

ハイパーツールを用いた統計解析コンサルテーション システムの構築

川崎医科大学 数学教室 林 篤 裕

1. はじめに

統計解析プログラムの開発・普及に伴って、これらプログラムが利用される頻度が増大し、多くの利用者を獲得するに至った。しかし、統計学の基礎知識が不足している利用者の中には、統計手法の不適切な使用や明らかな誤用を行っている例もみられるようである。

このような状況を解決する方法として、統計解析用のエキスパートシステムが望まれるようになり、研究・開発が行われている。

人工知能 (AI) の分野では、知識の表現モデルとしていくつかの物が知られているが、その中でも意味ネットワーク・モデルは、定義付けのはっきりしている階層的な問題に対して有効であり、知識の追加・修正も容易に行える。

そこで我々は、この意味ネットワーク・モデルをヒントに、統計の解析手順 (ストラテジと呼ぶ) をネットワークであらわしたストラテジ・マップを用いて統計解析のコンサルテーションを行うシステムを構築した。このシステムは利用者と対話する事によって、利用者の解析目的に適する解析手法を助言するもので、統計手法を熟知していないデータ解析者や統計学の手法を勉強中の者に対して有用であると思われる。

2. 意味ネットワーク・モデル

人工知能の分野では、知識の獲得・格納・利用を計算機に行わせ、より人間に近い判断や行動を計算機にさせる研究が進められている。

知識の格納に関する研究では、知識の表現モデルとしては、フレーム・モデル、ルール・モデル、論理モデル等が知られている。また、長期記憶の心理学的な構造モデルとして、意味ネットワーク・モデルが考えられた (Quillian (1968))。

このモデルは、ノードとリンクにより構成されており、前者で概念対象を、後者で概念対象間の関係を表し、ネットワーク全体としてある特定の意味を持つ知識体系を表現する方法である。意味ネットワーク・モデルの特徴を列挙すると以下のようなものである。

- ・構成要素はノードとリンクである
- ・リンクにはノード間の関係を示す意味ラベルがつけられている
- ・目的の (現在考えている) ノードを探して、そこから (へ)、出て (入って) いるリンクをたどる事で知識を利用できる
- ・2つのノード間に張られているリンク (複数) を順にたどる事で両者の関係がわかる
- ・意味ネットワークの利点は、
 - ・意味体系がはっきりと定義付けられているもの (宣言的知識) に対して有効
 - ・階層性・継承といった性質を表現することができる
 - ・直感的でわかりやすい
 - ・知識の追加・変更が視覚的に行える
- ・意味ネットワークの欠点は、
 - ・このような構造を本格的に使うのは難しい (定義が明確でない時)

- ・必要事項を漏れなく表現する事が難しい
- ・システムが大きくなると探索時間がかかる
- ・推論結果の妥当性の保証が問題

3. 統計解析コンサルテーションシステムへの応用

統計解析に関するコンサルテーションシステムに要求されるものとして、

- ・解析手順(ストラテジ)の書き換えの容易さ

が要求される。つまり、統計解析の技術は時間と共に改良されていくものであり、また個々の解析専門家によっても解析手順が異なる。この様な変化に対応できる柔軟なコンサルテーションシステムでなければならない。

一方、統計解析の基本作業部分に関する知識について考えてみると、

- ・定義付けのはっきりしている部分が多い
- ・階層的な解析手順である

といった性質がある。

これらの特性を考慮して、統計解析の知識表現に意味ネットワーク・モデルをモディファイしたストラテジ・マップが有効であると考えた。すなわち、意味ネットワークのノードに、「統計手法の特徴」、「統計用語」、「統計手法のプログラム」を割り当て、それらにそれぞれ「特徴の説明」、「用語の説明」、「プログラムの実行」を意味するリンクを張る事により、ノード間の関係を関連づけて、統計解析用のストラテジ・マップを構築する。

ストラテジ・マップを構築するにあたって、

- ・解析手法の分類…前提条件、対象となるデータの特徴
- ・データの分類……データのタイプ、データの採取分野

の2点を考慮した。ストラテジ・マップの例を Fig.1 に示す。

このコンサルテーションシステムの利用者は、システムから出される質問に答えることに

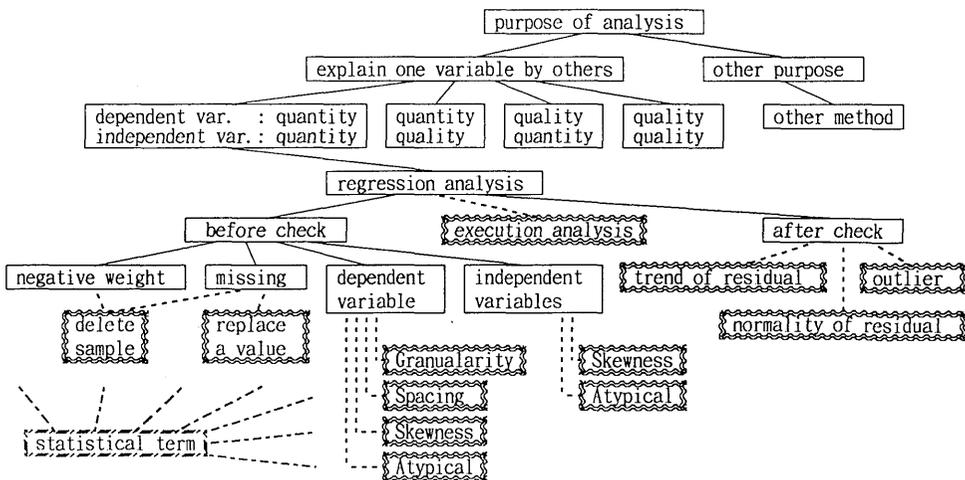


Fig.1. ストラテジ・マップの例.

よって、ストラテジ・マップをたどることができ、最終的に解析目的に適した統計手法を用いて解析を行うことが可能となる。

4. インプリメンテーション

系統立った知識を計算機に組み込んで利用することのできるツールとして、「HyperLink (発売: ウェブトレイン)」というソフトウェアを用いた。つまり、作成したストラテジ・マップを基に、階層的な統計解析手順を互いに関係付け、参照できるようにした。また、統計解析プログラムについては我々が開発したものを修正して組み込んだ。これらの作業の結果、統計解析コンサルテーションシステムを構築することができた。

5. ま と め

統計解析の知識体系は、階層的な関係が定義されており、また追加・修正が行われる頻度も高い。ストラテジ・マップは視覚的に捕らえる事ができ、また柔軟に構築できる点からも、今回の問題に適していると思われる。

今後の課題としては、システムが提示した手順に対して、なぜその様な決定に至ったかを説明する機能や、統計プログラムが計算した統計量に対する解釈方法に対して、どの様にシステムに取り入れていくかが残されている。

いずれにしても、試作と試用を繰り返して、本システムを鍛えていく必要があると思われる。

参 考 文 献

- Affi, A.A. and Clark, Virginia (1990). *Computer-aided Multivariate Analysis*, 2nd ed., Van Nostrand, Princeton, New Jersey.
- Chabris, Christopher F. (1989). *Artificial Intelligence and Turbo C*, Multiscience Press, New York (岩谷 宏 訳, 『Turbo Cによる人工知能』, 工学社, 東京).
- Gale, W.A. (1986). *Artificial Intelligence and Statistics*, Addison Wesley, Reading, Massachusetts.
- Gebhardt, F. (1990). *An Expert System Strategy for Selecting Interesting Results*, Proceedings of Compstat 90, Physica-Verlag, Heidelberg.
- 林 篤裕, 垂水共之 (1990). データ解析における手法選択のコンサルテーションシステム, 統計エキスパートシステム研究会配布資料.
- HyperLink ユーザーズマニュアル (1990). ウェブトレイン, 東京.
- Quillian, M.R. (1969). The teachable language comprehender, *Comm. ACM*, **12**, 459-476.
- 垂水共之, 林 篤裕 (1988). 『統計解析ソフトウェア Seto/B』, 共立出版, 東京.
- 上野春樹, 石塚 満 (1987). 『知識の表現と利用』, オーム社, 東京.
- Van den Berg, G.M. and Visser, R.A. (1990). *Knowledge Modelling for Statistical Consultation System; Two Empirical Studies*, Proceedings of Compstat 90, Physica-Verlag, Heidelberg.

データ解析チューターシステムの試作を終えて

北海道教育大学 函館分校 宇田川 拓 雄

1. はじめに

コンピュータを用いたデータ解析は、以前はコンピュータ言語によってデータ処理プログラムを自作して行っていた。従って統計学とコンピュータの知識を持った少数の人だけがそれを