



特集1 発足記念シンポジウム

グローバル情報ネットワーク (GIP)

原 正一郎

グローバル情報ネットワークについて発表させていただきます。ASEAN 研究プラットフォームが「人のネットワーク」を担当しているのに対して、グローバル情報ネットワークは「情報ネットワーク」を担当していると言えます。そのグローバル情報ネットワークの中身ですが、旧地域研がミッションとして進めてきた「資源共有化プロジェクト」を、新研究所の発足に合わせて発展させた研究プロジェクトと捉えています。その目的は大きく二つです。一つ目は、情報学を一つのドメインと見なした地域研究の確立です。もう一つは、今述べた目的と、これまでの資源共有化プロジェクトの展開を図るための、新しい情報基盤の再構築です。以下、簡単に説明させていただきます。

地域研究における、これまでの情報学といいますと、データベースの構築やデータの可視化など、いわゆる情報技術による研究支援が中心でした。これに対して、情報学を一つの研究ドメインと見なした地域研究とは、情報学の方法論を使って、これまでは無かったような地域研究を試みようというものです。キャッチフレーズは、「情報のレンズで地域を俯瞰する・理解する」です。情報のレンズとなる道具立はもちろんコンピュータです。これが得意とするのは、大量のデータを高速で処理できること、デー

タとプログラムが同じであれば同じ結果を再現できること、結果の善し悪しに関わらず、それを得るに至った経緯を明確に示せることです。このような特徴を活かして、次の4つの課題を設定しました。

最初の課題は、人手では読むことも分析することともできないほどの大量のデータ、最近ではビッグデータと呼ばれていますが、具体的にはネットワーク上の地域研究に関連するビッグデータから、何らかの知識を自動的に取り出してみようというものです。これは個別の課題というよりも目的そのものです。二つ目の課題は、知識を蓄積するデータベースの構築です。知識ベースと呼ばれることもあります。ビッグデータを分析するには、機械学習や推論などの情報処理が不可欠です。とくに言葉が主体である人文社会科学が対象となる場合は、基礎的な語彙の意味や語彙の関連をコンピュータが使える形で蓄積し、それを検索したり推論したりできる機能が必要です。このような機能をオントロジーと呼んでいます。三つ目の課題は、ネットワークから収集するビッグデータに知識ベースを適用して、実際に知識の抽出を行う機能の構築です。データの内容を自動的に分類したり、似た内容のデータを検索したり、曖昧な命令に基づいた検索をしたりできる機能の実現を目指しています。四つ目の課題は、以上の機能を組

み合わせて、計量的な分析を行うことです。たとえば、異常な状態の早期検出や、シミュレーションなど併用した将来予測などができればと考えています。

「情報のレンズで地域を俯瞰する・理解する」のイメージです。ネットワーク上には、地域研究に関する断片的データが浮かんでいます。それらを、ここではジグソーパズルのピースとして表しています。グローバル情報ネットワークの研究は、ネットワーク上から使えそうなピースを探し出すことから始まります。ここで、先ほどの知識ベースや推論の機能が発揮されます。これらのピースを繋いで新しい絵を作り出す作業が、計量的分析となります。幾つかの想定される分析ケースを示しています。

最初のケースはエネルギー分野への応用です。たとえば、化石燃料と太陽光発電といったエネルギーの構成比を一定量変更してみる、それに必要なコストはどれほどか、その結果として二酸化炭素排出量はどれだけ減るか、その成果と考えられる経費節減や医療費削減や QOL の改善はどれほど期待できるか、これらについて、シミュレーションを中心に様々な視点から分析して、都市計画などへの反映を目指します。次の保健医療の場合、たとえば地理情報システムを駆使して様々な時空間データを重ね合わせてみたら、疾病の発生や伝搬に関する空間パターンが見えてきて、ここから新しい研究仮説を構成したりすることができるかもしれません。さらに計量モデルを作って、疾病のリスク評価やパンデミックの予測などを行うことができないかと考えています。秩序変動、これは政治学的な研究テーマを想定しています。従来の計量政治学では統計的な分析が主体でしたが、例えば、様々なデータに機械学習を応用してみたら、これまで気がつかなかった現象のパターンや因果関係を見つけることができるかもしれません。将来予測などができればさらに面白いなと考えています。これらは、まだ「絵に描いた餅」で、今のところは基礎的な解析手法の実装を開始した程度です。最終的には、これらの機能を組み合わせることにより、地域をさまざまな視点から、さまざまなサイズで、計量的に俯瞰あるいは理解することを目指しています。やや夢の世界です。

いずれにしても、データベースがなければ仕事できませんので、情報基盤の再構築が先決です。こ

れがグローバル情報ネットワークのもう一つの目的です。最初の目的がやや夢物語あるのに比べますと、こちらはかなり現実的な話です。旧地域研では、情報基盤を使って様々な研究支援を提供してきましたが、主要なものは2つです。一つ目はデータベースの構築を支援するツールで、My データベースと呼んでいます。My データベースは、データベースシステムについての専門的な知識や技能がなくとも、簡単な操作だけでデータベースを公開できるツールです。旧地域研の殆どのデータベースは My データベースを使って作られています。もう一つは、インターネット上に分散している地域研究関連のデータベースをひとまとめに検索する機能で、資源共有化システムと呼んでいます。新研究所と人間文化研究機構に加えて、カリフォルニア大学バークレイ校やハーバード大学など、国内外の地域研究機関が公開している約 50 個のデータベースの統合検索を実現しています。グローバル情報ネットワークでは、セマンティック技術を使って、これらのサービスの再構築を進めています。これによって、オープンデータやオープンサイエンスなど、学会や社会から強まりつつあるニーズに応えていきたいと考えています。

新しい情報基盤の概要です。これまでの情報基盤は、論文など、学術成果として整理された情報を対象としてきました。ところが最近では、地域研究に関わる最新データのかなりの部分はサイバー空間上に流通しています。災害情報や選挙情報などはその典型例です。これらのビッグデータを使って地域研究を進めるということは、まずサイバー空間からデータのピースを集めて、それらを資源共有化システムに蓄積されている論文などの既存データと繋げて新しいビッグデータのセットを作り上げて、そこに機械学習や統計などの応用プログラムを適用して分析することです。しかし、データのピースごとに、構造はバラバラ、語彙もバラバラ、言語もバラバラで、このままでは繋ぐことができません。そこで、先ほどの知識ベースのオントロジーを使って、バラバラなデータを繋ぐこと目論んでいます。新しい情報基盤ですが、これまでのところデータの入れ物ができあがり、既存のデータ移行までは終了しています。現在、知識ベースと応用部分の設計を進めています。プロトタイプを早い段階で作り上げて、ビッグデータを利用した研究事例、たとえば地域の状況を可視化するモニタリング、統計的分析やシミュレーショ

ン、それらを発展させた異常検知などを試みていきたいと考えています。

グローバル情報ネットワークは、夢のような目標を掲げています。まだ始まったばかりで、海のものとも山のものとも分からない状況ですので、2、3年

ほどは温かい目で見守っていただければと思っています。また応用に関しては学内外の研究者との広汎な連携が不可欠です。これからもグローバル情報ネットワークへの御協力をお願いいたします。ありがとうございます。

＊「グローバル情報ネットワーク」Web サイト：<https://japan-asean.cseas.kyoto-u.ac.jp/glocal-2/>

＊本記事は、2017 年 6 月 2 日に京都大学で稲盛ホールで開催された京都大学東南アジア地域研究研究所発足記念シンポジウムでの講演内容を収録したものです。