Sparse Attention for Tabular QA: A Must-Have for Robust Table Encoding

Raphaël Mouravieff[1], Tristan Luiggi[1], Benjamin Piwowarski[1][0000−0001−6792−3262], and Sylvain Lamprier[2][0000−0002−2508−922X]

[1] Sorbonne Université, CNRS, ISIR, F-75005 Paris, France
{raphael.mouravieff, tristan.luiggi, benjamin.piwowarski}@isir.upmc.fr
[2] LERIA, Université d'Angers, France
sylvain.lamprier@univ-angers.fr

Abstract. The structured nature of tabular data poses significant challenges for deep learning models, which lose structural information when converting tables into linear sequences. Prior work has proposed methods to preserve structure, but they still fall short on generalization. In this study, we investigate the impact of encoding techniques on generalization. Ours results demonstrate that sparse attention mechanisms, focusing on key table components during encoding, significantly enhance model's structural understanding.

Keywords: Tabular Data · Sparse Attention · Generalization

0.1 Introduction

Transformers, originally designed for text processing, handle tabular data by flattening it into sequences [6], which results in the loss of crucial structural information. While existing methods attempt to preserve structure information using special tokens [7], structural embeddings [1], or attention bias [4], their individual impact remains unclear. Additionally, these architectures are prone to overfitting [8, 9]. Even Large Language Models [11] struggle with simple tasks, such as counting rows in large table. This paper examines mechanisms for preserving tabular structure and evaluate their impact on generalization across both synthetic and real-world datasets. We find that (1) tabular encoding choice significantly affects generalization, and (2) sparse attention masks improve robustness to table size, perturbations, and real-world datasets like WikiSQL [10].

0.2 Methods

We compare structure-preserving modules to assess their generalization impact. **Encoding Methods:** We evaluate four models that preserve table structure after flattening. TaPEx [2] introduces Structured **T**okens (**T**), TaPas [1] uses structural **E**mbeddings (**E**), TableFormer [4] applies attention **B**ias (**B**), and MATE [3] employs a sparse **M**ask (**M**) to restrict attention within rows and columns. For fair comparison, we use BART [5] as the backbone, integrating

each encoding module into a unified framework and accuracy [1] as metric.
Studying Generalization: We construct a SQL execution dataset with 10 query templates. Tables are randomly generated using a fixed vocabulary of 1000 numbers, with independent random generation for test sets. We evaluate **Out-Structure** (larger tables), **Robustness** (tables with repetition probability $r \in \{0.2, 0.4\}$), and **Compositionality** (unseen SQL query). We also use **WikiSQL** to confirm our results on a real-world dataset.

0.3 Experiments

Table 1 compares models w/wo masks, showing their key role in generalization and robustness to perturbations. Notably for structure generalization, with gains up to 10 points. To validate our results, we fine-tune all configurations on WikiSQL (Figure 1). Models without structural components perform worst (bottom left). Across all cases, models with masks consistently outperform those without.

Table 1. Model performance on generalization tests, highlighting the impact of mask **M** on synthetic datasets.

Model	Out Structure	Compositional	Robustness
TaPas	66.3	62.8	83.0
TaPas+M	76.0	62.8	88.6
TaPEx	66.3	62.9	71.3
TaPEx+M	71.5	62.6	90.2
TableFormer	78.0	62.7	80.2
TableFormer+M	81.1	62.4	86.2

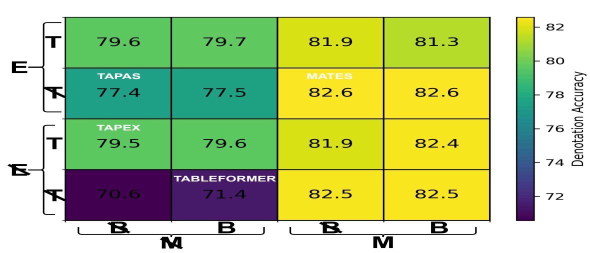

Fig. 1. Impact of structural encoding methods on WikiSQL performance. We use \ to indicate absence of the component. (**B**=**B**ias, **M**=**M**ask, **E**=**E**mbedding, **T**=**T**okens).

We **conclude** that sparse attention improves generalization by focusing on relevant table components and minimizing unnecessary token interactions, preventing spurious dependencies and biases.

Acknowledgments. This work was partly funded by the ANR-21-CE23-0007 ACDC project. Experiments were performed using HPC resources from GENCI IDRIS (Grant AD011014032R1, AD011014110 and A0151014638).

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Herzig, J., Nowak, P. K., Müller, T., Piccinno, F., Eisenschlos, J. M.: TaPas: Weakly supervised table parsing via pre-training. arXiv preprint arXiv:2004.02349 (2020)
2. Liu, Q., Chen, B., Guo, J., Ziyadi, M., Lin, Z., Chen, W., Lou, J.-G.: TAPEX: Table pre-training via learning a neural SQL executor. arXiv preprint arXiv:2107.07653 (2021)
3. Eisenschlos, J. M., Gor, M., Müller, T., Cohen, W. W.: MATE: Multi-view attention for table transformer efficiency. arXiv preprint arXiv:2109.04312 (2021)
4. Yang, J., Gupta, A., Upadhyay, S., He, L., Goel, R., Paul, S.: TableFormer: Robust transformer modeling for table-text encoding. arXiv preprint arXiv:2203.00274 (2022)
5. Lewis, M.: BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461 (2019)
6. Dong, H., Cheng, Z., He, X., Zhou, M., Zhou, A., Zhou, F., Liu, A., Han, S., Zhang, D.: Table pre-training: A survey on model architectures, pre-training objectives, and downstream tasks. arXiv preprint arXiv:2201.09745 (2022)
7. Jiang, Z., Mao, Y., He, P., Neubig, G., Chen, W.: OmniTab: Pretraining with natural and synthetic data for few-shot table-based question answering. arXiv preprint arXiv:2207.03637 (202
8. Zhao, Y., Zhao, C., Nan, L., Qi, Z., Zhang, W., Tang, X., Mi, B., Radev, D.: RobuT: A systematic study of Table QA robustness against human-annotated adversarial perturbations. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL), pp. 6064–6081. ACL, Toronto, Canada (2023). https://aclanthology.org/2023.acl-long.334, doi:10.18653/v1/2023.acl-long.334
9. Zhou, W., Mesgar, M., Adel, H., Friedrich, A.: FREB-TQA: A fine-grained robustness evaluation benchmark for Table Question Answering. In: Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), pp. 2479–2497. ACL, Mexico City, Mexico (2024). https://aclanthology.org/2024.naacl-long.137, doi:10.18653/v1/2024.naacl-long.137
10. Zhong, V., Xiong, C., Socher, R.: Seq2SQL: Generating structured queries from natural language using reinforcement learning. arXiv preprint arXiv:1709.00103 (2017)
11. Sui, Y., Zhou, M., Zhou, M., Han, S., Zhang, D.: Table meets LLM: Can large language models understand structured table data? A benchmark and empirical study. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining, pp. 645–654 (2024)