

# ESSAY

## 科学論理とデータサイエンスの流れ

山岡和枝

帝京大学法学部

最近、科学論理に関連する本を何冊か読んだ。長らくこの種の本は読んでいなかったが、datamining and knowledge in databases という論文も同じところに偶然いただき、ふと、この二つのことについて書き留めておこうという気になった。

アリストテレスによって体系化された三段論法をはじめ、現代の科学的考え方の多くは、演繹論理に基づいている。ご存じのように、演繹論理はいくつかの前提から1つの結論を導くやり方である。一般から個別へという命題の流れのなかでは前提が真である場合、結論も必ず真になるという単純なものである。たとえば、もし治療した者すべてがA治療法を行った者であり(大前提)、治療した者すべてがHLA〇〇型の遺伝子をもつならば(小前提)、A治療法を行った患者はHLA〇〇型の遺伝子をもつという結論が論理的に導かれるという論理構成である。

さらに、オーストリア生まれのイギリスの科学哲学者ポッパー(Karl Raimund Popper 1902~94)が「探求の論理」(1934)において、科学は基本的に帰納的学問だというそれまでの常識をくつがえし、「反証可能性」という考えを提出してから久しい。彼は、科学理論というのは仮説にすぎず、その仮説のもとで観察によってそれぞれの言明を確かめるのであり、もし別の観察によってその言明がまちがいだとわかった場合には、この科学理論

はただちに棄て去られる。

このような反証の試みによって理論がまちがいに限らない限り、この理論は受け入れられる。しかしその理論はあくまで仮説にとどまり、確実な科学理論などどこにも存在しないと主張した。これも演繹の一つの流れである。

そして、現代の医学の世界での検証に欠かせないのがこの演繹に基づいた「仮説検証」の考え方であり、パラダイムの構築である。もちろん医学の分野のみにとどまらないが、演繹の積み重ねによって現代医学は構築されてきているといっても過言ではないだろう。

しかし、と思う。現代ではこの演繹の考え方の方に大きな比重が置かれ過ぎてはいないだろうか。演繹に対して、フランシス・ベーコンにはじまり、ジョン・スチュアート・ミルの「論理学体系」(1843)によって完成された帰納論理があった。これは個々の事例から一般的な法則を導き出すという、演繹とは逆の方法であり、観察されたいくつかの事例にあてはまることは、観察されてはいない同じ種類のほかの事例にもあてはまるという考えに基づくものである。

現代の科学論理はさまざまな考え方があるが、帰納と演繹が両輪となって発見と確実な構築がなされるのではなからうか。演繹は、新しい情報はもたらさないが前提から必然的に導かれる命題を明らかにするというところで説得力があり、大きな意味をもつ。しかし、

限られた経験からより多くの内容を含む結論を引き出すという帰納的な推論により、多くの新しい発見がもたらされていることを忘れてはならないと思う。

帰納による確からしさは、観察された事例の数に左右されることになる。たとえば、限られた人数を調査し、それを母集団にあてはめる調査は、まさに帰納論理に基づくものといえよう。しかし、帰納には飛躍が伴い、ときには誤った飛躍を導き出すことも避けられない。これらを補完するのが確率論であり、統計学であったともいえよう。

はたして、現在の医学の世界ではいかがなものだろうか。医学では「仮説検定」という演繹的考え方が中心となっており、帰納的な考え方はなかなか受け入れられていないのではなかろうか。もちろん人の生命をあずかるのであるから確実なものが必要とされることがその一つの理由であろう。

しかし、ジョン・スノー (Snow J) の1854年にロンドンでのコレラの流行を終息させたエピソードにしても、コレラと水道水との関係の発見は何人かのコレラ患者が同じ水道会社の水を使っていたということから、帰納によりもたらされたのではなかろうか。この結果、彼は世帯調査によってS&V社の水道がコレラによる死亡と関連するという仮説を実際のデータから検証した。帰納により発見し、演繹により検証したという典型的な事例といえよう。

しかし、現代ではこの帰納が表に出にくくなっているのではないだろうか。薬の副作用などにしても臨床医の、形にはならないかもしれないが、直観的な情報がうまく収集されれば、もっとすばやい対応をもたらすことができるのではないだろうか。現代の医学の世界においては、個人には直観的な帰納的論理があるにもかかわらず、演繹に重点がおかれ過ぎているのではなかろうか。

データの科学においてはデータライブラリーにはじまり、データベースからデータウェアハウスの構築が叫ばれはじめている。データウェアハウスとはどんな種類のデータでもとにかく倉庫に詰め込んでおいて、あとから必要になったときに取り出せるようにしておくというものである。たとえば治療や投薬にまつわるさまざまな情報、治療や投薬の副作用や効果など、どんな些細な情報でもとにかくなんらかの形で保存しておき、それからなんらかの有用な情報を引き出していこうというものである。

現在のデータベース化においては情報の精度や正確さ、入力方法などが一定の形式や制約に基づいて集約されているが、そのさいに臨床医の印象や手書きで書きとどめておく程度のものであればそれらの情報は集約されないことになる。そのなかにはたとえば副作用に関する重要な情報が含まれているかもしれないし、あるいは効果的な治療法に対する情報が含まれているかもしれない。それらが生かされないまま消失していつてしまっている可能性もあるのではなかろうか。

データウェアハウスとはそのような情報をなんとかして掘り起こそう (data mining) という考えからはじまったものである。そのような情報にはある程度のあいまいさが含まれ、それと同時にデータ提供者のモラルに依存することが多くなってくる。データから情報を取り出す側の情報に対するセンスと知識が要求されるようにもなってくる。しかし、現にそのような動きがアメリカをはじめとし起こってきているということは、現在の情報の収集方法の限界を感じはじめた人々がいるということであろう。

これまで科学の中心を占めてきた演繹に基づく仮説検証という考え方そのものに対する人間のバランス感覚が働き出したのであろうか。ふと、そんなふうに思った。