TabRAG: Efficient Table Retrieval and Understanding with Large Multimodal Models

Zhuoyan Xu[1,2], Haoyang Fang[2], Boran Han[2], Bonan Min[2], Bernie Wang[2], Shuai Zhang[2]

[1]University of Wisconsin-Madison
[2]Amazon Web Services

October 2, 2024

Abstract

Table understanding problem presents a substantial challenge since it requires complex and holistic understanding to the table. Substantial research has been dedicated to exploring the relevant fields based on large language models (LLMs). However, in real world scenario, tables are often stored in a large data store, and given a general query requiring specific information, the relevant table may not be readily available by hand. Addressing a general user query typically involves two steps: first, retrieving the relevant table, and then solving the table understanding task. Current retrieval tasks in Retrieval-Augmented Generation (RAG) systems and table understanding frameworks typically focus on text modality alone, often disregarding the visual structure of tables. On the other hand, traditional Optical Character Recognition (OCR) methods for converting images to texts suffer from high computational cost and prone to errors, leading to inefficiencies in table understanding tasks. In this work, we introduce TabRAG, a novel framework that addresses retrieved table understanding challenges by directly utilizing table images in both retrieval and generation step. We first employ a RAG pipeline to retrieve relevant table images from extensive collections based on user queries. We then leverage the reasoning capabilities of large multimodal models (LMMs) to generate accurate responses. TabRAG enhances the effectiveness and real-world applicability of LMMs when dealing with tabular data, offering a more intuitive and effective solution for RAG on table reasoning tasks. Experimental validation is conducted using a newly constructed table image dataset from several public table understanding dataset, demonstrating the robustness and efficiency of our proposed framework.