



Title	放射線治療決定支援のための連続値属性を離散化する手法の研究
Author(s)	国府, 裕子
Citation	大阪大学, 2004, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/45353">https://hdl.handle.net/11094/45353</a>
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、<a href="https://www.library.osaka-u.ac.jp/thesis/#closed">大阪大学の博士論文について</a>をご参照ください。

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

氏 名 国 府 裕 子

博士の専攻分野の名称 博 士 (保健学)

学 位 記 番 号 第 18563 号

学 位 授 与 年 月 日 平成 16 年 3 月 25 日

学 位 授 与 の 要 件 学位規則第4条第1項該当

医学系研究科保健学専攻

学 位 論 文 名 放射線治療決定支援のための連続値属性を離散化する手法の研究

論 文 審 査 委 員 (主査)

教 授 手 島 昭 樹

(副査)

教 授 上 甲 剛 教 授 村 瀬 研 也 教 授 大 野 ゆ う 子

## 論 文 内 容 の 要 旨

## [目的]

病歴記録データベースの有効利用の一つに蓄積されたデータを用いた治療計画の決定支援などがあげられ、決定分析はその手法の1つである。しかし決定分析を行う際、照射線量などの連続値属性があらかじめ離散されている必要があり、連続値属性を最適な境界値で離散することが一つの重要な課題である。そこで連続値属性を離散値属性に変換するために治療結果に統計的有意差を生じる境界値を自動認識するアルゴリズムを考案する。

## [方法ならびに成績]

治療成績に統計的有意差が生じるか検討するために、対象となるデータセットは2つもしくは複数のデータサブセットに分けられる必要があり、かつ各データサブセットは治療成績の差を検定するのに適したデータ数を備える必要があった。提案された第一アルゴリズムは連続値属性において一定のクラスインターバルで等間隔分割することで複数のデータサブセットを作成した。このクラスインターバルは各データサブセットのデータ数に直接影響を与えるため、クラスインターバルを定める3つの方法が提案された。

方法 A : 最大値、最小値およびデータ数に基づきクラスインターバルを決定する方法

方法 B : データ数とクラスインターバルの対応表に基づきクラスインターバルを決定する方法

方法 C : スタージェスの公式に基づきクラスインターバルを決定する方法

以下は第一アルゴリズムの流れである。

- (1) 連続値属性を一定の値を持って最小単位であるデータサブセットに分割する。
- (2) (1)で作成された各データサブセットを隣接する順に比較検定する。
- (3) (2)の結果より、データサブセットの統合および分割を行う。
- (4) 上記のプロセス(2)および(3)を繰り返し最小単位データサブセットから作成された複数のデータサブセットの集合体に基づき境界値の検出を行う。

しかし上記のアルゴリズムはデータがノンパラメトリック分布する場合、クラスインターバルを定め等間隔で分割する第一アルゴリズムは(1)の過程で(2)の検定を行うに十分なデータ数を持たないデータサブセットを多く発生させる可能性があり、データ分布が不均一の場合に適した第二アルゴリズムが考案された。第二アルゴリズムは(1)の過程

で各データサブセットが任意のデータ数を備えるように非等間隔のクラスインターバルをもって分割することで上記の問題を回避した。あらかじめ境界値が確認されているサンプルデータセットを対象に検証を行った後に、上記のアルゴリズムを非小細胞肺癌の症例に適用した。

#### [総括]

第二アルゴリズムは第一アルゴリズムに比べて多くの計算量を要するが、データ分布の影響を受けにくいアルゴリズムとなった。計算時間は実用的であった。非小細胞肺癌の症例を対象に照射線量という連続値属性に関して統計的に有意な差を生じる境界値の検出を行った。

### 論文審査の結果の要旨

本研究では放射線治療の臨床データにおいて至適治療条件の自動検索手法を開発しその有用性を検証している。対象データ群を連続値属性の値にもとづいて複数の初期データサブセットに分割し、分割・統合を繰り返すことで最適な複数グループ化を行い統計的に有為な差を生じる適切な境界値を検出する。この時、初期データサブセットの生成を対象データ群のデータ分布の特徴等に基づいて行うことで、従来の手法よりも高い確率で統計的に有為な境界値の検出を可能とした。

研究内容そのものには新しい知見を含み、その有用性を全国からの集計データから抽出した数群のデータおよび現在までの一般的な他の統計手法と比較して証明しており、学位授与に値すると判定した。しかし、主論文として今回提出している Jpn J. Med. Phys. 23 (3) : 163-172, 2003 にはそれらの知見が十分記述されていない。したがって期限までに公聴会での発表内容を含む詳細な Thesis を主論文として再提出することを条件に、学位授与の価値があると認定する。