INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 4, Issue 1)

Available online at www.ijariit.com

Short Answer and Essay Scoring using Modeling Technique

Aratrika Saha
saha1995aratrika@gmail.com
Technique Polytechnic Institute, Hooghly, Panchrokhi West Bengal

ABSTRACT

A decorous grading system is very paramount in any institution, especially such institutes which gives short answers and essay type questions in the examination. The purpose of this research is to find an approach to automate the grading of short answers and essay type questions. In this work, Natural Language Processing algorithm is used to perform the above-said job using Modeling technique.

Keywords: *Essay Scoring Algorithm, Grading, Machine Learning, Natural Language Processing, Modeling.*

1. DETAILED DISCUSSION

The education community is growing endlessly with a growing number of students, curriculums, and exams. Such a growing community raised the need for scoring systems that ease the burden of scoring numerous numbers of exams and in same time guarantees the fairness of the scoring process. This primarily is the motivation behind this research work to develop an automated essay scoring system.

1.1 Introduction

Automated grading, if proved to be matched or exceed the ability of human graders, will reduce costs in a huge percentage. The purpose of this research is to implement and train machine learning algorithms to automatically asses and train machine learning algorithms to automatically asses and grade essay responses. These systems used for scoring the essays are commonly called as Automated Essay Scoring or AES. Some typical features of these AES are:

- Feedback – In many applications of AES, as in a "second reader" role on GMAT, feedback isn't as important. But in all the cases where a student will be learning based on how their response is scored, feedback is absolutely important.
- Iterative – One of the main advantages of AES is that students can submit their response as many times as they want, getting very quick feedback and scoring each time.
- Correct – Automated essay scoring is not useful if its scores are only marginally accurate, but its utility goes away if it can't score properly. So to get a perfect score automated essay scoring should be absolutely correct.

This work basically implements Machine Learning and Natural Language Processing techniques (NLP). A brief description of both is stated below.

Machine Learning – Machine Learning is a type of artificial intelligence (AI) that provides the ability to learn without being explicitly programmed. [1]

NLP – Natural Language Processing is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages. [2]

1.2 Overview

The basic idea is, a reference question and a reference answer will be given to the system; the system tries to build a model using some training dataset. The training dataset is nothing but some arbitrarily chosen answers which are received after taking the test. Once the model gets created then it is used to score the rest of the papers.

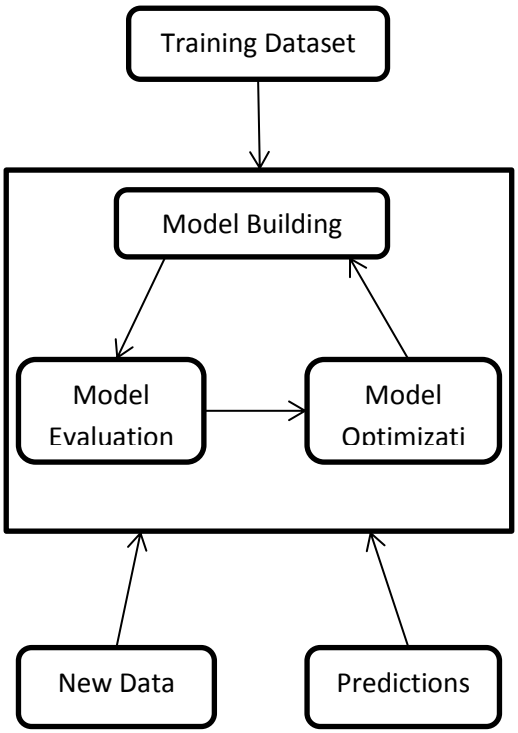

Chart 1:- Implementation Methodology

The backbone of this work is modeling. A learning model is first developed and then that model is used to score the student responses. The model creation is done semantically matching the reference answer with the student response (i.e. how semantically similar the student response is to the reference answer).Semantic means relating to the meaning in language or logic. Semantic matching is a technique used in computer science to identify information which is semantically related [3].The semantic checking is done by extracting some predefined features from both the reference answer and the student response. These features are cosine alignments and the length of the answer. Predefined weights are put for each and every word in the reference answer, then that is checked and a score is decided and put in the model. By this process, a regression model gets created. This means that a function gets created which is based on the characteristics of the data received from the training data set. The training data set in this case is the first few papers that are fed into the machine. The training dataset is not stagnant in this case. A detailed discussion for this is done in the next paragraph. The training dataset does not remain constant because once the model receives a student response it tries to match it with the model and find similar occurrences so that it can predict the score. If none of the occurrences match then it adds that particular occurrence in the trained model and updates the regression model. Since it always tries to update the model so the training dataset also does not remain constant. Another insight about the training dataset is that suppose same marks is getting awarded for more or less similar answers. The system will try to find out that which answer is been written using the words having more weight and which one is written with more words but not the words having weight. In this case, the one written with words having more weightage will be preserved.

1.3 Algorithm

Step1: A reference question and a reference answer are fed to the system as an initial step.

Step2: The first few student responses are treated as a training dataset. By this, the machine learns and creates a Regression Model.

Step3: Using the regression model further scores is predicted.

Step 4: If the student response received does not match with any response previously recorded in the model then that particular response gets recorded and the training model gets updated.

Step 5: Finally a score gets assigned to each student response.

1.4 Resources Used

Hardware Used:

Processor – Intel Core i7

RAM – 8.00 GB

System Type – 64 Bit Operating System.

Software Used:

Operating System – LINUX (Ubuntu)

Programming Language – Python 2.7

Packages – Stanford CoreNLP, NumPy

1.5 Illustration

This section contains the results observed for different student responses received by the system that has been checked using this system. All the observations are listed below.

1st Evaluation: This evaluation shows an example where the student response is exactly same as the reference answer. The input is shown in fig 1 clearly shows that that student response is exactly same as the reference answer. The expected score should be 100% since the student has exactly given the same answer as the reference given. Fig 2 shows the output which is same as the expected output i.e. 100%.

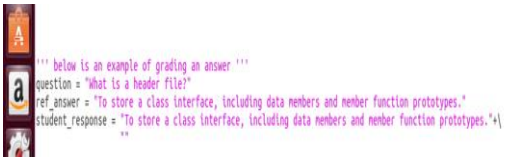

Fig 1: 1st Input

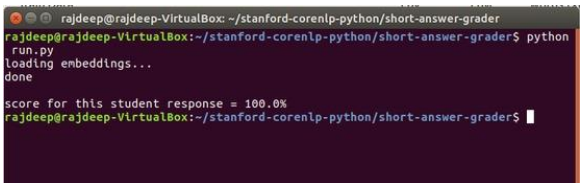

Fig 2: 1st Output

2nd Evaluation: This evaluation shows an example where the student response is not exactly same as the reference answer but the meaning remains same and all the important words are used too. The input is shown in fig 3 clearly shows that that student response is not exactly same but is same logically. The expected score should be 100% since the student has written an answer which contains all the keywords mentioned in the reference answer only the grammatically it is different from the reference answer. Fig 4 shows the output which is same as the expected output i.e. 100%.

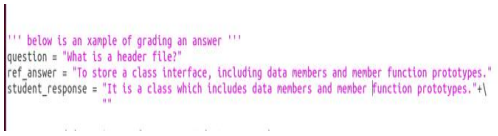

Fig 3: 2nd Input

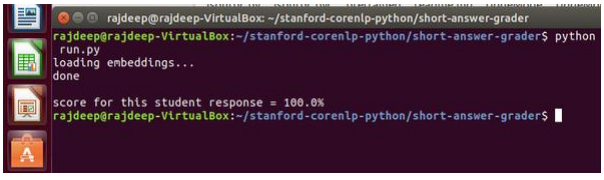

Fig 4: 2nd Output

3rd Evaluation: This evaluation shows an example where the student response is partially correct .The input shown in fig 5 clearly shows that that student response is partially correct which means that the student did not use all the keywords mentioned in the reference answer. The expected score should be less than 100% since the student has given an incomplete answer. The score completely depends on the weight of the keyword used by the student. This means that the system scores in such a way that it allots marks according to the importance of the keywords previously decided by the examiner. Fig 6 shows the output which is less than 100%. The score is completely decided by the keyword used by the student and its weight.

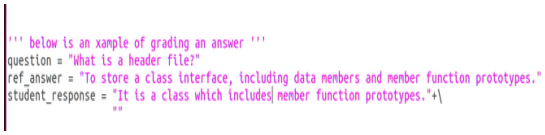

Fig 5: 3rd Input

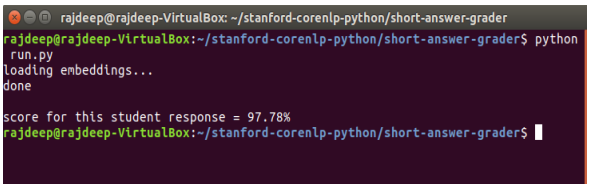

Fig 6: 3rd Output

4th Evaluation: This evaluation shows an example where the student response is partially correct. The input is shown in fig 7 clearly shows that that student response is partially correct which means that the student did not use all the keywords mentioned in the reference answer. The expected score should be less than 100% since the student has given an incomplete answer. The score completely depends on the weight of the keyword used by the student. This means that the system scores in such a way that it allots marks according to the importance of the keywords previously decided by the examiner. Fig 8 shows the output which is less than 100%. The score is completely decided by the keyword used by the student and its weight. The 3rd evaluation also shows that the student answer is partially correct and here also the answer is partially correct, but still, the score for both the answers are not same. This is because of the scoring completely depends on the weights of the keywords used. Since the keywords used for both the evaluations are different so the scores allotted are also different.

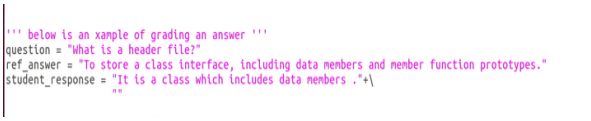

Fig 7: 4th Input

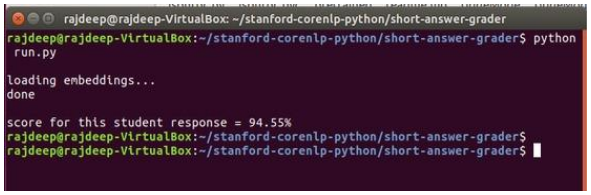

Fig 8: 4th Output

2. CONCLUSION

This process integrates Stanford NLP parser to get the results which are advanced in terms of time complexity. But Stanford NLP parser also has some bottleneck as it only works on a particular set of Parts of Speech and complex sentences can lead to failure. Apart from this bottleneck, the system works pretty efficiently and accurately.

It is safe to conclude that even though automation of scoring application for student-written essays and short answers is a challenging job but still this approach works pretty well in terms of time complexity to get output in real time and also gives outstanding results for simple sentences.

3. ACKNOWLEDGEMENT

I would like to thank Technique Polytechnic Institute, Hooghly, West Bengal, India for all their support without which this paper would not have been possible.

4. REFERENCES

[1] https://www.coursera.org/learn/machine-learning
[2]https://en.wikipedia.org/wiki/Natural_language_processing
[3] https://en.wikipedia.org/wiki/Semantic_similarity

