

A Experimental Settings

A.1 Dataset

TECHQA (Castelli et al., 2020). It is to stimulate transfer learning research to address non-factoid questions and long solution documents. The dataset is recently released with a competition leaderboard. The TECHQA dataset contains actual questions posed by users on the IBM Developer and IBM DeveloperWorks forums as well as a collection of IBM Technotes.

StackExchange Stack Exchange² is a network of QA websites composed by multiple communities. In our experiments, we choose two popular technical communities: AskUbuntu³ and StackUnix⁴.

A.2 Evaluation Metrics

As a standard retrieval task, our models are evaluated on retrieving and ranking answers for given

questions using two metrics: mean reciprocal rank (MRR) and recall at K ($R@K$). $R@K$ is the percentage of correct answers in top K out of all the relevant answers. MRR represents the average of the reciprocal ranks of results for a set of queries.

B Implementation Details

We use lite-BERT (ALBERT) as a pre-trained language model, which takes a maximum 384 input token sequence. For model fine tuning, we use AdamW (Kingma and Ba, 2014) with learning rate of $5.5e-6$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, L2 weight decay of 0.01, learning rate warm up over the first 10,000 steps, and linear decay of learning rate. Our models are trained with a single-card V100 GPU, and deployed on PyTorch with Flask web framework.

²<https://stackexchange.com/>

³<https://askubuntu.com/>

⁴<https://unix.stackexchange.com/>

Dataset	#Instance	#Answer*	#Train	#Valid	#Vocab	Q.Length	A.Length
AskUbuntu	30419	10951	7222	3729	33420	58.4	44.9
TechQA	750	607	450	157	7485	46.3	58.6
StackUnix	1394	1394	1115	279	7072	51.0	37.7
Internal \mathcal{D}	3256	3256	2148	1108	19189	61.9	36.8

Table 4: Statistics of three technical QA dataset.

Methods		Source QAs	Source Corpus	IBM Internal Dataset \mathcal{D}			
				MRR	R@1	R@5	R@10
SASE (LSTM)	M1	-	-	8.65	4.87	11.64	15.97
	M2	✓	-	10.58	6.98	13.25	17.95
	M3	-	✓	14.69	9.56	24.50	30.45
	M4	✓	✓	18.61	13.26	28.16	35.65
ALBERT	M1	-	-	19.44	12.81	25.27	32.85
	M2	✓	-	22.47	14.00	30.42	39.10
	M3	-	✓	19.64	11.46	26.53	35.20
	M4	✓	✓	24.85	16.60	32.58	41.78

Table 5: Results on an IBM internal technical dataset \mathcal{D} .