

(15) 遺伝的アルゴリズムを用いた斜面崩壊危険度診断事例からの知識獲得手法

Application of Genetic Algorithms to Rough Set Based Knowledge Discovery Method

原川 浩一^{*} ・ 広兼 道幸^{**} ・ 古田 均^{***}
Koichi HARAKAWA, Michiyuki HIROKANE and Hitoshi FURUTA

^{*} 関西大学 総合情報学部 (〒569-1095 大阪府高槻市霊仙寺町 2-1-1)

^{**} 工博 関西大学助教授 総合情報学部 (〒569-1095 大阪府高槻市霊仙寺町 2-1-1)

^{***} 工博 関西大学教授 総合情報学部 (〒569-1095 大阪府高槻市霊仙寺町 2-1-1)

Abstract In this paper, the application of the genetic algorithms is described for the sake of the solution of this problem. As for the application method of the genetic algorithms, the conditional attributes that should be reduced are considered as genes, and the number of the remaining conditional attributes, and the contradiction rules, are considered as the evaluation functions.

Key Words : Rough set, Genetic algorithms, Slope-failure, Knowledge Acquisition

1. はじめに

斜面の崩壊危険度の診断をシステム化するためのアプローチのひとつとしては、専門家の診断事例から経験的知識を抽出し、それを基にES (Expert System) を構築する方法が考えられる。診断事例から経験的知識を取り出す方法として、ラフ集合論を用いてルール型知識を獲得する方法^{1),2)}などが研究されている。これらの研究は、専門家の診断事例にラフ集合論の基本的概念のひとつである類別定義³⁾をもとに、不要な判定要因や判定要因に対する項目を取り除く事によって、専門家の経験的知識をルールとして抽出するというものである。しかし、ラフ集合論の適用においては、決定表に影響を与えていない判定要因、あるいは判定要因に対する項目の組み合わせを総当たりで検討する必要がある。そこで、診断に必要な判定要因の数や事例数など、データ数が多くなった場合に、実用面において適用が困難になる可能性が懸念される。

本論文では、専門家の診断事例から経験的知識を抽出する技術としては類別定義による方法を基本として、計算量削減のためにGA (Genetic Algorithms) を組み合わせた方法を提案する。GAは主にNP困難となるような組み合わせ最適化問題に対して、主に計算量を減らして準最適解を得る目的で利用されているアルゴリズムである⁴⁾。従来の類別定義による方法で必要となる組み合わせ最適化問題の準最適解をGAを用いて導くことで、計算量を減らすことが可能であると思われる。このとき、すべての組み合わせを検討して得られる極小決定アルゴリズムに充分近いものを得ることができれば、事例のデータ数が多くなった場合にも類別定義を経験的知識の抽出に適用しやすくなると考えられる。具体的には、不要な判定要因を取り除く過程において組み合わせの総当たりではなく、GAを適用して必要度の高い判定要因の組み合わせを得た。GAの適用に

あたっては、評価関数における評価項目の重みを数通りに変化させ、評価方法の違いによる結果の違いを検証した。さらに、GAを用いることによって得られた極小決定アルゴリズムと、総当たりで得られた極小決定アルゴリズムを比較し、データ数が多くなった場合の解決策として、ラフ集合論に基づく知識発見手法へGAを用いる方法の適用可能性を検証する。

2. 専門家の診断事例

本論文において、専門家の経験的知識を抽出するために使用した診断事例は、粘板岩を主体とする道路沿いの斜面の崩壊危険度を診断したものである。危険度の診断は高速道路調査会の診断方法⁵⁾に従っている。表-1はある地域の32箇所の斜面について、12の判定要因に対して選択された項目番号と、総合的な危険度の診断結果をまとめたものである。各斜面は、すべての要因に対して選択した項目から、対応する危険ランクをもとに点数制で危険度が診断されている。この表は、それぞれの斜面に対して、条件属性である判定要因の項目番号と決定属性である危険度との関係(決定規則)をまとめたもので、以後、決定表と呼ぶ。

3. 全数比較法による極小決定アルゴリズム生成

ラフ集合論に基づき、決定表から極小決定アルゴリズムを導く手順を以下に示す。

処理(1) 矛盾する決定規則のチェック

処理(2) 同一決定規則の取り除き

処理(3) 組み合わせの総当たりによる不要な判定要因の取り除き

処理(4) 同一決定規則の取り除き

処理(5) 不要な、判定要因に対する決定要因の取り除き

処理(6) 同一決定規則の取り除き

表-1. 専門家の診断事例

斜面 番号	判定要因												危険度
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	
1	②	①	①	②	②	②	②	④	②	②	③	①	A
2	④	②	①	②	②	②	②	④	②	②	②	①	B
3	③	②	②	③	②	②	②	④	③	④	③	②	C
4	④	②	①	②	①	③	②	②	②	③	③	③	B
5	③	②	①	①	②	②	②	④	①	③	③	①	A
6	③	①	①	①	②	②	②	④	③	②	①	①	A
7	④	②	②	②	②	②	②	④	②	④	③	③	C
8	④	②	②	②	②	②	②	④	②	④	③	③	C
9	③	①	①	②	②	④	②	③	①	④	③	③	B
10	③	①	①	③	②	④	②	③	①	④	③	③	B
11	④	②	①	③	②	④	②	④	②	③	③	③	C
12	④	②	①	③	②	④	②	③	①	③	③	③	B
13	④	②	①	③	②	②	②	④	②	③	①	①	B
14	③	②	②	③	②	②	②	①	④	①	④	②	B
15	④	②	①	③	②	④	①	④	②	④	③	③	C
16	④	②	①	③	②	②	②	④	③	④	①	①	B
17	④	①	①	③	②	④	②	③	①	③	②	③	B
18	③	②	①	③	②	②	②	①	④	②	③	①	B
19	④	②	①	③	②	②	②	④	②	③	③	①	C
20	④	②	①	③	②	③	②	③	①	③	③	③	B
21	③	①	①	③	②	②	②	④	②	③	③	③	B
22	④	②	②	③	②	②	②	④	③	③	③	③	C
23	④	②	②	③	②	③	①	④	①	③	③	③	C
24	④	①	①	②	②	④	②	③	①	④	③	①	B
25	④	②	②	①	①	②	②	④	①	④	①	①	A
26	④	②	②	②	②	②	②	①	③	①	④	①	B
27	④	①	②	③	②	②	②	④	③	④	③	③	C
28	②	①	①	①	①	②	②	④	②	④	③	①	A
29	②	①	①	①	①	②	②	④	②	④	③	①	A
30	②	①	①	①	②	②	②	①	④	③	③	②	A
31	③	①	①	②	②	②	①	④	①	③	②	①	A
32	④	②	②	②	②	②	②	④	③	③	③	②	C

表-2. 判定要因の組み合わせ

ケース	判定要因				
CASE-1	(1)	(2)	(8)	(9)	(11)
CASE-2	(1)	(3)	(4)	(8)	(11)
CASE-3	(1)	(3)	(4)	(9)	(11)
CASE-4	(1)	(3)	(8)	(9)	(11)
CASE-5	(1)	(4)	(8)	(9)	(11)
CASE-6	(1)	(4)	(8)	(10)	(11)
CASE-7	(1)	(4)	(8)	(11)	(12)
CASE-8	(1)	(8)	(9)	(10)	(11)
CASE-9	(1)	(8)	(9)	(11)	(12)

表-3. 極小決定アルゴリズム
(CASE-1)

斜面 番号	判定要因					危険度
	(1)	(2)	(8)	(9)	(11)	
1	②					A
2				②	②	B
3		②		③	③	C
4			②			B
5	③		④	①	③	A
6		①			①	A
7	④		④		③	C
9			③			B
13				②	①	B
14		②			②	B
16		②		③	①	B
18	③			②		B
25			④	①	①	A
27	④			③	③	C
31		①	④		②	A

この手順による方法を、以下、全数比較法と呼ぶ^{1),2)}。

表-2 は全数比較法において、不要な判定要因を取り除く過程（処理(3)）で得られた、決定表に矛盾を含まず、かつ、判定要因の数が最小となる判定要因の組み合わせである。表-2 の CASE-1 に示される判定要因の組み合わせを用いて、全数比較法の処理(4)以下の手順を行って得られたものが表-3 である。表-3 は、表-1 に示す専門家の診断事例に対し、全数比較法を用いて、不要な部分をすべて取り除いたもので、なおかつ表-1 と同様の診断を行うことのできる極小決定アルゴリズムである。表-2 の CASE-2 から CASE-9 に示される判定要因の組み合わせについても、それぞれ CASE-1 と同様に極小決定アルゴリズムを生成することが可能である。

4. GA を組み合わせた方法による極小決定アルゴリズムの生成

全数比較法による極小決定アルゴリズムの生成過程において、処理(3)では、すべての組み合わせを比較するという方法をとっているため、常に最適解が得られるが、データ数が増えれば増えるほど計算時間を要するという欠点もある。ここでは、処理(3)に代わる方法として、GA を用いて、矛盾が少なく、判定要因の数が極力少なくなるような組み合わせを求める方法について説明する。

4.1 GA の処理方法

(1)遺伝子列

本研究では、検討する判定要因の取り除き方を遺伝子列として表現するにあたって、一定の長さのビット列を用いた。遺伝子列となるビットの数は判定要因の数と等しく、表-1 の場合、12 となる。それぞれのビットは各判定要因に対応し、その値が 1 ならばそのビットに対応する判定要因を取り除き、0 であれば取り除かずに矛盾を検討することとした。たとえば、表-1 に対して、ビット列「110000000010」は、判定要因(1),(2)および(11)を取り除いた組み合わせを表している。

(2)交叉と突然変異

本論文では GA の交叉率は 0.6、突然変異率は 0.05 に設定した。交叉方法は一点交叉で、突然変異は任意の 1 ビットを反転させる方法を用いた。

(3)選択

選択はある個体が次の世代に残される確率を（その個体の適応度）/（全個体の適応度の和）とし、その確率に応じてランダムに選択されるルーレット戦略を基本とした。さらに、適応度の高い個体を保護するためにエリート戦略を併用し、各世代で適応度の最も高い個体 2 つを無条件で次の世代に残した。

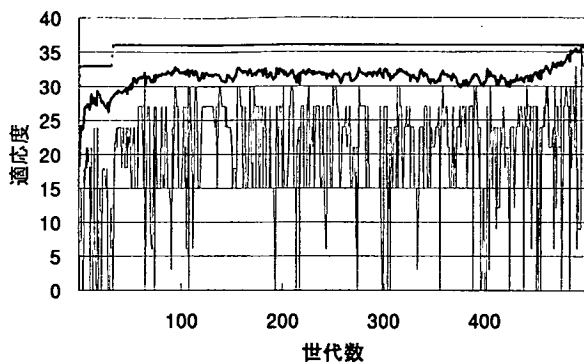


図-1.1. 適応度の変化 $w_1 = +3, w_2 = -3$

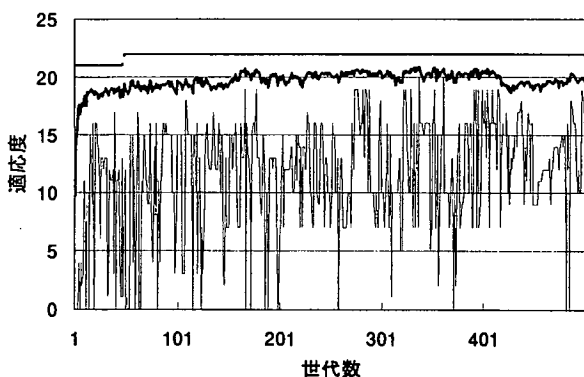


図-1.2. 適応度の変化 $w_1 = +1, w_2 = -3$

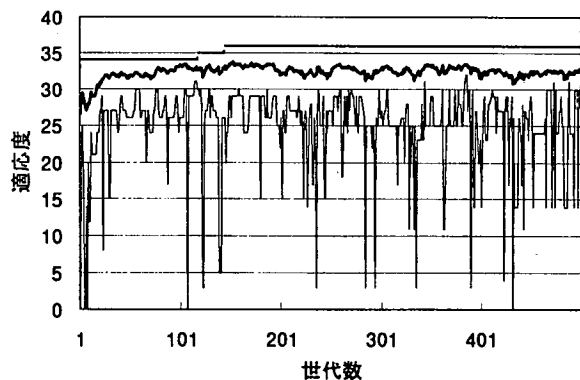


図-1.3. 適応度の変化 $w_1 = +3, w_2 = -1$

--- max. — min. — ave.

(4) 計算終了までの世代数と各世代の個体数

計算終了までの世代数は 500 世代、各世代における個体数は 50 個体に設定し、最終世代において、重複するものを除いて、適応度の高いほうから 10 個体を、判定要因の組み合わせの準最適解として採用した。

(5) 適応度の計算

各個体の適応度の計算は、一定の初期値を基準として、個体のビット列に含まれる 1 の個数が多いほど適応度を増加させ、その組み合わせで決定規則の矛盾を検討した後、矛盾が発見された数が多いほど適応度を減少させることにした。適応度計算の関数は、

表-4.1. 判定要因の組み合わせ

(GA, $w_1 = +3, w_2 = -3$)

ケース	判定要因						備考
CASE-11	(1)	(8)	(9)	(11)	(12)		CASE-9
CASE-12	(1)	(4)	(8)	(11)	(12)		CASE-7
CASE-13	(1)	(4)	(8)	(9)	(11)		CASE-5
CASE-14	(1)	(8)	(9)	(11)			矛盾含む
CASE-15	(1)	(8)	(9)	(10)	(11)		CASE-8
CASE-16	(1)	(4)	(8)	(9)	(11)	(12)	冗長
CASE-17	(1)	(6)	(8)	(9)	(11)		矛盾含む
CASE-18	(1)	(3)	(4)	(8)	(9)	(11)	冗長
CASE-19	(1)	(6)	(8)	(9)	(11)	(12)	冗長

表-4.2. 判定要因の組み合わせ

(GA, $w_1 = +1, w_2 = -3$)

ケース	判定要因								備考
CASE-21	(1)	(3)	(8)	(9)	(11)				CASE-4
CASE-22	(1)	(8)	(9)	(11)	(12)				CASE-9
CASE-23	(1)	(6)	(7)	(9)	(11)	(12)			
CASE-24	(1)	(5)	(6)	(7)	(9)	(11)	(12)		
CASE-25	(1)	(3)	(6)	(7)	(9)	(11)	(12)		
CASE-26	(6)	(7)	(9)	(10)	(11)	(12)			
CASE-27	(1)	(3)	(5)	(6)	(7)	(9)	(11)	(12)	
CASE-28	(1)	(5)	(6)	(7)	(8)	(9)	(11)	(12)	
CASE-29	(4)	(5)	(6)	(7)	(9)	(11)	(12)		

表-4.3. 判定要因の組み合わせ

(GA, $w_1 = +3, w_2 = -1$)

ケース	判定要因						備考
CASE-31	(1)	(6)	(9)	(11)			矛盾含む
CASE-32	(2)	(6)	(7)	(9)	(10)		矛盾含む
CASE-33	(2)	(6)	(9)	(10)	(11)		矛盾含む
CASE-34	(2)	(6)	(9)	(10)			矛盾含む
CASE-35	(1)	(2)	(6)	(9)	(10)		矛盾含む
CASE-36	(2)	(4)	(6)	(9)	(10)		矛盾含む
CASE-37	(2)	(6)	(8)	(9)	(10)		矛盾含む
CASE-38	(2)	(3)	(4)	(6)	(9)	(10)	矛盾含む
CASE-39	(2)	(4)	(6)	(7)	(9)	(10)	(11)

$$F = B + w_1 X + w_2 Y \quad (1)$$

とする。式(1)において、 X は個体のビット列に含まれる 1 の数、 Y は矛盾の発見された数とする。 B は適応度の初期値となる定数である。係数 w_1, w_2 は、判定要因を減らすことに対する加点と、矛盾に対する減点と同じ比重になるよう $w_1 = +3, w_2 = -3$ とした。さらに、比較のため、矛盾に対する減点を極端に重視した評価 ($w_1 = +1, w_2 = -3$) と、判定要因を多く取り除くことを極端に重視した評価 ($w_1 = +3, w_2 = -1$) を加え、合計 3 通りを試行した。

4.2 GA の適用結果

GA における適応度の変化を図-1.1 から図-1.3 に示す。これより、いずれの場合も、適応度の平均値は収束傾向にあるといえる。

表-4.1 から表-4.3 はそれぞれの GA の最終世代 50 個体のうち、重複するものを除いたものである。このうち、特に表-2 に含まれる全数比較法で得られた組み合わせと同一の組み合わせが存在する場合、その備考欄に、対応する「ケース番号」を記入した。さらに、得

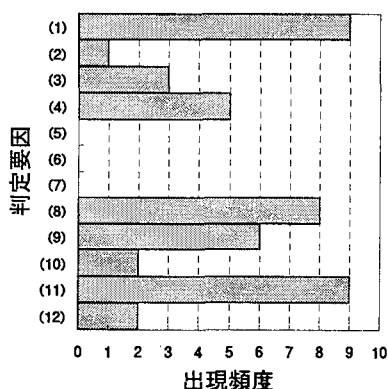


図-2. 判定要因の出現頻度 (全数比較法)

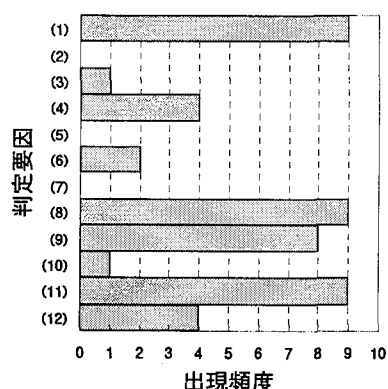


図-3.1. 判定要因の出現頻度 (GA, $w_1 = +3, w_2 = -3$)

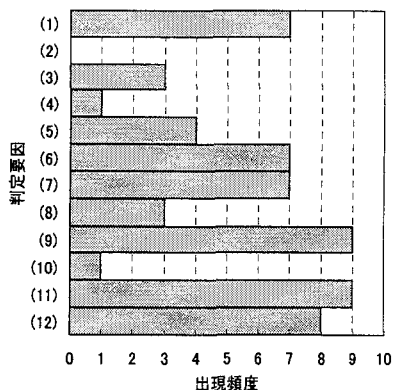


図-3.2. 判定要因の出現頻度 (GA, $w_1 = +1, w_2 = -3$)

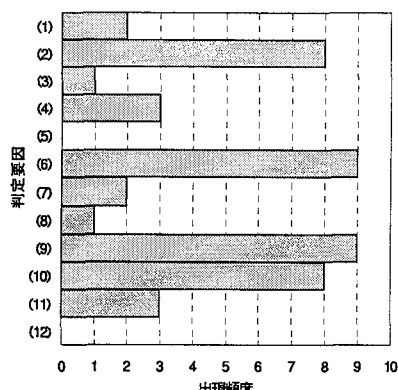


図-3.3. 判定要因の出現頻度 (GA, $w_1 = +3, w_2 = -1$)

られた極小決定アルゴリズムが矛盾を含むものに対しては「矛盾含む」、判定要因の組み合わせが、表-2に含まれる組み合わせのうちいずれかの組み合わせに余分な判定要因を付け加えた形になっているものには「冗長」と記入した。

表-4.1に示されるとおり、 w_1, w_2 を同じ割合にしたものでは、得られた9通りの判定要因の組み合わせのうち、4通りが全数比較法で得られたものと同じ組み合わせであった。

さらに、 w_1 に対して w_2 の割合を大きくすることによって、表-4.2に示されるとおり、矛盾を含む判定要因の組合せを少なくできたが、表-2や表-4.1に比べて判定要因が多く残る傾向があった。

また、 w_2 に対して w_1 の割合を大きくすることによって、表-4.3に示されるように、判定要因の数が多くなることは抑制することができたものの、ほとんどの場合が矛盾を含む組合せとなった。

5. 全数比較法とGAによる結果の比較

図-2および図-3.1から図-3.3は、それぞれ表-2(全数比較)および表-4.1から表-4.3(GA)に示す、極小決定アルゴリズムの生成に必要な判定要因の出現頻度を棒グラフで表したものである。図-2と図-3.1において、各判定要因の出現頻度の傾向は非常によく似ており、判定要因(1), (8), (9), (11)の出現頻度が高くな

っている。さらに、表-4.1に含まれる判定要因の組み合わせを検討すると、9通りの組み合わせの多くが、表-2に含まれる最適な組み合わせと同じものか、それらの組み合わせに余分な判定要因を1つ付け加えたり、必要な判定要因を1つ取り除いたものであることがわかる。

また、図-3.2, 図-3.3においては、全数比較法において出現頻度が高かった判定要因の出現頻度が低くなり、また逆に、全数比較法において出現頻度が低かった判定要因の出現頻度が高くなるという傾向が見られる。この傾向は、図-3.3でもっとも強く見られる。

以上のことから、GAによって極小決定アルゴリズムの生成に必要な判定要因の組み合わせを導いた場合でも、評価関数の設定によって、精度が若干低くなることを除けば、全数比較法によるものにかかなり近い傾向の結果が得られると考えられる。

表-2, 表-4.1から表-4.3に含まれるすべての判定要因の組み合わせを用いて極小決定アルゴリズムを生成した場合の、各極小決定アルゴリズムに含まれるルール(決定規則)の数と、評価項目(決定表内に残った項目)の数を表-5に示す。表-5の網掛け部は、その極小決定アルゴリズムが矛盾を含んだルールを持っていることを示している。

表-5. ルール数と評価項目数

方法	ケース	ルール数	評価項目数
ラフ集合論	CASE-1	15	35
	CASE-2	11	24
	CASE-3	16	37
	CASE-4	14	30
	CASE-5	15	32
	CASE-6	15	35
	CASE-7	15	34
	CASE-8	18	42
	CASE-9	18	41
ラフ集合論 + GA $(w_1=+3, w_2=-3)$	CASE-11	18	41
	CASE-12	15	34
	CASE-13	15	32
	CASE-14	15	38
	CASE-15	18	42
	CASE-16	19	45
	CASE-17	17	43
	CASE-18	16	37
	CASE-19	19	43
ラフ集合論 + GA $(w_1=+1, w_2=-3)$	CASE-21	14	30
	CASE-22	18	41
	CASE-23	22	53
	CASE-24	23	55
	CASE-25	22	53
	CASE-26	25	66
	CASE-27	23	55
	CASE-28	21	49
	CASE-29	23	58
ラフ集合論 + GA $(w_1=+3, w_2=-1)$	CASE-31	20	56
	CASE-32	23	68
	CASE-33	25	78
	CASE-34	21	65
	CASE-35	23	75
	CASE-36	21	66
	CASE-37	20	63
	CASE-38	20	54
	CASE-39	26	66

表-5 から、GA を用いた方法において、矛盾を出さないことを極端に重視したもの (CASE-21~29) や、判定要因を取り除くことを極端に重視したもの (CASE-31~39) では、全数比較法によるもの (CASE-1~9) や、判定要因の取り除きと矛盾の抑制のバランスを標準的な設定にした GA によるもの (CASE-11~19) と比べて、得られる極小決定アルゴリズムに含まれるルール数や評価項目数が、かえって大きくなる傾向があるということが読み取れる。

以上より、極小決定アルゴリズムの生成に必要な判定要因の組合せを得るために GA を利用する場合、判定要因を取り除くことに極端に比重を置いて、得られる極小決定アルゴリズムは矛盾を含み、正確さにかけるだけでなく、決定表の縮約化という観点においても、好ましい結果は得られにくいと考えることができる。逆に、矛盾に対する抑制を極端に強くしても、良い結果が得られるとは限らないといえる。

また、極小決定アルゴリズムのルール数が最小となる判定要因の組み合わせは、全数比較法によって導かれた CASE-2 の 11 であり、極小決定アルゴリズムの生成という観点からみれば、最終的に CASE-2 が最適解になるが、一方、GA を用いた方法の場合、残念なが

ら最適解に到達することはできなかった。しかし、評価関数の w_1, w_2 を同じ割合に設定した GA を用いても、導かれた判定要因の組み合わせによる極小決定アルゴリズムのルール数は、15 から 19 といった範囲に収まっており、これらは全数比較法によって導かれる他の組み合わせと比較しても、実用上問題のない、十分に小さいルール数であると考えられる。

6. あとがき

専門家が行った斜面の崩壊危険度の診断事例から極小決定アルゴリズムを生成するために、ラフ集合論の基本概念である類別定義に基づき、GA を利用して判定要因の組み合わせを導く方法を提案した。

類別定義に基づく極小決定アルゴリズム生成の過程に、GA を適用して明らかになったことを以下に示す。

(1) 必要性が高い判定要因の組み合わせを調べる過程において、GA を適用することによって得られた結果と、全数比較法で得られた結果を比較して、実用上十分に小さいと考えられる極小決定アルゴリズムを導くことができた。

(2) GA を適用した結果は、全数比較法によるものとよく似た傾向を示した。特に、図-2 と図-3.1 に示されるように、専門家がどの判定要因を重要視したのかを示すと考えられる判定要因の出現頻度傾向において、よく似た結果が得られた。

(3) GA の適用においては、極端に判定要因の数を減らすことを目的とする評価関数や、矛盾を必要以上に排除しようとする評価関数の設定は、かえって逆効果の場合があるという結果が得られた。

以上より、専門家の診断事例から類別定義に基づいて極小決定アルゴリズムを生成する際に、判定要因の取り除き方を GA を用いて導く方法を全数比較法の代わりに用いることができると考えられる。

また、GA を適用することに対して考えられる今後の課題を以下に示す

(1) GA を利用した方法の目的のひとつは、判定要因などデータ数が増えれば増えるほどに全数比較法では計算時間がかかるという問題を解決することである。本研究で利用したデータ数は全数比較法でも十分な量で、GA を利用することによる時間的利点は検証できなかった。今後、事例の充実をはかり、データ数が多くなった場合の GA の適用可能性について検証する必要がある。

(2) GA を適用した場合、全数比較法で得られる最適解 (CASE-2) と同じ判定要因の組み合わせを得ることができなかった。これは、CASE-2 が判定要因(3)を含むためだと考えられる。判定要因(3)は図-2、図-3.1 に示されるとおり、出現頻度が低い。このような判定要因を含む組み合わせは、本論文で提案した GA の適応度評

価法では比較的淘汰されやすいと考えられる。そのため、あまり重要視されていないと考えられる判定要因でも、すぐに淘汰される原因とならないよう、GA の個体の適応度評価法に何らかの新しい基準を組み込む必要がある。(3) 本論文では、GA の評価関数において、単純に、判定要因を取り除いた数と、矛盾が検出された数を評価項目とし、それぞれに重みをつける方法をとったが、より有効な評価方法を追求していく必要がある。

参考文献

1) 広兼道幸・古田均・中井真司・三雲是宏, 斜面の崩

壊危険度診断事例からのラフ集合を用いたルール型知識の抽出方法, 土木学会論文集, No.582/III-41, pp.285-294, 1997.

2) 古田均・広兼道幸・田中成典・三雲是宏, 橋梁の損傷要因診断事例からのラフ集合を用いたルール型知識の獲得方法, 構造工学論文集, Vol.44A, pp.521-528, 1998.

3) 松村英之, 集合論入門, 朝倉書店, pp.35-44, 1966.

4) 電気学会編, 遺伝的アルゴリズムとニューラルネットワーク, コロナ社, pp.1-34, 1998.

5) 高速道路調査会, 地滑り及び斜面崩壊の防止対策の調査手法に関する研究, pp.117-131, 1977.