

ラフ集合における矛盾データからルールを抽出する手法の橋梁点検データへの適用

(株)リョーセンエンジニアズ 正会員 ○八木 英樹
(有)ミツワ電器 正会員 江本 久雄

阪神高速技術(株) 正会員 塚本 成昭
山口大学大学院 フェロー会員 宮本 文穂

1. はじめに

近年、老朽化した橋梁の増加により劣化・損傷が発生する橋梁構造物が増加していくことが予想される。¹⁾このような背景で、既存橋梁の長寿命化を目的とした効率的な維持管理対策が求められている。その維持管理対策のひとつとして点検データからの損傷要因の抽出が期待されており、著者らはデータマイニング手法のラフ集合論²⁾を用いて橋梁の損傷要因を抽出しようと試みてきた。ラフ集合では矛盾したデータはルールとして抽出されないが、実際の点検データには矛盾したデータが多く含まれる場合があり、中にはデータの半分以上が矛盾している場合もある。これらの大量の矛盾データが活用されずに捨てられていることに着目し、本研究では矛盾データの多数派を救済する機能(矛盾救済機能)を提案し、その計算結果について検討を行ったものである。

2. データマイニング

ラフ集合論では、同値関係、類似関係などによる集合を知識と考え、与えられた集合を上近似と下近似の2つの近似方法により表現する。上近似とは可能的に近似を考え、下近似とは必然的に近似を考える手法である。この上近似と下近似の対をラフ集合といい、これによってあいまいな情報を近似すべきであるという考えが、ラフ近似と呼ばれる。これは、If-Then ルールで表わすことができ、対象の集合を粗く表現でき、非線形データや非数値データに強いという特徴を持つ。

3. 矛盾救済機能の提案

本研究では、矛盾データを多く持つ決定表を用いた場合のラフ集合の処理時に起こる問題について述べ、その問題を解決するための矛盾救済機能を提案する。

3.1 矛盾データを持つ決定表とその結果

矛盾データとは、条件属性が同じであるが、決定属性が異なるデータのことであり、表1に示す矛盾データを持つ決定表から{損傷の有無=損傷あり}となるデータマイニング結果の評価指標である被覆度(C.I.)値を求めると、表2が得られる。

表1 矛盾データを持つ決定表の例

ID	損傷の有無	床版種別	桁種別	直線・曲線主桁	桁区分
1	損傷あり	RC	H-1桁	直線主桁	鋼桁
2	なし	鋼床版	H-1桁	曲線主桁	鋼桁
3	なし	PC	箱桁	曲線主桁	PC桁
4	損傷あり	RC	箱桁	直線主桁	PC桁
5	なし	RC	箱桁	曲線主桁	PC桁
6	損傷あり	鋼床版	H-1桁	曲線主桁	鋼桁
7	損傷あり	鋼床版	H-1桁	曲線主桁	鋼桁
8	損傷あり	鋼床版	H-1桁	曲線主桁	鋼桁
9	なし	PC	箱桁	曲線主桁	PC桁
10	なし	PC	箱桁	曲線主桁	PC桁
11	損傷あり	PC	箱桁	曲線主桁	PC桁

表2 従来の手法によるC.I.値

【損傷の有無】(【損傷あり】となるルール)	床版種別	桁種別	直線・曲線主桁	桁区分	C.I.	1	4	6	7	8	11
直線・曲線主桁 = 直線主桁			直線主桁		2 / 6	0.33333	*	*			
床版種別 = RC and 桁種別 = H-1桁	RC	H-1桁			1 / 6	0.16667	*				
床版種別 = RC and 桁区分 = 鋼桁	RC			鋼桁	1 / 6	0.16667	*				

このとき、表1のID{6,7,8,11}は矛盾しているデータであるためルールに当てはめることができない。これにより有用なデータの消失が起っていることが考えられる。

3.2 矛盾救済機能の基礎概念

矛盾救済機能とは、3.1で述べたような決定ルールを当てはめることができないデータを決定ルールに当てはめることを可能にする機能のことである。矛盾救済機能の処理手順は以下のとおりである：

① データの抽出比率の設定

矛盾データの決定属性の比率が何%以上のデータを抽出するかを決定する。抽出比率>50%であること。

② 矛盾しているデータの精査

矛盾データを検出し、決定属性の比率を計算する。

③ 救済されるルールの決定

決定属性の比率が①で定めた抽出比率を上回る矛盾データを救済する。この際、決定属性の比率の高い属性値(=多数派)を救済するため、決定属性の比率の低い矛盾データ(=少数派)は削除する。

④ データマイニング

①~③により新たに作成されたデータで計算する。

キーワード データマイニング、ラフ集合論、矛盾、点検データ、橋梁、維持補修

連絡先 〒733-0036 広島市西区観音新町1-20-24 リョーコー・センタービル (株)リョーセンエンジニアズ 構造解析課 TEL 082-294-2651

3. 3 矛盾救済機能を適用したデータマイニング

前述表1のデータに対して矛盾救済機能を適用してデータマイニングを行う。この際の抽出比率は70%と仮定する。まず、矛盾データの精査を行う。表1より表3の矛盾データが得られる。

表3 矛盾データの精査の例

ID	条件属性				決定属性		
	床版種別	桁種別	直線・曲線主桁	桁区分	損傷の有無	レコード数	比率
2	鋼床版	H-1桁	曲線主桁	鋼桁	なし	1	25%
6,7,8					損傷あり	3	75%
3,9,10	PC	箱桁	曲線主桁	PC桁	なし	3	75%
11					損傷あり	1	25%

表3より、表1のID{2, 6, 7, 8}のデータを見ると、決定属性{損傷の有無=損傷あり}の比率が75%となり抽出比率70%を上回っている。よって多数派のID{6, 7, 8}を救済し、少数派のID{2}は削除する。同様に、ID{3, 9, 10, 11}のデータを見ると、決定属性{損傷の有無=なし}の比率が75%となり抽出比率70%を上回っている。よって多数派のID{3, 9, 10}を救済し、少数派のID{11}は削除する。矛盾データの精査により新たに表4の決定表が得られ、表4から{損傷の有無=損傷あり}となるタマイニング結果の評価指標である被覆度(C.I.)値を求めると表5の結果が得られる。

表4 矛盾データの精査を行った決定表の例

ID	損傷の有無	床版種別	桁種別	直線・曲線主桁	桁区分
1	損傷あり	RC	H-1桁	直線主桁	鋼桁
3	なし	PC	箱桁	曲線主桁	PC桁
4	損傷あり	RC	箱桁	直線主桁	PC桁
5	なし	RC	箱桁	曲線主桁	PC桁
6	損傷あり	鋼床版	H-1桁	曲線主桁	鋼桁
7	損傷あり	鋼床版	H-1桁	曲線主桁	鋼桁
8	損傷あり	鋼床版	H-1桁	曲線主桁	鋼桁
9	なし	PC	箱桁	曲線主桁	PC桁
10	なし	PC	箱桁	曲線主桁	PC桁

表5 矛盾救済機能を適用した場合のC.I.値

損傷の有無及び損傷ありとなるルール	床版種別	桁種別	直線・曲線主桁	桁区分	C.I.	1	4	6	7	8
桁種別 = H-1桁		H-1桁			4 / 6	0.666667	*		*	*
直線・曲線主桁 = 直線主桁			直線主桁		2 / 6	0.333333	*	*		
桁区分 = 鋼桁				鋼桁	4 / 6	0.666667	*		*	*
床版種別 = 鋼床版	鋼床版				3 / 6	0.5			*	*

表1のID{6, 7, 8}について、表2では決定ルールに当てはめることができなかったが、表5では矛盾救済機能を適用したことにより決定ルールに当てはめることが可能となった。以上の提案手法によって、今まで矛盾データとして決定ルールに当てはめることができずに消失していたデータを有効活用することが可能になると考える。

4. 実データでの検証

ある都市高速道路の橋梁支承に関する大量の点検データを

用いて検証を行った結果を表6および表7にまとめて示す。これより、矛盾救済機能により大量の矛盾データを救済できていることがわかる。

表6 実データ検証結果(矛盾救済なし)

〔支保, 損傷の有無及び損傷=コンクリートとなるルール〕

No.	支保					上部工				支持度	C.I.値		
	竣工年度	梁上区分	型式	固定可動	呼び寸法	桁区分	床版種別等	桁種別	ハ字・斜形状		直線・曲線主桁		
1	A							H-1桁	ハ字形状	曲線主桁	0.007492	71 / 3592	0.019921
2	A					鋼桁			斜形状		0.006796	55 / 3592	0.015265
3	A						RC		斜形状		0.006796	55 / 3592	0.015265
4	A								斜形状		0.006595	53 / 3592	0.014796
5									斜形状	直線主桁	0.006375	51 / 3592	0.014238
6	A								斜形状	直線主桁	0.006068	48 / 3592	0.0134
7	A							H-1桁	直線主桁	曲線主桁	0.004742	45 / 3592	0.012563
8		単橋桁部等						RC	直線主桁	ハ字・斜形状	0.004742	45 / 3592	0.012563
9	A							RC	直線主桁	直線主桁	0.004742	45 / 3592	0.012563
10	A								斜形状	斜形状	0.004637	44 / 3592	0.012294

表7 実データ検証結果(矛盾救済あり, 抽出比率70%)

〔支保, 損傷の有無及び損傷=コンクリートとなるルール〕

No.	支保					上部工				支持度	C.I.値		
	竣工年度	梁上区分	型式	固定可動	呼び寸法	桁区分	床版種別等	桁種別	ハ字・斜形状		直線・曲線主桁		
1	A							RC		直線主桁	0.112876	1016 / 3592	0.28364
2	A		単橋桁部等							直線主桁	0.086657	780 / 3592	0.217755
3	A						鋼桁			直線主桁	0.086324	777 / 3592	0.216918
4	A							RC		なし	0.061215	551 / 3592	0.153825
5			単橋桁部等							ハ字形状	0.048329	435 / 3592	0.121441
6	A								H-1桁		0.047884	431 / 3592	0.120224
7	A							RC		ハ字形状	0.047772	430 / 3592	0.120045
8	A		単橋桁部等							なし	0.046329	417 / 3592	0.115415
9	A									なし	0.046106	415 / 3592	0.115267
10	A								H-1桁	ハ字形状	0.041951	374 / 3592	0.104411

5. まとめ

以下に、本研究で得られた成果をまとめる：

- ① ラフ集合に矛盾救済機能を適用することで、消失していた矛盾データの救済を可能にした。
- ② 実データによる検証においても消失していた矛盾データが救済できており、有効性が確認できた。
- ③ 抽出比率を低くすることにより多くの矛盾データを救済可能であるが、抽出比率を低くしすぎると誤った結果を抽出してしまうことが考えられる。

謝辞

本研究を行うにあたって、多大な協力をして頂いた元山口大学工学部 安藤君に感謝致します。

参考文献

- 1) 道路橋の予防保全に向けた有識者会議：「道路橋の予防保全に向けた提言」について
<http://www.mlit.go.jp/common/000015315.pdf>
- 2) 江本久雄, 高橋順, 宮本文穂：ラフ集合を適用したデータマイニング汎用ソフトの開発と社会基盤構造物の維持管理への利用, 2009. 2