

Longest Common Subsequence: A Method for Automatic Evaluation of Handwritten Essays

Reshma.V.M¹, Linda Sara Mathew²

¹Computer Science and Engineering Department, Mar Athanasius College of Engineering Kothamangalam
Ernakulam Kerala, India)

²Computer Science and Engineering Department, Mar Athanasius College of Engineering Kothamangalam
Ernakulam Kerala, India)

Abstract Essays are used to evaluate student's knowledge from early before itself. The aim of the proposed system is to evaluate the handwritten essays automatically. The proposed method is to develop an automated system to evaluate handwritten student essays. For a single topic under study the students may refer more than one study material. According to the similarity of the contents a reference study material is constructed. The reference study material is used to evaluate the handwritten essays. The scanned image of the handwritten essays are used as the input in the proposed method. From the image texts are extracted for the evaluation. The Longest Common Subsequence (LCS) is used for the proposed system.

Keywords – Longest Common Subsequence

I. INTRODUCTION

From the early times itself, essays are used to check the knowledge of students. According to the concepts and contents present in the essay the marks or grades are assigned. The concepts and contents are evaluated according to the meaning of the sentences of essay. For a human reader, this process can do without having an extra effort. It is not requiring any extra training. If it is possible to construct an automatic system for the evaluation of the essay, it will be helpful for students as well as to the teachers. But, the evaluation of such essays are more difficult and demanding task than the evaluation of the single word answers like multiple choice questions.

The proposed system introduces the evaluation of handwritten essays like the normal human evaluation of the answer sheets. The aim of the proposed system is to check whether the automatic evaluation process is similar to human evaluation or not, how much accuracy can give by the system etc. For the proposed system, the scanned image of the handwritten essay is using as input. From this image, the texts are extracted and this extracted text is evaluated against the reference study material. Similar to the answer keys used in the exam evaluation, the reference study material is used to compare the student's handwritten essay. In the field of image processing and pattern recognition, handwritten recognition is a difficult task.

Student's knowledge is obtained by reading the study materials. For a single topic under study, the students may refer more than one study materials. If the student is referring more than one study material, the evaluation has to perform by comparing the essay with all these study material. The process will become more difficult and time consuming. In order to avoid this problem it is better to construct a single study material from all these study materials which contains all the points discussed in every study material. Sentences from different study materials are compared sentence by sentence by checking its similarity in meaning. According to the similarity of the sentences a new reference material is constructed which is used in the evaluation of the essay. For the proposed concept, the method called Longest Common Subsequence (LCS) is using.

Evaluation of the essay made in the next step. The grade is calculated by the use of Longest Common Subsequence method. The extracted text from the scanned image is compared with the constructed reference study material as they are doing in the reference material construction. Finally the grade is calculated and displayed.

II. BACKGROUND

Evaluation of essays is a demanding job for human beings. The methods of automatic assessment of documents are introduced in order to reduce the cost of the grading of those documents. Approaches based on the surface features and content have been introduced early before itself [1]. The system uses the method called Latent Semantic Analysis (LSA) for the essay content evaluation. Previously, the method LSA is made for information retrieval. Later it started to use for the essay evaluation in many applications. LSA will help to

compare the similarity of the concepts and the contents between essays. In the experiment the correlation between the scores given by the system and human grader varied from 0.78 to 0.82.

Another commercial software is Intelligent Essay Assessor (IEA). The IEA can be used to accurate assessment of the essay which performs similar to a well trained human being. It was using in many real life applications like Army's, Air force's, general staff colleges etc. This IEA is well suited for the distant education purposes. It gives a feedback for the essay within very few times [2].

In 1966 Ellis page developed Project Essay Grader (PEG).The purpose was to make an assessment of essays in large scale. It gives a very good quality for the essay evaluation procedure. An advantage of PEG is, the calculate grades are comparable with the human evaluated grades. It produces almost similar grades for the essay [3]. But, the drawback of the PEG is that , it does not give any importance to the contents present inside in the essay.

Another research tool is BETSY. BETSY is developed by Lawrence M Rudner. The other type of essay evaluation is implemented is AES. AES uses the concept of Bayesian theorem [4]. It has a number of applications. Some of them are identifying spam mails, sorting of resumes for job applications etc.

Another difficult and challenging job is the text extraction from the images. The process becomes more difficult because of the difference present in the images like scripts, styles, fonts, size, colour etc [5]. At first, image pre-processing is performed which include image gray scaling, noise removal, discontinuity removal, dot removal etc from the images. The next phase is the text segmentation of each character from the image. Finally, the separated character is recognized using the neural network and stored.

III. PROPOSED SYSTEM

Essays are used to evaluate student's knowledge from before itself. The aim of the proposed system is to evaluate essays automatically. During the time of essay evaluation, the concepts and contents present in the essay are used to give mark or grade for it. But the automatic evaluation of essays is more difficult and challenging task than evaluation of the single word answers like multiple choice questions. The proposed system introduces the evaluation of the handwritten essays similar to that of the normal answer sheet evaluation. The aim is to check whether the proposed system is similar to that of manual evaluation process, how much accuracy can give by the proposed system etc.

The proposed method has three phases. The first phase is the Reference Material Construction in which a single reference material is constructed from the multiple reference material. The second phase is the Text Extraction from the Handwritten Student Essay. From this phase, the text to evaluate against the reference material is constructed. The final phase is the Essay Grading. In this phase a grade is calculated and displayed.

3.1 Reference Material Construction

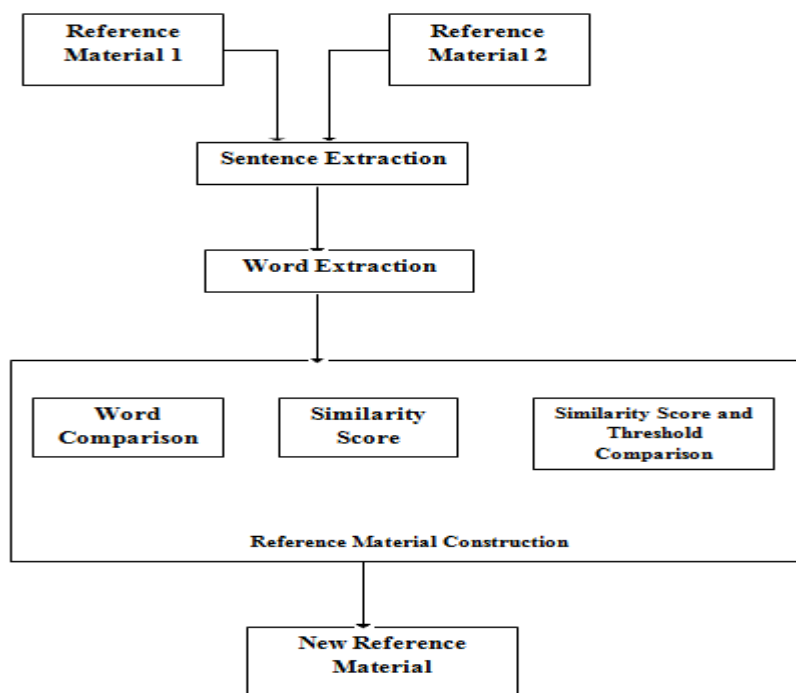


Fig. 1: Reference Material Construction

Knowledge for the students is obtained by reading the reference materials. For a single topic under study, there is a chance to refer more than one reference material. In order to evaluate the essay, it is required to compare the essay with all these reference material. This will be more time consuming and difficult to perform. In order to avoid this problem, it is better to construct a single reference material. Then the essay to be evaluated can be compare with the newly created reference material.

In the case of different study materials, the same content may be represented in different way. So that the concept or meaning of the two different reference material will be same. Because of this reason, the construction of the study material should perform by checking the sentence similarity or the by checking the meaning of the sentences. For the above stated problem, the concept of Longest Common Subsequence (LCS) is used.

Each and every sentence from both the study materials are compared semantically. For that each word are compared semantically. In order to compare each word of two different sentences synonyms of those words are compared. The synonyms of words are obtained from the synonym set of Microsoft word application. With this document the handwritten essays are compared and grade is evaluated. If the two comparing sentences are semantically similar, then any one of them is considered to the newly created reference study material. Similarly if the comparing strings are not semantically similar, then both sentences will add to the reference study material.

3.2 Text Extraction from Handwritten Essays

In the field of image processing and pattern recognition, handwritten recognition became the most fascinating and challenging research area.

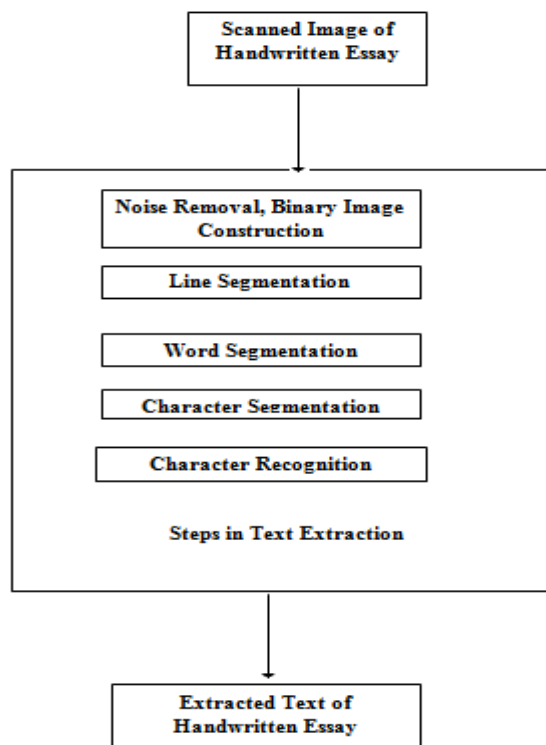


Fig. 2: Phases in Text Extraction from Image

The first stage is image pre-processing. In the pre-processing step, a clear binary image is constructed from the input image of the handwritten essay by removing the noises. This step will help to increase the performance of the text recognition task. Gray scaling is the first phase in the pre processing step. It will help to remove the noises and disturbances present in the scanned image of the student's essay. After that line removal and discontinuity removal are performed. Then dot removal is performed. The next phase is the text localization and the text segmentation. Both of these stages will help to separate each character from the entire word present in the image of the essay. By applying this method, the whole sentences comes in the essay are converted into single character image. Text localization is done for every single character. This will help to separate each character from the entire word.

The final stage is text recognition. In the text recognition stage, the segmented characters are thinned and scaled. After this, each character is compared with the stored neural network. From this the more nearest match is identified and stored.

3.3 Essay Grading

The system takes the summarized study material as input. The same document is used for calculating the score and this value is stored. In the analysis phase the student essay obtained from the second phase will take as the input. The study material and the essay are compared line by line and score is calculated. This score is compared with the previously calculated score and the grade is assigned.

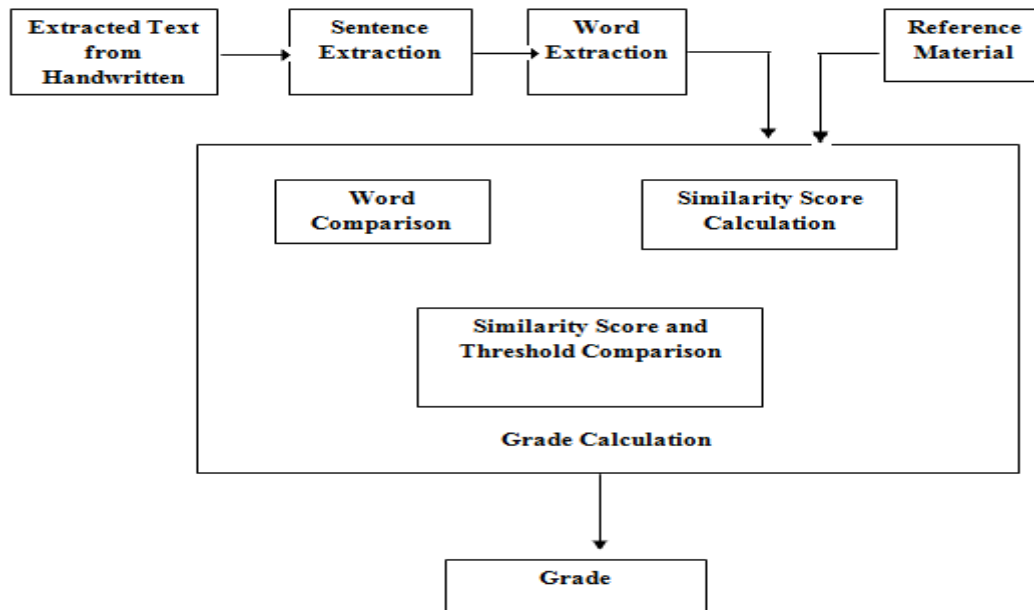


Fig. 3: Steps in Grade Calculation

The grades are calculated using by the approach called Longest Common Subsequence. The method is based on longest common subsequence (LCS) between a candidate translation and a set of reference translations. The candidate translation is a sentence of words from where it is trying find out the similar words by comparing with the reference translation. The term reference translation is a sentence of words from where comparison is made. Longest common subsequence identifies the longest occurring common words of given two sentences by using the semantic similarity. The proposed system is trying to find out the similar words by checking the meaning of the words present in the sentence.

3.4 Longest Common Subsequence (LCS)

Let A and B are two sentences consisting of words in which A and B can be represented as $A = (a_1, a_2, \dots, a_m)$ and $B = (b_1, b_2, \dots, b_n)$. For each $i = 1$ to m and $j = 1$ to n , a_i and b_j indicates the words present in the sentences. Then the longest common subsequence (LCS) of the two sentences A and B is the longest sequence C which is a subsequence of both sentences A and B.

Each and every sentence from both the study materials are compared semantically. For that each word are compared semantically. In order to compare each word of two different sentences synonyms of those words are compared. The synonyms of words are obtained from the synonym set of Microsoft word application. With this document the handwritten essays are compared and grade is evaluated.

By applying the LCS, a score is calculated for comparing each sentence in the extracted text of the handwritten essay. The score value obtained for each sentences in the essay are added to get the final score of the essay. The final score obtained for the essay is compared with the threshold score value. According to that, a grade is assigned for the essay. If the comparing two documents of sentences are exactly same, then the calculate final score value for the essay will be maximum. This maximum value of the final score is taken as the threshold score value.

IV. RESULTS

The following table gives the relation between the number of lines and the time required for combining the texts in the first and third phase.

TABLE I
RELATION BETWEEN NO: OF LINES AND THE TIME REQUIRED TO COMBINE SENTENCES

No: of Lines	Time Required to Combine Sentences (s)
1	184.54
2	342.87
3	735.46
4	1394.55

Fig. 5 shows the graphical representation of the details described in the TABLE 1. From the Graph it is clear that the time required to combine sentences is increasing according to the number of the lines increases.

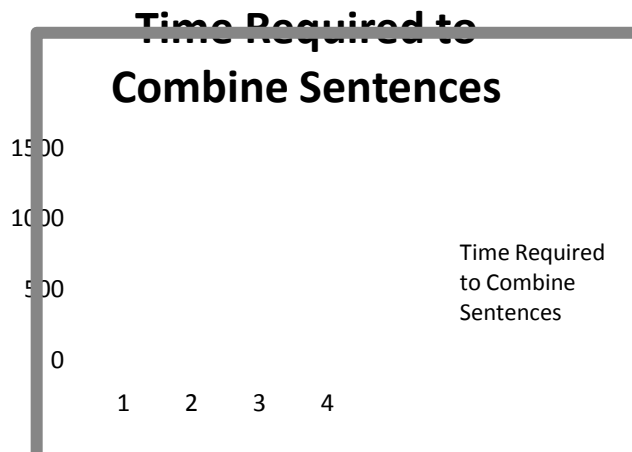


Fig. 5 Relation between No: of Lines and Time Required to Combine Sentences

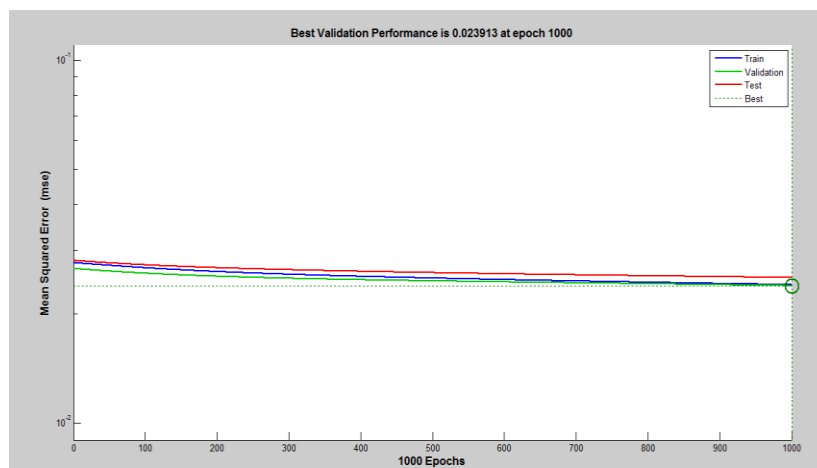


Fig. 6 Validation Performance of Neural Network

Since, the texts are semantically compared it requires to consider each and every sentence present in both documents. Thus the construction of the reference study material takes little more time than expected. The following diagrams indicate different performance measurements of the neural network data and its training. The Fig. 6 and Fig. 7 indicate the validation performance for the neural network and the training state of the neural network.

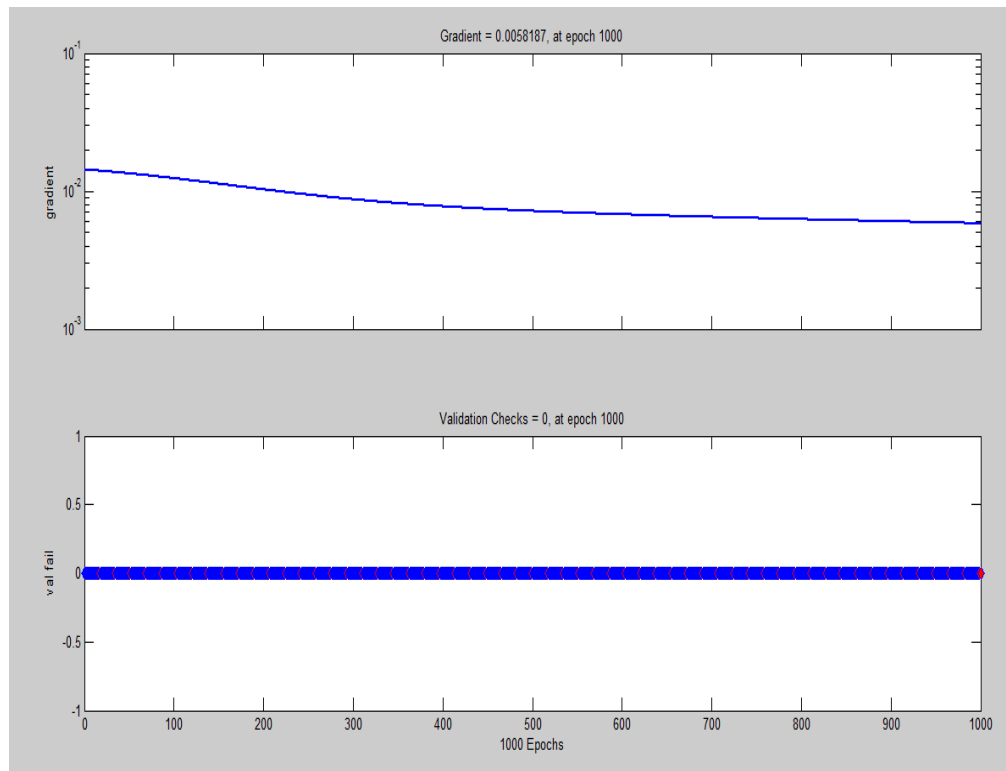


Fig. 7 Neural Network Training State

V. CONCLUSIONS

Essays are used for measuring student knowledge. Essays, is a demanding task even for a human. From previous times itself essays are evaluated according to conceptual content present in that. If there is an automated system for evaluating the essays, the evaluation can do faster, especially for evaluating handwritten essays which will perform similar to the manual human evaluation. The proposed system introduces evaluation of handwritten essays. Handwriting recognition is one of the most challenging research areas in field of image processing and pattern recognition. Student's knowledge is acquired by reading the study materials and the student's knowledge can be measured as the degree of semantic similarity between the essay and the parts of the textbook covering the topic under study. Students can refer more than on includes three different modules. The first module is text extraction from handwritten student essays. The next phase is and the final phase is grading of the students essays. The proposed system evaluates the essays successfully.

If there is an effective automated system for evaluating essays, it will be more helpful in the education field. Essays can evaluate more easily without requiring manual effort. The proposed system extracts texts from the image, but the texts extracted are not accurate always. The system results wrong words instead of giving the actual word present in the handwritten essay. In this case, the actual grade calculation is not possible. The grade calculated will not be effective. If it is possible to add a feature like automatic spell correction of the words along with the text extraction of the system, the more accurate texts can be extracted and thus more accurate grade also can calculate.

REFERENCES

- [1] Tuomo Kakkonen and Erik krisutinen, "Automatic Assessment of the Content of Essays Based on Course materials", Dept of Computer Science and Engineering
- [2] Lynn Streeter and Others, "The Credible Grading Machine: Automated Essay Scoring in DOD"
- [3] Semire Dikhi, "An Overview of Automated Scoring of Essays", Turkish Online Journal of Distance Education –TOJDE, January 2006, Vol: 7, No:1, Aritele: 5
- [4] Tuomo Kokken and Others, "Comparison of Dimension Reduction Methods for Automated Essay Grading", Educational Technology and Society, 11(3), PP: 275-288, 2008

- [5] Paraag Agarwal and Rohit Varma, "Text Extraction from Images", IJCSSET, April 2012, Vol:2, Issue: 4, PP: 1084-1087
- [6] Neeta Nain Subhash Panwar, "Handwritten Text Recognition System Based on Neural Network", Malavya National Institute of Technology, Jaipur
- [7] Chin-Yew Lin and Franz Josef Och, "Automatic Evaluation of Machine Translation Quality Using Largest Common Subsequence and Bigram Statistics", Information Science Institute
- [8] Md Monjurul Islam and A Sm Latiful Hoque, "Automated Essay Scoring Using Generalized Latent Semantic Analysis", Journal of Computers, Vol:7, No:3, March 2012
- [9] Chin Yen Li and Edward Hony, "Automatic Evaluation of Summaries Using N-Gram Co-occurrence Statistics", Proceedings in HLT-NAACL, 2003, pp: 41-48
- [10] Jessy Hansen, "A Matlab Project in Optical Character Recognition (OCR)", 2010
- [11] Chin Yen Lin and Edward Hony, "BLEU: A Method for Automatic Evaluation of Machine Translation", Proceedings on Computer Linguistics, PP: 311-318, July 2002
- [12] Chung, G.K.W.K and O.Neil, "Methodological Approaches to Online Scoring of Essays", CSE Technical Report 461, National Centre for Research on Evaluation, Los Angeles, USA, 1997