A Experimental Results Using Sentence-BERT

We have also experimented with using sentence embeddings output by Sentence-BERT (Reimers and Gurevych, 2019) instead of LDA in the CSG component. Specifically, we used bert-large-nli-stsb-mean-tokens from official repo[1], which is a BERT-based model fine-tune on NLI and STS tasks. The experimental results of ranking metrics are listed below in Table 1:

Instantiation		F1@3	P@1	P@3	R@3	MRR	NDCG@10	Semantic Sim.@3
CSG	DS							
WordNet	point-wise ranker	7.41	9.26	5.41	13.13	13.33	14.78	0.33
	pair-wise ranker	6.58	8.49	4.76	11.71	12.47	14.09	0.32
	list-wise ranker	7.34	8.11	5.15	13.51	12.21	14.31	0.32
Probase	point-wise ranker	6.16	6.56	4.63	10.29	11.95	12.56	0.36
	pair-wise ranker	6.18	6.18	4.77	10.03	12.64	13.63	0.35
	list-wise ranker	6.21	8.89	4.89	9.84	14.26	13.98	0.35

Table 1: Comparison of combinations of different choices of CSG and DS.

B Application

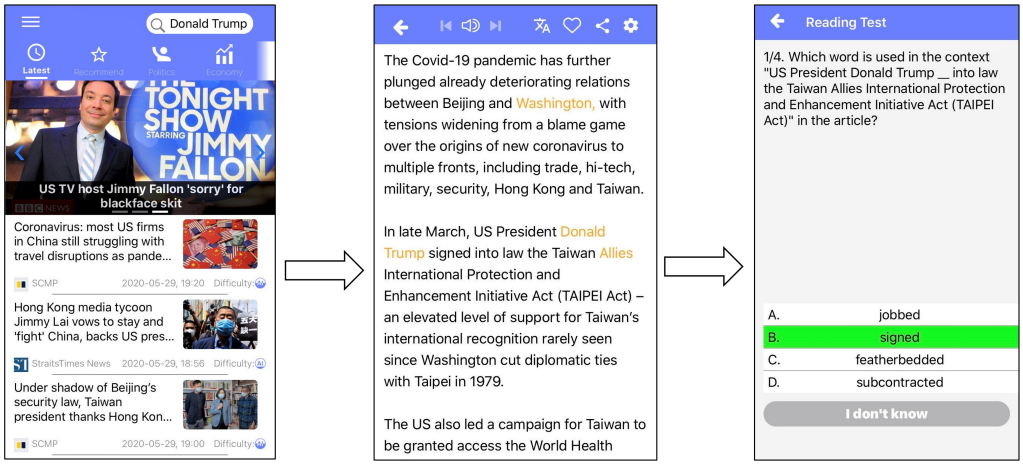

Figure 1: The mobile app in which the proposed distractor generation framework is deployed. The right-most screenshot shows one generated MCQ given the news context in the previous screenshot.

In this section, we demonstrate how the proposed framework is deployed in a real-world application. The most suitable scenario to implement our distractor generation system is AI-assisted education, especially in language learning. Our framework has been successfully deployed in an English learning and reading app. At the end of each chapter of book or news, it will generate several multiple-choice questions to assess how well users comprehend the content. In the implementation, after locating question-worthy context as question body and keyword as correct answer, our CSG+DS distractor generation framework is then applied to automatically generate distractive options that will be put together with correct answer to form a complete multiple-choice question.

[1] https://github.com/UKPLab/sentence-transformers

An overview of the app and application of our distractor generation framework is shown in Figure 1, where the user selects *signed* as the answer after reading the news:

The ideal distractors in MCQs are required to be plausible and reliable to fully test the users, which is to a large extent satisfied by our proposed framework. Recent statistics from the app shows that a user may take the quiz from either books or news 9.4 times per day on average and the chances of users making an incorrect answer is about 42%, which validates the effectiveness of our CSG+DS framework in practical use.

References

Reimers, N.; and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks.