

ラフ集合を用いた道路橋鋼桁における損傷ルールの獲得

山口大学工学部	学生会員	○脇本文
山口大学工学部	正会員	河村 圭
山口大学工学部	正会員	宮本 文穂
山口大学工学部	正会員	中村 秀明

1. まえがき

近年の情報処理技術の発展に伴い、データ蓄積方法は、従来の紙主体から、大量のデータを貯蓄可能なデータベースシステムへと移行してきている。このような背景から、大量のデータの中から有用な知識を発見することを目的としたデータマイニング技術が注目を集めている。そこで著者らは、橋梁の維持管理データベースから、ラフ集合の概念を用いた決定表簡約化手法を用いてルール型知識の抽出を行ってきた。しかしこの手法には矛盾データの取り扱いに起因する問題点があり、有用な知識の抽出に弊害をもたらしてきた。本研究ではその問題点の解決方法の提案を行う。

2. ラフ集合

ラフ集合とは、類似と近似を基本概念とし、ものごとをあらかく扱うための集合論である。ラフ集合では、関心のある属性の値に従って対象と他のものごとを識別する。関心のある属性の値が同じであれば同じものとして扱ってよい。この識別不能性がラフ集合の最も基本的な概念である。

3. 決定表簡約化手法¹⁾

決定表とは、ある条件下の元でどのような決定を下すべきかをまとめ上げたものである。主に複雑な条件判定を伴う意思決定に有効な道具として多くの分野で活用されている。

決定表にかかわる多くの応用領域において、与えられた決定表を簡約化するという問題は最も重要な話題である。簡約化の一例として、決定表における条件属性の簡約化が挙げられる。それにより簡約化された決定表においては、より少ない条件に基づいて簡約化前と同一の決定を下すことができる。

以下に、決定用簡約化手法の手順を示す。

決定表の簡約化手順

[STEP1] 矛盾している決定規則を削除する

同一の条件属性下で、異なる決定属性を持つ複数の決定規則が存在する場合、これら全てを削除する。

以下のSTEPは矛盾のない決定規則のみについて行われる。

[STEP2] 同一決定規則をまとめる

次に、1つの決定表の中に複数の同一な決定規則が存在する場合、1つの決定規則にまとめる。

[STEP3] 必要のない属性を削除する

続いて、決定表の分類能力を壊すことなしに削除できる条件属性を削除する。

[STEP4] 必要のない属性値を削除する

最後に、決定表の分類能力を壊すことなしに削除できる条件属性値を削除する。

この流れにより、得られた極小決定アルゴリズムは、if-then ルールとして表される。

4. 決定表簡約化手法の問題点と解決方法

決定表の簡約化が行われる際、矛盾する決定規則は[STEP1]において、不要な規則として削除されることを前節で述べた。そのため[STEP2]以降のルール抽出作業では無矛盾な決定表のみが使用され、矛盾を含んだ決定規則は全く考慮されないことになる。この結果、ルールの抽出作業に一部のデータしか使用されないことになり、簡約化される前の決定表にそぐわないルールが抽出される事がある。

そこで本研究では、元のデータに忠実なルールの抽出を可能にするため、①[STEP1]で削除されるデータの減少、②条件属性を簡約化する際に矛盾のある決定表も考慮に入れることを考えた。具体的な方法を以下に示す。

① [STEP1]において、発見された矛盾する決定規則が例外的なものでないか調べる。具体的には、決定表中に存在する矛盾する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の数を y とし、次のことが成り立つか否かを計算する。

$$x/(x+y) \geq \alpha$$

ここで α は、決定規則が矛盾するものであるかのしきい値である。(本研究では $\alpha=0.8$ と設定した)。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。

- ② [STEP1]において、矛盾とみなされた決定規則は矛盾のない決定表とは別に残しておく。

[STEP4]において、この完全に矛盾した決定表を参照しながら削除を許可するか否か判断する。具体的には、従来の手法で矛盾のない決定表から省略する属性値を決めた後、矛盾のない決定表、完全に矛盾した決定表を見て、両決定表中に存在する矛盾する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の数を y とし、①の式に当てはめる。これが成り立った場合にかぎり省略を認め、成り立たなかった場合省略を許可しない。

5. 実データを使った各手法の比較

旧手法と提案手法との比較のために、道路高架橋の鋼桁に関する維持管理データを用いた。今回は特にさびおよび腐食の損傷進行度に関する決定表を各手法によって簡約化した場合を比較する。

損傷判定を決定属性とし、条件属性には『橋齢』『桁種別』『床板種別』『床板厚』『路下条件』等、損傷に関連のあると思われるものを計 14 種用いた。

各手法の評価の方法として、抽出されたルールを評価し、それを比較することを考える。抽出ルールの評価は、抽出ルールを簡約化前の決定表へと当てはめ、次に示す信頼度、支持度を求めることで行った。

信頼度は抽出ルールがどれほど正確に意思決定を行っているかを数値化したもので、

$$\text{信頼度} = \frac{\text{すべての属性値が一致した決定規則の数}}{\text{条件属性値がすべて一致した決定規則の数}} \times 100$$

であらわされる。また、支持度は抽出ルールにどれほどの汎用性があるかを数値化したものであり、

$$\text{支持度} = \frac{\text{すべての属性値が一致した決定規則の数}}{\text{対象となった決定表に含まれる決定規則の総数}} \times 100$$

であらわされる。

旧手法と提案手法の信頼度、支持度の度数を図 1、表 1 に示す。支持度は 1% を切るような低い値が多いため、表の形で示した。

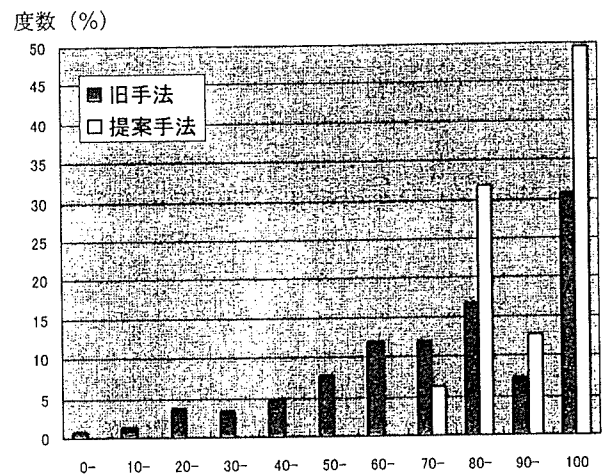


図 1 信頼度による度数の分布

表 1 支持度の比較

支持度(%)	旧手法(%)	提案手法(%)
0以上1未満	77.07	80.55
1以上2未満	12.89	12.01
2以上3未満	5.13	3.99
3以上4未満	2.32	1.71
4以上5未満	1.14	0.67
5以上6未満	0.66	0.54
6以上7未満	0.44	0.22
7以上8未満	0.22	0.19
8以上9未満	0.09	0.06
9以上10未満	0.04	0.06
平均支持度(%)	0.0074965	0.0064511

図 1 から、提案手法に現れるルールの多くは高い信頼度を示していることがわかる。また表 1 から、提案手法での抽出ルールは旧手法にくらべ支持度が低くなっていることがわかる。

6. まとめ

本研究では、ラフ集合の概念による決定表簡約化手法で問題となっていた矛盾するデータの扱いの改善を目的として、決定表簡約化手法のプロセスの変更を行った。

以下に本研究で得られた成果を示す。

- ① 既存の手法では不必要なデータとして除外されていたデータを使用し、簡約化に規制を作ることによって信頼度の高いルール抽出に成功した。
- ② 信頼度を意図的に向上させるには、支持度が犠牲になることが判明した。

[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 No373，サイエンス社，1994。