

大会企画

大会企画1

医療とAI ～ どこまで来たか、どこに行くのか

2021年11月19日(金) 09:10 ～ 10:40 A会場 (4号館1階白鳥ホール)

[2-A-1-02] 臨床が求める AIを探るための試み

An Attempt at Exploration of Artificial Intelligence in clinical practice

*山本 康仁¹ (1. 東京都立広尾病院 経営企画室)

*Yasuhito Yamamoto¹ (1. Tokyo metropolitan Hiroo Hospital Management Planning)

キーワード : Digital Transformation, Natural Language Processing, Artificial Intelligence

第三次 AIブームの流れに乗り、機械学習が注目されるとともに医療への応用が現実のものとなった。そして医師の仮説的演繹法に似た方法で EHRを照会し、これまでの統計的手法では見つけられなかった関連性を発見し、小児疾患の診断において経験豊富な小児科医に匹敵する AIも出た。DeepSOFAと呼ばれる試みは ICUで発生する大量のバイタルデータを学習し、どの患者に救命処置が必要かをいち早く示すことで医師の治療の時間を稼ぐ能力があるという。しかし、画像認識領域の AIは、結果精度に関する治験を行うことで診断根拠とし、追加学習は考慮されない。医師の代用判断するものは、専門家不足を補うものであり必ずしも求められるものとは限らないだろう。AIの判断がブラックボックスで、根拠を説明できないのであれば、そもそも利用できるかは不透明だ。

AIによる回答と説明

追加学習を容認し環境変化に対応可能で、判断をブラックボックス化せず、思考根拠を示すことができ、かつ医師が求める AIとは何か。その答えを求めるためには、実際に段階を踏んで臨床現場での試行が必要と考えた。そこで当院の取り組みを示す。検査結果や機器からの測定結果から直接臨床判断を行うのではなく、モジュールに分割した。モジュール間の連携は要約された情報で行ない、元の情報へのアクセスを制限した。そのため患者がもつコンテキストを考慮する必要があった。この時の動作過程をコンテキストと合わせて保存し、説明に利用した。

例えば生体モニターの出力は人と同じアルゴリズムでフィルタリングを行ない、統計学的機械学習によるノイズ除去を経て採用した。同時に患者基本情報やカルテ記載の文脈解析でコンテキストを用意、ソースを合わせて保存した。大量データを俯瞰して判断する時、最終判断を行うヒトが納得できるように途中結果を説明し、機械の誤りをヒトが判断できる AIが求められると考える。

臨床が求める AI を探るための試み

山本康仁*1

*1 東京都立広尾病院 経営企画室

An Attempt at Exploration of Artificial Intelligence in clinical practice

Yasuhito Yamamoto*1

*1 Tokyo metropolitan Hiroo hospital, management, and planning

If AI decisions cannot explain their reasons, clinicians will not be able to use them. To clarify what kind of AI doctors are looking for, we thought it would be necessary to conduct trials in actual clinical settings. Therefore, we divided the measurement results into modules without directly using them for clinical decisions by the machine. The linkage between modules was done with summarized information, and access to the original information was restricted. Under this restriction, the context of the patient had to be taken into account. Here we present our hospital's attempt to use this context and the original information to explain the machine's decisions.

Keywords: Digital Transformation, Natural Language Processing, Artificial Intelligence

1. はじめに

医療現場での AI は何のためにどのように使うのがベストプラクティスなのか。将棋や囲碁の世界では、人が思いもつかない指し手を提示する AI が出現、AI による肺移植の順位づけが専門医と違うという事態も出てきた。今後、専門職域集団である医療現場で AI をどのように利用していくのが良いのか、先進事例を参考に医師の立場からの方向性について考えてい。

2. 先進的な AI の取り組み

近年深層学習を取り入れた AI が画像認識の領域で飛躍的な発展を遂げたことを契機に再び見直され、医療領域でも数々の試みが積極的に行われるに至った。単純な画像認識にとどまらず、医師が使用する仮説的演繹法に似た方法で電子カルテを照会し、これまでの統計的手法では見つけられなかった関連性を発見することで、一般的な小児疾患の診断において経験豊富な小児科医に匹敵するものも現れている。¹⁾ 専門医の不足する領域では判断支援として活躍するかもしれないが、医師として自らを置き換えるような AI を求めているかという懐疑的ならざるを得ない。例えば深層学習の音声認識領域を拡張し、心雑音を聞き分ける AI も開発されているが、²⁾ 専門的診断能力の一部を置き換えるものであり、安価な聴診器のみで心エコーを用いることなく、ある程度の心疾患を見分ける能力を有するという。これらの技術は、AI で強化された心電計の自動診断装置と同様、検診などのスクリーニングを目的としており、臨床現場での活用を主とはしていない。実用化にもっとも近い各種画像認識も教師画像の収集と Annotation に主眼がおかれ、結果に対して治験を行うことで根拠としており、追加学習は考慮されない。病理標本の場合、施設が違い標本の「癖」が画像分析に影響与える可能性が指摘されているが³⁾、これらの癖の追加学習ができるかどうかは不透明である。

では医師が臨床現場で求める AI 像とはどんなものなのであろうか。その答えに近い先進事例も報告されている。その一つの DeepSOFA は、集中治療室のどの患者に救命処置が必要かを示すことで、致命的な状態を防ぎ、時間を稼ぐことができる道具であると Bihorac らは述べている。⁴⁾ AI は分散す

る大量のデータを分析し、従来の SOFA スコアよりも正確に患者の重症度を評価した。このように、24時間連続して高速に大量のデータを分析し、人間よりも早く事態を把握し医師に治療の時間を与えるタイプの AI は医師と協業し負担を軽減し、患者に利益をもたらすかもしれない

3. 求められる AI 像

リアルタイムで動作し、大量データを解析して人よりも早く異常に気づき伝える AI にはどのような要件が求められるだろうか。さらに医師と協業できるという絶対条件を加えるなら、AI の思考経緯を人が理解できる形として示される、あるいは AI が説明できる能力を持つ必要があるだろう。前述の先進事例は古典的な SOFA スコアと同じ領域のデータを用いるが、人に合わせて統計学的に導いた簡略化したデータセットと計算式を使うのではなく、多くのデータをもとに深層学習を用いたパターン認識で重症度を示している。要所に深層学習を使い、できるだけブラックボックスを減らし、従前の確立した手法に似せて説明を果たしたものとする。さらに AI がその思考ロジックを説明するならば、どのような手法があるのかを考えた。

3.1 AI による回答と説明

機械による判断の理由がブラックボックス化させないための手法として当院がとった方法を説明する。検査結果や機器からのリアルタイムの測定結果を演算するにあたり、いきなりそこから臨床判断を行うのではなく、モジュールごとに一度情報を要約した。要約するため、情報元と比較すると情報量も粒度も低下するが、その時に得られた周辺情報を「コンテキスト」として加えて、これを補う手法をとった。モジュール間の連携は、要約された情報をもとに行われ、次のユニットは元の情報にアクセスすることなく、コンテキストと合わせて適切な結果ができるように設計した。場合によっては元データを参照した方が簡単で早いですが、それを制限することで、モジュールの規模は冗長となったが、モジュール総体としては環境変化に強い例外に耐える仕組みとなった。モジュールの動作ログは説明のために保存し、その時点での演算に利用したコンテキストも同時に保存し分析できるよう保存した。

例えば、外部システムのレポートを含むすべてのカルテ記載はリアルタイムで日本語自然文解析し、その時点で存在する病態としてコード化した。同様に、手術オーダーや処置オーダーに含まれる疾病情報は、たとえば「手術を必要とする

骨折」という形で重み付けを含んで病態コード化、検査結果も数値情報をあえてその時の標準値と照らし合わせ、グレード別に「AST 上昇」という丸め処理を行いコード化している。これらのリアルタイムで刻々と変化する病態コードは、ソースの性格に合わせて有効期限を持った病態コード群として、その後の各処理のコンテキストとして活用した。

データクレンジング

看護師が測定結果を手動で書き込むバイタル情報は人がバリデーションしている情報であることからそのまま採用できるが、例えばモニターからそのまま送られる大量の数値情報は、人が数値を読む時と同じアルゴリズムでフィルタリングを行ない、その後統計学的機械学習によるノイズ除去を経て採用した。例えば、SpO2 値については、同時に測定する心電計による心拍数と、SpO2 による脈拍数を比較し SpO2 の測定が適切に行われているかを評価したうえでデータを採取し、それをもとに機械学習による異常値フィルタ処理を行っている。同様の仕組みとして観血的動脈血圧も脈波数と心電計の心拍数を比較し、一致するデータを採取した。機械学習に関しては、演算量が少なくて済む方法を選んだ。Box-Cox 変換で数値を圧縮、k 近傍法を用いてベクトルを求めた。k 近傍法で用いる連続パラメータ数 n は実データを用いて最適と思われるものをあらかじめ人が設定し、ベクトル集合体をホテリング理論でノイズ検知を行い、除外した。(図1、2、3)

$$y^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & (\lambda \neq 0) \\ \ln(x) & (\lambda = 0) \end{cases}$$

図1 Box-Cox 変換

後に処理をしやすくするため、xを圧縮する(伸張する)

$$d = \left[\sum_{k=1}^n (x_k^i - x_k^j)^2 \right]^{\frac{1}{2}}$$

図2 K 近傍法

n次元空間上の2つの位置ベクトルxi, xjの距離d、1分毎の数値であるため連続するn分のデータをn次元としてベクトル距離を各々求めていく。

$$a(x') = \frac{(x' - m)^2}{s^2}$$

図3 ホテリング理論

異常度は、サンプルから平均を引いた二乗を標準偏差の二乗で割ったもので示され、p=0.01としたとき、カイ二乗分布から9.21を超えるものとして識別した。

3.2 スコアリングを自動的に表示する試み

APACHE II score を自動的に算出し、さらに院内予測死亡率を複数のモジュールの出力を複合してもとめ、実績と比較するとともに、臨床医に公開している。こうした仕組みは、厳密には AI と呼べるレベルに達していないが、個々の仕組みが人の理解可能な形で中間出力を持ち、多くの業務に活用可能で、機械の考えがどこで間違ったかを理解でき、その結果の利用に躊躇がない状況を築いた。(図4)

図4 APACHE II 画面

APACHE II List Acute Physiologic And Chronic Health Evaluation ver.1.3

対象	Patient Name	最高体温	最低体温	最高血圧	最低血圧	最高心拍数	最低心拍数	最高呼吸数	最低呼吸数		
男	38.2	38.7	0	80	75	2	111	57	2	30	7
男	38.7	38.7	2	115	68	3	173	51	4	42	1
男	38.8	38.8	4	178	76	2	126	58	3	41	12
男	38.9	38.9	2	85	60	2	101	54	3	38	7
男	37.3	38.2	4	145	48	3	92	38	4	41	3
男	37.2	37.3	4	168	27	2	105	68	4	65	7
男	38.08	38.8	3	131	67	3	152	63	3	49	7
男	37.1	36.1	0	94	74	3	158	65	2	33	10

計算に使用したデータを丁寧に示すことで実際の値との乖離を示した。

3.3 略語処理を強化した HiPER NLP

カルテに記載される記録は、患者サマリーや症例報告と異なり、医師自身や診療科グループであれば理解できる略語が使用されていることがあり、他の診療科や医師以外の職種にとって難解である。略語は一般にアルファベット数文字で構成されているため、同じ略語が全く異なった複数の意味を持つ。例えば、AN は、神経性食思不振症、動脈瘤、無腐性骨壊死、アナフィラキシの意味をもつため、どの意味で使われているかを類推する必要がある。また文章自体もすべてが記載されているわけではなく、助詞や助動詞が省略され、箇条書きで書かれていることが多い。こうした文章を従来の形態素解析で分かち書きを行っても、省略されている語句が多く文脈を正確に把握することは困難である。HiPER NLP は分かち書きに関してシンプルに文字の種類で切断したが、TMN 分類などの記号が過度に分割されるため、これらに特化した分割アルゴリズムを作成した。箇条書きに対しては、全体の構造を類推し、省略されている語句を補完するなどの処理を加えている。時間軸の認識、相対時制表現も記載日から絶対的な時間軸を補完した。同様に、記載の主語に関しては本人以外を除外、否定や仮定も一部の特殊なルールをのぞき大胆に除外した。

前述のようにカルテ内の略語は NLP にとって問題とされるが、HiPER NLP は逆に略語は重要な記載が、高頻度かつ繰り返し使用されるからこそ略語化したと考え、この正しい解釈に重点を置いた。文章中に含まれる略語以外の記載に関して、略語を正しく処理するための文脈ベクトル(表1)を採取するために解釈した。記載した医師の診療科属性を考慮した上で、表1に示した文脈ベクトルを取得し、複数の意味を持つ略語の解釈は2パス目で文脈ベクトルを考慮して解釈している。文脈ベクトル以外にも、後ろに続く文字種類や、関連性の高い語句を検索するなどして略語の特定の精度を向上させた。形態素解析を行うと一般的に細かく単語が分割され、複合名詞などの複合語への対応が課題となるが、文字種類を基本として分割する HiPER NLP は最長一致で辞書を参照した。

こうして特に略語を正確に解釈することで病名、病態を抽出することに加え、検体検査の数値結果や、手術治療に関連する病態を有害事象共通用語基準 V4に照らして評価し、患者病態の特徴表現としてリアルタイムに取得し、ソースとなった記載、結果などと合わせて保存し、活用した。例えば図6のように利用したシステム側で根拠を示すことで、人が容易に利用価値を判断できるように工夫した。

表 1 文脈ベクトル一覧

A	感染症
C	新生物
D	内分泌
E	精神、行動障害
F	神経系
G	眼および附属器
H	耳
I	循環器
J	呼吸器
K	消化器
L	皮膚
M	骨格筋、皮下
N	尿路、性器
O	妊娠
P	周産期
Q	先天奇形
R	異常検査
S	損傷、外因
R	症状、兆候
ペ	ペースメーカー関連
術	手術
乳	乳癌
O	妊娠
病	病理
放	放射線検査
化	化学療法
R	エコー検査

4 結語

最終的に臨床現場で AI を何のために、どのように使うのかという問いに対して、24時間連続して観察しなければならない事案など、大量データを俯瞰して判断する場面において、最終判断を行う人が納得できるように途中結果を説明できるものが求められていると考える。言い換えれば、機械の誤りをしっかりと判断できるコンテキストを併せ持つ AI が求められる。

参考文献

- 1) Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence, Nature Medicine volume 25, pages 433-438 (2019)
- 2) Mayo, Eko team on machine learning to detect heart abnormalities, <https://www.healthdatamanagement.com/news/mayo-eko-team-on-machine-learning-to-detect-heart-abnormalities>
- 3) 病理診断におけるデジタル化と AI の現状, 肺癌.2020;60:81-89
- 4) University of Florida researchers develop artificial intelligence system for fast, accurate patient care, <https://ufhealth.org/news/2019/university-florida-researchers-develop-artificial-intelligence-system-fast-accurate>



図 6 総合評価の画面例

計総合評価対象を自動的にコンピューターが選別した時、チャット画面上にシステムからなぜ選別したかを示している。この場合根拠となる記載の一部が切り出されて引用される。