



災害時のソーシャルメディアと帰宅行動の関連性分析

榎本, 甫
桑野, 将司
小池, 淳司

(Citation)

土木学会論文集D3 (土木計画学), 70(1):102-112

(Issue Date)

2014

(Resource Type)

journal article

(Version)

Version of Record

(Rights)

©2014 公益社団法人 土木学会

(URL)

<https://hdl.handle.net/20.500.14094/90002812>



災害時のソーシャルメディアと 帰宅行動の関連性分析

榎本 甫¹・桑野 将司²・小池 淳司³

¹学生員 神戸大学大学院 工学研究科市民工学専攻 (〒 657-8501 兵庫県神戸市灘区六甲台町 1-1)
E-mail: hajime_enomoto@stu.kobe-u.ac.jp

²正会員 鳥取大学大学院 工学研究科社会基盤工学専攻 (〒 680-8552 鳥取県鳥取市湖山町南 4-101)
E-mail: kuwano@sse.tottori-u.ac.jp

³正会員 神戸大学大学院 工学研究科市民工学専攻 (〒 657-8501 兵庫県神戸市灘区六甲台町 1-1)
E-mail: koike@lion.kobe-u.ac.jp

東日本大震災時、ソーシャルメディアが活発に利用され注目を浴びた。なかでも Twitter は、従来のマスメディアでは扱いきれない情報が多くやり取りされ、被災者の情報支援に役立った。しかし一方で、Twitter で発信される情報の信頼性は必ずしも担保されていない。そのような状況下で、Twitter におけるテキスト情報が人々の行動にどう影響しているかは明らかではない。本研究では、ソーシャルメディア情報を用いて災害時の避難行動を誘導する新たな施策立案に資する情報を提供するという立場から、両者の間の関連性を定量的に把握することを目的とする。東日本大震災時の東京都内で発信された Twitter のテキスト情報と人口流動メッシュデータを用いた分析の結果、両者には相関関係があることを明らかにした。

Key Words : *social media, Twitter, text mining, canonical correlation analysis, the Great Eastern Japan Earthquake*

1. はじめに

不特定多数との情報のやり取りを可能にしたソーシャルネットワークサービス（以降、SNS）は、個人のライフスタイルにおいて無視できない重要なメディアとなっている。さらに昨今は、リアルタイムでメッセージを発信する Twitter、実名登録で誰とでもつながりが持てる Facebook、携帯電話の電話帳を利用したクローズドなコミュニケーションを提供する LINE など新しい SNS が急成長しており、日本国内での利用者は、2012 年 3 月時点で Twitter が 1,400 万人¹⁾、Facebook が 1,490 万人¹⁾、2012 年 11 月時点で LINE が 3,600 万人²⁾といわれている。スマートフォンなどのモバイルデバイスの急速な普及による利用機会や利用場所の拡大などもあり、ソーシャルメディア利用者は今後さらに拡大すると予想されている。

SNS を利用した情報（以降、ソーシャルメディア情報）は、誰でも容易に情報を発信でき、従来のメディアに比べ情報伝播力が強いという有用性から、日常時だけでなく、非常時の新たな情報交換ツールとして注目されている。これは東日本大震災時には、各個人が発生源となる SNS を利用したソーシャルメディア情報が多く行き交い³⁾、緊急時の情報伝達手段として機能し得ることが十分に証明されたためである。特に、Twitter は震災直後に広く利用された⁴⁾。Twitter とは、Tweet（つぶ

やぎ）と称される 140 字以内の短文を投稿することができる情報サービスである。他のユーザーはフォローすることでそれを閲覧できるだけでなく、140 字以内でコメントをつけることができ、これを繰り返すことでコミュニケーションをとる。また、他のユーザーの Tweet を ReTweet することで自分のフォロワー、すなわち自分の Tweet を見ることができるユーザーにも知らせることができる。リアルタイムでのメッセージとしてはチャットにも近く、速報的な情報のやり取りに特化したツールである。しかし一方で、情報発信源が不特定多数で、誰にでも手軽に情報を発信することが可能であることは、不確実性を孕んだ情報の拡散を誘発する要因にもなり得るという欠点も意味する。すなわち、誤った情報が拡散されやすく、情報の信頼性が必ずしも担保されるとは限らない。多くの論者が指摘するように、震災直後の Twitter はデマや流言が広まる母体ともなった。

本研究では、災害が起こった場合、ソーシャルメディア情報を適切に扱うことで、人々の混乱を避けるような誘導策の検討を行うことを目的とし、その第一段階として、災害時のソーシャルメディア情報と、人々の行動の関連性を明らかにすることを目的とする。具体的には、複数ある SNS の中でも、速報な情報のやり取りに特化したツールである Twitter に着目し、交通機関が麻痺したことで、日常的に利用している帰宅ルート

を利用できない人（以下、帰宅困難者）が多く発生したと考えられる東日本大震災時の東京首都圏を対象に、震災後に発信された Tweet のテキスト情報と災害時の行動データの分析を行い、ソーシャルメディア情報と人々の行動の関連性を明らかにする。

2. 既往研究と本研究の位置づけ

(1) ソーシャルメディアに関する既往研究

東日本大震災発生時、テレビやラジオといったマスメディアが災害に関する情報を多数報道する中、ソーシャルメディアによる情報のやり取りも活発に行われた⁴⁾。ソーシャルメディアでは、被災者自身が情報の発信源となりえるため、実際の体験に基づくリアルタイムな情報のやり取りが行われる。すなわち、既存のテレビやラジオに比べて、ソーシャルメディアによる情報伝達には、臨場感と即時性がある。そのため、近年、災害時の情報伝搬ツールとしての有用性について活発に議論されるようになった⁵⁾。さらに、ソーシャルメディアによる情報伝達には、情報伝播力が強いという特徴がある。それゆえ正しい情報だけでなく誤った情報（デマ）も拡散されてしまうという脆弱性がある⁶⁾。東日本大震災時にも、ソーシャルメディア内では、「コスモ石油のタンクが爆発し、有害物質が雨に溶けて降ってくる」といった誤った情報が出回った⁷⁾。このように、災害時のソーシャルメディアの情報には、臨場感と即時性などの有用性と、デマが拡散するなどの脆弱性があり、既往研究はそれら利点と欠点に着目した研究に大別できる。

ソーシャルメディアの有用性に関する既往研究として、鳥海ら⁸⁾は、震災時の Twitter において、同じ人が ReTweet した Tweet は類似した情報を持っていると仮定した上で、ReTweet したユーザーの重複度から Tweet をクラスタリングした。ReTweet とは、別の人の Tweet を自分のフォロワー、すなわち自分の Tweet を見ることができるユーザーにも知らせるときに用いる機能である。そして、クラスタリングされたそれぞれの Tweet 内容から、震災時に共有された情報の種類の把握を行った。その結果、震災時の Tweet は、震災直後の情報及び注意喚起、福島第一原発事故及び節電に関する情報、支援情報、話題提供型の 4 つのカテゴリーに分けられることができ、震災時の Twitter 上では ReTweet という優れた情報伝達機能を使うことによって、マスメディアなどの情報を仲介する役割と、個人が情報を発信する媒体としての役割の双方を担っていることを明らかにした。

一方、ソーシャルメディアの脆弱性に関する研究として、白井ら⁹⁾は、Twitter を対象とした情報伝搬シミュ

レーションを行い、情報ネットワークの変化によるデマ情報の拡散への影響を分析した。情報の拡散モデルには、SIR モデル（感染症流行モデル）を拡張したモデルを構築し適用した。具体的には、関連 Tweet によってコミュニケーションをとったユーザー同士をリンクで結ぶことによって、関連ネットワークと呼ばれるソーシャルネットワークを生成し、拡張した SIR モデルを用いた情報伝搬シミュレーションを行い、デマを否定する情報を流すことでデマ情報の拡散が抑えられることを明らかにした。

(2) テキストマイニングに関する既往研究

近年のソーシャルメディアの普及拡大によって各個人の情報の発信の自由度が大きくなった。それに伴い、テキスト形式のデータが量的に急増している。テキストマイニングに関する研究は、輻輳状態になっている膨大な量のテキストデータの中から有益な知見や情報を得ることを目的とし、主に自然言語処理分野で行われている^{10)~12)}。

風間ら¹³⁾は、Twitter の Tweet に出現する単語の出現頻度の時間的変化の類似性を、EMD (Earth Mover's Distance) を用いて判定することで、単語間の関連性を分析する手法を提案した。また、震災時の Tweet を対象にテキストマイニングを行い、「地震」「原発」の 2 つの単語とその関連名詞及びその単語間の関係の時系列変化を分析し、実際の事件などが Twitter に与える影響を明らかにした。笹原ら¹⁴⁾は、実世界において大きなイベントが生じると Tweet のバースト的な増加や不安定な振動が生じることに着目し、Tweet の定常状態と非定常状態の差を Jensen-Shannon ダイバージェンスで定量化してユーザーが大きく反応したイベントを検出した。さらに、2010 年と 2011 年に発信された約 5 億件の Tweet を対象にテキストマイニングを行い、単語ごとの頻度と頻度増加率を求め、検出された事象と対応するイベントを同定することで、著者らが提案した Jensen-Shannon ダイバージェンスを用いた手法の検出力の有効性を確認した。

(3) 本研究の位置づけ

(1) 節で述べたように、ソーシャルメディアには有用性と脆弱性という特徴が対立している。東日本大震災以降、この問題に関する研究は多くなされており、ソーシャルメディアの有用性、脆弱性について調査はされている。しかし、いずれの研究も、現状把握にとどまっており、将来同様の災害が起こった時の対策について議論するまでには至っていない。一方で、ソーシャルメディアの多くはテキスト形式で情報が蓄積されている。そのため、いずれ来たる災害時に備えるため、膨

大なテキストデータから減災のための有益な情報を取り出す必要があり、これはテキストマイニング技術を用いることで可能となると考えられる。

震災時にソーシャルメディアが活発に利用されたことが確認されており、ソーシャルメディア情報を受け取った人のその後の行動に影響を及ぼしているとも十分に考えられる。しかし、災害時のソーシャルメディア情報を対象としたテキストマイニングを行った分析はなされていても、ソーシャルメディアのテキストデータが人々の行動への影響に着目した研究はなされていない。本研究では災害時のソーシャルメディアのテキストデータをテキストマイニングによって分析し、災害時のソーシャルメディア情報と、帰宅困難者の行動の関係性を把握することを目的とする。

3. 使用データの概要

(1) ソーシャルメディア情報に関するデータ

本研究では、東日本大震災時に発信された Twitter のテキストデータを用いる。Twitter のテキストデータは全国各地で発言されたものが、どこで発言されたかという地理的属性を有さずに蓄積されている。そのため、本研究が対象とする帰宅困難者の分析には、何らかの方法で災害発生時に対象地域である東京首都圏で発言されたもののみを抽出する必要がある。そこで、当時の状況から公共交通の機能が麻痺したことで、多くの帰宅困難者で溢れたと予想される「駒沢通り」「目黒通り」および「六本木通り」をキーワードとする Tweet データを抽出することとした。データの詳細は、表-1 に示すとおりである。「通り名」でデータを抽出しているため、当時の帰宅困難者すべての発言を抽出されているとはいえず、また抽出された発言はすべて帰宅困難者によるものであるとは限らない点に注意が必要である。

表-1 の Twitter テキストデータから、1 時間当たりの Tweet 数の推移を図-1 に示す。地震発生の前後 5 時間を比較すると Tweet 数は約 6.7 倍に増加していた。このことから、震災直後に Twitter が活発に利用されたことがわかる。

(2) 帰宅行動に関するデータ

本研究では、帰宅行動に関するデータとして東日本大震災時の表-2 に示すような人口流動メッシュデータを用いる。(1) 節で述べたように、Twitter のテキストデータは「駒沢通り」「目黒通り」「六本木通り」の 3 つの「通り名」で抽出した。帰宅行動データは、3 つの通りと地理的に近い場所にある東横・目黒沿線に着目し、関東圏在住で、震災発生の約 1 週間前の 2011 年 3 月 4 日、5 日に東横、目黒沿線を 1 回以上利用した人のみを抽出

表-1 Twitter テキストデータ

対象	「駒沢通り」「目黒通り」「六本木通り」をキーワードとする Tweet を 1 回以上発信したユーザーの期間中の全 Tweet
期間	2011 年 3 月 11 日 9 時から 2011 年 3 月 12 日 23 時まで
アカウント数	1,182 (人)
Tweet 数	37,382 (件)

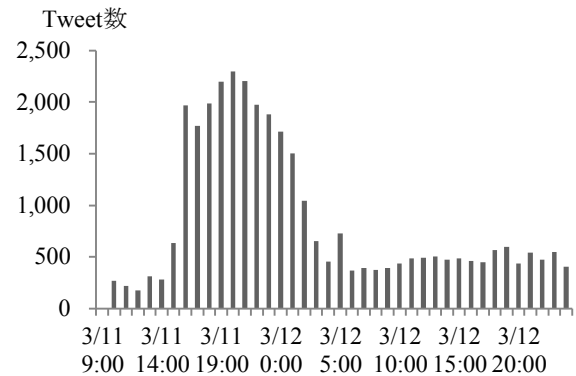


図-1 1 時間当たりの Tweet 数の推移

表-2 人口流動メッシュデータ

対象者	関東圏居住者
対象期間	2011 年 3 月 11 日 14 時から 2011 年 3 月 12 日 23 時まで
抽出方法	2011 年 3 月 4 日、5 日に 1 回以上東横・目黒沿線を利用した人
メッシュ間隔	250m
総メッシュ数	86,570 メッシュ
サンプル数	1,973
データ提供元	株式会社ゼンリンデータコム

した。東横、目黒沿線利用者に限定することで、震災による公共交通の麻痺によって徒歩で帰宅を検討する際、3 つのいずれかの通りを利用する可能性が高いサンプルの行動データを分析対象としており、旅行や出張で来訪者、あるいは、3 つの通りとは馴染みが薄いと考えられる人はサンプルから除外している。

4. テキストマイニングによる Tweet 内容分析

テキストマイニングとは、テキスト情報を定量的に分析する手法である。本研究では、形態素分析、共起分析を行い、1 時間ごとの単語間の関連性をネットワーク図（以降、単語間ネットワーク）として可視化したうえで、その時間的推移を把握する。

(1) 形態素分析

形態素分析とは、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である。ここで、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

表-3 不要語の例

助動詞	「だ」「ます」「です」
記号	「@」「。」「RT」
アカウント名	「aaaaa」

表-4 類義語結合の例

代表語	類義語
分かる	分かる, わかる, 理解する
携帯電話	携帯電話, 携帯, スマホ
帰る	帰る, 帰宅する, 帰宅
面白い	面白い, おもしろい, おもしろ
Twitter	Twitter, TWITTER, twitter, ツイッター
混む	混む, 混雑

表-5 各段階における単語の種類数

	単語の種類数 (割合)
全形態素	38,416 (100%)
品詞で選別	33,521 (87.26%)
不要語削除	16,546 (43.07%)
類義語結合	15,677 (40.81%)
出現頻度 10 回以上	2,392 (6.22%)

データを形態素に分解し、品詞による選別、不要語の削除、類義語の結合を行う。品詞による選別では、文の意味を直接表すような動詞・名詞・形容詞と動作の具体性を補う助動詞を抽出する。不要語の削除では、文書中で多く使用されているが、意味の把握を行う上であまり有用ではないと分析者が判断した語を除く。

不要語の削除では、表-3に示すような単独で意味を成さないアカウント名や記号などを対象とした。類義語の結合では、表-4に示すような似通った意味を持つ単語、あるいは、発音は同じだが表記が異なる単語を一つの代表語にまとめる。この操作により、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となる。各段階で抽出された単語の種類数を表-5に示す。

表-5に着目すると、類義語の結合を行った段階で全形態素数の約 40%となったが、このまま共起分析を行うには膨大な単語数である。そこで、単語自体の最低出現頻度 10 回という条件を新たに設定することで単語数を調整し、最終的に 2,392 語とした。あらかじめ類義語を結合しているため、最低出現頻度の設定によって、複数の表現方法が存在する同意語を過度に除外される可能性は低いと考えられる。

(2) 共起分析

テキストデータには否定の助動詞が含まれているため、単語のみに着目すると分析結果の解釈を誤る可能性がある。そこで、複数の単語に着目した共起分析を行う。共起分析では、対象とするテキストデータ全体で、同時に出現することが多い単語の組み合わせ（以降、共起パターン）を抽出する。ここで、共起するとは、ある単語が特定の単語を中心とする一定の範囲内に現れることを意味し、共起される単語を中心語、共起する単語を共起語と呼ぶ。本研究では、形態素分析で得られた各単語を中心語とし、各中心語と共起語の関係を分析する。なお、共起の範囲は中心語の後 3 語とし、この範囲で共起した回数を共起回数とする。ただし、共起が偶発的なものである可能性があるため、共起回数が統計的に有意であるかを検定する必要がある。本研究では t 検定を用いて、共起回数の統計的有意性を確認する。 t 検定の検定量である t 値は式 (1) で算出される。本研究では、 t 値の絶対値が 1.65 以上であればその共起が有意であるとする^{15), 16)}。

$$t = \frac{N_C - N_E}{\sqrt{N_C}} \quad (1)$$

ここに、 N_C ：共起回数、 N_E ：共起回数の期待値

全時点（3 月 11 日 15 時から 12 日 16 時まで）データを対象に、 t 値による共起判定の結果、8,113 組の有意な共起パターンが検出された。表-6に、有意な共起パターンの共起回数上位 50 組を示す。検出された共起パターンから多いため、この結果から直接、発言内容の特徴を把握することは困難である。そこで、次節以降では、単語間ネットワークの生成によって、視覚的に共起パターンを概観するとともに、その分析によって発言内容の時間的推移を明らかにする。

(3) 単語間ネットワーク

3.(1) 節で述べた Twitter テキストデータを 1 時間ごとに分割し、各々のテキストデータについて共起分析を行った。さらに、 (i, i) 成分は i 行目の単語の出現頻度を表し、 (i, j) 成分は i 行目の単語と j 列目の単語の共起回数を表すような行列を作成し、これを隣接行列とするネットワークで表現したものを単語間ネットワークと呼ぶ。本研究では、共起範囲を中心語の後方に限定しているため、単語間ネットワーク図を有向グラフで表示することができ、中心語と共起語の順序を把握することができる。また、単語間ネットワーク図におけるノードの大きさは単語の出現頻度、リンクの太さは共起頻度にそれぞれ比例する。以降は、紙面の都合上、全時点における単語間ネットワークを記載することが困難であるため、震災直後の単語間ネットワークのみを示し、考察する。

表-6 有意な共起パターン 50 組

中心語	共起語	共起回数	t 値	中心語	共起語	共起回数	t 値
歩く	帰る	388	18.5	つける	帰る	118	9.0
動く	ない	300	13.6	ない	思う	116	2.8
できる	ない	295	11.7	阪神	大震災	116	10.8
帰れる	ない	267	13.5	電話	繋がる	111	10.4
連絡	取れる	262	16.0	通り	歩く	111	6.0
ない	なる	252	12.8	確認	できる	109	9.9
連絡	ない	251	11.0	帰る	できる	109	8.5
ない	人	222	5.1	歩く	人	109	11.4
拡散	希望	220	14.8	東北	地方	106	10.3
なる	ない	217	12.8	公衆	電話	104	10.2
運転	再開	210	14.4	出る	ない	104	4.6
連絡	とれる	208	14.3	人	ない	104	5.1
わかる	ない	205	12.2	被害	ない	104	7.1
安否	確認	190	13.8	避難	場所	104	10.3
お気	つける	180	13.3	武蔵	小杉	97	9.8
通り	渋滞	167	4.5	電車	止まる	95	9.5
帰る	ない	166	2.1	寝る	ない	92	6.5
地震	速報	166	12.9	電車	ない	92	5.4
繋がる	ない	163	10.7	とれる	ない	90	7.4
電話	ない	160	8.5	渋滞	中	87	8.5
余震	続く	158	12.5	しれる	ない	86	9.0
電車	動く	142	11.6	やる	ない	86	5.2
山手	通り	136	11.8	使える	ない	85	8.0
人	多い	124	11.4	ガス	止まる	84	9.1
取れる	ない	121	8.8	人	歩く	82	11.4

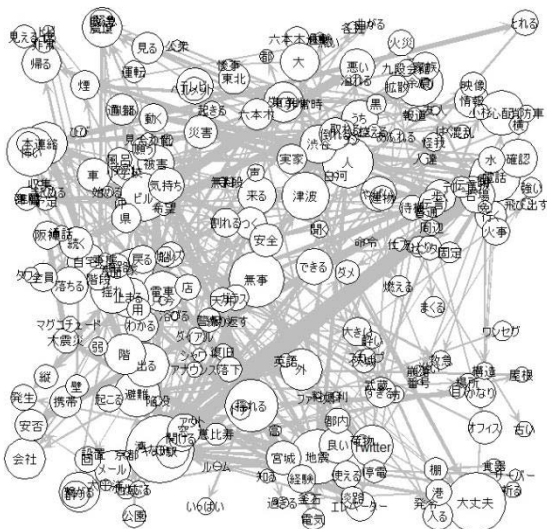


図-2 3月11日15時から16時の単語間ネットワーク

3月11日の15時から19時までの時点における単語間ネットワークを図-2から図-5に示す。形態素分析の段階で単語数(ノード数)を2,392語まで削減した。しかし、図-2から図-5に着目すると、ネットワーク図は

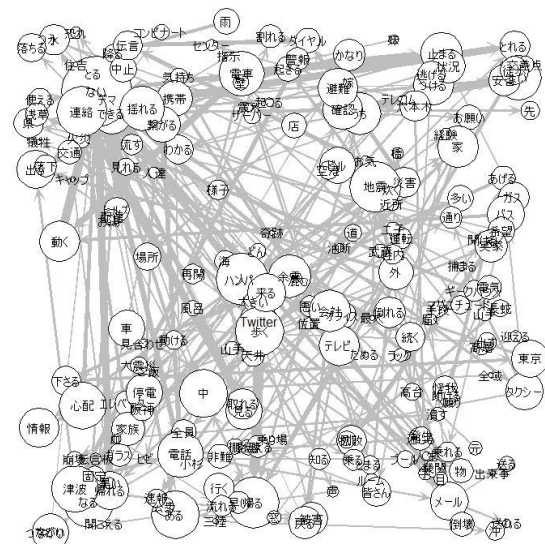


図-3 3月11日16時から17時の単語間ネットワーク

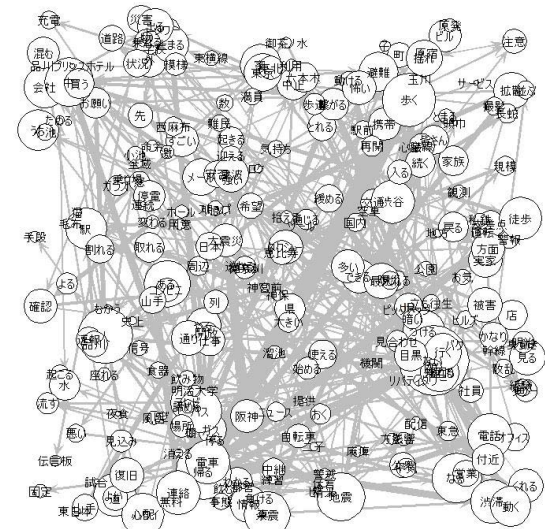


図-4 3月11日17時から18時の単語間ネットワーク

非常に煩雑であり、ネットワーク図から時点間推移を捉えることは困難である。そこで、時点間推移を概括的に把握するために各ネットワーク中でノードが大きいもの、すなわち出現頻度が高い単語上位10位までを表-7に示した。表-7に着目すると、15時から17時にかけては「地震」「揺れる」「余震」「津波」といった地震を連想させる単語が多く発言されているが、17時以降は前述のような地震に関連する単語が減少していることが分かる。また、地震発生から1時間後の16時になると「帰る」という単語が多く発言されるようになり、その後も同じ状態が続いている。15時から18時にかけて、「電話」「連絡」という単語が発言されているが、出現頻度は徐々に減少している。このことから、Twitterユーザーは地震発生直後に家族や友人の安否確認を積極的に行い、その後帰宅するか否かを決定する傾向があったのではないかと考えられる。

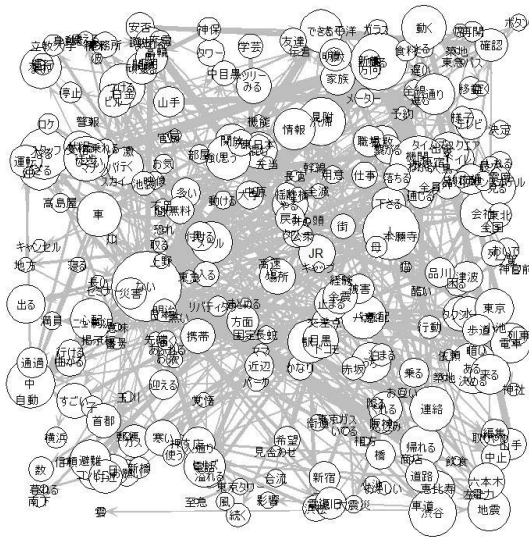


図-5 3月11日18時から19時の単語間ネットワーク

表-7 各ネットワーク中の出現頻度が高い単語

15時-16時	16時-17時	17時-18時	18時-19時
ない 495	ない 607	ない 679	ない 728
地震 168	人 135	人 205	帰る 251
揺れる 144	帰る 127	帰る 200	人 249
人 124	地震 124	歩く 171	歩く 249
電話 121	なる 119	なる 127	ある 147
なる 106	連絡 117	中 116	なる 146
中 106	歩く 102	地震 113	中 124
怖い 105	中 98	ある 97	渋滞 120
余震 101	ある 96	余震 94	渋谷 112
ビル 95	津波 89	電話 84	会社 99

(4) 単語間ネットワークの分析

(3)節ではネットワーク図の時点間推移を概括的に考察したが、単語間ネットワークが複雑なため十分にその特徴を把握することができない。本節では、複雑な単語間ネットワークの情報を集約し、ネットワーク類似性、およびネットワーク中心性の指標を用いることで、発言内容の時間的推移を定量的に分析する。

a) ネットワーク類似性

時点間で、Twitterの発言内容に関連性があるかを分析するために、ここでは、単語間ネットワークの類似性分析を行う。類似性、すなわち相関係数が高ければ、時点間で同じような話題が多くなされたことを、相関係数が低ければ異なる話題についての発言が多くなされたことを意味する。分析対象は、単語数が他時点と比べて非常に少ない3月12日0時以降を除く、2011年3月11日の15時から24時までの期間とし、当該期間

表-8 相関行列

	15時	16時	17時	18時	19時	20時	21時	22時	23時
15時	-	0.596	0.416	0.304	0.217	0.213	0.219	0.179	0.272
16時		-	0.617	0.562	0.479	0.440	0.429	0.375	0.464
17時			-	0.725	0.653	0.574	0.572	0.466	0.538
18時				-	0.760	0.689	0.657	0.589	0.609
19時					-	0.733	0.658	0.596	0.641
20時						-	0.794	0.724	0.718
21時							-	0.717	0.699
22時								-	0.709
23時									-

相関係数 0.5 以上

相関係数 0.7 以上

を1時間ごとに分割した全9時点とする。共起行列 X と共起行列 Y の間の相関係数は、式(2)によって求められる。

$$r_{XY} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2} \sqrt{\sum_i^n (y_i - \bar{y})^2}} \quad (2)$$

ここに、 x_i, y_i : 列ベクトル x, y の i 行目成分、 \bar{x}, \bar{y} : 列ベクトル x, y の相対平均、 n : 列ベクトル x, y の長さ

式(2)で算出した時点間の単語ネットワーク相関係数を表-8に示す。本研究では、相関係数が0.7以上のとき単語間ネットワークに強い類似性がある、0.5以上のとき類似性があると判定する。

表-8に着目すると、どの時点においてもその直後の時点との相関は高い。相関係数0.5以上の類似性があると判定された時間帯に着目すると、15時から16時までの時点に対しては1時間後、16時から17時までの時点に対しては2時間後、17時から18時までの時点に対しては4時間後までそれぞれ相関が高い。また、19時以降においては、その後の時点と0.5以上の相関がある。以上のことから、地震発生から時間が経過するほど、その後の単語間ネットワークとの類似性があり、引き続き同じ話題が発言される傾向があることが分かった。言い換えると、地震発生直後の1~2時間は状況の変化が目まぐるしく、それに伴って話題が急速に変化していると言える。また、20時から24時までの間では相関係数が0.7以上と強い類似性を示しており、20時以降は発言内容の変化が鈍化したことがわかる。

b) ネットワーク中心性

ネットワークの類似性分析から、3月11日の14時から24時までの間に話題が変化した可能性を見いだした。そこで、単語間ネットワーク図の内容から時点間推移を詳細に捉えるため、a)項で挙げた全9時点における単語間ネットワーク図に含まれる単語についてそれぞれ中心性を求め、上位10位までを抽出し、全時点間で比較する。ネットワーク中心性とは、ネットワー

表-9 各時点における次数中心性の上位 10 単語 (カッコ内の値は次数中心性)

	15 時-16 時	16 時-17 時	17 時-18 時	18 時-19 時	19 時-20 時	20 時-21 時	21 時-22 時	22 時-23 時	23 時-24 時
1	ない(294)	ない(324)	ない(299)	ない(316)	ない(306)	ない(261)	ない(189)	ない(155)	ない(151)
2	電話(199)	連絡(176)	帰る(119)	人(164)	人(255)	人(202)	運転(169)	帰る(139)	無事(108)
3	連絡(106)	電話(85)	歩く(118)	歩く(143)	帰る(175)	再開(186)	再開(137)	運転(110)	再開(77)
4	繋がる(92)	人(84)	人(109)	動く(140)	渋滞(129)	歩く(137)	帰る(122)	再開(102)	帰る(71)
5	人(80)	動く(82)	電話(106)	帰る(120)	歩く(124)	運転(131)	歩く(111)	人(84)	渋滞(66)
6	津波(79)	電車(75)	電車(97)	連絡(105)	地震(113)	動く(124)	無事(100)	渋滞(77)	連絡(65)
7	地震(68)	帰る(74)	連絡(96)	渋滞(104)	動く(107)	帰る(122)	動く(89)	無事(73)	運転(61)
8	災害(58)	地震(60)	地震(82)	通り(93)	大(102)	連絡(95)	九段下(89)	時間(63)	時間(60)
9	通じる(56)	できる(60)	渋滞(82)	駅(83)	通り(94)	渋滞(94)	押上(83)	動く(58)	動く(56)
10	確認(55)	歩く(54)	バス(81)	駒沢通り(80)	駅(87)	駅(94)	全線(79)	駒沢通り(57)	全線(47)

ク分析において用いられる指標で、ネットワークを構成する各頂点とそのネットワークにおいてどの程度中心的であるかを示す指標である。本研究では、複数あるネットワーク中心性指標の中から次数中心性¹⁷⁾を用いる。次数中心性とは、ネットワーク内でより多くの頂点と関係をもつ頂点を高く評価する指標であり、次数とは頂点に接続している辺の数を意味する¹⁷⁾。次数中心性は、有向グラフの場合、他の頂点から入ってくる辺の数である入次数と他の頂点へ出ていく辺の数である出次数の和から求められ、式(3)のように表される。本研究の場合、より多くの単語と共に起していれば、高い次数中心性を示す。

$$C_d = C_{id} + C_{od} \quad (3)$$

ここに、 C_d : 次数中心性, C_{id} : 入次数, C_{od} : 出次数

各時点における次数中心性の上位 10 位までを抽出した結果を表-9 に示す。14 時から 17 時に着目すると、「津波」「地震」といった災害に関する単語が上位に多く散見されるが、その後は少ない。これは、震災発生直後は津波や余震に対しての警戒が強かったためであると考えられる。一方で、17 時以降に着目すると、「駒沢通り」といった具体的な地名や「渋滞」「運転」といった交通状況に関する単語が上位を占めている。これは、交通機関の現状が被害状況を知る一つの指標となりえること、また、被災者にとって帰宅の可否が重大な関心事であることから交通に関する情報の収集および発信が盛んになされたためと考えられる。また、全時点において否定の助動詞「ない」が上位に抽出されており、地震発生後ネガティブな発言が続いていたことが推測される。

5. Twitter 情報と帰宅行動の関連性分析

本章では、正準相関分析を用いて 4 章で示した単語間ネットワークと災害時の帰宅行動の相関分析を行う。そして、得られた結果をもとに考察する。

Twitter のテキスト情報には、第 4 章で示した通り、多

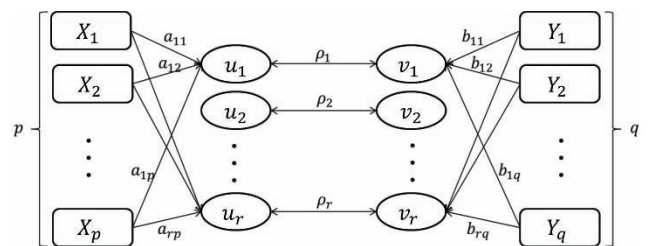


図-6 正準相関分析の概要図

様な発言内容が含まれる。また、帰宅行動も、移動を行う、勤務地に留まる、自宅に到着するなど、複数の行動要素で構成されている。そのため、1 つの変数 (目的変数) に対する多数の変数 (説明変数) の影響を分析する重回帰分析は、本研究の分析方法には適していない。そこで本研究では、重回帰分析の一般形であり、多数の変数が 2 つの変数群を構成するときに、変数群間の相互関係を包括的に分析することができる正準相関分析を適用する。

(1) 正準相関分析

正準相関分析とは、2 つの変数群の間の相関関係を把握するための手法である (図-6)。

2 つの変数群 $\{X_1, X_2, \dots, X_p\}$, $\{Y_1, Y_2, \dots, Y_q\}$ があるとき、各変数群の線形合成変量の間の相関係数 ρ_i ($i = 1, 2, \dots, r$; $r = \min(p, q)$) が最大になるような任意係数 a, b を求めることを正準相関分析といい、このときの相関係数 ρ_i を正準相関係数という。また、変数群の合成変量 u_i, v_i のことを正準変量という。ここで、2 つの変数群 X, Y についての分散共分散行列は、式 (4) のように表される。

$$S = \begin{pmatrix} S_{XX} & S_{XY} \\ S_{YX} & S_{YY} \end{pmatrix} \quad (4)$$

各変数群の正準変量は、任意の係数

$a_1 = \{a_{11}, a_{12}, \dots, a_{1p}\}$, $b_1 = \{b_{11}, b_{12}, \dots, b_{1q}\}$ を用いて、式 (5)、式 (6) のように表される。

$$u_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \quad (5)$$

$$v_1 = b_{11}Y_1 + b_{12}Y_2 + \cdots + b_{1q}Y_q \quad (6)$$

ここで、 u_1 と v_1 の間の相関係数 ρ_1 が最大になるような係数 $\mathbf{a}_1, \mathbf{b}_1$ を推定する。このときの係数を正準負荷係数という。 u_1 と v_1 の間の相関係数は、それぞれを定数倍した $C_1^{(1)}u_1$ と $C_1^{(2)}v_1$ の間の相関係数に等しいので、分散が 1 となるように標準化しても一般性は失われない。したがって、式 (7)、式 (8) のような制約条件のもとで、 u_1 と v_1 は標準化できる。

$$\text{Var}(u_1) = \mathbf{a}_1^T S_{XX} \mathbf{a}_1 = 1 \quad (7)$$

$$\text{Var}(v_1) = \mathbf{b}_1^T S_{YY} \mathbf{b}_1 = 1 \quad (8)$$

このとき、相関係数の分母は 1 となり、式 (9) のような関係式が得られる。すなわち、式 (9) において、正準変量 u_1 と v_1 の間の相関係数 ρ_1 を最大にするような $\mathbf{a}_1, \mathbf{b}_1$ を求める。

$$\rho_1 = \text{Cov}(u, v) = \mathbf{a}_1^T S_{XY} \mathbf{b}_1 \quad (9)$$

ラグランジュ未定乗数を λ_1, μ_1 とし、式 (9) のようなラグランジュ関数を設定する。

$$L = \mathbf{a}_1^T S_{XY} \mathbf{b}_1 - \frac{\lambda_1}{2} (\mathbf{a}_1^T S_{XX} \mathbf{a}_1 - 1) - \frac{\mu_1}{2} (\mathbf{b}_1^T S_{YY} \mathbf{b}_1 - 1) \quad (10)$$

これより、 $\mathbf{a}_1, \mathbf{b}_1, \lambda_1, \mu_1$ についての最大化問題に帰着する。式 (10) の両辺を $\mathbf{a}_1, \mathbf{b}_1$ で偏微分して 0 とおくと、次式が成り立つ。

$$\frac{\partial L}{\partial \mathbf{a}_1} = S_{XY} \mathbf{b}_1 - \lambda_1 S_{XX} \mathbf{a}_1 = 0 \quad (11)$$

$$\frac{\partial L}{\partial \mathbf{b}_1} = S_{YX} \mathbf{a}_1 - \mu_1 S_{YY} \mathbf{b}_1 = 0 \quad (12)$$

式 (11)、式 (12) の左からそれぞれ $\mathbf{a}_1^T, \mathbf{b}_1^T$ をかけ、式 (7)、式 (8) を利用すると、式 (13) のような関係が成り立つ。

$$\lambda_1 = \mu_1 = \mathbf{a}_1^T S_{XY} \mathbf{b}_1 \quad (13)$$

式 (11)、式 (12)、式 (13) より、整理すると、式 (14)、式 (15) のような関係式が得られる。

$$(S_{XY} S_{YY}^{-1} S_{YX} - \lambda_1^2 S_{XX}) \mathbf{a}_1 = 0 \quad (14)$$

$$(S_{XY} S_{XX}^{-1} S_{YX} - \mu_1^2 S_{YY}) \mathbf{b}_1 = 0 \quad (15)$$

式 (14)、式 (15) より、固有値 λ_1^2 に対応する固有ベクトル $\mathbf{a}_1, \mathbf{b}_1$ を用いて線形結合を構成すれば、相関係数 ρ_1 が最大になると分かる。また、このときの相関係数 ρ_1 を第 1 正準相関係数といい、式 (13) より固有値の平方根で求めることができる。

同様に、2 つの変数群 X, Y に対して、任意の係数 $\mathbf{a}_2 = \{a_{21}, a_{22}, \dots, a_{2p}\}, \mathbf{b}_2 = \{b_{21}, b_{22}, \dots, b_{2q}\}$ を用いると、式 (16)、式 (17) のような線形結合を構成できる。

$$u_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \quad (16)$$

$$v_2 = b_{21}Y_1 + b_{22}Y_2 + \cdots + b_{2q}Y_q \quad (17)$$

ただし、 u_2, v_2 は u_1, v_1 とは無相関という条件のもと u_2 と v_2 の間の相関係数 ρ_2 が最大になるように係数を求める。同様の最大化問題を解き、固有値 λ_2^2 を得る。

以下、同様の手順を $r = \min(p, q)$ 回繰り返すことで、大きい順に固有値 $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_r^2$ の平方根が正準相関係数 $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_r$ を与え、それぞれに対応する固有ベクトルを用いて正準変量を構成する。ここで、有効な正準相関係数の個数を決めるために χ^2 検定を行う。この場合、 i 番目までの正準相関係数が 0 でないとし、以下のような帰無仮説 H_0 を設定し検定する。

$$H_0 : \rho_1 \geq \rho_2 \geq \cdots \geq \rho_i \geq \rho_{i+1} = 0 \quad (18)$$

また、 χ^2 統計量は自由度 $(p-r)(q-r)$ の χ^2 分布に従うことを利用して、式 (19) で表される。

$$\chi^2 = -\left(n - 1 - \frac{p+q+1}{2}\right) \sum_{j=i+1}^r \log(1 - \lambda_j^2) \quad (19)$$

ここで、 n はサンプル数を表す。 $\chi^2 \geq \chi_{\alpha}^2((p-r)(q-r))$ ならば、有意水準 α で帰無仮説は棄却され、第 $i+1$ 正準相関係数は有効な値であると判断できる。すべての正準相関係数について χ^2 検定行ったらうで、正準変量の係数について考察する。

(2) 単語間ネットワークと帰宅行動の相関分析

本節では、正準相関分析を用いて、Twitter 発言内容と帰宅行動の関係性を明らかにする。具体的には、変数群 X を共起分析で得られた単語間の共起回数とし、変数群 Y を 3.(2) 節で述べた流動人口、自宅該当者、勤務地該当者に関する全メッシュの平均移動者割合の変化率とする。ただし、統計的有意性が高い共起回数の時間的分布を用いるために、2011 年 3 月 11 日 15 時から 12 日 16 時までを 1 時間ごとに区切った全 25 時点を用いる。共起回数については、表-6 に示した共起パターンの中から交通に関する共起パターンの上位 10 組を選定した。選定した共起パターンは表-10 に示す。変数群 X を 10 組の共起パターンの共起回数分布、変数群 Y を流動人口、自宅該当者、勤務地該当者に関する全メッ

表-10 正準相関分析で用いる共起パターン

	中心語	共起語
X_1	動く	ない
X_2	歩く	帰る
X_3	運転	再開
X_4	帰れる	ない
X_5	通り	渋滞
X_6	電車	動く
X_7	地震	速報
X_8	通り	歩く
X_9	人	多い
X_{10}	電車	止まる

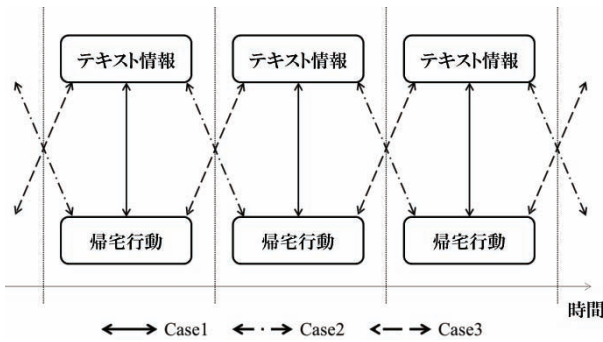


図-7 各 Case の相関関係の概要図

表-11 正準相関係数 (Case 1)

	正準相関係数	寄与率	累積寄与率	χ^2 値
ρ_1	0.598	69.85(%)	69.85(%)	52306**
ρ_2	0.346	23.35(%)	93.20(%)	14064**
ρ_3	0.186	6.80(%)	100.00(%)	3060**

(** : 1%有意)

シュの平均移動者割合の変化率とし、正準相関分析を行う。ここで、Twitter のテキスト情報と帰宅行動の関係性には、以下の3つが考えられる(図-7)。

- 1) Twitter のテキスト情報が帰宅行動に、あるいは帰宅行動が Twitter のテキスト情報に、瞬時に影響を及ぼす。このとき、同時刻の Twitter のテキスト情報と帰宅行動に関係性がある (Case 1)。
- 2) Twitter のテキスト情報の影響を受けて、帰宅行動が変化する。このとき、帰宅行動は1時点前の Twitter のテキスト情報と関係性がある (Case 2)。
- 3) 帰宅行動の結果を受けて、Twitter のテキスト情報が変化する。このとき、Twitter のテキスト情報は1時点前の帰宅行動と関係性がある (Case 3)。

正準相関分析は、因果を特定するものではないが、本研究では時点を考慮した3つのケースについてモデル分析を行い、モデル適合度を比較することによって、Twitter 情報と帰宅行動の時間的関係性を明らかにする。また、モデル適合度指標として AIC(Akaike's Information Criterion) を用いることとする。AIC とは、統計的モデリングにおいてバイアスを少なくするだけでなく、バイアスとモデルの単純さのバランスを評価する指標である¹⁸⁾。AIC は式(20)によって算出される。式(20)より、AIC が低いモデルほど適合度の高いモデルと言える。

$$AIC = -2 \ln L + 2k \quad (20)$$

ここに、 L : 最大尤度、 k : パラメータの個数

3つのケースごとに正準相関分析を行い、各ケースにおける正準相関係数の算出結果を表-11 から表-13 に、AIC を算出した結果を表-14 に示す。表-14 より、Case 2 の AIC が最も低く、次いで Case 1、Case 3 となった。すなわち、テキスト情報と1時間後の帰宅行動の間の最

表-12 正準相関係数 (Case 2)

	正準相関係数	寄与率	累積寄与率	χ^2 値
ρ_1	0.576	46.54(%)	46.54(%)	73582**
ρ_2	0.560	44.02(%)	90.56(%)	38654**
ρ_3	0.259	9.44(%)	100.00(%)	6027**

(** : 1%有意)

表-13 正準相関係数 (Case 3)

	正準相関係数	寄与率	累積寄与率	χ^2 値
ρ_1	0.646	60.55(%)	60.55(%)	72982**
ρ_2	0.472	32.34(%)	92.89(%)	26190**
ρ_3	0.221	7.11(%)	100.00(%)	4351**

(** : 1%有意)

表-14 各 Case におけるモデル適合度

	AIC
Case1	135.5380
Case2	122.4695
Case3	135.2035

表-15 各変数群に関する正準負荷係数 (Case 2)

変数群 X	u_1	u_2	u_3
X_1 : 動く-ない	1.869	-0.692	1.720
X_2 : 歩く-帰る	-0.502	3.468	0.680
X_3 : 運転-再開	-0.599	-0.208	-0.390
X_4 : 帰れる-ない	-2.241	-3.084	0.027
X_5 : 通り-渋滞	0.683	1.652	0.818
X_6 : 電車-動く	-1.476	-2.048	-0.984
X_7 : 地震-速報	0.295	-0.222	0.012
X_8 : 通り-歩く	0.340	0.788	-1.089
X_9 : 人-多い	0.752	-1.101	-1.065
X_{10} : 電車-止まる	1.591	1.322	-0.278
変数群 Y	v_1	v_2	v_3
Y_1 : 流動人口	-0.050	-0.833	-0.715
Y_2 : 自宅該当者数	-1.371	0.127	-0.016
Y_3 : 勤務地該当者数	1.033	0.906	-0.451

も強い相関関係が存在することが明らかとなった。これは、テキスト情報が帰宅行動に影響を及ぼしている可能性を示唆する結果である。以下では、最もモデル適合度が高かった Case 2 についてのみ着目する。このときの正準負荷係数を表-15 に示す。

表-12 において、すべての正準相関係数が有意な値を示した。すなわち、交通に関して発言された内容と帰宅行動には相関関係があることを表している。また、各正準相関係数の寄与率に着目すると、第2正準相関係数までの累積寄与率が90%を超えており、第1正準相関係数と第2正準相関係数でテキスト情報と帰宅行動の相関関係の90%を説明できている。したがって、以下では、第1正準変量と第2正準変量についてのみ考

察する。

第1正準変量においては、表-15の変数群Yの正準負荷係数に着目すると、自宅該当者数が負の値かつ負荷量が最大となっており、これが変数群Yにおいて第1正準変量を代表する変数であると言える。ここで、表-15の変数群Xの正準負荷係数に着目すると、「動く-ない」「電車-止まる」「通り-渋滞」「人-多い」の負荷量が正の値となった。一方で、「電車-動く」「運転-再開」の負荷量が負の値となった。このことから、交通状況に関して悲観的な情報が多いとき、あるいは楽観的な情報が少ないとき、自宅に帰る人の割合は低いということが分かった。また、「地震-速報」の負荷量が正の値となったことから、地震速報が多いとき自宅に帰る人が少ないと分かった。これは、地震情報が多く出回ると安全のためその場にとどまる人が多いからではないかと考えられる。「歩く-帰る」の負荷量は負の値を示しているため、帰宅に関する情報が少ないとき、自宅に帰る人の割合が低いと分かった。これは、Twitterユーザーが公共交通の麻痺を予想し帰宅を断念したからではないかと考えられる。

第2正準変量においては、表-15の変数群Yの正準負荷係数に着目すると、流動人口が負、勤務地該当者数が正の値を示している。さらに負荷量が二極化していることから、これらが変数群Yにおいて第2正準変量を代表する変数であると言える。ここで、表-15の変数群Xの正準負荷係数に着目すると、「電車-止まる」「通り-渋滞」「通り-歩く」の負荷量が正の値となり、「運転-再開」「電車-動く」の負荷量が負の値となった。これは、交通状況に関して悲観的な情報が多いとき、もしくは楽観的な情報が少ないとき、出歩かずに勤務地に滞在する人の割合が高いことを示している。また、「地震-速報」の負荷量が負であり、地震に関する情報が少ないときも無理に出歩かずに勤務地に滞在する人の割合が高いことを示している。一方、「歩く-帰る」「通り-歩く」の負荷量は正の値となっていることから、公共交通が機能せず帰宅困難を強いられることを連想する情報が多いときもまた、出歩かずに勤務地に滞在する人の割合が高いと分かった。「帰れる-ない」「動く-ない」の負荷量は負の値となった。このことから、帰宅できるかどうかの判断材料としての情報が少ないため、状況把握ができずにその場に留まる人が多いのではないかと考えられる。

6. 結論

災害時という情報の信頼性が担保されない状況下で、人々がどの程度ソーシャルメディアによって発信された情報の影響を受けるのかを明らかにすることは、将

来、災害発生時におけるソーシャルメディアの活用方法を考える上で有効なことである。災害時のソーシャルメディアにおけるテキスト情報の影響を顕著に表すものが、災害時の帰宅困難者の行動であると考えられる。本研究では、災害時のソーシャルメディアと帰宅困難者の行動に着目し、二者の間の関連性を明らかにすることを目的とし、分析を行った。

東日本大震災時に発信されたTwitterのテキストデータを対象にテキストマイニングを行った。さらに、単語間ネットワークを一定時間ごとに示し、ネットワークの類似性と中心性を用いることで単語間ネットワークの時点間推移を分析した。

ネットワークの類似性と中心性の分析の結果、地震発生直後と地震発生からある程度時間が経過した後を比較すると、発言された話題に異なる傾向があることが明らかとなった。具体的には、地震発生直後では災害に関する話題が中心的であったが、地震発生から2〜3時間経過すると交通に関する話題が主となり、ここでの発言内容はその後の発言内容と高い類似性を示すものであった。

Twitterのテキスト情報と災害時の帰宅行動の関連性の分析には正準相関分析を用いた。ただし、正準相関分析では因果関係を特定することは難しい。そこで、Twitterのテキスト情報と帰宅行動の時間的関係性を明らかにするために、同時点のTwitterのテキスト情報と帰宅行動に関係性がある場合(Case 1)、帰宅行動は1時点前のTwitterのテキスト情報と関係性がある場合(Case 2)、Twitterのテキスト情報は1時点前の帰宅行動と関係性がある場合(Case 3)の3通りを想定し、それぞれのケースについて正準相関分析を行った。その結果を、モデル適合度を指標として比較したところ、Case 2が最も低い値となり、テキスト情報と帰宅行動の時間的関係が示唆された。また、正準相関分析の結果から、交通状況に関して悲観的な情報が多いとき、あるいは楽観的な情報が少ないとき、自宅に帰る人の割合は低いということが分かった。また、交通状況に関して悲観的な情報が多いとき、もしくは楽観的な情報が少ないとき、出歩かずに勤務地に滞在する人の割合が高いことが分かった。これらは、直感に即した結果であり、本分析手法の妥当性を示すものである。さらに、これらテキスト情報と帰宅行動の影響を定量的に把握できたことは本研究の成果であると言える。

以上の成果をまとめると、災害時のソーシャルメディアのテキスト情報と帰宅行動の間には相関があることを明らかにした。すなわち、災害時という情報の信頼性が担保されない状況下でも、ソーシャルメディアで発信された情報は人々の行動を決定する影響因子に成りえることを意味する。これは災害時にソーシャルメ

ディアの有効な活用によって、帰宅行動による過度の混雑を分散する、あるいは混乱を減少させることができる可能性を示唆するものでもある。

本研究では、ソーシャルメディアのテキスト情報と帰宅行動の関連性にのみ着目しており、災害時のソーシャルメディア情報の信頼性、妥当性は無視している。実際の公共交通の運行情報と比較することで、災害時のソーシャルメディアの情報信頼性を考慮した分析が必要である。また、本研究では、ソーシャルメディア情報と帰宅行動の相関分析を行ったが、両者の間には、行動結果がソーシャルメディア情報に及ぼす影響と、ソーシャルメディア情報が行動意思決定に及ぼす影響の双方方向因果が考えられる。この因果影響を分析するためには、対象とするソーシャルメディア情報の種類の拡張を行ったうえで、長期間継続的に取得したデータを対象とした新たな分析手法を開発する必要がある、今後の重要な課題である。

参考文献

- 1) 総務省：平成 24 年版情報通信白書，「1.2.3 デジタルネットワーク完成が導くメディア新展開」
- 2) 株式会社 LINE：プレリリース・お知らせ，http://about.naver.jp/press/press_detail?docId=1838，2013 年 11 月 22 日閲覧
- 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，2011。
- 4) 吉次由美：東日本大震災に見る大災害時のソーシャルメディアの役割，放送研究と調査，2011 年 7 月号，NHK 出版，2011。
- 5) 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，放送研究と調査，2011 年 8 月号，NHK 出版，2011。
- 6) 総務省：情報通信統計データベース平成 23 年度版情報通信白書，東日本大震災における情報通信の状況
- 7) コスモ石油株式会社ホームページ，<http://www.cosmo-oil.co.jp/information/110312/index.html>，2011。
- 8) 鳥海不二夫，篠田孝祐，榊剛史：大震災，そのときソーシャルメディアは動いた…のか？，第 19 回社会情報システムシンポジウム学術講演論文集，pp. 1-6, 2013。
- 9) 臼井翔平，鳥海不二夫，石井健一郎，間瀬健二：震災による情報伝搬ネットワークの変化，第 26 回人工知能学会全国大会，2012。
- 10) 杉本健二，松原茂樹：消費者の意見に基づく商品検索，情報処理学会論文誌，Vol. 49, No. 7, pp. 2598-2603, 2008。
- 11) 小林のぞみ，乾健太郎，松本裕治，立石健二，福島俊一：テキストマイニングによる評価表現の収集，情報処理学会研究報告，NL-154，pp. 77-84, 2003。
- 12) 安藤俊幸：テキストマイニングと統計解析言語 R による特許情報の可視化，Journal of Information Processing and Management, Vol. 52, No. 1, 2009。
- 13) 風間一洋，鳥海不二夫，榊剛史，篠田孝祐，栗原聡，野田五十樹：東日本大震災時の Twitter データを用いた単語間の関係の時系列変化の分析，第 26 回人工知能学会全国大会，2012。
- 14) 笹原和俊，平田祥人，豊田正史，合原一幸：Twitter ネットワークにおける集団注意の創発ダイナミクス，Technical Report of the 3rd Workshop on Latent Dynamics, Sep 24, 2012, Tokyo, Japan, 2012。
- 15) 石田基広：R によるテキストマイニング入門，森北出版，pp. 80-81, 2008。
- 16) Church, K., Gale, W., Hanks, P. and Hindle, D. : Using statistics in lexical analysis, in Using On-line Resources to Build a Lexicon, Lawrence Erlbaum, pp. 115-164, 1991。
- 17) 鈴木努：R で学ぶデータサイエンス 8 (ネットワーク分析)，pp. 46-48, 共立出版，2009。
- 18) 赤池弘次，甘利俊一，北川源四郎，樺島祥介，下平英寿：赤池情報量基準 AIC - モデリング・予測・知識発見一，pp. 79-80, 共立出版，2007。

(2013. 6. 25 受付)

ANALYSIS OF RELATIONSHIP BETWEEN SOCIAL MEDIA INFORMATION AND RETURNING HOME IN DISASTERS

Hajime ENOMOTO, Masashi KUWANO and Atsushi KOIKE

The objective of this study is to clarify the relationship between social media information (i.e., Twitter) and returning home behavior in disaster. In order to analyze Twitter's textual data, which is tweeted in the Great East Japan Earthquake on 11 March, 2011, this study applies text-mining approach. The results of text-mining confirm the trend of collocation's network among terms after the Earthquake happened, and the distribution of term co-occurrence frequency. On the other hand, using the returning home data sets on 11 March, 2011 collected at Tokyo and Kanagawa, this study calculate the differences of estimated population between one point and the former point. Moreover, using the two data sets, this study analyzed the relationship between Social Media and returning home behavior using a canonical correlation analysis. Term co-occurrence frequency in Social Media and behavior of returning home is applied to two sets of variables in canonical analysis, and the estimation results demonstrated that these two are significantly interdependent.