

# EBSUM: 基於 BERT 的強健性抽取式摘要法

## EBSUM: An Enhanced BERT-based Extractive Summarization Framework

吳政育 Zheng-Yu Wu

國立臺灣科技大學資訊工程系

Department of Computer Science and Information Engineering

National Taiwan University of Science and Technology

M10615079@mail.ntust.edu.tw

陳冠宇 Kuan-Yu Chen

國立臺灣科技大學資訊工程系

Department of Computer Science and Information Engineering

National Taiwan University of Science and Technology

kychen@mail.ntust.edu.tw

### 摘要

目前大部分自動摘要方法，分為抽取式摘要(Extractive)與重寫式摘要(Abstractive)，重寫式摘要雖然能夠改寫文章形成摘要，但這並不是一種有效的方式，困難點在於語意不通順、重複字等。抽取式摘要則是從文章中抽取句子形成摘要，能夠避免掉語意不通順，重複字的缺點。目前基於 BERT(Bidirectional encoder representation from transformers)的抽取式摘要法，多半是利用 BERT 取得句子表示法後，再微調模型進行摘要句子之選取。在本文中，我們提出一套新穎的基於 BERT 之強健性抽取式摘要法(Enhanced BERT-based Extractive Summarization Framework, EBSUM)，它不僅考慮了句子的位置資訊、利用強化學習增強摘要模型與評估標準的關聯性，更直接的將最大邊緣相關性(Maximal Marginal Relevance, MMR)概念融入摘要模型之中，以避免冗餘資訊的選取。在實驗中，EBSUM 在公認的摘要資料集 CNN/DailyMail 中，獲得相當優良的任務成效，與經典的各式基於類神經網路的摘要模型相比，EBSUM 同樣可以獲得最佳的摘要結果。

關鍵詞：自動摘要，抽取式，BERT，強化學習，最大邊緣相關性

## Abstract

Automatic summarization methods can be classified into two major spectrums: extractive summarization and abstractive summarization. Although abstractive summarization methods can produce a summary by using different words and phrases that were not in the given document, it usually leads to an influence summary. On the contrary, extractive summarization methods generate a summary by copying and concatenating the most important sentences in the given document, which usually can provide a more continuous and readable summary. Recently, BERT (Bidirectional encoder representation from transformers)-based extractive summarization methods usually leverage BERT to encode each sentence into a fixed low-dimensional vector, and then a fine-tuned BERT model can be used to predict a score for each sentence. On top of the predicted scores, we can rank these sentences and form a summary for a given document. In this paper, we propose an enhanced BERT-based extractive summarization framework (EBSUM), which not only takes sentence position information and RL training into account, but the maximal marginal relevance (MMR) criterion is also be considered. In our experiments, we evaluate the proposed framework on the CNN/DailyMail benchmark corpus, and the results demonstrate that EBSUM can achieve a better result than other classic extractive summarization methods.

**Keywords:** Extractive, Summarization, BERT, MMR.

致謝

This work is supported by the Ministry of Science and Technology (MOST) in Taiwan under grant MOST 108-2636-E-011-005 (Young Scholar Fellowship Program), and by the Project J367B83100 (ITRI) under the sponsorship of the Ministry of Economic Affairs, Taiwan.