

■西澤正巳 情報学プリンシプル研究系 准教授

【タイトル】

論文データから優れた研究の原点を探る

【本文】

将来への可能性をもつ研究開発への適切な投資ができるようにする。これが私の研究目標です。このため、優れた成果を出している研究が、どんな支援・資金を受けたかを調べ、成果の原点が何かを研究しています。

成功した研究の履歴をたどる

具体的には、公開されている研究論文のデータベースを使って、過去の科学研究の流れを調べています。

例えば、ある優れた研究について、その研究者グループ（学部・学科・専攻でいえば、学科ごらの単位）が過去にどんな補助金を受けたかといった履歴を調べると、原点がどこにあったのか、どんな道をたどって成功したのかがわかります。

初期段階の研究を見て、どれが成功するかがわかるようになるわけではありません。しかし、支援する側の立場からは、何を重点的にやらなければならないか、ということが見えてきます。

言語処理技術を駆使して研究の関連性を調べる

優れた研究を探したり、その系譜をたどるには、論文の引用状況や、同一著者が書いた論文の履歴を調べます。

ある論文に別の論文が引用されているとしても、その引用のされ方がネガティブかポジティブかを見なければいけません。それに、引用された論文の主な著者が誰かを判断しなくてはなりません。著者が同姓同名というケースもかなりあるので、所属や論文のタイトルなどで区別します。論文同士の関連は、各論文で使われているキーワードのつながりの類似度などを言語処理的な方法で比較して見ていきます。

そうした作業を経て、論文数やキーワードの数などを取ると、割合きれいな統計分布になります。その分布から外れた言葉を除いたり使ったりして検索していくと、いろいろな関連がわかります。

各大学の国際競争力への寄与や特色がわかる

時代によってキーワードの意味が変化することもありますし、同じ言葉が分野ごとに違ったりすることもあります。こうした変化にはまだ手をつけていません。また、ある研究について一定数以上の論文が出揃うのは実施時期からしばらく経ってからですし、論文が引用されるまでには、発表後 2、3 年かかりますから、引用状況を調べる場合は、少し長い目で見ることになります。

こうした困難もありますが、大きく言えば、日本の国際競争力の強化にどの大学がどの程度寄与したか、さらにその中でどの学科がどんな影響を与えたか、などがわかります。最近のデータを見ると、全体的には東大、京大の貢献が大きいのですが、研究分野を絞り込んで見ていくと、単科大学や地方大学にも強い分野があることがわかります。さらにデータを細かく見ていけば、

ある研究が当初の想定以上に成功した理由がわかるかもしれません。

(取材・構成 齋藤 淳)