原著論文

タグ付きコーパスを用いた地震記事からの 動向情報抽出・可視化システム†

高間 康史*1・山田 隆志*2

動向情報の要約と可視化に関するワークショップ(MuST)から配布されているタグ付きコーパス中の地震記事を対象として、地震に関する動向情報を抽出・可視化するインタラクティブ地震動向情報可視化システムを提案する。ガソリン価格や株価の変動など、多くの話題が時間的動向情報に主に関心があるのに対し、本稿で対象とする地震では震源の位置や震度の広がりなど、空間的動向情報も重要となる。提案システムでは、空間的動向情報は日本地図を用いた可視化表現、時間的動向情報は折れ線グラフや棒グラフなどにより表現する。また、時間的・空間的要素が組み合わさった地震特有の動向情報として群発地震も対象とし、日本地図と棒グラフの組み合わせにより表示する手法を提案する。さらに、提案システムでは日本地図上でのマウスクリックなど、直感的な操作により可視化表現が得られるインタラクティブ性も提供する。実際に構築したシステムに基づき、タグ付きコーパスからの動向情報抽出精度、既存の地震データベース検索システムとの機能比較の観点から提案システムの有効性について検証すると同時に、新聞記事からの地震動向情報抽出の利点についても考察する。

キーワード：動向情報，情報可視化，地震情報，インタラクティブ情報検索

1. はじめに

動向情報の要約と可視化に関するワークショップ(通称MuST[†])[2]から配布されている研究用データセットであるタグ付きコーパス(以降MuSTコーパス)中の地震記事を対象として、地震に関する動向情報を可視化するインタラクティブ地震動向情報可視化システムを提案する。

「動向」とは、事態の変化していく方向や、社会や組織などの現状や今後のなりゆきなどを表すが、動向を正しく把握することは、組織、個人における意思決定において重要であり、動向情報を収集・生成し、わかりやすく提示することのできる動向情報可視化システムの開発は今後重要であると考える。

動向情報には様々なものが考えられ、MuSTコーパス中にもガソリン(全国平均店頭価格など)、パソコン(国内出荷台数など)、交通事故(全国交通事故死者数)、台風、地震、大リーグ(マグワイヤのホームランなど)、政治動向(内閣支持率など)、日経平均株価な

ど、計20話題についてのタグ付き新聞記事が含まれている。この中で、本稿では地震記事に含まれる動向情報を対象とする。日本は世界でも有数の地震多発国であり、一般人であっても地震への関心は高い。従って、地震動向情報を可視化することは、地震専門家にとって有用なだけでなく、防災意識の向上や防災対策への動機付けにもつながることが期待できる。

地震動向情報可視化システムにおいては、以下の二点を特に考慮する必要があると考える。一点目は、ガソリン価格の変動やホームラン本数の変化など、動向情報の多くが統計量の時系列な変化(時間的動向情報)であるのに対し、台風や地震は、台風の進路や震源と各地における震度の関係などの空間的動向情報も含んでいる点である。地震動向情報を扱うためには、時間的及び空間動向情報の両方を可視化する必要がある。二点目は、可視化表現は利用者が動向情報を理解するのに役立つだけでなく、新たな関心を生み出す役割も期待できる点である。そのためには、表示された可視化表現中の興味ある箇所をクリックすることにより、関連する興味へ直感的にアクセスできるインタラクティブ性を備えていることが望ましい。

以上より、提案システムでは、空間的動向情報は日本地図を用いた可視化表現、時間的動向情報は折れ線

† Extraction and Visualization System for Earthquake Trend Information from Tagged Corpus
Yasufumi TAKAMA and Takashi YAMADA

*1 首都大学東京システムデザイン学部
Faculty of System Design, Tokyo Metropolitan University

*2 首都大学東京大学院システムデザイン研究科
Graduate School of System Design, Tokyo Metropolitan University

1 http://www.kecl.ntt.co.jp/sci/workshop/must/must_index.html

グラフや棒グラフを用いてユーザに提示する。さらに、日本地図上でのマウスクリックなど、直感的な操作によりユーザがインタラクティブに可視化表現を得ることができる機能も提案し、実装する。

2節では、新聞記事における動向情報、地震記事から抽出可能な動向情報およびその可視化表現について考察し、3節で提案システムの詳細を述べる。4節では、既存の地震データベース検索システムとの比較などを通じて、提案システムの有効性、および新聞記事から動向情報を抽出することの利点について論じる。

2. 地震記事における動向情報

2.1 新聞記事における動向情報

本稿では、MuST[2]から配布されているタグ付きの新聞記事を動向情報の抽出対象とする。前節で述べた様な動向情報は、新聞記事を対象とする必要は必ずしもなく、各種機関が発表・公開しているデータを基に生成・可視化することも可能である。これらの一次情報源を基に生成されている二次情報源である新聞記事は、以下の特徴を持っている。

- ・記者(新聞社)によりオーソライズされた情報である。
- ・動向に関する解釈が与えられている。

新聞は、世の中で発生する膨大な情報の中から、限られた紙面へ掲載するために、読者にとって重要な情報を取捨選択している。個人的なBlogなどとは異なり、(特に大手)新聞社は思想的な偏りも少なく、広く一般に重要とみなされる情報を選択して新聞に掲載している。この点で、新聞はオーソライズされた一般人向けの情報フィルタであり、新聞記事はフィルタされた情報といえる。

また、単に統計量が記述されているだけでなく、「価格が減少傾向にある」、「前年と変わらない」、「最安値」など、その数値に対する意味、解釈も付与されており、それらが動向情報を表していることも多い。

以上より、新聞記事と動向情報の関係は図1にまとめる通り、世の中で発生した膨大な情報から、新聞というオーソライズされたフィルタを通すことによって

図1 動向情報のフィルタとしての新聞

重要な情報のみを取り出し、動向情報に関する解釈を付与したものが新聞記事とみることができる。新聞記事がこのような特徴を持つのは、専門家ではなく一般の読者が読むことを前提にしているためである。本稿では、これらの特徴を考慮して、新聞記事から動向情報を抽出、可視化を行う。

2.2 地震記事における動向情報

本稿で対象とする地震記事の一例を図2に示す。地震記事は一般に、地震の震源に関する情報(震源、マグニチュード)と、その地震により震度を観測した地域(震度観測地)に関する情報を含んでいる。その他、津波を伴うおそれのある地震の場合には津波に関する情報が含まれる場合もあるが、本稿で利用するMuSTコーパス中には含まれないため、ここでは考慮しない。また、図2の場合は交通機関への影響など、被害に関する記事などもあるが、動向情報としては定型的ではないため、本稿では対象外とする。

図2の様な定型的な地震記事から抽出可能で、かつ読者にとって有益と思われる動向情報について、本稿では以下のように分類する。

- (A) 1記事に含まれる動向情報
  - ・震源と震度観測地の関係(空間的動向情報)
- (B) 複数の記事を集約して得られる動向情報
  - (B-1) 時間的動向情報
  - (B-2) 空間的動向情報
  - (B-3) 時間的・空間的動向情報

大分などで震度4【大阪】

23日午前4時49分ごろ、中国地方を中心に地震があり、山口県光市や大分市などで震度4を記録した。震源地は瀬戸内海の伊予灘西部で、震源の深さは約90キロ、地震の規模を示すマグニチュード(M)は5.7と推定される。この地震の影響で山陽新幹線や四国内のJR線で遅れが出た。

主な各地の震度は次の通り。

震度4＝山口県光市、新南陽市、大分市▽震度3＝山口市、大分県別府市、島根県出雲市、広島市中区、香川県多度津町、松山市、高知県安芸市▽震度2＝岡山県倉敷市、広島県福山市、高松市、高知市、鳥取市、徳島市▽震度1＝兵庫県三木市、和歌山市、大阪府能勢町、同府岬町、奈良市

新幹線などに遅れ

山陽新幹線は広島―徳山間で、午前6時から約2時間、時速70～170キロの速度規制を行い、広島発博多行き「こだま561号」が32分遅れたのを最高に計10本が32～1分遅れ、3000人に影響が出た。JR四国によると、予讃線の伊予大洲―宇和島間と、予土線の宇和島―江川崎間で一時運転を見合わせ、特急など計3本が運休、6本が60～10分遅れ、約550人に影響した。

また、山口県内の山陽自動車道防府西―熊毛間が約2時間、通行止めになった。

図2 地震に関する新聞記事の例
(MuSTコーパスよりタグを除去して抜粋)

ここで、時間的動向情報とは、統計量の時間的変動や、出来事などの時系列的関係であり、空間的動向情報とは、統計量の地域における違いや、出来事などの空間的関係であると定義する。(A)は、通常は1記事に含まれる空間的動向情報であり、地震の規模や影響を知ることができる。また、(B)の動向情報を求める基にもなる。(B)は、集約の仕方によって空間的動向情報、あるいは時間的動向情報のいずれにもなりうる。例えば、「ある期間にどのくらい地震が頻発したか」を知りたい場合は、地震の発生頻度や発生間隔が関心の対象となるので、時間的動向情報になる。「ある地域で過去5年間に発生した地震を知りたい」場合なども、空間(地域)は固定し、時系列に地震情報を集約した時間的動向情報になる。

一方、複数の震源地の位置関係などを知りたい場合(「ある期間に、どこで地震が発生したか知りたい」など)には空間的動向情報の方がふさわしいといえる。

上記が空間軸、時間軸の一方を固定あるいは集約し、他方の軸上での動向を表示したものであるのに対し、時間、空間の両軸上で動向を同時に見たい場合も考えられる。例えば、「最近地震の多い地域を知りたい」場合などは、地震発生頻度の時間的変化(時間的動向情報)を地域ごとに集計し、比較する(空間的動向情報)必要がある。さらに、群発地震も、時間的・空間的に近い地震の集合と見なせるため、時間・空間的の動向情報の組み合わせといえる。

2.3 地震動向情報と可視化表現

2.2節で分類した各種の地震記事に関する動向情報について、その提示に用いるべき可視化表現、およびその可視化表現を呼び出すためのユーザ・システム間のインタラクションについて提案する。

空間的動向情報の表現においては、地図上での情報提示が有効であると考え、震源地や震度観測地の関係(2.2節(A))などは、震源や震度観測地の位置を地図上で表示することにより、地震による影響の範囲などを把握することが容易になる。Hi-net高感度地震観測網[5]や気象庁の震度データベース検索システム[3]、tenki.jp(日本気象協会)[6]などでも、日本地図上に震源地を表示し、各地の震度を数値あるいは色によって地図上に表示しており、一般的な可視化表現である。前節の例で示した複数の震源地の関係が知りたい場合(B-2)なども、複数の地震情報を同一の地図上に写像することで表現可能である。

一方、時間的動向情報は、時間を横軸とする折れ線グラフ、棒グラフによる表現が適していると考え、前節の例では、地震の頻度を知りたい場合は頻度を縦

軸とし、地震の時系列な発生の様子を知りたい場合にはマグニチュードを縦軸にとればよい。例えば、東京大学地震研究所において公開されている最新地震活動マップ[8]においても、横軸を時間(月日あるいは時刻)、縦軸をマグニチュードとする棒グラフを利用している。縦軸に扱いたい数量を用い、横軸として動向情報の集約に適切なスケールを採用することにより、各種の時間的動向情報を可視化可能である。

時間的動向情報の表現としては、他にアニメーション[4]を用いることや、各時点について作成した可視化表現(スナップショット)を時系列に並べるなどの表現[1, 7]が考えられるが、可視化可能な情報が特定のものに限定されてしまうため、本稿では対象としない。

新聞記事から抽出・可視化された動向情報は、読者が記事の内容を理解するのを助ける役割を果たすと同時に、新たな関心を生み出す役割も期待できる。そのためには、表示された可視化表現中の興味ある箇所をクリックすることにより、関連する情報へ直感的にアクセスできるインタラクティブ性を備えていることが望ましい。そこで本稿では、情報可視化表現の一つである地図から、インタラクションにより下記情報へのアクセスを可能にする。

- (i) 震源から関連する新聞記事へのアクセス
  - (ii) ある地域における時間的動向情報へのアクセス
  - (iii) 群発地震に関する時間的動向情報へのアクセス
- (i)は、可視化表現中のある地震に着目し、それについての新聞記事を確認したい場合である。(ii)は、フォーム形式のインタフェースで地域名などを指定するよりも、地図上で該当地域をクリックする方がより直感的であるため採用する。また、(iii)は2.2節の(B-3)に示したとおり、時間的動向情報と空間的動向情報の両者を組み合わせて表現する必要があるため、空間的動向情報から対応する時間的動向情報へアクセス可能とする。

3. 動向情報抽出・可視化システム

3.1 システム構成

本稿では、地震記事から震源、震度観測地に関する情報を抽出し、地震に関する時間的、空間的動向情報を可視化してユーザに提示する、インタラクティブ地震動向情報可視化システムを提案する。図3に、提案システムの構成を示す。

図に示す通り、提案システムは新聞記事データベースから動向情報抽出部により地震情報データベースを構築する事前処理(オフライン)と、ユーザからの要求に応じて可視化表現を生成し、インタラクティブに提

図3 インタラクティブ地震動向情報可視化システムの構成

示するオンライン処理に大別される。

事前処理において、地震記事データベースとしてMuSTコーパスを用い、不足する情報はWebで公開されている気象庁の震度データベース検索システム[3]から収集・補完する。また、空間的動向情報可視化の際に必要となる、震源や震度観測地の緯度経度情報などは、地震記事に含まれないためWebから収集し、市区町村データベースを更新する。

オンライン処理では、動向情報生成部により地震情報データベースと市区町村から可視化表現を生成し、Webブラウザを利用した動向情報表示インタフェースによりユーザへ提示する。以下、動向情報抽出部については3.2節、動向情報生成及びインタフェースについては3.3節で詳しく述べる。

3.2 地震記事からの動向情報抽出

本稿で用いるMuSTコーパスに含まれる、タグ付き記事を図4に示す。図4に示す記事は、図2と同じ内容のものである。なお、MuSTコーパスには配布時期によりいくつかのバージョンが存在するが、本稿では2005年6月に配布されたバージョンを用いている。

図において、文書番号(<DOCNO>)やカテゴリ(<SECTION>)、見出し(<HEADLINE>)など書誌情報に関するメタデータにタグ付けされているだけでなく、日付に関する情報(<date>)や震源地、マグニチュードについても<par>タグによりタグ付けされている。このタグを利用することにより、テキストからの動向情報抽出を容易に行うことができる。

地震情報データベースには、各地震について、震源地に関する情報(震源データ)と、震度観測地に関する

```

<?xml version="1.0" encoding="Shift_JIS"?>
<DOC>
<DOCNO>980523360</DOCNO>
<SECTION>総合</SECTION>
<AE>無</AE>
<WORDS>544</WORDS>
<HEADLINE>大分などで震度4【大阪】</HEADLINE>
<TEXT>
<unit event="地震"><date gra="日" abs="19980523">23日
</date><par>午前4時49分ごろ</par>、<par>中国地方</par>を中心に
<name>地震</name>があり、<par>山口県光市</par>や<par>大分市
</par>などで<par>震度4</par>を記録した</unit>。震源地は瀬戸内海
の伊予灘西部で、震源の深さは約90キロ、<unit event="地震"><name>
地震</name>の規模を示すマグニチュード(M)は<par>5.7</par>と推
定される</unit>。この地震の影響で山陽新幹線や四国内のJR線で遅れ
が出た。
主な各地の震度は次の通り。
震度4＝山口県光市、新南陽市、大分市▽震度3＝山口市、大分県別
府市、島根県出雲市、広島市中区、香川県多度津町、松山市、高知県安
芸市▽震度2＝岡山県倉敷市、広島県福山市、高松市、高知市、鳥取市、
徳島市▽震度1＝兵庫県三木市、和歌山市、大阪府能勢町、同府岬町、
奈良市
新幹線などに遅れ
山陽新幹線は広島―徳山間で、午前6時から約2時間、時速70～170
キロの速度規制を行い、広島発博多行き「こだま561号」が32分遅れた
のを最高に計10本が32～1分遅れ、3000人に影響が出た。JR四国に
よると、予讃線の伊予大洲―宇和島間と、予土線の宇和島―江川崎間で
一時運転を見合わせ、特急など計3本が運休、6本が60～10分遅れ、約
550人に影響した。
また、山口県内の山陽自動車道防府西―熊毛間が約2時間、通行止め
になった。
</TEXT>
</DOC>
  
```

図4 MuST コーパスのタグ付き地震記事の例

情報(震度データ)、関連記事に関する情報(記事データ)を格納する。震源データ、震度データ、記事データは、それぞれ以下のフィールドを持つ。

1. 震源データ
  - ・発生日時：数値
  - ・マグニチュード：数値
  - ・震源緯度：数値
  - ・震源経度：数値
  - ・群発地震の有無：0 or 1
2. 震度データ
  - ・発生日時：数値
  - ・都道府県名：文字列
  - ・市区町村名：文字列
  - ・震度：数値
3. 記事データ
  - ・発生日時：数値
  - ・記事ID：数値

本稿では、地震の発生日時を地震に関するIDとして用いる。これらのうち、発生日時、マグニチュード、記事IDの各フィールドは、コーパス中のタグを利用して、XMLパーサにより取得する。記事IDフィールドには、コーパス中の<DOCNO>タグで囲まれた値を指定する。震源データの発生日時フィールド

の値は、<date>タグの“abs”属性値、およびその直後の<par>タグで囲まれた文字列のうち、「〇時〇分」などの時間表現にマッチするものから取得する。マグニチュードについては、記事中の「マグニチュード」という単語の出現位置を手がかりに、その直後にある<par>タグで囲まれた数値を抽出する。

震源の緯度・経度に関する情報は、発生日時フィールドの値を基に、気象庁の震度データベース検索システム[3]から取得する。

図4に示すMuSTコーパスにおいて、震度観測地および震度に関する情報にはタグ付けがされていないため、簡単な自然言語解析処理を用いる。MuSTコーパスを分析した結果、地震記事での地域・震度情報の表記には以下の2パターンが存在することを確認した。

1. 震度→地域の順に記述

(例)「各地の主な震度は次の通り。 ▽震度4 新潟県小千谷市、六日町、．．．」(980221352)

2. 地域→震度の順に記述

(例)「1回目の地震の揺れは首都圏にも広がり、横浜市中区で震度3、千葉県館山市、東京都国分寺市、山梨県大月市などで震度2を観測した。」(980427160)

記事全体を形態素解析することによって、この2パターンの判定を行い、震度および震度観測地の情報を取得する。上記の例に示すとおり、震度観測地の情報は主に市区町村単位で記述されているため、同名の市区町村が複数存在する場合など、都道府県名を一意に特定できない。そのような場合、都道府県番号を利用して震源地に最も近い市区町村を特定する。JISコードで定められている都道府県番号は、北海道を01、沖縄県を47として全都道府県に固有の番号を割り当てたものであり、一般に近い位置にある都道府県は近い番号となるように定められている。この性質を利用し、同名の市区町村を含む都道府県と、震源地の都道府県の番号を比較して、その差が最小となる都道府県に震度観測地が存在すると判定する。

後述する動向情報の可視化表現生成時に、震度観測時の緯度・経度情報を取得する必要があるが、これは市区町村データベースに格納する。市区町村名の記事から抽出後、その緯度・経度が市区町村データベースにまだ登録されていない場合、全国市役所の緯度経度情報を公開しているWebサイト[9]より取得し、データベースを更新する。

2.3節で述べたとおり、本稿では時間的・空間的動向情報の組み合わせとして、群発地震についても動向情報の一種として可視化の対象とする。群発地震とは一般に、ある地域で一定期間継続して起こる一連の地震活動のことを指すが、継続期間の長さや頻度に関し

て明確な定義があるわけではない。そこで本稿では、2.1節で述べたとおり、新聞記事が解釈された情報であることに着目し、記事中に「群発地震」という表記がある場合、その地震を「群発地震候補」とし、前後30日、震源の緯度・経度が0.1度以内で異なる地震が発生している場合、それらを群発地震と判定する。震源データの「群発地震の有無」フィールドの値は、群発地震の時1、そうでない時0とする。

3.3 動向情報可視化インタフェース

オンライン処理においては、前節の処理により構築した地震情報データベース、市区町村データベースからユーザの要求に合致するデータを取り出し、動向情報生成部で適切な可視化表現を生成し、動向情報表示インタフェースに送ってユーザに提示する。

2.2節で考察した、地震に関する動向情報のうち、MuSTコーパス中の地震に関する記事から抽出可能であるものを考慮して、以下を可視化の対象とする。

・日本地図を用いた空間的動向情報表示

指定期間に発生した地震について、震源・震度観測地を日本地図上に写像して表示する。個々の地震(2.2節の(A))だけでなく、複数の地震を同時に表示することも可能にすることによって、時間軸方向の集約による空間的動向情報の提示(B-2)も可能とする。

・折れ線による時間的動向情報表示

日本全体、あるいは指定地域で発生した地震について、一定期間毎の発生回数を集計し、折れ線グラフで提示する。MuSTコーパスに含まれるのが1998年の地震記事のみであるため、今回は月単位の集約に固定しているが、今後コーパスが拡張された際に、複数の時間粒度に対応することは容易である。

・群発地震に関する動向情報表示

群発地震は、複数の地震の空間的動向(発生位置の関係、震度観測地の広がりなど)、時間的動向情報(発生間隔)の両要素を含んでいるため、空間的動向については上記の日本地図を用いた表示、時間的動向については横軸を時間、縦軸をマグニチュードとする棒グラフで提示する。

プロトタイプシステムはJSP(JavaServer Pages)として実装し、Webブラウザからアクセス可能とした。図5にインタフェースの画面例を示す。

画面左側に日本地図を用いた空間的動向情報提示部、右上部にフォーム形式のクエリ入力部、右下部に関連記事表示部が配置されている。

クエリ入力部を図6に示す。ユーザはグラフ種、可視化対象とする地震の規模(マグニチュード)の最小値、可視化対象期間を指定する。グラフ種として

図5 インタフェースの画面例

◎マップ ○グラフ	
マグニチュード 指定しない	
検索開始時刻	1998年 01月 01日 00時 00分
検索終了時刻	1998年 12月 31日 23時 59分
検索	

図6 クエリ入力フォーム

「マップ」を選択した場合、指定した条件(マグニチュード、発生時期)を満たす地震の震源・震度観測地を日本地図上にマップして表示する(図7)。日本地図はSVG(Scalable Vector Graphics)形式で生成されている。右下の凡例に示すとおり、震度は色分けされて地図上に表示される。左上には、該当地震の発生日時が上から時系列に表示されている。この表示と日本地図上の震源・震度観測地の表示は連動しており、左上の発生日時にカーソルを重ねると、該当地震の震源・震度が日本地図上で強調表示される。また、地図上で震源にカーソルを重ねると、関連する震度および発生日時が強調表示される。さらに、発生日時あるいは震源をクリックすることにより、関連記事が表示される。

グラフ種として「グラフ」を選択した場合、指定されたマグニチュード以上の地震の発生頻度を集計し、図8に示す折れ線グラフとして別ウィンドウに表示する。集計粒度はグラフに表示する期間によって変更することができ、図では1年間の地震発生頻度を2ヶ月毎に集計した結果を表示している。1週間分を表示する場合は、日毎に発生頻度を集計するなどといったことも可能である。クエリ入力部から「検索」を実行した場合は、日本全国で発生した地震を集計して表示するが、空間的動向情報提示部の都道府県をクリックすることにより、その地域での時間的動向情報を表示することも可能である。

指定した期間に群発地震が含まれる場合には、空間的動向情報提示部の左上において、該当地震の直前に赤字で「群発地震」と表示される。この文字にカーソルを重ねると、該当する一連の地震の発生日時が強調表示されるだけでなく、それらの震源・震度情報をまとめて日本地図上に強調表示する。同時に、発生間隔に関する時間的動向情報を、横軸に発生日時、縦軸にマグニチュードをとる棒グラフとして表示する(図9)。

図7 日本地図を用いた空間的動向情報表示

図8 折れ線グラフによる時間的動向情報の表示

図9 群発地震に関する時間的動向情報表示

4. 地震動向情報可視化システムの有効性の検証

本節では、本稿で提案・開発したインタラクティブな地震動向情報可視化システムの有効性について、MuSTコーパス中の地震記事からの動向情報抽出精度(4.1節)、既存の地震データベース検索システムとの機能比較(4.2節)、新聞記事からの地震動向情報抽出の利点(4.3節)の3点から考察・検証を行う。

4.1 MuSTコーパスからの動向情報抽出精度

MuSTコーパス中には地震記事が23記事含まれており、このうち3.2節で提案した手法により正しく震源および震度観測地の情報が抽出できた記事は19記事であった。残る4記事については、複数の地震についての記述が含まれているため、抽出できない地震情報が存在した。抽出できなかった地震は以下に大別できる。

- (1) マグニチュード、震度に関する情報が両方とも記述されていない地震。

該当記事2件(記事ID: 980504147, 980813002)

- (2) マグニチュード、発生時間に関してタグが付与されていない地震。

該当記事2件(記事ID: 980815139, 980822294)

- (3) 過去の地震に関する記事である。

該当記事1件(記事ID: 980813002)

(1)については、地震発生日時は抽出可能なため、気象庁の震度データベース検索システム[3]などから情報を補完することは可能である。ただし、本稿では地震記事からの抽出が目的であるため、対象外とした。

(2)に該当する記事(980815139)を図10に示す。午後7時36分に発生した地震(マグニチュード4.4)についてはタグが付与されているため、提案手法により正

```
<?xml version="1.0" encoding="Shift_JIS"?>
<DOC>
<DOCNO>980815139</DOCNO>
<SECTION>社会</SECTION>
<AE>無</AE>
<WORDS>227</WORDS>
<HEADLINE>上高地で震度4の地震</HEADLINE>
<TEXT>
  <unit event="地震"><date gra="日" abs="19980814">14
  日</date><par>午後7時36分ごろ</par>、岐阜県飛騨地方を
  震源とした<name>地震</name>があり、<par>長野県安曇村
  上高地</par>で<par>震度4</par>を観測した</unit>。気象庁
  の観測によると、震源の深さは約10キロ、<unit event="地震
  "><name>地震</name>の規模を示すマグニチュードは<par>
  4.4</par>と推定される</unit>。また、この地震と前後して、同
  日午後2時6分と同11時48分、同52分ごろ、上高地で震度3
  を記録する地震が3回起きた。マグニチュードはそれぞれ4.1、
  3.5、3.6と推定される。

  <unit event="地震">一連の群発地震で、<par>震度4以上
  </par>を記録したのは<date gra="日" abs="19980814">この
  日</date><par>夜</par>の<name>地震</name>で計3回に
  なった</unit>。
</TEXT>
</DOC>
```

図10 地震記事(980815139)

しく抽出可能であるが、午後2時6分(同4.1)、11時48分(同3.5)、11時52分(同3.6)についてはタグが付与されていないため抽出されなかった。これについては正規表現などを用いれば抽出可能であるが、タグ付きコーパスを利用するメリットがないため、本稿では対象としなかった。

(3)は、新聞記事の時点よりも過去に起こった地震についての記載であり、本稿では抽出対象としていない。

上記の様な、記述が定型的でない地震を除けば、提案手法はMuSTコーパスのタグに基づき地震情報を正しく抽出可能である。

4.2 既存地震データベース検索システムとの機能比較

提案システムが地震動向情報の可視化に必要な機能を備えていることを検証するために、気象庁で公開している震度データベース検索システム(既存システム)との機能比較を行う。

既存システムでは、1926年～2日前までの有感地震について、震度別回数表、有感地震検索が可能である。震度別回数表は、全国、都道府県、あるいは観測地において、指定した期間に発生した地震を震度毎に回数を集計し、表形式で提示する。集計の時間粒度は日別、月別、年別を指定可能である。この機能は、提案システムの時間的動向情報提示機能に対応するが、既存システムと比較した場合の違いは表1に示すとおりである。表において、「空間粒度」とは検索時の集計

表1 提案システムと既存システムの機能比較

機能	提案システム	既存システム
空間粒度	全国, 都道府県	全国, 都道府県, 観測地
震度指定	指定震度以上を集計	震度毎に集計
表示形式	グラフ	表形式

単位として指定可能な粒度を意味している。

空間粒度の指定では既存システムの方が詳細に設定可能だが, 今後MuSTコーパスが拡張される際に, 提案システムを拡張することは可能である。一方, 折れ線グラフによる提示は表形式による提示よりも直感的にわかりやすく, 提案システムの優位な点と言える。

既存システムの有感地震検索では, 都道府県(全国も指定可能), 震度, 期間を指定して, 該当する地震を期間先頭から指定個数まで検索する。結果は地震毎に, 震源情報(日時, 緯度・経度, 深さ, マグニチュード, 地域名), 震度観測地(震度毎に集計)を表形式で提示する。また, 図7と同様の可視化表現も提供している。この機能は, 提案システムの空間的動向情報提示機能に対応するが, 既存システムでは個々の地震についての可視化表現しか提供されないのに対し, 提案システムは複数の地震を同時に日本地図上に表示可能という大きな違いがある。これは, 空間的動向の把握を容易にする点で, 提案システムの大きな優位点と言える。

この他にも, 群発地震に関する動向情報の提示, 日本地図から新たな可視化表現をインタラクティブに生成可能であるなど, 既存システムにない機能も備えており, 地震動向情報可視化システムとして提案システムは十分な能力を有していると考ええる。

4.3 新聞記事からの動向情報抽出のメリット

表2は, MuSTコーパスに含まれる地震記事から, 規模(マグニチュードではなく, 震度で判断した)の大きな4地震を選び, 新聞記事に記載されている震度観測地と, 気象庁の地震データベースから取得した震度情報について, 震度毎に観測地数を比較したものである。表中で「>50」としている箇所は, 該当の観測地が50箇所より多く存在することを意味している。当然ながら, 地震データベースの方が多数の震度観測地の情報を含んでいるが, 空間的動向情報の可視化という観点からは, 地図上に表示される地点が多すぎるとかえって視認性が低下するおそれがある。従って, 地震の規模, 空間的な広がりという動向を可視化するためには, マクロな傾向を把握できるだけの, 代表的な地点の情報だけで十分と考える。この時, 高震度の地点

表2 新聞記事と地震データベースにおける震度情報の比較(上段:新聞記事, 下段:地震データベース)

記事ID/震度	6	5	4	3	2	1
980221352	0	0	4	6	7	1
	0	0	18	22	26	18
980423188	0	0	9	2	0	0
	0	0	10	37	>50	>50
980829289	0	0	11	13	11	0
	0	0	11	>50	>50	>50
980904003	1	0	1	7	15	0
	1	0	2	19	>50	41

だけではなく, その周辺に広がる低震度地域の情報も必要である。表2より, 低震度観測地が多数存在する様な場合に記事中では省略される場合もある(980423188)が, 全般的に高震度から低震度まで, 新聞記事はバランスよく含んでいることがわかる。これは, 2.1節で述べたとおり, フィルタされた情報としての新聞記事を動向情報可視化に用いる利点であると考える。

5. まとめ

MuSTコーパス中のタグ付けされた地震記事を対象とした, インタラクティブな地震動向情報可視化システムを提案した。提案システムは, 地震の発生頻度についての時間的動向情報だけでなく, 日本地図を用いた空間的動向情報の可視化, 群発地震などの時間的・空間的動向情報の組み合わせによる可視化が可能であることを示した。また, 既存の地震データベース検索システムとの比較などを通じて, 提案システムの有効性, 動向情報のリソースとして新聞記事を用いることの利点について論じた。

今後は, 新聞記事の解釈を裏付ける地震情報のWebからの獲得, 被害情報の新聞記事からの抽出およびWebからの収集などの拡張について検討する。

参考文献

- [1] Ed H. Chi, J. Pitkow, J. Mackinlay, P. Pirolli, R. Gossweiler, S. K. Card, "Visualizing the Evolution of Web Ecologies," Proceedings of the Conference on Human Factors in Computing Systems (CHI'98), 1998.
- [2] 加藤, 松下, 平尾, 動向情報の要約と可視化に関するワークショップの提案, 情報処理学会自然言語処理研究会, 2004-NL-164 (15), pp.89-94, 2004.
- [3] 気象庁, 震度データベース検索システム, http://www.seisvol.kishou.go.jp/eq/shindo_db/shindo_index.html, 2006.2.23 現在
- [4] A. V. Moore, "Information Flocking : Time-Varying

- Data Visualization using Boid Behaviors," 8th Int'l Conf. on Information Visualization (IV'04), pp. 409-414, 2004.
- [5] NIED 独立行政法人防災科学技術研究所, Hi-net 高感度地震観測網, <http://www.hinet.bosai.go.jp/>, 2006.2.23 現在
- [6] 日本気象協会, tenki.jp, <http://www.tenki.jp/qua/index.html>, 2006.2.23 現在
- [7] Y. Takama, K. Hirota, "Web Information Visualization Method Employing Immune Network Model for Finding Topic Stream from Document Set Sequence," J. of New Generation Computing, Vol.21, No.1, pp.49-59, 2003.
- [8] 東京大学地震研究所, 最新地震活動マップ, <http://eoc.eri.u-tokyo.ac.jp/harvest/eqmap/tkyMAP7.html>, 2006.2.23 現在
- [9] ゆめねっと, 全国市町村庁舎間直線距離方位方位角計算, <http://homepage2.nifty.com/itigo-club/yumenet.htm>, 2006.2.23 現在.
(2006年2月27日 受付)
(2006年4月20日 採録)
- [問い合わせ先]
〒191-0065 東京都日野市旭が丘6-6
首都大学東京システムデザイン学部
高間 康史
TEL: 042-585-8629
FAX: 042-585-8629
E-mail: ytakama@cc.tmit.sc.jp

 著 者 紹 介

たかま やすふみ
高間 康史 [正会員]

1994年 東京大学工学部電子工学科卒業。1999年同大学院博士課程修了。1999-2002 東京工業大学大学院総合理工学研究科助手, 2002-2005 東京都立科学技術大学助教授, 2005年より首都大学東京システムデザイン学部准教授。博士(工学)。Web Intelligenceや情報可視化, 知的インタフェースの研究に従事。主要著書は「AI事典第2版(編)」(オーム社), IEEE, 人工知能学会, 情報処理学会, 電子情報通信学会各会員。

やまだ たかし
山田 隆志 [非会員]

2006年東京都立科学技術大学工学部電子システム工学科卒業。同年首都大学東京大学院システムデザイン研究科に入学, 現在に至る。Web Intelligenceや情報可視化などの研究に従事。

Extraction and Visualization System for Earthquake Trend Information from Tagged Corpus

by

Yasufumi TAKAMA and Takashi YAMADA

Abstract :

A system for extracting and visualizing trend information about earthquakes from tagged corpus, which is distributed by a Workshop on Multimodal Summarization for Trend Information (MuST), is proposed. The topic of earthquakes does not contain only temporal trends, which are main concerns of typical topics such as gas price and stock price movement, but also spatial trends including the position of the seismic focus and the spatial distribution of its intensities. The proposed system employs the map of Japan for visualizing spatial trends, as well as bar and line graphs for temporal trends. The system also focuses on swarm earthquakes, which are visualized with the combination of the map and bar-graph representations. Furthermore, the system provides users with an interactive facility so that they can obtain several visualization results with intuitive operations such as mouse - click on the map. A prototype system is implemented, based on which the accuracy of trend information extraction from tagged corpus is evaluated, and its functionality is compared with existing earthquake database system. The merit of extracting earthquake trend information from newspaper articles is also considered.

Keywords : Trend Information, Information Visualization, Earthquake Information, Interactive Information Retrieval

Contact Address : **Yasufumi TAKAMA**

Faculty of System Design, Tokyo Metropolitan University

6-6 Asahigaoka, Hino, Tokyo 191-0065, JAPAN

TEL : 042-585-8629

FAX : 042-585-8629

E-mail : ytakama@cc.tmit.ac.jp