

言語処理学会第28回年次大会 併設ワークショップ
JED2022（日本語における評価用データセットの構築と利用性の向上）

学習データセット改善による アスペクト感情分析モデルの性能改善

2022.3.18

株式会社インテック

テクノロジー & マーケティング本部

先端技術研究所

亀谷 聡

1. はじめに
2. 学習データセットの改善方法一覧
3. アスペクト感情分析モデル概要
4. 学習データセットの改善方法試行
5. まとめ
6. 今後の課題

- 本発表の背景

- マーケティングオートメーション、コールセンター、CRM、医療、企業評価などでアспект感情分析を活用できる可能性がある。
- それら業務での活用可能性を検証するにあたり、学習データセットを作成し、改善していく方法を整理する必要がある。
- 公開データセットを用いてアспект感情分析の技術評価や学習データセットの作成及び改善方法を検討している。

- 本発表の目的

- アспект感情分析を業務活用する際の学習データセットの作成及び更新方法について気付きを得たい。
 - 検討した学習データセット改善方法の紹介。
 - 改善方法の試行結果の紹介。

2. 学習データセットの改善方法一覧

- モデルを運用する上で、学習データセットを改善していくことは必要不可欠。
 - 主目的の違いにより、改善方法は2つに分けられる。

No	主目的	方法	内容
1	継続的 データ更新	新旧データ の入替	✓ 期間を決めて新旧データを入れ替える。 ✓ 新規データについては人手で正解ラベルを付与。
2		無作為抽出を 基に追加	✓ 新規データが多い場合、無作為抽出により学習データセット に追加するデータを選別。
3	モデル 性能改善	誤判定傾向 の緩和	✓ 誤判定しやすい傾向を把握。 ✓ 学習データセットを追加して上記傾向を緩和する。
4		能動学習 による追加	✓ モデル判定結果が中程度スコアの文書を収集。 ✓ 上記文書に人手で正解ラベルを付与。
5		不適切な正解 ラベルの修正	✓ 不適切な正解ラベルを把握し、適切なラベルに修正。
6		リバランシング	✓ アップサンプリング/ダウンサンプリングを行う。

- 誤判定傾向の緩和（No.3）によりモデル性能を改善できるか試行。

• アスペクト感情分析とは？

- 文書中の**人・物・事に対する感情**（肯定的 or 中立的 or 否定的に書かれているか）を自動的に判定する技術。

ハンバーグはおいしいです。

肯定的

– 特徴

- 何に対する感情か判定**できる**。
- ひとつの**テキスト**に複数の感情が含まれる場合、それらを適切に判定**できる**。

肯定的

ハンバーグはおいしいですが、**店員**の対応は不親切でした。

否定的

3. アスペクト感情分析モデル概要

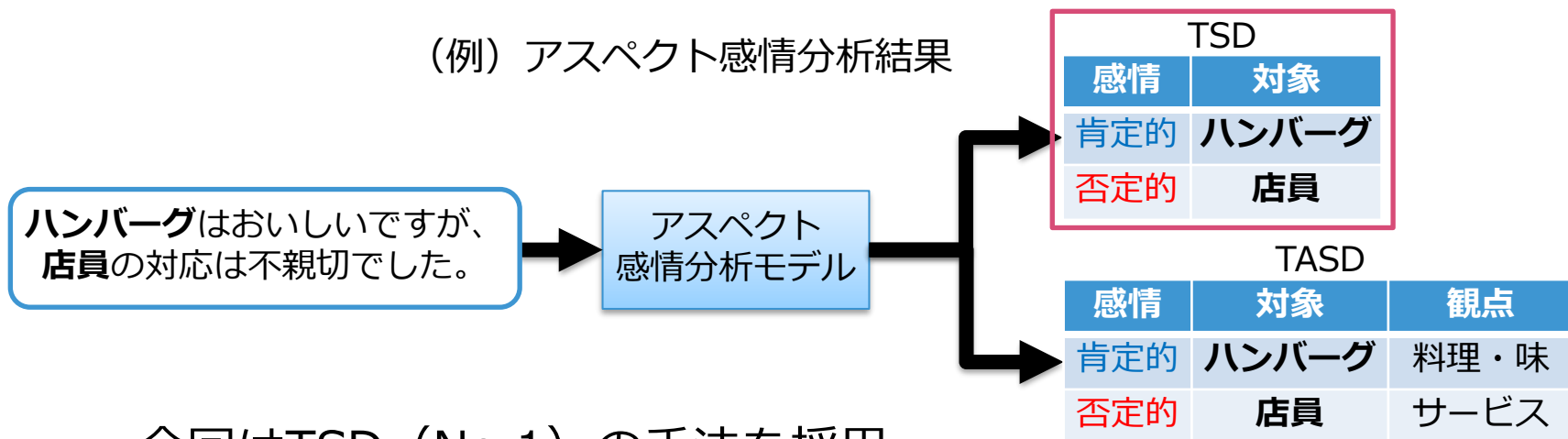
～ 適用手法 モデル作成方法 評価結果 ～

・ アスペクト感情分析の手法

– 出力の違いによって2つに分けることができる。

No	手法	出力	研究論文
1	TSD (Target-Sentiment joint Detection)	感情 対象	✓ Exploiting BERT for End-to-End Aspect-based Sentiment Analysis ✓ DOER: dual cross-shared RNN for aspect term-polarity co-extraction
2	TASD (Target-Aspect-Sentiment Detection)	感情 対象 観点	✓ Target-Aspect-Sentiment Joint Detection for Aspect-Based Sentiment Analysis

(例) アスペクト感情分析結果



– 今回はTSD (No.1) の手法を採用。

3. アスペクト感情分析モデル概要

～ 適用手法 モデル作成方法 評価結果 ～

・ 作成条件

項目	条件	備考
アスペクト感情分析手法	TSD	Exploiting BERT for End-to-End Aspect-based Sentiment Analysis
事前学習済みモデル	bert-base-japanese-whole-word-masking※1	東北大学 乾・鈴木研究室 日本語Wikipedia
学習データセット	chABSA-dataset※2	TIS株式会社 有価証券報告書

※1:<https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

※2:<https://github.com/chakki-works/chABSA-dataset>

・ 作成手順

No	実施項目	概要
1	データ変換	学習データセットをモデル入力形式に変換。
2	データ分割	訓練データ：検証データ：テストデータ = 6：2：2に分割。
3	モデル作成	訓練、検証データを使ってモデル作成。
4	モデル評価	テストデータにモデルを適用。 汎化性能（マイクロ平均F値）を確認。

3. アスペクト感情分析モデル概要

～

適用手法

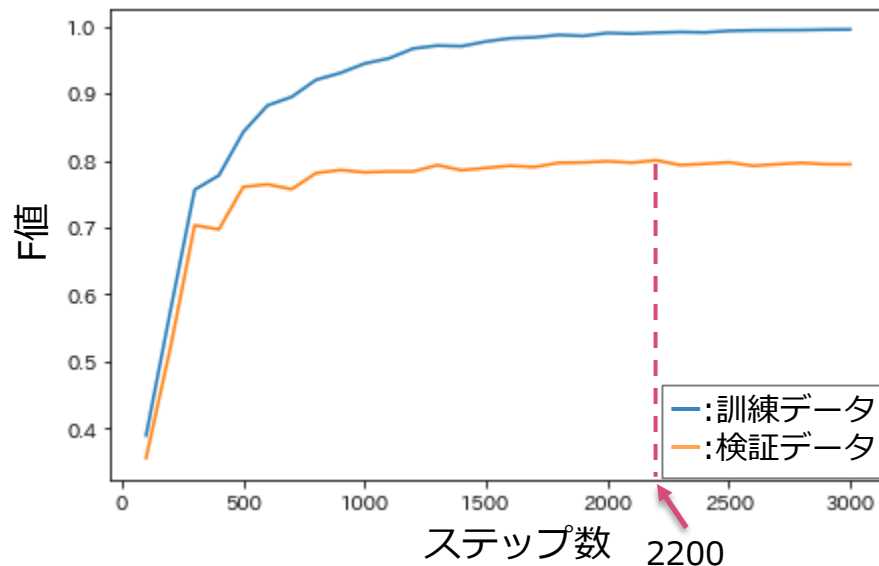
モデル作成方法

評価結果

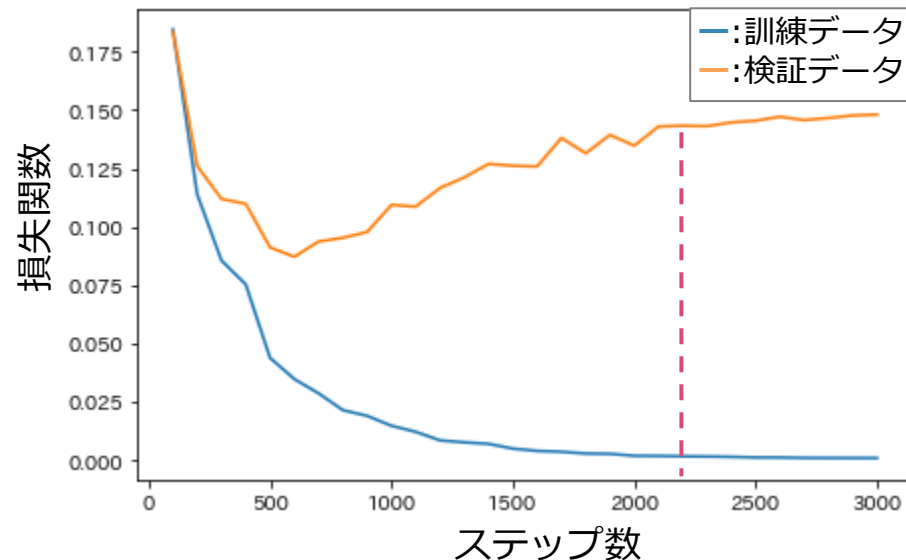
～

モデル選定：検証データF値最高モデル

学習曲線（F値）



学習曲線（損失関数）



テストデータ評価結果

指標	値
Precision	0.804
Recall	0.769
F値	0.786

3. アスペクト感情分析モデル概要

～

適用手法

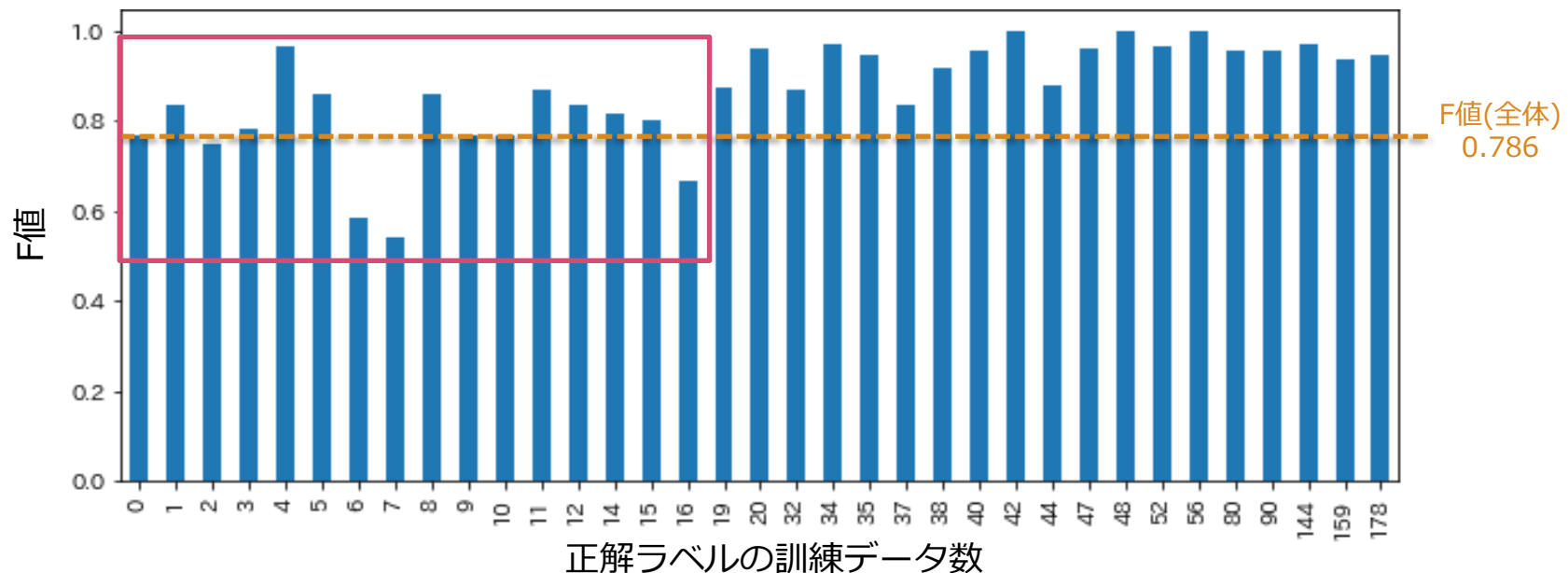
モデル作成方法

評価結果

～

・ 誤判定傾向

- 正解ラベル（対象：感情）の訓練データ数が少ないとF値は小さい。



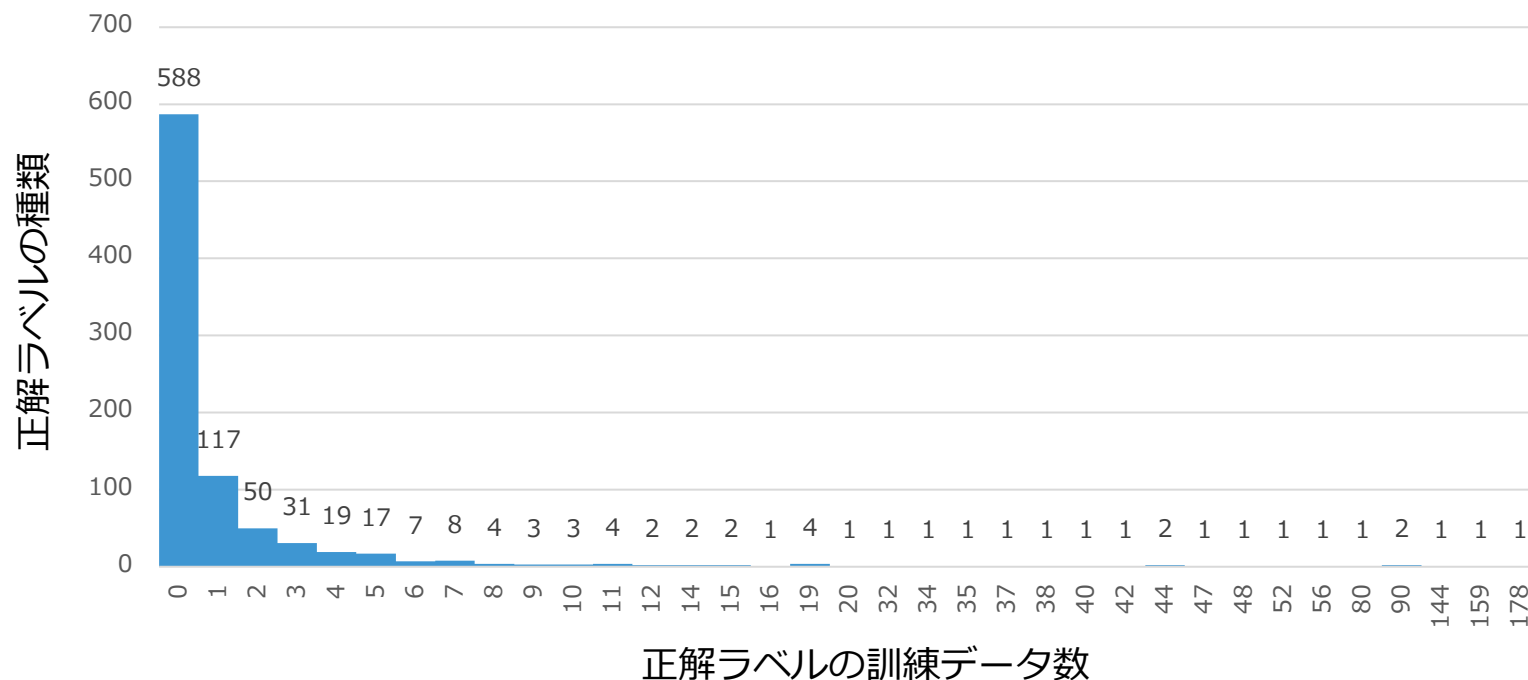
- 訓練データ数19以上の正解ラベルの汎化性能は高い。
- 訓練データ数の少ない正解ラベルは訓練データ数を増やすことで汎化性能を高められる可能性がある。

3. アスペクト感情分析モデル概要

～ 適用手法 モデル作成方法 評価結果 ～

• 誤判定傾向（続き）

- 今回、各正解ラベルについて訓練データを19個以上用意することで汎化性能の高いモデルを作れる示唆が得られた。
- 訓練データ数の多い正解ラベルは、正解ラベルの種類が数個と少ないため、種類を増やした検証が必要。



4. 学習データセットの改善方法試行

～ 改善方法検討 → 改善方法試行 → 改善効果確認 ～

• 誤判定傾向を緩和する方法の検討

- 訓練データ数0かつ偽陰性だった正解ラベルを含む文書を収集し、正解ラベルを付与後、訓練データへ追加。
- モデルを再学習することで偽陰性を真陽性に変え、汎化性能の改善を目指す。

(例) 訓練データ数0かつ偽陰性だった正解ラベル

正解ラベル (対象：感情)	訓練 データ	テスト データ	真陽性	偽陰性	偽陽性
新規店舗：肯定的	0	1	0	1	0
国内人材募集領域：肯定的	0	1	0	1	1
圧縮機：肯定的	0	1	0	1	0
業務用塩ビラップ：肯定的	0	1	0	1	0
流れ：肯定的	0	1	0	1	0
不動産業：否定的	0	1	0	1	0
大型タンカー：否定的	0	1	0	1	0
新政権：否定的	0	1	0	1	0
山陰製袋工業：否定的	0	1	0	1	0
機能樹脂製品：肯定的	0	1	0	1	0

(計215種類)

4. 学習データセットの改善方法試行

～ 改善方法検討 → 改善方法試行 → 改善効果確認 ～

・ 試行手順

No	実施項目	概要
1	データ収集	EDINET※1の全文検索機能を使い、「新規店舗」を含む文書を取得。
2	正解ラベル付与	アノテーション仕様書※2を参考に正解ラベルを付与。 収集した4文書に対して下記正解ラベルを付与。 新規店舗：肯定的 … 2文書 新規店舗：否定的 … 2文書
3	訓練/検証データ追加	訓練/検証データ（前回と同じ）へ上記文書を追加。 新規店舗：肯定的 → 訓練データ（1文書）、検証データ（1文書） 新規店舗：否定的 → 訓練データ（1文書）、検証データ（1文書）
4	モデル作成	訓練、検証データを使ってモデル作成。
5	モデル評価	テストデータにモデルを適用し、以下を確認。 ・ 汎化性能、正解ラベル（新規店舗：肯定的）の判定結果

※1:<https://disclosure.edinet-fsa.go.jp/>

※2:<https://github.com/chakki-works/chABSA-dataset/tree/master/annotation/doc>

追加文書

一方、店舗数の純増に加え、**新規店舗**が好調に推移したことから、増収を確保しました。

2020年、当行は、郡レベルの地域の店舗31を含め63の**新規店舗**を開設し、新たに8つの郡に進出した。

その他、当社が想定する海外の**新規店舗**の出店時期に遅れが生じるリスクがあります。

さらに、**新規店舗**が成功しない可能性もあり、1店舗当たり売上高を従前のように維持できない可能性もあります。

4. 学習データセットの改善方法試行

～ 改善方法検討 → 改善方法試行 → 改善効果確認 ～

・ テストデータ評価結果（1）

－ 汎化性能

指標	前回	今回
Precision	0.802	0.821
Recall	0.782	0.786
F値	0.792	0.803

－ 正解ラベル（新規店舗：肯定的）の判定結果

	正解ラベル (対象：感情)	訓練 データ	テスト データ	真陽性	偽陰性	偽陽性
前回	新規店舗：肯定的	0	1	0	1	0
今回		1	1	1	0	0

- － データ数の少ない正解ラベルを学習データセットに追加することで汎化性能が改善し、当該ラベルを正しく判定できた。

4. 学習データセットの改善方法試行

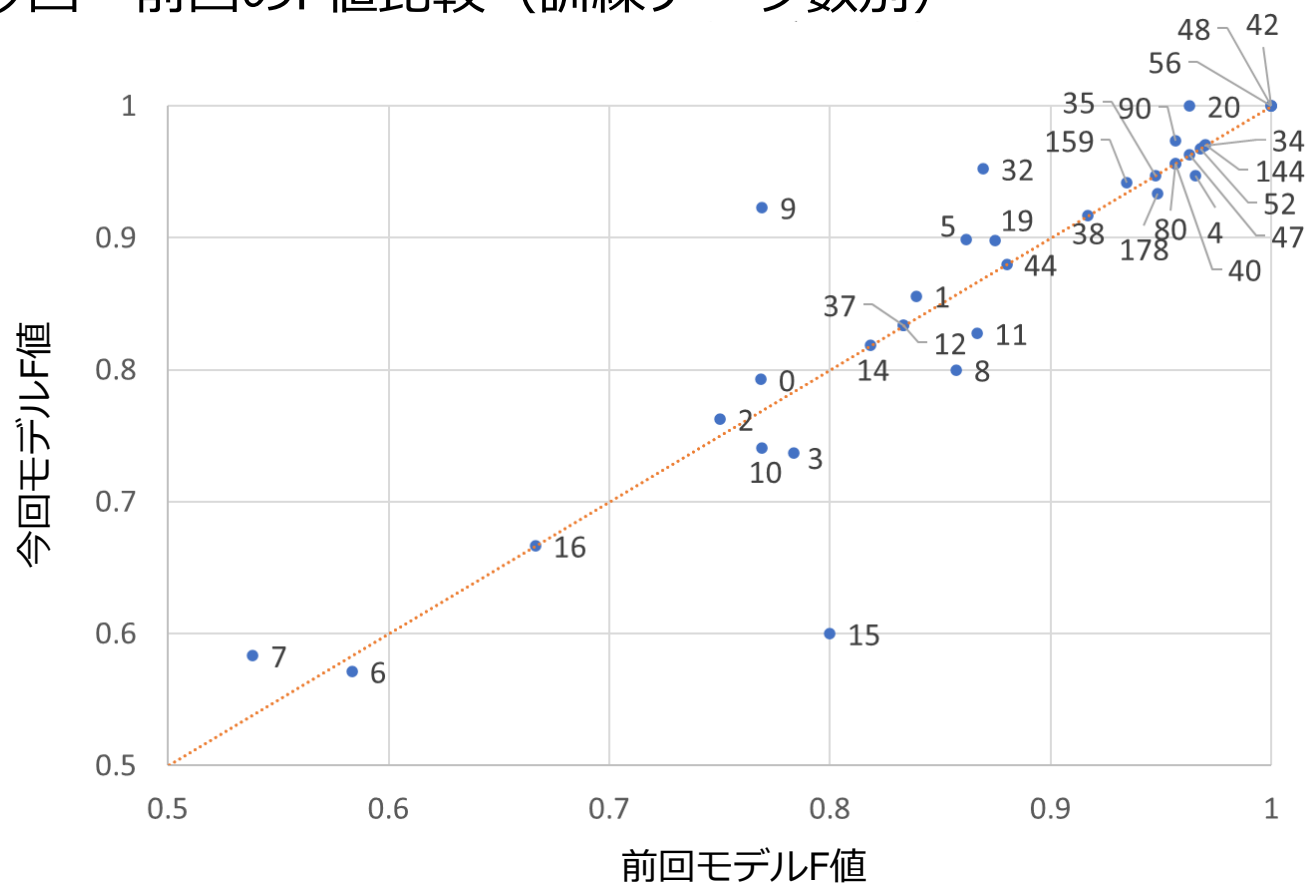
～ 改善方法検討

改善方法試行

改善効果確認

• テストデータ評価結果（2）

– 今回・前回のF値比較（訓練データ数別）



– 今回追加した正解ラベル以外の判定性能が変わるため、汎化性能が必ず改善するとは言い切れない。

- アスペクト感情分析の業務活用可能性を検証するために、技術評価や学習データセットの作成及び改善方法を検討している。
- アスペクト感情分析モデルの改善を目的に、学習データセットの改善方法を試行した。
 - 学習データセット（chABSA-dataset）を使いアスペクト感情分析モデル（TSD）作成。
 - 訓練データの少ない正解ラベル（新規店舗：肯定的）の文書を収集し、正解ラベルを付与。
 - 新規店舗：肯定的（2文書）、新規店舗：否定的（2文書）
 - 上記文書を訓練、検証データに追加してモデルを再作成。
 - 前回誤判定した正解ラベルを正しく判定し、汎化性能も改善。
 - 訓練データの少ない各正解ラベルの数を増やすことで、モデル性能を改善できる可能性がある。

6. 今後の課題（1）

- 学習データセット改善方法の課題と対策の整理
 - 今回、誤判定傾向を緩和する方法（No.3）を試行し、課題と対策を検討した。
 - それ以外の方法についても試行し、課題と対策を整理する。

No	主目的	方法	課題	対策
1	継続的 データ更新	新旧データ の入替	正解ラベル付け文書数が多い。 旧データ削除の条件、時機。	ユーザからフィードバックを収集。
2		無作為抽出を 基に追加		
3	モデル 性能改善	誤判定傾向 の緩和	正解ラベルを含む文書の収集。 正解ラベル付け作業負荷。	文書収集の自動化。 アノテーションツールの利用。
4		能動学習 による追加		
5		不適切な正解 ラベルの修正	学習データセットを1つ1つ確 認するのは負荷が高い。	
6		リバランシ ング		

- 各正解ラベルの学習データ数目標の決定
 - 今回、各正解ラベルについて訓練データを19個以上用意することで汎化性能の高いモデルを作れる示唆が得られた。
 - 訓練データ数の多い正解ラベルは、正解ラベルの種類が数個と少ないため、種類を増やして検証する必要がある。
- 各正解ラベルの学習データ数目標の削減
 - 訓練データ数19個は現実的な目標値ではない。
 - リバランシングや層化分割など試行し有効な方法を見つける。
- 正解ラベル付け作業負荷の削減
 - アノテーションツールを使っても作業負荷は高い。
 - ラベル付け作業負荷の軽減、自動化する方法を見つける。
- 学習データセット作成/改善に向けた関係者の動かし方
 - 作業負荷が高いため、お客様や社内メンバーに受け入れられ難い。
 - 作業負荷軽減に加え、説得しやすい根拠の示し方を検討する。

THANK YOU

ITで、社会の願い叶えよう。



TIS INTEC
Group