NTCIR 17 Conference: Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies, December 12-15, 2023, Tokyo, Japan
DOI: https://doi.org/10.20736/0002001310

Embedding Tables in Text Context: A New Approach to UFO Tasks

Hiroyuki Higa
Osaka Electro-Communication University
Japan
gp20a125@oecu.jp

Yuuki Maeyama
Osaka Electro-Communication University
Japan
gp20a140@oecu.jp

Keisuke Nanjo
Osaka Electro-Communication University
Japan
mi22a008@oecu.jp

Kazuhiro Takeuchi
Osaka Electro-Communication University
Japan
takeuchi@osakac.ac.jp

ABSTRACT

Text-to-Table Relationship Extraction (TTRE)[3] has emerged as a significant research topic. Although tables enable humans to comprehend complex data structures quickly, machines often struggle with such interpretations. The primary challenge of this paper lies in understanding the myriad intentions behind the table's creation and the possible ambiguity when viewed without context. We propose an approach to address these issues by embedding a table in a textual context. Specifically, we convert tables contained in HTML-formatted documents to the Markdown format and create training data that combine the tables with information about the associated question text and elements. Then, we use the training data to train a QLoRA model based on llama2-13b-chat-hf. This approach promotes holistic interpretation of tables and their associated texts within a single vector space.

KEYWORDS

table embedding, table understanding,LLM

TEAM NAME

tkl2023

SUBTASKS

UFO Text-to-Table Relationship Extraction (Japanese)

1 INTRODUCTION

In today ' s intricate business landscape, corporations continually process and assess vast quantities of data, with sources such as securities reports being pivotal. Securities reports contain a variety of tables, and these tables serve as an efficient medium for humans to swiftly grasp intricate data configurations and nuances. Their efficacy is accentuated in contexts that demand swift data comparisons, pattern discernment, and immediate detail absorption. Although the design and intent of a table are almost instinctive for humans to understand, machines find this endeavor challenging. Within this context, Text-to-Table Relationship Extraction (TTRE) has attracted significant attention.

In the quest to equip computers with the ability to interpret tables adeptly as humans, we are confronted with two main hurdles.

1) The potential multiplicity of underlying motives the author seeks to impart, and
2) The inherent ambiguity of a table's objective is based on its standalone presentation.

To overcome these impediments, our study introduces a novel methodology: positioning the subject table within a textual embedding realm. Our research methodology was divided into two pivotal stages.

- The table is transformed into a textual descriptor.
- Subsequently, this textual avatar was assimilated into a vector space, leveraging a pre-existing large language model.

Consequently, this facilitates the synergistic treatment of both tables and accompanying text within a unified vector space.

2 RELATED WORK

In the endeavor to enable computers to understand tables as effectively as humans, some previous research like [4] approaches have treated tables as images. Specifically, [4] pointed out the problem of using metadata on PDFs to detect tables, and proposed a method to recognize tables in a metadata-independent manner through image recognition. In other words, that method uses image recognition to extract the location information of tables, but does not learn about the relationship with the text.

In this paper, we propose a new method that treats text and tables as a single entity. Our proposed method does not require the construction of a dedicated architecture, but only additional learning on existing decoder models. The model used for this training is llama2, which is provided by META and trained on open data with a publicly available structure [6][5]. The training method also uses qlora, which is designed to allow large models to be computed on non-specialized personal computer resources [2].

3 METHODS

To treat both tables and text as unified entities, we undergo a process of converting all tables within securities reports into markdown format, using ChatGPT[1]. This conversion allows us to regard table cells as an integral part of the textual context, seamlessly blending them.

To bridge the gap between table values and the surrounding text descriptions, we utilize a versatile question-answering model, specifically the "llama2-13b-hf-chat" model provided by Meta, as our foundation. To associate the value of a cell with its corresponding description, we test whether that model predicts a particular value of the cell as an answer to a textual description presented as a question. With the qrola learning method, we tune the model that extracts the intricate relationships between textual content and tabular data.

For our tuning data's "answer" segment, we have implemented a method that involves splitting each cell value using a distinct character. This process serves as a key in the process of extracting relations.

The data in Figure1 was for tuning.

```
question:###“What is relevant to cost of sales?###

context:###

| Item                        | Amount (in thousand yen) |
|--------------------------------|--------------------------|
| Sales                          | 10,000                   |
| Cost of Goods Sold             | -5,000                   |
| Gross Profit                   | 5,000                    |
| Operating Profit               | 2,000                    |
| Financial Expenses             | -100                     |
| Pre-tax Net Profit for the Term | 1,900                   |
| Corporate Taxes, etc.          | -760                     |
| Net Profit for the Term        | 1,140                    |

        ###

answer:###”Cost of Goods Sold ”### ### -5,000 ###
```

Figure 1: TuningData

The "question" segment of the training data was formatted so that the annotated sentence was formatted to follow the phrase "what is relevant to." The "context" segment was truncated if it exceeded the upper limit of tokens for the table above and below the "question" segment. For the "answer" segment, a technique was employed in which each value was divided by a distinct character. The algorithm uses these letters to parse and extract cell values from the output. The same format of questions is used for inference, as shown in Figure2, where the answer: is truncated below.

.

4 EXPERIMENTS

Parameter Name	Value
name	0.0502
value	0.0404
total	0.0453

Table 1: Summary of experimental results

The results were not as good as expected, especially for "value". This suggests that there is room for further improvement in methodology and approach.

```
question:###“What is relevant to cost of sales?###

context:###

| Item                        | Amount (in thousand yen) |
|--------------------------------|--------------------------|
| Sales                          | 10,000                   |
| Cost of Goods Sold             | -5,000                   |
| Gross Profit                   | 5,000                    |
| Operating Profit               | 2,000                    |
| Financial Expenses             | -100                     |
| Pre-tax Net Profit for the Term | 1,900                   |
| Corporate Taxes, etc.          | -760                     |
| Net Profit for the Term        | 1,140                    |

        ###

answer:
```

Figure 2: InferenceData

5 CONCLUSIONS

The results indicate that there are significant opportunities for improvement in the methods employed. One critical factor influencing the results is the accuracy of converting tables from HTML to markdown, an underlying assumption of our experiment. Moreover, the accuracy of parsing table cell values from outputs generated by the language model can also be identified as an area of concern.

Future research will concentrate on implementing these improvements to enhance the method's accuracy and reliability. Specifically, we aim to refine the conversion and parsing processes by optimizing the conversion algorithm and enhancing the quality of the training data for the language model. Furthermore, we plan to assess the method's adaptability by testing it on diverse datasets and in real-world situations.

REFERENCES

[1] [n. d.]. introducing ChatGPT. https://openai.com/blog/chatgpt

[2] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. QLoRA: Efficient Finetuning of Quantized LLMs. *ArXiv* abs/2305.14314 (2023). https://api.semanticscholar.org/CorpusID:258841328

[3] Yasutomo Kimura, Hokuto Ototake, Kazuma Kadowaki, Takahito Kondo, and Makoto P. Kato. 2023. Overview of the NTCIR-17 UFO Task. *Proceedings of The 17th NTCIR Conference* (12 2023).

[4] Sebastian Schreiber, Stefan Agne, Ivo Wolf, Andreas Dengel, and Sheraz Ahmed. 2017. DeepDeSRT: Deep Learning for Detection and Structure Recognition of Tables in Document Images. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Vol. 01. 1162–1167. https://doi.org/10.1109/ICDAR.2017.192

[5] Hugo Touvron, Thibaut Lavril, and et al. 2023. LLaMA: Open and Efficient Foundation Language Models. *ArXiv* abs/2302.13971 (2023). https://api.semanticscholar.org/CorpusID:257219404

[6] Hugo Touvron, Louis Martin, and et al. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. *ArXiv* abs/2307.09288 (2023). https://api.semanticscholar.org/CorpusID:259950998