

應用強化學習與知識圖譜於故事共述生成之研究

Story Co-telling Dialogue Generation via Reinforcement Learning and Knowledge Graph

李聿鎧 Yu-Kai Lee 張嘉惠 Chia-Hui Chang

國立中央大學資訊工程學系

Department of Computer Science & Information Engineering,
National Central University

ykleee@g.ncu.edu.tw, chia@csie.ncu.edu.tw

摘要

模仿重述一則故事是一種培養學生敘事力的方法，但對於記憶力較差或是無法自己完成描述一個故事的學生來說，這可能帶來一些困難。因此，我們希望利用自然語言處理技術，開發一款故事共述對話模組，該模組能與學生共述一則英語故事，藉此培養學生的敘事能力。然而，故事共述是一項相對較少人涉及且相對新穎的任務，模型需具備兩大能力：(1) 理解故事的內容，以掌握故事劇情和資訊；(2) 根據目前對話討論其餘故事相關劇情。

我們採用開放領域資訊擷取技術來建構知識圖譜，並採用多代理人強化學習方法，讓兩個代理人根據對話歷史從知識圖譜中選擇相關的事實來生成回覆，並共同完成故事共述的任務。基於這些能力，我們能根據目前的對話歷史與候選回覆中，做出更明智的選擇。相較於僅依照時間順序回覆，我們的模型經由自我訓練的獎勵評估，性能從 67.01% 提升至 70.81%，上升了約 3.8%。

關鍵字：強化學習、知識圖譜、故事共述、對話機器人

1 Introduction

故事重述是提升學生敘事力的方式之一，然而並非每位學生都能完整闡述一則故事，這可能是因為他們語言掌握度較弱、或較難組織複雜劇情，或是在表達想法和感受的過程中遇到障礙。為解決這個問題，我們根據鷹架式學習理論 (Wood et al., 1976) 的概念，提出故事共述的任務來輔助學生故事重述的任務。所謂鷹架式學習理論的概念是當建築物還在建構中時，我們會給予鷹架作為暫時性支撐，當建築完工後或學習成熟後，即可將鷹架拆掉。如同我們在學腳踏車時，會架設輔助輪一樣，透過故事共述可以在必要時提供學生必要的支持，並於學生敘事能力增強時逐漸減少輔助。期望透過與機器人的互動，學生可以提升口語表達和書面寫作能力，同時享受個性化的學習體驗。

本研究目標開發一個故事共述對話模組，旨在協助與學生共同敘述長文本英語故事，藉以培養英語敘事能力。為了使研究目標更加明確，我們限制對話模組僅能參與與共述故事相關的對話，而非無目的的閒聊。因此，我們的對話模組被設計為聊故事機器人之支援模組，根據學生目前所述的故事，決定下一個要講述的情節，完成兩人的故事共敘。

故事共述是一項相對較少人涉及且相對新穎的任務，且與常見的故事生成任務有所不同。故事生成只專注於生成合理的後續情節，而故事共述則以原始故事的內容為基礎講述故事，這種差異需要我們重新思考模型的設計和訓練方法。我們預期採用強化學習技術來實作故事共述模組。

以強化式學習為基礎的故事共述對話系統設計有以下四個挑戰：

- 若要以在線強化學習訓練對話機器人，要求對話機器人能從實際與人互動中學習，可能會相當耗費時間與金錢成本。
- 若要以離線強化學習訓練對話機器人，目前沒有現成的故事共述對話語料集可供使用，這使得我們需要生成相關資料集。
- 強化學習模型需要有獎勵函數，如何針對每一輪的對話產出合適的環境獎勵，也是本文需要克服的問題。
- 當故事是一個長文本時，代理人如何理解全部的內容，如何選擇下一個可以闡述的故事情節或事件？

受到 Andrus et al. (2022) 的研究啟發，我們開發一個基於開放領域知識圖譜的故事共述對話模組。知識圖譜不但可以有效濃縮長文本故事的內容，也能使語言模型能夠理解長文本故事。為了提高生成回覆的連貫性和相關性，我們引入了多代理人強化學習 (Multi-Agent Reinforcement Learning, MARL) 技術，讓兩個代理人基於對話歷史和知識圖譜選擇最佳的

回覆。基於上述方法，我們的模型能在不同的決策情境中作出更明智的選擇。通過自我訓練的獎勵評估，我們觀察到相較於僅按照時間順序回覆，我們的模型的性能，從 67.01% 提升至 70.81%，約提升了 3.8%，這意味著我們的模型具有可行性。

2 Related Work

隨著自然語言處理技術的蓬勃發展，帶動了許多前沿性的研究方向與應用，其中，以人機對話方式進行交流的需求不斷增加。不過現今 ChatGPT 雖然已有一般回應使用者詢問的能力，但是大多以被動回應使用者的功能為主。要能扮演輔助學生、主導對話進行的特定目標仍有相當大的研究空間。

2.1 故事對話機器人

對話機器人在教育領域的應用受到廣泛的關注。不同教育工作者希望教育型機器人扮演的角色和提供的功能並不相同。例如加州大學 Irvine 大學的教育團隊 Zhang et al. (2022) 開發了一款陪伴家長與小孩共同閱讀故事的系統稱為 StoryBuddy，並在閱讀過程中搭配問答來增添親子互動。他們提出 FairyTaleQA 資料集 (Xu et al., 2022)，透過問答生成 (Question Answer Generation, QAG) 來解決不知如何問問題家長的困擾。透過實驗發現藉由提問和回答學生問題來實現陪伴式閱讀，孩子在回答問題時表現更具理解力 (Xu et al., 2021)。

另一方面，Chu and Min (2021) 開發一款用於用於小學英語故事書重述對話機器人，對話機器人通過提問協助學習者進行複述，並以規則式 (Rule-based) 判斷故事各場景 (Scene) 是否皆已正確提及，舉例來說，若第一個場景已經被正確提及，則機器人會提示學生講述後續場景，若沒有正確提及，則要求同學重新講述該場景。藉由前述方法，直到所有場景皆已被正確提及。透過這個方法，逐步引導學生獨立進行故事複述。上述研究結果凸顯了聊故事機器人在促進親子互動、培養閱讀興趣和提高學習成效方面的潛力。

延續上述陪養學童敘事力的主題，近來許多研究也致力於利用資訊擷取技術來進行長文本故事的理解與分析，透過資訊擷取技術將非結構化的文本轉為結構化資訊。例如，Xu et al. (2023) 開發一種文檔級 (Document-level) 敘事事件鏈擷取工具 (Narrative Event Chain Extraction Toolkit, NECE)，該方法利用語義角色標記 (Semantic Role Labeling, SRL) 等技術，從故事中提取角色和事件的相關資訊，並使用特定的 TF-IDF 演算法找出較重要的事

件。透過此架構，能有效地解析長文本故事中的敘事結構，提取出其中的角色和事件等元素。

同樣的，Andrus et al. (2022) 透過動態知識圖譜來應對長文本故事理解。不同於關於現實世界資訊的靜態常識知識圖譜，Andrus et al. (2022) 使用 OpenIE (Angeli et al., 2015) 開放領域資訊擷取技術建立動態知識圖譜，並將其應用於問答和故事完成等任務，解決文檔長度超過語言模型輸入限制時的故事理解，並證實其有效性。

3 Method

在本研究中，我們以 Andrus et al. (2022) 的動態知識圖譜故事理解演算法為基礎，進行一些修改和實作，架構如圖 1 所示。首先我們會說明如何把長文本故事轉換成知識圖譜 (Knowledge Graph)；接著將介紹代理人 (Agent) 如何利用對話歷史和知識圖譜，透過強化學習選擇接下來要共述的劇情；其次將介紹環境 (Environment)，說明如何在沒有現成故事共述語料的狀況下建構一個對話歷史評估模型 (Dialogue Evaluation Model)，該模型負責評估目前故事共述的表現，並將其回饋給代理人；最後我們會講解如何把上述部分用強化學習將其整合成一個能根據目前對話歷史進行決策的故事共述對話模組。

3.1 長文本故事表示法

建構知識圖譜的目的是濃縮長文本故事的資訊，並將非結構化的資料轉換為結構化的形式，讓我們的模型能夠有效地理解故事劇情內容。我們使用由 Stanford NLP Group 開發的自然語言處理套件 Stanford CoreNLP (Manning et al., 2014) (Qi et al., 2020)，其中包含了開放資訊擷取框架 OpenIE (Angeli et al., 2015)。OpenIE 能自動從文本中提取結構化的事實三元組 (Fact Triples)，這些三元組包含了主語 (Subject)、謂詞 (Relation) 和賓語 (Object) 的資訊。為了減輕開放領域資訊擷取模型可能產生過多重複事實三元組，我們移除重複的事實三元組，並留下較長的事實三元組，藉此儘量保留較多資訊。同時，我們可以使用 Stanford CoreNLP 所提供的指代消解 (Coreference Resolution) (Raghuathan et al., 2010; Lee et al., 2011; Recasens et al., 2013) 對文本進行處理，將代名詞替換成原本所提及的名詞。

給定的故事文本 D ，我們能夠利用 OpenIE 從長文本中逐句擷取所有事實三元組 $f = (subject, relation, object)$ ，並記錄句

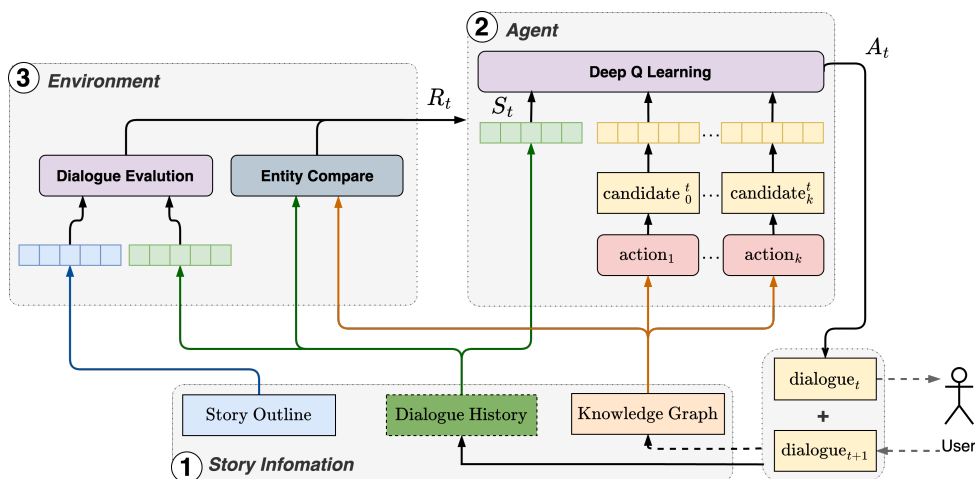


Figure 1: 應用強化學習與知識圖譜之故事共述模組架構

子索引 sid_x ，以建構故事知識圖譜 $G = [f_0, f_1, \dots, f_n]$ 。以「After a time there was another feast, and the Many-furred Creature begged the cook as at the last one to let her go and look on.」為例，本句雖然只有描述“多毛生物像上次一樣懇求廚師讓她去觀看”，但是因為受限於三元表示法，所以第二個受詞、時間、地點以及其他詞都要分開來記錄。因此可將前句表示為 [many furred creature, begged, the cook]、[many furred creature begged the cook, adv, as at the last one] 與 [many furred creature begged the cook, arg2, to let her go and look on] 三個事實三元組，將其以有向圖呈現如圖2所示。

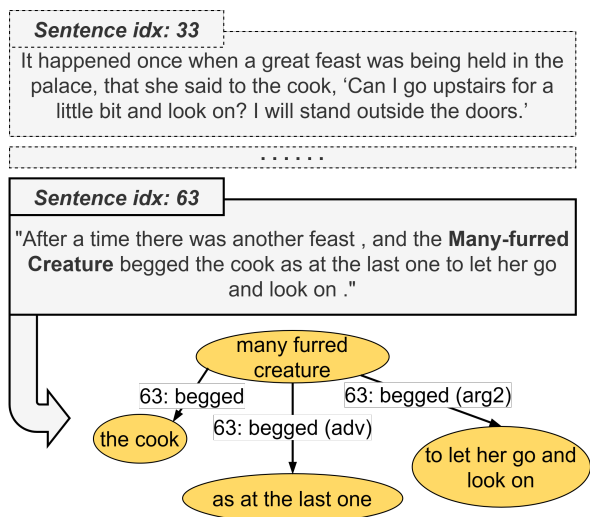


Figure 2: 透過 OpenIE 建構知識圖譜範例

在實作上，每個事實三元組除了儲存 (subject, relation, object) 之外，另外也記錄了該句子在原故事中的索引以理解故事情節的時間關係。此外，我們為每個事實保留一個狀態

指示器，以記錄它是否被提及。這有助於防止在共同講述過程中重複提及同一事實。

3.2 代理人

在本文中我們採用 Deep Q Learning 強化式學習來設計對話代理人，依據目前的狀態 S_t 進行接下來的行動 A_t ，並根據環境獎勵生成的回饋 R_t 來調整代理人的決策。 S_t 是由數個資訊連結起來的向量，分別是對話歷史 $D = [u_0, u_1, \dots, u_t]$ ，以及透過數種不同策略 $A = [a_0, a_1, \dots, a_k]$ 所產生的候選回覆 $C_t = [c_0^t, c_1^t, \dots, c_k^t]$ 。這些文字都會經過 Sentence Transformers 得到最後一層隱藏狀態，藉此代表這些文字各自的隱含資訊。這些資訊經過 Deep Q Learning 後選出要使用的候選回覆進行回應，也就是決策接下來的行動 $A_t = i$ ，其中 $i \in [0, k]$ 。接著我們介紹代理人動作設計及候選回覆的生成方法。強化學習細節將在後續小節討論。

3.2.1 知識圖譜篩選決策

為了讓雙人共述的故事有連貫性，代理人會根據對話歷史最新一則對話 u_t ，用 Sentence Transformer 在知識圖譜 G 找到前三個相關事實作為代表對手目前所述內容的參考點 p 。接著，以此參考點為基準，利用不同的篩選策略來選取故事其他情節，以生成候選回覆。我們將每個策略視為一種動作 a_i ，這些策略的動作如下所示：

- a_0 : 選擇參考點後續劇情，換言之，選擇的事實 f 的 $f.sid_x$ 必需大於參考點 $p.sid_x$ 。
- a_1 : 選擇跟參考點主詞 $p.subject$ 相似的事實，也就是補充跟上一句主詞 subject 相關的資訊。

- a_2 : 選擇跟參考點動詞 $p.relation$ 相似的事實，也就是補充跟上一句事實之 relation 相關的資訊。
- a_3 : 選擇跟參考點受詞 $p.object$ 相似的事實，也就是補充跟上一句事實之 object 相關的資訊。
- a_4 : 宣告結束。

3.2.2 生成候選回覆

我們可以利用先前使用 OpenIE 得到的故事句子與其對應的事實三元組，將輸入及輸出對掉，對 T5 模型進行微調，讓 T5 模型能夠根據事實三元組和故事大綱生成與原本句子相近的結果，其輸入範例如圖3所示。我們希望 T5 模型能夠根據給定的事實三元組和故事大綱，產生出與原本句子盡可能相近的結果，增強生成句子的一致性和品質。

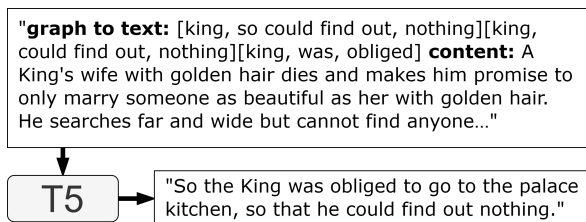


Figure 3: 微調 T5 模型以實現知識圖譜到文本

3.3 環境：獎勵函數之設計

環境結構主要分成兩個部分，包括對話歷史評估和實體關係評估，前者提供截至目前為止的整體敘事評分 DH ，後者計算本輪與前一句話的連貫性評分。我們會綜合這兩者的表現，並將其作為當前第 t 回合的總分數 $Score_t$ ，這個分數減去前一回合的總分數 $Score_{t-1}$ ，即為本輪的獎勵 R_t ，然後回饋給代理人。

$$\begin{aligned} Score_t &= S(P, H) + EC(U, G) \\ R_t &= Score_t - Score_{t-1} \end{aligned} \quad (1)$$

3.3.1 對話歷史評估

為了評估故事共述對話歷史 H 的成效，我們需要有好和不好的故事共述範例及其評分，可以用來訓練一個對話歷史迴歸模型來自動評估故事共述的好壞。由於目前沒有現成的故事共述對話資料集可供使用，因此我們採用 ChatGPT 對每個故事生成指定數量的條列式故事劇情重點，希望這些劇情重點可以成為故事共述對話的資料集，並訓練一個故事共述對話歷史評估模型，評價目前對話的品質。

如表1所示，我們設計了一個提示 (Prompt)，引導 ChatGPT 為故事文本生成指定數量的劇情重點。為了方便後續處理，生成結果以 JSON 格式輸出。由於 ChatGPT 的生成多樣性，同一個 Prompt 可以產生多種不同的結果。

Input
< Plots > = number of plots that you want to generate
< Story_text > = story corpus
Prompt
Please summarize the following story by outlining < Plots > plot points in JSON format in order. (example: [{"plot_id": 1, "plot_point": "first plot point"}, {"id": 2, "plot_point": "second plot point"}]) Do not provide additional information or comment.
—
< Story_text >

Table 1: 透過 ChatGPT 生成故事劇情重點的提示格式

我們使用 ChatGPT 生成劇情重點。接著，我們對已生成的劇情重點進行替換或刪除等操作，產生了品質較差的劇情重點。這種生成劇情重點的方法可以視為故事共述中的不良範例，因為它們可能會破壞故事的完整性和邏輯性。而根據替換或刪除的次數，我們給予不同的分數。

由於替換和刪除對劇情重點品質的影響程度不同，我們根據這兩種操作分別設定了不同的調整幅度公式和評價公式，如式2所示為刪除劇情重點的公式，式3所示為替換劇情重點的公式，其中 n 為異動次數， $|Plots|$ 為原劇情重點數量。我們認為將現有的某個劇情重點替換成其他的故事劇情重點，會對整體劇情重點的品質產生更大的影響，相較之下，刪除一個劇情重點的影響較小。因此，替換較多的劇情重點將獲得比刪除相同數量劇情重點更低的分數。

$$score = e^{(-1.6 \times \frac{n}{|Plots|})} \times 9 + 1 \quad (2)$$

$$score = e^{(-4 \times \frac{n}{|Plots|})} \times 10 + 1 \quad (3)$$

對話歷史評估本質上是一個迴歸問題，如圖4所示，我們將對話歷史 H 和故事大綱 P 分別輸入相同的 RoBERTa (Liu et al., 2019) 模型，並從模型中提取出 CLS 的隱藏狀態。接下來，將兩個隱藏狀態進行連接，將它們輸入到一個神經網絡中。讓模型輸出一個介於 0 和 10 之間的分數，以評估故事共述的品質。

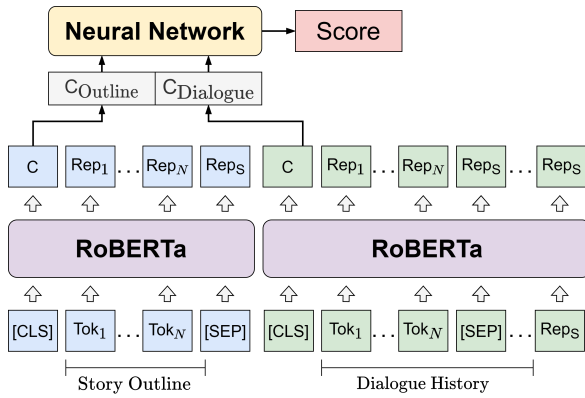


Figure 4: 對話歷史評估模型架構

這種架構的設計利用 RoBERTa 語言模型的能力來理解對話歷史和故事大綱。通過從兩個輸入中提取隱藏狀態，模型可以捕捉到兩者之間的相關資訊。然後，連接這兩個隱藏狀態，使模型能夠整合兩者的特徵資訊。最後，通過前饋神經網絡進一步處理這些特徵，模型可以生成一個 0 到 10 之間的分數，評估故事共述的品質。

3.3.2 實體關係評估

實體關係評估的目的是評估當前回覆的劇情 (U) 是否與前一句劇情的實體 (G) 有關聯。我們會透過 OpenIE 解析這兩句話，並運用 BFS 圖演算法來判定這兩個實體是否在知識圖譜中能夠連結。如果兩個實體在知識圖譜中能連結，我們就認定這兩句話有實體關係，並給予定量的獎勵作為回饋。

3.4 強化學習

在本節中，我們將整合前面所建構的環境、代理人和獎勵，使用 Deep Q Learning (下稱 DQL) 與多代理人強化學習 (Multi-Agent Reinforcement Learning, MARL) 方法，使兩個代理人能夠協同完成故事共述任務，透過獎勵分數的引導能學習如何接續講下面的故事。由於我們的故事共述不會只針對同一則故事進行共述，導致故事共述環境是不穩定的，因此必須仔細處理以避免災難性失憶 (Catastrophic Forgetting)，要在狀況隨機變化的情況下成功進行故事共述，演算法必須對環境有深入的理解，提升其泛化性。我們將選擇以經驗回放的 DQL 作為我們的強化學習方法。

將上述經驗回放 DQL 架構，結合多代理人強化學習場景的架構訓練演算法如式1所示。我們的目標是讓兩個代理人能夠相互共述一個故事，在每個對話輪次中，兩個代理人輪流透過環境傳遞故事共述的內容，但彼此不分享各自的知識圖譜狀態標記。這意味著每個代理人

只能透過故事共述對話歷史來理解故事共述劇情的發展並做出適當的回應。如果其中一方提前結束，整個對話也隨之終止，進行後續的分析和評估。這種設計能夠模擬現實世界中人機互動的情境，挑戰代理人的理解和回應決策能力。

在訓練開始前，我們會先初始化每個環境與模型 (第 1 至 5 行)。在每個 epoch 中 (第 6 行)，我們會對每個故事進行一次對話 (第 7 行)，同時在共述開始前初始化每個環境狀態 (第 8 至 11 行)。在第 12 至 23 行中，可以觀察到兩個代理人會輪流產生候選回覆，將其向量與對話歷史向量進行連接，形成目前的狀態表示 (第 13 至 14 行)。隨後，代理人會透過自身的 Q Network 來決定選擇哪個候選回覆 (第 15 至 16 行)。接下來，我們進行對話歷史評估模型與實體關係評估，以生成獎勵 (第 17 至 19 行)，同時產生下一個狀態 (第 20 至 21 行)。最後，將狀態轉換、行動、下一個狀態和對應獎勵 (s, a, s_{t+1}, r_{t+1}) 存放至各自的記憶中 (第 22 行)，以供後續的學習和更新過程使用。

4 Experiment

在本研究中，我們選擇使用 FairytaleQA (Xu et al., 2022) 中的故事作為故事共述的指定故事集。這些故事都是適合九年級以下學生閱讀的經典童話，且具有清晰的敘事結構。FairytaleQA 所使用的故事平均文本長度超過一千字。作者從古騰堡計劃網站¹使用「Fairytale」作為關鍵詞進行搜索，然後從下載次數最多的故事中進行篩選，共計 258 個故事。另外由 FairytaleQA 已有設計好的問題答案對，我們也可借由問題回答來評估故事共述的內容多元性。

我們設置了環境的結束條件，當其中一方符合以下任一條件時，對話將結束。首先，通過限制對話輪數超過 20 輪，我們能夠避免對話過長導致的資源浪費和訓練時間增加。此外，當其中一方提出結束關鍵字時，表示對話已達到一個合適的結束點，可以給予使用者更好的對話體驗。另外，我們設定了知識圖譜中的所有事實被使用完畢作為結束條件之一。這樣的設定確保了對話過程中的資訊利用率，並且避免了重複使用相同的事實。

4.1 對話歷史評估模型訓練成效

在對話歷史評估模型的訓練中，我們設定批次大小 (Batch Size) 為 1，並進行了 20 個訓練周期 (Epochs)。在這些訓練中，我們的訓

¹網站網址：<https://www.gutenberg.org/>

Algorithm 1: Story Co-telling MARL

Data:

$I = [(O_1, G_1), (O_2, G_2), \dots]$ Story info.;
 $O_j =$ Story outline;
 $G_j =$ Story knowledge graph;

Function:

$\mu =$ State embedding model;
 $\Phi =$ Candidate response generate func.;
 $\Theta =$ Dialogue evaluation model;
 $\Xi =$ Entity compare func.;

Training:

```

1 Initialize Agnest1 and Agnest2;
2 Initialize Q Network  $Q_1$  and  $Q_2$ ;
3 Initialize epsilon  $\varepsilon$ ;
4 Initialize replay memory  $M_1$  and  $M_2$ ;
5 Initialize environment  $E_1$  and  $E_2$ ;
6 foreach epoch do
7   foreach  $(O_j, G_j)$  in  $I$  do
8     Reset dialogue history  $D$ ;
9     Reset environment  $E_1$  and  $E_2$  by
       $(O_j, G_j)$ ;
10     $t = 1$ ;
11     $Score_t = 0$ ;
12    while ( $E_1$  is not done) and ( $E_2$ 
      is not done) do
13       $C_t \leftarrow \Phi(D, G)$ ;
14       $s_t \leftarrow \{\mu(D), \mu(C_t)\}$ ;
15       $a_t \leftarrow \operatorname{argmax}(Q_{t\%2}(s_t, \varepsilon))$ ;
16       $d_t \leftarrow C_t[a_t]$ ;
17      Append  $d_t$  to  $D$ ;
18       $Score_{t+1} \leftarrow$ 
         $\Theta(O_j, D) + \Xi(G_j, D)$ ;
19       $r_{t+1} \leftarrow Score_{t+1} - Score_t$ ;
20       $C_{t+1} \leftarrow \Phi(G)$ ;
21       $s_{t+1} \leftarrow \{\mu(D), \mu(C_{t+1})\}$ ;
22      Append  $(s, a, s_{t+1}, r_{t+1})$  to
         $M_{t\%2}$ ;
23       $t = t + 1$ ;
    end
24    Update  $Q_1$  by  $M_1$ ;
25    Update  $Q_2$  by  $M_2$ ;
  end
26 Update  $\varepsilon$ ;
end

```

練集損失值為 0.0197，這表示模型在訓練集上有很好的擬合效果；而最佳的驗證集損失值為 0.0299，在未見過的資料上也達到了一定的性能。此外，我們還計算了評分標籤與對話歷史評估模型輸出兩者的相關係數，結果為

0.8313，說明資料標籤和對話歷史評估模型輸出之間存在正相關。

表2展示了根據不同的劇情重點輸入，對話歷史評估模型給出的評分情況。其中，score欄位展示了從第一則對話到當前記錄的總分數，而整個對話的預設分數 Label 則標記在最底部。當輸入為高品質的劇情重點時，模型輸出結果接近預設的分數。這表明我們的對話歷史評估模型能夠對劇情重點進行準確評估，並給予適當的分數。如果在劇情重點中插入與故事無關的劇情，對話歷史評估模型給出的分數會顯著降低。這進一步說明了我們對話歷史評估模型的有效性和可行性，它能夠識別出與故事相關的重點並對其給予適當的評價。

4.2 故事共述成效

在本研究的實驗中，我們與規則式回覆進行了效能比較。在規則式回覆中，我們設定了一個固定的動作 a_0 ，也就是僅依照時間順序進行回覆。圖5展示了我們的訓練結果，可以看到透過不管是單環境強化學習 (1Env) 還是多環境強化學習 (2Env)，其表現都優於規則式回覆 (Rule-based)，而多環境強化學習表現最佳，以我們訓練的對話歷史評估模型回饋值來看，多環境強化學習的性能比規則式回覆還上升了約 3.8%，從 67.01% 提升至 70.81%。

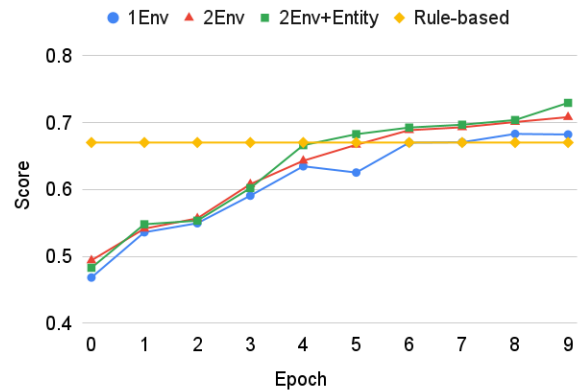


Figure 5: 故事共述以不同方法訓練之結果比較

此結果說明多代理人強化學習方法在故事共述任務上的可行性。相較於僅依照時間順序的規則式回覆，我們的模型通過多代理人的互動訓練，能夠更好地理解對話歷史並根據知識圖譜生成回覆。這使得我們的模型能夠提供更具連貫性和相關性的回覆，進一步提升了對話的品質和體驗。

4.3 比較不同獎勵對動作選擇之影響

最後，我們探討「僅使用對話歷史評估模型作為唯一的獎勵」與「對話歷史評估模型結合

turn	history	score
...
6	The Princess falls asleep in a hollow tree and is discovered by the King's huntsmen.	7.37
7	The King's huntsmen bring the Princess to the palace and she is assigned to work in the kitchen as the Many-furred Creature.	7.34
8	The Many-furred Creature lives in poverty and works in the kitchen doing all the dirty work.	7.79
9	The Many-furred Creature attends a feast at the palace and enchants the King with her beauty.	7.95
10	The Many-furred Creature cooks soup for the King and hides a gold ring in it.	8.06
...
14	The King and the Princess live happily ever after.	8.02
Final Score=8.02, Label=9.09		

turn	history	score
...
6	The Princess falls asleep in a hollow tree and is discovered by the King's huntsmen.	7.37
7	The Emperor takes Confucius' shoes and staff as a joke, but the tablet's warning comes true and he dies soon after.	6.34
8	The cock gets the garland and trades it for red silk from the brook.	4.82
9	The jackdaws and magpie eat the leftover pie-crust and gravy.	2.74
10	The Many-furred Creature cooks soup for the King and hides a gold ring in it.	2.17
...
14	The King and the Princess live happily ever after.	6.21
Final Score=6.21, Label=7.38		

Table 2: 比較對話歷史評估模型的表現。左半表為高品質的劇情重點在的表現，右半表為加入劇情無關內容致獎勵分數下降的表現。

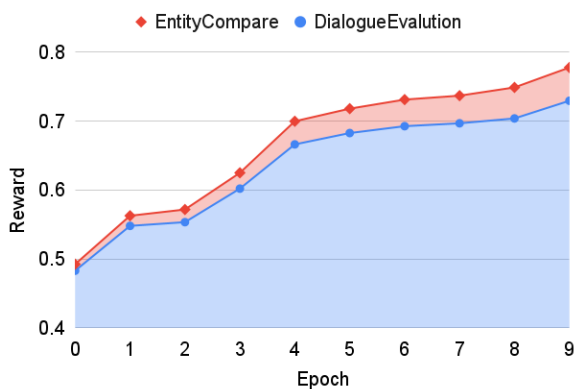


Figure 6: 實體關係獎勵在訓練過程之堆疊面積圖

實體關係評估作為獎勵對模型動作選擇的影響。如圖6所示，我們可以看到隨著訓練過程，平均實體關係評估獎勵也有所提升。除此之外，相較於只有使用對話歷史評估模型作為唯一的獎勵 (DialogueEvaluation)，在實體關係評估 (DialogueEvaluation + EntityCompare) 的鼓勵下，模型傾向於選擇與實體相關的動作 (如圖7)。顯示對話歷史評估模型中引入實體比較的方法對模型的決策有一定的影響。

4.4 故事大綱下的 FairytaleQA 表現

最後，我們嘗試評估共同講述的故事是好還是壞。一種可能的方式是使用問答來測試故事亮點是否能夠回答預先設計的問題。因此，我們使用基於故事摘要的微調 T5 問答模型 (Christian Di Maio, 2022) 進行了實驗。我們將 FairytaleQA 中問題對應的故事段落替換為故事摘要，以評估故事摘要是否能夠有效回答故事中的問題。

實驗結果如表3所示，這個經過微調的 T5 模型在故事全文大綱下的表現並不理想。主要

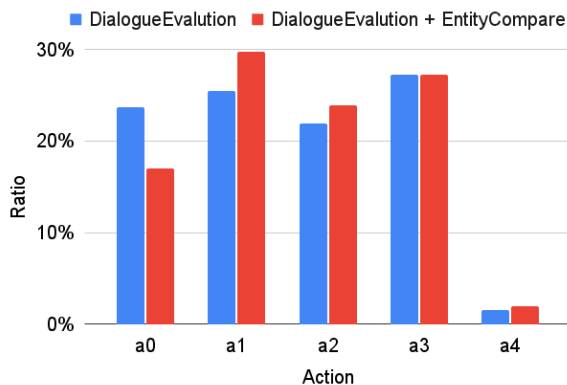


Figure 7: 不同獎勵方法對動作選擇比例直方圖

是因為故事的全文大綱較為簡短，缺乏細節和上下文，導致問題回答模型難以進行準確的回答。此外，故事的大綱可能包含一些隱含的信息，需要模型具備更強的推理能力來處理這些隱含內容。

	Train		Val		Test	
	F1	EM	F1	EM	F1	EM
character	24.11	16.53	27.33	18.69	20.41	11.65
action	11.85	2.19	13.64	3.00	13.27	2.54
setting	15.50	6.50	23.64	6.67	14.34	3.23
feeling	4.60	3.28	3.26	1.06	7.97	4.72
causal relationship	15.87	0.12	17.19	0.00	19.10	0.36
outcome resolution	12.18	0.12	14.22	1.03	17.39	0.00
prediction	16.34	3.55	19.23	1.82	16.30	0.00
All	14.09	3.46	15.93	3.51	15.63	2.78

Table 3: 故事全文大綱下的 FairytaleQA 在微調過的 T5 模型表現

5 Conclusion and Future Work

在本研究中，我們設計一個能夠進行故事共述對話模組，希望藉由這個對話模組，能協助提升學生個體的英文敘事能力。我們運用開放資訊擷取技術，為長文本故事建立了一個知識圖

譜。為了使模型能夠生成更具連貫性和相關性的回覆，我們引入了多代理人強化學習。通過訓練兩個代理人，在對話歷史的基礎上從知識圖譜中選擇最佳的回覆，我們的模型能夠在不同決策動作所生成的候選回覆中作出更明智的選擇。透過自我訓練的獎勵評估，我們觀察到相較於僅依照時間順序回覆，我們的模型的性能從 67.01% 提升至 70.81%。

有關未來的研究，我們的對話模組在處理長文本故事時仍然受限於知識圖譜的完整性和覆蓋範圍。其次，我們的對話模組在生成故事重述時有時會出現表達不夠流暢的情況。未來工作可以探索更先進的自然語言生成技術，例如預訓練的大型語言模型，以改進共述結果的品質和流暢度。綜合來，我們的研究還存在許多改進空間。

References

- Berkeley R Andrus, Yeganeh Nasiri, Shilong Cui, Benjamin Cullen, and Nancy Fulda. 2022. [Enhanced story comprehension for large language models through dynamic document-based knowledge graphs](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(10):10436–10444.
- Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D. Manning. 2015. [Leveraging linguistic structure for open domain information extraction](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 344–354, Beijing, China. Association for Computational Linguistics.
- Giacomo Nunziati Christian Di Maio. 2022. [Mariorossi/t5-base-finetuned-question-answering](#) (huggingface). <https://huggingface.co/MaRi0r0sSi/t5-base-finetuned-question-answering>.
- Seong Yeub Chu and Deok Gi. Min. 2021. [Development of an ai chatbot-based teaching model for english picture book retelling activities](#). *Modern English Education*, 22(4):37–50.
- Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2011. [Stanford’s multi-pass sieve coreference resolution system at the conll-2011 shared task](#). In *Conference on Natural Language Learning (CoNLL) Shared Task*.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. [Roberta: A robustly optimized bert pre-training approach](#).
- Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. [The Stanford CoreNLP natural language processing toolkit](#). In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.
- Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. [Stanza: A Python natural language processing toolkit for many human languages](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.
- Karthik Raghunathan, Heeyoung Lee, Sudarshan Rangarajan, Nathanael Chambers, Mihai Surdeanu, Dan Jurafsky, and Christopher Manning. 2010. [A multi-pass sieve for coreference resolution](#). In *Empirical Methods in Natural Language Processing (EMNLP)*.
- Marta Recasens, Marie-Catherine de Marneffe, and Christopher Potts. 2013. [The life and death of discourse entities: Identifying singleton mentions](#). In *North American Association for Computational Linguistics (NAACL)*.
- David Wood, Jerome S Bruner, and Gail Ross. 1976. [The role of tutoring in problem solving](#). *Journal of child psychology and psychiatry*, 17(2):89–100.
- Guangxuan Xu, Paulina Toro Isaza, Moshi Li, Akintoye Oloko, Bingsheng Yao, Cassia Sanctos, Aminat Adebisi, Yufang Hou, Nanyun Peng, and Dakuo Wang. 2023. [Nece: Narrative event chain extraction toolkit](#).
- Ying Xu, Dakuo Wang, Penelope Collins, Hyelim Lee, and Mark Warschauer. 2021. [Same benefits, different communication patterns: Comparing children’s reading with a conversational agent vs. a human partner](#). *Computers & Education*, 161:104059.
- Ying Xu, Dakuo Wang, Mo Yu, Daniel Ritchie, Bingsheng Yao, Tongshuang Wu, Zheng Zhang, Toby Li, Nora Bradford, Branda Sun, Tran Hoang, Yisi Sang, Yufang Hou, Xiaojuan Ma, Diyi Yang, Nanyun Peng, Zhou Yu, and Mark Warschauer. 2022. [Fantastic questions and where to find them: FairytaleQA – an authentic dataset for narrative comprehension](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 447–460, Dublin, Ireland. Association for Computational Linguistics.
- Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. 2022. [Storybuddy: A human-ai collaborative chatbot for parent-child interactive storytelling with flexible parental involvement](#). In *Proceedings of the 2022*

CHI Conference on Human Factors in Computing Systems, CHI '22, New York, NY, USA. Association for Computing Machinery.