Overview and Insights from Scope Detection of the Peer Review Articles Shared Tasks 2021

Saichethan Miriyala Reddy[1] and Naveen Saini[2(✉)]

[1] Indian Institute of Information Technology Bhagalpur, Bhagalpur, Bihar, India
miriyala.cse.1725@iiitbh.ac.in

[2] Université Toulouse III - Paul Sabatier, IRIT, UMR 5505 CNRS, Toulouse, France
Naveen.Saini@irit.fr

Abstract. In the current paper, we will present the results of our shared task at *The First Workshop & Shared Task on Scope Detection of the Peer Review Articles* (SDPRA) collocated with PAKDD 2021. It aims to develop system(s) which can help in the peer-review process in the initial screening usually performed by the editor(s). We received four submissions in total: three from academic institutions and one from the industry. The quality of submission shows a greater interest in the task by the research community.

1 Introduction

In recent years, scientific articles are published at a rapid rate and can be accessed using search capability of scholarly search engines like Google Scholar[1], dblp[2], among others. This continuous growth help the researchers in getting the recent developments in the respective domain [3] as well as discovering new research dire. But before publishing, peer review is the widely accepted method for the validation of the submitted papers to the conferences/journals. In the academic peer review process, the first stage begins at the editor's desk where task is to identify the in-scope and out-of-scope articles and then reject out-of-scope submissions or in other words, it is the job of the editor, who is also an expert in the particular field to take decisions whether an article should be rejected without further review or forwarded to expert reviewers for meticulous evaluation. Some of the common reasons for rejection are due to paper's language and writing/formatting style, results are not better than state-of-the-art, method is too simple (seriously? Isn't that a good thing?), too narrow or outdated or out of scope, among others.

Without any automatic system, the Editor may spends a substantial amount of time in finding the appropriateness of the submitted article or before passing it to the reviewers for review purpose. Inspite of having good quality of the submitted articles, many articles got rejection because of their out-of-score [18].

[1] https://scholar.google.com/.
[2] https://dblp.uni-trier.de/.

M. Gupta and G. Ramakrishnan (Eds.): PAKDD 2021 Workshops, LNAI 12705, pp. 73–78, 2021.
https://doi.org/10.1007/978-3-030-75015-2_7

Machine learning and artificial intelligence methods make it possible to identify the in-scope and out-of-scope of the scientific publications. However, in order to improve the performance of these methods and to carry out the experiments, researchers needs to access and use of large database of scientific publications. This shared task aims to bring undergraduate and master students, NLP/ML researchers and from other backgrounds who: (a) have a deep interest in mining scientific articles; (b) develop novel methods that able to improve the performance. The purpose of the shared task is to provide a platform for developing and evaluating such models. To this end, the shared task provides a dataset covering a broad range of topics within the computer science. The importance of the shared task can be easily understood from the recent papers [6,11,17].

In this paper, we present the task description, dataset, discussion about the participating systems followed by their results and insight from shared task.

2 Problem Definition and Datasets

This shared task focuses on identifying topics or category of scientific articles, which helps in pre-screening of the scientific articles at the Editor's desk to decide whether the article is out-of-score or not. For this task we collected a total of 35000 abstracts of the scientific article in computer science domain from ARXIV. Given an abstract of a paper, the objective of the shared task is to classify it into one of the 7 predefined categories listed in Table 1. The dataset is divided into training, validation and testing set. The statistics about the datasets is shown in Fig. 1. We have openly released our datasets at http://dx.doi.org/10.17632/njb74czv49.1 as a part of Mendeley data [16].

Table 1. Pre-defined categories for scientific article classification and their description. Here, Abb. means abbreviation of the categories used in this paper.

Category Name	Abb.	Description of the category
Computation & Language(CL)	CL	It can be defined as an application of computer science to the analysis, synthesis and comprehension of written and spoken language.
Cryptography & Security(CR)	CR	In this category, secure communication related articles are included.
Distributed & Cluster Computing	DC	It refers to the application of cluster and distributed computing systems in order to leverage processing power, memory, etc., of any computing system for an increase in efficiency.
Data Structures & Algorithms	DS	It refers to the study of using an entity that stores and organizes data in order to develop algorithms that help to decode and optimize computer programs.
Logic in Computer Science	LO	It refers to the overlap between the field of logic and that of computer science.
Networking & Internet Architecture	NI	It refers to the specification of a network's physical components and their functional organization and configuration.
Software Engineering	SE	It is the systematic application of engineering approaches to the development of software.

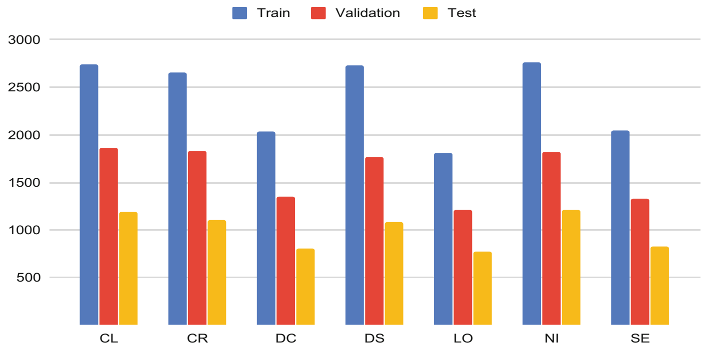

Fig. 1. Distribution of data

3 Participating Teams

For this shared task, 18 teams have registered, but only 4 were able to submit their systems and were able complete the final evaluation correctly. Description about the participated team and their methods are brief described below. The systems of all the teams are based on supervised techniques.

- **UTMN:** This team was from University of Tyumen, Russia. The method [8] developed by this team uses the SciBERT language model [1] along with topic modeling tool (Latent Dirichlet Allocation [2]) to obtain the good quality of predictions. Several other language models like RoBERTa [13], PubMedBERT [9], Topic-informed BERT-based models (tBERT) [14], among others, are also investigated to analyze the performance.
- **IIITT:** This system was submitted by the team members from Indian Institute of Information Technology Tiruchirapalli, India. The authors [10] utilize the transfer learning-based approach for the scope identification of the scientific articles. They fine-tuned the existing model namely, BERT, RoBERTa, and SCIBERT using the provided datasets and then, tested on the test dataset. Their results shows that ensemble approach over the three models perform the best.
- **parklize:** The author of this method [15] was from Maynooth University, Ireland and uses sentence embedding obtained using pre-trained SciBERT model along with entity embeddings mentioned in the text. The entities in the abstract are tagged using the *TagMe* model [5] and then, wikipedia2vec tool [19] is used to get the embeddings of the entities. For classification, seven classifiers are used followed by majority-voting ensemble approach.
- **FideLIPI:** The system by Ghosh et al. [7] developed four different subsystems for the shared task and ensemble the predictions of these four subsystems. The four sub-systems includes: (a) Abstract (of the scientific articles) level RoBERTa; (b) RoBERTa based model with additional features extracted using Latent Dirichlet Allocation [2]; (c) Sentence level RoBERTa model, and (d) Logistic regression [12] model utilizing features provided by tf-idf model [4].

4 Results and Discussion

As our task deals with classification; therefore, the official evaluation metric for the shared task is kept as weighted-average F1 score which is well-know in the field of Machine learning.

4.1 Results

Table 2 show the top run of each participating system. Although we allowed teams to submit an unlimited number of runs since this is an offline evaluation on a blind test set, we only tabulate the results from the top 3 runs. In order to compare between the systems, we considered the $F1$ scores. The ranking of systems based on their performance is also shown in the first column of Table 2. It has been observed that the system *UTMN* obtained the first rank and *FideLIPI* obtained the second rank. All teams have utilized transformer-based language models to develop their systems.

All systems shown to have a good performance. We have noticed that high classification performance of the abstracts are due to the use of single domain like computer science and also due to creating the test set from same sources as development and training. From a language perspective, most of the systems utilized transformer based language models, some even used pre-trained models on scientific corpora. However, we believe that more efforts should be focused developing computationally simpler models. In the future we plan to extend this corpus to different discipline & domains like history, astro-physics, etc., with the hope that SDPRA will help foster further research in this domain.

Table 2. The results reported by different submitted systems and their ranking

Rank	Team name	F1 Score		
		Run 1	Run 2	Run 3
1	UTMN	0.9370	0.9354	**0.9382**
2	FideLIPI	**0.9293**	0.9122	0.9163
3	IIITT	0.9227	0.9156	**0.9246**
4	Parklize	0.9173	**0.9206**	0.9151

4.2 Code Reproducibility

To improve code reproducibility and transparency in scientific community, all shared task participants are asked to submit their systems to our Github Repository[3]. The codes are open to use for the scientific community.

[3] https://github.com/SDPRA-2021/shared-task.

5 Conclusion and Future Scope

The first shared task on 'Scope Detection of the Peer Review Articles' comprises identification of *in-scope* and *out-of-scope* categories using only the abstract of the *arxiv* articles. We received four submissions. The reported high performance by different systems shows the high coherence among the abstracts, which needs expansion by including the abstracts of the other domains. Moreover, most of the submitted systems were based on utilizing a deep learning-based model, but we are interested in an elementary and unsupervised model. We are also interested in collaborating with NLP and AI researchers to interact and discuss the new tools to be developed for the scope identification.

Acknowledgment. We would like to thanks all steering committee members, Pushpak Bhattacharyya (IIT Bombay), Antoine Doucet (Universite de La Rochelle), Sriparna Saha (IIT Patna), Jose G Moreno (Universite de Toulouse, IRIT), Adam Jatowt (Kyoto University), Anita de Waard (Elsevier), Gael Dias (University of Caen Normandie) for their support. We are also thankful to all the members of our program committee for their time and valuable inputs.

References

1. Beltagy, I., Lo, K., Cohan, A.: SciBERT: a pretrained language model for scientific text. arXiv preprint arXiv:1903.10676 (2019)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
3. Cohan, A., Goharian, N.: Scientific document summarization via citation contextualization and scientific discourse. Int. J. Digit. Libr. **19**(2–3), 287–303 (2018). https://doi.org/10.1007/s00799-017-0216-8
4. Fautsch, C., Savoy, J.: Adapting the TF IDF vector-space model to domain specific information retrieval. In: Proceedings of the 2010 ACM Symposium on Applied Computing, pp. 1708–1712 (2010)
5. Ferragina, P., Scaiella, U.: TAGME: on-the-fly annotation of short text fragments (by wikipedia entities). In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 1625–1628 (2010)
6. Ghosal, T., Verma, R., Ekbal, A., Saha, S., Bhattacharyya, P.: Investigating impact features in editorial pre-screening of research papers. In: Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries, pp. 333–334 (2018)
7. Ghosh, S., Chopra, A.: Using transformer based ensemble learning to classify scientific articles. In: Gupta, M., Ramakrishnan, G. (eds.) PAKDD 2021. LNAI, vol. 12705, pp. 106–113. Springer, Heidelberg (2021)
8. Glazkova, A.: Identifying topics of scientific articles with BERT-based approaches and topic modeling. In: Gupta, M., Ramakrishnan, G. (eds.) PAKDD 2021. LNAI, vol. 12705, pp. 98–105. Springer, Heidelberg (2021)
9. Gu, Y., et al.: Domain-specific language model pretraining for biomedical natural language processing. arXiv preprint arXiv:2007.15779 (2020)
10. Hande, A., Puranik, K., Priyadharshini, R., Chakravarthi, B.R.: Domain identification of scientific articles using transfer learning and ensembles. In: Gupta, M., Ramakrishnan, G. (eds.) PAKDD 2021. LNAI, vol. 12705, pp. 88–97. Springer, Heidelberg (2021)

11. Kelly, J., Sadeghieh, T., Adeli, K.: Peer review in scientific publications: benefits, critiques, & a survival guide. Ejifcc **25**(3), 227 (2014)
12. Kleinbaum, D.G., Dietz, K., Gail, M., Klein, M., Klein, M.: Logistic Regression. Springer, Heidelberg (2002). https://doi.org/10.1007/b97379
13. Liu, Y., et al.: Roberta: a robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
14. Peinelt, N., Nguyen, D., Liakata, M.: tBERT: topic models and BERT joining forces for semantic similarity detection. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7047–7055 (2020)
15. Piao, G.: Scholarly text classification with sentence BERT and entity embeddings. In: Gupta, M., Ramakrishnan, G. (eds.) PAKDD 2021. LNAI, vol. 12705, pp. 79–87. Springer, Heidelberg (2021)
16. Reddy, S., Saini, N.: SDPRA 2021 shared task data V1 (2021). https://doi.org/10.17632/njb74czv49.1
17. Sergeeva, E., Zhu, H., Prinsen, P., Tahmasebi, A.: Negation scope detection in clinical notes and scientific abstracts: a feature-enriched LSTM-based approach. In: AMIA Summits on Translational Science Proceedings 2019, p. 212 (2019)
18. Trumbore, S., Carr, M.E., Mikaloff-Fletcher, S.: Criteria for rejection of papers without review (2015)
19. Yamada, I., et al.: Wikipedia2vec: an efficient toolkit for learning and visualizing the embeddings of words and entities from wikipedia. arXiv preprint arXiv:1812.06280 (2018)