

---

# 2014年総会シンポジウム「ビッグデータの可能性と課題 ——監視・シミュレーション・プライバシー」・論文

---

## 世界表象としてのビッグデータとビッグデータ・ガバナンス ～ 局所表象と分散表象の観点から ～

### Governance for Big Data as Representations of the World ～ Local Representation and Distributed Representation ～

キーワード：

ビッグデータ, 局所表象, 分散表象, 専門職制度, 技術ガバナンス

keyword：

Big Data, Local Representation, Distributed Representation, Professionalism, Technology Governance

静岡大学情報学研究科 吉田 寛

Shizuoka University, Department of Informatics Hiroshi YOSHIDA

---

#### 要 約

本稿は、ビッグデータに対する推進論や警戒論に対して、その議論の基盤となるような理論的観点を提案する。私はビッグデータを、世界についての情報を表現した「表象」とみなし、ビッグデータ・テクノロジーを世界についての表象を取り扱う技術であると考え。表象は解釈によって意味が確定する記号や表現であり、従ってビッグデータを扱うテクノロジーの評価とコントロールについては、こういった解釈をめぐる人文・社会的なアプローチが必須である。

表象について、認知科学や心の哲学では局所表象と分散表象という区別がある。分散表象とは、意味的な単位が分解されて、システム全体の中に分散的に存在している表象のことであり、その処理過程を意識化することは不可能である。ビッグデータがこうした分散表象として処理される場合、データの利用者にとっても処理の過程を意識化することはできない。従って、専門職としてのデータ・サイエンティストであっても、このテクノロジーについての透明性や説明責任を保証することは難しいだろう。そこで、分散表象的なビッグデータの社会的コントロールについては、専門職や政府、データ利用者だけでなく、データ対象となっている市民の参加を保証した、参加ガバナンスの構築が必要であると思われる。

## Abstract

I want to provide the base for debates between pro and con discussions on Big Data technology. My suggestion is to see Big Data as representations of the world, and to distinguish distributed representation from local representation. We can not be conscious of the distributed representations because they don't exist in the form of meaningful units. This is why responsible control of Big Data as distributed representations is very hard. Even Big Data professionals can not be responsible enough for controlling distributed Big Data. We should be more careful about the distinction between local and distributed representation. We also should construct open and effective Big Data governance which includes such actors as governments, companies, professionals, and public citizens.

## はじめに

本稿では、「ビッグデータ」をめぐる素朴な楽観論や悲観論には距離を置いて、ビッグデータの「表象」としての側面に光を当てることで、理論的ないし哲学的な観点から、このテクノロジーに対して社会情報学的に批判的に再考するための切り口を提示する。その上で、ビッグデータ・テクノロジーを社会的にコントロールする上で課題となるであろう論点について、テクノロジー・ガバナンスの観点から見通しを述べる。

## 1 ビッグデータ言説

### 1.1 ビッグデータ言説の混乱と課題

「モバイル」、「web2.0」、「ユビキタス」、「クラウド」などの流行語に続いて、2012年ごろより「ビッグデータ」という言葉が注目されるようになってきた。こうした一連のキーワードは、コンサルタントやライターたちのつくり出したパスワードにすぎないのかもしれない。しかしこうしたキーワードの変遷を、いつでもどこでもデータが産出され、そのデータが集約され利用されるようになってきたという大きな流れとして解釈するならば、「ビッグデータ」は確かに現在の情報社会の到達点をそれなりに捉えて表現していると言えるだろう。したがって、かりに「ビッグデータ」という言葉自体が一過性のパスワードに過ぎなかったということになったとしても、このテクノロジーの普及に伴う社会的課題自体は、現実的なものとして受け止めなければならない。

「ビッグデータ」をめぐる、巷には多くの言説が入り乱れている。2012年頃から『ビッグデータの衝撃』(城田, 2012), 『ビッグデータの覇者たち』(海部, 2013)といったタイトルで、ビッグデータの時代の到来を少々煽り気味に紹介するビジネス書が多く出版され、同様の趣旨の新聞記事も多く見られるようになった。2014年頃には

白書等でもビッグデータの活用が強調され、ビッグデータの具体的な活用法や活用事例の紹介、検討記事がメディアでも展開されるようになった<sup>(1)</sup>。

他方で、プライバシーや監視社会といった観点から、このテクノロジーに対する不信感や警戒感が表明されることもある。森健の『ビッグデータ社会の希望と憂鬱』(2012)は、インターネットや監視技術が人間の主体的な思考と意思決定を奪い、全体主義や衆愚政治を招きつつあるのではないかというオーソドックスなネット批判言説の延長線上にビッグデータ・テクノロジーを位置づけて警戒感を表明している。また、2014年6月の『現代思想』の特集「ポスト・ビッグデータと統計学の時代」における西垣通とドミニク・チェンの対談(西垣ら, 2014)においては、ビッグデータが因果関係でなく相関関係だけを見つける技術であるという理解の上で、西垣はビッグデータ・ブームに対して「反知性主義」を指摘して「根本的な疑問」を投げかけ、チェンもまた自動的なソフトウェアの更新に近い形で文化が駆動されていってしまう先には「虚無的な状況が広がっているのではないか」と不安を表明し、この状況の正体を言語化する必要を訴えている。

こうした賛否の言説は、もし慎重な検討を欠くならば、不毛な水掛け論、ないしすれ違いに終わってしまうことが懸念される。ブームを煽る言説はこのテクノロジーの光の部分にのみ言及し、不安を強調する言説は影の部分にのみ焦点を当てるといったパターンが懸念される。人文・社会研究者の警告がごく表面的にしか対応されずにテクノロジーが産業的に推進されていくことで、気がつく社会的な問題が蓄積し生活における人間性が脅かされてしまう状況になっているというのは、近代以降に繰り返されてきた技術史である。学としての社会情報学の使命の一つは、こうした表面的な言説の背後にある根拠や論理に光を当て、厳しく検討することで実質のある議論を社会的に提供

することにある。本稿は、こうした問題意識において、ビッグデータをめぐる言説の背後に隠れているロジックに目を向け、これを分析することで、曖昧なままに進行しているように見える社会的論議の現状の改善に寄与しようとする。従って本稿は、こうした脅迫的な積極論・歓迎論と頑強な拒否・警戒の立場、そして不安の入り乱れるビッグデータ言説の現状において、いずれかの立場を主張するものではない。データの性格に応じたより詳細で冷静な議論を可能にするため、「表象」という切り口から「ビッグデータ」の存在状態の分析を通して、より注意深い論議に向けた概念的区別を提案したい。

## 1.2 「ビッグデータ」の分類と分析

ビッグデータ・テクノロジーが扱うとされるときは、すなわち狭義の「ビッグデータ」と呼んでよいだろうが、これは一般に、大量で (Volume)、随時更新され (Velocity)、多様である (Variety) という性格を持つとされている。このデータをひとくくりせず、分類する観点には、すでにいくつか提言されている。

まず、データが人間に関するものか自然に関するものかという区別がある (西垣, 2014: p.46)。こうした区別を意識しない議論に対して、西垣は「初等的な混乱」として戒めているが、確かにそういった乱暴な議論は、例えばプライバシーへの懸念を人間の情報には関わらない「ビッグデータ」への批判にナイーブに拡大して、失笑を買ったり議論を混乱させたりしてしまうことだろう。逆に、人間に関わらないデータの分析技術の成果をそのまま人間や社会に対しても適用する乱暴な議論も同罪であろう。これに関連して、人間に関するデータに限定しても、それが個人にひもづけられたものなのか、匿名化されて集団的傾向の分析にしか使われないものなのかという区別も、データの社会的な利用を考える上で大きな違いをもたらす重要な区別である。こうした区別は、このテクノロ

ジーに関わる専門家にはすでに比較的意識されているように見えるが、いまだにこうした区別が無頓着な提案や議論は専門分化した研究者の言にもしばしば見られるし、メディアにあふれるビッグデータ言説では、こうした区別の一方だけを取り上げて分かりやすいメッセージにまとめられることが多いように見受けられる。

『現代思想』の同じ特集に収録された論文「ビッグデータの社会哲学的位相」で大黒 (2014) は、データ化を「主体」「対象」「目的」に注意して理解すべきであると論じている。大黒の観点では、誰が何の目的で何を対象としてデータ化したのかによって、ビッグデータは区別されるべきものである。ただし大黒は「ビッグデータ」の生成における「主体」と「対象」、「目的」が、結局は「データそのもの」に還元され、人間の判断がビッグデータの回帰運動となってしまうような近未来のSF的状况を想像する。この想像は、一つの極論として念頭に置く分には興味深い。商品やサービスをweb上で情報化して販売し、スマートフォンで情報を集めて自動的に消費行動を繰り返すという生活には、確かにそういった記述がある程度当てはまるのかもしれない。しかし本稿の関心からすれば、大黒の議論はむしろ、「主体」「対象」「目的」を考慮する必要性を示し、これによりむしろ「ビッグデータ」と一様に呼ばれているデータのより多様な分析の可能性を意識化してくれるものとして示唆的である。

大黒の指摘に近い形で社会学的観点から分析を試みているのが、同じ特集の和田 (2014) の論文「ビッグデータとビッグソサエティ」である。和田は、「ビッグデータ」を、「生産」、「市場」、「国家」、「社会」との関係性によって「多角的に」捉え、データをめぐる主体の権力関係、戦略などに注目した分析の可能性を示している。ただし、和田の議論は示唆的ではあるものの簡単なスケッチに止まっており、区別された各側面の相互作用や関係性の実効的な検討はこれからといったところ

である。柴田（2014）や檜村（2014）が同じ特集号に寄稿している論考は、それぞれ福祉情報論あるいは精神医学の関わる領域における、具体的な対象としての患者やサービス利用者、主体としての国家や学界とその戦略や目的に踏み込んだ実践的な分析であると言えるだろう。流行り言葉としての「ビッグデータ」が、現実社会に実装されるテクノロジーとして定着するのに遅れずに、こうしたビッグデータの具体的な多様性を踏まえた各分野での実効的な分析が急がれる。

「ビッグデータ」をもし技術的観点のみから見るなら、それは一つのまとまったテクノロジーであると見られうるかもしれない。しかし、もし社会的・人間的な観点から見るならば、上述のように詳細な分類があり、それに応じた議論によってしか、このテクノロジーに対する実効的な議論を提示することはできないだろう。一般に、テクノロジーを社会的観点からのみ見ることは、その技術がたまたまその社会において引き起こした結果をその技術の必然的な役割と見てしまい技術の可能性に盲目となることが考えられるが、技術的観点だけからテクノロジーを見ることはその技術の社会的評価の多様な可能性を認知できないままに技術の方向性が決定されてしまうことにつながる。ビッグデータの社会インフラとしての性格を踏まえるなら、今後は、技術的・産業的知見と人文社会学的な知の協同による、より具体的かつ実効的な議論が求められるフェーズを想定すべきであろう。

私としては、こうした実効的な議論のための道具立てを本稿で提供したい。すなわち、「ビッグデータ」を技術的観点から一様なものと見てしまわず、人間的・社会的関心に即してより詳細かつ実効的に捉える基本的な観点として、ビッグデータを「表象」として捉える視点を提示したい。

## 2 世界の表象としてのビッグデータ

### 2.1 「表象」を扱う技術

「ビッグデータ」は、世界についての表象の膨大な集積と見なすことができる。そして、そう見なすことで、ビッグデータを単に技術的観点からのみ分析していても見えないことがあり、表象の意味の解読を検討するという人文・社会的アプローチが必要であることが分かるだろう。

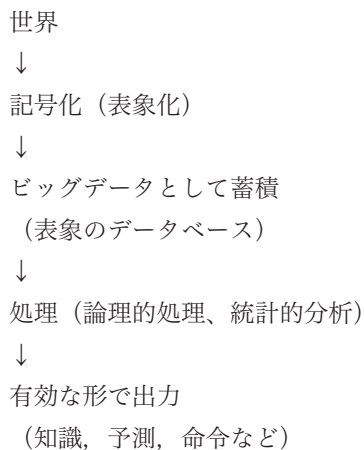
「表象」(representation)の本来の意味は、世界の再現前ということである。人間は、世界の一部をいわばコピーして再提示することができる。例えば、「富士山」という言語記号は、世界の中に存在する現実の富士山を記号によって再現前させていると見なすことができる。こうして、「富士山」という言語記号は、「表象」と呼ばれる。同じように、「富士山」と声に出してみたり、絵を描いて示してみることで、富士山を再現前させることができるので、これもまた表象である。こうした表象を用いて、世界について認識したり、これを用いて議論したり、意識的に新しい世界を構築するのは人間知性の本質的な能力であろう。

J.サールによる言語行為論の枠組み(Searle, 1969: pp.29-33)を用いて、表象と人間の社会生活の結びつきを一般的に捉えてみよう。世界を表現している表象は、言語においては、主張という志向性を伴って叙述文・事実命題として表現される。「Aは富士山に登る」という叙述文は、Aが富士山に登っているという表象に主張という志向性、すなわちその表象がどう使われるべきかという話者の意図を加えたものである。さらに過去や未来という時制を加えることで、「Aは富士山に登った」「Aは富士山に登るだろう」などの文を作成することができ、同じ表象に命令、疑問、意思などの異なる志向性を与えることで、「Aは富士山に登れ!」「Aは富士山に登るか?」「富士山に登ろう!」などの文が得られる。言語を通して計画され、実行され、評価される人間の営みや社



会的行為は、こうした意味で表象を媒介した多様な営みであるとみることができる。

広義にはじつは情報技術一般について言えることであるが、ビッグデータというテクノロジーは、こうしたいみでの表象を処理する技術である。すなわち、このテクノロジーは、まず世界のある状態をデータ化してシステムに大量に蓄積する。このデータは世界のある状態を記号的に表現するものであり、表象である。そして、このデータの統計的処理により、システムはある表象のある意図を伴って出力する。典型的には、例えば「この本を買ってはどうですか?」「このwebサイトがあなたに相応しいです」といった、記述・命令・推薦・疑問等々の志向性を伴った表象として出力される。時には、特定の出力に対して、温度調節を自動的に行うなど、機械的に命令を実行するタイプの出力も可能であろう。このプロセスを下記のように簡単にモデル化することができるだろう。



## 2.2 表象とAmazonとGoogleのビッグデータ

ビッグデータ・テクノロジーを用いた企業の代表格とされるAmazonとGoogleの想定されるサービスにおいて、世界を表現する表象が蓄積され利用される流れを簡単に確認しよう。例えば、Amazonのビッグデータには、膨大な本や商品についての情報が記号として集約されている。これ

らの記号は、システムにおいて本やユーザについての詳細な情報の表現であり、本と本をめぐる世界の表象であるということが出来る。この表象は必要に応じてAmazonのシステムによりダイナミックに処理されて、その一部がサービスに有効な形でブラウザに示される。例えば、あるユーザに対して、「あなたは、これらの本 {a, b, c, d, …} を購入してはいかがですか?」というようなリコメンドとして表示される (森, 2006 : pp.98-102)。

データ化する主体の責任者はもちろんAmazon社であるが、例えば書評を投稿したり、web上で閲覧・購買行動を残すユーザもまたデータ化に参加する主体であり、これらの主体間の権力関係やゲーム的構造についての社会的分析、作品の解釈や評価をめぐる人文学的分析は、このビッグデータの社会的影響を理解し評価する上で欠かせないだろう。また、Amazonでデータ化されている対象は、本だけではなく多くの家電や生活雑貨などにも及ぶが、これらを読み解くにもまた技術や経営だけの問題でなく、人文社会学的な相互作用やストーリー的な意味づけといった次元の分析が必要とされることが想定される。

Google検索の場合は、webにアップロードされた膨大な情報を中心に世界中のデータ化可能な情報を集積してデータベースを構築している (梅田, 2006 : p.50-52)。このデータベースには多様な情報が含まれているが、世界についての事実を伝えるニュース、科学的知識を伝える説明、社会現象・自然現象についての仮説や解釈、個人的な経験を記すブログ、政策や方針を提示する官公庁のサイト、タレントやスポーツチームなどに関する情報など、世界についてのさまざまな表象と見なすことができる記号群からなる。Google検索は、これをビッグデータとして保持し、ブラウザに入力されるキーワードに応じてそのデータベースから抽出したデータを整理された形での表象として表示するシステムである。Googleは

また、ユーザによってweb上に公開された情報だけでなく、Google自身がカメラを用いて撮影した街角情報、ユーザがGoogleの提供するサービスを用いることでGoogleに提供することになるメールやドキュメントの情報などについても蓄積し、これらをニーズに応える形で編集し、表象として表示するというサービスを提供している。

Google検索はweb上にデータ化されたものを集積してビッグデータとして蓄積し、それを編集するという意味で、いわば世界の表象化におけるメタ的な地位を占めているが、そうした情報編集主体としてのGoogle社の権力はしばしば「Google八分」などの言葉と共に指摘されてきた(吉本, 2006)。現代の情報インフラを代表する巨大企業に成長した同社の政治的な影響力から、Googleのサービスを拒否・制限する国家さえある。しかし、Googleのサービスによって、表象における意味的な変化が具体的にどのように生じているのかは、表象の変換サービスの提供者であるGoogle自身の秘密主義のため透明性を欠いたままサービスが流通しているのが現状であろう。

AmazonやGoogleの蓄積するデータだけでなく、例えば自然現象や工場・自動車などからセンサーで収集され蓄積されるセンサデータも代表的なビッグデータであるが、これらも世界の特定の状況を表現するものである。これについても、それぞれ、誰が何のために世界を表象化し、それを誰がどのように解釈しているのかといった、個別の分析が必要だろう。

こうした、表象を扱う情報テクノロジーに対する、人文・社会学における意味的アプローチの適用による分析や評価は、まさに社会情報学の課題である。一般的に言って、Google等による情報のメタ的な集積・編集・提供サービスが人間社会における世界表象の評価や意味づけに関わる度合いが高まれば高まるほど、私企業の立場からだけでなく、公的な立場からの人文・社会的な観点での分析や民主的マネジメントの必要性が高まるこ

とは間違いない。このテクノロジーの適切な推進ないし制限、あるいは社会的コントロールのために、社会情報学にはかなり大きな社会的責務がかかっていると言うべきだろう。

### 3 ビッグデータと局所表象／分散表象

#### 3.1 認知科学における局所表象／分散表象

前節では、ビッグデータを「表象」として見て人文・社会的にアプローチする必要性を訴えたが、本節では、「表象」といっても多様な区別があり、なかでもビッグデータの社会的評価やコントロールのためには「局所表象」と「分散表象」という表象の存在様態に関わる哲学的区別に注意することが重要であることを指摘したい。

局所表象(local representation)とは、ある意味のまとまりに応じて、ある程度まとまって存在している表象である。それに対して、分散表象(distributed representation)とは、表象が意味の単位にまとまっては存在していないケースである。この区別は、認知科学におけるコネクショニズムの立場を表現するために用いられた<sup>(2)</sup>。

認知科学における伝統的な立場であるファンクショナリズムは、人間の認知過程を、命令文逐次実行型の伝統的なコンピュータにおける情報処理過程になぞらえて、記号表象の逐次変換という局所表象的モデルで理解しようとした。この立場によれば、コンピュータにおいて情報が順次処理されていくように、まとまった意味を持つ文に相当する表象が人間の脳において順次処理されているのが人間の認知過程である。こうした過程においては、表象はその意味的まとまりを失わずに処理されるので、認知過程において局所性を失わないと考えられた。

それに対して、これに対抗する新しい認知モデルとしてコネクショニズムの提案した認知過程においては、入力された表象は特定の変換特性を持つネットワーク全体の中に一端分解され、ネット

ワーク全体において分散的に変換された後、ネットワーク全体から特定の出力が得られる（図-1）。このネットワークは人間の脳神経の構造を模したモデルとしてニューラル・ネットワークと呼ばれる。こうした有効な「判断」と解釈されうる形の出力を与えることのできるネットワークに対して、全体として認知過程をなしているとみなすのがコネクショニズムの立場である。その認知過程においては表象は局所的な存在様態ではなく、ネットワーク全体に分散した様態として存在していると見なされ、これが「分散表象」と呼ばれる。

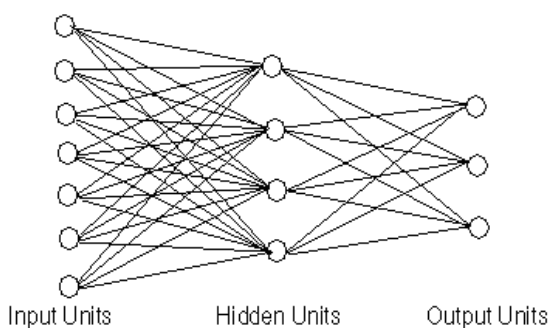


図-1 ニューラル・ネットワークのモデル  
"Connectionism", Stanford Encyclopedia of Philosophy, Feb19, 2015,  
<<http://plato.stanford.edu/entries/connectionism/>> Accessed 2015, March 30.

### 3.2 情報表現の存在様態としての局所性／分散性

表象が分散しているか局所的であるかは、認知科学における議論では脳やコンピュータにおける物理的な存在様態の区別であった。しかし原理的には、情報としての存在様態の区別と見なすことができるのではないか。言語を例にして、このアイデアを説明しよう。

言語の基本的な構成を、いくつかの語からなる文の集合と見なすことにしよう。すると例えば、「富士山は白い」という文は「富士山」「白い」という語の結合によるひとつの文表象であり、その意味するところは、富士山が白くなっているとい

う状態である。「富士山は白い」という言語記号は、こうした意味の単位に応じて局所性をもった表象として紙の上に物理的に存在しているように見える。絵画の場合も、紙の上において富士山の表現（形）と白さの表現（色）が近接していることで、白い富士山をという白さと富士山の意味的な結合が表現されているゆえ、物理的局所性が明らかであるように見えるかもしれない。

しかし、絵画の場合、遠近法を用いたり別の表現原理を用いたりする場合には、必ずしも紙の上の物理的距離が意味の近接性を表現するとは限らないことに注意したい。例えば、駿河湾の波の間から遠景として見える富士山を描いた絵画においては、近景となる波が紙上では富士山を越える大きさで描かれていても、多くの場合、波のサイズが富士山を越えていることを表現しているわけではないだろう。従って、絵画における表象の存在様態を局所表象とみなすなら、表象の物理的近接性ではなく、むしろその表現原理上の近接性が、その意味の近接性に対応している点が重要なのである。

言語記号の場合も、絵画の場合と同様である。つまり、紙の上で語同士が物理的に近いか遠いかが問題なのではなく、一定の表現規則の上で近接性が保持されているかどうかの問題なのである。例えば、「白く高くそびえる富士山と青い空」という表現は、紙の上の物理的距離では「青い」の方が「白く」よりも「富士山」に近接しているが、日本語の文法規則により表現的には「白く」こそ「富士山」に結合しているのである。

表象が局所性を持っているなら、その表象の意味をわれわれは直接に把握し、操作することが可能である。例えば、言語表象における局所性を前提にして、「富士山は白く高く、かつ、空は青かった」という表現から「富士山は白かった」を出力するという推論過程をわれわれは表象の逐次的な変換過程として把握・操作できる。逆にもし文という単位が局所性を持っていなければ、例えば「白



い」「青い」「富士山」「空」という局所性の認められる4語だけに限っても、これらの語の集合によってそもそも何を言いたいのか、いったい富士山が青いのか、空が青いのか、あるいはどちらかが白いのか、私たちにはさっぱり読み解くことが出来ないし、まして推論することなどできないのである。

こうして、表象が局所性を持つということ、ある意味の単位が物理的な単位にまとまっているかどうかではなく、むしろある意味の単位が表現原理上の単位として捉えられるかどうかという問題として捉え直したい。ここでもやはり表象が分散的である場合には、われわれがその表象を意味の単位で直接に把握・操作することができない。コネクショニズムのモデルにおいても、分散表象はニューロネットワークの中でいわばブラックボックスのように処理されるのであり、プログラマーはただネットワークの変換特性を調整することによって表象の処理過程に間接的にアクセスできるだけで、直接に個別の表象の処理を操作・説明することができないとされていた。今やそれは、分散表象の物理的な存在状態に起因するのではなく、分散表象がまとまった単位の情報表現として存在していないということに必然的に起因するものとして理解されるだろう。

### 3.3 分散表象のビッグデータ

私の提案は、この捉えなおされた局所／分散の区別を、ビッグデータの処理過程に対して導入することである。ビッグデータ自体を表象として理解するという観点は前節ですでに確認したが、この見方を推し進め、分散表象として処理される過程を含んでいるビッグデータを、局所表象としてのみ処理されるビッグデータから区別しよう。

AmazonやGoogleがじっさいにどのようにデータを処理しているのかは、各企業の秘密であり推測するほかないが、Amazonの提供するリコメンドやランキングの機能については、例えば次

のように想定することができる。特定の商品に対する評価の総数とレベルをカウントして、総合的に人気のありそうな本から順位付け、書評などをつけて表示する。購買履歴からユーザごとにジャンルの選好順位があり、順位付けにはこれも加点する。この場合、データは常に直接に意味が把握される状態で操作され、そのプロセスを経て最終的に編集されたものが出力される。こうしたシステムについては、ビッグデータは逐次意味を追いかけることのできる形で、局所表象としてのみ処理されていると見なすことができるだろう。

これに対して、Googleの提供する機械翻訳サービス（Google翻訳）については次のようなシステムと理解されている。クキエら（2013, pp.62-66）の解説によれば、従来の機械翻訳システムは二ヶ国語の文法と辞書を利用して、コンピュータによって逐語変換を行い、翻訳文を生成する仕組みだった。すなわち局所表象として情報を処理していた。しかし、Googleは、1980年代にIBMの研究チームが思いついたとされる「ある言語の単語やフレーズが別の言語のどの単語やフレーズに最も合致するのかを決定する際に、統計的確率をコンピュータで計算」という方式（前掲書, p.63）を、Googleの持つ大量の翻訳文データ（ビッグデータ）をコーパスにして実行するアプローチを採用している。ここでは、特定の文に対して特定の翻訳文が生成されるプロセスは、それらの表象の意味が考慮されずに計算・処理され、別の表象が生成されるというものである。もちろん、じっさいのGoogle翻訳のシステムにおいては、語の意味を考慮したり文法ルールを適用したりする局所表象のプロセスも併用されているだろうと推測される。しかし、Google翻訳システムの根幹が、表象の意味に関わらないこうした変換プロセスであるとしたら、そこでは表象は局所性を失い分散的に処理されていると言えるだろう。

自然界や人工物のセンサデータなどの情報もまた、データベース化され統計的に処理される過程

で、表象として特定の意味を持ちうる単位が分解され、人間には意識化不可能なレベルで処理されて一定の結果を生成されるということが考えられる(小林, 2015:p.138)。こうしたビッグデータの統計処理において使われるクラスタ分析においては、統計的な相関関係に基づいて新しい分類が作成される(水田, 2014)。その分類が、必ずしも人間にとって意味のある分類であるとは限らず、そのときには表象の局所性は失われているのである。

ビッグデータ・テクノロジーは、データを局所表象としてのみ扱うこともあるが、分散表象として扱うこともある。マイヤー＝ショーンベルガーとクキエ(2013)は、ビッグデータによる判断の変化の本質を、「因果関係から相関関係へ」として捉えている。これに対して私はこの変化を、因果関係を含む人間に意識化可能な結合・変換過程から、ある種の相関関係のように人間には意識化不可能な結合・変換過程を含む判断への拡張として理解すべきであろうと考える。すなわち、分散表象によるデータの処理過程によって、有意味な言語では理解不可能となる判断方式が導入されたことが、ポイントなのである。

## 4 ブラックボックスとしての分散表象

### 4.1 分析と言語化の困難

ビッグデータが分散表象として利用される場合、その処理過程はビッグデータの主体にとってさえ、ブラックボックスとなる。ここで、テクノロジーの社会的コントロールという観点からその問題性を指摘して、ビッグデータの分散性への注意を喚起したい。

ビッグデータ・テクノロジーにおける局所表象と分散表象の区別は、テクノロジーの社会的コントロールという観点からみると、人間には意識化可能な変換過程の社会的マネジメントと人間には意識化不可能な変換過程の社会的マネジメントの

区別に対応する。表象の変換過程を意識できるならば、そのシステムに問題があった場合には、変換過程をたどって問題の発生箇所を突き止め、それを修正することが想定可能である。もちろん、システムが問題なく有効に稼働している場合には特にこうした変換過程が問題視されることはないだろうが、システムのトラブルを修正する場合、現行のシステムにおける特定の問題点を意識化してそれを改良する新しいシステムを構築しようとする場合には、こうしたプロセスの意識化は必須である。

ところが、ブラックボックス化したシステムにおいては、こうした形でのシステム構築、メンテナンスやバージョンアップができない。もちろん、コネクショニズムの提唱する分散表象の認知モデルにおいても、システム全体を学習させることでシステムの精度を上げていくことができる。ただ、変換過程が意味的に意識化されない以上、表象の変換過程でトラブルの生じた箇所を突き止めてそこを修正することは難しい。例えば、Google翻訳のようなシステムにおいて、どうしても不適切な翻訳文が生じてしまう場合が考えられるが、出力時にフィルタリングをかけるなどの局所表象的アプローチを併用すれば応急的な対処は可能かもしれないが、それが生じないように特定の箇所を修正することは分散表象のシステムに関しては困難であろう<sup>(3)</sup>。

こうした分散表象のシステムにおいては、トラブルを言語化して社会的に有効に対応することもまた難しい。まず、プロセスが意識化されず、出力が言語化されないままに世界に反映される場合、その出力についての予めの批判的吟味は難しいだろう。また仮に、現象としてのトラブルが言語化されて社会的に認知されたとしても、プロセスがブラックボックスであるなら原因の同一性が確認できず、いわば「兆候の集合」として意識化されうるにすぎない。これでは、「兆候の集合」としてしか把握されていない新種の病気と同じ

く、有効で根本的な対処は困難である。

#### 4.2 分散表象と反知性主義・権力

ビッグデータ・テクノロジーが分散表象を利用する場合、第一節で言及した西垣とチェンによる「反知性主義」や「虚無的な状況」への懸念は、局所表象を利用したこのテクノロジーの場合に比べて、より深刻なものとなるだろう。

西垣(2014:p.42)は「反知性主義」を「コンピュータにデータを丸投げして、人間は考えなくてもよいのだという風潮」と表現しているが、分散表象のシステムにおいては、その処理過程への意識的な反省ということがそもそも不可能である。そこで、そもそも意識化できないのだから仕方ないという、より根本的な反知性主義的な開き直りが懸念される。同じ風潮の下では、チェンのいう「虚無的な状況」も容易に想定される。チェンは、ビッグデータを利用したわかりきったルーティンワーク的な処理によって、サービスが自動的に管理されるような状況を想定していた(チェン, 2014:p.45)。しかし、分散表象システムの場合は、サービス受容者だけでなくサービス提供者でさえ理解しようとしても原理的に理解できないままにシステムの出力に従って社会が制御されているという、より一層深い「虚無」が想定されるのである。

分散表象に対する知性的な反省やそれを踏まえた創造の困難は、大黒ら『現代思想』の論者たちが指摘していたビッグデータにおけるデータ化主体と客体の権力関係、データ化される対象の主体性などについても、より深刻な状況を突きつけるように思われる。

特に、データを選択したり処理したりする権力の格差だけでなく、データの意味を決定する解釈権力が、データ化・利用の主体とデータ客体との間に、決定的な断絶をもたらすと思われる。データ利用者は上記の虚無的な状況につけ込み、出力データの解釈を都合よく行うが、データ対象は

これに対して批判ができないという状況が容易に想定されるのである。Google検索は、ランク付けのプロセスがブラックボックス的に秘匿されていることによって、神がかったものとなっている。特に、この過程が分散表象によるビッグデータの処理の場合、Googleに限らずとも、データの独占的利用者は、データ化対象やサービス提供者に対して、ビッグデータのご託宣を解釈して取り次ぐ神官のような反知性主義的かつ絶対的な権威を持ってしまうことになるだろう。

原則として、ビッグデータが社会的に重要な役割を演じるようになればなるほど、データとデータを利用するテクノロジーの社会的ガバナンスが問題視されるべきだろう。そこでは、収集されているデータの範囲、処理過程、そして解釈における透明性や収集・利用主体による説明責任が求められることになるだろう。その上でデータの適正な利用、正統な管理について、関係者を広く含んで合意が形成されて、運用されるべきである。

しかし、この説明と合意のプロセスにおいて、分散表象ビッグデータの場合には、先の「反知性主義」や「虚無的な状況」に乗じて、ガバナンス構築の前提となるデータの解釈と説明が、データの収集・利用主体の手に批判不可能な形でもつぱら委ねられることになることが想定される。この場合、ビッグデータを社会的に論じるための言語化の次元にまで主体と客体の権力関係が浸透することになる。そして、データ化される対象である市民によるガバナンスへの主体的な参加がより困難になり、参加は形式化し、支配権力による誘導的なものになってしまうだろう。

## 5 ビッグデータのガバナンスに向けて

最後に、こうした分散表象によるビッグデータ・テクノロジーの社会的なコントロールについて、いくつかの実践的論点を指摘して稿を閉じることにはしたい。ビッグデータ・テクノロジーが社会的

に大きな影響力を持つ以上、それは適切にコントロールされなければならない。情報技術の未熟な時代においては、まず技術競争ありきで、そのコントロールはしばしば後回しにされがちだった。しかし、情報技術が基本的社会的インフラとしての地位を占めた近年は、情報テクノロジーに関してもこういった楽観論や無責任論はすでに成り立たなくなっていると思われる。

では、その処理プロセスがいわば無意識化・ブラックボックス化される分散的ビッグデータを、社会的にどう適切に管理していくことができるだろうか。マイヤー＝ショーンベルガー&クキエもまた、このテクノロジーの社会的コントロールの必要性を意識して、ビッグデータの専門職としての「アルゴミスト」の確立を提唱している（クキエら、2013：p.267）。

専門職とは、その分野の高度な専門的知識と技術に加えて、その技術の管理に関わる社会的知識や倫理的判断を備え、高度な社会的責任を担える社会的に認定された専門家のことである（黒田ら、2004：pp.72-77）。例えば、現代社会は一般市民にはどうも理解しきれない高度化した会計システムや法システムに依存している。こうしたシステムの正確で公正な運用を実現するスペシャリストとして「会計士」や「弁護士」などの専門職が制度化され、その社会的信頼を保証している。ビッグデータ・システムについても、「アルゴミスト」達が「データの監査人」の専門職として企業や政府の内外からこのテクノロジーの信頼性を保証する仕組みを作れば、「データ資本家の暴走」（クキエら、p.270）を牽制し、ビッグデータを人間の手で公正に、安全に、利用していくことができるというのがクキエらの提案である。

しかし、分散表象を扱い、その処理過程が専門家にも意識化され得ないことがあるというこのテクノロジーの特性を考えると、ビッグデータについては会計士や弁護士、あるいは技術士といった従来の専門職モデルは、そのままでは成立しない

のではないと思われる。専門職の具体的役割は、その専門領域における社会的・物理的トラブルの予期、トラブルの説明、トラブルの未然の防止などにより、専門領域の社会的透明性、説明責任、信頼性、公正性を保証することであるが、ビッグデータが分散表象を扱う場合には、すでに論じてきたように処理過程が専門家にとってさえブラックボックス化されるため、これは困難である。

また、クキエらの「アルゴミスト」の提案は、データの利用について「データの対象による個別の同意」から「データ利用者責任制」へのシフトを含んでいる。すなわち、データ利用者が政府や専門職と協力して良識を持ってサービスを提供すれば、素人であるデータ化対象たちのコントロールから解放されて、よりイノベティブなサービスを安全かつ比較的自由に試みられるというアイデアである。専門的なサービス提供者への責任を求めるこうした方式は、例えばPL法（製造物責任法）や自動車のリコール制度などに見られるいわば消費者安全の方式であり、必ずしも市民の権利を取り上げて市民を抑圧するというわけではない。ビッグデータの利用も、大規模化、複雑化するにつれて、市民が自分に関わるデータの利用を完全に理解しコントロールすることはより困難になるだろう。その状況化で、データ社会における市民の安全を保証するには、データ利用者責任制は、データ・サイエンティストの専門職化と合わせて、ビッグデータ・ガバナンスのひとつの選択肢となるだろう。

ただし、データを分散表象として扱う場合、データの解釈の問題を忘れてはならない。データの解釈においては、まずデータ化された対象の意思が尊重される必要があるだろう。問題はこの意思は極めて多様でありうるということである。自動車の安全や家電製品の不具合で損害を被ったという類の誰にとってもある程度共通の物理的・経済的なリスクに比べて、国家や専門職が、多様なデー



タ化対象・市民たちの意思や気持ちについて、一般的に推定することは極めて難しいだろう。

GoogleがStreet Viewというサービスで採用しているしくみは興味深い。Street Viewは、Googleが撮影した街角の映像をGoogleの提供する地図の上に表示するサービスだが、そこで撮影された街角データにはさまざまな画像が含まれる。Googleは今のところそれをいちいち解析せずに、その場所で撮影されたひとまとまりの画像データとして示すだけだが、その写真の中に自分が写っていた場合には、そのデータの対象となった個人はGoogleに対して自分の画像の修正・問題箇所の削除などを要求できる<sup>(4)</sup>。つまり、このシステムでは、データの解釈権はデータ対象の市民側にあり、Googleはそれを本人の申し出によりオプト・アウト方式で管理しているわけである。

もちろん、Googleのこの方式にも問題はあつた。本人が気付かず請求しないうちは勝手に自分の画像が利用されているという点は、本人からすれば不本意でありうる。もし本人に画像の権利があつたとするなら、本来Googleは勝手にそのデータを利用できなかったはずである。本人によるデータの解釈の優位性を認め、本人の参加を保証しようとするGoogleの発想と工夫は確かに興味深いだが、現状では分散的ビッグデータ・ガバナンスとして十分に正当性を確保できているとは未だ言えないのではないか。

ともあれ、データ解釈を含むビッグデータの社会的コントロールにおいては、多様な意思や気持ちを持つデータ化対象の参加をベースとしたガバナンスが不可欠であるということは指摘できるだろう。データの意味解釈に関しては、当事者を含む関係行為者全体の意思や気持ち、判断の総和となるべきであり、専門的な研鑽によって身につけられる一般的な専門知識のみで決定すべきものではないからである。従って、ビッグデータの専門職を想定するとしても、高度化された専門知識を活用して企業や国家をサポートする「上からの」

専門家に止まっているは足りない。むしろ、データ化対象の当事者を含む社会の多様な意思や気持ちを考慮して、データ利用の方法・技術について説明し、社会的合意形成をサポートする、あるいは当事者の権利の行使を確保するような、柔軟さやオープンさを備えた、より市民的な専門家でなければならないだろう。

## おわりに

分散的ビッグデータのガバナンスは、データ対象とされた人々を含む市民の、データ解釈における主体性と解釈内容の個性性にどう対応するのが問題である。このテクノロジーの管理は、当事者や多様なアクターの参加を前提としたガバナンスにならざるを得ないと思われる<sup>(5)</sup>。従って、仮に専門職がこうしたテクノロジーのガバナンスにおいて重要な役割を果たすとしても、それはパートナーリスティックなものではなく、多分に市民協同的なスタイルとなるべきだというのが私の考えである。

分散表象処理に伴う反知性主義や虚無主義に抗して、このテクノロジーが表象の領域でどのように機能しているのかを説明し、どのように機能すべきなのかを取りまとめるのは、他のテクノロジーの場合にも増して困難な課題である。ともあれ、必ずしも専門職という制度に依らねばならないわけではないが、データ化対象の主体的な参加を十分に保証することは、今後ビッグデータ・テクノロジーを人間的に利用していくべき情報社会における必須の課題であろう。

## 注

- (1) 田島逸郎「「情報通信白書」最新版に見るICTの現在(中編)——到来するデータ活用社会」, ZDNet Japan, 「情報通信白書」最新版を読む, 2014年10月15日, <<http://japan.zdnet.com/article/35053198/>>



Accessed 2015, March 30.

- (2) "Connectionism" Routledge Encyclopedia of Philosophy Version 1.0, Routledge, 1998.
  - (3) 機械翻訳をニューラルネットワークで分散的に処理する試みについては、人工知能の研究者やMicrosoft, Googleの技術者たちによってすでに始まっているが、この技術の意識化や制御の困難が同時に懸念されている(小林, 2015, pp.32-43)。
  - (4) グーグルストリートビュー「プライバシーとセキュリティ, 問題の報告」<<http://www.google.co.jp/intl/ja/maps/about/behind-the-scenes/streetview/privacy/#streetview>> Accessed 2015, March 30
  - (5) 以上のビッグデータ・ガバナンスにおけるデータ・サイエンティストの役割についての見解は、下記、2014年11月の第5回横幹連合総合シンポジウムにおける報告を反映したものである。吉田寛「ビッグデータ・ガバナンス」(2014年11月29日, 東京大学, 『第5回横幹連合総合シンポジウム予稿集』 pp.92-95)
- 『現代思想 (vol 42-9)』(『【特集】ポスト・ビッグデータと統計学の時代』収録) 青土社, 2014.
- 西垣通・ドミニク・チェン (2014) 「【対談】情報(データ)は人を自由にするか」, 『現代思想 (vol 42-9)』 pp.38-58.
- 水田正弘 (2014) 「ビッグデータブームを考える」(『現代思想 (vol 42-9)』, pp.69-79)
- 大黒岳彦 (2014), 「ビッグデータの社会哲学的位相」(『現代思想 (vol 42-9)』 pp.113-147)
- 柴田邦臣 (2014), 「生かさなく生-政治>の誕生 ビッグデータと「生存資源」の分配問題」(『現代思想 (vol 42-9)』, pp.164-189)
- 檜村愛子 (2014), 「「ネオ精神医学」を生み出した「トロイの木馬」: DSM アメリカにおける父殺しと科学への倒錯」(『現代思想 (vol 42-9)』, pp.190-201.)
- 和田伸一郎 (2014), 「ビッグデータとビッグソサエティ」(『現代思想 (vol 42-9)』 pp.219-229)
- 小林雅一 (2015), 『AIの衝撃 人工知能は人類の敵か』講談社
- 梅田望夫 (2006) 『ウェブ進化論』筑摩書房
- 森健 (2006) 『グーグル・アマゾン化する社会』光文社
- 吉本敏洋 (2006) 『グーグル八分とは何か』九天社
- John R. Searle (1969) , *Speech Acts, Cambridge UP.*
- 黒田光太郎・戸田山和久・伊勢田哲治 (2004) 『誇り高い技術者になろう』名古屋大学出版会
- 城山英明 (2007) 「科学技術ガバナンスの機能と組織」(城山英明編 (2007) 『科学技術ガバナンス』 東信堂, pp.40-72)

#### 参考文献

- ビクター・マイヤー＝ショーンベルガー&ケネス・クキエ (2013=2013) 『ビッグデータの正体』講談社
- 海部美知 (2013) 『ビッグデータの覇者たち』講談社
- 城田真琴 (2012) 『ビッグデータの衝撃』東洋経済新報社
- 森健 (2012) 『ビッグデータ社会の希望と憂鬱』河出書房