

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

多エンティティ多変量時系列データの 視覚的分析ツールの開発

包 涵

(コンピュータサイエンス専攻)

指導教員 三末和男

2011 年 3 月

概要

多エンティティ多変量時系列データは、複数のデータエンティティにおいて多変量時系列データを持つ。多変量時系列データとは、時間とともに変化する値を複数持つデータである。例えば、学生の成績データでは、一人の学生が複数の科目に対して成績を持ち、成績が学期ごとに変化する。ここで、我々は、一人の学生を一つのデータエンティティと呼ぶ。現実世界に学生の成績データや個人の購買データなど様々な多エンティティ多変量時系列データが存在する。データの過去の特徴を把握し、育成方針の作成や商売戦略の考案などのために、これらのデータを分析する必要がある。

多エンティティ多変量時系列データでは、エンティティに着目した分析がよく行なわれる。例えば、教師が教育方針を考える時に、成績が悪い学生に着目して分析を行う。しかし、従来の多エンティティ多変量時系列データを分析対象とした研究では、このような分析を支援していない。

本研究の目的は、エンティティに着目した多エンティティ多変量時系列データの視覚的分析を支援することである。

本研究では、エンティティに着目した多エンティティ多変量時系列データの視覚的な分析に焦点を合わせ、それに対しての分析要件を整理した。要件を満たすために、散布図と **small multiples** の組み合わせによりデータエンティティ全体の一覧と個々のデータエンティティの詳細情報を同時に把握できる可視化手法を開発した。そして、多変量時系列データを視覚的に表現できるように、既存の可視化手法を拡張した。さらに、提案した可視化手法をインタラクティブに分析する上で必要な操作を開発した。最後に、提案した可視化手法と操作を備えた多エンティティ多変量時系列データの視覚的分析ツール“MMTAnalyzer”を開発した。

本ツールの有用性を確かめるために、ケーススタディを行った。ケーススタディでは実際の購買データの分析を行ってもらった。ケーススタディによって、“MMTAnalyzer”を用いた多エンティティ多変量時系列データに対しての分析が有効であることが分かった。

目次

第1章	序論	1
1.1	多エンティティ多変量時系列データとは	1
1.2	多エンティティ多変量時系列データの分析	1
1.3	多エンティティ多変量時系列データの視覚的分析における問題	2
1.4	本研究の目的	3
1.5	本研究の貢献	3
1.6	本論文の構成	3
第2章	関連研究	4
2.1	単一多変量時系列エンティティの可視化研究	4
2.1.1	静的な多変量データの可視化から発展した研究	4
2.1.2	単一変量の時系列データの可視化から発展した研究	5
2.1.3	他の単一多変量時系列エンティティの研究	5
2.2	多エンティティ多変量時系列データの可視化研究	5
第3章	多エンティティ多変量時系列データの分析要件	7
第4章	多エンティティ多変量時系列データのための視覚的表現	8
4.1	視覚的表現の設計	8
4.2	概観の視覚的表現	9
4.2.1	散布図により全体データエンティティの分布を表現	9
4.2.2	Small multiplesにより各データエンティティの詳細情報を表現	10
4.3	詳細の視覚的表現	10
4.3.1	Line graphの視覚的表現	11
4.3.2	Stack graphの視覚的表現	12
4.3.3	Radar chartの視覚的表現	13
4.3.4	Sector graphの視覚的表現	15
4.4	視覚的表現要素一色の使用	17
4.4.1	情報の表現による色の使い分け	17
4.4.2	階層構造を持つ変数の色付け	17
4.5	視覚的分析の流れ	18
第5章	多エンティティ多変量時系列データの視覚的分析ツールの開発	19
5.1	ツールの設計	19
5.2	ツールのインタフェース	20
5.2.1	散布図パネル	20
5.2.2	Small multiples パネル	22
5.2.3	変数と時間の指定パネル	23
5.2.4	調整パネル	23
5.3	ツールの機能	24
5.3.1	平均値基準選択機能	24
5.3.2	手描き変数選択機能	25
5.3.3	エンティティフィルター機能	26

5.3.4	エンティティ集合分析機能	28
5.3.5	属性によりクラスタ分析機能	29
5.3.6	単一エンティティ分析機能	30
5.4	実装	31
5.4.1	実装言語	31
5.4.2	データの読み込み	31
第 6 章	ツールの利用例	32
第 7 章	ケーススタディ	36
7.1	データ	36
7.2	ケーススタディ設計	37
7.3	ユーザの分析	37
7.4	考察	44
7.4.1	ツールの有用性	44
7.4.2	視覚的表現及び機能の有用性	44
7.4.3	散布図と small multiples の組み合わせの有用性	45
第 8 章	結論と今後の課題	47
8.1	結論	47
8.2	今後の課題	47
8.2.1	ツールの操作とインタフェースの改善	47
8.2.2	大規模データへの対応	48
謝辞		49
参考文献		50

図目次

図 1-1	多エンティティ多変量時系列データの例	1
図 4-1	視覚的表現の設計	8
図 4-2	散布図の視覚的表現	9
図 4-3	radar chart の表示問題	11
図 4-4	line graph の基本表現	11
図 4-5	stack graph の基本表現	12
図 4-6	stack graph の縦軸のスケールの変化	12
図 4-7	stack graph のインタラクション操作	13
図 4-8	radar chart の基本表現	13
図 4-9	radar chart の軸の並びの変化	14
図 4-10	radar chart の色表現	14
図 4-11	radar chart の時系列情報の視覚的表現	15
図 4-12	sector graph の基本表現	15
図 4-13	sector graph の時系列情報の視覚的表現	16
図 4-14	多変量の階層構造の色付け	17
図 4-15	視覚的分析の流れ	18
図 5-1	散布図と small multiples でのハイライト表示	19
図 5-2	MMTAnalyzer のインタフェース	20
図 5-3	散布図パネル	21
図 5-4	散布図パネルでの選択操作	21
図 5-5	散布図パネルのサイズの調整	21
図 5-6	small multiples パネル	22
図 5-7	small multiples の表示スペースの調整	22
図 5-8	変量と時間の指定パネル	23
図 5-9	調整パネル	23
図 5-10	平均値基準選択	24
図 5-11	8 種類のデータエンティティ	25
図 5-12	手描き変量選択機能	26
図 5-13	エンティティフィルター機能	27
図 5-14	エンティティ集合の色付け	28
図 5-15	エンティティ集合の並び替え	29
図 5-16	属性によりクラスタ分析機能	30
図 5-17	単一エンティティ分析機能	31
図 6-1	散布図を用いた分析	32
図 6-2	radar chart を用いた分析	33
図 6-3	一つの変量に着目した分析	33
図 6-4	単一エンティティ分析	34
図 6-5	時間フィルターしてからの分析	35
図 7-1	ケーススタディデータのオーバービュー	38

図 7-2	買い物数が多い人の分析	39
図 7-3	部屋ごとの比較分析	40
図 7-4	食べ物の売れ行きに着目した分析	41
図 7-5	増加した人に着目した分析	42
図 7-6	radar chart を使った分析	42
図 7-7	月ごとに着目した分析	43
図 7-8	スクリーンショットを取る前の操作の回数	45

表目次

表 1-1	多エンティティ多変量時系列データのリストの表の例	2
表 4-1	詳細の視覚的表現が分析要件との対応	11
表 7-1	生データ	36
表 7-2	整理した後のデータ	37
表 7-3	詳細情報の視覚的表現への切り替えの回数	45

第1章 序論

1.1 多エンティティ多変量時系列データとは

各時刻において二つ以上の値を持つデータエンティティが多変量時系列データエンティティと呼ばれる。本研究では、データエンティティが複数存在する多変量時系列データを多エンティティ多変量時系列データと呼ぶ。

図 1-1 に示した学生の成績データを具体的な例として説明する。各学生について、複数の科目があり、これは多変量である。また、各科目に対して成績があり、成績が学期ごとに変化する。これは時系列である。一人の学生を一つのデータエンティティとした場合、一人の学生の成績データは多変量時系列データエンティティをなし、複数の学生の成績データの場合は多エンティティ多変量時系列データをなす。

データの特徴により、大まかに質的データ（四則演算ができないデータ）と量的データ（四則演算ができるデータ）二種類に分類することができる。本研究では、各変量の値を量的データとして扱う。

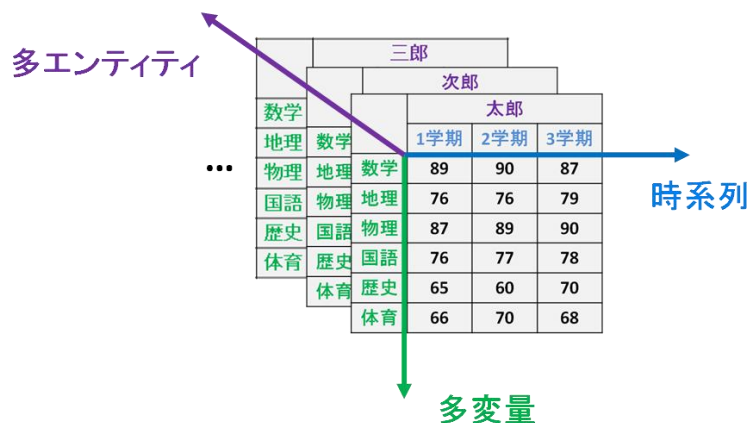


図 1-1 多エンティティ多変量時系列データの例

1.2 多エンティティ多変量時系列データの分析

現実世界にスポーツ選手の能力データや学生の成績データ、個人の購買データなど様々な多エンティティ多変量時系列データが存在する。データの過去の特徴を把握し、次の育成方針を作成したり、商売戦略を考えたりするために、これらのデータを分析する必要がある。

多エンティティ多変量時系列データを分析する際に、エンティティに着目した分析がよく行われる。1.1 節で説明した代表的な多エンティティ多変量時系列データの例である学生の成績データを挙げて、多エンティティ多変量時系列データの実際の分析について説明する。

クラス担任の先生が学生の成績を上げるために教育方針を立てる際に、以下のようなアプローチが考えられる。まず、全学生の中から優秀な学生を見つけ、これらの優秀な学生の成績を個別に分析する上で何かのヒントを得て、次の教育方針を作る。或いは、全学生の中から成績が悪い学生を見つけ、これらの学生の成績を分析する上で、各個人の苦手な部分に対して、個別な教育方針を作る。このように、学生の成績データに対しては、学生に着目した分析がよく行なわれる。

また、もう一つの多エンティティ多変量時系列データの代表例として、購買データの分析について説明する。購買データは、商品のカテゴリや商品が多変量である。各商品カテゴリや商品に対する購入数は時刻に従って変化する。これは時系列である。また、複数の購入者がいるので、このようなデータは多エンティティ多変量時系列データである。マーケティングリサーチ部門の担当者が次の商売戦略を立てる際に、次のアプローチがよく考えられる。まず、全購入者の購買履歴からよく買っている購入者を見つけ、次に、これらの購入者の履歴に対して詳しい分析を行う。そして、分析によって得られた商品の購買パターンを基に販売向上に繋がる商売戦略を立てる。個人購買データに対して、このような特定の購入者に着目した分析がよく行なわれる。

以上のような、特定の特徴を持つエンティティに対する分析がよく行われる。

1.3 多エンティティ多変量時系列データの視覚的分析における問題

多エンティティ多変量時系列データはいくつの表現形式が考えられるが、一般的には表 1-1 のようなリスト形式の表で現れる。このようなリスト形式の表に保存されたデータの値からデータの傾向を分析することは不可能ではないが、数値で埋め尽くされた表からそのデータの傾向を発見することは困難である。データを人間が直感的に分かりやすい形に視覚的に表現して分析するのは一つの有効的な手段だと考えられる[1]。

表 1-1 多エンティティ多変量時系列データのリストの表の例

	太郎			次郎			三郎			
	1学期	2学期	3学期	1学期	2学期	3学期	1学期	2学期	3学期	
数学	73	44	64	72	80	76	82	41	52	
物理	43	91	49	49	49	86	50	93	42	
地理	59	75	66	88	95	89	64	92	63	
国語	70	42	59	59	99	42	50	42	54	
歴史	66	97	81	42	86	75	85	64	51	
体育	46	44	93	81	79	61	99	44	56	

多エンティティ多変量時系列データを可視化することで、有益な情報の発見を支援する研究はこれまで数多く行なわれてきた。しかし、これらの研究はエンティティに着目した分析を支援することができなかった。

気候データ、脳電波データ（EEG データ）や患者の診断履歴データなどの多エンティティ

多変量時系列データの分析を通じて、各変量間の相関関係が時間とともにどう変化するかを明らかにする研究が多く行なわれたが[2][3][4][5]、これらの研究は各エンティティに着目した多変量時系列情報についての分析ができない。また、複数の多変量データを可視化する手法 **parallel coordinates** を拡張し、各変量の値の時間変化を視覚的に表現する研究も行なわれた[6]。しかし、この手法では各エンティティの変量の具体的な変化過程についての分析ができない。

本研究では、エンティティに着目した多エンティティ多変量時系列データの視覚的な分析を支援する視覚的表現及びツールを開発する。

1.4 本研究の目的

本研究では、エンティティに着目した多エンティティ多変量時系列データの分析を支援することを目的とし、それを支援する視覚的分析ツールを開発する。

1.5 本研究の貢献

本研究の貢献は以下の4点であると考えている。

一つ目はエンティティに着目した多エンティティ多変量時系列データの分析問題に着目し、それに対しての分析要件を整理したことである。

二つ目は散布図と **small multiples** を組み合わせた視覚的な表現手法を開発したことである。本手法により、多エンティティ多変量時系列データの全体データエンティティの一覧と各データエンティティの詳細情報を同時に分析することが可能になった。

三つ目は分析要件を満たすために、**radar chart** と **sector graph** の既存の視覚的な表現手法を拡張したことである。これにより、従来の静的な多変量データの可視化手法 **radar chart** と **sector graph** が時系列情報を視覚的に表現できるようになった。

四つ目は本研究で開発した視覚的な表現手法とインタラクティブな操作を用いて、多エンティティ多変量時系列データの視覚的分析ツールを開発したことである。本ツールにより、エンティティに着目した多エンティティ多変量時系列データの視覚的な分析が可能となった。

1.6 本論文の構成

本論文の以降の部分は以下のように構成される。

第2章で本研究に関連する研究を挙げ、本研究の位置付けを紹介する。第3章で多エンティティ多変量時系列データの分析要件をまとめる。第4章で多エンティティ多変量時系列データの分析要件に対する視覚的な表現について紹介する。第5章にて、多エンティティ多変量時系列データの視覚的分析ツール“**MMTAnalyzer**”について紹介する。第6章で利用例として“**MMTAnalyzer**”を用いた学生の成績データの分析を紹介する。第7章にて、実データのケーススタディを通じて、本研究で開発した“**MMTAnalyzer**”の有効性を示す。第8章が結論と今後の課題について述べる。

第2章 関連研究

2000 年前後は単一多変量時系列データエンティティに対しての可視化研究が数多く行われたが、多エンティティを対象としていなかった。近年、多エンティティ多変量時系列データに対しての可視化研究が盛んに行われている。ここで、まず単一多変量時系列データの可視化研究を紹介し、その次は最近の多エンティティ多変量時系列データの可視化研究を紹介する。

2.1 単一多変量時系列エンティティの可視化研究

多変量データと時系列データそれぞれに関する可視化研究がすでに大量に行われていた。ここで、単一多変量時系列データエンティティに関しての可視化研究を三つのカテゴリを分けて紹介する。2.1.1 節では、静的な多変量データの可視化をベースにして、時系列情報を視覚的に表現できるようにした多変量時系列データの研究について述べ、2.1.2 節では、時系列データの可視化をベースにして、多変量の情報を視覚的に表現できるようにした多変量時系列データの研究について述べる。最後に 2.1.3 節ではそのほかの多変量時系列データの研究について述べる。

2.1.1 静的な多変量データの可視化から発展した研究

一つの時刻の静的な多変量データに対して、2D and 3D scatter plots, Matrix of scatter plots, heat maps, radar graph, sector graph, parallel coordinates などのさまざまな可視化手法が提案された[7]。

parallel coordinates[8]は有効な静的多変量データの可視化技術であり、多くの研究がこの技術を拡張して時系列情報を視覚的に表現できるようにしている[5][6][9][10]。parallel coordinates の軸を繋ぐ線の代わりにポリゴンを使って値の時間変化を表現できるようにした可視化手法が提案された[6]、しかし、この手法では、値の変化範囲が分かるが、値の具体的な変化が分からない。さらに、データエンティティが多い場合にポリゴンが重なってしまい、各エンティティに対しての具体的な分析を行いにくいという問題が残されている。それ以外にも、時間軸をグラフの中心に置き、その周りに各変量を意味する軸を置く。各変量の値とその値と対応する時刻を線で繋げるにより多変量時系列データを視覚的に表現する手法も提案された[9]。この手法をさらに立体的な 3 次元空間に発展した手法も提案された[10]。この二つの手法は変量の値が多く、どの時間になど値になっているかが分かるが、値の時間においての変化過程が分からない、また、データエンティティが複数の場合の支持が言及されていない。

2.1.2 単一変量の時系列データの可視化から発展した研究

多変量データの可視化手法以外に単一変量の時系列データの可視化手法も数多く提案された[11][12][13]。折れ線グラフは時系列データを視覚的に表現する時に最適な手法と考えられる。多変量時系列データの場合には、複数の線により多変量を表せば、自然に多変量への対応もできる。しかし、同時に表示する変量の数が増加するにつれ、線が重なり合い、分析の妨げとなる。この問題に対して、線の代わりに、各変量を帯として表し、さらに、底から各帯を上から順に積み上げるスタックグラフが提案された。各変量の変化を重ならずに表示しつつ、値の和の変化も同時に把握できるという利点がある。それをベースにした ThemeRiver[14]は時間の流れの川のメタファーを用いて多変量時系列データを可視化した。

他にも、時間軸を螺旋状に表示して、変量の性質によって値を色や点などを使って螺旋に沿って描画することによってデータの周期的なパターンを発見しやすくする可視化手法も提案された[15]。

しかし、これらの手法はいずれも単一のデータエンティティのみ扱う。

2.1.3 他の単一多変量時系列エンティティの研究

以上の研究以外にも、鉛筆のメタファーを用いて、一つの“鉛筆”が一つデータエンティティを表し、“鉛筆”の各側面が各変量の時間変化を表す多変量時系列データの可視化手法が提案された[16]。“鉛筆”の異なる側面により、連続的な値と離散的な値を同時に表現できるという利点がある。

MultiCombs[9]はもう一つの多変量時系列データの可視化手法である。正多角形の各辺が各変量を意味する。各辺と垂直する放射上方向は時間軸を意味する。各変量の時間変化はこの時間軸により表す。

以上の可視化手法は一つの多変量時系列データエンティティに対しての分析が有用であるが、複数のデータエンティティの場合への対応が言及されていない。

2.2 多エンティティ多変量時系列データの可視化研究

Table Lens[17]はテーブルデータのスプレッドシートの各セルの値をグラフに切り替えることによりテーブルデータの分析をより直感的に行えるようにした。しかし、全体データのレイアウトはスプレッドシートのテーブル形式に制限され、データエンティティ間の関係を自由に分析することができない。

また、parallel coordinates、minmax plot と density map を組み合わせることにより脳電波データ（EEG データ）を可視化する手法 Tiled Parallel Coordinates (TPC) が提案された[4]。各変量の値の時間変化は minmax plot の特徴により最大値と最小値が表示され、density map の特徴により値の各時間帯における分布区間が表示される。しかし、値の時間変化の詳細を分析できない。さらに、多エンティティへの対応としては二人の脳電波データの可視化結果の比較までしか言及されていない。

気候データの多エンティティ多変量時系列データを分析対象とした研究が数多く行われた[2][3][5]。しかし、これらの研究では、各変量間の相関関係が時間とともにどう変化するかを明らかにすることを目的とし、エンティティに着目した分析を支援できないと考えられ

る。それらの中に、変量の値を離散的にいくつかのレベルを分けて、値の変化を色により表現し、一つの変量の値の時間変化を一つの色で視覚的に表現する手法がある[18]。異なるデータエンティティの多変量時系列情報は上下で並べて表現される。また、複数の視覚的表現のビューにより、多エンティティ多変量時系列データの視覚的分析を支援するツールも開発された[3]。

患者の身体状況データや診断履歴データの多エンティティ多変量時系列データを分析対象とした研究も多数行われた[19][20]。患者の診断履歴の多エンティティ多変量時系列データの視覚的なクエリ操作と結果表示のインタフェースが提案された[20]。事前にデータに対する知識を持つ上でクエリ操作を通じてデータを分析するのは本研究との違いである。一方、本研究では、データの視覚的な表現を通じて、データを理解しながら分析することが可能である。

第3章 多エンティティ多変量時系列データの分析要件

1.2 節で説明したように、エンティティに着目して多エンティティ多変量時系列データを分析する際に、ある特徴を持つ一つ或いは一部のデータエンティティについての多変量時系列情報を把握したい場面が多い。また、ある特徴を持つ一つ或は一部のデータエンティティの値について評価する時に、全データエンティティの値を参照する必要もある。例えば、学生 A は 1 学期に数学で 90 点を取ったが、多くの学生が 90 点以上を取っている場合、学生 A の数学の成績は相対的に低いという評価になる。つまり、エンティティに着目して多エンティティ多変量時系列データを分析する時に、全体データエンティティと個々データエンティティの分析を切り離せない。

また、エンティティに着目して多エンティティ多変量時系列データを分析する際に、Shneiderman が提案した「Overview first, zoom and filter, then detail-on-demand」の視覚的分析のガイドラインに従って[21]、まず、オーバービューとして全データエンティティの一覧を提供する必要がある。次に、ズームとフィルターの操作を行い、ユーザが着目したい特定の特徴があるデータエンティティの集合に対する多変量時系列情報についての詳細分析を支援する必要がある。そして、異なるデータエンティティの集合間の比較もよく行われるので、これを支援する必要がある。次に、ユーザの分析要求に合わせて、一つのデータエンティティに対する多変量時系列情報の詳細な分析を支援する必要がある。また、異なるデータエンティティ間の比較もよく行われるので、それに対しての支援の必要もある。

以上のエンティティに着目した多エンティティ多変量時系列データの分析の特徴に合わせ、六つの分析要件をまとめた。

- a) 全体データエンティティの一覧と各データエンティティの詳細情報を同時に把握できる。
- b) 全体データエンティティを一覧できる。
- c) 特徴がある部分データエンティティの集合の詳細を分析できる。
- d) 異なる部分データエンティティの集合の詳細を比較できる。
- e) 一つのデータエンティティの詳細を分析できる。
- f) 異なるデータエンティティの詳細を比較できる。

また、データエンティティの多変量時系列情報に対しての詳細分析について、本研究は以下のように定義した。

- 各時刻における各変量の値及び値の和を把握できる
- 各変量の値の時間変化及び値の和の時間変化を把握できる

本研究では、以上にまとめた分析要件を満たす視覚的表現やインタラクション操作、及びそれらを用いた分析ツールを開発する。

第4章 多エンティティ多変量時系列データための視覚的表現

4.1 視覚的表現の設計

一覧表示する可視化手法とは、多エンティティ多変量時系列データの多エンティティ、多変量、時系列三つの属性についての情報を一枚の図に表現する手法のことである。parallel coordinates を拡張して多エンティティ多変量時系列データを一覧表示する可視化手法が開発された[6]。しかし、限られた表示スペースの中に大量の情報を表示すると、大量の重なりが起こってしまう。人間の限られた知覚的な分析能力にとって、この表現を使った分析は難しいと考えられる[22]。また、人間が同時に処理できる情報量は限られているため、その量を超えた大量の情報が同時に表現されても、分析に役に立たないことがある。

本研究では、複数のビューの組み合わせにより、多エンティティ多変量時系列データを視覚的に表現する。

本研究では、散布図と small multiples の組み合わせにより、全体データの概観を視覚的に表現する。さらに、small multiples の各データエンティティの多変量時系列における詳細情報を line graph、stack graph、radar chart、sector graph の四つの可視化手法により視覚的に表現する（図 4-1）。

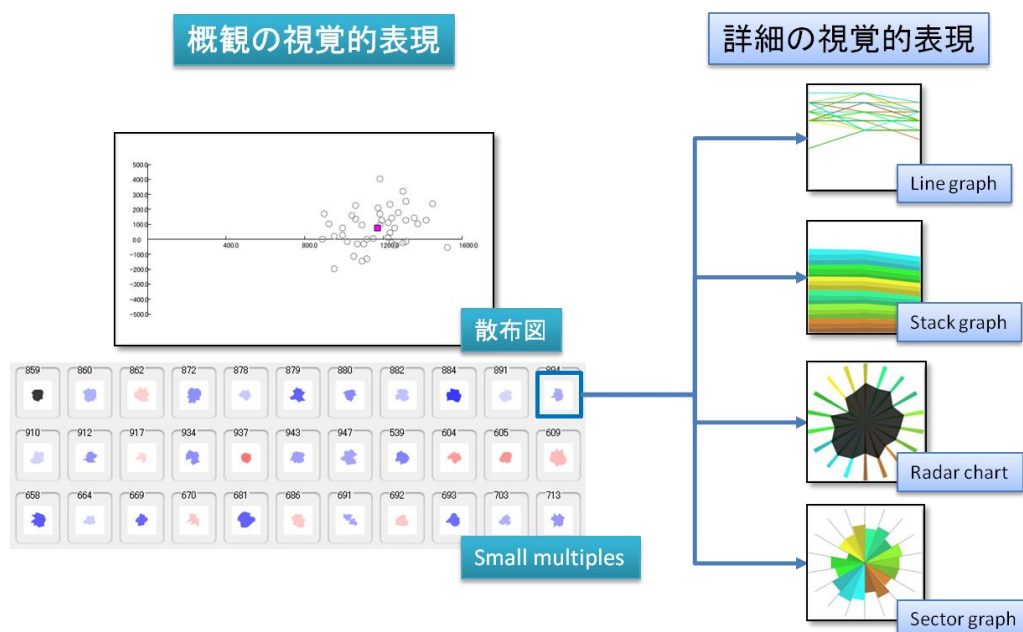


図 4-1 視覚的表現の設計

4.2 概観の視覚的表現

散布図は有用な可視化技術の一つである。垂直な二つの軸により構成された座標系に、値を対応させ、各データエンティティを記号で表示してプロットする。より直感的にすべてのデータエンティティの分布を把握できる特長がある。しかし、簡単な記号はただデータエンティティを代表することしかできない、各データエンティティ自身を持つ詳しい情報を同時に表すことができない。

一方、small multiples は複数のデータエンティティの詳細情報の視覚的な表現を同時にユーザに提示できる。しかし、grid のレイアウトに制限され、自由に各データエンティティの分布を表すことができない。

本研究では、散布図と small multiples を組み合わせることにより、お互いの短所を補い、全体データエンティティの分布と各データエンティティの詳細情報を同時に把握できる概観を提供する。

この概観を提供することにより、いくつかの分析要件を満たすことができる。まず、散布図で表現した全体データエンティティの分布より全体データエンティティの一覧を把握できる（要件 b）。さらに、small multiples の各データエンティティの詳細情報の視覚的表現と組み合わせることにより、全データエンティティの一覧と各データエンティティの詳細情報を同時に把握することができる（要件 a）。また、small multiples の各データエンティティの視覚的表現を通じて、各データエンティティの多変量時系列情報における詳細情報を分析することができる（要件 e）。また、small multiples の複数のデータエンティティの視覚的表現を同時に提示する特徴により、異なるデータエンティティの間の比較も行える（分析要件 f）。

4.2.1 散布図により全体データエンティティの分布を表現

多エンティティ多変量時系列データに対して、多変量と時系列の二つの属性における情報に基づいてデータエンティティ进行分类することがよくある。例えば、「全学生の中で、総合成績が高く、かつ 1 学期より成績がよくなった学生は誰か？」などの分析がよく行なわれる。多エンティティ多変量時系列データ分析のこの特徴に対して、散布図はデータエンティティの多変量と時系列の二つの属性における情報についての分布を表現する必要がある。本研究では、散布図の横軸と縦軸をそれぞれ多変量と時系列の二つの属性における情報に対応するようにした（図 4-2）。

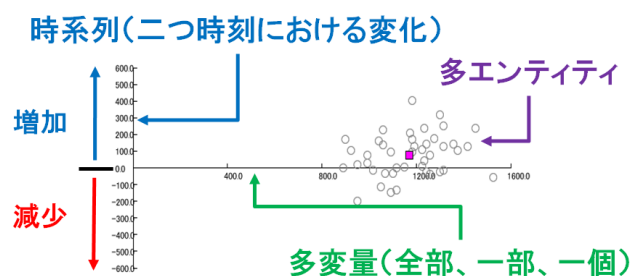


図 4-2 散布図の視覚的表現

散布図の中の各白点は各データエンティティを表す。紫の正方形の点は横軸と縦軸においての値の平均点を表す。横軸は多変量の属性における情報を表すのに使われ、ユーザが選択した変量の和を表す。ユーザの選択により、多変量の属性におけるすべての変量の和と一部の変量の和と一つの変量の値に切り換えられる。縦軸は、時系列の属性についての情報を表すのに使われ、ユーザが指定した二つの時刻の横軸の変量の和の差を表す。原点より上にあるデータエンティティは前の時刻より値が増加したことを表し、逆に、原点より下にあるデータエンティティは前の時刻より値が減少したことを表す。

学生成績データを例として説明すると、横軸が学生のある学期の選択された科目の総合成績を表し、縦軸がこの成績と一つ前の学期の成績の差を表す。

4.2.2 Small multiples により各データエンティティの詳細情報を表現

サムネイルサイズの画像を並べて表示する表現手法は「small multiples」と呼ばれる。データの時間変化の表現や異なるデータの比較によく使われる[23]。

本研究では、small multiples の各サムネイルサイズの画像は各データエンティティの多変量時系列の詳細情報を表す。各サムネイルサイズの画像を small multiples の子パネルと呼ぶ。各子パネルの左上に各データエンティティの ID を表示する。

4.3 詳細の視覚的表現

本研究では、line graph、stack graph、radar chart、sector graph の四つの可視化手法を用いてデータエンティティの多変量時系列における詳細情報を視覚的に表現する。

値の変化を表すために、line graph は最もよく使われている手法である。さらに、複数の値の変化を同時に表す多変量時系列情報の可視化手法としては、line graph から発展された手法が数多く提案されている。しかし、変量の数の増加に従って、最もシンプルな line graph のユーザパフォーマンスが優れているという結果が実験により示された[22]。本研究では、line graph により、各変量の値の時間変化を表す。しかし、分析にとって、線の重なりはやはり一つ無視できない問題である。ここで、本研究では、stack graph を採用する。各変量を表す帯の積み上げにより、線の重なり問題が改善される同時に変量の値の和の変化も把握できる。

さらに、本研究では、radar chart を使って、データエンティティのある時刻の多変量の情報を表す。データエンティティの radar chart の輪郭により、多変量の値の特徴（強みや弱み）をより直感的に読み取れ、面積により、多変量の値の和の大きさをより直感的に読み取れる。しかし、radar chart の軸の並び順は非常に表示結果を左右するので、本研究では、もう一つの多変量可視化手法 sector graph を採用する。図 4-3 は同じデータに対しての radar chart と sector graph の表示結果を示す。右の sector graph では、三つの変量に対して値があるということが読み取れるが、左の radar chart では、同じ情報が読み取れない。さらに、多変量時系列情報を視覚的に表現できるように、radar chart と sector graph を拡張する。

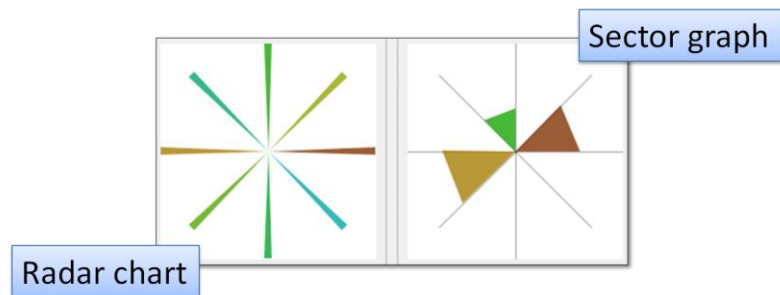


図 4-3 radar chart の表示問題

表 4-1 にこれらの 4 つの可視化手法の多変量時系列情報に対する「詳細分析」の分析要件との対応をまとめる。表 4-1 を見ると、同じ情報の分析に対して、複数の表現が使える。例えば、「一つの時刻の各変量の値」の分析に対して、4 つの可視化手法どちらでも使える。同じな情報の分析に対しても、いろいろな表現を使うことにより、意外な発見ができるかもしれないので、本研究では、できるだけ多くの視覚的表現をユーザに用意する。

本研究では、色により異なる変量を区別する。表現の一貫性を保つために、四つの視覚的表現手法の色と変量の対応関係が統一されている。例えば、灰色は変量 A を表すことが決まると、line graph、stack graph、radar chart、sector graph の全ての視覚的表現手法においての灰色が変量 A を表す。

表 4-1 詳細の視覚的表現が分析要件との対応

	一時刻の 各変量の 値	各変量値 の時間変 化	一時刻の 変量の値 の和	変量の値 の和の時 間変化
Line graph	○	○	×	×
Stack graph	○	○	○	○
Radar chart	○	×	○	○
Sector graph	○	○	○	×

4.3.1 Line graph の視覚的表現

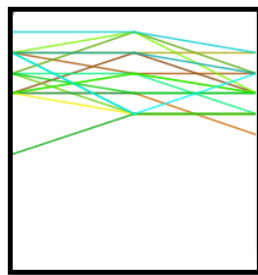


図 4-4 line graph の基本表現

本研究での line graph の基本表現を図 4-4 に示す。横軸は時間軸を表し、図の一番左が最初の時刻となり、一番右が最後の時刻となる。縦軸は変量の値を表し、図の一番下が 0 となり、一番上が変量の最大値となる。

ユーザが指定した時刻は時間軸と垂直する灰色の直線により表示される。また、各線にマウスオーバーすると、この線が表示する変量の名前がグラフの左上の所に表示される。

4.3.2 Stack graph の視覚的表現

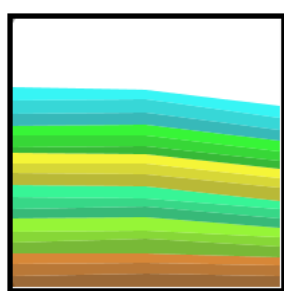


図 4-5 stack graph の基本表現

本研究での stack graph の基本表現を図 4-5 に示す。横軸は時間軸を表し、図の一番左が最初の時刻となり、一番右が最後の時刻となる。各変量を表現するスタックが縦方向に積み上げられる。各変量のスタックを色により区別する。縦軸のスケールは変量の最大値と選択された変数の数の掛け算となる。こうすることにより、数の少ない変数が選択された場合の分析がより容易に行える。図 4-6 には、左は 18 個の変数が選択された場合の stack graph の表現となり、右は 3 個の変数が選択された場合の stack graph の表現となる。各変量の最大値は 100 となる。

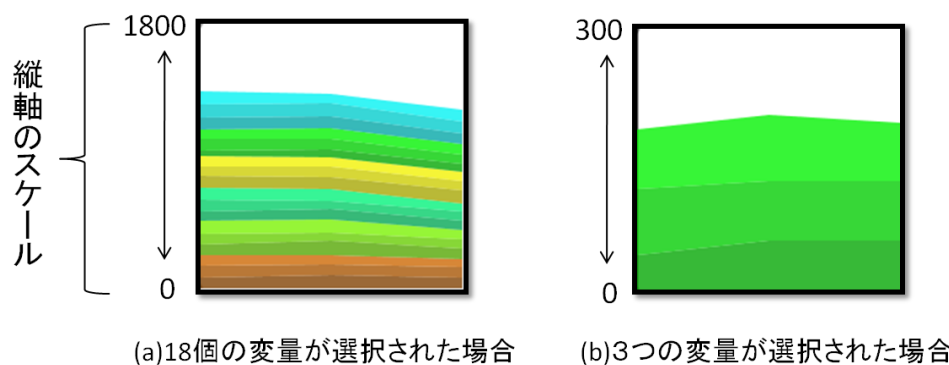


図 4-6 stack graph の縦軸のスケールの変化

下側に積み上げたスタックの形が上側のスタックの形に影響を与え、上側のスタックが表示する変量に対しての分析の障害になるのは **stack graph** の一つ視覚的表現の問題である [24]。本研究では、これを解決するために、以下のインタラクション操作を開発した。まず、あるスタックにマウスオーバーすると、このスタックがハイライトされる(図 4-7 (b))。さらに、マウスの左ボタンを押すと、このスタックの底が平らになる(図 4-7(c))。このようにすることで、下側に積み上げたスタックの影響が無くなり、一つの変量の変化に対しての分析や異なるデータエンティティの同じ変量の変化の比較がより容易に行える。

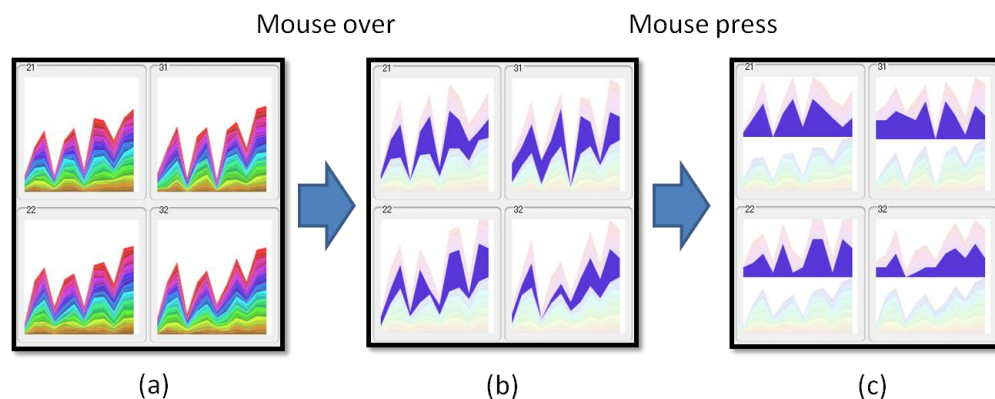


図 4-7 **stack graph** のインタラクション操作

4.3.3 Radar chart の視覚的表現

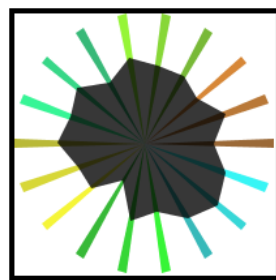
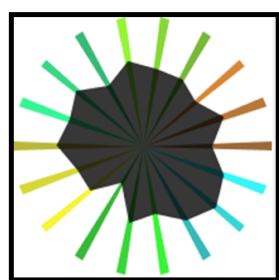


図 4-8 **radar chart** の基本表現

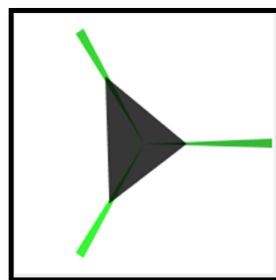
基本表現

Radar chart は有効な多次元データの可視化技術であり、**web chart**, **spider chart**, **star chart** なども呼ばれているが、表示方法が基本的に一緒である [25]。本研究での **radar chart** の基本表現は図 4-8 により示す。グラフの中心から複数の等角度に並べた軸が各変量と対応する。変量を表示する各軸に色を付ける。各変量の値により各軸に点を付ける。点の位置が中心と近いほど変量の値が小さい。次に、線でこれらの点を繋ぎ、一つ多変量時系列データの一つの時刻においての多変量情報の **radar chart** を成す。

数の少ない変数が選択された場合の分析をより容易にするために、変数の選択操作により変数の軸の並び位置を変える。図 4-9 には、左は 18 個の変数が選択された場合の radar chart の表現とであり、右は 3 個の変数が選択された場合の radar chart の表現である。付けられた色により、軸と変数の対応関係を維持する。さらに、radar chart の自身の表現制約により、選択された変数の数が 3 未満の場合の視覚的表現ができない。



(a)18個の変数が選択された場合



(b)3個の変数が選択された場合

図 4-9 radar chart の軸の並びの変化

時系列情報への発展

しかし、radar chart では時系列情報を表現できない。本研究では、radar chart に色を塗ることでデータエンティティの変数の値の和の二つの時刻における時間変化情報を表現する。青は値の和が増加したことを表し、赤は値の和が減少したことを表す。変化の度合いを色の彩度により表す。色が濃いほど変化が激しい。黒は値の和が変化していないことを表す。ここの値の和はユーザに選択された変数の値の和である。

図 4-10 が 6 個のデータエンティティの radar chart 表現を示す。左上と右下の三つのデータエンティティが前の時刻より値の和が増加したということが分かる。さらに、色の彩度からみると、左上のデータエンティティの増加の度合いが右下の二つのデータエンティティの増加の度合いより大きいことが読み取れる。

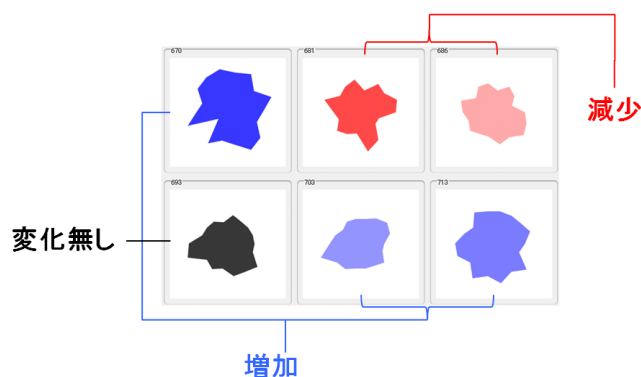


図 4-10 radar chart の色表現

二つの時刻の値の変化を表すだけでは、多変量時系列データの分析にとって不十分である。二つの時刻の間のすべての時刻においての変化を表すのは一つ重要な課題である。本研究では、以下のことにより、変量の値の和のすべての時刻においての変化を表す。

まず、データエンティティのすべての時刻の radar chart を重ねて表示する。ただし、中身の代わりに radar chart の輪郭に色を塗る。各時刻の radar chart の輪郭の色はこの時刻と一個前の時刻の値の変化を表す。点線となっている輪郭がデータエンティティの最新の時刻の radar chart となる。

次に、時系列順に従って、各時刻の radar chart の大きさを変える。一番古い時刻の radar chart を一番小さくし、一番新しい時刻の radar chart を一番大きくする。真ん中の radar chart の色から各時刻の radar chart の色を見れば、データエンティティの値の和の時間変化が分かる。さらに、各時刻の radar chart の形状を見ると、データエンティティの特徴（強みと弱み）の変化も同時に把握できる。

図 4-11 は一つのデータエンティティの三つの時刻の radar chart の表現となる。まず、三つの radar chart の色からみると、このデータエンティティの変量の値の和が最初に増加して、次に減少したという変化が分かる。さらに、形状からみると、最後の時に、このデータエンティティの左上の変量は弱みになったということも読み取れる。

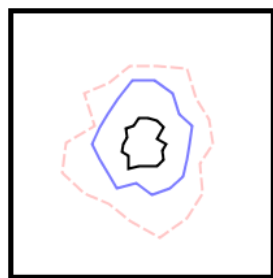


図 4-11 radar chart の時系列情報の視覚的表現

4.3.4 Sector graph の視覚的表現

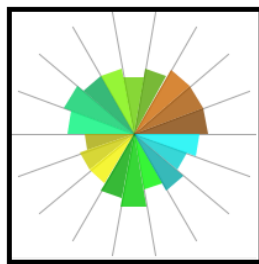


図 4-12 sector graph の基本表現

基本表現

本研究では、sector graph を使って各変量の値の詳細情報を表す。各等角度のセクタが各変量を表す。各セクタのサイズが指定された時刻の各変量の値により決められる。本研究では、セクタの外側の形を円弧の代わりに直線にした。このようにすることで、より容易に値の比較ができる。本研究での sector graph の基本表現は図 4-12 により示す。

時系列情報への発展

Sector graph は時系列情報を表現できない。本研究では、MultiCombs[9]の表現手法からインスピレーションを得て、sector graph を時系列情報が表現できるように拡張した。

各セクタの外側のブロックに色を塗ることで、各セクタに対応する時間変化情報を表す。また、提示する時間帯を指定することができる。時計周りで時間軸を表し、各ブロックの左端が最初の時刻、右端が最後の時刻の時間変化を表す。ここでの時間変化とは、一つ前の時刻との差のことである。各時刻の時間変化の色は等間隔に分布される。各色の間の色がグラデーションにより埋められる。色と変化の対応関係は radar chart の場合と同じであり、赤は一つ前の時刻の値より減少したことを表し、青は一つ前の時刻の値より増加したことを表し、黒は一つ前の時刻の値と変わっていないことを表す。

図 4-13 は、三つ時刻（1、2、3）を持つデータエンティティの sector graph 表現例である。時刻を 1 にした場合は、最初と最終時刻が同じなので、各セクタの色は無変化を意味する黒色となる(図 4-13(a))。時刻を 2 に指定した場合は、図 4-13(b)から、このデータエンティティの左下の六つの変量と右上の一つの変量の値が減少した以外ほとんどの変量の値が増加したということを読み取れる。また、最終時刻を 3 に指定した場合は、図 4-13(c)から、さき減少した変量がほとんど増加したということも読み取れる。

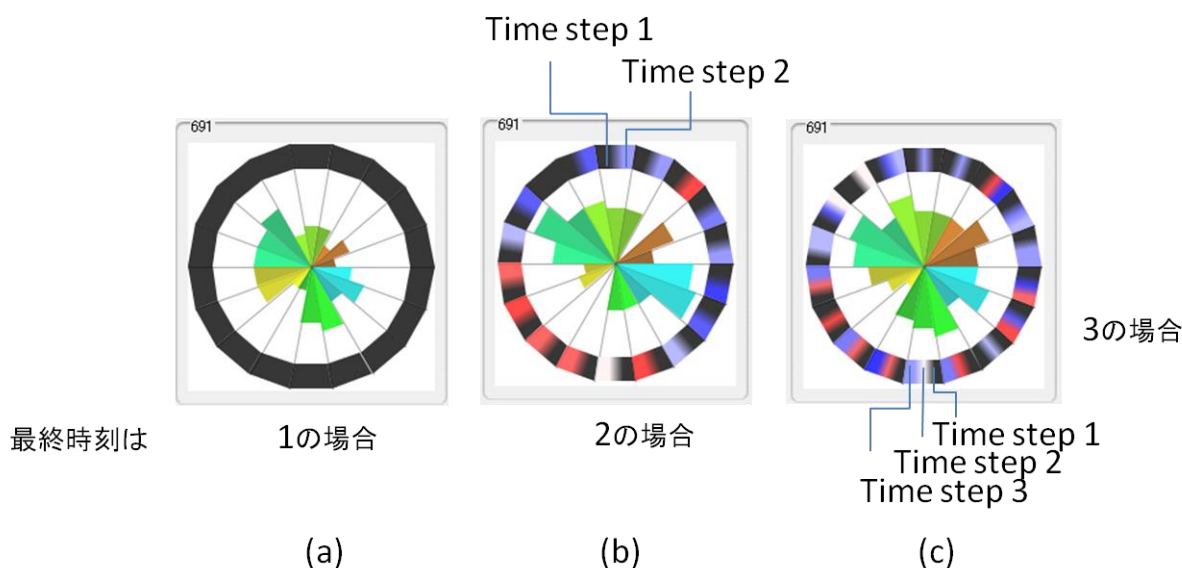


図 4-13 sector graph の時系列情報の視覚的表現

4.4 視覚的表現要素—色の使用

4.4.1 情報の表現による色の使い分け

本研究では、多くの情報の視覚的表現が同じ視覚的表現要素—色を使っている。例えば、**radar chart** と **sector graph** においての値の時間変化と異なる変量の区別の視覚的表現に色が使われている。情報分析を行う際に、ユーザを混乱させないために、異なる情報の視覚的表現に同じ色を使わないようにしている。例えば、青、赤と黒が **radar chart** と **sector graph** においての値の時間変化に使われるため、異なる変量の区別にこの三つ以外の色を使うようにする。

4.4.2 階層構造を持つ変量の色付け

多変量の属性において、階層構造を持つことがよくある。例えば、学生の成績データには、多変量を成す各科目は、まず、大まかに文科と理科の二つのカテゴリに分けることができる。文科には、さらに、国語、英語、歴史などの科目があり、理科には、数学、物理、生物などの科目がある。

本研究では、色相により異なるカテゴリを区別し、同じカテゴリに所属する変量を彩度 S により区別する。以上のことにより、多変量の属性においての階層構造を表す。さらに、**sector graph** の視覚的表現の時に、各カテゴリを区別しやすくするために、近い色相の色を隣に配置しないようにする。図 4-14 は六つのカテゴリに所属する 18 個の変量を表す **sector graph** の視覚的表現を示す。各カテゴリに三つの変量がある。図 4-14(a) は近い色相の並び替えが配慮されていない **sector graph** の表現となり、図 4-14(b) は近い色相の並び替えが配慮される **sector graph** の表現となる。

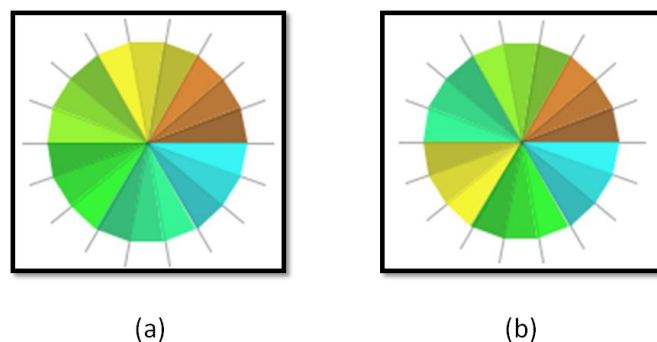


図 4-14 多変量の階層構造の色付け

4.5 視覚的分析の流れ

エンティティに着目した多エンティティ多変量時系列データの視覚的分析に対して、本研究では、Shneiderman が提案した「Overview first, zoom and filter, then detail-on-demand」の視覚的分析のガイドラインに従って[21]、多エンティティの属性に対して、「全体→部分→詳細」の視覚的分析流れを考えた。

まず、散布図により表示された全体データエンティティの分布と small multiples により表示された全データエンティティの詳細情報から注目した部分データエンティティを選択する。次に、選択された部分データエンティティの small multiples においての子パネルだけ残し、サイズを大きくする。最後に、四つの詳細情報の視覚的表現を使って、多変量時系列情報に対して詳しく分析する。また、ユーザの要求に合わせ、任意のステップに戻って分析することも可能である。

図 4-15 は本研究で開発した視覚的表現を用いた視覚的分析の流れを示す。

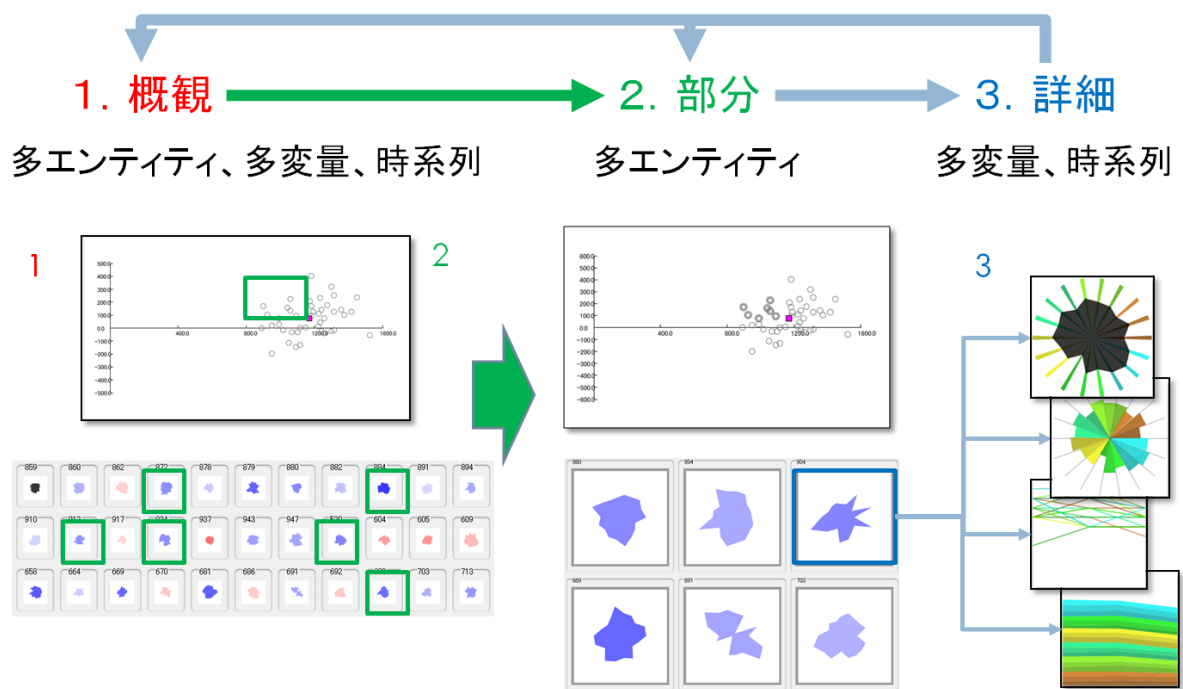


図 4-15 視覚的分析の流れ

第5章 多エンティティ多変量時系列データの視覚的分析ツールの開発

5.1 ツールの設計

本研究では、第4章で提案した複数の可視化手法を用いて、エンティティに着目した多エンティティ多変量時系列データの分析ツール“MMTAnalyzer”を開発する。異なる可視化手法を一つのツール上に統合するために、linking & Brushing[26]の手法を分析ツールに適応する。

Linking & Brushing とは、同じデータの複数のビューを連携させることで、一つのビューの変化を他のビューでもリアルタイムで反映させるインタラクション手法である。Linking[27]とは異なる可視化表現において、同じ情報の視覚的な提示方法を統一することである。Brushing[28]とは全体から着目したい部分を指定することによって、何らかの方法で指定された部分をハイライトして表示する手法である。

開発する分析ツールでは同じデータエンティティに対しての操作をリアルタイムに散布図と small multiples 両方でハイライトして表示することにより、散布図と small multiples を一つの可視化手法に統合させ、全体データエンティティの分布と各データエンティティの多変量時系列においての詳細情報を同時に把握できるような分析が行えるようにする。図5-1は散布図と small multiples それぞれにおいてのデータエンティティの視覚的な表現の通常状態、マウスオーバーされた状態と選択された状態のハイライト表示を示す。



図 5-1 散布図と small multiples でのハイライト表示

5.2 ツールのインタフェース

図 5-2 は MMTAnalyzer のスクリーンショットである。MMTAnalyzer は主に四つの部分から構成されている。図 5-2 の A の部分はデータの散布図の視覚的表現の散布図パネルであり、図 5-2 の B の部分はデータの small multiples の視覚的表現の small multiples パネルであり、図 5-2 の C の部分は全体データの変量の選択や時間の設定部分の変量と時間の指定パネルであり、図 5-2 の D の部分は散布図パネルと small multiples パネルに対しての操作や調整を行う部分の調整パネルである。

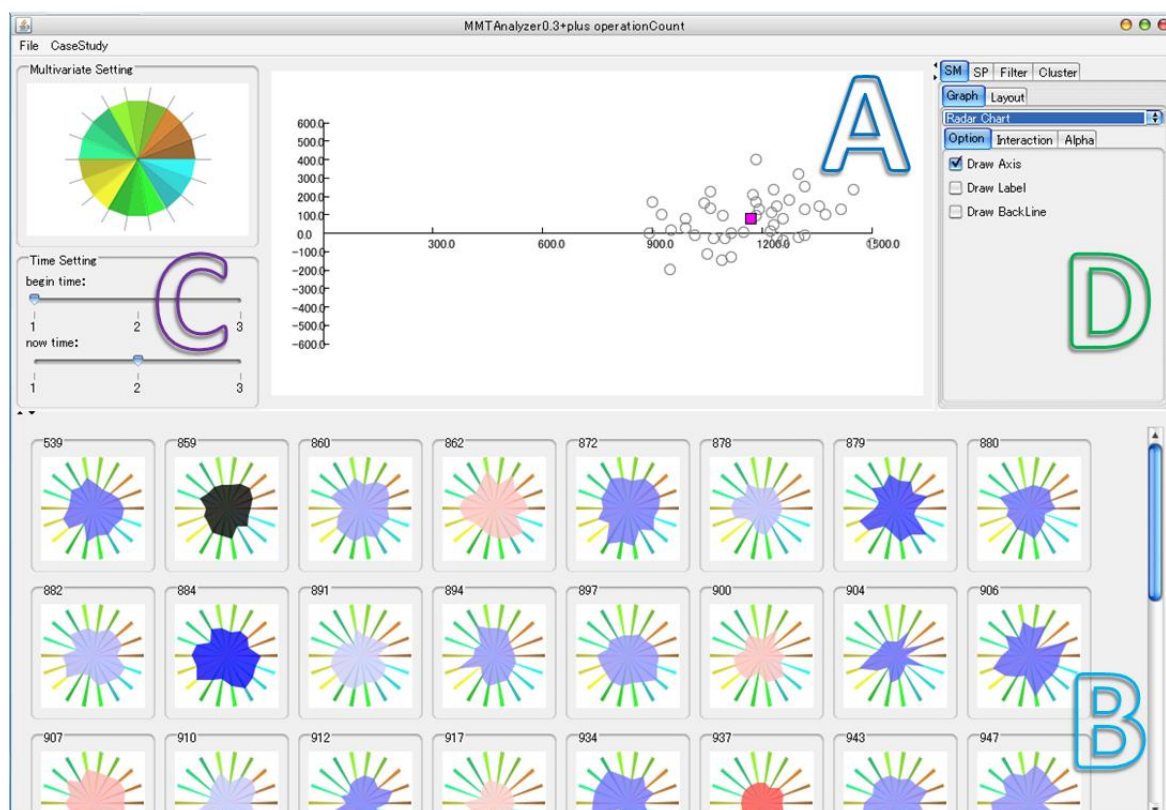


図 5-2 MMTAnalyzer のインタフェース

5.2.1 散布図パネル

散布図パネルは全データエンティティが散布図表現で表示される画面である（図 5-3）。

散布図パネルでは、データエンティティの矩形選択と任意形状の選択が行える（図 5-4）。二つの選択は散布図パネルにおいてのマウスの左ボタンのダブルクリックにより切り替えられる。

密集するデータエンティティの重なりを軽減するために、マウスホイールにより、散布図パネルの横と縦の大きさを調整することができる（図 5-5）。散布図パネルにおいてマウスの左ボタンをクリックしてからマウスホイールにより散布図パネルの横の大きさを調整する。散布図パネルにおいてマウスの右ボタンをクリックしてからマウスホイールにより散布図パネルの縦の大きさを調整する。

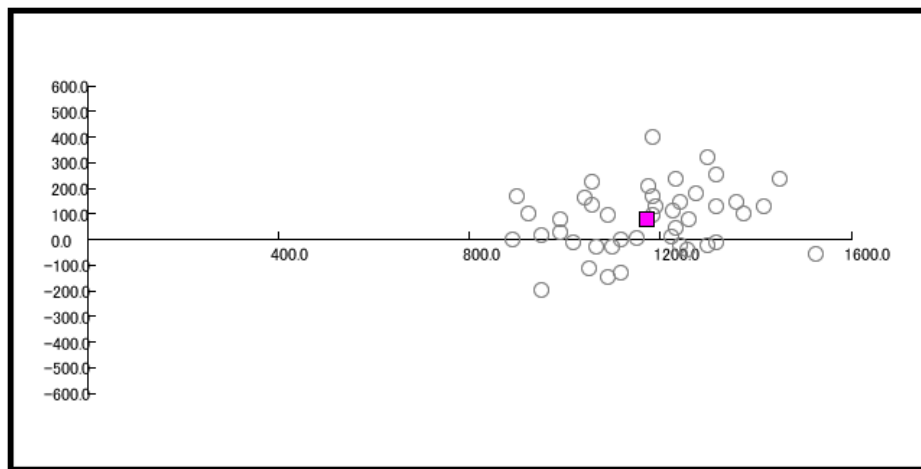
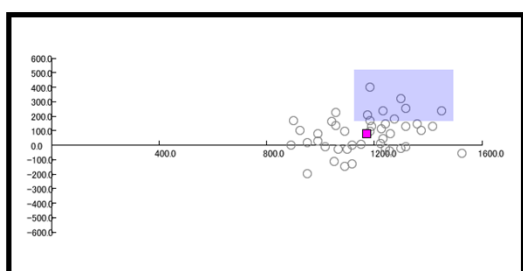
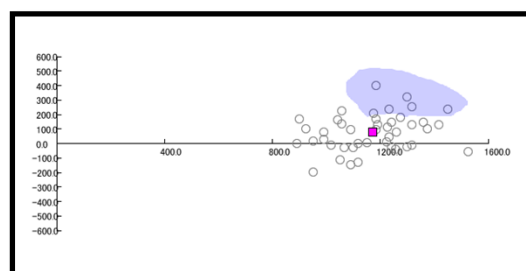


図 5-3 散布図パネル

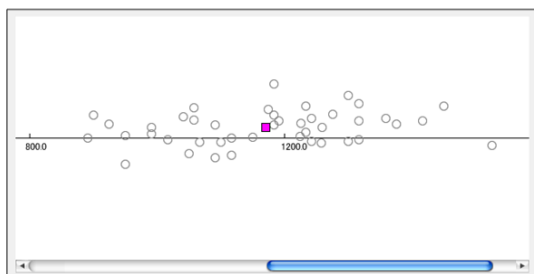


矩形選択

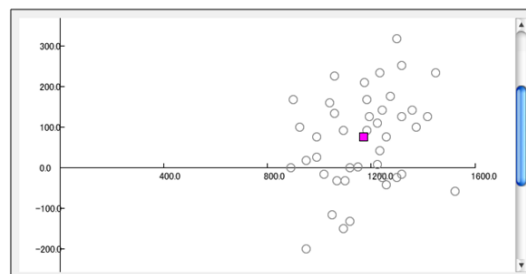


任意形状選択

図 5-4 散布図パネルでの選択操作



横の調整



縦の調整

図 5-5 散布図パネルのサイズの調整

5.2.2 Small multiples パネル

Small multiples パネルは全データエンティティが small multiples 表現で表示される画面である（図 5-6）。

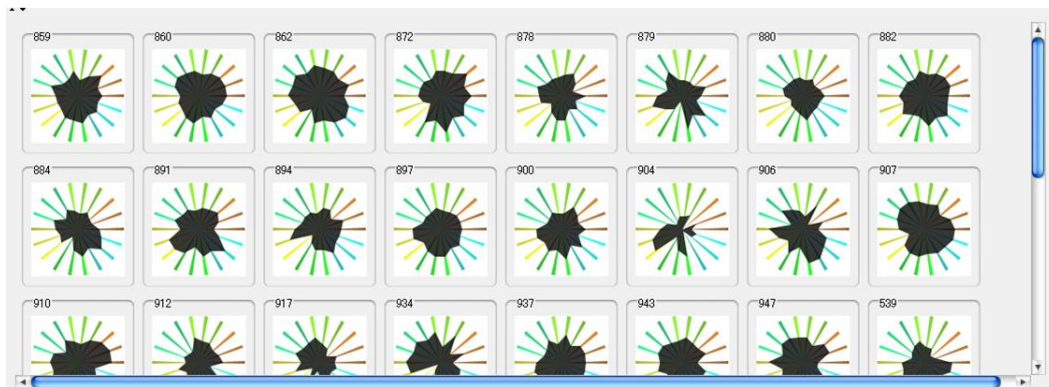


図 5-6 small multiples パネル

データエンティティの数が多く、一画面内に全ての子パネルが表示できない場合がある。この場合には、small multiples パネルの右のスクロールバーを調整し、表示しきれない部分のデータエンティティの子パネルを確認するか、各子パネルのサイズを小さくするか、或いは、small multiples パネルのツール全体での比重を大きくして、より広い表示スペースを取るようにする。図 5-7 は small multiples パネルをツールの全画面にした時のスクリーンショットである。

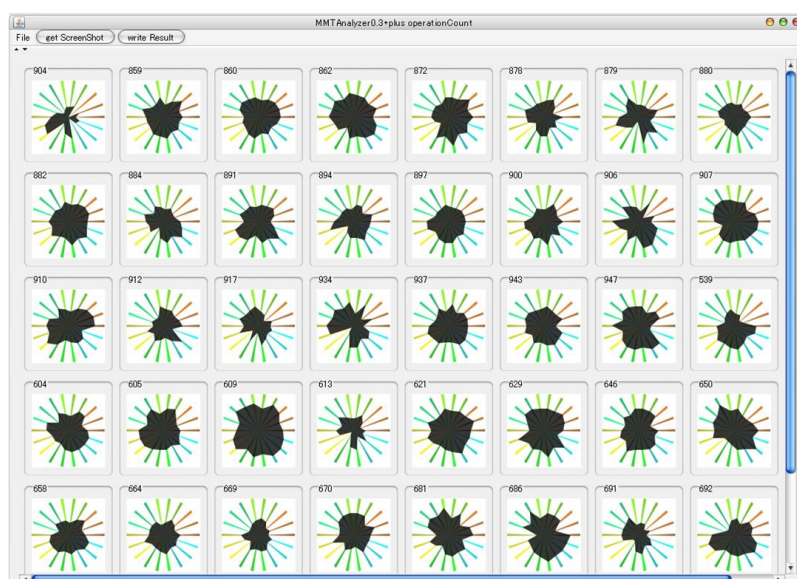


図 5-7 small multiples の表示スペースの調整

5.2.3 変量と時間の指定パネル

変量と時間の指定パネルは全体データの変量のフィルターと時間の指定を行うパネルである（図 5-8）。上側のパネルは変量のフィルターを行うパネルである。下側の二つスライダーは最初の時刻と最終の時刻を指定するのに使用する。

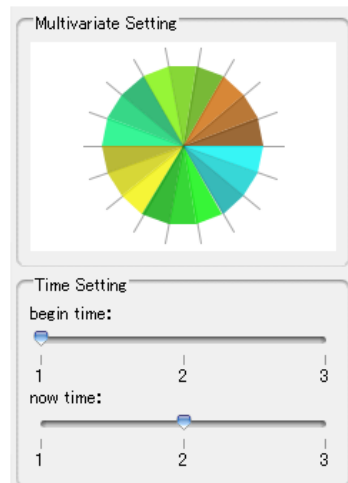


図 5-8 変量と時間の指定パネル

5.2.4 調整パネル

調整パネルは散布図パネルと small multiples に対する調整や操作を行うパネルであり（図 5-9）、四つのパネルにより構成されている。図 5-9 の a は Small multiples パネルに対しての操作を行う「SM」パネルである。図 5-9 の b は散布図パネルに対しての操作を行う「SP」パネルである。図 5-9 の c はデータエンティティのフィルター操作を行う「Filter」パネルである。図 5-9 の d はデータエンティティのクラスタ操作を行う「Cluster」パネルである。

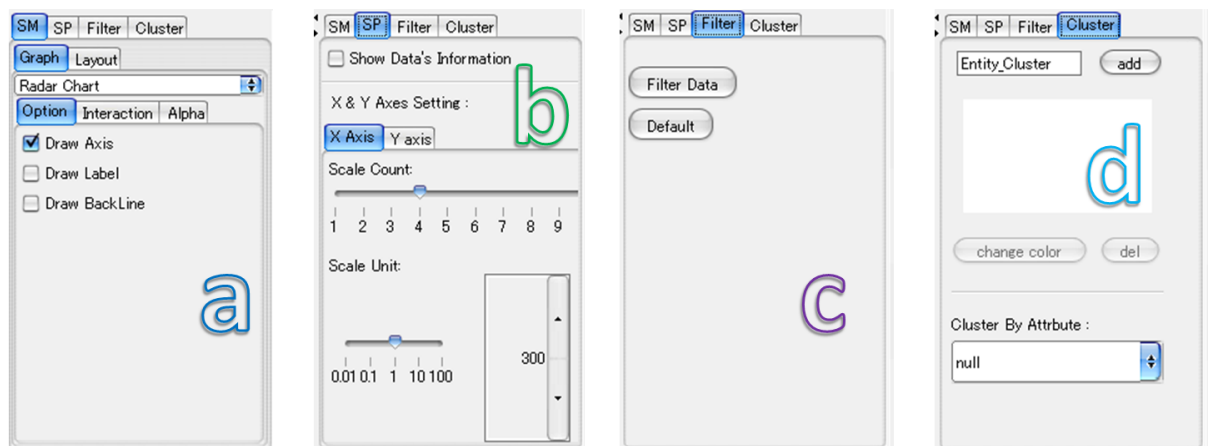


図 5-9 調整パネル

四つのパネルの中に、「SM」パネルはさらに、small multiples パネルの各子パネルの視覚的表現について操作を行う「Graph」パネルと small multiples パネルの各子パネルのサイズや並び順を調整する「Layout」パネルにより構成されている。また、「Graph」パネルは視覚的表現ごとにいくつかのパネルにより構成されている。

5.3 ツールの機能

5.3.1 平均値基準選択機能

平均値は統計データとして重要な意味を持ち、データ分析よく用いられる。例えば、学生の成績データの場合に、「クラスの平均より成績が悪い学生たちと成績良い学生たちに成績の特徴の違いがあるか」という分析が考えられる。本研究では、散布図において、通常の矩形選択と任意形状選択以外に、平均値基準選択という複数のデータエンティティ選択機能を開発した。

平均値基準選択機能は以下の手順で行う。まず、散布図での平均値を表す紫の正方形の点にマウスの左ボタンを押す。次に、左ボタンを押したままで、紫の点の周りにドラッグする。そうすると、紫の点とマウスの位置により散布図に一つの矩形が表示される。最後に、マウスの左ボタンを放すと矩形に含まれたデータエンティティが選択される。(図 5-10)

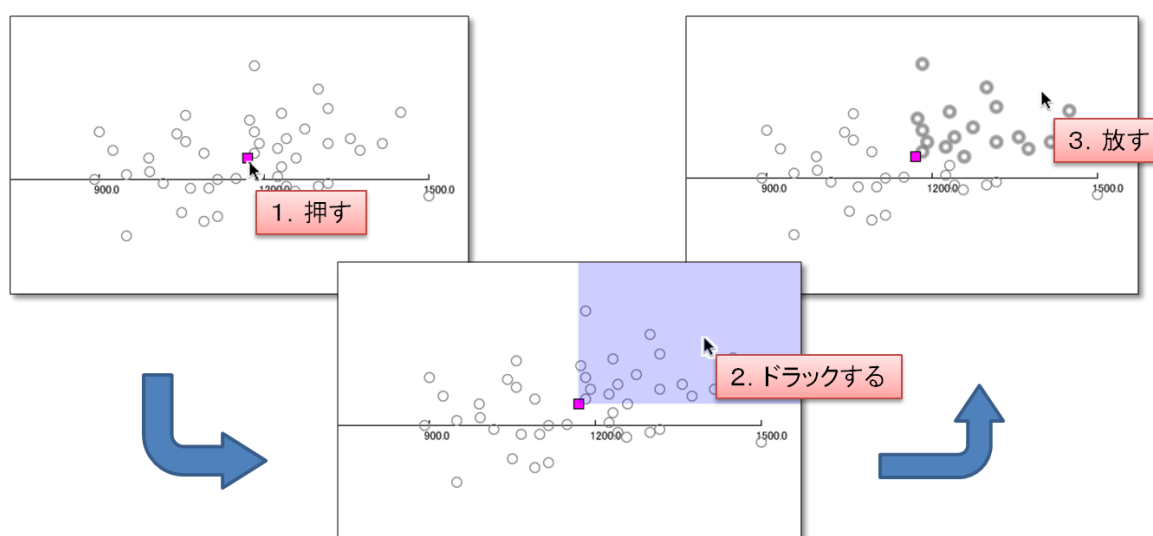


図 5-10 平均値基準選択

マウスの位置と紫の点を繋ぐ直線の角度により散布図の x 軸と y 軸の平均値を基準としての 8 種類のデータエンティティ集合を選択できる (図 5-11)。

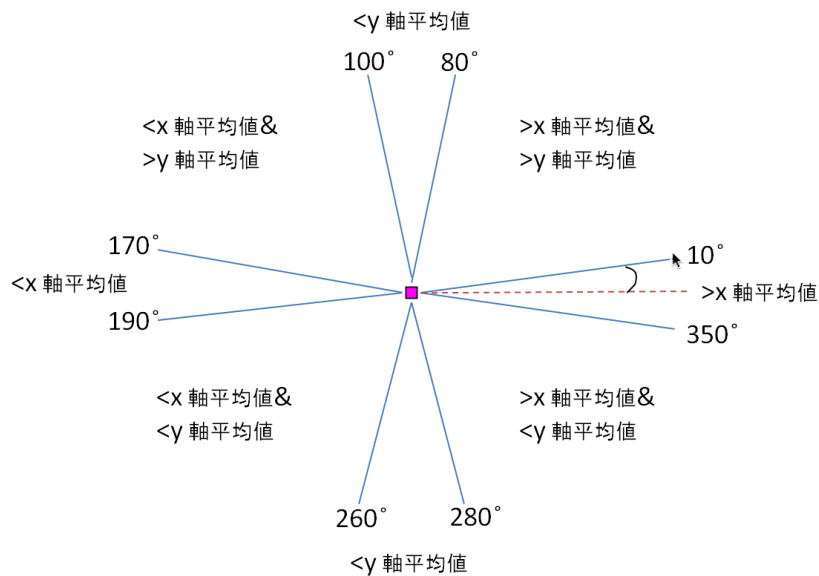


図 5-11 8種類のデータエンティティ

5.3.2 手描き変量選択機能

項目の選択や非選択操作は通常ラジオボタンあるいはチェックボックスによって実現される。しかし、複数の項目を操作する時に、一個一個のラジオボタンやチェックボックスに対して操作をしなければならないので、操作が面倒である。本研究では、複数の変量を選択や非選択する操作に対して、ラジオボタンとチェックボックスの代わりに、手描き変量選択機能を開発する。

5.2.3 節に紹介した変量と時間の指定パネルの上側のパネルは **sector graph** の視覚的表現を使って、多変量を表している。一個の変量に対して操作する場合は、セクタの上にクリックすることによりこのセクタが表示する変量を選択や非選択することができる。複数の変量に対して操作する場合は、ユーザがパネルに任意の線を描いたら、この線と交差するセクタが表示する変量が選択あるいは非選択される (図 5-12)。変量を表すセクタが非透明の時は変量が選択されているということを表し、透明になった時は変量が非選択されているということを表す。

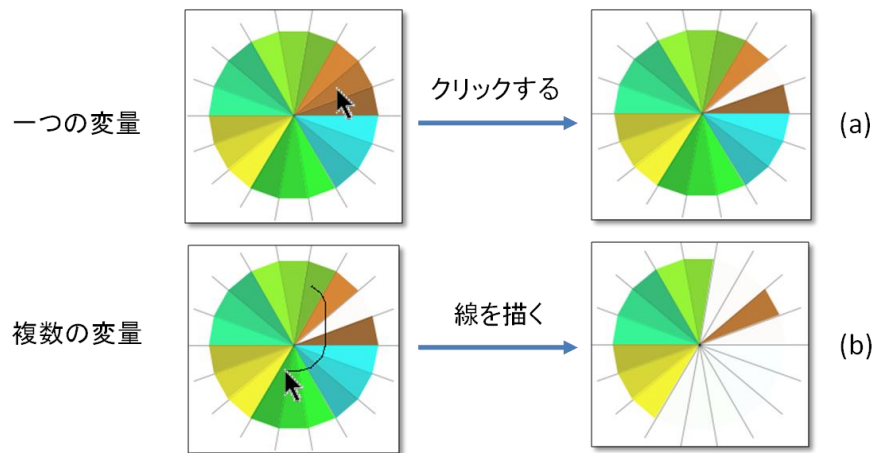


図 5-12 手描き変量選択機能

5.3.3 エンティティフィルター機能

本研究では、概観の視覚的表現から一部の特征があるデータエンティティをフィルターして詳しく分析する機能を開発する。このことにより、要件 c（特征がある部分データエンティティの集合の詳細を分析できる）を満たすことができる。

本研究で提供する概観は二つの部分により構成される。散布図で表現する全体データエンティティの分布から或いは **small multiples** で表現する各データエンティティの詳細情報から特征がある一部のデータエンティティを選択してフィルターすることができる。

散布図からのフィルタリング

まず、散布図において、一部のデータエンティティを選択する。次に「Filter」パネルの「Filter Data」ボタンを押すと、散布図で選択されていないデータエンティティの **small multiples** での子パネルの視覚的表現を消す。その次に、「SM」パネルの中の「Layout」パネルの「Sort Small Multiples」から「filter」を選択する。そうすると、先ほど選択されていなかったデータエンティティの **small multiples** での子パネルが隠される。最後に、同じ「Layout」パネルの「Panel Size」のスライダーを調整することにより **small multiples** で残されている子パネルのサイズを拡大して詳細情報について詳しく分析することができる。

図 5-13 により、以上の操作を示す。

Small multiples からのフィルタリング

Small multiples において、「Ctrl」キーを押しながら、各データエンティティの子パネルをクリックすることにより、複数のデータエンティティを選択することができる。その次の操作は散布図からのエンティティフィルター操作と同じである。

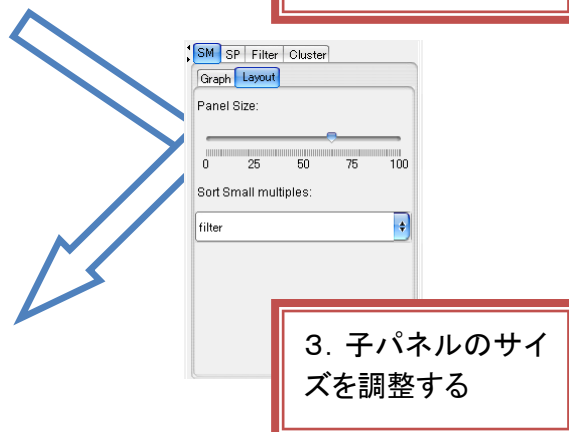
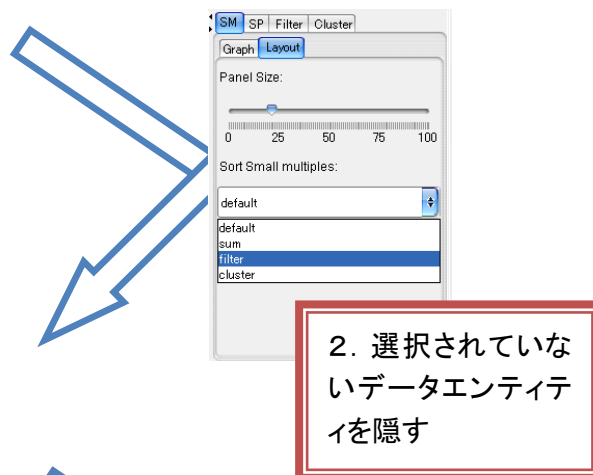
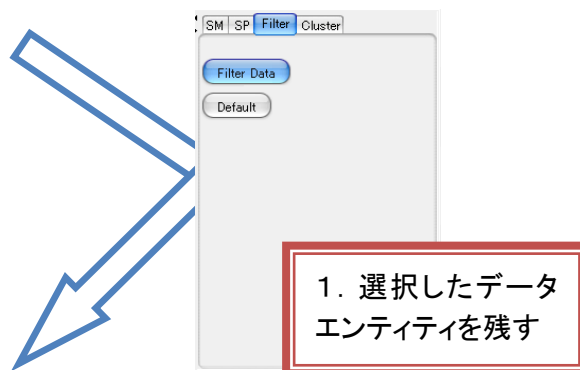
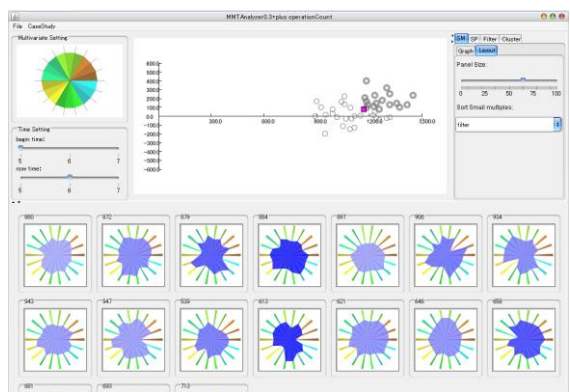
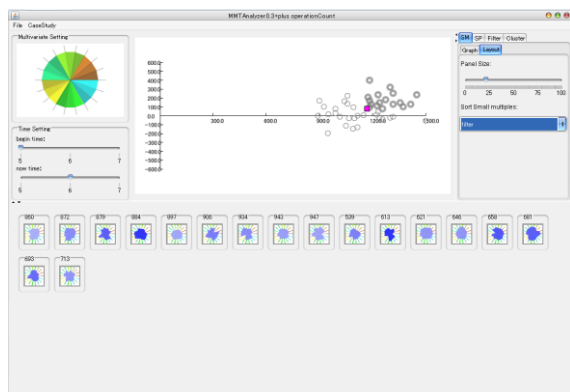
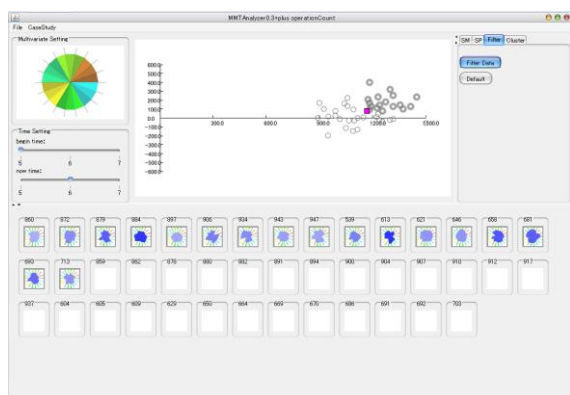


図 5-13 エンティティフィルター機能

5.3.4 エンティティ集合分析機能

本研究では、異なる特徴を持つデータエンティティの集合を比較する機能を開発する。このことにより、分析要件 d（異なる部分データエンティティの集合の比較を支援する）を満たすことができる。

まず、特徴がある一部のデータエンティティを選択する。次に、「Cluster」パネルの上側のテキストフィールドに集合の名前を入力する。その次に、横の「add」ボタンを押すと、入力されたテキストを使った名前のデータエンティティの集合が作られ、パネルの真ん中のリストボックスに集合の名前が追加される。リストボックスに追加された集合の名前を押すと、この集合に含まれるデータエンティティの散布図と small multiples での視覚的表現がハイライトされる。

異なるデータエンティティの集合を散布図での比較をしやすくするために、リストボックスに追加されたデータエンティティの集合の名前をクリックしてから、「Change color」ボタンを押して、色を選択して、集合のデータエンティティに色を付けることもできる（図 5-14）。

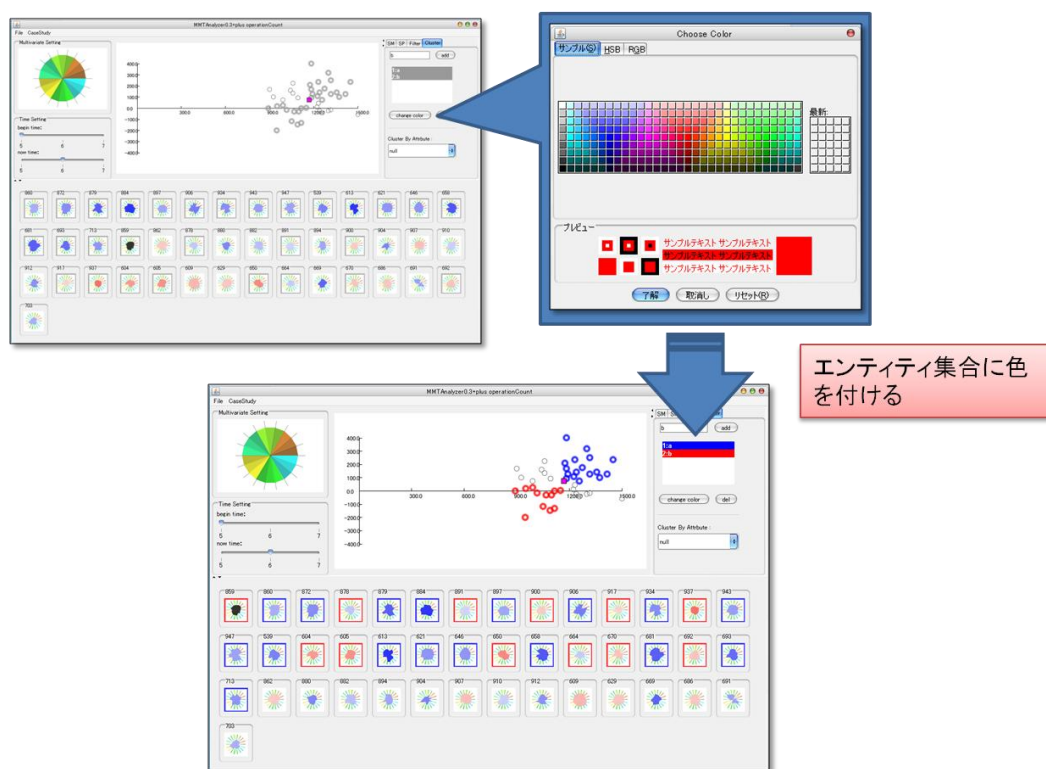


図 5-14 エンティティ集合の色付け

異なるデータエンティティの集合を small multiples での比較しやすくするために、「SM」パネルの中の「Layout」パネルの「Sort Small Multiples」から「Cluster」を選択して、small multiples でのデータエンティティの子パネルが集合ごとに詰めて表示できる（図 5-15）。

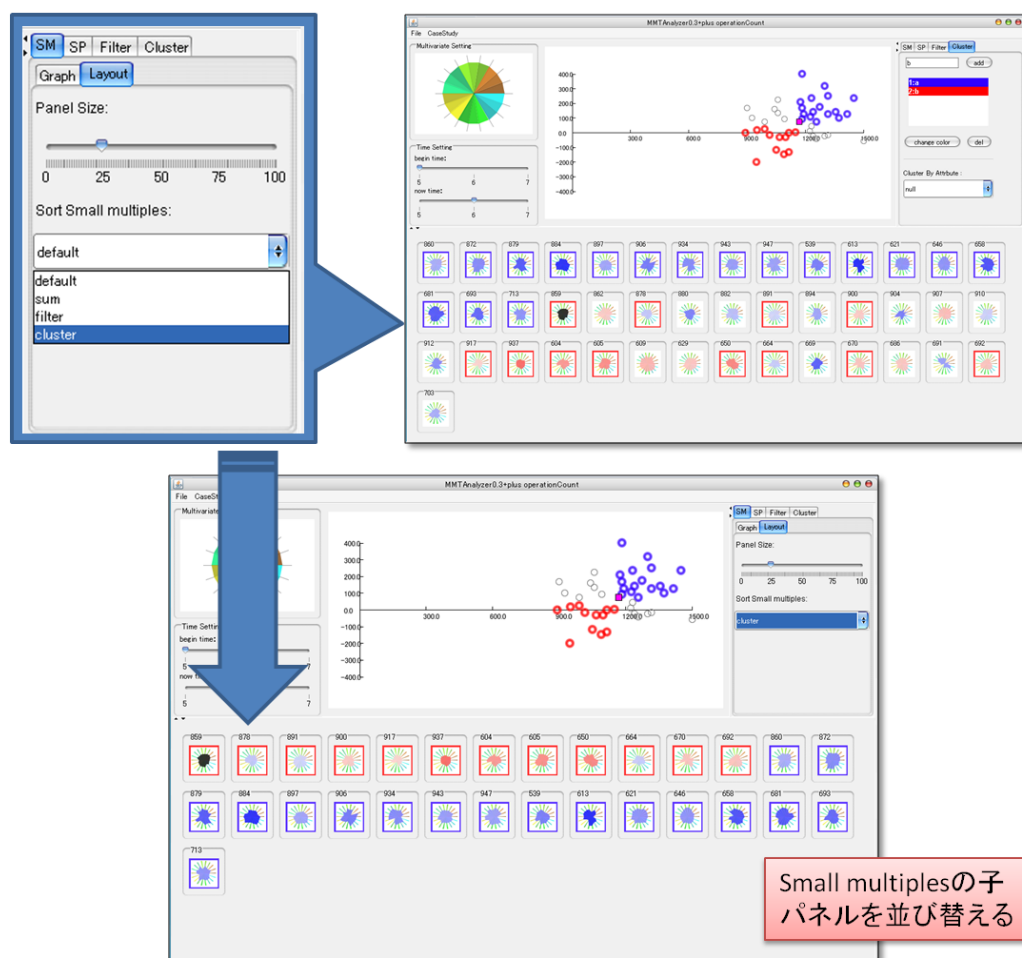


図 5-15 エンティティ集合の並び替え

5.3.5 属性によりクラスタ分析機能

各データエンティティは各変量に対しての値の以外にもいろいろな属性を持っている。例えば、学生の成績データの場合は、各学生は各科目に対しての成績以外にも、性別、入学年などの属性を持っている。

値の特徴や分布によりデータエンティティを分類してから分析する以外にも、属性によりデータエンティティを分類して分析することもよく行なわれる。本研究では、データエンティティの属性によりデータエンティティを分類する機能を備える。

「SM」パネルの「Cluster By Attribute」のコンボボックスからデータの一つの属性を選択すると、small multiples での各子パネルが選択された属性により分類される。異なる属性値を持つデータエンティティの子パネルの集合の前に属性パネルが挿入される。この属性パネルの左上はこれの後ろに詰めている子パネルの集合の属性値を表示している。属性パネル自身がこの集合に含まれたデータエンティティの散布図においての視覚的表現となる。さらに、各集合の名前になる属性値が「Cluster」パネルのリストボックスに追加される。異なる

集合を比較しやすくするために、各集合に色を付けることもできる。

図 5-16 は学生の学年によりデータエンティティをクラスタしたときの MMTAnalyzer のスクリーンショットである。2007 年に入学した学生は青色に付けられ、2008 年に入学した学生は赤色に付けられる。

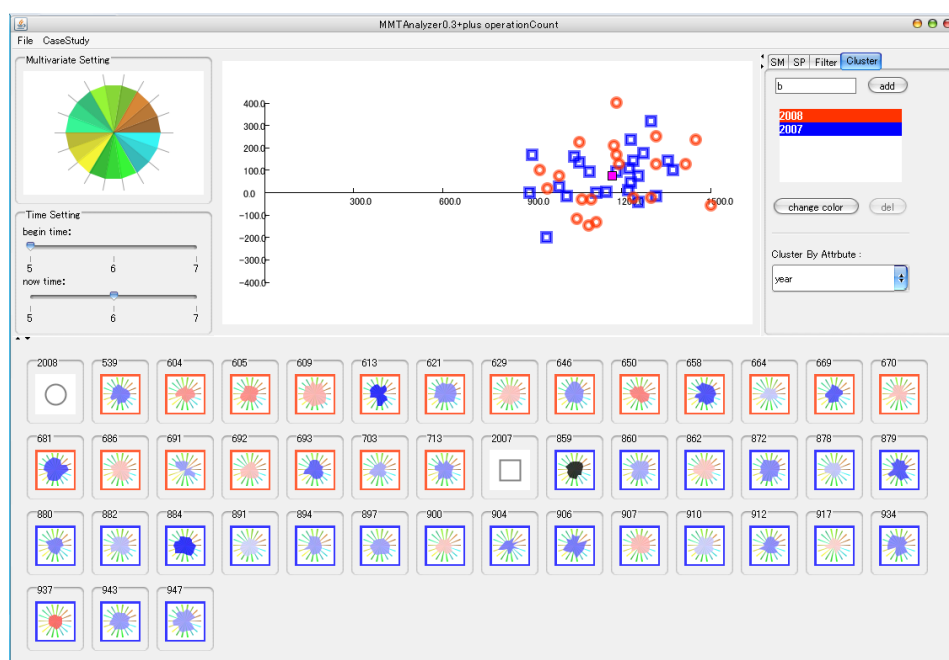


図 5-16 属性によりクラスタ分析機能

5.3.6 単一エンティティ分析機能

本研究では、四つの多変量時系列情報の視覚的表現を用いて一つのデータエンティティに対して詳しく分析できる機能を開発する。

small multiples の任意のデータエンティティの子パネルにダブルクリックすると、四つの異なる視覚的表現を使った子パネルが表示される。また、このデータエンティティが散布図における位置も同時にハイライトされる。図 5-17 は一つのデータエンティティに対して四つの視覚的表現を使って分析するときの MMTAnalyzer のスクリーンショットである。

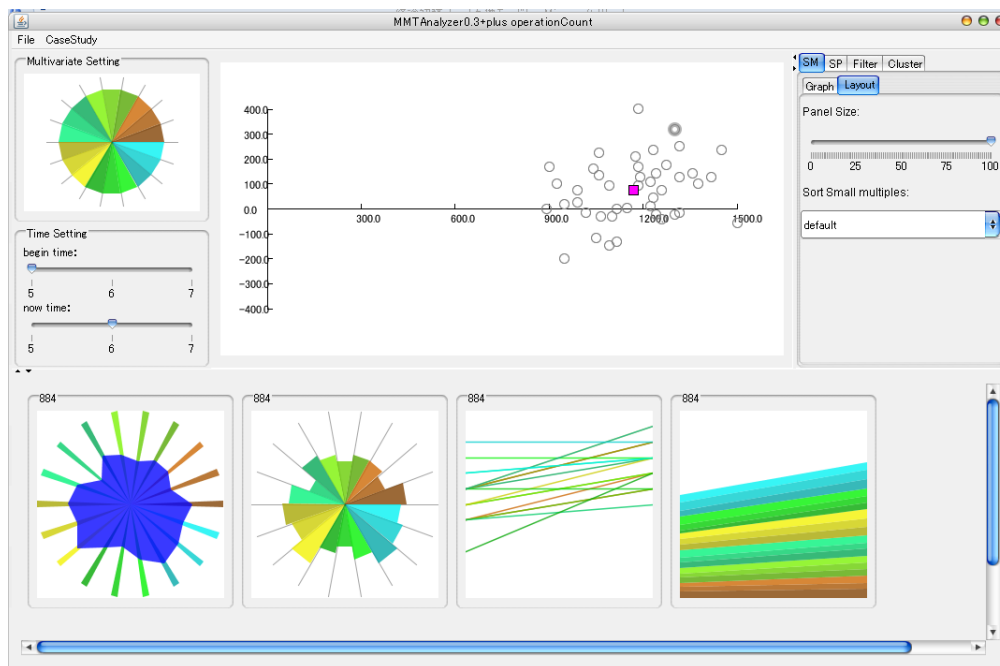


図 5-17 単一エンティティ分析機能

5.4 実装

5.4.1 実装言語

本研究では、多エンティティ多変量時系列データの視覚的分析ツール MMTAnalyzer を開発した。実装の言語としては、Java™ Platform, Standard Edition 6 Development Kit¹を用いた。

5.4.2 データの読み込み

多エンティティ多変量時系列データの読み込みは XLS 形式で行っている。XLS ファイルの読み込み部分には Java アプリケーションから Excel や Word などの Microsoft 製品のフォーマットファイルを読み書きするための API である Apache POI²を利用した。

¹ <http://java.sun.com/>

² <http://poi.apache.org/>

第6章 ツールの利用例

ここでは、学生の成績データに対しての分析をツールの利用例として挙げる。

今回の利用例は、43名の学生の3学期の18科目に対しての成績データを用いる。18科目は6個のカテゴリに分ける。各カテゴリに三つの科目が割り当てられる。43名の学生は学年、性別、国籍などの属性を持つ。ここでは、留学生（日本人ではない学生）の過去の学習成績について分析することにより、留学生に対する教育方針を考えるという場面を想定する。

まず、留学生の全体学生での分布を見たいので、「Cluster」パネルの「Cluster By Attribute」のコンボボックスから「nationality」を選択して、国籍による学生のクラスタを作成する。次に、「Cluster」パネルの中のリストボックスから日本人以外のクラスタを選択して、「Filter」パネルの「Filter Data」ボタンを押して、留学生だけの small multiples での子パネルを残すことにする。変量と時間の指定パネルの時間を指定する二つのスライダーを調整しながら、各学期での散布図を見た上で、937の学生は2学期の時に全体の学生の中で、成績が悪く、1学期よりも成績が大きく下がっていることが分かる（図 6-1）。

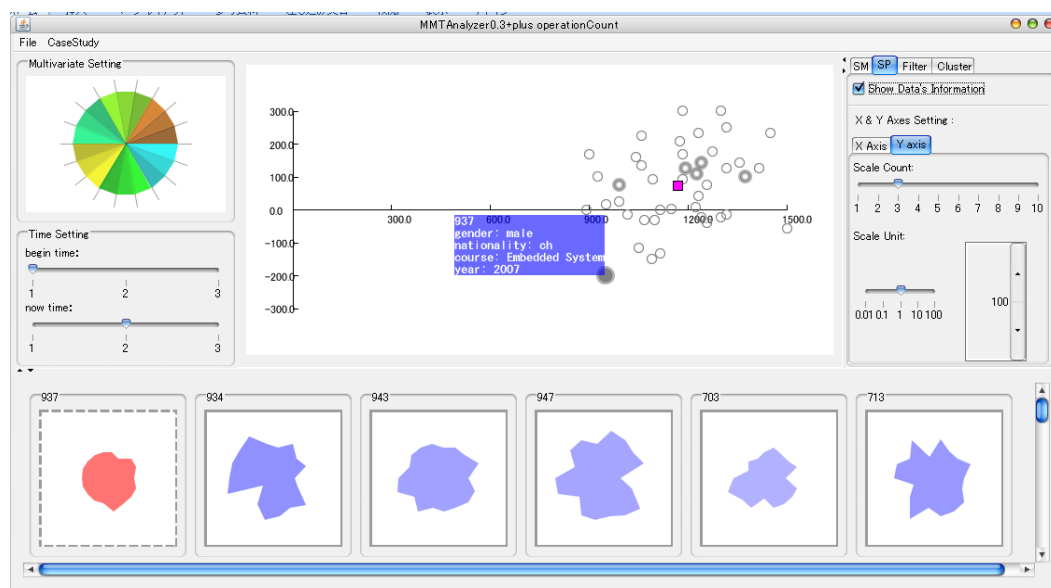


図 6-1 散布図を用いた分析

次に、留学生の各学期での成績の変化を見たいと考え、small multiples の各子パネルの可視化手法として「radar chart」を選択し、また、時系列の情報も視覚的に表現できるような拡張手法を選択する（図 6-2）。

図 6-2 が示したように、一つの radar chart の輪郭の色は赤色以外すべての radar chart の輪郭の色は青色或いは黒色となっている。以上のことにより、937の学生以外の留学生はすべての学期において成績が進歩していたということが分かる。

なお、各 radar chart の輪郭を観察してみると、科目 A について全員があまり良い成績を取得出来ていないことが分かる。科目 A について、より詳しく分析しようと考え、変量と時

間の指定パネルを使って科目 A 以外の変量を非選択にする。そうすると、散布図での各データエンティティが重なってしまう。「SP」パネルを使って、横軸と縦軸の目盛りの単位や数を調整して、各データエンティティを重ならないようにする。また、一つの変量の変化を分析しやすくするために、small multiples の子パネルの視覚的表現を「stacked graph」に変える（図 6-3）。

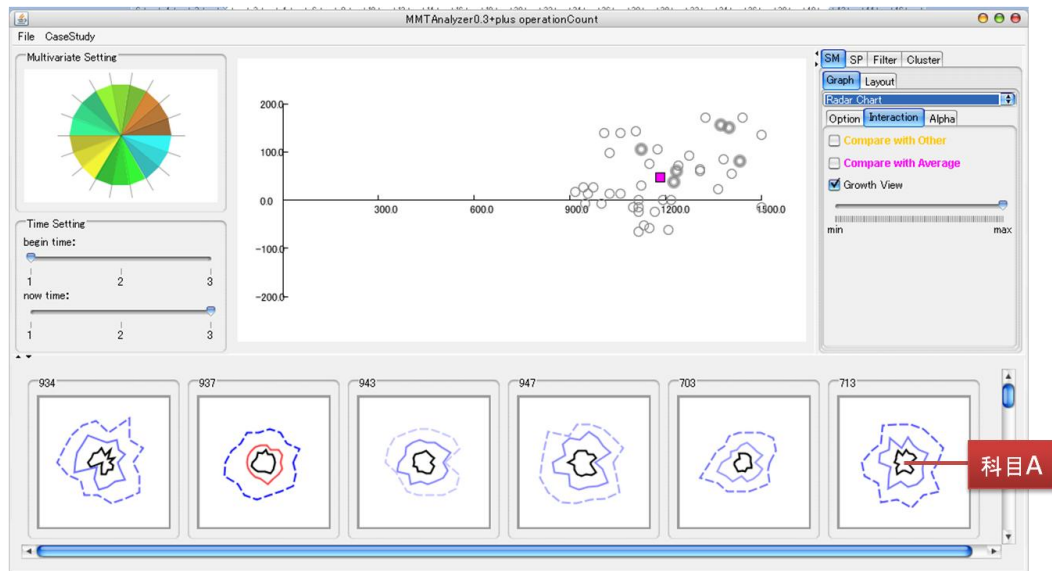


図 6-2 radar chart を用いた分析

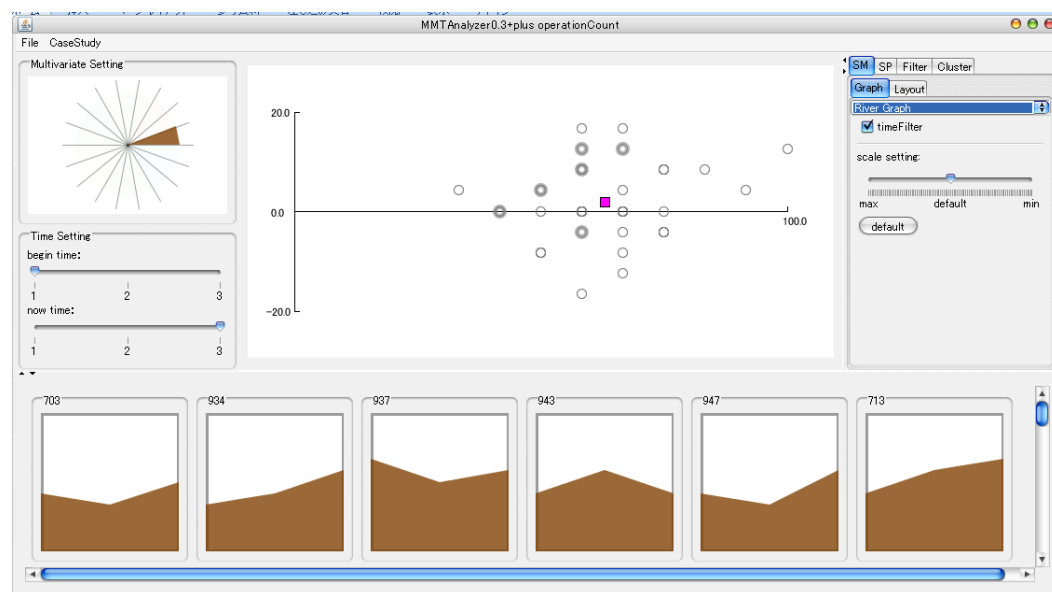


図 6-3 一つの変量に着目した分析

図 6-3 の散布図を観察すると、殆どの留学生が散布図の原点より上側に位置している。このことから、最後の 3 学期の時に、殆どの留学生の科目 A に対しての成績が最初の 1 学期より進歩したということが分かる。また、殆どの留学生が平均点より左側に配置されていることから、科目 A が留学生にとっての弱点だと考えられる。さらに、small multiples の各留学生の子パネルを見ると、各留学生の stack graph の一番右側の値が最大値の半分程度であることから、科目 A に対して進捗する余地がまだあることが分かる。以上のことにより、次の学期の留学生の教育に対して、科目 A についての教育を強め、また、新しい留学生が入る際に科目 A についての教育も重視するという方針を考える。

先の一時期に成績が下がった 937 の留学生の small multiples においての子パネルをダブルクリックすることにより、四つの視覚的表現を同時に使ってこの留学生を分析する。また、最初時刻を 1 学期に、最終時刻を成績が下がった 2 学期に指定した (図 6-4)。Sector graph の外側の各ブロックを見ると、一つの科目の成績が 1 学期より上がっているが、他の科目の成績がすべて 1 学期より下がった或は同等であることが分かる。

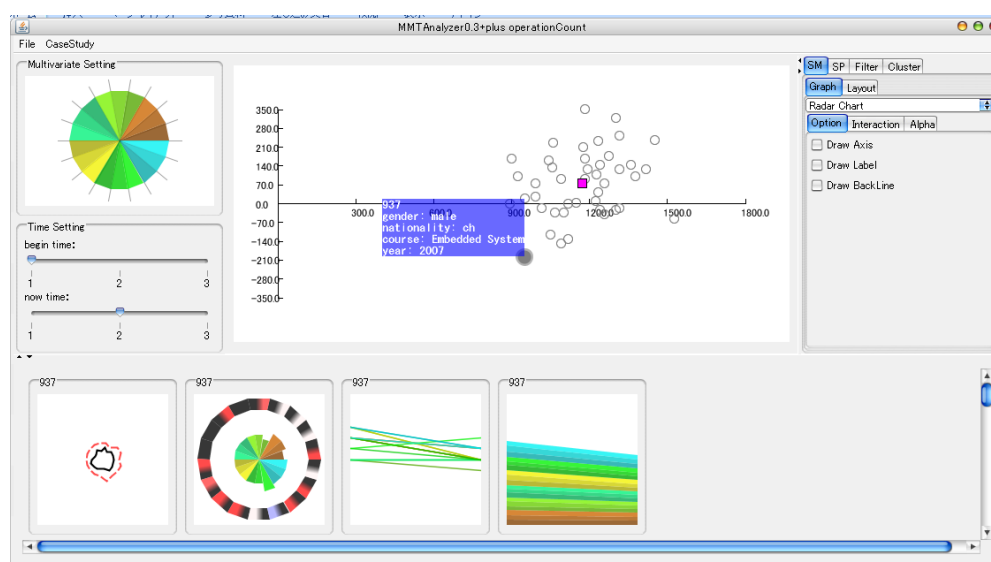


図 6-4 単一エンティティ分析

また、最近の成績の進捗状況を見たいので、最初時刻を 2 学期に、最終時刻を 3 学期に指定した (図 6-5)。Sector graph の外側の各ブロックと折れ線グラフを見ると、2 学期に成績が下がった科目は 3 学期に全部上がったことが分かる。さらに、散布図を見ると、937 の留学生が全体的に上側に位置していることから、937 の留学生は全学生の中でも成績がよく進歩していると考えられる。以上の分析から、「なぜ 937 の留学生は第 2 学期に成績が大きく落ちたのか」について疑問を持ち、同じようなことを他の留学生に繰り返さないように、937 の留学生と一度に話をすることにしたほうが良い。

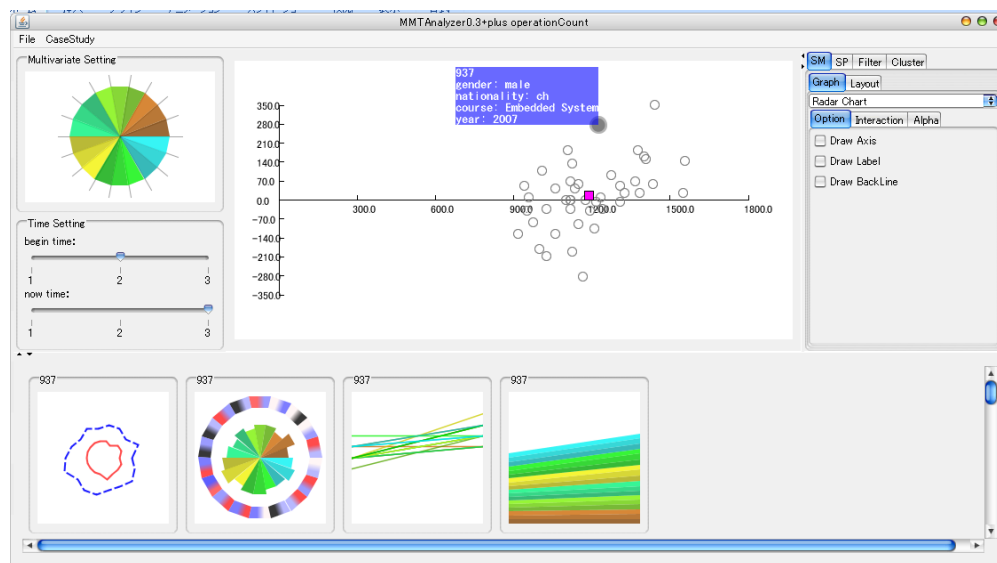


図 6-5 時間フィルターしてからの分析

第7章 ケーススタディ

本章では、本研究で開発したツール MMTAnalyzer を使用して実際の購買データを分析するケーススタディを通じて、本研究で開発したツールおよび視覚的表現や機能の有用性を示す。

7.1 データ

今回のケーススタディに用いたデータは著者が所属している研究室 IPLAB¹で運用している購買システムの2010年5月から2010年11月までの七ヶ月間の購買履歴データである。

著者が所属している研究室の学生は大学院生、学部生、研究生で構成されている。学生は四つの部屋（SB1024、3C316、SB1224、3E107）に分かれて活動している。この部屋の内、SB1024と3C316に買い物システムで稼働している。普段学生たちはこの購買システムを使って、飲料水、お茶やお菓子などの六つのカテゴリの商品の購買を行っている。

購買システムから取った生データは表 7-1 のようになっている。ここで、実際の購入者の名前を二文字のアルファベットに置き換えている。本ツールで扱える形式にするため、生データを集計し、月と購入者ごとの集計値として形式に整理した。ここで、「購入者」は多エンティティをなし、「商品のカテゴリ」は多変量をなし、買い物の数の月ごとの変化は時系列をなす。整理したデータは表 7-2 のようになる。

表 7-1 生データ

stimestamp	name	productname	category	room	team	home
2009-05-29 11:49:16+09	HB	BOSS レインボーマウンテン	コーヒー	3C302	NAIS	中国
2009-05-29 11:50:16+09	TU	BLEND COFFEE 微糖	コーヒー	3C316	NAIS	茨城県
2009-05-29 12:01:16+09	TI	BOSS レインボーマウンテン	コーヒー	3C316	NAIS	愛知県
2009-05-29 12:09:30+09	SD	BOSS レインボーマウンテン	コーヒー	3C316	NAIS	宮崎
2009-05-29 15:34:31+09	SD	BLEND COFFEE 微糖	コーヒー	3C316	NAIS	宮崎
2009-05-29 15:35:45+09	SD	BLEND COFFEE 微糖	コーヒー	3C316	NAIS	宮崎
2009-05-29 16:57:54+09	YS	BLEND COFFEE 微糖	コーヒー	3C316	NAIS	岡山
2009-05-29 20:22:32+09	TU	ポテトチップスリッチコンソメ	菓子類	3C316	NAIS	茨城県
2009-05-29 21:35:37+09	TI	BOSS レインボーマウンテン	コーヒー	3C316	NAIS	愛知県
2009-05-30 13:48:06+09	HB	日清 からしマヨネーズ やきそば	インスタント食品	3C302	NAIS	中国
2009-05-30 15:33:04+09	OW	クランキーチョコレート	菓子類	3C316	NAIS	日本
2009-05-30 16:47:22+09	TH	BOSS レインボーマウンテン	コーヒー	SB1024	NAIS	石川県
2009-06-01 13:29:23+09	TU	BLEND COFFEE 微糖	コーヒー	3C316	NAIS	茨城県
2009-06-01 13:37:27+09	YX	BLEND COFFEE 微糖	コーヒー	3C316	NAIS	中国
2009-06-01 14:54:07+09	TH	BOSS レインボーマウンテン	コーヒー	SB1024	NAIS	石川県
2009-06-01 16:20:29+09	TU	茶織	茶	3C316	NAIS	茨城県
2009-06-01 17:09:43+09	TU	Meiji Milk Chocolate	菓子類	3C316	NAIS	茨城県
.....						

1 <http://www.iplab.cs.tsukuba.ac.jp/index-j.html>

表 7-2 整理した後のデータ

name	time	清涼 飲料水	コーヒー	茶	水	菓子類	インスタント 食品	栄養 補助食品	その他
AM	2010-05	0	0	0	0	0	0	0	0
AM	2010-06	0	0	0	0	0	0	0	0
AM	2010-07	6	0	1	4	1	2	0	0
AM	2010-08	8	0	12	5	2	3	0	0
AM	2010-09	21	16	11	4	12	2	0	0
AM	2010-10	22	25	17	0	8	6	0	1
AM	2010-11	24	13	8	0	9	6	0	0
.....									

7.2 ケーススタディ設計

今回のケーススタディでは、購買システムの仕入れを担当している「TI」にツールを利用してもらった。

今回のケーススタディは以下の手順で行う。まず、ツールの使い方を説明する。次に、分析者に自由にツールを使ってもらう。また、時間制限を設けない。分析者がケーススタディを行う時に、分析者の操作を観察し、分析者からのツールについての質問に答える。全部の分析作業が終わった後に、分析者のツールに対する感想を聞く。

また、今回はケーススタディを行う時のツールの各機能の操作を記録し、ケーススタディの全過程を録画する。さらに、分析から得られた有益な情報を紙に記録してもらい、当時のスクリーンショットも取ってもらう。

7.3 ユーザの分析

今回のケーススタディはツール説明の 20 分を含めて、2 時間程度で行った。ここで、分析者が MMTAnalyzer を使って行った一連の分析を紹介する。

分析者は、まず、今回の購買データの概観を見たいと考え、「SM」パネルの中の「Layout」パネルの「Panel Size」のスライダーを使って、全ての購入者の small multiples での子パネルを一画面で表示できるように各子パネルのサイズを小さくした。その次に、ツールの左上の変量と時間の指定パネルの上側の変量フィルターパネルから全部の変量が選択されたことを確認し、次に、下側の時間指定スライダーを使って最初の時刻を 5 月に指定し、最後の時刻を 11 月に指定した。最後に、small multiples の各子パネルの視覚的表現を「Line graph」にした（図 7-1）。分析者は、散布図中で他の人と離れている購入者「TO」を発見した。また、「TO」の子パネルの line graph を確認したところ、「TO」の買い物数が連続に増加していることが分かった。

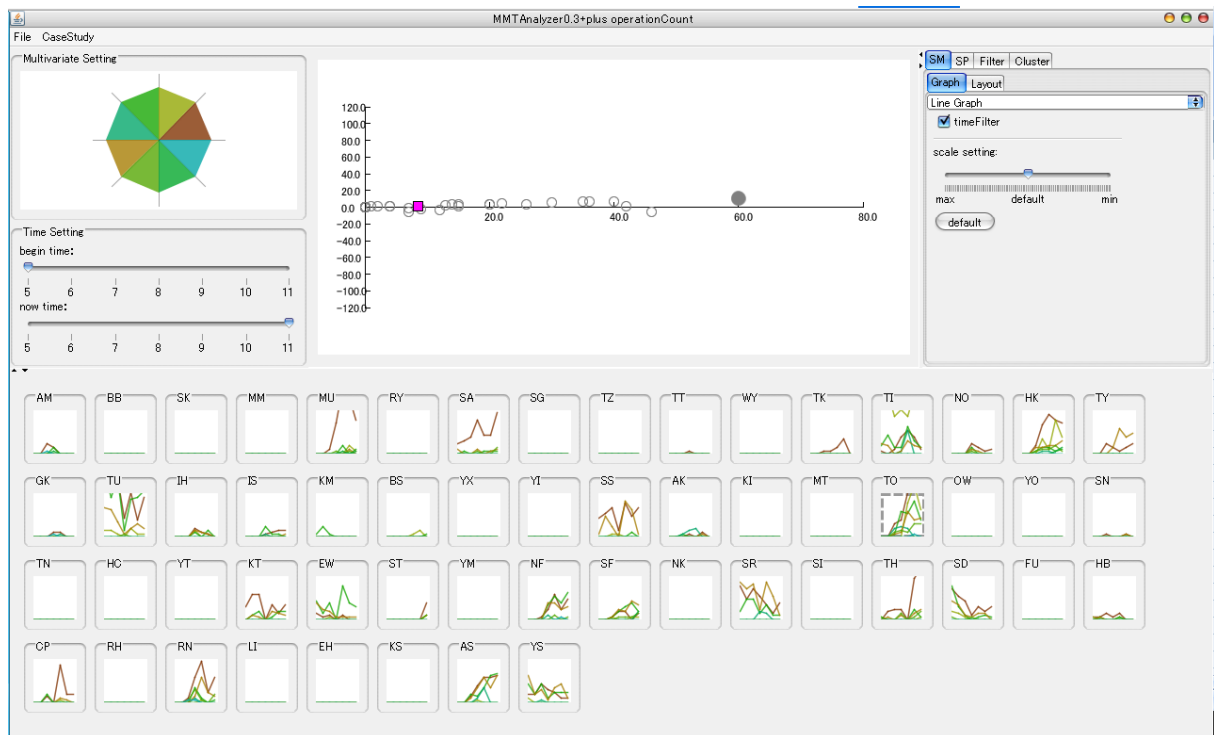


図 7-1 ケーススタディデータのオーバービュー

他の「TO」と同じく大量に商品を買っている購入者の状況を確認したいので、散布図の右側のいくつかのエンティティを選択した。また、各購入者の購入数の変化をもっと見やすくするために、子パネルの視覚的表現を「Stack graph」にした（図 7-2）。各購入者の stack graph を見ても、同じ特徴を持つ購入者を発見できなかった。しかし、選択された購入者の以外の何人かの small multiples での子パネルに興味を引いた。「SS」、「SR」、「SD」の stack graph の面積も大きいことから、商品を多く購入していたことが分かった。また、stack graph の形から、この三人は最近ではあまり商品を購入してなくなったことが分かる。特に、三人の中の「SD」は、ある時期に買い物の数が急激に減少したことが分かった。

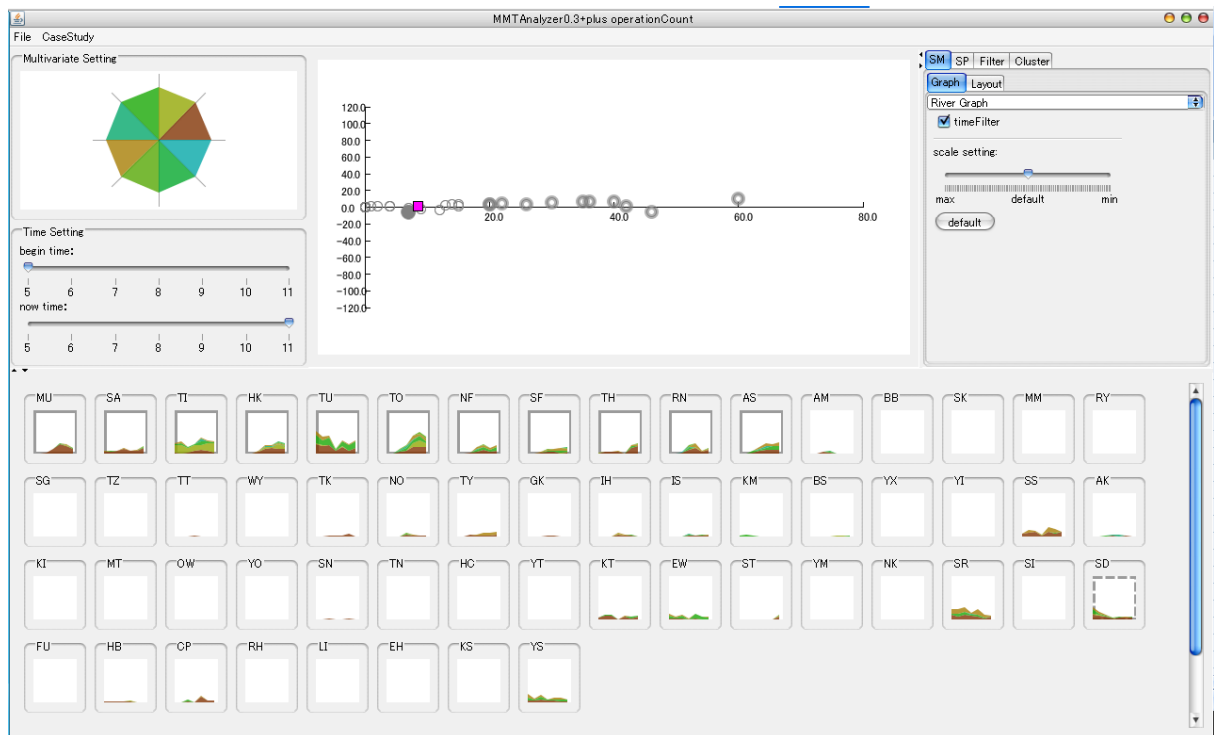


図 7-2 買い物数が多い人の分析

分析者は、次に、部屋ごとの買い物の状況を分析したいので、購入者を部屋ごとにクラスターに分けた。分析者は、SB1024 と 3C316 に買い物システムしか稼働していないことを知っており、SB1024 と 3C316 の購入者に着目して分析しようとしていた。各クラスターを区別しやすくするために、「SB1024」の人を赤い色を付け、「3C316」の人を青い色を付けた。また、各購入者の散布図の縦軸での分布をもっと分かりやすくするために、散布図の縦軸の目盛りの単位を小さくした（図 7-3）。各部屋の購入者の散布図から、SB1024 の購入者の購入数が上昇傾向であり、3C316 の購入者の購入数が減少傾向であることが分かった。

また、各部屋の商品カテゴリごとの違いを分析したいと考え、商品カテゴリのフィルターを行った。まず、飲み物（「清涼飲料水」、「茶」、「コーヒー」、「水」）に対して、「コーヒー」と「水」をフィルターした後の散布図が変わらないことから「コーヒー」と「水」があまり売れていないことが分かった。また、「茶」と「清涼飲料水」だけ残した散布図から、SB1024 の購入者と 3C316 の購入者の購入数の傾向が全体の購入数の傾向と変わらないことが分かった。次に、食べ物（「菓子類」、「インスタント食品」、「栄養補助食品」）に対して、散布図から SB1024 の購入者と 3C316 の購入者の購入数の傾向が全体の購入数の傾向と変わらないことが分かったが、small multiples の子パネルの stack graph から、3C316 の食べ物の売れ行きに波があり、SB1024 の食べ物の売れ行きが比較的に安定していることが分かった（図 7-4）。さらに、最初時刻と最後時刻を変えながら、9月から11月までの3ヶ月間に3C316 の購入者の食べ物の売れ行きが落ちていることも分かった。

分析者は、3C316 の食べ物の売れ行きに波がある原因として在庫切れと考え、安定な供給が売り上げ向上に繋がると考えた。また、small multiples の各購入者の子パネルを見たら、食べ物をよく買う購入者「SS」、「SR」、「SD」、「TU」、「YS」を食べ物のメインターゲット

と考えた。

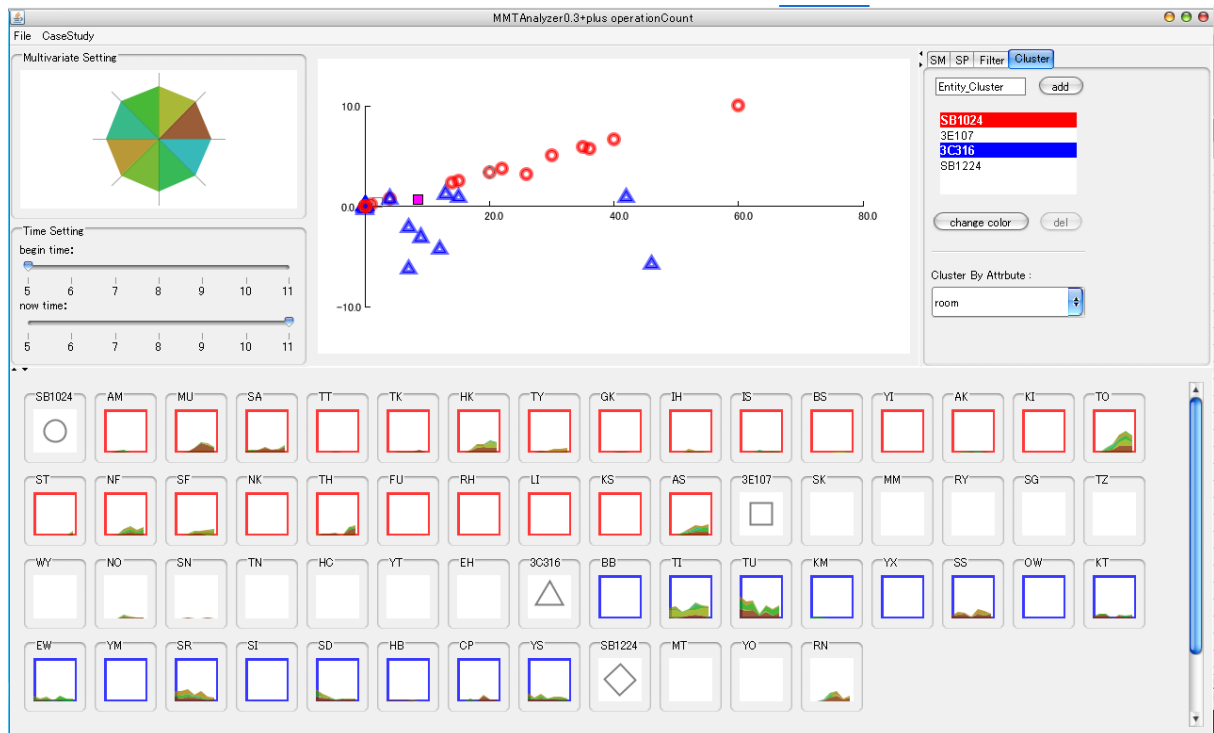


図 7-3 部屋ごとの比較分析

全体的な購入数に殆ど影響がない「コーヒー」と「水」についての分析も行った。商品カテゴリのフィルターを行い、「コーヒー」と「水」だけ残して分析したところ、各購入者の small multiples での子パネルから「コーヒー」について「TI」、「HK」、「TO」しか殆ど買っていないことが分かった。また、「水」について、夏ごろ一時的によく売れたことも分かった。

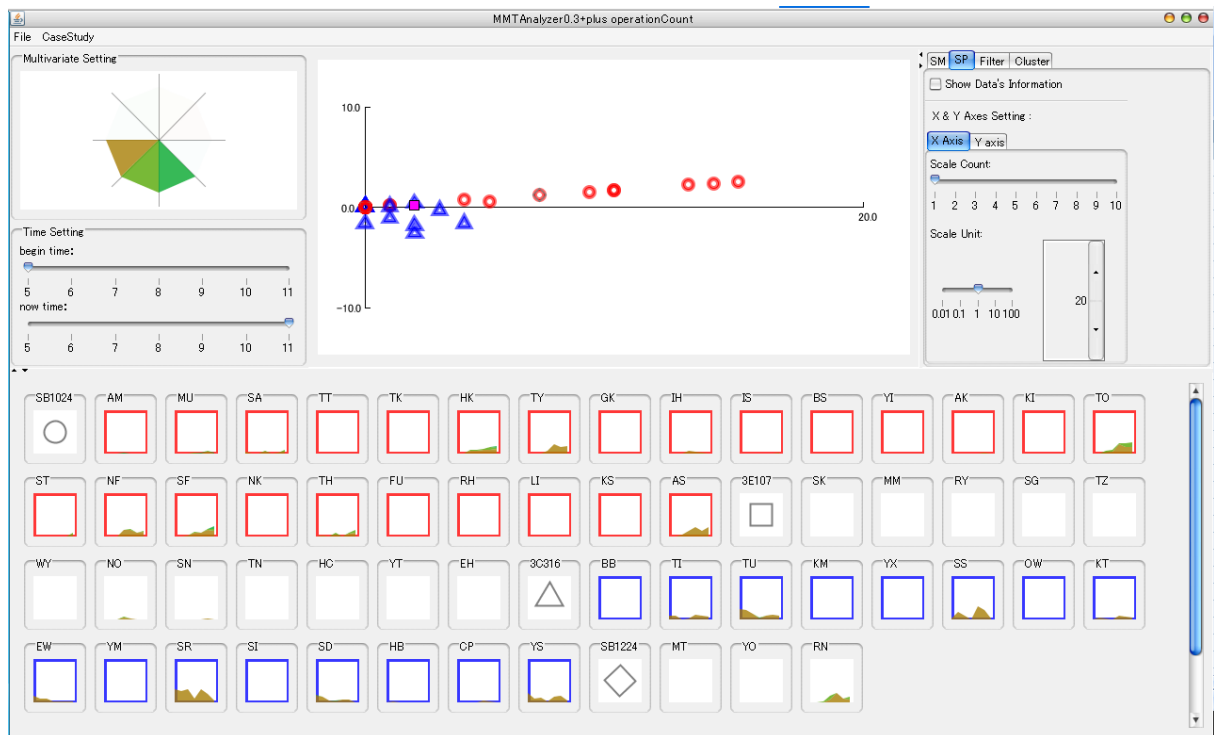


図 7-4 食べ物の売れ行きに着目した分析

次に、分析者は 3C316 の各購入者の買い物の特徴について分析した。分析者は、散布図から 3C316 の買い物の数が増加した人「TI」、「SS」、「KT」三人を選択して、そして、small multiples の子パネルの視覚的表現を「Line graph」にした。しかし、線の重なりが分析の支障となり、分析者は次に「Stack graph」にした。また、下に積み上げたスタックの形が上側のスタックの形に影響を与え、上側にある複数のスタックの変化を同時に分析するのは困難であるので、分析者は small multiples の子パネルの視覚的表現を「Sector graph」に切り替えた。Sector graph の外側の各ブロックの色がはっきり見えるように、small multiples の子パネルのサイズを大きくした (図 7-5)。「KT」と「SS」の「清涼飲料水」のセクタのサイズが大きく、また、最近の月の色が青くなっていることから、分析者はこの二人の購入数が増加した原因が「清涼飲料水」の購入が増えたためだと考えた。分析者「TI」は自分が殆ど「コーヒー」を買っていると思っていたが、「Radar chart」に切り替えて見ると (図 7-6)、一時的に他の飲みものも買うようになっていたことに気づいた。時刻を調整すると、9 月の夏の時に自分が「コーヒー」以外の「水」、「茶」など飲み物も買っていたことを思い出した。

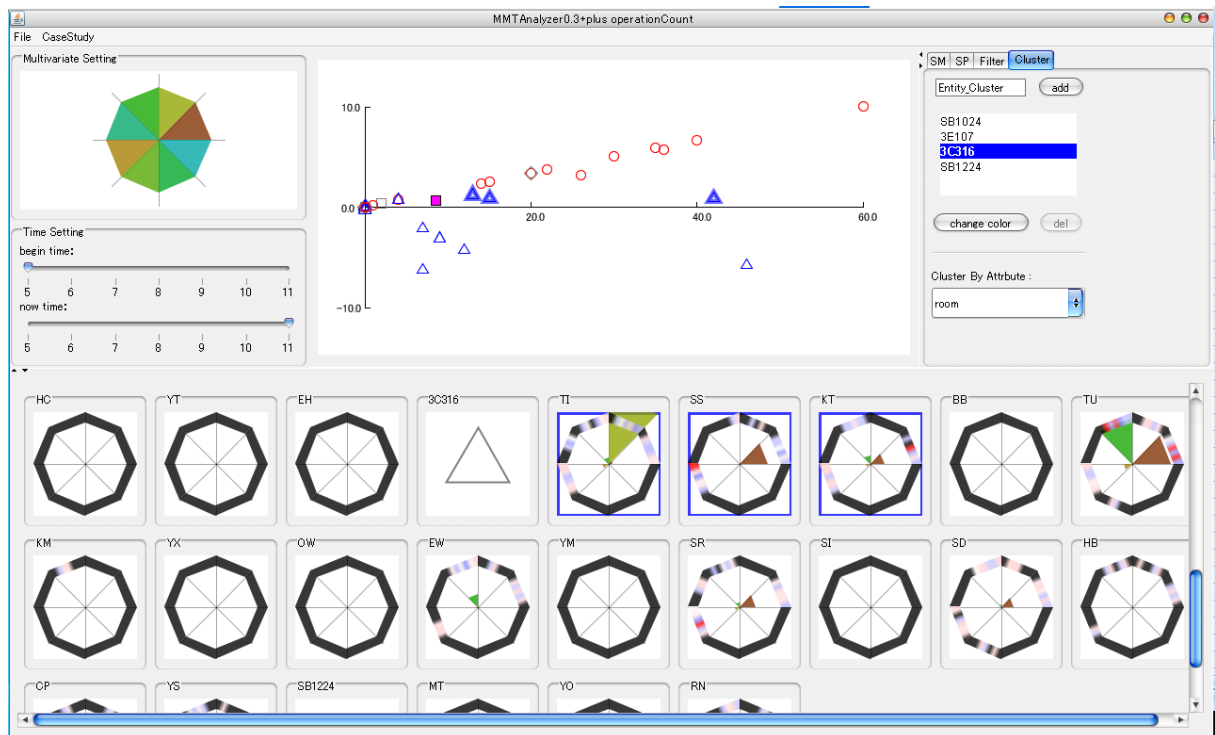


図 7-5 増加した人に着目した分析

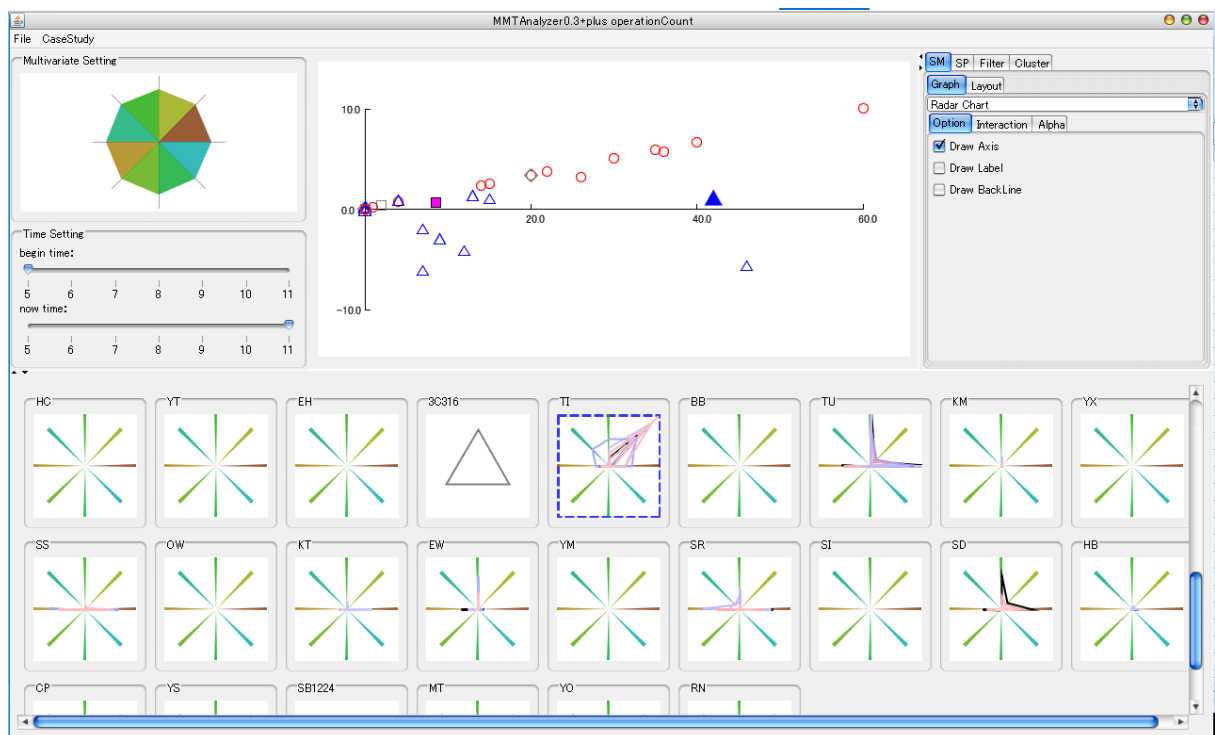


図 7-6 radar chart を使った分析

購入数が減少した 3C316 の購入者「EW」、「SR」、「SD」、「YS」に対して、「EW」は「茶」を中心に買い物しており、「SR」、「SD」は「清涼飲料水」の購入が多くなり、「YS」が「清涼飲料水」のかわりに「茶」の購入が多くなってきたことも分かった。

その後、時刻を調整し、連続した 2 月ごとの分析をした所、7 月と 8 月の散布図を見ると、3C316 と SB1024 の購入者は横軸により上下二つの部分に分かれたことを発見した(図 7-7)。つまり、7 月と比べると 3C316 の購入者の購入数が 8 月に減り、一方、SB1024 の購入者の購入数が増えたことが分かった。このことから、分析者は、3C316 の購入者は夏休みを取る人が多いと推測した。また、small multiples の各購入者の sector graph を見ると、SB1024 では「清涼飲料水」が売れていたことも分かった。しかし、3C316 の人の中に一人の買い物の数が増加したことも分かる。それは分析者自身だった。そして、「コーヒー」ばかり買っていたことも small multiples の sector graph から分かった。

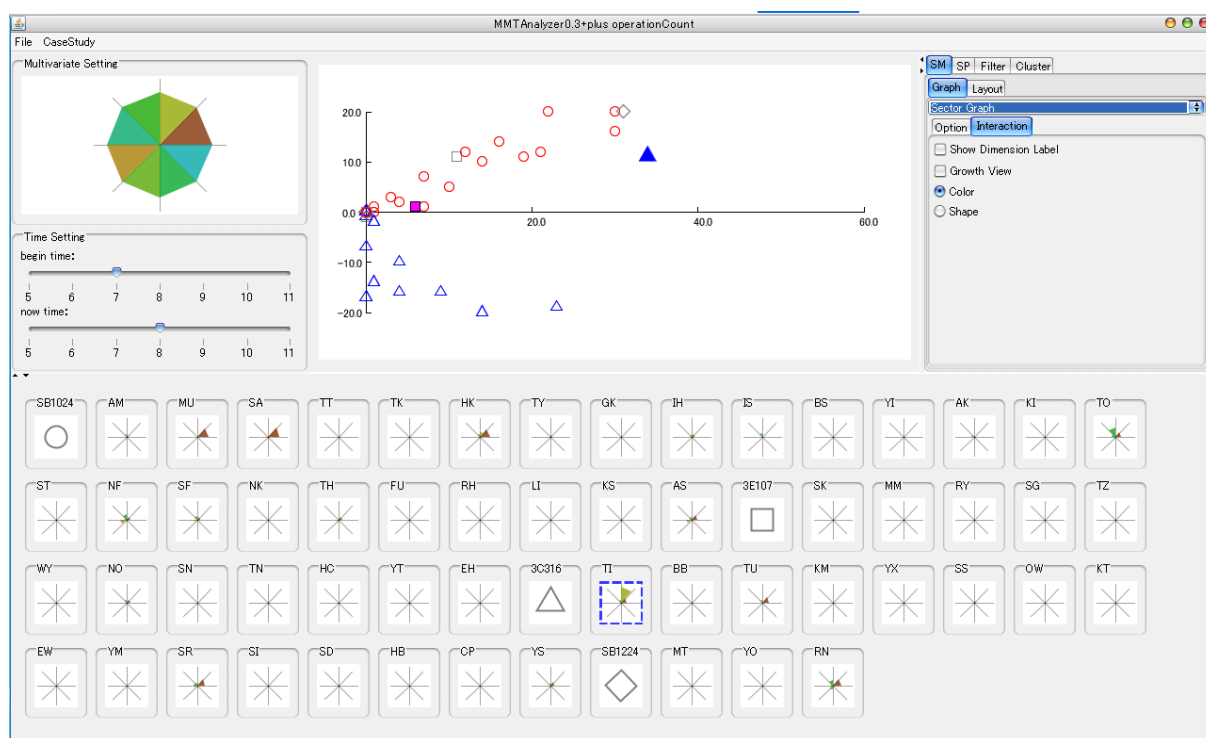


図 7-7 月ごとに着目した分析

7.4 考察

今回のケーススタディを通じて、以下の三つの事項について考察を行う。

1. ツール MMTAnalyer の有用性
2. 可視化手法及び機能の有用性
3. 散布図と small multiples の組み合わせの有用性

7.4.1 ツールの有用性

今回のケーススタディでは、大まかに以下のように分析が行なわれた。まず、概観を通じて、研究室全体の購買状況を把握した。次に、部屋ごとの購買者の購買状況を比較しながら分析を行った。最後に、3C316 の各人の買い物の特徴の分析を行った。以上の各段階で有益な情報が得られた。分析者の実際の使用状況から本研究で開発したツールがエンティティに着目した多エンティティ多変量時系列データの視覚的分析を支援することに対して有用であると言える。

分析の後に、分析者からツールについて、コメントをもらった。「自分が分析したいところに対して本ツールちゃんと支援できた。」とのコメントを頂いた。「商品カテゴリごとの時間変化の比較が行えなかった」のコメントがあった。

7.4.2 視覚的表現及び機能の有用性

表 7-3 は各詳細情報の視覚的表現へ切り替える操作の回数を示している。今回のケーススタディでは、line graph と stack graph が sector graph と radar graph より多く使われたことが分かる。各データエンティティの多変量時系列情報を分析する際に、分析者がよく「line graph > stack graph > sector graph > radar graph」の順に、4つの視覚的表現を試しながら分析を行うことが観察された。また、sector graph と radar graph を用いて分析を行う際にも、両方の時系列情報の拡張表現が主に使われていた。各データエンティティの多変量時系列の性質により、変数の一つの時刻に対しての分析より時間変化に対しての分析のほうが望まれると考える。

また、複数の人のすべての変数の変化を分析したい場合に、ユーザは4つの視覚的表現を試した後に「sector graph」を選択した。Line graph に線の重なりがあり、stack graph に下側のスタックが上にあるスタックの形への影響がある。以上のことは視覚的分析の支障となる。一方、本研究で拡張した sector graph の拡張表現が全ての変数を重ならず時間変化の分析ができるという利点があると考えられる。分析者から、「sector graph の時系列情報の視覚的表現は、複数の人の詳細情報を一望するのに便利だった」というコメントももらった。

分析者は自分の radar chart の時系列情報の視覚的表現を見て、自分の買い物の特徴の変化に気づいた。このことから、本研究で拡張した radar chart の拡張表現がデータエンティティの全体の特徴の変化を視覚的に分析できると考えられる。

表 7-3 詳細情報の視覚的表現への切り替えの回数

記録操作コード	操作	回数
SM_ChooseGraph_Line	Line graphへの切り替え	7
SM_ChooseGraph_Stacked	Stacked graphへの切り替え	7
SM_ChooseGraph_Sector	Sector graphへの切り替え	5
SM_ChooseGraph_Radar	Radar chartへの切り替え	4

ツールの利便性を向上するために開発した手描き変数選択操作について、分析者に「手描き変数選択機能が便利だった」というコメントをもらった。また、実際の操作を見ると、属性によるクラスタ分析機能を使って部屋ごとの比較がよく行なわれていたことが分かった。

しかし、エンティティフィルター機能とエンティティ集合分析機能に対して、「使いづらかった」とのコメントがあった。これらの機能を実現するために、いくつかの操作をしなければならぬので、ユーザが使いこなすまで時間がかかってしまったためだと考える。

7.4.3 散布図と small multiples の組み合わせの有用性

今回のケーススタディでは、27回のスクリーンショットが取られた。スクリーンショットが取られる前の操作は以下の四種類に分ける

- 1). 散布図パネルだけに対する操作 (sp)
- 2). small multiples パネルだけに対する操作 (sm)
- 3). 両方交代しずつの操作 (sp&sm)
- 4). 両方とも行なわなかった操作 (!sp&sm)

図 7-8 は以上の四種類の操作の回数を表す。両方交代しずつ操作した回数は片方しか操作しなかった回数の大体 2 倍となる。また、分析者からも「散布図と small multiples 両方を用いながらの分析が多かった」のコメントももらった。以上のことから、散布図と small multiples の組み合わせは多エンティティ多変量時系列データの分析に対して有用であると言える。

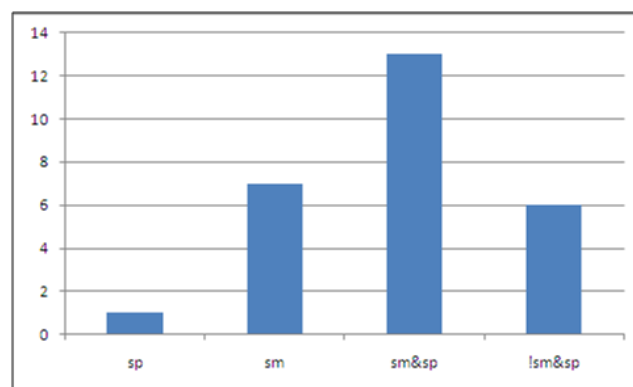


図 7-8 スクリーンショットを取る前の操作の回数

また、分析者の実際に分析する時のスクリーン録画から見ると、「まず、全体データエンティティの散布図から外れ値のデータエンティティを見つけ、さらに、small multiples を用いて詳細情報を分析する」と「まず、small multiples の子パネルの詳細情報から、特徴があるデータを見つけ、その次に、全体データエンティティの散布図での位置を確認しながら分析する」という分析流れが多かった。4.5 節で説明した視覚的分析流れに従った分析が見られた。

第8章 結論と今後の課題

8.1 結論

本研究では、多エンティティ多変量時系列データの視覚的分析を支援することを目的とし、従来手法の問題点としてエンティティに着目した視覚的分析を支援できていない点を挙げた。これを解決するアプローチとして、本研究では、まず、エンティティに着目した多エンティティ多変量時系列データの分析要件を整理した。次に、整理した分析要件を満たすために、全体データエンティティの分布と各データエンティティの詳細情報を同時に把握できる散布図と **small multiples** を組み合わせた可視化手法及び四つの詳細情報の視覚的表現を開発した。さらに、それを用いた多エンティティ多変量時系列データの視覚的分析ツール“MMTAnalyer”を開発した。操作の利便性を向上するために、平均値基準選択機能や手描き変量選択機能を開発し、また、分析要件を満たすために、集合エンティティ分析機能や属性によりクラスタ分析機能などのインタラクティブ操作を開発した。

実際の購買データを用いたケーススタディを通じて、本研究で開発したツールの有用性を示した。また、ケーススタディの分析者のコメント及び記録したデータから、本研究で開発した散布図と **small multiples** の組み合わせ表現が多エンティティ多変量時系列データの分析に対して有用であるとの結果を得た。

本研究で開発したツールにより、エンティティに着目した多エンティティ多変量時系列データの視覚的分析が可能になった。また、散布図と **small multiples** を組み合わせたことで全体データエンティティの一覧と各データエンティティの詳細情報を同時に把握することが可能になった。多エンティティ多変量時系列データだけではなく、他の多エンティティデータの視覚的分析にもこの手法も適応できると考えられる。

8.2 今後の課題

8.2.1 ツールの操作とインタフェースの改善

本研究で開発したツールでは、一つの機能を実現するために、複数の操作を行うことがよくある。複雑な操作の組み合わせによりユーザが全ての機能を使いこなすまでに時間がかかる。また、散布図パネルと **small multiples** パネルに対する操作を全て右上の調整パネルで行なわなければならない、操作するたびにパネルを頻繁に切り替え必要がある。

それを改善するために、二つの対策が考えられる。一つは一つの機能を実行するための操作回数を減らす。自動的に多変量時系列の情報によりデータエンティティをクラスタリングするアルゴリズム[29]を導入することにより、ユーザを面倒な選択やクリック操作から解放できると考えられる。もう一つは、操作のパネルをもっとユーザにとって分かりやすく配置することである。

8.2.2 大規模データへの対応

現在、大規模データへの要求が高まっている。また、今回のケーススタディで分析者が常に各エンティティの詳細情報を一画面で閲覧できるようにしながら分析する傾向が見られた。しかし、**small multiples** では、同時に複数のデータエンティティを表現するときに、多くの表示スペースが必要となる。

多くのデータエンティティの詳細情報を表示するためには、**small multiples** の子パネルのサイズを小さくしなければならない。各子パネルのサイズを小さくすると、詳細な分析ができなくなる恐れがある。もう一つの解決法としては **small multiples** パネルのサイズを大きくする。しかし、ツールの表示スペースも限られている。

本研究では、ユーザが選択したデータエンティティの子パネルを先頭に詰めて表示するようにしているが、限られたスペースの中に多くのデータエンティティの詳細情報を同時に表現するのは本研究の一つ今後の課題となる。

謝辞

私に日本に留学する大変貴重な機会を与えていただき、また、本研究を行うにあたり研究の進め方から論文の執筆に至るまで、多変貴重なご助言、ご指導をいただいた指導教員である三末和男准教授に心より感謝を申し上げます。また、研究に取り組む姿勢についても親身にご指導いただき、その中で多くのことを学ばせていただきました。この修士論文を書き上げることができたのはもちろん、人間的にも大きく成長したと実感できるのは、ひとえに先生のご指導のおかげです。この場を借りて厚く御礼を申し上げます。

田中二郎教授には、研究内容などについて終始丁寧なご指導をいただきました。また、研究設備などの面でも大きな支援をしていただきました。先生のおかげで大変有意義で充実した研究生活を送ることができました。

高橋伸准教授、志築文太郎講師には、研究発表や論文チェックなどの場で大変貴重なアドバイスをいただきました。先生方にいただいたご助言は大変有意義で、研究を大きく進歩させるものでした。深く感謝の意を申し上げます。

インタラクティブプログラミング研究室の皆さまには大変お世話になりました。特に、NAIS チームの皆さまに感謝してもしきれません。皆さまのおかげで 2 年間の研究室生活は有意義で忘れられないものになりました。深く感謝いたします。

何よりも家族の支え無くしてこのように充実した研究生活を送ることができませんでした。精神面、金銭面で多大なる支えをしてくれた両親に心から感謝します。

最後に、様々な面でお世話になった多くの友人、大学生活でお世話になった全ての方々に心より感謝申し上げます。本当にありがとうございました。

参考文献

- [1] B. B. Bederson and B. Shneiderman . The Craft of Information Visualization: Readings and Reflections. 2003.
- [2] L. Teng-Yok and S. Han-Wei. Visualization and Exploration of Temporal Trend Relationships in Multivariate Time-Varying Data. IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 1359-1366, 2009.
- [3] H. Akiba and K.-L. Ma. A tri-space visualization interface for analyzing time-varying multivariate volume data. In EuroVis '07: Proceedings of the Joint Eurographics-IEEE VGTC Symposium on Visualization 2007, pp. 115-122, 2007.
- [4] M. ten Caat, N.M. Maurits, and J.B.T.M. Roerdink. Design and Evaluation of Tiled Parallel Coordinate Visualization of Multichannel EEG Data. IEEE Transactions on Visualization and Computer Graphics, vol. 13, no. 1, pp. 70-79, 2007.
- [5] J. Blaas, C. Botha, and F. Post. Extensions of parallel coordinates for interactive exploration of large multi-timepoint data sets. IEEE Transactions on Visualization and Computer Graphics, vol. 14, no. 6, pp. 1436-1451, 2008.
- [6] J. Johansson, P. Ljung, and M. Cooper. Depth cues and density in temporal parallel coordinates. In EuroVis'07: Proceedings of the Joint Eurographics - IEEE VGTC Symposium on Visualization 2007, pp. 35-42, 2007.
- [7] G. Grinstein, M. Trutschl, U. Cvek. High-dimensional visualizations. 7th. ACM/SIGKDD Data mining Conference (KDD), 2001.
- [8] A. Inselberg. Multidimensional detective. In INFOVIS '97: Proceedings of the IEEE Symposium on Information Visualization '97, pp. 100-107, 1997.
- [9] C. Tominski, J. Abello, and H. Schumann. Axes-based visualizations with radial layouts. In SAC '04: Proceedings of the ACM Symposium on Applied Computing 2004, pp. 1242-1247, 2004.
- [10] C. Tominski, J. Abello, and H. Schumann. Interactive Poster: 3D Axes-Based Visualizations for Time Series Data. Poster Compendium of IEEE Symp. Information Visualization (InfoVis '05), 2005.
- [11] W. Muller, H. Schumann. Visualization methods for time-dependent data - an overview. Simulation Conference, 2003. Proceedings of the 2003 Winter, pp. 737-745 Vol. 1, 2003.
- [12] W. Aigner, S. Miksch, W. Muller, H. Schumann, C. Tominski. Visual Methods for Analyzing Time-Oriented Data. IEEE Transactions on Visualization and Computer Graphics, vol. 14, no. 1, pp. 47-60, 2008.

- [13] S. F. Silva, T. Catarci. Visualization of Linear Time-Oriented Data: A Survey. in: Proceedings of International Conference on Web Information Systems Engineering, pp. 310-319, 2000.
- [14] S. Havre, E. Hetzler, P. Whitney, L. Nowell. ThemeRiver: Visualizing Thematic Changes in Large Document Collections. IEEE Transactions on Visualization and Computer Graphics, vol. 8, no. 1, pp. 9-20, 2002.
- [15] M. Weber, M. Alexa, W. Muller. Visualizing Time-Series on Spirals. Proc. IEEE Symposium on Information Visualization 2001 (InfoVis '01), pp. 7-13, 2001.
- [16] F. Brian and J. Pritchard. Visualisation of historical events using Lexis pencils. In Case Studies of Visualization in the Social Sciences, 1997.
- [17] R. Rao, S. K. Card. The Table Lens: Merging Graphical and Symbolic Representations in an Interactive Focus+Context Visualization for Tabular Information. ACM CHI '94, 1994.
- [18] P. Roger. A Method for Visualizing Multivariate Time Series Data. Journal of Statistical Software, Code Snippets. pp. 1-17. vol. 25, no. 1, 2008.
- [19] P. Ordonez, M. desJardins, C. Feltes, C. U. Lehmann and J. Fackler. Visualizing Multivariate Time Series data to Detect Specific Medical Conditions. AMIA Annual Symposium Proceeding, pp. 530-534, 2008.
- [20] J. A. Fails, A. Karlson, L. Shahamat and B. Shneiderman. A visual interface for multivariate temporal data: Finding patterns of events across multiple histories. In VAST'06: Proceedings of the IEEE Symposium On Visual Analytics Science And Technology 2006, pp. 167-174, 2006.
- [21] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization. IEEE Symposium on Visual Language 96, pp. 336-343, 1996.
- [22] W. Javed, B. McDonnel, N. Elmqvist. Graphical Perception of Multiple Time Series. Visualization and Computer Graphics, IEEE Transactions on, vol. 16, no. 6, pp. 927-934, 2010.
- [23] Edward R. Tufte. The Visual Display of Quantitative Information. Graphics Press. 1983.
- [24] L. Byron, M. Wattenberg. Stacked Graphs----Geometry & Aesthetics. IEEE Transactions on Visualization and Computer Graphics, vol. 14, no. 6, pp. 1245-1252, 2008.
- [25] http://en.wikipedia.org/wiki/Radar_chart
- [26] D. A. Keim. Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics, Vol. 8, No. 1, pp. 1-8, 2002.
- [27] S. Eick and G. Wills. High interaction graphics. European Journal of Operational

Research, Vol. 81, No. 3, pp. 445-459, 1995.

- [28] A. Martin and M. Ward. High dimensional brushing for interactive exploration of multivariate data. In Proceedings of the 6th conference on Visualization '95, pp. 271-278, 1995.
- [29] A. Singhal, D. E. Seborg. Clustering of multivariate time-series data. American Control Conference, 2002. Proceedings of the 2002 , vol.5, pp. 3931-3936, 2002.