

まじめなデータから得られる「おもしろい」情報可視化

伊藤貴之¹

概要: 情報可視化は日常生活の一般的な情報やデータを視覚的に表現する技術であり、多くの場合において自然科学や社会科学などの専門知識や日常業務に関する「まじめな」データを分析あるいは俯瞰するための技術である。しかしその情報可視化結果の中から、しばしば笑える結果を得ることがある。それはまるで、本来なら真剣に勝負しているはずのプロ野球の「珍プレイ」から1つのバラエティ番組が成立するのに似ているといえるだろう。本報告では、そのようなまじめなデータ分析から、どのように「おもしろい」が得られるか、について議論する。

Funny Information Visualization from Serious Datasets

TAKAYUKI ITOH¹

Abstract: Information visualization is a visual approach to represent our general and daily information, which is mainly applied to "serious" datasets in the fields of natural and social sciences. However, we sometimes get interesting or funny something from the visualization results of such serious datasets. It may correspond to the funny TV programs collecting unexpected performance of professional baseball games. This paper discusses how we may get such funny results from the visual analytic processes of serious datasets.

1. はじめに

今年の本会議における芸術科学会オーガナイズドセッションのテーマは「おもしろい【の】研究」であり、面白さの定義、そして面白さを導く諸手法について論じるものである。「面白い」の定義や分類には多角的な考え方があるであろうが、本報告において著者は、

- 最初から狙って面白く制作したもの
- 狙ってないのに結果的に面白くなったもの

の2種類に「面白い」を大別して議論を進める。

著者が日常的に取り組んでいる研究テーマのひとつに「情報可視化」がある。情報可視化は日常生活の一般的な情報やデータを視覚的に表現する技術であり、多くの場合において自然科学や社会科学などの専門知識や日常業務上の「まじめな」データを分析するための手法である。ほとんどの場合において著者は、エンターテインメント的な「面白い結果」を得ることを最初から目的として情報可視化に取り組んでいるわけではない。あくまでもシリアスに、自然科学

学や社会科学の問題解明の一手段を開発することを目的としている。しかし著者の講演の機会にて情報可視化の事例を紹介すると、多くの講演においてそこには笑いが絶えない。そこには「最初から面白いものを狙って制作した」正統的なエンターテインメントではなく、「面白さを追求して制作したのではないのに、結果として面白くなってしまった」副産物のようなエンターテインメントが存在している。言い換えれば、私達のなんでもない日常生活や日常業務の一コマが、可視化されただけでエンターテインメントになる、という事例はいくらでも存在し得るとも考えられる。

そのような副産物としてのエンターテインメントは、すでにテレビ番組等のメディアにも存在する。典型的な例が「プロ野球珍プレイ集」の類のテレビ番組である。当然ながらプロ野球は真剣勝負の場であるが、その場面の中には多様なエンターテインメントが潜んでいる。そのような意外なエンターテインメントは、能力にあふれたプロ野球チームのような特殊な集団の中からでなくても引き出せる可能性が大いにある、と考えられる。

日本語では多様な意味をもつ「面白い」も、英語に訳す

¹ お茶の水女子大学大学院人間文化創成科学研究科

ると enjoyable, funny, interesting といった多様な単語に枝分かれする。本報告では情報可視化から得られる「面白い」をそのような角度から分類することで、私達の日常生活や日常業務からどのようにエンタテインメントを引き出せるかについて議論する。

2. 事例

本章では著者自身がしばしば講演で紹介する情報可視化の事例の中から、笑いが出ることの多い事例をいくつか抜粋する。

Facebook に見られる「別れ」の件数

図1は、Facebookにおいて各ユーザが「交際中」のフラグをオフにした件数、つまり男女の「別れ」の件数を1年間にわたって集計した結果であり、有名な学術講演会 TED の "The beauty of data visualization" [1] で発表されたものである。この可視化結果は「別れを集計する」という悪趣味な行為ゆえに画像を出す以前から笑いを引き起こすことがあるが、それにも増して可視化結果が非常に多くの現象を語っている。その中でも著者がこの可視化結果を紹介した際に特に聴衆から笑いが出るのは、「エイプリルフールに別れが多い」「4月の月曜に別れが多い」「クリスマスの2週間前に別れが多い」の3つの事象を議論するときである。

ある Twitter ユーザの時間別ツイート回数

図2は、あるユーザの Twitter におけるツイート回数を1時間ごとに集計したものである。このユーザは比較的規則正しい生活を送っているため、ツイート回数を見ただけで起床時刻、出勤時刻、会議などの多い時刻、朝食や夕食の時刻、就寝時刻、などが鮮明に可視化される。つまりこの可視化結果は、単に Twitter のツイート回数を可視化しているだけでなく、そのユーザの生活リズムそのものを可視化しているともいえる。この結果を講演等で示すと、そのユーザの生活リズムを知っている聴衆の中から内輪受けのような笑いがしばしば聞こえる。

カフェテリア出口付近の人の流れ

図3は、あるカフェテリアの出口付近の人の流れを高所からの動画追跡技術によって記録し、近くを通っている類似軌跡を集計した可視化結果 [2] である。このカフェテリアの出口の両側には、自動販売機とゴミ箱があるが、これらの前に立ち寄った人は非常に少なかったことが可視化されている。この結果の講演中、「せっかく自動販売機とゴミ箱を設置してくれたのに残念なことに...」という説明の時点で笑いが見られることがある。

ウェブサイトへのアクセスパターン

図4は、著者のウェブサイト構成するウェブページのリンク構造と、そのウェブページ群にアクセスする閲覧者のアクセスパターンを可視化した結果 [3] である。この可視化結果からは著者のウェブサイトへのアクセスに関する多くの知見が得られる。例えば著者は2000年にアメリカの大学に半年滞在しているが、その時の様子をまとめた全17回のレポートを全ページ読んでいた閲覧者が多数いることが見受けられる。また著者はPDFファイルとして講義資料を学内公開しているが、試験のある科目では試験前に全てのPDFファイルが再度アクセスされるのに、試験のない科目ではそのようなパターンが全く見られないことがわかる。また著者は研究室のプロジェクトを日本語と英語で公開しているが、日本語の研究紹介を全ページ読んでいた閲覧者が多数いるのに対して、英語の研究紹介を全ページ読んでいた閲覧者は少ないことがわかる。ここでも、「試験のない科目の講義資料へのアクセスが少ない」「英語の研究紹介を全ページ読んでいた閲覧者は少ない」という残念な結果への笑いがよく見られる。

化粧顔画像の印象評価

図5は、各種の化粧を施した顔画像を多数の被験者に提示し、その印象の良さを回答してもらった結果を集計して可視化したもの [4] である。この可視化結果にて最も目立つ点は、当該実験の被験者中の男性には「アイラインを施さない化粧顔画像」に関する印象評価が際立って高い傾向にある、という点である。アイラインの好きな女性からみたら、この結果はひょっとしたら残念な結果かもしれない。この可視化結果の講演においても、このような残念な結果に対して自然に笑いが発生することが多い。

クレジットカード不正決済の手口

図6は、クレジットカードによる決済ログの中から、不正決済であったものを抽出し、その不正種別と決済時間帯との相関性を3次元棒グラフとして可視化した結果 [5] である。この可視化結果から明らかになったことは、盗難されたクレジットカードは深夜から早朝にかけて多く使われ、偽造されたクレジットカードは正午から夕方くらいに多く使われている点である。この結果だけからは一見「笑える」とは思えないかもしれないが、そこには意外な背景が潜んでいる。日本においてクレジットカード盗難被害の多くは夜間に発生している。例えば繁華街で泥酔する人、電車の中で熟睡する人、などが狙われやすいらしい。その盗難被害の多くは翌朝に通報されるため、盗難されたクレジットカードは夜中から早朝に使われることが多いらしい。まったくもって笑いごとではないエピソードではあるが、聴衆はこのような意外な背景を聞いた時に、しばしば笑いでその反応を示すようである。

視覚的強調表現がもたらす「わざとらしさ」

ここまで、集計結果や実験結果を客観的に可視化した結果から笑いが起こった事例をあげてきた。一方で情報可視化やインフォグラフィックスなどの視覚表現においては、主観的に（あるいは恣意的に）情報の一部を不平等に強調表現することも可能である。あるいは情報可視化の実装が不十分である場合に、不適切な知見をもたらすような不平等な可視化結果をもたらすこともある。

例えばスマートフォンに関するウェブ上の記事 [6] を見てほしい。図中にて赤のドットは iPhone ユーザの存在を、緑のドットは Android ユーザの存在を示したものである。一番上の図を見ると圧倒的に iPhone ユーザ優勢であるように見えるが、これは単純に緑を赤で上描きしてしまった結果である可能性が高い。実際にこのウェブページ中の、赤だけ、緑だけを表示した結果をみると、どちらが優勢とは言い切れないことがわかる。このような可視化結果は現実以上に不当に iPhone が優勢であるかのような印象を与える可能性がある。しかし注意すべきことに、我々の日常はこのようにミスリーディングな図であふれている。

一方で、そのような強調表現はしばしば、情報発表者の意図を「見え見え」にすることになり、そのわざとらしさが笑いを引き起こすこともある。本稿中での紹介は控えるが、講演にていくつかの例を示したい。なお、わざと情報の一部を強調表現するウェブサービス [7] があるので参照されたい。

3. 面白さの分類

前章では情報可視化の事例講演から「笑いが出ることのある事例」をいくつか抜粋した。仮に「笑いが出ることのある事例＝面白い事例」とすると、そこには面白さの分類が可能であると考えられる。本章では情報可視化の結果に対する面白さを著者の主観によっていくつかに分類する。

3.1 狙った面白さ

情報可視化によって視覚的にもたらすデータ分析結果において、その分析に着手する前から「これから得られる可視化結果は笑える可視化結果であるに違いない」と確信して分析過程に入ることは可能である。

客観的な面白さ：これから処理するデータにどのような傾向があり、それを眺めることでどのような知見が得られるかについて、仮説をたてられるのであれば、最初からその仮説を客観的に検証する方向でデータを収集して可視化することができる。そして、その仮説に意外な結論や残念な結論が混じっているのであれば、それを客観的に示すことで計画的に笑いを引き起こすことは可能であると考えられる。特に近年の情報可視化技術には Visual Analytics というフレームワークが多用されており、仮説を検証するための反復作業が新しい知見を得るために重要であるとされて

いる。言い換えれば、あらかじめ予想されている笑いを仮説として計画的に可視化を遂行する、という行為が近年になって容易になっているとも考えられる。

主観的な面白さ：前章でも紹介した通り、情報発表者がわざとらしさをもって発表内容の意図や恣意を示すとき、そこには笑いが引き起こされることが期待される。つまり情報を描画する際の色選び、形選び、位置選びといった主観的な要因によって、意図的に笑いを引き起こすことが可能である。

3.2 狙っていない面白さ

本報告で紹介した「講演中に笑いが出る情報可視化」はいずれも、可視化に着手する以前から笑いや面白さを狙ったものではない。それは喩えるなら、プロ野球選手が最初から「珍プレイ集」に紹介されることを目標として野球に臨んでいるわけではないのに似ている。

ところで「プロ野球珍プレイ集」に登場するプレイはどのように分類可能であろうか。代表的なところでは、「ファインプレイや場外ホームランなどの絶賛プレイ」「エラーや暴力などの残念なプレイ」「ルールやマナーの盲点をついた意外なプレイ」「ファンに定着していて誰もが待っている定番プレイ」などがあげられるであろう。情報可視化の結果から得られる笑いや面白さも、これに似た考え方に基づき、図 7 に提案するような**称賛⇔残念**と**意外⇔当然**の 2 軸に沿った分類が可能であると著者は考える。

称賛に値する面白さ：例えば Twitter で単位時間あたりに出現した単語数を集計して棒グラフで可視化したら、きっと「バルス」という単語がとんでもない集計結果で突き抜けた棒グラフを形成するであろう。そのツイート数においても稼働し続けた Twitter のシステムは称賛に値するものであり、またその極端な可視化結果はある種の人には笑いを誘うものであろう。著者自身の可視化の事例ではこのようなものは現時点で特に見当たらないが、十分に作成可能であろうと思われる。

残念な結果に対する面白さ：本報告中で紹介した「カフェテリアの出口両側の自動販売機に立ち寄る人の少なさ」「試験のない科目の講義資料へのアクセスの少なさ」といった事例は、いずれも施設や資料の持ち主には残念な結果であろう。一方で、その残念さがプロ野球のエラーに似た笑いや面白さを誘う事例は多く存在するものであろう。

意外な知見に驚く面白さ：本報告中で紹介した「4月の月曜日や、クリスマスの2週間前に男女の別れが多い」という知見は、少なくとも著者自身は可視化結果を見るまで全く想像もしていなかった意外な知見であった。また「盗難されたクレジットカードは深夜および早朝によく使われる」という知見も同様に、背景を知るまでは理解できない意外な知見であった。かつて「へえ」という単語で意外な知識

を共有するバラエティ番組が一世を風靡していたが、このような「へえ」という声をあげたくなる知見の中には多くの笑いと面白さが存在すると考えられる。

当然な結果に同意する面白さ：意外な知見と全く逆に、当然の結果に「わかるわかる」と同意するような知見にも、笑いや面白さを引き起こす要因があると考えられる。本報告中で紹介した「あるユーザの1日のツイート数」の可視化結果が、そのユーザの生活リズムを知っている人に内輪受けする、というのがその典型例であろうと考えられる。

4. まとめ

本報告では著者自身による情報可視化の研究成果や講演内容を事例にして、まじめなデータからどのようなエンタテインメント的な面白い結果が導き出せる可能性があるかを議論した。

本報告中の議論は、情報可視化に限らずデータサイエンス全般に適用可能なものであると考える。例えばデータマイニングの学術的黎明期には、しばしば「コンビニでビールとおむつを同時に買う男性客が金曜の夜に多い」という事例が紹介されてきたが、著者が講義でこれを紹介するといまだに履修生から笑いと驚きの顔が見られる。そこで今後の展開として、情報可視化以外のデータサイエンス分野の専門家とも「エンタテインメント的な面白い結果」について情報を共有する機会がもてればと考える。

参考文献

- [1] The beauty of data visualization, <http://www.youtube.com/watch?v=pLqjQ55tz-U>
- [2] 藪下, 伊藤, 経路情報の要約と可視化の一手法, 芸術科学会論文誌, Vol. 10, No. 3, pp. 167-178, 2011.
- [3] 川本, 伊藤, ウェブのアクセスパターンとリンク構造の同時可視化の一手法と適用事例, 芸術科学会論文誌, Vol. 10, No. 2, pp. 68-78, 2011.
- [4] 野村, 伊藤, 山口, 画像ブラウザ「CAT」を用いた化粧の印象効果分析結果の可視化, 可視化情報学会可視化情報シンポジウム, 2009.
- [5] 迫田, 長崎, 伊藤, 伊勢, 宮下, クレジットカード不正検出のルール作成支援のための情報可視化システム, 画像電子学会 Visual Computing / 情報処理学会グラフィクスとCAD 合同シンポジウム, 2009.
- [6] 世界で本当に普及しているスマホはどれか 「iPhone」は都市部「Android」は郊外でシェア拡大 <http://rocketnews24.com/2013/06/25/343747/>
- [7] 不思議な円グラフを描くウェブサービス <http://d.hatena.ne.jp/aike/20130718>



図 1 Facebook に見られる「別れ」の件数 (文献 [1] より)

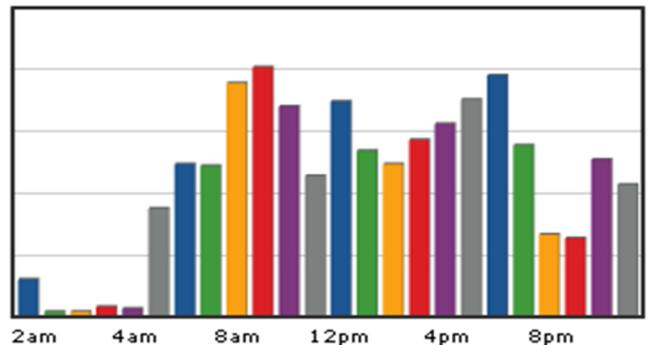


図 2 あるユーザの Twitter におけるツイート回数



図 3 カフェテリアの出口付近の人の流れ

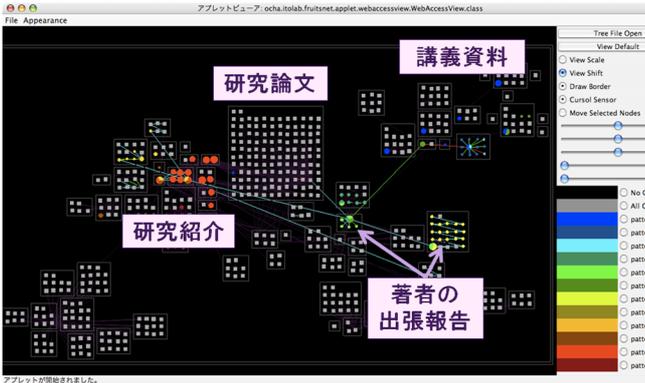


図 4 著者のウェブサイトにおける閲覧者のアクセスパターンの可視化



図 5 化粧顔画像の印象評価に関する集計結果の可視化

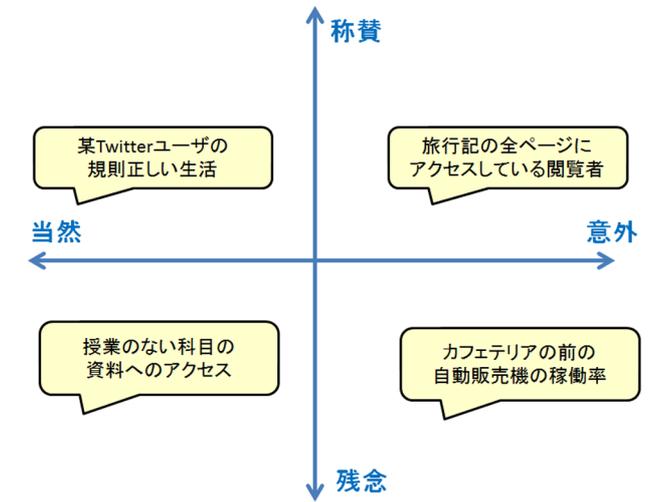


図 7 笑いを買ったことのある情報可視化を分類するための2軸の例

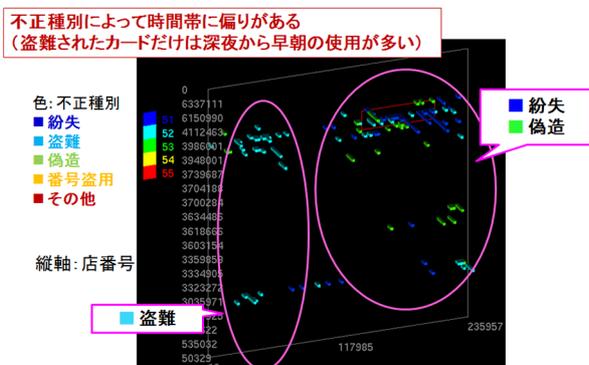


図 6 クレジットカード不正結果の不正種別と決済時間帯の可視化