

地震統計との相関に見るウェブセンサの可能性

服部 峻†

† 東京工科大学 コンピュータサイエンス学部

〒 192-0982 東京都八王子市片倉町 1404-1

E-mail: †hattori@cs.teu.ac.jp

あらまし 実世界で起きた様々な現象や事象に関する知識データを情報爆発し続けるウェブから、特にブログなどのCGMからマイニングする研究が盛んに行われ、一般向けへのサービス化も試行され始めている。しかしながら、大量のウェブ文書からマイニングされたデータが、実世界でのリアルなデータをどの程度正確に反映しているかの詳細な調査は未だ無く、精度が十分に保証されていないまま盲目的に利用するのは問題があると考えられる。そこで本論文では、実世界で起きた現象のリアルなデータとして気象庁の地震統計から震度情報を用い、ウェブから抽出した時空間依存データとの相関を評価することで、ウェブセンサによってマイニングされたデータの利用可能性や信頼性を検証する。**キーワード** ウェブマイニング, ウェブ抽出, ウェブの信頼性, ウェブセンサ, 時空間依存データ, 時系列データ, 地理情報システム.

Potentials of Web Sensors in Correlation with Earthquake Statistics

Shun HATTORI†

† School of Computer Science, Tokyo University of Technology

1404-1 Katakura, Hachioji, Tokyo 192-0982, Japan

E-mail: †hattori@cs.teu.ac.jp

Abstract Many researches on mining the explosively-growing Web, especially CGM (Consumer Generated Media) such as Weblogs, for knowledge about various phenomena and events in the real world have been done actively, and Web services with the Web-mined knowledge have begun to be developed for the public. However, there is no detailed investigation on how accurately Web-mined data reflect real-world data. It must be problematic to idola-trously utilize the Web-mined data in public Web services without ensuring their accuracy sufficiently. Therefore, this paper validates potentials and reliability of Web sensor's spatio-temporal data by measuring the correlation coefficient with earthquake statistics of Japan Meteorological Agency as real-world data.

Key words Web Mining, Web Extraction, Web Credibility, Web Sensor, Spatio-Temporal Data, Time-Series Data, Geographical Information System (GIS).

1. はじめに

かつてのウェブ世界は実世界とは互いに疎な関係で独立した存在と言っても過言ではなかったが、ウェブの利用が広く一般の老若男女に普及し、加えてブログやロコミサイト、SNSといったCGM (Consumer Generated Media) が非常に盛んになって来ており、少数のプロの書き手や編集者から成る新聞社などの組織だけでなく、大多数の一般消費者個人によって、実世界で実際に起きた (体験した)、或いは、今後起こるであろう様々な現象や事象 (イベント) に関して、ウェブ文書として記述されることが多くなり、ウェブ世界は実世界と互いにより密な関係になって来ている。

このようにウェブ世界と実世界との関係がより密になるに伴って、ウェブ世界はウェブ独自のサービスや活動だけでなく、実世界でどのような現象や事象が起きているのか、どのように変化しているのかをモニターするための情報源 (センサ) としての側面にも注目されており、実世界での様々な現象に関する知識をウェブ世界から抽出、可視化するための手法、その知識の活用方法などが盛んに研究されている。例えば、実世界で提供されている製品やサービスなどの評判抽出 [1], [2], 実世界のある場所である期間に味わうことが出来る体験 (イベント) の抽出およびマッピング [3]~[5] などが提案されている。これらのウェブから抽出された情報の一部は既にウェブサービスとして一般にも提供され始めており、実世界で製品やサービス、行

動を選択する際に多くの一般ユーザが参考にするようになって来ている。著者も、語概念の階層構造 (is-a/has-a 関係など) の抽出 [6], [7], 実世界オブジェクトの外観などの五感情報の抽出 [8]~[10] など, 実世界の様々な事象に関する知識を情報爆発し続けるウェブから, 特にブログなどの CGM からマイニングする研究を行っている。一方で, 実空間に様々な ICT 機器を埋め込み, 実世界でサービス配置や構造を最適化し, スマート (セキュア) 空間 [11] を実現するための研究も行っているが, 従来の物理的なリアルセンサだけでは, ある場所ある期間に起きた様々な現象に関して, その現象を人々がどのように認識しているか, 評判や印象などをセンサすることは困難である。そこで本研究では, 実世界での様々な現象や事象を人間自身がセンサ (体験) して記述したウェブ文書を活用して, 自動的に時空間依存データを高精度に抽出するウェブセンサ技術の開発を行っている。

しかしながら, 実世界で実際に起きた, 或いは, 今後起こるであろう現象や事象 (イベント) について, どの程度正確にウェブ世界にウェブ文書として記述されているのか, 実世界をモニターするための情報源 (センサ) としてのウェブの利用可能性や信頼性などについて, 詳細な調査は未だ不十分であると考える。マイニング技術の進歩によりウェブから何らかの知識らしいデータを抽出することは全く難しくなく, これらのデータを単に見て楽しむだけであるならば特に問題は無いかもしれないが, 実世界のセンサとしてのウェブの利用可能性や信頼性などが保証されていないままでは, 実世界で製品やサービス, 行動を選択する際に, これらのデータを本当に参考にしても良いのかは非常に怪しく, よりクリティカルなシステムへの活用には大きな問題があると考えられる。Web 世界の実態に詳しく「そもそもウェブは疑わしいもの」と認識している (研究) 者であればウェブから抽出されたデータを過信 (盲信) することは無いかもしれないが, 鵜呑みにしてしまう一般ユーザは少なくないであろう [12]。

ウェブから何らかの知識データを抽出する手法の研究やサービスの開発は既に多種多様に行われているが, 時空間依存データ (実世界上のある地理的位置である期間に起きた現象や事象に関するデータ) をウェブから抽出する手法 (本論文ではウェブセンサと呼ぶ) の研究は未だ少なく, ウェブセンサによってウェブ抽出された時空間依存データの利用可能性や信頼性についての詳細な検証も未だ不十分である。そこで, 実世界の様々な現象や事象に関して, 実世界に設置されたリアルセンサによって収集された統計データと, 人間自身がセンサとなって体験して記述したウェブ文書からウェブセンサによってウェブ抽出された時空間依存データとを多角的に照合および考察することで, ウェブセンサの利用可能性や信頼性を担保することは (社会的にも) 非常に重要であると考えられる。本調査により, ウェブセンサによってマイニングされたデータの利用可能性や信頼性が保証されれば, 一般ユーザが日常生活において製品やサービス, 行動を選択する際に誰もがより安心して気軽に参考にすることが出来るようになり, 同時に, ウェブセンサのデータを利活用した応用システムもより拡充されて行くと考えられる。また,

実世界に埋め込む必要があり, コストも掛かる物理的なリアルセンサを代替・補完でき, 実世界のある場所についてウェブ抽出された印象や要望などがこれまでのような参考程度のデータではなく, 潜在的に抱えている積極的に取り組むべき課題であるとの認識も広まり, ウェブ世界と実世界との間により良い循環を生むことにも貢献できると考える。

本論文では, 実世界で実際に起きた現象のリアルなデータとして, 気象庁の地震統計 [13] から震度情報を用いる。従って, まず, 実世界で起きた震度に関する時空間依存データをウェブ抽出するためのウェブセンサを構成する必要がある。ウェブから何らかの知識データを抽出する手法は様々な提案されているが, ウェブ検索エンジンのインデックス情報 (検索件数やスニペットなど) を用い, 特定の時空間で制限したウェブ文書の中で「地震」や「揺れ」といった言葉を含む頻度に基づいてウェブセンサを構成する。その上で, 実世界の震度に関してウェブ世界から抽出した時空間依存データと気象庁の地震統計との相関を多角的に評価することで, ウェブセンサによってマイニングされたデータの利用可能性や信頼性を検証する。これまでの研究においては, 実世界での現象として, 気温や降雨量, 事故件数などを取り上げ, ウェブセンサの可能性を調査している [14]~[16]。交通事故は非常に局所的かつ局所的な現象, 降雨はやや広域かつ長期にわたる現象, 気温はより広域で周期的な現象である。一方, 本論文で利用する震度は, 時間依存性 (局時性) や空間依存性 (局所性) の観点からも異なる現象であり, 新たな知見の獲得が期待できる。

本論文の以下の構成を示す。まず, 2 章では, 実世界の震度に関する時空間依存データをウェブ抽出するウェブセンサの構成方法を示す。次に, 3 章では, 2 章で構成したウェブセンサによってマイニングした時空間依存データと, 気象庁の地震統計の震度情報との相関を多角的に評価・考察する。最後に, 4 章で本論文をまとめ, 今後の課題についても述べる。

2. 震度に関する時空間依存データを抽出するウェブセンサの構成方法

次章では, ウェブセンサによってウェブ抽出した時空間依存データと, 気象庁の地震統計の震度情報との相関を多角的に評価・考察する。従って, 実世界の現象として震度に関する時空間依存データを抽出するウェブセンサを構成する必要がある。ウェブから何らかの知識らしいデータを抽出する手法は様々な提案されているが, 本論文では, ウェブ検索エンジンのインデックス情報の一つである検索件数だけを用い, 特定の時空間で制限したブログ文書の中で, 対象の現象 (地震) を連想させる「地震」や「揺れ」といった言葉を含む文書頻度に基づいてウェブセンサを構成する。一般のウェブ文書よりもブログ文書に限定した方が, より良いウェブセンサを構成できることはこれまでの研究で報告している [15]。

まず, 検索対象のブログ文書を特定の時間および空間で制限する必要がある。時間制約に関しては, Google ブログ検索 [17] がサポートしている期間指定オプションを利用する。但し, この期間指定オプションはブログ文書の最終更新日 (クロールさ

れた日)に基づいて制限するため、必ずしも真に必要な特定の時間に起きた実世界の現象について記述されたブログ文書であるとは限らないかもしれない。このようなブログ文書のメタデータとしての日付ではなく、ブログ文書の内容に含まれる日付表現を切り出す方法も考えられるが、日付表現は多様であり、検索件数を求めるために検索クエリを構成するのが容易ではないため、本論文では Google の期間指定オプションを利用している。例えば、「2011 年 3 月 1 日」という期間だけにブログ文書を限定するためには、開始日「2011/3/1」から終了日「2011/3/1」までという期間指定オプションを設定する。空間制約に関しては、Google の期間指定オプションのようにブログ文書のメタデータを用いて制限することができないため、特定の空間(地域)表現を内容に含むか否かでブログ文書を制限する。例えば、「宮城県」という地域だけにブログ文書を限定するためには、「宮城県」という言葉を検索クエリ自体に含める。

次に、時間 t および空間 s における実世界の現象「震度」に関するウェブセンサが出力する時空間依存データ(数値)を、「地震(quake)」「揺れ(shake)」「強い揺れ(forceful)」「弱い揺れ(weak)」といった記述を含むブログ文書の頻度で定義する。

$$\text{WebSensor}_q^0(t, s) := \text{bf}_t(["s" \text{ AND } "地震"])$$

$$\text{WebSensor}_s^0(t, s) := \text{bf}_t(["s" \text{ AND } "揺れ"])$$

$$\text{WebSensor}_f^0(t, s) := \text{bf}_t(["s" \text{ AND } "揺れ" \text{ AND } "強い"])$$

$$\text{WebSensor}_w^0(t, s) := \text{bf}_t(["s" \text{ AND } "揺れ" \text{ AND } "弱い"])$$

但し、 $\text{bf}_t(q)$ は、時間 t で期間指定した上で検索クエリ q で Google ブログ検索した結果のブログ文書の検索件数を表す。

また、文書頻度は空間 s (にいる人間の数の違い等) に依存し易いが、この空間依存性を出来る限り排除するために以下のように空間を表す記述のブログ文書頻度で正規化する。

$$\text{WebSensor}_x^1(t, s) := \frac{\text{WebSensor}_x^0(t, s)}{\text{bf}_t(["s"])}$$

$$\text{WebSensor}_x^2(t, s) := \frac{\text{WebSensor}_x^0(t, s)}{\sqrt{\text{bf}_t(["s"])}}$$

但し、 x は、 q, s, f, w のいずれかである。

さらに、上述の指標複数をパラメータで線形結合することによって、以下のようなウェブセンサも構成し、検証を行う。

$$\begin{aligned} \text{WebSensor}_{fw}^y(t, s) := & (1 - \alpha) \cdot \text{WebSensor}_f^y(t, s) \\ & + \alpha \cdot \text{WebSensor}_w^y(t, s) \end{aligned}$$

$$\begin{aligned} \text{WebSensor}_{sfw}^y(t, s) := & (1 - \alpha - \beta) \cdot \text{WebSensor}_s^y(t, s) \\ & + \beta \cdot \text{WebSensor}_f^y(t, s) \\ & + \alpha \cdot \text{WebSensor}_w^y(t, s) \end{aligned}$$

但し、 y は、 $0, 1, 2$ のいずれかである。また、前者の式においてパラメータ α は $0 \leq \alpha \leq 1$ を満たし、後者においてパラメータ α, β は $0 \leq 1 - \alpha - \beta \leq 1$ を満たす 0 以上の実数値である。

3. ウェブセンサと地震統計との相関の評価

本章では、前章で構成したウェブセンサによってウェブ抽出した時空間依存データと、気象庁の地震統計の震度情報との相関を多角的に評価することで、ウェブセンサによってマイニングされたデータの利用可能性や信頼性を検証する。特にウェブセンサの空間依存性や時間依存性について考察して行く。

時間制約としては「2011 年 3 月 1 日」から「2011 年 5 月 31 日」までの日毎、全部で 92 区間を用いる。空間制約としては 47 都道府県の名称を用いる。ウェブセンサの示す数値データとしては、各期間でオプション指定した上で「地震」や「揺れ」などを各都道府県名で拡張した検索クエリで検索したブログ文書の頻度に基づく。一方、実世界で実際に起きた現象のリアルなデータとしては、同日に何度地震があったとしても、気象庁の地震統計の震度情報 [13] のうち日毎の最大の震度階級を用いる。そして、気象庁震度階級表 [18] に基づいて以下の表 1 のように数値データに変換する。

表 1 気象庁震度階級から数値データへの変換

Table 1 JMA Seismic Intensity Scale to Numerical Data.

震度階級	0	1	2	3	4	5 弱	5 強	6 弱	6 強	7
数値データ	0	1	2	3	4	4.75	5.25	5.75	6.25	7

表 2 は、実世界の現象「震度」に関する記述を含むブログ文書の頻度のみで単純に構成した $\text{WebSensor}_x^0(t, s)$ 、及び、ウェブセンサの空間依存性を排除するために各空間を表す記述のブログ文書頻度で正規化を行った $\text{WebSensor}_x^1(t, s)$ や $\text{WebSensor}_x^2(t, s)$ によってウェブ抽出した時空間依存データと、リアルセンサによる統計データとして気象庁の震度情報との相関を比較している。実世界の現象「震度」に対して「弱い揺れ」という記述に基づき正規化しないで構成した $\text{WebSensor}_w^0(t, s)$ との相関が平均的には最も高いが、正規化していないため空間依存性はやや強い。しかし、Google ブログ検索を用いて取得した各空間を表す記述のブログ文書頻度はおよその概算になるケースが多いため、正規化しても必ずしも空間依存性が減少していない。

表 2 $\text{WebSensor}_x^y(t, s)$ と地震統計との相関

Table 2 Correlation of $\text{WebSensor}_x^y(t, s)$ with Quake Statistics.

	平均	最大	最小	標準偏差
$\text{WebSensor}_q^0(t, s)$	0.2323	0.5265	-0.1606	0.1946
$\text{WebSensor}_s^0(t, s)$	0.2534	0.5346	-0.1592	0.1832
$\text{WebSensor}_f^0(t, s)$	0.2385	0.5118	-0.1396	0.1802
$\text{WebSensor}_w^0(t, s)$	0.2668	0.7178	-0.0792	0.2002
$\text{WebSensor}_q^1(t, s)$	0.2437	0.6296	-0.1833	0.2217
$\text{WebSensor}_s^1(t, s)$	0.2617	0.5672	-0.0650	0.1786
$\text{WebSensor}_f^1(t, s)$	0.2390	0.5322	-0.1131	0.1746
$\text{WebSensor}_w^1(t, s)$	0.2035	0.6916	-0.1081	0.1771
$\text{WebSensor}_q^2(t, s)$	0.2407	0.5480	-0.1511	0.1940
$\text{WebSensor}_s^2(t, s)$	0.2607	0.5454	-0.1368	0.1823
$\text{WebSensor}_f^2(t, s)$	0.2442	0.5303	-0.1334	0.1796
$\text{WebSensor}_w^2(t, s)$	0.2563	0.7216	-0.0884	0.1957

図1から図3は、単純ブログ文書頻度またはそれを正規化した指標を二つ線形結合した $\text{WebSensor}_{\text{fw}}^y(t, s)$ と気象庁の震度情報との相関に関して、パラメータ α への依存性を示している。複数を線形結合した方が相関が強くなること、「強い揺れ」のブログ文書頻度よりも「弱い揺れ」のブログ文書頻度を優先すべきであることが分かる。

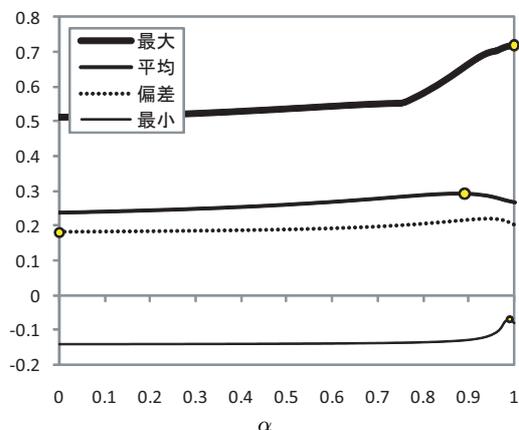


図1 $\text{WebSensor}_{\text{fw}}^0(t, s)$ と震度との相関の α 依存性 ($\alpha = 0.88$ のとき最大値 0.2907 をとる)

Fig. 1 Dependency of $\text{WebSensor}_{\text{fw}}^0(t, s)$ on α .

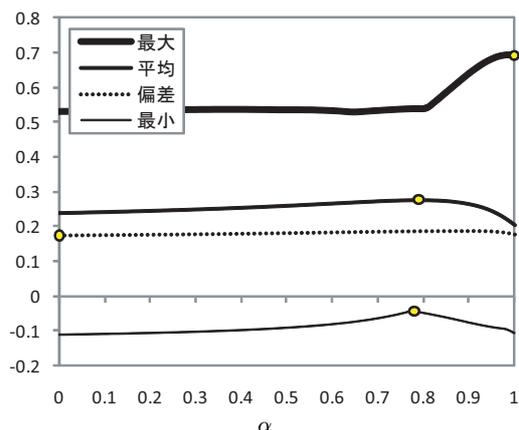


図2 $\text{WebSensor}_{\text{fw}}^1(t, s)$ と震度との相関の α 依存性 ($\alpha = 0.79$ のとき最大値 0.2769 をとる)

Fig. 2 Dependency of $\text{WebSensor}_{\text{fw}}^1(t, s)$ on α .

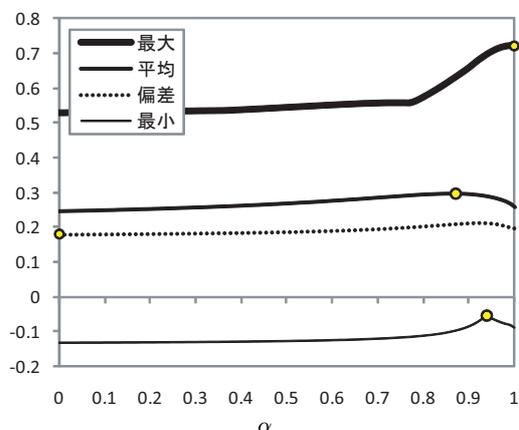


図3 $\text{WebSensor}_{\text{fw}}^2(t, s)$ と震度との相関の α 依存性 ($\alpha = 0.86$ のとき最大値 0.2959 をとる)

Fig. 3 Dependency of $\text{WebSensor}_{\text{fw}}^2(t, s)$ on α .

図4から図6は、単純ブログ文書頻度またはそれを正規化した指標を三つ線形結合した $\text{WebSensor}_{\text{sfw}}^y(t, s)$ と気象庁の震度情報との相関に関して、パラメータ α および β への依存性を示している。先の図と同様に「強い揺れ」のブログ文書頻度よりも「弱い揺れ」のブログ文書頻度を優先すべきであること、二つよりも三つ線形結合した方が相関が強くなること、つまり、形容表現を含まない「揺れ」のブログ文書頻度をわずかに加味した方が良いことが分かる。本論文で構成したウェブセンサの中では、 $\text{WebSensor}_{\text{sfw}}^2(t, s)$ ($\alpha = 0.84, \beta = 0.13$) が気象庁の震度情報と最も相関が強いという結果になった。

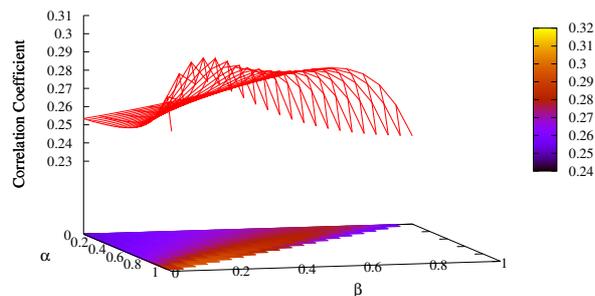


図4 $\text{WebSensor}_{\text{sfw}}^0(t, s)$ と震度との相関の α, β 依存性 ($\alpha = 0.85, \beta = 0.12$ のとき最大値 0.3045 をとる)

Fig. 4 Dependency of $\text{WebSensor}_{\text{sfw}}^0(t, s)$ on α, β .

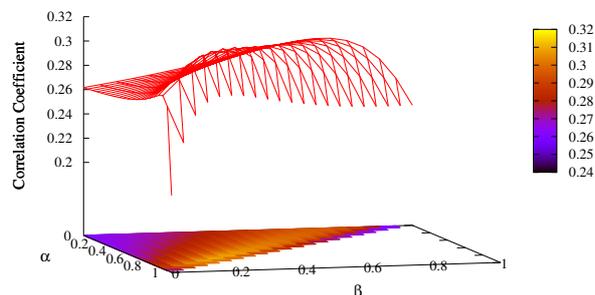


図5 $\text{WebSensor}_{\text{sfw}}^1(t, s)$ と震度との相関の α, β 依存性 ($\alpha = 0.73, \beta = 0.20$ のとき最大値 0.3133 をとる)

Fig. 5 Dependency of $\text{WebSensor}_{\text{sfw}}^1(t, s)$ on α, β .

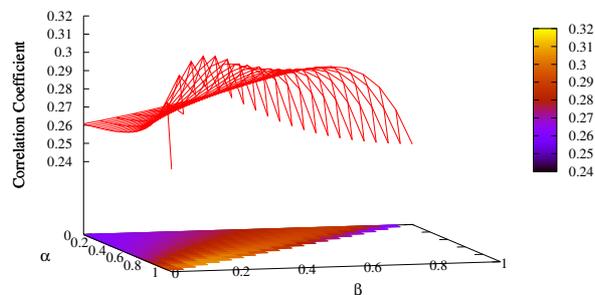


図6 $\text{WebSensor}_{\text{sfw}}^2(t, s)$ と震度との相関の α, β 依存性 ($\alpha = 0.84, \beta = 0.13$ のとき最大値 0.3145 をとる)

Fig. 6 Dependency of $\text{WebSensor}_{\text{sfw}}^2(t, s)$ on α, β .

図 7 は、 $\text{WebSensor}_{\text{sffw}}^2(t, s)$ ($\alpha = 0.84, \beta = 0.13$) と気象庁の震度情報との相関を都道府県毎に 92 日データに基づいて求めることで、ウェブセンサの空間依存性を示している。また、図 8 は、東日本大震災において最大の震度 7 を「宮城県」で記録した 2011 年 3 月 11 日の都道府県毎の最大震度の分布である。翌日にも「長野県」を中心として震度 6 強と非常に強い地震を記録していることから、大きな震度を記録した地域に対してウェブセンサで抽出した時空間依存データは相関が強く、有感地震をあまり記録しなかった地域ほど相関が弱いという傾向が見て取れる。図 9 は相関が強かった空間「宮城県」に対して、逆に図 10 は相関が最も弱かった空間「宮崎県」に対して、 $\text{WebSensor}_{\text{sffw}}^2(t, s)$ ($\alpha = 0.84, \beta = 0.13$) でウェブ抽出した時空間依存データと気象庁の震度データとを「2011 年 3 月 1 日」から「2011 年 5 月 31 日」までプロットしたグラフである。

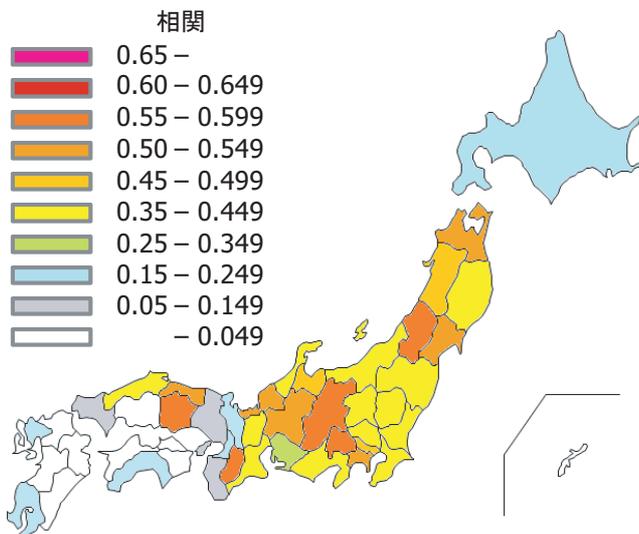


図 7 $\text{WebSensor}_{\text{sffw}}^2(t, s)$ と震度との相関の空間依存性
Fig. 7 Spatial Dependency of $\text{WebSensor}_{\text{sffw}}^2(t, s)$.

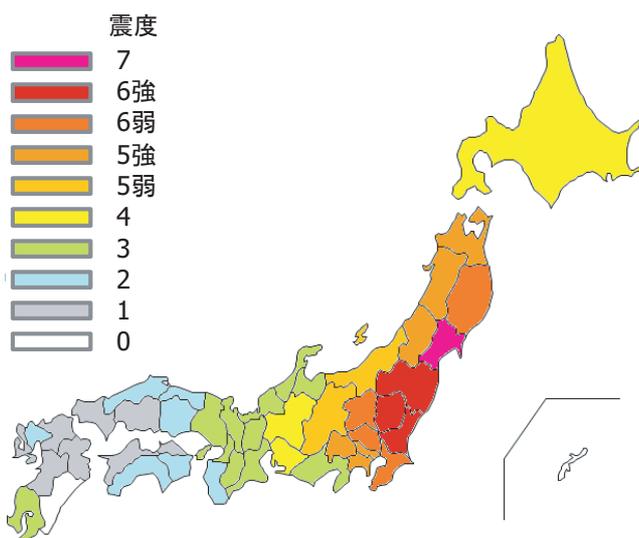


図 8 2011 年 3 月 11 日の都道府県毎の最高震度の分布
Fig. 8 Distribution of Seismic Intensity on March 11th, 2011.

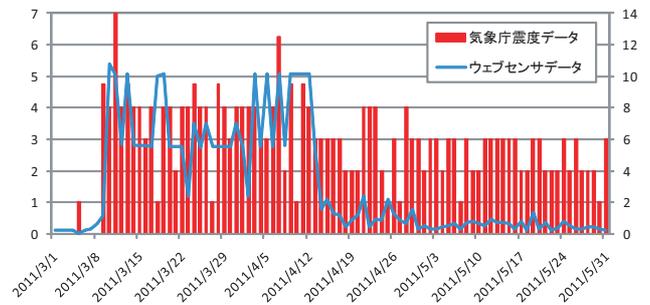


図 9 「宮城県」の $\text{WebSensor}_{\text{sffw}}^2(t, s)$ データと気象庁震度データ
Fig. 9 $\text{WebSensor}_{\text{sffw}}^2(t, s)$ and Seismic Intensity in Miyagi.

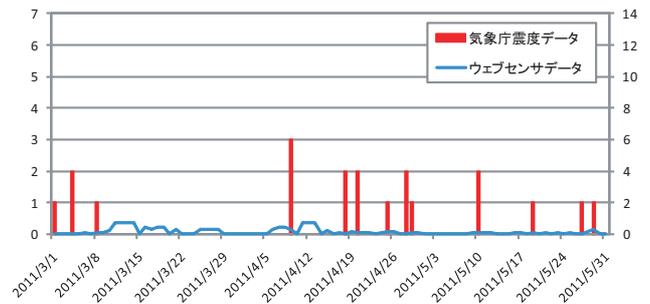


図 10 「宮崎県」の $\text{WebSensor}_{\text{sffw}}^2(t, s)$ データと気象庁震度データ
Fig. 10 $\text{WebSensor}_{\text{sffw}}^2(t, s)$ and Seismic Intensity in Miyazaki.

最後に、図 11 と図 12 は、 $\text{WebSensor}_{\text{sffw}}^2(t, s)$ ($\alpha = 0.84, \beta = 0.13$) と気象庁の震度情報との相関を日毎に 47 都道府県データに基づいて求めることで、ウェブセンサの時間依存性を示している。それぞれのグラフでは全国地震回数と最大震度との関係も見ているが、特に顕著な傾向は確認できない。かなり激しく変動しており、ウェブセンサの時間依存性の高さを示しており、今後優先的に取り組むべき課題であると考えられる。

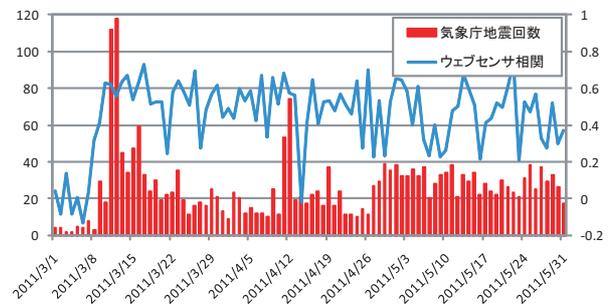


図 11 $\text{WebSensor}_{\text{sffw}}^2(t, s)$ の相関の時間依存性と気象庁全国地震回数
Fig. 11 Temporal Dependency of $\text{WebSensor}_{\text{sffw}}^2(t, s)$ with JMA Number of Earthquakes.

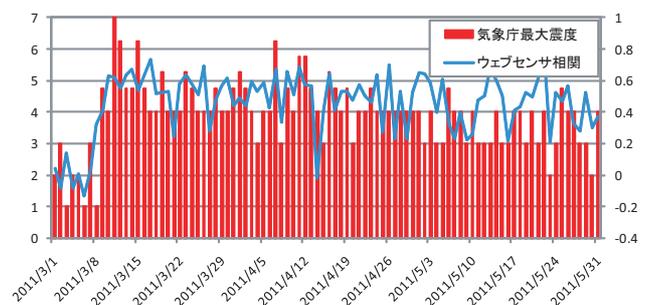


図 12 $\text{WebSensor}_{\text{sffw}}^2(t, s)$ の相関の時間依存性と気象庁全国最大震度
Fig. 12 Temporal Dependency of $\text{WebSensor}_{\text{sffw}}^2(t, s)$ with JMA Max Seismic Intensity.

4. まとめと今後の課題

実世界で起きた様々な現象や事象に関する知識データをブログなどのCGMからマイニングする研究が盛んに行われ、一般向けへのサービス化も試行され始めているが、大量のウェブ文書からマイニングされたデータが、実世界でのリアルなデータをどの程度正確に反映しているかの詳細な調査は少なく、精度が十分に保証されていないまま盲目的に利用するのは問題があると考えられる。そこで本論文では、実世界で起きた現象のリアルなデータとして気象庁の地震統計から震度情報を用い、ウェブから抽出した時空間依存データとの相関を評価することで、ウェブセンサによって抽出されたデータの利用可能性や信頼性を検証した。ウェブ検索エンジンのインデックス情報の一つである検索件数だけを用い、特定の時空間で制限したブログ文書の中で「地震」や「揺れ」といった対象の現象を連想させる記述を含む頻度に基づいてウェブセンサを構成し、実世界の地震に関する時空間依存データをウェブから抽出した。その上で、実世界の震度に関してウェブ世界から抽出した時空間依存データと気象庁の地震統計との相関を多角的に評価・考察を行った。

今後の研究課題としては、気象統計や地震統計以外の実世界のリアルなデータを収集し、自然現象以外のより多くの種類の実世界の現象についてもウェブセンサとの相関を評価する必要がある。また、本論文では、ブログ文書頻度および単純な方法で正規化した指標を用いてウェブセンサを構成したが、関数を工夫するなど抽出手法を高度化することで、より強い相関が得られる可能性がある。センサ対象の現象に依って時間的に前後にシフトさせたり、周りの時間にも影響を伝播させることによってロバスト性の向上を図るウェブセンサ技術[16]は既に考察しているが、時間的だけでなく空間的な伝播についても組み込み、一般ユーザがより安心して参考にでき、より多くの応用システムで活用可能にするため抽出手法の改善を図って行く。さらには、ウェブニュース記事[19]やマイクロブログの一種であるTwitter[20]のつぶやき文書など、他のウェブメディアからの時空間依存データの抽出実験も行い、ウェブセンサのマイニング対象メディアへの依存性についても詳細な調査を行う。Twitterのつぶやきを用いれば、月毎や日毎といった粒度よりも細かく、よりリアルタイムに数値データを出力するウェブセンサを構成できると考えている。

最後に、実世界に物理的に設置する必要があるリアルセンサとは異なり、あらゆる場所の時空間依存データを仮想的に抽出できる可能性があるウェブセンサを活用して、実世界に設置されたリアルセンサの代替あるいは補完センサと位置付け、ウェブ抽出された時空間依存データ（潜在的なニーズなど）をリアルタイムに実世界にフィードバックし、実空間の構造や実世界でのサービスの配置などを自動的に変えて行く新しい仕組みの研究にも取り組んで行きたいと考えている。

謝辞 本研究は科学研究費助成事業（学術研究助成基金助成金）若手研究（B）（研究代表者：服部峻，課題番号：23700129）「ウェブから時空間依存データを抽出するウェブセンサに関する研究」の助成を受けたものです。ここに記して謝意を表します。

文 献

- [1] 鈴木泰裕, 高村大也, 奥村学: “WebLog を対象とした評価表現抽出,” 第 6 回セマンティックウェブとオントロジー研究会, SIG-SWO-A401-02 (2004).
- [2] 藤村滋, 豊田正史, 喜連川優: “文の構造を考慮した評判抽出手法,” 電子情報通信学会 第 16 回データ工学ワークショップ (DEWS2005), 6C-i8 (2005).
- [3] Tezuka, T., Kurashima, T., and Tanaka, K.: “Toward Tighter Integration of Web Search with a Geographic Information System,” Proceedings of the 15th International World Wide Web Conference (WWW'06), pp.277–286 (2006).
- [4] 倉島健, 藤村考, 奥田英範: “大規模テキストからの経験マイニング,” 電子情報通信学会 第 20 回データ工学ワークショップ (DEWS2008), A1-4 (2008).
- [5] Inui, K., Abe, S., Morita, H., Eguchi, M., Sumida, A., Sao, C., Hara, K., Murakami, K., and Matsuyoshi, S.: “Experience Mining: Building a Large-Scale Database of Personal Experiences and Opinions from Web Documents,” Proceedings of the 7th IEEE/WIC/ACM International Conference on Web Intelligence (WI'08), pp.314–321 (2008).
- [6] 服部峻, 田中克己: “性質継承と概念の再帰的適用に基づく Web からの概念階層抽出,” 情報処理学会論文誌 (トランザクション) データベース, Vol.1, No.3 (TOD40), pp.60–81 (2008).
- [7] Hattori, S., and Tanaka, K.: “Extracting Concept Hierarchy Knowledge from the Web based on Property Inheritance and Aggregation,” Proc. of the 7th IEEE/WIC/ACM Int'l Conference on Web Intelligence (WI'08), pp.432–437 (2008).
- [8] 服部峻, 手塚太郎, 田中克己: “文書中の地物画像を言語的記述で代替するための地物の外観情報の Web からの抽出,” 情報処理学会論文誌 (トランザクション) データベース, Vol.48, No.SIG11 (TOD34), pp.69–82 (2007).
- [9] Hattori, S., Tezuka, T., and Tanaka, K.: “Mining the Web for Appearance Description,” Proceedings of the 18th International Conference on Database and Expert Systems Applications (DEXA'07), LNCS Vol.4653, pp.790–800 (2007).
- [10] 服部峻, 田中克己: “Web 抽出した特異な色名と色特徴量変換に基づく特異画像の Web 検索,” 情報処理学会論文誌 (トランザクション) データベース, Vol.3, No.1 (TOD45), pp.49–63 (2010).
- [11] Hattori, S.: “Context-aware Search Control for Secure Spaces,” Proceedings of the Joint 5th International Conference on Soft Computing and Intelligent Systems and 11th International Symposium on advanced Intelligent Systems (SCIS&ISIS'10), SA-D4-2, pp.1348–1353 (2010).
- [12] 情報検索に対する信頼性に関する調査および結果, <http://www.dl.kuis.kyoto-u.ac.jp/i-explosion/report/> (2011).
- [13] 気象庁 - 震度データベース検索, http://www.seisvol.kishou.go.jp/eq/shindo_db/shindo_index.html (2011).
- [14] Hattori, S., and Tanaka, K.: “Mining the Web for Access Decision-Making in Secure Spaces,” Proceedings of SCIS&ISIS'08, TH-G3-4, pp.370–375 (2008).
- [15] 服部峻: “気象統計との相関に見る Web センサの可能性,” 第 151 回 情報処理学会 データベースシステム研究会, 情報処理学会 研究報告「データベースシステム」, Vol.2010-DBS-151, No.4, pp.1–8 (2010).
- [16] Hattori, S.: “Secure Spaces and Spatio-Temporal Weblog Sensors with Temporal Shift and Propagation,” Proc. of the First IRAST Int'l Conference on Data Engineering and Internet Technology (DEIT'11), pp.1042–1047 (2011).
- [17] Google ブログ検索, <http://blogsearch.google.co.jp/> (2011).
- [18] 気象庁 - 計測震度の算出方法, <http://www.seisvol.kishou.go.jp/eq/kyoshin/kaisetsu/calc.sindo.htm> (2011).
- [19] asahi.com : 朝日新聞社の速報ニュースサイト, <http://www.asahi.com/> (2011).
- [20] Twitter, <http://twitter.com/> (2011).