EasyChair Preprint

№ 10463

The Mask One at a Time Framework for Detecting the Relationship Between Financial Entities

Sohom Ghosh, Sachin Umrao, Chung-Chi Chen and Sudip Kumar Naskar

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

June 28, 2023

The Mask One At a Time Framework for Detecting the Relationship between Financial Entities

Sohom Ghosh
Fidelity Investments, Jadavpur University
India
sohom1ghosh@gmail.com

Sachin Umrao
Fidelity Investments
Bengaluru, India
sachin.umrao.1512@gmail.com

Chung-Chi Chen
AIST
Tokyo, Japan
c.c.chen@acm.org

Sudip Kumar Naskar
Jadavpur University
Kolkata, India
sudip.naskar@gmail.com

ABSTRACT

In the financial domain, understanding the relationship between two entities helps in understanding financial texts. In this paper, we introduce the Mask One At a Time (MOAT) framework for detecting the relationship between financial entities. Subsequently, we benchmark its performance with the existing state-of-the-art discriminative and generative Large Language Models (LLMs). We use the SEC-BERT embeddings along with the one-hot encoded vectors of the types of entities and their relation group as features. We benchmark MOAT with three such open-source LLMs, namely, Falcon, Dolly and MPT under zero-shot and few shot settings. The results prove that MOAT outperforms these LLMs.

CCS CONCEPTS

• **Information systems** → *Information retrieval*; • **Computing methodologies** → **Information extraction**; *Natural language generation*.

KEYWORDS

relation extraction, financial texts, large language models

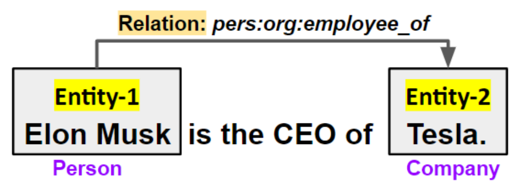

Figure 1: Relation between Financial Entities.

1 INTRODUCTION

Automatically determining the relationships between financial entities helps in interpreting financial information about people, organizations, transactions, market sentiment which changes over time. This contextual information can help analysts, investors and others make better informed decisions. Moreover, it can provide valuable insights for a range of stakeholders, from financial institutions to regulators and companies. These insights can also help in risk management, complying with regulations, and so on. For example, as presented in Figure 1, *"Elon Musk is the CEO of Tesla."*, "Elon Musk" and "Tesla" are two entities of type person and company respectively. The relation between these two entities is *"employment"*. Although this appears to be simple, in real scenarios relations between are difficult to comprehend. For instance, a person who is a board member need not necessarily be an employee. Moreover, understanding relations between entities helps in the construction of Knowledge Graphs, which have a variety of use cases in the financial industry ranging from fraud detection to stock market prediction. Financial documents tend to be long, with lots of complex relations between entities. Reading and understanding these documents is a tedious task. Thus, an automated system for extracting insights relating to relations between entities present in these documents help. In this paper, we introduce the Mask One At a Time (MOAT) framework for automatically detecting the relationship between financial entities. Subsequently, we benchmark its performance with existing open source generative Large Language Models (LLMs). Our codebase can be accessed from here.[1]

2 RELATED WORKS

Relation extraction is a well-studied field in natural language processing (NLP). It focuses on identifying the semantic relationships between entities in text. In the financial domain, relation extraction has been used to extract valuable insights from various financial documents, like news articles, financial reports, etc.

The task 8 of SemEval-2010 [4] was based on extracting 19 relations from 10717 instances. With the advent of transformers [9], researchers have started fine-tuning BERT [3] like architectures for the task of relation extraction [2]. Sharma et al. [8] released the FinRED dataset created from financial news and earning call transcripts. It comprises 29 relations and, 6767 instances.

Recently, Kaur et al. [5] released the REFinD dataset, consisting of more than 29 thousand instances involving 22 types of relations. They released few baseline models, among which Luke Large [11] fine-tuned using the Matching the Blanks architecture [2] performed the best. With the onset of LLMs [10, 12], researchers have started exploring their applications for relation extraction.

[1]https://github.com/sohomghosh/REFinD

Table 1: Distribution of categories of relation.

categories	# train	# valid	# test
no_relation	9128	1965	1953
pers:title:title	3126	671	671
org:gpe:operations_in	2832	606	605
pers:org:employee_of	1733	372	374
org:org:agreement_with	653	141	141
org:date:formed_on	448	96	96
pers:org:member_of	441	94	95
org:org:subsidiary_of	386	82	83
org:org:shares_of	286	61	61
org:money:revenue_of	217	47	47
org:money:loss_of	141	30	31
org:gpe:headquartered_in	135	29	29
org:date:acquired_on	134	28	24
pers:org:founder_of	92	19	20
org:gpe:formed_in	81	17	17
org:org:acquired_by	55	11	12
pers:univ:employee_of	53	11	12
pers:gov_agy:member_of	40	8	8
pers:univ:attended	30	6	7
pers:univ:member_of	23	5	5
org:money:profit_of	20	4	5
org:money:cost_of	16	3	4
TOTAL	**20,070**	**4,306**	**4,300**

3 PROBLEM STATEMENT

Given a financial entity pair $(e1, e2)$, our aim is to classify the relation between them into one of the categories mentioned in Table 1. For a given instance, $e1$ and $e2$ are subject and object, respectively.ORG, PER, UNI, GPE refers to organization, person, university and geopolitical entities respectively.

$e1 \epsilon \{ORG, PER\}$

$e2 \epsilon \{ORG, TITLE, UNI, GOV.AGENCY, DATE, GPE, MONEY\}$

$(e1, e2) \epsilon \{PER{-}TITLE, PER{-}ORG, PER{-}UNI, PER{-}GOV.AGENCY, ORG - GPE, ORG - DATE, ORG - ORG, ORG - MONEY\}$

4 DATASET

We use the Relation Extraction Financial Dataset (REFinD) [5] for our analysis. It comprised 29,000 instances with 22 types of relations among 8 types of entity pairs. The detailed distribution of categories across the training, validation and test set is presented in Table 1. Subsequently, as the evaluation of generative LLMs was compute-intensive, we created a smaller test set comprising 500 instances randomly selected from the test set for benchmarking these LLMs. We use weighted average Precision (P), Recall (R) and F1-score (F1) for evaluation.

5 SYSTEM DESCRIPTION

We present the architecture of MOAT in Figure 2. For a given text having two entities, we mask one entity at a time and extract the contextual SEC-BERT [6] based embeddings of the [MASK] tokens separately. We use these embeddings along with the one-hot encoded vectors of the types of entities and their relation group as features. A relation group is the concatenation of entity types, i.e. if enity-1 is *PERSON* and entity-2 is *TITLE*, the relation group is *PERSON-TITLE*. Using these features, we train a neural network (NN) for 300 epochs with two hidden layers having 512 neurons and 128 neurons respectively. The NN classifies the relationship between the entities into one of the categories mentioned in Table 1.

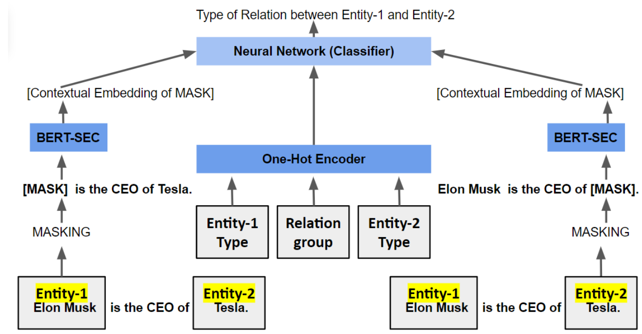

Figure 2: Architecture of MOAT.

6 EXPERIMENTS AND RESULTS

We firstly masked the entities. We extracted the contextual embeddings of these masks, concatenated them, and fine-tuned a SEC-BERT [6] for the task of classification. The performance was poor. We removed the masks, froze the underlying BERT model, and calculated the mean of the last hidden layer's embeddings of tokens constituting the entities. We concatenated these mean embeddings ($EMB_{E1,\ E2}$) and passed them through a feed forward neural network (NN). Few instances had more than 512 tokens. In those instances, we consider only 512 tokens around the entities, as the context length of SEC-BERT is 512 tokens only. This improved the performance steeply. Finally, on experimenting with the proposed MOAT architecture (described in §5) we obtained the best performance. We present the results on the test set in Table 2.

We conducted an ablation study in which we experimented by removing features constructed using the relation group (i.e. $(e1, e2)$) and entity types. We further added SEC-BERT [6] based sentence embeddings [7] of the shortest dependency path (SDP) provided in the dataset and one hot encoded vectors generated from POS tags of entities as features. Subsequently, we trained separate classifiers for each relation group and experimented by replacing SEC-BERT embeddings with that of Luke [11]. Training separate classifiers for each relation group improved the precision (P), but the recall (R) and f1-score (F1) fell drastically. The ablation study is presented in Table 3.

Recently, as generative-based LLMs have outperformed traditional methods of relation extraction [10], we benchmarked the performance of MOAT with three such open-source LLMs[2], namely Falcon, Dolly, and MPT under zero shot and few shot setting. We used the smaller test set comprising 500 instances selected randomly for inference because evaluating these LLMs is computationally

[2]More details are in the Appendix

Table 2: Performance of discriminative LLMs.

Model	P	R	F1
SEC-BERT	0.206	0.454	0.284
$\text{EMB}_{\text{E1, E2}}$+NN	0.731	0.701	0.709
MOAT	**0.748**	**0.743**	**0.736**

Table 3: Ablation Study.

Model	P	R	F1
MOAT	0.748	**0.743**	**0.736**
-relation group, entity types	0.736	0.725	0.720
+$\text{SBERT}_{\text{SDP}}$	0.694	0.687	0.679
+POS tags	0.747	0.738	0.737
MOAT (per relation group)	**0.839**	0.672	0.715
MOAT (LUKE)	0.467	0.545	0.497

Table 4: MOAT versus generative LLMs.

Type	LLM	P	R	F1
Zero Shot	Falcon	0.538	0.434	0.362
	Dolly	0.400	0.316	0.253
	MPT	0.295	0.380	0.255
Few Shot	Falcon	0.246	0.258	0.242
	Dolly	0.348	0.234	0.245
	MPT	0.296	0.156	0.128
Classifier	MOAT	**0.726**	**0.724**	**0.717**

expensive. We observe that MOAT outperforms all these LLMs in both zero shot and few shot settings. This is presented in Table 4.

7 CONCLUSION

In this paper, we discussed how language models can be used to determine the relation between a given pair of financial entities. We experimented with different LLMs and their variations. Finally, we propose MOAT, a novel architecture for determining relationship between financial entities.

Instruction fine-tuning generative LLMs, engineering better prompts, extracting entities and relations jointly, and benchmarking MOAT on other relevant datasets are a few directions for future work.

REFERENCES

[1] Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Merouane Debbah, Etienne Goffinet, Daniel Heslow, Julien Launay, Quentin Malartic, Badreddine Noune, Baptiste Pannier, and Guilherme Penedo. 2023. Falcon-40B: an open large language model with state-of-the-art performance. https://huggingface.co/tiiuae/falcon-7b-instruct

[2] Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the Blanks: Distributional Similarity for Relation Learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 2895–2905. https://doi.org/10.18653/v1/P19-1279

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[4] Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. 2010. SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals. In *Proceedings of the 5th International Workshop on Semantic Evaluation*. Association for Computational Linguistics, Uppsala, Sweden, 33–38. https://aclanthology.org/S10-1006

[5] Simerjot Kaur, Charese Smiley, Akshat Gupta, Joy Sain, Dongsheng Wang, Suchetha Siddagangappa, Toyin Aguda, and Sameena Shah. 2023. REFinD: Relation Extraction Financial Dataset.

[6] Lefteris Loukas, Manos Fergadiotis, Ilias Chalkidis, Eirini Spyropoulou, Prodromos Malakasiotis, Ion Androutsopoulos, and Paliouras George. 2022. FiNER: Financial Numeric Entity Recognition for XBRL Tagging. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022)* (Dublin, Republic of Ireland). Association for Computational Linguistics. https://arxiv.org/abs/2203.06482

[7] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://arxiv.org/abs/1908.10084

[8] Soumya Sharma, Tapas Nayak, Arusarka Bose, Ajay Kumar Meena, Koustuv Dasgupta, Niloy Ganguly, and Pawan Goyal. 2022. FinRED: A Dataset for Relation Extraction in Financial Domain. In *Companion Proceedings of the Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 595–597. https://doi.org/10.1145/3487553.3524637

[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (Long Beach, California, USA) *(NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 6000–6010.

[10] Somin Wadhwa, Silvio Amir, and Byron C. Wallace. 2023. Revisiting Relation Extraction in the era of Large Language Models. arXiv:2305.05003 [cs.CL]

[11] Ikuya Yamada, Akari Asai, Hiroyuki Shindo, Hideaki Takeda, and Yuji Matsumoto. 2020. LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention.

[12] Xiaoyan Zhao, Yang Deng, Min Yang, Lingzhi Wang, Rui Zhang, Hong Cheng, Wai Lam, Ying Shen, and Ruifeng Xu. 2023. A Comprehensive Survey on Deep Learning for Relation Extraction: Recent Advances and New Frontiers. arXiv:2306.02051 [cs.CL]

A PROMPTS

Details relating to the prompts are mentioned in Table 5.

B MODELS AND HYPER-PARAMETERS

We selected 3 open-source LLMs with number of parameters varying from 1 to 7 billion, such that these models could be effectively loaded and scored in Google Colab (free tier).More details about the models and their hyperparameters are mentioned in Table 6.

Table 5: Prompts for LLMs. The portion within box brackets, i.e. [content] is replaced by corresponding content from the dataset and is enclosed by backticks. <list of relation categories> refers to the categories mention in Table 1.

LLMs (Type)	**Prompt**
Falcon, Dolly, MPT (Zero Shot)	Determine the relationship between entities [e1] of type [e1-type] and [e2] of type [e2-type] in the text given below. Choose anyone from <list of relation categories> Input: [text] Response: Relation between [e1] of type [e1 type] and [e2] of type [e2 type] is
Falcon, Dolly (Few Shot)	Determine the relationship between entities [e1] of type [e1-type] and [e2] of type [e2-type] in the text given below. Choose anyone from <list of relation categories> Input: the capitalization table attached hereto as Exhibit H (the Cap Table) sets forth all of the outstanding capital of Microbot on a Fully - Diluted Basis as of the Effective Date , including the shares issued to Technion on account of Professor Shoham s involvement as a founder of Microbot ; . Response: Relation between ‘Shoham‘ of type ‘PERSON‘ and ‘Professor‘ of type ‘TITLE‘ is “pers:title:title“ Input: volatility in exchange rates between the U.S. dollar and currencies of the countries in which SAN DIEGO GAS & ELECTRIC CO operate , as SAN DIEGO GAS & ELECTRIC CO discuss below . Response: Relation between ‘SAN DIEGO GAS & ELECTRIC CO‘ of type ‘ORG‘ and ‘U.S.‘ of type ‘GPE‘ is “org:gpe:operations in“ Input: Dr. Fink joined Maxwell as President and Chief Executive Officer, and was appointed a director in May 2014, therefore his 2014 compensation in the table above reflects only a partial year . Response: Relation between ‘Fink‘ of type ‘PERSON‘ and ‘Maxwell‘ of type ‘ORG‘ is “pers:org:employee of“ Input: [text] Response: Relation between [e1] of type [e1 type] and [e2] of type [e2 type] is
MPT (Few Shot)	Similar to Falcon and Dolly with <|endoftext|>added at the end of each response.

Table 6: Models and Hyperparameters. We use the default values of the hyperparameters not mentioned here.

Model details	**Hyperparameters**
Falcon 7B-instruct[1]	max. new tokens=15, do sample=False, num. return sequences=1
Dolly-v2-3b	max. new tokens=15
MPT-1B-redpajama-200b-dolly	max. new tokens=15
SEC-BERT [6]	max. len = 512, batch size = 8, epochs = 5, learning rate = 2e-05, CrossEntropy loss, AdamW optimizer
Neural Network (NN)	hidden layer sizes = [512,128], max. no. of iterations = 300