

クレジットカード決済履歴テストデータの特徴発見のための可視化とGUIの一手法

迫田千華（指導教員：伊藤貴之）

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。本論文では、ルール作成支援を目標としたクレジットカード不正検出結果の可視化手法を提案し、その可視化の事例を報告する。

2 提案内容

2.1 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。

2.2 決済レコードの分類

本システムでは、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。ルールに該当し、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。ルールに該当したが、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。ルールに該当しなかったが、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。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。

[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。

2.3 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、「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、「店舗番号がX番」といったルールを設定可能であることを意味する。

2.4 散布図による可視化

2.4.1 追加機能

本システムではユーザに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、散布図でクレジットカード決済の分布を可視化する。その際、例えば 40 種類の属性の中から 3 属性を選択する場合、その組み合わせは、 $40^3 = 64,000$ 通りもあり、この膨大な組み合わせの中から、有用な可視化結果を得ることは非常に困難である。そこで本システムでは、属性間の相関性算出結果 [1] を算出した上で、独自の属性選択パネルを用いて属性の選択を行う。2.3 節で述べた属性の分類に加え、属性値の種類の数や、時系列、大小関係を考慮し、それぞれ X 軸、Y 軸、色へ割り当てる優先度が高い属性の選択を支援する GUI となっている。図 1 では、優先度が高い順にマゼンダ、ライトピンク、グレーの背景色となっている。優先度が高いものを選択すると、この属性選択パネルでの組み合わせは 80 通り程度で、より効率よく有用な可視化結果を得ることが可能である。

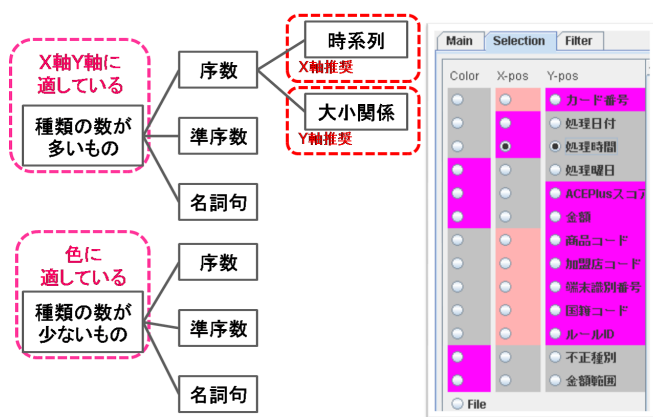


図 1: 属性の分類 (左) と属性選択パネル (右)

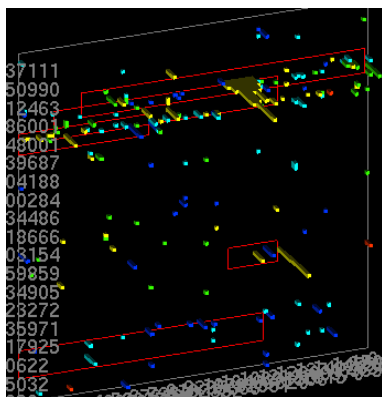


図 2: 長方形領域の提示例

また本システムでは、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。

さらに本システムでは [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、新規ルールの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 2 に示す。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 [2] 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。

2.4.2 適用事例

図 3 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。この集中部分に着目することで、ルール作成支援が可能となる。

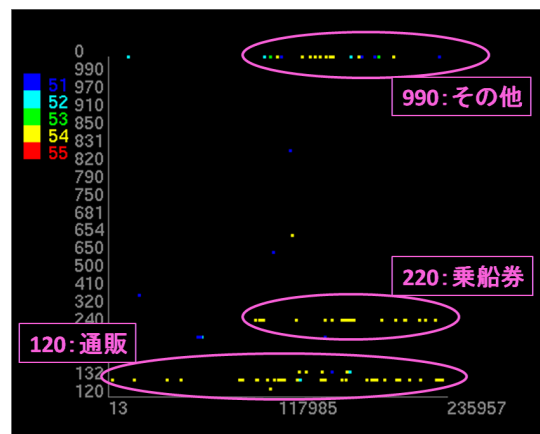


図 3: 適用事例

3 まとめ

本論文では、散布図を用いたクレジットカード決済履歴テストデータの可視化と、属性選択を支援するパネルや長方形自動抽出機能を備えた GUI の一手法を提案した。

謝辞

貴重なデータと数多くのご助言賜りました株式会社インテリジェントウェイブ様に感謝いたします。また、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] 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.
- [2] 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.