

研究推進潜在力

河本 大知

京都大学 学術研究支援室

kohmoto@kura.kyoto-u.ac.jp

Research Potential

Daichi Kohmoto

Kyoto University Research Administration Office

概要

研究推進潜在力（本稿では Research Potential をこのように訳す）の可視化・分析は、研究機関の研究推進戦略策定時に不可欠な情報・データである。本講演では、本学学術研究支援室に設置の分析チーム A2D（Analytics, Analysis, and Design Team）の業務から、研究推進潜在力の可視化・分析に関して、可能な手法ならびに事例・結果をいくつか紹介・議論する。本稿では、講演の背景として位置付けられる、私たちの見えている世界の一端について簡単に纏めている。

1 A2D とは

研究機関にとって、ビジョンに基づいた研究推進力の向上は重要なミッションの 1 つである。新たな方向性の策定・決定ではビジョンに沿った的確な現況把握・分析が要であり、私たち分析チーム A2D（Analytics, Analysis, and Design Team^{*1}）は、大学経営・研究者環境改善を念頭においた、本学ビジョン・ミッションに沿った形での研究力強化に向けた施策立案に資する能動的提案を行なっている。業務の軸は、目的・研究者指向のゼロベース思考によるトップダウン分析・組織デザインの観点であり、データに基づくボトムアップアプローチも併用している。

私たちが分析の際に用いる際には、具体的なデータだけでなく、形に落とし込めていない抽象情報や雰囲気といったものも分析の対象としている。そこから行える可能な構造化・モデリングをベースとして、今後に向けた提言を行っており、積み上げている。なお、ここで言うところの分析は、所謂「予測分析」とは異なることに注意する。

2 書誌データ分析

多くの研究者は研究成果を論文等に纏め学術誌に投稿し、採択されれば掲載される。掲載論文はデータとして出版社のデータベースに蓄積されており、これら

のデータを分析することで世界中の研究者・研究機関の現況を一挙にある程度知ることが可能である。有償の書誌データのデータベースとして、現在 Clarivate Analytics 社（旧 Thomson Reuters 社）の Web of Science、そして Elsevier 社の Scopus の 2 つがよく知られており、これらのデータベースに含まれる被引用数等のデータ・指標を使った研究パフォーマンスならびに研究推進潜在力分析が世界中で行われている。無償のデータベースの代表的なものとしては、Google Scholar が挙げられる。

一般に論文等は掲載・出版までに時間が掛かり、タイムリーな研究パフォーマンス・研究推進潜在力を測るには不向きである。また、そもそも研究成果を論文等以外の形で発信・蓄積する研究分野も存在し、当然ながらこの種の書誌データのデータベースにはこれらの研究成果は含まれていない。従って、特に本学のように多くの研究分野の研究者が集う研究大学の場合、何んらかの単一手法で網羅的に研究パフォーマンス・研究推進潜在力を測ることは不可能であり、分野や部局に応じて分析の手法や指標の選択を行う必要がある。少し話が逸れるが、例えば世界大学ランキングでは、研究パフォーマンスを測る指標として、書誌データの指標と併せて、世界中の一定層の研究者に対する評判調査の結果を（指標値算出アルゴリズムは公開されていないものの）指標・数値化し採用しており^{*2}、そ

^{*1} A2D についての更なる情報については、次を参照のこと。
URL:<http://goo.gl/eALmCf>

^{*2} 例えば Times Higher Education 世界大学ランキングにおける Methodology を参照: <https://goo.gl/ePRJjv>

それぞれの結果は一定程度に相補的である。

研究パフォーマンス・研究推進潜在力を測る際に、書誌データを殆ど用いない研究機関も存在する。例えばドイツでは書誌データ分析を行う専門機関 iFQ (Institut für Forschungsinformation und Qualitätssicherung) が存在し、研究大学はレポート等の形で提供される iFQ の分析結果を必要に応じて参照している。ドイツで多く見られる研究パフォーマンス・研究推進潜在力分析の手法は、プロジェクトの報告書やインタビュー等を通じた研究者によるピアレビューである。これは最もコストと労力が掛かる手法であるが、最も効果的に研究パフォーマンス・研究推進潜在力を測ることができ、研究者も安心して評価を受けられる等の状況を実現している。

ドイツの事例は、結局のところ最初に述べた書誌データを用いた研究指標も、評判調査の結果を数値化したランキング指標も、所詮本来測るべき研究パフォーマンスや研究潜在力の非常に粗い近似に過ぎず、ひいては研究者から「遠い距離」にある分析手法であることを示唆している。

3 人事・財務データ

人事・財務データと書誌データないし他の研究成果の情報とを紐づけてモニタリング・分析することで、研究機関の研究パフォーマンス・研究推進潜在力をある程度正確に把握できる。研究機関にとって最も大切なものは人材であって、人（研究者）を中心に適切な研究者環境や研究拠点が構築されていく。この意味で、研究機関に流入する若手を中心とする「人材の流れ」を把握すること、マーケティングも研究機関ひいては分析チームの重要な仕事であり、研究パフォーマンス・研究推進潜在力分析と大きく連動する。

「人材の流れ」の観点から、人材交流や協働状況の正確な把握も可能になる。実際のところ、例えば書誌データのみからでも共著論文等を調べることで、ある程度の協働状況の把握は可能であるが、前述の通り論文は掲載・出版までに時間が掛かるので、書誌データだけではタイムリーな研究パフォーマンスや研究推進潜在力について把握できない。近年協働状況は人材交流に裏付けられ把握できるケースも多く、この意味でも人事・財務データは私たちの目的には不可欠である。

4 モデリングの精度

データに基づいたモデリングは所詮現実世界の非常に粗い近似に過ぎず、多くの場合、そのままでは分析

目的に含まれる「捉えたい側面」を捉えるに足るモデルにはなっていない。それを補い肉付けする上で、研究者を取り巻く、データに落とし込まれていない様々な形で存在する「情報」や「研究の最前線の雰囲気」といったものも重要である。そのような情報は、分析者が様々な機会を通じて能動的に獲得に動くことで得られることが多い。また、最近では様々な機械学習の手法や大規模データ分析基盤 (Apache Spark 等)、統計的手法も、それぞれの角度からのモデリングの助けになる。

良いモデリングは、必ずしも現実世界の「精密な近似」を意味しない。分析目的、ひいてはビジョンやミッションに叶うモデリングであれば良く、これらに応じたモデリングの精度と労力とのトレードオフも分析業務において重要な観点である。

5 人的ネットワーク

データや情報、研究の最前線の雰囲気といったものは、全て何らかの人的ネットワークが無ければ手に入れることができない。また、大抵の場合、分析に用いるこれらのデータやネットワークが最初から整備されてはおらず、仮にデータ等が揃っていたとしても、データのクリーニングや名寄せに労力が掛かることは他の業界のデータ分析と同じである。

研究パフォーマンスや研究推進潜在力といったものには、フォーマルな定義はなく、いずれもビジョンやミッションから定まるものである。そして往々にして研究に関わるデータは思いの外少なく（「スモールデータ」）、目的に叶わないモデリングに終わる事例が多い。そのような状況でも、諦めず何らかの道筋を策定し、研究大学の成長に資する策を考え質の高い仕事をするためには、やはりビジョンやミッションを共有する同志から構成されるチームが必要である。私たちの業界においても、この種の人材は世界的に不足しており、私たちのチームでも私たちのビジョン・ミッションに同調してくださる協働者、もしくはプロジェクトにジョインしてくれるメンバーを常に探している状態である。