Subjective Evaluation using LSA Technique

M. SYAMALA DEVI 1 and HIMANI MITTAL 2
1. Professor, Department of Computer Science and Applications, Panjab University, Chandigarh, India
2. Research Scholar, Department of Computer Science and Applications, Panjab University, Chandigarh

ABSTRACT

Information and communication technologies are effectively used in educational administration and teaching-learning process. There is a need to utilize them for educational evaluation. Subjective evaluation and assessment are essential part of education system. For evaluation of subjective examinations, several statistical and mathematical techniques such as Latent semantic analysis, and maximum entropy are being used. Latent semantic analysis (LSA) is one such technique used for evaluation of non-technical prose. In this paper LSA technique is applied to evaluation of technical answers. The LSA algorithm clusters the text into groups on the basis of similarity. The detailed LSA algorithm is included in the paper. A prototype based on the LSA algorithm is developed using soft-computing platform MatLab and the programming language Java. The prototype so developed has been tested by conducting class tests of students of under-graduate classes of computer science courses. The results are analyzed and observed to be satisfactory. It is concluded that LSA technique can also be applied for the evaluation of technical answers.

KEYWORDS: Educational evaluation, Subjective evaluation, Latent Semantic Analysis

1. INTRODUCTION

Information and Communication Technologies (ICT) are being integrated with traditional classroom teaching. Several e-learning and virtual learning platforms have developed over the last few decades. An important part of learning process is evaluation of student knowledge. Subjective examinations are conducted to measure the critical thinking ability of the student. Evaluation of subjective examinations using computerized tools has been a topic of research for more than four decades. Several statistical and mathematical techniques have been proposed. However, the success of such endeavors has been limited. The application of these techniques has been limited to non-technical prose evaluation. The algorithms are not applied for technical prose evaluation. The needs of both the technical and non-technical prose are different. In non-technical evaluation style of writing, expression, grammar and correctness (to some extent, if not fiction) are desired. In technical prose correctness and keyword presence are most important. In this paper one such technique called Latent Semantic Analysis (LSA) is discussed. LSA technique is used for clustering of similar documents into groups. It is widely used for document indexing in search engines. It is also used for other information retrieval applications. Some researchers have applied LSA technique for answer evaluation of non-technical prose as discussed in section 2.

In this paper the technique is applied to evaluation of technical prose with slight modifications. A prototype is developed using programming language Java and soft computing platform MatLab implementing the modified LSA technique. Several class tests of undergraduate classes of computer science were conducted and the papers were evaluated. The evaluation was done by human evaluators as well as by using the prototype developed. It has been found that the modified LSA when applied to evaluation of technical prose gives satisfactory results.

The paper is organized as follows. Section 2 contains review of related work. Section 3 gives the detailed LSA algorithm for subjective answer evaluation. Section 4 includes the testing done to
evaluate effectiveness of LSA in subjective evaluation. Finally in section 5, it is concluded that LSA technique is effective for subjective answer evaluation and can be used to assist the examiner.

2. REVIEW OF RELATED WORK

In 1994, Project Essay Grader (PEG) [1] was developed for automated subjective evaluation. It performs the evaluation based on style analysis of the answer and does not take content into account. Its agreement with human graders is 87%. It measures features like essay length, word length and vocabulary used etc. PEG cannot evaluate technical validity of answers. It was used at primary and secondary levels of education for English subject essay evaluation.

In 1999, Foltz et al. [2] applied mathematical technique called Latent Semantic Analysis (LSA) to computerized evaluation in a tool called Intelligent Essay Assessor (IEA). It requires a database of pre-graded essays. The student answers are compared against these pre-graded essays and correlation is calculated between them. The success rate of IEA 80%.

In 1999, Hofmann [3] developed Probabilistic latent semantic analysis (PLSA). This method has strong statistical foundation in aspect modeling. It calculates document to word joint probability using estimation maximization algorithm. While LSA has its roots in linear algebra, PLSA is based on latent class model.

In 2003, Blei et al. [4] suggested a generalization of PLSA, by using mixture model and dirichlet allocation distribution for calculating co-occurrence probability of words. This technique is known as Latent Dirichlet Allocation (LDA).

In 2008, Kakkonen, Myller, Sutinen and Timonen [5], Automatic Essay Assessor (AEA) is a system that utilizes information retrieval techniques such as LSA, PLSA and LDA for automatic essay grading. They performed a series of experiments using LSA, PLSA and LDA for document comparisons in AEA for essay grading with empirical data. It was found that using LSA yielded slightly more accurate grading than PLSA and LDA.

In 2010, Islam and Hoque [6] proposed a system that makes use of Generalized latent semantic analysis (GLSA) technique for evaluation. In GLSA n-gram by document matrix is created instead of a word by document matrix of LSA. It has 89% of accuracy which show that the system is very closer to human grader.

For subjective evaluation, primarily statistical and mathematical techniques are used. Several statistical techniques used are discussed above. These statistical and mathematical methods classify, cluster and index the student answers by extracting essay features to calculate quantities like probability and latent variables. The success rate of these techniques for evaluation of non-technical prose is 80-90%. These techniques are applied to evaluation of non-technical answers only. In this paper, one of the techniques LSA is applied for evaluation of technical answers in the field of computer science.

3. LSA TECHNIQUE

Latent Semantic Analysis technique is used for clustering similar documents and finding if a given query exists in these clusters. For evaluation of technical answers, it requires a database of pre-graded essays. The it compares all the student answers together with the model answer keywords. It represents the standard answer and the student answer as a vector. It then calculates the cosine similarity between these two vectors. The figure-1 shows the diagrammatic representation of these
vectors and relation between them. If angle $\Theta$ between the $\vec{q}$ and $\vec{d}$ is zero degree, then the two are similar.

![FIGURE-1: Schematic representation of Latent Semantic Analysis Technique](image)

The cosine similarity is calculated by formula given in Equation-1. If $\cos \Theta = 1$, then $\Theta = 0$.

$$\cos \Theta = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}| \cdot |\vec{d}|} \quad \ldots \text{Equation-1}$$

The following steps are included in LSA algorithm for technical answer evaluation:

a) Construct the term to document frequency (tdf) matrix- $A_{mXn}$. Each element in the matrix gives the degree of participation of the document in the corresponding concept. Here $m$ is number of terms and $n$ is number of answers.

b) Calculate Singular Value Decomposition (SVD)

$$\text{SVD}(A_{mXn}) = S \cdot E \cdot U^T$$

Where,

$S$ = Left Eigen Vectors of $A^T A$. Terms in the concept space are represented by row vectors of $S$.

$E$ = Identity of square roots of eigen values of $A^T A$. It gives the degree of relationship between the $S$ and $U^T$ matrix.

$U^T$ = Right Eigen Vectors of $A A^T$. Documents are represented by column vectors of $U^T$

c) Calculate $S \cdot E$. It gives the participation of term in the query.

d) Calculate $E \cdot U^T$. It describes the participation of document in the query (d).

e) Compute the weight of the query $q$ as $\sum_{i=1}^{n} \sum_{j=1}^{n} se[i][j]$ of $S \cdot E$

f) Calculate cosine similarity between $q$ and $d_i$. It gives the probability of the query words appearing in the document. Higher the value more are the chances of occurrence.

This technique is largely used in information retrieval and information extraction applications. It is also used for plagiarism detection. Evaluation of examinations is one of its applications. LSA studies relation and similarity between student answers and standard answers. All the documents must be available when LSA is applied. It requires a database of pre-graded essays. Each student answer is compared to each standard answer.

4. IMPLEMENTATION AND TESTING

The LSA algorithm as discussed in Section 3 was implemented using programming language, Java and soft computing platform, MatLab. Before LSA technique was applied the student and stand answers were pre-processed. The following steps were peformed:

a) Tokenization: The individual words of the answers were extracted. They were all brought down to lower case representation to avoid any mismatches.
b) Stop word removal: The words which are common to English language and have only grammatical significance like prepositions, articles, etc were removed from both student and standard answers.

Tokenization, stop word removal and construction of term by document matrix was done in Java. The SVD calculations and cosine similarity were calculated using MatLab. JDK version 1.6 was integrated with MatLab version 7.9 (R2009) using MatlabControl library on Windows XP platform. A class test was conducted for undergraduate course Bachelor of Computer Applications part II. The subject chosen was Database management systems. The students had to write answer to question- “Explain the architecture of DBMS.” 17 students participated in the test and their answers were typed on computer. Then printouts of these answers were given to three anonymous evaluators for assessment. The answers were also evaluated using the developed prototype. The score assigned by the computer and human evaluators is given in table-1. The maximum marks for the answer were 15.

<table>
<thead>
<tr>
<th>Computer</th>
<th>Human1</th>
<th>Human2</th>
<th>Human3</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>5</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>5.5</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>5.5</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0.5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3.5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

**Table-1 Marks given by human and computer evaluators**

In table-2 the correlation between the various evaluators is given. It has been found that human to computer correlation between scores is as high as 77%, which is comparable to human to human correlation. There is a difference in the assigned score even between human evaluators.

<table>
<thead>
<tr>
<th>Human1</th>
<th>Human2</th>
<th>Human3</th>
<th>Computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human1</td>
<td>0.697209</td>
<td>0.815689</td>
<td>0.518398</td>
</tr>
<tr>
<td>Human2</td>
<td>0.697209</td>
<td>0.699455</td>
<td>0.564386</td>
</tr>
<tr>
<td>Human3</td>
<td>0.815689</td>
<td>0.699455</td>
<td>0.774931</td>
</tr>
<tr>
<td>Computer</td>
<td>0.518398</td>
<td>0.564386</td>
<td>0.774931</td>
</tr>
</tbody>
</table>

**Table-2 Correlation between various evaluators**
5. CONCLUSIONS AND FUTURE WORK

The results show that LSA agreement to human evaluators is quite high even for technical answers. As a practical method, LSA is a useful technique for evaluation of technical answers. It measures document to document similarity which imitates the human process of identifying similarity. This makes LSA suitable for use in evaluation tools.

The results can be enhanced by applying term weighting, term weighting helps account for the actual information value of that term within the text. This approach is commonly used in information retrieval in which the overlap between terms in a query and terms used in documents is weighted based on some transformation of the word frequency. The performance of the algorithm can be further enhanced by applying term conversion using lexicon like WordNet [7] to deal with the problems of polysemy and synonymy. Furthermore, it is believed that the combining syntax and semantics can be further explored, by designing sophisticated system of a combination of techniques.

6. REFERENCES


