日本語文学作品の感情分析 ── BERT と 日本語評価極性辞書 の比較

手順：①作品を選ぶ（Ctrl＋クリックで最大3件）　②「選択作品を解析する」　③「表示条件」のスライダーを動かす（再推論なしで結果がその場で変わる）

解析する作品（Ctrl＋クリックで最大3件）

- 夏目漱石／こころ（4340 文）
- 夏目漱石／吾輩は猫である（6387 文）
- 夏目漱石／坊っちゃん（2440 文）
- 夏目漱石／草枕（2790 文）
- 夏目漱石／三四郎（5935 文）
- 夏目漱石／それから（5295 文）
- 夏目漱石／門（3548 文）
- 夏目漱石／彼岸過迄（3999 文）
- 夏目漱石／行人（6193 文）
- 夏目漱石／夢十夜（641 文）
- 芥川龍之介／羅生門（136 文）
- 芥川龍之介／鼻（154 文）
- 芥川龍之介／蜘蛛の糸（53 文）
- 芥川龍之介／地獄変（425 文）
- 芥川龍之介／藪の中（263 文）

選択 1 件：羅生門

解析条件（変更すると再解析が必要　＝ BERT を推論し直す）

解析文数の上限（作品冒頭から）　600

最大系列長［サブワード］　512

推論バッチサイズ（速度とメモリのみに影響）　32

☑ GPU を使う（利用可能なとき）

デバイス: CUDA:0 NVIDIA GeForce RTX 3090 (24.0GB)

表示条件（解析後も即時反映　＝ 再推論せずに集計・グラフだけを作り直す）

NEUTRAL とみなす確信度の下限　0.34

NEUTRAL とみなす |文極性値| の上限 θ　0.5

移動平均の窓幅［文］　20

1場面あたりの段落数　10

文極性値の集計方式：◉ 合計　○ 平均　☑ 否定表現（ない・ぬ・ず 等）による極性反転

θ の意味　実効: 極性語の合計が ±1 以上の文だけを極性判定

選択作品を解析する　中断　グラフを表示　感度分析（θ・確信度）　結果を保存（CSV・図・レポート）

準備完了。作品を選んで解析してください。

集計レポート　判定が分かれた文　実行ログ

```
青空文庫コーパスを準備しています（初回はダウンロードに数分かかります）…
  コーパス読み込み完了：全 16951 作品
作品の文分割を準備しています…
  選択できる作品：30 件
形態素解析器（MeCab + IPAdic）を準備しています…
日本語評価極性辞書を準備しています…
    ダウンロード中: https://www.cl.ecei.tohoku.ac.jp/resources/sent_lex/wago.121808.pn
    ダウンロード中: https://www.cl.ecei.tohoku.ac.jp/resources/sent_lex/pn.csv.m3.120408.trim
  極性辞書読み込み完了：判定に使う極性表現 13569 件
学習済みモデルを準備しています（初回は約443MB のダウンロード）… 使用デバイス: CUDA:0 NVIDIA GeForce RTX 3090 (24.0GB)
準備完了。作品を選び、「選択作品を解析する」を押してください。
```

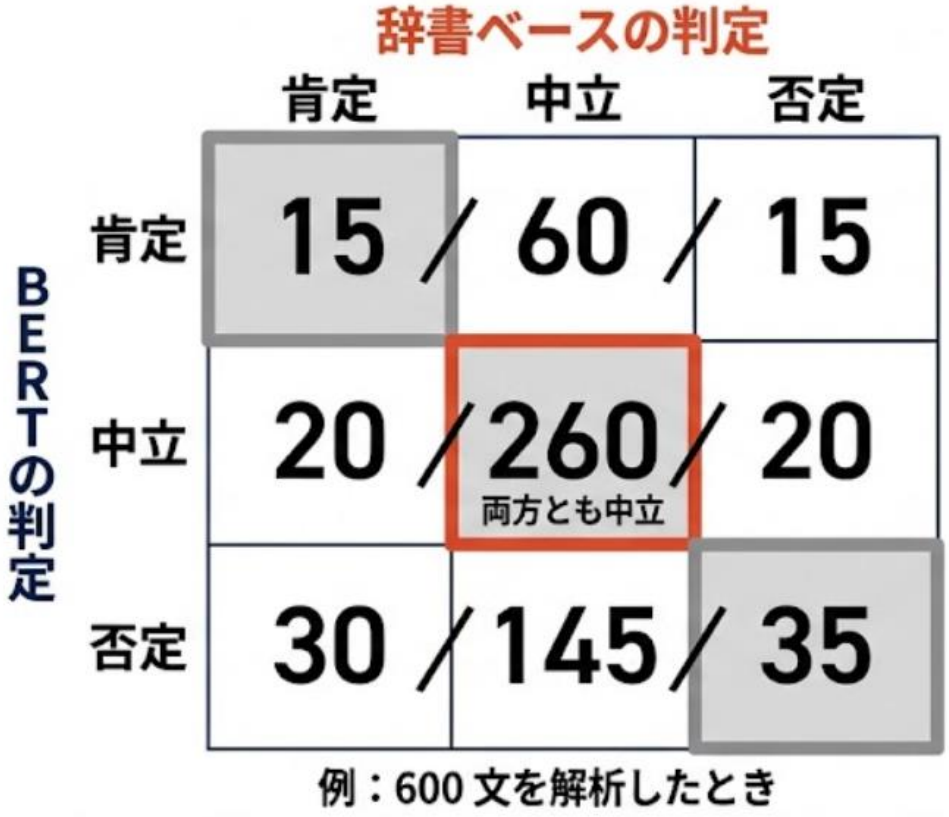
同じ文の集合に、原理の異なる2つの手法を当てて、判定を突き合わせる



	BERT	手法2　日本語評価極性辞書
文脈	語順や否定を含めて文全体で判断	語を数えるだけで語順は見ない
説明性	なぜその判定かは取り出せない	判定に効いた語を並べられる
調整	確信度の下限だけ	集計方式・θ・否定処理を変えられる

正解ラベルは存在しない。測っているのは正しさではなく、2手法の整合性である

一致率が高くても、その中身は『両方とも中立』かもしれない



例：600 文を解析したとき

51.7%
一致率　310文 / 600文

0.13
κ係数（偶然の一致を差し引いた値）
わずかな一致

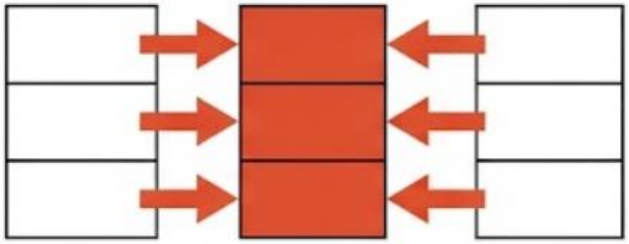
83.9%
一致した 310文のうち『両方とも中立』が占める割合

両方が判定を下した文だけで見ると、
一致は 95 文中 50 文（52.6%）
肯定と否定が食い違った文は 45 文

一致率だけを見ない。混同行列の中央のセルを必ず確認する

2つのしきい値は、どちらも判定を中立へ寄せる向きに働く

辞書側　しきい値θを上げる



辞書の判定が中立の列に集まる

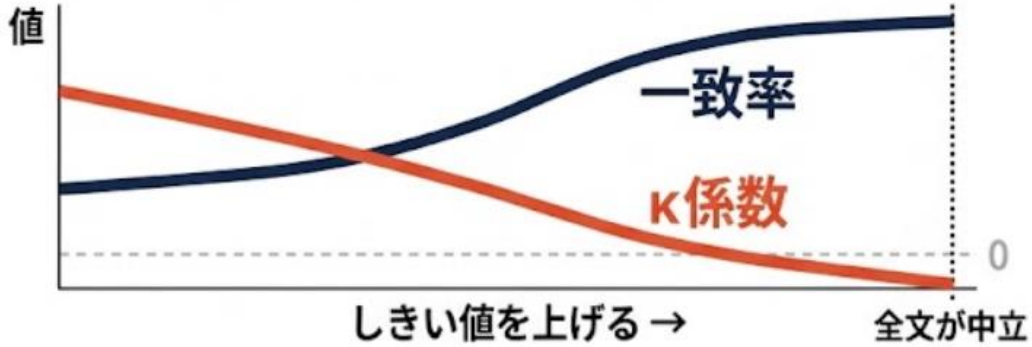
極性語が少ない文から順に中立になる

BERT側　確信度の下限を上げる



BERTの判定が中立の行に集まる

自信のない判定から順に中立になる



全文が中立になると、一致率は最大になるがκは0になる

θの効き方は集計方式で変わる

合計方式では文の極性値が整数になる

θ=0.5 と θ=1.0 は同じ判定。階段状に変わる

確信度の下限は 0.34 では効かない

3値分類では最大確率が 1/3 を下回らない

効果を見るには 0.5 以上に上げる

しきい値を1点だけ見ず、動かしたときの変化で判断する。再推論は不要で、その場で計算し直される