

ビッグデータはブラックホール!?

東京経済大学 コミュニケーション学部教授・東京大学名誉教授

にしがき とおる
西垣 通

人間の生活を楽しく豊かにしてくれるのが、そもそもIT（情報技術）の使命のはず。むしろ便利な面もたくさんありますが、よく考えると私たちの生活は、逆にITの機械処理によって振りまわされてはいないでしょうか。ビッグデータ時代が到来しつつありますが、データの大海の中でエネルギーを吸いとられ、溺死しないようにしたいものです。

● ITとは面倒なもの

先日、マイクロソフト社のパソコン用OSであるウィンドウズXPのサポートが、ついに中止されてしまいました。2001年に発売され、それから約14年も続いた長寿OSだそうです。継続して使うことは不可能ではないと言っても、これからはどんな危険なウイルスが忍び込んできても、もはやマイクロソフト社は対策をとってくれません。

幸い私は昨年、定年で職場を変ったこともあり、今はウィンドウズ8のユーザなので、それほど痛痒を感じないのですが、ずっとXPを継続使用してきた人たちからは大きな不満の声が聞こえてきます。これは単に、新しいOSやそれを動かすパソコンを購入するコストだけの問題ではありません。入れ替えのためにかなり大きなエネルギーが必要だからです。

とりわけ、XP上で動作する特殊なアプリケーション・プログラムを開発し、それに慣れ親しんでいたユーザは、いま悲惨な状況に追い込まれています。

ウィンドウズ8用のアプリケーション・プログラムを、新たに一から開発しないと仕事になりません。悲鳴が上がりますね。

本当は、OSを新バージョンに変えるときは、旧バージョンの機能をすべて包含することが望ましいし、厳密にこの原則が守られていれば、古いアプリケーション・プログラムは、原理上、新バージョンのOS上でも動作するはずなのです。しかし現実には、細かいインターフェイスがいろいろ変わったりで、うまく動作しないことのほうがずっと多いでしょう。

● コンピュータは石頭

まことにコンピュータとは、融通のきかない石頭に他なりません（図1）。それで結局、OSが変わるごとに膨大な作業が発生するという次第です。

だいたい、パソコンだけでなく、ケータイをはじめIT（情報技術）機器類の入れ替え作業はけっこう面倒なものです。私はしばらく前、ウィンドウズ8.0



イラスト：横山宇加
©uka yokoyama

図1 コンピュータは石頭

から8.1へバージョン・アップしたのですが、ダウンロードを開始してから、いろいろなセットアップやチューニングを行い、ひとまず完了するまでに半日かかりました。

自分自身は古いアプリケーション・プログラムの機能で十分満足していて、入れ替えの面倒を避けたいと願っていても、そうはいきません。仕事仲間や友人がネット経由でファイルを送ってきたとき、それが新バージョンのアプリケーション・プログラムで作ったものであれば、こちらの古いIT機器上ではうまく動作しないこともあるからです。仕方なく、新しいアプリケーション・プログラムを購入せざるをえなくなるのですが、それはたぶん新しいOS、新しいIT機器上でないと動作しないでしょう。

というわけで、メーカーの営業戦略にそって、不満を言いながらも新たなハードやソフトをせっせと購入し、入れ替え作業にいそしんでいる、というのが多くの人びとの現状ではないでしょうか。そう考えると、便利なIT

文明というのは、生活効率を大きく改善したようでいて、実は人間のエネルギーを吸い取っていくブラックホールのような気もしないではありません……。

いやいや、これはあまりに悲観的な見方だと叱られそうです。ハードもソフトも急速に進歩していて、だからこそOSのバージョン・アップが必要なのだと言われればその通りです。単に複雑で高い機能が増えたということだけではありません。プロセッサの処理速度が大きく向上し、メモリ価格が下落して大容量メモリが普及したことも一因で、データの量そのものがいちじるしく増加してしまったのです。

● ビッグデータとは

いま「ビッグデータ」という概念がIT業界を席卷しています(図2)。20世紀はスモールデータの時代だったが、21世紀はビッグデータの時代だ、というのがどうやら共通の了解事項になりつつあるようなのです。

確かにネット内のデータだけを眺め



図2 ビッグデータを構成する各種データ(例) 出典：総務省¹⁾

でも、データ量の伸びはすさまじいと言うほかはありません。昔は官庁や企業などの専門家が入力した文字や数表データが主でしたが、それに動画や音声といったマルチメディア・データが参入し、さらにウェブ2.0とともに一般ユーザが書き込むブログやツイッターといったソーシャル・データが追加され、そして今や、各種機器に備え付けられたセンサーで自動的に採集される膨大なデータが洪水のようにネットのなかに流れ込み始めています。

これは気象庁や地震観測所など専門的機関からのデータだけではありません。たとえば、一般ユーザが使っているスマートフォンの中には、GPS（全地球測位システム）による位置情報だけでなく、気圧や温度、方角、加速度

などの諸物理情報をセンサーで取得し、必要におうじて記録したり送信したりできる機種も出現しつつあるようです。

ある調査によれば、2011年における世界中のデジタル・データの総量は1.8ゼタバイト、つまりDVD映画（2ギガバイト）の約9,000億本に対応し、さらに2020年にはその20倍以上の40ゼタバイトにも達すると予想されています。まさに気の遠くなるほどのデータ量と言わねばなりません。

● 災害と事故の防止

それにしても、こういった大量のデータを使って、いったい何をしようというのでしょうか。

一つは物理的なデータを細かく採集す

ることで、災害や事故を防ごうというわけです。たとえば、ダムがある山岳地方の周辺一帯の降雨データをリアルタイムで測定し、コンピュータで適切に処理すれば、台風のときなど、ダムからどれくらいの水量を放流すべきか、最適な調整を行うことができるでしょう。同じようにデータ取得と分析の精度をあげることで、干ばつや山火事などにも対処することができます。

また、自然物だけでなく、人工物の保守にも有効でしょう。古くなった橋梁や高速道路、トンネルなどのあちこちにセンサーを付けておき、その状態を常時監視していれば、必要におうじて修理や改築をおこない、恐ろしい事故を未然に防ぐこともできるはずです。

東日本大震災における福島原発事故以来、この国の安全管理にたいして、一般の人びとのあいだには不信の念が広がってしまいました。そういう懸念を払拭するためにも、ビッグデータの活用は非常に期待されているのです。

とはいえ、これらはビッグデータ活用のほんの一部にすぎません。というより、この種の工学的応用は、規模は小さくとも以前から行われてきたとも言えるのです。

● マーケティングへの活用

IT業界がこれほど湧いている理由はむしろ、ビッグデータを商品マーケティングに活用したいというビジネス界の熱い要望に応えるためでしょう。一言でいうと、ビッグデータ処理によって、個別消費に着目し、ターゲッ

トをしぼった販売戦略をたてられるということです。

20世紀は大量生産大量消費の時代でした。テレビのコマーシャルは、不特定多数の人びとに共通の商品イメージを浸透させましたが、21世紀の宣伝活動は、そのレベルにとどまりません。新たなマーケティングの主流は、個々の消費者がもとめる商品ニーズを的確に把握し、予測し、ニーズに合った関連データを先回りして示すことで、効率よく消費者の購買行動を導くというものになっていくと考えられます。

そのためには、個人が日頃、どのような商品を購入し、何に興味をもっているか、またどのような仲間と付き合い合っているかを把握しなくてはなりません。人間は、とかく習慣的な購買行動をくりかえしがちです。さらに、親しい仲間が持っている商品と同じようなものを買いたいという傾向もあるからです。

こういった理由から、個人ごとに消費行動履歴を詳しく追跡して分析し、個別に宣伝広告データを提供するマーケティングが脚光をあびることになります。大都市には何百万人もの消費者がいますから、そのデータ量は大変なものになります。こうして、いわゆるビッグデータ処理技術が渴望されるわけです。

● ビッグデータに対する不安

以上のような販売戦略は、消費者にとって便利なこともあるでしょう。でも便利な反面、ちょっと不安を感じる

のではないのでしょうか。

たとえば何気なく街をあるいていて、ふとイタリア料理を食べたくなったとたん、ケータイの画面に近くのイタリアン・レストランの案内が出現したらギョッとするはずです。

誰にとっても、やはりプライバシーは大切なものです。個別マーケティングと個人情報保護とはなかなかうまく両立できません。とはいえ、ビッグデータにまつわるプライバシーの問題はすでに種々指摘されているので省略し、本稿では、ビッグデータをめぐる別の疑問点についてふれてみたいと思います。

ともかくやたらにデータ量が多いのがビッグデータなのですが、その分析処理は基本的にコンピュータがやってくれることになっています。時々刻々、流入してくる大量のデータを、高速コンピュータが自動的に統計処理、分析し、何をすべきか人間に指示してくれる仕掛けが現れるのです。平たく言えば、人間の代わりに機械が考えてくれるという次第です。

とすれば、どれほどデータ量が増えても対処できるわけで、人間にとっては手間が省け、ラクな時代になったという気がしてきます。事実、ビッグデータ信奉者の中にはそう明言する人も少なくありません。

彼らに言わせると、近未来には、人間がいちいち因果関係を探査する必要はなくなります。膨大なデータをコンピュータが勝手に分析し、自動的に相関関係を見出してくれるのがビッグデータ時代の特徴だというわけです。

しかし、本当にそうなのでしょうか……。

● 因果関係より相関分析？

情報工学者のビクター・マイヤー＝ショーンベルガーと技術ジャーナリストのケネス・クキエが二人で書いた『ビッグデータの正体』²⁾ という有名な本があります。これは、ビッグデータ関係者のあいだではまるでバイブルのように読まれていて、内容もそれなりに興味深いものです。とはいえ、ところどころに奇妙な記述が混じっているため、ビックリ仰天せざるをえません。

たとえば、「因果から相関の世界へ」というタイトルの第4章を眺めると、中古車の品質についての記述があります。データを分析して相関関係を調べた結果、オレンジ色のクルマは、他の色のクルマと比べて欠陥が明らかに少ないというのです。でも、いったいなぜでしょうか。

オレンジは特注色なので持ち主はカーマニアが多く、クルマを大切にしているとか、オレンジ色は路上で目立つので事故に遭いにくいとか、理由はいろいろ考えられます。しかし著者は、理由など探究する必要はないと断言します。相関関係は明確だが因果関係は複雑でわからない、理由など説明しようとしないうほうが賢明だというのです。「因果関係にこだわらない相関分析なら時間もかからず安上がりだ。因果関係と違って、数学や統計の手法で相関を調べ、デジタルツールで関係の深さだけを示せばいい」³⁾ というわけ

です。そして冗談めかしながら、ポンコツ車でもオレンジ色に塗ってはどうかと勧めています。

● 人間が考えるとは何か

オレンジ色の中古車の品質がよいとは、なかなか面白い話です。でも私は「コンピュータが答えを教えてくれたのだから、理由は分からなくていい」という結論にはどうしても同調できません。オレンジ色のクルマは、白や黒、灰色などの平凡なクルマより絶対数がずっと少ないので、統計処理に問題があったという可能性も否定できないでしょう。

「風が吹けば桶屋が儲かる」ということわざが示すように、二つの現象のあいだにたまたま相関があるという場合は決して少なくありません。強風が続いたからといって桶製作会社の株を買う人がいないのは、相関関係が成り立つ条件について考えるからです。

「人間は死ぬ」と「ソクラテスは人間だ」から、「ソクラテスは死ぬ」という命題を導くのは正当です。けれども「人間は死ぬ」と「ソクラテスは死ぬ」から「ソクラテスは人間だ」を導くことはできません。「ソクラテスという名の犬が死ぬ」のかもしれないからで

す。たとえば「事故の犠牲者はプラトン」という名の犬とその飼い主である」といった条件があるとき、はじめて「ソクラテスは人間だ」が導けるのです。

実を言うと、世の中で厳密な因果関係にもとづく分析など、ほとんど実行されてはいません。医師は、患者の体内で起きている生体反応を全て知り尽くしているのではなく「この薬を与えれば、患者の容態はたぶん好転する」と対症療法を行っているにすぎないのです。にもかかわらず、そういう相関分析において、なるべく因果的な関係が成り立つ条件を考え抜いていく努力が、恐ろしい誤りの防止につながるのです。それこそ「人間が考える」ということの本質ではないでしょうか。

いくらコンピュータが高機能になっても、全面的に相関分析の判断を任せるとはできません。データの活用は確かに大切です。しかし膨大なデータを正しく処理できるのは、究極的には、やはり人間なのです。

とすれば、ビッグデータがまるで人間のエネルギーを吸い取るブラックホールのように見えてくるのは、はたして私だけでしょうか。われわれはIT活用について、より深く考えるべきではないのでしょうか。

参考文献

- 1) 総務省：ビッグデータの活用の在り方について、ビッグデータを構成する各種データ（例）、情報通信審議会 ICT 基本戦略ボード ビッグデータの活用に関するアドホックグループ 取りまとめ、p.12 (2012) http://www.soumu.go.jp/main_content_1000160628.pdf
- 2) ビクター・マイヤー＝ショーンベルガー、ケネス・クキエ：ビッグデータの正体、(訳) 斎藤栄一郎、講談社 (2013)
- 3) ビクター・マイヤー＝ショーンベルガー、ケネス・クキエ：ビッグデータの正体、(訳) 斎藤栄一郎、講談社 p.106 (2013)