

2012年度 卒業論文

批判関係を利用したマイクロブログの信頼度推定

2013年3月31日

情報知能システム総合学科
(学籍番号: A9TB2112)

佐藤 雅宏

東北大学工学部

概要

近年、マイクロブログの普及に伴い、新聞やテレビのニュースの代わりにマイクロブログに投稿された文章から情報を得る人が増加している。マイクロブログ上には大量の情報が流れているが、その大半が真偽不明の情報であるため、その中から正しい情報を見極める必要がある。マイクロブログ上では各情報に対して、肯定的／否定的な態度（評価）が返信や引用などの手段で表明されている。本研究では、PageRank アルゴリズム [1] に負のエッジを組み込めるように拡張した PageTrust アルゴリズム [2] をベースに、各情報の信頼度を情報間の同意・反論関係に基づいて教師無しで推定する手法を提案する。

目次

第 1 章 序論	1
第 2 章 関連研究	2
2.1 PageRank アルゴリズム	2
2.2 PageTrust アルゴリズム	3
2.3 最適化問題による手法	4
2.4 分類器による手法	5
第 3 章 PageTrust の Twitter データへの適用	6
3.1 PageTrust の問題点	7
3.2 PageTrust の改善	7
3.3 グラフ作成	7
3.3.1 同意・反論エッジの張り方	8
3.3.2 仮ノードの追加	10
3.3.3 ソース URL の追加	11
第 4 章 実験	13
4.1 実験設定	13
4.2 評価方法	13
4.3 実験結果	13
4.4 分析	14
第 5 章 まとめ	17

第1章 序論

近年、マイクロブログの普及に伴い、新聞やニュースの代わりにマイクロブログに投稿された文章から情報を得る人が増加している。マイクロブログとは Twitter や mixi のボイスなどのことで、通常の記事と比べ一度に投稿する文章量を減らす代わりに投稿回数を増やし、今やっていることや思ったことをすぐに書き込むことが出来るブログである。誰でも気軽に投稿できるため、マイクロブログには様々なジャンルの大量の情報が存在するが、その情報の真偽は不明なものが多い。そのため、大量の情報の中から正しい情報を見極める必要がある。このように大量の情報の中から正しい情報を自動推定する手法としては、Lawrence Page らの PageRank アルゴリズム [1] や、Kleinberg らの HITS アルゴリズム [6] などが有名である。これらは Web ページのリンク関係をグラフとして考え、それをグラフ解析することで各 Web ページの信頼度を求めるアルゴリズムであるが、入力するグラフに正のエッジしか張ることが出来ないという問題が存在する。つまり、これらのアルゴリズムではマイクロブログで見られる、ある投稿内容に対する批判的な意見を反映させることが出来ない。

この問題を解決する手法としては、Kerchove らの PageTrust アルゴリズム [2] が存在する。これは、グラフ内の各 Web ページが他の Web ページをどの程度信頼していないのかという概念を新たに追加することで、PageRank アルゴリズム [1] に負のエッジを組み込めるように拡張したものである。このアルゴリズムにより、批判関係を考慮した信頼度推定を行うことが理論上可能になったが、実際のデータを用いた検証はまだされていない。

本研究では、代表的なマイクロブログである Twitter を用いて PageTrust[2] の実データ上での検証実験を行い、ツイートの信頼度推定の性能を評価する。具体的には、図 3.1 に示すように、まず入力されたツイートの集合から PageTrust で計算可能なグラフを作成する手法を提案し、次に PageTrust を用いて信頼度の高いツイートランキングを作成することで、信頼度推定の性能を評価する。グラフ作成時には、各ツイート間の同意・反論関係を判別し、同意関係にあるツイート間に正のエッジ、反論関係にあるツイート間に負のエッジを張ることで、ツイート間の関係を入力するグラフに反映させる。

以降、2章では信頼度推定アルゴリズムに関する関連研究を取り上げ、本研究で使用する PageTrust について詳しく説明する。3章では PageTrust の問題点と解決策を提示し、Twitter の内容を PageTrust に適用するためのグラフ作成手法について説明する。4章では実験設定と実験結果、その分析について記述し、5章では本研究で明らかになった点と今後の課題について述べる。

第2章 関連研究

信頼度を推定する手法には、初期の信頼度を定めてそれを伝搬させる手法 [7, 8]、初期値なしでグラフ解析をする手法 [1, 2, 6, 9]、教師あり学習で作成した分類器を用いる手法 [3]、半教師ありでグラフの最適化問題を解く手法 [4] などが存在する。しかし、今回は確実に正しい情報が存在しない Twitter を対象としているため、初期値や教師の必要としない手法を用いる必要がある。また、分類器による手法はグラフ解析による手法より信頼度推定の精度が低いため、本研究では初期値なしでグラフ解析をする手法 [1, 2, 6, 9] に焦点を当てた。

以下では、初期値なしでグラフ解析をする手法として PageRank アルゴリズム [1] を説明し、その上で本研究で用いる PageTrust アルゴリズム [2] を詳しく説明する。また、その他の信頼度推定アルゴリズムとして半教師ありでグラフの最適化問題を解く手法 [4] と教師あり学習で作成した分類器を用いる手法 [3] を紹介する。

2.1 PageRank アルゴリズム

PageRank アルゴリズム [1] は主に Google が Web ページの重要度を計算するために使用される教師なしグラフ解析アルゴリズムで、Google 検索の上位に表示される Web ページは PageRank のスコアが高いページとなっている。PageRank は確率を用いて計算するため、入力できるグラフは正のエッジのみで構成されたグラフとなる。PageRank に入力できるグラフの例を図 2.1 に示す。PageRank の基本的な性質としては以下の 3 つが挙げられる。

- 多くの Web ページからリンクが張られた Web ページは信頼度が高い
- 信頼度の高い Web ページからリンクが張られた Web ページは信頼度が高い
- ただし、リンク集のような Web ページからリンクが張られても信頼度はあまり高くない

計算方法としては、グラフに存在する各ノード $i \in N$ (N は全ノードの集合) の信頼度 x_i が収束するまで式 (2.1) を反復計算する。ただし、信頼度 x_i は $[0, 1]$ の実数で、各ノード $i \in N$ の初期値は入力するグラフの全ノード数 n を用いて $x_i^{(0)} = 1/n$ で計算される。

$$x_i^{t+1} = \alpha \sum_{j:(j,i) \in L^+} x_j^{(t)} / d_j + (1 - \alpha) z_i \quad (2.1)$$

L^+ は正のエッジの集合を表し、 d_j はノード j から張られたエッジの総数を表わす。ここで、パラメータ α は各グループごとに決められた $[0, 1]$ の定数で親ノードから得られる信頼度の割合を定めている。通常は $\alpha = 0.85$ を用いるが、意図的に PageRank のスコアを上げようとするグループには低い値が用いられる。また、 z_i は親ノード以外から得られる信頼度の値 (ランダム成分) を示している。 z_i は $\sum_{i=1}^n z_i = 1$ を満たし、通常は $z_i = 1/n, i \in N$ で与えられる。

ベクトル x は式 (2.2) のように行列 G の固有ベクトル π に収束することが証明されている。ここで行列 G の各要素は $G_{ij} = \alpha A_{ji}^+ / d_j + (1 - \alpha) z_i$, $i, j \in N$ で計算される。

$$\lambda \pi = G \pi \quad (2.2)$$

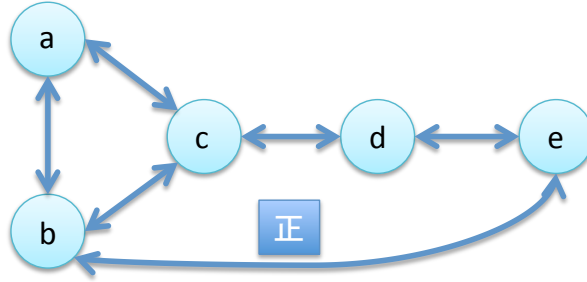


図 2.1: PageRank に入力できるグラフの例

2.2 PageTrust アルゴリズム

PageTrust アルゴリズム [2] は正のエッジのみを扱う PageRank に負のエッジを組み込めるように拡張したアルゴリズムである。正のエッジだけでなく、負のエッジを組み込むことによって、Web ページ A が Web ページ B を批判しているなどのネガティブな意見を取り入れることができ、より精度の高い信頼度推定を行うことができる。PageTrust に入力できるグラフの例は図 (2.2) のようになり、計算方法としては、各ノード $i \in N$ の信頼度 x_i が収束するまで式 (2.3) を反復計算して信頼度 x_i を求める。ただし、 x_i は $[0, 1]$ の実数である。

$$x_i^{(t+1)} = (1 - \tilde{P}_{ii})^\beta \cdot \left(\alpha \sum_{j:(j,i) \in L^+} x_j^{(t)} / d_j + (1 - \alpha) z_i \right) \quad (2.3)$$

\tilde{P}_{ii} は負のエッジを考慮した場合にノード $i \in N$ が信頼できない確率を表し、式 (2.4) で計算される。従って、 $(1 - \tilde{P}_{ii})$ は負のエッジを考慮した場合にノード $i \in N$ が信頼出来る確率を表している。また、 β は負のエッジの影響力を決めるパラメータである。

$$\tilde{P}^{(t+1)} = T^{(t)} \cdot P^{(t)} \quad (2.4)$$

ここで、 T と P はどちらも $n \times n$ の行列であり、 T を遷移行列、 P を反論行列と呼ぶ。それぞれ式 (2.5)、(2.6) で計算される。

$$T_{ij}^{(t)} = \frac{\alpha A_{ji}^+ x_j^{(t)} / d_j + M \cdot (1 - \alpha) z_i x_j^{(t)}}{\alpha \sum_{k:(k,i) \in L^+} x_k^{(t)} / d_k + (1 - \alpha) z_i} \quad (2.5)$$

$T_{ij}^{(t)}$ は時間 t の時のノード $i \in N$ の信頼度のうち、ノード $j \in N$ から得られた信頼度の割合を示す。また、 z_i は $\sum_{i=1}^n z_i = 1$ を満たし、通常は $z_i = 1/n$, $i \in N$ で与えられる。ここで、 M はラン

ダム成分に関して負のエッジの情報を記憶する ($M=1$) かしない ($M=0$) かのどちらかを表すパラメータで、 $M = 1$ の方が $M = 0$ の場合より負のエッジの影響力が高くなる。

$$P_{ij}^{(t+1)} = \begin{cases} 1 & (if (i, j) \in L^-) \\ 0 & (if i = j) \\ \tilde{P}_{ij}^{(t+1)} & (otherwise) \end{cases} \quad (2.6)$$

$P_{ij}^{(t)}$ は時間 t でノード $i \in N$ がノード $j \in N$ を批判している割合を示す。そのため、遷移行列 T と反論行列 P の内積をとることで \tilde{P} の対角成分 \tilde{P}_{ii} がノード $i \in N$ の信頼出来ない確率を表すことができる。初期値としては $P^{(0)} = \tilde{P}^{(0)} = A^-$ を用いる。

PageTrust アルゴリズムの原理を分かりやすく説明するために、random walker という考え方をを用いる。random walker とは、PageRank や PageTrust のもととなったもので、グラフの各ノードに walker と呼ばれる人が存在すると仮定し、各ノードに存在する walker の数が多いほどそのノードの信頼度が高いとする考え方である。初期状態では各ノードに一定の walker が配置され、各反復でエッジが張られたノードへランダムに移動する。そのため、信頼度の高いノードから多くのエッジが張られているノードにはたくさんの walker が集まることになり、そのノードの信頼度は高くなると考えられる。つまり、正のエッジのみを用いた図 2.1 のようなグラフを入力した場合に、random walker によって導かれたアルゴリズムが、2.1 節で説明した PageRank である。

では、ここで負のエッジを追加したグラフを考える。この時、PageTrust では random walker に次の 3 つの制約を加える事で負のエッジの影響力を計算する。

- 負のエッジは walker が通れない道である
- 負のエッジ $(i, j) \in L^-$ を持つノード i に辿り着いた walker はノード j が信頼出来ないノードだという意見を持つ
- walker は自分が辿り着いたノードのことを信頼できるノードだという意見を持つ

ここからは図 2.2 を用いた具体例で説明する。時刻 0 でノード a にいた walker は、ノード d が信頼出来ないという意見を持ち、時刻 1 でノード b とノード c に均等に移動する。時刻 2 では、ノード c にいる walker のうちの $1/2$ が時刻 1 でノード a からやって来た walker だと仮定すると、ノード c にいる walker のうちの $1/2$ がノード d を信頼していないため、walker の移動が均等ではなくなり、ノード a と b に $2/5$ 、ノード d に $1/5$ の walker が移動する。このように walker に意見を持たせることで、ノード d に移動する walker の数を減らすことができる。これを続けると全ての walker がノード d を信頼しなくなるように見えるが、3 つ目の制約により、ノード d に辿り着いた walker はノード d が信頼できないという意見を忘れるため、全ての walker がノード d に移動しなくなるという状況を防ぐことができる。

2.3 最適化問題による手法

信頼度推定アルゴリズムとして、グラフの各ノード $i \in N$ の信頼度 c_i を最適化問題を解くことにより求める手法 [4] が存在する。これは、少量の事実 (確実に正しいノード) を予め設定してお

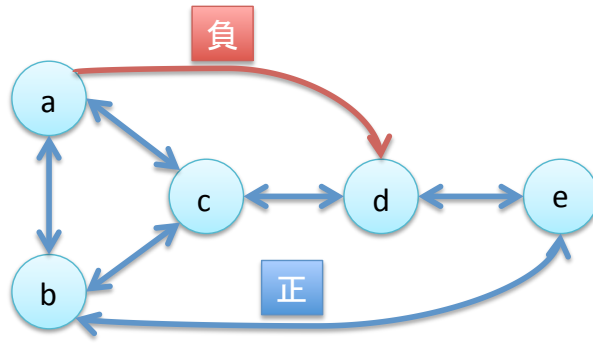


図 2.2: PageTrust に入力できるグラフの例

き、その事実情報を手がかりとして各ノードの信頼度を推定する手法である。具体的には、目的関数 $E(c)$ が最小になるように式 (2.7) を最適化問題として解き、信頼度 c を求める。

$$E(c) = \frac{1}{2} \sum_{i,j} |w_{i,j}| (c_i - s_{ij}c_j)^2 \quad (2.7)$$

$$s_{ij} = \begin{cases} 1 & (if\ w_{ij} \geq 0) \\ -1 & (if\ w_{ij} < 0) \end{cases} \quad (2.8)$$

ここで、信頼度 c は $[-1, 1]$ の値を取り、 $c = 1$ は信頼出来ることを示し、 $c = -1$ は信頼出来ないことを表している。 w_{ij} はグラフのエッジの重みで、同意関係のグラフには正の重み、反論関係のグラフには負の重みをつける。このアルゴリズムは式 (2.7) から分かるように、互いに同意し合うノードには同じような信頼度を割り当て、互いに反論し合うノードには符号の異なる信頼度を割り当てるという考え方を採用している。

2.4 分類器による手法

グラフ解析以外にも、分類器により信頼度を推定するアルゴリズム [3] も存在する。これは Twitter の信頼度を推定する手法で、教師あり機械学習により分類器を 2 つ作成し、入力ツイートに対して 2 つの分類器を使って分類することで信頼度推定を行う。作成する分類器は「入力ツイートが雑談のようなツイートなのか、ニュースのようなイベントが書かれたツイートなのかを判断するもの」と「入力ツイートが信頼できるのか信頼出来ないのか真偽不明なのかを判断するもの」の 2 つである。まず、入力データを一つ目の分類器を用いて雑談ツイートとイベントツイートに分類する。次に、雑談に分類されたツイートは破棄し、イベントに分類されたツイートを 2 つ目の分類器を用いて、信頼出来る、信頼出来ない、真偽不明の 3 つに分類する。これにより、各ツイートの信頼度を 3 値で推定することができる。

第3章 PageTrust の Twitter データへの適用

2章でも述べたが、本研究では確実に正しい情報の存在しない Twitter を対象として、各ツイートの信頼度を自動推定することを目的としているため、初期値や教師を必要としないアルゴリズムを用いる必要がある。そこで、Kerchove らの PageTrust アルゴリズム [2] に焦点を当てた。PageTrust は Web ページの信頼度を推定する有名なアルゴリズムである PageRank を、負のエッジが組み込めるように拡張したアルゴリズムであるが、実データによる検証と性能評価がまだ行われていない。そこで本研究では、PageTrust を Twitter データに適用させることで各ツイートの信頼度推定を行い、その性能を評価する。具体的には、PageTrust を Twitter データに適用させるためのグラフ作成手法を提案し、PageTrust を用いて計算した信頼度の高いツイートランキングを作成することで、性能を評価する。本研究の概要は図 3.1 に示す。

本章では、まず実データに PageTrust を適用した場合の計算方法に関する問題点を述べ、その解決方法を提案する。次に Web ページを対象としたアルゴリズムである PageTrust を Twitter に適用するためのグラフ作成手法について説明する。

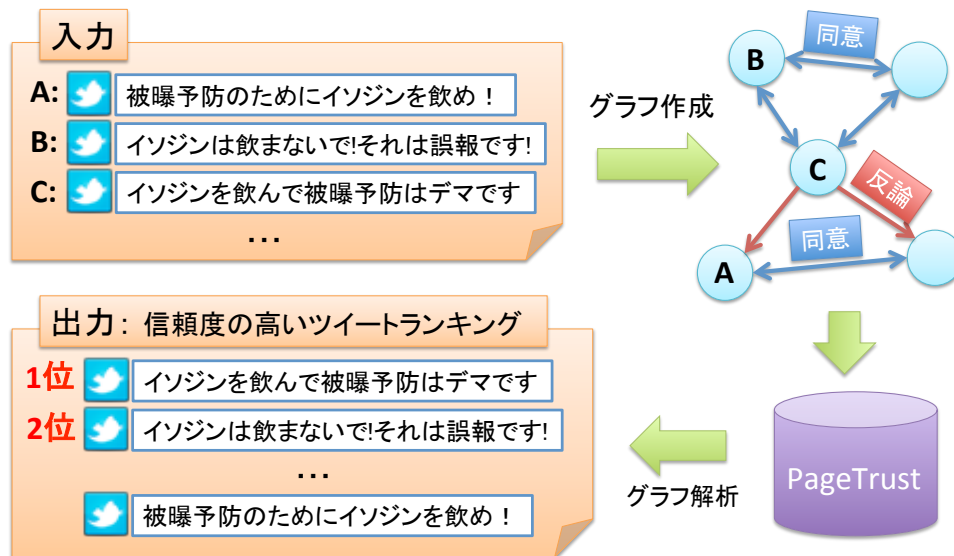


図 3.1: 本研究の概要

3.1 PageTrust の問題点

2章で述べたように、PageTrust では計算過程で反論行列 P と遷移行列 T の2つの行列を使用するが、反論行列 P は式 (2.6) で計算されるため、対角成分以外の全ての要素が0でない密な行列である。同様に、遷移行列 T も式 (2.5) から、一つ前の反復で計算した信頼度 $x_j^{(t)}$ を用いるため、全ての成分が0でない密な行列となる。ここで、行列 T, P は $n \times n$ の行列であるので、入力するデータ数に比例して行列を保持するメモリ量が爆発的に増加してしまう。また、式 (2.4) より $T \times P$ を計算する必要があるため、入力するデータ数に応じて計算時間も爆発的に増加してしまう。

3.2 PageTrust の改善

3.1節で述べたメモリと計算時間の問題点を改善するため、本研究では式 (2.5) のパラメータ M を $M = 0$ に固定するという手法をとった。 $M = 0$ に固定することによる影響は次のようなものが挙げられる。

- (1) 遷移行列 T の要素に $T_{ij} = 0, i, j \in N$ となる要素が追加される。
- (2) 反論行列 P の要素に $P_{ij} = 0, i, j \in N$ となる要素が増える。
- (3) $M = 1$ として負のエッジの影響力を高めることができなくなる。

(1) は式 (2.5) より、 $M = 0$ のとき遷移行列 T は正のエッジが張られた要素 $T_{ij}, (j, i) \in L^+$ のみ数値が計算され、正のエッジが張られていない要素 $T_{ij}, (j, i) \notin L^+$ は数値が0になることが分かる。(2) は式 (2.4), 式 (2.6) より、 $M = 0$ のとき反論行列 P は負のエッジが張られた要素 $P_{ij}, (i, j) \in L^-$ と正のエッジが張られた要素 $P_{ij}, (i, j) \in L^+$ の数値が計算され、それ以外の要素 $P_{ij}, (i, j) \notin L^+, L^-$ は0になることが分かる。これにより、プログラミング時に遷移行列 T と反論行列 P の0となる要素を取り除く事によって、メモリ使用量と計算時間を削減することができる。

(3) に関しては、負のエッジの影響力を変化させることが出来ないという大きなデメリットに見えるが、2.2節で説明したように、式 (2.3) のパラメータ β でも負のエッジの影響力を変化させることができるため、 $M = 0$ に固定しても大きなデメリットにはならない。

3.3 グラフ作成

本節では、入力データからグラフを作成する手法について説明する。グラフのノードには各ツイートデータ (ツイート本文と URL の組) を使用し、ツイート間の関係から同意のエッジと反論のエッジを張ることでグラフを作成する。ここで、同意エッジの集合を L^+ 、反論エッジの集合を L^- とすることで PageTrust に適用させる。また、信頼度推定の精度向上に必要なと思われる追加要素について説明する。

3.3.1 同意・反論エッジの張り方

Twitter 上では RT や QT、reply など、あるツイートに対して反応する機能があるが、それらの機能を用いて同意や反論を訴えるユーザーは少ないため、RT、QT、reply とは関係なく各ツイート間の同意・反論関係を判断する必要がある。そこで本研究では、「～はデマです」のような訂正ツイートに着目して同意・反論関係を判断する手法を用いた。例えば、表 3.1 のようなツイートを考える。

表 3.1: 訂正ツイートの例

放射能対策でヨウ素剤の代わりにイソジン 3 滴をコップ一杯の水に入れて飲み という全くのデマが流れているので止めてください。

表 3.1 のツイートには「全くのデマ」という記述があるため、このツイートがある情報を訂正するツイートであることが分かる。このとき表 3.1 のツイートが訂正しているのは、「という全くのデマ」の前の部分である「放射能対策でヨウ素剤の代わりにイソジン 3 滴をコップ一杯の水に入れて飲み」という情報である。つまり、表 3.1 のツイートは「放射能対策でヨウ素剤の代わりにイソジン 3 滴をコップ一杯の水に入れて飲み」という情報と反論関係にあるため、「放射能対策でヨウ素剤の代わりにイソジン 3 滴をコップ一杯の水に入れて飲み」という情報は正しいと記述しているツイートを検索し、マッチしたツイートと表 3.1 のツイートは反論関係であると判断することができる。また、同意関係に関しては、ツイートの内容がどれだけ似ていても「～はデマ」のような否定的な表現があるだけで完全に反対の内容となってしまうため、否定的な表現が含まれるツイートと含まれないツイートにグループ分けし、それぞれのグループ内で同じような内容をもつツイートが同意関係であると判断した。

以下では具体的なグラフの作成手法について説明する。グラフ作成手法の概要は図 3.2 に示す。また、訂正ツイートの判定には鍋島らの訂正パターンによる判定手法 [5] を用いた。

まず、鍋島らの訂正パターンによる判定手法 [5] を用いて訂正ツイートを検索する。この手法では正規表現を用いて表 3.2 のような表現を含むツイートを検索し、マッチしたものを訂正ツイートと判断する。検索では「というデマ」のように、表 3.2 に書かれた「接続パターン」+「訂正パターン」がツイートの本文内に存在するかどうかを調べる。この時、より多くのパターンとマッチさせるため、接続パターンと訂正パターンの間に 5 文字までどんな文字列が挿入されていてもマッチできるようにしておく。

次に、各ツイートを訂正カテゴリとその他カテゴリの 2 つに分類する。具体的には、正規表現でマッチしたツイート (以下、訂正ツイートとする) を訂正カテゴリに追加し、マッチしなかったツイート (以下、その他ツイートとする) をその他カテゴリに追加する。このとき訂正ツイートの場合は、表 3.1 の「放射能対策でヨウ素剤の代わりにイソジン 3 滴をコップ一杯の水に入れて飲み」のように、接続パターンの前部分 (以下、デマ情報とする) を抽出し、訂正ツイートの本文の代わりにデマ情報を訂正カテゴリに追加する。この操作を行うことで、訂正カテゴリに訂正ツイートが反論しているデマ情報を格納することができ、以下で説明する同意・反論のラベルが付けやすくなる。

さらに、2 つのカテゴリ内に存在する各ツイート間の類似度を計算し、類似度の高いツイート間

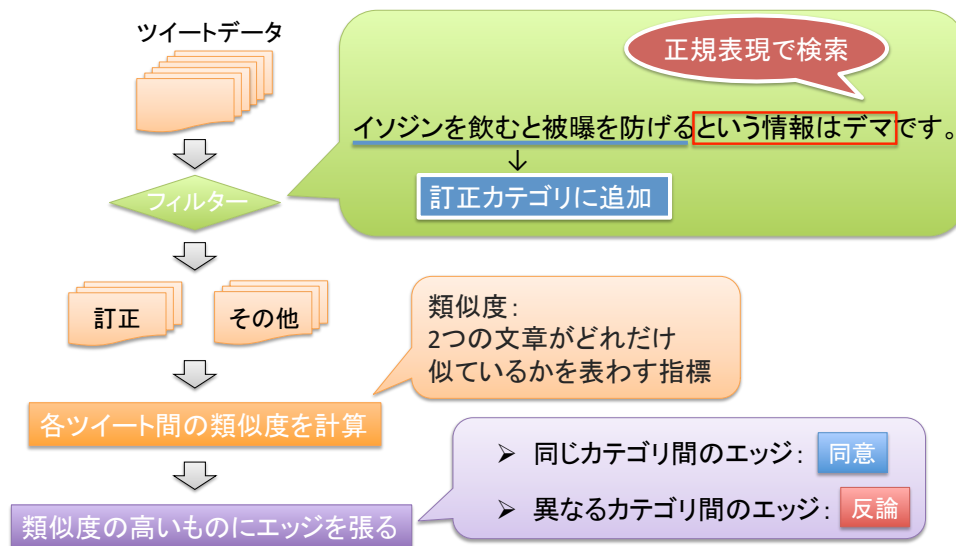


図 3.2: グラフ作成手法の概要

表 3.2: 正規表現リスト

接続パターン	訂正パターン
<p>は、的な、などの、などという、等という、との、とか、って、なんて、という、とかいう、っていう、とか言ってる、のような、の様な</p>	<p>デマ、誤報、誤り、誤情報、不確定情報、不確定な情報、嘘、虚偽、チェーンメール、事実はない、事実はありません、事実ではない、うそ、まちがい、ウソ、ガセ、誤った情報、誤解、真逆の情報、信じるな</p>

に同意もしくは反論のラベルをつける。類似度とは2つの文章がどれだけ似ているかを表わす指標で、本研究では文章に出現する単語のコサイン類似度を用いて計算した。類似度の高いツイート同士は内容が似ていると判断できるが、「放射能対策でヨウ素剤の代わりにイソジン3滴をコップ一杯の水に入れて飲め」というツイートと「放射能対策でヨウ素剤の代わりにイソジン3滴をコップ一杯の水に入れて飲めという情報はデマ」というツイートのように、内容が反論関係にあるものでも類似度は高くなってしまふ。そこで本研究では、同じカテゴリ内のツイート間で類似度が高かった場合は同意のラベル、異なるカテゴリ間で類似度が高かった場合は反論のラベルをつけた。同じカテゴリ内では、類似度を計算した2つのツイート両方に否定的な表現が含まれている、もしくは両方に否定的な表現が含まれていないため、類似度の高いツイート同士の内容は同意関係にあると判断できる。また、異なるカテゴリ間ではどちらか片方のツイートに否定的な表現が含まれ、もう片方には含まれていないため、類似度の高いツイート同士は反論関係にあると判断できる。これにより、先ほどの例のような否定的な表現の有無だけで内容が正反対になるツイートにも対応することができる。

最後に、先ほど求めた各ツイート間の同意・反論関係をグラフとして表現する。具体的には、同意・反論ラベルの付いているツイートをグラフのノードとして用い、同意ラベルの付いているツイート間に両向きの正のエッジ、反論ラベルの付いているツイート間には訂正ツイートからその他ツイートに向けて負のエッジを張る。この時、類似度は2つの文章がどれだけ似ているか、つまり2つのツイートがどの程度同意関係にあるかを表しているため、エッジの重みに類似度の値を用いた。この手法により作成されたグラフの例を図3.3に示す。

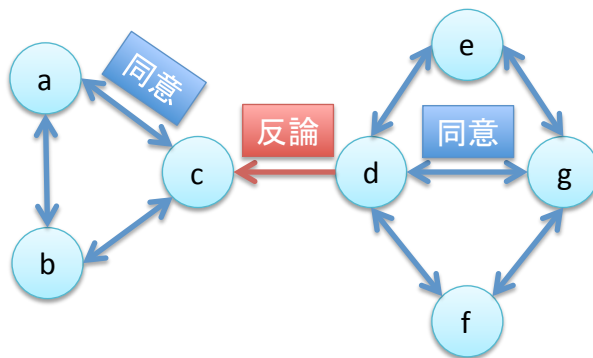


図 3.3: 作成したグラフの例

3.3.2 仮ノードの追加

PageTrust は入力として、図2.2のように負のエッジが存在するグラフを入力することができる。これは、負のエッジは通れない道と定め、負のエッジ $(i, j) \in L^-$ が張られたノード i にたどり着いた walker はノード j は信頼出来ないノードだという意見を覚えて移動を続けるからである。しかし、図3.3のようにお互いに同意関係にあるグループ間に反論のエッジを貼った場合は、同意関係にあるノード間だけで walker が移動し、信頼出来ないノードという情報を持っている walker が反対側のグループに移動することができない。つまり、図3.3のようなグラフを PageTrust に入力した場合、反論のエッジが張られていないのと同じような動きをしてしまうと考えられる。そこ

で本研究では、この現象を防ぐために新たなノードとして仮ノードを追加した。仮ノードは全てのノードに対して重み $1/n$ (n は仮ノードを含まないノードの総数) の同意エッジが張られたノードで、図 3.3 のグラフに仮ノードを追加したグラフを図 3.4 に示す。これにより、図 3.3 では左右の同意グループ間を walker が移動出来なかったのに対し、図 3.4 ではグループ間を仮ノードを経由することで walker が移動できるようになるため、右のグループでノード c が信頼出来ないという情報を持った walker が左のグループにも移動することができる。そのため、信頼出来ないノードの情報を持つ反論エッジが正しく作用されると考えられる。また、仮ノードを追加することによるデメリットとしては、仮ノードは全てのノードに均等に同意のエッジが張られているため、仮ノードの信頼度が高くなり、他のノードの信頼度が下がってしまうという点が考えられる。しかし、今回は信頼度の絶対値ではなく他のノードとの相対値を用い、信頼度の高いツイートランキングを作成するので、この欠点は無視することができる。

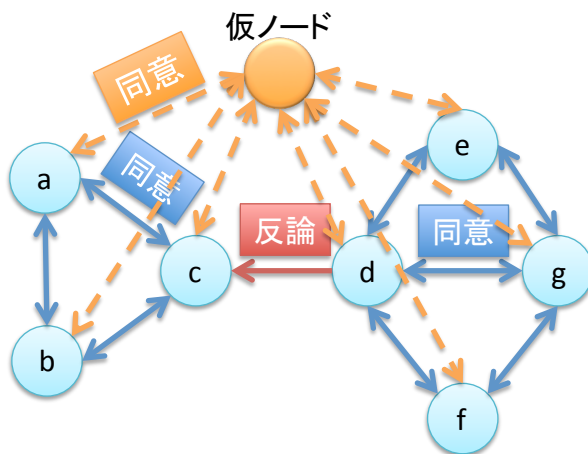


図 3.4: 仮ノードを追加したグラフの例

3.3.3 ソース URL の追加

作成したグラフに新たなノードとしてソース URL を追加した。ソース URL とはツイートの本文に記載された情報の提供元となる URL のことで、表 3.3 のようなものが対象となる。

表 3.3: ソース URL を含むツイートの例

<p>【拡散希望】コスモ石油の爆発で有害物質の雨が降る件はデマ。 / コスモ石油が否定「火災で有害物質降る」のメール連鎖 http://ow.ly/4cYQ9</p>

また、ソース URL を追加したグラフの例を図 3.5 に示す。図 3.5 に示すように、ソース URL をグラフに追加することで同じソース URL が記載されたツイート同士の結びつきが強くなるのが分かる。そのため、ソース URL が記載されたツイートの信頼度が高くなることが予想される。また、一般的にソース URL が記載されたツイートは、記載されていないツイートに比べ信頼度が高いため、ソース URL を追加することで信頼度推定の精度を向上させることができると考えられる。

さらに、ソース URL を追加することで各ツイートの信頼度だけでなく、ソース URL の信頼度も同時に推定することができるというメリットも存在する。今回はどの Web ページから情報を得ているのが重要と考え、細かいページではなくドメイン名までをソース URL としてグラフに追加した。

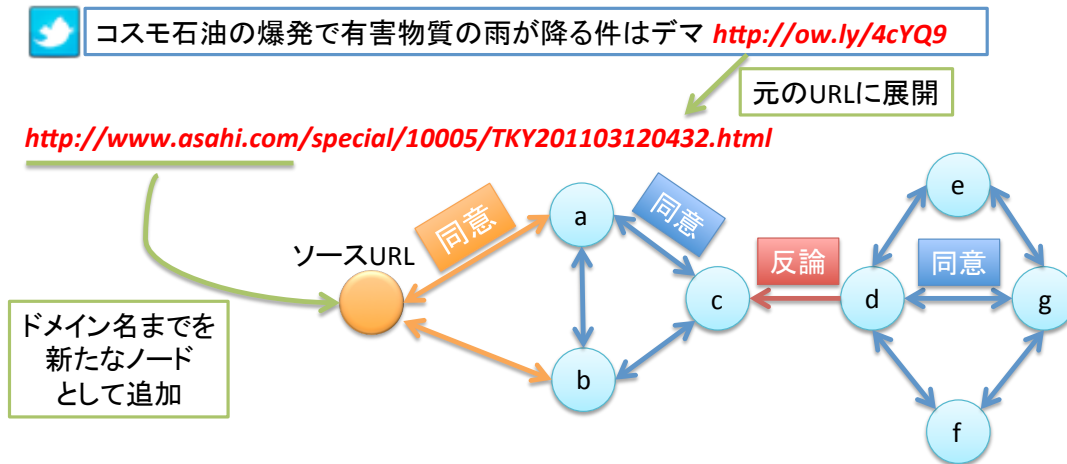


図 3.5: ソース URL を追加したグラフの例

具体的な追加方法としては、まず正規表現を用いてツイート本文からソース URL を抽出する。次に、表 3.3 の例でも分かるように、Twitter においてソース URL を記載する場合は短縮 URL を用いていることが多いため、短縮 URL を元の URL に展開する。最後に、展開した URL のドメイン部分のみを新たなノードとしてグラフに追加し、ソース URL が記載されたツイートとの間に同意のエッジを貼った。

第4章 実験

4.1 実験設定

実験に用いる入力データは東日本大震災時(2011/3/11 2011/3/29)のツイートのうち、「コスモ石油」を含むツイート 72887 件と、「イソジン」を含むツイート 24883 件を使用した。計算時間の膨大と、雑談のような真偽判断のつかないツイートを取り除くため、「コスモ石油」の場合は RT 数 50 以上、「イソジン」の場合は RT 数 10 以上という制限を加えた。実験は、先ほど述べた「コスモ石油」と「イソジン」に関するツイートを入力し、算出された PageTrust のスコアが高いツイートランキングを作成することで行う。また、仮ノードとソース URL を追加した場合と追加しなかった場合で実験結果を比較し、仮ノードとソース URL を加えることによる影響力を調べた。

ベースラインとしては、各ノードに張られた同意・反論のエッジの数をそれぞれ求め、(同意エッジの数) - (反論エッジの数) の値が高いツイートランキングを作成する。

正解データは、入力データに人手で正解、デマ、その他の 3 つのラベルをつけたものを用意した。正解のラベルは実際に起こったイベント等、一般的に正しいと考えられる内容のツイートであることを表わし、デマのラベルは実際に起こったイベントとは異なる事象のように、一般的に間違っていると判断できる内容のツイートを表わす。また、その他のラベルは雑談やユーザーの個人的な意見のように一般的には真偽を判別することができないツイートを表わす。

4.2 評価方法

信頼度の高いツイートランキング上位 m 件について *precision* と *recall* を計算し、しきい値 m を変化させながら PR 曲線を描くことによって評価を行う。*precision* と *recall* は次式で計算する。

$$precision = \frac{\text{上位 } m \text{ 件までに含まれる正解ツイート数}}{m} \quad (4.1)$$

$$recall = \frac{\text{上位 } m \text{ 件までに含まれる正解ツイート数}}{\text{正解データ内の全正解ツイート数}} \quad (4.2)$$

ソース URL を追加した場合は、入力するグラフのノード数がソース URL を追加していない場合よりも多くなってしまうため、ソース URL を追加した場合に増えたノードを無視して評価を行う。

4.3 実験結果

「コスモ石油」に関するデータセットで実験を行い、仮ノードとソース URL を追加した場合としなかった場合それぞれについて PR 曲線を描いた結果を図 4.1 に示す。また、「イソジン」に

関するデータセットで同様の実験を行った結果を図 4.2 に示す。図 4.1、4.2 における「URL」とはソース URL のことを指し、「URL あり」でソース URL をグラフに追加したことを表わす。また、図 4.1、4.2 のグラフは全て一番下の点のしきい値が $m = 5$ であり、下から上に向かってしきい値 m を 5 ずつ増加させながら *precision* と *recall* を計算している。今回作成した PR 曲線は一般的な PR 曲線とは異なる軌跡を描いているが、本研究では信頼度の高いツイートランキング上位 m 件について、*precision* と *recall* をそれぞれ式 (4.1), (4.2) を用いて計算するという独特な手法をとっているため、このような不自然な曲線になっている。例えば、図 4.1 のベースラインを見てみると、*recall* は式 (4.2) より、しきい値 m が増加するにつれ大きくなっていくのでグラフは下から上に向かって上がり続けていることが分かる。また、*precision* は式 (4.1) よりランキング上位 m 件の中の精度を計算しているため、激しく増減している。特に *recall* が 0.20 ~ 0.25 の範囲で *precision* が大きく下がっているが、これはランキング上位 10 ~ 15 件の範囲に存在するツイートが全てデマツイートであったためである。

図 4.1, 4.2 より、「コスモ石油」に関するデータセットでは仮ノードを追加した場合に信頼度推定の精度が大きく向上しているのに対し、「イソジン」に関するデータセットでは仮ノードを追加してもあまり精度が向上していないことが分かる。またソース URL については、追加することで「コスモ石油」に関しては精度が下がっているのに対し、「イソジン」に関しては精度が少し上がっていることが分かる。この違いは「コスモ石油」と「イソジン」の同意・反論関係の判別精度の差によるものと考えられるが、これについては 4.4 節で詳しく説明する。

4.4 分析

まず、仮ノードをグラフに追加することによる影響を考察する。図 4.1 より、「コスモ石油」に関するデータセットでは、仮ノードを追加することによって信頼度推定の精度が大きく向上していることが分かる。これは、3.3.2 節で述べたように図 3.3 のようなグラフを入力した場合、右のグループでノード c が信頼出来ないという情報を持った walker が左のグループに移動することが出来ないため、PageTrust では反論のエッジが張られていない状態と同様の動きをしてしまうという問題を、全てのノードにエッジが張られた仮ノードを追加したことで解決し、反論エッジの影響力を正しく反映できたからだと考えられる。一方「イソジン」に関するデータセットの場合、図 4.2 より、仮ノードを追加しても精度があまり向上していない。これは、同意・反論関係の判別精度の差によるもので、「コスモ石油」では同意・反論関係が高精度で判別出来ていたのに対し、「イソジン」では判別精度が低くなってしまった。そのため、正しいグラフを作成することができず、精度の向上が見られなかったと考えられる。また、「イソジン」に関するデータセットで同意・反論関係の精度が低くなってしまった原因としては、訂正パターンが多様過ぎたことが挙げられる。「コスモ石油」に関するツイートでは、訂正ツイートの多くが

- コスモ石油の爆発によって有害物質の雨が降るとするのは全くのデマです。
- 千葉市近辺で有害物質の雨が降るとするのは誤情報です。

のように直接的な否定表現を用いて書かれているのに対し、「イソジン」に関するツイートでは、

- イソジンを飲むバカ

コスモ石油

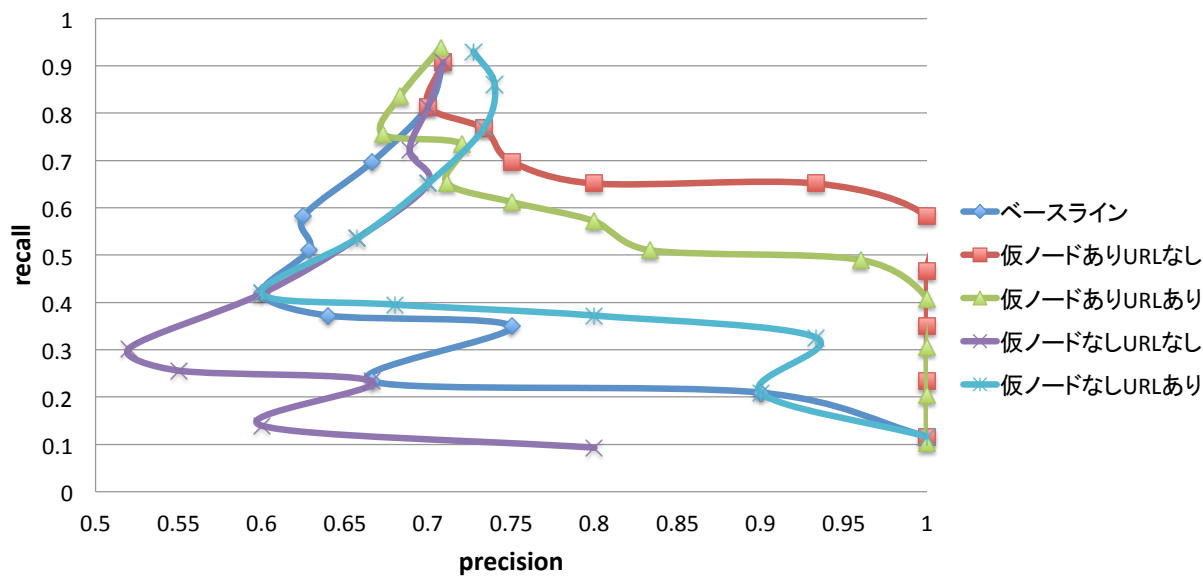


図 4.1: 実験結果: コスモ石油

イソジン

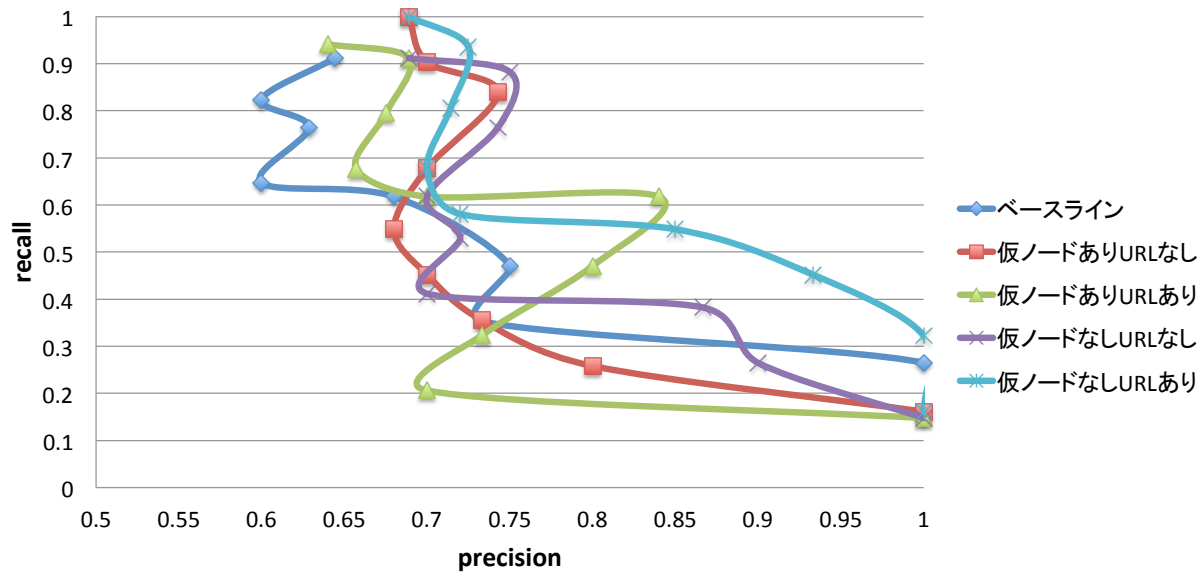


図 4.2: 実験結果: イソジン

- イソジンを飲んで被曝予防しようという情報を流した人は猛省すべき
- イソジンは体に悪影響を及ぼします

のように間接的な否定表現が多かったため、正規表現ではうまく同意・反論関係を判別することが出来なかった。以上より、PageTrust を用いた信頼度推定においてグラフに仮ノードを追加することは、同意・反論関係の判別精度が高い場合にのみ有用であると考えられる。そのため、今後の課題としてより高精度な同意・反論関係の判別方法を取り入れる必要がある。

次に、ソース URL を入れることによる影響を考察する。図 4.1 より「コスモ石油」に関するデータセットでは、仮ノードあり URL ありの場合は信頼度推定の精度が下がり、仮ノードなし URL ありの場合には精度が向上していることが分かる。これは、図 3.3 の左右のグループで同じソース URL が使われていたことが原因として挙げられる。仮ノードなしの場合には、ソース URL が図 3.4 の仮ノードと同様の働きをすることで反論のエッジの影響力を PageTrust に反映させることが出来き、精度が向上したと考えられ、仮ノードありの場合にはすでに左右のグループ間を移動するための道があるのにもかかわらず、ソース URL により道が追加されてしまったため、反論のエッジの影響力が小さくなり信頼度推定の精度が下がったと考えられる。また、図 4.2 より、「イソジン」に関するデータセットでも同様の結果が得られたが、「コスモ石油」の場合と比べ、「イソジン」の方がソース URL を加えた場合の精度の上がり方が大きく、仮ノードありの場合でもソース URL を追加した方が精度が高くなっている部分が存在する。これにより、PageTrust を用いた信頼度推定においてソース URL を追加することは、同意・反論関係の判別精度が低い場合に、有用であると考えられる。

第5章 まとめ

本論文では、Web ページの信頼度推定アルゴリズムである PageRank に、負のエッジを組み込めるように拡張した PageTrust の実データ (Twitter データ) による検証と性能評価を行った。これにより、PageTrust は入力するグラフの作成方法を工夫し、同意・反論関係を正しく反映させることで、実データにも適用できることが分かった。また、信頼度推定の性能も同意された回数と反論された回数を数えるというベースラインよりも高い精度を出すことが出来た。

グラフには追加要素として仮ノードとソース URL を取り入れた。仮ノードは図 3.3 のように、お互いに同意し合っているグループ間に反論のエッジのみを張るという特殊なグラフにおいて、PageTrust が反論のエッジの影響力を正しく反映しないという問題を解決し、信頼度推定の精度を向上させることが出来た。しかしながら、仮ノードが PageTrust において有用であるのは同意・反論関係の判別精度が高い場合だけであるため、同意・反論関係の判別精度を高めることが重要となる。ソース URL に関しては、ソース URL のみを追加した場合には信頼度推定の精度を向上させることができたが、仮ノードと同時に追加すると信頼度推定の精度を下げてしまうことが分かった。また、「イソジン」に関するデータセットの実験結果より、ソース URL を追加することで、同意・反論関係の判別精度が低い場合にグラフ内の同意・反論関係のミスを修正し、PageTrust による信頼度推定の精度を向上させることができるのではないかという可能性が浮上した。

今後の課題としては、多様な否定表現にも対応できる高性能な同意・反論関係の判別手法を取り入れることが挙げられる。今回は正規表現を用いた簡単な手法で同意・反論関係を判別したが、正規表現による手法では限界があり、限られたドメインでしか高精度の判別を行うことが出来なかった。同意・反論関係の判別精度は、PageTrust による信頼度推定の精度に直結するため、今後はより高精度の判別手法を実装し、グラフ作成時に使用していきたい。また、今回は2つのデータセットでしか実験を行うことが出来なかったため、できるだけ多くのデータセットを作成し、同様の実験を行うことで、仮ノードとソース URL を追加した場合の振る舞いをより詳しく調べていきたい。

謝 辞

本研究を進めるにあたり、適切なご指導をして下さった乾健太郎教授、岡崎直観准教授に深く感謝いたします。

日常的な議論や研究会での議論を通じて、多くの知識とアドバイスをして下さった乾・岡崎研究室の皆さんに感謝いたします。

参考文献

- [1] Lawrence Page and Sergey Brin and Rajeev Motwani and Terry Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report, 1999.
- [2] De Kerchove, Cristobald and Dooren, PV. The PageTrust algorithm: how to rank web pages when negative links are allowed. Proc. SIAM Int. Conf. on Data Mining, 346–352, 2008.
- [3] Castillo, Carlos and Mendoza, Marcelo and Poblete, Barbara. Information credibility on twitter. Proceedings of the 20th international conference on World wide web, 675–684, 2011.
- [4] Yin, Xiaoxin and Tan, Wenzhao. Semi-supervised truth discovery. Proceedings of the 20th international conference on World wide web, 217–226, 2011.
- [5] 鍋島啓太, 水野淳太, 岡崎直観, 乾健太郎. マイクロブログからの誤情報の発見と集約. 言語処理学会 第 19 回全国大会 発表論文集, 2013.
- [6] Kleinberg, Jon M. Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM), 604–632, 1999.
- [7] Yin, Xiaoxin and Han, Jiawei and Yu, Philip S. Truth discovery with multiple conflicting information providers on the web. Knowledge and Data Engineering, IEEE Transactions on, 796–808, 2008.
- [8] Gupta, Manish and Zhao, Peixiang and Han, Jiawei. Evaluating Event Credibility on Twitter. SIAM International Conference on Data Mining (SDM 2012), 2012.
- [9] Pasternack, Jeff and Roth, Dan. Knowing what to believe (when you already know something). Proceedings of the 23rd International Conference on Computational Linguistics, 877–885, 2010.