

29【B】1120

疫学データの解析における、決定木法への計算機集約型検定法の導入 A New Procedure of Validation for Binary Tree Models using Resampling Technique and Its Application to Epidemiology

○高木 達也^{1,2}, 岡本 晃典¹(¹阪大院薬,²阪大遺伝情報実験センター)

「ポスト IT」時代に入って、「データマイニング」と呼ばれる分野が盛んになってきた。相関ルール、SOM、SVM などの手法が中でも使われ始めてきたが、決定木法もその一つで、一定の予測変数から結果の予測が必要な分野で用いられている。しかしながら、医薬学分野では、各属性の結果に及ぼす決定力に関する情報…回帰分析における回帰係数の検定に相当するような…が乏しいことが、主要な問題点となり、普及できていない。例えば、ある疾患に対する複数の危険因子が、その疾患の罹患に対してどの程度の影響力を持っているのか、決定木法では、それら複数の危険因子が、どのように介在した場合に、どの程度の罹患の可能性があるかは明らかにすることができるが、その危険因子が「どの程度危険なのか」に対する情報は豊富ではない。

今回、我々は、決定木手法において特定の属性を削除した場合に及ぼす、結果に対する影響を、標本再抽出法の概念を応用することによって数値化し、各属性が結果に対して及ぼす決定力の検証が可能となる新規アルゴリズムを提案した。そして、そのアルゴリズムを人工データ、および脳腫瘍の疫学データへ適用することにより、手法としての有効性の検討を行った。その結果、人工データの結果により、新規アルゴリズムが各属性のクラス判断に対する重要性の検証に関して有益な情報を与えることが判明し、疫学データに対する適用では、妥当な検証結果を得ることができた。