Reconstructed Option Rereading Network For Opinion Questions Reading Comprehension

Delai Qiu[1], Liang Bao[1], Zhixing Tian[2,3], Yuanzhe Zhang[3], Kang Liu[2,3], Jun Zhao[2,3], and Xiangwen Liao[1]

College of Mathematics and Computer Science, Fuzhou University, Fuzhou, China
{noneqdl, caldreaming821618}@gmail.com
University of Chinese Academy of Sciences, Beijing, 100049, China
National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
{zhixing.tian, yzzhang, kliu, jzhao}@nlpr.ia.ac.cn

Abstract. Multiple-choice reading comprehension task has seen a recent surge of popularity, aiming at choosing the correct option from candidate options for the question referring to a related passage. Previous work focuses on factoid-based questions but ignore opinion-based questions. Options of opinion-based questions are usually sentiment phrases, such as "Good" or "Bad". It causes that previous work fail to model the interactive information among passage, question and options, because their approaches are based on the premise that options contain rich semantic information. To this end, we propose a **R**econstructed **O**ption **R**ereading **N**etwork (**RORN**) to tackle it. We first reconstruct the options based on question. Then, the model utilize the reconstructed options to generate the representation of options. Finally, we fed into a max-pooling layer to obtain the ranking score for each opinion. Experiments show that our proposed achieve state-of-art performance on the Chinese opinion questions machine reading comprehension datasets in AI challenger competition.

Keywords: Machine Reading Comprehension. · Opinion Analysis

1 Introduction

Multiple-choice reading comprehension(MCRC) is a major form of machine reading comprehension (MRC) task, which requires a system to read a given passage and a question for choosing the correct option from the candidate options. Questions of MCRC task are generally divided into factoid-based questions and opinion-based questions. Figure 1 shows two different types of question. As we can see, the options of factoid-based question usually have abundant context information, but the options of opinion-based question usually select a opinion with some short sentiment phrases such as "Yes" and "No". AI challenger 2018[1] defines a new sub-task called Opinions Question Machine Reading Comprehension, which needs to choose a opinion option from candidates.

[1] https://challenger.ai/competition/oqmrc2018

Passage:
... There is no problem if the intake of vitamin no more than the recommended amount ...

Question:
Can we take the vitamin C every day?
A. Yes B. No C. Indeterminacy

Answer: A

Passage:
...... In 1993, New York State ordered stores to charge money on beverage con-tainers. Within a year, consumers had returned millions of aluminum cans and glass and plastic bottles. Plenty of companies were eager to accept the alu-minum and glass as raw material for new products, but because few could figure out ...

Question:
What regulation was issued by New York State concerning beverage containers?
A. A fee should be charged on used containers for recycling.
B. Throwaways should be collected by the state for recycling.
C. Consumers had to pay for beverage containers and could get their money back on returning them.

Answer: C

Fig. 1. The above is a factoid-based question, and other is an opinion-based question.

Many previous studies have introduced neural-based models on multiple-choice reading comprehension [16, 24, 18, 21, 12, 23, 9], which typically have following pipelines. Firstly, they encode the passage, question and candidate options to generate the contextual representation of them respectively. Secondly, attention mechanisms are employed to acquire the interaction representation among them. Further more, they prove that consider option correlations contribute to the semantic representation. Finally, the final output module computes scores of options based on score function to generate the final predictions.

Though previous work achieve promising results in recent years, they cannot handle the opinion-based questions. When reading the passage, understanding the semantic information of options is a common strategy for human beings, which inspires most of existing models. But the strategy is ineffective in opinion-based questions, because options of opinion-based questions always do not contain the context information. Taking Figure 1 as a example, we are able to gain more context information from factoid-based options but we cannot access more from candidate options such as ”YES” and ”NO”.

In these paper, we present a novel model to trackle opinion-based questions in multiple choice reading comprehension task. Firstly, Our model introduces a simple but effective method to reconstruct opinion options to acquire the semantic information of options. Then, we employ BERT[2] as our encoder to obtain the context representation of reconstructed options, questions and passages. Then we apply the co-attention mechanism to fuse the information between each options and passage, the rereaded option representation for each option is computed with self-attention mechanism. Finally, we utilize the max pooling layer to make the final prediction.

We conduct experiments on Opinions Question Machine Reading Comprehension dataset in AI challenger 2018. Our experiments show that the validity of option reconstruction and option rereading with passage-aware information. Our contributions can be summarized as:

(1) We reconstruct options of the opinion-based questions, which supplement more semantic information of options.
(2) Attention mechanisms is employed to generate more subtle context-aware representation of the options.
(3) Experiments demonstrate that the model achieve state-of-the-art on Opinion Questions Machine Reading Comprehension datasets.

2 Related Work

Multiple-choice reading comprehension (MCRC) is a major form of machine reading comprehension (MRC) task, aiming to selecting the correct answer from candidate options given a question and a passage. There are some large-scale datasets for this task, such as MCTest [10] and RACE [4]. Differing from extractive machine reading comprehension datasets such as SQuAD [8]and NewsQA [15], the correct answer for most questions in MCRC may not directly appear in the original passage.

With the rapid development of deep learning, various neural networks have been proposed for MCRC in recent years [1, 22, 16, 3, 14, 5, 24, 18, 12, 23, 9]. The Stanford AR [1] and GA Reader [3] variants are used to encode question and passage independent of options, ignoring their correlations. Trischler et al. [16] incorporates hierarchy to compare passage, questions and candidate options. The model [24] observes that leveraging candidate options to boost evidence gathering from the passage play a vital role in this task. So go further, the DCMN[23] model the relationship among passage, question and options bidirectionally, and the OCN[9] incorporate the correlation of options to identify more sublte correlations between options to help reasoning. Their approach is based on the premise that options contain rich semantic information.

Recently, the pre-trained language models such as GPT[7], ELMo[6] and BERT[2] have achieved huge success on various nature language processing datasets, including SQuAD [8]. More and more models treat them as a strong encoder to generate contextual representation or even simple make a finetune on them[23, 9].

Almost all the models consider factoid-based questions, in which candidate options can boost the performance of these models. In our RORN model, we focus on opinion-based questions and utilize the strategy that reread options with the information of passage to answer question.

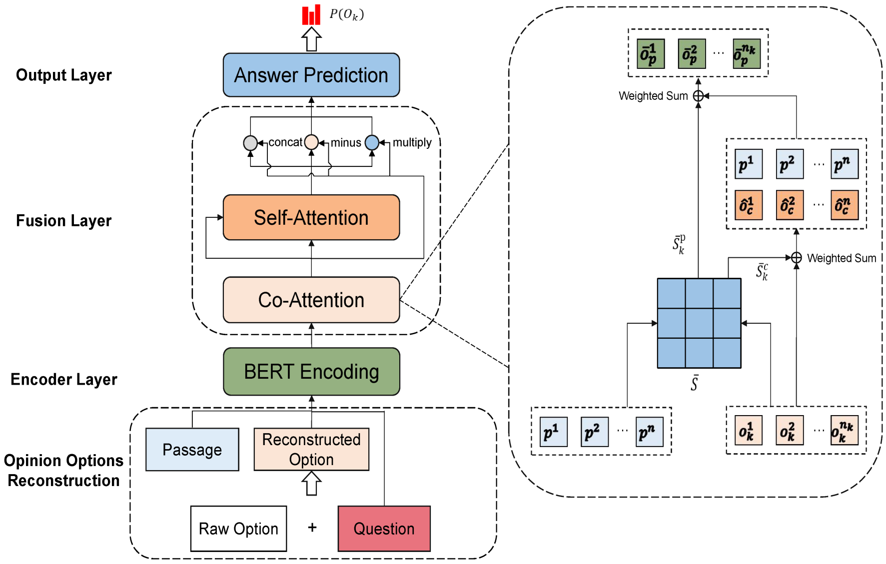

Fig. 2. Framework of our RORN model.

3 Model

The architecture of RORN is show in Figure 2. It contains four modules: (1) Opinion options reconstruction module, which reconstructs the opinion options based on question; (2)Encoder layer module, which extracts features with BERT for passage, question and option respectively; (3) Fusion layer module, which acquires the representation of context-aware options; (4) Output layer module, which is employed to generate the final answer. Then, we formally define MCRC task. There is a passage $P = \{p_1, p_2, ..., p_n\}$ with n tokens , a question $Q = \{q_1, q_2, ..., q_m\}$ with m tokens and a set of options $O = \{O_1, O_2, ..., O_k\}$ with k options, where each option is $O_k = \{o_{k_1}, o_{k_2}, ...o_{k_{n_o}}\}$ is a option with n_o tokens. Our model aims to compute a probability for each option and take the one with higher probability as the prediction answer.

3.1 Opinion Options Reconstruction

In this module, we reconstruct options according to the question to enhance the contextual information of options. For different type of question, we use templates to rewrite them.

We fisrt divide question to three types, which contain **Normal Question**, **Question with different opinions** and **Question with comparison**. For these types of questions, we do respectively:
Normal Question: This type of question always contain only one view of a specific object, like **"在上海读国际高中好吗"** (How about going to international

high school in Shanghai?). We first remove question words such as **“吗”** and **“么”** , then each opinion option replace the raw opinion word in the question to generate reconstructed options respectively.

Question with different opinions: This type of question includes two opposite opinions, like **“早上空腹吃芝士威化饼会不会发胖”** (Do you get fat if you eat cheese wafers on an empty stomach in the morning?), we transform the opposite opinions such as **“会不会”** to single opinion **“会”** or **“不会”** .

Question with comparison: This type of question compare two entities in the same view, like **“学数控技术好还是修车好”** (Is it better to learn NC technology or to repair cars in college?), we split it to two options with only one entity.

If the question cannot be overwritten by templates, we remove question words of question and then cat it with each raw option.

There are some examples shown in the Table 1.

Table 1. Some examples of reconstructed options

Question	Raw Option	Reconstructed Option
在上海读国际高中好吗？(How about going to international high school in Shanghai?)	好(Good)	在上海读国际高中好。(It is good at international high school in Shanghai.)
	不好(Bad)	在上海读国际高中不好。(It is bad at international high school in Shanghai.)
早上空腹吃芝士威化饼会不会发胖？ (Do you get fat if you eat cheese wafers on an empty stomach in the morning?)	会(Yes)	早上空腹吃芝士威化饼会发胖 (You will get fat if you eat cheese wafers on an empty stomach in the morning.)
	不会(No)	早上空腹吃芝士威化饼不会发胖 (You won’t get fat if you eat cheese wafers on an empty stomach in the morning.)
大专学数控技术好还是修车好？(Is it better to learn NC technology or to repair cars in college?)	数控技术 (NC technology)	大专学数控技术好。 (It is better to learn NC technology in college.)
	修车(Repair cars)	大专学修车好。 (It is better to learn to repair cars in college.)

3.2 Encoder Layer

We encode the tokens with BERT [2]. BERT has become one of the most successful natural language representation models in various NLP tasks. BERT’s model architecture is a multi-layer bidirectional Transfomer [17] encoder, which is pre-trained on large-scale corpus. We use BERT as an encoder. It takes as input passage P, question Q and each option O_k, then computes the context-aware representation for each token .

Specifically, given passage $P = \{p_i\}_{i=1}^{m}$,question $Q = \{q_j\}_{j=1}^{n}$, and the k^{th} option $O_k = \{o_j\}_{j=1}^{n_k}$, we pack them as a sequence of length $m + n + n_k + 4$ as follows:

$$S = [\langle CLS\rangle, P, \langle SEP\rangle, Q, \langle SEP\rangle, O_k, \langle SEP\rangle] \tag{1}$$

where $\langle CLS\rangle$ is a specific classifier token and $\langle SEP\rangle$ is a sentence separator which are defined in BERT.

Then the sequence is fed to BERT to generate the context-aware representation for each token in sequence. The output vectors of BERT are denoted as:

$$[\boldsymbol{P}; \boldsymbol{Q}; \boldsymbol{O}_k] = \mathrm{BERT}(S) \tag{2}$$

where $\boldsymbol{P} \in \mathbb{R}^{d\times n}$, $\boldsymbol{Q} \in \mathbb{R}^{d\times m}$, $\boldsymbol{O}_k \in \mathbb{R}^{d\times n_k}$, and $\mathrm{BERT}(\cdot)$ denotes the network defined in [2].

3.3 Fusion layer

This module aims to generate the passage-aware representation of each option and reread each option. We utilize the co-attention[20] mechanism to capture the context information of passage to option. Then the self-attention[19] mechanism apply to understand each option deeply.

First, we define our attention weight function. Given input matrices $\boldsymbol{U} = \{u_i\}_{i=1}^{N} \in \mathbb{R}^{d\times N}$ and $\boldsymbol{V} = \{v_j\}_{i=1}^{M} \in \mathbb{R}^{d\times M}$, We compute the similarity matrix $S \in \mathbb{R}^{N\times M}$, which contains a similarity score s_{ij} for each pair (u_i, v_j):

$$s_{ij} = \boldsymbol{v}^T[\boldsymbol{u}_i; \boldsymbol{v}_j; \boldsymbol{u}_i \circ \boldsymbol{v}_j] \tag{3}$$

where $\circ$ denotes the element-wise multiplication operation and $[\cdot;\cdot]$ denotes column-wise concatenation, And then the attention weight function $\mathrm{Att}(\cdot)$ is defined as:

$$\bar{\boldsymbol{S}} = \mathrm{Att}(\boldsymbol{U}, \boldsymbol{V}) = \left[\frac{\exp{(s_{ij})}}{\sum_i \exp{(s_{ij})}}\right]_{i,j} \tag{4}$$

and $\bar{\boldsymbol{S}} \in \mathbb{R}^{N\times M}$ is the attention weight matrix.

For each option O_k, the co-attention is performed as:

$$\bar{\boldsymbol{S}}_k = \mathrm{Att}(\boldsymbol{O}_k, \boldsymbol{P}) \tag{5}$$

$$\bar{\boldsymbol{S}}_k^p = \mathrm{Att}(\boldsymbol{P}, \boldsymbol{O}_k) \tag{6}$$

$$\hat{\boldsymbol{O}}_k^p = [\boldsymbol{P}; \boldsymbol{O}_k\bar{\boldsymbol{S}}_k]\bar{\boldsymbol{S}}_k^p \tag{7}$$

$$\bar{\boldsymbol{O}}_k^p = \mathrm{ReLU}(\boldsymbol{W}_p\hat{\boldsymbol{O}}_k^p + \boldsymbol{b}_p) \tag{8}$$

where $\bar{\boldsymbol{O}}_k^p \in \mathbb{R}^{d\times n_k}$, $\boldsymbol{W}_p \in \mathbb{R}^{d\times 2d}$ and $\boldsymbol{b} \in \mathbb{R}^d$ are the trainable parameters.

Then, mimicking humans, the options will be reread with passage via self-attention mechanism.

$$\bar{\boldsymbol{O}}_k^s = \mathrm{Att}(\bar{\boldsymbol{O}}_k^p, \bar{\boldsymbol{O}}_k^p) \tag{9}$$

$$\bar{\boldsymbol{O}}_k^f = [\bar{\boldsymbol{O}}_k^p; \bar{\boldsymbol{O}}_k^s; \bar{\boldsymbol{O}}_k^p - \boldsymbol{O}_k^s; \bar{\boldsymbol{O}}_k^p \circ \boldsymbol{O}_k^s;] \tag{10}$$

$$\boldsymbol{O}_k^f = \mathrm{ReLU}(\boldsymbol{W}_f \boldsymbol{O}_k^f + \boldsymbol{b}_f) \tag{11}$$

where $\boldsymbol{W}_f \in \mathbb{R}^{d \times 4d}$ and $\boldsymbol{b}_f \in \mathbb{R}^d$ is the trainable parameter and $\boldsymbol{O}_k^f \in \mathbb{R}^{d \times n_k}$ is the final representation of the k^{th} option.

3.4 Output Layer

To aggregate the final representation for each condidate option, a row-wise max pooling layer is employed to $\bar{\boldsymbol{O}}_k^f$:

$$\bar{\boldsymbol{O}}_k = \mathrm{maxpooling}(\bar{\boldsymbol{O}}_k^f) \tag{12}$$

where $\bar{\boldsymbol{O}}_k \in \mathbb{R}^d$.

And then the score s_k of option O_k to be the correct answer is computed as:

$$s_k = \mathrm{MLP}(\bar{\boldsymbol{O}}_k) \tag{13}$$

where MLP is a 2-layer full connect feed-forward network.

The probability $P(O_k|Q,P)$ of option Q_k to be the correct answer is computed as:

$$P(k|Q,P,O) = \frac{exp(s_k)}{\sum_i exp(s_i)} \tag{14}$$

And our loss function is computed as followed:

$$L(\theta) = -\frac{1}{N}\sum_i log(P(\hat{k}_i|Q_i,P_i,O_i)) \tag{15}$$

where θ denotes all trainable parameters, N is the training example number, and $\hat{k}_i$ is the ground truth for the i^{th} example.

4 Experiments

4.1 Experimental settings

Dataset

We conduct experiments on the Opinion Questions Machine Reading Comprehension dataset in the AIChallenger competition [2], in which questions are option-based questions. There are 270,000 and 30,000 examples in the training set and development set respectively. We divide the original development set into two parts evenly, one as a split development set for tuning model and the other one as a split test set, which contains 15,000 examples respectively.

Implementation Details

[2] The dataset can be downloaded in https://challenger.ai/competition/oqmrc2018.

Our model is implemented with pytorch[3], and uses the framework[4] for BERT model. We use pre-trained BERT on chinese corpus[5] to initialize our encoder. We use Adam optimizer and the learning rate uses the linear schedule to decrease from 3×10^{-5} to 0. Passages, questions and options are trimmed to 300, 30 and 30 tokens respectively. In this work, other hyper-parameter are shown in Table 2

Table 2. Hyper-parameter of our model

Paramete Name	**Value**
Train epochs	5
Batch size	12
Hidden units	786
Learning rate	0.00003
Dropout	0.8
Max sequence length	384

4.2 Baselines

We choose several baselines:

(1) **MwAN** [13] is a baseline for modeling sentence pair. It proposes the multi-way attention network which employ multiple attention function. It is provided as a baseline by the official[6].
(2) **BiDAF** [11] is a strong baseline for MRC tasks. It is a typical neural-based MRC model which utilizes bi-directional attention to obtain query-aware context representation. We compress the representation with max-pooling layer, then feed it into a 2-layer full connect feed-forward network for classification.
(3) **RNET**[7] [19] is one of the top MRC models. It introduces a self-matching attention mechanism to refine the representation by matching the passage against itself. The model is designed for SQuAD-style datasets. So we replace its output layer with a 2-layer full connect feed-forward network.
(4) **BERT** [2] is a powerful pre-trained language model based on Transfomer[17]. We finetune the model with a linear layer on top of the pooled output of BERT.

[3] https://github.com/pytorch/pytorch
[4] https://github.com/huggingface/pytorch-pretrained-BERT
[5] https://storage.googleapis.com/bert_models/2018_11_03/chinese_L-12_H-768_A-12.zip
[6] https://github.com/AIChallenger/AI_Challenger_2018
[7] https://github.com/HKUST-KnowComp/R-Net

Table 3. The results of different models.

Model	Dev(%acc)	Test(%acc)
MwAN [13]	69.98	69.16
BiDAF [11]	70.48	69.32
RNET [19]	72.84	72.42
BERT finetune [2]	74.18	73.75
RORN(ours)	**79.07**	**78.33**

4.3 Experimental results

Table 3 shows the results of our RORN model achieve better than other models, which is 4.58% higher in value than BERT finetune model. Our single RORN model achieves 78.33% in term of accuracy.

The modified RNET model is our original baseline for the competition, which can be at the top of the competition leaderboard with complicated data preprocessing. Our RORN model has huge improvement than RNET model, which demonstrates that our model can effectively handle opinion-based questions.

4.4 The Effectiveness of Option Reconstruction

To study the effectiveness of opinion-based option reconstruction, we conduct experiments on the dataset. Table 4 presents their comparison results. We can observe that the RORN model without option reconstruction shows performance drop with 1.81%. The results demonstrate that our reconstruction module can effectively obtain the context of opinion options to improve performance.

4.5 Attention Mechanisms Ablation

In this section, we conduct ablation stduy on attention mechanisms to examine the effectiveness of each attention mechanism. The experimental results are listed in Table 5.

From the best model, if we remove the co-attention mechanism, the accuracy drops by 1.16% on the test set, and if the self-attention mechanism is removed, the accuracy drops by appropriately 1.12% on the test set. The results suggests that rereading the options with passage has more important to guide the model.

When we remove all attention mechanisms, the performance of the model drops 2.25%, which demonstrate that attention mechanisms is indispensable for our model.

4.6 Error Analysis

Based on the analysis of misclassified our instances, we can find some main reasons for misclassification as follows:

Table 4. Effectiveness of option reconstruction.

Model	Dev(%acc)	Test(%acc)
RORN w/o Reconstruction	77.25	76.52
RORN	79.07	78.33

Table 5. Influence of different attention mechanisms.

Model	Dev(%acc)	Test(%acc)
RORN w/o attention	76.81	76.08
RORN w/o co-attention	77.24	77.17
RORN w/o self-attention	77.82	77.21
RORN	79.07	78.33

Table 6. Error instances

Instances
柠檬水的正确泡法用凉水还是热水？ (Cold or hot water is the correct way to soak lemonade?) 泡柠檬的水温一般在60到70℃比较合适。 (The water temperature to make soak lemonade is between 60 and 70 degrees centigrade.)
小孩头皮撞破皮了需要剃光头吗? (Does a child need to shave his head when his scalp breaks?) 宝宝头皮撞破最好是别沾水,会引起伤口发炎的。 (Baby scalp breakage is best not to touch water, it will cause wound inflammation.)

(1) Some example need external knowledge or complicated reasoning to infer answer. Taking the fisrt instance in Table 6 as a example, we need to know common sense like ”Water of 60 to 70 degrees is hot water.” to identify the right answer.
(2) There are some questions whose answers are uncertain according to the passage. The second instance in Table 6 shows a similar situation. We should introduce corresponding solutions to distinguish these questions which can not answer with context.
(3) Some manually annotated example are ambiguous or wrong, which mislead our model to predict wrong answer.

5 Conclusion

In this paper, we propose RORN model for opinion questions reading comprehension task. We use simple but effective method to reconstruct the opinion-based options, which can obtain the context information of options. Then the lastest breakthrough, BERT, is treated as our power encoder in our model. Mimicking humans, two attention mechanisms are employed to fuse semantic information among the passage, question and options. The experimental results demonstrate that our option reconstruction can boost our performance and two type of attention mechanism can influence the context-level fusion.

6 Acknowledgements

This work was supported by the National Natural Science Foundation of China (No.61772135, No.U1605251, No.61533018), the Natural Key R&D Program of China (No. 2018YFC0830101). This work was also supported by the Open Project of Key Laboratory of Network Data Science & Technology of Chinese Academy of Sciences (No.CASNDST201708 and No.CASNDST201606), the Open Project of National Laboratory of Pattern Recognition at the Institute of Automation of the Chinese Academy of Sciences (201900041).

References

1. Chen, D., Bolton, J., Manning, C.D.: A thorough examination of the cnn/daily mail reading comprehension task. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 2358–2367 (2016)
2. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. CoRR **abs/1810.04805** (2018), `http://arxiv.org/abs/1810.04805`
3. Dhingra, B., Liu, H., Yang, Z., Cohen, W., Salakhutdinov, R.: Gated-attention readers for text comprehension. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1832–1846 (2017)
4. Lai, G., Xie, Q., Liu, H., Yang, Y., Hovy, E.: Race: Large-scale reading comprehension dataset from examinations. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 785–794 (2017)
5. Parikh, S., Sai, A.B., Nema, P., Khapra, M.M.: Eliminet: a model for eliminating options for reading comprehension with multiple choice questions. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence. pp. 4272–4278. AAAI Press (2018)
6. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. arXiv preprint arXiv:1802.05365 (2018)
7. Radford, A., Narasimhan, K., Salimans, T., Sutskever, I.: Improving language understanding by generative pre-training. URL https://s3-us-west-2. amazonaws. com/openai-assets/research-covers/languageunsupervised/language understanding paper. pdf (2018)
8. Rajpurkar, P., Zhang, J., Lopyrev, K., Liang, P.: Squad: 100,000+ questions for machine comprehension of text. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 2383–2392 (2016)
9. Ran, Q., Li, P., Hu, W., Zhou, J.: Option comparison network for multiple-choice reading comprehension. arXiv preprint arXiv:1903.03033 (2019)
10. Richardson, M., Burges, C.J., Renshaw, E.: Mctest: A challenge dataset for the open-domain machine comprehension of text. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. pp. 193–203 (2013)

11. Seo, M., Kembhavi, A., Farhadi, A., Hajishirzi, H.: Bidirectional attention flow for machine comprehension. arXiv preprint arXiv:1611.01603 (2016)
12. Sun, K., Yu, D., Yu, D., Cardie, C.: Improving machine reading comprehension with general reading strategies. arXiv preprint arXiv:1810.13441 (2018)
13. Tan, C., Wei, F., Wang, W., Lv, W., Zhou, M.: Multiway attention networks for modeling sentence pairs. In: IJCAI. pp. 4411–4417 (2018)
14. Tay, Y., Tuan, L.A., Hui, S.C.: Multi-range reasoning for machine comprehension. arXiv preprint arXiv:1803.09074 (2018)
15. Trischler, A., Wang, T., Yuan, X., Harris, J., Sordoni, A., Bachman, P., Suleman, K.: Newsqa: A machine comprehension dataset. arXiv preprint arXiv:1611.09830 (2016)
16. Trischler, A., Ye, Z., Yuan, X., He, J., Bachman, P.: A parallel-hierarchical model for machine comprehension on sparse data. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). vol. 1, pp. 432–441 (2016)
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in neural information processing systems. pp. 5998–6008 (2017)
18. Wang, S., Yu, M., Jiang, J., Chang, S.: 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). pp. 746–751 (2018)
19. Wang, W., Yang, N., Wei, F., Chang, B., Zhou, M.: Gated self-matching networks for reading comprehension and question answering. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 189–198 (2017)
20. Xiong, C., Zhong, V., Socher, R.: Dynamic coattention networks for question answering. arXiv preprint arXiv:1611.01604 (2016)
21. Xu, Y., Liu, J., Gao, J., Shen, Y., Liu, X.: Towards human-level machine reading comprehension: Reasoning and inference with multiple strategies. arXiv preprint arXiv:1711.04964 (2017)
22. Yin, W., Ebert, S., Schütze, H.: Attention-based convolutional neural network for machine comprehension. In: Proceedings of the Workshop on Human-Computer Question Answering. pp. 15–21 (2016)
23. Zhang, S., Zhao, H., Wu, Y., Zhang, Z., Zhou, X., Zhou, X.: Dual co-matching network for multi-choice reading comprehension. arXiv preprint arXiv:1901.09381 (2019)
24. Zhu, H., Wei, F., Qin, B., Liu, T.: Hierarchical attention flow for multiple-choice reading comprehension. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)