8/12/2024

Finance LLM

An attempt to process financial reports

Dr. Avinash Kumar Singh

Table of Contents

Problem Premise

Let's first define the premise of the problem and see what we all have. Based on this understanding, we will further define the problem statement.

We have a ConvFinQA dataset that stands for **Conv**ersational **Q**uestion **A**nswering over **Fin**ancial Reports [1]. This dataset is the extension of the FinQA dataset released in October 2022, FinQA stands for **Q**uestion **A**nswering pairs over **Fin**ancial Reports [2][9]. The objective of this dataset to study the chain of numerical reasoning in conversational question answering. While the FinQA only has the question and answers pair, the ConvFinQA has the conversation (inter mediate steps) added to have better explainability in the whole process. In Figure 1, we can see that question, we have one answer, it comes in pair, while if we see Figure 2, steps section, then we can find that the question has one answer but it has divided over intermediate steps. Here it is divided as a two-step process to calculate the final answer.

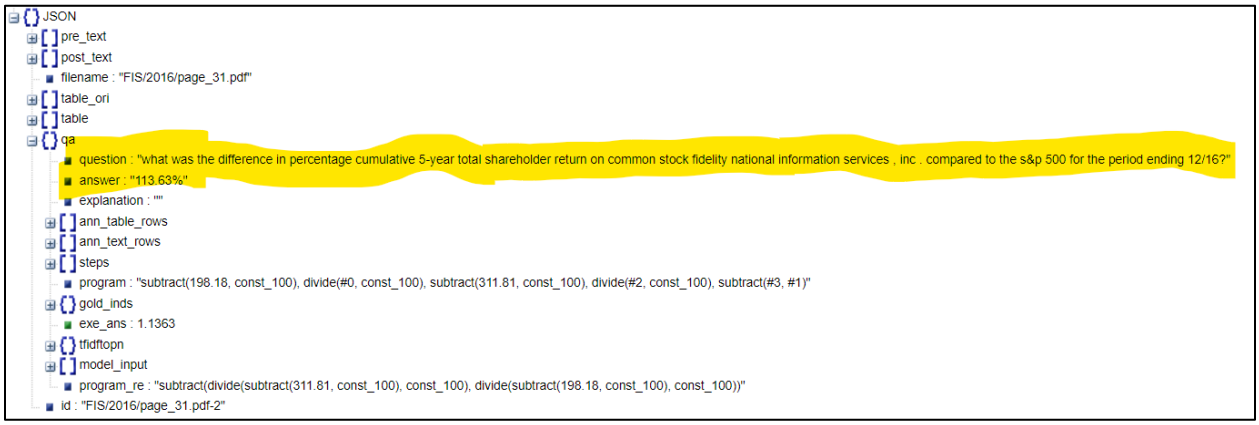


Figure 1. FinQA dataset representation [3]

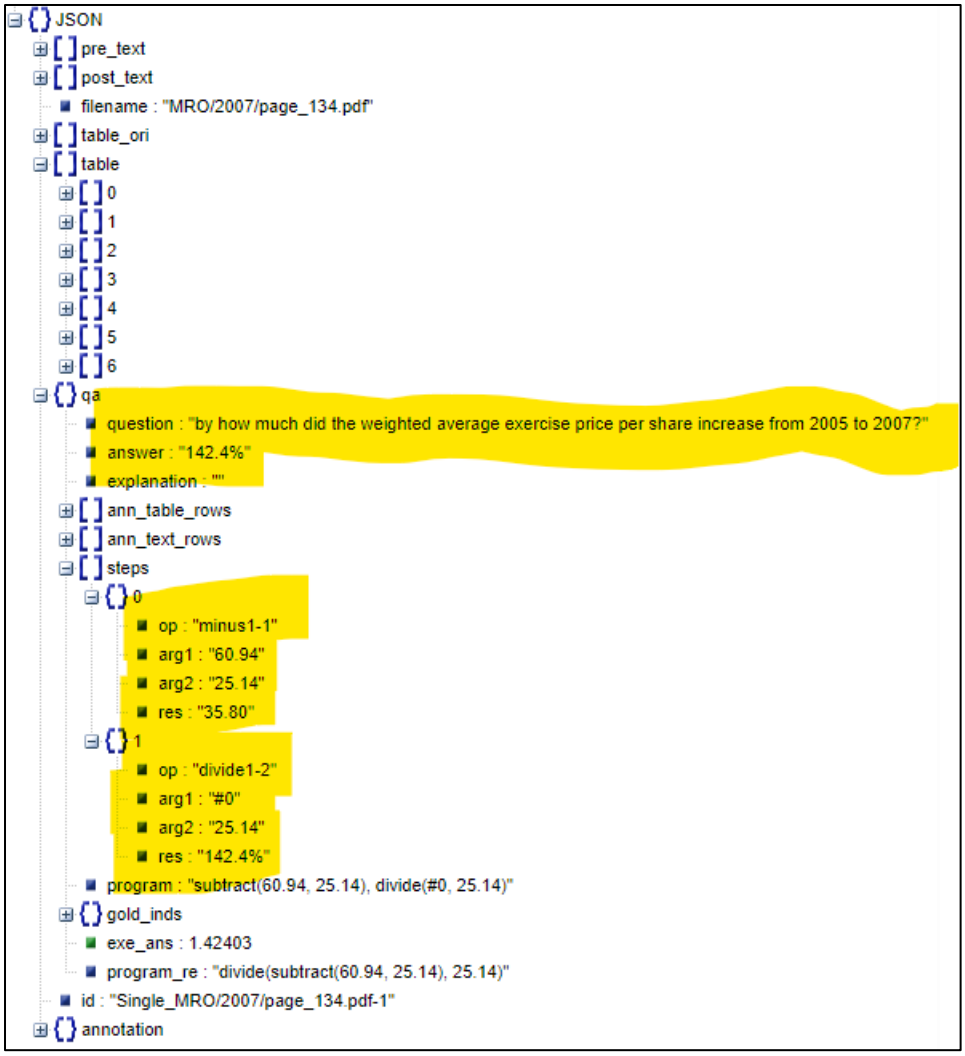


Figure 2. ConvFinQA Dataset Representation [8]

Problem Statement

Given the question, we need to find out the answer. For making the problem simple, rather generating the intermediate steps, we would predict the end outcome. The input of the program would be question, answer, look up table and program_re while the output would be the predicted expression to calculate the answer. One example of the input and output pair is explained below. [3]

Input to the system

The important part of the input is described below


```
{"qa":

{"question": "what was the difference in percentage cumulative 5-year
total shareholder return on common stock fidelity national information
services, inc. compared to the s&p 500 for the period ending 12/16?",

"answer": "113.63%",

"table": [ [ "", "12/11", "12/12", "12/13", "12/14", "12/15",
"12/16"],

          ["fidelity national information services inc .", "100.00",
```

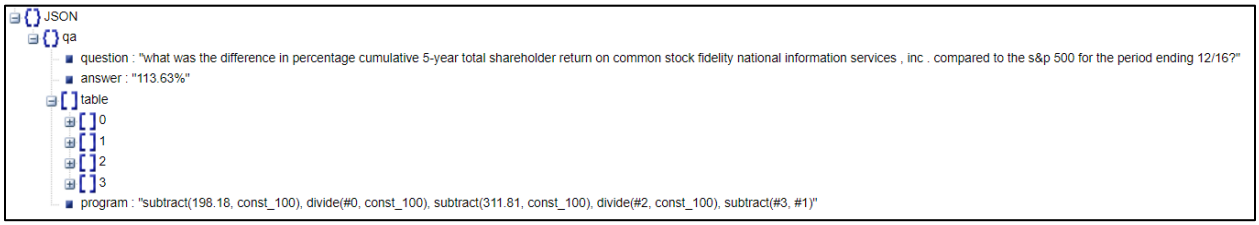


Figure 3. The JSON visualization to give better illustration of the input data

The table is presented in the below figure the same is represented in the above Json.

	12/11	12/12	12/13	12/14	12/15	12/16
Fidelity National Information Services, Inc.	**100.00**	**134.12**	**210.97**	**248.68**	**246.21**	**311.81**
S&P 500	**100.00**	**116.00**	**153.58**	**174.60**	**177.01**	**198.18**
S&P Supercap Data Processing & Outsourced Services	**100.00**	**126.06**	**194.91**	**218.05**	**247.68**	**267.14**

Figure 4. Lookup table (json representation is shown above)
URL: https://www.annualreports.com/HostedData/AnnualReportArchive/f/NYSE_FIS_2016.pdf, Page: 31 [4]

Output of the system

Output of the system should have a dictionary that would have id and the predicted expression. For the same input that shown in the previous section (input to the system), this would be the system output.

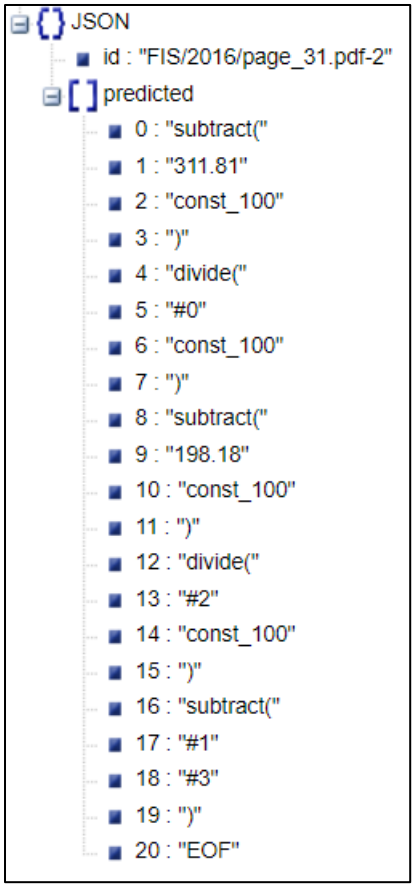


Figure 5. System prediction example.

Evaluation Metrices

There are two metrices that will be utilized to estimate the model accuracy. A good tutorial on evaluation of LLM is discussed here [5].

Program Accuracy:

Calculates the accuracy of the operator and operands between the predicted program and the golden program. Predicted program is the system prediction while golden program refers to “program”. See the above json for your reference. [6][7]

golden_Program =

Substract (198.18, Const_100),
divide (#0, Const_100),
Substract (311.81, Const_100),
divide (#2, Const_100),
Substract (#3, #1)

Predicted_Program =

["Substract (", "311.81", "Const_100", ")",
"divide (", "#0", "Const_100", ")",
"Substract (", "198.18", "Const_100", ")",
"divide (", "#2", "Const_100", ")",
"Substract (", "#1", "#3", ")", "EOF"]

We would process the predicted program to get the expression out of it.

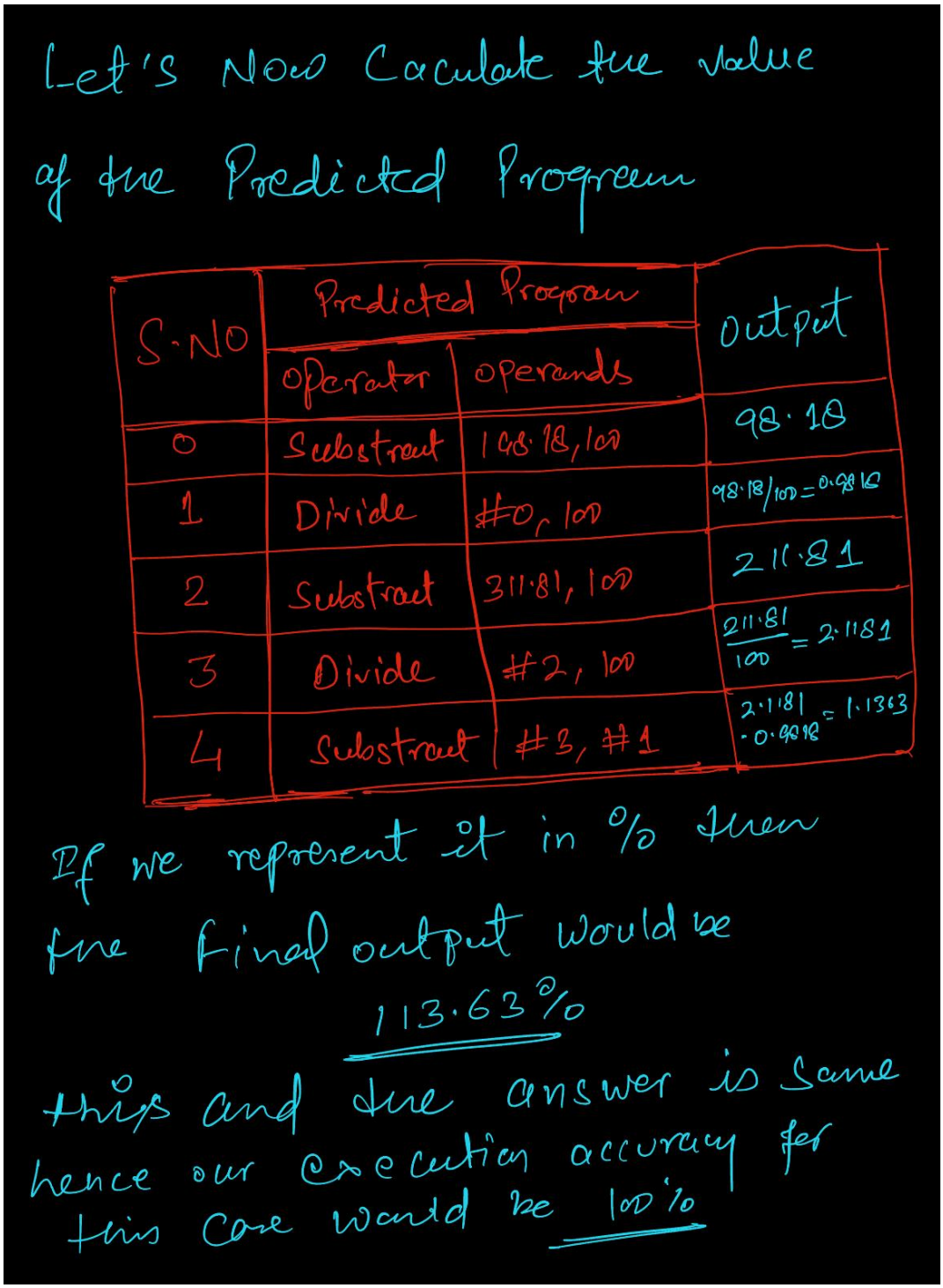
After Pre-Processing the final predicted Program would look like the same we need to also remove Eof to have the length same as of the golden Program.

	Golden Program		Predicted Program	
S.No	operator	operands	operator	operands
1	Substract	198.18, 100	Substract	198.18, 100
2	Divide	#0, 100	Divide	#0, 100
3	Substract	311.81, 100	Substract	311.81, 100
4	Divide	#2, 100	Divide	#2, 100
5	Substract	#3, #1	Substract	#3, #1

Since both tables are matching the Program accuracy will be 100%

Execution Accuracy

Evaluates the model by calculating the accuracy between the predicted program result and the golden executable results. The golden executable result is the answer of the question while the predicted program result is the output of the expression. We have described the same using the example below. [6][7]



References

[1]. Chen, Z., Li, S., Smiley, C., Ma, Z., Shah, S., & Wang, W. Y. (2022). Convfinqa: Exploring the chain of numerical reasoning in conversational finance question answering. arXiv preprint arXiv:2210.03849.

[2]. Chen, Z., Chen, W., Smiley, C., Shah, S., Borova, I., Langdon, D., ... & Wang, W. Y. (2021). Finqa: A dataset of numerical reasoning over financial data. *arXiv preprint arXiv:2109.00122*.

[3]. https://github.com/czyssrs/FinQA

[4]. Page 31, https://www.annualreports.com/HostedData/AnnualReportArchive/f/NYSE_FIS_2016.pdf

[5]. https://qa.fastforwardlabs.com/no%20answer/null%20threshold/bert/distilbert/exact%20match/f1/robust%20predictions/2020/06/09/Evaluating_BERT_on_SQuAD.html#Exact-Match.

[6]. Sun, J., Zhang, H., Lin, C., Gong, Y., Guo, J., & Duan, N. (2022). Apollo: An optimized training approach for long-form numerical reasoning. *arXiv preprint arXiv:2212.07249*.

[7]. Zhang, J., & Moshfeghi, Y. (2022). Elastic: Numerical reasoning with adaptive symbolic compiler. *Advances in Neural Information Processing Systems*, *35*, 12647-12661.

[8]. https://github.com/czyssrs/ConvFinQA

[9]. https://finqasite.github.io/index.html

[10].