Extractive Summarization and Multiple Choice Question Generation using XLNet

Dhanamjaya Pochiraju
Department of Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
dhanamjaysri@gmail.com

Abhinav Chakilam
Department of Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
abhinavchakilam24@gmail.com

Premchand Betham
Department of Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
premchandbetham@gmail.com

Pranav Chimulla
Department of Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
pranavchimmula@gmail.com

Dr. Govinda Rao S
Department of Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad, India
govind.griet@gmail.com

***Abstract*—Automated multiple-choice question (MCQ) generation is a rapidly growing field in natural language processing (NLP) that aims to assist educators and trainers in creating high-quality, efficient, and effective assessment materials. This is achieved by analyzing large amounts of textual data, such as educational content, and identifying key concepts and relationships between them. The process of generating MCQs automatically can be broken down into several steps, including text summarization, keyword extraction, distractor generation, and sentence mapping. A new approach is proposed for Text summarization is based on NLP models like XLNet and YAKE is used for keyword extraction. Distractor generation is done using lexical databases such as ConceptNet and WordNet. Sentence mapping is used to identify the main concepts and relationships within a text, which can then be used to formulate questions and options for the multiple-choice questions. The output is a set of MCQs that are semantically related to the input text and can be used for educational and training purposes.**

Keywords*—Semantically related, text Summarization, XLNET, BERT, Sentence Mapping, Distractors, Keywords, ConceptNet, WordNet

I. INTRODUCTION

Multiple Choice Questions (MCQs) are a type of assessment in which a question or statement is presented along with several options, and the student or test-taker is required to select the correct answer from the options provided. MCQs are widely used in education, professional certification, and standardized testing as they are a quick and efficient way to assess a student's knowledge or understanding of a particular subject. MCQs typically consist of a stem, which is the question or statement, and several options, of which only one is the correct answer. The options can be in the form of single words, phrases, or complete sentences. MCQs can be used to test a variety of subjects, including science, mathematics, history, literature, and more. MCQs are popular in education and training because they are relatively easy to grade and can be used to quickly assess a student's understanding of a subject. They are also efficient in terms of time and resources, as they can be administered and graded quickly. However, MCQs also have some limitations. They can be criticized for not testing higher-order thinking skills, such as analysis and evaluation, and for being too easy to cheat on. Also, MCQs can be limited in their ability to test complex or nuanced knowledge, as they often only have one correct answer, and the options are usually designed to distract the test-taker. This is where the field of Automatic MCQ generation comes in, as it aims to overcome these limitations by generating high-quality, semantically-related, and diverse options to the given question, based on the input text. The input must consist of huge amount of raw text which must be about a page.

II. LITERATURE OVERVIEW

Pritam Kumar Mehta, Prachi Jain, Chetan Makwana, Dr. C M Raut have proposed an architecture "Automated MCQ Generator using Natural Language Processing" [1]. In this research, a model that employs BERT (Bidirectional Encoder Representations from Transformers) as a summarizer is constructed. Each phase of the MCQ creation process and the architecture were thoroughly discussed. There was a detailed explanation of the keyword extraction and distractor generating processes. The mathematical model for deep learning was explained thoroughly.

Santhanavijayan et al. have proposed a system of "Automatic generation of multiple-choice questions for e-assessment" [2]. They've created MCQs for their proposed system using an ontology-based method and firefly-based preference learning. In order to make it possible to compose questions, they employed a web corpus. Hypernyms and hyponyms, two similarity measures, are used to generate the distractors. To assess the pupils' language skills, the system also generates analogy questions.

Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Mário Jorge, Célia Nunes and Adam Jatowt proposed a methodology "YAKE! Collection-Independent Automatic Keyword Extractor" [3]. The research deconstructs the mathematical concepts and architectural frameworks employed in keyword generation. The five characteristics that define a word have been discussed. This document also provides an explanation of YAKE's operation. The above reference has helped understand the logic behind the keyword generation and the unique way of generation of keywords.

Zhilin Yang, Zihang Dai, Yiming Yanz, Jaime Carbonell, Ruslan Salakhutdinov, Quoc V Le explained in their research "XLNet: Generalized Autoregressive Pretraining for Language Understanding" [4]. By maximizing the expected likelihood across all permutations of the factorization order, the generalized autoregressive pretraining approach XLNet

(1) permits learning bidirectional contexts and (2) circumvents the restrictions of BERT. Additionally, Transformer-XL, a cutting-edge autoregressive model, is incorporated into pretraining by XLNet. Empirically, on 20 tasks, including question answering, natural language inference, sentiment analysis, and document rating, XLNet outperforms BERT, frequently by a significant margin, in similar experiment settings.

Ayako Hoshino and Hiroshi Nakagawa "A real-time multiple-choice question generation for language testing: A preliminary study" [5] is built on machine learning to automatically generate questions. They use machine learning algorithms like Naive Bayes and KNearest Neighbors to extract vocabulary and grammatical problems in English from internet news articles. They created a program that can convert user input from an HTML file into a quiz session.

III. Existing method

As far as existing methods are concerned, we have a traditional method for question generation involving the manual scraping of textbooks and the manual creation of distractors based on keywords. However, these methods have become increasingly cumbersome and inefficient in the digital age. In order to address this issue, there is a growing need for automation in this process.

BERT (Bidirectional Encoder Representations from Transformers), a transformer model that tackles the shortcomings of conventional neural networks in handling long-term interdependence, is one method for automation. BERT is a model that has already been trained and may be customized for particular summarizing jobs in NLP. RAKE (Rapid Automatic Keyword Extraction), a Python module, is used to extract keywords, and the WordNet technique is used to generate distractor terms.

In our proposed system, we aim to improve upon the current automated methods by utilizing alternative libraries to achieve better results.

IV. Proposed method

This method has been built upon the existing architecture. The first step involves summarization of the data that will be given as an input by the user. Keywords are extracted from the summarized text and are simultaneously mapped to the sentences they have been extracted from. Distractors are generated based on the keywords.

The system that automatically creates questions is discussed in this paper. With the aid of NLP, questions are automatically generated using the Automatic MCQ Generator. The system receives any domain's text as input, which is then condensed using the XLNet technique. Since BERT predicts a missing word based on its context, XLNet predicts all the words in a phrase given their order. This is because BERT is a context-based language model. It has been demonstrated that XLNet performs at the cutting edge on a variety of natural language understanding tasks, including question answering and text classification. Using XLNet as a text summarizer in the context of automatic MCQ synthesis can be advantageous since it can extract the most crucial information from a given text, which can be utilized to create questions and alternatives for the MCQs.

Now the keywords are selected from the generated summarized text using YAKE (Yet Another Keyword Extractor) which is an unsupervised, automatic key-phrase extraction algorithm that is specifically designed to work with short, noisy text. It is based on a combination of feature-based and graph-based methods, and it does not require any training data or labeled examples.

This keyword is the correct answer of the MCQ, the next Phase is to generate Distractors. This is done by using ConceptNet and WordNet. These are linguistic Databases that help us in generating meaningful distractors for each MCQ. They contain information about the relationships between words, such as synonymy, antonymy, and hierarchies, which can be used to generate semantically-related distractors for MCQs. Using both ConceptNet and WordNet for generating distractors for multiple choice questions (MCQs) allows for a more comprehensive and diverse set of options.

V. Methodology

A. Input and Summarization

The initial step entails the loading of the data/input, which is utilized for the generation of MCQs by the user. The data may contain sentences that are trivial in nature, and therefore, incapable of being used for the generation of MCQs.

In order to address this, summarization is used to extract pertinent material from the data, which is then used to generate questions. For this, an upgraded variant of BERT called XLNet is used. XLNet is a generalized autoregressive model that uses "permutation language modelling" to incorporate bi-directional context. It overcomes the drawbacks of BERT while combining the benefits of auto-regressive models and bi-directional context modelling. In 20 tasks, including question answering, natural language inference, sentiment analysis, and document ranking, XLNet has outperformed BERT, frequently by a large margin. The model is used to extract crucial information from the given text.

B. XLNet vs BERT

Instead of being a completely different model, XLNet is similar to BERT. Yet it has a lot of potential and is quite promising. Generalized autoregressive pretraining is what XLNet is, to put it simply.

An example of an AR language model is one that predicts the following word based on the word's context. Yet, the context word in this instance is limited to only two directions: forward or backward.

Because it is challenging to train a fully auto-regressive model over a range of factorization orders, only a portion of the output tokens are used as targets in the pre-training of the XLNet network using the target mapping input.

Although XLNet employs a training method that builds on the standard autoregressive model, it is not one. It then allows the model to predict the token n+1 by using the last n tokens after permuting the tokens in the phrase. The sentence is fed into the model in the correct order because everything is done with masks, however XLNet employs a mask that hides the preceding tokens in a specific permutation of 1,..., sequence length rather than masking the first n tokens for n+1.

XLNet in terms of parameters can be divided into two future subgroups:

- XLNet English model is a twelve-layer model that has 768 hidden,12 heads,110m parameters.
- XLNet Large English model is a twenty-four-layer model that takes 1024 hidden,16 heads,340m parameters.

The maximum length of a token in BERT is 512. This cap was set in order to prevent output of poor quality. Transformers are autoregressive, and BERT's developers found that using documents longer than 512 tokens resulted in a considerable drop in performance. So, this cap was set to prevent the production of subpar work.

Contrary to BERT, which has a 512-token input limit, XLNET is one of the few models that has no sequence length limit, which is a significant benefit over BERT.

A simple example can be used to compare XLNet and BERT. For instance, consider the sentence 'San Jose is in Cali,' where the task is to predict the word 'San Jose.' The current permutation can be assumed as-

[is_1, in_2, $Cali_3$, San_4, $Jose_5$]

Tokens 4 and 5 are expected to be unrelated by ERT. As an auto-regressive model, XLNet on the other hand, predicts token 4 first, followed by token 5, in that order.

XLNet would compute-

logP(San| is in Cali) + logP(Jose|San, is in Cali)

BERT would compute-

logP(San| is in Cali) + logP(Jose| is in Cali)

C. *Keyword Extraction*

The next phase is utilizing YAKE (Yet Another Keyword Extractor), a simple unsupervised automatic keyword extraction technique based on text statistical data, to extract keywords from the condensed text. The answers to the queries created by these keywords. YAKE is independent of the text size, language, or domain and does not rely on any external corpus, dictionaries, or prior training on a particular set of documents.

Furthermore, because YAKE does not need training on a particular set of documents, it has been demonstrated to perform better than RAKE in terms of performance. YAKE employed 5 statistical techniques:

1. Preprocessing and Candidate Term Identification
2. Feature Extraction
3. Computing Term Score
4. Generating n-gram and computing keyword scores
5. Data Duplication and ranking

D. *Distractor Generation*

The erroneous answers to a multiple-choice question are referred to be distractions. We used both WordNet and ConceptNet to create these distractions. A massive lexical database called WordNet connects words into semantic relationships, such as synonyms, hyponyms, and meronyms. ConceptNet, on the other hand, is a freely accessible semantic network made to help computers understand the meanings of words spoken by people.

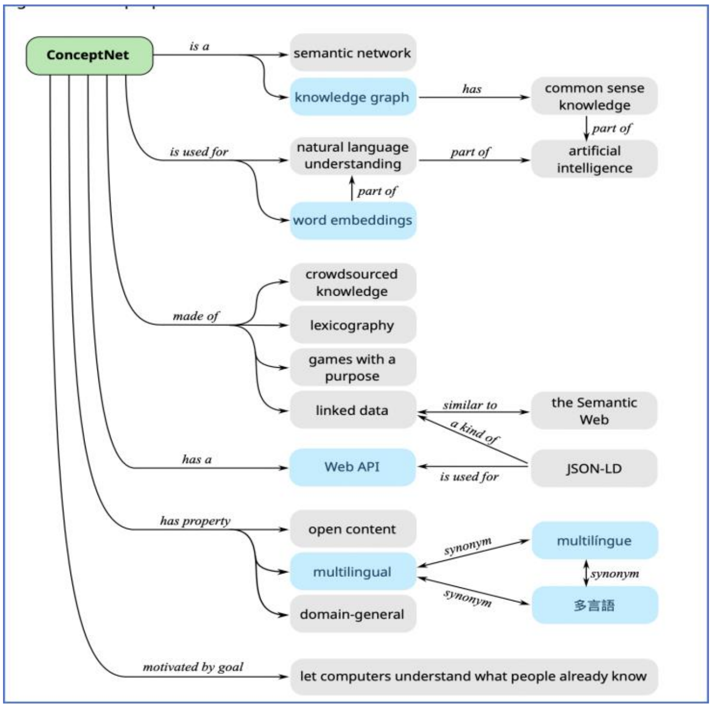

Fig. 1. ConceptNet breakdown [23]

VI. RESULTS ON DISCUSSION

When given a vast amount of text to work with, the proposed model's goal is to create a special model that, when used with the provided project, generates multiple-choice questions.
The following steps were used to implement the suggested model.

- Originally, a blank Jupyter notebook file was made, and all necessary dependencies, libraries, and modules were installed. Python was used to create this model.
- An XLNet model was built using the summarizer method from the Bert summarizer package and gave the huge text file as an input argument. We received a paragraph of the text that had been condensed after the XLNet model had finished its summary. The entire data set is condensed into a brief paragraph that contains crucial words and phrases.
- Using YAKE, we extract keywords from the input text and the summarized text in this stage. We link the statement to the keywords. The results are kept in a dictionary. Keyword extraction and sentence mapping is one of the key phases which help in the determining the quality of the questions generated.

- The user who creates generated MCQ's from this model has the option to choose some other distractors as well. The remaining words that can be used as distractors are mentioned in the array below each question.
- The next step here is to generate distractors which are the other keywords which must be given out as options to the multiple-choice questions generated. A completely new approach is used where both WordNet and ConceptNet both are used together for better results. A combination of two models is used here to increase the quality of the distractors making it a tough job for user to guess the answer.
- All the above-mentioned steps are simultaneously mapped and the end step is where the multiple-choice questions are generated and displayed to the user.

The MCQ generator was built and the results obtained were compared to the ones with the existing method. We have compared both the models by using the same raw text. The questions that are generated by the new model by using XLNet is way advanced than BERT model. The quality of the questions has been increased. The nature of questions has changed as the mathematical model behind each NLP model is different.

The quality of the distractors generated has improved significantly. There exists a huge correlation between the actual answer and the generated distractors.

The reason to choose XLNet is that XLNet has proven to have better results than BERT. The two-way context modeling and usage of semantically related words has led to better results and better results and better output.

The mode has improved significantly in each of the step used. By using YAKE the keyword extraction has enhanced. The usage of YAKE has been proved to be better as the keywords extracted are closer and have similar meanings. YAKE has a prebuilt library of over fifty million existing words which helps in better mapping. The below illustrated figure shows the comparison of all the other keyword extractors.

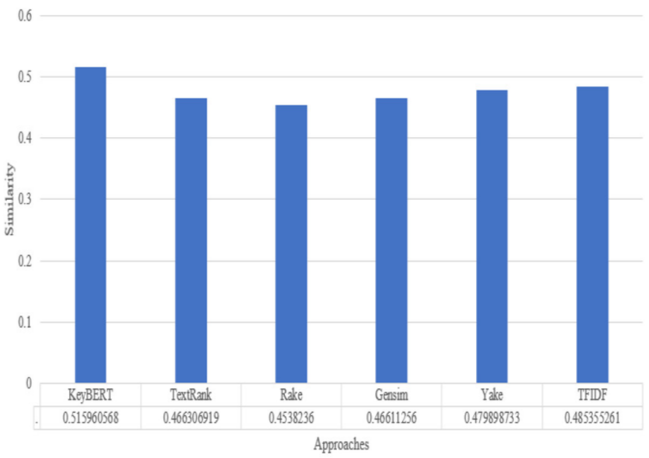

Fig. 2. XLNet vs BERT results [22]

VII. CONCLUSION

Multiple choice questions have gained immense popularity during the pandemic. All the educational institutions have moved from primitive mode of examination to an online mode. To reduce the burden of the educators a new method is developed which generates more sensible and intelligent questions which tests the skill of the test taker. The use of XLNet to summarize and generate MCQs has resulted in better quality of questions. However, the options that are generated by ConceptNet and WordNet can be improved. The distractors are generated from a pool of words available from the ConceptNet and WordNet library.

REFERENCES

[1] Pritam Kumar Mehta, Prachi Jain, Chetan Makwana and Dr. C M Raut "Automated MCQ Generator using Natural Language Processing" Internation Research Journal of Engineering and Technology, Vol:08, doi: 05 May 2021.

[2] Santhanavijayan, A., Balasundaram, S.R., Hari Narayanan, S., Vinod Kumar, S. and Vignesh Prasad, V. (2017) 'Automatic generation of multiple choice questions for e-assessment', Int. J. Signal and Imaging Systems Engineering, Vol. 10, Nos. 1/2, pp.54–62

[3] Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Mário Jorge, Célia Nunes and Adam Jatowt "YAKE! Collection-Independent Automatic Keyword Extractor",(LNISA,volume 10772) doi: 10.1007/978-3-319-76941-7_80

[4] Zhilin Yang, Zihang Dai, Yiming Yanz, Jaime Carbonell, Ruslan Salakhutdinov, Quoc V Le "XLNet: Generalized Autoregressive Pretraining for Language Understanding", arXiv:1909.12486v2 [cs.LG] 23 Dec 2019.

[5] Ayako Hoshino and Hiroshi Nakagawa "A real-time multiple-choice question generation for language testing: A preliminary study", doi: 10.5555/1609829.1609832.

[6] Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., Jatowt, A.: YAKE! collection-independent automatic keyword extractor. In: Pasi, G., Piwowarski, B., Azzopardi, L., Hanbury, A. (eds.) ECIR 2018, LNCS, vol. 10772, pp. 806–810. Springer, Cham (2018).

[7] D. R. CH and S. K. Saha, "Automatic Multiple Choice Question Generation From Text: A Survey," IEEE Transactions on Learning Technologies, vol. 13, no. 1, pp. 14-25, 1 Jan.-March 2020, doi: 10.1109/TLT.2018.2889100.

[8] AY. Satria and T. Tokunaga, "Automatic generation of English reference question by utilising nonrestrictive relative clause", Proc. 9th Int. Conf. Comput. Supported Edu., pp. 379386, 2017.

[9] Yang Liu,"Fine-tune BERT for Extractive Summarization", arXiv:1903.10318v2 [cs.CL] 5 Sep 2019

[10] Mihalcea, Rada. (2004). Graph-based Ranking Algorithms for Sentence Extraction, Applied to Text Summarization. 170-173. 10.3115/1219044.1219064.

[11] Gowon, Rahila & Yashim, Haddy. (2022). Effects of concept mapping strategy on sentence construction, paragraphing and editing in junior secondary schools students' composition writing achievements. Frontiers in Education. 7. 10.3389/feduc.2022.973844.

[12] Lan, Zhenzhong & Chen, Mingda & Goodman, Sebastian & Gimpel, Kevin & Sharma, Piyush & Soricut, Radu & Research, Google & de Carvalho, Mailton. (2019). ALBaGERT: A LITE BERT FOR SELF-SUPERVISED LEARNING OF LANGUAGE REPRESENTATIONS.

[13] Gorewar, Prof. (2021) "MCQ Generator using Machine Learning" International Journal for Research in Applied Science and Engineering Technology. 9. 4360-4362. 10.22214/ijraset.2021.35953.

[14] Hidayati, Nurul & Rosita, Rita. "Exploring new application of MCQ (multiple-choice question) assessment method in anatomy practical exam from students' satisfaction"

[15] Le, Nguyen-Thinh & Kojiri, Tomoko & Pinkwart, Niels. (2014). Automatic Question Generation for Educational Applications – The State of Art. Advances in Intelligent Systems and Computing. 282. 325-338. 10.1007/978-3-319-06569-4_24.

[16] Rao, Dhawaleswar & Saha, Sujan. (2022). Generation of Multiple-Choice Questions From Textbook Contents of School-Level Subjects. IEEE Transactions on Learning Technologies. PP. doi:1-13. 10.1109/TLT.2022.3224232.

[17] Aldabe, Itziar & Maritxalar, Montse & Martinez, Edurne. (2008). Evaluating and Improving the Distractor-Generating Heuristics.

[18] R. Mitkov, L. Ha, and N. Karamis. A computer-aided environment for generating multiple-choice test items. Natural Language Engineering: Special Issue on using NLP for Educational Applications, 12(2):177-194, 2006.

[19] Campos, Ricardo & Mangaravite, Vítor & Pasquali, Arian & Jorge, Alípio & Nunes, Célia & Jatowt, Adam. (2018). “A Text Feature Based Automatic Keyword Extraction Method for Single Documents”. doi: 10.1007/978-3-319-76941-7_63.

[20] Liu, Hugo & Singh, Push. (2004). “ConceptNet—A Practical Commonsense Reasoning Tool-Kit”. BT technology journal. 22. 10.1023/B:BTTJ.0000047600.45421.6d.

[21] Yunmar, Rajif & Setiawan, Andika & Tantriawan, Hartanto. (2020). “The Combination of YAKE and Language Processing for Unsupervised Term Extraction Ontology Learning”. IOP Conference Series: Earth and Environmental Science. 537. 012023. 10.1088/1755-1315/537/1/012023

[22] https://peerj.com/articles/cs-967.pdf

[23] https://www.cnblogs.com/TheTai/p/14393737.html