社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

Twitterにおける情報の広がりの可視化と分類

竹内　俊貴†　　谷川　智洋††　　西村　邦裕††　　廣瀬　通孝††

† 東京大学大学院学際情報学府
〒113–8656 東京都文京区本郷 7–3–1
†† 東京大学大学院情報理工学系研究科
〒113–8656 東京都文京区本郷 7–3–1
E-mail: †{take,tani,kuni,hirose}@cyber.t.u-tokyo.ac.jp

あらまし　“Twitter”とは，140 文字以内の「ツイート」と呼ばれるメッセージを共有することのできる Web コミュニケーションツールの一つである．Twitter には，特定の相手に対する返事を表す “Reply”や，引用を表す “Retweet (RT)”というツイートの種類が存在し，これらを利用することで情報が広がっていく．しかし，通常，ツイートは時系列的に表示されるため，ツイートの関連性が分かりにくく，情報の広がりも見えにくい．そこで我々は，Twitter 上の関連したツイートを再帰的にクラスタリングし，情報の広がりを可視化して表示するアプリケーション “Vital Atlas” を開発した．そして，実際に可視化を行い，いくつかの特徴的な可視化パターンを分類した．

キーワード　情報可視化，Twitter，関連性，再帰的クラスタリング

Visualization and Classification of Information Spreading on “Twitter”

Toshiki TAKEUCHI†, Tomohiro TANIKAWA††, Kunihiro NISHIMURA††, and Michitaka HIROSE††

† Graduate School of Interdisciplinary Information Studies, the University of Tokyo
Hongo 7–3–1, Bunkyo-ku, Tokyo, 113–8656 Japan
†† Graduate School of Information Science and Technology, the University of Tokyo
Hongo 7–3–1, Bunkyo-ku, Tokyo, 113–8656 Japan
E-mail: †{take,tani,kuni,hirose}@cyber.t.u-tokyo.ac.jp

Abstract　We visualized information spreading on the web service “Twitter”. Twitter is a kind of social networking service that users can send and read their short messages (Tweet) . There are tweets related to other ones (e.g. Reply and Retweet). Generally, tweets are lined in order of time, and it is called “Time Line”. Time Line is useful to read recent tweets, but it is difficult to see relations among tweets. Therefore, we developed a system “Vital Atlas” that visualized the relations using a technique of recursive clustering. Then, we found out that patterns and features about distribution of clusters existed.

Key words　Information visualization, Twitter, Relation, Recursive clustering

1. はじめに

近年，社会の高度情報化により，我々自身から発せられる情報量も増え続けている．同時に情報自体が大きな価値を持っていることが広く認識され，プライバシー問題を防ぐ意味においても情報管理が重要な意味を持つようになった．広い意味で，自分自身の発する情報を記録することを「ライフログ」という [1]．ライフログはその性質上，膨大なデータを有するため，記録した生のデータをユーザが扱うことは難しい．ライフログの管理・活用手段として，様々な情報に対して整理・提示を行う研究が多くなされている [2]．また，情報の効果的な提示方法の研究として，データやグラフの可視化を行う研究 [3] [4] も広く行われており，可視化技術はユーザの情報選択性の向上や情報の整理，一種のメディアアートとして用いられている．

本研究では，ユーザが発する情報源として，急速に広まりつつある Web サービス “Twitter”に注目した．Twitter とは，ユーザが 140 文字以内で発言（ツイート）を投稿し，共有することのできるサービスである．まだ比較的新しいメディアであ

るが，独特のコミュニケーション機能や手軽さから広く利用され，高い評価を得ている [5]．ユーザ同士のコミュニケーションや，ツイートの引用などにより，Twitter 上に投げられた情報はインターネット上で時として大きな広がりを見せる．しかし，一般的に，投稿されたツイートは投稿日時順に並べられることが多く，時系列にツイートを参照するには便利であるが，ツイート同士の関連性が見えにくいという問題がある．また，関係のないツイートに埋もれてしまい，特定の情報の規模・広がりが分かりづらく，ユーザからの情報選択性にとって非効率的であるといえる．そこで我々は，多数のツイートを可視化して表示することで，情報選択性の向上や，情報の広がりのパターンを見つけることができるのではないかと考えた．

本研究は，膨大なツイートを関連性を基にクラスタリング，情報の広がりを可視化し，情報選択性の向上や情報パターンの抽出を目的とする．我々は，Twitter 可視化アプリケーション “Vital Atlas” を制作し，複数人のユーザのツイートを可視化した．本論文においては，Vital Atlas の仕組みと可視化された情報パターンについて説明する．

2. Twitter における情報関連性

表 1　Twitter における関連性
Table 1　Relation on Twitter

Follow	自分がフォローしているユーザ
Follower	自分をフォローしているユーザ
Reply	他ユーザ宛のツイート
Retweet (RT)	他ツイートの繰り返し
Quote Tweet (QT)	他ツイートの引用
Favorite	お気に入りのツイート
ハッシュタグ	“#” を頭につけたタグ

Twitter おける関連性を含む情報としては，表 1 のようなものがある．Follow・Follower はユーザ同士の関係性を表すもので，自分がフォローしているユーザと自分をフォローしているユーザである．Reply・RT・QT はツイート同士の関連性を表すものである．Reply とは，他ユーザに対して返事をするように発言することで，Twitter 上での会話などに広く使われる．RT とは Retweet の略であり，他人のツイートをそのまま繰り返し投稿する機能で，興味を持ったツイートを自身のフォロワーに広める役割を持つ．QT とは，Quote Tweet の略であり，他人のツイートを引用した上で自身の発言を付加して投稿する機能である．Reply，RT と異なり非公式な機能であるが，使い勝手の良さから広く使われている．また，Favorite はお気に入りのツイートを登録することができるというものであり，ハッシュタグは “#” を頭に付けてタグで，同じテーマについてツイートする際に探しやすくするためのタグ機能である．

SNS における情報の分析・可視化は多く行われており [6]，Twitter に関しても例外ではない．Twitter における情報の可視化として，“Twitter Friends Network Browser” [7] や “mentionmap” [8] が挙げられる．これらは，ツイート同士の関連性を基に，Follow・Follower 関係を可視化して表示する．

特定の情報の広がりを知るには，これらの情報よりも，ツイート同士の関連性である Reply・RT・QT を利用するのが妥当である．起点となるツイートから，Reply・RT・QT により情報が広がり，一つの関連したツイート群であるクラスタができる（図 1）．このクラスタが，一つの情報の広がりを表す．

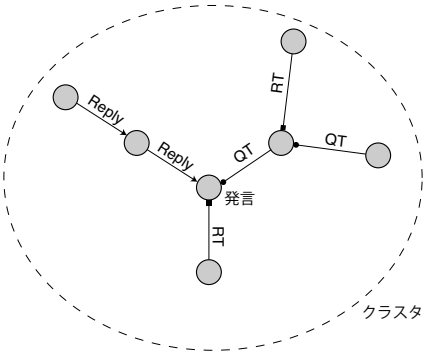


図 1　発言間の関連性
Fig. 1　Relation between tweets

本研究では，Twitter における関連性として Reply・RT・QT を用いて，再帰的なクラスタリングによりクラスタを生成し，それを可視化する．尚，RT は QT を基に作られた公式機能であり，両者は他発言を繰り返すという意味で同じであるため，本研究では区別せず，共に「RT」と呼ぶこととする．

3. 情報の広がりの可視化

可視化を行うにあたり，図 2 のような手順で処理を行った．ツイートの取得，再帰的クラスタリング，空間内における配置，そして可視化である．本章では，これらの処理を含む可視化手法について説明する．

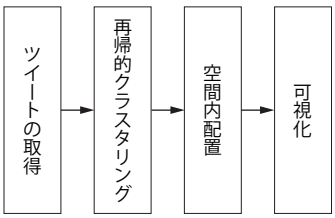


図 2　可視化までの一連の処理
Fig. 2　Flow chart of our system

3.1 発言の再帰的クラスタリング

まず，関連のある複数のツイートをまとめてクラスタを生成するため，再帰処理を用いてクラスタリングを行った（図 3）．最初に，ユーザ自身のツイート，mention（ユーザに向けられたツイート．Reply や QT を含む．），公式 RT されたツイートを取得する．ユーザ自身のある 1 つのツイートに対し，取得した全てのツイートを探索し，Reply・RT の関係を持つツイートを見つける．関連ツイートを見つけたら，そのツイートに対しさらに探索処理を行うという「深さ優先探索」で再帰的にクラ

スタを形成する．

個々のクラスタクラスは，1 つのメインツイートと複数の関連ツイートからできている．あるクラスタクラスの関連ツイートが，1 段階外側のクラスタクラスのメインツイートとなるように構成されており，クラスタの構造自体が再帰的な構造となっている．このように，再帰的なクラスタリング処理・構造を用いることで，柔軟に処理することが可能とした．

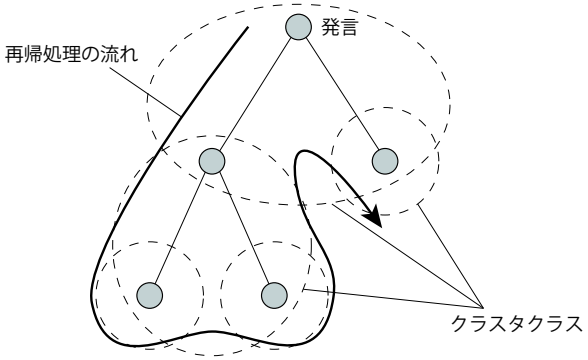


図 3　発言の再帰的クラスタリング
Fig. 3　Recursive clustering

3.2　空間内における配置

次に，クラスタ内の関連した発言を，どのように空間内に配置したかを説明する．図 4 は，配置の様子を示したものである．図 4 の中心が自身の起点ツイートであり，そこから関連したツイートが広がり，つながっていく．本研究では，個々のツイート間の距離は，各ツイートと起点ツイートとの時間差に比例するように配置した．このように配置することで，情報の時間的な広がりを表現することが可能となり，また，密度によって視覚的に情報の重要性を知ることができるのではないかと考えた．

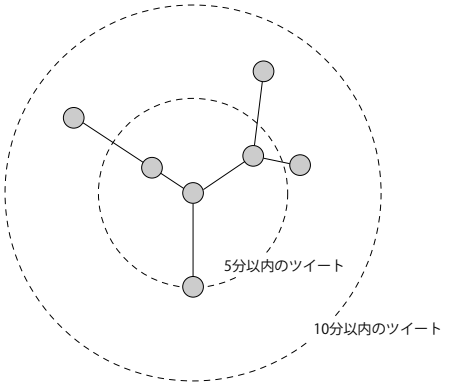


図 4　ツイートの配置
Fig. 4　Position of tweets

3.3　情報の広がりの可視化

そして，図 5 のように，実際に可視化を行う．図 5 において，クラスタの中心にある青白い光が，そのクラスタの関連性の起点となっている自身のツイートである．周囲に配置されている無数の光が関連ツイートであり，赤色は Reply，緑色は RT を表している．ツイート間の線がツイート間の直接的なつながりを示している．また，光の横に各ツイート内容が表示されており，発言・会話がどのように展開されていったのかを分かりやすく見ることができる．

図 5 の例では，左下と右下に向かって 2 回 RT されており，上方向に向かって Reply が 2 回続いているのが分かる．また，Reply のほうが距離が近く，短時間で会話をしている．．このように，視覚的に美しいだけでなく，情報の広がり，その性質を直感的に理解できることを念頭に可視化を行った．

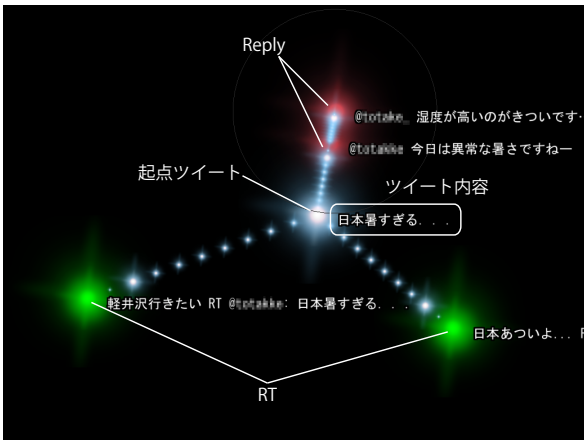


図 5　可視化の一例
Fig. 5　A sample of visualization

4.　可視化パターン

本章では，実際に可視化した様々な情報の広がりのパターンを紹介する．

4.1　Reply のパターン

まずは Reply の可視化パターンを紹介する．Twitter をチャットの様に用い，1 対 1 で対話すると図 6 のようになる．赤色の点が長くつながって可視化されている様子が分かる．また，複数人から話しかけられたり反応されると図 7 のように表示され，赤い点が分岐してつながる．

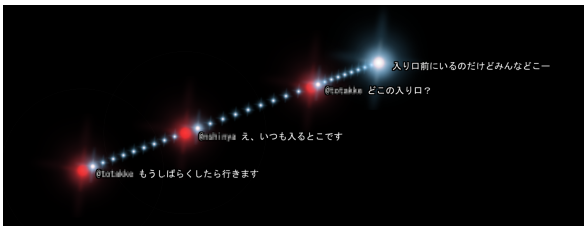


図 6　対　話　型
Fig. 6　Type: conversation

4.2　RT のパターン

図 8 は，著名人に RT され，1500RT 以上ものつながりを持ったクラスタの可視化の様子である．ある 1 つの発言が多く RT されているため，全体的に広がっている様子が見て取れる．対して図 9 では，つながりが深くなっていっている．RT された発言が再び RT され，それを繰り返して拡散していったことが分かる．

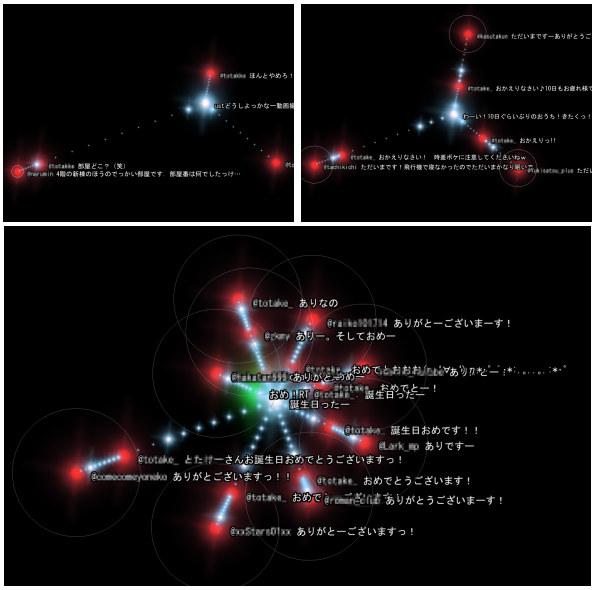


図 7　複数人会話型
Fig. 7　Type: Conversation among multi-users

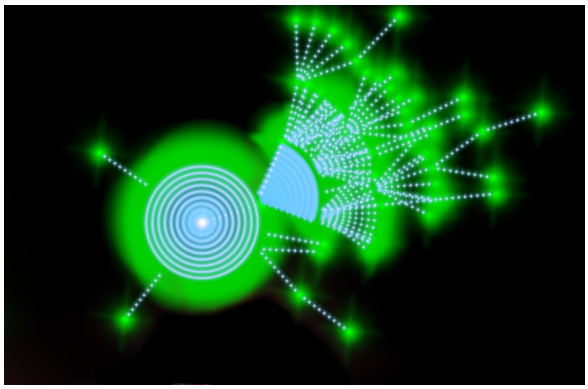


図 8　広く拡散した RT
Fig. 8　Widely spreading of RT

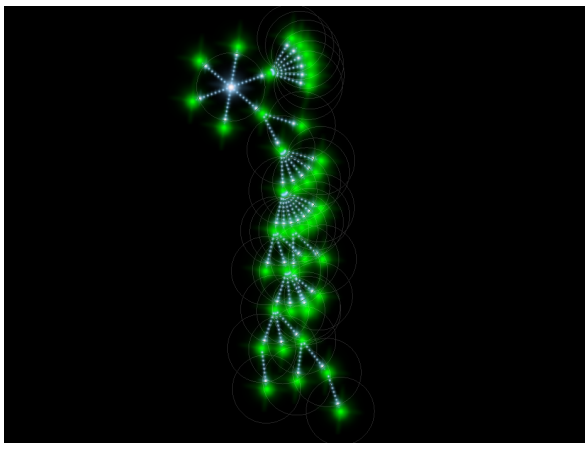


図 9　深く拡散した RT
Fig. 9　Deep spreading of RT

4.3　混在パターン

Reply と RT が混在した図 10 のようなパターンもある．

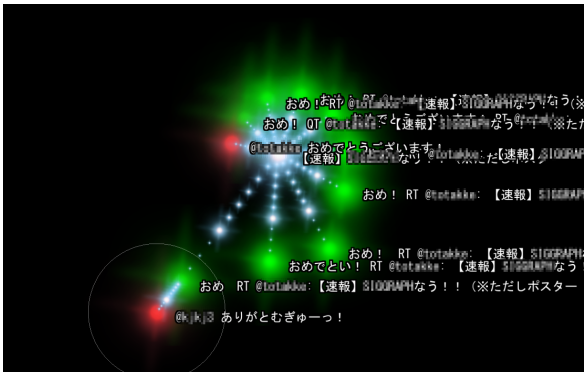


図 10　Reply と RT が混在した可視化パターン
Fig. 10　Mixed pattern

5.　結論と展望

本研究では，Twitter における情報の広がりを，Reply や RT といった関連性と時間的広がりに基づいて可視化するアプリケーションを開発した．実際に可視化されたパターンから，起点情報からの広がりが視覚的に様々な形で表れることが分かった．

今後の展望として，可視化パターンと起点ツイート，ユーザの性質などを解析し，何らかの法則を見つけ出すことがある．本研究において，主観的にではあるが，情報の種類と可視化パターンには規則性が見られるように感じた．加えて，ユーザが可視化された情報の広がりに直感的にアクセスできるようにすることで，Twitter における集合知を利用することができるのではないかと考えられる．

文　献

[1] J. GEMMELL,“MyLifeBits : A Personal Database for Everything,”Communications of the ACM，Vol. 49，No. 1，pp. 88-95，2006.
[2] 山口真弘，青木貴司，谷川智洋，廣瀬通孝，“5ZK-6 時間軸と空間軸を利用した情報提示インタフェースの構築 (情報爆発時代におけるマルチメディアデータと位置依存情報処理, 学生セッション,「情報爆発」時代に向けた新しい IT 基盤技術)，”全国大会講演論文集，Vol. 70, No. 5, pp. 5-225, March. 2008.
[3] Michael Bostock, and Jeffrey Heer, “Protovis: A Graphical Toolkit for Visualization, ” IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, Vol. 15, pp. 1121-1128, 2009.
[4] John Ellson, Emden R. Gansner, Eleftherios Koutsofios, Stephen C. North, and Gordon Woodhull, “Graphviz and dynagraph - static and dynamic graph drawing tools, ” Springer-Verlag, 2003.
[5] 三根慎二，“オープンアクセスをウォッチする 10 大ツール (<特集>オープンアクセス)，”情報の科学と技術, Vol. 60, pp. 156-161, Apr. 2010.
[6] 橋本康弘，陳 Yu，大橋弘忠，“ソーシャルネットワークからのコミュニティ時系列の抽出と可視化分析，”情報処理学会研究報告. MPS, 数理モデル化と問題解決研究報告，Vol. 2008，pp. 63-66，Sept. 2008.
[7] “Twitter Friends Network Browser, ” http://www.neuroproductions.be/twitter_friends_network_browser/
[8] “mentionmap, ” http://apps.asterisq.com/mentionmap/