

時刻情報付きデータの分析を支援する視覚的ツールの開発

白井 智子[†] 三末 和 男^{††} 田中 二郎^{††}

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などの様々な分野へ応用することができる。本研究では、時刻情報付きデータの分析を支援するための表現手法として、ChronoViewを開発した。ChronoViewは、アナログ時計の文字盤のような円を表示し、イベントをその中に配置することで、時刻情報付き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。また、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のツイートデータに着目したユースケースによって、ChronoViewの有用性を示した。

Development of visual tool to support analysis of data containing temporal information

SATOKO SHIROI,[†] KAZUO MISUE^{††} and JIRO TANAKA^{††}

Data containing temporal information represents data that records events such as behavior and time when appeared the events. If we analyze this data, we can understand behavior patterns, and we can apply our knowledge to many field such as the marketing field. We developed ChronoView, that is a visualization technique to support the analysis of data containing temporal information. ChronoView illustrates a circle-like analog clock dial which shows time-stamps and distributes events in this circle. In this way, our view presents in a visual manner an approximate distribution of the relationship between the events recorded in the data containing temporal information and the time-stamps. Moreover, our view enables further analysis of the details regarding the relationship between the events and the time-stamps, by adding the interaction according to the use. We showed the usefulness of ChronoView by performing use case focused on tweet data in Twitter.

1. はじめに

本論文で対象とする「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。これ以外にも、ショッピングサイトでの商品の購買履歴、Twitter のツイート、事件や事故の発生記録など、様々な時刻情報付きデータが存在する。このような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がわかる。これらの特徴の把握は、マーケティングの分野を初めとする様々な分野で期待されている。

一般に、イベントはひとつ以上の時刻で発生する。たとえば、ある Web ページが、10 時、11 時、12 時に閲覧されたとすると、その Web ページへのアクセスは 3 つの時刻で発生したことになる。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が、その多くは時刻の集合ではなく個々の時刻を視覚的に表現するものである。そのため、時刻への依存性や周期性といった時刻の集合から導かれる特徴の把握は容易ではない。

本論文では、時刻情報付きデータを分析することを支援するため、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を提案する。時刻の集合を位置に対応付けることで、イベントの時刻への依存性や周期性などの特徴発見が容易になる。本表現手法は、アナログ時計の文字盤のような円を基本とした手法である

[†] 筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻

Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba

^{††} 筑波大学システム情報系

Faculty of Engineering, Information and Systems, University of Tsukuba

ことから「ChronoView」と名付けた。円を基本とした可視化手法は古くから存在するが、近年においても様々なドメインで活発に利用されている¹⁾。

本論文では、第2章で、関連研究について述べる。第3章で、開発した ChronoView の表現手法と表現のバリエーションについて述べ、第4章で、それらを合わせて開発した分析ツールについて述べる。第5章で、ChronoView のユースケースについて言及し、第6章で、論文のまとめとする。

2. 関連研究

2.1 時刻情報付きデータの可視化

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示方法を開発した²⁾。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³⁾。これらの表示方法では、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個くらい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、イベントの周期性の概観を把握することや、複数のイベント同士の周期性の比較ができない。

Nguyen らは、時間参照されるタグを可視化する表現方法を開発した⁴⁾。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、時間のみに着目した場合のような特定の時刻に着目した場合のイベントの発生の比較は可能である。しかし、イベントの周期性の把握や、年月日、曜日、時間から複数の項目を組み合わせた場合のイベント発生の比較ができない。

2.2 円形の表現を用いた可視化

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⁵⁾。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。

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

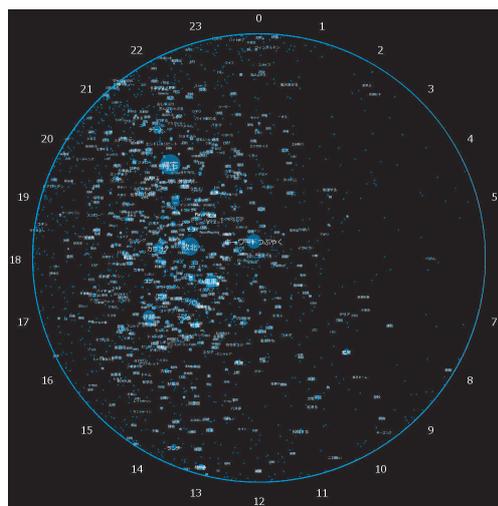


図1 ChronoViewの概観
Fig. 1 ChronoView

を開発した⁶⁾。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。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⁷⁾。しかし、このような螺旋表現を用いた場合、螺旋の巻き数が増加すると、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。

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⁸⁾。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。しかし、この手法では、長期的な時刻情報を持つイベントを視覚的に提示できない。

3. ChronoView

本研究では、イベントに付随する時刻の集合を二次元平面上の位置で表現する、ChronoView を開発した。ChronoView の概観を図1に示す。

3.1 時刻集合の表現

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 のときに、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 $f_0: U \rightarrow \mathcal{R}^2$ によって次のように定義できる。

$$f_0(t) = (r \cos \theta, r \sin \theta) \quad (1)$$

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 $f: 2^U \rightarrow \mathcal{R}^2$ によって次のように定義できる。

$$f(T) = \frac{1}{|T|} \sum_{t \in T} f_0(t) \quad (2)$$

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1の例に示す通り、本論文では 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。

このイベントの配置方法により、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。

3.2 イベントの属性の表現

イベントには、イベントが発生した時刻の集合以外にも、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⁹⁾ を利用する。

図1では、水色の円と白いラベルを用いて個々のイベントを表している。円の面積は、イベントの発生頻度に比例させた。さらに、文字サイズも頻度に応じて変えることで、頻発するイベントほど目立つようにしている。

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。

3.3 付加情報の表現

提案手法では、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。

提案手法では、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。そのため、この曖昧性を解消するための工夫を検討した。

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2は、「帰宅」というイベントが発生した時刻を示した例である。このような放射状の線を表示する

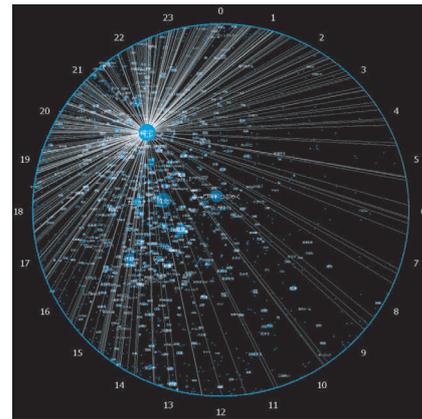


図2 イベントが発生した時刻を示す放射状の線の表示
Fig.2 Radiation to express the outbreak time of an event.

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。

提案手法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。

これらの付加的な図形の利用は、便利である一方、視覚的表現を繁雑にしてしまう可能性がある。そのため、マウスクリック等の操作に対応付けて、必要な時だけ一時的に表示する。

4. 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3章で示した ChronoView を利用した分析ツールを開発した。ツールの開発にあたり、ChronoView のユースケースとして Twitter のツイート扱うため、ツイートデータにフォーカスした。図3は、開発した分析ツールの外観である。

4.1 ChronoView

画面中央で、データを視覚的に提示する ChronoView を提供する。ここでは、all ビュー、weekly ビュー、day ビューの3つのビューに切り替えることができる。

all ビューでは、全データがまとめて一度に表示される。all ビューでは、気になるイベントをクリックすると、図2のような放射状の線が表示され、イベントが実際に発生した時刻を把握することができる。

weekly ビュー(図4)では、各イベントが平日と休日に細分化され、色分けされて配置される。この方法により、平日と休日で同じイベントの発生時刻の違い

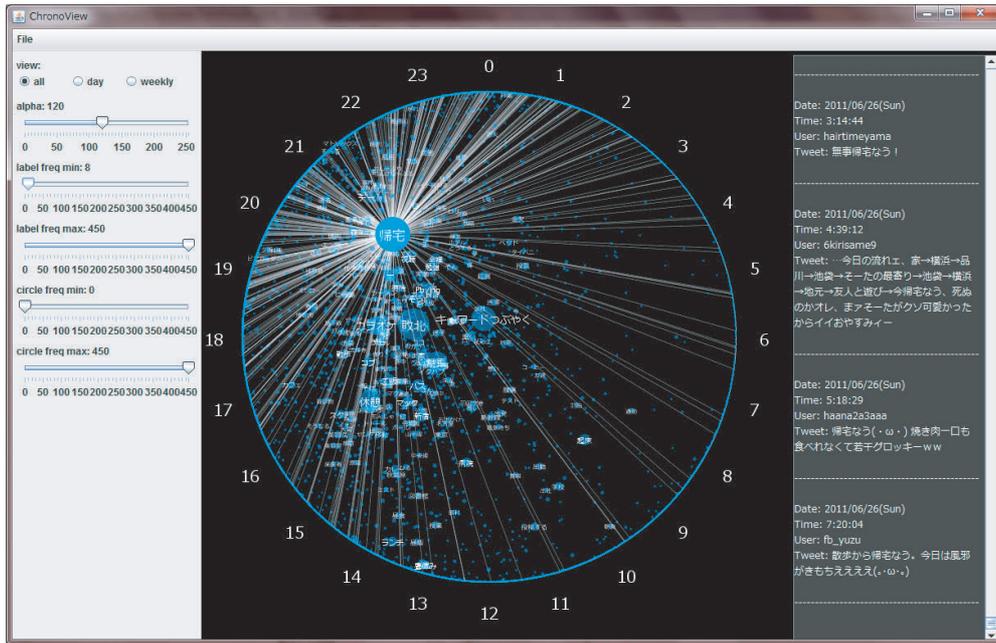


図 3 分析ツールの概観

Fig. 3 Overview of our analysis tool.

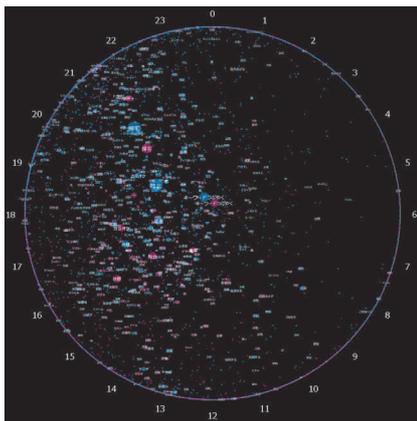


図 4 それぞれのイベントを平日と休日で色分けした場合のビュー
Fig. 4 Events in weekdays and holidays are distinguished by colors.

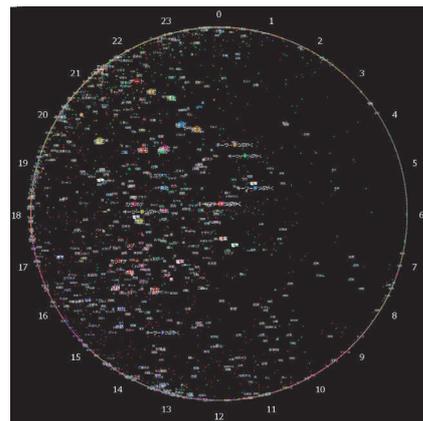


図 5 それぞれのイベントを曜日ごとに色分けした場合のビュー
Fig. 5 Events in days of the week are distinguished by colors.

を把握することができる。

day ビュー (図 5) では、各イベントが曜日別に細分化され、色分けされて配置される。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曲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。

4.2 詳細表示画面

画面右側のビューは、ChronoView の各ビューで気になったイベントをクリックすると、そのイベントを含むツイートを読むことができる詳細画面である。

4.3 表現の操作画面

画面の左側は、ビューを切り替えるためのラジオボ

タンや、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。スライダでは、イベントの発生頻度の最大値と最小値、イベント表示円の不透明度 (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

表 1 時刻情報付きデータの一例

Table 1 An example of data including temporal information

ユーザ名	年月日	曜日	時間	ツイート	イベント
white_luc	2011/06/20	Mon	17:30:24	カフェなう	カフェ
yamahakusyon	2011/06/20	Mon	17:33:09	帰宅なう	帰宅
adaajmdap	2011/06/20	Mon	17:35:12	新宿なう	新宿

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。

5. ユースケース

ChronoView のユースケースとして、Twitter のツイートデータを対象にした分析作業を行った。

5.1 利用データ

データは「起床なう」「新宿なう」のように「なう」というキーワードを含んだツイートである。「なう」を含むツイートは、人間の日常の何気ない行動の表れとして考えられ、その分析は、マーケティング分野への応用が期待されている。

ツイートデータは、Twitter の StreamingAPI を利用して収集した。このツイートデータは、ユーザ名、年月日、曜日、時刻、ツイート文を持つ「なう」を含むツイート文を、MeCab 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単語列を抽出した。ツイートデータの一例を表 1 に示す。

今回のユースケースでは、2011 年 6 月 20 日から 6 月 26 日までの 18584 ツイートのデータで、それぞれ異なるイベントが 6318 件含まれるデータと、2011 年 9 月 13 日から 9 月 19 日までの 19805 ツイートのデータで、それぞれ異なるイベントが 9494 件含まれるデータを扱う。それぞれ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、さらにこの 2 つのデータの表示結果を比較する。

これらのデータには、一部欠損がある。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。しかし、ユースケースを行うにあたり、差し支えない程度の欠損であると思われる。

5.2 観 察

6 月 20 日からのツイートデータ分析した時「起床」、「ランチ」、「お昼」というイベントの発生頻度は、円の大きさから、どれも同じくらいの発生頻度であった。

しかし、「起床」は右下中央あたりに位置しており、発生時刻にばらつきがありそうということがわかった。一方、「ランチ」や「お昼」は 13 時から 14 時近辺の円の縁に近い部分に位置し、特定の時間に集中しているそうということがわかった。

実際にいつ起床しているのか、いつランチを食べているのかを調べるため、各ラベルをクリックして放射状の線を表示した。「起床」は 6 時から 12 時あたりに線が集中しており、幅広い時間帯をカバーしつつ、18 時過ぎから 21 時の間にひかれる線も数本あった。また「ランチ」は 11 時から 14 時の間に線が集中しており、特定の時間に依存していたが、16 時から 19 時の間にひかれる線が数本あった。「起床」、「ランチ」を含むツイートを確認すると、起床時間が遅かったり、ランチが遅かったりしたのは、仕事や学業の関係で遅くなったというツイートを見つけた。

6 月 20 日からのデータと 9 月 13 日からのデータを比較したとき「起床」、「ランチ」、「お昼」、「帰宅」などの日常の行動が、どちらのデータにおいても、ほぼ同じような位置に配置されていた。しかし、それぞれのビューを、平日と休日のビューに切り替えると「起床」は 6 月 20 日からのデータが、平日よりも休日の方が起床が遅いという結果に対し、9 月 13 日からのデータでは、平日よりも休日の方が起床が早いという結果が見られた。

5.3 考 察

ChronoView で、円やラベルのサイズと位置を用いて時刻の集合を表示することで、あるイベントがどれくらいの頻度で、特定の時間帯のみに発生しているのか、幅広い時間帯に発生しているのかなどのおおよその特徴を把握できた。また、発生頻度が同じくらいのイベントであっても、特定の時間帯に依存しているイベントもあれば、幅広い時間帯に発生しているイベントもある、といった比較を可能にした。

気になったイベントのラベルをクリックして放射状の線の表示をすることで、イベントが実際にいつ集中して行われているのか、どのくらい発生時間が分散しているのかを把握することができた。この手段により、時刻の集合を位置で表しただけでは発生時間帯に曖昧

<https://dev.twitter.com/docs/streaming-api>
<http://mecab.sourceforge.net/>

性が残るようなイベントであっても，その曖昧性を改善することができた．

用途に応じてビューを切り替えることにより，曜日ごとや平日と休日での，各イベントの発生時間の違いを分析できた．実際にどのようなツイートがされていたのかを表示することで，イベントの詳細についても把握することができた．このように，ChronoViewでは，ビューから全体的な概観を把握しつつ，表現のバリエーションにより，さらに詳細を分析することで，日常的なイベントの中の異なる傾向の分析が可能である．

ビューの切り替えや放射状の線の表示により，日常的なイベントについても，異なる期日間で，異なる傾向を示していることがわかった．ChronoViewでは，異なる期日間の複数のデータ同士を比較した場合，同一のイベントの中でも，異なる傾向を示すことがわかった．

6. ま と め

本論文では，時刻情報付きデータの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，イベントの発生頻度とそのイベントの発生時刻を位置で示すことで，イベントと時刻の関係性を提示した．この表現方法により，イベントがどのような周期で発生しているのか，おおよその特徴を把握をすることを可能にした．また，放射状の線の表示やビューの切り替えといった表現のバリエーションを追加することで，イベントと時刻の関係性について，さらに詳細な分布の分析することができた．

本研究では，Twitter のツイートデータにフォーカスして，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．本論文ではツイートデータの分析に着目したが，ChronoView は，その他の分析においても汎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を適用した評価実験を行い，ChronoView の汎用性を明らかにしていくことを目指す．

参 考 文 献

1) 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).
2) Aleks Aris , Ben Shneiderman , Catherine Plaisant , Galit Shmueli , Wolfgang Jank : Representing Unevenly-Spaced Time Series Data for Visualization and Interactive Exploration , *Human Computer Interaction INTERACT 2005* , pp.835–846 (2005).
3) Paolo Buono , Aleks Aris , Catherine Plaisant , Amir Khella , Ben Shneiderman : Interactive Pattern Search in Time Series , *Proceedings of Conference on Visualization and Data Analysis (VDA 2005)* , pp.175–186 (2005).
4) Dinh Quyen Nguyen, Christian Tominski, Heidrun Schumann , Tuan Anh Ta : Visualizing Tags with Spatiotemporal References , *15th International Conference Information Visualisation (IV2011, July 12-15, 2011, London, UK)* , pp.32–39 (2011) .
5) Kazuo Misue: Overview of Networks Information by Using Anchored Maps, *11th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES2007)*, pp.1269-1276 (2007).
6) John V. Carlis , 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) .
7) Christian Tominski , Heidrun Schumann: Enhanced Interactive Spiral Display , *Proceedings of the Annual SIGRAD Conference 2008* , pp. 53-56 (2008) .
8) Pierre Dragicevic , 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) .
9) J. Bertin: *Semiology of Graphics: Diagrams, Networks, Maps*, Univ. of Wisconsin Press (1983).