

新聞記事におけるBERTに基づく意見文 抽出を用いた価値観推定

Human Values Estimation Based on Opinion Extraction Using BERT with News Articles

韓毅弘^{1*} 西原陽子² SHAN Junjie³
Yihong Han¹ Yoko Nishihara² Junjie Shan³

¹ 立命館大学情報理工学研究科

¹ Graduate School of Information Science and Engineering, Ritsumeikan University

² 立命館大学情報理工学部

² College of Information Science and Engineering, Ritsumeikan University

³ 立命館大学グローバル・イノベーション研究機構

³ Ritsumeikan Global Innovation Research Organization, Ritsumeikan University

Abstract: 本研究は価値観は意見文に反映されると仮定し、意見文抽出に基づく価値観推定手法を提案する。与えられた文書から意見文を抽出し、その意見文が反映する価値観を推定することで、文章全体の価値観分布を示す。提案手法は深層学習の仕組みを利用し、複数の文の連結によるアップサンプリング手法を採用する。毎日新聞の記事コーパスを用いて意見文抽出と価値観推定の実験を行い、意見抽出では92%の精度を得られ、アップサンプリングにより、価値観推定ではベースラインより50%以上の精度向上を確認した。

1 はじめに

Web 2.0 の普及とともに、UGC (User Generated Contents) の時代が幕を開けた。2000 年代以降、インターネットで自分の意見や感想を投稿することは一般的に行われるようになった。SNS (Social Network Service) が発展し、ユーザの興味を研究し、その嗜好に沿ったコンテンツを提供できるよう、あらゆる推薦システムが提案された。これらの技術は確かにユーザにとっての利便性を高め、サービスの体験を向上させた。実際に技術の恩恵を受け、それを支持している人も多いと思われる。

その一方で、「エコーチェンバー現象」という問題が生まれている。エコーチェンバー現象とは、価値観の似た者同士で交流、共感し合うことにより、特定の意見や思想が増幅する現象を指す。ユーザは自分の考え方似ている意見しか見えず、自分と異なる意見の存在を認識できない。その結果、視野が狭められ、社会全体の様子や別の立場の人の訴求が見えなくなる。特に社会的議題に関しては、考え方・価値観の違いで意見が分断され、建設的な議論が行きにくくなることも予想

される。本研究は「エコーチェンバー現象」という、実際に存在している様々な視点からの意見への限定的アクセスを、この問題の核心的問題点として捉える。このための第一歩として、テキストデータから人間の価値観の推定手法を検討する。

本研究はテキストデータから書き手の価値観を推定する手法を提案する。価値観を反映する文書のサンプルとして、新聞の社説を利用する。提案手法は「価値観は意見文の中に潜んでいる」という仮説に基づき、入力文書から意見文を抽出し、抽出された意見文をもとに、文書全体の価値観を推定する。本研究の主な成果は以下の 2 点になる：

1. 与えられた文書中の意見文を抽出し、各意見文の価値観を推定することで、文書全体の価値観を推定する手法を提案する。
2. 文書中の意見文を組み合わせ、連結することにより、ある価値観の意見文の訓練データ数を増やす。

2 関連研究

意見抽出及び価値観推定に関する既存研究を紹介する。

*連絡先：立命館大学情報理工学研究科
〒 525-8577 滋賀県草津市野路東 1 丁目 1-1
E-mail: is0387ps@ed.ritsumei.ac.jp

2.1 意見抽出

意見抽出は、コメント、レビューなどのテキストデータから、人々の観点や意見を探索する自然言語処理技術であり、あるトピック（政策、製品、サービスなど）に対するユーザの態度を把握することを目的とする [1]。意見抽出に関する先行研究の多くは、主観的な評価表現に着目している [2]。意見表現は常に対象、属性、評価の3つの要素で構成されるため、一般的な自然言語処理では、「意見表現」という概念は「評価表現」に記載されることが多い [3]。「この車のエンジンは良い」という意見文では、「車」が対象で、「エンジン」が属性、「良い」という形容詞が評価となる。また、客観的な記述から評価情報を抽出しようとする研究もある [4]。主観的な表現により記述された評価情報だけではなく、客観的な表現により記述された評価情報も評価表現として扱われる [4]。川田ら [5] は、Web テキストから評価情報を収集・分類し、評価表現のラベル付きコーパスを構築した。中川ら [6] は、川田らの研究をもとに、主観的評価表現と客観的評価表現の抽出手法を提案した。中川らは、評価表現を「当為」「要望」「感情」「批評」「採否」「出来事」という、7つのカテゴリーに分ける評価表現分類法を定義した。

本研究でも同様に意見文抽出を行うが、評価表現の抽出などは行わず、文が意見文であるか非意見文であるかの判断のみを行う。意見文の抽出には、機械学習手法の一種であるニューラルネットワークを用いる。

2.2 価値観推定

人間の価値観推定は、心理学・社会学・人類学・情報科学など幅広い分野で急速に発展しているトピックであり、主な課題は価値観カテゴリーの定義と、与えられたソースデータからの価値観検出である。Cheng ら [7] は人間の価値観のメタリストを提案した。彼らは価値観に関する既存研究をレビューし、人間の価値観の定義を作成した。さらに 12 の価値観リストをレビューし、人間の価値観のメタリストを作成した。石田ら [8] は、人間のアノテーターによる効率的な価値観アノテーションのアプローチを提案した。彼らはアノテーションの作成方法として、アノテーターが協力して各文書に対して最適なアノテーションを作成する方法と、各アノテーターがそれぞれ最大限の文書数をアノテーションする方法の 2 つを考案した。Kristoffer ら [9] は第 4 回 European Social Study (ESS-4) の調査データを分析するための、潜在クラス分析に基づく人間の価値観推定手法を提案した。Kristoffer らの研究は、調査結果に隠された人間の価値観要素を探ることに重点を置いている。この推定手法は ESS-4 のオリジナルデータにどのくらい適合しているかを評価ポイント

トにしている。高山ら [10] は、単語ベースの確率的潜在変数モデルを用いて人間の価値観を検出する手法を提案した。この手法は、学習データの各単語に 6 種類の価値観を表す 6 次元ベクトルを付与し、潜在変数モデルを用いて与えられた文章の価値観を推定するものである。

本研究では、単語ベースではなく、文ベースの価値観推定手法を提案する。単語ベースのものと比べ、文ベースの処理は文脈情報を利用できる。また、提案手法は文書ベースのアノテーションを利用することにより、各単語の価値観辞書の作成や、各単語のアノテーション作業は省かれる。これにより、アノテーション作業が軽減され、より効率的なモデル構築につながる。

3 提案手法

提案手法について説明する。提案手法は、与えられた文書に含まれる人間の価値観を推定する。図 1 は提案手法の処理の流れを示す。

価値観の推定のために、本研究では 1 つの仮説を置く。ある文書に 1 つの価値観が反映されているとき、文書内の意見文には同一の価値観が反映されているとする。文書内には意見文と非意見文があり、意見文には書き手の価値観が反映されており、非意見文には価値観は反映されていないとする。非意見文は事実や真実を述べている文を指す。例えば意見文「原子力の安全は、何よりも優先される」があるとき、ここには「Human Welfare」という価値観が反映されており、非意見文「原発への依存の見直し」という観点が加わった」には価値観は反映されていないとする。

この仮説に基づき、提案手法では、価値観推定の前処理として、入力文書から意見文の抽出を行う。

1. 提案手法は文書が入力として与えられる。文書は意見抽出器に送られる
2. Step.1: 意見抽出器は入力された文書から意見文を抽出する
3. Step.2: 価値観推定器は各意見文の価値観を推定する
4. 各意見文の価値観推定結果の分布を文書全体の価値観推定結果とする

本研究では 2 つの機械学習モデルを構築し、それぞれ意見文抽出と価値観推定のタスクに当てる。2 つのモデルとともに BERT 言語モデル [11] と Transformers フレームワーク [12] により訓練された。本研究では、東北大学が公開している pre-trained モデル¹と、石田ら [13]

¹<https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

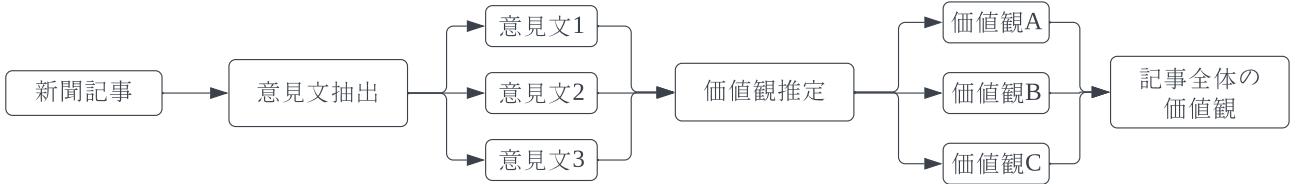


図 1: 提案手法の処理の流れ

が作成した価値観ラベル付き新聞記事データセットが利用されている。新聞記事コーパスは毎日新聞社 2011 年の原子力発電に関する社説記事を利用している。

3.1 意見文抽出

入力された文書から意見文を抽出する。意見文の抽出は機械学習モデルを作成し、モデルで文を意見文と非意見文に分類することにより行う。

機械学習を行うために、訓練用のデータセットを作成する。前述した中川ら [6] が提案した意見文の定義を参考に、毎日新聞の記事データを用いて、意見文と非意見文の訓練用データセットを構築した。30 件余りの記事に含まれた文を手作業でラベルづけし、224 文の意見文と 458 文の非意見文を訓練用データセットとして用意した。各文を BERT の日本語モデルでトークン化し、ベクトルとして表現する。トークン数の上限は 128 とする。ベクトルをニューラルネットワークに入力し、学習を行いモデルを作成する。

3.2 価値観推定

意見文の価値観を推定する。価値観の推定も機械学習モデルを作成し、推定を行う。

文書に 1 つの価値観が反映されているとする。一般に、ある文書に反映される価値観を推定することは複数の意見文を参照することにより達成可能だが、一つの意見文からその文書に反映される価値観を推定することは難しいと考えられる。そこで、機械学習を行う際に、各意見文の価値観を推定するのではなく、同一文書に含まれる意見文の組み合わせに対し価値観を推定するアプローチを取る。これにより 2 つの利点が生まれる。1 つは価値観の推定に用いることができる情報量が増えることである。1 つの文よりも 2 つ以上の文が与えられる方が、学習できる特徴が増える。もう 1 つは、モデルを構築する学習データが増えることである。文書中に n 個の意見文があり、2 文を組み合わせて学習するならば学習できるデータの数は $n(n - 1)/2$ となり元の n よりも多くのくなる。本論文では同一文書に含まれる意見文を組み合わせて学習することにより価値観の推定を行う。

本論文では、組み合わせる文数のことを「拡張レベル」と呼ぶ。文書を D 、文書内の意見文の数を $n(D)$ とする。拡張レベルを Lv とするとき、文書 D 内での意見文の組み合わせ数は $n(D)C_{Lv}$ となる。さらに、訓練データの構築に N 個の文書が使われるならば、最終的に作成される訓練用データセット内のサンプル数 $Dataset$ は式 (1) により与えられる。

$$Dataset = \sum_{i=1}^N (n(D_i)C_{Lv}) \quad (1)$$

4 評価実験

2 つの評価実験の実施方法と実験結果について説明する。1 つ目は意見文抽出の性能を評価する実験、2 つ目の実験は価値観推定の性能について評価する。

4.1 実験 1：意見文抽出

前述した意見文抽出手法の評価実験を行う。

4.1.1 実験準備及び利用するデータ

本研究では、毎日新聞社の 2011 年版の新聞記事から原子力発電に関する記事に価値観がラベリングされたデータセットを使用した [13]。このデータベースは 700 件のニュース記事と 234 件の社説記事の価値観ラベルが含まれる。ニュース記事には 2 人のアノテーターにより複数の価値観ラベルが付与されている。社説記事は 1 つの価値観ラベルが付与されている。社説記事の性質上、作者の評論や意見が多く述べられていることと、シングルラベル分類は比較的安定した分類性能を達成しやすいといった観点から、本論文では主に社説記事を使用した。

意見文抽出の段階では、価値観ラベルには触れず、テキストデータだけを利用している。我々は 234 件の社説記事のうち 23 件をランダムで選び、その 23 件の記事の中にある全部の文を意見文・非意見文で分類し、学習用のデータセットを作成した。学習データセットには意見文 224 件、非意見文 458 件が含まれている。検証

表 1: 学習用・検証用データセットにおいての意見文・非意見文の内訳

	意見文	非意見文
学習用データセット	224	458
検証用データセット	165	272

表 2: 訓練パラメータ

パラメータ	値
num labels	2
max length	128
max epochs	100
training set ratio	80%
validation set ratio	10%
test set ratio	10%

用のデータセットにおいては、同じ方法で、学習データとは別のデータセットを作成した。学習用データセットと検証用データセットの作成の際に使用された記事は異なっている。検証データセットには意見文 165 件、非意見文 272 件が含まれている。2 つのデータセットの内訳は表 1 に示す。訓練時のパラメーターの詳細は表 2 に示す。

モデルの性能は適合率(式(2))、再現率(式(3))、 F_1 値(式(4))により評価する。

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

4.1.2 実験結果

モデル訓練時の validation で得られた精度は最高 96% であった。テストデータで性能評価を行った結果を表 3 に示す。前述した計算法で計算した適合率、再現率、 F_1 値はそれぞれ 83.7%，87.3%，85.5% であった。この結果から、提案手法は意見文抽出を十分な精度で行えることがわかった。

4.2 実験 2：価値観推定

前述した価値観推定手法の評価実験を行う。

表 3: 意見文抽出の評価実験結果

アノテーション	分類結果	
	意見文	非意見文
意見文	144	21
非意見文	28	244

表 4: 価値観カテゴリと各価値観の記事数(降順)

価値観カテゴリ	記事数
Human welfare	57
Effectiveness	56
Importance	48
Power	33
Law and order	17
Wealth	7
Innovation	6
Nature	4
Personal welfare	3
Other	3
合計	234

4.2.1 実験準備及び利用するデータ

実験 2 は実験 1 と同じデータセットを使用した。実験 2 ではデータセットのテキストデータと価値観ラベルの両方を利用する。利用したデータセットには 10 種類の価値観ラベルが存在している、表 4 はその 10 種類の価値観及び各価値観にタグつけされた記事数を示す。

学習を十分に行うために、記事数上位 5 位の価値観を選び、合計 211 件記事を意見文抽出モデルに与えた。意見文抽出を経て、非意見文は削除され、もとの記事についている価値観ラベルを各意見文に割り当て、意見文と価値観ラベルで構成される訓練データセットを作成した。訓練データには 2,106 文が含まれている、表 5 は各価値観カテゴリの意見文数を表している。

1 つの意見文を 1 つの入力にする(拡張率=1)のパターンと、2 つの意見文を 1 つの入力にする(拡張率=2)パターンをそれぞれ用意し、複数の意見文を連結するという提案手法のアプローチの効果を評価するための実験を行った。訓練時のパラメータは表 6 に示す。

訓練時の入力とは別の、推定時の入力にも意見文の連結による効果を試した。そして、異なる入力長で訓練されたモデルの推定性能を評価するため、2 文 1 入力で訓練されたモデルに 1 文 1 入力での価値観推定を行った。テストは合計 3 つのケースで行われた。

表 5: 訓練データにおいての各価値観カテゴリの意見文数

価値観カテゴリ	意見文数
Human welfare	514
Effectiveness	594
Importance	449
Power	375
Law and order	174

表 6: 訓練パラメータ

パラメータ	値
num labels	5
max length	128
max epochs	100
training set ratio	80%
validation set ratio	10%
test set ratio	10%

4.2.2 実験結果

3つのケースの実験結果を表 7 に示す。

5 考察

5.1 意見文抽出の性能

意見文抽出は学習データが 682 件であり、データ量は多くなかったが、分類精度は 96% であり、有用性が示された。考えられる主な理由は、日本語の意見文には文法的な特徴が現れやすいことである。例えば、分類上「当為」に属する意見文の多くは「〇〇べき」「〇〇すべき」という形式で書かれている。「要望」に属する意見文は「〇〇てほしい」文法が使われている場合が多い。こういった文法的特徴の存在は意見文を発見しやすくしている。そして、今回使用している新聞記事のコーパスも意見文抽出の難易度を下げていること

表 7: 各テストパターンの精度結果

訓練時の文数 推定時の文数	1	2
1	40.5%	97%
2	-	99%

が考えられる。普段の話し言葉と違い、新聞記事は言葉使いや文法選びに一定のルールをかけている。こういう表現上の制限が存在しているため、テキスト的な多様性は狭められ、意見表現をより目立たせたと考えている。

5.2 価値観推定における意見文の連結の効果についての検証

表 7 から、訓練時及び推定時に意見文を連結して入力するという提案手法のアプローチの有効性が示された。訓練時に意見文を連結すると、価値観の推定精度は 40.5% から 99% に向上した。このことから意見文を組み合わせ、連結させて学習データを増やすことにより、推定精度を高めることができるとわかった。これは複数の意見文を 1 つの入力にまとめることにより、入力の情報密度が上昇し、ニューラルネットワークにより多い学習材料を提供し、モデルの性能が上がったためと考えている。

2 文を連結して学習し、1 文の価値観を推定するときも、精度が 97% から 99% に上昇し、意見文を連結する効果が確認できる。テキストデータから人間の価値観を推定する場合に、2 つの文を連結させて学習データ数を増やすことの効果が確認された。

6 終わりに

本研究では、意見文抽出を用いた価値観推定手法を提案した。人間の価値観は意見文に最も反映されていると仮定し、意見文から価値観を推定するアプローチを考案した。提案手法は、与えられた文書に対して、テキストから意見文を抽出し、その意見文に含まれる価値観を推定する。文を意見文と非意見文に分類することにより、意見文抽出を行う。価値観の推定では、提案手法は複数の意見文を選択し、それらを連結することで入力テキストの情報量を増やす。各意見文から得られる価値観の分布は、その文書の総合的価値観を表す。

意見文抽出と価値観推定について、2 つの評価実験を行った。毎日新聞の社説記事をコーパスデータとして使用した。意見文抽出の精度は 92% であった。また、評価実験では、 F_1 スコアが 85.5% であった。この結果から、意見文抽出が行えることがわかった。価値観推定の実験では、人間の価値観の推定性能を評価した。訓練時と推定時に複数の文を連結し入力テキストの情報量を増やすことで、精度が向上した。この結果から、提案手法により文の価値観推定が行えることがわかった。

今後の課題として、提案手法を SNS の投稿やネットニュースのコメントに応用し、価値観推定の活用範囲を広めることを計画している。

謝辞

本研究で利用した新聞記事の価値観アノテーションデータは高山泰博先生のチームが作成し、公開したものである。私達はそのデータの詳細について何度もメールで問い合わせしましたが、高山先生は迅速な対応で有益な情報を教えていただきました。高山先生の素晴らしい仕事と親切なサポートに感謝申し上げます。

本研究の一部は立命館グローバル・イノベーション研究機構によって支援されています。

参考文献

- [1] S. Chandrakala and C. S. Sindhu, “Opinion mining and sentiment classification: A survey,” 10 2012.
- [2] J. Wiebe, T. Wilson, R. Bruce, M. Bell, and M. Martin, “Learning subjective language,” *Comput. Linguist.*, vol. 30, no. 3, p. 277 – 308, sep 2004. [Online]. Available: <https://doi.org/10.1162/0891201041850885>
- [3] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, and 福島俊一, “意見抽出のための評価表現の収集,” *自然言語処理*, vol. 12, no. 3, pp. 203–222, 2005.
- [4] 中川哲治, 宮森恒, 赤峯享, 乾健太郎, and 黒橋禎夫, “Web上の客観的記述からの評価情報抽出に関する技術的検討,” *NLP2008 Archives*, 2008.
- [5] 川田拓也, 中川哲治, 森井律子, 宮森恒, 赤峯享, 乾健太郎, 黒橋禎夫, and 木俵豊, “Web テキストにおける評価情報の整理・分類およびタグ付きコードベースの構築,” *NLP2008 Archives*, 2008.
- [6] 中川哲治, 川田拓也, 乾健太郎, and 黒橋禎夫, “Web 文書からの主観的・客観的評価表現の抽出,” in *電子情報通信学会講演論文集*. 電子情報通信学会, 2008.
- [7] A.-S. Cheng and K. R. Fleischmann, “Developing a meta-inventory of human values,” *Proceedings of the American Society for Information Science and Technology*, vol. 47, no. 1, pp. 1–10, 2010.
- [8] E. Ishita, S. Fukuda, T. Oga, D. W. Oard, K. R. Fleischmann, Y. Tomiura, and A.-S. Cheng, “Toward three-stage automation of annotation for human values,” in *Information in Contemporary Society*, N. G. Taylor, C. Christian-Lamb, M. H. Martin, and B. Nardi, Eds. Cham: Springer International Publishing, 2019, pp. 188–199.
- [9] K. J. Albers, M. Mørup, M. N. Schmidt, and F. K. Glückstad, “Predictive evaluation of human value segmentations,” *The Journal of Mathematical Sociology*, vol. 46, no. 1, pp. 28–55, 2022. [Online]. Available: <https://doi.org/10.1080/0022250X.2020.1811277>
- [10] Y. Takayama, Y. Tomiura, E. Ishita, D. W. Oard, K. R. Fleischmann, and A.-S. Cheng, “A word-scale probabilistic latent variable model for detecting human values,” ser. *CIKM '14*. New York, NY, USA: Association for Computing Machinery, 2014, p. 1489 – 1498. [Online]. Available: <https://doi.org/10.1145/2661829.2661966>
- [11] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” *CoRR*, vol. abs/1810.04805, 2018. [Online]. Available: <http://arxiv.org/abs/1810.04805>
- [12] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, and A. M. Rush, “Transformers: State-of-the-art natural language processing,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Online: Association for Computational Linguistics, Oct. 2020, pp. 38–45. [Online]. Available: <https://www.aclweb.org/anthology/2020.emnlp-demos.6>
- [13] E. Ishita, T. Oga, A.-S. Cheng, K. Fleischmann, Y. Takayama, D. Oard, and Y. Tomiura, “Toward automating detection of human values in the nuclear power debate,” *Proceedings of the Association for Information Science and Technology*, vol. 54, pp. 714–715, 01 2017.