

# ラフ集合を用いたデータマイニング技術の橋梁データベースへの適用

株式会社 長大 正会員 神波修一郎 山口大学大学院 田中信也  
山口大学大学院 河村圭 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

## 1. 緒論

現在、橋梁維持管理の必要性が高まっており、各機関においてコンピュータによる橋梁維持管理支援ツールが開発されている。このように橋梁の維持管理をコンピュータを使って支援する場合、あらかじめ橋梁の維持管理に関する知識を知識ベースの形で整理しておく必要がある。これらの知識は、実験結果や橋梁管理に携わる専門技術者の経験から作成されたものがほとんどであるが、これらの情報のみが橋梁の維持管理に関する知識の全てであるとは考えにくい。

このような背景の下、本研究では、データマイニングによるデータベースからの橋梁部材劣化に関するルール抽出を試みた。データマイニングを行う手法としては、ラフ集合の概念を取り入れた決定表簡約化を採用した。このとき、事例データによる決定表を用いた場合、その中には例外的に発生した事例が含まれている可能性があるが、従来の決定表簡約化手法では、例外的に発生した事例は考慮されず、機械的に処理される。そこで、本研究では、例外的な事例も考慮した決定表簡約化手法を提案する。

## 2. ラフ集合の基本概念

ラフ集合の基礎概念は「類別 (classification)」と「近似 (approximation)」である。我々は外界からの複数の情報に対して知的に行動する場合、それらの情報における主語 (対象物) を属性に従って類別している。対象物の属性がこの部類分けに対して全て同じであるならば、それらの対象物は識別不可能であり、同じ物として取り扱われる<sup>1)</sup>。

ここで「橋梁」を例に挙げて説明してみる。橋梁の「主桁材料」という属性のみを類別の材料にした場合、例えばPC桁の橋梁は全て同じ物として認識される。同様に、RC桁、鋼桁の橋梁もこの場合は同じ物として認識される。この識別不能性が、ラフ集合の最も基本的な概念である。ここで、各橋梁を識別するためには、類別の材料である属性の数を増やしていけばよい。例えば上の例の場合、「主桁材料」の他に「床版材料」、「設計荷重」、「橋長」などという属性が加われば、より細かい類別が行われる。識別の材料とする属性を無限に増やしていけば、全ての橋梁は識別可能となる。

このような識別の性質を利用することにより、データベース中の膨大なデータを整理し、新たな知識が発見されると考えられる。

## 3. ラフ集合によるルール抽出手法

本研究では、ルール抽出を行う手法として、ラフ集合の概念を利用した決定表簡略化を採用した。

決定表簡略化の一般的な手法 (手順 1) を以下に記す<sup>2)</sup>。

### 手順 1

#### [STEP-1] 無矛盾な決定アルゴリズムの作成

与えられた決定表の各決定規則を順に調べ、矛盾するものが発見された場合、その決定規則を排除する。

#### [STEP-2] 同一規則のない決定アルゴリズムの作成

複数の同一決定規則が存在する場合、その中から一つだけ決定規則を残し、残りの決定規則は決定表から排除する。

#### [STEP-3] 条件属性集合の縮約化

様々な条件属性集合を取り除いていき決定表が無矛盾となる場合、その条件属性集合を不要なものとして排除する。

#### [STEP-4] 決定規則の縮約化

STEP-3 で縮約化された属性集合への射影による決定表において、ある条件属性値を取り除くことによりその決定規則が無矛盾かつ独立である場合、その条件属性値を不要なものとして排除する。

ここで、ある問題が考えられる。それは、「矛盾する決定規則の決定」にある。ルール抽出を行う決定表として事例データによるものを用いた場合、その中には例外的に発生した事例が存在する可能性がある。手順 1 の場合は、このような例外的な規則は考慮されずに決定規則の矛盾・無矛盾が決定されるため、最終的に適当なルールが抽出されない可能性がある。

そこで、本研究では、例外的規則を考慮した決定表簡略化手法 (手順 2) を考案した。以下に手順 2 を示す。

### 手順 2

#### [STEP-1] 例外的規則を考慮した無矛盾な決定アルゴリズムの作成

矛盾する決定規則が発見された場合、その決定規則が例外的なものであるか否かを判定する。具体的には、その決定規則と同一の決定規則の数を  $x$ 、それらの決定規則に対して矛盾する残りの決定規則 (すなわち、発見された矛盾する決定規則と

条件属性値は全て一致するが、決定属性値が異なる決定規則)の数を  $y$  としたとき、次の関係が成り立つか否かを計算する。

$$x / (x+y)$$

ここで、 $x$  は、決定規則が矛盾するものであるか否かを判断するための閾値である。これが成り立つ場合、発見された矛盾する決定規則は無矛盾であると判断される。この作業を全ての決定規則について行い、無矛盾な決定表を作成する。

[STEP-2] 同一規則のない決定アルゴリズムの作成

手順1のSTEP-2と同様。

[STEP-3] 例外的規則を考慮した条件属性集合の縮約化

属性の部分集合による決定表が矛盾であると決定された場合、例外的規則を考慮しつつその決定表の各決定規則が矛盾するか否かを決定する。つまり、STEP-1と同様の作業を行う。そして、例外的規則を考慮した結果、全ての決定規則が無矛盾であったとき、その決定表は無矛盾と判断される。この作業を、条件属性の全ての組み合わせによる部分集合について行う。

[STEP-4] 決定規則の縮約化

STEP-3で縮約化された属性集合への射影による決定表は、矛盾な場合があるため、この決定表に対してSTEP-1と同様の作業を行い、無矛盾な決定表を作成する。以降の作業は、手順1のSTEP-4と同様である。

#### 4. 実橋データへの適用

##### (1) 入力データ

先に記した手順1, 手順2の手法を用いて、実際にルール抽出を試みた。その際の入力データとなるデータ群には、高速道路高架橋に関する「資産データ」、「定期点検データ」を採用した。資産データは、各橋梁に関する環境、構造などの諸元的データであり、定期点検データは、各橋梁の定期点検時における部材別総合判定である。これらは、実際に使用されているデータベースに格納されているデータの一部である。

これらのデータから、ルール抽出の際に使用する決定表を作成した。本研究では、条件属性に「架設地区」、「橋齢」、「平面線形」、「車両進行方向」、「縦断勾配区分」、「縦断勾配」、「横断勾配」、「径間長」、「幅員」、「設計速度」、「設計荷重」、「桁材料」、「桁方式」、「桁形式」、「連続径間数」、「主桁数」、「主桁高」、「最大主桁間隔」、「床版材料」、「床版厚」、「総コンクリート量」、「交通量」の22項目、決定属性に定期点検の各部材における「総合判定」を採用した。そして、これらの属性のデータをデータ群の中から引き出し、それぞれの属性に対してカテゴリを作成しコード化を行った。引き出された事例は、主桁総合判定に関するものが159個、間詰め総合判定に関するものが77個である。

##### (2) 出力結果

表1は主桁総合判定および間詰め総合判定に関するSTEP-3までの出力結果である。表1より、手順2の方が少ない属性数の条件属性集合を出力し、たくさんの条件属性集合を出力したことが分かる。つまり、手順2の方が手順1に比べ、効果的に条件属性の縮約を行ったといえる。

また、最終的に抽出された多数のルールより、次のような傾向が読み取れた。

最大主桁間隔が小さいほど、主桁は健全である。架設地区が都市部よりも山間部のほうが、間詰めは健全である。

交通量が少ないほど、間詰めは健全である。横断勾配が小さいほど、主桁は健全である。

平面線形が曲線である場合、主桁は健全である。このうち、 $x$  は、一般的に知られているような知識である。このような一般的な知識も出力されたことより、ルール抽出は正常に行われたのではないかと考えられる。

表1 STEP-3での出力結果

	主桁総合判定		間詰め総合判定	
	手順1	手順2	手順1	手順2
縮約化された条件属性集合数	14975	36605	64874	69773
縮約化された条件属性集合中の最小属性数	7	6	3	2

#### 5. まとめ

本研究で得られた成果を以下に記す。

矛盾する決定規則の数を利用して、注目している決定規則が例外的に発生したものの有無を判断することにより、その決定規則が矛盾なものであるか否かを決定する決定表簡約化手法を新たに考案した。

実際に供用されている橋梁の事例データをよりルール抽出を行った結果、条件属性集合簡約化の段階において、例外的規則を考慮することにより、より効果的に条件属性集合の縮約化が行われたといえる。

決定表の簡略化の最終的な出力である極小決定アルゴリズムが出力された。この極小決定アルゴリズムより、一般的に知られている知識も読み取ることができたので、ルール抽出は正常に行われたといえる。

参考文献 1)吉田琢也：情報の脈を掘り当てるデータマイニングいよいよ本番，日経コンピュータ p70，1996.9.16

2)横森貴，小林聡：ラフ集合と意思決定，数理科学 No375，サイエンス社，p76，1994.9