**影片段落邊界辨識與段落描述生成探討**

**作者一 作者二 作者三**

**國立雲林科技大學 資訊管理系**

[**B11023000@yuntech.edu.tw**](mailto:B11023000@yuntech.edu.tw)[**B11023000@yuntech.edu.tw**](mailto:B11023000@yuntech.edu.tw)[**B11023000@yuntech.edu.tw**](mailto:B11023000@yuntech.edu.tw)

摘要

近期線上課程與網路影片興盛，大學端也養成了課程錄影的習慣，但大部分的課程錄影不會做太多的剪輯，一段錄影至少30分鐘起跳，未經整理的影片十分攏長且難以理解。針對此問題本研究將其拆解為兩個部分：邊界辨識和段落摘要，我們希望(本研究期望)將影片字幕檔STR送入邊界辨識模型取得段落切分點，並將每個切分點間的文本送入摘要模型得到段落摘要。邊界辨識的部分本研究使用BERT，透過其在Pre-train學習的NSP(Next Sentence Prediction)任務和段落向量間的餘弦相似性，自動選擇影片段落的切分點；而段落摘要模型我們分別在抽象式的Pegasus和抽取式的句向量K-means比較並選擇適合的模型。

**關鍵詞**：影片段落邊界辨識、影片段落摘要

Abstract

This paragraph describes the major work in your paper.

**Keywords**: not exceeding six keywords.

1. 研究動機與目的

最近眾多影音平台皆推出了許多方便的輔助功能，例如Youtube中的TIMECODE功能，其功用為讓閱聽人快速跳到影片中不同主題的段落，使閱聽人可迅速了解影片的整體架構，並達到快速理解且消化影片內容的效果。而TIMECODE目前多是上傳者在影片上手動標記，一但影片的長度變長，如課程錄影、立法院會期紀錄這種冗長且沒有腳本或大綱的影片，若使用人工標記的方法，不僅要完全理解影片的內容，還得自行切出好幾個不同主題的段落，這絕對是一個耗時的大工程，這也促使我們著手研究可以有效判別影片段落邊界並生成影片段落摘要的方法。

1. 相關研究與文獻探討

本研究主要和文本分割、文本摘要和TRANSFORMER BASE LM等領域有較大的關聯性，我們會在本節中間單回顧這些領域的相關論文。

* 1. Transformer Base LM

Transformer Attention Is All You Need 於2017年由Ashish Vaswani等人提出，為解決RNN（LSTM）或CNN在實現Seq2Seq架構時出現的主要缺陷：遞迴運算無法被平行化加速(或優化)和非常數操作距離導致的長距離依賴困難，Transformer使用自注意力機制（self attention）學習輸入與輸出序列的向量表示，Transformer藉著可平行化和常數操作距離（d=1）的優勢，可以在極大資料集中以快RNN數倍的速度擬和，使NLP領域的轉移學習（transfer learning）得到極大的成長，許多通用預訓練語言模型（Pre-trained Languge Model, PLM）問世，如BERT、GPT、RoBERTa，使用者只需依任務選擇合適的PLM並對其微調（fine tuning）便可達到優異效能。

* 1. 句嵌入

句嵌入是一種將文本中的句子映射為向量的一種技術之統稱，最早可以追朔到Zellig S. Harris在1954提出的詞袋（Bag Of Word, BOW）Distributional Structure 模型，使用詞頻來表示一段文本，但其不考慮詞的順序和意義，並且過於稀疏。 於是在2013由Tomas Mikolov等人提出word2vecEfficient Estimation of Word Representations in Vector Space使用非監督式的方法，從詞的前後文中學習，使用簡單的線性轉換將每個詞映射到向量空間，接著將段落中每個詞的向量取平均，以表達此段落，這個方法成功解決詞袋模型不考慮意義和過於稀疏的問題。 但使用word2vec的方法表達文本有著以下缺陷：

1. 和BOW一樣，不考慮詞序

2. 無視字詞在不同語境下多義性

為解決word2vec的缺陷，Peters等人在2018年初提出了ELMo（Embedding from Language Model）Deep contextualized word representations 同樣使用非監督式的方法，運用BILSTM組成的Seq2Seq架構，從輸入文本中學習推斷出詞在當前位置和語境中的意義，這個方法成功解決了word2vec的缺陷，並一直沿用到了現在，只是現在的模型架構通常會使用self-attention和feedfordward的組合取代BILSTM。

* 1. 文本分割

文本分割是將一段文本切分為多個段落的任務，段落內的主旨應該是近似的，相關研究最早在1990年代提出Multi-Paragraph Segmentation of Expository Text.以改善資訊檢索和自然語言生成等任務的效果，由於多領域的口語資料集目前並不存在，接下來主要會聚焦在非監督式的方法上做說明。

**Lexical Cohesion** Multi-Paragraph Segmentation of Expository Text.是早期的主要解決方案之一，對每個詞出現的位置分布及頻率進行分析，以判斷其詞彙凝聚性（Lexical Cohesion,不確定如何翻譯）。

**主題模型** Discourse Segmentation of Multi-Party Conversation（LDA）假設每個段落由不同主題組成，每個主題是不同的詞彙的集合，且不同主題集合之間沒有交集，而主題模型會從文本中自主學習每個段落的主題構成和每個主題的詞彙構成，透過了解每一個段落的主題構成，我們可以依據主題構成的相似性去決定段落之間是否意義近似。

**圖學分析** Unsupervised Text Segmentation Using Semantic Relatedness Graphs 使用word2vec取得句子或段落之間相似性，構成此文本的語意關係圖（Semantic Relatedness Graphs），並運用其進行分段。