# Associative Classifier におけるソフトな最小支持度の導入

日下部友飛† 菊地 真人† 大囿 忠親†

† 名古屋工業大学大学院 情報工学系プログラム 〒 466–8555 愛知県名古屋市昭和区御器所町
E-mail: kusayu@ozlab.org, {kikuchi,ozono}@nitech.ac.jp

**あらまし** 相関ルールマイニングで用いられる従来の最小支持度はハードなしきい値であり，値を低くすると低頻出ルールの信頼度が過大になるため，必要以上に高くせざるを得ない場合がある．Aoba らは，頻度のしきい値に近づくほど，条件付き確率を低め（保守的）に見積もる推定法を考案した．本研究では，この保守的な推定法を応用して “ソフトな” 最小支持度を実現する．すなわち，あるルールの支持度が最小支持度に近いほど，そのルールの関連強度は 0 に近づくよう低めに見積もられる．本論文では，相関ルールに基づく分類器である Associative Classifier へのソフトな最小支持度の導入方法，およびその有効性を分類精度，実行時間，ルールの品質の観点から評価した結果を示す．

**キーワード** ソフトな最小支持度， Associative Classifier，保守的な推定法，相関ルールマイニング，確率的データ処理

## 1 はじめに

相関ルールマイニングは，大規模データベースから関連する任意のアイテム集合の組み合わせ $A$，$B$ を相関ルール $A \to B$ として網羅的に発見するタスクである．ただし，何の制約もなくマイニングするとルール数が爆発的に増え，計算的・空間的コストが大きくなりすぎる．ルール生成の際は，アイテム間の可能な全ての組み合わせを考慮するため，低頻出ルールが非常に多く生成される．またアイテム集合間における関連の強さ（関連強度）は，ルールの出現頻度に基づき条件付き確率 $P(B \mid A)$ で推測される．ゆえに，低頻出ルールから推定される $P(B \mid A)$ は不確実性が高く，関連強度が過大に評価される場合がよくある．そこで相関ルールの出現率である支持度に対して最小支持度 $minsup$ というしきい値を設け，$minsup$ を上回る高頻出ルールのみを発見の対象とすることが単純な対策となっている．以降の議論を分かりやすくするため，単純な対策を設けることになった問題点を次にまとめる．

問題 (1) 計算的・空間的コストが指数関数的に増大する．

問題 (2) 低頻出ルールの関連強度が過大評価されやすい．

$minsup$ を設けることは，$minsup$ 以下の出現率しかない低頻出ルールを一律に発見対象から除くことを意味する．しかし除かれたルールの中には，まれにしか出現せずとも興味深いルールが含まれることがある．そのため，問題 (1)，(2) の両方が他の対処法で解消されるならば，$minsup$ を低くすることは価値があると考える．上記の単純な対策は，30 年前に提案された Apriori アルゴリズム [1] で導入された．近年ではコンピュータの性能が向上し，効率の良いマイニング手法も多数提案され，問題 (1) は緩和されている．一方で近年に提案されたマイニング手法でも，$minsup$ を設ける枠組みや $P(B \mid A)$ の推定法は Apriori アルゴリズムと変わらず，問題 (2) は着目されないケースがある [2], [3]．それゆえ，問題 (2) への対処が不十分なことが，$minsup$ を低くすることの障壁となっている．

表 1: 相関ルールに対する関連強度の推定例（MC は 3）.

| 相関ルール $A \to B$ | 出現頻度 $f(A)$ | 出現頻度 $f(A \cup B)$ | $\widehat{P}(B \mid A)$ | $\widetilde{P}(B \mid A)$ |
|---|---|---|---|---|
| $A_{\mathrm{ex1}} \to B_{\mathrm{ex1}}$ | 3 | 3 | — | — |
| $A_{\mathrm{ex2}} \to B_{\mathrm{ex2}}$ | 4 | 4 | 1 | 0.25 |
| $A_{\mathrm{ex3}} \to B_{\mathrm{ex3}}$ | 20 | 20 | 1 | 0.85 |

問題 (2) について，相関ルールの例を用いて説明する．表 1 に示す 3 つの相関ルールを想定する．Apriori アルゴリズムでは，アイテム集合 $A$，$B$ 間の関連強度を信頼度

$$\widehat{P}(B \mid A) = \frac{f(A \cup B)}{f(A)}$$

で測る．ここで $f(A \cup B)$ は $A$，$B$ をともに含むトランザクション数，$f(A)$ は $A$ を含むトランザクション数である．ただし問題 (1)，(2) を回避するため，$minsup$ を上回る支持度 $support(A \cup B)$ を持つルール，すなわち条件

$$minsup = \frac{minsupCount}{|D|} < support(A \cup B) = \frac{f(A \cup B)}{|D|}$$

を満たすルールのみ発見候補とし，信頼度を計算する．$|D|$ はデータベースにある全トランザクション数である．$minsupCount$（MC）は任意の整数であり，上の条件は $minsupCount < f(A \cup B)$ と置き換え可能である．いま MC を 3 とすると，表 1 において MC 以下の頻度しかない $A_{\mathrm{ex1}} \to B_{\mathrm{ex1}}$ は発見対象とならず，信頼度は計算されない．その一方で，MC を 1 だけ超える頻度の $A_{\mathrm{ex2}} \to B_{\mathrm{ex2}}$ は発見の候補となり，信頼度が 1 と計算される．MC を大きく超える頻度の $A_{\mathrm{ex3}} \to B_{\mathrm{ex3}}$ も同じく信頼度が 1 と計算される．これらの例から分かるように単純な対策では，MC をわずかに超える頻度の $A_{\mathrm{ex2}} \to B_{\mathrm{ex2}}$ 等のルールにも，1 といった大きな信頼度を付与する場合がある．そのため $minsup$ を低くすると，偶然に出現しただけで実際は関連のない，大量の低頻出ルールが優先的に発見されてし

まう．このことから，信頼度の使用が問題のない頻度になるまで $minsup$ を引き上げる必要がある．また，MC の制約により無視されるルールの信頼度は計算されないが，あえて値を付与するならば取り得る下限値 0 と考えることができる．このように，MC に近い出現頻度を持つルールは，頻度のわずかな差で信頼度が大きく異なるケースがある．本研究では，そのような信頼度の急激な変動は不自然と考える．なぜなら，アイテム集合間における関連強度は，頻度のごくわずかな差で大きく変化するとは考えにくいためである．

本研究では，Aoba ら [4] が提案した条件付き確率の推定量を，相関ルールマイニングへと応用する．Aoba らの推定量は

$$\widetilde{P}(B \mid A) = \frac{f(A \cup B) - minsupCount}{f(A)}$$

と定義され，信頼度の代わりに上式の推定値を求める．なおこの推定量は，二乗損失を最小化する理論的枠組みで導出される．導出過程は 3 章で述べる．上式では，$f(A \cup B)$ を MC で減算し，$P(B \mid A)$ を低め（保守的）に推定する．表 1 に示すように，低頻出な $A_{\mathrm{ex2}} \to B_{\mathrm{ex2}}$ は $\widetilde{P}(B \mid A)$ が 0.25 となり 0 に近づく．対して，高頻出な $A_{\mathrm{ex3}} \to B_{\mathrm{ex3}}$ は $\widetilde{P}(B \mid A)$ が 0.85 となり，信頼度の 1 に近づく．したがって問題 (2) を緩和し，MC 付近の頻度に対する関連強度の大きな変動も防ぐ．このことを図 1 でも確認する．図の横軸は頻度 $f(A) = f(A \cup B)$，縦軸は横軸の頻度から求まる $P(B \mid A)$ の推定値である．MC を上回る頻度では信頼度 $\widehat{P}(B \mid A)$ が常に 1 と最大になっている．一方で $\widetilde{P}(B \mid A)$ は，頻度が MC に近づくほど低くなる．以上から分かるように，信頼度による従来法では，支持度が $minsup$ を超えるかが重要で，$minsup$ は信頼度の計算に関与しない．それに対して本研究の枠組みは，$minsup$ が関連強度の推定に罰則をかけるとも解釈でき，支持度が低いほど罰則が強くなる．以上から，信頼度による従来法を “ハードな” 最小支持度と呼ぶのに対し，Aoba らの推定量による本研究の手法を “ソフトな” 最小支持度と呼ぶ．

本研究の評価実験では，マイニングされた各ルールの品質が，手法の良し悪しを決める重要な要因となる．ルールの品質とは，$A \to B$ を成す $A$，$B$ に対する真の関連強度である．真の関連強度が強いほど，$A \to B$ は良いルールと判断する．本稿のソフトな最小支持度は，信頼度ベースの相関ルールマイニングによる様々な枠組みに導入できる．しかし一般の相関ルールマイニングでは，計算時間や発見されるルール数が評価軸であり，ルールの品質を正確に測ることが難しい．そこで，相関ルールに基づく分類器である Associative Classifier [5] に，ソフトな最小支持度を導入する．分類タスクであれば分類精度を算出でき，良いルール集合から得られる分類精度は高いと考えられる [6]．ゆえに，Associative Classifier を用いて分類タスクを解き，ソフトな最小支持度の有効性検証を行う．

## 2 関 連 研 究

本研究では，Aoba ら [4] による条件付き確率の推定法を Associative Classifier に応用する．この推定法では，しきい値を

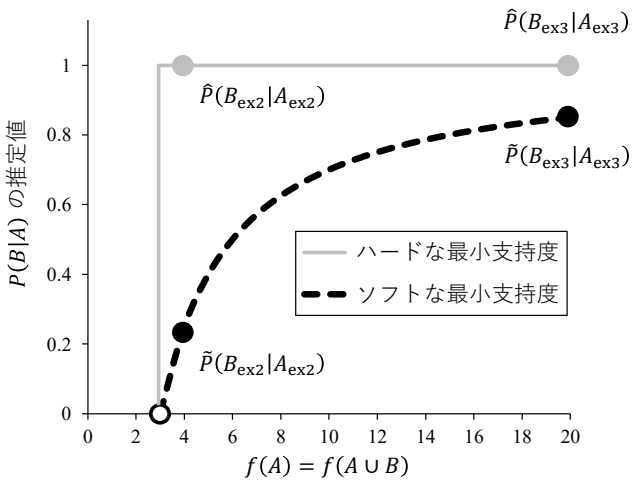


図 1: 頻度による推定値の変化（MC は 3）．図中の $\widehat{P}(B \mid A)$，$\widetilde{P}(B \mid A)$ は表 1 と対応している．

上回る頻度から条件付き確率を推定する際に，頻度の低さに応じて推定値をあえて保守的に見積もる．Aoba らは論文中で，新聞記事コーパス中から都道府県・市郡の地理的な包含関係を予測するタスクにより，手法の有効性を示した．しかしながら，このタスクでは頻度にしきい値を設けることは必須ではなく，再現率という単一の指標で有効性を評価した．本研究では，より実用的な分類タスクでの実験を行う．相関ルールを用いた分類タスクでは，計算的・空間的コストを抑えるため，しきい値 $minsup$ の導入が必須であり，分類精度に加えて実行時間とルールの品質も評価する．さらに，最適な $minsup$ が異なる複数のベンチマークデータセットを利用する．その結果，Aoba らの推定法における新たな課題が明らかになった．この課題は 6 章にて詳細に議論する．

ソフトな最小支持度の導入には，アイテム集合間に対する関連強度の推定法が深く関係する．関連強度の推定法は信頼度の他に，多数の手法が提案されてきた [7], [8]．そのうちのいくつかの推定法は Associative Classifier に導入され，分類性能の比較検討がなされている [9], [10], [11]．本研究で用いる Aoba らの推定量は，他の推定法と比較して次の特徴を持つ．まずこの推定量が，条件付き確率の真値と推定モデルとの二乗損失を最小化する理論的枠組みで導出される点である．そして，$P(B \mid A)$ をあえて保守的に見積もる考え方に基づき，推定式の内部に $minsup$ が含まれる点である．最後に，ルールの支持度が $minsup$ に近づくほど，$P(B \mid A)$ の推定値が 0 に近づく点である．特に，関連強度の推定で $minsup$ を考慮した手法はほとんど提案されていない．しかし低頻出ルールに対して形式上，関連強度の過大評価を回避できる推定法はある．そのため，6 章の課題を解決した後に，既存の推定法と提案法との比較実験を行い，ソフトな最小支持度の性質や利点を明らかにしたい．

$minsup$ の観点からも改良手法が提案されている．$minsup$ は高くしすぎると低頻出だが価値のあるルールが発見できず，低くしすぎると無価値なルールが大量に発見されてしまうト

レードオフを抱える．そこで，Liu ら [12] は様々なアイテムの性質や頻度を反映し，複数の *minsups* を設ける手法を提案した．この手法は分類タスクにも応用されている [13]．Lin ら [14] は，*minsup* をあらかじめ指定する必要があるという問題点に着目し，生成されるルール数が指定した範囲内に収まるよう，マイニング中に *minsup* を自動調整する手法を提案した．しかし，これらの手法で用いられるのは図 1 で示したハードな最小支持度である．ゆえに，それらをソフトな最小支持度に置き換えると，既存手法を改良できる可能性がある．

## 3　しきい値を設けた条件付き確率の保守的な推定法

Aoba ら [4] は，しきい値を上回る頻度から条件付き確率を推定する際に，頻度の低さに応じて推定値を保守的に見積もる手法を提案した．本章ではこの推定法を次の手順で説明する．まず，条件付き確率推定の問題設定を述べる．次に推定量の導出過程を述べる．最後に，相関ルールの関連強度を測るための応用方法を述べる．

あるデータセットが含む離散要素 $x$，$y$ の集合をそれぞれ $\mathcal{D}_X \subseteq \mathbb{U}_X$，$\mathcal{D}_Y \subseteq \mathbb{U}_Y$ とする．ここで，$\mathbb{U}_X$ は $x$ が取りうる全 $u$ 種類の離散値を含む集合，$\mathbb{U}_Y$ は $y$ が取りうる全 $v$ 種類の離散値を含む集合である．データに存在する要素ペア $(x, y) \in \mathcal{D}_X \times \mathcal{D}_Y$ の分布状況を表す同時確率関数を $P(x, y)$ とする．いま $q$ 個の要素ペア

$$\{(x_i, y_i)\}_{i=1}^{q} \overset{\text{i.i.d.}}{\sim} P(x, y)$$

からなる標本を得たと仮定する．本章では，二つの確率関数の比で表される条件付き確率関数

$$P(y \mid x) = \frac{P(x, y)}{P(x)} = r(x, y)$$

を標本 $\{(x_i, y_i)\}_{i=1}^{q}$ から推定する．$P(y \mid x)$ の素朴な推定法は，$P(x, y)$ と $P(x)$ を個別に推定して比を取ることである．しかしこの方法では，標本の要素数が低頻度のとき，推定結果が過大になる場合がある．この問題を回避するため，$P(x, y)$ と $P(x)$ の推定を介さずに尤度比関数 $r(x, y)$ を直接推定する手法 Least-Squares Conditional Density Estimation（LS-CDE）[15] が提案されている．

LS-CDE では，二乗損失の最小化により尤度比を直接推定する．$r(x, y)$ を線形和

$$\widehat{r}(x, y) = \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\phi}(x, y)$$

で表現する．$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_b)^{\mathrm{T}}$ は標本から学習されるパラメータ，$\boldsymbol{\phi}(x, y) = (\phi_1(x, y), \phi_2(x, y), \ldots, \phi_b(x, y))^{\mathrm{T}}$ は非負値を取る基底関数である．LS-CDE では扱う標本の要素が連続値のため，ガウスカーネルに基づく基底関数を用いる．これによって，連続的な標本空間の構造を尤度比推定に活用できる．しかし本研究では離散値を扱うため，ガウスカーネルが効力を発揮しない．そこで Aoba らが提案した基底関数 $\boldsymbol{\delta}(x, y) = (\delta_{1,1}(x, y), \delta_{1,2}(x, y), \ldots, \delta_{u,v}(x, y))^{\mathrm{T}}$ を代用する．$\boldsymbol{\delta}(x, y)$ の $\{(m-1) \times u + n\}$ 番目の要素は

$$\delta_{m,n}(x, y) = \begin{cases} 1 & (x = x_{(m)},\ y = y_{(n)}) \\ 0 & \text{otherwise} \end{cases}$$

と定義される．ここで $m$ と $n$ は，$x$，$y$ が取る離散値の種類を指定する添え字である．$x_{(m)}$ は $u$ 種類ある離散値のうち $m$ 番目の離散値を，$y_{(n)}$ は $v$ 種類ある離散値のうち $n$ 番目の離散値を表す．$\boldsymbol{\delta}(x, y)$ の代用に伴い，基底関数との線形和を取るパラメータも $\boldsymbol{\beta} = (\beta_{1,1}, \beta_{1,2}, \ldots, \beta_{u,v})^{\mathrm{T}}$ に置き換える．よって，$r(x_{(m')}, y_{(n')})$ に対する線形モデルは

$$\widehat{r}(x_{(m')}, y_{(n')}) = \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{\delta}(x_{(m')}, y_{(n')}) = \beta_{m',n'} \tag{1}$$

となる．線形モデル $\widehat{r}(x_{(m')}, y_{(n')})$ と真の尤度比 $r(x_{(m')}, y_{(n')})$ との二乗損失を最小化する $\boldsymbol{\beta}$ を学習する．その最適化問題は

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{u^2-u+v}} \left[ \frac{1}{2} \boldsymbol{\beta}^{\mathrm{T}} \widehat{\boldsymbol{H}} \boldsymbol{\beta} - \widehat{\boldsymbol{h}}^{\mathrm{T}} \boldsymbol{\beta} + \lambda \sum_{m=1}^{u} \sum_{n=1}^{v} \beta_{m,n} \right] \tag{2}$$

と定義される．上式には $\boldsymbol{\beta}$ に対する罰則項として $\lambda \sum_{m=1}^{u} \sum_{n=1}^{v} \beta_{m,n}$ が付与されている．$\lambda$ $(\geqq 0)$ は正則化パラメータである．なお，LS-CDE では $\boldsymbol{\beta}$ への罰則として，$\ell_2$-正則化を導入したが，本研究では Aoba らの定式化に従い，これを $\ell_1$-正則化へと変更している．また $\ell_1$-正則化項の $\beta_{m,n}$ は本来，$|\beta_{m,n}|$ であるが式 (1) に示すように，$\beta_{m,n}$ が尤度比の推定量であり非負のため，絶対値を外すことができる．$\widehat{\boldsymbol{H}}$ と $\widehat{\boldsymbol{h}}$ は

$$\widehat{\boldsymbol{H}} = \frac{1}{q} \sum_{n=1}^{v} \sum_{i=1}^{q} \boldsymbol{\delta}(x_i, y_{(n)}) \boldsymbol{\delta}(x_i, y_{(n)})^{\mathrm{T}}$$

$$\widehat{\boldsymbol{h}} = \frac{1}{q} \sum_{i=1}^{q} \boldsymbol{\delta}(x_i, y_i)$$

と定義される．式 (2) は拘束無し二次計画問題のため，目的関数を $\beta_{m',n'}$ で偏微分して 0 と置く．そして $\lambda = \frac{\lambda'}{q}$ とし，$\beta_{m',n'}$ について解くと

$$\widehat{r}(x_{(m')}, y_{(n')}) = \beta_{m',n'}(\lambda') = \frac{f(x_{(m')}, y_{(n')}) - \lambda'}{f(x_{(m')})} \tag{3}$$

が得られる．$f(x_{(m')}, y_{(n')})$ はデータセットにおける $x_{(m')}$ と $y_{(m')}$ の共起頻度であり，$f(x_{(m')})$ は $x_{(m')}$ の頻度である．本来ならば，$\lambda'$ の値によっては上式が負値となるため，負値をゼロに丸める近似が必要になる．しかし以下で述べるように，本研究では常に $\lambda' < f(x_{(m')}, y_{(n')})$ が成り立つため，上式の取る値は必ず正になり近似が不要である．

$A \to B$ に対する関連強度 $P(B \mid A)$ の推定に式 (3) を応用する．$\lambda'$ を MC とすると式 (3) は

$$\widetilde{P}(B \mid A) = \frac{f(A \cup B) - minsupCount}{f(A)} \tag{4}$$

と置換される．本研究で用いる分類器では，$minsupCount < f(A \cup B)$ を満たすルールのみ $P(B \mid A)$ を推定するため，上式の値は常に正となる．上式と信頼度との差異は分子から MC を減算するか否かのみのため，マイニングに要する時間は双方でさほど変わらず，上式は支持度と信頼度を用いた従来の様々な枠組みに応用できると考える．

## 4 使用する Associative Classifier

実験では，CBA [5] に基づく簡易な Associative Classifier を用いる．簡易な分類器を利用することで，複雑な要因による分類性能の変化をできるだけ排除する．本章では，分類器の全体をクラスアソシエーションルール（CARs）の生成とインスタンスの分類に分けて説明する．

### 4.1 分類器で扱う基本的概念

分類器について述べる前に，分類器で扱う基本的概念を説明する．本研究で扱うデータセット $D$ は，インスタンス $d \in D$ の集合である．各インスタンスは固定された複数の属性からなり，既知のクラスへと分類される．$D$ を Associative Classifier で扱うために，$d$ を複数の $(attribute,\ value)$ ペアおよび一つのクラスラベルからなる集合とみなす．ここで個々の $(attribute,\ value)$ ペアをアイテムと呼ぶ．$I$ を $D$ に含まれる全アイテムの集合，$L$ をクラスラベルの集合とする．分類には CAR と呼ばれる特殊な相関ルール $condset \to l$ を用いる．$condset \subseteq I$ はアイテムの集合，$l \in L$ はクラスラベルである．$condset \to l$ を表す $ruleitem$ は $\langle condset, l \rangle$ と定義される．$D$ において $condset \subset d$ となるインスタンス数を $condsupCount$ とする．また $condset \subset d$ であり，$l$ にラベル付けされるインスタンス数を $rulesupCount$ とする．以上より $condset \to l$ の支持度は

$$\frac{rulesupCount}{|D|}$$

と定義される．$|D|$ はデータセットの全インスタンス数である．最小支持度 $minsup$ を満たす，すなわち条件

$$minsup = \frac{minsupCount}{|D|} < \frac{rulesupCount}{|D|}$$

を満たす $ruleitems$ を頻出 $ruleitems$ と呼ぶ．CBA では，頻出 $ruleitems$ に対して信頼度を計算する．信頼度は，条件付き確率 $P(l \mid condset)$ の推定値

$$\widehat{P}(l \mid condset) = \frac{rulesupCount}{condsupCount} \tag{5}$$

として定義される．CARs の生成時には，同じ $condset$ を持つ全 $ruleitems$ に対して，最高の信頼度を持つ $ruleitem$ のみをそれらの $ruleitems$ が表す CAR として選ぶ．同じ $condset$ を持ち，かつ最高の同じ信頼度を持つ $ruleitems$ が複数ある場合は，その中から一つをランダムに選ぶ．

ソフトな最小支持度を導入する提案法では，信頼度の代わりに前章で導出した式 (4) を用い，

$$\widetilde{P}(l \mid condset) = \frac{rulesupCount - minsupCount}{condsupCount} \tag{6}$$

を計算する．同じ $condset$ を持つ複数の $ruleitems$ に対しては，信頼度を用いた場合と同様の方法で上式を用い，代表する $ruleitem$ を一つ選択する．

**Algorithm 1** CBA-RG に基づく CARs 生成アルゴリズム

```
 1: F_1 = { large 1-ruleitems };
 2: CAR_1 = genRules(F_1);
 3: for (k = 2; F_{k-1} ≠ ∅; k++) do
 4:    C_k = genCandidates(F_{k-1});
 5:    for each instance d ∈ D do
 6:       C_d = ruleSubset(C_k, d);
 7:       for each candidate c ∈ C_d do
 8:          c.condsupCount++;
 9:          if d.class = c.class then
10:             c.rulesupCount++;
11:          end if
12:       end for
13:    end for
14:    F_k = {c ∈ C_k | c.rulesupCount > minsupCount};
15:    CAR_k = genRules(F_k);
16: end for
17: CAR_all = ⋃_k CAR_k;
18: CARs_list = sortRules(CAR_all);
```

### 4.2 CARs の生成

最小支持度を超える支持度を持つ全ての頻出 $ruleitems$ を列挙し，それらを $P(l \mid condset)$ の推定値の降順でソートするアルゴリズムを Algorithm 1 に示す．このアルゴリズムは CBA-RG に基づくが，一部の処理を変更している．本節では，実験に用いる CARs 生成アルゴリズムを説明する．

Algorithm 1 において，$condset$ が $k$ 個のアイテムを持つ $ruleitem$ を $k$-$ruleitem$，頻出 $k$-$ruleitems$ の集合を $F_k$ とそれぞれ表記する．$F_k$ の各要素は

$$\langle (condset,\ condsupCount),\ (l,\ rulesupCount) \rangle$$

という形式で表される．アルゴリズムでは最初に，1-$ruleitems$ に関する頻度を計数し，$F_1$ を決定する（1 行目）．次に $genRules$ 関数を用いて，$F_1$ から $\mathrm{CAR}_1$ を生成する（2 行目）．この関数では，$F_k$ の各 $ruleitem$ に対して $P(l \mid condset)$ を推定するという，本研究で最も重要な処理を行う．また本来であれば，最小信頼度の条件判定も行い，最小信頼度を上回る信頼度を持つ CARs のみを生成する．しかし本研究では，$P(l \mid condset)$ の推定に信頼度と式 (6) のどちらを用いるかで推定値にずれが生じるため，最小信頼度に相当する均一な基準を設けることが難しい．よって，最小信頼度を設けずに $F_k$ の全 $ruleitems$ から $\mathrm{CAR}_k$ を生成する．代わりに後述する分類アルゴリズムにおいて，分類に用いるルール数 $\theta$ を指定する．

3 行目から 16 行目のループでは，CBA-RG と同様の処理を行う．まず，$genCandidates$ 関数を用いて $F_{k-1}$ から候補 $ruleitems$ の集合 $C_k$ を生成する（4 行目）．ある $k$-$ruleitem$ からアイテムを一つ除いた $(k-1)$-$ruleitems$ が，全て $F_{k-1}$ に含まれるならば，その $k$-$ruleitem$ は $C_k$ の要素になる．次に，$C_k$ の各要素に対して，支持度の計算に必要な頻度を計数する（5-13 行目）．なお，$ruleSubset$ 関数は $C_k$ と $d$ を受け取り，$condset \subset d$ を満たす $C_k$ の部分集合 $C_d$ を返す関数であ

表 2: 実験に使用するデータセット.

| データセット | インスタンス | 属性 | | クラス | 欠損値（%） |
|---|---|---|---|---|---|
| | | 連続 | 離散 | | |
| balance | 625 | 4 | 0 | 3 | 0.0 |
| breast-w | 699 | 0 | 9 | 2 | 0.3 |
| ecoli | 336 | 7 | 0 | 7 | 0.0 |
| glass | 214 | 9 | 0 | 6 | 0.0 |
| iris | 150 | 4 | 0 | 3 | 0.0 |
| lenses | 24 | 0 | 4 | 3 | 0.0 |
| mammo | 961 | 1 | 3 | 2 | 4.2 |

る．最後に $C_k$ から $F_k$ を決定し，*genRules* 関数により $F_k$ から $\mathrm{CAR}_k$ を生成する（14-15 行目）.

最終的なルール集合 $\mathrm{CAR}_{\mathrm{all}}$ を得る（17 行目）．分類を行う際にルール数 $\theta$ を指定する都合上，$\mathrm{CAR}_{\mathrm{all}}$ 中のルールをソートする処理を追加する．*sortRules* 関数により，$P(l \mid condset)$ の推定値に関してルールを降順にソートし，順序付き集合 CARs_list を得る（18 行目）．$P(l \mid condset)$ の推定値が等しいルールが複数ある場合，それらの順番をランダムに決める．

### 4.3 インスタンスの分類

CARs_list から上位 $\theta$ 件（すなわち $P(l \mid condset)$ の推定値が高い順に $\theta$ 件）のルールを指定し，分類に用いることにする．$\theta$ は実験条件として分類の前に指定する．分類したいインスタンス $d$ に対し，$\theta$ 件のルールを上位から順に走査し，$condset \subset d$ となる最初の CAR を探す．目的の CAR が見つかった場合，$d$ はその CAR のクラスに分類される．本研究において，分類を行う理由はルール集合の良し悪しを定量的に評価するためである．ゆえに，$condset \subset d$ となるルールが $\theta$ 件に含まれない場合，対象インスタンスを分類失敗として扱う.

## 5 評価実験

Associative Classifier を用いて分類タスクを解き，ソフトな最小支持度を用いた場合の有効性を検証する．一般の相関ルールマイニングでは，生成したルールの良し悪しに定量的な評価軸がなく，ルールの品質を測ることが難しい．しかし本研究では，ルールの品質が有効性の検証に重要となる．そこで相関ルールを利用した分類タスクを解き，分類精度をルール集合の良さとみなすことで，CARs の良し悪しを間接的に定量評価する．加えて，CARs の生成に要する時間を測り，ルール集合の上位にある相関ルールを直に観察する.

### 5.1 実験環境

実験環境を以下に示す.

- OS：macOS Ventura 13.0.1
- プロセッサ：1.1 GHz クアッドコア Intel Core i5
- メモリ：8.0GB
- Python：3.7.7

### 5.2 データと前処理

UCI リポジトリ [1]にある 7 種類のデータセットを分類に用いる．各データセットの情報を表 2 に示す．実験では，各データセットからインスタンスをランダムに選び，8 割を訓練データ，2 割を評価データとした．また，二つのデータセット balance と mammo は属性値の中に欠損値を含むため，連続の属性値は平均値，離散の属性値は最頻値を当てはめて欠損値補完を行った．本研究で用いる Associative Classifier の枠組みでは，離散の属性値のみを扱えるため，連続値は MDLP 法 [16] を用いて離散化した．なお実用上では，評価データを事前に得ることは困難な場合もある．そのため，欠損値と離散化に関する計算を訓練データのみに基づいて行い，その結果を評価データに適用することで，評価データ中の欠損値補完と離散化処理を行った.

### 5.3 実験手順

次の手順で分類精度を算出する．実験条件として MC は $0, 1, \ldots, 15$，使用するルール数 $\theta$ は 10，100，200，500，700，1,000 を試す．実験のランダム性を担保するため，指定した MC と $\theta$ の 2 条件に対し，次の 2 手法それぞれを用いてデータ作成から分類精度の算出までを 10 回繰り返す.

**ハードな最小支持度（Hard）：**

条件付き確率の推定に式 (5) の信頼度を用いる.

**ソフトな最小支持度（Soft）：**

条件付き確率の推定に式 (6) を用いる.

10 回に対する精度の平均値を，指定した 2 条件における各手法の最終的な分類精度とみなす．結果は $\theta$ ごとに表としてまとめ，最高の分類精度をそのときの MC と併記する．なお，同じ最高精度を持つ複数の MC がある場合は，最大の MC を記載する.

MC の各値（$0, 1, \ldots, 15$）において，2 手法が Algorithm 1 の完了に要する時間を計測する．MC を設けないとき（MC が 0 のとき）に生成される CARs が最も少ない balance，CARs が最も多いと考えられる breast-w の二つのデータセットを用いる．MC ごとに Algorithm 1 を 10 回繰り返し，10 回の平均値を算出する．なお，MC が 0 と 1 のとき，breast-w を用いた際は 24 時間経過しても Algorithm 1 が完了しないため，MC が 2 以上の場合に限り時間計測する.

Algorithm 1 を一度だけ実行し，2 手法それぞれで生成された CARs_list に含まれる上位 10 件のルールの有用性を調査する．データセットは breast-w と mammo を用い，MC は 2 と 14 で検証する．実験では，CAR の有用性を正解率によって定量化する．ある CAR を $condset \to l$ とする．正解率とは，$condset \subset d$ を満たす評価データ中のインスタンスのうち，真のクラスラベルが $l$ と一致するインスタンスの割合である．また，ここで言う有用な CARs とは，多数のインスタンスを正確に分類できる CARs である．そのため正解率とともに正解率を構成する分母と分子，すなわち $condset \subset d$ を満たす評価データ中のインスタンス数と，そのうち真のクラスラベルが $l$ と一致するインスタンス数も併記する.

1：https://archive.ics.uci.edu/ml/index.php

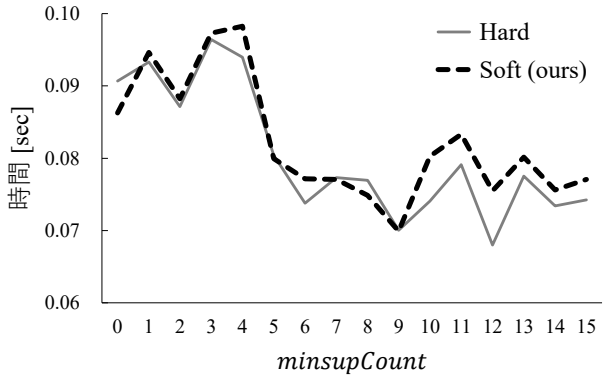


図 2: Algorithm 1 の実行時間（balance）.

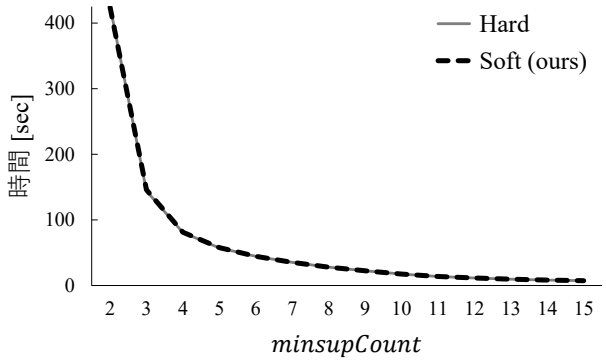


図 3: Algorithm 1 の実行時間（breast-w）.

### 5.4 実験結果

分類精度を表 3 に示す．2 手法の精度のうち，高い値に下線を引いている．MC が 0 のとき，Hard と Soft における $P(l \mid condset)$ の推定量は同じになるが，分類精度が異なる場合がある．これは Algorithm 1 で CARs_list を作成する際に，$P(l \mid condset)$ の推定値が同じ CARs は順位をランダムに決定することが影響している．ルール数 $\theta$ が 10，100，200 と小さい場合，Soft の精度が Hard を超えることが多い．よってソフトな最小支持度の導入により，インスタンスの分類に寄与するルールが上位に多く集まったことが示唆された．$\theta$ が大きい場合，Soft の MC が Hard の MC よりも低い場合が多い．これは Soft が，精度の低下を抑えつつ低頻出の CARs も扱えることを示唆する．しかしながら，Soft の精度が優れているデータセットの数は減少した．さらに，Hard よりも精度が低い場合でも，Soft の MC が低いことがある．この場合，Soft は分類精度と計算効率の両面で Hard よりも劣ることを意味する．

Algorithm 1 の完了に要する時間の計測結果を図 2 と図 3 に示す．なお，balance は生成される CARs が比較的少なく，ルール生成に要する時間が短いデータセットである．対して，breast-w は生成される CARs が比較的多く，ルール生成に要する時間が長いデータセットである．図 2 で示した balance に対する計測結果を見ると，Soft の方がやや時間がかかる傾向にあるが，両手法とも実行時間が非常に短いため，実用上でこの時間差はほとんど問題にならないと言ってよい．図 3 で示した breast-w に対する計測結果を見ると，両手法ともそれなりに実行時間はかかるが，手法間で時間の差はほとんど見られない．したがって，従来の支持度・信頼度の枠組みにソフトな最小支持度を応用しても，ルール生成に要する時間はほとんど変わらないと考える．

相関ルールを利用する際は，マイニングされた上位のルールが有用なことが求められる．そこでデータセット breast-w, mammo において，上位 10 件の CARs に対する正解率を調べた結果をそれぞれ表 4 と表 5 に示す．ここでは，多数のインスタンスによる比で構成され，1 に近い正解率を持つルールを有用とみなす．括弧内が 0/0 である場合，$condset \subset d$ を満たすインスタンスが評価データ中にないことを意味する．0/0 の CAR は良いルールとは言えないため，正解率を 0 とみなした．二つの表において MC=2 の場合，Hard は分類に寄与しない 0/0 の CARs が混在する一方，Soft は多くのインスタンスに関連し正解率が 1 に近い良質な CARs が上位にあることが分かる．MC=14 の場合は低頻出の CARs が排除され，Hard においても正解率が 1 のルールが上位のほとんどを占めた．しかし，これらのルールが関連するインスタンス数は一桁であることが多い．対して Soft は多くのインスタンスに関連し，正解率が 1 に近いルールが上位を占めた．以上より，相関ルールマイニングの観点では，Soft で生成されたルールの有用性が示唆された．

## 6 考　　察

実験ではルール数 $\theta$ が少ない場合，ソフトな最小支持度の導入によって分類精度が向上した．しかし $\theta$ を大きくし，より多くの CARs を分類に用いた場合，計算効率が落ち分類精度も低下したケースがあった．したがって，Soft は何らかの改良が必要である．そこで本章では，Soft が十分な有効性を示せなかった原因を考察する．

ソフトな最小支持度を導入する現状の方法では，条件付き確率を保守的に見積もる過程に問題がある．低い最小支持度であるほど，それをわずかに上回る支持度の CARs は信頼できず，$P(l \mid condset)$ をより保守的に見積もらなければならない．一方で高い最小支持度を設定すれば，$P(l \mid condset)$ はそれほど保守的に見積もらなくてよいと考える．しかし，本稿のソフトな最小支持度では，上記の関係が理想と逆になっている．例を用いて説明する． MC=1, 14 とし，それらを 1 だけ上回る支持度を持つ相関ルールをそれぞれ $A_{\mathrm{ex4}} \to B_{\mathrm{ex4}}$，$A_{\mathrm{ex5}} \to B_{\mathrm{ex5}}$ とする．各ルールの信頼度は両方とも 1 とする．各ルールに対して式 (4) の $\widetilde{P}(B \mid A)$ を計算すると

$$\widetilde{P}(B_{\mathrm{ex4}} \mid A_{\mathrm{ex4}}) = \frac{2-1}{2} = 0.50$$

$$\widetilde{P}(B_{\mathrm{ex5}} \mid A_{\mathrm{ex5}}) = \frac{15-14}{15} \approx 0.07$$

となり，$\widetilde{P}(B_{\mathrm{ex5}} \mid A_{\mathrm{ex5}}) < \widetilde{P}(B_{\mathrm{ex4}} \mid A_{\mathrm{ex4}})$ が成り立つ．例示したように，高頻出のルール $A_{\mathrm{ex5}} \to B_{\mathrm{ex5}}$ の推定値は 0.07 となり，信頼度の 1 よりもはるかに低くなってしまう．それに対して，2 回しか出現しない $A_{\mathrm{ex4}} \to B_{\mathrm{ex4}}$ の推定値が 0.50 と高いことは問題である．したがって，分子の頻度から MC を引く単

表 3: 各データセットに対する分類精度.

| データセット | $\theta = 10$ | | | | $\theta = 100$ | | | | $\theta = 200$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Hard | | Soft | | Hard | | Soft | | Hard | | Soft | |
| | 精度 | MC | 精度 | MC | 精度 | MC | 精度 | MC | 精度 | MS | 精度 | MC |
| balance | 55.2 | 15 | 72.8 | 15 | 75.2 | 15 | 76.8 | 15 | 75.2 | 15 | 76.8 | 15 |
| breast-w | 41.6 | 13 | 56.4 | 15 | 66.1 | 4 | 62.1 | 15 | 75.6 | 3 | 62.9 | 15 |
| ecoli | 22.5 | 15 | 20.6 | 15 | 45.6 | 15 | 60.3 | 3 | 60.3 | 15 | 64.7 | 8 |
| glass | 25.6 | 15 | 39.5 | 12 | 44.7 | 2 | 58.1 | 6 | 51.2 | 15 | 60.5 | 6 |
| iris | 74.0 | 13 | 96.7 | 15 | 86.7 | 15 | 96.7 | 15 | 86.7 | 15 | 96.7 | 15 |
| lenses | 40.0 | 9 | 40.0 | 9 | 54.0 | 0 | 50.0 | 0 | 54.0 | 0 | 50.0 | 0 |
| mammo | 7.3 | 15 | 51.3 | 4 | 80.8 | 10 | 81.3 | 14 | 82.4 | 10 | 82.4 | 2 |

| データセット | $\theta = 500$ | | | | $\theta = 700$ | | | | $\theta = 1,000$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Hard | | Soft | | Hard | | Soft | | Hard | | Soft | |
| | 精度 | MC | 精度 | MC | 精度 | MC | 精度 | MC | 精度 | MC | 精度 | MC |
| balance | 75.2 | 15 | 76.8 | 15 | 75.2 | 15 | 76.8 | 15 | 75.2 | 15 | 76.8 | 15 |
| breast-w | 85.2 | 2 | 78.6 | 5 | 89.0 | 3 | 83.3 | 2 | 92.4 | 3 | 87.9 | 3 |
| ecoli | 75.0 | 15 | 76.5 | 1 | 75.0 | 15 | 76.5 | 1 | 75.0 | 15 | 76.5 | 1 |
| glass | 60.5 | 15 | 60.5 | 6 | 60.5 | 15 | 60.5 | 6 | 62.8 | 7 | 62.8 | 2 |
| iris | 86.7 | 15 | 96.7 | 15 | 86.7 | 15 | 96.7 | 15 | 86.7 | 15 | 96.7 | 15 |
| lenses | 54.0 | 0 | 50.0 | 0 | 54.0 | 0 | 50.0 | 0 | 54.0 | 0 | 50.0 | 0 |
| mammo | 82.4 | 10 | 82.4 | 2 | 82.4 | 10 | 82.4 | 2 | 82.4 | 10 | 82.4 | 2 |

表 4: 上位 10 件の CARs に対する正解率（breast-w）.

| | $minsupCount = 2$ | | $minsupCount = 14$ | |
|---|---|---|---|---|
| 順位 | Hard | Soft | Hard | Soft |
| 1 | 0 (0/0) | 1 (67/67) | 1 (17/17) | 1 (67/67) |
| 2 | 1 (1/1) | 1 (66/66) | 1 (21/21) | 1 (67/67) |
| 3 | 0 (0/0) | 1 (62/62) | 1 (17/17) | 1 (66/66) |
| 4 | 1 (1/1) | 1 (61/61) | 1 (10/10) | 1 (65/65) |
| 5 | 0 (0/0) | 1 (60/60) | 1 (17/17) | 1 (62/62) |
| 6 | 1 (3/3) | 1 (59/59) | 1 (6/6) | 1 (61/61) |
| 7 | 1 (1/1) | 1 (59/59) | 1 (2/2) | 0.99 (69/70) |
| 8 | 0 (0/0) | 1 (57/57) | 1 (4/4) | 1 (60/60) |
| 9 | 0 (0/0) | 1 (58/58) | 1 (4/4) | 1 (59/59) |
| 10 | 1 (4/4) | 1 (56/56) | 1 (9/9) | 1 (68/68) |

表 5: 上位 10 件の CARs に対する正解率（mammo）.

| | $minsupCount = 2$ | | $minsupCount = 14$ | |
|---|---|---|---|---|
| 順位 | Hard | Soft | Hard | Soft |
| 1 | 1 (2/2) | 0.92 (12/13) | 1 (6/6) | 0.89 (72/81) |
| 2 | 1 (1/1) | 0.92 (11/12) | 1 (3/3) | 0.89 (66/74) |
| 3 | 1 (1/1) | 1 (6/6) | 1 (6/6) | 0.92 (12/13) |
| 4 | 1 (2/2) | 1 (6/6) | 1 (6/6) | 0.95 (37/39) |
| 5 | 1 (2/2) | 1 (6/6) | 1 (6/6) | 0.95 (39/41) |
| 6 | 0 (0/0) | 1 (6/6) | 1 (3/3) | 0.92 (11/12) |
| 7 | 0 (0/1) | 0.86 (12/14) | 0.92 (12/13) | 0.81 (13/16) |
| 8 | 0 (0/1) | 0.81 (13/16) | 0.92 (11/12) | 0.93 (26/28) |
| 9 | 0 (0/0) | 1 (3/3) | 1 (3/3) | 0.90 (27/30) |
| 10 | 1 (6/6) | 1 (3/3) | 0.80 (4/5) | 0.86 (12/14) |

純な方法は不十分であり，頻度の低さに応じて保守的な推定の度合いを調節する必要がある．この枠組みの考案と実現が今後の課題である．

表 4，表 5 の MC = 2 から分かるように，Hard による低頻出の CARs は，半数程度が正解率 1 であり，データセット中のノイズ（評価データ中で $condset \subset d$ を満たすが，真のクラスラベルが $l$ と異なるもの）が少ない可能性がある．そのため，扱ったデータセットは分類に適しており，解いた分類問題は簡単であったことが考えられる．この場合，Hard による信頼度を用いた枠組みでも，十分に高い分類性能を獲得でき，検証実験として不十分であったかもしれない．また一般の相関ルールマイニングでは，低頻出の相関ルールが正しい関係を表していないことが多い．したがって，より実態に即した難しい分類問題やマイニングタスクにおいて，Soft の有効性を再検証することも今後の課題である．

Soft を用いた際，個々のルールにおける分類の有用性が十分であったのに，分類精度は Hard とさほど変わらなかった理由として以下のことが考えられる．まず前段落で述べたように，低頻出の CARs でも分類に有益なものが多かったことである．次に，上位にある CARs が類似していたことである．各ルールが類似していることは目視にて確認した．二つの CARs を $condset_1 \to l_1$，$condset_2 \to l_2$ とする．CARs が類似するとは，$condset_1$ と $condset_2$ を構成するアイテム数が近く，$condsets$ 間で包含関係が成り立ち，$l_1 = l_2$ であることを指す．類似した CARs が上位に集まっていると，分類されるべきインスタンスが上位 1，2 件の CARs で全て分類されてしまい，残りの CARs が分類にほとんど使われないことがある．実験では分類に用いるルール数 $\theta$ を指定した．ゆえに分類に用いるルール集合に，類似する CARs が多く含まれると，分類性能の低下につながる．この問題は，CBA-RG で用いられていたルールの枝

刈り処理 [17] を導入することで解決できるかもしれない．また，マイニングされるルールの多様性を高めるための既存研究 [18] も活用したい．

## 7 おわりに

相関ルールマイニングにおいて，低頻出ルールはアイテム集合間の関係強度が過大推定されやすく，このことが *minsup* を低くすることの障壁であった．そこで，単に *minsup* を支持率のしきい値とする従来法（Hard）に代わり，関連強度の推定に *minsup* が罰則をかけるソフトな最小支持度（Soft）という枠組みを提案した．Soft では，*minsup* を利用して低頻出ルールに対する関連強度を保守的に見積もることができる．そして，生成されるルールに関して定量評価を行うため，Soft を Associative Classifier に導入した．複数のベンチマークデータセットを用いた実験により，Soft の有効性を分類精度，実行時間，ルールの品質の観点から評価した．その結果，ごく上位の CARs を用いる実験設定では分類精度が向上すること，分類精度の低下を抑えながら *minsup* を低くできる場合があることを示した．また，Soft が分類に要する時間は Hard とほぼ変わらず，分類に役立つ良質なルールを優先的に発見できることも示した．

しかし，使用するルール数を増やして分類を行うと，Soft は Hard よりも分類精度が劣るケースが見られた．加えて，分類精度が劣る場合でも *minsup* が Hard より低くなり，計算効率が悪くなった点も問題である．ゆえに，6 章で議論した次の 3 点の課題を踏まえ，再実験や Soft の改良が必要である．第一に，条件付き確率を保守的に見積もる度合いを，*minsup* の低さに応じて柔軟に変更できる方法の考案である．第二に，複雑な分類問題や現実的なマイニングタスクにおいて，Soft の有効性を再検証することである．第三に，マイニングされるルールの多様性を高めて，Soft を用いた際の分類精度・効率を向上させることである．以上の実験や改良が完了した後は，より新しい分類器に Soft を導入し，実用性を検証したい．また，アイテム集合間の関連強度を推定する他手法との比較も行い，Soft 独自の性質や利点を明らかにしたい．

## 謝　辞

本研究の一部は JSPS 科研費 JP19K12266，JP22K18006 の助成を受けたものです．

## 文　献

[1] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB'94)*, pp. 487–499, 1994.

[2] M. Huang, H. Sung, T. Hsieh, M. Wu, and S. Chung. Applying data-mining techniques for discovering association rules. *Soft Computing*, Vol. 24, pp. 8069–8075, 2020.

[3] E. Hikmawati, N. U. Maulidevi, and K. Surendro. Minimum threshold determination method based on dataset characteristics in association rule mining. *Journal of Big Data*, Vol. 8, pp. 1–17, 2021.

[4] T. Aoba, M. Kikuchi, M. Yoshida, and K. Umemura. Improving association rule mining for infrequent items using direct importance estimation. In *Proceedings of the 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA'20)*, pp. 1–5, 2020.

[5] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD'98)*, pp. 80–86, 1998.

[6] S. Mutter, M. Hall, and E. Frank. Using classification to evaluate the output of confidence-based association rule mining. In *Proceedings of the 17th Australian Joint Conference on Artificial Intelligence (AJCAI'04)*, pp. 538–549, 2004.

[7] R. Somyanonthanakul, M. Roonsamrarn, and T. Theeramunkong. Semantic-based relationship between objective interestingness measures in association rules mining. In *Proceedings of the 13th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP'18)*, pp. 1–8, 2018.

[8] R. Somyanonthahakul and T. Threeamunkong. Characterization of interestingness measures using correlation analysis and association rule mining. *IEICE Transactions on Information and Systems*, Vol. E103-D, No. 4, pp. 779–788, 2020.

[9] F Berzal, J. Cubero, N. Marìn, D. Sànchez, J. Serrano, and A. Vila. Association rule evaluation for classification purposes. In *Actas del III Taller Nacional de Mineria de Datos y Aprendizaje*, p. 135–144, 2005.

[10] P. J. Azevedo and A. M. Jorge. Comparing rule measures for predictive association rules. In *Proceedings of the 18th European Conference on Machine Learning (ECML'07)*, p. 510–517, 2007.

[11] M. Jalali-Heravi and O. R. Zaïane. A study on interestingness measures for associative classifiers. In *Proceedings of the 25th International Symposium on Applied Computing (SAC'10)*, pp. 1039–1046, 2010.

[12] B. Liu, W. Hsu, and Y. Ma. Mining association rules with multiple minimum supports. In *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'99)*, pp. 337–341, 1999.

[13] L. Hu, Y. Hu, C. Tsai, J. Wang, and M. Huang. Building an associative classifier with multiple minimum supports. *SpringerPlus*, Vol. 5, No. 528, pp. 1–19, 2016.

[14] W. Lin, S. A. Alvarez, and C. Ruiz. Efficient adaptive-support association rule mining for recommender systems. *Data Mining and Knowledge Discovery*, Vol. 6, pp. 83–105, 2002.

[15] M. Sugiyama, I. Takeuchi, T. Suzuki, T. Kanamori, H. Hachiya, and D. Okanohara. Least-squares conditional density estimation. *IEICE Transactions on Information and Systems*, Vol. E93-D, No. 3, pp. 583–594, 2010.

[16] U. Fayyad and K. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, pp. 1022–1027, 1993.

[17] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1992.

[18] J. Mattiev and B. Kavsek. Coverage-based classification using association rule mining. *Applied Sciences*, Vol. 10, No. 20, pp. 1–18, 2020.