

「デジタル台風」プロジェクト： 大規模時系列画像データベースのマイニングにむけて

国立情報学研究所 北本 朝展

kitamoto@nii.ac.jp

<http://www.digital-typhoon.org/>

気象衛星画像が毎日の天気予報に登場するようになってから、台風の渦巻き状の雲は誰もが知る存在となった。しかしその形状を仔細に観測してみると、台風ごとに多様な形状を示すことがわかる。その形状の多様さは見ているだけでも興味深いものであるが、実際的な応用においても、台風の形状と台風の勢力には相関があるとの仮定が台風の勢力推定に利用されている。つまり台風の形状には、台風の特徴が埋め込まれて現れていると考えてよいだろう。このような複雑な形状と、個々の台風の気象学的な特徴とはどのように関係しているのだろうか、という疑問を出発点として「デジタル台風」プロジェクトはスタートした。

プロジェクトの基本的な方針は、1) 関連データの網羅的な収集、2) データベースの構築、3) データベースからのマイニングである。まず最初に、台風経路データは1951年以降の1400個以上の台風経路、静止気象衛星画像（ひまわり・ゴーズ）は1995年以降の5万件以上の台風画像を網羅的に収集し、現在も更新を続けている¹。これらのデータを次にデータベース化し、さまざまな方法で検索したり解析したりすることが可能な状態に整備した。そして、このデータベースに隠されたパターンを、アルゴリズム的に発見しようというのが、プロジェクトの最終的な目標となる。ここで重要となるのが、画像パターンを扱うための画像解析・パターン認識技術であり、また時間的に変動する対象を扱うための時系列解析技術である。そこでこれらの領域を中心に研究を進めてきたのがこれまでの流れである。

これらの基本方針は、著者が今後推進しようと考えている「メテオ・インフォマティクス (meteoroinformatics)」という方法論に基づくものである。その基本的なアイデアは、データベースに基づく科学的探究というスタイルを、気象学に持ち込もうというものである。以前の科学的探究では、少量のデータを大事に利用するというスタイルが主流を占めていたが、現在では大量のデータを取得しデータベース化した後、そこから必要なデータを取り出すというスタイルが増加してきた。両者の研究の本質的な違いは、特定の研究対象をまず選んで観測するか、網羅的な観測の後に必要な研究対象をコンピュータ上で選ぶか、という違いにある。センサ技術とコンピュータ技術の発達により、現在ではむしろ後者の方が能率的な場合も増えてきたことが、いわゆる「X-インフォマティクス」研究の隆盛の背景にある。このような方法論を支えるためのデータ品質の一貫性保持、データ検索言語の開発などのテーマも、本プロジェクトの重要な研究課題である。

本プロジェクトはこのようにデータ中心主義のスタイルを採ってはいるが、観測されたデータを後付けで説明するだけでは研究として不十分であるとも考えている。観測されたデータには常に「たまたまそうだった」という側面があり、観測データを絶対視して説明するだけではモデルとして不十分であろう。換言すれば、観測されてデータとなった値は何らかの確率変数の実現値に過ぎず、観測されたデータを生成するプロセスに潜む確率的な構造を推定することが、本来やるべきことであると考えられる。本プロジェクトでは台風の雲パターンを特徴空間中の軌跡と捉え、そのダイナミクスを確率的にモデル化することでこの問題を追究する計画である。

上記の方針に基づき、現在までのところ、台風パターンの類似性検索とその時系列比較、台風雲パターンによる発達/衰弱の分類、台風ライフサイクルの類似時系列検索などについての結果を得た。その結果の詳細については発表で紹介する。これらのアルゴリズムはまだ実用的な台風解析に使えるものとはなっていないが、その大きな原因は、台風雲パターンの特徴解析と類似性に関する研究が不十分な

¹その他にアメダスデータやニュース記事、多数の人々による情報発信なども収集しているが、本発表では軽く触れるにとどめる。

点にあると考える。今後はこのテーマをさらに深く掘り下げることで、大量データを根拠とした新たな発見に結び付けていきたい。また、気象学はもともとシミュレーションが特に盛んな分野であるが、データベースとシミュレーションは、帰納過程と演繹過程というコインの表裏のような関係にあり、本来はお互いにフィードバックしあえる存在でもある。こうした二つの分野の融合も新たな発見に結びつく可能性がある。

最後に、研究のデモンストレーションとして作成したデータベース「デジタル台風」は、準リアルタイム更新の台風データベースとして多くの利用者を集めており、累計のページビュー数は1100万を越えている(2004年10月末現在)。このような活動は研究としては評価されにくい面もあるが、社会への還元として続けていきたいと考えている。