

クレジットカード不正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* 伊藤貴之* 伊勢昌幸** 宮下光輔**

*お茶の水女子大学理学部情報科学科

**株式会社インテリジェントウェイブ開発本部企画部

1. 概要

近年、クレジットカード等の不正を検出する様々な技術が開発されている。社団法人クレジット産業協会の調査によると、2002 年以降クレジットカードの不正使用の被害総額は減少傾向にある。しかし、それでもまだ十億単位での被害が発生しており、最近ではスキミングなど悪徳な技術が社会問題ともなっている。また、不正のパターンは組織性、期間、金額などの面で多様である。このように日々変化し多様化する不正パターンに対して、より高度な不正検出技術の研究が求められている。また、クレジットカード会社の多くが採用する不正検出システムには、会社ごとに固有のルールを設定することができるが、そのルール数は非常に多くなる傾向にある。その中で、時間経過によって不必要になったルールや、範囲変更が必要なルールの妥当性を、より効率的に管理できるシステムが望まれている。

本報告では、ルールの追加、更新、削除の手助けとなるグラフィカルユーザインタフェース(以下 GUI)を搭載し、システム管理者による不正パターンの視覚的な特徴発見を支援する可視化システムを提案する。可視化とは実体を持たない抽象的なデータを、コンピュータグラフィックスで表現することにより、ユーザの情報認識を手助けする技術である。本研究は、システム管理者の負担を減らし、効率的なルール管理に貢献できる可視化システムの構築を目指すものである。

2. 関連研究

本報告が対象とするクレジットカード不正履歴テストデータは、時系列情報を含む非常に多くの属性を有する。このような大規模な多変数データを可視化する手法は、従来から数多く存在する[1]。しかし、すべての属性値を同時に表示する手法は、視認性や画面解像度の点から、必ずしも使いやすいとは限らない。本報告で適用するような散布図をベースにした可視化手法においても、時系列情報の特徴に応じて適応的に属性を選択して可視化することが重要である。

また、著者らはクレジットカード不正履歴テストデータに対して、階層型データ可視化手法「平安京ビュー」[2]の適用

も試みている。平安京ビューは階層型データの葉ノードをアイコンで、階層構造を長方形の枠で表示する可視化手法であり、大規模データの可視化に適している。しかし平安京ビューは、時系列変化や時間的特徴の表現には不向きである。

時系列情報の特徴検出に関する研究は、可視化以外の分野でも多数報告されている。一例として文献[3]では、多次元時系列データに対して、まず各時系列のパターンクラスタリングを適用し、次に特徴的パターン間の相関ルールマイニングを適用し、得られたルール群を用いてシステムに生じた特徴を検知する手法を報告している。

3. 提案内容

3.1 データ構造

本報告で扱うクレジットカード不正履歴のテストデータは、列が各属性を表し、1 行が 1 件のイベントの単位であり、表の各欄に時系列情報を含む変数値が格納された表形式データである。クレジットカード不正履歴を構成する購買履歴は、不正検出システムのルールに該当するために不正の可能性が高いと判定された履歴に、後から不正でないことを確認できた履歴を削除し、さらに後から不正であると確認できた履歴を加えたものとする。本報告で扱うデータの一例を示したものが表 1 である。本報告で扱うデータでは、日時、金額、不正種別、会員番号など 18 の属性が格納されている。

表 1: クレジットカード不正履歴の一例

日付	利用金額	不正種別	国	会員番号	...
20080312	2000	52	424	213	...
20080312	150000	51	210	34322	...
20080312	15030	51	22	325	...
20080313	3429	53	105	328923	...
⋮	⋮	⋮	⋮	⋮	⋮

3.2 データの表現形式

提案手法では X 軸、Y 軸、色に対応する属性を指定し、散布図を用いてクレジットカード不正履歴テストデータを可視化する。ただし、大規模なデータを散布図で表示すると、点が重なり情報を失う可能性がある。そこで提案手法では、図 1 に示す 2 次元/3 次元表示の切り替え機能を用いる。3 次元表示には 2 次元的なヒストグラムを採用することで、ほぼ同一な属性値を有する不正の回数を視覚的に確認できる。例えば X 軸に時刻を指定した場合には、短時間で集中的に行われる不正を発見できると見込まれる。

Visualization and GUI for discovery of patterns from test log data of credit card frauds

Chika Sakoda* Takayuki Itoh*

Masayuki Ise** Kousuke Miyashita**

*Ochanomizu University **Intelligent Wave Inc.

{chika, itot}@itolab.is.ocha.ac.jp

3.3 GUI 概要

図2は現在開発中の本可視化システムのGUIである。2次元/3次元表示の切り替えの他に、軸、色に対応する属性の切り替えや属性値の絞り込み、ハイライトが可能である。更に、年月や時間帯、曜日毎など時系列幅を決定するGUIも備えることで、さまざまな角度からの検証が可能となる。

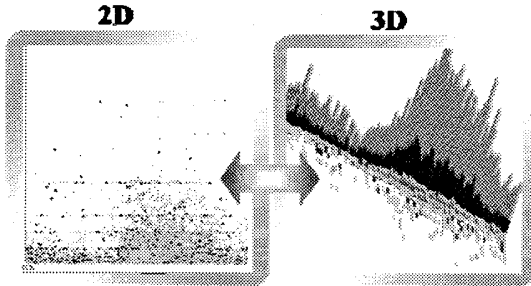


図1：2次元/3次元表示

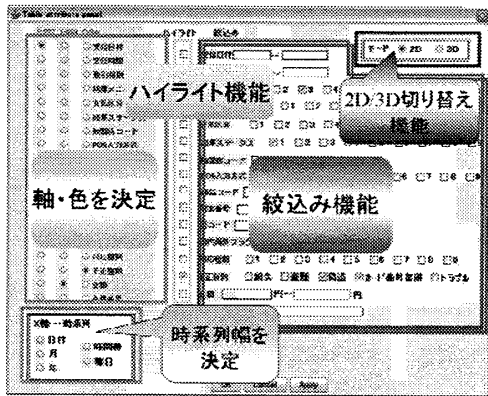


図2：クレジットカード不正履歴テストデータのためのGUI

4. 適用事例

4.1 金額の時系列変化

図3にX軸を日付、Y軸を金額、色を不正種別とした可視化結果例を示す。ただし、データは3ヶ月分で約21,000件分である。図3より、金額が10万、20万、30万、50万円の限度額のラインに不正使用が集中していることが視覚的に観察できる。これは、不正カード使用者が、不正発覚する前に一度で限度額までの使用を試みていると推測できる。

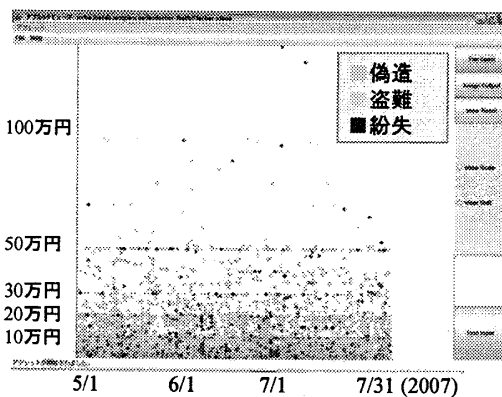


図3：可視化例1

4.2 時間帯における特徴

図4にX軸を時間帯、Y軸を商品コード、色を金額とした可視化結果例を示す。ただし、データは3ヶ月分で約21,000件分である。商品によって不正使用の時間帯が偏っているものがあることが観察できる。そして明らかに深夜である時間帯の不正も多く確認できる。カード加盟店の営業形態を考慮した上で、更に絞り込んで可視化していくことで悪意のある加盟店や犯罪組織を発見できる可能性もある。

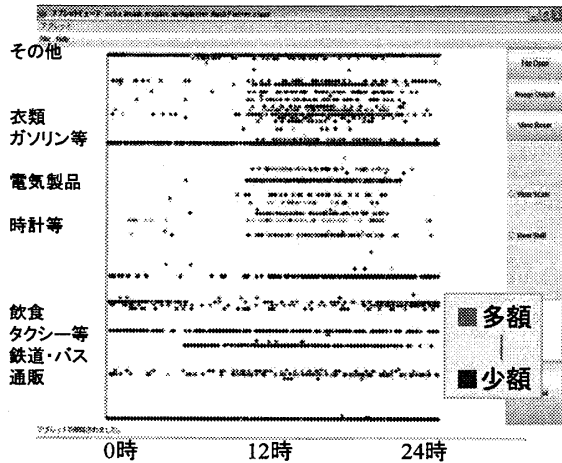


図4：可視化例2

5. まとめと今後の課題

本報告では、クレジットカード不正履歴を可視化する一手法を提案した。今後の課題として、まずGUIの改良と拡張を進めたい。そして、このGUIを用いて大規模な多変数時系列データを対話的に可視化し、その有用性を検証したい。また、文献[1]に提案されている属性間の相関性算出および可視化の半自動化手法を導入して、GUIと自動化手法を併用したシステムの構築を目指したい。また、3次元表示の性能向上について検討したい。更に、ルールの適用範囲を散布図上に同時に可視化し、不正履歴の分布と照合する機能を設けたい。

これらの機能を開発した上で、新たな不正パターンの発見、および有用なルールの追加、各ルールの妥当性の検証などに対して、本可視化システムがどのように貢献できるか、さらに検証を進めたい。

参考文献

- [1] Xie Z., Huang S., Ward M., Rundensteiner E., Exploratory Visualization of Multivariate Data with Variable Quality, IEEE Symposium on VAST 2006, pp.183-190, 2006.
- [2] 長崎, 伊藤, 伊勢, 宮下, 相関性を考慮した大規模階層型データの可視化 -クレジットカード不正履歴テストデータの可視化への応用-, 情報処理学会第70回全国大会, 2ZE-4, 2008.
- [3] 矢入, 加藤, 堀, 中須賀:時系列相関ルールマイニングに基づく人工衛星テレメトリデータからの異常検出法, The 15th Annual Conference of Japanese Society for Artificial Intelligence, 2001.