

# 災害情報基盤構築のための地理情報リソースに対する性能評価

## Development of Geographic Information Applications based on Information Infrastructure for Disaster Management

榊 剛史<sup>\*1,\*2</sup>  
Takeshi Sakaki

鳥海 不二夫<sup>\*2</sup>  
Fujio Toriumi

吉田 光男<sup>\*3</sup>  
Mitsuo Yoshida

<sup>\*1</sup>株式会社ホットリンク  
Hottolink, Inc.

<sup>\*2</sup>東京大学  
The University of Tokyo

<sup>\*3</sup>豊橋技術科学大学  
Toyohashi University of Technology

We are planning to develop information infrastructures for disaster management. As an example of those infrastructures, we develop a geographic information resource to assist digital volunteers who try to develop systems for disaster management. This system makes it possible to convert gps information and location name extracted from social media posts into administrative district information. In the final year of this session, we will evaluate the performance of the geographic information resource. Specifically, we verify how much improvement in performance can be obtained by applying the resource to the analysis we have performed in the past. Finally, we discuss the limits of the goals initially presented by this session and the efforts that will be needed in the future.

### 1. はじめに

本近未来チャレンジセッション「異種協調型災害情報支援システム実現に向けた基盤技術の構築」は、無事最終年度を迎えることができた。本セッションは、東日本大地震の直後に、図1のように災害時にも使える災害支援システム開発を支援するための情報基盤を構築することを目的に立ち上げられた。「災害支援システムの開発を支援する」ことを目的としたのは、以下のような理由からである。まず、「救援活動に必要な情報や機能などは、どこかに必ず存在している」ことを前提として、それらの情報をつなげるための技術が必要になると考えた。また、災害が発生した際にどのような情報が必要になるかを予測することは極めて困難であることに着目した。例えば、東日本大地震ではガソリンの不足が大きな問題となった。熊本地震では、損壊した建物に対するリスクアセスメントが重要となった。つまり、あらかじめ災害時に必要な情報を予測して平時からシステムを開発するのではなく、災害が発生し、必要な情報が明らかになった時点で迅速にシステムを開発できることが実用的であり、望ましい状況であると考えられる。そのために、「災害支援システム開発を支援する情報基盤の構築」を本セッションの目的とした次第である。

一方、著者は5年間の本セッションの取り組みにおいて、一貫してTwitterから位置に紐付いた情報を抽出することを試みてきた。その中で、ユーザが自由かつ安価に活用できる地理情報処理に関する基盤が十分に整備されていないことに着目し、地理情報リソースの整備およびそのリソースを用いて地理情報アプリケーションの開発を行った。

そこで最終年度にあたる本年は、本セッションを通じてどの程度の成果が得られたかの検証を行う。具体的には、2年目に行った東日本大地震データに関するTwitterデータからの地理情報抽出という問題設定に対して、本セッションを通じて整備した地理情報基盤を適用し、どの程度有用な情報が抽出できるかの評価を行う。最後に本セッションで掲げた目標の限界と今後必要となるべき取り組みについて議論を行う。

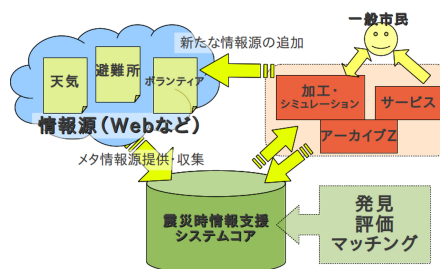


図1: 異種協調型災害情報支援システム実現に向けた情報基盤

表1: 異常値スコア上位5都道府県(3月11日15時)

| 順位 | 都道府県名 | $O_{rate}$ |
|----|-------|------------|
| 1  | 宮城    | 154.34     |
| 2  | 岩手    | 62.20      |
| 3  | 群馬    | 14.52      |
| 4  | 沖縄    | 14.28      |
| 5  | 長野    | 13.11      |

### 2. これまでの取り組み

本節では、著者の本セッションにおける取り組みについて述べる。

1年目においては、地域ごとのツイート数の推移から、被災程度が大きい地域を推定する課題に取り組んだ[榊剛 13]。具体的には、プロフィールから判定したユーザの居住地情報(都道府県単位, 市町村単位)を用いて、各投稿を各地域に紐付けて、地域ごと投稿数の時系列データを得た。そのデータに対し、異常検知の手法を適用し、極端に投稿数が減少した地域を被災地域として判定した。表2.のように一定程度の精度で被災地を推定することができたが、判定粒度として、都道府県単位や市町村単位では大きすぎるという問題点が浮き彫りとなり、結果として実用的な手法とはなり得なかった。

2年目は1年目の反省点から、ツイートデータに含まれる可能性がある3つの位置情報(GPS情報, ツイート内地名, ユーザの居住地情報)すべてを用いて、各投稿で言及されている位置情報を推定を行った[榊剛 14]。実際にはGPS情報はほとんど付与されていないため、主として用いたのはツイート内地名, ユーザの居住地情報である。ここでは特に、ツイート内地名を「粒度は細かいが、精度の低い位置情報」、ユーザの

表 2: 発信場所が推定できたツイートの例

|                                                                                                 |
|-------------------------------------------------------------------------------------------------|
| まみたすのお家も海の近くだから心配(;_;)石巻悪いで有名になりすぎだぁ                                                            |
| 再度。宮城県石巻市蛇田字西境谷地の川沿い、三菱自動車跡地後ろの二階建て住宅に取り残されています。二階まで浸水しそうです。子供が2人いるため屋根に登ることもできません。救助要請おねがいします。 |
| 大船渡のおともだち連絡つかないようしよづ y s                                                                        |
| 茨城大学、筑波大学、茨城県立医療大学ともに後期日程の学力検査は行わないことになりました。当初は延期が考えられていたようですが、この状況での実施は困難とのことです。               |



図 2: 位置情報を入力とした周辺施設検索システム

居住地情報を「粒度は粗いが、確度の高い位置情報」と捉え、その双方を組み合わせ、より詳細な位置情報を高い確度で推定することを試みた。具体的には、ツイート内地名を抽出した後、その地名の所在地とユーザ居住地が一致している場合に、ツイート内地名の確度が高いツイートとして抽出した。表 2 のように、確かに被災者からのものと思われる情報を抽出することができた。一方、ツイート内地名を抽出するための地名辞書の規模が小さかったため、十分な量の情報を抽出することが叶わなかった。

そこで、3 年目、4 年目は、地理情報リソースの整備およびその有用性を示すためのアプリケーション開発を行った [榎剛 15, 榎剛 16]。まずは地理情報リソースとして、Web 上に散在している地名と緯度経度情報が紐付けられた情報を集約・整理し、地理情報リソースとして本セッションのサイトで公開した\*1。また、地理情報処理 (GIS) サーバを構築するための手順とデータについても同時に公開した。そして、それらの地理情報リソースと GIS サーバを利用して、下記のような地理情報アプリケーションを開発した。

- 周辺施設検索システム  
現在の位置情報から、周囲の避難施設、教育施設、公園、医療機関を検索・可視化するシステム (図 2)。
- 過去ツイート検索システム  
位置情報と期間を指定して、収集したジオタグ付き過去ツイートを検索・可視化するシステム。
- リアルタイムツイート検索/集計システム  
位置を指定して、その周辺の位置情報が付与されたツイートをリアルタイムに検索・可視化するシステム。

これらのシステムを開発し、整備した地理情報リソースに実用性があることを検証した。特に「周辺施設検索システム」には、全国のコンビニ情報、観光地情報も追加し、学会参加者が周囲のコンビニ・観光地を探せるようなシステムとした。

### 3. 地理情報リソースの震災データへの適用

本節では、過去に行った分析方法と類似のアプローチについて、本地理情報リソースを適用し、どの程度性能が向上したかを検証する。手法自体は大きく変えず、手掛かりとして用いる地名辞書を大規模化するため、精度はそのままカバー率が向上することが期待される。

#### 3.1 分析手法

本稿では、本セッションの 2 年目に行った研究と類似のアプローチを適用する。すなわち、ツイート内地名を「粒度は細かいが、確度の低い位置情報」、ユーザの居住地情報を「粒度は粗いが、確度の高い位置情報」と捉え、その双方を組み合わせ、特定の地域に関連した投稿を選択的に抽出するアプローチである。

ここでは、特に特定の地域を入力として、その地域から投稿されたと思われる投稿を抽出することを目指す。手順は下記の通り。

##### 1. 特定地域ユーザリストの作成

あらかじめデータセット内のユーザの居住地を推定しておいた上で、特定の地域が居住地と推定されたユーザを集めて特定地域ユーザリストを作成する。ユーザの居住地推定は、Twitter ユーザプロフィールを手掛かりとする。具体的には Twitter ユーザプロフィールの「Location」に記載内容に地理情報リソースの地名が含まれていた場合に、その地名が含まれる地域 (区市町村単位) をユーザの居住地とする。

##### 2. 特定地域関連ツイート候補の抽出

1. で得られたユーザリストに含まれるユーザが投稿したツイートを、特定地域関連ツイート候補として抽出する。

##### 3. 特定地域関連ツイートの抽出

2. で得られたツイート群について、地名辞書を用いてツイート内に含まれる地名を抽出する。抽出された地名が入力された特定地域が含まれる場合に、当該ツイートを特定地域関連ツイートとして抽出する。地名辞書としては、整備した地理情報リソース (以下、地理情報リソース) と以前の研究で用いた都道府県名と全国の区市町村名のみが含まれる地名辞書 (以下、単純地名辞書) の 2 つを用いる。

上記の手順により、特定地域に強く関連したツイートが収集できると想定される。

#### 3.2 データセット

本研究では、過去の研究で用いた東日本大地震前後の日本語ツイートデータセットを用いる。収集手順は、以下の通り。

1. 予め日本語ツイートを投稿するユーザリストを作成

\*1 <https://sites.google.com/site/crepchidri/resources>

2. 1. のリストのユーザが震災前後に投稿したツイートを収集  
使用したデータセットの詳細は以下の通り。

- ユーザリストに含まれるユーザ数：130 万ユーザ
- ツイート数：356,118,522 ツイート
- 収集期間：2011 年 3 月 7 日～3 月 24 日

なお、このように位置情報を推定する研究は今までも数多く行われてきた [Hecht 11, Backstrom 10, Ji 09, Lin 04].

また、推定した位置情報を用いて、局所的なイベントを検出する手法が提案されてきた [Lee 11a, Walther 13, Lee 11b, Li 12].

本研究がそれらと異なる点は、位置情報の手掛かりについて、情報確度によって扱いを変えつつ、確度の違う情報を組みあわせて活用している点である。

### 3.3 評価実験

本研究で採用した分析手法を用いて、5つの地域を特定地域として選定し、それぞれから得られるツイート集合について適合率と再現率の評価を行う。ここでの正解とは、「人が投稿を読んで、確かに特定の地域から発信されていると判断できるもの」とする。5つの地域はそれぞれ、被害の大きかった地域として「石巻」「大船渡」、中程度の被害だった地域として「つくば」、原子力発電所事故が発生した場所として「双葉郡」、被害の少なかった地域として「本郷（文京区）」を選定した。

各地域毎に 200 件のツイートに対して地理情報推定を行った。結果は表 3 の通りである。5つの全ての地域において地名が推定できた割合が向上している。これより、大きくはないが、整備した地理情報リソースによる地名抽出のカバー率の向上を確認することができた。

また、抽出された詳細情報を見ると、2つの辞書で共に地名が抽出されている場合でも、地理情報リソースを用いた方がより詳細な位置情報を抽出できている。例えば、「中里小学校に避難した方もまだたくさんいるようですが、二階があるおうちの方は戻ってる人もいます。」というツイートからは、単純地名辞書では「中里」という地名が抽出されるだけだが、地理情報リソースを用いた場合は「中里小学校」という具体的な小学校が抽出されている。実際の災害支援においては、このような位置情報の粒度の違いが情報の重要度に大きな差を生むことが想定される。そのような意味で、地理情報リソースは単純なカバー率の向上以上の効果があると考えられる。

## 4. おわりに

昨年度までの成果により、整備した地理情報リソースが通常のシステム開発に活用可能であり、かつ災害支援に活用可能なシステムが構築できることを示していた。本稿では、本セッションの中で取り組まれてきた分析アプローチに対して、地理情報リソースを適用することで、既存のアプローチのカバー率を向上させられることを確認した。これにより、本セッションを通して整備したリソースにより、各種技術の性能向上を図ることが期待される。

本稿の成果に限らず、本セッションを通じて整備した地理情報リソースの貢献をまとめると下記ようになる。

- 多数の地名・施設名・ランドマーク名を含めた地理情報リソースを整備することで、ツイートからより詳細な位置情報をより多く抽出することができる。

- 既存の GIS サービス（各種地図 API サービス）と同等の機能を独立したサーバ単体で実現できることで、システムの大規模化・高速化が容易に図れる。

- 災害に限らず、地理情報を活用するための様々なシステムについて、必要な機能を容易に実装することができる。

本セッションが当初掲げた「異種協調型災害情報支援システム実現に向けた基盤技術の構築」について、全部を達成することはできなかった。しかし、地理情報に関する基盤データの整備という観点において一定の役割を果たせたと考えている。

他方、ただ基盤データを整備しただけでは多くの限界がある事も判明している。整備した基盤データの認知度を高めて、より多くの人に基盤データの価値を伝えと共、それを用いて実際に災害支援に役立つ情報の発信やシステムの開発を継続していく必要がある。例えば、IT × 災害や [ITx 災 16], 災害 IT 支援ネットワーク [災害 16], 情報支援レスキュー隊 IT Dart [情報 16], 減災 Info[gen 16] などの取り組みのように多数の人を巻きこんで大きな社会的な動きを作っていかなければ、いかに優れた基盤技術・基盤データを整備したとしても一過性の取り組みとして終わってしまうと考えられる。

本セッションは今年度で終わりであるが、本セッションに関わった人々の多くがこの取り組みを踏まえ、情報技術を災害支援に活かすための地道で継続的な社会的活動に取り組んでいくようになれば、それが本セッションの一番の成果となり得るのではないだろうか？

## 5. 謝辞

本研究は課題設定による先導的人文・社会科学研究推進事業、「リスク社会におけるメディアの発達と公共性の構造転換～ネットワーク・モデルの比較行動学に基づく理論・実証・シミュレーション分析」の助成を受けた。

## 参考文献

- [Backstrom 10] Backstrom, L., Sun, E., and Marlow, C.: Find Me If You Can: Improving Geographical Prediction with Social and Spatial Proximity, in *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pp. 61–70, ACM Press (2010)
- [gen 16] 減災インフォ, <http://www.gensaiinfo.com/> (2016)
- [Hecht 11] Hecht, B., Hong, L., Suh, B., and Chi, E. H.: Tweets from Justin Bieber's Heart: the Dynamics of the Location Field in User Profiles., in *Proceedings of the 2011 Annual Conference on Human factors in Computing Systems, CHI '11*, pp. 237–246, ACM Press (2011)
- [ITx 災 16] ITx 災害実行委員会：ITx 災害, <http://www.itxsaigai.org/> (2016)
- [Ji 09] Ji, R., Xie, X., Yao, H., and Ma, W.-Y.: Mining City Landmarks from Blogs by Graph Modeling, in *Proceedings of the Seventeen ACM International Conference on Multimedia, MM '09*, p. 105, ACM Press (2009)
- [Lee 11a] Lee, C.-H., Yang, H.-C., Chien, T.-F., and Wen, W.-S.: A Novel Approach for Event Detection by

表 3: ツイート内地名が推定できたツイートの割合

| 地名辞書     | 石巻    | 大船渡   | 双葉郡   | つくば   | 本郷    |
|----------|-------|-------|-------|-------|-------|
| 単純地名辞書   | 0.090 | 0.035 | 0.000 | 0.205 | 0.105 |
| 地理情報リソース | 0.105 | 0.080 | 0.005 | 0.245 | 0.145 |

Mining Spatio-Temporal Information on Microblogs, in *Proceedings of International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '11, pp. 254–259, IEEE (2011)

[Lee 11b] Lee, R., Wakamiya, S., and Sumiya, K.: Discovery of Unusual Regional Social Activities using Geo-tagged Microblogs, *World Wide Web*, Vol. 14, No. 4, pp. 321–349 (2011)

[Li 12] Li, R., Lei, K. H., Khadiwala, R., and Chang, K.-C.: TEDAS: A Twitter-based Event Detection and Analysis System, in *IEEE 28th International Conference on Data Engineering*, ICDE '12, pp. 1273–1276, IEEE (2012)

[Lin 04] Lin, J. and Halavais, A.: Mapping the Blogosphere in America, in *Workshop on the Weblogging Ecosystem at the 13th International World Wide Web Conference*, Vol. 18 (2004)

[Walther 13] Walther, M. and Kaiser, M.: Geo-spatial Event Eetection in the Twitter Stream, in *Proceedings of the 35th European conference on Advances in Information Retrieval*, ECIR'13, pp. 356–367, Springer-Verlag (2013)

[災害 16] 災害 IT 支援ネットワーク:<http://saigaiit.net/> (2016)

[榑剛 13] 榑剛史, 松尾豊, 鳥海不二夫, 篠田孝祐, 栗原聡, 風間一洋, 野田五十樹: ソーシャルメディアを用いた局所地域における異常検知手法の提案, 第 27 回人工知能学会全国大会, pp. 2B4–NFC–02a–1 (2013)

[榑剛 14] 榑剛史, 原久美子, 吉田光男, 鳥海不二夫, 篠田孝祐, 栗原聡, 風間一洋, 野田五十樹: 災害情報基盤構築に向けたテキストデータからの地理情報抽出システム, 第 28 回人工知能学会全国大会, pp. 1H2–NFC–02a–4 (2014)

[榑剛 15] 榑剛史, 大谷昭成, 鳥海不二夫, 吉田光男, 風間一洋, 篠田孝祐, 栗原聡, 野田五十樹: 災害情報基盤構築に向けた地理情報リソースの整備, 第 29 回人工知能学会全国大会, pp. 2B3–NFC–02a–5 (2015)

[榑剛 16] 榑剛史, 鳥海不二夫, 吉田光男: 災害情報基盤構築を活用した地理情報アプリケーションの開発, 第 30 回人工知能学会全国大会, pp. 1I2–NFC–01–1 (2016)

[情報 16] 情報支援レスキュー隊 (ITDART):<http://itdart.org/> (2016)