

気象現象を対象としたソーシャルメディアの取材・分析・可視化手法 Gathering, analysis, and visualization of meteorological phenomena using social media

北本 朝展^{1*}

KITAMOTO, Asanobu^{1*}

¹ 国立情報学研究所, ² 科学技術振興機構

¹National Institute of Informatics, ²JST

気象現象はすべての人に影響する事柄であり、ソーシャルメディアにおいても日々の会話における主要な話題の一つとなっている。一方で、台風などによる気象災害時にはソーシャルメディアは即時的に緊急情報を発信する場としても機能しており、現地からの生の情報が素早く集約され共有されるという災害対応のための有用なプラットフォームに成長していくことには根強い期待がある。そこで本稿では気象現象とソーシャルメディアに焦点を合わせ、ソーシャルメディアをどのように取材し、分析し、可視化することが可能か、筆者らの経験を踏まえて議論したい。

本稿でケーススタディとして主に取り上げるのは4つのシステム、すなわち「台風前線」(<http://front.eye.tc/>)、「ツイフーン」(<http://twiphon.eye.tc/>)、「台風なう!」(<http://typhoonnow.eye.tc/>)、「ふってきったー」(<http://agora.ex.nii.ac.jp/futtekitter/>)である。筆者らは台風情報を対象に、2004年にはブログから収集する実験、2009年にはツイッターから収集する実験を開始した。これらはユーザがイベントを報告するアクションを必要とする「通報型取材法」を用いている。この方法には、ユーザの通報の形式や質をある程度コントロールできるというメリットがあるものの、ユーザの積極的な関与が必要のため収集できる情報量に限界があるという問題があった。そこで2012年から開始した「ふってきったー」では「巡回型取材法」を採用し、ユーザの明示的な関与がなくてもイベントを積極的に取材し収集可能なシステムを構築した。これにより収集可能な情報量は大幅に拡大し、例えば関東地方の降雪イベントでは1時間あたり数千単位で情報を収集することが可能となった。

しかし巡回型取材法の問題点は、よく整理されたメタデータをユーザが付与してくれるとは期待できない点にある。この問題が端的に表れるのが固有表現の問題である。例えばそのツイートがどの場所の何の現象を述べたものなのか、そうしたメタデータは、例えばツイッターではハッシュタグの形で付与される場合もあるが、この種のメタデータには統一された形式が存在しないため、取材側からこれをコントロールすることは困難である。これを解決するには、テキストを自然言語処理して内容を解析し、固有表現を抽出して統合する必要がある。このような処理を地名に関して実現するため、自然言語文中の地名を抽出して一意に解決するジオコーディングツール GeoNLP (<http://agora.ex.nii.ac.jp/GeoNLP/>)を現在開発中である。まだ精度の問題は残っているものの、テキストを自動的にマッピングする処理が以前に比べて容易となった。

このような機能を備えた「ふってきったー」を構築し、雨と雪に関するツイートを取材、分析、可視化する実験に適用した。収集したツイートを分析することによって、どのくらいの時間・空間解像度で雨と雪に関する情報をソーシャルメディアから集めることができるのか、不確かな情報をどの程度信用できるのか? そうした問題意識に基づく分析結果を議論する。

キーワード: 気象現象, ソーシャルメディア, ツイッター, 固有表現, ジョコーディング, 自然言語処理

Keywords: Meteorological phenomena, Social media, Twitter, Named entity, Geocoding, Natural language processing