Twitter 特有のネットワーク構造を用いたユーザ重要度評価法の提案

石垣　藍睦†　　沼尾　雅之†

† 電気通信大学大学院情報理工学研究科情報・通信工学専攻　〒 182–8585 東京都調布市調布ヶ丘 1–5–1
E-mail: †i1431009@edu.cc.uec.ac.jp, numao@cs.uec.ac.jp

あらまし　近年，マイクロブログの一つである Twitter は，ユーザ間の情報のやりとりのツールとして急速に普及してきた．そのユーザ間の情報には，重要なユーザとそうでないユーザが発信したものが混在している．重要な情報を取得するためには，そのようなユーザを分類することが課題となる．Twitter の機能には，他のユーザへ情報を拡散するために再共有（リツイート，以後 RT）がある．そのため，RT する回数が多いユーザは情報を拡散させやすく，RT される回数が多いユーザは信頼度が高いと考えられる．また，RT の反応速度が速いユーザは情報に敏感で重要なユーザであると考えられる．そこで本論文では，RT の回数と反応速度を考慮したネットワークを基にユーザの重要度を推定する手法を提案する．

キーワード　ネットワーク分析，ソーシャルネットワーク，Twitter

1. はじめに

近年，マイクロブログの一つである Twitter は急速に普及してきた．現在（2015 年 9 月 30 日）の Twitter の月間アクティブユーザは，全世界に 3 億 2000 万人存在する [1]．Twitter では，ユーザが最大 140 文字の投稿（ツイート）で情報発信することで，ユーザ同士の情報の交換ができる [2] [3]．その情報には，ユーザの意見や感情が含まれることが多く，実社会に有益であるのではないかと注目されている．そのため，ユーザの情報を対象にした研究が盛んに行なわれている [4] [5]．

研究対象として大きく分けてツイートとユーザの 2 つがある．ツイートを対象にした研究では，各立候補者に言及したツイートからの選挙の各立候補者の当選予測や，災害時の緊急情報に言及したツイートのデマ判別がある [6] [7]．それらは，ツイート本文の特徴やツイートを投稿したユーザのフォロー数などの属性に注目している．そして，ツイートによる実社会への関係やツイート自体の重要性・信頼性を評価する．また，ユーザを対象にした研究では，無数のユーザから探し求めているユーザを推薦するものがある．Twitter では，企業の公式のアカウントや特定の分野の有名人の信頼度が高いユーザがいる．その一方で，機械的に無意味な情報を発信するようなユーザもいる．そのため，ツイートには重要なユーザとそうでないユーザが発信したものが混在している．重要な情報を取得するためには，そのようなユーザを分類することが課題となる．

ユーザの分類における研究では，ユーザの属性の情報や Twitter 特有の機能による情報を用い，ユーザの重要度を推定することがある．ユーザの属性を用いたユーザの分類における研究では，他のユーザから情報を取得するための登録数（フレンド数）や他のユーザから情報を取得されるための登録数（フォロワー数）を用いる [8]．フレンド数やフォロワー数などは静的な情報である．なぜならユーザのフレンド数やツイート数とは，他のユーザとの情報のやり取りの情報ではないためである．静的な情報では，刻一刻とユーザ間で情報がやり取りされている Twitter 上ではユーザの推定に適していないと考えられる．

また，Tiwitter の特有の機能によるユーザ分類における研究がある．それらは，他のユーザの情報を取得するための登録 (フォロー）や他のユーザへ情報を拡散するために再共有（リツイート，以後 RT）を用いる．これらの機能による情報は，ユーザ間のリンクとしネットワークと捉えることが多い．そのため，フォローや RT によるネットワークをフォローネットワークや RT ネットワークと呼ぶことがある．フォローネットワークは静的なネットワークであり，RT ネットワークは動的なネットワークであるといえる．なぜならフォローネットワークは，1 度ユーザ同士が繋がってしまうと，ユーザ間での情報交換の有無を知ることができないためである．また，RT ネットワークは，ユーザ間での情報交換の頻度や反応速度を知ることができるためである．RT は，情報間の頻度や反応速度の動的な情報を知るための数少ない機能といえる．ユーザの重要度推定において，ユーザの情報発信の頻度や速度は非常に有用であると考えられる．

そこで，本研究では，RT の回数を考慮した RT ネットワークと RT の反応速度と RT の回数考慮した RT ネットワークを提案する．頻繁に RT をされるユーザは，他のユーザからツイートを参照されやすく重要な情報発信源である．そのため，RT する回数が多いユーザは情報を拡散させやすく，RT される回数が多いユーザは信頼度が高いと考えられる．通常のスコアリンクアルゴリズムでは，リンクの重みが存在しない．そこで，リンクの重みを RT の回数にすることで，RT の回数を考慮した RT ネットワークを提案する．また，RT の反応速度が早いユーザは，情報に対して敏感で重要なユーザである．そのため，一回の RT にもそれぞれリンクに反応速度の重みを付加することによって，そのようなユーザを発見できると考えられる．そこで，はじめの RT ネットワークのリンクの重みに反応速度を考慮することで，RT の反応速度と RT の回数を考慮したネットワークを提案する．

本研究の目的は，Twitter のユーザから RT の反応速度や RT

の回数という動的要素を考慮した重要なユーザを発見することである．そのため，本提案の RT ネットワークを WEB のネットワークと捉え，スコアリンクアルゴリズムを適用することでユーザの重要度を推定する．

2. HITS アルゴリズム

ユーザの重要度推定では，スコアリンクアルゴリズムの一つである HITS アルゴリズムを用いられることがある．HITS アルゴリズムは，Kleinberg が考案したハイパーリンク構造を用いた WEB ページのランキング手法の一つである [9]．WEB のハイパーリンク構造は，評価されているリンク（被リンク）と評価をしているリンク（発リンク）で構築されている．HITS アルゴリズムは，ハイパーリンク構造においてオーソリティ，ハブの二つの概念を以下のように定義した．

（1） オーソリティ：重要な情報を発信しているページ

（2） ハブ：重要な情報を発信しているページに発リンクしているページ

オーソリティは，定義 1 から重要なハブからの被リンクを多く受けているほど，重要なオーソリティとなることを意味する．ハブは，定義 2 から重要なハブほど重要なオーソリティに発リンクすることを意味する．二つの概念から考案された評価値は，オーソリティスコアとハブスコアである．

3. 関 連 研 究

3.1 実世界の動向の予測

マイクロブログでは，リアルタイムなユーザの情報が入手しやすい．そのため，Twitter 情報を実世界と動向の予測に用いる研究が盛んに行われている [10].

筆者らは，以前為替取引に関するツイートの集合から為替予想に特化した評価表現辞書の構築法の提案した [5]．評価表現とはポジティブ・ネガティブの数値が付与された単語であり，評価表現辞書とは評価表現の集合である．筆者らは評価表現辞書を構築する際に，データセットを為替取引のツイートのみにした．それにより，為替のドメインに特化した評価表現辞書の構築した．そして，構築した評価表現辞書によって為替取引のツイートの評価し，その結果と為替レートとの関係調査があることを考察した．この研究では，為替取引をするユーザ間のツイートを全て対象にした．しかし，重要度が高くないユーザのツイートも含まれることもあり，為替のドメインに関係のない単語が評価表現辞書に登録されてしまった．そのため，そのような単語を登録しないようにするために，重要度を高いユーザを発見する必要があると考えらえる．

3.2 ユーザの重要度推定

Twitter の膨大なユーザから重要なユーザやツイートを発見することは，非常に困難である．そのため，重要だと考えられるそれらを推薦する研究が盛んに行われている [11] [12].

Jianshu らは，フォロワーが多いユーザとそのユーザのフォロワーを対象にユーザの影響力の推定を行った [13]．ユーザの影響力を推定する手法には，PageRank の拡張である TwitterRank を提案した．

TwitterRank では，はじめに LDA(Latent Dirichlet Allocation) を用いてユーザのツイートに含まれるトピックのユニークな単語をカウントする．そしてユーザに対応した各トピックのユニークの単語数は，図 1 のような特徴ベクトルとして扱う．

	トピックA	トピックB	トピックC	トピックD
ユーザA	1	0	10	9

図 1　TwitterRank で扱う特徴ベクトル

	トピックA	トピックB	トピックC	トピックD
ユーザA	0.05	0	0.5	0.45

図 2　正規化された特徴ベクトル

そして，図 2 のように特徴ベクトルを正規化し，その特徴ベクトルを DT と定義する．特徴ベクトル DT を対象のユーザに対して作成する．この特徴ベクトルを用いて，以下の式のようにユーザ i とユーザ j の類似度 $sim_t(i,j)$ を求める．ただし，t は任意のトピック，DT'_{it} とはユーザ i の特徴ベクトルにおけるトピック t の数値，DT'_{jt} とはユーザ j の特徴ベクトルにおけるトピック t の数値である．

$$sim_t(i,j) = 1 - |DT'_{it} - DT'_{jt}|$$

またフレンドからユーザ i への影響力 $P_t(i,j)$ は，以下の式で表す．ただし，T_j とはフレンドであるユーザ j の総ツイート数，$\sum_{a:s_i follows s_a} |T_a|$ はユーザ i のフレンドの総ツイート数である．

$$P_t(i,j) = \frac{|T_j|}{\sum_{a:s_i follows s_a} |T_a|} * sim_t(i,j)$$

TwitterRank では，ユーザ間のリンクの重みを $P_t(i,j)$ とする．そのため Jiansh らは，リンクにフォロー関係，情報伝播にツイートを用いたネットワークを構築した．そして，このネットワークをスコアリングアルゴリズムである PageRank に適用して，ユーザの重要度を推定した．. この研究により，ユーザの各トピックや全トピックの影響力を推定することができた．

3.3 問 題 点

節 3.1 では対象となるユーザの中で重要なユーザであるかどうかを見分けることが課題である．そこで，節 3.2 で述べた既存研究ではユーザの重要度の推定を行うことが有効であると考えられる．ユーザの重要度を推定するには，図 3 と図 4 より WEB と Twitter 上のリンクを同様に捉えることによってネッ

トワークを構築し，スコアリンクアルゴリズムに適用することが考えられる．

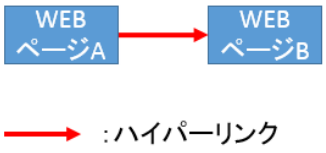

図 3　WEB 上のリンク

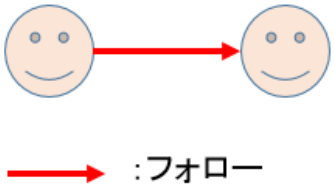

図 4　Twitter 上のリンク

Jiansh らの研究においては，フォロー関係やツイートといった静的な情報を用いてネットワークを構築した．しかしそれでは，WEB と同様にノード同士がどの程度やり取りが行なわれているかが不明であるという問題がある．そのためユーザの重要度の推定では，Twitter 特有の動的な要素が含めることが課題となる．

4.　RT ネットワークにおけるユーザの重要度推定

近年 Twitter の研究では，フォロー関係や RT による情報伝播で構築されたネットワークを分析することが多い．一般的にフォロー関係で構築されるネットワークのことは，フォローネットワークと呼ばれる．図 5 は，フォローネットワークの一つの例である．また，RT で構築されるネットワークのことは，RT ネットワークと呼ばれる．図 6 は，RT ネットワークの一つのである．

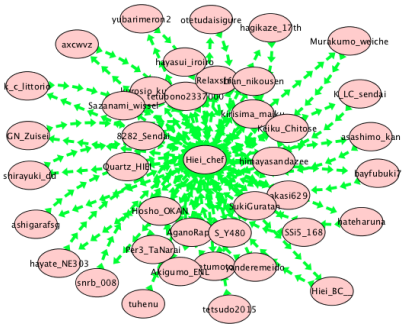

図 5　フォローネットワークの例

フォローネットワークは，Twitter でのユーザ間の静的な要素であるフォロー関係から成り立つ．そして，フォロー関係ではユーザ間で一度フォローしてしまうと，その後の情報のやり取りを知ることができない．そのため，図 5 のように有向グラフにはなるが，単なるリンクであるためリンクの重みがない．しかし，RT ネットワークでは，Twitter でのユーザ間の動的な要素である RT により成り立つ．そのため，RT ではフォロー関係とは異なり，ユーザ間の情報のやり取りをその都度知ることができる．さらに，RT ではユーザ間でのやり取りの反応速度も知ることができる．その動的な要素があるので，図 6 のように有向グラフのリンクに対して重みを付加することができる．そこで，ユーザの重要度を推定するには，フォロワーネットワークを用いるよりも RT ネットワークを用いる方が良いと考えられる．

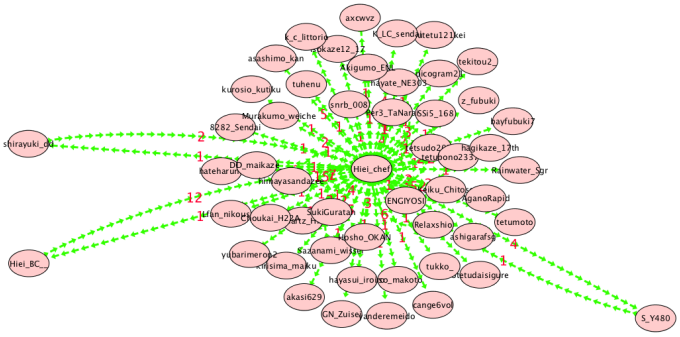

図 6　RT ネットワークの例

本提案の RT ネットワークでは以下の 3 つをユーザ間有向リンクとして定義する．

- リンクの重みを 1 とする RT ネットワーク (Normal Retweet Network，以後 NRN)
- リンクの重みを RT の回数とする RT ネットワーク (Retweet Count Network，以後 RCN)
- リンクの重みを RT の反応速度と RT の回数を考慮した RT ネットワーク（Retweet TimeWeight Count Network，以後 RTWCN）

この表現法の有効性については，第 5. 章の実験で評価する．

4.1　RT ネットワークの構成

以下では，本提案及び既存研究の RT ネットワークがフォローネットワークとどのように対応づけられいるかを説明していく．

4.1.1　既存研究の RT ネットワーク

RT は，フォロワーに RT したツイートを拡散するために行われる．さらに，その RT されたツイートをフォロワーも RT することが可能である．そのため，Twitter 上で行われる大規模な情報伝播は RT によるものである．そこで山本らは，図 7 のように RT による特定のツイートの情報伝播に注目した [14]．

Twitter では，図 7 のツイート番号のように各ツイートに ID が割り当てられる．図 7 では，ツイート番号を 001 としたツイートを RT によってユーザ A，ユーザ B，ユーザ C の順番で情報伝播されていくことがわかる．山本らは，このような RT の情報伝播で構築されたネットワークを RT ネットワークとして定義した．山本らの RT ネットワークは，図 8 で表現できる．このネットワークでは，あるユーザの特定のツイートがどのユーザによって情報が伝搬したかが見て取れる．そのためネッ

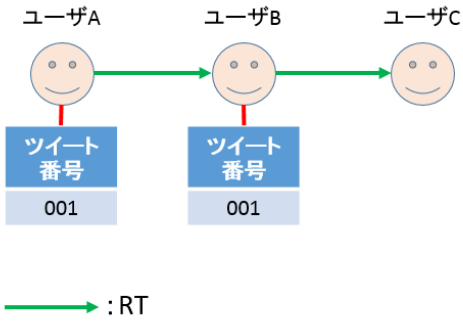

図 7　山本らの RT ネットワーク上でのユーザ同士のリンク

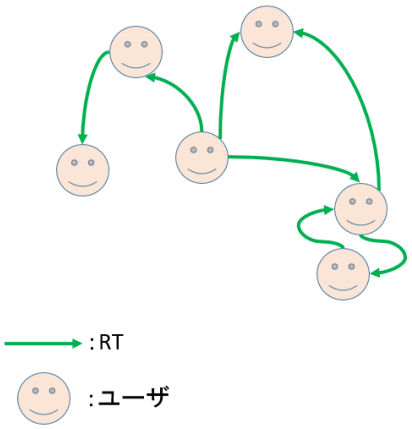

図 8　山本らの RT ネットワーク

トワークの規模によって，ツイートの自体の影響力がどの程度あるかどうかを知るために非常に有効である．フォローネットワークとの対応は，表 1 で示す．

表 1　フォローネットワークとの対応表

	フォローネットワーク	山本らの RT ネットワーク
ノード	ユーザ	ユーザ
ノードの属性	なし	ツイート内容
リンク	フォロー	RT
リンクの重み	1（固定）	1（固定）

4.1.2　RCN

本研究では，3 つの RT ネットワークを提案する．NRN の説明は，RCN のリンクの重みを 1 に固定した場合なので省略する．まず 1 つ目は，**RCN** を説明する．山本らは，特定のツイートの情報伝播された規模を RT ネットワークから分析した．本提案では，ユーザの重要度推定をすることを目的とした RT ネットワークを定義する．

まず，ユーザの重要度を推定するために必要だと考えらえたのは，情報のやり取りの頻度だと考えた．フォロー関係では，情報のやり取りを知ることはできない．RT では，RT の回数だけ情報のやり取りを行われたことがわかる．しかし，RT を用いてる山本らの RT ネットワークでは，1 回の RT のつながりしかなく，情報のやり取りを知ることはできない．そこで本提案の RT ネットワークでは，図 9 のような任意の期間中にユーザが RT した情報に注目した．ただし，Retweet Count(以後 RC) は，リンクの重みである．

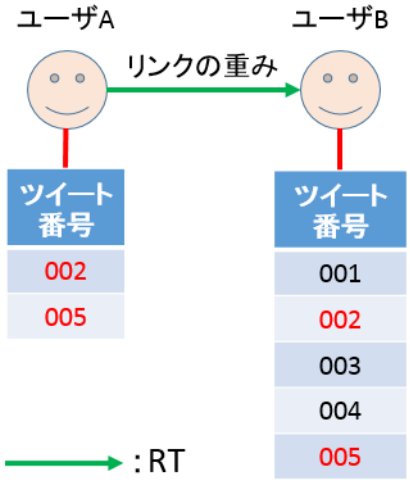

図 9　RCN 上でのユーザ同士のリンク

図 9 では，ユーザ B が任意の期間にツイート番号 001-005 のツイートをしている．そして，ツイート番号 002 と 005 のツイートをユーザ A が RT していることがわかる．つまり，任意の期間にユーザ A がユーザ B のツイートを 2 回 RT したことになる．そのため，RC は以下の式で表現する．ただし，$u(x, y)$ は x が発リンクするユーザと y が被リンクするユーザの組を表す．

$$RC(u(A, B)) = (B \text{ が } A \text{ から } RT \text{ された回数}) = 2$$

任意の期間における複数のユーザで RCN を構築する場合は，図 10 のようになる．

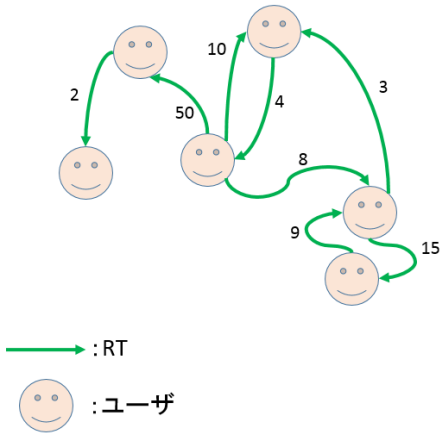

図 10　RCN

フォローネットワークとの対応は，表 2 で示す．

表 2　RCN とフォローネットワークとの対応表

	フォローネットワーク	RCN
ノード	ユーザ	ユーザ
ノードの属性	なし	ツイート番号
リンク	フォロー	RT
リンクの重み	1（固定）	RC

4.1.3　RTWCN

本提案の 2 つ目の RT ネットワークは，**RTWCN** である．RTWCN では RT の回数に加え，ユーザ間の RT の反応速度を考慮する．反応速度の考慮には，戸田らの時間類似度の考えを取り入れて以下のように定義する [15].

戸田らは，タイムスタンプを持つ文書集合に対する話題構造マイニングの提案した．なぜなら，近年ユーザは検索エンジンを用いて最新のニュースなどの情報を得ることが一般的になってきた．しかし，アクセス可能な情報が膨大になりすぎたために，ある一つの主要な話題や特定の話題に関する情報を把握することが困難である．そこで，文書内における複数の話題の関係性や主要な話題を特定する手法である話題構造マイニングを用いること解決しようと考えたためである．

戸田らの手法は，新聞記事のクラスタリングや話題抽出する際に文書間の内容の類似度に加え時間類似度を考慮するものである．時間類似度は，“**文書間のタイムスタンプが一定の時間離れる毎に，一定の割合で類似度が減少する**”の仮定のもと定義される．そして，時間類似度を求める式は，以下のように表現する．ただし，t は二つの記事のタイムスタンプの差，T_0 はタイムスタンプの差が 0 の場合の重み，$t_{1/2}$ は類似度が 50%になるタイムスタンプの差（半減期）である．

$$TimeWeight(t) = T_0 \times exp(-\frac{0.639}{t_{1/2}}t)$$

戸田らの研究では，適切なパラメータをセットすることで，時間類似度を考慮なしよりも精度の高いクラスタリングや話題抽出を行えるようになった．

戸田らの扱う文書は，異なる新聞記事の文書間であった．しかし，本研究で扱う RT は，同じ文書の情報伝播である．そのため，時間類似度 $TimeWeight(t)$ の仮定は RT に最適であると考えられる．そして，本提案に対して時間類似度を 1 回の RT の重みに適用する．付与の方法は，図 11，図 12 を用いて RCN と RTWCN を比較し説明する．

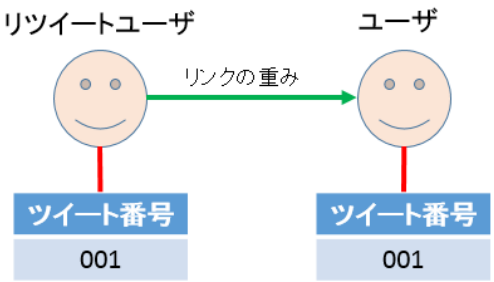

図 11　RCN の RT の重み

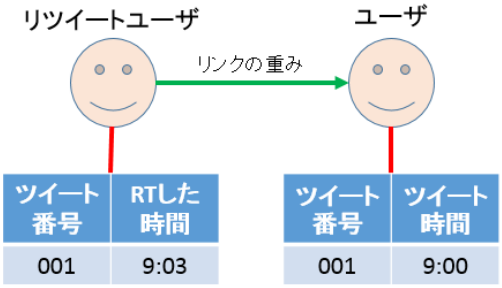

図 12　RTWCN の RT の重み

図 11 では，任意の期間中にリツイートユーザがユーザのツイートを 1 回の RT したことを表現している．RCN の 1 回の RT の重みは，常に 1 で固定されているため，ユーザとリツイートユーザ間のリンクの重みは 1 となる．図 12 では，図 11 と同様の状況を表している．しかし，ツイートの時間がユーザの属性に追加されている．そのため，RTWCN の 1 回の RT の重みは $TimeWeight(t)$ となる．$TimeWeight(t)$ を RT に適用するためにパラメータを次のように定義する．T_0 はツイート時間と RT した時間の差が 0 の場合の重み，$t_{1/2}$ は $TimeWeight(t)$ が 50%になるタイムスタンプの差（半減期），t はツイート時間と RT した時間の差である．本研究では，T_0 を 1 とし，$t_{1/2}$ を 60 分とした

図 12 では，RT した時間とツイート時間の差は

$$t = (RT\text{ した時間}) - (\text{ツイート時間}) = 3\text{ 分}$$

となる．そして，$TimeWeight(t)$ は以下のようになる．ただし，$T_0 = 1$，$t_{1/2} = 60$ とする．

$$TimeWeight(3) = 1 \times exp(-\frac{0.693}{60} \times 3) = 0.966$$

次に任意の期間中にリツイートユーザが，複数回の RT をされた場合を図 13 を用いて説明する．

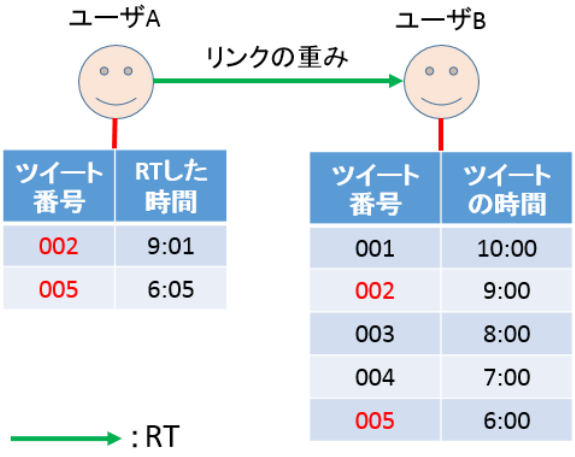

図 13　RTWCN 上でのユーザ同士のリンク

図 13 では，ユーザ A がユーザ B のツイート番号 001-005 の

中からツイート番号 002 と 005 を RT したことが表現されている．さらにユーザ B は，ツイート番号 002 と 005 のツイートをそれぞれ 9:00 と 6:00 にツイートしている．一方ユーザ A は，ツイート番号 002 と 0005 のツイートをそれぞれ 9:01 と 6:05 に RT している．RTWCN では，RT の反応速度を考慮するためにユーザ A のツイートに対するユーザ B の反応速度を求める．ツイート番号 002 におけるユーザ B の反応速度は，1 分である．ツイート番号 005 におけるユーザ B の反応速度は，5 分である．そのため，ツイート番号 002 と 005 の $TimeWeight(t)$ は，以下のように計算できる．

- ツイート番号 002 の場合

$$TimeWeight(1) = 1 \times exp(-\frac{0.693}{60} \times 1) = 0.99$$

- ツイート番号 005 の場合

$$TimeWeight(5) = 1 \times exp(-\frac{0.693}{60} \times 5) = 0.93$$

図 13 では，RT が複数回行われているためリンクの重みを Retweet Weight（以後 RW）と定義する．RW は，以下の式で定義する．ただし，x はツイートをしたユーザ，y は RT をしたユーザ，RC は RT された回数である．

$$RW(u(x,y)) = \sum_{i=1}^{RC} TimeWeight_i(t)$$

図 13 に適用すると，

$$RW(u(A,B)) = \sum_{i=1}^{2} TimeWeight_i(t) = 0.99 + 0.93 = 1.92$$

となる．そのため，任意の期間中に収集したデータでネットワークを構築すると，図 14 となる．

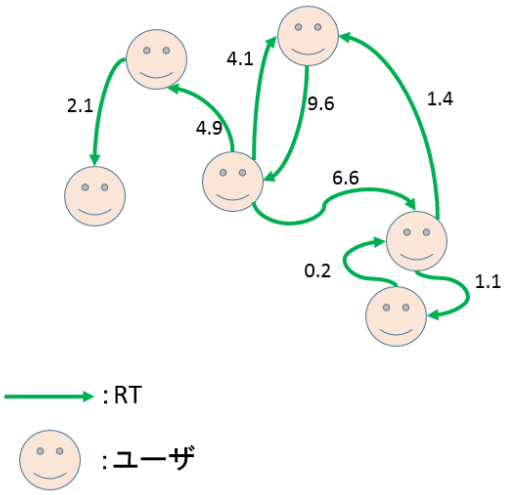

図 14　RTWCN

フォローネットワークとの対応は，表 3 で示す．

表 3　RTWCN とフォローネットワークとの対応表

	フォローネットワーク	RTWCN
ノード	ユーザ	ユーザ
ノードの属性 1	なし	ツイート番号
ノードの属性 2	なし	時間
リンク	フォロー	RT
リンクの重み	1（固定）	RW

4.2　リンクの重みの適用

本研究は，ユーザの重要度を節 4.1 のネットワークを HITS アルゴリズムに適用させ推定する．HITS アルゴリズムでは，有向グラフで表されるネットワークを行列 L で表現する．行列 L は隣接行列と呼ばれ，ある Web ページが他の Web ページをリンクしていることを表す．各 RT ネットワークで定義されたリンクの重みを，どのように隣接行列に適用するかを具体例を示し紹介していく．例えば，図 15 のようなネットワークがあるとする．

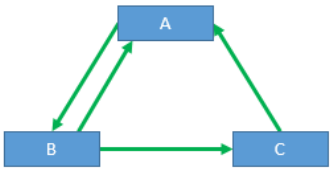

図 15　ネットワークの例

Web 上のネットワークでは，ノードは Web ページであり，リンクがハイパーリンクとすることができる．図 15 を隣接行列で表現すると，以下の行列のようになる．

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

この隣接行列では，Web ページ同士にリンクがあることを 1 で表す．また，行や列は Web ページごとに割り振られ対応している．たとえば 1 列目の要素は，すべて Web ページ A から他の Web ページに対する発リンクの有無を表す．図 15 では Web ページ A から Web ページ B に発リンクがある．2 行 1 列が 1 であるため，Web ページ A から Web ページ B に発リンクがあることを表している．Web ページ A から Web ページ C に発リンクはないため，3 行 1 列が 0 となる．

本研究での RT ネットワークは，節 4.1 で定義したものである．RCN のリンクの重みは RC であるため，図 15 を隣接行列で表現すると以下の行列のようになる．

$$\begin{bmatrix} 0 & RC(u(A,B)) & 0 \\ RC(u(B,A)) & 0 & RC(u(B,C)) \\ RC(u(C,A)) & 0 & 0 \end{bmatrix}$$

また，RTWCN のリンクの重みは RW であるため，図 15 を

隣接行列で表現すると以下の行列のようになる．

$$\begin{bmatrix} 0 & RW(u(A,B)) & 0 \\ RW(u(B,A)) & 0 & RW(u(B,C)) \\ RW(u(C,A)) & 0 & 0 \end{bmatrix}$$

5. ユーザの重要度推定

5.1 目的と環境

本実験の目的は，本提案の RT ネットワークである RCN と RTWCN の有効性を検証することである．

本実験では，図 16 のように RT ネットワークのデータを収集し構築する．ただし，収集する際に起点となるユーザのことをシードユーザと呼ぶ．

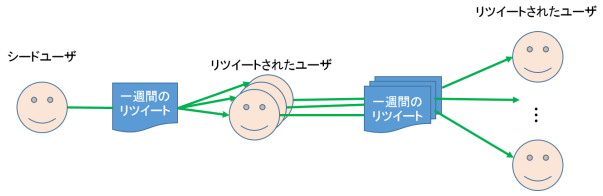

図 16　シードユーザからの RT ネットワークのデータ収集方法

図 16 では，シードユーザの 1 週間の RT の集合を取得する．そして，シードユーザの RT の集合から RT されたユーザを抽出する．次に，その RT されたユーザの 1 週間の RT を取得する．その取得された RT の集合からさらに RT されたユーザを抽出する．このようにシードユーザを起点に RT されたユーザと RT のデータを収集する．そのデータからユーザをノード，RT のデータをリンクにすることで RT ネットワークを構築する．リンクの重みは，NRN，RCN，RTWCN の定義のとおりである．それらを用いて，スコアリンクアルゴリズムに適用する．

5.2 方　　法

本実験でのシードユーザは，gaitame_com を選択する．シードユーザから各 RT ネットワークを構築し HITS アルゴリズムに適用することで，ユーザの重要度を推定する．そして，以下の 2 つの考察を行う．

- 各 RT ネットワークのスコアの重要度分布を考察
- 各 RT ネットワークのスコアが上位のユーザを考察

5.3 結　　果

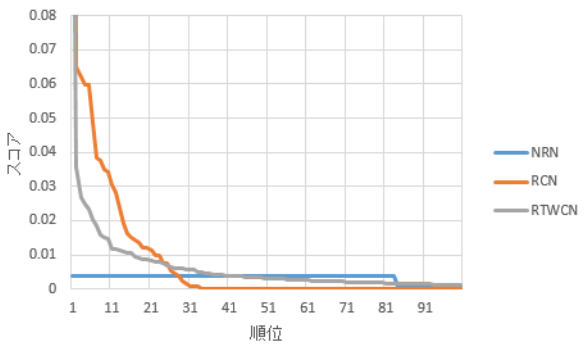

図 17　gaitame_com のオーソリティスコアの重要度分布

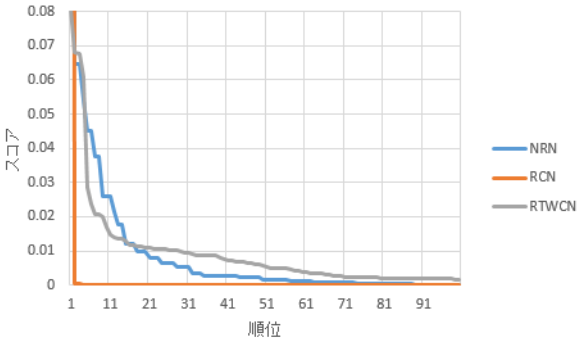

図 18　gaitame_com のハブスコアの重要度分布

5.4 考　　察

図 17 では，NRN での上位ユーザの重要度分布は一定の値を示している．しかし，図 18 では上位ユーザの重要度分布は変化している．そのため，一定の値である理由はハブスコアの高いユーザから発リンクされているユーザのオーソリティスコアが全て同じためであると考えれる．

図 18 では，RCN での上位ユーザの重要度分布は一定の値を示している．しかし，図 17 では上位ユーザの重要度分布は変化している．そのため，NRN とは異なりオーソリティスコアの高いユーザに発リンクしているユーザが多くいることがわかる．

図 17 と図 18 より，どちらのスコアも重要度の分布が変化していることがわかる．そのため，ユーザの重要度を推定するにあたって，RTWCN はユーザの重要度を明確に分かるため有効であると考えられる．

表 4 は，オーソリティスコアにおける上位のユーザである．NRN の上位ユーザの中には，犬の拉致情報やゲームに関する情報などの様々な情報ユーザが存在した RCN や RTWCN の上位ユーザの中には，投資やニュースの情報を発信するユーザが多く存在した．表 5 は，ハブスコアにおける上位のユーザである．NRN の上位ユーザの中には，オーソリティスコア同様に犬の拉致情報を発信するユーザや小説の情報を発信するユーザなどがいた．RCN や RTWCN では，オーソリティスコアの上位ユーザ同様に投資やニュースの情報を発信するユーザが存在した．表 4 と表 5 より，リンクの重みに RT 回数と反応速度を考慮することによって，投資やニュースを発信するユーザが上位に来ることがわかった．つまり，本実験でのシードユーザである gaitame_com が取り扱う為替の分野に近いユーザを知ることができた．さらに RTWCN では，情報の量が多く速いユーザを知ることができた．そのようなユーザは，為替の取引を行う際に非常に重要な情報源となり得ると考えられる．

6. ま　と　め

本研究では，本提案の RT ネットワークを HITS アルゴリズムに適用した．RTWCN では，各スコアの重要度分布がユーザごとに明確に異なるため，重要度を推定するにあたっては有効であると考えられる．また，RT の回数と反応速度を考量することで，為替に関する重要なユーザが上位ユーザに来ることがわかった．

表 4　gaitame_com のオーソリティの上位ユーザ

順位	NRN	RCN	RTWCN
1	0nanairo	okasanman	okasanman
2	18noname01	nhk_news	kabutociti
3	43_25_25_32_42	KandaTakuya	economic_bot
4	amnosick045855	kabutociti	rakuten_fx
5	AntiHero_o	zerohedge	SBILM
6	arpejjio	kirik	xRINGx
7	a_gale	SBILM	metabolic23
8	bluetempests	metabolic23	vkshy
9	Cafi_Nero	kigyo_hp_check	mikumo_hk
10	darkside_mao	kabumatome	KandaTakuya

表 5　gaitame_com のハブの上位ユーザ

順位	NRN	RCN	RTWCN
1	307cc19931113	07grell	kabutociti
2	takedayaofamily	6yamaguchigumi	07grell
3	JohnRentoul	2012_assd	chabuo11
4	imraansiddiqi	akshoukai	xRINGx
5	BreakTpp	anokotoscandal	harusmile
6	yamadataro43	26ooo	hitsuzikai
7	1loriking	advdesk	ny_blackswan
8	AndriiOlefirov	aka1you	kuma1618
9	vgvd	adatarayama	t1190165
10	sinzo_owarida	CuteAnimalsBaby	carl_vinson9

今後の課題としては，より良いデータセットを作成することが考えられる．今回はデータセットを作成する際に，あるユーザの RT の探索の深さを 4 とした．しかし，あるユーザから探索する深さ 4 よりも深い層に，重要なユーザ存在するが考えられる．このようなユーザを効率的に抽出するためにも，データ収集の際にフォーカスクローラーの考えを適用できると考えられる．フォーカスクローラーの考えを適用すると，以下のことが考えられる．

（1） ユーザのタイムラインや自己紹介の内容ので類似度でユーザを探索

（2） RT の回数に閾値を設けてユーザを探索

（3） RT の時間類似度の閾値を設けてユーザを探索

1 では，ユーザのタイムラインの名詞や形容詞などの単語から con 類似度など求めて，ユーザの取捨選択を行うことが考えられる．2 では，RCN ではユーザ間に 1 回でも RT の関係があった場合もリンクを構築している．しかし，それでは一時的な関係性しかないようなユーザでさえも取り扱っていしまう．そのため，複数回のリンクのみを扱うようにすれば，重要な抽出できるのではないかと考えられる．3 では，RT の時間類似度の閾値を設けることで情報に敏感なユーザのみでユーザの重要度を推定できる．

これら 3 つを取り入れることによって，高品質なデータセットでより重要なユーザを抽出できるのではないかと考えられる．

文　献

[1] Twitter Inc.：Twitter の利用状況/企業情報，入手先＜https://about.twitter.com/ja/company＞（参照 2016-1-6）.

[2] 石川哲也，近藤伸也，川崎昭如，大原, 美保，目黒公郎：災害時における Twitter 利用の特徴と課題の整理:-Twitter アカウント運用者の視点に立って-，生産研究，Vol.64(4)，pp.545-552，(2012)

[3] ザイ FX!:FX 実況ちゃんねる, 入手先＜ http://zai.diamond.jp/fxch/＞（参照 2015-6-5）.

[4] 奥村学：マイクロブログマイニングの現在，電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション 111(427), pp.19-24, (2012).

[5] 石垣藍睦, 沼尾雅之:Twitter からの為替予測に特化 したドメイン辞書構成法の提案,FIT2014 情報科学技術フォーラム講演論文集,RO-001,(2014).

[6] 船木洋晃, 佐々木彬, 岡崎 直観:インターネット上の 当選運動・落選運動の分析, 人工知能学会全国大会論 文集 28 回, pp.1-4, (2014).

[7] 梅島彩奈，宮部, 真衣，荒牧英治，灘本明代：災害時 Twitter におけるデマとデマ訂正 RT の傾向，研究報告 データベースシステム（DBS）,Vol.2011，No.4，pp1-6，(2011).

[8] 竹村光，田島敬史：情報発信の対象範囲に基づく Twitter ユーザの分類，DEIM Forum，B1-6,（2013）.

[9] J.M.Kleinberg.: Authoritative Sources in a Hyperlinked Environment, Journal of the ACM, vol.46, no.5, pp. 604-632, (1999).

[10] 荒牧英治, 増川佐知子, 森田瑞樹:Twitter Catches the Flu:事実性判定を用いたインフルエンザ流行予測, 研究報告音声言語情報処理（SLP）,Vol.2011, No.1, pp.1-8, (2011).

[11] Suh,B., Lichan,H., Pirolli,P. and Ed,H.: Want to be retweeted? large scale analytics on factors impacting retweet in twitter network, Social computing (socialcom), 2010 ieee second international conference on. IEEE, pp.177-184, (2010).

[12] 今森大地, 田島敬史：アーリーアダプター推定による優良 Twitter アカウントの早期発見,DEIM Forum 2015,(2015).

[13] Jianshu,W., Ee,P.L., Jing,J. and Qi,H.:TwitterRank: finding topic-sensitive influential twitterers,WSDM 2010,Association for Computing Machinery,pp.261-270,(2010).

[14] 山本雅人，小笠原寛弥，鈴木育男，古川正志，観光情報学：9. 東日本大震災時の Twitter における情報伝播ネットワーク，情報処理学会 ; 1960-，Vol.53，No.11，pp.1184-1191，(2012).

[15] 戸田浩之，北川博之，藤村考，片岡良治： 時間的近さを考慮した話題構造マイニング, 電子情報通信学会 第 18 回データ工学ワークショップ (DEWS2007) 論文集, L6-4 (2007).

[16] 山本雅人，小笠原寛弥，鈴木育男，古川正志:東日本大震災時の Twitter における情報伝播ネットワーク. 情報処理, vol.53, no.11, pp.1184-1191, (2012).