

自然災害などの緊急時における情報の信憑性計測システムの実現方式

本多 賢[†] 吉田 尚史[‡]

[†] 駒澤大学大学院グローバル・メディア研究科 〒154-0012 東京都世田谷区駒沢 1-23-1

[‡] 駒澤大学グローバル・メディア・スタディーズ学部 〒154-0012 東京都世田谷区駒沢 1-23-1

E-mail: [†] 3715101k@komazawa-u.ac.jp, [‡] naofumi@komazawa-u.ac.jp

あらまし 本論文では、自然災害などの緊急時において情報の信憑性を計測するシステムの実現方式を示す。本システムでは、情報の信憑性を計測する。任意の情報を WWW 上にある複数の情報源及びセンサーデータと比較する事で情報の信憑性を客観的に計測する。情報の信憑性が高いかどうかはその情報の客観性によって決まる。そのため本システムでは、1つの情報を複数の情報源及びセンサーデータを用いる事で客観性の程度を導き出し、信憑性を計測する。本論文では実験によって本システムの実現可能性を示す。

キーワード 信憑性, 客観性, web, センサーデータ

1. はじめに

東日本大震災では多くの情報機能が麻痺し、混乱した。その後も地震だけでなく土砂災害や火山の噴火等の災害が続いている。首都直下型の地震や富士山の噴火なども近い将来起きる可能性があると言われており、こういった状況に対応するためのシステム・ツールは増々重要になってくる。ICT・メディアのあるべき姿はいつ何時でも情報が提供される事である。例え被災地でも情報に格差があっても ICT・メディアの本来の力が発揮できているとは言えない。自然災害は一つの例であり、このような状況で役に立つシステムがあれば緊急時に限らず平常時の情報収集においても役に立つ。

不確かな情報によって実害が生じる場合は多々ある。例えば、2011年3月11日の東日本大震災で実際に起きた例として、支援物資に関するデマ情報の拡散がある。「岩手県が個人からの救援物資を必要としている」という情報が拡散されたがチェーンメールなどを通じてインターネット上で拡散されたが、このような事実はなく、のちに岩手県広報課によって否定されている[18]。正確な情報の不足によるデマの拡散は、支援の妨げや支援物資の無駄だけでなく、被災者の不安を煽ることに繋がりかねない。

このような背景から本稿では信憑性の計測を行うためのシステムの実現方式を提案する。情報の信憑性が高いかどうかは、その情報の客観性によって決まる。そのため本方式では、1つの情報を複数のメディア、及びセンサーデータを用いて比較する事で、情報の客観性の程度を導き出し、信憑性を計測する(図1)。

情報の信憑性に関する研究は数多くある。[12]の研究では、分析対象となる情報と、その関連情報間の類似性・共起関係(support)を分析し、support 関係にあるデータの対が多いものほど信憑性が高いとする、メデ

ィア間の類似性及びそれらの共起関係の計量により信憑性を判断する。[9]の研究では、twitter 上の情報から質疑応答形式で欲しい情報が手に入るようになっており、不適切な情報が抽出されないように、文中の表現やツイートの属性を分析するという、同種の SNS 内の情報間の比較という方法である。[7]の研究はマイクロブログにおけるデマ拡散防止のために、デマ情報とデマ訂正のリツイートの傾向と発信者の分析を、6つの仮説と共に行っている。

本稿では、web 上の信頼性の高い情報源やセンサーデータを利用してあらゆる情報の信憑性を計測し、特に緊急時において、情報収集を行う際の取捨選択の手助けとなるようなシステムの実現方式を提案する。ここでは緊急時として自然災害を想定している。

本方式の新規性は、入力された任意の情報を、異なる信頼性の高い情報と比較する事で計測し、入力された任意の情報と、複数のセンサーデータというリアルで客観的な情報を比較して信憑性を計測するという二点である。信頼性の高い情報が複数ある事、それに加えてセンサーデータを用いる事で客観的な情報とマッチするかを計測し、情報の客観性を導き出し、それを信憑性としている。

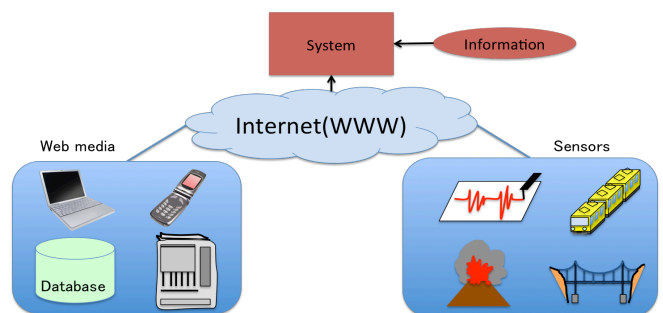


図 1: システム概要

2. 信憑性の計測方法の概要

本方式では信憑性の計測方法として、web 上にある複数の信頼性の高い情報源と、全国各地に設置されているセンサーからオープンデータとして公開されているものを利用して、「信憑性を測りたい任意の情報」と比較する事で、信憑性の計測を行う。図 2.1 にその概要を示す。

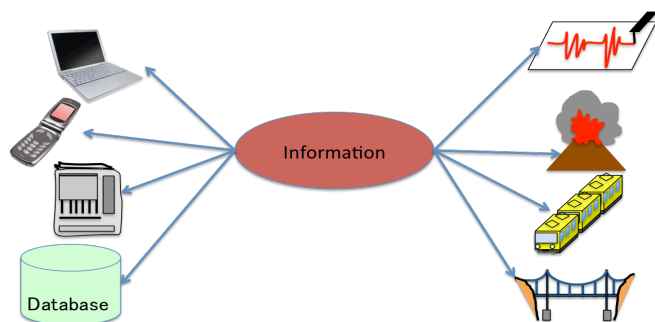


図 2.1: 任意の情報と web・センサーデータとの比較

具体的に信頼性の高い情報源とは、政府関係(内閣・各省庁・自衛隊・警察・消防など)、報道関係(全国紙・地方紙・通信社・企業など)、ボランティア団体(NPO・NGO・支援団体など)などの web サイト・情報発信ツールの事である。そして主なセンサーデータは、地震や火山活動を観測するセンサー、風雨の観測レーダー、災害観測用の人工衛星のような自然観測センサーや、橋の強度を監視する橋梁センサー、あらゆる緊急時に対応する鉄道などを利用する。

本方式における、入力された情報の信憑性の計測方法を図 2.2 に示す。これらの情報源(Resource)からリアルタイムデータを収集し、本システムのデータベースに格納する。そして「信憑性を測りたい任意の情報」をテキストとしてシステムに入力する。システム内部で任意の情報と Resource を比較し、どれだけマッチングしたかを計算し、最終的に 0% から 100% の数値を出力する。

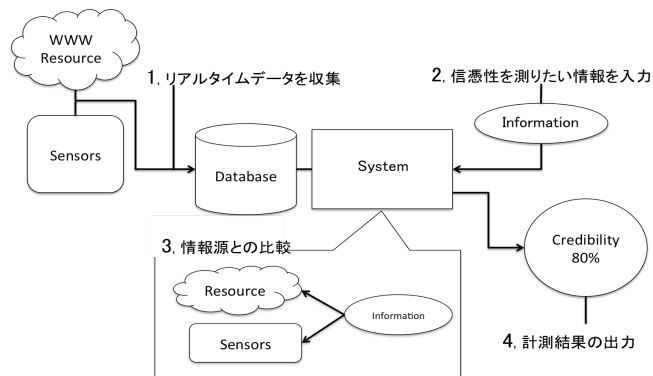


図 2.2: 計測方法

これにより、入力された情報がどれだけ客観的かがわかり、最終的な信憑性の度合いとなる。

3. 信憑性計測システム

本節では前節で示した提案方式を実現するために、システム内における Resource と任意の情報のマッチング、及び信憑性の計測に関する具体的な実現方法を説明する。

3.1 web 情報とのマッチング

はじめに、複数の Resource と任意の情報のマッチングの方法から説明する。現在の実装では Resource となる web サイトからリアルタイムな情報を HTML parser を用いて定期的に収集し、1 センテンス毎にデータベースに格納する。例えば「東京都で地震が発生」のような文章である。この文章を分解(split)する(図 3.1)。

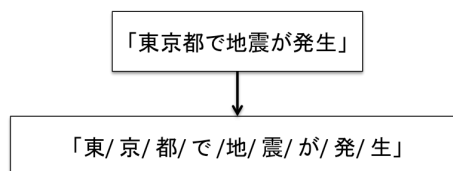


図 3.1: 文字の分解

次に信憑性を知りたい任意の情報を入力する。この情報も分解する。分解された Resource 情報と任意の情報をそれぞれ総当たりでマッチングし、値を取り出す。この操作をデータベース内のセンテンスの数だけ繰り返す(図 3.2)。

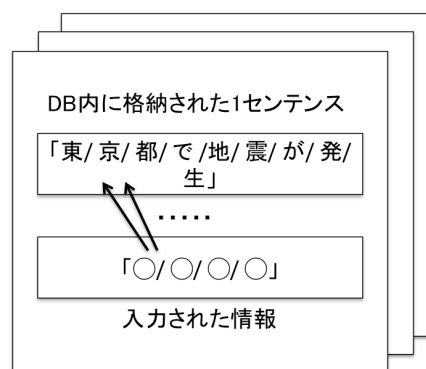


図 3.2: 文字のマッチング(現在の実装)

ここまでで得られたマッチングの値を、任意の情報とデータベースの Resource のセンテンスの文字数を掛けた数字で割った値が信憑性の度合いとなる。

具体的には、マッチングの値を n 、任意の情報の長さを $length_a$ 、データベースの Resource のセンテンスの長さを $length_b$ とすると、図 3.3 に示す式になる。

$$\text{credibility} = n \div (\text{length_a} \times \text{length_b})$$

図 3.3: 信憑性の計算

3.2 センサーデータとのマッチング

本節ではセンサーデータを用いた信憑性の計測方法について述べる。用いるセンサーによって手法は変わるが、ここでは例として防災科学技術研究所(NIED)が公開している強震モニターを利用した。強震モニターは、地震の発生場所を日本地図上に色で表示する。この画像をデータベースに取り込み、色を解析する。これにより、地震の規模と発生した場所が分かる。ここで表示される色は地震そのものの規模を表すマグニチュードではなく、地表がどれだけ揺れたかを表す「表面最大加速度」である。赤であれば表面最大加速度は 100gal 以上であり、震度でいうと、震度 5 以上である。これは 2 秒に一度更新される。このデータを用いる事で、発生した地震の規模や場所、時間などが分かる。単純に災害が発生したかどうかの信憑性に関しては、このセンサーを用いる事で計測する事ができる。具体的には、取得した画像内に現れている色を判定し、赤(震度 5 以上)があれば大きな地震が起きた事になるというものである。微弱な揺れの地震に関しては、かなり頻繁に起きており、さらに、人間の生活で生じる揺れである生活振動は日常的に発生しているので、これを含めてしまうと信憑性が計測できないため、データベースに取り込むものは地図上に赤が含まれるもののみとする。これにより震度 5 以上の地震が発生したかどうかを 0 か 1 で明確に判断する事が可能となる。センサーデータとのマッチングから得られた信憑性の値と web 情報とのマッチングで得られた信憑性の値を統合する事で、より客観的に信憑性の値を導き出す事を可能とする。

4.実験

実現方式の実現可能性を示すための実験について述べる。具体的には、構築したシステムを用いて、信憑性を計測するために有用なデータを導き出せるかどうかを検証する目的で実験を行う。

4.1 実験環境

web 情報を用いるシステムでは、実験用のデータセットとして、気象庁の震度情報、日本経済新聞の速報ページの見出し、Yahoo!ニュースの速報ページの見出しからリアルタイム情報を抽出し、データベースに格納しておく。入力する信憑性を計測する情報は 1:「東京都で震度 5 の地震」、2:「窃盗団が出没」、3:「東京で火災」の 3 つを用いる。1 は地震が起きたかどうか

についてのみの情報であり、データベース内に一致する内容が複数存在するものである。2 はデータベースには一つも格納されていない情報、3 は実際にデータベース内に一致度の高いものが一つだけ含まれる情報である。以上の 3 つの情報をシステムに入力し実験を行う。

4.2 マッチング実験

この実験は 1:「東京都で震度 5 の地震」を入力として行う実験である。実験によって得られた数字を下の表 4.1 にまとめる。db_text はデータベース内に格納されている Resource からの情報である。ratio はマッチングの割合、sensor はセンサーデータの値で地震が観測された場合を 1、観測されなかった場合を 0 で表したものである。表 4.1 ではセンサーは地震を観測せず 0 を示している。

表 4.1 センサーデータが 0 の場合の実験結果

db_text	ratio	sensor
平成27年12月24日17時13分24日17時10分頃宮城県沖M3. 6震度1	0.00093	0
平成27年12月22日21時38分22日21時35分頃長野県南部M2. 6震度1	0.00091	0
平成27年12月22日20時16分22日20時12分頃岩手県沿岸北部M3. 6震度3	0.00087	0
平成27年12月22日07時44分22日07時40分頃茨城県北部M3. 6震度2	0.00091	0
平成27年12月18日01時10分18日01時06分頃北海道南西沖M3. 3震度1	0.00089	0
平成27年12月17日18時15分17日18時11分頃宮崎県南部山沿いM2. 5震度2	0.00085	0

データベースに格納された Resource 情報の文字の長さが異なるために微妙な誤差が生じるものの「震度」という語句が必ずマッチングするため、どの情報もほとんど同じ割合であり、また、センサーデータは地震を観測していないため、地震が起きたかどうかの指標にはならなかった。次の表 4.2 ではセンサーが地震発生を観測し、1 を示した場合の表である。

表 4.2 センサーデータが 1 の場合の実験結果

db_text	ratio	sensor
平成27年12月24日17時13分24日17時10分頃宮城県沖M3. 6震度1	0.00093	1
平成27年12月22日21時38分22日21時35分頃長野県南部M2. 6震度1	0.00091	1
平成27年12月22日20時16分22日20時12分頃岩手県沿岸北部M3. 6震度3	0.00087	1
平成27年12月22日07時44分22日07時40分頃茨城県北部M3. 6震度2	0.00091	1
平成27年12月18日01時10分18日01時06分頃北海道南西沖M3. 3震度1	0.00089	1
平成27年12月17日18時15分17日18時11分頃宮崎県南部山沿いM2. 5震度2	0.00085	1

こちらの表ではセンサーが地震発生を裏付けているので、どの情報も信憑性が増す事になる。このようにセンサーからの情報を統合する事でweb情報だけで判別しづらかった情報の信憑性を判別しやすくできる。webとセンサーという異なる情報源を用いる事で情報の客観性を高める事となり、より正確な信憑性の計測を可能とする事が分かった。

次に 2:「窃盗団が出没」を入力とした実験を行い、結果を表 4.3 に示す。この入力ではデータベース内にマッチするものがほとんどないため、マッチングの割合は 0 もしくは非常に低いものであった。

表 4.3 データベースにない情報のマッチング

db_text	ratio	sensor
ウィーン当局「新年にかけ欧州の首都で襲撃の恐れ」	0	0
首相、3カ月ぶりにゴルフ	0	0
佐賀工や伏見工など勝つ 全国高校ラグビーが開幕	0.0008	0
東京・板橋でマンション火災 男性1人死亡産経新聞 23時28分	0	0
<ローマ法王>死刑廃止を呼びかけ…「人命最優先」鮮明毎日新聞 23時0分	0	0
函館本線でトンネル火災 明日28日も一部見合わせ続くレスキューナウ ニュース 22時45分	0	0

「が」という助詞が一文字だけマッチングした場合のみ 0 では無くなっているだけである。

最後に 3:「東京で火災」を入力とした実験を行い、結果を表 4.4 に示す。これはデータベース内に含まれる情報に近いものを入力として用いている。そのため、マッチングしたものに関しては割合が上がった。

表 4.4 データベースに一致するものがある場合の割合を示す

実験結果

db_text	ratio	sensor
ウィーン当局「新年にかけ欧州の首都で襲撃の恐れ」	0.00138	0
首相、3カ月ぶりにゴルフ	0	0
佐賀工や伏見工など勝つ 全国高校ラグビーが開幕	0	0
東京・板橋でマンション火災 男性1人死亡産経新聞 23時28分	0.00476	0
<ローマ法王>死刑廃止を呼びかけ…「人命最優先」鮮明毎日新聞 23時0分	0.00033	0
函館本線でトンネル火災 明日28日も一部見合わせ続くレスキューナウ ニュース 22時45分	0.00142	0

こちら「で」という助詞が一致する事で微妙に割合が上昇したものがあるが、入力した情報がほとんど、もしくは完全に一致した場合、割合は明らかに他と違うものとなった。今回使用したセンサーは地震発生を観測するものなので 2:「窃盗団が出没」と 3:「東京で火災」の入力に関してはセンサーは 0 を示している。

実験を重ねた結果マッチングの比率が高い、もしくは

は完全に一致するものは、割合が 0.003 を超える事が分かった。そのためマッチングの割合の閾値は 0.003 とする事ができる。

4.3 実験結果

これらの実験の結果から、本方式により、客観性を導き出し、信憑性を計測できる事が分かった。データベース内に同じような内容で一部の語句が違うだけの情報が複数存在した場合には、そのどれもが高い割合を出すため、センサーを含めた異なる情報源から情報を抽出してくる事が必須となり、客観的なセンサーの情報を統合する事の重要性を示す事ができた。

また、データベース内に「東京で震度 5 の地震は誤報」という過去の情報を訂正する情報が含まれる場合に 1:「東京で震度 5 の地震」を入力すると割合が上がる事が予想されるが、センサーの情報を統合する事でこの問題にも対応出来るという事が分かる。

本実験により、システム内での情報のマッチングが正確にできる事、そして多様性のある情報を集約する事の重要性が分かった。また情報の客観性を導き出す上でのセンサーデータの有用性を示す事ができた。

5. まとめと今後の課題

本稿では、情報の信憑性を計測する方法と、そのシステムの実現方式を示し、さらに実現可能性を実験により示した。人間が正常な判断を下しにくい状況下において、間違った情報に惑わされるというのは深刻な問題であり、収集した情報の信憑性を計測するというのは極めて重要な課題である。特に、誰もがインターネットに接続できる現代ではなおさらである。しかし、情報の信憑性を正確に判断する事は困難である。そのため本方式の目的は情報の客観性を示し、それを信憑性として、ユーザーが情報を取捨選択する際の手助けをする事である。このような背景から本研究は客観的に情報の質を判断し、収集した情報の信憑性を計測するための指標となるシステムの実現を目指すものである。

そのため、一つの情報ではなく複数の情報源を用いる方法と、人の手の入らない極めて客観的な情報を得るためにセンサーデータを利用する方法を提案した。

今後の課題として、テキストのマッチングの方法に形態素解析や機械学習などの技術を取り入れるなど、データベース上の信頼性の高いテキスト情報と入力されるテキスト情報をマッチングさせる方法の検討、センサーデータとより確実なマッチングができるよう、使用するセンサー毎にセンサーデータを獲得する方法を設計及び実現すべきである。さらに、web やセンサーの情報に時空間による重み付けを行う事によって、時間や場所によって重視する情報を変えるなどの工夫

を実現する方法やセンサーデータのイベント発生時刻と入力情報の時刻をマッチングさせる方法の検討も今後の課題となる。

謝 辞

本稿についての有益なコメントを頂いた駒澤大学グローバル・メディア・スタディーズ学部石橋直樹先生に深く感謝いたします。また本稿で利用しているセンサーの強震モニターのデータを提供していただいている防災科学技術研究所に感謝申し上げます。

参 考 文 献

- [1] Andrew.J.Flanagin, Miliam.J.Metzger (2008) *The Credibility of volunteered geographic information* 'GeoJournal' 72:137-148 Published online
- [2] M. J. Salganik, P. S. Dodds, and D. J. Watts (2006) *Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market* 'Science' vol. 311, pp. 854-856.
- [3] Shawn Tseng, B.J.Fogg (1999) *Credibility and Computing Technology* 'Communications of ACM' vol.42, No.5 ACM
- [4] Soo.Young, Brian.Hilligoss (2008) College Student's Credibility Judgements in the Information-Seeking Process 'MA' 49-72 The MIT press
- [5] Talal.H.Noor, Quan.Z.Sheng, Sherali.Zeadally, Jia.Yu (2013) *Trust Management of Services in Cloud Environments: Obstacles and Solutions* 'ACM Comput' Surv.46, 1, Article12 ACM
- [6] 打野拓 (2013) 「Twitter の非公式リツイートにおける信頼性評価と分析」
- [7] 梅島彩奈、宮部真衣、荒牧英治、灘本明代 (2011) 「災害時 Twitter におけるデマとデマ訂正 RT の傾向」 『IPSJ SIG Technology Report』 vol.2011-DBS-152, No.4, vol.2011-IFAT-103, No.4 情報処理学会
- [8] 長谷川友香、小口正人 (2013) 「緊急時判断に基づく状況に応じた個人情報へのアクセス制御」 DEIM Forum
- [9] 水野淳太、後藤淳、大竹清敬、川田拓也、島澤健太郎、クロエツェージュリアン、田中正弘、橋本力、奥村明俊(2015) 「対災害情報分析システム DISAANA 及びその質疑応答モードの性能評価」 情報処理学会研究報告 vol.2015-MBL-76No.14 vol.2015-CDS-14NO.14
- [10] 山岸佑己、斉藤和巳、武藤伸明 (2015) 「時間減衰を考慮した統計モデルに基づく動画のランキング手法」 DEIM Forum
- [11] 山本祐輔 (2011) 「ウェブ情報の信憑性分析に関する研究」 京都大学博士(情報学)
- [12] 山本祐輔、田中克己 (2010) 「データ対間のサポート関係分析に基づく Web 情報の信憑性評価」 『データベース』 vol.3, no.2, pp61-79 情報処理学会
- [13] 山本祐輔、田中克己 (2010) 「ウェブ検索結果の信憑性判断基準」 WebDB Forum 2010
- [14] ジェームズ・スロウィツキー (2009) 「みんなの意見は案外正しい」 株式会社角川書店
- [15] 関谷直也 (2011) 「風評被害 そのメカニズムを考える」 株式会社光文社
- [16] ドナルド・デイヴィッドソン (2007) 「主観的、間主観的、客観的」 株式会社春秋社
- [17] 岩手県広聴広報課 HP, <http://www.pref.iwate.jp/index.html>
- [18] 岩手県広聴広報課 公式 Twitter https://twitter.com/pref_iwate/status/46891690585370625
- [19] JAXA 宇宙航空研究開発機構 HP, <http://www.jaxa.jp>
- [20] NIED 防災科学技術研究所 <http://www.bosai.go.jp>
- [21] 国土交通省公式 HP, <http://www.mlit.go.jp>
- [22] 内閣府公式 HP, <http://www.cao.go.jp>
- [23] Mami Morimoto (2013) 東日本大震災とソーシャルメディア <http://www.res.kutc.kansai-u.ac.jp/~ckita/Links/Shin-sai-Media2013/index.html> 2015,8,4 閲覧