

潜在的ディリクレ配分法を用いた 平成 28 年熊本地震に関するブログ記事の分析

藤原 史一^{†1} 碓石 浩文^{†2} 村川 猛彦^{†1}

概要: 防災・減災の情報を、インターネット上のブログ記事より取得しようとしても、記事が多数存在し、その内容が自分にとって有用であるかの判断は容易ではない。そこで、記事内容を推測できるような分類や、不適切な記事の検出および除去を行えば、効率よく情報の獲得や理解ができるのではないかと考え、これまで災害記事データベースの構築に携わってきた。平成 28 年熊本地震の前後の Yahoo! ブログ災害カテゴリの記事を対象とし、潜在的ディリクレ配分法 (LDA) を用いた記事のトピック分析を実施したので報告する。

Analysis of Disaster Blog Articles around the Time of 2016 Kumamoto Earthquakes Using Latent Dirichlet Allocation

FUMIKAZU FUJIWARA^{†1} HIROFUMI SAKOISHI^{†2}
TAKEHIKO MURAKAWA^{†1}

1 はじめに

近年の自然災害は増加傾向にある。また、災害に対する防災・減災に対する意識は高まっており、災害発生時の備えが必要となっている。例えば、「ゲリラ豪雨」と呼ばれる 1 時間降水量が 50mm 以上の集中豪雨の回数は、1980 年からの 10 年間の平均で 198.5 回、2000 年からの 10 年間の平均で 220.9 回であり[1]、増加傾向にある。豪雨に限定することなく、災害が発生した際の避難方法や、避難時の装備などについて、個人的な対策が不可欠となっている。

防災・減災についての情報を入手する方法として、サーチエンジンや SNS、特定のサイトといったインターネットを用いた Web 上の記事から情報を取得する方法が考えられる。しかし Web 上の記事は、日本語に限っても多数存在し、検索を行って必ずしも欲しい情報が手に入るとは限らず、その内容が自分にとって有益であるかの判断がすぐにはできない。

そこで記事内容を理解できているような分類や、記事として不適切なものを取り除くことができれば効率よく情報の獲得・閲覧ができるのではないかと考えた。その際、ブログ記事を活用することにより、公的機関からの情報だけでなく、それ以外の不足する情報を収集できると判断し、災害記事の特徴を抽出、そして災害記事の分類のための調査・分析をこれまで行ってきた。

2 収集記事

2.1 記事の検索例

ブログを含むインターネット上の情報には、災害や防災に関するまとまった記事群が見られる。それらの取得方法として、Google や Yahoo! といったサーチエンジンや、Twitter、

Facebook といった SNS (Social Networking Service) などがあるが、それらはサイト名や記事のタイトルから内容の理解が必ずしもできない。図 1 に防災・減災記事の例を示す[2]。



図 1 防災・減災記事の例

この記事を得るため、サーチエンジン (Google) を用いていくつかのキーワードで検索した。本稿執筆の時点 (2016 年 8 月) では、「津波」を検索すると「約 16,000,000 件」、「防災」だと「約 121,000,000 件」、また「津波 防災」とすると「約 477,000 件」のヒット数となり、その上位から選び、読み進めていくことで、最終的に記事に到達した。一般的な情報取得の流れを図 2 に示す。実際には前のステップに戻ることもよく行われ、満足できる情報にたどり着くまでには多くの時間がかかってしまう。

^{†1} 和歌山大学
Wakayama University
^{†2} (株)日立システムズ
Hitachi Systems, Ltd.

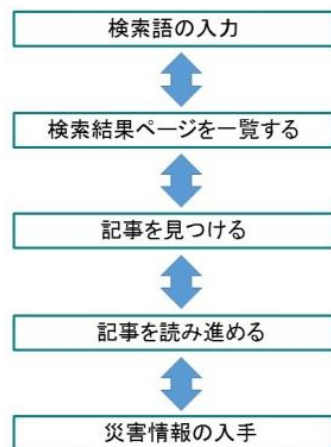


図2 情報取得の流れ

2.2 記事のタイトルと内容の相違

記事タイトルと記事の内容が異なっているものが、記事によっては見られる。例えば、Yahoo!ブログ災害カテゴリには、個人の体験や感想に関する記事、またニュースの転載記事などが、投稿日時の新しいものから順に並べられている。その中のある記事を開覧したところ、タイトルこそ「地域防災（火災 | 消火器説明）」という火災や消火器の取り扱いについての説明に関する記事だと推測できるものであるが、実際に記事本文を読み進めていくと、自分の子供に関する話しか書かれていなかった。このように、ブログ記事においてはタイトルから推測できる内容と実際の内容が異なる場合がある。

3 災害記事データベース

3.1 災害記事データベースとは

筆者らは和歌山大学独創的研究支援プロジェクトの一環である情報通信技術分野の災害関連記事自動収集システムとして「災害記事データベース」を構築中であり、PCでの閲覧を想定している[3]。現在大量の情報が更新し続けているYahoo!ブログ[4]から記事の取得を行い、記事分類プログラムを用いて分類し、インタフェースを通して利用者に記事を開覧してもらうことを想定している。この災害記事データベースでも同様に、個人の経験や意見などが書かれているブログを活用することによって公的機関からの情報だけでなく、それ以外の不足する情報を収集できることを期待している。

しかしWebから災害記事を収集し、分析を行うと記事のデータ量が多くなる。また収集した記事を保存させることが必要であり、その中から必要な情報だけを抽出させなければならない。そこで、データベースを用いることでその問題を解決できないかと考えている。データベースを用いることで収集記事の中から欲しい情報だけを抽出でき記事の検索、閲覧の手間を省くことが期待できる。

ここでブログ記事を対象とした理由について述べる。安部ら[5]は被災者の体験談や意見などの被災経験が有用な情報として活用され各省庁や地方自治体がアンケート調査によって収集している。しかしアンケートの配布や集計は労力を要するだけでなく、コストに見合った情報が必ずし

も得られるとは限らないことも述べられている。また、この問題に対し、個人の経験や意見などが書かれているブログの活用によって不足する情報量を補おうとしており、収集した記事の中から地震の震度を自動抽出する手法について検討している。しかし地震という限定されたもののみを対象としており、また収集データも「地震」、「震度」というキーワードを含む記事のみを評価対象としている。

災害情報の収集は過去の災害分析や、今後の災害予測などに用いることができる。そして足りない情報をブログ記事から収集できれば情報を効率的に集めることができる。

本研究では平成28年4月に発生した熊本地震について、一定の基準に基づき、災害発生前、災害発生直後、災害発生後の3期間に区切った分析を行った。

3.2 関連研究

災害記事データベースの関連研究・システムを取り上げる。

京都大学で構築された災害関連データベース[6]は災害発生時にオンライン情報を用いて被害と支援状況を示すシステムであり、また新聞などの報道記事を収集し地図上で示す機能もある。また防災科学技術研究所は台風災害情報データベース[7]を開発し、過去の台風災害概要や被害状況に関する情報を閲覧できるシステムである。これらは災害情報をリアルタイムで収集しているものであり、災害速報にすぎない。

秋常らは不特定多数が投稿できる災害ブログサイトを構築しており[8]、簡易GIS（地図を用いた特定場所に関する情報）が可能となっている。しかしカテゴリが固定であり更新がされにくく、また記事の優先度などについての記述は無く、資料内画像では災害速報のようなブログ記事群となっているため、個人の意見や感想が書かれているとは考えにくい。

藤原らはブログから抽出したイベント情報をデータベース化して、イベントとブログとの対応付けを行う手法を提案している[9]。その際、イベントデータベースで取り扱う日付情報は「イベントの日付」を採用している。災害記事データベースにおいては「災害発生日」だけでなく「記事公開日」も重要であり、イベントデータベースではこの点を考慮していない。

3.3 災害記事データベースの機能

災害記事データベースのシステム構成を図3に示す。

災害記事データベースは、サーバ上に構築されたデータベースシステムであり、収集した記事をデータベースに保存しておく。

災害記事データベースシステムに記事サイトを登録しておく、定期的にシステムに更新確認を行い、記事が新しく登録されていればそれを「収集記事データ」として保存する。その際それらの記事が災害記事であるかの判定を行い、記事の分類を行うといった「記事分類プログラム」にかけることで不要な記事や記事の振るい分けを行う機能を有する。

記事分類プログラムで処理された記事群は、利用者がシステムの「検索閲覧インタフェース」を通して閲覧し、知識の収集に役立ててもらうことができる。その際、閲覧さ

れる記事は「ファイル名検索」や「単一ファイル内の文字列検索」ではなく「複数文書にまたがって、タイトルや本文を含む文書全体を対象とした検索」である全文検索によって、キーワードや分類項目などから検索することができる。そのために災害記事データベースでは全文検索エンジンを利用する。

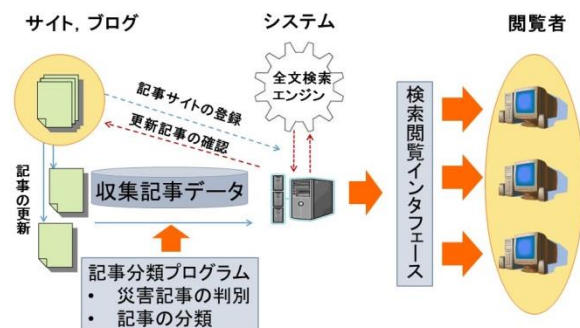


図3 災害記事データベースのシステム構成

3.4 データベースに用いる収集記事群

災害記事データベースの収集記事は、Yahoo!ブログ災害カテゴリから記事を収集することを想定している。Yahoo!ブログとは、Yahoo!Japan が提供するブログサービスであり、「趣味とスポーツ」や「ビジネスと経済」など多数のカテゴリから構成される。

各カテゴリは階層化されており、そこから多数の個人ブログやタレントのブログ、官公庁や地方自治体の情報などを効率よく発見、閲覧することができる。Yahoo!ブログの「災害カテゴリ」へは、Yahoo!ブログトップページから「カテゴリ」タブを選択し、「生活と文化」という大きなカテゴリの中にある「災害」カテゴリを選択することでアクセスができる。

このブログ記事群の「災害カテゴリ」には、災害に関する個人の体験や感想などが多数投稿されている。2015年6月～2016年7月の14か月間における災害カテゴリの記事登録数を図4-5に示す。

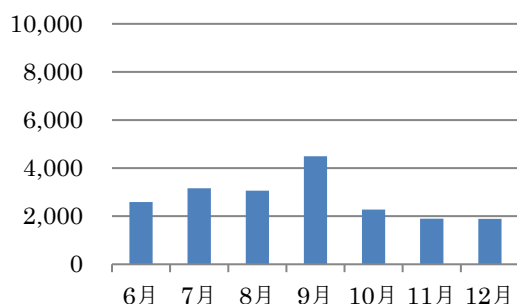


図4 2015年月別の記事投稿数

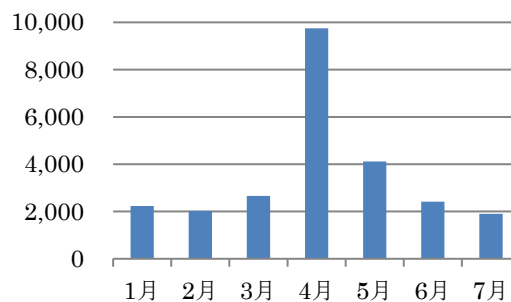


図5 2016年月別の記事投稿数

図4-5から1か月間の記事投稿数は1,000件を超えており、記事データとして収集するブログ記事としては十分な数が確保できることがわかる。また、2015年9月は平成27年関東・東北豪雨が発生したため、前後の月に比べて多くの記事数が得られた。2016年4月は平成28年熊本地震が発生したため、他の月と比べて突出した記事数が得られた。2016年5月については、2016年4月に発生した熊本地震の余震や、同地域の被災状況に関する記事が多く見られた。

3.5 災害記事分析における前処理

各ブログ記事はHTMLで構成されている。そこでは<HTML>、<HEAD>、<BODY>といったタグを用いて構造化されているが、記事の本文を抽出する際には不要な情報が多い。

そこで不要なタグなどを除去し、タイトルと記事本文だけを抽出する際、スクリプト言語であるRubyのNokogiriライブラリ[10]を使用した。このライブラリはHTMLやXMLの構造を解析して、特定の要素を指定して抽出できる。

本文を抽出したのちHTMLから必要な記事内容を抽出したが、その次は記事内容を形態素解析する。形態素解析とは自然言語処理技術の一つであり、文章を意味のある単語に区切られた形態素に分割し、辞書を利用して品詞や内容の判別を行う。本研究では形態素解析器MeCab[11]を採用した。出力例を図6に示す。

ファイル(E)	編集(E)	書式(O)	表示(V)	ヘルプ(H)
和歌山県でエレベーターに乗っているとき				
和歌山	名詞,固有名称,地域,一般,*,*,*和歌山,ワナ			
県	名詞,接尾,地域,*,*,*県,ケン,ケン			
で	助詞,格助詞,一般,*,*,*で,デ,デ			
エレベーター	名詞,一般,*,*,*エレベーター			
に	助詞,格助詞,一般,*,*,*に,ニ,ニ			
乗っ	動詞,自立,*,*,*五段・ラ行,連用タ接続,乗			
て	助詞,接続助詞,*,*,*て,テ,テ			
いる	動詞,非自立,*,*,*一段,基本形,いる,イル,			
とき	名詞,非自立,副詞可能,*,*,*とき,トキ,ト			

図6 形態素解析例

「和歌山県でエレベーターに乗っているとき」という文を形態素解析したところ、「和歌山」、「県」、「で」、「エレベーター」、「に」、「乗っ」、「て」、「いる」、「とき」という9つの形態素に分解され、品詞により、各形態素が地域のものか数値か、また固有のものかといった判別ができる。しかし図6で見られるように「和歌山県」という言葉を形態素解析したとき、「和歌山」と「県」に分かれてしまい、「和歌

山」が県名、市名、地域名として出現しても区別できない。

4 LDA を用いた記事のトピック分析

4.1 LDA

Latent Dirichlet Allocation (潜在的ディリクレ配分法, 以下 LDA, [12]) とはトピックモデルの一つであり, 1 つの文書が複数のトピック (話題) から表現されるという仮定からの教師なし推定である。文書, 単語, トピックの関連について, [12]の事例をもとに説明する。

文書 1 には「国会」「審議」「首相」「選挙」といった単語が出現している。文書 2 では「五輪」「経済」「景気」「球場」, また文書 3 では「景気」「国会」「審議」「対策」といった単語が出現している。そして文書群より, 「国会」「審議」「選挙」「内閣」といった単語で構成されるトピック 1, 「勝利」「五輪」「野球」「球場」で構成されるトピック 2, 「景気」「国会」「審議」「対策」で構成されるトピック 3 を獲得できたとする。

出現単語を比較したとき, 文書 1 とトピック 1 はほぼ重なっているのに対し, 文書 2 ではトピック 2 とトピック 3 が, また文書 3 ではトピック 1 とトピック 3 が混在しており, それらの文書には複数のトピックが含まれていると考えられる。このことにより各文書にはランダムなトピックの混合により表され, このトピックは単語の集合によって表される潜在的な意味に相当する。なお, トピックは明示的な話題を意味しているものではない点には注意が必要がある。

図 7 は LDA のグラフィカルモデルである。

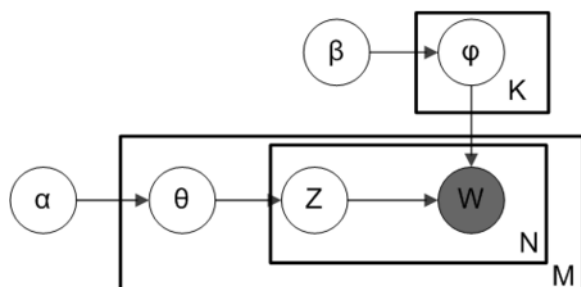


図 7 LDA のグラフィカルモデル

図 7 の説明として M は文書群で N が各文書, そして W は単語, Z がその単語のトピック, θ は文書に対してどのトピックが生じやすいかの多項分布であり, ϕ がトピックに対してどの単語が生じやすいかの多項分布である。 α と β はハイパラメータであり, LDA での推論とは Z や θ , ϕ そして α と β を求めることを意味する。

ここでトピック数 T を与えたとき推定された α は大きさ T のベクトルで出力され, β は $T \times V$ (総単語数) の行列として出力される。

この LDA の適用事例として, [13]では Wikipedia を知識源としてトピックの抽出を行い, 新規の記事に対して分類先の推定を試みている。

4.2 分析の概要

本研究では 2016 年 4 月の記事に対し, 同期間の記事をそれぞれ「平成 28 年熊本地震発生前 (以下「期間 a」)」「平成 28 年熊本地震発生直後 (以下「期間 b」)」「平成 28 年熊本地震の後, 1 日の記事数が落ち着いた後 (以下「期間 c」)」の 3 期間に区切って, それぞれ分析を行った。また事前に

与えるトピック数は 10 に, また各トピックのキーワードは上位 15 にそれぞれ限定した。

分析に使用したプログラムについて簡単に説明する。分析そのものは, Python で書かれたプログラム[14]を使用した。ただしこのプログラムの入力, 1 行が 1 ドキュメント (本研究ではブログ記事 1 件に対応) で, 行は「語の ID: カウント」の並びで表現する必要がある。データベースに格納された記事に対し, この形式に変換してからプログラムを実行するなどの雑多な処理については, Ruby で独自に実装した。

4.3 分析対象について

前項で述べたとおり, 本研究では 2016 年 4 月の記事を 3 期間に区切った上で, それぞれ分析を行った。ここで, 期間を区切った日付, およびその基準について述べる。まず, 熊本県で最初に震度 7 が観測された 4 月 14 日までは期間 a とした。また, 記事数を確認したところ 14 日以前は 1 日あたりの登録記事数が 100 件を下回っていたが, 登録記事数が 744 件を記録した 14 日を境に 1 日あたりの登録記事が増加し, 16 日には 1 日の登録記事が 1,000 件を上回っていることが確認できた。その後は徐々に登録記事数が落ち着きを見せ始め, 22 日には 1 日の登録記事数が 500 件を下回った。この 4 月 14 日から 4 月 22 日までは期間 b, 4 月 22 日以降を期間 c として, 分析を行った。

4.4 LDA のトピック抽出

トピックを抽出したときの出力の例を表 1 に示す。これはあるトピックの単語分布である。上位のものからそのトピックに出現しやすいものが並んでおり, これらのような単語の集合がトピックの数だけ得られる。またこれは図 7 の β に相当するものであり, これとは別に α が得られるが, 今回は使用しない。

LDA により各トピックの単語分布は得られるが, それだけではその単語群が何を表しているかわからない。そこで各ブログ記事を形態素解析し, それに対して LDA による解析を行いトピックの出現確率, 各トピックの単語分布とその単語の出現確率を抽出する。その抽出結果を用いてスコアリングを行い各トピックに合致する災害記事を判定させる。この記事判定までの流れを図 8 に示す。

表 1 「熊本地震直後」のトピック 1 の単語分布

単語	出現確率
地震	0.040005
熊本	0.243347
余震	0.016335
被害	0.015950
何	0.009421
九州	0.008927
事	0.008771
震度	0.007252
人	0.008194
心配	0.008177

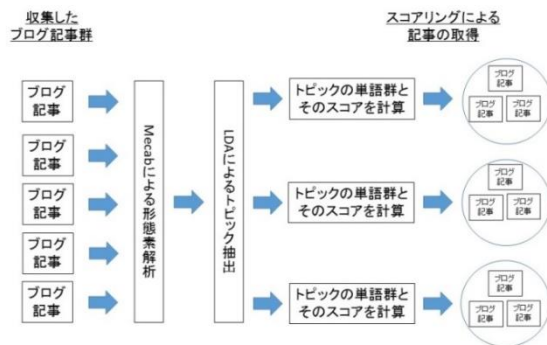


図8 記事判定までの流れ

4.5 スコアリング

分析では LDA で抽出された単語分布と出現確率を用いてスコアリングを行うことでそのトピックに影響を与えている記事の抽出を行った。具体的な方法はトピックの単語出現確率と単語の出現頻度（トピックの単語が各ブログ記事の本文、タイトル中に何回出現したか）を掛け合わせたものであり、式にすると以下の通りとなる。

$$\sum (\text{キーワードの数値}) \times (\text{出現頻度})$$

$$= (\text{記事がトピックに影響を与えている数値})$$

期間ごとの各トピックに対してスコアを算出した。最大スコアを表 2-4 に示す。異なる表の同一番号のトピックには関連がない。またどの期間およびトピックの組み合わせにおいても、最小スコアは 0 となった。

表2 期間 a のスコアリング結果

	最大スコア
トピック 1	2.005512
トピック 2	2.841160
トピック 3	3.992295
トピック 4	1.507309
トピック 5	3.585687
トピック 6	6.343006
トピック 7	16.156547
トピック 8	38.795914
トピック 9	39.260055
トピック 10	17.896467

表3 期間 b のスコアリング結果

	最大スコア
トピック 1	4.414322
トピック 2	2.920505
トピック 3	10.764650
トピック 4	11.332084
トピック 5	5.883466
トピック 6	13.531725
トピック 7	14.178559
トピック 8	206.280092
トピック 9	46.935031
トピック 10	186.058803

表4 期間 c のスコアリング結果

	最大スコア
トピック 1	5.360948
トピック 2	8.884246
トピック 3	11.665654
トピック 4	55.585100
トピック 5	101.817875
トピック 6	63.516756
トピック 7	6.011888
トピック 8	11.999442
トピック 9	14.544923
トピック 10	116.087251

表 2-4 から、期間 a の最大スコアは 39.260055、期間 b の最大スコアは 206.280092、期間 c の最大スコアは 116.087251 となり、指定期間、および各トピックの最大スコアに大きな差が生じていることが確認された。

4.6 結果と考察

まず、LDA による抽出トピックの比較を行う。

熊本地震前となる期間 a では「地震」「震源」「噴火」などのトピックが確認できるものの、特定の地名が特にトピックとして頻出することではなく、「大阪」「茨城」「福島」など、複数の地名がトピックとして出現した。それに対し、熊本地震が発生後となる期間 b、期間 c のトピックを確認すると、「地震」「震源」に加えて「被災」「避難」などトピックが確認でき、地名では「熊本」が頻出した。このことから、日時指定を用いずとも、「避難」や「被災」というキーワードを用いることで、記事が災害の発生前後のどちらに書かれたものなのかをある程度判別することができるのではないかと考えられる。

次に、熊本地震発生後となる期間 b と期間 c の比較を行う。特徴的なトピックとしては熊本地震発生直後となる期間 b のトピック 10 において「八代」「宇城」「天草」という九州の地名が抽出された。これは期間 c に対する分析では見られず、この結果により、期間を災害直後に限定した分

析を行うことで、より詳細な事象に関するトピックの取得、運用が可能になるのではないかと考えられる。

スコアリング結果より、100 を超える最大スコアが得られた期間 b のトピック 8 およびトピック 10、期間 c のトピック 5 およびトピック 10 においては、「市」「町」「村」「年」「月」「日」といった 1 文字からなるキーワードが上位に多く見られた。また、同トピックで最大スコアとなった記事を確認したところ、地名や発生年月といった災害の情報の羅列が行われていることが確認できた。このことより、「市」「町」「村」「年」「月」「日」などの 1 文字からなるキーワードを用いたスコアが著しく高い記事に対しては「災害についての情報が羅列されている記事」というデータを与えて分類に活かせるのではないかと考えられる。

期間 a のトピック 1、トピック 2、トピック 4、期間 b のトピック 2 では、最大スコアがいずれも 3 未満となった。これらの記事についてトピックと記事を照らし合わせて確認した。

まず期間 a のトピック 1 を確認すると、キーワードに「円」「建値」「決済」という、他と関連性が見られないものが見られた。このキーワードについて調査したところ、「建値」というキーワードが見られる記事 8 件すべてが同一ブログの別記事であった。また、これらの記事を確認したところ、いずれも「建値」と「決済」の金額に関する情報の羅列記事となっていた。このことから、ほかのものと関連性が見られない単語が出現した場合、今回と同様に災害と関連性の薄い類似記事をあぶり出すチェック機能を作ることができのではないかと考えられる。

期間 a のトピック 2 では、「日」「年」「者」「被災」などのキーワードが出現していた。またトピック 4 では、「地震」「噴火」「年」「回」などのキーワードが確認された。これらのトピックで最大スコアとなった記事[15]では、災害情報の羅列が行われていた。これらについては、100 を超える最大スコアが得られた記事と同様に、1 文字からなるキーワードを用いたスコアが高い記事に対して「災害についての情報が羅列されている記事」というデータを与えるという分類が行えるのではないかと考えられる。

期間 b のトピック 2 では、「被災」「人」「支援」「災害」などのキーワードが出現していた。また、最大スコアとなった記事を確認したところ、減災についての考察記事が得られた。この記事について「市」「町」「村」「年」「月」「日」などのキーワードが見られるトピック 8 およびトピック 10 によるスコアを確認してみたところ、「6.750850」「0.626421」という平均的なスコアが得られた。このことより、この記事は前段落で述べた「災害情報が羅列された記事」というデータを与えられない論述記事と言える。今後、この記事の類似例を調査することで、災害に関する論述記事抽出の助けとなるのではないかと考えている。

5 おわりに

本研究では 2016 年 4 月の災害記事郡を 3 つの期間に区切り、それぞれの期間に対して LDA 分析を行った。また、その結果を元に記事のスコアリングを試みた。

今後もスコアリングに関する調査・考察を進め、災害記事データベースに格納されている記事の分類を行っていきたいと考えている。

謝辞 本研究はJSPS 科研費 JP25242037 の助成を受けたものです。

参考文献

- [1] “気象庁 | アメダスで見た短時間強雨発生回数の長期変化について” .
<http://www.jma.go.jp/jma/kishou/info/heavyraintrend.html>, (参照 2016-08-23).
- [2] “津波の常識が変わった！東日本大震災の現場を見る [防災]All About” .<http://allabout.co.jp/gm/gc/379157/2/>, (参照 2016-08-23).
- [3] 碓石浩文, 村川猛彦: 防災・減災に関する Web 上の記事を対象とした分類の試み, 情報知識学会誌, Vol.24, No.2, pp.184-188 (2014).
- [4] “Yahoo! ブログ” .<http://blogs.yahoo.co.jp/>, (参照 2016-08-23).
- [5] 安部智也, 安藤一秋: 防災教育に向けた被災経験ブログの収集, 2012 年度 JSiSE 学生研究発表会 (2012).
- [6] “災害関連データベース | CIAS 京都大学地域研究統合情報センター” .
<https://www.cias.kyoto-u.ac.jp/database/category/40.html>, (参照 2016-08-23).
- [7] “台風災害データベースシステム” .
http://ccwd05.bosai.go.jp/DTT/search_jsp/login.jsp, (参照 2016-08-23).
- [8] 秋常秀明, 伊藤英恵, 加勢田毅, 中村直喜: ブログ技術を利用した防災情報共有システム, 平成 21 年度河川情報シンポジウム講演集, pp.6.1-6.10 (2009).
- [9] 藤原泰士, 難波英嗣, 竹澤寿幸, 石野亜耶: イベントデータベースとブログの自動対応付け, 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 (DEIM Forum 2015), C8-5 (2015).
- [10] “Nokogiri 鋸” .<http://www.nokogiri.org/>, (参照 2016-08-23).
- [11] “McCab: Yet Another Part-of-Speech and Morphological Analyzer” .<http://taku910.github.io/mecab/>, (参照 2016-08-23).
- [12] 岩田具治: トピックモデル, 講談社 (2015).
- [13] 牧田健作, 鈴木浩子, 小池大地, 宇津呂武仁, 河田容英: Wikipedia を知識源とする分野トピックモデルの推定と分析, 情報処理学会研究報告, 2012-DBS-155, pp.1-11 (2012).
- [14] “satomacoto: Python で LDA を実装してみる” .
<http://satomacoto.blogspot.jp/2009/12/pythonlda.html>, (参照 2016-08-23).
- [15] “地震頻発の阿蘇地方で地震活動が活発化そして南下? (地震) - 原典聖書研究 - Yahoo! ブログ” .
<http://blogs.yahoo.co.jp/semidalion/49259364.html>, (参照 2016-08-23).