

# 時刻付きデータの探索的分析を支援する 視覚的分析ツールの開発

武田 修平<sup>1,a)</sup> 蘇 紹華<sup>1,b)</sup> 濱田 康平<sup>2,c)</sup> 三末 和男<sup>3,d)</sup> 田中 二郎<sup>3,e)</sup>

概要：膨大なデータの活用のために、仮説を定めずに特徴的なパターンを探すような探索的分析が行われることが多い。時刻付きデータに関してもそのような分析作業ができれば、複数の商品の購買履歴から他の商品とは違う周期性を持つ商品や突発的に販売が伸びた商品など、予期しないパターンをもつ商品を探し出すことができそうである。さらには、その要因を探り新たな販売戦略につながる可能性もある。ただし、時刻付きデータの可視化手法の多くは、ある種の商品の販売のような一つの事象集合に対して、1年周期のような定められたパターンを効果的に見せるよう設計されている。そのため、周期を指定せずに周期性の強い商品を抽出するとか、何らかの周期性はあるが他とは違う傾向のある商品を抽出するといった作業には向いていない。このような問題に対して、我々は時刻付きデータの分析に有益だと考えられる特徴の洗い出しを行うとともに、定式化を行なった。さらに、それらの特徴を視覚的に観察しながら、時刻付きデータの分析が行えるツールを開発した。

キーワード：探索的分析, 情報可視化, 時刻付きデータ, 視覚的分析ツール, 複数ビュー.

## 1. はじめに

今日、様々な機関が膨大なデータを持っており、データはいくつも活用例がある。例えば犯罪発生データであれば、その日に犯罪が発生しそうな場所や時間帯の予測<sup>\*1</sup>、購買履歴データであれば、潜在的なニーズを捉えることで新規サービスの施策 [1] が可能となる。ただし、このようにデータを活用するためには、データを適切に分析する必要がある。

膨大なデータに対して探索的な分析が行われることが多い。探索的な分析とは、特定のモデルを前提としてデータを解釈するものではなく、仮説を定めずにデータの特徴を

探す分析である。また、データの特徴を表現する方法として可視化があげられる。犯罪発生データのような時刻付きデータを適切に可視化することにより、様々なパターンの把握が可能となる。もし、時刻付きデータに対して視覚的に探索的な分析を行えば、他とは違う周期性を持って発生する犯罪や、販売時刻の特徴が似ている商品の発見が可能になると考える。

しかし、既存の時刻付きデータの可視化手法の多くは特定の周期であったり、発生の順序などの定められた特徴を効果的に見せるように設計されている。したがって、特定の周期を決めつけずに周期性を持つ犯罪を探したり、似ている特徴を持つ商品を探したりといった作業が困難であると言える。

そこで、本研究では時刻付きデータの探索的な分析の支援を目的とする。そのために、分析の際に注目したいであろう発生の特徴の洗い出しを行うとともに、定式化を行った。さらに、特徴を視覚的に観察しながら、インタラクティブに分析が行えるツールを開発した

## 2. 関連研究

本研究では時刻付きデータの探索的な分析のために、時刻付きデータを様々な側面から観察する必要がある。そこ

<sup>1</sup> 筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻

Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba

<sup>2</sup> 筑波大学情報学群情報科学類

School of Informatics, University of Tsukuba

<sup>3</sup> 筑波大学システム情報系

Faculty of Engineering, Information and Systems, University of Tsukuba

a) takedas@iplab.cs.tsukuba.ac.jp

b) sushaohua@iplab.cs.tsukuba.ac.jp

c) hamada@iplab.cs.tsukuba.ac.jp

d) misue@cs.tsukuba.ac.jp

e) tanaka@cs.tsukuba.ac.jp

\*1 <http://www.nytimes.com/2011/08/16/us/16police.html>

で、関連研究として時刻付きデータの特徴の把握を支援している研究と、データの様々な側面への注目を支援している研究を紹介する。

## 2.1 時刻付きデータの特徴の把握

気温の変化や売上の変化のような時系列データを対象にした一般的な手法として、折れ線グラフや棒グラフ、折れ線グラフを塗りつぶした面グラフが挙げられる。これらは系列が少ないデータを対象としている古典的な手法であるため、現在はいくつか拡張された手法が存在する。SaitoらのTwo-Tone Pseudo Coloring [2] は値の大きさに色を割り当てることで、多くの系列を持つデータを表現できるようにした。JavedらのBraided graph [3] は面グラフの重なりを適切に表現することで、系列間の比較を容易にできるようにした。これらの手法によってより多くの系列を同時に見ることができ、お昼のみに多く発生している系列の探索などといった、指定した特徴を持つ系列を抽出するような作業を行うためには、1つ1つの系列に注目して判断しなければならない。ShiroiらのChronoView [4] は事象の集合の発生時刻の分布を2次元平面上で表すことで、数千以上の事象の集合のおおよその発生時刻を俯瞰することができる。しかし、あらかじめ周期を指定する必要があるため、周期が定まっていないような探索を行うことができない。

時刻付きデータは特定の分野で注目されているような様々な特徴が存在する。ChangらのWireVis [5] では、金融機関などの取引の不正を発見するために、共起性や発生頻度の変化などを表現している。MonroeらのEventFlow [6] は薬の投与の期間や回数、その後の患者の様子を把握するために、事象の発生の順番、発生の期間などを表現している。これらの手法はそれぞれの目的のためにあらかじめ定められた特徴を効果的に表現するように設計されている。したがって、複雑な特徴を表現することが可能だが、分析者が状況に応じて注目する特徴を選択するような、探索的な分析を行うことができない。

## 2.2 データの様々な側面に注目

ChangらのWireVisやMonroeらのEventFlowは同じデータを持つ複数の特徴を提示するために、いくつかのビューが用意されている。このビューはより多くの情報の俯瞰を提供することができるだけでなく、リンクを組み合わせてビュー間のデータの対応関係も把握しやすくする。同様に複数のビューを連携させることで、様々な側面に注目している手法として、ChevalierらのHistomages [7] が挙げられる。Histomagesは画像の加工を目的としており、画像のプレビューや色彩のヒストグラムなどの専用のビューを提供している。加工を行う人は自由にビューの追加と削除を行うことができるため、1つ1

つの側面に注目しながら画像の加工が可能である。本研究で開発したツールは時刻付きデータの特徴を複数のビューから観察することができ、ビューは自由に追加と削除が可能のため柔軟な探索を支援する。

## 3. 対象データ

対象データは時刻付きデータであり、1つの事象に発生時刻と複数の属性が付与されていることを想定している。犯罪発生データにおける例を以下に示す。

表 1 対象データの例

犯罪の発生時刻	属性 1(種類)	属性 2(場所)
07/02 06:00:00	CRIMINAL DAMAGE	APARTMENT
07/20 14:00:00	BATTERY	STREET
08/07 13:45:00	BATTERY	APARTMENT
08/14 19:00:00	BURGLARY	CAR WASH
09/20 23:30:00	OTHER OFFENSE	RESIDENCE

表1の「BATTERY」や「RESIDENCE」のような属性値は、発生した事象の内容を表す。このような同じ属性を持つ事象を以下のように定式化する。

事象を、その発生時刻とひとつ以上の属性値の組で表す。すなわち、事象  $e$  は次のように表される。

$$e = (t, a_1, \dots, a_k) \quad (1)$$

ここで、 $k \geq 0$  とする。

事象  $e$  の発生時刻を  $t(e)$  で表す。すなわち、 $e = (t, a_1, \dots, a_k)$  のとき、 $t(e) = t$  である。事象の集合  $E$  と時刻の集合  $T$  が与えられたとき、時刻集合  $T$  内で発生した事象からなる部分集合を式 (2) のように、 $E|_T$  で表すことにする。

$$E|_T = \{e \in E | t(e) \in T\} \quad (2)$$

さらに、時刻の集合である時間帯を導入する。時間帯は時刻  $t_0$  を起点として、時間  $d$  の幅をもつ時間帯の列を、 $R_0, R_1, \dots$  とし、それらの全体を  $U$  で表す。

$$U = \{R_0, R_1, \dots\} \quad (3)$$

このとき、 $R_i \in U (i \in \mathbb{N})$  は、式 (4) に示されるように、左閉半開区間とする ( $\mathbb{N}$  は自然数全体の集合。自然数には0を含むとする。)

$$R_i = [t_0 + id, t_0 + (i + 1)d) \quad (4)$$

事象の多くは周期性を備える。そのため時間帯に周期性を導入する。ここでは、 $c = md$  ( $m$  は2以上の整数) を周期とするために、式 (5) に示される同値関係を定義する。

$$R_i \simeq R_j \Leftrightarrow i = j \pmod{m} \quad (5)$$

周期性を考慮した時間帯を、 $U$  の  $\simeq$  による同値類として、式 (6) のように表す。

$$U_c = U / \simeq = \{[R_0], [R_1], \dots, [R_{m-1}]\} \quad (6)$$

#### 4. 分析で用いる特徴

探索的な分析であっても何かしらの指標を用いることになる。私たちは時刻付きデータの分析のための指標として発生の特徴を用いることにした。探索的な分析において有用と考えられる発生の特徴を洗い出すために、まず ChronoView を用いて時刻付きデータを分析し、分析の中で注目した特徴の列挙と注目したい特徴の考察を行った。そして、関連研究で紹介した既存の可視化手法 ([2, 4, 5]) で表現している特徴を調査し、表 2 のような特徴を洗い出した。

表 2 洗い出した特徴

注目したい特徴	特徴を表す指標
発生の頻度が高い	発生頻度
$x$ 時に最も多く発生する	最頻時間帯
$x$ 時周辺に発生しやすい	発生しやすい時刻
1 つの時刻に発生している	発生時刻のばらつき
$x$ 時に特に多く発生している	特に多く発生している時間帯
$x$ 時間のときに周期性が強い	周期性の強さ
発生しやすい時刻が近い	発生しやすい時刻の近さ
発生する時刻が似ている	発生時間帯の類似性
同時に発生している	同時に発生する回数

また、表の中の指標に加え、「頻度分布」を用いる。これは特定の特徴を表すものではないが、分析の中で必要だと考え追加した。

##### 4.1 発生頻度

時刻付きデータを分析する際に、より多く発生している犯罪やあまり売れていない商品の抽出など、事象の集合毎の発生の頻度に注目することが多い。このような探索を行うために発生頻度という指標を定めた。

事象の集合  $E$  が与えられたとき、事象の発生頻度を、式 (7) のように  $f(E)$  で表す。

$$f(E) = |E| \quad (7)$$

##### 4.2 頻度分布

事象の集合を何かしらの特徴で絞り込んだ後は、1 つ 1 つの事象の集合がどのような発生の仕方をしているのかを詳細に知りたいと考える場合がある。この頻度分布は時間帯毎の発生頻度を表しているため、発生頻度の推移を詳細に知ることができる。

事象の集合  $E$  が与えられたとき、発生頻度の分布を、式 (8) のように  $m$  次元ベクトル  $f_{U_c}(E)$  で表す。

$$f_{U_c}(E) = (f(E|_{[R_0]}), f(E|_{[R_1]}), \dots, f(E|_{[R_{m-1}]}) \quad (8)$$

頻度分布を見ることで多くの特徴を知ることができるが、特徴の比較や、指定した特徴を持つ事象の集合の抽出

といった作業が困難であると言える。また、同時に多くの事象の集合を表示すると、1 つ 1 つの事象の集合の特徴を知ることができなくなってしまう。

##### 4.3 最頻時間帯

指定した種類の商品を何かしらの時間帯に PR しようとした場合、その商品が確実に多く売れている時間帯を探す必要がある。このような探索のために、事象の集合がどの時間帯に最も発生しているのかを表す指標として、最頻時間帯を定めた。

事象の集合  $E$  が与えられたとき、最も頻繁に発生した時間帯を、式 (9) のように集合  $T_{mode}(E)$  で表す。

$$T_{mode}(E) = \{R \in U | \forall Q \in U, f(E|_R) \geq f(E|_Q)\} \quad (9)$$

なお、時間帯の周期性を考慮する場合には、式 (10) のように表す。

$$T_{mode}(E) = \{[R] \in U_c | \forall [Q] \in U_c, f(E|_{[R]}) \geq f(E|_{[Q]})\} \quad (10)$$

##### 4.4 発生しやすい時刻

時刻付きデータを分析しようとしたときに、それぞれの事象の集合が発生している時刻を把握することは重要だと言える。このような時刻を把握するために、ChronoView [4] の計算方法を元にして、事象の集合が発生しやすい時刻を指標として定めた。

事象の集合  $E$  が与えられたとき、発生しやすい時刻を、式 (11) のように  $t_{mean}$  で表す。

$$t_{mean}(E) = t_0 + \frac{c}{2\pi} \tan^{-1} \frac{\sum_{e \in E} \cos \frac{2\pi}{c} (t(e) - t_0)}{\sum_{e \in E} \sin \frac{2\pi}{c} (t(e) - t_0)} \quad (11)$$

ただし  $\tan^{-1}$  の引数が  $\frac{0}{0}$  のときには、発生しやすい時刻は無いものとする。

##### 4.5 発生時刻のばらつき

犯罪の発生時刻の予測をしようとした場合、発生時刻の偏りが強そうな犯罪の抽出が必要である。このような抽出を支援するために、事象集合が 1 つの時刻に集中して発生しているかを表す発生時刻のばらつきを指標として定めた。

事象の集合  $E$  が与えられたとき、周期性を考慮した発生時刻のばらつきを、式 (12) のように  $t_{dev}(E)$  で表す。

$$t_{dev}(E) = \sqrt{\frac{1}{|E|} \sum_{e \in E} (\sin^{-1} |\sin \frac{2\pi}{c} (t(e) - t_{mean}(E))|)^2} \quad (12)$$

##### 4.6 特に多く発生している時間帯

事象の集合が発生しやすい時間帯は複数存在する場合があります。犯罪の警備を行う際には複数の時間帯を把握することが重要である。これらの時間帯の把握を支援するために、事象の集合が特に多く発生している複数の時間帯を表す指標を定めた。

事象の集合  $E$  が与えられたとき、事象が多く発生している時間帯を、式 (13) のように集合  $T_{freq}(E)$  で表す。

$$T_{freq}(E) = \{R \in U_c | f(E|_R) > \bar{f} + a\sigma\} \quad (13)$$

ただし、 $\bar{f}$  と  $\sigma$  はそれぞれ、 $f_{U_c}$  の成分の平均と標準偏差とする。定数  $a$  は発生の「多さ」の程度を決める値である。

#### 4.7 周期性の強さ

時刻付きデータが持つ特徴の探索において、周期性は重要な要素だと言えるが、周期性が強い周期を把握していない場合がある。このような場合に使用する指標として、複数の周期毎に事象の集合の周期性の強さを定めた。

事象の集合  $E$  が与えられたとき、事象の発生時刻の周期性の強さを、式 (14) のように  $l$  次元ベクトル  $s_{cyc}(E)$  ( $l$  は 1 以上の整数) で表す。

$$s_{cyc}(E) = (s_1(E), s_2(E), \dots, s_l(E)) \quad (14)$$

ここで  $s_i$  は周期  $ic_s$  ( $i, c_s$  は 1 以上の整数) の際のスペクトルであり、 $f_{re}$  と  $f_{im}$  を用いて下のように表される。

$$s_i(E) = f_{re}(E, ic_s)^2 + f_{im}(E, ic_s)^2 \quad (15)$$

$$f_{re}(E, c) = \sum_{j=0}^{m-1} \{f(E|_{[R_j]}) \cdot \cos(j \frac{2\pi}{c})\} \quad (16)$$

$$f_{im}(E, c) = \sum_{j=0}^{m-1} \{f(E|_{[R_j]}) \cdot \sin(j \frac{2\pi}{c})\} \quad (17)$$

#### 4.8 発生しやすい時刻の近さ

犯罪を発生時刻でグルーピングしようとしたときに、発生しやすい時刻が近い犯罪をまとめることは警備などにおいて重要である。また発生しやすい時刻が特に遠い犯罪は独特の発生の仕方をしている可能性がある。このような事象の集合を探索できるようにするために、事象の集合の組み合わせ毎に発生しやすい時刻の近さを定めた。

事象の集合  $E_1$  および  $E_2$  が与えられたとき、ふたつの事象集合の発生時刻の近さを、式 (18) のように  $t_{diff}(E_1, E_2)$  で表す。

$$t_{diff}(E_1, E_2) = \begin{cases} |t_d(E_1, E_2)| & \text{if } |t_d(E_1, E_2)| \leq \frac{c}{2} \\ c - |t_d(E_1, E_2)| & \text{otherwise} \end{cases} \quad (18)$$

ここで  $t_d(E_1, E_2)$  は下のように表される。

$$t_d(E_1, E_2) = t_{mean}(E_1) - t_{mean}(E_2) \quad (19)$$

#### 4.9 発生時間帯の類似性

変則的な売れ方をする商品が見つかったとき、同じような売れ方の商品を集めることで共通点を発見できる可能性がある。このように発生の仕方を定めることなく、発生の仕方が似ている事象の集合の探索を行うために、発生時間帯の類似性という指標を定めた。

事象の集合  $E_1$  および  $E_2$  が与えられたとき、ふたつの事象集合の発生時間帯の類似度を、式 (20) のように  $sim$  で表す。

$$sim(E_1, E_2) = \frac{f_{U_c}(E_1) \cdot f_{U_c}(E_2)}{|f_{U_c}(E_1)| \cdot |f_{U_c}(E_2)|} \quad (20)$$

#### 4.10 同時に発生する回数

特定の場所で発生している犯罪の種類を知りたい場合など、事象の持つ異なる属性値の関係の把握が必要となることが多い。このような関係の把握を支援するために、2つの属性値の組み合わせ毎に事象の集合が何回発生したかを表す指標を定めた。

事象の集合  $E$  および二つの属性値  $a_i$  と  $a_j$  が与えられたとき (ここで、 $k \geq 2$ 、 $i < j$  とする) ふたつの属性値の両方を同時に備える事象の発生回数を、式 (21) のように  $n_{i,j}(E, a_i, a_j)$  で表す。

$$n_{i,j}(E, a_i, a_j) = |\{(t, x_1, \dots, x_i, \dots, x_j, \dots, x_k) \in E | x_i = a_i \wedge x_j = a_j\}| \quad (21)$$

### 5. 分析ツールの開発

時刻付きデータに対して探索的な分析を行うために、私たちは図 1 のような視覚的な分析ツールを開発した。このツールはメインの描画エリアと設定エリアに分かれており (図 1 A,B)、Shneiderman のマントラ [8] に従い、以下の手順で分析を行うことを想定している。

- (1) Overview : データに含まれる事象の集合が全体的にどのような時刻に発生しているかを把握するために、集合全ての発生時刻の分布を俯瞰する。俯瞰は ChronoView [4] で行う。
- (2) Filter : ツールの描画エリアに表示されている事象の集合から注目したい集合を選択する (図 1 C)。そして、設定エリアから見たい特徴を選択し (図 1 D)、新しいビューを追加する。
- (3) Detail-on-Demand : 事象の集合の選択によるリンクや、マウスホバーによる詳細表示 (図 1 E) を用いて探索を行う。さらに探索を続ける場合は (2)Filter に戻り、ここまでの分析の流れを繰り返す。

上記の手順での分析を支援するために実装した主な機能について説明する。

#### 5.1 専用のビューの追加

専用のビューは事象の集合の特徴を見るために存在し、分析者は任意のタイミングで描画エリア内に新しいビューを追加することができる。専用のビューを追加できることは、探索的な分析を 2つの点で支援している。

1つ目は、手順 (2) において分析の視点の柔軟さを向上させている点である。専用のビューで表示している特徴は

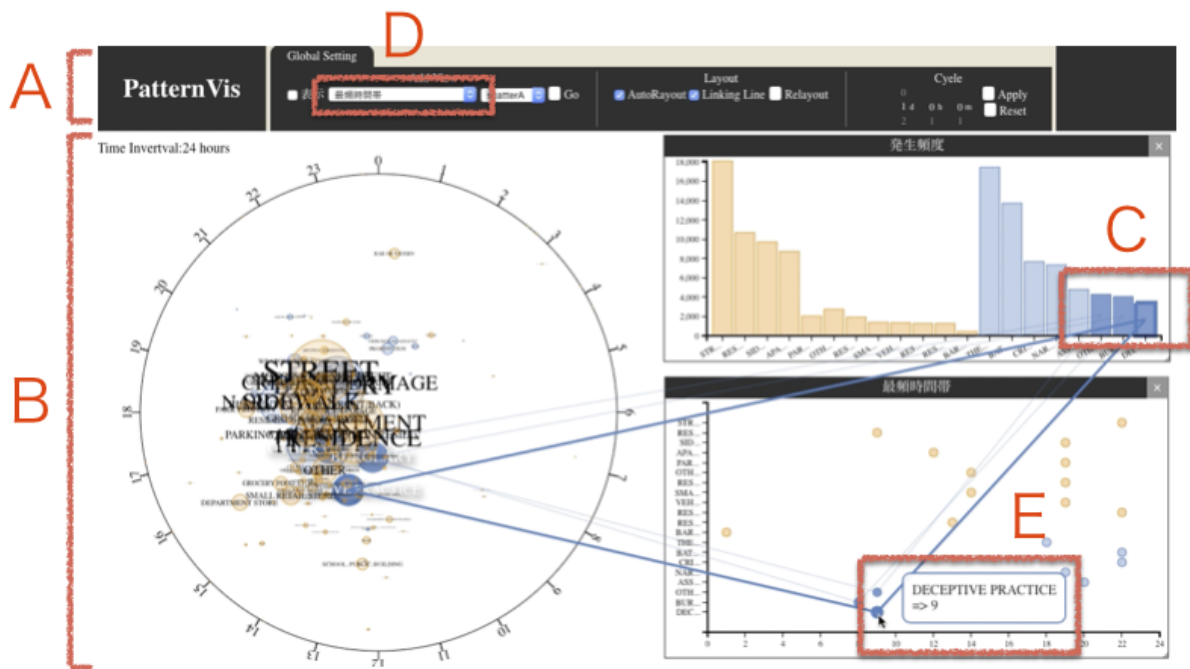


図 1 開発したツールの外観 (A) 設定エリア (B) 描画エリア (C) 事象の集合の選択 (D) 注目したい特徴の選択 (E) リンキングとマウスホバーの詳細表示

分析者が自由に決めることができる。また、ビューの数に制限は無く、それぞれのビューに異なる特徴を表示することができるため、分析者は様々な視点からデータの特徴を探索することが可能である。図 2 では同じ事象の集合を、異なる 3 つの特徴を選択したビューで表示している。マウスホバーや選択することにより同じ事象の集合がハイライトされるため、気になった事象の集合の持つ特徴を複数の視点から同時に知ることができる。

合わせて設計された視覚表現が表示されている。図 2 を見ると、上の発生頻度を表示しているビューでは、回数を比較しやすいように棒グラフを用いた。右下の同時に発生する回数を表示しているビューでは、回数が多い組み合わせを見つけやすいように、色の濃さで回数を表現する行列表現を用いた。これにより、ビューの数が増えた場合でも、分析者は直感的に特徴を把握することができる。

## 5.2 事象の選択

開発したツールでは、分析者が様々な視点から事象の集合を選択できるように、各専用のビューに対して選択が行えるようにした。事象の選択の方法は投げ縄選択を採用しており、ビュー内にある棒グラフのバーや、散布図の点といった要素を投げ縄で囲むことで事象の集合は選択されたことになる。事象の集合の選択は全てのビューで連動しハイライトされるため、分析者はビュー間での集合の対応関係を確認しながら、探索を進めることができる。

上記のような選択方法を採用することにより、グラフ上で近い位置にプロットされている事象の集合をまとめて選択したり、特定の値を持つものを選択したりといった、直感的な集合の選択が可能となる。また、専用のビュー内では事象の属性値の種類や、特徴を表す指標を元にソートが可能である。ソートを行うことで、発生頻度が非常に高い犯罪のみを抽出したり、販売時刻のばらつきが比較的小さい商品を選択するような、特徴がより強く出ている事象の集合の探索が容易となる。

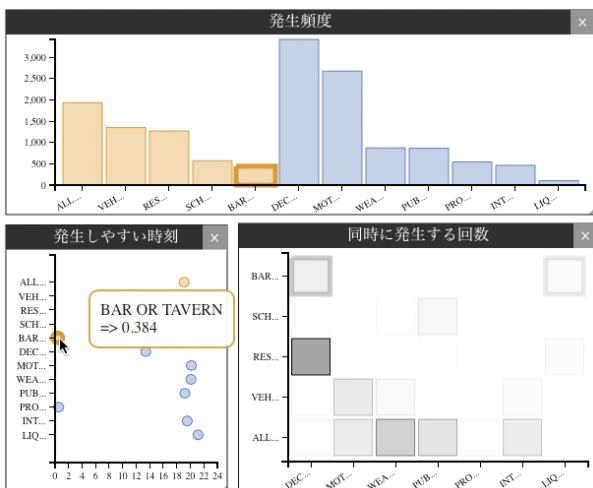


図 2 「BAR OR TAVERN」にマウスホバーした例 (上) 発生頻度 (左下) 発生しやすい時刻 (右下) 同時に発生する回数

2 つ目は、手順 (3) において 1 つ 1 つの特徴の把握を容易にしている点である。開発したツールでは探索を行うための特徴を 10 個提供しているが、ビュー内には各特徴に

## 6. ユースケース

ユースケースとして開発したツールを用いて、犯罪データの分析を行った例を示す。対象データはシカゴが公開しているオープンデータ<sup>\*2</sup>で、分析には以下のように抜き出したデータを用いた。

表 3 抜き出した犯罪データ

レコード数	74709
期間	2014/07/01 - 2014/09/30
レコード	犯罪の種類, 犯罪の場所, 発生時刻

分析は2つの状況を想定して行っている。

### 6.1 何らかの周期性を持つ犯罪の種類の探索

1つ目の分析では犯罪がどのような周期性を持つかわからないという状況を想定する。まず、周期性がありそうな周期を発見するために、全ての犯罪の種類の周期性の強さを表示する。3は周期性の強さを表すビューであり、犯罪の種類ごとの周期性の強さを、横軸を周期とした面グラフで表現している。また、犯罪の種類は発生頻度が高い順番に並んでおり、上に行くほど発生頻度が高い犯罪となっている。グラフの中の大きな値を示している場所に注目していく。

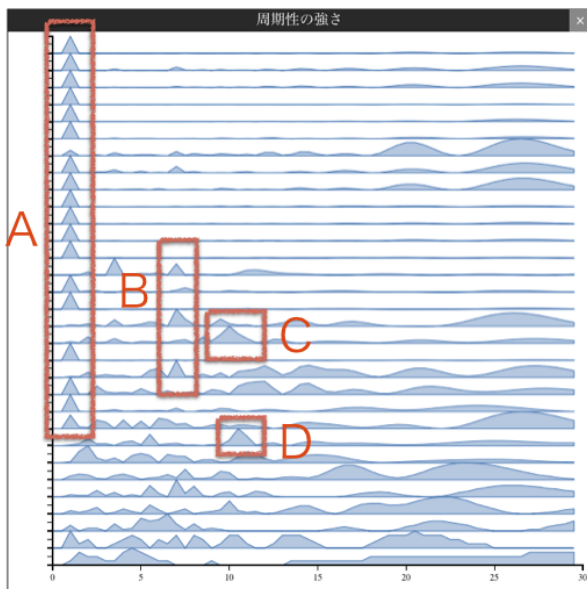


図 3 全ての犯罪の種類の周期性の強さ

図中の A を見てみると、多くの犯罪の種類が全く同じ周期のときに高い周期性を示している。また、B の場所も A ほど多くはないが複数の犯罪の種類が高い周期性を示している。これらの周期を見てみると、A は周期が 1 日、B は周期が 7 日のときであるため、犯罪は 1 日と 1 週間の周

<sup>\*2</sup> <https://data.cityofchicago.org>

期の場合に高い周期性を持つものが多いということがわかる。次に C と D に注目する。この 2 つの犯罪の種類は 1 日と 7 日のときはあまり高い周期性を示していないが、2 つとも 10 日あたりの周期で高い周期性を示している。そこでこの 2 つの犯罪の種類の特徴をさらに探索する。

2 つの犯罪の種類の周期性の強さを詳細に知るため、2 つのみを選択してもう一度周期性の強さを表示する。

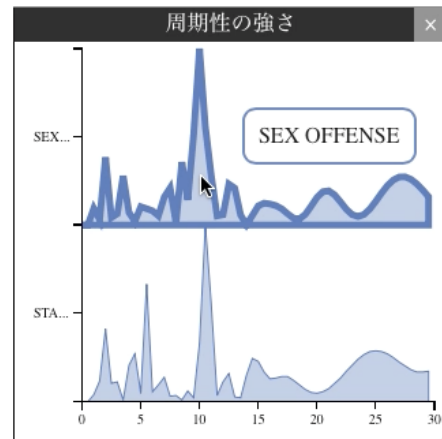


図 4 「SEX OFFENSE」と「STALKING」の周期性の強さ

表示した面グラフ(図4)を見てみると、SEX OFFENSE(性犯罪)は10日の周期で、STALKING(ストーキング)は11日の周期で高い周期性を示していることがわかる。実際にこれらの周期で何かしらの特徴があるのかを確かめるために、発生の頻度分布を複数の周期で表示してみる。

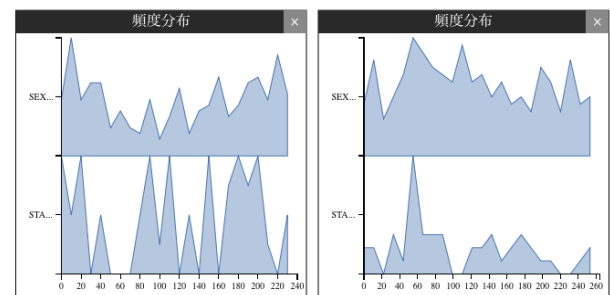


図 5 「SEX OFFENSE」と「STALKING」の2つの周期における頻度分布(左)10日周期(右)11日周期

10日で周期を設定したビュー(図5左)を見てみると、上のSEX OFFENSEは最初と最後の日に頻度が高くなっているが、下のSTALKINGはどの日でも同じぐらいの頻度である。11日で周期を設定したビュー(図5右)を見てみると、上のSEX OFFENSEは前半に少し頻度が高いように見えるが、下のSTALKINGは2日目あたりで顕著に頻度が高くなっていると言える。これらの結果から、SEX OFFENSEやSTALKINGは他とは違う周期性を持っている可能性があることがわかった。

## 6.2 通学時の犯罪防止ポスターの作成

2つ目の分析では学校向けの犯罪防止ポスターを作成するといった状況を想定する。まずは、通学時に発生しそうな犯罪を調べるために、全ての犯罪の種類の特によく発生している時間帯を表示する。

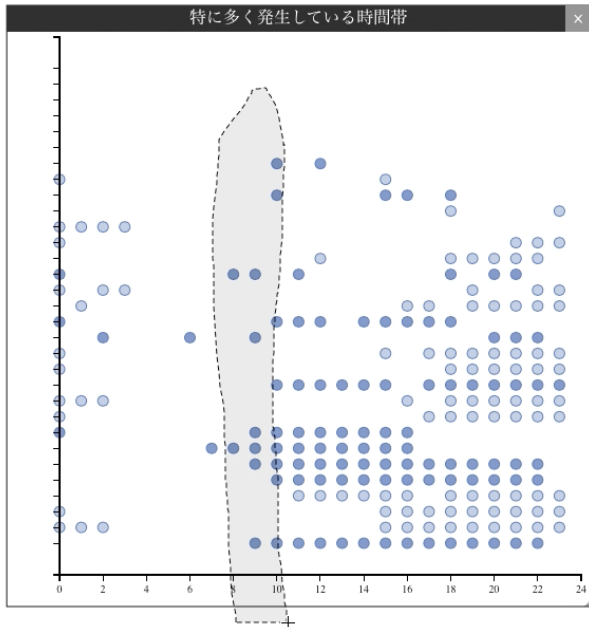


図 6 特によく発生している時間帯を元に犯罪の種類を選択

このビューの中から8-10時と15-18時に多く発生している犯罪の種類を選択する。そして、通学路で発生している犯罪に絞るために、SIDEWALKを選択し同時発生回数を表示する。

図7を見てみると、特に大きな値を示す犯罪の種類が5つありマウスホバーにより確認すると、BATTERY(脅迫)、NARCOTICS(麻薬)、ROBBERY(強盗)、THEFT(窃盗)、ASSAULT(暴行)が多いことが分かる。しかし、5つの犯罪の種類を選択し、同時発生回数のビューのハイライトを見てみると、通学時以外にも多く発生している犯罪であることがわかる。そこで、犯罪の種類の中から通学時間のみの特によく発生しているものを探すために、13-14時と0-6時にも多く発生している犯罪を除外する。するとGAMBLING(ギャンブル)、INTIMIDATION(脅迫)、KIDNAPPING(誘拐)、SEX OFFENSEなどが残り、これらが通学時に特によく発生している犯罪の種類だと言える。念のため、SIDEWALKでも発生しているかを確認するために、もう一度、同時発生回数を表示する。

そうすると、どの犯罪もSIDEWALKで発生していることがわかる。これらの分析結果より、通学路で多く発生しているという視点では5種類、通学時間に多く発生しているという視点では4種類の犯罪についてポスターを作れば効果的であるということがわかる。

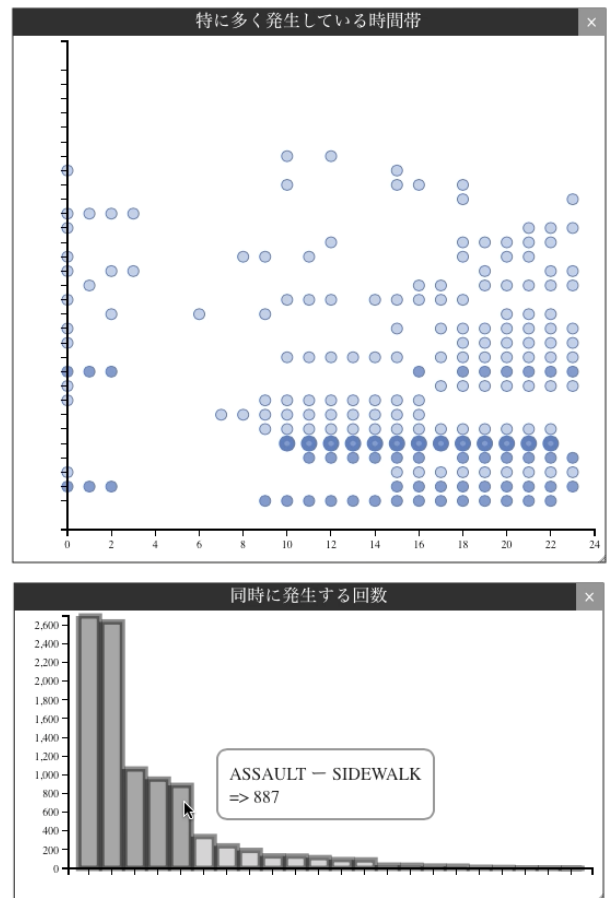


図 7 通学時に多く発生している種類とSIDEWALKの同時発生回数

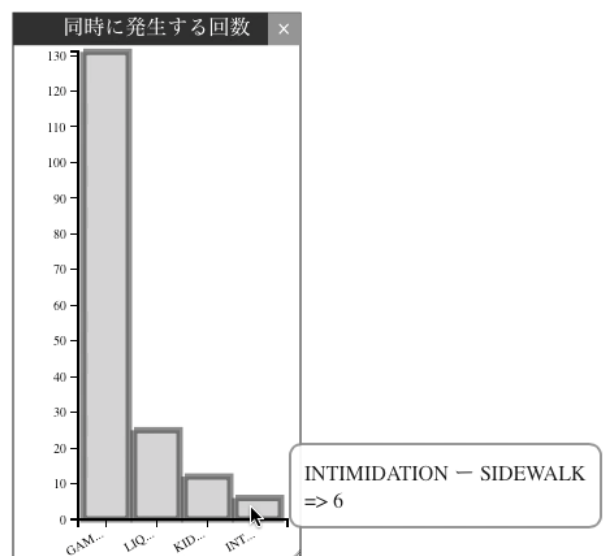


図 8 通学時のみに集中している種類とSIDEWALKの同時発生回数

## 7. まとめ

本研究では、時刻付きデータに対する探索的な分析を支援するために、分析に有益だと考えられる発生の特徴の洗い出しと定式化を行った。そして、特徴を視覚的に観察しながら分析を行うために、特徴毎の専用のビュー作成と柔軟に事象の選択が可能なるツールを開発した。開発したツールを用いたユースケースでは、分析者が状況に応じて分析を進めることができ、分析の目的に合った特徴の探索が可能であることを示した。

今後の課題としては、網羅的な発生の特徴の洗い出しと、視覚表現の検討が挙げられる。現在、注目することができる特徴は周期を考慮したものが多いため、様々な角度から探索可能な特徴が必要である。これにより、今までの表現やツールでは知りえなかった傾向の発見など、より探索的な分析が可能だと考えられる。

## 参考文献

- [1] T. Ishigaki, T. Takenaka, and Y. Motomura, "Customer Behavior Prediction System by Large Scale Data Fusion in a Retail Service", In *Transactions of the Japanese Society for Artificial Intelligence*, Vol.26, No.6, pp.670-681, 2011.
- [2] W. Javed, Student Member, IEEE, B. McDonnel, Student Member, IEEE, and N. Elmqvist, Member, IEEE, "Two-tone pseudo coloring: compact visualization for one-dimensional data", In *IEEE Transactions on Information Visualization*, pp.173-180, 2005.
- [3] T. Saito, H. N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya, and T. Kaseda, "Graphical Perception of Multiple Time Series", In *IEEE Transactions on Visualization and Computer Graphics*, Vol.16, No.6, pp.927-934, 2010.
- [4] S. Shiroy, K. Misue, and J. Tanaka, "ChronoView: Visualization Technique for Many Temporal Data", In *16th International Conference Information Visualization*, pp.112-117, 2012.
- [5] R. Chang, M. Ghoniem, R. Kosara, W. Ribarsky, and J. Yang, "WireVis: Visualization of Categorical, Time-Varying Data From Financial Transactions", In *IEEE Transactions on Visual Analytics Science and Technology*, pp.155-162, 2007.
- [6] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman, "Temporal Event Sequence Simplification", In *IEEE Transactions on Visualization and Computer Graphics*, Vol.19, No.12, pp.2227-2236, 2013.
- [7] F. Chevalier, P. Dragicevic, and C. Hurter, "Histomages: Fully Synchronized Views for Image Editing", In *Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST '12)*, pp.281-286, 2012.
- [8] B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations", In *Proceedings of the IEEE Symposium on Visual Languages*, pp. 336-343, 1996.