

I-18 平均圧縮情報量による鋼橋の補修方法選定のための知識抽出

Knowledge Discovery using Average Compressed Entropy for selecting retrofitting Method of Steel Bridges Damaged by Fatigue

皆川 勝

上谷 丈和

Masaru MINAGAWA

Takekazu KAMITANI

【抄録】疲労による損傷を受けた鋼橋の補修方法選定には熟練技術者の経験的な知識を必要とすることから、エキスパートシステムなどの利用が図られてきた。一方、エキスパートシステム構築の際にもっとも困難な課題は知識の専門家からの収集である。本研究では、知識獲得の初期におけるルールベース構築支援を目的として、データベースが持つ情報量の変化が知識の体系化によって生ずるという考えにもとづく平均圧縮情報量を基準として、データベースとしての事例群からルールを抽出し、知識ベースを同定できることを示す。

【ABSTRACT】 For the purpose of knowledge discovery, we evaluated average compressed entropies for a case-base virtually constructed through some inferences with the inference system that we proposed for selecting the retrofitting method. It is found from the analyses that the average compressed entropy is an effective measure for the discovery of knowledge that is implicitly buried into databases or case-bases.

【キーワード】 知識発見, データマイニング, エントロピー, 事例ベース, 補修, 鋼橋

【Key Words】 *knowledge discovery, data mining, entropy, case-bases, retrofitting, steel bridges*

1. はじめに

エキスパートシステムを構築する上で、最も困難なプロセスは知識の獲得作業である。システム構築の初期においては、専門家へのアンケート等から、主な特徴点を抽出してルール化する。また、知識の追加・修正においては、いかにルールベース内の整合性を保ち、信頼性の高い評価をするかが問題となる¹⁾²⁾。そこで、土木分野においても知識獲得問題に関する研究がなされている。

宮本らは、ファジィ集合論を用いた知識の更新手法³⁾や、経験的知識に内在する曖昧さをファジィエントロピーの概念を適用して定量的に評価する方法⁴⁾を提案している。三上らは、推論エンジンにニューラルネットワークの学習機能を持たせ⁵⁾、また、事例ベース推論の適用事例も報告している⁶⁾⁷⁾⁸⁾。また、皆川らは仮説をノードとし、仮説間の関係を表す結合で構成される相互結合型のネットワークでルールベースを構成し、知識洗練機能

付き推論システムを構築して、その有効性を示した⁹⁾¹⁰⁾。

一方、データベースからの知識発見 (*KDD: Knowledge Discovery in Databases*)、あるいはデータ発掘 (*Data Mining*) については、機械学習に関する Michalski¹¹⁾の研究や、Piatetsky-Shapiro¹²⁾によるデータベースからの知識発見手法の提案に端を発して、1995年に『知識発見とデータ発掘』会議が初めて開催されるなど、近年注目を集めている。人工知能の分野では上記のように知識獲得の困難を克服するための方法論としてこのデータマイニング手法の研究がなされている。特に相関ルールの抽出については数万件あるいはそれ以上の膨大なデータからその相関性を抽出することが主要な目的であることから、演算速度が速くノイズに対して頑健なアルゴリズムの開発が主たる研究目標である¹³⁾¹⁴⁾。しかし、実問題に対する適用に関しては、流通業¹⁵⁾、医療¹⁶⁾、製造業¹⁷⁾などで研究されているが、いまだ実用化には到っていない。また、土木工学の分野の応用はみられない。一方、

連絡先：東京都世田谷区玉堤 1-28-1 武蔵工業大学工学部土木工学科

Tel : 03-3707-3111 (内線 3252), Fax: 03-5707-2226 , E-mail: mminagaw@eng.musashi-tech.ac.jp

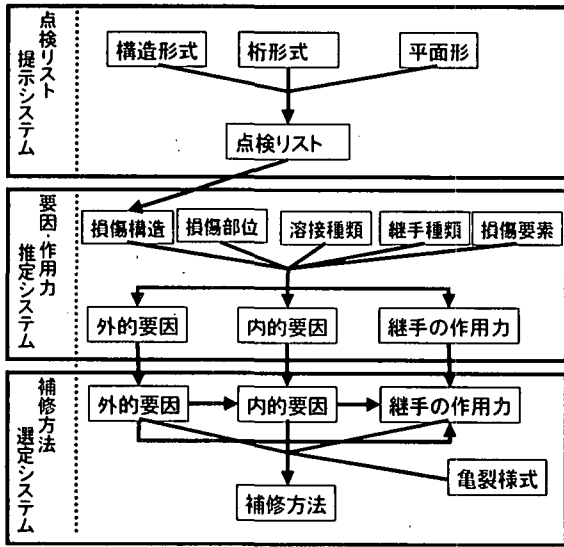


図-1 補修方法選定システムの概要²⁰⁾

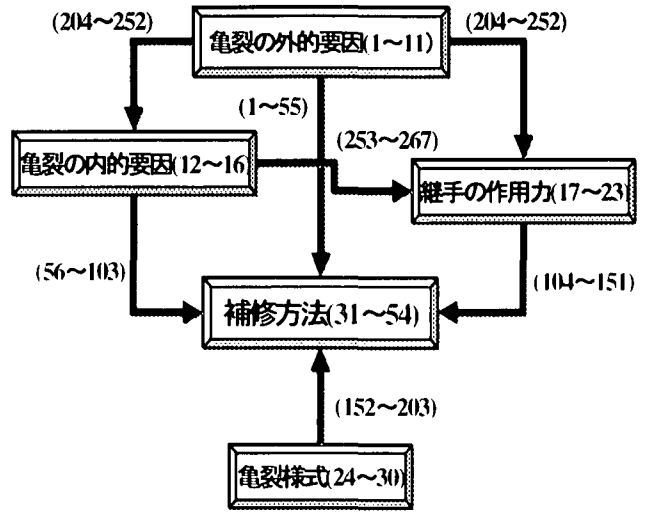


図-2 ネットワークの構成²⁰⁾

現在100種以上のKDDツールが開発されているが、1980年代のAIブームのときに発表されたAIツールに比較して、これらKDDツールを利用しても成熟した統計解析ツールほどの成果は得られないという指摘もある¹⁸⁾。したがって、安易にこの種のツールを利用することは避け、当面独自にソフトウェア開発をすることが望ましいと著者は考えている。

そこで、本研究では、知識獲得の初期におけるルールベース構築支援を目的として、データベースからの知識発見で用いられる平均圧縮情報量¹⁹⁾の概念を用いて、データベースからルールを自動生成するアルゴリズムをシステム化し、鋼橋の疲労損傷の補修方法選定問題に適用して、その有効性を検証する。

2. 鋼橋疲労損傷の補修方法選定のためのルールベース

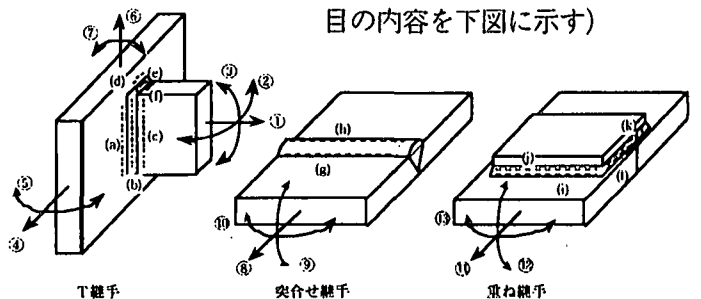
本研究では、田中らにより開発された鋼道路橋に発生した疲労損傷の補修・補強方法選定を対象問題としたエキスパートシステム²⁰⁾で用いられたルールベースをもとに、仮想事例を作成し、そこからの知識抽出を行った。

田中らによるシステムでは、図-1に示すように、点検リスト提示システムで、構造形式、桁形式、平面形から損傷事例に見られる損傷構造を探索し、点検すべき構造部分を点検リストとして提示し、このリストを基に点検、調査を行い、損傷が発見されれば、損傷構造、損傷部位、溶接種類、継手種類、損傷要素から要因・作用力推定システムにより、外的要因、内的要因、継手の作用力の推定を行っている。そして、補修の必要性があると判断された場合に、補修方法選定システムにより、要因・作用力推定システムで推定した外的要

表-1 補修方法選定システムにおける入力項目補修方法²⁰⁾

外的要因	補修方法項目
1 風による振動	31 ストップホール
2 活荷重の作用	32 ガウジング
3 地震による振動	33 グラインダ
4 低温	34 ビーニング
5 横分配作用	35 腹板ギャップの増大
6 ディテールの不適性	36 腹板厚の増大
7 2次変形	37 切り抜き
8 材質不良	38 再溶接
9 溶接不良	39 フランジと補剛材の溶接
10 製作誤差	40 再溶融
11 輸送架設荷重	41 補剛材付き添接板
	42 高力ボルト
	43 添接板
	44 挿入板
12 応力集中	45 ケーブルによる吊材相互の緊結
13 2次応力の応力集中	46 横桁と主桁のフランジ連結
14 2次応力	47 対傾橋と主桁のフランジ連結
15 応力集中による座屈2次応力	48 ダイヤフラムと主桁のフランジ連結
16 残留応力	49 横桁とアーチリブのフランジ連結
継手の作用力	50 縦桁支承の交換
17 T継手の作用力②	51 主桁の交換
18 T継手の作用力③	52 連結板の交換
19 T継手の作用力⑥	53 新補剛材の設置
20 T継手の作用力⑦	54 制振装置の設置
21 突合せ継手の作用力⑨	
22 突合せ継手の作用力⑩	
23 重ね継手の作用力⑫	
継手の亀裂様式	
24 T継手の亀裂様式(a)	
25 T継手の亀裂様式(b)	
26 T継手の亀裂様式(c)	
27 T継手の亀裂様式(d)	
28 T継手の亀裂様式(f)	
29 重ね継手の亀裂様式(i)	
30 重ね継手の亀裂様式(j)	

(表中の17から30の各項目の内容を下图に示す)



因, 内的要因, 継手の作用力の情報に加え観測された事実である亀裂様式の情報を用いて, 補修方法の選定を行っている。

このうち, 点検リスト提示システム及び, 要因・作用力推定システムでは, フレームを用いた推論を行っており, 推論過程が明らかにされていない。そこで, 亀裂の外的要因, 亀裂の内的要因, 継手の作用力を亀裂様式と同様, 観測された事実である既知情報とみなし, これから補修方法が選定される部分の知識ベースシステムを本研究の対象とした。図-2に仮説間の関係を示す。具体的な項目は表-1に示す。

3. 平均圧縮情報量によるデータベースからの知識発見

本節では, 情報理論の分野で用いられている平均圧縮情報量の知識発見への適用方法を述べる。

(1) データベースとしての事例群とルール

データベース中には, 有用性の有無に関わらず, 極めて多くの知識が存在する。この中から有用である知識を見分けるには, 対象領域の背景知識や, 対象領域固有の規準に基づいた重要度の評価を行うのが通常である。しかし, そのような専門家の固有知識に依存せず, 自動的にデータベースからルールを抽出することができれば, 知識獲得の困難が解消される。一方, 抽出した知識, あるいはルールの妥当性は, 客観的かつ定量的に評価されることが望ましい。

通常 $y \rightarrow x$ で表される相関ルールをデータベースから抽出する場合, 支持度及び確信度が用いられる²⁰⁾。支持度は事例全体に対して要素 x および要素 y を共に含む事例の割合であり, $support(y \rightarrow x)$ と表される。また, 確信度は要素 y を含む事例のうちの要素 x および要素 y を共に含む事例の割合であり, $confidence(y \rightarrow x)$ と表される。相関ルールを抽出することは, これらの両値が設定したしきい値を越えるルールを発見することを意味する。

今, データベース中に存在する各事例を e_i とし, 事例の総数を N とする。つまり, データベース中には, 以下のように N 個の事例が存在する。

$$\{ N: e_1, e_2, \dots, e_i, \dots, e_N \}$$

また, 事例 e_i は, m 個の要素 a_{im} によって構成されている。よって事例 e_i は, 次のように表すことができる。

$$\{ e_i: a_{i1}, a_{i2}, \dots, a_{im} \}$$

データベース内から, 相関ルールを生成する際に, 条

件部になり得る要素 a_{ij} を y , 結論部となり得る要素 a_{im} を x とし, 条件部と結論部になる要素を区別して表すことにする。なお, 各要素は, 条件部の要素にも結論部の要素にもなり得る。

次に, データベース内の各事例 e_i において, 条件部の要素 y を持つ事例の数を n_y とし, 結論部の要素 x を持つ事例の数を n_x とする。また, 条件部の要素 y と結論部の要素 x を共に持つ, すなわち, ルール $y \rightarrow x$ を抽出することのできる事例の数を $n_{y,x}$ とする。これは前述の支持度 $support(y \rightarrow x)$ に相当する。

(2) 平均圧縮情報量による相関ルールの抽出

はじめに, 結論部の要素 x の生起確率 $p(x)$ を式(1)より求め, 結論部の要素が \bar{x} でない場合, \bar{x} の生起確率を式(2)より求める。

$$p(x) = \frac{n_x}{N} \quad (1)$$

$$p(\bar{x}) = 1 - \frac{n_x}{N} \quad (2)$$

次に, 結論部の要素 x と条件部の要素 y の結合確率 $p(x, y)$ を式(3)より求め, また, 結論部の要素が \bar{x} で, 条件部の要素が y の結合確率 $p(\bar{x}, y)$ を式(4)より求める。

$$p(x, y) = \frac{n_{y,x}}{N} \quad (3)$$

$$p(\bar{x}, y) = 1 - \frac{n_{y,x}}{N} \quad (4)$$

また, 条件部の要素 y が起こった時に結論部の要素 x が起こる, 結論部の要素 x についての条件付確率を式(5)より求め, 条件部の要素 y が起こった時に, 結論部が \bar{x} である条件付確率を式(6)より求める。

$$p(x|y) = \frac{n_{y,x}}{n_y} \quad (5)$$

$$p(\bar{x}|y) = 1 - \frac{n_{y,x}}{n_y} \quad (6)$$

ここで, $p(x|y)$ は前述の $confidence(y \rightarrow x)$ であり, $p(\bar{x}|y)$ は $confidence(y \rightarrow \bar{x})$ である。

通常, エントロピーはある確率分布 $p(x)$ が与えられたとき, その不確かさの程度を表し, $-\sum p(x) \log(p(x))$ で与えられる²⁰⁾。ここで, 上式中の $\log(p(x))$ の底は 2 とする。不確かであるということは多くの情報量を含んでいるともいえることから, エントロピーはまた情報量であると

もいえる。 $-\log\{p(x)\}$ は $p(x)$ が $1/2, 1/4, 1/8, \dots$ となるにつれて、 $1, 2, 3, \dots$ となつて必要なビット数となることから、記述長と呼ばれる。

そこで、結論部の要素 x がデータベース内で持つ記述長を $-\log\{p(x)\}$ で示す。また、この時、結論部の要素 x を事例内の要素 $x = a_{im}$ として持つ事例が、同時に条件部の要素 y を $y = a_{ij}$ として持つ場合があった時、 $y \rightarrow x$ というルールが事例中に内在していることになる。この要素 y と要素 x を同時に持つ時の結論部の要素 x がデータベース内で持つ記述長は、 $-\log\{p(x|y)\}$ で示すことができる。この2つの記述長の差は、事例中に内在する $y \rightarrow x$ というルールによつてもたらされるものであり、条件部の要素 y が起こる前の記述長 $-\log\{p(x)\}$ と、条件部の要素 y が起こった後の記述長 $-\log\{p(x|y)\}$ との差によって表すことができる。

また、式(3)に示したように、ルール $y \rightarrow x$ を生成することが可能な事例は、データベース内に $N \cdot p(x, y)$ 個存在していることから、データベース内において結論部の要素 x が、条件部の要素 y によって圧縮される情報量は、 $N \cdot p(x, y)$ と上記の記述長の差との積により求めることができる。以上に示したことは、結論部の要素が \bar{x} の場合についても同様にいえる。

この $N \cdot p(x, y)$ あるいは $N \cdot p(\bar{x}, y)$ と、記述長の差の積は圧縮情報量 (Compressed Entropies: CE) と呼ばれ、式(7)により $CE(x, y)$ で表わされる。

$$CE(x, y) = [-Np(x, y) \log\{p(x)\} - Np(\bar{x}, y) \log\{p(\bar{x})\} - [Np(x, y) \log\{p(x|y)\} - Np(\bar{x}, y) \log\{p(\bar{x}|y)\}]] \quad (7)$$

この圧縮された情報量を評価値とするため、データベース内の事例の総数 N により平均値をとつて、評価値となる平均圧縮情報量 $ACE(x, y)$ は式(8)より求められる¹⁹⁾。

$$ACE(x, y) = CE(x, y) / N \\ = p(x, y) \log \frac{p(x|y)}{p(x)} + p(\bar{x}, y) \log \frac{p(\bar{x}|y)}{p(\bar{x})} \quad (8)$$

前述のように $confidence(y \rightarrow x) = p(x|y)$ であり、 $support(y \rightarrow x) = Np(x, y)$ であつて、平均圧縮情報量は支持度及び確信度を含んだ評価基準となっている。また、鈴木ら¹⁹⁾が指摘しているように、Blachman²³⁾およびSmyth and Goodman²⁴⁾もこれがルールを発見する際の望ましい評価基準であるとしている。

4. 適用結果及び考察

前述の平均圧縮情報量に基づく知識発見の適用例として、鋼橋疲労損傷の補修方法選定問題を選び、既存の推論システムで得られる事例群を生のデータベースと見な

し、ここからルールを抽出した後、それをルールベースとする推論を実施して、ルール抽出の可能性を検証する。

(1) データベースの作成

ここでは、平均圧縮情報量によつて、データベースから有効な知識を抽出することが可能であるかを検討するため、田中らによる鋼橋疲労損傷の補修方法選定システムのルールベースを基に、総数 2695 の仮想事例を作成し、これをデータベースとした。

亀裂の内的要因、外的要因、継手の作用力、亀裂様式の30個の入力情報項目について *true* または *false* を設定して、起こり得る可能性のある全組み合わせ、2695通りの仮想事実を発生させた。この仮想事実に対して、皆川らのルールベース洗練機能付推論システム^{9), 10)}より各補修方法の可能性を評価し、可能性のもっとも高い補修方法を唯一つ選定した。この補修方法と対応する仮想事実を示す入力情報項目とを組み合わせて仮想事例とした。なお、本来、データマイニング技術は既存知識の得られていないもの、容易には得られないものに対して適用するものであるが、本研究はその可能性を検証するために、仮想事例を用いたものである。生のデータベースからのルール発見が本来の目標であり、本研究の成果はその方向にゆく可能性を示すことにある。したがって、ここでの目標は仮想事例を作成するもとなつたルールベースを隠蔽して仮想事例に対してデータマイニングを実施して、隠蔽したルールベースを得られるか否かを示すことにある。

なお、データマイニングでは通常、ノイズを含んだ生のデータに対してクリーニングを施す^{18), 21)}が、ここで用いるデータは既存のルールベースから得られる因果関係を仮想的な事例データとしており、また、マイニングの目標はそのルールベース自体であることから、ノイズや人為的ミスを含まないデータを扱っていることになるた

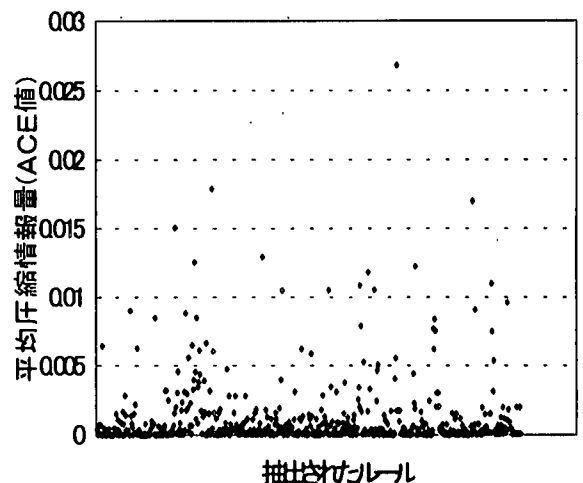


図-3 抽出されたルールの ACE 値

表-2 データマイニングにより抽出されたルール (b の場合)

	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	
1	0.4	0.2			0.6			0.2	0.6			0.2			0.8										
2		0.2	0.6					0.4				0.4	0.4												
3	0.2	0.4										0.8		0.6			0.4				0.6	0.8			
4	0.2	0.2	0.6	0.4				0.2	0.2			0.6									0.8	0.6			
5		0.8	0.6	0.4	0.4			0.4	0.4								0.8								
6	0.6	0.6	0.6	0.2	0.8			0.2	0.8			0.2	0.2				0.8								
7	0.4	0.2	0.4																			0.2			
8		0.6						0.4				0.2	0.6								0.6	0.6			
9	0.6	0.6		0.8								0.6													
10			0.6	0.4				0.6				0.4	0.6										0.8	0.8	
11		0.2	0.8	0.2	0.4			0.2	0.4			0.4													
12	0.4							0.6				0.8	0.8												
13	0.8	0.4	0.8					0.8				0.2	0.4												
14	0.6	0.2	0.8					0.8				0.4	0.2												
15	0.8	0.4	0.8	0.8	0.8			0.8	0.8			0.4	0.6				0.8								
16	0.2	0.6	0.8	0.8	0.8			0.8	0.8			0.6	0.8	0.6	0.8		0.8				0.8	0.8	0.8	0.8	
17	0.2	0.8	0.8	0.8	0.8				0.8			0.2					0.4								
18	0.8	0.8	0.2	0.4	0.8			0.6	0.8			0.8					0.2								
19	0.6	0.2			0.8			0.4	0.8			0.2	0.2		0.4						0.6	0.4	0.2	0.2	
20	0.6	0.6	0.8	0.4				0.2	0.8			0.6													
21	0.4	0.6	0.4	0.4	0.8			0.4	0.8			0.4	0.6		0.4		0.2				0.6	0.4	0.2	0.2	
22	0.4	0.6	0.4	0.2	0.6			0.2	0.8			0.2	0.6	0.4		0.2					0.4	0.4	0.2	0.2	
23	0.2	0.6	0.2	0.4	0.8			0.2	0.8			0.2			0.4		0.6				0.6	0.2	0.2	0.2	
24	0.8	0.2	0.4	0.4	0.2	0.6		0.8	0.6			0.6	0.4				0.6								
25			0.4		0.4			0.6	0.4			0.2					0.2								
26	0.6		0.8	0.6	0.2			0.6	0.4			0.4	0.8				0.6								
27	0.8	0.6	0.8	0.2				0.2	0.6			0.6	0.4												
28			0.2	0.6	0.6			0.6	0.6			0.6	0.2				0.2								
29	0.6		0.8	0.2				0.4	0.2			0.4	0.6	0.8		0.6					0.8		0.2	0.8	
30	0.4	0.2	0.4	0.2	0.6				0.4			0.6	0.6		0.8		0.6					0.8	0.8	0.2	

表-3 既存システムで用いたルールベース¹⁰⁾

	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	
1								0.4							0.6										
2	0.8		0.8	0.4				0.8		0.4		0.8	0.8								0.4				
3																									
4	0.4						0.4							0.4											
5	0.8	0.8			0.4			0.8				0.8				0.2	0.8					0.4			
6	0.8	0.8			0.8	0.4		0.8	0.8			0.8	0.8			0.4	0.8	0.4		0.4					
7	0.6											0.4													
8		0.2						0.2			0.4	0.4													
9	0.8	0.4	0.8	0.8			0.6	0.8		0.4		0.8	0.8	0.4							0.4				
10																							0.4	0.4	
11	0.8		0.8									0.8													
12		0.8	0.8	0.8			0.6	0.8		0.4		0.8	0.8	0.4	0.6					0.4	0.4	0.4	0.4	0.4	
13	0.8	0.8	0.8		0.8			0.8	0.8			0.8				0.4	0.8	0.4	0.2						
14	0.8		0.8					0.8				0.8	0.8												
15	0.8	0.4			0.4	0.4		0.8			0.4	0.8	0.8			0.2	0.2					0.4			
16	0.2					0.2					0.2	0.2	0.2												
17	0.8	0.8	0.8	0.8				0.8	0.8			0.8	0.8		0.6	0.4	0.6			0.4					
18	0.8	0.8	0.8		0.8	0.4	0.6	0.8	0.8		0.4	0.8	0.8			0.4	0.4	0.4	0.2						
19																									
20	0.8		0.8	0.4				0.8				0.8	0.6										0.4	0.4	
21	0.4	0.2						0.2				0.4	0.4												
22							0.4							0.4											
23				0.4				0.2		0.4		0.4									0.4	0.4			
24	0.8	0.6						0.6																	
25	0.8		0.6					0.6	0.4			0.8	0.4												
26	0.8	0.6	0.4		0.4	0.4	0.4	0.8	0.4		0.4	0.8	0.8		0.4	0.2	0.6						0.4	0.4	
27	0.8	0.6	0.8	0.4	0.6			0.8	0.4			0.8	0.4			0.4		0.4	0.2	0.4					
28	0.8		0.6	0.4				0.6	0.4			0.4	0.6		0.4										
29				0.4				0.2		0.4		0.4									0.4				
30																						0.4			

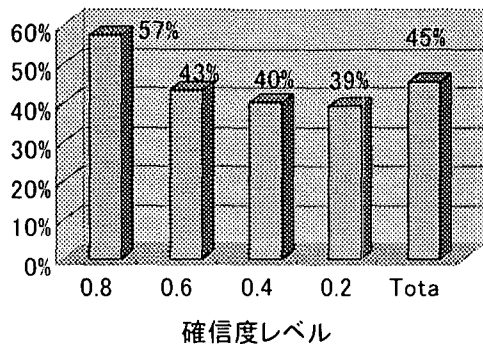


図4 データマイニングされたルールのうち、既存ルールベースに存在したルールの比率

め、クリーニングなどの前処理は行っていない。

(2) データベースからのルール抽出

平均圧縮情報量である式(8)を用いて、仮想事例より作成したデータベースからのルール抽出を試みた。その結果、620のルールがデータベースから抽出された。図-3には、横軸に抽出された620のルールをとり、縦軸に評価値である平均圧縮情報量(ACE値)をとって示す。

この抽出されたルールを基に新たにルールベースを構築し、鋼橋疲労損傷の補修方法の選定を行って、抽出されたルールの有用性を検討する。

抽出されたルールを、ACE値の高いルールから順に、次に示す3通りの規準で階層化して、ルールの確信度とし、それによって新たなルールベースを構築した。

- a) 全ルールを3分割し、上位40%を0.7、次の40%を0.4とした場合
- b) 全ルールを5分割し、上位から20%毎に0.8、0.6、0.4、0.2とした場合
- c) 全ルールを10分割し、10%毎に、0.8から0.1までとした場合

なお、3通りすべてにおいて、下位20%のルールについてはACE値が極端に小さく、切り捨てることとした。

表-2には、b)の場合のルールベースを示す。表中の番号1~30は、表-1の入力情報項目の番号に対応しており、また、31~54は補修方法項目の番号に対応している。比較のために仮想事例の作成に用いたルールベースを表-3に示す。

はじめに、データマイニングで得られた各確信度レベルのルールについて、仮想事例の作成に用いたルールベース中に存在していたルールの割合を図-4に示す。同図より、確信度レベルが高いほど一致する割合が高く、ルール抽出がより適切になされていることが分かる。逆に、仮想事例の作成に用いたルールベースで関連ルールが設定されているルールに対する、データマイニングで得られたルールの割合を図-5に示す。全体では57%であり、

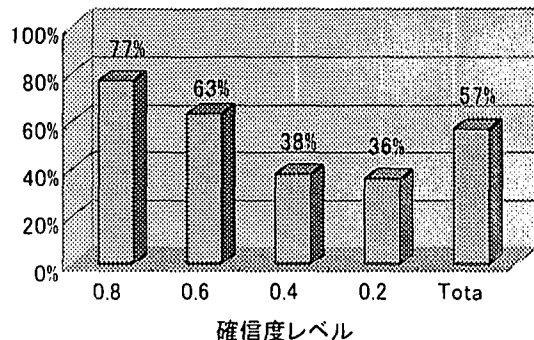


図5 既存ルールベースに存在したルールのうち、データマイニングされたルールの確信度レベル毎の比率

表-4 補修方法を推定する事例²⁾

橋梁名	事例1	事例2
国名	米国	米国
架設年度	1958年	1973年
損傷発見年度	12年	0年
構造形式	単純橋	連続橋
桁形式	格子合成プレートガーター	合成プレートガーター
平面形	直橋	直橋
損傷構造	主桁のカバープレートと取り付け部	主桁の垂直補剛材取り付け部
損傷部分	主桁下フランジ カバープレート	主桁腹板 垂直補剛材
溶接種類	隅肉溶接	隅肉溶接
継手種類	重ね継手	T継手
損傷要素	隅肉溶接	主桁腹板
外的要因	活荷重の作用	輸送架設荷重
内的要因	応力集中	2次応力
継手の作用力	重ね継手の作用力12	T継手の作用力3
亀裂様式	重ね継手の亀裂様式(i)	T継手の亀裂様式(d)
亀裂方向	溶接に平行な亀裂	溶接に平行な亀裂
補修方法	ピーニング 再溶融 高カポルト 添接板	ストップホール グラインダ

やはり確信度レベルが高いほど発見されたルールの比率は高い。仮想事例の作成に用いたルールのうち、確信度の低いものは推論結果に対する影響度が低いことから、仮想事例に内包された知識にその影響が生じにくいのは当然である。一方、確信度レベルが0.8の場合には80%近い一致度を示しており、ルール抽出が適切に行われていることを示している。

なお、個別の補修方法で見ると、補修番号36、37、40、41、46、48、49、50に関するルールがまったく抽出されていないことが表-2と表-3の比較からわかる。これは仮想事例作成の段階で各事例に対する補修方法をひとつに限定したために、情報がそこで失われたことによるものである。ひとつの方法としては、trueまたはfalseの2者択一形式で仮想事例を構築しているが、これをファジネスを考慮できるように修正することが考えられるが、これについては今後の課題とする。

(3) 補修方法選定結果

次に、新たに構築したルールベースを適用して、補修方法の選定を行った。抽出されたルールを用いて構築したルールベースによる推論結果(以下、推論結果DMとい

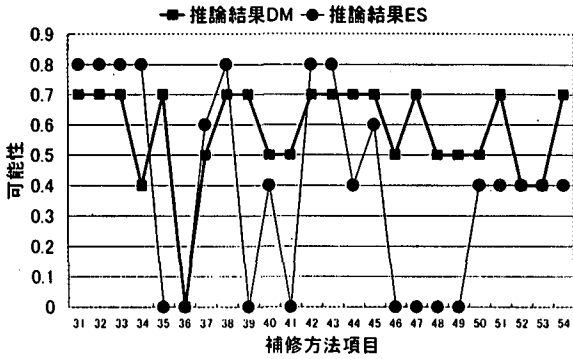


図-6 a)の場合の補修方法選定結果(事例1)

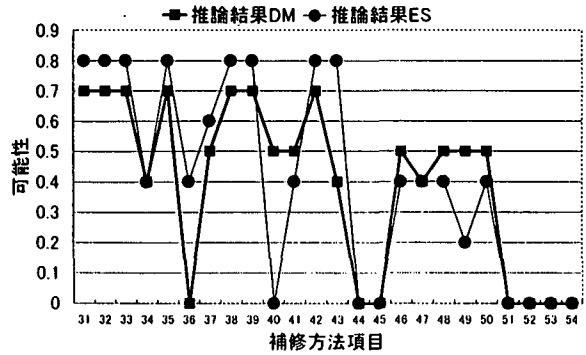


図-7 a)の場合の補修方法選定結果(事例2)

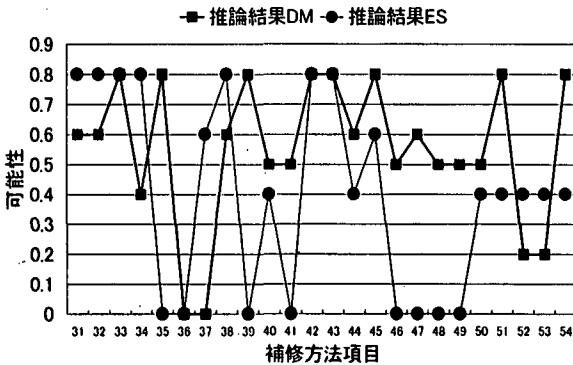


図-8 b)の場合の補修方法選定結果(事例1)

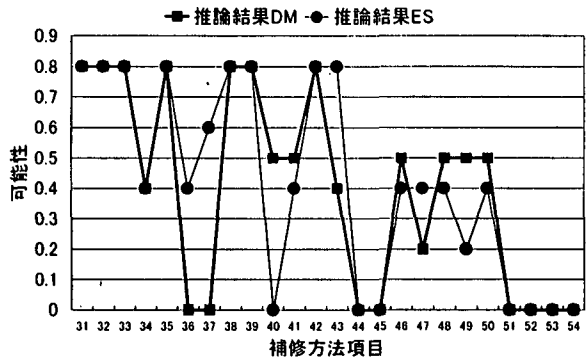


図-9 b)の場合の補修方法選定結果(事例2)

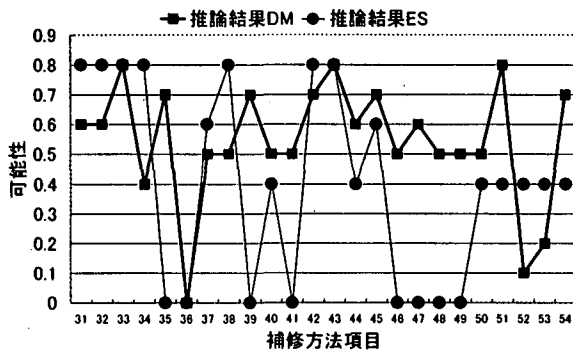


図-10 c)の場合の補修方法選定結果(事例1)

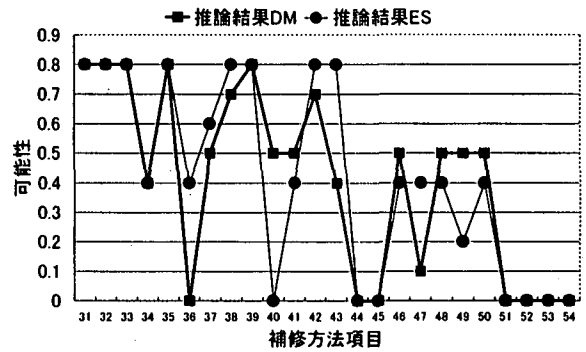


図-11 c)の場合の補修方法選定結果(事例2)

う)と、皆川らによる既存システムによる推論結果(以下、単に推論結果ESという)を比較することで、抽出されたルールの妥当性を間接的に検討することができる。推論に適用した事例は、表-4に提示した事例1,2である。

図-6,図-7にa)の場合における推論結果DMをそれぞれ事例1,事例2について、横軸に補修方法項目の番号をとり、縦軸に各項目の可能性であるノード値をとって示す。比較のため、推論結果ESを併せて示す。以下、図-8,図-9にはb)の場合の結果を示し、図-10,図-11にはc)の場合の結果を示す。

この結果、a)の場合、図-3のように、ACE値による評価によって幅広く分布するルールを0.7と0.4の2段階で分類したため、ノード値による可能性の分類が十分になされていないことがわかる。これに対し、b),c)の場合、それぞれ5段階、10段階に分類したことから、十分な可能性の分類を行って推論結果を提示している。図-8,図-10の事例1の結果では、推論結果DMとの間に

大きな違いが見られる。特に、補修方法項目の番号51,54で、推論結果ESは0.4,0.2と低い可能性を示しているのに対し、0.8と高い可能性を示しているなど、ノード値が0.8を示す項目に関しての違いが目立つ。しかし、このことは、推論結果ESについても同様である。また、図-9,図-11の事例2の結果では、ノード値0.8を示す補修方法項目は、推論結果ESとほぼ一致している。また、推論結果DM全体の傾向としても、可能性の値に違いは見られるが、推論結果ESの傾向と同様な結果を示している。

しかし、c)の場合、確信度を細分化したことにより、ノード値0.8を示す可能性の高い項目の提示が、特に事例1については3項目と少数であることから、本来システムに要求されている技術者支援システムとしての可能性の分類の観点からすると、利用者が提示された解を参照する以前に、過度に可能性の分類がなされる恐れがある。

5. おわりに

本論では、知識獲得の初期における知識獲得問題解決のため、データベースからの知識発見 (KDD) の手法に平均圧縮情報量 (ACE 値) を用いてルール抽出を試みた。また、抽出したルールの有用性を検討するため、抽出したルールを用いて鋼橋疲労損傷の補修方法選定のためのルールベースを構築し、実際の事例を適用して補修方法選定を試みた。この結果、抽出対象としたルール群の50%から60%程度のルールを抽出することができた。さらに、確信度の高いルールについては80%近い一致度でルールを抽出することができた。また、抽出されたルールを用いた推論結果は、既存システムによる推論結果と同様の傾向を示しており、知識発見におけるルール抽出に ACE 値を用いることができることを示すことができたと考えられる。現状では、confidence および support を算出する際に、事例に各要素が含まれているかいないかというクリスプな評価をしているが、それに対してファジイな評価を導入することを考えている。

謝辞：本研究を行うにあたり、関西大学の田中成典助教授に有益なご助言を頂きました。また、研究当時卒業研究生であった稲蔵聡氏（現在(株)横浜コンサルティングセンター勤務）ならびに関田竜典氏（現在(株)フジタ勤務）に協力いただいたことを記して謝意を表します。

参考文献

- 1) 本位田 真一, 市川 照久: エキスパートシステム基礎技術, オーム社, 1989. 12.
- 2) S.M. ワイス, C.A. クリコフスキ(森 健一訳): エキスパートシステムの設計, 近代科学社, 1987. 7.
- 3) 宮本 文穂, 森川 英典, 益成 一郎, 古川 正典: 橋梁診断エキスパートシステムにおける知識の更新手法, 構造工学論文集, Vol. 37A, pp. 643-654, 1991. 3.
- 4) 串田 守可, 徳山 貴信, 宮本 文穂: 橋梁診断における経験的知識に内在するあいまいさの定量化に対するファジイエントロピーの適用, 構造工学論文集, Vol. 38A, pp. 571-584, 1991. 3.
- 5) 三上 市蔵, 田中 成典, 倉地 晶: 鋼橋疲労損傷の補修方法選定のための学習機能のあるニューラルネットワークシステム, 構造工学論文集, Vol. 37A, pp. 655-668, 1991. 3.
- 6) 三上 市蔵, 田中 成典, 倉地 晶, 米田 慎二: 鋼橋疲労損傷の補修方法選定システムにおける類推推論と負の学習の実現, 構造工学論文集, Vol. 38A, pp. 557-569, 1992. 3.
- 7) 田中 成典, 三上 市蔵, 前田 秀典: 事例ベース推論を用いた鋼道路橋疲労損傷の補修方法の選定システム, 第2回ファジイ土木応用シンポジウム講演論文集, pp. 37-42, 1994. 12.
- 8) 田中 成典, 三上 市蔵, 前田 秀典: 動的矛盾知識の検出・解消処理を施した鋼橋疲労損傷補修方法選定システム, 構造工学論文集, Vol. 41A, pp. 587-596, 1995. 3.
- 9) 皆川 勝, 佐藤 茂, 上谷 丈和: 事例ベース推論を採用した知識洗練機能付診断エキスパートシステムの開発, 土木学会論文集, No. 595/VI-39, pp. 67-76, 1998. 6.
- 10) 皆川 勝, 上谷 丈和: 鋼橋疲労損傷の補修・補強方法選定に対するルール洗練機能付き推論システムの適用性, 構造工学論文集, Vol. 45A, pp. 523-531, 1999. 3.
- 11) Michalski, R.: Theory and Methodology of Inductive Learning, Machine Learning: An Artificial Intelligence Approach, Vol. 1, Morgan Kaufmann 1983.
- 12) Piatetsky-Shapiro, G.: Discovery and analysis of strong rules in databases, Advanced Database System Symposium, pp. 135-142, 1989. 12.
- 13) 喜連川 優: データマイニングにおける相関ルール抽出技法, 人工知能学会誌, Vol. 12, No. 4, pp. 19-26, 1997. 7.
- 14) 河野 浩之, 西尾 章治郎, Jiawei Han: データベースからの知識獲得技術, 人工知能学会誌, Vol. 10, No. 1, pp. 38-44, 1997. 7.
- 15) 沼尾 雅之, 清水 周一: 流通業におけるデータマイニング, 人工知能学会誌, Vol. 12, No. 4, pp. 34-41, 1997. 7.
- 16) 津本周作, 田中 博: データマイニングの医療応用, 人工知能学会誌, Vol. 12, No. 4, pp. 42-50, 1997. 7.
- 17) 落田 美紀, 行松 徹, 堀 聡, 滝 寛和: 製造業におけるデータマイニングの応用と課題, 人工知能学会誌, Vol. 12, No. 4, pp. 50-55, 1997. 7.
- 18) 寺野 隆雄: KDD ツールの動向と課題, 人工知能学会誌, Vol. 12, No. 4, pp. 27-33, 1997. 7.
- 19) 鈴木 英之進, 志村 正道: 情報理論的手法を用いたデータベースからの例外的知識の発見, 人工知能学会誌, Vol. 12, No. 2, pp. 129-136, 1997. 3.
- 20) 田中 成典: 橋梁工学への知識情報処理技術の応用に関する研究, 関西大学学位論文, pp. 25-248, 1996. 9.
- 21) 河野 浩之: データベースからの知識発見の現状と動向, 人工知能学会誌, Vol. 12, No. 4, pp. 3-10, 1997. 7.
- 22) 例えば, 堀部 安一: 情報エントロピー論, 森北出版, 1989
- 23) Blachman, N. M.: The amount of information that y gives about x, IEEE Trans. On information Theory, Vol. IT-14, No. 1, pp. 27-31, 1968.
- 24) Smyth, P. and Goodman, R. M.: Rule induction using information theory, Knowledge Discovery in Databases, pp. 159-176, AAAI Press/The MIT Press, 1991.