DOI 10.31891/2307-5732-2024-339-4-24
УДК 004.005:004.9

DOBROVOLSKY YURIY
Chernivtsi National University named after Yu. Fedkovich
https://orcid.org/0000-0002-1248-3615
e-mail: y.dobrovolsky@chnu.edu.ua
PROKHOROV PAVLO
Chernivtsi National University named after Yu. Fedkovich
https://orcid.org/0009-0008-0965-5771
e-mail: prokhorov.pavlo@chnu.edu.ua
PROKHOROV MYKOLA
Chernivtsi National University named after Yu. Fedkovich
https://orcid.org/0000-0001-5787-2676
e-mail: prokhorov.mykola@chnu.edu.ua
PAVLIUCHENKO OLEH
Chernivtsi National University named after Yu. Fedkovich
e-mail: oleh.pavl@gmail.com
MAKARUK OLEG
Chernivtsi National University named after Yu. Fedkovich
e-mail: makaruk.oleh@chnu.edu.ua

DEVELOPMENT OF RELIABLE QUESTION-ANSWERING SYSTEM FOR FINANCIAL ACCOUNTING STANDARDS STUDY AND ANALYSIS: CONCEPTS AND CHALLENGES

Current paper provides challenges and concepts that emerge during development of ChatGPT-like question-answering system for analyze and study financial accounting standards (such as IFRS, US GAAP, UK GAAP, etc.) differences and usage, including issues of compliance between different types of financial reporting in different standards, which called FinancialStandardTableGPT (or FST-GPT). Issues of processing data in various types and formats, as well as various cases with their sources and initial quality were considered during system development. To solve these issues, it was proposed to use solutions based on consequent usage of Deep Learning models which are able to effectively process financial reporting information from a large number of various formats because of this configuration. The paper also examines in detail the task of integrating data and commands formed on the basis of natural language and tabular data, including the issue of transforming natural language commands into a sequence of instructions over tabular data and final result forming, which is presented in the form of tabular data and an optional additional explanation, which is presented in natural language form. In addition, current paper examines the limitations of classic GPT architectures for working with tabular data and considers GPT-based solutions that allow to successfully solve this issue, their features and the specifics of forming the necessary dataset of tabular data, which is oriented towards the implementation in dialogue systems which using LLMs. As a result, proposed system is focused on processing of user queries in form of text combined with financial report tables in various formats. Architecture of the FST-GPT is based on set of Deep Learning and Machine Learning models used in succession and combined with fine-tuned domain-specific LLM.

Keywords: Financial accounting standards, Question-answering systems, LLM, GPT, Software Engineering, Software Reliability

ДОБРОВОЛЬСЬКИЙ ЮРІЙ, ПРОХОРОВ ПАВЛО,
ПРОХОРОВ МИКОЛА, ПАВЛЮЧЕНКО ОЛЕГ, МАКАРУК ОЛЕГ
Чернівецький національний університет ім. Ю. Федьковича

РОЗРОБКА НАДІЙНОЇ ДІАЛОГОВОЇ СИСТЕМИ ДЛЯ ВИВЧЕННЯ ТА АНАЛІЗУ СТАНДАРТІВ ФІНАНСОВОГО ОБЛІКУ: КОНЦЕПЦІЇ ТА ВИКЛИКИ

Поточна робота висвітлює виклики та концепції впровадження діалогових систем з використанням штучного інтелекту та Великих Мовних Моделей для задач вивчення та аналізу стандартів фінансової звітності (таких як IFRS, US GAAP, UK GAAP та інших), їх відмінностей та практик використання, у тому числі проблематика відповідності між різними видами фінансової звітності у різних стандартах. При розробці системи було розглянуто проблеми обробки звіту в різних видах та форматах, а також різні ситуації з їх джерелами та початковою якістю. Для вирішення цих задач було запропоновано використання рішень на основі Deep Learning моделей, що у результаті послідовного використання спроможні ефективно обробляти інформацію фінансової звітності з великої кількості різноманітних форматів. Також у роботі детально розглядається задача інтеграції даних та команд що сформовані на основі натуральної мови та табличної інформації, при цьому окремо розглянуто питання перетворення команд натуральної мови у послідовність інструкцій над табличною інформацією та формування кінцевого результату, що представлений у вигляді табличної інформації та опціональним додатковим поясненням у вигляді натуральної мови. Окрім того, у роботі розглядаються обмеження класичних GPT-архітектур по роботі з табличною інформацією та розглядаються засновані на GPT рішення що дозволяють успішно вирішувати цю проблему, їх особливості та специфіку формування необхідного датасету табличної інформації, що орієнований на впровадження у діалогові системи з використанням LLM. У результаті запропонована нами система, для якої обрано назву FinancialStandardTableGPT (або скорочено FST-GPT), заснована на обробці запитів користувачів у формі тексту з фінансовими таблицями у різноманітних форматах. Архітектура FST-GPT заснована на архітектурах глибинного навчання та моделях машинного навчання у комбінації з спеціально сконфігурованими та спеціалізованими LLM.

Ключові слова: Фінансова звітність, Великі Мовні Моделі, GPT, LLM, Програмна інженерія, Надійність програмного забезпечення

Introduction

Due to the development of technology and the openness of financial markets, the need for the development of uniform and common financial accounting and managerial accounting reporting standards has arisen. Even though in most of countries with a market economy compliance with internationally accepted standards (such as GAAP and IFRS) is established at the legislative level, in practice, each country in fact has its own financial reporting standards because of various economic and socio-cultural reasons. Such national standards are close to international standards to varying degrees, for example there are different GAAP implementations like US GAAP, UK GAAP, BR GAAP, etc. which have many specific differences in accounting and reporting between themselves [1, 2].

Also, progress in NLP/IR fields and vast amounts of unstructured data retrieval and processing techniques in recent years has increased significantly. This made it possible to effectively integrate applications based on these technologies into various stages of financial reporting processes [3].

Related work

Proposed work touches upon the problems of using dialogue systems based on Large Language Models to solve problems of analysis and study of financial statements within the framework of certain standards. A detailed overview of various methods and strategies for using Machine Learning and Informational Retrieval technologies is described in [3].

The issues of differences in specific implementations of financial statements within the framework of individual standards, their compliance and comparison practices were discussed in [1, 2].

Challenges and problems of constructing language models for use in economic and financial domains were discussed in papers [8, 9]. Specific work of forming a dataset for this domain are revealed in [15], development of a question-answering dialog system based on this dataset is shown in [14].

It is important to note that during solving problems of using GPT-based architectures in finance, the problem of processing tabular data arises, especially in case if it is combined with additional textual information. Development of GPT-based architectures that are capable of successfully performing such tasks is discussed in [5, 6].

Aim of the study

In current paper we propose a question-answering system that could help to study and analyze different financial accounting reports to understand various reporting standards. The system is designed to solve the following tasks:

- By given user text input answer to general questions about financial accounting and related financial reporting ("What is the purpose of the Balance Sheet?"), answer to standard-specific questions about financial reports ("How should we count the Goodwill according to US GAAP?")
- By uploaded financial report document and input text answer to standard-specific questions related to report ("Is uploaded report IFRS compliant?")
- By uploaded financial report document and input text perform general tasks on report table processing, summarization, counting, etc.

Although FST-GPT is designed to handle any kind of table in reports, it focuses mostly on standard financial reporting documents tables, such as Balance Sheet, Income Statement and Cash Flow Statement of different compliance categories. That's why this system in some cases would have worse results than known general-purpose models trained on general tabular data, such as Data-Copilot [4], TableGPT [5] and Microsoft Table-GPT [6].

Methodology

Architecture of FST-GPT (Figure 1) is based mostly on TableGPT approach [5]. Key differences are:

- Table extractor is an additional module that is designed to process many different formats of financial reports to appropriately restore, parse and extract tabular data in it.
- Instead of general-purpose Phoenix LLM [7], domain specific financial LLMs, that could be used through BloombergGPT [8] and FinGPT [9], are required to use.
- LLM fine-tuning process is based on Microsoft Table-GPT [6] approach, which requires to use table-tuning with the "(instruction, table, completion)" triples instead of the "(instruction, completion)" pairs.

Other parts of the system are similar in usage of BERT-like table encoder to extract vector representation from parsed table and generating the chain of SQL-like commands to perform on initial table to get the answer of the query in case of table manipulation tasks were given. General operation flow is as follows:

- User inputs text query and table (if required).
- Table extractor module extracts inner report table according to file format which user use to submit the report. The result of such extraction would be represented as a table in some software-defined form (e.g. Pandas DataFrame object for Python programming language).
- Table Encoder transforms table to vector representation.
- LLM process vector representation coupled with text input to generate answer and set of commands to perform on extracted table to get generated table (if required).

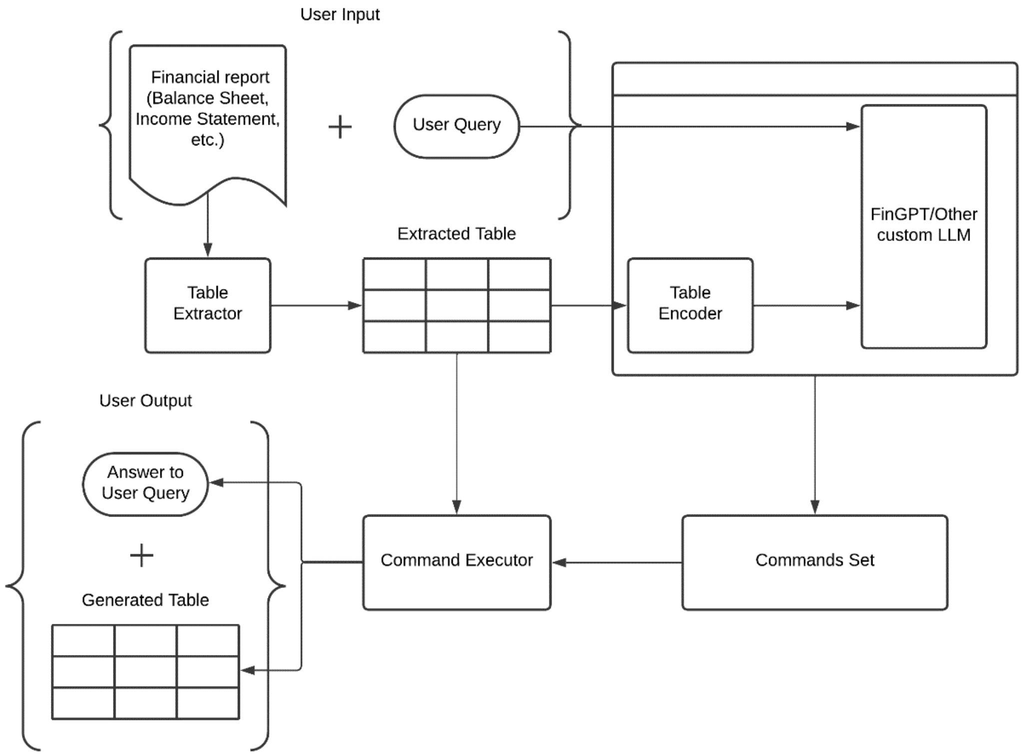

Fig. 1. Architecture of FST-GPT system

Table Extractor module (Figure 2) plays a very important role because of large number of possible financial report digital formats. We could divide all possible input report formats into 3 groups:

1. Structured tables (Figure 3a) – report table could be relatively easy parsed and processed by table formatting/processing scripts written in any general-purpose programming language using appropriate libraries (CSV, XML, XBRL, JSON, Spreadsheet files formats).
2. Weakly structured tables (Figure 3b) – report table doesn't appear in strict table format with explicit margins. Although such tables could be easily recognized and processed by humans, additional processing to extract content of the table is required. For this purpose, we use TSR-Net framework [10], which is based on fully convolution network and graph neural networks, to recognize table structure and perform form parsing to extract input report table. Possible file formats for this group include PDF and all plain text file formats.
3. Distorted structured, weakly structured or unstructured tables (Figure 3c) – are geometrically distorted structured and weakly structured report tables that were made of digital scans/photos, which require table structure restoration before any processing. For this case the TSRFormer framework is used [11]. Results of restored table are processed by TSR-Net framework from previous group. File formats of this group related to digital image files such PNG, JPEG, RAW, etc.

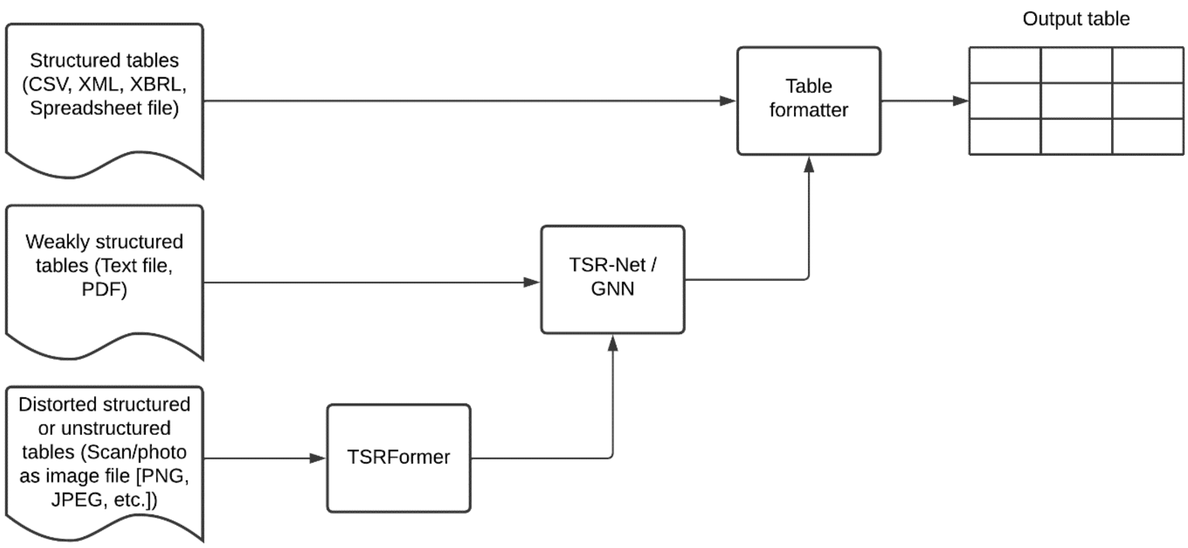

Fig. 2. Table Extractor Structure

Assets			2020	2019
Current assets				
		Cash	6.756	
		Accounts receivable	5.614	
		Inventory	42.783	
		Prepaid expenses	4.511	
		Short-term investments		
		Total current assets	59.664	

Fig. 3a. Structured table example

CONSOLIDATED STATEMENTS OF OPERATIONS
(In millions, except number of shares which are reflected in thousands and per share amounts)

	Years ended	
	September 29, 2018	September 30, 2017
Net sales	$ 265,595	$ 229,234
Cost of sales	163,756	141,048
Gross margin	101,839	88,186
Operating expenses:		
Research and development	14,236	11,581
Selling, general and administrative	16,705	15,261
Total operating expenses	30,941	26,842

Fig. 3b. Weakly structured table example [16]

(millions of dollars)

	Operating Leases	Capital Leases
2000	$ 134	$ 9
2001	93	9
2002	416	8
2003	50	7
2004	54	7
After 2004	315	14
Total lease commitments	$1,062	$54
Less interest		(8)
Present value of total capital lease obligations		$46

Fig. 3c. Distorted weakly structured table example

Because the use cases of proposed systems are limited to Financial Accounting domain and topics closely related to it, we decided to use financial domain specific LLM in for FST-GPT. At the moment of writing this paper only two production ready financial LLMs were available for use: BloombergGPT and FinGPT.

BloombergGPT is a 50 billion parameter LLM, that was trained on mixed domain specific and general data corpus of over 700 billion tokens. Despite its size and proven effectiveness this LLM is not open for academics use, so it could be added to system as primary only in case of commercial usage of proposed model.

FinGPT is an open-source and free for academics' usage framework for financial LLMs that provides different layers of usage, that makes it easy to integrate in our system with its layers APIs. Version v3.3 is based on llama2-13b model [12], which makes it efficient for declared tasks.

In case of available computational resources there is an option to create own LLM from scratch. According to FST-GPT architecture it could be also easily integrated, but to be efficient enough for declared tasks train data should contain datasets:

- FinTabNet – dataset that includes complex table from S&P 500 companies with detailed table structure annotations [13].

- ConvFinQA – large-scale dataset that models long-range, complex numerical reasoning paths in real-world conversations [14].
- FinQA – large-scale dataset with 2.8k financial reports for 8k Q&A pairs to study numerical reasoning with structured and unstructured evidence [15].

Since the vast majority of modern LLMs, by standard settings, are not suitable for processing complex table processing tasks (by standard configuration), in our system we propose to use the “table-tuning” approach [6] to fine-tune selected LLM. For that purpose, we propose to create a special dataset for fine-tuning process, that contains data in a form of (instruction, table, completion) triplets, grouped by different tasks, such as “question answering by just text input”, “question answering by text input and uploaded balance sheet report”, “performing balance sheet report data transformation” and other (Figure 4). Using this approach gives us the following advantages:

- The ability to balance the system evenly across all possible types of user input
- The ability to configure the table encoder in an optimal way for various document chunking approaches, for example in the Balance Sheet table we can logically separate parts associated with assets and liabilities, which allows us to isolate these parts into separate entities and even encode them with separate vectors, which will increase the performance of our model on specific tasks related to such report type.

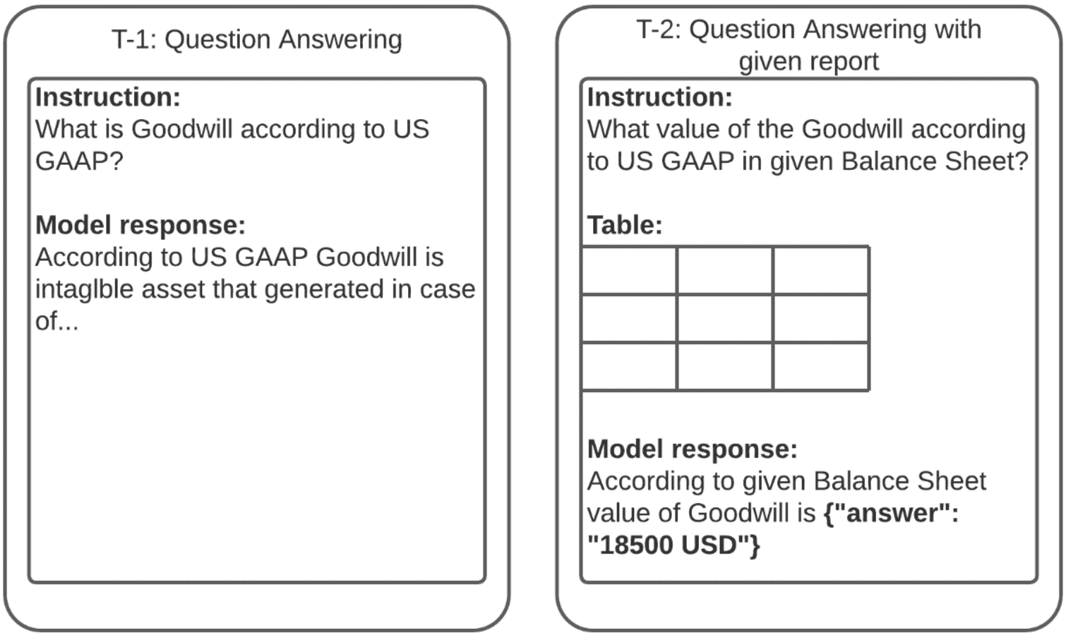

Fig. 4. Examples of dataset entries in the (instruction, table, completion) form

Conclusion

Proposed question answering system is designed to help with study and analysis of financial accounting reports and enables variety of additional options like report table manipulating, summarizing, making predictions based and other NLP tasks available for tabular data. Implemented Table Extractor module makes it robust and accessible to almost any type of possible report format, which is important because in educational practice good and refined datasets are not always available.

In our future work we’re going to provide detailed benchmarks of implemented system and present additional actions that make it possible to use this system using languages different from English using terms not fully related to traditional English GAAP implementations.

References

1. Alan Reinstein, Thomas R. Weirich. US and UK GAAP: Important differences for financial statement preparers and users. January 2002. Managerial Finance 28(1):59-72. DOI: 10.1108/03074350210767645

2. Sirlei Lemes, L. Nelson G. Carvalho. Comparability between BR GAAP and U.S. GAAP results: evidence from Brazilian companies listed on North American stock exchanges. August 2009. Revista Contabilidade & Finanças 20(50):25-45

3. Alessio Faccia, Pythagoras Petratos. NLP And IR Applications For Financial Reporting And Non-Financial Disclosure. Framework Implementation And Roadmap For Feasible Integration With The Accounting Process. NLPIR 2022: 2022 6th International Conference on Natural Language Processing and Information Retrieval, Bangkok, Thailand, December 2022. DOI: https://doi.org/10.1145/3582768.3582796

4. Wenqi Zhang, Yongliang Shen, Weiming Lu, and Yueting Zhuang. Data-copilot: Bridging billions of data and humans with autonomous workflow. arXiv preprint: 2306.07209, 2023.

5. Liangyu Zha et. al. TableGPT: Towards Unifying Tables, Nature Language and Commands into One GPT. arXiv preprint: 2307.08674, 2023

6. Peng Li et.al Table-GPT: Table-tuned GPT for Diverse Table Tasks. arXiv preprint: 2310.09263v1, 2023

7. Zhihong Chen, Feng Jiang, Junying Chen, Tiannan Wang, Fei Yu, Guiming Chen, Hongbo Zhang, Juhao Liang, Chen Zhang, Zhiyi Zhang, et al. Phoenix: Democratizing chatgpt across languages. arXiv preprint: 2304.10453, 2023

8. Shijie Wu et. al. BloombergGPT: A Large Language Model for Finance. arXiv preprint: 2303.17564v3, 2023

9. Hongyang (Bruce) Yang, Xiao-Yang Liu, Christina Dan Wang. FinGPT: Open-Source Financial Large Language Models. arXiv preprint: 2306.06031v1, 2023

10. Xiao-Hui Li, Fei Yin, He-Sen Dai, Cheng-Lin Liu. Table Structure Recognition and Form Parsing by End-to-End Object Detection and Relation Parsing. Pattern Recognition Volume 132, December 2022, 108946. DOI: https://doi.org/10.1016/j.patcog.2022.108946

11. Jiawei Wang, Weihong Lin, Chixiang Ma, Mingze Li, Zheng Sun, Lei Sun, Qiang Huo. Robust table structure recognition with dynamic queries enhanced detection transformer. Pattern Recognition Volume 144, December 2023, 109817 DOI: https://doi.org/10.1016/j.patcog.2023.109817

12. Hugo Touvron et. al. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv preprint arXiv:2306.06031v1, 2023

13. Xinyi Zheng, Douglas Burdick, Lucian Popa, Xu Zhong, Nancy Xin Ru Wang. Global Table Extractor (GTE): A Framework for Joint Table Identification and Cell Structure Recognition Using Visual Context. Winter Conference for Applications in Computer Vision (WACV), 2021. https://developer.ibm.com/exchanges/data/all/fintabnet/

14. Zhiyu Chen, Shiyang Li, Charese Smiley, Zhiqiang Ma, Sameena Shah, and William Yang Wang. 2022. ConvFinQA: Exploring the Chain of Numerical Reasoning in Conversational Finance Question Answering. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 6279–6292, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

15. Zhiyu Chen, Wenhu Chen, Charese Smiley, Sameena Shah, Iana Borova, Dylan Langdon, Reema Moussa, Matt Beane, Ting-Hao Huang, Bryan Routledge, and William Yang Wang. 2021. FinQA: A Dataset of Numerical Reasoning over Financial Data. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 3697–3711, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

16. Financial Statement Examples https://www.wallstreetmojo.com/financial-statement-examples/