

所属講座	知能数理学	指導教員	下園 真一 准教授
学生番号	05231009	氏名	伊藤 祐一郎
論文題目	規則と失敗事例にもとづく難解な手続きのサポートシステムの開発		

1 はじめに

事務管理業務では、運用のための規則が定められている。業務の従事者は、これを運用し続けるため、規則を修正・維持する必要がある。それを利用者に対し理解を促す作業が重要となる。しかし現実には、規則法規は規模が大きく、複雑かつ難解であり、様々な新制度や技術に対応するため変更も多く、利用者の理解を促す業務は後回しとなりがちである。その結果、運用が困難になり生じた問題を業務改善に生かすことが出来ず、業務効率の低下を引き起こす。

本研究では、従事者が規則や経験から抽象的な事例を作成し、C4.5の情報利得比基準[1]に基づくコンパクトな判別ダイヤグラムを作成する手法を提案する。これにより利用者にとって難解な手続きの内容を理解しやすくし、業務効率の向上に貢献する。

2 決定木と C4.5 の情報利得比基準

本研究では、分類モデルを「決定木 (Decision Tree)」[2] とよばれる形式で表現する手法を用いる。構造として、親ノードはテスト、子ノードはテストに対する答え、葉は各々クラス名を持つ。構造を決める重要な点は、どのテストから行えば、木の高さが小さくなるかということで、そのテストの選択に用いるのが「情報利得比基準」である。これは、各属性をテストとして選んだときの情報量を計算し、テスト結果を伝えるメッセージの情報量について利得が最大になるものを選ぶ、という基準である[1]。

3 抽象事例設定から C4.5 決定木を作成

抽象事例とは、一部の属性値が任意である事例である。任意の属性に属性値を代入することで、全属性値が定まっている事例を得られる。これを具体事例という。まず抽象事例の作成方法は、

- 特に重視したい属性だけ値を代入した Yes, No の事例を作る。値を代入していない属性にはすべて '*' (任意) が入る。
- クラスの異なる抽象事例から、全属性値の同じ具体事例が得られる場合は、適切な属性値を代入し、矛盾を解消する。

となる。次に具体事例の作成で、属性は、 $t_0=(OK, NG)$, $t_1=(1, 2)$ とし、クラスは、 $C=(yes, no)$ をとる。表 1 の

ような抽象事例を与えると、 $t_0="OK"$ となる場合、 C は "yes" になる。また、 $t_1="2"$ となる場合、 C は "no" となるが、矛盾が起きないように $t_0="NG"$ を代入している。この抽象事例から具体事例を生成すると、表 2 のような 3 つの事例が生成されることになる。

表 1: 抽象事例

t_0	t_1	C
OK	*	yes
NG	2	no

表 2: 具体事例

t_0	t_1	C
OK	1	yes
OK	2	yes
NG	2	no

4 実験と考察

授業料免除申請の受付を例として、抽象事例が出力された決定木に表れるか調べる。Yes 事例を 2 つ、No 事例を $t_{3,4}$ について値を入れ、5 つの抽象事例から作成した結果が図 1 である。図 1 を見ると、抽象事例で与えたテストはより上位の質問になりやすいといえる。よって、抽象事例を作成した人の意図を反映した決定木を作成できるといえる。

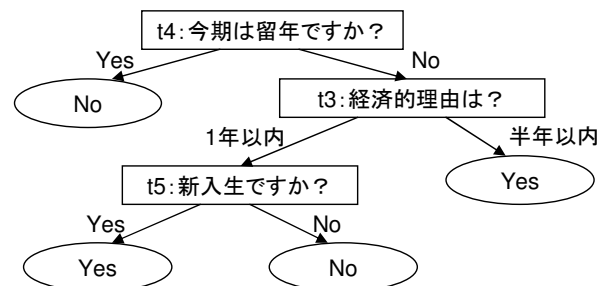


図 1: 授業料免除申請の受付判定木

5 まとめ

本研究では、情報利得比基準を用いたテスト選択によって、与えた抽象事例を反映させた決定木を作成するシステムの開発を行った。実験により、その有用性も確かめられた。ただし、情報利得比基準では一つのテストの利得のみで、その部分木の大きさや重複は考えていない。よりコンパクトな図にするには、テストの組合せによるテストも考慮するべきであろう。

参考文献

- [1] J.R.Quinlan 著, 古川 康一 訳, AI によるデータ解析 (初版), 株式会社トッパン, (1995).
- [2] T. Menzies, Y. Hu, Data Mining For Very Busy People. IEEE Computer, pp. 18-25, (2003).