國立臺灣師範大學理學院

資訊工程學系

碩士論文

Department of Computer Science and Information Engineering

College of Science

National Taiwan Normal University

Master's Thesis

改進提示學習訓練架構
以偵測社交媒體文本之心理健康面向

Improving Prompt-based Learning Framework for Mental Health Aspect Detection from Social Media Content

黃筱婷

Hsiao-Ting Huang

指導教授：柯佳伶 博士

Advisor: Jia-Ling Koh, Ph.D.

中華民國 113 年 2 月

February 2024

摘要

改進提示學習訓練架構以偵測社交媒體文本之心理健康面向

黃筱婷

本研究針對從中文社交媒體文本分析發文者心理健康面向的需求，探討如何從發文內容自動偵測發文者的心理健康狀態，包括是否有精神疾病或情緒問題，以及是否出現尋求協助的行為等三個面向。本論文改良提示學習中的訓練架構，提出漸次增加與策略挑選訓練資料的方法，對預訓練遮蔽語言模型微調的訓練方式進行改進，稱為 IS 訓練策略。此外，為了加強模型區別正負類別資料間的差異，額外採計邊界差異損失值。本研究在臺灣電子佈告欄上蒐集的三個心理健康面向資料集上進行實驗，結果顯示結合本論文所提出的 IS 訓練策略，可使 PET 和 iPET 所訓練出的面向偵測器，在 Precision 至少提升 10%。當控制訓練資料樣本數減少至一半的情況下採用 IS 訓練策略，相較於原訓練樣本數未採用 IS 訓練策略，各面向偵測器的 Precision 仍提升至少 5%，顯示 IS 訓練策略能從訓練集中有效挑選出對增進模型正確率學習有幫助的資料。在模擬開放環境測試集中，各目標面向的偵測效果皆可達 0.8 以上，顯示本論文所提出之提示學習訓練架構用於偵測社交媒體文本中心理健康面向的實用性。

關鍵字：提示學習、大規模預訓練語言模型、心理健康面向

目錄

附表目錄

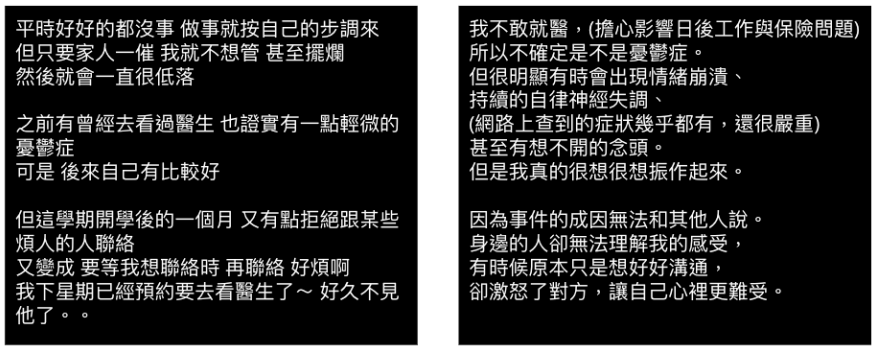
附圖目錄

第一章 緒論

1.1 研究動機與目的

由於社會型態的改變，現代人經常面臨情緒困擾，因此個人心理健康狀態逐漸受到重視。心理健康素養(Mental Health Literacy, MHL)評估是對個體進行評測，對於其是否了解如何保持積極的心理狀態，是否了解精神疾病與治療方式，以及尋求協助的態度等面向進行等級的評分。這些結果可用來分析心理健康素養中不同面向與尋求協助行為的關聯[13]。心理健康素養研究所需之評測資料，過往需要徵求受測者，以填寫問卷量表[2][19]的方式蒐集，再由專家評估其在心理健康素養各面向的態度分數，蒐集極為不易。隨著網際網路的普及，社交媒體蓬勃發展，使用者也常在社交平台上相互交流經驗並分享心情，因此開始有研究[25]嘗試從社交平台上發佈的文章內容(以下簡稱發文) ，理解該發文者描述的心理狀態與因應態度，並判斷其素養分數。

圖 1 所示為兩篇在電子佈告欄社交媒體平台的發文範例，其中圖 1 (a)所示發文範例一的內容提到「之前有曾經去看過醫生 也證實有一點輕微的憂鬱症」，顯示作者 A 自述具有精神疾病；此外，發文內容出現「我下星期已經預約要去看醫



(a) 發文範例一(作者 A)　　(b)發文範例二(作者 B)

圖 1 社交媒體平台的發文範例

生了」，顯示這位發文者面對精神疾病有尋求協助的行為。圖 1(b)所示發文範例二內容中出現「但很明顯有時會出現情緒崩潰」，顯示作者 B 自述具有情緒問題；此外，文中提到「我不敢就醫」，則顯示這位發文者面對情緒問題是負類別的因應態度。若能由社交平台上的發文內容自動判別出發文者自述有精神疾病或情緒問題，並能偵測出其是否有求助行為，可幫助心理健康素養專家更廣泛蒐集資料，用以找出可能有精神疾病與情緒問題的發文者，並分析其是否有尋求協助的行為[18]。

判別發文內容是否顯示發文者具有精神疾病、情緒問題與尋求協助行為，可視為對一段文字內容進行偵測任務。然而，採用監督式學習訓練類神經網路模型，需要大量的標示資料，對於上述的心理健康面向偵測任務，獲得這樣的標示資料通常需要專家反覆檢核確認，因此標示資料的蒐集相對困難。

近兩年，提示學習(Prompt-based learning)基於預訓練語言模型，運用其可產生通用語言理解表示法的特性，解決在少量標示樣本下(Few-shot scenario)的文本分類任務。論文[22][24]的實驗中顯示，在少量標示資料的情況下，在語言特徵空間採用提示學習的方式進行文本分類，相較直接採用預訓練語言模型取得語意表示法後另建立對應到類別的分類層，能大幅提升模型的預測效果。因此，本研究認為採用提示學習對於上述標示資料蒐集不易的文件分類任務是一個較可行的訓練方法。

使用克漏字形式的提示學習進行文本分類時，會將輸入文本內容接上提示模板文字(prompt template)，對於模板中填空空格的位置，透過語言模型預測較可能填入什麼詞彙。最後，將預測出的詞彙經過類別語言器(verbalizer)轉換成文本分類的判斷。以上述從發文中偵測心理健康面向的精神狀態為例，若要判斷圖 1 (a)

發文的文字段落中“之前有曾經去看過醫生 也證實有一點輕微的憂鬱症”是否顯示發文者具有精神疾病，先將此段內容後續加上提示模板文字“我[MASK]精神疾病”，產生“之前有曾經去看過醫生 也證實有一點輕微的憂鬱症，我[MASK]精神疾病”，形成一段具克漏字空格形式的文字段落作為模型輸入。預訓練語言模型在對這段文字內容進行語意編碼後，計算[MASK]位置可能填入「有」或是「沒」的機率值，選擇機率值高的詞彙作為模型輸出，再經由類別語言器轉換為類別標示(positive/ negative)。

iPET[22]採用半監督式(Semi-supervised)的學習策略建構文件分類器，訓練模型的過程分為兩個階段，即預訓練遮蔽語言模型(Masked Language Model, MLM)微調階段及分類器訓練階段。在預訓練語言模型微調階段，採用上述克漏字形式的提示學習對預訓練語言模型進行微調，再對未標記資料進行軟標記，接下來，根據這些具有偽標示的擴充資料訓練一個文字序列分類器，用以進行文本分類。

總結上述討論，採用 iPET 建構的模型對社交平台的發文進行心理健康狀態及求助行為自動偵測是一種可行的做法。然而本研究認為要採用 iPET[22]於此任務，在訓練時會面臨以下挑戰：社交媒體平台中的發文在心理健康面向資訊的呈現上，不具相關內容的負類別資料數量遠大於具相關內容之正類別資料。對預訓練語言模型採用提示學習調整模型時，所提供的少量訓練資料需正負類別數量接近，以避免語言模型預測類別時偏向特定類別，因此如何從較多數的負類別中挑選出有效的訓練資料是很重要的。提出 iPET 訓練架構的論文[22]中並未探討訓練資料的挑選機制。因此本研究將針對 iPET 訓練架構，探討如何系統化的選擇訓練資料，並對語言模型進行多回合的微調，以提升運用提示學習進行文本分類的訓練效果。

1.2 論文方法

本論文研究考慮的資料集來源為社交平台上的中文發文，對每一則發文內容進行三個心理健康面向的偵測任務，分別為是否具有精神疾病、是否具有情緒問題以及是否出現尋求協助行為。

本論文以 iPET[22]訓練架構為基礎，對預訓練語言模型微調階段的訓練方式進行改良，提出漸次增加與策略挑選訓練資料的方法微調語言模型。接著對大量未標記資料生成較高品質的偽標記結果，以擴充訓練文本分類器所需之標記資料。最後，透過這些擴充的標記資料訓練一個文字序列分類器，用以進行文本分類。

本論文認為提示學習的模型是基於原文字內容與提示模板文字在語言敘述的語意相關性，由於考慮的社交平台發文內容大部分很長，且內容中有許多個人生活描述，只有少量敘述顯出發文者的心理健康狀態及求助行為，若以整篇發文為單位進行訓練，由於訓練資料有限而影響模型訓練成效。因此本論文將每一篇發文進行前處理，切分為多個句組段落(segments)，以句組段落為單位進行偵測，判斷一篇發文的每個句組是否具有精神疾病、情緒問題或尋求協助行為。如果模型預測任何一個句組為正類別，則該篇發文最終被分類為正類別；反之，如果所有句組都被模型預測為負類別，則該篇發文最終預測結果為負類別。

本論文受到教育心理學中分散學習（Spaced Learning）[1]與反饋學習（Feedback Learning）[10]的啟發，在預訓練遮蔽語言模型微調階段(MLM Tuning)提出漸次增加與策略挑選訓練資料的訓練策略來結合 iPET，稱為 IS iPET。分散學習（Spaced Learning）是將學習內容分成多個小部分，然後重複學習這些內容，本論文採用分批逐漸加入訓練資料的方式對 MLM 模型進行多回合微調，以實現

分散學習的概念。至於如何有效挑選每一回合新增的訓練資料使模型更有效地學習，根據的基礎是反饋學習（Feedback Learning）：了解學習者已學習到的強項和弱點，並在後續學習過程中進行訓練資料的調整。本論文所提出漸次訓練的方式如下：初始回合對各類別隨機不重複採樣一些資料對 MLM 模型進行微調，調整後的模型對剩下的標示資料進行預測，藉由比對預測標示與實際標示，了解模型已學會的和仍然誤判的資料範例。為了調整模型學習上的偏差，下一回合的訓練資料會挑選模型正確判斷但確信度低的正類別資料，以及誤判為正類別但確信度高的負類別資料，加入以加強模型訓練。

此外，原先 iPET 在 MLM 模型時使用的損失函數為交叉熵，然而交叉熵損失的計算只考慮單一筆訓練資料，未能直接考慮不同類別資料的比較。為了使模型加強學習到區分不同類別資料間的特徵差異，本研究採用正負類別資料為一配對的訓練方式，除了各類別資料的交叉熵損失，並將正負類別訓練樣本在正類別的預測機率值差距列入損失計算，讓語言模型更有效地學習到正負類樣本間的特徵差異。

在效能評估部分，我們將實驗分成四個部分進行探討：包括

(1) IS 訓練策略對 PET 和 iPET 訓練架構的增進效果評估。

(2) IS 訓練策略在訓練樣本減少情境下的預測效果評估。

(3) IS iPET 訓練架構中各提出策略的效能評估。

(4) IS iPET 訓練架構於模擬開放環境測試集的預測效果。

本研究的主要貢獻有以下三點：

(1) 提出可結合 PET 及 iPET 的 IS 訓練策略，整體而言能使 PET 及 iPET 訓練架構所建構的面向偵測器在 Precision 提升 20%，F1-score 提升 10%。

(2) 於提示學習過程，在 MLM 模型損失函數額外考慮邊界損失值，以強化區別正負類別資料，並在實驗顯示此方式可有效提高模型偵測目標面向的準確度。

(3) 將 IS iPET 訓練策略於開放環境測試資料集進行評估，偵測效果皆可達 0.8 以上，顯示本論文改進提示學習訓練架構，用於偵測社交媒體文本中心理健康面向的實用性。

1.3 論文架構

本論文章節組織如下：第二章說明相關研究文獻，第三章說明本論文問題定義與資料處理，第四章介紹 IS iPET 訓練架構，第五章以實驗評估本論文提出方法在三個心理健康面向偵測任務的效果，最後於第六章總結及討論未來研究方向。

第二章 文獻探討

本章將對相關文獻進行探討：第一小節說明文本分類的相關技術演進，第二小節介紹使用提示學習處理自然語言任務的代表性論文，第三小節說明心理健康面向偵測目標的訂定及緣起。

2.1 文本分類的相關技術演進

2.1.1 採用非類神經網路模型之文本分類方法

早期文本分類方法通常採取兩個步驟：（1）制定特徵擷取方法(feature extraction)，例如詞袋模型(Bag-Of-Word Model, BoW) [6]，或是運用 TF-IDF[20]技術，以文本中的單詞及出現頻率作為文本表示特徵[15]，（2）將這些文本特徵輸入分類器建立模型並進行預測。然而，外在擷取特徵的方法由於事先制定特徵擷取方式而有其局限性，無法找出原始資料中更多潛在的特徵關係，而造成分類器預測效能的瓶頸。

2.1.2 採用類神經網路模型之文本分類方法

隨著類神經網路興起，可採用模型直接由文本內容學習特徵表示法。卷積神經網路(Convolutional Neural Networks, CNN) [14]與循環神經網路(Recurrent Neural Networks, RNN) [7]為兩種經常用來擷取文本語意的模型。運用卷積神經網路模型進行文本分類的研究[11][12]，經由卷積層擷取文本中的局部文字特徵，再透過全連接層將多個卷積層學習到的特徵進行融合，最後以輸出層對文本進行分類預測。採用循環神經網路的[26]及[29]則將文本視為一個單詞序列，運用循環層來學習文

本中文字序列涵蓋的語意，取序列最後一個文字輸入後的隱藏狀態經過輸出層進行文本分類。

2017 年，論文《Attention Is All You Need》[27]提出 Transformer 模型架構，解決傳統序列到序列模型在處理長序列時難以並行化的限制。Transformer 模型引入位置嵌入（Positional Embedding）來保留輸入序列中各詞彙出現的位置資訊。透過自注意力機制（Self-attention Mechanism），模型會計算出每筆詞彙在序列中的重要性分數，根據相對權重計算每筆詞彙的表示向量，使模型能同時融合輸入詞彙在序列中的前後資訊。接著經過前饋神經網路（Feedforward Network），計算出序列中每個詞彙的語意向量。為了得到文本的表示向量，可以將所有詞彙的語意向量取平均，或者採用句首標籤位置的語意向量，最後再將文本表示向量經過預測層得到文本分類預測[4]。

2.1.3 採用大規模預訓練語言模型之文本分類任務

近年來，為了避免針對不同的分類任務都需要從頭訓練模型，許多基於 Transformer-based 的大規模預訓練語言模型（Large-Scale NLP PTMs）問世，這些模型的出現有助於節省計算資源和資料標示成本。在大規模語料庫上進行語言模型預訓練，預訓練好的編碼器模型已學習到如何有效擷取出文本的語意表示法。因此可使用預訓練語言模型進行語意編碼後，另建立對應到類別的分類層(Linear Classifier)，利用下游任務的訓練資料，訓練分類器並微調預訓練語言模型的參數。以情緒分析任務為例，考慮句子“ I missed the bus, today is not my day.”，這是一個表達負面情緒的例句，分類器需要學習的是“I missed the bus, today is not my day.”

與負類別標示(negative)之間的映射關係。這樣的訓練方式會需要一定量的資料讓分類器學會資料與標示之間的映射關係。

在論文[28]中提出的 GLUE 資料集涵蓋情感分析(sentiment)，以及判斷一對文本關係為中立、矛盾、或包含的自然語言推論(Natural Language Inference, NLI)等文本任務所需資料，如圖 2.1 所示。GLUE 資料集中最小的資料集-WNLI 包含 634 筆訓練資料，其他資料集的訓練資料皆有上千筆。上述資訊顯示，即使是使用大規模預訓練語言模型進行文本分類任務，仍設定為具有上千筆的已標示資料作為訓練資料。然而，在特殊專業領域下，例如本論文所述之心理健康面向偵測任務，其標示資料需要經過至少兩位專家進行檢核確認，訓練資料取得成本高，要蒐集到上千筆的標示資料極為不易。

Corpus	|Train|	|Test|	Task	Metrics	Domain
			Single-Sentence Tasks		
CoLA	8.5k	**1k**	acceptability	Matthews corr.	misc.
SST-2	67k	1.8k	sentiment	acc.	movie reviews
			Similarity and Paraphrase Tasks		
MRPC	3.7k	1.7k	paraphrase	acc./F1	news
STS-B	7k	1.4k	sentence similarity	Pearson/Spearman corr.	misc.
QQP	364k	**391k**	paraphrase	acc./F1	social QA questions
			Inference Tasks		
MNLI	393k	**20k**	NLI	matched acc./mismatched acc.	misc.
QNLI	105k	5.4k	QA/NLI	acc.	Wikipedia
RTE	2.5k	3k	NLI	acc.	news, Wikipedia
WNLI	634	**146**	coreference/NLI	acc.	fiction books

圖 2.1 GLUE 資料集涵蓋任務[28]

2.2 提示學習

提示學習（prompt-based learning）的核心概念是以採用貼近語言模型預訓練方式的方法對模型進行微調。以 2.1.3 情緒分析任務為例，提示學習透過給定一個用來描述任務的文字提示模板，例如“I felt so [MASK] .”，將其與例句接合形成一

段文字：“ I missed the bus, today is not my day. I felt so [MASK]”。接著，設定正負類別標示可能在遮罩處出現的代表詞彙，例如以“happy”和“sad”分別對應到正面和負面情緒類別。語言模型的任務微調目標是理解句子與提示模板一整段的語意，然後學會在遮罩處對負面情緒句填入“sad”而不是“happy”。與 2.1.3 建立分類器的方式相比，提示學習方法只需微調語言模型在預訓練階段學得的模型參數，不需重新訓練分類器中的參數，因此可以用相對少量的訓練資料來進行。

論文[16]明確定義出在採用提示學習方法時需要提供的三項設定，分別是選擇預訓練語言模型、設計文字提示模板（prompt template），以及設計類別語言器（verbalizer）將類別映射到詞彙。透過這三個要件的組合，提示學習重新建構下游任務的輸入格式及輸出形式，使其符合預訓練語言模型在訓練階段的任務形式，以下將介紹幾篇代表性的提示學習研究。

2.2.1 PET 和 iPET 模型訓練架構

論文[22]提出的 PET（Pattern-Exploiting Training）和 iPET（iterative Pattern-Exploiting Training）模型訓練架構，採用半監督學習策略，其訓練過程分為三個階段：(1)使用少量標示資料微調一個預訓練遮蔽預測語言模型，使其適用於目標任務上，學習提示模板上遮罩位置的類別用語，並可根據不同提示模板及類別語言器的設計訓練多個模型$M_i^0(i = 1,2,\dots,n)$，如圖 2.2 中步驟(1)所示；(2)利用微調後的多個 MLM 模型在大量未標示的資料集D進行軟標記作為偽軟標示（pseudo soft label），以擴展訓練資料，如圖 2.2 中步驟(2)所示；(3)使用擴增後的訓練資料集 T_c 建構一個文字序列分類器C(sequence classifier)，如圖 2.2 中步驟(3)所示。

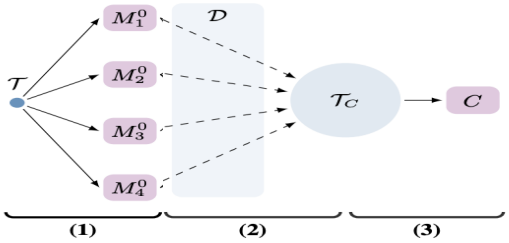


圖 2.2 PET 訓練架構[22]

對於一個文本分類任務，L表示分類任務的標示所成之集合，將給定的少量標示資料集以T表示，D表示未標示的資料集，V則表示文字資料集T及D中的字彙集合。在進行提示學習中，以M表示所採用的預訓練遮蔽預測語言模型；另外需定義一個文字提示模板Pat，提示模板Pat是一個含有[MASK]的克漏字形式文字段落；最後定義一個類別語言器(verbalizer) v，作為將模型M中的字彙表V映射到任務中的標示L之函數($V \rightarrow L$)。

PET 使用提示學習方式微調一個 MLM 模型的過程詳細說明如下：先將原始文字序列x，接合提示模板Pat後得到$Pat(x)$，再將$Pat(x)$輸入到預訓練遮蔽預測語言模型M，經過 softmax 層輸出在字彙表V中各個字可能出現在遮罩的機率分佈，透過類別言語化映射關係轉換得到對應的標示。根據 MLM 模型輸出的類別預測機率分佈與標準答案，使用交叉熵計算模型損失值，以L_{CE}表示，用來回調 MLM 模型的參數，如圖 2.3 中左側(1)所示步驟。

PET 將上一步驟得到的每個模型M_i^0在未標示資料集D的每筆資料上進行軟標記，其採用集成學習方式，將每筆未標示資料由n個微調後的 MLM 模型 M_i^0(i $= 1,2,\dots,n$)進行軟標記的結果，透過加權平均得到其偽軟類別標示。未標示資料集D及其偽軟標示形成訓練資料集T_c，如圖 2.3 中右側(2)所示步驟。最後，以資料集T_c訓練一個文字序列分類器，如圖 2.3 中右側(3)所示步驟。

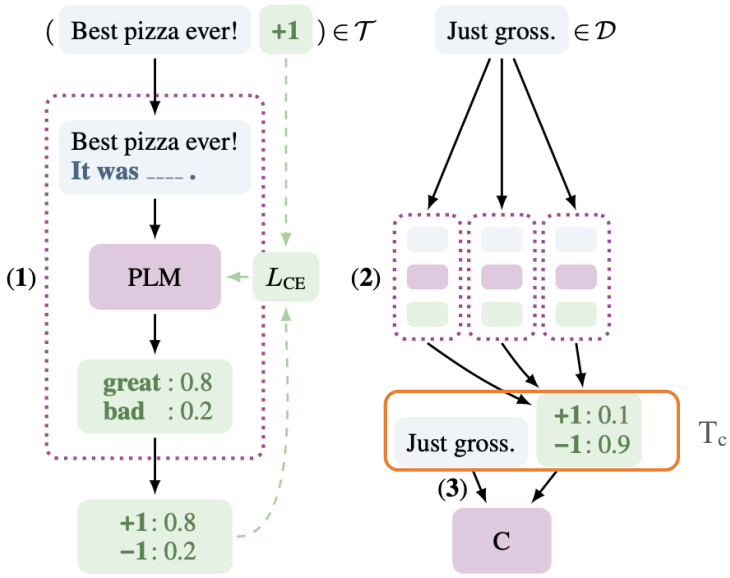


圖 2.3 PET 訓練過程[22]

在 PET 訓練架構中，由於T_c的偽軟標示可能因微調後的語言模型預測效果不夠好而涵蓋許多誤標的結果，進而影響所學習出分類器C的預測效果，因此論文中提出 iPET(iterative variant of PET)訓練架構來改善這個問題。其主要概念為：有次第的逐步增加偽標示訓練集的大小，先用來將預訓練遮蔽預測語言模型重複微調訓練幾回，用以提升語言模型預測效果，再對全部未標示資料集進行軟標示。每回合的處理過程是對前一回微調後的語言模型 $M_i^{t-1}(t = 1,2,\dots,k)$，隨機挑選該模型以外的多個模型，對$D$進行軟標記後挑選部分確定程度較高的資料及其偽標示加入前一回合的訓練資料T_i^{t-1}而形成新一回合的訓練資料集T_i^t，如圖 2.4 中(a)所示步驟；接下來以T_i^t微調模型 M_i^{t-1}得到模型 M_i^t，如圖 2.4 中(b)所示步驟。這樣的步驟會重複k次，得到模型M_i^k，如圖 2.4 中(c)所示步驟。iPET 接續的步驟(2)及步驟(3)則跟 PET 相同。

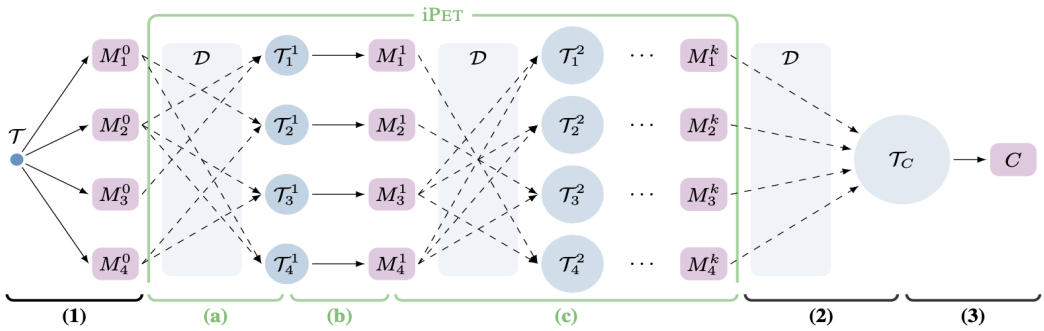


圖 2.4 iPET 訓練架構[22]

在論文[22]的實驗結果顯示，iPET 訓練架構在多數任務上的準確率較 PET 有所提升，上述結果顯示：iPET 改進 PET 提出的迭代訓練方式，藉由其他多個 MLM 模型產生的偽標示漸進挑選擴大訓練資料集，並實施多回合的微調策略是有效的。然而，對於負類別資料相對較多而造成類別分佈不平衡的資料集，隨機採樣各類別相同數量的資料對 MLM 模型進行一次微調，未必能達到好的訓練效果，進而導致 iPET 提出的迭代訓練方式從未標示資料中選出的偽標示資料有誤或是有偏差，無法使半監督式學習發揮增進效果。總結以上討論，本論文在 iPET 訓練架構的基礎上，將 iPET 的迭代概念提前引入到以標示資料微調 MLM 模型的步驟中，在圖 2.4 中步驟(1)提出了一種漸次增加與策略挑選訓練資料的方法，以因應上述的挑戰。

2.2.2 提示學習其他探討方向

關於提示學習的其他相關研究，由於不同的文字提示模板可能影響模型預測效果，且人工定義類別語言器，需要對分類任務有充分理解，因此其他一些研究聚焦於如何用系統自動制定文字提示模板和類別語言器。

論文[21]提出的 AutoPrompt 模型和論文[17]提出的 P-tuning 模型皆考慮提示學習中提示模板自動制定的方法，不同處在於 AutoPrompt 的提示模板是採用自然語言的文字形式制定(discrete prompt)，P-tuning 則是學習一個向量表示作為提示模板(continuous prompt)。

類似於提示模板，類別語言器(verbalizer)的自動制定方法可分為採用自然語言文字形式和向量表示兩大類。論文[23]提出的 PETAL 模型先從詞彙表中篩選出具有固定長度以上的詞彙，每一個候選詞彙作為答案與分類類別計算後，透過出現機率值最大化估計（Maximum Likelihood Estimation）的方法，自動找出類別的代表詞彙(discrete verbalizer)；論文[8]提出的 WARP 模型，是採用在模型中自動學習各類別所對應向量表示法(continuous verbalizer)，用於預測時與提示模板中[MASK]標籤位置進行點積(dot product)以算出類別的預測機率分佈，訓練時便可由計算損失函數回調各類別對應的表示法。

論文[5]則提出 LMBFF 模型，除了系統自動制定文字提示模板和類別語言器之外，提出可藉由示例(demonstrations)輔助語言模型的訓練和預測。該模型對少量標示資料中的每筆 x，由訓練資料中算出與 x 相似度前 50%高的資料，對每個標示類別各取出一筆資料，接續在 x 的後面作為示例。如圖 2.5 所示：“No reason to watch. It was [MASK]. ”這筆輸入資料的正負類別示例分別為，“A fun ride. It was great.” 和 “ The drama discloses nothing. It was terrible.” ，這些示例資料用以補充類別區分資訊，輔助語言模型更容易預測遮罩[MASK] 位置該填入的類別詞彙。

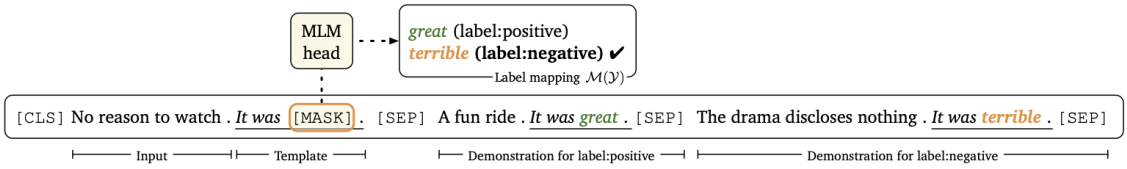


圖 2.5 LMBFF 模型訓練說明範例[5]

上述論文均探討如何自動制定提示模板和類別語言器以提升提示模型學習效能，但未考慮在「訓練資料的挑選策略」技術層面的討論。雖然多數提示學習研究強調僅需使用少量樣本即可微調語言模型，但如何挑選有代表性的訓練資料，使其在微調 MLM 模型時發揮最大效益，是本論文的研究重點。在提示模板和類別語言器方面，一些論文[3]實驗結果顯示，人工制定的提示模板和類別語言器即具有很好的表現，因此，本論文未就這方面的技術進行探究，而是採用人工制定的提示模板和類別語言器。

2.3 心理健康素養

心理健康素養（Mental Health Literacy, MHL）是指公眾對心理健康問題的認識、正確理解和處理能力。MHL 的早期概念由 1997 年澳大利亞心理學家 Tony Jorm 等人首次提出這一術語[9]，他們定義 MHL 為「對於精神疾病的知識與信念，可協助對於精神疾病的認識、處理及預防」，強調對精神疾病的正確理解和處理方式的重要性。後續心理健康素養的定義擴展到包括對心理健康問題污名化和尋求幫助的態度。

論文[2]提出的心理健康素養量表包含 26 個問題，涵蓋了五個主要面向：(1)瞭解如何獲取和保持良好的心理健康(Understanding how to obtain and maintain good mental health, M)；(2)瞭解心理障礙及其治療方式(Understanding mental

disorders and their treatment, R)；(3)解決與心理障礙相關的污名態度(Addressing stigmatized attitudes related to mental disorders, S)；(4)提高尋求協助的效能(Enhancing help-seeking efficacy, HE)；(5)增強尋求協助的態度(Enhancing help-seeking attitudes, HA)。每個主要面向都由多個次面向組成，以 HA(尋求協助的態度)為例，HA1 表示面對精神疾病的相關素養，包括對精神疾病的認識、理解和應對態度；HA2 則表示面對情緒問題的相關素養，包括對情緒問題的認識和理解，以及應對情緒問題的態度。此外，論文[19]提及的 HB 面向(Help-Seeking Behavior)則表示個體在面臨心理健康問題時，是否出現尋求支援或專業協助的行為。

上述定義可能會因研究設定目標而有差異，本論文是依照心理健康素養專家設定的研究背景，定義 HA1 為「我有精神疾病時，對於尋求專業協助的態度」；定義 HA2 為「我有情緒問題時，對於尋求專業協助的態度」；定義 HB 為「我發生精神疾病或情緒問題時，是否出現求助行為」。為了實現 HB 的自動偵測任務，分為兩個子任務：第一個子任務是識別出符合特定情景的相關發文，包含兩個二分類任務，即分別判定發文內容是否呈現發文者具有精神疾病及是否出現情緒問題；第二個子任務則進一步判斷發文者是否有尋求協助行為。因此，本研究可視為對每一則發文內容進行三個心理健康面向的偵測任務：(1)是否具有精神疾病，(2)是否具有情緒問題，以及(3)是否有尋求協助行為。

第三章 問題定義與資料處理

本章節於第一小節說明問題定義，接著在第二小節詳細說明資料集建構過程。

3.1 問題定義

本論文使用的資料集來自臺灣電子佈告欄批踢踢實業坊(PTT)蒐集的發文，每筆資料皆包含 5 個欄位，分別為文章編號(ID)、發文時間(Time)、發文者(Author)、標題(Title)、及內文(Content)，各欄位意義及範例如表 3.1 所示。

表 3.1 發文資料範例

欄位名稱	欄位意義	內容範例
ID	文章編號	116
Time	文章發佈時間	Fri Nov 7 23:46:25 2022
Author	文章發佈作者	doremi
Title	文章標題	[問題] 我好累喔
Content	內文	上週被主管唸，明明我已經很努力了，而且在報告前還覺得自己做得很好了，沒想到還是達不到他的要求…自從那天起覺得身心好疲憊，躺在床上眼淚一直滴 之後越來越嚴重被送到 1985 醫院，我說我睡不著 吃不下 心悸 乾嘔 暈眩想撞牆，醫生說我是中度憂鬱症，聽到這裡開始頭暈@@，聽不清楚醫生後來說了什麼，只剩下「我！是！憂！鬱！症！」～ 耳朵一直只聽見這五個字，後來我受不了尖叫了好久，被打鎮定劑才停止，住院觀察了幾天…

心理健康素養專家依據各目標面向的定義檢核發文，若發文中有某一連續段落符合目標面向的定義，則將該發文在該目標面向的類別標記為正類別，並將該

段落標記為判斷句；反之，若未發現任一符合目標面向定義的連續段落，則將該發文標記為負類別。由專家標記後的部分發文組成標示資料集，其中每一篇發文d_i都包含相對應的標示$d_i.label$，若標示為正類別，就會另指定一個判斷句$d_i.label_sentence$，如公式 3.1 所示。

$$d_i.label = \begin{cases} 'positive', & d_i.label\_sentence \neq \emptyset \\ 'negative', & d_i.label\_sentence = \emptyset \end{cases} \quad \text{(公式 3.1)}$$

以「具有精神疾病」的目標面向為例，表 3.1 所示範例d_{116}經專家標記為正類別，因此$d_{116}.label = 'positive'$且判斷句$d_{116}.label_sentence$為「醫生說我是中度憂鬱症」。

[文本面向偵測] 給定一篇包含多個句子所形成的文本$d_i = \langle s_{i,1}, s_{i,2}, \dots \rangle$，其中$s_{i,k}$代表第$i$篇文本中第$k$個句子。對輸入的發文$d_i$，給定一個目標面向$A_j$，**文本面向偵測任務**是預測該文本中是否呈現有目標面向，並對應到兩個類別標示：$\{positive, negative\}$。

本論文探討的心理健康面向包含「具有精神疾病」、「具有情緒問題」、「具有求助行為」三種不同面向，因此將分別進行三種文本的偵測任務。

通常一篇發文中的大部分敘述都是跟目標面向無關的生活描述，例如表 3.1 所示範例內文「上週被主管唸，明明我已經很努力了，而且在報告前還覺得自己做得很好了，沒想到還是達不到他的要求…自從那天起覺得身心好疲憊，躺在床上眼淚一直滴」，而文中跟目標面向相關的描述：「醫生說我是中度憂鬱症」，通常只會顯示在一句或一小段文字當中。因此以一篇發文為單位進行訓練和預測，模型可能因資訊量太繁瑣無法聚焦在具有目標面向的文字段落上；然而將發文斷句成句子的形式，可能會因為句子太短而造成語意資訊不夠完整。因此本論文將

斷句後的每一句子接合後面的句子直到一定長度以上，形成一個句組段落(segments)，本論文認為句組段落是一個相對完整敘述，適合用作模型訓練和預測的單位。

本論文將預測一篇發文是否呈現目標面向的問題，轉化為預測多個句組段落是否呈現目標面向的問題。在預測階段，如果模型預測任何一個句組段落為正類別，則該篇發文d_i最終被分類為正類別；反之，如果所有句組段落都被模型預測為負類別，則該篇發文d_i最終預測結果為負類別。

3.2 句組段落資料轉換

本小節將詳細說明如何將資料從發文轉換為句組段落。首先進行斷句，接著去除數字及標點符號，最後再接合後續句子形成句組段落。表 3.2 所示為三個步驟對一篇發文範例的處理過程。

社交平台中的每篇發文，常見用來表示語句停頓的標點符號為逗號、分號、句號、冒號、問號、驚嘆號、空白及換行。本論文會先將發文以上述語句停頓符號進行斷句，如表 3.2 步驟 1 所示，一篇發文經過斷句處理，由許多長短不一且含有雜訊的短句所形成。

為避免句子中非中文字的符號影響模型對文字語意的理解，接下來會移除標點符號（@、～、「、」...）和數字，如表 3.2 步驟 2 所示。

由於短句的語意不完整，句組段落是以每個句子為首，接續後面句子的內容，直到形成達到足夠長的文字段落。以文本$d_i = \langle s_{i,1}, s_{i,2}, \dots, s_{i,n} \rangle$為例，從第一個句子開始，當$s_{i,1}$的字數已達到指定長度，就形成一個句組段落以$seg_{i,1}$表示；若$s_{i,1}$的字數小於指定長度時，則接合句子$s_{i,2}, \dots, s_{i,k} (k \geq 2)$，直到字數大於等於$l$就停止接合，形成句組段落$seg_{i,1}$；然後再從下一個句子$s_{i,2}$為首，開始接合後續句子產生下個句組段落$seg_{i,2}$。一篇文本依上述方式可轉換成多個句組段落，$d_i = \{seg_{i,1}, seg_{i,2}, \dots, seg_{i,m}\}$，如表 3.2 步驟 3 所示。

表 3.2 處理流程範例

發文d_i	之後越來越嚴重被送到 1985 醫院，我說我睡不著 吃不下 心悸 乾嘔 暈眩 想撞牆，醫生說我是憂鬱症，聽到這裡開始頭暈@@，聽不清楚醫生後來說了什麼，只剩下「我！是！憂！鬱！症！」～ 耳朵一直只聽見這五個字，後來我受不了尖叫了好久，被打鎮定劑才停止，

步驟	處理方式	處理後
1	斷句	“之後越來越嚴重被送到 1985 醫院” “我說我睡不著” “吃不下” “心悸” “乾嘔” “暈眩” “想撞牆” “醫生說我是憂鬱症” “聽到這裡開始頭暈@@” “聽不清楚醫生後來說了什麼” “只剩下「我” “是” “憂” “鬱” “症」～” “耳朵一直只聽見這五個字” “後來我受不了尖叫了好久” “被打鎮定劑才停止”
2	移除數字及標點符號	“之後越來越嚴重被送到醫院” “我說我睡不著” “吃不下” “聽到這裡開始頭暈” “聽不清楚醫生後來說了什麼” “只剩下我” “是” “憂” “鬱” “症” “耳朵一直只聽見這五個字” “後來我受不了尖叫了好久” “被打鎮定劑才停止”
3	形成句組段落	“之後越來越嚴重被送到醫院 我說我睡不著” “我說我睡不著 吃不下 心悸 乾嘔 暈眩” “吃不下 心悸 乾嘔 暈眩 想撞牆 醫生說我是憂鬱症” … “耳朵一直只聽見這五個字 後來我受不了尖叫了好久” “後來我受不了尖叫了好久 被打鎮定劑才停止”

針對產生句組段落前需要指定的最小字數l，本論文決定l的方式如下：將各判斷句包含的字數進行離群值分析，找出第一四分位數(Q1)及第三四分位數(Q3)，計算 Q3-Q1 得出字數統計結果的四分位距(IQR)，找出判斷句的字數大於 Q3+1.5*IQR 或小於 Q1-1.5*IQR 的數據，即判斷為離群值。離群值分析有助於識別資料集為常態分佈中的極端值。因此去除判斷句字數的離群值後，計算平均字數取整數作為l的設定值。

對於從文本d_i取出的每個句組段落$seg_{i,j}(j = 1, 2, \ldots, m)$，若該文本的標示為‘positive’，則進一步比對$d_i.label_sentence$是否被包含在其中。若$seg_{i,j}$包含$d_i.label_sentence$，表示是專家判定具心理健康面向敘述的前後文段落，因此該句組段落標記為‘positive’；若$seg_{i,j}$不包含判斷句，只表示其中的敘述未被明列具目標面向，未必皆不具目標面向，因此視為未標示(none)；另一方面，若文本d_i的標示為‘negative’時，表示敘述內容不具面向，因此d_i中的所有句組段落標示皆設為‘negative’，如公式 3.2 所示，其中$Sub(seg_{i,j})$表示由$seg_{i,j}$的所有子字串所形成的集合。

$$seg_{i,j}.label =$$

$$\begin{cases} positive,\ d_i.label\_sentence \neq \emptyset \wedge d_i.label\_sentence \in Sub(seg_{i,j}) \\ none,\ d_i.label\_sentence \neq \emptyset \wedge d_i.label\_sentence \notin Sub(seg_{i,j}) \\ negative,\ d_i.label\_sentence = \emptyset \end{cases} \quad \text{(公式 3.2)}$$

給定一個文本d_i，令$L_p(d_i)$代表d_i中標示為‘positive’的句組段落所成之集合，如公式 3.3 所示；$L_n(d_i)$代表d_i中標示為‘negative’的句組段落所成之集合，如公式 3.4 所示；$U(d_i)$代表d_i中未經標示的句組段落所成之集合，如公式 3.5 所示。

$$L_p(d_i) = \{seg_{i,j} | seg_{i,j} \in d_i \wedge seg_{i,j}.label = 'positive'\} \quad \text{(公式 3.3)}$$

$$L_n(d_i) = \{seg_{i,j} | seg_{i,j} \in d_i \wedge seg_{i,j}.label = 'negative'\} \quad \text{(公式 3.4)}$$

$$U(d_i) = \{seg_{i,j} | seg_{i,j} \in d_i \wedge seg_{i,j}.label = 'none'\} \quad \text{(公式 3.5)}$$

給定一個文本集合D，$L_p(D)$代表D中各文本d_i標示為‘positive’的句組段落集合聯集所成之集合，如公式 3.6 所示；$L_n(D)$代表D中各文本d_i標示為‘negative’的句組段落集合聯集所成之集合，如公式 3.7 所示；$U(D)$代表D中各文本中未經標示的句組段落集合聯集所成之集合，如公式 3.8 所示。

$$L_p(D) = \bigcup_{d_i \in D} L_p(d_i) \quad \text{(公式 3.6)}$$

$$L_n(D) = \bigcup_{d_i \in D} L_n(d_i) \quad \text{(公式 3.7)}$$

$$U(D) = \bigcup_{d_i \in D} U(d_i) \quad \text{(公式 3.8)}$$

上述標記為‘positive’或‘negative’的句組為標示的句組段落，以$L(D)$表示，為$L_p(D)$和$L_n(D)$聯集而成，如公式 3.9 所示；$Seg(D)$表示文本集合D中所有句組段落$seg_{i,j}$所成之集合，為$L(D)$和$U(D)$的聯集，如公式 3.10 所示。

$$L(D) = L_p(D) \cup L_n(D) \quad \text{(公式 3.9)}$$

$$Seg(D) = L(D) \cup U(D) \quad \text{(公式 3.10)}$$

第四章 以提示學習訓練心理健康面向偵測模型

本章節介紹漸次增加與策略挑選訓練資料的 IS iPET 訓練架構，在第一小節說明提示學習訓練架構設定，在第二小節介紹如何採用漸次增加與策略挑選的 IS 訓練策略微調 MLM 模型的方式，並在第三小節提出對偵測候選資料的篩選方法。

4.1 提示學習訓練環境設定

使用提示學習方法訓練分類模型時，需要對應的預訓練語言模型、文字提示模板以及類別語言器。本論文所採用的 MLM 模型為 BERT，各心理健康面向使用的文字提示模板與類別語言器如表 4.1 所示。以「具有精神疾病」的目標面向為例，文字提示模板Pat為「我[MASK]精神疾病」；所採用之類別語言器v，類別標示 positive 對應到字彙「有」，而類別為 negative 則對應到字彙「沒」。

表 4.1 心理健康面向文字提示模板與類別語言器列表

目標面向	文字提示模板	類別語言器
具有精神疾病	我[MASK]精神疾病	$v(\text{positive})=$有 $v(\text{negative})=$沒
具有情緒問題	我[MASK]情緒方面的問題	
具有求助行為	我[MASK]主動尋求協助，來處理精神疾病或情緒問題	

舉例來說，以seg^+和seg^-分別代表標示句組段落集合L中的兩筆正負類別資料，其中 seg^+為「醫生說我是中度憂鬱症」，而seg^-為「天氣一變冷我的關節就開始疼痛」。兩筆資料與文字提示模板Pat接合後分別得到，$Pat(seg^+)=$「醫生說我是中度憂鬱症，我[MASK]精神疾病」，$Pat(seg^-)=$「天氣一變冷我的關節

就開始疼痛，我[MASK]精神疾病」。MLM 模型的任務目標是對這兩段文字內容進行語意編碼後，學會在$Pat(seg^{+})$的[MASK]位置填入「有」的機率高於填入「沒」的機率，在$Pat(seg^{-})$的[MASK]位置則反之。

4.2 提示學習訓練架構結合 IS 訓練策略

PET 及 iPET 訓練架構，在提示學習微調 MLM 模型階段，在實驗時會對各類別隨機取相同數目的樣本對 MLM 模型進行微調。然而在訓練資料集中類別分佈不平衡情況下，對數量較多的類別隨機採樣與較少量類別相同數量的資料對 MLM 模型進行一次微調，未必能達到良好的訓練效果。本論文提出 IS 訓練策略，以分批逐漸加入相同數量的正負類別訓練資料對 MLM 模型進行多回合微調，並且透過設計的挑選策略，來決定每回合新增的訓練資料。結合上述兩種概念，每回合新增對模型加強學習的訓練樣本，使 MLM 模型在微調階段能夠逐漸適應目標任務，提升模型的訓練效果。

IS 訓練策略可用來結合 PET 或 iPET 訓練架構，以結合 iPET 為例的訓練架構如圖 4.1 所示，並命名為 IS iPET 訓練架構。IS 訓練策略用於圖 4.1 步驟(1) MLM-tuning 處理過程，漸進且有策略地挑選訓練資料集來微調 MLM 模型。對於給定可用於訓練的標示句組段落集合L，IS 訓練策略會對 MLM 模型進行多回合的微調$(t = 0,1, \dots, I)$。各回合模型微調的訓練資料選取方式及訓練採用的損失函數將在以下小節說明。IS 訓練策略亦可結合 PET 訓練架構，而命名為 IS PET 訓練架構。

至於半監督式學習階段，包括圖 4.1 步驟(2)iterative tuning 所示以偽標示句組段落微調 MLM 模型，步驟(3)soft-labeling 對未標示資料進行軟標記，以及步驟

(4)classifier training 訓練目標面向偵測器，這些步驟的處理與 iPET 的原設計架構大致相同，修改的部分將於 4.3 中說明。

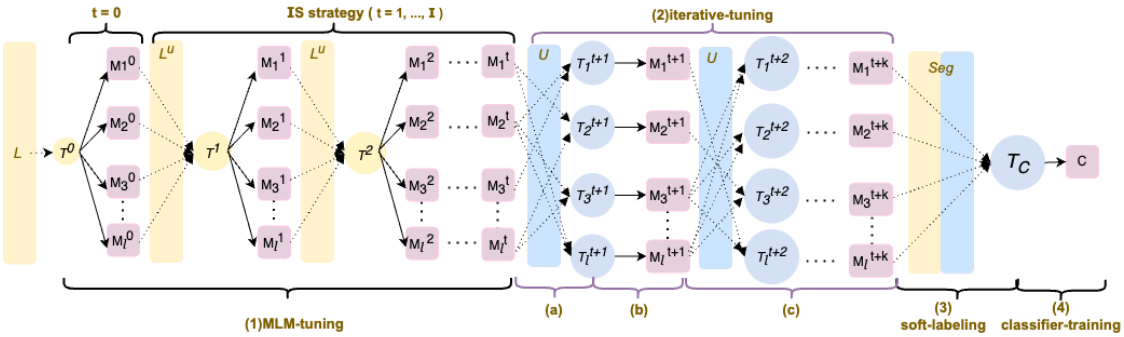


圖 4.1 IS iPET 訓練架構圖

4.2.1 模型微調初始回合訓練方式

本小節說明 MLM 模型微調階段的初始回合訓練資料如何選取，並介紹如何將訓練資料套入文字提示模板進行面向偵測，以提示學習方法對多個隨機初始化參數的 MLM 模型進行微調，建構出多個 MLM 模型，如圖 4.1 步驟（1）所示之 $M_1^0,\ldots,M_l^0$，本論文實驗l設為 4。

(1) 初始回合模型微調

初始回合選取訓練資料的方式，是從標示句組段落集合L中不重複取樣正負類別各 n 筆資料，形成 n 組正負類別資料配對，作為 MLM 模型的輸入；本論文實驗 n 值設定為 20。令L^s表示初始被選取的 $2n$ 個句組段落，即圖 4.1 中所示的 T^0；未被選入的句組段落以L^u表示(L^u=$L-L^s$)，會在接下來的回合再使用。

每一筆輸入 MLM 模型的句組段落$seg=\langle w_1^{seg},w_2^{seg},w_3^{seg},\ldots,w_{len(seg)}^{seg}\rangle$，會接續接合文字提示模板 $Pat=\langle w_1^{Pat},w_2^{Pat},\ldots,[MASK],\ldots,w_{len(Pat)}^{Pat}\rangle$，並在句首插入 [CLS] 符號項，在句尾新增 [SEP] 符號項，將輸入 seg 轉換成一個包含提示

文字及[MASK] 符號項的文字序列$Pat(seg) = \langle [CLS], w_1^{Pat(seg)}, w_2^{Pat(seg)}, \dots, [MASK], \dots, w_{len(Pat(seg))}^{Pat(seg)}, [SEP] \rangle$，其中$w_i^{Pat(seg)}$表示$Pat(seg)$中位置$i$的符號項$(i = 0, \dots, len(Pat(seg)) + 1)$。MLM 模型融合前後文語意得到每個位置相對應的語意表示法，取[MASK]位置的表示法$e_{[MASK]}$，經過一層全連階層，計算出對應到分類任務標示集合$L(l \in \{positive, negative\})$中正負類別之對應字彙$v(l)$的分數分佈，如公式 4.1 所示。

$$s_{MLM}(l|seg) = \text{MLM}([MASK] = v(l)|Pat(seg)) \quad \text{(公式 4.1)}$$

接著透過 softmax 函式進行正規化處理，計算出每個標示類別l對應的機率分佈，如公式 4.2 所示。

$$q_{MLM}(l|seg) = \frac{\exp\left(S_{MLM}(l|seg)\right)}{\sum_{l' \in L} \exp\left(S_{MLM}(l'|seg)\right)} \quad \text{(公式 4.2)}$$

最後透過最大值引數函式$arg\ max(\cdot)$輸出機率最高的類別$l_{predict}$，如公式 4.3 所示。

$$l_{predict} = arg \max_{l \in L} q_{MLM}(l|seg) \text{ (公式 4.3)}$$

(2) MLM 模型損失函數

邊界損失(margin-based ranking loss)是分類任務損失函數常採用的一種設計[30][31]，函數設計的概念是在訓練時同時考慮模型對正負樣本預測結果的差距，以增強模型對不同類別資料的區分能力。本研究的損失函數計算除了考慮資料的預測類別機率與標準答案間的損失值，在進行訓練時以一組正負類別資料配對(seg_i^+, seg_i^-)為單位，兩筆資料個別經模型預測後，為使 MLM 模型在微調時能加強考慮正負類別資料的區別，另將兩筆資料在預測為正類別之機率值差距達一定

邊界門檻值設為學習目標，因此加上一個邊界差距損失值至損失函數，實驗中將此策略以$Strategy_{\text{margin loss}}$表示。

根據 MLM 模型預測[MASK]標籤位置對應的類別機率與標準答案y，損失函數的計算方式採用二元交叉熵(binary cross-entropy)，如公式 4.4 所示。

$$L_{CE}(seg_i^*) = -(y * log(\hat{y}) + (1-y) * log(1-\hat{y})) \quad \text{(公式 4.4)}$$

其中 y 表示訓練資料seg_i^*的實際分類標示（正類別seg_i^+為 1，負類別seg_i^-為 0），$\hat{y}$表示系統預測為正類別的預測機率，相等於公式 4.2 的$q_{MLM}(l=1|seg)$。

為加強 MLM 模型區別正負類別資料間的差異，預期模型對於兩筆訓練資料中的正類別資料預測為正類別的機率，應該大於負類別資料預測為正類別的機率，且差距達到設定的 *margin* 值。因此定義此組訓練資料之邊界損失 (margin loss)以L_{pair}表示，如公式 4.5 所示。

$$L_{pair}(seg_i^+, seg_i^-) = max\left(0, margin - \left(\hat{y}_p - \hat{y}_n\right)\right) \quad \text{(公式 4.5)}$$

其中$\hat{y}_p$及$\hat{y}_n$分別表示模型對正負類別資料seg_i^+及seg_i^-分別預測為正類別的機率。如果$\hat{y}_p$大於$\hat{y}_n$且差距大於所設定的 margin 值，那麼損失函數的值計為零。否則損失函數的值為$margin - \left(\hat{y}_p - \hat{y}_n\right)$。本論文實驗中將 margin 值設定為 0.5。

綜合上述兩部分的損失計算，MLM 模型的損失函數，在訓練階段加總每一組訓練資料配對(seg_i^+, seg_i^-)各別之二元交叉熵損失，再結合其邊界損失L_{pair}，兩部分的損失值可以用α調整其相對佔比$(0 \le \alpha \le 1)$，加總 n 組訓練資料配對，如公式 4.6 所示。本論文實驗α設定為 0.5。

$$L = \frac{1}{n}\sum_{i=1}^{n} [\alpha\left(L_{CE}(seg_i^+) + L_{CE}(seg_i^-)\right) + (1-\alpha)L_{pair}(seg_i^+, seg_i^-)] \quad \text{(公式 4.6)}$$

4.2.2 以多回合微調訓練 MLM 模型

採用本論文所提出的 IS 訓練策略，接下來會將標示資料以多個回合逐步加入訓練資料，以多回合微調訓練 MLM 模型。每一回合的訓練，會保留上一回合已有訓練的資料，再新增一些新的訓練資料，並重新調整模型參數。

每回合漸次增加訓練資料的方法可分為兩種：一種是採用隨機漸次挑選，另一種方法則是有策略挑選訓練資料。挑選基準是根據前一回合微調過的 MLM 模型，對L^u中的資料進行預測並從中選取模型需要加強學習的資料在下一回合加入訓練，在訓練的過程中，有策略地新增訓練資料，以加強模型在特徵判斷的不足之處，並將先前的資料保留重複訓練，避免模型過分偏向下一回新選取的訓練資料。以下分別介紹無策略隨機增加與策略挑選訓練資料的兩種多回合微調訓練方式。

(1) 無策略隨機漸次增加訓練資料

每一回合新增的訓練樣本數量與初始回合相同，從初始回合未被選入訓練的句組L^u中，以不重複採樣正負類別各取 n 筆資料加入訓練資料。第t回合新增的正類別資料以$T^t_{positive}$表示，新增的負類別資料以$T^t_{negative}$表示。

漸次增加訓練資料是指在第t回合會保留第$t-1$回合的訓練資料，因此第t回合用於微調語言模型的訓練資料為T^{t-1}聯集$T^t_{positive}$及$T^t_{negative}$，如公式 4.7 所示。

$$T^t = T^{t-1} \cup T^t_{positive} \cup T^t_{negative} \quad \text{(公式 4.7)}$$

此外被選入的資料不會被重複選取，因此會從L^u中移除，如公式 4.8 所示。

$$L^u = L^u - T^t_{positive} - T^t_{negative} \quad \text{(公式 4.8)}$$

直到L^u中某一類別的樣本數不足 n，則選取所有可用樣本，對於樣本數較多的類別，從剩餘的資料中隨機取樣相同數量加入最後一個回合的T^t，完成對應微調，即停止 MLM 模型微調。

(2) 策略挑選訓練資料方法 $Strategy_{selection}$

策略挑選會以前一回合微調好的 MLM 模型$\mathcal{M}^{t-1}$對未被選入訓練的句組L^u進行預測，對L^u中每筆資料記錄預測的類別標示及預測機率值，產生L^u_{pseudo}。

由於有多個 MLM 模型$\mathcal{M}^{t-1} = \{M_1^{t-1}, \dots, M_{len(\mathcal{M})}^{t-1}\}$，因此這些模型對同一筆資料$seg$在各類別上的預測分數值，會以各模型的相對重要性進行比重加總，如公式 4.9 所示。

$$s_{\mathcal{M}^{t-1}}(l \mid seg) = \frac{1}{Z} \sum_{M_i^{t-1} \in \mathcal{M}^{t-1}} w(M_i^{t-1}) \cdot s_{M_i^{t-1}}(l \mid seg) \quad \text{(公式 4.9)}$$

其中$w(M_i^{t-1})$為語言模型M_i^{t-1}的權重，由模型M_i^{t-1}微調之前即對T^{t-1}進行預測的準確率決定，Z為各模型準確率的總和。

最後以 softmax 函式對各類別預測分數轉換，得到各類別l的預測機率，如公式 4.10 所示。

$$q_{\mathcal{M}^{t-1}}(l|seg) = \frac{\exp\left(S_{\mathcal{M}^{t-1}}(l|seg)\right)}{\sum_{l' \in L} \exp\left(S_{\mathcal{M}^{t-1}}(l'|seg)\right)} \quad \text{(公式 4.10)}$$

[負類別資料挑選策略]

第t回合負類別新增訓練集$T_{negative}^{t}(t = 1,2, \ldots, I)$的挑選策略如下：首先取出$L_{pseudo}^{u}$中預測標示與實際標示(true label)不一致的樣本，即模型誤判為正類別的資料，對這些誤判的負類別資料，本研究認為若誤判為正類別的預測機率值較高，表示是預測偏差大的樣本，應該優先進行訓練以調整模型偏差，因此根據其預測機率值區間化分成五個範圍：[0.9,1)、[0.8,0.9)、[0.7,0.8)、[0.6,0.7)、[0.5,0.6)。接下來從最高值區間[0.9,1)中依序選取，直到累積取到 n 筆資料。若當前區間的資料數不足 n，則全部選取並再從下一個區間補充至足夠數量，若下一區間數量高於需補足數量，則隨機取樣來決定。若誤判為正類別的負類別資料總數小於或等於 n 時，則全部選取。

[正類別資料挑選策略]

第t回合正類別新增訓練集$T_{positive}^{t}(t = 1,2, \ldots, I)$的挑選策略如下：首先取出$L_{pseudo}^{u}$中預測標示與實際標示(true label)相同的正類別資料。對這些已能正確判斷的正類別資料，本研究認為若預測機率值較低，表示是落在預測邊界的樣本，應該加強訓練以擴大正類別樣本涵蓋範圍。因此從中挑選預測機率值最低的 n 筆資料。

[以 IS 訓練策略微調 MLM 模型最終回合]

在進行多回合挑選策略之後，當誤判為正類別的負類別資料總數小於或等於 n 時，構成最後一個回合的$T_{negative}^{t}$，而最後一個回合的$T_{positive}^{t}$依正類別資料挑選策略挑選和$T_{negative}^{t}$相同數量的正類別資料，作為最後一個回合的訓練集T^{t}。待最後一個回合微調結束，以 IS 訓練策略微調 MLM 模型階段即停止。

4.2.3 半監督式學習階段

IS iPET 訓練架構中的半監督式學習階段包括圖 4.2 中的步驟(2)、(3)及(4)。

步驟(2)iterative tuning：此步驟會運用微調後的 MLM($M_1^t, \ldots, M_l^t$)對U產生偽標示，依照 iPET 的方法增加訓練取樣迭代微調 MLM 模型得到$\mathcal{M}^{t+1}$。

步驟(3)soft-labeling：原先 iPET 在此步驟是運用迭代微調後的 MLM($M_1^{t+k}, \ldots, M_l^{t+k}$)對未標示資料集$U$進行偽軟標記，獲得訓練分類器$c$所需的軟標示資料集(pseudo soft label)。由於本論文所採用標示資料集L有一定數量的資料，認為標示資料集L的資料也應該採用預測的軟標記方式加入訓練，以增加訓練樣本。因此進行軟標記的資料來源為Seg(也就是$U \cup L$)。實驗中將此處對軟標示資料集來源選取的改進策略以$Strategy_{soft-labeling\ data}$表示。

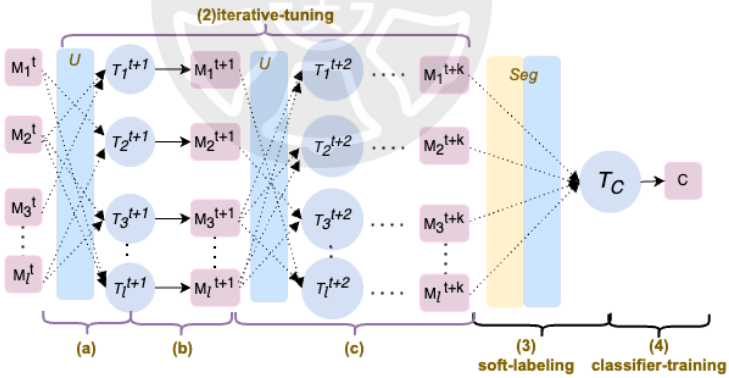


圖 4.2 IS iPET 訓練架構之半監督式學習階段

步驟(4)classifier training：最後一部份是目標面向偵測器訓練階段，以下說明所建構之二分類器架構及訓練目標。本研究採用 BERT 文字序列分類器作為分類器，其模型架構是在 BERT 預訓練語言模型上添加一個全連接層(線性分類層)。訓練資料集T_c中的每筆資料有文字內容seg及對應的軟標示$q_{\mathcal{M}}(l|seg)$，也就是具有其屬於正/負類別的機率值。將每筆訓練資料文字內容，在句首加入[CLS]符號

項，並在句尾加入[SEP]符號項，轉換為 BERT 預訓練語言模型的輸入格式後，輸入至 BERT 預訓練語言模型，取[CLS]位置的表示法代表該筆資料seg的語意向量$e_{[CLS]}$。最後將$e_{[CLS]}$向量經過一層全連接層(線性分類層)，經 softmax 函數預測seg屬於正負類別的機率值，如公式 4.11 所示。

$$q_{CLS}(l|seg) = \text{softmax}\left(W_{[CLS]}e_{[CLS]}\right) \quad \text{(公式 4.11)}$$

在分類器訓練階段，採用 KL 散度計算（Kullback-Leibler divergence）作為損失函數，如公式 4.12 所示。以軟標示$q_{\mathcal{M}}(l|seg)$的機率分佈作為基準，比較分類器預測seg的機率分佈$q_{CLS}(l|seg)$，計算兩個機率分佈間的差異。

$$\begin{aligned} \text{L}_{KL}(seg) &= \text{KL}(q_{\mathcal{M}}(l|seg)\|q_{CLS}(l|seg)) \\ &= \sum_{l\in L} q_{\mathcal{M}}(l|seg)\log\frac{q_{\mathcal{M}}(l|seg)}{q_{CLS}(l|seg)} \end{aligned} \quad \text{(公式 4.12)}$$

4.3 心理健康面向偵測之候選發文篩選方法

由於社交平台上具有心理健康面向的發文為相對少數，大部分發文不具有目標面向。為提高在開放資料集中未標示發文的偵測效率，本論文提出對未標示發文可先經過一個初步的篩選機制，選出較有可能具目標面向的候選發文，再採用目標面向偵測模型判斷是否具有目標面向。

透過目標面向之標示資料集中觀察對應的判斷句，可以發現一些經常出現的關鍵字。以「具有精神疾病」的目標面向為例，下列幾則正類別資料中的判斷句：「就學時有做過憂鬱症的檢測，結果是輕度的」、「像這種天生的憂鬱症能夠得到治療效果嗎」、「這是憂鬱症前兆嗎? 」、「覺得自己可能患有社交恐慌症」、「似乎跟恐慌症的症狀蠻吻合」、「我一直都有強迫症」、「自己有強迫症+社恐」。這些例子中出現了憂鬱症、強迫症和恐慌症等精神病症的關鍵字詞。如果發文中出現這些關鍵字，那麼很可能是正類別標示。

本論文使用自動化的方式找出某個目標面向A_j的代表關鍵字詞。我們採用詞頻（term frequency, TF）分析，比較字詞在面向A_j正類別資料的判斷句與負類別資料中出現的狀況，計算字詞對正類別資料判斷的重要性，找出目標面向A_j對應的代表關鍵字詞集合，以下將介紹處理過程。

首先，將一個目標面向A_j所有標示正類別資料的判斷句接合為一個文件，以D_p表示，而所有負類別文章也接合成一個文件，以D_n表示。接著D_p和D_n經過斷詞處理，並以$D.term$表示將文件D_p斷詞後出現的詞所構成的候選詞集合。

以下公式 4.13 用來計算$D.term$中每個候選詞$term_i$在文件D_*中的詞頻：

$$\mathrm{TF}(term_i, D_*) = \frac{D_*.count(term_i)}{|D_*|} \quad \text{(公式 4.13)}$$

其中$D_*.count(term_i)$表示候選詞$term_i$在D_*中的出現次數，$|D_*|$表示文件D_*中所有詞出現的次數總和；由於正類別資料的代表關鍵字詞在正類別文件的出現頻率應相對較高，而在負類別文件的出現頻率相對較低，因此本研究比較同一個候選詞$term_i$在正類別文件D_p和負類別文件D_n的TF值，將兩者相除，取該相對比較值落在$Top-N$的候選詞$term_i$形成面向A_j的代表關鍵字詞集合，以W^{A_j}表示，如以下公式 4.14 所示。

$$W^{A_j}(D_p, D_n) = Top - N \left\{ \frac{\mathrm{TF}(term_i, D_p)}{\mathrm{TF}(term_i, D_n)} \middle| term_i \in D.terms \right\} \quad \text{(公式 4.14)}$$

表 4.2 所示為本研究對三個心理健康面向A_j擷取出的面向關鍵字詞。

表 4.2 心理健康面向代表關鍵字詞擷取結果

目標面向	代表關鍵字詞
具有精神疾病	焦慮(症) 、強迫(症)、失眠、失調、思覺、恐懼(症)、妄想(症)、恐慌、檢測、自殘、自律、吻合、失調(症)、身心科、被害、傾向、季節、人格、自殺、鬱症、診斷、安眠藥、幻聽、厭食(症)、鎮定劑、社交、憂鬱(症)、精神病
具有情緒問題	低落、崩潰、焦慮、緊張、心情很差、煩躁、憤怒、情緒、恐懼、大哭、起伏、痛苦、鬱悶、低潮、恐慌、不穩定悶悶、不開心、不愉快、負面不快樂、胸口、沒來由、焦躁、心痛、生理、憂鬱、眼淚、心情、起伏很大、蟑螂自卑、深淵、起起伏伏、更糟、反反覆覆、不安、好慘、自責、現象、內心、哭泣、悲觀、糟糕、悲傷、出現、敏感、空虛、哭
具有求助行為	回診、心理師、乖乖、張老師、報到、先前、住院、鬱症、諮商、諮商師、近況、確切、劑量、網站、廣泛性、理所當然、瘋子、季節性、換藥、慮病、安寧、每週、耳鼻喉科、宗教、拿藥、服藥、出院、失真、長庚、病房、診斷、身心科、行屍走肉、實習、求診

根據找出的面向代表關鍵字詞集合，給定一個測試資料集DB^{Test}，其中的一篇發文d，若出現W^{A_j}中任一個關鍵字詞 w，則歸類至$DB^{Test}_{DB_{w/\,keywords}}$，如公式 4.15 所示；表示為具有面向$A_j$對應代表關鍵字詞的候選發文集合。

$$DB^{Test}_{DB_{w/\,keywords}} = \{d \mid d \in DB^{Test} \wedge \exists w \in W^{A_j} \wedge w \in d.term\} \quad \text{(公式 4.15)}$$

本研究將在實驗中評估採用測試資料集DB^{Test}直接進行目標面向偵測，以及經面向代表關鍵字詞篩選得到$DB^{Test}_{DB_{w/\ keywords}}$後再進行目標面向偵測，比較兩者的偵測效果。

第五章 實驗評估與討論

本章將針對本論文所提出的 IS 訓練策略結合 PET 和 iPET 之訓練架構，評估所建構出的心理健康面向偵測模型之預測效能。首先在第一小節介紹實驗採用的資料集，第二小節介紹 MLM 模型與分類器模型的參數設定，第三小節說明實驗採用的評估指標，第四小節說明在封閉標示資料集的實驗設計與結果討論，並於第五小節說明在模擬開放環境測試資料集的實驗結果。

5.1 資料集說明

5.1.1 標示資料集

本研究對發文內容進行三個心理健康面向的偵測任務：(1)判定是否具有精神疾病、(2)是否具有情緒問題，以及(3)是否出現尋求協助行為。考慮的資料集來源為臺灣電子佈告欄批踢踢實業坊(PTT)的中文發文，根據三個目標面向由人工選取部分發文並經標示後，是否具有精神疾病(mental disorder, MD)，是否具有情緒問題(emotional disorder, ED)，以及是否有尋求協助行為(help-seeking behavior, HB)的標示資料集分別以 DB_{MD}、DB_{ED}、DB_{HB} 表示，各資料集的統計資訊如表 5.1 所示。

表 5.1 心理健康面向資料集之統計資訊

	DB_{MD}	DB_{ED}	DB_{HB}
正類別發文數量	318	381	320
負類別發文數量	490	328	483
判斷句平均長度	15	14	14

5.1.2 開放環境測試資料集

為了評估真實開放環境時的偵測預測效果，本研究另外蒐集臺灣電子佈告欄批踢踢實業坊(PTT)Prozac、Psychiatry 和 WomenTalk 三個看板共 5500 則的發文。這些發文經標示資料集訓練過的偵測模型預測，再將預測結果交由專家進行確認。由於預測結果中的負類別數量極大，因此模型預測為正類別的發文全數確認，而預測為負類別的發文則抽樣檢核。經過人工確認標示的抽樣資料為本實驗測試集以 DB^{open} 表示，共 2122 筆發文，資料集中涵蓋各心理健康面向正負類別的發文數量統計如表 5.2 所示。

開放環境測試資料集 DB^{open} 在實驗中將先用來評估各目標面向偵測模型在面向 MD、ED、HB 的單獨偵測效能。接著再評估三個面向偵測模型的綜合預測效能：將各篇發文先經過「具有精神疾病」和「具有情緒問題」模型的篩選，任一模型認為具有精神疾病或情緒問題，才輸入「具有求助行為」偵測模型判別是否有求助行為。

表 5.2 開放環境測試資料集中正負類別發文數量

	DB^{open}		
目標面向	MD	ED	HB
正類別發文數量	283	827	220
負類別發文數量	1839	1295	1902

5.2 實驗參數設定

本論文採用深度學習系統 Pytorch 的模組進行實作，PET 與 iPET 訓練架構中的 MLM 模型與分類器模型相關實驗設定參數如表 5.3、表 5.4 所示。

表 5.3 MLM 模型參數設定列表

參數名稱	參數意義	PET 訓練架構	iPET 訓練架構
Hidden size	嵌入層向量維度	768	
Batch size	批次訓練大小	4	
Dropout rate	隨機關閉神經元比例	0.1	
Learning rate	學習率	1e-5	
Max_seq_length	輸入文本之最長字數	256	
Epochs	訓練週期	3	

表 5.4 分類器模型參數設定列表

參數名稱	參數意義	PET 訓練架構	iPET 訓練架構
Hidden size	嵌入層向量維度	768	
Batch size	批次訓練大小	4	
Dropout rate	隨機關閉神經元比例	0.1	
Learning rate	學習率	1e-5	
Max_seq_length	輸入文本之最長字數	256	
Epochs	訓練週期	3	

5.3 評估指標

本論文採用分類任務的 Accuracy(A)、Precision(P)、Recall(R)、F1-score(F1)作為面向偵測效果的評估指標，以下將分別說明四種評估指標的計算公式。

(1) Accuracy

計算模型預測正確的樣本數佔總樣本數的比例，計算方式如公式 5.1 所示。

$$A = \frac{\text{正確預測的樣本數}}{\text{測試總樣本數}} \text{ (公式 5.1)}$$

由於本研究的主要任務是偵測出目標面向的正類別發文，因此，以下評估指標 Precision 和 Recall 針對正類別(positive)的偵測結果進行評估。

(2) Precision

計算模型偵測正類別的準確率，計算方式如公式 5.2 所示。

$$P = \frac{\text{預測為正類別且標示為正類別的資料筆數}}{\text{預測為正類別的筆數}} \text{ (公式 5.2)}$$

(3) Recall

計算模型全部正類別資料中被偵測出的比率，此指標稱為偵測正類別資料的召回率，計算方式如公式 5.3 所示。

$$R = \frac{\text{預測為正類別且標示為正類別的資料筆數}}{\text{測試資料中所有正類別的筆數}} \text{ (公式 5.3)}$$

(4) F1-score

為了評估模型在正類別偵測之 Precision 和 Recall 的綜合表現，採用 F1-score 來評估，計算方式如公式 5.4 所示。

$$F1 = \frac{2 * P * R}{P + R} \text{ (公式 5.4)}$$

5.4 封閉標示資料集之實驗設計與結果討論

在封閉標示資料集的實驗分成以下四個部分：

[實驗1] IS訓練策略對PET和iPET的訓練增進效果評估。

[實驗2] IS訓練策略在訓練樣本減少情境下的影響評估。

[實驗3] IS iPET訓練架構中各提出策略的效能評估。

[實驗4]句組段落最小字數設定對IS iPET模型訓練架構的效果評估。

實驗 1 到實驗 4 使用的訓練和測試資料集為 5.1.1 所介紹的三個標示資料集 DB_{MD}、DB_{ED} 及 DB_{HB}。實驗中採用 5-fold 交互驗證進行測試，評估比較不同訓練架構在三個心理健康面向偵測的訓練效果。表 5.5 所示為其中一個訓練回合中作為訓練資料的發文經過資料處理，轉換為句組段落形式的數量資訊。

表 5.5 單一訓練回合中訓練資料發文之句組段落數量

	DB_{MD}	DB_{ED}	DB_{HB}
正類別句組段落數	642	859	579
負類別句組段落數	5704	3587	12048
未標示句組段落數	7888	9475	7022

[實驗 1] IS 訓練策略對 PET 和 iPET 的訓練增進效果評估

本實驗的目的在評估本論文所提出漸次增加與策略挑選訓練資料的 IS 訓練策略，對其在 MLM 模型和分類模型的效能進行評估。實驗再細分為兩部分：

[實驗 1.1] 評估 IS 訓練策略對單採用 MLM 模型提示學習的預測提升效果

[實驗 1.2] 評估 PET 和 iPET 結合 IS 訓練策略的預測提升效果

[實驗 1.1] 評估 IS 訓練策略對單採用 MLM 模型提示學習的預測提升效果

本實驗評估以提示學習進行微調的MLM模型，是否使用IS訓練策略對MLM模型在三個心理健康面向的偵測效果。

未使用 IS 訓練策略的 MLM 模型用來相對比較，其在資料集正負類別筆數不同的情況下，會以數量較少的正類別資料數量為基準，從負類別資料中隨機採樣相同數量的資料，對 MLM 模型進行一次微調。實驗結果如表 5.6 所示其中各評估指標的較佳數據會以底線標示。表 5.6 的結果顯示，當 MLM 模型採用 IS 訓練策略，三個目標面向偵測的 Precision 及 F1-score 均有明顯的提升。尤其在 Precision 的提升，顯示使用 IS 訓練策略漸進微調 MLM 模型，有助於模型更有效學習到正類別的判斷條件，減少負類別資料誤判為正類別。在 DB_{MD} 及 DB_{ED} 的 Recall 值有略微下降，但都在 7%以內；在 DB_{HB} 的 Recall 值雖然下降較多(12%)但 Precision 值提升更多(20%)，因此整體 F1-score 值仍然提升。

表 5.6 MLM 模型微調是否採用 IS 訓練策略的預測效果

	DB_{MD}			DB_{ED}			DB_{HB}		
是否使用 IS 策略	P	R	F1	P	R	F1	P	R	F1
否	0.605	<u>0.975</u>	0.747	0.725	<u>1</u>	0.841	0.579	<u>0.975</u>	0.726
是	<u>0.774</u>	0.906	<u>0.835</u>	<u>0.793</u>	0.989	<u>0.881</u>	<u>0.773</u>	0.853	<u>0.811</u>

[實驗 1.2] 評估 PET 和 iPET 結合 IS 訓練策略的預測提升效果

本小節評估 PET 及 iPET 訓練架構，是否結合 IS 訓練策略，對建構三個心理健康面向偵測模型的預測效果評估。

表 5.7 上半部顯示 PET 和 iPET 訓練架構未結合 IS 訓練策略在三個目標面向資料集的預測效果。iPET 相較於 PET 在三個目標面向資料集的 Precision 及 F1-score 都略微下降。這結果呼應本論文所述，用標示資料對 MLM 模型進行一次微調，未必能達到足夠的訓練效果。因此當 iPET 運用第一次微調的 MLM 模型逐步加入偽標示資料，提供額外資訊對 MLM 模型進行微調，因為錯誤的偽標示資料，導致 MLM 模型的學習產生偏差，使得以 iPET 訓練的目標面向偵測模型之預測效果反而比 PET 訓練的效果差。

表 5.7 PET 和 iPET 是否結合 IS 訓練策略的預測效果

	DB_{MD}				DB_{ED}				DB_{HB}			
	A	P	R	F1	A	P	R	F1	A	P	R	F1
PET	0.748	0.612	0.981	0.754	0.777	0.706	1	0.828	0.72	0.59	0.972	0.734
iPET	0.726	0.592	0.984	0.739	0.758	0.69	1	0.816	0.707	0.578	0.988	0.729
結合 IS 訓練策略												
IS PET	0.869	0.808	0.874	0.84	0.856	0.793	0.989	0.881	0.863	0.811	0.856	0.833
IS iPET	0.871	0.807	0.884	0.844	0.862	0.797	0.995	0.885	0.872	0.822	0.866	0.843

表 5.7 下半部顯示 IS PET 及 IS iPET 訓練架構在三個目標面向資料集的預測效果。比較 PET 和 IS PET 訓練架構的預測效果，可發現 PET 結合 IS 訓練策略後在 Precision 值的表現，在 DB_{MD} 提升將近 20%，在 DB_{ED} 提升將近 10%，且在 DB_{HB} 提升 22%。另一方面 iPET 和 IS iPET 訓練架構的預測效果，iPET 結合 IS 訓

練策略後 Precision 值，在 DB_{MD} 提升 21%，在 DB_{ED} 提升 10%，且在 DB_{HB} 提升近 25%。上述結果中不論是 PET 或是 iPET，結合 IS 訓練策略皆可大幅提升模型預測效果，顯示 IS 訓練策略的有效性。

由於 IS 訓練策略的目標是減少模型將負類別誤判為正類別的狀況，以提升正類別偵測的 Precision，因此挑選策略的訓練目標偏重於避免負類別誤判為正類別，且強化訓練已落在預測邊界的正類別樣本。結合 IS 訓練策略後，上述 IS PET 及 IS iPET 在 Precision 的提高顯示可達到上述目標。而 IS PET 及 IS iPET 的 Recall 值下降現象，表示在調整模型對正負類別的預測偏差時，仍難以避免排除部分落在預測邊界的正類別資料。整體來說，儘管模型可能會漏失一些正類別資料，造成 Recall 值下降，但相對能夠更有效地排除被誤判的負類別資料，實驗結果顯示結合 IS 訓練策略的 IS PET 及 IS iPET 皆可明顯提升 F1-score。

配合實驗 1.1 的觀察顯示， IS 訓練策略使 MLM 模型在標示資料微調階段的預測準確率達到 0.77 以上，因此 IS iPET 在隨後的迭代訓練中獲得品質較佳的偽標示句組段落繼續微調 MLM 模型，使得 IS iPET 與 IS PET 有相近的 Precision，並進一步提升 Recall 值。若比較 IS PET 及 IS iPET，從表 5.7 下半部顯示，IS iPET 和 IS PET 的 Precision 值相近或略有增進且在 Recall 值上皆有提升，因此整體 F1-score 皆提升。

[實驗 2] IS 訓練策略在訓練樣本減少情境下的影響評估

此實驗模擬當訓練資料中的發文數量降低時，觀察對 iPET 與 IS iPET 訓練架構的預測效果影響。

本實驗從實驗 1 的 5fold 交叉驗證中每回分配的訓練資料集中隨機取樣原訓練發文數量的 75%、50%、25%和 10%，將這些發文轉換為句組段落進行訓練，再使用和實驗 1 同樣的測試資料集進行測試。表 5.8 顯示 DB_{MD} 中某一回訓練集經過不同取樣比例，轉換為句組段落形式的數量統計，另外於附錄一說明 DB_{ED} 及 DB_{HB} 的數量統計。

表 5.8 5fold 交叉驗證單一訓練回合中的句組段落數量(DB_{MD})

	DB_{MD}				
發文取樣比例	100%	75%	50%	25%	10%
正類別句組段落數	642	485	317	177	66
負類別句組段落數	5704	4325	2780	1556	481
未標示句組段落數	7888	6187	3867	2627	546

本實驗以 Precision 作為評估指標，比較 iPET 訓練架構是否採用 IS 訓練策略在訓練集不同取樣比率下的效果，表 5.9 所示為三個心理健康面向偵測器的預測效果。在 IS iPET 的預測結果中，訓練集全採用的數據以粗體底線標示，訓練集採用 10%的數據則以底線標示；而在 iPET 的預測結果中，訓練集全採用的數據會以灰底標示。

隨著取樣比例降低，兩個訓練架構建構的模型在 Precision 均有下降趨勢。然而，當訓練樣本減少至 50%時，IS iPET 在各面向偵測器下降的幅度都在 5%之內；當訓練樣本減少至 25%時，各面向偵測器 Precision 下降的幅度都在 10%之內。當

訓練樣本減少至 10%的情況，IS iPET 的 Precision 值仍高於 iPET 在未減少訓練樣本時的表現。

表 5.9 訓練資料集不同取樣比率的預測效果評估

	DB_{MD}		DB_{ED}		DB_{HB}	
	IS iPET	iPET	IS iPET	iPET	IS iPET	iPET
100%	**0.807**	0.592	**0.796**	0.69	**0.822**	0.578
75%	0.783	0.583	0.761	0.687	0.808	0.552
50%	0.757	0.581	0.75	0.674	0.779	0.543
25%	0.708	0.543	0.736	0.663	0.747	0.485
10%	0.643	0.548	0.692	0.635	0.668	0.452

此外，對照表 5.8 的資料數量，可得知當發文取樣比例為 10%時，正類別句組段落數皆小於 100。此結果顯示 IS 策略在數量小於 100 的少量訓練樣本情境(Few-shot scenario)，仍能輔助 iPET 從訓練集中有效挑選出對增進模型正確率學習有幫助的資料，有效提升模型預測的 Precision。

[實驗 3] IS iPET 訓練架構中各提出策略的效能評估

本論文提出的 IS iPET 訓練架構中包含幾項主要的策略改進，本實驗透過移除單項策略來分析各策略對 IS iPET 訓練架構的效能影響。

三種用來比較的實作版本，分別是：

(1) w/o $Strategy_{selection}$：這個實作版本中，在逐步增加訓練資料的過程中沒有使用策略挑選，而是採用隨機挑選。

(2) w/o $Strategy_{\text{margin loss}}$：這個實作版本中，微調 MLM 模型時的損失函數僅考慮預測正確標示的機率，未加上正負類別資料在正類別預測機率的差異損失值。

(3) w/o $Strategy_{soft-labeling\ data}$：這個實作版本中，用來產生偽軟標示資料集的資料來源只採用未標示資料集U，而非採用所有句組資料集Seg。

表 5.10 移除單項策略實驗預測結果

	DB_{MD}			DB_{ED}			DB_{HB}		
	P	R	F1	P	R	F1	P	R	F1
IS iPET	<u>0.807</u>	0.884	<u>0.844</u>	<u>0.796</u>	0.995	<u>0.884</u>	<u>0.822</u>	0.866	<u>0.843</u>
w/o $Strategy_{selection}$	0.604	<u>0.987</u>	0.749	0.686	<u>1</u>	0.814	0.57	<u>0.975</u>	0.72
w/o $Strategy_{\text{margin loss}}$	0.738	0.928	0.822	0.768	0.995	0.867	0.786	0.897	0.838
w/o $Strategy_{soft-labeling\ data}$	0.747	0.893	0.814	0.758	0.995	0.86	0.787	0.866	0.824

表 5.10 顯示三個實作版本與採用 IS iPET 訓練架構在三個心理健康面向資料集的預測效果，結果顯示移除 IS iPET 中任一個策略都會使 Precision 和 F1-score 下降。在這些實作版本中，移除策略挑選的版本(w/o $Strategy_{selection}$)在 Precision 的預測效能下降最明顯，表示在訓練資料逐步增加的過程中，策略挑選對整體訓練效果有最主要的影響。

若再觀察各版本訓練結果對應的 Recall 值，可發現，w/o $Strategy_{selection}$版本訓練偵測器之 Recall 值皆很接近 1，表示未進行策略挑選可能擴大正類別學習到的資料偵測範圍，卻同時有更多的負類別資料被誤判。而進行策略挑選後偵測器雖然漏失部分正類別資料，在 Precision 則大幅提升，因此顯示挑選策略的有效性。

此外，$Strategy_{soft-labeling\ data}$對偵測器訓練效果的影響雖然最小，但加上此策略對 Precision 能提升超過 4%，顯示當具有一定數量的標示資料L，iPET 最後建立分類器的軟標示資料集來源除了採用未標示資料U，將標示資料L也採用軟標記方式加入訓練，對分類器的預測 Precision 仍有提升的幫助。

[實驗4] 句組段落最小字數設定對IS iPET模型訓練架構的效果評估

如本論文 3.2 小節，本論文指定句組段落最小字數l的方法如下：將各判斷句包含的字數進行離群值分析，去除判斷句字數的離群值後，計算平均字數取整數作為l的設定值。本實驗將句組段落最小字數要求分別設為l及 $2l$，比較 IS iPET 訓練架構所建構模型的預測效果。

表 5.11 句組段落最小字數不同設定對 IS iPET 建構模型的預測效果評估

	DB_{MD}				DB_{ED}				DB_{HB}			
最少字數要求	A	P	R	F1	A	P	R	F1	A	P	R	F1
l	0.871	0.807	0.884	0.844	0.862	0.797	0.995	0.885	0.872	0.822	0.866	0.843
$2l$	0.854	0.748	0.953	0.838	0.807	0.737	0.997	0.848	0.864	0.802	0.875	0.837

表 5.11 顯示不同句組段落最小字數設定所建構模型的預測效果。比較句組段落最小字數設定為l與 $2l$時所建構的面向偵測器預測效果，發現隨著句組段落長度增加，Recall 值會提升，但 Precision 有明顯下降。推測因為句組段落字數的增加，會因為資訊更完整而減少偵測的漏失，但也使得模型因資訊量過多無法聚焦在明確顯示目標面向的文字段落，而導致模型偵測面向正類別的準確度下降，因此上述實驗將句組段落最小字數設定為判斷句平均長度l。

5.5 開放環境測試資料集之實驗設計與結果討論

在開放環境測試資料集的實驗分成以下兩個部分：

[實驗 5] IS iPET 所建構之面向偵測器於開放環境測試資料集的預測評估。

[實驗 6] 經面向代表關鍵字詞篩選後在各心理健康面向的預測效果評估。

[實驗 5] IS iPET 所建構之面向偵測器於開放環境測試資料集的預測評估

由於開放環境測試資料集較具挑戰，本研究根據原封閉標示資料集訓練後，初步在開放環境測試資料集發現的錯誤案例分析(如附錄二)，歸納原因是封閉標示資料集中人工蒐集時偏重於正類別資料，而使負類別訓練資料缺乏某些類型內容敘述案例。因此在此實驗，會補充這些類型的負類別發文資料加入DB_{MD}、DB_{ED}及DB_{HB}，加入後的訓練資料集分別以DB_{MD}^{extend}、DB_{ED}^{extend}及DB_{HB}^{extend}表示，各資料集中的資料數量如表 5.12 所示。

表 5.12 心理健康面向擴增資料集資料數量

	DB_{MD}^{extend}	DB_{ED}^{extend}	DB_{HB}^{extend}
正類別發文數量	318	381	320
負類別發文數量	490	328	483
新增負類別發文數量	96	94	51

本實驗將上述擴展的資料集作為訓練集，對開放環境測試資料集進行預測。首先評估三個心理健康面向各自的偵測效果。為了符合心理健康素養 HB 面向的定義，選取符合條件後進行是否有求助行為的分析，因此在開放資料集綜合評估面向的偵測效果，將DB^{open}的發文先經過「具有精神疾病」和「具有情緒問題」模型的篩選，

任一模型認為其具有精神疾病或情緒問題時，才輸入「具有求助行為」模型判別是否有求助行為。

表 5.13 中先顯示三個目標面向偵測器在開放測試集中的各自偵測效果，結果顯示 MD 及 ED 面向偵測器之 Precision 值皆可到達 0.8。HB 面向偵測器的 Precision 值雖然為 0.67 略低於另兩個面向偵測器，但如果觀察接下來的綜合偵測效果：先經過 MD 及 ED 偵測器排除不具有精神疾病且不具有情緒問題的發文，再輸入進模型偵測 HB 面向，則可以有效排除在 HB 面向會被誤判為正類別的發文，使判斷求助行為的偵測 Precision 增進 10%而提高至 0.77，且 Recall 值維持不變。

表 5.13 開放環境測試資料集偵測效果

	心理健康面向偵測器											
	MD				ED				HB			
	A	P	R	F1	A	P	R	F1	A	P	R	F1
各自偵測效果	0.96	0.84	0.88	0.86	0.88	0.79	0.92	0.85	0.93	0.67	0.76	0.71
綜合偵測效果									0.9	0.77	0.76	0.76

上述結果是因為直接使用 HB 面向模型進行偵測時，可能將一些原因不明確去就醫或諮商的發文視為出現求助行為；但不符合心理健康素養中因為具精神疾病或情緒問題後就醫或諮商才視為出現求助行為。先進行 MD 和 ED 面向偵測，可多一層篩選將這些不具有精神疾病和情緒問題的發文排除，避免它們被 HB 偵測為出現求助行為。

表 5.14 單獨採用 HB 面向偵測器被誤判為出現求助行為的發文案例

id	發文內容
1711-40	每次到禮拜日晚上 總是特別的開心 因為明天又是全新的一個禮拜 **代表著諮商的日期靠近了 諮商是 3 個月之前就預約 所以我非常珍惜 有很多話要跟諮商師說** 有板友也是喜歡收假日嗎 期待 9/24 王心凌的演唱會 現在要開始練歌 希望特別嘉賓是蔡依林
1691-113	**不知不覺從心理師建議我這事**， 已經過了一百天， 剛開始練習的時候真的備受艱辛， 覺得自己一事無成又有什麼好紀錄的？ 但練習了這幕久，漸漸的我也開始習慣了這本本子的存在 每天書寫練習也讓自己越來越有成就感。 https//i.imgur.com/uXctLhR.jpg 剛開始都不知道要寫什麼， 只好把吃飯、睡覺、**回診**之類的都寫上去， https//i.imgur.com/EdqIzvI.jpg https//i.imgur.com/Tm4Q5gt.jpg 而到近期越來越能發現生活中的美好事物，逐漸變得豐富了起來， 能夠累積 100 天的紀錄真的覺得好開心， 接下來繼續朝著 200 天、300 天、一年邁進 謝謝你喜歡我的字 XD 謝謝(ω)/（這應該算是稱讚吧 XD） 謝謝你 o(//////)q 謝謝你，

表 5.14 所示為DB^{open}中單獨採用 HB 面向偵測器被誤判為出現求助行為的兩篇發文案例(1711-40 及 1691-113)。1711-40 因為提到「代表著**諮商的日期**靠近了 **諮商**是 3 個月之前**就預約** 所以我非常珍惜 有很多話**要跟諮商師說**」，1691-113 因為提到「不知不覺從**心理師建議我**這事」、「只好把吃飯、睡覺、**回診**之類的都寫上去」而被 HB 偵測模型偵測為出現求助行為。然而，這兩篇發文內容並未提及心理健康問題，不被認定為求助行為，因此經由 MD 和 ED 面向偵測器正確將其排除，提高 HB(出現求助行為)的偵測 Precision。

綜合以上討論，MD 和 ED 面向偵測的預測效果均達到實用水準。此外，經過 MD 和 ED 面向偵測器的篩選之後再進行 HB 面向的偵測效果，同樣也能夠達到實用水準。

[實驗 6] 經面向代表關鍵字詞篩選後在各心理健康面向的預測效果評估

針對不同心理健康面向，本小節將DB^{open}中的 2122 篇發文經面向代表關鍵字詞篩選得到$DB^{open}_{DB_{w/\,keywords}}$，探討其綜合預測效能。

經過面向代表關鍵字詞篩選後，在 MD 面向留下 590 篇發文，篩選掉 72%的資料；在 ED 面向留下 909 篇發文，篩選掉 57%的資料；在 HB 面向留下 305 篇發文，篩選掉 70%的資料。

表 5.15 為DB^{open}測試集是否經過面向代表關鍵字詞篩選在各目標面向的預測效果。表中的結果顯示，透過面向代表關鍵字詞篩選，在各目標面向偵測的 Precision 值皆有提升。儘管 Recall 值有所下降，但在真實開放測試時無法掌握實際正類別的確切數量，因此 Precision 值的提升更為重要。

綜合上述結果顯示，經本研究提出的面向代表關鍵字詞篩選步驟，能夠有效排除大量不具面向的發文，從而提升在真實開放環境測試資料集中的執行效率及偵測 Precision。

表 5.15 DB^{open}是否經面向代表關鍵字詞篩選之預測效果

	心理健康面向偵測器											
	MD				ED				HB			
測試資料集	A	P	R	F1	A	P	R	F1	A	P	R	F1
DB^{open}	0.96	0.84	0.88	0.86	0.88	0.79	0.92	0.85	0.9	0.77	0.76	0.76
$DB^{open}_{w/\,keywords}$	0.85	0.85	0.8	0.83	0.72	0.85	0.76	0.8	0.66	0.81	0.58	0.67

第六章 結論與未來研究方向

本論文提出一個可套用於 PET 及 iPET 訓練架構的 IS 訓練策略，在提示學習微調 MLM 模型階段，以分批逐漸加入訓練資料的方式對模型進行多回合微調，同時透過了解前一回合模型已學會的和仍然誤判的資料範例，決定每回合模型需加強學習的新增訓練樣本，多回合調整模型學習上的偏差，使模型在微調階段達到更好的訓練效果。此外，本研究對 MLM 模型損失函數提出改進的策略除了考慮正負類別資料各自的預測機率損失值，同時引入邊界差異損失值，以強化模型區別正負類別資料的能力。

實驗結果顯示：結合 IS 訓練策略的 PET 和 iPET 訓練架構所建構的分類模型，在三個心理健康面向資料集的評估結果中，Precision 能提升 20%，達到 0.8 以上。當控制訓練樣本數減少至一半，採用 IS iPET 訓練策略建構的面向偵測器之 Precision 值仍能維持在 0.75 以上。上述結果顯示本論文提出之 IS iPET 訓練架構對增進提示學習訓練文件分類器的有效性。在模擬開放環境測試資料集，本研究綜合運用 MD 及 ED 偵測器篩選出符合心理健康狀態有問題的發文並偵測其是否有求助行為(HB)，Precision 值可達到 0.81，顯示本論文方法於社交媒體文本，用來偵測心理健康面向的可用性。

本研究在實驗結果中發現此主題在未來有幾項可持續探討改進的方向。首先，由本研究在訓練樣本減少的實驗結果顯示，訓練資料的數量未必越多越有效果，而是挑選出讓模型有目標進行調整學習的訓練資料更具影響，因此除了本論文提出的方法，其他挑選策略值得持續探討。此外，本研究實驗也發現提供負類別更多不同類型的訓練資料，更有效提升偵測效果，因此未來可進一步考慮運用心理健康面向

候選發文的篩選方法，對不具有正類別關鍵字的發文資料進行分群再採樣，以擴展不同類型負類別訓練資料的自動蒐集。另一方面，可考慮主動學習(Active Learning)提出的概念，以原有標示資料建構模型後，對未標示資料進行預測，並設計預測結果的不確定性評估方法，挑選出對模型效能提升較有幫助的資料請專家進行標記，以適當補充標示資料，並盡量減少人工標示成本。

參考文獻

[1]. Cepeda, N.J., Pashler, H., Vul, E., Wixted, J.T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin, 132(3), 354-80.*

[2]. Chao, H., Lien, Y., Kao, Y., Tasi, I., Lin, H., & Lien, Y. (2020). Mental Health Literacy in Healthcare Students: An Expansion of the Mental Health Literacy Scale. *International Journal of Environmental Research and Public Health, 17.*

[3]. Cui, G., Hu, S., Ding, N., Huang, L., & Liu, Z. (2022). Prototypical Verbalizer for Prompt-based Few-shot Tuning. *In Proceedings of 60th Annual Meeting of the Association for Computational Linguistics, pages 7014–7024.*

[4]. Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186.*

[5]. Gao, T., Fisch, A., & Chen, D. (2021). Making Pre-trained Language Models Better Few-shot Learners. *In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, pages 3816–3830.*

[6]. Harris, Z.S. (1954). Distributional Structure*, WORD, 10:2-3 , pages 146-162.*

[7]. Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9, 1735-1780.*

[8]. Hambardzumyan, K., Khachatrian, H., & May, J. (2021). WARP: Word-level Adversarial ReProgramming. *In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Association for Computational Linguistics, 4921–4933.*

[9]. Jorm, A.F., Korten, A.E., Jacomb, P.A., Christensen, H., Rodgers, B., & Pollitt, P.A. (1997). “Mental health literacy”: a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. *Medical Journal of Australia, 166.*

[10]. Kluger, A.N., & Denisi, A.S. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119, 254-284.*

[11]. Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *In Conference on Empirical Methods in Natural Language Processing (EMNLP’14)*.

[12]. Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A Convolutional Neural Network for Modelling Sentences. *In Proceedings of the 52nd Annual Meeting of the*

Association for Computational Linguistics. Association for Computational Linguistics, pages 655–665.

[13]. Kutcher, S., Wei, Y., & Coniglio, C. (2016). Mental health literacy: Past, present, and future. *In The Canadian Journal of Psychiatry / La Revue canadienne de psychiatrie, 61(3), pages 154–158.*

[14]. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *In Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324.*

[15]. Liu, C., Sheng, Y., Wei, Z., & Yang, Y. (2018). Research of Text Classification Based on Improved TF-IDF Algorithm. *2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE), 218-222.*

[16]. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2021). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys, 55, 1 - 35.*

[17]. Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., & Tang, J. (2021). GPT Understands, Too. *ArXiv, abs/2103.10385.*

[18]. Mojtabai, R., Evans-Lacko, S., Schomerus, G., & Thornicroft, G. (2016). Attitudes Toward Mental Health Help Seeking as Predictors of Future Help-Seeking Behavior and Use of Mental Health Treatments. *Psychiatric services, 67 6, 650-7 .*

[19]. O'Connor, M., & Casey, L.M. (2015). The Mental Health Literacy Scale (MHLS): A new scale-based measure of mental health literacy. *Psychiatry Research, 229, 511-516.*

[20]. Salton, G., & Buckley, C. (1988). Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing & Management, 24(5):513–523*.

[21]. Shin, T., Razeghi, Y., Logan IV, R.L., Wallace, E., & Singh, S. (2020). Eliciting Knowledge from Language Models Using Automatically Generated Prompts. *In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 4222–4235*.

[22]. Schick, T., & Schütze, H. (2020). Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference. *In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume (EACL'21), Paola Merlo, Jörg Tiedemann, and Reut Tsarfaty (Eds.). Association for Computational Linguistics, pages 255–269.*

[23]. Schick, T., Schmid, H., & Schütze, H. (2020). Automatically Identifying Words That Can Serve as Labels for Few-Shot Text Classification. *In Proceedings of the 28th International Conference on Computational Linguistics (COLING'20), Donia Scott, Núria Bel, and Chengqing Zong (Eds.). International Committee on Computational Linguistics, pages 5569–5578.*

[24]. Scao, T.L., & Rush, A.M. (2021). How many data points is a prompt worth? *In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for*

Computational Linguistics, pages 2627–2636.

[25]. Turcan, E., & McKeown, K. (2019). Dreaddit: A Reddit Dataset for Stress Analysis in Social Media. *In Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019), pages 97–107, Hong Kong. Association for Computational Linguistics.*

[26]. Tai, K.S., Socher, R., & Manning, C.D. (2015). Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. *In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, pages 1556–1566, Beijing, China. Association for Computational Linguistics.*

[27]. Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. *In Advances in Neural Information Processing Systems, pages 5998–6008.*

[28]. Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S.R. (2018). GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. *In Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, pages 353–355, Brussels, Belgium. Association for Computational Linguistics.*

[29]. Zhu, X., Sobhani, P., & Guo, H. (2015). Long Short-Term Memory Over Recursive Structures. *In Proceedings of the International Conference on Machine Learning. 1604–1612*.

[30]. Gao, J., Pantel, P., Gamon, M., He, X., & Deng, L. (2014, October). Modeling interestingness with deep neural networks. *In Proceedings of the 2014 conference on empirical methods in natural language processing, pages 2-13.*

[31]. Goldberg, Y. (2016). A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research, 57, pages 345-420*

附錄

附錄一 在訓練樣本減少情境下的句組段落數量統計

附表 1.1 與附表 1.2 補充說明實驗 2 的 DB_{ED} 及 DB_{HB}，5fold 交叉驗證中每回分配的訓練資料集中隨機取樣原訓練發文數量的 75%、50%、25%和 10%，並將發文轉換為句組段落的數量統計。

附表 1.1 5fold 交叉驗證單一訓練回合中的句組段落數量(DB_{ED})

	DB_{ED}				
發文取樣比例	100%	75%	50%	25%	10%
正類別句組段落數	859	666	472	200	82
負類別句組段落數	3587	2749	1742	721	320
未標示句組段落數	9475	7321	5243	2675	864

附表 1.2 5fold 交叉驗證單一訓練回合中的句組段落數量(DB_{HB})

	DB_{HB}				
發文取樣比例	100%	75%	50%	25%	10%
正類別句組段落數	579	436	303	131	54
負類別句組段落數	12048	9020	6283	2767	1189
未標示句組段落數	7022	5321	3418	1936	733

附錄二 錯誤案例分析

本論文的主要任務是偵測發文是否具有心理健康面向，因此，本小節初步分析在開放測試資料集發現的錯誤案例。本小節將這些案例細分為不同類型，以深入討論模型容易誤判為正類別的情況。附表 2.1、附表 2.2 和附表 2.3 分別呈現三個目標面向資料集的錯誤案例分析。

附表 2.1 DB_{MD} 錯誤案例分析

案例分析	句組段落範例
提及精神疾病	“如果自己沒有持續調整心理狀態 **憂鬱症**也容易復發” “想請問一下 **思覺失調**是不是會導致記憶力退化”
非精神疾病的症狀	“我因為身體因素眼睛 很嚴重的**飛蚊症**” “因為我是**呼吸中止中度** 動眼期異常” “在網路上很逗趣的我好像**人格分裂**” “醫生只說是**更年期症狀** 有吃了兩周的中藥” “我身為一個邊**緣性人格障礙的患者**” “醫生跟我都覺得是**肌痛症**所以開這個”
提及情緒問題，未出現症狀名稱	“**焦慮**真的很痛苦 QQ 我的藥已經調好久” “我覺得很痛苦 我早就知道自己是**典型的焦慮依附**”
提及非精神疾病藥物	“距離上次頭痛還不到一個月 不得已又吃了一顆**普拿疼**” “請問吃這顆藥**大塚安立復**各位的副作用是什麼”

附表 2.1 呈現了「具有精神疾病」偵測模型容易錯誤預測為正類別的案例，主要源自兩大因素：一方面，受限於句組段落的文字內容有限，語意表達不夠完整；另一方面，負類別訓練資料中沒有或較少出現類似敘述。這兩大因素導致的錯誤案例又可歸納出四大類型，包括為「提及精神疾病」、「非精神疾病的症狀」、「提及情緒問題，未出現症狀名稱」、「提及非精神疾病藥物」。

在這四個類型中，「提及精神疾病」的案例可能是在陳述一個事實，例如“如果自己沒有持續調整心理狀態 憂鬱症也容易復發”，而非講述自身患有精神疾病。

然而，因為句組段落的文字內容有限，所能提供的語意資訊不充分，導致模型難以分辨。

其他三個類型的案例，包括「非精神疾病的症狀」、「提及情緒問題，未出現症狀名稱」、「提及非精神疾病藥物」，模型的誤判原因主要是負類別訓練資料中缺少類似的敘述。正類別訓練樣本中經常出現“症”這一字，或是負面情緒連同精神疾病一同出現在句組段落中，以及服用的精神藥物資訊，但相對在負類別訓練樣本中卻缺少非精神疾病症狀、非精神疾病用藥的資料，來告知模型應該將上述案例判斷為負類別。

附表 2.2 DB_{ED} 錯誤案例分析

案例分析	句組段落範例
有提到精神疾病，但沒有情緒問題	“而且台北有點冷 不知道為什麼好像有點**憂鬱症**” “而且因為**憂鬱症**的關係 有時候不得不中斷工作” “腦分泌失調**得到重度憂鬱** 我不知道或無法理解的事就是假的” “在這裡 你不孤單 我患有**恐慌症**”
情緒不佳程度未達專家認定水準	“我本身聲音不難聽只是**沒有自信**” “那我也沒什麼好過不去的 雖然**心裡還是過不去**” “當下常**覺得委屈** 但現在明白自己需要負很大責任” “完全**沒幹勁了** 現在的時間都在等待著” “體會到**過度自責**好耗盡我的精神”

附表 2.2 呈現了「具有情緒問題」偵測模型容易錯誤預測為正類別的案例，主要源自負類別訓練資料中沒有或較少出現類似敘述。可歸納出兩個類型，包括「有提到精神疾病，但沒有情緒問題」、「情緒不佳程度未達專家認定水準」。

對於「有提到精神疾病，但沒有情緒問題」這一類型，模型容易誤判的原因為，正類別訓練樣本中提及負面情緒時，經常連同精神疾病一同出現在句組段落中，導致模型偵測到精神疾病就容易預測為正類別。

對於「情緒不佳程度未達專家認定水準」，則是因為句組段落雖然呈現出情緒不佳的特徵，但其不佳的程度尚未達到專家認定的負面情緒水準，例如覺得委屈、沒幹勁了、過度自責。

附表 2.3 DB_{HB} 錯誤案例分析

案例分析	句組段落範例
語意不夠明確	“才發現社恐可以治療 也是有**心理諮商和吃藥**” “半夜又緊急**送醫** 那次我哭了一整個晚上”
非精神疾病就醫	“**眼科醫生**幫我撐開眼皮檢查眼睛” “高中**看了好多次腸胃科** 我都故意把要丟掉” “之前**看別科吃的藥** 醫生說副作用不會變胖但我還是變胖” “拖著我去看醫生 一開始是怕心臟問題**看了心臟血管科**” “**醫生**也**強調**很多次我**胃**不好很大的原因是因為容易緊張” “**醫生說**如果還會痛 就**得抽神經**” “**醫生說**我並沒有任何問題 有可能就是**胃食道逆流**”
提及藥物	“我四月中拿了九顆**普拿疼** 到現在已經吃了八顆了” “勉強喝了一瓶他喝的**亞培安素**和吃了一條小蛋糕”
尚未有具體求助行為	“無法控制的一直哭 終於**決定要找時間去**學校心輔室諮商” “憂鬱症困擾我好幾年了 **已經下定決心去**看醫生”

附表 2.3 呈現了「具有求助行為」偵測模型容易錯誤預測為正類別的案例，一方面，受限於句組段落的文字長度導致「語意不夠明確」；另一方面，負類別訓練資料中缺乏或較少出現類似敘述，又可分為三大類型，包括為「非精神疾病就醫」、「提及藥物」、「尚未有具體求助行為」。

「語意不夠明確」的案例像是“才發現社恐可以治療 也是有心理諮商和吃藥”，比較像是在陳述一件事實，而非自身尋求協助的行為；或是“半夜又緊急送醫”有

可能是他人將其送醫而非自發性。然而，由於句組段落的文字內容有限，所能提供的語意資訊不充分，導致模型難以分辨。

其他三個類型的案例，包括「非精神疾病就醫」、「提及藥物」、「尚未有具體求助行為」，模型的誤判原因主要是負類別訓練資料中缺少類似的敘述。正類別訓練樣本中經常出現醫生、就診和吃藥等詞彙，但相對在負類別訓練樣本中卻缺少非精神疾病就醫與非精神疾病用藥的類似敘述，以告知模型這樣的情境並非正類別的情形。