国际人工智能会议
AAAI 2021 论文北京预讲会

TSQA: Tabular Scenario Based Question Answering

Xiao Li, Yawei Sun, Gong Cheng
State Key Laboratory for Novel Software Technology, Nanjing University, China
{xiaoli.nju, ywsun}@smail.nju.edu.cn, gcheng@nju.edu.cn

Outline

- Motivation
- Task
- Dataset
- Approach
- Experiments
- Error analysis & Future work
- Conclusion

Motivation

Existing Table-based QA can be directly answered **solely on the basis of tabular data, based on sematic parsing or retrieval, with a certain gap from real life problem**

Year	City	Country	Nations
1896	Athens	Greece	14
1900	Paris	France	24
1904	St. Louis	USA	12
...	...	...	...
2004	Athens	Greece	201
2008	Beijing	China	204
2012	London	UK	204

x_1: *"Greece held its last Summer Olympics in which year?"*
y_1: $\{2004\}$

Pasupat et al. 2015, Compositional Semantic Parsing on Semi-Structured Tables

A sample of existing Table-based QA

(Scenario) The *educational level on products (ELP)* refers to the average educational level of workers that produce products. To compute it, educational stages are quantified as follows: 1 for lower secondary or below, 3 for higher secondary, 5 for junior college, 7 for undergraduate, and 9 for graduate or above. The following table shows the development of ELP in China. Please read the table and answer the question.

Year	...	1998	1999	2000	2001	2002	2003
ELP	...	2.465	2.476	2.504	2.490	2.482	2.473

(Question) The main reason for the change of ELP after year 2000 is
(A) improvement in the quality of education
(B) industrial upgrading
(C) rural labor migration
(D) counterurbanization

Real world problem with table

Task

TSQA (Tabular Scenario Based Question Answering):

Answering questions which contextualized by a scenario consisting of a textual passage and a set of tables.

Given:

passage scenario P, a set if table T, a question Q, a set of candidate options

O

Goal:

select an option from O as the answer to Q contextualized by $< P, T >$

Dataset

Dataset Construction of GeoTSQA

- Collecting Questions
  - Domain: geography problems from China’s high-school exams
  - Source: GeoSQA (Huang et al. 2019) and Web
- Identifying Tables
  - Tables, or charts that can be straightforwardly converted to tables (e.g., histograms, line charts)
- Extracting Tables
  - We recruited 15 annotators to annotate table from the raw images
- Filtering Questions
  - Filtered out questions that can be answered without using any table to ensure every question is contextualized by a tabular scenario

Dataset

- Statistics about GeoTSQA

Scenarios	556	
Chinese characters per passage	52.42	±32.99
Tables per scenario	1.58	±0.93
Cells per table	26.98	±17.51
Questions	1,012	
Chinese characters per question	44.02	±15.89

- Features
  - Most of tables (96%) in GeoTSQA is numerical
  - Question can be only answered from the fusing of table, question and scenarios

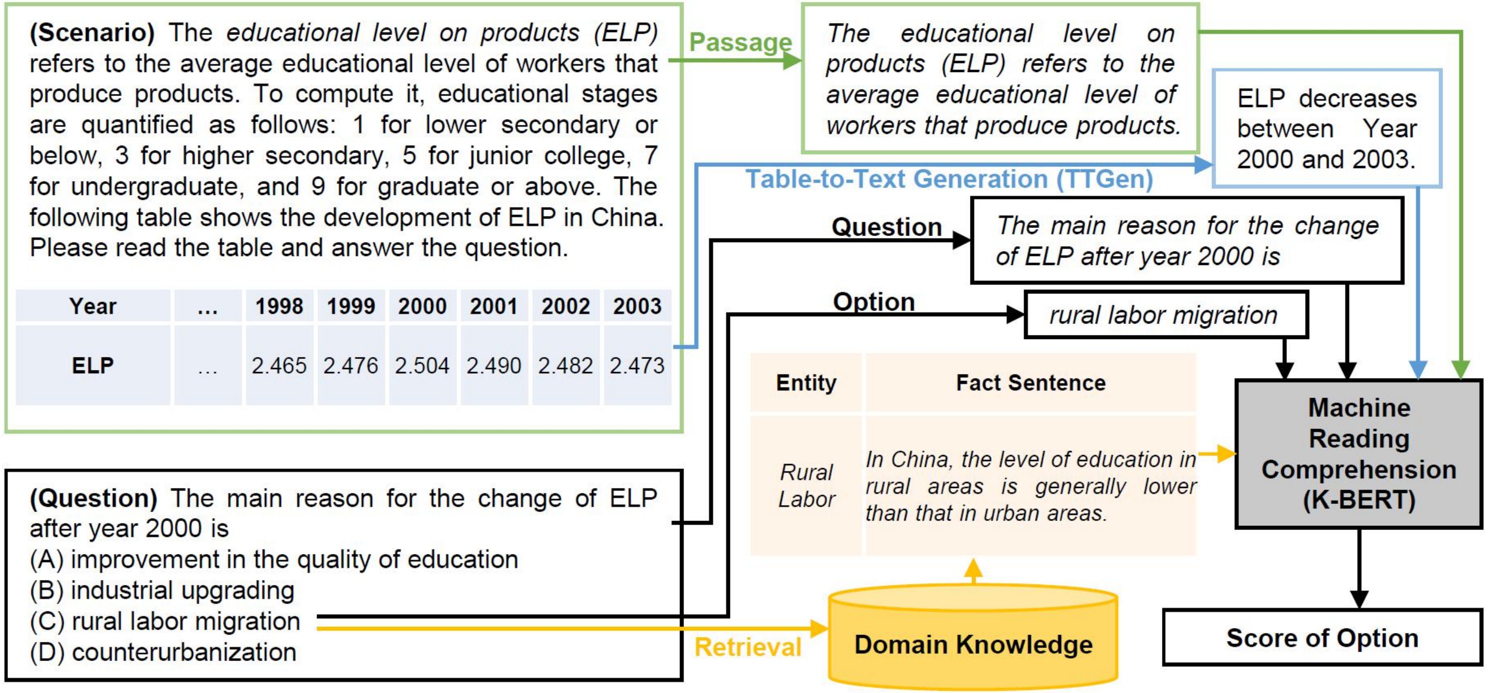
Approach

- Why do existing table-based QA models fail in GeoTSQA?
  - TSQA requires synthesizing data in multiple cells and combining tables with texts and domain knowledge to infer answers
  - Sematic parsing or retrieval based models do not work as answers do not appear in tables in most cases

- Can we extract information in table that needed for QA and then machine reading comprehension?

Approach

- Overview of our approach: a two stage model



Approach

- Template based sentence generation

Table

Year	ELP
…	…
1998	2.465
1999	2.476
2000	2.504
2001	2.49
2002	2.482
2003	2.473

Templates

- Extremum
- Special values
- Comparison with average
- Monotonicity
- Trend
- Range comparison

Sentences

- ELP reaches a maximum of 2.504 at Year 2000
- ELP reaches a minimum of …

- ELP at Year 2000 is 2.504
- …

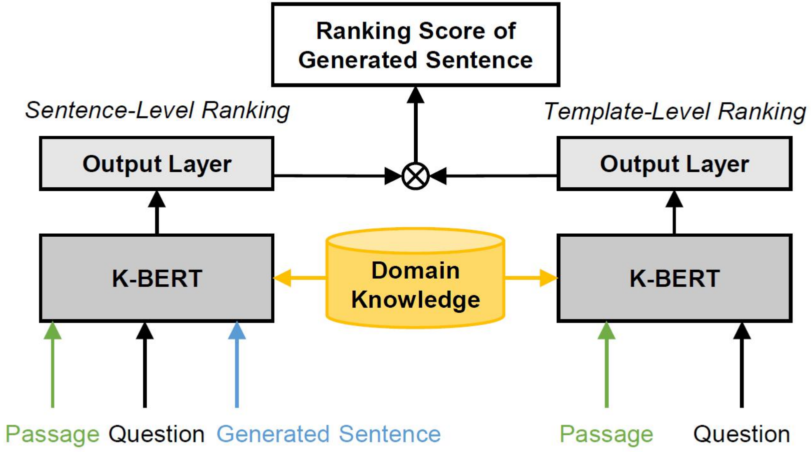
- ELP is relatively large between Year 2000 and 2002
- …
- ELP decreases between Year 2000 and 2003
- …

- ELP generally increases and then decreases
- …

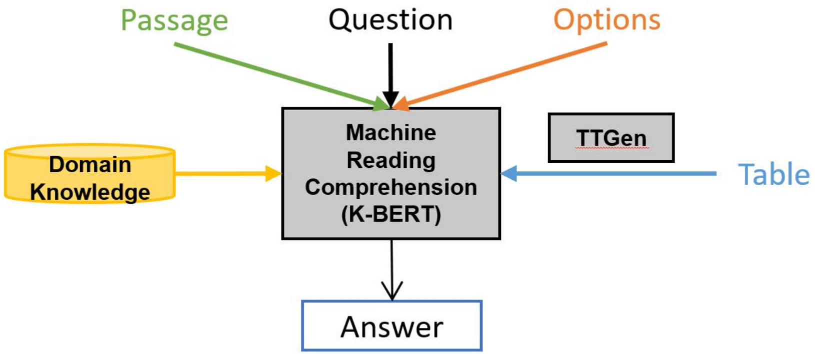
- …

Approach

- Not all table information is useful for QA.
- Two level sentence ranking: sentence-level ranking and template-level ranking
- Then feed the MRC model with top-k table sentences to select answer



Two level sentence ranking



Machine Reading Comprehension

Experiments

- Experiments on GeoTSQA
- We extend sota table-to-text method with MRC model
- Table-to-text Methods
  - Table-Infusing (Chen et al., 2020): encode table cells and then generate
  - GPT-Linearization (Chen et al., 2020): describe all cells of table and feed the GPT-2 with table passage to generate
  - Coarse-to-Fine (Chen et al., 2020): first generate template and then fill out template
  - Linearization: describe all cells of table into a passage
  - Templation: concatenate all sentences of $\boldsymbol{S}$ into a table passage

Experiments

Metric: Accuracy, i.e., the proportion of correctly answered questions

	Accuracy
Field-Infusing	0.353 $^{\bullet}$
GPT-Linearization	0.370
Coarse-to-Fine	0.367
GPT-Linearization^{+}	0.348 $^{\bullet}$
Coarse-to-Fine^{+}	0.359 $^{\circ}$
Linearization	0.235 $^{\bullet}$
Templation	0.243 $^{\bullet}$
TTGen	**0.397**
Gold-Standard Sentence	0.418

Table 4: Accuracy of TSQA. We mark the results of baselines that are significantly lower than TTGen under $p < 0.01$ ($^{\bullet}$) or $p < 0.05$ ($^{\circ}$).

Experiments

Varying k (top-k ranked sentences are used in MRC)

	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Accuracy	0.390	0.397	0.352	0.343	0.330

Table 5: Accuracy of TSQA by varying k in TTGen.

Ablation study

	Accuracy
TTGen	**0.397**
TTGen w/o tabular data	0.372
TTGen w/o domain knowledge	0.380

Table 6: Accuracy of TSQA (ablation study).

Error analysis & Future work

Error analysis:

- Lack of external knowledge base (76%)
- Weak Reasoning Capabilities (62%)
- Sentence Ranking Error (54%)

Future work:

- More powerful MRC model which is able to fuse scenario passage, table sentences and domain knowledge
- More powerful reasoning module
- More powerful question-aware table-to-text generator

Conclusion

- We constructed the first TSQA dataset: GeoTSQA.
- We use six templates encapsulating predefined operations for synthesizing tabular and then select the most useful sentences.
- We use MRC model, fusing scenario passage, question, table information and domain knowledge to select answer.

Resources: https://github.com/nju-websoft/TSQA

国际人工智能会议
AAAI 2021 论文北京预讲会

THANKS

2020.12.19