Learning on Structured Documents for Conditional Question Answering

CCL 2023

Zihan Wang, Hongjin Qian and Zhicheng Dou
Gaoling School of Artificial Intelligence, Renmin University of China

❖ Major Contributions of Our Work

▶ **New approach for conditional question answering:** We propose a self-supervised learning method on structured documents based on *conditional question generation* and *contrastive learning*, to resolve the challenge of insufficient supervision for conditional question answering.

▶ **Complex response generation:** We propose an end-to-end pipeline to provide controllable conditional answers for conditional question answering by: 1) selecting candidate answers and conditions, and 2) choosing the best matches by calculating the matching score of their corresponding query and key vectors.

▶ **Excellent results:** The experimental results show that our method can answer conditional questions more accurately Computational Linguistics compared to previous conditional question answering methods.

❖ Conditional Question Generation

Algorithm 1 Conditional Question Generation

Require: Structured doc set DATASET
Ensure: Cond. question q, scenario sc, answer a, condition c
procedure QUESTIONGEN(DATASET)
  Init: state gen. G_S, label gen. G_L
  Sample doc D from DATASET
  Select non-leaf text node $s \in D$ as potential answer
  Construct extracted $\overline{D}$ by selecting anc., child., sibl., and sibl. child. of s
  Gen. question q, scenario sc using $G_S(\overline{D})$
  Gen. cond. answers $A = (a_i, c_i)$ using $G_L(q, sc, \overline{D})$
end procedure

	answers	conditions
leaf node	86.93%	92.53%
text node	92.49%	98.33%

	a-a pairs	c-c pairs	a-c pairs
sibling-sibling	66.55%	53.67%	-
parent-child	-	-	39.59%

Algorithm: Conditional question generation with an augmentation corpus and a trained generator.

Tables: Statistics of the ConditionalQA train dataset for guiding selective extraction.

▶ *Selective extraction* are conducted as shown in Algorithm, based on the discoveries in statistics of ConditionalQA dataset:
1) leaf-text nodes are likely to be conditions and answers;
2) different answers to a conditional question are usually siblings;
3) the conditions for an answer is usually in the child nodes of it.

▶ Two models - a generative *state generator* and an extractive *label generator* are leveraged to generate diverse conditional question with user scenario and high-accuracy labels.

❖ Perturbative Contrastive Learning

Motivation: The *structure* of documents can be perturbed with little loss in its information. Similar representations of the corresponding parts can be learned to facilitate learning from such complex documents and benefiting the CQA task.

Operation	Description	Motivation
Node masking	Mask node with [MASK] of same length	Focus on structure & context
Node deletion	Delete non-root node & descendants	Learn node dependencies & importance
Node cloning	Clone node & descendants as another child	Identify semantically similar nodes
Node shuffling	Shuffle child nodes within parent	Understand impact of node order

Table: Four basic operations on the documents for Contrastive Learning with little loss in the document's information.

$$\mathcal{L}_{CL}(D, \hat{D}) = \sum_{i=1}^{m'} \frac{e^{\mathrm{sim}(t'_i, t_{k_i})}}{e^{\mathrm{sim}(t'_i, t_{k_i})} + \sum_{t^-_{k_i}} e^{\mathrm{sim}(t'_i, t^-_{k_i})}}$$

Formula: Contrastive learning loss, where m' is the total number of nodes in Document D, t'_i and t_{ki} represent a positive pair (i.e., tags of corresponding nodes), and t^-_{ki} represents tags of any nodes other than t_{ki} in D. sim computes the similarity of two nodes with dot-product.

❖ Overview of Our Idea & Approaches

Intuition: If we can provide Conditional Question Generator G with a more precise context S with sufficient information for a conditional question, then G can answer the question more accurately, and the obtained *Augmentation Dataset* can be used to train Conditional Question Answering Model M.

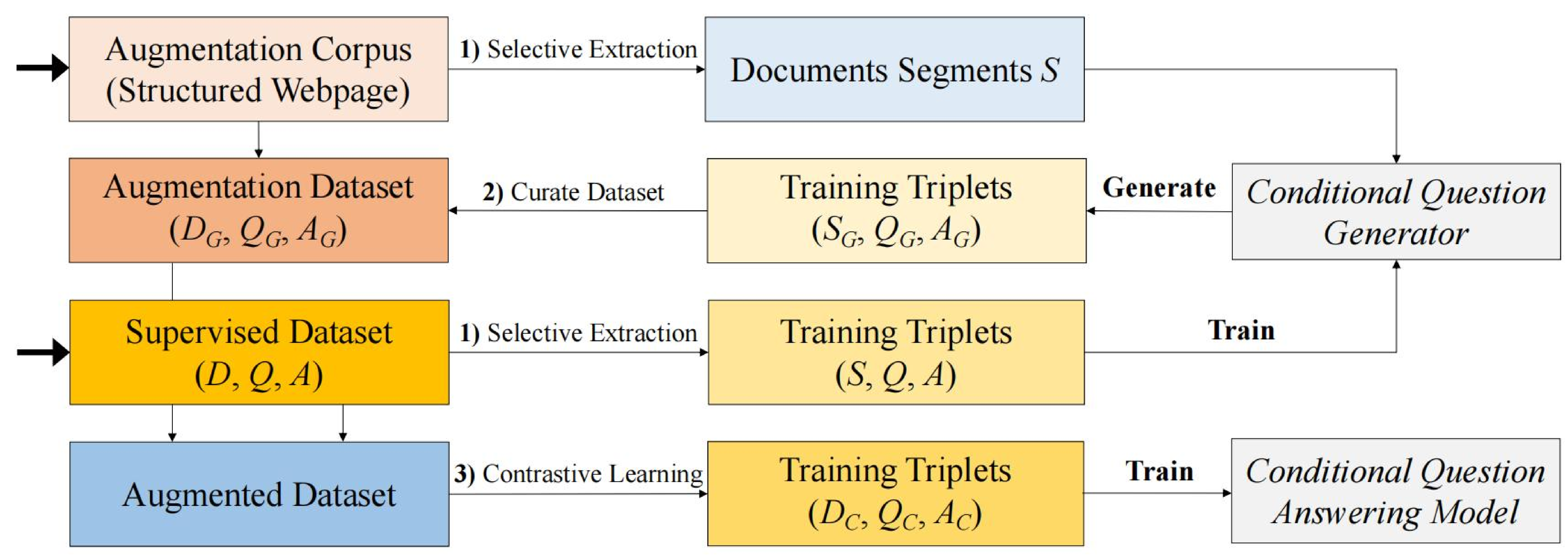

Figure: Overview of our self-supervised learning approach for Conditional Question Answering.

▶ The *Augmentation Corpus* are crawlled from web and we leveraged *selective extraction* to generate document segments. We leverage *Supervised Dataset* ConditionalQA to train G and generate *Augmentation Dataset* with G.

▶ Afterwards, M is trained on the union of *Augmentation Dataset* and *Supervised dataset* with *contrastive learning* which results in decent performance gain.

❖ Pipeline for CQA Response Generation

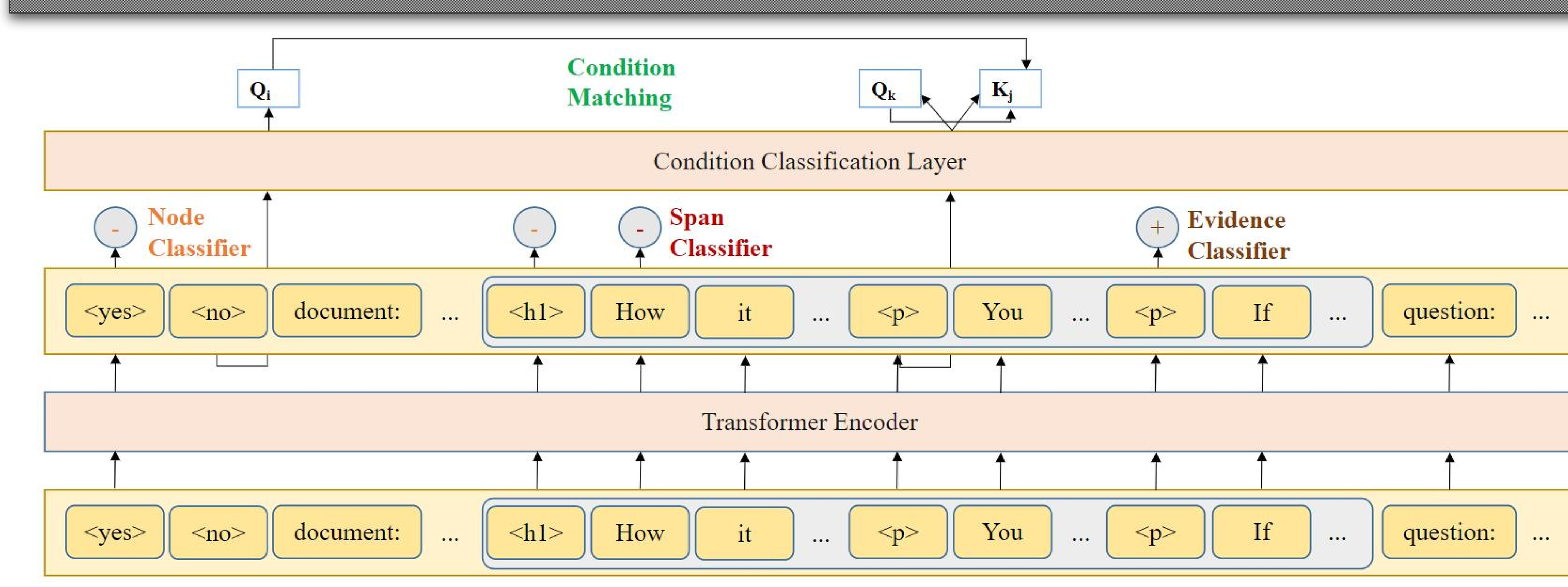

Figure: Pipeline to answer conditional questions with candidate answer-condition selection and matching score calculation.

❖ Evaluations on ConditionalQA Dataset

	Yes / No		Extractive		Conditional		Overall	
	EM / F1	w/ conds	EM / F1	w/ conds	EM / F1	w/ conds	EM / F1	w/ conds
ETC-pipeline	63.1 / 63.1	47.5 / 47.5	8.9 / 17.3	6.9 / 14.6	39.4 / 41.8	2.5 / 3.4	35.6 / 39.8	26.9 / 30.8
DocHopper	64.9 / 64.9	49.1 / 49.1	17.8 / 26.7	15.5 / 23.6	42.0 / 46.4	3.1 / 3.8	40.6 / 45.2	31.9 / 36.0
FiD	64.2 / 64.2	48.0 / 48.0	25.2 / 37.8	22.5 / 33.4	45.2 / 49.7	4.7 / 5.8	44.4 / 50.8	35.0 / 40.6
TReasoner	**73.2 / 73.2**	**54.7 / 54.7**	34.4 / 48.6	30.3 / 43.1	51.6 / 56.0	12.5 / 14.4	57.2 / 63.5	**46.1 / 51.9**
LSD (ours)	71.6 / 71.6	51.6 / 51.6	**39.9 / 56.4**	**31.6 / 43.8**	**57.3 / 61.8**	**21.4 / 25.1**	**58.7 / 66.2**	45.0 / 50.5

Table: Main results on the ConditionalQA dataset. "EM / F1" shows the standard EM / F1 metrics based on the answer span only. "w/ conds" shows the conditional EM / F1 metrics.
1) LSD outperforms all baselines on extractive and conditional subset.
2) LSD performs less well than TReasoner in Yes / No questions. LSD's is inclined to provide conditional answers due to training with our question generation system, which is penalized by the evaluation metric.

	Answer (w / conds)	Conditions (P / R / F1)
ETC-pipeline	/	/
DocHopper	/	/
FiD	3.2 / 4.6	98.3 / 2.6 / 2.7
FiD (cond)	6.8 / 7.4	12.8 / 63.0 / 21.3
TReasoner	10.6 / 12.2	34.4 / 40.4 / 37.8
LSD (ours)	**21.4 / 25.1**	**69.3 / 39.4 / 50.2**

Table: Result on the ConditionalQA dataset with conditional answers. LSD significantly outperforms all baselines in all metrics.

Email: {wangzihan0527, ian, dou}@ruc.edu.cn
Homepage: {zihanwang314, qhjqhj00}.github.io
http://playbigdata.ruc.edu.cn/dou/