Enhancing Video Understanding: NLP-Based Automatic Question Generation

JAGAN G, AKILA K

Department of Computer Science and Engineering,
SRM Institute of Science and Technology,
Vadapalani, Chennai, Tamil Nadu, India

Abstract - The goal of the abstract project is to improve video comprehension using an autonomous question creation system that is NLP-based. The system creates pertinent and contextually appropriate inquiries based on the content of a particular video using NLP (Natural Language Processing) techniques. This procedure entails transcribing the audio from the film, identifying the main ideas, and creating inquiries that speak to various facets of the material. The produced questions can help students interact with the content of the video more deeply, foster critical thinking, and act as a tool for summarizing the topic. This study demonstrates the potential of fusing NLP with video content to offer a more dynamic and thorough learning experience by bridging the gap among visual knowledge and textual comprehension. An ongoing research trend in natural language processing is automatic question generation (AQG). AQG is very beneficial for computer-assisted assessments since it lowers the cost of manually creating questions and meets the demand for a steady stream of fresh questions. The majority of exam-style questions produced by automated question generation are of the "WH" ("What," "Who," and "Where" types) or reading comprehension variety. Based on their degrees of assessment, the questions must be varied or semantically distinct, although the responses may stay the same, in order to be as natural or human-like as possible. As a result, creating different sequences as part of question production has emerged as a crucial NLP activity, particularly in the publishing and education sectors.

Keywords - NLP, AQG, Extract Audio, STT.

I. INTRODUCTION

Natural Language Processing, a field of study that has been active over the past few decades, is responsible for the task of question generation. It seeks to produce inquiries from a given passage of text when the passage itself contains the answers to the Questions. There are many potential and beneficial uses for automatic question generation across several industries. A significant area of use for AQG is within the publishing and education sectors, where it has the potential to significantly improve tasks for both teachers and their pupils [5]. Not only that, but AQG can also support conversations between agents or Chabot, assist with intelligent teaching, and even support self-learning tutorials by generating questions for testing and training purposes.

Our primary goal in this research is to create question pairs for educational purposes. We are motivated by the fact that manually creating questions is quite tedious and time-consuming. Additionally, based on the evaluation levels, students need to be exposed to inquiries for a certain topic that range in difficulty [7]. For this reason, numerous studies have been conducted by researchers to produce a variety of queries that are semantically distinct from one another from a single source. For instance, handled a scenario of this nature using an encoder-decoder approach. This model is one of the most well-liked machine learning models and has attained an accuracy that, in some situations, even exceeds that of humans.

When creating a variety of questions with the same response for the creation of various target sequences, Cho et al. suggested focused content selection in addition to a traditional encoder-decoder architecture. They employ the Stanford Question Dataset for training, testing, and data validation, along with several other AQG studies [10]. However, when trying to apply these methodologies in real-world circumstances, the inference-drawing process is not as straightforward as it may be. The data must always be in a particular format to be used with Squad or any other common dataset. For instance, the training set for Squad contains a few words, a question, and the appropriate response. This restriction is addressed by our model [4]. Our model's architecture is built to automatically generate inquiries from any text minus the need for manual intervention, making the process of testing and assessing the output fluid and approachable.

II. LITERATURE REVIEW

Ragasudha, I., & Saravanan, M. (2021, March) et al. Examination is important everywhere in the world. Exams are the basic approach used to gauge a student's knowledge and aptitude. A teacher is required to develop sets of exam questions in accordance with the institutions or universities. The preparation of the question paper is difficult for the teacher and requires a lot of their time.

Despite great attempts, mistakes are occasionally made by people. The taxonomy developed by Bloom and random package suggested for this project make the process of creating test questions simple. It covers the six hierarchy levels of Bloom's classification system in order to produce a question paper of the highest caliber. Users can quickly add questions to the database, which will break questions into sections. The question will generate itself in the format of a PDF and be sent to the teacher after you click the "Generate" button.

Nwafor, C. A., & Onyenwe, I. E. (2021) et al. The challenge of creating multiple-choice questions automatically is useful yet difficult in the field of Natural Language Processing (NLP). The objective is to automatically generate accurate and pertinent questions from textual input. Despite its value, manually generating substantial, significant, and applicable questions is a taxing and difficult effort for teachers. In this research, we provide an autonomous MCQG for Computer-Based Tests Evaluation (CBTE) system based on NLP. We extracted keywords—important words that are included in the instructional material—using the NLP technique. Five lesson materials were utilized to evaluate the system's effectiveness and efficiency in order to confirm that it is not perverse. The results demonstrate that the program was capable of generating keywords from instructional materials in setting visible questions when the traditionally extracted keywords by an instructor were compared to the automatically generated keywords. For ease of access, this result is displayed in a user-friendly manner.

Gangopadhyay, S., & Ravikiran, S. M. (2020, September) et al. An ongoing research field in processing natural languages is Automatic Questions Answer Generation (AQAG). AQAG is very beneficial for computer-assisted assessments since it lowers the cost of manually creating questions and meets the demand for a steady stream of fresh questions. Automatic Question Generation typically produces "WH" (What, Who, and Where) or reading comprehension-style questions for exams. Based on their degrees of assessment, the questions must be varied or semantically distinct, although the responses may stay the same, in order to be as natural or human-like as possible. As a result, creating different sequences as part of query generation has emerged as a crucial NLP activity, particularly in the publishing and education sectors. In a conventional "encoder-decoder" paradigm, this module directs the decoder to create questions depending on particular emphasis contents. To create solution tags and a collection of candidates concentration from which the three greatest focuses are selected based on the amount of information they include, we employ a keyword generating algorithm. After that, we create queries that are semantically distinct from one another using the focus content. Our research employs a straightforward design with a single "Set your sights Generator" module, and experimental findings reveal that our module achieves 1.2% BLEU4 score gains while requiring 20% less training time than the most recent state-of-the-art model. Our model is simple to use and makes drawing inferences simple.

Lyu, C., Shang, L. (2021) et al. The goal of Question Generating (QG) is to create a believable inquiry for a given pair. While supervised QG trains a system to create questions given passages and answers, template-based QG employs existing Query Addressing (QA) datasets to convert declarative statements into interrogatives. The generated questions for the heuristic technique are closely related to their declarative counterparts, which is a drawback. The supervised approach's dependence on the language and domain of the QA sample used as training data is a drawback. We suggest an unsupervised QG approach that employs questions produced heuristically from abstracts as an avenue of training information for a QG system in order to get around these drawbacks. In order to train a from end to end neural QG model, the resultant queries are then integrated with previous news items. Utilizing unsupervised QA, we extrinsically assess our strategy. Synthetic QA pairings are produced using our QG model in order to train a QA model.

Cho, J., Seo, M. (2019) et al. In many NLP applications, such as question formulation or summarization, where the source and destination sequences exhibit semantic one-to-many links, it is crucial to provide diverse sequences. With the use of a general plug-and-play module (named SELECTOR) that wraps over and directs an existing encoder-decoder model, we provide a technique to explicitly decouple diversification from generation. In the diversification stage, a variety of specialists sample various binary masks on the original sequence in order to pick diverse content. Given every piece of chosen content from the original sequence, the creation stage employs a conventional encoder-decoder paradigm. We use a proxy for the basic truth mask and use stochastic hard-EM to teach because discrete sampling is non-differentiable and binary mask lacks ground truth labels.

III. RESEARCH METHODOLOGY

The suggested system intends to integrate cutting-edge technology to create a sophisticated video understanding platform. Modern computer vision techniques will be used by the system to evaluate video information and extract important visual elements, objects, and scenes. While simultaneously collecting the semantic meaning of the audio, methods involving Natural Language Processing, or NLP, will be used to transcribe spoken content. A comprehensive knowledge of the footage context will be provided by combining these retrieved visual and audio characteristics. An automatic question-generation module will create a wide range of inquiries on the video content based on this framework. These inquiries will cover a range of difficulty levels and focal points, encouraging users to interact more deeply and critically. With the ability to provide more engaging, individualized, and thorough educational experiences, the suggested system has the power to change video-based learning.

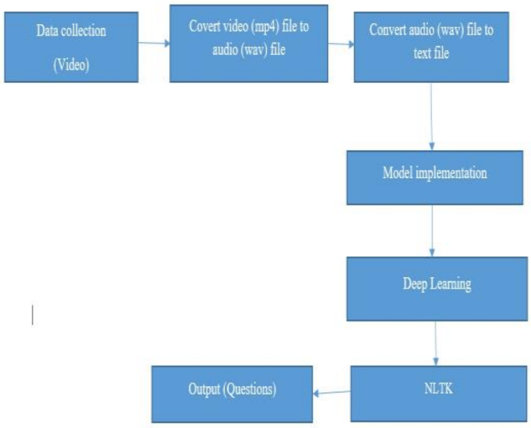


Figure 1 - Methodology

3.1. Video Analysis Module:

The video information in this module is analyzed using computer vision algorithms. It takes out visual elements including experiences, objects, motion, etc. spatial relationships [16]. To provide an extensive illustration of the visual context, sophisticated algorithms are required to locate and track pertinent features within the video frames.

3.2. Audio Transcription Module:

The spoken words in the video are converted into text using this module. Technologies called Automatic Speech Recognition, or ASR, capture the cultural significance of the uttered words as they are being converted from audio to text. The visual analysis [14] is complemented by the textual depiction, which improves the overall comprehension of the material.

3.3. Content Integration:

The previous modules' verbal and visual representations are combined to create a thorough knowledge of the context of the film [17]. The system obtains a stronger understanding of the connections between the visual features and the associated spoken content by coordinating the visual and aural signals.

3.4. Question Generation Module:

This module creates a variety of questions that depend on the integrated comprehension of the movie by utilizing NLP techniques. The questions range from factual queries [13] to higher-order thinking prompts, covering several cognitive levels. These inquiries are intended to promote analysis, participation, and enjoyment of the multimedia content.

3.5. Question Variability Enhancement:

This module makes use of strategies like paraphrase, rephrasing, and randomization to guarantee the quality and variety of questions [11]. This leads to a wider range of questions that investigate different facets of the multimedia content and accommodate varied learning preferences.

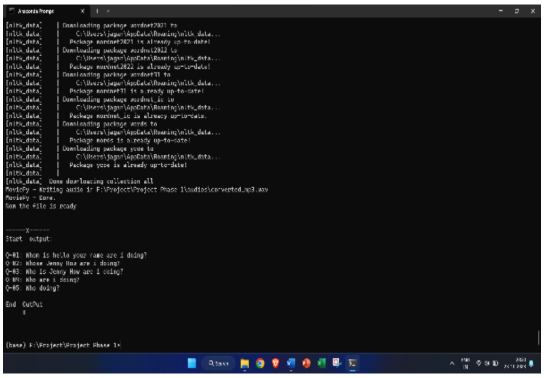
3.6. Model Implementation:

The goal of the artificial intelligence area known as "natural language processing" (NLP) is to make it possible for computers to comprehend, analyze, and produce human language. To enable robots to extract meaning, attitudes, and associations from language that is written or spoken, it entails creating mathematical models and algorithms that process and interpret text and speech data. Language translation, sentiment assessment, summarization of texts, speech recognition, and autonomous question production are just a few of the diverse activities covered by NLP. NLP plays a crucial role in technologies ranging from artificial intelligence and chatbots to text analysis and communication-driven automation by overcoming the gap between human conversation and computer understanding.

IV. RESULTS

The designed video understanding system has been put into practice with encouraging findings that show it has the ability to completely change how people interact with video information. A thorough representation of the movie's context was made possible by the video analysis module's successful extraction of pertinent visual elements. A thorough comprehension of the subject was made possible by the audio transcription module's accurate transcription of spoken materials. The system was better able to grasp the complex links across the two modalities thanks to the combination of visual plus textual signals. The question creation module demonstrated the capacity to generate a wide variety of inquiries that covered various levels of cognition and aspects of the movie's content. Due to the diversity, users were more engaged and encouraged to think critically. The strategies used to increase question variability expanded the pool of questions that were generated, resulting in a more thorough investigation of the issue. The user interface allowed for smooth communication and gave consumers a simple way to view video content, participate with questions, and get feedback. In order to promote the best understanding and skill development, the customized and adaptive education module successfully adapted the learning experience for distinct proficiency levels.

Additionally, the data collection and reporting part provided insightful information about user habits and educational outcomes, allowing teachers to improve techniques and content. The platform's integration and scaling capabilities demonstrated how easily it could be adapted to different educational environments and content delivery frameworks.



V. DISCUSSIONS

The potential repercussions, difficulties, and benefits of the suggested video understanding system are the main topics of discussion. The use of NLP and computer vision to improve video perception and engagement has considerable educational and non-educational potential. The system promotes intellectual inquiry and active engagement with video information by automatically creating contextually appropriate questions, leading to deeper learning experiences. But there are a few problems worth taking into account. It is essential to ensure the accuracy of computerized voice transcription because errors can result in incorrect comprehension. Additionally, maintaining user involvement without causing annoyance requires finding the ideal balance between question difficulty and user proficiency. Validating the variability augmentation module's capacity to produce interesting and useful queries is also important.

Positively, the system's mixture of textual and visual information provides a thorough grasp of video content and accommodates various learning preferences. The possibility for individualized content distribution and adaptive learning caters to the needs of individual students, increasing the overall efficiency of educational systems. The insights from the analytics component can help with data-driven enhancements in content development and user experience. Practically speaking, the system's wider acceptance would depend on its scalability and interoperability with current educational platforms. Additionally, privacy and security concerns for data are essential, particularly if the system engages with user-generated material.

VI. CONCLUSION

As a result, the proposed video understanding system has a great deal of potential to transform how we interact with video content for pedagogical and informational purposes. This system has the ability to close the gap across visual information and textual interpretation by seamlessly merging cutting-edge machine learning and Natural Language Processing, or NLP, methods. A key component of this system, the automatic question creation module, is an effective tool to promote engagement in learning, intellectual curiosity, and a deeper comprehension of video content.

Through its adaptive learning characteristics and capacity to extract and blend visual and aural signals, the system makes a substantial advancement towards individualized and interesting instructional experiences. This cutting-edge method has the potential to improve video-based learning as the digital world changes, making it more engaging, thorough, and available to a wide spectrum of learners. The suggested video understanding system creates new opportunities for dynamic and successful learning by fusing education and technology, meeting the varied needs of students in the digital age.

VII. REFERENCES

[1] Ragasudha, I., & Saravanan, M. (2021, March). Secure Automatic Question Paper with Reconfigurable Constraints. In 2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (Wisp NET) (pp. 16-20). IEEE.

[2] Nwafor, C. A., & Onyenwe, I. E. (2021). An automated multiple-choice question generation using natural language processing techniques. arXiv preprint arXiv:2103.14757.

[3] Gangopadhyay, S., & Ravikiran, S. M. (2020, September). Focused questions and answer generation by key content selection. In 2020 IEEE Sixth International Conference on Multimedia Big Data (Big MM) (pp. 45-53). IEEE.

[4] Lyu, C., Shang, L., Graham, Y., Foster, J., Jiang, X., & Liu, Q. (2021). Improving unsupervised question answering via summarization-informed question generation. arXiv preprint arXiv:2109.07954.

[5] Cho, J., Seo, M., & Hajishirzi, H. (2019). Mixture content selection for diverse sequence generation. arXiv preprint arXiv:1909.01953.

[6] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "Squad: 100,000+ questions for machine comprehension of text," arXiv preprint arXiv:1606.05250, 2016.

[7] A. Fan, M. Lewis, and Y. Dauphin, "Hierarchical neural story generation," arXiv preprint arXiv:1805.04833, 2018.

[8] X. He, G. Haffari, and M. Norouzi, "Sequence to sequence mixture model for diverse machine translation," in Proceedings of the 22nd Conference on Computational Natural Language Learning. Brussels, Belgium: Association for Computational Linguistics, Oct. 2018, pp. 583–592. [Online]. Available: https: //www.aclweb.org/anthology/K18-1056

[9] S. Subramanian, T. Wang, X. Yuan, S. Zhang, Y. Bengio, and A. Trischler, "Neural models for key phrase detection and question generation," arXiv preprint arXiv:1706.04560, 2017.

[10] J. H. Wolfe, "Automatic question generation from text an aid to independent study," in Proceedings of the ACM SIGCSE-SIGCUE technical symposium on Computer science and education, 1976, pp. 104–112.

[11] H. Kunichika, T. Katayama, T. Hirashima, and A. Takeuchi, "Automated question generation methods for intelligent English learning systems and its evaluation," in Proc. of ICCE, 2004

[12] R. Mitkov, H. Le An, and N. Karamanis, "A computer aided environment for generating multiple-choice test items," Natural language engineering, vol. 12, no. 2, pp. 177–194, 2006.

[13] V. Rus, Z. Cai, and A. C. Graesser, "Evaluation in natural language generation: The question generation task," in Proceedings of the Workshop on Shared Tasks and Comparative Evaluation in Natural Language Generation, 2007, pp. 20–21.

[14] I. Labutov, S. Basu, and L. Vanderwende, "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), 2015, pp. 889–898.

[15] I. V. Serban, A. Garc´ıa-Duran, C. Gulcehre, S. Ahn, ´ S. Chandar, A. Courville, and Y. Bengio, "Generating factoid questions with recurrent neural networks: The 30m factoid question-answer corpus," arXiv preprint arXiv:1603.06807, 2016.

[16] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in Advances in neural information processing systems, 2014, pp. 3104–3112.

[17] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014.

[18] O. Vinyals, M. Fortunato, and N. Jaitly, "Pointer networks," in Advances in neural information processing systems, 2015, pp. 2692–2700.

[19] C. Gulcehre, S. Ahn, R. Nallapati, B. Zhou, and Y. Bengio, "Pointing the unknown words," arXiv preprint arXiv:1603.08148, 2016.

[20] Q. Zhou, N. Yang, F. Wei, C. Tan, H. Bao, and M. Zhou, "Neural question generation from text: A preliminary study," in National CCF Conference on Natural Language Processing and Chinese Computing. Springer, 2017, pp. 662–671.