

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

拡張型の集約的可視化手法の分類  
および検索システムの構築

長谷川 貴紀

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

本論文は、“拡張型である”と定めた集約的可視化手法の検索を容易にし、積極的な利活用に繋げる目的で行った研究について述べるものである。目的のためのアプローチとして、拡張型の集約的可視化手法をデータの特徴から検索可能にする分類を試みた。

拡張型の集約的可視化手法の分類に際して、まずデータの特徴と結びつくような手法の特徴の分類項目を5つ考案した。次いで、情報可視化分野の主要な国際会議である InfoVis の採択論文20年分から30種類の該当手法を抽出し、考案した分類項目に基づき実際に分類を行った。

分類の結果から、マーカがもつ次元数や情報の尺度水準、用いられる視覚変数の傾向を把握し、拡張型の集約的可視化手法が情報を見やすく、正確に表現するべく設計されていることが推測できた。

また、分類の有用性や課題を見出すため、分類結果を元に拡張型の集約的可視化手法検索システムを構築し、ユーザスタディによって使用感を調査した。ユーザスタディの結果、可視化手法検索時において検索条件としては元データの特徴が重要視され、検索結果では見た目が重要視されるということが推測された。したがって、今後の可視化手法検索に関わる研究の方向性として、元データの特徴で検索し、ビジュアルな検索結果から元データの特徴を読み取れる仕組みづくりを提案する。

# 目次

<b>第1章</b>	<b>はじめに</b>	<b>1</b>
1.1	拡張型の集約的可視化手法	1
1.2	研究の動機	2
1.3	本研究の目的とアプローチ	2
1.4	本論文の構成	3
<b>第2章</b>	<b>関連研究</b>	<b>4</b>
2.1	可視化手法の分類に関する関連研究	4
2.2	可視化手法検索に関する WEB サイト	4
<b>第3章</b>	<b>拡張型の集約的可視化手法の分類項目</b>	<b>5</b>
3.1	拡張型の集約的可視化手法の分類に対する要件	5
3.2	考案した分類項目	5
<b>第4章</b>	<b>拡張型の集約的可視化手法の分類</b>	<b>8</b>
4.1	抽出した拡張型の集約的可視化手法	8
4.2	拡張型の集約的可視化手法の分類結果	9
4.2.1	レコード数およびスケーラビリティ	9
4.2.2	次元数	10
4.2.3	可視化されている情報およびその尺度水準	11
4.2.4	表現空間	11
4.2.5	視覚変数	12
4.3	分類結果に対する考察	12
<b>第5章</b>	<b>拡張型の集約的可視化手法検索システム</b>	<b>16</b>
5.1	拡張型の集約的可視化手法のモデル化	16
5.2	検索インタフェース	18
<b>第6章</b>	<b>ユーザスタディ</b>	<b>20</b>
6.1	使用したデータ	20
6.2	拡張型の集約的可視化手法検索タスク	20
6.3	アンケート調査	21
6.4	拡張型の集約的可視化手法検索タスクの結果および考察	21

6.5 アンケート調査の結果および考察 . . . . .	22
第7章 結論	25
謝辞	26

# 目 次

1.1	1995 年から 2014 年までの拡張型の集約的可視化手法数 . . . . .	2
4.1	Christian Tominski, Heidrun Schumann, Gennady Andrienko, and Natalia Andrienko, Stacking-based visualization of trajectory attribute data. . . . .	11
4.2	Diansheng Guo, Flow mapping and multivariate visualization of large interaction data. . . . .	12
5.1	拡張型の集約的可視化手法に関する ER 図 . . . . .	17
5.2	拡張型の集約的可視化手法検索システムの概観 . . . . .	18

# 表目次

4.1	レコード数およびスケーラビリティによる分類結果 . . . . .	10
4.2	次元数による分類結果 . . . . .	10
4.3	可視化されている情報およびその尺度水準による分類結果 . . . . .	14
4.4	表現空間による分類結果 . . . . .	15
4.5	視覚変数による分類結果 . . . . .	15
6.1	ユーザスタディに使用したデータのカラム構造 . . . . .	20
6.2	拡張型の集約的可視化手法検索タスクの結果 . . . . .	21
6.3	リッカート尺度のアンケート回答結果 . . . . .	22
6.4	検索条件のうち重要だと感じたもの上位3つとその順位の回答結果 . . . . .	23
6.5	検索結果のうち重要だと感じたもの上位3つとその順位の回答結果 . . . . .	24

# 第1章 はじめに

本研究では、“拡張型の集約的可視化手法”と呼ぶ可視化手法を対象とした分類および検索システムの構築を行った。

本章では、以下 1.1 節で拡張型の集約的可視化手法とは何かを説明し、1.2 節で本研究の動機となった背景、1.3 節で本研究の目的とアプローチ、1.4 節で本論文の構成を述べる。

## 1.1 拡張型の集約的可視化手法

Shneiderman [1] によると、可視化手法は1つのマーカに対応するデータレコードの数によって大きく2種類に分類することができる。マーカとは、可視化手法中に用いられ情報を表現する単位図形のことである。

アトミックな可視化手法と呼ばれる可視化手法では、1つのデータレコードを1つのマーカで表現する。アトミックな可視化手法には、元々のデータレコードの情報を厳密にプロットできるという利点がある。一方で、可視化領域上で同じ位置にプロットされるデータレコードが存在する場合や、データレコード数が膨大になった場合には、マーカの重畳による視覚的混雑や描画領域の不足を生じるという欠点がある。

集約的可視化手法と呼ばれる可視化手法では、複数のデータレコードを1つのマーカで表現する。集約的可視化手法は複数のデータレコードの情報を1つのマーカで表現するため、可視化領域上のマーカ数を削減することができ、重畳による視覚的混雑度の低減やデータ全体の視覚的な俯瞰が行えるという利点がある。但し、複数の情報が集約されて見えるため、元々のデータレコードの情報を厳密にはプロットできないという欠点がある。

上記の分類のうち、本研究では集約的可視化手法について更に“拡張型”であるものを独自に定義し、対象としている。“拡張型の集約的可視化手法”が満たす条件は以下のとおりである。

1. 1つのマーカが複数のデータレコードと対応する
2. マーカが2変数以上を視覚的に表現している

条件1については集約的可視化手法の定義である。条件2は、例えば集約したデータレコード数という1変数のみを表現するような単純な可視化手法は際限なく作成できると考え、それらを含まないために設けた。

## 1.2 研究の動機

本研究で拡張型の集約的可視化手法を対象とした背景として、大規模データや多次元データの流行がある。“ビッグデータ”と呼称されるようなデータレコード数の膨大なデータや、単一のデータレコードを見ても多数の属性を含むようなデータが積極的に蓄積・分析されている。

こうしたデータに対する可視化手法として、アトミックではなく集約的であり、多数の次元を表現可能な手法の需要が高まっていると考え、対象とした。実際に、後述の分類を進めていく中で過去 20 年間から抽出できた拡張型の集約的可視化手法の数は増加の傾向にあった(図 1.1)。

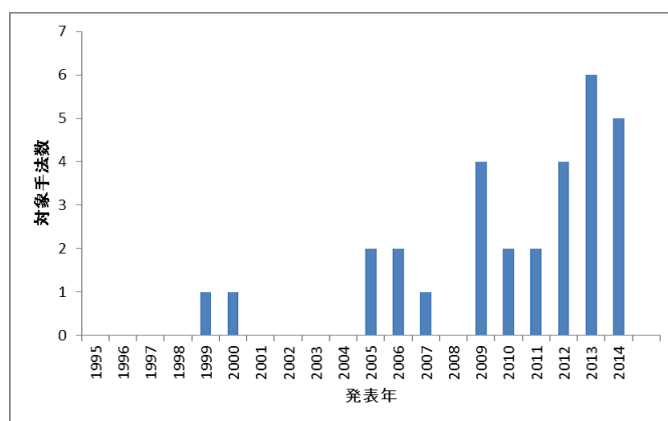


図 1.1: 1995 年から 2014 年までの拡張型の集約的可視化手法数

このようなデータの流行とそれに適した可視化手法への需要の高まりの中で、実際に開発された可視化手法を把握し、利用することを支援する仕組みが不十分なのではないかと考えた。特に、大規模・多次元データの流行については、それを所持していたり扱おうとする人びとは情報可視化の専門家に限らないため、可視化手法の検索がより困難であると考えられる。こうした情報可視化の専門家以外の人びとも含め、より多くの人びとが既存の可視化手法を積極的に利活用可能な仕組みを提供したいと考え、本研究に着手した。

## 1.3 本研究の目的とアプローチ

本研究の目的は、拡張型の集約的可視化手法の積極的な利活用のため、容易に可視化手法を検索することが可能な仕組みを提供することである。従来の可視化手法分類においては、分類結果を見ることで各可視化手法がどのような特徴を持っているかは把握できたものの、手法について未知の場合には、その特徴から手法を探すことが難しかった。

そこで本研究では、データの特徴と紐付くような可視化手法の特徴によって分類を行うというアプローチをとった。また、提案する分類の有用性を確認するため、分類結果に基づく



検索システムを作成した。

## 1.4 本論文の構成

本論文の構成は、以下第2章で関連研究を紹介し、第3章で拡張型の集約的可視化手法の分類項目についてその要件と具体的な項目を説明する。第4章では実際に行った拡張型の集約的可視化手法の分類について結果を示すとともに考察を行う。第5章では作成した拡張型の集約的可視化手法検索システムについて述べ、第6章で当該システムを用いたユーザスタディに関する内容および、それを踏まえた議論を行う。おわりに、第7章で本論文全体の結論を述べる。

## 第2章 関連研究

研究を進めるにあたり参考にすべき研究として、拡張型の集約的な可視化手法に限らず広く可視化手法の分類に関する研究と、可視化手法の検索に関する WEB サイトについて調査を行った。

### 2.1 可視化手法の分類に関する関連研究

Harris [2] は、様々な可視化手法について手法の図と簡潔な説明文を用いて説明し、また見た目の系によって分類を行っている。可視化手法の見た目を想像して、あるいは既知の手法に見た目の近い手法を検索する場合には、簡単に多くの手法を把握することができる。しかし、未知の手法や見た目の想像を伴わない可視化手法検索のためには、手法名や見た目の系による分類は手がかりになりにくい。

本研究では、可視化の見た目の想像を伴うか否かによらず、可視化したいデータに適用可能な幅広い可視化手法を提示したいと考えた。そこで、可視化したいデータの特徴に紐付くよう可視化手法の特徴を分類し、検索を可能にしている。

### 2.2 可視化手法検索に関する WEB サイト

Schultz は、[treeviz.net](http://treeviz.net) [3] に代表されるような、可視化手法の構造や対象データによって可視化手法をまとめ、検索が可能な WEB サイトを提供している。検索条件は見た目の系や対象データの視覚的エンコード方法を、予め用意されたものから選択する、検索結果の表示には可視化手法の画像のみを敷き詰める方式をとっている。

可視化手法の見た目がまず把握できることで、利用者の想像に近い手法を検索することに適していると考えられる。一方で、見た目が想像に近くとも、可視化したいデータに適用できるかどうかは見た目からだけでは判断が難しい。また、データの情報とそれを表現する視覚変数の組合せが予め用意されたものからの選択になるため、検索の自由度が低いと考えられる。

本研究では、データの特徴を出発点に可視化手法を検索するアプローチをとるため、検索結果がデータに適用できるかどうかの判断が容易になる他、組合せではなく個々の特徴の自由度が高くなると考えられる。自由な組合せによる検索は存在しないレコードを抽出する可能性もあるが、データベースに新たな組合せが追加されるときにも柔軟に抽出が可能であるという利点があると考えられる。

## 第3章 拡張型の集約的可視化手法の分類項目

拡張型の集約的可視化手法を実際に分類するにあたり、どのような項目で分類を行うかを検討した。以下、分類に対する要件とそれを満たすべく考案した分類項目について述べる。

### 3.1 拡張型の集約的可視化手法の分類に対する要件

本研究で提案する分類の要件を定めるにあたり、本研究の分類を利用する想定ユーザを“少なくとも可視化しようとするデータに対しては理解のある人”とした。その上で、提案する分類に対する要件を以下のように定めた。

1. 最低限データの特徴さえ把握していれば、適用可能性のある可視化手法を検索可能にする。
2. 簡単な見た目や使い方を指定して検索することができる。
3. 厳密に検索条件に一致しない場合にも、適用可能性のある可視化手法を提示する。

要件1については、想定ユーザが所持する最低限の情報から可視化手法を検索可能にすることで、可視化手法の検索を容易にするという目的を達成するものである。要件2は、情報を可視化する際には、情報可視化の専門家でもなくとも少なからず見た目や使い方を想像するであろうと考え、要件とした。また、可視化手法の中には具体的な対象データや使用用途を想定して開発されるものもあるため、そのような可視化手法が表現している情報とユーザの可視化したいデータの情報が厳密に一致しない可能性が考えられた。そこで、提案する分類では要件3として、厳密に検索条件に一致しない場合にも、適用可能性のある可視化手法を提示することを定めた。これにより、各可視化手法が開発されたときに元々想定されていた対象データ以外にも可視化手法を適用できる可能性を広げ、拡張型の集約的可視化手法の積極的利活用に繋げるという目的を達成するものである。

### 3.2 考案した分類項目

前述の目的と分類に対する要件を踏まえ、分類に用いる項目を検討した。まず、要件1を満たすため、データの特徴と紐付くような可視化手法の特徴によって分類する必要があると考えた。要件1に対応する分類項目は以下のとおりである。

### データレコード数およびスケーラビリティ

可視化したいデータのデータレコード数が可視化手法が表現可能なデータレコード数の範囲内であれば適用できる可能性がある。また、あらゆるデータレコード数を表現可能(スケーラブル)な手法であると述べられているものもあるため、スケーラブルであるかどうかについても分類を行う。

### 次元数

可視化したいデータレコードから何次元の情報を視覚的に表現できるかを把握することで、次元の取捨選択や可視化手法の選別が行える。

### 可視化されている情報

可視化手法で表現されている具体的な情報が一致すれば、適用できる可能性がある。

ここで、要件3を満たすべく、可視化されている情報に紐づく以下の分類項目を提案する。

### 可視化されている情報の尺度水準

可視化されている情報の尺度水準によっても分類することで、可視化したいデータの情報と可視化されている情報が一致しなくとも、尺度水準が一致していれば適用できる可能性があると考えた。

尺度水準とは、情報とその数学的・統計的性質によって分類するものであり、Stevens [4] によって以下のように提唱されたものが広く知られている。

#### 名義尺度

同一性のみに興味があり、値の大小を比較することや演算も不可能である。

#### 順序尺度

数字やアルファベットが割り振られ、同一性に加え順序の比較が可能である。しかし、加減算等の演算には意味がない。

#### 間隔尺度

数字が割り振られ、順序尺度の性質を全て満たす。その上で、加減算等の演算による差の等しさは間隔の等しさと一致し、意味があるものである。尺度上のゼロ点が任意であるため、直接の値同士の乗除算には意味が無いが、差の乗除算は意味をもつ。

#### 比例尺度

割り振られた数字は間隔尺度の性質を全て満たす。ゼロ点が絶対的に定まっているため、直接の値同士の乗除算も意味をもつ。

上記のような情報の性質による分類のため、尺度水準が一致すれば可視化手法を適用できる可能性があると考え、分類項目として採用した。

また、要件2を満たすために以下の分類項目を考案した。

### 表現空間

2次元平面上であったり3次元空間であったりといった、情報可視化の専門家でなくとも想像が可能な見た目として、可視化に用いられる表現空間によって分類する。

### 視覚変数

可視化する情報(値)を視覚的にエンコードするものである。つまり、可視化したい情報を形や色、大きさ等の視覚的属性のいずれを用いて表現したいかという指定に応えるための分類である。

以上6つの分類項目に沿って対象手法を分類していく。

## 第4章 拡張型の集約的可視化手法の分類

考案した分類項目に沿って実際に分類をするにあたり，本研究では情報可視化分野の主要な会議である InfoVis(IEEE Information Visualization Conference) について，第1回の1995年から2014年までの採択論文から該当手法を抽出し，分類することとした．本章では，20年間で通算539本の論文から該当手法を含む28本の論文(30手法)を抽出・分類した結果について述べる．

### 4.1 抽出した拡張型の集約的可視化手法

InfoVis20年間の採択論文から抽出できた拡張型の集約的手法一覧を以下に示す．

1. Aggregate Towers [5]
2. ThemeRiver [6]
3. Interactive Sankey Diagrams [7]
4. Parallel Sets [8]
5. ASK-GraphView [9]
6. Meta-Node (Perer ら) [10]
7. DAG Visualization [11]
8. Phrase Nets [12]
9. Wordle [13]
10. Flow Mapping and Multivariate Visualization [14]
11. Smooth Graphs [15]
12. Opinion Seer (Categorical Ring) [16]
13. Necklace Maps [17]
14. Angular Histograms [18]

15. TreeNetViz [19]
16. RankExplorer (Flow-in Bar and Flow-out Bar) [20]
17. RankExplorer (Changing Glyph) [20]
18. Outflow Visualization [21]
19. Time Lens [22]
20. SoccerStories (Corner Kicks and Crosses) [23]
21. Radial Sets [24]
22. Meta-Nodes (Vehlow ら) [25]
23. InProv [26]
24. StoryFlow (LOD technique 使用) [27]
25. Common Angle Plots [28]
26. RoseRiver (Overview) [29]
27. RoseRiver (Topic Bar) [29]
28. LiveGantt [30]
29. High-Level Infographic-Style Overview [31]
30. Origin-Destination Flow Data Smoothing and Mapping [32]

1つの論文に複数の該当手法が含まれる場合にも、それぞれを1手法として抽出し分類している。

## 4.2 拡張型の集約的可視化手法の分類結果

前節で挙げた30の手法について、考案した分類項目ごとに分類結果を示す。

### 4.2.1 レコード数およびスケーラビリティ

レコード数による分類は、論文中で明記されている対象レコード数あるいはユースケース等で用いられているデータレコード数によって行った。また、スケーラブルであるかどうか、論文中でスケーラブルであると明記されているかどうかによって分類している。1つの手

表 4.1: レコード数およびスケーラビリティによる分類結果

レコード数	該当手法番号
237	28
847	15
1000	23
1253	22
10000	23
18577	28
50000	21
56700	24
78294	29
93872	4
1048566	14
スケーラブルな手法	5, 6, 19

法は複数のデータレコード(ユースケース)をもつことがあり、また、スケーラビリティも同時に有する可能性がある。分類の結果を表 4.1 に示す。

ここで、スケーラブルな手法 19(Tominski, '12) の画像を図 4.1 に示す(掲載許可取得済み)。Time Lens は円環状に並ぶセルの幅(同心円方向)がデータレコード数を表すが、最大の幅(外側の太いグレーの線)を固定にすることでスケーラブルに利用できると考えられる。しかし、最大値と最小値の差が大きくなった場合には値の差が読み取りづらくなることが予想される。

#### 4.2.2 次元数

次元数による分類は、各可視化手法上で一度に表現されている情報の数のみをカウントして行った。即ち、インタラクション等を用いて異なる次元を見ることができたり、複数のデザインパターンをもつ場合にもそれぞれの状態で表現されている次元数のみによって分類している。分類の結果を表 4.2 に示す。

表 4.2: 次元数による分類結果

次元数	該当手法番号
2	3, 4, 5, 6, 9, 11, 12, 13, 15, 21, 23, 25
3	1, 2, 7, 14, 16, 17, 18, 19, 22, 26, 27, 29, 30
4	20, 24, 28
5	8, 10



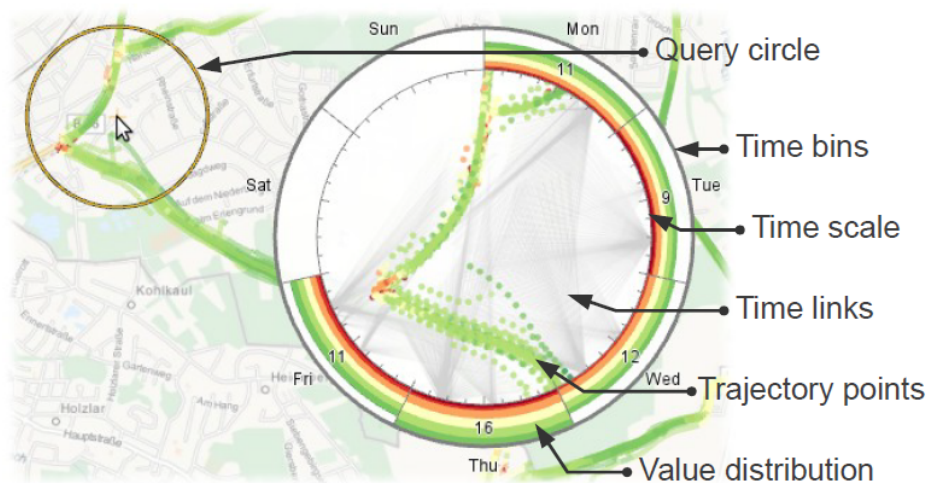


図 4.1: Christian Tominski, Heidrun Schumann, Gennady Andrienko, and Natalia Andrienko, Stacking-based visualization of trajectory attribute data.

ここで、もっとも多い5次元の情報を表現している手法10(Guo, '09年)の画像を図4.2に示す(掲載許可取得済み)。

図4.2では、円の中心位置が人口の重心、円の面積がモジュラリティ、エッジの幅がデータレコード数、エッジの明度がエッジの密度を表しており、加えてDirected Flowsモードであれば(画像はUndirected Flows)エッジが矢印形になりエッジの向きが表される。

#### 4.2.3 可視化されている情報およびその尺度水準

可視化されている情報の分類においては、集約した情報の量を論文内での記載名によらず広く“データレコード数”と見なすなど、表記の違いの範疇であると判断したものについては同一視し分類した。また、ASK-GraphView(手法5)においては不特定の名義尺度データが対象であったため、不特定として分類している。分類の結果を表4.3に示す。

#### 4.2.4 表現空間

表現空間による分類においては、2次元平面上での可視化が多く見られた。2次元平面以外の地図およびサッカーフィールドも広義には2次元平面といえるが、位置が表す値が座標ではなく、背景となっている表現空間上の位置である点が特徴的であると考え、それぞれ独立した分類とした。分類の結果を表4.4に示す。

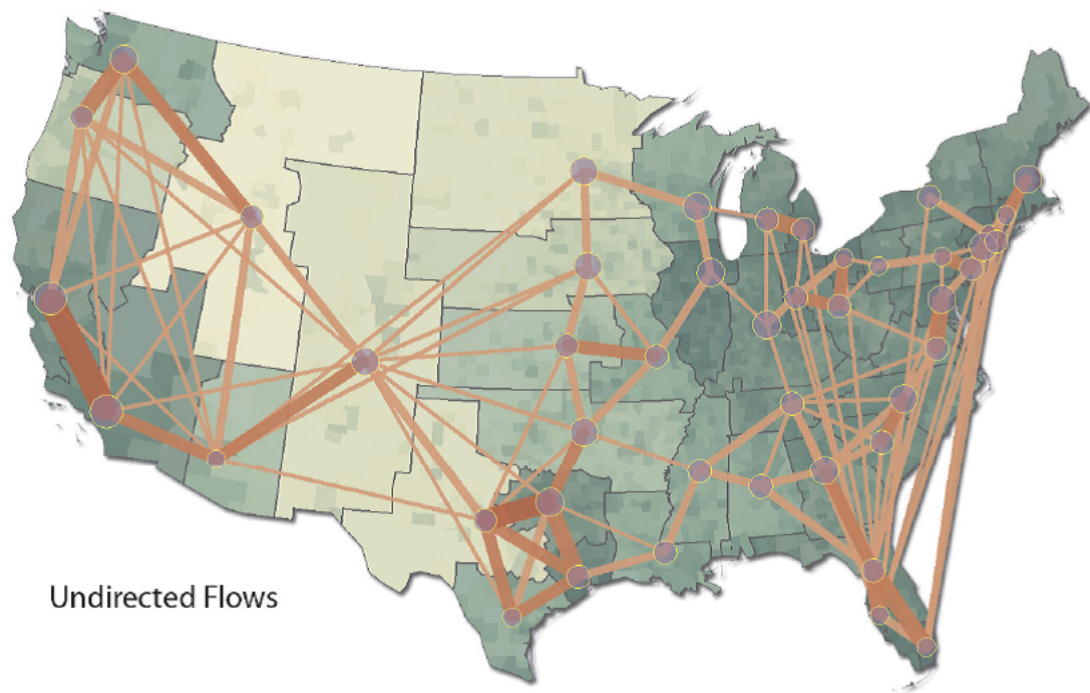


図 4.2: Diansheng Guo, Flow mapping and multivariate visualization of large interaction data.

#### 4.2.5 視覚変数

視覚変数による分類においては、色(色相)や長さ(幅)を用いている可視化手法が多く見られた。また、各可視化手法の次元数が2以上であるため、視覚変数間で該当手法の重複がある。視覚変数による分類の結果を4.5に示す。

### 4.3 分類結果に対する考察

まず、レコード数およびスケーラビリティによる分類結果について、具体的なレコード数やスケーラブルな手法であるということが明記されていたものがさほど多くなかった。したがって、可視化するデータのレコード数から手法を検索することがやや困難になることが考えられる。扱えるデータレコード数への言及やスケーラブルであることの断言が少ない原因としては、従来の手法と比較して“よりスケーラブル”といった表現がしばしば見受けられたり、収集・蓄積されるデータの規模が大きくなり続ける中でスケーラブルであると断言することの難しさがあるのではないかと考えられる。

次元数による分類の結果について、次元数が2および3の可視化手法が多くを占めることがわかる。これは、表現空間による分類結果が広義では全て2次元平面といえることから、2次元平面内で用いることのできる視覚変数の数の制約によるところが大きいと考えられる。ま

た、1つのマーカが表現する次元数が多くなるほど、表現されている情報の読み取りが直感的でなくなるため、設計の段階で次元数を抑える選択をしている可能性も考えられる。

可視化されている情報およびその尺度水準による分類結果について、可視化されている情報の分類を上手く統合することができず、可視化手法と1対1で対応するものが増えてしまった。したがって、可視化したいデータの情報と可視化されている情報が完全に一致する可視化手法を提示できる可能性が低くなることが予想される。

視覚変数による分類結果について、色(明度)を用いているものと長さ(幅)を用いているものが多いことがわかる。ここで、可視化されている情報の尺度水準による分類を参照すると、名義尺度と比例尺度の情報が多く可視化されていることがわかる。尺度水準と視覚変数の関係については、Mackinlay [33] が、視覚変数が各尺度水準を表す正確さを明らかにしており、各尺度水準で位置が最も精度が高いものの、比例尺度では次いで長さが、名義尺度では次いで色相が精度が高い。尺度水準による分類結果と視覚変数による分類結果の対応は表現の精度の高い組合せと一致しており、可視化手法の設計者が表現する情報の尺度水準に適した視覚変数を用いていることがうかがえる。

表 4.3: 可視化されている情報およびその尺度水準による分類結果

可視化されている情報	尺度水準	該当手法番号
レコード数	比例尺度	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29
任務タイプ	名義尺度	1
位置 (緯度経度)	間隔尺度	1, 10, 30
トピック (イベント)	名義尺度	2, 18, 20, 26, 27
時刻	間隔尺度	2, 16, 17, 18, 19, 21, 24, 26, 28
エネルギータイプ	名義尺度	3
クラス	名義尺度	4, 25
不特定	名義尺度	5
最大ランク	比例尺度	6
階層	順序尺度	7, 15, 23, 27
関係タイプ	名義尺度	7
クラスタ	名義尺度	8, 22
入次数に対する出次数の比	比例尺度	8
エッジの向き	名義尺度	8, 10, 11, 29, 30
単語	名義尺度	9
人口密度	比例尺度	10
エッジの重み	比例尺度	8
モジュラリティ	比例尺度	10
年齢	間隔尺度	12
国名	名義尺度	13
密度	比例尺度	26
関連	名義尺度	14
貢献度	比例尺度	16
変化パターン	名義尺度	17
速度	比例尺度	19
位置	間隔尺度	20
曖昧度	比例尺度	22
セッション	名義尺度	24
エンティティ	名義尺度	27
リソース	名義尺度	28
タスクタイプ	名義尺度	28
ノード	名義尺度	29
レコード数のランク	順序尺度	30

表 4.4: 表現空間による分類結果

表現空間	該当手法番号
2次元平面	2, 3, 4, 5, 6, 7, 8, 9, 11, 12, 14, 15, 16, 17, 18, 21, 22, 23, 24, 25, 26, 27, 28, 29
地図	1, 10, 19, 30
サッカーフィールド	20

表 4.5: 視覚変数による分類結果

視覚変数	該当手法番号
色 (色相)	1, 2, 3, 4, 5, 6, 12, 14, 15, 18, 19, 20, 21, 22, 24, 25, 26, 28
色 (明度)	8, 10, 20, 27, 29, 30
位置	1, 2, 10, 13, 16, 17, 18, 19, 20, 24, 26, 27, 28, 30
体積	1, 7
長さ (幅)	2, 3, 4, 8, 10, 11, 13, 14, 16, 18, 19, 22, 23, 24, 25, 26, 27, 28, 29
面積	5, 6, 7, 10, 12, 13, 15, 17, 21
形	7, 8, 10, 17, 29, 30
フォントサイズ	8, 9
テキスト (文字)	8, 9
アニメーションの方向	11
傾き	14

## 第5章 拡張型の集約的可視化手法検索システム

本研究では、前章で述べた分類結果を用いて拡張型の集約的可視化手法を検索可能なシステムを構築した。検索システムは分類の有用性をユーザスタディによって検証する目的と、モデル化によって分類を現実的に運用する目的で構築した。実装は Ruby on Rails を用いて行い、Ruby は ver. 2.2.2, Rails は ver. 4.2.5, データベースは Sqlite3 ver. 3.8.10.2 を使用した。

### 5.1 拡張型の集約的可視化手法のモデル化

拡張型の集約的可視化手法の分類結果をデータベース化するにあたり、拡張型の集約的可視化手法をモデル化した。設計し実装したモデルの ER 図を図 5.1 に示す。

#### Visualization モデル

画像 ID, 論文 ID, 手法名, 発表年, ユースケースの有無, スケーラブルか否か, インタラクションの有無, 次元数, 表現空間 ID をもつ。中間テーブルを介してデータレコード数, エンコード, と多対多のリレーションをもつ。

#### Paper モデル

論文タイトル, ファイルパスをもつ。1つの論文に対して複数の可視化手法をもち得る。

#### Image モデル

画像ファイルをバイナリでもつ。

#### Space モデル

表現空間名をもつ。1つの表現空間に対して複数の可視化手法をもち得る。

#### Record モデル

データレコード数をもつ。中間テーブルを介して可視化手法と多対多のリレーションをもつ。

#### VisualizationRecord モデル

可視化手法とデータレコード数の中間テーブルモデル。

#### Encode モデル

次元(可視化する情報)を視覚変数にエンコードする組合せをもつ。中間テーブルを介して可視化手法と多対多のリレーションをもつ。

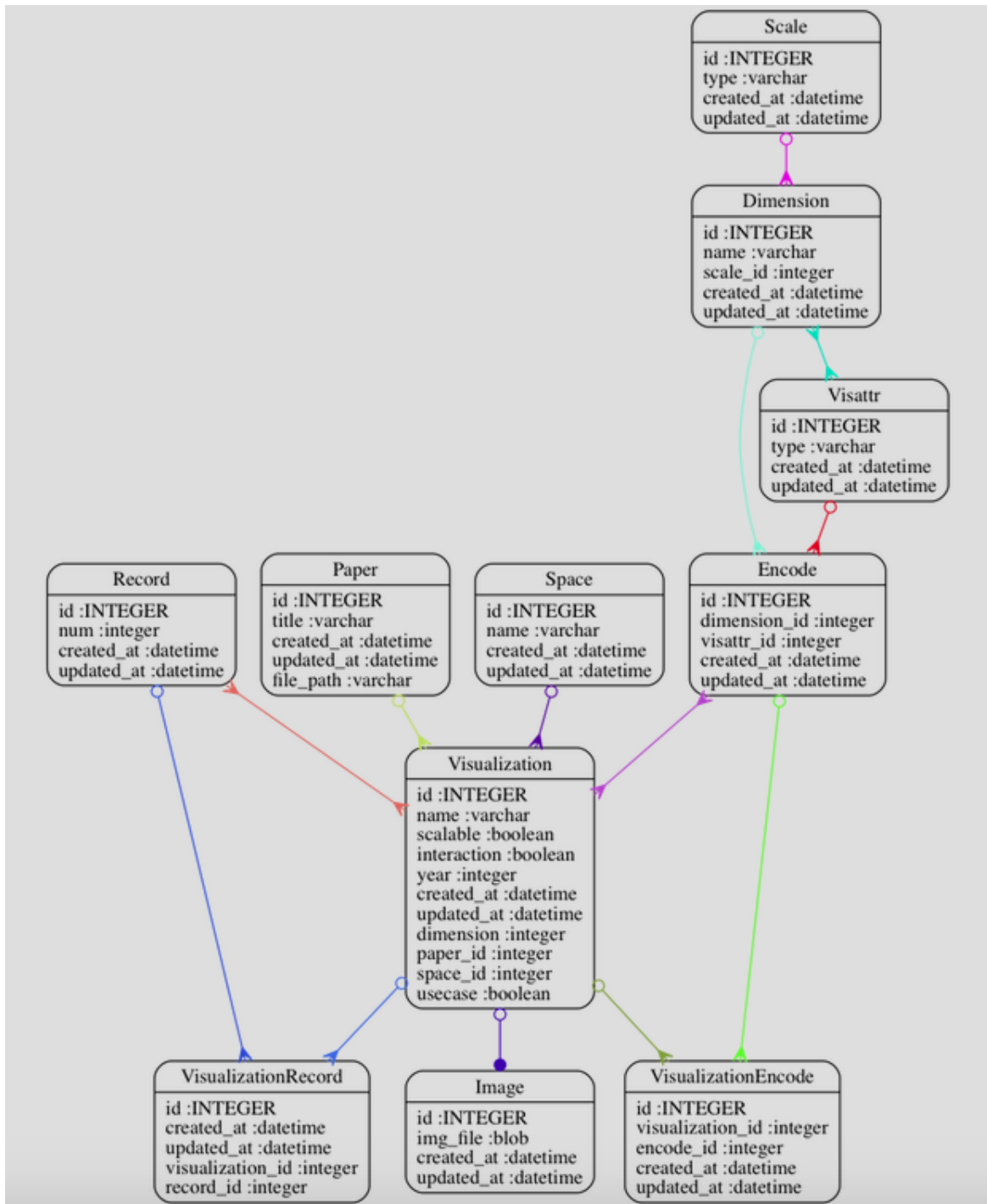


図 5.1: 拡張型の集約的可視化手法に関する ER 図

## Dimension モデル

次元(可視化する情報)名, 尺度水準 ID をもつ. エンコードを介して視覚変数と多対多のリレーションをもつ.

## Visattr モデル

視覚変数名をもつ. エンコードを介して次元と多対多のリレーションをもつ.

## Scale モデル

尺度水準名をもつ. 1つの尺度水準に対して複数の次元をもち得る.

## 5.2 検索インタフェース

本節では検索システムのインタフェースについて説明する. 検索システムの概観を図 5.2 に示す.

Extended Aggregate Visualization DB

データレコード数:

スケラブル

次元数

可視化する情報1

情報の尺度水準1

使用する視覚変数1

可視化する情報2

情報の尺度水準2

使用する視覚変数2

可視化する情報3

情報の尺度水準3

使用する視覚変数3

可視化する情報4

情報の尺度水準4

使用する視覚変数4

可視化する情報5

情報の尺度水準5

使用する視覚変数5

表現空間

インタラクション  有  無

30件の拡張型の集約的可視化手法

手法名: Aggregate Towers  
発表年: 1999  
ユースケース: なし  
スケラビリティ: -  
表現空間: 地図  
インタラクション: 有  
次元数: 3  
次元1: レコード数 (比例尺度)  
視覚変数: 体積  
次元2: 任務タイプ (名義尺度)  
視覚変数: 色(色相)  
次元3: 位置(緯度経度) (間隔尺度)  
視覚変数: 位置  
論文リンク: [Aggregate Towers: Scale Sensitive Visualization and Decluttering of Geospatial Data](#)

手法名: ThemeRiver  
発表年: 2000  
ユースケース: なし  
スケラビリティ: -  
表現空間: 2次元平面  
インタラクション: 無  
次元数: 3  
次元1: トピック(イベント) (名義尺度)  
視覚変数: 色(色相)  
次元2: 時刻 (間隔尺度)  
視覚変数: 位置  
次元3: レコード数 (比例尺度)  
視覚変数: 長さ(幅)  
論文リンク: [ThemeRiver: Visualizing Theme Changes over Time](#)

手法名: Interactive Sankey Diagrams  
発表年: 2005  
ユースケース: なし  
スケラビリティ: -

図 5.2: 拡張型の集約的可視化手法検索システムの概観

利用方法は, まず, 画面左側に並ぶテキストボックスやドロップダウンリスト等のフォーム群により手法の検索条件を指定する. ドロップダウンリストの選択肢はデータベースのレ



コードから自動生成されており，自由入力のデータレコード数や次元数，チェックボックスやラジオボタンにより送信される値によってデータベース内のレコードを検索する．

検索の結果はフォーム群の右側の領域に提示される．検索条件に一致した可視化手法群はそれぞれ DIV 要素で描画される．各可視化手法の DIV 要素には，検索条件に対応する情報の他，画像や手法名，発表名，論文ダウンロードリンクが配置されている．

## 第6章 ユーザスタディ

本研究の提案である拡張型の集約的可視化手法の分類および検索システムの有用性や課題を調査するため、提案システムを用いたユーザスタディを実施した。ユーザスタディの被験者は平均年齢24歳の合計4名であり、男性3人と女性1人、情報可視化を専門とする学生2人と特に情報可視化を専門としない学生2名という内訳である。

### 6.1 使用したデータ

ユーザスタディのタスクは予め準備したデータを用いて行った。使用したデータは、総務省公開「ブロードバンドサービス等の契約数の推移データ (<http://www.soumu.go.jp/johotsusintokei/field/csv/gt010103.csv>, 2016年1月12日閲覧)」を使用した。このデータのカラム構造は表6.1のとおりである。各ブロードバンドサービス名のカラムはその契約者数が格納されている。

表 6.1: ユーザスタディに使用したデータのカラム構造

西暦年	四半期	都道府県名	FTTH	DSL	CATV	FWA	BWA	LTE
string	string	string	integer	integer	integer	integer	integer	integer

### 6.2 拡張型の集約的可視化手法検索タスク

拡張型の集約的可視化手法検索システムを用いたタスクとして、前節で紹介したデータに適用可能な拡張型の集約的可視化手法を検索してもらうというタスクを実施した。タスクは以下の要領で実施した。

- データを自由に閲覧する (編集は行わない)。
- 適用できる可能性がある可視化手法を検索できるよう、検索条件を指定する。
- 検索は何度でもやり直し可能。
- 適用できそうな可視化手法が見つかった、或いは見つかりそうになく諦めた場合にタスク終了を宣言する。

ユーザが抽出しようとする可視化手法のイメージについて指示は行わなかったため、データからの次元の選び方などもユーザに一任した。タスクの結果としては、ユーザが適用できると判断した可視化手法を記録した。

### 6.3 アンケート調査

前節のタスクを実施後、拡張型の集約的な可視化手法の分類および検索システムの使用感を調査するため、質問紙によるアンケート調査を実施した。アンケートは情報可視化に対する意識に関する質問およびユーザスタディのタスクに関する質問の大きく2種類で構成され、それぞれに5段階のリッカート尺度で回答する質問が3問ずつ含まれる。情報可視化に対する意識に関する質問の意図としては、検索条件として使用する、尺度水準などあまり一般的ではない情報を理解するために、情報可視化に対するモチベーションが影響するのではないかと考えたからである。リッカート尺度で回答する質問リストを以下に示す。

1. データを可視化したことがある
2. データを可視化することは難しいと思う
3. 進んでデータを可視化したい
4. 検索システムは使いやすかった
5. 検索条件を理解できていた
6. 検索結果の項目を理解できた

また、ユーザスタディのタスクに関する質問には、検索条件と検索結果に含まれる情報についてそれぞれ重要だと感じたもの上位3つとその順位を回答とする質問を含んだ。

### 6.4 拡張型の集約的可視化手法検索タスクの結果および考察

すべてのユーザが、データに適用できる拡張型の集約的な可視化手法が見つかったと判断した。拡張型の集約的可視化手法検索タスクの結果を表 6.2 に示す。

表 6.2: 拡張型の集約的可視化手法検索タスクの結果

ユーザ ID	拡張型の集約的可視化手法 ID
1	2, 18, 24, 26
2	19
3	1, 10
4	10

ユーザ 2, 3, 4 の抽出した手法がいずれも地図上にマーカを描くタイプであり，データに地理情報を含むカラムが存在する場合，可視化に利用したいと考える人が多い傾向にあるのかもしれない．また，ユーザ 1 が抽出した手法は地図を用いないものではあるものの，そのいずれも横方向に時間軸をとり，色分けされたトピックなどの変動を描いているものであった．このことから，データの特徴を出発点に可視化手法を検索した場合にも，見た目が近い複数の手法を一度に検索できる可能性があるのではないだろうか．

同じデータを出発点に異なる傾向をもつ可視化手法を適用可能だと判断したということで，異なる見た目の系の可視化手法が同じデータセットを対象にできる可能性があり，これは本研究の目的の一つである拡張型の集約的な可視化手法の積極的な利活用に寄与するのではないかと考えた．但し，今回のユーザスタディでは可視化する次元を指定しなかったため，この部分の選択によって可視化イメージが大きく異なった可能性もあり，今後検証の必要があるかもしれない．

## 6.5 アンケート調査の結果および考察

リッカート尺度で回答を得た質問の調査結果を以下の表 6.3 に示す．質問の内容に最も同調的な回答値を 5，反対の回答を 1 としている．

表 6.3: リッカート尺度のアンケート回答結果

ユーザ ID	質問 1	質問 2	質問 3	質問 4	質問 5	質問 6
1	5	4	4	3	4	4
2	1	5	3	3	4	3
3	5	4	5	2	5	5
4	1	4	4	4	3	4
中央値	3	4	4	3	4	4

質問 1 の可視化経験以外の質問に関してはほとんど差が出なかった．即ち，可視化に対する意識に関する質問を設けた際に想定していた，可視化へのモチベーションの差は今回は生じていないと考えられる．質問 2 および質問 3 が概ね 4 以上の回答であることから，ユーザの情報可視化に対する意識は“可視化することは難しいが利用していくべきである”とする傾向にあるのかもしれない．

質問 4 の結果より，検索システムの使いやすさは良くも悪くもないといえそうである．自由記述による理由によれば，「どの項目がどの結果に影響したのか分からないので試しづらい」といったシステムのフィードバックの問題や，「可視化に慣れていない人は何の情報を検索条件にすれば良いのかわかりにくい」といった事前知識の差による問題があった．特に後者の問題については，質問 5 において「可視化のことを知らなかったため，初めて聞く単語ばかりだった」という意見があり，検索条件を多くの人に理解してもらえるような仕組みが必要だと感じた．

次に、検索条件のうち重要だと感じたもの上位3つとその順位の回答結果を表6.4に示す。

表 6.4: 検索条件のうち重要だと感じたもの上位3つとその順位の回答結果

	1位	2位	3位
データレコード数			1
スケーラビリティ			
次元数			
可視化する情報	2		
情報の尺度水準	1	1	2
使用する視覚変数	1	1	
表現空間		2	1
インタラクション			

1位を最多の2票獲得した“可視化する情報”および、その“尺度水準”にも票が集まっており、可視化したい情報そのものや似た性質の情報を扱っている可視化手法を検索したいという要求は存在すると考えられる。特に尺度水準での検索は「尺度水準が同じなら、ある程度応用できそう」という意見が得られ、本研究で尺度水準を採用した意図どおりに使用してもらえたユーザも存在した。

また、データレコード数やスケーラビリティが重要だと感じられなかったのは、分類結果に対する考察でも述べたように、これらを保持している可視化手法のレコードが少なかったため、指定すると検索結果が絞りこまれ過ぎてしまう影響もあるのではないかと考える。

次に、検索結果に含まれる情報のうち重要だと感じたもの上位3つとその順位の回答結果を表6.5に示す。

検索結果に含まれる情報のうち最も重要視されたのは可視化手法の画像であった。また、視覚変数も1位1票と3位1票を獲得しており、検索結果は可視化手法の見た目で把握したいという要求があると考えられる。

検索条件と検索結果で重要視される情報をまとめると、可視化手法を検索する際には手元のデータがもつ情報から可視化手法を検索し、検索結果は見た目で把握できることが求められているといえる。したがって、検索結果の見た目のどこに検索条件の情報が対応しているかを効果的に提示できれば、よりスムーズな可視化手法検索の仕組みが作られるのではないかと考える。

表 6.5: 検索結果のうち重要だと感じたもの上位3つとその順位の回答結果

	1位	2位	3位
画像	3	1	
手法名		1	
発表年			
ユースケース (データレコード数)			
スケーラビリティ			
表現空間			
インタラクション			
次元数			
可視化されている情報 (次元)		1	2
情報の尺度水準		1	
使用されている視覚変数	1		1
論文ファイル			1

## 第7章 結論

本研究は、大規模/多次元データの流行とそれらに対応可能な可視化手法である“拡張型の集約的可視化手法”に注目し、より幅広い利活用へと繋げる仕組みづくり目的として取り組んだものである。特に、従来の可視化手法分類・検索の方法はデータ起点で可視化する流れを考えた場合に利用しづらいと感じ、この問題解決に取り組んだ。そのためのアプローチとして、拡張型の集約的可視化手法の特徴をデータの特徴と結びつくように分類することを試みた。

6つの分類項目を検討し、実際に InfoVis20 年分の採択論文に含まれる拡張型の集約的可視化手法を分類した。すると、拡張型の集約的可視化手法は、可視化する情報をその尺度水準に適した視覚変数によって表現するように設計されているであろうことが推察できる結果が表れた。また、扱えるデータレコード数やスケーラビリティについて言及している手法が少なく、大規模データからそのデータレコード数を用いて拡張型の集約的可視化手法の検索を行うことの難しさが予想された。

上記のような難しさも予想されたが、実際の分類結果を用いて拡張型の集約的可視化手法を検索するシステムを構築し、ユーザスタディによって有用性や課題を確認した。ユーザスタディの結果、同一のデータからタイプの異なる拡張型の集約的可視化手法群を抽出した例が見られ、そのタスクの自由度の高さ故に可能性としかいえないが、本研究の目的である“幅広い利活用”へと繋がる期待を感じることができた。

また、可視化手法検索時の検索条件としては元データの特徴が重要視される可能性が推測されたことは、本研究のアプローチの一定の正しさを示すものであると考える。検索結果に対しては見た目で把握できることが重要視されるという仮設と合わせて、今後の可視化手法検索に関わる研究の方向性として、検索に使われた元データの情報をビジュアルな検索結果から読み取れるような仕組みを形成していくことを期待する。

## 謝辞

本研究を行うにあたり，指導教員である三末和男教授には，丁寧なご指導と適切なお助言を賜りました。幾度と無くご迷惑やご心配をお掛けした不肖な学生でしたが，本論文を書き上げるに至ることができたのは，偏に先生の手厚いフォローの賜物です。心より感謝申し上げます。

また，同期をはじめとする VisLab 学生の皆さまには，ゼミはもとより日々の研究生生活を豊かで実りあるものにしていただきました。大変お世話になりました。

研究室外でも私の学生生活を支えてくださった，数々の先輩方，友人たち，後輩たちにも，感謝の念が尽きません。今日まで私を支えてくださった皆さまのご厚意に改めて御礼申し上げます。

最後に，各々が遠く離れ離れになりながらも，強い絆で常に支え合ってきた家族に，深く，深く御礼申し上げます。ありがとうございました。



## 参考文献

- [1] Ben Shneiderman. Extreme visualization: squeezing a billion records into a million pixels. In *proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pp. 3–12. ACM, 2008.
- [2] Robert L Harris. *Information graphics: A comprehensive illustrated reference*. Oxford University Press, 1996.
- [3] Hans-Jorg Schulz. treevis.net - a visual bibliography of tree visualization 2.0. <http://vcg.informatik.uni-rostock.de/~hs162/treeposter/poster.html>, 2016年1月12日最終閱覽.
- [4] Stanley Smith Stevens. *On the theory of scales of measurement*, 1946.
- [5] James K Rayson. Aggregate towers: Scale sensitive visualization and decluttering of geospatial data. In *proceedings of IEEE Symposium on Information Visualization 1999 (Info Vis '99)*, pp. 92–99. IEEE, 1999.
- [6] Beth Hetzler Susan Havre and Lucy Nowell. Themeriver: Visualizing theme changes over time. In *proceedings of IEEE Symposium on Information Visualization 2000 (InfoVis '00)*, pp. 115–123. IEEE, 2000.
- [7] Manfred Hanfler Patrick Riehmann and Bernd Froehlich. Interactive sankey diagrams. In *IEEE Symposium on Information Visualization 2005 (InfoVis '05)*, pp. 233–240. IEEE, 2005.
- [8] Robert Kosara Fabian Bendi and Helwig Hauser. Parallel sets: Visual analysis of categorical data. In *IEEE Symposium on Information Visualization 2005 (InfoVis '05)*, pp. 133–140. IEEE, 2005.
- [9] Frank Van Ham James Abello and Neeraj Krishnan. Ask-graphview: A large scale graph visualization system. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 12, No. 5, pp. 669–676, 2006.
- [10] Adam Perer and Ben Shneiderman. Balancing systematic and flexible exploration of social networks. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 12, No. 5, pp. 693–700, 2006.

- [11] Martin Graham and Jessie Kennedy. Exploring multiple trees through dag representations. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 13, No. 6, pp. 1294–1301, 2007.
- [12] Martin Wattenberg Frank Van Ham and Fernanda B Viégas. Mapping text with phrase nets. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15, No. 6, pp. 1169–1176, 2009.
- [13] Martin Wattenberg Fernanda B Viegas and Jonathan Feinberg. Participatory visualization with wordle. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15, No. 6, pp. 1137–1144, 2009.
- [14] Diansheng Guo. Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15, No. 6, pp. 1041–1048, 2009.
- [15] Edward Grundy Mark W Jones Robert S Laramee Jorik Blaas, Charl P Botha and Frits H Post. Smooth graphs for visual exploration of higher-order state transitions. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15, No. 6, pp. 969–976, 2009.
- [16] Shixia Liu Norman Au Weiwei Cui Hong Zhou Yingcai Wu, Furu Wei and Huamin Qu. Opinionseer: interactive visualization of hotel customer feedback. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 16, No. 6, pp. 1109–1118, 2010.
- [17] Bettina Speckmann and Kevin Verbeek. Necklace maps. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 16, No. 6, pp. 881–889, 2010.
- [18] Robert S Laramee Jonathan C Roberts Zhao Geng, ZhenMin Peng and Rick Walker. Angular histograms: Frequency-based visualizations for large, high dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 17, No. 12, pp. 2572–2580, 2011.
- [19] Liang Gou and Xiaolong Zhang. Treenetviz: Revealing patterns of networks over tree structures. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 17, No. 12, pp. 2449–2458, 2011.
- [20] Shixia Liu Panpan Xu Wei Chen Conglei Shi, Weiwei Cui and Huamin Qu. Rankexplorer: Visualization of ranking changes in large time series data. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18, No. 12, pp. 2669–2678, 2012.
- [21] Krist Wongsuphasawat and David Gotz. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18, No. 12, pp. 2659–2668, 2012.

- [22] Gennady Andrienko Christian Tominski, Heidrun Schumann and Natalia Andrienko. Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18, No. 12, pp. 2565–2574, 2012.
- [23] Romain Vuillemot Charles Perin and Jean-Daniel Fekete. Soccerstories: A kick-off for visual soccer analysis. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 12, pp. 2506–2515, 2013.
- [24] Silvia Miksch Bilal Alsallakh, Wolfgang Aigner and Helwig Hauser. Radial sets: Interactive visual analysis of large overlapping sets. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 12, pp. 2496–2505, 2013.
- [25] Thomas Reinhardt Corinna Vehlow and Daniel Weiskopf. Visualizing fuzzy overlapping communities in networks. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 12, pp. 2486–2495, 2013.
- [26] Madelaine Boyd Peter Macko Krzysztof Z Gajos Margo Seltzer Hanspeter Pfister Michelle Borkin, Chelsea S Yeh. Evaluation of filesystem provenance visualization tools. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 12, pp. 2476–2485, 2013.
- [27] Enxun Wei Mengchen Liu Shixia Liu, Yingcai Wu and Yang Liu. Storyflow: Tracking the evolution of stories. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 12, pp. 2436–2445, 2013.
- [28] Heike Hofmann and Marie Vendettuoli. Common angle plots as perception-true visualizations of categorical associations. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 19, No. 12, pp. 2297–2305, 2013.
- [29] Zhuofeng Wu Weiwei Cui, Shixia Liu and Hao Wei. How hierarchical topics evolve in large text corpora. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20, No. 12, pp. 2281–2290, 2014.
- [30] Jonghun Park Bohyoung Kim Jaemin Jo, Jaeseok Huh and Jinwook Seo. Livegantt: Interactively visualizing a large manufacturing schedule. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20, No. 12, pp. 2329–2338, 2014.
- [31] Stef van den Elzen and Jarke J van Wijk. Multivariate network exploration and presentation: From detail to overview via selections and aggregations. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20, No. 12, pp. 2310–2319, 2014.
- [32] Diansheng Guo and Xi Zhu. Origin-destination flow data smoothing and mapping. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20, No. 12, pp. 2043–2052, 2014.

- [33] Jock Mackinlay. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics (Tog)*, Vol. 5, No. 2, pp. 110–141, 1986.