Automated Answer Grading

Lakshmi[1], B.Vaishnavi[2], P. Vamsi Chakravarthy[3]
[1]Assistant Professor, Cse Department, Anil Neerukonda Institute Of Technology And Sciences, Visakhapatnam, Andhra Pradesh, India
[2,3]4/4 Cse Student, Cse Department, Anil Neerukonda Institute Of Technology And Sciences, Visakhapatnam, Andhra Pradesh, India

Abstract- Automated answer grading (AAG) is a broadly used application of machine learning, that impacts decision-making for students. This project examines the issues related to Online Grading System. This would create a system of reliable and accurate to compute grades in different subjects. Manual grading takes significant amount of evaluator's time & hence it is an expensive process. Automated grading if proven effective will not only reduce the time for assessment but comparing it with human scores will also make the score realistic. A comparatively fast access of information of grades generate reports, and information of the input of teachers. Linear regression technique will be utilized for training the model along with making the use of various other classifications and clustering techniques. We intend to train classifiers on the training set, make it go through the downloaded dataset, and then measure the dataset, and then measure performance of our dataset by comparing the obtained values with the dataset values.

I. INTRODUCTION

1.1 What Is Automated Answer Grading

Automated answer grading system is developed for grading of short answer questions and providing useful feedback on their answers to students. We consider an approach that groups few answers graded by the teachers into clusters. Each cluster would be awarded the same mark, and the same feedback is given to each answer in the cluster. Then each of the remaining answers are evaluated by the model and is assigned to an existing cluster and will be graded accordingly. The feedback of the essay is also generated which may help the student.

1.2 Data Representation

In this model, each document is represented as a point in the n-dimensional space, where each feature points a dimension. The idea of this model is to extract features of word count, sentence unit, vocabulary, spelling mistakes from each of the collection of documents and position them in the feature space. Each document is represented as a vector of features. The magnitude of the vector in each dimension represent the strength of presence of the features. As the documents are represented in the space they are grouped as clusters. Each cluster will be given scores as per training.

II. METHODOLOGY

2.1 Working On Clustering:

k-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed, and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more.
In our work on clustering, we perform one of widely applicable distance measure: Euclidean distance the distance between these documents is described as follows:

$$d(\mathbf{r}_1, \mathbf{r}_2) = \sqrt{(r_{11} - r_{21})^2 + (r_{12} - r_{22})^2 + \cdots + (r_{1N} - r_{2N})^2}$$

As long answers have more words than short answers, the number of non-zero entities of features and, frequencies may be more compared with short answers with vector representation. To adjust the effect of length, term frequencies should be normalized. Given an answer r, the L2-normalization is defined as:

$$\hat{\mathbf{r}} = \frac{\mathbf{r}}{d(\mathbf{r}, \mathbf{0})},$$

After defining k centroids, each document is assigned to a cluster by using the distance d. Then, the centroids are recalculated until we find an optimal set of clusters based on some criterion function.

2.2 Steps Of K-Means:
Specify number of clusters K.
Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
Keep iterating until there is no change to the centroids. i.e assignment of data points to clusters isn't changing.
Compute the sum of the squared distance between data points and all centroids.
Assign each data point to the closest cluster (centroid).
Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster.

III. ALGORITHM:

3.1 During Training:

1. Grade few randomly selected answers by the teacher and provide those as input.
2. Decide the number of clusters by the max mark factor (Ex:10 for max marks of 100)
3. Each of the answers is extracted of its features and is plotted on the n-dimensional space based on the vector formed of its feature magnitude.
4. Apply k-means clustering algorithm to these spatial points as per the number of clusters decided in the step 2.
5. Assign each of the cluster a central score based on the marks awarded by the teacher.

During Testing:
All the remaining answers are the input of the test phase.

1. Each of the remaining answers are also extracted of its features and plotted on n-dimensional plane.
2. The model then decides on which cluster to assign this point.
3. The answer is graded based on the Euclidean distance from the assigned cluster.
4. The feedback is provided to the student.

3.2. Architecture

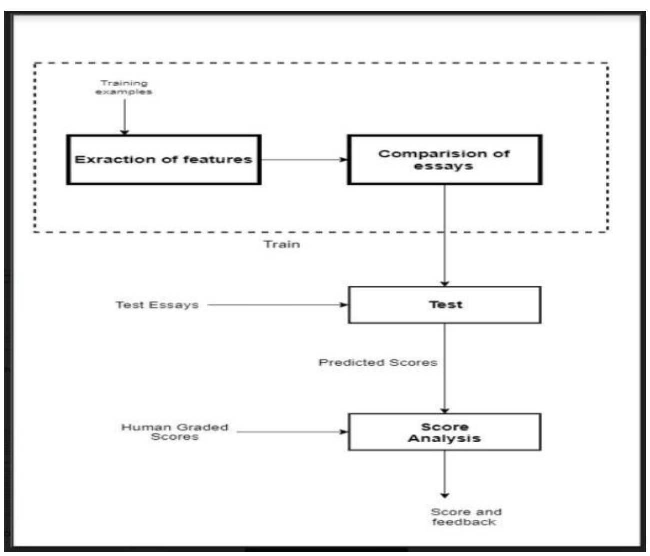

Fig 3.1 System Architecture

IV. FEAUTRES CONSIDERED:

4.1. Word count and Sentence count:

These are trivial features of any text document, but it influences the scoring of the document as well. Now, since these stop words are of not much use, we can skip them while we are calculating the word count of each document from the term document matrix. To get the sentence count, we simply break the document using '.' and thus, count the number of segments obtained.

4.2. POS Tag:

To evaluate the quality of content in an essay is the foremost important task. Along with it another important set of features for evaluating any piece of writing provide is the number of words in various syntactic classes like nouns, adverbs, verbs, adjectives etc. should also be evaluated. To get the counts of words in each POS (part-of-speech) class few library provides us the POS tag for each word in an essay, and thus, we can extract the number of nouns, adverbs, adjectives and verbs separately.

4.3.Spelling Mistakes:

An important parameter while grading an essay is the spelling mistakes. Therefore, number of spelling mistakes in an essay has also been included as a feature for our model.

4.4.Vocabulary Usage:

This is an important feature while evaluating an answer because similar words or the synonyms of the word is considered based on which the score will be predicted.While implementing this feature we will consider only few parts of speech like verb, noun ,adjective etc.

4.5.Feedback:

While evaluating, answers with the same grades are given the same feedback based on the spelling mistakes, word count, sentence count. Feedback may be positive or negative based on the scores that are been evaluated for the answers.

Sample Input :

The input of this model is in the form of text documents or with scanning through camera.

Sample Output :

The output is the grade or marks that is obtained for the given text i.e. the marks scored for each answer scaled on maximum marks allocated for that answer.

V. CONCLUSION AND FUTURE WORK

With this project we will have a great and a wonderful opportunity to work on machine learning. Our project will be useful when deployed in a real-time scenario as in schools, colleges, universities and examination boards. We can further improve our model to recognize and understand all styles of handwritten text.

VI. REFERENCES

[1] Valenti.S, Neri.F and Cucchiarelli, An Overview of Current Research on Automated Essay Grading

[2] Kaggle 2012 The Hewlett Foundation: Automated Essay Scoring

[3] Lukic A and Acuna V Automated Essay Scoring Rice University

[4] Preston D and Goodman D. Automated Essay Scoring and the Repair of Electronics.