

新聞記事を用いた災害後における生活の長期的な影響の把握手法に関する一考察

国土交通省 正会員 ○山口 真司
鳥取大学 正会員 谷本 圭志
鳥取大学 正会員 長曾我部まどか

1. はじめに

近年、わが国では災害の激甚化に伴い、被災地に生じる影響が長期化する傾向にある。インフラやライフラインが被る長期的な影響は管理者による記録で容易に把握できるのに対して、生活の影響を把握することは必ずしも容易ではない。今後は、建設業の人手不足に伴って影響の長期化が慢性的になることが懸念されることから、戦略的な復旧プロセスを構想するためには生活に関する長期的な影響の全体像を把握しておくことが有意義である。そこで本研究では、地元の新聞記事を用いて、長期にわたる生活の影響を俯瞰的に把握するための手法を提案するとともに、いくつかの災害事例を対象として災害の特徴を実証的に考察する。

2. 研究のアプローチ

(1) 対象とする災害

異なる特性をもつ複数の災害を対象とする。具体的には、平成30年7月豪雨（西日本豪雨）、平成26年8月豪雨（広島豪雨災害）を対象とし、広島県における地方紙の記事を用いる。このことは、広島県における被害とその影響に焦点を当てることを意味する。双方の災害では土砂流出による被害が多いという共通点がある一方、後者は災害規模が大きいとともに建設業の人手不足が大いに顕在化したため、生活の影響が長期化した事例である。

記事の情報源としては、広島県の地方紙である中国新聞の記事を用いる。具体的には、中国新聞の記事が閲覧できる「中国新聞データベース NIE 版」を利用して、記事を収集する。対象とする期間は災害発生後から一年間とする。

西日本豪雨災害では「西日本豪雨」のキーワードで、広島豪雨災害では「豪雨、土砂災害」の AND の

条件で検索し、それぞれの災害に該当する記事を収集する。その上で、ChaSen を用いて形態素解析を行う。対象とする品詞は名詞（一般、サ変、固有、組織名、地名）、形容詞、形容動詞であり、一つの記事を一つの文書として扱う。記事の合計は 10,610 件（西日本豪雨：7,271 件、広島豪雨災害：3,339 件）、総語彙は 8,762 語である。

(2) 分析の手順

記事には、生活の影響に直接的に関係する記事とそうでないものがある。関係しない記事の例としては、選挙、イベント、スポーツなどに関する記事が考えられる。そこで、自然言語処理の一つの手法であるトピックモデル¹⁾を用いてこれらを区別する。この区別によって生活の影響に直接的に関係しないと特定された記事は、以後の検討の対象外とする。

トピックモデルの適用に並行して、生活の影響を現す単語をリストアップする。それらの具体的な例としては通勤、失業、転居などが考えられる。これらの単語を含む記事の数が一年にわたってどのように累積していくのかに着目することで、個々の影響に関する時系列的な推移の特徴を明らかにする。

その上で、すべての影響（＝リストアップされた単語）に関する推移のデータを用いて、それぞれの影響の近接性を多次元尺度構成法²⁾により可視化する。なお、近接性にはいくつかの観点が想定され、例えば、影響の顕在化のタイミング、影響が消失するまでの期間などの観点が考えられる。これらの観点を軸として、二次元座標系にそれぞれの影響をマッピングすることで、生活に関する長期的な影響の全体像を俯瞰的に把握することができ、また、個々の災害の特徴もそこから読み取ることができると考えられる。

3. 分析手法

(1) 生活の影響に関する記事の特定

任意の記事を d ，単語を v ，トピックを k で表す．トピックモデルは，記事にそれぞれの単語が何回出現するのかをデータとし，トピックの数を事前に与えることで，個々の記事におけるトピックの出現確率 θ_{dk} ，各トピックにおける単語の出現確率 ϕ_{kv} を算出する．このうち， ϕ_{kv} に着目すると，トピック k において出現確率が高い単語が分かるため，どのトピックが生活の影響に関係するかを特定することができる．このようにして特定されたトピックの集合を K で表す．また，記事 d のトピック $k(d)$ は，次式で求めることができる．

$$k(d) = \underset{k}{\operatorname{argmax}} \theta_{dk} \quad (1)$$

よって， $k(d) \in K$ を満たすすべての記事 d が，生活の影響に関する記事である．

(2) 長期的な影響の可視化

以上のように特定された記事を対象に，長期的な生活の影響を可視化する．まず，生活の影響を現す単語をリストアップし，その任意の単語を w で表す．災害後の一年間を有限の期間に区切った上で，ある時点 t までに単語 w が出現する記事の数を $n_t(w)$ ，一年間に単語 w が出現する記事の数を $n(w)$ で表す．すると，ある時点 t までに単語 w が出現する記事数の累積相対度数は次式で求められる．

$$\pi_t(w) = \frac{n_t(w)}{n(w)} \quad (2)$$

よって，単語 w の時系列的な推移は，次式で表される．

$$x_w = (\pi_1(w), \pi_2(w), \dots, \pi_T(w)) \quad (3)$$

ただし， T は災害が発生した一年後の時点であり， $\pi_T(w) = 1$ である．すると，単語 w, w' の距離は次式で測定できる．

$$\|x_w - x_{w'}\| \quad (4)$$

表 1 生活の影響に関係するトピック

1	避難（避難，警戒，勧告）
2	住宅（仮設，入居，半壊）
3	ライフライン（施設，断水，再開）
4	ボランティア （相談，センター，ボランティア）
5	交通（バス，JR，再開）
6	住民の様子（思い，会社，元気）
7	医療・福祉（医療，病院，ケア）
8	復旧事業（ダム，工事，整備）
9	地震の被害（地震，震度，震災）
10	経済（企業，販売，工場）

※（ ）内の単語は，そのトピックに高い確率で出現する代表的な単語

表 2 生活の影響に関係しないトピック

11	自衛隊（岩国，基地，自衛隊）
12	ヒロシマ（平和，被爆，原爆）
13	災害の解説（説明，写真，教授）
14	観光（観光，倉敷，宿泊）
15	イベント（イベント，無料，主催）
16	スポーツ（選手，大会，試合）
17	選挙（選挙，現職，自民党）
18	救助（不明，消防，行方）
19	行政の予算（予算，対策，補正）
20	寄付のお願い（義援金，寄付，団体）

任意の単語のペアについて距離を測定した上で，多次元尺度構成法を用いると，影響の推移の特徴を二次元座標系で可視化することができる．

4. 実証分析

(1) 生活の影響に関する記事の特定

トピックモデルとして LDA (Latent Dirichlet Allocation) を用いるとともに，変分ベイズ法によってパラメータを推計した．トピック数については様々な決定方法があるが，本研究では長期的な影響を明らかにすることが目的であることから，1) なるべく詳細な単位でトピックを把握する，2) 分析者が解釈できるようにトピック数は過度に多くならないようにする，という相反する観点を総合して，トピック数を 20 とした．その結果，表 1, 2 に示す結果が得られた．これらのうち，1～10 を生活の影響に関係するトピックとして特定した．

表3 生活の影響を表す単語

区分	単語
生活機能	通勤, 通学, 買い物, 通院／診察／診療
復旧の 資源	借金／ローン, 物流／流通, 資金／融資, 人手不足
暮らし	住宅, 解体, 集落／コミュニティ／過 疎, 転居／転出
職業	雇用／就職, 失業／離農, 農地
産業	休業／休止, 廃業／倒産／撤退, 閉店

※1：例えば「物流／流通」とは「物流 or 流通」を表しており、複数の単語で影響を把握することを意味する。

※2：「人手不足」は、「工事」「人手」「不足」「入札」「不調」などのキーワードで記事検索し、記事の内容を確認した上で、その内容が建設業の人手不足に関するものであるのかを確認した。

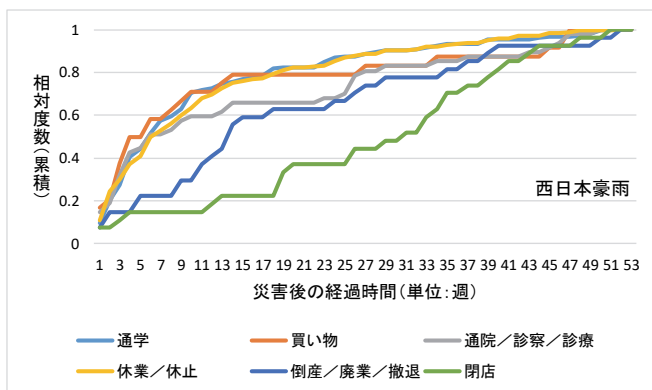


図1 影響の時系列的な推移（西日本豪雨）

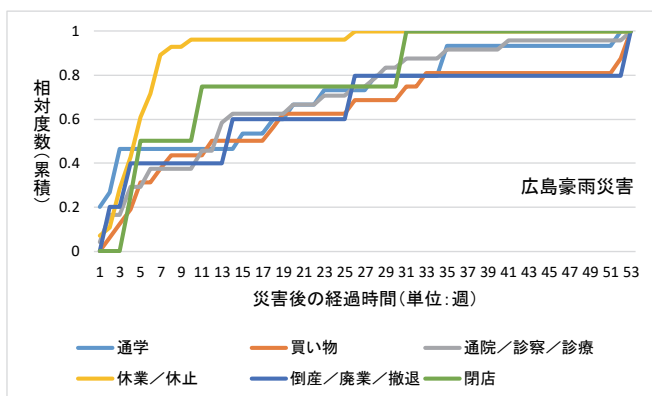


図2 影響の時系列的な推移（広島豪雨災害）

(2) 生活への影響の時系列的な推移

内閣府の復旧・復興ハンドブック³⁾等を参考に、生活の影響を表す単語をリストアップした（表3を

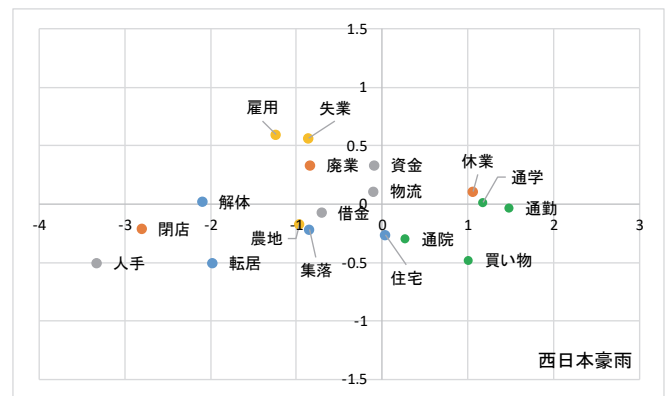


図3 長期的な影響の可視化（西日本豪雨）

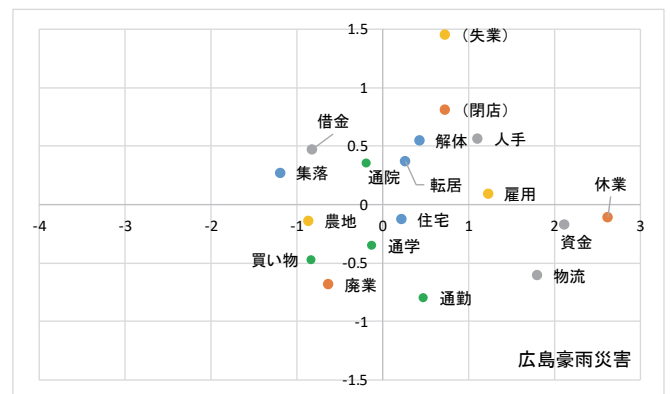


図4 長期的な影響の可視化（広島豪雨災害）

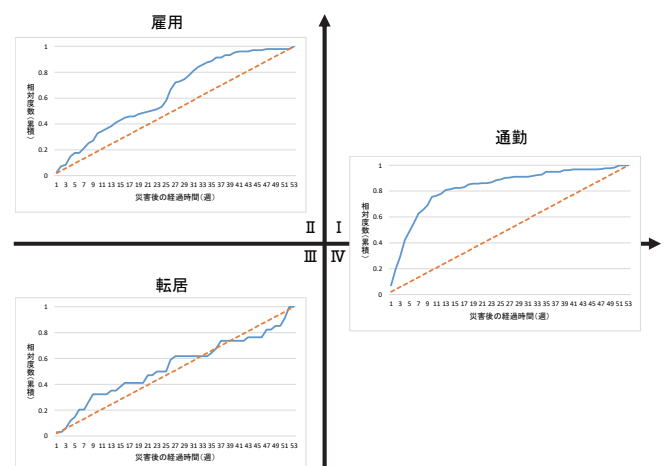


図5 各象限に位置する影響の時系列的な推移

参照）。その上で、それぞれの影響の時系列的な推移をデータ化した。図1, 2に、それらの一部を折れ線グラフで表す。この図より、影響によって推移の傾向が異なり、また、災害によっても傾向が異なることが見て取れる。なお、広島豪雨災害では不連続性が目立つ推移が見られるが、これは記事数が少ないことに起因している。

(3) 多次元尺度構成法による可視化

時系列的な推移のデータを用いて多次元尺度構成法を適用した結果を図3,4に示す。その際、式(4)に示す距離にはユークリッド距離を用いた。図中の()で囲われている影響は、記事数が5未満と非常に少なく、参考として示していることを表す。プロットの色の違いは、表3に示す区分の違いである。また、データラベルは、表3に示す単語の略語を表している。

図における二つの軸の意味を解釈するために、西日本豪雨を対象とし、各象限に特徴的な影響を選び、それらの時系列的な推移を図3に示す。なお、図の点線は原点と(53週目、累積相対度数が1.0)の二点を結んだ直線であり、毎週同じ数の記事があった場合の推移である。以下では、この線を「基準線」と呼び、各影響の推移とこの線との相対的な位置関係に着目して、二つの軸の特徴を把握する。

横軸は影響が災害後早期に集中して顕在化したかを表しており、右ほどそうであったことを表している。具体的には、「雇用」「転居」に比べて「通勤」は基準線よりも左側に時系列的な推移が見られることから、災害後早期に集中して多くの記事が観測されたことを意味する。災害後早期に影響が集中して顕在化したということは、早期に終息しているということでもあることから、右に位置する影響ほど生じたタイミングが早いと言える。

縦軸は、影響が消失するまでの期間の長さであり、上ほど短期に消失し、下ほど長引いたことを表している。図3の「雇用」と「転居」に着目すると、「雇用」は基準線と交差していないのに対して、「転居」では交差している。このことは、「雇用」に比べて「転居」の記事が災害後前期になかなか増えず、後期に増えたことを意味する。

以上より、図の左下方に位置している影響ほど、長期化した影響と考えられる。ただし、ここで言う影響はあくまで記事で把握できる範囲での影響であり、実際に生じている影響と厳密に一致するわけではなく、あくまで、記事という粗い精度をもつ情報で全体を俯瞰するという目的のもとでとらえた影響であることに留意されたい。

(4) 考察

先述のように、広島豪雨災害に比べて西日本豪雨

は災害規模が大きく、建設業の人手不足も顕在化したため、生活の影響が長期化した事例であることを踏まえつつ、図3と4を比較する。すると、住宅や農地に関する影響は、双方の災害において同じ位置にプロットされており、災害規模等の差異があっても影響の現れ方はほぼ同様であったことが分かる。

大きな違いが見られる影響は、解体、転居、人手、閉店である。広島豪雨災害ではこれらの影響は第二象限に位置していたのに対し、西日本豪雨では第三象限(解体については第二象限だが、第三象限に近い位置)にプロットされている。二つの軸の解釈に基づくと、図の左下方に位置している影響ほど、長期化した影響と考えられることから、第二象限に位置していた影響が第三象限に存在していることは、それらの影響が西日本豪雨で顕著に長期化したことを表している。これらの影響を踏まえると、住宅の解体が遅れるとともに、閉店が長期的に相次ぎ、生活の環境が整わない中、住民の転居が進んだという連鎖的な事象が想起される。また、これらの事象に拍車をかけたのが建設業の人手不足である可能性が推察される。

5. おわりに

本研究では、新聞記事をデータとしつつ、トピックモデルと多次元尺度構成法を用いて災害後における生活の長期的な影響を可視化する方法を提案した。その結果、それぞれの災害の特徴を二次元座標系で簡易に把握することができることが分かった。また、対象とした災害を比較することで、建設業の人手不足によって暮らしへの影響が長期化したとの推察が得られた。

ただし、ここでの推察は二つの災害の比較に基づくものであり、その信頼性は必ずしも十分ではない。その一方、より多くの災害を比較することで、人手不足の影響を追跡することも可能であると考えられる。以上、今後の課題としたい。

参考文献

- 1) 岩田具治：トピックモデル，講談社，2016。
- 2) 例えば、齋藤堯幸，宿久洋：関連性データの解析法 多次元尺度構成法とクラスター分析法，共立出版，2006。
- 3) 内閣府：復旧・復興ハンドブック，2016。