1V-7

Twitterにおけるツイートの関連性可視化システム

小川　貢平　　芝田　圭佑　　藤井　友紀子　　濱川　礼

中京大学　情報理工学部　情報システム工学科

1. 概要

本論文では，Twitter 上の，ユーザーが指定したツイートを用いてツイートの関連性と話題の特徴を可視化するシステムについて述べる．本論文での，関連性とは Twitter 上で近い話題をしているツイート間のつながりである．ユーザーに話題の種類，話題の規模，Twitter ユーザーの意見を容易に提供することを目的とする．

2. 背景

Twitter では Twitter ユーザー個人のタイムライン上を時系列順で見るとツイートには様々な話題が取り上げられている．しかし，個人のタイムライン上に話題の全ての内容が表示されることはない．従って，自分のタイムライン上に表示されたツイートに関して一目見ただけでは，そのツイートが何の話題について触れているか，どれぐらいの Twitter ユーザーが興味を持っているか，自分以外の Twitter ユーザーがどのような意見を持っているかが分からない．そこで我々は，Twitter 上のツイートの関連性と話題の特徴を分かりやすくするために本システムの開発を行った．

3. 本システムについて

本システムは，ユーザーの指定したツイートの話題の特徴を表現するため円状木構造グラフ及び話題値・キーワードの 3 つを提案し用いる．円状木構造グラフではツイート同士の関連性と時間の流れを可視化する．話題値は万人受けする話題であることを表す指標を，キーワードでは話題の中心と考えられる単語をそれぞれ示している．これにより，ユーザーは話題の種類，話題の規模，ユーザーの意見を容易に知ることが可能になる．

4. 関連研究

[1]では話題(趣味)の近いユーザーを検索し木構造で出力，[2]では同じ「話題」に触れているツイート群をグループ化し，タイムラインで出力している．またビッグデータの可視化を目的とした研究として[3]が挙げられる．[1][2]では単なる木構造やタイムラインで出力しているが本研究では[3]と同様に円状木構造を使用し関連性を可視化し，話題の特徴を話題値・キーワードとして出力する．

また，[1][2]ではリツイートや個人のタイムラインを用いて研究を行っているが本研究ではより多くのツイートの関連性を発見するためタイムライン・リプライ・リツイートを対象としている．

Visualization of connection using tweets in Twitter
Kohei Ogawa , Keisuke Shibata , Yukiko Fujii
and Rei Hamakawa

5. 提案手法

5. 1. 内部構成

本システムは通信部，解析部，ユーザーインターフェース部から成る．

5. 2. システム概要

ユーザーは検索したいツイートを入力する．通信部で入力したツイートに対して関連性のあるツイートを発見するためタイムライン・リプライ・リツイートを取得し，取得したリプライ・リツイートに関してはツイートの関連性を可視化するため伝播する経路の作成を行う．解析部で話題値の算出とキーワードの抽出を行う．ユーザーインターフェース部で解析部が作成したデータを読み取り円状木構造グラフ・話題値・キーワードを表示する．ツイート同士の関連性と時間の流れを表現するため円状木構造グラフを用いた．

5. 3. 内部処理

5. 3. 1. 通信部

ユーザーが入力したツイートに対するツイートの取得を行う．また，リプライ・リツイートに関しては伝播する経路の作成を行う．

5. 3. 2. 解析部

通信部が取得したツイートの定式化・保存を行う．また，話題値の算出とキーワードの抽出を行う．

5. 3. 2. 1. 話題値算出

リツイート数だけでは話題がどの範囲の Twitter ユーザーまで興味を持っているのか分からないため我々が定義した話題値の算出を行う．

Twitter ユーザーは自身のタイムライン上からリツイート・リプライを行うため，必然的にフォロー・フォロワー関係が成り立つ．また，自身に近い趣味や興味のある話題をツイートする Twitter ユーザーをフォローする傾向にあるため，元ツイートと直接繋がっている Twitter ユーザーは元ツイートに興味を持っている可能性が高い．逆に，元ツイートと繋がっていない Twitter ユーザーは普段はその話題にそれほど興味を示さないということである．普段興味を示さない Twitter ユーザーが多く拡散させている場合その話題は広い範囲の Twitter ユーザーに興味を引く話題であると言える．よって，全体のツイート数から元と繋がっていないツイート数の割合を算出することによって万人受けする話題かどうかを表すことが可能になる．

話題値は以下の式を用いて算出する．

$$\frac{\text{元ツイートと繋がっていないツイート数}}{\text{全体のツイート数}} = \text{話題値}$$

話題値は 0～0.999…の範囲で算出される．高ければ高いほどそのツイートは万人受けする値を表す．

5. 3. 2. 2. キーワード抽出

キーワードを抽出する方法として形態素解析 Yahoo!API を用いた．解析したい文字列を渡すことによ

り単語ごとに分解される．通信部が取得したツイート内で出現回数を求める．しかし，出現回数は相対数ではなく絶対数のため *tf-idf* 法を用いて正規化を行う．正規化した値の上位 5 位の単語をキーワードとして設定する．

5.3.3. ユーザーインターフェース部

解析部が作成したデータを読み込み円状木構造グラフ・話題値・キーワードを出力する．

5.3.3.1. ノード配置

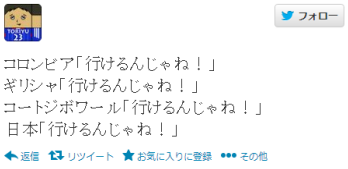
木構造データを円状木構造グラフになるようノードを配置する．中心から外側に向かって時間軸を設定した．よって中心に配置されるツイートは投稿時刻が一番早く投稿時刻が遅くなるにつれ中心のノードから距離が遠くなるよう配置する．Twitter の性質上ツイートは投稿した直後の時間帯にリツイートが集中してしまうため，そのまま時間に比例させて配置した場合中心にノードが集まってしまう．そのため今回は中心からの距離を求める時に対数を用いた．また，１つのツイートから繋がるツイートは１つから複数と様々なため，各ノードを親とした場合の葉ノード数を算出し，木構造グラフ全体のバランスをとるようノード配置する．

5.3.3.2. ノードの縮小・削除

木構造グラフ全体のバランスを保つためノード数に合わせて表示の際にノードの縮小･削除を行う．縮小する場合はノードを半分の大きさにし，ノードとノードを繋ぐ線の濃さを薄くする．削除する場合はノードを削除し他のノードが使うスペースを広くする．対象となる条件は，リツイート・中心のノードと繋がっている・繋がっているノード数が 1 の 3 つである．

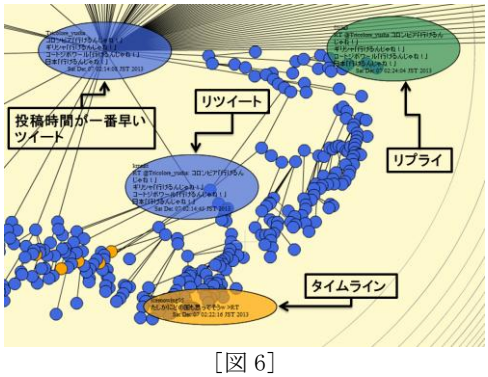
6. 評価・成果

大学生 16 名に対し，同じツイート[図 5]について本システムを用いて検索を行った場合と手動で検索を行った場合について評価を行った．



[図 5]

ツイート[図 5]を本システムを用いて可視化した結果[図 6]のようになった．取得できたツイート数は 2372 件であった．　拡散スピードとツイートの伝播する経路の様子を可視化し，Twitter ユーザーの関連性を表現することができた．



[図 6]

ツイート[図 5]を本システムを用いて話題値を算出した結果 0.319 となった．これは普段この話題について興味を示さない Twitter ユーザーが約 3 割しか興味を示さなかったことを示す．よって，このツイートは広い範囲の Twitter ユーザーに興味を引く話題ではないことが分かる．

キーワードについては評価者が考えたキーワードと本システムが出力したキーワード[図 7]の一致率は 20%だったが，「本システムで提供したキーワードは話題と合っているか」という問いに対して 85%の人が合っているとの回答を得ることができた．よって，本システムは人とは異なる視点で話題の分析が可能である．

評価者TOP5	本システムTOP5
日本	みんな
コロンビア	グループ
コートジボワール	コートジボワール
ギリシャ	国
サッカー	決勝

[図 7]

本システムが取得したタイムラインを分析することで Twitter ユーザーの意見を検索でき，手動でグループ分けすることができた．よって本システムではユーザーの意見の違いを判断する支援をすることができる．

「円状木構造グラフの見た目は綺麗か」という問いに対しては 44%の人が綺麗，「ツイートの内容が読めますか」という問いに対しては 88%の人が読めるという回答をした．

ツイート[図 5]に関するリツイート・リプライ・タイムライン各 10 件の検索について手動で検索を行った場合と本システムを用いて検索を行った場合との検索時間の比較を行った．手動で検索を行った場合は平均 56 分かかったが，本システムで検索を行った場合[*]は 11 分 5 秒であったため，本システムを用いた方が速くツイートを検索することができた．

7. 考察

上記に示すように，話題の種類，Twitter ユーザーの意見の違いを本システムユーザーが判断する支援を促すことができた．また、本システムを使用することで手動で行う場合より，リツイート・リプライ・タイムラインを数多く提供することができた．しかし，円状木構造グラフについては良い評価を得られなかった．

8. 展望

現状の円状木構造グラフではノードが並べられているだけであり，ユーザーがそのノードの内容を分析し Twitter ユーザーの意見の違いを判断する．そのため今後ツイートを解析し，グループ分けを行う必要がある．

9. 参考文献・関連研究

[1] 太田侑介：“Twitter におけるリツイート経路の可視化とユーザー発見支援”，電気通信大学平成 22 年度卒論

[2] 青島傅隼他：“文脈的なつながりを考慮したツイート群の効果的な抽出・提示手法の実現”，情報処理学会論文誌 2013-3, vol6, No. 2

[3]Fernanda Viegas 他：Google+Ripples: a native visualization of information flow,WWW'13

[*] 実験環境　OS:Windows7, CPU:Intel(R) Core(TM) i5 M520@2.40GHz, メモリ:4.0GB