3J3-3in

関連尺度に基づいた負の相関ルール抽出手法の高機能化

Effective Mining of Top-k Negative Association Rules Based on Relevance Measures

黒岩健歩 *1　岩沼宏治 *2　山本泰生 *2
Yasuho Kuroiwa　Kojo Iwanuma　Yoshitaka Yamamoto

*1山梨大学大学院医学工学総合教育部コンピュータ・メディア工学専攻
Computer Science and Media Engineering, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi

*2山梨大学大学院医学工学総合研究部
Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi

Positive association rules represent the co-occurrence relations of itemsets. In contrast, negative association rules represent some relationships between presence and absence of itemsets. The negative association rule mining has to deal with a huge amount of itemsets of absence. Therefore, negative rules mining is known as a quite difficult task. [Ide 14] was proposed some perfect and effective mining techniques of negative association rules to compute in the framework of support and confidence. The purpose of this study is to add new effective evaluation methods of negative rules with statistical measures. Thereby, valid negative rules can be further filtered out. Also, we propose a branch and bound mining of top-k rules and a filtering method using weak relevance to negative rules mining. Thus, we realize efficient and effective negative rules mining. We also show some good results of experiments for evaluating our proposed method.

1. はじめに

本論文では, 先行研究 [井出他 14] のトップダウン型の負の相関ルール抽出アルゴリズムの高度化を目的として, 統計的評価尺度および弱関連性を利用する上位 k 個の効果的な抽出法を考察し, 効率的に負の相関ルールを抽出する新たなアルゴリズムを提案する.

相関ルール発見問題は, データマイニングや知識発見の代表的な問題として知られている [TSK08]. 相関ルールとは, トランザクションデータベース中で同時に発生することの多い事象同士の強い共起関係を記述したものである. データベース中でアイテム集合 X が出現するトランザクションに同時にアイテム集合 Y が出現することが多いことを, $X \Rightarrow Y$ と記述する. これを**正の相関ルール**と呼ぶ. 本研究で扱う**負の相関ルール**は $X \Rightarrow \neg Y$, $\neg X \Rightarrow Y$, $\neg X \Rightarrow \neg Y$ と表記され, アイテム集合の出現と非出現の関係を表す規則である. 負の相関ルールは近年研究が盛んになった分野 [WZC08, WZZ04] であり, 正の相関ルールでは発見されない知識を提供し, 有益な情報を与える. しかし, 正の相関ルールに比べて, 負の相関ルールは非出現のアイテム集合を含むためにその数は膨大となる. そのため, 負の相関ルール抽出問題は困難であることが知られている.

先行研究 [井出他 14] ではトップダウン型の負の相関ルール抽出アルゴリズムが提案された. これは, 負の相関ルールを完全かつ効率的に抽出する手法であり, 著者の知る限り最も効率的な手法であるが, 負の相関ルールの評価尺度として支持度, 確信度のみを使用している. 本研究では, ルールの評価尺度に新たな統計的評価尺度を追加する. これによりルールをさらに絞り込み, より有効な負の相関ルールの抽出が可能になる. 本研究では更に関連研究 [亀谷他 11] の頻出パターン発見法を参考にし, 分枝限定法により探索空間の枝刈りを行い, 弱関連性を適用した上位 k ルール抽出法を提案し, 効率的に有効な負の相関ルールの抽出を行う. 最後に提案アルゴリズムを実装し, 性能評価実験を行った結果を示す.

本研究は, 負の相関ルールにおける評価尺度として lift, Φ 係数のみを考慮するが, これらは他の評価尺度を加える上での基盤となるものである.

論文の構成は以下の通りである. 第 2 章は, 準備として各種定義を行う. 第 3 章では, 負ルールにおける関連尺度を定義し, 有効な負の相関ルールの条件を定める. 第 4 章では, 関連尺度に基づく効率的な負の創刊ルールの抽出法を示す. 第 5 章では, 提案手法を実装し, 性能評価実験を行った結果および考察を示す. 第 6 章は, まとめとする.

連絡先: 黒岩健歩, 山梨大学大学院医学工学総合教育部
コンピュータ・メディア工学専攻,
〒 400-8511 山梨県甲府市武田 4-3-11
E-maill: t10kg010@yamanashi.ac.jp

2. 準備

2.1 正の相関ルール

$I = \{a_1, a_2, ..., a_n\}$ を**アイテム**の全体集合とするとき, **トランザクション** t をアイテムの集合 $t \subseteq I$ と定める. **トランザクションデータベース** D をトランザクションの多重集合とする. X をアイテム集合とするとき, $X \subseteq t$ となる D 中のトランザクション t を X の**出現**と呼び, その集合を $D(X)$ と略記する. 集合 A の大きさを $|A|$ と表記するとき, X の D 中の**支持度** $\sup(X)$ を, $\sup(X) = \frac{|D(X)|}{|D|}$ と定義する.

正の相関ルール (以下, 適宜 "正ルール" と略記) を $X \cap Y = \emptyset$ であるアイテム集合 X, Y からなる表現 $X \Rightarrow Y$ と定める. X と Y をそれぞれルールの**前件**, **後件**と呼び, $X \cup Y$ を**台集合** (underlying set) と呼ぶ. 正ルールに対する**支持度** sup と**確信度** conf を以下のように定義する.

$$\sup(X \Rightarrow Y) = \sup(X \cup Y), \quad \mathrm{conf}(X \Rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)}$$

最小支持度 ms と**最小確信度** mc とは, ユーザが支持度と確信度に関して与える閾値である. $\sup(X) \geq ms$ を満たす X を**頻出アイテム集合** と呼ぶ. また $\sup(X \Rightarrow Y) \geq ms$ と $\mathrm{conf}(X \Rightarrow Y) \geq mc$ の両方を満たす $X \Rightarrow Y$ を**有効** (valid) な正の相関ルールと呼ぶ.

2.2 負の相関ルールの定義

先行研究 [井出他 14] にならい, 負の相関ルールの定義を示す. **負の相関ルール** (negative association rule: 以下では適宜 “負ルール” と略記) は, X と Y を $X \cap Y = \emptyset$ であるアイテム集合とするとき, 以下のいずれかの表現である.

$X \Rightarrow \neg Y$ (**右否定形**もしくは**後件負形**),
$\neg X \Rightarrow Y$ (**左否定形**もしくは**前件負形**),
$\neg X \Rightarrow \neg Y$ (**両否定形**)

上記の $\neg X$ はアイテム集合の否定表現であり, **負アイテム集合**と呼ぶ. 負アイテム集合内のアイテムは論理積で関係づけられているとする. つまり, $X \Rightarrow \neg\{a, b\}$ は $X \Rightarrow \neg(a \wedge b)$ と解釈し, 「X が出現する場合, a, b のどちらか一方は出現しないことが多い」を表していると考える. $X \Rightarrow (\neg a \vee \neg b)$ と変形できるので, **否定和形**と呼ぶ. 否定和形の負ルールの支持度は, 下記に示すように正のアイテム集合の支持度を基に計算でき, 正の相関ルールマイニングで開発された技術を比較的容易に転用できる.

以下では C_X は, アイテム集合 X または負アイテム集合 $\neg X$ のどちらかを表すものとする. $C_X \Rightarrow C_Y$ に対して $X' \subseteq X$, $Y' \subseteq Y$ なる $C_{X'} \Rightarrow C_{Y'}$ を**部分ルール**と呼ぶ.

定義 1 (負ルールの支持度, 確信度) 負アイテム集合および負ルールの**支持度** sup と**確信度** conf を以下のように定める.

$$\begin{aligned}
\sup(\neg X) &= 1 - \sup(X) \\
\sup(X \Rightarrow \neg Y) &= \sup(X) - \sup(X \cup Y) \\
\sup(\neg X \Rightarrow Y) &= \sup(Y) - \sup(X \cup Y) \\
\sup(\neg X \Rightarrow \neg Y) &= 1 - \sup(X) - \sup(Y) + \sup(X \cup Y) \\
\mathrm{conf}(C_X \Rightarrow C_Y) &= \frac{\sup(C_X \Rightarrow C_Y)}{\sup(C_X)}
\end{aligned}$$

先に示した両否定形 $\neg X \Rightarrow \neg Y$ は, 一般に非常に数が多い. そのため, 両否定形の効率的な抽出は困難である. また, ルールとしての有用性も低いことが通常である. そのため本論文では, 右否定形 $X \Rightarrow \neg Y$ および左否定形 $\neg X \Rightarrow Y$ に焦点を絞って考察を進める.

3. 有効な負の相関ルール

3.1 関連尺度の定義

本節では, 負ルールの抽出に追加する統計的評価尺度を定義する. 以降, 統計的評価尺度を前件と後件の**関連尺度**と呼ぶ. 新たに加える関連尺度として, 2 つの評価尺度を考える. 1 つは lift[TSK08] である. lift は, 相関ルール発見問題で用いられる代表的な尺度であり, 前件と後件の独立性を測る指標である. もう 1 つは **Φ 係数** [TSK08] である. Φ 係数は, 2 確率変数間の関連性を測る尺度である. この 2 つの関連尺度を以下のように定義する.

定義 2 (関連尺度: lift, Φ 係数)

$$\mathrm{lift}(C_X \Rightarrow C_Y) = \frac{\mathrm{conf}(C_X \Rightarrow C_Y)}{\sup(C_Y)} = \frac{\sup(C_X \Rightarrow C_Y)}{\sup(C_X)\sup(C_Y)}$$

$$\Phi(C_X \Rightarrow C_Y) = \frac{\sup(C_X \Rightarrow C_Y) - \sup(C_X)\sup(C_Y)}{\sqrt{\sup(X)\sup(\neg X)\sup(Y)\sup(\neg Y)}}$$

lift, Φ 係数は, 正の相関ルールにおける表現の自然な拡張となっている. lift は確率の比で独立性を見ている. Φ 係数は, 差で独立性を見ている. 負ルールにおける関連尺度の定義として妥当なものと考える.

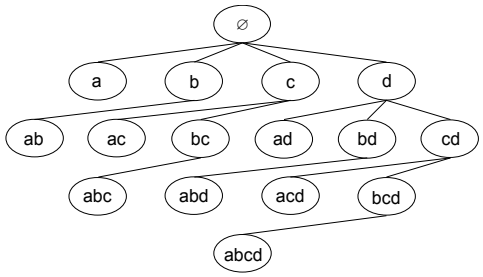

図 1: 接尾木の例

3.2 有効な負の相関ルールの定義

ルール $C_X \Rightarrow C_Y$ に対して関連尺度の値を $R(C_X \Rightarrow C_Y)$ とし, mr を関連尺度の閾値とする. 先行研究 [井出他 14] の有効な負の相関ルールの条件に関連尺度を加える.

定義 3 関連尺度に基づく有効な負の相関ルール $C_X \Rightarrow C_Y$ とは, 以下の 6 つの条件を満たすルールである.

(1) $X \cap Y = \emptyset$
(2) $\sup(X) \geq ms$ かつ $\sup(Y) \geq ms$
(3) $\sup(X \Rightarrow Y) < ms$
(4) $\sup(C_X \Rightarrow C_Y) \geq ms$
(5) $\mathrm{conf}(C_X \Rightarrow C_Y) \geq mc$
(6) $R(C_X \Rightarrow C_Y) \geq mr$

(6) が関連尺度の条件である. (3) は [井出他 14] で提案された無矛盾性条件である. これは, 正ルール $X \Rightarrow Y$ が有効である場合, 同様のアイテム集合を持つ負ルール $C_X \Rightarrow C_Y$ が同時に有効である状態を矛盾とし, 負ルールの抽出を行わない条件である.

4. 負ルール抽出手法

本章では, 先行研究 [井出他 14] の手法および本論文での提案手法について示す.

4.1 先行研究の提案手法

本手法は, 頻出アイテム集合を節点とする図 1 のような**接尾木** [亀谷他 11] の組合せ探索により, 負ルールを抽出する. アイテムの間には適当な順序 $\prec$ を仮定し, アイテム集合をアイテムの列として取り扱う. 図 1 ではアルファベット順 $a \prec b \prec c \prec d$ を仮定している. 各節点 N_c の親は, 長さが 1 つだけ短い接尾辞 (suffix) をもつ節点 N_p である. 子 N_c と親 N_p の差分アイテムは, $\prec$ 上において N_p 中のアイテムより前にあるものである. 兄弟関係にある節点は $\prec$ に基づく辞書順で左から右へ並ぶ. 接尾木上で左優先深さ優先探索を行うと, 節点 N を訪問する時点で N の部分集合は全て訪問が完了している. これは後で示す弱関連性において包含関係の検査をする上で都合が良い. また, 負ルールは接尾木を左優先深さ優先で探索する. 先行研究 [井出他 14] で利用しているトップダウン型のアルゴリズムは, 包含関係のあるルール間の関係性の検査が容易であるために, 効率よく探索空間を削減することを可能にする. 本手法も同様にトップダウン型のアルゴリズムを採用する.

本手法では, 重複性検査と次節で示す分枝限定法による枝刈り操作を使用する. 重複性検査は定義 3 の条件 (1) を保証するものであり前件, 後件のアイテム集合の独立性条件を用いた枝刈り手法である.

命題 1 (重複性の単調性) $X \cap Y \neq \emptyset$ ならば, $Y \subset Y'$ である Y' に対して $X \cap Y' \neq \emptyset$ が必ず成り立つ.

X と Y の重複性を検査し, 重複部分があれば子節点を枝刈りする. この枝刈は命題 1 よりその安全性が保証される.

4.2 分枝限定法

lift, Φ 係数を含め, 関連尺度は逆単調性を満たさないものが多い. そのような場合, 不用意に探索空間の枝刈りを行うと, 関連尺度の高いルールを見落としてしまう恐れがある. そこで分枝限定法を用いた枝刈りを使用する. [井出他 14] では, 確信度は下記の上界関数を定義して分枝限定操作を実現している.

定義 4 (上界関数: 確信度)

$$\overline{\mathrm{conf}}(\neg X \Rightarrow Y) = \frac{\sup(Y)}{1-\sup(X)}$$

同様に, 本研究で定義した lift, Φ 係数についても, 負ルールの包含関係に関して, 逆単調性を満たす上界関数を以下のように定義できる.

定義 5 (上界関数: 関連尺度)

$$\overline{\mathrm{lift}_R}(X \Rightarrow \neg Y) = \frac{1}{1-\sup(Y)}$$

$$\overline{\mathrm{lift}_L}(\neg X \Rightarrow Y) = \frac{1}{1-\sup(X)}$$

$$\overline{\Phi}(C_X \Rightarrow C_Y) = \frac{\sup(X)\sup(Y)}{\sqrt{\sup(X)\sup(\neg X)\sup(Y)\sup(\neg Y)}}$$

このとき, 以下が成り立つ.

命題 2 (上界関数の逆単調性) X, X', Y, Y' をアイテム集合とし, $X \subset X', Y \subset Y'$ と仮定する. R を $\overline{\mathrm{lift}_R}$, $\overline{\mathrm{lift}_L}$, $\overline{\Phi}$ のいずれかの関数とするとき以下が成り立つ.

1. $\overline{R}(C_X \Rightarrow C_Y) \geq R(C_X \Rightarrow C_Y)$
2. $\overline{R}(C_X \Rightarrow C_Y) \geq \overline{R}(C_{X'} \Rightarrow C_{Y'})$

命題 2 より, 上界関数は評価尺度の上界を成し, 逆単調性が成り立つことが保証されるため, 接尾木上での枝刈り操作に用いることができる. 即ち, 関連尺度に注目すると, 閾値 mr に対して $\overline{R}(C_X \Rightarrow C_Y) < mr$ ならば, $C_X \Rightarrow C_Y$ 自身および X, Y のアイテム集合を拡張したルール $C_{X'} \Rightarrow C_{Y'}$ は, 全て閾値を満たさないことが保証される. そのため, 即座に枝刈りを行うことができる.

4.3 上位 k 負ルール抽出

ユーザ指定の k に対し, 関連尺度に基づく上位 k 個の負ルールを抽出する手法を提案する.

相関ルールを抽出する際, 適当な閾値を設定されないと有益でないルールが大量に抽出される. それに伴い計算コストも増加する. この問題を解決する手法として, **上位 k 負ルール抽出法**がある. これは, ユーザがルール数 k を指定し, 関連尺度に基づく上位 k 個の負ルールを抽出する手法である. この手法の利点は, 関連尺度の閾値が, データベースに依存し自動調整されることにある.

本手法では, 左否定形, 右否定形に対し, それぞれ上位 k ルールの候補リストを持たせる. そのため, 左否定形と右否定形それぞれ別の閾値を持つことに注意していただきたい.

ここで右否定形に注目し, その閾値を mr_R とする. 上位 k 個の候補リストのうち, k 番目のルールの関連度を R^k とする. このとき mr_R は, 常に R^k の値で更新できる. なぜなら, 関連度が R^k 未満である場合, 最終的に有効な上位 k 個のルールとして抽出されることがない. そのため, R^k を閾値として更新できる (閾値上昇法 [亀谷他 11, HWLT02]). 閾値を上昇させることは, 上界関数の枝刈りを促す操作であるため, 分枝限定法による探索空間の削減を助長させる効果がある.

4.4 負ルールにおける弱関連性

相関ルールでは, 評価尺度が高いルールにつられて, 同じアイテム集合を含むルールの評価値も高くなる. そのため, 上位 k 抽出手法では, 類似したルールが抽出の多くを占めることが経験的に知られている. ここで, 似たルールを冗長と判断し, 非冗長なルールのみを抽出する手法を適用する. 以下では [亀谷他 11] によって提案された, パターン間の「より弱い (weaker)」という関係の自然な拡張として, 負ルールにおける弱関連性を次のように定義する.

定義 6 (負ルールの弱関連性) 右否定形の $rule1$: $X1 \Rightarrow \neg Y1$, $rule2$: $X2 \Rightarrow \neg Y2$ において, $X1 \subseteq X2$ かつ $Y1 \subseteq Y2$ で $R(rule1) \geq R(rule2)$ ならば, $rule2$ は $rule1$ より**弱い**.

左否定形の $rule3$: $\neg X3 \Rightarrow Y3$, $rule4$: $\neg X4 \Rightarrow Y4$ において, $X3 \supseteq X4$ かつ $Y3 \supseteq Y4$ で $R(rule3) \geq R(rule4)$ ならば, $rule4$ は $rule3$ より**弱い**.

負ルールは弱関連性に関して強いものを残す. よって右否定形は負ルールの弱関連性に関して極小なルール, 左否定形は極大なルールを有効なルールとして残す.

接尾木上でルールを探索する場合, ある負ルール $C_X \Rightarrow C_Y$ を探索する段階において, その部分ルールに対して全て探索済みである. そのため上位 k ルール抽出において弱関連性の検査は, 候補リストのルールのみと比較するだけでよい.

4.5 負ルール抽出アルゴリズム

本節では, 提案するアルゴリズムの概要を示す. 頻出アイテム集合の抽出には LCMver.2[宇野] を使用する. 以下では要素数 k の頻出アイテム集合の集合を FISS^k と表記する. mr_R, mr_L をそれぞれ左否定形, 右否定形の関連尺度の閾値とし, 負ルール抽出アルゴリズム [*1] の概要を以下に示す.

Input: トランザクションデータベース D, 最小支持度 ms, 最小確信度 mc, 抽出ルール数 k
Output: 上位 k ルールの右否定形 RL, 左否定形 LL

上界関数条件の真偽をみる F_L, F_R;
D から LCM によって, $\mathrm{FISS}^1, ..., \mathrm{FISS}^N$ を抽出し, FISS を要素とする接尾木を構築 (1, ...,N は接尾木を左優先深さ優先で探索した順序);
for $i = 1$ to N **do**
  $X := \mathrm{FISS}^i$;
  for $j = 1$ to N **do**
    $F_L := False$, $F_R := False$;
    $Y := \mathrm{FISS}^j$;
    if $X \cap Y \neq \emptyset$ **then**
      重複性により, Y の子孫節点を枝刈り;
    else if $\sup(X \Rightarrow Y) < ms$ and $\sup(X \Rightarrow \neg Y) \geq ms$ **then**
      if $\overline{\mathrm{conf}}(\neg Y \Rightarrow X) \geq mc$ and $\overline{R}(\neg Y \Rightarrow X) \geq mr_L$ **then**
        $F_L := True$;
        $\neg Y \Rightarrow X$ について, 確信度, 関連尺度, 弱関連性を検査し, 合格したら左否定形の候補リスト LL に追加;
      end if
      if $\overline{R}(X \Rightarrow \neg Y) \geq mr_R$ **then**
        $F_R := True$;
        $X \Rightarrow \neg Y$ について, 確信度, 関連尺度, 弱関連性を検査し, 合格したら左否定形の候補リスト LL に追加;
      end if
      if $F_L == False$ and $F_R == False$ **then**
        上界関数により, Y の子孫節点を枝刈り;
      end if
    end if
  end for
end for

5. 実験結果および考察

本研究の提案手法のアルゴリズムを実装し, 提案手法の効果を測定した実験の結果および考察を示す.

*1 支持度について $X \Rightarrow \neg Y$ のみを検査している. これは $X \Rightarrow \neg Y$ と $\neg Y \Rightarrow X$ には支持度の同値性 [井出他 14] が成り立つために, $X \Rightarrow \neg Y$ の支持度の有効性のみを検査するだけでよい.

表 2: lift に基づく上位 k 負ルール抽出の実験結果

データセット	ms	#(FIS)	sup 検査対	削減率 (%)	探索時間 (sec)	左否定ルール	右否定ルール	重複性枝刈り	上界関数枝刈り
T10I4D100K	0.01	385	146,792	0.97	22.65	0	100	412	110,853
	0.02	155	23,870	0.65	5.39	0	100	155	3,940
	0.03	60	3,540	1.67	1.06	0	100	60	0
T40I10D100K	0.02	2,293	4,954,451	5.77	941.51	0	100	86,867	4,090,603
	0.03	793	610,840	2.86	238.10	0	100	9,629	351,985
	0.04	440	191,476	1.10	105.98	0	100	1,536	47,844
	0.05	316	99,314	0.54	64.74	0	100	482	955
mushroom	0.35	1,189	334,318	76.35	20.08	100	100	117,094	28,074
	0.40	565	92,517	71.02	6.84	100	100	33,880	6,781
	0.45	329	30,406	71.91	2.81	100	100	12,710	463
retail	0.001	7,712	18,097,865	69.57	340.99	100	100	323,824	17,044,936
	0.002	2,715	2,926,877	59.58	100.75	100	100	58,267	2,757,285
	0.003	1,409	880,156	55.67	44.47	100	100	30,232	794,575
	0.004	837	351715	49.80	24.15	47	100	19,597	303,692

表 1: 実験に使用したデータ

データセット	#(item)	#(trans.)	ave(item)
T10I4D100K	870	100,000	10.1
T40I10D100K	942	100,000	39.6
mushroom	119	8,124	23
retail	16,470	88,162	10.3

Frequent Itemset Mining Dataset Repository[FIMI] から 4 種のデータセットを使用した．各データセットの詳細を表 1 に示す．そのうち T10I4D100K, T40I10D100K はランダムデータ, mushroom, retail は実データである．#(item) はデータセット中に含まれるアイテムの種類数を示し, #(trans.) はデータセット中のトランザクションの総数, ave(item) は 1 トランザクション中に出現するアイテムの平均数である．#(FIS) は頻出アイテム集合の総数である．

支持度の検査を行った頻出アイテム集合の対 $\langle X,\ Y\rangle$ を, 以下では sup 検査対と呼ぶ．重複性検査, 分枝限定法の 2 つの枝刈り手法により探索空間をどの程度減らすことができたかを**削減率**とする．削減率は以下のように示す．

$$\text{削減率} \quad = \quad 1 - \frac{\text{sup 検査対の総数}}{\text{直積 FISS}^2\text{の要素数}} \quad (\%)$$

評価尺度として lift を加え, 最小確信度 mc を 0.4, 抽出ルール数を 100 に固定し, 最小支持度 ms の値を変化させて負ルールを抽出した実験結果を表 2 に示す．

実験結果より, ランダムデータにおいて削減率は低いことがわかる．それに対して実データは, 削減率が高い値をとった．抽出される頻出アイテム集合を比較するとランダムデータは, ほとんどが要素数 1 ないし 2 のアイテム集合であった．これは接尾木の高さが低いことを示す．本手法は, 接尾木の親子節点の関係により探索空間が削減される．そのため接尾木の高さが低いと, 枝刈りによる削減の効果は低いことが考えられる．したがって, ランダムデータの頻出アイテム集合が小さいことから, 削減率が低いと推測される．明確な原因については今後検討する．

また [井出他 14] で指摘されたように, 右否定形に比べ左否定形の抽出数は少ない．これは確信度の特性によるものであり, 閾値を別々に抽出する必要性を述べていた．本研究も同様に改善されていないため, 今後の課題としたい．

6. おわりに

先行研究 [井出他 14] で提案された負の相関ルール抽出アルゴリズムに統計的評価尺度を導入し, 統計的尺度に基づく有用な負の相関ルールを効率的に抽出する手法を提案した．統計的評価尺度を用いて有効なルールを絞り込み, 有効な負ルールを抽出する．また, 効率的な抽出を実現するため, 分枝限定法により効率的に探索空間の枝刈りを行い, 弱関連性を適用した上位 k ルール抽出法を導入した．

謝辞

本研究は平成 24 年度電気通信普及財団研究助成および ISPS 科研費 25330256 の援助を受けたものです．

参考文献

[FIMI] Frequent Itemset Mining Dataset Repository, <http://fimi.ua.ac.be/> (2014-3-10).

[HWLT02] Han, J., Wang, J., Lu, Y. and Tzvetkov, P.: Mining top-K frequent closed patterns without minimum support. In Proc. of the 2002 IEEE Int'l Conf. on Data Mining (ICDM-02), pp. 211-218 (2002).

[井出他 14] 井出典子, 岩沼宏治, 山本泰生: 負の相関ルールを抽出する高速トップダウン型アルゴリズム, 人工知能学会研究会資料, SIG-FPAI, B303, pp.7-12, (2014).

[亀谷他 11] 亀谷由隆, 佐藤泰介: 最小サポート上昇法に基づく上位 k 関連パターン発見, SIG-DOCMAS, B101, pp.(2-24)-(2-32) (2011).

[TSK08] Tan, P., Steinbach, M. and Kumar, V.: Introduction to Data Mining, 769pages, Addison Wesley (2008).

[宇野] 宇野毅明: 宇野毅明と有村博紀による公開プログラム（コード）, <http://research.nii.ac.jp/ uno/codes-j.htm> (2014-3-10).

[WZC08] Wang H., Zhang, X. and Chen G.: Mining a Complete Set of Both Positive and Negative Association Rules from Large Databases. *Proc. the 12th Pacific-Asia conference on Advances in knowledge discovery and data mining (PAKDD'08)*, pp.777-784 (2008).

[WZZ04] Wu, X., Zhang, C. and Zhang, S.: Efficient Mining of Both Positive and Negative Association Rules. *ACM Trans. on Information Systems*, Vol.22(3), pp.381-405 (2004).