Reasoning over Tabular Data

CPSC677 - ANLP

Yilun Zhao (yilun.zhao@yale.edu)

Nov. 18th, 2021

How Does Table Understanding Interface with Natural Language Understanding?

We use tables to convey important information

Example: COVID cases

	Race/Ethnicity	Non-hospitalized	Non-fatal hospitalized	Confirmed deaths[1]	Probable deaths[2]	Total deaths
Count (% of known)	Black/African American	15927 (28.7%)	9432 (34.7%)	4239 (29.4%)	1330 (32.9%)	5569 (30.2%)
	Hispanic/Latino[3]	17036 (30.7%)	8780 (32.3%)	4513 (31.3%)	1183 (29.2%)	5696 (30.9%)
	White	18170 (32.8%)	6614 (24.3%)	3882 (26.9%)	1119 (27.6%)	5001 (27.1%)
	Asian/Pacific Islander	4088 (7.4%)	2128 (7.8%)	1143 (7.9%)	389 (9.6%)	1532 (8.3%)
	Other known[6]	237 (0.4%)	233 (0.9%)	630 (4.4%)	27 (0.7%)	657 (3.6%)
Count (% of total)	Total known	55458 (40.9%)	27187 (77.2%)	14407 (93.9%)	4048 (80.0%)	18455 (90.4%)
	Other/Unknown	80286 (59.1%)	8013 (22.8%)	942 (6.1%)	1009 (20.0%)	1951 (9.6%)
	Total	135744	35200	15349	5057	20406

Humans can quickly understand these tables and extract knowledge.

Pujara et al. [From Tables to Knowledge (KDD21).](#)

Express that information is tedious otherwise

Race/Ethnicity	Non-hospitalized	Non-fatal hospitalized	Confirmed deaths[1]	Probable deaths[2]	Total deaths
Black/African American	15927 (28.7%)	9432 (34.7%)	4239 (29.4%)	1330 (32.9%)	5569 (30.2%)

“Among African Americans there were 5,569 total deaths, representing 30.2% of total deaths. Of the 5,569 total deaths, 1,330 deaths were probable deaths (representing 32.9% of all probable deaths) while 4,239 were confirmed deaths (representing 29.4% of all confirmed deaths). Additionally, there were 9,432 non-fatal hospitalizations (representing 34.7% of all non-fatal hospitalizations) and 15,927 cases that did not require hospitalizations (representing 28.7% of all cases where hospitalization was unnecessary).”

Tables and text: two views of information, complementary sources of knowledge

Algorithm	SSL	RB	CIFAR-10 ($\gamma_l = 100$) $\gamma_u = 1$	$\gamma_u = 50$	$\gamma_u = 150$	$\gamma_u = 100$ (reversed)
Vanilla	-	-	$58.8_{\pm0.13}$ / $51.0_{\pm0.11}$	$58.8_{\pm0.13}$ / $51.0_{\pm0.11}$	$58.8_{\pm0.13}$ / $51.0_{\pm0.11}$	$58.8_{\pm0.13}$ / $51.0_{\pm0.11}$
Re-sampling [21]	-	✓	$55.8_{\pm0.47}$ / $45.1_{\pm0.30}$	$55.8_{\pm0.47}$ / $45.1_{\pm0.30}$	$55.8_{\pm0.47}$ / $45.1_{\pm0.30}$	$55.8_{\pm0.47}$ / $45.1_{\pm0.30}$
LDAM-DRW [8]	-	✓	$62.8_{\pm0.17}$ / $58.9_{\pm0.60}$	$62.8_{\pm0.17}$ / $58.9_{\pm0.60}$	$62.8_{\pm0.17}$ / $58.9_{\pm0.60}$	$62.8_{\pm0.17}$ / $58.9_{\pm0.60}$
cRT [22]	-	✓	$63.2_{\pm0.45}$ / $59.9_{\pm0.40}$	$63.2_{\pm0.45}$ / $59.9_{\pm0.40}$	$63.2_{\pm0.45}$ / $59.9_{\pm0.40}$	$63.2_{\pm0.45}$ / $59.9_{\pm0.40}$
VAT [30]	✓	-	$65.2_{\pm0.12}$ / $59.5_{\pm0.26}$	$64.0_{\pm0.31}$ / $57.3_{\pm0.66}$	$62.8_{\pm0.19}$ / $55.1_{\pm0.70}$	$59.4_{\pm0.36}$ / $50.6_{\pm0.61}$
Mean-Teacher [38]	✓	-	$73.9_{\pm1.19}$ / $71.7_{\pm1.42}$	$61.2_{\pm0.51}$ / $53.5_{\pm0.84}$	$59.7_{\pm0.50}$ / $50.0_{\pm1.61}$	$61.0_{\pm0.82}$ / $56.4_{\pm1.64}$
MixMatch [5]	✓	-	$41.5_{\pm0.76}$ / $12.0_{\pm1.34}$	$64.1_{\pm0.58}$ / $48.3_{\pm0.70}$	$65.5_{\pm0.64}$ / $51.1_{\pm2.41}$	$47.9_{\pm0.09}$ / $20.5_{\pm0.85}$
MixMatch + DARP	✓	-	$86.7_{\pm0.80}$ / $86.2_{\pm0.82}$	$68.3_{\pm0.47}$ / $62.2_{\pm1.21}$	$66.7_{\pm0.25}$ / $58.8_{\pm0.42}$	$72.9_{\pm0.24}$ / $71.0_{\pm0.32}$
			(-77.2% / -84.4%)	(-11.8% / -27.0%)	(-3.62% / -15.7%)	(-48.0% / -63.6%)
ReMixMatch [4]	✓	-	$48.3_{\pm0.14}$ / $19.5_{\pm0.85}$	$75.1_{\pm0.43}$ / $71.9_{\pm0.77}$	$72.5_{\pm0.10}$ / $68.2_{\pm0.32}$	$49.0_{\pm0.55}$ / $17.1_{\pm1.48}$
ReMixMatch*	✓	-	$85.0_{\pm1.35}$ / $84.3_{\pm1.55}$	$77.0_{\pm0.12}$ / $74.7_{\pm0.04}$	$72.8_{\pm0.10}$ / $68.8_{\pm0.21}$	$75.3_{\pm0.03}$ / $72.3_{\pm0.04}$
ReMixMatch* + DARP	✓	-	**89.7**$_{\pm0.15}$ / **89.4**$_{\pm0.17}$	**77.4**$_{\pm0.22}$ / $75.0_{\pm0.25}$	**73.2**$_{\pm0.11}$ / $69.2_{\pm0.31}$	**80.1**$_{\pm0.11}$ / **78.5**$_{\pm0.17}$
			(-31.4% / -32.5%)	(-1.72% / -1.49%)	(-1.53% / -2.64%)	(-19.5% / -22.5%)
FixMatch [37]	✓	-	$68.9_{\pm1.95}$ / $42.8_{\pm8.11}$	$73.9_{\pm0.25}$ / $70.5_{\pm0.52}$	$69.6_{\pm0.60}$ / $62.6_{\pm1.11}$	$65.5_{\pm0.05}$ / $26.0_{\pm0.44}$
FixMatch + DARP	✓	-	$85.4_{\pm0.55}$ / $85.0_{\pm0.65}$	$77.3_{\pm0.17}$ / **75.5**$_{\pm0.21}$	$72.9_{\pm0.24}$ / **69.5**$_{\pm0.18}$	$74.9_{\pm0.51}$ / $72.3_{\pm1.13}$
			(-53.1% / -73.8%)	(-13.3% / -17.0%)	(-10.9% / -18.4%)	(-31.3% / -60.3%)

Experiment result table

Table 3: Comparison of classification performance (bACC/GM) on CIFAR-10 across different distribution matching methods applied to ReMixMatch [4] under $\gamma_l = 100$.

Algorithm	$\gamma_u = 1$	$\gamma_u = 50$	$\gamma_u = 100$	$\gamma_u = 100$ (reversed)
[2]	81.4 / 80.5	76.0 / 72.7	72.5 / 67.8	72.9 / 67.1
[4]	85.0 / 84.3	77.0 / 74.7	73.8 / 69.5	75.3 / 72.3
DARP	**89.7 / 89.4**	**77.4 / 75.0**	**75.8 / 72.6**	**80.1 / 78.5**

estimation of $\{M_k\}_{k=1}^K$. In Table 2, one can observe that DARP consistently improves all the baselines. Surprisingly, the relative error gain from DARP increases as γ_u decreases, i.e., the overall class-distribution becomes more balanced. We believe that this is because SSL algorithms without DARP cannot fully enjoy this more balanced distribution as their pseudo-labels of minority class data are significantly biased toward majority classes. Meanwhile, DARP correctly refines pseudo-labels to (approximately) follow the true class-distribution, and hence it can take advantage of a more balanced class distribution of unlabeled dataset.

To further investigate this phenomenon, we also evaluate algorithms for unlabeled dataset with reversely ordered class-distribution, i.e., $M_1 \leq \cdots \leq M_K$ and $M_k = M_1 \cdot \gamma_u^{-\frac{k-1}{K-1}}$ for $\gamma_u = 100$, denoted by "$\gamma_u = 100$ (reversed)" in Table 2. As expected, SSL algorithms fail as they provide wrong pseudo-labels to the most of unlabeled data (which are majority in unlabelded data while minority in labeled data). In contrast, DARP successfully refines theses pseudo-labels and significantly improves baselines as in prior experiments. For example, DARP exhibits 19.5%/22.5% relative error gains of bACC/GM compared to the second-best method ReMixMatch* under $\gamma_u = 100$ (reversed).

Result Discussion

Core tasks in table understanding

Structure	Table segmentation	Cell role prediction	Functional block detection	Joint identification

Pujara et al. From Tables to Knowledge (KDD21).

Core tasks in table understanding

Structure	Table segmentation	Cell role prediction	Functional block detection	Joint identification
Knowledge Alignment	Semantic Typing	Semantic modeling	Schema mapping	Entity linking
	Representation	Completion	Knowledge extraction	Augmentation

Pujara et al. [From Tables to Knowledge (KDD21).]()

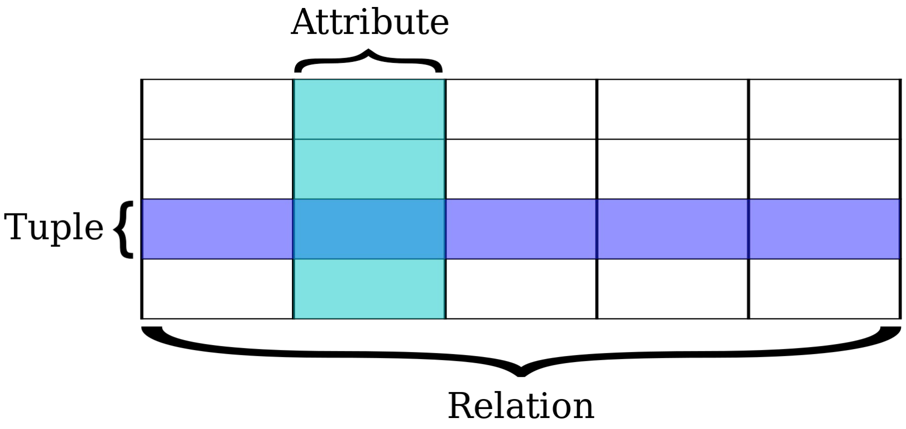
Core tasks in table understanding

Structure	Table segmentation	Cell role prediction	Functional block detection	Joint identification
Knowledge Alignment	Semantic Typing	Semantic modeling	Schema mapping	Entity linking
	Representation	Completion	Knowledge extraction	Augmentation
Downstream	Retrieval	Summarization / Text generation	Question answering	Verification

Pujara et al. [From Tables to Knowledge (KDD21).]

Table Structure Types

Relational Table (Database)



Name	Dry/Wet Food	Good Boy (Y/N)
Fido	Dry	Y
Rex	Wet	N
Bubbles	Dry	Y
Cujo	Wet	N

Tag #	Height (in)	Weight (lbs)
1573	15	21
2684	9	7
3795	27	130
4806	6	5

Tag #	Name	Breed	Color	Age
1573	Fido	Beagle	Brown/White	1.5
2684	Rex	Pekingese	White	9
3795	Bubbles	Rottweiler	Black	5
4806	Cujo	Chihuahua	Gold	4

Flat Table

	2006	2005	2004
Weighted average fair value of options granted	$20.01	$9.48	$7.28
Expected volatility	0.3534	0.3224	0.3577
Distribution yield	1.00%	0.98%	1.30%
Expected life of options in years	6.3	6.3	6.3
Risk-free interest rate	5%	4%	4%

Can be stored as 2-D list

Hierarchical Table

Table 1: Number of acceptances by category of work in conferences[1]

Year	**Conference**	**Papers**		**Workshops**	**Tutorials**	**Total**
		Research	*Industry*			
2021	KDD	800	500	25	40	1365
2021	WWW	700	400	25	40	1165
2020	KDD	600	300	25	35	960
2020	WWW	500	200	25	35	760

1. The source of this data is Jay's made-up data generation

Pujara, et al. [From Tables to Knowledge (KDD21).]

Parts of a table - cell, row, column

Table 1: Number of acceptances by category of work in conferences[1]					
Year	**Conference**	**Papers**		**Workshops**	**Tutorials**
		Research	*Industry*		
2021	KDD	800	500	25	40
row					
2020	KDD	600	300	cell	35
2020	WWW	500	200	25	35

1. The source of this data is Jay's made-up data generation

column

Parts of a table in dataset annotations

Metadata

Table 1: Number of acceptances by category of work in conferences[1]

Header

Left Attr.

Derived

Year	Conference	Papers		Workshops	Tutorials	Total
		Research	*Industry*			
2021	KDD	800	500	25	40	1365
2021	WWW	700	400	25	40	1165
2020	KDD	600	300	25	35	960
2020	WWW	500	200	25	35	760

Notes

1. The source of this data is Jay's made-up data generation

Data

Parts of a table from a knowledge perspective

The number of acceptances in the workshops category at the KDD conference for the year 2020 was 25 based on the source Jay's made-up data..

Table 1: Number of acceptances item by category of work property in conferences[1]

Year	Conference	Papers		Workshops	Tutorials	Total
properties		*Research*	*Industry*	attribute		
2021	KDD	800	500	25	40	1365
2021	WWW	700	400	25	40	1165
2020 attributes	KDD	600	300	25 value	35	960
2020	WWW	500	200	25	35	760

1. The source of this data property is Jay's made-up data generation attribute

Outline

- Joint representation learning for tables and text
- Diverse downstream tasks over tabular (& textual) data
- Discussion about selected papers
- Discussion

Joint Representation Learning for Tables and Text

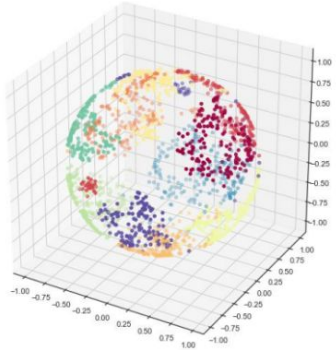
Tables

Rank	Title	Sales	Platform(s)
1	*Minecraft*	200,000,000	Multi-platform
2	*Grand Theft Auto V*	135,000,000	Multi-platform
3	*Tetris* (EA)	100,000,000	Mobile
4	*Wii Sports*	82,900,000	Wii
5	*PlayerUnknown's Battlegrounds*	70,000,000	Multi-platform
6	*Super Mario Bros.*	48,240,000	Multi-platform
7	*Pokémon Red / Green / Blue / Yellow*	47,520,000	Multi-platform

Natural Language

should be at the top of the list. When using the BEUrRE(rule+) model, the top 10 in all entities are *place, town, bed, school, city, home, house, capital, church, camp*, which are general concepts. Among the observed objects of the *atLocation* predicate, the entities that have the least coverage are *Tunisia, Morocco, Algeria, Westminster, Veracruz, Buenos Aires, Emilia-Romagna, Tyrrhenian sea, Kuwait, Serbia*. Those entities are very specific locations. This observation confirms that the box volume effectively represents probabilistic semantics and captures specificity/granularity of concepts, which we believe to be a reason for the performance improvement.

Goal



Joint (latent) representation

Relevance between NL and tabular content

Pujara et al. [From Tables to Knowledge (KDD21).]()

Common challenges for connecting tables and Natural language

- Handling diverse table layout structures
    - Relational table
    - Entity table
    - Matrix table
    - Nested table
    - ...

Lake	Area
Windermere	5.69 sq mi
Ullswater	3.86 sq mi
Derwent Water	2.06 sq mi

(a) Relational table

Country	United States
State	California
County	Los Angeles
Region	South California

(b) Entity table

	Right-handed	Left-handed
Males	43	9
Females	44	4
Totals	87	12

(c) Matrix table

		To		
		Solid	Liquid	Gas
From	Solid	Solid trans	Melting	Sublimation
	Liquid	Freezing	-	Boiling
	Gas	Deposition	Condensation	-

(d) Nested table

Pujara et al. [From Tables to Knowledge (KDD21).]

Common challenges for connecting tables and Natural language

- Precise table-text alignment rarely exists

Algorithm	SSL	RB	CIFAR-10 ($\gamma_l = 100$) $\gamma_u = 1$	$\gamma_u = 50$	$\gamma_u = 150$	$\gamma_u = 100$ (reversed)
Vanilla	-	-	$58.8_{\pm0.13}$ / $51.0_{\pm0.11}$	$58.8_{\pm0.13}$ / $51.0_{\pm0.11}$	$58.8_{\pm0.13}$ / $51.0_{\pm0.11}$	$58.8_{\pm0.13}$ / $51.0_{\pm0.11}$
Re-sampling [21]	-	✓	$55.8_{\pm0.47}$ / $45.1_{\pm0.30}$	$55.8_{\pm0.47}$ / $45.1_{\pm0.30}$	$55.8_{\pm0.47}$ / $45.1_{\pm0.30}$	$55.8_{\pm0.47}$ / $45.1_{\pm0.30}$
LDAM-DRW [8]	-	✓	$62.8_{\pm0.17}$ / $58.9_{\pm0.60}$	$62.8_{\pm0.17}$ / $58.9_{\pm0.60}$	$62.8_{\pm0.17}$ / $58.9_{\pm0.60}$	$62.8_{\pm0.17}$ / $58.9_{\pm0.60}$
cRT [22]	-	✓	$63.2_{\pm0.45}$ / $59.9_{\pm0.40}$	$63.2_{\pm0.45}$ / $59.9_{\pm0.40}$	$63.2_{\pm0.45}$ / $59.9_{\pm0.40}$	$63.2_{\pm0.45}$ / $59.9_{\pm0.40}$
VAT [30]	✓	-	$65.2_{\pm0.12}$ / $59.5_{\pm0.26}$	$64.0_{\pm0.31}$ / $57.3_{\pm0.66}$	$62.8_{\pm0.19}$ / $55.1_{\pm0.70}$	$59.4_{\pm0.36}$ / $50.6_{\pm0.61}$
Mean-Teacher [38]	✓	-	$73.9_{\pm1.19}$ / $71.7_{\pm1.42}$	$61.2_{\pm0.51}$ / $53.5_{\pm0.84}$	$59.7_{\pm0.50}$ / $50.0_{\pm1.61}$	$61.0_{\pm0.82}$ / $56.4_{\pm1.64}$
MixMatch [5]	✓	-	$41.5_{\pm0.76}$ / $12.0_{\pm1.34}$	$64.1_{\pm0.58}$ / $48.3_{\pm0.70}$	$65.5_{\pm0.64}$ / $51.1_{\pm2.41}$	$47.9_{\pm0.09}$ / $20.5_{\pm0.85}$
MixMatch + DARP	✓	-	$86.7_{\pm0.80}$ / $86.2_{\pm0.82}$	$68.3_{\pm0.47}$ / $62.2_{\pm1.21}$	$66.7_{\pm0.25}$ / $58.8_{\pm0.42}$	$72.9_{\pm0.24}$ / $71.0_{\pm0.32}$
			(-77.2% / -84.4%)	(-11.8% / -27.0%)	(-3.62% / -15.7%)	(-48.0% / -63.6%)
ReMixMatch [4]	✓	-	$48.3_{\pm0.14}$ / $19.5_{\pm0.85}$	$75.1_{\pm0.43}$ / $71.9_{\pm0.77}$	$72.5_{\pm0.10}$ / $68.2_{\pm0.32}$	$49.0_{\pm0.55}$ / $17.1_{\pm1.48}$
ReMixMatch*	✓	-	$85.0_{\pm1.35}$ / $84.3_{\pm1.55}$	$77.0_{\pm0.12}$ / $74.7_{\pm0.04}$	$72.8_{\pm0.10}$ / $68.8_{\pm0.21}$	$75.3_{\pm0.03}$ / $72.3_{\pm0.04}$
ReMixMatch* + DARP	✓	-	$\mathbf{89.7}_{\pm0.15}$ / $\mathbf{89.4}_{\pm0.17}$	$\mathbf{77.4}_{\pm0.22}$ / $75.0_{\pm0.25}$	$\mathbf{73.2}_{\pm0.11}$ / $69.2_{\pm0.31}$	$\mathbf{80.1}_{\pm0.11}$ / $\mathbf{78.5}_{\pm0.17}$
			(-31.4% / -32.5%)	(-1.72% / -1.49%)	(-1.53% / -2.64%)	(-19.5% / -22.5%)
FixMatch [37]	✓	-	$68.9_{\pm1.95}$ / $42.8_{\pm8.11}$	$73.9_{\pm0.25}$ / $70.5_{\pm0.52}$	$69.6_{\pm0.60}$ / $62.6_{\pm1.11}$	$65.5_{\pm0.05}$ / $26.0_{\pm0.44}$
FixMatch + DARP	✓	-	$85.4_{\pm0.55}$ / $85.0_{\pm0.65}$	$77.3_{\pm0.17}$ / $\mathbf{75.5}_{\pm0.21}$	$72.9_{\pm0.24}$ / $\mathbf{69.5}_{\pm0.18}$	$74.9_{\pm0.51}$ / $72.3_{\pm1.13}$
			(-53.1% / -73.8%)	(-13.3% / -17.0%)	(-10.9% / -18.4%)	(-31.3% / -60.3%)

estimation of $\{M_k\}_{k=1}^{K}$. In Table 2, one can observe that DARP consistently improves all the baselines. Surprisingly, the relative error gain from DARP increases as γ_u decreases, i.e., the overall class-distribution becomes more balanced. We believe that this is because SSL algorithms without DARP cannot fully enjoy this more balanced distribution as their pseudo-labels of minority class data are significantly biased toward majority classes. Meanwhile, DARP correctly refines pseudo-labels to (approximately) follow the true class-distribution, and hence it can take advantage of a more balanced class distribution of unlabeled dataset.

To further investigate this phenomenon, we also evaluate algorithms for unlabeled dataset with reversely ordered class-distribution, i.e., $M_1 \le \cdots \le M_K$ and $M_k = M_1 \cdot \gamma_u^{-\frac{k-1}{K-1}}$ for $\gamma_u = 100$, denoted by "$\gamma_u = 100$ (reversed)" in Table 2. As expected, SSL algorithms fail as they provide wrong pseudo-labels to the most of unlabeled data (which are majority in unlabelded data while minority in labeled data). In contrast, DARP successfully refines theses pseudo-labels and significantly improves baselines as in prior experiments. For example, DARP exhibits 19.5%/22.5% relative error gains of bACC/GM compared to the second-best method ReMixMatch* under $\gamma_u = 100$ (reversed).

Common challenges for connecting tables and Natural language

- Handling hierarchical structures of tables

hierarchical column headers

hierarchical row headers

	Year Ended December 31			
	2018		2017	
Segment	Sales	Expenses	Sales	Expenses
Innovation Systems				
Product	2,894	2,582	—	—
Service	382	351	—	—
Aerospace Systems				
Product	11,087	9,889	10,064	8,988
Service	2,009	1,796	2,067	1,854
Mission Systems				
Product	7,329	6,335	7,012	6,088
Service	4,380	3,854	4,458	3,940

1. Template based linearization

- Transform each row into sentence using template
- Simpler version of surface realization

Example: **The** *black / African American* **of** *Non-hospitalized* **is** *15927 (28.7%)*

Race/Ethnicity	Non-hospitalized	Non-fatal hospitalized	Confirmed deaths[1]	Probable deaths[2]	Total deaths
Black/African American	15927 (28.7%)	9432 (34.7%)	4239 (29.4%)	1330 (32.9%)	5569 (30.2%)

- Pros:
    - Effective for fine-tuning / small dataset
- Cons:
    - Lose the cell alignment information
    - For Large / complex table, might exceed LMs maximum input size

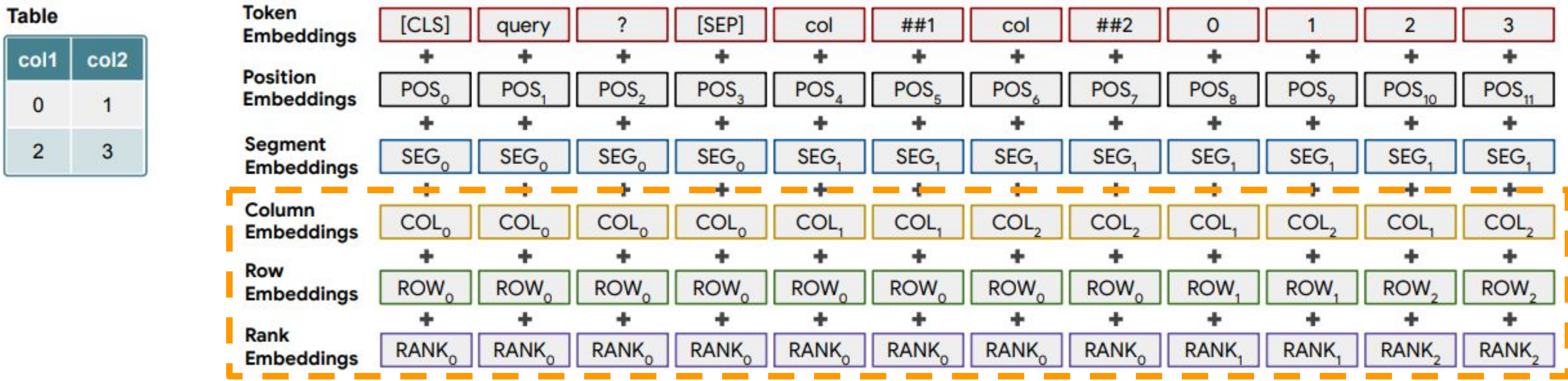
Table LMs - Discussion Questions -1

- How to adopt BERT to table pre-training?

2. Pre-training on language models (LMs) for structured tables

- Design additional embedding layers / TAGs
- Design additional pre-training objectives
- Redesign transformer structure

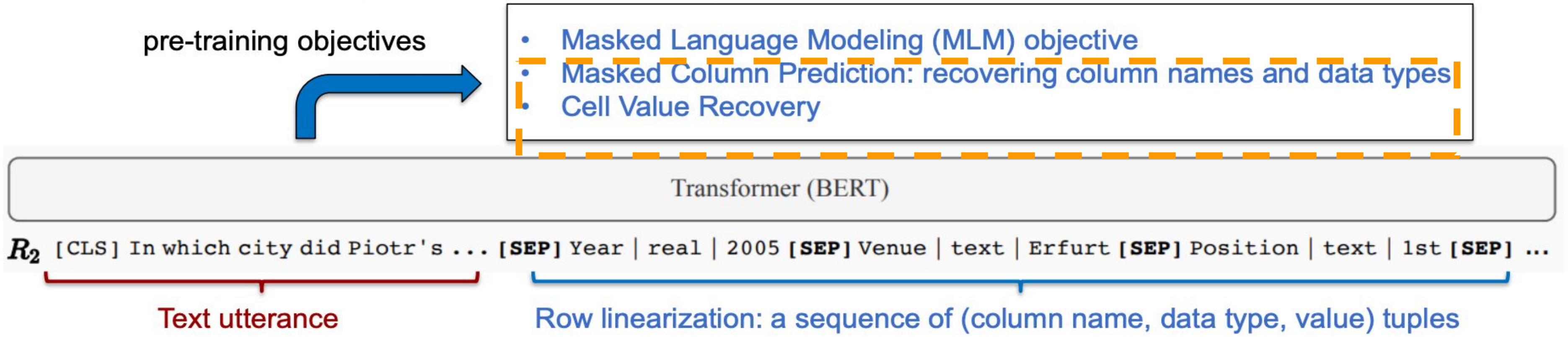
Additional embedding (TaPas)



Rank ID if column values can be parsed as floats or dates, sort them accordingly and assign an embedding based on their numeric rank. This can assist the model when processing questions that involve superlatives, as word pieces may not represent numbers informatively.

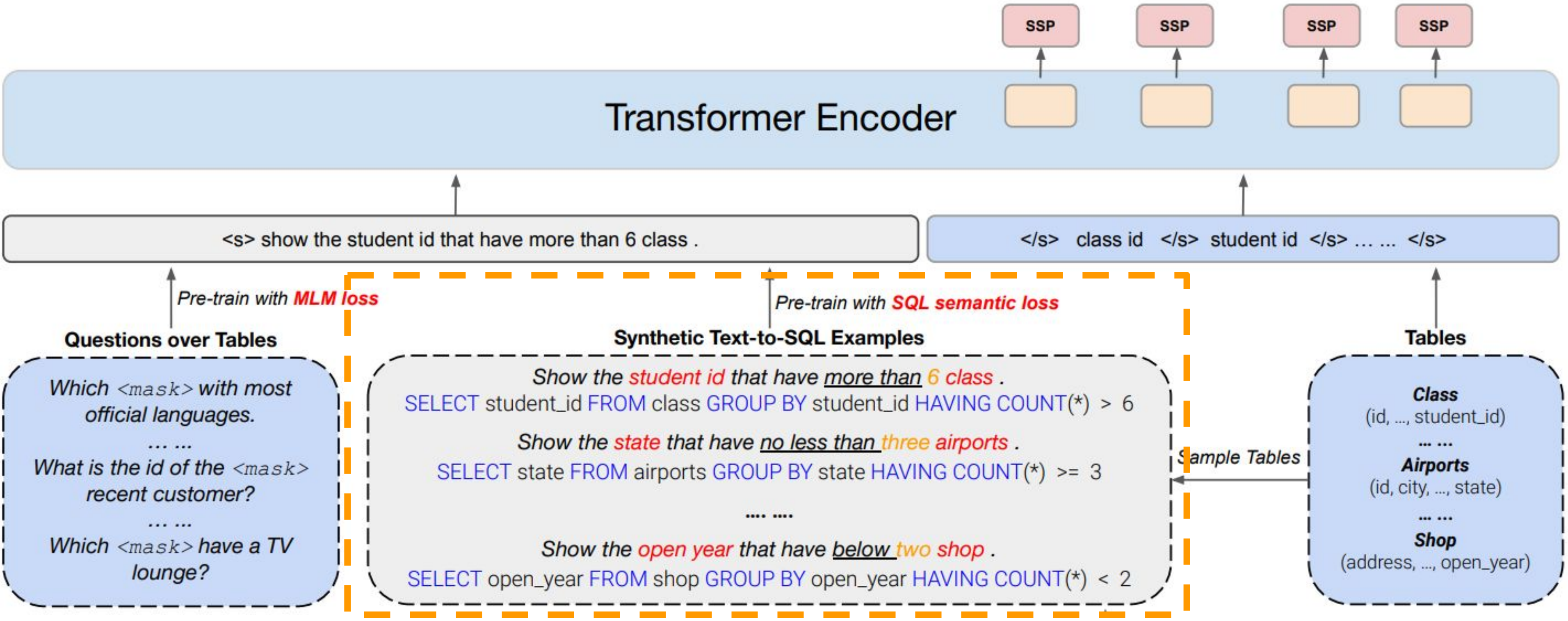
Herzig et al. [TaPas: Weakly supervised table parsing via pre-training]

Additional pre-training objectives (TaBERT)



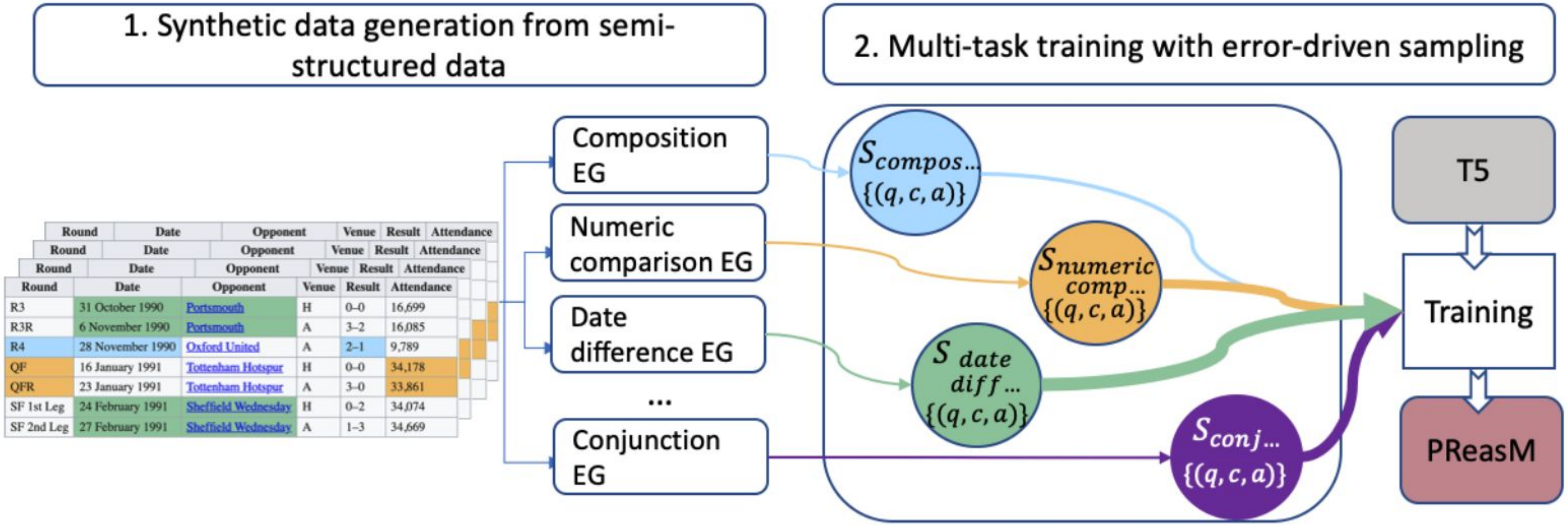
Yin et al. [TaBERT: Pretraining for Joint Understanding of Textual and Tabular Data.]

Additional pre-training objectives + synthetic data (GRAPPA)



Yu et al. [Grammar-Augmented Pre-Training for Table Semantic Parsing]()

Synthesize examples for data augmentation



Yoran O, Talmor A, Berant J. [Turning tables: Generating examples from semi-structured tables for endowing language models with reasoning skills]

Redesign transformer structure (TUTA)

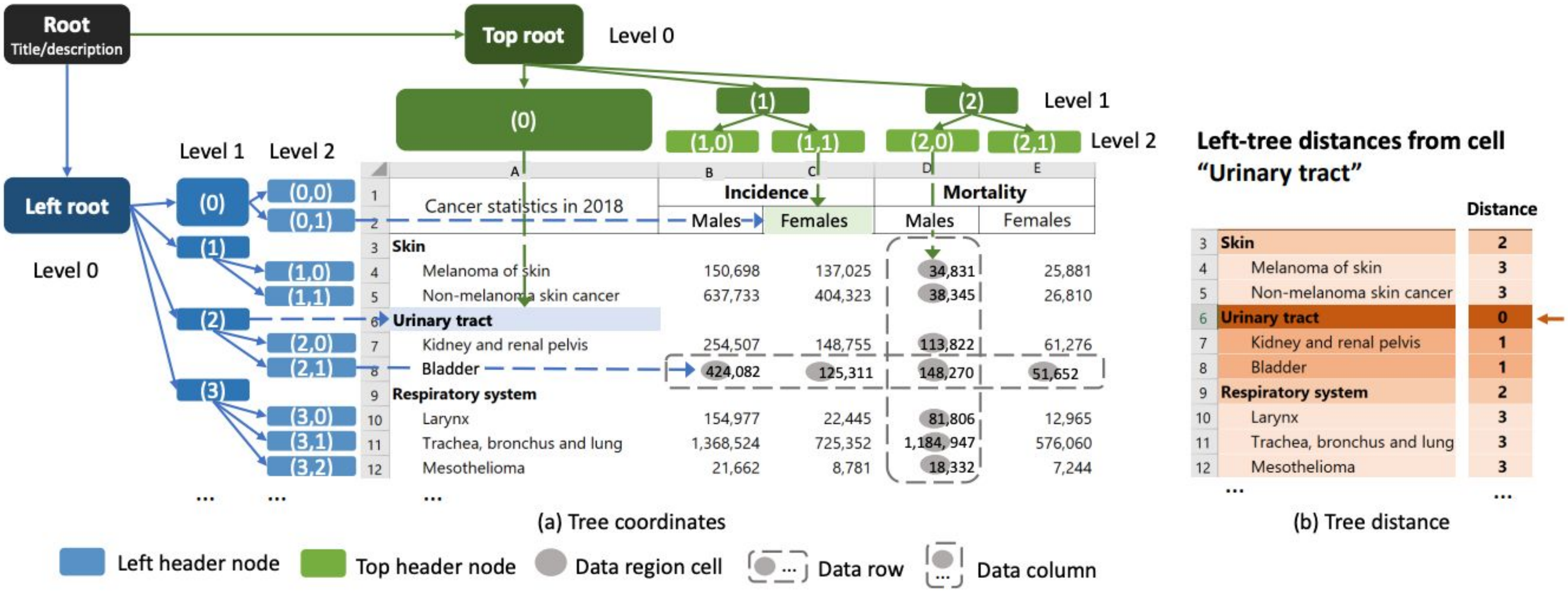
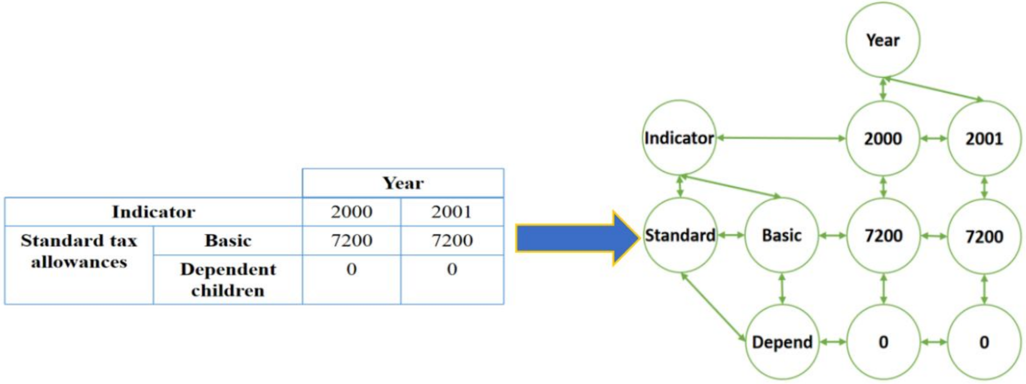
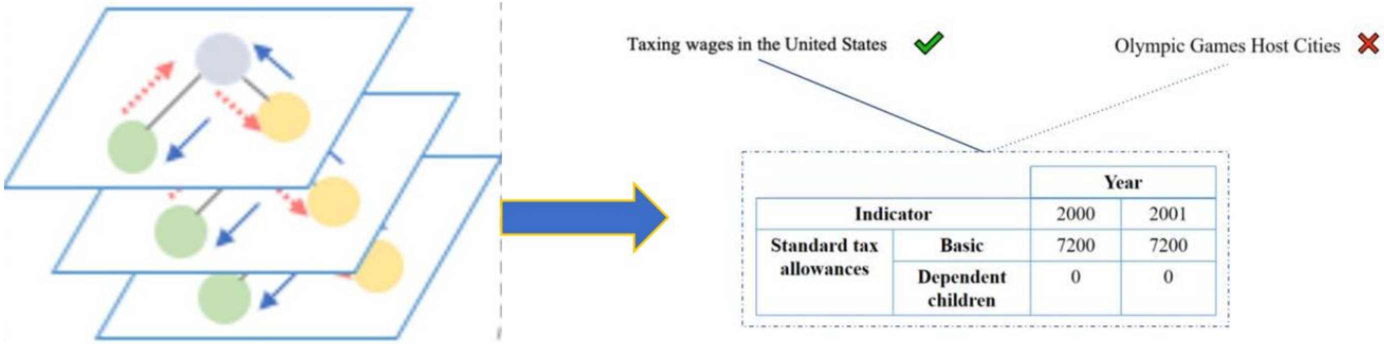


Figure 2: An example to illustrate our proposed tree coordinates and tree distance for generally structured tables. In this example, both the top tree and the left tree contain three levels. Cell "A6" ('Urinary tract') is the 'parent' node of cells "A7" and "A8", and has 'brothers' including "A3" and "A9".

Wang et al. [TUTA: Tree-based Transformers for Generally Structured Tables]

3. Graph Representation Learning





Zhang et al. [A Graph Representation of Semi-structured Data for Web Question Answering.]() COLING-20
Wang et al. [Retrieving Complex Tables with Multi-Granular Graph Representation Learning](). SIGIR-21

Diverse Downstream Tasks

Introduction of downstream Tasks

- Semantic Table Retrieval
- **Table-based Fact Verification**
- Question Answering
  - **Numerical Reasoning**
  - Single / Multi-hop Reasoning
  - Semantic parsing (text to SQL)
- Table to Text Generation

Table Retrieval

Changes of taxes in U.S.?

Taxing wages in the United States

Indicator		Year 2000	Year 2001
Standard tax allowances	Basic	7200	7200
	Dependent children	0	0

Olympic Games Host Cities

City	Country	Year	Continent
Los Angeles	U.S.	2028	North America
Milan–Cortina d'Ampezzo	Italy	2026	Europe
Paris	France	2024	
Beijing	China	2022	Asia

Input:

- A natural language query
- A set of tables, where each table consists of:
  - table body (headers, data cells, etc.)
  - context (captions, footnotes, etc.)

Output:

- A ranked list of semantically relevant tables

Table-based fact verification

United States House of Representatives Elections, 1972

District	Incumbent	Party	Result	Candidates
California 3	John E. Moss	democratic	re-elected	John E. Moss (d) 69.9% John Rakus (r) 30.1%
California 5	Phillip Burton	democratic	re-elected	Phillip Burton (d) 81.8% Edlo E. Powell (r) 18.2%
California 8	George Paul Miller	democratic	lost renomination democratic hold	Pete Stark (d) 52.9% Lew M. Warden , Jr. (r) 47.1%
California 14	Jerome R. Waldie	republican	re-elected	Jerome R. Waldie (d) 77.6% Floyd E. Sims (r) 22.4%
California 15	John J. Mcfall	republican	re-elected	John J. Mcfall (d) unopposed

Entailed Statement

1. John E. Moss and Phillip Burton are both re-elected in the house of representative election.
2. John J. Mcfall is unopposed during the re-election.
3. There are three different incumbents from democratic.

Refuted Statement

1. John E. Moss and George Paul Miller are both re-elected in the house of representative election.
2. John J. Mcfall failed to be re-elected though being unopposed.
3. There are five candidates in total, two of them are democrats and three of them are republicans.

1. **Table retrieval:** finding evidence table(s)
2. **NLI:** textual entailment using the table as premise and the statement as hypothesis

Chen et al. [TabFact: A Large-scale Dataset for Table-based Fact Verification.] ICLR-20

Numerical question answering

Page 91 from the annual reports of GRMN (Garmin Ltd.)

The fair value for these options was estimated at the date of grant using a Black-Scholes option pricing model with the following weighted-average assumptions for 2006, 2005 and 2004:

	2006	2005	2004
Weighted average fair value of options granted	$20.01	$9.48	$7.28
Expected volatility	0.3534	0.3224	0.3577
Distribution yield	1.00%	0.98%	1.30%
Expected life of options in years	6.3	6.3	6.3
Risk-free interest rate	5%	4%	4%

… The total fair value of shares vested during 2006, 2005, and 2004 was $9,413, $8,249, and $6,418 respectively. The aggregate intrinsic values of options outstanding and exercisable at December 30, 2006 were $204.1 million and $100.2 million, respectively. *(… abbreviate 10 sentences ...)*

Question: Considering the weighted average fair value of options , what was the change of shares vested from 2005 to 2006?

Answer: - 400

Calculations:

$$\left(\frac{9413}{20.01}\right) - \left(\frac{8249}{9.48}\right) = -400$$

Program:

divide (9413, 20.01)

divide (8249, 9.48)

substract (#0, #1)

Chen et al. [FINQA: A Dataset of Numerical Reasoning over Financial Data.]

Single-hop Table QA

Question

What languages do people in France speak?

Table Database from the Web

Country	Capital	Location	Main Language	Currency
Algeria	Algiers	Africa	Arabic, French	Dinar
France	Paris	Europe	French	Euro
Hungary	Budapest	Europe	Hungarian	Forint
Singapore	Singapore	Asia	Malay, Chinese, Tamil	Singapore Dollar

Answer

French

Evidence

Country	Main Language
France	French

Source: http://hasibul.info/gk/countries.php

Liang et al. [Compositional Semantic Parsing on Semi-Structured Tables.]()

Multi-hop Hybrid QA

Name	Year	Season	Flag bearer
XXXI	2016	Summer	Yan Naing Soe
XXX	2012	Summer	Zaw Win Thet
XXIX	2008	Summer	Phone Myint Tayzar
XXVIII	2004	Summer	Hla Win U
XXVII	2000	Summer	Maung Maung Nge
XX	1972	Summer	Win Maung

Q: When does the oldest flag Burmese bearer participate in Olympic ceremony.

Chen et al. [HybridQA: A Dataset of Multi-Hop Question Answering over Tabular and Textual Data.]

Multi-hop Hybrid QA

Multimodal Context

Steal This Movie!
The film follows Hoffman's (D'Onofrio) relationship with his second wife Anita (Garofalo) and their "awakening" and subsequent conversion to an activist life. The title of the film is a play on Hoffman's 1970 counter-culture guidebook titled "Steal This Book".

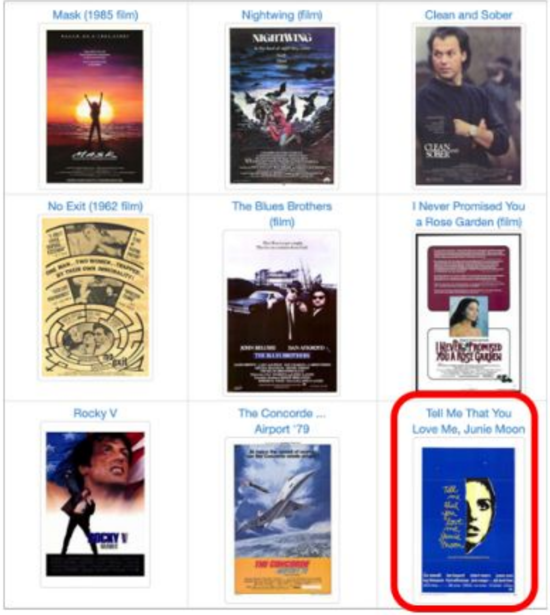
La liceale
La liceale (internationally released as The Teasers, "Under-graduate Girls", "Sophomore Swingers" and "Teasers") is a 1975 commedia sexy all'italiana directed by Michele Massimo Tarantini.
...
Guida. It was followed by "La liceale nella classe dei ripetenti".

Sage Stallone
Stallone made his acting debut alongside his father in Rocky V (1990), the fifth installment of the Rocky franchise, playing Robert Balboa Jr., the onscreen son of his father's title character. He did not, however,
...
After that, he acted in lesser profile films.

Pierino contro tutti
Pierino contro tutti (also known as "Desirable Teacher") is a 1981 comedy film directed by Marino Girolami. The main character of the film is Pierino, an
...
l as a short lived subgenre of joke-films in which the plot basically consists of a series of jokes placed side by side.

Ben Piazza - Filmography

Year	Title	Role
1957	A Dangerous Age	David
1959	The Hanging Tree	Rune
1962	No Exit	Camarero
1970	**Tell Me That You Love Me, Junie Moon**	Jesse
1972	The Outside Man	Desk Clerk
...	...	...
1985	Mask	Mr. Simms
1988	Clean and Sober	Kramer
1990	**Rocky V**	Doctor
1991	Guilty by Suspicion	Darryl Zanuck

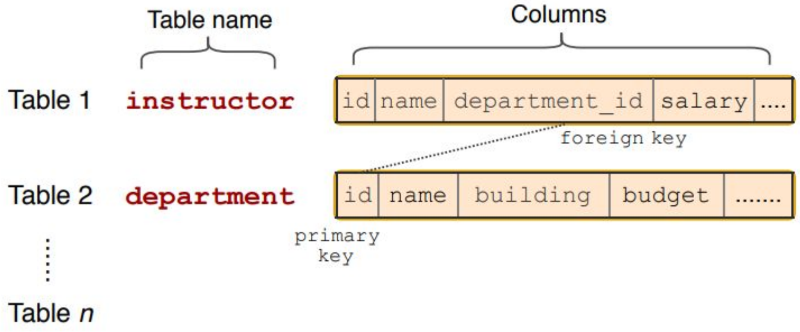


Q: Which B. Piazza title came earlier: the movie S. Stallone's son starred in or the movie with half of a lady's face on the poster?
A: Tell Me That You Love Me, Junie Moon

Chen et al. MultiModalQA: complex question answering over text, tables and images.

Semantic Parsing (text to SQL)

Annotators check database schema (e.g., database: `college`)



Complex question

What are the name and budget of the departments with average instructor salary greater than the overall average?

Complex SQL

```
SELECT T2.name, T2.budget
FROM instructor as T1 JOIN department as
T2 ON T1.department_id = T2.id
GROUP BY T1.department_id
HAVING avg(T1.salary) >
    (SELECT avg(salary) FROM instructor)
```

Yu et al. Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task.

Table-to-text generation

1. Surface-level realization
2. Logical text generation

Medal Table from Tournament

Nation	Gold Medal	Silver Medal	Bronze Medal	Sports
Canada	3	1	2	Ice Hockey
Mexico	2	3	1	Baseball
Colombia	1	3	0	Roller Skating

Surface-level Generation

Sentence: Canada has got 3 gold medals in the tournament.
Sentence: Mexico got 3 silver medals and 1 bronze medal.

Logical Natural Language Generation

Sentence: Canada obtained 1 more gold medal than Mexico.
Sentence: Canada obtained the most gold medals in the game.

Chen et al. Logical Natural Language Generation from Open-Domain Tables

SciGen

	ellipsis (infl.)	ellipsis (VP)
baseline	53.0	28.4
concat	**76.2**	76.6
s-hier-to-2.tied	66.4	65.6
CADec	72.2	**80.0**

Table 8: Accuracy on ellipsis test set.

Setting	Description
Gold	For ellipsis, both models improve substantially over the baseline (by 19-51 percentage points), with concat stronger for inflection tasks and CADec stronger for VPellipsis.
Few	table 8 shows the performance of our model on the ellipsis test set. our concat model outperforms all the other models that do not use concat embeddings.
Med.	the results show that both concat and cadec significantly outperform the baseline and s-hier-to-2.tied on the ellipsis test set. the concat model is better than the baseline on both inflected and verb-inflated forms, while cadec is better on the vp version. on the inflated version, cadec outperforms the baseline by a large margin and is within 3 points of the best model. the model with the best accuracy on the verb portion of the test set (80.0%) is very close to the best performance on the noun portion (75.6%). the results indicate that cadec has better generalizability than the other models.
Large	the concatenation model concatenates the output of the concatenated encoder and decoder, and the s-hier-to-2.tied model performs slightly better than concat on both ellipsis (infl.) and vice versa. cadec outperforms the other two models by a large margin, achieving 80.0% accuracy on the vp ellipsoids, which is the new state of the art.

Table 7: Sample outputs from the BART model on few-shot, medium, and large settings on the corresponding table from Figure 2. Factually correct statements are marked with green, factually incorrect statements and hallucinations are marked with red and blue, respectively.

Moosavi et al. Learning to Reason for Text Generation from Scientific Tables (NIPS 2021 benchmark track)

LogicNLG - coarse-to-fine generation

Medal Table from Tournament

Nation	Gold Medal	Silver Medal	Bronze Medal	Sports
Canada	3	1	2	Ice Hockey
Mexico	2	3	1	Baseball
Colombia	1	3	0	Roller Skating

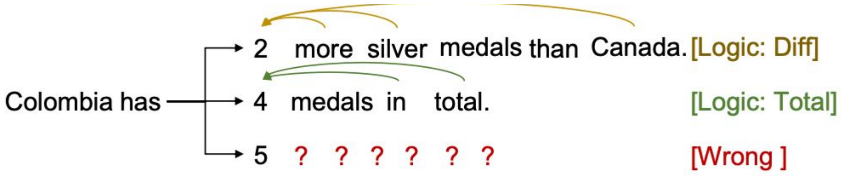
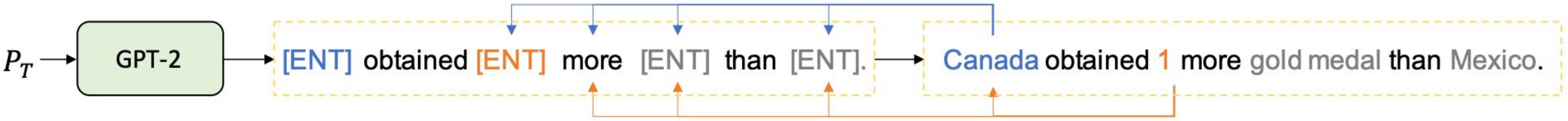


Figure 2: When making the decision at the third step, the model needs to foresee the future tokens to ensure logical consistency. There is no back-tracking once the model makes a wrong decision like “5”.



Discussion About Selected Papers

Paper 1

FINQA: A Dataset of Numerical Reasoning over Financial Data (EMNLP 2021)

Zhiyu Chen[1], **Wenhu Chen**[1], **Charese Smiley**[2], **Sameena Shah**[2], **Iana Borova**[1], **Dylan Langdon**[1], **Reema Moussa**[1], **Matt Beane**[1], **Ting-Hao Huang**[3], **Bryan Routledge**[4] and **William Yang Wang**[1]

[1]University of California, Santa Barbara
[2]J.P. Morgan
[3]Pennsylvania State University
[4]Carnegie Mellon University
{zhiyuchen,william}@cs.ucsb.edu

Related work - Math word problem solver

Context and Question

If Lily's test scores are 85 , 89 , 80 and 95 out of 100 in 4 different subjects , what will be her average score?

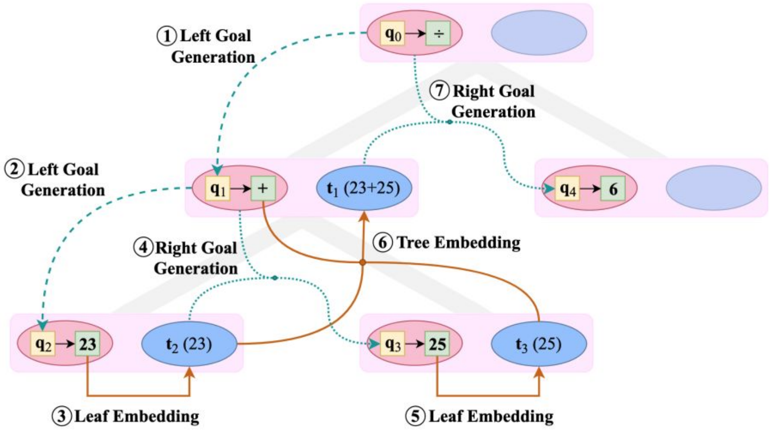
Equation

(85 + 89 + 80 + 95) / 4

Intermediate steps for solving math problem

Step 1	Step 2
a = 85 + 89	b = a + 80

Step 3	Step 4
c = b + 95	c / 4



Xie Z, Sun S. A Goal-Driven Tree-Structured Neural Model for Math Word Problems (IJCAI 2019).

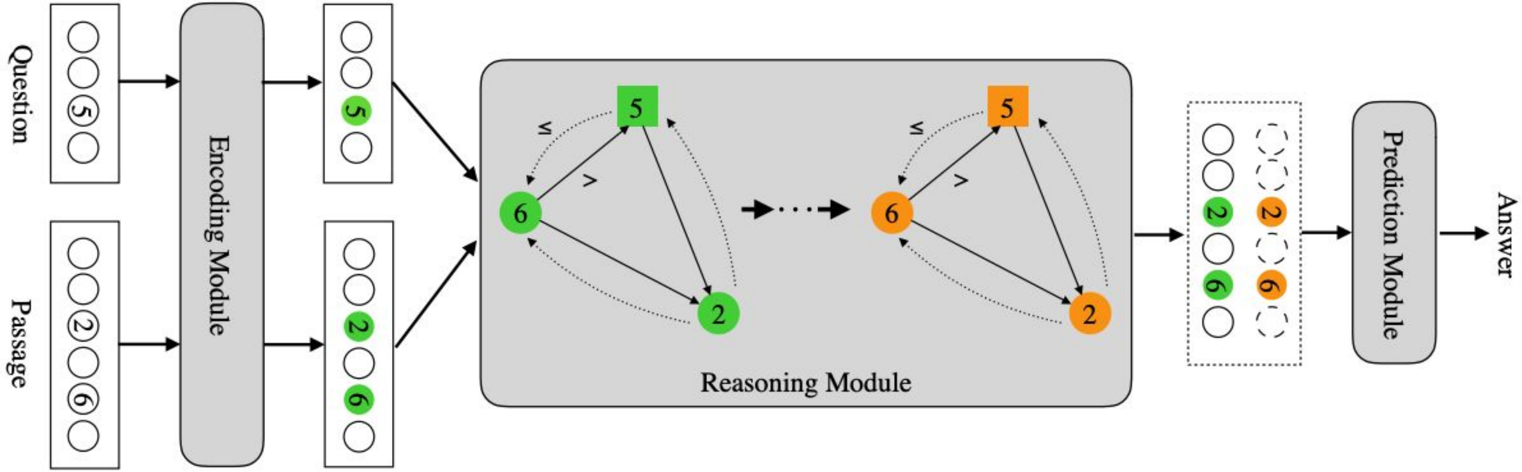
Related work - Machine Reading Comprehension

DROP dataset

Question	Passage	Answer
What is the second longest field goal made?	... The Seahawks immediately trailed on a scoring rally by the Raiders with kicker *Sebastian Janikowski nailing a 31-yard field goal* ... Then in the third quarter *Janikowski made a 36-yard field goal*. Then *he made a 22-yard field goal* in the fourth quarter to put the Raiders up 16-0 ... The Seahawks would make their only score of the game with kicker *Olindo Mare hitting a 47-yard field goal*. However, they continued to trail as *Janikowski made a 49-yard field goal*, followed by RB Michael Bush making a 4-yard TD run.	47-yard
How many age groups made up more than 7% of the population?	Of Saratoga Countys population in 2010, *6.3%* were between ages of 5 and 9 years, *6.7%* between 10 and 14 years, 6.5% between 15 and 19 years, *5.5%* between 20 and 24 years, *5.5%* between 25 and 29 years, *5.8%* between 30 and 34 years, *6.6%* between 35 and 39 years, *7.9%* between 40 and 44 years, *8.5%* between 45 and 49 years, *8.0%* between 50 and 54 years, *7.0%* between 55 and 59 years, *6.4%* between 60 and 64 years, and *13.7%* of age 65 years and over ...	5

Dua et al. [DROP: A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs.]

Related work - NumNet on DROP



Ran et al. [NumNet: Machine reading comprehension with numerical reasoning]() (EMNLP 2019).

GPT-3 demo

“Of Saratoga Countys population in 2010, 6.3% were between ages of 5 and 9 years, 6.7% between 10 and 14 years, 6.5% between 15 and 19 years, 5.5% between 20 and 24 years, 5.5% between 25 and 29 years, 5.8% between 30 and 34 years, 6.6% between 35 and 39 years, 7.9% between 40 and 44 years, 8.5% between 45 and 49 years, 8.0% between 50 and 54 years, 7.0% between 55 and 59 years, 6.4% between 60 and 64 years, and 13.7% of age 65 years and over …”

Q: How many age groups made up more than 7% of the population?

Correct answer: 5

GPT-3 demo

Race/Ethnicity	Non-hospitalized	Non-fatal hospitalized	Confirmed deaths[1]	Probable deaths[2]	Total deaths
Black/African American	15927 (28.7%)	9432 (34.7%)	4239 (29.4%)	1330 (32.9%)	5569 (30.2%)

“Among African Americans there were 5,569 total deaths, representing 30.2% of total deaths. Of the 5,569 total deaths, 1,330 deaths were probable deaths (representing 32.9% of all probable deaths) while 4,239 were confirmed deaths (representing 29.4% of all confirmed deaths). Additionally, there were 9,432 non-fatal hospitalizations (representing 34.7% of all non-fatal hospitalizations) and 15,927 cases that did not require hospitalizations (representing 28.7% of all cases where hospitalization was unnecessary).”

Q: What's the total amount of non-fatal hospitalizations for all races?

Correct answer: 9432 / 0.347

Q: Which kinds of deaths accounts the most portion to the total deaths among African Americans?

Correct answer: confirmed deaths

FINQA dataset

Page 91 from the annual reports of GRMN (Garmin Ltd.)
The fair value for these options was estimated at the date of grant using a Black-Scholes option pricing model with the following weighted-average assumptions for 2006, 2005 and 2004:

	2006	2005	2004
Weighted average fair value of options granted	$20.01	$9.48	$7.28
Expected volatility	0.3534	0.3224	0.3577
Distribution yield	1.00%	0.98%	1.30%
Expected life of options in years	6.3	6.3	6.3
Risk-free interest rate	5%	4%	4%

… The total fair value of shares vested during 2006, 2005, and 2004 was $9,413, $8,249, and $6,418 respectively. The aggregate intrinsic values of options outstanding and exercisable at December 30, 2006 were $204.1 million and $100.2 million, respectively. *(… abbreviate 10 sentences ...)*

Question: Considering the weighted average fair value of options , what was the change of shares vested from 2005 to 2006?
Answer: - 400
Calculations:

$$\left(\frac{9413}{20.01}\right) - \left(\frac{8249}{9.48}\right) = -400$$

Program:

divide (9413, 20.01)   divide (8249, 9.48)

substract (#0, #1)

Statistics:

- 8,281 financial QA pairs over 2,789 financial report pages. Only 1 flat table in each page.
- Supporting Facts coverage
  - 23.42% only text
  - 62.42% only table
  - 14.15% text + table
- Reasoning steps
  - 59.1% one step (mostly “diff / sum” question)
  - 32.71% two steps (mostly “change ratio” question)
  - 8.19% more than two steps

Annotation Process

- For each QA pair annotation:
  - Write a meaningful financial question
  - Compose a reasoning program to answer the question
  - Annotate the supporting fact

Name	Arguments	Output	Description
add	number1, number2	number	add two numbers: $number1 + number2$
subtract	number1, number2	number	subtract two numbers: $number1 - number2$
multiply	number1, number2	number	multiply two numbers: $number1 \cdot number2$
divide	number1, number2	number	multiply two numbers: $number1/number2$
exp	number1, number2	number	exponential: $number1^{number2}$
greater	number1, number2	bool	comparison: $number1 > number2$
table-sum	table header	number	the summation of one table row
table-average	table header	number	the average of one table row
table-max	table header	number	the maximum number of one table row
table-min	table header	number	the minimum number of one table row

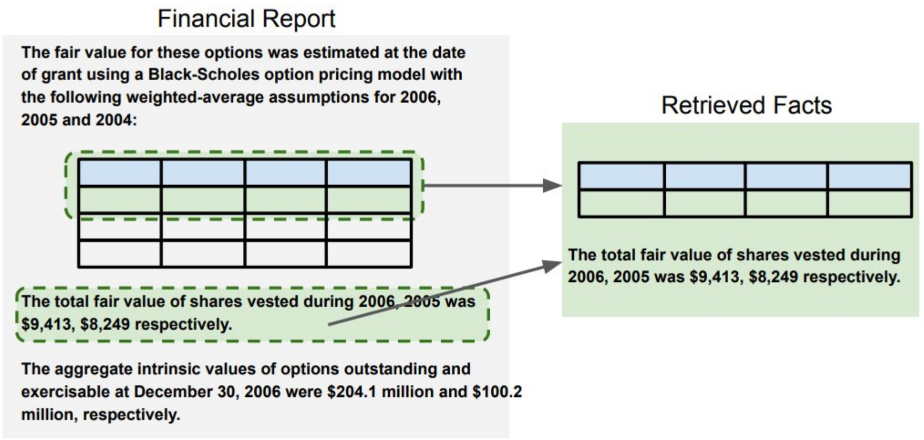
Table 4: Definitions of all operations

Discussion Questions -2

- Quality Control: For **crowd-source** annotation, how to ensure the diversity and meaningfulness of each QA examples?

Retriever - BERT-based bi-classifier

- Use template to translate each row of table into sentences
- Training Stage:
  - Input: Question + supporting fact sentence
  - Output: *Prob(True | Question + supporting fact sentence)*
- Inference Stage
  - Select Top-3 sentences with highest probability as retrieved facts



Program Generator

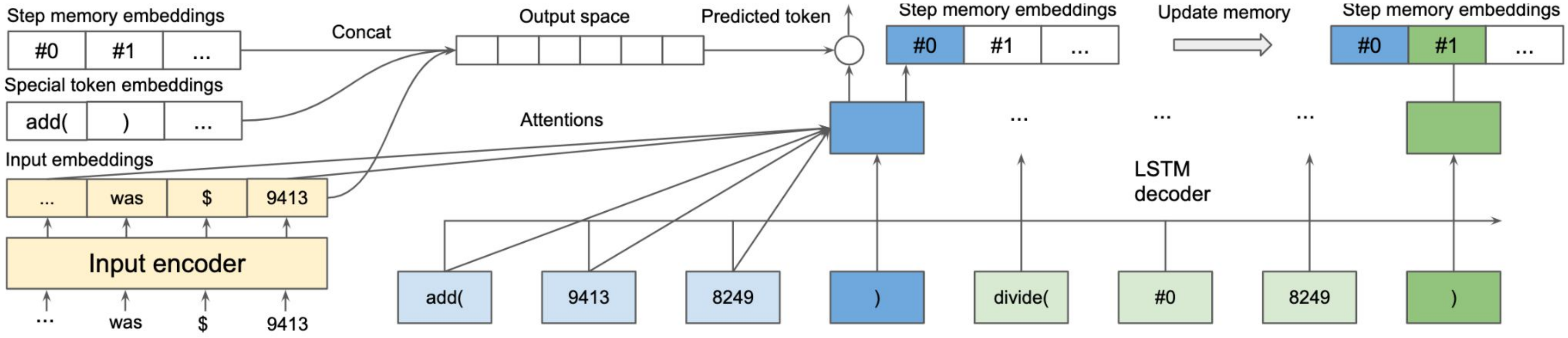


Figure 3: The program generator. The retriever results and the question are first encoded using pre-trained LMs. At each decoding step, the model can generate from the numbers or table row names from the input, the special tokens in the DSL, or the step memory tokens. At the end of the generation of each operation step, we update the step memory token embeddings.

Evaluation Metrics

- F1 score
- Execution Accuracy
  a. replace all the arguments in a program with symbols,
  b. evaluate if two symbolic programs are mathematically equivalent.

$$\text{add}(a_1, a_2), \text{add}(a_3, a_4), \text{subtract}(\#0, \#1)$$
$$\text{add}(a_4, a_3), \text{add}(a_1, a_2), \text{subtract}(\#1, \#0)$$

(a1 + a2) - (a3 + a4)

(a1 + a2) - (a4 + a3)

Results

Baselines	Exe Acc	Prog Acc
TF-IDF + Single Op	1.01	0.90
Retriever + Direct Generation	0.30	-
Pre-Trained Longformer (base)	21.90	20.48
Retriever + Seq2seq	20.40	18.29
Retriever + NeRd (BERT-base)	52.48	49.90
FinQANet (FinBert)	53.71	51.71
FinQANet (BERT-base)	54.95	53.52
FinQANet (BERT-large)	57.43	55.52
FinQANet (RoBERTa-base)	60.10	58.38
FinQANet (RoBERTa-large)	**65.05**	**63.52**
FinQANet-Gold (RoBERTa-large)	70.00	68.76
Human Expert Performance	91.16	87.49
General Crowd Performance	50.68	48.17

Methods	Exe Acc	Prog Acc
full results	**65.05**	**63.52**
Necessity of table and text		
table-only inference	41.62	40.48
text-only inference	16.38	15.33
Performances on table and text		
table-only questions	73.48	72.10
text-only questions	53.70	52.92
table-text questions	45.99	42.34
Performances regarding program steps		
1 step programs	70.27	68.77
2 step programs	63.69	61.79
>2 step programs	31.65	31.65
Programs with constants	39.80	39.80

Paper 2

Logic-level Evidence Retrieval and Graph-based Verification Network for Table-based Fact Verification (EMNLP 2021)

Qi Shi, Yu Zhang,* Qingyu Yin, Ting Liu
Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, Harbin, China
{qshi,zhangyu,qyyin,tliu}@ir.hit.edu.cn

Table-based fact verification

United States House of Representatives Elections, 1972

District	Incumbent	Party	Result	Candidates
California 3	John E. Moss	democratic	re-elected	John E. Moss (d) 69.9% John Rakus (r) 30.1%
California 5	Phillip Burton	democratic	re-elected	Phillip Burton (d) 81.8% Edlo E. Powell (r) 18.2%
California 8	George Paul Miller	democratic	lost renomination democratic hold	Pete Stark (d) 52.9% Lew M. Warden , Jr. (r) 47.1%
California 14	Jerome R. Waldie	republican	re-elected	Jerome R. Waldie (d) 77.6% Floyd E. Sims (r) 22.4%
California 15	John J. Mcfall	republican	re-elected	John J. Mcfall (d) unopposed

Entailed Statement

1. John E. Moss and Phillip Burton are both re-elected in the house of representative election.
2. John J. Mcfall is unopposed during the re-election.
3. There are three different incumbents from democratic.

Refuted Statement

1. John E. Moss and George Paul Miller are both re-elected in the house of representative election.
2. John J. Mcfall failed to be re-elected though being unopposed.
3. There are five candidates in total, two of them are democrats and three of them are republicans.

1. **Table retrieval:** finding evidence table(s)
2. **NLI:** textual entailment using the table as premise and the statement as hypothesis

Chen et al. TabFact: A Large-scale Dataset for Table-based Fact Verification. ICLR-20

Symbolic reasoning with logical operation

- Learn to parse NL statements into logical programs through **semantic parser**
- Execute the program on tables

Motivation

- Previous models only generate one or two programs for each statement
- Models may select spurious programs (i.e. wrong programs with correct returned labels) because there are weak supervised signals in the semantic parsing process.
- Additional evidence that contains logical operations is helpful to classify the entailment relation

Methodology

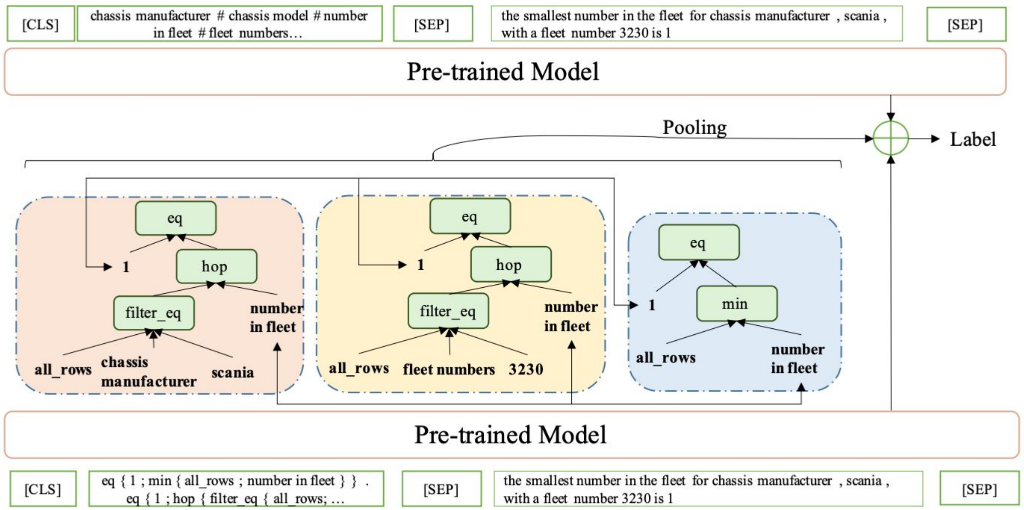
- Generate multiple programs, regards them as supplementary evidence for tables
- "Program generation - Retrieval - Verification" Pipeline
  - Logic-level Evidence Retrieval
  - Graph-based Verification Network

Logic-level Evidence Retrieval

1. Obtain Program set
    a. Given a table T and a statement S
    b. Performs entity linking to detect all the entities in the statement and link them to the table
    c. Collects a set of programs by executing sub-programs over the table and store the generated intermediate variables recursively
2. Filter the generated programs

Graph-based Verification Network

- Graph Construction and Initialization
  - Function nodes, Entity nodes
- Graph-based Verification Network
  - Graph-based Reasoning Process
  - Node Type Representations
  - Node Pruning
  - Label Prediction



Experimental results on TabFact

Model	Val	Test	Test (simple)	Test (complex)	Small Test
BERT classifier w/o Table	50.9	50.5	51.0	50.1	50.4
Table-BERT-Horizontal-F+T-Concatenate	50.7	50.4	50.8	50.0	50.3
Table-BERT-Vertical-F+T-Template	56.7	56.2	59.8	55.0	56.2
Table-BERT-Vertical-T+F-Template	56.7	57.0	60.6	54.3	55.5
Table-BERT-Horizontal-F+T-Template	66.0	65.1	79.0	58.1	67.9
Table-BERT-Horitonzal-T+F-Template	66.1	65.1	79.1	58.2	68.1
LPA-Voting w/o Discriminator	57.7	58.2	68.5	53.2	61.5
LPA-Weighted-Voting	62.5	63.1	74.6	57.3	66.8
LPA-Ranking w/ Transformer	65.2	65.0	78.4	58.5	68.6
LogicalFactChecker	71.8	71.7	85.4	65.1	74.3
HeterTFV	72.5	72.3	85.9	65.7	74.2
SAT	73.3	73.2	85.5	67.2	-
ProgVGAT	74.9	74.4	**88.3**	67.6	76.2
LERGV	**75.6**	**75.5**	87.9	**69.5**	**77.8**
Human Performance	-	-	-	-	92.1

Open Research Directions
& Discussions

Discussion Questions -3

- We mentioned 1) Semantic Table Retrieval, 2) Table-based Fact Verification, 3) QA, 4) semantic parsing, 5) logical text generation before,

  What else downstream tasks can be proposed for tabular & textual content?

Table Summarization / Dialogue

#	Stadium	Capacity	City	Country	Domed or Retractable roof	Tenant(s)
1	AT&T Stadium	80,000	Arlington, Texas	United States	RR	Dallas Cowboys (NFL)
2	Principality Stadium	74,500	Cardiff	Wales	RR	Wales National Rugby Union Team (Welsh Rugby Union)
3	Mercedes-Benz Superdome	73,208	New Orleans, Louisiana	United States	D	New Orleans Saints (NFL)

Figure 1: Table summarization in conversational search. The summary of the result table includes a leading sentence describing the table (red), answer to the question (blue), and a sentence helping to explore the table further (green).

Zhang et al. [Summarizing and exploring tabular data in conversational search (SIGIR 2020).]()

Discussion Questions - 4

- We mentioned 1) Semantic Table Retrieval, 2) Table-based Fact Verification, 3) QA, 4) semantic parsing, 5) table-to-text generation before,

  what else downstream tasks can be proposed for tabular & textual data?

- Current popular models rely on **entity linking match** to align table and text.

  E.g. “What is the type of the **document** named "David CV"?” If we replace “**document**” with “file”, which has similar semantic meaning but does not appear in table header, the model will not generate the correct answer. How to solve this problem?

TabFact demo

school year	class a	class aa	class aaa	class aaaa	class aaaaa
1992 - 93	lazbuddie	bangs	cameron yoe	dayton	abilene
1993 - 94	lazbuddie	rosebud - lott	mont belvieu barbers hill	sulphur springs	duncanville
1994 - 95	lazbuddie	rosebud - lott	mont belvieu barbers hill	rockwall	duncanville
1995 - 96	lazbuddie	rosebud - lott	mount vernon	rockwall	abilene
1996 - 97	lazbuddie	rosebud - lott	dalhart	sherman	duncanville
1997 - 98	trenton	rosebud - lott	dalhart	snyder	abilene
1998 - 99	lazbuddie	rosebud - lott	dalhart	los fresnos	abilene
1999 - 2000	lazbuddie	rosebud - lott	dalhart	sulphur springs	abilene
2000 - 01	lazbuddie	rosebud - lott	cameron yoe	snyder	abilene
2001 - 02	(unavailable)	(unavailable)	(unavailable)	(unavailable)	(unavailable)
2002 - 03	(unavailable)	(unavailable)	(unavailable)	(unavailable)	(unavailable)
2003 - 04	trenton	rosebud - lott	hamshire - fannett	cleburne	southlake carroll
2004 - 05	trenton	rosebud - lott	dalhart	brownwood	keller
2005 - 06	trenton	rosebud - lott	dalhart	brownwood	keller
2006 - 07	trenton	tuscola jim ned	hamshire - fannett	brownwood	keller
2007 - 08	happy	rosebud - lott	hamshire - fannett	brownwood	keller
2008 - 09	happy	rosebud - lott	snyder	magnolia	fort bend dulles
2009 - 10	happy	grandview	giddings	granbury	laredo united
2010 - 11	happy	rosebud - lott	giddings	granbury	fort bend dulles

Wikipedia Title : accounting (uil)

URL : 2-14630796-2.html.csv

Note: Entailed Statements are in RED, Refuted statements are in GREEN

rosebud - lott be class aa , with a class aaaa of sulphur spring and class aaaaa of duncanville

in the 2010 - 11 school year , rosebud - lott be the class aa with a class aaa of giddings

in the school year 2009 - 10 , grandview be the class aa with a class aaa of giddings

of the school year 1993 - 94 , duncanville be the class aaaaa with a class aaa of mont belvieu barber hill

the school year 1993 - 94 and 1994 - 95 with a class aa of rosebud - lott , class aaaaa of duncanville , and a class aaa of mont belvieu barber hill

duncanville as the class aaaaa , with a class aaaa of dayton and class aa of rosebud - lott

2009 - 10 be the year of rosebud - lott be the class aa with a class aaa of rockwall

in the school year 2009 - 10 , class aa winner be grandview and the class aaa winner be snyder

duncanville be the class aaaaa of the school year 1993 - 94 with a class aaa of mount vernon

1993 - 94 and 1994 - 95 be the school year with a class aa of rosebud - lott , a class aaaaa of duncanville , and a class aaa of mount vernon

hamshire - fannett be the class aaa of the 1993 - 94 school year

FINQA demo, github

pre paragrapgh:

15 .

debt the tables below summarize our outstanding debt at 30 september 2016 and 2015 : total debt .

30 september	2016	2015
short-term borrowings	$ 935.8	$ 1494.3
current portion of long-term debt	371.3	435.6
long-term debt	4918.1	3949.1
total debt	$ 6225.2	$ 5879.0
short-term borrowings		
30 september	2016	2015
bank obligations	$ 133.1	$ 234.3
commercial paper	802.7	1260.0
total short-term borrowings	$ 935.8	$ 1494.3

post paragrapgh:

the weighted average interest rate of short-term borrowings outstanding at 30 september 2016 and 2015 was 1.1% (1.1 %) and .8% (.8 %) , respectively . cash paid for interest , net of amounts capitalized , was $ 121.1 in 2016 , $ 97.5 in 2015 , and $ 132.4 in 2014. .

question:

considering the year 2016 , what is the short-term debt as a percent of total debt?

derivation: divide(add(935.8, 371.3), 6225.2)

LogicNLG demo, github

Explore LogicNLG

URL : 2-11545282-4.html.csv

player	nationality	position	years for jazz	school / club team
adrian dantley	united states	guard - forward	1979 - 86	notre dame
brad davis	united states	guard	1979 - 80	maryland
darryl dawkins	united states	center	1987 - 88	maynard evans hs
paul dawkins	united states	guard	1979 - 80	northern illinois
greg deane	united states	guard	1979 - 80	utah
james donaldson	united states	center	1993 , 1994 - 95	washington state
john drew	united states	guard - forward	1982 - 85	gardner - webb
john duren	united states	guard	1980 - 82	georgetown

- Transformer with Copy Mechnism and Field Infusing:

john dawkins and john dawkins are both from the utah - 95
john dawkins and john dawkins are both from the utah - 95
adrian dantley was drafted before john dawkins
john dawkins and john dawkins are both from the utah - 95
the utah all - time all - time all - time all - time fc

- GPT2-small with Linearized Table Input:

John Duren played for Utah Jazz for 2 Year
John Duren played for Utah Jazz for 2 Year
John Drew is the only Player listed who played for Utah Jazz
John Duren played for Utah Jazz for 2 Year
the Utah Jazz had 2 Guard - Forward in 1979 - 80

- GPT2-small with coarse-to-fine machanism:

John Drew played for Utah Jazz for the longest time
John Drew played for Utah Jazz for the longest time
John Drew was drafted before Greg Deane
John Drew played for Utah Jazz for the longest time
Utah Jazz All - Time Roster played Guard for the longest time

LogicNLG - coarse-to-fine generation

Medal Table from Tournament

Nation	Gold Medal	Silver Medal	Bronze Medal	Sports
Canada	3	1	2	Ice Hockey
Mexico	2	3	1	Baseball
Colombia	1	3	0	Roller Skating

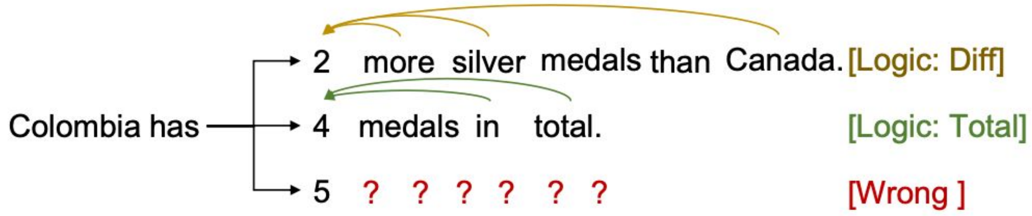


Figure 2: When making the decision at the third step, the model needs to foresee the future tokens to ensure logical consistency. There is no back-tracking once the model makes a wrong decision like “5”.