INSTITUTE OF INTELLIGENT SYSTEM AND ROBOTICS

Learning Relational Decomposition of Queries for Question Answering from Tables

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

Introduction

→ **Problem** : Current Table QA methods either focus on content extraction (a) semantic parsing (b) or direct answer generation (c)

→ **Our contribution** : Bridge the gap between semantic parsing and direct generation using a relational algebra like framework (d)

→ **Benefits** : Improved generalization and structural reasoning

Question : The player's career spanned less than 20 years ?

Date	Games	Yards	Team
1983	16	1,808	Los Angels Rams
1984	16	2,105	Los Angels Rams
...			
1993	4	91	Atlanta Falcons
Career	146	13,256	

Answer : True

(a) Extractive

Column cells Selection | **Aggregations**

1993 | .. | 1983 — Max

Answer : 1993 ✗

Error Explanation

struggles with Complex, multi-aggregation queries.

(b) SQL-Queries

SQL Generation

select (select max (Date) - min (Date) from w) < 20

Answer : None ✗

Error Explanation

invalid input syntax for type integer: 'Career'

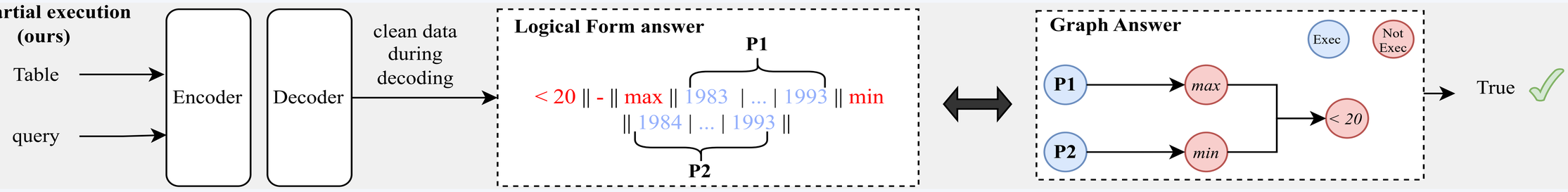
(c) Direct Answer Generation

Answer : False ✗

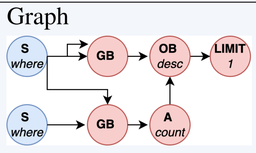
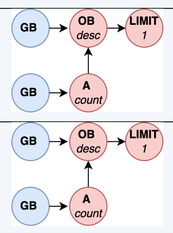
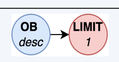
Error Explanation

Limited Numerical Reasoning in Transformers

d) Partial execution (ours)



Model

$\mathcal{O}^*$	Logical Form	Graph
{P, C, S}	Limit 1 ‖ OB desc ‖ GB ‖ newtongrange ǀ fauldhouse ǀ fauldhouse ‖ newtongrange ǀ fauldhouse ǀ fauldhouse ‖ COUNT ‖ GB ‖ newtongrange ǀ fauldhouse ǀ fauldhouse ‖ 1 ǀ 2 ǀ 7 ‖	
{P, C, S, GB, H}	Limit 1 ‖ OB desc ‖ fauldhouse,, fauldhouse ǀ newtongrange ‖ COUNT ‖ 2,, 7 ǀ 1 ‖	
{P, C, S, GB, H, OB}	Limit 1 ‖ OB desc ‖ fauldhouse,, fauldhouse ǀ newtongrange ‖ COUNT ‖ 2,, 7 ǀ 1 ‖	
{P, C, S, GB, H, OB, A}	Limit 1 ‖ fauldhouse,, fauldhouse ǀ newtongrange ‖	
Full	fauldhouse	

Algebra :

→ Defines operations like projection (P), comparison (C), selection (S), group-by (GB), aggregation (A).

Partial Execution of Computational Graph :

→ Executes parts of the computational graph using allowed operators $\mathcal{O}^*$.

→ Balances between direct generation (blue) and exeternal execution (red).

Linearizing the Graph :

→ Converts the computational graph into sequences for transformer models.

→ Uses pre-order and post-order linearization methods.

Results

Setup :

→ WikiTableQuestions Dataset : complex queries, mixed data types, sql supervision.

→ Denotation Accuracy (Da) and Flexible Denotation Accuracy (FDA) metrics. FDA ignores units for comparison.

Model	Projection (ALL)	Comparison	Selection	Group By	Order By	Aggregation	Operator	Limit	σ
#	500	367	363	30	151	206	75	153	
Tapas	52.6	51.8	52.3	16.7	53.0	43.7	30.7	52.3	13.5
Tapex	55.2	55.9	56.5	50.0	60.9	38.8	44.0	60.8	7.9
Omnitab	**58.8**	**59.7**	**59.8**	**56.7**	61.6	47.1	45.3	60.8	6.4
P	44.6	40.9	41.3	40.0	49.7	43.7	28.0	49.0	6.8
+C	51.6	50.1	50.7	23.3	48.3	50.0	38.7	47.7	9.7
+S	58.6	58.0	58.4	40.0	58.3	**52.4**	**52.0**	57.5	6.4
+GB+H	57.8	57.8	58.4	23.3	57.0	49.5	49.3	56.2	11.8
+OB	57.6	57.5	57.8	53.3	58.9	51.5	50.7	58.2	3.3
+A	58.0	57.8	58.4	**56.7**	**62.2**	47.1	49.3	**61.4**	5.4
+OP	56.6	57.8	58.4	50.0	60.3	46.1	42.7	60.1	6.8

(a) Using validation data

Model	Projection (ALL)	Comparison	Selection	Group By	Order By	Aggregation	Operator	Limit	σ
Tapas	42.6	41.7	42.2	16.7	38.4	37.9	18.7	37.9	10.6
Tapex	43.4	43.0	43.5	43.3	44.4	35.4	29.3	44.4	5.5
Omnitab	45.4	44.7	44.9	36.7	42.4	39.3	30.7	42.5	5.1
P	43.2	39.8	40.2	36.7	45.0	44.2	28.0	44.4	5.7
+C	49.0	46.3	46.8	23.3	45.7	48.5	38.7	45.1	8.5
+S	**53.6**	**51.0**	**51.2**	40.0	**49.0**	**51.9**	**50.7**	**48.4**	4.2
+GB+H	51.6	49.6	50.1	23.3	45.0	49.5	48.0	44.4	9.2
+OB	50.6	50.7	51.2	40.0	43.7	48.1	46.7	43.1	4.1
+A	47.2	46.0	46.6	**50.0**	45.7	41.8	40.0	45.1	3.1
+OP	47.8	47.7	48.2	**50.0**	45.7	43.2	30.7	45.8	6.1

(b) Using validation data with *random permutations*

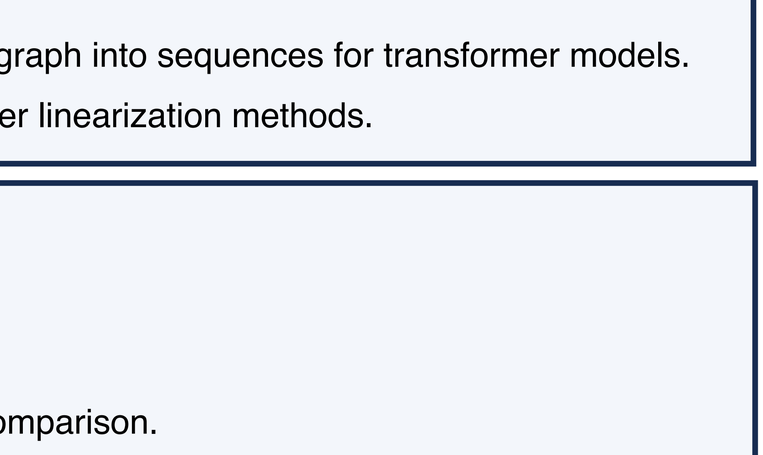
Performance Comparison :

→ Outperforms several state-of-the-art models.

→ Shows better generalization and robustness.

→ Combining models with different execution levels improves performance.

Model	SDA	FDA
Fine-Tuned Models		
TAPAS (Herzig et al., 2020)	48.8	50.2
TAPEX (Liu et al., 2021)	55.5	57.9
OmniTab (Jiang et al., 2022)	61.8	62.1
Semantic parsing on test with cleaned tables		
SQuALL (Shi et al., 2020)	50.4	54.3
Semantic parsing on test tables		
SQuALL (Shi et al., 2020)	23.2	27.2
Our models		
+P+C+S	59.0	61.4
Ensemble	63.3	66.3



References

1. Jiang, Z., et al. (2022). OmniTab: Pretraining with natural and synthetic data for few-shot table-based question answering.
2. Shi, T., et al. (2020). On the potential of lexico-logical alignments for semantic parsing to SQL queries.
3. Herzig, J., et al. (2020). TAPAS: Weakly supervised table parsing via pre-training.

Acknowledgments

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

[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)

Scan the QR code to visit our GitHub !