

目 次

第 1 章	はじめに	2
第 2 章	研究の目的	3
第 3 章	具体事例と抽象事例	4
3.1	属性と属性値	4
3.2	抽象事例と抽象事例による規則・知識の表現	5
3.3	抽象事例から具体事例への展開	6
第 4 章	決定木と C4.5 の情報利得比基準	8
4.1	決定木	8
4.2	テスト評価	9
4.2.1	利得基準	9
4.2.2	情報利得比基準	11
第 5 章	実験	13
5.1	実験準備	13
5.2	実験方法と評価基準	15
5.3	実験結果	17
第 6 章	考察	21
第 7 章	おわりに	23

第1章 はじめに

近年，事務管理業務においては，様々な新制度や新技術に対応する必要から変更や修正に伴った，大規模化や複雑化が進んでおり，その規則の従事者も利用者も理解に苦しむといった状況が生じる．業務の従事者が規則をきちんと理解していない場合，現状把握が出来ず業務改善が行われなままになるので，業務効率が低下につながる．また従事者の理解している規則を他人に伝える方法に問題があると，従事者が意図して書いた規則と違う意味で利用者は理解してしまい，手続き際に間違いを引き起こすといった問題が発生する．

この問題点は，従事者が規則を正しく理解し，規則を他者へ正確に伝えることが出来れば，解消すると考える．その方法として，規則を文章ではなく，図で表現する．従事者が経験的に持つ規則の重要部分を抽出し，規則の現状や規則中の手続きの手順を段階的な流れ図として表現することで，業務改善や規則を説明する際の補助に利用できる．

本研究では，業務の従事者が経験の中でつちかった知識や経験を視覚的に捉える，業務改善や現状把握に利用できるシステムの開発を行った．本システムを利用すれば，文章で表現された規則を，利用者が目で見て理解することが出来ると考える．

本論文では，まず第2章で研究の目的を述べる．第3章でシステムの入力となる抽象的な事例から具体的な事例の作成に至るまでの過程を説明し，第4章で，本システムの出力となる決定木の概念や構造と木構造を作成するためのC4.5についてを述べる．第5章で，そのシステムを用いて実際にある業務例を分析し，その考察を第6章で述べる．最後に第7章で本論文をまとめる．

第2章 研究の目的

本研究では、従事者が規則に関する経験や知識から重要なポイントや注意点を表現した抽象的な事例を作り、そこから具体的な事例を作成する。その事例を J.R.Quinlan らによる C4.5[1] に基づいて事例を分別し「決定木」という分類モデルを使って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。

このシステムにより業務の熟練者が持っている目には見えない経験や知識を形あるもので表現できるため、従事者にとっては現状把握や業務分析に利用でき、利用者にとっては難解な手続きの内容を理解しやすくするためのツールとして、全体的な業務効率の向上に貢献することを目的とする。

第3章 具体事例と抽象事例

この章では、規則の経験的な知識から特徴を抽出した抽象事例、その抽象事例を具体化することで生成される具体事例の2つの概念について説明する。

3.1 属性と属性値

まず具体事例、抽象事例を説明するにあたり、各事例が持っている「属性、属性値、クラス」について、ある授業の単位認定に関する規則を例に説明しよう。

「属性、属性値、クラス」とは、C4.5[1] で用いている用語である。それぞれを日常的な言葉で言い換えると、属性とは質問、属性値とは質問に対する答え、クラスとは最終的な結果のことである。表 3.1 の例の場合、属性 1 に対しての属性値は、[2/3 未満, 2/3 以上] の 2 通りがあり、属性 2 に対しての属性値は、[不提出, 提出] の 2 通り、属性 3 に対しての属性値は、[60 未満, 60 以上] の 2 通りがある。そしてクラスは、[合格, 不合格] の 2 通りがある。

表 3.1 単位認定の属性と属性値とクラスの関係

属性 1	出席 = [2/3 未満, 2/3 以上]
属性 2	課題 = [不提出, 提出]
属性 3	試験 = [60 未満, 60 以上]
クラス	合否 = [合格, 不合格]

3.2 抽象事例と抽象事例による規則・知識の表現

まずはじめに，具体事例と抽象事例の定義について説明する．具体事例とは，全属性値が既知の事例のことをいい，抽象事例とは，一部の属性値が未定になっている事例のことをいう．未定であることを'*'で表す．この2つの事例はどのようなものを前節の授業の単位認定の例を使い説明する．この授業の単位認定の基準には，合格と不合格を決める規則があるとする．その規則を表3.2に表す．

表 3.2 単位認定の規則

特徴 1	出席が 2/3 以上 and 課題が提出 and 60 以上	合格
特徴 2	出席が 2/3 以上 and 課題が提出 and 60 未満	不合格
特徴 3	出席が 2/3 以上 and 課題が不提出	不合格
特徴 4	出席が 2/3 未満	不合格

表 3.2 を経験的な規則として簡易表現するならば，その際の手順として，まずは属性を定めなければならない．この場合，属性は前節で述べた通り「出席，課題，試験」である．よって，それぞれの属性に対し，表 3.2 で分かっている属性値を代入していく．その際，値が不明または未定の属性値については，'*'を使用する．表 3.2 を簡易表現したものを表 3.3 に表す．

表 3.3 表 3.2 の簡易表現

	出席	課題	試験	クラス
事例 1	2/3 以上	提出	60 以上	合格
事例 2	2/3 以上	提出	60 未満	不合格
事例 3	2/3 以上	不提出	*	不合格
事例 4	2/3 未満	*	*	不合格

先ほど述べた具体事例と抽象事例の定義からすれば，表 3.3 の事例 1 と事例 2 は具体事例であり，事例 3 と事例 4 は抽象事例である．また表 3.2 の規則の特徴から，表 3.1 を用いて，表 3.3 を作成するまでの過程のことを抽象事例の設定という．

3.3 抽象事例から具体事例への展開

規則の特徴を表した抽象事例では，C4.5 を使って決定木を作成することは出来ない．これは C4.5 が属性値は未定な事例を扱えないからであり，そこで抽象事例から具体事例へと展開する必要がある．先ほどの授業の単位認定の例で作成した抽象事例，表 3.3 を使い，抽象事例から具体事例の展開方法について説明する．

表 3.3 の事例の中で，事例 1 と事例 2 は，すでに全属性値が既知であるため展開する必要はない．残りの事例 3 と事例 4 の抽象事例について考えていく．まず事例 3 は，試験の属性が”*”となっている．試験の属性値は [60 未満，60 以上] を取るので，これを展開すると表 3.4 となる．

表 3.4 事例 3 の試験について展開

出席	課題	試験	クラス
2/3 以上	提出	60 以上	合格
2/3 以上	提出	60 未満	不合格
2/3 以上	不提出	[60 以上]	不合格
2/3 以上	不提出	[60 未満]	不合格
2/3 未満	*	*	不合格

このように事例 3 からは，2 つの具体事例が作成されました．この作業を全抽象事例について”*”がなくなるまで繰り返すことにより，抽象事例を具体事例に展開することができます．そして事例 4 についても同様の展開を行い，全抽象事例を具体事例に展開したものが表 3.5 である．

表 3.5 具体事例

出席	課題	試験	クラス
2/3 以上	提出	60 以上	合格
2/3 以上	提出	60 未満	不合格
2/3 以上	不提出	60 以上	不合格
2/3 以上	不提出	60 未満	不合格
2/3 未満	提出	60 以上	不合格
2/3 未満	提出	60 未満	不合格
2/3 未満	不提出	60 以上	不合格
2/3 未満	不提出	60 未満	不合格

表 3.4 から表 3.5 へは，まず事例 4 の課題の”*”を展開します．すると試験の属性値が”*”の抽象事例が 2 つ作成されます．次に残った試験の”*”を展開することで 4 つの具体事例が作成され，すべての抽象事例を具体事例に展開することができました．

第4章 決定木とC4.5の情報利得比基準

この章では、本手法に用いる決定木 (Decision Tree) の構造とクラス分けに用いている C4.5 の情報利得比基準における分類手法について説明する。

4.1 決定木

本研究で使用した J.R.Quinlan らによる C4.5 は、情報利得比基準に基づく質問の選択により、コンパクトな判別ダイヤグラム、すなわち事例のクラス分けを行う決定木を作るためのヒューリスティックである。これは理論的に保証されることは多くないが、実際の用途には広く使われ高く評価されている。その分類モデルとして用いている「決定木 (Decision Tree)」[2] とよばれる形式で表現する手法は、データマイニングで良く用いられる方法でもある。その場合、決定木は、葉が分類を表し、枝がその分類に至るまでの特徴の集まりを表すような木構造を示す。

本システムの中で構造は、各分岐点または葉 (端点) のことをノードといい、特に親子関係になっているノードのことをそれぞれ親ノード、子ノードという。親ノードは質問を持っており、その答えの数だけ子ノードを持つ。子ノードは質問に対する答えを持っていて、特に分類された特徴の集まりが単一のクラスとなる場合、その子ノードを葉という。葉はクラス名を持っており、クラス名は決定木によって下された最終的な答えとなる。

4.2 テスト評価

決定木の構造は、質問の選び方に大きく依存している。そのためただ何でもよいから分割してしまうと、もとの規則と矛盾を生じたり、膨大な大きさの決定木になる可能性がある。それを避けるためにも、出来るだけ小さく簡潔であり、もとの規則と矛盾しない決定木を構成されるように、各段階でもっとも効果的な質問を選ぶ必要がある。

もっとも簡潔な決定木を求める方法として、「すべての可能な木を調べて、その中で最良のものを選ぶ」というものもあるが、矛盾しない最小の決定木を求める問題は、NP 完全であり [Hyafil and Rivest, 1976]、信じられないほど多くの木を調べる必要がある。このことから、C4.5 で用いている局所的に評価をし、その評価結果が最大になるような分割を行う質問を選ぶといった手法が必要になる。

4.2.1 利得基準

事例の集合 T を部分集合 T_1, T_2, \dots, T_n に分割する n 値のテストがあるとする。これらの T_i がのちの部分木を考えず、その質問だけで評価しようとするれば、 T 内の各 T_i 内のクラス分布の情報だけが利用できる。クラス分布についての記法で、 $\text{freq}(C_i, S)$ は S の中でクラス C_i に属する事例数を表す。また集合 S に含まれる事例数を $|S|$ と表す。

質問の評価基準の基礎となる情報理論的な考え方は、「メッセージによって伝えられる情報量は、メッセージの生起確率で決まる」という風に表せる。すなわち、情報量は、確率の値の 2 を底とする対数を取り、それを -1 倍にして計算され、「ビット」単位で表される。したがって、例えば、8 種類の等確率のメッセージであれば、その各々のメッセージによって伝えられる情報量は、 $\log_2(\frac{1}{8})$ ビット、すなわち 3 ビットとなる。

事例の集合 S からランダムに 1 つの事例を選び出し、それがクラス C_j に属してい

ると知らせるとすると，このメッセージ確率は，

$$\frac{freq(C_i, S)}{|S|}$$

であり，それが伝える情報量は，

$$\log_2 \left(\frac{freq(C_i, S)}{|S|} \right) \text{ ビット}$$

となる．このようなクラスの所属関係に関するメッセージの平均情報量を求めるために， S 内での頻度で重み付けしてクラス全体に対する平均を求めると，

$$info(S) = - \sum_{j=1}^k \frac{freq(C_j, S)}{|S|} \times \log_2 \left(\frac{freq(C_j, S)}{|S|} \right) \text{ ビット}$$

を得る．この量は集合 S のエントロピーとも呼ばれる．これを事例の集合 T に適用すれば， $info(T)$ は T 内のある 1 つの事例が属するクラスを同定するのに必要な情報量の平均値となる．

次に，質問 X の n 通りの結果に合わせて T が分割された後について，同様な評価を考えると，クラスを同定するのに必要な情報量の期待値は，部分集合上で荷重平均をとって，

$$info_X(T) = \sum_{j=1}^k \frac{|T_j|}{|T|} \times info(T_j)$$

となる．これらの差

$$gain(X) = info(T) - info_X(T)$$

は、質問 X で T 分割することによって獲得される情報量を表す。この情報量の利得は、質問 X とクラスとの相互情報量とも呼ばれる。これを最大にするように質問を選ぶ基準を、利得基準と呼ぶ。

4.2.2 情報利得比基準

先に述べた利得基準は非常に結果が得られるが、多数の値をとる質問を偏重する欠陥を持っている。例えば、質問として選んだ属性が氏名などの場合、1 事例からなるたくさんの部分集合が得られることとなり、 $info_X(T) = 0$ となる。したがって、この属性を用いて事例の集合を分割すれば、情報量の利得は最大になる。しかしながら、このような分割は全くの無益である。

利得基準に伴うこのような偏重は、一種の正規化によって矯正することができる。すなわち、多数の値を取ることによって得られた利得部分を調整すればよい。ある事例に関して、それがどのクラスに属するかではなく、その質問結果自体を伝えるメッセージの情報量を考える。info(S) の定義からの類推により、分割情報量を

$$split\ info_X(T) = - \sum_{j=1}^k \frac{|T_j|}{|T|} \times \log_2 \left(\frac{|T_j|}{|T|} \right) \text{ ビット}$$

と定める。これは T を n 個の部分集合へ分割することによって得られる全情報量を表す。一方、情報量利得は、そのうちのクラス分けにかかわる部分の情報量を表す。したがって、利得比

$$gain\ ratio(X) = gain(X) / split\ info(X)$$

は、分割によって得られる情報量のうち、有益な部分、すなわち、クラス分類に役

立つ部分の割合を表す．ところで，分割が自明な分割に近いときは，分割情報量の値が小さいために利得比の値が不安定になる．そこで，利得比基準では，全質問中で情報量利得が少なくとも平均以上であるという制約下でこの利得比を最大にする質問を選ぶ．

この評価基準の下では，先ほど述べた欠陥部分は高く評価されない．クラスの数 k とすると上式の分子 (情報量利得) は高々 $\log_2(k)$ となる．一方，事例数を n とすれば，質問結果も n 通りに分割されるので，分母は $\log_2(n)$ となる．ここで事例数 n は，クラス数 k よりずっと大きいので，利得比は小さな値となる．

第5章 実験

この章では、授業料免除制度を例に開発したシステムを用いて、申請可能かどうかを判定する決定木を作成を行う。

5.1 実験準備

決定木を作成する準備段階として、文章の規則から注意点（属性）を決めなければならない。規則については表 5.1 に表す。表 5.1 の規則の中に書かれている条件以外にも、実際に運用される場合の条件がいくつかある。その条件について述べると、

1. 申請者は、申請する学期は休学しない。
2. 申請者は、申請年度は留年した年ではない。
3. 申請者の在学年度は、4 回生以内である。

の 3 つがある。これらの 3 つの条件と表 5.1 を合わせて、この場合どのような属性が必要か考えると、属性と属性値、クラスは表 5.2 のようになる。

表 5.2 免除制度の属性と属性値とクラスの関係

属性 1	休学	[はい, いいえ]
属性 2	回生	[4 以下, 5 以上]
属性 3	理由	[半年以内, 1 年以内, 数年, なし]
属性 4	留年	[はい, いいえ]
属性 5	新入生	[はい, いいえ]
クラス	申請	[申請可, 申請不可]

このように授業料免除制度の申請規則を申請時の重要となる 5 つのポイントを属性とし、その属性値の取り方より、申請可能かを判定することにします。

表 5.1 授業料免除制度に関する規則

<p>(免除の対象及び額:新入生)</p> <p>第 3 条 次の各号に定める事項のいずれかに該当する者は、 入学料の全額又は半額を免除することがある</p> <p>(1) 大学院に入学する者で経済的理由により入学料の納付が困難であり、 かつ、学業優秀と認められる者であること。ただし、前記に該当しない者 であっても、次のいずれかに該当する特別な事情により入学料の納付が 著しく困難であると認められる場合には、免除の対象とすることができる</p> <p>ア 入学前 1 年以内において、入学する者の学資を主として負担している者 (以下「学資負担者」という)が死亡した場合又は入学する者若しくは 入学する者の学資負担者が風水害等の災害を受けた場合</p> <p>イ アに準ずる場合であって、学長が相当と認める事由がある場合</p> <p>(2) 学部に入学者で次のいずれかに該当する特別な事情により 入学料の納付が著しく困難であると認められる場合</p> <p>ア 入学前 1 年以内において、入学する者の学資負担者が死亡した場合 又は入学する者若しくは入学する者の学資負担者が 風水害等の災害を受けた場合</p> <p>イ アに準ずる場合であって、学長が相当と認める事由がある場合</p>	<p>(免除の対象及び額:在学生)</p> <p>第 11 条 学部及び大学院の学生(研究生、聴講生又は科目等履修生として 入学した者を除く。以下「学生」という)で、次の各号の一に 該当する者は、各期ごとの授業料の全額又は半額を免除することがある</p> <p>(1) 経済的理由により授業料の納付が困難であり、 かつ、学業優秀と認められる者</p> <p>(2) 授業料の各期ごとの納期前 6 月以内(新入学者に対する 入学した日の属する期分の免除に係る場合は、 入学前 1 年以内)において、学資負担者が死亡した場合 又は学生若しくは学資負担者が風水害等の災害を受け、 授業料の納付が著しく困難であると認められる者</p> <p>(3) 前号に準ずる場合であって、学長が相当と認める事由がある者</p>
--	---

5.2 実験方法と評価基準

実験方法について説明する．免除制度の規則の知識や経験から表 5.3 のような 6 つの特徴が考えられる．

表 5.3 授業料免除制度の規則の特徴

	条件	最終結果
特徴 1	新入生である and 理由が 1 年以内	申請可
特徴 2	4 回生以内 and 理由が半年以内	申請可
特徴 3	免除される学期に休学する	申請不可
特徴 4	免除時、5 回生以上	申請不可
特徴 5	支払い困難な事情がない or 数年前	申請不可
特徴 6	留年した年である	申請不可

このように申請可となるパターンが 2 通り，申請不可となるパターンが 4 通りの計 6 通りが考えられる．この 6 つの特徴があるとき，表 5.4～5.6 の 3 パターンで抽象事例を考える．それぞれの抽象事例の与え方の違いについて説明する．

まず表 5.4 は，表 5.3 の申請不可となる特徴を 1 つの抽象事例として，まとめて表現したものである．

表 5.4 免除制度の抽象事例 (1)

休学	回生	理由	留年	新入生	クラス
*	*	1 年以内	*	はい	申請可
*	4 以内	半年以内	*	*	申請可
はい	5 以上	数年前	はい	*	申請不可
はい	5 以上	なし	はい	*	申請不可

次に表 5.5 は，6 つの特徴を 6 つの抽象事例として与えて，C4.5 による決定木作成を行った結果，具体事例内に矛盾を生じる事例があったためその原因となった抽象事例である「留年」に関する特徴を削除することで，矛盾を解消させた例である．

表 5.5 免除制度の抽象事例 (2)

休学	回生	理由	留年	新入生	クラス
*	*	1 年以内	*	はい	申請可
*	4 以内	半年以内	*	*	申請可
はい	*	*	*	*	申請不可
*	5 以上	*	*	いいえ	申請不可
*	*	なし	*	*	申請不可

最後に表 5.6 は，矛盾が生じて決定木が作成されなかった場合の対処法として，抽象事例を削除するのではなく，もとの規則に反しないよう未定”*”だった部分に属性値を代入することで，矛盾を解消するといった手法により，抽象事例を作成した例である．

表 5.6 免除制度の抽象事例 (3)

休学	回生	理由	留年	新入生	クラス
*	*	1 年以内	*	はい	申請可
いいえ	4 以内	半年以内	いいえ	*	申請可
はい	*	*	*	*	申請不可
*	5 以上	*	*	いいえ	申請不可
*	*	なし	*	*	申請不可
*	*	*	はい	いいえ	申請不可

以上のような 3 つの方法で与えた抽象事例から決定木を作成する．そしてその作成された決定木について，(a). 「もとの規則に基づいた正しい規則を表す決定木が作成されたか」，(b). 「抽象事例の与え方により，決定木の構造がどのように変化し，規則運用での知識や経験を表した抽象事例がどのように反映されているか」の 2 点について評価を行う．

5.3 実験結果

表 5.4～5.6 までの抽象事例から C4.5 によって作成された決定木が図 5.1～5.3 である．それぞれの決定木について考察する．

まず表 5.4 の抽象事例から作成された図 5.1 の決定木について，どのような木の構造をしているかを確認する．図 5.1 は，理由の属性で 4 つに分別され，その子ノードは全て単一のクラスになっている．これは理由という 1 つの質問だけで，全具体事例を分別し，最終的な答えまで導けるということを意味する．しかし，もとの規則から考えると，この決定木で表現されているものだけでは不十分である．このように表 5.4 のような抽象事例の与え方では，もとの規則を反映させた決定木は得られないことが分かる．

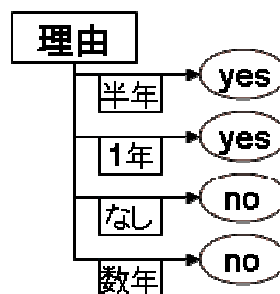


図 5.1 表 5.4 から作成された決定木

次に表 5.5 の抽象事例から作成された図 5.2 の決定木について，どのような木の構造をしているかを確認する．図 5.2 は，理由の属性で 4 つに分け，属性値が「数年」，「なし」の場合，申請不可になります．理由の属性値が「半年」の場合，次に回生，休学の順に属性を選び，[理由 = 半年 and 回生 = 4 以下 and 休学 = いいえ](関係 A) のときに申請可，回生と休学がそれ以外の属性値をとると，申請不可になります．また理由の属性値が「1 年」の場合，次に新入生，休学の順に属性を選び，[理由 = 半年 and 新入生 = はい and 休学 = いいえ](関係 B) のときに申請可，新入生と休学がそれ以外の属性値をとると，申請不可になります．このように図 5.1 の決定木よりも，もとの規則を詳しく表した決定木になり，さらに，文章で書かれた規則からでは読み取りづらかった関係 A や関係 B のような属性間の関係を明らかにすることができた．

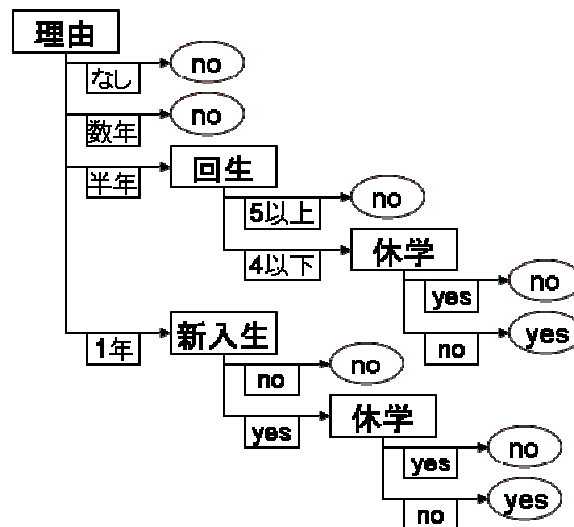


図 5.2 表 5.5 から作成された決定木

最後に表 5.6 の抽象事例から作成された図 5.3 の決定木について、どのような木の構造をしているかを確認する。図 5.3 は、新入生かどうかで分別されている。これにより、新入生である場合、何回生であるかと留年しているかの 2 つの属性は、聞くまでもなく答えが分かるため、のちの質問として選ばれることはない。よって、最初にこの属性で分別させるのは効果的である。新入生でない場合の部分木の方は、直線的な構造となっていて、質問の答えとして、1 つでも申請可の条件にふさわしくないものがあれば、申請不可の結論を出す構造になっている。このように unnecessary な属性を考えなくて済む質問の選択によって、木構造をコンパクトにしていることが分かる。また図 5.2 の決定木より抽象事例が増えていることで、全体的に木構造が大きくなっているものの、属性間の関係をより詳しく、もとの規則のポイント表した決定木になっているといえる。

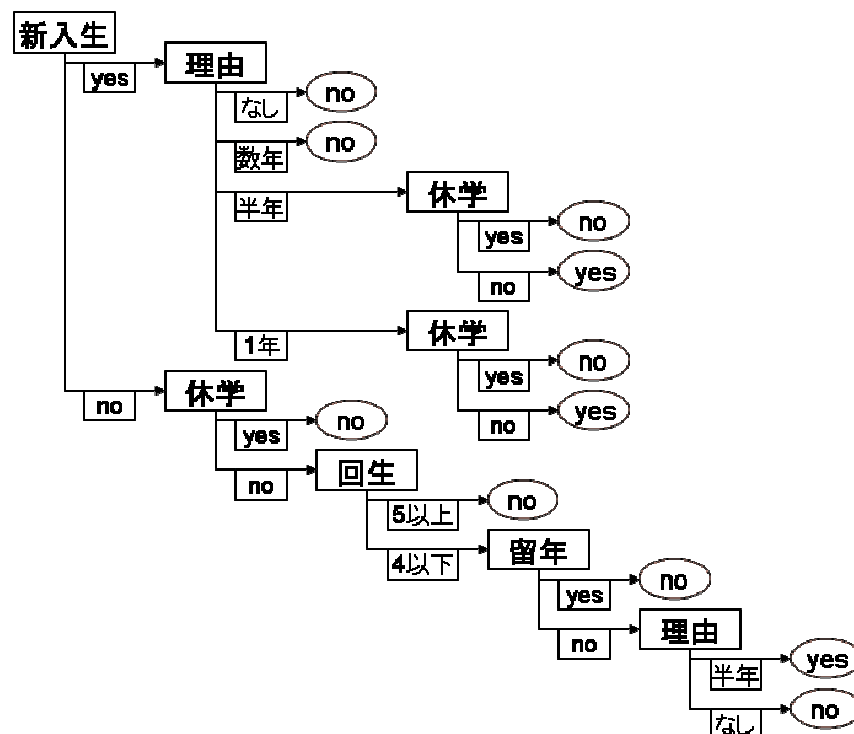


図 5.3 表 5.6 から作成された決定木

以上の3パターンの実験により、もと規則の表現の程度の差はあったものの、どの例も規則を表した決定木を作成することができた。また規則の特徴として与える抽象事例の数を増やして、規則の特徴を詳しくすることで、規則の核となるポイントを浮き彫りにし、知識や経験という目には見えない情報を可視化することができた。

第6章 考察

実験から，図 5.1 から図 5.3 の決定木は規則の表現の程度に差はあるが，どれも規則に反しないものが作成された．与えた抽象事例が少ないと，図 5.1 のような 1 つの属性しか考慮していない木構造になってしまい，規則の全体像は見えにくくなる．なので，与える抽象事例を増やした表 5.5 や表 5.6 のような例では，図 5.1 の決定木よりも規則の特徴が詳しく反映されたものが作成され，属性間の関係を明らかになることが分かった．これにより抽象事例の設定者が持っていた規則に関する知識や経験などを，決定木という図として表現し，設定者が曖昧な形で持っていた属性間の関係を明確に表すことができた．これを利用することで，抽象事例の設定者以外に対して，規則のポイントを規則の未熟練者へ伝えるためのツールとして用いたり，抽象事例の設定者が持っている知識や経験を視覚化することで，業務分析や業務改善に利用できる可能性がある．

今後の課題として，抽象事例の設定方法と C4.5 での質問評価部分の情報利得比基準の 2 点がある．一つ目の課題は，現在のシステムでは，規則の特徴を抽出した抽象事例を具体事例に展開して，C4.5 にかけて決定木を作るまで自動的に行う．その際，具体事例の中に，すべての属性値が同じ値を取るにも関わらず，最終的な答えのクラスが違ふ結果になるような事例があった場合，本システムでは，矛盾を生じてしまうため，途中で決定木を作成することなく，プログラムが停止してしまう．こうなってしまうと，展開した具体事例から矛盾を発生させた具体事例を探し出し，この矛盾を解消するために，抽象事例に修正を加えなければならない．この作業はすべて本システムの利用者がやらなければならない，すべての具体事例を調べる必要

があるため，大変手間のかかる作業となる．したがって，この手間を解消するために，抽象事例から具体事例を作成した際に，矛盾を生じる事例がないかを調べ，解消する機能が必要となる．この機能があれば，出力を出さずに停止するといったことが起こらなくなるはずである．次の課題として，C4.5 の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コンパクトな決定木を作成を行うが，情報利得比基準という評価は，与えた具体事例に対し，どの質問で分別すればよいかを，一つの質問の利得のみで計算している．この場合，部分木の大きさや重複は考えていない．実験結果の図 5.3 の決定木のように，新入生で，理由が半年と 1 年の所がある．この休学の属性で分別している部分木は，どちらも同じ構造で，さらにコンパクトな構造を目指すなら，重複部分を統合した方がよい．このように一段ごとの木の高さではなく，全体的な木構造の高さや重複を考慮した，質問の組み合わせの評価基準を提案する必要がある．

第7章 おわりに

本研究で目的として設定した，特徴を抽出した抽象事例を反映させた決定木が，実験結果により作成されることが確認された．さらに，抽象事例を詳しく与えるほど，規則を体系化した決定木が作成された．また文章の規則を決定木にしたことにより，属性間の論理積や論理和といった関係を明確に表現し，規則をあまり理解していない人に対しても，特徴を伝える手助けとして使用できたり，運用者が規則の業務改善や分析などの利用法が考えられる．

今後の課題として，抽象事例の設定の複雑さや質問の評価方法に対して考えが不十分な部分がある．しかし，その問題点を解決することで，本システムの利用者に対し，入力部分が簡単になり，規則を端的に重要なポイントを表現した判別ダイアグラムを作成できる．それにより，業務分析や業務改善に取り掛かるまでの時間の短縮につながるのではないかと考える．

謝辞

本研究ならびに論文作成にあたり、熱心に御指導して下さった下園真一准教授に深く感謝を致します。また、ゼミや研究の際、大変お世話になりました篠原武教授に心から感謝の言葉を申し上げます。最後に研究を重ねてきた篠原・下園研究室の学生諸氏に感謝の意を表します。

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).