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

July 11, 2024

方向性

やりたいこと

自然言語処理について行っていく

- レビューの文章のネガポジ分析
  → BERT と SHAP を用いてレビュー文の単語レベルでのネガポジ分析を行い、その結果を可視化するシステムを提案

システムの流れ

1. レビュー文のスクレイピング
2. BERT による文章の感情値を分析
3. SHAP による分析結果の解釈
4. 結果の出力

BERT とは

BERT（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。

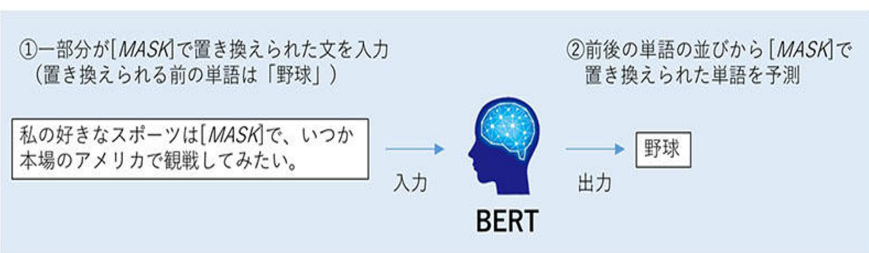
BERT とは

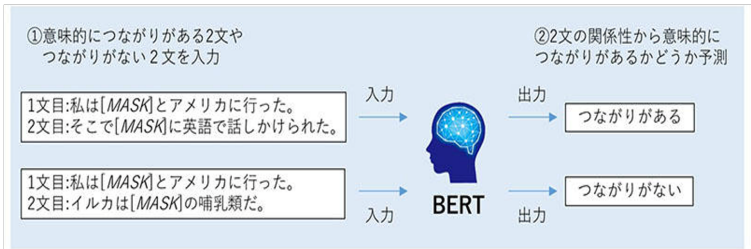
学習方法

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



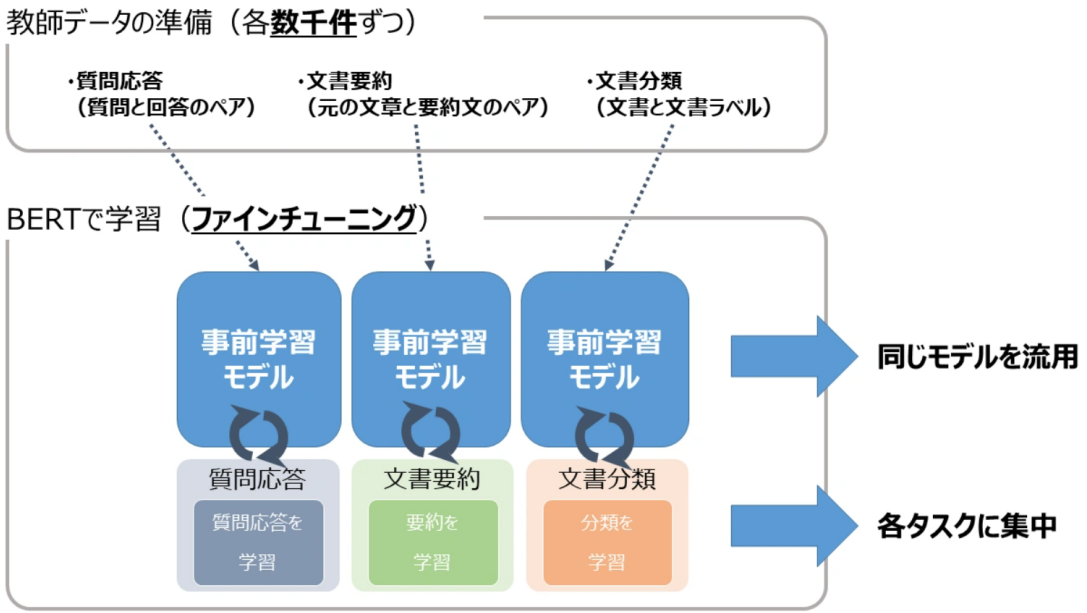


BERT とは

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

- 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 → 2 → 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のようになる

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 → 2 → 3	4	12	44
1 → 3 → 2	4	38	18
2 → 1 → 3	10	6	44
2 → 3 → 1	30	6	24
3 → 1 → 2	12	38	10
3 → 2 → 1	30	20	10

定式化

シャープレイ値の定式化

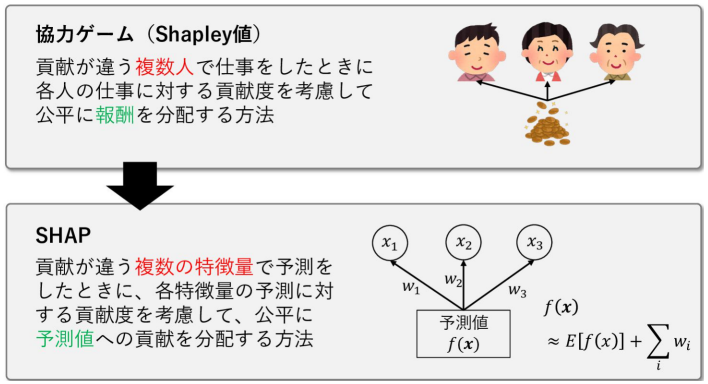
- 一般的には、プレイヤー i のシャープレイ値は次式によって定式化される。ただし、s は連携 S に含まれるプレイヤー数である。

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

SHAP について

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

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

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

進捗

実装に当たり必要なこと

- データセットの取得
  →解きたいタスクの学習用データに使用。
  ネガポジ分析の場合、ラベル付けしたネガティブな文章とポジティブな文章を大量に用意する
- 事前学習モデルの構築 (BERT)
  →「事前学習」「ファインチューニング」の二つの学習を行うモデルを構築する
- トークナイザーの構築 (BERT)
  →文章を語彙（トークン）に分割したうえで、BERT モデルに入力できる形に変換する処理
- SHAP の実装

進捗

行ったこと

BERT を用いて文章の感情値の分析を行った。

- モデル: 事前学習モデル「bert-base-japanese-sentiment-irony」を使用
- トークナイザー:「bert-base-japanese-whole-word-masking」を使用
- データセット:「multilingual-sentiments」を使用

```
from transformers import pipeline, AutoModelForSequenceClassification, BertJapaneseTokenizer,BertTokenizer, BertForSequenceClassification
from datasets import load_dataset


model = AutoModelForSequenceClassification.from_pretrained('kit-nlp/bert-base-japanese-sentiment-irony')
#model = AutoModelForSequenceClassification.from_pretrained('tohoku-nlp/bert-base-japanese-v3')
tokenizer = BertJapaneseTokenizer.from_pretrained('cl-tohoku/bert-base-japanese-whole-word-masking')
classifier = pipeline("sentiment-analysis",model=model,tokenizer=tokenizer)
dataset = load_dataset("tyqiangz/multilingual-sentiments", "japanese")

result = classifier("製品の品質は素晴らしいです。特にデザインが気に入りました。使いやすさも抜群です。")[0]
print(f"label: {result['label']}, with score: {round(result['score'], 4)}")

result = classifier("製品は期待外れでした。機能が不安定で、操作性も悪いです。購入を後悔しています。")[0]
print(f"label: {result['label']}, with score: {round(result['score'], 4)}")
```

```
label: ポジティブ, with score: 0.5552
label: ネガティブ, with score: 0.7491
```

- SHAP を適用することにより、出力結果のうち、どの単語が出力結果に影響しているかを可視化することができる。
- BERT のモデルを自分で作成していく予定

おわりに

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
- 自分ので BERT のモデルの構築
- SHAP の適応