Academic Reader: An Interactive Question Answering System on Academic Literatures

Yining Hong, Jialu Wang, Yuting Jia, Weinan Zhang, Xinbing Wang
Shanghai JiaoTong University
800 Dongchuan RD. Minhang District, Shanghai, China
{evelinehong, faldict, hnxxjyt, wnzhang, xwang8}@sjtu.edu.cn

Abstract

We present ACADEMIC READER, a system which can read academic literatures and answer the relevant questions for researchers. ACADEMIC READER leverages machine reading comprehension technique, which has been successfully applied in many fields but has not been involved in academic literature reading. An interactive platform is established to demonstrate the functions of ACADEMIC READER. Pieces of academic literature and relevant questions are input to our system, which then outputs answers. The system can also gather users' revised answers and perform active learning to continuously improve its performance. A case study is provided presenting the performance of our system on all papers accepted in KDD 2018, which demonstrates how our system facilitates massive academic literature reading.

Introduction

Recent years have witnessed an astounding growth in the number of published academic papers, which requires researchers to make much effort on surveys and investigations. In another word, researchers have to look through papers to find out their concerned questions, e.g., *What is the aim of this paper?*, *What method does the authors propose?*, *What is the advantage of the proposed model compared to previous works?* etc. Machine reading comprehension (MRC) system (Hirschman et al. 1999) can assist researchers in browsing through academic literatures, extracting valuable information, filtering and sorting works according to their specific needs and interests.

Teaching machines to accomplish reading comprehension tasks has made remarkable progress these years, focusing on corpus including news (Hermann et al. 2015), Wikipedia articles (Rajpurkar et al. 2016; Rajpurkar, Jia, and Liang 2018) and web documents (Nguyen et al. 2016). A variety of MRC models have been investigated to solve such tasks, including QANet (Yu et al. 2018), RNet (Wang et al. 2017), DCN (Xiong, Zhong, and Socher 2016). Similarly, MRC could also be applied into the domain of academic literatures, serving as an alternative to human efforts in reading and summarizing massive papers and making significant convenience for the researchers.

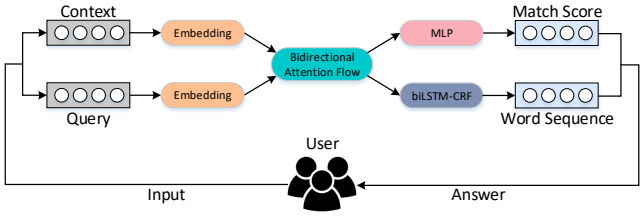


Figure 1: The overview of ACADEMIC READER

However, currently there are no such systems concerning reading comprehension tasks on academic literatures. In this paper, we build ACADEMIC READER, an interactive question answering system which automatically extracts essential information from academic literatures to answer given questions. An online platform[1] is established to demonstrate the results of our system. Researchers can upload academic literatures, such as paper abstracts, along with several questions, and review the predicted answers provided by our system.

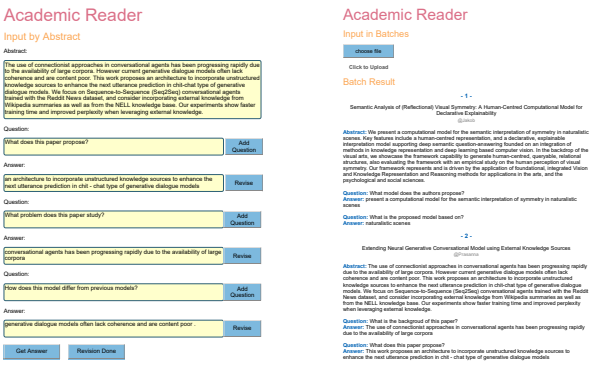
System Overview

ACADEMIC READER aims to build an interactive question answering system on academic literatures. As is shown in Fig 1, users provide pieces of academic literature as contexts, as well as any number of questions concerning the contexts, and then ACADEMIC READER will run our model and immediately generate answers to these questions. Users can revise the answers which fail to satisfy them, thus further optimizing our system through active learning.

User Input Interface

Since our system is mainly trained on academic abstracts in the field of computer science, it will be best for users to input texts of this kind. There are three ways for input. 1) Users can enter a single academic abstract followed by several questions about the abstract, as is shown in Fig 2a. 2) Users can enter an arXiv ID followed by several questions. Our system will generate the abstract by itself and provide answers. 3) ACADEMIC READER also provides batch processing option for users who need to process a large number of papers. To do this, users should upload a text file in required format, which includes all the abstracts and for each abstract, a set of questions. Our system will return a result page providing answers, as is shown in Fig 2b.

[1] http://bit.ly/AcademicReader



(a) Single Abstract Input (b) Batch Abstract Input

Figure 2: System demonstration for single abstract input and batch abstract input.

Reading Comprehension Model

The context (abstract) and queries (questions) will then be input to our system, which runs a reading comprehension model, as is shown in Fig 1. In our setting, answers are spans (i.e., sequences of words) in the abstract. To generate a span as the answer for each query, we locate an evidence sentence at first and then pinpoint consecutive words in the evidence to synthesis the final answer. To encode the context and query, our model applies biDAF (Seo et al. 2016) to acquire a query aware context representation, which is shared by the following sentence ranking and sequence tagging modules. Next, we build a multi-layer perceptron to conduct a match score corresponding to each sentence in the context. The sentence with the highest score will be taken as the evidence. Finally, to extract the specific span of words, we pass the evidence's query aware context representation into a biLSTM-CRF model to tag the answer sequence.

Demonstration and Revision

If a single abstract is input to our system, answers provided by our model will be shown directly after each question in the user input website, as is shown in Fig 2a. If our system is dealing with multiple papers, the result page contains listed abstracts as well as question-answer pairs according to each abstract, which is shown in Fig 2b. Users can revise the answers that are not good enough or answers that do not make sense. Our system will collect revised answers and active learning will be applied for further improvement of our model.

Case Study

ACADEMIC READER is especially useful when dealing with a large amount of papers, e.g., to read and comprehend all the newest papers in arXiv, or all papers accepted in one conference. We present a demonstration of ACADEMIC READER on all papers accepted in KDD 2018[2]. A part of the demonstration is shown in Fig 3. Two essential questions are asked for each abstract: *"What problem does this paper study?"* and *"What method/model/framework/etc. does this paper propose?"*. The content of each paper is summarized into two brief sentences concerning topic and method, i.e., the answers provided by ACADEMIC READER for the two

[2]http://bit.ly/AcademicReaderKDD

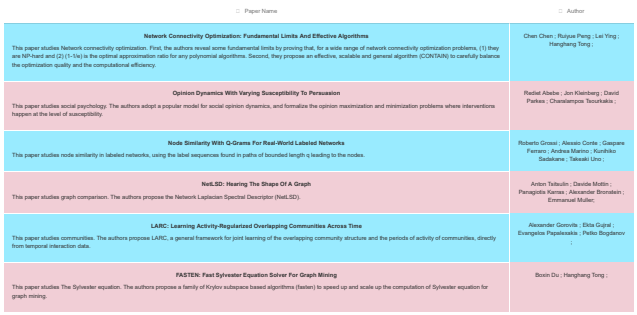


Figure 3: A case study of our system's performance on papers accepted in KDD 2018.

questions. This demonstration enables researchers to browse through all papers accepted in KDD, get the most important information in each paper, and filter papers according to their own research interests.

Conclusion and Future Work

We propose a question answering system called ACADEMIC READER, which massively provides answers to questions concerning the specific academic literature. It assists researchers in browsing through, filtering and sorting papers on their demands. As future work, we plan to expand our paper reading task from the domain of computer science to other domains like mathematics, physics, etc. Moreover, we will optimize our model by integrating the external knowledge in certain fields with context entities.

References

Hermann, K. M.; Kociský, T.; Grefenstette, E.; Espeholt, L.; Kay, W.; Suleyman, M.; and Blunsom, P. 2015. Teaching machines to read and comprehend. *CoRR* abs/1506.03340.

Hirschman, L.; Light, M.; Breck, E.; and Burger, J. D. 1999. Deep read: A reading comprehension system. In *ACL*.

Nguyen, T.; Rosenberg, M.; Song, X.; Gao, J.; Tiwary, S.; Majumder, R.; and Deng, L. 2016. Ms marco: A human generated machine reading comprehension dataset. *CoRR* abs/1611.09268.

Rajpurkar, P.; Zhang, J.; Lopyrev, K.; and Liang, P. 2016. Squad: 100, 000+ questions for machine comprehension of text. *CoRR* abs/1606.05250.

Rajpurkar, P.; Jia, R.; and Liang, P. 2018. Know what you don't know: Unanswerable questions for squad. In *ACL*.

Seo, M. J.; Kembhavi, A.; Farhadi, A.; and Hajishirzi, H. 2016. Bidirectional attention flow for machine comprehension. *CoRR* abs/1611.01603.

Wang, W.; Yang, N.; Wei, F.; Chang, B.; and Zhou, M. 2017. Gated self-matching networks for reading comprehension and question answering. In *ACL*.

Xiong, C.; Zhong, V.; and Socher, R. 2016. Dynamic coattention networks for question answering. *CoRR* abs/1611.01604.

Yu, A. W.; Dohan, D.; Luong, M.-T.; Zhao, R.; Chen, K.; Norouzi, M.; and Le, Q. V. 2018. Qanet: Combining local convolution with global self-attention for reading comprehension. *CoRR* abs/1804.09541.