Reproducing TableFormer: Enhancing Robustness in Table-Text Encoding with Transformer Models

Maksim Ploter
University of Tartu
Institute of Computer Science
maksim.ploter@ut.ee

Kaspar Metsa
University of Tartu
Institute of Computer Science
kaspar.metsa@ut.ee

Abstract

Understanding and reasoning over tabular data is important in many domains. Traditional transformer-based models like TAPAS (Table Parser) linearise tables into sequences, introducing row and column position biases. These models exhibit inconsistent performance when the table data order changes (perturbations). TableFormer, proposed by (Yang et al., 2022), addresses this by using learnable attention biases, making the model ignore row and column order. In this work, we assessed the importance of the pre-training step for the TableFormer model by omitting it entirely from the training procedure. Our evaluations showed that TableFormer can achieve the same accuracy as pre-trained TAPAS[1].

1 Introduction

Understanding and reasoning over tabular data is important in various fields, including scientific research, business, and daily activities. The advent of transformer-based models has significantly advanced the ability to process natural language, but their application to tables, especially those without a fixed schema, has revealed inherent limitations. Traditional approaches, such as those used in BERT-like models, convert tables into sequential word representations. This linearisation introduces unwanted biases due to the row and column positions, leading to inconsistent performance when the data order changes (Herzig et al., 2020), (Liu et al., 2021), (Eisenschlos et al., 2020).

Prior models, including TAPAS, have demonstrated vulnerabilities to such perturbations, where altering the order of rows or columns can significantly impact the model's predictions (Herzig et al., 2020). This inconsistency is a fundamental flaw, especially in applications requiring robust and reliable data interpretation (Chen et al., 2020), (Yin et al., 2020).

TableFormer, introduced by (Yang et al., 2022), addresses these limitations by incorporating learnable attention biases that make the model invariant to row and column order. This architecture ensures consistent performance across different table configurations, enhancing the model's robustness and reliability.

In their paper, (Herzig et al., 2020) claim that pre-training allows the TAPAS model to learn many interesting correlations between text and tables, as well as between the cells of a column and their header.

Our work aims to evaluate the importance of pre-training for the TableFormer model, and whether it can be omitted and still be able to outperform the pre-trained TAPAS model.

2 Related Work

The task of table-text encoding has seen significant advancements with the introduction of transformer-based models. This section reviews several key contributions in the field that have paved the way for our work on reproducing the TableFormer model.

2.1 Table Understanding with Transformers

The use of transformer architectures for understanding tables gained traction with models like TAPAS (Herzig et al., 2020). TAPAS adapts the BERT architecture to handle tables by linearising table rows and columns into sequences of words. While this approach leverages the powerful pre-training capabilities of BERT, it introduces biases related to the positional information of rows and columns, making the model susceptible to perturbations in the table's order.

[1]Code has been released at https://github.com/Maxvgrad/tapas

(Eisenschlos et al., 2021) extended this line of work by introducing MATE, a multi-view attention-based table transformer designed to improve the efficiency of table encoding. MATE incorporates multiple views of the table structure to enhance the model's understanding but still relies on row and column positional embeddings, which can lead to similar issues with order perturbations.

2.2 Addressing Structural Biases

Several approaches have been proposed to mitigate the biases introduced by linearisation. (Yin et al., 2020) introduced TaBERT, which integrates table structure by prepending column headers to cell contents and using a row encoder and a column encoder sequentially. However, this method is computationally expensive and still prone to biases from positional embeddings.

(Chen et al., 2020) tackled table-text entailment with TABFACT, a dataset for verifying facts from tables. Their model uses column headers as features for cells, which improves the understanding of table structures but does not eliminate the dependency on row and column order.

2.3 Learnable Attention Biases

(Yang et al., 2022) proposed TableFormer, a robust transformer model that incorporates learnable attention biases to capture the structural relationships within tables. This model eliminates the need for row and column positional embeddings, making it invariant to the order of table rows and columns. TableFormer outperforms existing models on various table reasoning tasks, particularly in scenarios involving perturbations of row and column orders.

(Shaw et al., 2018) introduced relative position representations in self-attention mechanisms, which inspired the learnable attention biases used in TableFormer. Their approach demonstrated the potential for improving transformer models' understanding of relational structures within input sequences.

2.4 Applications and Performance

The application of these models to tasks like question answering over tables has shown promising results. TAPEX (Liu et al., 2021) pre-trains a neural SQL executor for table question answering, further highlighting the potential of pre-trained transformer models in understanding tabular data. However, TAPEX still faces challenges related to order biases.

Our project aims to build on these foundational works by reproducing the TableFormer model's experiments and verifying its robustness and performance improvements over TAPAS in encoding table-text relationships for the question-answering task using the WikiTableQuestions dataset.

3 Methodology

3.1 Datasets

We use the following datasets in our experiments.

Wiki-table dataset from (Herzig et al., 2020), which consists of 6.2 million tables from WikiTable: 3.3 million of class Infobox[2] and 2.9 million of class WikiTable. Additionally, we subsampled the dataset by taking 5% and 10% of the tables from the original dataset, resulting in 317,478 and 634,956 tables, respectively. The dataset is used for masked language model pre-training (Devlin et al., 2019).

WikiTableQuestions (WTQ) dataset (Pasupat and Liang, 2015), a crowd-sourced dataset containing complex questions about Wikipedia tables. The dataset includes 22,033 training examples, 2,744 validation examples, and 4,344 test examples. Each example consists of a table and a corresponding natural language question. The dataset tests the model's ability to understand and reason over tabular data. The dataset is used for fine-tuning.

3.2 Models and Architectures

Our work involves two main models: TAPAS and TableFormer.

3.2.1 TAPAS Model

TAPAS (Table Parser) is a BERT-based architecture specifically designed for table question answering. It linearizes tables by converting rows into token sequences, appending the natural language question at the end, separated by [SEP] tokens. TAPAS uses positional embeddings to incorporate the table structure, including row and column IDs, which introduces unwanted biases (Herzig et al., 2020).

3.2.2 TableFormer Model

TableFormer enhances TAPAS by incorporating structural biases through learnable attention bi-

[2] en.wikipedia.org/wiki/Help:Infobox

ases, ensuring the model's robustness to row and column perturbations. TableFormer eliminates positional embeddings for rows and columns, using 13 types of attention biases to capture table structure and table-text relationships. These biases include "same row," "same column," "header to column cell," and others that facilitate better table-text alignment (Yang et al., 2022).

3.3 Experimental Setup

In the first experiment, we assess the importance of pre-training for the TableFormer model based on the BERT Tiny architecture. To achieve this, we trained two models: a model without pre-training and a model with 1M steps pre-training on 5% of the Wiki-table dataset. Fine-tuning on the WTQ dataset was the same for both models.

For the second experiment, we used a larger TableFormer model, based on the BERT Base architecture. We skipped pre-training, and fine-tuning remained unchanged. The training batch size was reduced to 32 (from 512), but we set gradient accumulation to 16 (see Appendix A) to replicate fine-tuning in (Herzig et al., 2020). TableFormer was trained for a smaller number of steps, only 61,000.

3.4 Evaluation Procedure

We evaluated the models using the denotation accuracy metric, which measures the percentage of questions for which the model's predicted answer matches the ground truth. Additionally, we assessed the robustness of the models to row and column perturbations by shuffling the table rows and columns during inference and observing the performance variation.

4 Results

Table 1 shows higher denotation accuracy on the development and test sets for TableFormer models without pre-training based on BERT (Tiny).

In the paper by (Herzig et al., 2020), there are no results reported for the TAPAS (Tiny) model. Therefore, we fine-tuned the TAPAS (Tiny) model, which was already pre-trained on the complete Wiki-table dataset. TAPAS (Tiny) achieved an accuracy of 5.80% and 5.92% on the dev and test datasets, respectively. These results are comparable to the TableFormer (Tiny) model without pre-training.

Model	**Dev**	**Test**
TableFormer (Tiny) no pre-training	5.59	5.87
TableFormer (Tiny) pre-training 1M steps	3.91	4.01

Table 1: Denotation accuracy on WTQ development and test set. The median of 5 independent runs is reported.

Table 2 shows that TableFormer (Base) without pre-training is capable of achieving similar accuracy as the TAPAS (Base) accuracy reported in (Herzig et al., 2020). See fine-tuning hyperparameters in Appendix A.

Model	Dev	Test
TAPAS (Base)	23.6	24.1
TableFormer (Base) no pre-training	18.64	19.92

Table 2: Denotation accuracy on WTQ development and test set. Accuracy for TAPAS (Base) is taken from (Herzig et al., 2020).

5 Discussion

Our experiments reveal several important insights into the performance of the TableFormer model compared to the TAPAS model on the WikiTableQuestions (WTQ) dataset. One of the key findings is that pre-training on a large dataset does not always guarantee superior performance. This is evident from our first experiment, where the TableFormer (Tiny) model without pre-training outperformed the pre-trained version, achieving higher denotation accuracy on both the development and test sets.

The results suggest that TableFormer's architecture can effectively capture the essential table-text relationships even without pre-training.

Furthermore, the comparison between TableFormer (Tiny) and TAPAS (Tiny) models shows that TableFormer (Tiny) without pre-training achieves comparable accuracy to TAPAS (Tiny), which was pre-trained on the entire Wiki-table dataset. This shows the efficiency of TableFormer's architecture in leveraging table structure for question-answering tasks.

In the second experiment, TableFormer (Base)

without pre-training managed to achieve denotation accuracies close to those of TAPAS (Base) reported in (Herzig et al., 2020). This further supports the idea that TableFormer's attention biases play a crucial role in understanding and reasoning over tabular data, potentially reducing the dependence on large-scale pre-training.

Future work could explore the application of similar attention bias mechanisms to other types of structured data, such as graphs or relational databases, to enhance the versatility and applicability of transformer models across different domains.

6 Conclusion

Our experiments demonstrate the advantage of TableFormer over TAPAS models on the WTQ dataset, as it achieves similar accuracies without the need for pre-training. Specifically, the TableFormer (Tiny) model without pre-training achieved denotation accuracies of 5.59% on the development set and 5.87% on the test set, outperforming the TableFormer (Tiny) model with pre-training. Comparatively, TAPAS (Tiny) achieved slightly higher accuracies of 5.80% on the development set and 5.92% on the test set.

For the larger models, TableFormer (Base) without pre-training achieved denotation accuracies of 18.64% on the development set and 19.92% on the test set. Although these are lower than TAPAS (Base)'s reported accuracies of 23.6% and 24.1%, they still demonstrate the potential of TableFormer's structural biases in handling tabular data effectively without extensive pre-training.

In summary, our study confirms that TableFormer's learnable attention biases contribute to a robust understanding of table structure, reducing the reliance on pre-training and showing strong performance relative to TAPAS models. These findings highlight the practical benefits of structural biases in transformer models for table-based question-answering tasks, offering insights for future research and applications in similar domains.

7 Contributions of the Group Members

Maksim and Kaspar collaboratively planned the reproduction of the paper and experimented with various approaches. They both implemented the pre-trained TAPAS model inference on HPC using the WikiTableQuestions (WTQ) dataset, with an unsuccessful attempt to run it on Google Colab. Maksim pre-trained and fine-tuned the TAPAS+TableFormer model, conducting inference on the WTQ dataset to compare it with plain TAPAS. Kaspar was responsible for writing the report and preparing the presentation slides.

References

Wenhu Chen, Hongmin Wang, Jianshu Chen, Yunkai Zhang, Hong Wang, Shiyang Li, Xiyou Zhou, and William Yang Wang. 2020. TabFact : A large-scale dataset for table-based fact verification. In *International Conference on Learning Representations (ICLR)*, Addis Ababa, Ethiopia, April.

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)*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Julian Eisenschlos, Syrine Krichene, and Thomas Müller. 2020. Understanding tables with intermediate pre-training. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 281–296, Online, November. Association for Computational Linguistics.

Julian Eisenschlos, Maharshi Gor, Thomas Müller, and William Cohen. 2021. MATE: Multi-view attention for table transformer efficiency. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7606–7619, Online and Punta Cana, Dominican Republic, November. Association for Computational Linguistics.

Jonathan Herzig, Pawel Krzysztof Nowak, Thomas Müller, Francesco Piccinno, and Julian Eisenschlos. 2020. TaPas: Weakly supervised table parsing via pre-training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4320–4333, Online, July. Association for Computational Linguistics.

Qian Liu, Bei Chen, Jiaqi Guo, Zeqi Lin, and Jianguang Lou. 2021. TAPEX: Table pre-training via learning a neural SQL executor. *arXiv preprint arXiv:2107.07653*.

Panupong Pasupat and Percy Liang. 2015. Compositional semantic parsing on semi-structured tables. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1470–1480, Beijing, China, July. Association for Computational Linguistics.

Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. 2018. Self-attention with relative position representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 464–468, New Orleans, Louisiana, June. Association for Computational Linguistics.

Jingfeng Yang, Aditya Gupta, Shyam Upadhyay, Luheng He, Rahul Goel, and Shachi Paul. 2022. TableFormer: Robust transformer modeling for table-text encoding. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 528–537, Dublin, Ireland, May. Association for Computational Linguistics.

Pengcheng Yin, Graham Neubig, Wen-tau Yih, and Sebastian Riedel. 2020. TaBERT: Pretraining for joint understanding of textual and tabular data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8413–8426, Online, July. Association for Computational Linguistics.

A Hyperparameters

Parameter	WTQ
Training Steps	61,000
Learning rate	1.93581e-5
Warmup ratio	0.128960
Answer loss cutoff	0.664694
Huber loss delta	0.121194
Cell selection preference	0.207951
Batch size	32
Gradient accumulation steps	16
Gradient clipping	10
Select one column	1
Reset cell selection weights	0

Table 3: Hyperparameters for WTQ.

B Training

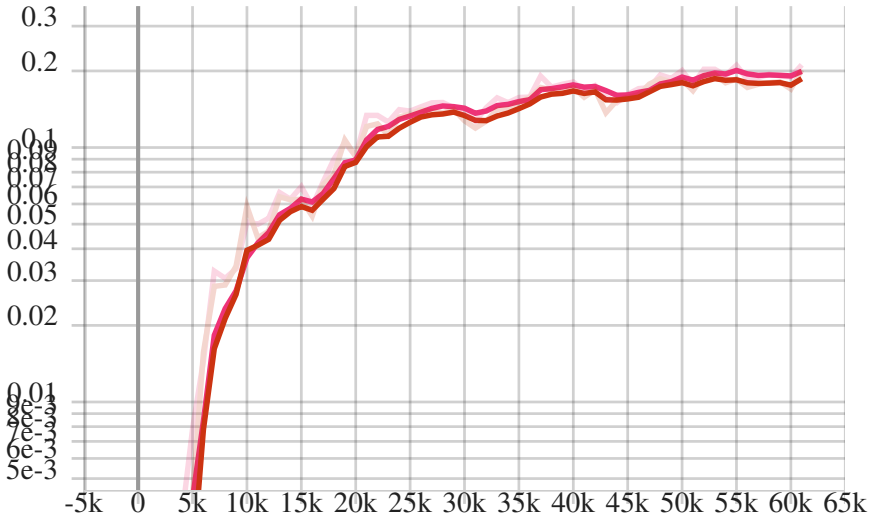

Figure 1: Denotation accuracy for test (light pink) and dev (dark red). Smoothing set to 0.6 .