研　究　報　告　書

「ベイジアンテレビ：取材・配信・編集を自動化した緊急情報メディア」

研究タイプ：通常型
研究期間：　平成 21 年 10 月～平成 25 年 3 月
研 究 者：　北本　朝展

1.　研究のねらい

情報発信源の爆発的な増大に伴って、インターネットは世界の動きをリアルタイムに伝えるメディアへと発展しつつある。しかし情報発信源が大幅に増加した結果、たとえ高性能な検索エンジンが存在しても、利用者が情報を引き出す（プル）積極的な行為をしなければ重要な情報にたどりつけない問題が生じている。特にクライシス時においては、短い時間で知らない分野に適切なキーワードを設定して効率的に検索せよ、というのは現実的ではない。

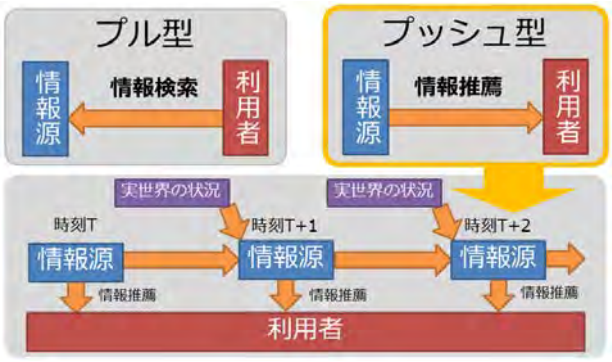

図1　ベイジアンテレビの概念図

そこで本研究では、特にリアルタイム性が重要となるクライシス時の情報爆発を念頭に、重要な情報が利用者に届くようなプッシュ型メディアの構築を目標とし、その概念モデルを取材・配信・編集を自動化した「ベイジアンテレビ」として提案する（図 1）。具体的な研究課題は、実世界の情報を取材・編集するエージェントの構築、データを実世界のエンティティと関連付ける解析ソフトウェアの開発、実世界の情報を取り入れつつ個別の利用者に情報を推薦するデータ同化アルゴリズム、プッシュ的・テレビ的なコンセプトに基づくデータ可視化インタフェースのデザインなどである。研究の対象とするデータは、当初は台風関連データを中心とする予定だったが、研究開始後に東日本大震災が発生したため急きょ震災関連データも研究対象に含め、事態の推移にリアルタイムで対応しながら研究課題に関する実践的な考察を深めることを狙った。

2.　研究成果

（1）概要

利用者が情報を引き出す行為がなくても重要な情報が利用者に届くようなプッシュ型メディアの構築を目標とし、取材・配信・編集を自動化した「ベイジアンテレビ」の開発を研究テーマとした。実世界の情報を取材・編集するエージェントの構築、データを実世界のエンティティと関連付ける解析ソフトウェアの開発、実世界の情報を取り入れつつ個別の利用者に情報を推薦するデータ同化アルゴリズム、プッシュ的・テレビ的なコンセプトに基づくデータ可視化インタフェースのデザインを具体的な開発項目とし、地名情報処理のためのソフトウェア GeoNLP の開発、ソーシャルメディアを用いた気象現況把握「ふってきったー」の構築、災害関連データの取材・編集システムの構築、プッシュ化・テレビ化インタフェース「311 メモリーズ」と「311TV」の開発・公開を果たした。

地名情報処理のためのソフトウェア GeoNLP の開発では、地名解決の方が GPS 情報付きの場合よりもマッピング可能性がはるかに高いことが分かった。また、災害関連データの取材・編集システムの構築では、東日本大震災後に重要となった電力データや放射線データなど、各種のオープンデータにも広がった。これらは、研究開始当初には対象としていなかったものである。これらは「311 メモリーズ」として公開し、第 16 回文化庁メディア芸術祭のアート部門審査委員会推薦作品に選ばれた。
ただし入力と出力をつなぐ部分、具体的には情報源と利用者とをつなぐ「情報の交換機」となる部分は、まだ課題が多く残っている。

2）詳細

研究テーマ1　地名情報処理のためのオープンソースソフトウェア GeoNLP の開発

実世界のイベントを報じるテキスト情報を位置と関連付けるジオタギング処理は、特にクライシス時のように迅速な状況認識が必要な場合に高い価値がある。こうした地名情報処理のためのオープンソースソフトウェア GeoNLP を構築し、近日中に公開する準備を進めている。例えば図 2 は東日本大震災関連ニュースから地名を抽出してマッピングした結果であり、円の色がその地名におけるニュース記事数を示す。

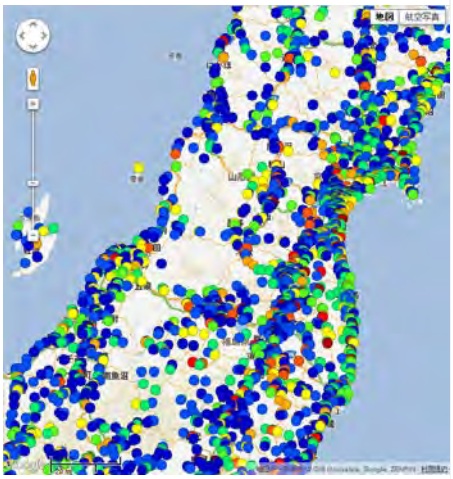

図2　GeoNLP による地名抽出

このソフトウェアが解決する問題は、自然言語文からの固有名抽出と固有名解決の 2 つである。前半の抽出では、地名辞書に含まれる地名とマッチングする単語を地名候補として抽出する。後半の解決では、複数の地名候補の中から文脈に最も適合する候補を決定する。また地名情報処理の精度は地名辞書の整備状況にも依存するため、地名辞書の共有も含めた情報プラットフォームを整備し、地名情報処理のハードルを下げることを狙う。

GeoNLP が高速な地名解決処理を実行できることを示すために、東日本大震災ビッグデータワークショップで提供された東日本大震災後 1 週間の日本語ツイート約 1 億 8000 万件に含まれる地名を解決したところ、1 時間に 1 コアあたり約 4 万ツイートを処理でき、112 コアを用いて 40 時間で全ツイートに対する地名解決処理を完了した。また、この解析結果では地名を含むツイートは全体の約 12%であった。この数字は GPS 情報付きツイートの割合 0.15%よりもはるかに大きいため、地名解決の方がマッピング可能なツイートは大幅に多いことがわかった。ただし GeoNLP は recall の向上を主に狙っているため、precision は満足できる水準には達しておらず、予備的な評価では precision は 60%程度にとどまっている。このような認識誤りを考慮すれば、地名を含むツイートは実際には 10%弱であり、12%という数字は過剰抽出によるものと考えられる。

研究テーマ2　ソーシャルメディアを用いた気象現況把握「ふってきったー」の構築

地名情報処理ソフトウェア GeoNLP の応用として、ソーシャルメディアから準リアルタイムで気象現況を把握するためのシステム「ふってきったー」を構築した。これはツイッターから「雨」や「雪」等の気象に関するキーワードを含むツイートを収集し、ツイートに含まれる地名を GeoNLP で解決し、その結果を地名ごとに集約してマッピングすることで、各地の雨や雪の状況を準リアルタイムで把握するシステムである。その結果、関東圏のようにソーシャルメディア人口が多い地域では、気象現象の準リアルタイム把握にソーシャルメディアが補完的に役立つ可能性があることを示した。例えば、雨と雪のツイート数を比較した図 3 は、どの地域が雨か雪なのかという既存のセンサ網では把握しづらい現況が、ツイート数の大小として観測できることを示唆している。このグラフから東京は雪、横浜は雨と考えられるが、実際の天候もそうであったことを気象庁官署の観測データから確認できた。その他の気象現象キーワードを用いた結果についても、「ふってきったー」では 10 分ごとに準リアルタイムで更新した結果を閲覧できる。

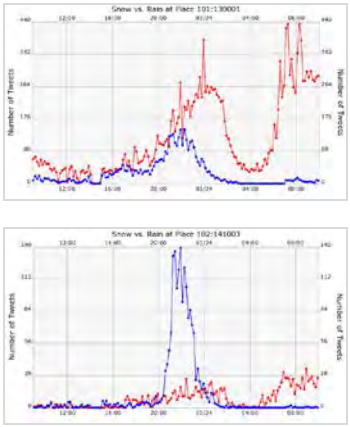

図 3　ツイッターによる気象現況把握。上段は東京、下段は横浜。赤線は雪、青線は雨ツイートの数を示す。

研究テーマ 3　災害関連データの取材・編集システムの構築

台風関連データや東日本大震災関連データをネット上から取得（取材）し、そこから情報を抽出して編集する各種のシステムを構築した。その取材対象は、当初から想定していたアメダス等の気象観測データや台風関連ニュース記事等のテキストデータだけではなく、東日本大震災後に重要となった電力データや放射線データなど、各種のオープンデータにも広がった。これらのデータは現状では個別対応が必要な部分が多いが、共通部分をフレームワーク的に構築すると同時に、解析・編集の基礎データとなる形態素解析辞書の整備や地名辞書の整備などの地道な作業も推進した。

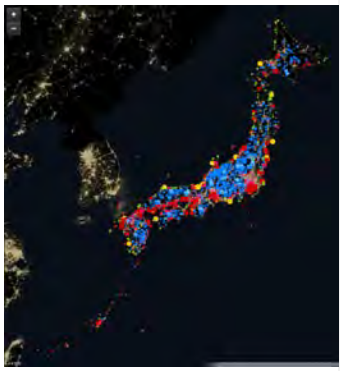

図4　電力データの可視化

またインターネット上の情報を機動的に取材するための能動的なウェブクローラーも開発した。平常時に利用するクローラーと異なる点は、気象警報等の外部からの情報に基づいてクローラーの収集範囲や頻度を動的に変更するための制御機能を備えている点にある。システムはウェブクローラーApache Nutch を改良して構築し、クローラーとの入出力にはドキュメント志向データベース MongoDB を用いた。緊急時はデータ更新が頻繁になるだけでなく、混乱の中でデータが後から検索不能になることも多い。長期的にデータをきちんと解析するためにも、クローラーの高度化による機動的な取材とアーカイブ機能を強化していく必要があると考えている。

研究テーマ4　プッシュ化・テレビ化インタフェース「311 メモリーズ」と「311TV」のデザイン

本研究を始めた最大の動機は、情報のプッシュ化とテレビ化という目標にある。プル型メディアは必要な情報を検索する時間がある平常時には有用だが、情報を検索する余裕がない緊急時には、テレビのように立ち上げるだけで必要な情報が流入してくるプッシュ型メディアが必要である、というのが本研究の重要な仮説である。このような方針に基づき構築したプロトタイプが、「311 メモリーズ」と「311TV」である。

図 5　311TV のインタフェース

まず「311 メモリーズ」は「静かに動く年表」というコンセプトを設定した。東日本大震災以後の震災関連オンラインニュース（Yahoo! News）を対象とし、自然言語処理を活用して日ごとの重要キーワードを抽出し、それを動く時間軸の上に表示した。これをプッシュ化・テレビ化インタフェースと呼ぶ理由は以下の 2 点である。第一に、利用者は能動的な行動を起こさなくても情報を眺めることができる。第二に、視覚的な情報と音楽という聴覚的な情報を合わせた視聴覚体験を味わえる。本システムを利用した結果から、能動的な情報探索体験よりも受動的な視聴覚体験の方が情報に対する思索を深めやすいのではないか、という新たな仮説を得た。能動的モードでは情報の中身の精査よりも情報の探索の最適化に意識が向くため、選択肢を減らした受動的モードの方が情報について深く考えられることが原因ではないかと推測している。なお本作品は、第 16 回文化庁メディア芸術祭のアート部門審査委員会推薦作品に選ばれた。

次に「311TV」では、スクリーンを仮想的な地域チャンネルに見立て、スクリーン表示範囲の地域情報がスクリーンにプッシュされ可視化されるという、テレビをメタファーとするインタフェースを構築した。スクリーン上ではテキストデータは自動的にスクロールし、ティッカーエリアには緊急情報が横に流れ、ラスターデータは地図上に表示される。利用者の能動的な行動は、地図や時間を移動するなど最小限で構わない。「311TV」は利用するデータのライセンスの都合で一般には公開できないが、これを発展させた「デジタル台風 TV」を近日中に公開する予定である。

またこの開発と並行する形で最近はオープンガバメントの流れが加速しており、2012 年 12 月からは気象庁が防災気象情報のプッシュ型試験配信を開始した。こうした公的サービスとも連携することで、重要な情報を即時的に届けるインタフェースの実現性も高まりつつある。その有効性を実際の台風情報で確かめることが今後の課題である。

3.　今後の展開

本研究ではベイジアンテレビの概念モデルの中でも、入力と出力の両側から研究を進めていったため、両側から成果が生まれてきた。研究開始当初「ベイジアンテレビ」は構想だけのものだったが、各種のコンポーネントが徐々につながって全体像が見え始めてきた感がある。ただし入力と出力をつなぐ部分、具体的には情報源と利用者とをつなぐ「情報の交換機」となる部分は、まだ課題が多く残っている。この部分については、外部の情報を取り入れ、内部の状態を更新し、

その時点で個別の利用者が必要とする情報を推薦するアルゴリズムを、データ同化などの観点からさらに研究していく必要がある。またこの課題は、近年の潮流であるデータジャーナリズムや計算論的ジャーナリズム等のジャーナリズムの変革といった文脈から捉えることもできる。メディアとしては「テレビ」は最大にして最強の旧来メディアであり、だからこそIT業界のプレイヤーもテレビを最終目標とした計画を練っている。こうした流れも見つつ、新しいメディアを創生するための研究課題を明確化して研究を進めていきたいと考えている。

4. 自己評価

本研究は、偶然ではあったが時宜を得たものであった。当初は台風関連データを対象としてクライシス・メディアを研究することを目標としたが、研究開始から 1 年半後の 2011 年 3 月に東日本大震災が発生し、本研究を活かすべき現場が突如として目前に出現した。この現場への対応は研究上も重要であると考え、東日本大震災関連データを対象に加えて事態の推移にリアルタイムで対応する取り組みを開始した。事前予測不可能な状況への対応を通して、実感をもって各種のメディアの限界を認識できたのは貴重な体験だった。また各種の貴重なデータを収集し、データを活用したサービスも開発し、その一つがメディアアートとして文化庁メディア芸術祭で受賞できたことは、研究対象を拡大した判断が適切だったことを示すと評価している。また GeoNLP を中心とするソフトウェア開発でも、震災時に使えなかったこと、そして大規模震災データを対象に評価を進めたことが、研究を進める上での強い動機として作用した。当初の研究課題の中に未達成部分が残ってしまったことは心残りである。しかし、この超大規模震災の期間にさきがけ研究を進めていたからこそ可能になった研究も多く、その巡り合わせには感謝の念を持っている。

5. 研究総括の見解

利用者が情報を引き出す行為がなくても重要な情報が利用者に届くようなプッシュ型メディアの構築を目標とし、取材・配信・編集を自動化した「ベイジアンテレビ」を開発した。

当初は台風関連データを対象としていたが、途中で東日本大震災が発生し、大震災関連データを対象に加えて事態の推移にリアルタイムで対応する取り組みに方針変更し、データを活用したサービスも開発している。入力と出力をつなぐ部分、具体的には情報源と利用者とをつなぐ「情報の交換機」となる部分は、課題未達成であるが、メディアアートとして文化庁メディア芸術祭で受賞という成果もあげている。

「テレビ」は最大にして最強のメディアとのことで、テレビを最終目標とした計画を維持しているが、今後はテレビというメタファーに囚われず、自由な発想で研究を進めてもらいたい。

6. 主な研究成果リスト

（1）論文（原著論文）発表

1. Asanobu KITAMOTO, Takeshi SAGARA, "Toponym-based Geotagging for Observing Precipitation from Social and Scientific Data Streams", Proceedings of the 2012 ACM Workshop on Geotagging and Its Applications in Multimedia, GeoMM'12 (co-located with ACM Multimedia 2012), Liangliang Cao, Gerald Friedland (編), pp. 23-26, ACM, 2012 年 11 月 (in

English)
2. Asanobu KITAMOTO, “Lessons Learned from Data Management Activities after the Great East Japan earthquake in March 2011”, Data Science Journal, (in press), 2013.
3. 北本 朝展, ″大規模マルチメディアデータの統合と検索による気象イベントのモニタリング″, 映像情報メディア学会誌, Vol. 66, No. 11, pp. 907-912, 2012 年 11 月
4. 北本 朝展, ″デジタル台風：気象現象のリアルタイム認識・理解に向けた大規模データの統合と分析″, 電子情報通信学会技術報告, Vol. 111, No. 222, pp. 5-5, 2011 年 10 月 （招待講演）
5. 北本 朝展, “センサデータとソーシャルメディアの統合によるリアルタイム状況認識の可能性”, 電子情報通信学会ヒューマンプローブ研究会公開シンポジウム「ヒューマンプローブの新たな展開」, 2012 年 11 月（招待講演）

（2）特許出願

研究期間累積件数：0 件

（3）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

- GeoNLP： http://agora.ex.nii.ac.jp/GeoNLP/
- デジタル台風： http://agora.ex.nii.ac.jp/digital-typhoon/
- 東日本大震災アーカイブ： http://agora.ex.nii.ac.jp/earthquake/201103-eastjapan/
- 311 メモリーズ： http://agora.ex.nii.ac.jp/earthquake/201103-eastjapan/311memories/
- ふってきったー： http://agora.ex.nii.ac.jp/futtekitter/
- デジタル台風 TV： http://digital-typhoon.tv/
- 第 16 回文化庁メディア芸術祭アート部門審査委員推薦作品「311 メモリーズ」