

医療データベースと統計パッケージの インターフェイスについて

開原 成允 (東大医・病院情報処理部)

前者は、病歴データの構造上の特徴及びその処理に要求される機能が、一般のデータベースのデータとは、かなり異なっていることを様々な機会に述べてきた。従って、この病歴データを扱うようなシステムは、データベース管理システムという情報処理技術の側から打てる、かなり興味ある問題を含んでいると考えられる。

病歴のデータには、現在、通院中又は入院中の病歴データと、すでに退院して、記録として残されている病歴データの二種類があり、この両者は、必要とされる処理形態が大きく異なっている。

この違いの概略は、表1のようなものである。すなわち、現在通院中の病歴データの処理として必要なのは、通常検索が主であり、その検索も、あらかじめ想定できるものである。

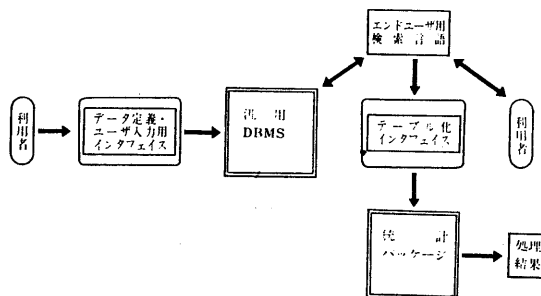
これに対し、後者のデータの処理は主として、臨床研究用であり、検索と共に統計解析の必要とすることが必須条件である。検索や統計解析は、あらかじめその内容を知ることは不可能で、*try and error*で検索や解析が進められていく。

さて、今日、二つの問題となっているのは、統計解析用プログラムの問題であるが、これは、後者の病歴データに関する問題と考えられる。それは、前者のデータに対しては、統計解析は通常必要ないからである。しかし、統計解析プログラムを論じる前に、医療上重要なのは、統計解析用プログラムが利用できるのに、病歴データを前処理する過程であって、この過程が、実際には一番難しい。

HOSPITAL DATABASE AND MEDICAL RECORD DATABASE

	OBJECTIVES	TYPE OF PROCESSING	CONTENTS OF PROCESSING	TIME FOR PROCESSING	NUMBER OF DATA
HOSPITAL DATA BASE	DAILY OPERATION OF THE HOSPITAL	SIMPLE PROCESSING MAINLY RETRIEVAL	PREDETERMINED	REAL TIME	LIMITED BY THE NUMBER OF THE PATIENTS OF THE HOSPITAL
MEDICAL RECORD DATA BASE	CLINICAL RESEARCH	COMBINATION OF VARIOUS PROCESSING	REQUIRED PROCESSING CANNOT BE PREDETERMINED	MAY WAIT FOR DAYS	UNLIMITED

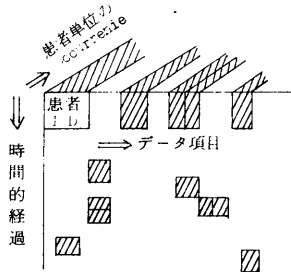
表 1



汎用医療データ処理システム

271

統計解析プログラムが利用できないまでデータが整理できれば、その解析の半分以上が終わったといっても言い過ぎではない。



患者に関するデータのひとつの特徴は時間軸を含む三次元の構造をもつことである。

図2

データベース管理システムだけあって、また、統計解析システムだけあってもちろん充分な機能が必要なのである。

そして、データベース管理システムと統計解析プログラムの間のインターフェイスの問題こそ、医学特有の問題として研究すべきであると思う。

ここで、病歴データの特徴にも一言触れておきたい。これは、演者がすでに何回も強調した所々をくり返しているが、才一は三次元のデータであることで、才二は空欄の著しく多いデータであることがその大きな特徴である。これをくり返しのべてきたので、ここでは図を示すに留める。(図2)

以下に、この病歴データの特徴を考えたつ、汎用データベースシステムのどれだけの部分について考察を進めたい。

a) データベース管理システム

今日、様々のデータベース管理システムがあり、ある程度の規模のコンピュータであれば、これを利用することが出来る。医療データにとって重要なのは、先に述べた三次元のデータが容易に検索できることと、検索されたデータを次の解析のためにファイル中に書き込む機能が必要である。これについては、IBM、ADABAS等を使用してしたが、いずれも、満足すべきものではなかった。

問題は、データベース管理システムそのものよりも、その周辺は、利用者とのインターフェイスのプログラムをどの程度もっているかによって、その使い易さが決まる。

b) 統計解析システム

これについては、今日、他の演者が詳細に述べると思われるので、ここでは述べない。

c) インターフェイス

問題は、前二者をつなぐインターフェイスである。すでに、演者が発表したように、統計解析システムは、二次元のデータしか扱えないから、三次元のデータを2

次元の意味から、ここでは、病歴データを統計プログラムを利用できるような形に変換する過程について考えてみたい。

演者は、このようなシステムに備えるべき機能として、図1のようなシステムを提案してきた。すなわち、医療データの解析とは、

図3

元のデータ						
case number	name	age	sex	date of visit	blood pressure	urine protein
1	Taro TIDAM	50	M	1977. 1. 15	160	-
				1977. 2. 15	180	-
				1977. 3. 1	130	+
2	Jiro OSAKA	46	M	1978. 2. 15	130	-
				1978. 4. 20	140	-
3	Hanako HESOSHIMA	20	F	1978. 4. 21	120	-

case number	age	sex	blood pressure	urine protein
1	50	M	177	+
2	46	M	135	-
3	20	F	120	-

case number	age	sex	blood pressure	urine protein
1	50	M	160	+
1	50	M	180	-
1	50	M	190	+
2	46	M	130	-
2	46	M	140	-
3	20	F	120	-

次元に変換する機能が必要である。

この変換の方法については、図に示るごとく、2つの方法がある。(図3-a,b)
すなわち、第1は、時間毎のくり返しを、全て、1つのデータと見做す場合で、
この場合は、患者個人の属性のデータは、くり返し出現することになる。

第2の場合は、時間的にくり返しを何らかの方法で1つにまとめ、患者単位の
データにまとめることである。

例えば、図でいえば、血圧は平均値がとられているし、尿蛋白は、1つでよい
があったものは1にするという約束の下に変換されている。

このように、時間的にくり返し出現するようデータは、どうやって1つにま
とめるかは、そのデータの医学的性質によって定まるもので、これは解析者自身
が選択する必要がある。

ここで、システムを設計する側から興味あることは、この場合どのような *operation*
を用意しておけばよいかという問題である。

われわれが、これを約10年間に扱ったデータの中からどのような *operation* が
必要であるかを考えてみると、次のようになる。

1. 平均値によって代表させる。
最も普通の場合である。
2. +, -, 一のようデータは、1つでよいがあればとする。これは、
例えば、検査法の感度が悪く、仲良、陽性とならないが、一度陽性とな
れば大きな意味をもつような場合である。例えば、LE細胞等がこれに
あたる。
3. 最大値又は最小値によって代表させる。例えば、肝炎のGOTの検査の
ように、その最高値に意味がある場合もある。最小値は、例えば、血液
疾患におけるHb、白血球数等にその例がみられる。
4. 最大値、最小値の差をもって代表させる。
理論上考えられるが、われわれは経験がない。
5. 時間的に表われた数値の総和をとる。
葉の総投与量のような場合である。
6. 時間軸を横軸としてグラフを著した時のカーブの下の面積。
薬剤の排泄量のようなもの。
7. 増加又は減少率。
典型的なものはいない。
8. 標準偏差、又は変動係数
精度管理のデータによくあらわれる。
9. 最初のデータ。
初診時のデータを重視するような場合。
10. 最後のデータ。
最新のデータを重視するような場合。
11. ある時期を定めて、それに近い時間のデータをとる場合。
12. データの存在している個数を数える場合。
測定回数のようなものを数えたとき。
13. データの値がある範囲内にあった期間。
例えば、肝炎でGOTが異常な高値を示した期間。

これからの理解されるように、パラメータとしては、測定値を変換したものである場合も、また、時間がパラメータに付く場合もあり、これを一般化しようとするとは、かなり難しい問題を含んでいる。

以上が最もよく表われる operation であるが、その他にも、非常に特殊で一般化しにくいようなものも沢山あった。

臨床家にとって、もっとも重要なことは、このような時間的経過をグラフの形で提示できる機能で、これを眺める間に、このデータの力の性格を知り、代表させるべき値を発見するようになる場合も多くあった。

さて、以上のような機能をあつた、汎用システムを実際に作り得たとしたら、それは医療関係者にとって、非常に便利なものとなるだろう。

データベースと統計プログラムが連絡したようなプログラムが、最近、種々開発されていることは、大変喜ばしいことであるが、これらは、まだ、SPSS とデータベース管理システムを結合したのみで、多次元のデータの扱いに充分な配慮が払われていないように思われる。

今後、このような機能を含めることが重要になると考えられ、われわれもその開発を進めている所である。

なお、詳細な議論は、又取1のオマケと参照されたい。

文 献

- 1) 岡原成允、稲田毅、医療情報学 情報処理学会(東京) 1980
- 2) 岡原成允他 汎用医療データベースシステムへのアプローチ
第18回日本エム・イー学会大会(東京) 1979
- 3) 上野 晴樹他 医学研究用データベースPODのユティリティ機能
第17回日本エム・イー学会大会(札幌) 1978