Augmenting ebooks with Recommended Questions using Contrastive fine-tuned T5

Shobhan Kumar
IIIT Dharwad
Karnataka, India
shobhank9@gmail.com

Arun Chauhan
IIIT Dharwad
Karnataka, India
aruntakhur@gmail.com

Pavan Kumar C
IIIT Dharwad
Karnataka, India
pavan@iiitdwd.ac.in

Abstract

The recent advances in Artificial Intelligence (AI) has made generation of questions from natural language text possible, this approach completely excludes human in the loop, while generating the appropriate questions which improves the students learning engagement. The ever growing rate of educational content renders it increasingly difficult to manually generate sufficient practice or quiz questions to accompany it. Reading comprehension can be improved by asking the right questions. In this work a transformer based question generation model specifically made for autonomously producing quiz questions from educational information, such as ebooks is introduced. This work proposes an contrastive training approach for "Text-to-Text Transfer Transformer" (T5) model where the model (T5-eQG) creates the summarised text for the input document and then automatically generates the questions. Our model shows promising results over earlier Neural Network based and rules based models for question generating task on benchmark datasets and NCERT ebooks.

1 Introduction

Textbooks are a primary source of information for students (Kumar and Chauhan, 2022). Besides this students tends to study ebooks, lecture notes, and MOOCs for further knowledge acquisition (Herrera et al., 2018; Kumar and Chauhan, 2020). With these reading materials students can only partially understand the material presented, and it does not makes their learning effective (Ebersbach et al., 2020). Asking questions about the reading content and evaluate the answer is a intuitive way of promoting learning (Xu et al., 2021; Ruan et al., 2019). This motivates us to look into ways to generate educational questions for ebook content to aid the students learning.

It is difficult to define the specific process for asking insightful educational questions about texts, which entails doing more than just writing fluid, natural-sounding texts. It is bit hard to generate relevant educational questions on text (Horbach et al., 2020). Typically, it involves gathering important instructional facts and turning them into questions. Few attempts have been made in the recent years to complete this task by using statistical and Neural Network based algorithms to choose crucial passages (Chen et al., 2019; Du et al., 2017) and concepts and produce insightful questions (Dong et al., 2019; Steuer et al., 2020). Education question generation on text, however, has not received much attention.

The proposed educational Question Generation (eQG) model presents a set of questions for each chapter of the ebook. Teachers might use these questions for self-study, before discussing the subject in class, which would helps them for deep knowledge transfer to the students (Kumar and Chauhan, 2019; Xu et al., 2021). We assess our methodology on QA dataset (HotPot QA(Yang et al., 2018), FairytaleQA[1]) as well as PRML (Bishop, 2006) and NCERT[2] eBooks. Table 1 presents sample results of our eQG model for the input text (refer Figure 1). The learner can use eQGs for self-assessment questions to gauge their conceptual understanding.

The contributions of this work are two fold:

- Text summarizer: We fine-tune the Text-to-Text transformer (T5) to extract the informative sentences that are most likely for educators to design questions for the original input.
- Contrastive training for T5-eQG: We fine-tune the T5 transformer on positive and negative training samples. A contrastive loss is added between the positive and negative training feature pairs during the fine-tuning process. It

[1]https://github.com/uci-soe/FairytaleQAData
[2]https://ncert.nic.in/textbook.php

Proceedings of the 19th International Conference on Natural Language Processing (ICON), pages 109 - 115
December 15-18, 2022 ©2022 Association for Computational Linguistics

helps in generating more complex questions on input document.

2 Background

Natural language creation has primarily evolved through statistical learning in recent years. To create manuscripts that resemble human writing, the models imitate linguistic conventions. In recent years, the natural language processing communities have shown a great deal of interest in the question generation (QG) task(Wang et al., 2017; Lyu et al., 2021), which creates a natural question corresponding to the supplied text or answer phase. The syntactic cues have been used in the rule based model to create Questions (De Kuthy et al., 2020). A back translation tool was paired with a syntactic question generator to eliminate grammatical errors and increase robustness(Dhole and Manning, 2020). Declarative sentences were transformed into natural questions by the emergence of sequence-to-sequence models (Radford et al., 2019). Applying pre-trained transformers or various optimization objectives (Qi et al., 2020) led to further advancements. Previous research has explored the importance of QG model in teaching learning process(Kurdi et al., 2020).

For QA and question generation, NarrativeQA (Kočiský et al., 2018) aims to incorporate important information from many places inside a paragraph. Similar to this, the MS MARCO (Nguyen et al., 2016) dataset combines many sources of responses to search queries. The employment of a reinforcement learning agent to align questions from various documents is proposed as a contrastive strategy, where supervised model is trained to produce questions on a text (Cho et al., 2021). To achieve good performance, questions with summaries and reports were generated using a rule-based methodology (Lyu et al., 2021). The solutions discussed above typically don't take the educational component into account and could not be effective in the real world edu QG task. Our research focuses on the generating question on e-book content, in this work we use FairytaleQA dataset (Xu et al., 2022). For each paragraph in FairytaleQA, experts typically create a different style of question. We propose that context is a key factor in determining the kinds of questions that ought to be made while reading e-books.

3 Methodology

Figure 2 depicts the overall architecture of our eQG system, which consists of two modules: i) Text summarizer ii) Question-generation(QG). We first create summaries of type s with the input paragraph d, and then generate the questions q on summarized text. The generated questions are said to be relevant if the question q_i can be answered with the paragraph d_i and this is formulated as maximizing the conditional probability $p(q|d)$:

$$q = argmax(p(q \mid d)) = argmax \prod_{i=1}^{L} p(w_i \mid d, q_{i'}) \tag{1}$$

where w_i is the i^{th} token of the generated question q, and $q_{i'}$ denotes the previous decoded tokens, i.e., $q_1, \ldots, q_{i\text{-}1}$.

T5 - Abstractive summarizer: In this work, we examined the text summarizer and QG as a task of text-to-text transformation. So, we first train a T5[3] summarising model to produce the abstract summary of the input text.

Edu Question Generation: Once the model produce the abstract summary of the input text, next step is to generate an educational question out of it. We train a T5-QG model directly on top of the summary using the annotated questions, since T5-summary model already has knowledge on rich informative text.

When the model is fine tuned with tiny dataset, fine-tuning process with QA task loss is generally insufficient to achieve satisfactory performance. To address this issue, we generated the negative samples for each document d_i and fine-tune using both the data samples.

Most of the existing QG model suffers from the exposure bias problem. Therefore, we created a negative sample and trained an end-to-end eQG model by introducing contrastive loss function. We trained two variants of T5_QG model. At first we fine-tune the T5_QG model by minimizing the cross-entropy loss. In the next step, model is trained on augmented data (both ground truth question and generated negative samples) with the contrastive loss and cross-entropy loss

$$T_{loss} = Q_{\text{task}} + Q_{\text{c\_loss}} \tag{2}$$

where T_{loss} is the total loss, Q_{task} and $Q_{\text{c\_loss}}$ are the QG task loss and the contrastive loss, respectively,

[3]https://huggingface.co/docs/transformers/model_doc/t5

1.2.4 The Gaussian distribution

We shall devote the whole of Chapter 2 to a study of various probability distributions and their key properties. It is convenient, however, to introduce here one of the most important probability distributions for continuous variables, called the *normal* or *Gaussian* distribution. We shall make extensive use of this distribution in the remainder of this chapter and indeed throughout much of the book.

Figure 1: The sample input text from PRML ebook (Section 1.2) by C.Bishop(Bishop, 2006)

Table 1: The generated questions using our eQG model (input text-PRML ebook).

Text: Refer Figure 1	
Abstract summary: The Gaussian distribution It is convenient, however, to introduce here one of the most important probability distributions for continuous variables	
Generated top k questions using eQG model (k=3)	
Q1	What is Gaussian distribution?
Q2	What is the most important probability distribution for continuous variables?
Q3	Where we use normal distribution?

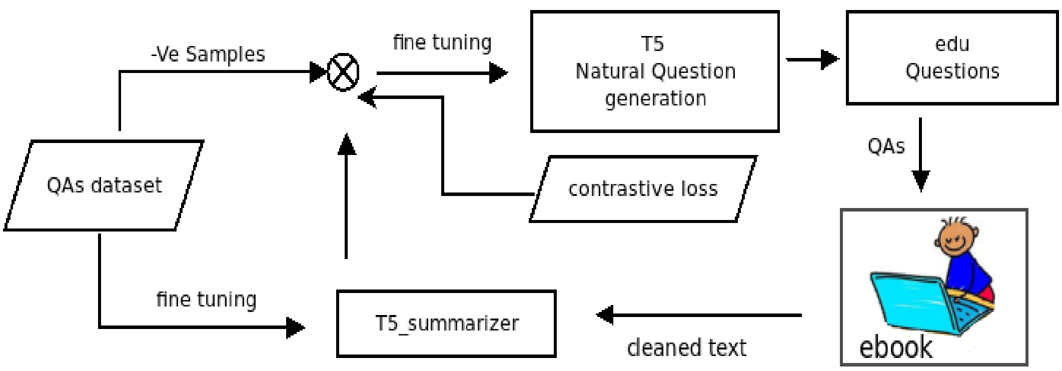

Figure 2: Overall architecture of our eQG model.

$$Q_{\text{c\_loss}} = q_l * D_w^2 + (1-q_l) * max(m-d_w, 0)^2 \quad (3)$$

where q_l is the ground-truth labels from our dataset, d_w is the Euclidean distance and m is the margin used for the contrastive loss function.

4 Experimental Results

In this work we trained T5 (Text-to-Text Transfer Transformer) base model from hugging face transformers[3]. An input sequence and a corresponding target sequence are required for every training run of T5. The model receives the input sequence via input ids. The target sequence is provided to the decoder using the decoder input ids after being prepended by a start-sequence token and moved to the right. The EOS token is subsequently attached to the target sequence in teacher-forcing fashion, which correlates to the labels. The start-sequence token here is the PAD token.

Data set: We used FairytaleQA[1] and HotpotQA data, the FairytaleQA has 10,580 QA-pairs, which were drawn from 278 different novels. The HotpotQA(Yang et al., 2018) has ≈100K QA pairings on Wikipedia articles. For FairyQA we divide the data into 8.5K/1K/1K, and for HotPot QA 84,512, 6K and 6K samples as train,validation and test data.

For fine tuning the T-5[3] model we used the AdaFactor optimizer and a maximum sequence length is set to 512, model is trained for 4 epochs. We follow the grid-search approach for choosing the best set of training parameters (learning-rate:{{2,3,5}e-3,4}} and batch size: {8,16,32,64}, warm-up ratio: {0.1}). During the experiment, we found that a mini-batch size of 32 (learning rate: 3e-3) produces acceptable results.

Table 1 and Table 2 highlight the generated questions for text from PRML (Bishop, 2006) and NCERT[2] CCT ebook. The generated questions makes the students learning more effective, and assist them in improving their conceptual understanding ability.

We validated the quality of questions generated on three experimental configuration. QG model trained i)only on HotpotQA, (ii) only on FairytaleQA, iii) on both FairytaleQA and HotpotQA. The third setting shows a significant improvement over the preceding setups, so this was chosen as our final QG model for further comparison to the earlier work. Results are shown in Table 3. We see that the model optimised on FairytaleQA alone shows significant improvement over the model trained on both the dataset. This is due to the disparities in domain and distribution between the two datasets. The third settings shows a decent results during the state-of-the-art model comparison.

Table 4 provides the comparison results of our model with state-of-the-art models. Our model achieves a comparable BLEU4, Rouge scores with the cutting-edge QG model in HotpotQA without using the answer information or any external linguistic knowledge. This illustrates the effectiveness of contrastive training of language model for the QG task.

5 Conclusion and Future scope

The use of AI is constantly evolving in diverse applications. This study investigates the potential advantages of a natural language processing approach for education. This research presented an education question generating (eQG) approach that augments the ebook content with generated edu-questions to provide students with an effective learning platform. Through experiments, we assessed the model's performance on a question generation task both before and after contrastive training. We discovered that a contrastive trained model can produce more pertinent questions on the input text and can comprehend key concepts more effectively. Experiments on QA dataset, PRML(Bishop, 2006) and NCERT [2] ebook shows that our model succeeds to produces complex questions at scale.

The possible future direction could be

- Design a context-aware QG model, where the generation of a new text is conditioned on previous generations as well as the ebook contents.
- Conduct a human evaluation to validate the appropriateness of the generated questions on ebook content.

Table 2: The abstract summary and generated questions using our eQG model(top k questions) for the input text NCERT[2] CCT ebook.

<table>
<tr><td colspan="2">Input Text:Use of innovative technologies like Silicon-On-Insulator (SOI), Complementary Metal-Oxide-Semiconductor (CMOS), capacitor -less memory, Micro-Optic-Electro-Mechanical-System (MOEMS) III-V compound materials-on-insulator and others have improved the performance and also reduced the size of consumer electronic devices...</td></tr>
<tr><td colspan="2">Abstract summary: Innovative technologies such as Silicon-On-Insulator (SOI), Complementary Metal-Oxide-Semiconductor (CMOS), capacitor-less memory, Micro-Optic...</td></tr>
<tr><td colspan="2">Generated Top k (k=3) questions</td></tr>
<tr><td>Q1</td><td>What are the benefits of using silicon-on-insulator (SOI)?</td></tr>
<tr><td>Q2</td><td>How is graphene expected to improve the processing speed of computers?</td></tr>
<tr><td>Q3</td><td>What is the advantage of using III-V compound materials-on-insulator in consumer electronic devices?</td></tr>
</table>

Table 3: Comparison of our contrastive eQG models with various experimental settings.

QG model	Evaluation metric:Rouge-L	
	Validation_data	Test_data
$T5_{base}$_HotpotQA	0.423	0.441
$T5_{base}$_FairytaleQA	0.512	0.526
$T5_{base}$_HotpotQA_FairytaleQA	0.507	0.518

Table 4: The comparison results of our model with prior work for QG task on HotPot dataset.

QG model	Evaluation metrics			
	BLEU-1	BLEU-4	Meteor	Rouge-L
RL_QG (Xie et al., 2020)	37.97	15.41	19.61	35.12
Deep_QG(Pan et al., 2020)	40.55	15.53	20.15	36.94
T5QG	40.96	17.54	19.21	42.36
Contrastive_T5QG	42.04	19.11	20.07	48.50

References

Christopher M. Bishop. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

Guanliang Chen, Jie Yang, and Dragan Gasevic. 2019. A comparative study on question-worthy sentence selection strategies for educational question generation. In *Artificial Intelligence in Education*, pages 59–70, Cham. Springer International Publishing.

Woon Sang Cho, Yizhe Zhang, Sudha Rao, Asli Celikyilmaz, Chenyan Xiong, Jianfeng Gao, Mengdi Wang, and Bill Dolan. 2021. Contrastive multi-document qg. In *Proc.16th European Conference, Chapter of the ACL*, pages 12–30.

Kordula De Kuthy, Madeeswaran Kannan, Haemanth Santhi Ponnusamy, and Detmar Meurers. 2020. Towards automatically generating questions under discussion to link information and discourse structure. In *Proc.28th International Conference on Computational Linguistics*, pages 5786–5798.

Kaustubh D. Dhole and Christopher D. Manning. 2020. Syn-qg: Syntactic and shallow semantic rules for question generation. *ArXiv*, abs/2004.08694.

Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. *Unified Language Model Pre-Training for Natural Language Understanding and Generation*. Curran Associates Inc., Red Hook, NY, USA.

Xinya Du, Junru Shao, and Claire Cardie. 2017. Learning to ask: Neural question generation for reading comprehension. *CoRR*, abs/1705.00106.

Mirjam Ebersbach, Maike Feierabend, and Katharina Barzagar B. Nazari. 2020. Comparing the effects of generating questions, testing, and restudying on students' long-term recall in university learning. *Applied Cognitive Psychology*, 34(3):724–736.

Jose Miguel Herrera, Barbara Poblete, and Denis Parra. 2018. Learning to leverage microblog information for qa retrieval. In *ECIR*, volume 10772 of *Lecture Notes in Computer Science*, pages 507–520. Springer.

Andrea Horbach, Itziar Aldabe, Marie Bexte, Oier Lopez de Lacalle, and Montse Maritxalar. 2020. Linguistic appropriateness and pedagogic usefulness of reading comprehension questions. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1753–1762, Marseille, France. European Language Resources Association.

Tomáš Kočiský, Jonathan Schwarz, Phil Blunsom, Chris Dyer, Karl Moritz Hermann, Gábor Melis, and Edward Grefenstette. 2018. The NarrativeQA reading comprehension challenge. *Transactions of the Association for Computational Linguistics*, 6:317–328.

Shobhan Kumar and Arun Chauhan. 2019. Enriching textbooks by question-answers using cqa. In *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*, pages 707–714.

Shobhan Kumar and Arun Chauhan. 2020. Recommending question-answers for enriching textbooks. In *Big Data Analytics*, pages 308–328, Cham. Springer International Publishing.

Shobhan Kumar and Arun Chauhan. 2022. Augmenting textbooks with cqa question-answers and annotated youtube videos to increase its relevance. *Neural Process Lett.*

Ghader Kurdi, Jared Leo, Bijan Parsia, Uli Sattler, and Salam Al-Emari. 2020. A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 30:121–204.

Chenyang Lyu, Lifeng Shang, Yvette Graham, Jennifer Foster, Xin Jiang, and Qun Liu. 2021. Improving unsupervised question answering via summarization-informed question generation. In *Proc.EMNLP*, pages 4134–4148.

Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. 2016. MS MARCO: A human generated machine reading comprehension dataset. *CoRR*, abs/1611.09268.

Liangming Pan, Yuxi Xie, Yansong Feng, Tat-Seng Chua, and Min-Yen Kan. 2020. Semantic graphs for generating deep questions. In *Proc. 58th Annual Meeting of the ACL*, pages 1463–1475, Online.

Weizhen Qi, Yu Yan, Yeyun Gong, Dayiheng Liu, Nan Duan, Jiusheng Chen, Ruofei Zhang, and Ming Zhou. 2020. ProphetNet: Predicting future n-gram for sequence-to-SequencePre-training. In *EMNLP 2020*, pages 2401–2410.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners.

Sherry Ruan, Liwei Jiang, Justin Xu, Bryce Joe-Kun Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth L. Murnane, Emma Brunskill, and James A. Landay. 2019. Quizbot: A dialogue-based adaptive learning system for factual knowledge. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–13, New York, NY, USA. Association for Computing Machinery.

Tim Steuer, Anna Filighera, and Christoph Rensing. 2020. Remember the facts? investigating answer-aware neural question generation for text comprehension. In *Artificial Intelligence in Education*, pages 512–523, Cham. Springer International Publishing.

Tong Wang, Xingdi (Eric) Yuan, and Adam Trischler. 2017. A joint model for question answering and question generation. In *Learning to generate natural language workshop, ICML 2017*.

Yuxi Xie, Liangming Pan, Dongzhe Wang, Min-Yen Kan, and Yansong Feng. 2020. Exploring question-specific rewards for generating deep questions. In *Proc. 28th International Conference on Computational Linguistics*, pages 2534–2546.

Ying Xu, Dakuo Wang, Penelope Collins, Hyelim Lee, and Mark Warschauer. 2021. Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. *Computers and Education*.

Ying Xu, Dakuo Wang, Mo Yu, Daniel Ritchie, Bingsheng Yao, Tongshuang Wu, Zheng Zhang, Toby Jia-Jun Li, Nora Bradford, Branda Sun, Tran Bao Hoang, Yisi Sang, Yufang Hou, Xiaojuan Ma, Diyi Yang, Nanyun Peng, Zhou Yu, and Mark Warschauer. 2022. Fantastic questions and where to find them: FairytaleQA – an authentic dataset for narrative comprehension. Association for Computational Linguistics.

Zhilin Yang, Peng, Saizheng Zhang, Yoshua Bengio, William W. Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. 2018. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. *CoRR*, abs/1809.09600.