

図 1: システム構成

3. 災害情報支援システムの構築

本研究では、2. 章で示した手法を用いて、災害情報を Twitter から収集し、自動的に分類し、わかりやすい形で関係者に提供することを目的し、システムの構築を行った。作成されたシステムは以下の様な流れで処理を行う。

- (1). 災害関連単語を含むリツイートデータを収集
- (2). それぞれの分類期間 (現在から 24, 12, 6, 3, 1 時間以内) で分類
 - (a) 選定された期間内に 10 回以上リツイートされたツイートの内、リツイート回数が多いものから 1000 ツイートを抽出
 - (b) 抽出されたツイート群のリツイートユーザ重複率を算出
 - (c) 情報共有行動ネットワークを構築
 - (d) Newman 法により情報共有行動ネットワークをクラスタリング
- (3). ユーザへ分類結果を提供

ツイートデータ収集, 分類, 提供の処理は図 1 に示すシステム上で行われる。

(1) の災害関連単語とは、地震、津波といった、災害に関連すると考えられる単語である。以下は収集に用いられた災害関連単語の一覧である。

災害関連単語

地震, 揺れ, 震度, マグニチュード, 活断層, 余震, 震源, 断層, 津波, 台風, 大雨, 集中豪雨, ゲリラ豪雨, 浸水, 洪水, 竜巻, 突風, 大雪, 豪雪, 雪崩, 寒冷前線, 噴火, 火山, 火砕流, 噴石, 溶岩, 震災, 警報, 注意報, 避難, 仮設住宅, 帰宅困難, 減災, ハザードマップ, 防災, 災害, 停電, 被災

得られたツイートデータについて、リツイートユーザ重複率の算出 (2b) には Hadoop^{*1} クラスタを使用した。Hadoop は大規模データのための分散処理フレームワークであり、長大なデータに対する逐次処理を高速・冗長化する。本研究で使った Hadoop クラスタは一般的なデスクトップ PC10 台と、ハイエンドな計算用 PC1 台で構成されている。

最後に、(3) の分析結果の提供では WEB アプリケーションとしてユーザに分析結果の提示を行う。情報提供ツイートを

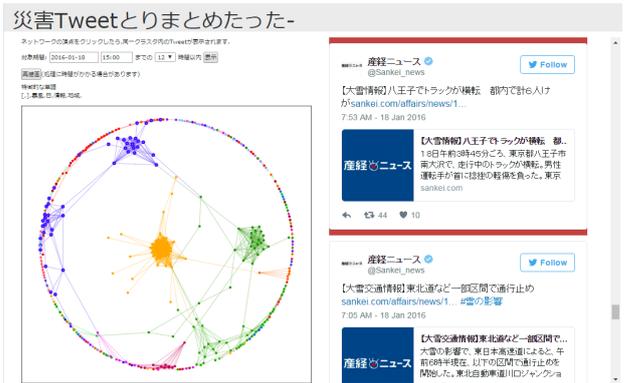


図 2: 本システムによる分類結果の提供



図 3: システムで提供される XML 形式のデータ例

ノードとし、関連ツイートとの間をリンクでつなぐことによってネットワークとして描画する。クラスタごとにことなる配色とし、分類された情報を可視化した。さらに、各ノードをクリックすることで、同じクラスタに含まれる関連ツイートの一覧を表示する。情報提供のシステムによる分類結果の例を、図 2 に示す。画面左がツイートクラスタを示したネットワークであり、画面右に選択したクラスタのツイートが表示される。なお、各クラスタの特徴を示すため、ネットワーク上のクラスタ上にマウスカーソルが載った際に、TF-IDF によって算出した特徴語を表示する。また、データを直接利用したい開発者や研究者のために、XML 形式でのデータの提供も同時に行っている。提供される XML の例を図 3 に示す。

本システムは WEB アプリケーションとして公開中^{*2}である。

4. 運用実験

本システムの運用実験を、関東で記録的な大雪が降った 2016 年 1 月 18 日のデータを対象として行った。当日は全国的に大雪が降り、東京でも数センチの積雪があり交通機関が乱れるなど大きな混乱が生じていた。

分析は 2016 年 1 月 18 日 15 時までの 12 時間のツイートを

*1 <https://hadoop.apache.org/>

*2 <http://www.torilab.sakura.ne.jp/disasterTweets/>

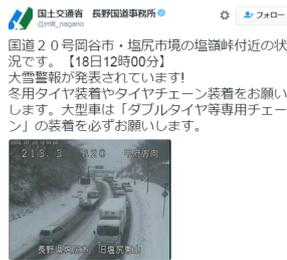


図 4: 大雪情報のツイート例

利用した。分類対象となったツイートは 387 あり、それらが 69 のクラスタに分類された。

もっともツイート数が多かったクラスタには国土交通省長野国道事務所 (@mlit_nagano) のツイートを中心とした長野県の大雪情報に関するツイートを中心に 50 ツイートが含まれていた。国土交通省長野国道事務所のツイートの例を図 4^{*3} に示す。

2 番目にツイート数の多かったクラスタは 45 ツイートを含み、新聞やテレビといったマスメディアのアカウントによる大雪情報が中心であった。

3 番目にツイートの多かったクラスタは、首都圏に関する NHK ニュースのアカウントによるツイートを中心に、停電や交通情報がまとまったツイート群であった。

そのほか、含まれるツイート数が多いクラスタには、交通情報を集めたツイート、個人による大雪情報ツイート、鉄道事故関連ニュースアカウントによるツイート、東北の大雪情報^{*4}などが分類された。

本システムによって分類された情報は、単に災害情報だけではなく、交通関連、あるいは地方ごとの情報に分割されていることが確認された。実際、求める災害情報は被災者によって異なっている。たとえば、首都圏で通勤しようとしている人であれば交通情報が重要であるが、同じ情報を長野県の人を受け取ってもあまり意味が無い。そのため、どのような情報なのか、そしてどの情報なのかの二点によって情報が分類されていることは、被災者にとっては重要となる。以上を踏まえると、本システムによって災害情報は適切に分類できているといえよう。

5. おわりに

本研究では、リツイート行動に基づいて情報共有行動ネットワークを構築し、ネットワーククラスタリングを適用することで、震災時に共有された情報の分類を行うシステムの構築を行った。運用実験によって、構築されたシステムは震災時に求められる情報を提供することが確認された。本システムは常時稼働しており、突発的な災害時にも活用されることが見込まれる。また、分類結果は XML 形式でダウンロード可能なため、他のシステムと本システムを統合することも容易に可能である。

今後の課題としては、ユーザの位置情報や、ツイート履歴などを活用することで、それぞれの個人の状況下に求められるで

あろう情報を、ユーザごとに提供するアルゴリズムの実装などが挙げられる。

これらの基礎技術についての研究は、人工知能学会近未来チャレンジ「異種協調型災害情報支援システム実現に向けた基盤技術の構築 (CHIDRI)」プロジェクトで行われている。最終的にこれらの基礎技術を統合した異種協調型災害情報支援システムを実現することも本研究の重要な課題の一つである。

参考文献

- [Baba 15] Baba, S., Toriumi, F., Sakaki, T., Shinoda, K., Kurihara, S., Kazama, K., and Noda, I.: Classification Method for Shared Information on Twitter Without Text Data, in *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pp. 1173–1178, Republic and Canton of Geneva, Switzerland (2015), International World Wide Web Conferences Steering Committee
- [Clauset 04] Clauset, A., Newman, M. E. J., , and Moore, C.: Finding community structure in very large networks, *Physical Review E*, pp. 1– 6 (2004)
- [Cobo 15] Cobo, A., Parra, D., and Navón, J.: Identifying Relevant Messages in a Twitter-based Citizen Channel for Natural Disaster Situations, in *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pp. 1189–1194, Republic and Canton of Geneva, Switzerland (2015), International World Wide Web Conferences Steering Committee
- [Frakes 92] Frakes, W. B. and Baeza-Yates, R. eds.: *Information Retrieval: Data Structures and Algorithms*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1992)
- [Imran 15] Imran, M. and Castillo, C.: Towards a Data-driven Approach to Identify Crisis-Related Topics in Social Media Streams, in *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pp. 1205–1210, Republic and Canton of Geneva, Switzerland (2015), International World Wide Web Conferences Steering Committee
- [Mendoza 10] Mendoza, M. and Poblete, B.: Twitter Under Crisis: Can we trust what we RT?, in *Proceedings of Social Media Analytics, KDD '10 Workshops, Washington, USA* (2010)
- [Sakaki 10] Sakaki, T., Okazaki, M., and Matsuo, Y.: Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, in *In Proceedings of the Nineteenth International WWW Conference (WWW2010)*. ACM (2010)
- [Small 73] Small, H.: Co-citation in the scientific literature: A new measure of the relationship between two documents, *Journal of the American Society for Information Science*, Vol. 24, No. 4, pp. 265–269 (1973)
- [Toriumi 13] Toriumi, F., Sakaki, T., Shinoda, K., Kazama, K., Kurihara, S., and Noda, I.: Information

*3 https://twitter.com/mlit_nagano/status/688876539744620544

*4 その他、人気ラーメンチェーン各店舗の営業状況なども抽出された

Sharing on Twitter During the 2011 Catastrophic Earthquake, in *Proceedings of the 22Nd International Conference on World Wide Web, WWW '13 Companion*, pp. 1025–1028, Republic and Canton of Geneva, Switzerland (2013), International World Wide Web Conferences Steering Committee

[平塚 12] 平塚 千尋：新版 災害情報とメディア，リベルタ出版 (2012)

[北田 15] 北田剛士, 風間一洋, 榎剛史, 鳥海不二夫, 栗原聡, 篠田孝祐, 野田五十樹, 斉藤和己：“東日本大震災時のツイートのトピック系列の可視化と分析”，人工知能学会 (2015)