

地震被害情報可視化のための拡張コーパス仕様の提案

Proposal of the Extended Corpus Specification for Visualizing Earthquake Damage Information

中野 純¹ 山田 隆志¹ 松下 光範² 高間 康史¹
Jun Nakano Takashi Yamada Mitsunori Matsushita Yasufumi Takama

1. 首都大学東京
Tokyo Metropolitan University

2. 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
NTT Communication Science Labs., NTT Corp.

概要: MuSTコーパスを拡張し、地震被害情報の抽出を行うことを可能とする地震被害情報コーパスの注釈仕様を提案する。日本は世界有数の地震多発国であり、大地震発生時の被害情報は、避難者数や負傷者数、断水や停電世帯数など、一種の統計量として、逐次、新聞などにより報道されている。これらの情報は、災害時において被害状況を把握するために必要であることはもちろん、復旧後も、今後の防災対策のために活用すべき重要な情報であり、地震被害情報コーパスの構築は意義あるものとする。提案仕様に基づくコーパスを実際に構築し、これを利用したインタラクティブな情報可視化システムによりその有効性を示す。

1. はじめに

日本は世界有数の地震多発国であり、最近でも震災時の帰宅支援マップがベストセラーとなるなど多くの人が地震災害に関心を持っている。大地震発生時の被害情報は、避難者数や負傷者数、停電や断水している世帯数などの一種の統計量として、逐次新聞などにより報道されている。これらの情報は、地震発生時においても被害状況を把握する上で重要であるが、復旧後も今後の防災対策に生かすために活用すべき、重要な情報であるとする。

我々はこれまでに、動向情報の要約と可視化に関するワークショップ (MuST) [2] の提供する研究用データセットであるタグ付きコーパス (MuSTコーパス) 中の新聞記事を対象とした地震動向情報可視化システムを開発している [7]。しかし、MuSTコーパスでは震源地および震度観測地のみを対象としており、地震によって引き起こされる被害状況については抽出・可視化の対象外としていた。そこで本稿では、新聞記事中の地震被害情報に注釈の付与を可能とするため、MuSTコーパスの注釈仕様を拡張した、地震被害情報コーパス [4] を提案・構築する。新潟中越地震に関して構築した地震被害情報コーパスを用いて、被害に関する時空間的動向をインタラクティブに表現する、地震被害情報可視化システムを開発し、被験者実験によりその有効性を示す。

次節では、被害情報を扱うために、MuSTコーパスの注釈仕様に対して施した拡張について提案し、MuSTコーパスとの比較を行う。3節では、地震被害情報可視化システムの詳細を述べる。最後に、構築した被害情報コーパスに基づく可視化システムの有効性を4節で示す。

2. 地震被害情報の抽出

2.1 新聞記事における地震被害情報

新聞記事に含まれる、地震被害に関する動向情報の記述には以下の2つが挙げられる。図1、2は毎日新聞2004年10月24日朝刊の一部である。

・人的被害に関する記述

地震による死亡者数、負傷者数、行方不明者数などの記述。死亡者数と行方不明者数は変動があれば市区町村単位で正確な値が発表されるが、負傷者数と避難者数はほぼ毎日概算で発表される。

同県内で10人が死亡、10人が行方不明となっている。毎日新聞の24日午前0時半現在のまとめでは、けが人は640人を超えた。

図1: 人的被害に関する新聞記事の例

・ライフラインの被害状況に関する記述

地震による停電世帯数、ガス供給停止世帯数、断水世帯数などの記述。被害件数の発表はほぼ毎日、市区町村単位で概算で発表される。

東北電力によると、午後10時現在、約9万5000戸が停電。都市ガスは約5万6000戸で供給が止まり、断水は約11万戸に達した。

図2: ライフラインの被害に関する新聞記事の例

2.2 地震被害情報のための注釈仕様

地震被害情報を抽出するにあたり、MuSTコーパスでは地震等の災害による被害情報に対する注釈付けは考慮されていないため、現状の注釈仕様ではそれらに対する注釈付けには不十分である。そこで本

稿では、地震被害情報に対する注釈付けが可能なように MuST コーパスを拡張することを提案する。具体的には、既存の要素により注釈が付与される対象を拡張すると共に、新たな属性を追加することにより、地震規模だけでなく、人的被害（死亡者数・負傷者数・行方不明者数・避難者数）やライフラインの被害状況など、震災による被害情報へ注釈付けを可能とする。以下に地震被害情報コーパスのために追加した仕様を説明する。

• unit 要素について

この要素が示す出来事や統計量の種類を type 属性によって表す。属性値として地震が発生したという事実について言及する部分には“地震”，マグニチュードなど地震の規模について言及している部分には“規模”，死亡者数など被害情報について言及している部分には“被害”をとる。“規模”と“被害”は of 属性を付与することにより，もととなる地震について言及する部分を参照する。

• val 要素

この要素が示す統計量の種類を type 属性によって表す。area 属性で対応する被害範囲を参照する。

• par 要素

この要素の示すパラメータの種類を type 属性によって表す。type 属性が“被災地”など地名をさす場合，地域範囲に応じて“地方”あるいは“市区町村”といった gra 属性を付与する。date 属性で対応する地震の起こった日付を参照する。

• date 要素

id 属性を付与することにより，後に（par 要素などから）参照を可能とする。

表 1 に MuST コーパスと地震被害情報コーパスの unit 要素，val 要素，par 要素について比較した表を示す。

表 1：MuST コーパスと地震被害情報コーパスの要素比較

	MuST コーパス	地震被害情報コーパス
unit 要素	地震発生の記事を event 属性を用いて表す	地震発生，地震規模，地震被害を type 属性を用いて表す
val 要素	なし	人的被害，ライフライン被害を表す統計量の値に付与。パラメータの種類を type 属性によって表す
par 要素	地震規模を表すパラメータに付与	地震規模を表すパラメータに付与。パラメータの種類を type 属性によって表す

毎日新聞 2004 年版 10 月 24 日朝刊記事の一部に、MuST コーパスに基づき注釈付けを行ったものを図 3、地震被害情報コーパスに基づき注釈付けを行ったものを図 4 に示す。

MuST コーパス（図 3）では注釈付け対象を地震発生時間、代表的な震度観測地とその地域での震度、マグニチュードのみとしている。そのため、震源地や震源の深さなど、地震規模を詳細に記述している節に del 要素を付与することにより，抽出対象外としている。

地震被害情報コーパス（図 4）では，unit 要素に「type=“規模”」属性を付与することにより，震源地や震源の深さなどを，地震の規模情報として抽出対象としている。また「type=“地震”」属性を付与することにより地震の発生に関する記述部分を指定し，被害情報に関する記述部分には「type=“被害”」属性を付与している。さらに地震被害（被害者数）を表す記述にもタグを付与することにより被害情報を抽出対象としている。

```

<unit event="地震">
<date gra="時" abs="20041023">2 3 日午後 5 時 5
6 分ごろ</date>
、
<par><rft id="041024003_1">新潟県</rft></par>
を中心に強い地震があり、
<par>新潟県小千谷市</par>
で
<par>震度 6 強</par>
の揺れを観測した。
<del type="src">気象庁によると、</del>
<del type="oot">震源地は同県中越地方で、</del>
<del type="oot">震源の深さは約 2 0 キロ</del>
、地震の規模を示すマグニチュード (M) は
<par>6・8</par>
と推定される。
</unit>
同県内で 1 0 人が死亡、1 0 人が行方不明となっ
ている。毎日新聞の 2 4 日午前 0 時半現在のまとめ
では、けが人は 6 4 0 人を超えた。住宅倒壊や土砂崩
れが各地で発生し、上越新幹線は浦佐―長岡間で脱
線した。

```

図 3：MuST 注釈仕様を用いた注釈付けの例

2. 3 地震被害情報コーパス

毎日新聞 CD-ROM 2004 年版を利用して，地震被害情報コーパスを作成した。本研究では，新潟中越地震に関する 44 記事（10 月 24 日～11 月 12 日）に対し，人手による注釈を付与した。

人的被害に関する 23 記事を対象に注釈付けを行った。各要素における注釈付与数は unit 要素 26 個，val 要素 40 個である。抽出結果によって得られるデータテーブルは以下の通りである。

```

<unit id="041024M01_1" type="地震">
<date id="041024M01_1_1" gra="時"
abs="20041023">2 3 日午後 5 時 5 6 分ごろ
</date>、
<rft id="041024M01_1_2">新潟県</rft>
を中心に強い地震があり、
<par type="被災地" gra="市区町村"
date="041024M01_1_1">新潟県小千谷市
</par>
で
<par type="震度" date="041024M01_1_1">
震度 6 強</par>
の揺れを観測した。
</unit>
<unit id="041024M01_2" type="規模"
of="041024M01_1">
<del type="src">気象庁によると、</del>
震源地は
<par type="震源地" date="041024M01_1_1">
<pro ref="041024M01_1_2">
同県</pro>中越地方</par>
で、震源の深さは
<par type="震源の深さ" date="041024M01_1_1">
約 2 0 キロ</par>
、地震の規模を示すマグニチュード (M) は
<par type="マグニチュード"
date="041024M01_1_1">6 . 8</par>
と推定される。
</unit>
<unit id="041024M01_3" type="被害"
of="041024M01_1">
<par id="041024M01_3_1" type="被災地">
<pro ref="041024M01_1_2">同県</pro></par>
内で
<val type="死亡者数" area="041024M01_3_1">
1 0 人</val>
が死亡、
<val type="行方不明者数"
area="041024M01_3_1">1 0 人</val>
が行方不明となっている。
<del type="src">
毎日新聞の 2 4 日午前 0 時半現在のまとめでは、
</del>
けが人は
<val type="負傷者数" area="041024M01_3_1">
6 4 0 人</val>
を超えた。
</unit>
住宅倒壊や土砂崩れが各地で発生し、上越新幹線は
浦佐―長岡間で脱線した。

```

図 4: 地震被害情報抽出のための注釈付け

- ・ 時間情報 (発表日時)
- ・ 都道府県名
- ・ 市区町村名

- ・ 被害の種類 (死亡, 負傷, 行方不明, 避難)
- ・ 被害人数

ライフラインに関する 21 記事を対象に注釈付けを行った. 各要素における注釈付与数は unit 要素 63 個, val 要素 63 個である. 抽出結果によって得られるデータテーブルは以下の通りである.

- ・ 時間情報 (発表日時)
- ・ 都道府県名
- ・ 市区町村名
- ・ ライフラインの種類 (電気, ガス, 水道)
- ・ 被害件数

地震発生に関する 9 記事を対象に注釈付けを行った. 各要素における注釈付与数は unit 要素 31 個, par 要素 73 個である. 抽出結果によって得られるデータテーブルは以下の通りである.

【震源に関するデータ】

- ・ 時間情報 (発表日時)
- ・ 震源地
- ・ マグニチュード

【震度観測地に関するデータ】

- ・ 時間情報 (発生日時)
- ・ 都道府県名
- ・ 市区町村名
- ・ 震度

3 . 地震被害情報の可視化

3 . 1 時間的動向情報の可視化

被害状況の時間的推移を図 5(a)のようなグラフを用いて表す. 時間的動向情報を, 縦軸に被害件数とマグニチュード, 横軸に被害日時をとる折れ線グラフと棒グラフを用いて表示する. このグラフによって人的被害 (死亡者数, 負傷者数, 行方不明者数, 避難者数), ライフライン被害 (停電戸数, ガス供給停止戸数, 断水戸数), 地震規模 (マグニチュード) が表示可能である. 関連する被害情報を時系列順にまとめて視覚化することで, 時間的な動向を把握することが容易となる. さらに, グラフ上の各点を選択することにより, その点における被害地域の分布を空間的動向情報として提示可能である.

3 . 2 空間的動向情報の可視化

被害地域の分布を図 5(b)のような日本地図を用いて表す. 本システムの地図表示は HieX [6] をベースに作成している. 図 5(b)は中越地震におけるガス供給停止地域の分布を表している. 空間的動向情報として人的被害 (死亡者数, 負傷者数, 行方不明者数, 避難者数) の分布, ライフライン被害 (停電戸数, ガス供給停止戸数, 断水戸数) の分布, 各地の震度が表現可能である. 被害の深刻さ (被害者数, 被害戸数) は赤色の濃度によって表現される. 空間粒度は市区町村単位で表示可能である. 地図を利用することにより, 地震による影響の範囲などを把握することが容易となる. また, 地図の被害地域を選択することにより, その地域における被害の時間的推移を提示可能である.

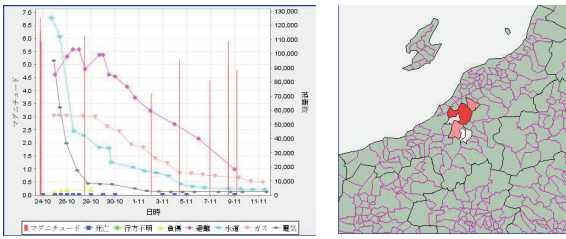


図 5(a) : 時間的動向情報の可視化(左),
(b) : 空間的動向情報の可視化(右)

3. 3 地震被害情報可視化システム

地震被害情報コーパスから被害情報を抽出し、地震被害に関する時間的、空間的動向情報を可視化してユーザに提示する、地震被害情報可視化システムを実装する。図 6 に、可視化システムの構成を示す。可視化システムは地震被害情報コーパスから動向情報抽出部により地震被害情報データベースを構築する処理（オフライン処理）と、ユーザの要求に応じた可視化表現を生成し、インタラクティブに提示する処理（オンライン処理）に分けられる。

地震被害情報コーパスから XML パーサを用い情報抽出を抽出し、被害地域を市区町村 DB より補完することにより、地震被害情報 DB を作成する。地震被害情報 DB と市区町村 DB の緯度経度情報を用いて地震被害情報の可視化を行う。本システムは Web 開発ツール grails を利用して実装することにより、ブラウザ上で閲覧可能な Web アプリケーションとなっている。

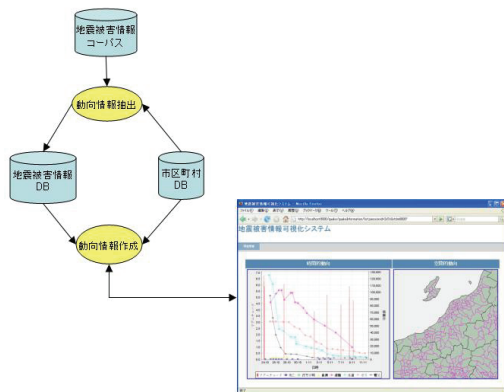


図 6 : 可視化システムの構成

4. 可視化システムの有効性の検証

地震被害情報可視化システムが、地震被害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。震災の復旧状況を表す資料 [1, 5] などに、震災被害の状況を示すものとして折れ線グラフと地図が用いられている。提案する被害情報可視化システムも同等の可視化表現を採用しており、さらに地域などの粒度をインタラクティブに変更可能であるため、実用上十分であるといえる。

また、地震被害情報を提示する既存システムとして、内閣府の地震被害想定支援ツール [3] がある。地震被害想定支援ツールは被害状況を地図のみで提

示するのに対し、提案システムはグラフと地図を用い、かつインタラクティブに両表現を対応付けることが可能である点で優れていると考える。

地震に対する知識・意識があまり高くない人に提案システムを使用してもらったところ、システム使用後のアンケートで「グラフ上の興味をもった点を選択することにより、その点における被害分布が地図を用いて表されており、被害状況がより一層伝わった」、「震災の恐ろしさが伝わった」、「防災に対する意識が高まった」などのコメントを得ることができた。これらの結果より、地震被害情報の抽出・可視化は防災意識の向上に貢献することがわかる。

5. おわりに

本稿では、MuST コーパスの拡張仕様である地震被害情報コーパスを提案し、新潟中越地震に関して構築したコーパスを利用したインタラクティブな情報可視化システムを実装することによりその有効性を示した。地震被害情報コーパスを構築することにより、地震規模だけでなく、地震によってもたらされた被害を抽出することが容易となり、震災の様子を様々な観点から提示し防災計画などに活用できることが期待できる。

今後は、被害の根拠となる要因を新聞記事から抽出可能な拡張を行うことにより、詳細な被害情報を伝えることが可能となると考える。

謝辞

本研究は日本電信電話株式会社 NTT コミュニケーション科学基礎研究所との共同研究の一部として行ったものであり、ここに謝意を表す。

参考文献

- [1] 財団法人電力中央研究所, <http://criepi.denken.or.jp/jp/>
- [2] 動向情報の要約と可視化に関するワークショップ (MuST), <http://must.c.u-tokyo.ac.jp>
- [3] 内閣府防災情報, <http://www.bousai.go.jp/>
- [4] 中野 純, 山田 隆志, 高間 康史: 地震被害情報コーパス作成のための注釈仕様の提案, 第 27 回ファジィ・ワークショップ/第 15 回北信越支部シンポジウム, pp. 27-30 2006.
- [5] 秦 康範, 目黒 公郎: 兵庫県南部地震後の各種ライフラインの復旧活動と復旧阻害要因の影響について, 第 25 回地震工学研究発表会論文集, 土木学会, pp.1077-1080, 1999.
- [6] 松下 光範, 加藤 恒昭: 自然言語要求に基づく対話的情報提供, 第 22 回ファジィ・ワークショップ, pp. 23-26, 2004.
- [7] 山田 隆志, 中野 純, 高間 康史: タグ付きコーパスを用いた地震記事からの地理的動向情報の可視化, 言語処理学会第 12 回年次大会ワークショップ「言語処理と情報可視化の接点」, pp. 9-12, 2006.