

# ビッグデータと心理学の接点を考える

兵庫県立大学社会情報科学部 教授

土方嘉徳

## ビッグデータと心理学の出会い

心理学とビッグデータ、相性がよい組み合わせのようにも思われるが、科学研究の方法論として一抹の不安も感じる。多くの心理学者は、このように感じているのではないだろうか。私はもともと情報科学の研究者であったため、ビッグデータを解析して何か知見を得ようというアプローチの研究をこれまで何度も行ってきた。しかし、研究対象をコンピュータではなく、人にしようとしたとき、すなわち心理学の分野に足を踏み入れた途端、その非力さにあっという間に失望することになった。本稿では、それについて話をするとともに、利点についても話をしたい。

## 工学的アプローチと科学的アプローチの違い

ビッグデータを信奉するのは主に情報科学の研究者である。情報科学は、電気工学や電子工学、通信工学に端を発する、発展目覚ましい分野である。すなわち工学の研究分野が基になっている。以前の情報工学科はほとんどが工学部に設置されていたが、この30年ほどで独立の情報学部や情報科学部になったケースが多い。工学の分野は、処理速度が速く、処理が正確で、扱える処理量が多ければよいという決まった評価指標をもつ。そのため、理論的背景よりも手法の性能向上に重点が置かれる場合があり、研究課題（Research

Question: RQ）が明示されないこともある。例えば、深層学習の一種である深層畳み込みニューラルネットワーク（Deep Convolutional Neural Network: DCNN）は、画像認識において飛躍的な精度向上を実現したが、その内部でどのように画像認識が行われ、なぜ高い精度が得られるのかを理論的に説明することは依然として難しい。このように、手法の有効性が先行し、理論的な理解が後回しになるケースもみられる。一方、心理学や経済学、物理学のような科学研究の分野では、これまでの研究の知見に基づき、何が分かっている、何が分かっているかを考え、それをRQとする。この基本的な研究の価値観や研究に対する取り組み方の違いにより、ビッグデータに対する評価も変わってくる。

## ビッグデータ研究の事例紹介

私は、Webとソーシャルメディアの研究者なので、ここからはソーシャルメディアにおける行動心理分析という研究を題材にして話を進めたい。この分野で代表的なビッグデータは、Twitter（現、X）のツイートデータである。情報科学の研究者は、まずはこのビッグデータを手に入れたところから研究を始める。解きたい課題があるわけではなく、まずはデータを見たいのである。そして、多くの研究者は、Twitterのランダムサンプリングデータを集めると

ころから始める（今はXとなり、簡単には収集できなくなった）。そして、このデータに対して、工学的な新しい分析手法を適用することで、何か面白い発見があるかどうかを探るのである。例えばTwitterでは、フォロワーの多いユーザはフォロワーの多いユーザと相互フォローになっているという社会的な同質性があることを示したり<sup>1</sup>、ツイート中の感謝の表現から犯罪の起きやすい地域を推定できることを示したり<sup>2</sup>してきた。これらの論文には、細かい研究の手順まで記されていないため、完全には分からないが、著者らは得られたデータを分析していく中で、友人数や感謝の表現に注目したのかもしれない。

## 心理学的視点からの課題

一方、心理学者は、まず心理学の研究分野で蓄積されてきた理論的背景に注目する。犯罪の起きやすい地域の推定であれば、犯罪心理学におけるソーシャルサポートという概念に注目するであろう。そして、ソーシャルサポートにおける互恵の意識が、外部者の攻撃意欲を緩和するかどうかというRQを立てる。このRQを解くためには、互恵とは何か、それを意識するとはどういうことかという構成概念（観察が難しい心理的特性や状態を理論的に定義し、測定可能にするための抽象的な概念）の定義を行い、それを取得するため

の質問を設計する（あるいはすでに提案されている心理尺度を用いる）。そして、その調査票を用いて社会調査を行う。すなわち、RQが決まってから、それを明らかにする調査方法や実験方法を設計し、それに従いデータを取得するのである。

このように、分析結果は同じになるかもしれないが、理論的背景や理論的枠組みの深みにはかなりの差がある。ビッグデータは、行動の結果だけを見ているので、2種類（また2種類以上）の行動の間の相関しか分からない。一方、社会調査や実験では、心理特性と行動の両方を取得することが可能となり、心理と行動の間にある因果関係をもつモデルも検証可能となる。すなわちビッグデータ分析では、行動データの背後にある心理を知りたくても、それを知ることが極めて困難である（ユーザにコンタクトすれば可能だが、倫理的な問題が発生するうえに、タイムラグからそもそも、そのときの心的状態を取得することはできないであろう）。ここにビッグデータの最大の欠点がある。すなわち、ビッグデータ中の1つのレコード（データ）と、それを生み出したユーザとのリンクに大きな問題があるのである。

### ビッグデータの利点

しかし、ビッグデータは短所ばかりではない。それについても説明する。上述の犯罪の起きやすい地域の推定では、住人と外部の犯罪者という2種類の人間が存在し、この間の関係を分析することになる。そのため、社会調査に基づく方法では、この両者に質問をしなければならない。住人はともかく、外部の犯罪者に質

問をするのは不可能に近い。社会調査では明らかにできないような関係を推定できるところがビッグデータの長所ともいえる。また、感謝の表現は普段の生活で得られたものであるのも、ここにバイアスはほとんどかかっていない。しかし、社会調査で「あなたは、近所の人を助けていますか?」のような質問をすれば、社会的望ましきバイアスが回答に乗ってくると思われる。実際の環境で発生した行動データであるということもビッグデータの長所ともいえる。最後に、近年は深層学習の発達により、機械学習によるテキストの極性分析や感情分析の精度が大きく向上している。そのため、ツイート本文から個々の感情を高精度に分類できるようになってきており、ユーザ数が十分に多ければ、社会全体としての感情の傾向や極性の分布を推定する手がかりにもなりうる。これらの利点を考えると、ビッグデータには不安があるからという理由で、それを使おうとしないのは、かなりもったいないと言える。

### 今後のビッグデータ研究のあり方

ビッグデータを用いた研究に、ある種の不信感を抱くのは、その手の研究における科学研究としてのシナリオの未熟さだろう。具体的には、基盤となる心理学の理論や、研究ギャップ（research gap: これまでの研究で十分に検討されていない、または解明されていない内容のこと）に基づくRQ（research question: 研究で明らかにしようとする学術的な問い）、仮説、そして得られた結果に対する議論が十分でないからだと思われる。これは、工学が起点となる

情報学と科学が起点となる心理学の差によるところが大きい。データに含まれていない構成概念を研究シナリオに含めることはできないが、対象の学問分野の文献調査をしっかりと行っていれば、この不安はある程度解消されるだろう。

また、情報科学では100万人や1000万人といったユーザ数が、分析のサイズ感になるが、果たしてこれだけ多くのデータが実際に必要なのであろうか。それよりは、1000人程度の規模であったとしても（社会調査や実験の参加者数と考えれば十分に多い）、本人の承諾を経て、SNSの行動データと社会調査の心理データの両方を取得していれば、高度な心理モデルの解明にも耐えうるデータになるのではないだろうか。今後は、研究倫理を考慮したうえで、小さくてもよいので、より強いデータを取得することを意識したほうがよいだろう。研究の方法は一つではない。自分が興味をもった研究課題やRQを解くことができそうなのであれば、既存のビッグデータを使ったり、自らSNSのデータを収集したりすることも検討することをお勧めする。心理学と情報科学は、異なる哲学をもちながらも、相補的に働こう。両者を架橋する研究こそ、これからの人間理解に必要とされているのかもしれない。



ひじかた・よしのり

大阪大学大学院基礎工学研究科システム人間系専攻博士課程修了。博士(工学)。専門は計算社会科学・社会心理学。著書に『ソーシャルメディア論：行動データが解き明かす人間社会と心理』（単著、サイエンス社）、など。

1 Kwak, H. et al. (2010) What is Twitter, a social network or a news media? *Proceedings of the 19th International World Wide Web Conference (WWW'10)*, 591-600. 2 White, A. M. et al. (2018) Does reciprocal gratefulness in Twitter predict neighborhood safety?: Comparing 911 calls where users reside or use social media. *Proceedings of the 12th International AAAI Conference on Web and Social Media (ICWSM-2018)*, 700-703. \*COI:本記事に関連して開示すべき利益相反はない。