

対災害情報分析システム DISAANA 及びその質問応答モードの性能評価

水野 淳太^{1,a)} 後藤 淳^{1,†1,b)} 大竹 清敬^{1,c)} 川田 拓也^{1,d)} 鳥澤 健太郎^{1,e)}
クロエツェー ジュリアン^{1,f)} 田仲 正弘^{1,g)} 橋本 力^{1,h)} 奥村 明俊^{1,i)}

概要：我々は、災害時に Twitter に投稿される膨大な情報を効率よく検索するために対災害 SNS 情報分析システム DISAANA を開発し、誰もが利用可能な Web アプリケーションとして試験公開している。本論文では、これまでにやってきたシステムの改善ならびに不適切な回答候補を抽出する事を回避するために新たに導入したモダリティ解析、ツイート属性判定、予報表現抽出について議論する。その上で、これまで東日本大震災関連の災害情報のみで行われてきた本システムの評価を、台風や大雪といった一般的な災害にまで拡張し評価を行う。その結果、さらなる改善の余地が残されているものの、実用可能な性能に達していることを確認できた。

1. はじめに

東日本大震災では、Twitter に膨大な量の災害関連情報が投稿された。米 Twitter 社によると、1 秒あたりのツイート数が 5,000 件を超えることが 5 回あり、日本からのツイート数は地震発生後に 500%増加した^{*1}。震災に限らず、災害時に Twitter に投稿される災害関連情報は、即時性が高く重要な情報が含まれる一方で、投稿数が膨大であるため、一般的なキーワード検索によって必要な情報を効率良く入手することは困難である。そこで、我々は災害関連情報をリアルタイムに効率的に検索することができるシステムとして、対災害 SNS 情報分析システム DISAANA を開発した。本システムは、スマートフォンおよびパソコン経由で誰でも利用可能な Web アプリケーションとして、<http://disaana.jp/> にて試験公開されている。

DISAANA は、災害時に発信される膨大な情報から必要

とする情報を効率的に見出し、災害状況等を俯瞰的に把握できるように質問応答を用いた情報アクセス手段を提供する。一方で、我々が事前に調査したところでは、こうした質問応答手段があったとしても災害時の逼迫した状況の中では、質問そのものを考えるのは困難であるとの指摘がある地方自治体よりあった。そこで、DISAANA では、市町村等のエリアを指定するとそこで起きているトラブルや問題を自動的に検出する機能も提供する。DISAANA では、前者を質問応答モード、後者をエリア検索モードと呼ぶ。

質問応答モードでは、自然文による質問を入力すると、その回答候補を一覧することができる。例えば、「X 市で何が不足していますか」という質問を入力すると、「X 市で毛布が不足している」や「X 市の病院で透析用チューブが足りない」といったツイートから、質問の答えとなる「毛布」「透析用チューブ」などが得られる。つまり、「不足する」と「足りない」といった表現の違いを吸収した上で、ピンポイントに質問の回答候補を網羅的に出力する。

エリア検索モードでは、県名や市区町村名を入力すると、その地域で発生している災害やトラブルを広く一覧することができる。例えば、「Y 県」のようにエリアを指定すると、「Y 県ではガソリンが枯渇している」や「Y 県 Z 市で停電が発生している」といったツイートから「ガソリンが枯渇」「停電が発生」といった情報が得られる。いずれのモードでも、回答候補は、災害やトラブル、犯罪などのカテゴリごとに分類表示されるので、必要な情報を効率良く入手することができる。

エリア検索モードの手法は István ら [1] によるものであ

¹ 情報通信研究機構

NICT, Seika-cho, Kyoto 619-0289, Japan

^{†1} 現在、NHK 放送技術研究所

Presently with NHK STRL, Kinuta, Setagaya-ku, Tokyo, 157-8510, Japan

a) junta-m@nict.go.jp

b) goto.j-fw@nhk.or.jp

c) kiyonori.ohtake@nict.go.jp

d) tkawada@nict.go.jp

e) torisawa@nict.go.jp

f) julien@nict.go.jp

g) mtnk@nict.go.jp

h) ch@nict.go.jp

i) okumura@nict.go.jp

*1 <https://blog.twitter.com/2011/global-pulse>

る。質問応答モードの手法は、後藤ら [2] によるものに基づいているが、さまざまな問題に対応するため、数多くの拡張、改良が施されてきている。後藤らの手法における問題の一つは、事実性の認定が不十分であったことである。例えば、「大雪になる見込み」や「酸性雨が降るといのはデマ」といった文の「大雪」や「酸性雨」は、現時点で実際に起きている事象として取り扱われていた。災害対応という観点からは、実際に起きている事象をいかに正確に捉えるかが重要である。また、予報は予報として区別されることが望まれる。さらに、後藤らの手法による質問応答システム（以下ではプロトタイプシステムと呼ぶ）は、東日本大震災という地震と津波を中心とした災害に対して調整、検証されており、その他の災害についての対応が不十分であった。そこで、本論文では、これらを解決するための手法を新たに導入し、地震のみならず、台風や大雪などの災害における質問応答モードの性能を評価し、議論する。

事象の事実性を認定することは、容易なように考えられるかもしれないが、否定の表現ひとつとっても、「火災は発生していない」といった単純なものから「火災発生というチェーンメールが来た」のように、チェーンメールの内容は通常真実ではないといった複雑な解釈プロセスを経て認識されるものまで多様である。また、単純に直近の「～ない」といった表現があるから否定であるという判断はできず、文、あるいは文を超えた広い範囲を考慮してはじめて判断できるものもある。我々はこうした多様な否定表現を扱えるモダリティ解析器を開発・導入する。

否定表現は、いわゆるデマを判断するための材料を提供するという意味でも重要である。例えば、「有毒物質が X 市で発生」というデマに対して「有毒物質は X 市では出ていない」という否定表現を検出できれば、ユーザにデマの可能性を判断する為の材料を提供することができる。DISAANA にはこのような機能が実装されている。

モダリティ解析の他に、災害対応の観点では、現在起きている事象が重要であるため、「東京では関東大震災があった」という表現の文については、過去、冗談といった属性を付与することで、回答候補の抽出源から除外する^{*2}。また、「X 市では今晚大雪の恐れ」といった、災害の予報情報は、現時点では非事実であるものの、近い将来に起きうる情報として有用であることから、予報情報であることを明記して表示する。属性および予報は、人手で整備した、過去、冗談、予報などを示唆するキーワードのリストと、ツイート中に含まれる時間表現によって判定される。

本論文の構成は以下の通りである。まず、2 節で質問応答システムなどの関連研究について述べる。3 節では、

DISAANA の構成を述べた後、本研究で新たに導入した言語解析モジュールについて詳述する。4 節では、東日本大震災時のツイートデータに対する質問応答モードの性能評価を行い、5 節では、大雪や台風など、4 つの災害に対する性能評価を行う。DISAANA の活用について自治体の防災訓練を通じた実証実験を行ったので、6 節では、それについて報告する。7 節で、結論を述べる。

2. 関連研究

近年では、検索エンジンや質問応答システムなど、情報アクセス手段の進歩はめざましく、IBM 社の Watson [3] がクイズ番組で人間のチャンピオンに圧勝し一躍有名になったことは記憶に新しい。Watson は、Wikipedia を含む辞書、辞典、台本など、クイズ番組の分野に関連する確かな知識をあらかじめ選別しデータベース化している。また、Watson が勝利したクイズ番組における質問はその回答が一意に定まるものだけであった。それに対し DISAANA は、リアルタイムに更新される質問応答用データベースから多数の回答候補を提供しようとする点が大きく異なる。

また、吉村 [4] は、実運用されている「しゃべってコンシェル」の構成について紹介しているが、このサービスでの質問応答では、質問のカバー率が低いものの、精度の高い回答を返すことができる DB 型質問応答システムと、質問のカバー率は高いが、精度はそれほど高くない検索型質問応答システムを組み合わせている。後者の Factoid 型と呼ばれる質問応答のアルゴリズムは、(1) 質問を分析し質問のタイプ、回答の属性（固有表現クラス）等を決定、(2) 検索を実行、(3) 検索結果に固有表現抽出器を適用し、求められる固有表現クラスと同一のものを回答候補として抽出し、スコアを算出しランキングする。一方で、DISAANA では、固有表現抽出器を用いず、構文パターンに基づいて回答候補を求めているため、期待している回答候補の属性とは異なる者が出力される場合があるものの、固有表現抽出器が対象とできない回答候補を出力できる。

本研究では、災害という観点から事実性の認定が重要であることは既に述べたとおりである。このような、事象の成否に対する著者の判断情報は、言語学ではモダリティと呼ばれる。その処理の難しさから、言語処理においてモダリティ解析が実用を前提に検討され始めたのは比較的最近であるといえる。言語処理におけるモダリティ解析の検討としては乾ら [5] と松吉ら [6] が詳しい。彼らは複雑なタグ体系を定義しているが、我々はそれら全てを解くことはせず、DISAANA において重要な否定と疑問・要求について取り組む。

本稿で提案するモダリティ解析は、学習データを用意し、機械学習にて対象となる事象の核となる述語のモダリティを推測するものである。そのような機械学習によるモダリティ解析として江口ら [7] の取り組みがある。江口ら

^{*2} より幅広く検索をしたい場合は、検索条件を変更することで、これらのツイートを回答候補の抽出源として利用することもできる。



図 1 DISAANA のスクリーンショット
質問「東京で何が発生していますか」の検索結果（2015/9/2 時点）

は従来論じられてきたモダリティに真偽情報、価値情報を統合した拡張モダリティと呼ぶ体系を整備し、その体系に基づいたアノテーションを行ったコーパスを構築し、Conditional Random Fields (CRF) を中心とした機械学習手法を用いたシステムと人手で作成した規則によるシステムを比較している。

Sauriら [8] は、事象に対する事実性を直接判定するべく、様々な手がかりを利用する Factuality Profiler と呼ばれるアルゴリズムを提案している。また Sauriらは、事実性を肯定・否定の極性と事象の真実性に関する書き手の自信の度合いに関するモダリティの 2 軸による事実性空間を定義し、その中で事実性を捉えようとしている。極性には positive, negative, unknown の 3 値を、モダリティには certain, probable, possible, unknown の 4 値を考え、これらの 2 つ組みで事実性を判定する。さらに Sauriらは、ニュース記事に対して時制の情報を付与した TimeBank という既存のデータ [9] に対してこれらの事実性の情報を付与したデータを作成し、FactBank として発表している [10]。

3. DISAANA の構成

DISAANA の質問応答モードの動作例を、図 1 に示す。この例は、「東京で何が発生していますか」という質問を入

力して検索した結果であり、「火事」「地震」「落雷がある」といった災害情報や、「運休がある」「事故」といったトラブルが検索されている。それぞれの回答候補をクリックすると、その抽出元となったツイートを閲覧することができ、誤情報でないかを確認することができる^{*3}。DISAANA が検索対象とするツイートは、日本語で書かれた全ツイートデータの 10%^{*4} のうち、間近の 4 日間に投稿されたものである。DISAANA では、間近の 4 日間の範囲であれば、任意の期間を設定して、検索対象となるツイートを限定することができ、最新の災害情報のみを得ることや、数日前の災害情報を俯瞰することが可能である。

スマートフォンなどのモバイル端末では、表示できる領域に限りがあり、操作方法もパソコンとは異なるため、専用のユーザインタフェースを提供している（図 2）。パソコン版で利用可能な機能はモバイル版でも全て利用可能である。さらに、モバイル端末の多くで利用可能な現在地情報を利用することで、エリア検索モードにおいて、現在地周辺で起きている災害・トラブル情報を素早く一覧することができる。

*3 表示する段階でユーザによって既に削除されたツイートは表示されない。

*4 https://nazuki-oto.com/twitter/service_menu.html

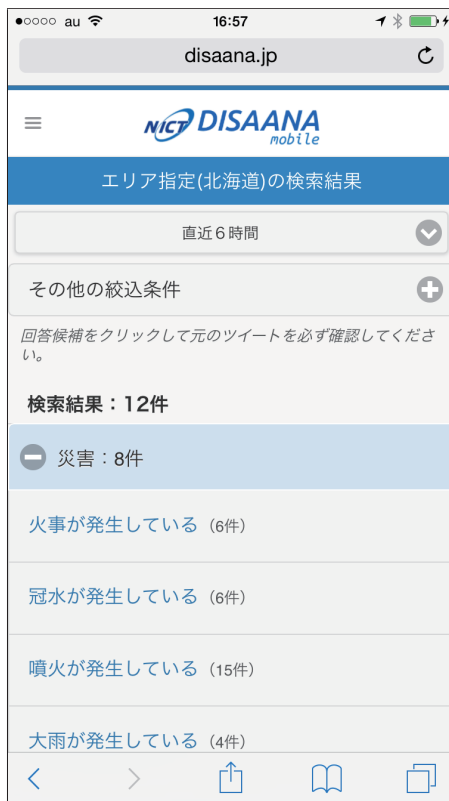


図 2 モバイル端末向けのユーザインタフェース
エリア検索モードにおける「北海道」の検索結果 (2015/9/2
時点)

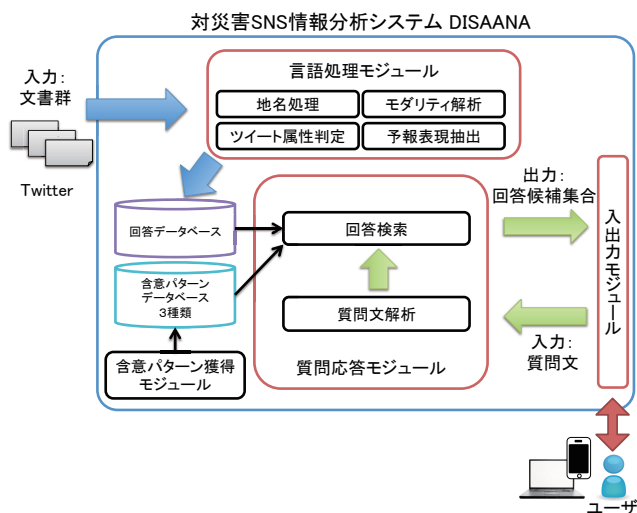


図 3 DISAANA の構成

DISAANA の構成を図 3 に示す。DISAANA は、大きく分けて、テキストを解析する言語処理モジュールと、入力された質問に回答する質問応答モジュールの二つの要素から構成される。言語処理モジュールは数百ツイートを同時に処理し、オンメモリの高速なデータベースに登録する。このデータベースを回答データベースと呼ぶ。質問応答モジュールは、入力された質問を分析し、回答データベースを検索することで、回答候補集合を得る。

まず、言語処理モジュールを構成する各言語処理器について概説する。

文分割 全ての文字を全角に変換し、句点などでツイート
を文単位に分割する。

形態素解析 全ての文について、単語分割し、品詞情報を
付与する。解析器には Juman 品詞体系 [11] の辞書を用いた MeCab[12] を用いる。

係り受け解析 全ての文について、単語を文節にまとめあ
げ、その修飾関係を解析する。解析器には J.DepP[13]
を用いる。

地名処理 「どこで火災が起きていますか」といった場所
を聞く質問に回答するために、文中に含まれる地名や
ランドマークについて、それらの完全な住所を辞書か
ら推定し、抽出する。地名やランドマークについてそ
の住所を扱えるようになるため、場所に関する階層性
を正しく処理できるようになる。たとえば、「宮城県の
どこで～」といった質問に対して、「仙台市で～」「南
三陸町で～」と記述されたツイートも対象として回答
候補を抽出できるようになる。また、辞書のエントリ
には、それぞれの緯度経度情報も付与されており、地
図上への場所の表示も容易になっている。

モダリティ解析 非事実の事象から回答候補を抽出するこ
とを防ぎ、誤情報を認知するための材料とするため、
文中の事象表現(動詞、形容詞、サ変名詞などの述語)
について、その成否を判定する。例えば、「X 市では
火事が《《 起きている》》の《《 起きている》》は実際に
起こったことであると判定され、「Y 市でのガソリン
《《 不足》》は誤報だった」の《《 不足》》は実際には起き
ていないことであると判定される。

予報表現抽出 ツイートの投稿時点では起きていないが、
近いうちに起きることが予想される災害情報に対して、
予報回答であると判定するために、予報であることを
示唆する表現を検出する。例えば、「X 市では今晚から
大雪になる見込み」における「今晚」「見込み」とい
ったキーワードや、11月5日に投稿されたツイート内の
「11月6日」といった未来の時間表現を抽出する。

ツイート属性判定 「平安時代に東北地方では大きな地震
があった」といった、事実ではあるものの過去の災害
情報であり、現時点では抽出されるべきではない情報
や、「東京が大きな地震に襲われる夢を見たんだ」と
いった冗談ととれる情報から、回答候補を抽出するこ
とを防ぐために、ツイートに対して、過去の情報、冗
談といった属性を付与する。

パターン抽出 以上の解析結果をまとめて、回答候補を検
索するための形式に変換する。例えば、「X 市で大雪
が降る」という文について、係り受け解析結果に基づ
いて、〈A で B が降る〉(ただし、A は X 市、B は大
雪)というパターンと、X 市は Y 県 X 市であること、

「降る」は成立事象であること、冗談や過去の災害ではないこと、予報情報ではないことが回答データベースに登録される。

これらの解析処理を、投稿されたツイートがシステムに入力されるたびに実施する。災害時に投稿される膨大な数のツイートをリアルタイムに解析することが要求されることから、一連の処理を複数の計算機を用いて数百並列で実行し、ツイートを分散させて処理をする。並列実行は、RaSC [14] を用いて実現した。

後藤らのプロトタイプシステムと大きく異なる点は、モダリティ解析、ツイート属性判定、予報表現抽出の新規導入と、パターン抽出のアルゴリズムの改善である。前者は、回答候補を抽出すべきではないツイートを除外することを目的として導入した。後者は、抽出される回答候補の数と種類を増やすために実施した。以下では、これらの要素技術について詳述する。また、3.5 節で言語処理モジュールの解析結果を用いた質問応答の方法について述べる。

3.1 パターン抽出

パターン抽出は、自然言語で書かれた文から、計算機が扱いやすい形として、述語を含む文節と、それと修飾あるいは被修飾関係にある二つの名詞(バイナリ)や、述語を含む文節と、それと修飾・被修飾関係にある一つの名詞(ユナリ)を抽出するモジュールである。バイナリは、「X市で大雪が降る」という文から、係り受け解析結果に基づいて、〈AでBが降る〉(ただし、AはX市、Bは大雪)というパターンが抽出される。バイナリは、前述のように、入力された質問文における疑問詞に相当する部分に入る名詞を回答候補として抽出できる。一方で、ユナリは、同じ文から、〈Aで降る〉(ただし、AはX市)と、〈Bが降る〉(ただし、Bは大雪)の2つのパターンが抽出される^{*5}。これらは、大きく二つの方法で利用される。一つ目は、「どこで大雪が降る」という質問が入力された場合に、地名情報がパターン外だが文中に存在するとき、〈Aが降る〉と組み合わせることで、バイナリと同様の方法で、回答候補であるAに入る名詞であるX市を抽出する。二つ目は、「何が降っている」のように、地域を指定せずに降っているものを聞く質問が入力されたときに、ユナリ〈Bが降る〉から、Bに入る名詞である大雪を回答候補として抽出する。

後藤らのプロトタイプシステムと比較して大きく異なるのは、サ変名詞以外の名詞についても災害表現については、必要に応じてパターンを抽出するように変更したことである。例えば、後藤らのパターン抽出では、「X市で地震」や「X市地震」、「X市で負傷者」といった、述語が省略された文からはパターンを抽出することができない。我々は、負

表 1 パターンを抽出する災害・被害を示唆する名詞

ツイート中で現れる名詞	より一般的名詞
集中豪雨	大雨
土砂降り	大雨
地滑り	土砂災害
強奪	盗難
凍傷	負傷者
靱帯損傷	負傷者
ダニ媒介性脳炎	病人

担・トラブル表現リスト^{*6}を元にして、災害や被害を示唆する約1000種類の名詞を手で整備し、それらの名詞がツイート中に現れたら、「発生する」という述語を補ってパターンを抽出する。これらの名詞の中には「雷火」や「ウイルス性食中毒」といった、あまり一般的ではない名詞も含まれる。「どこで火事が起きていますか」や「どこに病人がいますか」といった質問に対して、「X市で雷火」や「避難所でウイルス性食中毒」といったツイートも検索し、「X市」や「避難所」を回答候補として抽出できるように、「雷火」をより一般的な表現である「火災」に、「ウイルス性食中毒」をより一般的な表現である「病人」に置き換えたパターンも抽出する。整備した名詞の一部を表1に示す。

しかし、「北海道で地震はない」といったツイートから、〈AでBが発生する〉(ただしAは北海道、Bは地震)を抽出してはいけない。そこで、活性・不活性辞書[15]を利用する。名詞に続く述語があり、それが活性の述語(例えば、「が起きる」や「が降る」)である場合は抽出し、不活性の述語(例えば、「はない」や「は治まる」)である場合は抽出を抑制する。述語が続かず、災害を示唆する名詞で文が区切れている「X市で地震。」や「Y県で地震、…」といったツイートからは、無条件でパターンを抽出する。

3.2 モダリティ解析

モダリティ解析は、一般には事象に関する非常に多くの意味的側面を解析するタスクであるが、本稿では、文中の事象の成否を分類するタスクとする。事象は一般に一つ以上の名詞(何が)と述部(どうした)によって表現されると考え、本稿ではその核となる述語を含む文節を〈〉で示す。DISAANAで特に重要なのは、否定されている事象(例えば、「大雪は〈〈降っていない〉〉」)と、疑問・要求の事象(例えば、「X学校で携帯は〈〈充電できますか?〉〉」)である。これらの文節に対して事実、推量、否定、仮定、疑問・要求の5種類のモダリティを表すラベルを手で付与した学習データを用意し、機械学習によってこの問題を解く。上記の例示はすべて述部の1文節内の情報でそのモダリティを判定できたが、「降っていない」といった表現だけでなく、「大雪が〈〈降る〉〉」という予報は当たらなかつ

^{*5} システム上では、AとBはいずれもAで記述されるが、ここでは分かりやすくするために、AとBで表して区別する。

^{*6} <https://alaginrc.nict.go.jp/resources/nict-resource/li-info/li-outline.html#A-3>

た」や「X市が<<浸水している>>というチェーンメールが来た」といったより広い範囲の情報をを用いて複雑な解釈が必要な事例は少なくない。

機械学習に基づく日本語を対象としたモダリティ解析の先行研究 [7] には、こういった広い範囲にわたる事例を分類するための素性が含まれていない。また、広い範囲の情報を直接機械学習モデルに入れようとするとデータスパースネスの問題から膨大な量の学習データが必要となる。そこで、word2vec [16] を利用して、単語 N グラムを分散表現の類似度に基づいてクラスタリングを行い、事象に後続する表現を抽象化する。「チェーンメール」が「虚偽、デマ」といった誤情報を示唆する表現と同一のクラスタに属するようにできれば、「チェーンメールが来た」は、「[[誤情報を示唆する N グラムクラスタの ID]] が来た」に抽象化することができる。そして、学習データ中に同じように抽象化される表現が含まれていれば、正しく解析できることが期待できる。

単語 N グラムのクラスタは、日本語の Wikipedia の全記事 (2015/1/18 版)、2007 年にクロールした 6 億のウェブ文書のうち 0.125%、2015/2/14 から 2015/2/28 までに投稿されたツイートの 3 種類からそれぞれ構築した。これらのテキストはそれぞれ、4.2GB、4.5GB、4.3GB 程度の大きさであり、言語処理モジュールと同じ形態素解析器で解析した。次に、word2vec のツールに含まれる、出現頻度に基づいて隣接する単語を結合する word2phrase を 2 回適用し、1~4 グラムにまとめ上げる。その結果を word2vec に適用し、最後に k-means クラスタリングによってクラスタリングする。パラメータは、word2vec のベクトルの次元数と、クラスタ数を選択する必要がある。ベクトルの次元数 6 種類 (100, 500, 1000, 2000, 5000, 10000) と、クラスタ数 8 種類 (50, 100, 150, 200, 250, 300, 350, 500) の組み合わせを、学習データの 5 分割交差検定によって選択した。結果として、Wikipedia は 250 次元の 10000 クラスタが、ウェブ文書は 350 次元の 10000 クラスタ、ツイートは 300 次元の 10000 クラスタが選択された。クラスタ素性は、3 種類のクラスタそれぞれで単語 N グラムをクラスタ ID に抽象化する。従って、一つの単語 N グラムから、最大 3 つのクラスタ ID 素性が抽出される。抽象化する対象は、解析対象の文節に含まれる述語に後続する最大 7 形態素である。

word2vec と k-means クラスタリングによるクラスタ ID の他に、風間ら [17] によって構築された名詞の意味分類辞書の ID によっても同一条件の述部に後続する最大 7 形態素を抽象化する。この意味分類辞書は、名詞のみを 2000 個に分類したもので、2007 年にクロールされたウェブ文書全てから自動的に作成されたものであるため、前述のクラスタよりも、カバーされる名詞の種類が多くなることが期待できる。

モダリティ解析の機械学習に用いるその他の素性は、以下の通りである。

基本素性 解析対象の文節と、その係り先の文節に含まれる形態素の、表層、原形、品詞の 1~3 グラム。解析対象の文節には、その中心となる述語が含まれているが、それ自身は含めず、後続の表現のみを用いる。これは、事象自体を学習する、言い換えると否定されやすい事象の傾向を学習することを防ぐためである。

後続形態素 述語に後続する最大 7 形態素の表層をそのまま素性とする。基本素性と同様に、述語そのものは含めない。

否定表現 否定を示唆する表現が解析対象文節の周辺に存在する場合、否定されている可能性が高いと考えた。そこで、以下に示す、人手で整備した 33 個の誤情報を示唆する形態素について、それらが述語の前後のいずれかに現れるかと、現れる場合その形態素間の距離を、1~10 形態素または 11 形態素以上の 11 種類の素性とする。

デマ、でま、ガセ、ガセネタ、がせ、ネタ、風説、流言、流言飛語、流言蜚語、誤報、誤情報、誤解、嘘、うそ、ウソ、偽る、偽り、捏造、ねつ造、虚偽、間違、間違、出任せ、でまかせ、誤る、誤り、虚構、違う、違い、チェーンメール、チェンメ、ちえんめ

評価極性 「噴火が<<起きている>>無知な人が言っている」のように、「無知」というネガティブな単語によって否定される事象を認識するために、単語の評価極性を素性にする。基本素性を抽出した文節について、評価極性がポジティブまたはネガティブな単語が含まれる場合、それらを素性とする。評価極性は、人手で整備した辞書 [18] に基づいて判定する。

学習に用いるデータは、東日本大震災時のツイートデータからランダムサンプルした 96,824 事例、および人手で作文した否定表現を伴う文の 4,048 事例である。各事例は、述語を含む文節に対してそのモダリティを表現するラベルとして事実、推量、否定、仮定、疑問・要求の 5 種類のいずれかが付与されたものである。本稿ではサポートベクトルマシン (SVM) [19] の実装の一つである LIBSVM [20] を用いて、対象となる述語を含む文節のラベルを推定する。

学習データの 5 分割交差検定の結果を表 2 に示す。いずれも再現率がやや低いものの、否定、疑問・要求は、実用するのに十分な適合率が得られた。後述の予報表現抽出では、推量のモダリティを手がかりにすることが考えられるが、推量は他のモダリティと比べて適合率が低いため、本稿では利用しない。

3.3 予報表現抽出

「X市は今晚雪になりそう」や「予報では、Y県は明日には台風が上陸するらしい」といった災害の予測や、いわゆる

表 2 モダリティ解析の 5 分割交差検定による評価結果

モダリティ	適合率	再現率	F 値
事実	0.735 (39053/53117)	0.874 (39053/44672)	0.799
推量	0.690 (4137/5999)	0.484 (4137/8546)	0.569
否定	0.833 (9848/11827)	0.731 (9848/13480)	0.778
仮定	0.611 (10785/17651)	0.561 (10785/19239)	0.585
疑問・要求	0.7908 (9710/12278)	0.650 (9710/14935)	0.714
マクロ平均	0.732	0.659	0.694

天気予報に基づいた情報発信は、現時点では起きていない情報であるため、他の回答候補とは区別して表示する。それらを判定するために人手でキーワードを整備した。キーワードは、「来年、明日、将来」といった未来を示唆する単語だけでなく、「見込み、かも、恐れ」といった予測・予報を示唆する単語が含まれる。また、ツイート中に含まれる「7月8日」や「7/5」といった日付情報について、年月日の3値で正規化し、ツイートの投稿日より未来であることを判定する。日付の表現には「8日」のように年月の情報を省略したものもあるが、その場合は、投稿日と同じ年月を補完する。「来月の8日」といった表記の場合は、投稿日の月を一つ進めた月の「8日」とであると判定する。未来を示唆するキーワードや日付情報が含まれる場合、そのツイートに含まれる情報は未来を示唆する情報として取り扱われ、ユーザには予報情報であることを明記して提示する。

3.4 ツイート属性判定

Twitter には、「京都で大雪が降る設定のドラマ」といったフィクションを題材にしたツイートや、「東京では昔大きな地震があった」といった過去の事実を題材にしたツイートが投稿されることがある。こういったツイートから「京都」や「東京」を災害の起きた場所として回答候補とするのを避けるため、ツイート単位で冗談、過去、広告といった属性を付与するのがツイート属性判定である。いずれも人手で整備したキーワードによって判定する。冗談には、いわゆる冗談だけでなくフィクションや、夢の内容、絵文字や顔文字を使ったツイートなどが含まれる。これらを単語リストや正規表現によって判定する。過去は、過去の災害や過去の情報であることを示唆する表現によって判定する。過去の災害やそれに関連する情報は、上位下位関係抽出ツール^{*7}により Wikipedia から抽出した上位下位概念辞書から、「地震」などの下位概念を列挙して人手で精査した。過去の情報であることを示唆する表現には、「昔」「～年前」といったものがあるが、これらは人手で整備した。広告は、「～した人 RT」^{*8} や「【無料】」といった集客を目的とした表現や、会社名などが手がかりとなる。会社名は、それ自体は被害のあった場所である可能性があるた

^{*7} <https://alaginrc.nict.go.jp/hyponymy/>

^{*8} 投稿内容に同意する人に対してリツイートすることを求めるツイート。

め、URL または電話番号が同一ツイート内に含まれる場合のみ、広告であると判定する。これらの属性判定のためのキーワードとして現時点で 46,996 単語が整備されており、今後も拡張していく。

キーワードの他に、時間表現についても過去の情報であることを判定する。予報表現抽出によって列挙される、日付を示す表現の正規化された結果を利用する。ツイートは投稿時間が明らかであるため、それよりも3日以上前の日付が含まれる場合に、過去と判定する。例えば、7月8日に投稿されたツイートにおいて、「7月5日にX市に上陸した台風」といった記述がある場合、過去の情報であると判定する。

冗談や過去と判定されたツイートは、災害関連の用語も含まれる場合に限って、ツイートから抽出された回答候補を表示しないように、DISAANA の表示設定によって設定することができる。これは、災害情報の検索結果に冗談や過去の情報が混在することは大きな問題だが、その他の災害との関連性が低い回答候補については、冗談や過去の情報であっても表示されて構わないという判断によるものである。

3.5 質問応答

質問応答モジュールは、言語処理モジュールで構築された回答データベースから、質問の回答候補を抽出する。具体的には、入力された質問を平叙文に変形し、疑問代名詞に入りうる単語を回答データベースから検索して、回答候補として出力する。例えば、「どこで大雪が降っていますか」という質問は、平叙文の「どこで大雪が降る」に変形される。疑問代名詞は「どこ」であるため、〈A で B が降る〉というパターンを B が「大雪」であるという制約とともに抽出し、A に当てはまる地名を回答データベースから検索する。その際に、パターンと同義あるいはパターンを含意するパターンについても検索する [21], [22], [23]。この例では、〈A で B が降り続く〉や〈A で B が降り積もる〉といったパターンでも検索する。回答データベースに、「X市で大雪が降る」から抽出されたパターン〈A で B が降る〉(ただし A は X 市, B は大雪)や、「Y町で大雪が降り続けている」から抽出されたパターン〈A で B が降り続く〉(ただし A は Y 町, B は大雪)があるとき、パターンと B を照合することで、「X市」「Y町」を回答候補として抽出

する。

ユナリは、〈A が起きている〉のように名詞が一つだけのパターンだが、質問が「何が発生しているか」や「何が止まっているか」などのように、疑問代名詞と一つの述語からなる場合は、前述と同様の方法で回答候補を抽出することができる。例えば、質問「何が発生しているか」からは、〈A が発生している〉に変形され、「地震が起きている」から抽出される〈A が起きる〉(ただし A は地震)や「台風が発生した」から抽出される〈A が発生する〉(ただし A は台風)などから、「地震」「台風」を回答候補として抽出する。

後藤らのプロトタイプシステムと比べて、質問応答のアルゴリズムに大きな改善はないが、回答候補の抽出元の情報に、モダリティ解析結果などが付加されているため、それを考慮して回答候補を抽出する点が異なる。モダリティ解析では、否定と判定された事象からは回答候補を抽出しない。例えば、「X 市では雪は降っていない」というツイートがあるとき、「降る」は否定事象と判定されるので、「どこで雪が降っているか」という質問に対して「X 市」は回答候補として抽出されなくなる。また、疑問・要求と判定された事象が含まれるツイートからは回答候補を抽出しない。例えば、「Y 市でも雪が降るかな?」というツイートがあるとき、同様に「Y 市」を回答候補として抽出されない。

4. 東日本大震災データにおける性能評価

DISAANA の質問応答モードの性能を、東日本大震災時のツイートデータを用いて評価する。利用したツイートデータは、2011 年 3 月 9 日から 2011 年 4 月 4 日までのツイートデータ((株)ホットリンク提供)のうち、#tsunami や #jishin といった震災関連のハッシュタグや災害関連のキーワードを含み、さらにツイートの文字列比較によって同一と判断されたものは、時刻が最も古いもののみを残す^{*9} という処理を施した約 5,400 万件を用いた。後藤ら [2] でも、同じデータを用いて性能評価を実施している。本稿では、後藤らと同じ評価基準で評価を行い、比較する。

評価対象となる質問は、川田ら [24] が構築した 300 個の質問と、17,524 個の正答を用いる。300 個の質問それぞれに対して DISAANA で回答候補の集合を得る。再現率は、出力結果が正答に含まれるかどうかによって計算する。このとき、正答に含まれるかの判定は、システムが出力した回答候補が、正答の文字列に含まれるかによって判断する。正答は、各質問に対して最大 1000 件の関連ツイートを手で見て、抽出することで構築しているため、システムの出力が正解でも正答に含まれていない場合がある。後藤らと同様に、これらは正解として取り扱わない。適合率は、出力結果である質問・回答候補ペアからランダムに 250 ペ

表 3 東日本大震災データにおける質問応答モードの評価結果

	適合率		再現率	
後藤ら	0.608	(152/250)	0.519	(9,099/17,524)
本稿	0.448	(112/250)	0.707	(12,382/17,524)

アを選択して、人手で正解判定を行う。

後藤らのプロトタイプシステムとの大きな違いは、パターン抽出を含むシステム全体の改善と、モダリティ解析の導入である。また、後藤らは、質問応答の動作について、時間および回答候補数の上限を設けることなく回答候補を抽出したが、その場合は回答候補が出力されるまで 1 分程度かかることがある。我々は、より現実的な状況を想定して、10 秒の上限時間を設けた。なお、公開している DISAANA の時間制限はより短く 1 秒としているが、災害時などで再現率が優先される場合は、制限時間を緩めることを考えている。なお、ツイート属性判定および予報表現抽出は、後藤らとの違いが大きくなりすぎるため、本評価では利用しない。また、モダリティ解析の学習データは、本評価データとの重なりがあるため、モダリティ解析による否定回答の抽出の抑制は、5 節の評価実験に比べて、やや有利な設定となっている。

評価結果を表 3 に示す。パターン抽出によって回答候補の抽出数を増やすことができるため、再現率は大きく向上している。災害時には、災害情報の抽出漏れは致命的であるため、適合率をある程度犠牲にしても再現率を重視すべきである。その観点からすると、0.707 という再現率は、本システムが実用的なレベルに到達していると考えて良いだろう。適合率について、モダリティ解析は、否定情報からの回答候補抽出を抑制できるため、適合率の向上に寄与することが期待できる。しかしながら、適合率は後藤らより低い結果となった。この原因の一つは、ランダムサンプルした 250 の質問・回答候補ペアに含まれる質問の種類である。後藤らの方には、比較的正しく回答しやすい「何が発生していますか」という質問と回答候補のペアが 82 件含まれていたのに対して、本稿では 2 件であった。これは、本稿ではパターン抽出などの改善により、質問に対する回答候補の数が増えていることと、前述の質問応答の動作上限時間の違いが原因である。

適合率の評価に用いた 250 質問・回答候補ペアのうち正解であった 112 事例のうち、パターンによって回答候補が抽出されたのは 48 事例、周辺キーワードによって回答候補が抽出されたのは 64 事例であった。一方で、138 の誤り事例については、それぞれ 36 事例、102 事例であった。パターンによる回答候補抽出の誤りについて誤り分析を行った。

「どの X が～」質問への誤り 12 事例 「どの病院が開いていますか」「どんな危険がありますか」といった、疑問副詞と名詞の組み合わせからなる質問文を解析する

*9 リツイートを擬似的に削除するため。

際、「どの病院」「どんな危険」を抽象化して、それぞれ〈A が開いている〉、〈A があります〉のパターンにマッチする A が回答候補として抽出される。本来であれば、それらの候補に対して「病院」や「危険」の下位概念であるかを判定する必要があるが、「病院」は表層情報からある程度判定できることが期待できる一方で、「危険」は「～シーベルト」のような数値表現で判断される場合もあり、その下位概念を表層情報から判定することは難しい。特に災害に特化して上位下位概念を整備する必要がある。

パターン言い換えの誤り 9 事例 言い換えたパターンが誤っていた事例で、〈A を提供〉と〈A を提案〉は言い換え可能となっているため、質問「何を提供していますか？」に対して、「X を提案している」から X を回答候補として抽出しているが、誤りである。言い換えパターンは自動獲得されているため、言い換え不可能なパターンを人手で除外していくことが必要である。

質問の解析誤り 6 件 「流行っている病気は何ですか」や「ボランティアの作業は何になりますか」といった、疑問代名詞が文の後ろの方に現れる質問は、「どの X が～」質問への誤りと同様に、抽出された回答候補が「病気」や「作業」の下位概念であるかを精査する必要がある。

モダリティ解析の誤り 3 事例 「どこで安否確認ができますか」といった質問に対して、安否確認ができないとされる場所を回答候補としてしまった、モダリティ解析の誤り。誤った 3 事例を示す。それぞれ、下線部が抽出された回答候補である。

- 質問「どこで安否確認ができますか」に対する「安否確認のために 高速道路 を利用しようという考えだけは絶対に避けて欲しい。」
- 質問「何が発生していますか」に対する「地震で揺れた。最初 目眩 が起こったかと思った。」
- 質問「どこに遺体はありましたか」に対する「藤沢市が江ノ島に鳥葬の施設を作って遺体を受け入れようとしたが、周辺の鎌倉市と 茅ヶ崎市 の猛反対で中止になったらいい。」

それぞれ、述語「利用する」「起こる」「受け入れる」が後続の表現によって否定されているが、正しく解析することができなかった。このような複雑な表現は、大量のデータからも効率良く集めることができないため、誤った事例を学習データに加えていくことが有効である。

その他 6 事例 その他の誤り事例は、係り受け解析の誤りによるものや、地名を聞く質問に対して、正答ではない地名を出力した誤りなどである。

後藤らは、誤り分析において、単純な否定文や疑問文から回答候補を抽出してしまう誤りが散見されたと報告して

いるが、モダリティ解析を導入することによって、それらの誤りを大幅に減らすことができた。その結果、質問文の解析や、パターン言い換えに改善の余地が大きいことが明らかとなった。

5. 災害時の性能評価

東日本大震災以外の災害における DISAANA の性能を評価するため、長野県神城断層地震（以下、長野地震）の期間に投稿された 31,501,224 ツイート、2015 年台風 11 号（以下、台風）の期間に投稿された 22,254,542 ツイート、2014 年 12 月の大雪（以下、大雪）の期間に投稿された 73,815,442 ツイート、2015 年 6 月の東海道新幹線火災事件（以下、新幹線火災）の期間に投稿された 20,637,966 ツイートを用いて同様の評価を行う。

4 節との違いは、ツイートの属性判定および予報表現抽出を用いていることである。また、モダリティ解析の学習データおよびクラスタ素性の生成元となる文書は、本評価データとの重なりは全くないため、より厳密な評価となる。まず、川田ら [24] と同様に、各災害について、質問と正答のペアを構築する。質問は、一般ユーザと同等程度に DISAANA の使い方を知っているアノテータに対して、それぞれの災害が起きたときに調べたいことを列挙して作成した。表 4 に各災害のツイート投稿期間と、質問の一覧を示す。

正解データは、各質問につき最大 1000 件の関連ツイートから、正答をアノテーションすることによって構築した。川田ら [24] が対象としたデータは、東日本大震災関連の情報を優先してサンプリングしていたことと、当時東日本大震災関連のツイートが大半を占めていたことから、地震関連ではない正答が見つかることはまれであった。一方で、本データは、対象期間中の全ツイートの 10% が対象となるため、災害とは無関係だが、正答と判定する場合がある。例えば、「何が不足している」に対して「睡眠が不足」というツイートがあるとき、「睡眠」も正答とする。災害関連であるかを真に判断することは難しいこと、DISAANA では災害関連かどうかを区別する仕組みはないことから、このような基準とした。アノテーションは 3 名で行い、各質問について、1 名が全ての関連ツイートを見て正答のアノテーションを行った。正答が 1 つも見つけられなかった質問が 5 問あった。

表 5 に質問応答モードの評価結果を示す。適合率および再現率は、4 節と同様に評価した結果は表左上である。この結果に対して、冗談や過去と判定されたツイートから回答候補を抽出しない場合の評価結果を表右上に示す。同様に、予報と判定された回答候補を抽出しない場合の評価結果を表左下に示す。表右下には、属性および予報の両方を適用した結果を示す。

再現率は、表左上が最も高くなり、フィルタリングによ

表 4 4つの災害に対する質問

大雪 (2014/12/15 - 12/24)		正答数	新幹線火災 (2015/6/30 - 7/2)		正答数
どこでバスが運休		169	新幹線で何が燃えたか		4
避難所はどこ		91	新幹線内のどこで自殺をしたか		9
どこで雪崩が起きる		66	どこで新幹線事故が起きたか		29
どこが通行止めになっている		758	何を使って新幹線に火をつけたか		28
どこで大雪が降っている		58	自殺者はどこから新幹線に乗ったか		0
どこで事故が起きている		8	自殺者はどこまでの切符を買ったか		2
どこで電車が止まっている		0	どこで新幹線はとまっているか		152
どこが危険		12	何が新幹線の再開を知らせてくれるか		5
何が不足している		410	どこで火災事件の遅延証明を受けられるか		0
農作物で影響を受けるものは何		10	死体を降ろしたのはどこですか		8
除雪に必要な道具・機材は何		7			
確認されている人的被害は何		1			
ライフラインで影響が出ているものは何		19			
スタッドレスの何に問題がある		50			
台風 (2015/7/16 - 7/18)		正答数	地震 (2014/11/22 - 11/25)		正答数
どこの店が開いているか		20	どこに津波が到達する		4
交通機関はどこが動いていますか		65	どこが停電している		98
避難所はどこ		16	どこで通行止めが起きている		177
ペットが同伴できる避難所はどこ		0	震源地はどこ		62
台風では何に注意する		253	避難所はどこ		37
どこの病院があいているか		0	どこで断水している		84
どこに台風が上陸する		78	何が不足している		221
台風による影響は何		375	ボランティアに出来ることは何		93
どこで警報がでている		16	地震による被害で多いのは何		10
どこで大雨が降っている		138	何を持って避難する		7
どこで川が増水している		138			

表 5 4つの災害における質問応答モードの評価結果

災害	属性・予報によるフィルタ無し				属性によりフィルタリング			
	回答候補数	適合率	再現率		回答候補数	適合率	再現率	
新幹線火災	2,328	0.136 (34/250)	0.768 (182/237)		1,386	0.113 (19/168)	0.730 (173/237)	
長野地震	13,056	0.216 (54/250)	0.619 (491/793)		10,394	0.191 (44/230)	0.575 (456/793)	
台風	8,716	0.364 (91/250)	0.387 (425/1099)		4,978	0.321 (60/187)	0.268 (295/1099)	
大雪	10,612	0.616 (154/250)	0.576 (955/1659)		8,627	0.586 (106/181)	0.554 (919/1659)	
災害	予報によりフィルタリング				属性・予報によりフィルタリング			
	回答候補数	適合率	再現率		回答候補数	適合率	再現率	
新幹線火災	1,520	0.113 (16/141)	0.629 (149/237)		908	0.085 (8/94)	0.502 (119/237)	
長野地震	11,627	0.238 (46/193)	0.453 (359/793)		9,304	0.208 (37/178)	0.333 (264/793)	
台風	6,522	0.400 (58/145)	0.298 (327/1099)		4,072	0.367 (44/120)	0.177 (195/1099)	
大雪	7,994	0.610 (130/213)	0.380 (631/1659)		6,499	0.587 (91/155)	0.299 (496/1659)	

て再現率を維持したまま適合率が上がることが期待されたが、属性によるフィルタリングは正解事例を誤ってフィルタリングしてしまうことが多かった。誤り分析をしたところ、顔文字や絵文字を伴うが正解を含む事例が目立った。特に泣き顔や困り顔の顔文字は、災害の状況を伝える際に用いられる傾向が見られるため、顔文字の感情推定 [25] を取り入れることが考えられる。その他の事例として、他者への返信（リプライ）は、返信先のツイートも読まないという理解できない可能性が高いためフィルタリングの対象としていたが、返信文だけで内容を理解でき、かつ災害情報が

含まれている場合があった。長文の返信はフィルタリングの対象から除外することが考えられる。予報については、長野地震および台風において、適合率が改善された。新幹線火災については適合率が悪化したが、前例のない事故であったため、結果的に事実であったが、投稿時には憶測で書かれたツイートが予報と判定されたためである。

属性や予報によるフィルタリングは、DISAANA 上では、表示設定によって有効/無効を切り替えることができる。デフォルトでは、いずれも有効になっている。^{*10}DISAANA

*10 予報については Twitter に認定された公式アカウントのツイ

には、これらの設定を無効にすることで、冗談や予報が大量に表示される可能性があるものの、検索結果が増えることを明記している。従って、実用上は、まず両フィルタリングが有効な状態で災害情報を把握し、より多くの情報が必要な場合に、これらのフィルタリングを無効にすることになる。

次に、各災害について、質問応答の誤り分析を行う。分析の対象は、パターンによって抽出された回答候補のうち、属性、予報のいずれでもフィルタリングされなかったものである。

新幹線火災 分析対象の誤りは 16 事例である。パターン言い換えが正しくないといった典型的な誤り以外で重要な誤りとして、5 事例が「何を使って火をつける」に関する誤りであった。この質問の回答候補は、「X で火を付けた」といった表現から抽出することを期待しているが、特にこの災害では「油のようなものを被って火を付けた」という記述が大半であったため、手段のデ格による対応が取れなかった。また、デ格は場所を表す場合もあるため、「先頭車両で火を付けた」といった場所のデ格から「先頭車両」などの場所を出力した誤りもあった。

長野地震 分析対象の誤りは 120 事例である。この災害で誤った質問は、「何を持って避難する」と「何が不足している」の 2 種類だけであった。「何を持って避難する」の誤りは、79 事例あり、その全てが質問文の解析誤りに起因していた。この質問は、二つの述語を含むため、パターンによる回答が難しい。我々のアルゴリズムでは、パターン「<<A を持つ>>」にマッチし、かつ周辺に「避難」が現れる「A」を回答候補として抽出する。しかし、この方法によって抽出された回答候補数が少ない場合は、「避難」が現れない場合でも回答候補として抽出する。本評価では、「<<A を持つ>>」、「<<A を得る>>」といったパターンにマッチする「A」が回答候補として得られたが、いずれも誤りであった。周辺語の種類によってはその有無によって大きく質問の意図が代わってしまう場合があるため、今後は質問文解析の改善が必須である。「何が不足している」の誤りは、41 事例あり、いずれもパターン言い換えに起因する誤りであった。

台風 分析対象の誤りは 5 事例である。そのうち 4 件はパターン言い換えに起因する誤りである。残りの 1 件は複雑な否定であった。

大雪 分析対象の誤りは 48 事例であり、そのうち 22 件がパターン言い換えに起因する誤りであった。次に、7 件は冗談と判定されるべきツイートであった。7 件全てが、質問「何が不足している」に対する誤回答であっ

た。災害関連以外で、ゲームのアイテムが不足しているといった回答候補であった。次に、5 件は回答候補の地名選択の誤りであった。質問「どこで事故が起きている」に対して、「近鉄山田線は、近鉄大阪仙台で発生した人身事故の影響で一部遅れが出ています。」というツイートがあるとき、回答候補として「近鉄山田線」が選択された。事故に限らず、災害には、発生場所とその影響を受ける場所など、2 力以上の地名が現れることは珍しくない。複数の地名が回答候補となる場合、その選択には改善の余地がある。残りの誤りは、アノテーション誤りや、係り受け解析誤りに起因するものなどであった。

6. 実証実験

我々は、DISAANA の一般の利用者が利用する場合の問題点や、自治体等で実際の災害対応を前提として用いる場合の問題点を明らかにするために宮崎県において実証実験を行った。

宮崎県は、台風銀座と呼ばれるように地理的・自然的条件等から台風の接近に伴う風水害や土砂災害が発生しやすいところである。台風や集中豪雨による大きな土砂災害も近年では 1997 年、2004 年、2005 年と頻繁に発生し、県内各地に大きな爪痕を残している。また、南海トラフによる巨大地震の発生も懸念される。そのため県全体として防災意識が高く、防災士の育成に力を入れている。

これらの防災士や、防災意識の高い一般市民が SNS 上に災害関連情報を提供したとして、それらを DISAANA で分析し、自治体における災害対応の意志決定に有用な情報を DISAANA が提供できるかどうかについて机上訓練形式の防災訓練を通じた実験により検証した。

実証実験の概要を図 4 に示す。実験では、コントローラと呼ばれる人員を配置し、訓練開始後の時間経過にあわせて被験者および市役所担当者に被害状況等を与える形式をとる。すなわち、シナリオは存在するが、被験者および市役所担当者にそれは知らされておらず、訓練開始後に逐一被害状況が伝達される形式である。防災士を含む一般市民役の被験者は、設定された状況下において想定される被害状況を SNS を模した掲示板に自由に書き込む。そして、DISAANA がそれらの書き込みを解析する。市役所担当者は、コントローラより与えられる従来の情報チャネルで得られる被害情報に加えて DISAANA を用いて得られる SNS からの情報もあわせて、避難、救援等の意志決定を行う。また、被験者も、DISAANA を利用して自らの書き込みが期待するように処理されているかどうかを確認した。

実証実験は 2015 年 1 月 18 日に延岡市で、2015 年 2 月 7 日に宮崎市で実施し、延べ 115 名の被験者、15 名の自治体関係者、2 名のコントローラによる 5 時間半の訓練を通して計 4,400 件以上の書き込みと 1,760 回を超える

トについては表示される。

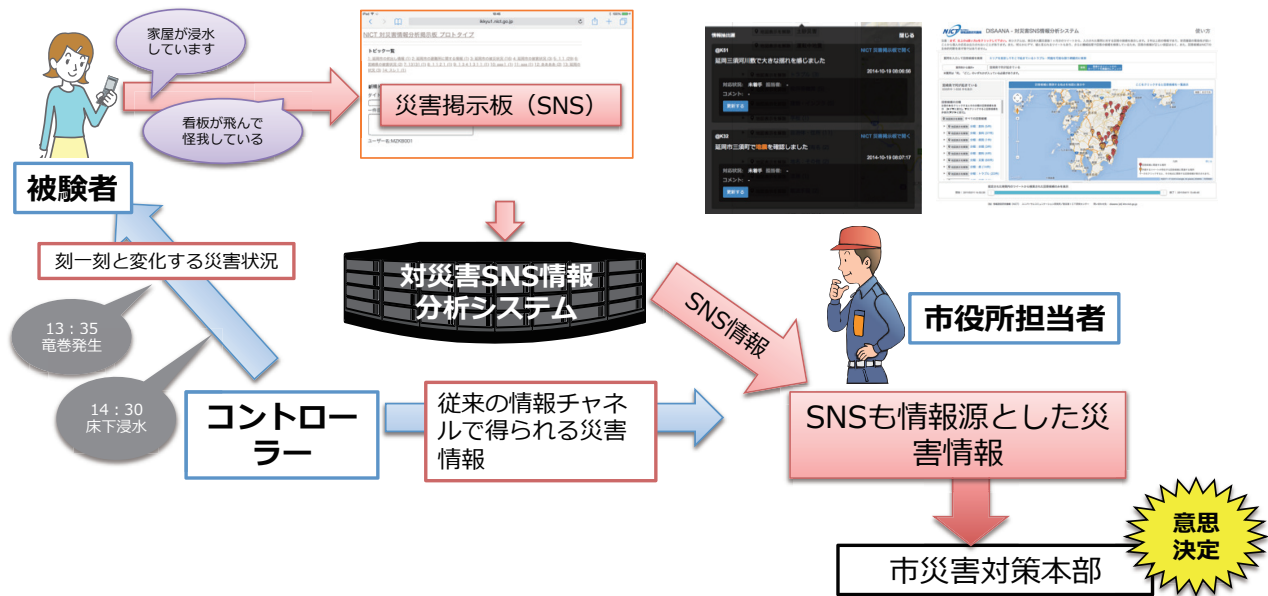


図 4 DISAANA を用いた実験実験概要

DISAANA での質問応答を得ることができた。被験者は、延岡市、宮崎市在住の防災士資格を持つものを含む 18 歳以上の一般市民である。被験者は、NICT にて用意したタブレット端末もしくはノート PC を用いて掲示板への書き込み、DISAANA の利用を行った。災害対策本部の市役所関係者は、全てノート PC を利用した。被験者全員を会議室に集め、2 時間程度操作方法等の説明を行ってから机上訓練（実験）を実施した。災害対策本部の市役所担当者にも同じく 2 時間程度の説明を行い、被験者とは別の部屋に設置した災害対策本部にて実験を行った。

災害対策本部で用いた DISAANA は、被験者が利用した通常のものとは異なり、自治体向けの機能を組み込んである。それは、書き込みに対して、(1) 着手、(2) 未着手、(3) 解決、(4) デマと認定という 4 状態を設定できるようになっており、更に特定の状態だけに絞り込んで書き込みを表示できる。また、状態の設定に加えてコメントを書き込むことができ、書き込まれたコメントは掲示板に災害対策本部からの書き込みとして反映される。なお、この自治体向けの機能は、開発中であり、現在試験公開している DISAANA には組み込まれていない。

実験後に、災害対策本部で災害対応にあたった市役所担当者と一般市民役の被験者にアンケートを実施した。被験者側からは、DISAANA に対する期待や、様々な改善すべき点の指摘があった。市役所関係者からも災害対応に役立つという好評を得る一方で、災害対応を実務とする者の視点から非常に貴重な改善点の指摘、コメントがあった。その一部を以下に示す。

被験者からのコメント・要望

- スマホから使う際には GPS と連動して欲しい (20 代女性)

- 登録されていない地名があった (60 代男性)
- 入力が災害毎にカテゴリー分けされ、エリアや人員などの記入で済むと楽 (40 代男性)
- 災害に限らず、防犯、交通情報、不審者情報にも使えるのでは (60 代男性)
- 災害時には是非利用したい (10 代女性)
- 慣れてしまうと意外と簡単だった (40 代女性)

市役所関係者からのコメント・要望

- 同一の災害事象に関する書き込みに一括して返信できると良い
- 結果が負傷者・火災・倒壊等の区分で分けられていると良い
- 結果に対して単語検索ができると良い
- 写真付きの情報は、状況把握がよりしやすくなるので良い
- 意志決定上、写真で把握できる場合はすぐに動ける
- 結果の表示の際に最新の書き込みが上に来るようにして欲しい

要望のいくつかは、現在試験公開している DISAANA に反映させることができた。また、実験を通して情報の信憑性をいかに捉えるべきかという課題も示された。

7. おわりに

本研究では、災害時に Twitter 上に投稿される情報を効率的に検索できるシステムとして DISAANA を開発し、誰でも利用可能な Web アプリケーションとして試験公開した。本システムは、後藤らが東日本大震災を契機に開発したシステムを改善したものであり、モダリティ解析、ツイート属性判定、予報表現抽出が新たに導入されている。質問応答モードの性能を、東日本大震災のツイートデータ

について、後藤らと比較したところ、適合率は下がったものの、災害時に重要な再現率は大幅に向上した。次に、東日本大震災以外に4種類の災害データに対しても評価した結果、適合率は限定的であるものの、再現率は実用レベルにあることが分かった。後藤らが大きな問題として報告していた、否定事象からの回答候補の誤抽出は、モダリティ解析を新たに導入することでほぼ解消された。誤り分析によって、パターン言い換えの適用や質問文解析に改善の余地があることが明らかとなった。今後は、これらの改善を行っていく。

謝辞

本研究で利用しているデータは、株式会社ホットリンク様よりご提供頂きました。ここに記して感謝いたします。

参考文献

- [1] Varga, I., Sano, M., Torisawa, K., Hashimoto, C., Ohtake, K., Kawai, T., Oh, J.-H. and De Saeger, S.: Aid is Out There: Looking for Help from Tweets during a Large Scale Disaster, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1619–1629 (2013).
- [2] 後藤 淳, 大竹清敬, Stijn, D. S., 橋本 力, Julien, K., 川田拓也, 鳥澤健太郎: 質問応答に基づく対災害情報分析システム, *自然言語処理*, Vol. 20, No. 3, pp. 367–404 (2013).
- [3] Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., Lally, A., Murdock, J. W., Nyberg, E., Prager, J., Schlaefler, N. and Welty, C.: Building Watson: An Overview of the DeepQA Project, *AI Magazine*, Vol. 31, No. 3, pp. 59–79 (2010).
- [4] 吉村 健: シャベってコンシェルと言語処理, *IPSSJ SIG Technical Report Vol. 2012-SLP-93, No. 4*, pp. 1–6 (2012).
- [5] 乾健太郎, 松吉 俊: 言語情報編集のための広義モダリティ解析に向けて, *Japio year book*, pp. 128–133 (2009).
- [6] 松吉 俊, 江口 萌, 佐尾ちとせ, 村上浩司, 乾健太郎, 松本裕治: テキスト情報分析のための判断情報アノテーション, *電子情報通信学会論文誌. D, 情報・システム*, Vol. 93, No. 6, pp. 705–713 (2010).
- [7] 江口 萌, 松吉 俊, 佐尾ちとせ, 乾健太郎, 松本裕治: モダリティ, 真偽情報, 価値情報を統合した拡張モダリティ, *言語処理学会第16回年次大会 発表論文集*, pp. 852–855 (2010).
- [8] Saurí, R. and Pustejovsky, J.: Determining Modality and Factuality for Textual Entailment, *Proceedings of the First IEEE International Conference on Semantic Computing* (2007).
- [9] Pustejovsky, J., Verhagen, M., Saurí, R., Littman, J., Gaizauskas, R., Katz, G., Mani, I., Knippen, R. and Setzer, A.: *TimeBank 1.2* (2006).
- [10] Saurí, R. and Pustejovsky, J.: FactBank: a corpus annotated with event factuality, *Language Resources and Evaluation*, Vol. 43, No. 3, pp. 227–268 (online), DOI: 10.1007/s10579-009-9089-9 (2009).
- [11] Kurohashi, S., Nakamura, T., Matsumoto, Y. and Nagao, M.: Improvements of Japanese morphological analyzer JUMAN, *Proceedings of The International Workshop on Sharable Natural Language*, pp. 22–28 (1994).
- [12] Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis., *EMNLP*, Vol. 4, pp. 230–237 (2004).
- [13] Yoshinaga, N. and Kitsuregawa, M.: Polynomial to Linear: Efficient Classification with Conjunctive Features, *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 1542–1551 (2009).
- [14] 田中正弘, 大竹清敬, 鳥澤健太郎, 田浦健次郎: RaSC: 高速なストリーム通信をサポートする言語処理プログラムの高速化・高並列化ミドルウェア, *言語処理学会第20回年次大会*, pp. 705–708 (2014).
- [15] Hashimoto, C., Torisawa, K., De Saeger, S., Oh, J.-H. and Kazama, J.: Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 619–630 (2012).
- [16] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J.: Distributed Representations of Words and Phrases and their Compositionality, *Advances in Neural Information Processing Systems 26* (Burgess, C., Bottou, L., Welling, M., Ghahramani, Z. and Weinberger, K., eds.), Curran Associates, Inc., pp. 3111–3119 (2013).
- [17] Kazama, J. and Torisawa, K.: Inducing Gazetteers for Named Entity Recognition by Large-Scale Clustering of Dependency Relations, *Proceedings of ACL-08: HLT*, Columbus, Ohio, pp. 407–415 (2008).
- [18] Oh, J.-H., Torisawa, K., Hashimoto, C., Kawada, T., De Saeger, S., Kazama, J. and Wang, Y.: Why question answering using sentiment analysis and word classes, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, Association for Computational Linguistics, pp. 368–378 (2012).
- [19] Vapnik, V.: *The nature of statistical learning theory*, Springer Science & Business Media (2000).
- [20] Chang, C.-C. and Lin, C.-J.: LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, pp. 27:1–27:27 (2011).
- [21] Kloetzer, J., Torisawa, K., Hashimoto, C. and Oh, J.-H.: Large-Scale Acquisition of Entailment Pattern Pairs by Exploiting Transitivity, *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2015) (to appear)* (2015).
- [22] Sano, M., Torisawa, K., Kloetzer, J., Hashimoto, C., Varga, I. and Oh, J.-H.: Million-scale Derivation of Semantic Relations from a Manually Constructed Predicate Taxonomy, *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 1423–1434 (2014).
- [23] 川田拓也, Kloetzer, J., 鳥澤健太郎: 時制・モダリティを考慮した含意パターンペアの生成, *言語処理学会第20回年次大会*, pp. 562–565 (2014).
- [24] 川田拓也, 大竹清敬, 後藤 淳, 鳥澤健太郎: 災害対応質問応答システム構築に向けた質問・回答コーパスの構築, *言語処理学会第19回年次大会*, pp. 480–483 (2013).
- [25] 奥村紀之, 大西智佳: 顔文字に含まれる感情成分の分析と感情極性辞書の構築, *言語処理学会第20回年次大会*, pp. 872–875 (2014).