

AUTOMATED SHORT-ANSWER GRADING SYSTEM USING MACHINE LEARNING: A REVIEW

Shivam Moray^{*1}, Sushila Choudhari², Nikita Salunke³ and Anand Dhawale⁴

^{*1}Department of Computer Engineering, Modern Education Society's College of Engineering, Pune, Maharashtra, India.
moray.shivam@gmail.com¹

²Department of Computer Engineering, Modern Education Society's College of Engineering, Pune, Maharashtra, India.
sushilasc101@gmail.com²

³Department of Computer Engineering, Modern Education Society's College of Engineering, Pune, Maharashtra, India.
nikitavsalunke@gmail.com³

⁴Assistant Professor, Department of Computer Engineering, Modern Education Society's College of Engineering, Pune, Maharashtra, India.
anand.dhawale@mescoepune.org⁴

Abstract: In this paper, a broad overview on the design and implementation of an automated evaluation system for short technical answers by using Machine Learning, along with a comparative study of the various approaches that have been used till current date, is provided. The prevailing system has its own pause in terms of volume, staffing, variation, correctness within the ways of assessing. The proposed system is an evaluation system which can identify the matching keywords in the textual answers and evaluating marks for the same based on some previous knowledge acquired by the machine. A syntactical relation-based feature extraction technique is proposed for automatic evaluation of descriptive type answers. Students are also benefited with a feedback system that can help them improve their score. It also provides a platform to academic institutes to enhance their system which can provide better results in assignment evaluation.

Keywords: Machine Learning, Natural Language Processing, Automatic assessment, Feature extraction, Textual pattern.

INTRODUCTION

The ability to communicate in natural language has long been considered a defining characteristic of Artificial Intelligence. In several cases, the answer rating task prices large human resources, however with less potency and therefore the score given by human authority is generally determined by his data, feeling and energy. There is also an enormous deviation between the scores evaluated by totally different rates. Thus, the correctness of grader rating system can't be guaranteed. In earlier times, the work of assessment in academic terms was heavily reduced by conducting a Multiple-Choice Questions (MCQs) test, whose answers only comprised of a single word or a short text. The modern era demands for the same efficiency in case of short technical answers. The automatic answer evaluation system has been dealt in numerous forms, like question responsive system, essay grading system over time. All the previous works related to this project throw light upon the key concept of evaluating the natural or linguistic language answers and providing a grade accordingly.

RELATED WORK

The basic concept of using computers to increase our understanding of textual features has long been considered as an additional benefit in comprehending written text.

Marti. A. Hearst [1] in her research developed a project called the Essay Grader, where she applied multiple linear

regression to determine optimal combination of weighted features. This system was vulnerable and faced many challenges. She then shifted her focus on developing another approach to assess more direct measures of writing quality called Latent Semantic Analysis (LSA).

Huang Houkuan [2] introduced a general process text machine learning at first, and thereby suggested a method for transforming text categorization problem to a series of binary classification problems using Support Vector Machine (SVM). For feature extraction the concept of stemming was used. However, this approach wasn't suitable for larger dataset as training time with SVMs can be high.

P. Selvi [3] presented a system based on combination of novel approach and latent semantic analysis. The novel approach separates composite and primitive features whereas the LSA module determines the number of words after applying stemming. The combination of the above 2 stated approaches shows greater efficiency and proves that combining various algorithms is a possible strategy asses' student's answer.

The TOEFL exam proved to be one of the best examples of Automated Essay Grading (AEG). Siddhartha Ghosh [4] proposed an AEG system which brought significant changes and gave a new shape to Indian Text Categorization and Machine Learning research work. Independent Bayesian Classifiers allow assigning probabilities to documents estimating the likelihood inferring they belong to specific classes. The diagnostic feedback was based on a suite of

programs that identify the essay’s disclosure structure, recognize undesirable stylistic features, and evaluate and provide feedback on errors in grammar, usage and mechanics. The proposed framework intended to capture the mental status of the student (Psychometric Analysis). Automated Marking System for Short Answer Examination (AMS-SAE) [5] is a system that has been developed to grade student’s answers based on given marking scheme. Fatimah Dato’ Ahmad developed an Automated Text Marker (ATM). ATM is a marking system that uses the language structure to compare the sentences.

P. A. A. Dumal proposed a system [6] to overcome the issues in existing similar systems. Answer extraction and comparison of similarity were the two methodologies which comprise of the core part of the project. Answer extraction of both student and tutor was done in three steps which were splitting of sentence, tokenizing and Part-of-Speech (POS) tagging. In this project solution the tutors can add questions to the question bank, making assignments and obtaining the final scores automatically. The learners are facilitated too in providing the answers for assessment and getting the feedback immediately.

The automated short answer evaluation system [7] proposed by Sijimol P. J., can identify the text in answer papers and evaluates marks for each short answer based on previous knowledge acquired by the model. In the proposed system Optical Character Recognition tools are used to extract the hand-written texts. The system works based on Machine Learning. It trains a model from the scored short answer paper dataset and a high weightage given key.

PROPOSED SYSTEM

In the proposed system an automatic grading system for descriptive answers is presented. The system takes in student’s answers as input and asses them based upon the predefined answers. The answer is scored depending upon the context of the answers and the structure of the answer.

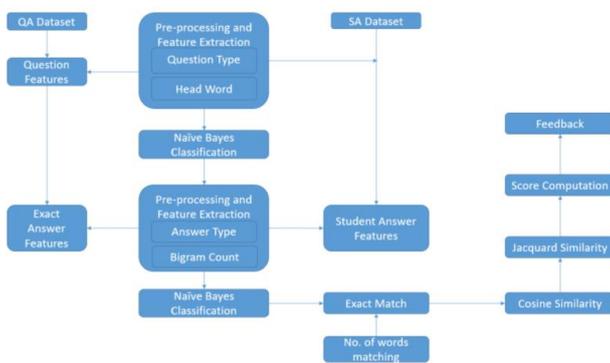


Figure 1. Proposed System Architecture

The Figure 1 essentially states the stages in the evaluation of the answers. To assess the performance of the student a novel approach is been presented. The system proposed predominantly focuses to evaluate the answers in the natural way of representing them. Every question has an answer

which is written in a form suitable for the question. The challenging part is, the machine should understand the exact inference of the information that is being conveyed by the student through his answers. The questions and their answers, with the student’s answers is given as the input. The system is organized into training and prediction sections. The training section includes:

- I. Question Classification
 - II. Answer Classification
- The prediction section includes:
- III. Answer evaluation, scoring and providing feedback.

I. Question Classification

It becomes mandatory to classify the question prior to classify and evaluate the answers. It would only be eloquent if the answers that are being written by the students are in correspondence with the question being asked.

Following are the steps involved in processing of the question module:

1. Finding the focus of the question.
2. Classifying the question.
3. Determining the expected answer.
4. Determining the semantic context of the answers

The preliminary stage is to find out the type of question which is asked, and this will help us to find out the nature of answer. This will help to elucidate the correspondence with the question.

In Feature Extraction the questions are chosen from the database and three key features are extracted from each question.

Key feature 1 corresponds to question-type analysis. It categorizes the questions into three types: Factual, Inductive and analytical types.

The factual question requires the student to retrieve facts given in the questions. This type of questions can be identified by the keywords who, where, what or which. These keywords can be isolated from the questions by entity recognition or using stemming process.

Factual questions

Example:
 What is Internet of Things?
 What is Intranet of Things?

The tag ‘what’ to the question in the above questions imply that the answer expected is the explanation of the entity mentioned in the question.

Inductive questions

Example:
 Why do we need Internet of Things?
 How are Things connected in an Internet of Things?

From these question tags such as the ‘why’ or ‘how’, it is evident that here the questions require the student to present the detail facts as well as other entities too.

Analytical questions:

These questions are having main keywords like Distinguish between..., Compare...etc. These questions not just ask the

students to present facts but also want to them to find out the relation between the entities mentioned in the question. It can test the analytical part of the student's answers.

Key feature 2 represents the identification of the headword from the question. Headword is a single word that specifies explanation in the question asked. Parses can be used to find out the headword of the sentence. The relationship of the headword with the question can play a role in finding out the headword.

Key feature 3 is the type of answer the student is expected to write. These features will help to provide meaning to the relationship between the question and answers.

II. Answer Classification

In this part of classification, we will check if the skill of student in writing skill and presenting of facts through his answers. There are two types of cognitive-based nonfiction textual pattern for textual structures namely, deep structure and surface structure. Here we are only going to consider the surface structure. Surface structure identifies pattern structures and phrases from the answers. It refers to the way in which student organizes his/her answer. The system notes the organizational pattern and hunts for signals or words to evaluate the content used by the learner. For this analysis the classification is focused on five categories:

1. Chronological ordering

It shows ordering of the texts from start till the end.

Example: First, Last, Next, etc.

2. Descriptive

The descriptive details asked in the question are identified.

Example: For example, For instance, Also, Another, etc.

3. Compare and contrast

The ideas are described to reader through the answers.

Example: Same as, Similar to, etc.

4. Problem/Solution

The question asked wants the students to address a problem. Words describing the solution are to be found in the answer.

Example: The solution is, one answers is, etc.

5. Cause and effect: The question demands the student to write the answers which show the occurrence of an event and the causes or effects of it are asked.

6. Example: because, since, yet, to, because of, etc.

III. Answer Evaluation and Feedback

The evaluation of the answers is performed by juxtaposing the model answer with the student answer. The comparison will extract all the necessary features from the student answers. The answer is then further classified into the categories of description, cause and effect, etc. The answer written by the student is compared with the model answer given. The criteria for the evaluation of the answer here is, first the answer type and the question type should match. Secondly, the keywords or facts which are asked in the question need to be addressed. Signal words as well as the phrases are used to score the answers.

Cosine similarity and Jacquard similarity are utilized to ascertain the comparison and relatedness between the

sentences. The final score is given to the student along with the feedback based on the analysis of the answers.

Analysis of the answers can give a detail view of his/her strong or weak areas in the test. This will create a good learning environment for the students. The feedback to the student can be in form of graph, charts or pie diagrams.

CONCLUSION

The present-day evaluation system confronts more challenges in grading the answers written by the students. There are a lot of issues associated with the scheme of the manual evaluation. Substantial resources are required as well as proper techniques and time-consuming to overcome these difficulties out of the current evaluation system, an automated evaluation system has been proposed for evaluating descriptive type of short answers. If the answer written by the student falls into the category of the question which is been asked, then it reflects that the student has certain knowledge in expressing about the topic. Further research can be made to improve the grammar for a language which includes the different ways in which students write.

ACKNOWLEDGMENT

We would like to thank our Guide Prof. Anand D. Dhawale and Computer Department Head of Department, Dr.(Mrs.) N. F. Shaikh. We thank them with all our heart and express our honour and deep gratitude for helping us in every way possible. We furthermore thank the teaching and non-teaching staff of Computer Department. We would like to thank all those, who have directly or indirectly helped us for the completion of the work during this paper.

REFERENCES

- [1] Hearst, Marti A. "The debate on automated essay grading." IEEE Intelligent Systems and their Applications 15.5 (2000): 22-37.
- [2] Jun, Hu, and Huang Houkuan. "An algorithm for text categorization with SVM."TENCON'02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering. Vol. 1. IEEE, 2002.
- [3] Selvi, P., and N. P. Gopalan. "Automated writing Assessment of Student's Open-ended Answers Using the Combination of Novel Approach and Latent Semantic Analysis."Advanced Computing and Communications, 2006. ADCOM 2006. International Conference on. IEEE, 2006.
- [4] Ab Aziz, M. J., Dato'Ahmad, F., Ghani, A. A. A., & Mahmod, R. (2009, October). Automated marking system for short answer examination (AMS-SAE). In Industrial Electronics & Applications, 2009. ISIEA 2009. IEEE Symposium on (Vol. 1, pp. 47-51). IEEE
- [5] Ab Aziz, Mohd Juzaidin, et al. "Automated marking system for short answer examination (AMS-SAE)."Industrial Electronics & Applications, 2009. ISIEA 2009. IEEE Symposium on. Vol. 1. IEEE, 2009.
- [6] Dumal, P. A. A., et al. "Adaptive and automated online assessment evaluation system."Software, Knowledge, Information Management and Applications (SKIMA), 2017 11th International Conference on. IEEE, 2017.

- [7] Sijimol, P. J., and Surekha Mariam Varghese. "Short Answer Scoring System Using Neural Networks." (2018).
- [8] Nandini, V., and P. Uma Maheswari. "Automatic assessment of descriptive answers in online examination system using semantic relational features."The Journal of Supercomputing(2018): 1-19.
- [9] Nehete, Charusheela, et al. "Checkpoint--An Online Descriptive Answers Grading Tool."International Journal of Advanced Research in Computer Science 8.3 (2017).
- [10] Qureshi, Jameel, and M. Rizwan. "a proposal of electronic examination system to evaluate descriptive answers."Science International 27.3 (2015).
- [11] Paul, Dimple V., and Jyoti D. Pawar. "Use of syntactic similarity based similarity matrix for evaluating descriptive answer. "Technology for Education (T4E), 2014 IEEE Sixth International Conference on. IEEE, 2014.
- [12] Chakraborty, Udit Kr, Rashmi Gurung, and Samir Roy. "Semantic Similarity Based Approach for Automatic Evaluation of Free Text Answers Using Link Grammar. "Technology for Education (T4E), 2014 IEEE Sixth International Conference on. IEEE, 2014.

SHORT BIODATA OF ALL THE AUTHOR

Mr. Shivam Moray is a student of final year engineering. His area of interest includes Machine Learning and Data Science.

Ms. Sushila Choudhari is a student of final year engineering. Her area of interest includes Artificial Intelligence and Text Retrieval.

Ms. Nikita Salunke is a student of final year engineering. She has interest in and Cryptography and Network Security.

Prof. Anand D. Dhawale is Assistant Professor at Modern Education society's College of Engineering, Pune. His research areas include Machine Learning and Artificial intelligence and Internet of Things. He has published over 10 research papers in various national and international journals.