

クレジットカード不正検出のルール作成支援のための情報可視化システム -An information visualization system for rule definition support of credit card fraud detection-

迫田 千華[†] 長崎 あずさ[†] 伊藤 貴之[†] 伊勢 昌幸[‡] 宮下 光輔[‡]

Chika SAKODA[†] Azusa NAGASAKI[†] Takayuki ITOH[†]

Masayuki ISE[‡] and Kousuke MIYASHITA[‡]

[†]お茶の水女子大学大学院

[†]Ochanomizu University

[‡]株式会社インテリジェントウェイブ

[‡]Intelligent Wave Inc.

1. 概要

クレジットカードの不正使用は深刻な社会問題となっており、その撲滅は非常に難しい。対策として近年、クレジットカードの不正使用をリアルタイム検出するシステムが構築されている。これらのシステムの多くは、例えば「ガソリンスタンドにて2000ドル以上のクレジットカードの使用があれば、不正の可能性があると検出せよ」というようなルールを複数設定し、このルールに該当する決済を順次警告する。しかし多くの現場においてこのルールは、新しい不正の手口が見つかるたびに、ユーザの主観によって次々と追加されているのが現状である。そのため不十分なルールや、逆に不必要なルールが設定されている事例も多い。その評価や再検討の手段として、不正使用とルールの関係を確認する情報可視化システムが有用であると考えられる。本報告では、我々が開発を進めてきたクレジットカード不正検出結果の可視化手法を体系化し、ルール作成支援を目標としてシステム化した事例を報告する。

2. 前提知識

クレジットカード決済システムは一般的に、「1決済=1レコード」として決済内容を記録する。各々のレコードには、クレジットカード番号などの個人属性、店舗番号や端末番号などの店舗属性、商品コードや金額などの購買属性、など40種類前後の属性を購買時に記録する。また、その決済が該当するルールの有無を自動判定し、存在する場合にはルール番号をレコードに付加すると同時に、必要に応じて店舗に連絡ができるようにシステム管理者に対して警告を発する。その決済が実際に不正であると発覚した時点で、当該レコードには不正種別（盗難、偽造などの種別）が追加記録される。

3. システム設計

3.1 決済レコードの分類

本システムでは前処理として、決済レコードを以下の4種類に分類する。

[検出・不正] ルール番号と不正種別の双方を有するレコード。つまりルールに該当すると検出され、実際に不正だった決済。

[検出・正当] ルール番号のみを有するレコード。つまりルールに該当すると検出されたが、実際には正当だった決済。

[不検出・不正] 不正種別のみを有するレコード。つまりルールに該当すると検出されなかったが、実際には不正だった決済。

[不検出・正当] ルール番号も不正種別も有さないレコード。つまりルールに該当せず、実際にも正当だった決済。

このように分類された決済レコードを用いて本システムでは、以下の2種類のタスクを支援する可視化手法を提供する。

[検出漏れ対策タスク] [不検出・不正]である決済に着目し、どのような新規ルールを追加設定すれば、不正決済の検出漏れを減らせるか、という考察を支援する可視化手法を提供する。

[誤検出対策タスク] [検出・正当]である決済に着目し、どのような既存ルールを修正または削除すれば、不正決済の誤検出を減らせるか、という考察を支援する可視化手法を提供する。

3.2 属性に基づくルール記述の分岐

本システムではレコードを構成する各属性を、以下の3つのタイプに分類する。

[序数] 大小関係や隣接関係に意味のある連続数値。決済年月日、決済時刻、金額などが該当する。

[名詞句] IDや選択肢としてのみ意味をもち、大小関係や隣接関係などの意味をもたない数値。ク

クレジットカード番号，不正種別などが該当する。

[準序数] 隣接関係に弱い意味のある数値。例えば商品コードには類似商品に近い番号を割り振る，店舗番号には同企業の店舗や同地域の店舗に近い番号を割り振る，といった運用が可能である。本システムではこれらの属性を，[序数]と[名詞句]の中間的な性質と位置づける。

本システムでは，[序数][準序数]に分類される属性に関するルールを，不等式で記述可能であるとする。これは具体的には，「朝 6 時以前」「10 万円以上」「商品コード X 番から Y 番まで」といったルールを設定可能であることを意味する。

また[名詞句][準序数]に分類される属性に関するルールを，等式で記述可能であるとする。これは具体的には，「クレジットカード番号が X 番」「端末番号が Z 番」といったルールを設定可能であることを意味する。

3.3 属性に基づく可視化手法の選択

本システムでは可視化手法として，1 決済を 1 ドットとして表示する以下の 2 手法を採用している。

本システムでは，[序数][準序数]に分類される属性に着目する際には，散布図による可視化画面を提供する。本システムの散布図は，あらかじめ選択された 3 属性値を，X 軸，Y 軸，色，の 3 つの視覚特性に割り当てることで不正分布を表示する。またこの散布図を 2 次元ヒストグラムに置き換えて，X 軸および Y 軸に割り当てられた 2 属性の一定区間における不正発生件数を高さで表現することも可能である。なお散布図を採用した理由は，属性の大小関係や順列関係と不正発生度の相関性を視覚的に確認しやすいからである。

また本システムでは，[名詞句][準序数]に分類される属性に着目する際には，その属性をもって決済レコードを階層的に分類し，その結果として得られる階層構造に対して可視化手法「平安京ビュー」[1]を適用する。本システムの平安京ビューは，あらかじめ選択された 3 属性値を，階層的分類指標，色，高さ，の 3 つの視覚特性に割り当てることで不正分布を表示する。平安京ビューを採用した理由は，階層的に分類された大規模点群の全体表示に適しているからである。同様な理由で別の階層型データ可視化手法 Quantum Treemap [2]を採用することも可能である。しかし文献[1]にて，平安京ビューがいくつかの数値評価指標において Quantum Treemap より望ましい結果を得ていることが示されている。

4. 属性間の相関性算出

本システムにて効果的な可視化結果を得るために

は，適切な属性の選択が重要である。本システムでは，表形式またはリレーショナルデータベース形式のデータに対して 2 属性間の相関性を算出する手法 [3]を適用して，相関性の高い属性ペアをリストアップし，そのリストの中からユーザが興味を持った属性ペアを選択することで可視化結果を得る，という仕組みを採用している。つまりユーザは，可視化の過程において全ての属性の組み合わせを試みる必要はなく，本処理において相関性が高いと判定された組み合わせの属性だけを可視化すればよい，と考えられる。

ここで属性の特徴を，以下のとおり再分類する。

- [序数].
- [準序数][名詞句]でかつ，属性値の大小関係に意味を有する属性。
- [準序数][名詞句]でかつ，属性値の大小関係に意味を有さないが，属性値間の距離に意味を有する属性。
- [準序数][名詞句]でかつ，属性値の大小関係に意味を有さず，属性値の距離にも意味を有さない属性。

2 属性間の相関性の算出には，その属性の特徴に合わせて，ケンドールの順位相関係数，標準偏差，エントロピー，ヒストグラム出現頻度上昇度，の 4 種類の算出手法を用いる。ケンドールの順位相関係数は，2 属性の両方が a)または b)に属するとき用いる。標準偏差は 1 属性が a)または b)に，他属性が a)以外に属するとき用いる。エントロピーは 1 属性が c)に，他属性が a)以外に属する時に用いる。ヒストグラム出現頻度上昇度は 1 属性が d)に，他属性が a)以外に属する時に用いる。

また，ケンドールの順位相関係数以外の 3 手法に関する処理手順は，以下のとおりである。詳細は文献[3]を参照されたい。

- 属性を 1 つ選び(以下属性 A とする)，属性 A の値に従ってデータ全体を階層的に分類する。
- 各分類について，別の属性 B の標準偏差あるいはエントロピー，ヒストグラム出現頻度上昇度を求める。
- それぞれの分類における最小値や分類全体の総和を，分類する前の標準偏差やエントロピー，ヒストグラム出現頻度上昇度と比較する。

なお本システムにおいて，必ずしも全ての属性間の相関性を算出する必要はない。40 種類前後の属性の中には，互いに独立でない属性が存在するからである。本システムでは従属関係にある 2 属性を相関性算出の対象から外し，結果として従属関係にある 2 属性を両方選んで可視化することを推奨しない。

5. 散布図による可視化

5.1 長方形領域の自動抽出

本システムでは、相関性算出結果を提示した上でユーザに3属性 P,Q,R を選択させ、その3属性から X 座標値, Y 座標値, 色相, の3変数を算出し、その結果として散布図でクレジットカード決済の分布を表示する。

本システムでは[検出漏れ対策タスク]を支援するために、[不検出・不正]に属する決済をプロットした散布図に対して、ドット密度の高い長方形領域を自動抽出し、新規ルール候補としてユーザに提示する機能を有する。この長方形領域との内外判定は、X座標値およびY座標値との不等式で表現可能である。よって散布図上で長方形領域を特定することは、上述の[序数][準序数]に分類される2属性に係るルールを特定することに等価である。散布図における長方形領域の提示例を図1に示す。

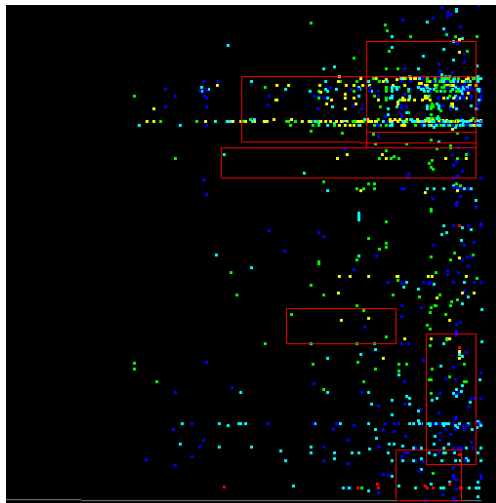


図1. 散布図における長方形領域の提示例。

長方形領域の自動抽出には、データマイニングの分野で有名な文献[4]に類似した手法を用いる。この手法では、長方形領域内部に高密度領域を含むことを「獲得」とみなし、長方形領域に低密度領域を含むことを「損失」とみなす。この考え方に基づいて長方形領域をスコア化し、できるだけスコアの高い長方形を生成する。我々の実装における処理手順を以下に示す。

1. ドット密度が高い微小長方形領域 $r_1 \sim r_n$ を生成する。
2. 既に確定している長方形領域 $R_1 \sim R_m$ のいずれかに内包されない r_i の中でドット密度が最も高いものを選ぶ。いずれも選ぶことができなければ処理を終了する。
3. 長方形領域内部の点群密度をスコアとして、 r_i

を膨張しながらスコアを更新し、損失が十分大きくなった時点で膨張処理を終了する。その時点での r_i を、確定長方形領域 R_j とする。

また[誤検出対策タスク]を支援するためには、[検出・正当]に属する決済をプロットした散布図に対して、正解率の低いルールを長方形領域として描画してユーザに提示すればよい。この長方形領域を観察することで、ルールの縮小や削除を検討できる。

5.2 適用事例

本節では、[不検出・不正]に属する決済を散布図で可視化した適用事例を報告する。



図2. X軸に時刻, Y軸に商品コード, 色に不正種別, を割り当て, 土曜の決済のみを絞り込んで可視化した事例。



図3. X軸に時刻, Y軸に商品コード, 色に不正種別, を割り当て, 木曜の決済のみを絞り込んで可視化した事例。

図2と図3に、X軸に時刻, Y軸に商品コード, 色に不正種別, の各属性を割り当て, 曜日ごとに絞り込んだ可視化事例を示す。図2は土曜, 図3は木

曜の決済のみを可視化した事例である。この可視化事例から、特定の曜日に検出漏れの不正が集中している商品コードが観察できる。この商品コードは“ガソリン・オイル類”と“インターネット通販”を表している。

さらに、この2つの商品コードについて詳しく検証した可視化結果を図4と図5に示す。

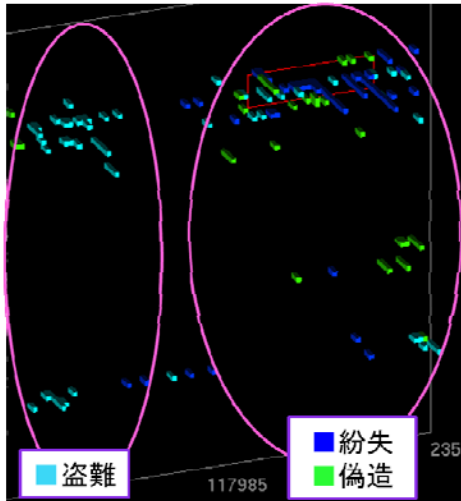


図4. X軸に時刻、Y軸に加盟店コード、色に不正種別、を割り当て、商品コードが“ガソリン・オイル類”に該当する決済のみを絞り込んで可視化した事例。

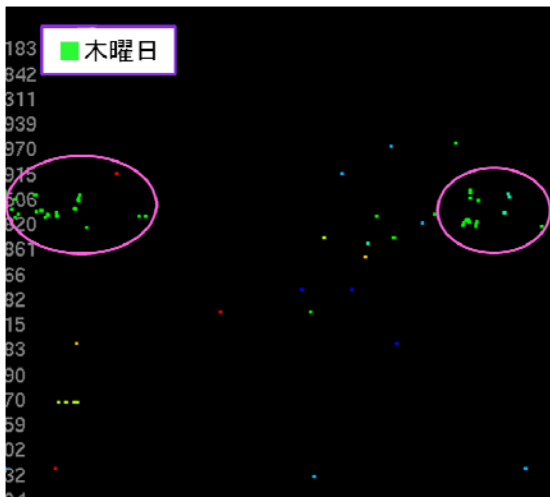


図5. X軸に時刻、Y軸にクレジットカード番号、色に曜日、を割り当て、商品コードが“インターネット通販”に該当する決済のみを絞り込んで可視化した事例。

図4は、X軸に時刻、Y軸に加盟店コード、色に不正種別、を割り当て、商品コードが“ガソリン・オイル類”に該当する決済のみを絞り込んで可視化した事例である。図4では、散布図の点の重なりを表すため、X軸、Y軸を一定間隔で分割し、その格子

状の点の個数を高さとして表示している。図4から、不正種別が盗難である決済は夜中から明け方にかけて集中していることが観察できる。一方、不正種別が紛失または偽造である決済は、午後集中していることが観察できる。

図5は、X軸に時刻、Y軸にクレジットカード番号、色に曜日、を割り当て、商品コードが“インターネット通販”に該当する決済のみを絞り込んで可視化した事例である。この可視化事例から、木曜の深夜に複数のクレジットカード番号による不正が多発していることが観察できる。

図6は、買い物の返品、店員の金額タイプミスなどに伴う取消処理によって、負の金額が記録された決済に絞って、X軸に時刻、Y軸に商品コード、色に不正種別、の各属性を割り当てて可視化した事例である。この可視化事例から、特定の商品コードに負額の決済が集中していることがわかる。さらに、不正種別は黄色で表される“カード番号盗用”に多いことが観察できる。

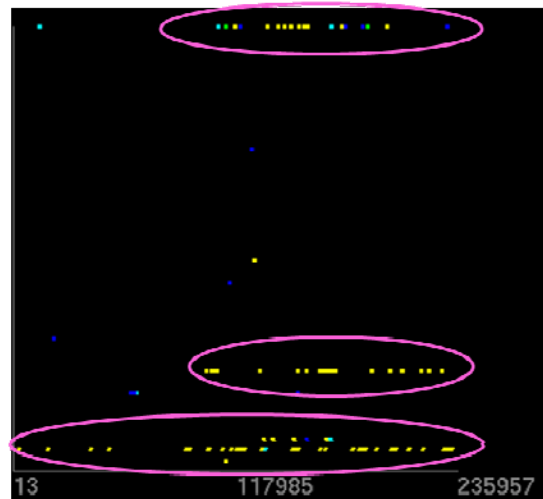


図6. X軸に時刻、Y軸に商品コード、色に不正種別、を割り当て、金額が負である負額決済のみを絞り込んで可視化した事例。

6. 平安京ビューによる可視化

6.1 可視化の処理手順

本システムでは、相関性算出結果を提示した上でユーザに3属性P,Q,Rを選択させ、Pに基づいて決済をグループ化し、Qから色、Rから高さを計算した上で、平安京ビューでクレジットカード決済の分布を表示する。

本システムでは[検出漏れ対策タスク]を支援するために、[不検出・不正]に属する決済を平安京ビューで可視化する。この際に、属性PQ間の相関性の高いグループに高い彩度、相関性の低いグループに低い彩度を与えることで、相関性の高いグループに

着目しやすい可視化結果を与える。この可視化結果において鮮やかに描画されたグループに該当するP,Qの値を読むことで、ユーザはルールを追加を検討しやすくなる。

また[誤検出対策タスク]を支援するためには、[検出・正当]に属する決済を平安京ビューで可視化して、同様に相関性の高いグループに高い彩度を与えればよい。この可視化結果において鮮やかに描画されたグループ内部の決済に注目し、その多くに該当するルールを検索することで、ユーザはルールの縮小や削除を検討しやすくなる。

6.2 適用事例

平安京ビューによる可視化事例を図7に示す。この可視化結果例では、[不検出・不正]に属する決済をクレジットカード加盟店ごとに分類し、その決済の曜日で色分け表示している。そしてグループごと時刻の標準偏差を算出し、値が小さいグループほど高い彩度を与えている。ここで図7にて(A)(B)(C)で表したグループは特に彩度が高く、可視化結果全体の中でも目につきやすい。これらのグループが示す加盟店においては、特定の曜日にクレジットカード不正利用が集中していることがわかる。これらの不正に着目して例えば「加盟店Bにおいてz曜日の決済」というような新規ルール作成を検討するために、本システムは有用であると考えられる。

別の平安京ビューによる可視化事例を図8に示す。この可視化結果例では図7と同じデータをクレジットカード加盟店ごとに分類し、決済の時刻で色分け表示している。赤色の枠で囲った店では午前1時から午前4時の間に不正使用が集中していることが読み取れる。このお店では全て商品コードが同一である。商品コードは通販を示すものであり、通販とインターネット通販は別に分類されているので、この場合の通販はカタログなどを通じた通販を示す。つまり本データにおいて、通販に関する不正は夜間に集中しているといえる。実際にクレジットカード業界では例えば、航空券や新幹線に関係ある不正被害は朝方に集中している、といった時間帯に関する知見があり、この可視化結果はその知見を裏付けているともいえる。これらの不正に着目して例えば「加盟店Aにおいてx時からy時の間の決済」というような新規ルール作成を検討するために、本システムは有用であると考えられる。また、黄色の枠で囲った店では、午前7時から午後0時の間と、特に22時台に不正使用が集中している。この店ではすべて不正種別はカード番号盗用であることから、同一犯による不正が行われた可能性がある、と予測できる。

図9は不正を商品コードで分類し、不正種別で色

分けし、平安京ビューで可視化した事例を示す。ここで赤色の枠で囲った商品では、黄色のドットが示すカード番号盗用による不正が非常に多くなっている。これらの商品はそれぞれ通販、インターネット通販、通販(流通券)、乗船券を示す。黄色で囲ったグループは、他の商品に比べて偽造による不正使用が多い。この商品は電気製品であり、電気製品には偽造による不正使用が多い、ということが言える。一方、青色の枠で囲ったレンタカー・タクシーの商品では、盗難による不正が多いと言うことができる。このように、商品によって不正種別に明確に差があることがわかる。

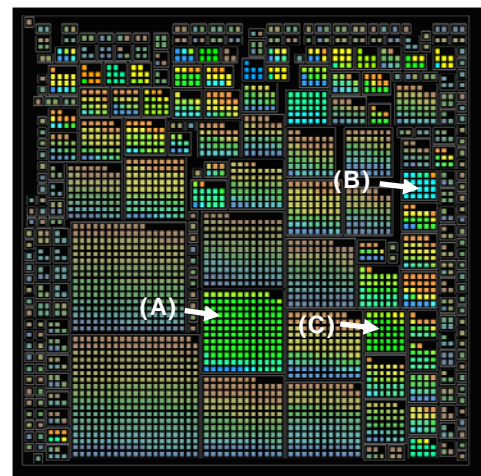


図7. 平安京ビューによる可視化事例 1.

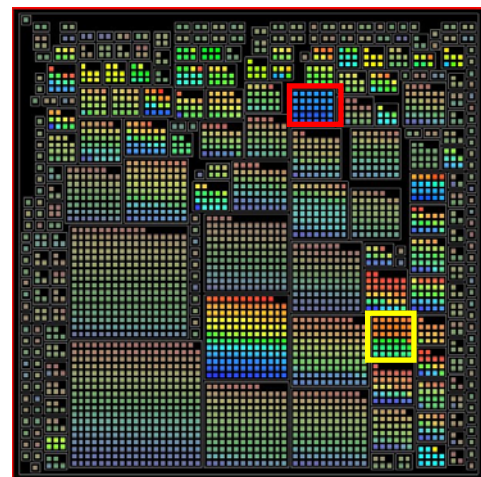


図8. 平安京ビューによる可視化事例 2.

図10には不正を対応種別で分類し、曜日で色分けして、平安京ビューで可視化した事例を示す。赤色の枠で囲った分類では、特定の対応種別が水曜日と木曜日に集中していることが見て取れる。

図11には不正を加盟店業種コードで分類し、不正種別で色分けして、平安京ビューで可視化した事例

を示す。赤色の枠で囲った分類では、特定の加盟店業種コードにおける不正種別がカード番号盗用に集中していることが見て取れる。

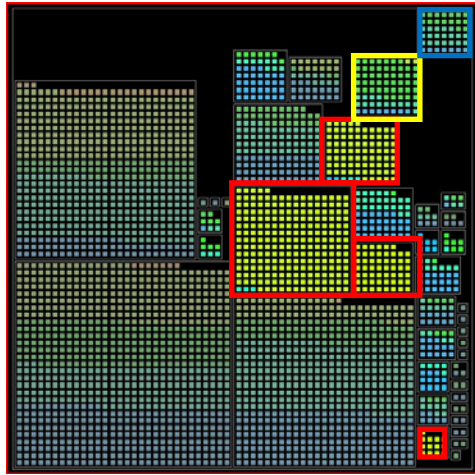


図 9. 平安京ビューによる可視化事例 3.

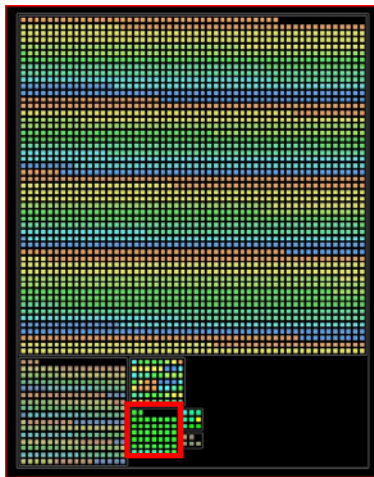


図 10. 平安京ビューによる可視化事例 4.

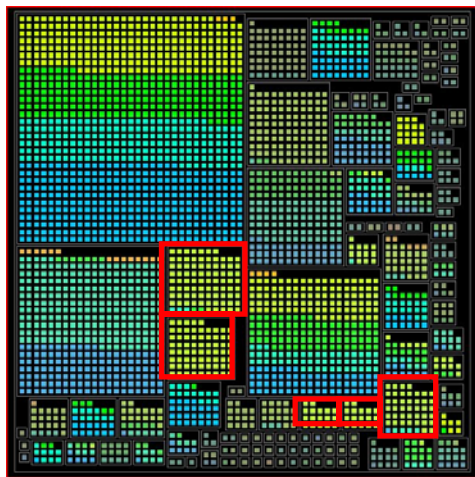


図 11. 平安京ビューによる可視化事例 5.

7. まとめ

本報告では、散布図と平安京ビューの2種類の可視化手法を適用した、クレジットカード不正検出のルール作成支援ツールとしての情報可視化システムの事例を報告した。

今後の課題としてまず、可視化手法の拡張を試みたい。特に、数十万件規模の決済を可視化するための詳細度制御技術の適用、4属性以上を同時に扱うための多変数可視化手法の適用、が重要であると考えられる。

またクレジットカード不正検出の運用上の課題として、ルールの作成、編集、削除に伴う副作用をも可視化できるようにシステムを拡張したい。例えばルールの作成や拡大によって、いままで[不検出・正当]に属していた決済の誤検出がどの程度増えるか、逆にルールの削除や縮小によって、いままで[検出・不正]に属していた決済の検出漏れがどの程度増えるか、といった点も可視化できるようにしたい。そのために重要な点として、3.1節に示した[検出・不正][検出・正当][不検出・不正][不検出・正当]の4種類の不正を別々に表示するのではなく、全て一画面に表示することが必要である。しかし現実には、[検出・不正]や[不検出・不正]の件数と[不検出・正当]の件数には1万倍程度の大きな差があり、そのまま両者を効果的に可視化することは現時点では難しい。この技術的課題を解決することが、クレジットカード不正検出のルール作成支援において重要であると考えられる。

謝辞

散布図上の長方形領域の自動抽出に関して貴重な意見をくださった、お茶の水女子大学瀬々潤准教授に感謝いたします。

参考文献

- [1] 伊藤, 山口, 小山田, 長方形の入れ子構造による階層型データ視覚化手法の計算時間および画面占有面積の改善, 可視化情報学会論文集, 6, 51-61, 2006.
- [2] Bederson B., Schneiderman B., Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies, ACM Transactions on Graphics, Vol. 21, No. 4, pp. 833-854, 2002.
- [3] A. Nagasaki, T. Itoh, M. Ise, K. Miyashita, A Correlation-based Hierarchical Data Visualization Technique and Its Application to Credit Card Fraud Data, 1st International Workshop on Super Visualization, 2008.
- [4] S. Brin, R. Rastogi, K. Shim, Mining Optimized Gain Rules for Numeric Attributes, the Fifth ACM SIGKDD (International Conference on Knowledge Discovery and Data Mining), 135-144, 1999.