あいまいさを考慮した知識獲得手法

Knowledge Acquisition Method from the Decision Table with a Quantitative Attribute

広兼道幸 *・古田均 **・原川浩一 ***・政森理恵子 ****
Michiyuki HIROKANE, Hitoshi FURUTA, Koichi HARAKAWA and Rieko Masamori

*博（工） 関西大学助教授 関西大学総合情報学部総合情報学科
**工博 関西大学教授 関西大学総合情報学部総合情報学科
***関西大学大学院総合情報学研究科
****タクトシステムズ株式会社 開発第2本部

Recently, the simplification method of the decision tables based on the rough set theory that is proposed by Pawłak in 1982 have greatly attracted attentions as one of the knowledge acquisition methods. By employing a concept of the rough set theory, it is possible to simplify a decision table and extract a set of some significant rules included in the practical cases. Though an application of the rough set theory is easy for a decision table that all attributes are classified into some categories, it is difficult to apply to a decision table with some quantitative attributes. From the point of view of applying rough set theory to the cases in the real world, the handling of the quantitative data is an important problem. In this study, we attempted to apply the rough set based simplification method to the decision table with a quantitative attribute. The numeric data for a quantitative attribute were transformed into the discrete data with the membership values by using fuzzy set theory. Moreover, the proposed method was applied to the diagnostic cases of the slope failure danger level, which contains the quantitative attribute "gradient of slope".

Key Words: *Knowledge Acquisition, Rough Set, Fuzzy Set, Slope Failure, Quantitative Attribute*

1. はじめに

知識獲得あるいはデータマイニングとは，大量のデータの中から自明ではない，潜在的に有用な知識を抽出する手法のことである[1]．ここでの知識とはデータ間の関連やそのパターンであり，その発見手法として近年注目を集めているものに，Pawłak によって提唱されたラフ集合の概念がある．ラフ集合の基本概念では，知識は対象を類別する能力に根差すものとして捉えられる[2]．我々の知的活動は，対象を様々な属性やその値に従って類別，識別することによって行われており，対象を識別するための最も単純な方法は類別に用いる属性を増やすことである．逆に，少ない属性数でもより多くの属性と同等の識別が可能であるならば，それらの属性の集合（=知識）は等価であると言える．このようなラフ集合の基本概念を利用することで，条件属性・決定属性とその値の関連（決定規則）をまとめた決定表から，その決定能力を損なうことなく不要な属性やその値を削除し，簡約化することが可能である．簡約化された決定表は，元の決定表と同等な知識を最小限のルールの集合で記述したものである．筆者らは，専門家による橋梁の健全性や斜面の崩壊危険度の診断事例からの，専門家の経験的知識の抽出にラフ集合の適用を試みてきた[3),4]．

一方で，ラフ集合の概念に基づいた決定表からの知識獲得手法においては，属性に対応する属性値はカテゴリに分類された離散的なデータであることが前提となる．測定データなどからの知識獲得を考えると，連続値の扱いは重要である．連続的な数値データを取り扱うためには，それらを何らかの方法でカテゴリ分けする必要がある．その手法は確立されていないのが現状であるが，最近では区間を設定して連続値を離散化している例が多くみられる[5),6]．しかし，専門家の診断などにおいて，数値データは必ずしも厳密にカテゴリ化されるとは限らず，柔軟な離散表現が必要な場合がある．本研究では，斜面の勾配という数値データを属性値として持つ斜面崩壊危険度診断事例をラフ集合によって簡約化し，そこに内在する専門家の経験的知識を簡潔なルールの集合として抽出することを試みた．その過程において，連続的な数値データを柔軟にカテゴリ化するためにファジィ理論[7]を適用し，帰属度を持つ離散的なデータと考えた．その結果得られた，帰属度を持つ決定規則の集合である決定表をラフ集合の概念に基づいて簡約化し，簡潔なルールの集合を得た．決定表を簡約化する際の手法には遺伝的アルゴリズム（GA: GeneticAlgorithm）[8]を用いた．

2. ラフ集合の概略

2.1 ラフ集合

ラフ集合の概念はポーランドの計算機科学者 Zdzisław Pawłak によって提案された．ラフ集合は類別による識別不能性と近似の概念による，集合論の拡張であると考え

表–1: 崩壊危険度の判定要因

判定要因	項目	DR
(1) 崩壊地の有無	1)大規模崩壊地がある	a
	2)崩壊地が多くある	b
	3)崩壊地が少しある	c
	4)崩壊地がない	d
(2) 崩壊前兆の有無	1)段落ち，亀裂，構造物の変位などの前兆がある	a
	2)上記の前兆がない	d
(3) 崖錐など不安定土塊の存在状況	1)厚く存在する	a
	2)薄く存在する	c
	3)存在しない	d
(4) 風化，変質の激しい岩の有無	1)風化，変質の激しい岩があり，上部は集水地形	a
	2)風化，変質の激しい岩があるが，1) 以外の場合	c
	3)風化，変質の激しい岩がない	d
(5) 破砕帯の有無	1)破砕帯がある	b
	2)破砕帯がない	d
(6) 斜面の勾配	数値データ	
(7) ガリーの有無	1)ある	b
	2)ない	d
(8) 斜面上沢状窪みの有無	1)沢状窪みの出口が道路より上部に位置する	a
	2)沢状部の表土，風化土が周辺部より比較的厚い	b
	3)沢状窪みはあるが，2) 以外の場合	c
	4)沢状窪みがない	d
(9) 斜面上部地形	1)凹型（集水地形）	b
	2)平型	c
	3)凸型	d
(10) 斜面の縦断形	1)オーバーハング状	a
	2)斜面途中および上部に平坦部がある	b
	3)明瞭な遷急点，遷緩点がある	c
	4)1)〜3) 以外	d
(11) 湧水の状況	1)多量の湧水がある	a
	2)汲み出し程度の湧水がある	b
	3)湧水がない	c
(12) 道路による斜面の切り取り状況	1)厚い不安定土塊を切り取っている	b
	2)風化，変質の激しい岩盤を切り取っている	c
	3)比較的新鮮な岩盤を切り取っている	d

られる[2),9)]．同値関係は類別と交換可能な概念であり，類別の代わりに同値関係を用いてラフ集合を説明することができる．

$U(\neq \emptyset)$ を関心のある対象の有限集合，R を U 上の同値関係とするとき，U/R によって U の R による全ての同値類の集合を表す．また $\mathbf{R}$ を U 上の同値関係の集合として，$\mathbf{P} \subseteq \mathbf{R}\,(\mathbf{P} \neq \emptyset)$ ならば $\bigcap \mathbf{P}$ も同値関係であり，これを $\mathbf{P}$ における識別不能関係と呼び，$IND(\mathbf{P})$ で表す．このような識別不能関係による全ての同値類の集合を知識ベース $K = (U, \mathbf{R})$ で表現する．また，同値関係と同等の属性の有限集合 $A \neq \emptyset$ を用いて，情報表現システムを $S = (U, A)$ と表現することも可能である．

$X \subseteq U$ が R のいくつかの同値類の和集合である場合，X は R-定義可能であると言い，そうでない場合，X は R-定義不可能，あるいは R-ラフであるという．ある知識ベース $K = (U, \mathbf{R})$ と $X \subseteq U$，$R \in IND(\mathbf{R})$ を仮定したとき，2つの近似集合

$$\underline{R}X = \bigcup\{Y \in U/R : Y \subseteq X\}$$
$$\overline{R}X = \bigcup\{Y \in U/R : Y \cap X \neq \emptyset\}$$

を定義し，それぞれ X の R-下近似または R-上近似と呼ぶ．$\underline{R}X$ は知識 R によって，確実に X の要素として類別できる対象の集合であり，$\overline{R}X$ は知識 R によって，X の要素として類別される可能性のある対象の集合である．

2.2 ラフ集合による決定表の簡約化

情報表現システム $S = (U, A)$ における $C, D \subset A\,(C \bigcap D = \emptyset)$ をそれぞれ条件属性，決定属性と呼ぶ．このような条件属性，決定属性を持つ情報表現システムを決定表と呼び，$T = (U, A, C, D)$ で表現する．識別不能関係 $IND(C)$，$IND(D)$ をそれぞれ条件クラス，決定クラスと呼ぶ．決定表は，ある条件クラスに分類される対象が，どの決定クラスに分類されるかという関係（決定規則）の集合である．ここで，決定表 $T = (U, A, C, D)$ において常に真である決定規則の集合は

$$POS_C(D) = \bigcup_{X \in U/IND(D)} \underline{IND(C)X}$$

によって表され，D の C への依存度すなわち決定表の矛盾の少なさは

$$\gamma_C(D) = \frac{card\ POS_C(D)}{card\ U}$$

によって表される．ラフ集合に基づいた決定表の簡約化手法においては，その依存度 $\gamma_C(D)$ を損なわないように条件属性やその値を取り除き，$POS_C(D)$ に含まれるような，与えられた決定表において常に真である決定規則を抽出することが主眼となる．また，このようにして得られる知識は，不要な条件属性や条件属性に対する値が取り除かれた，決定アルゴリズムと呼ばれる「不完全な」決定表の形で導かれる．決定アルゴリズムは通常複数の候補が得られるが，そのなかで最も縮約されたものを特に極小決定アルゴリズムと呼ぶ．

3. 斜面崩壊危険度診断事例

本研究では，ラフ集合に基づいた知識獲得手法を用いて，専門家による斜面崩壊危険度の診断事例から，そこに内在する専門家の経験的知識の抽出を試みた．この診断事例は，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[10)]に従っている．表–1 は，この診断方法を地質（粘板岩）や地形などの特性から，全ての斜面に対して選択される項目が共通となる判定要因は除外し，(1) 崩壊地の有無，(2) 崩壊前兆の有無 など，12 の判定要因にまとめたものである．各判定要因は複数の項目を持つ．例えば判定要因 (1) に対しては，1) 大規模崩壊地がある，2) 崩壊地が多くある，3) 崩壊地が少しある，4) 崩壊地がない という4つの項目があり，診断の対象となる斜面が，各判定要因に対してどの項目に該当するかを選択する．各項目には対応する危険ランク（a, b, c, d：表中の"DR"）が定められており，選択した項目に対応する危険ランク

表–2: 専門家による斜面崩壊危険度診断事例

斜面	判定要因												危険度
番号	(1)	(2)	(3)	(4)	(5)	**(6)**	(7)	(8)	(9)	(10)	(11)	(12)	
1	2	1	1	2	2	**38**	2	4	2	2	3	1	A
2	4	2	1	2	2	**37**	2	4	2	2	2	1	B
3	3	2	2	3	2	**35**	2	4	3	4	3	2	C
4	4	2	1	2	1	**27**	2	2	2	3	3	3	B
5	3	2	1	1	2	**36**	2	4	1	3	3	1	A
6	3	1	1	1	2	**39**	2	4	3	2	1	1	A
7	4	2	2	2	2	**37**	2	4	2	4	3	3	C
8	4	2	2	2	2	**38**	2	4	2	4	3	3	C
9	3	1	1	2	2	**23**	2	3	1	4	3	3	B
10	3	1	1	3	2	**22**	2	3	1	4	3	3	B
11	4	2	1	3	2	**25**	2	4	2	3	3	3	C
12	4	2	1	3	2	**23**	2	3	1	3	3	3	B
13	4	2	1	3	2	**37**	2	4	2	3	1	1	B
14	3	2	2	3	2	**35**	1	4	1	4	2	3	B
15	4	2	1	3	2	**22**	1	4	2	4	3	3	C
16	4	2	1	3	2	**37**	2	4	3	4	1	1	B
17	4	1	1	3	2	**21**	2	3	1	3	2	3	B
18	3	2	1	3	2	**39**	1	4	2	3	3	1	B
19	4	2	1	3	2	**35**	2	4	2	3	3	1	C
20	4	2	1	3	2	**28**	2	3	1	3	3	3	B
21	3	1	1	3	2	**37**	2	4	2	3	3	3	B
22	4	2	2	3	2	**35**	2	4	3	3	3	3	C
23	4	2	2	3	2	**30**	1	4	1	3	3	3	C
24	4	1	1	2	2	**19**	2	3	1	4	3	1	B
25	4	2	2	1	1	**39**	2	4	1	4	1	1	A
26	4	2	2	2	2	**38**	1	3	1	4	1	1	B
27	4	1	2	3	2	**36**	2	4	3	4	3	3	C
28	2	1	1	1	1	**40**	2	4	2	4	3	1	A
29	2	1	1	1	1	**39**	2	4	2	4	3	1	A
30	2	1	1	1	2	**35**	1	4	3	3	3	2	A
31	3	1	1	2	2	**38**	1	4	1	3	2	1	A
32	4	2	2	2	2	**39**	2	4	3	3	3	2	C

を数量化し，その合計得点により各斜面の総合的な危険度（A, B, C）を診断する．

判定要因 (6) 斜面の勾配については，斜面の勾配を数値データとして計測したものである．従来はこの数値データをその値によって4つのカテゴリに分類していた．

表–2 は，ある地域の 32 箇所の斜面について，表–1 に示される 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．例えば，斜面番号 1 は判定要因 (1) に対して項目番号 2 を選択し，判定要因 (2) に対して項目番号 1 を選択している．全ての要因に対して選択された項目から，対応する危険ランクをもとに点数制で診断した結果，斜面番号 1 の危険度は A と診断されている．この表は，12 の条件属性（判定要因）と 1 つの決定属性（危険度）の関係（決定規則）をまとめたもので，ラフ集合の概念によれば決定表と呼ぶことができる．

4. 極小決定アルゴリズムの抽出

4.1 実数データのファジィ集合化

ラフ集合の概念に基づいた決定表の簡約化手法では，扱うデータは表–1 の大部分がそうであるように，カテゴリに分類されていることが前提となる．連続的な数値データを含む決定表からルールを抽出する場合，それらをいくつかのカテゴリに分類しなければ十分な簡約化が期待できない．連続的な数値データをカテゴリ化する方法としては，データが取りうる値の範囲に，いくつかの閾値を設定する方法がもっとも簡単である．表–2 の判定要因 (6) に対する値は 19°～40° であり，例えば，25° と 35° に閾値を設定し，1)24° 以下，2)25°～35° 3)36° 以上 という 3 つのカテゴリに分類する方法が考えられる．しかし，特に測定データに基づいて専門家が判断を下す場合など，数値がどのようなカテゴリに属すかという判断にはあいまいさが含まれる場合が多い．このようなクリスプな集合モデルでは，実際の数値が僅差であってもカテゴリ分

表-3: 専門家による斜面崩壊危険度診断事例

斜面番号	判定要因												危険度	確信度
	(1)	(2)	(3)	(4)	(5)	**(6)**	(7)	(8)	(9)	(10)	(11)	(12)		
1A	2	1	1	2	2	**2**	2	4	2	2	3	1	A	0.3
1B	2	1	1	2	2	**3**	2	4	2	2	3	1	A	1.0
2A	4	2	1	2	2	**2**	2	4	2	2	2	1	B	0.5
2B	4	2	1	2	2	**3**	2	4	2	2	2	1	B	1.0
3A	3	2	2	3	2	**2**	2	4	3	4	3	2	C	0.8
3B	3	2	2	3	2	**3**	2	4	3	4	3	2	C	1.0
4A	4	2	1	2	1	**1**	2	2	2	3	3	3	B	0.6
4B	4	2	1	2	1	**2**	2	2	2	3	3	3	B	1.0
5A	3	2	1	1	2	**2**	2	4	1	3	3	1	A	0.7
5B	3	2	1	1	2	**3**	2	4	1	3	3	1	A	1.0
6A	3	1	1	1	2	**2**	2	4	3	2	1	1	A	0.2
6B	3	1	1	1	2	**3**	2	4	3	2	1	1	A	1.0
⋮						⋮							⋮	⋮
32A	4	2	2	2	2	**2**	2	4	3	3	3	2	C	0.2
32B	4	2	2	2	2	**3**	2	4	3	3	3	2	C	1.0

けでは別のカテゴリに分類されてしまうなど，そのあいまいさを表現することができない．

Zadeh によるファジィ集合[7]は，このような集合への帰属の問題に対して，メンバーシップ（帰属度）関数を導入し，対象が複数の集合へ「あいまいに」帰属することを可能にするものである．本研究では数値データを柔軟にカテゴリ化するため，ファジィ集合の概念を用いて，数値データを帰属度付きの離散値に変換した．具体的には以下の式 (1), (2), (3) に示されるファジィメンバーシップ関数によって表-2における各斜面の勾配の値をファジィ集合化した．

1) なだらか

$$\mu(x) = \int_{20}^{25} 1/x + \int_{25}^{30} \left(\frac{30-x}{5}\right)/x \quad (1)$$

2) 普通

$$\mu(x) = \int_{20}^{26} \left(\frac{x-20}{6}\right)/x + \int_{26}^{34} 1/x + \int_{35}^{40} \left(\frac{40-x}{6}\right)/x \quad (2)$$

3) 急

$$\mu(x) = \int_{30}^{35} \left(\frac{x-30}{5}\right)/x + \int_{35}^{40} 1/x \quad (3)$$

例えば，表-2において斜面番号1の勾配は38° であり，ファジィメンバーシップ関数の式 (2), (3) から，この斜面は帰属度0.3で「普通」のクラスに属し，帰属度1.0で「急」のクラスに属すことになる．同様に，すべての斜面の勾配をファジィ化した結果，表-3に一部を示す58のファジィ化された事例が得られた．ここでは，斜面の勾配の帰属度をルール全体の確信度として捉えている．これは，例えば表-3の斜面番号1Aと1Bは共に表-2の斜面番号1から得られたものであるが，これらの斜面は斜面1の勾配を帰属度0.3で「普通」，帰属度1.0で「急」であると分類したことによって生じた診断事例であると考えられるためである．条件属性と決定属性に対する値がすべて同じである斜面が複数得られた場合，それらの斜面のうち最も帰属度が高いものを残して他を削除した．

4.2 GA による極小決定アルゴリズムの抽出

斜面の勾配データをファジィ化することで得られた表-3をラフ集合の概念に基づいて簡約化し，極小決定アルゴリズムを得ることを考える．ラフ集合の概念に基づいた決定表の簡約化は，主に以下の手順によって行われる．

1. 不要な属性を見つけだし削除する
2. 重複する行を削除する
3. 属性に対する値のうち不要なものを削除する

ここで，1. 不要な属性の削除は，決定表 $T=(U,A,C,D)$ において，$POS_R(D) = POS_C(D)$ を満たす最小の条件属性の組み合わせ $R(R = C - X,\ X \subset C)$ を見つけだす作業である．R は C の D に関する縮約と呼ばれる．さらに2. で重複する行を削除した後，各決定規則の縮約を導き，3. その組み合わせで極小決定アルゴリズムを導く．このように，ラフ集合による決定表の簡約化手法では処理の初期の段階で属性の縮約を求め，その属性の組に基づいて以降の簡約化を行うという，段階的な方法がとられる．また，属性の縮約は元の属性集合から取り除けるものをすべて取り除いた，最小の属性の組み合わせとして定義されている[2]．

ここで，最終的に簡潔で有用なルールを抽出することを目的として考えたとき，取り除ける属性をすべて削除した組み合わせだけではなく，ある程度の余裕を持たせた属性の組み合わせからの簡約化も意義があるのではないかという疑問が残る．この疑問に対しては，縮約だけではなく，縮約にいくつかの余分な属性を付け加えた決定表から，それ以降の処理を行うことも可能である．しかし，ラ

フ集合による決定表の簡約化手法の計算量は属性数の増加に敏感であり，属性数の増加によって容易に組み合わせの爆発的増加を招くという問題が考えられる．筆者らは以前，このような組み合わせ探索問題の解決策の一つとして，GAを組み合わせた手法を提案した[11]．この手法は上記のような段階的な簡約化は行わず，与えられた決定表から直接的に極小決定アルゴリズムを導出するものである．GAにおける評価関数のパラメータによって，属性を簡約化する強さを変化させることも可能である．以下に，本研究における具体的なGAの適用方法を示す．

(1) 遺伝子列のコーディング

本研究では，決定表の要素を取り除くか否かを0，1の2値で表現した．0ならば取り除かず，1ならば取り除き，その結果をラフ集合の概念に基づいて評価した．1個体の遺伝子列の長さは表–3に含まれる属性値数と等しく，$58 \times 12 = 696$ビットである．1世代あたりの個体数は120個体とした．

(2) 選択

適応度の高い個体を確実に次の世代に残すため，選択演算にはルーレット選択とエリート戦略[8]を併用した．エリート数は4とした．

(3) 交叉

効果的な解空間の探索のため，交叉方法は2点交叉の変形手法を用いた．通常の2点交叉では，長さLの個体に対してm, n $(m, n \leq L,\ m < n)$の2点をランダムに決定し，その間の遺伝子列を入れ替える．これに対し，本研究ではひとつの個体のm, n間の遺伝子列をランダムにd移動し，他方の$m+d, n+d$間の遺伝子列と入れ替えた．交叉率は80%とした．

(4) 突然変異

突然変異演算は，任意の1ビットを反転するものとし，突然変異率は10%とした．

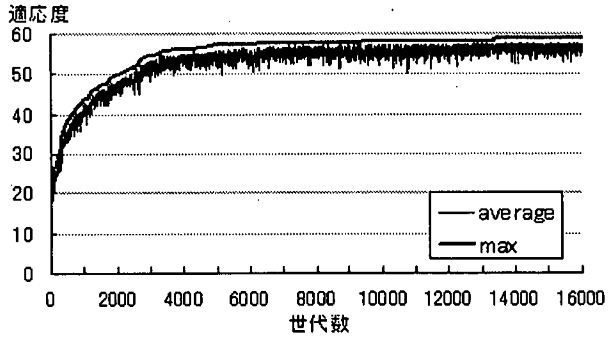
(5) 適応度評価関数

本研究では，得られた極小決定アルゴリズムが，与えられた決定表の事例をすべてカバーすることを最低条件とした．その上でより小さな決定アルゴリズムや，元の事例により多く適合するルールを含む決定アルゴリズムに高い適応度を与えるよう，以下に示す評価関数を定めた．

$$F = \left\{ R_{item} G_1 + (R_{rule} + R_{ca}) G_2 + R_{cf} G_3 + \frac{C}{10} + (1 - I) G_4 \right\} G_5^N$$

ここで，R_{item}, R_{rule}, R_{ca}はそれぞれ，その遺伝子列（決定アルゴリズム）における項目数，事例（ルール）数，条件属性数が，元の決定表に対してどのくらい減少したか，その比率を表す．R_{cf}は決定アルゴリズムに含まれるルールの帰属度の合計とその最大値の比であり，Cは決定アルゴリズムを元の事例に適応した際にヒットした事例の総数である．R_{item}, R_{rule}, R_{ca}, R_{cf}, Cを最大化することでより高い適応度が得られる．また，ペナルティとして，I, Nを与えた．Iは元の決定表に対して矛盾を起こすルール数，Nはカバーされなかった事例の数である．G_1〜G_5は重み付けの係数であり，ここではそれぞれ，1.0, 2.0, 1.0, 2.0, 0.2とした．

以上の設定で表–3からの極小決定アルゴリズムの抽出を行った．GAの収束までに必要とした世代数は16000世代であり，適応度の推移を図–1に示す．また，得られた個体群のうち，最も適応度が高い決定アルゴリズムを表–4に示す．



図–1: 適応度の推移

5. 得られた決定アルゴリズムの考察

表–4に示される極小決定アルゴリズムは，表–3に示す診断事例を全てカバーする，簡潔なルールの集合である．例えば表–4のルール1は「風化・変質の激しい岩があり，上部が集水地形であれば確信度1.0で危険度はAである」という簡潔な規則が診断事例から抽出されたことを示している．表–4は表–3と比較して，ルール数で約60%，属性に対する項目数で約92%縮小されており，元のデータと比較して大幅に少なく，簡潔なルールで専門家による診断事例を表現していると言える．また，得られた決定アルゴリズムにおいて，各条件属性がルール中に登場する頻度をまとめたものが表–5である．表–5から，判定要因(1), (4), (11)の出現頻度が比較的高いことが伺える．これらの判定要因は斜面の崩壊履歴や風化の度合い，水に関するものである．これらは斜面の特性や崩壊要因を知るために重要であると考えることができ[12]，得られた決定アルゴリズムは専門家の経験的知識の特徴を表現していると言える．一方で，実数値をその値に持つ判定要因(6)に関しては，出現頻度が1と少なかった．このことから，斜面の勾配という判定要因は斜面の危険度を判定する際に大きな影響を与えるものではなく，補助的に用いられる知識ではないか，と言う仮説を立てることも可能であろう．

6. おわりに

本研究では，数値データを条件属性として持つ斜面崩壊危険度診断事例から，ラフ集合の概念に基づいた知識獲得を試みた．数値データの離散化には，クリスプな集合ではなくファジィ集合の概念を用いて，数値を帰属度付きの離散データとして扱った．また，組み合わせ探索の手法として遺伝的アルゴリズムを用いた．その結果，決定表

表-4: 専門家による斜面崩壊危険度診断事例

ルール	判定要因												危険度	確信度
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)		
1				1									A	1.0
2		1					1						A	1.0
3	2												A	1.0
4									3	2			A	1.0
5	3								1	3			A	0.3
6					1					4			A	1.0
7					1							1	A	1.0
8								2					B	0.6
9								3					B	1.0
10				3							1		B	1.0
11		2				3	1						B	1.0
12		1		3						3			B	0.5
13							2		1			3	B	1.0
14	3	1									3		B	1.0
15	4		1	2									B	1.0
16	3		1	3									B	1.0
17				3							2		B	0.8
18	4				2						1		B	1.0
19	4							4		2			B	0.5
20	3								2				B	0.2
21	4	2	1								1		B	0.5
22			2								3		C	1.0
23	4							4	2		3	1	C	1.0

表-5: 判定要因の出現頻度

判定要因	出現頻度
(1)	10
(2)	5
(3)	4
(4)	6
(5)	3
(6)	1
(7)	3
(8)	4
(9)	4
(10)	5
(11)	7
(12)	2

に含まれる数値データを柔軟にカテゴリ化し，ラフ集合の概念に基づいた知識獲得手法を適用することが可能であった．

今後の課題として，以下のようなことが考えられる．

1) 今回利用した診断事例において，数値データとして与えられた斜面の勾配は，専門家の診断にあまり大きな影響を持たない可能性がある．そのため，ファジィ化によるルール抽出の影響が小さかったとも考えられる．今後データを拡充し，決定に対して影響が大きいと思われるデータへの適用を考える必要がある．

2) 決定アルゴリズムに含まれる決定規則に付随する確信度に対しては，現状では特に何らかの操作は考慮していない．今後の試みの一つとして，決定表の簡約化プロセスの中で，その平均値を最大化するするようにメンバーシップ関数をチューニングする方法を考える必要がある．

3)GA の遺伝子列コーディングが，事例数だけではなく，判定要因数にも依存しており，GA を用いることによるメリットが少ない．より効率的な適用方法が必要である．

参考文献

1) Pieter Adriaans, Dolf Zantinge 著, 山本英子，梅村恭司 訳：データマイニング, 共立出版, pp.6-8, 1998, 6.

2) Pawlak, Z., Rough Sets - Theoretical Aspects of Reasoning about Data, Kluwer Academic Pub., 1991.

3) 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，1998.3.

4) 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.285-294，1997.12.

5) Grzegorz Drwal, Rough, And Fuzzy-Rough Classification Methods Implemented in RClass System, RSCTC 2000.

6) Malcolm Beynon, An Investigation of β - Reduct Selection within the Variable Precision Rough Sets Model, RSCTC 2000.

7) 古田均，小尻利治，宮本文穂，秋山孝正，大野研，背野康英：ファジィ理論の土木工学への応用，森北出版，pp.1-17，1992.

8) 電気学会編：遺伝的アルゴリズムとニューラルネット，コロナ社，pp.1-61，1998.

9) 中村昭：ラフ集合 - その理論と応用, 日本ファジィ学会誌, Vol8, No.4, pp.594-603, 1996.

10) 石川芳治，古関潤一，佐々木康，三木博史，吉松弘之：土は襲う－地盤災害－，地盤工学会，pp.59-70，1995.

11) 原川浩一，広兼道幸，古田均：ラフ集合に基づく知識発見手法への遺伝的アルゴリズムの適用，ファジィシステムシンポジウム講演論文集，Vol.15th，pp.731-734 1999.

12) 奥園誠之：これだけは知っておきたい斜面防災一〇〇のポイント，鹿島出版会，pp.1-10，1986.