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

Task and Dataset

Task:

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

Challenges:

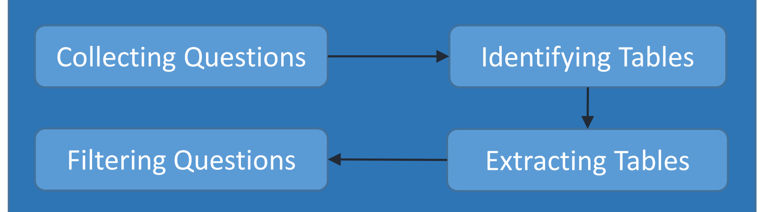
- Properly manipulating tabular data to obtain deep-level tabular information
- Reading comprehension with fusion of tabular data and textual information

Contributions:

- GeoTSQA: the first dataset dedicated to TSQA, posing new research challenges
- TTGen: extending machine reading comprehension methods with our table to text model TTGen to solve TSQA

Dataset:

- Construction



- Statistics

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:
  - Table information is needed for QA
  - Most of tables (96%) are numeric

Example:

(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.

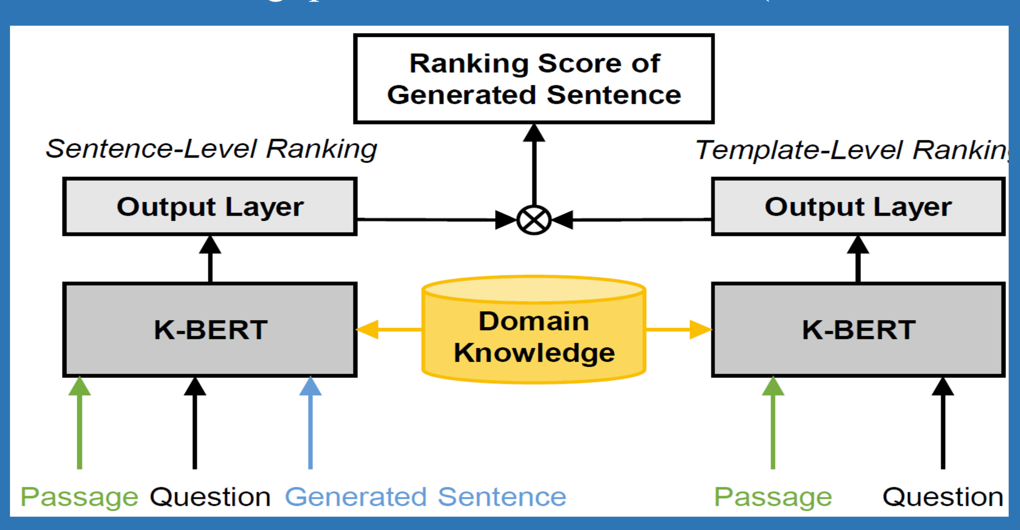
Year	…	1998	1999	2000	2001	2002	2003
ELP	…	2.465	2.476	2.504	2.490	2.482	2.473

(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

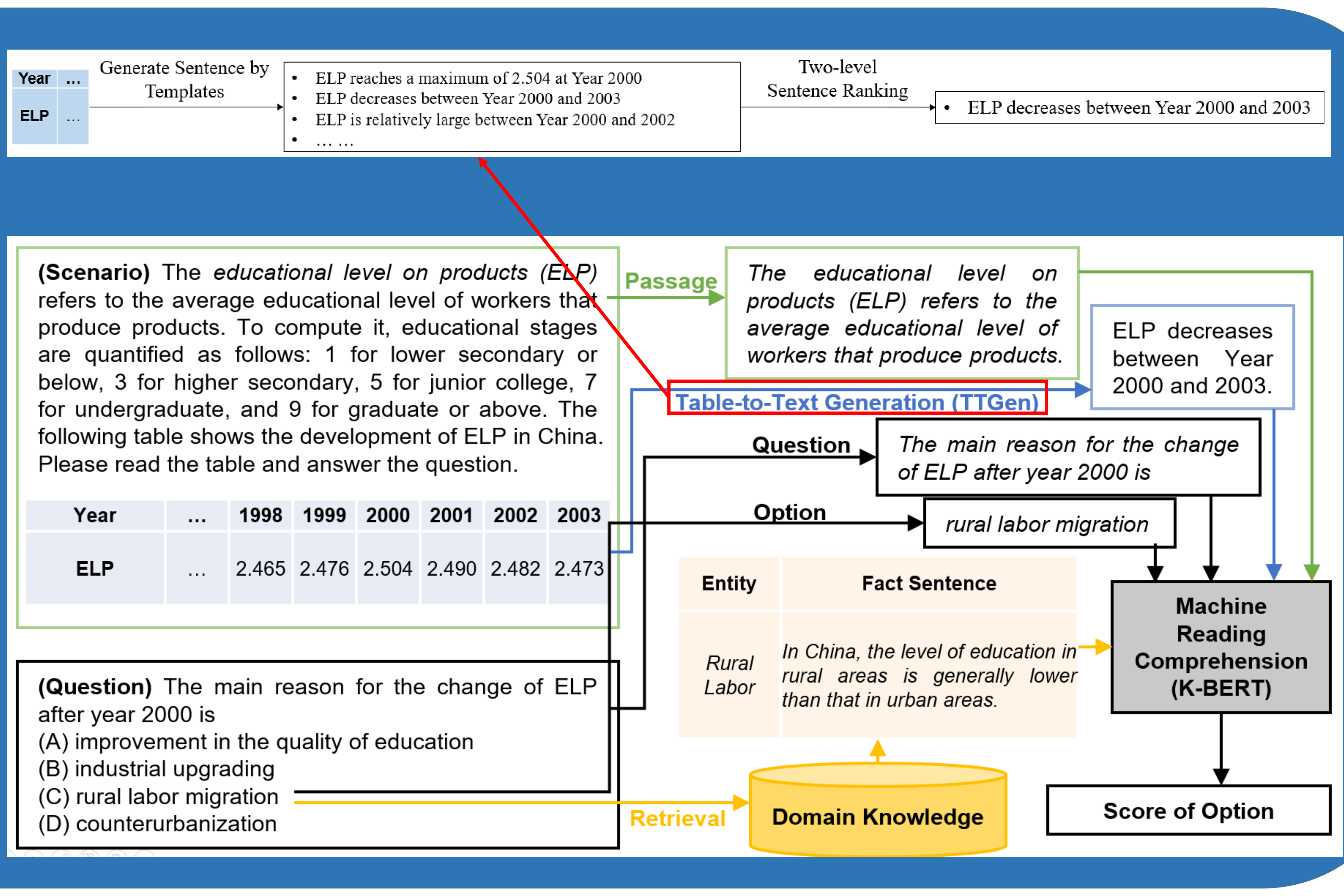
Approach

Approach:

- **TTGen: a table-to-text generator**
  - Sentence generation: we use six table-to-text templates to encapsulate different powerful operations for synthesizing numeric tabular data
  - Sentence Ranking: we get a set of table sentences $\hat{S}$ and only a subset of $\hat{S}$ is useful for answering question, we use K-BERT (Liu et al., 2020) for ranking



- **Machine reading comprehension with top-k $\hat{S}$**
  - h_i^{MRC} = K-BERT($question, option, domain\ knowledge,\ topk\ \hat{S}$)
  - The final of a candidate option is scored by h_i^{MRC} with a multi-layer perceptron (MLP)



Experiments

Evaluation Design:

- Experiments on GeoTSQA
- We extend sota table-to-text method with MRC model
- SOTA 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 $\hat{S}$ into a table passage

Results on GeoTSQA

	Accuracy
Field-Infusing	0.353 •
GPT-Linearization	0.370
Coarse-to-Fine	0.367
GPT-Linearization+	0.348 •
Coarse-to-Fine+	0.359 ∘
Linearization	0.235 •
Templation	0.243 •
TTGen	**0.397**
Gold-Standard Sentence	0.418

Results on GeoTSQA (ablation study)

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

Results on GeoTSQA by varying k (top-k subset of $\hat{S}$)

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

Error analysis:

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

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. At last, we use MRC model, fusing scenario passage, question, table information and domain knowledge to select answer.

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

主办方：中国中文信息学会青年工作委员会

承办方：智源社区