

シソーラス^{*}の標準化と共通システムの構築を

◎ 医療ビッグデータの活用を促進するためには

黒川 日本には活用し得る医療ビッグデータはすでにくつかありますが、それらはネットワーク化されておらず、アクセスも限られた形になっています。医療ビッグデータ活用によるメリットは前述した通り大きいことから、何らかの制限は設けるにしても基本は公開し、問題意識がある研究者や組織が自由にアクセスできるようにすることが望まれます。

さらに、ネットワーク化する際には、用語や情報の階層化におけるシソーラスの標準化が必要です。また、異なるベンダーがそれぞれでシステムを構築している現状では、ベンダーごとのシステムと並行して、互換性のある共通システムやコンバーターのようなものを作り、それらをつなぐ方が現実的だと考えます。医療ビッグデータの活用が進めば、治療も効率的に行えるため、無駄な医療費も削減できます。そのためにも1日も早く国や自治体などが主導し、より多くの情報が集まる共通システムを作ることが求められます。

*シソーラス

言葉を同義語や意味上の類似関係、包含関係によって分類した辞書またはDB。

国立情報学研究所所長 東京大学教授

喜連川 優氏



プロフィール

1983年東京大学工学系研究科情報工学専攻博士課程修了、工学博士。東京大学生産技術研究所教授、東京大学地球観測データ統合連携研究機構長、2013年4月より国立情報学研究所所長。データベース工学の研究に従事。内閣府最先端研究開発支援プログラム「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的サービスの実証・評価」を中心研究者として推進中。2013年6月より情報処理学会会長。情報処理学会功績賞、ACM SIGMOD E.F Codd Innovations Award受賞。2013年紫綬褒章。ACM、IEEE、電子情報通信学会ならびに情報処理学会フェロー

「Actionable Info」「Actionable Knowledge」

◎ データから知識を得るだけでなく、いかに活用するか？

喜連川 生のデータから情報、情報から知識へと抽象化されるプロセスにおいて、知識を得ることが最終目標だと考えがちですが、米国では、そこからもう一段階進んだ「actionable info」「actionable knowledge」という概念が重視されています。

これを直訳すれば、「行動に結び付けることが可能な情報や知識」となりますが、単に知識を得ることでよしとするのではなく、課題を前にして一体どういう行動をすればいいのかという、行動のための具体的方策を導出しようという動きです。

例えば、がん治療で原因や治療法の知識を得ることは重要ですが、いま社会が求めているのは、さらに一歩踏み込んだ予防法や早期発見の方法です。つまり、データを分析して情報や知識を得た後、それを基にがん予防や早期発見のシステムの構築が求められているのです。

また、「データ」「情報」「知識」を並べたとき、生のデータよりもそこから抽出された情報や知識の方が価値があると思いますが、近年はデータの重要性が見直されています。

一例を挙げると、自然エネルギー発電において、電力の安定供給には発電量の正確な予測が欠かせず、発電量を左右する天気の変化を把握することが重要です。太陽光発電であれば、パ

ネルに当たる太陽光の量まで予測できることが理想で、ここで求められるのは、従来のような晴れや曇りといった予測ではなく、細かい雲の動きを示す生のデータです。

このように、ある分野で集められたデータは、全く異なる分野でも利用価値を産むこともあるため、データそのものが重要なコンポーネントとして注目されてきています。

希少疾患の治療法研究が促進

◎ 医療ビッグデータを活用するメリットとは？

喜連川 ビッグデータという言葉は、2010年に米国政府が採り上げたことで一気に広がり、以来、世界中で関心を集めています。私は20年以上前から情報の蓄積と活用の必要性について主張してきました。

学生時代、情報通信の世界では通信が中心で、コンピュータはマイノリティでした。しかし、新しい技術であるだけに潜在的な可能性も大きいと感じてコンピュータを専攻し、しかも当時主流だった計算速度ではなく、記憶量（情報の記憶・管理）の分野を選んで研究を続けてきました。その中で感じたのは、コンピュータ技術の発達に伴い、取り扱う情報量が格段に増えれば、その管理や活用が大きな課題になるだろうということでした。ビッグデータの重要性が叫

ばれる昨今、まさにその予測が現実のものとなったわけです。

医療ビッグデータの活用で期待しているのは、希少疾患の治療です。症例が少なく、病態の変化が大きい希少疾患は、研究に使えるデータも限られています。しかし、全国規模で同じ病態のデータを集め、過去の事例も含めて解析すれば、治療に役立つ情報が得られる可能性も高まります。患者数が多い病気は、経済的インセンティブが働くため治療法の研究も進みますが、希少疾患はそれが期待できません。この分野こそ、ビッグデータを活用する意義があると考えられています。

もう一つ期待している分野が、治療情報の一元管理です。別々の医療機関で受けた治療や処方薬の情報が一元管理できれば、薬の飲み合わせ等の問題や病気の予防をはじめ、将来の個別化医療につながる道も拓けていくでしょう。

先にビジョンを示し、各分野が連携することが不可欠

◎ 情報システムの基盤は、独自性より統一性？

喜連川 しかし、これらを実現するためには、病院間のデータベースがつながることが前提です。現在、大多数の病院は、それぞれ独自のデータベースを構築しているため、病院間のデー

タ共有が進まず、有意義な研究も生まれにくい状況です。独自性をアピールする名目で、病院ごとのシステムを作った結果、病院間の連携が進まない上、システムの維持管理の負担も大きく、疲弊している病院も少なくありません。

情報システムの基盤は、今や橋や道路のように、社会にとって欠かせないインフラであり、独自性も重要ですが、一方で統一性を持たせることの重要性を再認識する必要があります。組織をまたいだデータの共有ができると同時に、システムの運営や維持管理も効率化できれば、医療機関だけでなく、社会にとつてのメリットにもなるはずです。システムを構築する企業も含め、関係者全員がその認識を持ち、情報システム基盤の統一に向けた議論を進めるべきと感じます。

ビッグデータの構築や活用に向け、研究対象の分野に通じたデータ分析の専門家である「データサイエンティスト」の育成を求める声がありますが、データを分析する人が対象分野に通じていることは理想ですが、学問の細分化が進んだ現在、異なる2つの分野の専門家になることは、現実的に難しいでしょう。また、業際、学際における活動には、両者の言葉を理解するコーディネイターが必要だといわれますが、意思疎通のしやすさを考えると、参加者は少ない方が望ましいのです。むしろITの専門家と、対象となる分野の専門家が密に連携することの方が現実的です。つまり、データサイエンティストを育成することも必要かもしれませんが、熾烈

な競争の中、いち早く一歩踏み出すためには、育成を待つのではなく、両者が歩み寄りという現実的な方法も忘れてはいけないと思います。

その際、重要になるのが、対象分野の専門家が個々の問題を提起する。つまり、「このデータを活用して何がしたいのか」を明確にすることです。

「これだけデータあるので、何かできないか」といった依頼がありますが、データを提供すれば、玉手箱を開いたように夢の世界が作れるわけではありません。

「こういうものを作りたい」というビジョンが先にあり、システムを使う側と構築する側の双方がその実現に必要な知識を習得することが肝心です。実際、これまでもさまざまな分野の情報化にこの方法で取り組んできましたが、別段、問題はありませんでした。

例えば、医療のデータベースを構築する際には、同じ病気でありながら、いくつも異なる表記があることなど、事前に情報をシステム構築側に伝えておけばいいのです。不足している知識は、システムを構築する中で活発にコミュニケーションを行うことで補えます。相手の領域を完全に理解することはとても難しく、それを要求することは困難です。むしろ互いに補う努力をすることこそが大切で、チームとしてある程度、共通言語があれば機能すると考えています。

情報を活用する上で個人のプライバシーを担保することは、人間が独創的な活動を行うため

に極めて貴重な要件です。特に医療情報の場合、希少疾患になるほど個人を特定しやすく、それを懸念する人が多いのも事実でしょう。ただ、希少疾患の患者が同じ病気になった人に対し、自身の情報を共有することに抵抗があるかという点、一概にそうとも言えないと思います。

むしろ、情報共有によって、新たな治療法が見つかったり、同じ病気の人の役に立つことができるのであれば、賛成するケースも多いのではないのでしょうか。

丁寧な意思確認をし、本人から同意を得ることで、当該問題を乗り越えることができる場合も十分にあると考えています。