IJARSCT
ISSN (Online) 2581-9429
IJARSCT
Impact Factor: 7.301
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 3, Issue 8, May 2023

Automatic Generation of MCQs Using Transformer Mode

Saloni Ingle[1], Chetan Naik[2], Soham Kamat[3]
Final Year Student(B.E) , Department of Computer Science and Engineering[1,2,3]
Vivekanand Education Society's Institute of Technology, Mumbai, India

Abstract: *Multiple Choice questions is the go to assessment technique used for competitive entrance exams, company's aptitude screening process and for the assessment exams of various institutions and universities. The rise of the online paradigm due to the pandemic has also changed the way universities and institutions conduct their assessment exams* Hence, MCQs play an important role in assessment of skill and knowledge in different domains and situations* However, it becomes very difficult for a human to generate large amount of MCQs along with good quality distractors in limited time. This problem can be solved using state of the art natural language processing and deep learning techniques. Our work is an attempt to generate multiple choice based questions which can be used for assessment in an exam setting. This paper proposes a system which uses T5 transformer for question generation and other state of the art deep learning technique for distractor generation generate multiple choice questions which resembles the human way of questioning.*

Keywords: Natural Language Processing, MCQs, T5 transformer, Deep learning, Distractors, Question generation

I. INTRODUCTION

Assessment is an important part of the learning process since it helps an individual to identify the areas of improvement and MCQs have emerged as an important tool for assessment [1]. Multiple Choice Questions (MCQs) are the set of questions that provides multiple options, usually four out of which one is correct and the other three are wrong. These wrong options are called distractors and the correct option is the key [2]. The sentence from which one question is generated forms the stem of the multiple-choice question. The manual generation of Multiple Choice Questions is a tedious task and takes up a majority of the time of an individual [3]. It's even more difficult to form questions that are tricky to solve since it needs to be formed with multiple layers of contextual depth. Both educators and students will be benefitted from this system since it will help them in the assessment process by eliminating human error and saving precious time.

The new online paradigm has changed the exam patterns from subjective to objective-based where MCQs play an important role in the assessment. Even competitive entrance exams tend to use MCQs as a part of their assessment. It becomes difficult for examiners to generate questions in short span of time. So there is a growing need for a system that can create questions with ease and less amount of time and requires less human effort. Along with this, parents of young students find it difficult to assess the knowledge of their wards on a particular topic since they don't have the time to form questions and find their answers. If there is an interface that presents the questions to the student and assesses their knowledge on a particular topic then it will be easy for both the parents and their children.

This paper is an attempt to solve this problem of manual generation of MCQs by presenting a system that generates questions from the provided text along with the key and the distractors. This system will make use of the recent advancements in deep learning techniques for the task of natural language processing. It also ensures that different questions are generated on every run of the system.

II. RELATED WORK

The recent advancements in the field of deep learning like RNN models and Transformer architectures have changed the way we solve natural language processing problems. The previous works on the task of Question Generation use

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, May 2023

traditional natural language processing techniques of writing rules and finding patterns [2]. Most of the question generation work includes a common workflow which includes - extracting phrases from the context and finding context-based options which are incorrect i.e. the distractors. The distractor generation task is usually trained using the RACE dataset [13]. A neural keyphrase extraction and detection are used to select the sequence of words that would be picked by humans [4]. A sequence to sequence approach with 2 RNN models in which the document and the answer act as sources of information are also used for question generation [5]. Models that include reinforcement learning generate the questions by including multiple documents in the corpus [6]. The introduction of the attention mechanism based transformer model made huge improvements in natural language processing tasks with a BLEU score of 41.8 in machine translation [16]. Transformer based models like BERT gives a significant performance at both sentence and paragraph level document [7]. One of the recent breakthroughs in the field of text to text generation is the T5 transformer which uses the idea of transfer learning [17]. SQuAD [8], HotpotQA [9] are some of the common datasets that are used for the training of this task. The evaluation measures used are BLEU [10], ROUGE [11], and METEOR [12]. The BLEU score of basic seq2seq model is 10.5, focusing on answer technique based on HotPotQA is 28.5 whereas pretrained models is 34. Hence pretrained models perform better compared to other basic seq2seq, external content QG, focusing on answer techniques, etc.

There are some online question generation services like QuestGen [14], Quillionz [14] which come with words limitation or paid licences. We aim to offer an open-source and free system for question generation by uploading a pdf or a text.

III. MODEL IMPLEMENTED

The model implemented in the system is the T5 transformer model. Standard RNN model [19] has a mechanism of "remembering" the previous inputs and producing an output based on all of the inputs and sequence to sequence RNN model has two parts - encoder and decoder. But this standard RNN sequence to sequence model is too slow to train. The 2017 paper [16] titled "Attention is all you need" proposes a transformer model which makes use of parallel computation power of GPUs and proposes the attention mechanism. Since the inception of the paper, many transformer models like GPT [21], BERT [20] etc have been created. The 2020 paper [17] titled " Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer" explores the idea of transfer learning and unifies the different transformer architectures to perform a downstream task.

T5 transformer defines a text-to-text framework (input and output both sequences are text) that uses the same hyperparameters, same loss function and even the same model(same checkpoints and layers) for all the NLP tasks. The inputs are modelled so that the model shall recognize a task, and the output is simply the "text" version of the expected outcome.

IV. PROPOSED SYSTEM

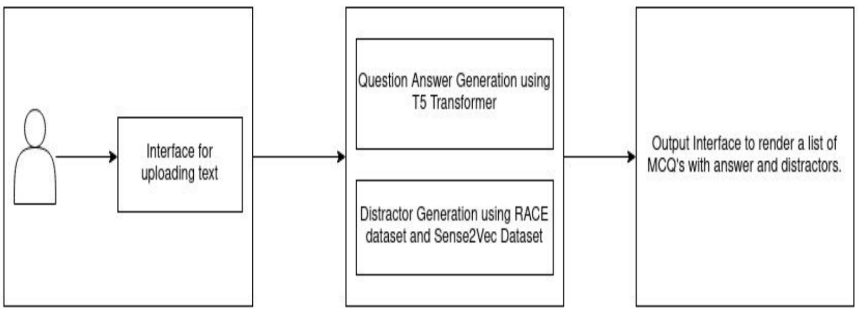

Fig.1 Modular Diagram

The proposed system takes in the input in the form of text or a pdf document with the number of questions to be formed and produces required number of questions along with it's correct answer and distractor as the output. The entire system can be divided into three modules as shown in the modular diagram (Fig 1). The first module is the user interface for uploading the text and the number of questions. Then comes the question and distractor generation module

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, May 2023

which uses the T5 transformer model trained on SquAD dataset to generate questions and T5 transformer model trained on RACE dataset and Sense2Vec Dataset for distractor generation. The third module is the output interface where the user can see the questions generated and use them according to his/her needs.

A. Input and Output Interface module

Users can upload a text or a pdf document along with the number of questions the user expects to be generated(Fig.2). This text or pdf, after submitting makes a REST API call to the Question Generation module. The API returns question generated along with the answer and distractors. Then the output interface displays these questions with the correct answer highlighted in green and wrong options in red(Fig.3). This frontend website interface is built using Next.js, a React Framework.

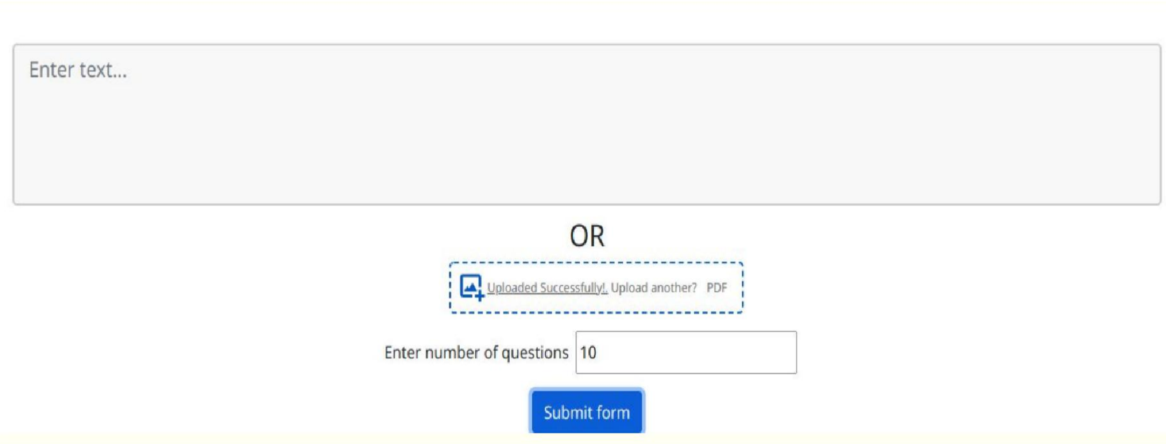

Fig. 2. Input Interface

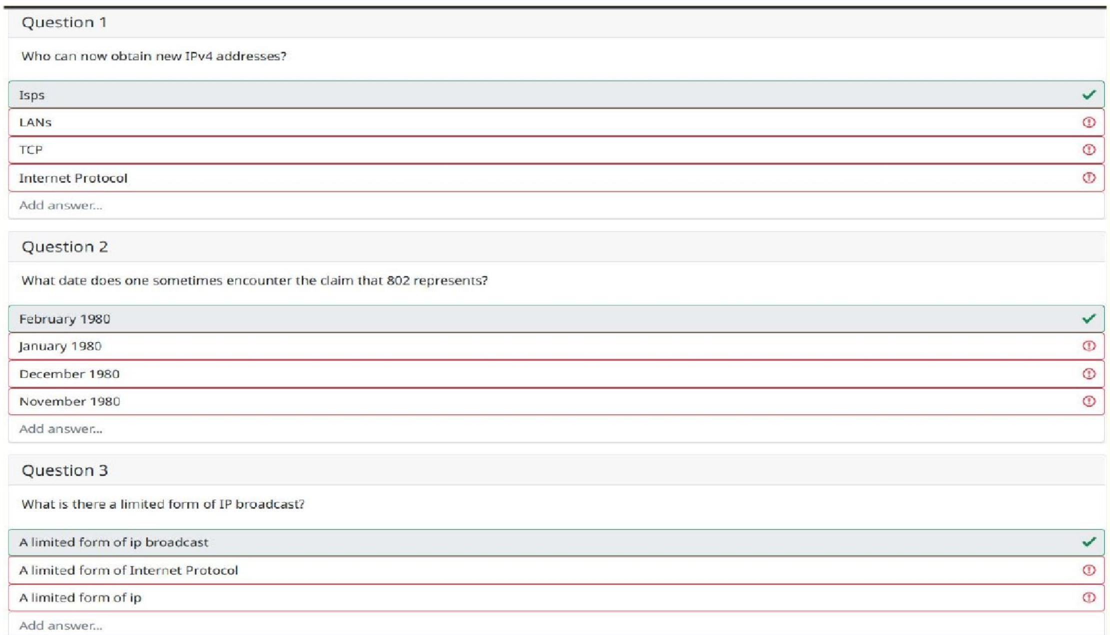

Fig. 3. Output Interface

B. Question Answers Generation Module

This module is inspired by the works of Leaf Question Generation [18]. We have built a similar model which is trained on SQuAD1.1 [8] dataset which includes 1,00,000 question pairs. We have used the T5 transformer model provided by huggingface that is tuned for the downstream task of question generation The training model accepts text and outputs

DOI: 10.48175/IJARSCT-10261

question and answer. The answer part is replaced with a mask token with 50% probability so that question can be generated without providing the answer. The SQuAD dataset gives F1 score of 51.2 with the trained model.

The distractors are generated using the same T5 model but trained on the RACE dataset. This model takes the question and the answer as input and outputs three distractors . The three distractors had BLEU score of 36.1, 35.2 and 33.5 respectively. If the T5 transformer model fails to generate the distractors then sense2vec [22] which helps in generating distractors that are similar to the answer.

The whole system can be summarized by the block diagram, as shown in Fig.4.

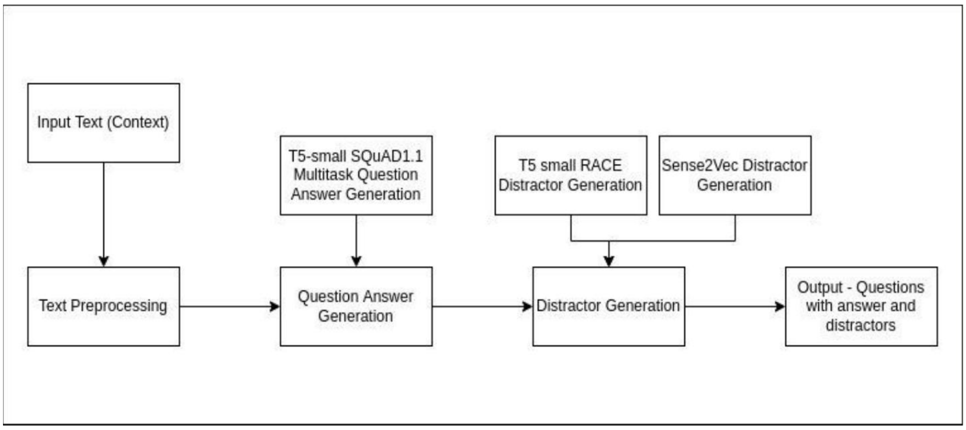

Fig. 4. System Block Diagram

Initially, the input provided to the Question Answer generation module is pre-processed i.e. the complete text is cleaned by removing irrelevant characters like punctuation marks, newline, etc. Further the text is processed using sentence tokenization. Then, from the pool of sentences, the required number of sentence is randomly picked. Question and answer are generated on each of these sentence using the T5 transformer.

The generated question and answer is passed to the distractor generator model which is also based on the T5 transformer. This model outputs the three distractors based on the question and it's respective answer.

V. CONCLUSION AND FUTURE WORK

The system is capable of generating MCQs that resemble the human way and helps educators and self-learners in the process of assessment. The system is an attempt to fill the gap between learning and assessment without much human intervention. Since our proposed system uses a pre-trained model which is continuously improved, the accuracy of the system could increase in the future with an increase in the performance of the model.

In future work, the system can incorporate generating questions based on given difficulty for which the Bloom's taxonomy can be used [23]. Question generation is a very nascent research topic in natural language processing in contrast to Question Answering tasks hence relevant dataset for this specific task can be generated. The quality of questions can be increased by incorporating multiple layers of context. This will also help in forming questions which are based on understanding and not just factual questions.

REFERENCES

[1]. Henry L. Roediger III, Adam L. Putnam, and Megan A. Smith. Chapter one – ten benefits of testing and their applications to educational practice. In Psychology of Learning and Motivation, volume 55, pages 1–36. Academic Press, 2011.

[2]. Rao, Dhawaleswar & Saha, Sujan Kumar. (2018). Automatic Multiple Choice Question Generation From Text: A Survey. IEEE Transactions on Learning Technologies. PP. 1-1. 10.1109/TLT.2018.2889100.

[3]. Yuni Susanti, Takenobu Tokunaga, Hitoshi Nishikawa, and Hiroyuki Obari. Evaluation of automatically generated english vocabulary questions. Research and practice in technology enhanced learning, 12(1):1–21, 2017.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 8, May 2023

[4]. Sandeep Subramanian, Tong Wang, Xingdi Yuan, Saizheng Zhang, Yoshua Bengio, Adam Trischler - "Neural Models for Key Phrase Extraction and Question Generation", 2018 Proceedings of the Workshop on Machine Reading for Question Answering

[5]. Yuan, Xingdi, et al. "Machine comprehension by text-to-text neural question generation." arXiv preprint arXiv:1705.02012 (2017).

[6]. Wang, Liuyin & Xu, Zihan & Lin, Zibo & Zheng, Haitao & Shen, Ying. (2020). Answer-driven Deep Question Generation based on Reinforcement Learning. 5159-5170. 10.18653/v1/2020.coling-main.452.

[7]. Chan, Ying-Hong & Fan, Yao-Chung. (2019). A Recurrent BERT-based Model for Question Generation. 154-162. 10.18653/v1/D19-5821.

[8]. Rajpurkar, Pranav, et al. "Squad: 100,000+ questions for machine comprehension of text." arXiv preprint arXiv:1606.05250 (2016).

[9]. Yang, Zhilin, et al. "HotpotQA: A dataset for diverse, explainable multi-hop question answering." arXiv preprint arXiv:1809.09600 (2018).

[10]. Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, ACL '02, pages 311–318, Philadelphia, Pennsylvania, USA, 2002. Association for Computational Linguistics.

[11]. Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In Proceedigs of the Workshop on Text Summarization Branches Out, pages 74–81, Barcelona, Spain, 2004. Association for Computational Linguistics.

[12]. Alon Lavie and Abhaya Agarwal. METEOR: An automatic metric for MT evaluation with high levels of correlation with human judgments. In Proceedings of the Second Workshop on Statistical Machine Translation, WMT '07, pages 228–231, Prague, Czech Republic, 2007. Association for Computational Linguistics

[13]. Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. RACE: Large-scale ReAding comprehension dataset from examinations. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP '17, pages 785–794, Copenhagen, Denmark, 2017. Association for Computational Linguistics.

[14]. Questgen. Questgen: AI powered question generator. http://questgen.ai/. Accessed: 2022-03-25.

[15]. Quillionz. Quillionz - world's first AI-powered question generator. https://www.quillionz.com/. Accessed: 2022-03-25.

[16]. Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[17]. Raffel, Colin, et al. "Exploring the limits of transfer learning with a unified text-to-text transformer." arXiv preprint arXiv:1910.10683 (2019).

[18]. Vachev, Kristiyan, et al. "Leaf: Multiple-Choice Question Generation." arXiv preprint arXiv:2201.09012 (2022).

[19]. Sherstinsky, Alex. "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network." Physica D: Nonlinear Phenomena 404 (2020): 132306.

[20]. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[21]. Brown, Tom, et al. "Language models are few-shot learners." Advances in neural information processing systems 33 (2020): 1877-1901.

[22]. Andrew Trask, Phil Michalak, and John Liu. sense2vec - a fast and accurate method for word sense disambiguation in neural word embeddings. ArXiv:1511.06388, 2015.

[23]. LW, Anderson & DR, Krathwohl & PW, Airasian & KA, Cruikshank & Mayer, Richard & PR, Pintrich & Raths, J.& MC, Wittrock. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives