A Compared Methods

BiDAF (Seo et al., 2016) is a representative network for machine comprehension. It is a multi-stage hierarchical process that represents context at different levels of granularity and uses a bi-directional attention flow mechanism to achieve a query-aware context representation without early summarization.

Co-matching (Wang et al., 2018) uses the attention mechanism to match options with the context that composed of paragraphs and the question, and output the attention value to score the options. It is used to solve the single paragraph reading comprehension task of a single answer question.

Multi-Matching (Tang et al., 2019) applies the Evidence-Answer Matching and Question-Passage-Answer Matching module to gather matching information and integrate them to get the scores of options.

SeaReader (Zhang et al., 2018) is proposed to answer questions in clinical medicine using knowledge extracted from publications in the medical domain. The model extracts information with question-centric attention, document-centric attention, and cross-document attention, and then uses a gated layer for denoising.

BERT (Devlin et al., 2019) achieves remarkable state-of-the-art performance across a wide range of related tasks, such as textual entailment, natural language inference, question answering. It first trains a language model on an unsupervised large-scale corpus, and then the pre-trained model is fine-tuned to adapt to downstream tasks.

RoBERTa (Liu et al., 2019) is based on BERT's language masking strategy and modifies key hyper-parameters in BERT, including changing the target of BERT's next sentence prediction, and training with a larger bacth size and learning rate. It has achieved improved results than BERT on different data sets.

ERNIE (Sun et al., 2019) is designed to learn language representation enhanced by knowledge masking strategies, which includes entity-level masking and phrase-level masking. It achieves state-of-the-art results on five Chinese natural language processing tasks.

B Relation Classification

We also show the dataset that used to pre-train on the relation classification task and the performance of the pre-trained models in this task. We compare several common text classification and matching models, including TextCNN (Kim, 2014), ESIM (Chen et al., 2017), DPCNN (Johnson and Zhang, 2017). For text classification, the input of the model is the concatenation of two entity words. For ESIM, the input layer is softmax multi-classification. Through learning with the relation classification task, pre-trained models achieve improved performance on the divided test set.

	TRAIN	DEV	TEST
# Knowledge facts	$1,129,780$	$50,000$	$50,000$

Model	**Accuracy (TEST)**
RoBERTa-wwm-ext-large (Cui et al., 2019)	89.4
RoBERTa-wwm-ext-large (w/o fine-tuning)	50.8
BERT-base (Devlin et al., 2019)	88.8
BERT-base (w/o fine-tuning)	50.6
DPCNN (Johnson and Zhang, 2017)	82.6
TextCNN (Kim, 2014)	67.8
ESIM (Chen et al., 2017)	77.8

Table 1: Data statistics of relation classification task and accuracy results.

C Introduction to Exam

The detailed statistics of exams in recent years are listed in Table 2. The professional qualifications for licensed pharmacists are subject to a national unified outline, unified proposition, and unified organized examination system (**?**). The qualification exam for licensed pharmacists is held on every October. The examination takes two years as a cycle, and those who take the examination of all subjects must pass the examination of all subjects within two consecutive examination years. The professional qualification examination for licensed pharmacists is divided into two professional categories: pharmacy and traditional Chinese pharmacy. The pharmacy exam subjects are (1) pharmacy professional knowledge (first part) (2) pharmacy professional knowledge (second part) (3) pharmacy management and regulations, and (4) pharmacy comprehensive knowledge and skills. The subjects for the examination of traditional Chinese medicine are (1) professional knowledge of traditional Chinese medicine (first part) (2) professional knowledge of traditional Chinese medicine (second part) (3) pharmaceutical management and regulations, and (4) comprehensive knowledge and skills of traditional Chinese medicine.

Years	# Applicants (k)	# Participants (k)	Exam ratio (%)	# Passing (k)	Pass ratio (%)
2018	687.5	566.6	82.41	79.9	14.10
2017	675.2	523.2	77.50	153.0	29.19
2016	884.7	728.6	82.38	151.0	20.74
2015	1121.4	937.7	83.62	235.0	25.16
2014	840.2	702.4	83.61	137.1	19.52
2013	402.3	329.8	81.99	51.8	15.72
2012	188.1	146.8	78.09	26.0	17.68
2011	145.9	109.7	75.16	14.4	13.13
2010	132.7	100.6	75.76	11.2	11.12

Table 2: Statistics of this exam in recent years

D Source of Questions

The source website and books of collected questions are (1) www.51yaoshi.com (2) Sprint Paper for the State Licensed Pharmacist Examination-China Medical Science and Technology Press (3) State Licensed Pharmacist Examination Golden Exam Paper - Liaoning University Press (4) Practicing Pharmacist Quiz App (5) The Pharmacist 10,000 Questions App (6) Practicing Pharmacist Medical Library App

References

Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang, and Diana Inkpen. 2017. Enhanced LSTM for natural language inference. In *ACL*.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Ziqing Yang, Shijin Wang, and Guoping Hu. 2019. Pre-training with whole word masking for chinese BERT. *CoRR*, abs/1906.08101.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Rie Johnson and Tong Zhang. 2017. Deep pyramid convolutional neural networks for text categorization. In *ACL*.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *EMNLP*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. 2016. Bidirectional attention flow for machine comprehension. *arXiv preprint arXiv:1611.01603*.

Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019. Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*.

Min Tang, Jiaran Cai, and Hankz Hankui Zhuo. 2019. Multi-matching network for multiple choice reading comprehension. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7088–7095.

Shuohang Wang, Mo Yu, Jing Jiang, and Shiyu Chang. 2018. A co-matching model for multi-choice reading comprehension. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 746–751.

Xiao Zhang, Ji Wu, Zhiyang He, Xien Liu, and Ying Su. 2018. Medical exam question answering with large-scale reading comprehension. In *AAAI*.