

## ルール群表現の発展による化学発ガン性知識の洗練

岡田孝

関西学院大学情報メディア教育センター  
okada@kwansei.ac.jp

化学構造と発ガン性の相関関係を明らかにすることは、非常に重要かつ困難な課題であり、データマイニング技術に大きな期待がかけられている。しかし、マイニングにより発ガン性を示唆するルールが提示された時、たとえ専門家であろうとその正否を即座に判断することはできない。ルールはさらに研究を進めるための材料に過ぎない。したがってマイニングを真に有効たらしめるには、利用者のアクティブな反応を誘発する、分かり易い表現を与える必要がある。本報告では、カスケードモデルによるルール発掘の枠組みに加えて、データスケープ概念にもとづくルールの提示法を採用することにより、有効な知識の発掘が効率的に行えることを示す。

### Polishing Knowledge on Chemical Carcinogenesis using a New Rule Expression

Takashi Okada

Kwansei Gakuin University, Center for Information & Media Studies

Data mining is expected to play a crucial role in the study of chemical carcinogenesis. As the problem is very complex, an expert cannot judge whether a pattern expressed as a rule is reasonable or not, immediately. A rule is just a material to formulate working hypotheses. The expression of rules must be easy to understand so that it invokes active user reactions. The cascade model was used to analyze the chemical carcinogenicity problem. In this report, we employ a new rule expression based on datascape concept to polish the knowledge on this problem. Inspection of the obtained rules has been an efficient work, leading to new valuable knowledge.

#### 1. はじめに

化学物質の構造とその物質が引き起こす生理活性間の相関を明らかにすることは非常に重要な課題であり、構造活性相関と呼ばれる研究分野がすでに確立されている。この分野では対象とする化合物構造を同一骨格を有するものに限定し、各化合物の特徴を各種の物理化学的およびグラフ理論的な特性量によって記述して、活性との相関を各種の多変量解析の方法により解析することが一般的であった。近年 high throughput screening 技術の革新により、データ量が急激に増加している。それに伴い、より多様な構造を持つ多数の化合物群のアッセイデータから、薬理活性や毒性の原因を研究者が理解できる形式で説明することが求められており、データマイニング技術への期待が高まっている [1]。

しかし、問題の複雑性とその結果が直接に人類の健康に影響を与えることから、マイニングによって得られた知識をそのまま利用できる可能性はない。ルールは、研究者のさらなる調査・実験を計画する際に、作業仮説を考え出すための材料である。まさに利用者のアクティブな反応が必要とされるマイニング課題であるといえる。反面、もしマイニング結果が組織化されないままに多数のルールとして提示されるならば、利用者はその中から有効な材料を探すために多大の労

力と注意を費やすことになり、アクティブな反応を期待することは難しい。

本報告では Predictive toxicology challenge 2000-2001 ワークショップにおいて用いられた化合物群をマイニングの対象とする [2]。筆者はこのワークショップに参加するにあたり、カスケードモデルを用いて出力されるルール群の解析を行った。しかし、百近いルールを精細に調査することは困難であり、高強度の 10 個以内のルールのみを考察の対象とせざるを得なかった [3]。

このような経験に基づき、筆者はデータスケープ概念とそれにもとづく新しいルール群の表現を提唱している [4, 5]。本報告では、この新しい表現を用いることにより、発ガン性と化学構造の間の相関関係探索において、どのような利点を実際に得られるか確認することを目的とする。次節では、カスケードモデルと新しいデータスケープ概念にもとづくルール表現を簡単に紹介する。第3節では、実際に解析した結果を提示するとともに、利用者によるアクティブな反応が可能となったことを示す。

## 2. カスケードモデルとルール群表現

### 2.1 カスケードモデル

このモデルは、相関ルールと同様に itemset のラティスを構築し、ラティス内で識別力の高いリンクをルールとして選択し提示する。その際、item としてはすべて [attribute: value] の形式を採用する。ルールの識別力は Gini によるカテゴリー変数に対する平方和から導かれる(1) 式の  $BSS$  (Between-groups sum of squares) で与えられる [6]。

$$BSS_i^g = \frac{n^L}{2} \sum_a (p_i^L(a) - p_i^U(a))^2 \quad (1)$$

ここで、 $p_i(a)$  は目的属性  $i$  の値が  $a$  である確率を示し、 $g$  がグループを、 $U$  はグループに分割前の節点、 $L$  は分割後の節点を表す。また、ルールは下記のように表現される。

```
IF main_condition ADDED ON preconditions
THEN class_variable: pre-distribution → post-distribution
THEN other_variable: pre-distribution → post-distribution
```

ここで、ルール左辺は、リンク上位節点のアイテム群からなる前提条件部と、リンクにおいて付加されるアイテムの主条件部に分けて表現される。他方、ルールの右辺は主条件適用前後の目的変数値の分布を表現する。なお、対象事例群において、主条件と相関の高い説明変数についても、その値の分布の変化を表示する。

### 2.2 データスケープ概念にもとづくルール群表現

データスケープとは読んで字のごとくデータの風景を意味する造語である。大きな山地全体を理解しようとするとき、尾根筋等の概念図による大まかな把握から、その一部の詳細表示に至るまで、レンズの倍率を変えるように特徴を提示できれば、山歩きをする人にとっては親切である。データを解析する場合も、ルール群を概念図と考え、比較的少数のルール群による全体像の提示から、その一部の詳細に至るまで、できる限り組織化して提示することにより、データ風景の全体像を提供しようとする試みである。具体的には上記の形式で得られたルールから出発して、以下のような操作により最終的なルール群を提示する。

- (1) ルールの主条件節および前提条件節の形式を [属性: 値] から [属性: 値域] に拡張し、 $BSS$  値を規範とした貪欲的探索によって、局所最適なルールを探索する。これにより、多くのルール候補が同一の最適ルールに収束し、ルール数の減少に効果がある [4]。
- (2) 一对のルール A, B 間で、主条件適用前後の支持事例集合の関連度を次式で定義し、その値が一定値 ( $min\text{-}rlv$ ) よりも大きい場合に、 $BSS$  値の大きい方を主ルール、他方を関連ルールとして組織化する [4]。

$$rlv^{UL}(A, B) = \max\left(\frac{cnt(A^{UL} \cap B^{UL})}{cnt(A^{UL})}, \frac{cnt(A^{UL} \cap B^{UL})}{cnt(B^{UL})}\right) \quad (2)$$

ここで、 $A, B$ は支持事例集合を表し、 $cnt$ は集合の要素数を返す関数である。また、 $UL$ は主条件適用前か後かを指示する。適用前後の双方ともに関連すると判断された場合を  $ULrelative$ 、適用後だけが関連する場合を  $Lrelative$  と呼ぶ。なお、適用前だけが関連する場合も  $Urelative$  と呼ぶが、この場合は主ルール：関連ルールへの組織化は行わない。

- (3) 目的変数の値の分布が、前提条件または主条件の特定の属性値で急激に変化することがある。このような値域を  $ridge$  と呼び、次式で定義する  $\Delta BSSrate$  が負で大きな値を取る。そこで、これが一定値 (default: -0.3) よりも小さい場合に、ルールの付帯情報として提示する [5]。

$$\Delta BSSrate = \frac{BSS(X + \Delta X) - (n(X + \Delta X)/n(X)) \cdot BSS(X)}{|n(\Delta X)|} \bigg/ \frac{BSS(X_0)}{n(X_0)} \quad (2)$$

ここでは、ルールの値域の変化： $X \rightarrow X + \Delta X$ に伴う  $BSS$  値の変化を、 $ridge$  領域  $\Delta X$  中の事例数  $n(\Delta X)$  と、元のルール値域  $X_0$  における  $BSS$  値で規格化している。 $ridge$  領域中の事例数が一定値以上 (default:  $\max(0.1 \cdot n(\Delta X), 5)$ ) の時  $ridge$  と認定し、領域中の事例群における主条件適用前後の目的属性値の分布を表示する。

### 3. 化学発ガン性への適用

昨夏に開催された Predictive Toxicology Challenge 2000-2001 のワークショップにおいて、ネズミに対する化学発ガン性の推定が共通課題として設定され、多数の参加者が各自の方法での解析結果をもちより、その予測精度を競った。筆者はこの challenge に参加し、幸いにして良い評価を得ることができた。本節では、この challenge 全体の概要を紹介した後、筆者が前節の方法論により得た結果を、これまでのカスケードモデルにより得られた結果と比較して紹介する。

#### 3.1 Predictive Toxicology Challenge 2000-2001 [2]

このワークショップでは、NTP による化合物 417 種に対する発ガン性データ (male, female の rat, mouse に対するもの) を主催者が公開し、これらに対する記述子を一般から募集した。その結果、7種類の記述子群 (総数は 7000 以上におよぶ) が利用可能となった。また、FDA による 185 化合物の発ガン性検査結果を別に主催者が準備しておき、その化学構造や記述子のみをあらかじめ公開した。チャレンジの参加者は、NTP による化合物群を学習データとしてモデルを作成し、FDA の化合物群に対してその発ガン性を予測した。ワークショップでは、主催者がそれらの予測を集計して ROC 解析により各モデルの評価を行った。また、毒性の専門家が各レポートを読み、その理解容易性や妥当性を評価した。これらすべてのデータや、レポートはワークショップのホームページに記載されており、またこの試みに対する *Bioinformatics Journal* の特集号が近く発刊される予定である。

#### 3.2 カスケードモデルによる解析

解析に当たっては、提供された記述子群の中から、Kramer による 63 種の線形フラグメント構造、および TReymers の記述子群内のよく知られた 9 種の物理化学的性質を利用している。また本稿では解析対象を male mouse データに限った。

カスケードモデルによるラティスの生成は、 $minsup = 0.01$ ,  $thres = 0.1$  のパラメータ値を指定して行った。Male mouse データの場合、4 アイテムまでのラティスにほぼ 4 万個の節点を作成された。これまでの解析では、このラティス中からその  $BSS$  値が 1.68 (事例数の 0.5%) よりも大きな 526 リンクをルール候補として選択する。その中で強いルールから順に事例のカバーリングを考慮して 3 種類のルールセットを選択し、利用者の解析に提供していた [7]。この例では、各ルール群に 35, 26, 30 種、計 91 種のルールが提示される。解析者の立場からは、これらのルール群すべてを詳細に点検することは難しい。そこで、これらのルール中で  $BSS$  値が 3.0 以上の値を示す、表 1 の 3 種のルールのみを取り出し、点検の対象としていた。この表の第 1 のルールには、前提条件の異なる 3 種の付随するルールが記載されているが、この組織化は解析者が手作業により行ったものであった。

今回の解析においても、同一のパラメータを用いて生成したラティスが出発点となる。その中で *BSS* 値が 1.68 (事例数の 0.5%) よりも大きな 526 リンクをもとに、ルール最適化を行ったところ、*BSS* 値が 3.36 (事例数の 1%) よりも大きな値を有するルール 53 種が得られた。(2)式の値が 0.5 以上の条件で、これらのルール群を組織化したところ、8 種の主ルールとそれらに付随する合計 36 種の *ULrelative*, および 6 種の *Lrelative* ルールで結果を表現することができた。また、それぞれのルールごとに、*ridge* 情報も計算した。

### 3.3 ルールの解釈

表 1 中の Rule1 に対して、主条件と相関の高い変数値の分布、各種分布の可視化、事例となる化学構造の視察により、筆者は先に「flexibility に富む構造を持つハロゲン化合物は、水素結合受容体でないならば発ガン性である。また不活性化合物には低い HOMO レベルを持つものが多い。」という化学的に意味のある仮説を導き出している。

今回の解析で得られた 8 種の主ルールを表 2 に示す。この中で表 1 の Rule1 に対応するルールは、表 2 中の Rule1 である。比較すると、最適化により *BSS* 値は増加しているが、前提条件が複雑になっている。このルールには多数の関連ルールが存在するが、その内の 5 種は HBA および FLEX の条件節を主ルールと共有しており、最適化時の収束位置の差によるものと考えられた。また、他の 2 種の関連ルールも、属性の意味を考慮すれば、主ルールと同一の内容を表すものと見なすことができた。

図 1 に、仮説生成に有効と考えられる 2 種の関連ルールを選択し、主ルールと共に示す。最初の主ルールから主条件[HBA=0]が属性値[HBD: 低], [HOMO: 低], [C-C: y], [C-C-Cl: y], [c:c: n]

Table 1. Strong rules (*BSS*>3.0) derived from male mouse data set by the conventional method.

No.	Main condition	Preconditions	Changes in distribution †	<i>BSS</i>
1	[HBA = 0]	[FLEX > 0.5]	(.43 .57) / 79 → (.95 .05) / 19	5.08
	If no precondition is applied, (.38 .62) / 336 → (.65 .35) / 55, <i>BSS</i> =4.03.			
	If [C:c:c:c: n] is the precondition, (.40 .60) / 228 → (.69 .31) / 49, <i>BSS</i> =4.26.			
	If [c:c-N: n] is the precondition, (.39 .61) / 227 → (.65 .35) / 55, <i>BSS</i> =3.79.			
2	[C-Cl: y]	[C-O: n]	(.40 .60) / 214 → (.68 .32) / 44	3.56
	The percentage of [HBD = 0] also changes 52%→98%, [c:c-N: n] changes 62%→98%.			
3	[Dipole > 4]	[Cl: y]	(.54 .46) / 94 → (.07 .93) / 14	3.11

† In (*pos neg*)/#*case*, *pos* and *neg* show the probabilities of positive and negative cases, respectively. #*Case* denotes the number of cases. Distributions before and after the application of the main condition are shown.

Table 2. Principal rules derived from male mouse data set at *min-rlv*=0.5.

No.	Main condition	Preconditions	Changes in distribution †	<i>BSS</i>	UL	L
1	[HBA = 0]	[FLEX>0.05] [HBD≤2] [C-C-C-C: n] [C-C-C-O: n] [Cl-c:c: n]	(.35 .65) / 125 → (.95 .05) / 19	6.73	21	6
2	[S: y]	[CLOGP>0] [HBD≤2] [C-C-c:c: n] [C-C=C: n]	(.42 .58) / 184 → (.3 .97) / 33	5.11	10	0
3	[HBD=1-2]	[CLOGP>0] [LUMO>-0.15] [C-C=O: n] [C-O: n] [Cl-c:c: n] [c:c:c-N: n]	(.52 .48) / 82 → (0.0 1.0) / 18	4.95	1	0
4	[N: y]	[HBD=0] [HBA≤2] [C-C-C-C: n] [C-C=C: n] [C-C=O: n] [Cl-c:c: n] [c:c-O: n]	(.62 .38) / 58 → (0.0 1.0) / 11	4.24	0	0
5	[Dipole>4]	[CLOGP≤4] [HBD≤2] [LUMO>-0.15] [HOMO≤-0.25] [C: y] [c:c-O: n]	(.45 .55) / 117 → (.04 .96) / 24	4.06	3	0
6	[C-Cl: y]	[C-C-C: y] [C-C-O-C: n]	(.32 .68) / 75 → (.75 .25) / 20	3.70	1	0
7	[C-N: y]	[HOMO≤-0.2] [C: y] [C-C=O: n] [N-O: n]	(.46 .54) / 127 → (0.0 1.0) / 17	3.67	0	0
8	[N: y]	[CLOGP>0] [C-C-C-C: n] [c:c-N: n] [c:c-O: n]	(.48 .52) / 111 → (.11 .89) / 27	3.62	0	0

† Notations are the same as those used in Table 1. UL and L denotes the number of *ULrelative* and *Lrelative* rules.

と高い正の相関を持つことが判る。また、ridge 情報から、前提条件中の[FLEX>0.05]が非常に強い制約でことが判る。実際、[FLEX<=0.05]の ridge 領域中にある 53 事例においては、主条件の適用が発ガン性の有無に殆ど影響を与えていない。

Rule1-UL7 で示される *ULrelative* ルールは、主条件適用前には 87 事例中の 51 事例を主ルールと共有し、適用後の 14 事例ではその内の 12 事例が主ルールと同じである。主ルールにおいては付加的な相関として表された HOMO と LUMO が、ここでは主条件、前提条件として現れている。Ridge 情報を見ると、低い LUMO 値では発ガン性の分布が逆転しており、これが重要な情報であることが判る。また、主条件適用後に[c:c: n]が 100%となっていることは、このルールが芳香族以外の分子に適用されるべきことを示している。最後の Rule1-UL9 も主条件適用後の 12 事例を共有しており、他の 2 ルールでは付加的な相関であった[C-C-Cl: y]が主条件となっている。さらに、[FLEX<=0.05]が 3 ルールを通じて ridge として現れており、重要な前提条件であることが理解できる。

これらの知見は、厳密に個々のルールごとの前提条件、主条件と解釈するよりも、不活性化化合物のセグメントを特徴づけるパターン群である、と解釈することが適当であろう。得られた知見の大部分はすでにこれまでの解析で知られていたものであるが、(1) [LUMO: 高], [HOMO: 低]で不活性化化合物を特徴づけられること；(2) [FLEX: 低]が重要な ridge であること；(3) 芳香族以外の分子を対象としていること；これら 3 点が新たに判明したと言える。

表 1 の Rule2 は、すでに表 1 Rule1 の反映であることが判っている。実際に今回の解析では、表 2 Rule1 の関連ルールとして、同じ内容が示されている。また、表 1 の Rule3 は、[C-Cl: y]

```

Rule 1:      Cases: 125 -> 19; BSS= 6.735
IF [HBA=0] added on [FLEX>0.05] [HBD<=2] [C-C-C-C: n] [C-C-C-O: n] [Cl-c:c: n]
THEN MM:      BSS:6.73      0.35 0.65 ==> 0.95 0.05
THEN FLEX:     BSS:3.19      0.00 0.46 0.15 0.39 ==> 0.00 0.11 0.05 0.84
THEN HBD:      BSS:3.68      0.56 0.44 0.00 ==> 1.00 0.00 0.00
THEN HBA:      BSS:10.3     0.15 0.36 0.49 ==> 1.00 0.00 0.00
THEN HOMO:     BSS:2.73     0.20 0.52 0.28 ==> 0.63 0.37 0.00
THEN C-C:      BSS:3.73     0.50 0.50 ==> 0.05 0.95
THEN C-C-Cl:   BSS:4.91     0.82 0.18 ==> 0.32 0.68
THEN c:c:      BSS:3.32     0.42 0.58 ==> 0.84 0.16
Ridge information      dBSSrate      NewBSS      Ridge Upper      --> Ridge Lower
Pre:left-outside      [FLEX<=0.05]      -1.84      4.34      0.45 0.55 / 53      --> 0.50 0.50 / 22
Pre:left-outside      [HBD>2]          -1.02      6.69      0.36 0.64 / 36      --> 0.00 0.00 / 0

*UL:Rule1 [Up] rlv: 0.59 Tanimoto: 0.32 R&S: 51 R-S: 36 S-R: 74
           [Low] rlv: 0.86 Tanimoto: 0.57 R&S: 12 R-S: 2 S-R: 7
Rule 1-UL7: Cases: 87 -> 14; BSS= 4.402
IF [HOMO<=-0.25] added on [FLEX>0.05] [LUMO>-0.15] [C-C=C: n] [C-O: n]
THEN MM:      BSS:4.40      0.37 0.63 ==> 0.93 0.07
THEN FLEX:     BSS:7.39      0.00 0.66 0.13 0.22 ==> 0.00 0.00 0.00 1.00
THEN VOLUME:   BSS:3.35     0.23 0.51 0.26 ==> 0.79 0.14 0.07
THEN HBD:      BSS:3.14     0.46 0.34 0.20 ==> 1.00 0.00 0.00
THEN HBA:      BSS:6.05     0.24 0.36 0.40 ==> 1.00 0.00 0.00
THEN HOMO:     BSS:7.45     0.16 0.48 0.36 ==> 1.00 0.00 0.00
THEN C-C:      BSS:2.81     0.45 0.55 ==> 0.00 1.00
THEN C-C-Cl:   BSS:5.92     0.79 0.21 ==> 0.14 0.86
THEN c:c:      BSS:6.66     0.31 0.69 ==> 1.00 0.00
Ridge information      dBSSrate      NewBSS      Ridge Upper      --> Ridge Lower
Pre:left-outside      [LUMO<=-0.15]     -2.62      2.35      0.32 0.68 / 25      --> 0.17 0.83 / 6
Pre:left-outside      [FLEX<=0.05]     -1.60      2.44      0.46 0.54 / 65      --> 0.45 0.55 / 11

*UL:Rule1 [Up] rlv: 0.59 Tanimoto: 0.31 R&S: 50 R-S: 35 S-R: 75
           [Low] rlv: 0.75 Tanimoto: 0.52 R&S: 12 R-S: 4 S-R: 7
Rule 1-UL9: Cases: 85 -> 16; BSS= 4.623
IF [C-C-Cl: y] added on [CLOGP>0] [FLEX>0.05] [C-c:c:c:c: n] [C-O: n]
THEN MM:      BSS:4.62      0.40 0.60 ==> 0.94 0.06
THEN FLEX:     BSS:6.19      0.00 0.58 0.08 0.34 ==> 0.00 0.00 0.00 1.00
THEN VOLUME:   BSS:2.78     0.25 0.58 0.18 ==> 0.69 0.19 0.12
THEN HBD:      BSS:2.55     0.54 0.27 0.19 ==> 1.00 0.00 0.00
THEN HBA:      BSS:5.97     0.25 0.28 0.47 ==> 0.94 0.06 0.00
THEN HOMO:     BSS:3.04     0.25 0.46 0.29 ==> 0.75 0.19 0.06
THEN C-C-Cl:   BSS:10.5     0.81 0.19 ==> 0.00 1.00
THEN c:c:      BSS:4.69     0.46 0.54 ==> 1.00 0.00
Ridge information      dBSSrate      NewBSS      Ridge Upper      --> Ridge Lower
Pre:left-outside      [FLEX<=0.05]     -1.41      3.46      0.50 0.50 / 52      --> 0.57 0.43 / 7

```

Figure 1. Output for Rule 1 (part)

であっても[Dipole: 高]ならば不活性である」ことを述べているが、同じ内容は表 2 Rule5 の関連ルールに含まれている。他方、表 2 中の他のルールからは、これまで見過ごされてきた多くの新しい知見を得ることができる。

### 3. 結論

データスケープを志向した新しいルール群表現により化学発ガン性データを解析したところ、これまでのルール群表現と比較して、仮説生成のための知見を容易に得ることができるようになった。新しいルール表現法の利点をまとめると、次のようになる。

- (1) ルール数が少なく、かつ相互に関連するルールをまとめて視察することができる。従って、ルール群を視察する際に、一つの特徴的セグメントを順に多方面から眺めることとなり、思考が中断することなく考察を深めることができる。これまでの方法では、何らかの基準で選択したルールのみを考察の対象とし、他のルールは捨て去っていた。今回のルール群では、その全体を一通り読むことができるようになった。
- (2) **ridge** 情報に従って、注目すべき属性の選択が簡単に行える。多数の前提条件が存在しても、**ridge** として現れる属性に沿って、仮説を考えればよい。従って、解析者の思考効率が格段に向上したと言える。
- (3) 出力されるルール数が少ないにもかかわらず、ルール最適化の効果により、これまで得られなかった強いルールを得ることができる。

実際に専門家が解析可能な数のルール群内に、これまでは得られなかった新しい知見が見いだせることは、ルールによるデータ解析を行う立場からすれば大きな進歩であるといえる。実際、本稿では省略したが、これまでは強いパターンが見つからなかった **rat** データに対しても、有効なルールを検出することができた。

他方、今後改良すべき点も存在する。第 1 の問題点は、ルールの最適化により、本質的ではない前提条件節がルールに入り込み、表現が複雑化してしまうことである。この問題は AIC や MDL の方法を適用することにより、最適なルールを見出す必要性を示している。第 2 の点は、**BSS** 値の局所的な極大値が多数存在し、特に意味のない関連ルールが出現することである。関連ルールとして組織化されているため、視察時における解析者への負荷はそれほど高くないが、今後何らかの方法で無意味なルールを除去することが望まれる。

### 参考文献

- [1] 岡田: データマイニングが拓く分子・生体の相互作用, 化学と工業, 54 pp.560-563 (2001).
- [2] Helma, C. et al.: The Predictive Toxicology Challenge for 2000-2001, <http://www.informatik.uni-freiburg.de/~ml/ptc/index.html>.
- [3] Okada, T.: Characteristic Substructures and Properties in the Chemical Carcinogenicity Studied by the Cascade Model, Proc. Int. Workshop on Predictive Toxicology Challenge 2001, Freiburg (2001), also submitted to Bioinformatics.
- [4] Okada, T.: Datascape Survey using the Cascade Model, submitted to KDD2002, (2002).
- [5] Okada, T.: Topographical Expression of a Rule for Active Mining, in press, IOS press (2002).
- [6] Okada, T.: Rule Induction in Cascade Model based on Sum of Squares Decomposition, in Zytkow, J.M. & Rauch, J. (Eds.): *Principles of Data Mining and Knowledge Discovery PKDD'99*, pp.468-475, LNAI 1704, Springer-Verlag (1999).
- [7] 岡田: アクティブマイニングのためのルール群表現法, 人工知能学会, 第 46 回基礎論 (SIGFAI) 第 54 回知識ベースシステム (SIG-KBS) 合同研究会, SIG-FAI/KBS-J-36 (11/14), pp. 219-224, 函館 (2001).