

筑波大学大学院博士課程

システム情報工学研究科修士論文

時刻付きデータの探索的分析を支援する
視覚的ツールの開発

武田 修平

修士（工学）

（コンピュータサイエンス専攻）

指導教員 三末 和男

2016年3月

概要

時刻付きデータの探索的分析を支援するために視点切り替え時の作業の削減と、分析の過程や状況の把握の支援を目的とした研究を行った。本研究では目的を達成するために、時刻付きデータの特徴の選択による視点の切り替えが可能な視覚的分析ツールを開発した。視点の切り替えは分析に有益だと考えられる特徴の洗い出しと定式化、視覚表現の割り当てをあらかじめ行うことで機能として提供される。開発した分析ツールは視点の変更などの操作の履歴や、データの参照関係を俯瞰するための視覚表現を備えている。さらに、インタラクティブな操作によりデータを直感的にフィルタリング可能である。開発した分析ツールの有用性を示すためにケーススタディを実施した。ケーススタディでは犯罪の発生データの分析を行うことで、分析する範囲の広さだけでなく質も向上し、分析の過程や現状の把握ができることを示した。

目次

第1章	序論	1
1.1	時刻付きデータの活用	1
1.2	探索的分析の課題	1
1.3	目的とアプローチ	2
第2章	関連研究	3
2.1	時刻付きデータの特徴の把握	3
2.2	探索的な分析の支援	4
第3章	探索的分析の支援の要求事項	5
3.1	観察する特徴の切り替えのための要求事項	5
3.2	分析の過程を把握するための要求事項	6
3.3	分析の状況を把握するための要求事項	6
第4章	時刻付きデータの分析で用いる特徴	8
4.1	発生頻度	9
4.2	発生頻度分布	10
4.3	最頻時間帯	10
4.4	発生しやすい時刻	10
4.5	発生日時のばらつき	11
4.6	特に多く発生している時間帯	11
4.7	周期性の強さ	11
4.8	発生しやすい時刻の近さ	12
4.9	発生時間帯の類似性	12
4.10	同時に発生した回数	12
第5章	分析ツールの開発	13
5.1	分析ツールの設計	15
5.2	Feature View Area	16
5.3	Small View Area	19
5.4	Overview Area	20
5.5	インタラクション	21

第6章 ケーススタディ	23
6.1 犯罪の種類と犯罪が発生した場所の種類から傾向を得る	23
6.2 分析結果のまとめと考察	29
第7章 議論	30
第8章 結論	32
参考文献	34

目次

5.1	開発した分析ツール	14
5.2	ビューの自動レイアウト	16
5.3	グラフのソート	17
5.4	選択した履歴の表示	18
5.5	Small View Area と Feature View Area の連携	19
5.6	Overview Area の表示	20
5.7	エリアごとの選択の様子	21
5.8	グラフの要素のブラッシング	22
6.1	発生頻度を表示	24
6.2	「最頻時間帯」と「発生しやすい時刻」、「特に多く発生している時間帯」を表示	25
6.3	「発生頻度」と「発生しやすい時刻」の連携の様子	25
6.4	頻度が高い事象の集合の「最頻時間帯」と「発生しやすい時刻」、「特に多く発生している時間帯」を表示	26
6.5	朝が最も頻度の高い事象の集合の頻度分布を表示	27
6.6	朝が最も頻度の高い事象と同時に発生した回数を表示	28

表目次

4.1 対象データの例	8
6.1 犯罪の種類と犯罪が発生した場所の種類	23

第1章 序論

1.1 時刻付きデータの活用

今日、様々な機関がシステムのログや取引の記録などの時刻付きデータを持っている。時刻付きデータとはレコードの中にタイムスタンプを含んでいるデータである。例えば購買履歴データの場合、購買時刻などがタイムスタンプとして用いられているものが時刻付きデータであり、さらに商品名や店名といった情報が付随することが多い。

世の中の多くの事象は発生時刻に特徴があり、水曜日に特に多く購入される商品や、同じ時間帯に発生しやすい犯罪のように事象ごとに固有の発生のパターンを持っている。この発生パターンを把握することにより、商品が購入されるであろう時間帯や発生に関係がありそうな犯罪といった予測が可能となる。実際の例として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¹、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 [2] が可能となる。これらのようなデータの活用は時刻付きデータを分析することにより行われる。

データの分析手法の1つに探索的分析というものが存在する。探索的分析とは、あらかじめデータを観察する内容や順番を定めことなく、分析の中で観察する側面を選んで分析を進める手法である。探索的分析は実際にデータを観察してから次に観察する内容を決定するため、データに合わせた特徴をより深く広く観察していくことが可能である。もし、時刻付きデータの探索的分析を支援することができれば、様々な視点から発生のパターンを探ることができると考えられる。

1.2 探索的分析の課題

探索的分析を行う場合、分析者はデータの特徴の観察と観察する特徴の切り替えを繰り返す必要がある。観察する特徴の切り替えとは、次に観察したい特徴が定まった際に、その特徴を観察するために行うデータのフィルタリングや可視化といった一連の作業である。これらの作業はデータの形式や分析者の意図に大きく左右されるため、状況に柔軟に対応可能な分析ツールが必要である。既存のツールの例としては Microsoft の表計算ソフトウェアである Excel や、Ross らの統計分析フリーソフトである R が挙げられる。これらのツールを探索的分析に用いた場合、分析者は特徴の切り替えのためにデータのフィルタリングや計算方法、可視化

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

表現の検討などが必要となる。しかし、特徴の切り替えのたびにこれらの作業をしては、次の特徴を観察するまでに毎回多くの手間が発生してしまう。

特徴の観察と切り替えを繰り返していると、分析者は分析の過程や状況がわからなくなってしまうことがある。分析の過程とはどのような特徴を観察してきたのかや、どのような特徴をもとにフィルタリングを行ったかといった、過去に分析者が行った作業の流れを表す。分析の状況とはデータのどのぐらいのレコードを参照しているのかや、どの特徴を持つレコードに注目したのかといった、分析者がデータ全体に対して注目した部分を表す。過程と状況は分析の内容を振り返る場合に必要だが、既存のツールではこれらの振り返りを支援していない。

整理すると、以下が探索的分析を支援する際の課題だと言える。

- 異なる特徴に注目しようとした際に必要な手順を減らす
- 現在注目している時間的特徴にたどり着いた過程が分かるようにする
- データ全体に対してどの部分を参照しているのかが分かるようにする

1.3 目的とアプローチ

本研究では、時刻付きデータの探索的分析を支援するために視点切り替え時の手順の削減及び、分析の過程と状況の把握を支援することを目的とする。目的を達成するためのアプローチとして、観察したい特徴を選択するだけで観察する視点を切り替えることができ、過程と状況を視覚的に提示するツールを開発する。そのために、時刻付きデータの分析で有益だと考えられる特徴の洗い出しと定式化、視覚表現の割り当てを行った。そして、注目した特徴の順番やフィルタリングの履歴と、データの全体に対する参照状況を俯瞰するための視覚表現を設計した。

本論文では、分析で用いる時刻付きデータの特徴と機能を備えたインタラクティブな視覚的分析ツールを開発について述べ、犯罪発生データを用いたケーススタディによりその有用性を示す。

第2章 関連研究

時刻付きデータの探索的な分析のためには、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、探索的な分析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。これらの古典的な手法は多くのデータ項目を一度に表現することが困難であるため、拡張した手法がいくつか開発されている。SaitoらのTwo-Tone Pseudo Coloring[3]は値の大きさに2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[4]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、発生が集中している時刻や、発生の頻度が変化する様子などの一部の特徴を把握するために特化した表現だと言える。ShiroyiらのChronoView[5]は事象の集合の発生日時の分布を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日時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、特徴がありそうな周期の長さがわかっているデータに対して有用だと言える。

時刻付きデータはデータの種類に応じて、様々な特徴を持っている。ChangらのWireVis[6]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。Wangらの電子カルテの分析ツール[7]では、薬の投与といったイベントを元にソートを行い、発生の期間と回数を表示している。VrotsonらのActiviTree[8]では、大規模なイベントベースのデータを分析するために、事象の発生の順番のまとまりを表示している。MonroeらのEventFlow[9]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、探索的な分析には向いていない。

2.2 探索的な分析の支援

探索的な分析を支援するために、複数の側面を同時に観察できるようにしている研究は多く存在している。North の Snap-Together[10] や前述した WireVis、EventFlow は同じデータを持つ複数の異なる特徴を提示するために、いくつかのビューを用意している。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、リンクングを組み合わせることでビュー間のデータの対応関係も把握しやすくする。

Eick ら [11] や Boukhelifa ら [12] は複数のビューを用いているだけでなく、分析者がビューを自由に編集できるようにした。彼らが設計したモデルではビューで表示するデータだけではなく、視覚表現も分析者が変更可能であり、分析者は自由な視点でデータを見ていくことができる。Chevalier らの Histomages[13] 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。

これらのツールはデータの持つ複数の側面に注目することは可能であるが、時刻付きデータ特有の側面に注目したものではない。

第3章 探索的分析の支援の要求事項

本論文では、時刻付きデータを分析する際の要求事項を以下の3点に大別する

1. 観察する特徴の切り替えのための要求事項
2. 分析の過程を把握するための要求事項
3. 分析の状況を把握するための要求事項

3.1 観察する特徴の切り替えのための要求事項

観察する特徴の切り替えとは、次に観察したい特徴が定まった際に、その特徴を観察するために行うデータのフィルタリングや可視化といった一連の作業である。具体的には以下のような手順を踏む必要がある。

1. データのフィルタリング。分析するデータの中から、注目したい犯罪の種類や時間帯などに一致するレコードのみに絞る。
2. 計算方法の検討。発生の頻度や周期性の強さなど、分析者が気になった特徴を知るためにどのような計算をすべきかを考える。
3. 計算式の記述。Excelのセルの中などに、検討した計算方法を実際を書いていく。
4. 可視化表現の検討。計算結果から特徴を知るために、棒グラフや面グラフといった視覚表現をどのように用い、視覚変数をどの値に割り当てるかを考える。
5. 可視化。使用するツールの中から、検討した可視化表現のための機能や関数を使用してグラフを表示する。

もし、これらの作業を1人でこなそうとした場合、分析者は考えていることを切り替える必要があったり、他の人に作業を任せただけの場合でも、作業の依頼などに時間がかかってしまうと言える。そうすると、前の特徴で観察していた内容を忘れてしまったり、切り替えのタイミングで分析が終了してしまう可能性が考えられる。

以上より、特徴を切り替える際の手順はできるだけ少ない方が望ましい。

3.2 分析の過程を把握するための要求事項

データから何かしらの傾向を得た際に、どのような分析を行ったかかを振り返る必要がある。例えば、購入される時間帯に差がある複数の集団を発見した場合、どのような差があるかだけではなく、それぞれの集団がどのような集団かを知らなくてはならない。そのため、探索的分析ではその集団を見つけるために、発生頻度や周期性の強さなどのどのような特徴を観察して、頻度の高さや周期の長さなどのどのような部分を選択したかを覚えておくことになる。しかし、次々に観察する特徴を切り替える探索的分析では次々に新しいことを考えるため、分析の内容を覚えておくことは困難である。

振り返るべき内容はいくつか存在する。1つ目は、どのような特徴をどのような順番で観察したかである。分析者は特徴を観察して考えたことを元にして次に観察する特徴を決定するといったように、探索的分析の観察は連続的である。特徴の決定は1つの特徴を観察することによって行われるときもあるが、複数の特徴を元に決定する場合も存在する。同様に1つの特徴を観察した結果、他の複数の特徴を観察しようとする場合も存在する。分析の振り返りを行うとき、このような観察の分岐の様子を把握できるべきである。

2つ目はどのような特徴のどの部分を選択したかである。例えば発生の頻度といった特徴を観察していた場合、分析者は発生の頻度の高いものや低いものなど、一部の部分を選択して次に観察する特徴を選択することが多い。選択した部分を振り返ることによって、観察してきた特徴のどの部分に注目した結果なのかということを知ることができるため、選択の様子も把握できるべきである。

3.3 分析の状況を把握するための要求事項

探索的分析は次々と観察する視点を変えていく。そうすると、現在データのどの部分を参照しているのかや、どの程度網羅的に観察したのかといった分析の状況が把握できなくなってしまう。分析の状況が把握できないと、非常に少ないレコードのみを参照していたり、同じ特徴を繰り返し注目してしまう可能性があると言える。

データの参照状況を知るとは以下のような利点がある。

- 観察しているレコードが全体に対してどれぐらいの割合かを知ることによって、観察で得た傾向がどの程度全体に影響があるかがわかる。例えば、犯罪の発生のデータを観察している場合、発生の時間帯がはっきりしている犯罪を見つけたとしても、あまりに発生回数が少ない犯罪の場合その時間帯の防犯はあまり意味がないと言える。
- 観察しているレコードが全体に対してどの時刻に発生しているかを知ることによって、観察で得た傾向がどの程度信頼して良いかがわかる。例えば、商品の購買履歴のデータを観察している場合、何かしら周期性などを発見したとしても、ある1日だけにレコードが集中していた場合その傾向は信頼できないと言える。

このように、レコードの割合やレコードの分布は把握できるべきである。

何も傾向がわかっていないデータを分析する場合、ある程度網羅的に特徴を観察していくべきである。そのためには、今までの観察ごとにどの部分を注目したのかといった情報を把握できるべきである。

第4章 時刻付きデータの分析で用いる特徴

探索的分析では、データの持つ複数の特徴を観察する。本論文では時刻付きデータの特徴を時間的特徴と呼ぶ。時間的特徴の説明のために、まず対象データの定式化を行う。

対象データは時刻付きデータであり、1つの事象に発生日時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¹における例を表4.1に示す。

表 4.1: 対象データの例

犯罪の発生日時	属性 1(種類)	属性 2(場所)
07/20 14:00	BATTERY	STREET
08/07 13:45	BATTERY	APARTMENT
08/14 19:00	BURGLARY	CAR WASH
09/20 23:30	OFFENSE	RESIDENCE

表4.1の「BATTERY」や「RESIDENCE」のような属性値は、発生した事象の種類を表す。事象をその発生日時と1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。

$$e = (t, a_1, \dots, a_k) \quad (4.1)$$

ここで、 $k \geq 1$ とする。

事象 e の発生日時を $t(e)$ 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(4.2)のように、 $E|_T$ で表すことにする。

$$E|_T = \{e \in E | t(e) \in T\} \quad (4.2)$$

さらに、時刻の集合である時間帯を導入する。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。

$$U = \{R_0, R_1, \dots\} \quad (4.3)$$

このとき、 $R_i \in U$ ($i \in \mathbb{N}$) は、式(4.4)に示されるように、左閉半开区間とする (\mathbb{N} は自然数全体の集合。自然数には0を含むとする。)

$$R_i = [t_0 + id, t_0 + (i + 1)d) \quad (4.4)$$

¹<https://data.cityofchicago.org>

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 (4.5) に示される同値関係を定義する。

$$R_i \simeq R_j \Leftrightarrow i = j \pmod{m} \quad (4.5)$$

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 (4.6) のように表す。

$$U_c = U / \simeq = \{[R_0], [R_1], \dots, [R_{m-1}]\} \quad (4.6)$$

探索的な分析において有用と考えられる時間的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、第 2 章で紹介した既存の可視化手法 ([3, 5, 6, 7, 8, 9]) で表現している特徴を調査し、特徴を表す指標を下のように洗い出した。

- 発生頻度
- 最頻時間帯
- 発生しやすい時刻
- 発生日時のばらつき
- 特に多く発生している時間帯
- 周期性の強さ
- 発生しやすい時刻の近さ
- 発生時間帯の類似性
- 同時に発生した回数

また、これらの指標に加え、「発生頻度分布」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 (4.7) のように $f(E)$ で表す。

$$f(E) = |E| \quad (4.7)$$

4.2 発生頻度分布

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 (4.8) のように m 次元ベクトル $f_{U_c}(E)$ で表す。

$$f_{U_c}(E) = (f(E|_{[R_0]}), f(E|_{[R_1]}), \dots, f(E|_{[R_{m-1}]})) \quad (4.8)$$

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。このような特定の特征に注目する探索を多くの事象の集合に対して行う場合は、他の特徴を用いる。

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 (4.9) のように集合 $T_{mode}(E)$ で表す。

$$T_{mode}(E) = \{R \in U | \forall Q \in U, f(E|_R) \geq f(E|_Q)\} \quad (4.9)$$

なお、時間帯の周期性を考慮する場合には、式 (4.10) のように表す。

$$T_{mode}(E) = \{[R] \in U_c | \forall [Q] \in U, f(E|_{[R]}) \geq f(E|_{[Q]})\} \quad (4.10)$$

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[5] 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 (4.11) のように t_{mean} で表す。

$$t_{mean}(E) = t_0 + \frac{c}{2\pi} \tan^{-1} \frac{\sum_{e \in E} \cos \frac{2\pi}{c} (t(e) - t_0)}{\sum_{e \in E} \sin \frac{2\pi}{c} (t(e) - t_0)} \quad (4.11)$$

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 発生日時のばらつき

発生日時の偏りが大きい犯罪の発生日時の予測は容易であると考えられる。このような発生日時の偏りに特徴がある事象の集合を抽出するために、事象集合が1つの時刻に集中して発生しているかどうかを表す発生日時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日時のばらつきを、式 (4.12) のように $t_{dev}(E)$ で表す。

$$t_{dev}(E) = \sqrt{\frac{1}{|E|} \sum_{e \in E} (\sin^{-1} |\sin \frac{2\pi}{c} (t(e) - t_{mean}(E))|)^2} \quad (4.12)$$

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 (4.13) のように集合 $T_{freq}(E)$ で表す。

$$T_{freq}(E) = \{[R] \in U_c | f(E|_{[R]}) > \bar{f} + a\sigma\} \quad (4.13)$$

ただし、 \bar{f} と σ 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、事象の発生日時の周期性の強さを、式 (4.14) 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。

$$s_{cyc}(E) = (s_1(E), s_2(E), \dots, s_l(E)) \quad (4.14)$$

ここで s_i は周期 $ic_s(i, c_s)$ (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} を用いて下のよう表される。

$$s_i(E) = f_{re}(E, ic_s)^2 + f_{im}(E, ic_s)^2 \quad (4.15)$$

$$f_{re}(E, c) = \sum_{j=0}^{m-1} \{f(E|_{[R_j]}) \cdot \cos(j \frac{2\pi}{c})\} \quad (4.16)$$

$$f_{im}(E, c) = \sum_{j=0}^{m-1} \{f(E|_{[R_j]}) \cdot \sin(j \frac{2\pi}{c})\} \quad (4.17)$$

4.8 発生しやすい時刻の近さ

犯罪を発生日時でグルーピングしようとしたときに、発生しやすい時刻に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日時の近さを、式 (4.18) のように $t_{diff}(E_1, E_2)$ で表す。

$$t_{diff}(E_1, E_2) = \begin{cases} |t_d(E_1, E_2)| & \text{if } |t_d(E_1, E_2)| \leq \frac{c}{2} \\ c - |t_d(E_1, E_2)| & \text{otherwise} \end{cases} \quad (4.18)$$

ここで $t_d(E_1, E_2)$ は下のように表される。

$$t_d(E_1, E_2) = t_{mean}(E_1) - t_{mean}(E_2) \quad (4.19)$$

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 (4.20) のように $sim(E_1, E_2)$ で表す。

$$sim(E_1, E_2) = \frac{f_{U_c}(E_1) \cdot f_{U_c}(E_2)}{|f_{U_c}(E_1)| \cdot |f_{U_c}(E_2)|} \quad (4.20)$$

4.10 同時に発生した回数

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 (4.21) のように $n_{i,j}(E, a_i, a_j)$ で表す。

$$n_{i,j}(E, a_i, a_j) = |\{(t, x_1, \dots, x_i, \dots, x_j, \dots, x_k) \in E | x_i = a_i \wedge x_j = a_j\}| \quad (4.21)$$

第5章 分析ツールの開発

洗い出しと定式化を行った時間的特徴を用いて、探索的分析を支援するツールを開発した。本ツールの開発言語には JavaScript (ECMAScript Edition 5) を使用している。また、グラフの表示や選択のために複数のライブラリをしており、それぞれ以下のような役割を持つ。

- D3.js : データに基づいたドキュメントの操作のための JavaScript のライブラリ¹。SVG を用いた視覚化のためのコンポーネントが多く用意されておりアニメーションも可能である。ツール内ではグラフの描画やリンクングのために使用している。用いたバージョンは 3.4.6 である。
- jQuery : HTML の操作とイベントの処理を簡易にするための JavaScript のライブラリ²。高速ながら機能が豊富であり、ブラウザ間の処理の差異の吸収も行う。ツール内では新たなビューの生成やクリックやマウスオーバーのイベントを追加するために使用している。用いたバージョンは 1.11.2 である。
- jQuery UI : ユーザーインターフェースに特化した jQuery の派生の JavaScript のライブラリ³。jQuery を読み込む必要があるが、より複雑なインタラクションやエフェクトを使用することができる。ツール内では領域の選択やドラッグの操作のために使用している。用いたバージョンは 1.11.4 である。
- Material Design Lite : フラットデザインをより簡単に用いることが可能なデザインフレームワーク⁴。ボタンやスライダーなどのパーツだけではなく、ページ全体のレイアウトもテンプレートして提供している。ツール内の UI のパーツやアイコンはこのフレームワークによるものである。用いたバージョンは 1.0.4 である。

読み込む時刻付きデータは csv 形式のものを対象としている。実行環境としては、Safari (バージョン 9.0.2) を利用した。開発した分析ツールのスクリーンショットを図 5.1 に示す。各エリアについては 5.2 節から 5.4 節で説明する。

¹<http://d3js.org>

²<https://jquery.com>

³<https://jqueryui.com>

⁴<http://www.getmdl.io>

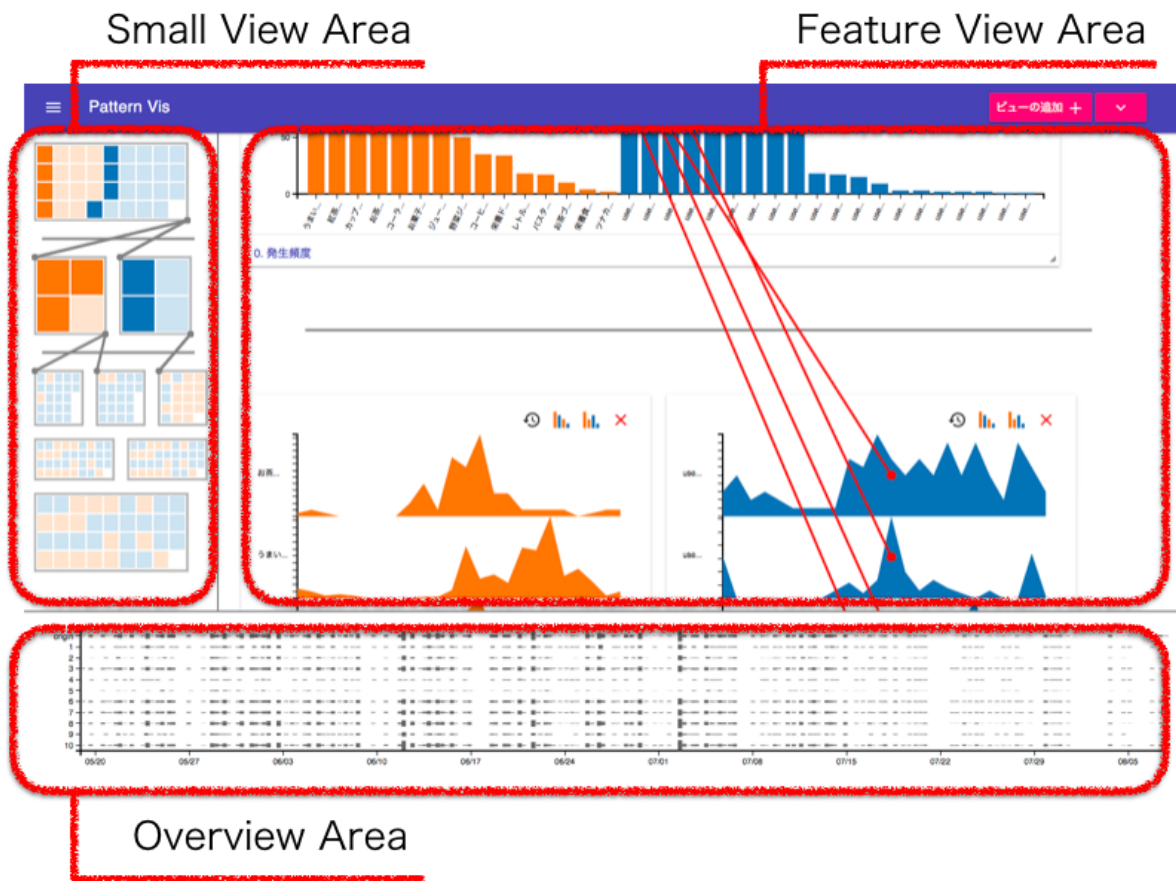


図 5.1: 開発した分析ツール

5.1 分析ツールの設計

分析ツールは探索的分析を支援するために、第3章で示した要求事項を満たす必要がある。そこで、1つめの「観察する特徴の切り替えのための要求事項」のために、分析者は以下の手順を踏むだけで分析を行えるようにした。

1. 観察したいレコードを選択する
2. 観察したい時間的特徴を選択する
3. 特徴を観察する
4. 状況に応じて1に戻る

このような手順で分析を行うことができれば、観察する特徴の切り替えはレコードと特徴の選択のみで可能となると言える。さらに、レコードの選択を支援するために、本ツールでは各エリアごとにインタラクションを備えている。2つ目の「分析の過程を把握するための要求事項」は、Feature View Area と Small View Area で満たす。それぞれのエリアでは観察した特徴の順番や、選択したグラフの要素を視覚的に提示する。3つ目の「分析の状況を把握するための要求事項」は、Small View Area と Overview Area で満たす。それぞれのエリアではデータの参照の状況や、選択したレコードを視覚的に提示する。

以降の5.2節では Feature View Area について、5.3節では Small View Area について、5.4節では Overview Area について、5.5節ではツールのインタラクションについて、それぞれ詳細に説明する。

5.2 Feature View Area

Feature View Area は選択した特徴を観察するためのエリアである。エリア内には分析者が選択した特徴を表すグラフを含むビューが追加される。ビューは特徴の観察を支援するために複数の機能を持つ。

ビューのレイアウトの機能について説明する。Feature View Area に表示されるそれぞれのビューはグラフの種類ごとに決まった縦横比を持ち、この比を元に図 5.2 のように自動でレイアウトされる。もし、非常に多くのビューを追加したとしても、分析者はビューを 1 つ 1 つレイアウトする必要はなく、スクロール機能により昔のビューも簡単に観察可能である。さらに、画面内は自由にスクロールできるだけでなく、分析者はビューのサイズや位置を自由に変えることができる。この機能によって、2 つ以上の特徴を比較したくなった場合は、ビューを近くに並べることができ、1 つの特徴をより大きく表示したい場合は、ビューを大きくすることにより 1 つのグラフを詳細に見ることができる。このように状況に応じた分析が可能であると言える。

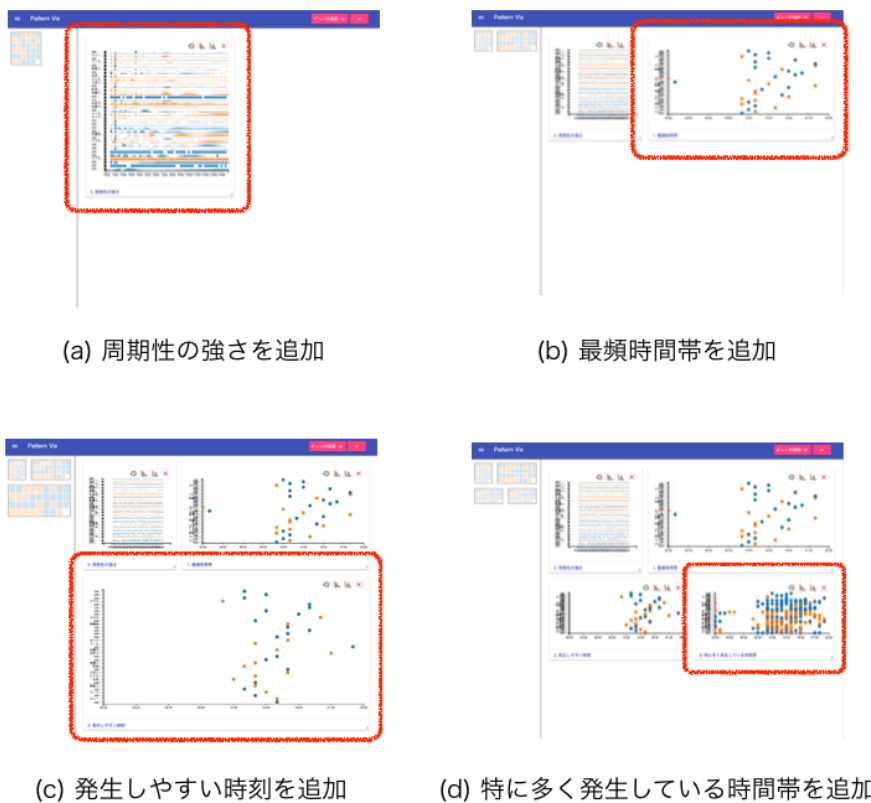
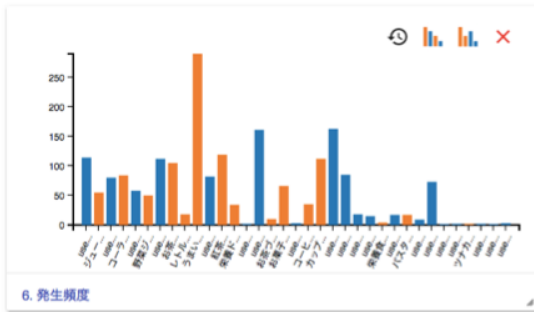
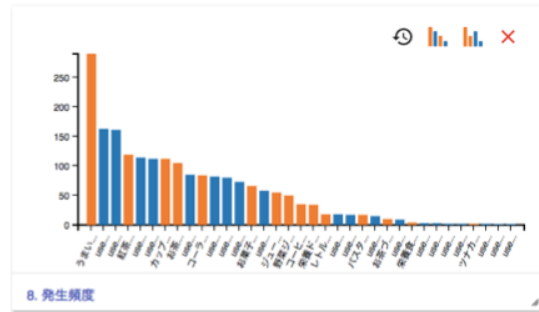


図 5.2: ビューの自動レイアウト

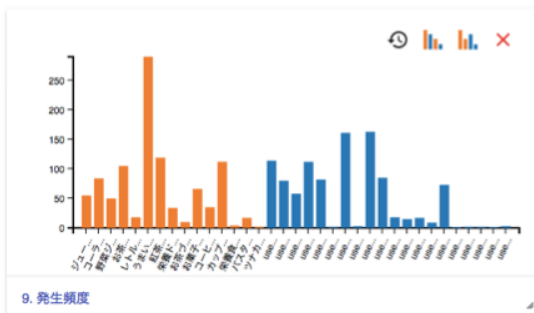
ビュー内のグラフは2種類のソートが可能である。1つ目は特徴に応じたソートである(図 5.3 b)。発生頻度を表すビューであれば発生頻度が高い順番に並ぶようになり、発生しやすい時刻を表すビューであれば時刻が早い順番に並ぶようになる。2つ目はレコードに含まれる変数の種類である(図 5.3 c)。犯罪の発生を表すデータで発生した犯罪の種類と場所という変数を持っていた場合、種類と場所で分けて表示するといったことが可能となる。



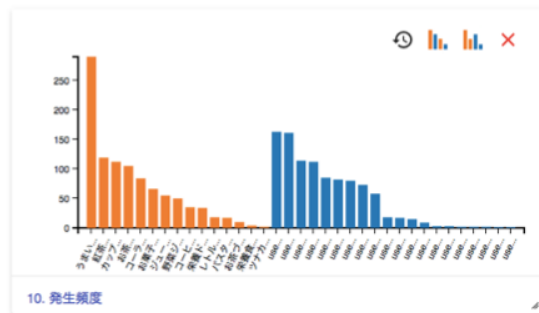
(a) ソートなし



(b) 特徴でソート



(c) 変数の種類でソート



(d) 特徴と変数の種類でソート

図 5.3: グラフのソート

本ツールは今までの選択の内容を記憶しているため、図 5.4 のように、選択の履歴を提示することができる。選択の履歴は線によって表され、線は選択されたグラフの要素をつなぐように描かれる。例えば、発生頻度がある数以上のものを選択して発生頻度分布のビューを追加した場合、発生頻度のビューの選択した要素 1 つ 1 つを始点、発生頻度分布のビューの対応するそれぞれの要素を終点として線は描かれるということである。

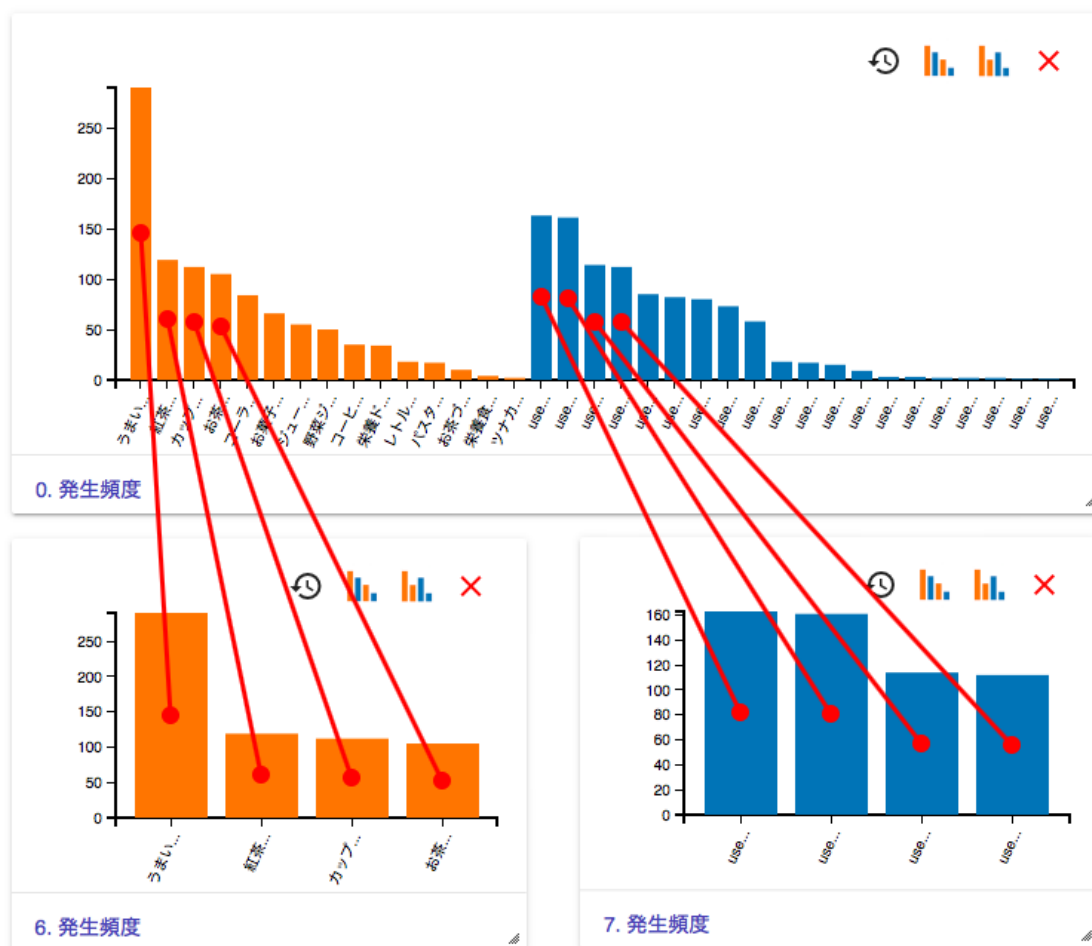


図 5.4: 選択した履歴の表示

5.3 Small View Area

Small View Area は表示したビューを俯瞰するためのエリアである (図 5.5 左)。Small View Area には Feature View Area に追加したビューを縮小したものが表示され、それぞれのビューの位置や大きさは連携している。さらに、ビュー内の要素を選択して新しいビューを追加した場合、Small View Area 内の選択されたビューと新しいビューは線で繋がる。これにより、今までどのようなビューを追加したかを俯瞰可能なだけでなく、それぞれのビューはどのビューに注目した結果追加されたのかも把握することができる。

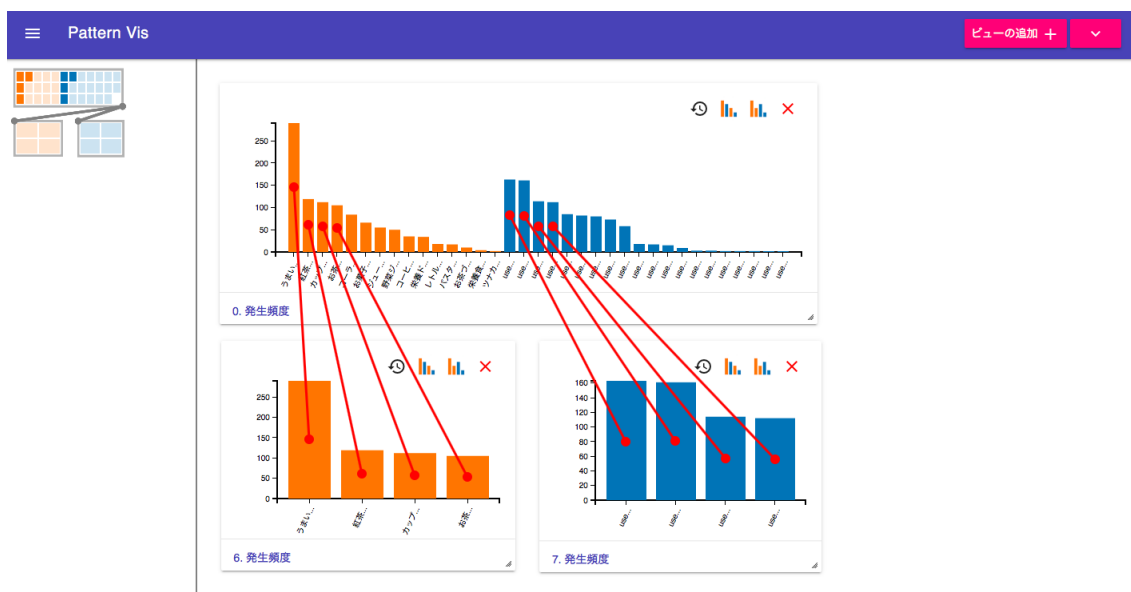


図 5.5: Small View Area と Feature View Area の連携

Small View Area に表示されるビューには、タイルが敷き詰められている。このタイル1つ1つは事象の集合を表しており、それぞれのビューで事象の集合がいくつ表されているかを把握することができる。さらに、ビュー内の要素を選択して新しいビューを追加した場合、このタイルの色はより濃いものとなる。これにより、各ビューの事象がどの程度網羅的に観察されたのかを把握することができる。

5.4 Overview Area

Overview Area はビューで観察しているデータの参照状況を把握するためのエリアである (図 5.6 下)。Overview Area ではデータに含まれるレコードの開始時刻から終了時刻までの各時間帯における発生回数を、グラフ内の灰色の棒の高さで表現している。新しいビューを追加すると、そのビューで参照されているレコードは分けて表現されるため、分析者はビューごとにどれぐらいのレコードを参照しているかを把握できるだけでなく、時間帯ごとに表示されているため 1 つの時間帯のみに発生している場合もすぐに発見することができる。



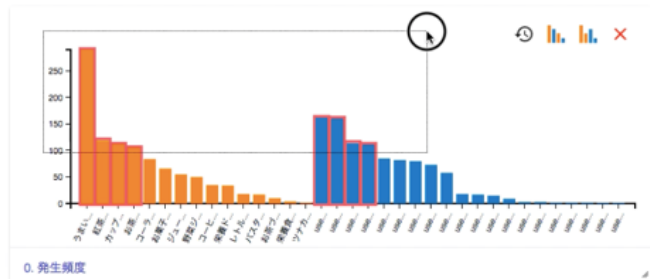
図 5.6: Overview Area の表示

5.5 インタラクション

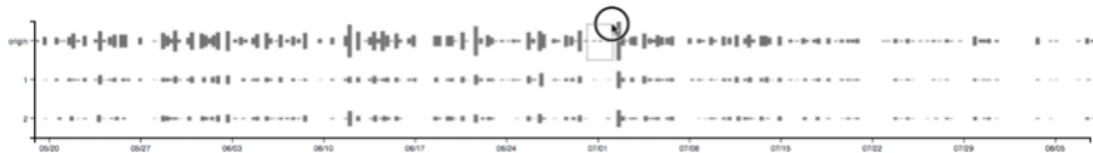
本ツールはレコードの選択やレコードの関係性の把握を支援するために、幾つかのインタラクションを備えている。1つ目は領域選択によるレコードのフィルタリングである。前述したすべてのエリアでは図 5.7 のようにマウスをドラッグすることにより矩形選択をすることができる。この矩形に含まれた青色やオレンジ色の要素はすべて選択された状態となり、要素に紐づくレコードをフィルタリングすることが可能である。例えば、Feature View Area で選択を行えば、ある時刻に集中して発生している事象の集合のみを選択することができたり、Small View Area で選択を行えば、まだ選択されていない事象のみを選択することが可能となる。Overview Area は時間帯ごとのレコードを表示しているため、10月1日から10月31日といったある特定の期間で発生した事象の集合のみを選択することもできる。



(a) Feature View Area



(b) Small View Area



(c) Overview Area

図 5.7: エリアごとの選択の様子

2つ目はマウスオーバーや選択時のブラッシングである(図 5.8)。本ツールではマウスオーバーした要素は赤の線で、選択した要素はピンクの線で囲むようになっている。囲まれる要素はすべてのビューで連携しており、どれか1つのエリアの要素にマウスオーバーしたら、他の2つのエリアの要素にも反映される。本ツールではエリアが複数存在するため、レコードの対応関係を把握する必要が出てきてしまうが、このブラッシングの機能を用いることで、分析者は容易に対応関係を把握できる。



図 5.8: グラフの要素のブラッシング

第6章 ケーススタディ

ケーススタディとして、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¹である。それぞれの分析では特徴が分かっていない状態から、何かしらの傾向を知ることが目的として分析を進める。

6.1 犯罪の種類と犯罪が発生した場所の種類から傾向を得る

この分析では表 6.1 のようなデータを抜き出して分析を行う。ビューのグラフの中で犯罪の種類はオレンジ、犯罪が発生した場所の種類は青の要素と対応している。最初にデータを読み込ませた時点では、まだ分析者は発生の期間やレコードに含まれる属性というような、表 6.1 に記述してある情報しか知らない状態である。

表 6.1: 犯罪の種類と犯罪が発生した場所の種類

レコード数	63746
期間	2015/9/1 - 2015/11/30
属性	犯罪の種類, 犯罪が発生した場所の種類, 発生時刻

まず、犯罪の種類と犯罪が発生した場所の種類はどのようなものがあり、それぞれ何回くらい発生しているのかを知るために「発生頻度」の特徴を選択する。すると図 6.1 のように「発生頻度」を表すビューが新しく追加される。

¹<https://data.cityofchicago.org>

「発生頻度」は棒グラフを用いており、横軸が犯罪の種類や場所の種類といった事象の集合を、縦軸は発生の頻度を表している。グラフを見てみると、犯罪の場所の種類は4つ飛び抜けて多いものがあり(図 6.2 a)、犯罪の種類の方は10個ぐらいまではほどほどに多いことがわかる(図 6.2 b)。また、犯罪の種類は場所の種類よりも3倍ほど種類が多く、犯罪の種類のほとんどは2000回以下であることがわかる。

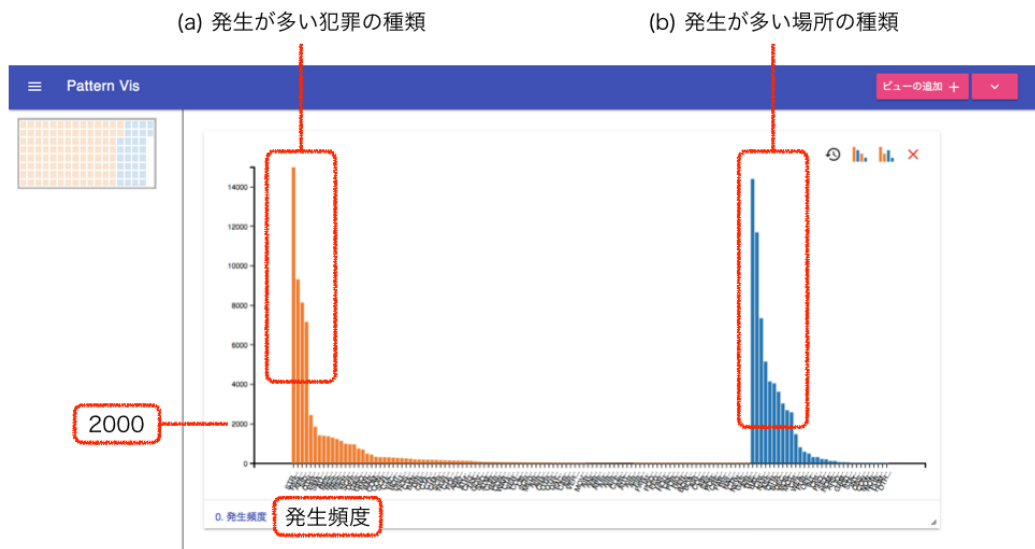


図 6.1: 発生頻度を表示

「発生頻度」を観察することにより、発生した回数の偏りは知ることができたが、この特徴だけでは発生している具体的な時刻は知ることができない。分析者はデータの全体の傾向をさらに知るために、それぞれの事象の集合が何時ぐらいに多く発生しているのかを調べたいと考える。選択できる特徴を見てみると「最頻時間帯」や「発生しやすい時刻」、「特に多く発生している時間帯」という特徴を選択できることがわかる。そこで、これらの特徴のビューを一通り表示する(図 6.2)。

新しく表示した3つの特徴は散布図を用いており、横軸は時刻か時間帯を、縦軸は横軸が犯罪の種類や場所の種類といった事象の集合を表している。

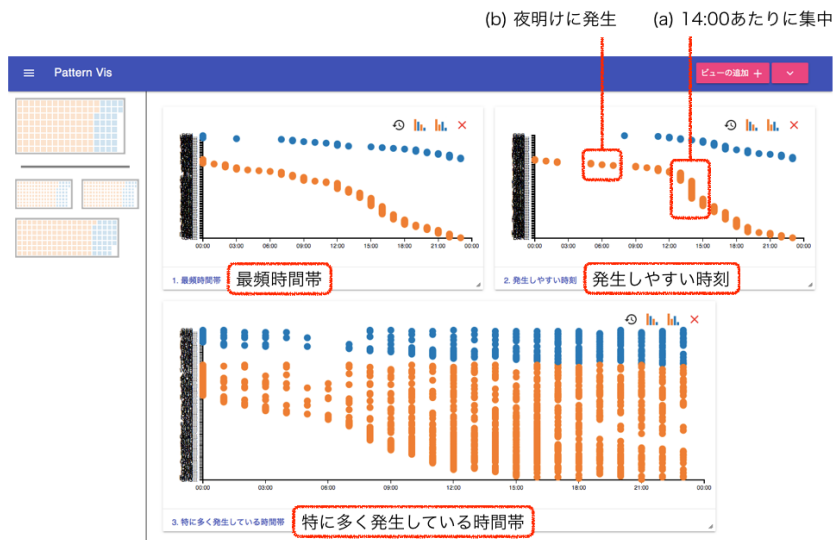


図 6.2: 「最頻時間帯」と「発生しやすい時刻」、「特に多く発生している時間帯」を表示

図 6.2 の「最頻時間帯」を見てみると直線に近い形でプロットされているため、どの時間帯も何かしらの事象の集合が最も高い頻度で発生していることがわかる。「発生しやすい時刻」は図 6.2 a のあたりに固まってプロットされていることから、14 時頃に発生しやすい事象の集合が多いと言える。また、図 6.2 b を見てみると夜明けに発生しやすい事象の集合が存在している。「発生頻度」と連携させてみると頻度はあまり多くないが(図 6.3)、何か特徴があるかもしれないため、これらの事象の集合である VEHICLE(車両)や FIRE STATION(消防署)などに絞ったデータをより多く抜き出すべきだと言える。

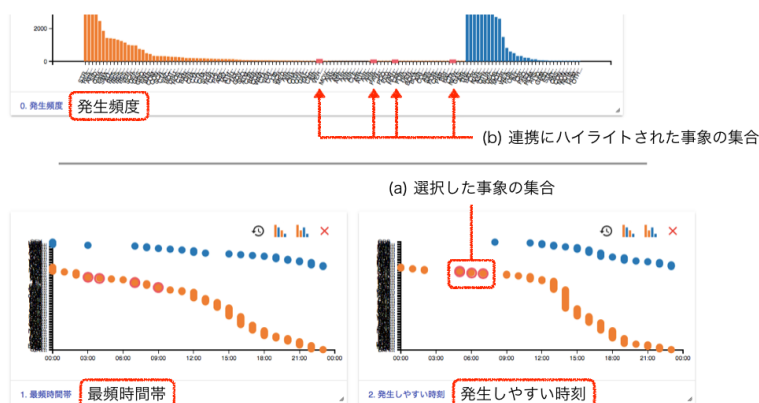


図 6.3: 「発生頻度」と「発生しやすい時刻」の連携の様子

また、このことから発生頻度が高いものに注目してみようとする。そこで最初に追加した「発生頻度」にもう一度注目し、頻度が高いものを選択しそれぞれの特徴をもう一度選択する。

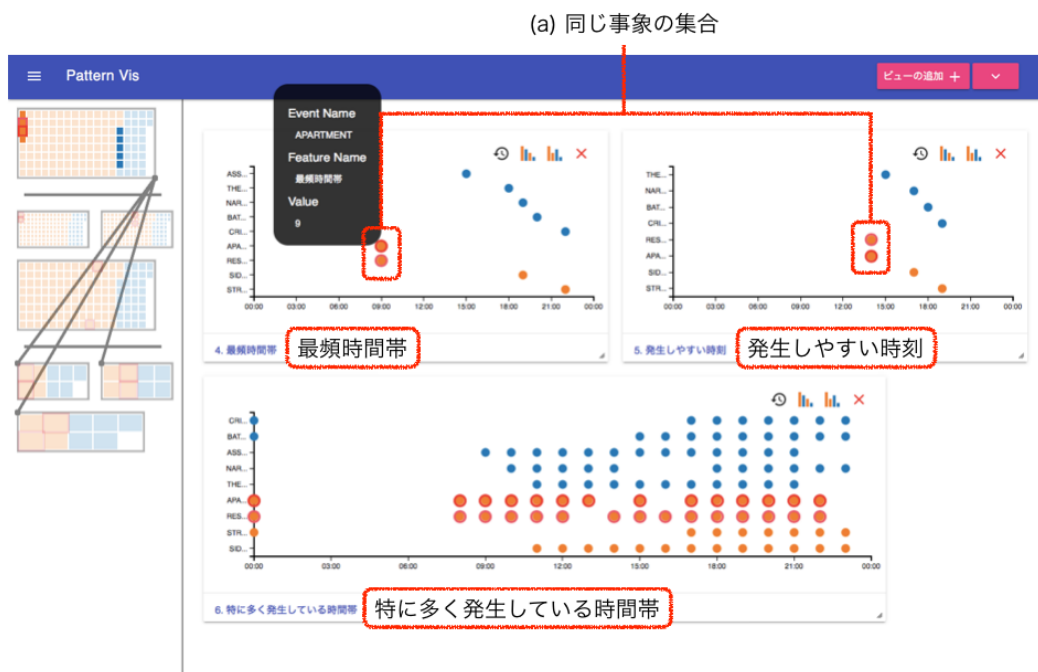


図 6.4: 頻度が高い事象の集合の「最頻時間帯」と「発生しやすい時刻」、「特に多く発生している時間帯」を表示

今度は詳細の表示や選択などを行うことで、事象の集合の複数の特徴を同時に観察していく。すると、図 6.4 a のように「最頻時間帯」では朝の 9 時頃の時間帯を示しているが、「発生しやすい時刻」は昼の 14 時頃の時間帯を示している事象の集合が存在することがわかる。しかし、これらの特徴を見るだけでは実際に 1 日の中で発生頻度がどのように変化しているかわからない。そこで、現在注目している事象の集合 (APARTMENT と PRESIDENCE) を選択し、「発生頻度分布」を新しく表示する。

「発生頻度分布」は面グラフを用いている。面グラフの横軸は時間帯、縦軸は発生の頻度を表し、犯罪の種類や場所の種類といった事象の集合ごとに描かれる。

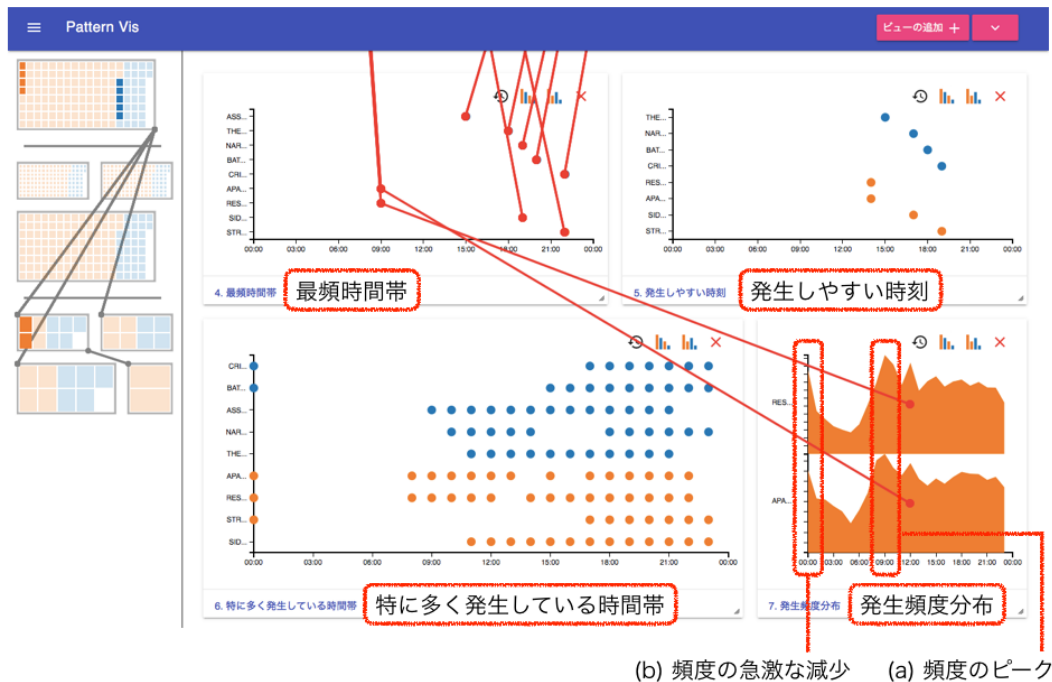


図 6.5: 朝が最も頻度の高い事象の集合の頻度分布を表示

図 6.5 の「発生頻度分布」を見てみると、選択した事象は朝の 9 時頃が頻度のピークとなり (図 6.5 a)、そのあとはだんだんと減少していくが、0 時までにはある程度の頻度で発生していることがわかる (図 6.5 b)。この犯罪が発生した場所ではどのような犯罪の種類が発生しているのかと考える。そのために、注目している場所の種類を 2 つとすべての犯罪の種類を選択し、「同時に発生した回数」の特徴を表示する。

「同時に発生した回数」は行列表示を用いている。行列表示は縦軸も横軸も事象の集合を表しており、発生した回数を変現するためにそれぞれのセルの色の濃さを用いている。

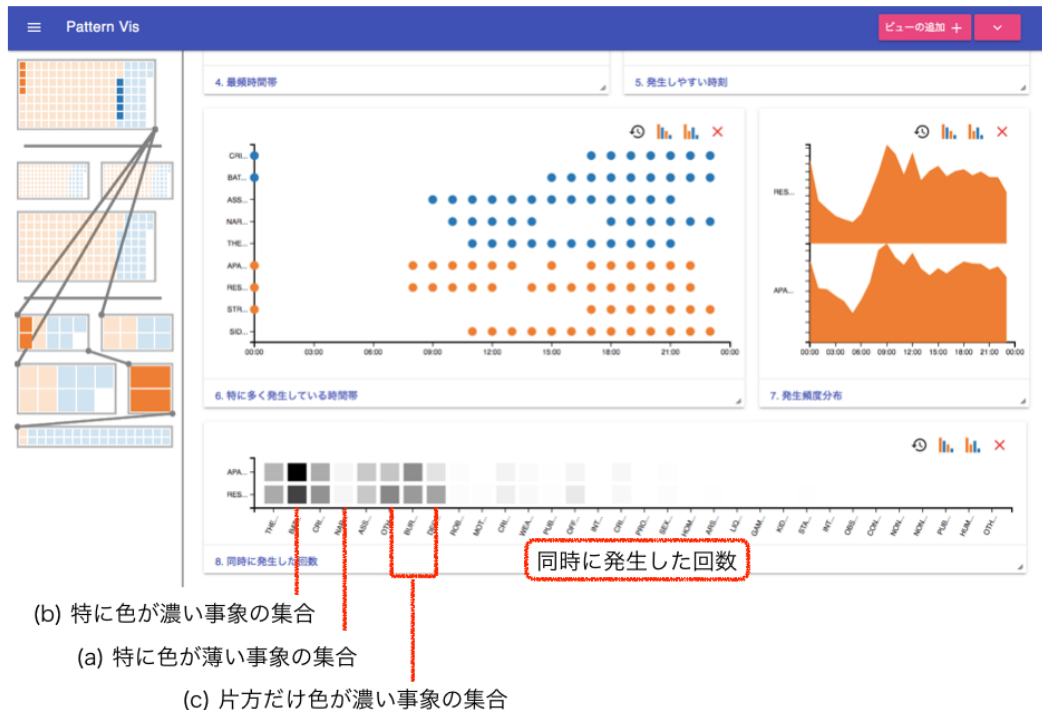


図 6.6: 朝が最も頻度の高い事象と同時に発生した回数を表示

図 6.5 では縦軸が場所の種類、横軸が犯罪の種類を表しており、それぞれの事象集合の「発生頻度」を元にソートされている。最初に表示した「発生頻度」と連携させながらそれぞれの特徴を見てみると、場所の種類として選択されている 2 つの事象の集合は頻度に差がないことがわかる。犯罪の種類は事象集合ごとに大きな差があるため、「発生頻度」を元にソートされている図 6.5 の「同時に発生した回数」のグラフも、左に行くほど色が濃くなっている。しかし、図 6.5 a の示している部分は色が非常に薄くなっていることから、NARCOTICS(麻薬)という犯罪は頻繁に発生しているが、RESIDENCE や APARTMENT ではあまり発生しないと言える。逆に図 6.5 b の示している部分は特に色が濃くなっていることから、RESIDENCE や APARTMENT では BATTERY(暴行)という犯罪が特に多く発生していると言える。図 6.5 c のあたりを見てみると RESIDENCE と APARTMENT で少し色の濃さが異なる部分がある。色の濃さが異なる犯罪は OTHER OFFENSE(その他の犯罪) と DECEPTIVE PRACTICE(欺瞞的慣行)であり、両方とも RESIDENCE の方が回数が多い。

6.2 分析結果のまとめと考察

犯罪発生データに対して探索的分析を行うことで以下のことが分かったと言える。

1. 犯罪は種類や場所に応じて発生回数が大きく異なり、犯罪の種類は上位4個、場所の種類は上位10個ほどに集中する。
2. 発生する頻度がピークに達する犯罪はどの時間帯においても満遍なく存在するが、発生のしやすさという観点では14時あたりを示す犯罪が比較的多い。
3. 6時あたりの夜明けに発生しやすい犯罪も幾つか存在し、これらの犯罪の場所の種類はVEHICLEやFIRE STATIONなどである。
4. 場所の種類であるAPARTMENTとPRESIDENCEは朝の9時頃が頻度のピークとなり、そのあとはだんだんと減少していくが、0時まではある程度の頻度で発生している。
5. APARTMENTとPRESIDENCEでは特にBATTERYが特に多く発生するが、NARCOTICSは比較的発生しにくい。
6. APARTMENTとPRESIDENCEは発生する犯罪の種類や時間帯が似ているが、OTHER OFFENCEとDECEPTIVE PRACTICEはPRESIDENCEの方が多い。

開発したツールでは、これらの分析結果の検証まではサポートしていないが、次の分析の流れとしては分析結果が一般的に言えることかどうかを検証する作業となる。特に3つ目で述べている夜明けに発生しやすい場所の種類は、発生回数が全体に対して少なく、より多くのサンプルを用意する必要があると言える。5つ目でNARCOTICSについて述べているが、実際は発生しないというわけではなく、検挙できていないという可能性も考えられる。データ上では予想することしかできないが、この予想を元に検挙の仕方やデータの収集方法を検討すべきである。

第7章 議論

本章では、探索的分析の支援のために開発した分析ツールの有用性と課題について議論する。開発したツールは3章の要求事項に従い開発したため、まずはそれぞれの要求事項について考察する。

1つ目の「観察する特徴の切り替えのための要求事項」に関しては、十分に満たせていると言える。3.1節で述べている手順では特徴の切り替えに5つのステップを踏む必要があるが、本ツールではレコードの特徴の選択の2つのステップを踏むだけで切り替えが可能である。これにより分析全体の時間を圧縮することができただけでなく、より幅広い特徴を選択することで複数の側面から観察することが可能となったため、観察する量という面だけではなく、質という面でも向上していると言える。さらに、あらかじめ観察できる特徴をメニューとして提供しているため、どのような特徴を見るべきなのかといった分析者のための道しるべとしても役に立つと考えられる。

2つ目の「分析の過程を把握するための要求事項」に関しては、3.2節で述べた内容に関しては満たせていると言える。この節では、特徴の注目した順番と選択した部分について述べていた。注目した順番はSmall View Areaの俯瞰によって、どのような特徴を注目してきたかを即座に把握することができると言える。選択した部分はFeature View Areaの選択の履歴を表示することにより、ある程度分析が進んでしまった後でも、どのようなグラフの要素を選択してきたかを把握することができる。しかし、実際に分析を進めていく中で、その時に考えていたことも知りたいという要望も出てきた。これはツール側で自動的に意図を汲み取るという形か、コメントを残すといった形で実現できると考えられるため、今後の課題である。

3つ目の「分析の状況を把握するための要求事項」に関しては、Overview Areaの改良が必要だと言える。3.3節で述べているレコードの分布に関しては、Overview Areaのグラフを見ることによって、事象の集合ごとにどの時間帯に発生しているかを知ることが可能である。しかし、レコードの割合に関しては幾つか課題が残っている。まずは、データによってはレコードの割合がほとんど見えなくなってしまう点である。現在は棒の高さにより時間帯ごとの発生回数を表しているが、事象の集合ごとに発生頻度に大きな差があるデータでは、発生頻度が低い事象の集合は非常に棒が小さくなってしまい、どれくらい小さいのかを把握することができない。改善策としては、幅ではなく面積などで発生回数を表現するか、ビューごとに絶対的な数を表示する方法が考えられる。

本研究では、時刻付きデータの持つ特徴を観察できる形であらかじめ用意したが、特徴を網羅するためにはさらに特徴を増やす必要があると考えられる。具体的には発生の順序や頻度の変化した時刻など、タイムスタンプを含むデータ特有のものがまだ存在しているため、これらの特徴を観察できるようにしなくてはならないと言える。

第8章 結論

時刻付きデータの探索的分析をするために視点切り替え時の手順の削減及び、分析の過程と状況の把握を支援することを目的として研究を行った。目的を達成するためのアプローチとして、観察したい特徴を選択するだけで観察する視点を切り替えることができ、過程と状況を視覚的に提示するツールを開発した。

観察する視点の切り替えのため、時刻付きデータの分析で有益だと考えられる特徴の洗い出しと定式化、視覚表現の割り当てを行った。開発したツールでは洗い出した特徴を選択することができ、分析者は視点の切り替えを容易に行うことができるようになった。さらに、ツールの持つ複数のエリアは、選択の履歴やデータの参照状況を視覚的に提示するだけでなく、インタラクションを備えることで直感的なデータのフィルタリングを可能とした。

開発した分析ツールのケーススタディとして、犯罪の発生データの持つ傾向を探索した。ケーススタディを行うことで、分析ツールが3章の要求を満たしているだけでなく、さらなる利点を持つことを示した。具体的には観察の量という面だけではなく質という面も向上しているという点と、提供した特徴のリストが分析の道標になるという点である。しかし、コメントを保存やオーバービュー俯瞰は改善が必要だということもわかり、今後の課題だと言える。

今回は犯罪の発生データを対象として分析を行ったが、開発したツールは購買履歴のデータや株価の変動データなど、多様な時刻付きデータに適応可能である。今後はより多くの時刻付きデータの分析を行うことで、分析ツールの可能性や有用性を示したい。

謝辞

本論文を執筆するにあたり、指導教員である三末和男先生には研究方法から研究内容、論文執筆に至るまで丁寧なご指導と適切なご助言をいただき、心より感謝致します。ビジュアルリゼーションとインタラクティブシステム研究室の皆様にはゼミなどを通じて貴重なご意見を数多くいただきました。また、卒業年度には別々の研究室になってしまいましたが、インタラクティブプログラミング研究室の田中二郎先生をはじめ、志築文太郎先生、高橋伸先生、Simona Vasilache 先生には学部生の頃から多くのご助言をいただきました。ここに深く感謝致します。最後に、様々な面で力になってくれた家族や多くの友人、大学生活でお世話になったすべての方々に心から感謝いたします。

参考文献

- [1] Stanley Smith Stevens, “On the Theory of Scales of Measurement”, *Science*, Vol. 103, No. 2684, pp. 677–680, 1946.
- [2] T. Ishigaki, T. Takenaka, and Y. Motomura, “Customer Behavior Prediction System by Large Scale Data Fusion in a Retail Service”, *Transactions of the Japanese Society for Artificial Intelligence*, Vol.26, No.6, pp.670-681, 2011.
- [3] T. Saito, H. N. Miyamura, M. Yamamoto, H. Saito, Y. Hoshiya, and T. Kaseda, “Two-tone pseudo coloring: compact visualization for one-dimensional data”, *IEEE Symposium on Information Visualization*, pp.173-180, 2005.
- [4] W. Javed, Student Member, IEEE, B. McDonnel, Student Member, IEEE, and N. Elmqvist, Member, IEEE, “Graphical Perception of Multiple Time Series”, *IEEE Transactions on Visualization and Computer Graphics*, Vol.16, No.6, pp.927-934, 2010.
- [5] S. Shiroy, K. Misue, and J. Tanaka, “ChronoView: Visualization Technique for Many Temporal Data”, *Proc. of 16th International Conference on Information Visualisation*, pp.112-117, 2012.
- [6] R. Chang, M. Ghoniem, R. Kosara, W. Ribarsky, and J. Yang, “Aligning Temporal Data by Sentinel Events: Discovering Patterns in Electronic Health Records”, *Proc. of the SIGCHI Conference on Human Factors*, pp.457-466, 2008.
- [7] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, B. Shneiderman, and S. Murphy “Wire-Vis: Visualization of Categorical, Time-Varying Data From Financial Transactions”, *IEEE Symposium on Visual Analytics Science and Technology*, pp.155-162, 2007.
- [8] K. Vrotsou, J. Johansson, and M. Cooper “ActiviTree: Interactive Visual Exploration of Sequences in Event-Based Data Using Graph Similarity”, *IEEE Transactions on Visualization and Computer Graphics*, Vol.15, No.6, pp.945-952, 2009.
- [9] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman, “Temporal Event Sequence Simplification”, *IEEE Transactions on Visualization and Computer Graphics*, Vol.19, No.12, pp.2227-2236, 2013.

- [10] C. North, and B. Shneiderman “Snap-Together Visualization: A User Interface for Coordinating Visualizations via Relational Schemata”, *Proc. of the working conference on Advanced visual interfaces*, pp. 128-135, 2000.
- [11] S. G. Eick, Senior Member, IEEE “TVisual Discovery and Analysis”, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 6, No. 1, pp.44-58, 2002.
- [12] N. Boukhelifa, J. C. Roberts, and P. J. Rodgers “A Coordination Model for Exploratory Multi-View Visualization’, *Proc. of International Conference on Coordinated and Multiple Views in Exploratory Visualization*, pp. 76-85, 2003
- [13] F. Chevalier, P. Dragicevic, and C. Hurter, “Histomages: Fully Synchronized Views for Image Editing”, *Proc. of the 25th annual ACM symposium on User interface software and technology (UIST '12)*, pp.281-286, 2012.