

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

大量の時刻情報付きデータの可視化手法の開発

白井 智子

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

2013年3月

概要

時刻情報付きデータとは、行動や事件などのイベントと、それらのイベントが発生した時刻を記録したデータである。このデータを分析すると、行動パターンがわかり、マーケティング分野やセキュリティ、リソースマネジメントなどの様々な分野へ応用することができる。本研究は、大量にある時刻情報付きデータの分析を支援するため、大量のデータの俯瞰を提供することを目的とする。この目的を達成するため、イベントのもつ時刻の集合を二次元平面上の位置で表す視覚的な表現手法である ChronoView を開発した。本研究では、Web へのアクセスや商品の購入、事故の発生などの動きがわかるものをイベントとする。イベントは、ひとつ以上の時刻で発生しているものとする。ChronoView は、アナログ時計の文字盤のような円を表示し、イベントの時刻の集合をその中に配置することで、データに記録されているイベントと時刻の関係性のおおよその分布を視覚的に提示する。ChronoView を適用した分析ツールは、用途に応じたインタラクションを提供することで、イベントと時刻の関係性についての詳細をさらに分析していくことを可能にする。本論文では、Twitter のツイートデータとファイアウォールログに着目したユースケースによって、本表現手法の有用性を示した。

目次

第1章 序論	1
1.1 時刻情報付きデータとは	1
1.2 時刻情報付きデータの分析における問題	2
1.3 研究の目的	2
1.4 本研究の貢献	2
1.5 本論文の構成	5
第2章 関連研究	6
2.1 データの俯瞰を提供する可視化	6
2.1.1 折れ線グラフを用いた手法	6
2.1.2 折れ線グラフを拡張した表現手法	6
2.1.3 タイルのような表現を用いた手法	7
2.1.4 色の変化を用いた手法	7
2.2 イベントの周期性を提示する可視化	8
2.2.1 螺旋を用いた表現手法	8
2.2.2 円グラフを用いた手法	8
2.2.3 円周に時間軸を持つ表現手法	8
2.3 データの俯瞰とイベントの周期性の両方を提供する可視化	8
第3章 大量データの可視化における要求	10
3.1 データの俯瞰	10
3.2 イベント発生の特徴の把握	10
第4章 ChronoView: 視覚的表現の提案	12
4.1 データの俯瞰とイベント発生の特徴の把握における要件	12
4.2 位置を使った時刻の集合の視覚的表現	12
4.3 各種属性の表現	15
4.4 付加情報の表現	15
4.5 金平糖表現	16
第5章 分析ツールの開発	19
5.1 ChronoView ウィンドウ	20
5.2 詳細画面	20

5.3	表現の操作画面	20
第6章	ユースケース	25
6.1	ファイアウォールログの可視化	25
6.1.1	利用データ	25
6.1.2	観察	27
6.1.3	考察	29
6.2	ツイートデータの可視化	29
6.2.1	利用データ	29
6.2.2	観察	30
6.2.3	考察	30
6.3	議論	33
第7章	まとめ	34
	謝辞	35
	参考文献	36

図目次

1.1	面グラフを 2×2 で並べた場合	3
1.2	面グラフを 5×5 で並べた場合	3
1.3	2 個のイベントを同じ軸上に並べた場合	4
1.4	10 個のイベントを同じ軸上に並べた場合	4
3.1	2 個のイベントを一度に表示した面グラフの例	11
3.2	10 個のイベントを一度に表示した面グラフの例	11
4.1	ChronoView の概観	13
4.2	線分上に時間軸を取った場合のイベントの配置	14
4.3	ChronoView で時刻の集合を配置した場合	15
4.4	実際にイベントが発生した時刻を示す放射状の線の表示	16
4.5	金平糖表現の表現方法	16
4.6	金平糖表現	18
5.1	分析ツールの概観	19
5.2	All View で気になるイベントをクリックした場合、放射状の線を表示する。	21
5.3	イベントを曜日ごとに色を塗り分けたビュー	22
5.4	イベントを平日と休日別に色を塗り分けたビュー	23
5.5	イベント「起床」をクリックした場合に表示される詳細画面の例	24
6.1	2011 年 1 月 1 日のデータの表示	28
6.2	1 つのイベントに着目した場合の表示	28
6.3	12 月 1 日からのデータを使用した表示	31
6.4	1 月 1 日からのデータを使用した表示	32

第1章 序論

1.1 時刻情報付きデータとは

本研究で対象とする「時刻情報付きデータ」とは、行動や事件などのイベントとそれらが発生した時刻を記録したデータである。たとえば、ある Web 閲覧者 (IP アドレス) が、いつ、どの Web ページにアクセスしたかを記録した Web アクセスログは、時刻情報付きデータである。時刻情報付きデータの一例を表 1.1 に掲げる。

本研究で扱うイベントは、Web へのアクセスや商品の購入、事故の発生などの動きがわかるものとする。イベントは、ひとつ以上の時刻で発生し、時刻の集合をもつものとする。イベントの一例として、Web ページへのアクセスをあげる。ある Web ページへのアクセスが、10 時、11 時、12 時に閲覧されたとすると、その Web ページは、3 つの時刻でアクセスされたことになる。この場合、ある Web ページへのアクセスを 1 つのイベントとすると、このイベントは、3 つの時刻で発生したことになる。

表 1.1: 時刻情報付きデータの例：Web 閲覧データの場合

日時	Web ページ アドレス
2013/01/10 10:00:20	http://www.iplab.cs.tsukuba.ac.jp/index-j.html
2013/01/10 11:00:09	http://www.iplab.cs.tsukuba.ac.jp/mlist.php
2013/01/10 12:00:12	http://www.iplab.cs.tsukuba.ac.jp/index-j.html
2013/01/10 12:10:18	http://www.iplab.cs.tsukuba.ac.jp/papers-j.html
2013/01/10 12:15:12	http://www.iplab.cs.tsukuba.ac.jp/international-j.html
2013/01/10 12:18:32	http://www.iplab.cs.tsukuba.ac.jp/paper/international/shiroi_iv2012.pdf
⋮	⋮

このデータ以外にも、ショッピングサイトでの商品の購買履歴、事件や事故の発生記録、SVN の管理ログなど、様々な時刻情報付きデータがある。

このような大量にある時刻情報付きデータを分析していくと、イベント発生の特定期刻への依存性や周期性などのイベントの特徴がわかる。これらの特徴の把握は、マーケティングやセキュリティ分野を初めとする様々な分野で期待されている。

1.2 時刻情報付きデータの分析における問題

データを可視化することは、データ分析を効率化する。時刻情報付きデータを分析するための可視化手法には、既に様々な提案がある [1, 2]。ただし、従来の可視化手法は時刻の集合の表現に関して、スケーラビリティ上の問題を抱えている。従来の表現手法の多くは時間軸を直線分で表現し、その線分上に個々のタイムスタンプをプロットする。例えば、図 1.1 や図 1.2 のような面グラフは、その一例である。この図は、横軸に時間軸、縦軸にイベントの発生頻度をとっている。これらの図は、Twitter 上で各時間帯にどのくらいつぶやかれているかを示したグラフである。特に、ここでは、今何をしているかを表す「新宿なう」、「カフェなう」などの「なう」を含むつぶやきを視覚的に提示している。グラフ上に書かれているキーワードは、「なう」の直前にある人間の行動にあてはまる単語である。ここで、グラフ上に書かれているキーワードは、本研究であつかうイベントである。このようなグラフ以外に、データの周期性を明示するために直線分の代わりに円周やらせんの時間軸を用いるものもある [3]。

いずれの手法においても、本研究で扱うイベント、すなわち時刻の集合を表現するためには、相応の面積を占める事になる。これでは、第 1 に、一度に多くのイベントをディスプレイ上で俯瞰することができない。例えば、数個から数十程度のイベント数であれば、ディスプレイ上に全てのイベントをおさめることができ、一度に俯瞰することが可能である。しかし、数千以上ものイベントがあるデータを同様に観察しようとしても、通常のディスプレイにはおさまらず、縦や横へのスクロールが避けられない。第 2 に、たとえ、ディスプレイ上に配置できたとしても、イベント間の特徴を把握することは困難である。例えば、図 1.3 と図 1.4 を見比べる。この図は、横軸に時間軸、縦軸にイベントの発生頻度をとっている。ここで、2 個のイベントを比較することは可能だが、10 個になると、グラフが繁雑化し、それぞれのイベントを見分けることが難しい。類似の時間的特徴を備えたイベントを探し出す、特定の周期性をもつイベントを探し出すといった作業に対する支援がなされないからである。

1.3 研究の目的

本研究は、大量の時刻情報付きデータの分析を支援するため、大量のデータを一度に俯瞰できるよう、データの俯瞰を提供することを目指す。特に本研究では、時刻情報を持つデータの俯瞰を提供するために、イベントが発生した時刻の集合に着目する。本研究で扱う時刻情報付きデータは、数千件以上のイベントを含んでおり、それらはいずれも時刻の集合に関連づけられている。

1.4 本研究の貢献

我々は、アナログ時計の文字盤のような円を表示し、時刻の集合をその中の位置で表現することで、データに記録されているイベント発生のおおよその分布を視覚的に提示する。本研究で提案する可視化手法を ChronoView と呼ぶ。

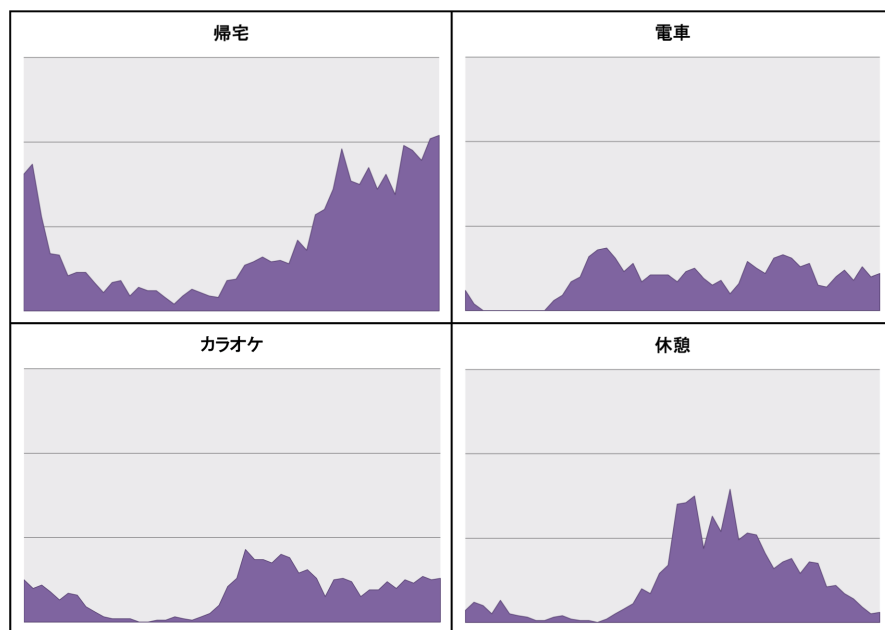


図 1.1: 面グラフを 2×2 で並べた場合



図 1.2: 面グラフを 5×5 で並べた場合

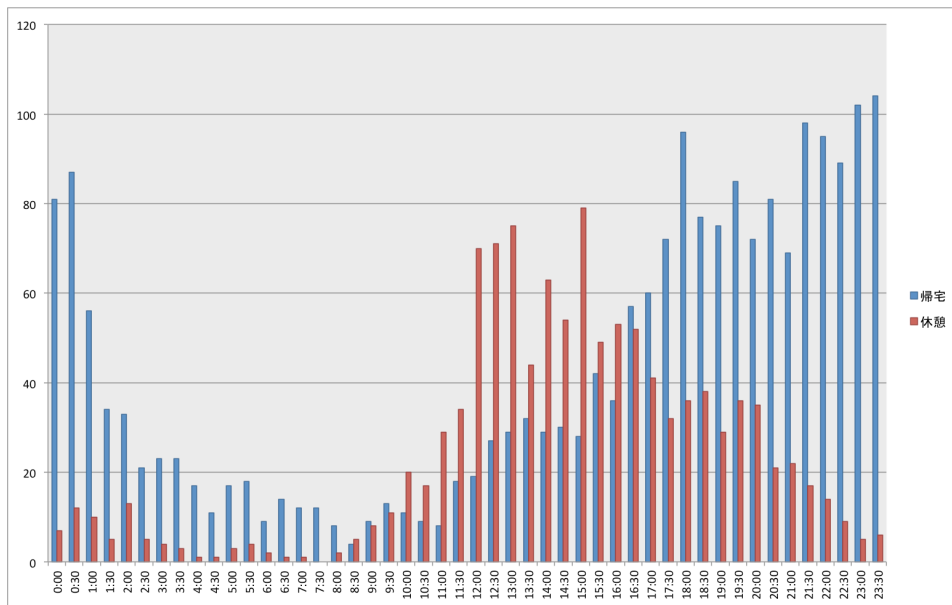


図 1.3: 2 個のイベントを同じ軸上に並べた場合

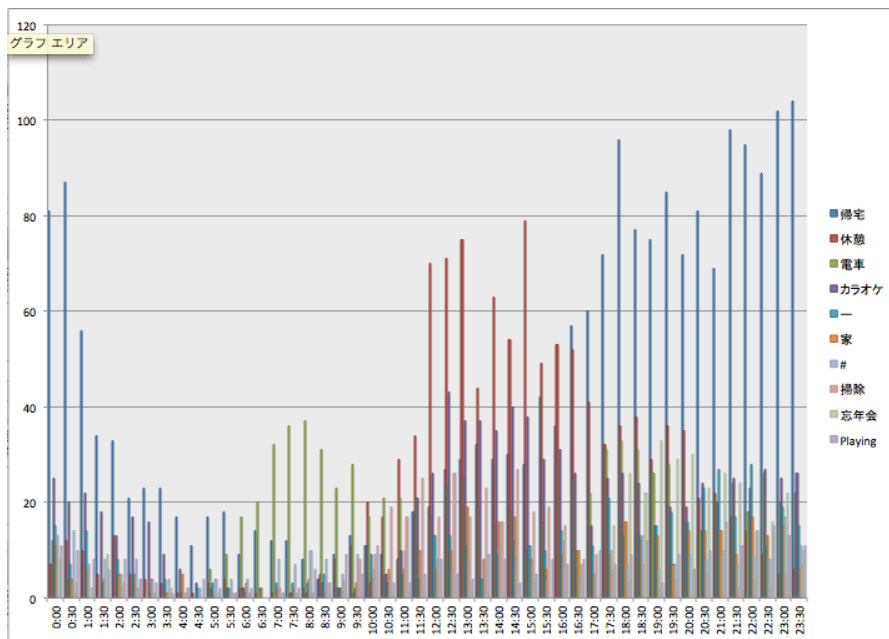


図 1.4: 10 個のイベントを同じ軸上に並べた場合

大量の時刻の集合を視覚的に表現することで、我々はイベント発生に関する様々な特徴を把握することができる。例えば、事象に関するトレンドや周期性、因果関係などである。このような特徴は、マーケティング、セキュリティ、リソースマネジメントなど、様々な分野に応用していくことができる。

1.5 本論文の構成

以下、第2章で、時刻情報付きデータを可視化する関連研究を紹介する。続いて、第3章で、時刻情報付きデータの可視化について、どのような要求があるかをまとめる。ChronoViewの表現方法について、第4章で述べ、第5章では、ChronoViewを適用した分析ツールについて述べる。第6章で、ChronoViewを用いた2つのユースケースを提示し、最後に、第7章をまとめとする。

第2章 関連研究

時刻情報付きデータを可視化する手法として、データの俯瞰を提供するものがある。この手法では、一度に複数のイベントを視覚的に表示し、各イベントの特徴を把握できるようにする。時刻情報付きデータの可視化手法の中でも、特に、データ中のイベントの周期性に着目した手法がある。この手法では、注目したイベントの発生の仕方について、どのような周期性があるかを詳細に提示する。また、データの俯瞰とイベント発生の周期性の両方を示す可視化手法もある。この章では、これらの手法について議論する。

2.1 データの俯瞰を提供する可視化

2.1.1 折れ線グラフを用いた手法

Playfair は、折れ線グラフや棒グラフを提案した [4]。折れ線グラフや棒グラフは、時系列データの表現として、最も一般的なものであり、時系列データを正確に表現する [5]。これらのグラフを拡張した手法がある。Aris らは、突発的に発生したイベントとその時刻を記録した時刻情報付きデータについて、イベント別の表示方法とイベント索引を使った表現方法を開発した [6]。Buono らは、インタラクティブに時系列データに問い合わせつつ、データ分析を可能にするツールを開発した [7]。これらの表現方法では、個々の時刻を視覚的に表現し、ある1つのイベントがいつ発生したのかということや、2、3個のイベント発生の共起性を把握することができる。しかし、これらの表現では、1つのイベントを表示するために多くのスペースを必要とするため、たくさんあるイベントを一度に表示することができない。

2.1.2 折れ線グラフを拡張した表現手法

折れ線グラフに類似しているが、各線に太さを付加することで時間に沿った変化を視覚的に提示する手法がある。Shi らは、積み上げグラフのような ThemeRever[8] を拡張し、時系列データの中でも特にイベントが順位付けされているデータを可視化する手法を開発した [9]。この手法では、ThemeRever 内の各時間ごとに、色で塗られた四角を用いて、順位付けされたイベントの注目のされた方の変化を時間の流れに沿って表現する。Wongsuphasawat らは、フローチャートを拡張して、時刻情報付きデータを可視化する手法を開発した [10]。この研究では、サッカーの試合結果の良い結果と悪い結果を、時間の流れと共に一度に俯瞰できるような概観を提供する。Tanahashi らは、映画や本などのストーリーの流れを可視化する手法を

開発した [11]。この手法では、登場人物を一本一本の線で表し、話しの流れに応じて、線の位置を入れ替えるように表現している。これらの表現では、数個から数十個のイベントを同時に表示することは可能だが、イベント数が大量になった場合、全てのイベントを俯瞰するために、ディスプレイ内に全イベントを収めることは難しい。

2.1.3 タイルのような表現を用いた手法

タイルのような表現を用いてデータの俯瞰を提供する手法がある。Mintz らは、縦軸に曜日、横軸に月をとるタイルのような表現する Tile Maps を開発した [12]。Tile Maps は、1つのイベントについて、その日その日の特徴をタイルの色で表し、カレンダーのように敷き詰められている。この表現では、1つのイベントの時間経過とともにどのような変化をしているかを示すことができる。Shimabukuro らは、1日、1ヶ月、1年を表す複数のセルを用意し、そのセルの中の色を塗り分けることで、イベントの時間に応じた変化を示す表現を開発した [13]。この表現では、1つのイベントについて、色を塗り分けることで、数十のイベントの特徴を示すことができる。これらの表現では、長期のデータを視覚的に提示することができるが、1つのイベントを表示するために多くのスペースを必要とするため、たくさんあるイベントを一度に表示することができない。Wattenberg らは、色付きの小さなタイルを敷き詰めて、Wikipedia の記事が、いつ、どのように変更されたかを表す Chromogram を開発した [14]。Chromogram は、記事がどのように変更されたかを色で表し、時間軸に沿って並べて表示する。Lammarsch らは、カレンダー風の表現で、1ヶ月分のデータを5分間隔でイベントの特徴を表現する GROOVE を開発した [15]。GROOVE は、5分間隔で1つのイベントの特徴の変化を各ピクセルの色で表現する。小さなタイルと色を用いることで、数百程度のイベントを一度に表現することはできるが、それ以上になると色のみでは識別できなくなってしまう。

2.1.4 色の变化を用いた手法

色の变化を用いて、データの俯瞰を提供する手法がある。Nguyen らは、時刻情報を持つタグを可視化する表現方法を開発した [16]。タグの時間情報の表現として、テキストに大きさや輝度などを利用した視覚的な外観表現方法と、タグに背景色やカレンダー風のセルの色分けなどの視覚的な図を付加した表現方法を提案した。この表現手法では、時間のみの場合といった、特定の項目に着目した場合に、大量のイベント発生の特徴の比較は可能である。しかし、イベントの周期性を把握することや、年月日、曜日、時間のような項目を複数組み合わせ、イベント発生の特徴を比較することができない。

2.2 イベントの周期性を提示する可視化

2.2.1 螺旋を用いた表現手法

Carlis らは、時間の連続性と週、月、年といった周期性の両方を同時に表示する螺旋を用いた可視化手法を開発した [17]。螺旋の上に時間の連続性、半径に月、巻き数に年といった属性を割り当て、その時々イベントの発生頻度の調査を可能にした。Weber らは、螺旋表現で、周期を様々に変更しながら表現する手法を開発した [18]。この表現では、螺旋上の線の太さに応じて、イベント発生の特徴の変化を示す。Dragicevic らは、アナログ時計のような可視化手法である SpiraClock を開発した [19]。この手法では、アナログ時計と螺旋表現の組み合わせで、直近の 2、3 時間程度のイベントを表示することができる。Tominski らは、既存の螺旋表現を改良し、2 トーンの疑似カラーリングとインタラクション手法を利用して、多変量データの分析がしやすくなる表現手法を開発した [20]。しかし、これら表現の場合、長期的なデータを表現しようすると螺旋の巻き数が増加し、その上にのせられる情報数が限定されるため、一度に複数のイベントの周期性を比較することは難しい。

2.2.2 円グラフを用いた手法

Keim らは、円グラフを改良し、円の中心から外側に時間の流れを取り、円周に沿って分割したセルの色を塗り分けて、イベント発生の特徴を示す CircleView を開発した [21]。CircleView は、数個のイベントについての周期性を色で表現する。しかし、1 つのイベントに多くのスペースを使用するため、大量のイベントの周期性を一度に比較することは難しい。

2.2.3 円周に時間軸を持つ表現手法

Muniandy らは、アナログ時計のような円の内側に、正方形のセルを配置し、注目したイベントについて実際に発生した時刻を示す表現を開発した [22]。この表現では、数個のイベントを円の内側に配置できるが、大量のイベントの周期性を比較することはできない。Misue が提案するアンカーマップは、2 部グラフのノード群の一方を円周上に等間隔で固定し、それらと関係性を持つ他方のノード群を自動配置する手法である [23]。アンカーマップでは、スプリングモデルを利用して固定されたノードと他方のノードをつなぐエッジの距離から、その関係性を把握することができる。ChronoView は、アンカーマップの変形と考えることもできるが、ノード間の距離ではなく、フリーノードに相当する図形の位置によって時刻とイベントの関係性を提示するものである。

2.3 データの俯瞰とイベントの周期性の両方を提供する可視化

Paley は、本や Web テキストなどのストーリーの概観とストーリー中に出てくる単語の周期性を示す TextArc を開発した [24]。TextArc は、特定のページ内の文章を順番に螺旋状に並

べ、螺旋状にある単語の位置の重みに応じて、螺旋の内側にそれぞれの単語を配置する。螺旋の内側に配置される単語は、文字の明度や大きさにストーリー中の単語の発生頻度を示し、位置で周期性を示す。しかし、時間軸を示す円周が螺旋状になっているため、単語の位置だけでは、単語発生の周期性を把握しづらい。

Krstajić らは、限られたスペースで、大規模で動的なイベントデータを時間を基本として表示する CloudLines を開発した [25]。この表現手法は、各イベントの時間ごとの関係性に応じて、線の色が徐々に減衰していく関数を適用し、線の太さの変化でデータの特徴を表現する。Zhao らは、リング状の表現と折れ線グラフを組み合わせた KronoMiner を開発した [26]。KronoMiner では、リング上にイベントが発生した区間を表す扇型を用意し、その中に折れ線グラフを埋め込むことで、データの概観とイベントの周期性を表現する。この手法では、数個のイベントを一度に表現することができる。しかし、これらの表現では、1つのイベントを表現するのに多くのスペースを必要とするため、一度に大量のイベントの周期性を比較したりすることは難しい。

Fischer らは、円形の TreeMap[27] に大量の円グラフのような表現を埋め込む表現である ClockMap を開発した [28]。円グラフのような表現は、円周を1日周期とし、円グラフ内の色を塗り分けることで、各イベントがどのような行動を起こしたかを示す。この円グラフのような表現と通常の TreeMap やピクセルベースの TileMap を組み合わせた Multi Display のツールの開発をした [29, 30]。これらの手法は、データの概観とイベントの周期性の両方を提示するが、1つのイベントを表現するのに多くのスペースを要するため、大量のイベントを一度に表現するのは難しい。

第3章 大量データの可視化における要求

大量の時刻情報付きデータを分析する際、一度に全てのデータを俯瞰できることが望ましい。全データを視覚的に一度に提示すると、どのようなイベントが、どのような時間に、どれくらい発生しているのか、データの大まかな特徴が把握しやすくなるからである。そこで、大量データの可視化における要求について考える。

3.1 データの俯瞰

大規模な時刻情報付きデータには、数千以上のイベントが含まれる場合がある。しかし、Excel やメモ帳などでレコードを眺めただけでは、データ全体のおおまかな特徴が把握しづらい。通常、大量データを分析する場合、データの一部を削り、分析をする場合がある。例えば、発生頻度の低いイベントを削除するといった方法である。しかし、削られてしまった部分にも、新たな知見が含まれている場合がある。

ここで、データの俯瞰について、

要求1 データから新たな知見を得るために、数千以上あるイベントを一度に俯瞰したい。
という要求あげられる。

3.2 イベント発生の特徴の把握

データに含まれるイベントには、様々な発生特徴がある。例えば、あるイベントは、特定の時刻に依存して周期的に発生している、時間に依存することなく常に発生し続けている、などである。Excel には、グラフ機能があり、ヒストグラムや棒グラフでイベント発生の特徴を表示できる。図 3.1 と図 3.2 は、Excel のグラフ機能で、面グラフを使った一例である。図 3.1 では、2 個のイベントを比較することはできるが、図 3.2 では、グラフが重なっている部分があるため、全てのデータを比較ができない。このような表示では、数千以上のイベント発生の特徴を一度に比較することは難しい。

ここで、イベント発生の特徴の把握として

要求2 数千以上のイベントを俯瞰しながら、それらのイベント発生の特徴を知りたい。

要求3 数千以上のイベントについて、全イベントの特徴を一度に比較したい。

という要求があげられる。

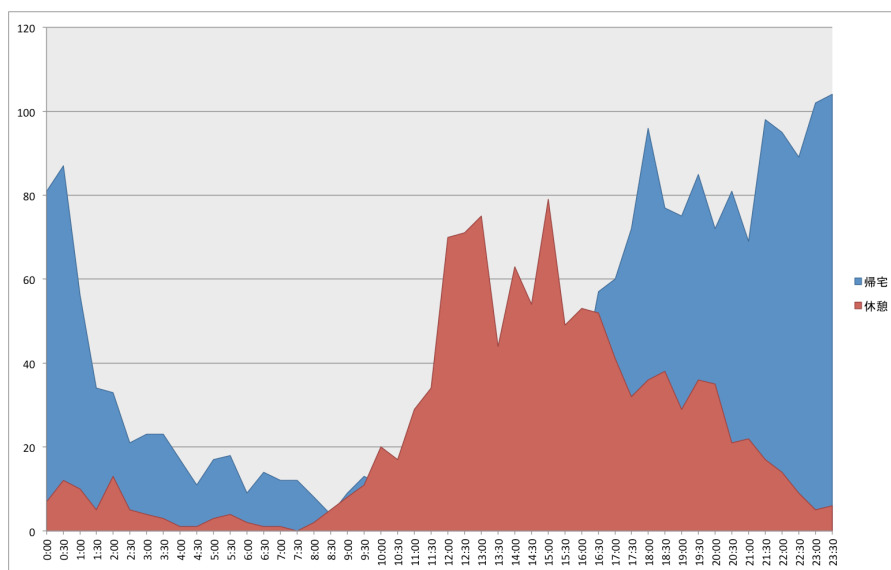


図 3.1: 2 個のイベントを一度に表示した面グラフの例

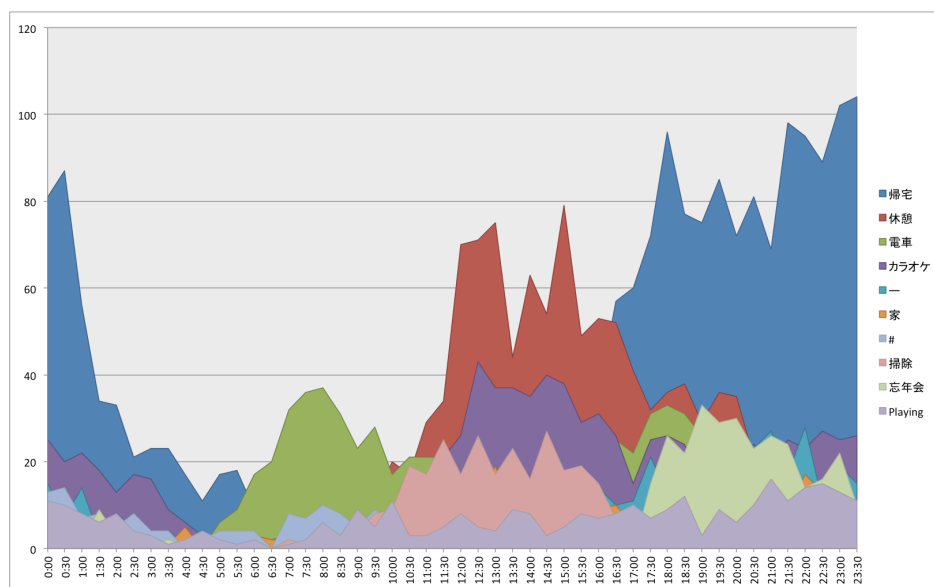


図 3.2: 10 個のイベントを一度に表示した面グラフの例

第4章 ChronoView: 視覚的表現の提案

4.1 データの俯瞰とイベント発生の特徴の把握における要件

第3章で掲げた要求を満たす大量データの可視化の要件として、

要件1 1つのイベントについて表現するためのスペースをできるだけ抑える。

要件2 全イベントをディスプレイ内に納め、スクロールなどの操作をしなくても一度に俯瞰できる。

要件3 1つのイベントを表現するスペースをできるだけ抑えつつも、イベントが、どのような時間に、どれくらい発生しているのかといった特徴を把握できる。

要件4 数千以上のイベントを一度に表示しても、全イベントの特徴を比較できる。

の4つの要件が考えられる。

ここで、これらを要件を満たしながら、大量データを可視化する表現として、

- 位置を用いた表現
- ピクセルベースでの表現

の2つのパターンで可視化する方法が考えられる。本研究では、前者の位置を利用した表現手法を考える。

4.2 位置を使った時刻の集合の視覚的表現

ChronoView は、アナログ時計の文字盤のような二次元平面上にイベントを配置し、位置でイベントの発生特徴を示す表現である。

まず、ひとつの時刻を二次元平面上に配置することを考える。時刻が t_0 のとき、アナログ時計の文字盤の12時の位置に配置し、半径 r の円周上に時計回りに周期 c で配置する。すべての時刻の集合を U とすると、このような配置は関数 $f_0 : U \rightarrow R^2$ によって次のように定義できる。

$$f_0(t) = (r \cos \theta, r \sin \theta) \quad (4.1)$$

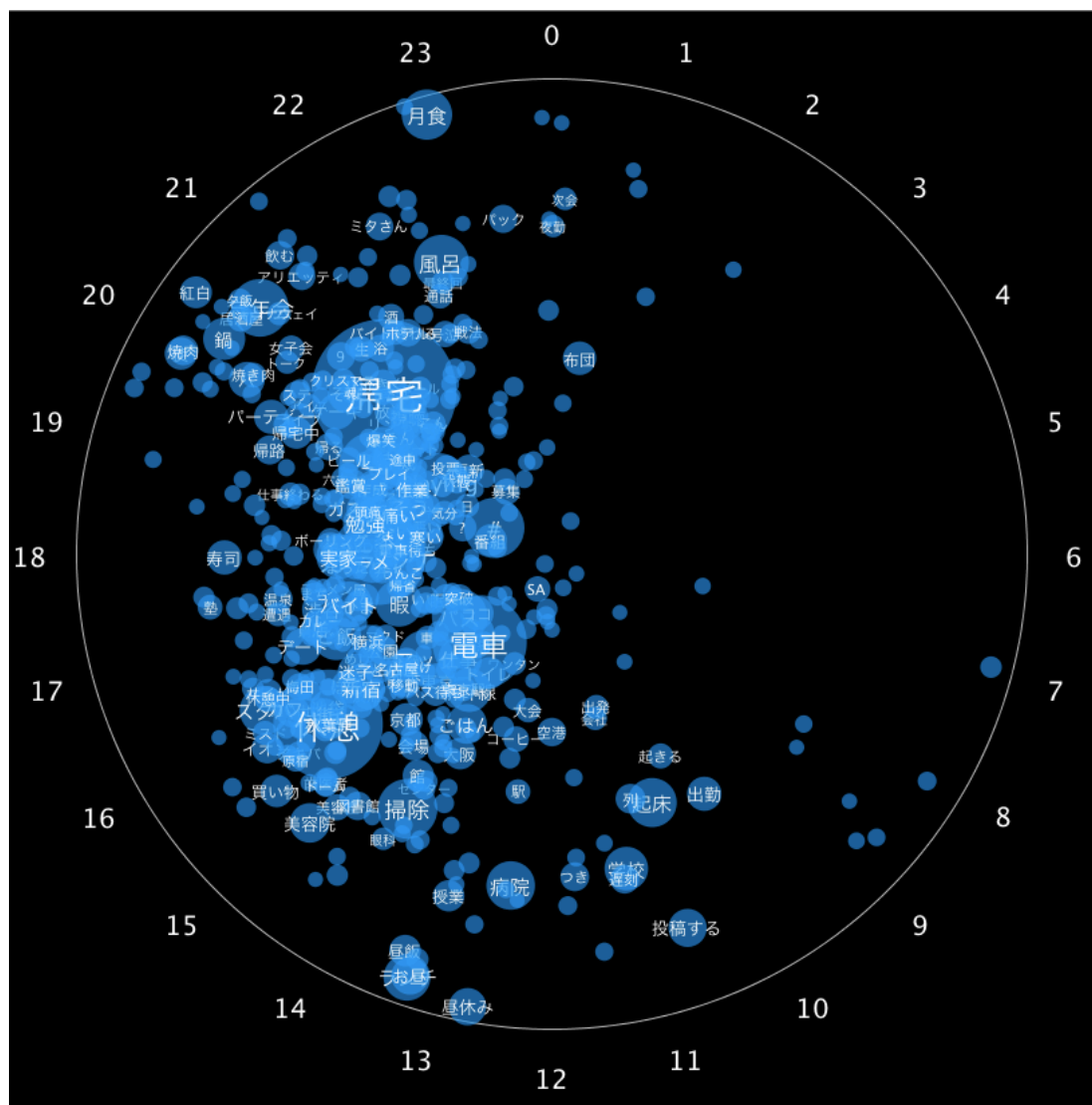


図 4.1: ChronoView の概観

ただし、

$$\theta = \frac{\pi}{2} - 2\pi \frac{t - t_0}{c}$$

時刻の有限集合 $T = \{t_1, t_2, \dots, t_n\} \in 2^U$ は、要素 t_i ($i = 1, \dots, n$) に対する平面上の点の重心に配置する。このような配置は関数 $f: 2^U \rightarrow R^2$ によって次のように定義できる。

$$f(T) = \frac{1}{|T|} \sum_{t \in T} f_0(t) \quad (4.2)$$

このように、あるイベント e が時刻の有限集合 T を持つとき、 e を $f(T)$ に配置する。

図 4.1 では、 t_0 を午前 0 時、 c を 24 時間としている。 c を 1 時間とすれば、1 時間周期の図を表現できる。また、 t_0 を日曜日の午前 0 時、 c を 7 日間とすると、1 週間周期の図を表現できる。同様に、1 か月周期や 1 年周期の図を表現することもできる。しかし、月の長さや年の長さは一定ではないため、実際の図を作成する際に若干の補正を加える。

時刻の集合の配置について、具体的に説明する。個々の時刻を線分上にとることを考える。今回、直線分は、0 時から 23 時 59 分まで時間軸を表すものとする。時間軸を直線分にとった場合、イベントは、その時間軸上に点をうつ。図 4.2 は、時間軸を直線分にとった一例である。もし、0 時にイベントが発生した場合、線分の 0 時の軸にイベントは配置される。一方、8 時にイベントが発生した場合、8 時の軸の上にイベントは配置される。

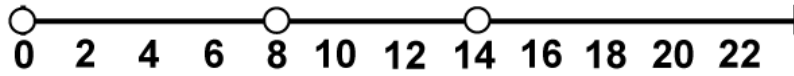


図 4.2: 線分上に時間軸を取った場合のイベントの配置

ChronoView は、時間軸を円周にとって表現する。図 4.3 は、ChronoView でイベントを配置した場合の一例である。ここでは、1 周を 1 日周期としている。イベントが一回だけ発生した場合の例を考える。もし、0 時に発生したイベントがあった場合、円周上の 0 時の位置にイベントは配置される。一方、8 時に発生したイベントがあった場合、円周上の 8 時の位置にイベントは配置される。次に、イベントが二回以上発生した場合の例を考える。もし、0 時、8 時、14 時に発生したイベントがあった場合、このイベントは、円の中心よりに配置される。一方、18 時、22 時に発生したイベントがあった場合、18 時と 22 時の間の円周よりに配置される。

この配置方法により、イベントが広範囲に分散して発生しているのか、特定の時刻に集中して発生しているのかなど、おおよその特徴把握が可能になる。

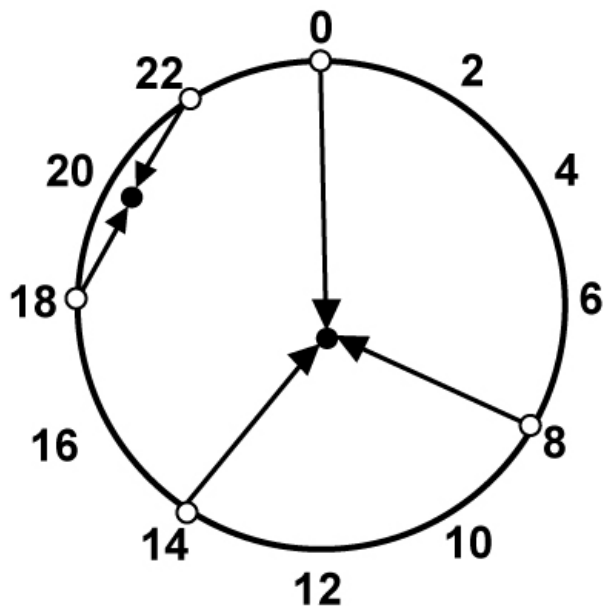


図 4.3: ChronoView で時刻の集合を配置した場合

4.3 各種属性の表現

イベントには、イベントが発生した時刻の集合以外に、それを識別するラベル、発生頻度、各種カテゴリのような属性も付随する。属性の表現には、位置以外の視覚変数 [31] を利用する。データによっては、イベントをカテゴリ分けできるものも多い。カテゴリは、色相や形状によって表現できる。

図 4.1 では、イベントのラベルを白い文字で表す。また、イベントの発生頻度は、円の面積で表現する。

4.4 付加情報の表現

ChronoView では、二次元平面上の特定の位置に配置した図形でイベントを表現するが、さらに付加的な図形を配置することで表示情報の詳細化や明確化が可能になる。

イベントが実際にどの時刻に発生しているのかを示すため、イベントが発生した時刻を放射状の線で表現する。図 4.4 は、放射状の線のイメージ図である。このような放射状の線を表示することで、イベントがいつ発生したのか、発生時刻の分布の特徴を正確に把握することができる。

ChronoView で表現されるイベントには、意味的な関連性が存在する場合がある。たとえば、ある日時の前後の比較を行う場合、その時刻前のイベントと後のイベントを異なるイベントとして表示し、さらに関連のあるイベントとして表現できると比較しやすい。そのため、そ

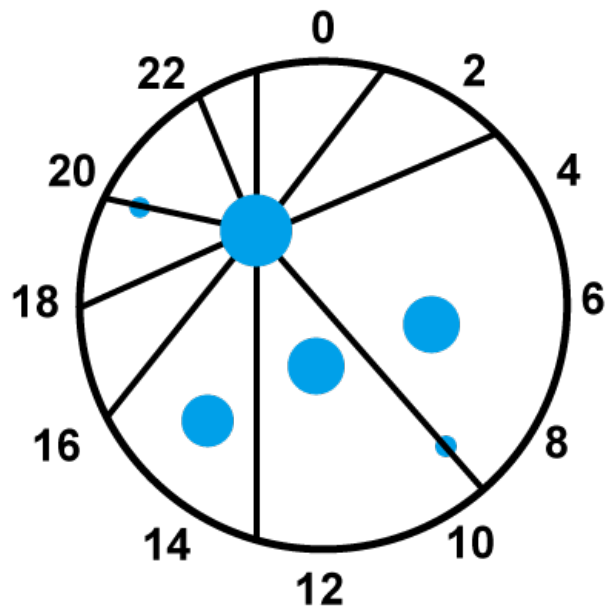


図 4.4: 実際にイベントが発生した時刻を示す放射状の線の表示

のような関連のあるイベント群は、線でつないだり、囲んだりすることで表現する。

4.5 金平糖表現

ChronoView は、異なる時刻の集合が同じ位置を占める可能性がある点で、表現上の曖昧性がある。本研究では、この曖昧性を解消し、複数イベントを一度に比較できる表現として、金平糖表現を考えた。

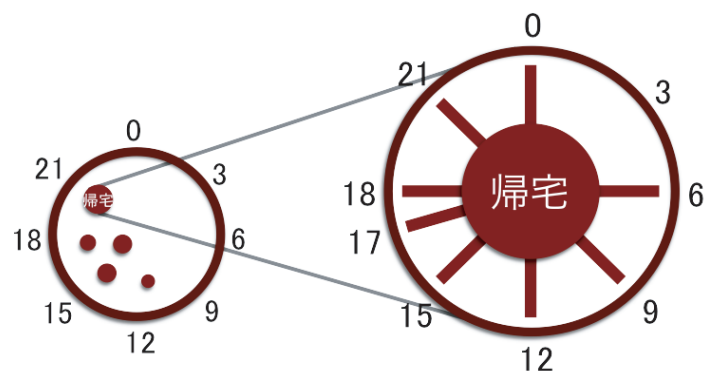


図 4.5: 金平糖表現の表現方法

まず、1つのイベントに着目し、そのイベントにおいて、アナログ時計の文字盤のような円

を考える。次に、注目したイベントの中心から、アナログ時計の文字盤のような円において、実際にイベントが発生した時刻の方向へ、1本1本直線を描画していく。図4.5は、描画方法の一例である。

図4.6は、本研究で開発した金平糖表現を適用した例である。ここでは、各時間帯ごとにイベントの発生頻度を取り、線の長さで各時間帯ごとの発生頻度を表している。この表現により、形を見る事で、おおまかな発生時刻の特徴が読み取る事ができる。また、一度に複数のイベント発生の特徴を把握することが可能である。

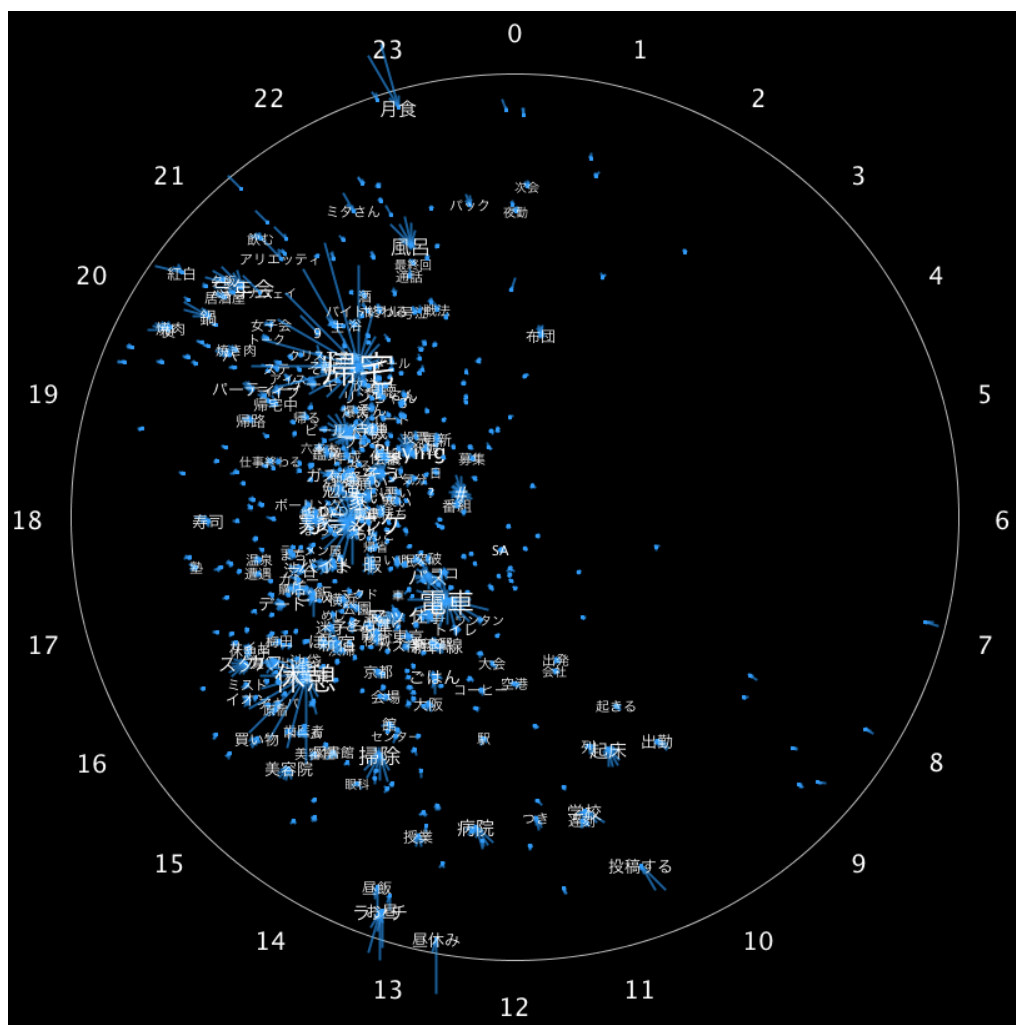


図 4.6: 金平糖表現

第5章 分析ツールの開発

我々は、時刻情報付きデータの分析を支援するため、第4章で示した ChronoView を利用した分析ツールを開発した。図 5.1 は、開発した分析ツールの概観である。

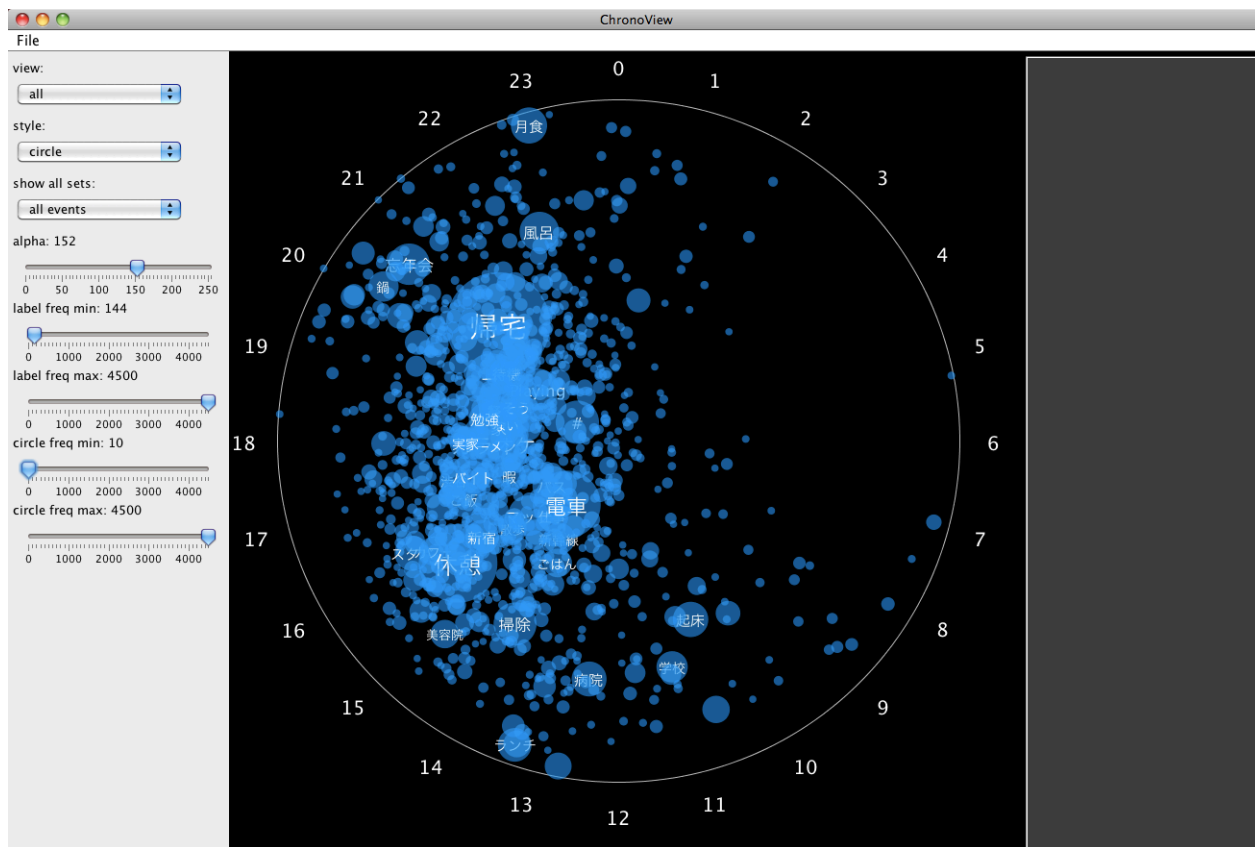


図 5.1: 分析ツールの概観

5.1 ChronoView ウィンドウ

画面中央では、データを視覚的に提示する ChronoView を提供する。ChronoView は、All View、Day View、Weekday&Holiday View の3種類に切り替えることができる。

All View は、全データをまとめて表示する。気になるイベントをクリックした場合、イベントが実際に発生した時刻を表す放射状の線が表示される。図 5.2 は、イベント「起床」をクリックした場合に表示される放射状の線の一例である。

Day View(図 5.3) では、各イベントを曜日別に分けて集計し、重心を求めてイベントを配置する。各イベントは、曜日別に7つの色で塗り分ける。イベント毎のばらつきやまとまりを把握しやすくするために、特定のイベントをクリックすると同じイベント同士を直線で結んで表示する。この方法により、イベント毎に曜日間のばらつきがどの程度あるかを把握することができる。

Weekday&Holiday View(図 5.4) では、平日と休日別に分けて集計し、重心を求めて配置する。各イベントは、平日が水色、休日がピンク色の2色に塗り分けられる。ここで、平日のイベントと休日のイベントは、それぞれ一日分の平均頻度を計算し、イベントを表す円の面積に対応させる。配置された各イベントのまとまりがわかりやすくなるように、同じイベント同士は直線で結ぶ。この方法により、イベントが平日と休日別によって、ばらつきがあるのか、まとまっているのかを把握することができる。

5.2 詳細画面

画面右側では、ChronoView の各ビューで気になったイベントをクリックした時に、そのイベントを含む詳細情報を見ることができる詳細画面である。図 5.5 は、イベント「起床」をクリックした場合に表示される詳細画面の一例である。

5.3 表現の操作画面

画面左側は、ビューを切り替えるためのリストや、各イベントの表現に関するパラメータを変更するためのスライダを設けている。スライダは、イベントの発生頻度の最大値と最小値、イベントを表す円の不透明度(アルファ値)の変更を行うことができる。同程度の発生頻度のイベントが近くに密集している場合には、発生頻度でフィルタリングしても図が繁雑になることを解消できない。そのような場合、イベントの円の不透明度の調整が有効である。密集度が高い場合は、不透明度を低めることで、イベントの分布が把握しやすくなる。イベント数が膨大になると、一度に全データを表示すると図が煩雑化し、分析が困難になってしまうことがある。その際、イベントの発生頻度によるフィルタリングや不透明度の調整を行い、イベントの分布を把握しやすくする。

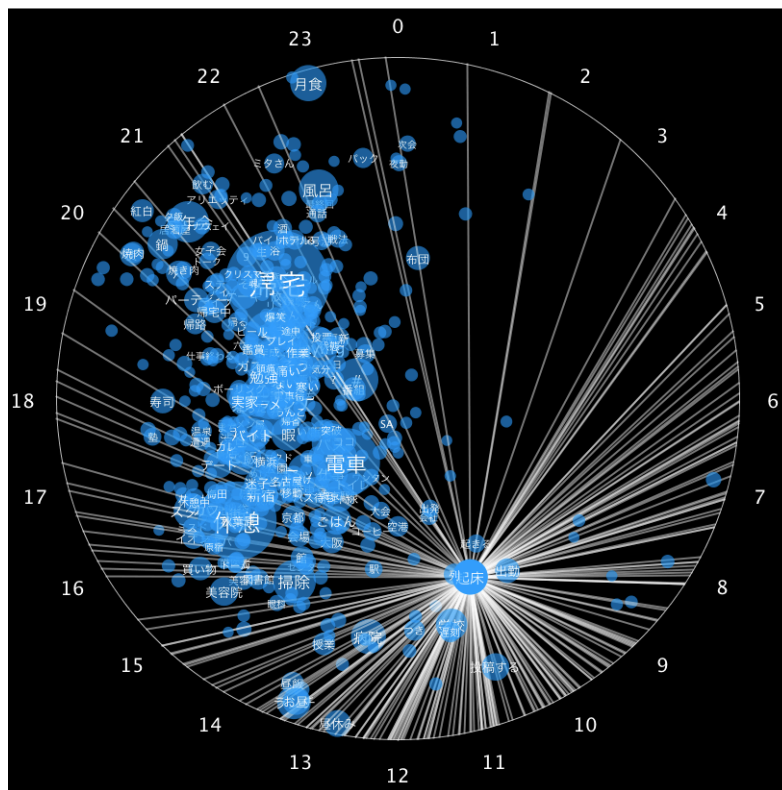


図 5.2: All View で気になるイベントをクリックした場合、放射状の線を表示する。

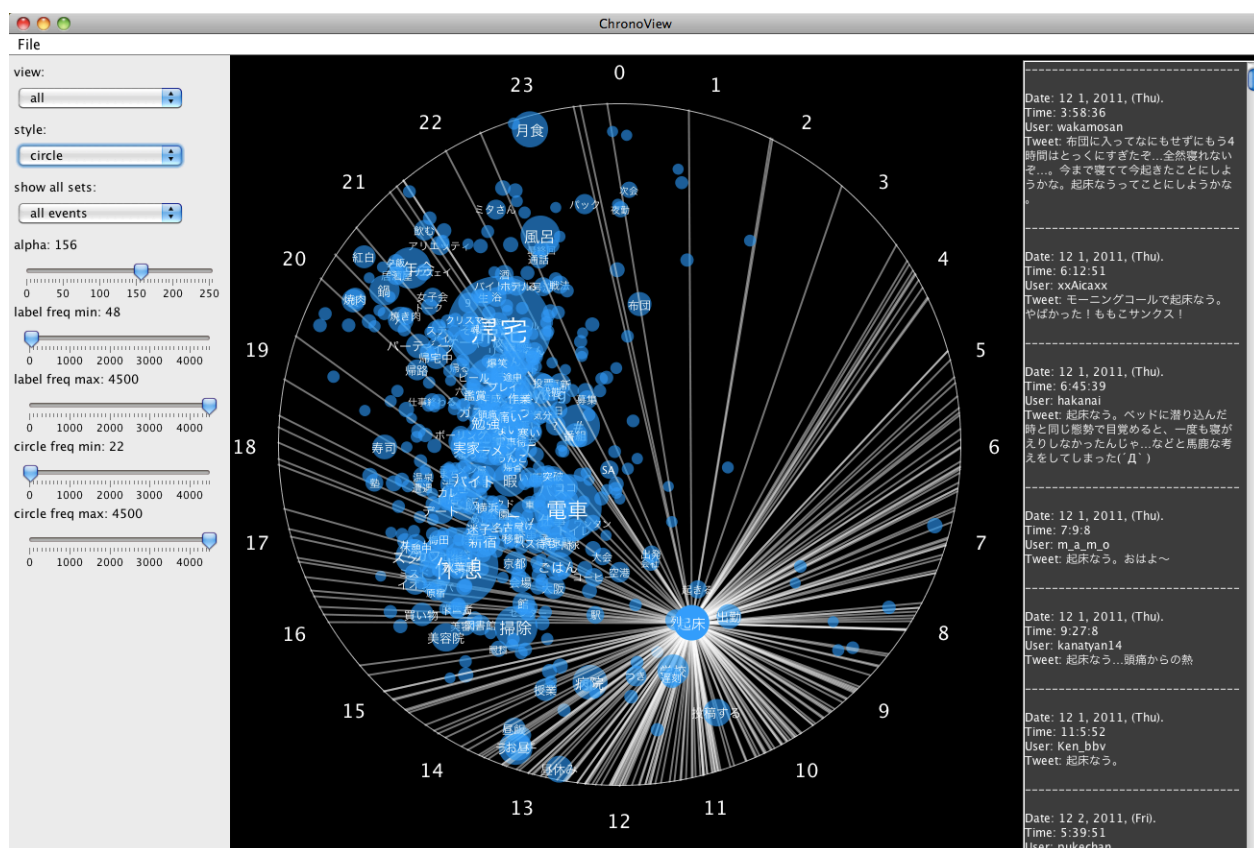


図 5.5: イベント「起床」をクリックした場合に表示される詳細画面の例

第6章 ユースケース

ChronoView の有用性を示すため、2つのユースケースを行った。

1つ目は、ファイアウォールログを使用したユースケースである。ファイアウォールログは、数千以上のイベントを含む、大規模データである。このデータを一度に俯瞰できれば、インターネットアクセスへの攻撃といった事象を発見しやすくなり、セキュリティ分野へ応用できる。このユースケースでは、イベントを位置で表現することで、どのような特徴を把握することができるか、調査する。

2つ目は、マイクロブログである Twitter のツイートを使用したユースケースである。特に、今回のユースケースでは、「新宿なう」、「カフェなう」といった今の行動を表す「～なう」というツイートに着目する。このツイートデータは、1日に数万以上のイベントが含まれることがある。この大量にあるイベントの中から、人間の行動の特徴が把握できると、マーケティング分野へ応用できる。

6.1 ファイアウォールログの可視化

ChronoView のユースケースの1つ目として、ファイアウォールログデータを対象とした分析を行った。この分析では、送信元 IP アドレスに注目し、ユーザがどのような時間にインターネットを利用しているのかを調べる。

6.1.1 利用データ

データは、筑波大学の宿舎ネットワークのファイアウォールログデータである。このデータは、タイムスタンプ (UNIX タイム)、接続所要時間、サービス、プロトコル番号、送信バイト数、受信バイト数、送信元 IP アドレス、受信元 IP アドレス、送信元ポート、受信元ポート、切断理由を持つ。ファイアウォールログデータの一例を表 6.1 に示す。

表 6.1: ファイアウォールログデータの例

タイムスタンプ (UNIX タイム)	接続所要時間	サービス	プロトコル番号	送信バイト数	受信バイト数	送信元 IP アドレス	受信元 IP アドレス	送信元ポート	受信元ポート	切断理由
1293861319	4	tcp	6	70	64	192.0.2.1	198.51.100.2	56604	13896	Close - AGE OUT
1293888801	70	http	6	1731	522	192.0.2.2	203.0.133.4	54556	80	Close - TCP RST
1293892329	66	http	6	3804	62851	192.0.2.3	198.51.100.34	54584	80	Close - TCP RST
1293899980	3	dns	17	82	330	192.0.2.4	198.51.100.10	49939	53	Close - RESP
1293900230	61	udp	17	183	88	192.0.2.2	203.0.133.23	57662	1178	Close - AGE OUT
1293923012	22	tcp	6	210	0	192.0.2.3	198.51.100.102	54047	9064	Close - AGE OUT
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

今回のユースケースでは、2011 年 1 月 1 月のデータを扱う。このデータには、レコードが 10,030,270 件、異なる送信元 IP アドレスが 1,268 件含まれる。このデータを利用し、インターネット利用者がどのような時間帯にインターネットを利用しているのか、その行動の特徴を調べる。今回は、送信元 IP アドレスがインターネットへアクセスをしたことを対象のイベントとする。

ユースケースを行うため、ファイアウォールログから送信元 IP アドレスと、そのアドレスがインターネットへアクセスした時間を示すタイムスタンプを抽出した。加工後のファイアウォールログデータの一例を表 6.2 に示す。

表 6.2: 加工後のファイアウォールログデータの例

タイムスタンプ (UNIX タイム)	送信 IP
1293861319	192.0.2.1
1293888801	192.0.2.2
1293892329	192.0.2.3
1293899332	192.0.2.1
1293903401	192.0.2.4
1293912234	192.0.2.2
⋮	⋮

図 6.1 は、実際に ChronoView で描画した画面の一例である。図 6.2 は、1 つのイベントに着目した場合の画面例である。ここでは、着目した IP アドレスとインターネットへの接続回数を白いラベルで表現している。表 6.1、表 6.2、図 6.2 と文中に記載した IP アドレスは、個人の特定を避けるため、テスト用のダミー IP アドレスで表記している。

6.1.2 観察

ChronoView でデータ全体を俯瞰した。イベントを表す円は、ビューの左上の方にややかたよっていることが見えた。円周上には、ほとんどイベントは配置されていないが、22 時から 2 時の間に、円周のかなり近い部分にもイベントが配置されていた。この円周近くに配置されているイベントの中でも、22 時と 23 時の間、1 時と 2 時の間に配置されているイベントは、円の大きさがやや大きめであった。それぞれのイベントをクリックしてみると、22 時と 23 時の間のイベントは、インターネットへの接続回数が 104,137 回、1 時と 2 時の間のイベントは、167,190 回と表示された。

ビューの右下に、やや円周よりで、ビューの中で一番大きな円を表示しているイベント「192.0.2.1」を見つけた。このイベントをクリックすると、インターネットへの接続回数が 1,735,131 回と表示された。一方、ビューの中央あたりに、2 番目に大きい円であるイベント「192.0.2.2」を見つけた。このイベントをクリックすると、インターネットへの接続回数が、1,177,276 回と

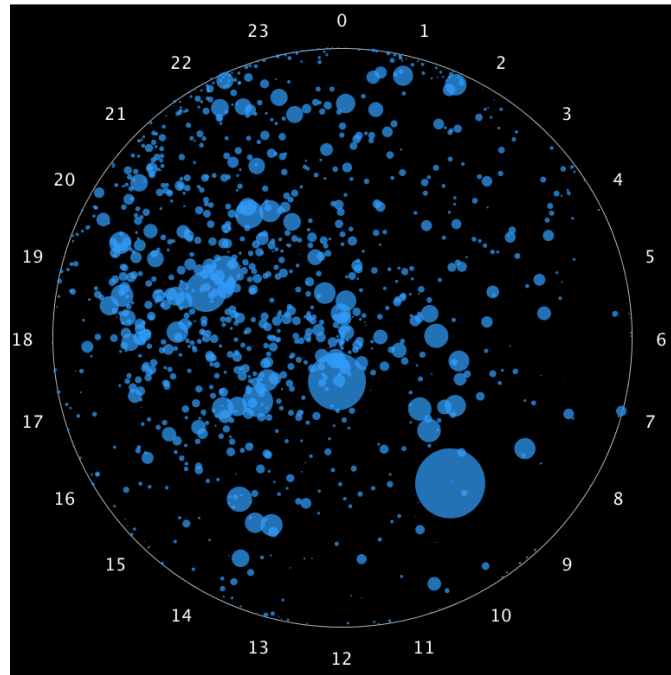


図 6.1: 2011 年 1 月 1 日のデータの表示

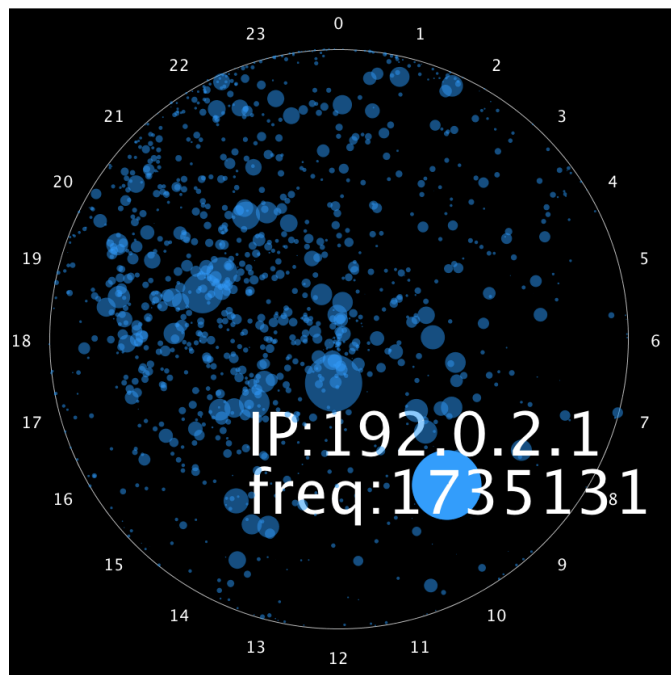


図 6.2: 1 つのイベントに着目した場合の表示

表示された。

6.1.3 考察

イベントがビューの左上にかたよっていることから、この日は、夕方から真夜中にかけて、インターネットへよく接続されていたことがわかった。また、円周のかなり近い部分に、やや大きめの円であるイベントが配置されていたことから、ごくわずかな時間帯に集中してインターネットを利用していたことがわかった。

一番大きな円で表示されていた 192.0.2.1 と二番目に大きな円で表示されていた 192.0.2.2 を比較すると、インターネットへの接続回数にそこまで大きな差がないことがわかった。しかし、イベントの配置されている位置から、192.0.2.1 の方が特定の時間帯にのみ、インターネットに接続していたことがわかった。この結果から、192.0.2.1 は、何らかの攻撃をしていた可能性があると考えられるため、元データをさらに詳細に調べる必要がある。

6.2 ツイートデータの可視化

ChronoView のユースケースの 2 つ目として、Twitter のツイートデータを対象とした分析作業を行った。この分析作業では、人間の行動の特徴に着目する。

6.2.1 利用データ

データは、「起床なう」、「忘年会なう」のように「なう」というキーワードを含むツイートである。ツイートデータは、Twitter の StreamingAPI¹ を利用して収集した。このデータは、ユーザ名、日時、時刻、ツイート文を持つ。「なう」を含むツイートを MeCab² を利用して形態素解析し、「なう」の直前にある人間の行動にあてはまる名詞、動詞、形容詞を抽出した。今回は、抽出した単語列を対象のイベントとした。ツイートデータの一例を表 6.3 に示す。

表 6.3: ツイートデータの例

ユーザ名	日時	ツイート	イベント
white_luc	2011/12/18 19:30:20	忘年会なう	忘年会
yamahakusyon	2011/12/18 20:33:09	帰宅なう	帰宅
adajmdap	2011/12/18 20:35:12	新宿なう	新宿
⋮	⋮	⋮	⋮

¹<https://dev.twitter.com/docs/streaming-api>

²<http://mecab.sourceforge.net/>

今回のユースケースでは、2011年12月1日から12月31日までの103,194 ツイートのデータセット (dataset #1) と、2012年1月1日から1月31日までの99,923 ツイートのデータセット (dataset #2) を扱う。このデータを利用して、どのようなイベントが起こっているかを分析し、人間の行動の特徴を分析する。図 6.3 は、dataset #1 を使用したイベントを曜日別に色を塗り分けた表示、図 6.4 は、dataset #2 を使用したイベントを曜日別に色を塗り分けた表示である。

このデータには、若干の欠損がある。これは、StreamingAPI を利用したデータ収集プログラムに不安定な部分があったためである。厳密な分析を行う場合、データの欠損がないことが望ましいが、ユースケースを示すにあたり、差し支えない程度の欠損であると考ええる。

6.2.2 観察

dataset #1 と dataset #2 を ChronoView で俯瞰し、比較した。

Weekday&Holiday View でそれぞれのデータを俯瞰すると、「起床」や「ランチ」、「帰宅」、「風呂」などの日常的な人間の行動を示すイベントは、どちらのデータにおいても、ほぼ同位置に配置されていた。これらのイベントについて、平日と休日別に色分けした場合のそれぞれのイベントを結ぶ線の長さを見た時、どのイベントも平日と休日での距離は短かった。

Day View に切り替えて、「起床」、「ランチ」、「帰宅」、「風呂」を示すイベントをもう一度見てみると、それぞれのイベントの曜日ごとの配置で違いが見られた。「起床」、「帰宅」、「風呂」は、曜日ごとに表示されたイベントの位置に、ばらつきがあった。一方、「ランチ」は、曜日ごとの位置にばらつきがなく、ほぼ一カ所にまとまっていた。dataset #1 と dataset #2 とで、この違いは似ていた。

ここで、「ランチ」に着目し、曜日別に金平糖の形がどのようになっているかを調べた。dataset #1 で、「ランチ」は、12時と13時の線が特に長かった。しかし、木曜日は、11時と14時の線も長く、日曜日は14時の線も長かった。一方、dataset #2 では、どの曜日も12時と13時の線の長さが特に長かったが、曜日によっての違いはほとんどなかった。

dataset #1 と dataset #2 の俯瞰を比較すると、dataset #1 にのみ発生しているイベント「月食」を発見した。All View で見た時、このイベントは、ビューの左上、23時の円周近くに位置していた。発生頻度を表す円の大きさは、日常の行動の「風呂」や「起床」とほぼ同じ大きさであった。イベントをクリックして放射状の線を表示し、イベントの発生時間帯を確認すると、22時から0時に線が集中していた。Day View に切り替え、金平糖表現を表示すると、月食は、ピンク色の金平糖が大きく表示されており、土曜日の22時から23時、日曜日の0時の線が長くなっていた。詳細画面でツイートを確認すると、12月10日土曜日の22時から12月11日日曜日の1時近くまで、「月食なう」というツイートが多く見られた。

6.2.3 考察

曜日別、平日と休日別にイベントを表現し、時刻情報付きデータを俯瞰することで、日常的な人間の行動について、それぞれの特徴の違いを把握することができた。Weekday&Holiday

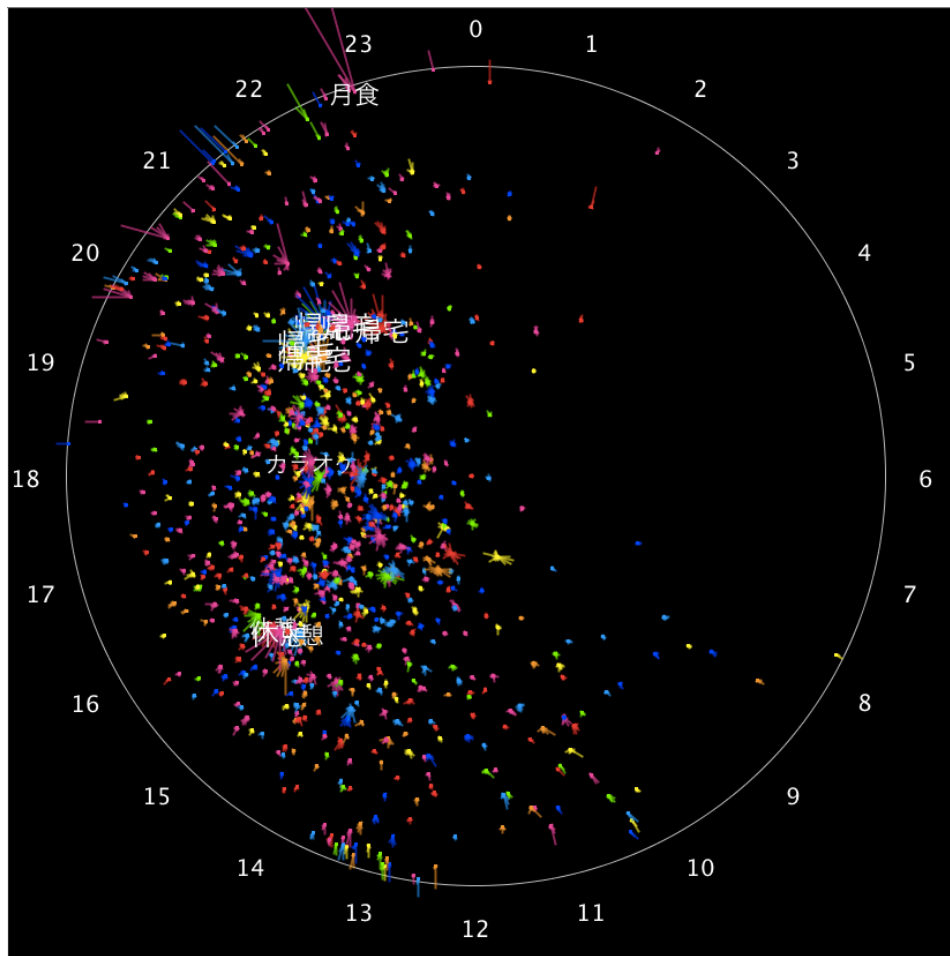


図 6.3: 12 月 1 日からのデータを使用した表示

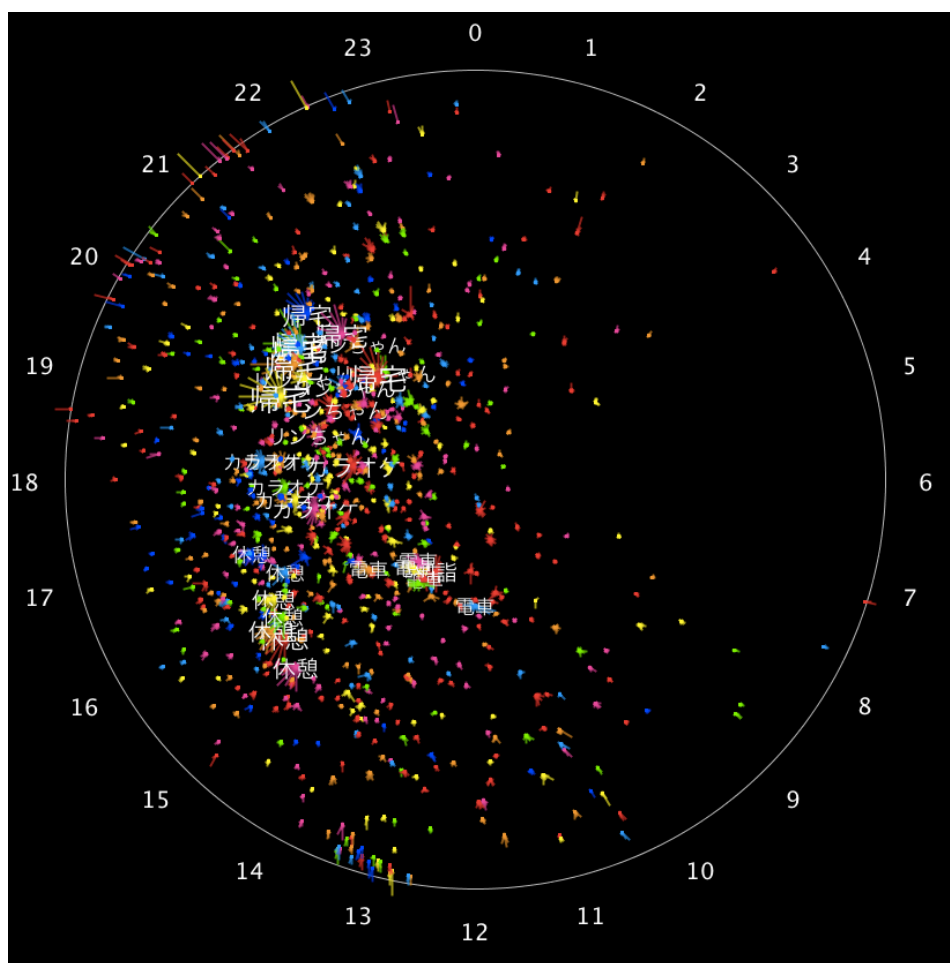


図 6.4: 1月1日からのデータを使用した表示

View で見た場合、「起床」、「ランチ」、「帰宅」、「風呂」というイベントは、それぞれ平日と休日で発生する時間に大きな違いがないように見えていた。Day View に切り替えて見た場合、「起床」、「帰宅」、「風呂」は、曜日ごとの発生時間にばらつきがあった。しかし、「ランチ」は、曜日ごとに関係なく、ほぼ同位置にあることがわかった。「ランチ」に着目し、曜日別の金平糖の形状の違いを比較した時、金平糖の形状から「ランチ」のピークは 12 時から 13 時まででありそうだとことがわかった。しかし、木曜日と日曜日は、11 時の線の長さが長かったことから、その曜日の「ランチ」のピークは、他の曜日よりも広いことがわかった。

これらの結果から、「起床」、「帰宅」、「風呂」は、出勤や通学などの個人の活動による影響で、人によってばらつきが出るが、「ランチ」は、複数人で昼食をとることが多いために、ほぼ同じような時間に発生するのではないか、という仮説がたてられる。もし、この仮説が成立すれば、「ランチ」のピークと考えられる 11 時から 13 時の間に、レストランで友人割引サービスなどのクーポンを提供することで、更なる売り上げ向上につなげられる可能性がある。このように、ChronoView は、人間の行動パターンや特徴の発見に効果があることが期待でき、マーケティング分野へ応用できる。

dataset #1 と dataset #2 を比較した時、dataset #1 にのみ発生しているイベント「月食」を見つけることができた。このイベントは円周近くに配置されており、特定の時刻に依存していることがわかった。また、Day View で金平糖表現を表示した時、土曜日の 22 時から翌日の 0 時すぎまで多くつぶやかれていることがわかった。詳細画面でツイートを確認することで、12 月 10 日土曜日の 22 時から 12 月 11 日日曜日の 1 時近くまで、「月食なう」というツイートが多く見られたことから、この日に多くの人が月食を見ていたことがわかる。このように、ChronoView では、日常の行動パターンだけでなく、数千以上あるイベントの中から、その時のみ起こるイベントを発見することが可能である。

6.3 議論

イベントの持つ時刻の集合を位置で表したことにより、数千以上のイベントを一度に俯瞰することができた。これは、1 つのイベントを表示するためのスペースをできる限り抑えるという要求を満たしている。

イベントが発生した時刻と位置を関連づけたことにより、イベント発生の特徴を表すことも可能にした。ビューの中央に配置されるイベントについて、発生時間帯について曖昧性が残ってしまう可能性があるが、発生した実際の時刻を表す放射線表現や金平糖表現により、この曖昧性を解消した。これは、数千以上のイベントを俯瞰しながら、イベントの特徴が知りたいという要求、数千以上のイベントを一度に表示しても、全イベントの特徴を比較したいという要求を満たしている。

ChronoView で気になったイベントを詳細に調べたい場合、一般的なグラフを用いて調査をすすめた方がより詳細に調べられる場合がある。この点についてはトレードオフが生じるが、ChronoView は、データの俯瞰を提供することで、大量のデータの中から特異なイベントを発見でき、そのイベントの詳細な調査につなげる手段として利用することができる。

第7章 まとめ

本研究では、大量の時刻情報付きデータからの特徴分析を支援するための表現手法である ChronoView を提案した。ChronoView では、時刻の集合をアナログ時計を模した二次元平面上の位置で示すことで、数千件以上のイベントを一度に表示できる、データの俯瞰を提供した。さらに、ChronoView の曖昧性を改善する表現を検討し、放射線表現と金平糖表現を提案した。放射線表現により、1つのイベントに着目した場合、そのイベントが実際にいつ発生したかを提示することを可能にした。また、金平糖表現により、一度に複数イベントの発生時刻の特徴を比較することを可能にした。また、各イベントを曜日別や平日と休日別に分類し、視覚的に表現することで、各イベントのさらに細かい特徴まで把握することを可能にした。

本研究では、Twitter のツイートデータとファイアウォールログにフォーカスし、ChronoView の試用を行った。この試用により、ChronoView の有用性、今後の可能性を示した。本論文では、人間の行動の特徴を分析するということを考えたツイートデータの分析、インターネットへのアクセスの特徴を分析するということを考えたファイアウォールログの分析に着目したが、ChronoView は、その他の分析においても応用できる可能性がある。例えば、Web のアクセスログの分析や、ショッピングにおける購買状況の分析、SVN 管理データの分析などである。今後は、ChronoView に Web のアクセスログやショッピングの購買履歴、医療データを適用した評価実験を行い、ChronoView の応用性をさらに明らかにしていく。

謝辞

本研究を行うにあたり、指導教員である三末和男准教授には、丁寧なご指導と助言を頂きました。心より感謝を申し上げます。また、志築文太郎准教授、高橋伸准教授、田中二郎教授には、研究室のゼミを通して、様々なご助言を頂きました。本当にありがとうございます。

インタラクティブプログラミング研究室の皆様には、ゼミでの発表や研究室での日常生活の中で、様々な貴重なご意見を頂きました。特に、NAIS チームの皆様には、ゼミ以外でも、研究の相談を聞いていただいたり、研究生活全体においてもたくさんのご意見やご指摘を頂きました。深く感謝しております。

ユースケースで使用した大学宿舎のファイアウォールログデータは、筑波大学学術情報メディアセンターから提供していただきました。また、ユースケースを進めるにあたり、和田耕一教授、佐藤聡講師に、貴重なコメントをして頂きました。本当にありがとうございます。

本手法を開発するにあたり、富士通研究所の皆様には様々な議論をしていただきました。この議論を通し、研究を有意義に進めることができました。本当にありがとうございます。

そして最後に、家族や友人をはじめ、大学生活の中でお世話になった全ての人々に、心より感謝を申し上げます。本当にありがとうございました。

参考文献

- [1] Wolfgang Aigner, Silvia Miksch, Heidrun Schumann and Christian Tominski. Visualization of Time-Oriented Data. Springer (2011).
- [2] Sônia Fernandes Silva and Tiziana Catarci. Visualization of Linear Time-Oriented Data: A Survey. *Proceedings of the First International Conference on Web Information Systems Engineering (WISE'00)*, Volume 1, pp. 310–319 (2000).
- [3] Geoffrey M. Draper, Yarden Livnat, Richard F. Riesenfeld. A Survey of Radial Methods for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 15 Issue 5, pp. 759–776 (2009).
- [4] William Playfair and James Corry. The Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Progress of the Commerce, Revenues, Expenditure and Debts of England during the Whole of the Eighteenth Century. Printed for J. Debrett; G. G. and J. Robinson; J. Sewell; the engraver, S. J. Neele; W. Creech and C. Elliot, Edinburgh; and L. White, London, UK (1786).
- [5] Ian Spence. William Playfair and the Psychology of Graphs. *Proceedings of the American Statistical Association, Section on Statistical Graphics*. Alexandria VA: American Statistical Association, pp. 2426–2436 (2006).
- [6] Aleks Aris, Ben Shneiderman, Catherine Plaisant, Galit Shmueli and Wolfgang Jank. Representing Unevenly-Spaced Time Series Data for Visualization and Interactive Exploration. *Human Computer Interaction INTERACT 2005*, pp. 835–846 (2005).
- [7] Paolo Buono, Aleks Aris, Catherine Plaisant, Amir Khella and Ben Shneiderman. Interactive Pattern Search in Time Series. *Proceedings of Conference on Visualization and Data Analysis (VDA 2005)*, pp. 175–186 (2005).
- [8] Susan Havre, Beth Hetzler, and Lucy Nowell. ThemeRiver: Visualizing theme changes over time. *Proceedings of IEEE Symposium on Information Visualization (InfoVis 2000)*, pp. 115–123 (2000).
- [9] Conglei Shi, Weiwei Cui, Shixia Liu, Panpan Xu, Wei Chen, and Huamin Qu. RankExplorer: Visualization of Ranking Changes in Large Time Series Data. *Proceedings of IEEE Symposium on Information Visualization (InfoVis 2012)*, pp. 2669–2678 (2012).

- [10] Krist Wongsuphasawat and David Gotz. Exploring Flow, Factors, and Outcomes of Temporal Event Sequences with the Outflow Visualization. *Proceedings of IEEE Symposium on Information Visualization (InfoVis 2012)*, pp. 2659–2668 (2012).
- [11] Yuzuru Tanahashi and Kwan-Liu Ma. Design Considerations for Optimizing Storyline Visualizations. *Proceedings of IEEE Symposium on Information Visualization (InfoVis 2012)*, pp. 2679–2688 (2012).
- [12] David Mintz, Terence Fitz-Simons and Michelle Wayland. Tracking Air Quality Trends with SAS/GRAPH. *Proceedings of the 22nd Annual SAS User Group International Conference (SUGI)*, (1997).
- [13] Milton H. Shimabukuro, Edilson F. Flores, Maria Cristina F. de Oliveira and Haim Levkowitz. Coordinated Views to Assist Exploration of Spatio-Temporal Data: A Case Study. *CMV '04 Proceedings of the Second International Conference on Coordinated & Multiple Views in Exploratory Visualization*, pp. 107–117 (2004).
- [14] Martin Wattenberg, Fernanda B. Viégas and Katherine Hollenbach. Visualizing activity on wikipedia with chromograms. *INTERACT'07 Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction - Volume Part II*, pp. 272–287 (2007).
- [15] Tim Lammarsch, Wolfgang Aigner, Alessio Bertone, Johannes Gartner, Eva Mayr, Silvia Miksch and Michael Smuc. Hierarchical Temporal Patterns and Interactive Aggregated Views for Pixel-Based Visualizations. *IV '09 Proceedings of the 2009 13th International Conference Information Visualization*, pp. 44–50 (2009).
- [16] Dinh Quyen Nguyen, Christian Tominski, Heidrun Schumann and Tuan Anh Ta. Visualizing Tags with Spatiotemporal References. *15th International Conference Information Visualisation (IV2011, July 12-15, 2011, London, UK)*, pp. 32–39 (2011).
- [17] John V. Carlis and Joseph A. Konstan. Interactive Visualization of Serial Periodic Data. *Proceedings of the 11th annual ACM symposium on User interface software and technology (UIST'98)*, pp. 29–38 (1998).
- [18] Marc Weber, Marc Alexa and Wolfgang Müller. Visualizing Time-Series on Spirals. *Proceedings of IEEE Symposium on Information Visualization (InfoVis 2001)*, pp. 7–13 (2001).
- [19] Pierre Dragicevic and Stephane Huot. SpiraClock: A Continuous and Non-Intrusive Display for Upcoming Events. *CHI'02 Extended Abstracts on Human Factors in Computing systems (CHI EA'02)*, pp. 604–605 (2002).
- [20] Christian Tominski and Heidrun Schumann. Enhanced Interactive Spiral Display. *Proceedings of the Annual SIGRAD Conference 2008*, pp. 53–56 (2008).

- [21] Daniel A. Keim, Jörn Schneidewind and Mike Sips. CircleView A New Approach for Visualizing Timerelated Multidimensional Data Sets. *AVI '04 Proceedings of the working conference on Advanced visual interfaces*, pp. 179–182 (2004).
- [22] Kovalan Muniandy. Visualizing Time-Related Events for Intrusion Detection. *Proceedings of IEEE Symposium on Information Visualization (InfoVis 2001)*, IEEE Computer Society (2001).
- [23] Kazuo Misue. Anchored Map: Graph Drawing Technique to Support Network Mining. *IEICE Trans. Inf. & Syst., Vol. E91-D, No. 11*, pp. 2599–2606 (2008).
- [24] W. Bradford Paley. TextArc: Showing Word Frequency and Distribution in Text. *Proceedings of IEEE Symposium on Information Visualization, Poster Compendium, IEEE CS Press*, (a poster) (2002).
- [25] Miloš Krstajić, Enrico Bertini and Daniel A. Keim. CloudLines: Compact Display of Event Episodes in Multiple Time-Series. *IEEE Transactions on Visualization and Computer Graphics, Vol. 17*, pp. 2432–2439 (2011).
- [26] Jian Zhao, Fanny Chevalier and Ravin Balakrishnan. KronoMiner: Using Multi-Foci Navigation for the Visual Exploration of Time-Series Data. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems(CHI'11)*, pp. 1737–1746 (2011).
- [27] B. Shneiderman. Tree visualization with tree-maps: 2-d space-lling approach. *ACM Transactions on Graphics*, 11(1):92—99, Jan. (1992).
- [28] Fabian Fischer, Johannes Fuchs and Florian Mansmann. ClockMap: Enhancing Circular Treemaps with Temporal Glyphs for Time-Series Data. *Proceedings of the Eurographics Conference on Visualization (EuroVis 2012)*, pp. 97–101 (2012).
- [29] Fabian Fischer, Johannes Fuchs, Florian Mansmann and Daniel A. Keim. BANKSAFE: A Visual Situational Awareness Tool for Large-Scale Computer Networks. *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST Challenge 2012)*, pp. 157–158 (2012).
- [30] Fabian Fischer, Johannes Fuchs, Pierre-Antoine Vervier, Florian Mansmann and Olivier Thonnard. VisTracer: A Visual Analytics Tool to Investigate Routing Anomalies in Traceroutes. *9th International Symposium on Visualization for Cyber Security (VIZSEC 2012)*, pp. 80-87 (2012).
- [31] J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. Univ. of Wisconsin Press (1983).