

運用超圖注意力網路於中文幽默文本多標籤分類

Multi-Label Classification of Chinese Humor Texts Using Hypergraph Attention Networks

高浩銓¹ Hao-Chuan Kao¹, 洪滿珍¹ Man-Chen Hung¹,
李龍豪¹ Lung-Hao Lee¹, 曾元顯² Yuen-Hsien Tseng²

¹ 國立中央大學 電機工程學系

² 國立臺灣師範大學 圖書資訊學研究所

¹Department of Electrical Engineering, National Central University

²Graduate Institute of Library & Information Studies, National Taiwan Normal University
{108521098, 109521068}@ncu.edu.tw, lhlee@ee.ncu.edu.tw, samtseng@ntnu.edu.tw

摘要

我們運用超圖注意力網路 (HyperGAT) 模型來辨識中文幽默文本的多重類別，文本以超圖表示，並用順序超邊和語意超邊構建超邊結構，接著使用注意力機制取得語境資訊，最後預測文本的多標籤分類。我們在中文幽默多標籤資料集比較模型效能，無論哪個評測指標，HyperGAT 模型皆優於其他基於序列 (CNN, BiLSTM, FastText) 和基於圖形 (Graph-CNN, TextGCN, Text Level GNN) 的深度學習模型。

Abstract

We use Hypergraph Attention Networks (HyperGAT) to recognize multiple labels of Chinese humor texts. We firstly represent a joke as a hypergraph. The sequential hyperedge and semantic hyperedge structures are used to construct hyperedges. Then, attention mechanisms are adopted to aggregate context information embedded in nodes and hyperedges. Finally, we use trained HyperGAT to complete the multi-label classification task. Experimental results on the Chinese humor multi-label dataset showed that HyperGAT model outperforms previous sequence-based (CNN, BiLSTM, FastText) and graph-based (Graph-CNN, TextGCN, Text Level GNN) deep learning models.

關鍵字：超圖神經網路、幽默辨識、多重分類
Keywords: hypergraph neural networks, humor recognition, multi-label classification.

1 介紹

幽默是指令人感到好笑、高興或滑稽的言行舉止，有助於幫助化解敵意、緩和摩擦與安慰他人，也是人際交往的重要溝通的元素。在商業應用中，幽默的溝通可以消除用戶的抱怨 (Bellegarda, 2014; Binsted, 1995)。在教育方面，Bryant & Zillmann (1989) 和 McGhee & Frank (2014) 研究發現課堂上適當的使用幽默，可以吸引學生的注意力並改善課堂互動，幫助學生在樂趣中能夠有更好的學習成效。在高壓的環境下，像是在公共場合發表演講時，幽默可以幫助演講者減輕焦慮並提升他們的表現。除此之外，廣告、娛樂等商業領域也是幽默應用的領域。

幽默的內容可能包含挑戰社會規範或禁忌、攻擊或嘲諷人或事物，又或是用來安慰他人的元素。這些內容可以引發不同的情感 (例如：釋放自我約束，或是增加個人優越感)，從而化解內部壓力並產生愉悅感。隨著科技進步，人類的對話系統介面或聊天機器人變得更廣泛使用，將幽默引入人機交流之中變得非常重要。因此，我們運用機器學習來辨識人類產生幽默的話語。

幽默辨識 (humor recognition) 本質上是一個多標籤分類問題，因為一則笑話可能屬於多個意向或動機，也可能屬於多個建構類別。因此，我們運用 Kaize et al. (2020) 所提出的超圖注意力神經網路模型 (Hypergraph Attention Network, HyperGAT) 來辨識人類幽默文本的多重類別。藉由超圖注意力網路找出字詞與

句子的關聯，並用超邊聚合結構與邊層級注意力，建構超邊與獲得上下文信息，用以預測文本的多標籤結果。在中文幽默多標籤資料集 (Tseng et al., 2020)，我們採用的 HyperGAT 模型可以達到 Macro F1-score 0.2419、Micro F1-score 0.4695、Weighted F1-score 0.4084 與 Subset Accuracy 0.1215，整體效能皆優於其他相關研究模型。

本文章節如下，第二章探討相關研究，第三章敘述我們使用的超圖神經網路模型架構，第四章為模型效能評估，最後是結論。

2 相關研究

2.1 中文幽默種類

在創造笑話時會運用一些幽默的技巧，Chen et al. (2017) 研究發現大多數建構笑話的方法類別與心理學上的不協調有關。建構笑話的方法分為八種，分別是雙重含義、誇飾法、擬人化、聯繫性推論、缺乏邏輯、諷刺、模仿和其他技巧。說明如下：

- (1) 雙重含義：雙重意味著不只一種解釋文本的方法，其中一種是隱藏在普遍能意會到的解釋方式，而一旦發現另一個解釋方式，就會產生一種幽默的感覺。雙重含義有同音異義詞、雙關語、語義學、語法和短語等類型。例如：女員工：「老闆，您必須幫我加薪，已經有三家公司在找我了！」老闆：「哪三家？」員工：「自來水公司，台電，天然氣公司。」
- (2) 誇飾法：是指將情境或描述的層次最大化，以打動人們或表達創造力。例如：某醫生在家接到同事電話：「打麻將，三缺一！」醫生說：「我馬上來！」妻子在旁邊問：「情況嚴重嗎？」醫生嚴肅地說：「很嚴重，已有三位醫生在那了。」
- (3) 擬人化：是指將人類的特徵套用在非人事或事件上的解釋。例如：「0 認為自己是一個優雅的人。當她遇見 8 時，她批評 8 是一個戴腰帶的假胖子。」
- (4) 聯繫性推論：隱含著不同角度的關聯，以有趣的方式用喂給讀者/聽眾帶來驚喜。例如：妻子：「你很少在外面喝酒，你為什麼在家喝很多酒？」丈夫：「有人告訴我酒精讓我勇敢。(沒有酒精而要面對妻子太可怕了。)」

(5) 不合邏輯：在錯誤的情況下，使用合乎邏輯的方式，來嘲笑一個人的愚蠢或行為。例如：「我的妻子總是鼓勵我盡力而為，所以我盡我所能使她放棄讓我做所有的家務。」

(6) 諷刺：描述與預期相反的消極/積極的情況。例如：「我窮得只剩下錢。」

(7) 模仿：按照設置情景的邏輯來做一個相似的情景。例如：「一個好的配偶是你在暴風雨中休息的港灣；糟糕的配偶是港口的暴風。」

(8) 其他技巧：不能歸入以上七類的，像是俗諺或語錄。例如：「什麼手術可以把眼睛變成耳朵？(讀唇技術)」

另外，根據 Chen et al. (2017) 所提出的觀點，笑話的意圖可以分為六類，分別為親和力、自我提升、攻擊、自我抑鬱、禁忌和其他動機。介紹如下：

- (1) 親和力：扭轉局面，用友好和善的話語擺脫尷尬，使別人感到安慰，或是說一些有趣和輕鬆的事情，讓彼此更親近或緩解團隊中的衝突。例如：「大家都希望有個和平的世界，但我只想要你的世界。」
- (2) 自我提升：接受糟糕的情況並改變觀點或自我鼓勵，讓自己打起精神來面對問題，這是一種幽默的應對方式。例如：「如果長得好看是一種罪，那我就是有罪。」
- (3) 攻擊：透過嘲笑別人的缺點讓自己開心，說別人的缺點或刁難的話，或是讓別人感到不舒服使其降低自己在群體的地位。例如：「我已經等了一個小時的餐點了，廚師是睡著了嗎？」
- (4) 自我抑鬱：用自嘲的方式來取悅他人，例如：「我無意犯下天生醜陋的罪刑。」
- (5) 禁忌：嘲笑與性、死亡、排泄物、被禁止的行為或思想有關的事物。例如：「18 歲少女 20 年前離奇失蹤。」
- (6) 其他意圖：不能歸入上述五類的。

2.2 多標籤文本分類

多標籤文本分類 (Multi-Label Text Classification) 是一種多重分類，各文本可能同時屬於幾個預先定義的標籤。不同於多元文本分類任務 (Multi-Class Text Classification) 標籤是互斥的，多標籤分類的每個標籤可以來自不同的分類任務，而標籤在某種程度上是相關的。

近年來，各種基於神經網路的模型被用於多標籤文本分類任務中。卷積神經網路 (Convolutional Neural Network, CNN) 最早使用於圖像分類，運用卷積濾波器 (filters) 提取圖片特徵，可以同時將不同核 (kernel) 定義的卷積應用於序列的多個區塊，大大提升了影像辨識的表現。CNN 應用於文字上是將輸入文本的詞向量拼接成矩陣，再將矩陣送入卷積層 (convolution layer)，該層包含多個不同維度的濾波器。最後，卷積層的輸出經過池化層 (pooling layer) 將池化結果拼接起來，得到文本的最終向量表示，並預測文本歸屬的標籤。Yoon Kim (2014) 所提出的 TextCNN 模型使用無偏差值 (bias) 的 CNN，它可以通過一層卷積，更好地決定最大池化層中有區別性的詞語，並將詞向量保持靜態以學習除了詞向量之外的超參數。

長短期記憶 (Long Short Term Memory, LSTM) 模型加入了遺忘閘、更新閘、輸出閘三個控制閘來強化記憶的儲存與使用，提升了其在長期記憶中的表現。但是利用 LSTM 對句子進行模型的運算，無法編碼從後到前的語意資訊，雙向長短期記憶 (Bidirectional LSTM, BiLSTM) 模型，便是由前向 LSTM 與後向 LSTM 組合而成，通過前後的結合可以捕獲更好的雙向語義依賴。

FastText (Joulin et al., 2016) 是一個開源資料庫，由 Facebook 人工智能研究實驗室開發，它的目標是簡單且快速地完成文本分類任務，並且有詞向量的訓練生成。它的結構類似於 Word2Vec 中的 CBOW 架構，通過一個全連接層將句子特徵映射到向量空間，再直接對詞向量進行平均進行預測。模型當中使用 Huffman 演算法建立用於表徵類別的樹形結構以加速運算，並用 n-gram 的特徵加強句子的表達。

Graph-CNN (Defferrard et al., 2016) 利用柴比雪夫多項式 (Chebyshev polynomials) 來近似擬合卷積核，來解決起初基於拉普拉斯 (Laplace) 及傅立葉 (Fourier) 的圖形卷積網路 (Graph Convolutional Network, GCN) 的缺點。

TextGCN (Yao et al., 2019) 是使用圖卷積神經網路 (GCN) 為整個資料集建構了一個基於文本和詞的異質 (heterogeneous) 圖，可以用來取

得全局詞的共現信息，使 GCN 能夠對文本進行半監督分類。

Text Level GCN (Huang et al., 2019) 為每個輸入的文本建構獨立，但具有全局參數共享的圖，而不是為整個訓練、測試語料庫建立一個巨大的單圖，並透過滑動窗口 (sliding-window) 來構建圖形，當中可以設定 n 元語法 (n-gram) 的數量，用以提取更多的局部特徵，並減少大量的計算資源，這也是使圖神經網路能夠從已有的資料中歸納出模式，並應用於新的資料與任務。

總而言之，基於序列的方法 (sequence-based methods) 有 CNN、LSTM、BiLSTM 和 FastText。而基於圖形的方法 (graph-based methods) 有 Graph-CNN, TextGCN 和 Text-Level-GCN。我們採用的超圖注意力網路，也是一種圖神經網路模型。

3 實驗方法

我們使用的模型架構是超圖注意力網路 (Hypergraph Attention Networks, HyperGAT) (Kaize et al., 2020)，在此章節之中，我們會介紹如何將文本以超圖表示，並用兩種方式構建超邊結構，接著使用注意力機制在超邊與節點上，捕獲文本上下文信息，最後預測文本的多標籤分類結果。

3.1 圖表示

超圖 (hypergraph) 是一種圖的結構，可以將其定義為 $G = (N, E)$ ， $N = \{n_1, n_2, \dots, n_v\}$ 為超圖中的節點 (node)， $E = \{e_1, e_2, \dots, e_m\}$ 為超圖中的超邊 (hyperedge)，其中對於每個超邊 e ，可以連結兩個或兩個以上的節點 n ，即 $\sigma(e) \geq 2$ 。超圖 G 中的拓撲結構 (topological structure) 亦可以表示為關聯矩陣 (Incidence matrix) $A \in R^{v \times m}$ ，其中 v 是節點數量， m 是超邊數量，定義如下：

$$A_{ij} = \begin{cases} 1, & \text{if } n_i \in e_j \\ 0, & \text{if } n_i \notin e_j \end{cases} \quad (1)$$

我們將實驗資料中每一篇文本，建立成獨立的超圖結構，為了考慮超圖結構中每個單一節點的屬性，把文本中的單詞 (token) 作為超圖中的節點後，需要先初始化節點的屬性。

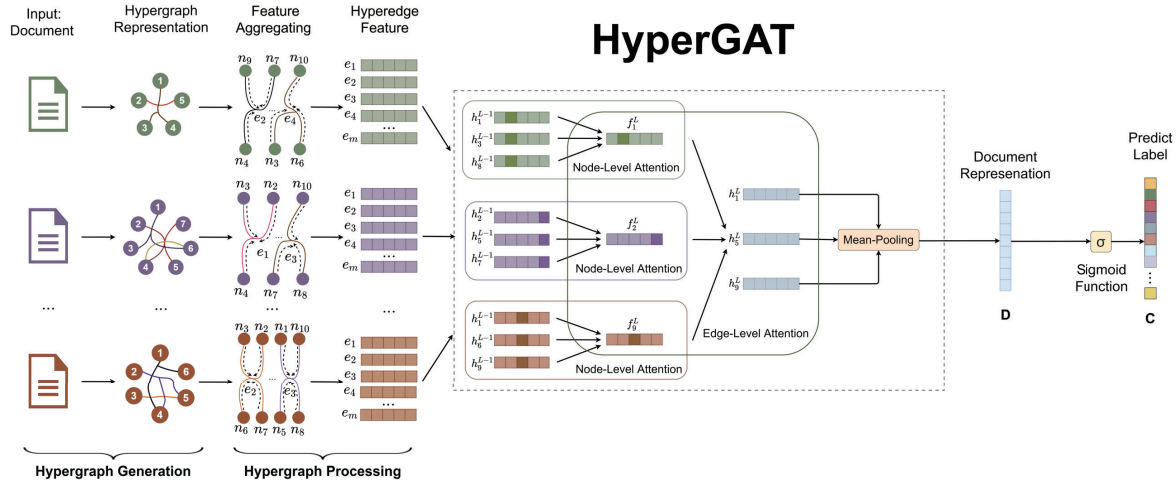


圖 1: 超圖注意力網路 (HyperGAT) 模型架構

將每個節點定義成維度 d 的向量，並以矩陣 $X = [x_1, x_2, \dots, x_n]^T \in R^{n \times d}$ 來表示所有的節點屬性，其中 n 是節點數量， d 是節點屬性之維度，並以 $G = (A, X)$ 表示整個超圖結構。

3.2 超邊聚合結構

為了捕獲文本之中，高階全局之上下文信息，我們使用兩種多關係超邊，分別是順序超邊跟語意超邊。

3.2.1 順序超邊 (sequential hyperedge)

文本中的語句是有順序性的，透過上下文我們可以知道文字之間，局部共同出現的語言特徵。為了利用每個詞的順序信息，我們需要為資料中的每個文本建構順序超邊：將每個文本中的句子，使用標點符號分隔，並將其視為超邊，此超邊結構稱為順序超邊，它連接了句子中的所有詞，如此一來，能使模型捕獲整個文本每個句子的結構信息。

3.2.2 語意超邊 (semantic hyperedge)

除了文本的順序信息，我們也希望能找到詞的語意信息。在本模型中，我們使用隱含狄利克雷分布 (Latent Dirichlet allocation, LDA) (Blei et al., 2003) 去建構語意超邊，將文本文檔中的潛在主題 T 找出之後，每個主題 $t_i = (\theta_1, \theta_2, \dots, \theta_w)$ 可以表示成單詞的機率分佈， w 為詞彙量的大小。接著我們將每個主題中機率最大的前 K 個單詞連接成語意超邊，通

過這些與主題相關的超邊，取得單詞與主題的相關性，能夠豐富每個詞彙的語意訊息。

3.3 超圖注意力網路層

為了有效的聚合文本超圖中的超邊與節點，我們使用超圖注意力網路層訓練模型。本層使用兩種不同的注意力聚合方法，分別是節點層級注意力 (node-level attention) 以及邊層級注意力 (edge-level attention)。 $AGGR_{node}^l$ 是節點聚合函數，將節點的特徵聚合到超邊； $AGGR_{edge}^l$ 是另一個聚合函數，將超邊的特徵聚合到節點。兩種注意力聚合方式的定義如方程式(2)和(3)：

$$f_j^l = AGGR_{node}^l(\{h_k^{l-1} | \forall n_k \in e_j\}) \quad (2)$$

$$h_i^l = AGGR_{edge}^l(h_i^{l-1}, \{f_j^l | \forall e_j \in E_i\}) \quad (3)$$

其中， E_i 為連接到節點 n_i 的超邊集合、 f_j^l 是超邊 e_j 在 l 層中的表示、 h_i^l 是節點 i 在 l 層的節點表示，一般將初始化向量 x_i 作為首層 h_i^0 之節點特徵。方程式(2)與(3)的細項分別於 3.3.1 小節與 3.3.2 小節做介紹。

3.3.1 節點層級注意力

我們將特定的節點標示為 n_i 、超邊標示為 e_j ，並將所有的超邊集合表示為 $e_j \in E_i$ ，由於每個超邊中的節點對超邊的重要程度不盡相同，我們使用注意力機制來強調那些對超邊意義較為重要的節點，將其聚合 (aggregate) 之後計算超邊表示 f_j^l ：

$$f_j^i = \sigma \left(\sum_{n_k \in e_j} \alpha_{jk} W_1 h_k^{l-1} \right) \quad (4)$$

其中 σ 是非線性函數 ReLU、 W_1 是可以訓練的權重矩陣、 α_{jk} 為節點 n_k 與超邊 e_j 的注意力參數，並用以下方式去做計算：

$$\alpha_{jk} = \frac{\exp(a_1^T u_k)}{\sum_{n_p \in e_j} \exp(a_1^T u_p)} \quad (5)$$

$$u_k = \text{LeakyReLU}(W_1 h_k^{l-1}) \quad (6)$$

其中 a_1^T 是權重向量，亦可以說是上下文向量。注意力當中使用的激勵函數為 LeakyReLU。

3.3.2 邊層級注意力

將節點信息傳遞至超邊後，為了強調那些較具有資訊的超邊，我們將所有超邊 $\{f_j^l | \forall e_j \in E_i\}$ 應用邊層級的注意力 (edge-level attention) 機制，將信息繼續傳播給下一層節點 n_i 。過程如下：

$$h_i^l = \sigma \left(\sum_{e_j \in E_i} \beta_{ij} W_2 f_j^l \right) \quad (7)$$

其中 h_i^l 是節點 n_i 的輸出表示、 W_2 是可以訓練的權重矩陣、 β_{ij} 為超邊 e_j 與節點 n_i 的注意力參數，並用以下方式去做計算：

$$\beta_{ij} = \frac{\exp(a_2^T v_j)}{\sum_{e_p \in E_i} \exp(a_2^T v_p)} \quad (8)$$

$$v_j = \text{LeakyReLU}([W_2 f_j^l || W_1 h_i^{l-1}]) \quad (9)$$

其中， a_2^T 是另一個權重(上下文)向量用以計算超邊的重要程度，而「||」符號為串接 (concatenation) 的操作。注意力當中使用的激勵函數為 LeakyReLU。

以上提出的雙重注意力機制，可以使超圖注意力網路層不僅能取得單詞間的交互關係，還可以在節點表示的學習過程中，強調顯示不同精細度的關鍵信息。

3.4 歸納文本分類

對於每個文本，經過單個超圖注意力網路層後，我們能夠計算構建的文本超圖上的所有節點表示。然後，我們將學習到的節點表示使用均值池化 (mean-pooling) 的操作，以獲得

文本表示 z ，並將其饋送到 sigmoid 層進行多標籤文本分類：

$$\hat{y} = \text{sigmoid}(W_c z + b_c) \quad (10)$$

其中 W_c 是將文檔表示映射到輸出空間的參數矩陣， b_c 是偏差 (bias)。 \hat{y} 表示預測的標籤機率。最後，多標籤文本分類的損失函數定義為交叉熵損失：

$$L = \sum_{c=1}^C y^c \log(\sigma(\hat{y}^c)) + (1 - y^c) \log(1 - \sigma(\hat{y}^c)) \quad (11)$$

其中 C 為標籤數，我們假設文本的真實標籤是 $y \in R^C$ ，其中 $y^i = \{0, 1\}$ 表示第 i 標籤是否存在。

我們使用的 HyperGAT 模型，對於新加入的測試文本可以直接將它們對應生成的文本超圖，提供給先前訓練好的模型預測它們的標籤，解決了訓練期間需要對測試文本的強制訪問，讓模型可以用歸納 (inductive) 的方式處理新添加的數據，而不是重新訓練模型。

4 效能評估

4.1 實驗資料

我們使用的實驗資料為中文幽默多標籤資料集 (Tseng et al., 2020)，資料來源為網路蒐集的笑話，經由人工標記類別標籤後分別是：雙關、誇飾、擬人化、僞介推論、法則誤用、悖理、模仿、其他技巧、親和、自我提昇、攻擊、自我貶抑、禁忌以及其他意圖，總共有 14 個標籤。資料集文本總數是 3,365 則笑話，平均字數是 114.47、平均詞數是 76.98，標籤總數是 7,227，每篇文章平均 2.15 個標籤。

4.2 實驗設定

在將資料輸入進模型前，我們先將笑話以 ckiptagger 系統進行斷詞，並將斷詞後的結果作為模型的輸入。每個實驗會執行 k-fold cross-validation (k 值取 5) 以確保模型訓練的可信度，我們使用的 HyperGAT 模型參數值設置如下：潛在主題抽取 T 為標籤數量 14；而機率最大的前 K 個單詞設定為 10；批次大小 (batch size) 的大小設定為 8；學習率 (learning rate) 設定為 0.005；epoch 取 10；文本表示起

Method	Macro F1-score	Micro F1-score	Weighted F1-score	Subset Accuracy
CNN (Yoon, 2014)	13.95	45.83	32.30	7.88
BiLSTM (Liu et al., 2016)	12.42	46.29	33.28	8.50
FastText (Joulin et al., 2016)	21.92	46.20	40.50	13.16
Graph-CNN (Defferrard et al., 2016)	21.27	40.64	39.54	8.44
TextGCN (Yao et al., 2019)	21.27	38.78	37.33	9.45
Text Level GNN (Huang et al., 2019)	22.42	44.48	40.55	12.78
Our used model (HyperGAT)	24.19	46.95	40.84	12.15

表 1: 中文幽默多標籤分類資料集的實驗結果

始向量是 300 維的隨機向量而文本表示之最後一層隱藏層維度為 100；L2 正規化懲罰 (L2 penalty) 參數設定為 $1e-6$ ；dropout rate 設定為 0.3 以防止過擬合 (overfitting)。

4.3 評估方法

多標籤文本分類的結果是一篇文本不僅僅只有單一標籤，無法單純的以二元的方法評估。目前主要的評估方法需要計算出每一個類別的 F1-score，根據不同的方式綜合各個標籤的 F1-score 以評估多標籤分類器的效能。其中最常見的是 Macro F1-score，將所有標籤視為平等，計算方式是將各標籤的 F1-score 先計算出來之後，再取其平均值，此評估方法提升樣本數較少的類別對分類器性能評估之影響。Micro F1-score 先計算所有類別加總的 Precision 和 Recall，然後再計算兩者的調合平均 F1，此評估方法會使樣本數較多之類別對性能評估的影響較大。Weighted F1-score 也是需要先將各標籤的 F1，根據每個標籤真實樣本的數量，賦予每個標籤不同的權重，是一種類似加權平均的 F1-score。最後，Subset Accuracy 是最嚴格的指標，表示所有標籤都正確的樣本百分比，舉凡有一個標籤分類錯誤，則 Subset Accuracy 不將其判斷為正確結果。

4.4 模型比較

表 1 為的實驗結果。在基於序列的方法之中，CNN 表現比 BiLSTM 來得好。CNN 以不同的

核定義的卷積，來抽取序列單詞中的多個區塊信息，並透過池化層將結果拼接起來，可以得到比 BiLSTM 模型較好的文本表示結果。FastText 是繼 Word2Vec (Mikolov et al., 2013) 與 GloVe (Pennington et al., 2014) 之後，較為新穎的單詞表示模型，在執行文本分類的同時也會輸出由文本訓練的詞嵌入向量，其 Macro-F1 score 比基於序列的方法提升。接著，比較基於圖形的方法可以發現 TextGCN 雖然比 Graph-CNN 新穎，但表現上沒有差距太多。由於 TextGCN 為整個實驗資料建構了一個基於文本和詞的巨大圖結構，訓練時會消耗大量的記憶體空間，以及每次測試時都需要重新訓練分類器，因此沒辦法執行在線測試 (online testing)。Text Level GNN 改善了記憶體需求過多的問題，並使用滑動窗口取得文本相對位置，比 TextGCN 模型的 Macro F1-score 高出 1.15%。我們採用的 HyperGAT 模型為文本建立超圖建構，使用兩個不同的超邊抓取句子順序信息與語意信息，用注意力機制取得文本的特徵，以上優勢是一般圖神經網路架構模型無法獲得的。由實驗結果得知，無論哪個效能指標，HyperGAT 皆比其他基於序列和圖形的深度學習方法好。

4.5 錯誤分析

我們想要找出測試資料當中，哪些標籤容易被我們使用的 HyperGAT 模型混淆，具體作法是將各標籤的 FN (False Negative) 算出之後，

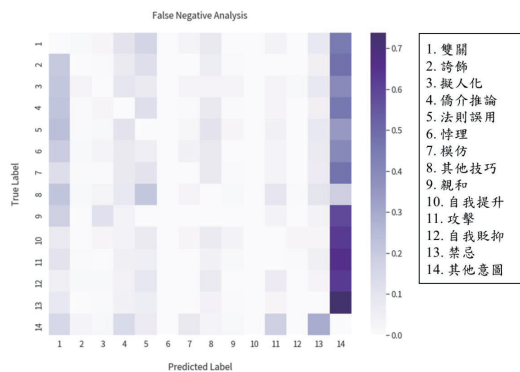


圖 2: HyperGAT 模型錯誤分析

再將與其他標籤的 FP (False Positive) 重疊的索引統計成表格，並用視覺化的方法將數值的大小以顏色深淺作為表達，若顏色越深，表示當某特定標籤無法被正確預測時，模型更傾向歸類出此標籤。

圖 2 為錯誤分析結果。首先，可以發現對角線顏色皆為空白，數值皆為 0 表示其標籤無法被正確預測。本模型的錯誤分析可以歸納以下幾點發現：1) 當標籤沒有被判斷正確時，容易誤判成標籤最多的「其他意圖」，這是因為模型傾向分類到標記數量多的標籤。2) 幽默技巧中，標籤「其他意圖」之直行方格的顏色較深是因為標籤的 FN 數量較小，因此誤判後所占比例較高。3) 在沒有判斷出「其他意圖」這個標籤時，分類器會將標籤分到「攻擊」和「禁忌」。

5 結論

本研究我們運用一種新的基於超圖結構的深度學習方法，解決中文幽默多標籤文本分類問題，透過超圖找出更好的文本表示，我們使用的超圖注意力神經網路模型，效能優於其他基於序列和基於圖形的深度學習模型，達到 Macro F1-score 0.2419、Micro F1-score 0.4695、Weighted F1-score 0.4084 與 Subset Accuracy 0.1215。

未來研究方向，我們將持續改善 HyperGAT 提升模型的表現，並且應用在其他領域，例如：健康照護和醫療問題等。

致謝

This work was partially supported by the Ministry of Science and Technology, Taiwan under the grant MOST 108-2218-E-008-017-MY3 and MOST 109-2410-H-003-123-MY3.

參考文獻

- Jerome Bellegarda. 2014. Spoken Language Understanding for Natural Interaction. In *Proceedings of the SIGDIAL 2013 Conference*, page 203. <https://aclanthology.org/W13-4033.pdf>.
- Kim Binsted. 1995. Using humour to make natural language interfaces more friendly. *Paper presented at the AI, ALife and Entertainment Workshop, Montreal, Canada*. <https://www2.hawaii.edu/~binsted/papers/BinstedI JCAI1995.pdf>.
- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *The Journal of machine Learning research*, vol. 3, pages 993-1022. <https://dl.acm.org/doi/10.5555/944919.944937>.
- Jennings Bryant and Dolf Zillmann. 1989. Chapter 2: Using Humor to Promote Learning in the Classroom. *Journal of Children in Contemporary Society*, 20(1-2), pages 49-78. https://doi.org/10.1300/J274v20n01_05.
- Hsueh-Chih Chen, Yu-Chen chan, Ru-Huei Dai, Yi-Jun Liao, and Cheng-Hao Tu. 2017. Neurolinguistics of Humor. In S. Attardo (Ed.), *The Routledge Handbook of Language and Humor*, pages 282-294. <https://doi.org/10.4324/9781315731162>.
- Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *Proceedings of the 30th International Conference on Neural Information Processing Systems December 2016*, pages 3844-3852. <https://doi.org/10.5555/3157382.3157527>.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. *Human Language Technologies*, pages 4171-4186. <https://doi.org/10.18653/v1/N19-1423>.
- Kaize Ding, Jianling Wang, Jundong Li, Dingcheng Li, and Huan Liu, 2020. Be More with Less: Hypergraph Attention Networks for Inductive Text Classification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural*

- Language Processing*, pages 4927–4936. <https://doi.org/10.18653/v1/2020.emnlp-main.399>.
- Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK. <https://doi.org/10.1017/CBO9780511574931>.
- James Goodman, Andreas Vlachos, and Jason Naradowsky. 2016. Noise reduction and targeted exploration in imitation learning for abstract meaning representation parsing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, pages 1–11. <https://doi.org/10.18653/v1/P16-1001>.
- Mary Harper. 2014. Learning from 26 languages: Program management and science in the babel program. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin City University and Association for Computational Linguistics, page 1. <http://aclweb.org/anthology/C14-1001>.
- Sepp Hochreiter and Jurgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, vol. 9, no. 8, pages 1735–1780. <https://www.bioinf.jku.at/publications/older/2604.pdf>.
- Lianzhe Huang, Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng WANG. 2019. Text level graph neural network for text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 3444–3450. <https://aclanthology.org/D19-1345.pdf>.
- Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (Volume 2, Short Papers)*. Association for Computational Linguistics, pages 427–431. <https://aclanthology.org/E17-2068>.
- Yoon Kim, 2014. Convolutional Neural Networks for Sentence Classification, In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, page 1746–1751. <https://doi.org/10.3115/v1/D14-1181>.
- Pengfei Liu, Xipeng Qiu, and Xuanjing Huang, 2016. Recurrent neural network for text classification with multi-task learning. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, July 2016, pages 2873–2879. <https://doi.org/10.5555/3060832.3061023>.
- Paul E. McGhee. 1989. *Humor and Children's Development: A Guide to Practical Applications*. Oxford, UK: Routledge.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. <https://arxiv.org/pdf/1301.3781.pdf>.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543. <https://aclanthology.org/D14-1162.pdf>.
- Yuen-Hsien Tseng, Wun-Syuan Wu, Chia-Yueh Chang, Hsueh-Chih Chen, and Wei-Lun Hsu. 2020. Development and Validation of a Corpus for Machine Humor Comprehension. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, pages 1346–1352. <https://aclanthology.org/2020.lrec-1.168>.
- Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Graph convolutional networks for text classification. In *Proceedings of the AAAI conference on artificial intelligence, 2019*, vol. 33, no. 01, pages 7370–7377. <https://doi.org/10.1609/aaai.v33i01.33017370>.