

マイクロブログへの投稿に基づく政治家の立場推定

Estimation of a politician's stance based on his/her posts to microblogs

岩井 宏道 *¹ 道満 恵介 *² 井手 一郎 *³ 出口 大輔 *⁴ 村瀬 洋 *⁵
Hirohito Iwai Keisuke Doman Ichiro Ide Daisuke Deguchi Hiroshi Murase

*¹名古屋大学 工学部 *²中京大学 工学部
School of Engineering, Nagoya University School of Engineering, Chukyo University

*³*⁵名古屋大学大学院 情報科学研究科
Graduate School of Information Science, Nagoya University

*⁴名古屋大学 情報連携統括本部
Information and Communications Headquarters, Nagoya University

We propose a method for estimating a politician's stance based on sentiment analysis of his/her posts to microblogs. This method calculates the average semantic orientation of a politician's posts in general, and compares it to that of the posts regarding a particular theme, to estimate the politician's stance regarding the theme. Through an experiment with actual posts to "Twitter", the effectiveness of proposed method is shown.

1. はじめに

インターネット上での選挙活動が 2013 年に解禁され、多くの議員や議員候補者がソーシャルネットワークサービス (SNS) を通じて情報発信を行っている。SNS を通じて議員候補者の考えを知ることは、投票意思決定などにおいて重要である。しかし、多数の候補者の情報発信を随時確認するのは負担が大きい。そこで、過去の発言に基づいて、個々の議員候補者が関心を持つトピックや、それに対する立場を自動的にまとめて提示することができれば、この負担を軽減することができる。本研究では、議員候補者によるマイクロブログへの投稿の感情極性を計算することで、あるトピックに対するその候補者の立場 (本研究では「肯定的」または「否定的」のいずれか) を推定する手法を提案する。

2. 立場の推定

一般に、我々が何らかのトピックについて言及する際には、そのトピックを肯定的に捉えていれば肯定的な語が、否定的に捉えていれば否定的な語がより多く出現すると考えられる。マイクロブログは 1 投稿あたりの文書長が短いため、1 投稿中で複数のトピックに言及することは少ないと考えられる。そこで、1 投稿が 1 トピックに対応していると仮定し、各投稿の感情極性を計算する。そして、あるユーザの投稿すべてを母集団とした感情極性の平均値に対して、そのユーザのあるトピックに関する投稿のみに限った感情極性の平均値を比較することで、そのトピックに関する立場を推定する。

2.1 処理手順

処理手順を図 1 に示す。まず、対象とするユーザの過去の投稿を収集し、各投稿に対して形態素解析を行う。形態素解析には MeCab^{*1} を用いた。

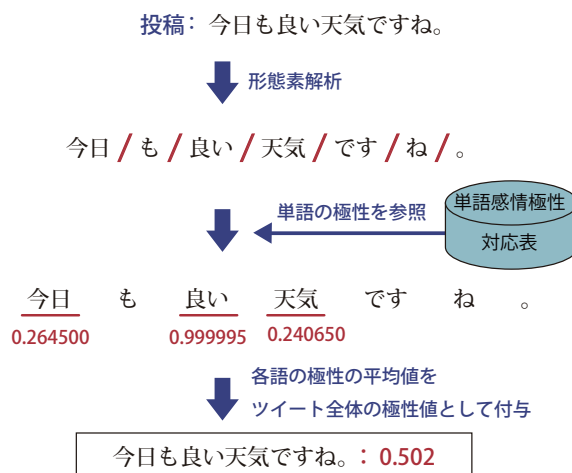


図 1: 投稿に対する感情極性付与の例

次に、形態素解析により分割された各単語について、高村 [1] が作成した単語感情極性対応表を参照し、感情極性を付与する。この単語感情極性対応表には、ある語と、その語が一般に肯定的に使われるか、否定的に使われるかの極性との対応が列挙されている。[-1, 1] を値域として、否定的な印象の語ほど -1 に近く、肯定的な印象の語ほど 1 に近くなるような値が割り振られている。なお、単語感情極性対応表の語彙にない単語については感情極性の計算に含めないことにした。

続いて、図 2 に示すように、投稿中に出現する単語の感情極性の平均値を求め、これをその投稿全体の感情極性とする。そして、あるユーザの投稿すべての感情極性を計算し、その平均値を算出する。これをそのユーザの肯定・否定を判定する際の基準値とする。その後、注目トピックに関連する特定のキーワードを含む投稿を抽出したうえで、それらの感情極性の平均値を計算する。最終的に、基準値よりも高ければそのトピックに対して肯定的、低ければ否定的であると判定する。

連絡先: 岩井宏道, 名古屋大学, 愛知県名古屋市千種区不老町,
iwaih@murase.m.is.nagoya-u.ac.jp

*1 本研究では MeCab-Python の第 0.993 版を用いた。
<https://code.google.com/p/mecab/>

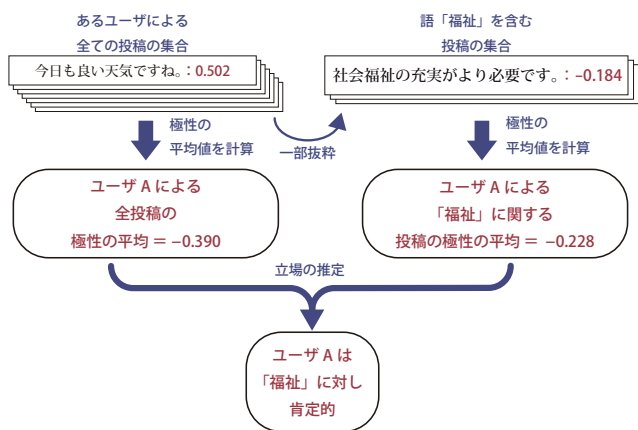


図 2: 提案手法による立場推定の例

表 1: 注目したトピックと検索に用いたキーワード

トピック	キーワード
TPP (環太平洋経済連携協定)	「TPP」
消費税増	「消費税」
憲法改正	「憲法改正」 または 「改憲」
原子力発電所	「原子力発電所」 または 「原発」

3. 評価実験

提案手法を評価するために、2つの実験を行った。実験1ではマイクロブログを利用している議員候補者を対象に、提案手法の推定精度を確認した。実験2では単語感情極性対応表の語彙に制限を加え、推定精度の変化を調べた。

3.1 データセット

2013年参議院議員選挙における議員候補者のうち、マイクロブログサービス“Twitter”を利用していた281名を対象として投稿を収集した*2。2013年11月22日から2014年1月25日までの期間に、合計22,995件の投稿を収集した。そのなかで、「新しい写真をFacebookに投稿しました」のような自動生成された定型文は、ユーザの感情とは無関係であるため削除した。さらに、他のユーザによる投稿文を引用している場合には、引用部分を削除した。本研究では、281名のユーザのうち、投稿数が100件を超えていた40名について、その立場を推定した。

3.2 提案手法の推定精度に関する評価

提案手法により、各ユーザの特定トピックに関する立場を推定した。また、各ユーザの基準値と比べるのではなく、絶対値を用いて、特定のキーワードを含む投稿の平均極性が負であれば否定的、正であれば肯定的とする方法と比較した。推定したトピックおよび当該トピックに関する投稿の検索に用いたキーワードを表1に示す。例として、あるユーザの投稿の感情極性の分布を図3に示す。推定された立場について、文献[2]を正解データとして評価を行った。文献[2]は、自己申告により、議員候補者の各政策に関する立場を「賛成」から「反対」までの5段階で調査したものである。本実験では「賛成」および「どちらかといえば賛成」を肯定的な立場、「反対」および「どちらかといえば反対」を否定的な立場とみなした。「どちらでもない」や「無回答」となっている項目については推定から除

表 2: 「肯定的」が正解である立場の推定精度 (語彙制限なし)

推定手法	推定精度と件数
比較手法	0.00 (0/10)
提案手法	0.20 (2/10)

表 3: 「否定的」が正解である立場の推定精度 (語彙制限なし)

推定手法	推定精度と件数
比較手法	1.00 (78/78)
提案手法	0.69 (54/78)

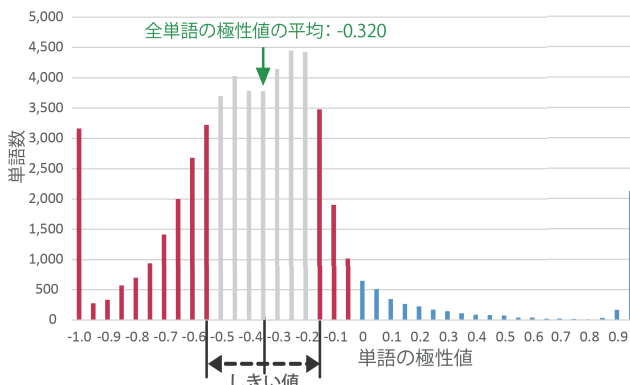


図 5: 単語感情極性対応表の語彙制限

外した。

実験結果を表2, 3に示す。比較手法では「肯定的」である立場を1件も正しく推定できなかったが、提案手法では一部といえども推定できるようになった。

3.3 単語感情極性対応表の語彙に制限を加えた際の推定精度に関する評価

単語感情極性対応表には極性が-1または1に近いような、強い否定・肯定の特徴をもつ単語が含まれる一方で、極性値が0に近い、実際には否定・肯定どちらの文脈にも登場し得る単語も数多く含まれる。このような中間的な極性値の単語を除外することで、極端な感情極性がその投稿の感情極性へと反映されやすくなると考えられる。そこで、図5に示すように、ある感情極性の範囲に含まれる単語を除外したうえで実験1と同様の推定を行った。除外する範囲は、単語感情極性対応表に含まれる全単語の感情極性の平均値である-0.32を中心として、±0.1~0.6までの6通りを設定した。

実験結果は表4, 表5のようになった。適切なしきい値を設定することで、推定精度が向上することがわかった。また、例として、中心±0.6までの単語を除外した際の、図3と同じユーザの投稿の感情極性の分布を図4に示す。図3と比べて図4では投稿の感情極性の分散が大きくなり、否定・肯定の特徴がより際立って感情極性に表れていることがわかる。

4. まとめ

ユーザのマイクロブログへの投稿に基づき、投稿の感情極性を計算することで、特定のトピックに関する立場を推定する手法を提案した。極性値の正負を基準にした立場推定では「肯定的」な立場が全く推定できなかったのに対し、提案手法では正しく推定できるようになった。実験2では、単語感情極性対応表の語彙に適切な制限を加えることで、推定の精度が向上することがわかった。

*2 投稿の収集には Twitter から公開されている API を利用した <https://dev.twitter.com/>

表 4: 「肯定的」が正解である立場の推定精度 (語彙制限あり)

極性値のしきい値	推定精度と件数
しきい値なし (実験 1 と同様)	0.20 (2/10)
中心± 0.1	0.10 (1/10)
中心± 0.2	0.20 (2/10)
中心± 0.3	0.20 (2/10)
中心± 0.4	0.20 (2/10)
中心± 0.5	0.20 (2/10)
中心± 0.6	0.40 (4/10)

表 5: 「否定的」が正解である立場の推定精度 (語彙制限あり)

極性値のしきい値	推定精度と件数
しきい値なし (実験 1 と同様)	0.69 (54/78)
中心± 0.1	0.71 (55/78)
中心± 0.2	0.71 (55/78)
中心± 0.3	0.76 (59/78)
中心± 0.4	0.68 (53/78)
中心± 0.5	0.72 (56/78)
中心± 0.6	0.68 (52/78)

今後は、渡邊ら [3] のように潜在的 Dirichlet 配分法や協調フィルタリングを導入するなどして、ユーザが直接言及していないトピックに関して間接的に立場を推定したり、推定精度の向上を図る必要がある。

参考文献

- [1] 高村大也, 乾孝司, 奥村学, “スピンモデルによる単語の感情極性抽出,” 情報処理学会論文誌, vol.47, no.2, pp. 627–637, Feb. 2006.
- [2] (株) 朝日新聞社, “朝日・東大谷口研究室共同調査—2013 参院選: 朝日新聞デジタル,” <http://www.asahi.com/senkyo/senkyo2013/asahitodai/> [2014/3/10 アクセス]
- [3] 渡邊恵太, 加藤昇平, “潜在的ディリクレ配分法に基づく協調フィルタリングを用いたマイクロブログユーザの興味対象分析,” 第 12 回情報科学技術フォーラム講演論文集, vol.2, pp. 273–276, Aug. 2013.

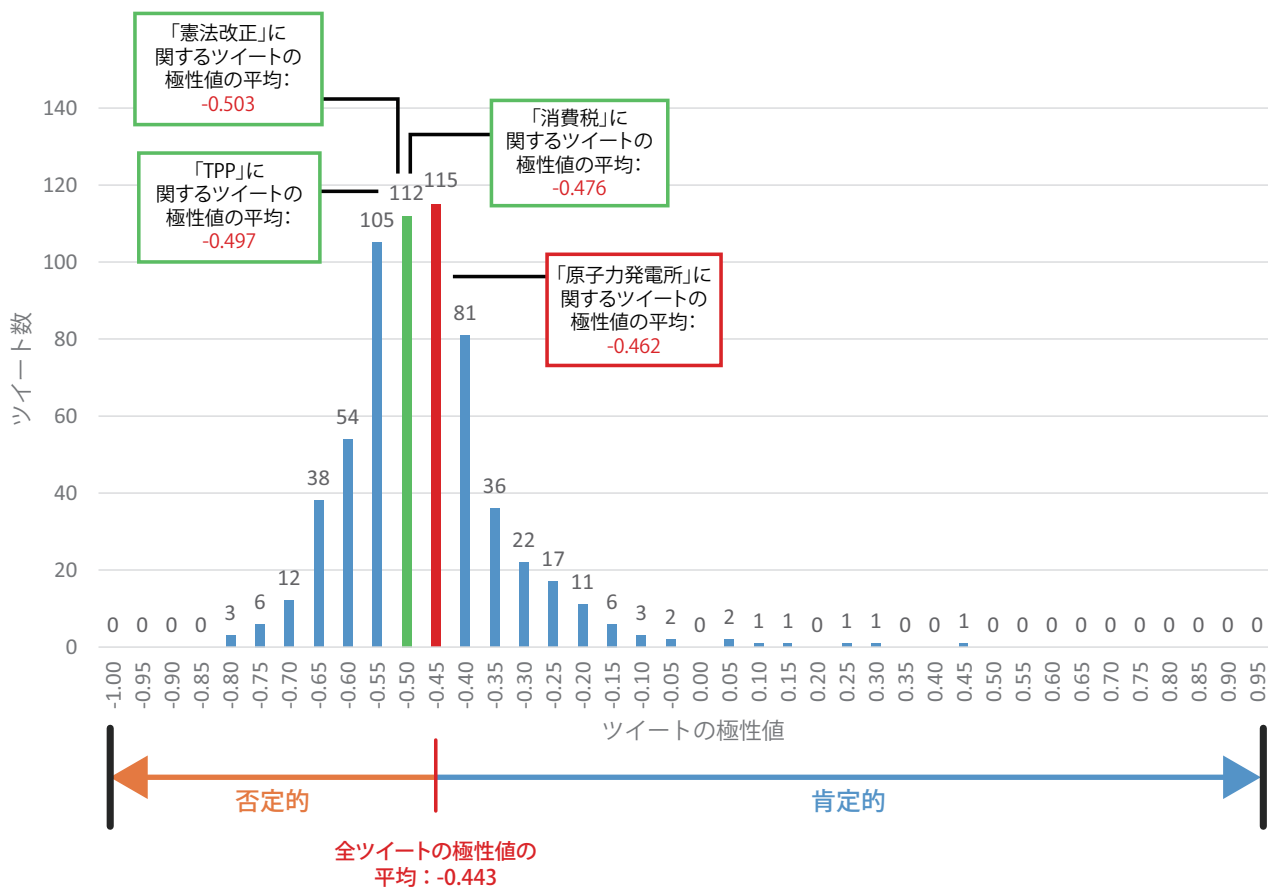


図 3: あるユーザの投稿の極性の分布 (語彙制限なし)

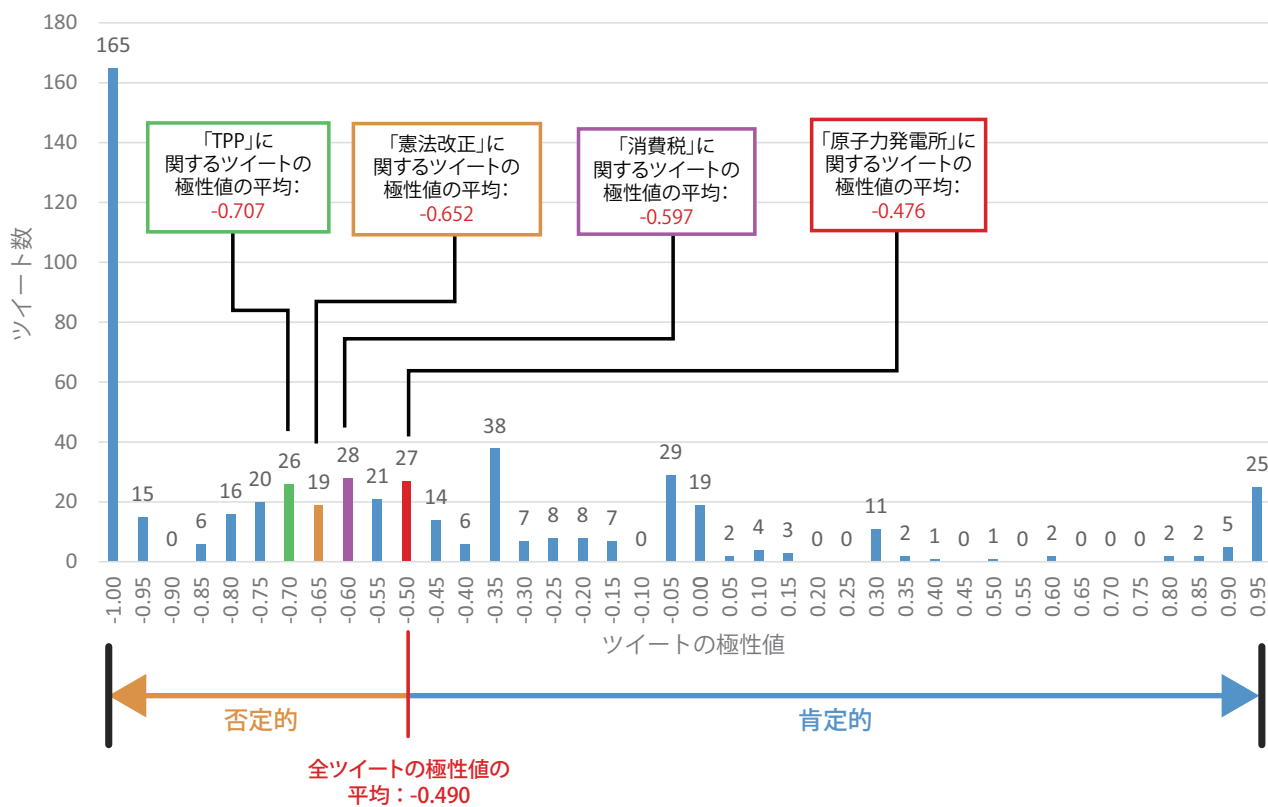


図 4: あるユーザの投稿の極性の分布 (語彙制限あり)