

東日本大震災後の被災者ニーズの分析

——データマイニング技術による時系列トピック抽出——

橋 本 隆 子

1. INTRODUCTION

2011年3月11日の東日本大震災から3年余りが経過した。震災発生後、多くの人々が自宅や仕事、そして家族との生活を失い、きわめて困難な状況に陥った。行政機関やNPO法人は、避難所の設置、物資の提供、仮設住宅の建設・斡旋など種々の支援を行っていたが、かつて我々が経験したことのない大きな災害であったため、多くの支援は手探りであった。支援する側のNPO法人も自分たちの支援が被災者のニーズを的確に捉えていたかについて不安を抱えていたという。被災者支援を振り返り、評価するためにも大震災後、被災者のニーズがどのように変化していったかを解析することは、重要であると我々は考える。そこで本稿では、被災者の声が記載されたブログデータを対象とし、大震災発生後、被災者のニーズ・関心がどのように変化したかを時系列で解析する。対象とするブログは、岩手県最大のNPO法人SAVE IWATE [1] が提供している「番屋日報」[2] である。「番屋日報」には2011年6月5日から現在に至るまでの被災者の要望や不安、不満が綴られている。このブログに対して、時系列でニーズの抽出を行い、被災者のニーズ変化の解析を行う。解析手法としては、3次元行列構造に基づく時系列トピック解析及びメジャーなトピック抽出技術の一つである Latent Dirichlet Allocation (LDA) [3] を用いる。

さらに解析結果と、被災者の状況を比較し、本論文の手法により抽出されたニーズの有効性について評価する。本手法が被災者への適切な支援を行うための意思決定に有効であるかについても考察する。本研究により、3次元行列構造及びLDAによるトピック抽出の効果が検証できるとともに、トピック抽出に基づく意思決定支援システムの可能性について示すことができる。

本稿は以下のような構成になっている。第2章において、トピック抽出に関する関連研究について述べる。第3章では、研究協力者であるNPO法人SAVE IWATEの本研究に対する期待と、被災者のニーズ抽出の素材となる対象ブログについて説明する。第4章で本稿で提案する手法について述べる。第5章で実際のブログデータから抽出できたニーズを実験結果として示す。第6章で、LDA手法により抽出されたニーズと被災者の状況を比較し、解析結果の有効性について評価する。さらにNPO法人に対してインタビューを行い、本手法が被災者への適切な支援を行うための意思決定に有効であるかについて考察する。最後にまとめと今後の展開について述べる。

II. 関連研究

ソーシャルメディアを対象とした解析技術には種々のものがある。ここでは、単語の共起や時系列上の変化をベースとしたデータマイニング技術について述べる。

A. 単語の共起をベースとしたデータマイニング技術

ソーシャルメディア上の書き込みは、自然言語で記述されており、それをコンピュータで解析できるような形式に変換する必要がある。自然言語処理技術としては、形態素解析(意味を持つ最小の単語に分割し、品詞を判定)、キーワード抽出(単語の重要度をスコアリング)、センチメント分析(好き・嫌い、といった感情を評価)などが挙げられる。たとえば災害発生後に不足しそうな品物をいち早く検知するために、ソーシャルメディア上の書き込みから「買えない」、「買占め」といった単語を含む書き込みを抽出し、「買えない」品物が何なのかを言語処理技術を利用して特定する。あるいは特定の製品に対して、「便利」、「楽」、「おもしろい」、「好き」といったポジティブな感情を表す単語と、「不便」、「面倒」、「つまらない」、「嫌い」といったネガティブな感情を表す単語の数や発生頻度・割合を比較し、製品に対する評判を判定するといった手法などが良く用いられている。関口ら[5]は、ブログの投稿を対象とし、単語の共起関係と頻度をベースに単語間の関係と話題を可視化する手法を提案している。Asurら[6]は、Twitterのツイート数に基づく単純なモデルを用いることで、ユーザによる内容の共鳴が流行を発生させる主要因となっていることを見つけている。Wangら[8]は、日本語を対象としたグラフィカルな評判解析システムを提案している。彼らのシステムは、製品とそれに対する人々の評価を単語の共起に基づいた単純なグラフを用いて可視化するものである。

Pennacchiotte[9]らは、Twitterを対象として、LDAを用いてトピック抽出を行っている。一つ一つのツイートを1ドキュメントとするのではなく、書き込みユーザ単位でドキュメントを作成し、トピック抽出するほうが良い結果が得られることを示している。

B. 時系列上の変化分析をベースとしたデータマイニング技術

Radinskyら[7]は、時系列の意味解析モデル(Temporal Semantic Analysis, TSA)を提案している。これはニューヨークタイムズの記事を対象として、単語の共起変化を時系列に評価するDynamic Time Warping(DTW)[10]によって、予め定義した単語間の関係を評価し、意味を発見する手法である。Dingら[11]は、ソーシャルネットワーク上でコミュニティを発見する手法について提案している。彼らはコミュニティネットワークグラフ分割アルゴリズムを提案し、コミュニティが時間推移に従って構成されていく様子を可視化している。Jiangtaoら[12]は、階層的コミュニティモデルを用いて、動的なソーシャルネットワークから組織構造を抽出する手法を提案している。

ネットワーク上のコミュニティ追跡として、Greeneら[13]は、動的なネットワーク上のコミュニティの追跡モデルを提案している。彼らはネットワークからサブグラフを抽出し、あらかじめ遷移していそうなコミュニティを特定し、Jaccard係数[14]を計算することでコミュニティ間の類似度を計算している。

C. 我々のアプローチ

上記で挙げた既存研究は、ソーシャルメディアに記載されたメッセージから、特定の製品・会社等や、特定のコミュニティ等に関する情報を抽出し、その関係変化に注目することで、その構造や時系列変化を解析するものである。解析対象とする製品・会社、コミュニティ（に関連する単語群）をあらかじめ決定している。一方、我々のアプローチは、災害（大震災）という大きな枠を決めつつも、その中で特定の話題に絞るのではなく、不特定の話題を対象として抽出しようとしている。我々はこれまでも「消費者の口コミから話題を抽出し、想定外の消費行動を発見する手法」[15] や、「東日本大震災の口コミから風評を可視化する技術」[16] について研究を行い、不特定の話題を動的に抽出し、その時系列変化の解析を行ってきた。データ構造としては、単語の共起行列構造を採用し、行列構造の時系列な変化から、構造上の距離を算出し、話題の推移や活発化しつつある話題をフレキシブルに発見することを試みてきている。またメジャーなトピック抽出モデルの一つである LDA も用いている。不特定な話題を対象として時系列な評価を行なっているという点が我々のアプローチの新規性であると言える。

Ⅲ. 本研究に対する期待と対象ブログ

本論文で対象とするデータは、NPO 法人 SAVE IWATE が提供しているブログ「番屋日報」[2] である（図1）。SAVE IWATE（東日本大震災被災地支援チーム）は震災直後の平成23年3月13日、盛岡市で設立されたNPO 法人である。震災発生以降、被災者の安否情報の収集、整理、提供や被災地実情についての調査、支援物資の搬送、避難生活の支援などを積極的に行なっている。そうした活動を通じて、SAVE IWATE は被災者のニーズ変化を的確に掴むことの難しさを感じるようになったとのことである。たとえば震災直後から実施された「衣服の支給」のような支援活動は、大変効果的であり、必要性の高い支援であった。しかし、震災後時間が経過するにつれ、こうした支援をいつまで続けるべきか、いつ止めるべきかということが課題となってくる。SAVE IWATE の人々には、支援の必要性が薄れてきたような感覚はあるが、実際に「支援を止める」決断をすることは難しい。そんな時、何らかのツールで被災者ニーズの変化を定量的に示すことが出来れば、支援を打ち切る、あるいは新しい支援を開始するといった意思決定を後押しすることが可能となる。この「被災者支援のための意思決定支援ツール」が彼らのモチベーションであり、我々の研究に期待するところである。「番屋日報」は、被災者の困りごと、要望などが日々綴られているブログである。震災直後から開始され、被災者のさまざまな状況が自然な文章で記載されている。被災者のニーズ変化を解析するニーズ抽出エンジン開発の第一ステップとして、本ブログが重要な素材になると我々は考えた。



図 1. 番屋日報画面例

IV. 提案手法

提案手法は、以下の 5 ステップからなる (図 2)。

- Step A : データ収集
- Step B : 言語処理
- Step C : トピック抽出
- Step D : 時系列トピック推移解析
- Step E : 可視化

提案手法は筆者らの従来研究 [17], [18] に既に紹介されているので、そちらを参照いただきたい。ここでは提案手法のフレームワークについて簡単に紹介する。

Step A : データ収集

Step A は、対象とするソーシャルメディアから書き込みメッセージ $D = \{d_i\}$ をクロールする。対象ソーシャルメディアの特徴に応じて、「大震災」などの時事問題を表現する単語を入力クエリとし検索結果を得る場合もある。本論文では NPO 法人 SAVE IWATE のブログ「番屋日報」をターゲットとし、2011年6月3日から2012年7月27日までのブログ全体 (502件) を収集した。一つの書き込みを 1 ドキュメント d_i として定義し、 D を処理対象の全ドキュメント集合と考えている。

$$d_i = (MID_i, Posted_i, Title_i, Content_i).$$

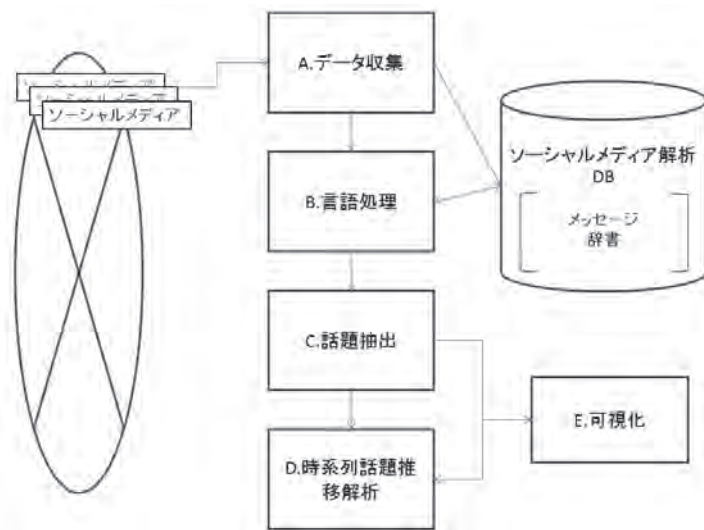


図2. 提案手法フレームワーク

ここで、 MID_i はメッセージの ID、 $Posted_i$ は投稿日時、 $Title_i$ はタイトル、 $Content_i$ は内容である。

Step B：言語処理

次に、形態素解析技術を用いてドキュメント d_i の内容 ($Content_i$) からキーワードリスト $KW = \{kw_i\}$ を抽出する。形態素解析には日本語用形態素解析ソフト Mecab [20] を利用している。

$$kw_i = (MID_i, Posted_i, \{w_{ij}\}).$$

Step C 及び Step D：トピック抽出と時系列トピック推移解析

Step C では、ドキュメント集合 D におけるキーワードの共起関係に注目し、トピック抽出を行う。Step D はドキュメント集合 D を投稿日時 $Posted_i$ によって、適切な周期（日別、週別、月別など）で部分ドキュメント群 $\{X_k\}$ に分割し、トピックの時系列解析を行う。

解析手法としては、3次元行列構造に基づく時系列トピック解析 [18] 及び Latent Dirichlet Allocation (LDA, 潜在的ディリクレ配分法) モデルを用いる。ここでは、手法ごとに Step C 及び Step D で行っている処理について説明する。

1) 3次元行列構造による分析：

ここで紹介する手法は、研究者の従来研究 [18] の抜粋である。詳細は、[18] を参照されたい。

Step Cでは、文書・単語行列 M を Step B のアウトプットである $\{kw_i\}$ から生成する。ドキュメント集合 D は投稿日時 $Posted_i$ によって、適切な周期（日別、週別、月別など）で部分ドキュメント群 $\{X_k\}$ に分割されている。すなわち文書・単語行列は、時系列の文書・単語行列群 $\{M_k\}$ となる。各行列において、行は文書群を、列は Step B において抽出されたキーワード群となる。行列の各値は、内容語を抽出するのに適していると言われるスコアリング手法 RIDF [19] を用いて算出した各ワードのスコア（RIDF 値）とする。

次に、LSA (Latent Semantic Analysis) [21] を適用し、時系列に生成された文書・単語行列から隠れトピックを抽出する。LSA は文書・単語行列を次元圧縮することで、潜在的に隠れているコンセプトを解析可能なテクニックである。LSA により解析されたコンセプトを、ここでは隠れトピックと考え、時系列の文書・単語行列に適用することで、時系列の隠れトピック群を得ることができると考える。

もしも、 $\{X_k\}$ の中にトータルで m 個の文書があり、文書群 D の中にトータルで n 個のキーワードが合った場合、全体として $m \times n$ 次元の文書・単語行列ができる。行列 M_k に対する特異値分解は以下のように表される。

$$M_k = U \Sigma V^T$$

ここで U と V は正規直交行列、 $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ は対角行列であり、 $\sigma_1, \sigma_2, \dots, \sigma_n$ は特異値である。ランク $\lambda (>= 1)$ を決定し、その値に従って文書・単語行列に特異値分解を適用することで、潜在的な隠れトピック構造の抽出が可能となる。この処理は元の文書・単語行列 $\{X_k\}$ を λ ランクの基底ベクトル U_λ とすることにより、隠れ文書を抽出し、時系列の文書・単語行列をクラスタリングしていると考えられる。

隠れトピックの推移を時系列で解析するためには、静的な LSA では不十分である。隠れトピック・単語の行列を時系列に評価できる技術が必要となる。そこで Step D では、隠れトピック・単語行列の時系列集合を、3次元（隠れトピック・単語・時間軸）の行列（図3）として表現する。行列間の類似度を算出することで、トピックの時系列遷移を解析可能となると考える。

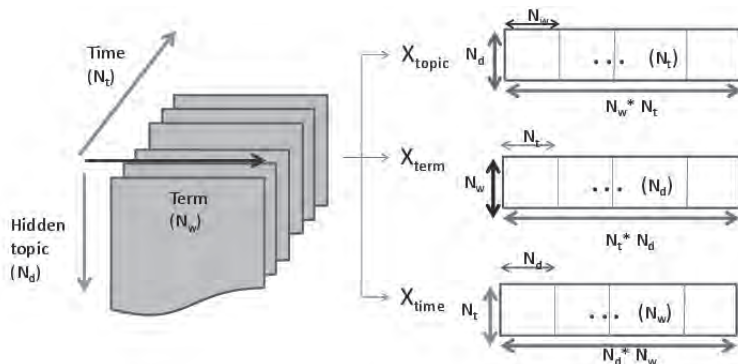


図3. 提案手法フレームワーク

2) LDAによる分析：

LDA はベイズの推定 (Bayesian inference) に基づくメジャーなトピック抽出モデルである。LDA 方式の詳細については、オリジナル論文 [3] を参照して頂きたい。この LDA 方式の探索空間は広く、計算量が大きいため、現在の多くの LDA を用いた研究では、Gibbs サンプリングを用いたランダム・ウォークによる探索が広く用いられている [22]。我々も Gibbs サンプリングを用いる計算方式を取っている。LDA 方式では、トピック数を最初に決める。一般にトピック数は5から10個とされている。あらかじめ決めたトピック数が適切であるかの判定が非常に重要となるが、トピック変化が顕著に現れるトピック数を試行錯誤で決めていくこととなる。トピック数決定後、得られたトピックを時系列に評価する。時系列ごとにクラスタリングされたドキュメント群に対し、推定トピックの適合率を求める。推定トピックの適合率から、トピックが時系列上で発生している確率を求め、時系列トピックの推移を解析する。

Step E：可視化

トピック抽出及びその時系列推移をグラフ構造等を用いて可視化する。可視化することで、どのようなトピックが発生し、それがどのように推移していったかを直感的に理解できる。手法に合わせて可視化のやり方も柔軟に対応させることとする。

V. 実験結果

STEP A では、2011年6月3日～2012年7月27日までに「番屋日報」に投稿されたメッセージをクロールし、結果として $D = 502$ メッセージを得た。表1はその一部である。STEP B では、形態素解析によりキーワードの抽出を行った。表2は抽出されたキーワード (ユニークキーワード総数2741個) の一部である。

次に得られたキーワードに対して、3次元行列構造に基づく時系列トピック解析及び LDA 手法によりトピックの分析を行った。以下では、それぞれの手法を適用した結果について述べる。

表Ⅰ. STEP A の結果の例

MID	投稿日	ブログ内容
1	2011/06/03	*物資配給場所に車で行かなくてはならないが、車がない。*店が遠すぎる。*元住んでいた場所に戻りたい。.....
2	2011/06/04	*避難所にストーブがなくて寒い。*2-3年後の将来が不安。*洋服を買うお金がない。*クルマがないので、買い物に行けない.....
3	2011/06/05	*避難所に情報が届かない。*十分な食料がない。ベッドがない。*マタニティグッズがない*トイレが十分にない。*洋服がない。*子供のおやつがない。*毎日カップヌードルばかり.....
4	2011/06/08	*家がない。仕事がない。将来が不安。*仕事に戻れない。*支援金が必要。*車が必要。移動が難しい.....
5	2011/06/10	*必要な情報が届かない。*物資の配給が不十分。*将来が不安。政府の支援, 仕事, 復旧が不十分。*情報が不足.....
6	2011/06/11	*仮設住宅についての説明がない。*収入がない。*物資の配給を継続してほしい。*政府の支援をもっと素早く。*物資が不十分。*情報が不足.....
...
15	2011/06/30	*仮設住宅に入りたい。*将来が不安。*夏用の服が必要。*年寄りでも出来る仕事があるのか.....
...

表Ⅱ. キーワード抽出の例

MID	Posted	Keywords
1	2011/06/03	車, 場所, 物資, 供給, ない, 店, 遠い
2	2011/06/04	配給, ない, 不安, 避難所, 中心, 寒い
3	2011/06/05	情報, 避難所, 中心, 食べ物, 物資, 子供
4	2011/06/08	家, 仕事, 不安, もどる, 必要, 支援金,
5	2011/06/10	物資, 配給, 将来, 支援, 情報,
6	2011/06/11	ない, 物資, 支援, 情報, 仮設, 住宅, 継続,
...
15	2011/06/30	移る, 不安, 欲しい, 仕事, 仮説, 住宅,

1) 3次元行列構造によるトピック解析の結果:

STEP Bでは、2011年12月までのブログデータにおいて、150行(文書)×2741列(単語)の文書・単語行列が生成された。この行列を時系列データとするために、月ごとにグルー

表Ⅲ. 文書単語行列の例 (2011年6月)

MID	Posted	Title	Terms					
			車	配給	避難所	ない	不安	...
1	2011/06/03	Problem	0.13	0.13	0	0.83	0	...
2	2011/06/04	Problem	0	0	0.12	0.33	0.16	...
3	2011/06/05	Problem	0	0	0.41	0	0	...
4	2011/06/08	Fear	0	0	0	0	0.31	...
5	2011/06/10	Problem	0	0.12	0	0	0	...
6	2011/06/11	Problem	0	0	0	0.18	0	...
...
15	2011/06/30	Request	0	0	0	0	0.26	...

表Ⅳ. 隠れトピック (2011年6月)

Posted	TID	Terms	Supposed Contents
Jun., 2011	0	避難所, 支援, 物資, ほしい, ない, 仮設, 不安, 生活, 継続	避難所の人々は継続的な物資の支援を求めている。
	1	避難所, 不安, 仮設, 仕事, 家, ない, 配給, 購入, 子供, 寒い	人々は仕事や配給がないこと, 寒いことなどに不安をいだいている。
	2	仮設, 欲しい, 家, 行く, 買い物, 行事, 開く, 移動, 購入, 残る	人々は仮設住宅に写りたがっており, 買い物や行事に出かけたい。
	3	支援, 欲しい, 理解, 仕事, 見つける, 不安, 生活, 仮設, 物資, 将来	人々は仕事を見つけたがっており, 支援を必要としている。将来への不安がある。
	4	欲しい, 避難所, 食べ物, 子供, 政府, 対策, 悪い, 十分	人々は政府の食料や子供に対する対策が悪いと感じており, もっと欲しいと願っている。

ピングする。表3は2011年6月の文書・単語行列(抜粋)である。各セルの値は、算出された各単語のスコア(RIDF値)である。

次に我々は、特異値分解により、 $\lambda (>= 1)$ ランクの次元圧縮を行い、各月ごとの隠れトピックを行列 U_λ を求めた。表4は2011年6月の隠れトピックの一覧である。表4において、TIDは隠れトピックID、Termsは各トピックの特徴的な単語10種を表している。Supposed Contentsは我々が判断したトピックの内容である。2011年6月においては、人々の物資や支援、仮設住宅、仕事などに対する要望や、仮設住宅の気温や政府対応に関

する不満といった隠れトピックが抽出されているとの結果になった。

2011年6月から12月の7ヶ月間では、39の隠れトピックが発見され、以下の12種類のトピック遷移が見られることを確認した。

- T1：物資を供給してほしい
- T2：仕事を見つけない
- T3：避難所が不満。仮設住宅へ移りたい
- T4：政府の対策が不満
- T5：金銭的なサポートがほしい
- T6：移動手段が必要
- T7：新しい住宅（仮設ではなく）に移りたい
- T8：仮設住宅は不便、不公平
- T9：車がほしい
- T10：メンタルケアを受けたい
- T11：将来が不安
- T12：家族と住みたい

図4は6月から12月の隠れトピックの遷移の様子を示している。各行がトピックのタイプを示し、列が各月を示している。色のついたセルは、その月にそのトピックが発生したことを示している。色が濃いほど、同じ内容を示す隠れトピックが多かったことを示している。図4から以下を考察することができる。

- T1（物資を供給してほしい）及びT2（仕事を見つけない）は被災者にとって基本的なニーズである。このニーズは長期間に渡って現れている。図4において、下部に示された矩形がこのニーズの推移を示している。

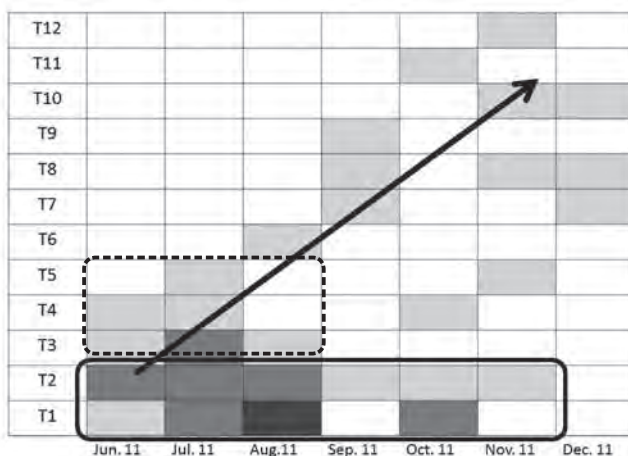


図4. 被災者のニーズ変化（2011年6月12月）

- T3（仮設住宅へ移りたい）、T4（政府の対策が不満）and T5（金銭的なサポートがほしい）は比較的早い時期に現れているニーズである。これらは被災者にとって直近の課題であったためと思われる。図4において、点線で示された矩形が、このニーズの推移を示している。
- 時間が経つに連れ、人々のニーズは徐々に変遷してきている。たとえば、T7（新しい住宅（仮設ではなく）に移りたい）、T8（仮設住宅は不便、不公平）、T9（車がほしい）、T10（メンタルケアを受けたい）、T11（将来が不安）and T12（家族と住みたい）は時間が経つに連れ現れてきている。図4において、右斜め上方向に向かって矢印がこのニーズの変遷を示している。これは、状況が落ち着くに連れ、人々のニーズが変わってくることを示している。

2) LDAによる時系列トピック解析の結果：

次にLDA手法によって、トピックを抽出し、時系列での評価を行った。トピック数は5とし、25回のイテレーションを実施した。表5は、STEP Cで抽出された5つのトピックと、そこに属する主なキーワードのリストとキーワードから想定される内容である。5つのトピックに属する主なキーワード群から、以下の様な内容のトピックが抽出されたと考えられる。

- (1) 仮設住宅への避難、情報や行政への不安、
- (2) 仮設住宅の生活の不便さ、仕事や情報に対する欲求、
- (3) 用品や食料の継続、生活への希望、
- (4) 生活再建や仕事、健康への不安、
- (5) 孫や子供、イベント開催、心のケア、

上記の抽出結果について、パープレキシティと呼ばれる言語モデルの評価パラメータを算出する。一般にパープレキシティは次式で計算される。

$$PP = 2^{H(p)}, H(p) = -\frac{\sum_x \log_2 p(X)}{N}$$

ここで $X = \{w_1 \dots w_N\}$ は評価用のテキスト集合である。上記では各語の生起確率をもとめ、そのすべての情報量の平均を exp の係数としている。すべての語の生起確率を求めするには、トピックの確率×そのトピックにおける語の生起確率を用いる。パープレキシティ値が高いほど単語の特定が難しく、言語として複雑であると言われている。今回の実験において、パープレキシティの値は 3.58 となった。一般にパープレキシティが10以上となると認識困難な文書となると言われており、ターゲットブログの文書集合が言語モデルとして十分機能していると言うことが考えられる。

次にSTEP Dにおいて、上記5つのトピックの月ごとの発生状況を解析し、STEP Eで可視化を行った。表6は5つのトピックの月ごとの発生率であり、図5は時系列の可視化の結果である。

トピック（1）は震災後初期（6－8月）の頃に多く見られている。トピック（2）は

表V. 抽出されたトピック

#	Keywords	Supposed Contents
1	仮設, 避難, 町, 不安, 市, 住宅, 盛岡, 仕事, 情報, 何, 大槌, 山田, 人, 行政, 配給, 釜石, 金, 車, 子供, 時	仮設への避難, 情報や行政への不安,
2	仮設, 仕事, 生活, 不便, 避難, 情報, 年, 家, 不安, ため, 居住, 地元, 住宅, 入居, 子供, 家族, 自宅, 宅, 内陸, 車	仮設の生活の不便さ, 仕事や情報に対する欲求,
3	用品, 継続, 食料, 沿岸, 生活, 希望, 日, 声, 子ども, 今, 等, おむつ, アンケート, 被災, 用, 内訳, 消耗, 夏物, 個別, 衣	用品や食料の継続, 生活への希望,
4	生活, 仕事, 自宅, 沿岸, 仮設, 再建, 関係, 不安, 家, 声, 健康, 今後, 内訳, 関連, 居住, 子ども, 住宅, 内陸, 事, 趣味	生活再建や仕事, 健康への不安,
5	孫, 子ども, イベント, 仮設, ケア, 心, 義援金, 友人, 開催, 配送, 収集, 番屋, 家族, 仕事, 募集, 会, 事, 知人, 人, 子供	孫や子供, イベント開催, 心のケア,

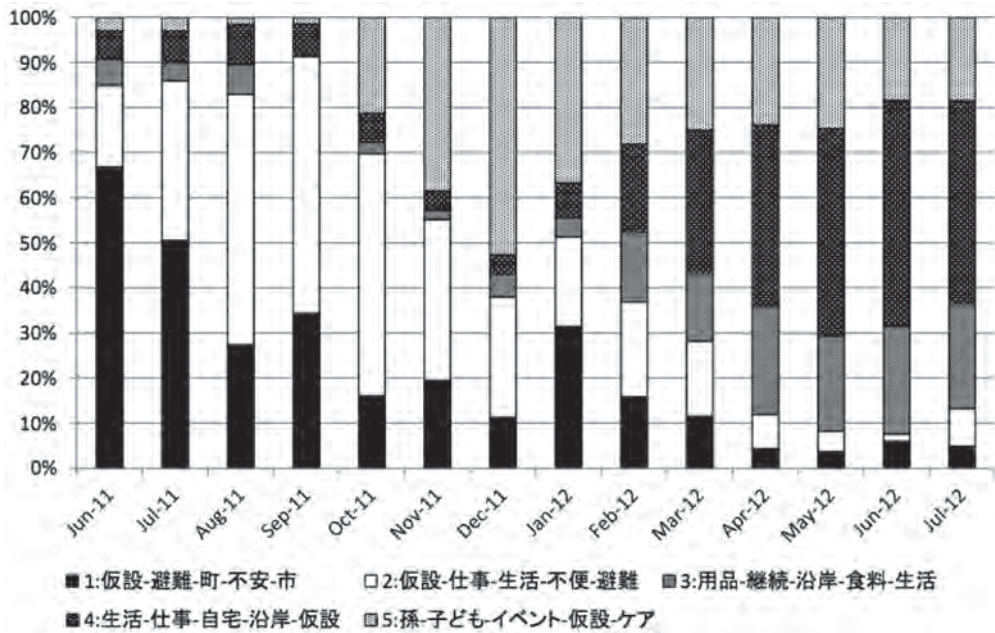


図5. トピックの時系列推移

2011年の8月～11月に増加している。トピック（3）は2012年になってから発生している。トピック（4）も2012年になってから多く見られるようになっており、特に2012年3月以降に増えている。トピック（5）は2011年11月から増加しており、特に年末年始に多く見られている。これはトピック（1）が、震災発生後まだ日が浅い時期には、多くの人が体育館などの避難所に滞在しており、仮設住宅へ避難したいというニーズや情報がなかなか手に入らないという不満・不安を抱いていたことを示していると考えられる。またトピッ

表Ⅵ. TOPIC OCCURRENCE RATIO BY MONTH

period	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Jun-11	0.67	0.18	0.06	0.06	0.03
Jul-11	0.51	0.35	0.04	0.07	0.03
Aug-11	0.27	0.56	0.06	0.09	0.02
Sep-11	0.34	0.57	0.00	0.07	0.01
Oct-11	0.16	0.54	0.02	0.07	0.21
Nov-11	0.20	0.36	0.02	0.04	0.38
Dec-11	0.11	0.27	0.05	0.04	0.53
Jan-12	0.31	0.20	0.04	0.08	0.37
Feb-12	0.16	0.21	0.15	0.19	0.28
Mar-12	0.12	0.17	0.15	0.32	0.25
Apr-12	0.04	0.08	0.24	0.40	0.24
May-12	0.04	0.05	0.21	0.46	0.25
Jun-12	0.06	0.02	0.24	0.50	0.18

ク（２）は仮設住宅に移動後のニーズであると考えられる。仮設に移動したものの、仮設の生活が不便であり、相変わらず情報が不足しており、仕事に対する不安も継続していることが想定される。トピック（３）は、2012年になってからのものであり、用品や食料支援の継続に対するニーズではないかと考える。震災後、1年近くが経過する中で、さまざまな支援が打ち切れつつあった状況にあり、被災者が支援の継続を願っていたのではないかと想像する。トピック（４）は2014年3月以降に増加しているトピックであり、震災から1年が経過し、生活再建や仕事の再開などに対して、被災者が不安を感じていることを示していると考えられる。さらにトピック（５）は年末年始に増加しているトピックである。これは年末年始という時期に、孫や子供と過ごす時間やイベントに対する要望や、震災から8ヶ月ほどが経過し、心のケアの必要性を実感し始めているということを示していると考えられる。

Ⅵ. 抽出ニーズの有効性と意思決定ツールとしての可能性の考察

図5で示したトピックの時系列推移結果が実際の被災者の状況変化やニーズ変化を捉えているかの評価を行う。評価にあたっては、NPO法人SAVE IWATE [1]の方々からの協力を仰いだ。

表7は震災発生後に実際に起きた事象や観察された状況をSAVE IWATEの方々から提供いただき、月別にまとめたものである。3月11日の震災発生後、多くの人々は避難所に移動した。中には2次避難として、別の避難所に移動する人もいた。仮設住宅への引ッ

越しが始まったのは7月に入ってからである。7－8月の約2ヶ月で、多くの人々が仮設住宅へ移動したとのことであった。9月に入ると、失業保険の失効により、多くの被災者が将来やしごとへの不安を感じるようになった。また冬を迎え、寒さが身にしみてくるようになると、被災者には仮設住宅に対する不満が多く見られるようになった。仮設住宅はクオリティが一定しておらず、低品質の仮設住宅に移動した人を中心に多くの不満が起きた。冬製品に対するニーズも見られたとのことである。2012年度3、4月頃は、新学期、新入学の時期ということもあり、被災者は前向きな気持ちを抱くとともに、自立再建の希望と不安を感じるようになった。さらに6月には支援センターの閉鎖が決まり、多くの被災者が支援継続に対して不安を感じていたとのことである。

上記の実際に起きた事象や観察された状況を図5の上に重ねたものが図6である。トピックの大きな変化が起こるタイミングと、被災者の状況変化に関連があり、被災者の状況変化がトピックの変化を引き起こしていると考えられることができる。

抽出されたニーズは、被災者の状況変化をよく表しており、被災者の気持ちの変化を確認することが可能である。たとえば仮設住宅に対する不満が発生していることや、年末年始に対する期待、メンタルケアの必要性など、被災者の気持ちの変化を掴むことで、適切

表Ⅶ. 被災者の状況の推移

月	観察された被災者の状況	実際に起きた事象
Mar-11	避難所へ移動	震災発生
Apr-11		
May-11	2次避難（避難所間の移動）	
Jun-11	2次避難（避難所間の移動）	
Jul-11	仮設住宅への移動開始	
Aug-11	ほぼ全員の被災者が仮設住宅への移動終了	お盆
Sep-11	失業保険が切れる→仕事への不安	
Oct-11		
Nov-11	寒さに対する不満	冬到来
Dec-11	内陸部の仮設住宅の寒さに対する不満	年末
Jan-12	家族と過ごす、冬用品に対するニーズ	正月
Feb-12		
Mar-12	新生活への期待	卒業
Apr-12	新生活への期待、自立再建への希望と不安	新入学、新学期
May-12		支援所の閉鎖通知
Jun-12	支援の継続の要望	支援所の閉鎖通知
Jul-12		衣料品支援打ち切り

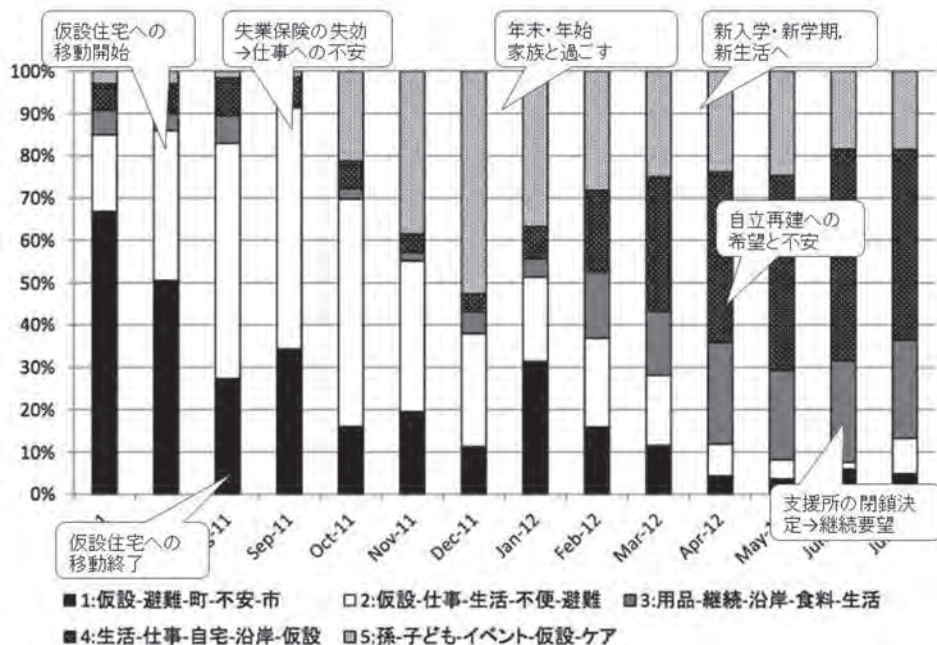


図6. トピックの時系列推移と被災者の状況

な支援を行っていくための参考とすることが可能となると考えられる。支援センターの閉鎖がアナウンスされた時に、支援継続に対する不安が出てくるなどの反応を知ることで、被災者の不安を少しでも減らせるような活動（支援を継続して受けるための方法の告知など）につなげていくことができると考える。今回の手法により得られた結果が、支援を提供するNPO法人等にとっても有効な情報となり、意思決定支援ツールとして利用できる可能性があると言える。

VII. まとめ

被災者の声が記載されたブログデータを対象とし、大震災発生後、被災者のニーズ・関心がどのように変化したかを3次元行列構造によるトピック解析やLDA手法を用いて、時系列で解析した。被災者のニーズをトピックとして捉え、時系列にトピック抽出を行うことで、被災者のニーズがどのように変化していったかを可視化することができた。

さらに得られた結果を、NPO法人の担当者とともに評価を行い、実際に起きた事象や、観察された被災者の状況と比較することで、LDAにより抽出されたニーズが被災者の状況変化を良く表していることが確認できた。それとともに、本手法が被災者への適切な支援を行うための意思決定に有効であるかについて考察も行った。

今後は、他のデータ・ソースを対象とした実験を実施するとともに、得られたデータを震災後の情報としてNPO法人の方々と協力し、アーカイブしていく予定である。また支

援にあたっての意思決定ツールとしての可能性をより探索するために、ほぼリアルタイム（日毎）にトピック変化を感知していく手法についても検討を行っていく。

ACKNOWLEDGMENT

被災時の状況調査、データ提供、解析結果の評価に対して、多大なるご協力をいただいたNPO法人SAVE IWATEの皆様にご心より感謝いたします。本研究はJSPS科研費24500180、及び千葉商科大学 学術研究助成金の助成を受けたものです。

REFERENCES

- [1] SAVE IWATE, <http://sviwate.wordpress.com/in-english/>
- [2] Bannya Nippou, <http://sviwatebanya.wordpress.com/>
- [3] D. M. Blei, A. Y. Ng, and M. I. Jordan, *Latent dirichlet allocation*, *Journal of Machine Learning Research*, vol.3, pp.993-1022, 2003.
- [4] T. L. Griffiths and M. Steyvers, *Findings cientific topics*, *Proceedings of the National Academy of Sciences*, vol.101 (Suppl.1), pp.5228-5235, 2004.
- [5] Y. Sekiguchi, H. Kawashima and T. Uchiyama, *Discovery of related topics using series of blogsites'entries*, Proc. of JSAI2008, 2I1-1, 2008.
- [6] S. Asur, B. A. Huberman, G. Szabo and C. Wang, *Trends in socialmedia: Persistence and decay*, ICWSM2011, pp.434-437, 2011.
- [7] K. Radinsky, E. Agichtein, E. Gabrilovich and S. Markovitch, *A word at a time: Computing word relatedness using temporal semantic analysis*, WWW2011, pp.337-346, 2011.
- [8] Y. Wang, E. Agichtein and M. Benzi, *TM-LDA: efficient online modeling of latent topic transitions in social media*, KDD2012, pp.123-131, 2012.
- [9] M. Pennacchiote and S. Gurumurthy, *Investigating Topic Modelsfor Social Media User Recommendation*, WWW2011, pp.101-102, 2011.
- [10] R. Bellman and R. Kalaba. *On adaptive control processes*, IRE Transactionson Automatic Control, 4, pp.1-9, 1959.
- [11] Z. Ding, C. Isaac, Z. Hongyuan and C. L. Giles, *Discovering temporal communities from social network documents*, ICDM2007, pp.745-750, 2007.
- [12] Q. Jiangtao, L. Zhangxi, T. Changjie and Q. Shaojie, *Discovering organi-zational structure in dynamic social network*, ICDM2009, pp.932-937, 2009.
- [13] D. Greene, D. Doyle and P. Cunningham, *Tracking the Evolutionof Communities in Dynamic Social Networks*, Prof. of ASONAM2010, pp.176-183, 2010.
- [14] P. Jaccard, *The distribution of flora in the alpine zone*, New Phytologist, 11 (2), pp.37-50, 1912
- [15] T. Hashimoto, T. Kuboyama and Y. Shirota, *Detecting Unexpected Correlation between a Current Topic and Products from Buzz Marketing Sites*, Proc. of the

- DNIS2011, pp.147—161, LNCS7108, Springer, Heidelberg, 2011
- [16] T. Hashimoto, T. Kuboyama and Y. Shirota, *Rumor Analysis Framework in Social Media*, Proc. of IEEE TENCON2011, pp.329-333, 2011.
- [17] S. Higuchi, T. Hashimoto, T. Kuboyama and K. Hirata, *Exploring Social Context from Buzz Marketing Site-Community Mapping Based on Tree Edit Distance*, Proc. of PerCol 2013 (Fourth International Workshop on Pervasive Collaboration and Social Networking) (2013), pp.187-192, 2013.
- [18] T. Hashimoto, B. Chakraborty, T. Kuboyama and Y. Shirota, *Temporal Awareness of Needs after East Japan Great Earthquake using Latent Semantic Analysis*, Proc. of EJC2013 (23rd European-Japanese Conference on Information Modelling and Knowledge Bases) (2013), pp.214-226, 2013.
- [19] Church, K. W. Church, and W. A. Gale, *Poisson mixtures*, Natural Language Engineering 1, pp.163-190, 1995.
- [20] T. Kudo, *MeCab: Yet Another Part-of-Speech and Morphological Analyzer*, [Online] Available: <http://mecab.sourceforge.net/>, [retrieved: Jan., 2014]
- [21] T. K. Landauer and S. T. Dutnais, *A solution to Platofs problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge*, Psychological Review, 104 (2), pp.211-240, 1997.
- [22] T. L. Griffiths and M. Steyvers, Finding scientific topics, Proceedings of the National Academy of Sciences, vol.101 (Suppl.1), pp.52285235, 2004.

(受理日：平成26年7月23日)

(校了日：平成26年9月10日)

〔抄 録〕

2011年3月11日に発生した東日本大震災により、多くの人々が自宅や仕事、家族との生活を失い、きわめて困難な状況に陥った。行政機関やNPO法人は、物資の提供、避難所の設置、仮設住宅の建設など種々の支援を行ったが、一方で、それらの支援が適切に提供されていたかを確認することは容易ではなかった。大震災後、被災者のニーズがどのように変化していったかを解析することは、被災者支援を振り返り、評価するためにも重要であると考えられる。そこで本論文では、被災者の声が記載されたログデータを対象とし、データマイニング技術を用いて、大震災発生後、被災者のニーズ・関心がどのように変化したかを時系列で解析する手法を提案する。解析手法としては、3次元行列構造に基づく時系列トピック解析及びLatentDirichletAllocation (LDA) によるトピック抽出技法を用いる。得られた結果（ニーズ変化）を被災者の実際の状況と比較することで、本論文の手法により抽出されたニーズの有効性について評価する。さらに本論文の本手法が被災者への適切な支援を行うための意思決定支援に効果があるかについて、NPO法人へのインタビューも行い考察する。