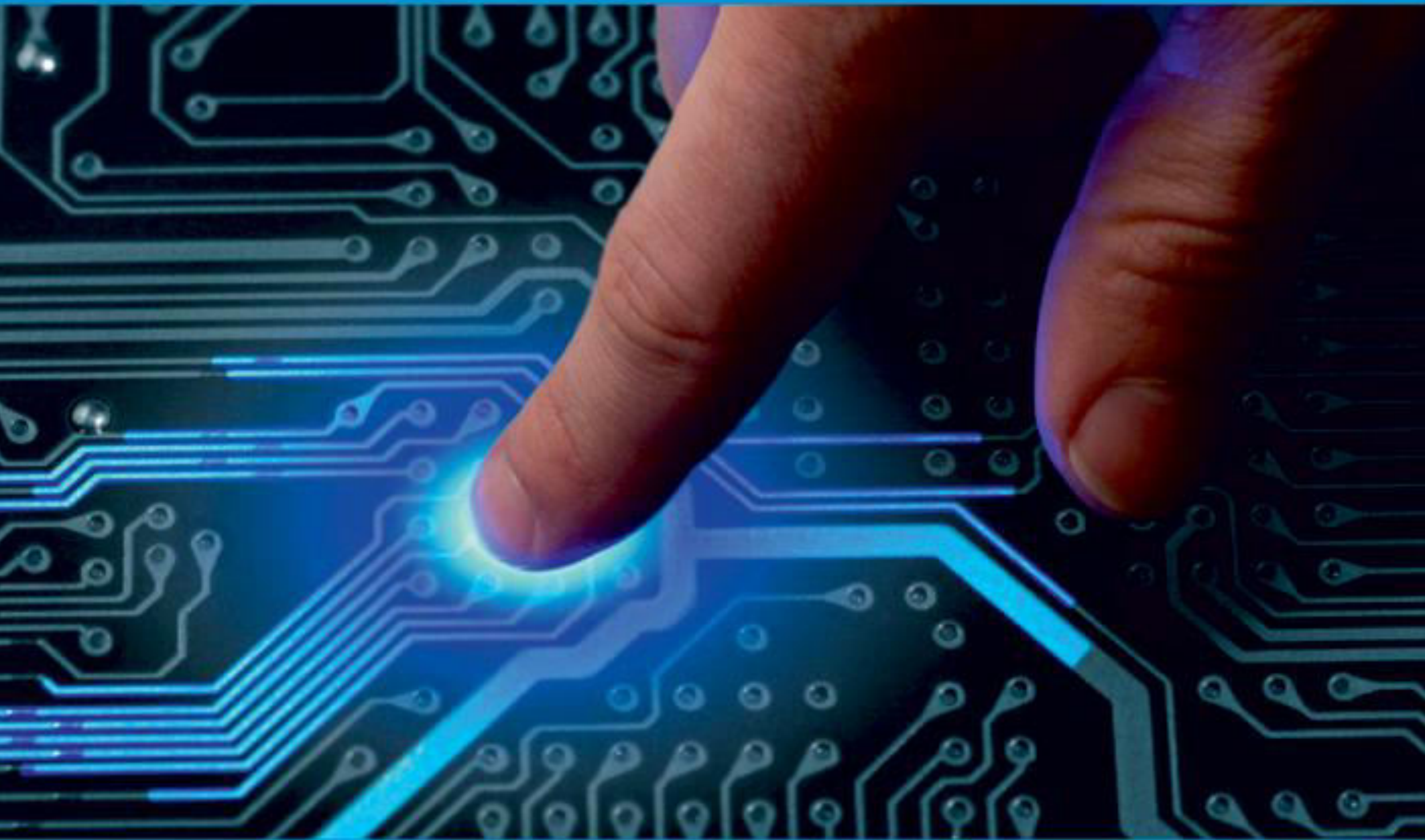
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 2, February 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002011 |

A Survey of work on Automatic Generation of Questions from Text

Chetan Naik[1], Saloni Ingle[1], Soham Kamat[1], Veena Trivedi[2], Pallavi Gangurde[2]

B.E. Student, Department of Computer Engineering, Vivekanand Education Society's Institute of Technology, Chembur, Mumbai, India[1]

Assistant Professor, Department of Computer Engineering, Vivekanand Education Society's Institute of Technology, Chembur, Mumbai, India [2]

ABSTRACT: Automatic question generation specifically MCQ based is a widely appreciated application of natural language processing and deep learning. Since the dawn of the pandemic, the world has shifted towards the online paradigm along with universities and institutions. Hence, examinations and assessments are going through a huge change. Most of the institutions have changed their examination pattern towards objective-based questions. However, the manual generation of objective questions and designing specific options/ distractors is a tedious task. This paper thus reviews the work of various system that automatically generates question and answer with text as input. A comparative study of different relevant systems done in this paper gives an overview of the traditional workflow in solving the automatic QG problem.

KEYWORDS: Question Generation, Deep Learning, Machine Learning, e-learning, Automatic QG.

I. INTRODUCTION

Multiple-choice questions are the type of questions with one correct answer which needs to choose from multiple given options. The incorrect options are called distractors and are closely related to the correct answer. The aim of this survey paper is to find relevant research papers which claim to generate question answers from the text. Question answering has been a very active area of research in the field of natural language processing for a long time. Any major advancement in this field will positively impact the educational assessment paradigm around the world. Multiple Choice question is a specific use case of the broader Question Answer generation problem. The Wikipedia definition of QA is "Question answering (QA) is a computer science discipline within the fields of information retrieval and natural language processing (NLP), which is concerned with building systems that automatically answer questions posed by humans in a natural language" [1].

It's important to improve the progress of learning by removing the gap between assessment and self-learners. Hence, there is a dire need for the generation of questions and answers automatically using modern tools of machine learning and deep learning. At the same time, the pattern of assessment is majorly shifting towards the objective assessment i.e. MCQ based, it is very hard to construct and requires a considerable amount of time for setting numerous questions. As in any education system, the examination is conducted to judge the calibre of the students. Hence, to eliminate the manual effort to generate the questions, a system needs to be established which will help educators automatically generate the multiple-choice questions along with their answers. The above discussion provides two important aspects, one from the student's side and the other from the educator's side where the need to generate questions for assessment is imminent and both the parties will be benefitted from an automatic system to do the work for them. The recent advancements in the field of natural language processing and deep learning can be utilised to create a system that can automatically generate relevant multiple-choice questions. This paper tries to provide an introduction to the work done in this particular area of question-answer generation.

A. Terminology

Some important terminology related to MCQ's [2]:

1. Stem - It is the sentence or the phrase from the passage which acts as the question [2].
2. Key - It is the word that acts as the correct answer to the stem [2].
3. Distractors - These are the words that closely relates to the key but are not the actual answer [2].

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002011 |

B. Common workflow in generating MCQ's

To generate multiple-choice questions, the system uses various NLP techniques. The most common workflow in this area can be explained as follows [2] -

1. Find the summary of the text.
2. Extract keywords from the text.
3. Map sentences to the keyword.
4. Find relevant distractors using the keyword.
5. Replace the keyword with a blank.
6. Generate the MCQ using the stem, keyword and distractor.

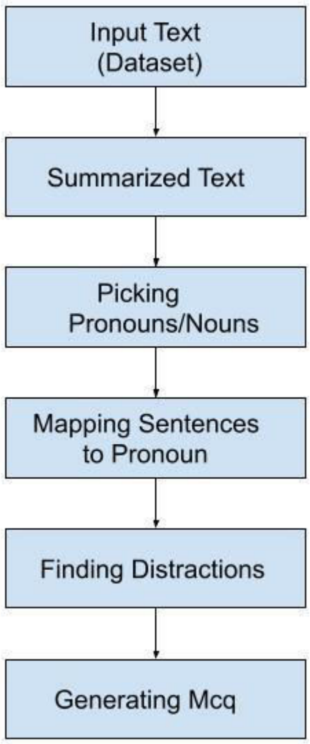


Fig. 1 Architecture of the common workflow

II. REVIEW METHODOLOGY

A systematic review is a summary of studies addressing a clear question, using systematic and explicit methods to identify, select, and critically appraise relevant studies, and to collect and analyze data from them. In order to ensure a systematic review process, this study followed the following steps for conducting a systematic review [2]:

1) Formulate the research problem.
2) Literature Search.
3) Information Gathering.
4) Analyze and integrate the outcomes of research.

III. RELATED WORK

A. *Neural Models for Key Phrase Detection and Question Generation*

In this paper [3], the problem of question generation is solved using a two-stage neural model. The first part is basically predicting the probability that the sequence of the words would be picked by humans using a neural keyphrase extractor on the dataset. Then the next part uses these predicted key phrases for question generation using a sequence to sequence with a copy mechanism. spaCy is used to predict the entities in the document. This is

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002011 |

supported by the fact that entities constitute the largest proportion of answers in the SQuAD dataset [4]. The key phrase extraction part is divided into three parts - Entity tagging baseline, neural entity selection and pointer networks[5].

B. *Machine Comprehension by Text-to-Text Neural Question Generation*
This paper [6] follows the seq2seq approach, which uses 2 RNN that work together to predict the next state sequence using the previous sequence. The first step is to encode the document along with answers and then output questions sequentially using the decoder [7]. Hence the document and the answers act as two sources of information. Rather than using the maximum likelihood of the document and answer tuple, a policy gradient optimization [8] is used, which acts as a slight modification in the standard encoder-decoder technique. The policy gradient modification on the standard seq2seq model does produce better results in the question generation system. The implementation was done in Keras with an initial learning rate of 2e-4. GloVe [9] was used for word embeddings.

C. *Answer-driven Deep Question Generation Based on Reinforcement Learning*
This paper [10] proposes better use of the target answer to generate questions. It also aims at generating the questions using a very difficult technique that is not much explored i.e. by reasoning through multiple documents. The author claims that most of the available research focuses on generating questions without considering the answer to the question that makes the question irrelevant to the context of the answer and such question lacks practical use. To tackle this problem the proposed model in this paper is based on the encoder-decoder mechanism. It tries to generate the question based on the answer, thus trying to integrate the context of the answer in the process of question generation. Two encoders are used in this paper one is the word encoder and the other is the graph encoder for documents and answer respectively. Then, the decoder is fed with an answer aware initialization module with a gated connection layer. This helps to introduce the documents and answer information to the decoder. Further, the answer with document and the semantic graph [11] integrates with a semantic rich fusion attention mechanism. The out of the vocabulary problem is solved by the copy mechanism [12]. Reinforcement learning used in this paper is to remove the exposure bias and RL is found to be very effective in such cases [13].

D. *Improving Neural Question Generation using Answer Separation*
This paper [14] addresses the problem that most question generation systems rely on the RNN Seq2Seq model which generally does not give high-level variability [15]. Most of the neural question generation model uses the copy mechanism [12] which increases the number of words from the passage to appear in the question. To overcome this problem, the paper suggests a novel architecture called answer separated seq2seq which utilizes information from both the target answer and the passage. The target answers from the passage is replaced by a token which provides a way to capture the contextual information of the question. The keyword net module which is part of the seq2seq architecture extracts information from the target which is passed before. The architecture comprises two encoders one for the passage and the other for the target answer. When the output of the encoder is passed to the decoder, it uses both the contextual feature of the passage from the attention mechanism and the information about the target answer using the keyword net. The author claims that this technique outperforms state of the art neural question generation model by a considerable margin on the SQuAD dataset [4].

E. *Asking Questions the Human Way: Scalable Question-Answer Generation from Text Corpus*
This paper [16] proposes Answer-Clue-Style-aware Question Generation (ACS-QG), a technique that is clear from the title claims to ask questions in a human way. The paper talks about a three-part technique - an information extractor that extracts information from the text (most of the information is assertive) so as to guide the generation of the question, a neural question generator which generates questions using the extracted information and a quality controller again built on neural networks that removes bad quality questions using text entailment methods. The author claims that compared to the exisiting work, their technique achieves significantly better results which also support their use of the clues and style information. One unique thing about this paper is the usage of the quality controller which is rare in other question generation techniques.

F. *AnswerQuest: A System for Generating Question-Answer Items from Multi-Paragraph Documents*
The proposed system [17] is an end-to-end model which has Question-Answer and Question-Generation to improve human comprehension. This is established by creating a list of question-answer pairs which captures all aspects of the provided document. It also discusses the individual tasks of question generation and question answering. The paper focuses on the challenging task of QA applied to multi-paragraph documents and shows the impact of incorporating a pre-trained text encoding model into an existing approach. Moreover, the paper reports a new set of

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002011 |

results for QA that assesses generalizability between datasets that are typically evaluated separately. For QG(Question Generation) they have used data augmentation by seeding the model with automatically-generated questions. This resulted in producing more fluent and answerable questions as observed by original human-authored data. In combining the two tasks into a single pipeline, the paper shows that the information targeted when humans make questions is the same as the information captured by the model. A web application showing the demo of the implementation of this technique can be found at qna.sdl.com.

G. *A Recurrent BERT-based Model for Question Generation*

This paper [18] tries to solve the problem of Question Answer generation using the BERT model [19]. This paper employs the BERT model in three different implementations. One of them is the direct implementation of the BERT that shows some of the difficulties and shortcomings of the direct employment of the BERT model. The other two models are improvement upon the BERT model in which sequential information is used from the previous decoded information. The implementation is tested on the SQuAD [4] dataset. The author claims to achieve significant performance at both sentence-level and paragraph-level input.

H. *Questionator - Automated Question Generation using Deep Learning*

This paper [20] generates Multiple choice based questions not only from text provided but also from images using image captioning techniques. It uses a Convolutional neural network (CNN) [21] as an encoder and Long short term memory [22] (LSTM) as a decoder. The CNN part helps in extracting features from the image and the decoder part helps in turning those features into natural language. This encoder decoder mechanism is part of the image captioning part [23] to extract textual information about the image and then using natural language processing technique, question answers are generated. Distractors are generated using the GLoVe [9] representation. The author claims to get outstanding results for the above technique. One of the shortcomings as discussed by the author is that the natural language processing techniques like extracting subject, object and predicate from the text are not useful in generating complex questions.

IV. CONCLUSION AND FUTURE WORK

According to our analysis, the models in the paper can broadly be classified as - Basic seq2seq model, Selection of content, Focusing on the answer, Incorporating external context QG and Pretrained QG Models. The scores considered in the evaluation measure are BLEU and F1. BLEU [24] is a standard in machine translation, which computes {1,2,3,4}-gram matches between generated and ground-truth questions. The F1 score is the harmonic mean of the precision and recall [25]. These scores in the above table are directly taken from the papers and all the models are trained on the SQuAD dataset unless specified otherwise.

The accuracy of these models can be formulated as follows-

Model	BLEU	F1
Basic seq2seq model	10.5%	40.1%
Selection of content	34.5%	92.97%
Focusing on answer	28.5% (Hotpot QA dataset)	-
Incorporating external context QG (Rulemimic model)	-	50.3%
Pretrained QG Models	34%	86.82

In this paper, we have tried to discuss some of the state of the art approaches to generate questions from the text. During the study, we have observed that the available datasets only help in generating factual questions and not understanding level questions based on bloom's taxonomy. Bloom's taxonomy is a measure of question difficulty,

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002011 |

defining a process to assess what difficulty level the question belongs to [26]. Our future work involves proposing a system that incorporates the difficulty level in the questions. Such a system will be more practical for use in educational assessment. Some of the research we found in this area involves - (1) Position embeddings are learned to capture the proximity hint of the answer in the input sentence. (2) Global difficulty variables are learned to control the overall "difficulty" of the questions [27]. Another framework can be progressively increasing the difficulty of the generated questions through step-by-step rewriting [28]. Future research in this area should be focused on incorporating more practical applications of the question-answer generation problem like educational assessment and creating a relevant dataset for the same.

REFERENCES

1. en.wikipedia.org. 2022. Question answering - Wikipedia. [online] Available at: <https://en.wikipedia.org/wiki/Question_answering> [Accessed 15 September 2021].
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. 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
4. Rajpurkar, Pranav, et al. "Squad: 100,000+ questions for machine comprehension of text." arXiv preprint arXiv:1606.05250 (2016).
5. Vinyals, Oriol, Meire Fortunato, and Navdeep Jaitly. "Pointer networks." arXiv preprint arXiv:1506.03134 (2015).
6. Yuan, Xingdi, et al. "Machine comprehension by text-to-text neural question generation." arXiv preprint arXiv:1705.02012 (2017).
7. Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078
8. Kirthevasan Kandasamy and Yoram Bachrach. 2017. Batch policy gradient methods for improving neural conversation models. Proc. of ICLR
9. Jeffrey Pennington, Richard Socher, and Christopher D.Manning. 2014. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP). pages 1532–1543. http://www.aclweb.org/anthology/D14-1162.
10. 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.
11. Liangming Pan, Yuxi Xie, Yansong Feng, Tat-Seng Chua, and Min-Yen Kan. 2020. Semantic graphs for generating deep questions. CoRR, abs/2004.12704.
12. Gu, Jiatao, et al. "Incorporating copying mechanism in sequence-to-sequence learning." arXiv preprint arXiv:1603.06393 (2016).
13. Ranzato, Marc'Aurelio, et al. "Sequence level training with recurrent neural networks." arXiv preprint arXiv:1511.06732 (2015).
14. Kim, Yanghoon, et al. "Improving neural question generation using answer separation." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. No. 01. 2019.
15. Serban, I. V.; Sordoni, A.; Lowe, R.; Charlin, L.; Pineau, J.; Courville, A. C.; and Bengio, Y. 2017. A hierarchical latent variable encoder-decoder model for generating dialogues. In AAAI, 3295–3301
16. Liu, Bang, et al. "Asking questions the human way: Scalable question-answer generation from text corpus." Proceedings of The Web Conference 2020. 2020.
17. Roemmele, Melissa, et al. "AnswerQuest: A System for Generating Question-Answer Items from Multi-Paragraph Documents." arXiv preprint arXiv:2103.03820 (2021).
18. Chan, Ying-Hong & Fan, Yao-Chung. (2019). A Recurrent BERT-based Model for Question Generation. 154-162. 10.18653/v1/D19-5821.
19. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
20. A. Srivastava, S. Shinde, N. Patel, S. Deshpande, A. Dalvi and S. Tripathi, "Questionator-Automated Question Generation using Deep Learning," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), 2020, pp. 1-5, doi: 10.1109/ic-ETITE47903.2020.212.
21. Yann LeCun, Patrick Haffner, Léon Bottou, and Yoshua Bengio. Object Recognition with Gradient-Based Learning, pages 319–345. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 10, Issue 2, February 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1002011 |

22. Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural Comput., 9(8):1735–1780, November 1997.
23. Zhongliang Yang, Yu-Jin Zhang, Sadaqat Rehman, and Yongfeng Huang. Image Captioning with Object Detection and Localization, pages 109–118. Springer International Publishing, 12 2017.
24. Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics.
25. Pado, Ulrike. (2017). Question Difficulty – How to Estimate Without Norming, How to Use for Automated Grading. 1-10. 10.18653/v1/W17-5001.
26. 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.
27. Gao, Yifan & Bing, Lidong & Chen, Wang & Lyu, Michael & King, Irwin. (2019). Difficulty Controllable Generation of Reading Comprehension Questions. 4968-4974. 10.24963/ijcai.2019/690.
28. Cheng, Yi & Li, Siyao & Liu, Bang & Zhao, Ruihui & Li, Sujian & Lin, Chenghua & Zheng, Yefeng. (2021). Guiding the Growth: Difficulty-Controllable Question Generation through Step-by-Step Rewriting. 5968-5978. 10.18653/v1/2021.acl-long.465.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING