

決定木とQM法による職業意思決定過程分析

河野 康成
山下 利之

1. はじめに

1.1. 研究の背景

職業の選択と準備は青年期後期の最も重要な発達課題の一つである(平出, 1988)。また、Erikson (1980)によれば、自我同一性(ego identity)の確立で重要であるのは、青年期における社会的役割の獲得であり、その中心的な役割を果たすのが職業決定である。そこで、職業選択に関する意思決定支援システムの構築が急務となるが(山下, 1996a, 1996b)、それらは一般に、卒業後に就業(進学)することを意思決定していることを前提としている。しかし、最近、アパシー(apathy)や留年など自我同一性の発達が不十分なため、職業についての自己決定ができないという消極的、心理的問題の職業未決定の問題(下山, 1986)、フリーター志望などそもそも特定の職業に就く意思がないなどの問題が増大している。効果的な職業指導等を行うためには、職業決定・職業未決定に至る心理的要因あるいは意思決定過程を十分に理解しなければならない(山下他, 2003)。

下山(1986)は大学生に対する質問紙調査と因子分析による解析により、職業未決定をもたらす因子を抽出し、それをもとにして職業未決定尺度を作成した。このような因子分析による研究は、職業未決定状態をもたらす心理的要因を明らかにするが、個人においてどのような心理的要因がどのように組み合わせられたときに職業未決定の問題

を引き起こすかに関しては、他の分析を組み合わせる必要がある。

1.2. 研究目的

本研究は、職業決定・未決定の心理的構造を把握するために、決定木とQM法(河野, 2001)を組み合わせることを試み、その有効性を考察することを目的とする。

職業未決定の状態は、さまざまな心理的要因が組み合わせられて生じていると考えられることから、どのような心理的要因がどのように組み合わせられて生じているかを把握し、それに応じた職業カウンセリング等の心理的配慮が必要である。

そこで、職業未決定尺度(下山, 1986)によって捉えることのできる職業決定・未決定の心理的要因の組み合わせに関して決定木分析を行い、さらに、決定木分析の結果に対し、分岐点を0,1としてクロス表を作成し、QM法によって考察を行った。すなわち、決定木分析の結果に対して、信頼性・妥当性があるかどうかを判断するために、決定木の分岐基準を使用し、総当りアルゴリズムを採用しているQM法によって確認するといった方法を取った。

最終的に、これらの分析結果を比較考察し、提唱手法の有効性を探った。

2. 方法

2.1. 被調査者

被調査者は、公立大学工学部学部生計 102 名（男性 91 名、女性 11 名）とした。

内訳（表 1）は、1 年生 22 名（男性 21 名、女性 1 名）、2 年生 39 名（男性 30 名、女性 9 名）、3 年生 31 名（男性 30 名、女性 1 名）、4 年生 10 名（男性のみ）であった。被調査者の学年、性別の構成が不均等であるため、本研究ではすべての被調査者をプールして分析、考察を行った。

2.2. 質問紙

本研究では、下山（1986）の職業未決定尺度を用いた。なお、40 項目のうち職業決定に関する 5 項目を除いた 35 項目から成る職業未決定尺度は、次のような 5 つの下位尺度から構成されている。

（1）未熟 尺度：

「19. これまで、自分自身で決定するという経験が少なく、職業決定のことを考えると不安になる」、「35. 自分が職業としてどのようなことをやりたいのかわからない」などの 7 項目から成る。職業意識が未熟なため、将来の見通しがなく、職業選択に取り組めないでいる状態を表す。

（2）混乱 尺度：

「12. 職業決定のことを考えると、とてもあせりを感じる」、「26. 将来の職業のことを考えると気が滅入ってくる」などの 8 項目から成る。職業決

定に直面して不安になり、情緒的に混乱している状態を表す。

（3）猶予 尺度：

「3. せっかく大学に入ったのだから、今は職業のことは考えたくない」、「4. できることなら職業決定は、先に延ばし続けておきたい」などの 7 項目から成る。職業決定を猶予して当面のところは職業については考えたくない状態を表す。

（4）模索 尺度：

「20. 職業に関する情報がまだ十分ないので、情報を集めてから決定したい」、「39. 職業は決まっていないが、今の関心を深めていけば職業につながってくると思う」などの 6 項目から成る。職業決定に向かって積極的に模索している状態を表す。

（5）安直 尺度：

「7. 生活が安定するなら、職業の種類はどのようなものでもよい」、「14. 自分を採用してくれる所なら、どのような職業でもよいと思っている」などの 7 項目から成る。自らの関心や興味を職業選択に結びつけていこうとする努力をしない安易な職業決定態度を表す。

下山（1986）の職業未決定尺度では項目 18 と項目 38 が、項目分析の結果に基づき、尺度から削除されている。そこで、本研究では、項目 18 と項目 38 に、以下の項目を挿入して、職業決定の

表 1. 被調査者の内訳

	1年生		2年生		3年生		4年生		合計
	男性	女性	男性	女性	男性	女性	男性	女性	
就職先希望決定	6	1	15	2	13	1	3	0	41
就職先希望未決定	15	0	15	7	17	0	7	0	61
合計	21	1	30	9	30	1	10	0	102

基準とした。

「18. 卒業後、どのような職業（職種）に就きたいかを既に決めている。」

「38. 卒業後、就職希望先（会社名、官庁名）をすでに決めている。」

これらの質問項目に対して、「まったくあてはまらない」を-2、「ややあてはまらない」を-1、「どちらともいえない」を0、「ややあてはまる」を1、「非常にあてはまる」を2とする5件法により回答を求めた。

3. 分析例

3.1. 決定木分析

職業決定・職業未決定の心理的要因について、尺度化された下位尺度に関して、決定木分析を行った。

決定木分析では、QM法が現在のところ2値データしか扱えない（マルチデータの場合、ダミー変数化する必要がある）ため、決定木アルゴリズムには、2進木を採用した。分析ソフトウェアには、SPSS Answer Tree 3.1のC&RTを用いた。

説明変数には、未熟、混乱、猶予、模索、安直の各下位尺度に含まれる項目の平均評定値を用いた。

目的変数に関して、「18. 卒業後、どのような職業（職種）に就きたいかを既に決めている」の項目について、2と1を1に、0、-1、-2を0にそれぞれ変換した。

また、QM法では論理変数を必要とするため、これらの説明変数 未熟（Immaturity）、混乱（Confusion）、猶予（Moratorium）、模索（eXploration）、安直（Easiness）を論理変数I、C、M、X、Eとした。

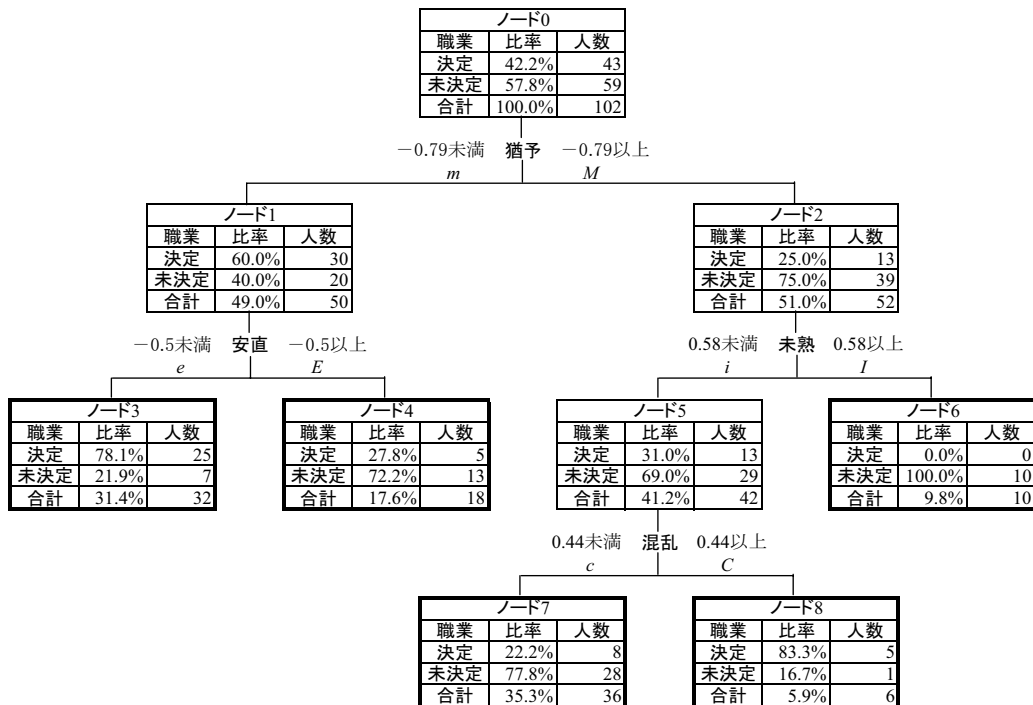


図1. 決定木分析による結果

図1は、職業決定に関する決定木による分析結果である。

職業を決定している学生のパターンを示すターミナルノードは、ノード3とノード8である。ノード3は、「猶予 M が -0.79 未満」かつ「安直 E が -0.50 未満」の学生を表す。また、ノード8は、「猶予 M が -0.79 以上」かつ「未熟 I が -0.58 未満」かつ「混乱 C が 0.44 以上」の学生を表す。

一方で、職業を決定していない（職業未決定）学生のパターンを示すターミナルノードは、ノード4、ノード6、ノード7である。ノード4は、「猶予 M が -0.79 未満」かつ「安直 E が -0.50 以上」の学生を表す。ノード6は、「猶予 M が -0.79 以上」かつ「未熟 I が -0.58 以上」の学生を表す。さらに、ノード7は、「猶予 M が -0.79 以上」かつ「未熟 I が -0.58 未満」かつ「混乱 C が 0.44 未満」の学生を示す。

この結果を論理変数によって示すと、職業決定 (R_1) と職業未決定 (R_0) のパターンは、

$$R_1 = me + iCM, \quad (1)$$

$$R_0 = mE + IM + icM \quad (2)$$

となった（この式において、大文字は各項目の分岐値以上を、小文字は分岐値未満を示す。また、積は論理記号 AND を、+ は論理記号 OR を示す）。

3.2. Q 法による分析

次に、決定木の分析結果に対して QM 法を用いて考察する。QM 法は、応答確率に従ってデータからサブグループを探索する手法である（河野，2001）。この方法は、ターゲット値（設定した応答確率）よりも二値型説明変数の応答確率が高いターゲット集団を検出しているため、データから興味あるパターンを抽出する場合に効果的である。

また、QM 法による分析では 2 値データに変換する必要があり、決定木の分岐値を元に 1 と 0 の値に変換した。2 値変換後の職業決定者および職

表2. 職業決定者のクロス表

Condition				Number of $Y = 1$	Cases	Response probability
I	C	M	E			
0	0	0	0	23	30	76.7%
0	0	0	1	4	15	26.7%
0	0	1	0	7	24	29.2%
0	0	1	1	1	12	8.3%
0	1	0	0	2	2	100.0%
0	1	0	1	1	2	50.0%
0	1	1	0	1	1	100.0%
0	1	1	1	4	5	80.0%
1	0	0	0	0	0	-
1	0	0	1	0	0	-
1	0	1	0	0	1	0.0%
1	0	1	1	0	2	0.0%
1	1	0	0	0	0	-
1	1	0	1	0	1	0.0%
1	1	1	0	0	2	0.0%
1	1	1	1	0	5	0.0%

業未決定者のクロス表は、表2と表3である。この表において、模索 (eXploration) は、決定木の結果に表れなかったためこの変数を削除した。

これらクロス表に基づき、QM 法を行った結果、職業決定・未決定のパターン（応答確率 75% 以上）は、

$$R_1 = ime + iCM, \quad (3)$$

$$R_0 = cME + IM + ICE \quad (4)$$

となった。

式(3)を見ると、職業決定者のパターンは、決定木の場合とほぼ同様な結果を示しており、 iCM は式(1)と同じサブグループが検出されている。唯一、決定木では式(1)の中で、 me （「猶予 M が -0.79 未満」かつ「安直 E が -0.50 未満」）と検出されたサブグループが、QM 法では「未熟 I が -0.58 未満」の項目が加わり、下位グループの ime に絞られている程度である。

また、職業未決定者の場合は、式(2)と式(4)

表3. 職業未決定者のクロス表

Condition				Number of Cases	Response probability	
<i>I</i>	<i>C</i>	<i>M</i>	<i>E</i>			
0	0	0	0	7	30	23.3%
0	0	0	1	11	15	73.3%
0	0	1	0	17	24	70.8%
0	0	1	1	11	12	91.7%
0	1	0	0	0	2	0.0%
0	1	0	1	1	2	50.0%
0	1	1	0	0	1	0.0%
0	1	1	1	1	5	20.0%
1	0	0	0	0	0	-
1	0	0	1	0	0	-
1	0	1	0	1	1	100.0%
1	0	1	1	2	2	100.0%
1	1	0	0	0	0	-
1	1	0	1	1	1	100.0%
1	1	1	0	2	2	100.0%
1	1	1	1	5	5	100.0%

を比較すると、*IM*（「未熟 *I* が -0.58 以上」かつ「猶予 *M* が -0.79 以上」）は共通している。

しかしながら、残りのサブグループは、式(2)で、*mE*（「猶予 *M* が -0.79 未満」かつ「安直 *E* が -0.50 以上」）と *icM*（「未熟 *I* が -0.58 未満」かつ「混乱 *C* が 0.44 未満」かつ「猶予 *M* が -0.79 以上」）が検出されているが、式(4)では、*cME*（「混乱 *C* が 0.44 未満」かつ「猶予 *M* が -0.79 以上」かつ「安直 *E* が -0.50 以上」）と *ICE*（「未熟 *I* が -0.58 以上」かつ「混乱 *C* が 0.44

以上」かつ「安直 *E* が -0.50 以上」）と全く別のグループが検出されている。

さらに、職業未決定者のパターンに、応答確率 75% 以上から 70% 以上に、閾値 (cutoff) を変更した場合の結果は、

$$R_0 = cM + IM + ICE + icE \quad (5)$$

となった。

同様に、式(2)と式(5)を比較すると、先の式(4)との比較と同様に、*IM*（「未熟 *I* が -0.58 以上」かつ「猶予 *M* が -0.79 以上」）が共通している以外は、まったく別のサブグループが検出されている。唯一、式(2)で、*icM*（「未熟 *I* が -0.58 未満」かつ「混乱 *C* が 0.44 未満」かつ「猶予 *M* が -0.79 以上」）と検出されたサブグループが、式(5)では、*cM*（「混乱 *C* が 0.44 未満」かつ「猶予 *M* が -0.79 以上」）と上位のサブグループが検出された程度である。

3.3. 信頼区間の設定

以上の決定木とQM法の結果に対して、80%の信頼区間を用いて、信頼区間の下限 (Lower of the confidence) を満たしているかどうかをチェックする。

表4を見ると、職業決定者に関しては、決定木及びQM法のいずれの結果のサブグループも信頼区間の下限が70%未満となり、設定応答確率を70%とした場合の信頼区間を満たしていない。すなわち、いずれの結果も信頼性が低いことを示し

表4. 職業決定者の結果に対する信頼区間 (80%)

Target group	Method	Number of Cases	Y=1	Response probability	Lower of the confidence	Upper
<i>ime</i>	QM	25	32	78.1%	68.8%	87.5%
<i>iCM</i>	QM, Tree	5	6	83.3%	63.8%	102.8%
<i>me</i>	Tree	25	32	78.1%	68.8%	87.5%

表5. 職業未決定者の結果に対する信頼区間 (80%)

Target group	Method	Number of Y=1	Cases	Response probability	Lower of the confidence	Upper
<i>cME</i>	QM75	13	14	92.9%	84.0%	101.7%
<i>IM</i>	QM75, QM70, Tree	10	10	100.0%	100.0%	100.0%
<i>ICE</i>	QM75, QM70	6	6	100.0%	100.0%	100.0%
<i>cM</i>	QM70	31	39	79.5%	71.2%	87.8%
<i>icE</i>	QM70	22	27	81.5%	71.9%	91.1%
<i>mE</i>	Tree	13	18	72.2%	58.7%	85.8%
<i>icM</i>	Tree	28	36	77.8%	68.9%	86.7%

注 QM75 は閾値 75% 以上、QM70 は閾値 70% 以上を示す

ている。

次に、表5を見ると、職業未決定者に関して、決定木とQM法の共通に検出されたサブグループである*IM*以外は、異なった結果となっている。決定木によって検出された残り2つのサブグループは、信頼区間の下限が70%未満となり、設定応答確率を70%とした場合の信頼区間を満たしていない。一方で、QM法によって検出された残りのサブグループは、信頼区間の下限が全て70%以上となり、設定応答確率70%の信頼区間を満たしている。

4. おわりに

決定木とQM法の比較考察の結果、双方のサブグループの共通項は、より強い要因となっていることを示している。それは、決定木では、ノード分割基準にジニ係数を用いて重要な要因の組み合わせが検出 (Breiman *et al.*, 1984) され、さらに、QM法では、応答確率の高いグループが検出されているからである。

決定木とQM法を比較すると、QM法は、総当り法であるため要因抽出にあたって取りこぼしが

ないという優位性がある。また、応答確率によってパターンを抽出できるため、式(3)、(4)以外にも、応答確率を分析の用途に従って100~0%の間で変更することが可能である。

しかしながら、QM法は、どのような要因がどのように組み合わせられたときに、当該事象を生起するかにに関する情報は示すことができるが、どの要因がどれほど重要であり、どのような過程により各要因が組み合わせられていくかにに関する示唆は得られない。この点については、重要な要因順に検出している決定木の方が優れている。

本手法のメリットのひとつに、実際に存在しないサブグループを削除することが可能である点が挙げられる。たとえば、職業決定者に関する決定木の結果として検出された *me* は、「猶予 *M* が -0.79 未満」かつ「安直 *E* が -0.50 未満」で、他の未熟 (Immaturity) と混乱 (Confusion) は、1と0の両方の場合を含む。しかし、実際には、*Icme* と *ICme* の2つのサブグループは実在しない (表6)。これに対して、QM法の結果は、応答確率に従ってサブグループを検出しているため、実際に存在している下位のサブグループである *ime* を示している。つまり、サブグループを絞り込むこと

表6. 実在しないパターンに関するクロス表

Condition				Number of Cases	Response probability	
<i>I</i>	<i>C</i>	<i>M</i>	<i>E</i>			
0	0	0	0	23	30	76.7%
0	0	0	1	4	15	26.7%
0	0	1	0	7	24	29.2%
0	0	1	1	1	12	8.3%
0	1	0	0	2	2	100.0%
0	1	0	1	1	2	50.0%
0	1	1	0	1	1	100.0%
0	1	1	1	4	5	80.0%
1	0	0	0	0	0	-
1	0	0	1	0	0	-
1	0	1	0	0	1	0.0%
1	0	1	1	0	2	0.0%
1	1	0	0	0	0	-
1	1	0	1	0	1	0.0%
1	1	1	0	0	2	0.0%
1	1	1	1	0	5	0.0%

が可能となる。

今回の分析で、決定木とQM法を組み合わせたアプローチは、意思決定過程における要因分析に有効であることがわかった。

今後は、QM法にもApriori アルゴリズム(Agrawal & Shafer, 1996)で採用されているサポートルール(Agrawal et al., 1993)等のアルゴリズムを導入し、データ量に係わらず分析可能な手法に改善する。さらに、本手法に関するソフトウェアを開発する予定である。

引用文献

Agrawal, R. and Shafer, J.(1996) Parallel mining of association rules. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 962-969.

Agrawal, R., Iielinski, T., & Swami, A.(1993) Mining association rules between sets of items in large databases. *Proceedings of the ACM SIGMOD Conference on Management of Data*, 207-216.

Breiman, L., Friedman, J., Olshen, R. S., & Stone, C. (1984). *Classification and Regression Trees*. Wadsworth International Group. Belmont, CA.

Erikson, E. H. (1980). *Identity and the life cycle*. Norton.

平出彦仁 (1988). 発達心理学序説 . 八千代出版 .

河野康成 (2001). Quine-McClusky アルゴリズムを用いた目的別対象集団の探索手法 . 応用社会学研究 43, 95-101.

下山晴彦(1986). 大学生の職業未決定の研究 . 教育心理学研究 34(1), 20-30.

山下利之(1996a). 職業生活設計に関する意思決定支援 - ISM法の適用の試み - . 日本労働研究雑誌 438, 40-48.

山下利之(1996b). ファジィ構造モデルによる職業選択動機および職業生活設計の分析 . 日本ファジィ学会誌 8(5), 861-870.

山下利之・河野康成・葛原茂一郎 (2003). 大学生の職業未決定をもたらす心理的要因の組み合わせに関する質的比較分析 . 日本教育工学会論文誌 27(Supple.), 109-112.