

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[○](お茶の水女子大学), 伊藤 貴之(お茶の水女子大学),
伊勢 昌幸((株)インテリジェントウェイブ),
宮下 光輔((株)インテリジェントウェイブ)

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

Chika SAKODA, Takayuki ITO, Masayuki ISE and Kousuke MIYASHITA

ABSTRACT

Information visualization is a research field which visualizes various data applying computer graphics techniques. We are developing a visualization system and GUI to interactively visualize large multi-dimensional time series data. We expect users can discover troubles and trends by watching the visualized large and complicated data on their screens. Specifically, we visualize the test log data of credit card frauds. The test log data of credit card frauds form tables, and contain some attributes, such as time, amount of money, type of frauds, and personal numbers. In this paper, we assign the attribute to the X-axis, the Y-axis, and a color. And, it visualizes with a scatter plot. We use GUI of an attribute change, and 2-dimensional / 3-dimensional change function. And, we report the discovery which time change of data and the correlation between each attribute.

Keywords: Visualization, GUI, Scatter plot, Credit card frauds

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とともに、具体的な適用事例をいくつか示す。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関連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

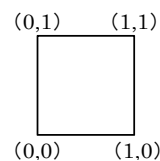
本報告で扱うクレジットカード不正使用履歴のテスト

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 Table1 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。

Table 1 Test log data of credit card frauds

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	...
:	:	:	:	:	:

仮定する。X 軸に i 番目の属性、Y 軸に j 番目の属性を割り当てる場合、正規化された 2 値から得られる座標値 $(V_{i,n}, V_{j,n})$ で、任意の値を正方形内部にプロットする。



また色の算出の際には、HSI 表色系の Hue(色相)を属性値から算出する。本報告では $V_{i,n} = 0$ のとき青、 $V_{i,n} = 1$ のとき赤になるように Hue を算出する。

また、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、Fig.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。具体的には、散布図の X 軸、Y 軸を一定間隔で分割し、その格子上の点の個数を 3 次元表示の高さとする。3 次元表示を採用することで、ほぼ同一な属性値を有する不正の件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。

3.3 GUI 概要

Fig.2 は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を行う GUI を備えており、さまざまな角度からの検証が可能である。

Fig. 1 2D and 3D representations.



Fig. 2 GUI for switching and narrowing attribute values.

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。具体的な座標の算出方法を次に述べる。

i 番目の属性における値の最大値を $V_{i,max}$ 、最小値を $V_{i,min}$ とする。このとき、任意の値 V_i を区間 $[0,1]$ に正規化した値 $V_{i,n}$ を求める式は以下ようになる。

$$V_{i,n} = \frac{(V_i - V_{i,min})}{(V_{i,max} - V_{i,min})} \quad (1)$$

ここで、4 点 $(0,0)$ $(0,1)$ $(1,0)$ $(1,1)$ を四隅とする正方形を

4. 適用事例

4.1 金額の時系列変化

Fig.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。Fig.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、不正が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。

4.2 時間帯における特徴

Fig.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することで、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。

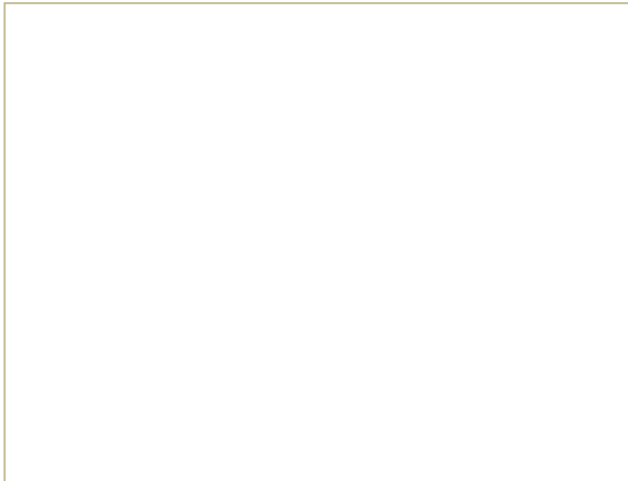


Fig. 3 Result (1).

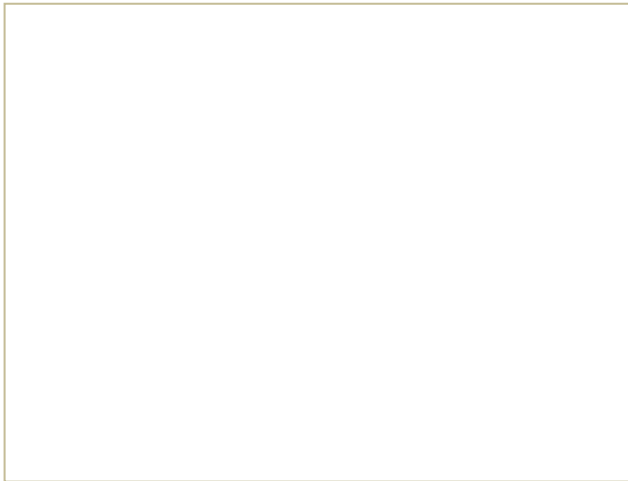


Fig. 4 Result(2).

4.3 特徴的事項の発見事例 1

Fig.5 に X 軸を時間帯、Y 軸を加盟店コード、色を商品コードとした可視化結果例を示す。ただし、データは 12 ヶ月分で約 113,000 件分である。一例として、加盟店コード“1234”はすべて同じ色であることが観察できる。つまり、加盟店コード“1234”での不正使用はすべて同じ商品であるといえる。次節 4.4 では、更にこの加盟店コード“1234”で絞込みを行った可視化結果について考察する。

4.4 特徴的事項の発見事例 2

Fig.6 に X 軸を時間帯、Y 軸を日付、色を会員番号とし、加盟店コード“1234”で絞り込みを適用した可視化結果例を示す。ただし、データは前節 4.3 と同一である。特定の会員番号の不正が集中していることが観察できる。一例として、会員番号“56789”の不正は、長時間数日に及ぶことが確認できる。このように、複数回の不正をもつ会員番号や加盟店コードを可視化し検証することで、個人の不正の手口や加盟店における不正の特徴、または、加盟店自身の不正の手口等の発見が見込まれる。今後、不正の回数が多い、または被害額の大きい会員番号や加盟店コードを自動でリストアップし、機能の拡張を行う。

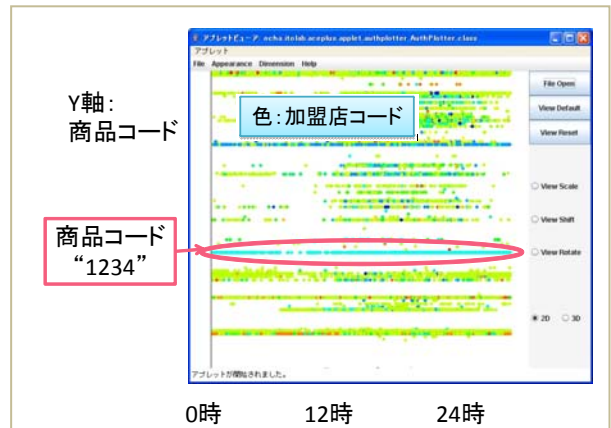


Fig. 5 Result(3).

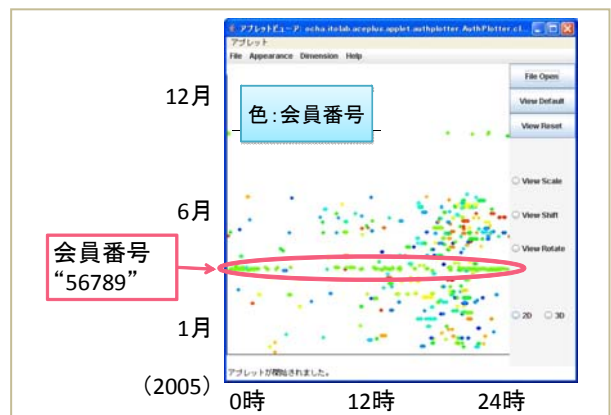


Fig. 6 Result(4).

4.5 高額の不正の特徴

Fig.7 に X 軸を時間帯、Y 軸を加盟店コード、色を不正種別とし、金額が 500,000 円以上で絞り込んで表示した可視化結果例を示す。ただし、データは前節と同一である。高額の不正のうち、不正種別が偽造又は盗難であるものが、それぞれ特定の加盟店コード群に集中していることが観察できる。この要因を追求解明することで、不正パターンや、犯罪組織の発見が見込まれる。

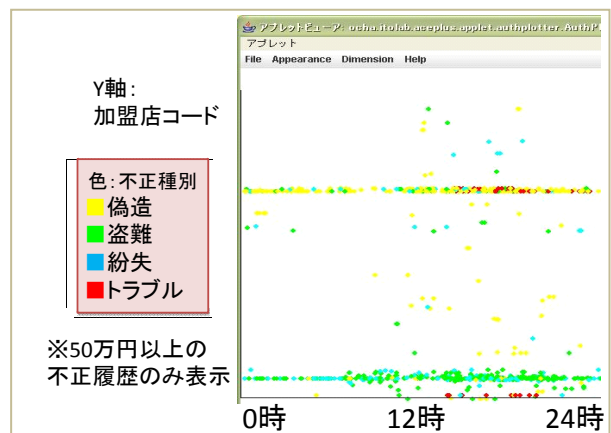


Fig. 7 Result(5).

4.5 ルールの可視化 1

続いて、実験用として実ルールデータに擬似した架空のルールデータを作成し、それらのルールに該当するものに“1”，該当しないものに“0”，の 2 値を有する属性を

新たに付け加えて、可視化実験を行った。

Fig.8にX軸を日付、Y軸を金額、色をルール該当/非該当とした可視化結果例を示す。さらに、Fig.9にルール非該当のものだけを絞り込んだ可視化結果例を示す。ただし、データは前節と同一である。Fig.8, Fig.9より、金額が10万、20万、30万、50万円の限度額のラインにルール非該当の不正使用が集中していることが視覚的に観察できる。この場合、限度額のラインに新たなルールを追加することで、効率よくルールの追加を行うことが可能である。

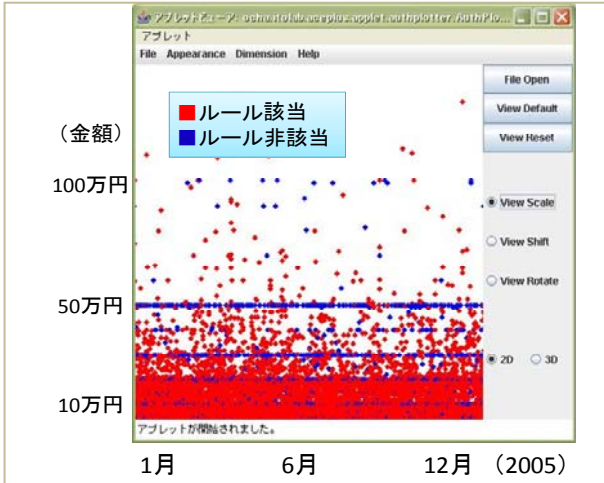


Fig. 8 Result(6).

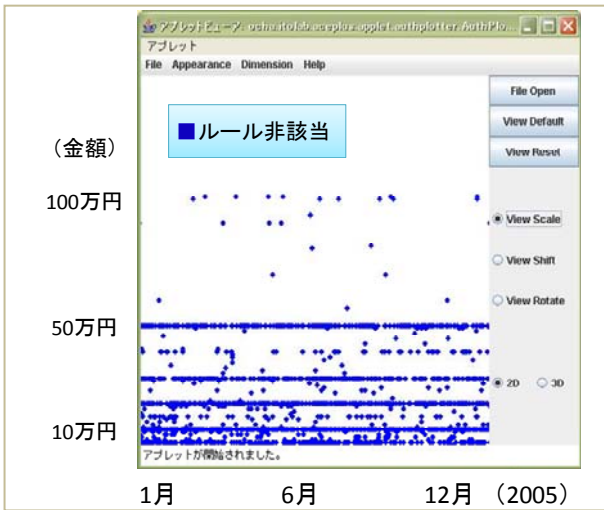


Fig. 9 Result(7).

4.6 ルールの可視化 2

本節ではX軸を時間帯、Y軸を加盟店コード、色を金額帯とした可視化結果を考察する。実験用の架空ルールデータを用い、Fig.10にルール該当のものだけを絞り込んで表示した可視化結果例、Fig.11にルール非該当のものだけを絞り込んで表示した可視化結果例を示す。ただし不正履歴データは12ヶ月分で約113,000件分である。Fig.10, Fig.11より、特定の加盟店コード群がルールに該当/非該当であることが確認できる。この場合、特定の加盟店コード群に対してルールの改善を行う必要があるといえる。このように、ルールを改善する必要がある範

囲の視覚的判断が可能である。

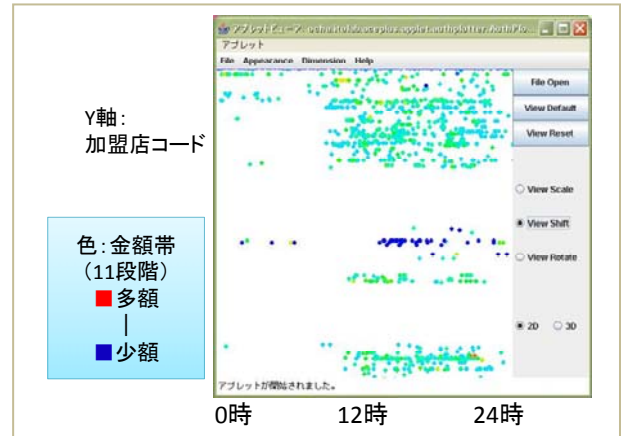


Fig. 10 Result(8).

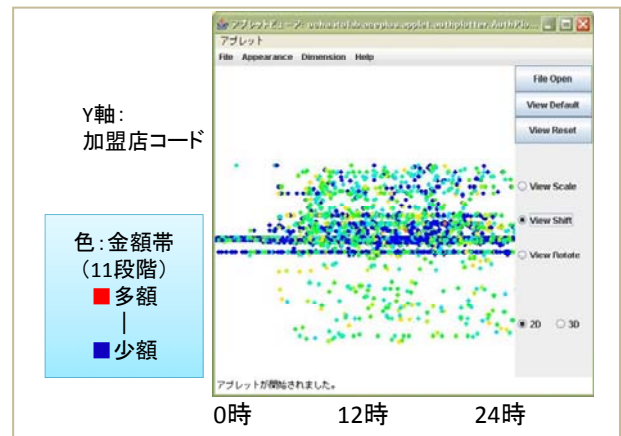


Fig. 11 Result(9).

5. まとめと今後の課題

本稿では、クレジットカード不正履歴を可視化した適用事例を報告した。今後の課題として、文献[2]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次元表示の有用性について評価を進めたい。その上で、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-, 第36回可視化情報シンポジウム, 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.