

知識抽出のための橋梁点検データへのデータマイニング技術の適用

山口大学工学部 ○三角俊介
(有) ミツワ電器 正会員 江本久雄

(株) リョーセンエンジニアズ 絹谷一郎
山口大学工学部 正会員 河村 圭

1. はじめに

近年、老朽化した橋梁の増加が危惧されており、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¹⁾が期待されている。

そこで、本研究では、データマイニング技術を用いて鋼橋の点検データから鋼橋の損傷要因に関するルール抽出を行う。データマイニングの応用例としては、医療、財務、品質管理、マーケティング、経済予測など、あらゆる方面で利用されている。ここで、データマイニング技術として、決定木とラフ集合理論を橋梁の点検データからのルール抽出に適用し、それぞれの手法の特徴を明らかにした。両手法とも専門家により作成されたサンプルデータと実際の点検データに適用し、それぞれの妥当性、特徴と問題点について検討を行った。

2. データマイニング手法の概要

データマイニング²⁾とは、データから以前には知られていない有用な知識(ルール)を導き出すプロセスのことである。

2.1 決定木

回帰分析のような明示的な関数を用いずに、一連の手順に沿ってデータを仕分けしていくことによって予測や判別を行うことを決定木という。

決定木は、訓練例集合を使い、学習過程を経て構築される。構築された決定木は、分類クラスの知られていない新たな例を分類するために使われる。決定木の代表的なアルゴリズムとしては、「C5.0」、「CART」などがある。

2.2 ラフ集合

ラフ集合では、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。

3. 点検データの概要

以下に、今回マイニングに用いた橋梁の点検データの概要について述べる。

3.1 サンプルデータの概要

サンプルデータは、条件属性数を10、サンプル数を84とし、その条件属性は表1のように設定した。また、条件属性中の最大主桁間隔、最小主桁間隔と塗装面積は、連続値であるので、表2に示すようにカテゴリーを分けた。さらに、決定属性としては「損傷の有無」とし、その値は、「有り」・「無し」の2値である。なお、このデータは、専門家によって実験用に作成された架空の特徴的なサンプルデータである。サンプルデータの主な特徴としては、以下のように設定した。

- ① 塗装種別で、塩化ゴム系の損傷はまれとした。
- ② 主桁では、主桁数が多いほど損傷が多いとした。
- ③ 塗料という項目を作り、任意に値を与えた。

3.2 実データの概要

実データは、サンプル数が65、属性項目数が18とし、そのうち「内面」、「損傷部材」、「個別判定」の1つを決定属性として用い、残りの2つは条件属性とした。「内面」は桁内面に損傷が発生しているかどうか、「損傷部材」は

表1 サンプルデータ

ID	主桁数(本)	最大主桁間隔 ^(*)	最小主桁間隔 ^(*)	区線・曲線主桁	縦断勾配(%)	裏面板等	上部工立地条件/路下	塗装種別	塗装面積 ^(*)	塗料
p1	2	4	3	曲線主桁	0.3	化粧板(側面板あり)	河川	塩化ゴム系	3	A
p2	3	3	2	直線主桁	0.3	橋防止ネット	河川	ポリウレタン樹脂	5	B
p3	2	4	3	直線主桁	0.3	橋防止ネット	公園用地	ポリウレタン樹脂	4	B
~										
p83	2	3	3	直線主桁	3.5	なし	その他用地	塩化ゴム系	3	C
p84	2	2	2	曲線主桁	0.4	化粧板(側面板あり)	一般道路	塩化ゴム系	3	A

表2 最大主桁間隔、最小主桁間隔、塗装面積のカテゴリー分けの一覧

最大主桁間隔	(m)	最小主桁間隔	(m)	塗装面積	(m ²)
1	0 ~ 1	1	0 ~ 1	1	0 ~ 100
2	1 ~ 5	2	1 ~ 5	2	101 ~ 500
3	5 ~ 10	3	5 ~ 10	3	501 ~ 1000
4	10 ~ 15	4	10 ~ 15	4	1001 ~ 2000
(*)1	5 : 15 ~	(*)2	5 : 15 ~	5	2001 ~

表 3 実データの抜粋

ID	勾配(%)	桁種別	直線・曲線主桁	主桁数本	...	塗装系	塗料	経年数年	損傷部材	内面	個別判定
p1	0.1	I桁	直線・曲線主桁	8	...	a-2	A	35	鋼桁本体	F	C
p2	0.1	I桁	直線・曲線主桁	8		a-2	A	35	鋼桁本体	F	C
p3	0.1	I桁	直線・曲線主桁	8		a-2	A	35	鋼桁端部	F	B
~											
P63	0.4	H桁	直線主桁	5	...	a-2	A	33	鋼桁端部	F	C
P64	0.4	H桁	直線主桁	5		a-2	A	33	鋼桁端部	F	C
P65	0.4	I桁	直線主桁	6		a-2	A	33	鋼桁端部	F	C

「鋼桁本体」と「鋼桁端部」のどちらに損傷が発生しているか、「個別判定」は発生している損傷の度合いを「A」～「C」で示している。損傷の度合いは「A」の方が重度の損傷であり、「C」の方が軽度の損傷であることを示している。また、「内面」と「損傷部材」をひとまとめにした「内面\本体」も決定属性とした。その際、「内面」と「損傷部材」は条件属性から除いた。実データの一部を表 3 示す。属性項目は「勾配」、「桁種別」、「直線・曲線主桁」、「主桁数」、「主桁高さ」、「最大主桁間隔」、「最小主桁間隔」、「最大床版支間」、「裏面板等」、「上部工立地条件/路下」、「設計速度」、「塗装種別」、「塗装系」、「塗料」、「経年数」、「損傷部材」、「内面」、「個別判定」である。このうち、「設計速度」はダミー属性であり、属性値は1つしか与えられていない。また、実データはすべて損傷が発生しているデータである。

3.3 適用結果とその考察

本研究では、決定木とラフ集合の両手法をサンプルデータ、実データにそれぞれ適用しその結果の比較を行った。以下に、適用結果とその考察をまとめる。

- ①決定木でサンプルデータをマイニングした結果、さび・腐食が起りやすい橋梁条件が「主桁数=3」、「塗料=A or B」、「直線・曲線主桁=直線主桁」であることが分かった。反対に、さび・腐食の起りにくい橋梁条件として「主桁数=2」、「直線・曲線主桁=曲線主桁」、「塗装=C」であることが分かった。
- ②決定木で実データをマイニングした結果、経年数が「33年以下」と「33年より大きい」で分類されている。これは、鋼桁本体よりも鋼桁端部の方が短い期間で損傷が発生していることを示しており、鋼桁端部桁端伸縮装置からの漏水が影響していると考えられる。漏水箇所が明確にできれば、解析結果が有用であったと裏付けできるため、今後、点検結果の詳細な分析を行う必要がある。
- ③ラフ集合でサンプルデータをマイニングした結果、さび・腐食が起りやすい橋梁条件が「主桁数=3」、「塗装

種別=ポリウレタン樹脂」、「塗料=B」であることが分かった。反対に、さび・腐食の起りにくい橋梁条件として「主桁数=2」、「塗装面積=2 (101~500 m²)」、「塗装会社=C」であることが分かった。

④ラフ集合で実データをマイニングした結果、「損傷部材が鋼桁端部であるならば損傷が発生している」となるルールで、塗装種別の「塩化ゴム系」と「ポリウレタン樹脂」の両方が損傷に影響を与えていることが分かった。塗装種別によって損傷の発生頻度に違いがあると予測していたため、新たな知見である。

⑤決定木とラフ集合ともにサンプルデータを解析した結果、さび・腐食が起りやすい橋梁条件・起りにくい橋梁条件の両方に共通した複数の条件属性が抽出された。このことより、両手法ともルール抽出に適用可能だと考えられる。また、条件属性に塗料が含まれているということは新しい発見である。

⑥実際の点検データに決定木を適用した結果、ツリーが枝分かれせずルール抽出が不可能となる場合があった。これは、データに特徴がないものと考えられる。一方で、同じ条件でラフ集合に適用した結果、ルールの抽出が行え、ラフ集合の適用が有効であることが分かった。

4. まとめ

本研究では、橋梁の点検データからのルール抽出を行った。以下に、本研究で得られた成果をまとめる。

- ①本研究で用いた両データにおいて、橋梁の損傷要因となるルールが得られ、両手法とも適用が有効であることが分かった。
 - ②決定木では分類できない特徴の無いデータからでも、ラフ集合では縮約およびルールを抽出することができ、ラフ集合の適用が有用であることが分かった。
- 今後の課題としては、データの性質により、ルールが抽出できない場合があったので、種々のデータに適用を行い、さらに検討を行う必要がある。

謝辞

データを提供して頂いた阪神高速技術(株) 加賀山泰一氏に深く感謝致します。

参考文献

- 1)古田均, 広兼道幸, 田中成典, 三雲是宏: 橋梁の損傷要因診断事例からのラフ集合を用いたルール型知識の獲得方法, 構造工学論文集 Vol.44A, 1998.3
- 2)Pieter Adriaans, Dolf Zantinge : データマイニング, 共立出版, 1998.6