

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 17 日現在

機関番号：32504

研究種目：基盤研究(C)

研究期間：2012～2014

課題番号：24500180

研究課題名(和文) ソーシャルメディアの横断的解析に基づく東日本大震災の評判分析

研究課題名(英文) Reputation Analysis related to the East Japan Great Earthquake by Cross Media Analysis

研究代表者

橋本 隆子 (Hashimoto, Takako)

千葉商科大学・商経学部・准教授

研究者番号：80551697

交付決定額(研究期間全体)：(直接経費) 4,100,000円

研究成果の概要(和文)：ソーシャルメディアにおける評判形成プロセスの明確化を目的として、東日本大震災に関するブログ、動画サイト、口コミサイト、マイクロブログ(Twitter)上の書き込みを横断的に解析する評判分析を行った。NPO法人から提供を受けたブログに対し、LSA、LDA等のデータマイニング手法を活用し、時系列の話題抽出を行い、被災者のニーズの変化を可視化した。さらに動画サイトに対して、トリガとなる重要単語を抽出し、それがマイクロブログ上でどのようにBurstしたかの解析も行い、ソーシャルメディア間の関係を示した。さらにタイとのプロジェクトも開始し、よりグローバルな横断解析に取り組んでいる。

研究成果の概要(英文)：The aim of this research was to conduct the cross media research for blogs, web video sites, micro-blogs (twitter), online bulletin boards related to the East Japan Great Earthquake, to make clear the reputation formation process in social media. We analyzed and visualized the damaged people's requirements transition from blogs provided by NPO in Iwate Prefecture, using the data mining techniques such as Network Community Graph, LDA, LSA, and so on. From the web video sites, we extracted trigger words and explored how those words activated burst micro blogs. The results were analyzed as cross media relationships. In addition, as the part of cross media analysis, we started the new collaborative research with Thai to analyze the relationships between Thai and Japan social media.

研究分野：ソーシャルメディア解析

キーワード：ソーシャルメディア解析 データマイニング ビッグデータ 話題抽出

1. 研究開始当初の背景

2011年3月11日、関東・東北地方において東日本大震災が発生した。電気・ガス・電話等のライフラインが寸断される中、人々はソーシャルメディアを通じて震災に関する情報を取得し、その行動はソーシャルメディアの情報に大きく影響をうけることとなった。東日本大震災のような大災害が発生した際に、ソーシャルメディア上の評判がどのように生まれ、広まっていくのか、その様子を解析し可視化を行うことは、大災害後の人々の反応分析及びそのモデル化において、きわめて重要であると考えられる。従来の研究が主として単一メディアを対象とした評判分析を行っているのに対し、本研究では動画サイトブログ、Twitterなどの複数のソーシャルメディアを対象とし、横断的に解析を行うことを特徴とした。

2. 研究の目的

ソーシャルメディアにおける評判形成プロセスの明確化を目的として、東日本大震災における動画サイト、口コミサイト、マイクロブログ上の書き込みを横断的に解析する評判分析を行う。本研究では、まず1) メディアごとにそこで議論されている話題を構造化、可視化する(話題抽出)、次に2) その結果として得られた話題を横断的に時系列で比較・評価する(横断解析)ことで、ソーシャルメディアから話題が生まれ、評判として形成されていく様子を示す。本研究により大災害後の話題の推移やその関係が把握できるようになり、大災害後の人々の反応分析及びそのモデル化へ貢献できると考える。

3. 研究の方法

本研究の実施にあたり、我々は、図1、図2のようなフレームワークを提案し、そのフレームに従って研究を行った。以下、各ステップについて説明する。

(1) Step A: データ収集 (図1)

口コミ掲示板、動画共有サイト、Twitterなどのソーシャルメディアから、東日本大震災に関する対象データを選定し、ユーザの書き込みを収集する。

(2) Step B: 言語処理 (図1)

Step Aの収集結果に対して、形態素解析技術を用いてキーワードを抽出する。

(3) Step C: 話題抽出 (図1)

Step Bの収集結果に対し、ネットワークコミュニティグラフ、LSA(行列分解)[1]、LDA(言語モデル)[2]による手法などを用いて、それぞれのデータの特性に応じた話題を抽出する。

(4) Step D: 時系列話題推移解析 (図1)

Step Cで抽出された話題の結果をベースに、

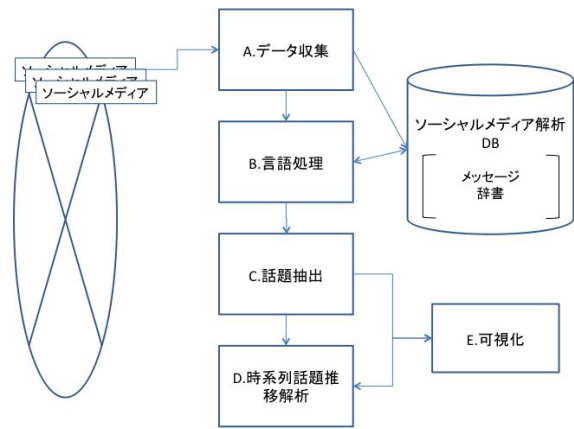


図1 提案フレームワーク(話題抽出)

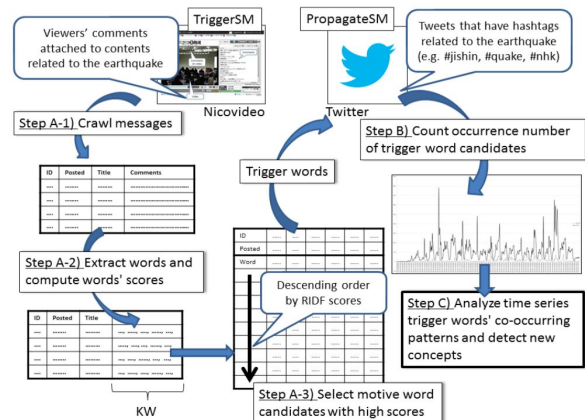


図2 提案フレームワーク(横断解析)

時系列上での推移を観察。グラフ編集距離[3]による手法や、各話題の類似度算出による手法などを用いて、話題の時系列変化を分析する。

(5) Step E: 可視化 (図1)

話題抽出及びその時系列推移をグラフ構造等により可視化する。

(6) 横断分析 (図2)

Step C及びDの結果をベースに、ソーシャルメディアを横断的に解析する。ソーシャルメディア上での評判形成のきっかけをつくるソーシャルメディア、それを受けて反応(Burst)するソーシャルメディアとしてモデル化し、それぞれの関係を分析する。

(7) 評価

分析を行うだけでなく、実際に大震災の被災地で活動された方々に対して解析結果を開示し、共に評価を行う。

4. 研究成果

各ステップにおける結果を述べる。

(1) Step A: データ収集

対象ソーシャルメディア及び取得件数は表1のとおりである。

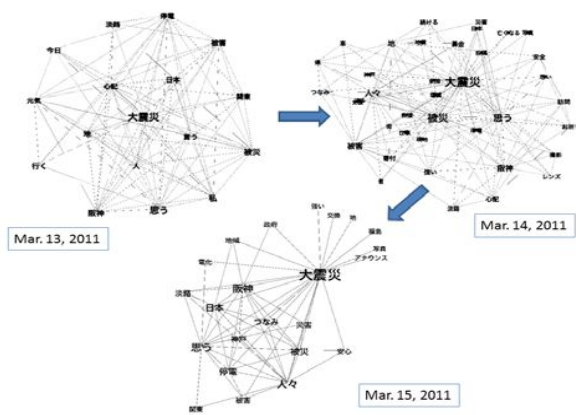


図 3 ネットワークコミュニティグラフによる話題抽出

(2) Step B : 言語処理

形態素解析には、日本語用形態素解析ソフト Mecab[4]を用いた。震災関連、原発関連等の用語に対して、適切な用語・辞書を用いて、形態素解析のための辞書トレーニングも行った。

結果としてNPO 法人 SAVE IWATE 提供の東日本大震災の被災者のブログからは、約 3000 ワードのキーワードを抽出した。また 2 億件強の Twitter データ (ホットリンク社提供) からは、1 日毎に約 23 万~25 万件のキーワードを抽出した。

表 1 対象ソーシャルメディア一覧

ID	対象ソーシャルメディア	件数
1	大震災に関して言及している kakaku.com 上の書き込み (2011 年 3 月 11 日~4 月 5 日)	約 500
2	大震災の被災者のブログ(NPO 法人 SAVE IWATE 提供)(2011 年 6 月~2012 年 7 月)	約 750
3	大震災に関連した Twitter データ ((株)ホットリンク社提供)(2011 年 3 月 9 日~29 日)	約 2 億
4	ニコニコ動画で生放送配信された政府・東京電力の記者会見映像及びユーザコメント (2011 年 3 月 11 日~24 日)	約 97000 (67 コンテンツ)
5	タイ国ソーシャルメディア Pantip, Kapook, Sanook (2011 年 3 月)	約 3000

(3) Step C : 話題抽出

表 1-# 1 のデータに対してネットワークコミュニティグラフ手法によって話題抽出した結果を示す(図 3)。図 3 において、各ノードはキーワードを、エッジは共起を示す。グラフ構造から話題を抽出するために、ネットワーク構造の分割の「質」を表現する指標で「モジュラリティ[5]」を用いた。また表 2、表 3 はそれぞれ表 1-# 2 のデータに対する話題抽出結果である(表 2 LSA) 表 3 LDA)。

(4) Step D : 時系列話題推移解析

図 3 の話題抽出結果に対して、話題の階

層構造を根付きラベル付き無順序木とみなし、グラフ編集距離を用いて、時系列推移を解析した。図 4 はその手法を示した図であり、図 5 は結果として得られた話題の時系列推移である(「(5) Step E : 可視化」処理を含む)。階層構造を時系列で並べ、各階層における編集距離を算出して、話題推移を判定する。表 1-# 1 のソーシャルメディアにおける話題として、「被災者に義援金を送る」、「地震と津波による被害」、「被災者へ祈り」といった話題が継続しており、「計画停電」や「車の生産・納品」に関する話題が時折現れたことがわかった。

LSA から得られた結果(表 2)に対しては、ある月の各トピックと次の月の各トピック間の類似度を算出し、しきい値以上の類似度をもつトピックを時系列上で推移したと判定した。

LDA の解析結果(表 3)については、Step C で得られた話題が各月ごとでどれくらいの割合を占めるかを算出し、時系列推移として可視化((5) Step E) した(図 6)。

表 2 LSA による話題抽出結果 (被災者ブログより、2011 年 6 月)

ID	キーワード	内容
1	避難所, 支援, 物資, ほしい, ない, 仮設, 不安, 生活, 継続	継続的な物資の支援
2	避難所, 不安, 仮設, 仕事, 家, ない, 配給, 購入, 子供, 寒い	仕事や配給がないこと, 寒いことなどへの不安
3	仮設, 欲しい, 家, 行く, 買い物, 行事, 開く, 移動, 購入, 残る	仮設住宅に移りたがっており, 買い物や行に出かけたい。
4	支援, 欲しい, 理解, 仕事, 見つける, 不安, 生活, 仮設, 物資, 将来	仕事を見つけない, 支援を必要としている。将来への不安。
5	欲しい, 避難所, 食べ物, 子供, 政府, 対策, 悪い, 十分	政府の食料や子供に対する対策が悪い, もっと欲しいと願っている。

表 3 LDA による話題抽出結果 (被災者ブログより、2011 年 6 月-2012 年 7 月)

ID	キーワード	内容
1	仮設, 避難, 町, 市, 不安, 情報, 行政, 大槌, 盛岡, 配給, 釜石	仮設へ避難したい。市や行政に対する思い, 不安,
2	仮設, 仕事, 生活, 不便, 避難, 家, 居住, 地元, 住宅, 入居	生活の不便さ, 仕事や家に対する欲求がある。
3	用品, 継続, 食料, 沿岸, 生活, 希望, 日, 声, 子ども, おむつ, 被災, 内訳, 消耗, 夏物	おむつなどの具体的な用品への要望や食料等の支援の継続を願っている。
4	生活, 仕事, 自宅, 沿岸, 仮設, 再建, 不安, 家, 健康, 今後, 子ども, 住宅	生活再建や仕事への思いが現れている, 健康への不安も感じている。
5	孫, 子ども, イベント, 仮設, ケア, 心, 義援金, 友人, 開催	孫や子供に対する思い, イベント開催, 心のケアへの欲求がある。

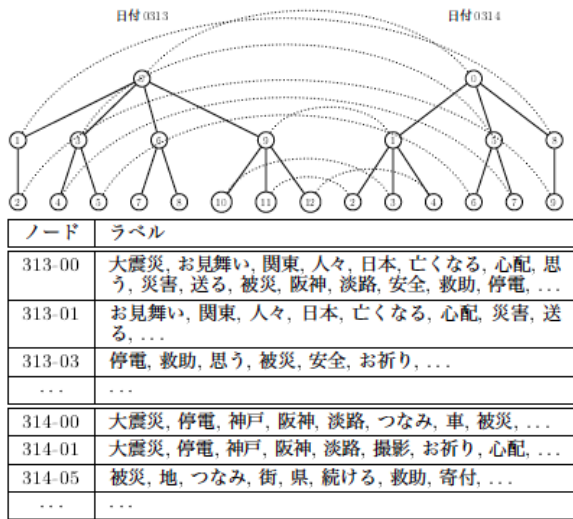


図 4 グラフの階層構造に基づく時系列推移解析方法

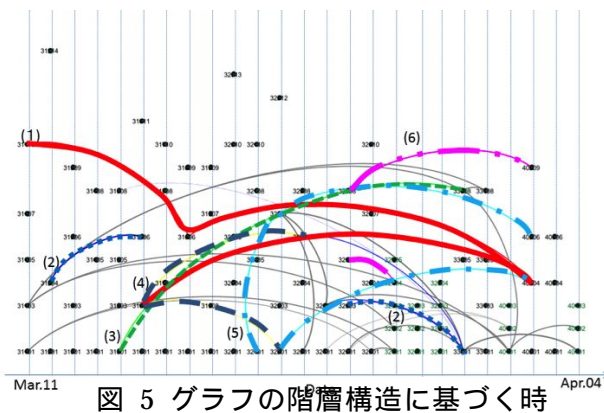


図 5 グラフの階層構造に基づく時

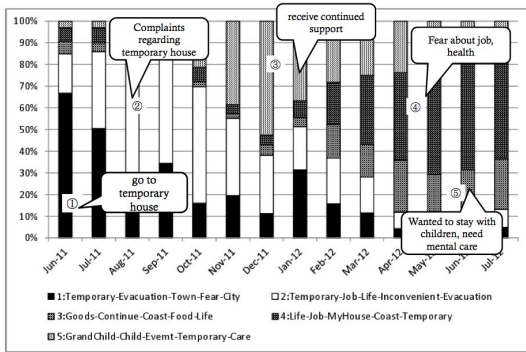


図 6 東日本大震災の被災者のニーズ変化 (LDA 手法)

(5) Step E : 可視化

図 2, 図 4, 図 5, 図 6 等が可視化の結果である。この他にも動画等を用いて, グラフ構造が推移していく様子の可視化等を行った。

(6) 横断分析

評判形成のきっかけとなるソーシャルメディアを表 1-#4 のニコニコ動画のコメントとし, それを受けて反応するソーシャルメディアを表 1-#2 の Twitter データとして, 横

断解析を行った。手法としては, shifted wavelet 木を利用した柔軟なバースト検知手法[6] を用い, ニコニコ動画のコメント上で解析された特徴的なキーワードが, Twitter 上でも同時刻に発生しバーストしているかどうかを解析した。結果として, 以下のような話題がニコニコ動画の生放送と Twitter データ上で現れ, 評判として生成されたのではないかということがわかった。

- ・ 福島原発から逃げるという話題
- ・ 東京電力の社長, 副社長, 部長, 等に対する不平, 不満
- ・ 消防隊, 自衛隊の原発における活動に関する話題
- ・ 原発における放射能, 電源対策等に関する話題

横断解析の一貫として, 表 1-#5 で示したタイ国のソーシャルメディアデータに対しても話題抽出フレームワーク(図 1)を適用し, 日本のソーシャルメディア上で発生した話題との比較を行った。図 7 はタイ国のソーシャルメディア上の話題と日本のソーシャルメディア(表 1 の #1 と #2)上の話題の相関を示した図である。タイ国のソーシャルメディアにおいても, 表 1#-1 のソーシャルメディアで見られたような, 「被災者に義援金を送る」, 「地震と津波による被害」といった話題が発生していたことが確認できた。これらの話題は, 日本における話題と同じく, 震災に対する反応として認識することができる。

(7) 評価

図 6 に示した結果について, これらが被災者の状況変化やニーズ変化を適切に表しているかの評価を行った。岩手県の NPO 法人 SAVE IWATE の方々に結果を見ていただき, 実際に発生した出来事とも関連付けて確認を行った。結果として, 話題の大きな変化が起こる

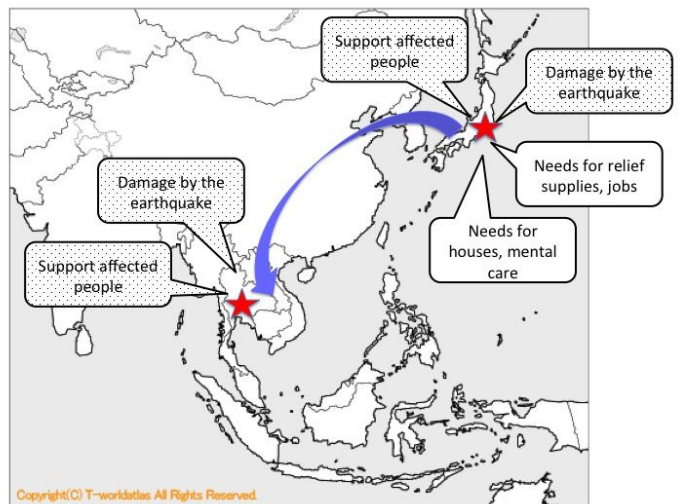


図 7 タイ国と日本のソーシャルメディア上の話題の関連

タイミングと、被災者の状況に関連があり、被災者の状況変化が話題の変化を引き起こしていることがわかった。支援を行う NPO にとっても、こうした情報は意思決定の参考となる有用な情報になりうると考えられる。

5. まとめ

ソーシャルメディアからの話題抽出は、性能面でも、精度の面でも課題が多く、結果としてソーシャルメディアからの話題抽出とその時系列分析に研究期間の多くのリソースを費やすこととなった。ソーシャルメディアデータに対する時系列話題分析は、今後、2億件強の Twitter データに正面から取り組み、高性能・高精度の話題抽出手法を開発することで、継続して取り組んでいく。またソーシャルメディアの横断分析についても、これまでに得られた Twitter, Blog, オンライン掲示板、さらにタイ国のソーシャルメディアデータの結果などを活用し、引き続き解析・可視化を行っていく予定である。

<引用文献>

- [1] Landauer, T. K., Foltz, P. W., and Laham, D., *An introduction to latent semantic analysis*, Special Issue: Quantitative Approaches to Semantic Knowledge Representations, Vol. 25, Issue 2-3, 1998
- [2] Blei, D. M., Ng, A. Y., and Jordan, M. I., *Latent dirichlet allocation*, J. of Machine Learning Research, vol. 3, pp. 993-1022, 2003.
- [3] Tai, K. C., *The tree-to-tree correction problem*, J. of ACM 26, 422-433, 1979.
- [4] Mecab, <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>
- [5] Newman, M. E. J., *Modularity and community structure in networks*, Proc. Of National Academy of Science USA 103(23), 8588-8696, 2006.
- [6] Zhu, Y., Shasha, D., *Efficient elastic burst detection in data streams*, Proc. Of KDD 2003, 336-345, 2003.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 5 件)

Takako Hashimoto, Basabi Chakraborty and Yukari Shirota, *Developing a Framework for an Advisory Message Board for Female Victims after Disasters*, J. of Digital Scholarship in the Humanities, 16 pages, 2015 (to appear). 査読有

Takako Hashimoto, Basabi Chakraborty and Yukari Shirota, *Topic Transition Detection about the East Japan Great*

Earthquake based on Emerging Modularity Over Time, Int. J. Computational Science and Engineering, 1-16, 2015 (to appear). 査読有

橋本隆子, 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, 千葉商大紀要 52(1), 127-144, 2014-09, 2014. 査読無
<http://ci.nii.ac.jp/naid/110009837238>

Takako Hashimoto, Tesuji Kuboyama and Yukari Shirota, *Topic Detection about the Great Japan East Earthquake based on Emerging Modularity*, Information Modelling and Knowledge Base XXIV, IOS Press, 110-126, 2013. 査読有
DOI: 10.3233/978-1-61499-177-9-110

Takako Hashimoto, Basabi Chakraborty and Yukari Shirota, *Social media analysis - determining the number of topic clusters from buzz marketing site*, Int. J. Computational Science and Engineering, Vol. 7, No. 1, 65-72, 2012. 査読有
DOI: 10.1504/IJCSE.2012.046181

[学会発表](計 20 件)

Takako Hashimoto, Basabi Chakraborty, Supavadee Aramvith and Tetsuji Kuboyama, *Affected People's Needs Detection after the East Japan Great Earthquake - Time Series Analysis using LDA*, The Asia-Pacific Signal and Information Processing Association, Dec. 10, 2014, Siem Reap (Cambodia).

Takako Hashimoto, Supavadee Aramvith, Teeranoot Chauksuvanit and Yukari Shirota, *Framework for Language Independent Social Media Analysis Platform to detect Reactions on Global Topics*, TENCON2014, Oct. 24, 2014, Bangkok (Thailand).

Takako Hashimoto and Yukari Shirota, *Framework of an Advisory Message Board for Women Victims after Disasters*, The 6th Joint International Conference of the Alliance of Digital Humanities Organizations, July 9, 2014, Lausanne (Switzerland).

Takako Hashimoto, Supavadee Aramvith, Teeranoot Chauksuvanit and Yukari Shirota, *Human Reaction in Thailand based on Social Media Analysis after the East Japan Great Earthquake*, The 24th International Conference on Information Modelling and Knowledge Bases, Jun. 4, 2014, Kiel (Germany).

Takako Hashimoto, David Shepard, Basabi Chakraborty and Tetsuji Kuboyama, *Detecting New Concepts in Social Media using Co-burst Pattern Mining*, The Eighth International Conference on Digital Society,

Mar. 27, 2014, Barcelona (Spain).

Shoichi Higuchi, Takako Hashimoto,
Tetsuji Kuboyama, Kouichi Hirata,
*Exploring Social Context from Buzz
Marketing Site - Community Mapping
Based on Tree Edit Distance* ; PerCol 2013
(Fourth International Workshop on
Pervasive Collaboration and Social
Networking), Mar. 18, 2013, San Diego
(USA).

〔その他〕

ホームページ:

<http://www.cuc.ac.jp/~takako/research.html>

動画コンテンツ:

http://www.gakushuin.ac.jp/univ/rioc/project/past_shirota.html

6. 研究組織

(1) 研究代表者

橋本 隆子 (HASHIMOTO Takako)
千葉商科大学・商経学部・准教授
研究者番号: 80551697

(2) 研究分担者

白田 由香利 (SHIROTA Yukari)
学習院大学・経済学部・教授
研究者番号: 30337901

(3) 研究分担者

久保山 哲二 (KUBOYAMA TETSUJI)
学習院大学・付置研究所・教授
研究者番号: 80302660

(3) 研究協力者

バサビ チャクラボルティ
(CHAKRABORTY Basabi)
岩手県立大学・ソフトウェア情報学部・
教授
研究者番号: 90305293