Automatic Question Answer Generation Using NLP Techniques

Keshav Vardhan Narkudi, Hari Haran Juluri and Srujana Inturi

January 6, 2025

Automatic Question Answer Generation using NLP Techniques

Narkudi Keshav Vardhan
B.E. 4/4 CSE
Chaitanya Bharathi Institute of Technology
Hyderabad, India
keshav.narkudi@gmail.com

Juluri Hari Haran
B.E 4/4 CSE
Chaitanya Bharathi Institute of Technology
Hyderabad, India
hariharanjuluri@gmail.com

Srujana Inturi
Assistant Professor, CSE
Chaitanya Bharathi Institute of Technology
Hyderabad, India
isrujana_cse@cbit.ac.in

Abstract- ***The rising demand for scalable educational assessments and efficient question generation exposes the limitations of manual question construction, which is labor-intensive and difficult to scale with style and across domains. As a panacea to this, the "Automatic Question Answer Generator" uses the fine-tuned transformer-based T5 model on the SQuAD and RACE datasets to generate different types of questions: multiple-choice, fill-in-the-blank, factoid, and match-the-following. Advanced natural language processing techniques are employed: Keyphrase Extraction through PKE Library and TextRank for fill-in-the-blank questions; Named Entity Recognition and Sense2Vec for generating distractors validated by cosine similarity in multiple-choice questions; and finally, for refining match-the-following pairs, using WordNet and BERT. Factoid questions are generated using wh-type structures to target important concepts. The proposed system is a sturdy solution that bolsters contextual relevance, question variety, and its adaptability across domains and thus may possibly rise to the occasion for their dynamic needs associated with modern educational assessments.***

***Keywords-- Automated Question Generation, Deep Learning, Natural Language Processing (NLP), T5 Model, Convolutional Neural Networks (CNNs), SQuAD Dataset, RACE Dataset*.**

I - INTRODUCTION

The development in automated question-and-answer generation has become very valuable for teachers and trainers, as it allows scalable, low-cost, and efficient assessment of the comprehension and retention of knowledge in students. Traditional question generation relies solely on human inputs that are tedious and limited in scope. New developments in the fields of natural language processing and machine learning have produced models such as T5 to automatically generate questions in multiple-choice, fill-in-the-blank, factoid, and match-the-following formats. By utilizing transformer models and hyperparameter tuning on a number of datasets such as the SQuAD and RACE datasets, these QAG systems are now in a position to generate contextually coherent, diverse, and adaptive questions across different subjects, thus promoting a more dynamic and engaging learning experience [1][3].

However, the focus of the state-of-the-art QAG systems has been the introduction of neural networks focused on transformer-based models to provide the flexibility, relevance, and complexity of synthesized questions. For instance, the Handover QG architecture adopts a multi-task learning procedure where separate decoders would generate factoid or exam-type questions while switching on the fly to maximize the relevance and difficulty of each question in question [1]. Likewise, Opinerium introduces subjective question generation into the QAG process using user opinion to elicit engagement in open-domain contexts, such as media content [2]. Not with standing the advancement, these QAG systems are vulnerable to non-diversity, relevance, and distractor generation challenges when confronted with complex domain questions or difficulty scale modifications [3][5].

To remedy these challenges, we include the T5 model in the sysytem with paraphrasing integrated with the SQuAD and RACE datasets, as well as TextRank for keyphrase extraction, Sense2Vec for distractor generation, and WordNet for semantic disambiguation. These integrated NLP techniques make it a flexible, efficient, scalable QAG system which can accommodate all educational settings and respond to the increasing call for effective automated-tests in the learning and training arenas [6][10].

II – Related Work

In recent years, Automated Question Generation (QG) has gained substantial traction within the educational and content review sectors, primarily due to the potential of NLP and machine learning techniques to efficiently produce meaningful and diverse questions. QG systems aim to automate the labor-intensive task of question creation, offering scalability and adaptability across domains such as academic learning, corporate training, and digital content assessment. These systems employ a variety of methods, including rule-based and model-based approaches, to generate questions that are relevant, contextually accurate, and tailored to the desired complexity level. Advanced models, such as transformer-based architectures, have enabled the creation of multiple question type from simple factoid questions to complex, exam-style assessment making QG a transformative tool in modern educational technology.

A. Handover QG: A Framework for Multi-Task Question Generation

Chung et al. [1] introduced the Handover QG framework, a pioneering hybrid approach for automated question generation that combines exam-type and factoid question generation tasks. The Handover QG framework is a new hybrid automated question generation approach helping students by combining generation of two kinds of questions: exam-type and factoid questions; multi-task learning to enhance the generalization across the different question types is an important building block of this framework based on a transformer approach such as BART. The system switches off between two specialized decoders-one for factoid questions and one for exam-type questions, thus balancing complexity and contextual relevance of the generated questions. This novel decoder fusion approach makes the model adaptable to various educational contexts by producing questions of varying cognitive challenge.

Reinforcement Learning (RL) is central to training statistical structure conversion, thus optimizing the decoding process inside Handover QG. The ability for the model to dynamically alternate between its two decoders means questions produced can be in varying degrees of complexity in relevant contexts, thus maximizing engagement and usability. Tuned by RL, the decoding sequences are specialized towards the generation of complex questions for higher-order assessments. That capability of dynamic adjustment becomes a virtue to enhance scalability of the program, thereby placing Handover QG squarely within the arsenal of tools to automate complex question generation in personalized learning solutions [1].

B. Opinerium: Subjective Question Generation Using Large Language Models

Babakhani et al. [2] proposed Opinerium, a subjective question generation system aimed at eliciting opinions from users based on media content. Opinerium distills subjective information from news articles to generate opinion-based questions, unlike conventional QG systems that focus on factual questions. The model is trained on a custom-built dataset of 40,000 articles, with flan-T5 and GPT-3 backbone architectures dedicated to question generation. The model was evaluated using several metrics, including BLEU, ROUGE, and BLEURT, where flan-T5's overall quality was better in respect to question relevance and user engagement compared to GPT-3.

This system expands the ambit of QG to cover public opinion mining and user interaction in media-driven contexts. Opinerium resolves the historical gulf between knowledge-driven QG and user engagement by adapting its language models for semantic relevance. The generation of reflective, opinion-driving questions signifies that large language models can drive QG to a beyond-historical context, making it a valuable asset to the field of interactive natural language processing [2].

C. Automatic Question Answer Generation using T5 and NLP

Virani et al. [3] introduced an automatic question-answer generation system based on the T5 model, which supports multiple question formats, including Wh-questions, multiple-choice, fill-in-the-blank, and true/false. The system was designed for applications in e-learning, customer service, and digital assessments, stressing personalized learning with question designs that feature user knowledge levels. Preprocessing steps, like tokenization and removal of stop-words, have enhanced the ability of the model to generate coherent question-answer pairs.

Experimental results have revealed the system's efficacy in increasing student engagement while minimizing the workload of educators. Despite pedagogical setting limitations, the Create-a-Test function of valid test question customization allows real-time feedback that supports adaptive learning environments. This flexibility and customization make the system the perfect fit for various educational and business settings; thereby validating its varied applicability in automated assessments.

D. Context-Based Question Answering System with Suggested Questions

Kumari et al. [4] have resorted to a hybrid QA-QG model that produces contextually relevant question-answer pairs. The model integrates BERT for QA and T5 for QG to formulate so that the resulting questions are grounded in the answers, thus improving contextual accuracy. The BM25 algorithm ranks questions in relation to the information they hold, allowing the user most relevant information in real time.

On the other hand, facilitating multi-modal PDF processing with the incorporation of image-based text significantly boosts this system for use in the education and customer support sectors. This hybrid model endorses the benefit of combining QA and QG in one framework, thus providing an intuitive interface that helps dynamic and interactive information retrieval [4].

E. Automated Question Generation using Deep Learning:

Srivastava et al. [5] presented the Questionator-a deep learning based system for the generation of multiple-choice questions from text and visuals. Its approach incorporates a CNN-LSTM pipeline for image captioning, dependency parsing for question generation, and GloVe embeddings for distractor generation, making the tripartite architecture viable to generate holistic questions that embrace text and visual data.

However, still, it represents an empirical approach to question-and-answer generation, which has its tender zone in being able to implement semantic alignment between questions and answer choices in multi-modal settings. The dependency parsing and word embedding strongly indicate necessary work and improvement required for quality and consistency of the generated question. However, the Questionator is a milestone toward multi-modal QG to a range of educational and entertainment applications.

F. Automatic question-answer pair generation using Deep Learning

Kumar et al. [6] established a deep-learning framework to produce high-quality question-answer pairs, using a fine-tuned BERT model as the core mechanism. It goes through a systematic approach, including answering extraction, question generation, and validation of outcomes derived, ensuring highly standardized and satisfactory outputs where it befits. The system was assessed with the SQuAD dataset, demonstrating impressive generation of Wh-questions aligned with the extracted answers.

This framework gives the possibility of undertaking QA tasks with the help of transformer-based architecture, which is promising, especially for automating assessments in the educational and professional domains. The dependence on the pre-trained models again emphasizes more on the fine-tuning of domain-specific aspects if working with additional contextualities, thus promising further avenues for future research in automated QA [6].

G. Automatic Question Generation from Textual data using NLP techniques

Rathi et al. [7] have developed a web application for question generation from textual data, both subjective and objective, automatically. The user feeds the system with a particular input which then performs tokenization, POS tagging, and syntactic tree parsing to extract vital information for the question. Subjective questions are created using topic templates, while objective questions are derived from keywords that lead to fill-in-the-blank creations.

The evaluation of the system has shown that the rate achievability of questions generated by it to make logical and grammatically correct inquiries is 73%. Areas for improvement were noted within an answer's correctness and context relevance. The system, while promising, is faced with flexibility and scalability issues in complexity question types and across diverse domains. Future additions include integration of advanced NLP techniques and expansion on the ability to conduct domain-based assessments.[7]

H. Knowledge Map Construction and Question Answering System Design Based on NLP and Neural Network Algorithm

Xu et al. [8] proposed a knowledge map-based QA system, combining NLP and neural network algorithms for processing large datasets with entity recognition, relationship extraction, and classification as key features of the system that add on to its capacity for arranging and interpreting complex data. Thereby, this enables it to generate accurate and contextually relevant questions, which is an asset to information-dense domains. The integration of knowledge maps with neural networks demonstrates the system's scalings and potential for real-world applications, such as education and knowledge management. However, challenges in ensuring data diversity and adapting tovaried input formats suggest areas for future improvement [8].

I. Generation of English Question-Answer Exercises from Texts Using Transformers-Based Models

De Souza et al. [9] developed a framework-a deep learning convolutional neural network (CNN)-to carry out standard based grading of programming assignments in an automated way. The CNN architecture scored the programming exercises from two perspectives-structural correctness and semantic alignment. This work overcomes some of the challenges of rule-based grading, for instance, using machine learning coupled with conventional paradigms to obtain approaches to instance learning in inputs such as logic used by learners, establishing full structural correctness checks, and announcing or counting patterns that appeared in the learner submissions as referred in Figure 2.1 .

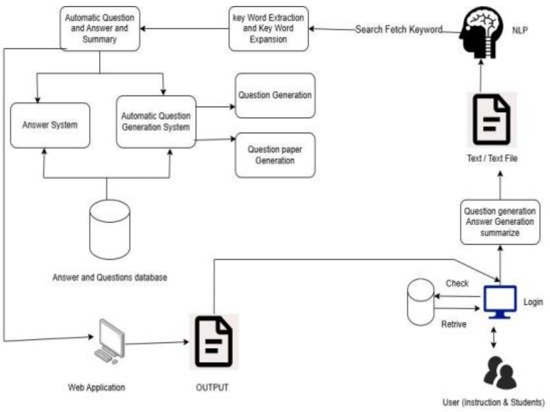

Figure 2.1 System Architecture

J. Automated Question and Answer Generating System for Educational Platforms

Rahaman and Hoque [10] brought in another way of doing it by grading programming assignments via computer-like information retrieval techniques, in particular, Term Frequency-Inverse Document Frequency (TF-IDF) technology. With this method, not only does the instructor check if the submission is word-for-word with the answer key, but also this is a new way of automated grading which content relevancy is the main part and not syntactic similarity.

Even though this system has a new way of classifying, it still was not able to deal with such syntax changes as different names of identifiers and layered constructions, which influenced its trustworthiness. The authors highlighted one of the potential benefits of their model might be its ability to handle different programming styles (that is robustness is increased) and also deep learning techniques might be used to greater accuracy in terms of grading. This method may lead to success in resolving the difficulties of automated assessment in programming education [10].

III- INSIGHTS FROM LITERATURE AND EXISTING CHALLENGES

Recent advancements in question generation (QG) and answering (QA) systems have led to a diverse set of frameworks and models that address various types of educational and interactive needs. The models and systems that are referred to as reinforcement learning further boost these other systems to not only produce accurate questions, but also to adjust the difficulty accordingly. Reinforcement learning improves QG models' ability to dynamically balance between complexities, making them well-suited for the higher-order thinking assessments. But still, in maintaining the relevance of questions across the domains with maximum complexity, the technology is not up to the mark as switching information through decoders may tend to over-enhance complexities; on different occasions, it can lose contextual relevance in certain educational domains.

Systems like Opinerium have explored opinion-based QG, using large language models (LLMs) such as Flan-T5 and GPT-3 to formulate questions that solicit users' opinions rather than factual responses. This research expands the applications of QG into engaging scenarios such as news and media but is met with the challenges of optimizing semantic relevance and ensuring questions represent diverse opinions accurately. With systems like Opinerium achieving commendable performance in opinion-based QG, the avenue for future work still exists in the form of training a single model that captures a larger range of subjective perspectives on matters touching on nuanced topics in real-world media.

The hybrid models that integrate the works of QA and QG tasks include the Context-Based Question and Answering System, which combines BM25 ranking and BERT-based question generation to ensure effective retrieval of information. Such systems give a contextual question and answer pair to the desirable cases in terms of customer support and e-learning. The challenge, however, is that the integration of heterogeneous content types, such as PDF and images of questions, could challenge the contextual awareness during the extraction of questions out of complex documents. This calls for yet more effort for research on mediating multimodal integrations to boost system adaptability across a variety of input modes for answering and other relevance measures.

Table 3.1 Comparative Analysis of Automatic Question Answer Generators

S.No.	Author(s)	Method	Gap/Remarks
1	Chung et al. [1]	Multi-task learning with decoder fusion and reinforcement learning (RL).	Complexity balancing is context-specific; RL optimization can struggle with diverse datasets.
2	Babakhani et al. [2]	Subjective QG using flan-T5 and GPT-3 fine-tuned on a custom dataset.	Semantic relevance of opinion-based questions could improve for nuanced topics.
3	Virani et al. [3]	Automatic question-answer generation using T5 and NLP techniques.	Limited scalability for highly domain-specific or complex question types.
4	Kumari et al. [4]	Hybrid QA/QG system integrating BERT and T5 with BM25 ranking.	Integration of multi-modal inputs (e.g., PDF, images) needs better contextual alignment.
5	Srivastava et al. [5]	CNN-LSTM for image captioning, dependency parsing for QG, GloVe for distractors.	Semantic alignment between distractors and questions in multi-modal contexts needs refinement.
6	Kumar et al. [6]	Question-answer pair generation using a fine-tuned BERT model.	Domain-specific fine-tuning required for better contextual accuracy.
7	Rathi et al. [7]	Web-based application using NLP techniques for subjective/objective QG.	Contextual accuracy of generated answers and support for complex domains require improvement.
8	Xu et al. [8]	Knowledge maps integrated with neural networks for QA.	Scalability and adaptation to diverse datasets remain challenging.
9	De Souza et al. [9]	CNN for programming assignment grading, focusing on structure and semantics.	Feature extraction and support for diverse programming styles need enhancement.
10	Rahaman and Hoque [10]	TF-IDF for evaluating semantic similarity in programming assignments.	Syntax variations (e.g., identifiers) reduce reliability; integration with deep learning could help.

IV- Automatic Question Answer Generation

The automatic question-answer generation project is divided into several key stages, which should lead to the creation of a reliable and efficient system for educational and content review applications. The methodology hence involves data preprocessing, feature extraction, question generation, and evaluation, ensuring that all these approaches come into play in developing a sophisticated automatic question-answer system.

DATA PREPROCESSING

This is the first step which pre-processes the input text into a processible condition to enable the analysis and question generation. Techniques at work are: it starts with tokenization wherein the input text gets broken into individual words or tokens thus making the data easier for the model, followed by sentence splitting wherein the text is divided into sentences so that this very system can generate specific questions relevant to the text. Stop words are common words that do not add much meaning to the content. They really help filter out noise and focus on meaningful terms needed for developing a question into contextually appropriate wording.

FEATURE EXTRACTION

This is core to the understanding of the information carried in the text. The project makes use of the T5 model-the Text-to-Text Transfer Transformer-on the extraction of key features by its capabilities in other NLP tasks. The choice of using a method known as NER for proper detection of things such as names, dates, and locations used as a basis to formulate fact-based questions. The T5 model conducts a more in-depth analysis of the text context, especially relationships and connections that contribute to the construction of accurate and relevant questions.

WH-TYPE QUESTION – ANSWER GENERATION

It produces factual as well as inferential questions over the processed input. These Wh-questions are created using the pre-trained T5 model. The pre-trained T5 model develops these questions using "What," "Where," "Who," and "When," thus ensuring contextual alignment with the input text. Questions are meaningful in this stage and therefore directly bound to the content. The system also predicts accurate answers and makes complete question-answer pairs that are contextually appropriate and useful for education.

MCQ-BASED QUESTION-ANSWER GENERATION

It is designed to generate multiple-choice questions using distractors that target different users. The system then makes use of the ability of NER to identify relevant entities and then Sense2Vec in order to suggest words that may potentially become distractors. Since they rely on cosine similarity, these are selected so that the semantic closeness with regard to the right answer is duly ensured in being relevant but distinct. Answer choices are randomly ordered for any given question so that the generated questions are not predictable. This hence makes the assessment more or less fair and robust.

FILL IN THE BLANKS QUESTION GENERATION

It forms fill in the blank's questions using techniques of keyword extraction, filling missing words. The PKE library-which is a Python Keyword Extraction library-is used to extract the most important words in the text. It, in turn, substitutes these important words with blanks and so fill-in-the-blank questions will be generated. This will ensure only selected important terms are utilized, which is due to the fact that it ranks the importance levels of the keywords within a context with Text Rank Algorithm. This approach focuses on some of the major concepts of the topic so that the fill-in-the-blank questions have a strong reinforcement effect.

MATCH THE FOLLOWING QUESTION ANSWER GENERATION

It is designed for generating questions with terms or concepts that match. In the process, PKE is used in identifying the key terms that are suitable for matching exercises from the given text. The context of every term is analyzed very keenly, and Word Sense Disambiguation (WSD) is performed using the BERT model. It ensures that each word meaning is well interpreted so that the system can generate appropriate and relevant matches. A set of matching pairs from the output results in several matching pairs facilitating a deeper understanding of the content

MODEL TRAINING AND EVALUATION

The process of training and testing the model allows efficiency across several question types by using datasets such as SQuAD and RACE. Data are divided into training, validation, and testing subsets for more effective learning without overfitting. To achieve generalization, multi-task learning is adopted in training, which allows the model to be trained on shared parameters for different question types. Techniques applied include NER, Sense2Vec, and WordNet to enhance the quality of distractors in multiple-choice questions. The model fine-tunes its hyperparameters, namely learning rate and batch size using grid or random search. The AdamW optimizer is used to stabilize the training process. metrics for evaluating lexical and semantic quality include BLEU, ROUGE, METEOR, and BERTScore. The model learns to classify question types using categorical cross-entropy, and reinforcement learning is incorporated into the generation process to ensure that the outputs are both relevant and of high educational quality.

V- CONCLUSION

This survey provided a thorough review of approaches used for automatic question generation by categorizing them into three main use cases: standalone question generation, visual question generation, and conversational question generation. It also analyzed the datasets commonly adopted for this task, followed by a discussion on the challenges and the applications of these systems. While the community has made great strides, especially in question generation from text, this task is still riddled with problems that need to be looked into.

Of course, many generated questions are often less natural and also lack contextual relevance, thereby limiting their ability to extract significant information. Achieving semantic coherence and making questions informative and interesting have been established as ongoing challenges. That said, the evaluation metrics for determining question quality are still maturing and need somewhat more standardization. Multimodal-Input-powered advancement has garnered a lot of excitement, yet both its scientific underpinnings and subsequent embedding within question generation systems remain for tomorrow's research. Models that combine these diverse techniques in an application-specific context need to be built to take this field further. The evolution smoothed by this process would allow the development of new-age question generation systems that are more robust and context-aware in serving a wide range of user needs.

REFERENCES

[1] Ho-Lam Chung , Ying-Hong Chan, and Yao-Chung Fan, " Handover QG: Question Generation by Decoder Fusion and Reinforcement Learning ," IEEE/ACM Transactions Onaudio, Speech, Andlanguageprocessing, Vol.32,2024.

[2] Pedram babakhani fikret sivrikaya 1,andreas lommatzsch1, torben brodt 2,doreensacker2, 3,andsahinalbayrak Opinerium: Subjective Question Generation Using Large Language Models. IEEE Access, vol. 12, pp. 131152-131167.

[3] Altaj Virani, Rakesh Yadav, Prachi Sonawane, Smita Jawale , "Automatic Question Answer Generation using T5 and NLP," IEEE Sensors Journal, vol. 22, no. 12, pp. 12064-12078, Jun. 2022.

[4] Vijay Kumari, Srishti Keshari, Yashvardhan Sharma, Lavika Goel, "Context-Based Question Answering System with Suggested Questions," IEEE Transactions on Information Forensics and Security, vol. 18, pp. 5651-5664, Aug. 2023.

[5] Animesh Srivastava Shantanu Shinde, Naeem Patel, Siddhesh Despande , Anuj Dalvi and Shweta Tripathi, " Questionator- Automated Question Generation using Deep Learning," IEEE Access, vol. 11, pp. 138986-139003, Dec. 2023.

[6] Alok Kumar, Aditi Kharadi, Deepika Singh, Mala Kumari ,"Automatic question-answer pair generation using Deep Learning," IEEE Access, vol. 11, pp. 16984-16993, 2023.

[7] Snehal Rathi, Pawan wawage, Amrut Kulkarni, "Automatic Question Generation from Textual data using NLP techniques," IEEE Transactions on Cognitive and Developmental Systems, vol. 14, no. 2, pp. 672-683, June 2022.

[8] Nannan Xu; Jiangfeng Chen; Chenguang Hu, "Knowledge Map Construction and Question Answering System Design Based on NLP and Neural Network Algorithm," IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 7, pp. 4626-4638, July 2022.

[9] Gonzalo Berger, Tatiana Rischewski, Luis Chiruzzo, Aiala Ros´, "Generation of English Question Answer Exercises from Texts using Transformers based Models," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 3238-3253, 2022.

[10] Maheraj Thiruvanantharajah; Nawanjana Hangarangoda; Samantha Rajapakshe "Automated Question and Answer Generating System for Educational Platforms," IEEE Transactions on Information Forensics and Security, vol. 18, pp. 4258-4271, 2023