

1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに

## 方向性と進捗

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

October 21, 2024

# 1. はじめに

2/17

## 研究の背景 (仮)

- SNS や Web サイトの利用者は増加し、個人のインターネット利用率は 80 % を越えており、大勢の人がインターネット上の情報に触れることができるようになった。
- 現代の消費行動の特徴としてインターネット上の口コミや評価を重視し、意思決定や行動に大きな影響を与えている。
- しかしインターネット上の口コミは対面のクチコミよりも利用可能な非言語的手がかりが乏しく、相手に伝わりにくい。

## 研究の目的 (仮)

BERT でレビューの感情分析を行い、分析結果に大きく影響を及ぼしている単語を SHAP で可視化を行う。そして、トピックごとにクラスタリングを行うことでユーザーにとってより解釈可能なレビューの分析を行うシステムを提案する。

## 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。

1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに

## 2. 関連研究?

3/17

1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに

## 出力

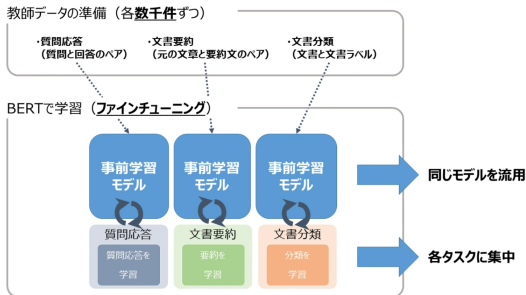
# BERT とは

5/17

## 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



1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに

# 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

7/17

- 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
$1 \rightarrow 2 \rightarrow 3$	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

1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに

## 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

### 協力ゲーム (Shapley値)

貢献が違う複数人で仕事をしたときに  
各人の仕事に対する貢献度を考慮して  
公平に報酬を分配する方法



### SHAP

貢献が違う複数の特徴量で予測を  
したときに、各特徴量の予測に対  
する貢献度を考慮して、公平に  
予測値への貢献を分配する方法



## 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  $\phi_i$  とすると以下のようにあらわす

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

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

1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに

1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに

## 文章のカテゴリ分類について

- レビューの文章を文節ごとに区切り、カテゴリ (味に関する文、価格に対する文、接客態度に関する文、など) ごとにクラスタリング分けを行う。

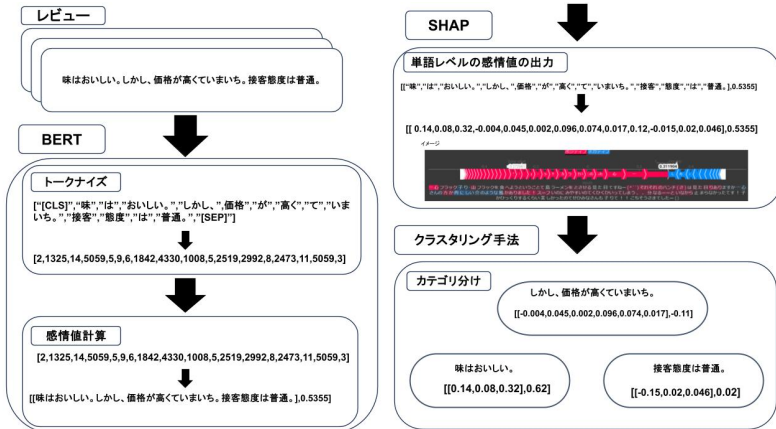
## クラスタリング手法

クラスタリング手法で定番の k-means 法か LDA を検討中

## 実装の流れ

- 1 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
- 2 分析レビューの取得  
→感情分析を行うためのレビューをスクレイピングする
- 3 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
- 4 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
- 5 SHAP の実装  
→ BERT で分析した結果を可視化する
- 6 レビュー文をクラスタリング
- 7 出力  
→可視化した結果をわかりやすく表示する

1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに



## 進捗 1(データセットの取得)

- 1 TIS が公開している感情分析を行うためのデータセット「chABSA-dataset」を取得できるようにした。
- 2 「chABSA-dataset」は上場企業の有価証券報告書 (2016 年度) をベースに作成されたデータセットで、各文に対してネガティブ・ポジティブのラベル付けがされている
- 3 このデータセットを BERT に読み込ませ、モデル構築を行う

text	labels
売上収益は、抗がん剤「レンビマ」(欧州における腎細胞がんに係る製品名「Kisplyx」)および抗てん	1
セグメント別には、日本医薬品事業およびアジア医薬品事業が増収となりました	1
また、すべての海外セグメントにおいて現地通貨ベースで成長を果たしました	1
グローバルブランド4品目合計では、為替の影響を受ける中で、前期から14.5%増の728億22百万円と	1
営業利益は、オペレーションの高質化・効率化に加え、EAファーマ株式取得に伴う一時収益(割安購	1
親会社の所有者に帰属する当期利益は、393億58百万円(前期比28.4%減)となりました	0
基本的1株当たり当期利益は、137円63銭(前期より54円61銭減)となりました	0
当期利益にその他の包括利益を加減した当期包括利益は、前期に円高の進行によって為替換算差額が	1
売上収益の内訳は、医療用医薬品が2,439億99百万円(同4.3%増)、ジェネリック医薬品が280億27百万	1
品目別売上収益については、ニューロロジー領域で、不眠症治療剤「ルネスタ」が80億12百万円(前期	1
ファイザー社と共同販促を展開している疼痛治療剤「リリカ」の共同販促収入は242億68百万円(同1.8	0
オンコロジー領域では、「ハラヴェン」が77億63百万円(同14.2%増)、「レンビマ」が27億15百万円	1
さらに、ヒト型抗ヒトTNFαモノクローナル抗体「ヒュミラ」は376億62百万円(同15.4%増)と順調に	1
売上収益は、1,172億17百万円(前期比4.1%減、現地通貨ベースでは6.3%増)となりました	0
セグメント利益は、効率的なマーケティング活動による販売管理費の減少および前期に発生した米国	1
品目別売上収益については、ニューロロジー領域で、抗てんかん剤「Banzel」が138億48百万円(前期	1
「Belvq」は37億14百万円(同16.0%減)でした	0
オンコロジー領域では、制吐剤「Aloxi」が480億82百万円(同12.1%減)、「ハラヴェン」が166億19百	0
売上収益は492億74百万円(前期と同等水準、現地通貨ベースでは17.0%増)、セグメント利益は136億96	1

1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに

## 進捗 2(分析レビューの取得)

1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに

- 1 食べログのラーメンのレビューをスクレイピングし、利用できるようにした。
- 2 1つの店に対して上位 20 件分を取得し、感情分析をして、その店の分析結果を表示する

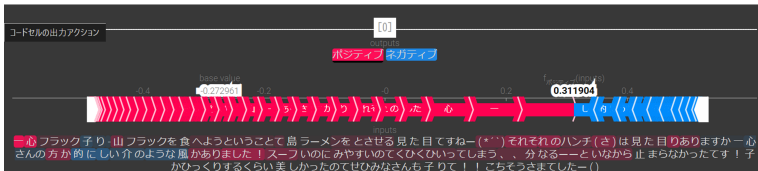
**Source** Nani Review  
貴族食堂 水見屋煮子ラーメン 1000円今日は、富山県を代表する富山の人気を誇るラーメン屋である貴族さんにお邪魔しました！初見で読むことが不可能なお店の名前にもちゃんと旅行中、通りかかった水見にて美味しさを伝えているところ。断った点の客が店に遭遇、すくじに向かいました。平日13:20:15の列に接続、すぐに由緒あるおやつで貴族の朝に伺いました。6時半到着で7番目まで入行。朝は涼しいので待ち時間苦しくなりました。1回目でお店。7時に20人以上の待ち。スプーンが大量にある！自覚めたはず貴族 富山、富山旅行の初日は水見！湯通と丼物と美味しい、折角だから10名ほど見てみたいかな？という事で水見駅からタクシーで店へ。11時45分到着する58人(赤ちゃん含む)待ち 今日のお店は富山県水見町の貴族(としん)さんにお伺いします。営業時間：11:00-15:00 / 23:00-23:00 定休日：月曜 日曜日：カウンター10席 / テーブル席 駐車場 貴族 〇に入った点・旨みだけを抽出した透明感のある煮汁スープ〇追加情報：御一考があった専用メニューのように：軽便車乗る〇ルール：列は道路を挟んだ向かい側の外〇に入った点・旨みだけを抽出した透明感のある煮汁スープ・様々な具材の彩りある旨味をまよくまとめた白濁スープ〇追加情報：1時間待ち待たせ方をしよう！ 貴族 〇に入った点・旨みだけを抽出した透明感のある煮汁スープ●注意：1時間は待つ覚悟をしましょう！！〇駐車場の〇ルール：列は道路を挟んだ向かい側から並ぶ 貴族 貴族前に訪ねました。今回朝ラーメンしているのとこの朝に訪問風ラーメンを注文したスプーンは美味しかったように思いますが美味しくてうしろでは糖は糖味を使用して 貴族前には訪問白濁油の外売りに付けた状態です……………湯通しに介介が付けて美味しく使います。また、白濁油にコウが良いです。福中は味噌味噌用紙が便利です。チャイ 水見屋煮子ラーメン1000貴族前水見水見に移動して「貴族」に念願の初訪問。店の前に二人、向かいに一人居てましたが店内案内係実質無し。営業家オスメスの 貴族 水見3スプーン対応38番間4.0レビュー★あり更新されていなかったようであり適当営業家はゆずやぶとの噂においては最近始まった朝ラーメンを試そうと 水見屋煮子ラーメン1000★★★★★☆☆☆☆☆☆富山県を取りつて富山訪問。富山県No.1の評価で、水見らBMWにたちまち。平日12時半の時点で客と25人の行列が。1時間前 最高に美味かつてごちそうでしたBuono!-私はお店を評価できるような者ではないのでジャッジし人数を付けての評價は控えてさせていただきます。 貴族 最初に美味かつてごちそうでしたBuono!-私はお店を評価できるような者ではないのでジャッジし人数を付けての評價は控えてさせていただきます。 貴族 土日は朝ラーメンをやっていると聞きました。こちら午前百名店「貴族(としん)」さんです。開店以前前の時刻から11時に準備していました。ちょっと物を置いて 貴族 さっさと来たに当たらないと思ひ、貴族さんへ質問しました。富山駅から高岡駅乗り換えで約60分、水見駅に約10分の距離です。お店は開店2時間前の 以前から行きたいと思っていた。貴族(としん)さん、富山のラーメン名店の本店です。水見屋の前で近くは国道415号線一本となっており、人気の多い。幸々 幸々車庫内「水見」のトヨタレンタカーが日曜に朝ラーメンを食べたのが行方不明でした。顔の写真をみて近づく道でなくと奪い取りました。ほんとは……、開店後

## 進捗 3(BERT モデルの構築、SHAP の実装)

- 1 自分でモデルを作成し、レビュー文に適応させ、感情値を出力できるようにした。
- 2 BERT で出力した結果対して、SHAP で可視化できるようにした。

※ テストデータ

```
test_texts = ["***【一心ブラック煮玉子入り ¥980】 富山ブラックを食べようということで訪問徳島ラーメンを彷彿とさせる見た目ですね(・_・)味それぞれ醤油のパンチ(濃さ)は見た目通りありますが一心さんの方が全体的に優しい魚介のような風味がありました！スープ濃いのに飲みやすいのでくびくびいってしまう、塩分過多なるーと思いながら止まらなかったです！煮玉子がびっくりするくらい美味しかったのでぜひみなさんも煮玉子入りで！！ごちそうさまでしたー(・_・)"]
test_encodings = tokenizer(test_texts, truncation=True, padding=True, max_length=128, return_tensors='pt')
```



- 1 出力結果の精度がまだよくないので、モデルの設定を変えて精度を上げる
- 2 文章をカテゴリごとに分類できていないため、レビュー文をカテゴリごとに分類する手法を検討

1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに

1. はじめに
2. 関連研究?
3. BERT と SHAP について
4. 提案手法
5. 数値実験ならびに考察
6. おわりに

## 有効性の検証

- 1 ほかの感情分析の手法も実装し、どれだけ優れているか比較する
- 2 実際にシステムを利用してもらい、どれだけつかいやすかったかアンケートを取る
- 3 ほかに有効性を示す手法があれば検討



### やったこと

- レビューサイトのレビューをスクレイピングした
- 自分で BERT モデルを構築した
- BERT を用いて文章の感情 (ポジティブ、ネガティブ) を分類した
- 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した

### これからやること

- BERT のモデルの調整
- レビュー文をカテゴリ (味に対する文、価格に対する文) ごとに分類分けをする
- SHAP の出力結果をより分かりやすく表示できるようにする