

Automated Evaluation of Telugu Text Essays Using Latent Semantic Analysis

M Varaprasad Rao^{1*}, B Kavitha Rani², K Srinivas³, G Madhukar⁴, A.Anusha⁵

^{1,2,3} Professor, CMR Technical Campus, Hyderabad

⁴ Assistant Professor, CMR Technical Campus, Hyderabad

⁵ Assistant Professor, JB Institute of Engineering & Technology, Hyderabad

¹varam78@gmail.com, ²phdknr1@gmail.com, ³phdknr@gmail.com, ⁴madhu.mani536@gmail.com,

⁵anushampavathi@gmail.com

Article History: Received: 11 January 2021; Accepted: 27 February 2021; Published online: 5 April 2021

Abstract: The most productive strategy to improve students' ability to write is to have direct and as much as possible teacher input. However, the workload of the teacher is greatly increased. Automated systems are increasingly required to help students write essays. In the field of educational assessment technology, automated test evaluation is becoming more and more common. We present a framework that is modelled on the programme, following which the school-teachers in the BPDVA School and Govt. High School Hyderabad, Telangana, India present the automatic evaluator of student essays in the Telugu language. Language skills; the structure of the essay and the contents that fit the subject are the principal requirements for evaluating the essays. In this context, we have established a scheme focused on latent semantical analysis and the theory of rhetorical structure. The method has been evaluated in more than 600 different essays, written in different manuscripts by schoolchildren. Overall 0.82 with the teacher's assessment was achieved in our method.

Keywords: Automatic evaluation of essay, latent semantic analysis, rhetorical structure theory

1. Introduction

Automatic evaluation of essays (AEE) is a computer program that evaluates essays written by students. Automated essay feedback has appeared with the advent of online education systems. There is an increasing need for such systems to help students draft essays. A field of Natural Language Processing (NLP), AEE automatically provides feedback for essays that students write in natural language. The AEE system is also known as automated essay scoring and, automated essay grading.

In 1966, Ellis Page published a paper, "The Imminence of ... Grading Essays by Computer," which discusses the use of computers to evaluate essays and provide feedback [6]. Page published the article to explain his ideas for the development of Project Essay Grade (PEG) [7]. Some additional forces at work ultimately facilitated the future of AEE. These included the creation and widespread adoption of the AEE systems to evaluate writing in different languages.

We can only describe AEE as a means of automated computer evaluation of the written prose[10]. Evaluation means that the computer system will do the job of scored or assigned a number to an essay. To improve writing quality more specifically and because the large scale testing programmes for English, such as TOEFL and GMAT, are required, AEE systems are available. Effective AEE systems are applied in many ways, and many active AEE implementations exist. Effective AEE systems are implemented using different NLP techniques which include elements of IR and machine learning (ML).

AEE systems for the Telugu language need to be developed. The school performs an Aptitude Test (AT), a kind of online test, which is required for any student seeking school admission. While AT assesses Telugu's competence, there are no essay writings.

The lack of an automated test grading system is responsible. State-wide, this exam is performed annually by over ten lakh students. It is impractical to consider the manual evaluation of the essays because of the sheer number of students taking AT exams. Our system will allow the school to take seriously the automated evaluation of Telugu-language essays.

In this paper, we briefly describe our proposed system for evaluating school children essays in the Telugu language. The paper is organized as follows. In Section 2 the related work is summarized. The design of the system is explained in Section 3. In Section 4, a discussion of evaluation and discussion of the proposed model performance, and finally, the conclusion of the work described in Section 5.

2. Related work

There are many contributions for essays in the English language, see for example [9][2][8], and for other languages, e.g. [3][4]. Latent Semantic Analysis (LSA) has produced promising results in content analysis of essays, e.g. [9][3]. With very few methods, the work on the automated evaluation of Telugu essays lags.

Alghamdi et al. [1] presented a hybrid AEE system for evaluating Arabic essays that makes use of efficiently reduced-dimensionality for LSA (LSAD). Some of the features that were used for the assessment are spelling

mistakes and the proportion of spelling mistakes for the given length of the essay. For the dataset, the authors collected around 600 essays written by the two schools. The essays were part of a test in the Telugu language course. The length of an essay ranged between 100-200 words. It is a two-phase system, the training phase and the testing phase. The training phase involves some pre-processing which includes Buckwalter stemming. This phase is made up of three parts: a bag of words, a vector of spelling mistake, and the LSA concept space. In the runtime phase, the input essays pass through several processes. These processes make it possible to get the minimum cosine distance (cosine distance LSAD) between the input essays and the training essays. The size of the LSAD vector is the essay's score. In this particular system, there are six marks. During this phase, they use a linear regression approach to obtain features that reflect the human senses. The authors reported an accuracy of 96.72% on the test data. The correlation result between this system's score and the evaluation by humans was 0.82.

Nahar and Alsmadi [5] presented a system for grading online exams in Arabic involving essay questions. Unlike multiple-choice questions, where grading is straightforward, this is more challenging. The idea is to score the student answer against the model answer by the instructor. The authors used different statistical distributions to give weights to the keywords in the model answer.

The instructor determines the weights, which tells how important the keyword is. There is a provision to handle synonyms in the student's answer; this, however, requires synonym words to be manually added into the system. To score the student's answer, the system needs to measure the distance between both answers (student and the model). The paper does not go beyond the schemes; it does not evaluate the system on some real exam dataset, so to compare the automatic grading with manual grading.

3. Proposed System

The objective is to develop a system to automate the evaluation of school children essays written in the Telugu language. All the children belonged to the middle level of school. That is covering grades 6 to 9 inclusive. The assessment criteria are based on an online survey of middle-level school teachers in Telangana state. According to the survey, the criteria are: spelling and grammar mistakes, the coherence and organization of the essay, the essay should be related to the topic, and sticking to Standard Telugu words. There was no general agreement on how much weight to assign to each of the criteria, however, the consensus was 3 marks (out of 10) for spelling mistakes, 2 marks for grammar mistakes, and 5 marks for the organization of the essay.

To solve the problem at hand we opted for a hybrid approach that combines latent semantic analysis (LSA), rhetorical structure theory (RST), and some other features that we will cover later in the paper. One reason for this approach is the need to assess essays by focusing on elements such as cohesion. This hybrid approach applies LSA for the semantic analysis of the essay, and the RST to assess the cohesion and the writing style of the essay. In our design, we assign 40% of the total score on the cohesion of the essay, 40% for writing style and the remaining 20% for spelling mistakes.

We already noted that LSA has been successfully applied to automate giving grades and feedback on free-text responses in several systems. The basic assumption behind LSA is that there is a close relationship between the meaning of a text and the words in that text. The power of LSA lies in the fact that it can map the essays with similar wordings closer to each other in the vector space. The LSA method can strengthen the similarity between two texts even when they do not contain common words. The general architecture of multiple processes AEE system is shown below in Figure 1.

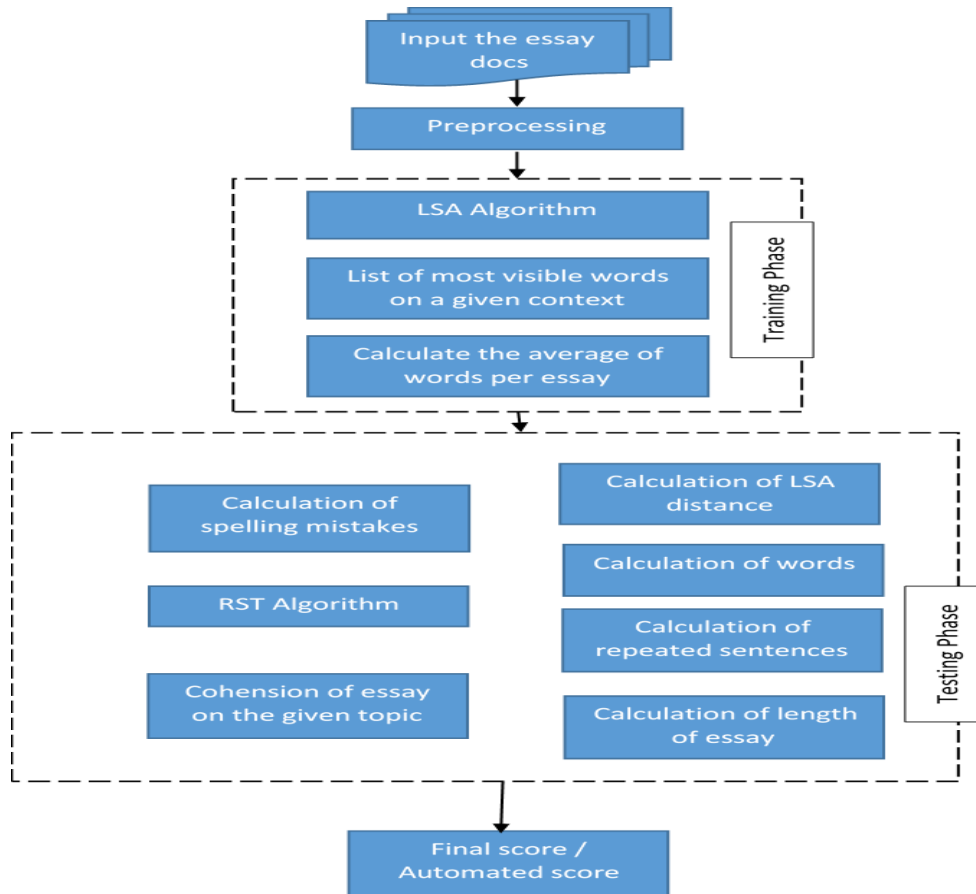


Figure1: Automated Evaluation of Essays

4. Experimental Results

The handwritten essays were part of a typical assignment, so they were graded by the class teacher out of 10 marks. For our evaluation, we had to retype the collected essays on the computer – as is – including any mistake the student might have committed. The total number of essays collected is slightly over 600.

For the evaluation, we use 10-fold cross-validation (CV). The entire set of essays was divided into 10 distinct sets. We use nine sets for training and the tenth for testing. The entire process is repeated ten times, each time picking a different set for testing. We measure the performance of our system using accuracy. For each run, we count the number of differences between the teacher score and the autoscore given by the system. We consider two scores being the same if the absolute difference is less than a threshold, which we set at 1.5 marks. Alghamdi et al. [1], set the threshold at one mark where the essays were graded out of 6 marks. Our essays were marked out of 10 marks, so if we follow Alghamdi et al. [1], the threshold should be 1.67 marks. This means we are using a slightly tighter threshold. The accuracy is given by,

$$\text{Accuracy} = \frac{\text{Number of predictions within a specific range}}{\text{Total number of predictions}} \quad \text{--- (1)}$$

Pearson's correlation coefficient is used to calculate between the teacher's marks (x) and an automated score (y) on a given number of essays (n) respectively using the following formula.

$$\text{Correlation} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2]} \sqrt{[n(\sum y^2) - (\sum y)^2]}} \quad \text{--- (2)}$$

The average accuracy of this problem is 83.36%, for the 10-fold CV, whereas the automated score is 0.828.