

# Q&A サイトを用いた地域に対する関心の推移の可視化

藤田 秀之・柴崎 真理子・木實 新一・有川 正俊

## Visualization of changing interest in geographic areas based on a Q&A site

Hideyuki FUJITA, Mariko SHIBASAKI, Shin'ichi KONOMI and Masatoshi ARIKAWA

**Abstract:** A large amount of geographic information is accumulated in Web space. But, it is not clear what user's interest and demand led to accumulation of such information. However, question/answer pairs on geographic issues submitted to a Question & Answer site directly reflect some interest and demand of users. This paper discusses a framework of revealing trends of user's interest and demand by displaying them on a map, based on question/answer contents. We visualize user's interest and demand in the area that was affected by the 2011 Tohoku earthquake and tsunami using question and answer from the Japanese Yahoo! Answers site, which is one of the most popular Q&A sites in Japan.

**Keywords:** ソーシャルメディア(social media), 可視化(visualization), Q&A サイト(Question and Answer site), ヒートマップ(heat map)

### 1. 序論

近年、ネットワーク上には地理空間に関する情報が大量に蓄積されている。しかしながら、これらの地理空間情報は、どのようなユーザの要求・関心に基づいて生成されたか明らかではなく、膨大な情報の中に容易に埋もれてしまい、その価値が次第に失われてしまう可能性がある。

地理空間関連情報の利活用が進み、有用な知識を共有・集約して人々の行動を支援するサービス(Ushahidi, 2011; Keay, 2010)や、地域に関する情報を効率良く検索できるサービス(以下、ローカルサーチとする)が利用されている。これらのサービスの利便性を高めるための研究も盛んになっており、たとえば、Jones 他 (2007)は、周辺にいるユーザが

使用した検索語を参照することで検索の利便性が上がると考え、周辺ユーザの検索語を表示する検索インターフェースを考案している。

しかし地域に関する要求や関心は、社会情勢の変化や大きな事象の発生に対して、広範囲に、また急激に変動する場合もある。そのような変化する要求や関心を扱うためには、一時的な空間分布(Jones, 2007)だけではなく、その推移にも着目すべきである。このような事象に対する地理空間情報の推移に着目した関連研究には Twitter を用いた Sakaki 他(2010)の取組みがある。

本研究では、Q&A サイトに投稿される“地域”に関する質問および回答は対象地域に対するユーザの要求・関心をより直接的に反映していると考え、代表的な Q&A サイトである Yahoo!知恵袋<sup>1</sup>を

柴崎真理子 〒277-8568 千葉県柏の葉 5-1-5  
東京大学 空間情報科学研究センター

Phone: 04-7136-4291

E-mail: ms@csis.u-tokyo.ac.jp

<sup>1</sup> <http://chiebukuro.yahoo.co.jp/>

対象に、東日本大震災被災地域に関する質問を収集し、その内容をユーザの要求・関心の推移を考慮して地図上に可視化する。

Yahoo!知恵袋とは、2008年時点での利用者数は1,261万人(ニールセン、2008)の利用者がおり、質問回答形式による知識・情報共有サービスである。過去の質問と回答はネット上で蓄積・共有される。回答の質を上げるために、一番適切な回答をした回答者にポイントを与えることで、ユーザに適切な回答をするインセンティブを与えている。Yahoo!知恵袋に投稿される質問に、質問者の現在位置情報は付与されていない。

本研究では、以下の可視化ツールを提案する。Yahoo!知恵袋から、「地域」を対象とした「ある特定のキーワードを含む質問」のトレンドを抽出し、地図上に可視化することで、地理空間に関連するユーザの要求・関心およびその推移を表す。具体的には、2つの可視化ツールを提案する。1つ目はユーザが指定した検索語を含む質問の分布を、県レベルの地域別に可視化するツールである。2つ目は市町村レベルの地域別に、ヒートマップを用いて可視化するツールである。これらのツールを用いれば、時空間的なパターンに基づき要求や関心の高い情報に焦点をあてることが容易になり、ユーザの要求や関心に合致する情報を提供することも容易になると見える。本稿ではこの可能性についても議論する。

## 2. 地域に関するデータ取得

Yahoo!知恵袋サイトでは、位置情報や詳細地名による検索APIは提供されていない。しかし、条件指定による検索フォーム<sup>2</sup>とキーワードによる質問検

索API<sup>3</sup>が提供されており、本収集システムではこれらを利用した。

### 2.1 県レベルのデータ取得

都道府県別の質問を取得するには、条件を指定する検索を用いた。この条件指定による検索は、キーワード、カテゴリ、質問・回答の投稿日時と解決日時、エリア(都道府県別)などの条件を指定できる。質問取得に用いた指定条件はキーワードと質問・回答の投稿日時・解決日時とエリアの3条件である。収集方法は3段階からなる。

(Step 1) 条件指定の検索フォームのリクエストURLのgetパラメータに、キーワード、期間、県のパラメータを設定し、html形式のデータを取得する。図1の赤文字はキーワード、青文字は期間、オレンジ色は県の条件指定を表しており、実際はURLエンコードを行う。

```
http://chiebukuro.search.yahoo.co.jp/search?p=放射能&vaop=a&search=all&id=dfrom_y=2011&dfrom_m=07&dfrom_d=01&dto_y=2011&dto_m=07&dto_d=31&type=&geo=13&nogeo=&region=10&prefecture=13&f_adv=1
```

図1. リクエストURLの構造

(Step 2) 取得したhtml形式のデータを解析し、検索キーワードヒット数や質問IDリストを取得する。

(Step 3) (Step 1)の指定条件の期間を指定し、開始日をずらしてリクエストを繰り返すことで、各期間の集計データを取得する。

以上の方法により、地域に関する質問文のキーワードヒット件数と質問IDリストを取得した。

### 2.2 市町村レベルのデータ取得

市町村ごとの質問文の取得には、キーワードに

<sup>2</sup><http://chiebukuro.search.yahoo.co.jp/advanced?fr=common-navi>

<sup>3</sup><http://developer.yahoo.co.jp/webapi/chiebukuro/chiebukuro/v1/questionsearch.html>

による質問検索 API を用いた。質問検索 API は、質問本文に指定したキーワード(以下、検索語とする)を含む質問の質問 ID(各質問文固有の ID), 質問本文, 回答(ベストアンサー), カテゴリ, 投稿・回答日時などを取得する。収集システムでは、質問 ID と質問本文, 投稿・回答日時, 投稿の種類(pc か mobile)を取得した。

これら上記の 2 つの質問文取得方法には、以下のようないが存在する。

条件指定による検索フォームは、都道府県の指定のみで市町村レベルの指定はできないが、2004 年からさかのぼって質問を取得することができ、かつ条件指定に期間を設定できる。また、取得できる質問数に制限はない。

質問検索 API はクエリに市町村名を入れることで各市町村に関連する質問を取得できると考えられる。しかし、質問取得の条件に期間設定ができない、API を動かした時点から最大 1,000 件のみの質問しか取得できない。そのため、1,000 件を超える過去の質問は取得できないという制約がある。

### 3. 実装

#### 3.1 収集システム

##### 3.1.1 県別の質問収集システムの実装

2.1 節にて解説した Yahoo! 知恵袋の条件指定による検索フォームによる収集方法を利用する。

本収集システムでは、代表的な被災地域を含む岩手県、宮城県、福島県と関東近郊の茨城県、栃木県、千葉県、群馬県、東京都と大阪府、福岡県の 10 都道府県ごとの「原発」「原子力」「セシウム」をキーワードとした 1 ヶ月間ごとのヒット件数を 2010 年 3 月から 2011 年 7 月まで取得した。

条件指定による検索フォームのリクエスト URL を用いて取得した html ページから検索ヒット件数が何行目にあるか解析し、該当行を切り出すこと

で検索ヒット件数を取得する。ただし、検索ヒット件数が 1 件のみの場合、html ページの構造が変化し、検索ヒット件数の位置がずれるため、従来の方法では検索ヒット件数が 1 件の場合と 0 件の場合の区別ができない。そのため、検索ヒット件数が 1 件かどうか確認する check プログラムを動かし、検索ヒット件数が 1 件となる期間を調べる。

##### 3.1.2 市町村別の質問収集システムの実装

2.2 節で解説した質問検索 API による市町村レベルの質問収集システムを構築する。今回は一例として、岩手県に関する質問文を収集し可視化を行う。質問本文に“地名語”を含む質問は、その“地域”に関する質問だと考えられる。したがって、質問検索 API の質問取得条件である検索語に“地名語”を設定し、質問文に“地名語”を含む質問を取得する。

以下のように実装した。

###### (1) 地名語リストの作成

質問検索 API の検索語に用いる地名語は、国立情報学研究所の北本朝展准教授と東京大学空間情報科学研究センターとの共同研究として開発を行っている「地名語辞書管理システム」で使用されている地名語辞書(国交省作成「街区レベル位置参照情報」(国土交通省, 2011))を用いた。この地名語辞書から対象とする県内の全市区町村名を抽出し、地名から「市」「区」「町」「村」を省いた地名語リストを作成する。全 76 語からなるリストが作成された。

###### (2) 質問検索 API による質問取得

質問検索 API を用いて、(1)で作成した地名語リストを検索語として質問を取得する。収集された岩手県に関する総質問数は 19,932 件となる。取得された質問集合には、以下のように目的と異なる質問も取得される。たとえば、地名「大迫」で取得された質問集合には、「大迫選手」という人名を

含む質問もある。また、同じ地名で異なる地域を指している質問も取得されてしまう。たとえば、岩手県にある山形町は「山形」という検索語で取得されるため、「山形(県)」の質問も含まれてしまう。さらに、取得結果の重複も起こる。「二戸市軽米町」を含む質問文は、「二戸」と「軽米」の検索語で重複して取得している。このような重複は質問IDによって抽出され、削除できる。重複を削除した結果、岩手県の市町村名を含む質問数は17,301件となった。

### (3) 目的と異なる質問の除外とジオコーディング

(2)で取得し重複を削除した質問本文に「地名語辞書管理システム」のジオパース(文章中から地名語抽出)とジオコーディング(地名語への緯度経度を付与する)機能を用いる。

このシステムのジオパース機能は、以下の特徴を備えている。文章の前後関係から地名語らしさを評価する。例えば、前述の「大迫」が人名なのか、地名なのかを推定する。また、部分的な地名語を補完し、住所文字列を生成する。上位の地名語の推定は、文章中に出現する他の地名語を反映して行われる。例えば、文章中に岩手県の地名が複数出現する場合、前述の「山形」を「岩手県山県町」と推定する。これらの特徴により、ジオパースすることで、

(2)で示した目的と異なる質問の多くが除外される。

Yahoo!知恵袋の質問検索APIと「地名語辞書管理システム」のジオパースAPIとジオコーディングAPIを用いて、市町村レベルの質問収集システムを作成した。

```
- {
  Id: "1370031028",
  Content: "東京駅周辺にいる方で今から盛岡方面に行く方いませんか??い",
  CategoryPath: "地域・旅行・お出かけ|交通・地図|鉄道・列車・駅",
  PostedDate: "2011-08-29T00:41:58+09:00",
  SolvedDate: "",
  PostedDevice: "mobile",
  Locations: [
    [
      name: "東京駅",
      lat: "35.681391",
      lng: "139.766103",
      count: 1
    ],
    [
      name: "盛岡駅",
      lat: "39.701547",
      lng: "141.136599",
      count: 1
    ]
  ]
}
```

図2. 市町村に関する質問収集システムの取得結果

この質問収集システムは、検索語に「市」「町」「村」「区」を省いた地名を使うことで、地名らしき語を含む質問収集の網羅性を高めることを優先し、不要な質問を「地名語辞書管理システム」のジオパース機能を用いて、選択的に削除している。

## 3.2 可視化ツールの実装

2つの可視化ツールを実装した。1つは、3.1.1節の県別の質問収集システムで集められたキーワードのヒット件数を時系列的な折れ線グラフで表現する。2つ目は、3.2.1節の市町村に関する質問収集システムで収集された地域に関する質問集合から、ユーザが入力したキーワードを含む質問の時系列的な件数推移を市町村ごとに地図上にヒートマップとして可視化する。

## 4. 実装結果

### 4.1 県レベルの可視化

3.1.1節の県別の質問収集システムで収集した検索キーワードヒット件数を時系列な折れ線グラフで表現する。図3は「原発」と「原子力」の各キーワードのヒット件数合計の推移を表した。ここから、原発や原子力に関連する質問が福島に集中していることがわかり、5月以降から関心が離れていることが分かる。図4は「セシウム」のヒット件数である。放射性物質セシウムが社会的問題となった時期

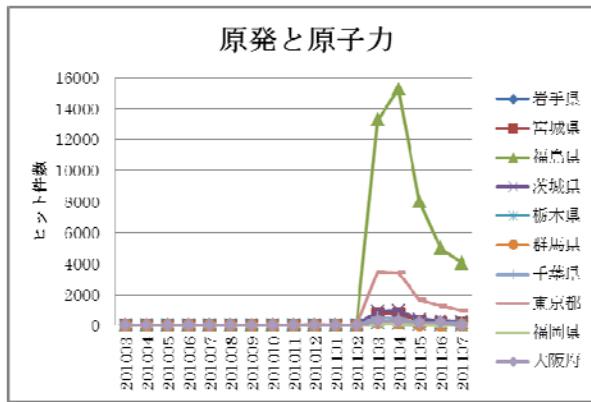


図3. 「原発」と「原子力」の県別ヒット件数

から福島県と東京都に関連する「セシウム」のヒット件数が増加していることが分かる。

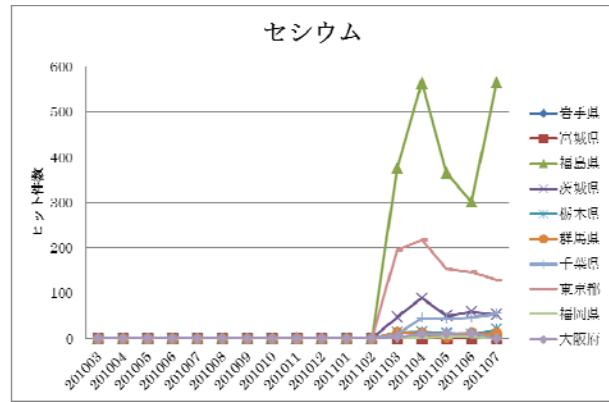


図4. 「セシウム」の県別ヒット件数

#### 4.2 市町村レベルの可視化

3.2節で実装した可視化ツールの市町村レベルの可視化のデモとして岩手県の質問集合から「安否」という検索キーワードを含む質問の時系列的な推

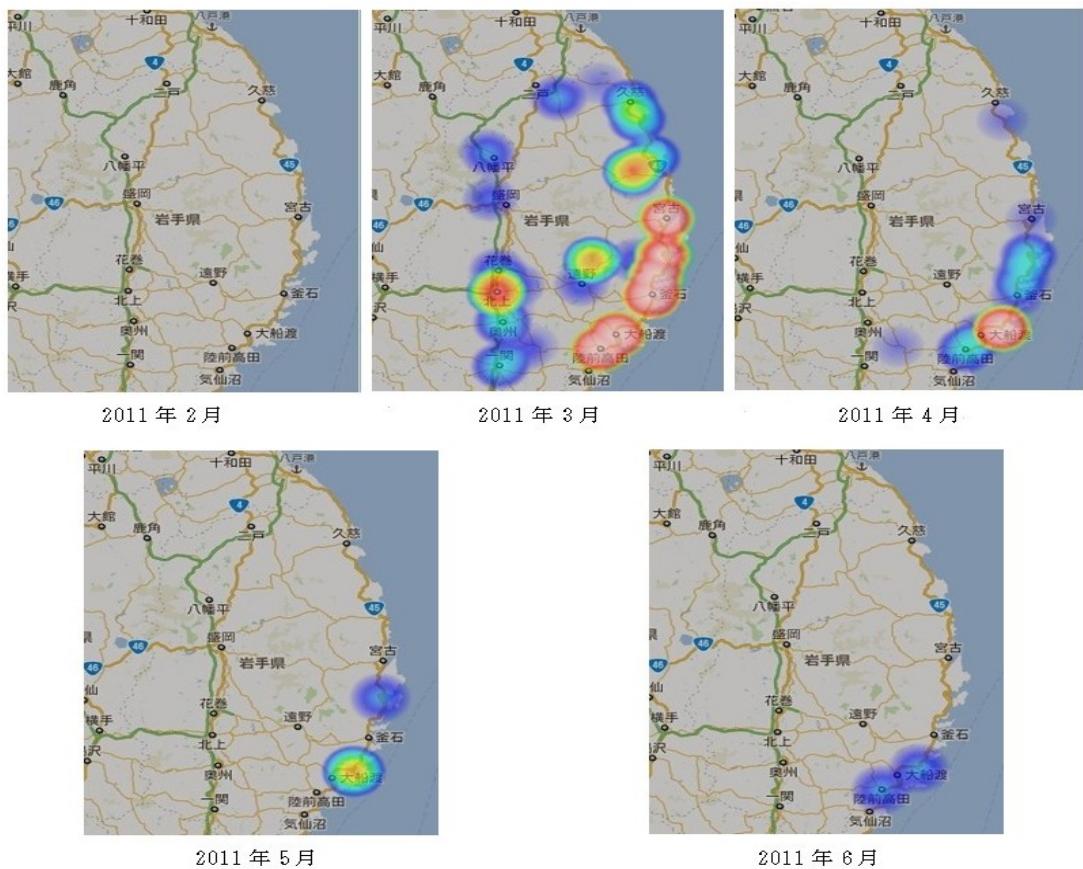


図5. 「安否」に対する岩手県市区町村レベルのヒートマップ

移を、以下のヒートマップに可視化した。

「安否」という語を含む質問数の推移は、2011年3月11日に発生した東日本大震災時に、ユーザの知人の安否情報に対する要求を表していると考えられる。

図5のヒートマップでは、3月には岩手県全域に「安否」を含む質問が分布しているが、時間が経つにつれ、津波による不明者の多い沿岸地域のみに分布が残っている。

#### 4.3 結果の概要と考察

本研究で扱った岩手県の例に見られるように、東日本大震災のような社会的関心の高い、大きな事象が発生したとき生成される地域に対する要求や関心は、事象の変化に伴い即座に変動する。これらの推移をヒートマップ、折れ線グラフを用いた時系列表現に基づき地図上に可視化し、他のユーザに緊急性の高い回答が求められている要求や関心を示すことで、優先的に回答を求めることができる。これにより、質の高い回答ができるだろうユーザを質問者と結びつけることができる可能性がある。

このように、ヒートマップおよび時系列の可視化を統合して、マクロなパターンに基づき要求や関心の高い情報に焦点をあてることができれば、ローカルサーチ等においてもユーザの要求や関心に合致する情報を提供することが容易になると見える。

また、ユーザの要求や関心の推移を考慮すれば、災害時などの大きな事象が発生した場合に、発信される膨大な情報の中から、コミュニケーションや情報伝達に使いやすいユーザフレンドリーな地理空間情報システムを構築することが可能と考える。

#### 5. 今後の課題

本システムで収集された質問が地名語辞書管理システムによってどれほど正確にジオコーディングされているのか、もしくは質問内容に基づいた可視化がされたか定量的な評価が必要である。また、デモで用いたキーワードはいずれも東日本大震災に密接に関連した語であるため、ユーザの要求や関心の推移を比較的明確に表すヒートマップを地図上に可視化できたと考える。「セシウム」「安否」などとは違い、より日常的に用いられる語を用いたユーザの要求や関心の可視化についても検討する。

#### 謝辞

本研究では、国立情報学研究所の北本朝展准教授と東京大学空間情報科学研究センターとの共同研究として開発を行っているジオコーダを利用させていただきました。

#### 参考文献

- Ushahidi, 2011. Ushahidi Home Page.  
<http://www.ushahidi.com/>. 参照日時: 2011/08/28.
- Keay,A., 2010. Ushahidi and Crowdmap:  
micro-Streaming as time-binding media. *3PM Journal of Digital Research & Publishing*, 116-125.
- Jones,M., Buchanan, G., Harper, R., Xech, P.-L., 2007. Question Not Answers: A Novel Mobile Search Technique. *Proc. CHI 2007*.
- Sakaki, T., Okazaki, M., Matsuo, Y., 2010. Earthquake Shake Twitter Users: Real-time Event Detection by Social Sensors. *Proc. WWW2010*.
- ニールセン, 2008. [http://www.netratings.co.jp/news\\_release/2008/04/Newsrelease2008422.html](http://www.netratings.co.jp/news_release/2008/04/Newsrelease2008422.html).  
参照日時 : 2011/08/28.
- 国土交通省, 2011. 位置参照情報ダウンロードサービス, <http://nlftp.mlit.go.jp/isj/>. 参照日時 2011/08/28.