

医療ビッグデータ解析技術による 医療革新への貢献

Contribution of Medical Big Data Analysis Technology to Medical Innovations

● 勝田江朗 ● 一色隼人

あらまし

近年、ICTによるビッグデータの活用は、国内外の様々な分野において非常に大きな注目を集めている。これまで医療分野では個別化医療の実現に向け、電子カルテの普及から診療情報の集積、更にはゲノム情報の統合化といった革新が進められてきた。しかし、これらにより集積された大量の情報を、人手で処理するには限界があり、効果的・効率的に処理する医療ビッグデータを活用した技術が必要と考えられている。そこで富士通は、医療ビッグデータ解析技術を開発した。本技術は、迅速で高信頼かつ正確性を担保した解析の実行、そして利用者の解析スキルに依存しない適切なインターフェースなど、医療分野固有の要件に対応している。本技術は、独立行政法人国立病院機構 長崎川棚医療センター様と共同で実際の診療情報解析を実施し、医療革新に貢献できることを確認した。

本稿では、富士通が開発した医療ビッグデータ解析技術について述べる。

Abstract

Recently, utilization of big data by information and communications technology (ICT) has been attracting a great deal of attention in various fields in Japan and overseas. In the medical field, for the purpose of realizing personalized medicine, innovations ranging from diffusion of electronic medical records to accumulation of medical care information and further to integration of genome information have been promoted up to now. However, there is a limit to manually processing the large volume of information accumulated by these and technology utilizing medical big data that can be effectively and efficiently processed is considered necessary. Accordingly, Fujitsu has developed medical big data analysis technology. This technology meets the requirements peculiar to the medical field such as execution of analysis with promptness, high reliability, and accuracy ensured and an appropriate interface independent of the analysis skills of users. This technology has been confirmed to be capable of contributing to medical innovations through actual medical care information analysis conducted jointly with the National Hospital Organization's Nagasaki Kawatana Medical Center. This paper presents the medical big data analysis technology developed by Fujitsu.

ま え が き

近年、ICTによるビッグデータ活用に関し、世界規模でのパラダイムシフトが起きている。2012年以降、米国や欧州・アジアの主要国がビッグデータ活用に関し、具体的な政策や政府主導プロジェクトの実施といった動きを見せている。⁽¹⁾例えば、米国では大統領府の科学技術政策局が主導し、2012年からの5年間で総額2億ドル超の研究開発予算を六つの政府機関に割り当てるビッグデータ研究開発イニシアティブが実施されている。また韓国でも、ビッグデータマスタープランをまとめ、2012年からの4年間で5,000億ウォンの予算をビッグデータ活用の促進に充てた。このように、国によるビッグデータ利活用施策の実行例は枚挙に暇がない。

一方、日本は国家戦略や法制度を整備して、データ立国を目指している。日本再興戦略2016では、GDP 600兆円の実現に向けて「官民戦略プロジェクト10」を、人工知能 (AI)、IoT (Internet of Things)、そしてビッグデータを鍵として進めている。⁽²⁾総務省の白書においても、製造業、運輸業、サービス業などが活用先として挙げられており、様々な産業でビッグデータ活用が広がっている。⁽³⁾

医療分野においても、ビッグデータ活用の効果が期待されている。現在、少子高齢化や医療費の増大など、日本の社会保障は深刻な状況にある。政府はこれらの問題に対し、有効な医療を効率的に国民に広く提供する技術開発や研究開発を支援する仕組みとして、医療ビッグデータの活用を検討してきた。2017年には、医療ビッグデータ法と呼ばれる医療分野の研究開発に資するための匿名加工診療情報に関する法案が政府から提出された。この法案では、国が認定する機関を対象に、匿名加工情報を本人の同意を得ることなく大量に集積することが可能となる。これにより、基礎研究や医薬品開発だけでなく、個別化医療の実現を後押しすることで社会保障状況の改善が期待されている。

本稿では、個別化医療を軸に、近年のビッグデータ活用のトレンドについて述べる。次に、医療分野におけるビッグデータの活用に必要な技術について述べ、富士通が開発した医療ビッグデータ解

析技術の特徴と実証研究の結果を紹介する。

医療ビッグデータ活用のトレンド

電子カルテシステムの普及に伴い、蓄積された診療情報を活用することで、個別化医療への貢献が期待される。電子カルテシステムは、従来医師が診療の経過を記入していた紙のカルテを電子的なシステムに置き換え、電子情報として編集・管理し、記録する仕組みである。現在、400床以上の病院における電子カルテシステムの普及率は72.9%に達している。⁽⁴⁾これにより、患者を中心とした診療情報が大量に蓄積されており、これらを活用することで、患者個人に有効な医療を提供する個別化医療への貢献が期待されている。

また個別化医療では、診療情報に加えてゲノム情報の活用が重要である。例えば、同じ症状の風邪の患者に同じ薬を投与しても、効果が違う場合がある。これは、患者個々で異なる遺伝的特性や体質的な個性が生んだ結果であるが、診療情報だけでは捉えにくい事象である。そのため、蓄積された診療情報に加えて、ゲノム情報が重要となる。ゲノム情報は、生命の設計図であるDNA配列情報など、ヒトを構成する生命情報である。したがって、患者個々の差異をゲノム情報から把握できる。

近年、ゲノム情報解析に要する費用の低価格化に伴い、診療情報とゲノム情報の統合化が進められている。次世代シーケンサーと呼ばれるゲノム情報を超高速に解析する技術の登場により、ゲノム情報解析に要する費用の低価格化が実現した。⁽⁵⁾その結果、医療機関において診療情報にゲノム情報を統合化する事例が増えている。例えば、東北大学東北メディカル・メガバンク機構では、東日本大震災の被害から地域医療の復興と個別化医療の実現を目指し、15万人分の地域住民の診療情報とゲノム情報の統合化を進めている。⁽⁶⁾

このように、電子カルテの普及率の増加とゲノム情報の統合化によって、医療分野におけるデータ量は飛躍的に増加したが、大量情報を人手で処理するには限界がある。これらの大量情報を処理し、個別化医療を支援する手段の一つが、医療ビッグデータ解析である。

医療ビッグデータ解析に求められること

近年の電子カルテシステムのデータ活用において、大量に蓄積された診療情報は数万から数十万レコードに達し、解析する際には大きな負荷がかかる。また、ゲノム情報の統合化から、解析に用いる変数が数万から数千万にも及ぶ場合がある。変数が増えるほど不要な情報が混在する。例えば、疾患にかかっていない人でも検査で陽性を示す「偽陽性」や、本当は陽性なのに検査方法によって陰性と判定してしまう「偽陰性」といったものが解析結果に含まれることもある。更に、患者ごとに行う検査種別が異なるため、多くの検査項目値に欠損値が生じる。このように欠損値が含まれると、機械学習などの解析エンジンが動作しない。しかし、人命を扱う医療現場において、解析結果が出るまでのレスポンス遅延、偽陽性・偽陰性が含まれる解析結果、データによって動作しなくなる解析エンジンは許されない。したがって、医療ビッグデータ解析には、解析結果を医療現場に返す際、迅速性と高い正確性が必要となる。

同時に、医療分野ではビッグデータ解析により得られた知見に対して、高い信頼性の確保も必要となる。特に、個別化医療を実現する解析アルゴリズムは、人命に影響を及ぼす可能性があるため、知見には明確な根拠が要求される。しかし、機械学習で得られた知見の生産過程の多くはブラックボックスであり、根拠が不明確である。このことから、医療ビッグデータ解析により得られた知見は信頼性を担保する必要がある。

更に、利用者の解析スキルに依存せず、医学的知識に基づいて解析を実行できるインターフェースが必要となる。ビッグデータ解析には、機械学習などの複雑な数式や処理に基づいた手法が適用される。このため、得られた結果を解釈するには、複雑な数式や処理の理解を前提とした解析スキルが必要となる。しかし、多くの医療機関では医学的知識を有した人材が中心であり、解析スキルまで持った人材は少ない。したがって、利用者が医学的知識に基づいたビッグデータ解析を行い、結果を解釈できるインターフェースが必要となる。

開発した医療ビッグデータ解析技術の機能

富士通は個別化医療の実現に向けて、大量に集積された診療情報などを用いて対象患者の類似症例群を抽出し、病態の進行や治療効果を予測する「医療ビッグデータ解析技術」を開発した。

● 類似症例検索機能と変数選択機能

類似症例検索機能は、欠損値を含む項目を持つ大量のデータを、メモリ上に置いて高速に解析処理する。図-1に示すように、100万のレコードに対して、1CPUあたり約2秒で類似度を計算する。また、ゲノム情報を扱う際、前章で述べたように解析に用いる数千万もの変数の中から、医師が注目するような関連変数を抽出できる変数選択機能を実装している。この機能を用いて不要な変数を減らすことで解析精度を向上させる。これらの機能によって、多くの欠損値を含む巨大な医療ビッグデータに対して、迅速かつ正確に解析結果を医療現場に返すことが可能となる。

● 知見抽出機能

解析結果に対して、ある前提条件の結果を求め際に逆に観察結果から前提条件を推測する尤もらしさを表す尤度をを用いた有意確率(p値)を提示する知見抽出機能によって、解析結果の信頼性を高めている。ここで言う知見とは、類似患者情報を要約したものである。そのため本機能では、知見の根拠として類似患者情報の生データを提示し、信頼性を確保している。また、類似症例群を統計集団として扱うことにより、統計的検定が可能になる。ここから知見を抽出し、上述のように尤度を用いて有意確率(p値)を算出する。有意確率(p値)

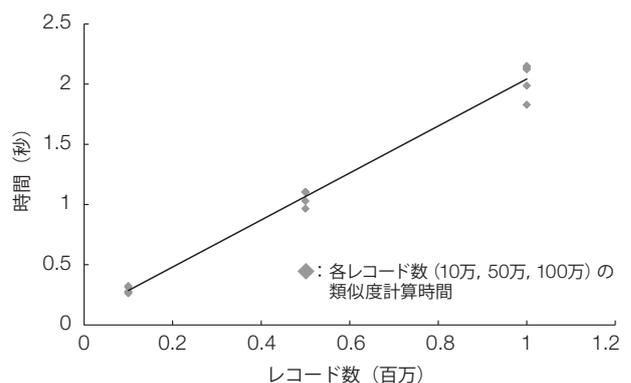


図-1 レコード数と類似症例検索の解析処理時間

の値が小さいほど知見の信頼性が高く、大きいほど知見の信頼性は低いことが示せる。

● 知見抽出機能の活用例

上述の知見抽出機能には、病態進行予測機能と治療効果予測の機能が実装されている。この二つの機能の活用例として、糖尿病とその合併症である糖尿病性腎症の発症したケースを説明する(図-2)。病態進行予測結果から、この類似症例は類似患者群の発症期間の中央値が2,200日となり、そのほかの患者群と比較すると500日短いという顕著な差が見られた。つまり、この対象患者は確かに不良であると判断される。一方、治療効果予測から、予後不良の類似症例中の患者に代表的な注射剤であるインスリン注射剤とGLP1注射剤を併用した場合には、ほかの治療行為と比較して糖尿病性腎症の発症を強く抑えることが分かる。つまり、この患者は糖尿病性腎症も合併症として発症する可能性が高いが、インスリン注射剤とGLP1注射剤を併用することで、糖尿病性腎症の発症を予防する可能性が高いことが予想される。

なおこの例では、利用者が医学的知識のみで簡

易的に操作し、容易に解析結果を解釈できるインターフェースが実装されていることも付言しておく。その特徴として、類似症例検索画面と病態進行予測画面を図-3と図-4に示す。いずれの画面でも、複雑な数式や処理の理解を含む解析スキルを必要としない、類似患者や解析結果を医学的知識から解釈可能な画面としている。例えば、図-3の類似症例検索画面は、患者間の検査項目値の違いから類似性を判断できるインターフェースとしている。また、図-4の病態進行予測画面は、医学ガイドラインや症例報告で扱われるログランク検定やKaplan-Meier曲線の統計的検定法を採用し、解析結果を表現している。こうした操作仕様や画面表示は、現場の医師から理解しやすく操作しやすいと高い評価を頂いた。

実証研究による評価

日本の糖尿病患者数は拡大傾向にあり、医療費の増加や専門医不足が問題となっている。富士通は2014年度に、地域間で医療情報などをセキュアに交換する地域医療連携ネットワークにおいて、

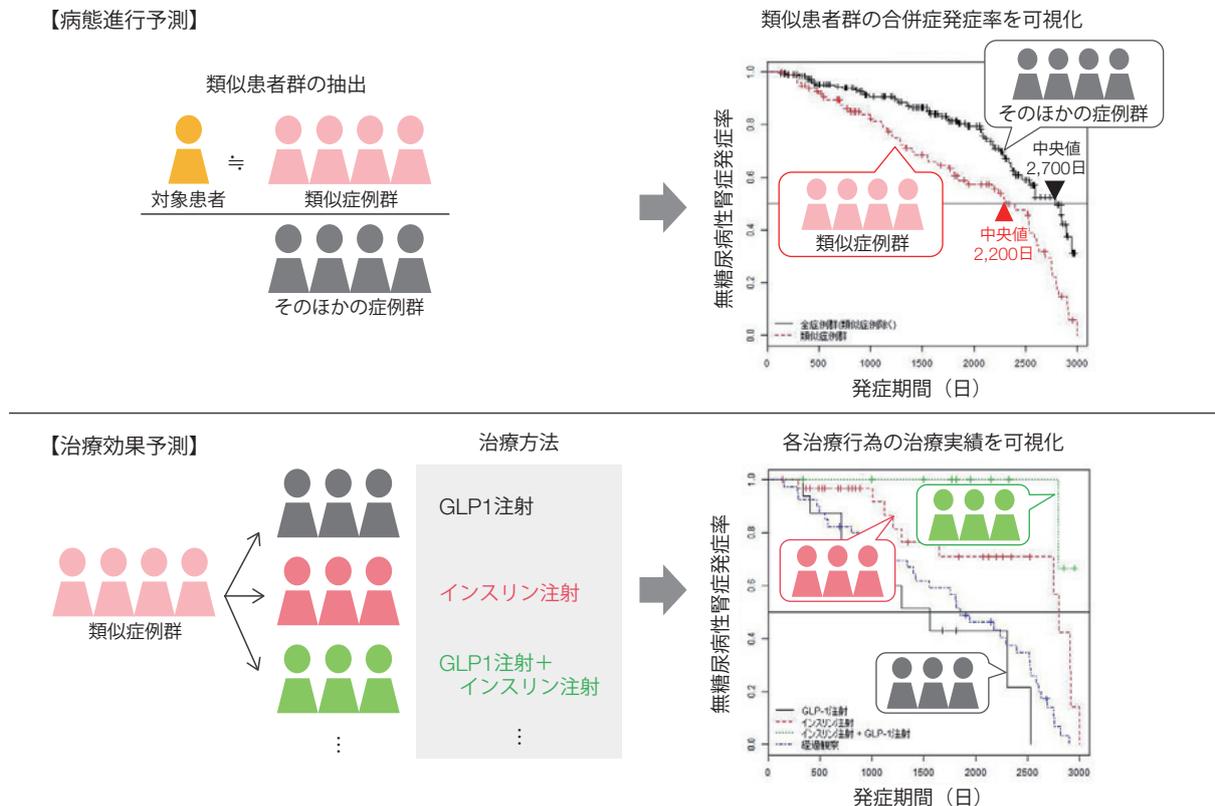


図-2 病態進行予測と治療効果予測の例



図-3 類似症例検索画面

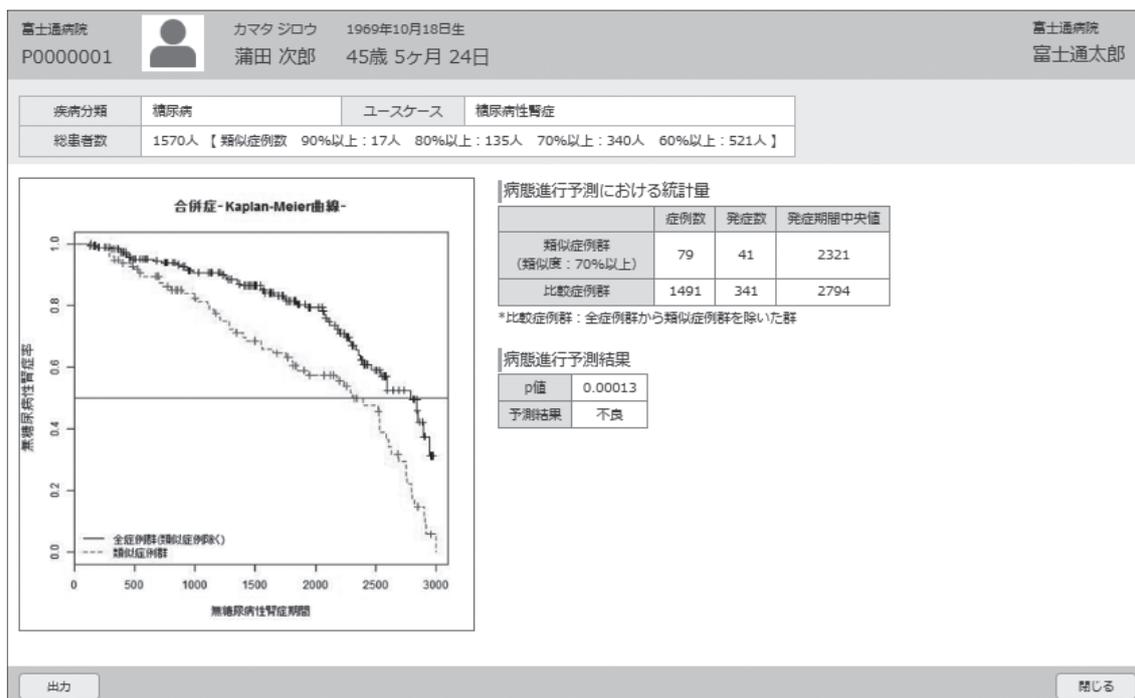


図-4 病態進行予測画面

糖尿病患者の治療内容などの情報を集積・管理するデータベース基盤となる疾病管理システムを開発した。2015年10月、独立行政法人国立病院機構長崎川棚医療センター様（以下、長崎川棚医療センター）と富士通は、疾病管理システムに集積された診療情報を活用し、糖尿病の重症化予防や効

果的治療に役立てるための研究を開始した。そして、医療ビッグデータ解析技術を適用した類似症例検索システムを開発した。2016年度より、長崎川棚医療センターと四つの医療機関の協力のもと、糖尿病患者から本研究の同意を得た上で、実証研究を行っている。⁽⁷⁾

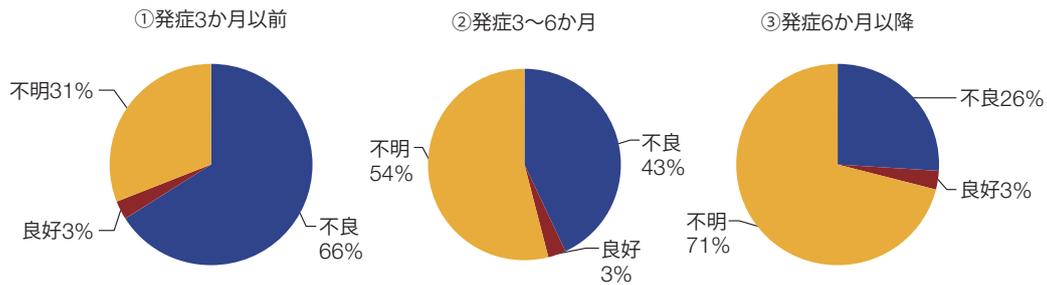


図-5 実証研究による病態進行予測結果

その結果、医療ビッグデータ解析技術は、実際の診療情報においても糖尿病合併症の発症を高精度に予測できた。2014年から現在までに、糖尿病性腎症を発症した30症例の患者を対象とし、同一患者の診療情報を以下の期間別にレコードを分けて管理し、病態進行予測を実施し、解析結果を評価した(図-5)。

- ①発症3か月以前
- ②発症3～6か月
- ③発症6か月以降

30症例は糖尿病性腎症を発症した症例であるため、解析結果が「不良」の場合は「正解」、「良好」の場合は「不正解」を意味する。そして、解析結果が「不明」の場合は「解析不能」、つまり類似症例数が少ないなどの理由により、その結果は統計的検定により信頼できないことを意味する。図-5に示すように対象の診療情報には欠損値が約50%も含まれていたにも関わらず、「不明」を除くと解析可能なほぼ全ての症例が「不良」、つまり「正解」として判定された。この結果から、医療ビッグデータ解析技術は、欠損値が多く含まれる診療情報においても高精度に病態進行を予測できることが確認できた。

今後の展望

現在、医療ビッグデータ解析技術は、血液検査結果などの数値で表された構造化情報のみを扱っている。しかし、診療情報には、カルテのテキスト情報やCTの画像情報などの非構造化情報が多く含まれており、疾患によっては非構造化情報がより重要な情報となる場合がある。例えば、癌では血液検査結果よりもCTの画像が重要である。今後、多様な疾患に対応するために、非構造化情報を解

析対象として扱うことが必要となる。

また、医療ビッグデータ解析技術にAIを活用していく。画像情報やテキスト情報などの非構造化情報は、AIが得意とする領域である。特に、Deep Learningの活用により、画像や音声、テキストなどに含まれる特徴量を機械自身が発見し、分類できるようになった。富士通には、Deep Learning技術を提供するサービス「Zinraiディープラーニング」がある⁽⁸⁾。医療ビッグデータ解析技術は、これらの技術やサービスを活用し、非構造化情報を取り扱うことによって、多様な疾患や利用シーンに対応していく。

む す び

本稿では、医療ビッグデータ解析技術による医療革新への貢献について、個別化医療の実現を軸に近年のビッグデータ活用のトレンド、求められる技術や考え方と富士通の取り組み、および実証研究の評価を述べた。医療分野において今後ますます巨大な情報が集積されていく中、ヒトを中心に豊かで快適な生活をサポートする医療・健康ソリューションやサービスの実現に向けて、富士通の医療ビッグデータ解析技術が一つの有効なアプローチになると考えている。

最後に、実証研究の評価にご協力を頂いた国立病院機構 長崎川棚医療センター様、国立病院機構 長崎医療センター様、中村医院様、伊崎脳神経外科・内科様、ちくばクリニック様に深く感謝いたします。

参考文献

- (1) 総務省:平成26年度版情報通信白書, 第3章, 第1節.
<http://www.soumu.go.jp/johotsusintokei/whitepaper/>

ja/h26/pdf/

- (2) 内閣府：日本再興戦略2016.
<http://www.kantei.go.jp/jp/singi/keizaisaisei/kettei.html>
- (3) 総務省：平成27年度版情報通信白書，第5章，第4節.
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/pdf/>
- (4) 保健医療福祉情報システム工業会：JAHIS オーダ
エントリ・電子カルテ導入調査報告2016年調査.
https://www.jahis.jp/action/id=57?contents_type=23
- (5) Kris Wetterstrand：DNA Sequencing Costs.
<https://www.genome.gov/sequencingcostsdata/>
- (6) 長神風二ほか：日本人1,070人の高精度全ゲノムデー
タの統合的な解析に成功. 東北メディカル・メガバン
ク機構.
<http://www.megabank.tohoku.ac.jp/news/11873>
- (7) 富士通：糖尿病の重症化予防を支援するシステムの
実証研究を開始.
<http://pr.fujitsu.com/jp/news/2016/07/8.html>
- (8) 富士通：世界最速クラスのディープラーニング基盤
システムを販売開始.
<http://pr.fujitsu.com/jp/news/2017/05/16-3.html>

著者紹介



勝田江朗 (かつだ ただあき)

パブリックサービスビジネスグループ
第二ヘルスケアソリューション事業本部
第五ソリューション事業部
医療ビッグデータ解析ソリューション
開発に従事。



一色隼人 (いっしき はやと)

パブリックサービスビジネスグループ
第二ヘルスケアソリューション事業本部
第五ソリューション事業部
医療ビッグデータ解析ソリューション
開発に従事。