

GAI 2024 Final Project 第 6 組

Written Judgment Summarization 報告

資訊 114 F74106254 熊貫仲

資訊 114 F74101254 張曄俊

資訊 114 F74106084 林宸哲

資訊 114 F74106026 李星辰

Data Preprocessing

1. 繁體轉簡體

由於模型預訓練的資料來源大多是簡體中文，將現有繁體中文資料轉換為簡體中文會有助於模型訓練與生成，確保資料的一致性以提高模型的準確性。我們使用 OpenCC 進行轉換。

2. 移除噪聲

觀察裁判原文，可以發現兩個問題：

- 實際需要摘要的地方是從「如下：」到「判決如主文」，故僅擷取這兩段文字之間的內容。
- 裁判原文中包含許多非一般文字的符號（如 0、0、(一)、o），以及全形半形標點混用的問題，這些符號可能會影響模型的訓練效果。因此，我們對這些符號進行以下處理：

- 數字轉換：將數字符號（如 0、0）轉換成中文數字（如零）。
- 符號處理：將特殊符號（如(一)、o）轉換成中文數字（如一、零）。
- 標點處理：將標點統一為全形標點，以確保資料的一致性。

3. Sliding Window

由於裁判原文長度較長，直接使用可能會導致模型過載，影響訓練效果。因此，我們使用 sliding window 將資料切分成固定長度的子序列。這樣不僅可以有效利用所有資料，還能避免模型在處理過長資料的困難。

完整前處理程式碼如下：

```

1. import opencc
2. num_to_chinese = {
3.     '0': '〇', '1': '一', '2': '二', '3': '三', '4': '四',
4.     '5': '五', '6': '六', '7': '七', '8': '八', '9': '九',
5.     '0': '〇', '1': '一', '2': '二', '3': '三', '4': '四',
6.     '5': '五', '6': '六', '7': '七', '8': '八', '9': '九',
7.     'o': '〇', ',': ',', '.': '。', ';': ';', ':': ':',
8.     '!': '!', '?': '?'
9. }
10.
11. def digit_to_chinese(match):
12.     return ''.join(num_to_chinese[digit] for digit in match.group())
13.
14. def clean_text(text):
15.     if "如下：" in text:
16.         text = text.split("如下：", 1)[-1]
17.
18.     if "判決如主文" in text:
19.         text = text.split("判決如主文", 1)[0]
20.
21.     return text
22.
23.
24. def renoise(text):
25.     text = re.sub(r'<[^>]+>', '', text)
26.     text = re.sub(r'\n', '', text)
27.     text = re.sub(r'[0-90-9o,.;!?!?]+', digit_to_chinese, text)
28.     text = re.sub(r'^\u4e00-\u9fa5\u0001\u000c\u001a\u001b\u001f\u0061', '', text)
29.     text = clean_text(text)
30.     converter = opencc.OpenCC('t2s')
31.     text = converter.convert(text)
32.
33.     return text

```

Model & Training Method

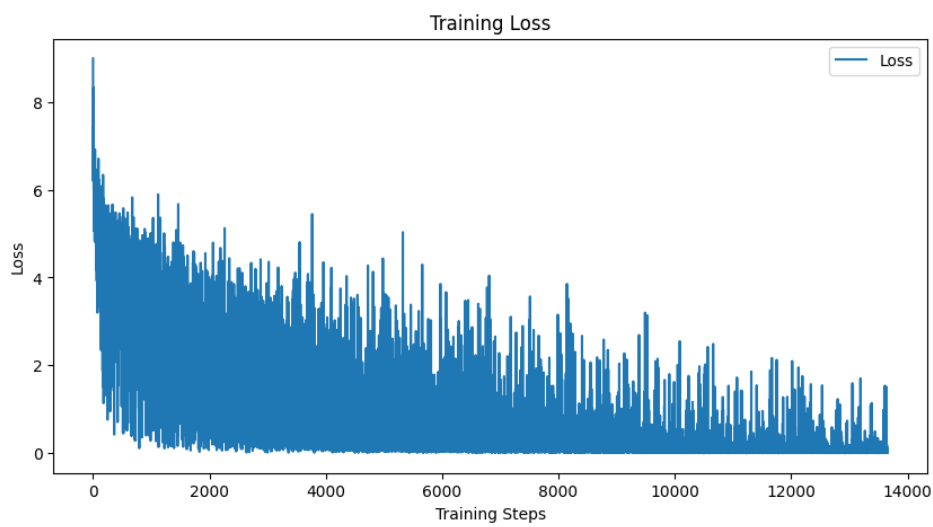
我們使用 IDEA-CCNL/Randeng-BART-139M-SUMMARY。我們嘗試過 T5 模型，發現其訓練速度較慢、Rouge 分數較低、生成結果較不理想，所以我們選擇使用 BART 作為本次使用的模

型。

我們總共訓練 10 個 epoch，最終 rouge 分數如下：

Rouge-L-P	Rouge-L-R	Rouge-L-F
0.6770542939247607	0.7404029746660421	0.7404029746660421
Rouge-2-P	Rouge-2-R	Rouge-2-F
0.23837758058892203	0.24345459245632428	0.24345459245632428

訓練過程損失值如下：



Analysis

在進行這項作業時，我們遇到了以下幾點困難：

1. 生成字數過短
2. 生成內容出現英文與<unk>字元
3. 生成內容不會出現標點符號

為了解決這些問題，我們分別採取對應的解決辦法：

1. 生成字數過短

推測是因為判決原文過長，超過模型最大輸入長度，導致模無法正常運作。因此，我們使用 **sliding window**，將輸入文本裁切成適當的長度，同時也增加了訓練資料數量，從而提升模型的訓練效果。

2. 生成內容出現英文與<unk>字元

推測是因為判決原文內含有模型無法辨認的繁體字或符號。因此，我們在輸入前將判決原文轉換成簡體中文，使模型的輸入與輸出都是簡體中文，最後再自行將結果轉回繁體中文。

3. 生成內容不會出現標點符號

推測是因為判決原文中有太多不同的符號（全形半形混用、數字符號等），造成模型混淆。因此，我們對符號進行前處理，統一轉為全形標點以及中文數字，以減少模型的混淆。

以下是進行前處理與未進行前處理的比較：

有前處理

Rouge-L-P	Rouge-L-R	Rouge-L-F
0.6770542939247607	0.7404029746660421	0.7404029746660421
Rouge-2-P	Rouge-2-R	Rouge-2-F
0.23837758058892203	0.24345459245632428	0.24345459245632428

生成結果：

```
model generate examples:
output: ['按不动产权经登记者,推定登记权利人适法有此权利,民法第七七五九之一第一项定有明文该登记之推定效力,乃登记名义人除不得援以对抗其直接前手之
target: ['按不动产权经登记者,推定登记权利人适法有此权利,民法第七五九条之一第一项定有明文该登记之推定效力,乃登记名义人除不得援以对抗其直接前手之
```

無前處理

Rouge-L-P	Rouge-L-R	Rouge-L-F
0.811365442803453	0.842998488689003	0.8429984886890034
Rouge-2-P	Rouge-2-R	Rouge-2-F

0.20319951490811378	0.20401557252777355	0.20401557252777355
---------------------	---------------------	---------------------

生成結果：

```
model generate examples:
output: ['</s> <unk>捐稽徵法第28<unk>第2<unk>定之退<unk>溢<unk>款<unk>求<unk>, <unk>人<unk>行政救<unk>期限<unk>止, 法律特<unk>予之程序重<unk>
target: ['一、<unk>彩券(或<unk>特<unk>公益彩券), 指以各<unk>技<unk>的, <unk>事<unk>程及其<unk>果<unk>方式之彩券; 此<unk>以「<unk>事<unk>果」
```

可以看到進行前處理後，生成的文字品質與評估分數皆有顯著提升。

Proposal: 法律裁判原文摘要練習小幫手

為了協助法律相關人員提升閱讀與撰寫能力，我們推出了一款專為法律判決主文設計的閱讀與摘要練習程式。這個程式旨在幫助使用者更加精準地理解判決主文內容，並有效地撰寫出正確的摘要。

程式特色：

- **判決原文展示**

程式介面會顯示完整的判決原文，讓使用者可以詳細閱讀和分析判決書的每一部分。

- **摘要撰寫欄位**

使用者可以在欄位中撰寫自己的摘要，並根據閱讀判決原文的理解進行歸納和整理。

- **正確摘要對照**

程式內建正確的摘要範例，讓使用者在完成撰寫後，可以對照範例檢查自己的摘要是否正確和完整。

目標：

- **提升閱讀能力**

通過反覆練習，使用者將能更快、更準確地抓住判決主文的重點，提升法律文書的閱讀能力

- **強化摘要技巧**

程式提供的範例，能幫助使用者逐步掌握撰寫摘要的技巧，並應用到實際工作中。

- **自我評估與成長**

讓使用者能夠隨時了解自己的進步情況，為後續的學習和提高提供方向。

Bonus (Implement of Proposal):

[下載連結](#)

HW1

法律裁判原文摘要練習小幫手

裁判原文

最高行政法院判決105年度判字第537號上訴人林朝明訴訟代理人莊守禮律師被上訴人桃園市政府環境保護局代表人沈志修訴訟代理人林宗燭律師上列當事人間廢棄物清理法事件，上訴人對於中華民國104年11月5日臺北高等行政法院104年度訴字第569號判決，提起上訴，本院判決如下：主文原判決廢棄。其中關於如附表所示原處分1中被上訴人所為103年6月3日桃環事字第1032328981號函說明第六項（即依廢棄物清理法第71條規定等之處分）部分及該部分訴願決定，發回臺北高等行政法院。其餘廢棄部分，訴願決定及原處分均撤銷。第二項廢棄部分第一審及上訴審訴訟費用由被上訴人負擔。理由一、緣被上訴人於民國103年4月25日派員會同行政院環境保護署（下稱環保署）北區環境督察大隊（下稱北區督察大隊）、環保警察及訴外人即上訴人之子林昇輝至上訴人所有坐落桃園市○○區○○段000○號等土地（下稱系爭土地，其上建物門牌號碼：桃園市○○區○○路0段000○0號）勘查，查獲該址違規露天堆置處理大量事業廢棄物。前經北區督察大隊於103年4月8日於系爭土地進行採樣分析，發現部分事業廢棄物之毒性特性溶出程序檢測結果，銅含量已超過有害事業廢棄物認定標準，屬有害事業廢棄物，其餘未超出有害事業廢棄物認定標準者，係屬一般事業廢棄物，其貯存方式均未依規定辦理，核已違反廢棄物清理法（下稱廢清法）第36條第1項暨事業廢棄物貯存清除處理方法及設施標準（下稱設施標準）第6條、第7條、第10條及第11條規定，依行政

下一題

請輸入摘要

送出

文章摘要

HW1

法律裁判原文摘要練習小幫手

裁判原文

最高行政法院判決105年度判字第537號上訴人林朝明訴訟代理人莊守禮律師被上訴人桃園市政府環境保護局代表人沈志修訴訟代理人林宗燭律師上列當事人間廢棄物清理法事件，上訴人對於中華民國104年11月5日臺北高等行政法院104年度訴字第569號判決，提起上訴，本院判決如下：主文原判決廢棄。其中關於如附表所示原處分1中被上訴人所為103年6月3日桃環事字第1032328981號函說明第六項（即依廢棄物清理法第71條規定等之處分）部分及該部分訴願決定，發回臺北高等行政法院。其餘廢棄部分，訴願決定及原處分均撤銷。第二項廢棄部分第一審及上訴審訴訟費用由被上訴人負擔。理由一、緣被上訴人於民國103年4月25日派員會同行政院環境保護署（下稱環保署）北區環境督察大隊（下稱北區督察大隊）、環保警察及訴外人即上訴人之子林昇輝至上訴人所有坐落桃園市○○區○○段000○號等土地（下稱系爭土地，其上建物門牌號碼：桃園市○○區○○路0段000○0號）勘查，查獲該址違規露天堆置處理大量事業廢棄物。前經北區督察大隊於103年4月8日於系爭土地進行採樣分析，發現部分事業廢棄物之毒性特性溶出程序檢測結果，銅含量已超過有害事業廢棄物認定標準，屬有害事業廢棄物，其餘未超出有害事業廢棄物認定標準者，係屬一般事業廢棄物，其貯存方式均未依規定辦理，核已違反廢棄物清理法（下稱廢清法）第36條第1項暨事業廢棄物貯存清除處理方法及設施標準（下稱設施標準）第6條、第7條、第10條及第11條規定，依行政

下一題

請輸入摘要

在最高行政法院105年度判字第537號判決中，法院審理林朝明上訴桃園市政府環保局的案件，做出了具有重要意義的決定。林朝明因不服桃園市政府環保局的處分，經訴願後向臺北高等行政法院提起行政訴訟，該法院判決駁回林朝明的訴訟請求。林朝明不服，再次向最高行政法院提出上訴。在本案中，最高行政法院仔細審查了相關事實與法律適用，認為臺北高等行政法院在審理過程中，對於事實的認定及法律的適用有欠妥之處，尤其是在環保法規的適用上存在爭議。最高行政法院指出，原審法院未能全面考量林朝明所提的證據和法律觀點，從而影響了裁判的公正性。最高行政法院在本案判決中，首先決定廢棄臺北高等行政法院的原判決，這意味著原

送出

文章摘要

HW1

—

○

×

法律裁判原文摘要練習小幫手

裁判原文

最高行政法院判決105年度判字第537號上訴人林朝明訴訟代理人莊守禮律師被上訴人桃園市政府環境保護局代表人沈志修訴訟代理人林宗竭律師上列當事人間廢棄物清理法事件，上訴人對於中華民國104年11月5日臺北高等行政法院104年度訴字第569號判決，提起上訴，本院判決如下：主文原判決廢棄。其中關於如附表所示原處分1中被上訴人所為103年6月3日桃環事字第1032328981號函說明第六項（即依廢棄物清理法第71條規定等之處分）部分及該部分訴願決定，發回臺北高等行政法院。其餘廢棄部分，訴願決定及原處分均撤銷。第二項廢棄部分第一審及上訴審訴訟費用由被上訴人負擔。理由一、緣被上訴人於民國103年4月25日派員會同行政院環境保護署（下稱環保署）北區環境督察大隊（下稱北區督察大隊）、環保警察及訴外人即上訴人之子林昇輝至上訴人所有坐落桃園市○○區○○段000○號等土地（下稱系爭土地，其上建物門牌號碼：桃園市○○區○○路0段000○0號）勘查，查獲該址違規露天堆置處理大量事業廢棄物。前經北區督察大隊於103年4月8日於系爭土地進行採樣分析，發現部分事業廢棄物之毒性特性溶出程序檢測結果，銅含量已超過有害事業廢棄物認定標準，屬有害事業廢棄物，其餘未超出有害事業廢棄物認定標準者，係屬一般事業廢棄物，其貯存方式均未依規定辦理，核已違反廢棄物清理法（下稱廢清法）第36條第1項暨事業廢棄物貯存清除處理方法及設施標準（下稱設施標準）第6條、第7條、第10條及第11條規定，依行政

下一題

請輸入摘要

在最高行政法院105年度判字第537號判決中，法院審理林朝明上訴桃園市政府環保局的案件，做出了具有重要意義的決定。林朝明因不服桃園市政府環保局的處分，經訴願後向臺北高等行政法院提起行政訴訟，該法院判決駁回林朝明的訴訟請求。林朝明不服，再次向最高行政法院提出上訴。在本案中，最高行政法院仔細審查了相關事實與法律適用，認為臺北高等行政法院在審理過程中，對於事實的認定及法律的適用有欠妥之處，尤其是在環保法規的適用上存在爭議。最高行政法院指出，原審法院未能全面考量林朝明所提的證據和法律觀點，從而影響了裁判的公正性。最高行政法院在本案判決中，首先決定廢棄臺北高等行政法院的原判決，這意味著原

送出

文章摘要

一按合法行政處分，係指由有管轄之行政機關，得對以行政處分為規制之事項，以正確之程序及依規定之方式作成，在內容上亦無瑕疵之行政處分。缺一要件，均屬不法。二行為時廢棄物清理法第七條第一項：不依同法第九四條第二項第二款規定清除處理之廢物，直轄市縣主管機關或執行機關得命事業有重大改變，但並非完全不能履行其職責義務者，亦同普通行政處分以特定人為相對人，行政法院並無裁量權限，則基於依法行政及規範有利原則，應認系違法行政處分於法理上雖有定位為違法事由，然依我國行政程序法第一一條第二項規定，仍應適用依行政訴訟法第二二條規定：當事人有下列情形之一，不得執行職務：五曾參與該訴訟事件之前審裁判九九年五月一日修正公佈，同年一月二十七日施行，及一年六月一日施行之修正發布，一年七

目前的成果仍有可進步的地方，我們希望未來有幾點可改善與拓展：

- 多語言支援

擴展到多語言版本，支援更多國家的法律專業人士，提高全球法律從業者的學習效果。

- 資料庫擴展

增加更多不同類型和難度的判決案例，涵蓋更多法律領域，讓使用者能夠更全面地練習和提升。

- 人工智慧技術應用

結合預訓練過的大語言模型（如 LLaMA、ChatGPT），給出更準確流暢的摘要答案。

- 使用者介面優化

簡化操作流程，提升介面友好性，讓使用者在使用程式時更加便捷和舒適。

- 跨平台兼容性

開發針對不同操作系統和設備的版本，如手機應用程式和網頁版，讓使用者可以在不同的設備上方便地使用。

我們期待能夠不斷提升應用程式的實用性和使用者體驗，並為法律相關人員提供更好的學習和成長平台。