Assessment of Bangla Descriptive Answer Script Digitally

Preprint · September 2019

DOI: 10.13140/RG.2.2.12203.39204

5 authors, including:

Md Gulzar Hussain
Green University of Bangladesh
19 PUBLICATIONS **11** CITATIONS
SEE PROFILE

Sumaiya Kabir
Green University of Bangladesh
13 PUBLICATIONS **11** CITATIONS
SEE PROFILE

Tamim Al Mahmud
Aston University
17 PUBLICATIONS **16** CITATIONS
SEE PROFILE

Ayesha Khatun
Chittagong University of Engineering & Technology
20 PUBLICATIONS **17** CITATIONS
SEE PROFILE

Assessment of Bangla Descriptive Answer Script Digitally

Md Gulzar Hussain*, Sumaiya Kabir†, Tamim Al Mahmud‡, Ayesha Khatun§, Md Jahidul Islam¶
Department of Computer Science and Engineering*
Green University of Bangladesh, Dhaka-1207, Bangladesh
gulzar.ace@gmail.com*, sumaiya@cse.green.edu.bd†, tamim@cse.green.edu.bd‡,
ayesha@cse.green.edu.bd§, jahidul.jnucse@gmail.com¶

***Abstract*—Answer script evaluation is an essential part of the student evaluation process in the education system. In an exam, students need to answer subjective and objective questions. In educational institutes, instructors need to evaluate the answer script manually to evaluate the students. In Bangladesh, the number of students and institutes are increasing day by day. For this reason, it is becoming hard to evaluate the answer script in a perfect way by the instructors. So it is necessary to find a way to evaluate the answer script automatically. Many techniques are proposed for the English language, but we didn't find any for Bangla language. Our paper proposed a way to evaluate Bangla subjective answer scrips automatically by keyword matching and linguistic analysis. Using our proposed model we have tested on answer scripts of 20 questions and found the minimum relative error of 1.8%. 15 teachers and 10 students volunteered to evaluate the answer scripts.**

Keywords: Bangla Language, Bangla Subjective Script, Evaluation, Keyword Matching, Automatic Evaluation, Answer Script.

I. Introduction

To test the skills and knowledge of individuals, examination systems are designed. There are various kinds of tests or examination systems all over the world, such as multiple choice questions, subjective, etc. Objective evaluation is a way of questioning that has one correct answer. In the other hand, subjective evaluation can have more than one correct answer. Like other countries, Bangladesh also follows both of these evaluations. In Bangladesh there are around 23,907,151 students overall in primary, secondary, post secondary level of education in 2015 and primary language used in these education levels are Bangla and English [1]. But there are not enough teachers or instructors for those students. It is becoming hard to evaluate these students for these teachers or instructors. In 2015, the teacher & student ratio was 1:41 [2] [3]. Teaching students, setting questions, evaluating the answer script of the students is difficult for the teachers due to this ratio. If some of these tasks can be performed then it will make life easier for the instructors. Not only in the education system but also in the various job entrance examination is done in a subjective way. These answer scripts also need to be evaluated. So if we can asses the answer script automatically then the evaluation system will be more smooth for Bangladesh.

The rest of the paper is organized as follows: **Section II** discusses previous works. Methodology discussed in **Section III**. Result analysis and discussion demonstrated in **Section IV**. And finally **Section V** refers the conclusion.

978-1-7281-5242-4/19

II. Previous Works

Generally, answer scripts evaluation is a difficult task. It becomes harder for Bangla language. Various works are found for English but none is found for Bangla Text. Here we will discuss some previous works to evaluate subjective answer script.

Authors of [4] suggest alternative sentence generator method to produce an alternative model response by linking the technique to a synonymous dictionary. In the matching stage, they proposed combination of three algorithms, Commons Words (COW), Longest Common Sub-sequence (LCS) and Semantic Distance (SD), which were used effectively in many Natural Language Processing systems and produced effective outcomes. Hyperspace Analog to Language (HAL) procedure and Self-Organizing Map (SOM) method is used to evaluate students answer script in paper[5]. Kohonen Self-Organizing Maps clustering technique is applied to the vector in their suggested system. Authors of [6] observed that the semantic Enhanced NLP based technique outperforms easy lexical matching techniques. This application mechanism offers an automatic answer assessment based on the keyword given by the moderator to the application in the form of the input that will provide equal mark distribution.

In paper [7], authors proposed a Natural Language Processing (NLP) based method for evaluating of the answer script. They used a keyword based summarizing technique to generate a summary of the answer script. Authors of [8] proposed a syntactical-relation based feature extraction technique to evaluate descriptive type answer scripts. Their method contains steps like question-classification, answer-classification, and answer-evaluation of subjective answers of students and grade with a suitable score. Advanced machine learning techniques and methodologies based on a new model is proposed by Prakruthi et. al. in [9]. They did it for Optical Character Recognition based work involving supervised learning technique. An implementation that uses machine learning to evaluate the answers scripts is proposed in paper [10].

Authors of [11] proposed evaluation procedure in a semi-automated manner where subjective questions are supple-

mented with model answer points. Their suggested framework also includes provisions of reward systems and penalties.

III. Methodology

Working methodology for our proposed system is discussed in this section. Our system follows the given system diagram in Fig 1 to evaluate the Bangla answer script of the examinee.

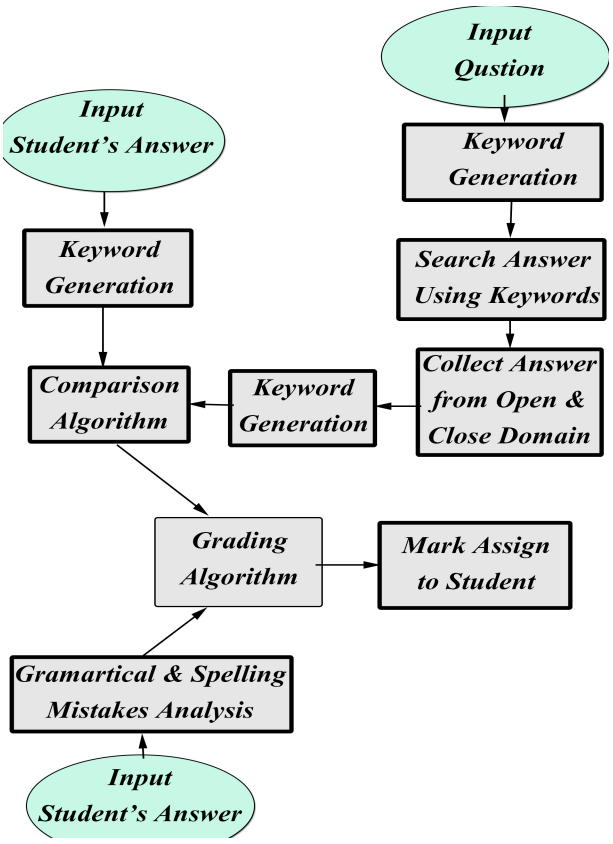

Fig. 1. System Workflow of proposed system

In the system flow diagram, we can see that step "Keyword Generation" repeated for the analysis of question, students answer script, and answer collected from open & close domain. This is an important step in our proposed system. We can also see that to evaluate the student's answer, we searched keywords generated from question in the open domain and closed domains. Here open domains include World Wide Web, various web pages, blogs, Wikipedia etc. and closed domain consist of a specific category or specific answer to the question. A comparison algorithm is proposed for comparing our generated answer keywords and keywords generated from the student's answer. These various steps are elaborated below:

A. Keyword Generation

Our proposed system simply takes an answer from the students as a text or document file. These answers are written by students in a document file. We are not taking handwritten answer scripts in consideration due to reducing the complexity. In this step, we process the text very carefully. Sub steps of Fig 2 are taken in this step.

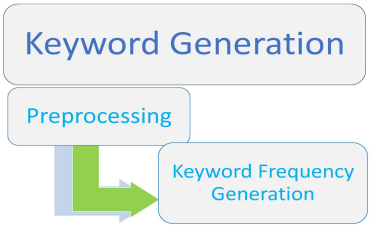

Fig. 2. Sub-steps in the Keyword Generation

1) Preprocessing: As answers collected from various domains and students contain many unnecessary words, stop symbols, punctuation symbols, emotions symbols etc. These are noise for our unprocessed data. We remove these special characters like #, &, % etc. emojis like :), :P etc, stop words etc. We also remove the articles, punctuations etc. to simplify the text. After preprocessing keywords are generated and we use the following sub-steps to generate their frequency.

2) Keyword Frequency Generation: After preprocessing all the text data we need to identify every keyword and count their frequency ratio. To do that we used algorithm 1.

Algorithm 1: Algorithm to Generate Keyword Frequency

```
Answer = Answer Script after preprocessing;
Frequency_Array = NULL;
Number_Keywords = Number of keywords in the
  Answer;
for Words in Answer do
    if Word found in Frequency_Array then
        Frequency_Array[word] =
          Frequency_Array[word] + 1/Number_Keywords;
    else
        In the Frequency_Array add that word as index ;
        Frequency_Array[word] = 1/Number_Keywords;
    end
end
```

Frequency Array is formed after applying algorithm 1 where frequency of every word is calculated. There will be no stop words, punctuation, irrelevant words in this array as all of them will be removed after preprocessing.

B. Searching & Collecting Answers Using Keywords

To assessment, student answer system will need a standard answer for a question. Answers can be from an open or close domain. An instructor can provide the answer manually or system can collect the answer from the internet or other resources. Our system will parse data from online knowledge-based websites like Wikipedia etc. The MediaWiki action API is a good example of web service that provides access to certain wiki features such as authentication, page operations, and search where the entry point is bn.wikipedia.org/w/api.php or many others. Our proposed system will collect data depending on the keywords from the World Wide Web or any other local resources. As an example, if the question is Bangladesh

sampark likhuna (Write about Bangladesh.) then it will search about Bangladesh and collect answers from these domains.

C. Comparison Algorithm to compare Two Results

When keywords and their frequency ratio is calculated for a student's answer script and searched answer, then we need to compare them for scoring the student's answer. To do that algorithm 2 is proposed.

Algorithm 2: Comparison Algorithm

KeywordScore = 0;
SAFrequency = Frequency Result Of Student's Answer;
SRFrequency = Frequency Result Of Searched Answer;
LengthSR = Size of SRFrequency;
WeightSA = Empty Array;
Sort SRFrequency in decending order;
for *word in SAFrequency* **do**
  if *word is in SRFrequency* **then**
    'word' added as index in WeightSA;
    if *SAFrequency[word] > SRFrequency[word]* **then**
      $WeightSA[word] = 1$;
    else
      $WeightSA[word] = \frac{SAFrewuency[word]}{SRFrequency[word]}$;
    end
    $SAFrewuency[word] = 0$;
    $SRFrequency[word] = 0$;
  else
  end
end
NeededWord = 0;
for *word in SRFrequency* **do**
  if *SRFrequency[word]* $! = 0$ **then**
    $WeightSA[word] = -1 * SRFrequency[word]$;
    $SRFrequency[word] = 0$;
    $NeededWord+ = 1$;
  else
  end
end
UnnecessaryWord = 0;
for *word in SAFrequency* **do**
  if *SAFrequency[word]* $! = 0$ **then**
    $WeightSA[word] = -1 * SAFrequency[word]$;
    $SAFrequency[word] = 0$;
    $UnnecessaryWord+ = 1$;
  else
  end
end

D. Grammatical & Spelling Mistake Analysis

Grammatical and Spelling mistake is very important for evaluating any kind of answer script. It is also important because students can write just the keywords required for a correct answer. If grammatical mistakes are checked then it will come to the knowledge of the system that the student is tried to cheat. We can see in the evaluation process grammar and spelling checking is done in many proposed method. But as we are working with Bangla language in which case such resources are limited and we are using Akkhor Bangla Spell and Grammar Checker [12] in our proposed method, which one is a tool to check spelling and grammatical mistakes. Hence we are proposing a simple keyword-frequency-based process with the use of this tool. This tool will return a score depending on the grammatical and spelling mistakes. That score will help the evaluation system to evaluate the student's answer script. The score is calculated based on algorithm 3.

Algorithm 3: Algorithm for Scoring Grammatical and Spelling Mistakes

NumberOfWord = Number of total words in Student Answer Script;
NumberOfSentence = Number of total words in Student Answer Script;
SMistakes = 0;
GMistakes = 0;
TGSMScore = Total Grammatical & Spelling Mistake Score which is initially 0;
for *Words in Student Answer Script* **do**
  if *word is in Student Answer Frequency* **then**
    $SMistakes = \frac{WeightSA[word]}{NumberOfWord} + SMistakes$;
  else
    $SMistakes = \frac{1}{NumberOfWord} + SMistakes$;
  end
end
for *Sentences in Student Answer Script* **do**
  $GMistakes = \frac{1}{NumberOfSentence} + GMistakes$;
end
$TGSMScore = \frac{SMistakes+GMistakes}{SMistakes*GMistakes}$

E. Mark Assigning to Students

After performing algorithm 2 and 3, we will get some values of parameters LengthSR, NeededWord, Un-necessaryWord, WeightSA, and TGSMScore. We need to assign marks to the students depending on the answer script. These values will be used in the grading algorithm step to assign mark to the student depending on his or her answer script. In the grading algorithm we are considering that out of the total mark 80% will be allocated to the answer and 20% will be for the grammatical and spelling mistakes. Exam authority will be able to change the value as their need.

In this step, calculation of the final mark of a student is done using the weights which are set in the comparison algorithm and score obtain from algorithm 3. Mark assignment is done by following grading algorithm 4. This algorithm takes consideration the score obtain from total grammatical and spelling mistakes and the number of necessary and unnecessary words to assign the mark to the answer.

Algorithm 4: Grading Algorithm to Assign Mark

StudentMark = 0;
FullMark = N;
ActualMark = $0.8 * N$;
GSMark = $0.2 * N$;
for *Words in WeightSA* **do**
  if *WeightSA[word] is positive* **then**
    $ActualMark = (N * WeightSA[word]) + ActualMark$;
  else
    $ActualMark = \frac{WeightSA[word]}{NeededWord+UnnecessaryWrod} + ActualMark$;
  end
end
$GSMark = TGSMScore * GSMark$;
FullMark = ActualMark + GSMark;

TABLE I
ABSOLUTE AND RELATIVE ERRORS FOR PROPOSED SYSTEM

Number of Teachers	Average Score given by Teachers	Score given by the system	Absolute Error	Relative Error
5	8.56	8.1	0.46	5.37%
10	8.25	8.4	0.15	1.81%
15	8.4	8.7	0.3	3.57%

IV. Result Analysis

For experimenting our system, we gave 20 questions for answering to 10 students. Every students answered two questions and each answer contain 300-350 Bangla words. Each of the answer scripts are evaluated by our volunteered teachers and assigned a score to each of the answer. Each of the questions was of 10 marks. We also evaluate the answer scripts using our proposed system. Based on the scores, we have calculated the Absolute and Relative Errors, and found table I.

We calculated the Absolute error using the formula 1 and relative error using formula 2.

$AbsoluteError = ActualValue - MeasuredValue$ ...(1)

$RelativeError = \frac{AbsoluteError}{ActualValue} * 100$ .....(2)

Where in formula 1 actual value is the score given by the teachers and measured value is the score given by the system. We also find the relative error of answer script of each or the questions are given in fig 3 which is the graph of Questions Vs Relative error.

As resources in Bangla Language are hard to find and there is no work found in past our work is just an initiative. But the relative error which is observed in our system can be accepted as it is just the beginning. We hope to improve our work in the future. In the future implementation of the whole idea is needed to make it automated and faster. New parameters can be added to the system to make it more reliable and synonyms of words can be checked. Machine Learning algorithms can be used to make the system more efficient and effective. Our proposed method should be implemented for various type of sentences. Input should be taken from handwritten answer scripts to make it more applicable in the real word.

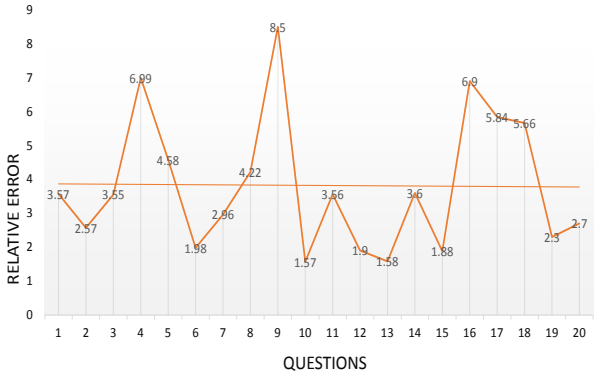

Fig. 3. Questions Vs Relative error

V. Conclusion

We proposed a basic method to evaluate Bangla subjective answer script for the first time. The relative error is under 10% of our proposed methodology which is quite satisfactory. As it is just the beginning, there are still many options or ways to improve the experimental methodology we proposed. We know that working for the Bangla language is more difficult then English language due to its minimum amount of resources. We hope our proposed system will save some times and utilize the current resources.

References

[1] Wikipedia. (2015) Education in bangladesh. [Online]. Available: https://en.wikipedia.org/wiki/Education_in_Bangladesh

[2] BANBEIS. (2017) Bangladesh education statistics 2016. [Online]. Available: http://lib.banbeis.gov.bd/

[3] AsiaNewsNetwork. (2017) Teacher-student ratio worsens in bangladesh. [Online]. Available: http://annx.asianews.network/content/teacher-student-ratio-worsens-bangladesh-46376

[4] A. Benomran and M. Ab Aziz, "Automatic essay grading system for short answers in english language," *Journal of Computer Science*, vol. 9, pp. 1369–1382, 09 2013.

[5] K. Meena and L. Raj, "Evaluation of the descriptive type answers using hyperspace analog to language and self-organizing map," in *2014 IEEE International Conference on Computational Intelligence and Computing Research*, Dec 2014, pp. 1–5.

[6] K. Wangchuk, "Automatic answer evaluation: Nlp approach," 05 2016.

[7] M. Rahman and F. Hasan Siddiqui, "Nlp-based automatic answer script evaluation," vol. 4, pp. 35–42, 12 2018.

[8] V. Nandini and P. Uma Maheswari, "Automatic assessment of descriptive answers in online examination system using semantic relational features," *The Journal of Supercomputing*, 05 2018.

[9] M. Prakruthi S T, "Automated students answer scripts evaluation system using advanced machine learning techniques," *International Journal for Research in Applied Science and Engineering Technology*, vol. 6, pp. 1794–1797, 06 2018.

[10] P. Sinha and A. Kaul, "Answer evaluation using machine learning," 03 2018.

[11] C. Roy and C. Chaudhuri, "Case based modeling of answer points to expedite semi-automated evaluation of subjective papers," in *2018 IEEE 8th International Advance Computing Conference (IACC)*, Dec 2018, pp. 85–90.

[12] Akkhor. (2019) Akkhor bangla spell and grammar checker. [Online]. Available: http://www.akkhorbangla.com/