

データマイニングにおけるルール抽出の考察と トンネル覆工の微細ひび割れ発生要因への適用

A Basic Consideration on Association Rules with Data Mining Algorithms
and Reasoning cause of Tunnel Lining Crack.

須藤 敦史*・吉尾 薫光**・三上 隆***

Atsushi SUTOH, Tadamitsu YOSHIO and Takashi MIKAMI

*正会員 博士(工学) 株式会社 札幌工業 土木技術部 主席研究員 (〒105-8488 東京都港区西新橋2-23-1)

**正会員 修士(工学) 札幌市土木部 (〒060-8611 札幌市中央区北1条西2丁目)

***フェロー会員 工博 北海道大学大学院教授工学研究科社会基盤専攻(〒060-8628 札幌市北区北13条西8丁目)

In this study consists of the following two topics, one is a basic consideration on data mining algorithm which is a powerful data-processing procedure, and the other is introduce and/or discuss to a data mining procedure using neural network. The notion of data mining is classified and interpreted to a non-structural inverse problem. Data mining means to discover objectively knowledge hidden in vast amount of data. It is found that the usefulness of this data mining procedure using neural network, for finding the cause of tunnel lining.

Key Words: data mining, neural network, non-structural system, identification, tunnel lining crack

1. はじめに

寒冷地に施工されたNATMトンネルにおいて二次覆工の施工直後もしくは数年以内に微細なひび割れが確認された事例が報告されている¹⁾。

このような二次覆工の微細ひび割れは施工状況・観測記録などから剥離・崩壊に至るような有害なひび割れに進展するとは考えられないが、寒冷地トンネルの厳しい冬期の環境²⁾を考慮するとできる限り抑制することが望ましい。

しかし、このような微細ひび割れは地山の緩みなどの外力に伴うものとは異なり、自然環境(風や気温)、材料(乾燥³⁾・熱応力⁴⁾)特性や施工状況など、種々の要因の複合作用により発生してくるものと考えられるため、発生要因・誘引やそのメカニズムの特定は非常に難しいのが現状である。

一方、コンピュータ技術の発達により大量のデータ処理が可能となっているが、データ様式・形態や相互関係が複雑であるため、『知識やルールもしくは現象や事象間の因果・相関関係』を客観的に探索する手法の開発が望まれている。

このような状況下、知識発見(KDD: Knowledge Discovery in Databases)あるいはデータマイニング(DM: Data Mining⁵⁾、⁶⁾以下DMと記述)が最近注目されてき

ており、流通・マーケティングや医療などで数多く応用⁷⁾などされ、またルール探索には集合理論・決定木(Decision Tree)・Genetic Algorithm: GA・Neural Network: NN・人工生命(Artificial Life)技術など様々な解析ツールが用いられている^{8),9),10),11)}。

しかし、DMは現実問題として a)現象・事象の生データを対象とし、b)現象・事象の関係式や支配方程式がどのように記述されるか、また c)それに対する境界条件や各種パラメータの数や存在の有無までも明確ではなく、加えて d)入出力値などが特定できないデータ分析・解析であるため、データベースからの新たな知識やルールの発見・獲得は容易ではなく、図-1 に示すように従来個別に用い

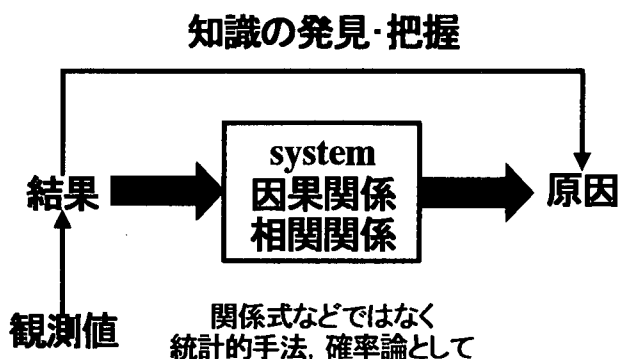


図-1 データマイニング

られている解析技術や手順・手法などを融合した総合的データプロセッシング技術(非常に特殊な逆問題)^{12),13)}とも言える。

そこで,本研究では寒冷地トンネルにおける二次覆工の微細なひび割れを対象として,その気象環境や施工状況などの種々の要因とひび割れ発生との因果(相関)関係を,施工状況や観測データを入力としたニューラルネットワーク(Neural Network: NN¹⁴⁾)を用いたDMにより推定・把握することを目標としている。

ここで,NNは適用例の多い階層型ネットワーク¹⁵⁾を採用して,ノード重みの大きさからデータ間の相関・関連性の強さを算出している。さらにノード・中間層の数と学習誤差の関係および抽出ルールの単純さに対する基礎検討も行っている。

2. データマイニング

データからの有用な情報抽出は確率・統計・機械学習など以前から行われているが,データマイニングはこれらの技術を融合してシステム化された新たなデータ解析の枠組み,あるいはデータ処理・解析における新しい考え方として注目されており¹³⁾,背景には以下の要求が影響している。

(1) 膨大なデータの有効活用

コンピュータ技術の発展により,多種多様のデータウェアハウスが蓄積されていく反面,データを有効に活用することが求められている。

(2) データ理論・技術の統合

確率統計,機械学習,データ解析理論・技術はそれぞれ別途にあるいは重複して研究がなされているため,これらを融合して,すべてを見通せるような新しいデータ解析手法が求められている。

(3) 技術のソフトウェア化

人工知能やデータベース技術をデータマイニングという考え方により統合し,ユーザーフレンドリーでかつ実際のデータ処理に適用する汎用化されたソフトウェアが求められている。

したがって,データマイニングは様々な解析手法を用いることから特別な知識を必要とせず手軽に素早く,なおかつ簡便にデータ解析(相関関係や因果関係を発見)を行うことを目的としている。

3. ニューラルネットワーク

(1) 情報伝達

脳における神経細胞の結合を工学的にモデル化したものであり,神経細胞をシナプスがつなぎ,情報の伝達を行っている。ここで興奮状態により発生するのがインパルスであり,この大きさの度合いで次の細胞への情報伝達の決定がなされる。

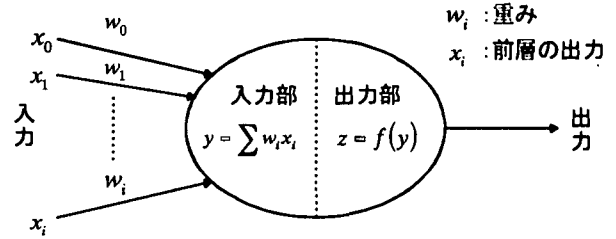


図-2 ニューロンのモデル化

それを工学的にモデル化してニューロン(図-2)としたのがニューラルネットワークであり,ニューロンは入力信号の積和を計算する入力関数部と入力に従って答えを決定する出力関数(しきい値関数)部とがある。

いま1つのニューロンが \$n\$ 個から入力を受け,その出力 \$x_i\$ としてシナプス結合の強さを \$w_i\$ とする。このとき \$i\$ 個の入力を単純に加算したものは式(1)となる。

$$y = \sum w_i x_i \quad (1)$$

ここで入力の総和 \$y\$ がしきい値 \$\theta\$ を超えたとき式(2)インパルス \$u\$ が発生して,出力関数 \$z\$ は式(3)となる。

$$u = \sum_{i=1}^n w_i x_i - \theta \quad (2), \quad z = f(u) \quad (3)$$

また \$\theta\$ の代わりに -1 を与えると抑制入力する式(4)となる。

$$u = \sum_{i=1}^n w_i x_i + w_o(-1) = \sum_{i=1}^n w_i x_i - w_o \quad (4)$$

(2) バックプロパゲーション法(BP法)

階層型ネットワーク(図-3)は入力された信号が次々に伝播しながら一方向で出力層に到達するモデルであり,学習方法は出力値と教師信号との誤差を最小にするようにノード間の結合重みを調節するものである。

各ノードは式(5)で示す出力関数(シグモイド関数)を有しており,出力値を一般化すると式(6)となる。

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (5)$$

$$O_i^k = f\left(\sum_j W_{ij}^k O_j^{k-1} - \theta_i^k\right) \quad (6)$$

\$O_j^{k-1}\$: \$k-1\$ 層の出力, \$W_{ij}^k\$: 重み, \$\theta_i^k\$: しきい値

時刻 \$t\$ における入力信号 \$p\$ は絶対誤差を式(7),またネットワークの平均誤差は式(8)となる。

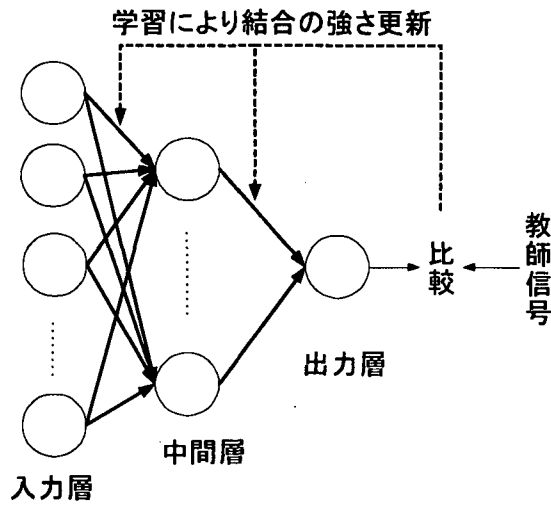


図-3 階層型ネットワーク

$$E_p(t) = \frac{1}{2} \sum_i (T_{pi} - O_i^n)^2 \quad (7)$$

T_{pi} : 入力信号 p に対する出力 i の教師信号

$$E(T) = \frac{1}{P} \sum_p E_p(t) \quad (8)$$

ここで誤差が小さくなるように式(9)により重み W を調整する。

$$\Delta W_{ij}^k(t) = -\varepsilon \frac{\partial E(t)}{\partial W_{ij}^k(t)} + \alpha \Delta W_{ij}^k(t-1) \quad (9)$$

ここで α は誤差の振動を減衰させる効果がある反面、学習速度の低下や局所解滞留の恐れが生じ、一般的に α は $0 < \alpha < 1$ の範囲が望ましい。

また誤差が振動する時には ε を減少、収束が遅い時には ε を増加させればよい。

(3) 仮想インピーダンス法¹⁶⁾

仮想インピーダンス法は高速学習するように重みの変化率を2次振動系として扱っている。

ここで重みの変化率を式(10)のように振動系の速度と仮定して、離散近似すると式(11)となる。

$$M \dot{v}_{ij}^k(t) + D \dot{v}_{ij}^k(t) + K v_{ij}^k(t) = - \frac{\partial E(t)}{\partial W_{ij}^k(t)} \quad (10)$$

M : 仮想質量, D : 粘性係数, K : スティフネス

$$v_{ij}^k(t) = - \frac{1}{M + D + K} \frac{\partial E(t)}{\partial W_{ij}^k(t)} + \frac{2M + D}{M + D + K} v_{ij}^k(t-1) - \frac{M}{M + D + K} v_{ij}^k(t-2) \quad (11)$$

式(11)に従って重みの変化量を求め、時間刻みを1とす

ると式(12),(13)となる。

$$\Delta W_{ij}^k = v_{ij}^k(t) \Delta t = v_{ij}^k(t) \quad (12)$$

$$\Delta W_{ij}^k(t) = - \frac{1}{M + D + K} \frac{\partial E(t)}{\partial W_{ij}^k(t)} + \frac{2M + D}{M + D + K} v_{ij}^k(t-1) - \frac{M}{M + D + K} v_{ij}^k(t-2) \quad (13)$$

$$\frac{1}{M + D + K} = \varepsilon, \frac{2M + D}{M + D + K} = \alpha, \frac{-M}{M + D + K} = \beta \quad (14)$$

$$\Delta W_{ij}^k(t) = -\varepsilon \frac{\partial E(t)}{\partial W_{ij}^k(t)} + \alpha \Delta W_{ij}^k(t-1) + \beta \Delta W_{ij}^k(t-2) \quad (14)$$

ここで学習法は式(14)となり、2次の微分項に対応するパラメータ β が追加され、 β が強制振動することで局所解からの脱出を期待する。

(4) 成長側抑制学習

成長側抑制学習はノードの結合を際立たせるために用いる方法で、重みの成長に抑制をかけて重要な結合だけ残るようにするため、相関関係やルールが読み取りやすくなる。ここで重み更新に成長側抑制の項 S (式(16))を追加した式(15)で示される。

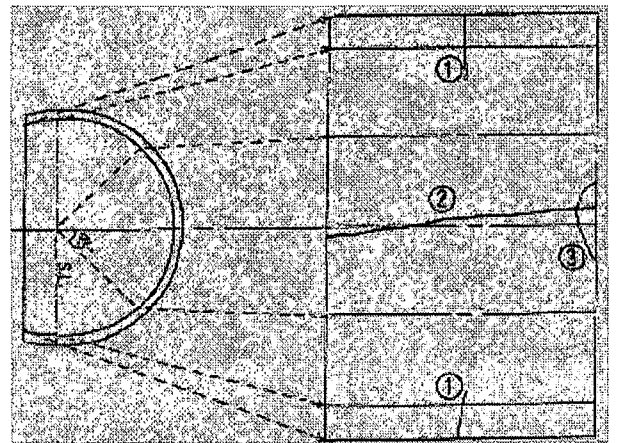


図-4(a) ひび割れタイプ

表-1 ひび割れパターン

パターン	向き	長さ	幅	パターン	向き	長さ	幅
A・②	延長	長	広	E・①	横断	長	狭
B・②	延長	長	狭	F・②	横断	短	広
C・②	延長	短	広	G・①	横断	短	狭
D・②	延長	短	狭	H・③	他	短	狭

長さは3~4m、幅は最大値0.4mmが基準

$$\Delta W_{ij}^k(t) = -\varepsilon \frac{\partial E(t)}{\partial W_{ij}^k(t)} + \alpha \Delta W_{ij}^k(t-1) + \beta \Delta W_{ij}^k(t-2) + S \quad (15)$$

$$S = -s \frac{1}{m-1+1} \operatorname{sgn}(W_{ij}^k(t)) \left\{ \sum_{i=1, \neq j}^m |W_{ij}^k(t)| + |\theta_i^k| \right\} \quad (16)$$

S :成長側抑制項, s :成長側抑制係数,
 m : $k-1$ 層ノード数, W :重み, θ :しきい値
 $\operatorname{Sgn}(x)$: $x < 0 \rightarrow -1, x = 0 \rightarrow 0, x > 0 \rightarrow +1$

(5)ネットワークの再構築

さらに各ノードの結合を明確にするために小さい重みの結合を削除してネットワークの再構築を行う.このことにより簡素化され規則やルールを容易に読取ることができる.

4. 二次覆工のひび割れ

(1)ひび割れパターン

二次覆工に発生するひび割れを要因・発生箇所別(図4(a))に整理すると以下となる.

- ① 脚部付近から周方向:コンクリート内部が低温・乾燥もしくは地山側が高温・高湿である時に拘束部(インバートや地山)の内部側に発生
- ② 天端部付近の延長方向:坑内の温度変化,コンクリートの温度降下,乾燥収縮加えて断熱材を施工している箇所では地山の拘束力や外方向反力の不足(自重に対する)
- ③ 打継目付近に発生:型枠設置時に既設コンクリートに生じた弱面が温度差あるいは乾燥よりひび割れに進展

(2)収縮ひずみの発生する要因

コンクリートにひび割れが発生する要因は,地山からの外力・施工環境・コンクリート性状などが挙げられるが収縮ひずみの発生要因としては以下に示す事項が考えられる.

- ① コンクリートにおける発熱・硬化温度の降下による収縮
- ② トンネル坑内温度および湿度の降下・低下による温度収縮

よって,湧水区間や外気の影響を受けやすい坑口付近や延長の短いトンネルではひび割れが発生しやすく,加えて冬期気温の低い寒冷地トンネルの影響は顕著になると考えられるが,これら要因と実際のひび割れとの相関関係は明確になっていないのが現状である.

ここで,ひび割れには図4(a)に示すタイプに加えてひ

び割れの有無・方向(縦・横・他3種類),長さ(長・短2種類),幅(広・狭2種類)のパターンを設定した.

(3)調査項目

調査項目とは表-2に示すように1998年5月から観測を継続している9項目(春・夏)の調査データを使用している.

表-2 観測項目

			最大値, 最小値, 基準値
A	季節	夏	6月, 7月, 8月, 9月
		冬	12月, 1月, 2月, 3月
B	スパン長	5.5m~10.5m	8m
C	コンクリート温度	12℃~29℃	平均値を基準: 17.5℃
D	スランプ	8.5cm~14cm	平均値を基準: 12.0cm
E	空気量	3.6%~5.5%	平均値を基準: 4.4%
F	7日	11.5~20.3kgf/cm ²	平均値を基準: 15.4kgf/cm ²
G	28日	24.2~30.3kgf/cm ²	平均値を基準: 27.3kgf/cm ²
H	インバートの有無	0 or 1 データ	
I	断熱材の有無	0 or 1 データ	
J	クラックの有無	0 or 1 データ	

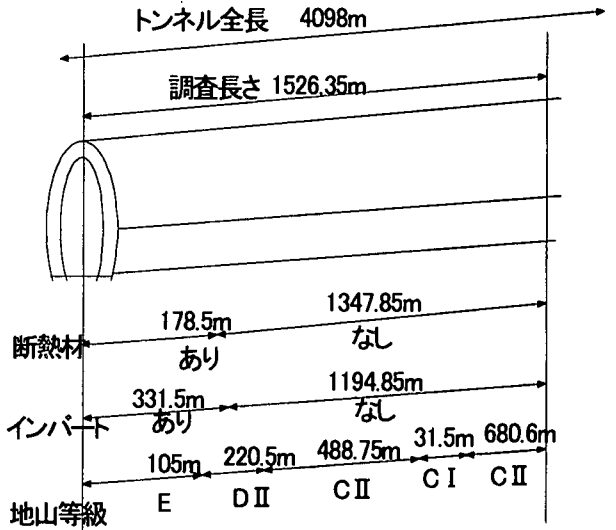


図-4(b) 観測値の位置図

ここでデータ解析を簡略化するために数値属性データ B~G は平均以下 0, 以上は 1 としてプール化を施し, 各観測項目とひび割れ発生との関連性を最も単純な「If ~ then ...」ルールとして抽出する.

また使用したデータ総数は 184 個であり, その各項目の概略位置を図4(b)に示す.

(4)中間層の考察

ニューラルネットワークではデータの性質や量などによって目的に即した中間層のユニット数およびノード数を決めることが膨大なデータ解析を行う上で有利であり, 効果的・合理的であるため, 数値的検討により中間層のユニット数を決定した.

ここで中間層のユニット数を変えたときの収束回数と

学習誤差の変化を表4に示す。

表より、中間層のユニット数を増加させるとネットワークの学習の平均誤差は減少するが、逆にネットワークが誤差を小さくすることを優先して（自由度が大きくなって観測誤差に過剰に適応）しまい規則やルールが判別できない状況に至った。

そこで、本研究ではネットワーク結合の明確化と誤差が比較的小さくなる（学習効果が現れる）中間層のユニット数となる2個を採用している。

表-3 入・出力値

入力項目	出力項目
地山等級	クラックの有無
打設季節	
インバート有無	
断熱材の有無	
コンクリート打設温度	
スランプ	
空気量	
7日強度	
28日強度	

表-4 収束・誤差値

中間層 ユニット数	入力 項目数	結合状態がわかる 入力項目の個数	収束数	平均誤差
1	9	9	0	0.21048
2	9	8	48	0.15239
3	9	3	76	0.14807
4	9	3	120	0.09196
5	9	1	130	0.08099
6	9	1	150	0.07653
7	9	0	157	0.03438

(3) 学習定数の決定

ニューラルネットワークでは小さい学習誤差（学習効果）を得るために、適切な中間層のユニット数や ε, α を適切に設定する必要がある。

そこで今回は解析ですべて誤差収束が確認されたパラメータの組み合わせ（学習回数＝100000回、 $\varepsilon = 5.0, \alpha = 0.4$ ）を用いて行っている。

5. データマイニングによる相関解析

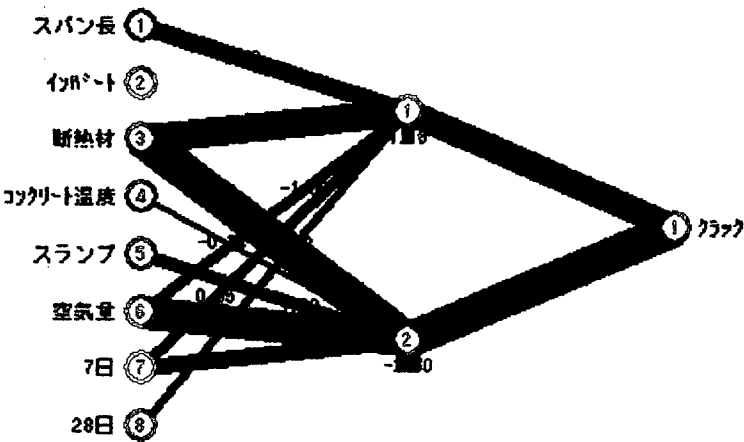


図-5 1次相関(成長側抑制学習)

(1) 一次相関解析

成長抑制学習を用いた相関解析では図-5 に示すように明確な結合が得られ、表-5 に示す「スパン長・スランプ」に正の相関、「空気量・28 日強度」に負の相関が現れている。また「断熱材・7 日強度」は正・負の同等な相関関係が得られ、この観測データでは解析は難しいと考えられる。

ここで相関関係の強さはネットワーク結合・数字の大きさで、正の相関・負の相関はその符号で判断される。

表-5 1次相関

出力	正の相関	負の相関
クラックの有無	スランプ	空気量
	スパン長	28 日強度

□ は相関の強いもの

(2) 季節別の相関解析

打設季節別の相関解析を行った結果では図-6,7 に示すネットワーク結合が得られた。

ここで「スパン長・スランプ」（正の相関）、「空気量・断熱材の有無」（負の相関）は夏・冬のいずれも大きな相関を示しており、求められた相関ルールを表-6 に示す。

表-6 季節別の相関

正の相関		負の相関	
夏	冬	夏	冬
スパン長	スパン長	空気量	空気量
スランプ	スランプ	断熱材	断熱材
インバート		コンクリート温度	28 日強度

□ 両方から抽出されたルール
□ 片方だけから抽出されたルール

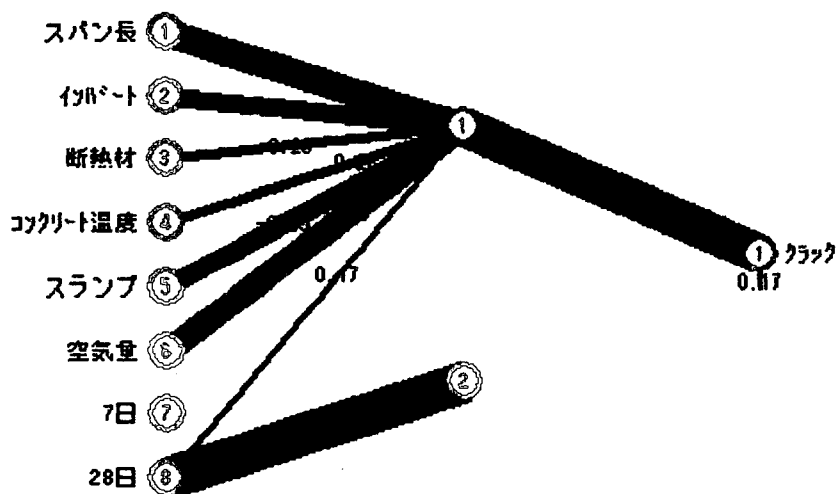


図-6 季節別の相関 (夏)

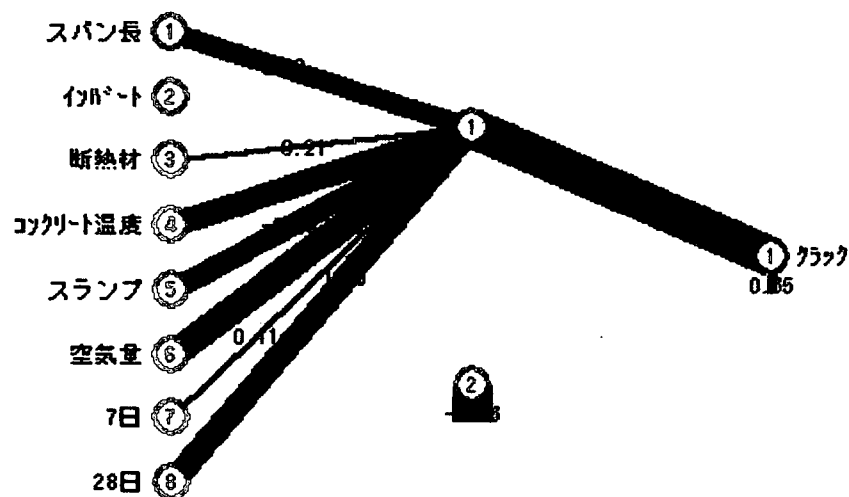


図-7 季節別の相関 (冬)

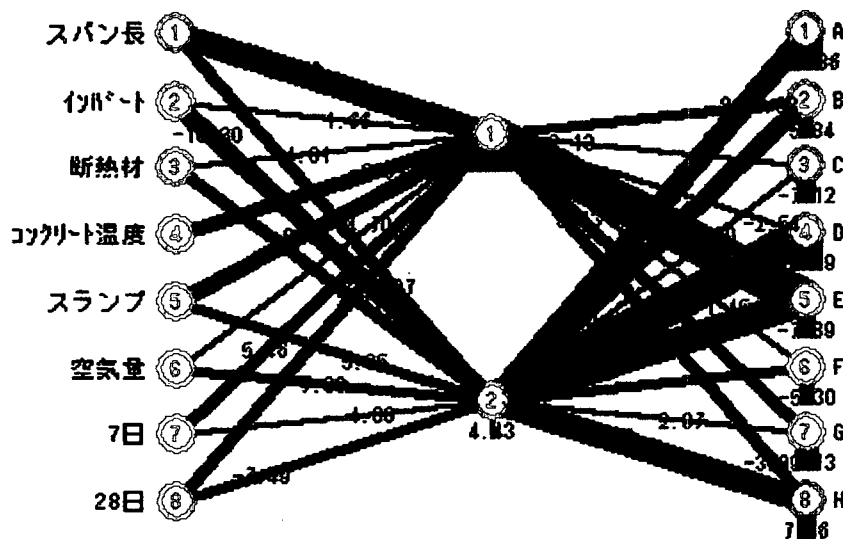


図-8 ひび割れ解析 (1)

(3) ひび割れパターンの相関解析

ひび割れ方向・長さ・幅などひび割れパターン別の解析結果は図-8～10に示すように3種類の解析を行った。

解析(1)では、全観測項目を入・出力項目に設定して解析を行ったが、図-8に示すように結合が複雑になったため相関関係の判別はかなり難しい結果が得られている。

そこで解析(2)では図-9に示すように、各ひび割れパターン別に1つずつ出力項目に設定して解析を行った。

図-9から得られた相関ルールを表-7に示す。この解析においてもほとんどのパターンで一次相関解析とほぼ同じ結論が得られたが、簡便化・観測誤差・定性記述などの影響により矛盾した結果も得られている。

一方、パターン解析(3)では、図-10に示す例のように観測項目を1つずつ入力としてそれぞれのひび割れパターンを出力項目に設定して相関解析を行った。

ここで図-10から抽出された相関ルールを表-8に示す。

パターン解析(3)においても結果は解析(2)と同様にほとんどのパターンで相関解析(1)と同じ結論が得られているが、簡便化・観測誤差・定性記述などの影響により一部矛盾した結論もまた得られている。

このひび割れパターン解析(2)・(3)の結果は同じになると考えられるが、実際には前述した理由などにより矛盾した結果も得られる。

しかし、実際には相関ルールをより多く抽出することができたパターン解析(2)の方

がこのような条件のついた相関解析には有効であると思われる。

表-7 ひび割れパターン解析 (2)

出力	正の相関	負の相関
パターン A	7日強度 スランプ 断熱材の有無	コンクリート温度 空気量 28日強度
パターン B	コンクリート温度 スパン長 スランプ	インパートの有無
パターン C	断熱材の有無	空気量 スランプ 28日強度
パターン D	インパートの有無 28日強度	断熱材の有無 7日強度 空気量 スランプ
パターン E	インパートの有無 28日強度	断熱材の有無 7日強度
パターン F	スパン長 断熱材の有無	7日強度 インパートの有無 空気量 コンクリート温度
パターン G	断熱材の有無 7日強度 スパン長	28日強度 空気量 コンクリート温度
パターン H	コンクリート温度 スランプ	インパートの有無 7日強度 28日強度 空気量

□ は相関の強いもの

表-8 ひび割れパターン解析 (3)

出力	正の相関	負の相関
パターン A	7日強度 スランプ	コンクリート温度
パターン B	スパン長 コンクリート温度	インパート スランプ 7日 28日
パターン D	28日強度	断熱材 空気量 7日強度
パターン E	インパート	7日強度
パターン F	断熱材 スパン長 スランプ	7日強度

パターン C, G, Hはルールの抽出なし

6. 結果の検討

寒冷地トンネルの二次覆工の微細なひび割れの解析からより複雑な相関(因果)関係の発見の可能性が示唆されたことより、ニューラルネットワーク等の人工生命技術

を用いたデータマイニングを用いた、この種の問題(定性・定量データが混在するデータベースにおける相関(因果)解析)に対する適用性の一端を確認したと言える。

また、このような二次覆工における状況・環境、施工性等とそのひび割れとのデータマイニングによる相関関係に関する研究は、初めて間もなく、かつまだ試行錯誤の段階であるため、ひび割れ発生との調査項目との明確な相関関係の抽出は今後の課題と言える。

参考 文献

1) 杉浦高広, 中井健司, 佐藤京, 西村克弘, 三上隆: 二次覆工のひび割れ発生に対する施工状態・環境の影響について, トンネル工学研究論文・報告集第9巻, pp.177~182, 1999.

2) 坂本稔, 川北稔, 五十嵐敏彦: 道路トンネルの変状実態-北海道の場合-, トンネルと地下, 第20巻5号, pp.31~35, 1989.

3) 中井健司, 小林将, 湯浅浩介, 三上 隆: 北海道におけるトンネルライニングの温度依存性と施工に関する提案断熱材設計のための実用的な温度算定式の提案, 土木学会北海道支部論文報告集, 第54号(A) I-72, pp.344~347, 1998.

4) 三上隆, 小池明夫, 吉村俊彦, 須藤敦史, 岡田正之: 水和熱に起因する寒冷地コンクリートの温度応力解析, 第16回寒地技術論文・報告集, Vol.16, CTC2000- I-044, pp.242~246, 2000.

5) Pieter Adrians, Dolf Zantinge 著(山本英子, 梅村恭司訳): データマイニング, 共立出版, 1998.

6) 中林三平: データマイニング(価値ある情報を掘り当てる), NIKKEI COMPUTER, pp.142~147, 1996.

7) 大規模データベースからの知識獲得, 人工知能学会誌, Vol.12, No.4, pp.496-549, 1997.7

8) 徳山豪: データマイニングに使われる最適化の数理, 応用数理, VOL.6, NO.4, pp.303-313, 1996.12

9) Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R.: Advances in knowledge Discovery and Data Mining, MIT Press, 1996.

10) 須藤敦史, 高須光朗, 星谷勝: ニューラルネットワークを用いたデータマイニングによる非構造システムの同定, 応用力学論文集, Vol.2, pp.83-90, 1999.

11) 須藤敦史, 星谷勝: 決定木・GAを用いたデータマイニングによる赤潮発生要因の同定, 応用力学論文集, Vol.3, pp.99-106, 2000.

12) 須藤敦史, 佐藤大介, 星谷勝: 非構造システムの同定・逆問題におけるデータマイニングによる相関ルール抽出について, 応用力学論文集, Vol.3, pp.33-40, 2001.

13) 河野博之: データベースからの知識発見の現状と動

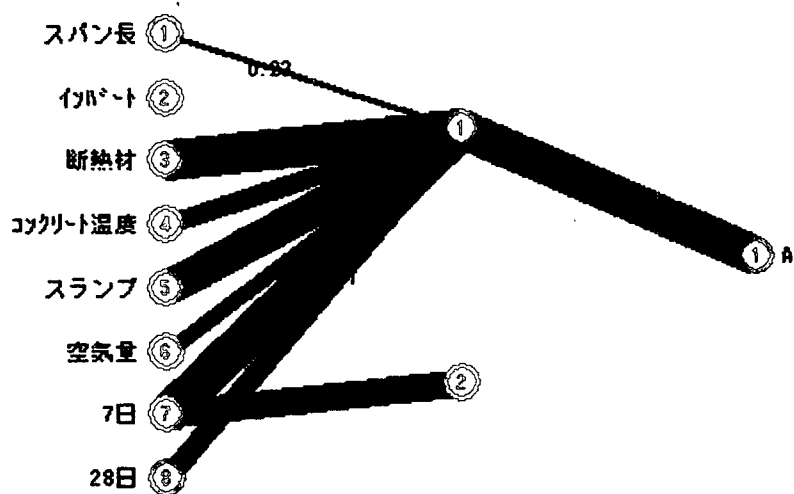


図-9 ひび割れパターン解析 (2)

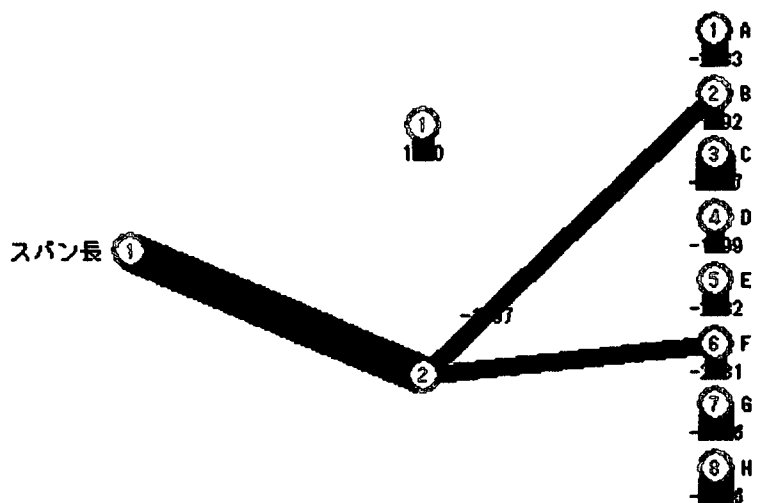


図-10 ひび割れパターン解析 (3)

向, 人工知能学会誌, Vol.12, No.4pp.496~504.1997.

- 14) (社) 日本工業技術振興協会(ニューロコンピューター研究部会):ニューロコンピューティングの基礎理論,海文堂出版,1990.
- 15) 萩原将文:ニューロ・ファジィ・遺伝的アルゴリズム,産業図書,1995.
- 16) NEUROSIM/L light,富士通,1996.

(2002年4月19日)