RoBERTa-based-clickbait Siamese Network for Clickbait Classification

111423040李泳輝,111423072姜道宣

1、實驗簡介

由於社群媒體發展成熟，人們每天都要接收大量的資訊。然而，人一天能接受的資訊量有限，因此我們會根據一篇文章的標題、圖片挑選有興趣的內容。為了從大量的資訊中吸引大眾的注意力，媒體常會利用聳動的標題吸引大眾的目光。這造成了當人們點擊文章後，往往會發現這些內容並沒有如標題描述的那麼精采甚至與標題毫不相干。因此，如何找出利用誇大不實的標題浪費人們注意力的文章，幫助閱聽人排除沒有價值的資訊，也越來越重要。目前大多數處理點擊詐騙的方法是依據標題進行辨識，但若能同時考慮標題和內文間的差異性，將能增加辨識的準確度。另外，我們可以對文章內文做摘要，減少模型訓練的複雜度。在本篇論文，我們提出以 roberta-base-clickbait 組成的孿生網路框架，對 webis-clickbait-17 資料中的標題和文章內文或摘要進行相似性的比較，進行點擊詐騙的辨識。

2、介紹

在網路與社群媒體發展蓬勃的時代，不論是商業或是個人用戶都會將新聞資訊轉發連結並附上自己對於新聞內文有意或無意的解讀與感想，然而單若觀看這些推文標題常造成對內文誤解的案例，也使多數人成為"標題黨"的受害者。我們認為在忙碌且資訊暴漲的時代，能以有效辦法減少誤判和並免於在誤導的文章花費時間與資源，也是在未來更需倍受重視的問題。

目前，大多數處理點擊詐騙的方法和資料集還是基於標題的辨識。(Anand et al., 2019) 建立基於 BILSTM 的框架，並嘗試使用 word2vec 和 CNN 對文章標題作嵌入。(Naeem et al., 2020) 提出一個以LSTM為基底的深度學習框架，針對點擊詐騙新聞標題的語言特徵建立模型。(Pujahari and Sisodia, 2021) 使用了一種混合分類的技術，考慮多個不同的特徵、句子結構和 t-SNE 進行分群來區分點擊詐騙和非點擊詐騙的文章。

(Dong et al., 2019) 使用了兩個 Bi-GRU 分別對新聞的標題和內文做嵌入，學習局部相似性和原始輸入的特徵，以注意力集中的方式進行預測。(Mintamanis and Mandala, 2022) 提出使用 ColBERT 的孿生網路，針對印度新聞的標題和內文間的相似性來判斷新聞是否是點擊詐騙，相較於使用 M-BERT 對新聞的標題做辨識並取得更加的結果。這些文章表示如果能同時考慮標題和內文之間的差異性，將能提高辨識的準確度。

為了處理更長的文字序列和上下文訊息，(Hartl and Kruschwitz, 2022) 使用 BERT 對文章的內文進行摘要，以減少模型訓練的複雜度。(Sepúlveda-Torres et al., 2021) 將 Fake News Challenge FNC-1 競賽的任務拆成兩個子任務，文中使用 TextRank 對新聞進行摘要，接著比較與新聞標題的相似性，減少模型需要處理的訊息量，並從摘要捕捉特徵，進行新聞標題與內文相關性的辨識任務。TextRank是基於co-occurence的計算相對重要性，並將文中每一個句子視為一個節點，依據計算出的TextRank值對所有結點排序，選擇TextRank值較高者作為文章摘要。雖然前述做法能夠有效找出文中最據代表性的文章片段，可是卻未能考慮到上下文對於本句含意造成的影響，也無法表現出排名最高句子結點外的文章內容。因此在本實驗中將以 BART、PEGASUS 模型進行文章摘要作為特徵，以比較考慮上下文產生的摘要是否能在辨識成果上有更良好的表現。

在此文章，我們提出一個以 roberta-base-clickbait 為基底的孿生網路，我們使用 TextRank、PEGASUS、BART 對新聞的內文做摘要，比較直接使用內文或摘要與標題進行相似度比較後，對於模型區分點擊詐騙的影響。我們的框架使用了兩個 Loss Function，分別是 consine embedding loss 和 cross entropy，以多任務學習（multi-task learning）的方式訓練模型。

3、相關工作

3-1. 內容摘要比較文章標題相似度 :

普遍 clickbait 的作法使用誇大或與事實不相符的文章標題吸引瀏覽者進入，以騙取點閱率。為了分辨標題與文章內容是否一致，在(Sepúlveda-Torres et al., 2021)中提到一種以摘要的方式判斷文章內容與標題的相關程度，並判斷內文與標題是一致、相牴觸、或是正反參半的討論方式。文中以 TextRank 進行文章摘要，後續將內文以及摘要作詞嵌入後相連接轉換為數值向量，以此進行訓練和分類。在 Transformer 出現後，在文本理解任務使用預訓練模型得到了很大的進展，但是使用 BERT 在文本生成(Seq2Seq)的應用中結果並不理想。在 (Lewis et al., 2019)BART(Bidirectional and Auto-Regressive Transformers)，其建立 Transformer model 之上的基礎之上並吸收了 BERT 的 biderection encoder 以及 GPT 的 left-to-right decoder 方法作為靈感，這使得 BART 更適合文本生成的應用，並可以在一些理解任務上取得 SOTA。Google 在(Zhang et al., 2020) 中提出 PEGASUS (Pre-training with Extracted Gap-sentences for Abstractive Summarization)，由於當時未有抽象文本摘要(abstractive text summarization)為目標的文本模型，其提出一種訓練方式 GSG(Gap Sentences Generation)搭配大量的訓練資料，將輸入文檔中的重要句子刪除或是遮蔽，再利用剩餘的文本輸出這些被遮蔽的句子。在實驗結果中，PEGASUS 刷新了12個資料集的 ROGUE 分數紀錄，並顯示在多個資料集上能以低於一千筆的樣本資料完成 fine-tune 並擁有良好的表現。

3-2. 文字詞嵌入 :

受到 BERT 的啟發，近年來有著相同概念的預訓練模型接踵而至，其中受到廣為使用的 RoBERTa(Robustly optimized BERT approach)(Liu et al., 2019)顧名思義，是由 Facebook提出

他們認為 BERT 在各方面表現出訓練不足，因此在訓練過程中加大訓練資料集和訓練時間、增大批次訓練集(batch size)、並使用動態產生的遮罩。RoBERTa 另有 base、large 等不同版本。Large 相較於 base 在輸出的張量、隱藏層、multi-attention heads 、參數量上皆有更大量的配置。

3-3. 孿生網路 :

孿生網路是使用兩個或多個子神經網路使用相同的架構、參數以及權重，藉由將兩筆不同資料同時輸入相同架構的網路轉換為相同維度的向量空間，並將兩向量進行相似度計算。過去研究(Li et al., 2022)顯示孿生網路較適合使用於目標擷取、圖像匹配、重新辨識、變化檢測、產品推薦等等問題。孿生網路藉由比較輸入資料得到辨識所需的資訊。而近期應用於點擊詐騙判斷標題與內文的相關性的研究(Mintamanis and Mandala, 2022)得到成果顯示，在文中實驗孿生網路能夠較M-BERT更有效的辨識詐騙點擊。其架構是將標題與新聞內容分別輸入ColBERT作詞嵌入後得到轉換的向量並計算兩者間的Cross-Entropy，藉以使詞嵌入的模型進行學習。

4、模型

4-1. 問題描述

在本文，要處理的主要任務是根據文章的標題和內容辨識這篇文章是否是點擊詐騙。該問題可以表示成給定一組標題 $H = \{h_1, h_2, \cdots, h_N\}$，以及這些標題的內容 $B = \{b_1, b_2, \cdots, b_N\}$，目標是根據這些資訊預測一個標籤 $Y = \{y_1, y_2, \cdots, y_N\}$，其中如果標題是網路詐騙，則$y_i$為1；否則為 0，其中N代表的是文章的總數。我們提出的架構可以分成三個部分 : 學習標題和文章的表示、學習標題和內文的相似程度、根據標題和文章的表示進一步預測標籤。我們的架構如圖 (1) 所示。

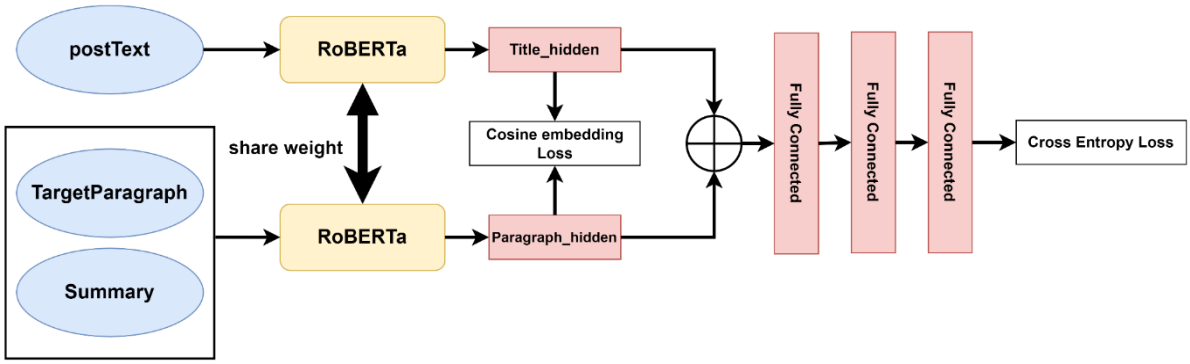

圖 (1)

4-2. 學習文章標題和內文的表示

在學習文章的標題和內文的表示之前，可以將處理內文的過程區分為有做摘要和沒做摘要。我們選擇使用TextRank、BART、PEGASUS 分別對文章內文做摘要，並將摘要表示為 $S = \{s_1, s_2, \cdots, s_N\}$，如圖 (2) 所示。

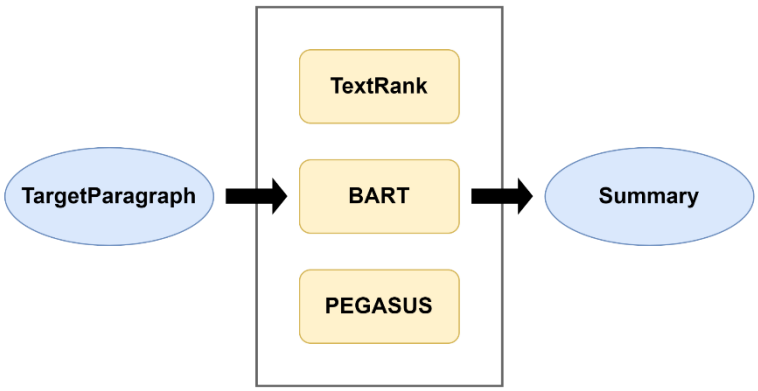

圖 (2)

接著，使用以roberta-base-clickbait為基底的孿生網路分別將標題 $H = \{h_1, h_2, \cdots, h_N\}$ 和內文 $B = \{b_1, b_2, \cdots, b_N\}$ 或摘要$S = \{s_1, s_2, \cdots, s_N\}$做嵌入，學習它們的隱性表示。標題經過 roberta-base-clickbait 處理後的嵌入表示為 $E_H = \{e_{h_1}, e_{h_2}, \cdots, e_{h_N}\}$，內文的嵌入為 $E_B = \{e_{b_1}, e_{b_2}, \cdots, e_{b_N}\}$，摘要的嵌入則是為 $E_S = \{e_{s_1}, e_{s_2}, \cdots, e_{s_N}\}$。roberta-base-clickbait 是在 Hugging Face 上一個用於區分點擊詐騙文章的 RoBERTa fine-tune 模型。它使用 Webis-Clickbait-17 數據集進行訓練，模型輸入為文章的 "postText"，並在測試集上 F1-score 達到約 0.7 的表現。

4-3. 學習標題和內文的相似程度

我們使用一層的全連接網路 (1) 分別對標題的嵌入 $E_H = \{e_{h_1}, e_{h_2}, \cdots, e_{h_N}\}$ 和 內文的嵌入 $E_B = \{e_{b_1}, e_{b_2}, \cdots, e_{b_N}\}$ 或摘要的嵌入$E_S = \{e_{s_1}, e_{s_2}, \cdots, e_{s_N}\}$進行降維，其中$A$是全連接網路需要學習的權重，b則是要學習的偏移量。接著，計算經全連接網路降維後的標題嵌入與內文嵌入或摘要嵌入的 consine embedding loss (2)。如果 y_i 是 點擊詐騙，那標題與內文或摘要應該距離較遠；如果y_i 是 非點擊詐騙，那標題與內文或摘要應該距離較近。

$$l_k = e_k A^T + b, k \in \{h, b, s\} \quad (1)$$

$$cosineembeddingloss\left(l_{h_i}, l_{m_i}, y_i\right) = \begin{cases} 1 - \cos\left(l_{h_i}, l_{m_i}\right), \text{if } y_i = 0 \\ \max\left(0, \cos\left(l_{h_i}, l_{m_i}\right) - margin\right), \text{if } y_i = 1 \end{cases}, m \in \{b, s\} \quad (2)$$

4-4. 預測標籤

我們將標題的嵌入 $E_H = \{e_{h_1}, e_{h_2}, \cdots, e_{h_N}\}$ 和 內文的嵌入 $E_B = \{e_{b_1}, e_{b_2}, \cdots, e_{b_N}\}$ 或 摘要的嵌入$E_S = \{e_{s_1}, e_{s_2}, \cdots, e_{s_N}\}$合併(3)起來，在交給以三層的全連接層的分類器做處理，其中B、C、D是需要全連接網路學習的權重，b_1、b_2、b_3則是要學習的偏移量。

$$e_t = e_{h_t} \oplus e_{m_t}, m \in \{b, s\} \quad (3)$$

$$c_t = LeakyReLU(e_t B^T + b_1)$$

$$\text{dt} = LeakyReLU(c_t C^T + b_2) \quad (4)$$

$$\widetilde{y_t} = d_t D^T + b_3$$

4-5. 損失函數

CrossEntropy Loss 和 Cosine Embbeding Loss 是此次框架使用到的兩個損失函數。Consine Embedding Loss 目的是學習標題和內文或摘要的相似程度，而 CrossEntropy Loss (5)目的是比較預測結果與實際標籤間的差別，而這個損失函數的值都是越小越好。另外，因為這個資料集有嚴重的類別偏移，因此我們加入損失權重到 cross-entropy中，其中權重的設定為 [$\frac{negative\_class\_\#}{negative\_class\_\#}$, $\frac{negative\_class\_\#}{positive\_class\_\#}$]。最後，我們將這兩個損失函數(6)合併起來，作為訓練的損失函數。

$$\text{CrossEntropyLoss} = -\sum_i w * \widetilde{y_i} * \log(y_i) + (1 - \widetilde{y_i}) * \log(1 - y_i)) \quad (5)$$

$$Loss = cosineembeddingloss + CrossEntropyLoss \quad (6)$$

5、實驗設計

本次的實驗，我們將只用標題當作 roberta-base-clickbait 輸入，當作本次模型的 baseline，以比較同時考慮標題與內文對點擊詐騙的影響。另外，文章內文會經過 TextRank、BART、PEGASUS 處理，將擷取出摘要當作模型的的輸入。我們會比較內文有做摘要對模型的影響，以及比較 TextRank、BART、PEGASUS 這些摘要方法哪個比較適合用於我們的框架。模型表現是使用四個常見的評估指標，分別是 accuracy、recall、precision 和 F1-score。

5-1. 資料集

本次實驗資料採用發布於 Webis-Clickbait-17上的詐騙點擊混和資料集，包含了來自27家美國主要新聞，文體總共38,517則 Twitter 標題，以及標題中連結的文章的相關信息。這些推文發布於2016年11月至2017年6月期間。為了避免出版媒體基於主題的偏見，每天和每個媒體最多只抽樣了十個推文。標題的標註方式來自於五位來自 Amazon Mechanical Turk 的標註者以四點評分標準進行標註以其程度以0至1的區間分為四個等級:非點擊誘餌（0.0），輕微點擊誘餌（0.33），相當程度的點擊誘餌（0.66），非常強烈的點擊誘餌（1.0）。根據大多數標註者的認定，共有9,276則推文標題被認為是點擊誘餌，約資料集總數的四分之一。為了能夠客觀評估點擊誘餌檢測系統，目前測試集可通過 Evaluation-as-a-Service平台 TIRA 獲取。

實驗是使用資料集的 ‘postText' 和 ‘targetParagraphs’ 欄位，當作模型的輸入，並且以資料集的 'truthClass' 為最終預測的標籤。我們的模型是用Webis-Clickbait-17 中的 webis_train.csv 作為訓練集，並隨機將其中的 20% 當作驗證集，剩下的 80% 為訓練集，

且分割過程中會確保驗證集和訓練集中點擊詐騙的資料比例是相同的，並比較模型在 webis_test.csv 的表現。

5-2. 參數設定

5-2-1. TextRank文章摘要:

本實驗中使用SUMMA套件以TextRank方式計算個句子節點產生文本摘要，詳細輸入參數如下: "text" 輸入文本為刪除特殊符號的文章內容製 targetParagraphs_no_label 欄位，"ratio" 為文章摘要對比原文長度的比例，採預設值0.2，"words" 為限制摘要字數。因部分文章句子字數較長採用預設值None不限制摘要句子字數，"splitter"指定句子分割方式，採用默認值"."句點符號分割文本句子。

5-2-2.PEGASUS文章摘要

模型使用 transformer 套件，使用 PegasusTokenizer 作為文本詞嵌入轉換器，模型建立使用 PegasusForConditionalGenerationalGenenration，並分別導入google/pagesus-xsum 作為 pre-trained 版本。tokenizer 的 input 參數如下: "text" 為輸入文本，使用 targetParagraphs_no_label 欄位資料，"truncation” 為 True，當輸出超過最大長度默認值512時截斷處理，"padding"採用"longest"使每個輸出與最長的序列長度一致，"return_tensors"使用"pt"輸出為pytorch張量。model 輸入 input參數如下 : ” input_ids” 為 tokenzier 中輸出的標記序列，"attention_mask" 使用 tokenizer 回傳數值中的 attention_mask 作為輸入，” temperature” 採用默認值1.0控制產生文本的多樣性保持在較為保守的狀態，"early_stopping" 設定為 False，確保模型在產生完整文本後結束。 decoder 的input參數如下 : "sequences" 輸入為 mode l的輸出，"skip_sepcial_tokens"設定為 True表示不保留CLS、SEP、PAD等特殊記號。

5-2-3. BART文章摘要

使用 simpletransformer 套件中 seq2seq model，輸入 input 如下所示:"necoder_decoder_type" 設定為 bart，” encoder_decoder_name” 設定為"facebook/bart-large" 使用多參數與大量訓練資料集的摘要版本，” num_trina_epochs” 設定為 10，"evaluate_generated_text" 設定為Tru e，在訓練時顯示準確度、時間等資訊。

5-2-4.模型

模型的輸入會先經過 roberta-base-clickbait 處理，裡面的參數是使用模型的預設值。接著，使用一個維度為 256 的全連接網路，將標題和內文或摘要表示降維，再比較降維後的標題和內文或摘要表示的 Cosine Embbeding Loss。另外，標題和內文或摘要表示會合併，再交給由三層全連接網路組成的分類器做處理，其維度分別為 768, 256, 2。最後，資料的 batch_size 為16，優化器為 Adam，模型總共會訓練15回，取準確度最高的模型當作最後的結果。

6、實驗結果討論

表格 1 為實驗的結果，圖 (2) ~ 圖 (7) 是這些方法的混淆矩陣。根據表格一，可以觀察到只用 postText 作為模型輸入，在 recall 和 F1-score 的表現是最好的，而且在 accuracy 和 precision 上的結果與其他模型差不多。我們推論造成這樣現象的原因有幾個可能 : 首先，可能是 postText 這個特徵就足以完成點擊詐騙的任務，而增加內文對於這個任務的影響並不大。第二可能是 RoBERTa 並不適合用來擷取標題和內文的特徵或是應該對 RoBERTa 表示適用其他的模型，像是 LSTM、CNN ，進一步的擷取特徵。

當我們直接使用內文當作模型的另一個輸入時，可以看到模型結果的 accuracy 和 precision 都比只用 postText 當作輸入時好，而以TextRank、PEGASUS、BART對內文做摘要的結果在 precision 上也有相同的現象，但 BART 在 accuracy 上表現就略比其他他模型來的差。整體而言，TextRank、PEGASUS、BART這些摘要方法中，TextRank是表現最好的，它只有 recall 的結果比其他兩個摘要方法差。雖然 TextRank 在其他指標上只有略勝於用內文當輸入的結果，但其只有在 accuracy 的表現比用內文來的差。因此，我們推論這三種摘要方法，用 TextRank 擷取內文特徵會是比較好的方式，而用摘要來代表內文可能不是一個最恰當的選項。

Model	Accuracy	Recall	Precision	F1-score
postText_only	0.858	0.799	0.672	0.730
Paragraph	0.863	0.757	0.694	0.724
TextRank	0.859	0.758	0.696	0.726
PEGASUS	0.858	0.764	0.678	0.719
BART	0.857	0.773	0.674	0.720

表格 1 實驗結果

所有模型的對於點擊詐騙的 recall 都沒超過 0.8，這顯示出我們的模型在區分點擊詐騙還有很大的進步空間。造成此結果的主因我們推測與資料類別偏移有關，在本次實驗的資料集中點擊詐騙類別的數量與總資料集佔比約1:3，此訓練集結構造就少數類別在判別上的劣勢。依此我們推論孿生網路在訓練資料集類別比例偏移的情況下，並無法有效維持少數類別在判別時的準確率。

根據實驗的結果，若是提供一般用戶瀏覽社群網路時用來檢測點擊詐騙的 plug-in 工具，可以選擇只考慮 postText 的模型，也就是這次實驗的 baseline，因為它有最高的 recall ，能夠幫助使用者盡可能的去除可能為點擊詐騙的文章。如果是像 Facebook 或 Twitter 這類的社群平台想要過濾點擊詐騙的文章，但又想盡可能避免因審核太過嚴格導致投放廣告者或創作者的不滿，此時可以選擇同時考慮以TextRank得到的摘要或直接用文章內文和 postText 的模型，因為它們分別在實驗中有最高的和次高的 precision。

7、消融實驗

為了驗證孿生網路是否是有效的，我們使用了兩個不同的 RoBERTa 去擷取 postText 和內文的特徵，結果如表二所示。當模型以非孿生網路的架構進行訓練時，模型整體的表現都有明顯的下降。這說明模型的輸入為 postText 和內文時，我們提出的孿生網路架構是有效的。

另外，為了檢驗使用 Cosine Embedding Loss 和 Cross Entropy Loss 的多任務學習是否能幫助我們提出的框架學習。我們以表格 2 'postText' 和內文做為輸入並只做 Cross Entropy Loss 來訓練的模型當作比較參考。根據 的結果，可以看到如果只用 Cross Entropy Loss進行訓練，模型的 recall 雖然有提升，但其他的三個指標的表現都下降了。因此，可以推論多任務學習對於我們提出的框架是有幫助的。

Model	Accuracy	Recall	Precision	F1-score
Paragraph	0.863	0.757	0.694	0.724
no_siamese_P	0.830	0.648	0.641	0.645
one_loss_p	0.855	0.788	0.665	0.722

表格 2 消融實驗

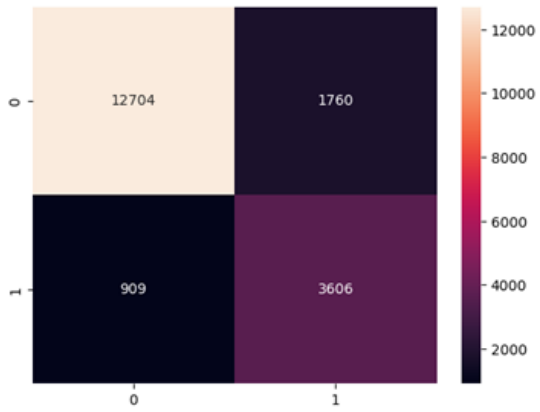

圖 (3) postText-only

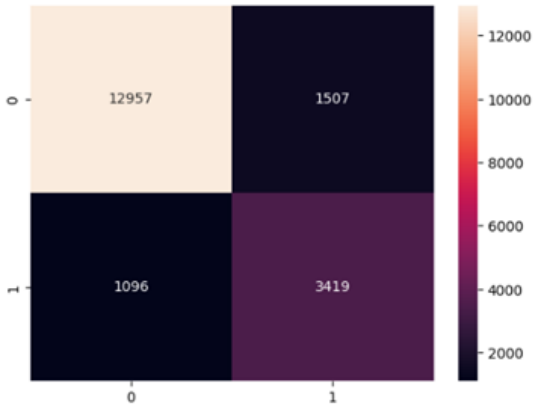

圖 (4) paragraph

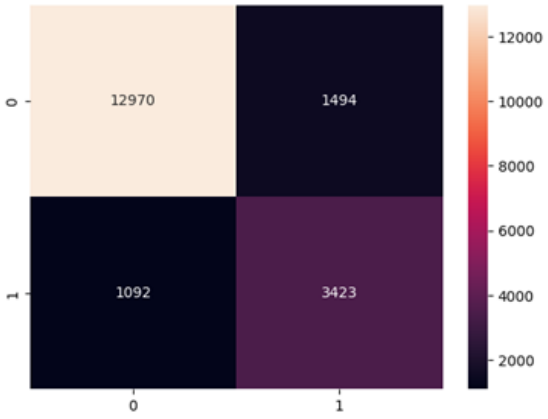

圖 (5) TextRank

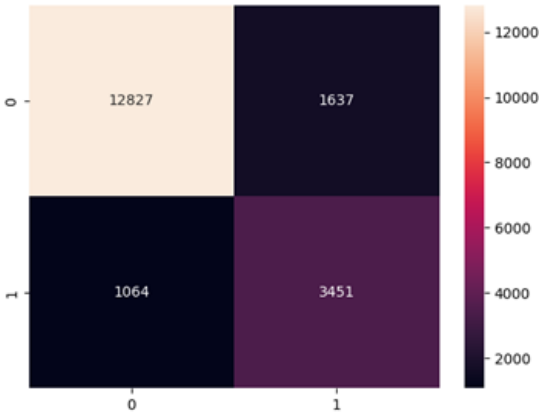

圖 (6) PEGASUS

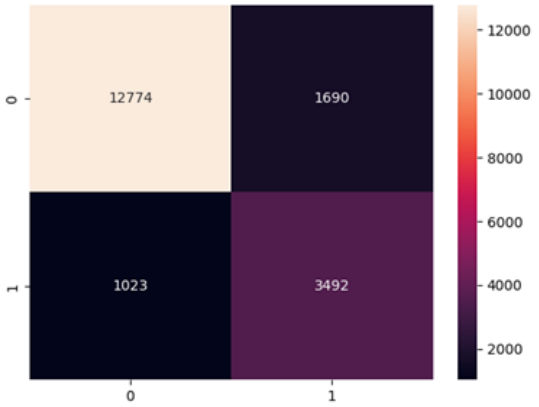

圖 (7) BART

8、結論與展望

本次實驗我們提出了一個用於判斷推文標題是否為以誇大、不實的方式增加新聞點擊率，並使用對原文內容產生誤解或是無關的聳動文字。此模型以孿生網路為基礎，分別比較標題與內文、標題與摘要經過詞嵌入轉換為數值向量的相似度。與原本預期相異的是，使用標題與內文組合進行辨識的準確度最高，在我們的想法中原文應包含了最完整得內容足以做出正確的判斷，但未能反映在實驗成果上，這可能是與詞嵌入和資料特徵處理有關。因此未來應該嘗試使用 LSTM、GRU 等特徵萃取或是 RoBERTa以外的詞嵌入方式，嘗試有效地捕捉內文特徵。

此外，我們藉由比較有無使用孿生網路的實驗中得到驗證，孿生網路在比較推文標題和內文的相關性中能較其他模型表現的更加有效，且在 recall 方面受到資料組成結構不平衡的影響也較小。最後，在驗證 Cosine Embedding Loss 和 Cross Entropy Loss 的多任務學習是否能有助於本次提出框架的學習實驗中，使用兩種 loss學習的方式較僅使用

Cross Entropy Loss更能維持多數指標的水準，因此我們認為多任務學習有助於達成本次目標。

考慮到本次實驗中 clickbait 辨識成果的 recall 尚有進步空間，資料集組成結構的平衡應可以為辨識成果帶來顯著提升。為提升點擊詐騙的辨識率，未來可嘗試對少數類別做 text augmentation 或使用孿生網路的變形版本Triple Network。Triple Network是以兩個正例一個反例或是兩個反例一個正例所組成架構，使相似性計算達到同類別差異最小，異類別差異最達的差距盡可能大，達成判別上區隔的最大化。

9、參考資料

Anand, A., Chakraborty, T., Park, N., 2019. We used Neural Networks to Detect Clickbaits: You won’ t believe what happened Next!

Dong, M., Yao, L., Wang, X., Benatallah, B., Huang, C., 2019. Similarity-Aware Deep Attentive Model for Clickbait Detection, in: Advances in Knowledge Discovery and Data Mining: 23rd Pacific-Asia Conference, PAKDD 2019, Macau, China, April 14-17, 2019, Proceedings, Part II. Springer-Verlag, Berlin, Heidelberg, pp. 56 – 69. https://doi.org/10.1007/978-3-030-16145-3_5

Hartl, P., Kruschwitz, U., 2022. Applying Automatic Text Summarization for Fake News Detection.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L., 2019. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. https://doi.org/10.48550/arXiv.1910.13461

Li, Y., Chen, C., Zhang, T., 2022. A Survey on Siamese Network: Methodologies, Applications and Opportunities. IEEE Trans. Artif. Intell. PP, 1 – 21. https://doi.org/10.1109/TAI.2022.3207112

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. https://doi.org/10.48550/arXiv.1907.11692

Mintamanis, J.C., Mandala, R., 2022. Clickbait Indonesian News Classification using ColBERT with Siamese Network on Headline and Content News, in: 2022 9th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA). Presented at the 2022 9th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), pp. 1 – 6. https://doi.org/10.1109/ICAICTA56449.2022.9933005

Naeem, B., Khan, A., Beg, M., Mujtaba, H., 2020. A deep learning framework for clickbait detection on social area network using natural language cues. J. Comput. Soc. Sci. 3. https://doi.org/10.1007/s42001-020-00063-y

Pujahari, A., Sisodia, D.S., 2021. Clickbait detection using multiple categorisation techniques. J. Inf. Sci. 47, 118 – 128. https://doi.org/10.1177/0165551519871822

Sepúlveda-Torres, R., Vicente, M., Saquete, E., Lloret, E., Palomar, M., 2021. Exploring Summarization to Enhance Headline Stance Detection, in: Natural Language Processing and Information Systems: 26th International Conference on Applications of Natural Language to Information Systems, NLDB 2021, Saarbrücken, Germany, June 23 – 25, 2021, Proceedings. S

pringer-Verlag, Berlin, Heidelberg, pp. 243 – 254. https://doi.org/10.1007/978-3-030-80599-9_22
Zhang, J., Zhao, Y., Saleh, M., Liu, P.J., 2020. PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. https://doi.org/10.48550/arXiv.1912.08777