

AIを用いたSNS上の災害関連情報を要約・分析するシステムの利活用



国立研究開発法人情報通信研究機構
耐災害ICT研究センター 応用領域研究室 上席研究員

おおたけ きよりの
大竹 清敬

1. はじめに

東日本大震災では、ツイッター等のSNS上に有用な災害関連情報が投稿される一方で、それらの情報を利活用するための分析・検索手段が充分ではなく、必要とする情報を短時間に得ることが容易ではなかった。情報通信研究機構(NICT)では、東日本大震災を契機として、これらのSNS上に存在する災害関連情報を迅速に分析し、被災者はもとより、復旧、救援活動を行う人々に対しても有用な情報を提供するシステムを研究開発してきた。その成果の一つが現在投稿されているツイートをリアルタイムに分析するシステムとして2015年4月より<http://disaana.jp>にて試験公開している対災害SNS情報分析システムDISAANA(ディサーナ; DISAster-information ANALyzer)である。

2016年4月に発生した熊本地震では、損壊した家屋が多く発生した。避難所が不足したため、指定避難所以外へ避難する人、車中泊にて過ごす人なども多数発生した。一方で、東日本大震災以降の通信インフラの強靱化により、携帯電話等が使用できないエリアや時間帯は限定的であった。そのため、個人からの情報発信等により災害関連情報があふれ、それによる混乱も多数発生したと聞いている。

内閣官房は、ツイッター分析班を設置し、指定避難所以外のニーズを把握するためにDISAANAを活用して情報収集にあたった。集められた災害関連情報は、現地災害対策本部に伝えられた。担当した参事官によると、被災自治体が混乱している発災直後は特に有用だったとのことである。

本稿では、熊本地震でも活用されたDISAANA及び、DISAANAでは把握が困難な災害状況の全体像を見渡すことができるシステムとして2016年10月より試験公開を開始した災害状況要約システムD-SUMM(ディーサム; Disaster-information SUMMArizer)について紹介し、その利活用について述べる。

2. 対災害SNS情報分析システムDISAANA

キーワード検索は、多くの人が情報を探するために日常的に利用している検索手段であろう。日常的な情報検索においては、求める情報が1つでも見つければそれで済む場合が多く、キーワード検索で充分かもしれない。しかしながら、

災害時の様々なニーズあるいは、救助を求める情報等を探すためには、回答を網羅的に拾い上げる必要がある。このような場合、キーワード検索では、その検索結果の一つ一つをつぶさに確認する必要がある、膨大な時間を要する。そこで、NICTでは、東日本大震災を契機として、質問応答技術を用いて求める情報をピンポイントに、しかも漏れなく提供する事ができるシステムとしてDISAANAの研究開発を進めてきた。その詳細については、NICT NEWS No. 452, 2015 (http://www.nict.go.jp/data/nict-news/NICT_NEWS_1506_J.pdf) 等をご覧いただきたいが、以下にその特徴をまとめる。

- 文の構文構造の解析結果をパターンにまとめ、それを用いて回答をピンポイントに提供する質問応答機能を実現
- 表現の多様性をカバーする3億件の同義パターン辞書をAI(Artificial Intelligence;人工知能)技術により構築し活用
- 400万件の地名データベースにより、地名処理にて位置情報を提供。それを用いて、回答と関連する地点を地図上に表示
- 矛盾する情報を同時に検索するための矛盾パターン辞書(250万件)を同様にAI技術により構築し、これによりデマの判断材料を提供
- エリアを指定するだけでそこで挙がっている被災報告を自動的に抽出する機能も提供。被災報告に対応する対策(たとえば、「ミルクが足りない」という被災報告に対して「ミルクが届いた」が対策になる)も自動抽出し、対応づけて提示
- 災害オントロジーと呼ぶ意味カテゴリー辞書を整備し、これを用いてシステムからの出力をわかりやすく提示

DISAANAの質問応答エンジンは、従来の検索エンジンを発展させたものと言えるが、回答候補を探すその検索において、異なる表現だがほぼ同じ意味を表しているものを捉える、あるいはその反対に矛盾する表現を捉えるための辞書を構築するためにAIを活用している。こうした辞書を構築するためには、例えば、同義パターン辞書であれば、種となる同義パターンを集めた学習データを用意し、AIに

学習させ、AIに大量の言語データの中から同義パターンを抽出させる。この抽出結果には、合っているものもあれば、間違いも含まれるため、それらの間違いを訂正し、より大きな学習データとし、再度学習を行う。これらを繰り返し、AIによる抽出結果の精度を高め、最終的に大規模な同義パターン辞書の抽出が可能となる。このように、AIはDISAANAや、これから説明するD-SUMMにおいても非常に重要な役割を担っている。こうした特徴を持つシステムを<http://disaana.jp>にて公開しており、誰もが無償で利用できる。必要なのは、Webブラウザだけであり、パソコン、スマートフォン等で利用可能である。なお、スマートフォンからアクセスした場合は、スマートフォンに適したインターフェースで利用できる。

3. 災害状況要約システムD-SUMM

DISAANAを活用することによって災害関連情報を容易に得ることができるようになったものの、大規模災害時にはその出力結果も膨大となるため、どこで、どのような被害が発生しているかを把握することが困難であった。NICTでは、内閣府総合科学技術・イノベーション会議の戦略的イノベーション創造プログラム(SIP)「レジリエントな防災・減災機能の強化」(管理法人・JST)の支援を受けて、これらの問題を解決するために災害状況要約システムD-SUMMの研究開発を進め、2016年10月より<http://>

disaana.jp/d-summ/にて公開を開始し、誰もが利用できるようになってきている。

D-SUMMは、要約を作成する対象となるエリアが指定されるとそのエリアで挙がっている被災報告を自動的に抽出し、要約、整理して提示するシステムである。H28年熊本地震の前震から1時間のツイートを対象に熊本県について要約した結果の例を図1に示す。被災報告を自動的に抽出する機能は、DISAANAも有しているが、この機能は、AIを用いて実現している。まず、どういった表現が被災報告であるかの学習データを大量に用意し、その表現が用いられる文やツイートの情報も含めてAIにそれらを学習させる。そうすると、AIは一度学習した表現を抽出できるようになるのはもちろんのこと、初めて見る表現も、周辺にあらわれる情報から被災報告を表していると推定できる場合は、被災報告として抽出できるようになる。

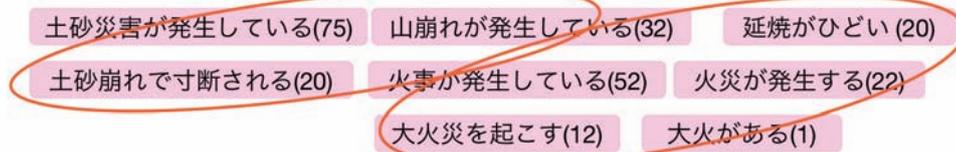
DISAANAでは、このような自動抽出結果が要約・整理されることなくそのまま提示される。例えば、DISAANAでは「土砂災害が発生する」「土砂崩れで寸断される」「山体崩壊で壊滅する」などの多様な表現がそのまま出力される。一方D-SUMMでは、これらを「土砂災害がおきている」という1つの被災報告に要約し、さらにこの被災報告に係る地名を列挙し、それぞれ何件の被災報告があるかを整理して出力する。D-SUMMにおける要約の概要を図2に示す。



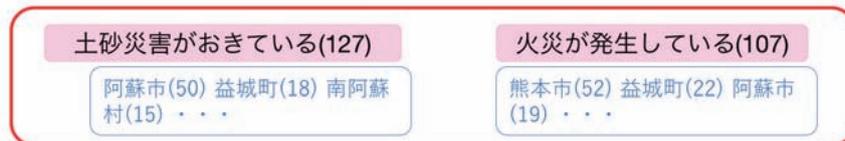
図1. H28年熊本地震前震後1時間のツイートを対象にしたD-SUMMの出力結果例(対象エリア:熊本県)



(1) 被災報告の抽出 : DISAANAと同様の仕組みで抽出



(2) 類似表現の集約による要約と場所情報の整理



■ 図2. D-SUMMにおける要約処理の概要

これを可能にしているのは、災害オントロジーと呼ぶ意味カテゴリー辞書の細分化である。DISAANAでは、「土砂崩れ」や「山体崩壊」といった単語が「災害」というカテゴリーだけに分けられるのに対し、D-SUMMでは、このカテゴリーを細分化し、災害の下に地震、火災、土砂災害、といったカテゴリーを設け、既存の意味カテゴリー辞書を整理した。その結果、上記の「土砂災害」、「土砂崩れ」、「山体崩壊」には「災害：土砂災害」というカテゴリーが与えられる。

加えて、述語とそれにかかる助詞の組合せに対して、その組合せが、ある名詞Nに対してその機能をオンにする(活性)かオフにする(不活性)かを記述した述語の意味的極性辞書を整備した(約5万件)。例えば、「Nがある」「Nがおきる」は活性であり、「Nがない」「Nを失う」は不活性である。D-SUMMでは、これを用いて要約を行う。要約の方法は、同一の意味カテゴリー、同一の意味的極性を有する被災報告を1つの被災報告にまとめる。上記の土砂災害の例では、「土砂災害」、「土砂崩れ」、「山体崩壊」いずれの表現も名詞の意味カテゴリーは「災害：土砂災害」であり、また、そこで用いられる助詞と述語の意味的極性はいずれも活性であることから、意味カテゴリー辞書のラベルの下位部分(土砂災害)を直接用いて「土砂災害がおきている」と要約する。要約における述語部分(上記の場合は「がおきている」)については、意味的極性が活性の場合には「がおきている」としたり、要約対象の被災報告における頻度が最も大きい被災報告の術語を用いたり、出力結果が可能な限り自然になるよう規則に基づき決定する。

また、D-SUMMでは、DISAANA同様に地名処理結果

に基づく地図表示が可能となっているが、細分化した意味カテゴリー単位で、しかもそれらを組み合わせで地図上に結果を表示できる。そのため、例えば、救助カテゴリーと道路カテゴリーを合わせて地図上に表示した際の結果から、もし両方とも同じ地域で被災報告が挙がっている場合には、その地域の救助に向かうためには、道路が使えない状況も考慮すべきであるといった、災害状況全体を見渡した意志決定を支援することが可能である(図3)。

DISAANA及びD-SUMMは解析モジュールを共有している。このモジュールは、ツイートを受け取ってから約1秒で解析を終了し、解析結果を格納する。また、質問等のリクエストを受け取ってから1秒以内にその結果を返す。解析に関しては、最大、毎秒5,000ツイートほどを処理し続ける事が可能な規模で運用しており、DISAANAの質問応答に関しては、最大で毎秒400質問を処理可能な規模で運用している。こうした、リアルタイム性が要求されるシステムを実現するためには、大規模な計算機サーバを効率よく利用する枠組みが必要である。また、研究開発段階からこうした実運用段階へスムーズに移行する上でも、研究で用いたソフトウェアを改変せずに、実運用規模にスケールさせる仕組みが望まれる。NICTでは、そのようなミドルウェアとしてRaSC(Rapid Service Connector)を開発し、<http://alaginrc.nict.go.jp/rasc/>にて一般公開している。DISAANA及びD-SUMMでは、解析モジュールを共有していることは既に述べたが、システムを構成する様々なモジュールにおいてRaSCを活用しており、RaSCを通して様々なコンポーネントが接続されている。RaSCはDISAANAやD-SUMMのような大規模データをリアルタイムに処理するシステムを構築する上で非常に有用な



図3. H28熊本地震熊本震災後12時間のツイートを対象にD-SUMMにて熊本県の被災状況を地図表示した例

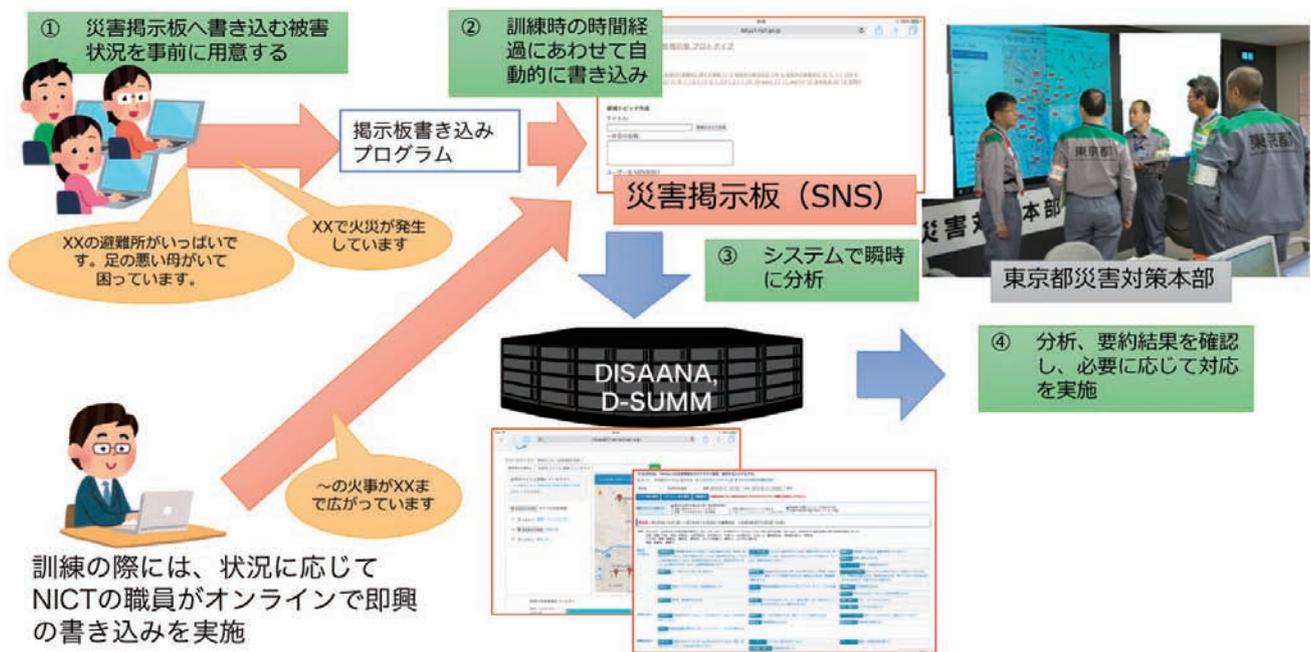
ソフトウェアであり、興味がある方には是非一度お試しください。

4. 地方自治体の防災訓練における利活用

NICTでは、研究開発成果の社会実装を目指して、実証実験等を実施している。我々が公開しているいずれのシステムもWebサイトにアクセスするだけで利用できるが、有効に利活用するためには、実際の大规模災害時にどのような情報がどの程度流れるのか、どういった問題があるのかなどを事前に体験することが望まれる。その一部に役立てていただこうと、DISAANAとD-SUMMでは、過去の大規模災害として、東日本大震災ならびにH28年熊本地震の際のツイートを対象とした試用版を提供している。しかしながら、これらの試用版は、いずれも、一部の地域での被害について体験できるのみである。そこで、自らが住む町で大きな地震が発生するとどういった投稿がSNSに流れるかといったことを体験してもらえよう、防災訓練等でのリアルなシミュレーションを実施できる体制を整えている。

この体制の検証の場として、我々が研究開発を進めてきたD-SUMM及びDISAANAが平成28年度東京都図上訓練

にて活用されたのでその概要について述べる。この訓練は2017年1月31日に首都直下地震を想定して実施され、東京都職員等200名、政府現地対策本部として各府省職員等85名が参加した。この訓練におけるD-SUMM及びDISAANAの活用の概要を図4に示す。この訓練では、まず、SNSを模した掲示板を用意し、掲示板への投稿をリアルタイムに分析するようシステムを構成した。次に想定される被害状況（状況付与）等に関する書き込みを事前に人手で約4,000件、テンプレートを用いたプログラムによる自動生成により約3,000件の合計約7,000件用意しておき、これを訓練の時間経過にあわせて自動的に掲示板に投稿するようにした。これらの投稿をD-SUMM及びDISAANAで、集約、分析、要約することで、時間経過にあった被害状況の分析結果を提供した。2015年に宮崎県で実施した類似の訓練では、防災士や学生等のボランティアを50人以上集め、訓練時に状況に応じて掲示板へ投稿するという形式でシステムを活用した。それに比べて、短時間に大量の投稿を投入することも可能で、よりリアルな状況に近づけることができたと考える。また、事前に投稿を用意することで、システムでの分析結果を事前に確認しておくことができ、システムの辞書等の不備を事



■図4. 東京都図上訓練でのD-SUMM・DISAANAの活用概要

前に修正できるという利点もある。

訓練は、参加者に訓練内容は伝えられず、コントローラと呼ぶグループが進行に合わせて逐一情報を与える（状況付与する）形式で実施された。また、訓練時には、東京都側の要請に合わせて、オンラインでNICT職員がその場で投稿する事も実施した。訓練に参加した東京都の職員からは、発災直後の状況把握におけるSNS情報の集約が重要な情報源であることが認識できたとのコメントをいただき、好評であった。また、システムの活用を通して、システムの機能上の改善を要する点の発見、データ整備におけるノウハウの蓄積等、今後のシステム開発における重要な経験を得ることができた。

さらに、2017年4月25日に大分県にて実施された平成29年度大分県総合防災訓練（図上訓練）においても同様の形式でD-SUMM、DISAANAが活用された。大分県総合防災訓練では、大分県側が1,800件の投稿を事前に用意し、NICTが4,000件ほどの投稿をプログラムで自動生成して訓練を実施した。特に大分県での訓練では、デマ情報の検出等に期待が寄せられており、期待に応える情報を提供できた一方で、一部の機能には問題があることが分かった。今後は、

こうした防災訓練における活用を通して得られた問題点等を改善するとともに、全国各地を対象として、地震や台風等の風水害といった訓練内容に応じた投稿用データを半自動的に生成するソフトウェアの開発等、可能な限り低いコストで防災訓練におけるシステムの活用ができる体制作りを進める予定である。

5. おわりに

本稿では、NICTにて研究開発し、現在一般公開している2つのシステム、DISAANAとD-SUMMについてその概要を紹介し、その利活用例として地方自治体の防災訓練における活用について述べた。今後は、これらの訓練で判明したシステムの問題点を改善するとともに、システム機能を外部プログラムから呼び出して使うためのインタフェース（API: Application Programming Interface）の公開等を検討し、より確実に社会実装を進めていく予定である。また、全国各地を対象として、様々な条件での防災訓練等での活用が容易に実行できる体制作りを進め、平常時においても、災害時の情報収集や意志決定をシミュレーションできる環境を整えていく予定である。