという問題があった。そこで予備実験2では、Plutchikが提唱する基本感情8つからなる分類[11]を採用する。具体的には「Joy・Acceptance・Fear・Surprise・Sadness・Disgust・Anger・Anticipation」の8つである。ここで少しニュアンスが異なるがそれぞれの和訳として「喜び・好き・恐れ・驚き・悲しみ・嫌い・怒り・望み」を採用する。「喜び」と対極をなす感情は「悲しみ」、また「喜び」に近い感情として「好き・望み」があるなど、この分類方法には感情同士の関係されており、循環性がある。図で表すと図8のようになり、隣り合っていることでその意味合いの近さを、点对称の位置にあることでその意味合いの対極性を表している。感情の種類を、4つの極性をもって4種類とするものもあるが、例えば「喜び」と「悲しみ」がともに高まっていることと、何も感情を抱いていないことは等価でないと考え、この8種類とした。

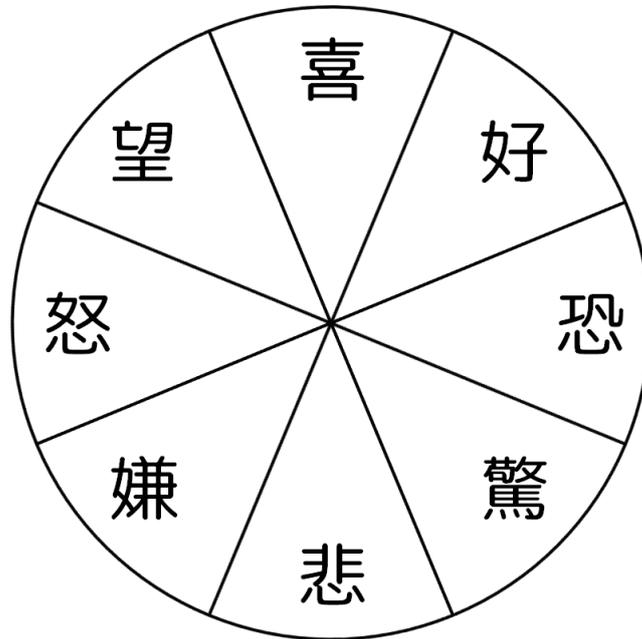


図 8 Plutchik による基本感情の環

さらに、実際の感情抽出方法について検討する。前述の感情の分類方法をもとに、感情語辞書を作成する。今回は WordNet を用いずに、該当するであろう語を経験的に筆者が選んだ。表 2 は実際の感情語辞書の一部である。予備実験 1 と同様、投稿された文章を MaCab によって形態素解析し、感情語と文章に含まれる単語とを照合して、等しい場合は辞書から感情概念を導き、投稿はその感情を含むとする。ここでも、1つの投稿で異なる感情概念を複数個含む可能性があること、その感情の強さについては判定しないこと、文や文章単位での特徴については考慮しない。

表 2 予備実験 2 の感情語辞書

感情概念	感情語
喜び	喜ぶ、幸せ、嬉しい、笑う、愉快、笑、……
好き	愛、好き、恋しい、憧れ、可愛い、おめでとう、……
恐れ	恐い、怖い、怯える、恐怖、不安、心配、ホラー、……
驚き	驚く、焦る、困る、驚愕、仰天、ビックリ、……
悲しみ	悲しい、惜しい、嘆く、泣く、不憫、絶望、すみません、……
嫌い	嫌、憎む、恨む、悩む、苦しい、嫌悪、軽蔑、……
怒り	怒る、叱る、激怒、我慢、むかつく、……
望み	望み、欲しい、楽しみ、希望、期待、わくわく、……

ここで、実データを用いて、感情がどの程度分類されているのかをしてみることにする。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日間で、前述の日本領域内・日本語で投稿されたことが確認できるものについて表3に示す。総数は6,800,080個であった。そのうち、いずれかの感情を含むと分類されたものは1,876,058個で、総数に対しておよそ28%であった。具体的な感情それぞれを見てみると、「喜び」の感情を含むものが総数に対しておよそ16%と他の感情と比べて高い。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、次で述べる感情的なバイアスがかかっている可能性がある。

表3 実データの感情の分類の様子

	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	4.99 %
「望み」を含むもの	90,492	1.33 %

感情的バイアスとは、8つに分類した感情同士を比較する際のバイアスで、3つの要因による。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。この個数の差を把握したいのならば、このバイアスを解消しないままにする必要がある。しかし、その個数の差ではなく普遍状態との差を見たいのならば、このバイアスを解消する必要がある。続いてはメディアの要因である。言葉や表情などといった概念的なものや、手紙やソーシャルメディアといった具体的なものなど、感情はこれら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にくくなるというバイアスがかかっている可能性がある。例えば今回は感情の抽出対象にTwitterへの投稿文章を選んだが、ほかのメディアのほうが悲しみを投稿しやすいというようなことがありうる。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。日本語の自然言語から感情抽出を行う試みをした菅原も「感情カテゴリ毎の抽出精度にバラつきが出き、抽出しやすい感情と抽出しにくい感情が出てきていた」と言及している[14]。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需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。

続いて、前述と同じ時間帯のデータについて、投稿自体の個数の時間的分布を見てみることにする。図9は1時間ごとの時間帯で集積した個数を示したヒストグラムである。朝方には少なく、夜中に多いことが分かる。また、図10は1日ごとに集積したものである。1時間ごととはほどではないが、比較的多い日と少ない日があることが分かる。このことを考えると、時間的バイアスがかかっていることが考えられる。投稿の多い時間帯は、抽出される感情の絶対的な個数も多くなることが予想される。

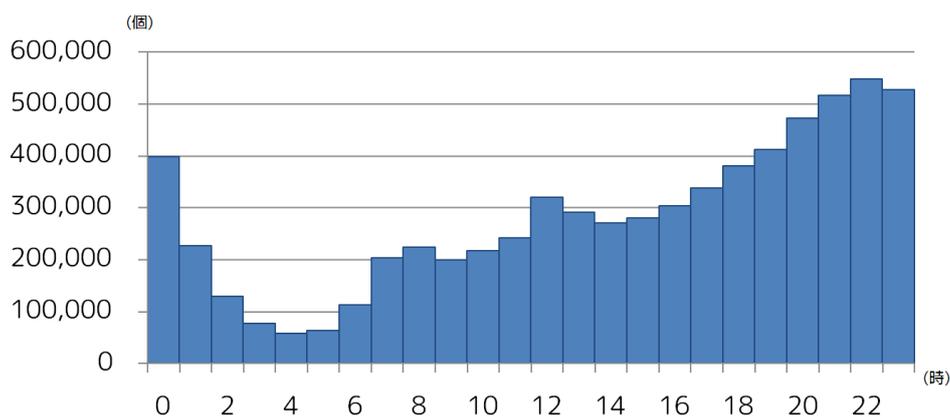


図9 1時間単位の投稿数の分布

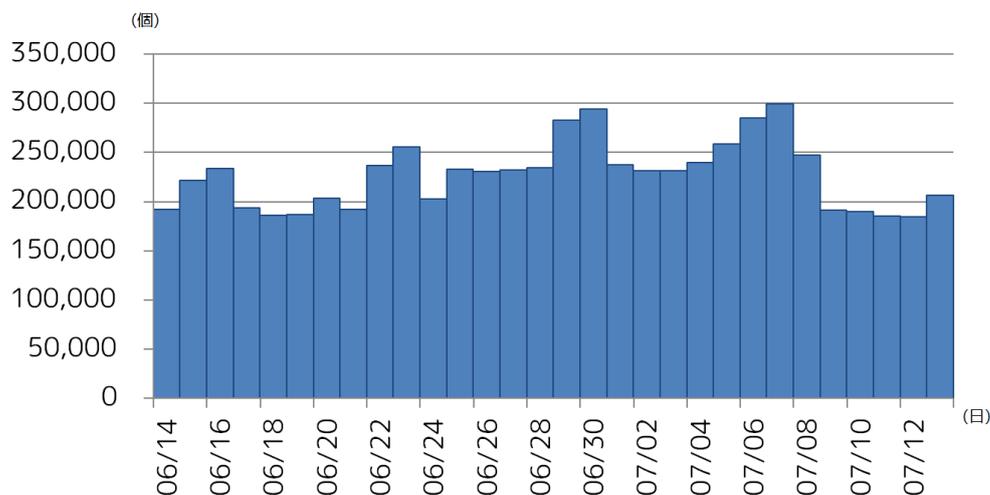


図10 1日単位の投稿数の分布

さらに、前述と同じ時間帯のデータについて、投稿自体の個数の空間的分布を見てみることにする。図 11 は緯度経度 1 度ずつの領域に分け、その領域に該当する投稿数について、最大のを基準に、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。都市部は人口が多く、それに比例して投稿数も多い。そうなるとう抽出される感情の絶対的な個数も多くなることも予想され、ここに空間的バイアスがかかっていることが考えられる。

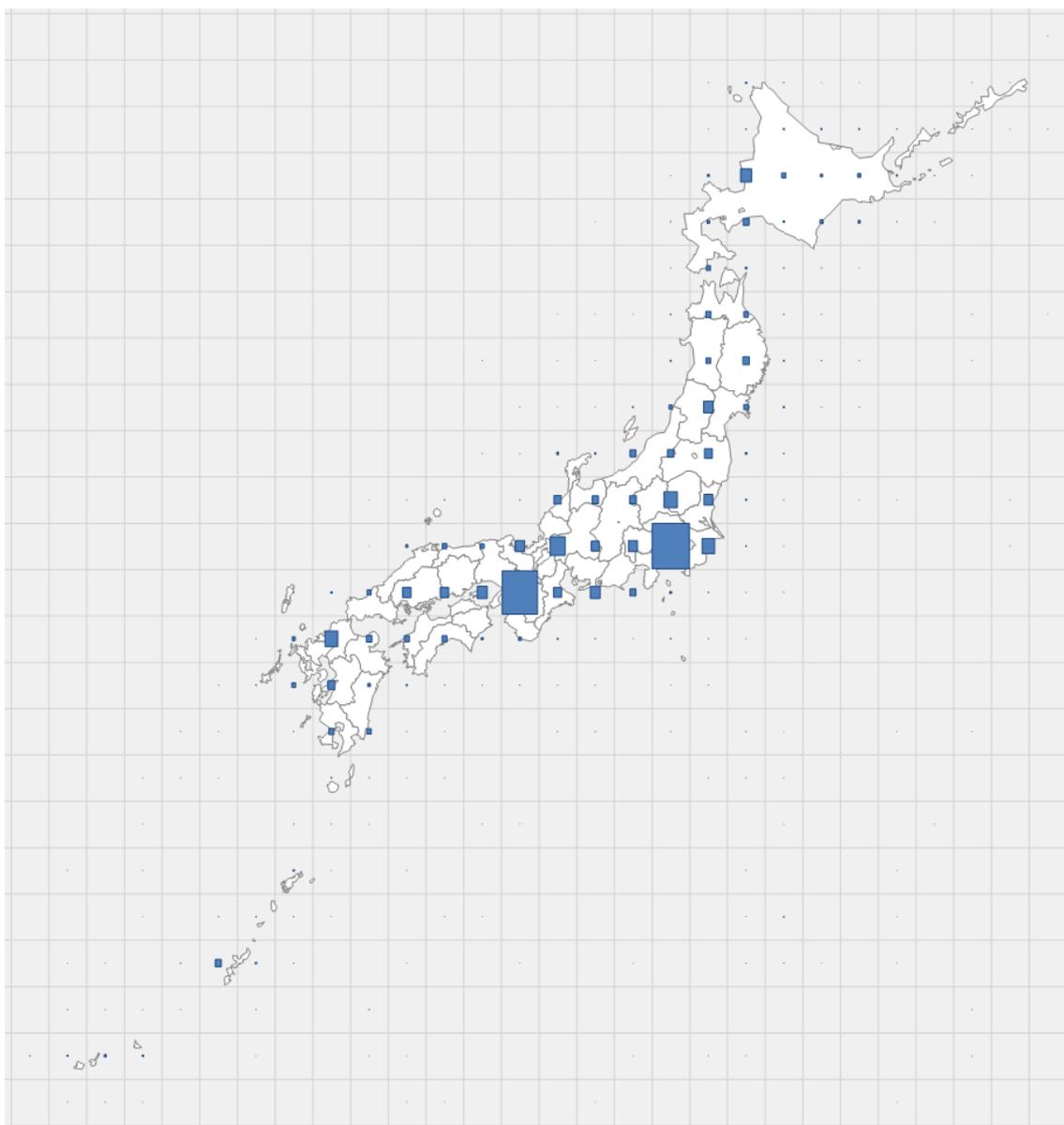


図 11 緯度経度 1 度の矩形領域を単位とする投稿数の分布

図 11 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[25]を利用した。正距円筒図法をベースに、経度方向に 2 割ほど縮めたものである。また、以降の空間的分布の提示にも、この地図データを用いる。正距円筒図法は面積などが正しく表現できないが、日本の範囲は狭くあまり影響は受けないと判断した。

時間的バイアスと空間的バイアスは、投稿自体の絶対数に関するバイアスである。ここで、2 つの投稿の集合 $A \cdot B$ について、 A の投稿総数が 100 個でそのうち「喜び」が 10 個、 B の投稿総数が 20 個でそのうち「喜び」が 5 個とする。 A のほうが B よりも喜びの数が 5 個多いため A の「喜び」の特徴量がより高くなるのがバイアスを解消しないもので、 B のほうが A よりも喜びの割合が 15 ポイント高いため B の「喜び」の特徴量がより高くなるのがバイアスを解消したものである。どちらもそれぞれ意味のある値なので、分析目的に応じてそれを切り替える必要がある。

以上を踏まえたうえで、データセット内での、特定の時間帯・地理領域で、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、8 つに分類されるどの感情を含むかで分割を行う。続いて、一定の時間間隔 (例えば 1 時間ごと) で区切った時間帯に分割する。さらに、特定の地点を基準に、一定の大きさ (例えば、緯度経度 1 度ずつの領域) で地理的領域に分割する。この (感情の種類、時間帯、地理的領域) の組を 1 つの単位として、それぞれ該当する投稿の個数を数え上げる。数え上げた個数が最大となるものを基準として 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が導かれる。これは、前述した感情的・時間的・空間的バイアスが全てかかった値である。

続いて、前述の 3 つのバイアスの解消方法について検討する。まず、それぞれのバイアスがかかっている強さの度合いを算出する。

感情的バイアスの強さの度合いの算出方法について示す。全ての投稿について、それぞれの感情別にそれを含む投稿を数え上げる。次にその最大値を基準として、それぞれの感情が何%になっているかを算出する。最終的にその逆数が、その感情にかかる感情的バイアスの強さとなる。

時間的バイアスの強さの度合いの算出方法について示す。全ての投稿について、前述の一定時間間隔で区切った時間帯別に、その時間帯内で投稿されたものを感情を含むかどうかを問わずに数え上げる。次にその最大値を基準として、それぞれがその何%になっているかを算出する。最終的にその逆数が、その時間帯にかかる時間的バイアスの強さになる。

空間的バイアスの強さの度合いの算出方法について示す。全ての投稿について、前述した一定の大きさで区切った地理的領域別に、その領域内で投稿されたものを感情を含むかどうかを問わずに数え上げる。次にその最大値を基準として、それぞれがその何%になっ

ているかを算出する。最終的にその逆数が、その矩形領域にかかる空間的バイアスの強さになる。ただし、その領域での投稿数が少なすぎると空間的バイアスの強さが非常に大きくなってしまふ。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領域として空間的バイアスの強さを 0 と定めることでこれを回避する。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が少なすぎる領域を算出したのち、前述した基準となる感情スコアに解消したいバイアスの強さの度合いをかけ、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が導かれる。

最終的に提示するデータは、(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である。

(3) 視覚的表現への対応付け

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。

時間帯は、それぞれビューを順番に切り替えることで対応付ける。

感情の種類は、色と地図の位置で表すことにする。具体的な配色を図 12 と表 4 に示す。これは Plutchik の本の表紙[26]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。さらに、金箱ら[27]によって、黄色と喜び・青と悲しみ・赤と怒り・緑と落ち着きが、山内ら[28]によって、赤と怒り・橙と喜び・緑と信頼・青紫と悲しみ・紫と恐れ・赤紫と嫌悪などが関連していることが明らかにされており、一部がよく対応している。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。明らかにされ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 12 のようになるのは適当である。地図の位置で表すことについては次の項で示す。



図 12 予備実験 2 における感情と色の対応付け

表 4 予備実験 2 における感情と色の対応付け

感情概念	色 (R,G,B)
喜び	黄色 (255, 255, 0)
好き	黄緑色 (0,255,0)
恐れ	緑色 (0, 127, 0)
驚き	水色 (0, 255, 255)
悲しみ	青色 (0, 0, 255)
嫌い	紫色 (127, 0, 127)
怒り	赤色 (255, 0, 0)
望み	橙色 (255, 127, 0)

(4) 表示領域への適用

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、下段に移って左下から右に、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り、地図の位置で感情の種類と相互関係を表すことになる。また、地図 1 枚の左上に文字列でも提示した。

時間帯のビューを切り替えるには、閲覧者が操作することで行う。

5.3 可視化例

図13～図16に予備実験2の可視化結果の一例を示す。この4つの図のデータは共通で、日本時間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、時間帯の単位を午前0時を基準に24時間とする。全てのデータが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13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いる。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14は感情的バイアスだけを解消したもので、抽出された感情それぞれがどう空間分布しているのかを見ることができるようになっている。どの感情も東京都心を含む領域が最も高いことが分かる。

図15は空間的バイアスだけを解消したもので、感情スコアはほぼ一様に広がっているように見える。深い緑に対応する「恐れ」は、宮城県の辺りで高いスコアであることが分かる。

図16は感情的・時間的・空間的バイアスを解消したものである。上記の「恐れ」の分布についてさらにはっきりと宮城県の辺りが高いスコアであることが分かる。

図17・図18にもう1つの可視化結果の例を示す。データセットは前述のものと同じだが、この例では時間帯の単位を午前0時を基準に1時間としており、データが時間方向に24個に分割されている。

図17はデータセットのうち午前4時台のものについて、感情的バイアスのみを解消して可視化したものである。この時間帯は他の時間帯に比べ絶対数が少なく、どの感情も高い値の分布は見られない。

図18はデータセットのうち午前4時台のものについて、感情的・時間的・空間的バイアス全てを解消して可視化したものである。「驚き」と「嫌い」について、高い値になった領域が観測できる。

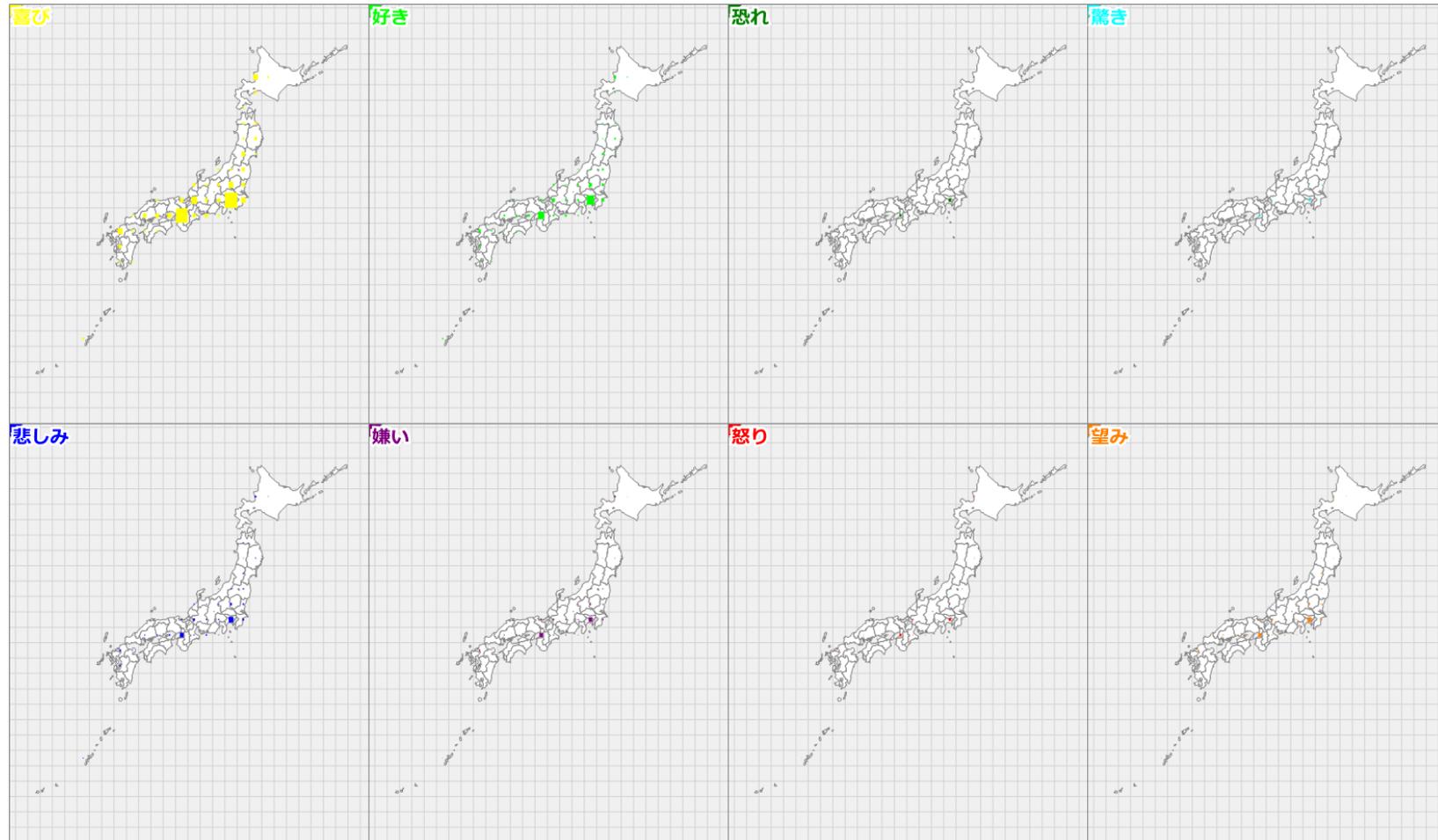


図 13 全てのバイアスがかかったもの (予備実験 2)

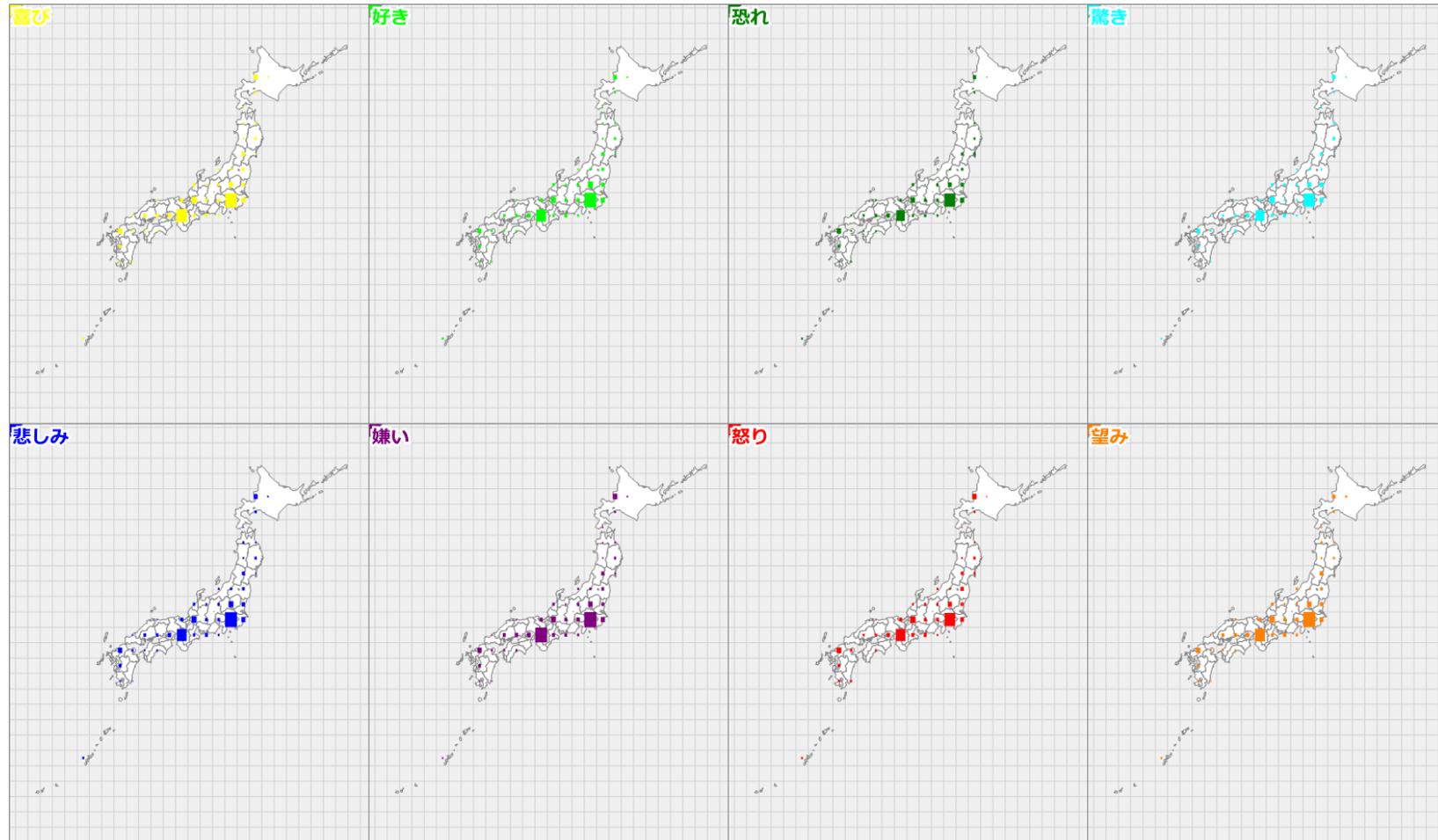


図 14 感情的バイアスのみを解消したもの（予備実験 2）

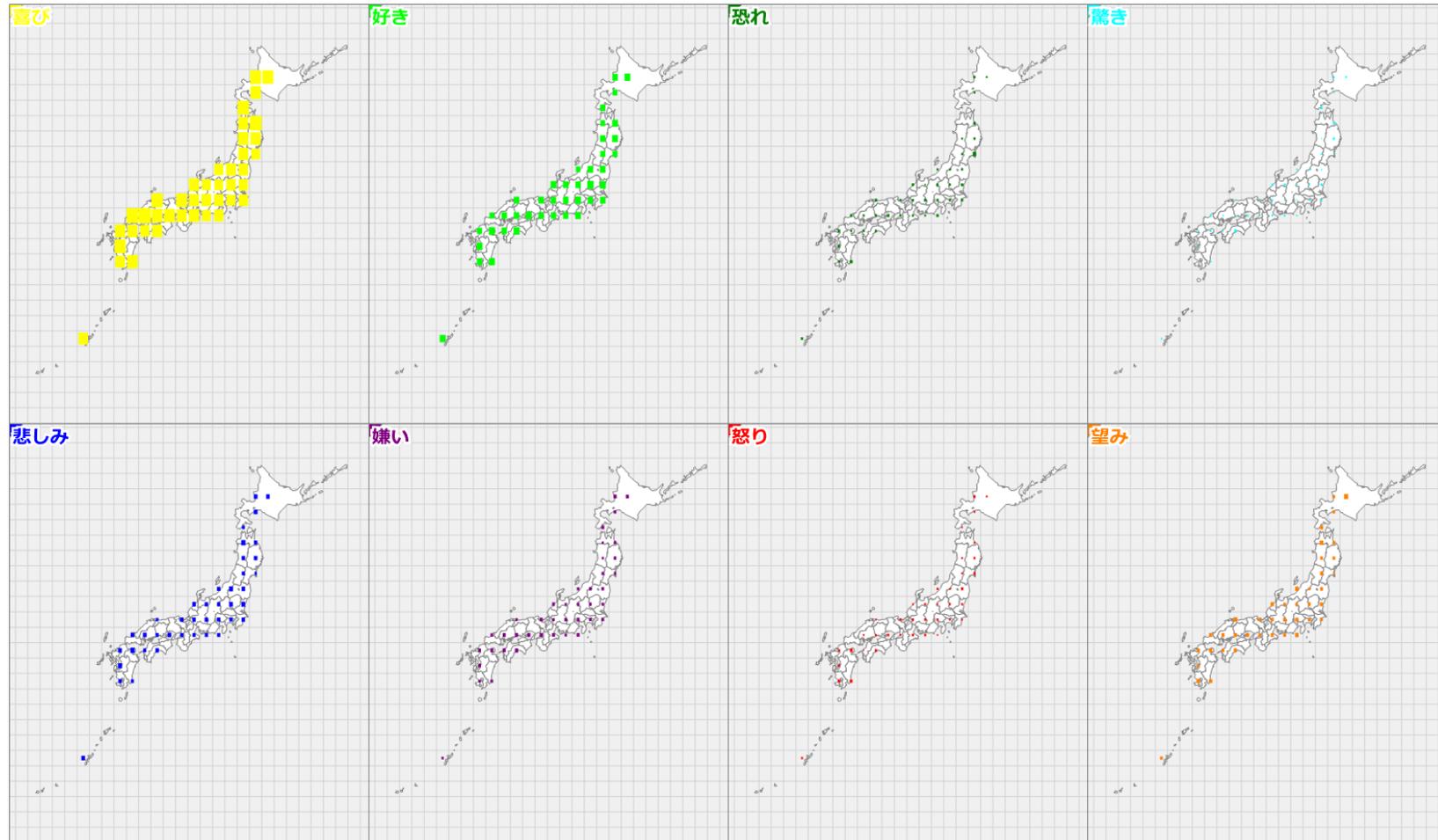


図 15 空間的バイアスのみを解消したもの（予備実験 2）

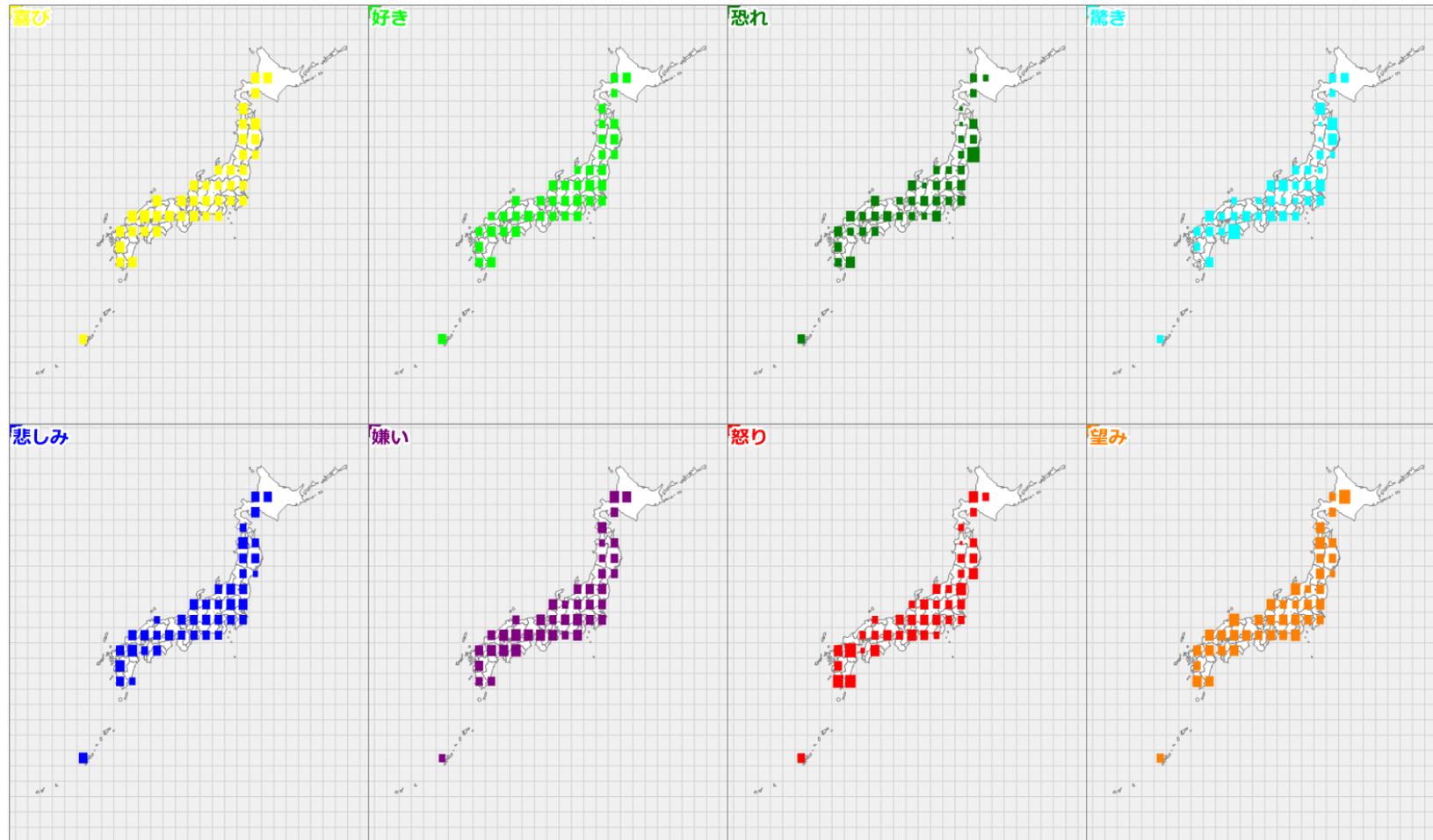


図 16 全てのバイアスを解消したもの (予備実験 2)

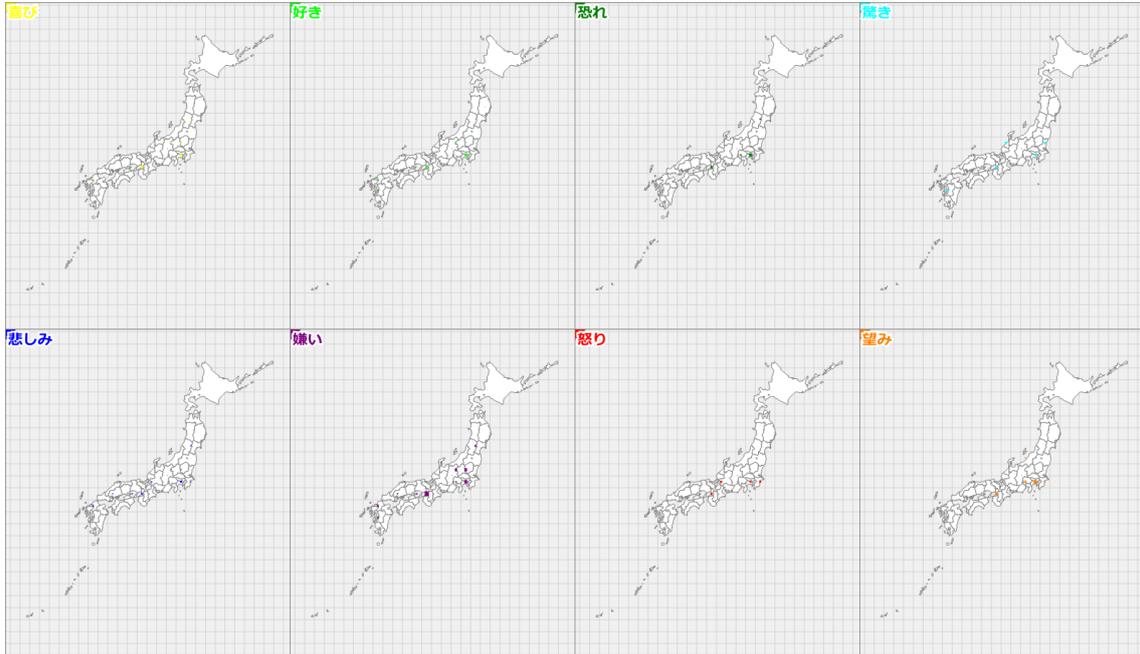


図 17 感情的バイアスのみを解消したもの（2013年6月14日午前4時 / 予備実験2）

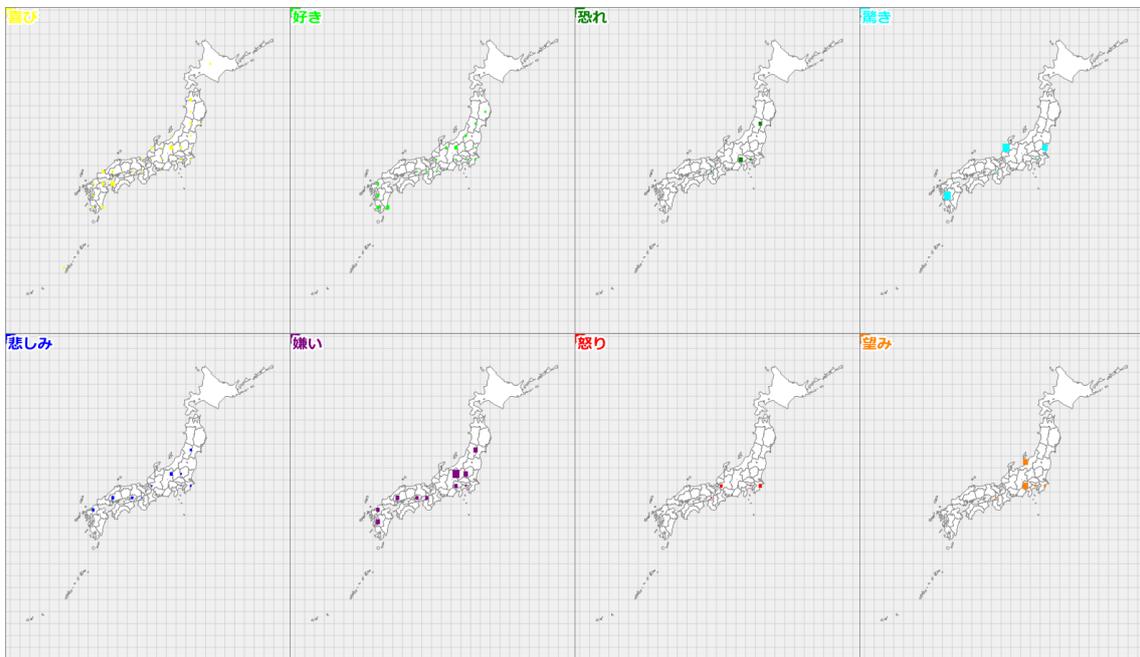


図 18 全てのバイアスを解消したもの（2013年6月14日午前4時 / 予備実験2）

5.4 考察

バイアスの存在について議論する。実験の過程で感情的・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た。ただし、それぞれバイアスを解消するのかもしれないのかは閲覧者が何を知りたいかと関係するため、それに応じて切り替える必要がある。具体的には、どんな感情が表出されやすいのかを知りたいのならば、感情的バイアスは解消すべきでないし、感情それぞれが他の時間帯・地域と比べたときにどれほど表出されているのかを知りたいのならば、解消したものを提示すべきだろう。時間的・空間的バイアスについては、絶対数が知りたいのならば解消しないデータを、割合を知りたいのならば解消したデータを提示すべきだろう。

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。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、配色には問題があると感じた。色によって知覚的な明度が異なることで、背景色とのコントラストもそれぞれ異なり、感情によって表現している量を読み取りにくくなっている。解決方法の1つに、背景の色を変えてコントラストをそろえることがあるが、これでは地図1枚ごとに受ける印象が異なり、また本来は存在しない何らかの分類があることを暗示してしまう。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。

第6章 感情天気図

2つの予備実験で得た知見をもとに、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。

6.1 3つの分析方法

「感情天気図」は、感情情報の時間分布を見る、空間分布を見る、時空間分布を見るという3つの分析方法を可能にする。それら分析にはそれぞれ、時間分布ビュー・空間分布ビュー・時空間分布ビューが対応し、それぞれのビューを切り替えながら感情情報の分析を進める。これらは生データは同じだが、提示するデータやその対応付けなど、可視化の手順に異なる部分がある。まず、共通する部分について説明をする。

(1) 生データの収集

生データの収集は、予備実験2と同じ手順で行う。Twitterを対象メディアとして、日本の地理情報・日本語のテキストを含む、通常のツイート（・非公式リツイート・メンション・リプライ）を集める。生データは投稿文章・緯度・経度・投稿時刻からなる。

(2) 提示データへの変換

提示データへの変換も、基本的に予備実験2と同じ手順で行う。投稿文章に自然言語処理をすることによって、8つに分類された感情の抽出を行う。各バイアスについては、解消するかどうかをGUIによって切り替えることができるようにした。

(3) 視覚的表現への対応付け

感情の種類について、予備実験2とは異なる色に対応付けする。予備実験2ではHSV色空間の色相に割り当てたが、「感情天気図」では改め、L*a*b*色空間の色相に対応付ける。実際の対応付けを図19と表5に示す。L*a*b*色空間は、L*（明度：色の明るさを示し、0~100の値を取る）、a*（正が赤、負が緑の色味を示す）、b*（正が黄、負が青の色味を示す）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色を表す座標のユークリッド距離が実際の知覚的距離と等しく対応付けられている。これを採用したのは、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ことに加えて、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ことを狙ったためである。また、これは予備実験2で用いた色相とよく対応している。



図 19 感情天気図における感情と色の対応付け

表 5 感情天気図における感情と色の対応付け

感情概念	色(R,G,B)
喜び	(205, 179, 107)
好き	(146, 194, 128)
恐れ	(71, 198, 177)
驚き	(24, 198, 232)
悲しみ	(128, 185, 254)
嫌い	(209, 165, 234)
怒り	(250, 153, 183)
望み	(245, 160, 131)

具体的には、喜び・好き・恐れ・…・望みを $0 \cdot 1 \cdot 2 \cdot \dots \cdot 7$ 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。参考として、図 20 に L*が 75, 50, 25 のときの一般的なディスプレイで正しく表示できる色領域を示す。周囲の灰色部分は一般的なディスプレイでは表示不可能な色領域である。「喜び」を黄色にすることを基準に、8つの値を色相で見分けるために図 20 中の正八角形の重心から頂点までの長さをこの向きのまま最大にしたい。L*が 74 のとき、重心から頂点までの長さが最長で 40 となるため、前述の式となる。図 20 はクリエイティブコモンズライセンス (CC BY-SA 3.0) で公開されているもの[29]を再編集している。図 20 に関しては、CC BY-SA 3.0 とする。

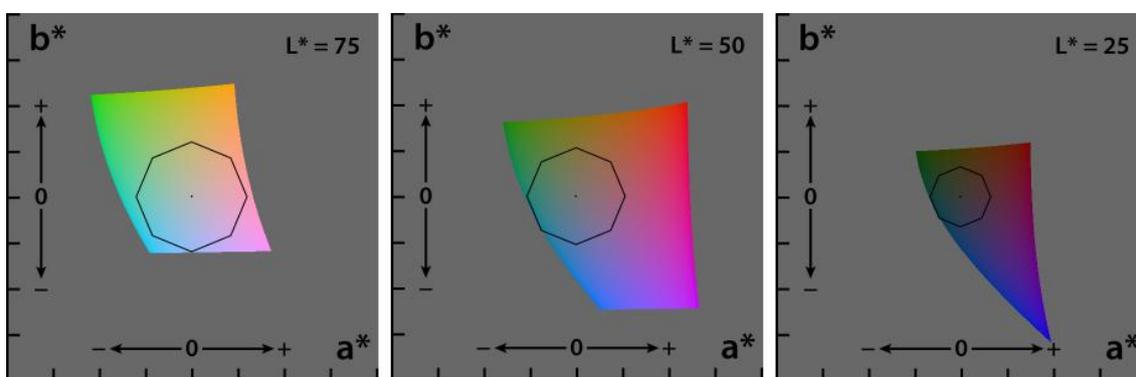


図 20 L*a*b*色空間において一般的ディスプレイが表示可能な色領域

6.1.1 時間分布の分析

感情情報の時間的分布を分析するために時間分布ビューを提示する。時間分布ビューは、それぞれの感情を含む投稿について、単位時間帯ごとの分布をヒストグラムで示すものである。ここでは地理的領域は表現されない。

(2) 提示データへの変換

生データの分割・集積を、(感情の種類、時間帯)の組で行う。空間的バイアスについては考慮する必要が無い。

(3) 視覚的表現への対応付け

感情ごとにヒストグラムを描画する。X軸が時間帯、Y軸が感情スコアと対応している。図 21 は時間帯を 24 個に分割したときの「怒り」の感情スコアを表すヒストグラムの例である。

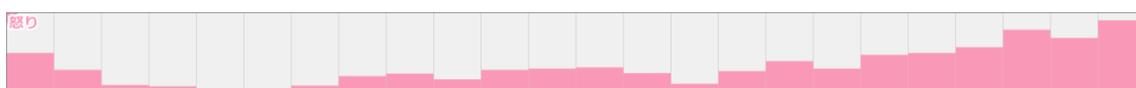


図 21 ヒストグラムの例

(4) 表示領域への適用

画面を縦に8分割し、それぞれの感情のヒストグラム描画する。また、一番下段に時間を示す文字列も描画する。

6.1.2 空間分布の分析

感情情報の空間的分布を分析するために空間分布ビューを提示する。ここでは時間帯は表現されない。

(2) 提示データへの変換

生データの分割・集積を、(感情の種類、地理的領域)の組で行う。時間的バイアスについては考慮する必要が無い。

(3) 視覚的表現への対応付け

地理的領域とそこでの感情スコアについて、該当する領域を塗りつぶす範囲を対応付ける表現だと、差の量がはっきりと分かるが、周囲の領域との接続がはなれ、関係が繋がって見えづらくなる。解決方法の1つに不透明度を含む色を対応付け、領域全体を塗りつぶす表現があるが、これは周囲との関係が分かりやすいが、具体的な量がはっきりとは分かりづらい。そこで、2次元メタボールを与える影響力の境界線を描画することによって感情スコアの領域表示を行うことにする。図22に感情スコアを領域ごとに塗りつぶして表現したものから、メタボールレンダリングによる表現に変えたものを示す。スコアが近く、領域も近いものを緩くつなげることによって、地理的広がり把握しつつ、大まかな量を把握できることを狙った表現である。これは天気図でいう等圧線のようなものである。

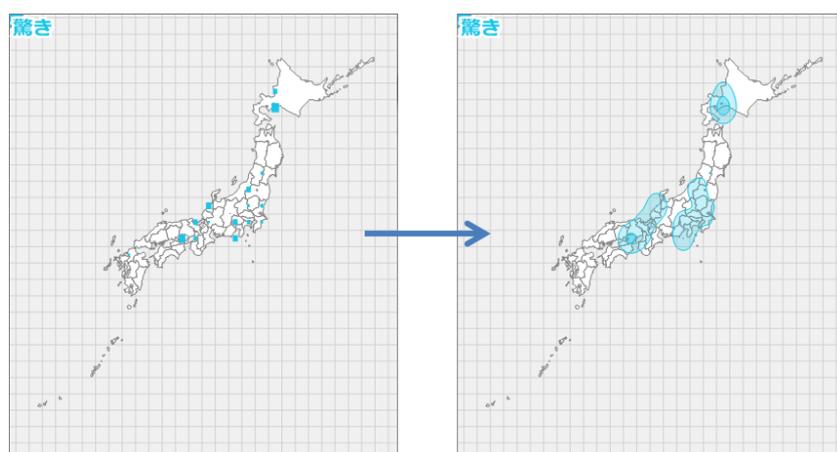


図 22 感情スコアの表示

(左) 領域を塗りつぶす範囲で表現 (右) 2次元メタボールレンダリングで表現

以降、具体的な描画方法について示す。メタボールの境界線を描画するために、**Marching Squares** 法を用いる。まず、地図上に 2 次元のスカラー場を設ける。本研究では地理的領域を経緯 1 度ずつに分割するのに対して、スカラー場はその 1/4 となる経緯 0.25 度ずつに座標を取る。これが小さいほど滑らかな線が描けるが、計算コストが高くなる。続いて、このスカラー場に、感情スコアを影響力の大きさ、座標を地理的領域の中心とする 2 次元メタボールを追加していく。ただし、その影響範囲はある円の中に限り、その半径は全てのメタボールで統一で、地理的領域の分割より大きな値とする。本研究では地理的領域を経緯 1 度ずつに分割するのに対して、影響範囲を 1.5 倍の 1.5 度を半径とする円とする。これは滑らかに緩くつなげるための倍率で、大きいほどつながり方が緩くなる。1 倍未満にしてしまうと、隣り合う領域と影響し合わないため、緩くつなげる目的が達成されない。ただし、領域の分割の際に重複を許している場合はこの限りではない。1 つのメタボール m についての中心点からあるスカラー場の座標 p までの距離 l と影響力の関係を表す濃度分布関数は、次の式で与えることにする。

$$f(m,l) = \begin{cases} S_m \times \frac{(r-l)^2}{r^2}, & l \leq r \\ 0, & r < l \end{cases}$$

ここで、メタボール m の影響力が S_m 、影響範囲の半径が $r (= 1.5^\circ)$ である。ここでいう距離は、経緯度の格子上での距離であって、実際の地理的距離でない。グラフで表すと図 23 となる。このようにメタボールからの距離が遠いほど受ける影響力は小さくなる。

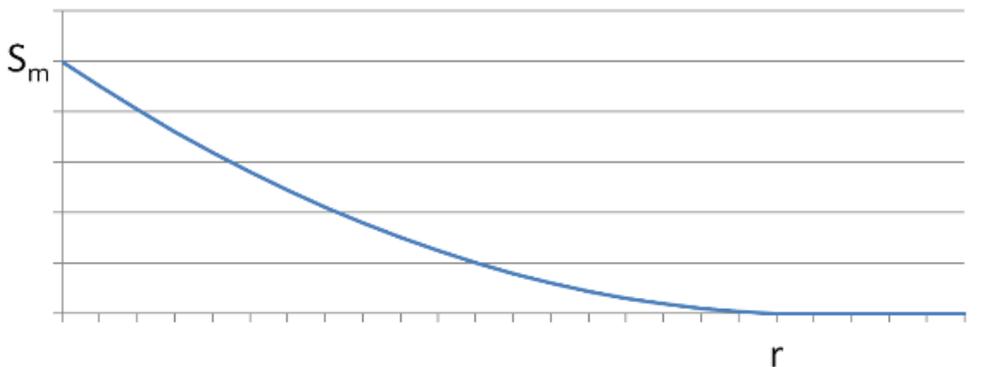


図 23 影響力と距離の関係

最終的にあるスカラー場の座標 p における影響力は、次の式のように全てのメタボールによる影響力を合計したものになる。

$$S_p = \sum_n f(m_n)$$

続いて、メタボール境界線を算出する。1つのメタボールが持ちうる影響力の大きさ、つまり感情スコアの値が0~1なのに対して、境界を作る閾値には $1/5 \cdot 2/5 \cdot 3/5 \cdot 4/5$ の4段階を採用する。ある座標 p とその右・下・右下の4つの座標からなる正方形で、それぞれの座標の影響力が閾値を超えているかいないかで、境界線の一部となる線分が引かれるかどうかが決まる。4点に対して、そのパターンは $2^4 = 16$ 通りのみに想定される。閾値を超えているものと超えていないもの間に線分の端が来ることになり、その2つの影響力の値の比をもって、線分端の詳細な座標が導かれる。こうして境界線となる線分を各4点から算出していき、つながるものをつなげることで領域が導かれる。1つの閾値をもとに得られた複数領域が一部包括関係にある場合がある。その領域全ての排他的論理和のとなる領域がその閾値以上の影響力となる領域である。

これらの領域を、閾値 $1/5$ で得られる領域を不透明度 $1/5$ で、閾値 $2/5$ の領域を不透明度 $1/4$ で、閾値 $3/5$ の領域を不透明度 $1/3$ で、閾値 $4/5$ の領域を不透明度 $1/2$ で塗り重ねることで、最終的に、不透明度 $1/5 \cdot 2/5 \cdot 3/5 \cdot 4/5$ で塗ったときと同じ不透明度で描画することができる。ただし、アルファブレンディングの結果の色が、一般的なディスプレイでは表現できない $L^*a^*b^*$ 色空間の値となる可能性がある。また、境界線は不透明度100%の色で描画する。

(4) 表示領域への適用

表示領域への適用は予備実験2と同様である。画面を8つに分け、「喜び」を始めとして「好き」・「恐れ」…と、感情環の順番でそれぞれの主題地図を描いていく。

6.1.3 時空間分布の分析

感情情報の時空間的分布を分析するために時空間分布ビューを提示する。

(2) 提示データへの変換

生データの分割・集積は、予備実験2の手順に準ずる。

(3) 視覚的表現への対応付け

地理的領域とそこでの感情スコアについては、空間分布ビューと同様に2次元メタボールレンダリングによる領域提示を行う。

時間帯については、ビューを切り替えていくことで表現する。

(4) 表示領域への適用

表示領域への適用についても、空間分布ビューと同様である。

ただし、時間帯のビューを切り替えるには、閲覧者が操作することで行う。

6.2 補助的機能

実際の分析において、前述の視覚提示では不十分なものがある。本節では、それを補うために導入した機能について紹介する。

6.2.1 感情天気記号の提示

空間ビュー・時空間ビューのいずれかの地図上にて、マウスで左クリックしたとき、そこに該当する地理的領域のそれぞれの感情のスコアをパイチャートで表示する。もう一度左クリックすることで非表示にする。実際の例を図 24 に示す。知りたい対象の領域があるとき、感情の主題地図のみではそれぞれが離れているのに対して、パイチャートで示すことで感情スコアの比率を把握しやすくなることを狙ったものである。

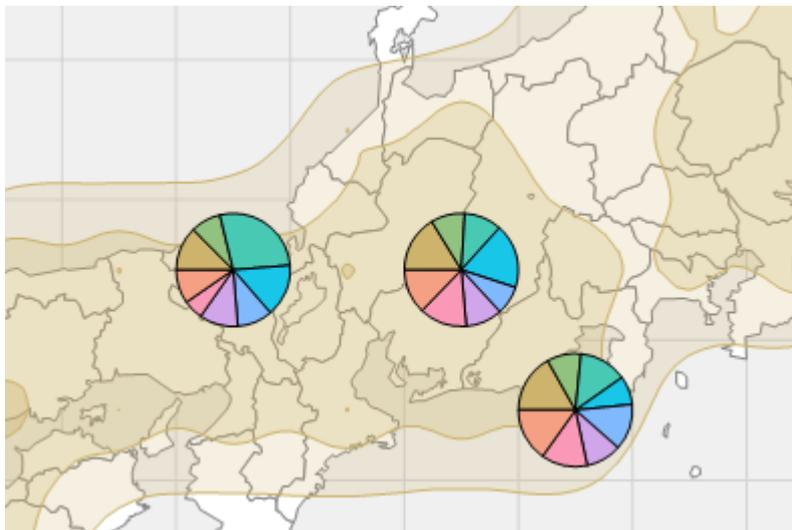


図 24 感情天気記号の表示

6.2.2 トピックの提示

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。

投稿文章に自然言語処理をする過程で、ハッシュタグ*と品詞が一般名詞とサ変接続名詞で、かつ感情語でない単語を控えておき、感情語と共起したものをその感情の対象のトピック候補として新たにデータを持つ。具体的には、(感情の種類、時間帯)の組とそれに該当する(トピック候補、その頻度)、(感情の種類、地理的領域)の組とそれに該当す

* Twitter において、投稿者が投稿文章中で話題を明確に提示する表記

る（トピック候補、その頻度）、（感情の種類、時間帯、地理的領域）の組とそれに該当する（トピック候補、その頻度）という3種類のデータである。

（感情の種類、時間帯）の組については、その感情スコアが4/5以上になる組で、トピック候補の頻度が最も高く、かつ2以上のものをその組のトピックと定義する。これは時間分布ビューに、その感情の種類・時間帯に対応する場所に吹き出しで重畳表示する。

（感情の種類、地理的領域）と（感情の種類、時間帯、地理的領域）の組については、メタボールレンダリングの結果4/5以上を示す領域で、トピック候補の頻度が最も高く、かつ2以上のものをその領域のトピックと定義する。前者は空間分布ビューに、後者は時空間分布ビューに吹き出しでその領域を代表する場所に重畳表示する。図25は「恐れ」の感情において、京都府の辺りに「大雨」としてトピックが表れている様子である。

トピックの重畳表示をするかしないかはGUIによって切り替えることができる。

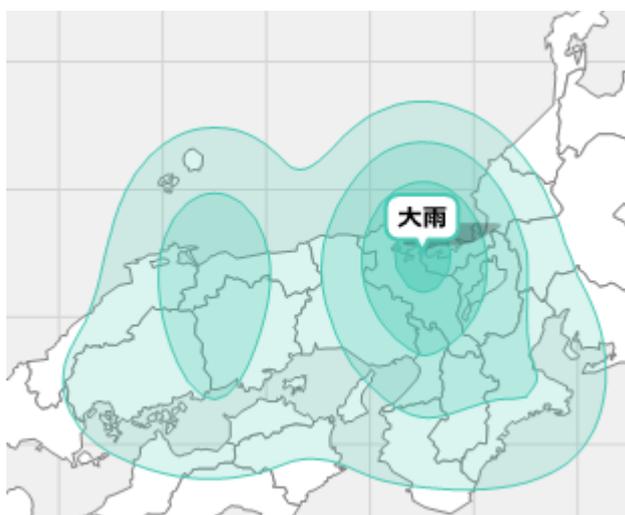


図25 トピックの表示

6.2.3 投稿の詳細ウィンドウ

それぞれのビューにおいて、右クリックして「この部分の詳細」という項目を選ぶと、右クリックした位置が該当する組に対応する投稿を「投稿の詳細」ウィンドウに表示する。このウィンドウには、該当した組の情報（上部）と投稿の表（中部）、投稿によるトピック候補のタグクラウド（下部）が表示される。実際のを図26に示す。

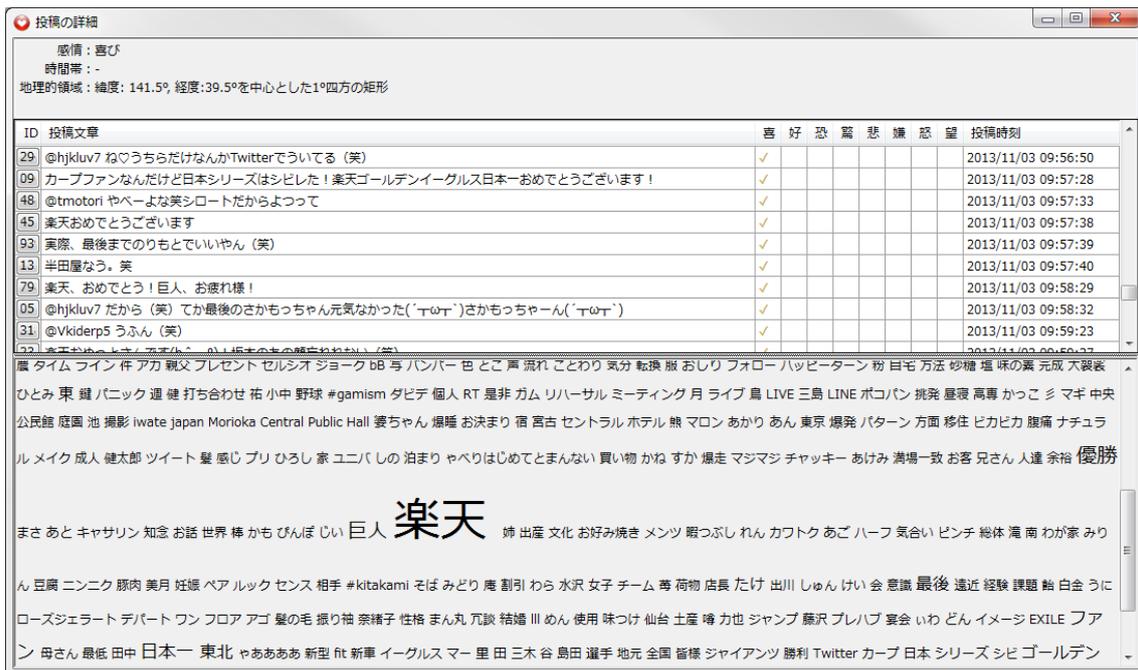


図 26 投稿の詳細ウィンドウ

6.3 実装

「感情天気図」の自動作成は3つのプログラムからなる。生データとして地理情報付きの投稿を収集するクローラと、投稿文章を単語に分割すると同時に投稿文章から感情を抽出するデータ変換器、データを集約したのち設計をもとに描画を行うビューワである。どのプログラムも、開発言語はC#で、ライブラリに.NET Framework 4.0を用いている。また、前述の通り、文章を単語に分割するにはMeCab[22]を、地理情報のデータはクリエティブコモンズのもの[25]を用いている。

第7章 適用例

「感情天気図」の有用性を示すため、この章ではいくつかの適用例を示す。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。

7.1 前処理

本章では2つの事例を紹介するが、2つの事例ともに、感情抽出の際にその感情を含むとするのにふさわしくないと考えたものについて、前処理をしている。処理の対象は、ある Twitter クライアントで特定のボタンを押すと投稿される「きつねかわいい!!!」という文章から始まる投稿であり、この投稿は遊び目的で連続して行われることがしばしばある。本来「かわいい」は「好き」の感情語としているが、この投稿は「好き」という感情を表現するために行う投稿ではないと判断し、投稿が「きつねかわいい!!!」から始まる文章である場合は、感情を含まないものとして処理した。ただし、投稿の総数には含む。

7.2 事例 1:地震

7.2.1 生データと提示データへの加工

事例1で扱うのは、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大きな地震があった。

生データから提示データへの分割は、時間帯として午前0時を基準に1時間ずつ、地理的領域として緯度0度・経度0度を基準に緯度経度1度ずつの矩形とした。

7.2.2 可視化結果の観察と得られた知見

図27～図29に示したのは、感情的・時間的・空間的バイアスがかかったままのデータを可視化したものである。

図27は時間分布ビューである。絶対数としては「喜び」が多く、また長い時間にわたってある程度分布していることが分かる。

図28は空間分布ビューである。いくつかの感情が都市部で高いことが分かる。また、ここでも「喜び」の感情スコアの高さが分かる。

図29は時間分布ビューの一部で、大きな地震のあった午後0時台を示したものである。

いくつかの感情で、都市部でスコアが高いことが分かるのみである。

図 30～図 33 に示したのは、感情的・時間的・空間的バイアスを全て解消したデータを可視化したものである。

図 30 は時間分布ビューである。午後 0 時台に「恐れ」と「驚き」の感情スコアが高くなっていることが分かる。

図 31 は空間分布ビューである。宮城県を中心に東北で「恐れ」と「驚き」の感情の高まりが見てとれる。また、トピックの表示により、それが主に「地震」によってもたらされた感情であることが分かる。

図 32 と図 33 は時間分布ビューの一部で、それぞれ午後 0 時台と午後 1 時台を表したものである。この 2 つから、「恐れ」の感情スコアが 2 時間にわたって高く広く分布している一方で、「驚き」の感情スコアの高まりは 1 時間かぎりで治まったことが見てとれる。

各種バイアスをそのままにしたものと解消したものとを比べたとき、得られる知見が異なった。全てのバイアスを解消しない場合は感情の種類、時間帯、地理的領域に該当する投稿の絶対的な個数の比のみが表現され、少なくともこのデータの時間的・地理的範囲では、「喜び」の感情が、夕方、都市部で多いことが見てとれた。全てのバイアスを解消した場合は、感情の種類、時間帯、地理的領域として、他のそれらと比べたときに特徴となりえるかという値が表現され、少なくともこのデータの時間的・地理的範囲で、「恐れ」や「驚き」の感情が、午後 0 時台に、宮城県付近で高いことが見てとれた。

また、「恐れ」が 2 時間にわたって続いた一方で「驚き」が 1 時間で治まったように、感情の違いによってその時空間的分布も異なる様子が確認できた。

ただし、「地震」による感情の高まりだと思われる時間帯、地理的領域、およびその組について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。

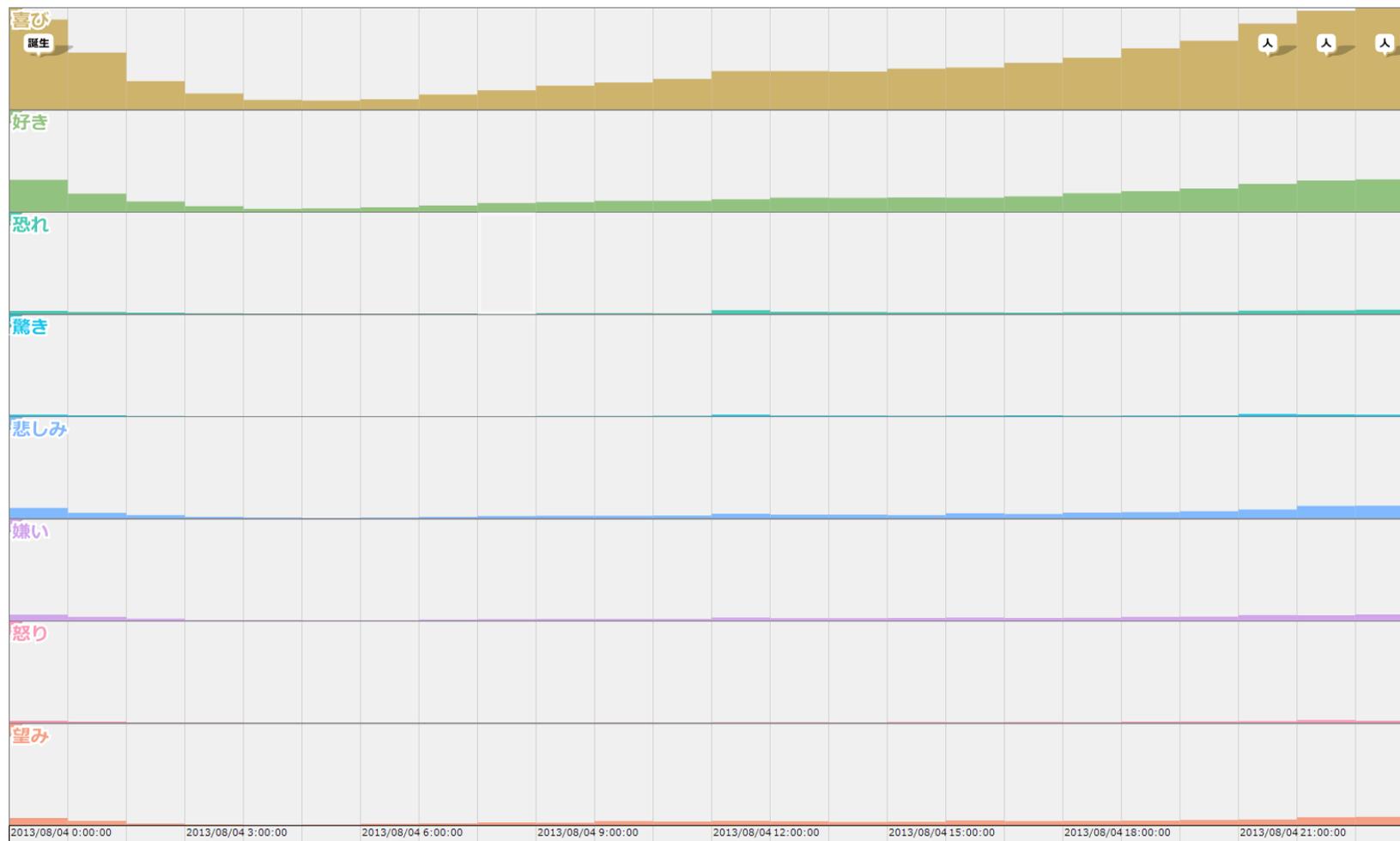


図 27 事例 1 の全てのバイアスがかかった時間分布ビュー

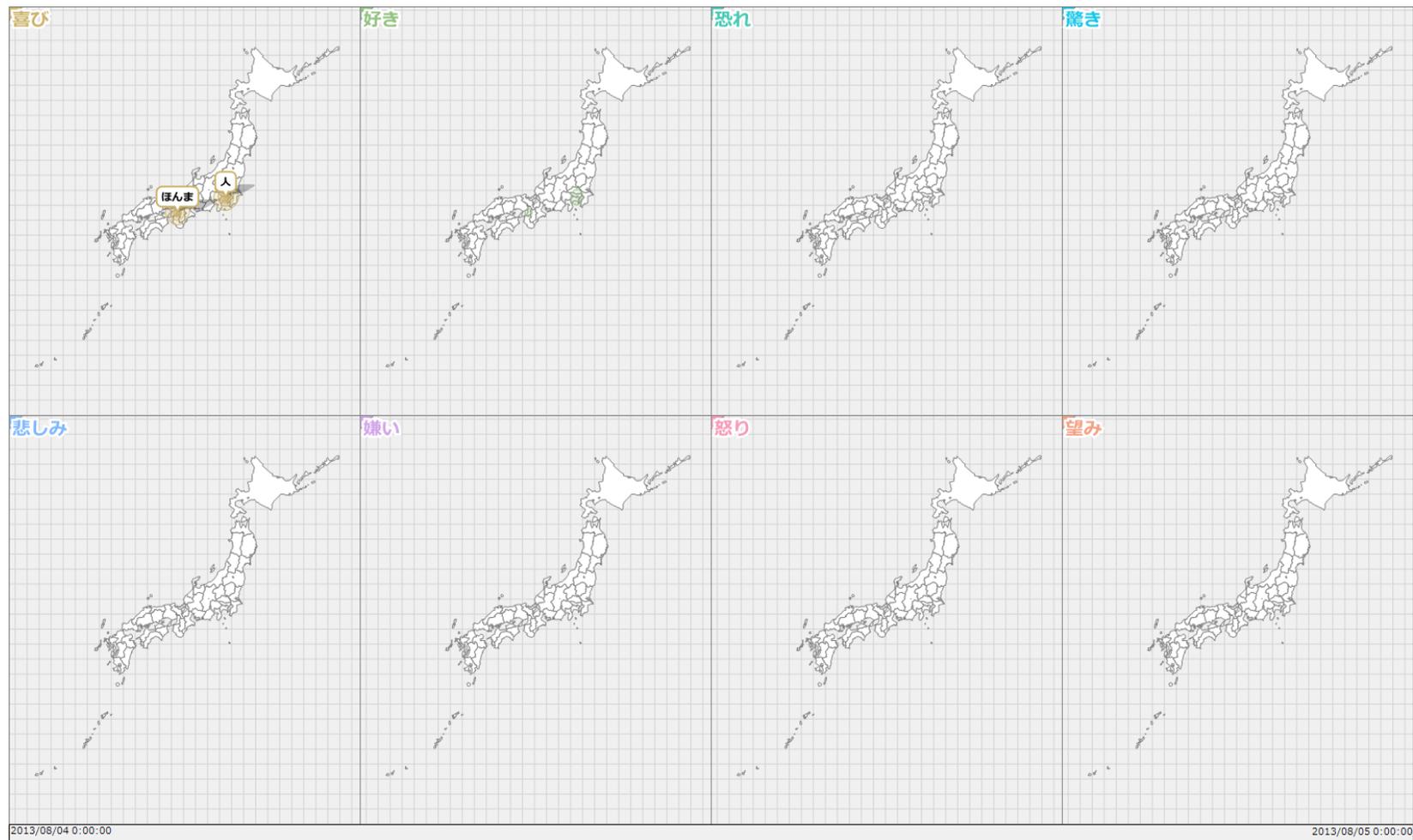


図 28 事例 1 の全てのバイアスがかかった空間分布ビュー

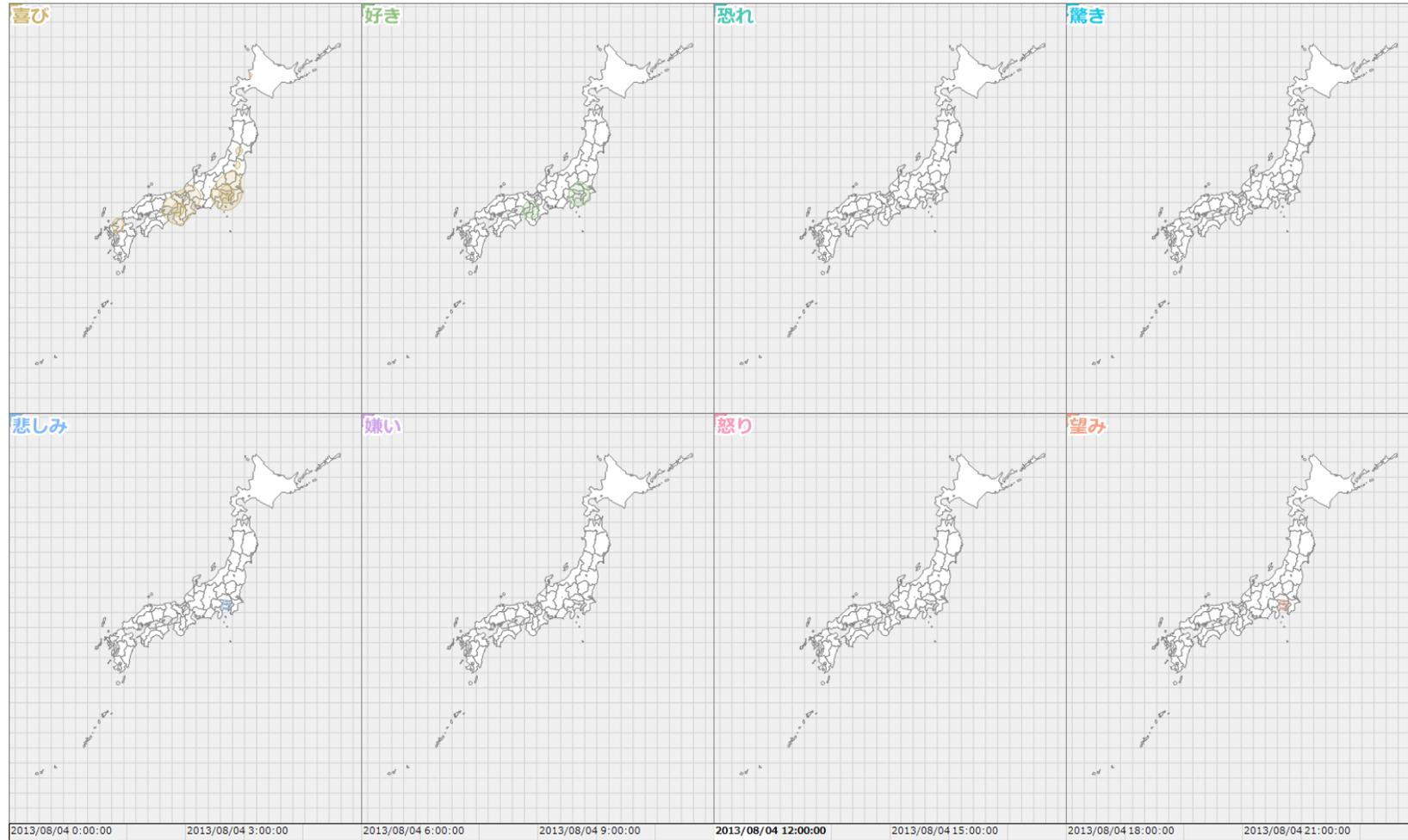


図 29 事例 1 の全てのバイアスがかかった時空間分布ビュー（午後 0 時台）

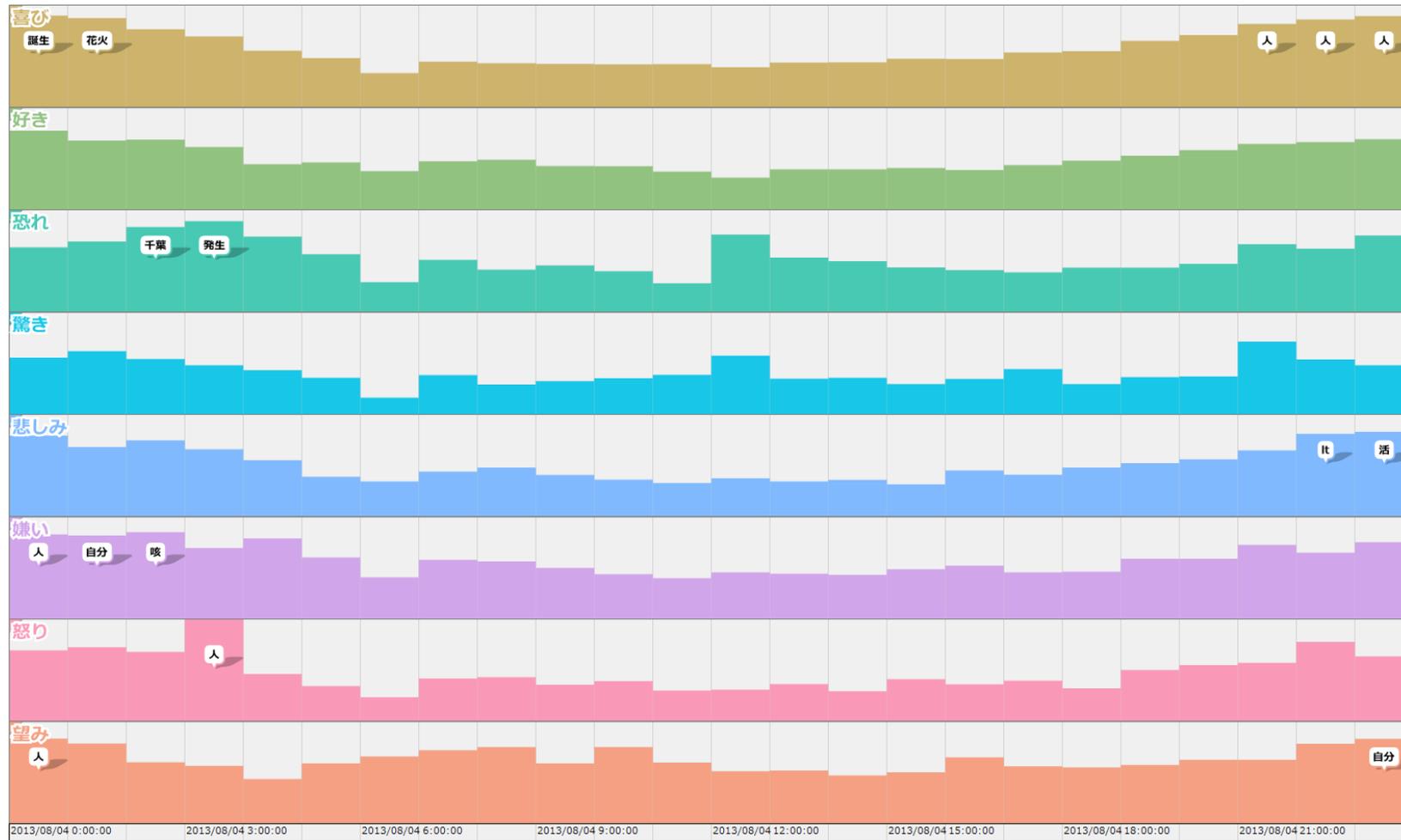


図 30 事例 1 の全てのバイアスを解消した時間分布ビュー

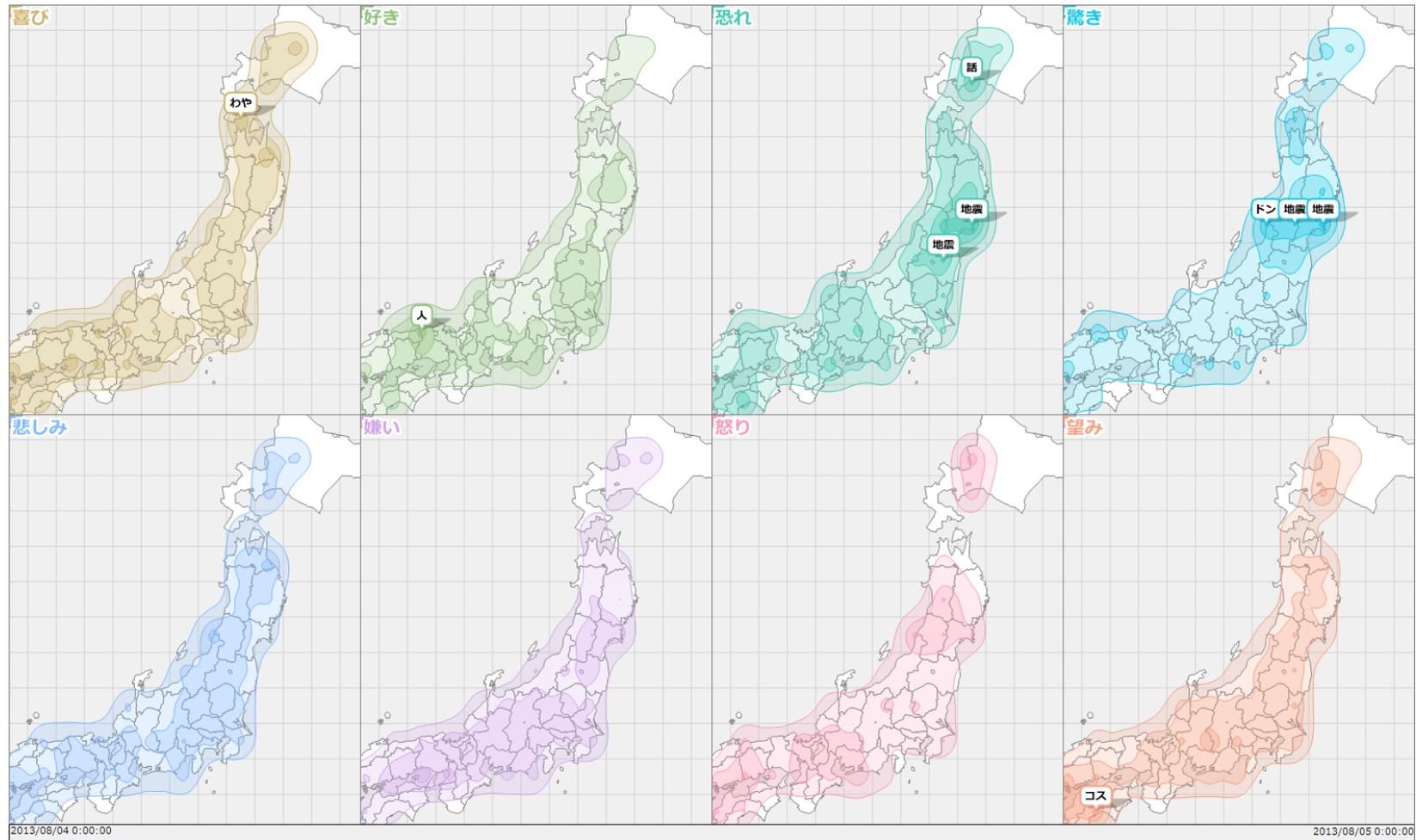


図 31 事例 1 の全てのバイアスを解消した空間分布ビュー

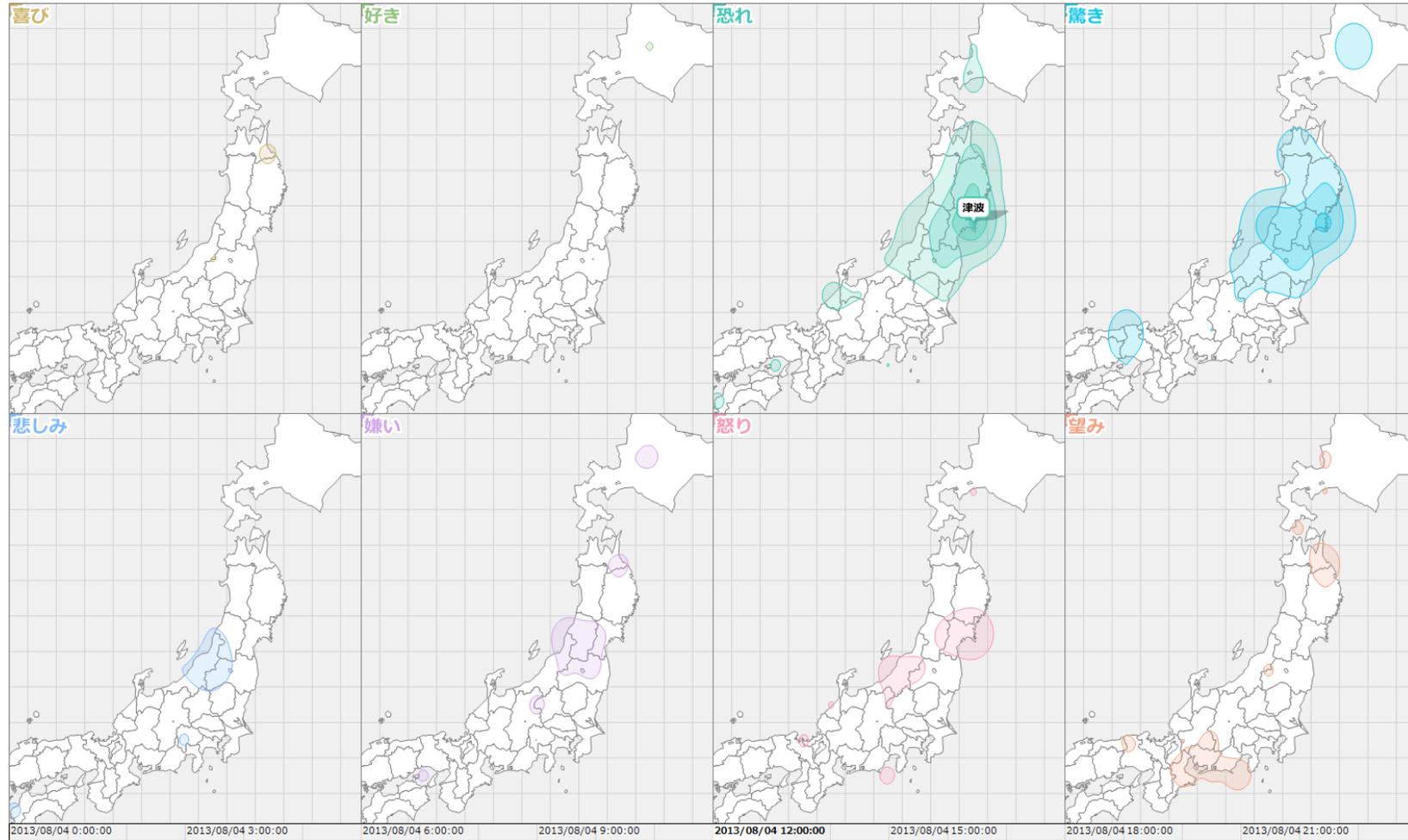


図 32 事例 1 の全てのバイアスを解消した時空間分布ビュー（午後 0 時台）

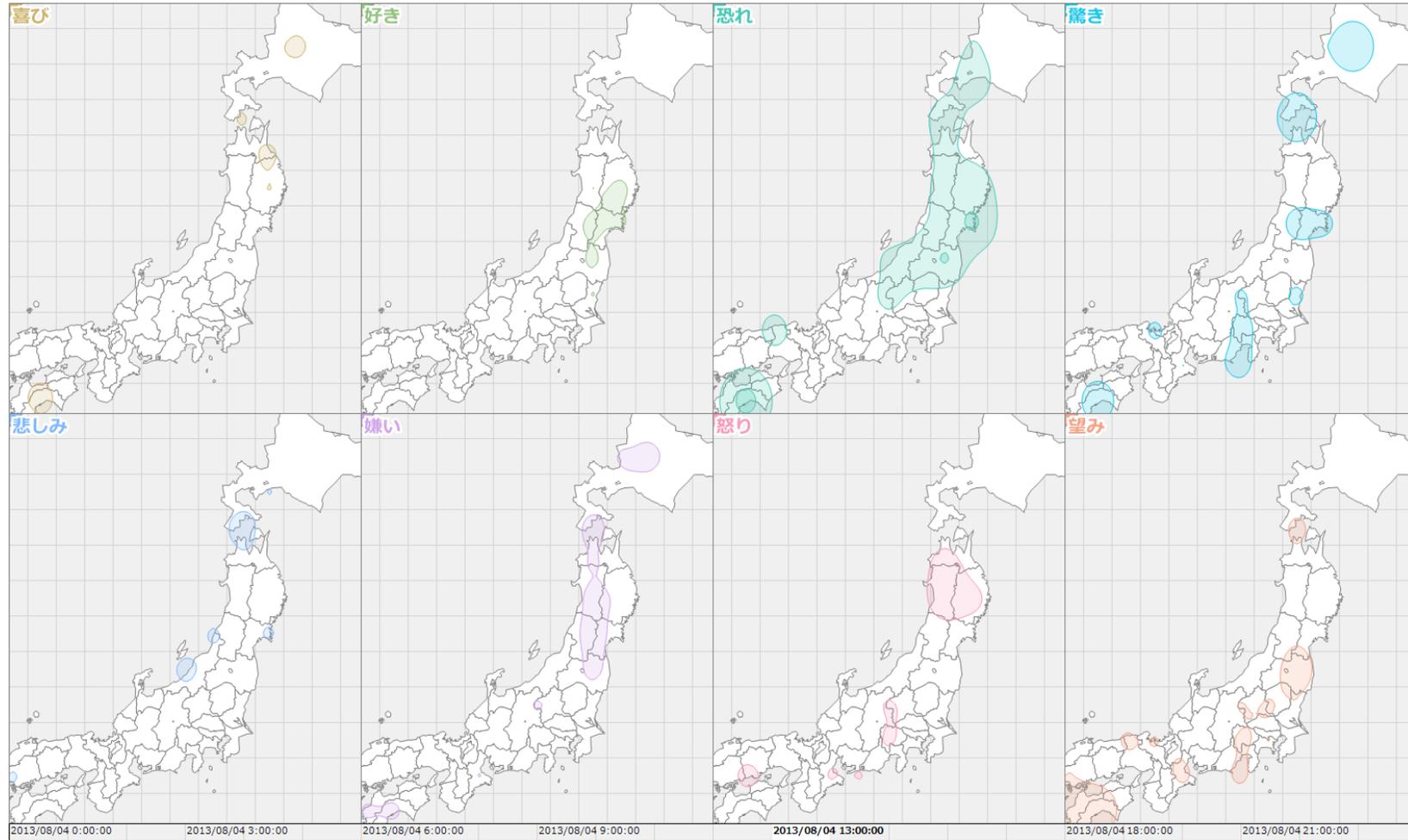


図 33 事例 1 の全てのバイアスを解消した時空間分布ビュー（午後 1 時台）

7.3 事例 2:高校野球大会

7.3.1 生データと提示データへの加工

事例 2 で扱うのは、2013 年 8 月 22 日午前 0 時から、翌日 23 日午前 0 時までの 24 時間分である。投稿総数は 205,660 個であった。この日は、第 95 回全国高等学校野球選手権大会の決勝戦が行われた。群馬県と宮崎県の代表校が戦い、群馬県の代表校が勝利した。

生データから提示データへの分割は、時間帯として午前 0 時を基準に 1 時間ずつ、地理的領域として緯度 0 度・経度 0 度を基準に緯度経度 1 度ずつの矩形とした。また、感情的・時間的・空間的バイアスは全て解消した。

7.3.2 可視化結果の観察と得られた知見

図 34～図 38 に可視化結果の一部を示す。

図 34 は時間分布ビューである。午後 11 時台の「悲しみ」の感情スコアが高くなっていることが分かる。図 35 は午後 11 時台の「悲しみ」についての投稿の詳細ウィンドウである。これを見ると、日中にあった試合のテレビ番組が放送されたことが大きく影響していることが分かった。一方で勝敗の決まった午後 2 時台は「悲しみ」の感情スコアが特に高くなった様子は確認できない。

図 36 は空間分布ビューである。宮崎県辺りで「悲しみ」の感情スコアが高くなっている様子が分かる。宮崎県の代表校が決勝戦で敗れてしまったことを惜しむ様子が表現されている。

図 37 は勝敗の決まった午後 2 時台を表した時空間分布ビューである。九州地方、なかでも特に宮崎県辺りでの「悲しみ」の感情スコアが高くなっていることが分かるが、同時に「喜び」と「好き」の感情スコアもわずかながら高いことが見てとれることから、宮崎県の辺りの投稿を総合した感情としてアンビバレントな感情を抱いていたことが分かる。また、代表校が勝利した群馬県辺りでの「喜び」の感情スコアも高まりも見てとれる。

図 38 は前述のテレビ番組が放送された午後 11 時台を表した時空間分布ビューである。ここでは「悲しみ」の感情スコアは高めだが、それが特定の地域ということではなく、この時間帯は日本全体が「悲しみ」を抱いたことが分かる。

このように、大会決勝戦による大衆感情の分布を一部見てとれた。アンビバレントな感情は、「喜び」と「悲しみ」を別次元として設計したからこそ出てきたものである。しかし、「喜び」の感情スコアはもっと高くなることを期待した。これは、「喜び」に分類される投稿数自体が多く、その状態からさらに大きな「喜び」となるにはさらに多くの投稿数が必要になることから生じたものだと考えられる。

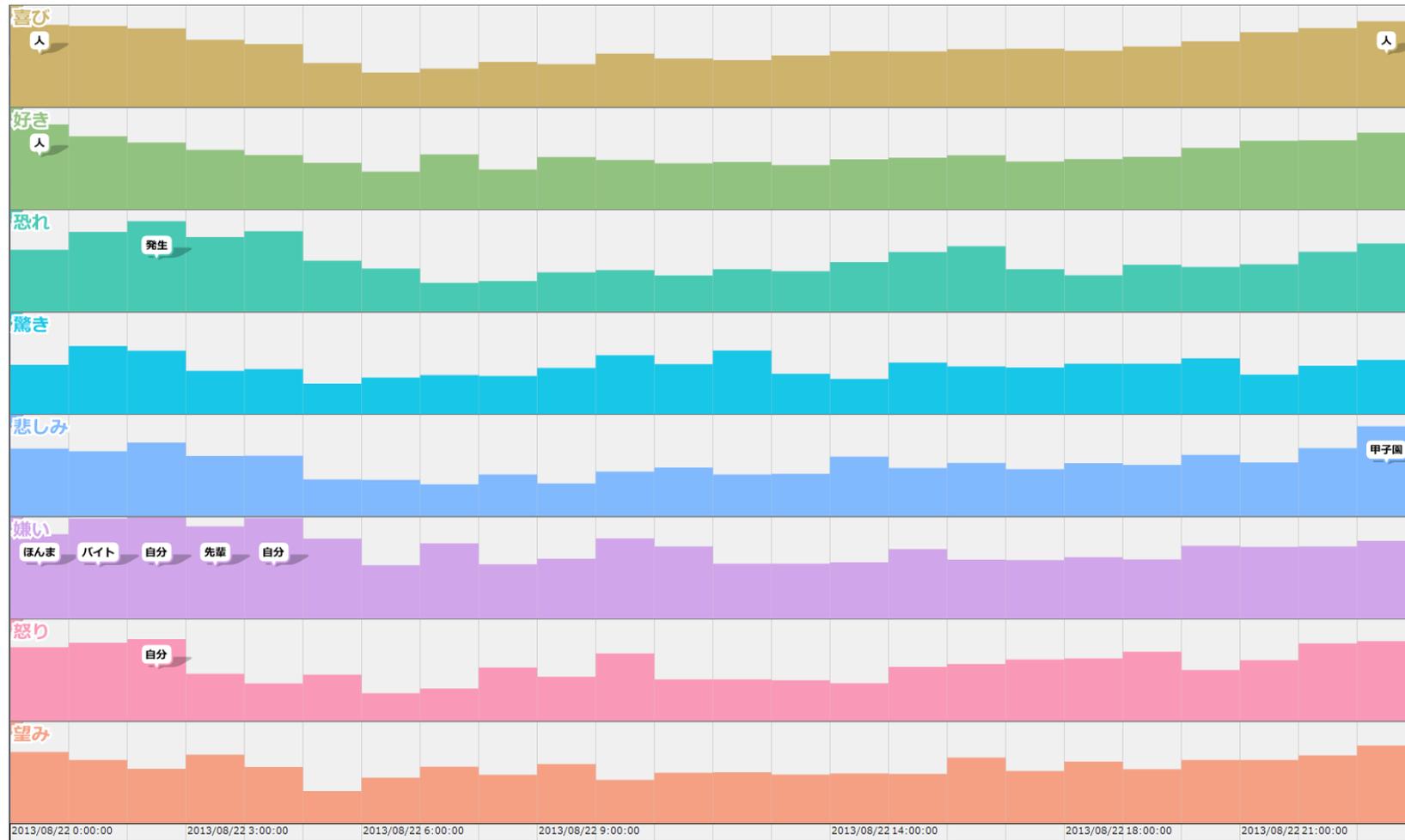


図 34 事例 2 の時間分布ビュー

投稿の詳細

感情: 悲しみ
 時間帯: 2013/08/22 23:00:00 ~ 2013/08/23 0:00:00
 地理的領域: -

ID	投稿文章	喜	好	恐	驚	悲	嫌	怒	望	投稿時刻
48	@mv_yuna あたしが泣いても一たならななの応援が足りなかったからな、笑笑	✓				✓				2013/08/22 11:33:11
00	ほんまにごめん、うざいって。そこの2人 まちでやること子供でしょーもないことせんとちやっちゃ考え方だけでも大人になれぼ?					✓				2013/08/22 11:33:38
11	使い方わかんなくてきつねかわいい! 連呼したらごめんね		✓			✓				2013/08/22 11:33:38
12	あかん熱闘甲子園泣いた					✓				2013/08/22 11:33:55
91	これ観て泣かない人いるんかなー					✓				2013/08/22 11:33:57
40	工藤の涙にやられる件ww					✓				2013/08/22 11:34:05
05	熱闘甲子園マジ泣いた					✓				2013/08/22 11:34:52
01	@kyosuke0908 見てん! 4sの子みんな今家で泣いとるで口 (爆笑)	✓				✓				2013/08/22 11:35:16
00	もーやばいなw高校野球泣いてるの見たら鬼泣けてくるw浦和のピッチャー泣けるwどこのチームもすんごいな♪					✓				2013/08/22 11:35:33
52	ほんま喧嘩ばっかやな記念日やったのになほんまごめんやで					✓				2013/08/22 11:35:41
04	@canytk0411 あ、はい、わら ごめんやん? 言うてしらたまのスタンプ攻撃いらっとする、わら					✓				2013/08/22 11:35:47
64	熱闘甲子園ほんま毎年感動しちゃうよね。涙ぼろぼろ。あー。ステキ。					✓				2013/08/22 11:35:48
16	なんか小春やっこいみたいやけどなんもできんね。ごめんておかしいけどこーゆー時どーするもんなん??					✓				2013/08/22 11:36:11
38	@09Miie yuyupで検索してみて(((o(*マ*o)))明日行きたいところやねんけど予定はいつてるねーん(>_<);まちごめんっ(>_<);					✓				2013/08/22 11:36:16
20	長崎は、明子と花見の写真が胸にイキッてくるのが残念だわ					✓				2013/08/22 11:36:54

甲子園

い 延 ライダー キック 凄み かつこ ゼクロスメイン 戦い 派手め スカイ オムニバス 話 っくやしいのっ 楽天 夏 準備 Dirty ぽにつくろっ BY 雷太 女 わが

熱闘

まま 一緒 既読 作り方 ぎだろっ 鼻水 胸 キュン 自身 道 人間 何とか チャニヨル #ハリネコ沙知さんの名字を大胆予想 月亭 林家 三遊 亭 お知らせ 祭り 出店 gt lt リブ

図 35 事例 2 の投稿の詳細ウィンドウ

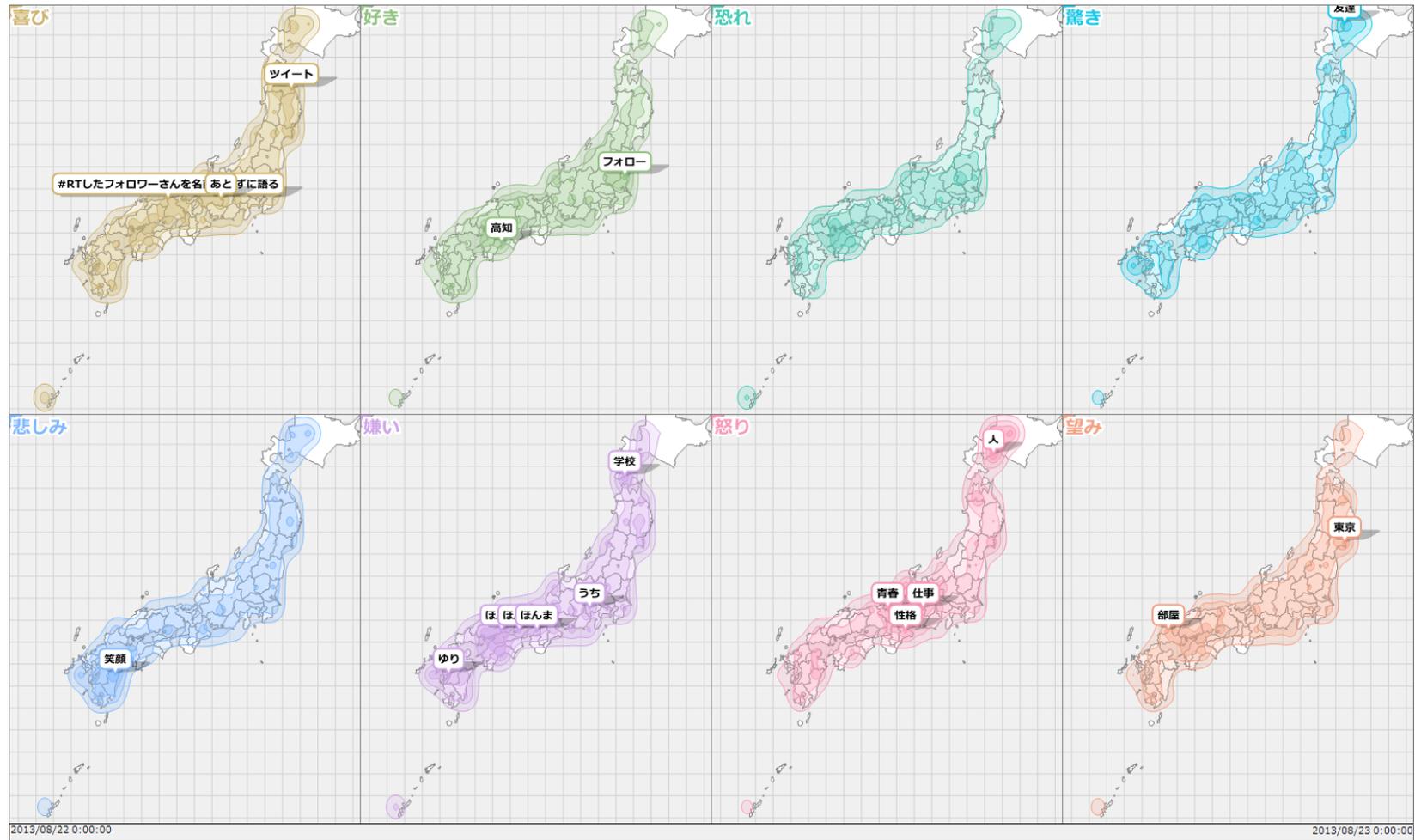


図 36 事例 2 の空間分布ビュー

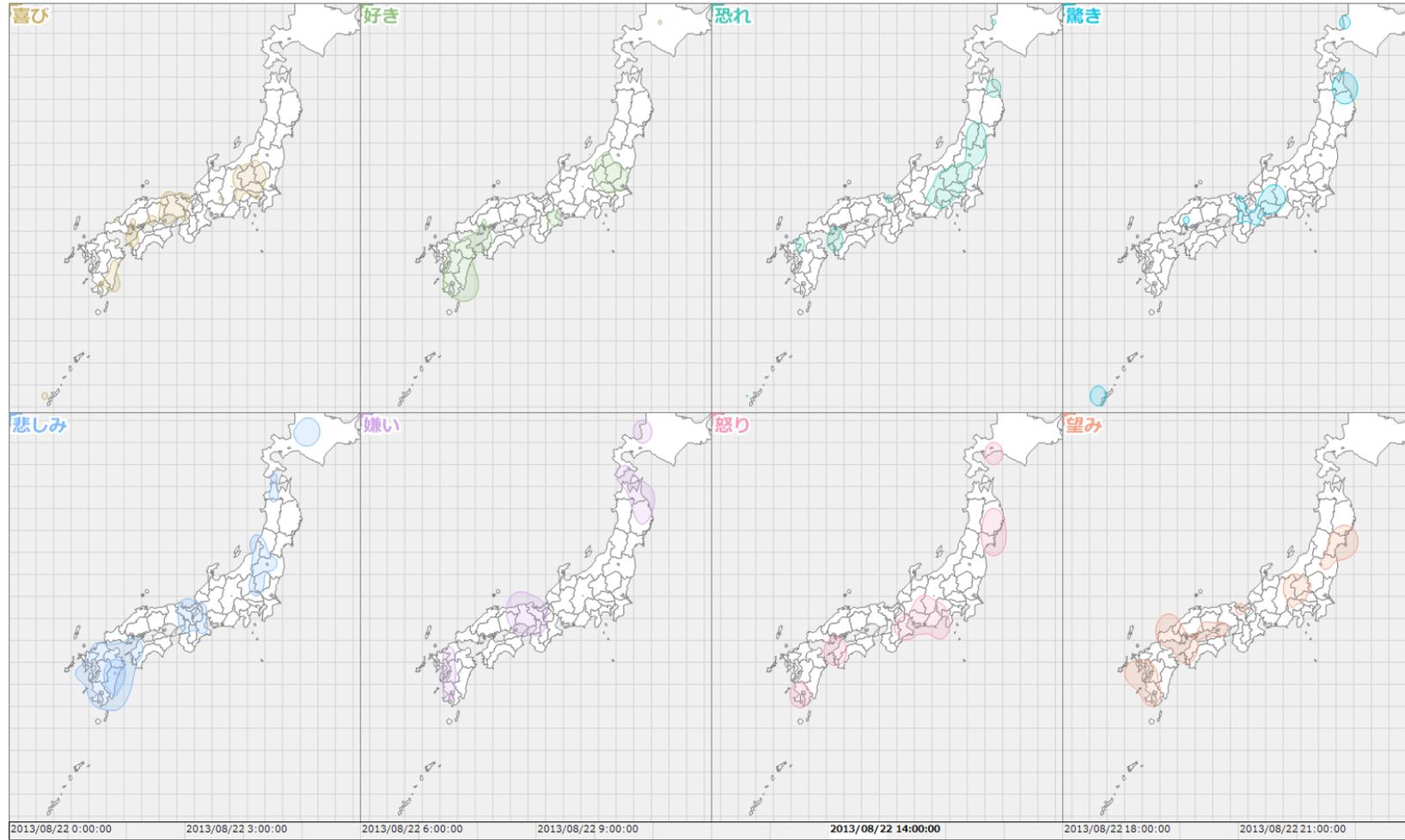


図 37 事例 2 の時空間分布ビュー（午後 2 時台）

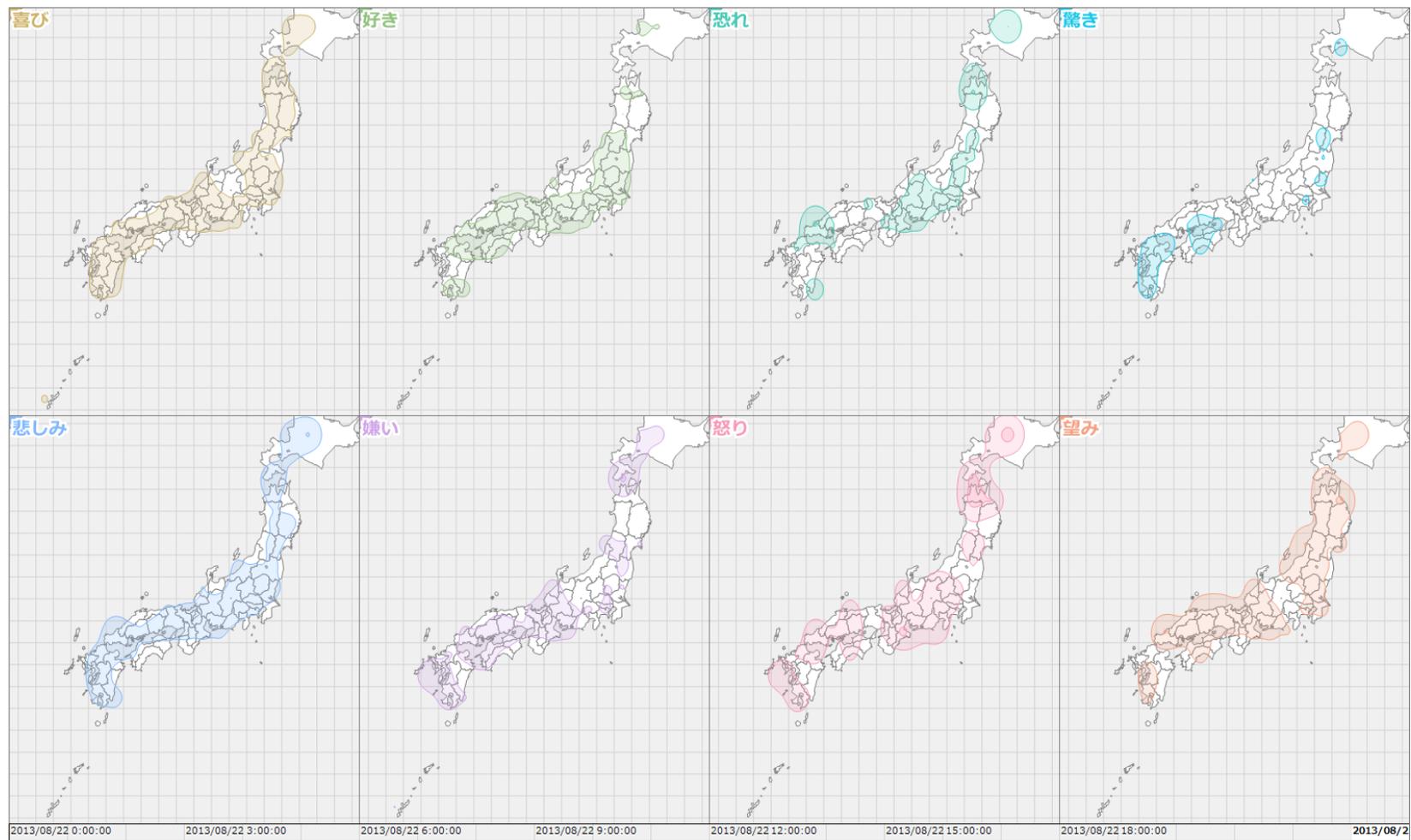


図 38 事例 2 の時空間分布ビュー（午後 11 時台）

第8章 まとめ

本論文では、世界中の人々が抱く気持ちを捉えることを目的に、感情の可視化をするときに生じる問題について検討する予備実験を2つ行い、最終的に日本で大勢の人々が抱く感情が時間的・空間的にどう分布しているのかを把握することを目指した「感情天気図」の設計および開発をした。感情情報を収集しただけのデータでは、そこに感情的・時間的・空間的バイアスの存在が考えられる。それを解消すべきかどうかという選択肢があることと解消するためのデータの変換方法を示した。視覚的表現においては、感情の種類を知覚的等距離にある8つの色に対応付けることで、Plutchikが示した感情同士の関係を再現し、その感情スコアを2次元メタボールレンダリングをすることで、閲覧者がその地理的広がりを把握しやすくなることを狙った設計をした。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、実データの適用をすることで、「感情天気図」自体とその設計の有用性について議論した。感情によって時空間的分布が異なることが確認できた。

続いて、感情の主題地図の自動作成としての課題を述べる。まず、データの変換方法について述べる。例えば、「喜び」の感情はその投稿数自体が多く、「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、算出した感情スコアはデータセット内での特徴量であり、感情的変化があるだろう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。感情天気図では、バイアスを解消することで普遍状態との差を見いだすことを狙ったが、それとは違う、「喜び」の感情スコアがいつでも高めになることを踏まえた新たな普遍状態を定義し、その差を見せるデータ変換方法を開発することが課題である。次に視覚的対応付けについて述べる。一般の人々をエンドユーザと仮定したとき、色覚異常に配慮するべきであるが、今回用いた配色は色覚異常を持つ人には判別が難しい。これには、ハッチングと呼ばれる模様を付けるなどの対応が必要である。また、感情それぞれの主題地図を描くことで提示を行ったが、8枚の地図を描くにはスペースを要する。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の開発が課題である。さらに、本研究は時間軸の対応付けにビューを切り替えるという時間を対応付けている。時間への対応付けは分析にそのまま時間を要することになる。感情の主題地図の要件3として挙げた、「時間にしてどのぐらいの間、その感情が特徴的であり続けるか」という情報を、ほかの対応付けを妨げることなく、より短い時間で表現する方法の開発が課題として挙げられる。

また、本研究はデータ変換による加工と、感情の分類方法とその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。今後の課題として、より精度の高い感情抽出方法によって得られるデータの適用が望まれ

る。また、本研究では Plutchik が提唱する基本感情 8 つのみを対象にしている。Plutchik によれば、感情には分類のほかに、その強さがあると述べている。また、例えば絶望は恐れと悲しみにと、全ての感情は基本感情に分解できるとも述べている。自然言語処理による抽出の段階で、その強さの算出と分解を行うことが課題である。これが達成されれば、Plutchik が提唱する分類方法により近く沿った結果が得られるだろう。さらに本研究では、分析対象を Twitter への投稿、その中でも日本の投稿に限定している。本来、感情の高まりは誰でもいつでもどこでも発生するあるいは表出されるものである。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」が出来上がることに期待する。

謝辞

本研究を行うにあたり、指導教員である三末和男先生には、丁寧なご指導と適切なお助言をいただきました。心より感謝申し上げます。また、田中二郎先生、志築文太郎先生、高橋伸先生、**Simona Vasilache** 先生、嵯峨智先生には、ゼミを通して貴重なお助言をいただきました。深く感謝申し上げます。

また、インタラクティブプログラミング研究室の皆さまには、たいへんお世話になりました。そのご厚意に感謝申し上げます。

最後に、さまざまな面から支えてくれた家族に、心より深くお礼申し上げます。ありがとうございました。

参考文献

- [1] Ming C Hao, Christian Rohrdantz, Halldor Janetzko, Daniel A Keim, Umeshwar Dayal, Lars erik Haug, Meichun Hsu and Florian Stoffel, “Visual sentiment analysis of customer feedback streams using geo-temporal term associations”, Information Visualization Volume 12, pp.273-290, 2013.
- [2] Eric Shook, Kalev Leetaru, Guofeng Cao, Anand Padmanabhan and Shaowen Wang, “Happy or not: Generating topic-based emotional heatmaps for Culturomics using CyberGIS”, 2012 IEEE 8th International Conference on E-Science, pp.1-6, 2012.
- [3] Hansen Andrew Schwartz, Johannes Eichstaedt, Richard Lucas, Lukasz Dziurzynski, Margaret L Kern, Gregory Park, Megha Agrawal, Shrinidhi K Lakshmikanth, Shneha Jha, Martin Seligman and Lyle Ungar, “Characterizing Geographic Variation in Well-Being using Tweets”, Seventh International AAAI Conference on Weblogs and Social Media (ICWSM 2013), 2013.
- [4] 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点差異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.
- [5] ツイッター社, “Twitter”, <http://twitter.com/>, 2006-, 閲覧: 2013/1/15.
- [6] Tetsuro Takahashi, Shuya Abe and Nobuyuki Igata, “Can twitter be an alternative of realworld sensors? ”, Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments, pp. 240–249. 2011.
- [7] ニフティ株式会社, “ニフティ、「みんなの花粉症なう！β」を提供開始 | ニュースリリース | ニフティ株式会社”, <http://www.nifty.co.jp/cs/newsrelease/detail/100303003821/1.htm>, 2010-, 閲覧: 2013/1/15.
- [8] 株式会社ウェザーニューズ, “SAKULiVE ニッポンの桜をみんなでライブ！ ウェザーニューズ”, <http://weathernews.jp/sakura/sakulive/>, 2011-, 閲覧: 2013/1/15.
- [9] Joel Robert Davitz, “The language of emotion”, Academic Press, 1969.
- [10] Carroll Ellis Izard, “The Face of Emotion”, Appleton-Century-Crofts., 1971.
- [11] Robert Plutchik, “Emotion: A Psychoevolutionary Synthesis”, Harper & Row, 1980.

- [12] Tetsuya Nasukawa and Jeonghee Yi, “Sentiment Analysis: Capturing Favorability Using Natural Language Processing”, Proceedings of the 2nd international conference on Knowledge capture, pp.70-77, 2003.
- [13] Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka, “Textual Affect Sensing for Sociable and Expressive Online Communication”, Affective Computing and Intelligent Interaction (ACII’07), pp.220-231, 2007.
- [14] 菅原 久嗣, “感情語辞書を用いた日本語テキストからの感情抽出”, 東京大学大学院情報理工学系研究科電子情報学専攻 修士学位論文, 2010.
- [15] Klaus R. Scherer, “Vocal Affect Expression: A Review and a Model for Future Research”, Psychological Bulletin Volume 99 Issue 2, pp.143-165, 1986.
- [16] Bethany McDaniel, Sidney D’Mello, Brandon King, Patrick Chipman, Kristy Tapp, and Art Graesser, “Facial Features for Affective State Detection in Learning Environments”, Proceedings of 29th Annual Meeting of the Cognitive Science Society, pp.467-472, 2007.
- [17] Christine Lætitia Lisetti and Fatma Nasoz, “Using Noninvasive Wearable Computers to Recognize Human Emotions from Physiological Signals”, EURASIP Journal on Applied Signal Processing, pp.1672–1687, 2004.
- [18] Jeonghee Yi and Wayne Niblack, “Sentiment Mining in WebFountain”, Proceedings of the 21st International Conference on Data Engineering (ICDE 2005), pp.1073-1083, 2005.
- [19] Gilad Mishne and Maarten de Rijke, “MoodViews: Tools for blog mood analysis”, AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW 2006), pp.153-154, 2006.
- [20] 青島 さやか, 青木 惇季, 宮下 芳明, “60万人の感情実況”, インタラクション 2010, pp.107-112, 2010.
- [21] Stuart K. Card, Jock Mackinlay and Ben Shneiderman, “Readings in Information Visualization: Using Vision to Think”, Morgan Kaufmann, 1999.
- [22] Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama and Kyoko Kanzaki, “Development of Japanese WordNet”, Language Resources and Evaluation Conference 2008, Marrakech, 2008.
- [23] 京都大学情報学研究科・日本電信電話株式会社コミュニケーション科学基礎研究所, “MeCab: Yet Another Part-of-Speech and Morphological Analyzer”,

<http://http://mecab.googlecode.com/>, 閲覧: 2013/1/15.

- [24] Semiocast, “Semiocast — Twitter reaches half a billion accounts — More than 140 millions in the U.S.”,
http://semiocast.com/en/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US, 2012, 閲覧: 2013/1/15.
- [25] “日本の白地図”, <http://www.kabipan.com/geography/whitemap/>, 閲覧: 2013/1/15.
- [26] Robert Plutchik, “Emotions and Life: Perspectives from Psychology, Biology, and Evolution”, Amer Psychological Assn, 2002.
- [27] 金箱 淳一, 藤田 ハミド, “色と音の感情を介したマッピング-マッピング規則の抽出と作曲支援システムへの応用-”, 日本ソフトウェア科学会大会講演論文集 第22巻, 2005.
- [28] 山内 厚志, 寺田 和憲, 伊藤 昭, “動的な発色変数によるロボットの感情モデル”, 第27回ファジィシステムシンポジウム, pp.1015-1020, 2011.
- [29] Wikipedia , “File:Lab color space.png”,
http://en.wikipedia.org/wiki/File:Lab_color_space.png, 閲覧: 2013/1/15.