

ネット上の情報コンテンツの信頼性を分析・検証する言語情報処理技術

乾健太郎 岡崎直観 (情報科学研究科システム科学専攻,)

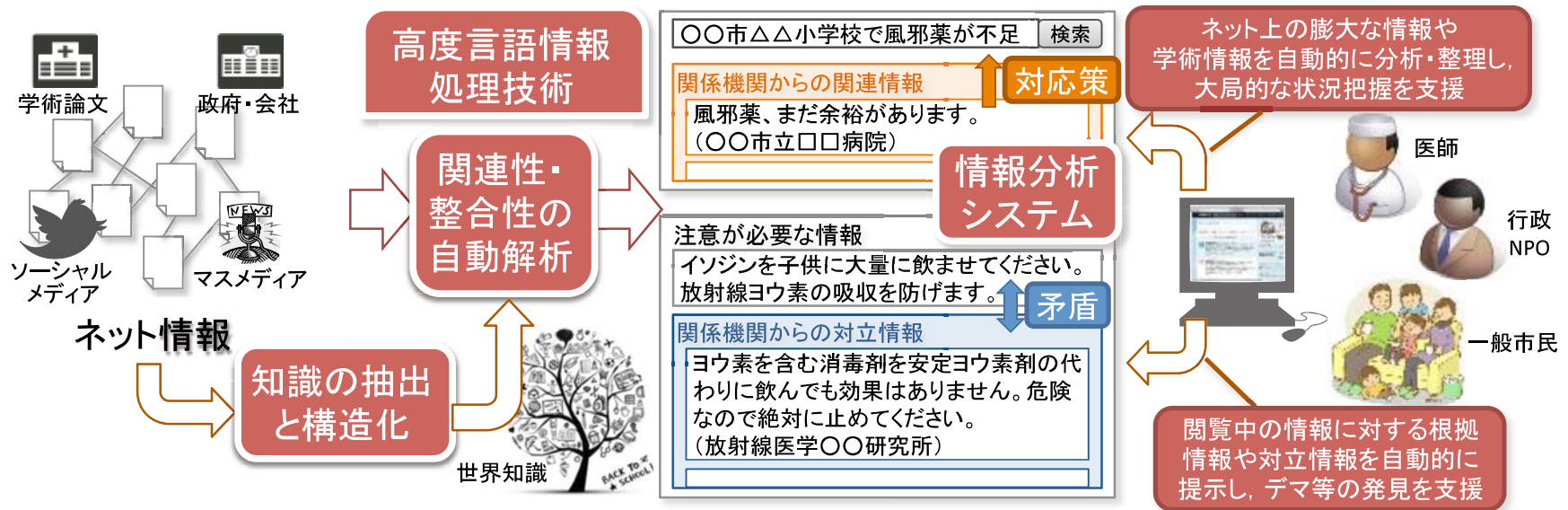
電話 022-795-7091
email inui@ecei.tohoku.ac.jp

研究の背景:

東日本大震災では、Twitterや様々な支援サイトなど、インターネットを介して安否情報や被災状況等の情報の集約・流通が進み、家族との再会や被災地への物資の供給等の支援に大きく貢献した。こうした情報共有を促進する情報技術は将来の災害への対策としても極めて重要である。一方で、今回の震災では偏った情報やデマが拡散し、社会の混乱や不安も広がった。ネット上の情報を有効に活用し、安心して便利な社会を実現するには、情報の信頼性を自動的に分析・検証する技術が不可欠である。

研究の目的:

- ネット上の言語情報(自然言語で書かれた非構造情報)を自動的に解析・集約し、流通している情報の背後にある論理構造を解析・分析することで、ユーザが安全かつ有効に利用できる情報環境を構築する。
- 例:保健医療情報分析システムの開発(下図参照)
 - 医療関係の学術文献DB, 関係機関のWebサイト, Facebook・Twitter等のソーシャルメディアなど、多種多様な情報源から恒常的に情報を収集する仕組みを作る。
 - さらに収集した情報同士の相互の関連性(例えば、問題に対する対応策)や整合性(根拠情報や矛盾/対立する情報)を自動解析し、俯瞰的な分析結果をユーザに提示する。
 - これらにより、災害時の保健医療情報の収集・伝達, デマ等の低品質情報の発見・検証などを支援する。



● 研究のコアとなる技術・特徴・特許 (下図参照)

- 知識ベース構築: ウェブなどの大規模文書集合から対象領域の知識を自動抽出し、大規模知識ベースを構築する技術
- 文書内の論述構造の解析: 対象文書に書かれた情報を知識ベースに照らして解釈し、その論述構造を解析する技術
- 文書間の整合性のモニタリング: 文書間で論述構造を比較し、矛盾等の不整合情報を自動検出する技術

● 研究の波及効果・応用

- 専門家(例えば、医療機関や行政機関、支援団体)に対する効果: 被災地情報や学術情報等へのアクセシビリティが向上し、大局的な保健医療状況などの把握が容易になる。
- 一般市民(例えば、患者とその家族、一般市民)に対する効果: 専門家が提供する情報へのアクセシビリティが向上するとともに、デマ・風評など低品質情報の発見・検証が容易になる。保健医療分野で言えば、市民がデマや民間療法に惑わされず安心して医療を受けられるようになり、不足する医療コーディネーションの補完にも繋がる。

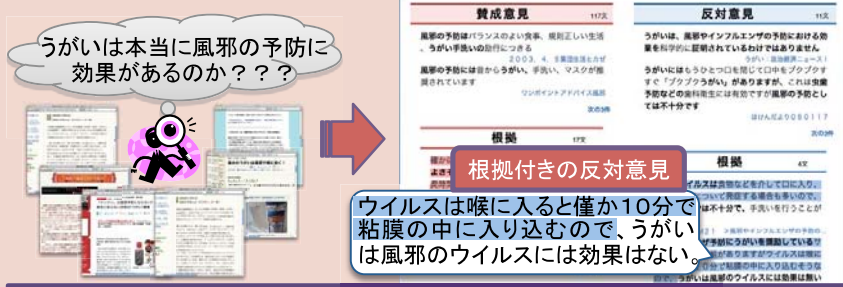
● 研究期間・予算(目安)・共同研究先:

- 3~5年, 8千万~1億円, ネットIT企業(Yahoo!, ニフティなど), ITベンチャー(mixi, PFIなど)

言語情報の自動分析

玉石混淆のネット情報を自動解析し、信頼性の検証を支援する高度な言語意味処理技術

- ネットの多様な言い回しや専門用語を認識する意味解析技術
- 膨大な知識を駆使して同意・対立の関係を認識する推論技術



賛成・反対意見を根拠情報とともに俯瞰できる

言語解析の基盤技術

多様な言語表現を実世界に対応づける言語処理

