

[論文紹介]BERT ベース分類器と SHAP を 用いた地方議会議員の議会発言における 特徴分析+方向性

水上 和秀 (Kazuhide Mizukani)
u355020@st.pu-toyama.ac.jp

富山県立大学 工学部 電子情報工学専攻

June 28, 2024

やりたいこと

自然言語処理について行っていく

- 文章に含まれる感情分析
→レビューの文章の感情値を判定

手段

- BERT を用いた感情分析
- SHAP による分析結果の解釈

BERT とは

BERT (Bidirectional Encoder Representations from Transformers) は、Google が開発した自然言語処理のための深層学習モデル。文の前後を考慮した文脈理解ができる。これにより、高度な自然言語処理タスクを行うことが可能になる。

BERT とは

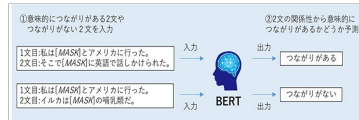
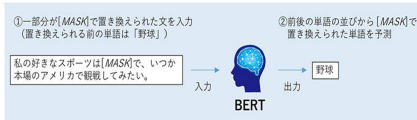
3/13

学習方法

BERT の学習方法は「事前学習」「ファインチューニング」の 2 段階がある。事前学習は、ラベル無しデータを用いて、複数のタスクで事前学習を行うことであり、ファインチューニングは事前学習の重みを初期値として、ラベルありデータでファインチューニング（微調整）を行う。

BERT では MLM と NSP の二つの事前学習を行う

- ある文章において一部のトークンを特殊トークンである [MASK] に置き換えて、その [MASK] に入るトークンを予測する言語モデルのこと。
→単語に対応する文脈情報を獲得できる
- 「意味的につながりがある 2 文」、または「意味的につながりのない 2 文」を入力し、2 文の関係性を考慮することで「入力された 2 文に意味的につながりがあるかどうか」を予測する
→単語の関係性だけでなく文章の関係性の情報も獲得できる



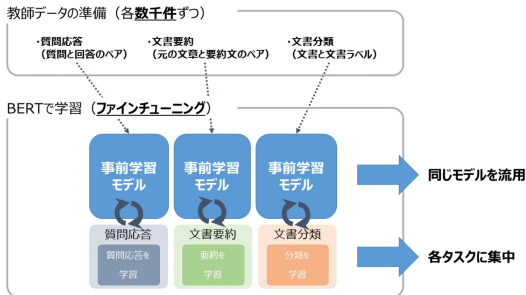
BERT とは

4/13

ファインチューニング

BERT の学習済みモデルは、そのまま使うことは珍しく、一般に、解きたいタスクに応じて特化するようにする。

ファインチューニングを行うときにはモデルの初期値として、事前学習で得られたパラメータを用い、新たに加えられた分類器のパラメータにはランダムな値を与える。そして、ラベル付きデータを用いて BERT と分類器の両方のパラメータを学習する → ファインチューニングの際事前学習で得られたパラメータを初期値として用いることで比較的少数の学習データでも高い性能のモデルを得ることができる



SHAP とは

SAHP とは

機械学習で導出した予測値に対して各特徴量がどのくらい寄与しているかを算出する手法で、シャープレイ値の考え方に基づいている

シャープレイ値とは

協力ゲーム理論において複数のプレイヤーの協力によって得られた利得を各プレイヤーの貢献度に応じて構成に分配するための手段の一つ

- 3 人のプレイヤー (1.2.3) が協力してゲームに挑戦し、利得として、以下の賞金
が得られるとする
- このときの 1.2.3 にそれぞれどのようにお金を分配するか。

表1 協力ゲームの例

参加プレイヤー	賞金/万円
1	4
2	6
3	10
1, 2	16
1, 3	22
2, 3	30
1, 2, 3	60

シャープレイ値とは

6/13

- このとき各プレイヤーの限界貢献度を導入する。限界貢献度とは、プレイヤー i が参加したときの利得の増加分である。
- 例えば、プレイヤーの参加順「 $1 \rightarrow 2 \rightarrow 3$ 」のときのプレイヤー 3 の限界貢献度は、 $v(1, 2, 3) - v(1, 2) = 60 - 16 = 44$ のように計算できる。
- 各プレイヤーのシャープレイ値は以下ようになる

プレイヤー 1: $(4 + 4 + 10 + 30 + 12 + 30)/6 = 15$

プレイヤー 2: $(12 + 38 + 6 + 6 + 38 + 20)/6 = 20$

プレイヤー 3: $(44 + 18 + 44 + 24 + 10 + 10)/6 = 25$

表2 限界貢献度

プレイヤーの参加順	各プレイヤーの限界貢献度		
	1	2	3
1 \rightarrow 2 \rightarrow 3	4	12	44
1 \rightarrow 3 \rightarrow 2	4	38	18
2 \rightarrow 1 \rightarrow 3	10	6	44
2 \rightarrow 3 \rightarrow 1	30	6	24
3 \rightarrow 1 \rightarrow 2	12	38	10
3 \rightarrow 2 \rightarrow 1	30	20	10

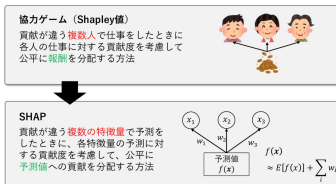
シャープレイ値の定式化

- 一般的には、プレイヤー i のシャープレイ値は次式によって定式化される。ただし、 s は連携 S に含まれるプレイヤー数である。

$$\phi_i = \sum_{S: i \in S \subset N} \frac{(s-1)!(n-s)!}{n!} \{v(S) - v(S - \{i\})\}$$

シャープレイ値と SHAP について

協力ゲーム理論のシャープレイ値の概念を応用して、特徴量の貢献度を計算



SHAP の定式化

解釈したい予測モデルを f 、バイナリ変数 (0 か 1 の変数) を z 、各特徴量に対する貢献度を ϕ_i とすると以下のようにあらわす

$$g(z) = \sum_{i=1}^p \phi_i z_i$$

$$\phi_i(f, x) = \sum_{z \subseteq x} \frac{|z|!(p - |z| - 1)!}{p!} [f(z) - f(z \setminus i)]$$

背景

議会会議録には議会におけるすべての発言が記録されており、多くの自治体が各々の web サイト上で議会会議録を公開している。議会会議録をもとに議員の発言内容を比較することで議会における会議の取り組みや政治的な態度を分析することができる。

目的

会議録中の各発言の発言者を推定する BERT ベース分類器と SHAP を用いて産出されるトークン単位の分類貢献度を利用し、発言文から文節単位で政治的関心を持つ特徴的な表現を抽出する手法を提案する

データセット

分析対象は 2015 年から 2019 年の福岡県議会本会議の会議録である。これを福岡県議会データセットと呼び、発言者の分類モデルおよび特徴表現抽出に用いる。本研究では、政治的主張を含むと考えられる議員の発言に着目し、それらを対象に分析を行う。発言者の分類およびその結果を基にした特徴の抽出には一定のデータ量が必要であるため、発言文が 500 個以上の議員 26 名（合計 24,707 文）を対象とする。これを無作為にシャッフルしたうえで、訓練データ用として 80 %，検証データ用として 10 %，評価データ用として残り 10 % に分割した。

分類モデル (BERT)

本研究では事前学習モデルとして、東北大学が公開している“bert-base-japanese-v2”を用いた。福岡県議会データセットの訓練データを用いて、発言者分類のファインチューニングを行った。

結果の可視化 (SHAP)

評価用データを分類モデルに適用した結果に対して SHAP を用いることで、入力文における各文節の分類への貢献度を算出する。SHAP は分類貢献度である SHAP 値をトークン単位で算出するため、算出された各トークンの SHAP 値を用いて、“[SEP]”トークンで区切られる各文節に属するトークンの SHAP 値の中で最大の値を文節の貢献度と見なす。

研究事例

分類精度

福岡議会データセットによる発言者分類の正解率は 50.4 % となった。

議員の政治的関心の抽出

SHAP 値が高い文節は分類器が発言者を推定する際の手がかりとして貢献していることから、その発言者の発言内容の特徴（他の発言者による発言内容との違い）が表出する可能性があると考えられる。そこで、SHAP 値が高い文節に発言者（議員）の政治的関心（取り組もうとしている施策）が現れているかを観察する。

表3 SHAP による出力と評価（中村誠治議員の例）

No.	SHAP 値が高い文節	係り先文節	評価
1	生物多様性の	保全に	1.75
2	生物多様性の	重要性を	1.75
3	生態系の	機能の	1.75
4	光合成による	生産量は	0.75
5	中小企業へ	行います。	0.75
6	世界の	人口が	0.75
7	マイナンバー制度における	本県の	1.25
8	ASEAN 各国の	邦人会、	1.25
9	農業大学の	必要性和	1.75
10	可能性は	ある。	0
11	設定されています。		0
12	寄附を	集めることができるのか	0.5
13	輸血と	いえば、	0.75
14	バイオマス燃料等への	作物利用法の	2
15	登壇	ございます。	0
16	予備です。		0
17	まいりました。		0
18	貴重な	血液製剤です。	1.25
19	十人も減っています。		0

議員の政治的関心の抽出の評価方法

著者ら 4 人が、観察対象文において SHAP 値が最も高かった文節とその係り先（親文節）を「政治的関心を適切に表している」「政治的関心を適切に表しているとまでは言えないが、関連するキーワードが含まれている」「政治的関心を表すキーワードは含まれていない」の 3 段階で評価し、それぞれを 2 点、1 点、0 点で得点化する。

議員の政治的関心の抽出の評価方法の結果

SHAP による手法に対する評価の平均は 0.835 であった。

具体例を見ると、「生物多様性」「生態系」「農業大学」「バイオマス燃料」などの政治課題に関連したキーワードが抽出されている。係り先の文節と組み合わせることによって、「生物多様性の保全」「農業大学の必要性」のように発言者の政治的関心を適切に読み取ることができる文節に高い評価点が付与されていた。

まとめ

- BERT ではテキストの特徴量 (感情値など) 抽出、SHAP では特徴抽出した結果の、単語レベルの特徴の可視化を行うことができる

できそうなこと

レビューサイトの単語レベルのネガポジ度を分析するシステムの提案

- レビューサイトのレビューをスクレイピング
- レビューサイトのレビューについて BERT を用いて文章の感情 (ポジティブ、ネガティブ) を分類
- SHAP を用いて、文章のどの部分が感情に強く影響しているか可視化する

やってること

- レビューサイトのレビュー文のスクレイピング
- BERT を用いてサンプル文章の特徴分析を行う
- SHAP の勉強