



HOKKAIDO UNIVERSITY

Title	医療データにおけるシンボリックデータ解析の適用に関する研究 [論文内容及び審査の要旨]
Author(s)	高木, 諒
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第14175号
Issue Date	2020-06-30
Doc URL	https://hdl.handle.net/2115/79041
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	Ryo_Takagi_abstract.pdf, 論文内容の要旨



学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称 博士（情報科学） 氏名 高木 諒

学 位 論 文 題 名

医療データにおけるシンボリックデータ解析の適用に関する研究
(A study on applications in symbolic data analysis for medical data)

近年の国際的な医学データの公開化の流れにより、医学分野におけるデータの二次利用のニーズが高まってきている。しかし、公開されるデータが蓄積されてはいるが、それらを適切に活用できてはいない。一般的に、医学データの解析は、安全性や有効性の評価が目的で実施されるため、探索的な解析が十分に行われている事例は少ない。さらに、医学の発達により得られるデータが多様化し、従来の方法での解析が難しくなっている。

このような様々な形で与えられているデータについて探索的に解析するシンボリックデータ解析と呼ばれる方法論が提案されている (Diday E, 2008)。シンボリックデータ解析は、その解析対象であるクラスを、データそのものの個体としてではなく、個体の集合体として解析する。その集合体は、区間値や分布値データなどとして表現でき、それらの解析を通じて、従来の方法では見いだせなかった知見を得ることを目的とする。

医学データの解析例のひとつに、放射線治療における、がん治療評価の方法がある。治療評価は、放射線照射による損傷度合いを定量化して行う。従来、世界的に使われている方法として、治療評価を対象臓器の総線量とする評価指標がある。しかし、放射線はその種類に応じて、生体内で線量が分布することや、がん以外への放射線の線量に関しても、一様ではないことが一般的に知られている。このことから、解析対象として、線量そのものを分布とした、新しい観点での線量分布の解析が肝要である。

そのほかに、医学において重要視されていることとして、信頼性の高い医療の提供がある。高い信頼性を得るには、相応の根拠ある結果が必要であり、その最たるものとして、メタアナリシスがある。これは、異なる様々な研究結果を適切に統合して、高い根拠を持つ結果を導くための方法論である。メタアナリシスの適切な実行には、もともと研究間に存在する違いの考慮が必要である。探索的な解析において、その異質性の発見は、メタアナリシスをより適切に実行するために重要である。

以上のことから、本論文では、医学データにおいて、従来のような解析対象ではなく、より広い形での解析対象として、データを分布値や、集計データとして記述した場合のシンボリックデータ解析の適用について提案する。

本論文は、5つの章から構成される。各章の概要は以下の通りである。

第1章では、本論文の背景、目的、および構成について説明している。

第2章では、本研究の主題である、シンボリックデータ解析について、その概念と用語の定義を与え、解析対象の種類、記述の方法について説明している。

第3章では、放射線治療におけるシンボリックデータ解析の適用について説明している。放射線によるがん治療において、治療対象であるがんや、それ以外の正常臓器に照射される線量分布が放射線治療の計画段階から dose-volume histogram とよばれるグラフから把握できる。臓器に対して

照射される総線量そのものを利用した放射線の照射計画を立てる (Clark BG, 1998) 方法論が、現在、世界的に主流である一方、dose-volume histogram で与えられる線量そのものを分布として正常臓器の評価を行う (O'Donoghue JA, 1999) ような、シンボリックデータ解析としての枠組みを用いた評価指標も存在している。放射線治療の対象となるがんには様々な特性があること、正常臓器に照射される線量についてもいくつかのパターンが存在することから、多くの放射線照射の状況においても、適切な評価が可能な指標であるかの検証は重要である。ここでは、実際のがん治療における比較として、肺がんと前立腺がんにおける性能比較、ならびに、放射線治療における典型的な正常臓器への線量分布を仮定した場合における評価の妥当性について、総合的に検証した。

第4章では、メタアナリシスにおけるシンボリックデータ解析の適用法について提案する。一般的なメタアナリシスは、「メタアナリシスにおける研究結果」を統合したうえで、治療効果の推定を行う。統合する研究間の異質性については慎重に考慮する必要がある。しかし、治療効果に影響を及ぼす因子についての発見的な解析がなされることは少ない。そこで、本章では、「メタアナリシスにおける研究結果」をクラスとした、階層的シンボリッククラスタリングの方法を提案する。メタアナリシスを実施する多くの場合、2値の治療に関するデータと2値の結果に関するデータが与えられており、2元分割表として変換できる。そのため、「メタアナリシスにおける研究結果」である、2元分割表として記述されたクラスについて、カット距離 (Bolla, 2010) を基とした、非類似度を用いた階層的シンボリッククラスタリングを提案した。実データの解析例として、慢性頭痛に対する鍼治療と標準治療の効果を比較したメタアナリシスのデータに対して提案手法を用いることにより、この手法の有益性を確認した。

第5章で、本論文の総括と成果について述べている。