

生態系研究におけるビッグデータの動向と学際研究の可能性

“Big-data” in ecological science and seeds for interdisciplinary researches

大場 真^{*1}
Makoto Ooba

^{*1} 名古屋大学エコトピア科学研究所
EcoTopia Science Institute, Nagoya University

生態系研究分野においても、測定の最密化やリアルタイム化、過去のアナログデータ(標本や紙地図など)のアーカイブ化、これらを収納するデータベースの巨大化と複雑化、そしてそれらを組織化する様々なレベルの技術とネットワークが急速に成長を始めている。本発表では生態系研究におけるビッグデータの状況を素描し、それに対する生態系情報学(エコインフォマティクス)の展望、今までは考えられなかった情報科学分野との学際研究の可能性について紹介する。

1. はじめに

エコインフォマティクス(ecoinformatics)は 2000 年代より注目され始めた分野で、“eco”と情報学 informatics の造語である。解釈によるが、eco は、ecology(生態学)や ecosystem(生態系)を指している。日本語では生態系情報学となるだろうか。

本発表では、大量データを生態系科学へもたらした技術について概観した後で、それに関連して発展中であるデータベースとそのネットワークについて説明する。

またエコインフォマティクスには諸自然科学や工学、農学、社会科学などの一層の連携が求められている。従って既存の手法のみならず、エコインフォマティクスならではの新しい情報学的、知識工学的手法の研究・開発が求められていると言え、今後必要となると考えられるより深い学際研究についても提案する。

2. 生態系・生物多様性のデータ

2.1 観測・観察

生物、生態系の対象は野外に存在するためデータの取得には様々な困難が付きまっていたが、電子デバイスを活用した測定機器の発達によって、この問題は機器の設置撤収、メンテナンス時へと限定されつつある。

データロガーと呼ばれるセンサーや機器の電気出力を記録する機器も珍しいものではなくなった。省電力消費機器を大容量化された電池で長期間動作させ、データは近年飛躍的に進歩したフラッシュメモリーに記録させることが可能となった。これらは有線や無線のネットワークを利用してリアルタイムにデータを収集することもできる。

単純なデータだけではなく、より複雑な構造をもったデータ、例えば DNA, RNA, 同位体、フラックス情報なども生態系・生物多様性研究で欠かせないデータとなってきた。

これら技術革新の相乗効果により、生態系科学研究にも大量データの時代が到来しつつある。

2.2 データベース

電子技術革新前においても、様々な種類の生物学・生態学に関わるデータが収集されてきた。例えば集計表やチャートに

記録された環境条件や生物学的属性、印刷された種名リスト、地形図に手書きされた観察地点、あるいは標本そのものなどである。こういったアナログな資料のデジタル化が進んでおり、前述の現在データと合わせてデータベース化が進んでいる。

大規模なデータを格納し、検索によって必要データを提供する直接的なデータベースはもちろん、存在する様々なデータベースを示すカタログ、メタデータベースのプロジェクトが急速に重要性を増し、大規模化が進んでいる[小川 2007, 小宮 2009]。研究や観測のネットワーク・プロジェクトが同時にデータベースなどを整備している場合が多い。国内国外で著名なものを表 1 にリストアップした。これは主なものであり、大小さまざまなサイトが存在し、相互にリンクされている。

表 1 国内外観測ネットワーク、データベースの例

略称	内容	URL
GBIF	生物多様性	http://www.gbif.org
KNB	観測, 実験 データ	https://knb.ecoinformatics.org/index.jsp
ILTER	生態系観測 ネットワーク	http://www.ilternet.edu
JaLTER	〃	http://db.cger.nies.go.jp/JaLTER/
NEON	米国での生態系の観測 ネットワーク	http://www.neoninc.org/
DataONE	環境データ も含む	http://www.dataone.org/
OBIS	海洋生物	http://www.iobis.org
GEO Portal	地球観測	http://www.geoportal.org/web/guest/geo_home
FluxNET	フラックスサイトデータ	http://fluxnet.ornl.gov/obtain-data
J-IBIS	生物多様性	http://www.biodic.go.jp/J-IBIS.html

3. 解析

3.1 GIS(Geographical Information System)

従来手作業で行っていた地図作製とそれをもとにした解析を、コンピューターで行うものが GIS である。GIS は生態系科学においても、単なる観測地・地域を地図にするソフトではなく、生

物や生態の地理情報を高速に表示解析できるフィールド研究に欠かせないツールとなっている。この際、GPS の普及がこの動きを加速させている点も見落とせない。

3.2 RS(Remote Sensing)

GIS と同様 RS による解析も生態系科学分野では現在当たり前の技術になりつつある。衛星画像データの低価格化が進み、有名な Landsat 衛星シリーズの画像も複数サイトから無料で入手できるようになった。また航空機から撮影した空中写真もカタログを閲覧しながら、デジタルデータを入手できるようになった。

RS による分析は、画像分類による被覆分類、植生判別だけでなく、生理活性やバイオマスの量を推定することも可能になってきている。

3.3 メタ解析, モデリング

生態系科学の中では統計諸分野、特に空間統計解析という分野も大きな進展を遂げている。

データの大規模化に伴い、解析手法も必然的に急変した。メタ解析と呼ばれる植物や動物の様々な属性、特徴量を広く収集し、統計解析を行う研究がある。従来は出版された文献を相当な労力で収集しデータを入力する必要があった。しかし、観測・観測データベースが完備され、それがメタデータベースなどで接続されているため、メタ解析が容易となっている。

また従来、ある現象に対する数理モデルを開発した上で、実測データを使ってモデルの正当性が評価されることが多かった。しかし最近ではパラメーターフィッティングに代表されるようにデータありきの解析(Data-driven)が主流となりつつある。つまり予測や補間などの精度を高めるモデルの応用面に研究のテーマが多くなった。

4. エコインフォマティクス

4.1 現在と将来

フィールド研究、数理生態学、あるいは大規模データを伴う統計解析は一種のエコインフォマティクスであり、その重要性が薄れることはない。しかし、全く新しいニッチとして大量のデータを複雑な解析方法で処理する研究が顕在化してきた。

データの大量化・解析の複雑化は進展中であり、その相乗効果は予測することすら難しいが、以下のような、なるべく控えめなエコインフォマティクスの将来像が考えられる。

リアルタイムデータ、完備されたアーカイブデータはメタデータつきでデータベースに格納されている。研究者は自分の興味や社会的要請によって、大量データを処理する。例えば外来種の分布や温暖化による植生の変化などのモデルをビジュアルに開発し、二三の操作によってモデルに必要なデータがメタデータベースで検索され、パラメーターがチューニングされる。その際に、生態学オントロジーなどによって、適切に種名や環境条件が再解釈され、必要に応じて適切な変換や補間が行われる(例えば気象条件の単位変換や空間補間)。結果は自動的にされた不確定性評価とともに GIS 上に表示され、スライダーをつまむと 10 年後、100 年後の変化が表示される。これらの結果はすぐに論文化できるだけでなく、視覚的な資料が多いので政策決定者から環境教育まで、社会へアウトリーチも容易である。これらの結果は必要な調査や研究が示唆されており、研究者だけでなく、研究資金や調査資金を配分する意思決定者にも有益な情報である。

4.2 課題

データに関する困難性として生物・生態系特有の「多様性」「非平衡・非線形性」「観測・観察の制約」が挙げられるだろう。しかし問題はこれら 3 点だけではなく、むしろ無数にある。これらの解決策を情報科学や人工知能学から待つのではなく、エコインフォマティクスが積極的に解決策を提案することもまた必用とされている。

データの公共性にも焦点が当たっている。グレー文献ならぬグレーデータとも言うべきもので、半公開データが研究室や博物館に多数眠っていると考えられる。また研究機関の競争のため、公的機関が行った観測でもデータが秘匿される場合がある。

生態系や生物多様性の研究において、観測・観察だけでも少なくとも 1 年間は時間が必要な場合が多い。解析を終えて、その論文を出版するには少なくとも 2~3 年、場合によっては 10 年単位の時間が必要と言えらる。

こういった問題への様々な解決策が試みられている。フェアユースなどの条項を定めているデータベースもある。また、論文ではなく、「データペーパー」という基準で、観測業績を認めた上で、二次利用を促進する動きもある。

5. 学際的研究の方向性

これらの潮流は生物学、生命科学内の分野間の協力を促しているだけでなく、それ以外の専門の科学者との連携が求められている。特に求められているのは環境問題などに関する環境科学の分野であるといえるだろう。大規模な事業に際して、生物、生態系への影響を調査するアセスメントが義務づけられているが、エコインフォマティクスの進展による環境影響評価だけでなく、より進んだ保全計画が立てられる。また環境計画に生態系サービスや生物多様性の視点を入れることによって、都市計画との補完関係も見込むことができる。

エコインフォマティクスの分野ではデータベース、統計学的ツールは広く認知され、受け入れられつつある。しかし情報科学分野で先端的な技術の導入はまだ初期的段階であり、この分野への本格的な参入が望まれているとも言える。

しかし、バイオインフォマティクスと比較して、エコインフォマティクスが対象とするデータは前章で指摘したように様々な問題を抱えている。このことも含め、情報科学関連の研究者が生態系科学の研究に参加しにくい状況であることもまた指摘できる。

生物や生態系など種とらえどころのないシステムを情報学的対象として捉え直し、特に環境問題や社会の持続可能性問題と関連づけて、総合・統合的に研究する場が強く求められている。さらにこういった場では、情報科学、人工知能科学そのものの新しい発展も十分期待できる。

成長したエコインフォマティクスが描く生態系像は、「ただの動く地図」以上のものであるに違いない。

謝辞 先端研究助成基金助成金(最先端・次世代研究開発支援プログラム)による支援を受けた。

参考文献

- [小川 2007] 小川 安紀子, 藤原 章雄: USLTER のエコロジカル・インフォマティクス技術の動向. 日本森林学会誌, 89, 360-364 (2007)
- [小宮 2009] 小宮 圭示, 真板 英一, 小川 安紀子: 長期生態系モニタリングデータの蓄積と利用について: JaLTER の取り組み. 北方森林保全技術 27, 16-19 (2009)