

ビックデータは新たなビジネスになり得るか

3738 横田光悦

1. 研究の動機（背景）・目的

今日の社会は、情報社会といっても過言ではない。その中にいる我々にとって、その横溢する情報をどうにか、付加価値のあるものにしたいと思ったから。

また、煎じ詰めることで、情報化社会に伴って起こる情報格差が広がり、情報に疎いものすなわち情報弱者と呼ばれる人たちの擁護に励みたいと思ったから。

そして、ビックデータのマイナス面を解消することで新たなビジネスを切り開きたいと思ったから。

2. 研究の方法・過程

始めは、ビックデータが誕生した背景を調べ、それによりもたらされた恩恵の例を挙げる。良い面が存在する一方付随するマイナス面をあげ、ビックデータが完全ではないであることを示し、問題解決に努める。

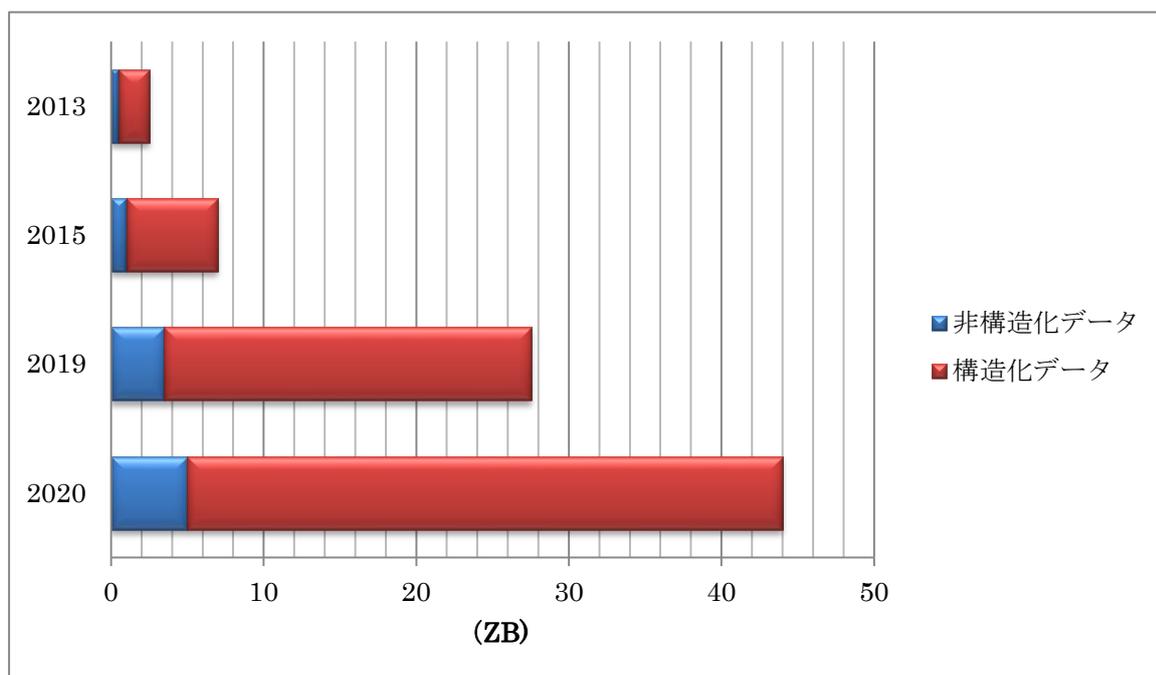
主題に基づき、実際にビックデータをどのような方法で運用するのか、具体的に述べる。

3. 研究の成果・考察

ビックデータとは、そのままの意味で解釈すれば巨大なデータとなる。しかしながらその意味だけではもの足りないので

- 1, 途方もない量であること
- 2, 非構造化データを含むこと
- 3, リアルタイム処理をすることができること

この3点を含むことが大事である。ここでますますデータが増えるにつれ、ビックデータの価値が高まる。2013年に4.4 ZB（ゼタバイト：10の21乗）だった年間生成データ量は、2020年には44 ZBに達すると予測される。



※出典：EMC、IDC 調査報告「デジタルユニバース」より

ここで構造化データというのは、顧客情報などデータベースに簡単に格納されるデータの様式のこと、非構造化データというのは、電子メールやテキストファイルなどの文書や、画像、動画といった複雑でデータベースに収めることができないデータのことである。しかしながら従来のデータ管理ソフトウェアと別のソフトウェアを併用することにより、「非構造化データ」も取り扱えるようになった。また高速なデータ処理と膨大なストレージの確立も相まって、これまでの物事の仕組みについて仮説をたてるといった所謂仮説主導型から、仮説ではなく豊富なデータを使う、データ主導権型へ、重心が移った。またビックデータの随伴現象として、従来のやり方では、発見しうることのできない、理由を伴わない答え、つまり因果から相関の世界への転換がされていること…かみ砕いて説明すると、過去には見えなかった関連が見え始めどれほど努力しても把握しきれなかった技術や社会の複雑な力学が把握できるようになったのである。そこで「データフィクション」（すべてのものがデータ化され、ビジネスになる時代）が新たに誕生。あらゆる分野でビックデータが使われている。例えば…



など多岐にわたる。

身近な例としては、インターネットの検索内容から、各検索語の使用頻度と、インフルエンザ感染の時間的・空間的な広がりとの間の相関関数を見て、インフルエンザ・ウイルスの感染状況をリアルタイムに特定し2009年のH1N1ウイルスによる新型インフルエンザ危機の際に役立った「インフルトレンド」また女性客の購入履歴から、女性の妊娠を特定できたりもした。

医学面に関していえば、病院で患者にさまざまな管や電極コード、装置を取り付けると、膨大なデータが発生する。心電図だけでも測定回数は1000回にも及ぶ。例えば、早産児の診断を的確に判断するのを目的に心拍、呼吸、体温、血圧、血中酸素濃度ほか16種類のデータを収集する。収集したデータをもとに早産児のわずかな変化を検知し、感染症の明らかな兆候が現れる24時間前に発症を警報で知らすことができる。

ビックデータの収集方法に至っては、先ほどの患者に装置を取り付けバイタルサインを収集したり、車などにセンサーを取り付けたり、TwitterやFacebookから情報を集めたりと(個人情報は伏せられているらしいが…)多岐にわたる。

個人情報に関するものを実際にビジネスに生かしている企業の権威であるアマゾン、グー

グルなどの例をとると、アマゾンでは商品購入や商品閲覧、レビュー投稿といったデータをフル活用し、グーグルは検索と無料アプリケーションによって蓄積した膨大なデータをもとに広告ビジネスを行っている。一見個人情報の取り扱いがうまく守られている気がするが、往々にして悪用もしくは流出といった事例が発見されている。例えば、最近の例として、ベネッセの個人情報流出や、日本年金機構の顧客情報の流出などがある。また、匿名化されている情報でさえも、特定されるといった悪質なケースが見られる。よってビックデータに順応した新たな法の改正が必要である。現在データ収集の際には「どういう目的でどの情報を集めるのか」を本人に説明することになっている。つまり本人が同意しなければ収集が始まらない。合法的に個人情報を収集・処理する手続き方法は「告知に基づく同意（告知と同意）」方式だけではないが、やっかいなことに今やこの「告知と同意」方式が世界中でプライバシー保護の基本になっている。なぜやっかいかというと、ビックデータは、二次利用にデータの価値の大部分がある。したがって、データ収集時には想定されていないケースが多い「告知に基づく同意」ではまともにプライバシーを守ることはできない。なぜなら一度同意をすれば、個人情報を守る枠組みがなくなるからだ。また企業にとっても二次利用ができない。そこで新たな法の「データ利用責任制」がある。そのような仕組みになれば、企業は、個人情報が処理される際、個人にどのような影響が及ぶのか慎重に検討したうえで、データ再利用を正式に評価することになる。ビックデータの用途を適切に評価し、分析結果を正しく利用すれば、データ利用者にはメリットがある。利用のたびに本人から明示的な同意を毎回もらわなくても、個人利用の二次利用を自由に推進できる点である。

一方、用途の評価や保護対策が不十分だと、データ利用者は法的責任を問われ、規制当局からの命令、罰金、ときには刑事訴追の対象になる。ただし、このデータ利用責任制は、この利用者に実際的な権限があったと見なされる場合にのみ適用される。また、現行のほとんどのプライバシー保護法では、当初の目的に沿って個人情報を利用した後は情報の削除が義務づけられている。しかしながら、先ほどの新たなプライバシー保護の枠組みがあれば、情報削除の必要もなくなる。さしあたっては、この法律で個人情報を守る観点では、問題ないだろう。ビックデータを取り扱う企業、新たな法の改正によって守られるプライバシー、双方に利点があるだろう。

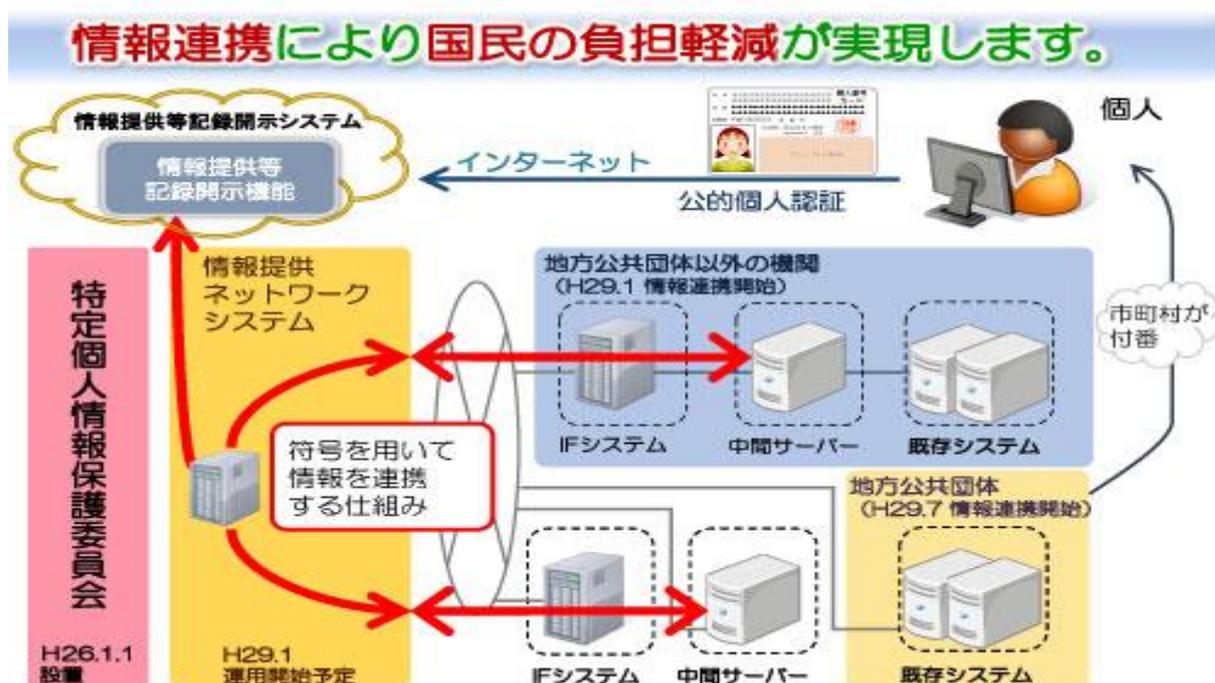
また、ビックデータのマイナス面として、人間が物事を判断するあらゆる場面で未来の責任を問うかどうかやビックデータで決まってしまうという問題がある。たとえば、経営者が従業員を解雇するかどうかや、医師が患者の手術をやめるかどうか、夫婦が離婚に踏み切るかどうかやあげられる。

人間は意外に‘データの独裁’に支配されやすく、明らかに変だと思っけていても、よく考えずに分析結果を鵜呑みにしてしまう。あるいは、データのためのデータ作りが自己目的化となってしまう、本懐を忘れ、数字集めにのみめり込んでしまうこともある。確かにデータが全ての原動力になって、その戦略が今日の成功に大きく寄与していることは間違いない。だが、ビックデータ予測に一切切を委ねるといことは、傾向や習性で罰する道具に使われる恐れもある。そうなれば、自由意志は否定され、人間の尊厳も破壊される。人は往々にして、メリットを聞いているうちに、本来不向きな分野にもビッグデータを応用したり、分析結果に過度な自信を持ったりする。そして、ビッグデータ予測が向上すればデータの利用はますます魅力的になり、データ至上主義に拍車がかかる。ビッグデータの予測が正確になるにつれ、実際の行為ではなく、予測される行為で人を判断する誘惑に駆られる。これはあまり好ましくなく、実際の行為についてのみ責任を問われるべきであり、将来の行動の統計的な予測について責任を問われることがあってはならない。

社会は安全になり、効率化するが、いっぼうで人が人間たる所以はもろく崩れ去る。よって、ビックデータは、社会での人間の選択の集団主義化をもたらし、自由意志を断念させる道具になる。(予測アルゴリズムに取って代わられる)

ビックデータの独り歩きを防ぐためには、我々がビックデータをコントロールできる範囲に留めるべきだ。つまり、予測に人間の関与が確実に含まれていることが大切であり、あくまでも「助言する立場」でいるべきである。

また情報管理の観点から見て、来年に施行されるマイナンバー制度を引き合いに出すと、個人情報と同じところで管理されることはない。例えば、国税に関する情報は税務署に、児童手当や生活保護に関する情報は市役所に、年金に関する情報は年金事務所になど、これまでどおり情報は分散して管理される。また、役所の間で情報をやり取りする際には、マイナンバーではなく、役所ごとに異なるコードを用いるので、一か所での漏えいがあっても他の役所との間では遮断される。したがって、仮に一か所でマイナンバーが漏えいしたとしても、個人情報が芋づる式に抜き出せない仕組みとなっている。



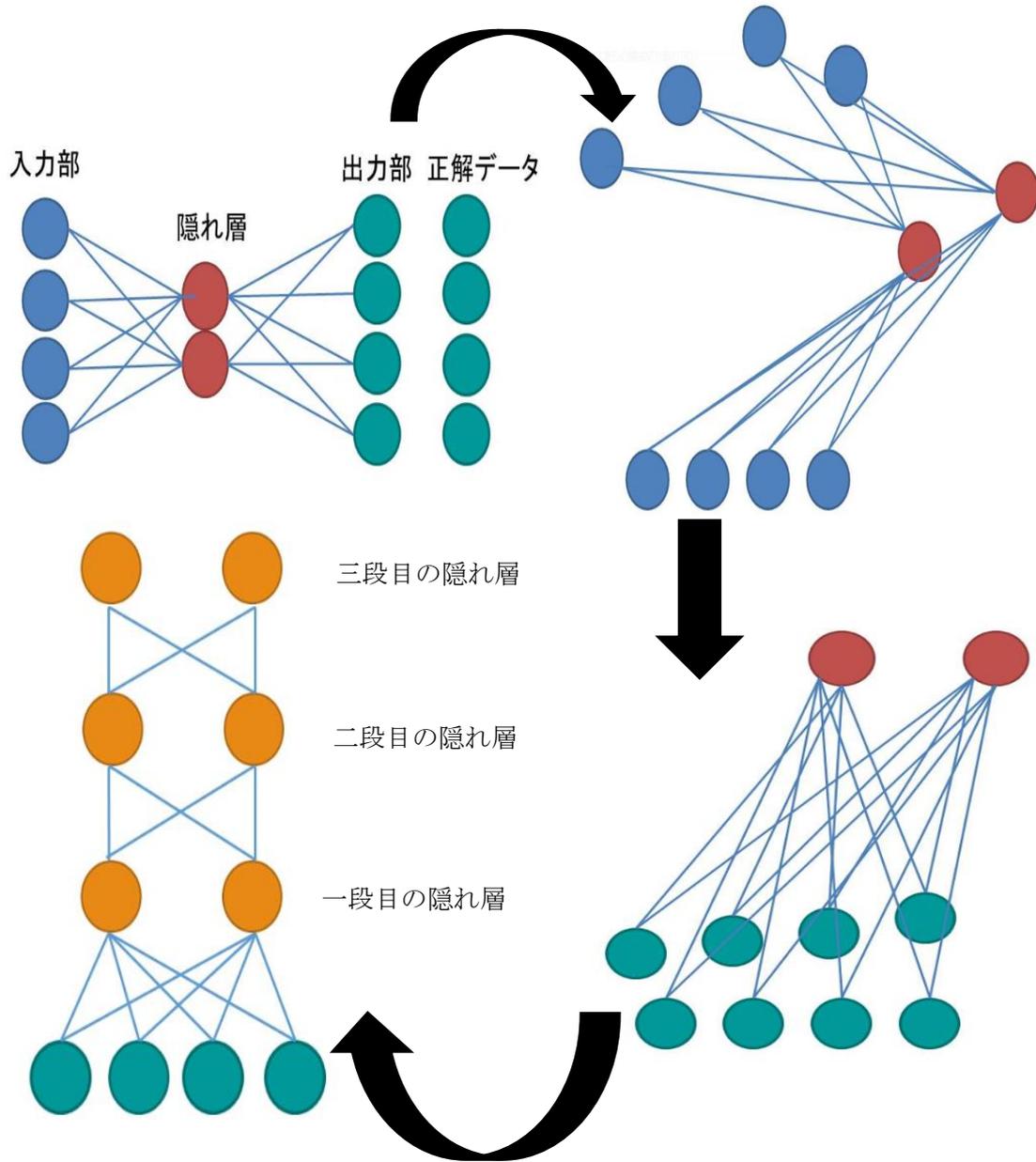
主題に基づき、実際にビックデータをビジネスに生かせるのかという点、ある企業ではビックデータを人工知能 (Artificial Intelligence) に取り込ませて、学習させて活用している。人工知能という言葉は、一度は耳にしたことはあると思うが、最初に人工知能とおぼしきものができてかれこれ 50 年経つ。あらゆる先達方が研究してきたが、まだ完全または完璧な AI はできていない。

1960 年の第 1 次 AI ブームでは、一見すると知的に見える様々な課題を解いていったが、煎じ詰めると、迷路を解くのも、パズルを解くのも、チェスや将棋に挑戦するのも明確な定義されたルールの上で解けばよいが、現実の問題はもっと複雑だった。そして頓挫した。冬の時代を経て 1980 年代になり再び第 2 次 AI ブームが起こった。知識を集約すればそれらしく振舞えるが、基本的に知識以上のことはできないといった問題があった。あまつさえ情報をインプットするのは人間であったため、人手とコストがかかりすぎてしまった。

しかしながらビックデータによって再び AI が注目された。(第 3 次 AI ブーム)

第 3 次 AI ブームはディープラーニング (特徴表現学習) を筆頭とする機械学習である。ディープラーニングとは、多階層のニューラルネットワークである。

またディープラーニングは従来の機械学習とは大きく異なる点が2点ある。1つは、1層ずつ階層ごとに学習していく点、もう1つは、自己符号化器（オートレコーダー）という「情報圧縮器」を用いることだ。



自己符号化器では、少し変わった処理をする。ニューラルネットワークを作るには、正解を与えて学習させる学習フェーズが必要だった。その場合、例えば手書きの「3」という画像を見せれば、正解データとして「3」を与える。ところが自己符号化器では「出力」と「入力」を同じにする。つまり正解も同じ「手書きの3」の画像として答え合わせをする。一見意味のないような行為に思えるが、入力と出力の間にある隠れ層にその画像の特徴を表すものが自然に生成される。またさらにこの作業を1段、もう1段と重ねていく。1段目の隠れ層を2段目の入力（および正解データ）として、コンピュータに学習させる。次々と繰り返して、多階層にしていくのである。よって多階層にすることによって抽象度を増し、高次の特徴量が生成される。

噛み砕いて説明すると生後すぐの赤ちゃんは、目や耳から入ってくる情報の洪水の中から、何と何が相関し、何が独立な成分かという「演算」をすごいスピードで行い、情報の洪水の

中から、予測しては答え合わせを繰り返すことでさまざまな特徴量を発見し、やがて「お母さん」という概念を発見し、まわりにある「もの」を見つけ、それらの関係を学ぶ。そして少しずつ世界を学習していく過程とほとんど同じである。よって画像を見せれば「これは人間だ」「これはネコだ」といった区別がつけられるのである。

また、ビックデータも相まってさらに正確・精密なものができる。人工知能が自発的に選択することができなかったのはコンピュータが概念を獲得しないまま、記号を単なる記号表記としてのみ扱っていたからだ。記号を「概念と記号表記がセットにしたもの」として扱ってこなかった、あるいは扱うことができなかったからである。そのために、現実世界の中から「何を特徴表現とするか」はすべて人間が決めてきた。決める他はなかった。ディープラーニングの登場は、少なくとも画像や音声という分野において、「データをもとに何を特徴表現すべきか」をコンピュータが自動的に獲得できるという可能性を示している。簡単な特徴量をコンピュータが自ら見つけ出し、それをもとに高次の特徴量を見つけて出す。その特徴量を使ってあらわされる概念を獲得し、その概念を使って知識を記述するという、人工知能の最大の難関に、ひとつの道が示された。いったん人工知能のアルゴリズムが実現すれば、人間の知能を大きく凌駕する。そこでビックデータを人工知能に学習させることで、人工知能が発展するのではないかと思った。しかしながらデメリットでも述べたように、データを過信するデータ至上主義が起りかねない。人工知能は「万人のためのもの」で、「人間の尊厳」を犯してはならない。あくまで最終決定権は人間がすべきである。

4. 今後の研究課題

現在、ディープラーニング画像からネコを認識あるいは概念の獲得が出来た。今後は概念の獲得から発展して、行動データと観測データを入力データとし、「行動と結果」の特徴表現と概念の獲得をして、さらに言語データから、言語を通じて知識を獲得できればなと思った。将来、今よりはるかに進化した人工知能が人間に対して何らかのアドバイスを行うようになった時、疑いなく受け入れることができるのか、人類社会と調和した社会像を目指すために人工知能に取って代わられた労働者…幸福の追求の観点から見て、どのような措置をとるべきかと懸案事項はあげるときりがない。しかし人工知能が今後の社会において重要な役割を担うことは十分にあるので、ニーズに応えるのではなく、自身の「こうであれば・未来はこうなってほしい」とい思いから、大学で人工知能を根底から学びなおし「ディープラーニング」の研究ができればいいなと思う。ひいてはそれを生業とし、世の中をもっと便利かつ豊かにしていきたい。

5. 参考文献

- V. M=シヨーンベルガー&K・クキエ 『ビックデータの正体』 講談社 2013年
マイナンバー社会保障 <http://www.cas.go.jp/jp/seisaku/bangoseido/faq/faq4.html>
ビックデータへの道 第1回「ビックデータとは」日立
<http://www.hitachi.co.jp/products/it/bigdata/column/column01.html>
東京大学准教授 松尾豊 『人工知能は人間を超えるか』 角川選書 2015年