AUTOMATED QUESTION GENERATOR SYSTEM: A REVIEW

Ms. Priti S Gumaste, Ms. Shreya S Joshi, Ms. Srushtee A Khadpekar, Shubhangi R Mali
Final Year B.E. (Computer Engineering)
Sandip institute of Technology & Research Centre, Nashik, Maharashtra, India
(Affiliated to Savitribai Phule Pune University, Pune)

Abstract— **Question generation, the task of automatically creating questions that can be answered by a certain span of text within a given passage, is important for question-answering and conversational system in digital assistants. Automatic generation of questions from text plays a key role in two domains - interactive question answering sessions and educational assessment. Recent sequence to sequence neural models have outperformed previous rule-based system. Existing models mainly focus on using one or two sentences as the input. In proposed system the admin can add text or paragraphs of his/her choice. User will operate that system hence user can enter the paragraph in English language with grammatically correct sentence. The sentence is selected and separated and Stanford POS tagger for POS Tagging is applied. After the input is given, keywords from a data-set are matched to the input text so as to find the sentence/context on which the question can be created. In feature extraction the system will identify the questionable term from that sentence and rearrange the words and automatically create question from entered sentence or paragraph using Bloom's Taxonomy. However, it often requires the whole paragraph as context in order to generate high quality questions. Proposed system uses Stanford tagger for tagging the sentences with gated self-attention encoder to address the challenges of processing long text input for question generation. With sentence-level input, this model outperforms previous approaches with either sentence or paragraph input. Furthermore proposed model can also effectively utilize paragraphs as inputs.**

Keywords— **Paragraph, grammar, Sentence separation, Stanford POS tagger, question word, Bloom's Taxonomy**

I. INTRODUCTION

Question generation, aiming at creating questions from natural language text, e.g. a sentence or paragraph, is an important area in natural language processing. This issue has a challenging interests from both industrial and academic communities, due to the booming of Question-and-Answer and conversation systems, such as Alexa, Cortana, Google Assistant and Siri, the advancement of question generation or machine comprehension technologies together. A traditional system can be proactive by asking questions, while proposed system can benefit from a large scale question-answering corpus which can be created by an automated system. Answering the questions by the information received through reading comprehension is the key application of education. Automatic question generation concept mainly tackled by two approaches: 1) rule-based approach 2) neural-based Question Generation approach. It is also assumed the answer is comprised of certain spans of the text from the given passage. Similar problems have been addressed by many researchers. A paragraph often contains much richer context than a sentence pointed out that about questions in paragraph-level information to be asked and using the whole paragraph can improve question generation performance on those questions. However, a paragraph can contain irrelevant information w.r.t. the answer for generating the question. The challenge is thus how to effectively utilize relevant information at paragraph-level for question generation.

Generating set of questions automatically through a system deals with Natural Language Processing. NLP is an interesting and prominent area of research. Many research papers were referred along with materials related Boom's taxonomy to prepare background to work with this area. Many distinguished researchers across the globe have worked in this context, there is enough scope in achieving better results with respect to performance measures like accuracy, precision, recall etc. Even in digital era, on one side academicians/ teachers / professors / tutors from university / academic institutions / industries spend a lot of energy and time in preparing question papers, quizzes manually through traditional systems and on other side students too spend lots of time on self-analysis. Learners are mainly dependent on their mentors for improvement and their self-analysis. Hence, it is need of an hour to work in this area, which has wide scope of development. Learners can be benefited for calibrating themself and removing dependencies on mentors using this proposed system. Here, user can give the information in the form of text material that they referred to, and on basis of information provided (may be paragraph/page etc), system generates series of questions from which one can do his/her self-analysis. Academicians can also use similar approach for

creating test papers and quizzes by providing information to the system.

The rest of the paper is organized as follows. Through literature survey is discussed in section II. Bloom's Taxonomy is discussed with keywords classification in section III. Proposed model is highlighted in section IV. Section V concludes the proposed model with references listed in Section VI.

II.   LITERATURE SURVEY

Zhang and Lee (2003) [1] have proposed a method based on machine learning techniques in order to categorize questions into Bloom's cognitive levels in which, several research studies have been presented in the field of exam question classification. For example, the authors have utilized the n-gram features in order to address each term located in the question separately. Authors classified the questions into several semantic based categories. However they used Open domain factual questions are considered for experimentation carried over publicly available data sets from USC (University of Southern California data set available from libguides.usc.edu/az.php), UIUC (University of Illinois at Urbana-Champaign data set available at databank.illinois.edu/datasets), TREC (Data set fro question classification available at Text REtrieval Conference www.trec.nist.gov). Authors suggested two layered question taxonomy covering 6 coarse grained categories and 50 fine grained categories. Authors considered surface text features as bag-of-words and bag-of-ngrams (all continuous word sequence in the question) for experimentation. However authors tried with SVM, NN, NB DT and SNoW (Sparse network of Winnows) classification algorithms in their research and found that, SVM classification gives out better results of accuracy with respect to other four classification algorithms over bag-of-ngrams feature. Finally authors conclude their research work by saying syntactic structure of questions are really helpful to question classification. They proposed tree kernel function to enable SVM classifier to take advantage questions with syntactic structure/nature.

Chang and Chung (2009) [2] proposed a combination of machine learning and keyword-based approach for classifying questions based on Bloom's cognitive level. Authors suggested five functionality architecture in their proposed system. Keyword management module can add, modify, delete and query the keywords. Question management module performs two functions mainly classifying the identified keywords to related Boom's cognition level and managing test item attributes. Examination maintain and Analysis model deals with authoring and analysis of paper as per Bloom's taxonomy. Test module deals with generating and conducting online examination and provides result analysis. Lastly System Management level can update, add, delete user accounts through administrator and provides various levels of authorities to each user. Such an approach utilizes the

keywords that distinguish the level of the question, for instance, the keyword 'define' is related to remembering level. Authors conclude their research work with 75% correct keywords matching for knowledge level category of Blooms taxonomy, however percentage results are lower for other category levels.

Yusof and Hui (2010) [3] proposed a machine learning technique with a statistical feature called Category Frequency-Inverse Document Frequency (CF-IDF). As per author's work flow, initial feature set is formed as a result of preprocessing steps like word extraction, stop word removal and stemming over long stream of characters from the paragraph or material supplied which reduces size of initial feature set. Porter's stemming algorithm is then applied for stemming process to create valid stem. Stemming is a process of converting different morphological variants of the same word into their most common form, called a stem. Size and complexity of the feature vector that represents a question item is further reduced due to stemming process, and thus the performance of the question classifier can be improved. To improve the scalability, authors suggested two potential feature reduction methods, the DF method and the CF-DF method. The DF method uses local feature ranking technique that reduces the size of the initial feature set. The CF-DF method introduces a quantity called the category frequency that allows the discrimination value of a feature to be considered in the feature reduction process. The proposed method aims to utilize the frequency of each class label of Bloom in order to provide a probability mechanism for new questions.

Haris and Omar (2012) [4] proposed a rule-based approach for classifying questions that are related to the computer science domain. Such rule-based approach is based on Natural Language Processing (NLP) techniques such as normalization (i.e., eliminate the stopwords) stemming (i.e., retrieving the root of each word), Part-Of-Speech (POS) Tagging (i.e., provide the syntactic class for each word). However, this approach is not dynamic as the machine learning techniques and produced poorly results of classification.

Yahya et al. (2015) [5] have addressed the use of machine learning techniques in classifying the questions into Bloom's Cognitive Level (BCL). Authors suggested four steps of using ML to tackle the problem of questions classification into BCLs.First step is Question representation step - as question text cannot be directly interpreted by Machine Learning techniques, it is a need of a conversion procedure to map the question text into a compact representation of its content. Questions were represented by vector of term weight <w1j, … wTj> where T is a feature. For this, initial steps conducted by authors were converting the question text to lower case, removing punctuation's, tokenization and these token were stemmed with porter stemmer algorithm. Second step is Term Selection Step - selects a subset of terms with highest score according to a function, e.g. Term Frequency (TF), which measures the importance of a term for a classification task. Third step is Classifier construction step - a general inductive

process automatically builds a classifier for a given class ci by observing the characteristics/features of a set of text, called training set. Final step is Classifier evaluation step - Common measures like Precision, recall, Fβ measure and accuracy are taken to evaluate ML performance of system. Rule based, KNN, NB, SVM classifiers were used in the process and authors conclude their work with the statement that, SVM outperforms the other classifiers in terms of Fβ measure and accuracy.

Results showed that SVM has superior performance compared to the other classifiers. Abduljabbar and Omar (2016) [6] have proposed a combination classification using voting technique among multiple classifiers including NB, SVM, and KNN. In fact, the proposed method has been created to classify programming questions by tokenizing and removing stopwords from the material supplied. . In this manner, multiple feature extraction techniques have been used including chi-square (measures lack of flexibility between the category and the term), mutual information (for eliminating tokens with less beneficial attribute), and odds ratio (general idea was that the distribution of relevant document features differs from the distribution of irrelevant document features). Authors conclude their research work with results of KNN classifier has greater results in performance measures like accuracy and F1 measure.

 OMAR J et all (2015, 2019) [7][12] articulates a combination method for identifying Bloom's categories for questions. The proposed combination method has utilized a semantic and syntactic approach. The semantic approach can be represented by using an external knowledge of the WordNet dictionary with the Lesk algorithm. While the syntactic approach can be represented by using the POS tagging in order to retrieve frequent patterns of terms. Finally, three classifiers have been applied in order to classify the questions. In order to examine the effectiveness of the proposed method, the three classifiers have been carried out twice; first with the proposed combination method, and second without the proposed combination. Results revealed that the three classifiers with the proposed combination method shown better performance rather than without using the combination. For future directions, addressing large-scale data-set of questions would be an interesting effort. In addition, examining recent technologies such as word embedding would yield promising results.

Ozar et all (2011) proposed category weighting mechanism. According to their model, weights are assigned to the conflicting categories. The weight is calculated based on question's category from subject matter experts (SMEs).[11]

 Taqi M K, Ali Rosma (2016) also proposed types of text classifications based on multiple categories to get question classification [13].

III. BLOOM'S TAXONOMY

Bloom's taxonomy plays role in defining teaching and learning objectives. Bloom's taxonomy is a set of three hierarchical models used to classify educational learning objectives into levels of complexity and specificity based on the student's understanding level. Bloom's Taxonomy consists of six levels, with each level having a different degree of competency. Levels with original and revised action verbs are discussed with few question verbs as example. (2009, 2015, 2016) [9],[8],[10]

1. Knowledge / Remembering: Capability of a student to reproduce / recall the information by remembering what they have learned. Some of the question verbs of this category are - define, state, identify, list, label.

2. Comprehension / Understanding: Comprehension involves demonstrating an understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions, and stating the main ideas. Some of the question verbs under this are - explain, summarize, interpret, convert, predict.

3. Application / Applying: Application involves using acquired knowledge—solving problems in new situations by applying acquired knowledge, facts, techniques and rules. Learners should be able to use prior knowledge to solve problems, identify connections and relationships and how they apply in new situations. Students are required to apply what they have learned to solve problems using suitable formula and algorithms.

4. Analysis / Analyzing: Analysis involves examining and breaking information into component parts, determining how the parts relate to one another, identifying motives or causes, making inferences, and finding evidence to support generalizations. Some of popular question verbs are - analyze, compare, contrast, relate, differentiate.

5. Synthesis / Evaluating: Questions on this level allow students to invent or build a new structure from various elements and integrate the sections / parts together to form a complete system. Some of the question verbs are - compile, create, design, construct, formulate.

6. Evaluation / Creating: Students are required to make judgments regarding the values of materials or ideas, method or solution to problems using skills learned. Examples of question verb are - defend, evaluate, justify, relate, support. [8]. Table 1 summarize keywords as per Bloom's Taxonomy, we are considering all these keywords in our proposed system.

IV.   PROPOSED MODEL

In proposed system the admin can simply add users. User can enter the paragraph in English language with grammatically correct sentence. System will pre-process the entered material, and sentences are separated. Feature extraction process will extract the question words, then apply Stanford Pos tagger for Pos Tagging. In feature extraction the system will identify the question word from that sentence and rearrange that question

words and automatically get question from entered sentence or paragraph. However, it often requires the whole paragraph as context in order to generate high quality question. Proposed model uses Stanford tagger for tagging the sentences with gated self-attention encoder to address the challenges of processing long text input for question generation. During the process it is proposed to adopt all basic classifier techniques to find better results. Research plan also has an idea of comparing hybrid approach for better performance.
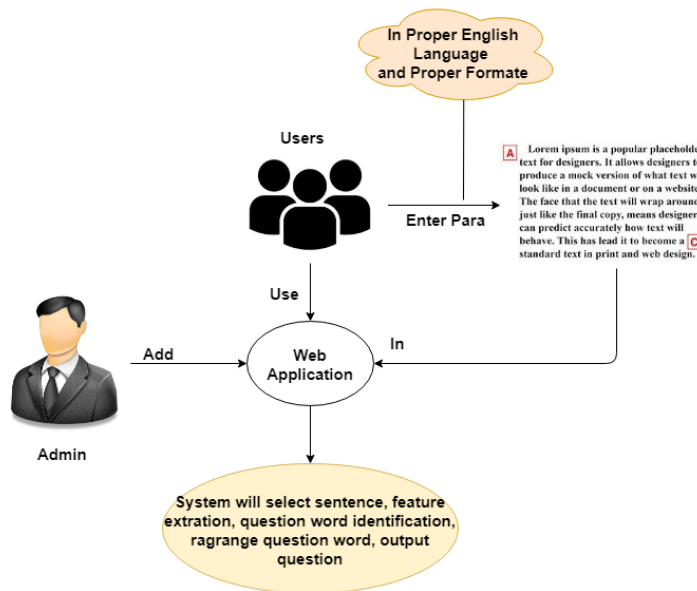

Figure 1: Abstract Architecture.

V. CONCLUSION

During the discussion, useful and great research papers were collected and discussed along with methodologies adopted so far and advantages and research gap. We conclude here by observing enough scope of improving performance measures for analysis of system. Hybrid approach may provide better result.

Table 1: Bloom's Taxonomy Action Verbs [9]

Definitions	I. Remembering	II. Understanding	III. Applying
Bloom's Definition	Exhibit memory of previously learned material by recalling facts, terms, basic concepts, and answers.	Demonstrate understanding of facts and ideas by organizing, comparing, translating, interpreting, giving descriptions, and stating main ideas.	Solve problems to new situations by applying acquired knowledge, facts, techniques and rules in a different way.
Verbs	Choose Define Find How Label List Match Name Omit Recall Relate Select Show Spell Tell What When Where Which Who Why	Classify Compare Contrast Demonstrate Explain Extend Illustrate Infer Interpret Outline Relate Rephrase Show Summarize Translate	Apply Build Choose Construct Develop Experiment-with Identify Interview Make use of Model Organize Plan Select Solve Utilize
Definitions	IV. Analyzing	V. Evaluating	VI. Creating
Bloom's Definition	Examine and break information into parts by identifying motives or causes. Make inferences and find evidence to support generalizations.	Present and defend opinions by making judgments about information, validity of ideas, or quality of work based on a set of criteria.	Compile information together in a different way by combining elements in a new pattern or proposing alternative solutions.
	Analyze Assume Categoriz	Agree Appraise Assess	Adapt Build Change

Verbs	e Classify Compare Conclusion Contrast Discover Dissect Distinguish Divide Examine Function Inference Inspect List Motive Relationships Simplify Survey Take part in Test for Theme	Award Choose Compare Conclude Criteria Criticize Decide Deduct Defend Determine Disprove Estimate Evaluate Explain Importance Influence Interpret Judge Justify Mark Measure Opinion Perceive Prioritize Prove Rate Recommend Rule on Select Support Value	Choose Combine Compile Compose Construct Create Delete Design Develop Discuss Elaborate Estimate Formulate Happen Imagine Improve Invent Make up Maximize Minimize Modify Original Originate Plan Predict Propose Solution Solve Suppose Test Theory

VI. REFERENCE

[1]. Zhang, D.; and Lee, W.S, (2003) "Question classification using support vector machines", Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval. Toronto, Canada, (pp 26-32)
https://www.comp.nus.edu.sg/~leews/publications/p31189-zhang.pdf

[2]. Chang, W.-C.; Chung, M.-S. (2009). "Automatic applying Bloom's taxonomy to classify and analysis the cognition level of English question items". Joint Conferences on Pervasive Computing (JCPC). Tamsui, Taipei, Taiwan, (pp 727-734)
https://ieeexplore.ieee.org/document/5420087
DOI: 10.1109/JCPC.2009.5420087

[3]. Yusof, N.; Hui, C.J.(2010). "Determination of Bloom's cognitive level of question items using artificial neural network". Proceedings of the 10th International Conference on Intelligent Systems Design and Applications (ISDA). Cairo, Egypt, (pp 866-870).
https://ieeexplore.ieee.org/document/5687152

[4]. Haris, S.S.; and Omar, N. (2012). "A rule-based approach in Bloom's Taxonomy question classification through natural language processing". Proceedings of the 7th International Conference on Computing and Convergence Technology, Seoul, South Korea, (pp 410-414) *https://ieeexplore.ieee.org/document/6530368*

[5]. Yahya, A.A.; Osman, A.; Taleb, A.; Alattab, A.A. (2013). "Analyzing the cognitive level of classroom questions using machine learning techniques". Procedia-Social and Behavioral Sciences, 97, (pp 587-595).
https://www.sciencedirect.com/science/article/pii/S1877042813037221

[6]. Abduljabbar, D.A.; and Omar, N. (2015). "Exam questions classification based on Bloom's taxonomy cognitive level using classifiers combination". Journal of Theoretical and Applied Information Technology, 78(3), (pp 447-455). *https://pdf.sciencedirectassets.com/277811*

[7]. Omar J. Mohamed, Nawaf A. Zakar, Basel Alshaikhdeeb; (2019) "A combination method of syntactic and semantic approaches for classifying examination questions into Bloom's taxonomy cognitive", Journal of Engineering Science and Technology Vol. 14, No. 2 (2019) (pp 935 - 950) © School of Engineering, Taylor's University. *https://www.academia.edu/38919824*

[8]. Benjamin S. et al (1950-1975) "A Committee of College and University Examiners edited by "Bloom Taxonomy of Educational Objectives - The Classification of Educational" Goals"*https://www.uky.edu/~rsand1/china2018/texts/Bloom%20et%20al%20-Taxonomy%20of%20Educational%20Objectives.pdf*

[9]. Anderson, L. W.,& Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing, Abridged Edition. Boston, MA: Allyn and Bacon

[10]. Soozandehfar S M A, Ali M R (2016), "A Critical Appraisal of Bloom's Taxonomy", American Research Journal of English and Literature(ARJEL) , vol 2 (pp 1-9)

[11] Omar Nazila, Haris S S, Hassan Rosila, (2011), "Automated analysis of exam questions according to bloom's Taxonomy", doi: 10.1016/j.sbspro.2012.09.278, Elsevier Ltd. Selection and/or peer reviewed under responsibility of the UKM Teaching and Learning Congress 2011 (pp 297-303)

[12] Kusuma S F, Siahaan D, Yuhana U L (2015), "Automatic Indonesia's Questions Classification Based On Bloom's Taxonomy Using Natural Language Processing A Preliminary Study", DOI: 10.1109/ICITSI.2015.7437696, 2015 International Conference on Information Technology Systems and Innovation (ICITSI) Bandung – Bali, (978-1-4673-6664-9/15 ©2015 IEEE)

[13] Taqi M K, Ali Rosma (2016), "Automatic Question Classification Models for Computer Programming Examination: Asystematic literature review" Journal of Theoretical and Applied Information Technology, Vol 93, Number 2, (pp 360-374)