

デジタル台風：複合災害の現状認識を目指すビッグデータ解析
Digital Typhoon: Big Data Analysis toward Situational Awareness of Complex Disasters

○北本朝展
○Asanobu KITAMOTO

“Digital Typhoon” is a website that became a data infrastructure for scientists, citizens and machines. The original purpose of the website was a demonstration prototype for our technology on content-based image retrieval, but later evolved into a large-scale infrastructure with the variety of typhoon-related data. One of a relevant research question in Digital Typhoon is to use big data analysis of typhoon-related data for reducing the problem of information overload and scarcity. The paper closes with detailed discussion about the future direction of Digital Typhoon, especially from the viewpoint of open science, to transform the website from a personal website to a sustainable website supported by communities and organizations.

1. はじめに

ウェブサイト「デジタル台風」は、開設して十数年の間に、通算2億ページビュー以上を記録するほど多くの利用者を抱えるサイトに成長した。しかし最初から人気サイトを目指したわけではなく、2001年頃にサイトを開設した目的は、「内容に基づく画像検索」技術のデモンストレーションであった。自らの研究をデモするときに使うつもりだったサイトが、様々なデータや機能を加えて増築に増築を重ね、今や作成者自身も全貌を把握するのが難しいほどの巨大サイトに成長した。

こうした成長は、気象学者からのリクエストに応えた結果として生じたのではなく、自らの知的興味とサイト運用ニーズに沿って生じたものである。筆者は気象や台風に関して小さい頃から興味を抱いており、その後は気象予報士の資格も取得した。さらに専門分野である情報学の観点から、気象データをどのように整理し、アクセス可能とすべきかについて研究上の関心を有していた。このような背景から、情報学的興味を主、気象学的興味を従として生まれたのがデジタル台風というサイトの特徴である。このことが、気象学的興味を主としたサイトとは大きく異なる、独特のデータ構造を生み出す要因になっていると言える。

2. デジタル台風の歴史

筆者が有していた情報学的興味とは何か。それは今なら「ビッグデータ解析」とシンプルに呼ぶことができよう。しかし、デジタル台風の研究を

開始した1999年、あるいは2000年前後[1, 2, 3]には「ビッグデータ」という便利な用語が存在しなかったため、筆者はmeteo-informatics（気象情報学）という言葉に代わりて提唱していた。当時既に発展していたbio-informatics（生物情報学）を参考に、大規模データの統合的な解析から新しい現象を発見するというアプローチを、気象学に適用することを目論んだのである。そして台風を「モデル現象」と位置づけることにした。生物学では、大腸菌やマウスなどの「モデル生物」を詳細に分析することで、典型的な現象に対する理解を深めると共に、一般の生物へと分析を広げる足場とするアプローチがある。そのアナロジーとして、「台風」という研究に好適な性質を備えたモデル現象を、気象データの典型例として集中的に分析するというアプローチを考えたのである。

研究に好適な台風の性質とは何か。それは、1) 台風番号というユニークなIDが付与されること、2) 気象庁が長期間にわたって責任を持って分析したデータが存在すること、3) 社会的に重要な現象であるだけでなく、多種類のデータで共通のIDが利用されること、などである。こうした性質は他の気象現象や災害事象では一般に成立しないため、研究対象としての台風には様々な利点がある。こうした考察を踏まえ、台風データの収集を開始したのが1999年であった。

実は台風に注目した理由がもう一つある。それは当時の研究テーマであった「内容に基づく画像検索」に関係する。画像内容に基づく検索とは、

画像に付与されたキーワードを使わずに、画像という画素配列データのみを分析することで、検索要求に合致する画像を探すという問題である。典型的なタスクは、キーワードではなく画像をシステムに与え、それに類似した画像を検索するというタスクである。このテーマに大学院時代から取り組んでいた筆者は、職を得てからも同じ問題に取り組んでいたが、多少の行き詰まりも感じていた。そこで頭に浮かんだのが台風だったのである。

筆者が求めていたのは、画像検索の精度を客観的に評価する手法、あるいは客観的な評価を可能とする問題設定であった。ここで着目したのが「ドボラック法」と「台風ボーガス」という2つの概念である。まずドボラック法が、専門家による目視判断に依存する点に着目した。これは筆者が取り組んでいた画像解析の問題そのものである。次に台風ボーガスが、人間が適当に(?) 定めた人工的なデータである点に着目した。もしその巧拙がシミュレーション精度を左右するなら、台風ボーガスの制御が予測性能に与える影響を評価できるのではないか。この2つのアイデアを組み合わせたストーリーは以下ようになる。現在の台風に類似した過去の台風を検索することで、過去の台風(再解析済み)を参照してより高度なボーガスを生成し、それを埋め込んだ場合の予測性能を比較すれば、検索の良さを定量的に評価できるのではないか?つまり、過去事例に基づく高度なボーガス生成を用いて初期値を改善することで、予測性能という観点から画像検索の精度を評価するという、無謀な(?) 研究目標が誕生した。

この研究目標は、今から考えれば遠回りすぎる問題設定だったと言えるだろう。ボーガスは確かに予測性能に影響を与えるだろうが、それ以外にも影響を与える要因は多数あり、調べたい要因だけを取り出して論じることは簡単ではない。しかしいったん目標を決めた以上、進むしかないと考えた筆者は、できるだけ多くの事例を、統一的な基準で、可能な限り網羅的に収集する必要に迫られた。例えば統一的な基準として、台風の雲パターンを正規化するという処理を取り上げてみよう。これは、台風の中心を画像の中心と一致させ、地球上で同じ大きさになるように衛星画像を切り取るという処理である。単に緯度経度の範囲で画像を切り出すだけでは、この条件を満たすことはできない。湾曲した地球表面において、単位緯度が一定の距離とはならないためである。そこで円形

の雲パターンの形状歪みが小さくなり、かつ面積を保存するような特殊な地図投影法を用いて、静止気象衛星画像から台風画像を切り抜く方法を開発した。ここまで厳密な処理は、人間が見るだけなら不要かもしれないが、コンピュータが統一的な処理を行うには不可欠なのである。

こうして大規模化へのモチベーションが高まったおかげで、デジタル台風は15年にわたって着実な発展を続けてきた。その発展段階を大まかにまとめると、以下の8段階となる。

1. 気象衛星「ひまわり」の網羅的データベース。
2. アメダスやレーダーなど気象データへの拡大。
3. ニュース記事などテキストデータへの拡大。
4. ツイッターなどソーシャルメディアへの拡大。
5. 気象庁防災情報XMLなど防災データへの拡大。
6. GPVなど数値予報データへの拡大。
7. 伊勢湾台風や天気図など歴史データへの拡大。
8. コラボレーションデータへの拡大(今後)。

15年間にわたるデータベース構築の成果として、データ収集というフェーズは大きな山を越えたのではないかと筆者は感じている。一方、データ活用というフェーズは、未だに問題が山積している。データがあまりに多様となった結果、データ探索の難易度がどんどん上がってきた。せっかくサイトで提供しても、利用者に気づかれなければ、それは存在しないのと同じなのである。

こうした問題を解決し、データ活用を推進する方法として、筆者は2つの解決策を考えている。第一は、データの存在を知らせるためのデータカタログやチュートリアルを整備である。第二は、ユーザの状況(コンテキスト)に合わせて、見るべきデータをプッシュするという情報推薦技術の導入である。特に第二の方向は、モバイル端末の普及に伴って重要性を増している。狭いスクリーンと乏しい情報入力方法という問題点を乗り越えて必要な情報を届けるには、コンピュータが人間を支援することが不可欠だからである。

3. ビッグデータ解析

ビッグデータとは、単に巨大なデータを意味するわけではない。デジタル台風で扱うデータは数十テラバイト程度に達するため、これを処理するにはそれなりの規模の情報基盤が必要となる。とはいえ、気候予測データのようにペタバイト規模に達するほど巨大ではないため、一研究室の規模で扱うのが不可能なほどではない。むしろデジタ

ル台風が重視するのは多様性である。画像、テキスト、数値、グリッドデータなど、多種多様な台風関連データをどうやって統合し、そこから価値を生み出していくか。ビッグデータにおいては、データの「掛け合わせ」や「読み替え」が新しい価値を生み出す駆動力となる。そうした発想を支えるアルゴリズムや方法論を提案することは、デジタル台風の主要な研究テーマである。

そうしたテーマの一つに、クライシス時の情報爆発の解決がある。多種多様なデータが大量にシステムに流入してきたとき、限られた時間と限られた注意力という制約の中で、人間がデータの意味をすべて理解することは不可能である。この限界を緩和するためには、システムが情報を集約して1枚の絵にまとめ、情報への負荷を減らすとともに情報の共有を進めることで、現状認識の能力を高めることが必要である。

そこでデジタル台風では、多様なデータを対象として、データの意味を理解しやすくして情報負荷を減らす方法の提案に取り組んでいる。本発表で紹介したのは、気象データ、テキストデータ（マスメディア・ソーシャルメディア）、災害・防災データであるが、その詳細についてはプレゼン資料や別論文[4, 5]等を参照いただきたい。

ただしすべての試みに共通する方針として、「データを解釈する文脈を提示する」という考え方があることは指摘しておきたい。これはデータを単独で表示せず、可能な限り他のデータと一緒に表示するという考え方である。例えば2つのデータを一緒に表示することで、どちらが大きいかといった付加的な情報が得られる。またデータと平均と一緒に表示することで、平均より上か下かといった情報が得られる。このような「データの文脈化」を通して、データの意味をより明確にすることで、情報としての理解と活用を支援することができる。また地図へのマッピングやグラフの可視化が狙うのも同様の効果である。位置や時間という文脈の中で個々のデータはどのような位置づけなのか、そうした情報を視覚的に理解可能とすることで可視化ツールの価値がある。

4. 今後の方向性

先述のように、デジタル台風のデータ収集フェーズは、15年を費やしてようやくゴールがうっすら見える段階に到達できたと感じている。もちろん、定常的な更新にまつわる作業や、細かな機能

向上のための作業は永遠に続いていくし、ひまわり8号などの新たなデータが登場すれば、それに対応するための作業も新たに発生する。とはいえ、これらは持続的な運用に関わる問題であり、デジタル台風の姿を大きく変えるものではない。より大きなインパクトを与えうる残されたフロンティアとしては、以下の2つの領域に注目している。

第一が、クラウドソーシングあるいはシチズンサイエンスによって生み出されるデータである。デジタル台風が解くべき問題は情報過多であると既に述べたが、一方で情報過少の問題も忘れてはならない。情報過多が問題なのに、なぜ情報過少も問題なのか。それは、前者がデータの総量に着目するのに対し、後者は細分化されたデータの量に着目するからである。例えば日本全体では情報過多であっても、特定の市区町村に限定すれば情報が十分ではないことがある。情報過多の状態から、時間・空間・テーマなどの軸で情報を細分化していくと、情報が不足する「セル」が生まれてくる。この問題は災害における被災地域で特に顕著となる。こうした問題を解決するには、機動的に観測を増やす手段が必要であり、そこで頼りになるのがクラウドソーシングやシチズンサイエンスの方法論である。これらの方法論は気象庁や民間企業等でも活用が進むが、並立するプロジェクトの間でデータを共有することは簡単ではない。しかし、限られた人的資源を有効に活用するためには、人々の負担にならない形で相互に情報を共有していくことが重要な課題となる。

第二が、研究成果データである。デジタル台風が扱う対象は、従来は研究の素材となる観測データやテキストデータに限定され、それらの分析から得られた研究成果データは対象ではなかった。しかし、研究素材と研究成果をリンクさせれば、研究素材データから得られた研究成果データを一覧できるだけでなく、そのデータを再利用して別の研究を展開するという、研究データの循環にもつなげていくことができる。

このようなリンクは、伝統的には論文を単位としたコミュニティ単位の論文データベースの中で構築されてきた。すなわち、あるコミュニティ（例えば台風研究コミュニティ）の人々が投稿しそうな学術雑誌に掲載された論文を収集し、それを読み込みながら論文が対象とした現象（例えば台風〇〇号）を人間が注釈付けするのである。そして、個々の台風ごとの研究論文リストを見れば、その

台風についてこれまで何が研究され、そして何がわかったのかを、研究者はすばやく理解できるようになる。もちろん、現代の電子ジャーナルを使えば、こうした注釈付けをコンピュータが代替することもある程度は可能だろう。このようなデータのアノテーションやキュレーションは、研究を支援する情報として重要な役割を果たしてきた。

それに対して、近年は研究データの管理という問題にも注目が集まり始めている。これは論文だけではなく、論文の成果の元となる研究データも共有するという考え方である。その背景には2つの目的がある。第一が利便性である。研究成果データを他者が再利用可能にすれば、重複した研究を減らして研究成果を増すことができる。第二が透明性である。研究成果を他者が再検証可能にすれば、不正を働いた研究を検出して研究成果の信頼性を向上させることができる。こうした文脈では、研究素材データと研究成果データのリンクの構築が、デジタル台風の新たな役割になりうることをわかる。そうなるデジタル台風は、研究コミュニティにおけるデータポータルやデータリポジトリという役割を果たすことになる。

デジタル台風とは何のためのウェブサイト（ウェブサービス）なのかという点は、結局のところこのサイトが研究コミュニティや社会の中でどういう役割を果たすのか、という点にかかってくる。現在のデジタル台風は、様々なサービスを支える存在となっており、ある種のインフラとしての役割を果たすようになった。しかし実態は個人が運営する「個人サイト」であり、この点は最初から現在に至るまで変わっていない。デジタル台風の運用を支援する研究費は存在しないため、他のプロジェクトの枠内にデジタル台風を位置づけていかねばならない。そのため、デジタル台風に専属の研究者・技術員を雇用することは困難であり、文字通りすべてを一人で運用するという個人サイトという状態が続いている。これはサイトの永続性が、極めて脆弱な状態にあることを意味する。

ここで問題になるのは、こうした研究データ基盤を維持していくべきなのは誰で、それをどのように支援するのが望ましいかという点であろう。これはデジタル台風固有の問題ではなく、あらゆる学術分野で生じている問題でもあり、そのことは筆者が関わる「オープンサイエンス」のムーブメントにおいても重要な課題として意識されている。オープンサイエンスは、科学研究のあり方

や進め方を変革するムーブメントとして幅広いトピックを含むが、その中には研究データをどう管理し、研究データの貢献にどうクレジットを与え、それをどう定量化できるか、といった問題も含まれる。つまり、デジタル台風のような研究基盤を維持する問題は、オープンサイエンスの枠組みを使って考察していくことができる。とはいえ、オープンサイエンスの進む先がまだ曖昧模糊としているのと同様、デジタル台風の進む先も霧に包まれている。霧の中を迷わず航海していくことは難しいが、その羅針盤となるのは、生み出す価値が大きい方向に進むという方針ではなかろうか。

冒頭に述べたように、もともとデジタル台風は気象学者の研究を支援する意図はなく、あくまで自らの知的興味を満たすためのサービスであった。しかしそろそろ、この考えを改めるべき時なのかもしれない。最初は個人サイトとして始まったものの、徐々に重要性が高まってきたサイトを、どうやってコミュニティや組織による永続的な運用につなげるか。それが多くの研究者の悩みの種でもある。個人サイトから永続的网站に移行すれば、運用方針も個人の興味からコミュニティのコンセンサスへと移行せざるを得ない。両者の間に横たわる深い谷をどう越えるか、それがこれから数年の重要な課題であると考えている。

参 考 文 献

- [1] 北本 朝展, "「デジタル台風」 -- 人工知能的アプローチに基づく台風解析", 情報処理学会技術報告, Vol. CVIM123-8, pp. 59-66, 2000.
- [2] 北本 朝展, 小野 欽司, "台風画像コレクションの構築および台風解析への応用", NII Journal, No. 1, pp. 7-22, 2000.
- [3] 北本 朝展, "Holistic Analysis を用いた台風雲パターンの解析", 電子情報通信学会技術報告, Vol. PRMU2000-240, pp. 129-136, 2001.
- [4] 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.
- [5] 北本 朝展, "オープンな地名情報システム GeoNLP~曖昧なテキストの地名を解析し共有するためのツール~", 月刊「測量」, Vol. 64, No. 9, pp. 6-11, 2014.