

学 位 論 文 題 名

データ解析における知識支援システムに関する研究

学位論文内容の要旨

近年、コンピュータ指向の高度なデータ解析の手法が数多く提案されている。また、データ解析に用いられる計算機は大型機からワークステーションなどの小型機に移りつつあり、グラフィック表示等を通じてユーザとのインタラクションを取り入れるなど、データ解析を行なうための計算機環境は量、質ともに改善されてきている。しかし、統計学に関する専門的な知識をほとんど持たないユーザが必要に迫られてデータ解析システムを利用する機会は少ない。そのため、手法の誤用や無意味な手法の適用がなされる可能性が高い。

データ解析に関する知識を持たないユーザが少なくともデータ解析や統計学の観点からみて意味のある解析を実行するために支援を試みる研究として、知識工学などの成果を用い、データ解析の専門家が有する知識を計算機上に実装し、支援に役立てることを目的としたデータ解析エキスパートシステムの研究がある。しかし、その研究例の多くは特定の解析手法を利用するための支援を目的としたものであり、観測データに対し適切な解析手法を提供するものは少ない。また、知識処理において知識ベースと推論機構は本来分離されているのが望ましいがデータ解析エキスパートシステムとして報告されているもののほとんどは両者が明確に区別されずに用いられている。

以上の背景から、本論文ではデータ解析に対する計算機支援を目的とし、特に多次元データの解析に適切な手法を提案し作業実行を支援するシステムの実現についてデータ解析およびエキスパートシステムなどの知識工学の両面から考察し、解析作業や解析に用いられる知識の定式化を行なうとともに推論機構から独立した知識ベースの構築を行なった。

本論文は、以下の7章から構成されている。

第1章では、本研究の背景とデータ解析支援システムの研究の現状について概説し、本研究の目的について述べた。

第2章では、本研究が支援対象とする多次元データ解析とその手法について解説した。またデータ解析のためのシステムとその具体的な利用例を示し、最近の動向について詳述した。

第3章では、データ解析支援システムに関する従来の研究についての概観と具体的な説明を行ない、従来の研究における解析支援形態が特定の手法を対象に詳細な情報を提供するものと観測データに対して複数手法から適切な手法を選択するものの二種類に類別されることを明確にした。また、研究例で用いられている知識処理手法について、知識表現と推論機構に関する検討を行なった。

第4章では、本研究が手法選択の支援を目的としたものであるという立場を明確にし、解析支援実現のためのさまざまな提案や定式化を行なった。まず、データ解析において用いられる知識についての考察から、解析時に必要な知識がデータに関する形式的知識、処理に関する知識、解析対象における領域知識の三種類に細分されることを示し、さらにデータ解析システムの持つ汎用性を考慮して領域知識を除いた上記二種類の知識が支援に際して

利用可能であることを示した。また、データに関する形式的な情報を記述する枠組として Metadata の概念を導入し、データおよび処理に関する知識が Metadata を用いて記述できることを示した。つぎに、データを解析する過程を細分し、データに関する形式的知識と処理に関する知識が領域知識に比して大きな影響を及ぼす部分として解析計算部分を支援対象として特定した。さらに、観測データを初期状態、解析処理を操作子とすることにより解析処理列の構築が前向き推論による問題解決として定式化できることを示した。これらの提案に基づき、観測データに関する Metadata と解析目的から前向き推論による問題解決を行なうことにより、観測データから出力結果までの適切な処理列を構築するようなシステムについて、解析手法を研究するために公開されたデータを適用した例を通して、実現可能性と有用性を確認した。また、システムの動作過程について検討を行なった。

第5章では、第4章で定式化された知識の構成に基づいた上での、推論機構から独立した知識ベースについて考察を行なった。既存の研究例における知識表現からプロダクションルール、評価表、フレーム、オブジェクト指向の四種類を具体的に取り上げて詳細な検討を加えた上で、本研究が対象としている手法選択に適した知識ベースとしてフレーム理論に基づく枠組の提案を行ない、多次元データ解析で用いられる基本的手法に関する知識ベースを構築した。更に、構築した知識ベースを第4章で用いた推論機構に適用し、第4章の実行例と同一のデータによる利用例を通して知識ベースとしての有効性を示した。

第6章では、第4章で用いた推論機構に関して再検討を行なった。データ解析作業という観点から後向き推論の適用が困難であることを指摘した上で、より自然な推論方法の実現を目的として、非単調推論の一実現形式である TMS (Truth Maintenance System) を応用した推論系の導入を提案した。また、構築したシステムの動作を確認するため、あらかじめ解析結果が得られているデータを対象としてシステムによる解析を試み、指示した解析過程との対比を行なった。また、第4章で用いた前向き推論との推論過程の比較を行ない、TMS 利用による改善点について考察を行なった。さらに、第5章で提案された知識ベースの適用例も合わせて示し、推論機構と知識ベースの独立性を確認することで、データ解析におけるエキスパートシステムにおいて両者が分離可能であることを示した。

第7章では、結論としてデータ解析における形式的知識を用いた問題解決手法によって手法列を提案する支援が可能であること、解析支援における従来の研究例ではほとんど明確な区分がなされていない知識ベースと推論機構が明確な形で分離可能であること、実際の解析過程に近い形での推論機構として TMS の導入が有効であることなど、本研究の有効性および適用可能性などについて得られた知見を述べた。また、領域知識の取り扱いなど本研究において残された問題点について触れ、それらの問題点へのアプローチに対する指針について考察を行ない今後の解析支援システムの方向性を示した。

学位論文審査の要旨

主査	教授	佐藤	義治
副査	教授	宮本	衛市
副査	教授	大内	東
副査	助教授	水田	正弘

学位論文題名

データ解析における知識支援システムに関する研究

近年、知識工学の発展に伴い、データ解析の分野においてこれを効果的に利用し、統計学に関する知識をそれほど有しない利用者に対して、統計学的に不合理ではない結果が得られるような計算機による支援を行なうための研究が進められている。しかし従来の研究においては支援する解析手法が限定されたものがほとんどであり、さらに出力処理までを含めた一連の解析作業の実行にまで踏み込んだ支援を行なうものは少ない。

本論文はこのような背景に基づき、与えられたデータや解析目的に応じて適切な解析手法を選択し一連の作業を行ない解析結果を得るようなデータ解析システムに関する計算機支援を目的として、データ解析、知識工学の両側面から研究を行なったものである。

本論文では、まずデータ解析における作業過程の細分を試み、対象手法を限定しない場合の支援が可能な解析作業を明らかにした上で、解析手法列の作成および実行までを支援対象としている。解析作業を行なう処理列を得るため、まず観測データと解析目的から解析作業の中心となる手法を定め、データがそれに適した入力となるよう解析処理の部分列を作成する。そのためにここでは解析処理を操作子として問題解決を行なう枠組を提案している。問題解決などの知識工学的手法においては定性的な知識の記述が必要とされることが多く、定量的情報を取り扱うデータ解析作業において何らかの関係を導入する必要があるため、本論文ではデータの形状や項目数などを表す Metadata を定性的記述のために用いている。

データ解析に関するエキスパートシステムに関する従来の研究では推論機構と知識ベースが分離されていないものが多くエキスパートシステムの構成として問題とされていたが、本論文では推論機構と独立した知識ベースを作成し、実行例によりその有効性を確認している。

さらに本論文では、解析処理列を実行する際、指標値などから得られた結果が不適切である場合の再実行について、データ解析作業が定量的情報を主とすることを踏まえて推論方式に関する検討を加え、状況依存後戻り推論方式の導入を行なっている。推論機構としては非単調推論の一実現形式である TMS(Truth Maintenance

System) を導入し、柔軟に処理を再実行するための枠組と実行例が述べられている。また、TMS による推論機構においても先の知識ベースが利用可能であることを示し、推論機構との独立性を確認している。

これを要するに、著者は、データ解析の支援システムに関する知識ベースの構築および推論方式の導入に対して有益な新知見を得たものであり、情報工学の発展に貢献するところ大なるものがある。

よって著者は、北海道大学博士（工学）の学位を授与される資格あるものと認める。