

イベントデータ可視化のためのグリフ配置手法の開発

安齋 康宏^{1,a)} 三末 和男^{1,b)}

概要： イベントの周期性や発生時刻の類似性を表現することは人々の行動分析など様々な目的において有用である。我々は大量のイベントの特徴を空間効率良く表現するための可視化手法を開発した。我々はグリフベースの可視化によりイベントを小面積で表現する。加えてイベントの時間的な特徴と類似性を同時に表現するグリフの配置方法を求める。本論文では ChronoView と多次元尺度構成法を組み合わせたグリフの配置方法について説明する。

Visualization of Event Data by Arranging Glyphs

YASUHIRO ANZAI^{1,a)} KAZUO MISUE^{1,b)}

Abstract: Representing the periodicity and similarity of occurrence time of events is useful for various purposes such as behavior analysis of people. We developed a visualization technique to represent event data space-efficiently. We represent events by glyph-based visualization. In addition, we find glyph-arrangement techniques which simultaneously represent the temporal features and similarity of events. In this paper, we describe a glyph-arrangement technique that combines ChronoView and multidimensional scaling.

1. はじめに

本論文で扱うイベントデータとは行動内容とその発生時刻（タイムスタンプ）を持つデータである。イベントデータの例として購買履歴データが挙げられる。「X 時に商品 A を購入」というデータがあるとき、これが行動内容「商品 A を購入」と発生時刻「X 時」を持つひとつのイベントである。同様の行動内容を持つイベントの集まりをイベント群と呼ぶ。例えば「X 時に商品 A を購入」「Y 時に商品 A を購入」というイベントは「商品 A を購入」という行動内容を持つ。これらのイベントをまとめて「商品 A を購入」イベント群と呼ぶ事ができる。

イベント群に含まれるイベントの発生時刻に着目すると、発生時刻の分布や周期性などに特徴が見られることがある。そういったイベントの特徴や特徴の類似性を分析することで、ビジネスやセキュリティなど様々な領域における知見の獲得と応用が可能になる。

イベント群の時間的な特徴を見るには、イベント群に含まれるイベントのタイムスタンプを表現することが有効である。Strip chart, Bar chart, Spiral chart などがタイムスタンプの集合を表現するために利用される。ただしこれらの手法はいずれも一定の表示領域が必要となる。大量のイベント群の描画に伴い、一定サイズを持つイベント群の表現が視覚的な煩雑さを生じさせる。この視覚的煩雑さを回避するにはイベント群を小面積で表現する必要がある。したがって、チャート類よりも小さい表示領域で表現可能なグリフを用いると都合が良い。我々はタイムスタンプの集合を発生時間帯で集計することによって、イベントが発生した時刻の分布（時刻分布）を表現する。グリフによってこの時刻分布を空間効率良く表現する。

たくさんのグリフを表現する手法としては Small Multiples が主に用いられる。ただし、Small Multiples ではイベント群間の時刻分布の類似性やはずれ値の発見を積極的に支援してくれない。これは格子状の配置がグリフ間の関係やはずれ値の存在を示さないためである。そこで、我々はグリフの位置を利用する。我々は次に定める条件を満たすようにグリフの配置を工夫する。

¹ 筑波大学
University of Tsukuba

a) anzai@vislab.cs.tsukuba.ac.jp

b) misue@cs.tsukuba.ac.jp

R1 グリフの位置が各イベント群の時刻分布を表現する

R2 グリフの位置の近さがイベント群の時刻分布の類似性を表現する

我々は多次元尺度構成法 (MDS) [1] を用いて類似性表現のためのレイアウトを得る。しかし MDS によって得られるレイアウトはイベント群間の類似性のみを提供し、時刻分布の特徴はレイアウトに反映しない。一方で ChronoView [2] は発生時刻の周期的な特徴を表現するレイアウト手法の一つである。だが ChronoView は表現に曖昧性を持ち、イベント群間の類似性をレイアウトに反映しない。我々は二つの条件を満たすために MDS と ChronoView を融合することを試みた。

2. 関連研究

2.1 タイムスタンプの表現

イベント群に含まれるイベントのタイムスタンプを表現する手法として Strip chart, Line chart, Bar chart といったものがある [3]。これらは横軸に時間軸を取り、縦軸に一定時間間隔で集計したイベントの頻度を取ることで一つのイベント群を表現する。多くのイベント群を同時に表現する際はイベント群の表現を重ねるといった手法がとられるが、イベント群の数が多くなると表現が煩雑となり個々のイベント群の時刻情報が読み取りづらくなる。ThemeRiver [4] は積み上げグラフにおける値の変化を連続的にしたような表現を行う。多くのイベント群を同時に空間効率よく表現するが、イベント群の数が多くなると視覚的に比較する組の数が増え、類似した時系列パターンを発見しづらくなる。

他の手法として螺旋状の時間軸にタイムスタンプを配置する手法がある [5][6]。螺旋状の時間軸にタイムスタンプがひとつひとつ配置される。これらの手法ではアナログ時計のように角度が発生時刻を表現する。同時刻に発生したイベントのタイムスタンプは直線状に並ぶため周期性を把握し易い。しかし 1 つのイベント群を表現するために大きな描画領域が必要となってしまう、多くのイベント群を並べて比較するといったことは難しい。

2.2 グリフの配置方法

多変量データを空間効率良く表現するためにグリフがよく用いられる。多変量データにおける変量のまとまりをレコードと呼び、各レコードをひとつのグリフで表現する。多くの場合レコードのパターンが表現されるようにグリフの配置方法を定める。グリフの配置はレコードの変量あるいはレコード間の関係性によって決められる [7]。

よく利用されるグリフの配置方法は、変量をグリフの配置座標へ割り当てるという方法である。この配置方法としては各レコードの任意の変量を座標値に割り当てる、変量の組を 2 次元に写像するといったものがある。例えば散布

図のように 2 変量をそのまま座標値に割り当てる手法がある。2 変量の関係が直接表現され、座標点の密度からレコードの分布が表現される。しかしこのような手法では割り当てない変量について説明できない。2 変量でレコードを十分に説明しない場合、多変量データの特徴を効果的に表現できない。Nguyen らは頻度の時系列変化が表現されたグリフを、レコードの発生場所という属性を用いて地図上に配置している [8]。これはレコードの地理的な関係を反映したレイアウトを提供するが、レコード間の類似関係は反映しない。レコードの変量の組を 2 次元座標へ写像する主要な方法として次元削減手法がある。次元削減手法はレコード間の類似性を 2 次元上で表現する。グリフは写像前の変量を表現するために用いられ、多変量データを効率的に表現する [9][10]。しかし多変量の特徴はレイアウトに反映されない。

データの変量でなく、レコード間の暗黙的あるいは明示的な関係性によって配置を求める手法がある [7]。例えばレコード間に、ある順序関係が定義される場合、その順序関係に従ってグリフを並べることができる。Infuse [11] は特徴量でランク付けした順序関係で大量のグリフを整理して表現する。他の例として、ある変量がレコード間の接続関係や階層構造を説明する場合、グラフやツリー構造による接続関係を表現した配置が可能である。ClockMap [12] はネットワークなどの階層構造を円状のツリーマップ構造で表現しており、個々のグリフが時刻情報を表現する。我々が設計した手法は配置の際に発生時刻の類似関係を考慮する点が共通する。それに加えてイベントの発生時刻を座標に割り当て、イベント群の位置が時間的な特徴を表現する点が異なる。

3. グリフ配置の条件

グリッド状のレイアウトは Small Multiples の最も単純なレイアウトである。しかしそのような単純なレイアウトでは時刻分布の傾向や類似性の把握を支援できない。例えば、多くのイベント群がある特定の時刻に集中して発生したことや、いくつかのイベント群が類似した時刻分布を持つかどうかを見つけることは難しい。これらの理由を受けて、我々は時刻分布を効果的に示すための二つの条件を定めた。

R1 グリフの位置が各イベント群の時刻分布を表現する。グリフの位置が時刻分布を表現できる場合、発生時刻の傾向をより容易に掴む事が出来る。

R2 グリフ位置の近さがイベント群の時刻分布の類似性を表現する。近さで類似性が表現されれば、類似したグループやはずれ値を容易に発見できる。

4. 時刻分布表現のための ChronoView

図 1 は ChronoView の配置規則を示したものである。ChronoView は時計の文字盤のような座標系を持ち、時間軸は円周で表現される。つまり、ChronoView は時間が持つ周期性を表現するものである。図 1 では、周期は 24 時間としている。ひとつのタイムスタンプは円周上つまり時間軸上の位置で表現される。イベント群はそのイベント群に含まれるイベントのタイムスタンプの位置の重心に配置される。

プロットされた円とその色はイベント群の情報を表現するのに用いられる。本論文では円の面積をイベント群に含まれるイベントの数を表現するために用いている。

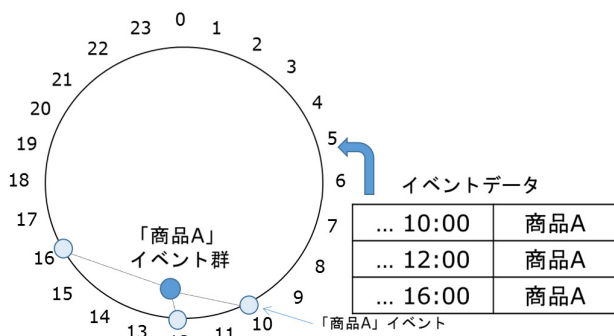


図 1 ChronoView の配置規則

イベント群の位置がイベント群に含まれるイベントの発生時刻を大まかに表現する。ある時刻に集中してイベントが発生したイベント群は円周に近い位置に配置される。一方で、まばらな時間にイベントが発生したイベント群は円盤の中心近くに配置される。そのため、ChronoView によるグリフのレイアウトは条件 R1 をある程度満たす。

ChronoView はタイムスタンプの集合を 2 次元平面上に落とし込む、一種の次元削減を行う。しかしこれが表現における曖昧性を生じさせてしまう。近くに配置された 2 つのイベント群が必ずしも似た時刻分布を持っているとは限らない。例えば 0 時と 12 時に 1 回ずつイベントが発生したイベント群は円盤の中央に配置される。だが 6 時と 18 時に 1 回ずつ発生した場合も中央に配置されてしまう。これは時刻分布の類似性が小さいイベント同士が近くに配置されうるということを意味する。このため ChronoView は条件 R2 を満たさない。

5. 類似性表現のための MDS

多次元尺度構成法 (MDS) は次元削減手法の一種である。MDS は項目間の (非) 類似度が与えられているとき、低次元空間上で項目間の類似関係を表現する。二つの項目が類似している場合、それらは近接して配置される。二つ

の項目が非類似であるほどそれらは遠く離されて配置される。各項目が n 次元データに関連付けられている場合、二つの項目の (非) 類似度はしばしば n 次元空間上のユークリッド距離やコサイン距離などで与えられる。

タイムスタンプを 1 時間毎に集計することによって、イベント群のタイムスタンプを 24 次元データとして集約できる。我々は MDS を用いることで、イベント群の時刻分布の類似性を保持したレイアウトを得る事が出来る。

MDS は視覚的な距離で項目間の類似性を表現するため、条件 R2 を満たす。しかし座標系に意味を持たない MDS が提供するものは類似関係のみである。イベント群の位置はイベント群の時刻分布の特徴を表現しない。つまり MDS ではイベント群の時刻分布を 2 次元平面状で表現できない。これは MDS が条件 R1 を満たさないことを意味する。

6. レイアウト構築のプロセス

我々は二つの条件を満たすレイアウトを得るために、条件 R1 を満たす ChronoView と条件 R2 を満たす MDS を融合することを試みた。本論文では、ChronoView と MDS による 2 種類のレイアウトをそれぞれ「ChronoView-Layout」と「MDS-Layout」と記す。さらに、それらを融合したレイアウトを「Fusion-Layout」と記す。

始めに ChronoView によるグリフのレイアウトを作成し、それからそのレイアウトを MDS によって歪ませる。これは ChronoView のレイアウトを歪ませることであり、分析者の誤解を生む恐れがある。そのような誤解を避けるために、我々はその歪みを提示する表現を導入する。

6.1 グリフの設計

我々はイベント群が持つイベントの発生時刻の特徴を表現するグリフを設計した。図 2 は我々が設計した 2 種類のグリフを示したものである。これらは 1 時間ごとにイベントの頻度を集計した 24 次元データを表現する。方向と色が発生時刻を表し、色分けされた図形の面積が 1 時間ごとの頻度を表す。

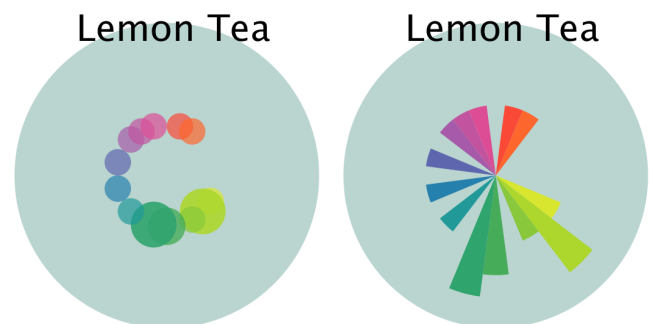


図 2 設計した 2 種類のグリフ

グリフの設計において、ドメイン上の知識が効果的に知覚されるように変量を視覚的変数へ割り当てることが重要

である [13]. 時間の周期性は時計に見られるように円周や螺旋といったイメージで表現される. また具体的なイメージを表現するグリフ形状を用いることが効果的であることがいくつかの実験で示されている [14]. 今回設計したグリフでは, 周期的なイメージが知覚されるように方向と色を発生時刻を意味する視覚的変数とした. 加えて各方向に円を配置し, 円の面積でイベントの発生頻度を表すことで発生時刻の頻度パターンを表現する. 同様の形状を持つグリフとして StarGlyph や ClockGlyph などがある [15]. 図 2 はこれらのグリフと比べてより小さい面積で発生時刻の分布を表現することを目指したものである. 設計したこれらのグリフが ChronoView における個々の時刻分布の曖昧性を解消する.

6.2 類似性表現のためのレイアウト

我々は MDS を用いて時刻分布 $v_i \in R^{24}; i = 1, \dots, m$ を $p_i \in R^2$ へと写像する. ここで m はイベント群の総数である. 時刻分布は 1 時間ごとにイベントの数を集計した 24 次元データとする. 我々はイベントの絶対的な頻度でなく発生時刻のパターンを強調したいため, MDS を適用する際のイベント群間の類似度としてコサイン類似度を用いる.

上記で得られる MDS-Layout はイベント群間の発生時刻の類似性を反映するが, 時刻分布の特徴は反映しない. そのため, 我々は MDS-Layout の座標系を時刻分布を表現する座標系へと変形する. 我々は MDS-Layout を Procrustes 回転によって回転させる. Procrustes 回転は 2 点群間の位置に関する 2 乗誤差が最小となるように点群を移動させる手法である. Procrustes 回転は対応の取れた 2 点群に対して並進, 回転, 一様スケーリングを行う. 回転された MDS-Layout を Rotated-Layout と呼ぶ.

Procrustes 回転は MDS-Layout を ChronoView の座標に近づくよう回転させる. しかし回転による変換のみでは個々のイベント群の時刻分布を十分に表現できない. この問題を解決するためにばねモデルを用いる.

6.3 グラフの構築

ばねモデルによる変換を行うためにイベント群をノードとしたグラフを構築する. Rotated-Layout におけるノード間のユークリッド距離に応じてエッジを接続する. 具体的にはノード間の距離が閾値 L 以下であるノード間にエッジを接続する. 閾値 L が大きいほど変形によってエッジの長さはイベント群間の類似性を表現する理想的な距離とする.

6.4 ばねモデルによる Fusion-Layout の作成

各ノードの位置が ChronoView-Layout 上の位置へ近づくようにグラフを変形する. このときノード間のエッジ長

を保つ力を適用することで ChronoView-Layout にイベント群間の類似性を反映する. 各イベント群の時刻分布を表現する配置を求めながらも, 類似性を読み取り可能な Fusion-Layout が作成される.

始めにグラフの各ノードが ChronoView 上の座標に近づくための力を与える (式 (1)). ここで, $p_i \in R^2$ は Rotated-Layout 上のノード座標であり, $g_i \in R^2$ は ChronoView 上のノード座標である. c_c はノード同士を近づける度合いを表す定数である. 描画フレーム毎に座標 p_i が g_i に微小移動する.

$$f_c = c_c |g_i - p_i| \quad (1)$$

次にイベント群間の類似性を反映するために Eades のばねモデル [16] を拡張した力学モデルを適用する. ノード間の距離が一定に保たれるよう各ノードに斥力とばね力を与える. 式 (2) のばね力 f_g がエッジで接続されたノード間に働き, 式 (3) の斥力 f_r が全ノード間に働く. ここで c_g , c_r は定数である. d は変形中のレイアウト上のノード間距離である. d_0 はレイアウト作成時におけるばねの理想的な長さである. d_0 には Rotated-Layout 上のノード間のエッジ長が与えられる.

$$f_g = c_g \log \frac{d}{d_0} \quad (2)$$

$$f_r = c_r \frac{1}{d^2} \quad (3)$$

f_g は $d > d_0$ の場合はノード同士を近づけ, そうでない場合はノード同士を離すばね力となる. ノード間距離が d_0 を保つような変形が行われるため, ChronoView-Layout へのイベント群間の類似性の反映が可能である.

6.5 空間のゆがみ表現

Fusion-Layout は ChronoView-Layout に MDS-Layout の特徴を反映したものである. 時刻分布の類似性を反映したことにより, ChronoView の座標系が歪められるため, 発生時刻の読み取りに関して誤解を生む可能性がある. 我々は ChronoView の座標系を表す同心円を描画することで, ChronoView の座標系がどの程度歪められたのかを提示する. 図 3 では, 歪められた同心円が座標系の歪みを表現している. 時刻分布の類似性の反映によってイベント群の位置が変化した場合, そのイベント近くの座標軸が歪む. 歪められた同心円表現が ChronoView-Layout 上から Fusion-Layout 上へのイベント群の位置変化の把握を助ける. このため類似性の表現が優先されたレイアウト上であっても, 分析者はイベント群の時刻分布を誤解なく読み取ることができる.

7. レイアウト結果

本節では Fusion-Layout が効果的に時刻分布の特徴と類

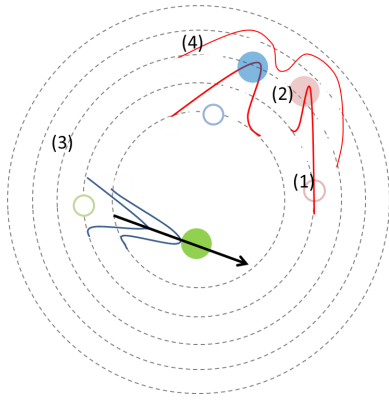


図 3 ゆがみ表現. (1)ChronoView-Layout 上での座標. (2)Fusion-Layout 上での座標. (3) ChronoView の座標軸 (4) 座標軸のゆがみの表現

似性を表現するかを人工的に作成したデータセットを用いて説明する。我々は発生時刻に偏りがあるイベントを生成し、それにランダムに作成したイベントを加えたデータ群をデータセットとして用いた。図 4(a), (b), (c) は、そのようにして作成したデータセットを、ChronoView-Layout, MDS-Layout, Fusion-Layout それぞれでレイアウトした結果を示している。

図 4(a) では ChronoView の座標系を示す同心円が並べられており、その平面上にイベント群が配置されている。例えばイベント群「25」「6」「17」は 22 時～0 時付近の円周近くに配置されている。また、イベント群のグリフを見ると、22 時～0 時を表す方向に円がプロットされていることがわかる。これはイベント群に含まれるイベントがその時間帯に発生したことを意味する。図 4(a) を俯瞰して見ると、8～10 時の時間帯と 17 時から夜中にかけて発生したイベントが多いことがわかる。中央付近のイベント群は位置から発生時刻を読み取ることが難しくなるが、グリフで表現された時刻分布を読み取ることでより正確な発生時刻の分析が可能である。

図 4(a) において「0」「1」「4」イベント群が近接している（点線内）。これらのイベント群のグリフを見てみると、それぞれ円がプロットされている時間帯が異なっている。これは非類似な時刻分布を持つイベントが近接してしまっていることを意味する。ここで、図 4(c) を見てみる。「0」「1」「4」イベント群はそれぞれ離れており、より似た形のグリフを持つイベント群近くへ移動している。その位置関係は MDS-Layout(図 4(b)) で見られる位置関係に近い。これはイベント群間の時刻分布の類似性が効果的に反映されていることを意味する。

図 4(c) と図 4(a) を見比べると、各イベント群の位置は少しずつ変化している。だが、多くのイベント群の位置関係はある程度保たれており、また各イベント群の時刻分布はグリフによって表現されている。このため、発生時刻の類似性が表現された Fusion-Layout (図 4(c)) で同時に発生

時刻の特徴を読み取ることができる。イベント群「0」「1」「4」は Fusion-Layout (図 4(c)) において大きく位置を変化させているが、歪んだ同心円を観察することで本来のイベント群の位置を把握することができる。例えばイベント群「1」に関しては、円盤の中央から 16 時の方向へ大きく移動したことを同心円の形状が示している。

8. まとめ

我々はイベントの時間的な特徴とその類似性を表現するグリフベースの可視化手法を開発した。この研究の重要な貢献は、グリフによって時間的特徴を小面積で表現したこと、そのグリフを効果的に配置する手法を提案したこと、イベント群の時間的特徴と類似性を同一平面上で表現したことである。このレイアウト手法の実現のため、我々は ChronoView のレイアウトと MDS のレイアウトを融合した。また、我々は融合によって生じるレイアウトの歪みを補うための表現を提案した。この表現は分析者に時刻分布の正確な読み取りを可能にすると期待される。本論文で提案した手法によって、分析者は大量のイベントに関して発生時刻のパターンや似た発生傾向を持つイベントの分析が可能になる。

今後の課題は以下の 2 点である。第 1 の課題はスケール合わせの問題を克服することである。同一平面上で各レイアウトの特徴を正確に表現するには各レイアウトの座標系を同じスケールで扱う必要がある。しかし異なる座標系を持つ ChronoView と MDS のスケールを正確に合わせる事は現状難しい。第 2 の課題として、今回設計したグリフが時刻分布の曖昧性の解消に有効であるかを被験者実験によって検証する必要がある。

謝辞 本研究は JSPS 科研費 JP17K00264 の助成を受けたものです。

参考文献

- [1] J. B. Kruskal and M. Wish. *Multidimensional Scaling*. Sage, London, 1978.
- [2] S. Shiroy, K. Misue, and J. Tanaka. Chronoview: Visualization technique for many temporal data. in *Proceedings of 16th International Conference on Information Visualisation*, pp. 112–117, 2012.
- [3] W. Javed, B. McDonnell, and N. Elmqvist. Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 16, No. 6, pp. 927–934, 2010.
- [4] S. Havre, B. Hetzler, and L. Nowell. Themeriver: Visualizing theme changes over time. in *Proceedings of IEEE Symposium on Information Visualization*, pp. 115–123, 2000.
- [5] John V. Carlis and Joseph A. Konstan. Interactive visualization of serial periodic data. in *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pp. 29–38, 1998.
- [6] M. Weber, M. Alexa, and W. Müller. Visualizing time-series on spirals. in *Proceedings of the IEEE Symposium*

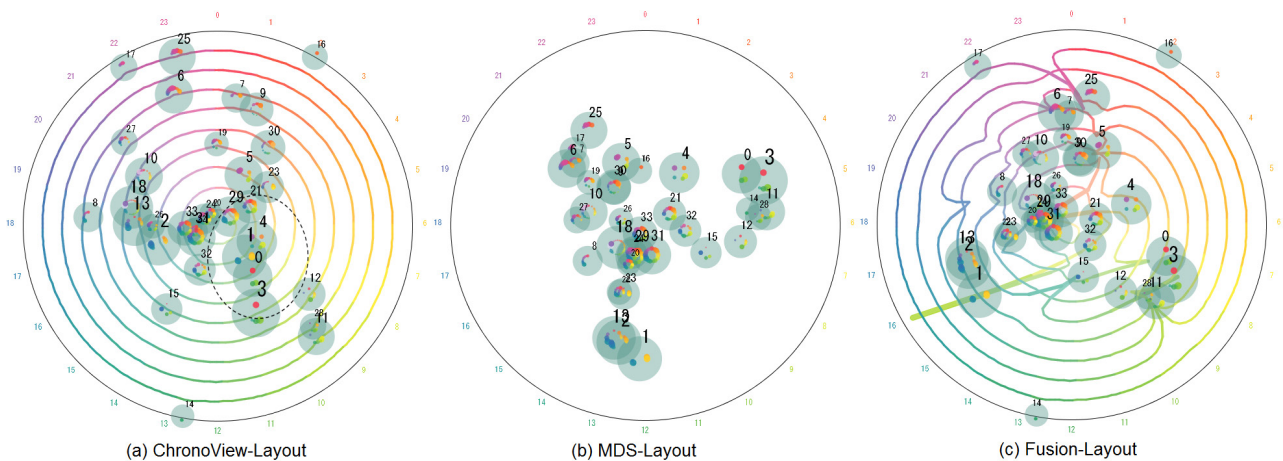


図 4 (a) ChronoView-Layout: 似た時刻分布を持つイベント群同士が近くに配置される。類似性が必ずしも反映されない。 (b) MDS-Layout: 時刻分布の類似性が表現されるが時刻分布の特徴は表現されない。 (c) Fusion-Layout: 時刻分布の類似性が ChronoView-Layout に反映されている。

on *Information Visualization*, pp. 7–14, 2001.

- [7] M. Ward. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, Vol. 1, No. 3–4, pp. 194–210, 2002.
- [8] D. Q. Nguyen, C. Tominski, H. Schumann, and T. A. Ta. Visualizing tags with spatiotemporal references. in *Proceedings of the International Conference on Information Visualisation*, pp. 32–39, 2011.
- [9] M. Keck, D. Kammer, T. Gründer, T. Thom, M. Kleins-teuer, A. Maasch, and R. Groh. Towards glyph-based visualizations for big data clustering. in *Proceedings of the 10th International Symposium on Visual Information Communication and Interaction*, pp. 129–136, 2017.
- [10] W. Cui, Y. Wu, S. Liu, F. Wei, M. X. Zhou, and H. Qu. Context preserving dynamic word cloud visualization. in *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis)*, pp. 121–128, 2010.
- [11] J. Krause, A. Parer, and E. Bertini. Infuse: Interactive feature selection for predictive modeling of high dimensional data. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20, No. 12, pp. 1614–1623, 2014.
- [12] F. Fischer, J. Fuchs, and F. Mansmann. Clockmap: Enhancing circular treemaps with temporal glyphs for time-series data. in *Proceedings of EuroVis Short Papers, Eurographics*, pp. 97–101, 2012.
- [13] M. O. Ward. Multivariate data glyphs: Principles and practice. In *Handbook of Data Visualization*, pp. 179–198, Berlin Heidelberg, 2008. Springer.
- [14] J. Fuchs, P. Isenberg, A. Bezerianos, and D. Keim. A systematic review of experimental studies on data glyphs. *IEEE Transactions on Visualization and Computer Graphics, Institute of Electrical and Electronics Engineers*, Vol. 23, No. 7, pp. 1863–1879, 2017.
- [15] J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM*, pp. 3237–3246, 2013.
- [16] P. Eades. A heuristic for graph drawing. *Congressus Numerantium*, Vol. 42, pp. 149–160, 1984.