

斜面崩壊危険度診断事例からの支持度と条件数に基づく決定アルゴリズムの導出

関西大学 総合情報学部 学生会員 西村 文宏
関西大学 総合情報学部 正 会 員 広兼 道幸
関西大学 総合情報学部 正 会 員 古田 均
オムロンソフトウェア 原川 浩一

1. はじめに

知識獲得に関する研究の目的は、知識ベースシステムを構築する際に必要となる知識、すなわち様々な属性間の関連や重要度を発見することにあるといえる。現在の知識獲得研究における最大の関心事は、データベースなどに蓄えられた膨大なデータから、如何に知識を抽出するかという事である。ラフ集合論は、近年このような知識獲得の分野で注目されている手法のひとつである。そこで、著者らは斜面崩壊危険度診断事例からルール型知識を抽出するためにラフ集合の適用を試みた¹⁾。しかし、ここではデータサイズの増加に伴い計算時間が指数関数的に必要となる問題が残された。そこで、計算時間の問題を解決するため、個々の属性値を1つの遺伝子座と考え、遺伝的アルゴリズムとラフ集合を併用した知識獲得手法を提案した²⁾。しかし、この手法においてもデータサイズの増加に伴い、準最適解への収束が困難になるという問題が示された。そこで本研究では、これらの問題を解決するための方法について提案する。まず、専門家による斜面崩壊危険度診断事例に存在するすべての決定規則を抽出する。ここで得られる決定規則数は依然とデータサイズに依存するため、これらの決定規則の前件部に含まれる条件数、および決定規則の支持度をもとに、決定規則の絞り込みを実施した。絞り込みの結果、残された決定規則群に対して遺伝的アルゴリズムを適用して、簡潔な決定アルゴリズムの導出を試みた。

2. 決定規則の絞り込み

本研究では、決定規則の絞り込みを行う際の基準として、決定規則の条件数と支持度を用いた。決定規則の条件数は、決定規則の前件部における条件の数である。決定規則の条件数は、決定規則の簡潔さを示す指標であると考えられる。決定規則の条件数は、RSL³⁾においても、最適な決定アルゴリズムを導くための基準として用いられている。RSLにおいては、単純に短い決定規則が良いとされている。一方、決定規則の支持度は、その決定規則が、元の決定表においてどれだけ多くの事例に当てはまるかを示す指標である。支持度が高ければ、その決定規則は少なくとも元の決定表において頻繁にあてはまる規則であると言える。本研究では、表1に示す7つの絞り込み方法を用いて、簡約化した決定アルゴリズムの導出を試みた。これらの絞り込み方法は、条件数と支持度の両基準で共通とした。決定規則は、度数分布の階級ごとに取り除かれる。本研究では全事例を被覆した無矛盾な決定アルゴリズムの導出を目的とするため、ある階級に属する決定規則を取り除いた結果、残った決定規則だけでは元の事例をすべて被覆できなくなった場合、その階級に属する決定規則は取り除けないと考えた。

表1. 決定規則の絞り込み方法

番号	絞り込み方法	絞り込みの停止条件
0	絞り込みを行わない	-
1	基準値の小さいものから	はじめに取り除けない階級を発見した時点
2	基準値の大きいものから	はじめに取り除けない階級を発見した時点
3	基準値の小さいものから	すべての階級に対して検査する
4	基準値の大きいものから	すべての階級に対して検査する
5	基準値の中間値から	はじめに取り除けない階級を発見した時点
6	基準値の中間値から	すべての階級に対して検査する

3. GA による決定アルゴリズムの導出

上記の絞り込みによって得られた決定規則群を元に、GA を用いて簡潔な決定アルゴリズムの導出を行う。具

Keywords: 斜面崩壊, 知識獲得, ラフ集合, 遺伝的アルゴリズム, 条件数, 支持度

連絡先: 〒569-1095 高槻市霊仙寺町 2-1-1 関西大学総合情報学部 TEL 0726-90-2402 FAX 0726-90-2402

体的には、決定アルゴリズムは任意の決定規則の集合であると考え、絞り込みによって得られた決定規則の数を遺伝子長とした。すなわち、各遺伝子座の値は1つの決定規則に対応し、その値が1のときその決定規則を決定アルゴリズムの候補として考慮することとした。これらの遺伝子に対する評価関数は、

$$F = \frac{BF}{R_{val} \cdot R_{attr} \cdot R_{rule}} \cdot 0.2^{N_{nc}} \quad (1)$$

を考えた。ここで、 BF は基準となる適応度であり、本研究では200という値を用いた。 R_{val} , R_{attr} , R_{rule} はそれぞれ、決定アルゴリズム中の値(条件)の数(val)、条件属性数($attr$)、決定規則数($rule$)が、元の事例に対してどれだけの割合かを示すものである。決定表に残された値の数、条件属性数、決定規則数などが少ないほど簡潔な決定アルゴリズムであると評価できる。

4. 斜面崩壊危険度診断事例への適用

決定規則の条件数と支持度を基準とした、前述の7通りの絞り込み方法を、条件数12、決定規則数32である専門家による斜面崩壊危険度診断事例に適用した。条件数と支持度の2つの基準に対して、それぞれ表1に示す絞り込み方法を実施した。条件数により絞り込んだ結果を表2に、支持度により絞り込んだ結果を表3に示す。

表2. 条件数による決定規則の絞り込み結果

条件数 絞込方法	条件 属性数	決定 規則数	条件数	平均 条件数
0	6	9	18	2.0
1	9	11	33	3.0
2	6	9	18	2.0
3	10	10	30	3.0
4	7	9	23	2.6
5	8	10	18	1.8
6	9	11	21	1.9

表3. 支持度による決定規則の絞り込み結果

支持度 絞込方法	条件 属性数	決定 規則数	条件数	平均 条件数
0	6	9	18	2.0
1	8	9	21	2.3
2	10	14	37	2.6
3	11	10	29	2.9
4	10	16	46	2.9
5	8	11	24	2.2
6	10	12	29	2.4

上記の結果から、それぞれの絞込方法によって得られた結果の決定規則の数は大きく変化していることがわかる。条件数による決定規則の絞り込みにおいて、最も絞り込みの効果が高かったのは、方法2で絞り込んだ場合である。結果として得られた決定アルゴリズムは、簡潔さの点で、すべての決定規則から得られる結果に近いものとなっている。また、支持度による決定規則の絞り込みにおいて、最も絞り込みの効果が高かったのは、方法1で絞り込んだ場合である。結果として得られた決定アルゴリズムは、簡潔さの点で、すべての決定規則から得られる結果に対して多少は劣るが近いものとなっている。上記以外の方法で絞り込んだ決定規則から得られた決定アルゴリズムは、ほとんどの場合、すべての決定規則から得られた結果よりも、簡潔さの面で劣っていると言える。

5. まとめ

本研究では、RSLを用いて導出された決定規則群を、条件数と支持度の値によって絞り込み、その絞り込んだ決定規則群を元に遺伝的アルゴリズムを用いてより簡潔な決定アルゴリズムを導出することを試みた。絞り込み方法は、条件数と支持度に対してそれぞれ7種類を考慮した。その結果、条件数による絞り込みでは、条件数の多い順に取り除いていく方法が最も良い結果を示し、支持度による絞込では、支持度の低い順に取り除いていく方法が最も良い結果を示した。

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) 原川浩一・広兼道幸・古田均：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，システム最適化に関するシンポジウム，土木学会，pp.79-84，1999.
- 3) M.Gawrys', J.Sienkiewicz, ROUGH SET LIBRARY USER'S MANUAL (Ver. 2.0, September '93) : Institute of Computer Science Warsaw University of Technology.