

## 非負値テンソル因子分解を用いた災害時におけるメディア分析

## Analyses of Topics on Media in Disaster using Non-negative Tensor Factorization

細江 美欧<sup>†</sup>桑野 将司<sup>†</sup>長曾我部 まどか<sup>†</sup>福山 敬<sup>†</sup>石井 晃<sup>†</sup>

Mio Hosoe

Masashi Kuwano

Madoka Chosokabe

Kei Fukuyama

Akira Ishii

## 1.はじめに

2000 年代中頃から、ブログや電子掲示板、動画共有サービス、Social Networking Service (以下、SNS) などのソーシャルメディアの利用者の増加が著しい。たとえば、過去 1 年間に利用した経験のある SNS を訪ねた総務省の調査によると、Facebook は 35.3%、Twitter は 31.0% と、日本人の約 3 人に 1 人は SNS を利用しているといわれている<sup>1)</sup>。さらにスマートフォンの普及が、生活の中でソーシャルメディアをいつでもどこでも利用可能にし、さらには位置情報などのスマートフォンの多様な機能と連携して、その活用の幅が広がっている。このようなソーシャルメディアの普及によって、かつては情報を受信するだけの立場であった人が、リアルタイムに、かつ不特定多数に向けて情報を発信することが可能となった。これにより新聞や TV、ラジオといった限られた発信者から多くの受信者に向けた単方向の情報発信であったマスメディアの寡占的時代は終わり、発信者と受信者が双方向性を持って情報を媒介していくソーシャルメディアによる情報共有時代に移り変わったといえる。

2011 年 3 月 11 日に発生した東日本大震災は、ソーシャルメディアが普及した時代に発生した最初の大規模災害といわれている<sup>2)</sup>。これを機に平常時だけでなく災害発生時などの非日常時におけるソーシャルメディアによる情報の受発信の役割が活発に議論されるようになった<sup>3,4)</sup>。特に、ソーシャルメディアの 1 つである Twitter は、報道機関や政府、自治体、企業だけでなく一般のネットユーザーまで幅広く利用されたことから、社会学分野、交通分野、防災分野、情報処理分野など様々な研究分野において、3 月 11 日に発信された情報内容に関する研究が盛んになされた<sup>5,6,7)</sup>。これら研究から、リアルタイムの災害情報の取得や安否確認などのために Twitter が有効に機能したことが確認された一方、デマの拡散防止やノイズ対策が必要であることも明らかにされた。

災害時のソーシャルメディアでの発信内容に関する既往研究の多くは、ある 1 つのメディア、主に Twitter にのみ着目し、複数のメディアでの関連性に着目した研究は極めて少ない。誰もが限られた種類の TV 番組や新聞を見て、類似した情報を入手していた時代は終わり、インターネットとソーシャルメディアの隆盛によって、利用者が情報を取捨選択できるようになった現在においては、ある 1 つのメディアにのみ着目しても、その特徴を十分に把握できるとはいえない。各メディアの特徴を把握するためには、複数のメディアとの比較分析を行う必要があり、それを理解することによって、初めてマスメディアとソーシャルメディアのそれぞれが持つ特性を活かしながら、相互に補完し合って情報を発信する、すなわち、各メディアの特性を活かした災害報道のあり方を検討することが可能になると考えられる。

<sup>†</sup>鳥取大学 Tottori University

本研究では、2016 年 10 月 21 日 14 時 07 分に鳥取県中部で発生した最大震度 6 弱の地震 (以下、本震) を対象に、マスメディアとして TV、ネットニュースを、ソーシャルメディアとして Twitter、ブログ、掲示板、2ちゃんねるの合計 6 種のメディアから発信された情報内容とタイミングの関連性について比較分析を行い、各メディアの特性を把握することを目的とする。具体的には、どのようなメディアから、どのタイミングで、どのような情報が発信されたか、すなわち「メディアの種類」と「情報発信の時刻」および「情報内容」の 3 要素を同時に分析可能とする非負値テンソル因子分解を用いて、各メディアの特性を明らかにし、考察を行う。

## 2. 既往研究の整理と本研究の位置づけ

災害時にメディアが果たす役割は大きく、特にソーシャルメディアの活用が近年盛んに議論されている。東日本大震災時には、災害時の情報源として TV 放送を重視する人が多かっただけでなく、ポータルサイトやソーシャルメディアの情報を重視する人も多かったことが報告されている<sup>3,4)</sup>。特に、ソーシャルメディアは災害時に発生した不特定多数の情報を整理する際に活用されたことから<sup>8)</sup>、各メディアの特性に着目した研究も多く報告されている。三浦ら<sup>9)</sup>は、災害時の Twitter について、感情に関する内容を分析している。地震・津波に言及したツイートは、夜間に不安反応のピークがあり時間経過と共に減衰したことを明らかにしている<sup>9)</sup>。さらに、災害に関するツイートのうち、「怖い」、「恐ろしい」や「気がかりな」といったネガティブな感情語や活性度の高い感情語が含まれるツイートほど多数リツイートされていたことを明らかにしている<sup>10)</sup>。また、沼田・目黒<sup>11)</sup>は、Twitter、地方紙、全国紙、TV が発信した内容を分析し、Twitter は注意喚起に関する内容、地方紙は生活情報、全国紙は訃報に関する情報、TV は経済活動に関する報道がそれぞれ特徴的だったことを明らかにしている。さらに、相田ら<sup>12)</sup>は、2ちゃんねる、ブログ、Twitter が連携した支援活動について設立の経緯と活動内容を報告している。

多様な種類の SNS について、武田<sup>13)</sup>は「価値観」、「現実生活」、「情報交換」、「関係構築」の 4 つの観点から特性を整理している。既往研究<sup>8)-12)</sup>と武田<sup>13)</sup>を参考に、災害時における情報と各種メディアの関係を図 1 のように整理した。災害時には、情報が緊急かつ重要であるか否かが人々の行動に影響を与えると考えられるため、新たに「緊急性」の観点を加え、「緊急性 (高い・低い)」、「情報提供・収集 (一方向)」、「関係構築・情報交換 (双方向)」の 3 つの観点を設けた。図中の白抜き楕円は情報の内容を、四角の枠線は各メディアが主に取り扱う情報の範囲を示している。

災害直後に、避難指示や注意喚起といった緊急性が高い情報を提供するメディアは、TV、ネットニュース、Twitter であり、その後、被災状況や生活情報、被災地の商業や観光

について情報を提供するメディアは、新聞、TV、ネットニュース、掲示板などであると考えられる。一方、物資の要請やボランティアといった、緊急性が高く情報の発信者と受信者のコミュニケーションを必要とする情報を数多く扱うメディアは、Twitter、ブログ、2ちゃんねるが重要な役割を果たすと考えられる。このように、各種メディアがそれぞれの特性を活かし、多様な情報を発信・受信する役割を担っていると整理することができる。しかし、このようなメディア特性に関する整理は、既往研究でも多くなされているが、実際の災害時に発信された情報を定量的に分析し、各種メディアの特徴を同時に分析しながら比較した研究はない。本研究は、メディア別の発信情報の内容と発信されたタイミングを同時分析することによって、各メディアの特徴を把握する点に新規性がある。

メディアが取り上げる話題や交換された情報内容の特徴を定量的に分析する際には、第一段階の処理として、テキストマイニングが行われることが多い。ここで、テキストマイニングとは、金<sup>14)</sup>によると「蓄積された膨大なテキストデータを何らかの単位に分解し、これらの関係を定量的に分析すること」である。先に挙げた既往研究<sup>9),10),11)</sup>においても、テキストマイニングの形態素解析を行い、文章を単語単位に分解したうえで、クラスター分析や共分散構造分析、ネットワーク分析、コレスポネンス分析などの各種分析を行っている。松井<sup>15)</sup>は、20年分の雑誌記事タイトルを収集し、「癒し」という語の使われ方について分析している。コレスポネンス分析を用い、男女によって「癒し」の意味が異なること、時代の変化に伴い「癒し」の意味が世俗的、物質的なものに変化したことを明らかにしている。また、神田<sup>16)</sup>は福島原発事故に係る新聞報道の見出し文に対し、クラスター分析、共起ネットワーク分析、コレスポネンス分析を適用している。複数の多変量解析の結果より、新聞は「不安・不満」、「安心・大丈夫」といった概念的語彙で、健康影響の有無を間接的に説明している場合が多いことを考察している。以上のように、テキストマイニングの形態素解析を行った後に、データの持つ特徴を抽出する手法として、さまざまな手法が用いられてきた<sup>17),18)</sup>。これらの手法は、ある特定のメディアに注目して時間の経過と出現する単語の関係を分析する（「時間帯×単語」の分析）、あるいは、ある単語に注目し複数のメディア間での出現頻度の推移を比較する（「メディア×単語出現頻度」の分析）など、データ構造の次数を2次に変換した上で、分析を行っている。これは、本来の高次数で表されるデータの構造から、分析者の分析目的に応じて低次数のデータに変換、抽出し、分析していると言い換えることができ、データの持つ特徴のすべてを捕捉できていないとはいえない。

本研究では、本来の高次数のデータ構造を保持したまま、多要素によって構成される複雑なデータの特徴を抽出できるテンソル因子分解を用いて、災害時にメディアから発信された情報の特徴を分析する。テンソル因子分解とは、2次テンソルである行列データの次元を圧縮する際に因子分解が用いられるように、3次以上のテンソルデータの各モード（各要素軸）の次元を圧縮する際に用いられる。具体的には、どのメディアから、いつ、どのような内容が発信されたか、つまり「メディア×時間帯×出現語」で構成される3次数データを分析することにより、それら3つの関係性から災害時のメディア情報の特徴を明らかにする。

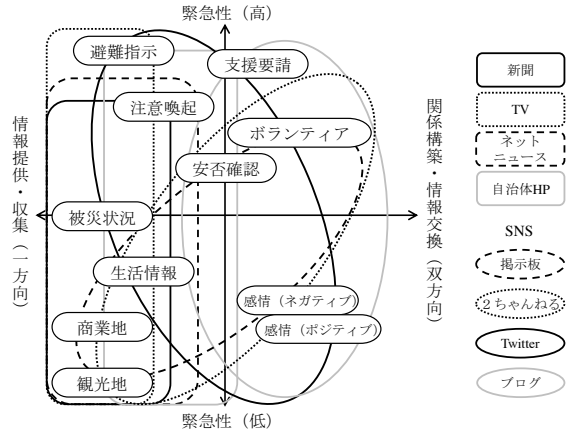


図1 災害時の情報とメディアの種類

### 3.使用データの概要

#### 3.1 データの収集方法

本研究では、クチコミ@係長<sup>19)</sup>から収集したメディアの1時間単位の発信件数、および発信情報のテキストデータを用いる。クチコミ@係長とは、株式会社ホット・リンクが提供しているソーシャル・ビックデータを収集するためのクラウドサービスでありTwitter、2ちゃんねる、ブログ、掲示板、ネットニュース、TVに関するデータが蓄積されている。

「Twitter」とは、つぶやき（Tweet）と称される140字以内の文を投稿し、他の利用者がそのつぶやきに対して反応することでコミュニケーションをとる情報サービスである。クチコミ@係長では、10%サンプリングされたデータが収集できる。「2ちゃんねる」とは、テーマ別に掲示板を作成し、利用者間でコミュニケーションを図る電子掲示板システムである。クチコミ@係長では、2ちゃんねる以外にYahoo!知恵袋や教えて!Gooなど4つの掲示板からデータを収集・蓄積しており、これらは2ちゃんねると区別して「掲示板」と呼ぶ。「ブログ」とは、個人が日々の出来事や感じたことなどを日記形式で投稿し、公開するウェブサイトである。閲覧者は投稿文にコメントを付記することができる。クチコミ@係長では、アメーバブログやYahoo!ブログなど国内ブログの約90%を網羅している。

「ネットニュース」に関するデータは、日刊アメーバニュースやNHK WEB NEWSなどニュースポータルサイトに掲載された日時とその記事内容を用いる。「TV」についてはNHK2局および関東キー局（日本テレビ、テレビ朝日、TBS、テレビ東京、フジテレビ）で放送された番組に関するデータである。

#### 3.2 分析対象と情報発信数の推移

2016年10月21日14時07分に発生した鳥取県中部の地震を分析対象に、10月21日14時から24時までの10時間分の各メディアで発信された情報を用いる。なお、関連する情報は、鳥取県内の地名である「鳥取」、「倉吉」、「湯梨浜」、「北栄」、「三朝」を検索語として抽出した。

各メディアの時間帯別発信件数の推移を図2(a)、(b)に示す。なお、図2(b)は、図2(a)の縦軸のスケールを変

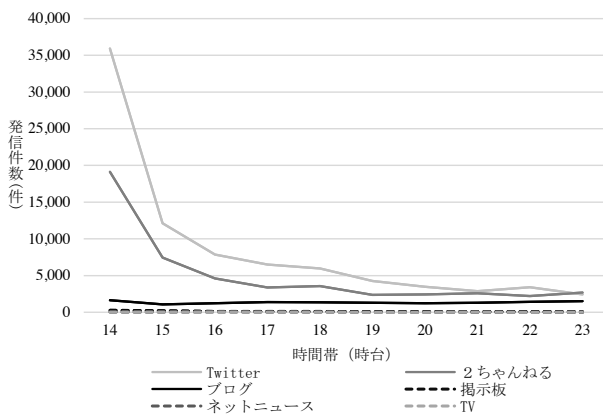


図2 (a) 各メディアの時間帯別発信件数の推移

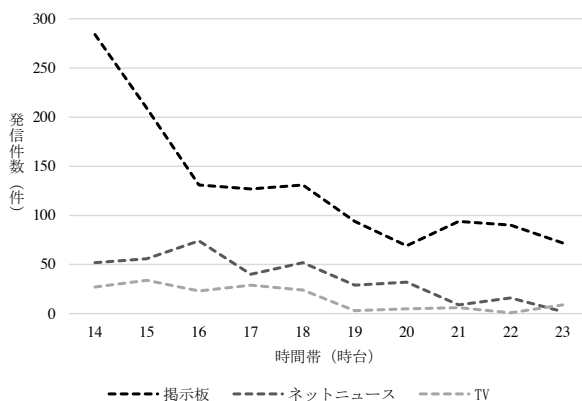


図2 (b) 各メディアの時間帯別発信件数の推移 (掲示板・ネットニュース・TVのみ)

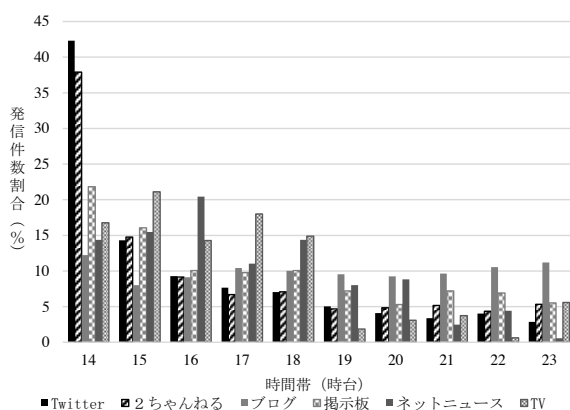


図3 各メディアの時間帯別発信件数割合の推移

更した図である。14時から24時までの10時間で発信された情報の総件数は、Twitterが84,893件、2ちゃんねるが50,418件、ブログが13,503件、掲示板が1,301件、ネットニュースが362件、TVが161件であった。時間帯別の発信件数に着目すると、本震が発生した時間帯である14時台はTwitterや2ちゃんねるの発信件数が顕著に多く、それぞれ35,918件/時間、19,119件/時間であった。

各メディアの時間帯別発信件数割合の推移を図3に示す。ここで発信件数割合とは、メディアごとの時間帯別を、分析

対象期間である10時間の総発信件数で除した値である。図3より、Twitter、2ちゃんねるは、14時台の発信件数割合が高く、その後急激に減少していることがわかる。すなわち、これらメディアでは、他のメディアと比べて、本震発生直後に多くの情報が発信されたが、時間の経過とともに発信件数が減少したことがわかる。一方で、ブログやネットニュース、TVは14時から18時まで割合の変化は比較的小さく、定期的に新しい情報が発信され続けていたことがわかる。全体の発信件数割合に着目すると、19時台以降は、発信件数割合に大きな変化が見られない。よって、各メディアにおいて、鳥取中部に関する地震の情報発信が、19時台以降には収束(安定)したと推察できる。

### 3.3 主な出現語の抽出

各メディアから発信された情報のテキストデータを用いて形態素解析を行い、発信内容における主な出現語を抽出する。ここで形態素解析とは、テキストデータを統計的に分析可能な状態に加工する処理のことであり、形態素とは意味の最小単位のことを指す。本研究では、テキストデータを形態素に分割し、品詞による選別を行い、名詞のみを抽出することとした。なお、既往研究では、分析とは直接関係しない単語(不要語)の削除、類似した意味を持つ単語あるいは発音は同じだが標記が異なる単語を1つの代表語にまとめる類義語の結合などを施すことが一般的である。この操作は、代表語の出現頻度が相対的に高くなり、頻出する話題の抽出が容易になるという利点がある。特に、Twitterや2ちゃんねるなどのSNSでは、必ずしも整理された文章で情報が発信されておらず、不要語削除や類義語結合の実施、およびその程度によって、その後に行われる結果が左右される。言い換えると、不要語削除や類義語結合の方法や程度は、分析者に依存しており、それによって得られる知見が異なる。

しかし、そもそもTwitterや2ちゃんねるなどのSNSでの発言には、分析者が注目する発言(単語)とは無関係な発言や話題が大量に含まれている。ある話題から始まり、次々と話題が展開し、結果的に無関係に見えるような発言が生まれることがSNSの本質ともいえよう。さらに、大量のデータから、分析者が意図しないような特徴や傾向を抽出することを分析目的とする場合、事前に分析者が、不要語を定めて削除したり、類義語を結合したりすることは、予期せぬ結果の抽出を試みるデータマイニング分析の視点からは必ずしも適切とはいえない。そこで本研究では、不要語削除や類義語結合などで生じる分析者の恣意性を排除するために、これら操作は行わず、不要語や類義語を残したまま分析することとした。

表1に形態素解析結果の概略を示す。なお、表中の割合とは全形態素に占める割合のことであり、6メディア、10時間分のテキストデータ数は10,506件であり、単語の種類は25,461単語、総語数は910,339語であった。このうち形態素解析によって、検索語とした「鳥取」、「倉吉」、「湯梨浜」、「北栄」、「三朝」を除く名詞のみを抽出した結果、名詞の種類は12,428単語、名詞単語総数は215,272語であった。名詞のみを抽出した段階で全形態素数の約49%となったが、このまま分析を行うには膨大な単語数がある。そこで、名詞の種類数の頻出回数上位1%にあたる124種類の単語を抽出し、さらにその中から独立で意味をなさないアカウント

表1 形態素分析による単語の抽出

	単語の種類数 (割合)	単語の数 (割合)
全形態素	25,461 (100%)	910,339 (100%)
名詞	12,428 (48.81%)	215,272 (23.65%)
上位1%の名詞 (独立で意味をなさない単語は除く)	114 (0.45%)	116,630 (12.81%)

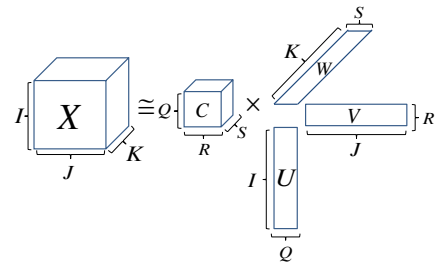


図4 NTD概念図 (3次テンソル)

表2 分析で用いる単語の種類と出現回数

出現語	総出現回数	出現語	総出現回数
地震	15,642	M6.6	379
震度	14,971	投入	378
観測	6,024	建物	377
震源	3,788	原発	377
発生	3,269	北部	375
禁止	2,916	発表	356
転載	2,910	様子	355
地方	2,891	テレビ	328
無断	2,889	施設	323
岡山	2,871	兵庫	309
心配	2,770	避難	307
津波	2,422	京都	298
気象庁	2,206	温泉	296
中部	2,141	断層	286
揺れ	2,104	場所	282
マグニチュード	2,016	火災	279
中部	1,872	国宝	277
鏡野	1,796	南海トラフ	277
被害	1,761	住宅	269
九州	1,446	中心	269
関東	1,425	松江	268
情報	1,418	熊本地震	268
島根	1,052	電話	266
速報	1,027	中継	258
クラシック	1,007	近畿	256
最大	969	携帯	243
余震	888	映像	240
各地	785	島町	239
総合	750	市内	237
規模	723	東経	226
琴浦	635	北緯	221
停電	634	西日本	218
災害	615	東京	216
実況	610	会見	214
広島	593	県民	213
注意	592	中国電力	213
真庭	588	ミヤ	211
推定	561	仕事	209
大阪	551	市役所	206
確認	534	西部	206
日本	530	皆様	205
運転	519	智頭	205
広島	519	皆さん	202
影響	487	被災	201
隠岐	484	出雲	201
倒壊	476	中国地方	200
ニュース	463	付近	198
地域	459	大山	198
日吉津	430	境港	194
土下	423	けが人	189
葵	419	大震災	189
米子	417	四国	189
県内	414	関係	187
鹿野	414	南部	187
状況	406	空港	186
龍島	393	青谷	185
熊本	389	八頭	185

ルとは多次数の配列を表すものであり、スカラーは0次のテンソル、ベクトルは1次のテンソル、行列は2次のテンソルといえる。

本研究では、テンソル因子分解手法の1つである、Non-negative Tucker Decomposition (NTD) (20),(21),(22))を用いる。NTDは、非負値行列因子分解 (Non-negative Matrix Factorization) を3次元以上の多次元構造を持ったデータの分解に拡張した手法であり、脳波解析や画像識別の分野で用いられている。NTDでは、各次元 (各要素軸) の特徴を表す行列 (以下、特徴行列) とコアテンソルに分解し、それらからデータの特徴を把握する。

3次テンソルを例に NTDの概念図を図4に示す。次数が3で  $I \times J \times K$  の大きさを持つ3次テンソル  $X$  を考える。次元圧縮とは、各要素軸について低次元空間に写像することであり、 $I$  のモードに関して次元を  $Q$  ( $Q \leq I$ ) に、 $J$  のモードに関して次元を  $R$  ( $R \leq J$ ) に、 $K$  のモードに関して次元を  $S$  ( $S \leq K$ ) に分解し、情報を落とし込む方法がNTDである。具体的には、それぞれ次元が  $I, J, K$  で与えられる3つのモードを持つ  $X$  を、 $I$  のモードに関する情報をもつ  $I \times Q$  の  $U$ 、 $J$  のモードに関する情報をもつ  $J \times R$  の行列  $V$ 、 $K$  のモードに関する情報を持つ  $K \times S$  の行列  $W$  と、次元  $Q \times R \times S$  のコアテンソル  $C$  に分解する。この分解をランク ( $Q, R, S$ ) タッカー分解と呼び、 $Q, R, S$  を各モードに対する因子数と呼ぶ。具体的には次式で定式化される。

$$X \cong C \times U^T \times V^T \times W^T \quad (1)$$

ここで、 $C$  はコアテンソル、 $U, V, W$  は特徴行列である。式 (1) を要素ごとに書くと、式 (2) となる。

$$x_{ijk} \cong \sum_{q=1}^Q \sum_{r=1}^R \sum_{s=1}^S c_{qrs} u_{iq} v_{jr} w_{ks} \quad (2)$$

なお、特徴行列において圧縮する次元数 ( $Q, R, S$ ) は分析者が任意に決定する。

NTDはパラメータとして  $C, U, V, W$  を持っており、2乗誤差を最小にするように、非負値制約を与えてパラメータを決定する。つまり、パラメータ推計は以下の問題として定式化される。

$$\begin{aligned} & \min_{C,U,V,W} f(C, U, V, W) \\ & f(C,U,V,W) = \|X - C \times U^T \times V^T \times W^T\|_F^2 \\ & \text{s.t. } c_{qrs} \geq 0, u_{iq} \geq 0, v_{jr} \geq 0, w_{ks} \geq 0 \end{aligned} \quad (3)$$

ここで、 $\|\cdot\|_F$  はフロベニウスノルムを表す。

名や記号などを削除した 114 単語を分析対象とした。分析で使用する単語の種類、および出現回数を表2に示す。

#### 4.非負値テンソル因子分解

テンソル因子分解とは、3要素以上で構成される多次元データの特徴を抽出する手法である(20),(21),(22))。ここで、テンソ

## 5.各種メディアの発信内容とタイミングの関連性分析

本研究では、6種のメディアにおいて、2016年10月21日14時台から23時台までの10時間帯に、発信された情報における出現語114単語を用いて、メディア×時間帯×出現語の3次テンソルを構築することにより、これら3要素軸を考慮した同時分析を行う。すなわち、3次テンソル( $I \times J \times K$ )のサイズは $6 \times 10 \times 114$ となる。この3次テンソルデータに、非負値テンソル因子分解を適用し、メディア、時間帯、出現語の各要素の特徴行列とコアテンソルを抽出する。分析に際して、各要素の特徴行列の因子数を分析者が予め設定しなければならない。本研究では分析結果が解釈しやすく、モデル適合度が高い因子数を試行錯誤的に探索し、各要素の因子数を決定することとした。その結果、因子数( $Q, R, S$ )はそれぞれメディア( $Q$ )を4因子、時間帯( $R$ )を3因子、出現語( $S$ )を5因子とした。コアテンソルのデータサイズは特徴行列の因子数と対応するため、 $4 \times 3 \times 5$ のサイズを持つ3次テンソルとなる。式(2)、式(3)からわかるように、各要素の特徴行列とコアテンソルは同時推定されるが、便宜上、以降では各要素の特徴行列とコアテンソルの結果について順に考察する。

### 5.1 特徴行列に関する結果の考察

#### 5.1.1 メディアに関する特徴行列

メディアに関する特徴行列の分析結果を表3に示す。

第1因子の各メディアに対する成分値は「Twitter」での値が0.988と最も高く、次いで「2ちゃんねる」での値が0.300となっている。その他の「ブログ」、「掲示板」、「ネットニュース」、「TV」での値はゼロである。また「2ちゃんねる」に関しては、第4因子以外の因子で0.261~0.300の値を示した。これは「2ちゃんねる」は、第4因子(TV)以外のメディアと類似性があることを表している。ただし、各因子の最大値に対する「2ちゃんねる」の成分の相対値に着目すると、第1因子において最も高い値を示していることから、本分析では「2ちゃんねる」は「Twitter」と類似性が高いメディアと分類する。

同様に考察すると、第2因子は「ネットニュース」と「ブログ」、第3因子は「掲示板」、第4因子は「TV」の特徴を表した成分であるといえる。

以上の結果から、6種のメディアは「Twitter・2ちゃんねる」、「ネットニュース・ブログ」、「掲示板」、「TV」の4グループに分類できることが明らかとなった。上述したように、2ちゃんねるは掲示板の一種と考えることもできる。しかし、発信された情報のタイミングと情報の内容によって分類した場合、2ちゃんねるは、掲示板ではなく、Twitterと類似性が高いことが明らかとなり、これは本研究で得られた新たな知見であるといえる。さらに、マスメディアのネットニュースとソーシャルメディアのブログとの間には類似性が見られ、マスメディア(TV、ネットニュース)で一群、ソーシャルメディア(Twitter、2ちゃんねる、ブログ、掲示板)で一群という一般的な分類は必ずしも適切ではないことが明らかとなった。

表3 メディア因子

第1因子	値	第2因子	値
Twitter	0.988	ネットニュース	0.678
2ちゃんねる	0.300	ブログ	0.402
ブログ	0.000	2ちゃんねる	0.261
掲示板	0.000	Twitter	0.000
ネットニュース	0.000	掲示板	0.000
TV	0.000	TV	0.000

第3因子	値	第4因子	値
掲示板	0.918	TV	0.796
ブログ	0.352	ネットニュース	0.210
2ちゃんねる	0.299	掲示板	0.010
Twitter	0.000	Twitter	0.000
ネットニュース	0.000	2ちゃんねる	0.000
TV	0.000	ブログ	0.000

表4 時間帯因子

第1因子	値	第2因子	値	第3因子	値
14時台	1.354	15時台	0.707	19時台	0.485
15時台	0.094	16時台	0.637	20時台	0.371
16時台	0.000	17時台	0.580	18時台	0.292
17時台	0.000	18時台	0.221	22時台	0.278
18時台	0.000	23時台	0.110	21時台	0.274
19時台	0.000	21時台	0.015	23時台	0.172
20時台	0.000	22時台	0.009	16時台	0.081
21時台	0.000	14時台	0.000	15時台	0.056
22時台	0.000	19時台	0.000	14時台	0.000
23時台	0.000	20時台	0.000	17時台	0.000

#### 5.1.2 時間帯に関する特徴行列

時間帯に関する特徴行列の分析結果を表4に示す。表4の考察は、表3と同様に行うが、冗長性を避けるため、ここでは分析結果の特徴についてのみ説明する。

第1因子は「14時台」の値が顕著に高く、第2因子は「15時台」、「16時台」、「17時台」の値が高い。第3因子は「18時台」、「19時台」、「20時台」、「21時台」、「22時台」、「23時台」の値が高いことが表4よりわかる。以上の結果より、時間帯別での特徴は、「本震直後」の14時台、本震から1時間程度経過した15時台から17時台までの「夕方」、18時台から深夜までの「夜」の3つに分類できることがわかった。これら時間帯に関する特徴は、直感に即したものであり、妥当な結果であるといえる。

#### 5.1.3 出現語に関する特徴行列

出現語に関する特徴行列の分析結果を表5に示す。ただし、出現語に対する特徴行列は、単語数が114種類あり、全ての値を比較、考察することは複雑である。そこで、ここでは各因子の最大成分値に対して70%以上の単語のみを、その因子の特徴を表現する単語として注目する。例えば、第1因子では「国宝」の成分値が1.054なので、成分値が0.738以上の単語のみを考察の対象としている。

第1因子の上位には「国宝」と「投入」の単語が含まれており、これら単語は他の因子には含まれていない。これらは鳥取県三朝町の国宝に指定されている三徳山三佛寺・投入堂を示す単語であると考えられ、この単語は第1因子のみ抽出されている。その他「関東」、「南海トラフ」といった語が上位にあることが第1因子の特徴であるといえる。よって、第1因子は、「国宝の被害状況、および南海トラフ関連」の情報と考えられる。これら国宝の投入堂と南海トラフ

表 5 出現語因子

第 1 因子	値	第 2 因子	値	第 3 因子	値
国宝	1.054	九州	1.181	近畿	1.689
西日本	0.944	四国	1.163	推定	1.612
皆さん	0.819	西部	1.140	四国	1.599
大震災	0.813	火災	1.114	速報	1.520
関東	0.811	付近	1.092	地方	1.410
投入	0.808	北部	1.078		
建物	0.795	中国地方	1.011		
様子	0.784	広島	0.999		
仕事	0.777	各地	0.990		
ミヤ	0.777	推定	0.965		
会見	0.759	規模	0.948		
南海トラフ	0.752	中国電力	0.943		
総合	0.745	近畿	0.918		
県民	0.741	兵庫	0.912		
施設	0.739	京都	0.910		
		隠岐	0.908		
		葵	0.866		
		土下	0.856		
		龍島	0.852		
		M6.6	0.851		
		境港	0.851		
		大阪	0.837		

第 4 因子	値	第 5 因子	値
智頭	1.614	琴浦	0.981
南部	1.594	被災	0.944
青谷	1.521	関係	0.884
鹿野	1.502	災害	0.854
龍島	1.371	日本	0.832
土下	1.280	火災	0.831
葵	1.256	付近	0.794
島町	1.255	中継	0.788
大山	1.195	熊本地震	0.780
八頭	1.150	広島	0.777
		市内	0.752
		地域	0.744
		大山	0.735
		発表	0.734
		電話	0.734
		テレビ	0.724
		影響	0.715
		ニュース	0.714
		携帯	0.703
		県内	0.698
		会見	0.688
		大山	0.735
		発表	0.734
		電話	0.734
		テレビ	0.724
		影響	0.715
		ニュース	0.714
		携帯	0.703
		県内	0.698
		会見	0.688

に関する情報が同時に情報発信される傾向が強いことは、事前に分析者が予見することは困難であり、共起性が高い 2 つの異なる話題が抽出できたことは、本分析手法適用の有用性を示す結果といえる。

第 2 因子は「九州」、「四国」、「中国地方」、「近畿」などの語から鳥取だけでなく、西日本全体の広範囲における地震情報と考えられる。また本震の程度を表す「推定」、「規模」、「M6.6」といった単語、および「火災」、「中国電力」の単語の数値が相対的に高いことから、第 2 因子は「西日本の地震状況」に関する情報と考えられる。

第 3 因子は、「近畿」、「推定」、「四国」、「速報」、「地方」の 5 単語のみで構成される。特に「速報」が含まれているのは第 3 因子のみであることから、これは「鳥取県とその一帯

の地震速報」に関する情報と考えられる。

第 4 因子は「智頭」、「南部」、「青谷」、「鹿野」など鳥取県内に位置する市町村名のみが含まれ、他の因子で検出された地震の規模や被害状況に関する単語は抽出されなかった。しかし、本震直後に発信された情報を分析していることから、地名だけを発信するとは考えにくく、文脈を推察すると、第 4 因子は「鳥取県内の地震状況」に関する情報と考えられる。

第 5 因子は「被災」、「災害」、「火災」などの災害状況に関する単語が上位を占めている。また、他の因子の上位には挙がらなかった「熊本地震」が高い数値を示したことから、他の因子と比べて多くの単語で構成されていることが第 5 因子の特徴といえる。よって、第 5 因子は「被災状況と熊本地震関連」の情報と考えられる。

## 5.2 コアテンソル値による各メディアの特徴把握

コアテンソルの分析結果を図 5 に示す。図 5 はコアテンソルの組合せに特徴行列因子の結果を対応付けたメディア因子別の結果である。コアテンソルは各特徴行列の因子の組合せ ( $Q \times R \times S$ ) 分の  $4 \times 3 \times 5 = 60$  の要素が存在する。図中の数字は因子の組合せであり、メディア因子 ( $Q$ )、時間帯因子 ( $R$ )、出現語因子 ( $S$ ) の順となっている。例えば、(1,2,3) はメディアが第 1 因子、時間帯が第 2 因子、出現語が第 3 因子の組合せを表現している。

図 5 より、コアテンソル値が最も大きいのは、(4,2,5) の組合せであることがわかる。すなわち、メディア第 4 因子、時間帯第 2 因子、出現語第 5 因子の組合せであり、これは「TV」から「夕方」に「被災状況と熊本地震関連」の情報が最も多く発信されたことを示している。次いで、コアテンソル値が大きい、すなわち発信された情報量が多いのは (2,1,4) の「ネットニュース・ブログ」から「本震直後」に「鳥取県内の地震状況」に関する情報であり、3 番目に大きいのは (1,1,2) の「Twitter・2ちゃんねる」から「本震直後」に「西日本の地震状況」に関する情報であることが分析結果からわかる。

図 5 の結果を、メディア因子毎 ( $Q=1,2,3,4$ ) に整理した結果を表 6~表 9 に示す。

表 6 のメディア第 1 因子 (Twitter・2ちゃんねる) に関する結果に着目する。時間帯因子別の値が大きい組合せに着目すると、「本震直後」は「西日本の地震状況」の値が高く、「夕方」と「夜」は「国宝の被害状況、および南海トラフ関連」の値が大きいことがわかる。これは、Twitter や 2ちゃんねるでは「本震直後」は「西日本の地震状況」に関する情報発信が多かったが、「夕方」になると国宝の被害状況や南海トラフに話題が移り変わったことを意味している。さらに、「国宝の被害状況、および南海トラフ関連」の値を「夕方」と「夜」で比較すると、「夜」のほうが値は低いため、発信される情報数が「夜」になって減少したことがわかる。

次に、表 6~表 9 間での値の比較を行う。表 6~表 9 の「本震直後」の値を比較すると、メディア第 1 因子 (Twitter・2ちゃんねる) では、「西日本の地震情報」、メディア第 2 因子 (ネットニュース・ブログ) では、「鳥取県内の地震状況」、メディア第 4 因子 (TV) では「鳥取県とその一帯の地震速報」の値が高い。すなわち、「本震直後」の「Twitter・2ちゃんねる」、「ネットニュース・ブログ」、「TV」は異なる種

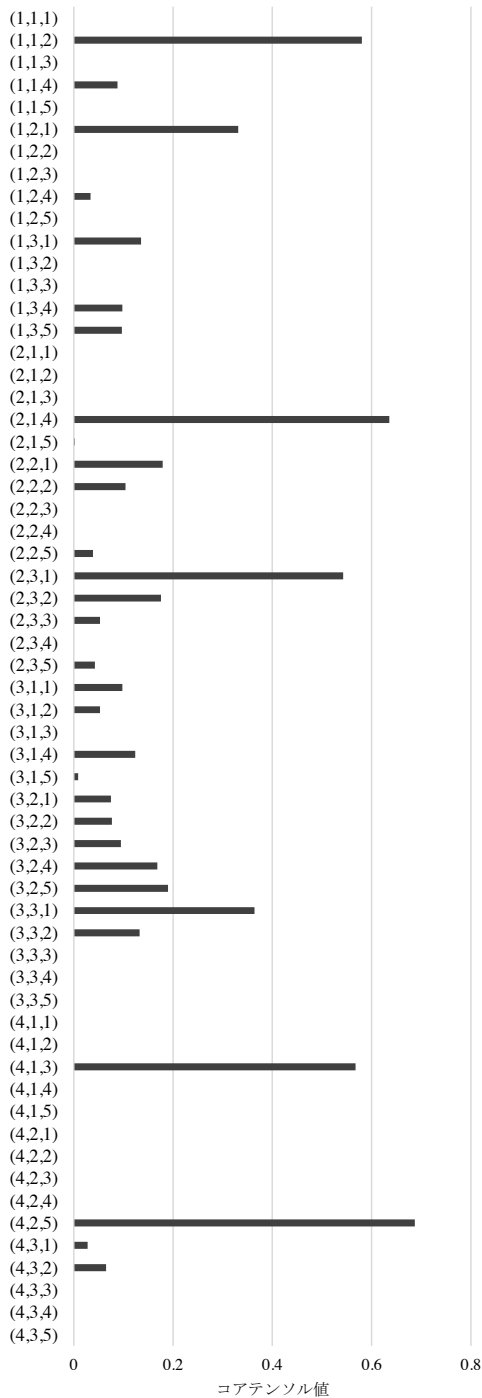


図5 コアテンソル

類の情報を発信しており、それぞれ西日本、鳥取県内、鳥取県周辺地域に関連する情報を特徴的に発信したことが明らかとなった。

一方、メディア第3因子(掲示板)では「本震直後」に大きな値は見られない。すなわち、「本震直後」は「掲示板」からの情報発信は少ないことも、本分析結果よりわかる。

また、表6～表9の「国宝の被害状況、および南海トラフ関連」に着目すると、「本震直後」はメディア第3因子(掲示板)のみで0.098と値は小さいが正値を示しており、他の

表6 メディア第1因子 (Twitter・2ちゃんねる)

	本震直後	夕方	夜
国宝の被害状況、および南海トラフ関連	0.000	0.331	0.136
西日本の地震状況	0.580	0.000	0.000
鳥取県とその一帯の地震速報	0.000	0.000	0.000
鳥取県内の地震状況	0.088	0.034	0.098
被災状況と熊本地震関連	0.000	0.000	0.097

表7 メディア第2因子 (ネットニュース・ブログ)

	本震直後	夕方	夜
国宝の被害状況、および南海トラフ関連	0.000	0.179	0.543
西日本の地震状況	0.000	0.104	0.176
鳥取県とその一帯の地震速報	0.000	0.000	0.053
鳥取県内の地震状況	0.636	0.000	0.000
被災状況と熊本地震関連	0.002	0.039	0.043

表8 メディア第3因子 (掲示板)

	本震直後	夕方	夜
国宝の被害状況、および南海トラフ関連	0.098	0.075	0.364
西日本の地震状況	0.053	0.077	0.133
鳥取県とその一帯の地震速報	0.000	0.095	0.000
鳥取県内の地震状況	0.124	0.168	0.000
被災状況と熊本地震関連	0.009	0.190	0.000

表9 メディア第4因子 (TV)

	本震直後	夕方	夜
国宝の被害状況、および南海トラフ関連	0.000	0.000	0.028
西日本の地震状況	0.000	0.000	0.065
鳥取県とその一帯の地震速報	0.568	0.000	0.000
鳥取県内の地震状況	0.000	0.000	0.000
被災状況と熊本地震関連	0.000	0.687	0.000

メディア因子ではゼロである。「夕方」になると、「国宝の被害状況、および南海トラフ関連」の値は、メディア第1因子(Twitter・2ちゃんねる)、メディア第2因子(ネットニュース・ブログ)で、それぞれ0.331、0.179となり、メディア第3因子(掲示板)では、0.075となっている。以上の結果は、「国宝の被害状況、および南海トラフ関連」に関する情報は、「本震直後」は情報量は少ないが「掲示板」のみで発信され、「夕方」になって「Twitter・2ちゃんねる」、「ネットニュース・ブログ」へと拡散したと推察できる。さらに、「夜」のメディア第3因子(掲示板)では、値が0.364となっている。よって、「掲示板」では「震災直後」に「国宝の被害状況、および南海トラフ関連」の話題が発信され、「夕方」に情報発信量は減少したが、「夜」になって再度活発に話題になったことが明らかとなった。

以上の結果を整理すると、「本震直後」は「Twitter・2ちゃんねる」から「西日本の地震状況」、「ネットニュース・ブログ」から「鳥取県内の地震状況」に関する情報が顕著に発信された。「夕方」になると「TV」からの「被災状況と熊本地震関連」の情報発信が増え、本震当日に発信された情報量としてはこれが最も多い。一方、「夜」になると「TV」や

「Twitter・2ちゃんねる」からの情報発信量は減少し、再び「ネットニュース・ブログ」からの情報発信量が増加する。また、「掲示板」からの情報発信量も増加し、南海トラフに対する懸念など、鳥取県外での地震のリスクやその備えに関する話題へと移り変わったと推察できる。

## 6.おわりに

本研究では、2016年10月21日に発生した鳥取県中部の地震を事例に、メディアと時間帯、発信された情報の内容を同時に分析するために、非負値テンソル因子分解を用いた方法論の提案を行った。既往研究において、災害時のメディア情報を分析した研究は多いが、複数のメディアを同時に分析した研究は筆者らの知る限り存在しない。分析に際して、メディアの種類、情報発信された時間帯、情報の内容という3つの次数(要素軸)を考慮すると、その組合せによって要素数は爆発的に増加するため、データ全体の特徴を把握することは容易ではない。本研究では、6メディア×10時間帯×114出現語で6,840の要素数があるデータに、本手法を適用することによって4メディア×3時間帯×5出現語の60要素に縮約した。実証分析の結果から、データ特性の特徴解釈が容易となることが示され、非負値テンソル因子分解が複数メディアの比較分析に適用可能であることが明らかとなった。

提案手法によって抽出される特徴行列の結果より、本研究が対象とした6種のメディアは「Twitter・2ちゃんねる」、「ネットニュース・ブログ」、「掲示板」、「TV」の4種類に分類可能であることがわかった。また、情報発信のタイミングとして「本震直後(14時台)」、「夕方(15時台~17時台)」、「夜(18時台~23時台)」の3時間帯に分類でき、さらに発信された情報は大きく5つの内容に分類できることが示された。

また、特徴行列と同時に推定されるコアテンソルの値より、どのメディアから、どの時間帯に、どのような内容の情報が、どの程度発信されたかを把握することが可能となった。具体的には、本事例においては「TV」から「夕方」に「被災状況と熊本地震関連」の情報が最も多く発信されたことが明らかとなった。また、コアテンソル値からメディア別あるいは時間帯別に発信された情報の内容の比較が可能となり、各メディアにおける発信内容の変遷や他メディアへの情報の拡散や伝搬などの分析に応用可能であることが示唆された。これは、複数のメディア情報を同時に分析した本研究の成果であるといえる。

本研究では、テキストデータの事前処理として名詞のみを抽出し分析を行った。今後は、動詞や助動詞などの品詞を同時に分析できるフレームワークを構築し、発信された情報内容をより詳しく分析することによって、災害時の各メディアで発信される情報の質を考慮した分析に展開する必要がある。

### 参考文献

- [1] 総務省, 平成27年度版 情報通信白書,SNSの利用率, <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/html/nc242220.html>(2017).
- [2] 林 春男, “東日本大震災と災害報道”, [https://www.nhk.or.jp/bunken/book/media/pdf/2014\\_11.pdf](https://www.nhk.or.jp/bunken/book/media/pdf/2014_11.pdf)(2014).
- [3] 野村総合研究所, “震災に伴うメディア接触動向に関する調査を実施”, <https://www.nri.com/jp/news/2011/110329.html>(2011).
- [4] 執行 文子, “東日本大震災・ネットユーザーはソーシャルメディアをどのように利用したのか”, [https://www.nhk.or.jp/bunken/summary/research/report/2011\\_08/20110801.pdf](https://www.nhk.or.jp/bunken/summary/research/report/2011_08/20110801.pdf)(2011).
- [5] 安田 雪, “ソーシャルメディア上の情報拡散の特性—東日本大震災時のデマの事例とハブの役割”, 関西大学社会学部紀要, Vol.45, No.1, pp.33-46(2013).
- [6] 白井 翔平, 鳥海 不二夫, 石井 健一郎, 間瀬 健二, “震災による情報伝播ネットワークの変化”, 人工知能学全国大会論文集, Vol.26, pp.1-4(2012).
- [7] 濱岡 豊, “震災におけるe流言の発生と消滅:ソーシャルメディアとマスメディアの影響比較”, 情報処理学会全国大会論文集, Vol.75, pp.4-517-4-518(2013).
- [8] 小林 慎太郎, “ソーシャルメディアに期待される「新しい公共」としての役割”, NRIパブリックマネジメントレビュー, Vol.94, pp.1-6(2011).
- [9] 三浦 麻子, 小森 政嗣, 松村 真宏, 前田 和甫, “東日本大震災時のネガティブ感情反応表出—大規模データによる検討—”, 心理学研究, Vol.86, No.2, pp.102-111(2015).
- [10] 三浦 麻子, 鳥海 不二夫, 小森 政嗣, 松村 真宏, 平石 界, “ソーシャルメディアにおける災害情報の伝播と感情:東日本大震災に際する事例”, 人工知能学会論文誌, Vol.31, No.1, pp.NFC-A\_1-9(2016).
- [11] 沼田 宗純, 目黒 公郎, “東日本大震災におけるTwitter・地方新聞・全国新聞・テレビ報道のキーワード分析による基礎的な比較”, 生産研究, Vol.65, No.4, pp.379-385(2013).
- [12] 相田 慎, 新堂 安孝, 内山 将夫, “「東日本大震災関連の救助要請情報抽出サイト」による救助活動支援”, 自然言語処理, Vol.20, No.3, pp.405-422(2013).
- [13] 武田 隆, “ソーシャルメディア進化論”, ダイアモンド社(2011).
- [14] 金 明哲, “テキストデータの統計科学入門”, 岩波書店(2009).
- [15] 松井 剛, “言語とマーケティング—「癒し」ブームにおける意味創造プロセス”, 組織科学, Vol.46, No.3, pp.87-99(2013).
- [16] 神田 玲子, 辻 さつき, 米原 英典, “東電福島第一原発事故関連の放射線に関する新聞記事見出し及びインターネットコンテンツのテキストマイニング解析”, 保健物理, Vol.49, No.2, pp.68-78(2014).
- [17] 樋口 耕一, “現代における全国紙の内容分析の有効性—社会意識の探索はどこまで可能か”, 行動計量学, Vol.38, No.1, pp.1-12(2011).
- [18] 樋口 耕一, “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”, ナカニシヤ出版(2014).
- [19] 株式会社ホット・リンク, クチコミ@係長, <https://www.hottolink.co.jp/service/kakaricho>.
- [20] 松林 達史, 幸島 匡宏, 林 亜紀, 澤 田宏, “非負値テンソル因子分解を用いた購買行動におけるブランド選択行動”, 人工知能学会論文誌, Vol.30, No.6, SPI-C, pp.713-720(2015).
- [21] 矢野 肇, 滝口 哲也, 有木 康雄, 神谷 勝, 中川 誠司, “脳磁界計測を用いたエアコン音の聴感印象推定の試み—非負値テンソル分解による関連脳活動の抽出—”, 日本音響学会研究発表会講演論文集, CD-ROM, pp.2-P-48(2016).
- [22] 石黒 勝彦, 林 浩平, “関係データ学習”, 講談社(2016).