Journal of Information Systems Engineering and Management
2025, 10(1s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Dynamic Question Paper Generation using Bloom and NLP based Heterogeneous Feature Extraction and Machine Learning Techniques

Kanchan Babaji Dhomse[1], Dr. Sandhya Sharma[2]

[1]*Research Scholars, SGVU, Rajsthan, India.*

[2]*Professor of SGVU, Rajsthan, India.*

[1]*kdhomse@gmail.com,* [2]*sandhya.sharma@mygyanvihar.com*

ARTICLE INFO

Received: 04 Oct 2024

Revised: 29 Nov 2024

Accepted: 18 Dec 2024

ABSTRACT

Automatic question generation is a very difficult task in the field of Natural Language Processing (NLP) and Machine Learning (ML). The procedure requires bidirectional language processing, including both Natural Language Understanding to comprehend the input text and Natural Language Generation to produce queries in textual form. This article presents our system for generating factual inquiries from unstructured English text. The strategy used combines conventional linguistic methodologies that rely on sentence structures with several machine learning techniques. Initially, we extract lexical, syntactic, and semantic data from the input text. Subsequently, we create a hierarchical collection of patterns for each phrase. The characteristics are derived from the patterns and then used for the automatic acquisition of new transformation rules. The learning technique we use is entirely data-driven, since the transformation rules are derived from a collection of initial sentence-question pairings. The benefits of this technique are twofold. Firstly, it allows for the straightforward addition of additional transformation rules, enabling the generation of many sorts of questions. Secondly, the system may be continuously enhanced via bloom algorithm with collaboration of NLP and machine learning. The framework furthermore has a question assessment module that assesses the calibre of created questions. It functions as a filter to discern the most optimal inquiries while removing erroneous or duplicate ones. We conducted many tests to assess the accuracy of the produced questions and also compared our system with other cutting-edge technologies. Our findings recommend that the generated queries surpass the performance of existing algorithms and are on par with questions produced by people. In addition, we have developed and released an interface that includes all the datasets we have produced and the questions we have analysed.

Keywords: bloom filter, NLP, Name entity recognition, question generation model, machine learning, bag of words, feature extraction, classification.

Introduction

In the realm of education, the creation of question papers is a critical and often labour-intensive task for educators. Traditional methods of question paper generation involve manually reading through textbooks, lecture notes, and other educational materials to formulate relevant and challenging questions. This process is not only time-consuming but also prone to human error and bias. With advancements in technology, particularly in the fields of Machine Learning (ML) and Natural Language Processing (NLP), there is a growing potential to automate this process, thereby enhancing efficiency and accuracy.

Automatic Question Paper Generation leverages the capabilities of ML and NLP to analyze input documents and generate a variety of questions based on the content. This system aims to transform the manual process into an automated one, ensuring that the generated questions are relevant, diverse, and aligned with the educational objectives. The examination and organization of address papers is the method that is used to carry out an evaluation of an understudy, which may be an essential component of the guideline. The display day innovations make it easier

for the instructor to save the questions in a computer database. However, the problem that has arisen is how the display day innovations would also assist the instructor in creating a variety of sets of questions without causing the instructor to worry about virtually replicating and duplicating questions from previous examinations. One of the most important and essential ways to achieve achievement is by providing instruction. At the time that we are evaluating the training, it is imperative that we have a suitable examination paper and arrangement. At this point, the manual method of producing address paper has become the standard procedure. The creation of address paper has been offered by us as a robotized handling that is not only rapid but also beautiful, disorganized, and secure. We came up with this framework in order to facilitate efficient work on the production of programmed address sheets with less effort and with less control from the human intervention. A taking after algorithm is something that we have suggested. The Blooms algorithm is a scientific classification that may be a collection of three progressive models that are used to categorize educational learning objectives into different degrees of complexity and specificity. There are three recordings that address the learning objectives in the areas of cognitive, general sensation, and tactile development. Within the realms of computer science and operations research, the calculation of insect colony optimization has the potential to be a probabilistic approach to comprehending computing difficulties.

This approach can be reduced to numerous discovering fantastic approaches via charts. When we talk about manufactured ants, we are referring to multi-agent methods that are inspired by the behavior of real ants. Anything that can effectively reduce the amount of time spent on something and manage how much it is used is recognized and praised. In this manner, we are demonstrating a computerized address creator of paper framework that is capable of reducing the amount of time that is used in order to replace the conventional method with their address paper era framework. For all of its activities, our architecture is designed to provide rapid operations, a large capacity for information, and a high level of security. The architecture that has been developed is capable of supporting a wide variety of instructional setting and competitive test associated organizing systems. Because of this, the Address Paper Generator framework generates paper, plans, and document records in accordance with the paper format that is selected. You should send an email to all of the institutions. The integration of ML and NLP in the automatic generation of question papers holds immense potential to revolutionize the educational assessment process. By automating this task, educators can save time, reduce bias, and enhance the overall quality of assessments, ultimately contributing to a more effective and efficient educational system. The subsequent sections of this article are organized in the following manner:

The second section of our paper presents contemporary methodologies employed in question generation systems. We also propose a classification of cutting-edge systems based on multiple criteria and conduct an assessment of the questions generated by these systems, with a specific emphasis on factual questions derived from unstructured text. In section third, we provide our suggested framework for generating questions from unstructured text. This framework includes a comprehensive explanation of each step involved in the process, starting with dataset preparation to question production, and concludes with an assessment that involves exhaustive comparisons. In the fourth part, we present and elucidate our experiments and analyses the assessment of outcomes from many perspectives. In the concluding part, we provide a summary and put up recommendations for future endeavours.

Literature Survey

According to Dhomse KB and Ishan Ranjan [1] proposed an automatic question paper generation using various natural language process and machine learning techniques. This article briefly describes numerous techniques that used for data processing, feature extraction and module training as well as testing etc. Ramli R. et. al. [2] the purpose of this program is to provide opportunities for persons from different historical periods and geographical locations to interact, study, and collaborate with one another via the use of various instructional administrative services. This program is designed to incorporate language learners throughout the era. As a result of the fact that the existing system for question paper generation involves human staff members checking out questions that appear inside the address paper, which is a process that takes a significant amount of time, we presented a mechanized approach for address paper generation that is implemented as a real-time application in our system. A computerized framework is shown in the work that is being offered. This framework represents an advancement from the traditional strategy of the paper period to a mechanized method.

Shihua Xu and many more [3] Both in the field of IR and in the field of Characteristic Dialect Handling (NLP), one of the most popular topics is programmed address reply. In the past, the conventional address structure comprised of questions, comprehension, and responses that were somewhat complicated and included representation. For the purpose of identifying the center of address, it must be supporting. The question is classified with the help of this

device. Utilizing these frameworks allows for the classification of the various questions, which will be of great assistance when it comes to planning.

In addition to Juee Gosavi, [4] The unused level of data sharing is enabled by various online communications such as wikis, blogs, organize, and so on. These online communications provide a platform for interaction, which offers services such as searching and posting questions with the notoriety of websites such as Yahoo!, answers, cross-approved, stack flood, Quora, and so on. Currently, an increasing number of people are turning to these online meetings in order to get answers to the issues they have. This site is able to provide solutions to all inquiries; but, in order to organize, they were required to take examinations in order to verify the knowledge of the students. Since this is not possible with online gatherings, we developed our framework so that it offers the capability of producing programmed address paper for examinations.

Xingboxie, [5] The path of natural language processing (NLP) is used in this study to construct a framework for programmed address answering that is centered on philosophy. Through the process of evaluating the address of the client, the framework is able to extract watchwords and transform the purposeful intent of the address into the investigation of vital components in ontology. The asking address in natural language processing is supported by this programmed address answering framework, which also has a few systems in other areas.

Surbhi Choudhary [6] in this modern age, the e-book has become an essential need for the candidates to appear and prepare for their competitive examinations inside the confines of the college campus. In this study, we propose a display framework for the period of colleges that are characterized by the use of shrewd address papers. The tool that underpins this framework is the creation of a large number of random question papers in conjunction with the difficulty level of the address in terms of rate. In this framework, address papers are made according to the difficulty level and weight age of the students. However, this framework does not cover all questions; thus, in order to cover all types of questions, we offered our framework, which may be very accommodating for colleges.

This is Yoichi Matsuyama [7]. In this research, we suggested components for an agreeable conversational system that are based on the concept of programmed expressive conclusion sentence creation. Despite the fact that it is helpful for automatically creating phrases, this system is not suitable for educational purposes. In order to assist our educational society, we presented a new method that creates automated question papers.

Within the context of Ehsan Sherkat [8], the task of determining the exact response to an address written in common dialect is referred to as "programmed address replying." Within the domain of programmed address reply, classification is an extremely important component. Within the scope of this research, we have fundamentally presented the half-breed approaches that are used in a web-based automated question answering framework. This is helpful for locating the particular address of the respondent, which is beneficial for students to have in order to get the response that they want.

According to L. Bednarik and L. Kovács [9], they integrated a few specific instruments with the purpose of developing a model in order to generate questions as a result. Based on the premise of the clustering technique, the proposed method selects a few watchwords from the input information and forms on the other aspect of the data. The framework is responsible for the linguistic preparation of the whole yield structure in order to enable the creation of a computerized address paper.

The framework that was presented by N. Omar and colleagues [10] is now working on a few specified principles for the creation of the final yield. The natural language processing (NLP) system is responsible for managing the full handle, which includes the arrangement of verbs and a few really important watchwords taken from the database. According to this framework, the watchwords and other aspects of the yield paper are dependent on the kind of paper that is being used. When it comes to discovering the exact subject matter and the kind of output that the user requires, the framework that is used makes a difference.

Jun Araki et. al. [11] is aimed at locks in dialect, which are learners who produce multiple-choice questions (MCQs) Address, their particular, notably coreference determination and summary detection.Additionally, this framework offers the ability to modify distracters in response to their choices. Exam questions are an essential tool for teachers to use in order to evaluate the breadth and depth of their students' knowledge of fabric. Consequently, the computer-generated address generation from content has the potential to be an important feature language learning technology that may assist teachers in evaluating the reading comprehension of their students. A problem that addresses their greater portion of the question that was made, which was a single word, as well as a specific and less level of properly

degree understandings of their content in general. All of our examinations consist of a total of 200 questions derived from each framework, as well as three distracters for each address. In this paper, the

y generated what might be considered questions that were useful but were not as valuable as the automated generate address paper that we ought to generate with this that we have.

According to Tirath Prasad Sahu and colleagues [12], the notions of marking connected points to content on the websites are very widely recognized. This is due to the fact that it makes a difference in classifying and finding the content that is available on the websites. Despite the fact that this framework can, in a sense, give answers to questions, it is not more supportive for establishing or instructive reasons. In order to overcome this disadvantage, we developed our own framework that is capable of creating programmed address paper for all organizing. This paper presents that one client inquires the question from other clients who have the information to reply the question.

An open data extraction (OIE) may be a method of extracting significant data from a collection of unstructured information in a structured manner, according to Thenmozhi, D. et al. [13]. The architecture is flexible enough to accommodate the receiving of responses to any inquiry. In the future, they will provide a tool that will allow them to get rid of the incorrect responses that are also extracted without any purpose; semantics may be introduced in the future in order to extract correct answers for the questions. This framework is useful for obtaining responses to questions, but it has a significant drawback in that it allows questions to be repeated. This is a significant disadvantage, and in order to mitigate this drawback, we suggested our framework for the creation of programmed address paper, which avoids the repetition of questions.

Abderrazzak et. al. [14] this study exhibit the Address responding framework (QAS) that makes a difference learners to discover the leading solutions to their queries and makes a difference mentors to reply the address questioned by their understudies in an e-learning environment. This framework makes it possible to discover the answers to a question in their database; however, it has a drawback in that some questions can be repeated, and the process of finding those questions can be extremely time consuming. In order to overcome this drawback, we developed our framework for the creation of programmed address paper, which allows it to anticipate questions from being repeated or duplicated.

Zalte et al. [15] often difficult because of the growth that has taken place within the field of computer science (CS) and the increase in the need for confronting in the modern day; accordingly, analysis plays an important role in evaluating the performance of students. We have suggested a computerized handling of address paper period, which would be beneficial for institutions. Our framework, which also generates programmed address paper with dodges reiterations of questions and with more security useful for all institutes, was developed to overcome this issue. This is the major drawback of this system, so in order to overcome this drawback, we developed our system that also generates automatic question paper and it is useful for all institutes. There is a possibility that some questions will be repeated in multiple address papers because the teacher has a personal bias towards them. As a result, there is no guarantee that the address paper will be produced in an immaculate and arbitrary manner.

According to Nguyen-Thinh Le et al. [16], in recent times, analysts from a variety of fields have been presenting their shared interest in the field of programmed address technologies for the aim of educational objectives. We conduct a study of the current state of the art of several methods to the creation of instructional apps of address era in this paper. We have arrived at the conclusion that despite the fact that there is a vast array of approaches to the generation of programmed addresses. During this time period, a very small number of instructional frameworks have been developed. The goal of this study is to demonstrate how questions can be generated automatically from a variety of inputs and databases. However, the system can only produce questions and not a question paper. In order to address this limitation, we designed a system that can generate questions and question papers automatically for educational purposes.

According to Route [17] a result of the rapid development of data technology and the need of various groups of people for data that is both faster and more precise, the development of address responding technology has evolved. In this study, the primary focus is on illustrating the address classification technique of the Tibetan online programmed address answering framework. A look at the different types of addresses is shown in the context of various coordinating performances. If an address and a run the show coordinate for this address, and if this address has a position in the comparing reply category, then it will be extremely accommodating to push real information.

An example of a programmed paper era architecture that adhered to the B/S standard was shown by G. Cen et al. [18] in J2EE. It is compatible with a variety of administration modules, each of which provides aid in managing client

information and other administrations of the framework. Some of the components of the module include the ability to classify information, the capacity to deal with address information and administration client information, the option to choose between paper and address, and so on. There are a variety of situations that need inputs in order to manufacture address paper, and the suggested instrument makes a difference in planning and supervising these circumstances.

Wang bo [19] Within the realm of regular language training, the programmed address responding structure is becoming an increasingly popular topic of discussion, and it is also playing an important role in the teaching process. Within the scope of this article, an ontology-based programmed address responding framework section is proposed. Due to the fact that students and teachers are unable to converse face-to-face, it is necessary to organize learning and instruction via different arrangements. For the purpose of resolving this problem, an application of a framework for responding to programmed addresses has been developed.

In their study, T. Fei et al. [20] presented a framework that is capable of providing aid in supplying the essential address information for their paper base. A considerably better learning framework was developed as a result of the whole handle preparing their premise content categorization and giving output with back engendering training to their neural arrangement. This resulted in the creation of a learning framework that was significantly better significantly better higher more grounded. The computation of compound angles is employed in the development of their framework, and neural calculations are used to prepare this information. When it comes to the content need, their client information is utilized, and the framework extracts their highlights for the client inquiry. Additionally, the framework provides results that are dependent on the extraction of the stored highlights. The investigation suggests that there was a conspiracy during the age of high precision based address paper.

The authors Dhawaleswar Rao et al. [21] programmed a variety of multiple-choice questions (MCQ) to their material, which is included in the prevalent investigate range. Multiple-choice questions (MCQs) are widely considered to be a key evaluation that is comprised of a variety of spaces and constraints. To the survey, this system displays the findings that we have discovered. The graphic depicts a monotonous flow from the MCQ period structure to the uninteresting step-by-step approach. The similar approach has developed in [22]. The design of a programmed generating framework of examination papers that is based on the internet was developed by Liancheng Guan [23] in conjunction with the development of technological innovations in the field of organization. It is based on an examination address bank that is always being updated in accordance with the needs of the instructors, which may naturally result in the creation of the test paper with a certain degree of complexity involved in monitoring and maintaining the examination address.

Bindra, S. K. et al. [24] said that in the present scenario, the address paper period may be pushing for the unnecessary duplication to the times and safety bases. As a result, data mining had been very widespread in order to provide a facility in a variety of distinct fields. The database stores a vast quantity of information components in its storage space. According to Gauri Nalawade et.al. [25], the examinations are becoming an increasingly important component in the world of showcase private businesses for the purpose of evaluating the mental growth of their students. As a matter of fact, their quality students determine their examination based on the address for their particular education. Because of this framework, a wonderful address project is being created, and the goal pattern is being learned. We are interested in their vital teachers; nevertheless, there is a shortage of permitted instructors in the establishments and colleges. In order to overcome this disadvantage, we suggested our framework, which is capable of producing programmed address paper without the need for instructors and does not need any human efforts.

Xie et al. [26] created a model that utilizes TEEM for entity recognition. The model then used a deep structured semantic model to compute the semantic similarity between the query and predicate in the candidate knowledge triples. Liu et al. [27] introduced a technique that relies on entity sorting and joint fact selection. This approach employs a BiLSTM-CRF model to extract patterns, followed by similarity matching to rank the candidate entities. Finally, a joint fact selection model is used to improve the selection of the optimal entity relationship pair by multi-level coding. Luo et al. [28] introduced a relationship detection model that utilizes a multiangle attention mechanism. This mechanism extracts the connection between question patterns and potential relationships by using the attention mechanism at both the word level and relationship level. The response generator utilizes the BERT-word frequency semantic feature pointer generation network paradigm to achieve answer creation. During the word vectors acquisition step, the semantic properties of BERT and word frequency are combined together as the input sequence for the response generator. Furthermore, during the word vector acquisition phase, the impact of including Term

Frequency-Inverse Document Frequency (TF-IDF) [29] on the resulting output is taken into account. To assess the efficacy of our approach, a comparison experiment was conducted.

Research Methodology

Figure 1 illustrates the general implementation method of this research, which has three components: the construction module for knowing vocabulary, the module for acquiring word vectors, and the module for generating the generative model. In this generative model, the vocabulary of knowledge is used to choose things with a high likelihood in the vocabulary as the subsequent word to be created. The understanding vocabulary constructing module utilizes the bloom algorithm for named entity recognition. This algorithm identifies entities present in both the source text and question sentences. It also determines the relationship between these entities and the tail entities they refer to in the knowledge base. The module then counts the entities individually and calculates the frequency of each component in all triples associated with the entity. The word vectors acquisition module utilizes a pre-trained language model to extract the word vectors of the question phrase. These word vectors are then combined with the word frequency semantic attributes of the entity in the questioning sentence to form the input sequence. The generative model construction module introduces a suggested model to find the final answers to different WH type questions.

Bloom Model

The BLOOM (BigScience Large Open-science Open-access Multilingual) model is a state-of-the-art multilingual language model developed by the BigScience project. While specific "BLOOM feature extraction" is not a standardized term, we can discuss how to leverage the BLOOM model or similar transformer-based language models for feature extraction in NLP tasks. By leveraging the powerful capabilities of the BLOOM model, you can extract rich, contextualized features that enhance various NLP applications. Below is the procedure that we used for entire process.

Step 1 : Load the BLOOM Model and Tokenizer and Use the Hugging Face transformers library to load the BLOOM model and tokenizer.

tokenizer = AutoTokenizer.from_pretrained("bigscience/bloom")

model = AutoModel.from_pretrained("bigscience/bloom")

Step 2 : Tokenize the text to prepare it for the model.

text = "This is a sample sentence for feature extraction."

inputs = tokenizer(text, return_tensors="pt")

Step 3 : Pass the tokenized inputs through the model to obtain the hidden states.

*outputs = model(**inputs)*

hidden_states = outputs.last_hidden_state

Step 4 : Use the hidden states as features. Typically, the last hidden state or a pooled representation is used.

features = hidden_states[:, 0, :].detach().numpy()

Step 5 : Each word in the input text is represented by a dense vector that captures its contextual meaning.

word_embeddings = hidden_states.detach().numpy()

Step 6 : Then methods for aggregation include mean pooling, max pooling, or using the [CLS] token representation.

sentence_embedding = hidden_states.mean(dim=1).detach().numpy()

Step 7 : Use sentence embeddings as input features for classifiers such as any ML classifier

training_args = TrainingArguments(output_dir="./results", num_train_epochs=3)

trainer = Trainer(model=model, args=training_args, train_dataset=train_dataset, eval_dataset=eval_dataset)

trainer.train()

Step 8: end procedure

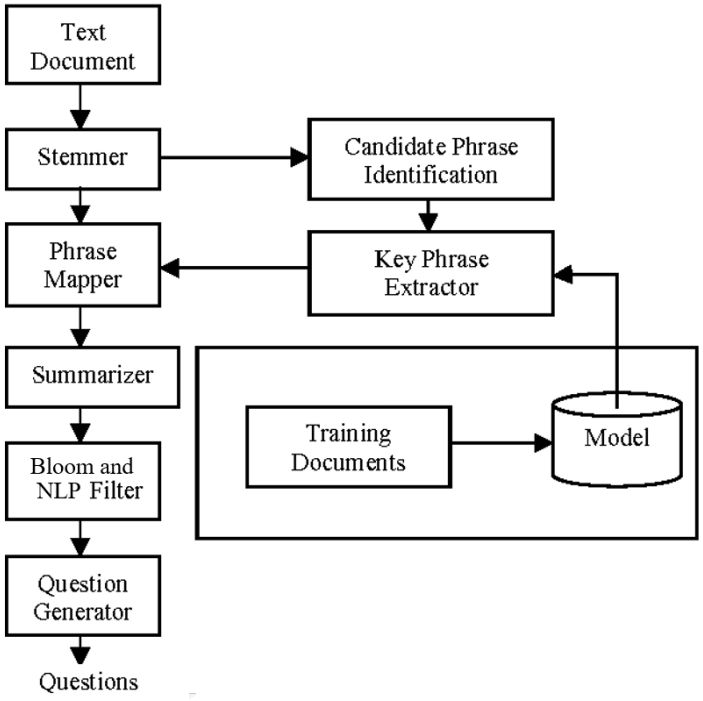

Figure 1: Dynamic question generation proposed framework

NLP Model

Natural Language Processing (NLP) models are designed to understand, interpret, and generate human language. These models range from simple rule-based approaches to complex deep learning architectures.

Lemmatization: Lemmatization is the process of reducing words to their base or root form, known as the lemma. This process considers the context and morphological analysis of the words to ensure that they are reduced to their appropriate base forms. Unlike stemming, which simply chops off prefixes and suffixes to generate the root form, lemmatization uses vocabulary and morphological analysis to accurately transform words into their lemmas. The words "running," "ran," and "runs" are all reduced to the lemma "run."

N-gram features: N-gram features are a key concept in NLP and text analysis. An n-gram is a contiguous sequence of n items from a given sample of text or speech. The items can be phonemes, syllables, letters, words, or base pairs according to the application. N-grams capture local context and can distinguish between different usages of the same word.

NLP features: In NLP, features are specific pieces of information extracted from text data that are used to represent and analysed the text. By extracting and utilizing these features, NLP systems can better understand and process human language, leading to more accurate and effective applications.

Table 1 : Example NLP based relationships

Initial word	Next word
JJ	NNS/NN
RBR/RB/RBS	JJ
JJ	JJ
NNS/NN	JJ
RBR/RB/RBS	VB/VBD/VBG
VBN	NNS/NN
VBG/VB	JJR/JJ/JJS
JJ	VBN/VBG
RBR/RB/RBS	RB/RBR/RBS

VBN/VB/VBG/VBD represents the verb, JJ/JJR/JJS represents the adjectives, RB/RBR/RBS represents the adverb, and NN/NNS represents the noun.

POS tag-based features: Part-of-Speech (POS) tagging is a crucial preprocessing step in natural language processing (NLP) that involves assigning parts of speech to each word in a text. POS tagging helps in understanding the syntactic structure of a sentence and is often used to extract various features for different NLP tasks.

Negation: Negations play a crucial role in linguistics since they have an impact on the binary nature of other words. Examples of negative phrases include words such as no, not, shouldn't, and similar expressions. When a negation is present in a phrase, it is crucial to ascertain which words are impacted by this term. Negation may either be restricted to the word immediately after it or expand to include additional words following the negation.

Semantic based features: Semantic-based features in natural language processing (NLP) focus on capturing the meaning and context of words, phrases, and sentences. These features go beyond the syntactic structure and aim to understand the underlying semantics of the text. Semantic-based features provide a deeper understanding of text, enabling more sophisticated and accurate NLP applications by capturing the meaning and context beyond surface-level representations.

Classification Model:

We combine the Blooms filter features and the customized NLP features to train and test this module. We used a customized implementation of the ANN classification to build the final module. All WH types of tokens feed into the hidden layer, while the input layer receives the extracted tokens generated from the test data. The system produced more accurate results by adjusting the different hyperparameters of the ANN using both feed-forward and backpropagation networks.

Results and Discussions

The subsequent Table 2 provide a comprehensive account of a performance evaluation conducted with different quantities of input layers, hidden layers, batch size, and random alpha. Nevertheless, the beta1 and beta2 values have been calculated dynamically during the module's training, testing, and execution phases.

Table 2 : Performance matrix for training and Testing of proposed ANN model

No. of input layers	Hidden Layer size	Batch size	Training Accuracy	Testing Accuracy
10	**10**	16	95.32	95.31
		32	95.82	95.98
		64	95.88	95.98
		128	95.90	95.97
	20	16	95.87	95.85
		32	95.40	95.48
		64	95.80	95.85
		128	96.45	96.45
	30	16	95.88	95.81
		32	95.23	95.30
		64	96.24	96.22
		128	93.88	93.88
	10	16	95.30	95.56
		32	95.56	95.56
		64	95.56	95.56
		128	95.5	95.5
	20	16	95.56	95.56

20		32	95.89	95.89
		64	95.75	95.75
		128	95.83	95.83
	30	16	95.59	95.59
		32	95.58	95.58
		64	95.77	95.77
		128	95.79	95.79
30	**10**	16	95.56	95.56
		32	95.56	95.56
		64	95.56	95.56
		128	95.56	95.56
	20	16	95.56	95.56
		32	95.56	95.56
		64	95.56	95.56
		128	95.86	95.86
	30	16	95.18	95.18
		32	95.89	95.89
		64	95.07	95.07
		128	95.99	95.99

The Table 2 describes an accuracy of purposed module outcomes with bloom filter features extraction and ANN based machine learning classification module.

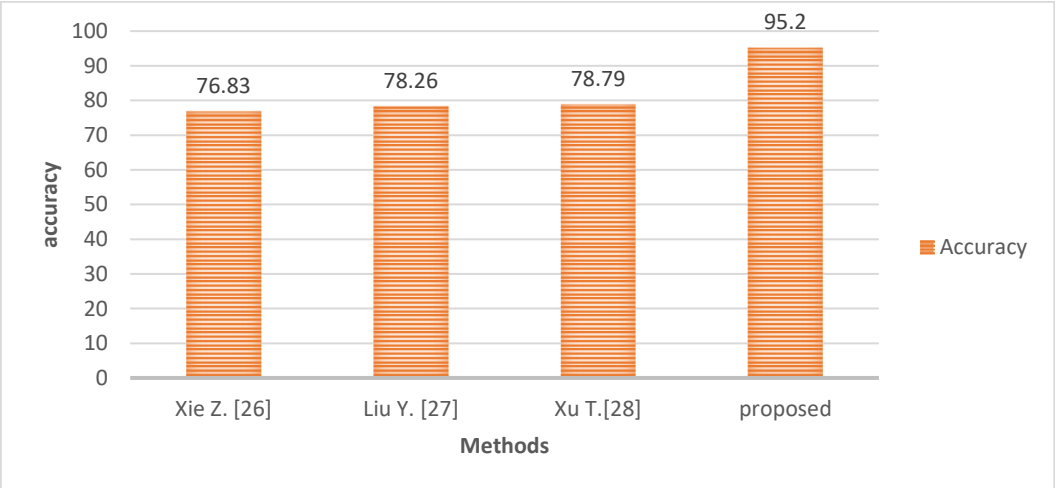

Figure 2 : comparative analysis of proposed model

Due to the limited availability of extraction-generative techniques using knowledge graphs, a comparison model of extraction methods is selected for experimentation. Figure 2 clearly demonstrates the successful outcomes of our strategy. Through thorough research and analysis, it has been determined that the errors in the experiment primarily stem from the insufficient presence of entities in the knowledge vocabulary, the presence of incorrect entities, and the use of words that describe relationships in both the knowledge base and the answers. These errors are a result of

the entity recognition module's recognition capabilities. The amount of the vocabulary setting has a direct influence on the performance of our model.

Conclusion

This article introduces our innovative method for automatically generating questions from unorganized material. It integrates a linguistic methodology that incorporates different phrase patterns (such as blooming, NLP, and semantic patterns) with machine learning techniques. The process of obtaining transformation rules that convert declarative sentences into questions is automated by using an initial collection of sentence-question pairings. The benefit of this technique is its ability to easily expand transformation rules and consistently enhance the model. Consequently, the need for human specialists to intervene is reduced compared to alternative methods. The primary contributions of our technique may be summarized as follows: A data-driven method enables the model to train and update question generating processes using sentence-question pairings. This eliminates the requirement for human experts in linguistics to manually design transformation rules or human experts in software engineering to alter the source code. The enhancement of the system relies on explicit feedback provided by users, which is used in reinforcement learning and by incorporating fresh sentence-question combinations into the supervised learning model. The framework utilizes a hierarchy of patterns to generate different sorts of queries based on the similarity match between texts at different stages of processing.

References

[1] Dhomse KB. Automatic question paper generation using ML: A review. Turkish Journal of Computer and Mathematics Education (TURCOMAT). 2021 Apr 11;12(1S):239-45.

[2] Ramli R, Sivan S, Razalli H. A review on automated examination question paper template generator. In2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA) 2020 Feb 28 (pp. 180-185). IEEE.

[3] Shihua Xu et al, “Research on question classification for automatic question answer”,IEEE,pp[218-221],2016.

[4] Juee Gosavi, B. N. Jagdale“Answer Selection In Community Question Answering Portals”,IEEE,2018.

[5] Xingboxie,“Automatic question answering system based on ontology.”,IEEE,pp[1366-1367],2015.

[6] Surbhi choudhary, “Question paper generator system”, ijcst, vol-3, pp [147-176], Sep-Oct 2015.

[7] Yoichi matsuyama et al, “Automatic expertise opinion sentence generation for enjoyable conversational system”, IEEE, vol-23,pp[313 -326],Feb-2015. Automatic Question Paper Generation using ML: A Review245

[8] Ehsan sherkat et al, “A hybrid approach for question classification in persian automatic question answering system”,IEEE,pp [279-284],2014.

[9] L. Bednarik and L. Kovács, “Implementation and assessment of the automatic question generation module,” in 3rd IEEE International Conference on Cognitive Info communications, CogInfoCom 2012 -Proceedings, pp. 687–690, 2012.

[10] N. Omar, S. Haris, R. Hassan, H. Arshad, M. Rahmat, N. Zainal and R. Zulkifli,, “Automated Analysis of Exam Questions According toBloom’s Taxonomy,” Procedia -Social and Behavioral Sciences, vol. 59, no. 1956, pp. 297–303, 2012.

[11] Jun Araki, Dheeraj Rajagopal, Sreecharan Sankaranarayanan, Susan Holm, Yukari Yamakawa, Teruko Mitamura “Generating Questions and Multiple-Choice Answers using Semantic Analysisof Texts” International Conference on Computational Linguistics: Technical Papers, pages 1125–1136, Osaka, Japan, December 11-17 2016.

[12] Tirath Prasad Sahu, Reswanth Sai Thummalapudi & Naresh Kumar Nagwani, “Automatic question tagging using multi-label classification in community question answering sites”, IEEE, pp [63-68], 2019.

[13] Thenmozhi, D., & Kumar, G. R. (2018). “An Open Information Extraction for Question Answering System.” 2018 International Conference on Computer, Communication, and Signal Processing (ICCCSP), Icccsp, pp[1–5].

[14] Abderrazzak SAMADI, EL FAZAZI Hanaa, Mohammed QBADOU, Mohamed YOUSSFI, Fatiha AKEF “A syntactic and semantic multi-agent based Question Answering System for collaborative E-learning”,IEEE,2018.

[15] Zalte, S. V, Jadhav, C. C., Mangier, A. A., Hole, A. D., & Tulshi, A. R. (2018). “Automatic Question Paper Generator System”.Cim,pp[545–548]. Https://doi.org/10.17148/IJARCCE.2018.73103

[16] Nguyen-Thinh Le, Tomoko Kojiri, Niels Pinkwart, "Automatic question generation for educational application-The state of art.", Springer, 2011.
[17] Ms. Priti SGumaste, Ms. Shreya S Joshi, Ms. Srushtee A Khadpekar, Shubhangi R Mali, "Automated Question Generator System using NLP Libraries", International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 06 | June 2020 pp 4568-4572]
[18] G. Cen, Y. Dong, W. Gao, L. Yu, S. See, Q.Wang, Y. Yang, H. Jiang, , "A implementation of an automatic examination paper generation system," Mathematical and Computer Modeling, vol. 51, no. 11–12, pp. 1339–1342, 2010.
[19] Wang Bo, "Research on the design of the ontology-based automatic question answering system",IEEE,pp[871-874],2008.
[20] T. Fei, W. J. Heng, K. C. Toh, and T. Qi, "Question classification for e-learning by artificial neural network," in ICICS-PCM 2003 -Proceedings of the 2003 Joint Conference of the4th International Conference on Information, Communications and Signal Processing and 4th Pacific-Rim Conference on Multimedia, vol. 3, pp. 1–5, 2003.
[21] Dhawaleswar Rao Ch and Sujan Kumar Saha,"Automatic Multiple Choice Question Generation from Text: A Survey",IEEE,2018.
[22] Ms. Priti S Gumaste, Ms. Shreya S Joshi, Ms. Srushtee A Khadpekar, Shubhangi R Mali, "Automated Question Generator System: A Review", International Journal of Engineering Applied Sciences and Technology, 2019 Vol. 4, Issue 8, ISSN No. 2455-2143, Pages 171-176 Published Online December 2019 in IJEAST (http://www.ijeast.com)
[23] Liancheng guan, "The design of the automatic generative system of examination papers",IEEE,pp[110-113], 2017.
[24] Bindra, S. K., Girdhar, A., & Bamrah, I. S. (2017). "Automatic Question Paper using Data Mining."Icces, 629–634.
[25] Gauri Nalawade, Rekha Ramesh" Automatic Generation of Question Paper from User Entered Specifications using a Semantically Tagged Question Repository" IEEE 8th International Conference on Technology for Education,2016.
[26] Xie, Z.; Zeng, Z.; Zhou, G.; He, T. Knowledge base question answering based on deep learning models. In Natural Language Understanding and Intelligent Applications; Springer: Cham, Switzerland, 2016; pp. 300–311.
[27] Liu, Y.; Zhang, L.; Yang, Y.; Zhang, G.; Zhang, C. Simple question answering with entity ranking and joint fact selection. Appl. Res. Comput. 2020, 37, 3321–3325.
[28] o Luo, D.; Su, J.; Li, P. Multi-view Attentional Approach to Single-fact Knowledge-based Question Answering. Comput. Sci. 2019, 46, 215–221.
[29] Xu, T.; Wu, M. An Improved Naive Bayes Algorithm Based on TF-IDF. Comput. Technol. Dev. 2020, 30, 75–79.