

# ソーシャルメディア上の発言と ユーザー間の関係を利用した批判的ユーザーの抽出

高瀬翔<sup>†</sup> 村上明子<sup>‡</sup> 榎美紀<sup>‡</sup> 岡崎直観<sup>†§</sup> 乾健太郎<sup>†</sup>  
東北大学<sup>†</sup> 日本アイ・ビー・エム株式会社 東京基礎研究所<sup>‡</sup>  
科学技術振興機構 さきがけ<sup>§</sup>

{takase, okazaki, inui}@ecei.tohoku.ac.jp {akikom, enomiki}@jp.ibm.com

## 1 はじめに

ソーシャルメディア上には、「A社の手口は本当に酷い」など、日常的に特定の個人や企業に対する非難を行うユーザーが存在する。このようなユーザーを認識することは、悪評を広めようとしているユーザーの認識につながるため、風評被害の防止に有用である。本論文では、特定の個人や企業、団体などの実体に対して日常的に非難を行うユーザーを批判的ユーザーと呼び、ソーシャルメディア上の批判的ユーザーを自動で抽出する手法を提案する。すなわち、特定の实体の名前と、ソーシャルメディア上でのその実体のアカウントを入力とし、その実体に対する批判的ユーザーを抽出する。

あるユーザーが入力の実体に対する批判的かどうかは、ユーザーの発言内容を検証することで判定できる。特定の实体を非難する発言かどうかは、人間にとっては判断が容易な場合が多い。しかしながら、批判的ユーザーの中には揶揄を行うユーザーもあり、計算機による非難の判断は難しい。例えば、対象の実体は何らかのミスをした際、それに対し「さすがですね」などの発言で非難を行うことがある。また、ソーシャルメディア上の発言には省略が多く、非難の対象が発言中で明示されていないことがある。例えば、あるユーザーの「A社の新製品が使いづらい」という発言に対し「馬鹿だから」という発言が行われたとする。この場合「使いづらい新製品を出すA社が馬鹿」と言っているのか、「A社の新製品を使いづらいと言うユーザーが馬鹿」と言っているのかの判定が難しい。

本論文では、言語的な解析にもとづく評判分析 [9, 10] に加え、ユーザー間の結びつきを利用し、言語的な解析だけでは判定が難しい批判的ユーザーも発見できる手法を提案する。ソーシャルメディア上では、同じ意見のユーザーが結びつきを持つ傾向がある [2, 7]。そこで、本研究ではユーザー間の結びつきをグラフで表現し、言語的な解析で抽出した批判的ユーザーを利用したグラフ解析を行う手法を提案する。実験では、ソーシャルメディアのひとつであるツイッターにおける、特定の实体に関する発言から批判的ユーザーの抽出を行う。実験結果によって、言語的な解析では多くの批判的ユーザーを取りこぼすこと、ユーザー間の結びつきを利用することで批判的ユーザーの抽出精度が向上することを示す。

## 2 関連研究

本研究は、オンラインコミュニティ上での評判分析 [1, 4, 5] と関連が深い。O'Connor らはソーシャルメディア

上の発言に対し、極性辞書を用いた単純な評判分析を行い、世論調査による大統領の支持率など、実社会との相関が見られることを示した [4]。Sperious らは極性辞書や分類器の極性判定結果などとソーシャルメディア上でのユーザー間の結びつきにもとづくグラフを用い、発言の極性判定精度を向上させる手法を提案した [5]。これらの研究は個々の発言の極性判定を目的としており、ユーザーの極性判定を行うわけではない。また、各発言がどの実体に対して極性を表明しているのか解析していない。Conover らはソーシャルメディアにおけるユーザー間の結びつきや使用するタグを元に、ユーザーの政治的な極性を予測する手法を提案した [1]。彼らは政治を対象としているが、本研究は対象とする実体を制限しておらず、より一般的である。

特定の实体を非難する発言を日常的に行うユーザーを抽出することは、スパマーの発見 [6, 7] と関連する。Wang は結びついているユーザー数などを素性にした教師あり学習によって、スパマーを判定する分類器を作成した [6]。Yang らはスパマーのコミュニティを分析することで、スパマーはソーシャルメディア上で、互いに密接な結びつきを持つことを明らかにし、少数のスパマーを入力として、スパマーを抽出する手法を提案した [7]。Wang [6] や Yang ら [7] はユーザー間の結びつきという、グラフ的な特徴を利用しているが、誰が何に対してどのような意見なのかという言語的な解析は行っていない。

## 3 提案手法

提案手法の概要を図1に示す。提案手法ではまず、ユーザーの発言を言語的に解析することで、入力として与えられた実体を非難する傾向の強いユーザーを批判的ユーザーとして抽出する。次に、ユーザーをノード、ソーシャルメディア上での結びつきをエッジとしたグラフを作成し、言語的な解析で抽出した批判的ユーザーと強く結びついているユーザーを批判的ユーザーとして抽出する。本節では、言語的な解析手法、およびその結果を用いたグラフ解析の手法について、それぞれ述べる。

### 3.1 言語的な解析

批判的ユーザーとは、入力として与えられた実体に対し、非難する発言を日常的に行うユーザーである。提案手法では、ユーザーの投稿した発言について、入力の実体を非難するものであるかどうかを判定する。実体を非難する発言を非難発言と呼び、非難発言の回数の多いユーザーを批判的ユーザーとして抽出する。

ソーシャルメディア上のユーザー間のやりとりでは、

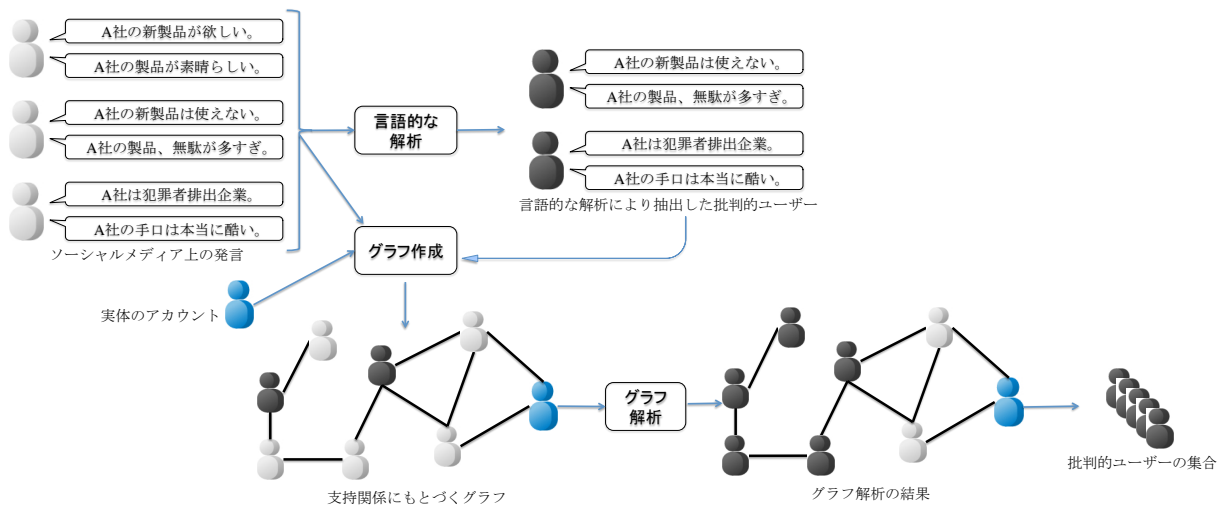


図 1: 提案手法の概要

省略が頻繁に発生し、さらに、文脈も失われることが多い。そこで、ユーザー間の直接的なやりとりは解析対象から除外する。すなわち、ツイッターにおいて、あるユーザーへの応答であるリプライ（“@ユーザー名 発言”）、あるユーザーの発言の引用である非公式リツイート（“引用への言及 RT @ユーザー名: 引用した発言”）、あるユーザーの発言の再投稿であるリツイート（“RT @ユーザー名: 発言”）はここでの言語的な解析の対象としない。

ある発言が特定の実体を非難するものかどうかは、非難表現とその対象を調べることで判定する。非難表現とその対象は、統語的な関係で結ばれているとする。提案手法では、金山らの評価表現抽出エンジン [10] における、不評タグの付与された表現を非難表現とした。金山らの評価表現抽出エンジンは、構文解析と格フレームから評価対象を特定できる。しかしながら、ソーシャルメディア上には、格が明記されていない発言も多く、金山らの評価表現抽出エンジンでは対象を抽出できない場合がある。これに対処するため、金山らの評価表現抽出エンジンの抽出した対象に加え、非難表現を含む文節と係り受け関係で距離が 2 以下の文節に含まれる単語も非難の対象とする。なお、このとき金山らの評価表現抽出エンジンにおける不評タグに加え、ソーシャルメディア上に多く見られる軽卑表現も非難表現として扱う。軽卑表現の判定には、荻野らの軽卑表現辞書 [9] を用いた。

例えば、図 2 のような文について考える。この図では、文は文節に区切られており、文節間の係り受け関係は矢印によって表されている。また、四角で囲っている「酷い」という単語が非難表現である。この場合「酷い」から係り受け関係で距離が 2 以下の文節を非難の対象とするため「手口は」と「A 社の」が「酷い」の対象となる。したがって、この文は A 社を非難する発言であると判定する。なお、発言の係り受け解析には KNP [3] を用いた。

ある実体に対する批判的ユーザーの発言には、非難発言が多いと考えられる。したがって、ある実体についてユーザーが発言した総数に対する、非難発言の比を計算し、しきい値を超えたユーザーを批判的ユーザーとする。

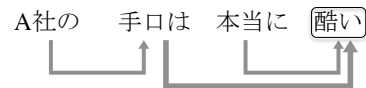


図 2: 非難発言の例

すなわち、ユーザー  $i$  の発言のうち、ある実体に対する発言の総数が  $N_i$ 、非難発言の数が  $n_i$  であるとき、以下の式により、ユーザーの批判的ユーザーらしさ  $s_i$  を計算する。

$$s_i = \frac{n_i}{N_i} \quad (1)$$

この  $s_i$  がしきい値を超えたユーザーを批判的ユーザーとして抽出する。

### 3.2 グラフ解析

揶揄や省略が発生している場合には、言語的な解析により非難発言と認識するのが難しい。そのため言語的な解析だけでは、多くの批判的ユーザーを取りこぼしてしまう。これに対処するため、提案手法では、ソーシャルメディア上でのユーザー間の結びつきをグラフで表し、言語的な解析結果をシードとしたラベル伝播を行う。

ソーシャルメディア上では、ユーザー間で議論が生じること多い。このため、ユーザー間の結びつきについて、支持しているのか、反対しているのかという、ユーザー間の関係を考慮する必要がある。ソーシャルメディアにおいて、ユーザーは自分の支持している情報が他人にも伝わるよう、その拡散に寄与すると考えられる。したがって、あるユーザーの発言に対して、その拡散に寄与する行為を行ったユーザーは発言の投稿者を支持しているユーザーと考えられる。ツイッターでは、あるユーザーの発言を再投稿するリツイートが、拡散への寄与と考えられる。Conover らはリツイートをを行ったユーザー間にエッジを張ることにより、同じ意見のユーザー同士が結びついたグラフが作成できることを示している [2]。このことから、リツイートが支持関係を表すと考えられる。

あるユーザーの発言を引用する非公式リツイートも、情報の拡散への寄与の可能性はある。しかし非公式リツイートにおいては、“こんなものありえない RT @hoge:

新製品できました” というように、引用した発言を非難する場合がある。このようなユーザー間のやりとりでは、3.1 節で述べたように、省略が多い<sup>1</sup>ため非難かどうかの解析が難しい。このため提案手法では、リツイートによって表される支持関係のみを利用してグラフを作成する。

$m$  人のユーザー  $U = \{u_1, \dots, u_m\}$  をノードとし、 $u_i$  と  $u_j$  間には支持の度合いに応じて、重み  $w_{ij}$  ( $0 \leq w_{ij} \leq 1$ ) でエッジを張る。 $u_i$  が  $u_j$  の発言をリツイートした回数を  $r_{ij}$ 、 $u_i$  の総リツイート回数を  $R_i$  としたとき、重み  $w_{ij}$  を以下の式によって計算する。

$$w_{ij} = \frac{1}{2} \left( \frac{r_{ij}}{R_i} + \frac{r_{ji}}{R_j} \right) \quad (2)$$

すなわち、重み  $w_{ij}$  は  $u_i$  と  $u_j$  が互いの発言をリツイートする回数が多いほど 1 に近くなり、互いの発言をリツイートする回数が少ないほど 0 に近くなる。また、 $w_{ij}$  は対称性があり、 $w_{ij} = w_{ji}$  である。

言語的な解析によって抽出した批判的ユーザーとエッジの重み、さらに実体のアカウントを利用し、グラフのノード  $U = \{u_1, \dots, u_m\}$  について、批判的ユーザーと考えられる度合い（確信度）を付与する。各ユーザーの確信度の値は  $c = (c_1, \dots, c_m)$  として表す。本研究では確信度は、-1 から 1 までの実数とする。批判的ユーザーと考えられるノードほど 1 に近い値を取り、批判的ユーザーとは異なると考えられるノードほど -1 に近い値を取るとする。なお、言語的な解析によって抽出した批判的ユーザーの確信度は 1 とし、実体のアカウントは確信度 -1 に固定する。すなわち、これらのユーザーの確信度をシードとして、極性の分からないユーザーの確信度をラベル伝播で求める。

ラベル伝播では、Zhu ら [8] と同様に、次の目的関数を最小化する。

$$E(c) = \frac{1}{2} \sum_{i,j} w_{ij} (c_i - c_j)^2 \quad (3)$$

$E(c)$  を最小化するためには、 $w_{ij} > 0$  のとき  $c_i$  と  $c_j$  が近い値であれば良い。すなわち、支持関係にあるユーザーの確信度が近い値であれば良い。 $E(c)$  を最小化することにより、言語的な解析とユーザー間の支持関係を考慮した確信度を求めることができる。得られた確信度について、しきい値を超えたユーザーを批判的ユーザーとする。

グラフ上ではあるユーザーと直接支持関係にあるユーザーほど同意見である可能性が高く、遠いユーザー間には同意見である可能性が低くなる。しかしながら、言語的な解析によって抽出した批判的ユーザーを含み、かつ実体のアカウントの存在しない部分グラフについては、すべてのノードの確信度を 1 とすることが  $E(c)$  を最小化する解であり、ユーザー間の距離は考慮されなくなってしまう。シードとして確信度を固定したノードからの距離を考慮し、確信度を減衰させるために、Yin らは中立ノード<sup>2</sup>を導入した [8]。本研究もこれと同様に、中立ノードを導入する。

<sup>1</sup>実際の引用への言及としては、“ありえない RT @hoge: 新製品できました” のような場合が多い

<sup>2</sup>シード以外の全ノードと結びついている、確信度 0 のノード

## 4 実験

### 4.1 実験設定

実験に用いるデータは、ツイッターのサーチ API を利用して収集した。サーチ API への入力クエリは対象とする実体と、その実体のツイッター上でのアカウント名とした。今回対象とする実体は、ツイッター上に批判的ユーザーが多数見られるある会社とした。この会社は社長がツイッター上にアカウントを持っており、これを実体のツイッター上でのアカウントとした。クエリをサーチ API に入力し、その検索結果のうち日本語ツイートのみを一ヶ月間収集した。

実体についての発言が盛んなユーザーを絞り込むため、実体を含む発言が 4 回以下のユーザーを除去した。さらに、自動投稿する bot を対象から除去するため、ユーザー間でのやりとりが 1 回以下のユーザーを除去した。残ったユーザーのうち、発言の多い順に 300 人を実験対象とした。なお、実験対象のユーザーの発言の総数は約 3 万件であった。

提案手法の精度を測定するため、実験対象としたユーザーに批判的ユーザーであるかそうでないかを人手で付与し、正解データを作成した。データを収集した一ヶ月間における各ユーザーの実体を含む発言を調べ、実体に対する悪評や非難を積極的に発言していると思われるユーザーを批判的ユーザーとした。この際、対象にとって都合の悪い報道や非難発言を多く再投稿しているユーザーも、悪評を広めようとしていると考えられるため、批判的ユーザーとした。なお、対象に関する報道を内容に関わらず再投稿するなど、一時的に批判的な発言をしているだけと考えられるユーザーは批判的ユーザーとしていない。この基準の妥当性を検証するため、実験対象としたユーザー 300 人のうち 50 人をランダムに抽出し、二人で批判的ユーザーの判定を行ったところ、カッパ係数は 0.90 となった。したがって、この基準による批判的ユーザーかどうかの判定は妥当であると言える。50 人の判定を行った二人のうち一方が、実験対象とするユーザー 300 人に対し、批判的ユーザーであるかそうでないかを付与した結果を、正解データとした。正解データにおける批判的ユーザーは 300 人中 117 人であった。

実験では、言語的な解析手法とグラフ解析手法の組み合わせによる効果を検証する。このため、言語的な解析のみの手法、グラフ解析のみの手法および言語的な解析とグラフ解析の組み合わせ（提案手法）を比較する。

### 4.2 結果

各手法において、しきい値を変化させたときの適合率と再現率を図 3 に示す。言語的な解析ではしきい値を 0 から 0.1 までの範囲で変化させ、グラフ解析では 0.5 から 0.7 までの範囲で変化させて適合率と再現率を計算した。言語的な解析のみの手法は破線で、グラフ解析のみの手法は点線で、提案手法は実線でそれぞれ記されている。提案手法では、言語的な解析において適合率と再現率が共に高い値となっている、しきい値が 0.06 のとき（図 3 において、丸で囲った点）の結果をグラフ解析のシードとした。なお、その際のユーザー数は 42 人であった。提案手法における、言語的な解析の結果をグラフ解析のシードとすることの効果を確認するため、グラフ解析のみの手法との比較を行う。グラフ解析のみの手

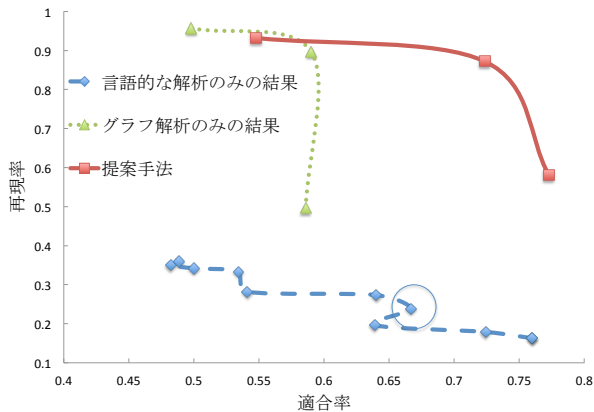


図 3: 各手法における適合率と再現率

法では、言語的な解析結果と同数のユーザーをランダムに抽出し、批判的ユーザーとしてグラフ解析へのシードとした。

図 3 より、言語的な解析のみの手法では、高い適合率を達成できるが再現率が上がりにくい。反対に、グラフ解析のみの手法では、高い再現率を達成できるが適合率が上がりにくいことがわかる。提案手法は言語的な解析のみの手法と同程度の適合率で、高い再現率を達成できている。すなわち、言語的な解析のみの手法の適合率を維持したまま、より多くの批判的ユーザーを抽出できている。これにより、言語的な解析手法とグラフ解析手法の組み合わせは効果的であると言える。

言語的な解析では再現率が低いことが図 3 より言える。これは、今回利用した非難表現をドメインに依存しない、一般的なものとしたことにより、「マスゴミ」や「反日」などドメインに特有の表現による非難発言を、非難と判定できないためであると考えられる。また、ツイッターは発言の文字数が 140 文字と制限されているため、非難する文書を書いたブログ記事へのリンクを張り、悪評を広めようとするユーザーも存在する。これはツイッターへの投稿発言からだけでは判定することができない。なお、図 3 において、言語的な解析のみの結果は適合率 0.65 付近で一度下がった後に上昇している。これは、出力数が少ないため、わずかな違いが適合率に与える影響が大きいためである。

提案手法は言語的な解析のみの結果に比べ、再現率が上昇している。これは、ソーシャルメディアにおけるユーザー間のつながりを利用することにより、言語的な解析では抽出できなかった批判的ユーザーを抽出できているためであると考えられる。すなわち、ユーザー間の支持関係を利用することにより、言語的に抽出した批判的ユーザーと同じ意見を持つユーザーを批判的ユーザーとして抽出できている。

グラフ解析では、対象とする実体のアカウントも入力とし、実体を強く支持しているユーザーを批判的ユーザーとして抽出しないようにしている。これにより、グラフ解析のシードとしている、批判的ユーザーの選択の誤りが緩和されていると考えられる。このため提案手法では、言語的な解析のみの結果より、適合率も上昇したのだと考えられる。しかし、図 3 のグラフ解析のみの手法の結果より、グラフ解析のシードとする批判的ユーザー

ランダムに選択した場合、高い適合率は達成できない。したがって、グラフ解析への入力としては、ある程度高い適合率の批判的ユーザーの集合とする必要があると言える。

## 5 まとめ

本論文では、ソーシャルメディア上において、特定の個人や組織などの実体に対し、日常的に非難を行う批判的ユーザーを、言語的情報とユーザー間の関係の情報を用いて抽出する手法を提案した。言語的な解析では周辺の文脈から話者の意図を読み取ることや、ドメイン特有の表現を解析することが難しい。しかし、ソーシャルメディアにおけるユーザー間の関係を利用することにより、言語的な解析だけでは抽出できない批判的ユーザーの抽出が可能であることを示した。

今回はユーザー間の関係として、支持関係のみを用いた。しかしユーザー間には支持以外にも反対、反論している関係が存在する。今後はこのような、ユーザー間の反対関係などもグラフ解析に取り入れたい。

謝辞 本研究は、文部科学省科研費 (23240018)、文部科学省科研費 (23700159)、および JST 戦略的創造研究推進事業さがけの一環として行われた。本研究の実験において、日本アイ・ビー・エム (株) 東京基礎研究所の金山博氏による評価表現抽出エンジン、ならびに荻野紫穂氏による軽卑表現辞書を利用させて頂きました。ここに感謝いたします。

## 参考文献

- [1] Michael D. Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. Predicting the political alignment of twitter users. In *Proceedings of the 3rd IEEE Conference on Social Computing*, pp. 192–199, 2011.
- [2] Michael D. Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. Political polarization on twitter. In *Proceeding of the 5th International AAAI Conference on Weblogs and Social Media*, pp. 89–96, 2011.
- [3] Sadao Kurohashi and Makoto Nagao. Kn parser : Japanese dependency/case structure analyzer. In *Proceedings of the Workshop on Sharable Natural Language Resources*, pp. 48–55, 1994.
- [4] Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth International AAAI Conference on Weblogs and Social Media*, pp. 122–129, 2010.
- [5] Michael Speriosu, Nikita Sudan, Sid Upadhyay, and Jason Baldridge. Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proceedings of the First workshop on Unsupervised Learning in NLP*, pp. 53–63, 2011.
- [6] Alex Hai Wang. Don't follow me - spam detection in twitter. In *Proceedings of the 5th International Conference on Security and Cryptography*, pp. 142–151, 2010.
- [7] Chao Yang, Robert Harkreader, Jialong Zhang, Seungwon Shin, and Guofei Gu. Analyzing spammers' social networks for fun and profit: A case study of cyber criminal ecosystem on twitter. In *Proceedings of the 21st international conference on World Wide Web*, pp. 71–80, 2012.
- [8] Xiaoxin Yin and Wenzhao Tan. Semi-supervised truth discovery. In *Proceedings of the 20th international conference on World Wide Web*, pp. 217–226, 2011.
- [9] 荻野紫穂, 那須川哲哉, 金山博, 榎美紀. 軽卑表現の情報を活用した知識発見. 言語処理学会第 18 回年次大会発表論文集, pp. 58–61, 2012.
- [10] 金山博, 那須川哲哉, 渡辺日出雄. 木構造変換を利用した評判分析手法. 人工知能学会論文誌, Vol. 26, No. 1, pp. 273–283, 2011.