

検査データを用いたデータ解析手法の提案

高知大学理学部数理情報科学科情報科学コース村岡研究室

山下 真 村岡 道明

1. はじめに

医療情報の電子化に伴い、診療情報、検査結果などが長期間にわたり蓄積されつつあり、最近では蓄積された情報の有効活用が求められている。本研究では医療情報の有効活用として投薬が患者に与える影響を調査する手法を提案し、提案手法を用いて抗うつ薬 SSRI が患者に与える影響を導出する。

2. 提案手法

本研究では投薬前と投薬中の患者の血液検査データを比較し、投薬による影響を導出する。提案手法ではまず、決定木学習を用いて投薬前と投薬中で主に異なる検査項目を導出する。本研究ではこれを主要項目とし、関連性分析により主要項目と共に異常値をとる項目を導出する。本研究ではこれを関連項目とする。関連性分析で導出されたルールの有用性の指標として、confidence が 0.9 以上、lift 値が 1 以上をとるもので、決定木学習で導出された主要項目が発生した場合のルールであることとした。以上の手順で導出した主要項目、関連項目から副作用を考察する。

3. 解析システム

本研究では様々な医療機関から対象となるデータを収集し、対話的に解析を行うために図 3.1 のようなシステムを提案した。次章ではこのシステムを用いて SSRI の解析を行う。

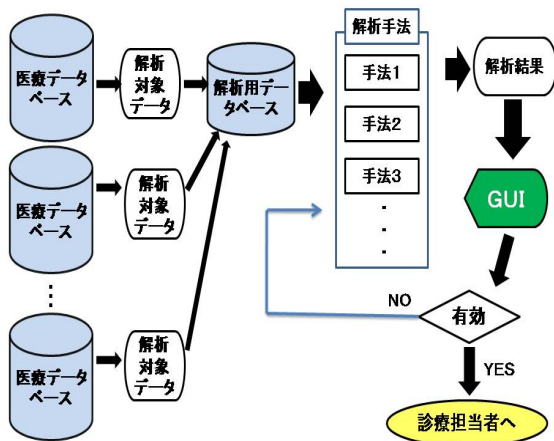


図 3.1 システム構成図

4. SSRI の解析

提案手法を用いて抗うつ薬 SSRI の一種、デプロメールの解析を行う。対象データはデプロメールの投薬を行った患者 54 人の血液検査データ 130 件であり、1 検査を 1 件としている。本来ならば患者ごとのデータの重みを揃えるために、データの統一を行う必要があるが、データ数の問題から今回はデータの統一を行わずに解析を進める。

決定木学習により生成した決定木を図 4.1 に示す。

```
LDH <= 326 :
| GLB <= 2.3 : 投薬前 (10-2/5-1)
| GLB > 2.3 : 投薬中 (41-2/18-9)
LDH > 326 :
| GOT_AST <= 49 : 投薬前 (61-7/16-2)
| GOT_AST > 49 : 投薬中 (17-0/8-6)
```

図 4.1 決定木学習結果

図 4.1 より投薬中は GOT が増加する傾向が見られる。LDH や GLB は基準値内の分岐であったため投薬による変化として扱わなかった。さらに有効な結果を出すために対象データの欠損値を 1 ヶ月以内の値で補完を行った。1 ヶ月以内に値が存在しなかった場合は欠損値として扱う。補間後のデータを対象として決定木学習を行った結果を図 4.2 に示す。

```
S_CL <= 103 : 投薬前 (38-6/11-8)
S_CL > 103 :
| TBIL > 0.7 : 投薬中 (8-2/2-7)
| TBIL <= 0.7 :
| | CRP > 0.2 : 投薬前 (16-7/6-2)
| | CRP <= 0.2 :
| | | GLB <= 2.3 : 投薬前 (13-1/5-9)
| | | GLB > 2.3 :
| | | | S_CL <= 104 : 投薬前 (6-6/3-0)
| | | | S_CL > 104 : 投薬中 (46-7/24-0)
```

図 4.2 補間後決定木学習結果

図 4.2 より投薬中は S_CL が増加する傾向が見られる。また、CRP が減少する傾向も見られる。TBIL, GLB に関しては基準値の範囲内での分岐であり、件数も少ないため投薬による変化として扱わなかった。

図 4.1, 4.2 の決定木学習結果より得られた GOT, S_CL, CRP を主要項目として扱い、投薬により異常値をとると思われる GOT, S_CL に関するルールを関連性分析で導出する。

分析の結果、ルールの有用性の指標に達するものは 7 件であった。有効なルールからは、GOT, S_CL が増加することで GPT が増加する。GOT, S_CL, GPT が増加することで LDH が増加するという傾向がみられた。

以上の解析で、投薬により GOT, S_CL, GPT, LDH が増加し、CRP が減少する傾向があると考えられる。これらの検査値は肝機能の指標となる検査値であり、これらが増加していることから投薬を続けると肝機能が悪化すると考えられる。また、CRP の減少も見られる。うつ病患者はストレスのため炎症が進むという仮説があり、デプロメールによってうつ病が回復と共に CRP が減少するのではないかと考えられる。

5. おわりに

今回は患者 54 人分の小規模なデータを解析した。今後はさらに有効な結果を出すために大規模なデータを対象に解析を進める。