

共起成分の含意関係に基づくデータマイニングの実験と考察

Experiments and Considerations on Data Mining

Based on Cofactor Implication

二木 克也*

Niki Katsuya

湊 真一†

Minato Shin-ichi

Abstract: We report some unknown relations among the items in the database can be extracted by focusing on cofactor implication. Cofactor implication can be found by using ZDD which is efficient algorithm for processing data. However, as the means has strict restrictions for irregular data, some meaningful relations might be left unfound. In this paper, we propose the way for extracting such linkage, and studied what can be extracted, especially from the database which consists of one-hot type items.

Keywords: BDD, ZDD, data mining, transaction database, cofactor implication, sets of combinations

1 まえがき

近年、大規模記憶装置の発展などによって、大規模なデータベースの中から有用な規則を発見するデータマイニングの研究が盛んになっている。頻出アイテム集合マイニング (Frequent Itemset Mining) は、最も基本的なデータマイニングの問題の1つであり、Agrawal 等 [2] による Apriori アルゴリズムの研究に始まり、現在までに様々なアルゴリズムが提案されている。

我々はこれまでに、VLSI CAD の分野で大規模論理関数データの表現法として広く用いられている二分決定グラフ (BDD: Binary Decision Diagrams) [3], その中でも「ゼロサプレス型 BDD (ZDD: Zero-suppressed BDD) [4]」と呼ばれるデータ構造を用いて、トランザクションデータベースにおける頻出アイテム集合を効率よく生成する手法に関する研究を進めている。ZDD を用いることにより、大規模な組合せ集合データを非明示的に列挙し、頻出アイテム集合の発見から解析に至る多様な演算を効率よく実行することができると期待されている。

一方、ZDD を用いることにより、「共起成分の含意関係」を検出する手法が湊らにより提案されている [1]。共起成分の含意関係が成り立つとは、2つのアイテム a ,

b に関して、 a と共起する成分の集合と、 b と共起する成分の集合との間に含意関係が存在することをいう。共起成分の含意関係を抽出し、直接的な含意関係と比較することにより、アイテム同士に直接的な含意関係がないにも関わらず、共起成分には含意関係があるようなアイテムの組を検出することができる。これらのアイテムの組には、これまで発見されていなかった、何らかの興味深い関係が存在している可能性がある。一例として、mushroom [7] というデータベース例題より共起成分の含意関係を抽出した場合、このデータベースは n 者択一型のデータベースで、抽出された関係は全て、それぞれ同じ選択肢に属する複数のアイテム間の関係であった [11]。

しかしながら、現在提案されている手法では、 a と共起する成分と b と共起する成分の包含関係に、わずかでも破れがあれば、どんなに強く共起成分の含意関係に準ずる性質をもつ関係であっても、これを抽出していない。そこで、本稿では、このような例外を許容した共起成分の含意関係を抽出する手法を提案し、抽出した結果が持つ意味、ならびに、その有用性について考察する。

2 共起成分の含意関係と ZDD

2.1 ZDD とトランザクションデータベース

BDD は、図 1 に示すような論理関数のグラフによる表現である。一般に、論理関数のそれぞれの変数につい

*北海道大学大学院情報科学研究科アルゴリズム研究室, 060-0814 札幌市北区北 14 条西 9 丁目, e-mail niki@mx-alg.ist.hokudai.ac.jp
Algorithm Laboratory, Graduate School of Information Science and Technology, Hokkaido University, Sapporo, 060-0814 Japan.

†e-mail minato@ist.hokudai.ac.jp

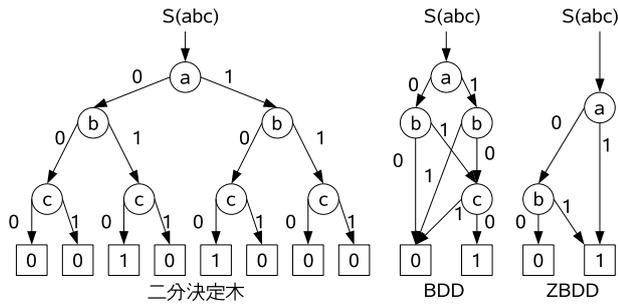


図 1: 二分決定木と BDD, ZDD

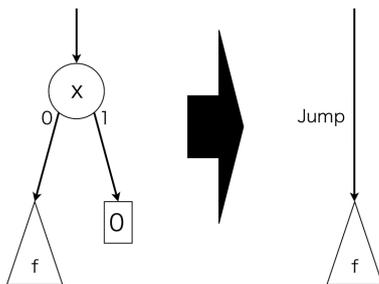


図 2: ZDD の簡約化規則

て、0, 1 の値を代入した結果を、二分岐の枝 (0-枝/1-枝) で場合分けし得られる論理関数の値を、2 値の定数節点 (0-終端節点/1-終端節点) で表現すると、図 1 のような二分木状のグラフになる。このとき、場合分けする変数の順序を固定し、冗長な節点の削除と等価な節点の共有という 2 つの縮約規則を可能な限り適用することにより、「規約」な形が得られ、論理関数をコンパクト且つ一意に表せることが知られている。

BDD は元々は論理関数を表現するために考案されたものだが、これを用いて組み合わせ集合データを表現・操作することも出来る。組み合わせ集合とは、「 n 個のアイテムから任意個を選ぶ組み合わせ」を要素とする集合である。これを BDD で表現するとき、類似する組み合わせが多ければ、部分的に共通する組み合わせがグラフ上で共有されて、記憶量や計算時間が大幅に削減される場合がある。さらに、組み合わせ集合に特化した「ゼロサブレス型 BDD (ZDD)」[4] を用いると、より簡潔な表現が得られ、一層効率よく扱うことが出来る。

ZDD では、冗長な節点を削除する簡約化規則が通常の BDD と異なり、1-枝が 0-終端節点を直接指している節点を取り除く (図 2)、という規則になっている。これにより ZDD では図 1 のように、組み合わせ集合に一度も選ばれないアイテムに関する節点が自動的に削除されることになり、BDD よりも効率よく組み合わせ集合を表現・操作することが出来る。

アイテム	a	b	c	d	e	f	g	h	i
レコード1	○	○		○		○			
レコード2	○		○					○	
レコード3	○	○			○		○		
レコード4				○		○			
レコード5									○
レコード6	○		○				○		
レコード7	○	○	○			○			○

図 3: トランザクションデータベースの例

本稿では、複数個のアイテムの組合せが 1 つ 1 つのレコードとなっている「トランザクションデータベース」を対象とする。図 3 に簡単な例を示す。トランザクションデータベースでは、個々のレコードは各アイテムが出現するかどうかの組合せのみを記録しており、アイテムの順序は問わない。つまり、 ab と ba を区別しない。一般的なデータマイニングでは、1 つのデータベースに同じアイテム組合せが複数回出現するとき、その出現回数を「頻度」として意味を持たせるが、本稿では、問題をまず単純化して、1 回以上表れたアイテム組合せを集合として扱うこととする。なお、頻度を扱う拡張も可能である。

2.2 共起成分の含意関係

共起成分の含意関係 [1] が成り立つとは、2 つのアイテム a, b に関して、 a と共起する成分の集合と、 b と共起する成分の集合との間に含意関係が存在するをいう。

データベースに含まれる全てのアイテムの組合せの集合をあるアイテム a を含むか含まないか、および、アイテム b を含むか含まないかで $S_{a\bar{b}}, S_{\bar{a}b}, S_{a\bar{b}}, S_{ab}$ の 4 通りに分類する (図 4)。このうち、 a のみと共起する成分の集合 ($S_{a\bar{b}}$) が、 b のみと共起する成分の集合 ($S_{\bar{a}b}$) に含まれるとき、すなわち、 $S_{a\bar{b}} \supseteq S_{\bar{a}b}$ が成り立つとき、アイテム a はアイテム b に対して共起成分の含意関係にあると (以下、 $a \rightarrow b$ のように記述) いうことができる。また、このとき、 $b \rightarrow c$ が成り立つならば、 $a \rightarrow c$ も成り立つ。一方、 $S_{a\bar{b}}$ が空集合であるとき、 a そのものが必ず b と共起するので、すなわち、 a は b に対して直接的含意関係にあるということができる。

例えば、以下のような S という集合があれば、

$$S = \{bc, acd, ce, ac, abe, bd, bcd\}$$

次のような集合へ分けられる。

$$S_{a\bar{b}} = \{ce\}$$

$$S_{\bar{a}b} = \{c, d, cd\}$$

$$S_{a\bar{b}} = \{cd, c\}$$

$$S_{ab} = \{e\}$$

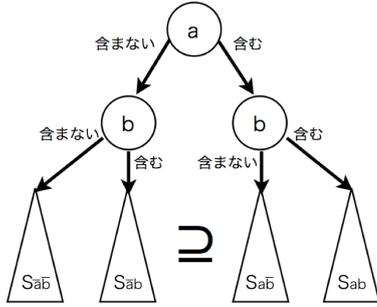


図 4: 共起成分の含意関係

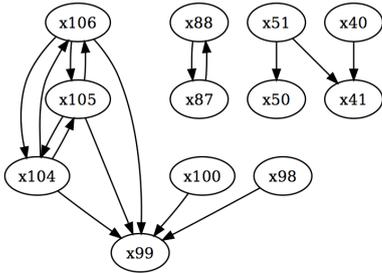


図 5: グラフ化した場合のクラスタ例

この例では、 $S_{\bar{a}b} \supseteq S_{a\bar{b}}$ が成り立つので、 $a \rightarrow b$ が成り立つ、一方、 $S_{\bar{a}\bar{b}} \supseteq S_{ab}$ は成り立たないので、 $b \rightarrow a$ は成り立たない。

共起成分の含意関係を抽出し、直接的な含意関係と比較することにより、アイテム同士に直接の含意関係がないにも関わらず、共起成分には含意関係があるようなアイテムの組を検出することができる。既存の研究の多くは、直接的な含意関係に基づいているのに対して、共起成分の含意関係に基づいた研究は少ないので、これらのアイテムの組には、これまで発見されていなかった、何らかの興味深い関係が存在している可能性がある。直接的な含意関係とは、図 4 でいうところの $S_{\bar{a}\bar{b}}$ 、もしくは、 S_{ab} のどちらか一方が空集合である状態なので、本稿では、 $S_{\bar{a}\bar{b}}$ と S_{ab} がどちらも空集合ではない場合のみを指して、共起成分の含意関係として、取り扱うこととする。

特に、抽出された関係について、グラフ化すると、いくつかのクラスタを形成する場合があることがわかっている。図 5 は、本稿で実験に使用したデータベース例題 mushroom [7] より、共起成分の含意関係を抽出し、グラフを作成したものから、一部を抜粋したものである。このクラスタに関するデータ構造には、何かしらの有意な情報が含まれている可能性が高い。

なお、 n 個のアイテムが存在するならば、共起成分の含意関係を抽出するためには、 $n(n-1)$ 通りのアイテムのペアから制約条件を満たすものを全て抽出する必要があるが、ZDD を用いることにより、データベース中の

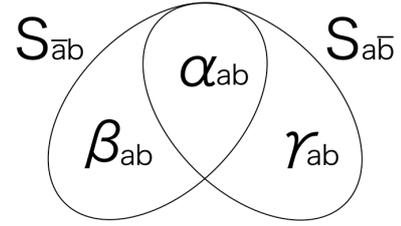


図 6: $S_{\bar{a}b}$ と $S_{a\bar{b}}$ の関係

全てのアイテムのペアに対して、約 n 倍高速なアルゴリズムで検出することが可能である [1]。

これまでの実験結果としては、現実のインフルエンザのアミノ酸配列のデータベースから、全ての共起成分の含意関係を、現実的な時間で抽出することに成功している [5]。また、そのようなアルゴリズムを実装したプログラムとして、VSOP [6] が公開されている。

3 誤差を許容した共起成分の含意関係

共起成分の含意関係を別の図 (図 6) を用いて説明する。この図は、先ほどの 4 グループのうち、共起成分の含意関係を判定するにあたり必要となる $S_{\bar{a}b}$ と $S_{a\bar{b}}$ の関係を表したものである。 $S_{\bar{a}b}$ にも $S_{a\bar{b}}$ にも含まれるものを α_{ab} 、 $S_{\bar{a}b}$ に含まれるが、 $S_{a\bar{b}}$ には含まれないものを β_{ab} 、反対に $S_{a\bar{b}}$ に含まれるが、 $S_{\bar{a}b}$ には含まれないものを γ_{ab} という 3 つの集合に分けられる。すなわち、

$$\alpha_{ab} = S_{\bar{a}b} \cap S_{a\bar{b}},$$

$$\beta_{ab} = S_{\bar{a}b} \setminus S_{a\bar{b}},$$

$$\gamma_{ab} = S_{a\bar{b}} \setminus S_{\bar{a}b}$$

とする。このとき、 $\alpha_{ab} = \emptyset$ 、かつ、 $\gamma_{ab} = \emptyset$ であるならば、 $S_{\bar{a}\bar{b}} = \emptyset$ なので、直接的な含意関係となり、 $\alpha_{ab} \neq \emptyset$ 、かつ、 $\gamma_{ab} = \emptyset$ であるならば、 $S_{\bar{a}b} \supseteq S_{a\bar{b}}$ が成り立つので、 $a \rightarrow b$ という共起成分の含意関係が成り立つ。例えば、前章で用いた例では、

$$S_{\bar{a}b} = \{c, d, cd\}$$

$$S_{a\bar{b}} = \{cd, c\}$$

であるので、次のように分類できる。

$$\alpha_{ab} = \{c, cd\}$$

$$\beta_{ab} = \{d\}$$

$$\gamma_{ab} = \{\emptyset\}$$

γ が空集合となっていることから、 $a \rightarrow b$ が成り立つことが確認できる。また、 $\alpha_{ab} = \alpha_{ba}$ 、 $\beta_{ab} = \gamma_{ba}$ である。

ところで、もし、 γ_{ab} に一つでも要素があったとしたら、どうだろう。仮に、ある組み合わせについて、上のよう分類したところ、 α_{ab} が数千パターン存在する一方で、 β_{ab} と γ_{ab} がともに一つずつ要素を持っていたと

カテゴリ	A			B		C			
アイテム	a ₁	a ₂	a ₃	b ₁	b ₂	c ₁	c ₂	c ₃	c ₄
レコード1		○		○		○			
レコード2			○	○				○	
レコード3			○		○		○		
レコード4	○			○		○			
レコード5		○			○				○
レコード6	○			○			○		
レコード7			○	○					○

図 7: 択一式データベースの例

すると、その組み合わせは、共起成分の含意関係にはないので、従来手法では、その関係を抽出することができなかった。しかしながら、そのような場合においても、共起成分の含意関係に準ずる関係にあると推認できるし、また、 β_{ab} や γ_{ab} に含まれているデータ自体が誤りであったり、例外的なデータであったりする可能性は少ない。したがって、このような関係の中にも、なにかしらの重要な関係が隠れているかもしれない。そこで、本稿では、 α_{ab} に何かしらの要素が含まれているものの、 β_{ab} や γ_{ab} が空集合でないために共起成分の含意関係としては、抽出できなかった関係を抽出し、そのような関係が持つ意味について、議論する。

なお、本稿では、共起成分の含意関係を抽出する場合と同様に、VSOP [6] を用いて実験を行ったが、このような例外を許した共起成分の含意関係に準じた関係を抽出する処理は実装されていないので、3つの段階に分けて、関係を抽出した。

まず、 α_{ab} が空集合ではない組み合わせを、VSOP により抽出し、それぞれに含まれるパターン数を数える。次に、 α_{ab} が空集合でない組み合わせに関して、 β_{ab} と γ_{ab} それぞれに含まれるパターン数を数える。そして最後に、 α_{ab} 、 β_{ab} 、 γ_{ab} のパターン数を比較することで、例外を許した共起成分の含意関係を抽出する。 α_{ab} に対して、 β_{ab} 、 γ_{ab} が十分に小さいことをもって、誤差の範囲であるとし、例外を許した共起成分の含意関係にあるとするが、比較する基準を変更することで、抽出できるパターン数がどのように変化するかについても考察する。本稿では、 α_{ab} が β_{ab} もしくは γ_{ab} の x 倍以上のサイズを持つアイテムの組み合わせについて、 x を変化させて、複数の条件で抽出する。また、誤差を最大限に許す場合として、 α_{ab} が空集合でない組み合わせを全て抽出する。

4 択一式データベースにおける共起成分の含意関係

択一式データベースでは、図 7 のように、データベースに含まれるアイテムがいくつかのカテゴリに分類でき、全てのカテゴリから一つずつ選択されることで、1 件のレコードが複数の排他的選択肢の集合であるようなデータ構造になっており、それぞれのレコードに含まれるアイテム数はカテゴリ数に等しい。ここで、この択一式データベースより共起成分の含意関係を抽出した場合、カテゴリを跨いだクラスタが検出されることが理論上あり得ないことを示す。例えば、図 4 で a が A というカテゴリに属するアイテムで、 b が B というカテゴリに属するアイテムだとすると、 S_{ab} には、カテゴリ B に属するアイテムがない組み合わせだけが含まれるのに対して、 $S_{\bar{a}\bar{b}}$ は、カテゴリ A に含まれるアイテムがない組み合わせの集合となることから、これらに共通する α_{ab} は、絶対に存在し得ないからである。このことから、択一式データベースにおいては、本手法を用いた上で、例外を最大限許容することで、完全ではないものの、アイテムをカテゴリごとに分類することができる可能性がある。

5 実験と考察

5.1 択一式データベースにおける実験

まず、誤差を許容することにより、新たに抽出することができるようになった関係が、どのような意味を持つのか調べるために、データベース中の各アイテムの意味が分かっているデータベースである mushroom (FIMI ベンチマーク [7] より) に対して、本手法を適用した。mushroom はキノコの生態に関するデータベースで、各アイテムは色やにおい、生息場所などを表している。データベース中には 119 個のアイテムが存在し、それらは 23 のカテゴリのいずれかに分類され、それぞれのカテゴリの中からアイテムが一つずつ選択されることにより、8124 件のレコードが表現されている。例えば、赤いキノコのレコードであれば、色のカテゴリから赤というアイテムだけがそのレコードに含まれ、紫や橙といったアイテムが同時に含まれることはない。においなど、他のカテゴリについても、同様であるので、各レコードは 23 個ずつのアイテムを含む。また、例外を許さない共起成分の含意関係は 53 組である。

本手法を実行したところ、 α_{xy} が空集合ではない (x, y) のペアが 106 組存在した。これらのペアを全て許容し、グラフを描写した場合、23 のカテゴリの内、6 カテ

	$\beta=0$ $\gamma=0$	$\alpha \geq \beta$ $\alpha \geq \gamma$	$\alpha \geq 2^* \beta$ $\alpha \geq 2^* \gamma$	$\alpha \geq 5^* \beta$ $\alpha \geq 5^* \gamma$	$\alpha \geq 10^* \beta$ $\alpha \geq 10^* \gamma$	$\beta \geq \gamma$ $\gamma \geq \beta$
検出数	53	97	82	68	62	120
新しい関係	共起成分の 含意関係	44	29	15	9	67

図 8: mushroom からの抽出結果

ゴリーで、全てのアイテムを含むクラスタが抽出でき、その他のカテゴリーに関しても、過半数のアイテムがクラスタとして抽出しているものが多かった。

次に、 α と β , γ の差が 10 倍を超えるほどの差がある場合を考えてみる。これを例外を許した共起成分の含意関係とするならば、抽出された関係は 62 組、すなわち、新たに抽出できるようになった関係は 9 組である。この内、6 組は例外を許さない共起成分の含意関係を抽出し、グラフ化した場合でも存在していたクラスタ内のアイテム同士の関係で、残りの 3 組は、全く新しい関係であった。中でも、several と solitary の関係などは、双方の意味が近いことから、データベースへ登録した観察者が同種のキノコを別の表現を用いてデータベースへ記録したことに由来している可能性がある。

なお、実行環境は 2GHz Intel Core2 Duo, 4GB 1067MHz DDR3 である。また、以降の実験についても同様の環境で実行した。

次に、択一式で、より大規模、かつ、実用的なデータベースに対して、本手法を適用した。このデータベースには、インフルエンザウイルスのアミノ酸配列が登録されている。登録されているレコード数は 1750 件で、アイテム数は 1182 個である。また、共起成分の含意関係は 77 組抽出できている。このデータベースでは、位置情報とその場所を占めるアミノ酸の種類が一つのアイテムとなっていることから、特定の位置に関して、唯一のアミノ酸の種類を選択するという択一式のデータベースとなっている。

本手法を適用すると、 α が空集合ではないペアが 337 組存在した。これらのペアは、やはり、同じ選択肢に含まれるアイテム同士の組み合わせであった。このデータベースでは、mushroom の場合と比べ、 β , γ が α よりも圧倒的に大きい組み合わせが目立った。これは、インフルエンザウイルスのアミノ酸が変化する中で、当該組み合わせが変異する前と後、両方の配列が登録されている場合、この二つの配列が α の成分として、検出されている可能性がある一方で、当該組み合わせ以外でも変化が起こっている場合、その結果として生まれた配列が β や γ の成分として分類されるためである可能性があると考えられる。このため、例外を許した共起成分の含意関係として、新しい関係を抽出することはできなかったが、先の仮定が正しいという前提に立てば、 β と γ の大

	$\beta=0$ $\gamma=0$	$\alpha \geq \beta$ $\alpha \geq \gamma$	$\alpha \geq 2^* \beta$ $\alpha \geq 2^* \gamma$	$\beta \geq \gamma$ $\gamma \geq \beta$
検出数	77	116	77	338
新しい関係	共起成分の 含意関係	39	0	261

図 9: インフルエンザウイルスデータからの抽出結果

	$\beta=0$ $\gamma=0$	$\alpha \geq \beta$ $\alpha \geq \gamma$	$\alpha \geq 2^* \beta$ $\alpha \geq 2^* \gamma$	$\alpha \geq 3^* \beta$ $\alpha \geq 3^* \gamma$	$\beta \geq \gamma$ $\gamma \geq \beta$
検出数	25	525	69	25	91289
新しい関係	共起成分の 含意関係	500	44	0	91264

図 10: BMS-Web-View-1 からの抽出結果

小関係に基づいて、グラフを作成すると、各クラスタ内の関係性に注目することにより、有意な情報を抽出できる可能性があるのではないかと思料される。

5.2 非択一式データベースにおける実験

択一式ではないデータベースに本手法を適用した場合は、どのような関係が抽出できるのか調べるために、インターネット通販サイトの閲覧履歴のデータベースである BMS-Web-View-1(FIMI ベンチマーク [7] より) へ適用した。このデータベースには、497 個のアイテムからなる、42629 件のレコードが登録されている。一つのアイテムはの通販サイト上のページ一つ一つに対応しており、一つのレコードは、一人の顧客が訪問した際、閲覧したページの情報で構成されている。なお、例外を許さない共起成分の含意関係は 25 組抽出されている。

本手法を適用したところ、 α が空集合ではないペアは 91139 組存在した。この、データベースもインフルエンザウイルスのデータベースと同様に、 α が β , γ よりも小さいことから、例外を許した共起成分の含意関係として、新たに抽出された関係はなかった。また、 α が空集合でないペアの数もかなり大きな数字であることから、少なくともこのデータベースに関しては、本手法を用いることによって、有意な関係を抽出することは難しい可能性が高い。

6 結論

本稿では、例外を許容した共起成分の含意関係を検出する手法を提案し、この手法を実際のデータベースへ適用する方法について述べた。択一式のデータ構造を持つデータベースに適用したところ、そのデータ構造を大まかではありながらも、抽出することができた。しかしながら、例外を許さない共起成分の含意関係からの拡張という観点からみると、観察者の視点というアナログ

的要素を持つ mushroom においては、有意な関係を抽出することに成功したが、インフルエンザウイルスのアミノ酸配列データでは、新しい関係を抽出することはできなかった。むしろ、例外を許さない共起成分の含意関係が成り立つ状況こそが特異な状況であると言えるかもしれない。また、非択一式構造のデータベースであるネット通販サイトのページ閲覧履歴データでは、本手法を用いて、有用な関係性を抽出することはできなかった。

なお、今後の展開としては、以下のようなものが挙げられる。

- 択一式データベースで、本手法により、抽出した結果をグラフ化した際、カテゴリーごとのクラスタが形成されるが、そのクラスタ内での関係性から、有用な情報が抽出できるか検討し、抽出できるのであれば、既存の手法と比較する。
- 本稿では、例外を許すことが有用であるか否かに着目したために、全てのアイテムのペア（アイテム数の2乗個）についての検査を単純に繰り返し実行しているが、アルゴリズムを工夫することにより計算回数を削減し高速化できる可能性があると思われる。
- 本稿では、個々のアイテムのペアに関する共起成分の含意関係について考察したが、アイテムの組合せのペアに関する共起成分まで議論を拡張し、その有用性や計算時間について検討する。

7 謝辞

ご指導をいただいた北海道大学アルゴリズム研究室 Thomas Zeugman 教授に感謝いたします。またご討論いただいた岩崎玄弥さんを始めとする同研究室の方々に感謝いたします。

参考文献

- [1] 湊: “「共起成分の含意関係」を満たすアイテム集合のデータマイニング,” 人工知能学会 第65回人工知能基本問題研究会 資料, SIG-FPAI-A603-10, pp. 53–58 (Mar. 2005)
- [2] R. Agrawal, H. Mannila, R. Strikant, H. Toivonen and A. I. Verkamo, “Fast Discovery of Association Rules,” In *Advances in Knowledge Discovery and Data Mining*, MIT Press, 307–328 (1996)
- [3] R. E. Bryant : “Graph-based algorithms for Boolean function manipulation,” *IEEE Transactions on Computers*, Vol. C-35, No. 8, pp. 677–691 (1986)
- [4] S. Minato: “Zero-Suppressed BDDs for Set Manipulation in Combinatorial Problems,” In *Proc. of 30th ACM/IEEE Design Automation Conference (DAC’93)*, pp. 272–277 (Jun. 1993)
- [5] S.Minato and K.Ito: “Symmetric Item Set Mining Method Using Zero-suppressed BDDs and Application to Biological Data,” *Trans. of the Japanese Society of Artificial Intelligence*, Vol. 22, No. 2, pp. 156–164 (Feb. 2007)
- [6] 湊: “VSOP:ゼロサプレス型BDDに基づく「重み付き積和集合」計算プログラム,” 電子情報通信学会コンピュータシミュレーション研究会, 信学技報 Vol. 105, No. 72, COMP2005-10, pp. 31–38 (May. 2005)
- [7] B. Goethals, M. Javeed Zaki (Eds.), *Frequent Itemset Mining Dataset Repository, Frequent Itemset Mining Implementations (FIMI’03)*, <http://fimi.cs.helsinki.fi/data/> (2003)
- [8] 湊, 有村: “ゼロサプレス型二分決定グラフを用いたトランザクションデータベースの効率的解析手法,” 電子情報通信学会論文誌, Vol. J89-D, No. 2, pp. 172–182 (Feb. 2006)
- [9] 湊: “データベースの頻出アイテム集合を表すゼロサプレス型 BDD の変数順序付けの理論的考察,” 電子情報通信学会コンピュータシミュレーション研究会, 信学技報 Vol. 106, No. 566, COMP2006-55 pp. 37–42 (Mar. 2007)
- [10] 湊: “二分決定グラフ (BDD) を活用したデータマイニング・知識発見技術の最近の話題,” 電子情報通信学会 人工知能と知識処理研究会, 信学技報, Vol. 107, No. 78, AI2007-6, pp. 27–32 (May. 2007)
- [11] 湊: “「共起成分の含意関係」に基づくデータベースからの知識発見,” 人工知能学会 第4回 データマイニングと統計数理研究会 (SIG-DMSM) 講演報告集, SIG-DMSM (July. 2007)