

# トピックモデルの適用による 地域防災ワークショップ討議の地域特性評価

榊原 弘之<sup>1</sup>・森 翔太<sup>2</sup>・長曾我部 まどか<sup>3</sup>・西江 哲哉<sup>4</sup>・三橋 洗道<sup>5</sup>・  
山中 亮<sup>6</sup>・宮国 敏秋<sup>7</sup>・神谷 大介<sup>8</sup>・塚井 誠人<sup>9</sup>

<sup>1</sup>正会員 山口大学教授 大学院創成科学研究科 (〒755-8611 山口県宇部市常盤台2-16-1)  
E-mail:sakaki@yamaguchi-u.ac.jp

<sup>2</sup>非会員 広島市 (〒730-8586 広島市中区国泰寺町一丁目6番34号)

<sup>3</sup>正会員 鳥取大学助教 大学院工学研究科 (〒680-8550 鳥取市湖山町南4丁目101)  
E-mail:mchoso@tottori-u.ac.jp

<sup>4</sup>非会員 トマト銀行 (〒700-0811 岡山市北区番町2丁目3番4号)

<sup>5</sup>非会員 千葉県 (〒260-8667 千葉市中央区市場町1-1)

<sup>6</sup>正会員 中央建設コンサルタント (〒901-2126 沖縄県浦添市宮城5丁目12番11号)  
E-mail:yamanaka@cyuo.co.jp

<sup>7</sup>正会員 中央建設コンサルタント (〒901-2126 沖縄県浦添市宮城5丁目12番11号)  
E-mail:tmiyaguni@cyuo.co.jp

<sup>7</sup>正会員 琉球大学准教授 工学部 (〒903-0213 沖縄県中頭郡西原町字千原1番地)  
E-mail: d-kamiya@tec.u-ryukyu.ac.jp

<sup>7</sup>正会員 広島大学准教授 大学院工学研究科 (〒739-8527 東広島市鏡山一丁目4番1号)  
E-mail: mtukai@hiroshima-u.ac.jp

ワークショップ等の討議においては、参加者らによる発話内容に参加者の問題意識、関心事などが反映されていると考えられる。しかし、発話内容がテキスト化されることが少ないため、KJ法等で意見集約された結果を除けば、討議内容が計画プロセスに十分反映されるとは限らない。本研究では、沖縄県国頭村の村内各地区で実施されている地域防災ワークショップについて発話音声データをテキスト化し、トピックモデルを適用した。発話内容から主要な話題を抽出することができたほか、地区ごとの話題の重みの強弱も明らかとなった。これにより、各地区共通の関心事や、特定の地区特有の関心事などを明らかにすることを試みる。

**Key Words** : community disaster risk managemen, text mining, topic model, topic extraction

## 1. はじめに

ワークショップ (以下WS) 等の討議においては、参加者らによる発話内容に参加者の問題意識、関心事などが反映されていると考えられる。しかし、発話内容が逐一テキスト化されることは少ない。またテキスト化の目的も主として議事録作成のためであり、テキストが分析されることが多いとは言えない。そのため、KJ法等で意見集約された結果を除けば、討議内容が計画プロセスに十分反映されるとは限らない。そこで、討議における参加者の発言内容から、主たる話題を抽出する分析技術

が必要となる。

筆者のうち長曾我部・榊原の既往研究<sup>1)</sup>においては、WSの討議記録に形態素解析<sup>2)</sup>を実施した上で、因子分析を適用し、語彙群の共起傾向から話題の特定を試みている。しかし因子分析を用いた場合、因子負荷量及び因子得点が負の値を取ることがあり、解釈が困難になるという課題が存在した。一方筆者のうち塚井ら<sup>3)</sup>は、機械学習による自然言語処理の分野で近年開発が進んでいるトピックモデル<sup>4,5)</sup>を地域公共交通会議の議事録データに適用し、話題抽出が可能であることを示している。また川野ら<sup>6)</sup>も、熊本地震の被災地である益城町の仮設住宅

住民を対象にした聞き取り調査での発言録に対してトピックモデルを適用している。トピックモデルにおけるトピック分布，語彙分布の重みは常に非負の値を取り，解釈がしやすいという利点が存在する。

本研究では，沖縄県国頭村の村内各地区で実施されている地域防災ワークショップについて，発話音声データをテキスト化し，塚井ら<sup>3)</sup>と同様のトピックモデルを適用する。発話内容から主要な話題の抽出を行うとともに，地区ごとのトピックの重みの強弱から，各地区共通の関心事や，特定の地区特有の関心事などを明らかにすることを試みる。

## 2. 沖縄県国頭村の地域防災WS

### (1) 沖縄県国頭村の概要

沖縄県国頭郡国頭村は沖縄本島の最北端に位置し，人口 4922 人(2015年 国勢調査)である。高齢化率は，沖縄県全体の 17.4%に対し，国頭村は 27.5%(2015年 国勢調査)であり，高齢化が進行している。

国頭村内には，20地区の集落が存在している。村役場の所在地である辺土名地区に人口の約3割が居住する一方，その他の集落は，東西両海岸に沿った小規模河川河口付近に主に分布している。

国頭村における主要な災害リスクは台風による風水害である。沿岸部を中心に越波等に伴う通行止めが毎年のように発生している。また，過去に津波被害の記録はないものの，沖縄県は，地域で起こりうる最大クラスの地震を想定し，津波浸水想定を実施している。その結果によれば，国頭村の集落の大半が被災し，幹線道路が冠水し道路ネットワークが寸断されることが予想されている。また，広大な村の領域に集落が点在しており，かつ高齢化が進み人口が減少傾向にあるため，村単位での防災活動にも限界があることが指摘されている<sup>7)</sup>。

### (2) 国頭村の地域防災WS

2013年から2016年までの4年間に，国頭村内の8地区において地域防災WSが開催されている。これらのWSの目的としては，防災の専門家から地域住民への情報提供，地区内の潜在的なリスク要因の特定，地域住民の災害リスク，防災に関する関心事項の把握，地区単位での共助の取り組みのきっかけづくり，等が挙げられる。

表-1に各地区におけるWSの実施年を示す。毎年，1～3地区でWSが開催された。また表-2に各WSにおける班数を示す。地域内の点検が行われたため，討議データ自体が存在しない安田地区の第1回WSを除いて，2，3班に分かれて討議を行っている。

表-1 各地区における地域防災WS実施年

地区名	与那	安田	伊地	鏡地
WS実施年	2013	2014	2015	2015

地区名	辺野喜	宇嘉	桃原	安波
WS実施年	2015	2016	2016	2016

表-2 各WSにおける班数

地区名	与那	安田	伊地	鏡地
第1回WS	3	0	3	3
第2回WS	3	3	3	3

地区名	辺野喜	宇嘉	桃原	安波
第1回WS	3	3	2	3
第2回WS	3	2	3	3

## 3. トピックモデルの適用

### (1) トピックモデルの概要

本研究では，2. で説明した沖縄県国頭村の地域防災WSの討議データに対してトピックモデルを適用し，話題抽出を試みる。まず本節では，トピックモデルの概要について説明する<sup>3,4,5)</sup>。

トピックモデルによるテキスト分析において，分析の基本単位は，バッグ・オブ・ワーズ(Bag of Words, BOW)と呼ばれる，一連のテキスト中で用いられる語彙の多重集合である。BOW表現においては，文書と呼ばれる最小分析単位内における語彙の出現順は分析上考慮されていない。まず，以下の定義を行う。

$d \in D$  : 文書 (BOW) ( $D$ は文書の総数)

$k \in K$  : 話題 ( $K$ は話題の総数)

$v \in V$  : 語彙 ( $V$ は語彙の総数)

$w_{dn}$  : 文書 $d$ 中の $n$ 番目の出現語句

$N_d$  : 文書 $d$ 中の語句総数

トピックモデルにおいて，各文書 $d$ は次式で示されるトピック分布 $\theta_d$ を有する。

$$\theta_d = (\theta_{d1}, \theta_{d2}, \dots, \theta_{dk}) \quad (\theta_{dk} \geq 0 \text{ and } \sum_{k=1}^K \theta_{dk} = 1) \quad (1)$$

一方各話題 $k$ は以下の語彙分布 $\phi_k$ を有する。

$$\phi_k = (\phi_{k1}, \phi_{k2}, \dots, \phi_{kv}) \quad (\phi_{kv} \geq 0 \text{ and } \sum_{v=1}^V \phi_{kv} = 1) \quad (2)$$

$\theta_{dk}$ は文書 $d$ における話題 $k$ の「重み」を示し， $\phi_{kv}$ は話題 $k$ における語彙 $v$ の「重み」を示す。(1), (2)式に示されるように，各文書は複数の話題を含み，各語彙は複数の話題に含まれ得る。実際に観察された文書 $d$ の出現確率は次式で示される。

$$w_d = \prod_{n=1}^{N_d} \sum_{k=1}^K \theta_{dk} \phi_{kn} \quad (3)$$

$\theta_{dk}$ 及び $\phi_{kv}$ は潜在変数であり，推定の必要がある。

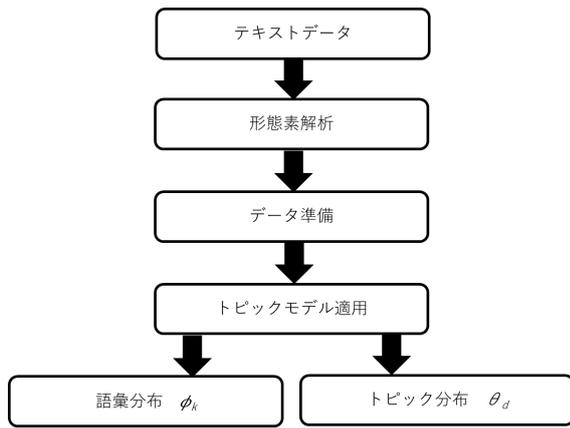


図-1 分析の手順

(2) 分析の手順

図-1に分析の手順を示す。初めに、テキストデータに対して形態素解析を適用し、文をBOWに分解する。次に、使用語句を絞り込んだ上で、分析のためのデータを準備する。その上でトピックモデルを適用し、語彙分布  $\phi_k$  とトピック分布  $\theta_d$  を推定する。

(3) データの準備

形態素解析によって得られたBOWデータから、本研究では名詞、動詞、形容詞を分析に使用する。まず、意味が不明瞭な語彙を除外した上で、使用頻度順に語彙を並び替える。その上で、最頻使用語彙からの累積使用頻度が約90%となるように、最低使用頻度を決定した。その結果、最小使用頻度は10回、累積使用頻度は91.8%となった。分析に使用する語彙は1163種、語句総数は81,236語となった。

トピックモデルを適用するためには、話題の総数  $K$  と、1つの文書中の最大の語句数  $U$  を決定する必要がある。複数の  $K, U$  の組み合わせを試行し、対数尤度を比較した結果、 $K = 20$ 、 $U = 50$  の組み合わせを使用することとした。

4. 分析結果

(1) 語彙分布からの話題抽出

前述のように、計20件の話題を抽出した。以下、その一部について、重み上位20位以内の語彙分布と、その話題に関する解釈を示す。

話題1 (図-2) : 「子供」、「娘」、「息子」等の家族を示す語彙の重みが高いことから、住民の家族形態に関する話題であると解釈した。各世帯の状況の把握は、災害時の避難方法を考える上で重要である。この話題は全20件の中で最も重みが大い。また「車」、「運転」等の重みも高く、自動車の運転の可否についても併せて確

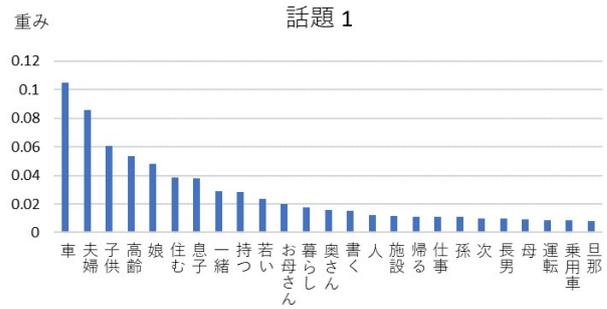


図-2 話題1 (家族形態) の語彙分布

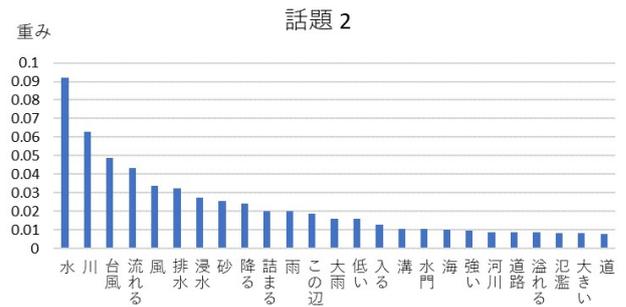


図-3 話題2 (水害) の語彙分布

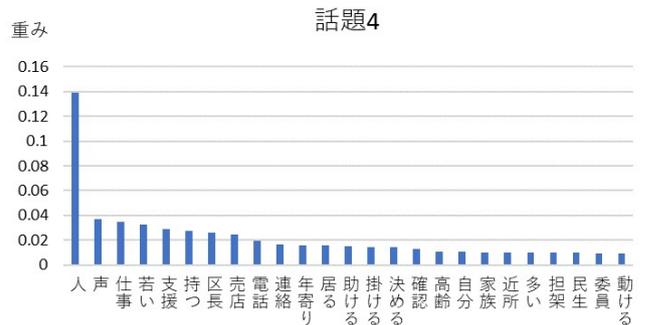


図-4 話題4 (共助) の語彙分布

認していたものと考えられる。

話題2 (図-3) : 「台風」、「風」、「浸水」等の語彙の重みが高いことから、水害に関するトピックであると解釈した。台風による風水害が多いことから、地域防災WSにおいても主要な話題となったと考えられる。

話題4 (図-4) : 「声」を「掛ける」、「若い」、「支援」、「連絡」等の語彙の重みが高いことから、災害時の共助に関するトピックであると解釈した。地区内における共助の仕組みの構築は、地域防災WSの目的の1つであり、そのための議論がなされたと考えられる。

以上のようにして解釈された20件の話題のタイトルの重みの大きい順に以下に示す。

家族形態、水害、自治会、共助、建物の状況、地図の確

認, 津波避難, 進行, 高潮・津波, 要支援者, 道路交通, 消防, 危機感, 自主対策, 炊き出し, ブロック塀, 防災無線, 避難路, 地域の状況, 地震

(2) トピック分布からの地域特性評価

50語句(U=50)で構成される文書ごとに, トピック分布が形成される. 各WSの班ごとに, トピック分布における各話題の重みを集計し, 全話題の重みに対する比率を算出する. 地区ごとのこの比率の平均値を, 当該地区におけるその話題の相対重みとする.

各地区における各話題の相対重みを比較評価するために, レーダーチャートを作成した. 図-5に, 8地区の平均相対重み, 平均相対重み+標準偏差, 平均相対重み-標準偏差の値をレーダーチャートとして示す.

図-5において, 標準偏差は地区間のその話題への言及量のばらつきの程度を示す. 標準偏差の大きい話題は, 地区間で関心の差が大きい話題であると推測される.

「家族形態」, 「高潮・津波」等がこれに該当する. 一方, 平均相対重みが大きく, 標準偏差が小さい話題は, 各地区で同様に言及される, 共通的な話題であると考えられる. 「共助」等がこれに該当する.

図-6に与那地区における各話題の重みを示す. 各話題の相対重みが比較的近似しており, 多くの話題が言及されていたと考えられる. また「地域の状況」, 「要支援者」の重みが特に大きく, 地域コミュニティに関する関心が高いものと推測される. 与那地区は, 全8地区の中でもコミュニティ活動が盛んであることから, 日頃のコミュニティ活動の延長線上で地域防災について議論がなされていたものと考えられる.

一方, 図-7は安田地区における各話題の重みを示す. 与那地区と比較すると, 平均相対重み+標準偏差及び平均相対重み-標準偏差を超過する話題が複数存在し, 与那地区と比較して話題間の言及量のばらつきが大きいことがわかる. 特に, 「高潮・津波」, 「津波避難」, 「道路交通」, 「ブロック塀」, 「共助」などの話題への言及量が, 他地区と比較して多かったものと考えられる. このうち, 「道路交通」は地区内細街路の自動車による通行可能性などに関する話題であり, 「ブロック塀」はそれらの細街路における住宅のブロック塀の倒壊可能性に関する話題である.

安田地区は, 東海岸に面しており, 津波に関するシミュレーションにおいては最短津波到達時間が17分とされている. また地区の多くの領域が津波により浸水すると予想され, 高台への避難が必要であると考えられている. 地域防災WS開催時点でその情報は既に地域住民に伝達されていた. そのため, 大地震時の津波リスクや, その際の迅速な避難のための体制づくり, 避難経路やその障害要因に関する話題に議論が集中したものと考えら

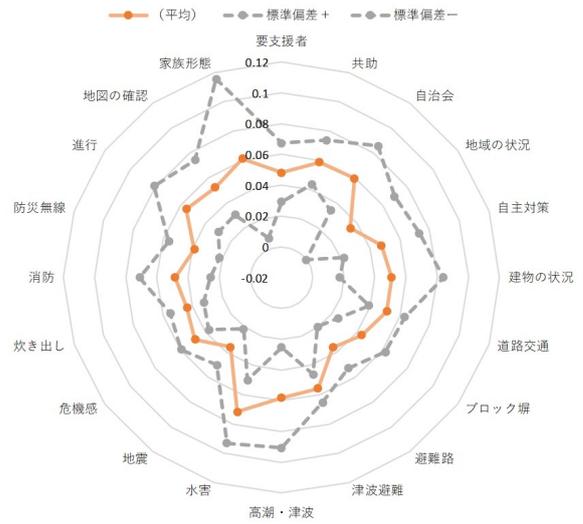


図-5 各話題の相対重みと標準偏差

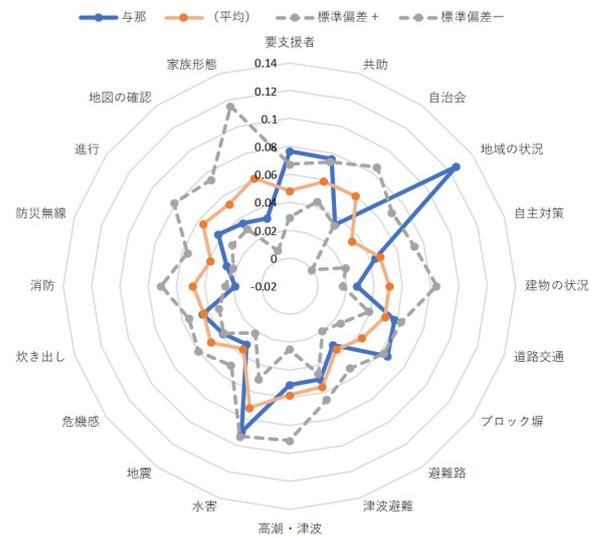


図-6 与那地区における各話題の重み

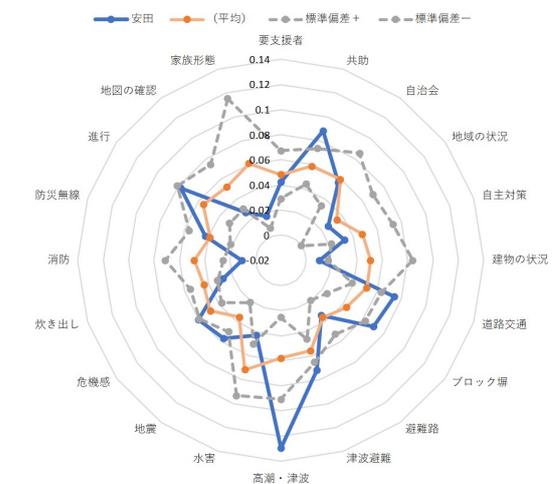


図-7 安田地区における各話題の重み

れる。

## 5. おわりに

以上、本研究では、沖縄県国頭村の地域防災WSを対象に、トピックモデルを適用し、討議における主要な話題を抽出した。また地区ごとの話題の重みの強弱に差異が存在することも認められ、各地区に共通する話題、地区特有の話題が存在することも明らかとなった。

今後の課題としては、地区間の話題の分布の差異に関する情報を、地域防災にどのようにフィードバックするかが挙げられる。

### 参考文献

- 1) 長曾我部まどか，榊原弘之：ワークショップにおける相互補完的対話の分析，都市計画論文集，Vol.50, No.1, pp. 28-36, 2015.

- 2) 奈良先端科学技術大学院大学情報科学研究科自然言語処理講座：ChaSen（茶筌），<http://chasen-legacy.sourceforge.jp/>（最終参照日 2018 年 4 月 26 日）
- 3) 塚井誠人，椎野創介：討議録に対するトピックモデルの適用，土木学会論文集 D3, Vol.72, No.5, pp. I\_341-I\_352, 2016.
- 4) 佐藤一誠，奥村学：トピックモデルによる統計的潜在意味解析，コロナ社（自然言語処理シリーズ），2015.
- 5) 岩田具治：トピックモデル，講談社（機械学習プロフェッショナルシリーズ），2015.
- 6) 川野倫輝，佐藤嘉洋，円山琢也：トピックモデルを用いた益城町仮設住宅聞き取り調査の自由回答の基礎分析，土木計画学研究・講演集，Vol. 56, 2017.
- 7) 山中亮，宮国敏秋，神谷大介，中山貴喜，榊原弘之，長曾我部まどか：地域防災計画の見直しにおける事前危険度評価の取り組み，土木計画学研究・講演集，Vol. 49, 2014.

(2018.4.27 受付)