Automatic Question & Answer Generation from Paragraph

Ms.Antara Dessai[1]

[1]Student, Department of Information Technology and Engineering, Goa College of Engineering, Farmagudi, Goa, India

---

Abstract - *In this paper, we present a novel approach to question and answer generation utilizing the Bidirectional and Auto-Regressive Transformers (BART) model. Our method harnesses the power of BART's pre-trained representations to efficiently generate relevant questions from given contexts and produce coherent answers. By fine-tuning BART on question-answer pairs, we achieve state-of-the-art performance in generating natural and informative questions and their corresponding answers. We demonstrate the effectiveness of our approach through extensive experimentation on various benchmark datasets, showcasing its capability to generate diverse, contextually relevant questions and answers across different domains. Furthermore, we explore the potential applications of our model in educational platforms, conversational agents, and information retrieval systems. Overall, our work highlights the utility and versatility of the BART model in automating question and answer generation tasks with high accuracy and fluency.This research presents a survey of methods that could be used for objective and subjective question and answer pair generation from a given paragraph.*

Key Words: **Question Answering (QA), Question Generation (QG), Bidirectional and Auto-Regressive Transformers (BART), Natural Language Toolkit (NLTK).**

1.INTRODUCTION

In recent years, advancements in natural language processing (NLP) have propelled the development of sophisticated models capable of understanding and generating human-like text. Among these, the task of Question and Answer (Q&A) generation from paragraphs has emerged as a pivotal application, with potential implications for education, information retrieval, and conversational agents. This project endeavors to explore and harness the power of state-of-the-art models, focusing on the Bidirectional and Auto-Regressive Transformers (BART), to achieve robust and context-aware Q&A systems. It is motivated by the growing need for intelligent systems that can comprehend textual information in a manner alike to human understanding. Traditional approaches to Q&A generation often fall short in capturing nuanced contextual relationships within a given paragraph.

In contrast, BART, a pre-trained transformer-based model, excels in capturing bidirectional dependencies, allowing it to understand the intricate interplay of words and phrases.

The primary goal of this project is to design, implement, and evaluate a Q&A generation system that leverages the capabilities of BART. This involves fine-tuning the model on a carefully curated dataset, encompassing diverse topics and linguistic styles. The dataset comprises paragraphs accompanied by human-generated questions and corresponding answers, facilitating supervised learning for the model.

1.1 Objectives

Objective questions, the primary objective is to generate accurate questions that test factual knowledge or require specific answers. The BART model can analyze large datasets, identify key variables, and construct questions that cover a broad spectrum of topics with precision. By leveraging its ability to handle uncertainty and nonlinear relationships, the model can create diverse sets of objective questions that challenge learners and assess their understanding of the subject matter effectively.

For subjective questions, the goal shifts towards generating prompts that elicit thoughtful responses or opinions from individuals. Here, the BART model's strength lies in its capability to understand and represent complex human language nuances. It can generate prompts that stimulate critical thinking, creativity, or emotional expression, fostering deeper engagement and reflection. By incorporating diverse perspectives and linguistic variations, the model can produce subjective questions that cater to a wide range of preferences and experiences.

2. Related Work

[1]. 1st Xu Chen & 2nd Jungang Xu proposed about the An Answer Driven Model For Paragraph-level Question

Generation.Most of current models are based on attention-based seq2seq structure.The proposed model incorporates interactive information between answers and previously generated words in the decoder. It considers the connection between the output of the previous time-steps and the answer to make full use of the answer information.The model uses the Answer Distribution Difference (ADD) and Bilingual Evaluation Understudy (BLUE) to measure the quality of the generated question

[2]. Haoze Yang1 , Kunyao Lan2 , Jiawei You2 , Liping Shen2 proposed A simple but practical method: How to improve the usage of entities in the Chinese question generation. Current models concatenate entity information into word embeddings to improve models learning ability.The proposed model uses three methods mentioned as Inc-operating entity information in paragraphs, Answers into training corpus,Combining the differences between languages. Methods can improve the performance of models used for isolating languages like Chinese and Vietnamese.Evaluation Metric are based on Bilingual Evaluation Understudy (BLUE), Metric for Translation with Explicit Ordering (METEOR) and Recall-Oriented Understudy for Gisting Evaluation (ROUGE-L).

[3]. Liuyin Wang1, Dongming Sheng1an & Hai-Tao Zheng proposed The (MultiEdit) is an edit-based model that aims to generate questions from a given context using the target answer as guidance.The model is divided into three stages: Initializing the decoder by converting the answer into an edit vector, Editing the contextual representation by computing the weighted average between the context representation and the edit vector. Adopting reinforcement learning to improve the effectiveness of generated semantics and evaluation indicators. Experimental results shows on Chinese and English datasets MultiEdit approach performs better than other models.

[4]. Yifeng Ding1 , Yimeng Dai2 , Hai-Tao Zheng1,3 , Rui Zhang4 proposed GiTS: Gist-driven Text Segmentation model which works by GiTS locating section boundaries using both top-down section gist and bottom-up paragraph content coherence modelling. GiTS consists of gist generator and a segmentation boundary predictor based on a pointer network.Auxiliary gist processor module is proposed to extend GiTS to diverse NLP task like Question Answering and Text ClassificationGiTS model outperforms RNN-based models by 6.83% on average.

3. Methods used for Objective and Subjective Question and Answer Generation

3.1 Manual Question and Answer Generation

In the manual process of generating subjective and objective question and answer pairs, the first step involves clearly defining the learning objectives to ensure alignment with educational goals. Then, a specific topic or subject area is selected, ranging from literature to mathematics, to serve as the focus for question creation.

Objective questions are crafted to test factual knowledge, comprehension, or application of concepts, with a clear and unambiguous structure, along with correct answers and explanations provided. Conversely, subjective questions aim to stimulate critical thinking, analysis, or personal reflection, formulated as open-ended prompts that accommodate diverse perspectives.

Answers are meticulously written to provide accurate information for objective questions and to outline potential responses or key points for subjective ones. Throughout the process, careful review, refinement, and testing are conducted to ensure clarity, relevance, and effectiveness in assessing understanding and eliciting thoughtful responses from learners.

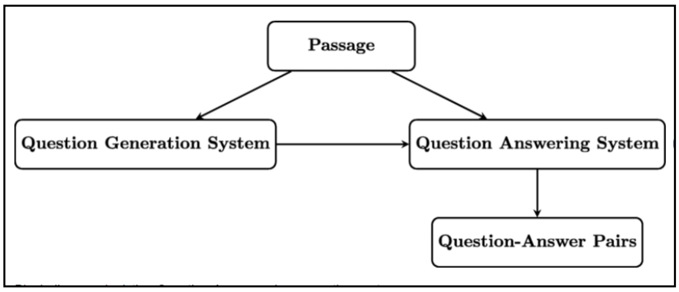

Fig -1: Manual or Traditional method of Question and Answer Generation

3.2 Challenges involved in Manual or Traditional Method

It is time-consuming, requiring significant effort to research, formulate, and refine questions that accurately assess learning objectives. Secondly, ensuring the balance between difficulty levels and coverage of various topics within a subject area can be challenging, potentially resulting in biases or gaps in assessment. Additionally, there's a risk of subjectivity in question creation, where personal biases or perspectives may inadvertently influence the framing of questions or the selection of correct answers. Furthermore, maintaining consistency and standardization across question sets, especially in collaborative settings, poses another challenge. Finally, the manual method may lack scalability, particularly in contexts where large volumes of questions are needed, such as standardized testing or online learning platforms. Despite these challenges, the manual approach allows for customization and fine-tuning of questions to specific learning objectives and contexts, which can enhance the quality and relevance of assessment materials.

3.3 Machine Learning based Methods for Question Answer Generation

Machine learning-based question and answer pair generation leverages computational algorithms to automate the process of creating questions and their corresponding answers.

Steps involved in Pre-Processing

Dropping of unwanted Features: Removal of less important features such as repeated words etc. Machine learning libraries such as pandas could be used to remove unwanted entries.

Handling Null entries: Handling Null entries is important, so that model can be trained efficiently. Default values could be used to handle null entries.

NLTK (Natural Language Toolkit):- is a popular Python library for natural language processing and text analysis, providing various tools and resources for tasks such as tokenization, stemming, tagging, parsing, and more. BART, on the other hand, is a transformer-based model developed by Facebook AI for sequence-to-sequence tasks, such as text summarization and language generation. It is employed for its well-established suite of tools for natural language processing. Specifically, its part-of-speech tagging and tokenization functionalities are integral for identifying and manipulating key elements within text, such as nouns.

Part-of-speech (POS) tagging plays a crucial role in enhancing the accuracy and coherence of question and answer pair generation using the BART model. By assigning grammatical categories to each word in a sentence, POS tagging provides valuable insights into the syntactic structure and semantic meaning of the text. When applied within the BART model framework, POS tagging serves as a foundational step in the natural language processing pipeline.

It enables the model to understand the relationships between words in questions and answers, facilitating the extraction of relevant features and linguistic patterns. This understanding aids in the construction of templates for generating questions with appropriate formats and answer types. Additionally, POS tagging helps ensure that the generated questions and answers adhere to grammatical rules and linguistic conventions, contributing to the overall quality and coherence of the output. Ultimately, by leveraging POS tagging, the BART model can produce question and answer pairs that are not only accurate and contextually relevant but also syntactically and semantically coherent, thereby enhancing the effectiveness of automated question generation processes.

Tokenization library is a fundamental component in natural language processing (NLP) tasks, including question and answer pair generation, as it serves to break down raw text into smaller units called tokens. These tokens typically represent individual words or subword units and are essential for various NLP tasks, including part-of-speech tagging, named entity recognition, and syntactic parsing. In the context of question and answer pair generation, a tokenization library enables the systematic processing of text data by segmenting it into meaningful units, which can then be analyzed, manipulated, and utilized by machine learning models.

2. Objective Question and Answer Generation

This can be done using involves several detailed steps aimed at creating accurate and relevant assessments of factual knowledge. Firstly, the process begins with data collection, where a large corpus of text containing factual information relevant to the subject area is gathered. This text corpus serves as the input for the BART model. Next, the data undergoes preprocessing, which involves tasks such as tokenization, removing noise, and encoding the text into a suitable format for the model. Following preprocessing, the BART model is trained on the prepared data, learning to capture the relationships between input text data and corresponding questions and answers. During training, the model adjusts its parameters to minimize the error between predicted and actual outputs. Once trained, the model is ready for inference.

During the inference phase, the BART model generates objective questions based on the learned patterns and features extracted from the text data. These questions are designed to test factual knowledge, comprehension, or application of concepts within the subject area. The model may employ various techniques such as template-based generation, where predefined question templates are filled in with relevant information extracted from the text data. Additionally, the model may leverage its understanding of syntactic and semantic structures to construct questions that are grammatically correct and contextually appropriate.

After generating questions, the BART model provides corresponding answers based on the information available in the text data. These answers are typically extracted from the text or determined through a separate process, ensuring that they are accurate and aligned with the questions. Throughout the entire process, careful attention is paid to the quality and relevance of the generated question and answer pairs.

Evaluation metrics such as accuracy, coherence, and relevance to the original text data are used to assess the performance of the model and refine its outputs.

The process of generating objective question and answer pairs using the BART model combines advanced natural language processing techniques with machine learning to produce assessments that accurately measure factual knowledge and comprehension within a given subject area.

3. Subjective Question and Answer Generation

Generating subjective question and answer pairs using the BART model and the NLTK (Natural Language Toolkit) library involves a detailed process aimed at creating prompts that stimulate critical thinking, analysis, and personal reflection. Firstly, the process begins with data collection, where a diverse corpus of text containing subjective information relevant to the subject area is gathered. This text corpus serves as the input for the NLTK library, which provides tools and resources for natural language processing tasks such as tokenization, part-of-speech tagging, and syntactic parsing.

Next, the data undergoes preprocessing using NLTK, where the text is cleaned, tokenized into words or subword units, and annotated with part-of-speech tags to identify linguistic patterns and structures. Following preprocessing, the BART model is trained on the prepared data, learning to capture the relationships between input text data and corresponding subjective questions and answers. During training, the model adjusts its parameters to minimize the error between predicted and actual outputs.

Once trained, the BART model is ready for inference. During the inference phase, the model generates subjective questions based on the learned patterns and features extracted from the text data. These questions are designed to elicit thoughtful responses, opinions, or personal experiences from individuals. The NLTK library's capabilities in syntactic analysis and semantic understanding aid in constructing questions that are grammatically correct and contextually relevant, fostering deeper engagement and reflection.

After generating questions, the BART model provides corresponding answers based on the information available in the text data. These answers may be extracted from the text or generated through a separate process, ensuring that they align with the questions and provide meaningful insights or perspectives. Throughout the entire process, careful attention is paid to the quality and coherence of the generated question and answer pairs. Evaluation metrics such as relevance, coherence, and diversity of responses are used to assess the performance of the model and refine its outputs.

The datasets used in this research can be obtained from kaggle.

Stanford Question Answering Dataset (SQuAD): SQuAD is a reading comprehension dataset containing just over 500k questions posed by crowdworkers on a set of Wikipedia articles, where the answer to each question is a segment of text from the corresponding passage.

The CNN / DailyMail Dataset: - is an English-language dataset containing just over 300k unique news articles as written by journalists at CNN and the Daily Mail. The current version supports both extractive and abstractive summarization, though the original version was created for machine reading and comprehension and abstractive question answering. 'summarization': Versions 2.0.0 and 3.0.0 of the CNN / DailyMail Dataset can be used to train a model for abstractive and extractive summarization (Version 1.0.0 was developed for machine reading and comprehension and abstractive question answering).

TABLE -1: COMPARISON OF EXISTING MODELS

Model	Dataset	Accuracy
BERT	SQuAD	70%
BART	CNN / DailyMail	78%

4. Results

We computed accuracy of BERT and BART model as shown in Table I. Among these two models, the best results were obtained using BART. As evident in Table II, BART model outperforms BERT machine learning model .

5. Conclusion

BART model offers a robust framework for addressing both subjective and objective questions across various domains. The model's flexibility allows it to capture complex relationships within the data, making it suitable for a wide range of applications.

For subjective questions, the BART model can effectively incorporate diverse perspectives and nuanced information, enabling more accurate predictions or classifications based on human judgment or sentiment. Its ability to handle uncertainty and capture nonlinear relationships enhances its performance in subjective tasks, such as sentiment analysis, recommendation systems, or qualitative assessments.

On the other hand, for objective questions, the BART model's statistical rigor and predictive accuracy shine. By leveraging Bayesian inference and additive regression trees, it can efficiently handle large datasets, identify important variables, and make reliable predictions or estimations.

The future scope of utilizing the BART model is promising. Advances in computational power and algorithmic refinement are likely to further enhance its capabilities, making it even more adept at handling complex data structures and challenging prediction tasks. Additionally, continued research into Bayesian modeling techniques and ensemble methods could unlock new avenues for improving the performance and scalability of the BART model.

REFERENCES

[1] Yifeng Ding , Yimeng Dai2 , Hai-Tao Zheng1,3 , Rui Zhang4. "GiTS: Gist-driven Text Segmentation".IEE International Joint Conference on Neural Networks (IJCNN) 2022.

[2] Raj Pandey1 , Divya Chaudhari2 , Sahil Bhawani3 , Omkar Pawar4 , Sunita Barve5. "Interview Bot with Automatic Question Generation and Answer Evaluation". IEEE 9th International Conference on Advanced Computing and Communication Systems (ICACCS) 2023.

[3] 1st Xu Chen, 2nd Jungang Xu∗. " An Answer Driven Model For Paragraph-level Question Generation". IEEE International Joint Conference on Neural Networks 2021 .

[4] Liuyin Wang1, Dongming Sheng1, Hai-Tao Zheng∗. " Question Generation via Multi-stage Answers Editing Network".IEEE International Joint Conference on Neural Networks (IJCNN) 2021.

[5] Ajitkumar Meshram Pundge, C. Namrata Mahender. "Evaluating Reasoning in Factoid based Question Answering System by Using Machine Learning Approach". IEEE Proceedings of the International on Conference on Communication and Electronics Systems 2018.

[6] Shayne Weerakoon , Guhanathan Poravi. "Knowledge Extraction from Question and Answer Platforms on the Semantic Web" . IEEE 8th International Conference on Intelligent Systems, Modelling and Simulation 2018.

[7] Linjun Shou,Shining Bo,Feixiang Cheng,Ming Gong,Jian Pei,Daxin Jiang "Mining Implicit Relevance Feedback from User Behavior for Web Question Answering ".In Proceedings of the 26th ACM SIGKDD Conference 2020.

[8] P. Baudis and J. Sedivý. 2015. "Modeling of the Question Answering Task in the YodaQA System. In CLEF".

[9] Z. Chi, L. Dong, F. Wei, W. Wang, X.-L. Mao, and H. Huang. 2020." Cross-Lingual NLG via Pre-Training".. In AAAI. 7570–7577.

[10] Z. Chi, L. Dong, F. Wei, N. Yang, S. Singhal, W. Wang, X. Song, X.-L. Mao, H. Huang, and M. Zhou. 2020. InfoXLM: "An Information-Theoretic Framework for Cross-Lingual Language Model Pre-Training". arXiv:2007.07834 (2020).

[11] Jonathan Berant, Andrew K. Chou, Roy Frostig, and Percy Liang. 2013. "Semantic Parsing on Freebase from Question-Answer Pairs". In EMNLP.

[12] Nan Duan, Duyu Tang, Peng Chen, and Ming Zhou. 2017. "Question generation for question answering". In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Copenhagen, Denmark, 866–874. https://doi.org/10.18653/v1/D17-1090

[13] Vrindavan Harrison and Marilyn Walker. 2018. "Neural generation of diverse questions using answer focus, contextual and linguistic features". In Proceedings of the 11th International Conference on Natural Language Generation. Association for Computational Linguistics. https://doi.org/10.18653/v1/w18-6536

[14] Michael Heilman and Noah A Smith. 2009. "Question generation via overgenerating transformations and ranking". Technical Report. CARNEGIE-MELLON UNIV PITTSBURGH PA LANGUAGE TECHNOLOGIES INST.

[15] V. Joshi, Matthew E. Peters, and Mark Hopkins. 2018. "Extending a Parser to Distant Domains Using a Few Dozen Partially Annotated Examples". In ACL.

[16] Payal Khullar, Konigari Rachna, Mukul Hase, and Manish Shrivastava. 2018. "Automatic question generation using relative pronouns and adverbs". In Proceedings of ACL 2018, Student Research Workshop. 153–158.

[17] Yanghoon Kim, Hwanhee Lee, Joongbo Shin, and Kyomin Jung. 2019. "Improving neural question generation using answer separation", In Association for the Advancement of Artificial Intelligence (AAAI). Proceedings of the AAAI Conference on Artificial Intelligence 33, 6602–6609. https://doi.org/10.1609/aaai.v33i01.33016602

[18] Kettip Kriangchaivech and Artit Wangperawong. 2019. "Question Generation by Transformers". https://doi.org/10.48550/ARXIV.1909.05017

[19] Igor Labutov, Sumit Basu, and Lucy Vanderwende. 2015. "Deep Questions without Deep Understanding". In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, Beijing, China, 889–898. https://doi.org/10.3115/v1/P15-1086

[20] Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. 2017. "RACE: Large-scale ReAding Comprehension Dataset From Examinations". arXiv preprint arXiv:1704.04683 (2017).