Technical Question Answering across Tasks and Domains

Wenhao Yu†, Lingfei Wu‡, Yu Deng‡, Qingkai Zeng†, Ruchi Mahindru‡, Sinem Guven‡, Meng Jiang†
†University of Notre Dame, USA ‡ IBM Thomas J. Watson Research Center, USA

Introduction/Motivation (Given a question, our goal is to identify the best document and find the appropriate answer from the document.)

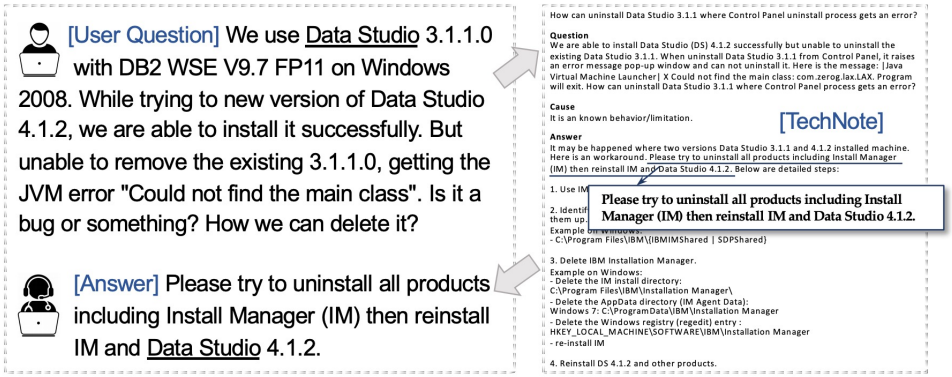

Figure: A question-answer example in the IT support domain

Proposed Method (TransTD jointly learns snippet prediction and document matching, applying it on both general domain QA and technical domain QA.)

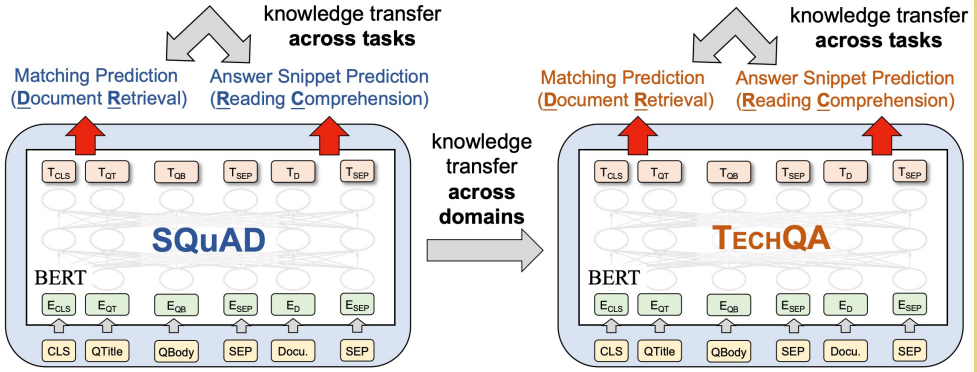

Figure: The overall architecture of TransTD

Proposed System (knowledge transfer across tasks and domains

Table 1: Ablation study on knowledge transfer across tasks and across domains on TechQA.

Methods		Adjustable	Source task(s)	Target task(s)	Reading Comprehension			Document Retrieval		
					Ma-F1	HA-F1@1	HA-F1@5	MRR	R@1	R@5
BERT_{DR}	-	✘	-	DR	-	-	-	55.80	45.58	58.23
BERT_{RC}	-	✘	-	RC	52.49	24.92	37.26	51.20	48.13	56.25
TransD	-	✘	RC	DR	-	-	-	60.63	58.13	64.38
	-	✘	RC	RC	55.31	34.69	50.52	64.60	60.63	68.23
TransT	CLS	✘	-	RC+DR	53.43	26.83	38.50	51.19	46.88	56.88
	Mean	✘	-	RC+DR	52.30	26.28	41.50	52.68	47.50	59.35
TransTD	CLS	✘	RC+DR	RC+DR	56.43	39.12	52.30	66.79	64.38	70.63
	Mean	✘	RC+DR	RC+DR	56.88	37.96	49.83	67.55	**67.50**	69.38
TransTD^{+}	CLS	✔	RC+DR	RC+DR	56.66	38.33	50.95	67.80	65.00	72.50
	Mean	✔	RC+DR	RC+DR	**58.58**	**40.28**	**52.57**	**67.98**	66.88	**73.13**

Table 2: TransTD outperforms two-stage retrieve-then-read methods that retrieve document based on semantic alignment. k is the number of retrieved documents.

Method	Setting	Ma-F1	HA-F1@1	R@1
BERTserini (Yang et al., 2019a) (with BM25 as retriever)	k=1	51.34	15.23	30.00
	k=5	56.60	28.31	48.75
DPR (Karpukhin et al., 2020) (w/o pre-trained retriever)	k=1	53.22	15.57	26.25
	k=5	56.47	30.40	47.50
DPR (Karpukhin et al., 2020) (with pre-trained retriever)	k=1	54.82	19.46	30.63
	k=5	58.56	33.03	53.13
TransTD-Mean^{+} (Ours, S_{with})	-	**58.58**	**40.28**	**66.88**

Table 3: Our proposed snippet ranking function can bring additional improvements. Using $(p_s[0] + p_e[0])$ reflects the degree of misalignment between Q and D.

Snippet ranking method	Ma-F1	HA-F1@1	R@1
MP-BERT (Wang et al., 2019) ($S_{\text{MP-BERT}} = p_{\text{DR}} \cdot p_s \cdot p_e$)	49.45	24.65	43.75
WKLM (Xiong et al., 2020) ($S_{\text{BERT}} = \alpha \cdot p_{\text{DR}} + p_s + p_e$)	57.82	39.71	66.25
Ours (w/o document score) ($S_{w/o} = p_s + p_e - p_s[0] - p_e[0]$)	**58.58**	**40.28**	65.00
Ours (with document score) ($S_{with} = \alpha \cdot p_{\text{DR}} + S_{w/o}$)	**58.58**	**40.28**	**66.88**

Acknowledgement

This work was supported by Thomas J. Watson Research Center, and National Science Foundation grants, IIS-1849816 and CCF-1901059.