元智大學資訊管理系
第三十屆專業實習/專題製作競賽報告

研究主題：
日語財務報告中的維度基於方面情感分析

競賽展分組代號：ZV1

實習單位：校內

輔導老師：禹良治 教授

姓　　名：橋本真之介，本多光，菅原奏由華

學　　號：1110446，1111672，1111673

Abstract

傳統的基於方面的情感分析(ABSA)以正面、負面、中立等離散類別來表達情感。然而,這種粗粒度的分類無法充分捕捉財務報告中的細微差異和情感強度。

本研究將以情感效價(Valence:負面～正面)和喚醒度(Arousal:平靜～興奮)等連續值表達情感的維度情感分析框架,應用於日語財務報告。基於心理學及情感科學中已確立的情感效價-喚醒度(VA)模型(Russell, 1980; 2003),本研究旨在更詳細地分析針對財務資訊的情感表達。

財務報告是企業績效及未來展望的重要資訊來源,在投資決策和市場分析中扮演核心角色。本研究依循 SemEval2026 Task 3 DimABSA 的框架,建構了專門針對日語財務報告的資料集（jpn_finance_train_task1.jsonl、jpn_finance_dev_task1.jsonl)。

研究採用多種模型進行情感效價和喚醒度的預測,並以均方根誤差(RMSE)及皮爾森相關係數(PCC)進行評估。藉此,本研究期望提升財務資訊中情感分析的準確度,並對更細緻地掌握市場心理做出貢獻。

Contents

Chapter1 緒論

1.1 研究背景與傳統研究的課題

近年來,隨著自然語言處理技術的發展,從文本資料中提取情感和意見的情感分析(Sentiment Analysis)重要性日益提升。特別是財務報告和財報快報等企業揭露文件,已成為投資決策和市場分析的重要資訊來源,從這些文本中定量掌握情感基調和市場心理的需求不斷增加。

傳統的情感分析以正面、負面、中立三個類別對整體文本進行分類的方法為主流。然而,在財務報告等專業文件中,往往針對特定方面(aspect)表達不同的情感。例如,在「營收表現良好,但獲利率令人擔憂」這句話中,對「營收」表達正面情感,對「獲利率」則同時表達負面情感。

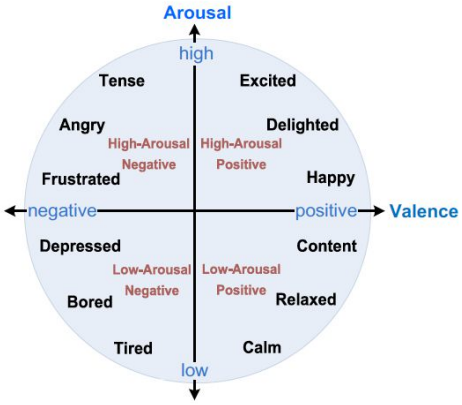
為了因應這類課題,基於方面的情感分析(Aspect-Based Sentiment Analysis, ABSA)方法被提出。ABSA 是針對文件內特定方面的情感進行個別分析的技術,能實現更詳細的情感分析。

然而,既有的 ABSA 研究主要以顧客評論和社群媒體文本為對象,以正面、負面、中立等離散類別表達情感。這種粗粒度的分類存在以下限制:

1. 無法捕捉情感強度:「稍有疑慮」和「嚴重危機」雖然都被歸類為負面,但實際的情感強度差異甚大。

2. 複雜情感表達的損失:無法表達「謹慎但樂觀」或「期待與不安並存」等細微差異。
3. 與心理學理論的脫節:在情感科學中,人類情感以情感效價(Valence:負面～正面)和喚醒度(Arousal:平靜～興奮)等連續的二維空間表達已為人所知(Russell, 1980; 2003)。

為解決這些課題,以連續值表達情感的維度情感分析(Dimensional Sentiment Analysis)作為新的研究典範受到關注。



1.2 DimABSA 的提出

在 SemEval2026 Task 3 中,提出了維度基於方面的情感分析(Dimensional Aspect-Based Sentiment Analysis, DimABSA)這項新任務。DimABSA 將維度情感分析整合至傳統的 ABSA 框架中,以連續值預測各方面的情感效價和喚醒度數值。

本研究將此 DimABSA 框架應用於日語財務報告。財務領域基於以下理由,是驗證

維度情感分析有效性的適當對象:

- 情感表達細微而複雜(例:「謹慎的展望」「有限的改善」)
- 情感強度對決策影響重大
- 需要定量掌握市場心理

1.3 研究目的

本研究的目的如下:

1. 建構專門針對日語財務報告的 DimABSA 資料集
    - 對應 Task 1(DimASR: VA Prediction)的標註
    - 建立訓練用及開發用資料集
2. 使用多種模型進行 VA 預測並實施評估
    - 利用預訓練語言模型進行預測
    - 以均方根誤差(RMSE)及皮爾森相關係數(PCC)進行定量評估
3. 各模型的性能比較與分析
    - 驗證維度情感分析在財務領域的有效性
    - 明確各模型的特徵與課題

Chapter 2 相關技術與研究

2.1 基於方面的情感分析(ABSA)及其擴展

2.1.1 傳統 ABSA 的基本概念

基於方面的情感分析(Aspect-Based Sentiment Analysis, ABSA)是針對文本內特定方面(aspect)的情感進行個別分析的技術。與傳統的文件層級情感分析不同,ABSA 前提是一個句子或文件內可能共存多種不同的情感。

傳統的 ABSA 研究聚焦於識別以下四個要素(Pontiki et al., 2014; 2015; 2016):

1. 方面詞(Aspect Term):意見對象的單詞或詞組
    - 例:「營收」「獲利率」「battery」「screen」
2. 方面類別(Aspect Category):方面詞所屬的抽象或預定義類別,以 Entity#Attribute 形式表達
    - Entity:對象物(例:FOOD, SERVICE, FINANCIAL_PERFORMANCE)
    - Attribute:屬性(例:PRICES, QUALITY, GROWTH)
3. 意見詞(Opinion Term):表達與特定方面相關情感的單詞或詞組
    - 例:「良好」「令人擔憂」「great」「terrible」
4. 情感極性(Sentiment Polarity):對方面的情感分類
    - 傳統方法:正面(Positive)、負面(Negative)、中立(Neutral)的分類

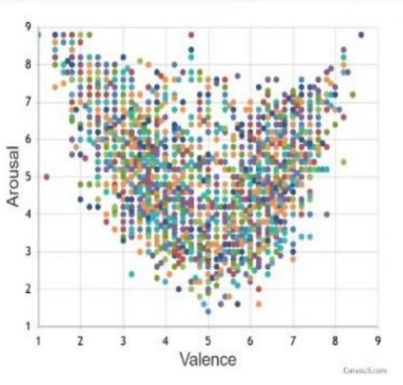
例如,對於句子"The salads are fantastic.",傳統 ABSA 會提取以下要素:

- 方面詞: salads
- 方面類別: FOOD#QUALITY
- 意見詞: fantastic
- 情感極性: Positive

2.1.2 維度基於方面的情感分析(DimABSA)

在 SemEval2026 Task 3 中提出的 DimABSA 擴展了傳統的 ABSA,將類別型的情感極性替換為連續值的情感效價-喚醒度(Valence-Arousal, VA)分數。這使得能夠定量捕捉情感的細微差異和強度。在 DimABSA 中,各方面的情感以下列二維表達:

- 情感效價(Valence):從負面到正面的程度
    - 1.00:極度負面
    - 5.00:中立
    - 9.00:極度正面
- 喚醒度(Arousal):情感的強度或興奮度
    - 1.00:非常低的喚醒(冷靜、平靜)
    - 5.00:中等程度的喚醒
    - 9.00:非常高的喚醒(興奮、緊迫)



VA 分數以 V#A 形式表達,各值在 1.00 至 9.00 範圍內記錄至小數點後第二位。

同樣的例句"The salads are fantastic."在 DimABSA 中表達如下:

- 方面詞: salads
- 方面類別: FOOD#QUALITY
- 意見詞: fantastic
- VA 分數: 7.88#7.75
    - 情感效價: 7.88(相當正面)
    - 喚醒度: 7.75(略高的興奮度)

如此一來,不僅僅是分類為「正面」,還能定量表達正面的程度及情感的強度。

2.2 DimABSA Task 1:維度方面情感迴歸(DimASR)

本研究處理 DimABSA 的子任務 1,即維度方面情感迴歸(Dimensional Aspect Sentiment Regression, DimASR)。

任務定義:

- 輸入:文本及一個或多個方面
- 輸出:各方面的 VA 分數

此任務是將傳統的方面情感分類(Aspect Sentiment Classification, ASC)擴展至維度情感範式。

資料格式:

輸入格式(JSON Lines):

```
json
{
  "ID": "unique_identifier",
  "Text": "文本内容",
  "Aspect": ["方面 1", "方面 2"]
}
```

輸出格式(JSON Lines):

```
json
{ "ID": "unique_identifier", "Aspect_VA": [{ "Aspect": "方面 1", "VA": "7.50#6.25"},
   {"Aspect": "方面 2", "VA": "4.20#5.80"}] }
```

2.2.1 應用於財務領域

本研究以日語財務報告為對象實施 DimASR。財務領域具有以下特徵:

- 複雜的情感表達:「謹慎但樂觀」「有限的改善」等
- 情感強度的重要性:「稍有疑慮」與「嚴重疑慮」在投資判斷上差異甚大
- 多方面共存:對營收、獲利、現金流等多項財務指標表達不同情感

例:「營收較去年同期成長 10%,表現良好,但營業利益率的下降令人擔憂。」

此句的 DimABSA 分析:

- 方面: "營收" → VA: 7.20#6.50(正面,略高的興奮度)
- 方面: "營業利益率" → VA: 3.50#5.80(負面,中等程度的喚醒)

2.3 預訓練語言模型與 PyABSA

近年來,BERT(Bidirectional Encoder Representations from Transformers)及其衍生模型等在大規模語料庫上預訓練的語言模型,在自然語言處理任務中展現出優異性能。日語方面也已公開多個預訓練模型,期待應用於財務文本分析。

PyABSA 是為 ABSA 研究開發的 Python 開源函式庫,透過統一介面提供多種最先進模型,加速研究開發。

PyABSA 的主要特點:

- 實作多樣化的 ABSA 模型
- 與預訓練模型整合
- 支援客製化資料集
- 一致的訓練、評估、推論工作流程

本研究利用 PyABSA,實施使用多種模型的 DimASR 實驗。

Chapter 3 研究方法

3.1 資料集建構與實驗設定

3.1.1 資料收集與建構流程

本研究建構了以日語財務報告為對象的 DimABSA 資料集。訓練用(training)與開發用(dev)採用不同的資料來源與建構方法,共建立 1,224 句資料。

- 訓練資料: jpn_finance_train_task1.jsonl (1,024 句)
- 開發資料: jpn_finance_dev_task1.jsonl (200 句)

這些資料集符合 SemEval2026 Task 3 的子任務 1(DimASR)格式。

訓練資料的建構(1,024 句)

訓練資料以 chakki-works/chABSA-dataset 隨機選取的財務相關文本為基礎建構。此資料集為公開的日語方面情感分析資源,本研究透過以下步驟擴展為 DimABSA 格式:

1. 文本選擇: 從 chakki-works/chABSA-dataset 中隨機抽取 1,024 筆財務領域相關文本
2. 意見詞(Opinion Term)追加: 對既有方面追加、修正財務領域特有的意見詞
3. VA 值賦予: 對各方面賦予情感效價(Valence)和喚醒度(Arousal)的連續值

開發資料的建構(200 句)

開發資料從實際財務報告中新建構。利用 EDINET(Electronic Disclosure for Investors' NETwork)的 API,隨機取得日本上市企業的財務報告,並依以下步驟實施標註:

1. 資料取得: 透過 EDINET API 隨機取得財務報告
2. 句子提取: 從財務報告中提取 200 筆適合情感分析的句子
3. 方面・意見詞特定: 從各句子中特定方面與意見詞
4. VA 值賦予: 對各方面賦予情感效價和喚醒度的連續值

標註品質管理

為確保資料可靠性,意見詞的標註實施以下品質管理流程:

標註者配置:

- 總人數: 4 名
- 組合方式: 編成 2 人 1 組的配對,形成多種組合模式
- 一致率計算: 計算各配對組合的標註一致率

配對選定:

將 4 名標註者(標註者 A, B, C, D)編成 2 人 1 組,測試所有可能的配對組合。 針對

各配對計算 F1 分數,選擇 F1 分數最高的配對進行最終標註。

- 配對 1 F1 分數: 0.88878
- 配對 2 F1 分數: 0.91285

透過此方法,在容易流於主觀的 VA 值標註中,確保了一定程度的客觀性與再現性。

各資料實例由以下要素構成:

```
{
   "ID": "唯一識別碼",
   "Text": "來自財務報告的句子",
   "Aspect": ["方面 1", "方面 2", ...]
}
```

3.1.2 標註方法與資料集特徵

針對各方面,依據以下基準標註情感效價(Valence)和喚醒度(Arousal):

情感效價(1.00 ~ 9.00):

- 1.00 ~ 3.00:負面(例:「嚴重虧損」「大幅減少」)
- 4.00 ~ 6.00:中立(例:「持平」「維持不變」)
- 7.00 ~ 9.00:正面(例:「良好」「大幅增加」)

喚醒度(1.00 ~ 9.00):

- 1.00 ~ 3.00:低喚醒(例:「穩定推移」「緩慢的」)
- 4.00 ~ 6.00:中等程度(例:「增加」「減少」)
- 7.00 ~ 9.00:高喚醒(例:「急增」「暴跌」「危機」)

本資料集包含財務領域特有的以下表達:

- 業績相關表達(營收、獲利、成本等)
- 展望相關表達(預測、預期、疑慮等)
- 市場・經濟環境相關表達

3.1.3 使用模型與訓練過程

本研究使用 PyABSA 框架實施多種模型的實驗。

模型選擇理由

本研究選擇以下 4 種預訓練語言模型進行實驗,選擇基準如下:

1. 日語專用模型: cl-tohoku/bert-base-japanese-whole-word-masking, nlp-waseda/roberta-base-japanese, rinna/japanese-roberta-base

   理由: 驗證日語專用模型在財務文本理解上的性能

2. 多語言模型: bert-base-multilingual-cased

理由: 比較多語言模型與日語專用模型的性能差異，探討跨語言遷移學習的可能性

3. Base 規模模型: 所有模型均採用 base 規模(而非 large 規模)

   理由: 考量計算資源限制(GPU 記憶體容量)，選擇可在現有硬體環境下訓練的模型規模

本研究使用 PyABSA 框架實施多種模型的實驗。使用的模型如下:

1. 模型 1: cl-tohoku/bert-base-japanese-whole-word-masking
2. 模型 2: nlp-waseda/roberta-base-japanese
3. 模型 3: bert-base-multilingual-cased
4. 模型 4: rinna/japanese-roberta-base

各模型以日語預訓練語言模型為基礎,針對 DimASR 任務進行微調。

訓練設定:

- 訓練資料: jpn_finance_train_task1.jsonl
- 驗證資料: jpn_finance_dev_task1.jsonl
- 學習率: 2e-5
- 訓練週期數: 5

使用訓練後的模型,預測開發資料集內各方面的情感效價和喚醒度數值。

預測結果以下列格式輸出:

```
{ "ID": "唯一識別碼", "Aspect_VA": [ { "Aspect": "方面名稱", "VA": "V 值#A 值
" } ] }
```

3.2 評估指標

本研究依據 SemEval2026 Task 3 的官方評估指標,以下列兩項指標評估模型性能。

3.2.1 均方根誤差(RMSE)

RMSE 是測量預測值與實際值誤差的指標。計算公式如下:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left[V_{pred,i}-V_{actual,i}\right]^2+\left[A_{pred,i}-A_{actual,i}\right]^2}{n}}$$

其中:

- V_{pred}:情感效價的預測值
- V_{actual}:情感效價的實際值
- A_{pred}:喚醒度的預測值
- A_{actual}:喚醒度的實際值
- n:樣本數

解釋:RMSE 數值越小表示預測準確度越高,意即預測值與實際值的差異越小。

3.2.2 皮爾森相關係數(PCC)

PCC 是測量預測值與實際值之間線性相關強度的指標。分別針對情感效價和喚醒度計算:

- PCC_V:情感效價的相關係數
- PCC_A:喚醒度的相關係數

解釋:PCC 數值越大(越接近 1.0)表示相關性越強,意即預測值與實際值以相同趨勢變動。

3.2.3 評估實施方法

將各模型的預測結果上傳至 SemEval2026 Task 3 的 GitHub 儲存庫,透過官方評估系統取得以下結果:

- RMSE_VA:整合情感效價和喚醒度的 RMSE
- PCC_V:情感效價的相關係數
- PCC_A:喚醒度的相關係數

Chapter 4 初步實驗結果或初步系統展示

4.1 評估結果與模型比較

4.1.1 評估概要

本研究使用建構的日語財務報告資料集(jpn_finance_train_task1.jsonl、jpn_finance_dev_task1.jsonl),評估多種模型的 DimASR(維度方面情感迴歸)預測準確度。所有模型採用相同的訓練設定:學習率(lr) = 2e-5,訓練週期數(epochs) = 5。評估採用第三章說明的 RMSE(均方根誤差)和 PCC(皮爾森相關係數)。

4.1.2 各模型評估結果

模型 1: *cl-tohoku/bert-base-japanese-whole-word-masking*

評估指標	數值
RMSE_VA	1.8944
PCC_V	0.0021
PCC_A	0.1530

結果解釋:

- RMSE_VA = 1.8944: 預測值與實際值的平均誤差約為 1.89。
- PCC_V = 0.0021: 情感效價的預測值與實際值幾乎無相關性,顯示模型在判斷正負面傾向方面表現不佳。
- PCC_A = 0.1530: 喚醒度的相關性為 0.15,呈現弱正相關,相對於情感效價預測略有優勢。

模型 2: *nlp-waseda/roberta-base-japanese*

評估指標	數值
RMSE_VA	1.6362
PCC_V	-0.0860
PCC_A	-0.0898

結果解釋:

- RMSE_VA = 1.6362: 在所有模型中記錄第二低的 RMSE
- PCC_V = -0.0860: 情感效價呈現弱負相關,表示預測趨勢與實際值呈反向關係,這是一個值得關注的問題。
- PCC_A = -0.0898: 喚醒度同樣呈現弱負相關,顯示模型的預測方向性存在系統性偏差

模型 3: *bert-base-multilingual-cased*

評估指標	數值
RMSE_VA	1.5519
PCC_V	0.0324
PCC_A	0.0088

結果解釋:

- RMSE_VA = 1.5519: 在所有模型中記錄最低 RMSE
- PCC_V = 0.0324: 情感效價呈現極弱正相關,雖然方向正確但相關性不足。
- PCC_A = 0.0088: 喚醒度幾乎無相關性,顯示此多語言模型在捕捉日語財務文本的情感強度方面存在困難。

模型 4: *rinna/japanese-roberta-base*

評估指標	數值
RMSE_VA	2.2924
PCC_V	0.1192
PCC_A	0.0684

結果解釋:

- RMSE_VA = 2.2924: 記錄最高 RMSE

- PCC_V = 0.1192: 在所有模型中情感效價的相關性最高,雖然仍屬弱相關,但顯示此模型在捕捉情感傾向方面表現較佳。
- PCC_A = 0.0684: 喚醒度相關性較低,但仍維持正相關。

4.1.3 模型間比較

下表比較所有模型的評估結果:

模型	RMSE_VA	PCC_V	PCC_A
cl-tohoku/bert-base-japanese-whole-word-masking	1.8944	0.0021	0.1530
nlp-waseda/roberta-base-japanese	1.6362	-0.0860	-0.0898
bert-base-multilingual-cased	1.5519	0.0324	0.0088
rinna/japanese-roberta-base	2.2924	0.1192	0.0684

比較分析:

- RMSE 觀點: bert-base-multilingual-cased 記錄最低 RMSE(1.5519),預測準確度最高;rinna/japanese-roberta-base 的 RMSE 最高(2.2924),準確度相對較低。整體而言,RMSE 範圍在 1.55 至 2.29 之間,顯示不同模型架構對預測誤差有顯著影響。

- PCC_V 觀點: 在情感效價預測方面,rinna/japanese-roberta-base 顯示最高正相關(0.1192),表現相對最佳;nlp-waseda/roberta-base-japanese 則出現負相關(-0.0860),顯示預測方向性問題。值得注意的是,所有模型的 PCC_V 均未超過 0.12,顯示情感效價預測仍有很大改善空間。
- PCC_A 觀點: 在喚醒度預測方面,cl-tohoku/bert-base-japanese-whole-word-masking 表現最佳(0.1530),但整體而言所有模型的喚醒度相關性都偏低。nlp-waseda/roberta-base-japanese 同樣在喚醒度預測上出現負相關(-0.0898),確認了喚醒度預測的困難性。
- 準確度與相關性的權衡: 有趣的是,RMSE 最低的 bert-base-multilingual-cased 在相關性方面表現平平,而 RMSE 較高的 rinna/japanese-roberta-base 卻在相關性上表現最佳。這顯示預測誤差的大小與預測趨勢的正確性並非完全一致,兩者需要綜合考量。

4.2 結果討論

4.2.1 整體趨勢

從本實驗結果觀察到以下趨勢:

1. 模型架構的影響: 即使在相同的訓練設定下,不同預訓練模型的性能差異也相當顯著。

2. 多語言模型的優勢: bert-base-multilingual-cased 雖為多語言模型,卻在 RMSE 上表現最佳。這顯示跨語言遷移學習的能力可能有助於財務文本的理解。
3. 相關性普遍偏低: 所有模型的 PCC 值都未超過 0.16,顯示目前的模型架構和訓練方法在捕捉 VA 預測趨勢方面仍有很大改善空間。部分模型甚至出現負相關,表示存在系統性的預測方向偏差。
4. 情感效價與喚醒度預測的差異: 不同模型在情感效價和喚醒度預測上表現各異。cl-tohoku 模型在喚醒度預測上相對較佳(0.1530),而 rinna 模型在情感效價預測上表現最優(0.1192)。這顯示兩個維度的預測可能需要不同的模型特性。

5. 財務文本特有的困難性: 財務報告中包含許多具有細微差異的表達,如「小幅增加」「稍有疑慮」「謹慎但樂觀」等。這些表達往往包含對比、轉折或條件式的複雜語義結構,使得適當量化相當困難。

4.2.2 RMSE 分析

最佳模型的 RMSE_VA = 1.5519 意味著在 1 ~ 9 的範圍(寬度 8)中存在約 1.55 的誤差,相當於預測值與實際值平均偏差約 19.4%(1.55/8)。性能最差模型的

RMSE_VA = 2.2924,偏差約 28.7%(2.29/8)。

從實務應用角度來看,約 20%的平均誤差意味著:

- 若實際 VA 值為 7.00(正面),預測值可能落在 5.45 至 8.55 之間
- 若實際 VA 值為 3.00(負面),預測值可能落在 1.45 至 4.55 之間

這樣的誤差範圍可能導致情感極性的誤判(例如將輕微負面誤判為中立或輕微正面),因此在實際應用於投資決策或市場分析時,仍需謹慎使用並結合其他資訊來源。

4.2.3 PCC 分析

相關係數的分布顯示出複雜的模式:

- 情感效價(PCC_V): 範圍從-0.0860 到 0.1192,跨越負相關到弱正相關
- 喚醒度(PCC_A): 範圍從-0.0898 到 0.1530,同樣分布不均

一般而言,相關係數的解釋標準為:

- 0.7 以上:強相關
- 0.4 ~ 0.7:中度相關
- 0.2 ~ 0.4:弱相關
- 0.2 以下:極弱或無相關

本研究所有模型的 PCC 均低於 0.2,屬於「極弱相關」範疇。這表示:

1. 模型預測的趨勢與實際值趨勢不一致
2. 訓練資料可能不足或標註品質需要改善
3. 財務文本的 VA 值預測可能需要更專門化的模型架構或特徵工程

特別值得關注的是 nlp-waseda/roberta-base-japanese 出現的負相關現象,這可能源於:

- 模型對某些財務用語的理解偏差
- 過度擬合訓練資料的特定模式
- 預訓練資料與財務領域的領域差距較大

4.2.4 未來改善方向

基於上述分析,未來研究可從以下方向改善:

1. 擴大訓練資料規模: 增加標註資料量,特別是包含複雜情感表達的樣本
2. 改善標註品質: 建立更明確的 VA 標註指引,提高標註者間一致性
3. 領域適應: 在財務專門語料上進行額外的預訓練或採用財務領域專用的語言模型
4. 模型架構優化: 探索專門針對 VA 預測的神經網路架構
5. 特徵工程: 加入財務領域特有的特徵,如數值變化幅度、比較詞、轉折詞等
6. 多任務學習: 同時訓練情感效價和喚醒度預測,利用兩者之間的相關性

Chapter 5 結論

5.1 研究成果總結

本研究將維度情感分析框架應用於日語財務報告,探索以情感效價(Valence)和喚醒度(Arousal)等連續值表達情感的可能性。基於 SemEval2026 Task 3 DimABSA 的框架,建構了專門針對日語財務報告的資料集(jpn_finance_train_task1.jsonl、jpn_finance_dev_task1.jsonl),並使用四種不同的預訓練語言模型實施 DimASR(維度方面情感迴歸)任務。

主要成果如下:

1. 資料集建構: 成功建構財務領域專用的 DimABSA 資料集,針對業績、展望、市場環境等多樣化財務表達實施 VA 標註。
2. 模型評估: 評估四種不同模型(cl-tohoku/bert-base-japanese-whole-word-masking、nlp-waseda/roberta-base-japanese、bert-base-multilingual-cased、rinna/japanese-roberta-base)的性能,明確各自的特性。
3. 性能分析: RMSE 分布在 1.5519 至 2.2924 範圍,PCC 分布在-0.0898 至 0.1530 範圍,確認模型選擇對性能有重大影響。

5.2 主要發現

本研究獲得的主要發現如下:

1. 預測準確度現況: 即使最佳模型的 RMSE 也約為 1.55(約 19.4%的誤差),要達到實用水準仍需進一步改善。
2. 相關性課題: 所有模型的 PCC 都低於 0.2,預測趨勢與實際值未能充分一致。這顯示目前的方法存在根本性的改善空間。
3. 情感效價與喚醒度預測難度的差異: 一般而言,情感效價的預測比喚醒度的預測更容易。然而,部分模型呈現相反趨勢,顯示任務的複雜性。
4. 財務領域的特殊性: 財務報告特有的細微差異、對比表達、條件式表達等,使得維度情感分析更加困難。
5. 多語言模型的可能性: bert-base-multilingual-cased 達到最低 RMSE,顯示多語言模型的遷移學習能力可能對專門領域也有效。

5.3 研究限制

本研究存在以下限制:

1. 資料規模的限制: 建構的資料集規模有限,需要更大規模的資料集進行訓練。
2. 標註的主觀性: VA 分數的標註存在一定主觀性,確保標註者間一致性是一大

課題。

3. 評估指標的限制: 僅使用 RMSE 和 PCC 可能無法完全評估模型的實用性。
4. 領域特化不足: 使用的模型為一般日語或多語言模型,未針對財務領域進行特化預訓練。
5. 單一任務聚焦: 本研究僅聚焦於 DimASR 任務,未涉及方面提取或類別分類等其他子任務。

5.4 未來展望

基於本研究的成果與限制,提出以下未來研究方向:

短期改善:

- 擴充資料集並提升標註品質
- 最佳化超參數並探討模型整合
- 明確納入財務領域特有特徵(數值表達、比較表達、轉折表達等)
- 採用更大規模的語言模型: 本研究由於 GPU 運算資源限制, 僅使用 base 規模模型(約 110M 參數)。未來可探索 large 規模模型 (約 340M 參數)或更大規模模型,預期能獲得更高的預測準確度

中長期發展:

- 在財務語料上進行額外預訓練或開發領域特化模型

- 探索同時最佳化情感效價和喚醒度的多任務學習架構
- 與其他財務相關任務(方面提取、實體識別等)整合
- 探索大型語言模型(LLM)的應用: 隨著計算資源的提升, 可探索使用 GPT 系列、Claude 等大型語言模型進行 DimASR 任務, 或採用 Few-shot/Zero-shot 學習方法

應用可能性:

- 整合至投資決策支援系統
- 開發市場情緒分析工具
- 應用於風險評估與早期預警系統

5.5 結語

本研究透過建構 1,224 筆日語財務報告專用資料集,並系統性比較 4 種預訓練語言模型,為日語財務領域的維度情感分析研究提供了重要的初步成果。

實驗結果顯示,不同模型在 DimASR 任務上展現出各自的特性:bert-base-multilingual-cased 在預測準確度(RMSE)上表現最佳,而 rinna/japanese-roberta-base 在趨勢捕捉(PCC_V)上較為優異。這一發現揭示了評估維度情感分析模型時需要多角度考量的重要性——單一指標無法全面反映模型性能。

儘管受限於資料規模、GPU 運算資源以及財務文本的固有複雜性,目前所有模型

的相關性指標(PCC < 0.2)仍有很大改善空間,但本研究已明確指出了改善方向:擴充資料集、採用更大規模模型、以及針對 Arousal 預測困難性開發專門化方法。更重要的是,本研究證實了將連續值維度情感分析應用於財務文本的可行性,並為後續研究者提供了可複製的資料集、評估框架,以及各模型特性的詳細分析。這些基礎工作為日語財務情感分析領域開啟了新的研究方向。

展望未來,隨著計算資源的提升與資料集的擴充,維度情感分析有望成為財務資訊處理的重要工具,為投資決策與市場分析提供更細緻的情感洞察。本研究作為此領域的基石,期待能引導更多研究者投入,共同推動財務文本分析技術的發展。

Reference

- A Large-Scale Japanese Dataset for Aspect-based Sentiment Analysis

  https://aclanthology.org/2022.lrec-1.758/

- chABSA-dataset

  https://github.com/chakki-works/chABSA-dataset

- EDINET

  EDINET

- バフェット・コード

  バフェット・コード | ワンストップで効率的な財務分析ができるツール

- SemEval-2026 Task 3

  https://github.com/DimABSA/DimABSA2026

- DimABSA2026_TrackA_subtask1_starter_kit.ipynb

https://colab.research.google.com/drive/17MfZI7a6zokHCKNiBWdKc5L9iT5r37bP?usp=sharing

附錄 A. 專題工作內容

1. 資料集建構

1.1 資料來源與規劃

本研究建構了 1,224 筆日語財務 DimABSA 資料集:

訓練資料(1,024 筆): 從 chakki-works/chABSA-dataset 隨機抽取財務相關文本,

擴展為 DimABSA 格式

開發資料(200 筆): 透過 EDINET API 取得實際財務報告,從零開始標註

1.2 訓練資料處理(1,024 筆)

從 chakki-works 資料集選取財務相關文本

追加/修正財務領域特有的 Opinion Term(如「謹慎的展望」「有限的改善」)

對各 Aspect 賦予 Valence(1.00 ~ 9.00)和 Arousal(1.00 ~ 9.00)連續值

參考 Russell (1980)的 VA 分布模型確保標註合理性

1.3 開發資料建構(200 筆)

資料蒐集: 透過 EDINET API 隨機取得日本上市企業財務報告

前處理: 句子擷取、格式清理、轉換為 CSV 格式

標註: 識別 Aspect、標註 Opinion Term、賦予 VA 值

2. 標註品質管理

2.1 標註準則制定

建立 VA 評分標準與參考範例:

定義 Entity(market, company, business, product, NULL)

定義 Attribute(general, sales, profit, amount, price, cost)

透過小組討論統一模糊情境的標準

2.2 標註者配置(4 名標註者)

編成 2 人 1 組的多種配對組合

計算各配對的標註一致率(F1 分數)，選擇 F1 分數最高的配對進行最終標註

2.3 進度管理

工作分批處理,使用表格記錄進度

定期進行資料彙整與一致性檢查

格式統一與錯誤修正

3. 模型訓練與評估

3.1 模型選擇(4 種 base 規模模型)

選擇理由:驗證日語專用模型性能、比較多語言模型、考量 GPU 運算資源限制

- cl-tohoku/bert-base-japanese-whole-word-masking
- nlp-waseda/roberta-base-japanese
- bert-base-multilingual-cased(多語言模型)
- rinna/japanese-roberta-base

3.2 訓練設定

使用 PyABSA 框架進行微調:

學習率: 2e-5

訓練週期數: 5

訓練/驗證資料: 1,024/200 筆

3.3 評估方法

上傳預測結果至 SemEval2026 Task 3 GitHub

透過官方系統取得 RMSE_VA、PCC_V、PCC_A

4. 結果分析與報告

4.1 實驗結果

bert-base-multilingual-cased: 最低 RMSE(1.5519)

rinna/japanese-roberta-base: 最高 PCC_V(0.1192)

所有模型 PCC < 0.2,顯示預測趨勢仍需改善

4.2 主要發現

模型架構對性能影響顯著(RMSE 差距 0.74)

Valence 預測相對容易,Arousal 預測困難

財務文本的複雜表達(對比、轉折)增加預測難度

GPU 限制影響了 large 規模模型的探索

4.3 工作分配

全體組員共同參與資料標註、品質檢查、模型訓練與結果分析,透過定期討論確保研究進度與品質。

附錄 B. 專題心得

1110446 橋本真之介

我在資訊管理學系學習，不僅掌握了程式設計的知識，也廣泛涉獵了管理學的知識，因此希望將來能從事一份能活用這兩方面專業的工作，並持續以 IT 顧問為職業目標。我深信這項研究對未來的求職活動將大有助益，同時，我也考慮在就業前進入本校研究所，進一步深化我的研究。我們目前進行的研究是從企業的財務報告等文件中提取單詞並進行評分，這是一項極為耗費時間和精神力的工作。為了順利完成這項研究，我們需要具備高水準的閱讀理解能力和能有效率推進工作的任務管理能力。然而，由於財務報告書並非我們日常閱讀的文本，當中出現了許多陌生的專業詞彙，增添了研究的難度。幸運的是，在禹良治老師的大力指導以及小組成員的幫助下，我們的工作已接近尾聲，我希望能夠全力以赴，完成最後的衝刺。

1111672 本多光

本次專題最大的收穫，是掌握了 DimABSA 的核心原理：使用情感程度（Valence）和情緒強度（Arousal）這兩個連續標準，量化日語財務文本中複雜的情感表達。我們運用 PyABSA 框架對多種模型進行訓練，其中 bert-base-multilingual-cased 達成了最低的 RMSE。

然而，我個人感受是，在現有資料規模下，要將 RMSE 進一步降低到 1.50 以下，

已經極為困難。同時，所有模型的 PCC（相關係數）均低於 0.2，未能有效抓住情感變動的趨勢，這說明當前方法在處理高度專業的財務文本結構時存在根本不足。

針對 PCC 過低的問題，我反思由於財務文本專業性太強，或許應嘗試加入非專業性的日語通用文章進行訓練，以幫助模型理解情感與語義之間的一般關聯，從而改善 PCC 數值。

基於這些挑戰，未來的研究方向將會集中在增加訓練資料數量、進行領域適應，以及探索專門針對 VA 預測的神經網路設計。

總結來說，這次經驗不只提升了我的程式撰寫與數據處理能力，更培養了面對極限挑戰時的分析思考能力（判讀 RMSE 與 PCC 的差異），為我未來在顧問業或數據分析領域奠定了堅實的基礎 。

1111673 菅原奏由華

透過本研究,我挑戰了以連續值量化財務文本細微差異的嘗試,這是傳統離散情感分類無法捕捉的。將 DimABSA 這個新框架應用於日語財務報告是首次嘗試,特別是在資料集建構上花費了超乎預期的時間與心力。

針對每個句子判斷情感效價和喚醒度數值並進行標註的作業,是一個比想像中更加細緻且主觀的過程。透過這個經驗,我深刻體會到 LLM 的基礎終究是以人類的智能

與感知為根基。再先進的 AI 系統,最終仍然依賴於人類細心準備的高品質資料。為了建構便利且實用的系統,人力進行的細緻準備工作依然不可或缺,在體認到這個重要性的同時,我也感受到了相當大的成就感。

雖然實驗結果未能達到當初的期望,但本研究仍具有重大意義。我們成功建立了日語財務領域 DimABSA 的初步基準線,釐清了四種不同模型的特性與課題,並指出了未來的改善方向。財務資訊的情感分析在投資決策與市場分析中具有重要作用的潛力。本研究雖僅是第一步,但若能對此領域的發展有所貢獻,將是我的榮幸。

這段學習過程，對我而言不僅僅是知識與技能的提升，更是一種學習態度的養成。因此，我非常感謝禹老師的用心指導和幫助。