

精神疾患、症状の定量化をAIで一慶應義塾大学医学部専任講師の岸本泰士郎氏講演レポート



東京大学政策ビジョン研究センター(組織改編により2019年4月より未来ビジョン研究センター)、慶應義塾大学メディカルAIセンター、エムスリー株式会社m3.com編集部が2019年1月から開催する「医療×AIセミナーシリーズ」のイベントレポートです。

東京大学未来ビジョン研究センター、慶應義塾大学メディカルAIセンター、エムスリー株式会社m3.com編集部が2019年6月15日に開催した医療×AIセミナーシリーズ第6回シンポジウム「医療AIの臨床への実装とトラスト」に慶應義塾大学医学部専任講師の岸本泰士郎氏が登壇し、「精神科におけるAI活用の取り組み」と題して講演した。症状の定量的な評価が難しい精神疾患の領域での人工知能(AI)の可能性や、新しい技術を使う際のルール作りなどについて話した。



講演する岸本氏

■ 定量的な評価が難しい精神疾患

うつ病や認知症などの精神疾患は、人類にとっての代表的な疾病負荷となっている。高齢化が進む現在では認知症の方が462万人と見積もられており社会費用も甚大だ。

一方、例えば雑誌「NEW YORKER(ニューヨーカー)」の2013年記事で報じられたように、精神疾患の治療薬の研究開発から大手製薬企業が撤退しているという危機的な状況がある。精神疾患の根本的な病理生理が分かっていないため、病態に直接作用する薬の開発が難しいからである。一方で、長年の開発の最後の関門である治験という場面において、薬の効果を証明するのが難しいことも少なからず影響している、と岸本氏は指摘する。

少し前のデータで、アメリカで最終的に米食品医薬品局(FDA)が薬として認可した薬がそれまでの治験でどれくらい失敗していたかという割合を見ると、抗うつ薬は半分以上、抗不安薬も半分以上、向精神薬で4分の1程度というように、プラセボと比較して有効性を示せていない現状がある。岸本氏の話では、これらの薬が臨床現場で効かないと感じることはなく、むしろ多くの患者の症状を改善できる有用なものだと実感しているという。こうした現状の背景にあると考えられる理由の1つが、精神疾患の症状は判断基準が曖昧で、専門医であっても定量的な評価が難しいという点だ。

例えば抗不安薬の治験で、HAM-Aという評価尺度で20点以上の患者に参加してもらう場合、専門医による評価で条件に合致していた人が選ばれる。だが患者自身で自己評価をしてもらうと、本来治験に参加すべきでない人が入っていると示唆されるようなことがあるという。多くの精神疾患でも同じような評価尺度があり、誰が点数をつけても同じ点数になるような仕組みを整えたり、評価者が訓練を積んだりしているが、中には「どの程度気持ちが沈んでいるのか」という質問など、数値で定量的に答えにくいものもある。

精神疾患の患者に限らず、気持ちというのは移ろいやすく、質問するタイミングや聞き方でも答えが変わってしまうという難しさがある。一方の医師側としても、治験に参加できる該当者をできるだけ紹介してほしいと頼まれることがあり、迷った時には該当する方向になるように点数をつけてしまっても不思議ではないという。各項目の点数を加算していく評価手法の場合、4-5点程度の差はすぐに出てしまう。こうして本来は対象ではない患者が治験に入ってしまう、プラセボと候補薬の効果に差が出なくなるという現象が起きるのではないかとされている。

もう1つ、治験で候補薬の有効性を示すのを困難にしているのが「多次元性」だという。患者によって点数が高く出る項目や低く出る項目は異なり、症状の現れ方にも差があるが、合計点数としてまとめられる。臨床症状が異なる患者も評価尺度では同じように判断されるという問題点は昔から指摘されてきたという。

また、認知機能検査に関する課題も多くある。検査には認知機能を広く浅く測る尺度もあれば、特定の認知領域を中心に詳細に検査するものもあるが、簡便なものは大雑把な結果になるという問題があり、一方で、詳しい検査は1~2時間かかって患者にも医師にも負担を強いてしまうという問題がある。さらに、もともと高い能力があった場合はある程度機能低下が進まなければ症状が把握できない「天井効果」、能力がある程度下がってしまうとそれ以上の変化は追跡できない「床効果」、同じ検査を繰り返すと患者側が慣れてしまう「学習効果」などが生じる。

カメラやマイク、表情検知ソフトで精神疾患患者の症状を定量化

こうした状況に対して、IT機器などのテクノロジー、ビッグデータとAIを使った解析を活用して、症状の特徴量を抽出しようとする研究も増えてきている。精神科領域の研究では脳で何が起きているのか、病態生理を直接反映するようなバイオマーカーを捉えられるようになる事が理想だが、こうした新たな取り組みの中にも可能性が感じられるものがあるという。精神疾患の症状は日常の行動と密接に関連している場合が多く、センシング技術などで簡単に患者に負担をかけずに変化を見られるようになってきたことも、こうした展開に関連していると岸本氏は見ている。

例えば海外では、スマートフォンを使って躁うつ病患者の活動状態を定量化したり、会話の仕方を解析して患者の気分を定量化したりする研究があるという。実際の臨床でも、躁状態の患者はよく話すというような特徴があるため、こういった手法で特徴量を定量化できると感じるという。岸本氏自身も日本医療研究開発機構 (AMED) の支援を受けたPROMPTというプロジェクトに3年半ほど前から取り組んでおり、うつ病や認知症の患者の精神症状の定量化を目指している。うつ病では重症になると体の動きも重くなり、話しかけても返事が返ってくるまで時間がかかるようになり、話す時の声の抑揚もなくなるという特徴が見られるという。

カメラや指向性の高いマイク、人の表情を読み取るソフトなどを使って、診察中の患者の表情、体の動きのスピード、患者の声やトーン、話すスピードなどのデータを収集し、動きが遅くなる「精神運動抑制」の定量化を試みている。こうしたデータを精神科でこれまで利用されてきた評価尺度と照合して、機械学習に活用できれば、新しい患者の重症度を機械が自動的に推定できるようになる。現状でもある程度の精度で重症度や症状の有無を推定できることは確認できており、研究チームが目指していた相関係数に到達したという。同じように、ウェアラブル装置を使って数日間の日常行動からうつ病の症状の強さを推定するシステムも開発している。

言葉遣いの特徴から患者の重症度を推定

これまでとは違う新しい着眼点で精神状態の定量化を目指す取り組みもある。共同研究先の東京工業大学の研究チームは声の音響学的な特徴から認知症患者と健常者を区別する方法を探っている。検証ではかなりの精度で認知症が発症しているかどうかを推定できた。臨床医がこれまで気付いていた特徴だけではなく、これまで気付いていなかった新しい要素も、今後の診断や重症度の評価に役立つ可能性があるという期待を込めた。

また、岸本氏は科学技術振興機構 (JST) の支援を受けているUNDERPINというプロジェクトで、自然言語処理の技術を使って言葉遣いの特徴から患者や重症度を推定しようとしている。どんな語彙を使っているか、指示語はどれくらいの頻度で出てくるか、文章全体の構造はどうなっているかなど、様々な特徴が数値化できるようになってきている。

このプロジェクトでは患者一人につき30分から1時間のインタビューを受けてもらう。岸本氏らは、今までに350時間分のデータを収集した。海外からも類似の研究が報告されているが、例えばうつ病の患者は一人称代名詞が増えて、話すことを断言する傾向があるという報告があり、患者の話し方を定量化したデータが利用できる可能性がある。岸本氏の研究においても類似の傾向があるという。

また、岸本氏のチームは言葉や文章を「ベクトル化」する技術を使い、10分程度の会話を書き起こしてAIにテキストから認知症の可能性を推定する研究にも取り組んでいる。会話の特徴をベクトルで表し、データ空間にプロットすると特徴量が似たケースが集まり、症例の分類が可能になる。ある程度の精度で認知症やMCIの患者が推定できる段階まできているという。

AIを使って、患者に合う治療法を選びやすくするための手がかりにしようという研究もある。電気痙攣療法は精神科領域で治療効果が高い治療法だが、患者は何週間も入院して週2〜3回施術を受けるなど手間もコストもかかる。岸本氏らはMRIで撮像した脳領域の構造に対する機械学習によって、この治療法が有効かを予測しようとしており、今まで反応性の指標として知られていた人口統計学的な因子や症状の特性に脳領域の構造に関する情報を加えることで、その予測精度が大きく伸びたという。

医師は機械の結果をどのくらい信用すべきか？

以上のようなデータの蓄積や機械学習を用いた取り組みを通じて、良い医療を患者に提供できるようになるかもしれないと岸本氏は大きな可能性を感じている。自身の取り組みは暗黙知の形式知化、もしくは情報空間の可視化だと表現し、定量化しにくかった精神疾患の重症度を客観的に示すことができるようになれば日常の治療の効果が判断しやすくなり、新薬の開発にも役立つかもしれないという。精神科領域にはもともと精緻なラベル付きのデータがなく、重症度や病気の判定に使えるバイオマーカーもなかったため、全てを一から蓄積している状況だ。

集めているデータがこれで適切なのか、ラベルの付け方は正しいのか、精度はどの程度に設定すべきなのかなど迷うことも多い。機械の出す結果をどのくらい信用すべきなのかという難しい課題もある。例えば、今までのデータから治療の効果が望めそうにないという推定がなされれば、その患者には効果があるかもしれないのに、その治療を選択肢から外してしまうことがありうるからだ。また、このような技術が多くの人々の社会場面で幅広く利用されるようになる際には、その社会的影響についての検討も必要だ。例えば、検査を希望していない人の音声から勝手に認知症の進行を推定できてしまうといったことが生じてしまうかもしれない。今後、規制や開発の責任についての議論も必要になると岸本氏は指摘した。

講演の最後には、会場からの音声認識の技術の使い勝手についての質問があり、岸本氏は現状では患者の会話に適用できるレベルではないとした。結局10分程度の会話であれば書き起こすためのコストはさほどかからないため、将来的には診療報酬との兼ね合いもあるものの技術を実用化する際のハードルにはならないとの見通しを示した。



鴻知佳子
ライター

大学で人類学、大学院で脳科学を学んだ後、新聞社に就職。バイオを中心とする科学技術の関連分野を主に取材する。約10年の勤務後に退社。ずっと興味があった現代アートについて留学して学び、現在はアートと科学技術の両方を堪能する方法を模索中。