

医療におけるビッグデータ活用

帝京大学医療情報システム研究センター
教授 澤 智博 氏



■医療におけるビッグデータ活用

ビッグデータを特徴付けるのは4つのV (Volume (量)、Variety (種類)、Velocity (速度)、Value (価値))で、ビッグデータに期待されるのはこれまでに扱うことのなかったデータ量、データ種をこれまでより短時間でデータ処理し、これまでになかった価値を生み出せるかもしれない、ということだろう。病院で行われることは診断もしくは治療なので、医療におけるビッグデータ活用で可能になると考えられるのは主に、1. 新たな診断方法の開発、2. 新たな治療方法の開発、3. 効率の良い診療方法の模索、4. 診療水準の向上、5. 病院や医師による診療水準のばらつきをなくすことである。

■病院が扱うデータ種と量

2000年代、わが国では電子カルテが診療効率向上、研究データ蓄積等に大きな役割を果たすと期待されていたが、期待通りの成果は上がらなかった。その主な原因の1つはコンピュータに人間を超える能力を発揮させるために必要な大規模データが十分蓄積されていなかったためである。当院は1,154床で、板橋区の本院開院時から約5年間で登録カルテ数延べ110万件、オーダーが6,600万件、生体モニタが24億8,000万件、放射線画像等は78テラバイトに上っている。紙媒体の記録物もスキャンにより電子化することで多くのデータの蓄積は実現しており、病院内のあらゆるデータはコンピュータ上で閲覧可能である。

■コンピュータにより可能となったゲノム医療

近年ビッグデータとともに活用が期待されているのがバイオデータである。2003年時点でヒトゲノムの全塩基配列の解読が完了し、2013年にはほぼ1,000ドルで一人の人間のDNA解読が可能な状態になっている。人間のDNAは31億塩基対あると言われ、1塩基2ビットと換算すると一人当たり約6.4ギガ、全人類70億人分を計測してもコンピュータには処理可能な量である。遺伝子発現の状態を表したヒートマップを比較すると類似の機能を果たすDNAを見つけることができるようになっている。

急性骨髄性白血病 (AML)、急性リンパ性白血病 (ALL)、慢性骨髄性白血病 (CML)、慢性リンパ性白血病 (CLL) とさまざまな種類がある白血病も遺伝子発現を調査し、コンピュータによるクラスタリング分析から ALL と AML が目視の病理診断を超えて遺伝子発現データとして鑑別可能になっている。また、かつては1種類の病気と考えられていた diffuse large B-cell lymphoma (びまん性大細胞型 B 細胞リンパ腫) も、遺伝子発現データをクラスタリング分析すると生存率に差がある2種のクラスタに分かれることがわかり、ゲノム関連データが実際の診療で活用され始めている。

■医療ビッグデータを活用するテクノロジー

病院の診療行為は殆どすべてがオーダーという概念によって成り立っており、当院も1日約200万件のトランザクションが動いている。トランザクションのDBに誤ったSQLを流して病院システム全体が停止する、といった事態は避けねばならないので、トランザクション用DBとは別に分析用のDB(In-Memory データベース)を使用してスタースキーマ、スノーflakeスキーマといったデータモデリングを行っている。

私が所属する日本麻酔科学会では、麻酔や手術に関する問題点を調査するため全国の症例を集めており、2011年時点で330万件から成るDBを持っており、それによると手術後7日以内の死亡率が1万症例当たり3.14となっている。しかしこれは、通常死亡率が高くない健常人の小手術から出血多量で搬送された重症のけが人といったあらゆる状態の患者全体のデータなので、手術を受ける患者にこの数値を話したところで現実味が薄い。このDBもスタースキーマを作ってモデリングすると、細分化が可能になり、各症例に合わせた偶発症の発生率の話もできるようになる。

Hadoopは大規模データの分散処理に向いているので、血圧や心拍数などのストリーミングデータをパターン分析することで、医師のスキルに合わせた支援が可能となり、手術中の適切な麻酔管理を実現する、といった役立て方が可能だろう。

■医療ビッグデータの活用に不可欠な仮説と検証

医療ビッグデータの活用に不可欠なのは、大量のデータをコンピュータが自動的に処理して何かが見れるのを待つのではなく、人間が自らの脳で、データ処理から導かれること(仮説)を生み出し、実際の処理から仮説を検証、シミュレーションすることだ。どのデータに対してどういう処理をするかは人間自身が考えるしかない。医療にしても企業にしても、現場の感覚を持った人間がデータを処理できるかが肝要である。

■医療現場と個人情報

病院には、名前が書かれていなくても個人を識別できる情報がたくさんある。例えば当院の5番手術室で、2014年1月30日14時に開始された手術は1件しかなく、施術された個人名の特定はされなくても識別可能である。医療現場には、個人情報に関わる問題が多数存在する。先ほどゲノム情報の解析が非常に進んだという話をしたが、将来例えば、リスク細分型保険が適用され、がん発生率80%の遺伝子を持っている人は通常の遺伝子を持つ人の3倍の保険料を払わされる、という可能性が絶対ないと言えるだろうか。また、遺伝子データというのは、先祖や子孫にも関連する一族情報なので、現在の社会状況に照らして問題ないという判断から公開した遺伝子情報によって、子孫が30年、50年後の社会で差別を受けるかもしれない。さらに、データや判断が誤りである可能性もゼロではなく、現在でも誤診、取り違えや誤差はあり、データだけが独り歩きして他者が自分のことを決定するような社会は怖い。人間をidentifyするのはDNAが31億塩基対集まったものであり、コンピュータにとっては70億件(全人類)から1件を見つけるのは造作もないことだ。テクノロジーによってこれだけ容易に個人を特定可能になった現在だからこそ、データを何でも一方的に分析し、ありとあらゆる判

断を下して良いのか、と思う。情報の分析、利用、そして医療に携わる者はその危険性についても十分配慮する必要がある。

尚、本日の講演内容の詳細は

◆ビッグデータ活用の“特効薬”はあるのか？

<http://itpro.nikkeibp.co.jp/article/COLUMN/20130704/489370/>

◆iPS細胞とITの接点はどこにあるのか？

<http://itpro.nikkeibp.co.jp/article/COLUMN/20131009/510005/>

◆人間は誰でもミスをする、システムは必ず障害を起こす——トラブルを減らす“6つの知恵”

<http://itpro.nikkeibp.co.jp/article/COLUMN/20131212/524346/>

◆「何か見つかるかもしれないから全部出して」でいいの？——ビッグデータについて考える

<http://itpro.nikkeibp.co.jp/article/COLUMN/20140226/539660/>

に掲載しているので、参照されたい。