

7. 医療分野におけるAI開発の実際

大田 信行 Preferred Networks America, Inc.

現在、人工知能 (AI)、特にディープラーニング (深層学習) という機械学習の手法は、画像認識、音声認識、自動翻訳ばかりでなく、製造業のロボットや自動運転まで、さまざまな社会インフラに应用が始まっている。近い将来、医療に深層学習を適切に応用することにより、さまざまな医療分野に大きな変革をもたらすであろうAI医療への期待が高まっている。株式会社Preferred Networks (以下、PFN) は、深層学習をはじめとする機械学習技術の医療応用への研究開発と医療産業への展開を進めている。特に、国立がん研究センターと、がんの早期診断・予後予測・創薬などへ機械学習技術を応用し、AI医療実現をめざす共同研究を進めている。本稿では、これまでの国立がん研究センターとの共同研究を通して経験したAI医療の開発の実際と可能性、そして、課題を検討してみたい。

Liquid biopsy (血液による) がん早期 診断技術の開発

PFNは、2015年から国立がん研究センターと共同で、13種類のがんを血液1滴から深層学習で予測する研究を行っている。この研究では、血中のmiRNAの発現パターンを深層学習で分析することにより、さまざまながん種の早期診断において、感度、特異度共に画期的に向上させることを可能としている。特に、乳がんにおいては、感度、特異度共に99%以上を達成している。この大型プロジェクトは、日本医療研究開発機構 (AMED) の支援の下、国立がん研究センター分子細胞治療研究分野長、落谷孝広博士主導により2014年から行われている。深層学習は、大量のデータを学習し、そのデータに潜む特徴を自動抽出することによって予測精度が画期的に向上する。この大型プロジェクトにより、血中のmiRNA発現データを大量に利用・分析することが可能になった。通常、特定のがん種の血液データは、100件のデータでも論文に発表することは可能であるが、深層学習の利点を引き出すためには、どうしても大量のデータが必要になる。このプロジェクトでは、最終的に約8万検体ものデータを集める計画であり、膨大なデータを深層学習で分析することにより、さまざまながん種を精度良く予測することが可能になる。今後ともさらに精度を上げていくよう、改良を

進めていく予定である。

機械学習には大きく2つの学習方法がある。「教師あり学習」と「教師なし学習」である。深層学習は、大量のデータを用いることによりその威力を発揮することができるが、医療やヘルスケアで利用する場合は、これまでのところ教師あり学習が基本となる。

教師あり学習とは、医用画像や医療バイオデータ (ゲノムや血液データ) などとともに正解データ (例えば、がん種のデータ) を一緒に与えることで、ニューラルネットワークが自動的に特徴を学習していく手法である。上述した血液によるがん早期診断においては、大量のがん患者と健常者の血中のmiRNAプロファイルデータをニューラルネットワークに入れて、どの種のがんであるかを予測させる。そして、それぞれ予測がどの程度合っているかを正解データ (この場合、がん種や健常など) と損失関数を用いて比べることにより、より正しい予測ができるようにニューラルネットワークの重みを更新していくのである。よって、深層学習を医療に使う場合は、単にデータ数が多いばかりでなく、カルテなどの患者の健康状態がわかる医療データと紐付いていないと正解データとして用いることができない。今回のプロジェクトでは、大量の血液データとそれに紐付いた患者医療データを正解データとして利用することが可能であったが、たいていの場合、深層学習に利用できる質と量が担保された医療データはなかなか世の中に存在しないのが現実である。