A Comparative Study of Modified Principal Component Analysis and Generalized Latent Semantic Analysis Approach to Automated Marking of Theoretical-Based Exams

Odunayo Esther Oduntan[1] Ibrahim Adepoju Adeyanju[2]

[1]*Computer Science Department. The Federal Polytechnic Ilaro-Ogun State, Nigeria*
[2]*ComputerEngineeringDepartment Federal University Oye Oye-Ekiti, Nigeria*

Abstract: *Examination is a means by which individual's ability on an acquired skill or knowledge is evaluated. Some examiners useMultiple Choice Questions (MCQ), while others use the theory-based Questions. This study performed a comparative study of two techniques: Modified Principal Component Analysis and Generalized Latent Semantic Analysis used in grading students' responses in theory-based examination. Softcopy of students' responses and examiner's marking scheme were captured in an electronic format of .txt. file. The inherent stopwords and stemming in the .txt document were pre-processed to address morphological variations using standard stopwords list and porters stemmer algorithm, respectively. N-gram terms were derived for each student's response and the marking schemes (MS) using the vector space model. A Document Term Matrix (DTM) was generated with n-gram terms of MS and students response representing columns and rows, respectively. MPCA and GLSA algorithms were used to reduce the sparseness of the DTM to obtain a vector representation of the students' answers and the marking scheme. The reduced vector representation of the students' answers were graded according to the mark assigned to each question in the marking scheme using cosine similarity measure. The developed Automated Theory-Based Marking System (ATBMS) was implemented in Matrix Laboratory 8.1 (R2013b). Performance of MPCA was compared with GLSA to determine the effectiveness of ATBMS on the grading of students' answers in COM 317 and COM 325 courses, in terms of Pearson correlation coefficient(r) and coefficient of determination (R^2).These performance evaluation shows that MPCA is a better feature extraction techniques.*
Keywords: *Examination, Theory-based Questions, Multiple Choice Question, Document Term Matrix, Generalized Latent Semantic Analysis, Modified Principal Component Analysis, Feature Extraction*

I. Introduction

Evaluation of student's performance which is carried out by periodically testing students is a key issue in the educational sector. When students are evaluated either through continuous assessment or examinations, they need to be graded to rate their performances (Allen, 2004). In education, the term assessment refers to the wide variety of methods or tools that educators use to evaluate, measure, and document the academic readiness, learning progress, skill acquisition, or educational needs of students. Assessments also are used to identify individual student weaknesses and strengths so that educators can provide specialized academic support**,** educational programming, or social services. In addition, assessments are developed by a wide array of groups and individuals, including teachers, district administrators, universities, private companies, state departments of education, and groups that include a combination of these individuals and institutions.Assessment of students' ability and understanding of various course concepts have been done using tools such as paper based test and computer based test (Adeyanju, Wiratunga, Recio_Garcia and Lothian, 2010).

The mode of the questions administered to students' are usually multiple choice questions, multiple choice multiple answers, multiple choice single answer e.t.c. The multiple choice questions allow the examinee to select the right answer from a least of answers provided by the examiner usually referred to as objective questions. The multiple choice multiple answer allows the examinee to select from a list of options, the best combinations of answers that suits a particular question. With the Paper Based Test, computer answer sheets were provided to the students to use in shading the correct answer. The shaded computer sheets are marked or scored either by the examiner manually or by the computer system. In a computer based test, the examinee are made to make use of computer application consisting of questions to be answered, the examinee submits the page containing the answered questions electronically (Rudner and Gagne, 2001). This study is focusing on the type of questions that are theoretical in nature; that is the examinee is given questions that requires an open ended answer, the examinee expresses his or herself as well as taking cognizance of the keywords that are required in the context of the questions.

According to Oduntan, et. al. (2015), researchers in the field of automated grading system have made use of various approaches raging from PAGE(1996), E-rater, Intelli-Metrics , AETG-MPCA, LSA and GLSA.

In this study, a comparative analysis of the use of Generalized Latent Semantic Analysis(GLSA) and the Modified Principal Component Analysis Approach(MPCA) was performed by simulating electronic copy of students' responses in theory-based examination.The vector space model was used to extract n-grams terms from students' responses; the similarity score was computed using the cosine similarity score and the machine generated student score was computed as a weighted aggregate of documents' similarity scores; where the weight is the examiners marks assigned for each question in the marking scheme.

Performance of the MPCA was compared with GLSA to determine the effectiveness of Automated Theory-Based Marking System (ATBMS) on the grading of students' responses in theory-based examination, in terms of Pearson correlation coefficient(r) and coefficient of determination (R^2).

II. Literature Review

A theory is a contemplative and rational type of abstract or generalizing thinking, or the results of such thinking. Depending on the context, the results might for example include generalized explanations of how nature works. The word has its roots in ancient Greek, but in modern use it has taken on several different related meanings. A theory can be *normative* (or prescriptive),meaning a postulation about what ought to be. It provides "goals, norms, and standards" (Mason and Grove-Stephenson, 2002). Hence, a theory-based Examination is an assessment method in which examinees express their understanding of a particular question having the in mind the underlying terms of the context.

Natural Language Processing (NLP) is a technology that concerns with natural language generation (NLG) and natural language understanding (NLU). NLG uses some level of underlying linguistic representation of text, to make sure that the generated text is grammatically correct and fluent (Adeyanju, 2012).

The context of the sentence includes the situation in which the sentence is used, the immediately preceding sentences, and so on (Rao, 2003), context interpretation concerns how the context affects the interpretation of the sentence. Natural Language Processing uses approaches such as clustering and hybrid in automated essay grading system. The clustering approach entails grouping essays that have similar words patterns to form a cluster with the same score.Pattern matching technique is used to interpret input utterances as a whole, rather than building up their interpretation by combining the structure and meaning of words or other lower-level constituents (Fenwa, et. al., 2012)

2.1 Text Representation

In text processing, terms that constitute a sentence can be represented making use of phrases and n-grams. Phrases are more precise than single words as topic descriptions and usually less ambiguous (e.g., "rotten apple" vs. "apple"). However, the impact of phrases on retrieval can be complex (Jusoh and Alfawareh, 2007). An n-gram is a subsequence of n items from a given sequence. N-gram models can be imagined as placing a small window over a sentence or a text, in which only n words are visible at the same time. A simple definition is that a phrase is any sequence of *n* words. Sequences of two words are called *bigrams* and sequences of three words are called *trigrams* (Salton, Wong and Yang, 1975).

2.2 Automated Scoring Model

This refers to a statistical formula that is used, usually with the help of computers to estimate future performance of prospective borrowers and existing customers. A scoring model calculates scores based on the data extracted. Features used may be exact match or best match, in exact match, the query specifies precise retrieval criteria in which every document either matches or fails to match query resulting in a set of document without order. In best-match, the query describes good or best matching documents resulting in a ranked list of documents and estimation of quality. The vector space model and the probabilistic models are reviewed in this study.

2.2.1   Vector Space Model

Vector space model otherwise known as term vector model is an algebraic model for representing text documents as vectors of identifiers, such as index terms. It is used in information filtering, information retrieval, indexing and relevancy rankings. Each dimension corresponds to separate terms. Terms are orthogonal and they form a vector space. This model is used in document representation to specify the details of the document. The general equation for vector space model according to Salton, Wong and Yang (1975) is illustrated in Equation (2.1):

$$d_j = (w_{1j}, w_{2j}, w_{3j}, ...w_{ij}) \qquad\qquad 2.1$$

where$d_j$denotes the jth essay and w_{ij}denotes the weight of the ith feature in the jth essay, which is the weight of the feature.

2.2.2   The Probabilistic Model

The probability essay d_i receives score classification c_j as follows (Foltz et al., 1999):

$$P(d_i \mid c_i) = \prod_{t=1}^{v} [B_{it} P(w_t \mid c_j) + (1 - B_{it})(1 - P(w_t \mid c_j))] \qquad 2.2$$

where v is the number of features in the vocabulary

$B_{it} \in (0,1)$ indicates whether features t appears in the essay I and $P(w_t/c_j)$ indicates that features w_t appears in a document whose score is c_j. It is calculated using Equation (2.3)

$$P(w_t \mid c_t) = \frac{1 + \sum_{i=1}^{D_j} B_{it}}{J + D_j} \qquad 2.3$$

where$D_j$is the number of essays in the training group scored C_j and j is the number of score groups. It scored the trial essays by determining the probability that essay d_i should receive score classification $c_j$given the equation is multiplied by the prior probabilities. The score with the highest posterior probability is then assigned to the essay. This scoring model also uses the binary independence model to represent documents.

2.3 Feature Extraction

Feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (e.g. the same measurement in both feet and meters) then the input data will be transformed into a reduced representation set of features (also named features vector) (Guyon, Gunn and Nikravesh, 2006). Transforming the input data into the set of features is called *feature extraction*. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input (Hasan, 2009).The most widely known linear feature extraction methods are principal component analysis (PCA) and, especially for textual data, Latent Semantic Analysis (LSA) and the independence component analysis (Tsai, 2011). In this study, two techniques was reviewed; Principal Component Analysis (PCA) and Latent Semantic Analysis (LSA).

2.3.1   Principal Component Analysis

PCA was invented in 1901 by Karl Pearson, as an analogue of the principal axes theorem in mechanics; it was later independently developed (and named) by Harold Hotelling in the 1930s (Hotelling, 1933). The method is mostly used as a tool in exploratory data analysis and for making predictive models. PCA can be done by eigenvalue decomposition of a data covariance (or correlation) matrix or singular value decomposition of a data matrix, usually after mean centering (and normalizing or using Z-scores) the data matrix for each attribute (Abdi and Williams, 2010). The results of a PCA are usually discussed in terms of component scores, sometimes called factor scores (the transformed variable values corresponding to a particular data point), and loadings (the weight by which each standardized original variable should be multiplied to get the component score) Omidiora, (2006). The PCA is generated with the following algorithm (Hotelling, 1933):

Step 1: input keyword from the document vector that is n = 1
Step 2: Subtract the mean
For PCA to work properly, mean is subtracted from each of the data dimensions.
The mean subtracted is the average across each dimension. This produces a data set whose mean is zero.
Step 3: Calculate the covariance matrix
Step 4: Calculate the eigenvalue and eigenvector of the covariance matrix
Step 5: Choosing components and forming a feature vector
This is the final step in standard PCA, and is also the easiest. Once the components
(eigenvalue) that are to be kept in our data and formed a feature vector have been chosen, then the transpose of the vector is taken then it is multiplied with the left of the original data set, to derive the generalised document vector.

2.3.2   Latent Semantic Analysis

This is a machine learning method that infers the meaning relations among words and passages through mathematical computations applied to a large corpus of text (Landauer et al, 1997). It is also a complex statistical technique that was initially developed for indexing documents and information retrieval (Deerwester*et al*., 1990). Nevertheless, it can also be applied to automated essay grading (Landauer*et al*., 1997). In this field, this technique serves to extract the conceptual similarity between the student's candidate text and the teacher's reference text by looking for repeated patterns between them. This approach is quite robust and proves its name by finding the hidden relationships between words that could be in different documents or between documents

that do not share words. The reason for this fact is given by Landauer who said that what causes two words to have similar meanings is that they change the meaning of passages in which they occur

According to Whittingdon and Hunt (1999) Latent Semantic Analysis might be described in the following steps:
1. The training phase: In this step, it is necessary to calculate the weights for the vectors that represent the reference texts. It has the advantage of not requiring a manually marking. Moreover, they could be as big as a textbook or as small as a short paper.
2. The test phase: Each student text has to be represented in LSA, which implies several transformations (Valentiand Cucchiarelli, 2003) such as:
(a) The matrix representation:
First of all, certain stopwords are removed and words are stemmed, so that there are less words and all of them will be different and meaningful. LSA is a procedure to reduce the dimensionality of the space, i.e. to reduce the size of the matrix, where each row represents a word and each cell its frequency for each context (e.g. a sentence or a paragraph). Therefore, the number of dimensions will be as large as the size of the vocabulary in the language. This matrix can be considered as the Vector Space Model representation of the contexts (sentences, paragraphs, or documents) studied.
(b) The tuning of the matrix weights:
Now the relevance of each word in the passage is measured. The measurement is based on the frequency of the word in all the contexts. This idea is very similar to the used by the inverse document frequency (IDF) weighting.
(c) Singular Value Decomposition (SVD):
The original matrix is decomposed into the product of orthogonal matrices. One of them is diagonal and its values are the singular values (the eigenvalues) of the original matrix.
(d) Dimensionality reduction transformation:
        In order to find the relations between words and contexts, it is necessary to reduce the rank of the diagonal matrix. It is in this transformation where the hidden relationships are detected.
(e) The reconstruction:
This new diagonal matrix is multiplied by the other two matrices outputs of the SVD and the result is the weighted version of the original matrix that is the LSA representation of the text.
3. The result phase: The LSA representation of the student text is compared against the LSA model representations and their similarity is computed.
In summary, the LSA model first transforms a large representative corpus to matrix consisting of co-occurrence-based measures. This is followed by a dimension reduction process to produce a semantic space at an optimal dimensionality, in which any text unit is mapped to another through their respective relevance to dimensions or concepts extracted from the original corpus.

2.4   Related Works
        To understand how AES tools work, it may be helpful to take a look at how AES tools were used by some researchers, in 1966, Page, the inventor of Project Essay Grader (PEG) and the pioneer of AES, published an article entitled "The Imminence of Grading Essays by Computer." In this article, Page described his invention of using computer technology to grade essays and expressed his optimism about the promising future of relieving English teachers from the burden of grading papers (Wresch, 1993).
        Page's PEG used three steps to generate scores (Yang, Buckendahl, Juszkiewicz, and Bhola, 2002). First, it identifies a set of measurable features that are approximations or correlates of the intrinsic variables of writing quality (proxes); second, a statistical procedure (linear multiple regression) was used to find out the "optimal combination" of these proxes that can "best predict the ratings of human experts" (Yang et al., 2002); third, the proxes and their optimal combination are then programmed into the computer to score new essays.
        Other AES tools use similar three-step strategies to score essays. Intelligent Essay Assessor (IEA), which is used by the ETS to score the Graduate Equivalency Diploma essay test, grades essays by using the technique of latent semantic analysis. It first processes a large body of the texts in a given domain of knowledge, establishing a "semantic space" for this domain. Then, it analyses a large amount of expert-scored essays to learn about the desirable or undesirable qualities of essays. Finally, it uses a factor-analytic model of word co-occurrences to find the similarity and semantic relatedness between the trained essays and the new essays at different score levels (Rudner and Gagne, 2001; Yang *et al*., 2002).
        E-rater, which was also adopted by ETS, uses natural language processing and information retrieval to develop modules that capture features such as syntactic variety, topic content, and organization of ideas or rhetorical structures from a set of training essays pre-scored by expert raters. It then uses a stepwise linear regression model to find the best combinations of these features that predict expert raters' scores. These combinations are processed into the computer program to score new essays (Yang *et al.,* 2002).

Building on the strategies utilized by PEG, IEA, and e-rater, IntelliMetric™, developed by Vantage Learning, incorporates the technologies of artificial intelligence and natural language processing, as well as statistical technologies. These combined approaches are treated as a "committee of judges," and "potential scores" from these judges are calculated by using proprietary algorithms to achieve the most accurate score possible (Vantage Learning, 2003). The algorithm was capable of analyzing more than 300 semantic, syntactic, and discourse level features. IntelliMetric system functions by building an essay scoring model firstly by using samples of essays with scores already assigned by human expert raters are processed into the machine, which would then extract features that distinguish essays at different score levels. Once the model is established, it is validated by another set of essays. Finally, it was used to score new essays (Elliot, 2003).

Ade-Ibijola, Wakama and Amadi (2012) developed an expert system for automated essay scoring , they built a knowledge base system and populated it with answer templates from Lecturers on a specific course, designed an inference engine using Information Extraction (a shallow NLP technique which is an hybrid of Statistical Keyword Analysis technique and Pattern Matching with Domain Specific Dictionary), attached a Fuzzy-Module for correctness evaluation and developed two-web applications; one as a user-interface for lecturers to set their test questions and supply answer templates, and the other for Students' to write open-ended tests online and obtain an instantaneous feedback of their performance. IE was used to extract dependencies between concepts, the dependencies found are compared against the human experts to give the student's score.8

Islam and Hoque (2012) developed a generalized latent semantic analysis (GLSA) based automated essay scoring system in which n-gram by document was created instead of word by document matrix of LSA, GLSA system involves two stages: The generation of training essay set and the evaluation of submitted essay using training essay set. Essays were graded first by human grader, the average value of human score is treated as training score for a particular essay. The first stage involves: preprocessing the training essay set which is done in three steps: removal of stopwords, word stemming, selecting of the n-gram index terms, computing of the SVD of n-gram by document, the n-gram by document matrix contains orthogonal, diagonal and orthogonal matrices, reduce dimensionality and determine the document similarity using the cosine formula.

Oduntanet. al., (2016) developed a modified principal component analysis based automated essay-type grading system in which the vector space model was used to extract n-gram terms from electronically captured students' response and examiner marking scheme to derive Document Term Matrix (DTM) of both the students' response and the examiner marking scheme, modified principal component analysis algorithm was used to sparse the DTM, to generate document vectors. The generated document vectors of the students' response was compared with that of the examiner marking scheme using the cosine similarity measure and the machine score was an aggregates of the similarity score and the marks assigned by the examiner.

In this study, it has been observed that GLSA based automated grading system and MPCA based automated essay type grading system made use of n-gram terms but different feature extraction techniques. GLSA used the Latent Semantic Analysis while MPCA used the Modified Principal Component Analysis approach.

III. Materials And Methods

In this study, an experimental research approach was used; this involves the design and implementation of an automated essay type grading system making use of the Generalised Latent Semantic Analysis(GLSA) and Modified Principal Component Analysis (MPCA) approach.The developed Automated Essay-Type Grading System (AETGS) was implemented in Matrix Laboratory 8.1 (R2013b). Performance of MPCA was compared with GLSA on the grading of students' answers in terms of Pearson Correlation Coefficient(r) and Coefficient of Determination (R^2). COM 317: Management Information System and COM 325: Research Methodology courses were used as case study.

Figure 1 gives a description of the framework of automated theory type grading system using Modified PCA generalised LSA, which involved the collection of data set comprising of the theory-type marking scheme and softcopy students' answers. The hardcopies of the marking scheme were transcripted into electronic form and the softcopies theory-type students' answers in *.txt* file format. The inherent stopwords and stemming in the *.txt* document were pre-processed to address morphological variations using standard stopwords list and porters stemmer algorithm, respectively. N-gram terms were derived for each student's response and the marking schemes (MS) using the vector space model. A Document Term Matrix (DTM) was generated with n-gram terms of the marking scheme and students' response representing columns and rows, respectively. The GLSA and MPCA were used to reduce the sparseness of the DTM to obtain a vector representation of the students' answers and the marking scheme. The reduced vector representations of the students' answers were graded according to the mark assigned to each question in the marking scheme using cosine similarity measure. Figure 2 is a flow diagram describing the step by step process involved in obtaining the automated student score.
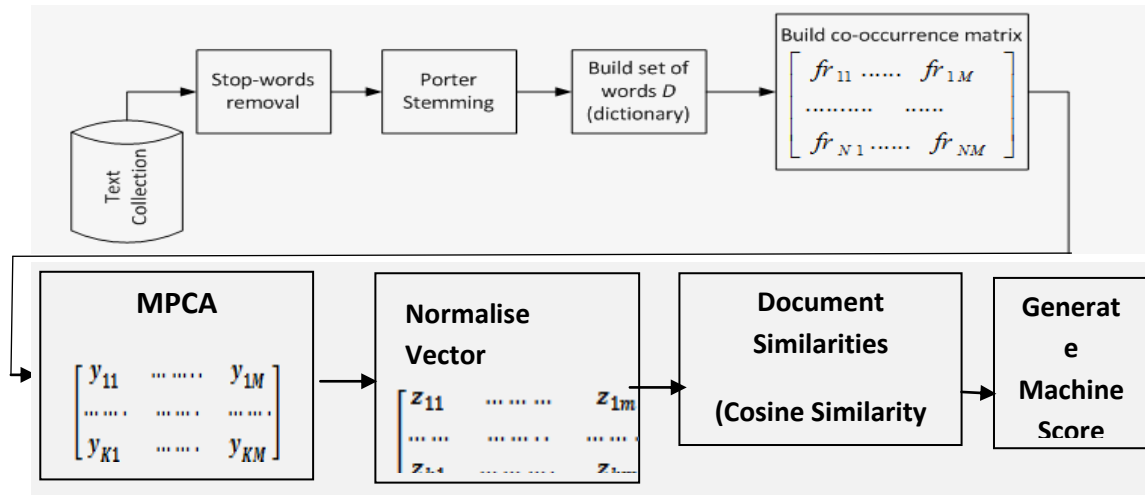
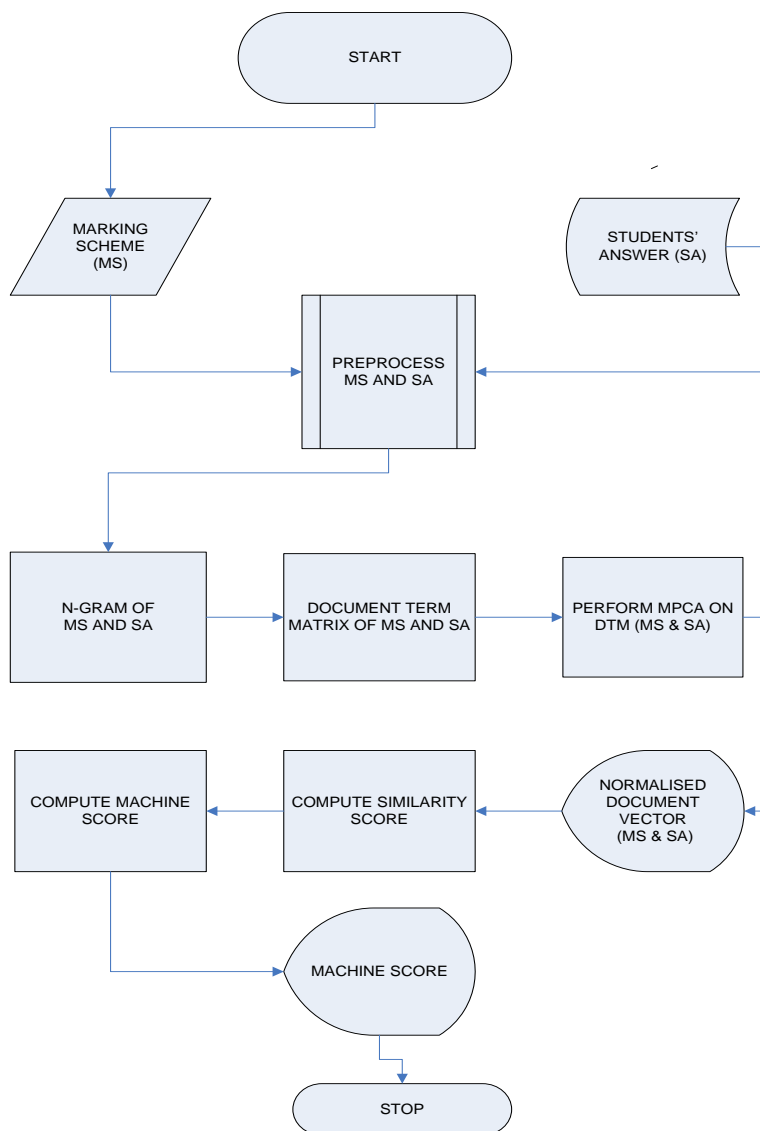**Figure 1:** Framework of the Automated Essay-Type Grading System

Figure 2: Flow Diagram of the Developed MPCA-AETG System

IV. Results

4.1 Experimental Dataset
Two data sets were used in the experimental set up of this research comprising of the students answers and the marking scheme. Experimental data sets are:
i. COM 317 Data Set: Management Information System I
ii. COM 325 Data Set: Research Methodology

i. COM 317 Data Set: Management Information System I
This is a course undertaken by the Higher National Diploma Students of Nigerian Polytechnics under the authority of the National Board for Technical Education. The course is taken by the 300 levels students of Management Studies which include Accountancy, Marketing and Business Administration and taught by lecturers of the Department of Computer Science. The undergraduates in these Departments write the examination of this course in the second semester of the session. The specific question chosen for this study is a second semester course of 2012/2013 academic session of the Federal Polytechnic Ilaro in Ogun State. The number of students answers used for this experiment is thirty-five and one marking scheme.

ii. COM 325 Data Set: Research Methodology
This is a course undertaken by the Higher National Diploma Students of Nigerian Polytechnics under the authority of the National Board for Technical Education. The course is taken by the 300 levels students of Computer Science Department and taught by lecturers of the Department of Computer Science. The undergraduates in these Departments write the examination of this course in the second semester of the session. The specific question chosen for this study is a second semester course of 2013/2014 academic session of the Federal Polytechnic Ilaro in Ogun State. The number of students scripts used for this experiment is thirty-five (35) and the marking scheme is one (1). Student answers were loaded into the system by entering the student identity such as std 1 into the student script button, questions to be graded are loaded by typing the range of question into: Enter total number of question section of the framework.

4.2 Text Preprocessing
The preprocessing operation taken in this research involves the transcription of examiner's marking scheme, removal of stopwords and stemming operation. Stopwords and stemming preprocessing operations were performed on the transcripted marking scheme and the student answers by selecting the preprocessing operation button. Figure 3 shows the snapshot of transcripted text in its .txt file format. Figure 4 is a snapshot of the output of the preprocessed operation. From the result, frequently used words have been removed and the reduction of different forms of word to a common stem was performed. The standard stopword list and the Porter stemmer algorithm were used respectively. The output of the text preprocessing stage are the extracted keyword and bigram which serves as input into the vector space model to derive an n-gram document term matrix.
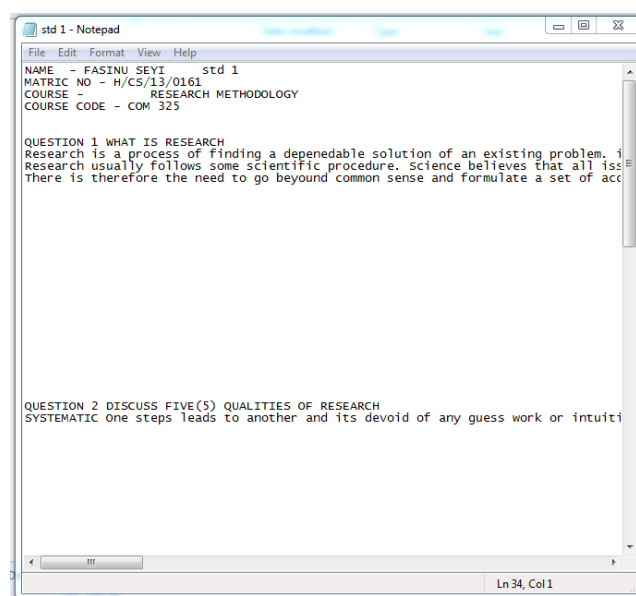

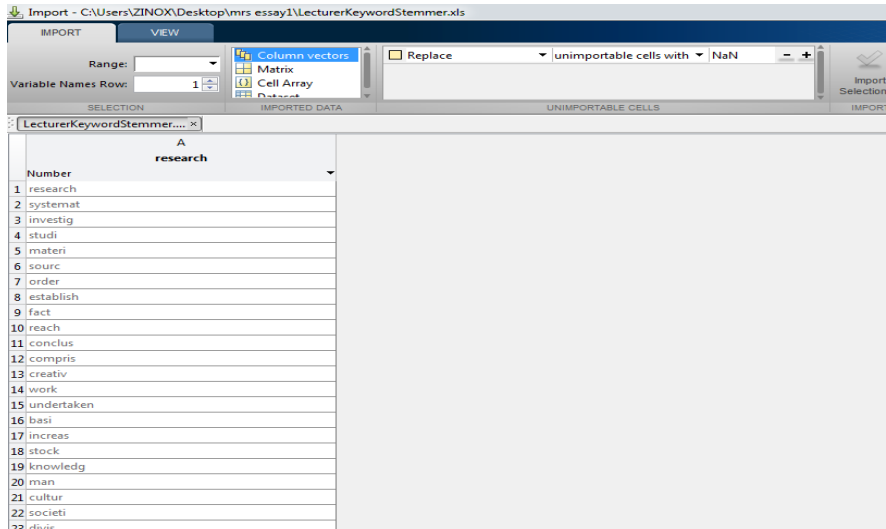**Figure 3:** A Screenshot of Student's Response in .txt file format

Figure 4 Snapshot of Pre-processed Text

4.3 Generating of N gram Document Vectors

The vector space model was used to generate the document vectors, this was achieved by loading the marking scheme from the text editor into the application by selecting the marking scheme button, document extractions were done using the keyword and bigrams buttons. Terms were weighted using binary (0, 1) scheme. When a relevant term is found in the essay-type student script, it will be weighted 1, if the relevant term is not found it will be weighted 0. In the document term matrix, each question will be used to represent a document. Each document is represented as a row in the document term matrix. Figure 5: shows the extracted bigram terms from the pre-processed text.
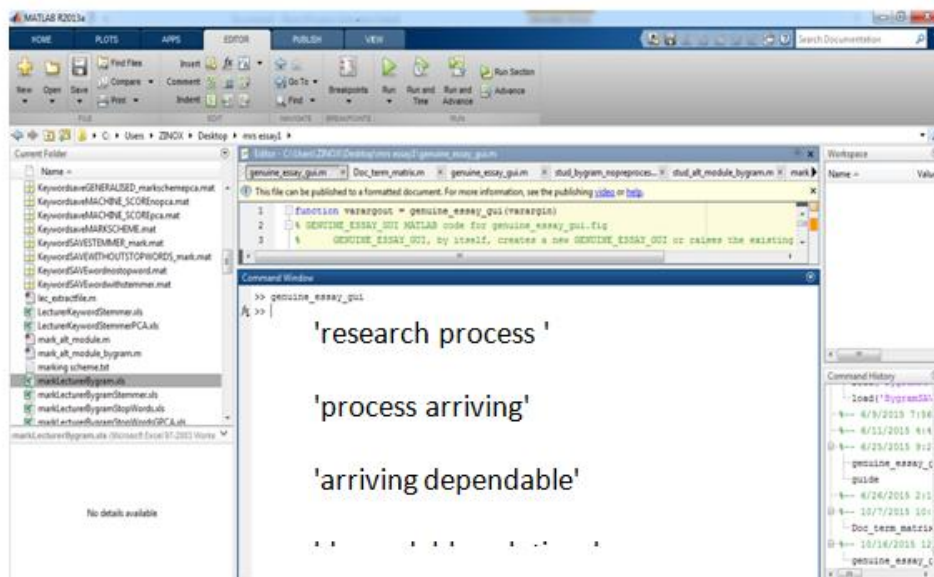

Figure 5 A screenshot of N-grams document representation

V. Discussion

In this study, performance evaluation of the Modified Principal Component Approach (MPCA) and the Generalised Latent Semantic Analysis Approach (GLSA) to Automated Theory-type Marking System was carried out using Pearson Coefficient Correlation(r) and Coefficient of Determination (R^2).

5.1 Pearson Coefficient Correlation

Pearson Product Moment Correlation (r) signifies the degree of relationship that exists between dependent variables and independent variable. The dependent variable is the human score denoted as X, while the independent variable is the machine score. Valid result for r lies between -1 and +1. If the result lies between 0 and 1, it shows there is a positive correlation that is X increases as Y increases. If r = 1, it shows that the result

is perfect positive. If r is between 0.5 and 1, it shows a high positive correlation, when r is between 0 and 0.49, it exhibits a low positive correlation. When r = -1, it shows a perfect negative correlation that is the rate at which the dependent variable increases is exactly equal to the rate at which the independent variable decreases. When r is between -0.5 and 0, it shows a weak negative correlation, when r is between -0.49 and -1, it exhibits a strong negative correlation. The formula for deriving the Pearson Correlation Coefficient as stated by Islam and Hogue (2012) is:

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

5.1

where X represent the human score and Y represent the machine score and N is the number of student scripts processed.

5.2 Coefficient of Determination

The coefficient of determination (R^2) indicates how well data points fits a statistical model. It is a measure of the model's predictive power and it can be useful for evaluating the statistical importance of the independent predictor variables. It also provides a measure of how the observed outcomes are replicated by the model. R^2 is derived by the formula in Equation 5.2. The coefficient of determination value ranges from 0 to 1. The better the linear regression fits the data in comparison to the simple average, the closer the values of R^2 is to 1, if $R^2 = 1$, it indicates that the regression line perfectly fits the data. It gives information about the goodness of fit of a model (Gujarati and Porter, 2009).

$$R^2 = \left(\frac{\sum_{i=1}^{n}(O_I - \overline{O})(P_I - \overline{P})}{\sqrt{\sum_{i=1}^{n}(O_I - \overline{O})^2}\sqrt{(P_I - \overline{P})^2}} \right)$$

5.2

where O_I is the actual estimate or value of the original assessor, P_I is the predicted value generated by the system, $\overline{O}, \overline{P}$ are their respective mean and n is the number of data.

5.3 Performance measure of MPCA with GLSA

Generalized latent semantic analysis (GLSA) was used to develop automated essay grader by Islam and Hogue (2012). This existing state of Art (GLSA) was implemented in this study. A comparison of the result generated by the modified principal component analysis method and the generalized latent semantic analysis method for automated essay grading system was performed.

Table I shows the Pearson correlation coefficient for MPCA to be 0.7 while GLSA have 0.4 for COM 317 dataset. MPCA had 0.75 while GLSA have 0.45 for COM 325 data set. MPCA had a high positive correlation while GPCA had a low positive correlation. Figure 6: is a chart showing the Pearson Correlation analysis of the evaluation result of MPCA and GLSA. Table II represents the coefficient for determination performance measure, for COM 317 data set, MPCA gave 0.49 while GLSA gave 0.16. For COM 325 data set, MPCA had 0.56 while GLSA had 0.20. R2 results indicated that the regression line of MPCA Approach model fits the n-gram term representation than the GLSA Approach Model. Figure 7: is a chart showing the coefficient of determination R^2 analysis of the evaluation results for MPCA and GLSA systems. These performance evaluation shows that MPCA is a better feature extraction techniques.

Table I: Pearson Correlation Result For MPCA And GLSA

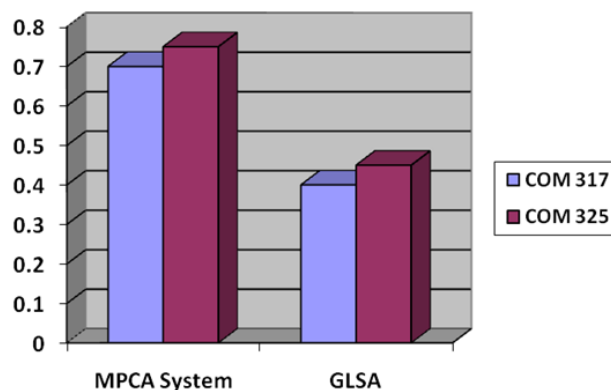|	MPCA System	GLSA
COM 317	0.7	0.4
COM 325	0.75	0.45

Figure 6: A chart of pearson corellation evaluation measure of MPCA and GLSA

Table II: Coefficient of Determination Result for MPCA And GLSA

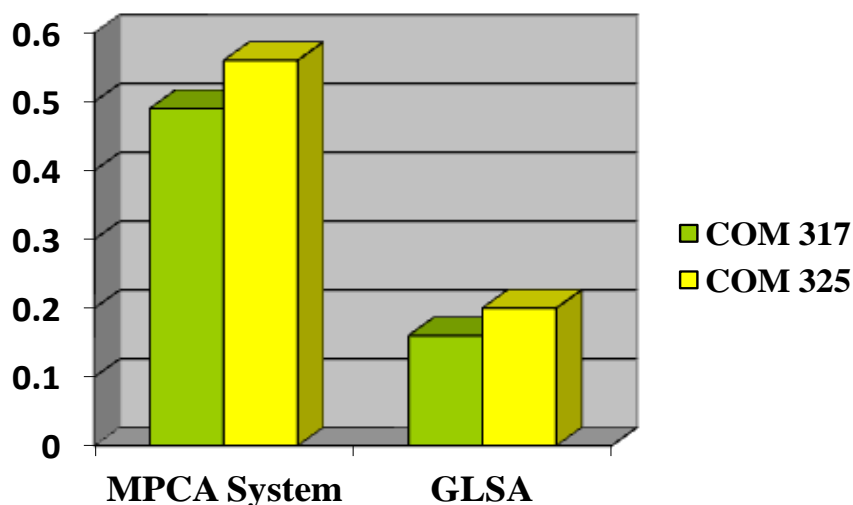|	**MPCA System**	**GLSA**
COM 317	0.49	0.16
COM 325	0.56	0.20

Figure 7: Coefficient of determination Chart for GLSA and MPCA Systems

VI. Conclusion

In this study, it has been established that Modified Principal Component Analysis MPCA approach has a high positive correlation with the original assessor when compared with the Generalized Latent Semantic Analysis technique (GLSA) for automated theory type marking.

References

[1]   Allen M.J. (2004): Assessing Academic Programs in Higher Education. Bolton M.A, Anker Publishing, Alexandria: Association for Supervision and Curriculum Development, pp.70-75

[2]   Ade-Ibijola, A.O., Wakama, I. and Amadi, J.C. (2012): An Expert System for Automated Essay Scoring(AES) in Computing using shallow NLP Techniques for Inferencing. International Journal of Computer Applications Vol. 51, pp. 37-45.

[3]   Adeyanju, I.A. (2012): Generating Weather Forecast Texts with Case Based Reasoning. International Journal of Computer Application Vol. 45 pp 35-40.

[4]   Adeyanju, I.A., Wiratunga, N., Recio-Garcia, J. and Lothian, R. (2010): Learning to Author Text with Textual (CBR) Proceeding of European Conference on Artificial Intelligence(ECAIA), pp. 777-782

[5]   Attali Y. and Burstein, J. (2006): Automated essay scoring with e-raterR v. 2. *The Journal of Technology, Learning and Assessment*, 4(3).

[6]   Burstein, J., Leacock, C., Swartz, R. (2001): Automated evaluation of essay and short answers. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborough University, Loughborough, UK, pp. 55

[7]   Elliot, S. (2003): Intellimetric™: From here to validity. In M.D. Shermis& J.C. Burstein (Eds.), *Automatic essay scoring: A cross-disciplinary perspective* (pp. 71-86). Mahwah, NJ: Lawrence Erlbaum Associates.

[8]   Fenwa, O.D., Omidiora, E.O. and Fakolujo O.A. (2012): Development of a Feature Extraction Technique for Online Character Recognition Systems, Innovative Systems Design and Engineering, UK, 3(3); pp. 10-23.

[9]   Guven, O. B, and Kahpsiz O. (2006): "Advanced Information Extration with n-gram based LSI" in Proceedings of World Academy of Science, Engineering and Technology, vol. 17 pp. 13-18.

[10]   Islam, M.M., Hogue, A.S.M.L.,(2012): "Automated Essay Scoring Using Generalized Latent Semantic Analysis" *Journal of Computers vol 7 no 3 pp.616-626.*

[11]   Jolliffe, I.T. (2002): Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28

[12]   Jusoh S. and Alfawareh, H.M.(2007) "Natural language interface for online sales," in *Proceedings   of   the   International Conference on Intelligent and Advanced System (ICIAS2007)*. Malaysia: IEEE, pp. 224–228.

[13]   Mason, O. and Grove-Stephenson, I. (2002): Automated free text marking with paperless school. In M. Danson (Ed.), *Proceedings of the Sixth International Computer Assisted Assessment Conference*, Loughborough University, Loughborough, UK, pp. 85-95

[14]   Oduntan, O. E., Adeyanju, I. A., Olabiyisi, S. O. and Omidiora, E. O.(2015) "Evaluation of N-gram Text Representations for Automated Essay-Type Grading Systems", *International Journal of Applied Information Systems* 9(4):25-31, July 2015. Published by Foundation of Computer Science, New York, USA.

[15]   Oduntan, O. E., Adeyanju, I. A., Olabiyisi, S. O. and Omidiora, E. O. (2016), A Modified Principal Component Approach to Automated Essay-Type Grading in Proceedings of the Future Technology Conference USA, pp 96 -100

[16]   Page, E.B. (1996): Grading essay by computer: Why the controversy? Handout for *NCME Invited Symposium*.

[17]   Rao, R. (2003): "From unstructured data to actionable intelligence,"in*Proceedings of the IEEE Computer Society*.pp 28

[18]   Rudner, L., and Gagne, P. (2001): An overview of three approaches to scoring written essays by computer. *ERIC Digest*, ERIC Clearinghouse on Assessment and Evaluation. (ERIC Document Reproduction Service No. ED458290).

[19]   Rudner, L.M. and Liang, T. (2002): Automated essay scoring using Bayes' Theorem. *The Journal of Technology, Learningand Assessment, 1*(2)*, pp.*3-21.

[20]   Salton, G., Wong, A., Yang, C.S. (1975): A vector space model for automatic indexing Communications of the ACM. Vol 18 no 11 pp.613-620

[21]   Vantage Learning (2003): *IntelliMetric™ : From here to validity.* (Report No. RB-504). Newtown, PA: Vantage Learning, pp. 45 - 70

[22]   Wresch, W. (1993): The imminence of grading essays by computers—25 years later. *Computers and Composition 10*(2), pp. 45-58.

[23]   Yang, Y., Buckendahl, C. W., Juszkiewicz, P. J., and Bhola, D. S. (2002): A review of strategies for validating computer automated scoring. *Applied Measurement in Education, 15*(4), pp.391-412.