Vol.2013-HCI-152 No.22
2013/3/14

順位と量の推移の視覚的表現を用いたトレンド分析支援ツールの開発

大久保 心織[1,a)]　三末 和男[2]　田中 二郎[2]

概要：
時刻と量を伴った順位データの可視化によって，流行の期間や始まりなどのトレンドを分析するためのツールを開発した．提案する可視化手法では，順位と量の推移を同時に表現することで，従来の表現手法に比べてデータの全体像が把握しやすくなり，個々の事象の動向についてより詳細な分析が可能になった．また，事象の持つテキストラベルから自動的に色相の決定を行うことで，視覚的に事象のカテゴリを確認しやすくなった．本ツールを用いて，twitter のトピックデータを対象としたトレンド分析を行った．その結果，トレンドやデータの全体像についての理解が得られたと同時に，メディアに対する人々の反応のデータをツールに適用することで，メディアごとの受け手の反応の差も示せるという可能性も提示した．

Development of a Tool for Analyzing Trends with Simultaneous Visualization of Changes of Rank and Value

Abstract: We developed a tool for analyzing trends featured by visualization of ranking data with time stamp and value. Compared to existing visualizing techniques, representing change of value and ranking at the same time in our visualization enables us to grasp trends of both the entire data set and each occurrence more easily. Also, it tells us category of event with hue of color which estimated automatically from its text label. Using our tool, we analyzed trends of topic data from twitter. As a result, we can observe the trend of topics and have an overview of the data. In addition, our tool allows the possibility to present differences between the users' ideas of different media by using data regarding the reactions of users in media.

1. はじめに

我々の暮らしの中で，学校のテストの成績や商品の売上，流行歌など，さまざまなものに順位が付けられ提示されていることは非常に多い．このような順位は，テストの点数や売上金額などの量的値を元に決定されることがしばしばである．さらに，順位の推移を観察することで，その時における流行や世間の様子についての情報が得られる．こうしたトレンドを分析することは，マーケティングや政治情勢分析などの幅広い分野で必要とされている．すなわち，何が流行しているか・していないか，あるいはいつごろ流行し始めたか・流行が終了したかということを，過去・現在・未来にわたって知りたいという要求が存在する．

トレンドを詳細に分析するためには，順位と量の両方の推移を観察する必要がある．これに対して，従来多くの可視化手法が提案されてきた．しかし，折れ線グラフや表を用いた一般的な視覚的表現は，順位または量どちらかの推移のみを表現するようなものであり，それらの推移を同時に表現したい場合には適さない．

そこで本研究では，上記のような，ある事象の順位が時間によって変化する量を伴ったデータ（以下「時刻付き量的順位データ」）を用いたトレンド分析の支援を行うことを目的とする．また，それに対し，視覚的表現を用いるものとする．加えて，本研究で提案する視覚的表現により，順位及び量の推移を同時に可視化し，より詳細なトレンド分析を行えるようにすることを目標とする．

本研究の貢献として次の 2 点が挙げられる．

(1) 順位と量的値を同時に表現することで，データの全体像が把握しやすい表現手法を開発したこと

[1] 筑波大学 情報学群 情報メディア創成学類
College of Media Arts, Science and Technology, School of Informatics, University of Tsukuba
[2] 筑波大学 システム情報系
Faculty of Engineering, Information and Systems, University of Tsukuba
[a)] okubo@iplab.cs.tsukuba.ac.jp

(2) 事象の持つ文字列を利用して，事象のカテゴリを反映した色相値の自動算出技術を開発したこと

1 では，両方の変量の推移を観察することで，片方のみの観察では得られなかった特徴も見えるようになると期待される．

2 について，事象に付随するテキストラベルの文字列を利用して色相の割り当てを行った．これによって文面が似たものに近い色相が割り当てられるようになり，事象を色相で見分けたり，似たもの同士を識別することができるようになった．

2. 関連研究

2.1 量的データ推移の可視化

時刻付き量的データの可視化で最も一般的なものは線グラフであり，量をより的確に捉えられるように改良したものが積み上げ面グラフ (Stacked Area Chart) である．積み上げ面グラフを利用した時刻付き量的データの可視化は Wattenberg の NameVoyager[1] に代表され，これはインタラクティブ操作でユーザの見たい情報を絞込みながら積み上げ面グラフで可視化を行なっているものである．

積み上げ面グラフを発展させた Harve らの ThemeRiver[2] もまた，よく用いられる手法である．これは「川 (River)」のモチーフを用いることで，時系列に沿った量変化を直観的に見られるようにしたものである．積み上げた図形を上下対称に配置することで，推移の様子が形の変化として把握しやすくなっている．

このような表現において，本研究で扱う時刻付き量的順位データに対して，順位を表現するには層を積み上げる順番を利用することが考えられる．しかし，積み上げ面グラフや ThemeRiver のような各事象を層として重ねて表示するものは，時刻に沿って局所的に層を入れ替えていくことができない．つまり，順位を表現するには層全体も入れ替えなくてはならず，順位の変化を提示することが困難である．

2.2 順位推移の可視化

実世界に存在する順位データの多くは時刻付きデータであり，我々の生活の中にもこうした時刻付き順位データ，およびその可視化は広く浸透している．例として，Yahoo!Japan の検索ランキング [3] や The Economist の 2009 年 9 月 11 日付の記事 [4]，The Telegraph の 2011 年 8 月 13 日付の記事 [5] など，一般向けのウェブサイトでも，さまざまな形で順位の推移が表現されている．同様に，駅伝の現在までの順位表示などに利用されているのが，折れ線グラフを拡張した Brinton の Rank Charts[6] である．

順位と量を同時に表現したものとして，ThemeRiver を拡張した Shi らの RankExplorer[7] が存在する．これは事象となる検索キーワードを順位でカテゴリ分けしており，順位の推移とともに各カテゴリへの流出入などを棒グラフを埋め込んで表したものである．RankExplorer は順位でカテゴリ分けすることで，層の入れ替えを発生させないようにしているが，個々の事象の詳細な順位推移を観察することは難しい．

2.3 時刻付きテキストデータの可視化

折れ線グラフに準拠する Sparkline をタグクラウドに埋め込んで，単語のトレンドを表現するのが Lee らの SparkClouds[8] である．これはどの事象が関連性を持っているかなどの比較をすることは難しい．

同様にタグクラウドを利用したものとしては Collins らの Parallel Tag Cloud[9] が挙げられる．これはカテゴリに分けられたタグクラウドが並行軸状に配置されており，このカテゴリ分けを時間で行えば，時系列テキストデータの表現も可能になる．この表現は単語が順位順ではなくアルファベット順に並べられている．また，実際の文字列を画面上に表示しており，量によって文字の大きさが変化しているため，領域の確保や重なりといった問題に対応しなければならない．

3. 時刻付き量的順位データの概要

本研究で扱うデータは「事象が量的値により順位付けされた時刻付き順位データ」である．ここでいう量的値とは出現回数等である．また，それぞれの事象を区別するために，各々が一意なテキストラベルを持つものとする．具体的に言うと，商品の販売データであれば，事象はある商品が購入されたこと，量的値は購入された数や売上金額となり，事象を区別するためのテキストラベルは商品名から付けられる．また，政党支持率であれば，事象は政党に対する支持，量的値は実際の支持率であり，テキストラベルは政党名から付けられる．

このような量的値を元に順位付けを行い，データを採取した時刻を付加したものを実際のデータとする．量的値の多い順に順位を定めるのが最も一般的だと考えられるが，他の定義の仕方も存在する．なお，ここでは，ある時刻において同順位のものは存在しないとする．

データは単位時間ごとに計測される．例えば単位時間を 1 日，あるいは 1 時間と設定した場合，1 日おき，あるいは 1 時間おきにデータが計測される．

4. 視覚的表現の概要

視覚的表現の概要を**図 1** に示す．以下，これについて詳しく説明していく．

4.1 順位と量の推移の表現

1 つの事象は 1 つの矩形として描かれ，その時刻に発生した事象が垂直方向に並べられている．事象の持つ順位が

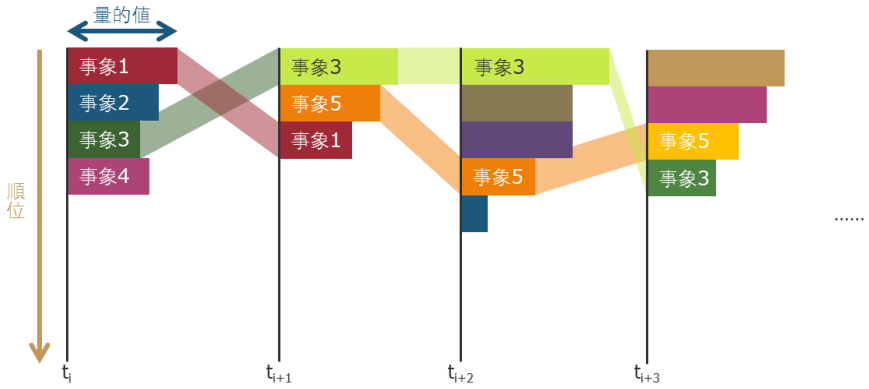

図 1 表現の概要
Fig. 1 Overview of visualization.

高いほど，矩形は上の方に配置される．次の時刻で計測されたデータにおいても同じ事象が観測されていた場合，その矩形同士を太い線（リボン）でつなぐ．反対に，連続していなければリボンは描画されない．これによって，事象の連続性や断続性が把握しやすくなる．

4.1.1 順位の表現

順位を表現する要素として縦方向の座標を用いる．各軸において上にある矩形ほど，その時刻において高い順位を持つことを表し，上から順に 1 位，2 位…と並ぶ．順位が高いものを物理的に高い位置に置くことで，より直観的な表現を提示する．

4.1.2 量的値の表現

矩形の横幅が量的値に対応している．

例えば Wikipedia の編集履歴を可視化した HistoryFlow[10] では，縦幅で量的値を表現しており，これによって全体の量の推移も一目で分かるようになっている．一方，本研究のように縦方向の並びで順位を同時に表すことを考えた場合，同じく縦方向で事象の量を表すと，横に並ぶもの同士では順位が必ずしも一致しない．これは縦幅を固定していないために生じる現象であるが，本研究ではこの事態を避けるために，矩形の横幅で量的値を表すことにした．

ただし，量変化の方をより重点的に観察したい場合もあるということを考慮し，矩形の縦幅が量を表現するように表示を切り替えることができるようにもする．これによって各事象の量だけでなく，その時刻における事象の総量が連結された矩形の縦幅で表現されることになる．

4.1.3 時刻情報（時間軸）の表現

時刻を表現するのは横方向の座標であり，画面の左から右へと古い順に並べる．このように，ある時刻における表現を並列させることで，時間の推移に伴う事象の変動を観察することができる．

4.2 事象の区別

事象をそれぞれ区別して表示することについては，矩形を塗りつぶす色相を利用する．矩形を塗りつぶす色相については，その矩形が表す事象に予め付けられた一意なテキストラベルによって決定される．また，この色は，次の時刻において順位が上昇していれば彩度と明度を上昇させ，より目立つようにしている．

事象に付けられたテキストラベルから色相を算出する方法として，以下のような計算を行う．

(1) テキストラベルとなっている文字列の各文字（文字コード）を数値に変換する．
長さ n の文字列 s を $s = \langle c_1, c_2, \ldots, c_n \rangle$ と定義する．また，文字 c の文字コードを数値に変換したものを $D(c)$ とおく．

(2) それらの平均値を求める．
すなわち，文字列 s の数値は以下によって表される．

$$D(s) = \frac{1}{n} \sum_{i=1}^{n} D(c_i) \qquad (1)$$

(3) 求めた値を色相の最大値で割り，余りを色相値とする．
色相の最大値を h とおけば，求める色相値 $H(s)$ は

$$H(s) = D(s) \bmod h \qquad (2)$$

この手法は Wattenberg らの Chromograms[11] を参考にしている．彼らは単語の頭 3 文字を利用して色相，彩度，明度の 3 値を定義しているが，本研究では彩度および明度は順位の上昇を表すことに利用しているため，上記のようになるべく色相のみを用いて，文字列を一意に表現できるようにした．

4.2.1 数値の平均化

提案する手法による色相割り当てでは，テキストが示す数値に対して平均化を行なっている．これは，同じ事象に対する表記のバリエーションや，似た事象に対する似た言葉を含む文字列について，差異を減らしなるべく近い色相

値を算出するために行うものである．

具体的に説明すると，テキストラベルは事象を区別するために付けられているが，同じような事象に対してはしばしば同じようなテキストラベルが付けられる．例えば，商品の売上データを考えたとき，「ミルクチョコレート」や「ビターチョコレート」，「アーモンドチョコレート」などは，テキストである商品名こそ違うが，すべて「チョコレート」というカテゴリに属する商品であると言える．これらを表現する矩形に対して似たような色相を割り当てるために，平均化を行うものである．

また，カテゴリごとに事象が把握できることで，カテゴリ全体の流行の推移も見えるようになる．先の商品の売上データで言えば，2 月 14 日のバレンタインデーにあわせ，その付近において「チョコレート」というカテゴリの商品の売上が伸びていることが確認できると期待される．

4.2.2 剰余を利用したハッシュ法

剰余を利用することで，文字列の数値がどのような最大値をとるか分からない場合においても，色相の最大値以下の数値を割り当てることができる．これは剰余を用いたハッシュ法である．

5. ツールの開発

5.1 ツールの概要

時系列量的順位データを 4 章で述べた手法により可視化し，それを利用してトレンド分析を支援するツールを開発した．視覚的表現からトレンドの全体的な流れを把握し，そこからさらに各事象について詳細な分析が行えるように，インタラクティブ操作を利用しながらトレンド分析を行えるように設計を行った．

本ツールは Processing[*1]によって開発を行い，GUI のライブラリとして ControlP5[*2]を利用している．

5.2 基本操作

ツールの画面は大きく 3 つに分けられる（**図 2**）．

5.2.1 視覚的表現

画面の中央には，第 4 章で設計した視覚的表現を配置する（図 2 の 1）．この図では，31 の時刻において，それぞれ 100 の事象を表している．

ひとつの事象を表す矩形にマウスカーソルを乗せると，事象の量に応じた大きさの円および事象のテキストラベルが表示される．

5.2.2 テキストラベルの一覧表示

キーボードの矢印キーの左右キーを操作してある時刻を選択すると，その時刻のデータ中にある事象の順位とテキストラベルの一部が画面右部に一覧表示される．なお，文字の大きさは事象の量に対応しており，文字列の左端には事象に割り当てられた色相の円と順位を表示している．

5.2.3 事象の詳細表示

画面中央において，矩形をクリックすると，その矩形が表現する事象の詳細が画面下部に一覧表示される．画面に描画しきれていない分については，右側のスクロールバーを利用して見ることができる．

5.3 表現の変更

各事象が持つ属性の値に応じて，矩形に対し強調したいものを変更することができる．本ツールにおいて，強調することに利用する視覚的要素として矩形の幅と明度・彩度を採用する．ここで，特に明度と彩度で強調する場合を「ハイライト表示する」と呼ぶ．また同様に，上下の配置を決定する属性を変更することもできる．

5.3.1 矩形のハイライト

事象の持つある属性について，どのような値や特徴を持つものをハイライト表示するかをひとつ選択できる．これによって，特定の事象を強調することができ，観察したい動向を示す事象をより観察しやすくなる．

順位の上昇と下降　デフォルトでは順位が上昇している事象をハイライトするように設定した．トレンドを探索するにあたり，もっとも参考になるのは順位の上昇であると考えたためである．

反対に，順位が下降しているものをハイライトすることも可能である．この強調によって，流行が終息しつつあるような事象を確認することができる．

事象の連続性と断続性　順位の上昇・下降に関係なく，連続して出現しているものをハイライトすることができる．同様に，他の時刻では全く出現しなかったような，単発的な事象もハイライト表示ができる．この表示によって，長期にわたり人々に支持された事象や，ある時刻においてのみ人々が興味を持った事象などが分かる．

事象の属性の大小　事象が順位と量的値以外に何らかの属性を持っており，それが二値であったり，あるいは量的なものであった場合，値にしたがってハイライト表示する．これにより，その事象の持つ順位と量的値以外の属性に着目した分析を行うことができる．

5.3.2 矩形の整列順序

各事象を表す矩形の上下の配置を決定する．整列はすべてにおいて昇順か・降順かを選択できる．

順位順　デフォルトでは，事象の順位が高いものほど，それに対応する矩形が画面上部に配置される．本研究で用いるデータは様々な属性がついているが基本的には「順位データ」であり，これを観察するためにデフォルトではこの設定にしている．

量的値順　事象の持つ量的値にしたがって上下の配置を決定できる．事象の順位は量的値のみで決定しない場合

*1 http://processing.org/
*2 http://www.sojamo.de/libraries/controlP5/

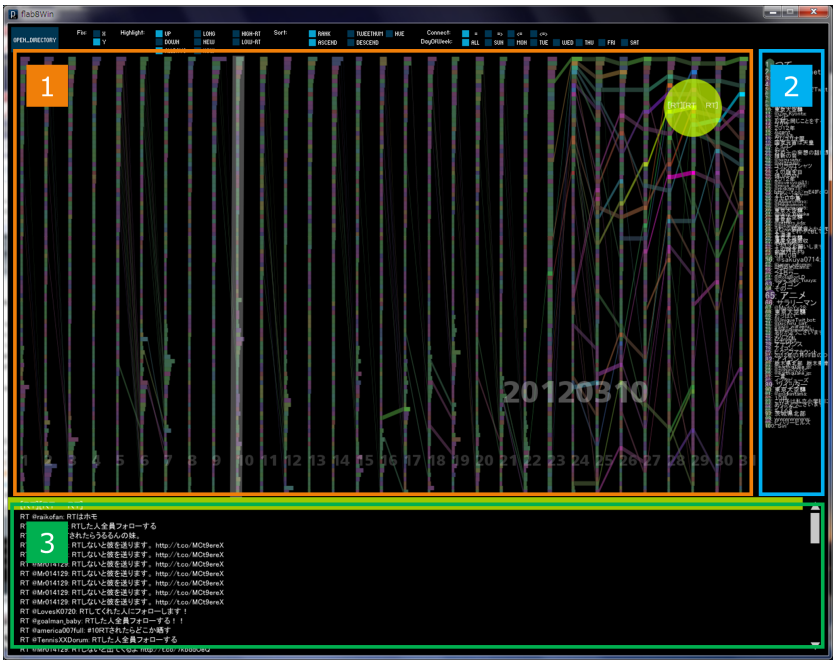

図 2 ツールの概要
Fig. 2 Overview of tool.

もあるため，量的値に着目したいときに用いると，値の多いもの，あるいは小さいものが把握しやすくなる．

カテゴリ順 事象に割り当てられた色相にしたがって上下の配置を行う．文字列表現として近いものに似た色相が割り当てられるため，同じカテゴリに属する事象が近くに配置される．しかし，提案手法の性質上，近くに配置されているからと言って，同じカテゴリに属する事象であるとは限らない．

この整列方法により，対象のデータにおいて事象がどのようなカテゴリに分けられるかや，カテゴリごとの事象の数などを知ることができる．

6. ユースケース

6.1 データの概要

本ユースケースで用いるデータは短文投稿サービス twitter[*3]に投稿された文書（ツイート）を整形したトピックデータである[*4]．データの具体例を表 1 に示す．

ある 1 日のツイートに含まれる共通の単語群を「トピック」と定義する．そのトピックの順位，トピックを構成する単語群，関係する実際のツイート数，および実際のツイートが 1 つのレコードとなる．これが 1 日ごとにまとめられており，2011 年 7 月から 2012 年 6 月まで 1 年分のデータが用意されている．このデータにおいて，量的値はツイート数，テキストラベルはトピックとなる単語群である．なお，「実際のツイート」に含まれる属性は twitter のユーザ ID，ツイート ID，本文，時刻情報，GPS 情報であるが，今回は本文のみを画面上に表すことにした．

データの属性として，順位・量的値・時刻の他に「リツイート (RT)[*5]率」を加える．これは，あるトピックに関する実際のツイートのうち，RT であるものがどれだけの割合を占めているかということを計算したものである．これを元にして，RT 率が高いもの・低いものをハイライトできるようにした．

なお，テキストラベルを用いた色相算出については，トピックとなっている単語群を頭文字の文字コード順にソートし，その 1 単語目となった単語を利用した．これは，各トピックについて，同じ事柄を話題にしているものは，なるべく同じ単語を用いて色相の算出が行えるように配慮したものである．

トピックの判別及び順位の算出方法は以下の通りである．

6.1.1 トピックの判別

処理の対象となるツイートの集合（1 日分など）において，指定回数以上出現したキーワードを抽出してトピックとする．このキーワードとは 2 つ以上の単語からなる組み合わせであり，単語とは形態素解析によって名詞と判断されたもの，オンライン百科事典 Wikipedia[*6]に登録されている 3 文字以上のエントリー，URL，ハッシュタグ[*7]である．

6.1.2 順位の算出

トピックに関するツイート数が多いものから順位を 1 位，2 位…と定めていく．

*3 http://twitter.com/
*4 株式会社富士通研究所から提供を受けた
*5 そのツイートが他人のツイートを参照したものであることを表す
*6 http://ja.wikipedia.org/wiki/
*7 twitter において，投稿したツイートに対して投稿者が付けるタグ

表 1　twitter トピックデータの例
Table 1　Example of topic data.

時刻	順位	トピック	ツイート数	実ツイート（本文のみ）
2012/1/1	3	あけおめ, メール	586	#あけおめメールが来ない
				あけおめメールきたっ
				あけおめメールがこない
				...

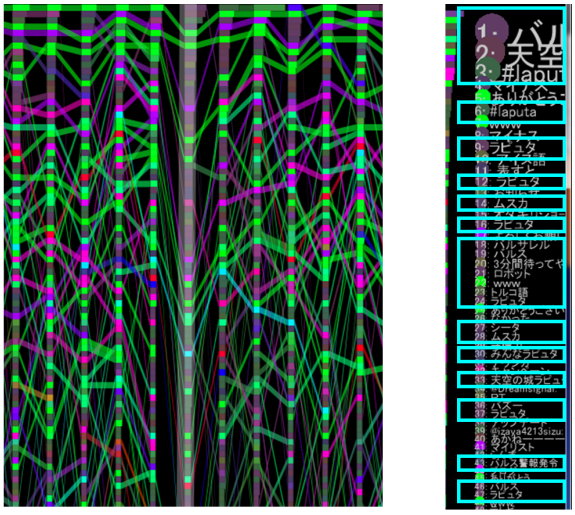
(a) 2011年12月9日のデータに見受けられる谷間　(b) 該当日を選択した時の表示

図 3　『天空の城ラピュタ』に関するトピックの発見と観察
Fig. 3　Discovery and Observation of a topic about *Castle in the Sky*.

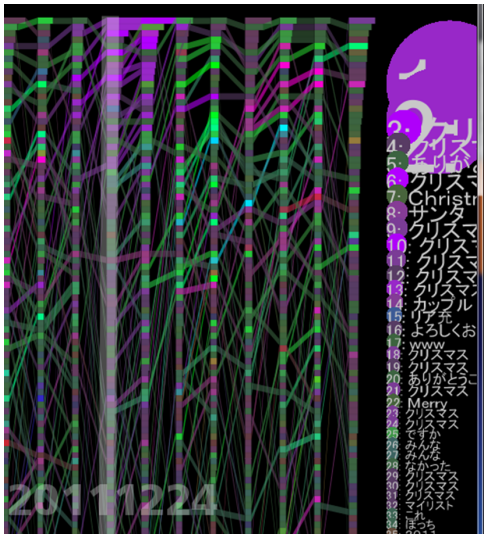

図 4　「クリスマス」を含むトピック群（紫色のもの）
Fig. 4　Topics including “クリスマス (Christmas)” which filled with purple.

6.2　観察

6.2.1　流行の把握

データを観察すると，上位には「フォローありがとう*8」というような挨拶に関わる単語がほとんどを占める．これはどのような時期においても見られた．しかし，そうした定型句以外の単語で構成されるトピックが，定型句を含むトピックよりも上位に，かつ多くのツイート数を伴って現れる場合がいくつか観察された．

図 3(a) は 2011 年 12 月のデータを表現したものの一部である．この図の中央を見ると，上記のような現象が発生し，「谷間」のような表現になっていることが見受けられる．これについて詳しく観察してみると，この日（2011 年 12 月 9 日）はアニメーション映画『天空の城ラピュタ』が地上波で放映され，その実況をするようなツイートが多かったということが分かった．図 3(b) は該当する日を選択したときの画面右部のトピック一覧表示である．水色の枠で囲んでいるトピックが，この作品に関するものであると推測される（なお，水色の枠については筆者らが描画した）．

6.2.2　トピックのカテゴリの把握

2011 年 12 月 24 日から 25 日にかけて，上位に紫色の矩形が集中していることを発見した．これらのトピックは「クリスマス」という単語を含むものが多かった（**図 4**）．

6.2.3　データの全体像の把握

2011 年・2012 年にどのような出来事が起こったかという事実と見比べながら分析を行った．その結果，3 月 11 日の東日本大地震や 8 月 15 日の終戦記念日に関するツイートは筆者らの予想よりも少なく，また 2011 年 8 月の民主党代表選，同年 12 月の北朝鮮の金正日総書記死去といった政治的なニュースに関するツイートはさらに少ないと感じた．

反して，2011 年 10 月に Apple のスティーブ・ジョブズ元 CEO が死去したという IT 系のものや，バレンタイン・正月などの季節の行事，また 2011 年 11 月 11 日の江崎グリコが提唱する「ポッキー&プリッツの日」，2012 年 5 月の金環日食といったイベント的なものに関するツイートは非常に多かった．

実際にトピックの数を調べてみると，上位 100 トピックのうち，民主党代表選については 2 トピック・最大ツイート数 72 であったのに対し，ジョブズ死去については 53 トピック・最大ツイート数 429 であった．

6.3　考察

6.2.1 節で観察できた「谷間」は，詳しく説明すれば，矩形同士をつなぐリボンが途切れたり，あるいはその時刻を

*8　あるユーザがそのユーザとつながりを持とうと働きかけたことに対する感謝の意

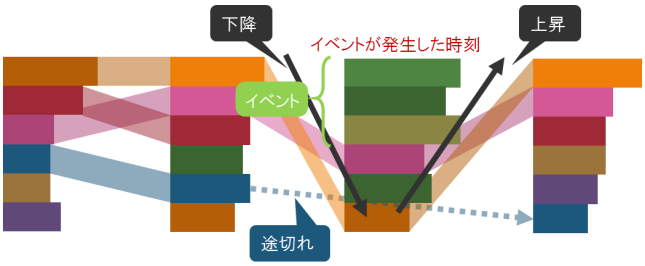

図 5 単発的なイベントの発生
Fig. 5 Occurrence of a discontinuous event.

挟んで全体的に下降・上昇しているものである（**図 5**）．これは，単発的なイベントが上位に挿入されたことに起因すると考えられ，その時刻における一時的な流行（ファッド）を表しているといえる．また，図 3 で発生した「谷間」を引き起こした現象について調査を行った．その結果，物語中に登場する「バルス」というセリフが，1 秒間における世界中のツイート投稿数 25,088 という当時の新記録を達成するなど，twitter 上でもこの作品のテレビ放映は大きな話題となっていた [12]．

図 4 に見られる，「クリスマス」という事柄に関する話題への類似色の割り当てについて，同じような話題に，つまり同じカテゴリに属するトピックに対して同じような色相が割り当てられていると言える．これは，提案する色相算出手法がカテゴリの把握に有用であることを示す．

また，6.2.3 節から，twitter 上で流行していた話題は，インターネット上のソーシャルメディアらしく，IT 系のサブカルチャー的なものが多いということが分かった．さらに，季節のイベント，あるいは皆で盛り上がれるようなイベントに対しても多くの関心が寄せられていた．これは，twitter というメディアを通じて，ユーザ同士がつながりを持ち，それを大いに楽しむ人が多いということだと言える．すなわち，このツールを用いて twitter のトピックデータを分析することで，twitter というメディアにおけるユーザの反応の傾向を得ることができた．

6.4 ツールの有用性に関する議論

本研究で開発したツールを用いることで，twitter というソーシャルメディアでの流行について，それぞれの特徴および全体的な特徴を把握することができた．

6.4.1 事象の連続性について

図 5 のような現象は，ある時刻において，それまで流行していた事象の順位が下がったり，あるいは事象自体が出現しなかったりしたいうことを表す．このような表現の「谷間」を探すことは，瞬間的な流行の発見につながる．

個々のトピックがあまりにも長く続いているものは「慢性的な」ものでもあり，挨拶などの頻出するトピックは常に出現し続けるからと言って流行しているとは言えない．よって，「話題が長く続いているものをハイライト表示する」という機能は，流行の事象を検出するというよりも，そのデータにおけるスタンダードなものを検出するという意味で，データの全体像を把握できる機能だと言える．

逆に，単発的なものをハイライト表示すると，その時刻で瞬間的に流行したものが見える．つまり，その事象が示されたその瞬間にしか興味が持たれないような，継続した話題性がないようなものである．そのような事象がどういった目的で人々に提示されたのかを併せて考えることで，人々の興味関心はどこにあるのかということを探ることができる．

6.4.2 データの傾向把握について

6.2.3 節から，開発したツールによって，トレンドを分析する一方で，twitter というメディアの全体的な傾向も把握できた．

メディアの傾向とは，例えば，掲載された記事に寄せられたコメントなどを利用し，同じようなデータを新聞社のサイトやニュースサイトからも抽出したとする．すると，おそらく今回のユースケースで得られたような視覚的表現とは違った図が得られ，twitter とこれらのメディアのユーザは性質が違うということが分かるだろう．

すなわち，ツールから提示された結果を分析することで，そのメディアの発信する情報のうちどのようなものがユーザの興味を引いているか，どのようなユーザがそのメディアを好んでいるかといった，メディアとそのユーザの特性を知ることができる．

6.4.3 テキストラベルからの色相自動算出について

色相の自動算出によるカテゴリ分けについては，6.2.2 節からある程度はうまくいったと考えられる．

このユースケースで用いた twitter のトピックデータにおいて，カテゴリとは「そのトピックが共通して話題にしている事柄」である．例として図 4 に見られる「クリスマス」について考える．「クリスマスプレゼント」や「メリークリスマス」は文字列としては完全には一致しないが，すべて「クリスマス」という共通の事柄（カテゴリ）に関するトピックである．これらは今回提案した色相の割り当てによって似たような色相で表されており，この手法は有益であると言える．

このように，提案手法においては，平均値から算出することで，多少はトピック内の単語にばらつきがあったとしても，それを補うことができた．しかし，現状は文字列を表現としてしか扱っておらず，その意味を全く考慮していない．したがって，今回のようなテキストデータに対して用いるにはさらなる改良が必要だと考える．上記の例でいえば，「サンタクロース」や「Christmas」は明らかに「クリスマス」というカテゴリに属するトピックのキーワードであるが，表現としての文字列が「クリスマス」とは全く異なるため，同じような色相を割り当てられない．このような場合，辞書データの参照によるカテゴリ分けなどが改

善案として考えられる．

6.4.4 日本語以外のテキストに対する色相自動算出

今回用いたデータは日本語で記述されたデータ，つまり2バイト文字を多く使用するデータである．欧文のような文字数が少ないデータを用いたとき，今回提唱した色相値の算出方法は適しているとは言いがたいと考える．さらに言えば，日本語であっても，ひらがな・カタカナ・数字・アルファベットに比べ，漢字の数は文字通り桁違いに多い．よって，カタカナやひらがなのみで構成された単語に対しては，どうしても似たり寄ったりな色相が割り当てられてしまう．

これを解決するために，データに利用されている言語別に色相算出方法を変えたり，意味を考慮して色相を割り当てたりするなどを行うとよいと思われる．

7. おわりに

7.1 まとめ

時系列量的順位データから，順位と量の両方の時間変化を可視化し，それを利用してトレンド分析を支援するツールを作成した．これは，それらの推移を俯瞰でき，かつ，インタラクティブ操作を用いることでより詳細な情報を得られるものである．本ツールを用い，目的に応じて表現を変更することで，詳細なトレンド分析を行えるようにした．さらに，事象の区別に色相を利用することを考え，その色相値を自動で算出する手法を提示し，この手法の有用性を述べた．

ユースケースとしては，twitter のトピックデータを利用したものを示した．このユースケースから，そのデータにおけるトレンドについてだけでなく，データの持つ全体的な流行の特性も観察することができたという結果が得られた．加えて，例えば新聞や各ニュースサイトなどの他のメディアから同様にデータを抽出し分析することで，各メディアとユーザの特性も把握できるという可能性を示した．

7.2 今後の課題

重大な課題として事象のカテゴリ分けが挙げられる．今回はテキストラベルをもとに色相を算出し，それが近いものをカテゴリ的に近いとみなすようにしているが，この手法では字面が同じものが同様のカテゴリだとみなされ，意味的なことは全く反映されていない．そこで，事象のカテゴリ分けをルールにしたがって正確に行い，それを元に色分けを行うことで，流行の全体像が一層明確に分かるようになる．加えて，日本語以外の言語を含むデータを用いた分析をする場合においては改善をすべきである．

表現上の課題としては，順位が極端に離れているもの同士をつなぐリボンが非常に細くなり，どことどこが対応しているかが見づらいというものがある．リボンの傾き具合がランクの上下の幅をより明確にしている場合もあるため，この特性を失わないよう改良策を考えるなくてはならない．また，同順位を含むデータにおける表現についても，検討する必要がある．

謝辞 本研究の一部は，株式会社富士通研究所からの受託研究によるものである．

参考文献

[1] Wattenberg, M.: Baby Names, Visualization, and Social Data Analysis, *IEEE Symposium on Information Visualization (InfoVis 05)*, pp. 1–7 (2005).
[2] Havre, S., Hetzler, E., Whitney, P. and Nowell, L.: ThemeRiver: Visualizing Thematic Changes in Large Document Collections, *Visualization and Computer Graphics*, Vol. 8, No. 1, pp. 9–20 (2002).
[3] Yahoo! Japan: Yahoo!検索ランキング，Yahoo Japan Corporation（オンライン），入手先 ⟨http://searchranking.yahoo.co.jp/⟩（参照 2013-1-21）．
[4] The Economist online: Cutting red tape, The Economist (online), available from ⟨http://www.economist.com/⟩ (accessed 2013-1-11).
[5] The Telegraph: The top 100 baby names in England and Wales in 2011, Telegraph Media Group (online), available from ⟨http://www.telegraph.co.uk/⟩ (accessed 2013-1-11).
[6] Brinton, W. C.: *Graphic methods for presenting facts*, The Engineering magazine company (1914).
[7] Shi, C., Cui, W., Liu, S., Xu, P., Chen, W. and Qu, H.: RankExplorer: Visualization of Ranking Changes in Large Time Series Data, *Visualization and Computer Graphics*, Vol. 18, No. 12, pp. 2669–2678 (2012).
[8] Lee, B., Riche, N. H., Karlson, A. K. and Carpendale, S.: SparkClouds: Visualizing Trends in Tag Clouds, *Visualization and Computer Graphics*, Vol. 16, No. 6, pp. 1182–1189 (2010).
[9] Collins, C., Viégas, F. B. and Wattenberg, M.: Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora, *Visual Analytics Science and Technology*, pp. 91–98 (2009).
[10] Viégas, F. B., Wattenberg, M. and Dave, K.: Studying Cooperation and Conflict between Authors with history flow Visualizations, *CHI '04 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 575–582 (2004).
[11] Wattenberg, M., Viégas, F. B. and Hollenbach, K.: Visualizing Activity on Wikipedia with Chromograms, *Lecture Notes in Computer Science*, Vol. 4663, pp. 272–287 (2007).
[12] 藤井 涼：CNET Japan:「天空の城ラピュタ」が世界新記録–1 秒間のツイート数，CNET Japan（オンライン），入手先 ⟨http://japan.cnet.com/⟩（参照 2013-1-11）．