

20240718Meeting

GENERATE RATHER THAN RETRIEVE: LARGE LANGUAGE MODELS
ARE STRONG CONTEXT GENERATORS

Introduction

- Retrieve-then-read
 - 從大型資料庫中檢索相關文檔，然後閱讀這些文檔以回答問題
- Generate-then-read(GENREAD)
 - 利用大型語言模型（InstructGPT）來生成問題的上下文文檔

GENREAD

- two main advantages
 - generated contextual documents contain the correct answer more often than the top retrieved documents
 - outperforms directly generating answers from large language models
- recall performance does not scale as well with generated documents
 - clustering-based prompt method

GENREAD

1. 生成文檔（Generate）

- 通過大型語言模型（如InstructGPT）生成基於給定問題的上下文文檔

2. 閱讀文檔（Read）

- 使用生成的文檔和問題來預測最終答案
- 零樣本設定（Zero-shot Setting）
 - 在沒有訓練數據的情況下，使用生成-閱讀管道可以提升大模型的回答精度
- 監督設定（Supervised Setting）
 - 在有訓練數據的情況下，小型閱讀模型（如FiD）可以通過生成的文檔進行微調，以提高性能