

# Word-level Polarity is All You Need?: 解釈可能なニューラルネットワークモデルを利用した 単語極性変換による効率的な金融センチメント適合

伊藤 友貴<sup>1</sup>

<sup>1</sup> 三井物産株式会社

{Tomok.ito@mitsui.com}

## 概要

金融分野へのサービス提供時等において、深層学習モデルのドメイン適合実施時において、「計算コスト」や「更新されたパラメータのブラックボックス性」が課題になることがある。特に金融分野では単語レベルでのポジネガが時代やドメインが変わると変化することも多く、本課題は当該分野では重要な課題である。そこでニューラルネットワークモデル SINN を活用した単語極性変換手法 Word-level Polarity Adaptation framework based on SINN (WPAS) を提案する。景気センチメントに関するデータセットを利用した検証の結果、提案手法 WPAS により (1) 少ないパラメータ数の更新、かつ (2) 更新されたパラメータが解釈可能な形で高性能なドメイン適合ができることを実証した。

## 1 はじめに

### 金融分野におけるドメイン適合時の課題

深層学習モデルは強力なモデルである一方、学習時の計算コストが課題になることが多い。特に、金融分野では単語レベルでのポジネガが時代やドメインが変わると変化することも多く、例えば、「特定ドメイン (ソースドメイン) のデータ」を用いて学習されたモデルが別のドメインでは使えないことも十分にあり得る。このような状況の場合、再度新しくターゲットドメインのデータを追加データとしてモデルを再学習し、ドメイン適合をさせることが多い。一方、このような「愚直な」ドメイン適合には計算コストがかかり、実サービスの提供時等にはユーザーの利便性を損ねる危険がある他、運用負担もかかる。特に近年、言語モデルのサイズは増加傾向 [1] であり、この学習時の計算コストに関する課

題は今後も拡大すると思われる。このような背景の下、近年、LoRA[2] や知識編集 [3, 4] を始め、「少ないパラメータ数」で効率よくニューラルネットワークモデルを学習させる方法がいくつか提案されている。これらの手法は「学習の効率的性」という面で有用であるものの、「更新されるパラメータの解釈が難しい」という側面もあり、サービス運用時や提供時に支障をきたす可能性もある。

## 本研究の目的

そこで、本研究では、ソースドメインの学習後、ターゲットドメインへの適合時に、以下の2つの要件を予測性能を保持しつつも満たすようにドメイン適合を行えるフレームワークの構築を目指す。

**要件 1:** 少ないパラメータ数のみの更新でターゲットドメインへの適合できる

**要件 2:** ターゲットドメインへの適合時に更新されるパラメータが解釈可能である

## 提案アプローチ

本目的達成のため、本研究では「解釈可能なニューラルネットワークモデル [5]」を活用した単語極性変換手法 Word-level Polarity Adaptation framework based on SINN (WPAS) というドメイン適合のフレームワークを提案する。WPAS 法では、予測モデルとして [5] や [6], [7] 等で提案されている Sentiment Interpretable Neural Network (SINN) を採用する。SINN とは [5] にて提案された、各層が「単語のセンチメント」「極性反転」「大域的重要度」の意味で解釈可能なニューラルネットワークである (図 1)。提案手法では SINN における「各層毎にセンチメントを種類別に分離されている」性質を活用し「要件 1」及び「要件 2」を満たすことを試みる。一般に、「単語のセンチメント」や「単語の意味」

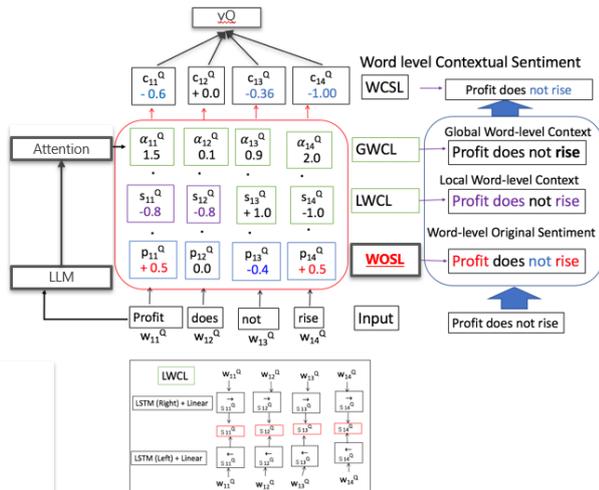


図1 SINN. 提案手法 WPSA ではターゲットドメインへの適合時に WOSL の層のみ更新する。

に比べ、文脈由来の「否定」や「反転」, 「強調」等の表現はどのドメインでも共通である可能性が多い。よって、これらを表現する層はどのドメインでも共有可能である可能性が高い。本仮説の下, 提案手法では「否定」や「反転」, 「強調」等を表現する層のパラメータは更新せず, 「単語のセンチメント」を表す層 (WOSL) のみを更新する。WPAS により, 少ないパラメータ数のみの更新でターゲットドメインへの適合 (要件 1) ができる。また, ドメイン適合時に更新されるパラメータは WOSL のみ, 即ち「単語センチメント値のみ」なので, 解釈可能 (要件 2) となることが期待される。

本研究の貢献は以下の通りである。

- (1) 要件 1 及び要件 2 を満たすドメイン適合フレームワーク WPAS 法を提案した。
- (2) 提案手法の有効性を実テキストデータを用いて実証した。

## 2 関連研究

深層学習モデルのブラックボックス性に関する研究として「ニューラルネットワークモデルの解釈」に関する研究 [8, 9, 10, 11] や解釈可能なニューラルネットワークの構築 [5, 6, 7, 12, 13, 14] が挙げられる。また, 効率的な言語モデルのドメイン適合に関する研究として, 知識編集 [3, 4] やモデルの量子化 [15], 更新パラメータの次元削減等により少ないパラメータ数で学習させる [2, 16], あるいはモデルマージ [17] 等のアプローチが提案されている。これらは様々な場面で有用である一方, 「更新されるパラメータの解釈が難しい」という側面もあり,

サービスの運用時や提供時に支障をきたす可能性もある。

## 3 SINN

提案フレームワークの説明の前に, 解釈可能なニューラルネットワークモデル SINN を紹介する。

SINN は訓練データ  $\Omega_{train}^{source}$  及び小規模な単語のセンチメントスコア辞書を用いた学習 Lexical Initialization Learning (LEXIL) [5] により構築可能である。ここで,  $\mathbf{Q}_n$  はレビュー,  $d_i^{\mathbf{Q}_n}$  はセンチメントタグ (1: ポジティブ, 2: ネガティブ) である。

SINN は Word-level Original Sentiment layer (WOSL), Local Word-level Contextual layer (LWCL), Global Word-level Contextual layer (GWCL), Word-level Contextual Sentiment layer (WCSL) から構成され, 入力テキスト  $\mathbf{Q} = \{w_t^{\mathbf{Q}}\}_{t=1}^n$  のポジネガ予測結果  $y^{\mathbf{Q}} \in \{0(\text{negative}), 1(\text{positive})\}$  を出力する NN である。本論文ではコーパスに出現する語彙数  $v$  の語彙集合を  $\{w_i\}_{i=1}^v$ , 単語  $w_i$  の語彙 ID を  $I(w_i)$ ,  $w_i^{em} \in \mathbb{R}^e$  を単語  $w_i$  の次元  $e$  の用意されたコーパスから計算された分散表現とし, さらに  $\mathbf{W}^{em} \in \mathbb{R}^{v \times e} := [w_1^{emT}, \dots, w_v^{emT}]^T$  とする。

**WOSL:** この層ではコメント  $\mathbf{Q} = \{w_t^{\mathbf{Q}}\}_{t=1}^n$  の各単語をその単語が文脈に左右されずに持つセンチメント値, オリジナルセンチメント値に変換する。

$$p_t^{\mathbf{Q}} := w_{I(w_t^{\mathbf{Q}})}^p \quad (1)$$

ここで,  $\mathbf{W}^p \in \mathbb{R}^v$  は各単語のオリジナルセンチメント値を表す。  $w_i^p$  は  $\mathbf{W}^p$  の  $i$  番目の要素を表し,  $w_i^p$  の値が  $w_i$  のオリジナルセンチメント値に対応する。

**LWCL:** この層では各単語  $w_t^{\mathbf{Q}}$  をセンチメントの反転に関する値へ変換する。

まず, レビュー  $\mathbf{Q}$  内の単語  $\{w_t^{\mathbf{Q}}\}_{t=1}^n$  を埋め込み表現  $\{e_t^{\mathbf{Q}}\}_{t=1}^n$  に変換する。

その後, 以下のように, 順方向及び逆方向の単方向の言語モデル  $\overrightarrow{\text{CLM}}_1$  及び  $\overleftarrow{\text{CLM}}_1$  によって, 前後の単語列が各単語  $w_t^{\mathbf{Q}}$  へ与える「反転」に関する影響を表す値  $\overrightarrow{s}_t^{\mathbf{Q}}$  及び  $\overleftarrow{s}_t^{\mathbf{Q}}$  へ変換する。

$$\overrightarrow{h}_t^{\mathbf{Q}} := \overrightarrow{\text{CLM}}_1^{DEC}(w_1^{\mathbf{Q}}, w_2^{\mathbf{Q}}, \dots, w_{t-1}^{\mathbf{Q}}), \quad (2)$$

$$\overleftarrow{h}_t^{\mathbf{Q}} := \overleftarrow{\text{CLM}}_1^{DEC}(w_{t+1}^{\mathbf{Q}}, w_{t+2}^{\mathbf{Q}}, \dots, w_n^{\mathbf{Q}}), \quad (3)$$

$$\overrightarrow{s}_t^{\mathbf{Q}} = \tanh(v^{leftT} \cdot \overleftarrow{h}_t^{\mathbf{Q}}), \quad \overleftarrow{s}_t^{\mathbf{Q}} = \tanh(v^{rightT} \cdot \overrightarrow{h}_t^{\mathbf{Q}}). \quad (4)$$

ここで,  $v^{right}, v^{left} \in \mathbb{R}^e$  はパラメータであり,  $\overrightarrow{\text{CLM}}_1^{DEC}$  及び  $\overleftarrow{\text{CLM}}_1^{DEC}$  はそれぞれ  $\overrightarrow{\text{CLM}}_1$  及び

$\overleftarrow{\text{CLM}}_1$  によって出力される最終層への変換を表し、LSTM 等を利用することを想定する。その後、

$$s_t^{\mathbf{Q}} := \overrightarrow{s}_t^{\mathbf{Q}} \cdot \overleftarrow{s}_t^{\mathbf{Q}} \quad (5)$$

の形で前後の単語列が各単語  $w_t^{\mathbf{Q}}$  へ与える「反転」に関する影響を表すスコアを算出する。

**GWCL:** この層は各単語  $w_t^{\mathbf{Q}}$  をその文脈内での強調スコアへ変換する。まず、順方向及び逆方向の単方向の言語モデル  $\overrightarrow{\text{CLM}}_2$  及び  $\overleftarrow{\text{CLM}}_2$  によって、前後の単語列が各単語  $w_t^{\mathbf{Q}}$  へ与える「強弱」に関する影響を表す値  $\overrightarrow{\alpha}_t^{\mathbf{Q}}$  及び  $\overleftarrow{\alpha}_t^{\mathbf{Q}}$  へ変換する。

$$\overrightarrow{\alpha}_t^{\mathbf{Q}} = \tanh(\overrightarrow{\text{CLM}}_2^{\text{att}}(w_1^{\mathbf{Q}}, w_2^{\mathbf{Q}}, \dots, w_{t-1}^{\mathbf{Q}})), \quad (6)$$

$$\overleftarrow{\alpha}_t^{\mathbf{Q}} := \tanh(\overleftarrow{\text{CLM}}_2^{\text{att}}(w_{t+1}^{\mathbf{Q}}, w_{t+2}^{\mathbf{Q}}, \dots, w_n^{\mathbf{Q}})). \quad (7)$$

ここで、 $\overrightarrow{\text{CLM}}_2^{\text{att}}$  及び  $\overleftarrow{\text{CLM}}_2^{\text{att}}$  は  $\overrightarrow{\text{CLM}}_2$  及び  $\overleftarrow{\text{CLM}}_2$  による各単語へのアテンションへの変換をし表し、GPT 等の利用を想定する。その後、

$$\alpha_t^{\mathbf{Q}} := \overrightarrow{\alpha}_t^{\mathbf{Q}} \cdot \overleftarrow{\alpha}_t^{\mathbf{Q}} \quad (8)$$

の形で前後の単語列が各単語  $w_t^{\mathbf{Q}}$  へ与える「強調」に関する値を算出する。

**WCSL:** この層では WOSL 及び WCL の値を各単語の文脈センチメント  $\{c_t^{\mathbf{Q}}\}_{t=1}^T$  へ変換する。

$$c_t^{\mathbf{Q}} := p_t^{\mathbf{Q}} \cdot s_t^{\mathbf{Q}} \cdot \alpha_t^{\mathbf{Q}}. \quad (9)$$

ここで、 $s_t^{\mathbf{Q}}$  及び  $\alpha_t^{\mathbf{Q}}$  は以下で定義する。

$$s_t^{\mathbf{Q}} := \overrightarrow{s}_t^{\mathbf{Q}} \cdot \overleftarrow{s}_t^{\mathbf{Q}}, \alpha_t^{\mathbf{Q}} := \overrightarrow{\alpha}_t^{\mathbf{Q}} \cdot \overleftarrow{\alpha}_t^{\mathbf{Q}} \quad (10)$$

**出力:** 最後に文の極性  $y^{\mathbf{Q}}$  を以下のように出力する

$$y^{\mathbf{Q}} = \sum_{t=1}^T c_t^{\mathbf{Q}}.$$

ここで、 $y^{\mathbf{Q}} > 0$  は  $\mathbf{Q}$  がポジティブであることを表し、 $y^{\mathbf{Q}} < 0$  は  $\mathbf{Q}$  がネガティブであることを表す。

## 4 WPAS (提案手法)

今回提案するドメイン適合フレームワーク Word Polarity Domain Adaptation Framework based on SINN(WPAS) は Algorithm 1 の通りである。1 行目から 5 行目がソースドメインの学習、6 行目以降がターゲットドメインへの適合に該当する。

### ソースドメインの学習

ソースドメインの学習 (1-5 行目) では 2 行目のように予め用意した単語の極性 (ポジティブ 1, ネガティブ -1) に関する辞書 (極性辞書) による初期化を用いた学習 LEXIL を利用する。ここで、 $PS(w_i)$  は

### Algorithm 1 Word Polarity Domain Adaptation Framework based on SINN(WPAS)

---

```

1: for  $i \leftarrow 1$  to  $v$  do
2:    $w_i^p \leftarrow \begin{cases} PS(w_i) & (w_i \in S^d) \\ 0 & (\text{otherwise}) \end{cases}$ 
3: for  $i \leftarrow 1$  to  $epoch$  do
4:   for  $\mathbf{Q} \in \Omega_{train}^{source}$  do
5:     Update  $\mathbf{W}^p, \overrightarrow{\text{CLM}}_1, \overleftarrow{\text{CLM}}_1, \overrightarrow{\text{CLM}}_2, \overleftarrow{\text{CLM}}_2, \mathbf{v}^{right}, \mathbf{v}^{left}$  based on the gradient value by  $SCE(y^{\mathbf{Q}}, d^{\mathbf{Q}})$ ;
6: for  $i \leftarrow 1$  to  $epoch$  do
7:   for  $\mathbf{Q} \in \Omega_{train}^{target}$  do
8:     Update only  $\mathbf{W}^p$  based on the gradient value by  $SCE(y^{\mathbf{Q}}, d^{\mathbf{Q}})$ ;

```

---

単語  $w_i$  のセンチメント辞書値であり、 $S^d$  はセンチメント辞書内単語の集合である。 $SCE(a, b)$  は  $a$  と  $b$  のシグモイド損失である。LEXIL を使うことで、 $S^d$  内の各単語のセンチメント辞書値が正しいとき、十分小さい  $\delta$  に対し、式 11 を満たす単語については SINN の各層における解釈性が担保され、WOSL から取得できるオリジナル単語センチメントが正しくなることが保証されている [5]

$$\min_{w_i \in S^d} \|w_i^{em} - w_j^{em}\|_2 < \delta \quad (11)$$

## ターゲットドメインへの適合

ターゲットドメインへの適合 (6-8 行目) では 8 行目のように  $\mathbf{W}^p$  のみ更新させる。

## 5 評価実験

「ソースドメインの学習 → ターゲットドメインへの適合 (ドメイン適合タスク)」を実施する際の提案手法 WPAS の有効性を実データを用いて評価した。

### 5.1 評価方法

今回の検証では、下記の「Now → Future タスク」と「Future → Now タスク」という二つのドメイン適合タスクに関する検証を行った。

**Now → Future タスク:** 「現状の景気」に関するポジネガ分析のデータセット (EcoRev1) をソースドメインとして学習し、「景気見通し」に関するポジネガ分析のデータセット (EcoRev2) へのドメイン適合がうまくできるかを評価するドメイン適合タスク。

**Future → Now タスク:** EcoRev2 をソースドメインとして学習し、EcoRev1 へのドメイン適合がうまくできるかを評価するドメイン適合タスク。

**データセット:** 本検証のため、景気ウォッチャー調査の現状に関する日本語コメントの「現状の景気」に関するポジネガ分析のデータセット [5] をもとに EcoRev1 及び EcoRev2 を構築した。EcoRev1 及び EcoRev2 には訓練データ、検証データ、テストデータから構成され、各データセットにポジティブコメント及びネガティブコメントがそれぞれ 20,000 件、2,000 件、4,000 件格納されている。

提案手法を以下の 3 つの切り口から評価した

**評価 1: 更新パラメータ数が少ないか?:** 本評価ではターゲットドメインへの適合時に更新対象となるパラメータ数をもとに評価した。

**評価 2: 「解釈性」が担保されているか?:** 本評価では [5] にて提供されている人手で作成された、単語レベルでのポジネガリスト及び極性反転に関するデータセットをもとに WOSL を評価した。即ち、WOSL から得られるリスト内単語の極性（ポジティブ 149 件、ネガティブ 165 件）と単語極性リスト内の極性の一致度 (macro  $F_1$  値) をもとに評価した。

**評価 3: 予測性能が十分に高いか?:** 本評価ではターゲットドメインのテストデータに関するポジネガ分類性能をもとに評価した。評価指標として  $F_1$  値 (マクロ平均) を用いた。

## 5.2 ベースライン

提案手法の性能評価、特に WPSA の特徴が「ターゲットドメインの適合時に  $W^P$  のみを学習対象とすること」にあるため、その有効性を評価するため、提案手法のアプローチ (WPSA) と下記に挙げるベースラインとなるアプローチの結果を比較した。

**Word Tuning (WT):** SINN を用いてソースドメインのデータセットを学習後、ターゲットドメインのデータセットを学習するアプローチ。但し、更新パラメータは  $W^P$  のみとする。

**ソースのみ:** SINN を用いてソースドメインのデータセットのみを学習するアプローチ

**Full Fine-Tuning (FFT):** SINN を用いてソースドメインのデータセットを LEXIL により学習後、ターゲットドメインのデータセットを学習するアプローチ。更新対象は SINN の埋込表現以外の全パラメータ。

## 5.3 その他の実験設定

Lexicon Initialization においては「”上がる”, ”回復”, ”上方”, ”増加”, ”上昇」及び「”減少”, ”低下”, ”損失”, ”遅れ”, ”リスク”」についてそれぞれ

+1 と -1 を入れる形で初期化を実施した。また、Tokenizer は rinna/japanese-roberta-base [18] を利用、 $LLM_1$  には LSTM [19] (中間層の次元 768),  $LLM_2$  には rinna/japanese-gpt2-small [20, 21] を利用した。

## 5.4 実験結果・考察

表 1 及び表 2 が評価結果である。提案手法 WPSA により、少ないパラメータ数の更新でも FFT と同等の予測性能 (評価 3) を解釈性を担保 (評価 2) しつつ出すことに成功していることがわかる。

表 1 評価結果 (Now2Future タスク)

	パラメータ数 (評価 1)	解釈性 L (評価 2)	予測性能 (評価 3)
WT	0.03 M	0.826	0.867
ソースのみ	-	0.860	0.914
FFT	256 M	0.793	0.929
<b>WPSA</b>	<b>0.03M</b>	<b>0.839</b>	<b>0.932</b>

表 2 評価結果 (Future2Now タスク)

	パラメータ数 (評価 1)	WOSL (評価 2)	予測性能 (評価 3)
WT	0.03 M	0.834	0.905
ソースのみ	-	0.821	0.902
FFT	128 M	0.817	0.921
<b>WPSA</b>	<b>0.03 M</b>	<b>0.818</b>	<b>0.920</b>

また、Appendix A のように SINN を活用することで更新されたパラメータの可視化ができる。

## 6 結論

本研究では、大規模なニューラルネットワークのドメイン適合には (1) 計算コストがかかり、かつ (2) 更新されたパラメータがブラックボックスであるという課題解決のため、解釈可能なニューラルネットワークモデル SINN を活用した単語極性変換手法 Word-level Polarity Adaptation framework based on SINN (WPAS) を提案した。景気センチメントに関するデータセットを利用した検証の結果、提案手法により (1) 少ないパラメータ数の更新のみでドメイン適合可能、(2) 更新されるパラメータが解釈可能、かつ (3) 予測性能を保持可能であるようなドメイン適合が可能であることを実証した。

今後の展開として、異なる言語間のドメイン適合や、離れたドメイン間での適合性能の検証や逆向き言語モデル [22] を利用した性能改善、センチメント分析以外のタスクへの拡張が考えられる。

## 参考文献

- [1] Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. A survey on evaluation of large language models. **ACM Trans. Intell. Syst. Technol.**, Vol. 15, No. 3, 2024.
- [2] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In **International Conference on Learning Representations**, 2022.
- [3] Knowledge mechanisms in large language models: A survey and perspective. In **EMNLP 2024 Findings**, 2024.
- [4] Editing large language models: Problems, methods, and opportunities. In **EMNLP 2023**, 2023.
- [5] Tsubouchi K. Sakaji H. Yamashita T. Ito, T. and K. Izumi. Word-level contextual sentiment analysis with interpretability. In **AAAI 2020**, 2020.
- [6] Tsubouchi K. Sakaji H. Yamashita T. Ito, T. and K. Izumi. Csn: Contextual sentiment neural network. In **IEEE ICDM 2019**, 2019.
- [7] Tsubouchi K. Sakaji H. Yamashita T. Ito, T. and K. Izumi. Sentiment shift neural network. In **SDM 2020**, 2020.
- [8] S. Bach, A. Binder, G. Montavon, F. Klauschen, K. R. Muller, and W. Samek. On pixel-wise explanations for nonlinear classifier decisions by layer-wise relevance propagation. **PLOS ONE**, Vol. 10, No. 7, pp. 1–46, 2017.
- [9] L. Arras, G. Montavon, K. R. Muller, and W. Samek. Explaining recurrent neural network predictions in sentiment analysis. In **EMNLP Workshop**, 2017.
- [10] M. T. Ribeiro, S. Singh, and C. Guestrin. "why should i trust you?" explaining the predictions of any classifier. In **KDD**, 2016.
- [11] A. Shrikumar, P. Greenside, and A. Kundaje. Learning important features through propagating activation differences. In **ICML**, 2017.
- [12] T. Ito, H. Sakaji, K. Tsubouchi, K. Izumi, and T. Yamashita. Text-visualizing neural network model: Understanding online financial textual data. In **PAKDD 2018**, 2018.
- [13] Q. Zhang X. Huang Y. Zou, T. Gui. A lexicon-based supervised attention model for neural sentiment analysis. In **COLING 2018**, 2018.
- [14] Z. Quanshi, Y. N. Wu, and S. C. Zhu. Interpretable convolutional neural networks. In **CVPR 2018**, 2018.
- [15] Ruihao Gong, Yang Yong, Shiqiao Gu, Yushi Huang, Chengtao Lv, Yunchen Zhang, Dacheng Tao, and Xianglong Liu. LLMC: Benchmarking large language model quantization with a versatile compression toolkit. In **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track**, pp. 132–152, 2024.
- [16] Sourab Mangrulkar, Sylvain Gugger, Lysandre Debut, Younes Belkada, Sayak Paul, and Benjamin Bossan. Peft: State-of-the-art parameter-efficient fine-tuning methods. <https://github.com/huggingface/peft>, 2022.
- [17] Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. In **The Eleventh International Conference on Learning Representations**, 2023.
- [18] Tianyu Zhao and Kei Sawada. rinna/japanese-roberta-base.
- [19] S. Hochreiter and Jurgen Schmidhuber. Long short-term memory. **Neural computation**, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [20] Kei Sawada, Tianyu Zhao, Makoto Shing, Kentaro Mitsui, Akio Kaga, Yukiya Hono, Toshiaki Wakatsuki, and Koh Mitsuda. Release of pre-trained models for the Japanese language. In **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, pp. 13898–13905, 5 2024.
- [21] Tianyu Zhao and Kei Sawada. rinna/japanese-gpt2-small.
- [22] Takumi Goto, Hiroyoshi Nagao, and Yuta Koreeda. Acquiring bidirectionality via large and small language models. In **Proceedings of the 31st International Conference on Computational Linguistics (COLING2025)**, 2025.

# A SINNによるレビューの解釈結果の可視化例

参考として、SINNによるレビューの解釈結果の可視化例を記載する。例示されているように、SINNを使うとセンチメント分析の流れが理解できる。

```
-----WOSL-----
・ここ数か月間、夜の来客数が激減している。それと共に客単価も下がっているため、非常に良くない。
-----LWCL-----
--> (LWCL)
・ここ数か月間、夜の来客数が激減している。それと共に客単価も下がっているため、非常に良くない。
<-- (LWCL)
・ここ数か月間、夜の来客数が激減している。それと共に客単価も下がっているため、非常に良くない。
total (LWCL)
・ここ数か月間、夜の来客数が激減している。それと共に客単価も下がっているため、非常に良くない。
-----GWCL-----
--> (GWCL)
・ここ数か月間、夜の来客数が激減している。それと共に客単価も下がっているため、非常に良くない。
<-- (GWCL)
・ここ数か月間、夜の来客数が激減している。それと共に客単価も下がっているため、非常に良くない。
total (GWCL)
・ここ数か月間、夜の来客数が激減している。それと共に客単価も下がっているため、非常に良くない。
-----WCSL-----
・ここ数か月間、夜の来客数が激減している。それと共に客単価も下がっているため、非常に良くない。
=====
```

図2 SINNによるポジティブコメントの解釈結果の可視化例(2)

```
-----WOSL-----
・客の動きが鈍かったが、ここにきて鈍いどころでなく止まっている。熊本地震の影響もあるが、九州方面はひたりと止まっている。為替変動の影響もあり、企業側の収益減と今後の先行き不安からくるものと思われる。前年同時期と比較して、受注も6割程度となっている。
-----LWCL-----
--> (LWCL)
・客の動きが鈍かったが、ここにきて鈍いどころでなく止まっている。熊本地震の影響もあるが、九州方面はひたりと止まっている。為替変動の影響もあり、企業側の収益減と今後の先行き不安からくるものと思われる。前年同時期と比較して、受注も6割程度となっている。
<-- (LWCL)
・客の動きが鈍かったが、ここにきて鈍いどころでなく止まっている。熊本地震の影響もあるが、九州方面はひたりと止まっている。為替変動の影響もあり、企業側の収益減と今後の先行き不安からくるものと思われる。前年同時期と比較して、受注も6割程度となっている。
total (LWCL)
・客の動きが鈍かったが、ここにきて鈍いどころでなく止まっている。熊本地震の影響もあるが、九州方面はひたりと止まっている。為替変動の影響もあり、企業側の収益減と今後の先行き不安からくるものと思われる。前年同時期と比較して、受注も6割程度となっている。
-----GWCL-----
--> (GWCL)
・客の動きが鈍かったが、ここにきて鈍いどころでなく止まっている。熊本地震の影響もあるが、九州方面はひたりと止まっている。為替変動の影響もあり、企業側の収益減と今後の先行き不安からくるものと思われる。前年同時期と比較して、受注も6割程度となっている。
<-- (GWCL)
・客の動きが鈍かったが、ここにきて鈍いどころでなく止まっている。熊本地震の影響もあるが、九州方面はひたりと止まっている。為替変動の影響もあり、企業側の収益減と今後の先行き不安からくるものと思われる。前年同時期と比較して、受注も6割程度となっている。
total (GWCL)
・客の動きが鈍かったが、ここにきて鈍いどころでなく止まっている。熊本地震の影響もあるが、九州方面はひたりと止まっている。為替変動の影響もあり、企業側の収益減と今後の先行き不安からくるものと思われる。前年同時期と比較して、受注も6割程度となっている。
-----WCSL-----
・客の動きが鈍かったが、ここにきて鈍いどころでなく止まっている。熊本地震の影響もあるが、九州方面はひたりと止まっている。為替変動の影響もあり、企業側の収益減と今後の先行き不安からくるものと思われる。前年同時期と比較して、受注も6割程度となっている。
```

図3 SINNによるポジティブコメントの解釈結果の可視化例(2)

```
-----WOSL-----
・2018年度の新卒採用活動の本格化により求人が増え、景気の盛り上がりが見える。また、雇用形態も正社員雇用の割合が増加している。
-----LWCL-----
--> (LWCL)
・2018年度の新卒採用活動の本格化により求人が増え、景気の盛り上がりが見える。また、雇用形態も正社員雇用の割合が増加している。
<-- (LWCL)
・2018年度の新卒採用活動の本格化により求人が増え、景気の盛り上がりが見える。また、雇用形態も正社員雇用の割合が増加している。
total (LWCL)
・2018年度の新卒採用活動の本格化により求人が増え、景気の盛り上がりが見える。また、雇用形態も正社員雇用の割合が増加している。
-----GWCL-----
--> (GWCL)
・2018年度の新卒採用活動の本格化により求人が増え、景気の盛り上がりが見える。また、雇用形態も正社員雇用の割合が増加している。
<-- (GWCL)
・2018年度の新卒採用活動の本格化により求人が増え、景気の盛り上がりが見える。また、雇用形態も正社員雇用の割合が増加している。
total (GWCL)
・2018年度の新卒採用活動の本格化により求人が増え、景気の盛り上がりが見える。また、雇用形態も正社員雇用の割合が増加している。
-----WCSL-----
・2018年度の新卒採用活動の本格化により求人が増え、景気の盛り上がりが見える。また、雇用形態も正社員雇用の割合が増加している。
```

図4 SINNによるポジティブコメントの解釈結果の可視化例(1)

```
-----WOSL-----
・住宅展示場への来場客数が3か月前と比べて10%増、相談件数2.5%増と商談内容が充実してきている。
-----LWCL-----
--> (LWCL)
・住宅展示場への来場客数が3か月前と比べて10%増、相談件数2.5%増と商談内容が充実してきている。
<-- (LWCL)
・住宅展示場への来場客数が3か月前と比べて10%増、相談件数2.5%増と商談内容が充実してきている。
total (LWCL)
・住宅展示場への来場客数が3か月前と比べて10%増、相談件数2.5%増と商談内容が充実してきている。
-----GWCL-----
--> (GWCL)
・住宅展示場への来場客数が3か月前と比べて10%増、相談件数2.5%増と商談内容が充実してきている。
<-- (GWCL)
・住宅展示場への来場客数が3か月前と比べて10%増、相談件数2.5%増と商談内容が充実してきている。
total (GWCL)
・住宅展示場への来場客数が3か月前と比べて10%増、相談件数2.5%増と商談内容が充実してきている。
-----WCSL-----
・住宅展示場への来場客数が3か月前と比べて10%増、相談件数2.5%増と商談内容が充実してきている。
=====
```

図5 SINNによるポジティブコメントの解釈結果の可視化例(2)