

特集 : CKD Big Data

医療における人工知能の活用と将来展望

Utilization and future perspective of artificial intelligence in healthcare

大江和彦

Kazuhiko OHE

最初の医療人工知能ブーム

医療の分野で人工知能(AI)が注目を浴びたのは1975年である。当時スタンフォード大学の大学院生であったShortliffe EHは、感染症患者の症状や検査所見を入力して最適な抗菌薬の推薦を出力とするコンピュータプログラムMYCINを発表した¹⁾。MYCINでは、数百件の「もし…ならば…の確からしきで…の可能性はある」といった条件部と結論部からなるルール型の知識が計算機に知識ベースとして記述されている。対話形式でシステムに入力される患者の症状や所見をこの条件部と順に照合することで、合致するルールの結論部を推論結果として導出し、その結果をさらに他のルールの条件部と照合していくという形で論理推論していく。その結果として可能性の高い起炎菌を推論し、それに対応する抗菌薬グループを推薦するという形で動作した。知識を記述する部分と、その知識を使って推論するプログラム部分とが分離されていることが特長で、これにより、知識の記述を増やしていくことで動作を改良していくことができた。こうしたシステムは、当時、ルールベース駆動型のエキスパートシステム(専門家を支援するシステム)と呼ばれ、第二次人工知能ブームの先鞭となるとともに、関節リウマチ、感染症、緑内障、急性腎炎、内科の初期診断領域など、医療分野で多くのエキスパートシステムが国内外で開発され、個々の領域では一般医師の能力を上回るような性能を示したものも多い。しかし、より性能を向上させようとして知識の記述を増やすにつれ、他のルールとの整合性を維持するなどの知識メンテナンスが困難になるとともに、ルール相互間の矛盾を解消する論理的方法が確立できなかつた。また、記述した知識の限

られた領域でしか性能が発揮できないことや、そもそも医療の場での患者情報や検査情報がデジタル化されておらず、情報を対話的に逐次入力して使用していくことも現実的でなかつたことなどから、実用にはほど遠いことがわかり、全体として当時の人工知能ブーム自体が下火になっていった。しかし、こうした知識ベース駆動型のAIシステムは、論理的に推論する過程や結論が導出される過程をシステムが説明できるため、医療応用に向いている面があつた。

医療情報のデジタル化の進展

情報技術、ネットワーク技術の進歩により医療の場でも21世紀に入り診療情報のデジタル化が進み、電子カルテシステムの導入が加速した。その結果、処方や検体検査、画像検査データがデータベースに蓄積されるようになった。電子カルテ導入率は、米国ではここ数年で80%を超えるようになっているのに対して、日本ではまだ35%程度であり、全医療データが電子的に利用可能であるとはいえないが、レセプトの電子化はほぼ100%になっている。

しかし電子カルテシステムでは、重要な診断情報である各種の画像検査や病理検査の結果レポートが自由記載形式で作成されることが多く、現在のコンピュータ処理では自由記載文章から正確な所見・診断情報を取得することは困難である。また、病名情報は保険請求上必要な病名として登録されるため、登録されている多くの病名のうちどれが診断確定病名であるかを判定できない状況にある。そのため、デジタル化された診療情報のうち、統計処理システムや後述べる人工知能応用システムで使用できるのは、処方や注射の実施情報による医薬品投与情報、数値で得られる検体検査結果、画像や波形データに限られている状況で

ある。

医療情報のデジタル化における標準化の問題

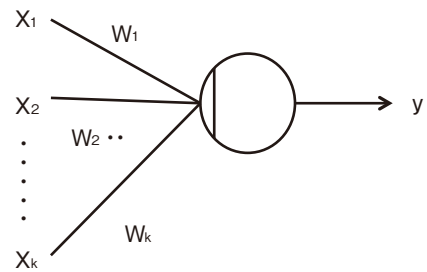
医薬品投与情報や検体検査結果のデータを多施設から収集して1つの巨大なデータベースとして活用するには、個々の医薬品、検体検査項目を識別するための識別コードが共通であることが求められる。それぞれ厚生労働省標準コードが規定されているが、ほぼすべての医療機関の情報システムの運用においてこのコードは使用されておらず各医療機関が固有に割り当てたコードを使用しているのが現状である。そのため、各医療機関からデータを集める際に、個々のコードを共通の厚生労働省標準コードに変換する必要があり、そのためには、各医療機関が固有コードから標準コードへの変換対応表を作成しメンテナンスする必要がある。しかしこれを各医療機関が作業するには、臨床検査コードや医薬品コードに関してある程度の専門知識が必要で、検査部や薬剤部のスタッフだけでできるとは限らず、それが障害となっており、学会団体などが支援する枠組みの構築が求められている。

現在の人工知能技術—機械学習

ここ数年話題をさらっている人工知能であるが、その技術の根幹を成すのは機械学習、とりわけ多層ニューラルネットワークを用いた深層学習である。非常に乱暴に言ってしまうと、機械学習とは、本来は複数のグループに分類できるはずの一群のデータセットとそれぞれの正しい分類結果データを与えられて、そのデータだけから計算により正しく複数グループに分類できるような計算モデル(つまり複雑な数式など)を作りあげ(学習)、その計算モデルを用いて新たなデータに対しても正しく分類ができるようにする手法である。学習のためのデータには正しい分類結果が含まれている必要がある。

例えば、1型糖尿病(T1DM)と2型糖尿病(T2DM)の2グループからなる計5,000人の患者の発症から1年間の臨床検査データ10項目の検査結果と正しい分類情報(T1DMかT2DMか)を学習のためのデータセットとする。「あらかじめ想定する多変数からなる数式モデル」に含まれる多数のパラメータについてデータを入力するたびに微調整していき、5,000人のデータのどれについてもできるかぎりT1DMかT2DMかを正しく出力できるようにしていく。これにより正解と出力との間の差の総合的な和(例えば差の自乗和)

W_i は各入力 の重み



ノード(神経細胞に相当)への総入力 $S = W_1 \cdot X_1 + W_2 \cdot X_2 + \dots + W_k \cdot X_k$ とする。出力 y は総入力 S が一定値(例えばゼロ)以上になれば入力値 S をそのまま出力するようなノード

図1 ニューラルネットの1単位ノード

が最小になるまでパラメータが調整できれば終了とする。こうして実際の多数のデータで最適の結果が得られるようにパラメータを調整していく過程を学習(Learning)と読んでおり、それを計算機により行っているのが機械学習(Machine Learning)と呼んでいる。このようにして得られたパラメータによる数式モデルは、与えられた5,000人のデータについては正しく2群を分類できるように調整されたものと言える。この数式モデルに学習には用いなかった新たな1,000人の臨床検査データだけを1人ずつ入力した場合に、1,000人すべてについて正しくそれぞれの患者をT1DMかT2DMか分類できる出力が得られるかどうかで、その性能が決まる。学習に用いたデータ(学習データセット)で例えば95%の分類が正しくできるように学習した数式モデルに対して、新たなデータ(テストデータセット)については85%の性能しか出ないというようにある程度性能が落ちることが多い。このように学習データセットでは良い性能が出るのに、テストデータセットでは性能が出ない減少を過学習と呼び、いかに過学習を解消するかは機械学習での重要な課題である。

機械学習では、「あらかじめ想定する多変数から成る数式モデル」に何を使うかで性能が変わるため、少しずつ異なるモデルが提唱され成果を上げていったが、劇的に進歩するきっかけとなったのが、このモデルに多層ニューラルネットワークを採用した機械学習、すなわち深層学習(Deep Learning)の出現である。多層ニューラルネットワークは図1のように、複数の入力にそれぞれ異なる重みを掛けて入力し、この総和が一定値以上になるとそれが出力になるような活性化関数を通して出力する1単位(ニューロン、ノード)を

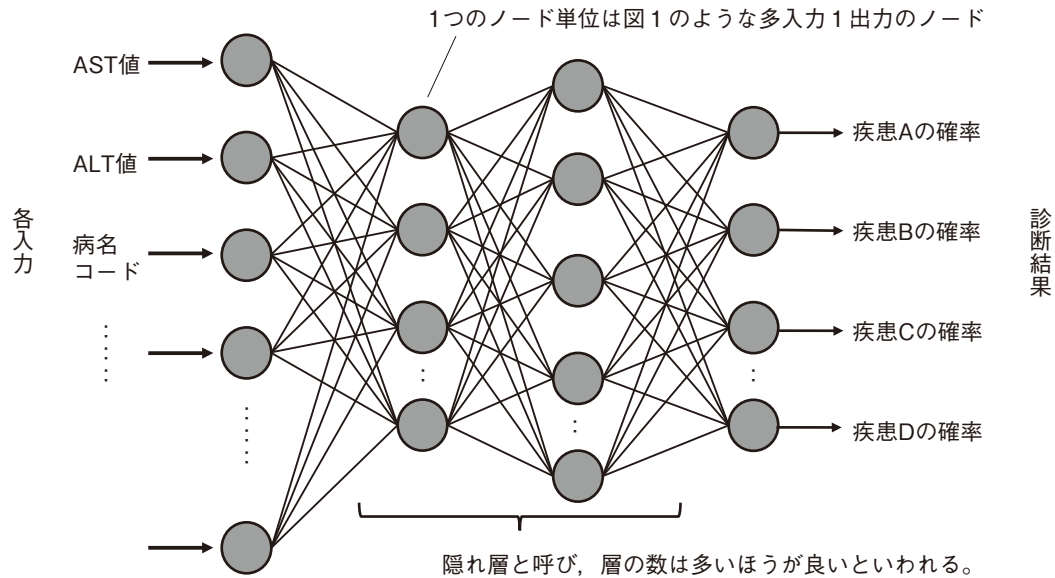


図2 多層ニューラルネット

作成し、それを図2のように多層ネットワークとして構成するものである。

多層ニューラルネットに関しては、近年急速にさまざまな工夫がなされるようになり、畳み込みニューラルネットワーク(CNN)は画像分類や画像認識などに、再帰型ネットワーク(RNN)は画像区画識別や自然言語処理(文章情報の分析や自動翻訳など)に優れた性能を示すようになってきた。

医療におけるAI技術の活用と課題

医療におけるAI技術の活用は、高速な医学知識検索、医療画像診断、診療経過や推移予測、治療機器の自動制御など多岐にわたる。高速な医学知識検索では、膨大な医学文献から遺伝子変異、表現型、医薬品候補との関係知識を自然言語処理と人手を組み合わせることで知識データベースとして構築し、診断困難で治療方針を立て難い腫瘍性疾患の患者の遺伝子解析結果の入力からこのデータベースを検索し、治療効果の可能性のある医薬品候補を提示するシステムがある。また医療画像診断では、前述のCNNやRNNを活用したものが多く、例えば胸部X線写真の所見分類、眼底写真画像からの網膜症の所見分類、病理画像写真や皮膚病変写真^{2,3)}の良性・悪性の判別など、世界中で多くの研究開発がされており、診断補助機器として実用化されるものが出てくる日も近いと思われる。

診療経過や推移予測においては、数カ月以上の時系列の臨床情報と重症化の有無を学習データとして機械学習を行うことで予測モデルを構築し、テストデータに対してもある程度高い精度で重症化を予測するといった研究成果を出している例もある⁴⁾。筆者らの研究室でも、糖尿病の受診を中断してしまう患者を外来で予測するモデルを機械学習により構築し、正確度75%を得ている⁵⁾。

こうした医療におけるAI技術の活用においては、①知識記述における定式化と標準化、②データの大規模収集、③個人情報保護や匿名化、などの課題の解決が必要になっている。例えば高速な医学知識検索では、人が読むことを前提に自然言語文で書かれている論文に含まれている医学的知見や知識を、計算機が処理可能なデータベース形式に変換するとともに、そのデータベースに記述されるさまざまな医学概念を、同一概念は同一の語で表現するとか、語同士の意味関係(意味の上限関係など)を表現するなどした標準化用語辞書とオントロジーの整備が求められる。自然言語文を計算機で処理する技術は飛躍的に改善されているものの、専門用語の標準化辞書やオントロジーの整備は不完全であり、全自動で質の高い知識データベースを構築することはできておらず、まだまだ膨大な人手処理が必要である。

現在の深層学習手法では、正解分類ごとに数千～数万症例のデータが学習には必要とされている。囲碁ソフトウェアなどでは膨大な過去の対戦データが利用できるうえに、

ソフトウェア同士が短時間で対局することができ、その勝敗データを学習データとして利用できるため、短時間で膨大な対局データを学習することができる。しかし、医療における診断可能性のある分類は数千以上に及ぶものが多く、すべての分類ごとに症例データを学習に必要な件数集めることは困難で、特に稀な疾患の症例データを必要な数集めることは不可能である。また、医療画像や病理画像に対して詳細な所見データは人間が読んでわかる文章形式で書かれていることが多く、コンピュータが正解分類データとして利用できる形でコーディングされているデータは少ないため、現在の電子カルテデータをそのまま深層学習の学習データとして利用することは難しい。

また、医療における診断や意思決定は単純に結果だけが得られればよいのかという議論がある。機械学習、深層学習では、なぜその分類に到達したのかを論理的に説明できず、ブラックボックスのように計算結果として解が出力される。そのため、結果を医師や患者が納得して受け入れることが難しい場合がある。

AI技術、とりわけ機械学習を利用した医療機器や診断補助システムでは、学習に使用するデータにより性能が大きく影響を受けるため、学習データセットを企業が囲い込もうとする動きもみられる。また、学習データセットの著作権や機械学習に使用したアルゴリズムの知財権の帰属など新しい課題も見え始めているが、膨大なデータを医療者が処理するのが不可能になりつつあるこれからの医療において、AI技術は必要不可欠なものとなることは確実である。現在のAIブームを牽引している機械学習では、これまで医療者が自分でも必ずしもすべてを論理的にその思考過程を説明できないような種類の意思決定プロセスや診断プロセスにおいて、データに基づいて計算モデルが構築され、

結果が出力される。画像診断や波形診断、経験データにもとづく直感的なパターン判断や機器操作などに依存している医療プロセスのうち、大量の学習データが入手できるような領域では、補助システムや支援システムとしてこのような機械学習システムが大きく活躍していくことになるであろう。一方で、論理的な推論の積み重ねで意思決定をしていく必要のある領域では、医学知識やルールをデータベース化して推論していくような1970～80年代の人工知能システムの発展形も不可欠である。したがって、これからの医療における人工知能の活用は、この2種類の技術がハイブリッドする形で発展していくことになると考えられる。

利益相反自己申告：申告すべきものなし

文 献

1. Shortliffe EH, Davis R, Axline SG, Buchanan BG, Green CC, Cohen SN. Computer-based consultations in clinical therapeutics : explanation and rule acquisition capabilities of the MYCIN system. *Comput Biomed Res* 1975 ; 8(4) : 303-320.
2. Litjens G, et al. A survey on deep learning in medical image analysis. arXiv : 1702.05747v2 [cs.CV] <https://arxiv.org/pdf/1702.05747.pdf> (accessed July30, 2017)
3. Esteva A, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017 ; 542(7639) : 115-118.
4. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient : An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. *Scientific Rep* 2016 ; 6 : 26094.
5. Kurasawa H, Hayashi K, Fujino A, Takasugi K, Haga T, Waki K, Noguchi T, Ohe K. Machine-Learning-Based Prediction of a Missed Scheduled Clinical Appointment by Patients with Diabetes. *J Diabetes Sci Technol* May 2016 ; 10(3) : 730-736.