



Title	感性及び医療データにおけるプロダクショナルルール指向な統計解析手法に関する研究 [論文内容及び審査の要旨]
Author(s)	万, 可
Citation	北海道大学. 博士(情報科学) 甲第14583号
Issue Date	2021-03-25
Doc URL	<a href="http://hdl.handle.net/2115/81204">http://hdl.handle.net/2115/81204</a>
Rights(URL)	<a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>
Type	theses (doctoral - abstract and summary of review)
Additional Information	There are other files related to this item in HUSCAP. Check the above URL.
File Information	Ke_Wan_abstract.pdf (論文内容の要旨)



[Instructions for use](#)

## 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 万 可

### 学位論文題名

感性及び医療データにおけるプロダクションルール指向な統計解析手法に関する研究  
(Studies on Production-Rule Oriented Statistical Method for Affective and Medical Data Analysis)

統計的データ解析において、結果に対する精度は重要であるが、それを追求するあまり、解釈が困難となつては、実用面において、有益な知見にたどり着くことができない。特に、実用面において解釈の付与そのものが極めて重要な意味をもつ分野におけるデータの解析では、結果の解釈が可能であることが最優先事項である。本研究では、解析結果に対する解釈可能性の向上を目的として、プロダクションルール指向の統計解析手法に焦点を当て、問題点の改善を含めた方法論を構築する。さらに、感性データ及び医療データへの適用を通じ、その有用性を確認する。

プロダクションルールは人工知能において黎明期から使われており、「if-then」の形式で与えられる。Quinlan (1984) は決定木からプロダクションルールを生成する手法を提案した。ここではプロダクションルールを、決定木のルールとしてそのまま用いており、解釈が非常に容易である。本研究では、Quinlan (1984) におけるプロダクションルールの条件部をあらゆる要素を、変数と定数の大小関係の積集合に制限し、このようなプロダクションルールを導入した統計手法を、プロダクションルール指向な統計解析手法と呼ぶこととする。具体的な対象手法には、決定木、サブグループ同定法及び指示関数に基づく回帰手法などが含まれる。特に、決定木はグラフィカルな表現を可能とするため、解釈が非常に容易であり、データ解析においても頻繁に使われている。

感性データ解析における決定木の適用例として、楊ら (2012) は街路評価の影響要因を決定木である CART(Classification and Regression Tree; Breiman, 1984) で探索を行った。因子分析や回帰分析より簡単に影響要因を解釈することができたが、連続応答において CART による分類では、お互いに異なる評価のクラスに分けられるとは限らないため、各クラスターの解釈が困難であった。それに対し、指示関数に基づく回帰手法である適応的指示モデル (Adaptive Index Model; AIM; Tian and Tibshirani, 2010) は、決定木と同じく説明変数に基づきプロダクションルールを構築することだけでなく、プロダクションルールに基づき得点化もなされる。また、高い(低い)得点は高い(低い)応答と対応するため、点数に基づき分類された各クラスターの解釈が容易にできる。一方、AIM ではデータが線形構造を持つ場合、線型モデルを用いた解析法と比べて精度を損ねることが Tian and Tibshirani (2010) で示されているため、本研究では既存手法 AIM の予測精度の改善も目的とする。

医療分野において、精密医療や個人化医療が注目される中、臨床試験で得られた生存時間データを用いて新規治療と既存治療の比較を行う際、新規治療に対して有効・安全な患者像が求められることがあるため、本研究では治療効果が顕著に高いサブグループを生存時間データから抽出する手法に着目した。これらの手法において、PRIM 法 (Patient Rule Induction Method; Friedman and Fisher, 1999) のアルゴリズムに基づき生存時間データに拡張した研究が行われている (Kehl and Ulm, 2006) が、生存時間に関して、新規治療と既存治療の生存時間分布間に比例ハザード性、すなわち両群のハザードの比が時間の経過によらず一定であることを仮定しなければならない。例えば、PD-1 抗体あるいは PD-L1 抗体のような免疫チェックポイント阻害剤による治療と殺細胞性抗癌剤治療を比較する臨床試験では、免疫チェックポイント阻害剤の作用機序により治療群間の比例ハザード性の仮定を満たさないことは広く知られている。したがって、本研究では比例ハザード性の仮定を満たさない生存時間データにおいても適用可能な PRIM 法に基づくサブグループ同定法

を提案する。

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。

第一章では、本論文の背景、目的、及び構成について説明する。

第二章では、本研究で扱うプロダクションルール指向な統計解析手法、感性データ解析及び医療データ解析の既存研究について説明する。

第三章では、AIMの解析例を示したのち、線形構造を仮定できるデータに適用した場合の予測精度の向上を目的として、ハイブリッド型適応的指示モデル (HAIM; Hybrid AIM) を提案する。本章で手法の検討のために用いるデータセットとして、街路利便性評価データを用いる。このデータの解析目的は (1) 街路利便性評価の影響要因、(2) 各影響要因をどの程度評価すれば街路利便性評価に繋がるか、(3) 街路利便性評価に基づく類型化である。既存の研究で用いられている決定木 (CART) では目的 (1)(2) を達成できるが、目的 (3) を達成することが難しい。本研究では、目的 (1) ~ (3) のすべてを達成できる AIM を適用した。その結果、中国成都市の街路利便性評価の影響要因を明らかにし、各影響要因と街路利便性の関係を簡単に解釈することができ、推定された得点に基づき、異なる街路利便性評価を行った幾つかのクラスターに分類することができた。また、AIMの予測精度の改善については、説明変数が制御可能変数と制御不可能変数に分けられる場合に対して、AIMと線形モデルを結合したハイブリッド型適応的指示モデル HAIM を提案した。ここで、制御可能変数はアルコール摂取量、喫煙者か否かなど制御できる変数で、制御不可能変数は身長、性別など、人為的に制御できない変数として定義している。したがって、実用上、制御可能変数に基づいた影響要因の探索がより重要であることが考えられるため、本研究では、線形モデルと AIM を結合し、制御不可能変数に対して精度の高い線形モデルをあてはめ、制御可能変数のみに AIM を適用し解釈を求める手法を開発した。また、数値検証においては、データ構造に線形傾向がある場合、提案手法の予測精度は既存手法を上回った。事例検証においては、街路利便性評価データは説明変数を制御可能変数と制御不可能変数に分けることが困難であるため、ここでは、中国成都市における伝統的な街並みに訪れる観光客の満足度評価データ (Liu et al., 2017) を事例として用いた。そのうち、歴史的建築物や河川景観などランドスケープに関わる項目を制御不可能変数とし、残りの項目を制御可能変数として HAIM に適用し、既存手法の AIM の適用結果と比較した。その結果、提案手法では観光客の満足度影響要因を明らかにすることができ、また既存手法である AIM より高い予測精度が確認された。

第四章では、比例ハザード性を満たさない生存時間データにも適用可能なサブグループ同定法として、境界内平均生存時間 (Restricted Mean Survival Time; RMST) に基づく生存時間 PRIM 法を提案する。ここでは、生存時間データの比例ハザード性に関わらず、治療効果が顕著なサブグループを抽出することを目的とする。また、近年では比例ハザード性を満たさない生存時間データも考慮され、新規治療群 (介入群) と既存治療群 (対照群) の RMST の差を用いて治療効果の差を評価している研究が幾つか行われている (Delmar et al., 2017; Ngo et al., 2020; Zhang et al., 2020)。したがって、既存手法である生存時間 PRIM 法 (Kehl and Ulm, 2006) に RMST を用いて改良し、比例ハザード性を満たさない生存時間データ解析にも適用可能なサブグループ同定法、RMST に基づく生存時間 PRIM 法を提案する。また、数値検証においては、生存時間データの比例ハザード性に関わらず、提案手法が優位であることを示した。事例検証においては、乳がん罹患している患者に対するホルモン療法に関する無作為化比較第 III 相試験データ (Schumacher et al., 1994) に提案手法及び既存手法を適用した。その結果、提案手法は既存手法と比べ、治療効果の差がより大きいサブグループを抽出することができ、優位性が確認された。

第五章では、本論文のまとめ及び成果について述べている。