

東海大学大学院令和2年度博士論文

テキスト情報を含む大規模データの分析と
可視化に関する研究

指導 山本 義郎 教授

東海大学大学院 総合理工学研究科
総合理工学専攻

船山 貴光

目次

第1章	序章	1
1.1	近年のデータ活用の現状と課題	1
1.2	テキストマイニング	1
1.3	インタラクティブな可視化	10
第2章	大規模大学における研究分野の研究実績の可視化	13
2.1	本研究の課題と概要	13
2.2	データセットの作成	14
2.3	トピックモデルを用いた論文の研究領域の推定	15
2.4	自己組織化マップによる研究実績の可視化	23
2.5	自己組織化マップを用いた可視化の改良	28
第3章	災害時の Twitter データの活用	33
3.1	災害時のツイートデータ及び有益情報の可視化システムの開発	33
3.2	災害時のツイートデータの解析と可視化	39
第4章	気象データの活用	66
4.1	研究背景	66
4.2	数値データ提供方法の提案と可視化	68
4.3	任意の領域のデータ活用	73
第5章	まとめと展望	80
	謝辞	82
	参考文献	83
	業績一覧	91

第1章 序章

1.1 近年のデータ活用の現状と課題

諸説あるが2011年頃からビッグデータが企業を始め世界各国で注目され、問題解決に活用される様になった(水田, 2016). ビッグデータは、サイズが大きいデータのみを表している訳ではない。従来のデータ解析は、その解析を目的として収集されたデータを用いたが、ビッグデータ解析では、目的のために収集されたデータだけでなく、そのほかのサービスや外部情報源から得られた多種多様なデータを大規模に解析することで、目的の達成を試みるという違いがある(原, 2017). これを可能にした要因は、近年のコンピューターの発達である。これにより大規模データを解析することが容易になった。また、一般に利用可能なデータは、自治体が公開しているオープンデータだけでなく、TwitterなどのSNSデータや、企業や団体、個人が公開しているデータなど様々なデータがある。ただし、アカウントの作成や契約、営利目的での使用の不可や有償など条件付きの場合もある。これらの多くは、インターネット経由で利用することができる。Webページのボタンをクリックしてファイルをダウンロードする仕様が多くのURLに取得したいデータの情報を決められたフォーマットで記載して送信してダウンロードするWeb APIという仕様もある。Web APIではシステム内で入出力を自動で行うことが可能になる。このように多種多様なデータが利用でき問題解決をすることが可能な時代となった。多種多様なデータを使用できる一方、目的にあったデータの選択や大規模で複雑なデータの理解や解析方法などの課題がある。また、有効的な解析結果をわかり易く発信することも問題解決には重要である。これらの課題を解決することで、ビッグデータという資源を有効活用し、社会の問題解決に活用することができる。

1.2 テキストマイニング

テキストマイニングとは、大量のテキストデータから有益な情報を導き出す技術の総称である(小林, 2017). テキストマイニングには、自然言語処理とデータマイニングの2つの工程がある。自然言語処理とは、日本語や英語などの言語を処理す

る技術や学術分野の総称である(奥野 他, 2016). 最初に行うのが形態素解析である. 形態素解析は, 単語(もしくは形態素)への分割(分かち書き), 品詞の推定, 語形変化の処理(原形を求める)の一連の処理を呼ぶ(工藤, 2018). 英語の正書法には, 単語と単語の間に空白を挿入する「分かち書き」の習慣があるが, 日本語は分かち書きしないため単語に分割する必要があり(金 他, 2003)わかり易さを実装するシステムとして, テキストマイニングツールにおけるシステムの実装に加え, 形態素解析ソフトが, いくつか開発されている. 日本語に対応した形態素解析ソフトとして, MeCab, JUMAN, Chasen, KyTee が有名である. また, 英語などの外国語の形態素解析ソフトは TreeTagger(Schmid, 1994, 1995) などある. テキストマイニングツールとしては, 株式会社 NTT データ数理システムの Text Mining Studio や KH Coder(樋口, 2020) などがある. Text Mining Studio は有償であるが, KH Coder は無償で使用することができる. また, R や Python などでもテキストマイニングを行うためのパッケージが開発されている.

構文解析は, 文章を句単位に分割して句の係受け関係(修飾・被修飾)を分析することである. 構文解析ソフトとしては CaBoCha や KNP などがある.

データマイニングとは, 大量のデータから有益な情報を掘り起こす技術である(小林, 2017). データマイニングでは, 最初に基礎集計となる単語の頻度集計などを行う. そして, 統計処理, 可視化を行い大量のデータから有益な情報を導き出す.

1.2.1 トピックモデル

トピックモデルは, 文章を分類する手法の1つであり, 1つの文章は, 複数のトピック(話題)を持つと仮定し, それぞれのトピックに属する確率をモデル化した手法である. トピックモデルにおいては, 様々なモデルが提案されているが本論文では, LDA(Latent Dirichlet Allocation)(Blei et al., 2003)を用いた. LDA は, 単語の共起性を統計モデルとして数理的に扱うために提案された. 各文章には, 潜在トピックがあると仮定し, 統計的に共起しやすい単語の集合が生成される要因をこの潜在トピックという観測できない確率変数で定式化している. 文章ごとのトピックの出現確率はパラメータ $\theta = (\theta_1, \theta_2, \dots, \theta_K)$, トピックごとの単語の出現確率はパラメータ ϕ によって決められ, θ と ϕ は, Dirichlet 分布によって生成される. Dirichlet 分布によって生成される θ は, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)$, $\alpha_K > 0$ をパラメータとして定義すると以下のようなになる.

$$\begin{aligned}
 p(\boldsymbol{\theta}|\boldsymbol{\alpha}) &= \text{Dir}(\boldsymbol{\theta}|\boldsymbol{\alpha}) \\
 &= \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_k^{\alpha_k-1}
 \end{aligned}$$

ここで, $\Gamma(n)$ はガンマ関数である. また, Dirichlet 分布の期待値と分散は $\alpha_0 = \sum_{k=1}^K \alpha_k$ のとき, 次の様になる.

$$\begin{aligned}
 E(\theta_k) &= \frac{\alpha_k}{\alpha_0} \\
 V(\theta_k) &= \frac{\alpha_k(\alpha_0 - \alpha_k)}{\alpha_0^2(1 + \alpha_0)}
 \end{aligned}$$

次に, トピックを独立に生成し, 各単語の出現確率をトピックに基づいて生成する.

$$\begin{aligned}
 p(\mathbf{z}) &= \prod p(z_i) \\
 p(\mathbf{w}|\mathbf{z}) &= \prod p(w_i|z_i) \\
 p(\mathbf{w}, \mathbf{z}) &= p(\mathbf{z})p(\mathbf{w}|\mathbf{z}) = \prod p(z_i)p(w_i|z_i)
 \end{aligned}$$

トピックモデルのによる解析は, R の `topicmodels` パッケージ (Grün and Hornik, 2011) などを使用することで容易に可能である. また, パラメータの推定は, 変分ベイズ法 (Blei et al., 2003) や Gibbs sampler 法 (Griffiths and Steyvers, 2004), collapsed 変分ベイズ法 (Teh et al., 2007, Asuncion et al., 2009) などが提案されている. 佐藤 (2015) では, 各推定方法の計算コストや予測性能等について比較が行われている.

1.2.2 対応分析

形態素解析により分割された文章中の単語を文章単位や書物単位, 著者単位などで単語の出現頻度を集計するとクロス集計表を作成することができる. このクロス集計表から各文章と単語の出現頻度や各著者と使用される単語の特徴などを分析する方法の1つとして対応分析 (CA:Correspondence Analysis) がある. 対応分析は, Benzecri et al.(1973a, 1973b) によって提唱された外的基準がない場合の数量化法である. 類似する手法として, 林の数量化理論のⅢ類 (林, 1956) や双対尺度法 (Dual Scaling) (西里, 1982) がある.

対応分析の基礎となる r 行 c 列のコレスポネンス行列 Z の成分は、以下の様に定義される。

$$z_{rc} = \frac{f_{ij} - f_i \times \frac{f_j}{n}}{\sqrt{f_i \times f_j}}$$

このとき、 $i = 1, 2, \dots, r$, $j = 1, 2, \dots, c$ で、 f_{ij} はクロス集計表の i 行 j 列成分、 f_i はクロス集計表の i 行目の行和、 f_j はクロス集計表の j 列目の列和、 n はクロス集計の総和である。また、対応分析で得られる行得点 X と列得点 Y は以下の式である。

$$\begin{aligned} X &= D_r^{-\frac{1}{2}} V \\ Y &= D_c^{-\frac{1}{2}} U \end{aligned}$$

このとき、 D_r は $p_i (= \frac{f_i}{n})$ を要素とした対角行列、 V は ZZ^t の固有ベクトルからなる行列 D_c は $p_j (= \frac{f_j}{n})$ を要素とした対角行列、 U は $Z^t Z$ の固有ベクトルからなる行列である。

例として、R の MASS パッケージ (Venables and Ripley, 2002) の中のデータセット `caith` を用いて紹介する。このデータセットは、目の色 (blue(青), light(淡褐色), medium(濃褐色), dark(灰色)) と髪の色 (fair(金髪), red(赤毛), medium(茶髪), dark(栗毛), black(黒髪)) についてまとめたクロス集計 (表 1.1) である。表 1.1 についてコレスポネンス行列 Z を作成すると、表 1.2 ができる。その後、行得点 X (表 1.3) と列得点 Y (表 1.4) を作成する。そして固有値、寄与率、累積寄与率については表 1.5 である。第 2 軸までの累積寄与率が 99.6% であるため、今回は第 2 軸までを使用して biplot を行った (図 1.1)。黒字が行得点 (目の色) を示し、赤字が列得点 (髪の色) を示している。この結果から、目の色「blue」と「light」は髪の色「fair」と関連が強く、目の色「dark」は髪の色「black」と関連が強いことがわかる。

表 1.1: 目の色と髪の色のカロス集計

目の色/髪の色	fair	red	medium	dark	black
blue	326	38	241	110	3
light	688	116	584	188	4
medium	343	84	909	412	26
dark	98	48	403	681	85

表 1.2: 目の色と髪の色のコレスポネンス行列 Z

目の色/髪の色	fair	red	medium	dark	black
blue	0.12921	-0.00026	-0.03538	-0.07544	-0.04372
light	0.17230	0.04777	-0.02328	-0.14838	-0.07088
medium	-0.08474	-0.01429	0.10542	-0.02932	-0.02810
dark	-0.18592	-0.03557	-0.07078	0.25246	0.14265

表 1.3: 行得点 X

目の色	x_1	x_2	x_3
blue	-0.89679252	0.9536227	2.1884132
light	-0.98731818	0.5100045	-1.0837859
medium	0.07530627	-1.4124778	0.1894089
dark	1.57434710	0.7720361	-0.1482208

表 1.4: 列得点 Y

目の色	y_1	y_2	y_3	y_4
fair	-1.21871379	1.0022432	0.4271282	-0.8692696
red	-0.52257500	0.2783364	-4.0268545	-1.3400421
medium	-0.09414671	-1.2009094	0.1103959	-0.8453208
dark	1.31888486	0.5992920	0.3450676	-1.2251588
black	2.45176017	1.6513565	-1.5736976	1.1609621

表 1.5: 固有値, 寄与率, 累積寄与率

	第1軸	第2軸	第3軸
固有値	0.19924	0.03009	0.00086
寄与率	0.86556	0.13070	0.00373
累積寄与率	0.86556	0.99627	1.00000

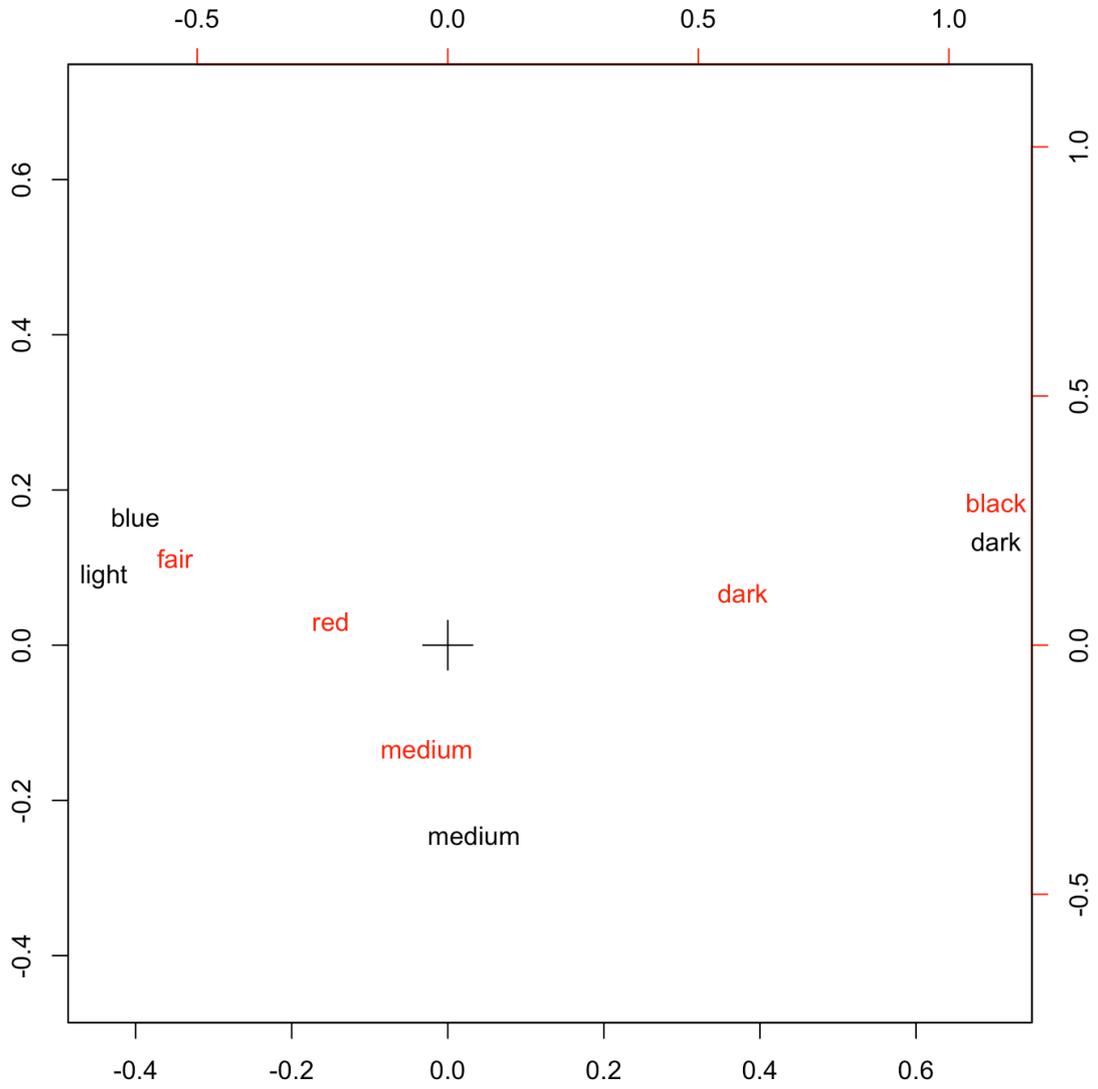


図 1.1: 対応分析の biplot の例

1.2.3 自己組織化マップ

統計データの分類及び可視化手法の1つとして2章で扱う自己組織化マップ(SOM:Self-Organizing Map)について紹介する. 自己組織化マップは, kohonen によって提案されたニューラルネットワークにより, あらかじめ推定した2次元構造にマッピングするクラスター分析の解析法であり, この結果を用いてクラスター構成を2次元に可視化できる (Kohonen, 1982, 2000). 自己組織化マップによる解析は, R の kohonen パッケージ (Wehrens and Buydens, 2007) を使用することで実装できる.

n 個の p 次元観測ベクトルを $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jp})$, ($j = 1, 2, \dots, n$) に対して, k 個のユニットからなる2次元平面上にマッピングするとき, 入力層が観測ベクトルであり, 出力層が1つのユニットとなる. このとき, ユニット i の重みベクトルを $\mathbf{m}_i = (m_{i1}, m_{i2}, \dots, m_{ip})$, ($i = 1, 2, \dots, k$) とすると自己組織化マップのアルゴリズムは, 以下の様になる.

Step1: 入力層と出力層における全てのユニットの中から, 最も類似しているユニット c を探し出し, そのユニットを勝者とする.

$$\|\mathbf{x}_j - \mathbf{m}_c\| = \min_i \{\|\mathbf{x}_j - \mathbf{m}_i\|\}$$

Step2: 下記の式を用いて, 勝者のユニットおよびその近傍のユニットの重みベクトル \mathbf{m}_i を更新する.

$$\mathbf{m}_i(t+1) = \begin{cases} \mathbf{m}_i(t) + h_{ci}(t)[\mathbf{x}_j - \mathbf{m}_i(t)] & (i \in N_c) \\ \mathbf{m}_i(t) & (i \notin N_c) \end{cases}$$

$$h_{ci}(t) = \alpha(t) \exp\left(-\frac{\|\mathbf{r}_c - \mathbf{r}_i\|^2}{2\sigma^2(t)}\right)$$

ここで, $h_{ci}(t)$ は近傍関数で, ユニット c とユニット i の近さによって \mathbf{x}_j の影響を調整する. $\alpha(t)$ は学習率係数, \mathbf{r}_c と \mathbf{r}_i はユニット c とユニット i の2次元上の座標ベクトル, $\sigma^2(t)$ は座標の分散, t は学習回数である.

Step3: Step1 と Step2 を任意の回数繰り返す.

Step4: Step1 から Step3 を全ての観測ベクトル \mathbf{x}_j に対して行う.

例として, iris データセット (Fisher, 1936) を用いて紹介する. このデータセットには, 150 件のデータが収録されている. 今回は, アヤメのがくの長さ (Sepal.Length), がくの幅 (Sepal.Width), 花びらの長さ (Petal.Length), 花びらの幅 (Petal.Width) のデータを使用する. このデータを 10×10 の六方細密構造の自己組織化マップ上

に振り分ける. 図1.2が150件のデータをマッピングし、アヤメの種類 (Species) ごとに色分けした結果である. また、図1.4は、各ユニットの重みを円グラフで可視化したマップである.

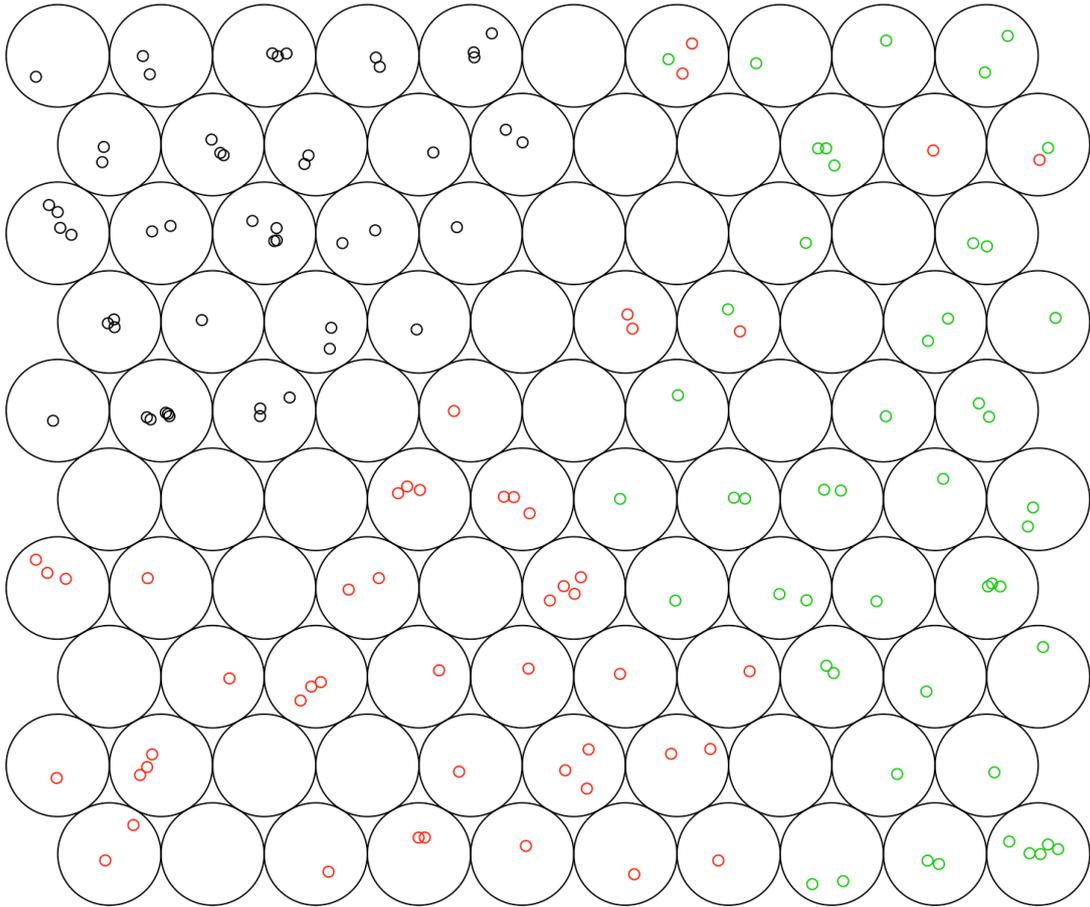


図 1.2: 自己組織化マップの個体のマップ

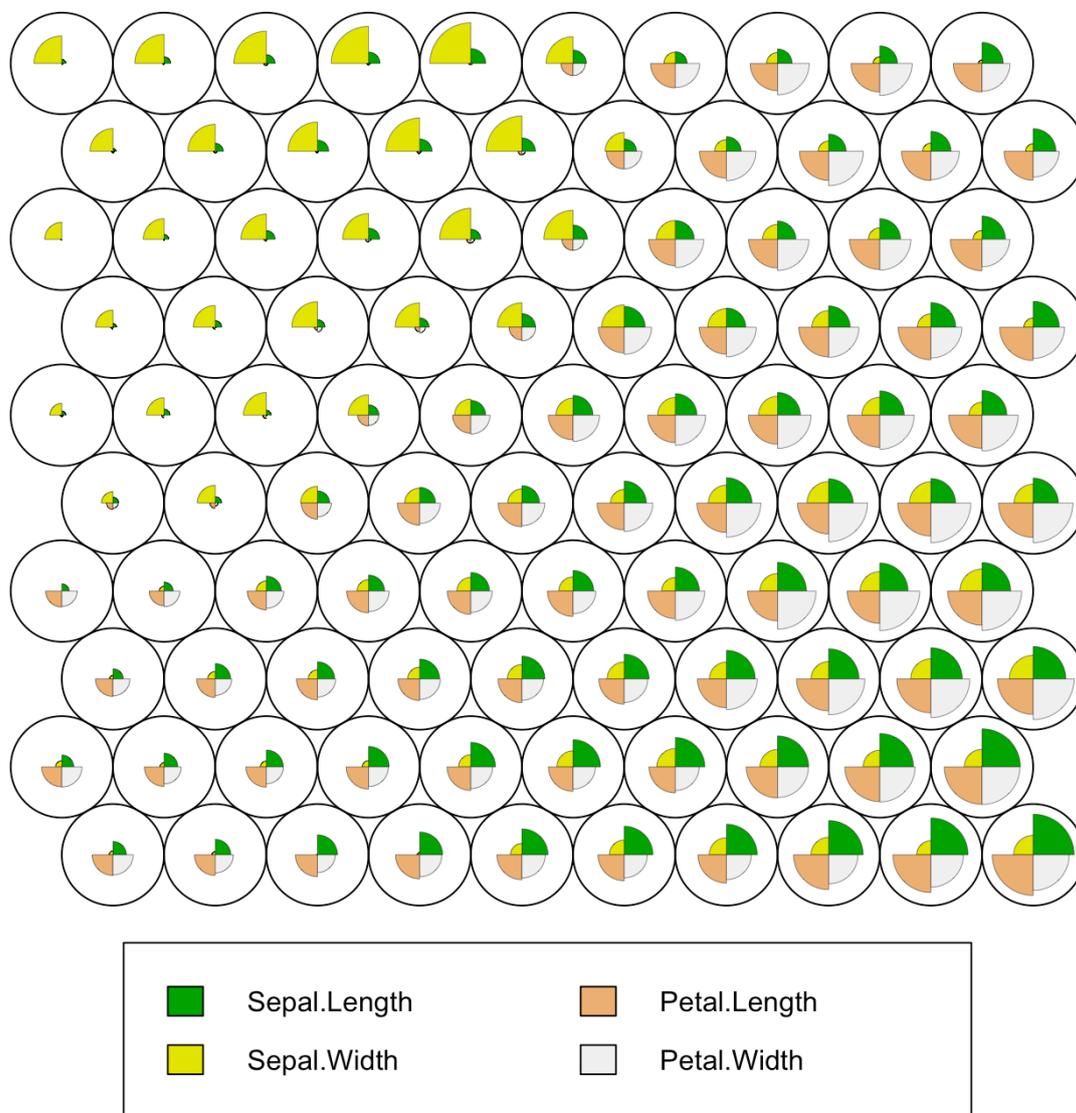


図 1.3: 各ユニットの重みの可視化

1.3 インタラクティブな可視化

静的なグラフを用いた可視化は、あらかじめ作っておいたグラフィイメージを見せることしかできない。そのため、特定のグラフ表現において、質的変数の値との関係との関係を示すような場合には複数のグラフィイメージが必要となる。ひとつのイメージで多面的に観察するにはインタラクティブな(対話的)可視化により可能となる。インタラクティブな可視化は、データを探索的に分析することもできる。また、インタラクティブな可視化に限らず可視化は実際に多くの人に見てもらうことが重要である。そのための手段のひとつとして Web での公開がある。Web 公開することで容易に多くの人に見てもらうことや利用してもらうことができる。本節では、Web 公開の方法として D3.js(Bostock et al., 2011) と R の Shiny パッケージ(Chang et al., 2019) について紹介する。

D3.js は、データビジュアライゼーション作成用の JavaScript ライブラリである。また、BSD ライセンスのもとでリリースされているため商用や非商用問わず無償でコードの使用や改造、修正することができる。D3.js は、ブラウザのメモリにデータをロードしてグラフ化することもできる。また、推移をアニメーションで表示したりグラフを拡大や回転などもすることができる。図 1.4 が、D3.js を用いて作成したインタラクティブにビン数を変更することができるヒストグラムの可視化である。

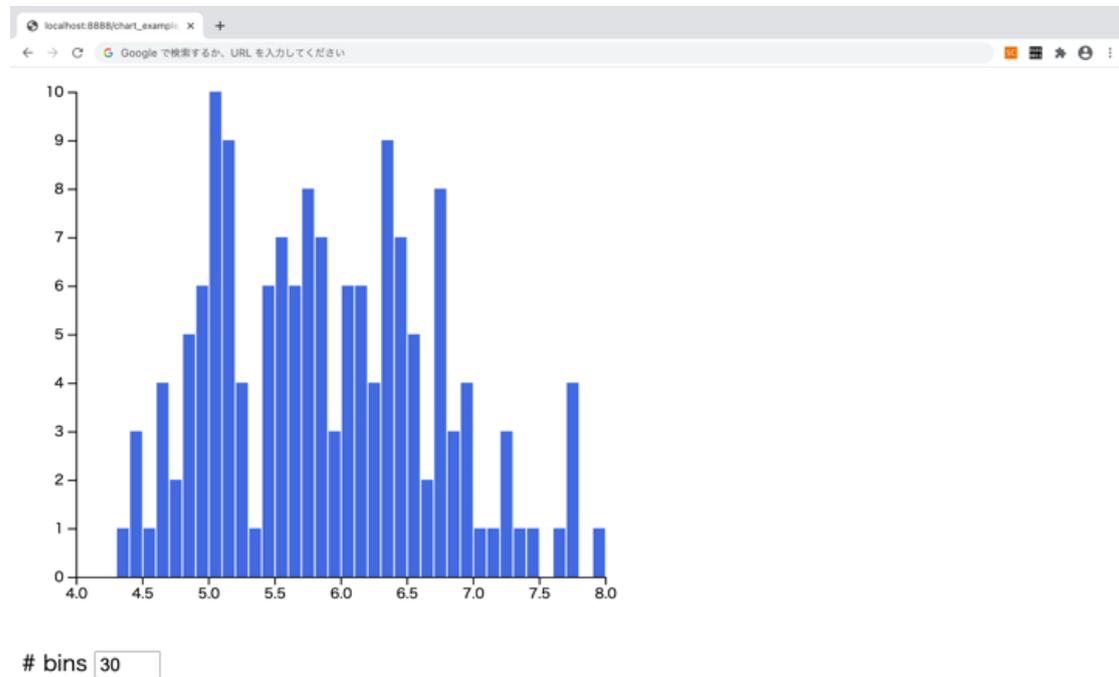


図 1.4: D3.js によるインタラクティブな可視化

R の Shiny パッケージは, Web アプリケーションを簡単に作るためのライブラリである. このパッケージにより HTML や JavaScript などのプログラミング言語の知識がなくても Web アプリケーションを作成することができる. Shiny では, R で作成したグラフを表示することができるだけでなく, パラメータなどの条件の変更やデータの入出力などインタラクティブな可視化が可能である. 作成した Web アプリケーションは, GitHub や自身の Linux 環境に Shiny Server を構築することで公開することができる. 図 1.5 が, Shiny を用いて作成したインタラクティブにビン数を変更することができるヒストグラムの可視化である. 左側のスライダーを動かしてヒストグラムのビン数を変更することができる.

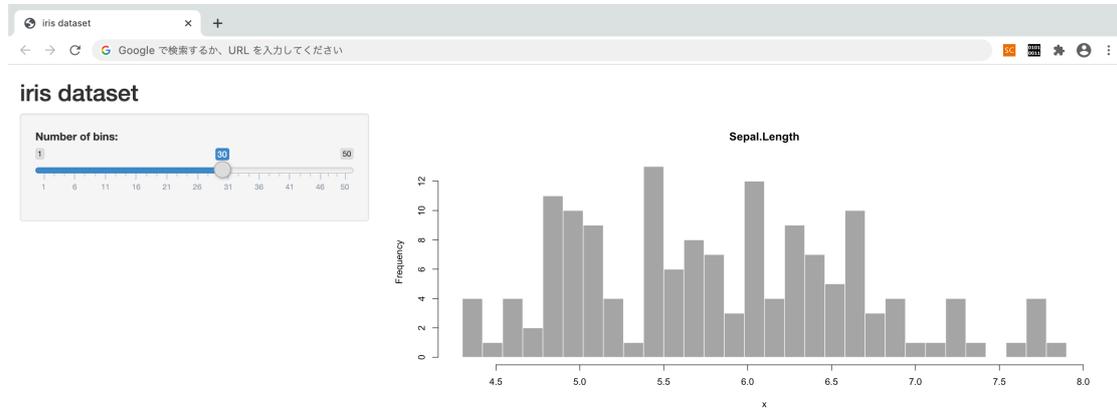


図 1.5: Shiny パッケージによるインタラクティブな可視化

第2章 大規模大学における研究分野 の研究実績の可視化

2.1 本研究の課題と概要

近年、大学ではIR(Institutional Research)に関心が高まっている。IRは、現在でも発展を続けているなどの理由から共通の定義がない(小林 他, 2016)。現在では、大学の管理運営、戦略計画、財務計画などに必要なデータを提供し、これらの策定を支援することがIRに課せられた重要な使命となっている(小林 他, 2016)。IRは、大きく分けて教学IRと研究IRの2つの取り組みがある。教学IRでは、教育改革や教育改善などの教育質保証の取り組みである。特にアメリカでは、先行して教学IRが発展している。研究IRは、学内の研究状況の把握や研究支援などの研究強化が目的である。研究IRの方法論としての研究開発は乏しく、確立してない。また、統計数理研究所では、2016年度から2019年度まで共同利用研究の重点テーマとして研究が行われていた。2016年度から2017年度は、「学術文献データ分析の新たな統計科学的アプローチ」(統計数理研究所, 2017, 2018)、2018年度から2019年度は、「IRのための学術文献データ分析と統計的モデル研究の深化」(統計数理研究所, 2019)という研究テーマで取り組まれた。この共同研究では、IRの研究者や実務者、統計解析の研究者が集まり、研究力評価に貢献するような新たな指標の開発や学術論文データベースの活用などの研究が行われた。

本章は、統計数理研究所で行われた共同研究の学術論文データベースの活用に関して行った研究成果である。学術論文データベースを用いた関連研究として、交通研究の論文のアブストラクトに対してトピックモデルを適用して研究動向を調査(Sun and Yi, 2017)やオペレーションズリサーチと経営工学の論文のアブストラクトに対してトピックモデルを適用して研究動向調査(Gatti and Brooks, 2015)などがある。特定の研究分野内の分析は活発に研究されているが、複数の研究分野が混在する学内全体の分析に関する研究は少ない。本研究では、学術論文データベースに収録されている論文のタイトルとアブストラクトのテキストデータから論文の研究領域を推定し、大規模な総合大学における学内の研究活動状況の把握を行った。研究領域の推定にトピックモデルを用いることで、各論文の推定されたトピック

クへの所属確率を得られるため、各論文は特定のトピックのみに関係するのではない。そのため、各論文の推定されたトピックへの所属確率の値に自己組織化マップを適用することで、複数のトピックとの関連により研究分野について特定が可能であることを示す。また、自己組織化マップの結果を用いることで研究分野の業務評価について様々な可視化が可能となることを実例により示す。

本研究では、ケーススタディとして大規模な総合大学である T 大学を例に分析を行った。また、学術文献データベースは、クラリベイト・アナリティクス社の提供する Web of Science(WoS) を使用した。本研究は、統計数理研究所の共同研究として、WoS のデータ活用及びデータベースの利用ができるようになったものである。

本章の研究成果は、統計数理研究所が刊行している「統計数理」に原著論文として掲載された。掲載された論文の正式の全文 pdf ファイル(出版社版)は、<http://www.ism.ac.jp/editsec/toukei.html> で閲覧できる。また、著作権は統計数理研究所に帰属する。

2.2 データセットの作成

本研究で扱った WoS のデータは、統計数理研究所の共同研究として提供された WoS のデータベースをもとに Neo4j に格納されたネットワークデータベースから抽出した。Web の WoS では、「著者の所属(大学名)」の変数があるが、提供された Neo4j データベースには無く、著者の所属を示す変数は、「著者の住所」のみであった。著者の住所とは、研究機関名(大学名など)、部署(学部学科など)、住所の順に記述された文字列データである。本研究では、著者の住所が“T Univ.”(T 大学)から始まる著者を T 大学に所属する研究者とし、2007 年から 2016 年に発表された T 大学に所属する研究者(大学院生を含む)を著者に含む論文(4,261 編、以下では「T 大学著者論文」とそれらの論文を引用した論文(36,604 編、以下では「T 大学引用論文」)のタイトルとアブストラクトを抽出し、T 大学著者論文と T 大学引用論文を結合したデータセットを作成した。データセットの論文のタイトルとアブストラクトのテキストデータを形態素解析し、過去形などの単語を原形に統一し、さらにピリオドやカンマなどの記号やストップワードは除外した。今回は、英語で書かれた論文データであるため、形態素解析ソフトウェアとしては、TreeTagger(Schmid, 1994, 1995)を使用した。以上の前処理後のデータセットは、総論文数は 40,865、データセット内で使われている単語数は 133,430 であった。T 大学著者論文の T 大学所属の著者は述べ 23,001 人(同一著者を含む)であった住所翻訳から判別できた各論文の主要研究者の学内の所属は、学部 20,390、センター・研究所 809、大学院 762、付属病院 625、短期大学 14、不明 401 であった。不明は、大学名以外に住所しか記述さ

れていないものである。更に、各論文においてT大学の著者が複数含まれている論文にそれらの著者が異なる学部、センター等に所属するものについては「複数学部著者論文」として学内の組織間の共同研究による成果として扱うこととする。

2.3 トピックモデルを用いた論文の研究領域の推定

論文の研究領域を推定するため、トピックモデルを用いた。桂井 他(2015)は、日本の論文データベースである CiNii に登録されている論文に対して、LDAによるトピックモデルを用いて著者同定を行った。藤野 他(2016)では、分析対象の研究組織に所属する研究者名と同名の研究者が WoS において所属が示されていない論文について、学内データを加えることで推定された著者の特徴ベクトルを用いて組織の研究者か否かについて判別を行なった。武井 他(2019)では、統計科学分野の論文のアブストラクトに対してトピックモデルを適用して研究評価への活用の有効性を示した。

LDA は、事前分布に Dirichlet 分布を用いるため、特定のトピックへの所属確率が高くなる傾向がある。その性質を利用して最も所属確率が高いトピックに論文を割り当てることで各論文と1つの研究領域に振り分けられるようにした。トピックモデルは、あらかじめトピック数を設定する必要がある。本研究では、トピック数の評価指標の1つである Perplexity(Blei et al., 2003) を使用して妥当なトピック数について判断した。Perplexity は、トピック数を増やすほど良い評価(小さな値)となる傾向があること、トピック数が多過ぎると研究領域を極端に細分化してしまう。また、膨大な計算量になってしまうため実用的でなくなることから、Perplexity が減少から増加に転じたトピック数、図 2.1 から 53, 55, 59, 67, 75 などを妥当なトピック数の候補と考えた。本論文では、計算コストを考慮してトピック数を 59 とした場合の結果を示す。LDA により推定された各トピックが示す研究領域は、各単語の出現確率が高い単語から判断した。図 2.2 は、各論文のトピック出現確率のヒートマップであり、行に論文に対するデンドログラムで列にトピックに対するデンドログラムを示している。表 2.1 には 59 のトピックに振り分けられた論文数および各トピックに出現した単語上位 5 単語、それらの単語から類推されるそのトピックの示す研究領域、さらに、T大学の異なる学部の著者が含まれている論文数の割合を示した。

第2章 大規模大学における研究分野の研究実績の可視化

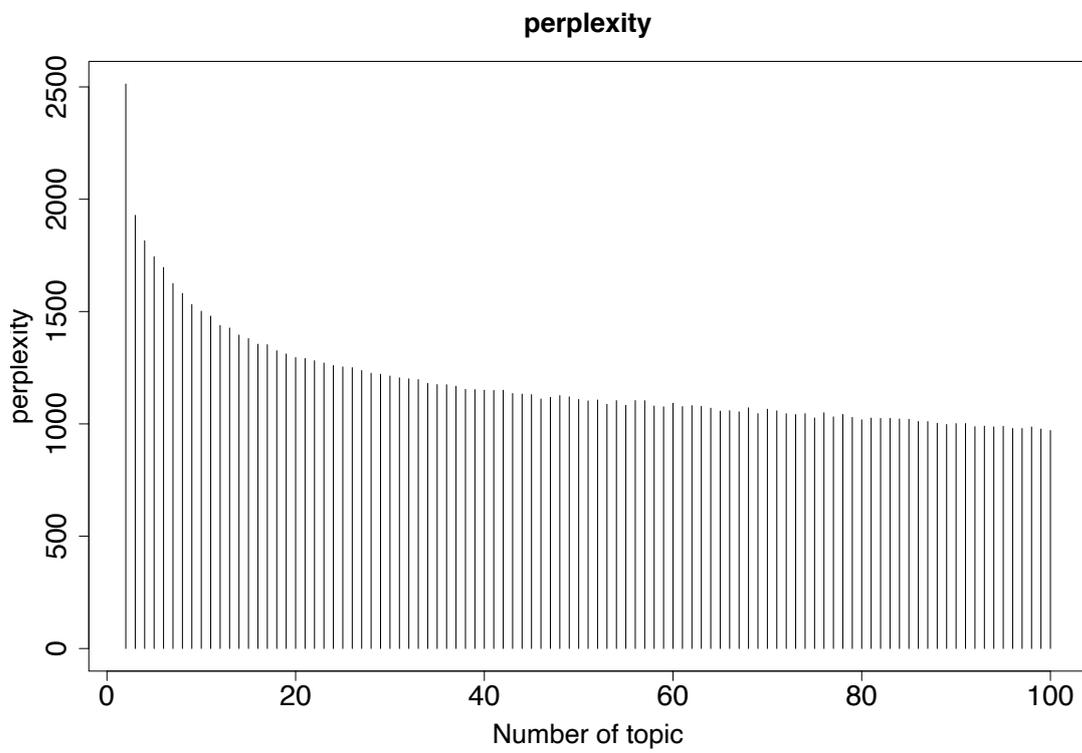


図 2.1: Perplexity の推移

第2章 大規模大学における研究分野の研究実績の可視化

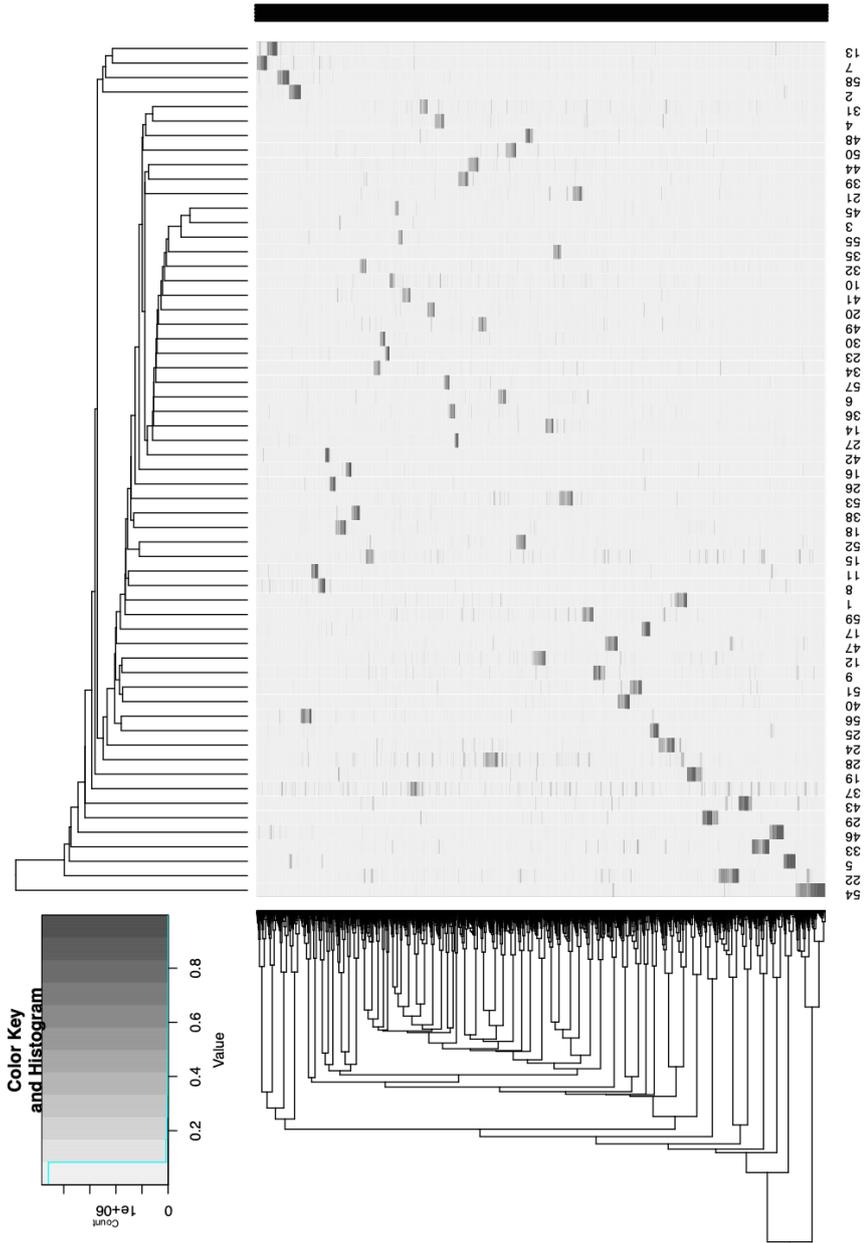


図 2.2: 各論文のトピックの出現確率のヒートマップ

第2章 大規模大学における研究分野の研究実績の可視化

表 2.1: 各トピックの出現確率が高い上位5単語

トピック番号	単語 1	単語 2	単語 3	単語 4	単語 5	研究領域	T 大学著者論文数	複数学部著者論文
トピック 54	patient	atrial	fibrillation	oral	anticoagulant	抗凝固	48	0%
トピック 22	patient	risk	disease	stroke	cardiovascular	心臓血管	110	0%
トピック 5	neutrino	matter	mass	dark	model	ニュートリノ	59	5%
トピック 33	cancer	patient	lung	treatment	survival	癌	190	1%
トピック 46	surface	property	alloy	use	high	金属・工学	142	3%
トピック 29	patient	transplantation	cell	donor	leukemia	移植	165	1%
トピック 43	coronary	stent	cardiac	patient	artery	心臓	105	1%
トピック 37	review	use	new	research	clinical	関連研究についての説明	32	0%
トピック 19	association	gene	study	polymorphism	genetic	ゲノム (応用)	100	0%
トピック 28	role	mechanism	disease	function	cell	メカニズム	11	0%
トピック 24	autophagy	cell	induce	apoptosis	increase	オートファジー・細胞	24	4%
トピック 25	center	dot	reaction	compound	synthesis	化合物・合成	98	5%
トピック 56	bind	protein	peptide	binding	acid	タンパク質	141	1%
トピック 40	case	pancreatic	tumor	lesion	neoplasm	膵臓	109	2%
トピック 51	image	imaging	tomography	use	magnetic	画像・磁気・MRI	97	4%
トピック 9	age	patient	aging	study	care	健康	83	0%
トピック 12	disease	patient	syndrome	disorder	clinical	臨床医療	62	0%
トピック 47	platelet	antiplatelet	patient	dose	therapy	血小板	74	0%
トピック 17	sea	water	ocean	high	surface	海洋	78	0%
トピック 59	cell	expression	liver	beta	pathway	細胞・肝・増殖	80	3%
トピック 1	protein	mitochondrial	autophagy	mutant	function	細胞	43	19%
トピック 8	model	use	function	result	method	モデリング	103	1%
トピック 11	water	temperature	relaxation	dielectric	solution	溶液を使っての実験	90	1%
トピック 15	card	patient	study	group	analysis	調査デザイン・調査の結論	71	0%
トピック 52	level	serum	patient	ckd	kidney	透析	84	1%
トピック 18	sequence	gene	class	allele	hla	ゲノム (基礎)	70	0%
トピック 38	gene	vertebrate	evolution	genome	species	ゲノム・遺伝子・進化	42	19%
トピック 53	cell	stem	differentiation	human	progenitor	細胞・造血・再生医療	63	2%
トピック 26	cloud	use	data	aerosol	satellite	気象・衛星	62	2%
トピック 16	population	asian	genetic	asia	human	遺伝・地理	49	0%
トピック 42	delta	lake	sediment	card	change	自然・火山	55	0%
トピック 27	star	abundance	card	line	expose	星	19	0%
トピック 14	bone	tissue	use	cartilage	fracture	骨・整形外科	51	2%
トピック 36	plant	compound	acid	extract	isolate	工場・化合物・毒	94	1%
トピック 6	cell	notch	development	signale	neural	細胞・遺伝子・ニューロン	36	6%
トピック 57	strain	assay	bacterial	pylorus	bacterium	細菌・病原体	54	2%
トピック 34	brain	rat	diabetes	injury	cerebral	脳・糖尿病・虚血	53	4%
トピック 23	monitor	monitoring	earthquake	test	time	地震	42	2%
トピック 30	muscle	skeletal	stimulation	motor	exercise	筋・骨格・運動	52	4%
トピック 49	liver	acid	metabolic	fatty	diet	肝	33	6%
トピック 20	dna	imprint	epigenetic	gene	methylation	DNA・遺伝・マウス実験・ゲノム	44	7%
トピック 41	infection	virus	human	viral	marmoset	感染症	53	0%
トピック 10	expression	gene	protein	express	analysis	DNA・遺伝子	38	11%
トピック 32	group	blood	level	effect	significantly	血液	108	0%
トピック 35	disc	degeneration	intervertebral	cell	ivd	細胞	47	0%
トピック 55	mouse	use	modify	model	modified	マウス実験	59	4%
トピック 3	elsevier	right	reserve	reserved	publish	引用文献について	3	0%
トピック 45	male	female	hormone	pituitary	gland	性・ホルモン	31	0%
トピック 21	cell	endothelial	vascular	epcs	progenitor	細胞・血管	80	0%
トピック 39	renal	kidney	nephropathy	injury	glomerular	腎臓	49	2%
トピック 44	inflammation	mouse	effect	oxidative	stress	炎症	57	0%
トピック 50	cell	immune	complement	response	activation	免疫・抗体	77	0%
トピック 48	lymphoma	lcell	tcell	patient	case	リンパ腫	57	0%
トピック 4	cancer	tumor	breast	alpha	cell	乳癌・卵巣癌	54	0%
トピック 31	treatment	drug	therapy	target	therapeutic	薬剤治療	24	0%
トピック 2	energy	telescope	gammaray	source	emission	光工学	55	4%
トピック 58	use	method	laser	sample	pulse	レーザー・分析手法	105	1%
トピック 7	system	propose	proposed	use	method	情報系・システム系	162	0%
トピック 13	film	plasma	thin	ion	temperature	プラズマ・イオン・磁気	184	1%

第2章 大規模大学における研究分野の研究実績の可視化

T大学の場合、医学部の研究者の発表論文の割合が高いことがあり、推定された研究領域は、医学系分野に含まれる割合が多かった。トピック数を多くしたことで、医学系分野について論文数に応じて区分することができた。また、工学系分野(トピック 2, 46, 51, 58)、情報系分野(トピック 7)、理学系分野(トピック 5, 25)、自然科学系(トピック 17, 26, 27, 42)などの研究領域も推定されているためT大学において、研究業績がある程度ある研究領域を特定することができたと考えられる。また、研究方法・手順についての記述が多いアブストラクトを用いた影響によりトピック 3 や 37 のような研究領域とは考えにくいトピックも生成された。

図 2.3 に各トピックに属している T 大学著者論文の被引用数の箱ひげ図を示した。外れ値があり箱ひげ図が潰れてしまったため、縦軸の上限を 250 とし、上部に × と外れ値の値を示した。また、各論文について T 大学の異なる学部の著者が含まれている割合(表 2.1 の左から 3 列目)については、トピック 1 とトピック 38 が 19%、トピック 10 が 10% の割合であった。トピック 10 が医学部と工学部、トピック 38 が医学部と海洋学部、生物学部、健康科学部、工学部などの共同研究であることがわかった。この様にトピックモデルで研究領域を特定することで、学内連携が多くみられる分野についても特定することができた。

第2章 大規模大学における研究分野の研究実績の可視化

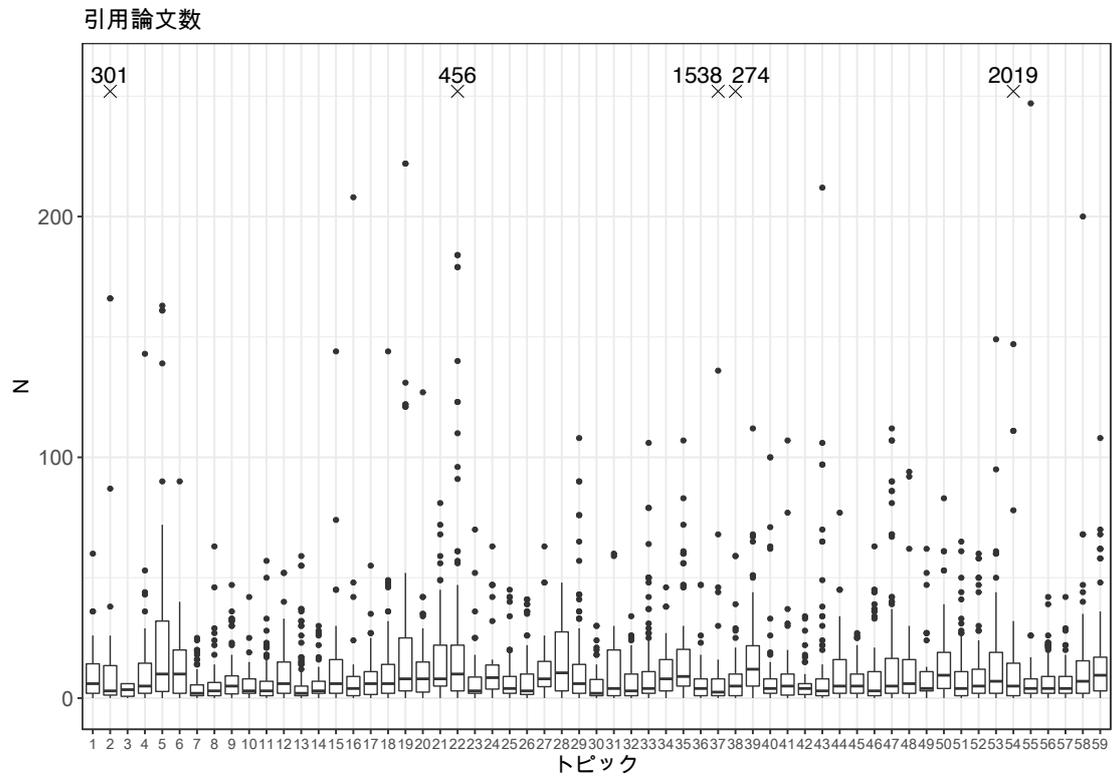


図 2.3: 引用論文数の箱ひげ図

第2章 大規模大学における研究分野の研究実績の可視化

次に、T大学引用論文の研究領域について考察する。T大学引用論文の研究領域は、T大学著者論文の研究成果が貢献した研究領域と考えることができる。T大学著者論文と被引用論文の研究領域を比較することで同一研究領域または複数の研究領域での貢献であるかを把握することができる。そのためにT大学著者論文と被引用論文の研究領域の関連性をアルビアルダイアグラム (Alluvial diagram) を用いて可視化した (図 2.4)。図 2.4 の上段は被引用論文のトピック、下段がT大学著者論文のトピックであり、それらを繋ぐ線の太さが対応する論文数を示している。図 2.5 は、図 2.4 の可視化を D3.js (Bostock et al., 2011) を用いて実現したインタラクティブな可視化である。図 2.5 では、左側のトピック 22 の領域にカーソルを置くことでトピック 22 に属する T 大学論文を引用している T 大学引用論文が属するトピックの連結線が強調されるインタラクティブ機能を示している。多くの T 大学著者論文は、元の論文と同じトピックに引用されているが、異なるトピックに引用されている割合が高いトピックもある。例えば、トピック 22 (心臓血管) は、同一トピックだけでなく、トピック 43 (心臓)、47 (血小板)、54 (抗凝固) など多くの研究領域の影響があるトピックの論文に引用されておりこれらは互いに関係の強いトピックと考えられる。同様にトピック 8 (モデリング) も同一トピックだけでなく、トピック 5 (ニュートリノ)、トピック 7 (情報系・システム系)、トピック 58 (レーザー・分析手法) などの研究領域の影響があるトピックの論文に引用されていた。この様に T 大学引用論文の研究領域も同時に推定することで研究成果が貢献している研究領域についての考察ができることを示した。

第2章 大規模大学における研究分野の研究実績の可視化

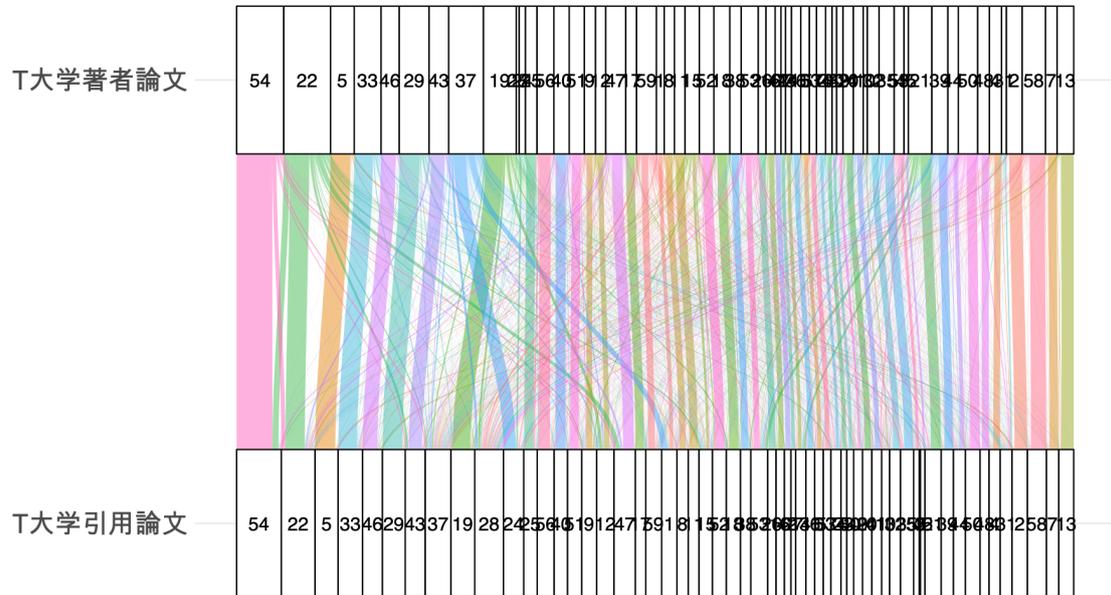


図 2.4: T 大学著者論文と T 大学引用論文のトピックの関連性

第2章 大規模大学における研究分野の研究実績の可視化

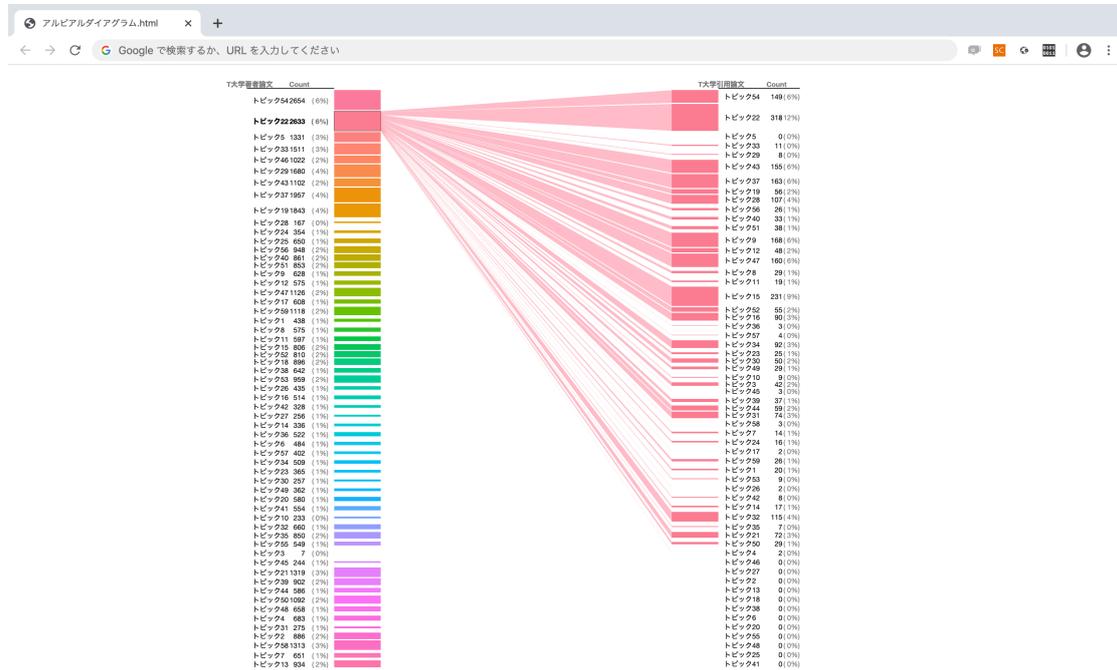


図 2.5: D3.js による T 大学著者論文と T 大学引用論文のトピックの関連性のインタラクティブな可視化

2.4 自己組織化マップによる研究実績の可視化

トピックモデルを用いて研究領域を推定することができたので、次に研究領域が類似しているトピックについて考察する。そのため推定された研究領域の類似性が判別できるようなクラスタリングを考えた。図 2.2 のクラスタリングの結果で表 2.1 のトピックは並べているが、階層型クラスタ分析の結果によりクラスターを分けると、特定のトピックは単一のクラスターに含まれることになる。トピックモデルでは各トピックの所属確率が得られているので、所属確率を入力データとして自己組織化マップによりクラスタリングする。Tian et al. (2014) では、SOM のユニット数として $5\sqrt{N}$ を目安としており、T 大学著者論文数 $N = 5,893$ について $5\sqrt{5893} \approx 384$ であるので 20×20 の出力ユニットとした。図 2.6 は、 20×20 の六方最密構造の自己組織化マップ上に各ユニットに振り分けられた論文のトピック番号を示し可視化した結果である。自己組織化マップでは、類似性が高いサンプルが同一ユニット及び隣接するユニットにマッピングされる。各トピック毎にトピック所属確率を標準化した値を入力データとして、自己組織化マップを適用した(図

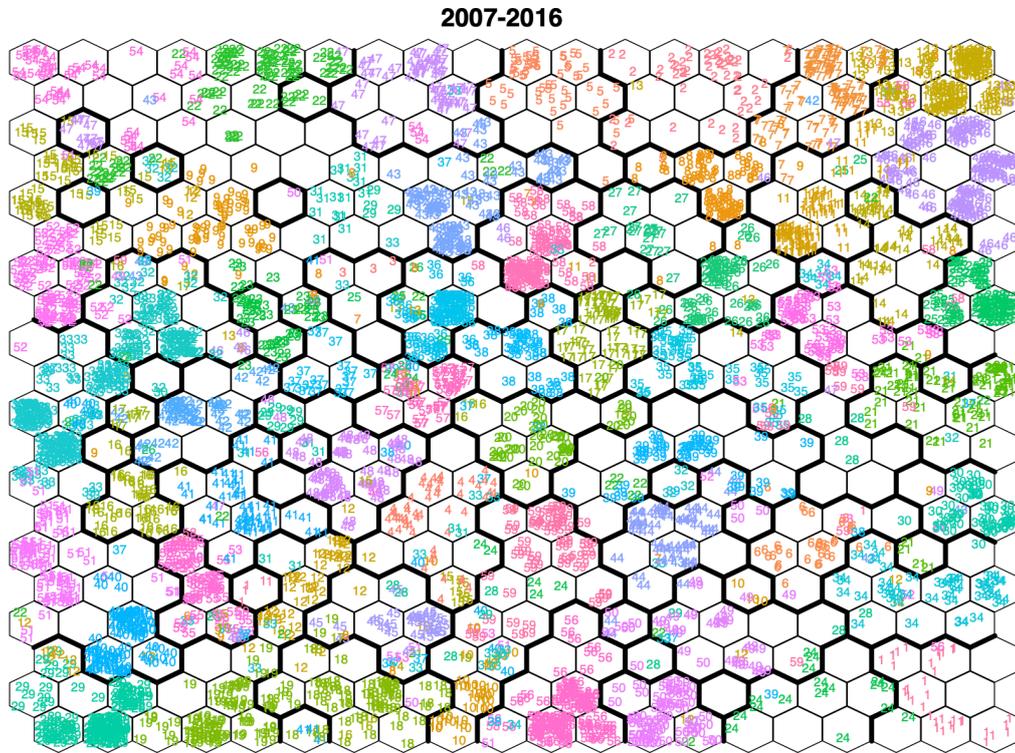


図 2.6: 入力データを標準化した自己組織化マップによる可視化

2.6). 図 2.6 から、隣接しているトピック同士が類似性が高いトピックであるものが見受けられる。実際に下部の真ん中あたりにトピック 18(ゲノム (基礎)), トピック 19(ゲノム (応用)), トピック 10(DNA・遺伝子)があり、ゲノムに関するトピックが隣接してマッピングされている。また、左下にはトピック 29(移植)とトピック 40(臓器)が隣接したユニットにマッピングされている。このことから臓器移植の研究の中でも臓器移植の研究が対象期間に T 大学において盛んに行われていたことを推測できる。

図 2.6 におけるユニットの境界線の太線は、各ユニットに関する自己組織化マップの重みベクトルをデータとして全 400 ユニットを Ward 法によりクラスタリングしたクラスター境界を示している。クラスター間の距離としては、ユークリッド距離を用いた。クラスター数は、Upper Tail 法 (Mojena, 1977) を用いて最適なクラスター数を算出し、37 クラスターとした。

第2章 大規模大学における研究分野の研究実績の可視化

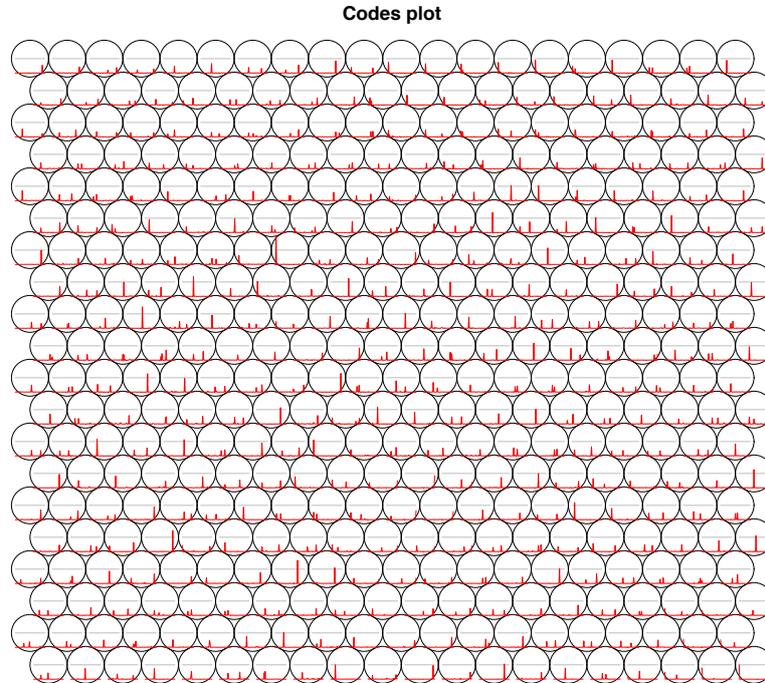


図 2.7: 自己組織化マップの各ユニットにおける各トピックに対する重みベクトル

図 2.6 において上から 2 段目左から 4 列目のユニットは、太線で囲まれているが、その多くはトピック 22 であり近隣のユニットにもトピック 22 がある。このユニットのトピック分布 (各トピックに対する重みベクトル)(図 2.7) を見るとトピック 22 と同程度にトピック 34 の確率も高いことがわかった。その為、周囲のトピック 22 が多いユニットと異なるクラスターに分類されている。また、図 2.6 において上から 1 段目左から 9 列目付近のユニットにトピック 47 の論文が固まっているが、3 段目左から 2 列目に同じトピック 47 からなるユニットがあり、太線で囲まれている。3 段目左から 2 列目のトピック 47 のユニットのトピック分布を見るとトピック 47 の他にトピック 15 の確率も高いことがわかった。トピックモデル数を 67 や 73 とより大きくすることでこのクラスター境界で別れているトピックは、より小さな分野に分かれると思われる。このように自己組織化マップを用いることでトピックモデルの結果及び各論文の特徴をより細かく分析することができる。また、前節で示した引用関係があった、トピック 22(心臓血管)、トピック 43(心臓)、47(血小板)、54(抗凝固) は、上部左から中央へ繋がってマッピングされた。

第2章 大規模大学における研究分野の研究実績の可視化

さらに、経年的な変化を把握するために、特定の期間に発表された論文のみをマッピングした。ここでは、2007年から2010年、2010年から2013年、2013年から2016年の重なりのある4年間ごとに3期間に分けて可視化した。さらにトピック番号の色を各トピックに属す論文数が多いほど濃く、少ないほど淡い色とすることで、各トピックの論文数の比較ができるようにした。各期間のマップを比較することで各トピックに属する論文数の推移を捉えることができた。3期間のマップ(図2.8)を比較すると左上にマッピングされたトピック54, トピック47, トピック15の血液に関する研究領域の論文が増加していることがわかった。また、下部の中心から左側にマッピングされたトピック18(ゲノム(基礎)), トピック19(ゲノム(応用))のゲノムに関する研究領域は、応用研究の論文数が減少傾向にある一方、基礎研究の研究論文が増加傾向にあることがわかる。右上には、医学系以外の情報系や工学系などの研究領域が多くマッピングされており、それらは、3期間とも論文数が多いことがわかる。

第2章 大規模大学における研究分野の研究実績の可視化

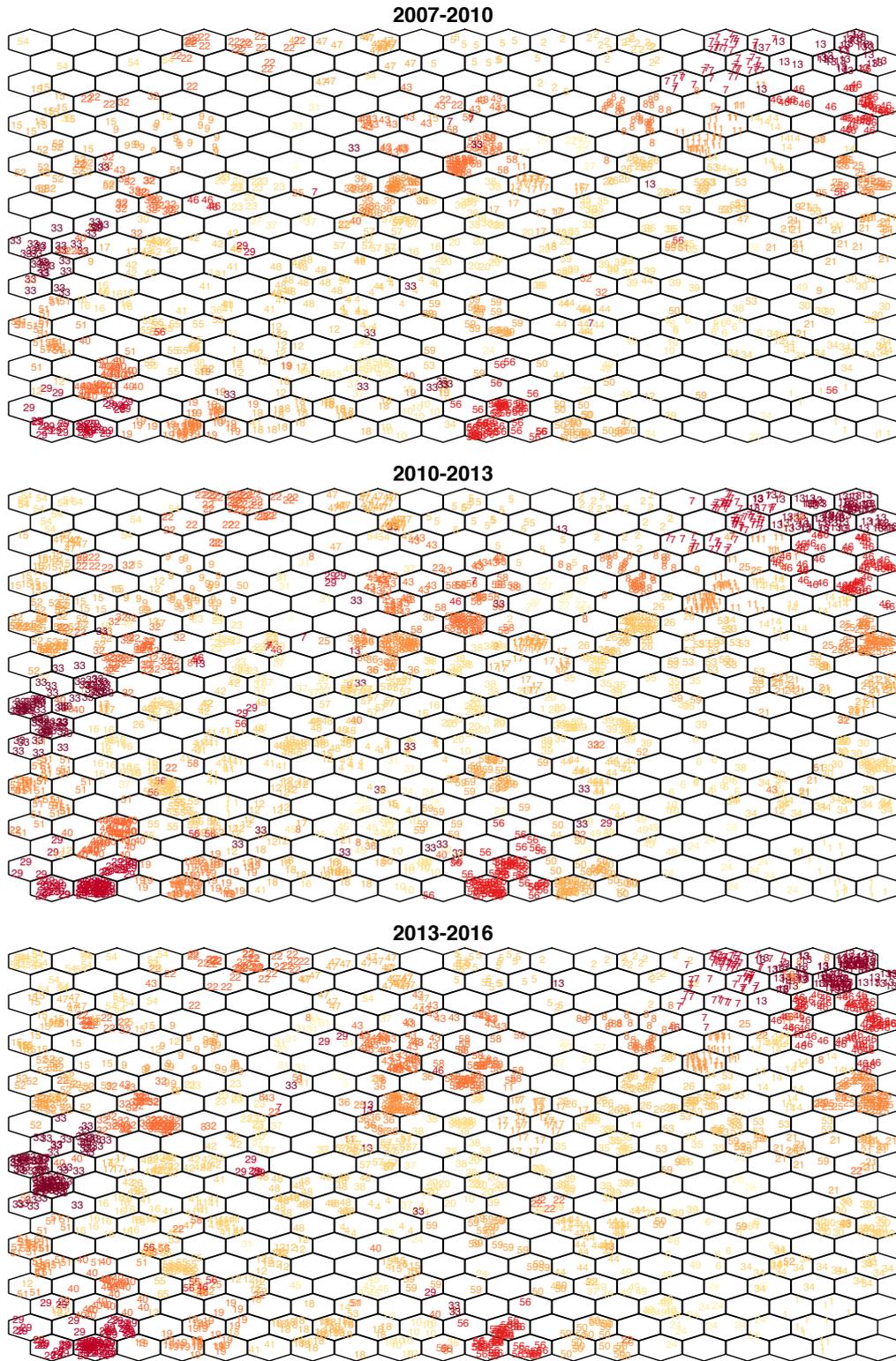


図 2.8: 経年的な変化の可視化

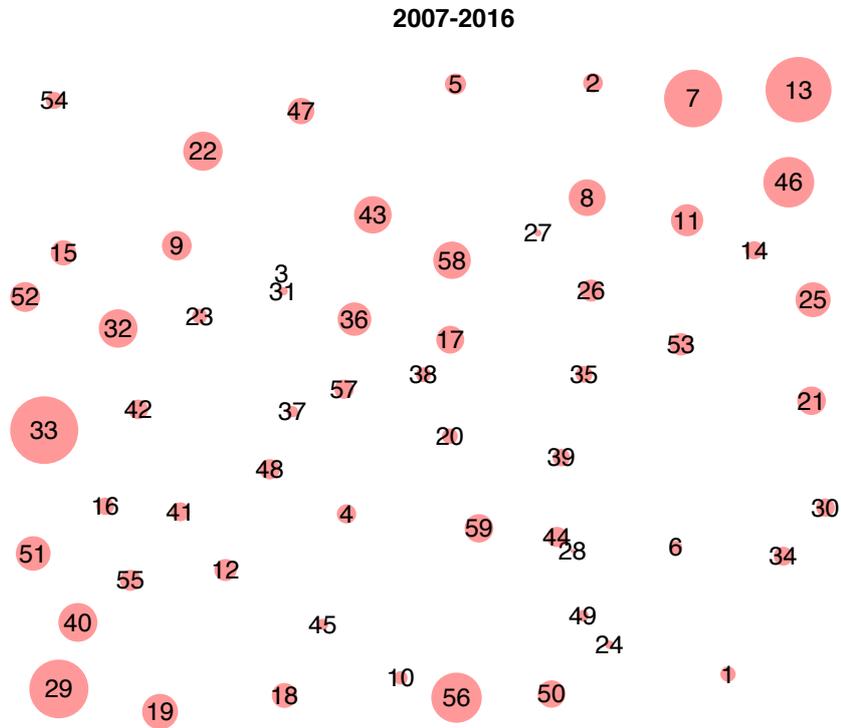


図 2.9: 各トピックの重心による可視化

2.5 自己組織化マップを用いた可視化の改良

前節では、自己組織化マップを用いて、論文や研究分野を2次元平面上に可視化したものについて、期間を限定して表示したものを比べることで(図 2.8)、研究が活発になった分野についても把握できることを示したが、この目的の為に有用な可視化について提案する。各トピックに属する各論文が振り分けられたユニットの中心座標をその論文の座標とし、トピックごとに重心を求め、その重心座標をトピックの代表点とする。さらに各トピックの論文数をバブルチャートを用いて自己組織化マップ上に可視化した(図 2.9)。この可視化により全トピックを2次元平面上にマッピングすることで、業績の多い研究領域について業績の量や研究領域の関連について把握しやすくなった。

研究成果は、論文数だけでなくその研究の影響力や他研究者からの評価も大切な指標である。論文の影響力は、被引用論文数で示することができる。図 2.3 の被引用論文数の箱ひげ図から大きな外れ値があることがわかっているため、各トピックの

第2章 大規模大学における研究分野の研究実績の可視化

代表値として被引用論文数の中央値を用いた。この被引用論文数の中央値を図 2.9 のグラフの円の色 (濃度) で示した (図 2.10)。被引用論文数が多いほど色を濃く、少ないほど淡くした。図 2.10 を見ると円が大きいく色が薄い、つまり論文数が多く被引用論文数が少ないトピック、逆に円は小さいが色が濃い、つまり論文数は少ないが被引用論文数が多いトピックなど各研究分野について多面的に捉えることができる可視化を示した。前節の3つの期間に対する、経年変化を見ることができるようモーショングラフを作成した (図 2.11)。図 2.11 のモーショングラフは、R の plotly パッケージ (Sievert, 2018) を用いて実現しており、3つの期間それぞれの各トピックの論文数、各トピックの論文の被引用論文数を各トピックの座標にバブルチャートで示し、[play] を押すことで、3期の変化の様子が確認でき、またスライダーを動かすことで1期から2期の変化など一部の变化について確認することもできる。

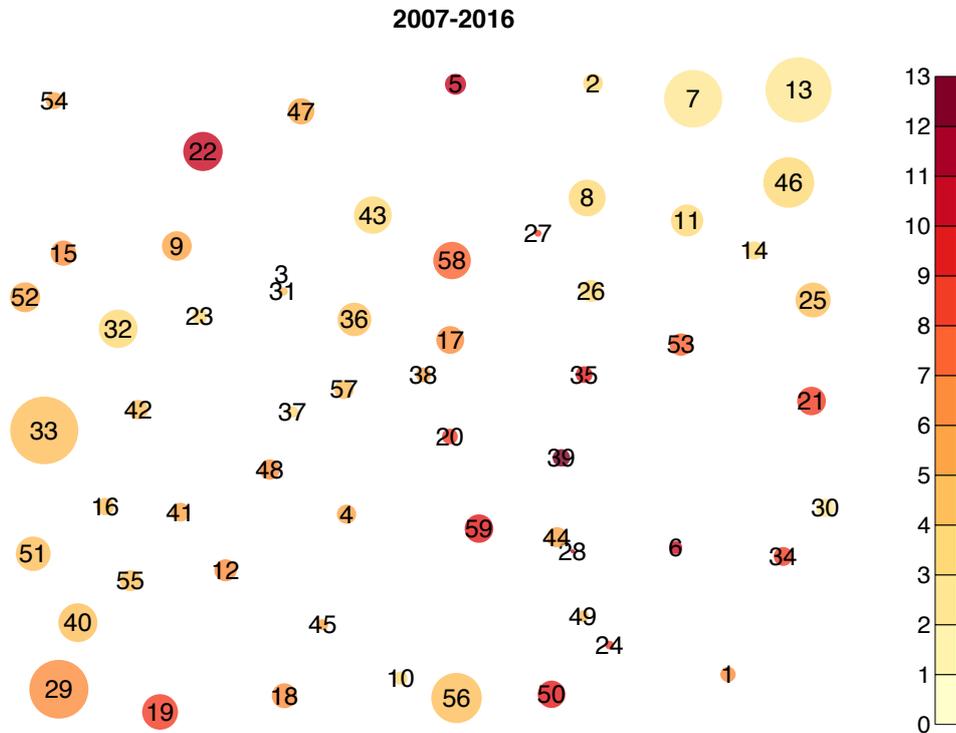


図 2.10: 被引用論文数の情報を追加した可視化

第2章 大規模大学における研究分野の研究実績の可視化



図 2.11: モーショングラフによる経年変化の可視化

第2章 大規模大学における研究分野の研究実績の可視化

さらに経年的な論文数の変化を1つのプロットで示す可視化を提案する。図2.10のグラフでは、各トピックの全体の論文数を円の大きさで表しているためそのトピックの論文がどの時期に増加したのかまたは減少したのかを捉えることができない。そのため図2.11のモーショングラフで見えるような論文数の推移について、3つの期間における各トピックの論文数の増減を折れ線グラフで示し、それを図2.8の自己組織化マップ上に反映させた。図2.12は、2.4節の経年的な変化の可視化と同様の3期間の論文数の推移を可視化した。折れ線グラフは、第2期(2010年から2013年)を基準に前後の期間の増減を示している。この可視化により、各トピックの論文数について全体量とともに経年変化についても把握できるプロットを作成することができた。経時的な変化の可視化については、論文数だけでなく、被引用論文数の中央値、異なる学部の著者の割合などについても有用できる。T大学の場合、自己組織化マップの右上のトピック13(プラズマ・イオン・磁気)、トピック46(金属・工学)、トピック14(骨・整形外科)、トピック25(化合物・合成)、トピック53(細胞・造血・再生医療)の研究領域は、引用論文数はまだ少ないが論文数は経年的に増加していることがわかる。

研究成果の貢献という観点から考えると論文数の多さよりも被引用論文数が多い方が評価が高い。例えば、トピック38の様に論文数は少なく近年減少傾向にあるが、被引用論文数が多い研究領域である。この研究領域の論文数の減少傾向を増加傾向に転じる様に支援することで大学を代表する様な研究領域にすることができる。また、トピック13の様に論文数が多く、増加傾向にあるが被引用論文数が少ない研究領域の場合は、近年発展している研究領域と考えることができる。この研究領域は、今後の被引用論文数の増減で評価することになる。

第2章 大規模大学における研究分野の研究実績の可視化

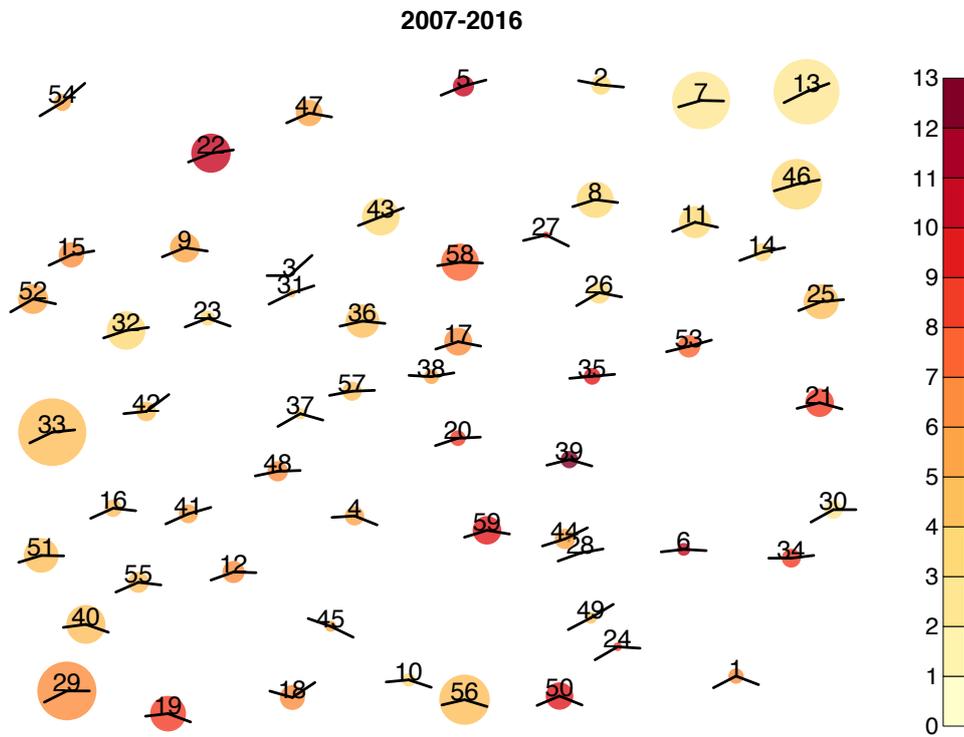


図 2.12: 経年的な変化の情報を追加した可視化

第3章 災害時のTwitterデータの活用

3.1 災害時のツイートデータ及び有益情報の可視化システムの開発

3.1.1 ツイートデータの可視化の研究背景

近年日本では、地震や台風、噴火などの災害が多発している。災害発生時は、電話などのインフラが不安定になり家族や知人と連絡をとることができなかった。その時インターネットは、電話と比べて通信が安定していたためSNSを利用した情報の発信と収集が頻繁に行われた。SNSの利用は、個人ユーザのみでなく企業や地方自治体も活発に利用して情報発信を行なっている。そのため災害状況に関するSNS上には有益な情報が集まる。また、SNSでは速やかに情報を拡散することもできる。このような背景からSNSを活用及び分析することで、現在や将来の災害に対する減災に繋がると考えられる。災害時のツイートデータの可視化に関する研究では、ツイートされた位置を地図上にマッピングして災害や救援要請などの位置を示す研究が盛んに行われている。例えば、Uchida et al.(2015a)では、現在位置の市町村に関する災害情報のツイートを地図上に表示している。また、画像の添付の有無などによってマッピングに使用するアイコンを変えている。渡邊(2014)は、「Real-time Typhoon Watcher」という、ツイート情報と台風に着目した減災支援システムを開発している。このシステムは、Google Earth上にツイートや雲、台風の軌道などを可視化するシステムである。しかし、多くのツイートデータには緯度経度からなる位置情報であるジオタグが付加されていない。ジオタグを付加することで自宅や職場などが特定されてしまうため犯罪に巻き込まれる恐れがあることも要因のひとつと考えられる。橋本 他(2012)の調査では、ジオタグ付きの日本語のツイートは全体の0.18%であったと報告している。そのため地図上にマッピングできるツイートの割合が少ない。また、災害発生場所や救援が必要な場所を正確に伝えるためにもジオタグが重要となる。支援をする側も場所がわからないと支援のしようがない。そこで、Uchida et al.(2015b)では、開発したアプリケーション

からの投稿を示すハッシュタグや位置情報のハッシュタグ、現在地の住所を自動でツイートの中に付加するアプリケーションを開発している。このアプリケーションによりツイート内に位置情報が統一した形式で記述されるため、ツイート内から位置情報を容易に抽出することができる。

本研究では、ツイート内の地域名を活用して災害時に役立つ地理情報システム (GIS) のプロトタイプを作成する。さらに Web サービスを活用して、災害時に役立つ情報を収集できることを示す。

3.1.2 GIS 上へのツイート情報の可視化

デジタル地図は、Google や Yahoo! Japan, 国土地理院など様々な企業や団体が有償または無償で提供している。これらの地図の多くは、Web API により使用することができる。本研究では、Yahoo! Japan が提供している地図を使用する。この地図は、Yahoo! JavaScript map API から容易に使用することが可能である。

災害が発生したとき多くのユーザが一斉にツイートするため、ツイート数が多い地域が被災地であることが分かる。そこで、ツイートした位置を地図上にマッピングした。また、地図上にツイートした位置をマッピングすることで、どの地域でどのような支援が必要なのかなどの情報を整理することもできる。ツイートされた位置は、ジオタグにより把握することができる。しかし、ジオタグが付いてない場合マッピングする位置を決める必要がある。そこで、ジオタグが付いてないツイートについてツイート内に位置情報が示されているものは、その位置情報からマッピングする位置を決めた。文章中から地名を抽出する方法は、いくつかある。例えば、形態素解析により地名と推定された単語を抽出する方法がある。また、Uchida et al.(2015b) で開発された DITS では、ツイート内に「#平塚市災害」のような市区町村名を示すハッシュタグや現在地の住所（「神奈川県平塚市北金目」周辺）が自動で付加されているため、容易に地名を抽出することができる。地名（テキスト）を取得しても GIS 上にマッピングするには、地名を緯度経度などの座標データに変換する必要がある。本研究では、地名を座標データに変換するために Yahoo! ジオコード API という Web API を使用した。これにより各ツイートを GIS 上にマッピングすることができる (図 3.1)。

災害情報



図 3.1: GIS のイメージ図

マッピングすることで災害発生地域や支援が必要な地域などを把握することが可能となる。さらにマーカーをクリックするとツイートと投稿日時、ユーザ名が表示される仕様にすることで、情報の内容や日時を把握することができるため各地域の状況把握や対策に活用することができると思われる。

3.1.3 Web サービスを活用した情報収集と GIS への可視化

災害時に各々の情報を別個に収集すると手間がかかり減災につながる情報の提供により迅速な行動を促すことに影響がある。また、被災時は精神的にも円滑に情報収集することが難しいと想像することができる。そこで、前節のツイートがマッ

ピングされた GIS に災害や減災に有益な情報追加してひとつの可視化で完結するようにする。情報の収集は、Web API を活用することで新旧の情報を迅速に収集及び更新をすることが可能となる。本研究では、災害状況の把握と被災時に命を守るための情報収集を目的として考え、気象情報を GIS に追加した。災害が発生する気象現象として台風やゲリラ豪雨などの大雨が考えられる。雨による災害の場合、雨量がある地域や雨雲の位置や雨雲の推移が分かれば避難する方向や今後の雨量を予測することができる。また、大雨、強風、雷などの警報や注意報の情報も減災のための意思決定に有益な情報となる。そこで雨雲の情報収集は、Yahoo! が提供する Web GIS に雨雲レーダーを地図に重ねるサービスを使用することで実現した (図 3.1)。警報と注意報の情報は、Yahoo! Japan が配信している、Yahoo! 天気・災害という Web サービスから収集することができる。このサービスは、RSS ファイルで提供される。このデータには、各地域の警報・注意報の有無とその詳細が書かれた Web ページの URL とこの情報の発表日時が XML 形式で格納されている。このデータから地域名 (テキスト情報) を抽出して地域名をもとに地図上にその地域に色を付けて警報や注意報を可視化した。図 3.1 において、警報が出ているエリアを赤く塗り、注意報が出ているエリアを黄色に塗った。さらに、警報・注意報の詳細が書かれた Web ページへのリンクを集めたページも作成した。

3.1.4 減災支援システムのプロトタイプ作成

様々な気象などに関する Web サービスでは、欲しい地域の情報を得るには、地域を選択したりする作業をする必要がある。そこで都道府県と市町村区を入力することで参照したい地域情報が表示されるページに行くことができる仕様にした (図 3.2)。さらに、GPS を利用することで現在地の地域を特定してその地域情報を表示する機能も組み込んだ。GPS の緯度・経度を都道府県に変換することは、Yahoo! リバースジオコード API を使用することで可能になる。URL の引数を変えることで地域を指定することができる仕組みを利用し、Web サイトのリンク先を決めた。また、ツイートからの情報は、Uchida et al.(201b) で開発されたアプリケーションから投稿された情報を活用することにする。このアプリケーションでは、ツイートに投稿した住所と救援が必要かどうかを自動で追加される。また、このアプリケーションから投稿したことを示すハッシュタグも追加される。このハッシュタグを含むツイートを定期的に取り得して、データベースに取り込むことで、このアプリケーションから投稿した情報を更新することができる。

減災支援システムの GIS のページが図 3.3 である。GIS 上では、警報や注意報が出ていることが分かるが、何の警報や注意報が出ているのか把握することができない。Yahoo! 天気・災害では、警報や注意報が出ている地域だけでなく、詳細が書

第3章 災害時の Twitter データの活用

かかれている Web ページの URL も取得することができるため、その URL のページへのリンクを画面右側に作成した。これにより警報や注意報の詳細な情報を把握することが可能となった。また、このシステムで利用している API の概要図が図 3.4 がある。Web サービスを活用することで、ツイートデータや気象情報、緯度経度の座標データを使用することができる。この様に複数のサービス（データ）を組み合わせることで有益な情報を生み出すことができる。



図 3.2: 都道府県と市町村区を入力するページ

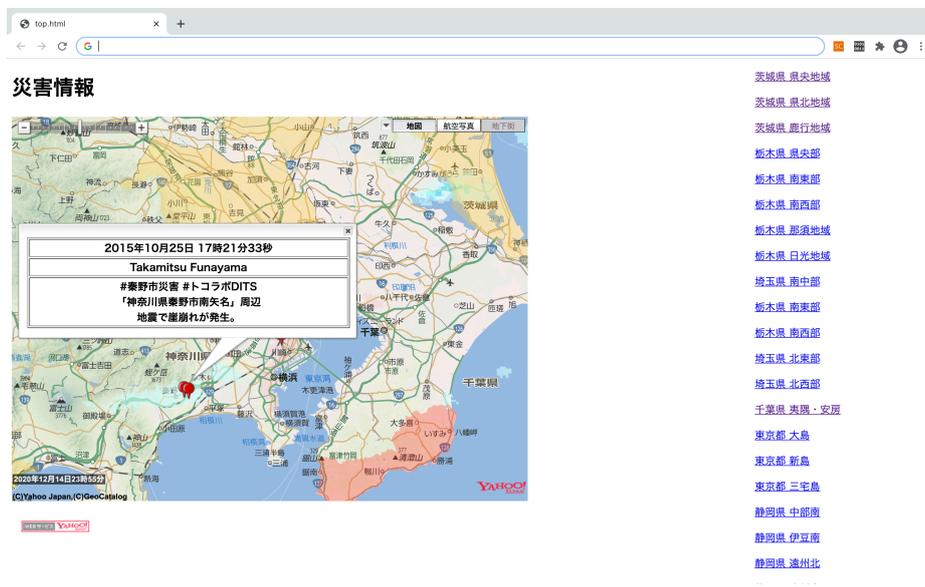


図 3.3: 減災支援システム

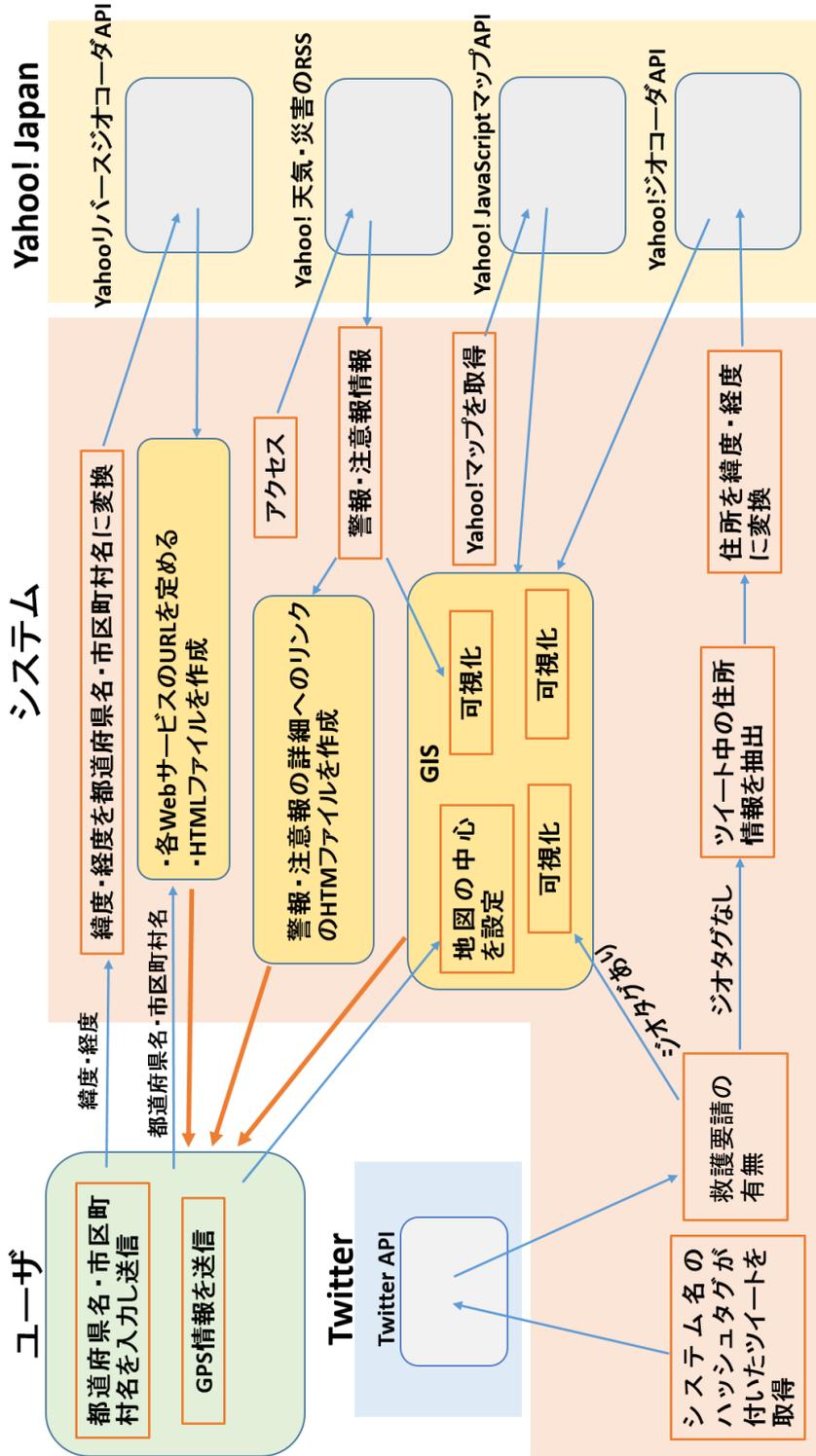


図 3.4: システムの概要図

3.2 災害時のツイートデータの解析と可視化

3.2.1 災害時のツイートデータの解析と可視化の研究背景

近年日本では、地震や洪水などの自然災害が多く発生している。最も被害が大きかったのが、2011年3月に発生した、東日本大震災である。この震災では、地震と津波により広範囲に被害が及ぼした。このとき、多くの人々が、交通機関が乱れ、帰宅ができなかったり、住居が倒壊し避難所での生活を余儀なくされた。被災者は、ラジオやテレビ、携帯電話から情報を得ることができたが、電話が利用できなかったため、家族と連絡を取ることができなかった。その時に、TwitterなどのSNS(Social Networking Service)が情報伝達に利用され、Twitterからの情報により命を救われた人もいる。多くの人々が電話で、家族と連絡を取ろうとしたため、電話が使えなかったが、インターネットは、通常通り利用することができたからである。そのあとに起きた地震や洪水などの自然災害の時にも情報伝達にSNSが活躍した。Twitterは、ユーザ数も多いいため様々な情報を得ることもできる。しかし、Twitterから有益な情報を抽出することは、容易なことではない。なぜならば、災害時に全国のTwitterユーザが一斉にツイートするため、被災者の救助を求めるツイートや被災状況などの有益なツイートが埋もれてしまうからである。有益な情報のツイートを見逃さずに抽出できれば、被災状況や崖崩れが起きそうな危険箇所、自力避難が困難な人などの情報を速やかに把握でき、一人でも多く人命を救うことができるかもしれない。そのため災害発生時のTwitterデータの活用に関する研究は、活発に行われている。山陰地方豪雪災害時には、あるユーザが「#sanin_snow」というハッシュタグを使った情報共有を呼びかけ、大規模な災害コミュニティが形成された事例もある。このTwitterデータを気象情報や被害報告、呼びかけ、被災者への励ましなどに分類し、災害状況の把握や情報の信頼性、ユーザの特性などを分析する研究も行われている(石川 他, 2012)。Yamada et al.(2018)では、大阪府北部地震のときのツイート数、ハッシュタグの使用傾向、絵文字の使用傾向の分析を行なっている。

このように災害時にTwitterが活用されているため、ツイートデータには、様々な情報が集まっている。本節では、このツイートデータから有益な情報を抽出することを試みた。そこで、本研究では、テキストマイニングを用いたツイートデータの可視化による情報の抽出を行なった。また、ツイートされた日時によりデータを区切り、その時間帯ごとの抽出される情報の違いや特徴の考察も行った。本研究では、2016年4月に熊本県を震源とした熊本地震後のツイートデータを用いて説明する。

表 3.1: 収集したツイートデータ

キーワード	ツイート数
#救助	48,222
デマ	150,314
救助	372,800
熊本	2,310,704
地震	9,235,067
東海大	11,795
避難	2,734,384
安否	30,077
合計	14,893,363
重複削除後	5,537,928

3.2.2 被災状況の情報抽出

本研究では、Twitter API を用いて、熊本地震に関するキーワードとして「地震」「避難」「安否」「救助」「#救助」「熊本」「東海大」「デマ」の8つ(表 3.1)を設定して収集した。収集したツイートは、14,893,363あり、重複するデータを除くと4,868,308ツイートになった。また、収集されたツイートデータは、2016年4月12日1時58分から2016年04月21日14時20分の期間にツイートされたデータである。

まず、収集した全てのツイートデータを用いて、分析を行い有益な情報の抽出を試みたが困難であった。理由は、全てのツイートデータには、様々な情報が含まれているため、欲する情報以外の情報がノイズになるからである。例えば、「地震」をキーワードにしたため、自動投稿の地震速報のツイートがたくさんあり、このツイートがノイズになる。また、災害発生時、多くの人がツイートすることによって、ツイートデータは、大規模になり、計算コストも大きくなる問題もある。つまり、高スペックな計算機が必要になる。このような問題があるため、本研究では、抽出する情報を決め、それに関係するキーワード(表 3.2)を設定し、そのキーワードを含むツイートデータを分析することとした。熊本地震では、車の中での避難生活について、マスメディアからの報道が多かったため、車での避難生活に関する情報を抽出した。また、本研究ではリツイートを除いたオリジナルのツイートのみを解析対象とした。

表 3.2: 車での避難生活に関係すると考えられるキーワード

車	車中泊	駐車場	車中	車内
電車	車両	車内泊	車中泊避難	駐車場不足
自家用車	自動車	車内避難	車避難	車生活
車中避難	駐車	満車	車泊	車内生活
自動車内	避難車	乗用車	車移動	車待機

3.2.3 車での避難生活に関する情報抽出

避難に関する情報を抽出するため、「避難」をキーワードに設定して収集したツイートデータを使用する。そして、そのデータを形態素解析して、対応分析を用いて、情報の抽出を行う。形態素解析では、NTT データ数理システムの Text Mining を使用する。車での避難生活に関する単語は、「車」、「車内」、「車中泊」、「駐車場」など色々考えられる。これらの多くの単語に「車」という文字が含まれているので、「車」を含む単語を集計し、出現頻度が多い単語の中から車での避難生活に関係すると考えられる単語(表 3.2)をキーワードとする。そして、そのキーワードを含むツイートデータを分析対象データとする。これによって、車での避難生活に関係ない情報を減らすこととデータサイズを削減することができる。このデータのツイート数の時系列の推移(図 3.5)を見ると、前震と本震の直後は、ツイート数が多く、前震後の日中より本震後の日中の方がツイート数が多いことがわかる。次に、表 3.3 のように前震から 6 時間ごとに区切って、期間ごとに情報の抽出を行い、時間帯での抽出される情報の変化を考察する。

第3章 災害時の Twitter データの活用

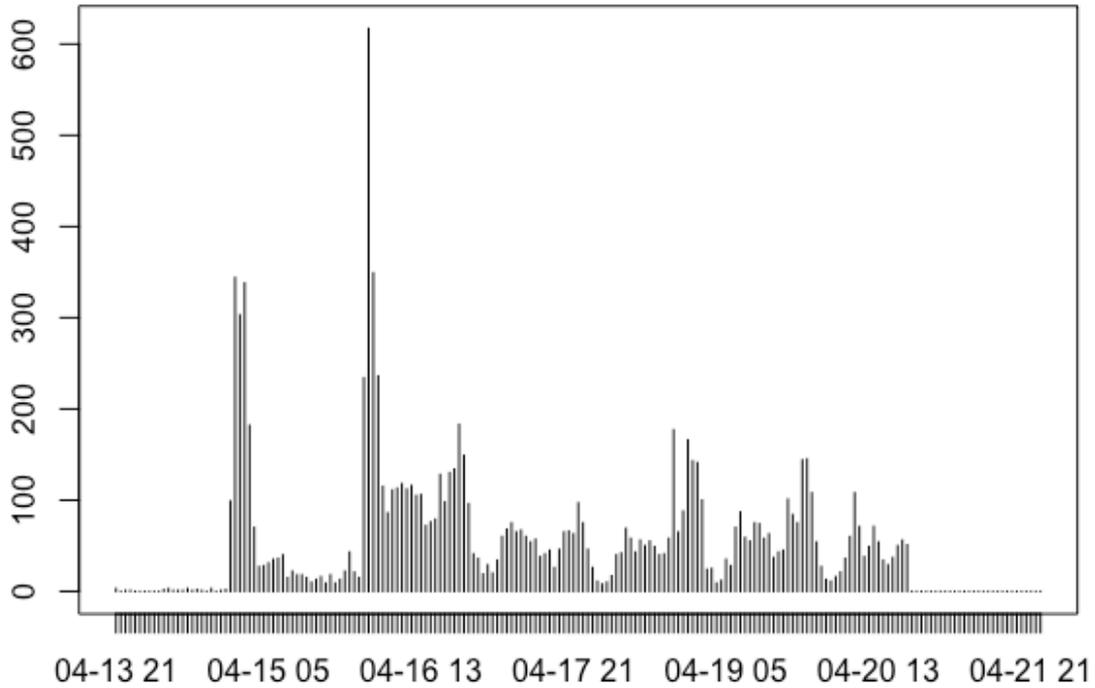


図 3.5: 「避難」を含むツイートデータのツイート数の時系列変化

表 3.3: 期間

期間番号	期間	備考欄
1	2016年4月14日 21:00 ~ 2016年4月15日 3:00	2016年4月14日 21:26 前震発生
2	2016年4月15日 3:00 ~ 2016年4月15日 9:00	
3	2016年4月15日 9:00 ~ 2016年4月15日 15:00	
4	2016年4月15日 15:00 ~ 2016年4月15日 21:00	
5	2016年4月15日 21:00 ~ 2016年4月16日 3:00	2016年4月16日 1:25 本震発生
6	2016年4月16日 3:00 ~ 2016年4月16日 9:00	
7	2016年4月16日 9:00 ~ 2016年4月16日 15:00	
8	2016年4月16日 15:00 ~ 2016年4月16日 21:00	夜間雨の予報

第3章 災害時の Twitter データの活用

本研究では、各期間で名詞と形容詞の共起行列を作成し、対応分析を用いて可視化して、関係性がある単語を抽出した。全単語を用いて分析を行うと共起行列のサイズが大きくなり、計算時間がかかり実用的ではないため名詞と形容詞を各上位 50 単語からなる共起行列を作成した。ただし、頻出単語が 50 単語に満たない場合は、その品詞の全ての単語を用いた。

図 3.6 が期間 1 の対応分析の結果の 1 軸と 2 軸を用いた可視化である。赤色の単語が名詞であり、黒色の単語が形容詞である。この可視化では、同じツイートデータに含まれる傾向がある単語が近くに表示される。また、出現頻度が多い単語は、原点付近に布置される傾向がある。例えば、「駐車場」と「痛い」、「駐車場」と「広い」、「心強い」と「揺れ」などである。これでは、「駐車場」と「痛い」がどのような意味をなす情報なのか理解できない。そこで、「駐車場」と「痛い」を含むツイートデータを抽出して、ツイートから情報を得ることにする。「駐車場」と「痛い」を含むツイートデータの一部が表 3.4 である。ツイートを確認することで、駐車場にブルーシートを敷いて避難している人がいるという情報をツイートデータ(表 3.5)から抽出することができる。また、「駐車場」は、「広い」とも近いので、同じツイートに含まれる傾向がある。このツイートデータを抽出すると「広い駐車場に避難している」という内容のツイートがたくさん抽出される。これにより車で広い駐車場に避難している人がたくさんいることがわかる。

表 3.4: 期間1の「駐車場」と「痛い」を含むツイート

ツイート
@marira55 ほんとだよ、命懸け。だんだん被害が明るみになってきていて心が痛い。益城町、対応大変そうだよ、避難所どこなのか知らないが駐車場にブルーシートって。
なんで学校の体育館とか開けないんだろう…なんでずっと市役所前や老人ホームの駐車場が避難所なんだろう…公共施設より個々の事業団のが判断早いのか… アスファルト座ってたらビニールシートの上では痛いよね、
ブルーシートが敷かれた駐車場で毛布に包まって…胸が痛いね 土日が雨降りそうな気配だから、できるだけ早くちゃんとした避難場所が確保される事を願うばかり https://t.co/WOpHtciBfd

表 3.5: 期間1の「駐車場」と「広い」を含むツイート

ツイート
外の広い駐車場に避難してるなう
広い駐車場に避難中。まだ揺れてやがる。
@marucat1119 ありがとー!!! o(T□T)o いちお、念のため広い駐車場に避難しよる (T_T) そこで妹とも合流予定＝
@karagirl625 正直やっばいです。棚のものの全部落ちました…w 震源地からめっちゃ近んで広い駐車場に避難してます！
一応近くのお店の広い駐車場に家族みんなで車で避難してます＝(＝°＝＝°)＝ マンション最上階だから何かあってからじゃ遅いからね
@umeshiho 今、帰宅途中なんだけど疲れたから広い駐車場にいる。避難してる人いっぱい。でも早くおうち帰りたい
職場の駐車場が広くてたくさんの方が車で避難してきてます。私はそこにいます。大丈夫。

期間6(図3.7)では、「外」と「寒い」が同じツイートに含まれる傾向がある。これらに単語を含むツイートデータ(表3.6)を抽出すると、車を持っていなくて避難していない人や寒いので避難所から車に避難場所を変えた人がいる情報を得る

表 3.6: 期間6の「外」と「寒い」を含むツイート

ツイート
車もってないから避難しようにも勇気がない 身一つで外でるのも寒いのできついのが本音
避難場所の中学校は危険だと言われた. ショッピングモールの駐車場が一時避難場所だと言われたけど, 外寒いし部屋で寝る
さすがに外での避難は寒いので, 一旦アパートに戻り, 車が動かせるのを確認. そのまま龍田中学校に避難場所を移動しました.

表 3.7: 期間6の”揺れ”と”長い”を含むツイート

ツイート
寝てたら昨日よりも激しくて長い揺れ. 停電して今も外は真っ暗なまま. 近くのスシローの駐車場が広いから避難中. ほんとに怖い.
@misepsycho みせせんさん, ご心配本当にありがとうございました. 前回よりも酷く長い揺れで, 車に避難して夜を明かしました. お気持ちだけでも, すごく救われました…!
車中泊してやっと夜が明けた. 昨日の揺れは長く強く怖かった… もう家の中に居るのが怖過ぎて 今から家の状況見て避難所行こうと

期間8(図3.8)では、「風」と「強い」が同じツイートデータに含まれる傾向がある。「風」と「強い」を含むツイートデータが表3.8である。このツイートデータから車から避難所に避難場所を変える人や自宅から車に避難する人がいることがわかった。この日の夜の雨に備えて避難している情報を抽出することができた。

表 3.8: 期間8の「風」と「強い」を含むツイート

ツイート
風も強くなってきたし 雨も降ってきたので 近くの駐車場じゃやっぱ怖いねってことで 小学校に避難して来ましたー 体育館の中はいっぱいに入れてないから 車の中
@jiyongsarage88 風も強いのが恐怖すぎる泣 今, 避難する準備した! 車に詰めたよ!
また夜がくる, 今日は風も強いし雨も降るって怖すぎる. ひとまず家族で道の駅に車で避難してるけど, ちゃんとした避難所がいいのかな. 避難所の建物が壊れないかも心配. どうしよう..
高専に避難. 車にいるけど風が強すぎて気持ち悪い. 体育館いこうかな
@caori_co 雨, 意外と遅い降りですね. 風やばくないですか?今日は私だけ避難所の駐車場にいます. 私の住んでる地域は土砂崩れ危険地域でさすがに強かったの. ただ父と犬は残してきたので心配です. 一人ぼっち. 何事もないように祈ってます
とりあえず中学校のグラウンドに避難してる…体育館人いっぱいらしいから車内に!でも風が強くて車が余震で揺れてるのか風なのかわからない…

対応分析は、頻度が多い単語が原点付近に布置される傾向がある。そのため、原点付近に単語が密集してしまい、原点付近に布置された単語を把握することができない。原点付近に布置された情報は、頻度が多い情報のため多くの被災者の声が集まっている可能性がある。図 3.9 が、図 3.8 の原点付近を拡大した図である。原点付近を観察すると「近く」と「危ない」や「よかった」と「物資」などが近くに布置されていることがわかる。「近く」と「危ない」を含むツイートを確認すると被災状況や被災者へのメッセージであった（表 3.9）。また、「よかった」と「物資」を含むツイートを確認すると避難所の状況や情報であった（表 3.10）。

第3章 災害時の Twitter データの活用

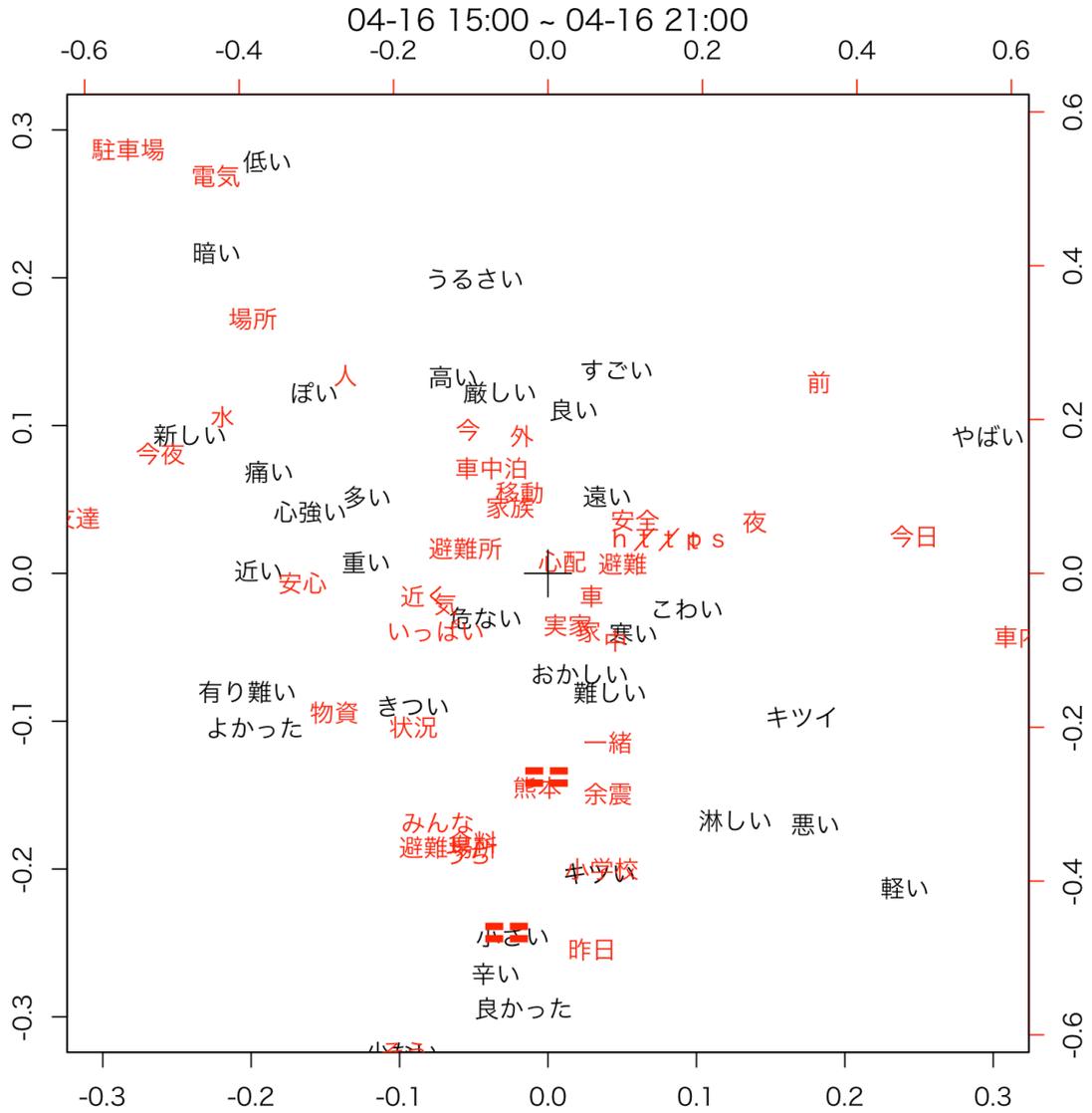


図 3.9: 期間 8 の対応分析による可視化の原点付近

表 3.9: 期間8の「近く」と「危ない」を含むツイート

ツイート
@MArFA5vUsQmHXP9 家危ないし近くに土砂崩れしそうなどこあるんで避難して車の中で過ごしてます。(; ㄥ `)
@akaponzzz あっきーねさんとお子さん達が無事でとにかく良かった！家の中の悲惨な状況に驚いています。車中生活は大変だと思いますが、まだ余震も続いているし、家の中は危ないから致し方ないですね。近くにペット可の避難所があればいいのに。あっきーねさん、ファイトだよ！！
正規の避難所ではないものの、自分も待機した最寄りの大き目の駐車場から「今はまだ動くと危ない、一旦ここで待機しよう」と呼びかけ、ある程度落ち着くまで2時間弱、近くを通っていく人を呼び集めて、しばらく待機し、ある程度落ち着いてからみんなで移動を行いました。(続く)
私の大事な妹が、熊本の、震源地のすぐ近くに住んでいて、もう家の中メチャクチャになって、危ないから避難所の駐車場に車停めて寝泊りしてるらしくて、もう心配で仕方ない…！考えただけで泣きそう。お願いだから無事でいて！ニュース見る度に胸が痛くなる…

表 3.10: 期間8の「よかった」と「物資」を含むツイート

ツイート
今避難所がいっぱいで、南高江のベルエアの駐車場に避難所に入れなかった人たちがたくさん避難してきてます。近くのコンビニなどもすべて品切れで近くのトイレも詰まって不便な状態です。よかったらこちらにも物資を、わけて下さったら嬉しいです。#拡散希望 #熊本地震
@tamami0202 そういうことか。情報共有の時だけでも同じ車になよ。そっか、よかったよかった。室小は割としっかりした避難所みたいだから物資も届くみたいだし動けるときに避難しておいた方が、いいかも

この様に対応分析の可視化より、関係性のある単語を抽出し、それらの単語を含むツイートを確認することで情報を把握することが可能であることを示した。つ

まり、大規模なツイートデータをトピックごとに分類することができたということである。これにより大規模なツイートデータから効率的に情報を抽出することができるようになった。

3.2.4 対応分析による可視化の改良

図 3.6 の様な対応分析の可視化は、単語間の距離から情報を把握するすることができるが、その情報の信憑性や対象者の量を表す単語の出現頻度を把握することができない。また、単語間の距離は、近いがそれらの単語を含むツイートが存在しないことがある。これは、誤った情報の抽出である。この様なことを防ぐために、ツイート数と共起情報を可視化に追加した。単語の出現頻度は、円の大ききで示し、同じツイート内にある単語は、線で繋いだ。さらに共起数は、線の太さで示している。この改良により図 3.10 は、原点に近いほど単語の出現頻度が多く、原点から離れるほど出現頻度が少なくなっていることがわかる。また、原点付近の単語は、周辺の単語と共起関係があり、共起数も多いことがわかる。一方、原点から離れた単語は、周辺の単語と共起関係が無く、原点付近の共起関係がある。従って、近くに布置された「食料」と「確保」から「食料の確保」に関するツイートが多いという様な誤った情報の抽出を防ぐことが可能になる。

第3章 災害時の Twitter データの活用

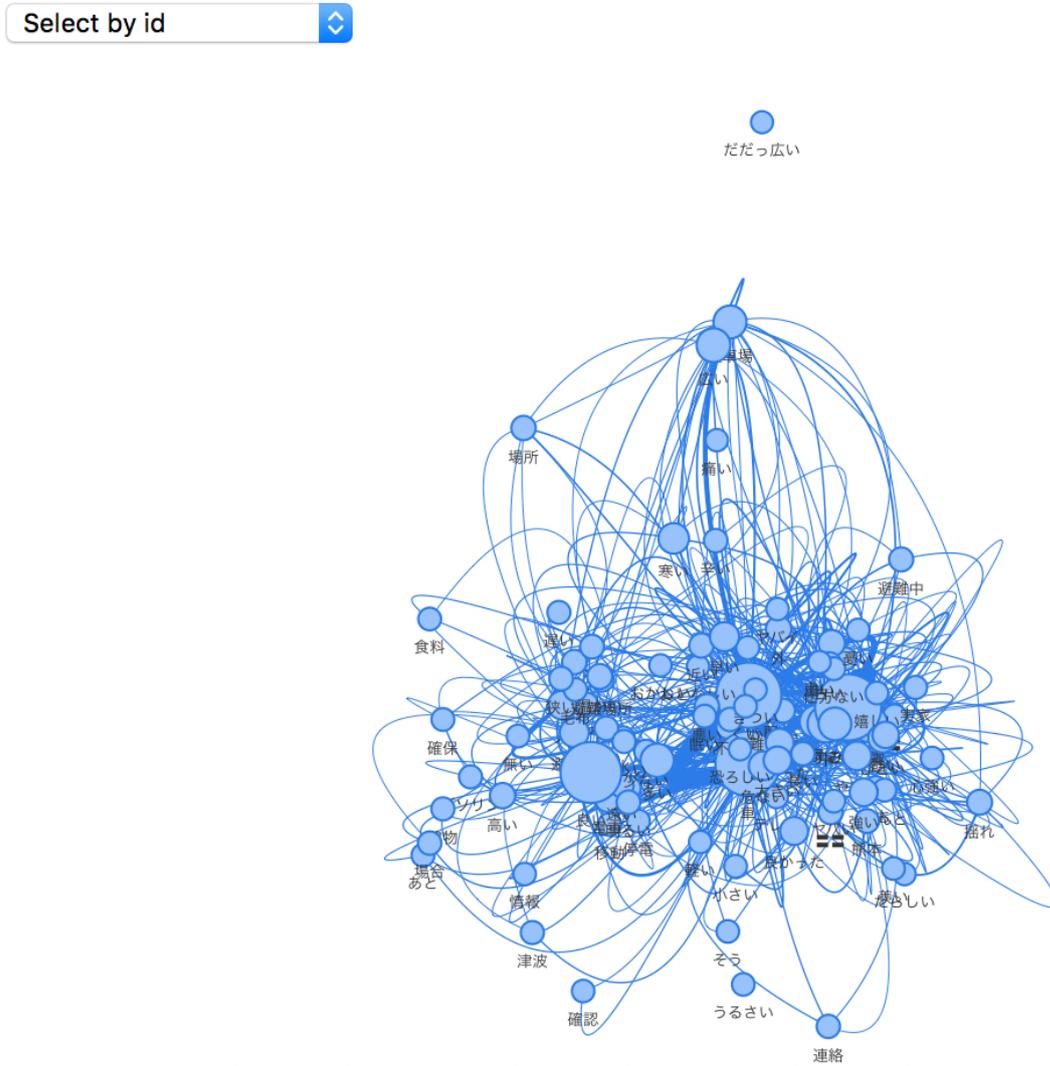


図 3.11: R の visNetwork パッケージを使ってインタラクティブな可視化

ることが可能となった。

本節で提案した可視化を実装するための Web アプリケーションを RStudio の Shiny パッケージ (Chang et al., 2019) を用いて作成した。図 3.13 が、本アプリケーションのインターフェースである。このアプリケーションでは、データ収集したときのキーワードを選択することで、解析対象のデータセットを選択することができる。さらにデータを絞り込むためのキーワードを CSV ファイルで入力する仕様にした。これにより複数のキーワードを設定することを可能にした。また、表示する単語数を変更したり、共起数の下限値や期間などを選択することもできる様にした。これらの設定により可視化された図が表示される。さらに対応分析の可視化から関係性がある単語を抽出し、その単語が含まれるツイートを表示できるようにした。また、2つの時間帯の対応分析の可視化を比較できるように並べて表示するページも作成した (図 3.14)。このページは、タブで表示を切り替えることができる。このアプリケーションにより様々な条件でツイートデータをフィルタリングして対応分析の可視化ができるため、探索的に情報を抽出して、被災状況の把握などが可能となった。また、2つの時間帯の対応分析の可視化を並べて表示することで、経時的な状況の変化も把握することも可能である。

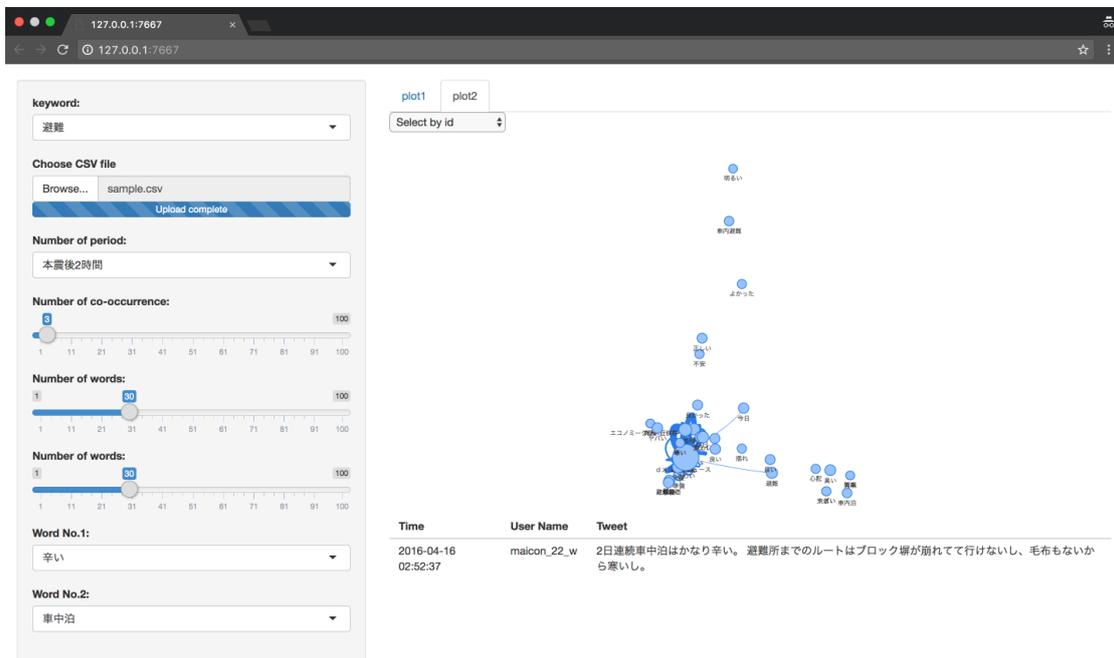


図 3.13: 探索的に情報を抽出するための Web アプリケーション

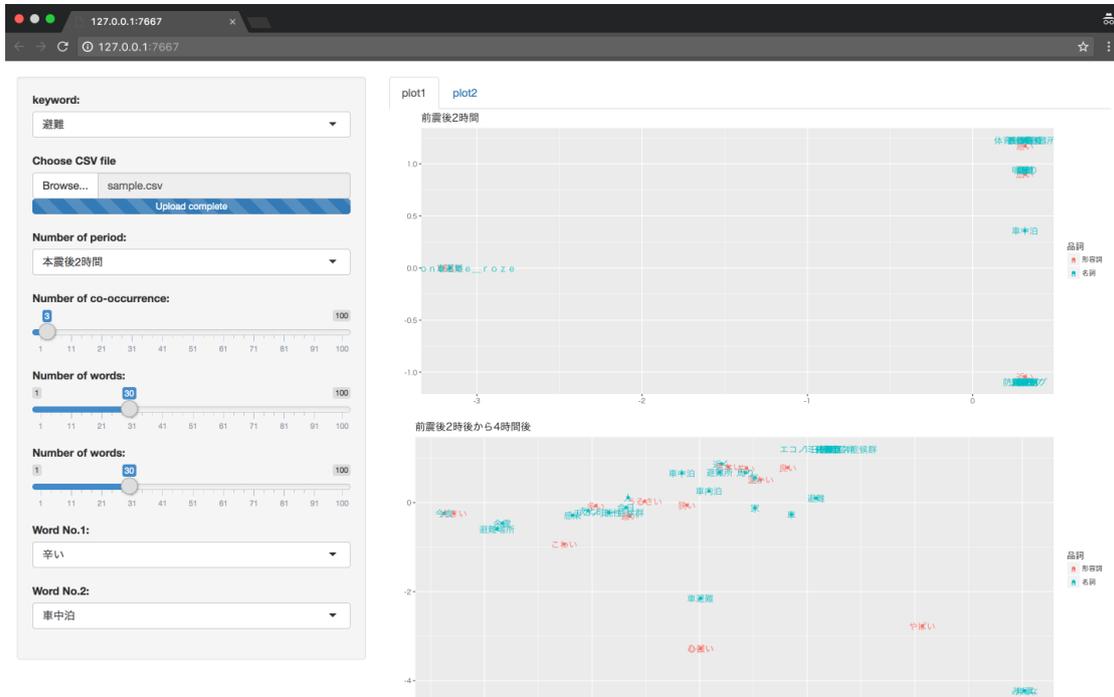


図 3.14: 2つの時間帯の対応分析の可視化を並べて表示する Web アプリケーション

3.2.5 経時的な可視化の改良

各時間帯の可視化を比較する場合、毎時単語の布置が異なるため各時間帯の特徴を理解することが難しい(図 3.15)。そこで、主要単語の布置を類似させる改良を行った。主要単語が毎時類似する位置に布置されることで主要単語の周辺に布置された単語の違いや距離を比較することが容易になる。本研究では、出現頻度が多い名詞上位 10 単語、「車」「避難」「避難所」「中」「家」「駐車場」「余震」「人」「今」「熊本」の布置を類似させることとした。

第3章 災害時の Twitter データの活用

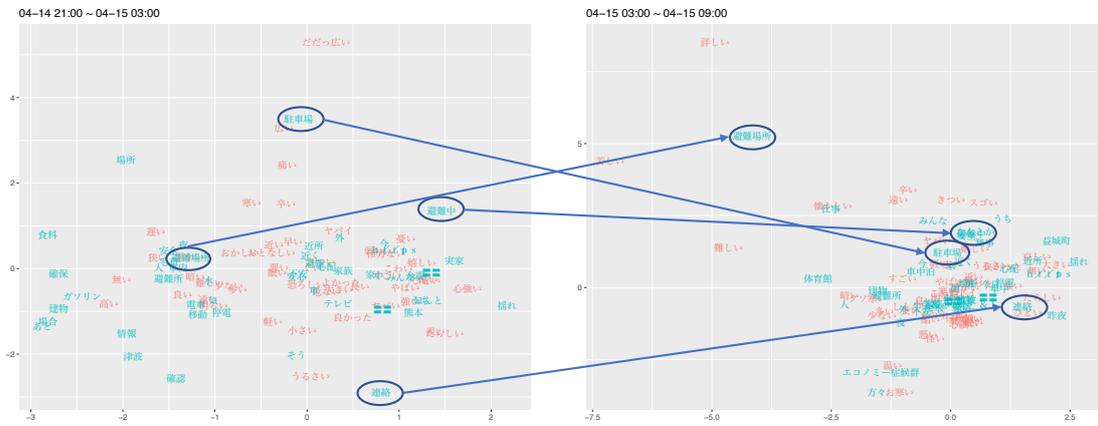


図 3.15: 可視化の比較

まず1つ目の方法が、軸の反転による調整である。縦軸と横軸の反転により布置させる象限が最初の時間帯の布置と一致する割合が最も高い反転方法をもとに決めることにした。例えば期間2(図 3.17)の「熊本」と「避難所」が布置された位置と他の期間の布置(図 3.16, 図 3.18, 図 3.19)を比較すると全く類似しなかった。この原因として基準とした布置が他の布置と比較して特徴的である可能性がある。そのため類似させることが困難となる。そこで直前の布置を基準とすることにした(図 3.20, 図 3.21, 図 3.22, 図 3.23)。

第3章 災害時の Twitter データの活用

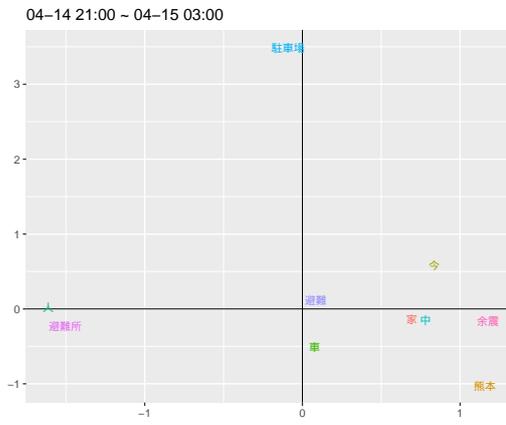


図 3.16: 期間 1

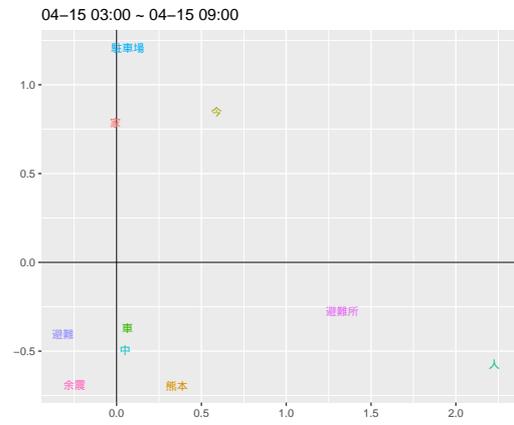


図 3.17: 期間 2

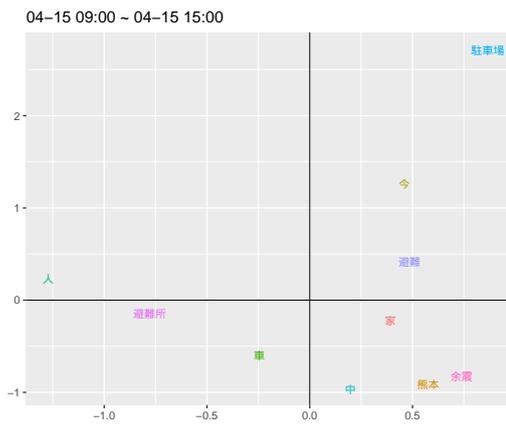


図 3.18: 期間 3

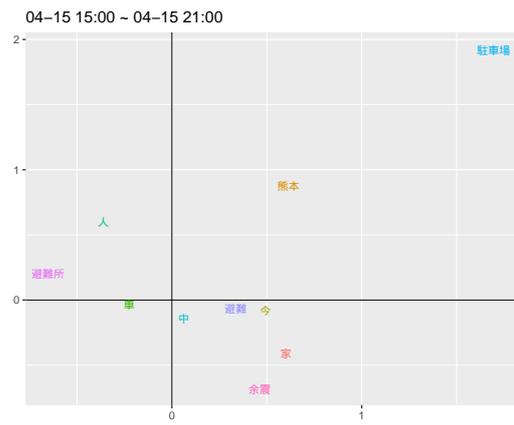


図 3.19: 期間 4

第3章 災害時の Twitter データの活用

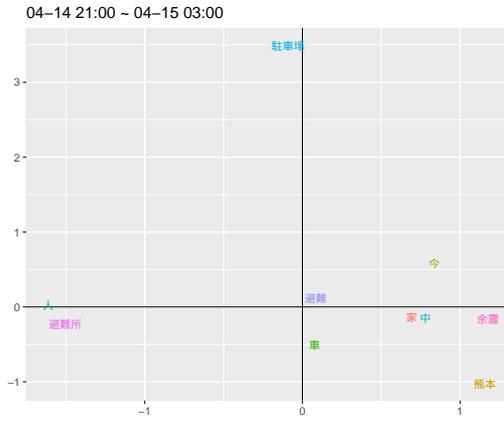


図 3.20: 期間 1

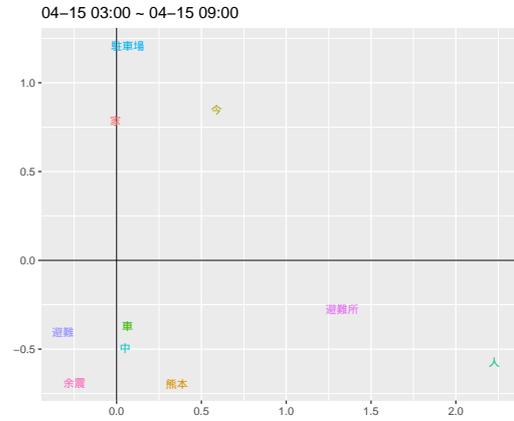


図 3.21: 期間 2

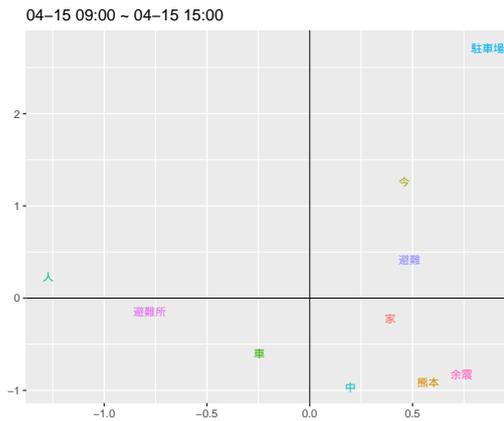


図 3.22: 期間 3

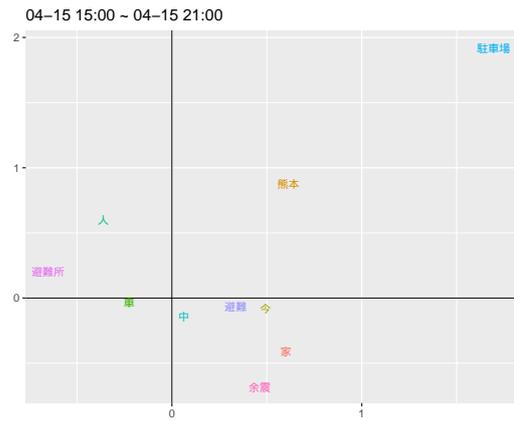


図 3.23: 期間 4

しかし、今回のデータでは改善させることができなかった。「避難所」は、期間 1 で第 3 象限に布置されたが、期間 2 で第 2 象限、期間 3 で第 3 象限、期間 4 は、第 4 象限に布置され、期間 2 以外では左側に布置される。しかし、「熊本」は、全期間で右側に布置されている。軸の反転では、ある単語の布置を類似させても違う単語の布置が異なってしまった。

2 つ目の方法として、原点を中心に回転させて調整することで近付ける。 t 時の布置の調整方法は、次のようになる。 α 度回転したとき t 時と $t-1$ 時との布置との

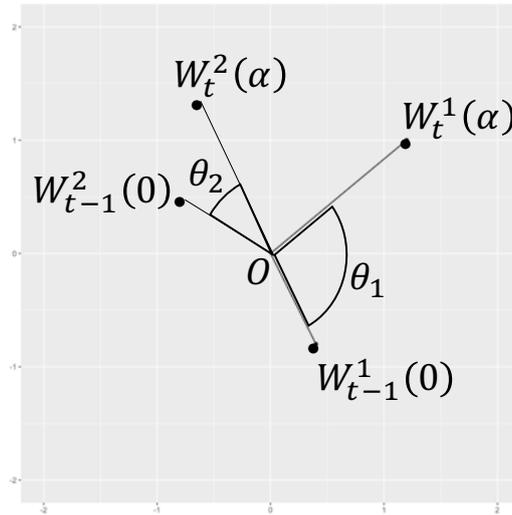


図 3.24: 回転による調整

コサイン類似度の和を S_α は、次のように定義する。

$$S_\alpha = \sum_{i=1}^N \cos\theta_i$$

このとき、 S_α が最大となる角度 θ を最も類似する布置とする。図 3.33 を用いて説明すると、基準となる $t-1$ 時の単語 1 と単語 2 を $W_{t-1}^1(0)$, $W_{t-1}^2(0)$ とする。 t 時の α 度回転したときの単語 1 と単語 2 が $W_t^1(\alpha)$, $W_t^2(\alpha)$ である。このとき単語 1 と単語 2 の $t-1$ 時と t 時のなす角が θ_1 と θ_2 となる。この θ_1 と θ_2 のコサイン類似度の和が S_α である。本研究では、 α を 1 度ずつ変化させて S_α が最小となる α を見つけた。

回転により調整した布置が図 3.25 から図 3.28 である。全期間で「駐車場」が上部、「避難所」を左側に布置することができた。さらに布置を類似させるために、軸の反転と回転による調整の組み合わせることでより多くの布置のパターンが生み出すことができるため、最適な布置を見つけ出すことができると考えた。図 3.29 から図 3.32 が、軸の反転と回転による調整を組み合わせた最適布置である。「駐車場」が上部、「避難所」が左側、「余震」が右側など類似する位置に布置することができた。

図 3.29 の期間 1 では、原点の近くに「避難」があるため期間では、「避難」が多くのツイートに含まれていたことがわかる。図 3.30 の期間 2 では、「避難」が右端にある「余震」の近くに布置されているため「避難」は「余震」と一緒にツイートされる傾向に変化したことがわかる。図 3.31 の期間 3 では、「余震」の近くに「熊

本」が布置されているが、図 3.32 の期間 4 では、「家」が「余震」の近くに布置された。このように軸の反転とコサイン類似度による回転とを組み合わせることにより各時間帯の布置を類似させて各時間帯の比較をすることが容易になった。図 3.33 から図 3.36 が、最適布置にした可視化である。黒字が、最適布置するための基準(重要単語)とした単語であり、赤字が形容詞、緑字が名詞である。図 3.34 の様に外れ値があると原点付近に密集してしまうが、重要単語の位置関係は類似するため各期間との比較には影響がない。「避難所」の近くに布置された単語を比較すると、期間では「狭い」「毛布」「車内」「暗い」「温かい」など初期の避難所の様子や必要な物資に関することだと思われる単語が近くに布置されている。期間 2 では、前震発生後の朝方のツイートのため「建物」「寒い」「しんどい」など夜間の避難所生活の様子に関することだと思われる単語が布置された。期間 3 では、昼間の時間帯であるため「情報」「良い」「多い」「3.11」「すごい」など東日本大震災の時のことなど情報の発信が盛んだったのだと思われる。期間 4 では、「有り難い」「人たち」「明るい」などの単語が近くに布置されているため、支援が届いていることを想像することができる。この様に重要となる単語の布置を類似させることで、経時的な比較が容易になる。

第3章 災害時の Twitter データの活用

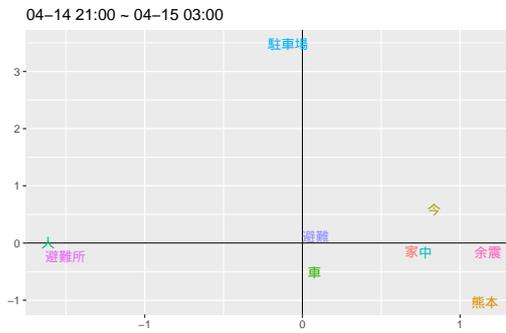


図 3.25: 期間 1

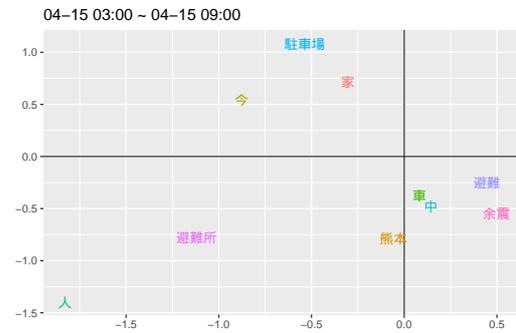


図 3.26: 期間 2

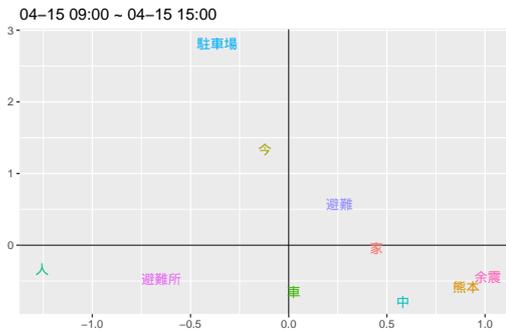


図 3.27: 期間 3

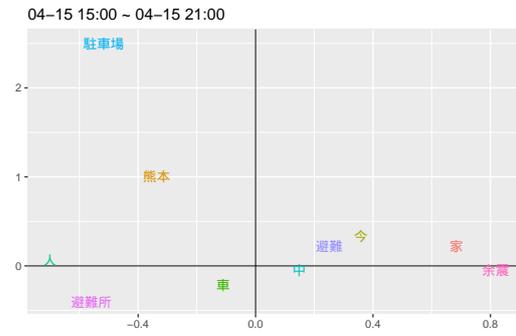


図 3.28: 期間 4

第3章 災害時の Twitter データの活用

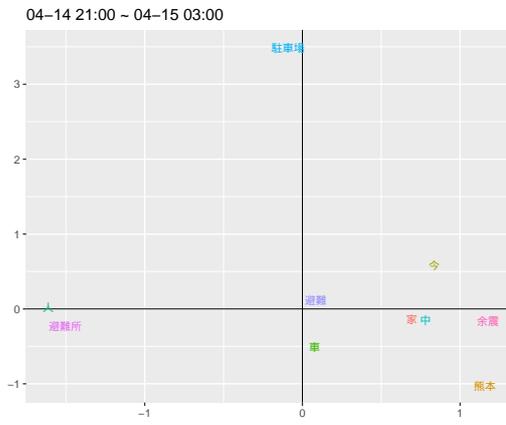


図 3.29: 期間 1

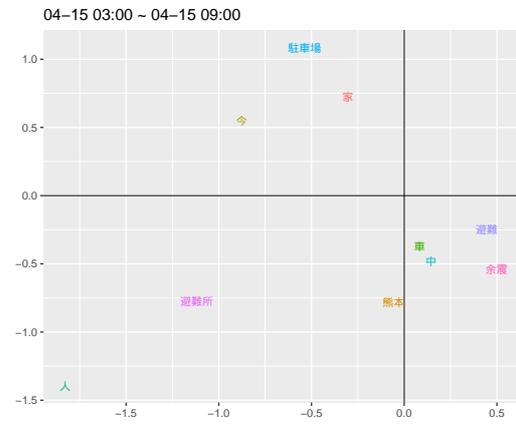


図 3.30: 期間 2

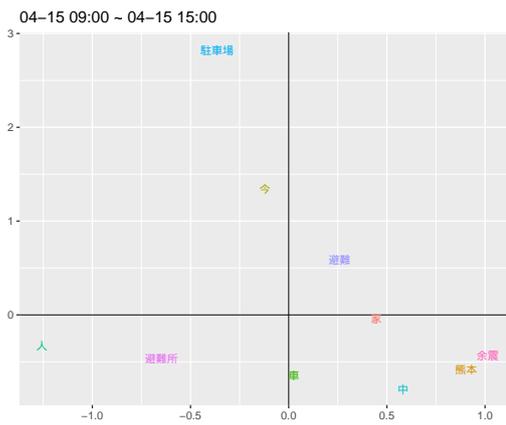


図 3.31: 期間 3

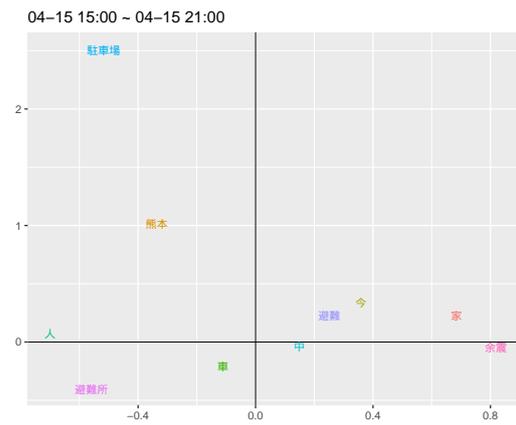


図 3.32: 期間 4

第4章 気象データの活用

4.1 研究背景

近年、電力システムのスマート化や再生可能エネルギー導入、エネルギー市場の自由化などが注目されている(内田, 2014)。これらの背景として化石燃料をやめて再生可能エネルギーを使用する社会的な動きがある。化石燃料は、安定したエネルギー源として使用されてきたが、有限な資源であり、特定地域でしか採取できないため経済問題もはらんでいる。化石燃料を資源とするエネルギーは、地球温暖化などの環境問題の原因と考えられている。また、東日本大震災のときに起きた福島第一原子力発電所事故もあり、原子力エネルギーの比重を減らし、再生可能エネルギーに比重を増やす方向に向かっている。更に電力事業に市場競争を導入して経済的効率化を目的として、電力事業を官営から民営へ比重を移す動きがある。これらを実現するためには、エネルギーの需要と供給などを管理・制御するエネルギーマネジメントシステムが必要となる。特に再生可能エネルギーは、気象状況の影響を大きく受けるため、制御するために気象情報が重要な役割を果たす。

国立研究開発法人 科学技術振興機構 (JST) で行われている戦略的創造研究推進事業のプログラムの1つ CREST において、2015年度から2019年度に「分散協調型エネルギー管理システム構築のための理論及び基盤技術の創出と融合展開」(藤田, 2020) という研究プロジェクトが遂行された。この研究では、「再生可能エネルギーをはじめとした多様なエネルギーの需給の最適化を可能とする、分散協調型エネルギー管理システム構築のための理論、数理モデル及び基盤技術の創出」を戦略目的としていた。このプロジェクトでは、5つのチームが編成され、電力工学、制御工学、情報科学、経済学、気象学などの分野間連携を強め、「分散協調型エネルギーマネジメントシステム」の基盤技術の創出と社会実装に向けた研究が行われた。5つのチームの中に、「分散協調型 EMS における地球科学情報の可用性向上とエネルギー需要モデルの開発」(中島, 2020) を研究テーマとして、エネルギーに関する地球情報とエネルギー需要の状態や変動をそれぞれ理論で説明できる科学要素として捉え、これらの学理基盤の確立を通じてエネルギーマネジメントシステム (EMS) に貢献することを目的とする研究チームがあった。この研究チーム

は、地球科学とエネルギー需要科学の異なる研究分野の研究者たちが共同研究を行っていた。それぞれの研究分野の研究成果や生成物をやり取りしながら研究を進めていた。その中で特にデータのやり取りが重要になる。地球科学分野で生成されるデータは、大量で複雑なデータ構造であり、CSV形式などのテキストファイルではサイズが大きくなってしまいうため日常的にバイナリ形式が採用されている。しかし、エネルギー需要科学分野では、バイナリ形式のデータを扱うことは稀であるためエネルギー需要科学の研究者にとって難易度が高い。更に、バイナリ形式は、一般的な表計算ソフトで開くことができないためプログラミング技術が必要となる。これもまたエネルギー需要科学の研究者にとってバイナリ形式を扱うのが困難な原因の1つである。この様に日常的に使用されているデータ形式が異なるため異分野間のデータのやり取りがスムーズに行われることが少ない。そこで本研究では、地球科学分野のデータをエネルギー需要科学分野の研究者が使用しやすくするための研究を行った。本研究は、静止気象衛星ひまわり8号による高頻度、高解像度の観測データをもとに日射量を高速に推定するシステム (Takenaka et al., 2011) から推定されたデータをエネルギーマネージメントに活用するためのデータインタフェースの開発を行う。Takenaka et al.(2011)のシステムでは、日射推定量の他に気象業務支援センターが配信している全球数値予報モデルからの出力データ、GPV/GSMをもとに算出された気温、風速、風向などのデータも同じデータフォーマットで出力される。このシステムでは、アジア+オセアニア領域と日本領域の2種類の領域を推定する。アジア+オセアニア領域のデータは、東経85度から西経155度、北緯60度から南緯60度を0.04度間隔の3000×3000のメッシュデータが1次元配列に数値が格納されたデータがバイナリ形式で10分間隔で生成される。日本領域のデータは、東経120度から東経150度、北緯47.6度から北緯22.4度を0.01度間隔の2521×3001のメッシュデータが1次元配列に数値が格納されたデータがバイナリ形式で2.5分間隔で生成される。これらのデータは、FTPサーバを用いてバイナリ形式で提供されている。利用者は、FTPサーバからデータをダウンロードして使用する地点または領域のデータを抽出して解析に利用する。ある地点のデータの抽出や時系列データを生成するためのFORTRANやC言語、Java、MATLABのスク립トが公開されている。また、バイナリ形式をCSV形式に変換するツールも公開されている。

このようにツールも用意されているが多くの利用者にとっては、利用する難易度が高い。その原因は、プログラミング技術が必要なこととデータサイズが大きいことである。多くの利用者は、表計算ソフトを使用する。しかし一般的な表計算ソフトでは、バイナリ形式のファイルを扱うことができない。そのためバイナリ形式をCSV形式のようなテキストファイル形式に変換する必要がある。バイナリ形式からCSV形式に変換するツールは用意されているが、時系列データを作成するには、



図 4.1: 気象データ抽出 Web システム

複数ファイルを CSV 形式に変換する必要があり手間と PC への負荷がかかる。バイナリ形式だと 1 ファイルのサイズが小さいが CSV 形式に変換するとファイルサイズが増加するため PC の容量を圧迫することや同時に複数ファイルを開くとメモリへの負荷もかかってしまう。しかし、用意されているスクリプトを使用して時系列データを作成することでこの問題は解決するが、そのためにはプログラミングの知識と技術が必要となる。プログラミングができない利用者にとっては、プログラミングができる人のサポートが必須となる。

本研究では、プログラミング技術がなくても利用できるデータの提供方法の提案と気象データの可視化システムの開発を行った。

4.2 数値データ提供方法の提案と可視化

多くの利用者は、任意の地点の時系列データを解析に使用している。このデータは、地点の緯度経度情報と期間の情報が決まると作成することができる。そこで緯度経度情報と期間を入力すると時系列データを CSV 形式でダウンロードできる Web システムを作成した。この Web システムによりプログラミングせずに任意の地点の時系列データを CSV 形式で取得できるようになった。また、期間の指定も月単位で指定する方法と日単位で指定する方法の 2 パターンを作成した。半年や年単位でデータを抽出するときは、月単位で指定する方 (図 4.1) が適しているが、1 週間や 1 日のデータを取得するときには日単位の指定する方 (図 4.2) が適している。

第4章 気象データの活用

160.7.27.139 x +
Google で検索するか、URL を入力してください

気象データダウンロード

[月を指定して取得](#) [日にちを指定して取得](#) [月を指定して取得 \(家数複数選択\)](#) [日にちを指定して取得 \(家数複数選択\)](#) [ある市町村のデータを取得](#)

緯度 :

経度 :

取得可能期間 : 2016年04月01日 ~ 2016年07月31日

年 月 日 ~ 年 月 日

変数名 :

領域 :

図 4.2: 気象データ抽出を日単位の指定による Web システム

このシステムでは、位置情報として緯度経度を用いている。一般に太陽光パネルが設置されている住宅などの住所は把握しているが緯度経度は把握していない。また、住宅の屋根以外に設置された太陽光パネルの場合、正確な住所を把握していない可能性もある。そこで、Web-GIS を使用して地図をクリックするとその地点の緯度経度が自動で入力される機能も追加した (図 4.3)。これにより正確な位置情報を入力することができるようになった。



緯度 :

経度 :

取得可能期間 : 2015年11月01日 ~ 2016年01月16日

年 月 日 ~ 年 月 日

変数名 :

領域 :

図 4.3: Web-GIS を利用した気象データ抽出 Web システム

また、日本領域のデータの場合、2.5 分間隔のデータが生成されるが、分析するときの時間スケールとしては細かすぎるといった意見が利用者からあったため 1 時間単位の積算値を算出し提供することも行なった。この他にもある市区町村の平均値や標準偏差などの基本統計量のデータの需要があったため基本統計量の時系列

第4章 気象データの活用

データを作成して提供することも行なった(図4.4)。これは、地理情報データを使用することで市区町村内のデータ点を抽出することができる(図4.5)。図4.6がある日の神奈川県横浜市の日射量の基本統計量の推移である。最大値は、概ね綺麗な曲線であるが、最大値と最小値に差があることがわかる。よって、データの使用目的によって、統計量を適切に選択する必要がある。また、基本統計量を図4.6のように可視化することで地域内のデータに散つきなどがあることを利用者が把握することができる。

気象データダウンロード

[月を指定して取得](#) [日にちを指定して取得](#) [月を指定して取得 \(変数複数選択\)](#) [日にちを指定して取得 \(変数複数選択\)](#) [ある市町村のデータを取得](#)

都道府県 :

市町村 :
例 : 横浜市

取得可能期間 : 2015年11月01日 ~ 2016年01月25日
(2015年) (12月) (1日) ~ (2016年) (1月) (9日)

変数名 :

領域 :

図 4.4: 市区町村の基本統計量を作成する Web システム

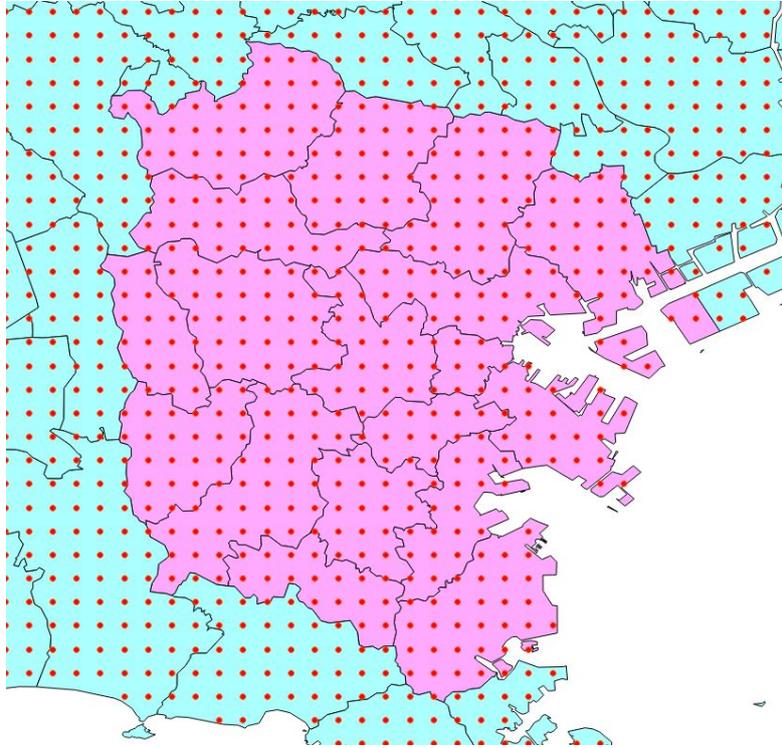


図 4.5: 神奈川県横浜市内のデータ点

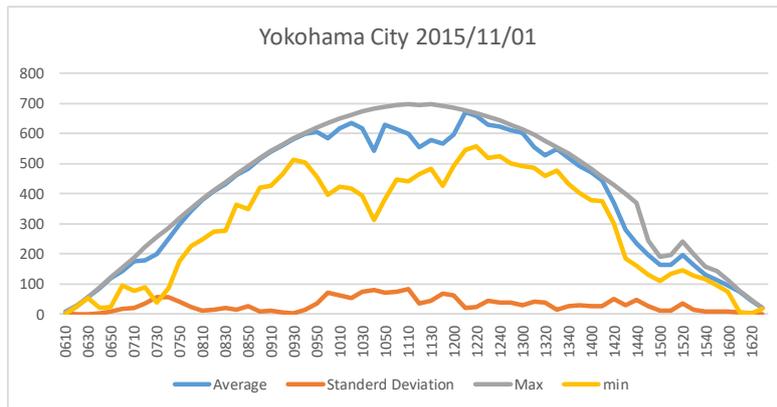
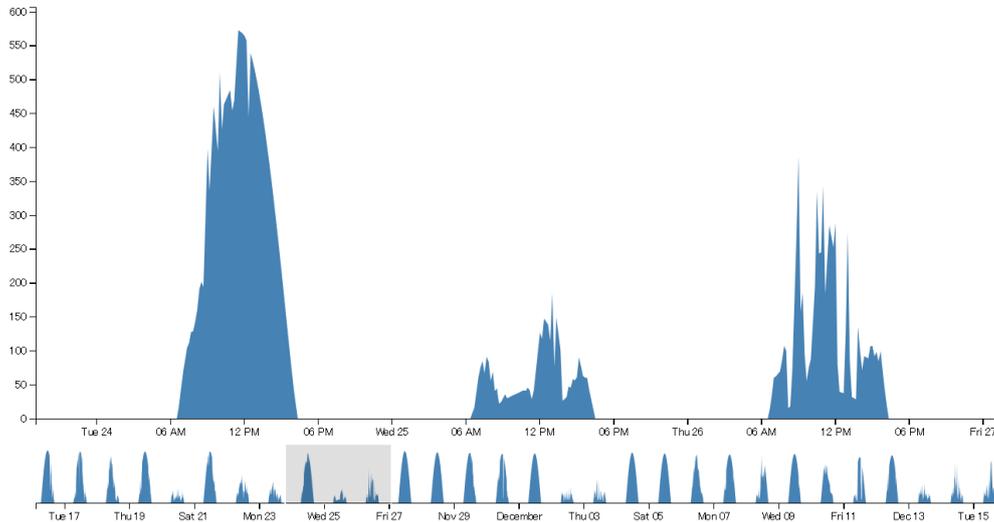


図 4.6: 神奈川県横浜市内の日射量の基本統計量の推移

全データの推移



[1時間ごとの推移を表示](#)

図 4.7: D3.js を用いた気象データの可視化

この Web システムは、抽出した時系列データを可視化する機能も追加した。この機能によりデータのダウンロードとグラフ化をせずに、本システム内で時系列データを可視化して探索的データ解析ができるようになる。この可視化は、D3.js(Bostock et al., 2011) を用いて実装した。長期間のデータを時系列データを折れ線グラフや面グラフで表現するとある 1 日の推移が圧縮されてしまう。ある 1 日の推移のデータを抽出し直すことで見易いグラフにすることが可能だが、毎回抽出することとなり手間がかかる。そこで、推移のグラフのある期間を選択するとその期間の推移が拡大されるようにした(図 4.7)。これにより 1 度のデータ抽出で探索的な可視化をすることができた。

4.3 任意の領域のデータ活用

広範囲の気象データを活用する方法として、全領域を使用する以外に任意の領域のデータを抽出して使用方法がある。本研究では、任意の領域のデータを抽出して活用する事例として、任意の領域のデータを可視化した画像ファイルを

作成して提供をする。本研究は、オーストラリアで開催されたソーラーカーレース、WORLD SOLAR CHALLENGE 2015 及び WORLD SOLAR CHALLENGE 2017 に参戦した東海大学チームへ気象情報を提供する支援の際に行った研究である。ソーラーカーレースにおける気象データの活用は、木村 他 (2011) において日射強度のデータを可視化してソーラーカーレースの戦略に活用している。本支援では、衛星通信を利用して気象情報を渡すため、ファイルサイズに制限がある。このことも考慮した可視化及びデータ提供を行う。本可視化では、オーストラリア全体とコース周辺のデータをヒートマップを用いた可視化（図4.8, 図4.9）とコース上の気温と全天日射推定量の折れ線グラフ（図4.10）の作成を行った。オーストラリア全体とコース周辺のデータの可視化では、4点の緯度と経度を定めることで、その4点からできる四角形内のデータを抽出することができる。このデータを地図上に可視化する。また、風向データは、風速のヒートマップ上に矢印を用いて風向を可視化する。

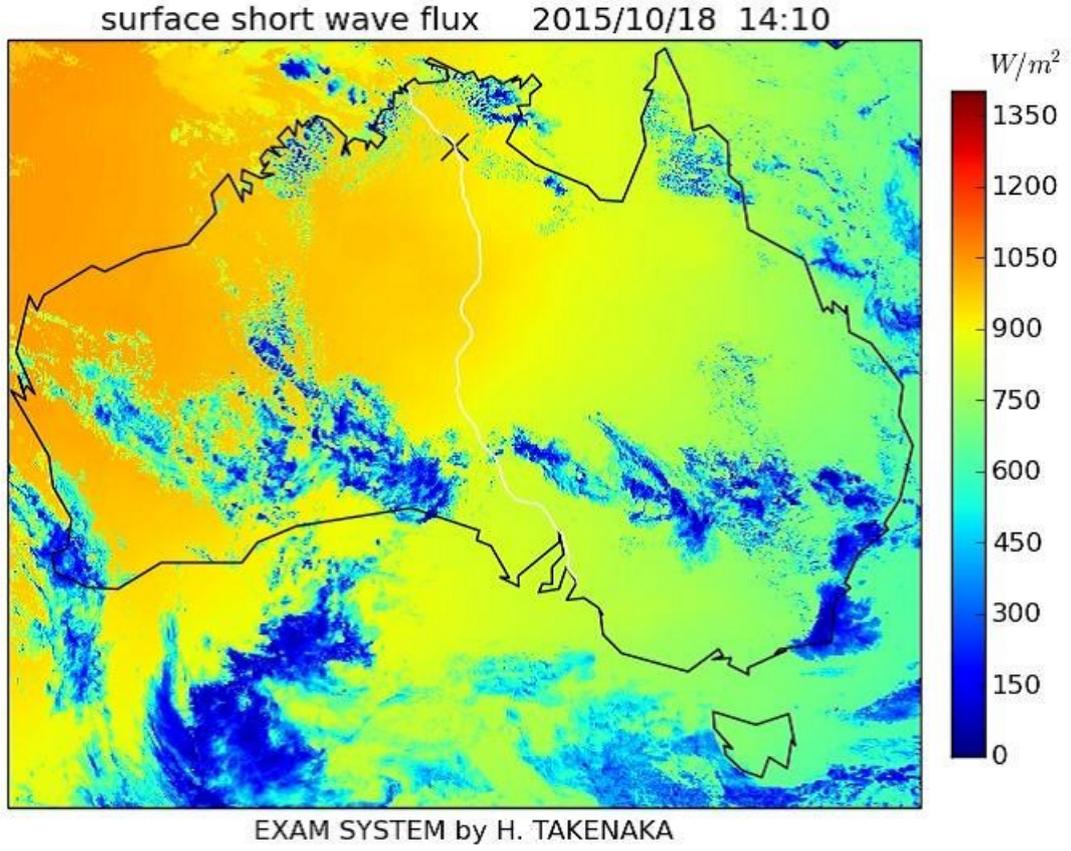


図 4.8: オーストラリア全体の日射量の推定値の画像

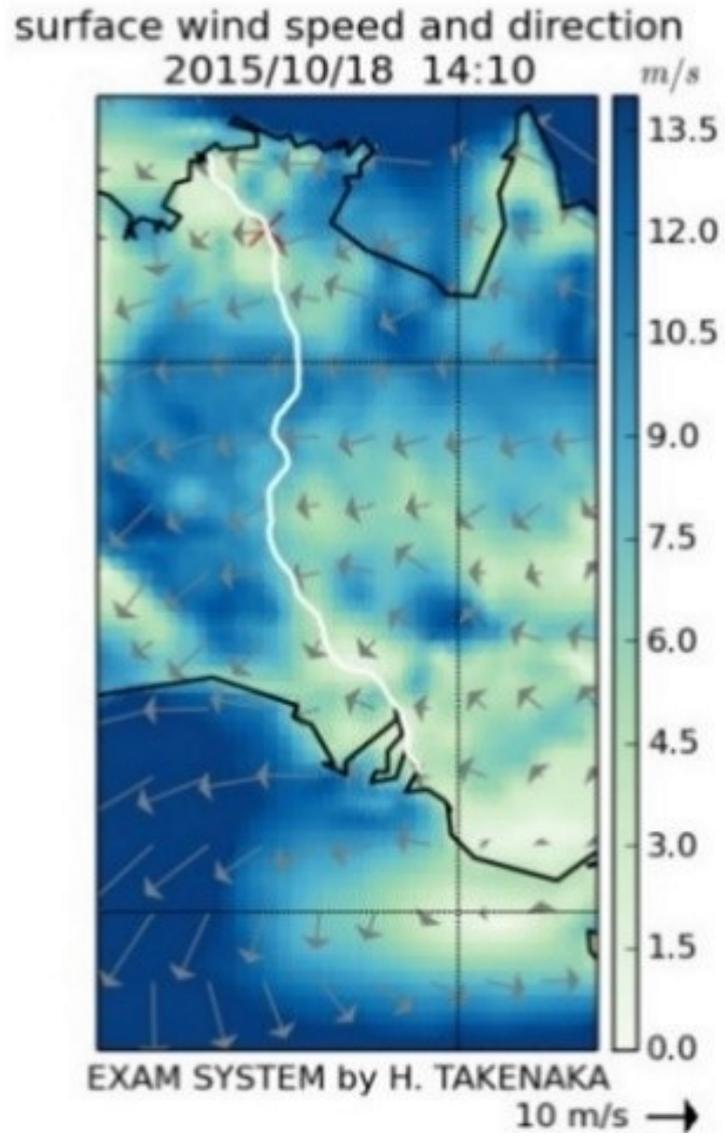


図 4.9: コース周辺の風向・風速の画像

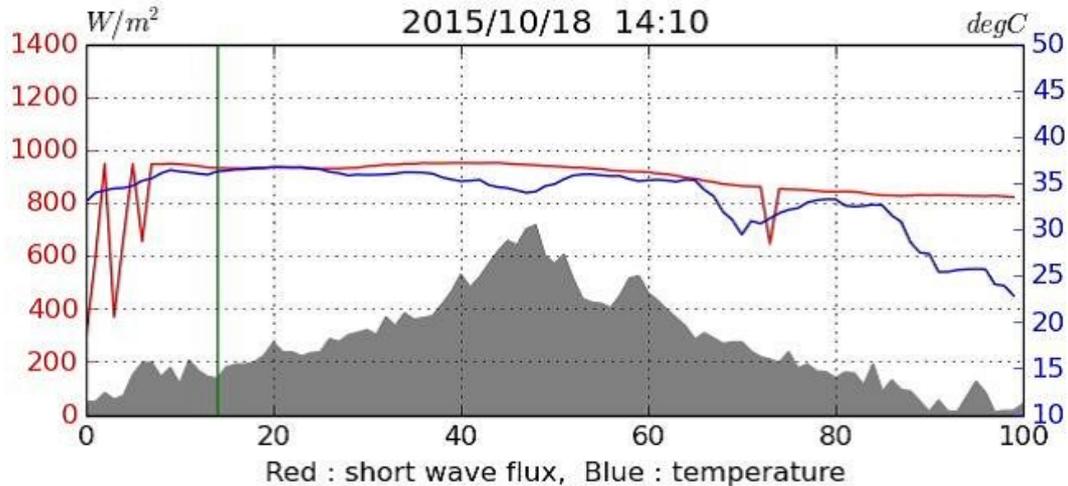


図 4.10: 日射量の推定値と気温のグラフ 横軸はスタートからのゴールまでの経路を 100 等分した点である. 実線はそれぞれ日射量 (赤色), 気温 (青色), 車両の現在位置 (緑) を示す. 影は標高を表す.

4.3.1 予測データの提供

全天日射量の予測データの提供も行った. 予測データは, 毎時 1 回算出され 10 分間隔で 6 時間先まで予測される. 全時間帯の予測データを提供すると処理時間がかかるため, 30 分後, 1 時間後, 1 時間 30 分後, 2 時間後, 3 時間後, 4 時間後, 5 時間後の 7 つの時間帯のみを提供した. また, レース 3 日目からは, 1 時間 30 分後までの 10 分間隔のデータも提供した. 日射量の予測データは, 各時間の予測値をヒートマップの形で可視化した. 可視化する地理的範囲を固定し, 時系列順に見たときに雲の動きを把握できるようにした. 本支援では, オーストラリア全体 (図 4.8), コース周辺 (図 4.9), その日のスタート地点から 1000km 先まで (図??) の 3 種類の範囲の可視化画像を作成した. それぞれの画像にコースと GPS からのソーラーカーの位置座標を追加した. その日のスタート地点から 1000km 先までの画像には, その日のスタート地点から 100km 毎に印を付けて, その日の走行距離や 1 時間後のソーラーカーの位置などを把握しやすくした. また, その画像のデータが予測された日時を判断できるように図のタイトルに予測をした日時と予測された日時を追加した. タイトルの最初の日時が予測をした日時, 最後の日時が予測された日時とした. つまり図??は, 10 時 30 分の 1 時間後 (11 時 30 分) の予測データである.

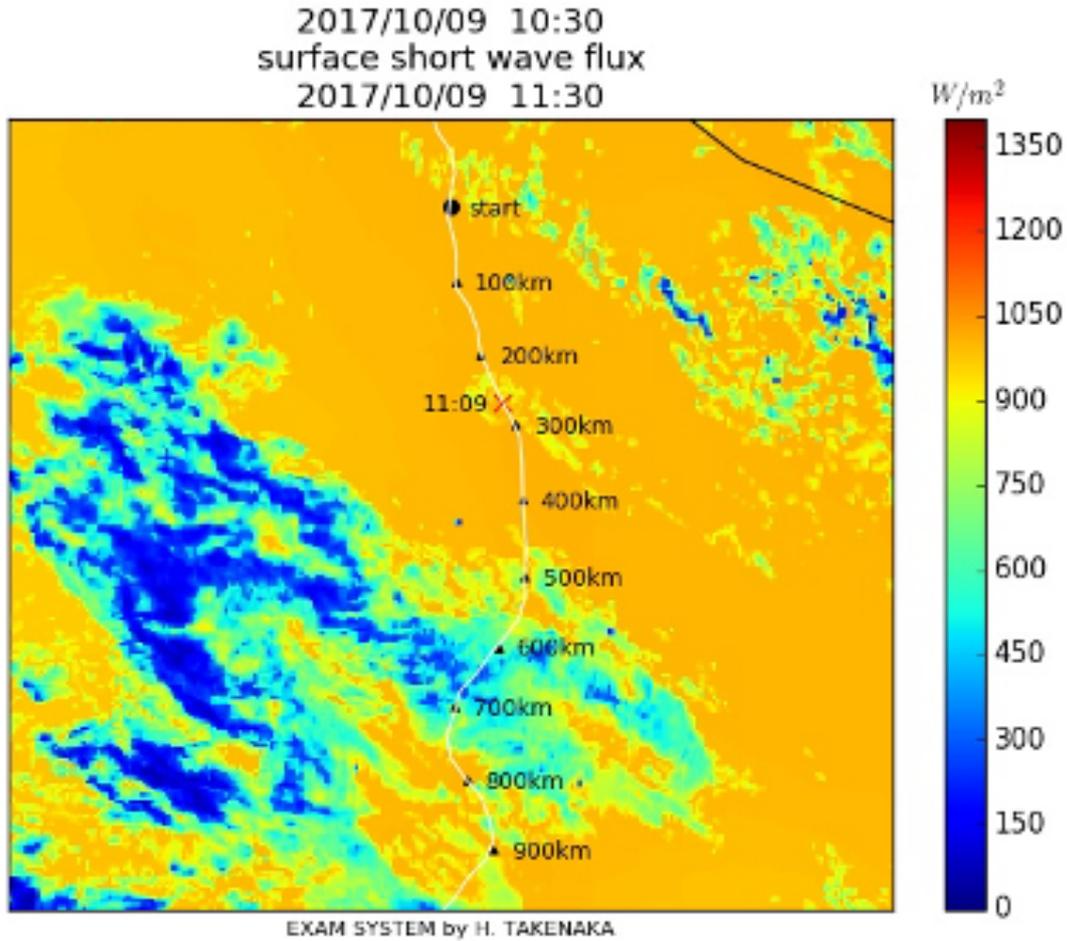


図 4.11: その日のスタート地点から 1000km 先までの画像

4.3.2 システムの概要

本システムの処理工程の概要が図 4.12 である。本システムは、気象データをダウンロードし、気象データを可視化した画像データを作成し、ダウンロードする Web ページを更新するシステムと GPS サービスからソーラーカーの位置座標の取得するシステムからなる。ソーラーカーチームは、主にインマルサット車載用衛星電話を使用して通信するため、画像サイズを考慮する必要がある。本システムは、LinuxOS で構築されたサーバ上で、Python 言語で実装した。準リアルタイムデータと予測データが生成される時間及び公開されるサーバも異なるためデータのダ

ダウンロードと可視化のプログラムをデータごとに分けた。そして、それぞれのデータが生成される時間に合わせて cron を用いて各プログラムを定時実行した。また、ソーラーカーチームが画像データのダウンロードと画像を切り替える作業を省くことができるように画像を時系列順に動画として表示する Web ページも作成した。

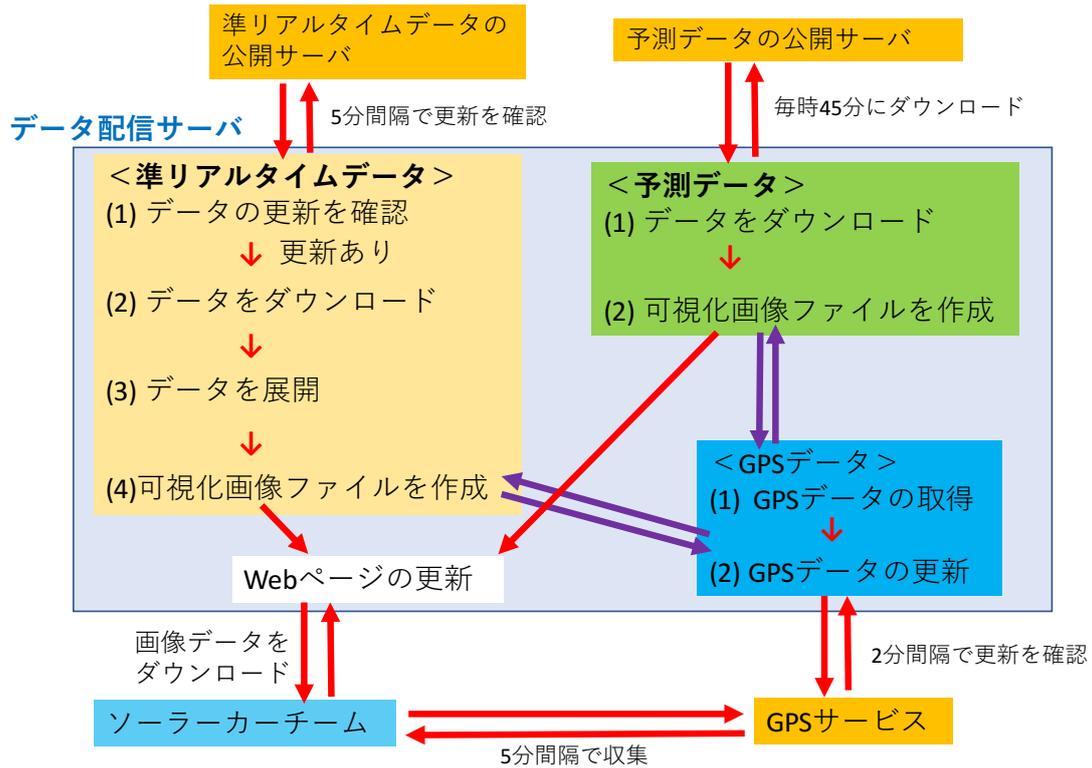


図 4.12: システムの処理工程

第5章 まとめと展望

本論文では、大規模データや分析結果の理解に可視化が有用であることを示した。学術論文データベースの活用では、テキストデータからなる各論文のアブストラクトとタイトルに対して、トピックモデルを適用することで各論文の研究領域(トピック)を推定を行った。さらにトピックモデルの結果を自己組織化マップを用いた経時的な可視化などいくつかの可視化を提案した。これらの可視化を用いることで、学内の研究状況の把握をすることが可能であり、研究支援に役立てることができることを示した。また、提案した可視化は、潜在的な話題(トピック)を持つテキストデータに適用することが可能なため、幅広い分野での活用が期待できる。

災害時のTwitterデータの活用では、Twitterの情報を地図上に反映して、減災支援システムのプロトタイプを作成した。地図上にマッピングする際の位置情報であるジオタグが付いてないツイートに対して、ツイート内の地名などの位置情報からマッピングする位置を決める方法を提案した。また、Webサービスを活用することで情報収集ができることも示した。

災害時のツイートデータ分析では、大量で複雑なツイートデータに対して、収集する情報のキーワードを設定し、そのキーワードを含むツイートデータのみにすることでデータ量を少なくするとともに効率的な情報の抽出を行うための前処理を行った。そして、対応分析を用いて関係性が強い単語の組み合わせを抽出することでツイートデータから情報を採取することができることを示した。さらに、対応分析の可視化とその可視化から関係性が強いと思われる単語が含まれるツイートを表示するアプリケーションを作成した。また、対応分析による可視化を経時的に比較することを容易にするための布置の最適化法を提案した。提案した布置の最適化法は、対応分析以外にも主成分分析や多次元尺度構成法などの次元圧縮法のプロットにも適用することが可能である。

気象データの活用では、プログラミング技術がないと利用することが難しいバイナリ形式のデータを使い易くて、理解し易いデータの提供方法の開発を行なった。また、任意の領域のデータの活用として、オリジナルデータから任意の領域のデータを抽出して可視化した画像をソーラーカーレースに参戦したチームに提供するという行なった。

今後の展望として、第2章で提案したトピックモデルの結果を自己組織化マップを用いて可視化する手法は、ツイートデータにも適用が可能である。そのため、対応分析では抽出することができなかった情報(トピック)を抽出することができるかもしれない。このように本論文で提案した可視化手法は、経時的な属性を持つテキストデータに適用が可能であるため、色々な分野のデータに適用して問題解決に役立てることができることを示して行きたい。本論文では、医学部を持つ大規模な総合大学をケーススタディとして学内の研究状況の把握を行なったが、小規模大学や単科大学など様々な特色を持つ大学の研究状況を提案手法で分析及び比較を行い、提案手法の潜在的な特徴なども明らかにして行きたい。また、大規模データの解析では、コンピュータの性能への依存度が大きくなる。本論文の研究でもコンピュータの性能の影響で計算ができないという問題が度々あった。今後は、コンピュータの性能などの解析環境も考慮した解析手法の開発についても研究を行いたい。

謝辞

本博士論文を作成するに当たり、多くの方々からのご支援とご指導を賜り心より御礼申し上げます。特に指導教員の山本義郎教授には、終始熱心にご指導頂き深く感謝しております。山本義郎教授の下で様々な研究に参加させて頂いたことで、多くの知識や技術の修得だけでなく研究者の心得も学ぶことができました。学位申請の主査を務めていただいた情報理工学部 内田理教授及び、審査員を務めていただいた工学部 梶田佳孝教授、理学部 北林照幸教授、情報通信学部 宇津圭祐准教授には、論文をより良くするための有益な助言を下さり深く感謝致します。また、共同研究者である、福岡女子大学 藤野友和准教授、教養学部 富田誠准教授、現代教養センター 田島祥准教授、情報理工学部 中島孝教授をはじめとする JST/CREST/EMS TEEDDA チームのメンバー、工学部 木村英樹教授、工学部 福田紘大准教授には、多大なるご指導ご鞭撻を賜り、厚く御礼申し上げます。日頃より研究に対する刺激を与えて頂いたと共に激励を下さった、同級生の山田実俊氏、修士課程時の同級生の八木圭太氏、宗像昌平氏、後輩の亀岡瑠氏、研究室のメンバーにも感謝致します。更に、博士論文を作成するに当たり、ご配慮頂いた職場の東北大学東北メディカル・メガバンク機構 田宮元教授をはじめとする田宮研究室の皆様にも心より感謝致します。最後に、博士課程まで進学させて下さった両親に感謝の意を表します。

また、本博士論文は、統計数理研究所共同利用研究重点テーマ4「学術文献データ分析の新たな統計科学的アプローチ」における「学術文献データベースを用いた環境系および文理統合系研究組織における研究成果の評価に関する基礎研究」(28-共研-4401)、統計数理研究所共同利用研究重点テーマ2「IRのための学術文献データ分析と統計的モデル研究の深化」における「学術文献DBにおける著者識別問題と研究組織学術文献DBにおける著者識別の精度向上に関する研究」(30-共研-4202)、JST/CREST/JPMJCR15K4、JSPS 科研費 JP15K11997 の助成を受けたものであり、Web of Science のデータベースは、統計数理研究所共同利用研究重点テーマの下で利用許可を受けている。

参考文献

- [1] 石川哲也, 川崎昭如, 目黒公郎 (2012). 山陰地方豪雪災害時の Twitter ユーザによる情報発信行動に関する分析と考察, 地域安全学会論文集, **17**, 135–143.
- [2] 岩田具治 (2015). トピックモデル, 講談社.
- [3] 岩坪秀一 (1987). 数量化の基礎. 朝倉書店.
- [4] 内田健康 (2014). エネルギーマネージメントシステム: 環境, 社会・経済システム, 物理システムのインテグレーション, 計測と制御, **53**(1), 2–5.
- [5] 梅津雄一, 中野貴広 (2018). R と Shiny で作る Web アプリケーション, C&R 研究所.
- [6] 大隅昇, Lebart, L., Morineau, A., Warwick, K. M., 馬場康維 (1994). 記述的多変量解析法, 日科技連.
- [7] 奥野陽, グラム ニュービグ, 荻原正人 (2016). 自然言語処理の基本と技術, 翔泳社.
- [8] 桂井麻里衣, 大向一輝, 武田英明 (2015). 大規模学術論文データベースにおける研究者のトピック推定と著者同定への応用, 第7回データ工学と情報マネジメントに関するフォーラム (DEIM2015), A5-2.
- [9] 金明哲, 村上征勝, 永田昌明, 大津起夫, 山西健司 (2003). 言語と心理の統計, 岩波書店.
- [10] 金明哲 (2018). テキストアナリティクス, 共立出版.
- [11] 佐藤一誠 (2015). トピックモデルによる統計的潜在意味解析, コロナ社.
- [12] 工藤拓 (2018). 形態素解析の理論と実装, 近代科学社.
- [13] 小林雅之, 山田礼子, 浅野茂, 森利枝, 劉文君 (2016). 大学の IR 意思決定支援のための情報収集と分析, 慶應義塾大学出版会.

- [14] 小林雄一郎 (2017). Rによるやさしいテキストマイニング [機械学習編], オーム社.
- [15] 武井美緒, 藤野友和, 中野純司, Frederick Kin Hing Phoa(2019). トピックモデルを用いた研究動向の分析, 日本計算機統計学会第33回大会, 77-80.
- [16] 統計数理研究所 (2017). 共同利用実施報告書 平成 28(2016) 年度, https://www.ism.ac.jp/kyodo/index_i/H28_Report_web.pdf.
- [17] 統計数理研究所 (2018). 共同利用実施報告書 平成 29(2017) 年度, https://www.ism.ac.jp/kyodo/index_i/H29_Report_web.pdf.
- [18] 統計数理研究所 (2019). 共同利用実施報告書 平成 30(2018) 年度, https://www.ism.ac.jp/kyodo/index_i/H30_Report_web.pdf.
- [19] 徳高平蔵, 岸田悟, 藤村喜久郎 (1999). 自己組織化マップの応用—多次元情報の2次元可視化, 海文堂出版.
- [20] 中島孝 (2020). 戦略的創造研究推進事業 CRST 研究領域「分散協調型エネルギー管理システム構築のための理論及び基盤技術の創出と融合展開」研究課題「分散協調型 EMS における地球科学情報の可用性向上とエネルギー需要モデルの開発」研究終了報告書, https://projectdb.jst.go.jp/file/JST-PROJECT-15653379/JST_1111069_15653379_2019_中島_PER.pdf.
- [21] 西里静彦 (1982). 質的データの数量化 - 双対尺度法とその応用. 朝倉書店.
- [22] 橋本康弘, 岡瑞起 (2012). 都市におけるジオタグ付きツイートの統計 (特集:人と環境に見る高次元データフローの生成と解析), 人工知能学会誌, **27**(4), pp. 424-431.
- [23] 林知己夫 (1956). 数量化理論とその応用例 (II). 統計数理研究所彙報, **4**(2), 19-30.
- [24] 原隆浩 (2017). ビッグデータ解析の現在と未来 -Hadoop, NoSQL, 深層学習からオープンデータまで-. 共立出版.
- [25] 樋口耕一 (2020). 社会調査のための計量テキスト分析 —内容分析の継承と発展を目指して— 第2版, ナカニシヤ出版.

- [26] 藤田政之 (2020). CREST「分散協調型エネルギー管理システム構築のための理論及び基盤技術の創出と融合展開」研究領域事後評価報告書, https://projectdb.jst.go.jp/file/JST-AREA-1111069/JST_1111069_2019_PAEE.pdf.
- [27] 藤野友和, 山本由和, 船山貴光, 山本 義郎 (2016). 学術文献DBにおける著者識別問題について, 日本計算機統計学会 第30回シンポジウム講演論文集, 45–48.
- [28] 船山貴光, 今田一希, 竹中栄晶, 木村英樹, 福田紘大, 山本義郎, 中島孝 (2018a). WSC2017 支援のための気象情報可視化・配信システムの構築, 情報処理学会 第80回全国大会.
- [29] 船山貴光, 山本義郎, 富田誠, 梶田佳孝, 田島祥, 宇津圭祐, 内田理 (2015). WebサービスとSNS情報を活用した減災支援システムの開発, 科研費シンポジウム「空間データと災害の統計モデル」.
- [30] 船山貴光, 山本義郎, 藤野友和 (2018b). トピックモデルを用いた学内の研究分野の分類, 日本計算機統計学会第32回大会.
- [31] 船山貴光, 山本義郎, 藤野友和 (2019). トピックモデルを用いた大学の研究の強み領域の分析と可視化, 日本計算機統計学会第33回大会.
- [32] 船山貴光, 山本義郎, 藤野友和 (2020). 大規模大学における研究分野の研究実績の可視化, 統計数理, **68**(2), pp.233–246.
- [33] 船山貴光, 山本義郎, 内田理 (2017a). 熊本地震の Tweet データの解析, 科研費シンポジウム「空間データと災害の統計モデル」.
- [34] 船山貴光, 山本義郎, 内田理 (2017b). 熊本地震後の Tweet データからの有益な情報の抽出と分類, 日本分類学会第35回大会.
- [35] 船山貴光, 山本義郎, 内田理 (2017c). 災害時の情報収集のための Twitter データの解析, 第16回西東京統計研究会.
- [36] 船山貴光, 山本義郎, 内田理 (2018a). 熊本地震の Tweet データの可視化, 科研費シンポジウム「空間データと災害の統計モデル」.
- [37] 船山貴光, 山本義郎, 内田理 (2018b). 災害時の Tweet 情報の経時的な変化を把握するための可視化, 日本分類学会第37回大会.

- [38] 船山貴光, 山本義郎, 内田理 (2018c). 災害時の Tweet 情報の変化に関する分析, 第 17 回西東京統計研究会.
- [39] 船山貴光, 渡邊武志, 竹中栄晶, 木村英樹, 福田紘大, 山本義郎, 中島孝 (2016). ソーラーカーレース支援のための気象情報可視化・配信システムの構築, 情報処理学会第 78 回全国大会.
- [40] 水田正弘 (2016). ビッグデータに対する統計学の役割. 日本情報経営学会誌, **36**(4), 12–17.
- [41] 山本義郎, 船山貴光, 今田一希, 中島孝, 木村英樹, 福田紘大, 竹中栄晶 (2018). ソーラーカーレース支援のための気象情報可視化・配信システムの改良, 2017 年度東海大学情報技術センター研究・開発報告会.
- [42] 山本義郎, 船山貴光, 渡邊武志, 中島孝, 木村英樹, 福田紘大, 竹中栄晶 (2016). ソーラーカーレース支援のための気象情報可視化・配信システムの構築, 2015 年度東海大学情報技術センター研究・開発報告会.
- [43] 渡邊英徳 (2014). 台風リアルタイム・ウォッチャー, <https://typhoon.mapping.jp/index.html>
- [44] Almende B.V., Thieurmel, B. and Robert, T. (2019). visNetwork: Network Visualization using 'vis.js' Library, R package version 2.0.9., <https://CRAN.R-project.org/package=visNetwork>
- [45] Asuncion, A., Welling, M., Smyth, P., and Teh, W. Y. (2009). On smoothing and inference for topic models. Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, 27–34.
- [46] Benzecri, J. P. et al.,(1973a). L' Analyse des Donnees. Tome 1, La Taxinomie, Dunod.
- [47] Benzecri, J. P. et al.,(1973b). L' Analyse des Donnees. Tome 2, L' Analyse des Crrispondances, , Dunod.
- [48] Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). Latent Dirichlet Allocation, *Journal of Machine Learning Research*, **3**, 993–1022.
- [49] Bostock, M., Ogievetsky, V. and Heer, J. (2011). D3: Data-Driven Documents, *IEEE Transactions on Visualization and Computer Graphics*, **17**(12), 2301 – 2309.

- [50] Chang, W., Cheng, J., Allaire, J., Xie, Y. and McPherson, J. (2019). shiny: Web Application Framework for R, R package version 1.3.2, <https://CRAN.R-project.org/package=shiny>
- [51] Clausen, S. (1998). *APPLIED CORRESPONDENCE ANALYSIS: An Introduction*, SAGE(藤本一男 訳・解説 (2015). 『対応分析入門 原理から応用まで 解説◆Rで検算しながら理解する』, オーム社).
- [52] Fisher, R. A. (1936). THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS, *Annals of human genetics*, **7**(2), 179–188.
- [53] Funayama, T., Yamamoto, Y., Tomita, M., Kajita, Y., Tajima, S., Utsu, K. and Uchida, O.(2015). Disaster Mitigation Support System using Web Services and SNS Information, The 13th International Conference on ICT & Knowledge Engineering, pp.42-45, DOI:10.1109/ICTKE.2015.7368469.
- [54] Funayama, T., Yamamoto, Y., Tomita, M., Uchida, O. and Kajita, Y.(2014). Disaster mitigation support system using Twitter and GIS, The 12th International Conference on ICT & Knowledge Engineering, pp.18-23, DOI:10.1109/ICTKE.2014.7001528.
- [55] Funayama, T., Yamamoto, Y. and Uchida, O.(2017a). Extraction of the Information from Tweet Data in Case of the Disaster using Visualization, Proceedings of the 2017 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA' 17), pp.226-231.
- [56] Funayama, T., Yamamoto, Y. and Uchida, O.(2017b). Information extraction and its visualization using the quantification method from Tweet information for a disaster, The 2017 conference of the International Federation of Classification Societies (IFCS-2017).
- [57] Funayama, T., Yamamoto, Y. and Uchida, O.(2017c). Development of Visualization Application of Tweet Data for Extracting Information in case of Disaster, The 15th International Conference on ICT & Knowledge Engineering, pp.45-49, DOI:10.1109/ICTKE.2017.8259620.
- [58] Funayama, T., Watanabe, T., Takenaka, H., Kimura, H., Fukuda, K., Nakajima, T. Y. and Yamamoto, Y.(2016). Improvement of the methods of providing meteorological data for energy management systems, The 22nd International Conference on Computational Statistics (COMPSTAT 2016).

- [59] Gatti, C. J. and Brooks, J. D. and Nurre, S. G. (2015). A Historical Analysis of the Field of OR/MS using Topic Models, arXiv:1510.05154.
- [60] Griffiths, T. L. and Steyvers, M. (2004). Finding Scientific Topics. *PNAS*, **101**(1), 5228–5235.
- [61] Grün, B. and Hornik, K. (2011). topicmodels: An R Package for Fitting Topic Models, *Journal of Statistical Software*, **40**(13), 1–30.
- [62] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, **4**(1), 59–69.
- [63] Kohonen, T. (2000). *Self-Organizing Maps*, 3rd ed., Springer(徳高平蔵, 堀尾恵一, 大北正昭, 大藪又茂, 藤村喜久郎 訳 (2005). 『自己組織化マップ』, シュプリンガー・フェアラーク東京).
- [64] Mojena, R. (1977). Hierarchical grouping methods and stopping rules:an evaluation, *The Computer Journal*, **20**, 359–363.
- [65] Murray, S. (2013). *Interactive Data Visualization for the Web*, O’Reilly(長尾高弘 訳 (2014). 『インタラクティブ・データビジュアライゼーション —D3.js によるデータの可視化』, オライリー・ジャパン).
- [66] Nakajima, T. Y., Watanabe, T., Funayama, T., Yamamoto, Y., Takenaka, H., Nakajima, T., Irie, H. and Higuchi, A.(2015). Satellite remote sensing and energy management, The 23rd CERES International Symposium.
- [67] Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees, *Proceedings of International Conference on New Methods in Language Processing*.
- [68] Schmid, H. (1995). Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*.
- [69] Sievert, C. (2018). plotly for R. <https://plotly-r.com>.
- [70] Sun, L. and Yafeng Yin, Y. (2017). Discovering themes and trends in transportation research using topic modeling, *Transportation Research Part C: Emerging Technologies*, **77**, 49-66.

- [71] Takenaka, H., Nakajima, T. Y., Higurashi, A., Higuchi, A., Takamura, T., Pinker, R. T. and Nakajima, T. (2011). Estimation of solar radiation using a neural network based on radiative transfer, *J. Geophys. Res.*, **116**(D08215), 1–26, doi:10.1029/2009JD013337.
- [72] Teh, W. Y. Newman, D. and Welling, M. (2007). A Collapsed Variational Bayesian Inference Algorithm for Latent Dirichlet Allocation. In *NIPS*, **19**, 1353–1360.
- [73] Tian, J., Azarian, M. H. and Pecht, M. (2014). Anomaly Detection Using Self-Organizing Maps-Based K-Nearest Neighbor Algorithm, *Second European Conference of the Prognostics and Health Management Society 2014*.
- [74] Uchida, O., Rokuse, T., Tomita, M., Kajita, Y., Yamamoto, Y., Toriumi, F., Seaan, B. and Robertson, S. (2015a). *Classification and Mapping of Disaster Relevant Tweets for Providing Useful Information for Victims*, *IEEEJ Transactions on Image Electronics and Visual Computing*.
- [75] Uchida, O., Kosugi, M., Endo, G., Funayama, T., Utsu, K., Tajima, S., Tomita, M., Kajita, Y., and Yamamoto, Y. (2015b). *A Real-Time Disaster-Related Information Sharing System Based on the Utilization of Twitter*, *The Fifth International Conference on Social Media Technologies, Communication, and Informatics (2015)*.
- [76] Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth Edition. Springer, New York.
- [77] Yamada, S., Utsu, K. and Uchida, O. (2019). An Analysis of Tweets During the 2018 Osaka North Earthquake in Japan -A Brief Report, 2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), DOI: 10.1109/ICT-DM.2018.8636393.
- [78] Yamamoto, Y., Funayama, T., Konda, K., Takenaka, H., Kimura, H., Fukuda, K. and Nakajima, T. Y. (2017). Data provision and visualization system for EMS using cloud, NZSA/IASC-ARS2017 (New Zealand Statistical Association and the International Association of Statistical Computing (Asian Regional Section) Joint Conference 2017).

- [79] Yamamoto, Y., Funayama, T., Konda, K., Takenaka, H., Murata, K. T. and Nakajima, T. Y.(2019). Provision and Visualization of Solar Radiation Data for Energy Management System, 2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE), pp.1–4, DOI:10.1109/ICTKE47035.2019.8966805.
- [80] Wehrens, R. and Buydens, LMC. (2007). Self- and Super-Organizing Maps in R:The kohonen Package. *Journal of Statistical Software*, **21**(5), 1–19.
- [81] Text Mining Studio, <https://www.msi.co.jp/tmstudio/>
- [82] Yahoo!ジオコーダ API, <http://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/geocoder.html>
- [83] Yahoo! JavaScript マップ API, <http://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/js/>
- [84] Yahoo! Open Local Platform(YOLP), <http://olp.yahoo.co.jp/>
- [85] Yahoo!リバースジオコーダ API, <http://developer.yahoo.co.jp/webapi/map/openlocalplatform/v1/reversegeocoder.html>
- [86] Yahoo! Japan 天気・災害, 提供 RSS, <http://weather.yahoo.co.jp/weather/rss/>

業績一覧

査読あり

1. 船山貴光・山本義郎・藤野友和, 大規模大学における研究分野の研究実績の可視化, 統計数理, 第 68 巻, 第 2 号, pp.233-246, 2020/12, 統計数理研究所.
2. Y. Yamamoto, T. Funayama, K. Konda, H. Takenaka, K. T. Murata, T. Y. Nakajima, Provision and Visualization of Solar Radiation Data for Energy Management System, 2019 17th International Conference on ICT and Knowledge Engineering (ICT&KE), pp.1-4, DOI:10.1109/ICTKE47035.2019.8966805, 2019/11, IEEE.
3. T. Funayama, Y. Yamamoto & O. Uchida, Development of Visualization Application of Tweet Data for Extracting Information in case of Disaster, The 15th International Conference on ICT & Knowledge Engineering, pp.45-49, DOI:10.1109/ICTKE.2017.8259620, 2017/11/23, IEEE.
4. 山田実俊・船山貴光・河野克哉・服部敦・山本義郎, 売り上げ傾向による店舗の分類と購買傾向の分析と可視化, 計算機統計学, 第 29 巻, 第 2 号, pp.159-167, 2017/2, 日本計算機統計学会.
5. 亀岡瑤・船山貴光・宗像昌平・山田実俊・八木圭太・山本義郎, 条件付きアソシエーションルールによる顧客の購入特徴の抽出, 計算機統計学, 第 29 巻, 第 1 号, pp.57-64, 2016/ , 日本計算機統計学会.
6. O. Uchida, M. Kosugi, G. Endo, T. Funayama, K. Utsu, S. Tajima, M. Tomita, Y. Kajita & Y. Yamamoto, A Real-Time Information Sharing System to Support Self-, Mutual-, and Public-Help in the Aftermath of a Disaster Utilizing Twitter, IEICE Transactions on Fundamentals, Vol.E99-A, No.8, pp.1551-1554, 2016/8, 電子情報学会.

7. S. Yamada, T. Funayama & Y. Yamamoto, Visualization of relations of stores by using Association Rule Mining, The 13th International Conference on ICT & Knowledge Engineering, pp.11-14, DOI:10.1109/ICTKE.2015.7368463, 2015/11, IEEE.
8. T. Funayama, Y. Yamamoto, M. Tomita, Y. Kajita, S. Tajima, K. Utsu & O. Uchida, Disaster Mitigation Support System using Web Services and SNS Information, The 13th International Conference on ICT & Knowledge Engineering, pp.42-45, DOI:10.1109/ICTKE.2015.7368469, 2015/11, IEEE.
9. K. Yagi, T. Funayama & Y. Yamamoto, The Digitizing of the Characteristic and Visualization of the Wave of the Condition of Batter of the Professional Baseball, The 12th International Conference on ICT & Knowledge Engineering, pp.43-47, DOI:10.1109/ICTKE.2014.7001533, 2014/11, IEEE.
10. T. Funayama, Y. Yamamoto, M. Tomita, O. Uchida & Y. Kajita, Disaster mitigation support system using Twitter and GIS, The 12th International Conference on ICT & Knowledge Engineering, pp.18-23, DOI:10.1109/ICTKE.2014.7001528, 2014/11, IEEE.
11. K. Yagi, T. Funayama, Y. Yamamoto & Y. Hoshi, Use of stand-alone and web-based systems for DNA analysis - Construction of an analysis environment for distant laboratories performing joint research -, The 11th International Conference on ICT & Knowledge Engineering, pp.43-45, DOI:10.1109/ICTKE.2013.6756273, 2013/11, IEEE.

査読なし

1. 船山貴光・山本義郎, 勝敗を分ける要因の解析とパスの可視化と解析, 統計数理研究所共同研究リポート 380 スポーツデータ解析における理論と事例に関する研究集会 第4巻, pp.147-152, 2017/3, 統計数理研究所.
2. 船山貴光・山田実俊・山本義郎・内山秀一, 選手の動きによるチーム戦略と試合の流れの可視化, 統計数理研究所共同研究リポート 363 スポーツデータ解析における理論と事例に関する研究集会 第3巻, pp.91-94, 2016/3, 統計数理研究所.

3. 河野克哉・亀岡瑤・船山貴光・山田実俊・山本義郎, 審判と投手の特徴の分析と可視化, 統計数理研究所共同研究リポート 363 スポーツデータ解析における理論と事例に関する研究集会 第3巻, pp.41-46, 2016/3, 統計数理研究所.
4. 船山貴光・河野克哉・八木圭太・池上拓人・山本義郎, 打者の好調不調の原因分析と守備能力の分析, 統計数理研究所共同研究リポート 334 スポーツデータ解析における理論と事例に関する研究集会 第2巻, pp.29-36, 2015/3, 統計数理研究所.
5. 船山貴光・八木圭太・宗像昌平・池上拓人・亀岡瑤・宮前貴基・山本義郎, 選手の特徴および調子の波を把握する為の修正 OPS とその活用例, 統計数理研究所共同研究リポート 314 スポーツデータ解析における理論と事例に関する研究集会 第1巻, pp.51-56, 2014/3, 統計数理研究所.
6. 山田実俊・八木圭太・宗像昌平・船山貴光・山本義郎, サッカー試合のゾーン使用率・ゾーン支配率の可視化, 統計数理研究所共同研究リポート 314 スポーツデータ解析における理論と事例に関する研究集会 第1巻, pp.45-50, 2014/3, 統計数理研究所.

国際会議

1. Y. Yamamoto, T. Funayama, K. Konda, H. Takenaka, H. Kimura, K. Fukuda & T. Y. Nakajima, Data provision and visualization system for EMS using cloud, NZSA/IASC-ARS2017 (New Zealand Statistical Association and the International Association of Statistical Computing (Asian Regional Section) Joint Conference 2017), Auckland, New Zealand, 2017/12.
2. T. Funayama, Y. Yamamoto & O. Uchida, Extraction of the Information from Tweet Data in Case of the Disaster using Visualization, Proceedings of the 2017 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA' 17), pp.226-231, Monte Carlo Resort, USA, 2017/07/18
3. T. Funayama, Y. Yamamoto & O. Uchida, Information extraction and its visualization using the quantification method from Tweet information for a disaster, The 2017 conference of the International Federation of Classification Societies (IFCS-2017), Tokai University, Tokyo, Japanese Classification Society, 2017/08/08.

4. T. Funayama, T. Watanabe, H. Takenaka, H. Kimura, K. Fukuda, T. Y. Nakajima & Y. Yamamoto, Improvement of the methods of providing meteorological data for energy management systems, The 22nd International Conference on Computational Statistics (COMPSTAT 2016), Auditorium/Congress Palace Principe Felipe, Oviedo, Spain, 2016/8/25.
5. T. Funayama, S. Uchiyama & Y. Yamamoto, Visualization of tracking data for a soccer game and Discussion of evaluation indices of the defensive ability, The Fifth KMITL-TKU Joint International Symposium on Mathematics and Applied Mathematics (MAM2016), TKU-KMITL, 2016/3/21.
6. T. Y. Nakajima, T. Watanabe, T. Funayama, Y. Yamamoto, H. Takenaka, T. Nakajima, H. Irie & A. Higuchi, Satellite remote sensing and energy management, The 23rd CEReS International Symposium, Chiba University, 2015/12/2.
7. O. Uchida, M. Kosugi, G. Endo, T. Funayama, K. Utsu, S. Tajima, M. Tomita, Y. Kajita & Y. Yamamoto, A Real-Time Disaster-Related Information Sharing System, 5th International Conference on Social Media Technologies, Communication, and Informatics, pp.22-25, IARIA, Spain, 2015.
8. T. Funayama & Y. Yamamoto, Analysis of batter's condition change using moving OPS, The Fourth TKU-KMITL Joint Symposium on Mathematics and Applied Mathematics (MAM2014), TKU-KMITL, 2014/3/19.

国内会議

1. 船山貴光・山本義郎・藤野友和, トピックモデルを用いた大学の研究の強み領域の分析と可視化, 日本計算機統計学会 第33回大会, 日本計算機統計学会, 2019/6/1.
2. 船山貴光・山本義郎・内田理, 災害時の Tweet 情報の経時的な変化を把握するための可視化, 日本分類学会 第37回大会, 日本分類学会, 2018/6/10.
3. 船山貴光・山本義郎・藤野友和, トピックモデルを用いた学内の研究分野の分類, 日本計算機統計学会 第32回大会, 日本計算機統計学会, 2018/5/27.

4. 今田一希・船山貴光・山田実俊・山本義郎, Shiny アプリケーションを用いたインタラクティブな顧客情報の分析と可視化, 日本計算機統計学会 第32回大会, 日本計算機統計学会, 2018/5/27.
5. 船山貴光・今田一希・竹中栄晶・木村英樹・福田紘大・山本義郎・中島孝, WSC2017 支援のための気象情報可視化・配信システムの構築, 情報処理学会 第80回全国大会, 情報処理学会, 2018/3/14.
6. 船山貴光・山本義郎・内田理, 災害時の Tweet 情報の変化に関する分析, 第17回西東京統計研究会, 西東京統計研究会, 2018/3/13.
7. 山本義郎・船山貴光・今田一希・中島孝・木村英樹・福田紘大・竹中栄晶, ソーラーカーレース支援のための気象情報可視化・配信システムの改良, 2017年度東海大学情報技術センター研究・開発報告会, 東海大学情報技術センター, 2018/3/3.
8. 船山貴光・山本義郎・内田理, 熊本地震の Tweet データの可視化, 科研費シンポジウム「空間データと災害の統計モデル」, 2018/1/27.
9. 船山貴光, 災害時の情報抽出のための Twitter データの解析と可視化, 日本計算機統計学会 2017年度若手セミナー, 日本計算機統計学会, 2017/9/2.
10. 船山貴光・山本義郎・古山恒夫, Shiny によるソフトウェア開発プロセスの要因分析, 日本計算機統計学会 第31回大会, 日本計算機統計学会, 2017/5/11.
11. 日比颯・船山貴光・山本義郎, 熊本地震における土砂崩れの始点の推定, 第16回西東京統計研究会, 西東京統計研究会, 2017/3/15.
12. 船山貴光・山本義郎・内田理, 災害時の情報収集のための Twitter データの解析, 第16回西東京統計研究会, 西東京統計研究会, 2017/3/15.
13. 船山貴光・山本義郎・内田理, 熊本地震後の Tweet データからの有益な情報の抽出と分類, 日本分類学会 第35回大会, 日本分類学会, 2017/3/7.
14. 日比颯・船山貴光・山本義郎, 土砂移動分布図からの土砂崩れ状況の考察, 科研費シンポジウム「空間データと災害の統計モデル」, 2017/1/28.
15. 船山貴光・山本義郎・内田理, 熊本地震の Tweet データの解析, 科研費シンポジウム「空間データと災害の統計モデル」, 2017/1/28.

16. 藤野友和・山本由和・船山貴光・山本義郎, 学術文献 DB における著者識別問題について, 日本計算機統計学会 第 30 回シンポジウム, 日本計算機統計学会, 2016/11/24.
17. 河野克哉・亀岡瑤・船山貴光・山田実俊・山本義郎, 売り上げ向上のための注目商品の抽出, 日本計算機統計学会 第 30 回大会, 日本計算機統計学会, 2016/5/20.
18. 日比颯・船山貴光・山田実俊・服部敦・山本義郎, 外部データを用いた時間毎来庁パターンの分類, 日本計算機統計学会 第 30 回大会, 日本計算機統計学会, 2016/5/20.
19. 船山貴光・渡邊武志・竹中栄晶・中島孝・山本義郎, EMS のための気象データ提供方法の改良, 日本計算機統計学会 第 30 回大会, 日本計算機統計学会, 2016/5/20.
20. 船山貴光・渡邊武志・竹中栄晶・木村英樹・福田紘大・山本義郎・中島孝, ソーラーカーレース支援のための気象情報可視化・配信システムの構築, 情報処理学会 第 78 回全国大会, 情報処理学会, 2016/3/12.
21. 日比颯・河野克哉・船山貴光・梶田佳孝・山本義郎, オープンデータと地域活動データの可視化, 第 15 回西東京統計研究会, 西東京統計研究会, 2016/3/9.
22. 船山貴光・山田実俊・山本義郎, サッカートラッキングデータの可視化と守備状況の指標の考察, 第 15 回西東京統計研究会, 西東京統計研究会, 2016/3/9.
23. 山本義郎・船山貴光・渡邊武志・中島孝・木村英樹・福田紘大・竹中栄晶, ソーラーカーレース支援のための気象情報可視化・配信システムの構築, 2015 年度東海大学情報技術センター研究・開発報告会, 東海大学情報技術センター, 2016/2/13.
24. 船山貴光・山本義郎・富田誠・梶田佳孝・田島祥・宇津圭祐・内田理, Web サービスと SNS 情報を活用した減災支援システムの開発, 科研費シンポジウム「空間データと災害の統計モデル」, 2015/12/25.
25. 日比颯・河野克哉・船山貴光・梶田佳孝・山本義郎, オープンデータと地域活動データの可視化への試み, 科研費シンポジウム「空間データと災害の統計モデル」, 2015/12/24.
26. 山本義郎・船山貴光・山田実俊, R を利用した災害情報の収集と解析・可視化, 2015 年度統計数理研究所共同研究集会「データ解析環境 R の整備と利用」, 統計数理研究所, 2015/12/5.

27. 内田理・小杉将史・遠藤岳・船山貴光・宇津圭祐・田島祥・富田誠・梶田佳孝・山本義郎, Twitter を用いた災害情報共有システム, 2015 年電子情報通信学会基礎・境界ソサイエティ大会, 電子情報通信学会, 2015/9/10.
28. 船山貴光, 多様な指標を用いた打者の調子の波の可視化, 統計サマーセミナー 2015, 2015/8/8.
29. 船山貴光・山田実俊・河野克哉・服部敦・山本義郎, 売り上げ傾向による店舗の分類と購買傾向の分析と可視化, 日本計算機統計学会 第 29 回大会, 日本計算機統計学会, 2015/5/15.
30. 船山貴光・山本義郎, 多様な指標を用いた打者の調子の波の可視化と比較, 第 14 回西東京統計研究会, 西東京統計研究会, 2015/3/17.
31. 池上拓人・八木圭太・船山貴光・山本義郎, 状況別の打撃成績による打者の分類, 第 14 回西東京統計研究会, 西東京統計研究会, 2015/3/17.
32. 河野克哉・船山貴光・八木圭太・山本義郎, 打者の好調不調の原因分析と守備能力の分析, 第 9 回 日本統計学会春季集会, 日本統計学会, 2015/3/8.
33. 池上拓人・八木圭太・船山貴光・山本義郎, セイバーメトリクスの指標を用いた打者の分類, 日本分類学会 第 33 回大会, 日本分類学会, 2015/3/3.
34. 船山貴光・山本義郎, ツイッター情報を活用した減災支援システムの構築へのオープンソース活用, 科研費シンポジウム「空間データと災害の統計モデル」, 2014/12/21.
35. 船山貴光・山本義郎, 減災支援システム構築の為にオープンソースの利用, 日本計算機統計学会 第 28 回シンポジウム, 日本計算機統計学会, 2014/11/14.
36. 船山貴光・山本義郎, 移動 OPS を用いた打者の調子の波の分析, 応用統計学会 2014 年度年会, 応用統計学会, 2014/5/22.
37. 宗像昌平・山田実俊・船山貴光・八木圭太・趙斌・山本義郎, アンケートの可視化によるメディア層の嗜好の分析, 日本計算機統計学会 第 28 回大会, 日本計算機統計学会, 2014/5/18.
38. 亀岡瑤・船山貴光・八木圭太・山田実俊・宗像昌平・山本義郎, 条件付きアソシエーションによる購入の特徴抽出, 日本計算機統計学会 第 28 回大会, 日本計算機統計学会, 2014/5/18.

参考文献

39. 船山貴光・山本義郎, 移動 OPS による打者の調子の波の可視化, 第 13 回西東京統計研究会, 西東京統計研究会, 2014/3/11.