

時事問題の論点を Web からマイニングするシステム

井上 結衣[†] 藤井 敦[†]

[†] 筑波大学図書館情報メディア研究科 〒305-8550 茨城県つくば市春日 1-2

あらまし World Wide Web には、「赤ちゃんポスト」のような時事問題に対して様々な人々が自分の考えを発信している。これらの意見情報から人々の考え方に対する傾向や法則を発見することができれば、個人や組織の意思決定に役立つ可能性がある。本論文では、Web から意見情報をマイニングし、可視化するシステム「OpinionReader (オピニオンリーダー)」を提案する。OpinionReader では、ある問題について賛否両論が対立する場合は「論点」が存在する点に着目する。Web 上の意見情報から頻出する語句を論点として抽出し、それぞれの論点がどれだけ重要かを表す「重要度」と、どれだけ賛成または反対に固有かを表す「固有度」を計算し、賛成派と反対派が対立する構図を 2 次元グラフ上に可視化する。それぞれの立場でよく論じられる論点を知ることによって、後から議論に参加した人でもこれまでの議論を効率的に把握でき、また議論が不十分な点を発見できるといった利点がある。さらに OpinionReader は、論点が似ている別の時事問題を検索することができる。

キーワード Web マイニング, 意見情報, 論点, 可視化

A Web mining system for points at issues of current topics

Yui INOUE[†] and Atsushi FUJII[†]

[†] University of Tsukuba Kasuga 1-2, Tsukubashi, Ibaraki, 305-8550 Japan

Abstract On the World Wide Web, the volume of opinions for current topics, such as "baby hatches", has been increasing. The trends and rules latent in a large set of opinions can potentially be useful for decision-making purposes. In this paper, we propose a system, called "OpinionReader", for mining and visualizing pro and con arguments for a target topic. For visualization purposes, OpinionReader arranges points at issue for a target topic in a two-dimensional space, in which the x and y axes correspond to the polarity and the importance of points at issue, respectively. Users can easily summarize the main points and how they are used in arguing about the topic in support of one standpoint or the other. In addition, OpinionReader can search for related topics whose distribution of points at issue is similar to the target topic.

Key words Web mining, Opinions, Points at issue, Visualization

1. はじめに

World Wide Web には、報道記事のように客観性が高い情報だけでなく、意見、評判、感想などの主観情報も存在する。例えば、「赤ちゃんポスト」や「ワーキングプア」などの時事問題に対して、様々な人々が掲示板やブログ等を通して自分の考えを発信している。これらの意見情報から人々の考え方に対する傾向や法則を発見することができれば、個人や組織の意思決定に役立つ可能性がある。例えば、「赤ちゃんポスト」に対する賛否両論が含まれる意見群を読んで、その問題に対する自分の態度を決定する場合がある。

筆者らは、Web から時事問題に対する意見情報をマイニングし、その傾向を可視化することによってユーザの意思決定を支援するシステム「OpinionReader (オピニオンリー

ダー)」[6], [8], [11] について研究している。

本研究において、意思決定とは、ある話題に対する賛否両論を網羅的に洗い出し、対立させて、より合理的な立場を採用する過程ととらえる。

ある問題について賛否両論が対立する場合は「論点」が存在する。例えば、「株式会社の病院経営」という話題では、賛成派は「情報公開が促進される」、反対派は「病院が利益追求に走る」などを論拠として主張を展開する。ここでは、この論拠を「論点」と呼ぶ。

そこで OpinionReader は、時事問題に対する論点を Web からマイニングし、各話題ごとに賛否両論が対立する構図を論点に基づいて可視化する。具体的には、各論点がいかに重要かを表す「重要度」と、どれだけ賛成または反対に固有かを表す「固有度」を計算し、賛成派と反対派が対立する構図を二次元

グラフ上に可視化する。また、論点ごとの代表的な賛否両意見を対峙させて提示する。

その結果、ユーザは大量の意見情報を読まなくても、その話題に関する議論の全容を把握することができる。この手法は、議論に途中から参加したユーザにとっても有用性が高い。また、時事問題ごとに特有の論点をマイニングすることで、論点の分布が類似している別の時事問題を検索することができる。この機能を用いると、まだあまり議論されていない新しい時事問題と、問題の本質が似ていて、歴史が古く以前から議論がなされている時事問題の意見とを読み比べることができ、より合理的な意思決定が可能になる。

2. 関連研究

近年、主観情報に対するテキストマイニングの研究が行われている。主観情報の抽出では、辞書、抽出パターン、機械学習などを用いる手法が提案されている [5], [7]。主観情報の分類では、「肯定」か「否定」に二値分類する手法 [9] や公定と否定の程度に応じて多値分類する手法 [2], [4], [10] がある。

主観情報の分類に関する既存の手法は、商品や映画に対する批評を対象としていることが多い。批評の記述には、特定の商品とは無関係に「満足した」や「不具合」のような肯定や否定に特有の表現が存在する。既存の研究ではこのような肯定や否定に関する普遍的な表現を学習することが中心的な課題である。

それに対して、本研究は「赤ちゃんポスト」などの時事問題に対する Web 上の意見を対象とする。この場合「賛成」と「反対」という言葉以外には、話題とは無関係にどちらかの立場に特有の表現を見つけることが難しい。このことは、話題の選び方によって賛成と反対が入れ替わることから分かる。例えば、「詰め込み教育」という話題に対する反対意見は「ゆとり教育」という話題に対する賛成意見になる可能性が高い。

Eguchi ら [3] は、この問題に取り組んだ。しかし、Eguchi らが新聞記事から意見を収集するのに対して、本研究は雑多な Web から意見を収集する点が異なる。また、本研究は、検索モデルに依存しないため、既存の検索エンジンを利用することができる。

また、主観情報を集約して閲覧を支援する研究もある。blog-Watcher^(注1)は、入力したキーワードに関するブログ記事を検索し、記事中の評判情報を肯定と否定に分類して、ヒストグラフや円グラフで表示する。

Hu ら [1] は、商品に対する批評を要約するために、批評テキストから抽出した特徴を重要度という 1 次元の軸によって可視化した。ここで、Hu らの研究における「特徴」は、本研究における「論点」に相当する。それに対して、本研究は固有度と重要度によって 2 次元平面に論点の分布を可視化する点が異なる。この点は、本研究で提案するシステムの主要な新規性である。また、Hu らが商品に対する批評を対象にしている点に対して、本研究では時事問題に対する意見を対象にしている点が異なる。

3. OpinionReader の概要

OpinionReader の概要を図 1 に示す。はじめに、ユーザは時事問題のキーワードをシステムに与える。例えば「赤ちゃんポスト」、「ワーキングプア」などである。システムは与えられた時事問題のキーワードに対する意見情報を Web 上から収集し、賛否に分類する [6]。

本システムにおける意見収集手法は、高精度の初期検索と、網羅性が高い再検索の 2 段階検索に基づく。初期検索では「赤ちゃんポストに賛成です」などの具体的な検索質問を用いて検索を行う。次に初期検索で検索されたページに頻出する言葉を関連語として抽出し、関連語を検索質問に追加して再検索を行う。さらに、初期検索で賛成意見として収集された段落を正例とし、反対意見として収集された段落を負例として、SVM (サポートベクターマシン) で分類器を学習する。検索された段落を分類器によって賛否に分類し、収集の精度を向上させる。

上記の手順で図 1 の「Web からの意見抽出」が完了する。

次に、収集した意見情報から時事問題に対する論点をマイニングし、二次元グラフ上に可視化する [8], [11]。この処理は、図 1 における「論点の抽出」以降の手順に該当する。

本研究の目的は、時事問題について、賛否両論が対立する構図を論点に基づいて可視化することである。本研究は、各論点ごも性質の違いに着目して可視化を行う。

まず、賛成意見と反対意見の両方で論じられる論点と、どちらか一方の立場で特に論じられる論点がある。こうした違いを表す尺度として、論点の「固有度」を考える。また、ある立場において、重要な論点とそうでない論点がある。こうした違いを表す尺度として、論点の「重要度」を考える。

上記の固有度と重要度に基づいて、論点の分布を二次元平面上に表示する。ユーザは対象の話題に関する論点を俯瞰して全容を把握することができる。

しかし、各論点について、どのように議論されているのかわからない場合がある。そこで、各論点に関する複数の意見を順位付きリストで表示する。ユーザは重要な意見から選択的に読むことができる。

さらに、ある時事問題と論点の分布が類似する別の時事問題を検索し、ユーザに提示する。ユーザは問題が本質的に似ている他の時事問題の傾向を参考にすることができる。

4. OpinionReader の実行例

図 2 は、OpinionReader で「赤ちゃんポスト」という時事問題を扱った時の出力内容である。図 2 は、論点の分布を可視化する領域と、右端のパネルで構成されている。図 2 では、「育児放棄」、「虐待」などが論点として表示されている。

操作パネルを操作すると、固有度のしきい値を変更して、表示される論点を制限することができる。また、名詞句や動詞句などの言語的な単位を指定することで、表示される論点を制御することができる。

論点の分布を可視化する領域において、横軸と縦軸はそれぞれ固有度と重要度である。中央付近の論点は、賛成派と反対派

(注1): <http://blogwatcher.pi.titech.ac.jp>

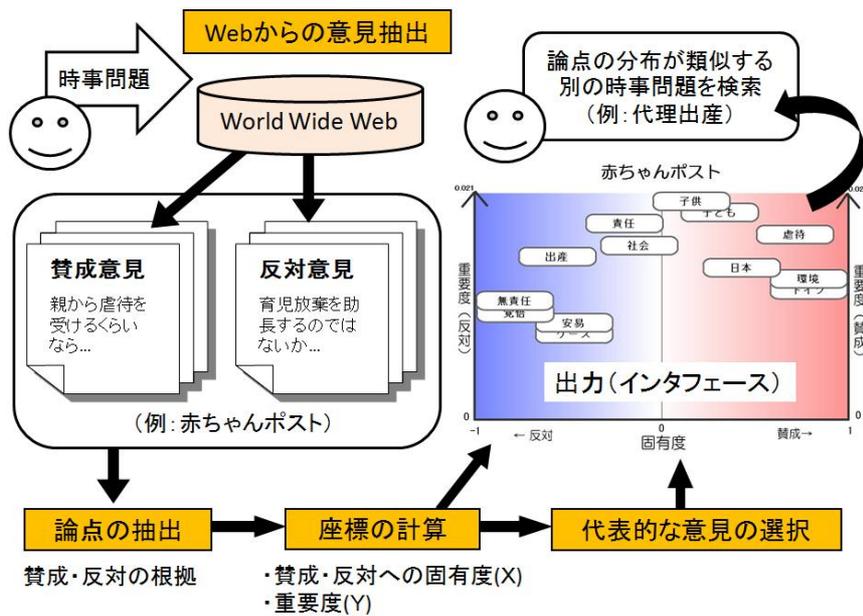


図 1 OpinionReader の概要

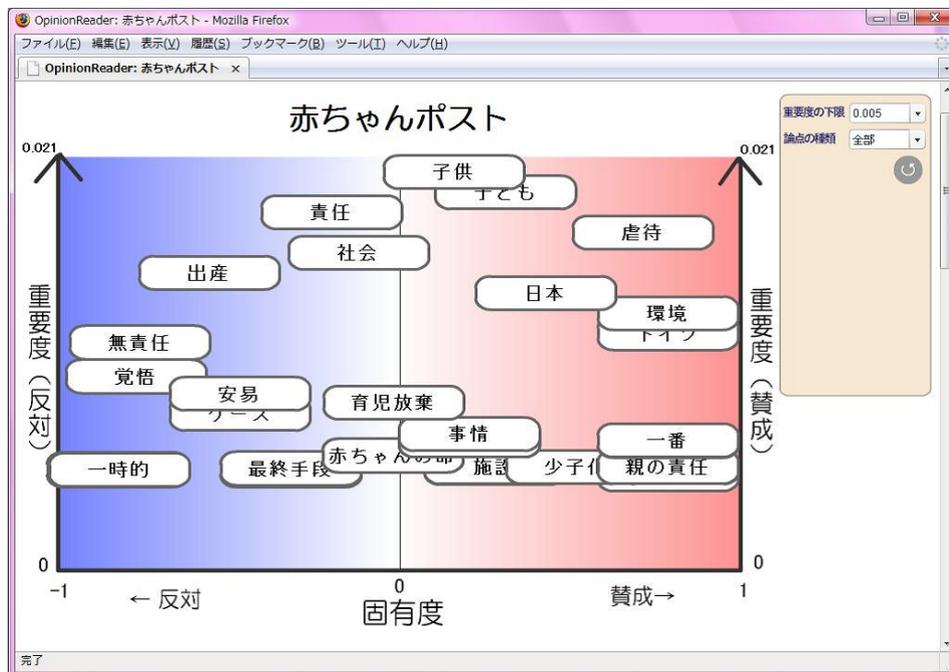


図 2 「赤ちゃんポスト」に対して論点の分布を可視化した例

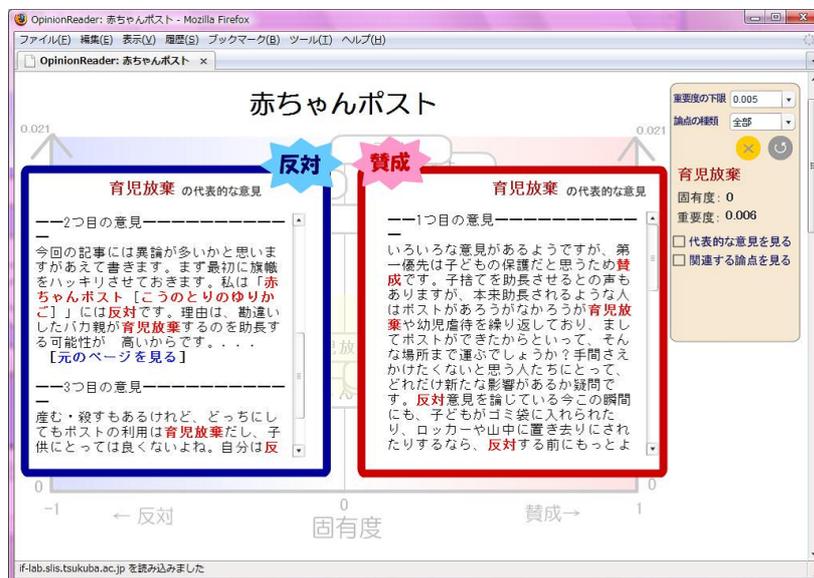


図3 「育児放棄」という論点に対する意見を表示した例

の両方で均等に論じられていることを表す。固有度が一方の立場に偏っている論点は、逆の立場にとっては不利な論点であるか、若しくは議論が不十分なために逆の立場では論じられていない可能性がある。論点の分布を可視化することで、それぞれの立場で何が論拠となっているのかがわかると同時に、それぞれの立場にとって強みになり得る論点や議論が不十分な可能性がある論点を発見することができる。

図2の論点をマウスでクリックして、操作パネルを操作すると、該当する論点に対する意見が順位付きリストとして表示される。図3は、図2の「育児放棄」を選択した場合の例である。図3には、「育児放棄」に対する反対派と賛成派の意見が左右に表示されている。この例では、反対派の意見として「育児放棄を助長する」と述べられている一方、賛成派の意見では「ポストに関係なく育児放棄をする人はする。それより子供の保護が優先」と述べられている。ただし、一方の立場だけで論じられている論点を選択された場合は、逆の立場における意見は表示されない。

以上をまとめると、ユーザは論点の分布を俯瞰し、特定の論点について重要な意見を選択的に読むことによって、議論の要点を効率よく把握することができる。

また、出力インタフェースの下方に、表示している時事問題と論点が類似している時事問題が関連する話題として表示される。例えば「赤ちゃんポスト」に関連する話題として表示される「中絶」という時事問題では、「責任」、「命」、「母親」など「赤ちゃんポスト」と共通する論点が多く含まれる。「赤ちゃんポスト」に関連する時事問題は、他に「代理出産」や「臓器移植」などがある。

表示された時事問題をマウスでクリックすると、選択した時事問題に対する意見の二次元グラフが表示される。この機能を用いて、ユーザはより多角的に時事問題に対する意見を読むことができる。例えば「赤ちゃんポスト」は設置されてまだ歴史が浅いため、議論がまだ不十分かもしれない。しかし、「赤ちゃんポスト」と論点が類似する「中絶」、「代理出産」などの意見は、「赤ちゃんポスト」と問題が本質的に類似しており、議論されてきた歴史も「赤ちゃんポスト」より長い。このようにまだ議論が浅い時事問題について、他の時事問題に対する議論を参考に議論を予測したり、自分の立場を決定することができる。

5. おわりに

本論文は、時事問題に対する意思決定の支援を目的として、Webから時事問題に対する論点をマイニングし、傾向を可視化するシステム「OpinionReader」を提案した。ユーザは可視化された内容を吟味し、合理的な立場を選択することで意思決定を行う。

今後の課題として、時事問題を自動的に抽出するために、時事問題の単位や種類を決定する基準について検討する必要がある。

また、システムの応用範囲を拡張するための項目として、時間による議論の推移を可視化する機能、対立する論点同士の強弱を特定する機能が必要である。

謝 辞

本研究の一部は、文部科学省科研費特定領域研究「情報爆発時代に向けた新しいIT基盤技術の研究」(課題番号: 19024007)によって実施された。

文 献

- [1] M. Hu and B. Liu. Mining and summarizing customer reviews. *Proc. Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168–177, 2004.
- [2] Soo-Mim Kim and Eduard Hovy. Determining the sentiment of opinions. *Proceeding of Conference on Computational Linguistics*, pp. 1367–1373, 2004.
- [3] Eguchi Koji and Victor Lavrenko. Sentiment retrieval using generative models. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 345–354, 2006.
- [4] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationship for sentiment categorization with respect to rating scales. *Proceeding of the 43th Annual Meeting of the Association for Computational Linguistics*, pp. 1367–1373, 2005.
- [5] 小林のぞみ, 飯田龍, 乾健太郎, 松本裕治. 照応解析手法を利用した属性・評価値対および意見性情報の抽出. 言語処理学会第11回年次大会講演論文集, pp. 436–439, 2005.
- [6] 井上結衣, 藤井敦. Web世論からの意見抽出と賛否に基づく分類. 言語処理学会第14回年次大会発表論文集, pp. 364–367, 2008.
- [7] 立石健二, 福島俊一, 小林のぞみ, 高橋哲朗, 藤田篤, 乾健太郎, 松本裕治. Web文書集合からの意見情報抽出と着眼点に基づく要約生成. 情報処理学会研究報告, No. NL-163-1, pp. 1–9, 2004.
- [8] 佐々木千晴, 藤井敦, 石川鉄也. 意思決定支援のための主観情報マイニング. 言語処理学会第12回年次大会発表論文集, pp. 77–80, 2006.
- [9] 鈴木泰裕, 高村大也, 奥村学. Semi-supervisedな学習手法による評価表現分類. 言語処理学会第11回年次大会講演論文集, pp. 668–671, 2005.
- [10] 岡野原大輔, 辻井潤一. 評価文に対する二極指標の自動付与. 言語処理学会第11回年次大会講演論文集, pp. 664–671, 2005.
- [11] 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, Vol. J91-D, No. 2, pp. 459–470, 2008.