Overview of NLPCC Shared Task 2: Multi-perspective Scientific Machine Reading Comprehension

Xiao Zhang, Heqi Zheng, Yuxiang Nie, and Xian-Ling Mao(✉)

School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
maoxl@bit.edu.cn

Abstract. In this report, we give an overview of the shared task about multi-perspective scientific machine reading comprehension at the 12th CCF Conference on Natural Language Processing and Chinese Computing (NLPCC 2023). Scientific machine reading comprehension (SMRC) aims to understand scientific texts through interactions with humans by given questions. In this task, questions about scientific texts include perspectives from beginners, students and experts. It requires different levels of understanding of scientific texts. We describe the task, the corpus, the participating teams and their results.

Keywords: Machine reading comprehension · Multi-perspective · NLPCC 2023

1 Instruction

In today's fast-paced world, there are countless articles and information created around the world every day in the news field, self-media field and even technology field. Therefore, it is impossible to fully digest every article for us. Machine Reading Comprehension (MRC) can help us understand this information more quickly and obtain useful information from it. Based on machines' ability to understand natural language, MRC can extract relevant content from a large amount of information based on the questions we ask and make answers after understanding the content in a short time.

Scientific machine reading comprehension (SMRC) aims to understand scientific texts through interactions with humans by given questions. The ability of machines to understand and make sense of scientific texts is crucial for many applications such as scientific research [1,4,8], education [2,5] and industry [3,7,11]. With the increasing amount of scientific literature being produced, the need [6,9,10] for machines to understand these texts is becoming more pressing.

F. Liu et al. (Eds.): NLPCC 2023, LNAI 14304, pp. 102–107, 2023.
https://doi.org/10.1007/978-3-031-44699-3_10

2 The Task

As far as we know, there is only one dataset [6] focused on exploring full-text scientific machine reading comprehension, which is proposed to improve MRC models in seeking information from specific papers with questions. However, the dataset has ignored the fact that different readers may have different levels of understanding of the text, and only includes single-perspective question-answer pairs from annotators whose background is NLP, which leads to a lack of consideration of different perspectives, especially for beginner's and expert's perspectives. Different perspectives correspond to different types of problems, which requires different levels of understanding. It will help us analyze and explore machine reading comprehension from a more comprehensive perspective. Therefore in NLPCC 2023, we offer a multi-perspective scientific machine reading comprehension task.

3 The Dataset

The provided dataset is referred as the SciMRC corpus in the following. It contains a training set, a validation set, and a test set. For the training set, it contains a large set of scientific papers from top conferences in natural language processing (e.g. ACL, EMNLP, NAACL, etc.) as well as corresponding human-written question-answer pairs (QA pairs) and their evidence paragraphs/figures/tables, which denotes the specific information in the paper that can support the answer to the question. The data is used for machine reading comprehension on scientific papers. The training and validation datasets include 4,873 QA pairs with their evidence while the test set contains 1,169 QA pairs with their evidence. As shown in Table 1, we collect QA pairs from different perspectives (i.e. BEGINNERS, STUDENTS, EXPERTS) to enhance the diversity of the data in the SciMRC and calculate the average of the paper length, the figure/table number, the question length and the evidence sentence number for each perspective.

Table 1. Representative features from SciMRC categorized by different perspectives

Type	Paper	Figure/Table	Question		Evidence
PERSPECTIVE	Avg Paper Length	Avg Figure/Table Number	Avg Question Length	Avg Answer Length	Avg Evidence Sentence Number
BEGINNERS	3725.6	5.32	10.0	17.2	1.39
STUDENTS			9.8	11.7	1.08
EXPERTS			22.4	95.9	4.56
ALL			11.0	21.8	1.56

3.1 Data Format

The training data contains a file and a directory, one file for the scientific papers with evidence and the other directory contains images and tables. In the training file, each json item contains six fields: "id" "title" "abstract" "full_text" "qas" and "figures_and_tables".

For evaluation, every line (in the json format) contains a paper with its question and the answer and evidence are absent. Each submission must contain a single json file with the name `answer.json`, with each key corresponding to a question id in the test set and its value is the answer to the question.

All files are encoded in UTF-8.

Obtaining the Dataset: You may download the training data from https://drive.google.com/file/d/1ewbgZOy6CEpjzoVxnkQPPVItj6yslUi1/view?usp=sharing. The test data is available at https://drive.google.com/file/d/1N2fVmr-InkIA8rdEoXrtIj6ENmDaGkrw/view.

Use of the Data: You are free to use the data for research purpose and please cite the dataset paper with the following bib entry (Tables 2 and 3).

```
@article{zhang2023scimrc,
  title={SciMRC: Multi-perspective Scientific Machine Reading
  Comprehension},
  author={Zhang, Xiao and Zheng, Heqi and Nie, Yuxiang and Huang,
  Heyan and Mao, Xian-Ling},
  journal={arXiv preprint arXiv:2306.14149},
  year={2023}
}
```

Table 2. A total of 16 teams from the global industrial and academic sectors are participating in our competition

Team ID	System Name
1	Evay Info AI Team
2	Dependency Graphs For Reading Comprehension
3	OUC_NLP
4	Langdiaozheyang
5	Emotional damage
6	Mirror
7	huawei_tsc_zeus
8	Lastonestands
9	cisl-nlp
10	CUHK_SU
11	its666
12	zutnlp-wujiahao
13	MPSMRC_cup
14	IMU_NLP
15	Nicaiduibudui
16	PIE

4 Evaluation Metric

In this paper, we utilized RougeL as our evaluation metric. RougeL is a commonly used metric for assessing the quality of text summarization systems. It measures the overlap between the generated summary and a reference summary using the longest common subsequence (LCS) algorithm. RougeL computes the length of the LCS between the two summaries and normalizes it by the length of the reference summary. This metric allows us to quantitatively evaluate the performance of our summarization system based on the similarity and coverage of the generated summaries compared to the reference summaries. The formula for RougeL can be expressed as:

$$\mathcal{R}_{LCS} = \frac{LCS(Prediction, Golden)}{len(Golden)} \tag{1}$$

$$\mathcal{P}_{LCS} = \frac{LCS(Prediction, Golden)}{len(Prediction)} \tag{2}$$

$$\mathcal{F}_{LCS} = \frac{(1+\beta^2)\mathcal{R}_{LCS}\mathcal{P}_{LCS}}{\mathcal{R}_{LCS}+\beta^2\mathcal{P}_{LCS}} \tag{3}$$

5 Participating Teams

A total of 16 teams from the global industrial and academic sectors are participating in our competition.

6 Evaluation Results

The teams were ranked based on their performance in the evaluation, and the final scores represent their respective achievements. The team 'Nicaiduibudui' secured the top position with a score of 0.5459, followed by 'IMUNLP' with a score of 0.4519. 'PIE' and 'OUC_NLP' also performed well, obtaining scores of 0.4181 and 0.3574, respectively."

Table 3. Final Leaderboard

Team ID	System Name	Final Score
1	Nicaiduibudui	0.5459
2	IMUNLP	0.4519
3	PIE	0.4181
4	OUC_NLP	0.3574

7 Conclusion

We had a total of 16 teams participating in the competition and 4 of them submitted their final results. Each team developed their own system for the task at hand. The evaluation of the systems was performed using the RougeL metric, which is a widely used measure for assessing the quality of text summarization. In the field of machine reading, there are still significant challenges to overcome, but there is also considerable room for future development.

Acknowledgments. We thank the colleagues from Beijing Institute of Technology, especially the DataHammer research group to write potential questions for scientific papers. The annotation of the Multi-perspective Scientific Machine Reading Comprehension dataset is supported by National Key R&D Plan (No. 2020AAA0106600) and National Natural Science Foundation of China (No. 62172039). We also thank the participants for their valuable feedback and outstanding results.

References

1. Beltagy, I., Lo, K., Cohan, A.: SciBERT: a pretrained language model for scientific text. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 3615–3620. Association for Computational Linguistics, Hong Kong, China, November 2019. https://doi.org/10.18653/v1/D19-1371, http://aclanthology.org/D19-1371
2. Bianchi, N., Giorcelli, M.: Scientific education and innovation: From technical diplomas to university stem degrees. Microeconometric Studies of Education Markets (Topic), ERN (2019)
3. Bruches, E., Tikhobaeva, O., Dementyeva, Y., Batura, T.: TERMinator: a system for scientific texts processing. In: Proceedings of the 29th International Conference on Computational Linguistics, pp. 3420–3426. International Committee on Computational Linguistics, Gyeongju, Republic of Korea, October 2022. www.aclanthology.org/2022.coling-1.302
4. Cachola, I., Lo, K., Cohan, A., Weld, D.: TLDR: extreme summarization of scientific documents. In: Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 4766–4777. Association for Computational Linguistics, November 2020. https://doi.org/10.18653/v1/2020.findings-emnlp.428, www.aclanthology.org/2020.findings-emnlp.428
5. de la Chica, S., Ahmad, F., Martin, J.H., Sumner, T.R.: Pedagogically useful extractive summaries for science education. In: International Conference on Computational Linguistics (2008)
6. Dasigi, P., Lo, K., Beltagy, I., Cohan, A., Smith, N.A., Gardner, M.: A dataset of information-seeking questions and answers anchored in research papers. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4599–4610 (2021). www.aclanthology.org/2021.naacl-main.365/
7. Erera, S., et al.: A summarization system for scientific documents. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing

(EMNLP-IJCNLP): System Demonstrations, pp. 211–216. Association for Computational Linguistics, Hong Kong, China, November 2019. https://doi.org/10.18653/v1/D19-3036, www.aclanthology.org/D19-3036
8. Marie, B., Fujita, A., Rubino, R.: Scientific credibility of machine translation research: a meta-evaluation of 769 papers. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 7297–7306. Association for Computational Linguistics, August 2021. https://doi.org/10.18653/v1/2021.acl-long.566, www.aclanthology.org/2021.acl-long.566
9. Sadat, M., Caragea, C.: SciNLI: a corpus for natural language inference on scientific text. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 7399–7409. Association for Computational Linguistics, Dublin, Ireland, May 2022. https://doi.org/10.18653/v1/2022.acl-long.511, www.aclanthology.org/2022.acl-long.511
10. Wadden, D., et al.: Fact or fiction: verifying scientific claims. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 7534–7550. Association for Computational Linguistics, November 2020. https://doi.org/10.18653/v1/2020.emnlp-main.609, www.aclanthology.org/2020.emnlp-main.609
11. Zulfiqar, S., Wahab, M.F., Sarwar, M.I., Lieberwirth, I.: Is machine translation a reliable tool for reading German scientific databases and research articles? J. Chem. Inf. Model. **58**(11), 2214–2223 (2018)