

Web 文書からの人の安全・危険に関わる情報の抽出

岡崎 直観^{†‡}

成澤 克麻[§]

乾 健太郎[†]

{okazaki, katsuma, inui} at ecei.tohoku.ac.jp

[†] 東北大学大学院情報科学研究科

[‡] 科学技術振興機構 さきがけ

[§] 東北大学工学部知能システム総合学科

1 はじめに

インターネットやモバイル通信などの情報通信技術により、一般の人がいつでも、どこでも、誰とでも繋がりが、情報発信・伝達を行えるようになった。平成 23 年版の情報通信白書 [19] によると、日本のインターネット利用者は平成 22 年末の時点で 9,462 万人（人口普及率 78.2 %）に達し、情報通信技術が生活の中に浸透したことが伺える。また、mixi, Twitter, Facebook などの新しいソーシャルメディアにより、リアルタイムな情報の流通が促進された。

2011 年 3 月に発生した東日本大震災では、安否確認や被災者支援のために、ソーシャルメディアが大活躍した [20]。東日本大震災後の Twitter の利用動向、交換された情報の内容、情報の伝搬・拡散状況などの分析・研究も進められている [1, 3, 12, 18]。一方で、3 月 11 日の「コスモ石油のコンビナート火災に伴う有害物質の雨」に代表されるように、インターネットやソーシャルメディアがいわゆるデマ情報の流通を加速させたという指摘もある。東日本大震災に関するデマをまとめたツイート¹では、2012 年 1 月時点でも月に十数件のペースでデマ情報が掲載されており、情報の信頼性の担保や、メディア・リテラシーの養成が急務である。

東日本大震災とそれに関連する福島第一原子力発電所の事故では、多くの国民の生命が脅かされる事態となった。以前より、健康に関する身近な問題を解決するためにインターネットが活用されており [14]、今回の震災でも多くの国民が健康危機管理に関する情報をネット上から収集したものと推定される。震災に関するデマとして、人間の安全・危険に関するもの（例えば「放射性物質から甲状腺を守るにはイソジンを飲む」）が大勢を占めた。これらのことから、健康危機管理に関する情報源としてのネットの重要性が垣間見える。

2 本研究のねらい

2.1 安全・危険に関する情報の構造化

こうした背景のもと、我々は人間の安全・危険に関する行動決定を支援するため、インターネット上に鑿められている情報を獲得・集約し、情報の背後にある論述構造を解析する研究に取り組んでいる。「知識の自動獲得・構造化に基づく情報の論理構造とリスクの分析」と名付けた本課題では、利用者が達成したい目的（例えば「放射能の影響を減らしたい」「インフルエンザを予防したい」）や、取ろうと思案している行動（例えば「イソジンを飲む」「インフルエンザワクチンを接種する」）に対して、リスクの見積もり（目的の達成度、安全/危険度）と、各リスクの見積もりに関連する理由、根拠、条件などを情報を抽出する。

図 1 に「放射線の影響を減らすためにイソジンを摂取する」という目的・行動に対する分析例を示す。ウェブに対して情報検索を行い、「イソジン摂取」と「放射能」に関連する文章を取得し、各文章から目的達成や人の安全・危険に関する箇所を抽出する。図 1 上段の文では、例えば「止めてください」という表現が危険を直接的に示唆しており、その理由と条件が「と」や「ので」などの談話マーカーで示されているので、危険な事態とその根拠を抽出できる。同様の分析を複数の文章に対して適用することにより、同じ行動・根拠の文章を集約したり、安全/危険の極性が食い違っている根拠対を提示したり、行動・根拠・リスクの支持率をウェブ上において定量的に評価することができる。

現実には、図 1 の上段のような理想的な構成の文は少ない。例えば「止めてください」が欠け「嘔吐が引き起こされます」という表現だけで危険な事態を示唆する場合、理由や条件が複数の文で構成されている場合など、恣意的な結論を導くため論理に飛躍がある場合などが問題となる。これまでの自然言語処理では、テキ

¹https://twitter.com/#!/jishin_dema

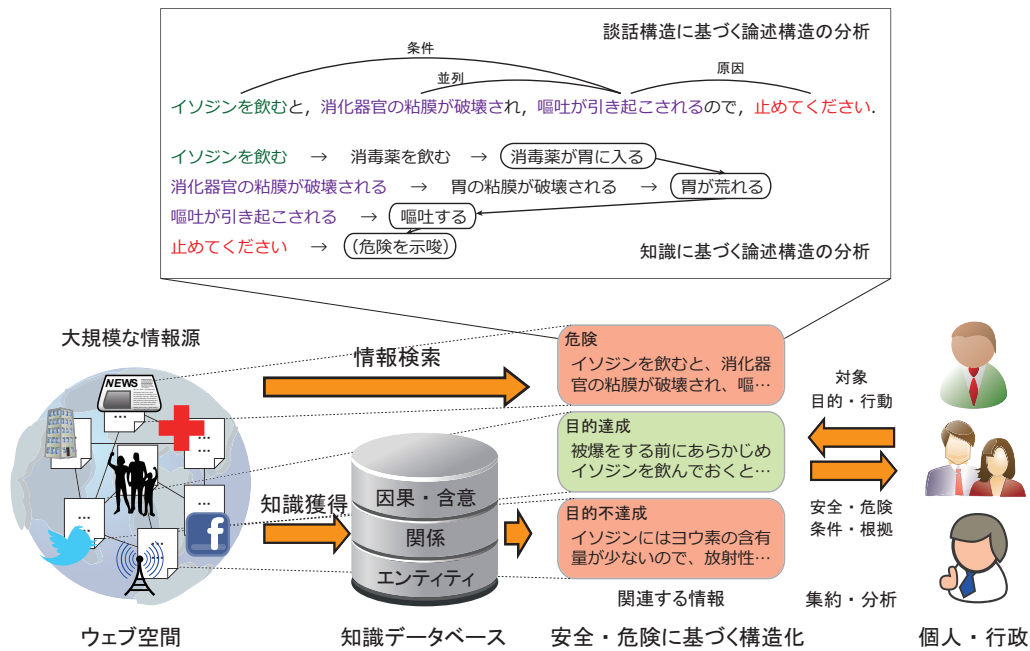


図 1: 安全・危険に関する情報の構造化

ストに明示的に書かれていない談話構造を解析することは難しかった。本課題では、大規模な文書集合から語彙知識、関係知識、関係間（事態間）知識を自動的に獲得する研究を進め、これらの利活用を行うことで、論述構造を頑健に解析する手法の確立を目指す。図 1 の「知識に基づく論述構造の分析」では、「胃があれば嘔吐する」「嘔吐するのは危険な事態」などの知識を利用することで、論述構造をさらに頑健に推定し、危険な事態に至る根拠としての筋の良さを評価している。

本研究で構造化した安全・危険に関する情報は、利用者の行動決定に役立つだけでなく、行政や個人が情報の信憑性を検証するための手がかりとしても利用できる。健康危機管理の専門家ではない個人でも、安全・危険に関する行動決定の背後にある根拠を俯瞰することで、他人の助言を鵜呑みにするだけでなく、各人の状況に合わせて情報を分析する能力が養われる。そして、流通している情報に対して、多くの人が冷静な判断・行動が取れるようになれば、ネット上のみならず社会全体の健全性を維持できると期待している。

2.2 関連研究

本研究は、情報の信頼性、評判分析、臨床判断支援などと関連が深い。ただし、情報の信頼性はユーザが信頼すると思われる情報を選ぶタスク、評判分析は明示的に評価を述べた箇所を抽出するものであり、安全・

危険の根拠まで分析する本研究の方が、より広いゴールを設定している。臨床判断支援は電子カルテや文献を分析し、医者や研究者などの専門家の意思決定を支援するものであるが [2, 8]、紙面の都合上、情報の信頼性と評判分析について、最新の研究動向を紹介する。

情報の信頼性 (trust) や事実調査 (fact finding) に関する研究は、ウェブやソーシャルメディアの分野で盛んに行われている [4, 9]。また、ツイッター上のデマを検出する試みも始められている²。梅島ら [16] は、東日本大震災後に多くリツイートされたツイートを分析し、デマ情報に特有な現象を検証した。Gupta ら [5] は、各ツイートのメタ情報の素性 (ツイートしたユーザーのフォロワー数やプロフィールの有無など) から学習したツイートの信頼度を、ツイート間の類似度などで構成されたネットワーク上で拡散させ、情報の信頼性の推定精度を改善できたと報告している。Pasternack ら [11] は、一階述語論理で記述された事前知識を、事実調査アルゴリズムに統合する枠組みを提案した。

安全・危険やユーザの目的達成に関する言及を、ポジティブ/ネガティブな表現と考えれば、本研究の一部は評判 (感情) 分析として扱うことができる。最近では、評判分析の対象ドメインとして、ウェブやソーシャルメディアに的を絞った研究も多く報告されている [6, 10, 13]。O'Connor ら [10] は、極性辞書を用いた単純な感情分析

²例えば「でまったー」のようなアプリケーションも公開されている。http://taretw4u.appspot.com/Dema.do

表 1: ラベルの定義とアノテートされた表現例

ラベル	基準	例	箇所数
SAFE	安全を直接的に示唆する箇所	「全く問題にする必要はありません」「普段通りの生活をしていて差し支えありません」「健康上何ら問題ないレベルが続いています」	276
SAFE_REASON	安全を示唆する理由, または安全を間接的に示唆する箇所	「今のレベルは自然放射線のレベルと比べてあまり変わらないレベルです」「市場に出回る前に暫定基準値以下であることがチェックされて」	184
SAFE_COND	安全が成立する条件	「現時点の放射線による線量や診断で受ける線量では」「現在の時点でこれ以上の大きな変化がなければ」「1年以内に収まれば」	82
DANG	危険を直接的に示唆する箇所	「最終的にガンになるリスクが増えると考えられています」「無用な外出は控えた方がよいでしょう」「死に至ります」「中毒症状が出て」	38
DANG_REASON	危険を示唆する理由, または危険を間接的に示唆する箇所	「治療能力が追いつかなくなり」「暑さにのぼせたりしたためであって」「いづれ海藻や魚介類に取り込まれます」「放射性物質が少しずつもれています」	16
DANG_COND	危険が成立する条件	「最初の放射能濃度が高ければ」「放射線を1度に浴びると」「放射性ヨウ素を呼吸や食品を通じて大量に体に取り込んだ場合」「風向きによっては」	13

をツイートに適用し, 消費者信頼感指数や大統領の支持率などの世論調査の結果と感情分析結果との間に, 相関を見いだした. Kouloumpis ら [6] は, Twitter のハッシュタグや顔文字の情報を用いて感情分析の学習データを自動獲得するアプローチを提案した. Speriosu ら [13] は, 分類器の極性判定結果と Twitter のフォロワー関係の情報をラベル伝搬で統合する手法を提案している.

3 予備実験

3.1 実験に用いたコーパス

安全・危険に関する情報やその根拠を抽出する際の技術的な課題を確認するため, 予備実験を行った. この予備実験では, 自動獲得した知識に基づく論述構造の分析は行わず, 表層的な情報抽出のアプローチで論述構造を解析する. 今回は, 安全・危険に関して言及している文書データとして, 日本保健物理学会が提供している「専門家が答える暮らしの放射線 Q & A」³を用いた. このサイトは, 利用者から寄せられた放射線に関する質問に対し, 専門家の知見や判断を踏まえて一つ一つ丁寧に回答を行うものである.

今回は, 「暮らしの放射線 Q & A」からランダムに選んだ 300 件の回答部分に対して, 表 1 に示す 6 種類 (安全 / 危険に関して直接的な言及, 理由, 成立条件) の記述の箇所にラベル付けを行った. 表 1 にラベル付けの基準を示したが, 安全 / 危険の理由を述べている箇所 (SAFE_REASON / DANG_REASON) は, 文脈や背景知識などによる人間の推論が働くと, 直接的に示唆する箇所 (SAFE / DANGER) と紛らわしいので, これらを区別するようにアノテーション作業者に指示を与えた. 最終的に, 162 件の回答に何らかのラベルが付与され

³<http://radi-info.com/>

た⁴. 今回のデータでは, 安全に関して直接的に言及する箇所が圧倒的に多く, その言い回しも多様であった.

3.2 実験

3.1 節でラベル付けしたテキストを Cabocha⁵で形態素解析・係り受け解析を行った. 文節に対して IOB2 記法でラベル付けの領域を表現することで系列ラベリング問題に変換し, CRFsuite⁶で条件付き確率場を学習した. 学習に用いた素性テンプレートを表 2 に示した. ここで, 位置 t の文節に対し, w_t , pos_t , pos_detail_t , $phrase_t$ はそれぞれ, 形態素のユニグラム, 品詞のユニグラム, 詳細な品詞分類のユニグラム, 文節のフレーズ全体を表す. また, $t.parent$ は位置 t の文節の係り先, $t.children$ は位置 t の文節の係り元を表す. 辞書素性 d_t は, 位置 t の語が評価極性の判定に使われる辞書に収録されているかを示す. 今回は, 評価極性の辞書として, ALAGIN の「意見 (評価表現) 抽出ツール用モデル」⁷, 用言と名詞の日本語評価極性辞書 [15, 17] を用いた. 二段階予測素性 [7] は, 1 次マルコフ連鎖のグラフィカルモデルで離れた位置のラベル間の依存関係を取り入れるもので, 1 段階目のラベルの予測結果 y'_t を, 2 段階目の予測の素性として利用する.

安全・危険に関するラベルを 10 分割交叉検定で予測したところ, マイクロ平均 F1 スコアは 52.1 であった. 各ラベル毎の F1 スコアを調べると, 71.4 (SAFE), 44.3 (SAFE_REASON), 25.7 (SAFE_COND), 10.2 (DANG), 13.2 (DANG_REASON), 23.5 (DANG_COND) であった.

⁴300 件の Q & A の中には, 安全・危険とは関係なく, 一般的な知識を問う質問 (例えば「放射性ヨウ素は気体ですか, 固体粒子ですか?」) が含まれており, 今回の実験ではこれらを除外した.

⁵<http://code.google.com/p/cabocha/>

⁶<http://www.chokkan.org/software/crfsuite/>

⁷<http://alaginrc.nict.go.jp/resources/nictmaster/resource-info/abstract.html#C-3>

表 2: 素性テンプレート

種類	定義
基本	$\{x_{t-2}, x_{t-1}, x_t, x_{t+1}, x_{t+2}\} \times y_t$ $\{\textit{phrase}_t, w_{t.\textit{parent}}, w_{t.\textit{children}}\} \times y_t$ (ただし $x = (w, \textit{pos}, \textit{pos_detail})$)
辞書	$y_{t-1} \times y_t$ $\{d_t\} \times y_t$
二段階予測	$\{y'_{t-10}, y'_{t-9}, \dots, y'_t, \dots, y'_{t+10}\} \times y_t$

表 3: 安全に関するラベル付けの F1 スコア

素性セット	SAFE	REASON	COND	全体
基本	71.8	43.8	27.9	54.9
+ 辞書	72.3	43.9	28.1	55.2
+ 辞書 + 二段階	69.9	47.3	31.6	55.9

危険に関する予測性能が極端に低いのは、学習データの不足によるものと考えられる。

そこで、今回のデータから危険に関するラベルを削除し、安全に関するラベル付けを行った結果を、表 3 に示した。評判分析では評価極性辞書を導入することで分類の性能が大幅に向上していたが、今回のデータでは性能の上昇幅が小さかった。今後、人間の安全・危険に関するドメインで、極性辞書を整備する必要があると考えられる。また、二段階素性を導入することによって、学習データの少なかった SAFE_REASON や SAFE_COND のラベル付け性能が改善した。1 次マルコフ連鎖のグラフィカルモデルでは文の構造を陽に表現できないが、安全の示唆と理由・条件との依存性をモデルに取り込むことによって、論述構造の解析がさらに改善する可能性がある。

4 おわりに

本稿では、Web 文書から人の安全・危険に関する情報の構造化を行う研究に関する我々の取り組みを紹介した。今後は、ウェブから大規模な知識を獲得する研究を進め、知識に基づく論述構造の解析に取り組む予定である。また、一般のウェブ文書や、ツイートなどにもアノテーション作業を拡大し、実用的なプロトタイプシステムの開発も並行して進める予定である。

謝辞

本研究は、文部科学省科研費 (23240018)、文部科学省科研費 (23700159)、および JST 戦略的創造研究推進事業さきがけの一環として行われた。

参考文献

- [1] A. Acar and Y. Muraki. Twitter for crisis communication: lessons learned from japan's tsunami disaster. *International Journal of Web Based Communities*, 7(3/2011):392–402, 2011.
- [2] D. Demner-Fushman, W. W. Chapman, and C. J. McDonald. What can natural language processing do for clinical decision support? *Journal of Biomedical Informatics*, 42:760–772, 2009.
- [3] S. Doan, B.-K. H. Vo, and N. Collier. An analysis of Twitter messages in the 2011 Tohoku Earthquake. In *4th ICST International Conference on eHealth*, 2011.
- [4] M. Gupta and J. Han. Heterogeneous network-based trust analysis: a survey. *ACM SIGKDD Explorations Newsletter*, 13:54–71, 2011.
- [5] M. Gupta, P. Zhao, and J. Han. Evaluating event credibility on twitter. In *SIAM International Conference on Data Mining (SDM 2012)*, p. (to appear), 2012.
- [6] E. Kouloumpis, T. Wilson, and J. Moore. Twitter sentiment analysis: The good the bad and the omg! In *Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*, pp. 538–541, 2011.
- [7] V. Krishnan and C. D. Manning. An effective two-stage model for exploiting non-local dependencies in named entity recognition. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics (ACL 2006)*, ACL-44, pp. 1121–1128, 2006.
- [8] S. M. Meystre, G. K. Savova, K. C. Kipper-Schuler, and J. F. Hurdle. Extracting information from textual documents in the electronic health record: a review of recent research. *IMIA Yearbook of Medical Informatics*, pp. 128–144, 2008.
- [9] J. R. C. Nurse, S. S. Rahman, S. Creese, M. Goldsmith, and K. Lamberts. Information quality and trustworthiness: A topical state-of-the-art review. In *2011 International Conference on Computer Applications and Network Security (ICCANS 2011)*, 2011.
- [10] B. O'Connor, R. Balasubramanyam, B. R. Routledge, and N. A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth International AAAI Conference on Weblogs and Social Media (ICWSM 2010)*, pp. 122–129, 2010.
- [11] J. Pasternack and D. Roth. Knowing what to believe (when you already know something). In *23rd International Conference on Computational Linguistics (Coling 2010)*, pp. 877–885, 2010.
- [12] T. Sakaki, F. Toriumi, and Y. Matsuo. Tweet trend analysis in an emergency situation. In *Special Workshop on Internet and Disasters (SWID 2011)*, pp. 3:1–3:8, 2011.
- [13] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldridge. Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proceedings of the First workshop on Unsupervised Learning in NLP*, pp. 53–63, 2011.
- [14] K. Zickuhr. Generations 2010. Technical report, Pew Research Center, 2010.
- [15] 小林, 乾, 松本, 立石, 福島. 意見抽出のための評価表現の収集. *自然言語処理*, 12(2):203–222, 2005.
- [16] 梅島, 宮部, 荒牧, 灘本. 災害時 twitter におけるデマとデマ訂正 RT の傾向. *情報処理学会研究報告*, 第 2011-IFAT-103 巻, pp. 1–6, 2011.
- [17] 東山, 乾, 松本. 述語の選択嗜好性に着目した名詞評価極性の獲得. *言語処理学会第 14 回年次大会論文集*, pp. 584–587, 2008.
- [18] 宮部, 荒牧, 三浦. 東日本大震災における twitter の利用傾向の分析. *情報処理学会研究報告*, 第 2011-DPS-148/2011-GN-81/2011-EIP-53 巻, 2011.
- [19] 総務省. 情報通信白書平成 23 年版. <http://www.soumu.go.jp/johotsusintokei/whitepaper/>, 2011.
- [20] 野村総合研究所. プレスリリース: 震災に伴うメディア接動向に関する調査. <http://www.nri.co.jp/news/2011/110329.html>, 2011.