

II 医療AIを加速させる研究開発の動向

5. 医療AI人材育成と医療オントロジー

高山 順 東北大学大学院医学系研究科AIフロンティア新医療創生分野/
 理化学研究所革新知能統合研究センター遺伝統計学チーム

東北大学は、北海道大学、岡山大学とともに、文部科学省プロジェクト「『Global×Localな医療課題解決を目指した最先端AI研究開発』人材育成教育拠点—Clinical AI—」に取り組んでいる。本プロジェクトの目的は、高度人工知能(AI)技術に精通した医療人材を育成し、地方への医師不足や少子高齢化など、先進国が共通して抱える地域医療の課題を解決することである。本稿では、本プロジェクトの一環の講義でも扱われており、実用化されたAI技術である医療オントロジーについて概説する。

オントロジーとは

オントロジー(ontology)は、人類の知識を計算機で表現し処理するための技術である。Gruber(1993年)によれば、オントロジーとは「概念化の明示的な記述(仕様)(explicit specification of conceptualization)」であり、概念化とは特定の関心領域における“実体”と、それらの間に成立する“関係”であると定義される¹⁾。例えば、「白血球は造血幹細胞から発生する」という知識を計算機で扱うことを考える。この時、“白血球”と“造血幹細胞”は“実体”として、“から発生する”は“関係”として、それぞれ扱うことができる。“関係”のほかの例では、“具体-抽象関係(is-a関係)”が挙げられる。is-a関係は、さまざまなオントロジーで用いられる基本的な関係である。例えば、「リンパ球は白血球の一種である」という知識を表現できる。

これら特定の形式で知識を表す、すなわち“知識を構造化する”ことで、人間にも理解でき、計算機にも処理が可能な表現が可能となる。もちろん、すべての医学知識を特定の形式のみに従って表すことは困難である。しかし、これらの形式を用意するだけでも、造血系細胞の発生系列という医学的知識を柔軟に表現し、推論などの処理を計算機に行わせることが可能となる²⁾。

知識の計算機処理におけるオントロジーの利点は何か。人類の知識は、例えば、書物や論文のような文章(フリーテキスト)形式で表せる。フリーテキスト形式は表現力が高く、これを対象とした自然言語処理技術による知識抽出が、近年大きく発展している。一方で、多義語や表記ゆれ、分野固有の専門用語、文意のあいまいさなど特有の困難があり、また、大規模な教師データを必要とするなど利活用に向けた課題も多い。他方、知識を表現する手法の一種として、統制された用語集であるところの語彙(vocabulary)の利用も考えられる。語彙に定義された用語のみを用いた知識表現は、表記ゆれを排した計算を可能にするものの、用語間の関係に関する情報がない点で表現力に乏しい。オントロジーは、統制された用語(すなわち概念)とその間の関係を定義することで、高い表現力と容易な計算機処理の両方を可能にする知識の表現形式である。

医療分野において、患者の示す表現型を計算処理可能な形式で表す際には、オントロジーを利用することが事実上の標

準となっている。特に、Robinsonらが提唱したHuman Phenotype Ontology(以下、HPO)は盛んに利用されている³⁾。HPOは、もともと希少難病の表現型を記述するために開発されたが、今では神経疾患、腎疾患、免疫系疾患をはじめとするコモンディゼースをも対象に含むまで拡張されている⁴⁾。また、大規模コホートでの利用や表現型のマッチング、診断支援システムに、実際に利用されている。希少難病は有病率がきわめて低い上に、7000とも言われる疾患数の多さから、長い経験を有する医師でも診断が困難なことがある。しかし、表現型のオントロジーを用いることで、ほかの疾患や、(しばしば未診断の)ほかの患者との類似性の検索が行えるようになる。

オントロジー技術がすでに実用的なレベルにまで発展しており、また、診断支援に適用可能であるという特徴は、人類の蓄積してきた知識を利用しやすい形で整備したことの賜物である。したがって、オントロジー技術は、地域医療で顕著な医師不足の問題解決にも有効と考えられる。さらに、一度でもオントロジー技術に触れたことのある医師は、カルテや退院サマリーなどの二次利用価値が高い医療文書を作成する際にも、機械可読性を意識した文章を執筆することが可能となるだろう。これらの意味で医療オントロジー技術についての講義が、冒頭で述べた人材育成プロジェクトに取り入れられていることはきわめて意義深いと言える。次節以降でHPOの詳細、表現型間の類似性計算の原理を概観する。