Structural Deep Encoding for Table Question Answering

Mouravieff Raphaël [1], Piwowarski Benjamin [1,2], Lamprier Sylvain [3]

Introduction

Problem: Flattening tables **breaks their structure,** and their **size exceeds what transformers can encode** due to quadratic complexity.

Method: We systematically analyze table encoding techniques and introduce **novel sparse attention masks** to enhance both generalization and efficiency.

Contribution:

- Identify key encoding factors that improve model robustness.
- Show that absolute positional encoding is essential for table processing.
- Propose efficient sparse attention masks, reducing computation while preserving structure.

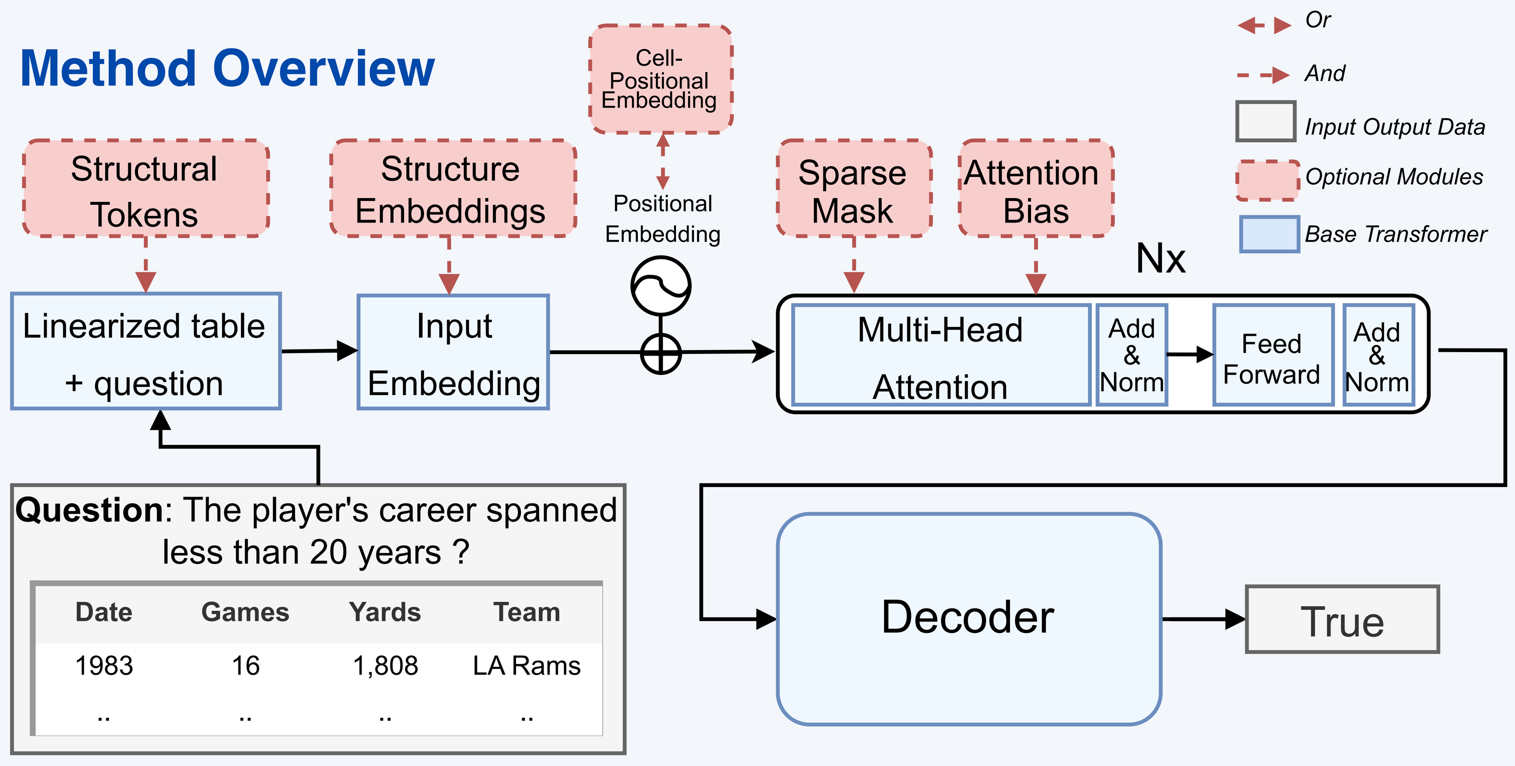
Structural encoding components

Backbone Model: BART and segment embeddings.

Structural Encoding:

- **Attention Bias (B):** Learnable biases for relational modeling. **[1]**
- **Embeddings (E):** Row-column embeddings. **[2] [4]**
- **Special Tokens (T):** Row, column, and cell tokens markers (T0, T1, T2). **[3]**
- **Sparse Mask (M):** Six attention masking strategies for table structure. **[4]**
- **Positional Embeddings (PE):** Cell-based (CPE) or table-wide (TPE) positional IDs. **[1] [4]**

Method Overview



Generalization Results on Synthetic Data

Synthetic Data: Generated to test generalization across structural changes, missing values, compositional reasoning, and correlations (mixability).

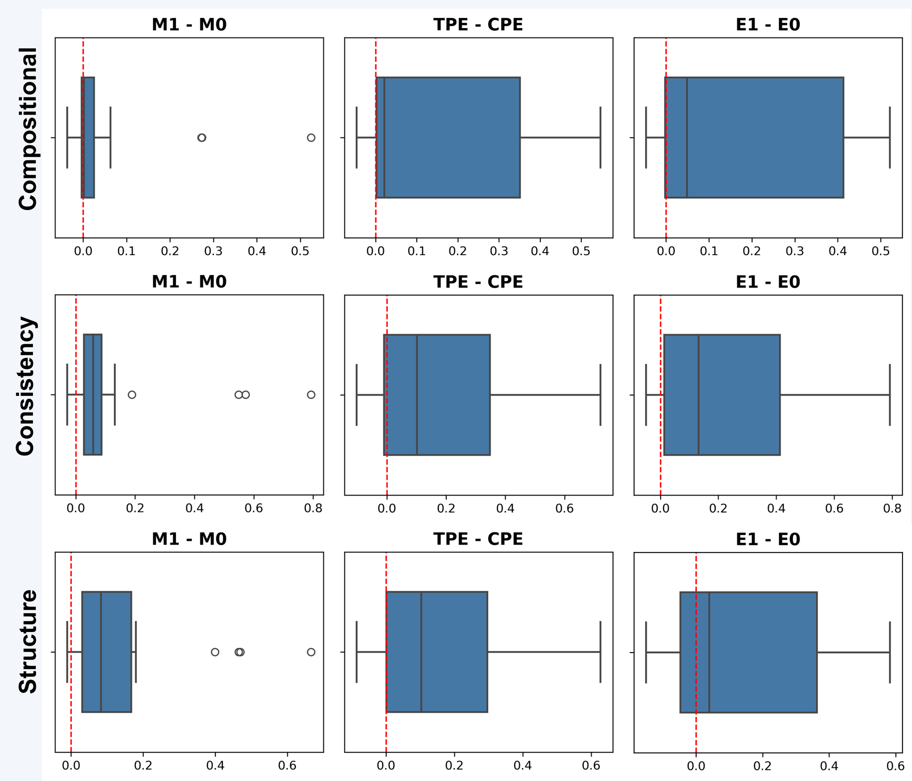
Results:

- Absolute encodings (TPE, E) improve performance, outperforming relative ones.
- Sparse masking (M) enhances generalization, especially with PE and E.
- Special tokens (T) and bias (B) have minimal impact on performance.

ANOVA of all factors on synthetic data

Factor	In Domain η^2 (p-value)	Structure η^2 (p-value)	Consistency η^2 (p-value)	Compositional η^2 (p-value)
T	0.00	0.00	0.00	0.00
M	0.04	0.07	0.01	0.00
PE	**0.19**	**0.27**	**0.19**	**0.26**
B	0.01	0.01	0.00	0.00
E	**0.20**	**0.15**	**0.30**	**0.26**
TM	0.00	0.01	0.01	0.02
T×PE	0.00	0.00	0.00	0.01
T×B	0.00	0.00	0.00	0.00
T×E	0.00	0.00	0.00	0.00
M×PE	**0.08**	**0.05**	**0.07**	**0.04**
M×B	0.02	**0.04**	0.02	0.01
M×E	0.09	0.04	0.07	0.04
PE×B	0.01	0.00	0.00	0.00
PE×E	**0.19**	**0.21**	**0.18**	**0.26**
B×E	0.01	0.01	0.01	0.00

Differences structural encoding



Validating Structural Encoding Insights

Results:

- Sparse masks improve generalization, with M1 outperforming the baseline (M0).
- Sparser masks (M3, M5, M6) remain competitive, demonstrating the effectiveness of structured sparsity.
- Absolute positionning (row/column indices provided by TPE or E1) is required for decoding performance.
- Findings align with synthetic data experiments, reinforcing the importance of structured encoding.

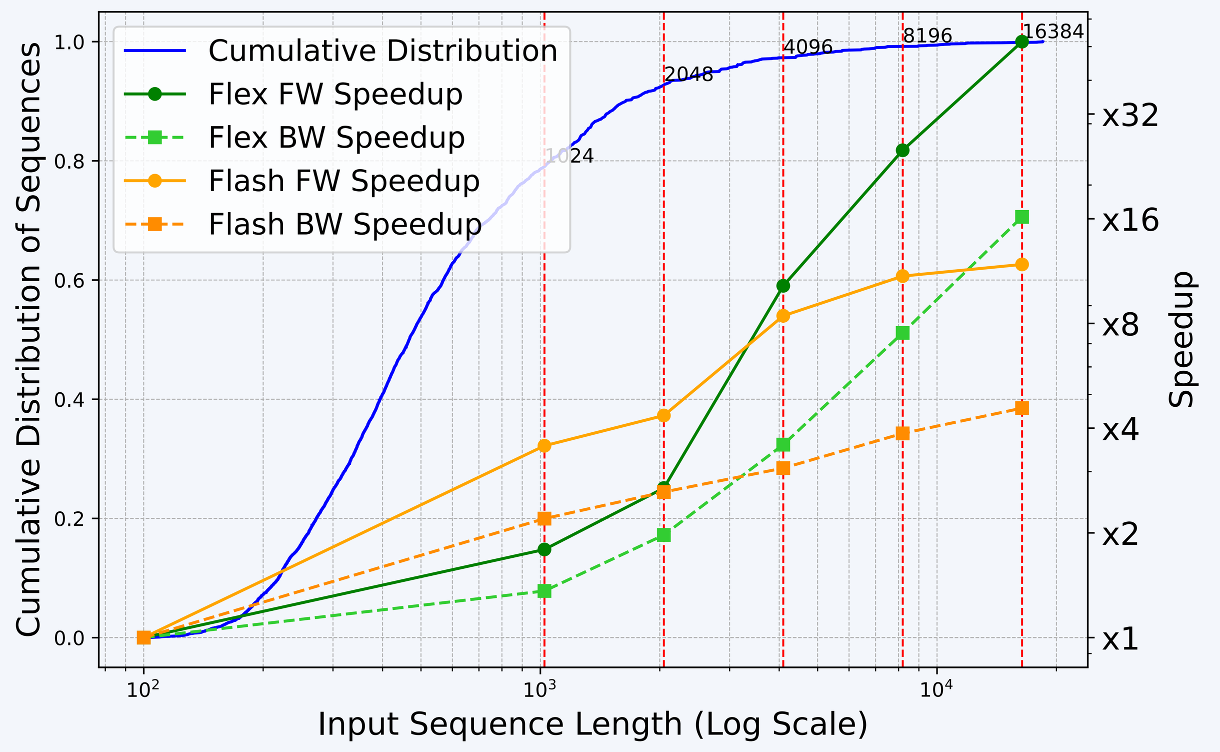
Results on WIKISQL dataset

PE	E1	T	M0 B0	M0 B1	M1 B0	M1 B1	M2 B0	M2 B1	M3 B0	M3 B1	M4* B0	M4* B1	M5* B0	M5* B1	M6* B0	M6* B1
CPE	E0	T2	26.1	38.6	81.3	81.6	30.2	29.7	58.0	56.3	29.9	29.9	61.4	57.5	29.2	28.8
		T1	26.1	26.6	82.3	81.5	30.0	30.0	57.1	54.7						
		T0	22.7	23.1	79.6	76.8	29.8	29.5	36.5	34.4						
	E1	T2	75.7	73.0	82.1	81.6	78.5	77.6	81.8	81.2	79.2	78.6	81.9	81.2	29.3	29.6
		T1	79.0	79.0	82.2	81.5	78.5	78.0	82.4	81.4						
		T0	29.8	28.8	79.2	78.3	77.2	78.0	71.0	66.2						
TPE	E0	T2	79.4	79.6	82.0	82.0	78.9	78.2	78.7	79.1	80.1	80.0	79.0	78.5	80.5	80.4
		T1	79.5	79.6	81.9	82.4	79.3	78.8	79.2	79.5						
		T0	70.6	71.4	82.5	82.5	79.3	78.9	75.3	73.9						
	E1	T2	79.4	79.3	82.2	82.3	78.6	78.0	79.8	79.4	80.5	80.5	79.7	79.3	80.4	79.8
		T1	79.6	79.7	81.9	81.3	79.0	78.5	79.1	79.9						
		T0	77.4	77.5	82.6	82.6	79.2	79.1	77.2	76.8						

An Efficient Sparse Attention Mask for Faster and improved performance

Results:

- **Efficient Sparse Mask:** M3 enhances computational efficiency while preserving accuracy.
- **Up to 50× Speedup:** Achieves significant forward acceleration for long sequences (16,384 tokens).
- **Scalable for Large Tables:** Handles long table sequences efficiently, overcoming token limits.
- **Tested on Real & Synthetic Data:** Maintains strong performance across diverse datasets.
- **Optimized with Flex & FlashAttention2:** Leverages sparse matrix operations for faster execution.



References

[1] Yang, J., et al. (2022). TableFormer: Robust transformer modeling for table-text encoding.
[2] Herzig, J., et al. (2020). TAPAS: Weakly supervised table parsing via pre-training.
[3] Liu, Q., et al (2021). TAPEX: Table pre-training via learning a neural SQL executor.
[4] Eisenschlos, J. M et al (2021). MATE: multi-view attention for table transformer efficiency.

Acknowledgments

This work was partly funded by the ANR-21- CE23-0007 ACDC project. Experiments were performed using GENCI-IDRIS.

Machine Learning - Deep Learning and Information Access (MLIA)

[1] Machine Learning - Deep Learning and Information Access (MLIA)
[2] The National Centre for Scientific Research (CNRS)
[3] The Laboratory for Study and Research in Computer Science of Angers (LERIA)