

第3263号

いいね! 30

G+

Share

チェック

ツイート

関連書

【インタビュー】医療ビッグデータの活用(喜連川優)

【寄稿】患者自己評価で有害事象を測定するツール開発(山口拓洋, 川口崇, 宮路天平)

【インタビュー】内科医が知っておきたい, 母性内科の視点(村島温子)

【連載】高齢者の「風邪」の診かた(3) 鼻症状は薬剤性, 喉症状はカンジダやヘルペスを疑う!(岸田直樹)

【連載】栄養疫学者の視点から(12) アルミニウム・認知症・HPVワクチン(今村文昭)

MEDICAL LIBRARY 書評・新刊案内

第3263号 2018年3月5日

【interview】

個々の医師, 病院の枠を超えて始まる
医療ビッグデータの活用

喜連川 優氏(国立情報学研究所所長/東京大学生産技術研究所教授)に聞く

2017年11月, 国立情報学研究所(NII)に「医療ビッグデータ研究センター」(以下, 研究センター)が新設された。医療ビッグデータを活用した人工知能(AI)開発は, 診療補助, 医療の均てん化, 医療の質向上, エビデンスの創出に向け, 近年大きな期待が寄せられている。中でも医療画像は, 日本は他国と比較して画像機器の設置台数, 撮影回数ともに多いこと, 正しい診断名が付与された質の高い画像データの収集が学会によって進められていることなどから期待が高い。そこで本紙では, 日本のビッグデータ研究をけん引してきた喜連川優氏に, 活用に向けた展望を聞いた。

——研究センター設立の目的を教えてください。

喜連川 ネットワークやクラウド, セキュリティ, AIなどの最先端情報技術の活用により, 医療分野の課題解決を推進することです。

現在, 大きく分けて2つの事業に取り組んでいます。1つは, 医療画像ビッグデータを集積するクラウド基盤の構築(図)。もう1つは, 収集した大量の医療画像を解析し医師の診断を助けるAI基盤の開発です。これらは, AMED「臨床研究等ICT基盤構築・人工知能実装研究事業」に採択された3学会(日本消化器内視鏡学会, 日本病理学会, 日本医学放射線学会)と連携して進めています。さらに, 日本眼科学会とも協議を始めています。



図 医療画像ビッグデータクラウド基盤の概念図(クリックで拡大)

大学や病院で蓄積されたデータを各学会が集め, 匿名化とフォーマット統一を行う。そのデータを各学会サーバーからNIIが構築する「医療画像ビッグデータクラウド基盤」へアップロードし, 保存。研究者がクラウド上でデータを解析できるようにする。

——AIによる医療画像診断の研究開発は, すでに海外でも進んでいます。日本で取り組む意義は何ですか。

喜連川 AIの性能を決めるのはデータです。日本には質の高い大量の画像データが保存されています。その優位性を活用すれば, 高い性能の画像認識AIを作ることをめざせます。また, 疾病の傾向は人種ごとに異なり, 生活環境も影響します。日本人の特性に合わせた早期の病変検出システム構築に向けて, 自国データを用いて研究開発することは大きな意義があると考えます。もちろん, 構築するフレームワークはグローバルを意識しています。

——厚労省「保健医療分野におけるAI活用推進懇談会」が17年6月に出した報告書では, AI開発を進めるべき重点領域に画像診断支援が挙げられていました。日本は診断系医療機器の開発能力が高く, 同機器の貿易収支も黒字という面で強みがあります。

喜連川 産業への転換の道筋がつけやすいという点はその通りかと思いますが, まだまだ研究開発中です。

日本一の超高速回線により膨大な医療画像データを収集

申込・変更

メールマガジン

紙媒体

広告のご掲載をお考えの方へ
医学書院 AD BOX

週刊 医学界新聞 アプリ

今日の治療指針
TODAY'S THERAPY 2018 web電子版付

治療薬マニュアル
2018 web電子版付

今日の診療 症例項目数 約100,000件
国内最大級の総合診療データベース

がんかん!
癌検診のためのwebマガジン by 医学書院

Gastropedia
消化器にかかわる最新医療情報のために



《標準理学療法学・作業療法学・言語聴覚障害学 別巻》
脳画像



デジタルマンモグラフィ品質管理マニュアル
(第2版)



腹部血管画像解剖アトラス



誰も教えてくれなかった胸部画像の見かた・考えかた



仮想気管支鏡作成マニュアル
迅速な診断とVAL-MAPのために

——研究センターが現在取り組んでいる2つの事業について教えてください。まず、クラウド基盤構築とはどのようなものでしょうか。

喜連川 医療ビッグデータを活用するためには、当然ながらまずデータを収集し、格納する基盤が必要です。それにはデータを安全に転送・保存できなければなりません。さらに、膨大な量のデータの転送や利活用の際の取り出しが円滑にできることも求められます。

本事業では、NIIが構築・運用する学術情報ネットワークSINET 5を活用します。100 Gbpsという超高速回線で日本全国、現在850以上の大学や研究機関が接続されています。この回線速度は家庭用光回線の約1000倍に相当します。

——医療ビッグデータを扱うにはそんなにも速い回線が必要なのですか。

喜連川 100 Gbpsは独り占めするものではなく、皆で使うものです。とりわけ最近では多くの学問でデータインテンシブになる傾向があります。ICTの歴史を振り返ると、文字列、音声、画像、動画と、よりリッチなメディアへと進化してきています。その背景には、通信だけでなく、プロセッサの高速化や大容量ストレージ技術の進展も同時にありました。ICT環境が成長すると、今度はその環境を利用した新しいアプリケーションや機器が生み出されます。

今日の医療機器は著しくIT化され、現時点でも膨大なデータを生み出しています。上述の歴史が示すように今後ますます大量のデータを生み出すようになることは必至です。100 Gbpsを超える通信技術も生み出されつつあります。例えば、8K画像の活用は医療から始まるだろうと政府でも議論されました。手術映像の8K伝送や遠隔医療をはじめとした広域ネットワークが果たす役割は大きそうです。

——人間一人が一生に生み出すヘルスケアデータは100万ギガバイトとの分析もあります。ビッグデータ研究というと収集したデータの解析に注目がちですが、環境整備も重要なですね。

喜連川 人の活動をデジタルデータで記録することをLife Logと言います。健康に一生を過ごせることは大切ですが、人生の目標は健康だけではありません。匿名化した人生データの解析は国連の提唱する持続可能な開発目標(SDGs)実現のための、人類の究極のテーマだと私は考えています。膨大なデータとなりますが、挑戦可能な時代はそれほど遠くないかもしれません。“IT屋”の夢は大きく広がります。

約12万症例の画像データ登録予定。次年度以降さらに拡大

——次に、医療画像データ収集の現状を教えてください。

喜連川 2017年11月にクラウド基盤が整い、運用が可能になりました。画像データ登録開始は今年度中をめざしています。

——医療画像データはどれくらいの規模で収集する予定なのでしょう。

喜連川 17年度中の画像データ登録症例数目標は、消化器内視鏡が1万症例、病理が11万症例と学会から聞いています。今後各学会がより多くの病院の協力を得て収集規模を拡大することを期待しています。

——AIの性能は教師付データ約5000の学習で許容できる性能に達し、1000万学習すれば人間の能力に匹敵するとの推計がありますね。

喜連川 AIの性能はデータの質と量に依存します。医療画像AIは昔から研究されていましたが、データ量の少なさが一番の課題でした。今回は学会によって従来では入手困難だった規模で収集できていますので、どこまで性能が上がるか楽しみです。

——今回の事業と近年のディープラーニング(深層学習)の進歩をどうご覧になっていますか。

喜連川 今回、画像に焦点が当てられ、学会との連携によりAI解析が事業として取り上げられたことは、大変時宜を得た、素晴らしいことです。近年急速に発達した深層学習技術により画像認識精度が飛躍的に改善されました。深層学習は多様な領域に適用されていますが、例えば言語処理などに比べますと画像領域において圧倒的に高い性能が達成されています。

——解析は万能でしょうか？ 課題もありそうですね。

喜連川 いろいろな疾患があり、それぞれに多様なパターンがあります。典型的な病変ばかりでなく、医師にも判別が難しい例がたくさんあります。今回の事業においては医師の方々と議論をずっとしていますが、なかなか一筋縄ではいきません。まだ始まったばかりで、多様な疾患への対応はこれからです。しかし、一人の医師が一生で数回診るかどうかというようなまれな症例も、日本全国のデータを集約できれば、その解析は大きく進展します。人間もそうですが、AIも勉強すればするほど概ね賢くなります。ロングテールの価値を引き出すことがビッグデータの強みです。

——画像以外の診断支援システムの研究開発に取り組む予定はありますか。

喜連川 医療にはいろいろなビッグデータがありますので、可能性はあります。実際、東大の私の研究室では内閣府のプロジェクトにおいて、日本全国のレセプトデータの解析システムを医療経済研究機構と連携して構築しています。ここでは約2000億レコードを保有し解析を進めています。今までは感じ取れなかった事実が多々見いだされつつあります。世界的に見てもこのようなビッグデータはオンリーワンだと思います。日本には貴重な情報がたくさんあります。

解析AI、数年以内に実用段階をめざす

——研究センターが取り組むもう1つの事業、解析AI構築の現状はいかがですか。

喜連川 データの収集と解析に着手し始めたところです。今後、学会と協力し、収集基盤の強化とともに、収集された画像データを基にした認識精度の向上を一步一步進めていきます。2～3年以内を目標に、医師の先生方に「これなら使いものになる」と言っていたレベルにしたいものです。

——厚労省の「AIの活用に向けた工程表」でも、画像データベース構築の目標が2020年とされています。

喜連川 冒頭述べたように、本事業はデータが全てです。現在、データ収集に関して各学会に多大な努力をしていただいています。どのようなデータを集めるか、ターゲットのロードマップデザインを含め、学会と総合的な検討を深めています。

——今回の事業で、NIIが参加した意義は何でしょうか。

喜連川 NIIはIT研究とITサービスの運用を同時に行う世界的にもまれな研究機関です。今の時代、IT“も”研究している研究所はたくさんありますが、その多くはもとの研究分野にITを活用する研究です。ITは今日40以上の分野に細分化されます。AIはその一つにすぎません。例えば次世代の“ITホスピタル”を考えると、画像処理技術だけでなく、カルテの文字列を読み解く言語処理技術、ユーザーへの情報提示をつかさどるヒューマンマシンインタフェース技術、AIを作るためのソフトウェア工学技術、患者の訴えを上手に聞く音声対話技術をはじめ、多様なIT基盤が必要です。日本で唯一IT「だけ」を総合的に研究する国研であるNIIだからこそ、ITの基礎から応用までほぼ全領域を網羅し、各領域の最先端の知見を融合できます。繰り返しますが、NIIが有するIT全体の総合的研究力が参加を求められた理由と考えております。

——研究は、NIIだけでなく多数の大学も参加して進めているそうですね。

喜連川 今回のプロジェクトの最大の特徴はオールジャパンで臨むことです。NIIはもちろん、医療画像診断AI開発の経験を持つ、東大、名大、九大などの先生方に参画していただいております。今後さらに拡大したいと思っています。NIIは大学共同利用機関です。オープンな研究環境をめざしてシステムを構築しています。学会からは将来限りなく多くの病院の参画を期待していると聞いていますので、ITサイドも広く日本の英知を招くことのできる場のデザインが必要だと感じています。多様な研究者が、時に異なる方法論を持ち寄り、よりよい方法を模索し、いろいろなことが試せる環境を作りたいです。

——AIの現状での性能や、AI完成後の将来的な活用の予定は？

喜連川 まだ発表前なので正確には言えませんが、認識率の性能値はかなりいい数字が出ています。活用については具体的な予定はまだ立っていませんが、支援をいただいているAMEDや各学会と今後協議させていただく予定です。

医療ビッグデータの面白さ

——最後に、情報学の専門家である喜連川先生から見た医療ビッグデータの面白さを教えてください。

喜連川 モビリティ、通信、金融、メディア、環境、災害、政策をはじめ、いろいろな分野のビッグデータを対象とした研究を進めてきましたが、やはり最も難しく、同時にやりがいのあるのは人間を対象としたデータの解析です。医療だけではなく、例えば教育もそうです。米国では学習の度合いを解析することでドロップアウトを大きく低減する成果があったと耳にします。東大の私の研究室では25ペタバイトもの地球環境のビッグデータ解析基盤を運用しています。多くのビッグデータがありますが、最終的には「人間を起点とした統合化」がなされるのではないかと夢想しています。人間の健康を対象とする医療データは最重要テーマの一つと言えるでしょう。

——喜連川先生の東大の研究グループでは、医療・介護・健診のヘルスケアビッグデータを統合した解析システムを構築していますね。エビデンスに基づいた地域包括ケア実現をめざし、近々三重県名張市と協力して、具体的な施策立案を行うと聞きます。

喜連川 医療データは多方面の活用が考えられます。名張市での試みは行政への利活用を指向しています。例えば、住民がどの医療機関に向いているかを解析すると、医療サービスが満足のいく状態か、住民に遠方の医療機関へ出向くことをどの程度強いているかを定量的に把握可能となります。また、医療と介護を連結した大量データの解析により、ターミナルケアの導入が医療費の削減に資することがわかりつつあります。後期高齢者の健康状態の推移

の解析とそれに伴う自治体の医療負担を検討するシステムなど、具体的な利用も進みつつあります。医療費の増大は大きな課題であり、今後ますますデータを用いたエビデンスベースの施策が重要になると考えています。

—技術面で難しさはありますか。

喜連川 全ての分野に共通する課題ですが、用語やオントロジーの整備が実は最も悩ましい問題です。

—Allに学習させる際は、表現揺れや記載ミスのクレンジングが必要と聞きます。コストを減らすため、標準化を進めることが重要でしょうか。

喜連川 もちろん標準化は重要です。しかし、デジュールにせよデファクトにせよ、一般にかなり時間がかかります。加えて、技術の進歩は昨今極めて速く、新しい表現がどんどん生まれます。そもそも技術が生まれる最初の時点で用語が整理されるわけがありません。つまり、これはエンドレスな戦いです。表記が不ぞろいだと「データが汚い」と言う方が多くいます。その際私は、「私は人生で、きれいなデータを見たことがない。きれいなデータがあるなら見せてほしい」と言います。不ぞろいの中で何ができるか、どうすればノイズを軽減できるかが腕の見せ所だと認識する必要があります。

—改正個人情報保護法や次世代医療基盤法により、データの取り扱い基準が明確になりました。研究が進めやすくなることを期待します。

喜連川 医療データのオプトアウト利用が可能になったことは画期的だと思います。これから細やかな制度調整がなされると思いますが、データを提供する患者に便益をフィードバックできるフレームワークとなることを祈念します。ITによる高精度画像診断支援がそれに貢献できればと考えます。

—ありがとうございました。

(了)

きつれがわ・まさる氏

1978年東大工学部電子工学科卒、83年同大学院工学系研究科情報工学専攻博士課程修了(工学博士)。2003年同大生産技術研究所戦略情報融合国際研究センター長、08年文科省科学官、10年東大地球観測データ統合連携研究機構長、13年より国立情報学研究所所長。09年ACM SIGMOD E.F. Codd Innovations Award, 12年 IEEE Fellow, ACM Fellow, 13年紫綬褒章, 15年21世紀発明賞, 16年レジオン・ドヌール勲章シュヴァリエなど受賞歴多数。内閣府最先端研究開発支援プログラム(FIRST)にて非順序実行方式による高速データベースエンジンを開発。文科省「情報爆発プロジェクト」や経産省「情報大航海」などの国家プロジェクトをけん引してきたビッグデータ活用研究の第一人者。ライフワークは「吉本工学(笑いを学問する)」。

