

決定木を用いた医療データの解析手法

The Analysis Method of Medical Data Using a Decision Tree

林 佳宏 村岡 道明

高知大学大学院 理学研究科 村岡研究室

1. 研究背景

近年、大規模病院では電子カルテの導入が進められるなど医療データ(診療データや検査データ)の電子化が進み、長期間にわたり大量の検査結果や診療情報が収集・蓄積されてきた。しかしながら、それらのデータは、人手で分析するためには時間も労力もかかってしまうために、十分に活用ができていないといえず、その活用が今後の課題となっている。そのため、人手で行っていた作業を解析の自動化をすることにより、時間と労力を削減して、所望の分析結果を得ることが求められている。

2. 研究目的

本研究では、情報科学分野で使用されているデータベース技術やデータ解析技術を医療データに適用することにより、精神科学の領域において、従来では見逃されていたルールや新たな知見を発見するとともに、そのための解析手法を提案することを目的とする。また、本研究ではうつ病患者に投薬される抗うつ薬SSRIの一種であるデプロメールの薬効についてデータマイニングを行い解析する。

3. 解析の前処理

医療データには、欠損値が大量に存在する、患者毎の検査回数や検査間隔が異なるなどの特徴がある。そこで、患者一人毎の欠損値に対して、その日から過去一ヶ月以内の検査日までさかのぼり、一番近い検査日の値でデータを補間した。一ヶ月以内の値で補間できない場合は、NULL値としてそのデータを扱う。図1に補間の例を示す。

項目名	19991015	19991107	19991218	補間値
TP	6.0	NULL	6.5	6.5
GLU	NULL	104	120	104
ALP	153	187	NULL	NULL

図1 欠損値の補間例

また、データ統一として、各患者の投薬が開始された直前の日の検査データと投薬が終了した直前の検査データのみを抽出するという処理を行った。

4. 解析対象

解析対象のデータは、抗うつ薬SSRIの一種であるデプロメール(マレイン酸フルボキサミン)を投薬したときの検査データ(高知大学医学部精神科学教室所有)とする。また、対象データは1検査当たりを1レコードとし、130レコード、31属性であった。その内訳としては、投薬前のものが78レコード、投薬後のものが52レコードであった。

5. 解析結果

欠損値の補間を行ったデータに対して、データ統一前後の解析を行った。図2と図3にそれぞれの解析結果を示す。



図2 データ統一前の解析結果



図3 データ統一後の解析結果

データ統一前は、UAが増加し、CRPが減少したのに対して、データ統一後は、CPK(CK)が増加し、CRPが減少するという違いが見られた。

6. まとめ

データ統一前後ともに、CRP(蛋白の一つであり、炎症性疾患の存在や病変の活動性、障害程度を鋭敏に反映する代表的な炎症マーカー)という検査項目が抽出されたので、評価できる結果であると考えられる。また、うつ病患者はストレスで炎症が酷くなると最近になって指摘がされており、医学的な観点からも有効な結果であると考えられる。

参考文献

[1] 津本周作, 平野章二, “複合医工学としてのデータマイニング”, 人工知能学会誌, Vol. 22 No. 2, pp. 201-207, 3月, 2007年