

遺伝的アルゴリズムおよびラフ集合によるデータマイニング技術の
橋梁データベースへの適用

山口大学大学院 学生会員 ○田中信也 山口大学工学部 正会員 中村秀明
山口大学工学部 正会員 宮本文穂 (株) 長大 神波修一郎

1. はじめに

近年、橋梁維持管理の分野においても橋梁データベースが構築されてきているが、データベース全体が有効に利用されている例は少ない。そこで、近年注目を集めているデータマイニング技術(ラフ集合)を用いて橋梁データベースからのルール型知識の抽出を試みた。また、橋梁データベースのデータをコード化する上での属性のカテゴリー区分を遺伝的アルゴリズムによって最適化し、ルール型知識の抽出を試みた。

2. ラフ集合によるルール型知識の抽出

2.1 ラフ集合とは¹⁾: ラフ集合は、類似性と近似性を基本概念としている。また、大量のデータを扱うことに適している、内部知識が表形式で表現できる、ルール型知識の抽出に有効、といった利点がある。

2.2 情報システム²⁾: ラフ集合によるルール型知識の抽出を行う際、まずデータを抽象化して、情報システム(S)を考える。その情報システム(S)は以下のように表現される。

$$S = (U, A, \{V_a\}_{a \in A}, f)$$

ただし、

U : 対象の集合, A : 属性の集合, V_a : 各 $a \in A$ に対する値の集合, f は $U \times A$ から $\bigcup_{a \in A} V_a$ への写像

また、条件となる属性の集合(条件属性)を C 、結論となる属性の集合(決定属性)を D とおく。 C, D は A の部分集合である。一般に、このような情報システムを考えた場合、 ψ を C -基本式、 ξ を D -基本式とすると、 $\psi \rightarrow \xi$ を CD -決定規則と呼び、 CD -決定規則の有限集合を CD -決定アルゴリズム(決定表)と呼ぶ。

3.3 決定表の簡約化: 次に、決定表に関する重要な問題とされている、決定表の簡略化について説明する。

Step 1 重複および矛盾のない決定表の作成

決定属性 C が、条件属性 D に従属しているか否かを決定し、決定表を矛盾のない決定表と矛盾する決定表に分離する。

Step 2 不要な属性の削除

矛盾のない決定表について、条件属性の集合が不要なものを含んでいるか否かを決定する。ここでは、決定表から各属性を取り除き、残りの属性による CD -決定アルゴリズムが無矛盾であるかどうかを調べる。

Step 3 不要な属性値の削除

次に、取り除けることが可能な条件属性値を決定する。従って、抽出されたルールが、真かつ独立でなければならない。ここでは、決定表から各属性値を取り除き、その値と同じ行の残りの値がその行の決定属性を一意的に決定しているかどうかを調べる。

以上のような手順で得られたルールの集合は、極小なものとなり、極小決定アルゴリズムと呼ばれる。

3.4 ラフ集合によるルール型知識の抽出: ルール型知識の抽出に用いた入力データは、『都市部』、『山間部』、『湾岸部』の異なる 3 つの地区に架設されている橋梁データの主桁部 159 個、間詰部 77 個である。データの全属性から条件属性として表 1 の 21 項目を選択し、決定属性として『主桁判定』、『間詰部判定』を選択した。

表 2 属性のカテゴリー区分

	属性	カテゴリー区分						
		1	2	3	4	5	6	7
1	A01 架設地区	都市部	山間部	湾岸部				
2	A02 橋齢(年)	1~8	9~16	17~24				
3	A03 平面線形	直線	緩和曲線	曲線				
4	A04 車両通行方向	2方向	上り	下り				
5	A05 縦断勾配区分	上り勾配	レベル	下り勾配				
6	A06 縦断勾配(%)	0	0.1~1.0	1.1~2.0	2.1~3.0			
7	A07 横断勾配(%)	0	0.1~2.0	2.1~4.0	4.1~6.0			
8	A08 径間長(m)	10.0以下	10.1~20.0	20.1~30.0	30.1~40.0	40.1~50.0	50.1~60.0	60.1~70.0
9	A09 幅員	30.0以下	30.1~36.0	36.1~42.0	42.1~48.0	48.1~54.0		
10	A10 設計速度(Km)	60	80					
11	A11 設計容量	TL-20	TL-25	TL-43				
12	A12 桁形式	T桁	I桁	桁(床版連続)				
13	A13 連続径間数	0	1	2	3	4		
14	A14 主桁数(本)	1~2	3~4	5~6	7~8	9~10	11~12	
15	A15 主桁高(m)	1.30~1.40	1.41~1.60	1.61~1.80	1.81~2.00			
16	A16 最大主桁間隔(m)	2.30以下	2.31~2.60	2.61以上				
17	A17 床版材料	PC	RC	コンクリート				
18	A18 床版厚(mm)	160以下	161~180	181~200	201~220	221~240		
19	A19 バネ数(個)	0	1~8	9~16	17~24	25~32	33~40	
20	A20 縦コンクリート量	100以下	101~200	201~300	301~400	401~500		
21	A21 交通量(台/平日)	20000以下	20001~40000	40001~60000	60001~80000	80001~100000		
22	決定 RANK	点検判定						

また、ルール型知識の抽出を行う前処理として、これらデータをコード化することから、本研究ではコード化の上でのカテゴリー区分を作成し、コード化を行った。

4. GAs によるカテゴリー区分最適化

4.1 カテゴリー区分最適化の目的および評価関数

3.3 の Step 1 において、矛盾する決定規則をすべて削除すると、ルール抽出に重要となる決定規則も削除される可能性が生じる。本研究では、矛盾のない決定規則の数 (F) およびカテゴリー幅の合計 (G) を最大にすることを目的に各属性の最適なカテゴリー区分を GAs によって探索した。

目的関数: $F = x \rightarrow \text{Max}$

$$G = \sum_{i=1}^n y_i \rightarrow \text{Max}$$

制約条件: $1 \leq x \leq 159$ (主桁部), $1 \leq x \leq 77$ (間詰部)

$$0.625 \leq y \leq 13$$

ここで、 x : 矛盾のない決定規則の数, y_i : 各条件属性のカテゴリー幅, n : 条件属性の数, i : 属性番号

4.2 カテゴリー区分最適化の流れ

入力データは、3.4 において述べた 21 項目のうち連続値である 13 項目を選択した。選択した項目とカテゴリー区分を表 2 に示す。初期個体は、1 項目について 3 ビットの 2 進数で表すことにする。世代交代回数は、100 回、選択方法は適応度比例方式、エリート保存方式、突然変異率は 10%、交叉方法は、一点交叉を採用した。

5 出力結果

5.1 カテゴリー区分最適化の結果

3.3 の Step 1 において最適化を行っていない初期カテゴリー区分では、無矛盾なルール数が主桁部で 27、間詰部で 21 であり、カテゴリー区分最適化によって得られた矛盾のないルール数は主桁部で 81、間詰部で 34 であった。従って、カテゴリー区分最適化は成功したと考えられる。

5.2 ラフ集合によるルール抽出結果の考察

ここでは、初期カテゴリー区分と最適カテゴリー区分を用いてラフ集合によるルール抽出を行った結果(主桁部のみ)を考察する。

・初期カテゴリー区分によるルール型知識の抽出結果・・・ルール数 85

『橋齢』および『最大主桁間隔』、『設計荷重』が深く『判定』に関わっていると考えられる。

・最適カテゴリー区分によるルール抽出型知識の抽出結果・・・ルール数 155

『橋齢』、『横断勾配』、『幅員』、『最大主桁間隔』が深く『判定』に関わっていると考えられる。

また、抽出されたルールより、一般的な知識も読み取れることから、最適なカテゴリー区分を用いたルール抽出は成功したといえる。

6 結論

- ① 橋梁データベースにデータマイニング技術を適用することにより、ルールを抽出することに成功した。
- ② GAs による最適なカテゴリー区分を作成することにより、矛盾のないルール数を増やすことに成功した。
- ③ 最適カテゴリー区分による決定表を入力として、ラフ集合によるルール抽出を行った結果、ルールの傾向および一般的な知識を発見することができた。

参考文献

- 1) 中村 昭：ラフ集合 その基本概念と知識情報，数理科学 No.373, 1994
- 2) 中村 昭：ラフ集合と意思決定，数理科学 No.375, 1994

表 2 属性のカテゴリー区分

	カテゴリー区分							
	000	001	010	011	100	101	110	111
橋齢	3年毎	5年毎	7年毎	10年毎	15年毎	20年毎	25年毎	30年毎
横断勾配	0.1%毎	0.2%毎	0.5%毎	0.7%毎	1.0%毎	1.2%毎	1.5%毎	2.0%毎
横断勾配	0.1%毎	0.2%毎	0.5%毎	0.7%毎	1.0%毎	1.2%毎	1.5%毎	2.0%毎
橋間長	1m毎	2m毎	5m毎	7m毎	10m毎	12m毎	15m毎	20m毎
幅員	1m毎	2m毎	5m毎	10m毎	15m毎	20m毎	25m毎	30m毎
連続橋間数	1区間毎	2区間毎	3区間毎	4区間毎	5区間毎	6区間毎	7区間毎	8区間毎
主桁数	1本毎	2本毎	3本毎	4本毎	5本毎	6本毎	7本毎	8本毎
主桁高	0.1m毎	0.2m毎	0.3m毎	0.4m毎	0.5m毎	0.6m毎	0.7m毎	0.8m毎
最大主桁間隔	0.1m毎	0.2m毎	0.3m毎	0.4m毎	0.5m毎	0.6m毎	0.7m毎	0.8m毎
床版厚	10mm毎	20mm毎	30mm毎	50mm毎	60mm毎	80mm毎	100mm毎	120mm毎
パネル数	1個毎	2個毎	5個毎	8個毎	10個毎	12個毎	16個毎	18個毎
膨張剤入量	20kg毎	50kg毎	80kg毎	100kg毎	150kg毎	300kg毎	400kg毎	600kg毎
交通量	5000台毎	10000台毎	15000台毎	20000台毎	25000台毎	30000台毎	40000台毎	50000台毎