

Application of Big Data and prospect of medical care and nursing

メタデータ	言語: jpn 出版者: 公開日: 2021-02-10 キーワード (Ja): キーワード (En): 作成者: 勝部, 憲一 メールアドレス: 所属:
URL	https://doi.org/10.50818/00000041

【総説】

ビッグデータ活用と医療・看護の将来

Application of Big Data and prospect of medical care and nursing

勝部 憲一

Ken-ichi KATSUBE

要 旨

ビッグデータとはインターネットで接続されたあらゆる器機が収集するデジタルデータで構成されるストレージを基に、多種類の解析を同時並行でおこない、あらゆる将来変化の予測をおこなうことである。近年のコンピュータ技術の急速な発展で可能になった大規模データの蓄積と超高速な演算解析で可能になった解析がその基盤になっている。我々の日常生活一般に今後広く影響すると考えられるが、医療・看護も例外ではない。医療の広範な進歩に貢献し、患者の診療体制も近い将来大幅に変化するだろう。

キーワード：ビッグデータ

I. 諸言

近年ビッグデータという言葉がニュースなどでマスメディアが盛んに取り上げるようになった。ビッグ「大きな」データ「記録」となるわけだが、その実体は素人には依然として判然としない。しかしその技術は人間の行動解析から企業の効率的なマーケティング活動、気候の変動解析から地球温暖化の進行予測などあらゆる複雑系解析の局面で導入が試みられている。本稿ではビッグデータについての概念と今後の見通しを述べ、医療や看護学の中にいる者がそれをどう活用し、将来の発展につなげられるか考察をおこないたい。

II. 私からみたビッグデータに至るまでのデジタルデータの歴史

私たちがパソコンを日常のツールとして使うようになってから既に30年以上経つ。私が学部生時代だった1980年代前半までは、パソコンはまだ非常に高価で一部の理数が得意な学生が趣味的に買って使うものだった。教科レポート作成も手書きが普通で、文字情報をテキストファイルとして蓄積することもなかった。専ら先生や先輩からもらうコピーな

どを手がかりに勉強するアナログ世界だったと言える。大学院生となった1980年代後半になると、論文を含めて文章は基本的にパソコンでテキストを作成して、整理・投稿するのが普通になった。しかし写真などの図表に関しては当時のパソコン（NECのPC98シリーズ）では演算子やメモリの処理能力を超えており、ワークステーションと称する中型以上のコンピュータ（UNIXなど）が必要だった。それでもA4版程度の大きさの白黒の単純なオートラジオグラフの描画（500kバイト程度）にも5分以上かかり、イライラしたことを思い出す（富士フィルムのBASシステム）。検出器のIP（イメージングプレート）の性能の方も高いとは言えなかったが、しかし当時のワークステーションの処理能力は2メガバイト程度だったので、それもやむを得ないことだったのである。

一方ゲノムシーケンスはAGCTから成る文字配列で基本的にテキストだが、その類似性解析や照合も大変だった。1990年代前半当時、フランスの国立研究所に留学していた私は、自分が得た新しい遺伝子の塩基配列解析のためにパリ中心部のパスツール研究所にデータが入ったフロッピーディスクを送った。つまり当時のコンピュータはデータベースに使われるものであるかにかかわらず、今で言うスタンドアローン（インターネット接続なし）が普通だったのである。パスツール研の遺伝子解析室に

あるコンピュータのデータベースに照合してもらおうのだが、わずか数百ベースの配列解析（100 キロバイト程度）になんと二日かかりだった。ひとつデータを送って解析してもらい、またしばらく順番待ちだった。しかしちょうどその頃、在籍していた研究所にもインターネット回線が入り、電子メールのアドレスをもらった。「あれ、これは fax より簡単。ミニテルより便利そうだな（ミニテルはフランステレコムの子会社で、単純なミニコン。1980～2000 年代活躍。電話線に接続した9インチくらいの白黒モニターで切符予約などができた）。データとか送信できるならわざわざフロッピーを郵送する手間もないし、回答もメールでもらえるかも」と思ったのを思い出す。しかし LAN 回線の設定が悪く、LAN 端末が各室に配置されていても実際にインターネットにアクセスできるパソコンは所内に2台くらいしかなかったことを思い出す。この直後私は日本に帰国することになり（1995年）、帰国後インターネットの爆発的な発展に出会うことになった。

帰国後の東京医科歯科大着任から本年（2015年）ですでに20年経った。今は皆様もご存じのように我々の生活はパソコンとそれに接続するインターネットなしでは成り立たない。パソコンの性能も飛躍的に向上し、周辺付属器も含めて随分安価に買えるようになった。論文作成に必要な図表作成は、精密な写真イメージを含めてすべてパソコンでおこなえるようになった。投稿については紙媒体での郵送は2005年頃にはほぼすべてなくなり、数メガバイトの電子ファイルをインターネットのジャーナルサイトに送る。査読の方もこのサイトにアクセスして論文ファイルをダウンロードして読み、レビューを書いて送る。遺伝子解析にしても、今ではアメリカのNCBI（National Center for Biotechnology Information 米国立バイオテクノロジー情報センター）の遺伝子データベースにアクセスして手持ちの遺伝子データをアップロードすれば、類似性のみならずタンパクとして機能予測なども瞬時に回答されるようになった。

こういった電子化の恩恵は基礎医学・生命科学系といった研究方面ばかりでない。病院カルテも2005年くらいには日本にも電子カルテシステムが導入され、CTやMRIの画像診断データもすべて電子化されるようになった。カンファレンスなどがあると、昔は大量の重いX線写真フィルムを運んだものだったが、今はデスクの端末モニターに瞬時に映し出され

るようになった。しかも的確な写真が既にピックアップされており、あれこれ探す手間が大幅に省けるようになった。日本全体での病院電子カルテ普及率は2014年末現在まだ3割程度だが、医療情報デジタル化の拡大は続くと考えられる。この便利さ・速さにはまさに隔世の感がある。このような変化で現在では大量の大型データが無数にインターネット上にあり、それらを容易に見ることもできれば、ダウンロードして操作することもできるようになった。しかしその膨大なデータはさまざまな形式で保存されており、全体を通底する解析や理解は難しいと私は思っていた。それを乗り越えようとしているのがビッグデータ概念である。

Ⅲ. ビッグデータとは何か？

ヒトゲノム解析を例に考えてみよう。ゲノムとは「ある生物種を規定するのに必要かつ十分な遺伝子の総体」である。19世紀の遺伝学の概念から由来しているので随分抽象的な定義だが、早い話が我々ヒトならヒトという生物が持っているDNAの塩基配列がすべて判れば、それがゲノムの実体となるわけだ。正確に言うと遺伝子としてタンパク生成に直接関わらないDNAも非常に多いが（というよりヒトの場合90%近いDNAは直接タンパク生成（翻訳）に関係しない）、その多くがDNAの構造安定性や非翻訳小型RNA生成を通じて遺伝子制御に貢献していると考えられている。その配列はおおよそ30億塩基対（6ギガバイト相当）と考えられ、1980年代後半からアメリカを中心とした各国の共同作業で進められた。その結果最初のヒト全塩基配列が明らかになったのは2003年であった。DNA塩基配列が有限とわかってはいてもそのギガ級の膨大な情報がすべて判ったことはすごいと当時私は感銘を受けた。しかしながら、これを利用して各個人の病気と関連する因子を見つけて将来の健康を予測したり病気を予防したりするのは大変なことだとも感じた。計画が本格化してから、ある一人のヒトの全DNA配列が明らかになるまで、実に10年以上かかった。そもそも一人のゲノム解析にかかる費用は、約100億円だった。この成果を利用できるのは世界でも少数の大金持ちだけで、大多数の普通人の福祉に関係しないであろう。それに人一人のゲノム解析に10年もかかるなら、病因がわかるころには大抵のひとは寿命が来てしまう

とも思った。

ところがその後 DNA 塩基配列解析（シーケンシング）の技術に画期的な進歩が起こる。いわゆる次世代シーケンサーと呼ばれる手法で、短く断片化した DNA の配列を同時大量に読み取り、得られた配列を大容量・高速度のコンピュータで解析してたちまちにしてつなぎ合わせる方法である。塩基配列読み取りの実験手法はここ 30 年以上基本的に同じなので、この技術革新はパソコンおよびそのソフトの進歩に拠るところが大きい。このようにして今ではわずか数日あればある個人の塩基配列はすべて調べることができるまでになった（しかも 10 万円程度で！）。おかげで現在ある種の癌など遺伝子変異が主たる原因の病気では、個人の将来的な罹患を予測できる段階に達している。しかし人類の病気全体からみればそれはごく一部に過ぎず、例えばあるひとが肥満になったり高血圧になったりするよくある病気のメカニズムやリスクの詳細は依然として明らかでない。そのため今得られている各個人の DNA 塩基配列をすべて解析し、その個人の検診データや病歴を照合して因果関係を追求する動きが加速している。このため病院や研究機関、検査会社などに蓄積しているさまざまな種類のデータをインターネットを通じて一挙に収集し、その関連を調査する手法が本格化している。このような動きがビッグデータの源流のひとつになっている。現在ビッグデータと呼ばれる技術が扱う情報量の大きさはペタバイト級以上と考えられている。論文情報量は数メガバイトで、ヒトのゲノム情報量はその千倍のギガバイト級になる。ペタバイトはギガバイトの百万倍のオーダーとなるわけで、その巨大さがわかると思う。この蓄積している情報をストレージ（storage）と呼ぶが、ただ蓄積しているだけでは意味がない。巨大な情報を解析する新たなアルゴリズムの開発やそれを高速で運用できるコンピュータの性能向上があって、初めて有用な解析や予測をおこなえる。当然それを専門に開発する技術者の育成も不可欠といえる。

ビッグデータはもちろん医療関係ばかりでない。2013 年初夏、ビッグデータに関連して話題になったのは、JR 東日本が Suica の情報を販売する計画を発表したことだった。首都圏を JR で移動する人のほとんどが Suica を使うと考えられるが、その移動情報は JR 東日本のデータベース上に記録されていく。これを個人名が特定されない形式に変換して販売する

というものだった。データはある駅で列車に乗るひとがどこの駅に向かうのかということだけだが、それが天候の局所的变化、イベント情報など他のビッグデータと組み合わせていけば膨大な解析をおこなえる。Suica の情報が提供されれば、人間移動の社会的な研究だけでなく企業のマーケティング調査や震災のような大規模な事故・事件発生時のセキュリティ対策にも大きな貢献をしたはずだった。ところがこのビッグデータの販売は 2013 年 6 月から開始されたが、なんと 7 月末に停止されてしまう。このデータを利用することで Suica 利用者の個人情報が漏れることが強く懸念されたことが大きな理由だった。実際には個人情報が漏れないよう入念な操作をされた上で情報が提供されることになっていたが、データを収集される側の JR 利用者の懸念が予想以上に強く、批判が殺到したのだ。この一件は、ビッグデータに関してその内部に含む可能性がある個人情報がどう取り扱われるのか、一般人はほとんど理解できていないという問題を露呈させた点では意味のある出来事だったと思う。

IV. ビッグデータに必要なデータ基盤と解析法

ビッグデータストレージとしてペタバイト以上という超巨大な情報がさまざまな形式で蓄積され続け、しかもその内容は時々刻々と変わっていく。今話題になっているのはクラウドサービスである。例えば本学でも使用されているサイボウズがクラウドサービス的一种である。サイボウズで提示されるメールは web メールで、いつでもどこでもアクセスして読める点が良い。掲示板やスケジュール管理などもすべてサーバ上で記録され、我々はそこに端末パソコンでアクセスして読んだり書き込みをおこなったりしている。インターネットでつながれたパソコンならどこからでもアクセス可能で、そのネットワーク構成は変幻自在である。この可変なネットワーク構成は曖昧模糊としており、それに「雲（cloud）」のイメージをあてはめて呼んだのが、クラウドサービスである。先ほどの Suica にしても Suica それ自体に記録があるわけではなく、クレジットカードと同じで個人情報や使用履歴はデータベースを保持する JR 東日本のサーバコンピュータの中にある。当然ながらインターネットはさまざまなクラウドサービスを含んでおり、その全体は無限の広がりとも組み合わせを

もつ。これがビッグデータの基盤のひとつと考えて良い。そしてその基盤を支えているのが、デジタルデータを収集する色々な測定器機の進歩と活用である。

このような整理されたデータベースとは別に、個人や企業がHP上に掲載するブログもビッグデータの情報源として重要である。Googleなどの検索エンジンがロボット探索の進歩を背景に強力になった現在、個人でも容易にそれらの雑多な情報を収集できるようになった。情報量としては整理されたデータベースにくらべて桁違いに多く、おそらく数万倍以上あるだろう。

またコンピュータとして認識されていない器機もビッグデータを構成すると考えられている。例えば家電製品でもインターネットで結ばれているものが多くある。いわゆるデジタル家電と称するもので、その源流は1984年当時のキャプテンシステムにある。今ではデジカメやiPad、DVDレコーダなどといったものだけでなく、「見守り家電」と呼ばれるインターネットで繋がった電気ポットやテレビも、人的活動の情報を提供するデジタル家電である。つまりこれらネットに接続できる機械はすべてビッグデータの一部を構成しようと考えられる。

しかし従来こういった情報は全体として規格化されておらずジャンク（ゴミ情報）であり、役に立つものとはみなされていなかった。それを「宝の山」と考えて発掘しようとする動き（マイニング）が近来盛んになってきた。ファイル形式もさまざまなこれらを解析して意味のある相関性や方向性を見いだす作業をデータマイニングと呼ぶ。たとえばブログやツイッターで発信される個人のメッセージはそれだけでは意味がほとんどない。しかし発信日時を関連づけ、個人の背景を継続的な発信記録から推定することで、断片的な情報の集合から意味のある相関性や方向性を推測することができるようになる。このデータマイニングには、我々がよく使用する表計算アプリであるExcelの通常運用では到底追いつかないのは明らかである（Excelが処理できるのは100万件まで）。そのため特殊なプログラム開発が必要となるが、これは情報科学に精通する専門家でなければ、手も足も出ない。そのため現在ビッグデータ解析の専用システムを販売する会社が続々と出ているが、まだ高価である（年あたりの契約で数百万円以上）。ここが今ビッグデータを取り扱う最大の難点で、

特に個人の研究者が直面する壁である。今後開発競争が活発になって低価格に移行すると思われるが、現状ではそれがいつになるかははっきりしない。ただまったく解決法がないわけでもない。実はExcelを応用する方法である程度容量以上の大規模な情報を扱えるようになった。

V.ビッグデータと医療・看護,そして我々の将来

医療・看護にビッグデータを活用した場合、どのような展望が開けるのだろうか。まず言えることは、今まで医療・看護に携わる者が個人個人で蓄積したノウハウを合体させることである。その結果として患者との対応で早い段階から最適な道を選ぶことができるようになり、また患者との関わりの将来についても予測が可能になる。何よりも大きいのは経験に基づいて集積していたノウハウが標準化され、だれでもそれが利用可能になる点である。経験が少ない者でもその利用でさまざまな経験をヴァーチャルに学ぶことができる。

ただしそのビッグデータの基盤となりうる医療のデジタル情報の蓄積に関しては課題がある。先ほど述べたように病院カルテの電子化は国内ではまだ十分でない。先ほど述べたように病院でのカルテ電子化は全体として3割程度だが、大規模病院では70%近いのに対して中小規模の病院では20%程度にとどまっている。デジタル化されていない情報はビッグデータにとって無に等しい。ゲノム解析では次世代シーケンサー開発で膨大な塩基配列情報が集まって初めてビッグデータ構築が可能になったことを考えると、医療における情報電子化はもっと急がなければならない。しかも医療情報でクラウドサービスの利用はまだ一般的ではなく、各病院のコンピュータ上のデータベースに蓄積している。いかに大病院といえども、患者数には限界がある。今電子化が遅れている中小規模の病院に対して、電子カルテ業界は地域医療連携を見据えた上で、クラウドサービスでの提供を進めている。在宅医療・訪問介護といった患者と直接関わる医療から地域包括ケアマネジメントや医療用画像システムを応用した遠隔医療（テレパソロジーなど）など大規模な地域医療システムを推進する上でもクラウドサービスの活用は今後必須であり、医療費の抑制（低価格化）を促す効果もある。業者の参入が増えてサービスの提

供価格も徐々に低下しており、ここ数年で急速に進行すると思われる。大病院も含めた医療情報のクラウド化がいつまでに完成するかがビッグデータ構築成功の鍵である。しかしJR東日本のSuica情報販売の混乱でもわかるように、個人情報の保護が徹底されること及びその保護が医療利用者に十分周知されることが前提条件として重要である。

看護に話を絞ると、その情報デジタル化は電子カルテ普及とほぼ歩調を合わせていると考えられる。看護記録の電子化とその情報集約がビッグデータ構築に必須だが、病院の支援や協力体制が十分とはまだ言えない。看護記録は患者との接触がもっとも密接な内容なのが普通で、その匿名化を図って情報共有するのはなかなか難しいといえる。しかし医療情報の収集は医療者ばかりからでない。医療利用者すなわち患者自身が発信する情報もある。つまりブログやツイッターに発信される医療に関係したメッセージは、医療行為を反映したものといえる。特に看護に関する発信は多い。このような情報をどのようにして組み込み、ストレージを構築するかが重要な問題になってくる。またこういったビッグデータの集積には国内の情報だけでは十分と言えない。欧米など先進国やアジア各国のビッグデータを集めれば、もっと大きな流れを把握することが可能になる。そのためには各国の看護医療に関わる研究者を糾合し、データの合体を図るようにすることができる。英語がその基本になると考えられるが、各国の情報をどのように規格化するかという議論は遠からず始まると考えられる。

デジタル家電類のネット接続を取り上げたが、最近Internet of Things（通称IoT）という概念が急拡大している。直訳すると「モノのインターネット」となるが、要するに人間が使用する器機のデータをインターネットを通じてすべて集め、解析と利用を試みる考え方である。そのためあらゆる製品にインターネットと接続できるセンサーを埋め込み、利用状況などのデータを集める。また解析を容易にするためデータの規格化の動きが広がっており、現在GoogleやMicrosoftなどがその規格を巡って激しく覇権を争っている。この規格化が成功すればその後のビッグデータ活用は急速に進歩するだろう。

ひとが仕事を終えて帰途につこうとすると、車のモニターは道路の混雑や渋滞を予測して最適なルートを提示する。帰宅すればエアコンがちょうど良い

時間にオンになっていて、吹き上がる汗もがたがたの震えもない。ちょうど上がった風呂に入って出てくれば、炊飯も完了。炊きたてあつあつご飯を食べていると、掃除・洗濯も完了した。そのような夢のような自動生活の時代も目前となっている。医療や看護でもそういった変化が大いに期待される。患者からの採血でゲノム解析をおこない、疾病に対する脆弱性を予測する。患者が自宅生活するなら、血圧や心電図・呼吸の変化を自動的に解析して、ゲノム解析を合わせて近い将来の病態を予測する。もし病態が悪化しそうだ判断されれば、夜間なら緊急入院に備えた救急車の手配や対応できる病院の自動的な選択をおこなう。病院に当直ないしオンコールで対応できる医師や看護師などの医療スタッフも、病院が決まれば打診して、到着次第治療にあたるようにする。判断が難しいと思われる病態なら、遠隔の専門医を待機させて送ったデータをもとに適宜アドバイスを受ける。こういったことをひとを介することなく自動的にかつ速やかにおこなえる日はそう遠くないと私は考える。

受理日：2015年1月26日

