

Twitter における議論の論点可視化に向けた賛否分類に関する研究

情報科学科 半田 雄士

指導教員：小林 邦和

1 はじめに

近年 SNS の利用者はさらに増加し、政治家らも SNS を活用する動きが広がっている。SNS 上の政治的論争は、選挙での投票を左右する重要な判断材料である。しかし、SNS はフォローしたアカウントの意見が多く表れるシステム上、意見が偏る可能性がある。そのような偏った意見による判断は非合理的であり、特に意思決定の際は問題となる。この問題を解決するためには、意見を整理し、ユーザーに分かりやすく伝えるインターフェースで、議論の論点を可視化をすることが必要である。

上記の問題に対処する既存のシステムとして、Yahoo!リアルタイム検索 [1] や、OpinionReader[2] が挙げられる。しかし、これらのシステムは文章中のポジティブ/ネガティブな言葉から感情を決定しており、意見の観点が無視されている。観点とは、物事を見る立場のことであり、例えば「携帯電話」というカテゴリの観点には「iPhone」など固有名詞が挙げられる。さらに「iPhone のカメラ」など細かい観点もあり、このような観点がわかることで議論の詳細な賛否分類ができるようになる [3]。

本研究の目的は、意見の観点を考慮したシンプルな賛否分類手法を示すことである。具体的には、観点と感情の 2 種類の分類器を組み合わせることで、観点を考慮した賛否分類を行う。

2 提案手法

賛否分類を行うには、「観点に対する感情分類」をする必要がある。ここでは、賛否分類のために 2 種類の分類器を用意する (図 1)。1 つ目の分類器では、文章の「観点」を判別するためにカテゴリの多値分類を行う。2 つ目の分類器では、文章の「感情」を判別するために 2 値分類を行う。これらの分類をすることで、ある文章がどの観点に対してポジティブかネガティブか分類できる。つまり、賛否分類が可能になる。

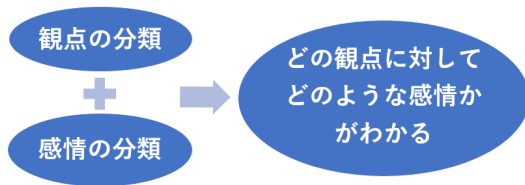


図 1 賛否分類のシステム図

3 実験

提案手法の性能評価のため、計算機シミュレーションを行った。データセットは、Twitter 日本語評判分析データセット [4] を用いる。このデータセットから Xperia、Aquos、iPhone の 3 種類のカテゴリに属する、且つポジティブ/ネガティブいずれかのラベルがあるツイートを 12000 個抽出した。抽出したデータのうち、ランダムに抽出した 9000 個を訓練データ、3000 個を検証データとする。なお、カテゴリ別の比率は 1:1:1 である。

上記のデータセットを用いて、提案手法で示した 2 種類の分類器を学習する。分類器は LightGBM を用いる。LightGBM は決定木ベースの勾配ブースティングの機械学習フレームワークで、処理速度が速く、精度が高いという特徴がある。テキストは

形態素解析エンジンの MeCab によりわかち書きしたのち、ベクトル表現に変換する。文書ベクトル表現手法は、予備実験で精度の高かった Bag-of-Words を用いる。

観点の分類では、入力文章がどのカテゴリであるかを出力する。感情の分類では、入力文章がポジティブ、またはネガティブであるかを出力する。

4 結果

計算機シミュレーションの結果、観点の分類では精度 99.5 % (図 2)、感情の分類では精度 84.8 % (図 3) であった。観点の分類が約 100 % になった理由は、特定の単語を含むかどうかをベクトルで表した Bag-of-Words が有効に働いたことによるものと思われる。一方、感情の分類が約 85 % になった理由は、Bag-of-Words の否定形を認識できないという欠点によるものと思われる。

Truth	Xperia	991	8	1
	Aquos	4	996	0
	iPhone	3	0	997
		Xperia	Aquos	iPhone
Truth	Negative	1770	168	
	Positive	288	774	
		Negative	Positive	

図 2 カテゴリ分類の結果 図 3 ポジネガ分類の結果

5 おわりに

本論文では、賛否分類を行うためには「観点」と「感情」の分類が必要であると仮定し、2 種類の分類器の実装を行った。観点の分類において、今回は教師あり学習による分類を行ったが、Twitter の生データはラベル付けされていないため、教師なし学習で分類できるほうが望ましい。そのため発表会までに、k 平均法やトピックモデルの LDA などクラスタリング手法を用いた観点の分類について検証する予定である。教師なし学習で観点の分類ができるのであれば、手動で見つけづらい細かい観点でも文章を捉えることができ、賛否分類に有用であると考えられる。

本研究が今後目指すものは、議論を可視化するインターフェースの開発である。つまり、1 章で紹介した Yahoo!リアルタイム検索や OpinionReader での問題点を解決し、意思決定に役立つシステムの構築を目指す。

参考文献

- [1] 野畑周, 内藤弘朗, 清水徹. "ヤフージャパンのリアルタイム検索における感情分析" 電子情報通信学会技術研究報告, Vol.114, No.211, pp.31-35, 2014.
- [2] 藤井敦. "OpinionReader: 意思決定支援を目的とした主観情報の集約・可視化システム" 電子情報通信学会論文誌 D, Vol.91, No.2, pp.459-470, 2008.
- [3] S. Somasundaran, J. Wiebe. "Recognizing stances in online debates" Proceedings of ACL2009, Vol.1, pp.226-234, 2009.
- [4] Y. Suzuki. "Filtering Method for Twitter Streaming Data Using Human-in-the-Loop Machine Learning" Journal of Information Processing, Vol.27, pp.404-410, 2019.