Deanship of Graduate Studies
Al-Quds University

Automatic Essays Scoring

Hamzeh Abdel Hamid Mujahed Mujahed

M.Sc. Thesis

Jerusalem-Palestine

1430 / 2009

Automatic Essays Scoring

Prepared By:
Hamzeh Abdel Hamid Mujahed Mujahed

B.Sc.: Computer Systems Engineering, 2002, Palestine Polytechnic University, Palestine

Supervisor: Dr. Labib Arafeh

A thesis submitted in Partial fulfillment of requirements for the degree of Master of Electronics and Computer Engineering/ Department of Electronics and Computer Engineering/ Faculty of Engineering/ Graduate Studies

Al-Quds University

1430 / 2009

Al-Quds University
Deanship of Graduate Studies
Electronics and Computer Engineering Department

Thesis Approval

Automatic Essays Scoring

Prepared By: Hamzeh Abdel hamid Mujahed
Registration No.: 20520140

Supervisor: Dr. Labib Arafeh

Master thesis submitted and accepted, Date: June, 20, 2009
The names and signatures of the examining committee members are as follows:

1. Head of Committee: Dr. Labib Arafeh Signature: ....
2. Internal Examiner : Dr. Amjad RATTROUT Signature: ....
3. External Examiner : Dr. Adi Anani Signature: ....
4. External Examiner : Dr. Mahmoud SAHEB Signature: ....

Jerusalem - Palestine

1430 / 2009

Dedication

This thesis is dedicated to:

Words fail me to express my appreciation to my wife Rabab whose dedication, love and persistent confidence in me, has taken the load off my shoulder. I owe her for being unselfishly let her intelligence, passions, and ambitions collide with mine.

My dear three children, Taima', Basel and Saja have been very tolerant of three years worth of hours schooling to obtain this degree. They always encourage me through their love to improve myself and I greatly appreciate that.

My parent's who offered me unconditional love and support throughout the course of this thesis.

To Dr. Labib Arafeh, Dr. Hussien Jaddu and Dr. Ali Jamoos from Al-Quds University; you have all been extraordinary teachers, dedicating your valuable time and energy to the growth and development of your students.

Finally, this thesis is dedicated to all those who believe in the richness of learning.

January, 2009
Hamzeh Abdel Hamid Mujahed,

Declaration:

I certify that this thesis submitted for the degree of Master, is the result of my own research, except where otherwise acknowledged, and that this study (or any part of the same) has not been submitted for a higher degree to any other university or institution.

Signed:........Hamzeh.........................

Hamzeh Abdel Hamid Mujahed Mujahed

Date:.........20, June, 2009...........................

Acknowledgement

There are a number of people without whom this thesis might not have been written, and to whom I am greatly indebted.

Dr. Labib Arafeh has been the ideal thesis supervisor. His sage advice, insightful criticisms, and patient encouragement aided the writing of this thesis in innumerable ways.

I gratefully acknowledge Professor Younis Amro (president of Al-Quds Open University) for his advice and support during my study.

I am grateful to Mr. Yousef Oriqat from Al-Quds Open University for his support and help.

I would like to show my special thanks to the staff of the Engineering College -Department of Electronic and Computer Engineering in Al-Quds University who taught me during my Master degree study.

I am grateful to Dr. Nadia Qawasmi, from Al-Quds Open University (QOU) and Mr. Mohammad Jabary from Palestine Polytechnic University(PPU) for their fruitful cooperation to support us with data to check the developed models. The valuable support and encouragement is so important. Your assistance went far beyond the call of duty, and I appreciate it immensely.

I would like to express my gratitude to all those who gave me the possibility to complete higher studies especially my colleagues at Al-Quds Open University.

Finally, I would like to thank everybody who was important to the successful realization of thesis, as well as expressing my apology that I could not mention personally one by one.

Abstract

In this study, an AES system has been developed. The idea behind our proposed AES is to grade the essays by identifying the main keywords in the essays and its synonyms that determined by the teacher, and processing these keywords using the modeling approach-based techniques including Fuzzy Logic, Clustering, and Nuero-Fuzzy. Three models have been developed; the first model is Multiple Input Single Output (MISO) Mamdani Model; the second model is MISO Sugeno with back propagation optimization technique and the last one is Sugeno subtractive clustering with hybrid optimization technique. These developed AES models are capable to identify up to 15 keywords, each of which has up to 4 synonyms. A 100-word history essay has been used to test the developed AES. A 1080-datasets have been generated using an automatic answers generator depending on 13 questions. Using the cross-validation method, the data has been splitted into 718 samples for training and 362 samples for testing. To check the adequacy of the models, we have used the correlation coefficient to measure the agreement between the theoretical (actual) and predicted marks. Two error measures have also been used to check the accuracy including; Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE).

Different scoring dimensions (content, structure, syntax, etc.) commonly used in developing AES systems. In our developed models, we adopted content dimension in scoring the written essay. One of the notable gaps in AES based on content dimension is ignoring the negation's issue in the sentences and the order of the words in the sentence. In our models we have succeeded to address the problem of negation issue and to identify the necessary procedures to solve the problem of words order in the sentence by suggesting the development of language parser. The obtained results for the answers to some questions containing negation are promising and show high agreement between actual and predicted marks.

The obtained correlation coefficient between the theoretical (actual) and the predicted marks for the first model(Multiple Input Single Output (MISO) Mamdani Model) ranges between 0.887 and 0.9969 with an average value of 0.9863 for training and 0.9599 for the testing data. The MAPE values range between 0.017 and 0.1619 with an average value of 0.0854 for the training and 0.1189 for the testing data. The RMSE values range between 0.0378 and 0.2947 with an average value of 0.087 for training and 0.2073 for testing data.

The correlation coefficient obtained between the theoretical and predicted marks in the second model (MISO Sugeno with back propagation optimization technique) ranges between 0.9347 and 0.9966 with an average correlation of 0.9906 for training and 0.9675 for the testing data. The MAPE value ranges between 0.015 and 0.399 with an average value of 0.1039 for the training and 0.1071 for the testing data. The RMSE ranges between 0.043 and 0.3775 with an average value of 0.0809 for training and 0.2069 for testing data.

The third model is subtractive clustering Sugeno. The obtained correlation coefficient between the theoretical and predicted marks ranges between 0.9121 and 0.9977 with an average correlation of 0.9948 for training and 0.9712 for the testing data. The MAPE value ranges between 0.015 and 0.1516 with an average value of 0.0399 for the training and 0.1024 for the testing data. The RMSE ranges between 0.0328 and 0.2763 with an average value of 0.0526 for training and 0.1837 for testing. It is noted that the results obtained with subtractive clustering model are the best results since the correlation between predicted and

actual marks are the highest even though the number of Membership Function(MF) are fixed(seven MF for each input), the type of MF are triangular one and the number of rules is less than other two models. However, other researchers have obtained correlation between actual and predicted marks ranges between 0.87 (i.e. Project Essay Grader PEG) to 0.98 (i.e. Modeling Techniques Applied to Short Essay Auto-grading Problem), although the sample essays under testing and other factor (modeling techniques used, number of samples, scoring dimension used)are not similar. The results show that we can adopt these models for AES purposes with high correlation between actual and predicted marks when we tested it using the online system. Thus, we may conclude that the preliminary and promising results demonstrate the suitability, adequacy and competitive of using the modeling techniques to solve the automated essay scoring problem.

Further more, a powerful online-supported graphical user interface system(Fuzzy Automatic Essay Scoring System (FAESS)) have been developed to allow the user to feed the system with keywords and synonyms to score the essays in a more usable and flexible way. The interface is simple to use and has the ability to score the essays in normal mode and fuzzy based scoring mode. Stand alone application was built to make the system easy and simple to use for testing by different teachers. We have examined the stand alone application using the unseen data obtained from the local universities and the results are promising.

For further investigation, we have established a contact with other local universities (Al-Quds Open University (QOU) and Palestine Polytechnic University (PPU)) to support us with data. From QOU we obtained a sample of data (18 samples) related to one essay-type question in subjects other than history such as English literature. When this data was tested using FAESS and subject to the developed models (MISO Grid partition Sugeno), the calculated correlation coefficient value between theoretical and predicted marks in QOU data was 0.9954, the value for MAPE is 0.1292 and the value of RMSE equal to 0.348. We have been also in contact with PPU. From PPU we obtained answers for two questions related to information technology with 11 sample of answers for the first question and 23 samples for the second one. The obtained correlation value between actual and predicted marks for the first sample was 0.9755. whereas the a correlation value for the second sample equal to 0.9745 with error measures MAPE equal to 0.0696 and RMSE equal to 0.266. The obtained results from unseen data from local universities support us with promising results since we had a high correlation coefficient between actual and predicted marks and small values in the error measures (MAPE and RMSE). We may notice that the differences between the actual (human) score and the predicted one range between 0% and 8.3%. That is in marks, it ranges between 0 to 2.5 marks out of the total 30 marks. Although, the number of these samples are not so large (23, 11, and 18), we may conclude that this approach is adequate and suitable to address and solve the AES problem.

Further investigation is still required with more samples of data (essays) and of different types of field other than historical questions. Further investigation required also to identify and to implement the required procedures to solve the problem of words order in the sentence using language parser.

Table of contents

List of Tables

List of Figures

List of Appendices

Chapter One

Introduction

1.1 Introduction

Essays scoring fall into two categories: short answers and open ended essays answers. Short answer essays compromised of few words to few lines mainly related to specific topic and mainly evaluated for their content dimension. For example define questions. Open ended answers essays constitute of several lines or paragraphs and several writing dimensions (i.e. content, style, vocabulary, rhetorical structure and syntax /grammar) must be combined to produce an effective scoring. Educational institutions have strong interest in the development of scoring methods and constantly exploring new techniques to effectively score essays (Oriqat, 2007).

Automated Essay Scoring (AES) has been a topic of research for over four decades. A limitation of all past work is that the essays have to be in computer readable form (Srihari, et al, 2007). Automated scoring of exams consisting of written text would be doubtless of advantage to teachers. However, recent advances in computer techniques have opened up the possibility of being able to automate the scoring of free text responses typed into a computer without having to create systems that fully understand the answers (Pulman and Sukkarieh, 2005).

When large numbers of answers are submitted to be scored, teachers find themselves bogged down in their attempt to provide consistent evaluations and high quality feedback to students within as short a timeframe as is reasonable, usually a matter of days rather than weeks. Educational administrators are also concerned with quality and timely feedback, but in addition must manage the cost of doing this work. Clearly an automated system would be a highly desirable addition to the educational tool-kit, particularly if it can provide less costly and more effective outcome (Palmer, et al, 2002).

A number of studies have been conducted to assess the accuracy (measurement of the degree of agreement between actual marks and predicted marks) of the AES systems with respect to writing assessment. The results of several AES studies reported high agreement rates between AES systems and human raters. AES systems have been criticized for lacking human interaction, and their need for a large corpus of sample text to train the system. Despite its weaknesses, AES continues attracting the attention of public schools, universities, testing companies, researchers and educators (Dikli, 2006).

There is a previous thesis work done at A-l-Quds University (Oriqat, 2007). The work concentrated on using fuzzy logic to score the short essays based on short answers by determining five main keywords and synonyms (inputs). These inputs have been processed by developing models based on fuzzy and Neuro-Fuzzy approaches. The obtained result from the models was promising and showed high agreement between actual and predicted marks. Our Fuzzy Automated Essays Scoring System (FAESS) is represented in fig. (1.1). we have pre-process stage to determine the fifteen main keywords and synonyms necessary for the systems to predict the mark for longer answers. The main difference between the two approaches relies on number of keywords which are important factors to deal with longer answers. Also in our work we have concentrated in scoring the essays on content dimension and we have explored the importance / impact of words' order in the sentence and we have also explored negation's issue in the sentence, and ways to solve.

We have focused on building models for Automatic Essays Scoring. The purpose of these models is to examine and to explore the adequacy of using fuzzy techniques to score written essays. In order to build fuzzy inference systems, we need to prepare and collect a reasonable size of data to train and test the models. In our models, we generated 1080 dataset related to 13 questions (see section 4.3.1). The dataset has been processed by developing three models including Multiple Input Single Output (MISO) Mamdani model, Grid Partition Sugeno with back propagation optimization Model, subtractive clustering Sugeno model. The preliminary obtained results are promising and show high agreement between actual and predicted marks. For further investigation, we received dataset from Palestine Polytechnic University (PPU) and Al-Quds Open University (QOU) in order to compare the results obtained from our approaches with those obtained from other universities.

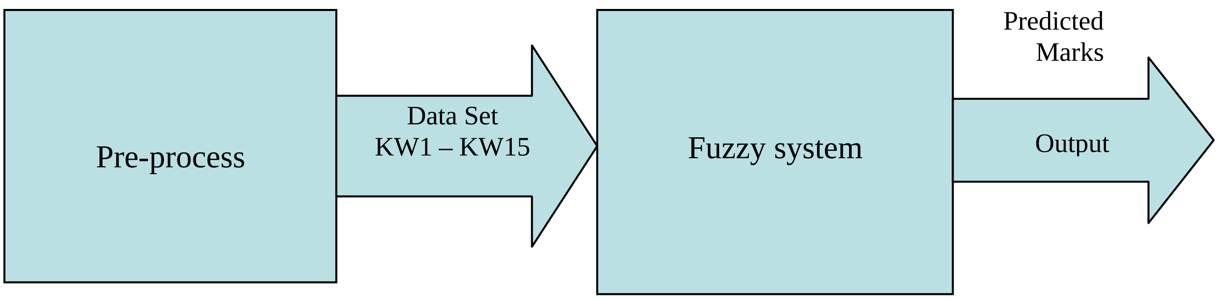


Fig. 1.1: FAESS block diagram

1.2 Rationale

Essays are considered by many researchers as the most useful tool to assess learning outcomes, implying the ability to recall, organize and integrate ideas, the ability to express oneself in writing and the ability to supply merely than identify interpretation and application of data. It is in the measurement of such outcomes, corresponding to the evaluation and synthesis levels of the Bloom’s (1956) taxonomy that the essay questions serve their most useful purpose (Valenti, et al, 2003).

Reasoning based on fuzzy approaches has been successfully applied for inference of multiple attributes containing imprecise data; in particular, Fuzzy Rule-Based Systems (FRBS) provide intuitive methods of reasoning based on linguistic models. It is worth mentioning that the applications of FRBS in numerous real world classification problems have been mentioned frequently in most of the literature in the area of fuzzy systems (for example; medical instrumentation, and decision-support systems). Recent developments in this area also show the availability of FRBS, which allow interpretation of their inference results and have high accuracy rates. This is very important in providing a platform for the application of FRBS in student academic performance evaluation (Rasmani, 2002).

Teaching staff around the world are faced with a perpetually recurring problem: how do they minimize the amount of time spent on the relatively monotonous tasks associated with scoring their students' essays? With the advent of large student numbers, often counted in thousands in first year common core units, the scoring load has become both times consuming and costly(Chung and O'neil,1997 ;Cohen et al ,2003). A system that can automate the tasks is currently just a dream for most staff. One of the most thankless tasks in all academia is that of scoring, particularly when there is no need to supply individual feedback as in the case of examination scoring (Palmer, et al, 2002).

Important reasons exist for creating a new evaluation method for student academic performance, they include:

a. *Evaluation of academic performance using natural language.* Instead of using numerical values, linguistic terms can be awarded to represent student achievement in each of the assessment components. The use of natural language such as "good", "very good" and "extremely good" would allow more flexible ways to make judgment on students' performance. Therefore a method that is able to aggregate information given in the form of natural language is needed. In our developed models, we used the numerical values (marks) instead of natural language such as "good" and "very good".

b. To handle uncertain scores. The primary method of assessment usually involves awarding numerical marks by an evaluator. Such marks are usually given according to a given scoring scheme. These marks are usually numerical values that may fluctuate a little as different evaluators may award different marks. Evaluation of a student's work may be affected by the evaluator's experience, sensitivity and the standard used. To avoid this, cooperation with other institutions and teaching experts will make the evaluation process more efficient and reliable. Thus marks awarded by an evaluator to represent student performance are only an approximation. Although linguistic terms (e.g. bad, good, very good, excellent etc.) have also been widely used to represent the final student's performance, their inherent nature of vagueness is often ignored. However, academic performance evaluation involves the measurement of ability, competence and skills. Ability, competence and skills are fuzzy concepts and can be approximately captured in fuzzy terms (Rasmani, 2002).

1.3 Problem Definition

The use of computer generated-scores contributes significantly to the standardization accuracy in the test scores. There are several important reasons why AES is needed. Firstly, the automatic scoring can be used as an indication of a student's level of understanding. Secondly, AES is an important purpose for making academic decisions about the performance issue. Finally, the academic staffs need to minimize the amount of time spent on the relatively monotonous tasks associated with scoring their students' essays and to find a new method and approach to score the written essays.

Traditionally, automatic marking (scoring) has been restricted to item types such as multiple choice that narrowly constrain how students may respond. More open ended items have generally been considered unsuitable for machine scoring because of the difficulty of coping with the myriad ways in which credit-worthy answers may be expressed. Successful automatic scoring of essays would seem to presuppose an advanced level of performance in automated natural language understanding. However, recent advances in Natural Language Processing (NLP) techniques, which extract the meaning from a given text with varying degree of success, have opened up the possibility of being able to automate the scoring of free text responses typed into a computer without having to create systems that fully understand the answers (Sukkarieh, et al, 2004).

The main goal of our approach is to explore the application of the fuzzy modeling approach and Nuero –Fuzzy modeling techniques to contribute solving the problem of essay scoring and obtain effective measurements for essays scoring based on content dimension.

1.4 Research Objectives

The objectives of the study research include:

- Developing several models for Automatic Scoring System, including:
  - Multiple Input Single Output (MISO) Mamdani Fuzzy Inference System (FIS) model to evaluate its adequacy to solve essays scoring problems.
  - Sugeno FIS in association with other techniques including subtractive clustering or Grid Partition with back propagation optimization technique models.
  - Combining different techniques to score essays.
- Maintaining a database of history related answers to train and test the developed model.
- Exploring the importance / impact of words' order in the sentence.
- Exploring the negation's issue in the sentence, and ways to solve.
- Checking the capability of the proposed models using three measures (Correlation coefficient, Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE)).
- Checking the developed models using data from other international institution.
- Developing User-Interface System for stand alone application /online application using web technology.

1.5 Contributions

In this study, we have developed several models based on Fuzzy Inference Systems with different optimization techniques. The importance of our approaches is to check the adequacy of these models to score the written essays automatically. We have generated data to train and test the models, and we have used data from other institutions to check the capability of the models to score essays. Three models have been developed; the first model is Multiple Input Single Output (MISO) Mamdani Model; the second model is MISO Sugeno with back propagation optimization technique and the last one is Sugeno subtractive clustering with hybrid optimization technique.

In developing our AES models, we adopted content dimension in scoring the written essay. One of the notable gaps in AES based in content dimension is the ignorance of the negation's issue in the sentences and the order of the words in the sentence. In our models we succeeded to solve the problem of negation issue and to identify the necessary procedures to solve the problem of words order in the sentence by using language parser as it will be discussed later in the next chapters.

We have used the correlation coefficient to check the adequacy of the developed models, and to measure the degree of agreement between theoretical (actual) and predicted marks. MAPE and RMSE, which are two error measures, have been used to check the accuracy of the developed model. These developed models are capable to identify up to 15 keywords, each of which has up to 4 synonyms. A 100-word history essay has been used to test the developed AES models. A 1080-datasets based on 13 questions have been generated using automatic answers generator. Using the cross-validation method, the data has been splitted into 718 for training and 362 for testing.

The obtained correlation coefficient between the theoretical (actual) and predicted marks in the first model ranges between 0.887 and 0.9969 with an average correlation of 0.9863 for the training and 0.9599 for the testing data. The correlation coefficient obtained between the theoretical and predicted marks in the second model ranges between 0.9121 and 0.9977 with an average correlation of 0.9948 for the training and 0.9712 for the testing data. Finally, for the third model, the correlation coefficient between the theoretical and predicted marks ranges between 0.9347 and 0.9966 with an average correlation of 0.9906 for the training and 0.9675 for the testing data.

Table (1.1) summarizes the obtained average results for correlation, MAPE and RMSE for the three developed models.

Table 1.1: Average value for Correlation, MAPE and RMSE for the three developed models

Models	Average Correlation		Average MAPE		Average RMSE	
	Training	Testing	Training	Testing	Training	Testing
Sub. Clustering with hybrid optimization model	0.9948	0.9712	0.0399	0.1024	0.0526	0.1837
MISO Mamdani model	0.9863	0.9599	0.1854	0.1189	0.087	0.2073
Grid partition Sugeno with back propagation	0.9906	0.9675	0.1039	0.1071	0.0809	0.2069

The preliminary and promising results demonstrate the adequacy of using the modeling techniques in solving the automated scoring systems. The results show that we can adopt these models for AES purposes with high degree of agreement between theoretical and predicted marks. To simplify the use of our models, we have developed an online system supported by a powerful graphical user interface to allow the user to feed the system with keywords and synonyms to score the essays in flexible way. The interface is simple to use and has the ability to score the essays in normal mode and fuzzy based scoring mode. For further investigation, we obtained unseen data from local universities and we have examined it on our stand alone application and the results are promising.

1.6 Structure of the thesis

The thesis contains the following chapters:

Chapter one starts with the introduction of Automatic Essays Scoring systems, and their importance for institutions and educational centers. While chapter two briefly overviews other AES systems and approaches by comparing these systems and the ways and technologies used to score the essays. At the end of this chapter we summarize these systems in table by considering scoring methods and performance achieved. Chapter three briefly introduces the various modeling approaches and techniques including fuzzy logic, Fuzzy Inference System, Adaptive Nuero-Fuzzy Inference Systems (ANFIS) and Artificial Neural Networks (ANN). Chapter four presents the generation of the dataset used and the various developed models which include Multiple Input Single output Mamdani Model, Sugeno FIS with back-propagation optimization model and Sugeno with subtractive clustering FIS. Chapter five discusses the obtained results, comparison between developed models and comparison with other AES system. Finally, chapter six concludes and represents the future work in the field.

Chapter Two

Literature review of Automated Scoring Systems

2.1 Introduction

A number of studies have been conducted to assess the accuracy and reliability of the AES systems with respect to writing assessment. The results of several AES studies reported high agreement rates between AES systems and human raters (Dikli, 2006).

An essay is usually a short piece of writing and composed of one paragraph or more. A paragraph is a series of sentences that are organized and coherent, and are all related to a single topic. An essay is often written from an author's personal point of view. Essays can be literary criticism, student’s answers to specific questions, political manifestos, learned arguments, observations of daily life, recollections, and reflections of the writer.

2.2 Types of Questions in AES Systems

We can classify questions for computer scoring of free text answers according to three different criteria:

• ***According to the number of right/correct answers:***

1. Convergent questions: When there is only one correct answer. They deal with concrete facts. For example, “Who is the author of The Quixote? ” (Mar´ın, 2004).
2. Divergent questions: When there are many right answers, which are usually opinion or hypotheses, such as “What do you think about nuclear energy?” (Mar´ın, 2004).

• ***According to the answer of the question:***

1. Open questions: When there are many right answers but they are based on textual facts. For example, “Describe the water cycle”.
2. Closed questions: Questions that expect a yes or no, true or false or a choice among several options. For example, “Would you use processes or threads to solve this problem?”
3. Counter questions: Questions asked by the student to clarify the teacher’s question. For instance, the answer to “How would you manage an organization?” might be “It depends, what kind of organization?”
4. Numerical questions: When the answer involves numerical calculations.

• ***According to its function, the question may ask the students to:***

1. Make a choice: The simplest version is a multiple question, but it could be rewritten as in the following example: “If you mix black and white you get grey, pink or yellow?”
2. Find out if a sentence is true: It is the same case as above. Thus, if the question is formulated like “Say if the affirmation “Distribute operating systems are more robust than others” is true”. From the point of view of fully assessing the student knowledge, the teacher could also require a justification.
3. Develop ideas: The student has to fill a given text area to response the question.
4. Calculate: The answer is the outcome of a numerical analysis.
5. Study a graphical image: For example, by clicking with the mouse on some region.
(Mar´ın, 2004).

2.3 Latent Semantic Analysis (LSA)

A powerful approach to AES is based on a technique developed in the information retrieval community known as latent semantic indexing. Automated essay scoring with latent semantic analysis (LSA) has recently been subject to increasing interest. Although previous researchers have achieved grade ranges similar to those awarded by humans, it is still not clear which and how parameters improve or decrease the effectiveness of LSA (Wild, et al, 2004). A matrix for the essay is built, and then transformed by the algebraic method of singular value decomposition (SVD) to approximately reproduce the matrix using reduced dimensional matrices built for the topic domain. Using SVD new relationships between words and documents are uncovered, and existing relationships are modified to represent their significance. Using LSA the similarity between two essays can be measured despite differences in individual lexical items (Srihari, et al, 2007).

2.4 Overview of Automated Essays Scoring Systems

We will provide an overview of current approaches to the automated assessment of essays scoring. There is currently some systems available either as commercial systems or as the result of research in this field, we will give brief explanation about some of these systems: these include Project Essay Grade (PEG), Intelligent Essay Assessor (IEA), Electronic Essay Rater (E-Rater), Modeling techniques Applied to Short Essay Auto-Grading Problem and BETSY. For each system, the general structure and the performance are described.

2.4.1 Project Essay Grade (PEG):

PEG is one of the earliest (1966) and longest-lived implementations of automated essay grading. It was developed by Page and others and primarily relies on style analysis of surface linguistic features of a block of text. Thus, an essay is predominantly graded on the basis of writing quality (grammar, punctuation, etc), taking no account of content (Valenti, S., et al, 2003). Page and his colleagues use the terms trins and proxes while explaining the way PEG generates a score. While trins refer to the intrinsic variables such as fluency, diction, grammar, punctuation, etc., proxes denote the approximation (correlation) of the intrinsic variables. Thus, proxes refer to actual counts in an essay (Dikli, 2006). Specific attributes of writing style, such as average word length, number of semicolons, and word rarity are examples of proxes that can be measured directly by PEG to generate a grade. For a given sample of essays, human raters grade a large number of essays (100 to 400), and determine values for up to 30 proxes. The grades are then entered as the criterion variable in a prediction equation with all of the proxes as predictors, and beta weights are computed for each predictor.

The scoring methodology in PEG employs is simple. The system contains a training stage and a scoring stage. PEG is trained on a sample of essays in the former stage. In the latter stage, proxy variables (proxes) are determined for each essay and these variables are entered into the prediction equation. Finally, a score is assigned by computing beta weights (coefficients) from the training stage. PEG needs 100 to 400 sample essays for training purposes. PEG has been criticized for ignoring the semantic aspect of essays (Dikli, 2006). The PEG system is a good example of how a holistic score essay prediction system works. First, a group of training essays were collected and assigned a grade by multiple graders. In the initial PEG study, four graders were used to determine a holistic score of overall quality of the essay. The electronic versions of the essays were analyzed and features were collected. Correlation and regression analysis determined which variables were most closely correlated with the essay holistic score. Multiple variable regression techniques resulted in a prediction equation with the independent variables being the features and the dependent variable the predicted holistic score (Millett, 2006).

Regarding PEG's performance, Page's latest experiments achieved results reaching a multiple regression correlation as high as 0.87 with human graders (Valenti, et al, 2003).

2.4.2 Intelligent Essay Assessor (IEA):

Intelligent Essay Assessor was developed in the nineties and is based on the Latent Semantic Analysis (LSA) technique that was originally designed for indexing documents and text retrieval. LSA represents documents and their word content in a large two-dimensional matrix semantic space. Using a matrix algebra technique known as Singular Value Decomposition (SVD), new relationships between words and documents are uncovered, and existing relationship are modified to more accurately represent their true significance(Valenti, et al,2003). IEA uses the Knowledge Analysis Technologies™ (KAT) engine, a patented technology based on over twenty years of research and development. KAT provides a sophisticated computer analysis of text. It assesses the total content of an essay as well as the correlation between the essay's content and that of training essays previously scored by expert human readers (www.knowledge-technologies.com, 2008).

LSA is defined as "a statistical model of word usage that permits comparisons of the semantic similarity between pieces of textual information (Dikli, 2006). LSA first processes a corpus of machine-readable language and then represents the words that are included in a sentence, paragraph, or essay through statistical computations. LSA measures of similarity are considered highly correlated with human meaning similarities among words and texts. Moreover, it successfully imitates human word selection and category judgments. The underlying idea is that the meaning of a passage is very much dependent on its words and changing even only one word can result in meaning differences in the passage. On the other hand, two passages with different words might have a very similar meaning.
The underlying idea can be summarized as:
Meaning of word$_1$ + meaning of word$_2$ + ……… + meaning of word $_k$ = meaning of passage (Dikli, 2006).

In the LSA based approach, the text is represented as a matrix. Each row in the matrix represents a unique word, while each column represents context. Each cell involves the frequency of the word. Then, each cell frequency is considered by a feature that denotes not only the importance of the word in that context but also the degree to which the word type carries information in the domain discourse. Unlike other AES systems, IEA require smaller number of essays (answers) to train and only requires 100 pre-scored answers.

Regarding IEA performance, a test conducted on Graduate Management Admissions Test GMAT essays using the IEA system resulted in correlation with human graders between 85%-91% (Valenti, et al, 2003).

2.4.3 Electronic Essay Rater (E-Rater):

E-Rater uses a combination of statistical and NLP (mentioned in chapter three)techniques to extract linguistic features from the essays to be graded. Essays are evaluated against a benchmark set of human graded essays(Valenti, et al,2003) .E-Rater was initially used by Educational Testing Service (ETS) for operational scoring of the Graduate Management Admissions Test Analytical Writing Assessment GMAT. *E-rater* 2.0 uses 12 features when scoring an essay. Eleven of these reflect essential characteristics in essay writing and are aligned with human scoring criteria. The importance of this alignment is that it increases the validity of the scoring system. Validity here refers to the degree to which the system actually does what is intended. The E-Rater program has also been evaluated for testing of college level essays for non-native speakers of English for the Test of Written English (TWE) exam (Burstein and Chodorow, 1999). Although there were significant differences between the scores of native English speakers and native Chinese, Arabic or Spanish speakers writing in English, the E-Rater system, which is tuned for each topic using about 52 syntactic, discourse and other analysis variables, was able to predict the holistic scores exactly (Millett, 2006).

The features of E-rater include a syntactic module, a discourse module, and a topical-analysis module. These modules provide outputs for model building and scoring. E-rater has been trained on a set of essays scored by at least two human raters on a 6-point holistic scale to build models. The general procedure for a vector-spec model, which is used to capture the topic or vocabulary usage, is described as follows:
Training essays are converted into vectors of word frequencies, and the frequencies are then transformed into word weights. These weight vectors populate the training space. To score a test essay, it is converted into a weight vector, and a search is conducted to find the training vectors most similar to it, as measured by the cosine between the test and training vectors. The closest matches among the training set are used to assign a score to the test essay.

Here, a vector can be described as the mathematical representation of an essay. Moreover, word frequencies can be calculated by counting the words in a paragraph and dividing by the number of their occurrence (each time a word appeared in a paragraph). While word weights refer to the frequency divided by the number of words in an essay, training space refers to the entire set of vectors that were generated from the training essays. Finally, in this context cosine is the distance between testing and training vectors (Dikli, 2006).

Regarding the performance of E-rater, Over 750000 GMAT essays have been scored, with correlation agreement between human expert and system above 97% (Valenti, et al, 2003).

2.4.4 Bayesian Essay Test Scoring sYstem (BETSY):

The goal of the system is to determine the most likely classification of an essay into a four point nominal scale (e.g. extensive, essential, partial, unsatisfactory) using a large set of features including both content and style specific issues(Valenti, S., et al,2003). The underlying idea of BETSY is the classification of texts based on trained materials. BETSY uses a large set of items. The "items" refer to a large set of essay features. These essays features include content related features such as specific words and phrases, frequency of certain content words, form related features including number of words, sentence length, number of verbs, number of commas and others, e.g., the order certain concepts are presented and the occurrence of specific noun verb pairs. It is claimed that BETSY™ includes the best features of PEG, LSA, and E-rater along with its own essential characteristics (Dikli, 2006).

Regarding the performance of BETSY; two text classification models that were calibrated using 462 essays with two score points. The calibrated systems were then applied to 80 new pre-scored essays, with 40 essays in each score group. An accuracy of over 80% (The measuring of accuracy based on probability in math) (Valenti, et al, 2003).

2.4.5 Modeling techniques Applied to Short Essay Auto-Grading Problem:

The work in this model concentrated on developing a system that tries to mimic the way in which human kind think, reason, evaluate, etc. The main purpose from this AES is to explore the various modeling techniques approaches to automatically evaluate short answer. The works focused in building models and examining these techniques aiming to produce an accuracy measures between actual and predicted marks.

The system adopted the short answers and based on input, process and output. The input is assessed subject, which is related to information technology. The output will be the predicted marks. The process was divided into pre process system and the developing techniques. The pre process system was used to construct data set (900 dataset). Five input main keywords with specific constraints. The FIS and neural approaches were used to developed models.

The main criticism to this system is neglecting the negation (Not, nor, etc) issue in the sentences and the order of the words in the sentences. The second criticism for this system is the difficulty to use it since it doesn't have Graphical user interface or stand alone application.

Regarding the performance of modeling techniques applied to Short Essay Auto-Grading Problem. Over 900 short answers have been scored, with correlation agreement between actual and predicted marks equal 98.15% (Oriqat, 2007).

2.5 AES Summary and Discussion

There have been several studies over the past three decades that have examined ways to apply technology to writing assessment. More recently, increasingly sophisticated computer technology has enable writing performance to be assessed using AES technology (Dikli, S., 2006). In 1996, Page introduced a distinction between grading essays for content and for style, where the former refers loosely to what an essay says, while the latter to "syntax and mechanics and diction and other aspects of the way it is said" . Some of the systems discussed in previously evaluate essays primarily either for content (IEA, Modeling techniques Applied to Short Essay Auto-Grading Problem and our models) or for style (PEG). Finally, some of the systems evaluate essays taking in account both aspects (BETSY) (Valenti, et al, 2003). During this study, we have developed models based on fuzzy and Nuero-fuzzy systems with different optimization techniques. The importance of our approaches is to check the adequacy of these models to score the written essays automatically based on content dimension. The preliminary results are promising.

There is a discussion and disagreement about which system can be considered as the best. For example (Valenti, et al, 2003) stated that, among IEA, E-rater and PEG, the best choice for evaluating writing style is PEG. This is because it relies on writing quality feature to determine the marks. Besides, it is simpler and it is consumes less CPU, on the other hand IEA and E-rater are better for scoring content.

For further discussion about our developed models (FAESS) and Oriqat model (modeling techniques Applied to Short Essay Auto-Grading Problem), the two systems are based on content dimension but the first one can deal with longer answers and the second just for short answer. Also the first system has the ability to deal with negation issue in the sentences whereas the second system neglects it. The most important distinguish between the two systems is the GUI in the first system and the stand alone application which allows the user to test the system online and provide the feedback (marks).

Each AES system needs different numbers of essays to train the system. Table (2.1) below compares the AES systems(Dikli, 2006).

Table 2.1: Comparison between AES Systems

AES Systems	Developer	Technique	Main Focus	Number of Essays used for Training
PEG	Page (1966)	Statistical	Style	100-400
IAE	Landauer, Foltz, & Laham,(1997)	LSA	content	100-300
E-rater	ETS development team (Burstein, et al.,1998)	NLP	Style and content	465
BETSY	Rudner 2002	Bayesian text Classification	Style and content	1000
Modeling techniques applied to short answers	Oriqat,Y.,2007	Fuzzy /Neuro-fuzzy system	content	900
FAESS	Mujahed,H.,2008	Fuzzy /Neuro-fuzzy system	content	1080

The demand for incorporating AES systems in writing assessment is increasing. Although some teachers and educators may fear that AES technology will eventually substitute humans, the producers of classroom based AES systems claim that the main role of these systems is not to replace teachers in writing classes but to assist them. Although AES is a developing technology the search for better machine scoring is ongoing as investigators continue to move forward in their drive to increase the accuracy and effectiveness of AES systems (Dikli, 2006) .The main weakness of all Automatic scoring systems is the lack of very large corpus of essays that may become a reference for every one interested in Automatic Essay scoring. It is a very important issue because most of these systems performances are strongly affected by reference texts.

In conclusion, there is no current system that could be highlighted as the best one by all comparisons. However, they resumed that the technology exist. In fact, just the combining of the main advantages of each one, the resulting system might be the ideal AES of free text answer tool.

2.6 Advantages and disadvantages in AES

This chapter finishes with an overview of the main advantages and disadvantages that it has been argued that computer scoring systems of free text answers systems have.

• Some advantages that have been cited are that computer automatic scoring can:

1. Achieve a fairer and more efficient assessment process: The computer does not get tired, bored, irritated or inattentive.
2. Provide more feedback to the students: Computerized scoring could avoid the problem of teachers that do not have time to assess their students' work.
3. Develop more effective instructional materials: These materials could be used for improving reading, writing and other communication abilities.
4. Detect plagiarism: An automatic grading system could detect plagiarism by looking the database of students' answers more easily than a human expert.
5. Reduce costs: If the system can be used in several departments and without considering the initial implantation phase, it could reduce some costs.
6. Improve the learning process: By engaging students with interesting questions that are given instant feedback once the students have answered them.
7. Challenge the students: The core idea is that whoever student able to fool the machine to achieve a good score, it is a student with a good knowledge of the domain and hence, he or she deserves the score obtained.

• Some disadvantages that have been cited include:

1. The computer is not a credible grading machine: Computer-based analyses are sometimes based solely on symbols manipulation according to some simple previously schemes programmed by humans. Hence, students with a peculiar writing style could be prejudiced.
2. The computer lacks human common sense and intelligence: There are still many things that only humans can do.
3. The student-teacher relationship gets lost: If teachers stopped reading the students' works they would ignore their students' progress.

4. The computer is useless for high-stakes assessment: For students one point can be crucial. It would not be acceptable for them that the computer has failed and hence, it has given them a lower grade.
5. The system’s scores might not be legally defensible: if a student’ exam automatically assessed would be legally considered (Mar´ın, 2004).

Chapter Three

Soft Computing Modeling Techniques

3.1 Introduction

Soft computing refers to a collection of computational techniques in computer science, machine learning and some engineering disciplines, which study, model, and analyze very complex phenomena: those for which more conventional methods have not yielded low cost, analytic, and complete solutions (zadeh, 1994).In recent years, the number and variety of applications of soft computing have increased significantly. The applications range from consumer products such as washing machines, cameras and microwave ovens to industrial process control, medical instrumentation, and decision-support systems. Intelligent system is performed with the help of Fuzzy Logic (FL) as a tool. Fuzzy Logic enables the development of rule-based behavior. The knowledge of an expert can be coded in the form a rule-base, and used in decision making. The main advantage of Fuzzy Logic is that it can be tuned and adapted if necessary, thus enhancing the degree of freedom of control (Rajagopalan., et al, 2003).

Artificial Neural Networks (ANN) and Fuzzy Inference Systems (FIS) (will be explained later on this chapter) have attracted the growing interest of researchers in various scientific and engineering areas due to the growing need of adaptive intelligent systems to solve the real world problems (Abraham, 2004). The success of FIS is mainly due to their closeness to human perception and reasoning, as well as their intuitive handling and simplicity, which are important factors for acceptance and usability of the systems (Castellano, et al, 2003).

In engineering and science, complex systems are usually described by mathematical models (kruse, et al, 1994). System modeling based on conventional mathematical tools (e.g., differential equations) is not well suited for dealing with will-defined and uncertain systems. By contrast, a fuzzy inference system employing fuzzy if then rules which model the qualitative aspects of human knowledge and reasoning processes without employing precise quantitative analyses. Fig.(3.1) below presents the architecture of FIS (Jang, 1993).

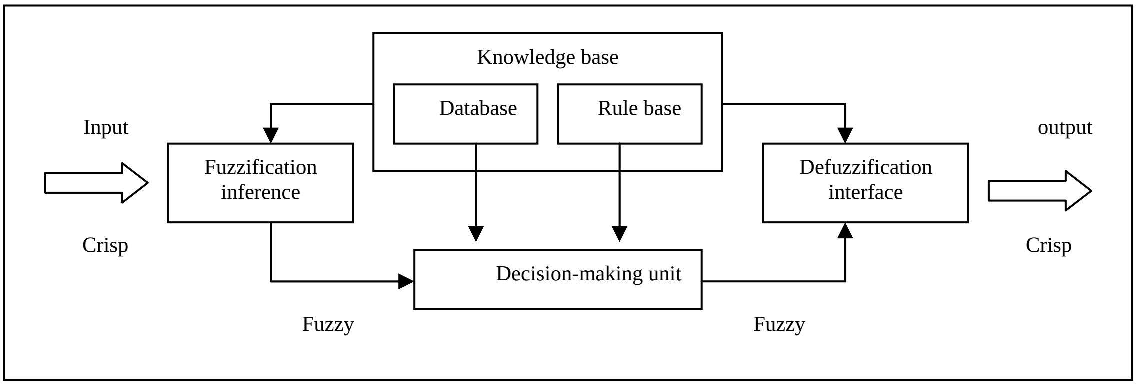


Fig. 3.1: Fuzzy Inference System

3.2 Fuzzy Sets

A classical set is a set with a crisp boundary. For example, a classical set A of real numbers greater than 6 can be expressed as: A = {x | x>6}.Fuzzy sets are sets whose elements have degrees of membership. Fuzzy logic starts with the concept of a fuzzy set. A fuzzy set is a set without a crisp, clearly defined boundary. It can contain elements with only a partial degree of membership. In fuzzy set the transition from " belong to a set" to " not belong" is gradual and this smooth transition is characterized by membership function that give fuzzy set flexibility in modeling commonly used linguistic expressions such as " the water is hot " or " the temperature is high"(Jang, 1997).

3.3 Membership Functions

A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The input space is sometimes referred to as the *universe of discourse*, a fancy name for a simple concept. Fig.(3.2) shows the membership functions for the linguistic values Fail, Pass, Good, V. Good and Excellent

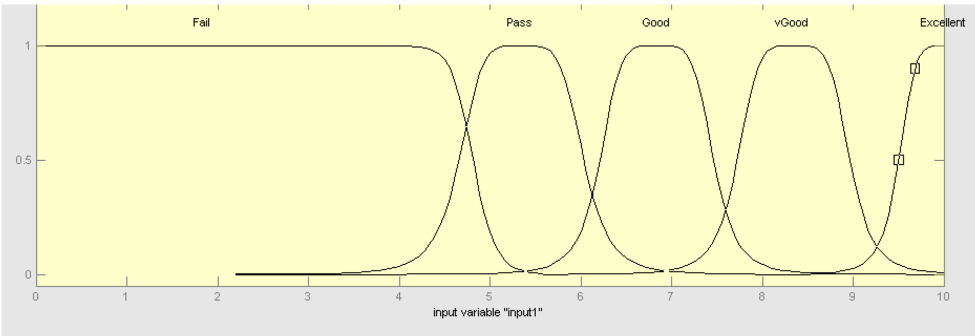


Fig 3.2: MF example for linguistic value (Fail, Pass, Good, V. Good and Excellent)

3.4 IF -THEN Rules

A rule-based system utilizes a model that represents human knowledge in the form of "IF-THEN" rules. Fuzzy sets and fuzzy operators are the subjects and verbs of fuzzy logic. These if-then rule statements are used to formulate the conditional statements that comprise fuzzy logic. A single fuzzy if-then rule assumes the form if x is A then y is B

Where A and B are linguistic values defined by fuzzy sets on the ranges (universes of discourse) X and Y, respectively. The if-part of the rule "x is A" is called the *antecedent* or premise, while the then-part of the rule "y is B" is called the *consequent* or conclusion. An example of such a rule might be:

If service is good then tip is average

The concept *good* is represented as a number between 0 and 1, and so the antecedent is an interpretation that returns a single number between 0 and 1. On the other hand, average is represented as a fuzzy set, and so the consequent is an assignment that assigns the entire fuzzy set B to the output variable y. In the if-then rule, the word "is" gets used in two entirely different ways depending on whether it appears in the antecedent or the consequent. In general, the input to an if-then rule is the current value for the input variable (in this case, service) and the output is an entire fuzzy set (in this case, average). This set will later be defuzzified, assigning one value to the output (Matlab R14, 2008).

3.5 Fuzzy Inference System (FIS)

Fuzzy Inference Systems are currently being used in a wide field of applications. In recent years, fuzzy modeling technique have become an active research area due to its successful application to complex system model, where classical methods such as mathematical and model-free methods are difficult to apply because of lack of sufficient knowledge (Priyono, 2005). One popular approach is to combine fuzzy systems with learning techniques derived from neural networks; such approaches are usually called Neuro-Fuzzy systems (Singh, et al, 2005). For the most complex system where few numerical data exist and only ambiguous or imprecise information may be available, fuzzy reasoning provides a way to understand system behavior by allowing us to interpolate approximately between observed input and output situation. Reasoning based on fuzzy approaches has been successfully applied for inference of multiple attributes containing imprecise data (Balog and Berta, 2001).

Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned. There are two types of Fuzzy Inference Systems that can be implemented in Fuzzy Logic Toolbox: Mamdani-type and Sugeno-type. These two types of inference systems vary somewhat in the way outputs are determined (Dogan, 2005). One of the basic differences between the Mamdani and Sugeno fuzzy structures is the fact that the consequents are, respectively, fuzzy and crisp sets, respectively. Hence, the procedures involved in the computation of the output signals are distinct. While in the case of Sugeno fuzzy structure the output is computed with a very simple formula (weighted average, weighted sum). Mamdani fuzzy structure requires higher computational effort because one is required to compute a whole membership function which is then defuzzyfied. This advantage to the Sugeno approach makes it highly useful in spite of the more intuitive nature of Mamdani fuzzy reasoning in terms of dealing with uncertainty (Schnitma, et al, 2000).

3.5.1 Mamdani's Fuzzy Inference System:

Mamdani's fuzzy inference method is the most commonly seen fuzzy methodology. Mamdani's method was among the first control systems built using fuzzy set theory. It was proposed as an attempt to control a steam engine and boiler combination by synthesizing a set of linguistic control rules obtained from experienced human operators .Mamdani's effort was based on Lotfi Zadeh's 1973 paper on fuzzy algorithms for complex systems and decision processes (Sharma, 2007).

Mamdani-type inference, as defined for Fuzzy Logic Toolbox, expects the output membership functions to be fuzzy sets. After the aggregation process, there is a fuzzy set for each output variable that needs defuzzification. It is possible, and in many cases much more efficient, to use a single spike as the output membership functions rather than a distributed fuzzy set. This type of output is sometimes known as a *singleton* output membership function, and it can be thought of as a pre-defuzzified fuzzy set (Matlab R 14, 2008). The general formula for the rule in Mamdani type is:

$$\text{If } x \text{ is } A \text{ and } y \text{ is } B \text{ then } z \text{ is } C \qquad (3.1)$$

Where *A*, *B*, *C* are fuzzy sets of the universes of discourse *X*, *Y*, and *Z*, respectively and *x* and *y* are values of the input variables (Jang 1995).

The formula for rules in our developing model has the form:

$$IF\ (KW_i \text{ is } MF_j) \text{ and } (KW_{i+1} \text{ is } MF_j) \text{ and } \ldots \text{ and } (KW_m \text{ is } MF_j)\ THEN\ (Mark \text{ is } MF_k) \qquad (3.2)$$

Where i = 1 to m represent the i^{th} keyword.

m = 15, number of keywords or synonyms (15 input to the system).

MF_j is the J^{th} membership function where j=1 to 7(the number of MFs for each input); and k= 1 to 16 represent the k^{th} output membership function for the predicted mark, and KW is the abbreviation for keyword or one of its synonyms. Fig.(3.3) shows the basic structure of output membership functions .

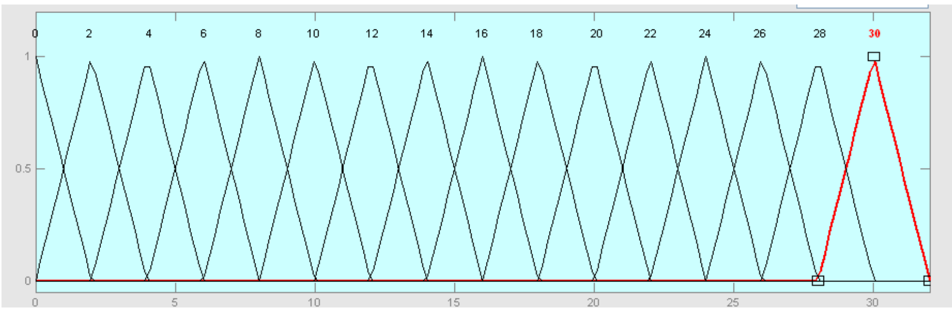


Fig. 3.3: Output membership functions

3.5.2 Sugeno Fuzzy Inference System:

It is similar to the Mamdani method in many respects. The Sugeno output is a crisp function that may be either linear or constant. The general form is characterized with functional type conclusions. A first-order Sugeno MISO model has the form (Jang 1995):

$$\text{If x is } A \text{ and y is } B \text{ then } z = f(x, y) \qquad (3.3)$$

The rule in our developed Sugeno fuzzy model has the form:

R_i : IF (KW_j is MF_{i1}) and … and (KW_{j+1} is MF_{im}) and (KW_n is MF_{im}) THEN $Y_i = a_{i1}KW_1 + \ldots + a_{im}KW_m + a_{i0}$ (3.4)

Where R_i (i =1, 2,…, c) denotes the i^{th} fuzzy rule, are the input (antecedent) variables, Y_i are the rule output variables, and a_{i1}, …, a_{im}, a_{i0} are the model consequent parameters that have to be identified in a given data.j= 1 to n represent the j^{th} keywords, and n=15 represents the number of keywords or synonyms(Matlab R14, 2008).

3.6 Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANN, like people, learns by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons (Stergiou and Siganos, 1996).

There has recently been intense, and fast-growing, interest in 'artificial neural networks'. These are machines (or models of computation) based loosely on the ways in which the brain is believed to work. Neurobiologists are interested in using these machines as a means of modeling biological brains, but much of the impetus comes from their applications. For example, engineers wish to create machines that can perform 'cognitive' tasks, such as speech recognition, and economists are interested in financial time series prediction using such machines. Inevitably, there is a certain amount of hype associated with the subject, particularly in relation to neurobiological modeling (Martin, 1999)

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. This expert can then be used to provide projections given new situations of interest and answer "what if" questions.

Other advantages include:

1. Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.
2. Self-Organization: An ANN can create its own organization or representation of the information it receives during learning time.
3. Real Time Operation: ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.
4. Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to the corresponding degradation of performance. However, some network

capabilities may be retained even with major network damage (http://www.doc.ic.ac.uk, 2008).

Neural networks learn by example, thus, they cannot be programmed to perform a specific task. The examples must be selected carefully otherwise useful time is wasted or even worse the network might be functioning incorrectly. The disadvantage is that because the network finds out how to solve the problem by itself, its operation can be unpredictable (Stergiou and Siganos, 1996). Fig.(3.4) shows the basic NN architecture.

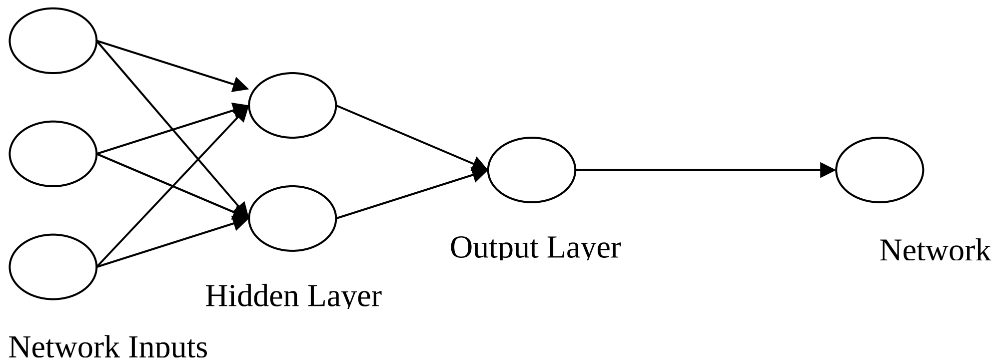


Fig. 3.4: Basic Neural Network Architecture (Abdi, 2003)

3.7 Artificial Neural network Vs Fuzzy Inference System

Table 3.1 provides the basic differences between ANN and FIS (Abraham, 2004).

Table 3.1: Comparison between Neural Networks and Fuzzy Inference Systems

Artificial Neural Network (ANN)	Fuzzy Inference System (FIS)
Difficult to use prior rule-based knowledge	Prior rule-base can be incorporated
Learning from scratch	Cannot learn (use linguistic knowledge)
Black box	Interpretable (if-then rules)
Complicated learning algorithms	Simple interpretation and implementation
Difficult to extract knowledge	Knowledge must be available

3.8 Adaptive Neuro-Fuzzy Inference System (ANFIS)

The acronym ANFIS derives its name from adaptive Neuro-Fuzzy inference system. Using a given input/output data set, the toolbox function anfis in matlab constructs a fuzzy inference system whose membership function parameters are tuned (adjusted) using either a back-propagation algorithm alone or in combination with a least squares type of method. This adjustment allows your fuzzy systems to learn from the data they are modeling

(Matlab R14, 2008). ANFIS is a multilayer feed forward network which searches for fuzzy decision rules that perform well on any given task. The fuzzy decision rules are implemented as MFs and the model learns the best fitting parameters of the MFs. The architecture of ANFIS is shown in fig. (3.5).

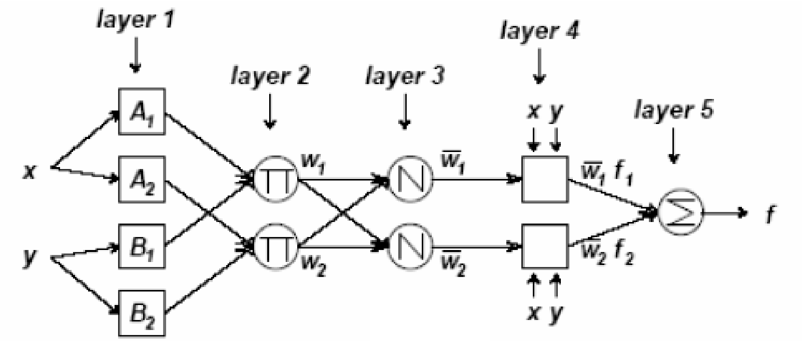


Fig. 3.5: The general architecture of ANFIS (Jang 1993)

Even thought ANFIS is a five layer neural network, only two of these layers have adjustable weights (here represented by squares). The first layer is composed of *n* MFs, each implementing a fuzzy decision rule. Any type of distributions can be modeled by MFs and the set of parameters to minimize is determined accordingly. The second layer computes every possible conjunction of the *n* decisions rules. The third layer normalizes the conjunctives MFs in order to rescale the inputs. The fourth layer is a standard perceptron and associates every normalized MF with an output (weights are called consequent parameters). Finally, the fifth layer sums the evidences. The output is a real number. The consequent parameters and the MF's parameters are learned by standard back propagation (Hélie, et al, 2003).

3.9 Data Clustering

Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). Clustering has emerged as a popular technique in many areas. Clustering of numerical data forms the basis of many classification and system modeling algorithms. The purpose of clustering is to identify natural groupings of data from a large data set to produce a concise representation of a system's behavior. Fuzzy Logic Toolbox is equipped with some tools that allow you to find clusters in input-output training data. You can use the cluster information to generate a Sugeno-type fuzzy inference system that best models the data behavior using a minimum number of rules. The rules partition themselves according to the fuzzy qualities associated with each of the data clusters (http://www.mathworks.com, 2008). The most challenging step in clustering is feature extraction or pattern representation. Pattern recognition researchers conveniently avoid this step by assuming that the pattern representations are available as input to the clustering algorithm. In small size data sets, pattern representations can be obtained based on previous experience of the user with the problem (Jain, 1999). One important clustering is subtractive clustering*;* Subtractive clustering is technique for automatically generating fuzzy inference systems by detecting clusters in input-output training data. The idea of fuzzy clustering is to divide the data space into fuzzy clusters, each representing one specific part of the system behavior. *Subtractive clustering* is a fast, one-pass algorithm for estimating the number of clusters and the cluster centers in a set of data.

3.10 Natural Language Processing (NLP)

Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications (Liddy, 2003). Many Natural Language Processing (NLP) techniques have been used in Information Retrieval. In theory; NLP is a very attractive method of human-computer interaction.Natural Language Processing (NLP) is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things. NLP researchers aim to gather knowledge on how human beings understand and use language so that appropriate tools and techniques can be developed to make computer systems understand and manipulate natural languages to perform the desired tasks (Chowdhury, 2002). Currently, most practical applications in NLP have been realized via symbolic manipulation engines, such as grammar parsers.

Artificial intelligence (AI) is defined as the science of making intelligent machines. AI has several applications including game playing, speech recognition, understanding natural language processing, computer vision, and so on. NLP is considered to be one of the most challenging areas of AI (Dikli, 2006).

At the core of any NLP task there is the important issue of natural language understanding. The process of building computer programs that understand natural language involves three major problems: the first one relates to the thought process, the second one to the representation and meaning of the linguistic input, and the third one to the world knowledge. Thus, an NLP system may begin at the word level to determine the morphological structure, nature (such as part-of-speech, meaning) etc. of the word and then may move on to the sentence level to determine the word order, grammar, meaning of the entire sentence, etc. and then to the context and the overall environment or domain. A given word or a sentence may have a specific meaning or connotation in a given context or domain, and may be related to many other words and/or sentences in the given context (Chowdhury, 2002).

NLP is claimed to be a complex task to comprehend since it contains several levels of processing as well as subtasks. It has four categories of language tasks including speech recognition, syntactic analysis, discourse analysis and information extraction, and machine translation. Speech recognition focuses on diagramming a continuous speech signal into a sequence of known words. Syntactic analysis, on the other hand, determines the ways the words are clustered into constituents like noun and verb phrases. Semantic analysis employs diagramming a sentence to a type of meaning representation such as a logical expression. While, discourse analysis focuses on how context impacts sentence interpretation, information extraction locates specific pieces of data from a natural language document. Finally, the task of machine translation is to translate text from one natural language to another, i.e., English to German or vice versa (Dikli, 2006).

Chapter Four

Datasets used and the Developed Models

4.1 Introduction

The most important factor that must be considered when developing AES system is the availability of data set for training and testing the developed models. This chapter has two main parts; the first part describes the data set used in building the models and the second part explains and clarifies the steps and the procedures done to develop the models. In our models, we scored the essay in two different approaches. In the first method we scored the students' answers by comparing it with reference answers (main keywords and synonyms) and this is the normal scoring. In the next method we scored the students answer by using fuzzy approaches (developed models; will be explained later on this chapter).

An essay is usually a short piece of writing and composed of one paragraph or more. In our models we adopted the short essay with 100-word length. The main idea from our approach is to score the essay by identifying the main keywords or synonyms in the short essay (100-word length). The developed models are capable to identify up to 15 main key, each of which has up to 4 synonyms which identified by the teacher. The developed models are capable to read the student answer (an essay) and determining the keywords in the text. In the next stage the weighting for each keyword (Input to the fuzzy system) are identified. The weighting of keywords is processed by the developed fuzzy models. The output from the developed fuzzy models is the predicted marks (the details procedures in the coming subsections).

4.2 Impacts of Words order and negation in the sentence on Scoring Method in Content-based AES systems

We have succeeded in our AES models to address the problem of word's negations in the sentences by adopting a technique based on negative weights and by using special membership function for that purpose. Section four of this chapter (the developed model) will elaborate more on this issue. Regarding the order of the words in the sentence, we have suggested a model to build a language parser as it will be discussed in the following subsections.

4.2.1 Impact of words' order in the sentence:

We know that AES content-based system produces measures of word-word, word-passage and passage-passage relations that are well correlated with several human cognitive phenomena involving association or semantic similarity. The correlations demonstrate close resemblance between what AES content-based system extracts and the way peoples' representations of meaning reflect what they have read and heard, as well as the way human representation of meaning is reflected in the word choice of writers. AES allows us to closely approximate human judgments of meaning similarity between words and to objectively predict the consequences of overall word-based similarity between passages, estimates of which often figure prominently in research on discourse processing (Landauer, 1998). One critical objection that is raised against the AES content-based system is that not only does it ignore the syntactic structure of sentences, it even ignores word order. In other words, content based treats a text as a bag of words (Wiemer-Hastings, 2004).

The great evidence that deal with information about the meaning of passages may be carried by words independently of their order. If you asked the students to write short essays that would demonstrate their knowledge of scientific topics. The amount and correctness of topic relevant knowledge displayed in the essays was determined either by judgments of two expert human readers or by measures derived from the text by content dimension. The principal findings were content-based measures which take no account of word order were as closely related to human judgments as the latter were to each other. These results and analyses demonstrate that most of the meaning derived by people in reading the texts was also extracted by the content method without recourse to syntax (Thomas K, et al, 1997). Even the best current methods in automatic essay scoring use little or no syntactic information in representing documents, relying primarily on “bag-of-words” methods.

From our literature review to previous AES content-based systems, we recommend to use language parser that will help to identify the structure of the sentences in proper way. Parsing is the process by which phrases in a string of characters in a language are associated with the components names of the grammar that generated the string , a parse is often depicted as a tree(Burstein, et al,2000). Although, the proposed parser requires implementation and testing, we may propose a general architecture of the suggested parser as shown in fig. (4.1).

Fig.(4.1) has the following major elements:

- Natural Language Processing: to convert a plain text file containing the essays to be scored into the parse tree format.
- Morphology stripping: to strip morphology from the input parse tree. This stage open and read the input file (parse form) and automatically strips the morphology from all the words and the output will be morphology stripped words. One of commercially available program used is WordNet (http://wordnet.princeton.edu, 2009)
- Concept extraction: to extract the concepts from the morphology-stripped parse tree. It reads the data from morphology stripping stage and use special symbols to separate the sentences and it is also searches the lexicon file to search for required data(for example, matches between words).
- Rule matching /scoring: to use rule matching for extracted concepts. It is used for searching for any matches between the rules and to score the essays based on matches (using rule file).
- Lexicon file: this file is created by hand (i.e. linguistics people) and include predefined data (plurality of metonyms for selected words). The metonyms being predetermined in accordance with particular domain of the essay being scored.

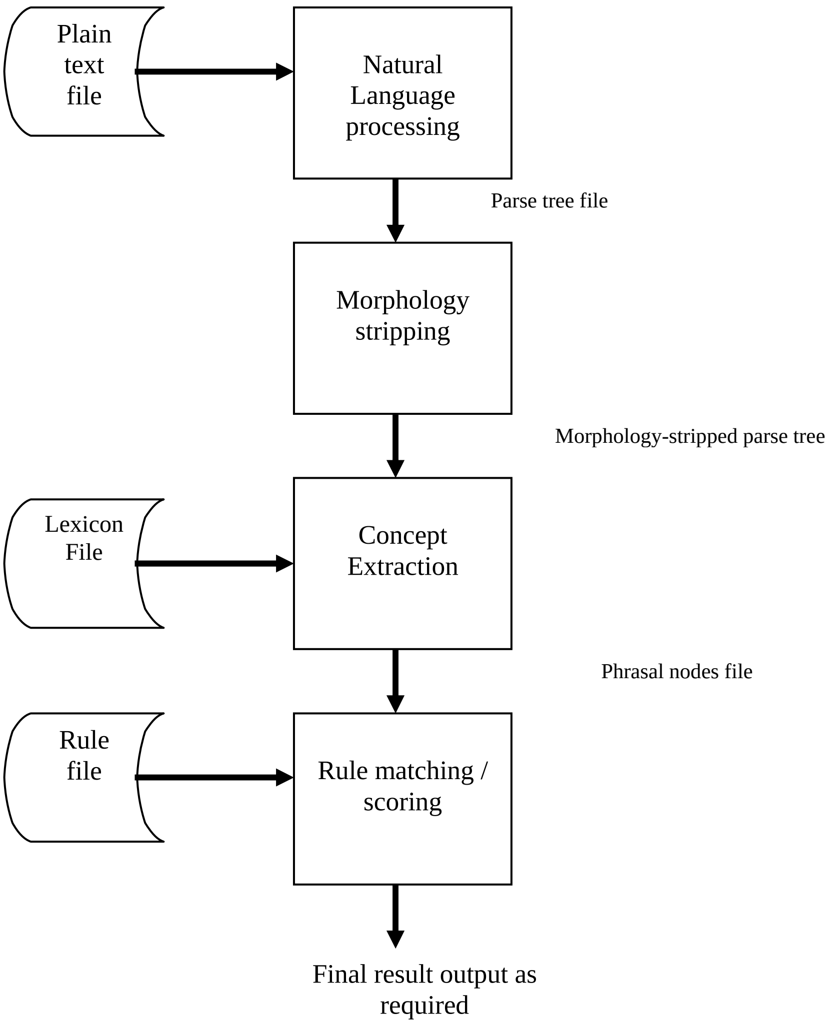


Fig. 4.1: Proposed Parser Architecture

4.2.2 Negation's issue in the sentence:

Another notable gap in AES content-based system is negations. Although no satisfactory approach yet exists for dealing with negations, a possibility would be to treat them as an essentially syntactic component that can process (Wiemer-Hastings, 2004). For example, AES systems based on LSA technique (i.e. IEA) ignore the syntactic structure of the sentences, and it is considered as a bag of word approach, so there are a lot of misses in IEA include: attachment, modification, predication, anaphora and negation. LSA ignores negations, either because they are omitted from the LSA training via a "stop words" list, or simply because their widespread use throughout a corpus renders them representation ally depleted (Sukkarieh, 2004).

To solve the problem of words negation issue in the sentences, we have applied the following criteria: when the user of the system import the student answer in text format, the system will search for the given keyword identified by the teacher, if these main keywords or one of the synonyms encounter before the main keywords, the negative weight will be given to this keywords. In the first stage we just adopted two negation (Not and nor).So any keywords or synonyms with negations (not, nor) before it, negative weights will be assigned. In the normal scoring method, this will reduce the summation process for the whole marks. Later on fuzzy scoring method, the negative input was processed by suitable membership function to illustrate that there is a negation in the student answers. Fig.(4.2) shows the membership function for negative weight. The membership function that processes the negation issue is the general bell membership function. The data set used (answers to questions) for training and testing the developed models has negation to the meaning in the sentences, and we have succeeded to score these answers by adopting the negation criteria mentioned above and the result are promising as shown in the results obtained from training and testing the developed models.

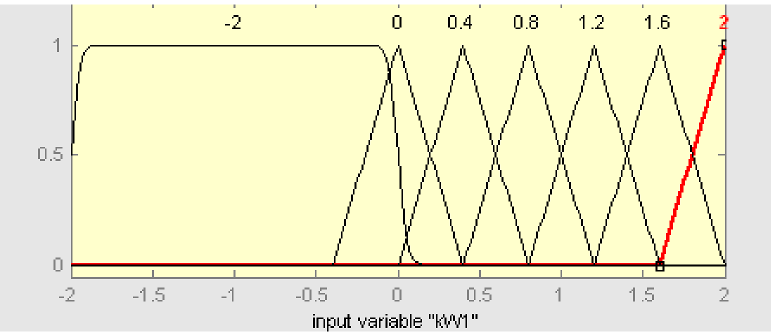


Fig. 4.2: Membership function for negative weight

4.3 Data Set Used

Any system needs input data to produce the desired output. Through out our deep literature review to the available AES system we have concluded that in order to develop effective AES system; we need sufficient size of data to train and test the system. The size of data depends on different factors such as the techniques (LSA, NLP, etc.) used in developing the systems and the performance required (if you increase the training data, the performance of the system will be better). For example the developer of E-rater (based on NLP technique) used 465 essays to train the system. Whereas; the developer of BETSY (Based on Bayesian classification technique) used 1000 essay to train the model. In our model we used 1080 data set to train and test the developed model and the result are promising.

One of the basic difficulties in developing effective AES is how to get sufficient data to train the models; none of the educational centers or institutions provide freely or voluntarily data to other researchers in order to use it for developing and testing. However, we have established cooperation with Palestine Polytechnic University (PPU) and Al-Quds Open University (QOU) to support us with data so we can examine it on our developed models. From QOU we obtained a sample of answers (18 samples of answers) for one question in subjects other than history, namely English literature. From PPU we obtained two samples of answers for two questions (11 sample of answers for the first question and 23 of answers for the second question) related to Information Technology subject.

4.3.1 Data sets Collection, Generation and Classifications

The availability of data is very critical aspect in developing AES systems. We used resources available on web (Example: www.answers.com) to obtain answers for the questions. Moreover we succeed to establish a good relation with other local universities to obtain data for testing the models. We have used 13 different history-related questions as detailed in Appendix c. A sample questions and their typical answers are illustrated below:

Q1) Define the United Nation (UN)?

Typical reference answer:

The United Nations was established to preserve peace through international cooperation and collective security. The United Nations is central to global efforts to solve problems that challenge humanity. Cooperating in this effort are more than 30 affiliated organizations, known together as the UN system. The UN and its family of organizations work to promote respect for human rights, protect the environment, fight disease and reduce poverty. Throughout the world, the UN and its agencies assist refugees, help expand food production and lead the fight against AIDS. UN headquarters in New York City. (See appendix c for more answers)

A1.1) answer 1:

The United Nations (U.N.) is an organization of 185 states that strives to attain international peace and security, promotes fundamental human rights and equal rights for men and women, and encourages social progress. By October 24, 1945, China, France, the United States, the Soviet Union, the United Kingdom, and a majority of the charter's other signatories had ratified it, and the United Nations was officially established. Shortly thereafter the U.S. Congress unanimously invited the United Nations to set up headquarters in the United States, and the organization chose New York City as its permanent home.

A1.2) Answer 2:
United Nations: The United Nations officially came into existence on 24 October 1945. The purposes of the United Nations, as set forth in the Charter, are to maintain international peace and security; to develop friendly relations among nations; to cooperate in solving international economic, social, cultural and humanitarian problems and in promoting respect for human rights and fundamental freedoms; and to be a centre for harmonizing the actions of nations in attaining these ends.
(See appendix c for more answers)

Q2) Define: Arab League

A2.1) Answer1: Regional organization formed in 1945 and based in Cairo. It initially comprised Egypt, Syria, Lebanon, Iraq, Transjordan (now Jordan), Saudi Arabia, and Yemen; joining later were Libya, Sudan, Tunisia, Morocco, Kuwait, Algeria, Bahrain, Oman, Qatar, the United Arab Emirates, Mauritania, Somalia, the Palestine Liberation Organization, Djibouti, and Comoros. The league's original aims were to strengthen and coordinate political, cultural, economic, and social programs and to mediate disputes; a later aim was to coordinate military defense. Members have often split on political issues; Egypt was suspended for 10 years (1979 – 89) following its peace with Israel.

A2.2) Answer2: popular name for the League of Arab States, formed in 1945 in an attempt to give political expression to the Arab nations. The original charter members were Egypt, Syria, Lebanon, Transjordan (now Jordan), Iraq, Saudi Arabia, and Yemen. A representative of Palestinian Arabs, although he did not sign the charter because he represented no recognized government, was given full status and a vote in the Arab League. The Palestine Liberation Organization (PLO) was granted full membership in 1976. Other current members include Algeria, Bahrain, Comoros, Djibouti, Eritrea (pending in 1999), Kuwait, Libya, Mauritania, Morocco, Oman, Qatar, Somalia, Sudan, Tunisia, and the United Arab Emirates.

A2.3) Answer3: The League of Arab States, also known as the Arab League, is composed of twenty-two independent Arab states that have signed the Pact of the League of Arab States. Palestine, represented by the Palestinian Authority, is included as an independent state. The multipurpose League of Arab States seeks to promote Arab interests in general, but especially economic and security interests. It also works to resolve disputes among members and between member states and nonmember states. It has the image of unity in the protection of Arab independence and sovereignty. It promotes political, military, economic, social, cultural, and developmental cooperation among its members

(See appendix c for more answers)

Q3) Define: Palestine Liberation Organization (PLO):

A3.1)Answer1: Umbrella political organization representing the Palestinian people in their drive for a Palestinian state. It was formed in 1964 to centralize the leadership of various groups. In 1969 Yasir 'Arafat, leader of Fatah, became its chairman. From the late 1960s the PLO engaged in guerrilla attacks on Israel from bases in Jordan, from which it was expelled in 1971. PLO headquarters moved to Lebanon. In 1974 'Arafat advocated limiting PLO activity to direct attacks against Israel, and the Arab community recognized the PLO as the sole legitimate representative of all Palestinians. It was admitted to the Arab League in 1976. In 1993 Israel recognized the PLO by signing an agreement with it granting Palestinian self-rule in parts of the West Bank and Gaza Strip.

A3.2) Answer2: (PLO), coordinating council for Palestinian organizations, founded (1964) by Egypt and the Arab League and initially controlled by Egypt. Composed of various guerrilla groups and political factions, the PLO is dominated by Al Fatah, the largest group, whose leader, Yasir Arafat, was chairman of the PLO from 1969 to 2004 and established Palestinian control over the organization. Other groups in the PLO include the Syrian-backed As Saiqa and the Marxist-oriented Popular Front for the Liberation of Palestine (PFLP).
(See appendix c for more answers and questions)

The total number of generated and collected answers to build the models (1080 answers) as detailed in table (4.1). In the first stage of data collection and generation, we have collected answers from different resources on the web (for example: www.answers.com) and this sample of answers was used later for generation more answers using developed random answers generator. The generator uses the keywords and synonyms to produce suggested answers to train and test the models. As a result, we have two types of data (answers to questions); answers collected from the web and the rest of answers have been generated.

Using cross-validation method, the answers are divided into training (Tr) and testing (Ts). The idea from cross validation method is to choose the first two answers for training purpose and to choose the third one for testing purpose and so on. The training data is about 66 % of whole data generated. We have generated 1080 answers related to 13 questions. In fact, there is no specific rule why we have chosen 13 questions and 1080 sample of answers. It is possible to use more or less number of questions and answers. But from our literature survey we noticed that the size of data is important together with techniques used and the required accuracy. For example (Oriqat, 2007) used 900 sample of data to train the systems and the degree of agreement (correlation) between actual (theoretical) and predicted marks is high. The developer of E-rater (based on NLP technique) used 465 essays to train the system. Whereas; the developer of BETSY (Based on Bayesian classification technique) used 1000 essay to train the model. In our study, we suggested to use 13 questions related to historical subject. For each question we determined the total number of answers and as a result we have 1080 answers for the 13 question as shown in table (4.1). As a final conclusion, there are no clear reasons or rules to determine the number of sample required for training AES system. The number of samples depends on the technique used and the required accuracy.

The total answers for all 13 questions were 1080 with 718 of datasets for training and 362 for testing (using cross-validation method mentioned before). The importance of generated data is to train the models.

Table 4.1: Fifteen main keywords dataset collection, generation and classification

Question Number	Number of answers	Training data (Tr)	Testing data (Ts)
Q1	100	67	33
Q2	70	46	24
Q3	60	40	20
Q4	80	53	27
Q5	110	73	37
Q6	180	120	60
Q7	100	67	33
Q8	50	33	17
Q9	50	33	17
Q10	60	40	20
Q11	80	53	27
Q12	50	33	17
Q13	90	60	30
Total	**1080**	**718**	**362**

4.3.2 PPU Dataset

We have been in contact with Mr. Mohammad Jabary, one of the academic staff at PPU, and he supported us with student answers to the two questions related to information technology. The first question asked the student to define the "Programming", while the second question asked the student to define the "Operating System OS". For the first question we obtained 11 answers and for the second question we obtained 23. PPU use MOODLE as an e-Learning environment and using this technology, the following questions and answers illustrate the data obtained from student using MOODLE:

Q1: *What is the programming?*

Typical answer: Computer programming (often shortened to programming or coding) is the process of writing, testing, debugging/troubleshooting, and maintaining the source code of computer programs. Computer program is a set of instructions. The source code syntax is written in a programming language. The code may be a modification of an existing source or something completely new. The purpose of programming is to create a program that exhibits a certain desired task (customization). The process of writing source code requires expertise (programmers) in many different subjects, including knowledge of the application domain, specialized algorithms and software.

Std#1: The process of writing, testing, debugging, and maintaining the source code of computer programs. This source code is written in a programming language .The code may be a modification of an existing source or something completely new. The purpose of programming is to create a program that exhibits a certain desired behavior (customization). (A set of instructions telling the computer what to do also referred as software).

Std#2: computer program is a set of instructions for a computer to perform a specific task. Programs generally fall into these categories applications, utilities or services Programming is planning how to solve a problem. No matter what method is used pencil and paper, slide rule, adding machine, or computer -problem solving requires programming. Of course, how one program depends on the device one uses in problem solving?"

Std#3: Computer Programming is the art of making a computer do what you want it to do. Programming is a creative process done by programmers to instruct a computer on how to do a task. Hollywood has helped instill an image of programmers as super thief who can sit down at a computer and break any password in seconds or make highly tuned warp engines improve performance by 500% with just one tweak. Sadly the reality is far less interesting!

(Note: see appendix C for complete students' answers and data set)

Q2: *Define the Operating System*?

Typical answer: Is the software component of a computer system that is responsible for the management and coordination of activities and the sharing of the resources of the computer. The operating system acts as a host for application programs that are run on the machine. As a host, one of the purposes of an operating system is to handle the details of the operation of the hardware(input /output) .This relieves application programs from having to manage these details and makes it easier to write instructions for applications. Operating system acts as an interface window for communication.

Std#1: An Operating System is a computer program that manages the resources of a computer. It accepts keyboard or mouse inputs from users and displays the results of the actions and allows the user to run applications, or communicate with other computers via networked connections

Std#2: Software designed to handle basic elements of computer operation, such as sending instructions to hardware devices like disk drives and computer screens, and allocating system resources such as memory to different software applications being run. Given uniformly designed operating systems that run on many different computers, developers of software do not need to concern themselves with these problems, and are provided with a standard platform for new programs.

Std#3: is the software component of a computer system that is responsible for the management and coordination of activities and the sharing of the resources of the computer. The operating system acts as a host for application programs that are run on the machine. As a host, one of the purposes of an operating system is to handle the details of the operation of the hardware.

(Note: see appendix C for complete students answers and data set).

4.3.3 QOU Dataset

We have been in contact with Dr. Nadia Qawasmi, she is an English supervisor in QOU, and she supported us with data related to English subject. Dr. Nadia asked the students to define the concept of "Allegory", and for this question we have 18 sample of answers(see appendix C).The students in QOU use academic portal to stay in touch with academic staff in addition to MOODLE. Data from both PPU and QOU are evaluated manually from instructors themselves. The results of these data will be discussed in the next chapter. The following question and answers illustrate the data obtained from QOU:

Q: What do we mean by Allegory?

Typical Answer:
In literature, symbolic story that serves as a disguised representation for meanings other than those indicated on the surface. The characters in an allegory often have no individual personality, but are embodiments of moral qualities and other abstractions. The allegory is closely related to the parable, fable, and metaphor, differing from them largely in intricacy and length. The medieval morality play *Everyman,* personifying such abstractions as Fellowship and Good Deeds, recounts the death journey of Everyman. John Bunyan's *Pilgrim's Progress,* a prose narrative, is an allegory of people spiritual salvation. Although allegory is still used by some authors, its popularity as a literary form has declined in favor of a more personal form of symbolic expression.

Std#1: Allegory is a symbolic narrative in which the surface details imply a secondary meaning. Allegory often takes the form of a story in which the characters represent moral qualities. The most famous example in English is John Bunyan's *Pilgrim's Progress*, in which the name of the central character, Pilgrim, epitomizes the book's allegorical nature. Kay Boyle's story "Astronomer's Wife" and Christina Rossetti's poem "Up-Hill" both contain allegorical elements.

Std #2: allegory: "A story or visual image with a second distinct meaning partially hidden behind its literal or visible meaning. In written narrative, allegory involves a continuous parallel between two (or more) levels of meaning in a story, so that its persons and events correspond to their equivalents in a system of ideas or a chain of events external to the tale."

Std #3: Allegory: the saying of one thing and meaning another. Sometimes this trope works by an extended metaphor ('the ship of state foundered on the rocks of inflation, only to be salvaged by the tugs of monetarist policy'). More usually it is used of a story or fable that has a clear secondary meaning beneath its literal sense. Orwell's *Animal Farm*, for example, is assumed to have an allegorical sense.

Std #4: Work of written, oral, or visual expression that uses symbolic figures, objects, and actions to convey truths or generalizations about human conduct or experience. It encompasses such forms as the fable and parable. Characters often personify abstract concepts or types, and the action of the narrative usually stands for something not explicitly stated. Symbolic allegories, in which characters may also have an identity apart from the message they convey, have frequently been used to represent political and historical situations and have long been popular as vehicles for satire.
(Note: see appendix C for complete answers and data set).

4.4 FAESS Design

In the development stages of our models, we began by preparing question and write its reference (typical) answer. Data collection is important factor in AES design; we need to prepare reasonable data (if you increase the training data, the performance and accuracy will increase) to achieve the required results. Fig. (4.3) shows the main stages for dataset collection and classification in our model.

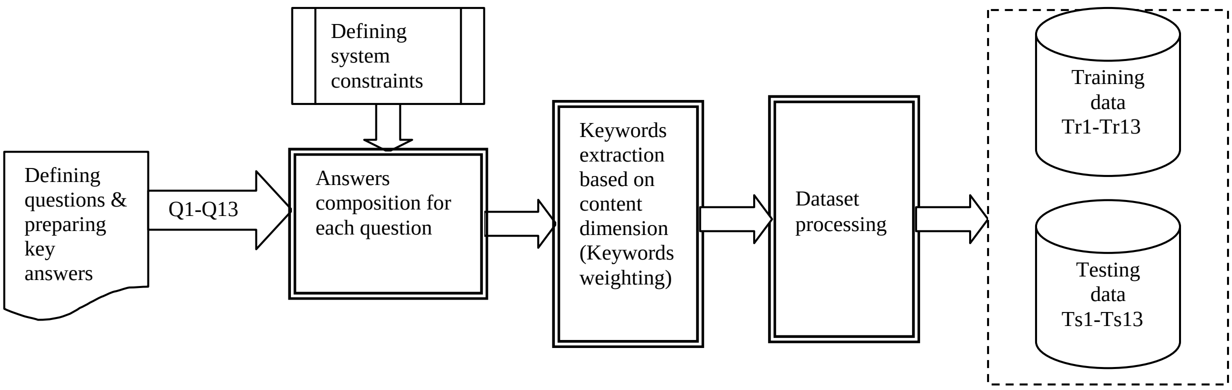


Fig. 4.3: Stages required for dataset collection and classification

Fig. (4.3) could be briefly explained as follow:

- Defining question and key answers:
  - Determine the type of questions(Historical question)
  - Identify the key answer(reference answer)

- Determine the system constraints:
  - An answer with mathematical equation is not accepted.
  - Scoring is depending on content dimension, i.e. keywords will be extracted irrespective of their order in the student answer.

- Answers composition :

- In the first stage, we collected the answers (about 80 answers) from the World Wide Web; for example: www.answers.com (www.answers.com, 2008).
- In the second stage, we generated the data using automatic answer generator (about 1000 answers).

- Keyword extraction based on content dimension:
  - The teacher will determine the keywords and synonyms of the reference answer.
  - The teacher will determine the weight for each keyword and synonym.

- Dataset processing:
  - Use cross-validation method to divide the dataset for training and testing.
  - About 66% of data will be used for training purpose, and the rest data will be used for testing purpose(the first two sets of data will be selected as the training sets, and the third set will be used as the testing set, and so on).

Fig.(4.4) shows how the systems deal with stored training and testing weighted data.

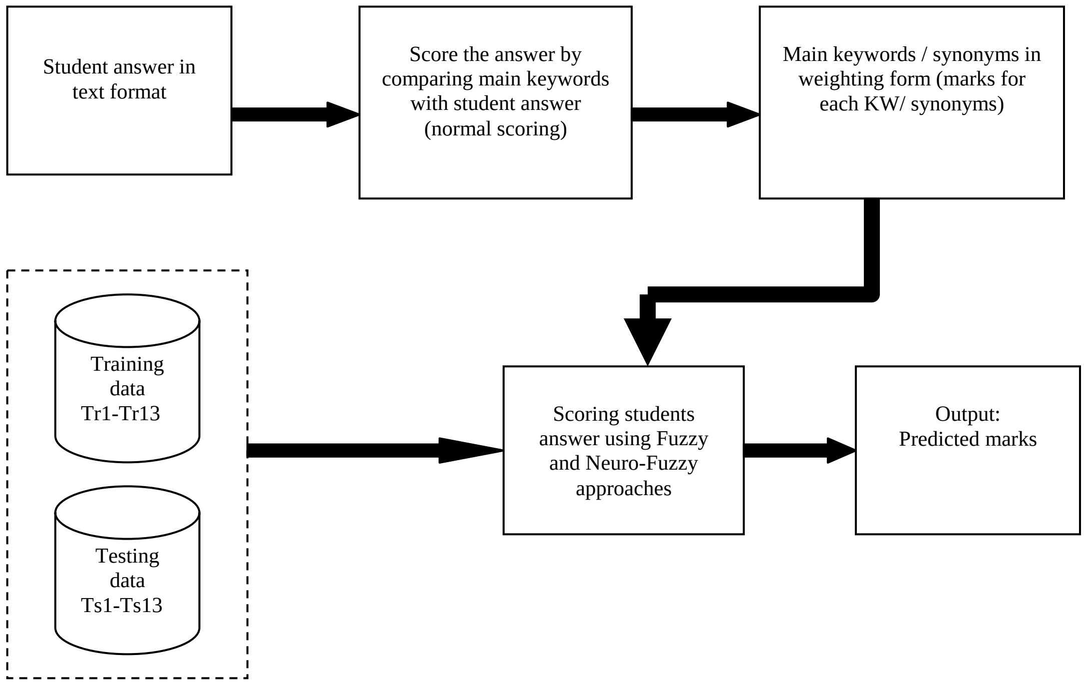


Fig. 4.4: processing of weighted stored data and scoring methods

In fig.(4.4) above, the three developed models have the ability to read the student answers (text format) and provide the feedback(mark) using two methods: normal scoring and Fuzzy method. The developed models also having the ability to train and test data stored in DAT format (Tr and Ts data). The normal scoring method based on extracting the main keywords and synonyms from the key answers or reference answers described by instructors. When the instructors determine the keywords and synonyms and weights for these keywords, the system will compare these keywords with given student answers. The system will produce the final mark by generating linear summation for the weights of words in student answer. The negation issue is adopted in the above design, and we apply the techniques discussed in section 4.2.2 above. The output from these models is the predicted marks.

4.5 Accuracy measurements used

In order to check the accuracy and the adequacy of the system, three measurers (Arafeh et al, 2008) were used:

(a) The correlation coefficient measure between actual and predicted marks

$$r_{xy=}\sqrt{1-\frac{\sum (y_i-\mathrm{f}(\mathrm{x}_\mathrm{i}))^2}{\sum (y_i-y)^2}} \qquad (4.1)$$

Where y_i is the i^{th} actual data.

y is the average of all actual data.

$f(x_i)$ is the i^{th} predicted data.

(b) The mean absolute percentage Error (MAPE), as calculated in:

$$\text{MAPE}= \sum_{i=1} \left| \frac{\text{Actual}_{\text{I}} - \text{Predicted}_{\text{I}}}{\text{Actual}_{\text{I}}} \right| * \frac{100}{N}\% \qquad (4.2)$$

Where N is the total number of data point under construction.

(c) The Root Mean Square Error(RMSE) is used to evaluate the error between a series of marks founds in the pairs of same–dimension vector arrays of dimension [1 x n]. For two distinct arrays actual Mark (Y) and predicted marks (X) of dimension [1 x n], the general form of the RMSE equation is presented below:

$$\text{RMSE}= \frac{\sqrt{\sum_i (actual_i - predicted_i)^2}}{(N-1)} \qquad (4.3)$$

N is the total number of datasets.

4.6 The Developed Models

In this research we have presented a description and evaluation models for Automated Essay Scoring Systems. The development stage began with defining question and typical answer. The kind of the questions is historical subject. The purpose of the model was to assess whether Automated Essay Scoring using fuzzy and Nuero-fuzzy was feasible, economically viable and as accurate as manually scoring the essays. During the first trial of an Automated Essay Scoring system development, we had 1080 students answer for the 13 questions (for example: Define the United Nation (UN)).

In the following subsections, three models based on fuzzy and Neuro-Fuzzy have been constructed based on data set collected and generated and from suggested answers produced from the models.

4.6.1 Multiple Input Single Output (MISO) Mamdani Fuzzy Inference System (FIS):

The MISO model has fifteen inputs as suggested in design stage, each input represent one main keyword or its synonyms and each input have number of membership functions were each function correspond to a weighting value from an answer document that are suitable to the input(the weighting process will be discussed later).
The general architecture for the MISO Mamdani model is shown in fig. (4.5).

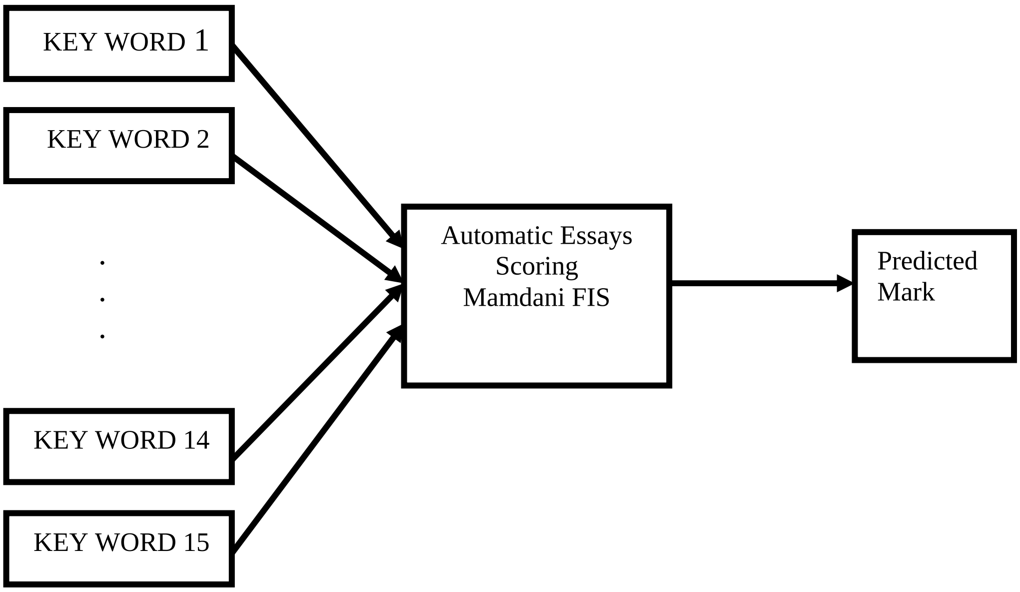


Fig 4.5: Multiple input single output Mamdani model.

The output from the model is the predicted marks which processed according to table (4.2) below; then these linguistics terms defuzzified to get numerical predicted marks (0 for the lowest mark and 30 for the highest mark).

Table 4.2: Linguistics Term for outputs

Linguistic term (MISO output)	Fail	Pass	Fair	Good	V. Good	Excellent
marks	0-15	15-18	18-21	21-24	24-27	27-30

In order to train the MISO model, the system parameters need to be processed manually i.e. the rules constructions and the type of membership function's selection (trapezoidal, triangular, general bell, etc). The purpose of tuning the rules and membership functions is to produce the best results .In fig. (4.6), the predicted marks will be measured against the actual marks. To get the best results, the MFs will be altered by choosing the suitable MF.

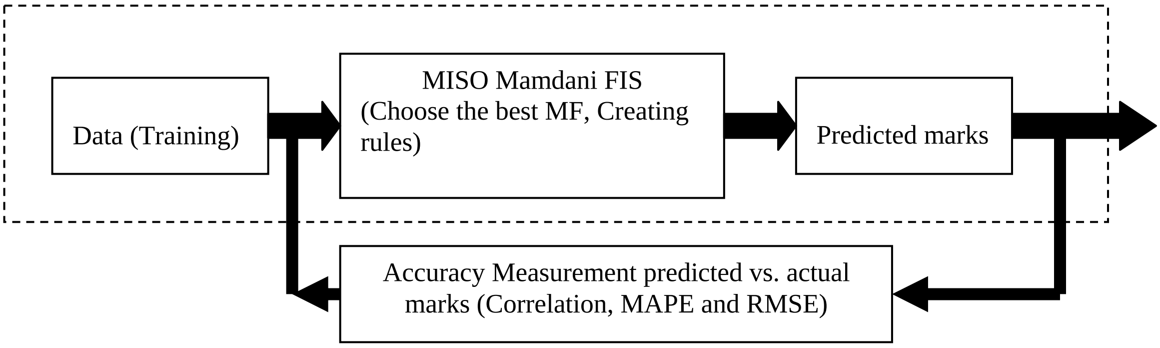
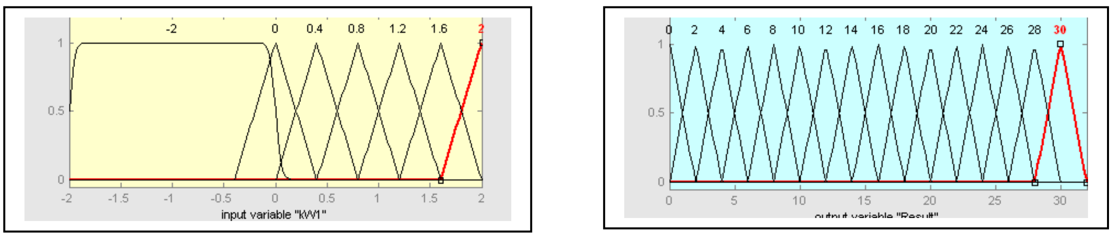


Fig. 4.6 Mamdani FIS training system

In our Mamdani model, we have tried more than one kind of MF including trapezoidal, Triangular, Generalized bell and Gaussian curve. The MF that produces the best result in our model was triangular MF. As a result, we adopted triangular MF since it produced the lowest error and highest correlation agreement between actual and predicted when we used the testing data.

Fig.(4.7) shows the basic input and output membership functions that were used to develop MISO Mamdani model. The triangular membership function was selected because the best result achieved when we use it. The fuzzy sets are learned offline (the weight of each KW) and used with predefined rules constructed according to data.



a. Input (keyword) b. Output (Marks)

Fig.4.7: input and output membership functions

To have a clear idea about fig.(4.7);

For fig. (4.7) part a:

- Represent one input(keywords)
- Each input has seven membership functions (weights are: -2, 0, 0.4, 0.8, 1.2, 1.6, 2).

- Each main keyword has four synonyms (according to our suggestion).
- Each synonym has a value for 0.4 to 1.6 (synonyms suggested weights).
- No match represents 0(No keyword presents at all) and full match represent 2(keyword found).

For fig.(4.7) part b:

- Represent single output (predicted marks).
- Has 16 membership functions(predicted marks distribution)

The general MISO Mamdani model was constructed using the dataset. The dataset was used to train the system. We can summarize the construction of the model in the following steps:

- Building and determining the input and output membership function based on table 4.2 and fig.(4.6).
- Building rule base.
- Checking the adequacy of the models (degree of agreement between actual marks and predicted marks using three measures (Correlation, MAPE and RMSE)).
- If the obtained results are not good, try to alter the membership functions type and shape. Also you need to alter the rules if necessary.
- Checking the model using testing data.

Table (4.3) shows the correlation coefficient between theoretical (actual) and predicted marks for the training and testing data, the values of MAPE and RMSE are also included. The average results for all data set also calculated.

Table 4.3: MISO Mamdani model results

Question No.	Training/Testing answers	Training			Testing		
		Corr.	MAPE	RMSE	Corr.	MAPE	RMSE
1	67/33	0.9902	0.085	0.0704	0.9822	0.0866	0.1414
2	46/24	0.9904	0.0849	0.0846	0.9761	0.1502	0.1719
3	40/20	0.9950	0.1438	0.081	0.9844	0.1619	0.2137
4	53/27	0.9962	0..0932	0.0627	0.9819	0.1184	0.1608
5	73/37	0.9334	0.1093	0.1818	0.9896	0.1067	0.1318
6	120/60	0.9963	0.0735	0.0378	0.9796	0.1556	0.1232
7	67/33	0.9904	0.0163	0.046	0.887	0.0742	0.2688
8	33/17	0.9557	0.0275	0.2315	0.9178	0.0505	0.2341
9	33/17	0.9969	0.1009	0.0765	0.9673	0.1119	0.2595
10	40/20	0.9956	0.1066	0.0747	0.9745	0.1221	0.2947
11	53/27	0.9956	0.0774	0.0584	0.9722	0.1559	0.2141
12	33/17	0.9956	0.1155	0.0796	0.9634	0.1493	0.2296
13	60/30	0.9914	0.017	0.046	0.9027	0.0624	0.2516
Average		0.9863	0.0854	0.0870	0.9599	0.1189	0.2073

Some results obtained for testing data set are better than the results obtained from trained data that is because we have training a general model for all sets. For example, answers for question NO. 5 show correlation (0.9896) value for the testing dataset higher than correlation for training dataset (0.9334).

The results obtained in table (4.3) are promising and the average correlation between predicted and actual mark approximately more than 0.98 which best describe the agreement between actual and predicted marks. The error measured (MAPE and RMSE) indicate the percentage errors in two different ways.

Fig. (4.8) shows a sample of the agreement plot between actual and predicted marks related to one of the questions TR1 (Training data for question 1). The stars in fig.(4.8) represent the actual marks; whereas the circles represent the predicted marks. As we can see from the graph, there is some total match between actual and predicted marks (for example; answers No. 30, while we can see approximately a small agreement for answers No. 60)

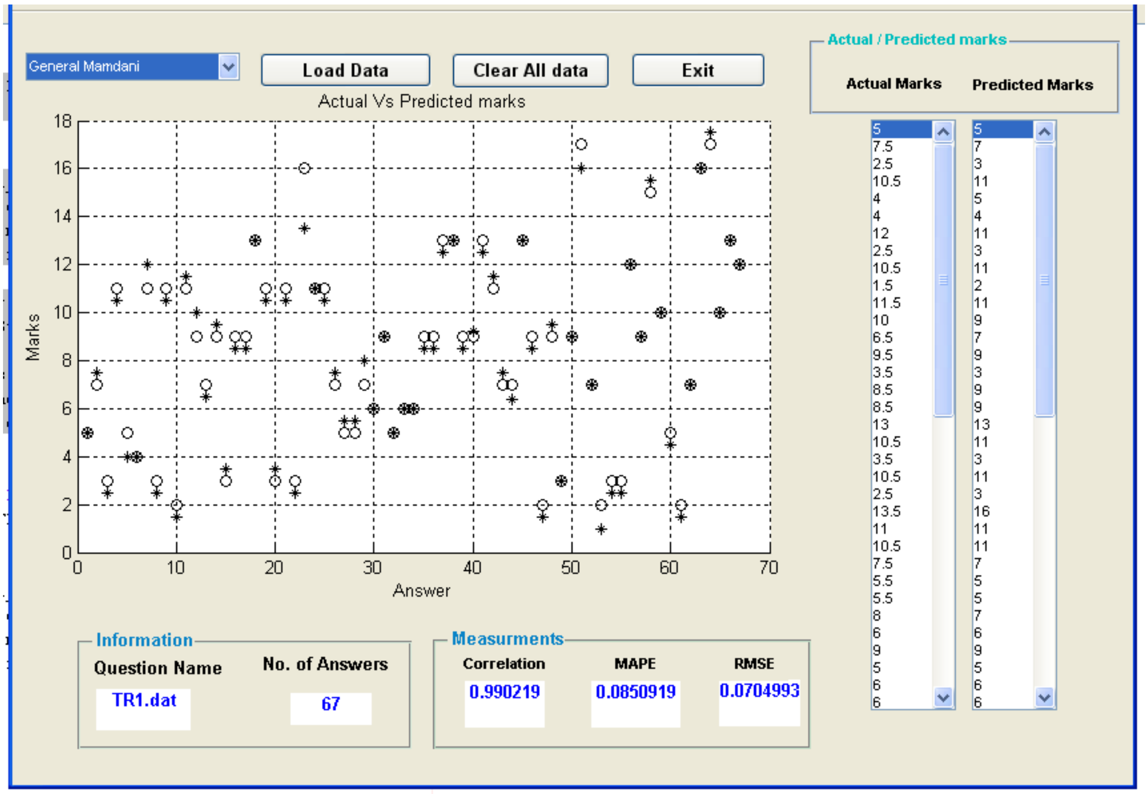


Fig. 4.8: MISO Mamdani Model plot (Actual vs. predicted marks) for TR1 training dataset.

4.6.2 MISO Sugeno-type FIS with Back Propagation optimization technique:

The second model built was Sugeno model with back propagation optimization technique. As we mentioned in chapter three, the difference between Mamdani and Sugeno FIS is the consequent of the fuzzy rules and hence the aggregation and defuzzification procedure accordingly. In Mamdani, the consequent is fuzzy set but in Sugeno it is crisp value. The general architecture for this model is represented in fig.(4.9).

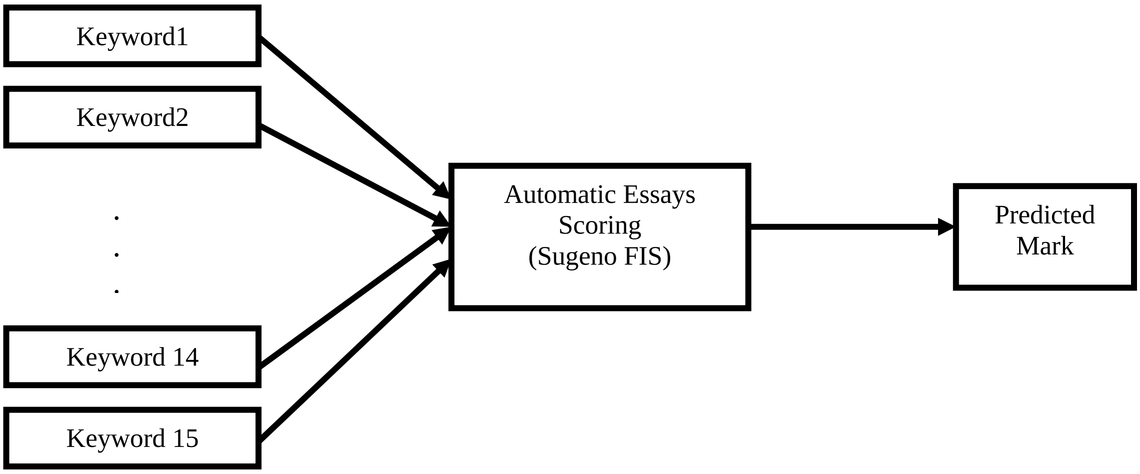


Fig. 4.9: MISO Sugeno-type FIS with Back Propagation model

In building this model, we have used the same procedures that were used in building Mamdani model (pre-process stage); but in Sugeno the model was trained using Adaptive Neuro-Fuzzy Inference System (ANFIS). The main important characteristic in ANFIS is that you can tune the FIS parameter from input and output data.

Using ANFIS techniques, we have built one general model for the dataset in order to predict the marks, so we trained the MISO Sugeno using ANFIS for all available data. Table 4.4 shows the correlation coefficient between theoretical (actual) and predicted marks for the training and testing data, the values of MAPE and RMSE are also included. The average results for all data set also calculated.

Table 4.4: Grid partition Sugeno model results

Question No.	Training/Testing	Training			Testing		
		Corr.	MAPE	RMSE	Corr.	MAPE	RMSE
1	67/33	0.9957	0.0663	0.0538	0.9837	0.1128	0.1307
2	46/24	0.9929	0.0818	0.072	0.9744	0.1552	0.1773
3	40/20	0.9766	0.3376	0.2098	0.9864	0.109	0.1933
4	53/27	0.996	0.074	0.065	0.9856	0.076	0.1465
5	73/37	0.9785	0.1934	0.1108	0.9753	0.3041	0.207
6	120/60	0.9946	0.0735	0.043	0.9805	0.1127	0.1223
7	67/33	0.992	0.015	0.0407	0.9619	0.0577	0.1739
8	33/17	0.9836	0.0205	0.0662	0.9108	0.0584	0.2341
9	33/17	0.9933	0.1312	0.1126	0.9719	0.1381	0.244
10	40/20	0.996	0.116	0.0653	0.9725	0.1239	0.3049
11	53/27	0.9932	0.099	0.0798	0.9686	0.2476	0.2401
12	33/17	0.9966	0.1256	0.0855	0.9347	0.399	0.3775
13	60/30	0.9889	0.0174	0.0478	0.9718	0.0386	0.139
Average		0.9906	0.1039	0.0809	0.9675	0.1071	0.2069

The obtained results in table (4.4) are promising and we can notice that there is nearly total agreement between actual and predicted marks (High correlation). Also the values of errors measures (RMSE and MAPE) are small. The correlation values between predicted and actual marks for training data ranging from 0. 9766 to 0.9966 with 0.9906 average correlation. Fig.(4.10) shows the agreement plot between actual and predicted marks for answers to question No. 5.

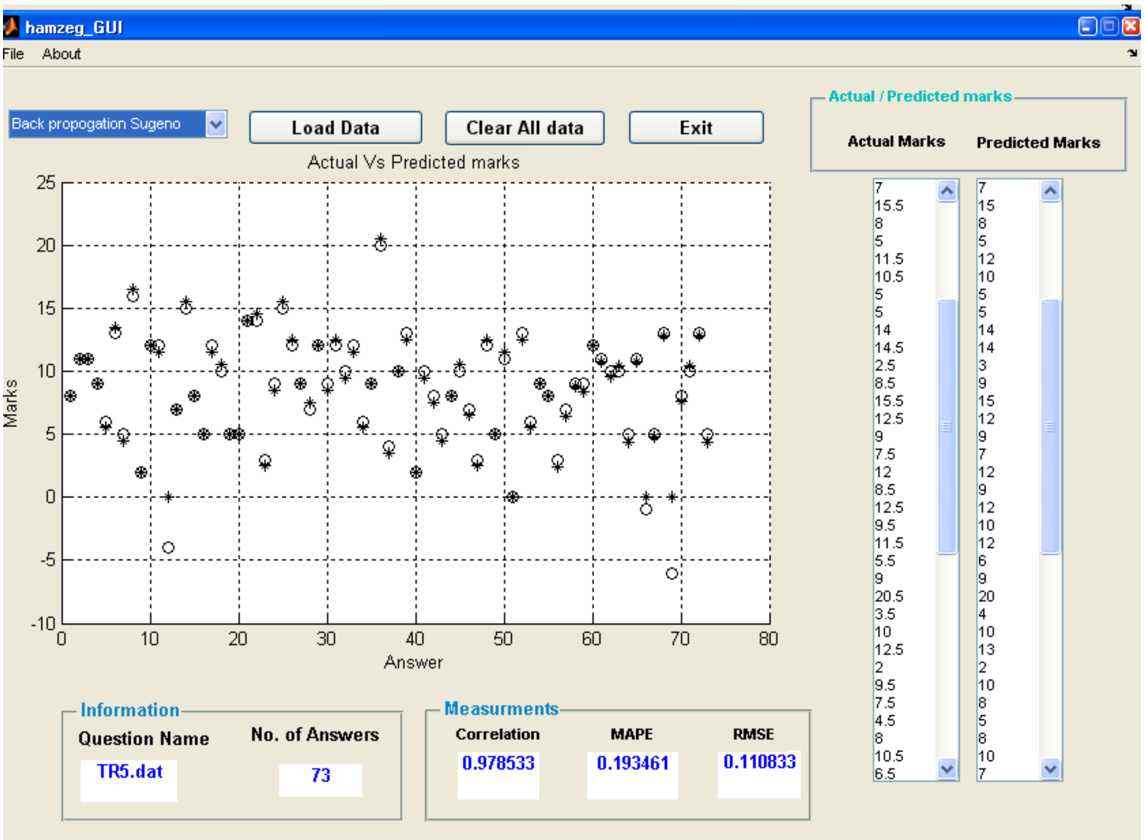


Fig. 4.10: MISO Sugeno Model plot (Actual vs. predicted marks) with back propagation for TR5 training dataset.

Fig. (4.10) shows a sample of the agreement plot between actual and predicted marks related to one of the questions TR5 (Training data for question 5). The stars in fig.(4.10) represent the actual marks; whereas the circles represent the predicted marks. As we can see from the graph, there is some total match between actual and predicted marks (for example; answers No. 40, while we can see approximately a small agreement for answers No. 30).

4.6.3 Sugeno Subtractive Clustering Fuzzy Inference System:

The main advantage of fuzzy clustering is to divide the data space into fuzzy clusters, each representing one specific part of the system behavior. We used the cluster information to generate Sugeno model that best represent the data behavior using the minimum number of rules. The most important benefit of using clustering to find rules is that the resultant rules are more tailored to the input data than they are in a fuzzy inference system generated without clustering.

The following steps are used to build the model:

- Preparing the training data and subjecting it to sub-clustering technique.
- The data must be already divided into two parts one for fifteen keywords input and one actual marks.
- After that you need to check the adequacy of the model by subjecting it to all testing data.

The general block for subtractive clustering model is presented in fig. (4.11).

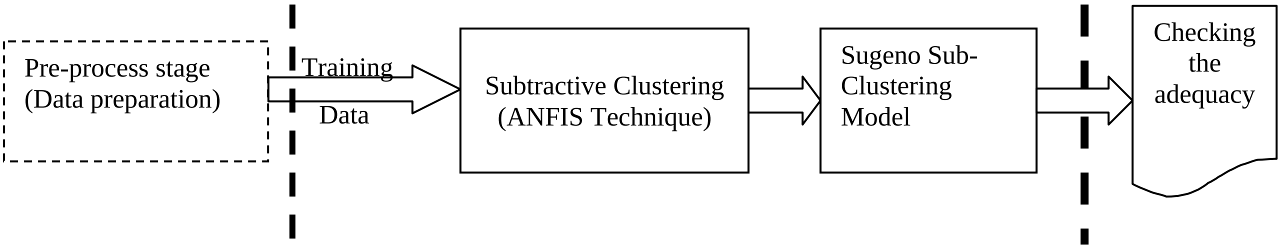


Fig.4.11: Sugeno Subtractive Clustering Model

The clustering model was build using 1080 dataset for training and testing the model. Table (4.5) shows the correlation coefficient between theoretical (actual) and predicted marks for the training and testing data in additions to the values of MAPE and RMSE . The average results for all data set also calculated.

Table 4.5: Sugeno sub-clustering model results

Question No.	Training/Testing	Training			Testing		
		Corr.	MAPE	RMSE	Corr.	MAPE	RMSE
1	67/33	0.9946	0.074	0.0512	0.9824	0.116	0.1379
2	46/24	0.9936	0.0746	0.0684	0.9767	0.0927	0.1686
3	40/20	0.9977	0.0376	0.0573	0.9913	0.0964	0.1622
4	53/27	0.9973	0.0362	0.0524	0.9816	0.1238	0.1608
5	73/37	0.9963	0.0396	0.0432	0.9903	0.1499	0.1289
6	120/60	0.997	0.0318	0.0328	0.981	0.1516	0.128
7	67/33	0.9915	0.015	0.0407	0.9429	0.0675	0.2209
8	33/17	0.9844	0.0205	0.0662	0.9121	0.0584	0.2341
9	33/17	0.9981	0.0403	0.0625	0.9726	0.1013	0.2275
10	40/20	0.997	0.066	0.0543	0.978	0.1123	0.2763
11	53/27	0.9971	0.0424	0.0 552	0.9786	0.1212	0.1923
12	33/17	0.9975	0.0246	0.0584	0.9745	0.0942	0.1926
13	60/30	0.9905	0.0165	0.0447	0.9642	0.0466	0.1589
Average		0.9948	0.0399	0.0526	0.9712	0.1024	0.1837

The obtained results in table 4.5 are promising and showed a high agreement between actual and predicted marks (High correlation). Also the values of errors measures (RMSE and MAPE) are small. The correlation values between predicted and actual marks for training data ranging from 0. 9844 to 0.9977 with 0.9948 average correlation. The RMSE is also small value, the average RMSE is 0.0526 for trained data and 0.1873 for untrained data (testing). The values for the MAPE measure error is also small value which has average value of 0.0399 for trained data and 0.1024 for untrained data (testing).

Fig.(4.12) shows the agreement plot between actual and predicted marks for answers to question No. 6(training).

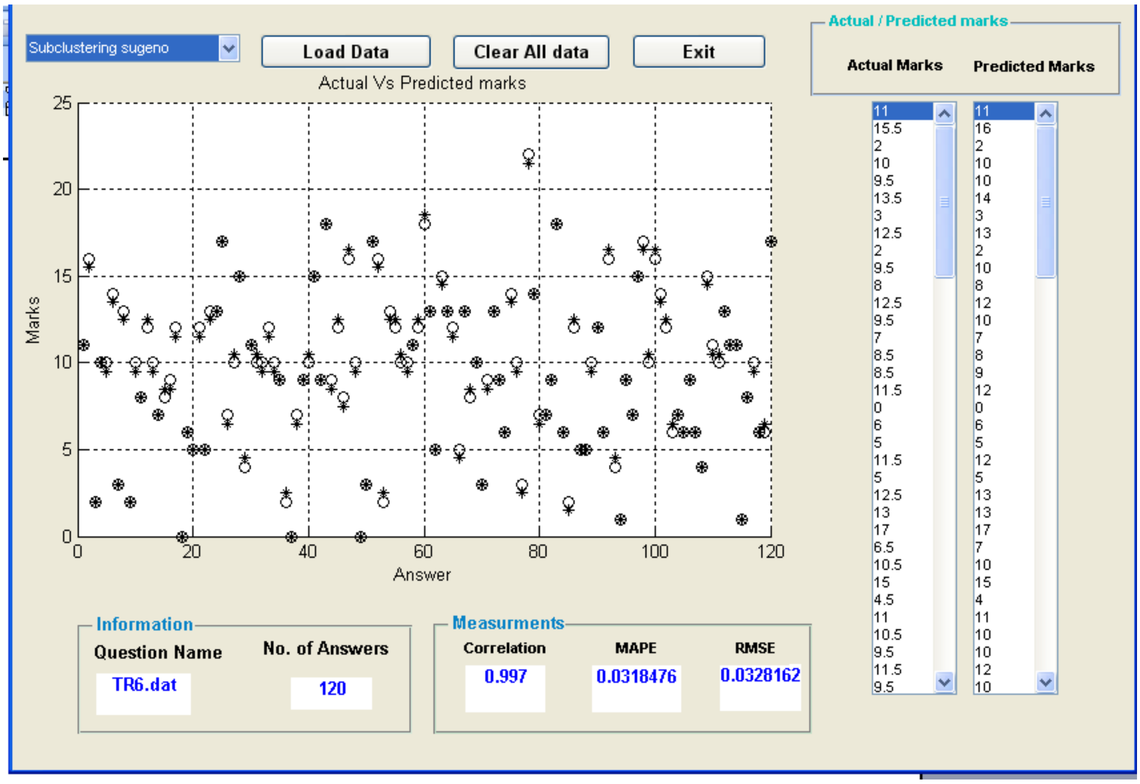


Fig: 4.12: Subtractive clustering Sugeno Model plot(actual vs. predicted marks) for TR6 testing dataset.

Fig. (4.12) shows a sample of the agreement plot between actual and predicted marks related to one of the questions TR6 (Training data for question 6). The stars in fig.(4.12) represent the actual marks; whereas the circle represent the predicted marks. As we can see from the graph, there is some total match between actual and predicted marks (for example; answer No. 4, while we can see approximately a small agreement for answer No. 10).

4.7 Graphical User Interface (GUI): Automatic Essay Scoring System

Different modeling techniques for AES purposes have been developed. The developed models based on fuzzy and Neuro-fuzzy techniques and different measures were used to check the adequacy of the developed models. To simplify the use of these developed models we built stand alone application. The application can score the essay in both normal form and fuzzy form as shown in fig. (4.13).

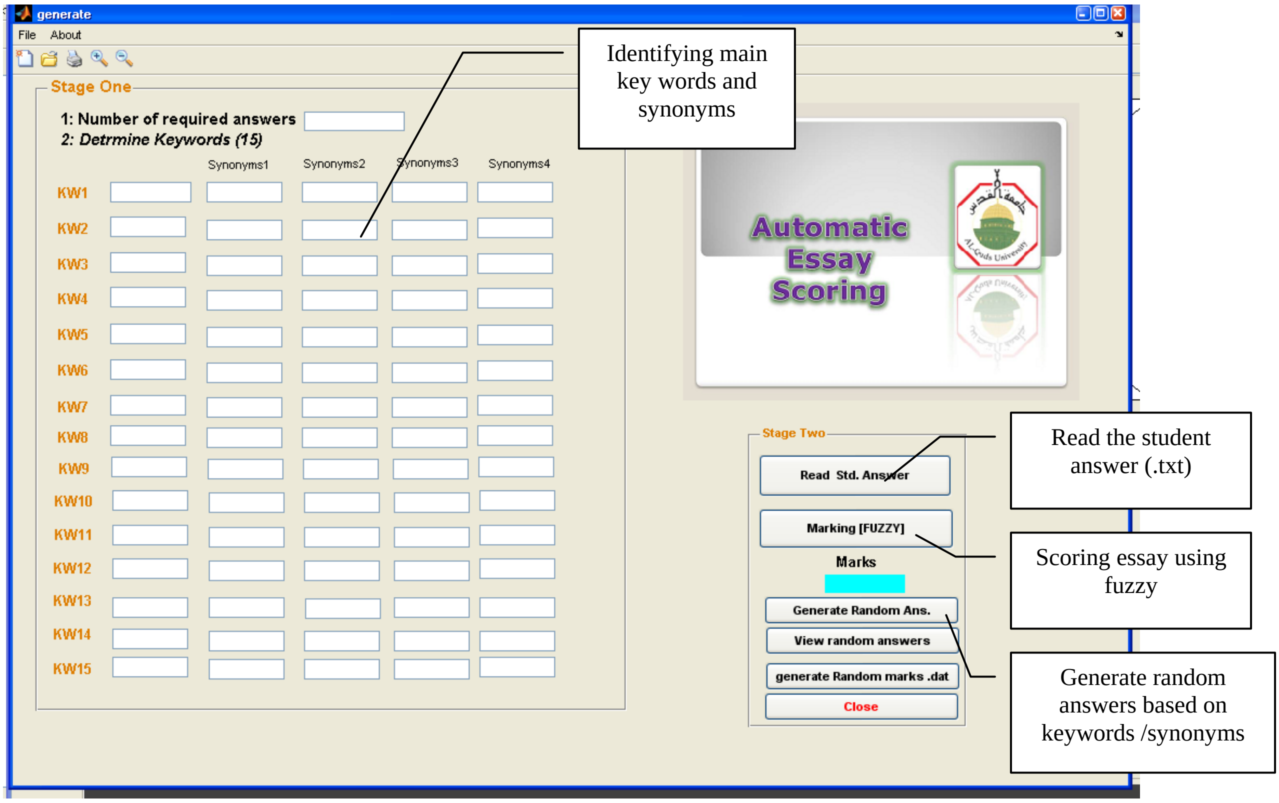


Fig. (4.13): Graphical User Interface (main screen)

Figure (4.14) shows the GUI for the essay scoring based on fuzzy and Neuro-fuzzy approaches. Using this GUI, we can load the data, choose the required FIS and plot the theoretical (actual) marks versus predicted marks. The three different measures are also appearing in the bottom of the right corner, the correlation, MAPE and RMSE.

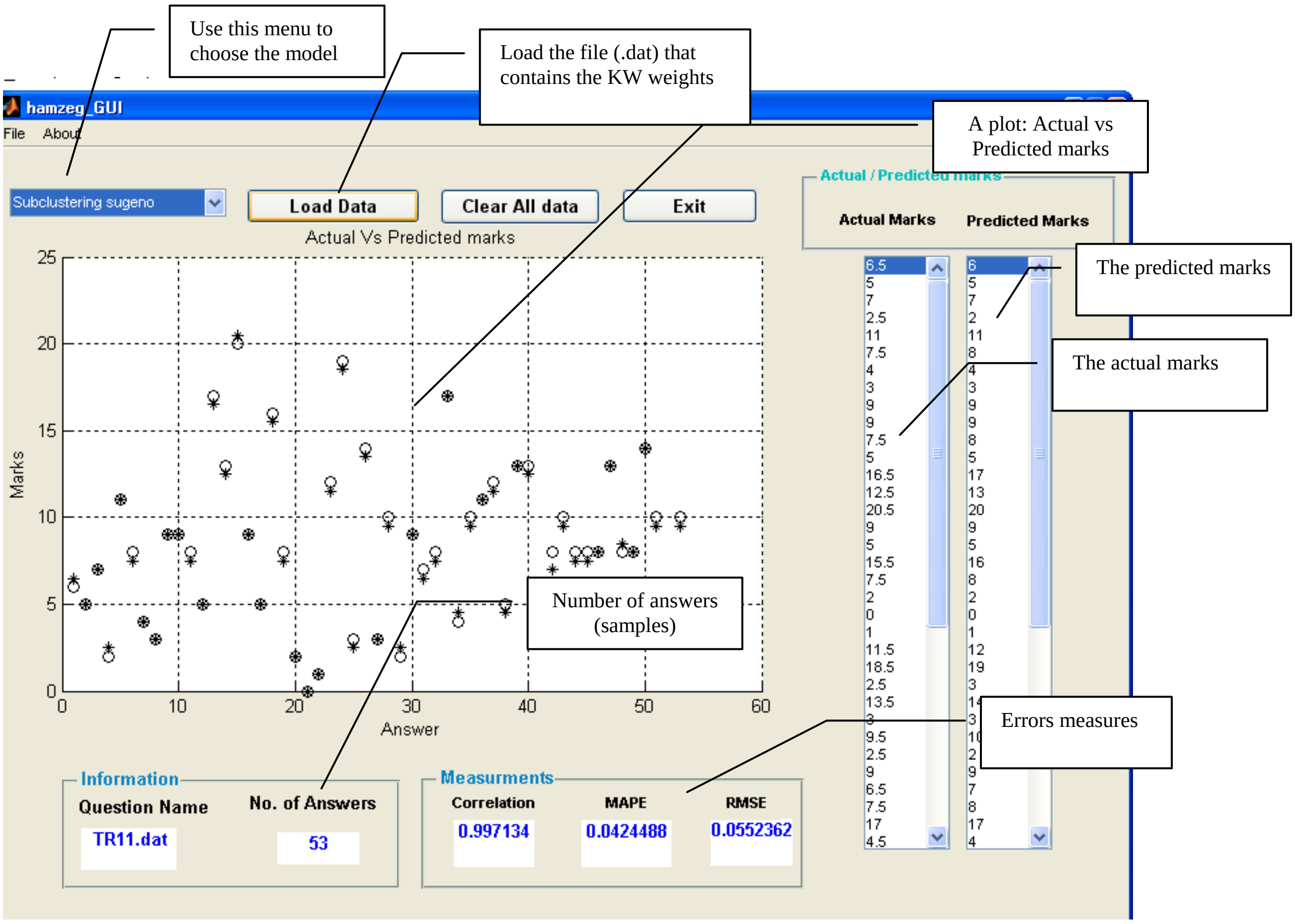


Fig. (4.14) GUI for FAESS

Fig.(4.14) represents a snapshot for the FAESS when it is used to score the answers for training set No.11. The stars represent the actual marks whereas the circles represent the predicted marks.

4.8 Testing the FAESS using unseen data from local universities (PPU and QOU)

As we mentioned in section 4.3.1 in this chapter, we obtained data from local university (PPU and QOU) in order to check it on our developed models. The detailed discussion and comparisons about these untrained data will be discussed in chapter five section three.

Chapter five

Results Discussion and Comparison

5.1 Introduction

The results of this study comprise several types of discussions and comparisons. This chapter is divided into three major sections: section one presents comparisons between the developed models. While section two covers the comparisons between the achieved results with untrained data from other institutions, section three demonstrates a comparison with other systems and results. All the measures collected and used for the analysis of this study are presented in figures and tables. In our study, we have started by developing general model called Multiple Input Single Out put (MISO) Mamdani model; then we developed two Sugeno models using different optimization technique. The obtained results are promising and showed high agreement between actual and predicted marks.

5.2 The Developed Models: Comparisons and results discussion

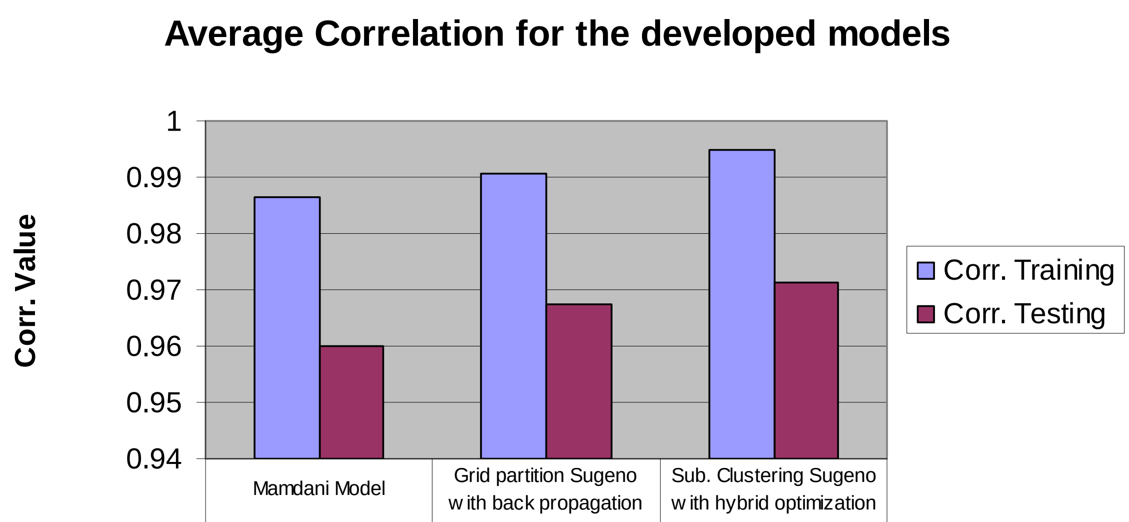
In our research study, we have developed three models based on fuzzy logic and Nuero-fuzzy approach. The developed models are used to score the essays based on content dimension. The first developed model; MISO Mamdani model was built using manual tuning for the parameters (the tuning of parameter including changing the type of memberships function manually and updating the rules). The other two models are MISO Sugeno models with different optimization techniques based on ANFIS as a major building criterion.

In table (5.1) we will present the average correlation between actual and predicted marks for the three developed models (average for training and testing data in three models), also the overall correlation average will be calculated. The general form of the rules for both Mamdani and Sugeno FIS was mentioned in chapter three section five.
Table 5.1: Average correlation for the developed models

Model	Average Correlation between actual and predicted marks		Average correlation between training and testing data for the models
	Training data	Testing data	
Multiple Input Single Output (MISO) Mamdani Model	0.9863	0.9599	0.9731
Grid partition Sugeno with back propagation	0.9906	0.9675	0.979
Sub. Clustering Sugeno with hybrid optimization	0.9948	0.9712	0.983
The average correlation for the three models	0.9905	0.9662	0.9783

The average correlation found in table 5.1 represents the results obtained from the models developed by subjecting the models to 1080 dataset for the fifteen main keywords with 718 of data for training and 362 of data for testing. In general, the average correlation values between actual and predicted marks ranges between 0.9599 and 0.9948 with average correlation for all models equal to 0.9783. Logically speaking the correlation coefficient between theoretical (actual) and predicted marks are promising and has a value more than 95%.

Fig.(5.1) shows the average correlation between actual and predicted marks for the three developed models based on 1080 data set used to develop these models.



	Mamdani Model	Grid partition Sugeno with back propagation	Sub. Clustering Sugeno with hybrid optimization
Corr. Training	0.9863	0.9906	0.9948
Corr. Testing	0.9599	0.9675	0.9712

Fig. (5.1): Graphical Representation for obtained average correlation between actual and predicted marks for both training and testing data

As we can notice from graph in fig.(5.1), the highest correlation obtained was for Sugeno subtractive clustering with hybrid optimization techniques (average correlation is 0.9948 for trained data and 0.9712 for testing data). In Mamdani model, with lowest average correlation, the average correlation for trained data is 0.9863 and 0.9599 for testing data. Grid partition Sugeno with back propagation optimization technique provide us with promising results with average correlation equal to 0.9906 for trained data and 0.96753 for untrained data (testing).

Fig. (5.2) shows the correlation obtained between theoretical and predicted marks for the three developed models. We noticed from the graph the high values of correlation which exceed 90% for all trained data.

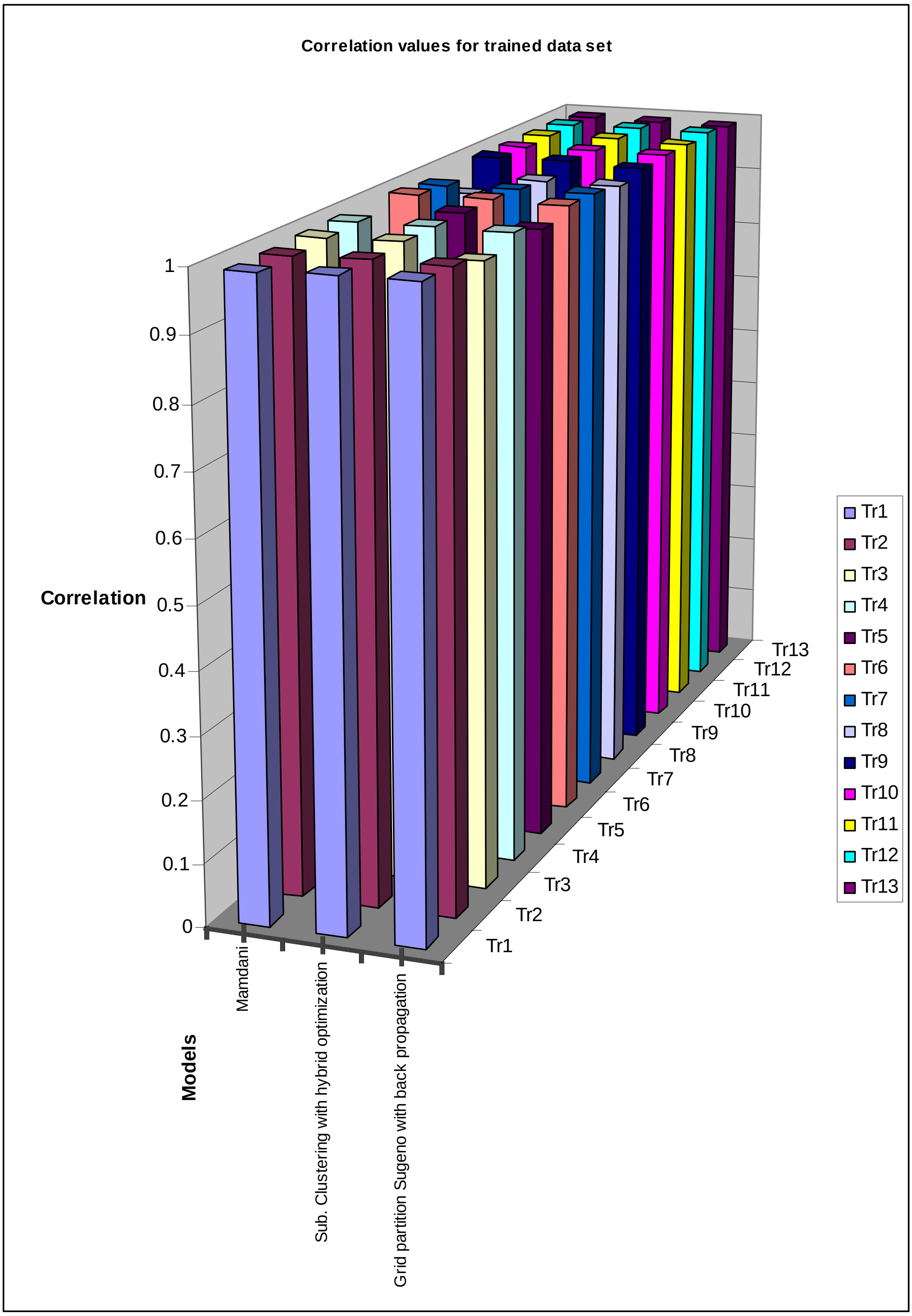


Fig. (5.2): The correlation measures obtained for trained data in the three models

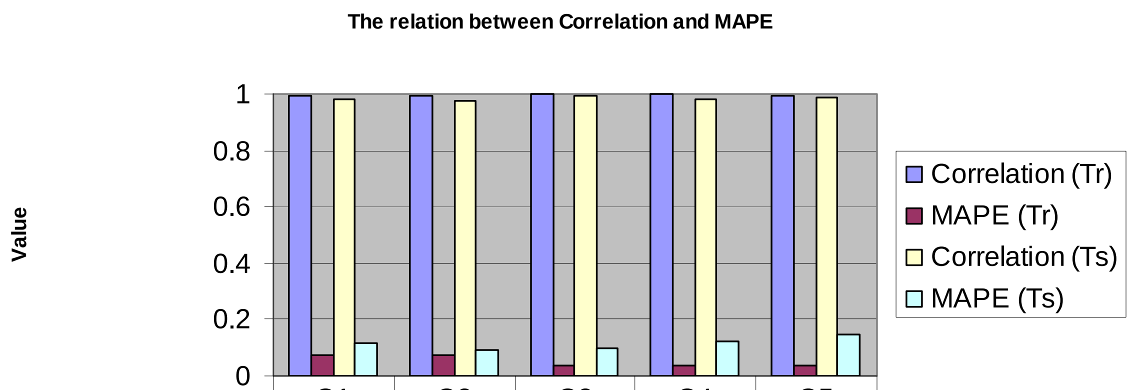
The Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) have also been used for further checking. In table 5.1, the correlation values supported us with agreement measures between actual and predicted marks, whereas error measures (RMSE and MAPE) give us an indication about the accuracy of the developed models. The obtained results of MAPE and correlation in Sugeno subtractive clustering for the data set (answers to Q1-Q5 in table 4.1) are presented in table (5.2) and graphically in fig.(5.2).

Later on this chapter we will compare the values of correlation with MAPE and RMSE to show the consistency.

Table 5.2: Correlation and MAPE values for data set Q1 to Q5

Data set	Correlation (Tr)	MAPE (Tr)	Correlation (Ts)	MAPE (Ts)
Q1	0.9946	0.074	0.9824	0.116
Q2	0.9936	0.0746	0.9767	0.0927
Q3	0.9977	0.0376	0.9913	0.0964
Q4	0.9973	0.0362	0.9816	0.1238
Q5	0.9963	0.0396	0.9903	0.1499

The values of MAPE in table (5.2) give us promising results about the success of these developed models to score essays based on content dimension. As we can see that if the correlation values increases, the error value (MAPE) decreases and vice versa.



	Q1	Q2	Q3	Q4	Q5
Correlation (Tr)	0.9946	0.9936	0.9977	0.9973	0.9963
MAPE (Tr)	0.074	0.0746	0.0376	0.0362	0.0396
Correlation (Ts)	0.9824	0.9767	0.9913	0.9816	0.9903
MAPE (Ts)	0.116	0.0927	0.0964	0.1238	0.1499

Fig. 5.3: The correlation and MAPE consistency (Sugeno subtractive)

Another important measures that show the adequacy of the develop models is RMSE. The results of RMSE for the developed modeling technique are shown in table (5.3), while fig.(5.3) presents the results graphically.

Table 5.3: RMSE values for the three developed models

Model	Average RMSE		Average RMSE for training and testing data
	Training data	Testing data	
Mamdani Model	0.087	0.2073	0.1471
Sub. Clustering Sugeno with hybrid optimization	0.0526	0.1837	0.1181
Grid partition Sugeno with back propagation	0.0809	0.2069	0.1439
The average RMSE for the three models	0.0735	0.1993	0.1364

The average value for RMSE for the three developed models is 0.1364, and subtractive clustering with hybrid optimization technique produced the least RMSE value which is 0.1181. In general, the results are promising and showed the ability of this kind of modeling technique to handle this kind of essay scoring.

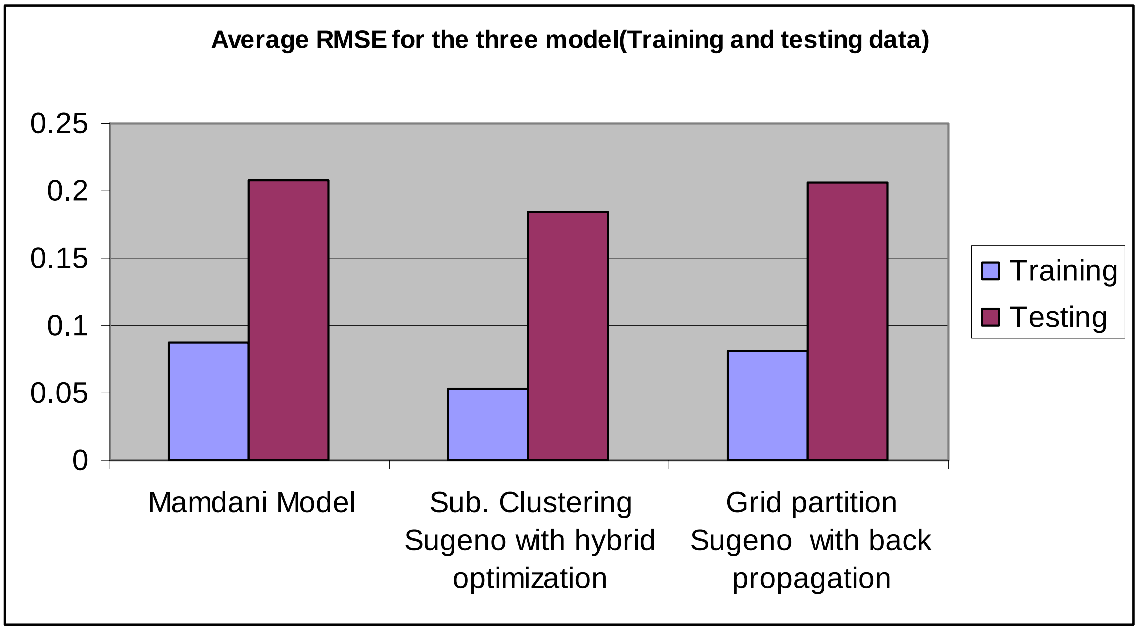


Fig.5.4: Average RMSE for the three models (training and testing data)

To check the consistency between the correlation, MAPE and RMSE, fig.(5.5) shows the correlation, MAPE and RMSE for both training and testing data. If the degree of agreement (correlation) between theoretical (actual) and predicted marks increase, the value of error measures (MAPE and RMSE) will decreases as shown in fig. (5.5)

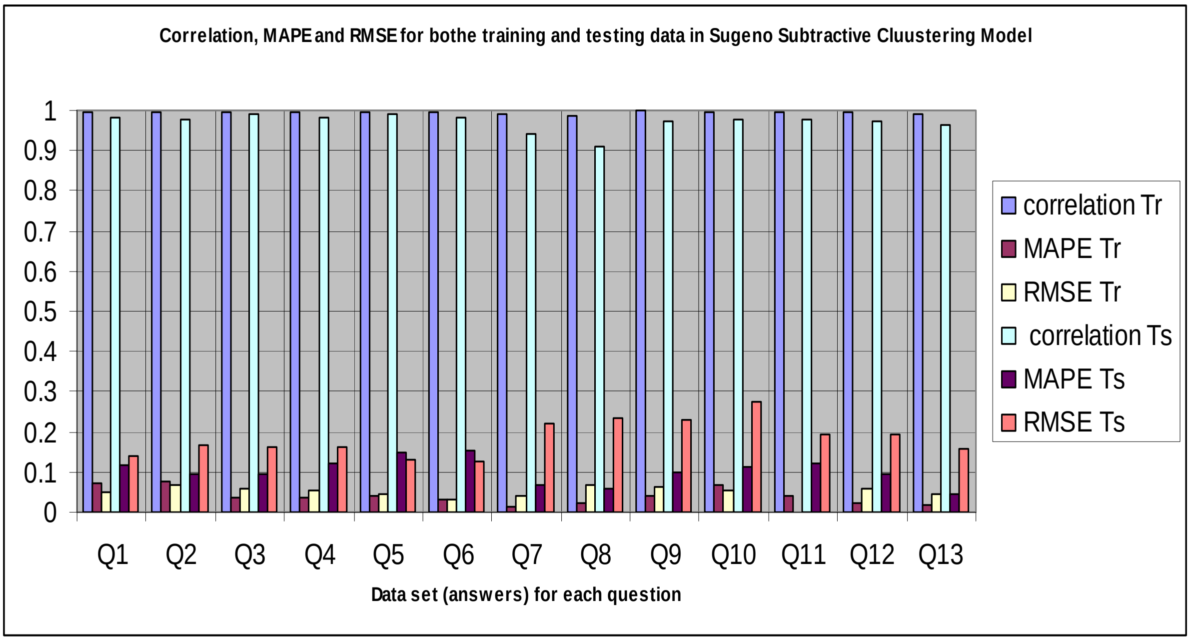


Fig. 5.5: Correlation, MAPE and RMSE for Sugeno subtractive clustering model

Fig.(5.6) shows a graphical representation for the obtained results using grid partition Sugeno for two data set(answers to Q1 and Q2 data set in table 4.1 (generated data)). The idea from this graph is to have a clearer idea about the consistency in the developed models, and how Correlation, MAPE and RMSE are complement to each other (when Correlation increase, the MAPE and RMSE value decrease).

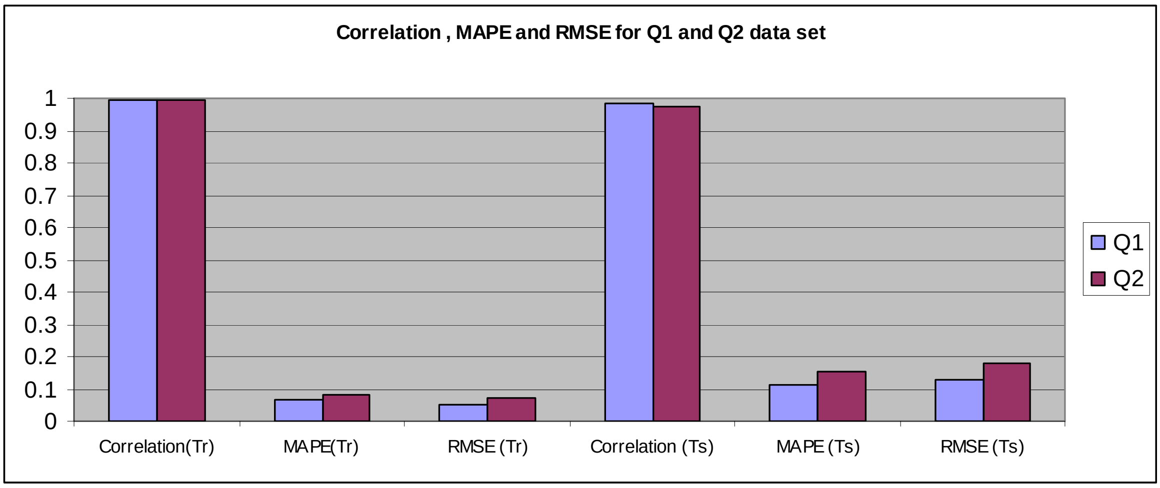


Fig. 5.6: Correlation, MAPE and RMSE for Grid partition Sugeno (Dataset one and two in the generated data; table 4.1)

5.3 The Developed Models: Comparison with data from other institutions

In order to have the ability to judge on the reliability and adequacy of your developed modeling techniques, you need to establish cooperation with experts, specialist and academic staffs. We have successfully established a contact with academic staffs from Palestine polytechnic University (PPU) and Al-Quds Open University (QOU) to support us with data in order to check it using our developed models. From PPU we obtained sample of answers for two questions. The first sample has 11 answers and the second one has 23 samples. From QOU we obtained sample of answers for one question, and the number of sample is 18.The data from PPU and QOU have been introduced in chapter four.

5.3.1 Unseen data from PPU

As we mentioned before, we got sample of answers from PPU related to subject other than history and related to two questions about information technology. We checked the students answer in two different ways. In the first question; the instructor asked the students to define the concept of “Operating System”. Then we asked the instructor to score the essays manually and at the same time we have scored the answers using our developed model; the obtained result for sample of data is shown in table 5.4;

Table 5.4: Human scoring vs. computer scoring result (PPU Q1 untrained)

Answer No.	Human marks (0-30)	Computer score (0-30)	Percentage difference ((actual-predicted)/30)*100%	Correlation
1	17.5	15	8.3%	0.9745
2	24	22	6.7%	
3	16	15	3.3%	
4	14	12	6.7%	
5	11	12	3.3%	
6	19	17	6.7%	
7	15.5	16	1.7%	
8	15	15	0.0%	
9	17.5	17	1.7%	
10	14.5	12	8.3%	
11	11	11	0.0%	
12	12	11	3.3%	
13	9	9	0.0%	
14	6.5	6	1.7%	
15	11.5	11	1.7%	
16	10.5	10	1.7%	
17	11	10	3.3%	
18	9	10	3.3%	
19	16	15	3.3%	
20	11.5	12	1.7%	
21	9	10	3.3%	
22	25	24	3.3%	
23	16.5	17	1.7%	

As we can note from table 5.4, the difference between actual and predicted marks is small, and the percentage degree is less than 9% in all students’ answers despite the fact that this data is untrained data. Also the value off correlation coefficient between predicted and actual marks is high and equal to 0.9745.

In fig.(5.7) a sample plot between actual and predicted marks for the data that have been presented in table 5.4(PPU data).

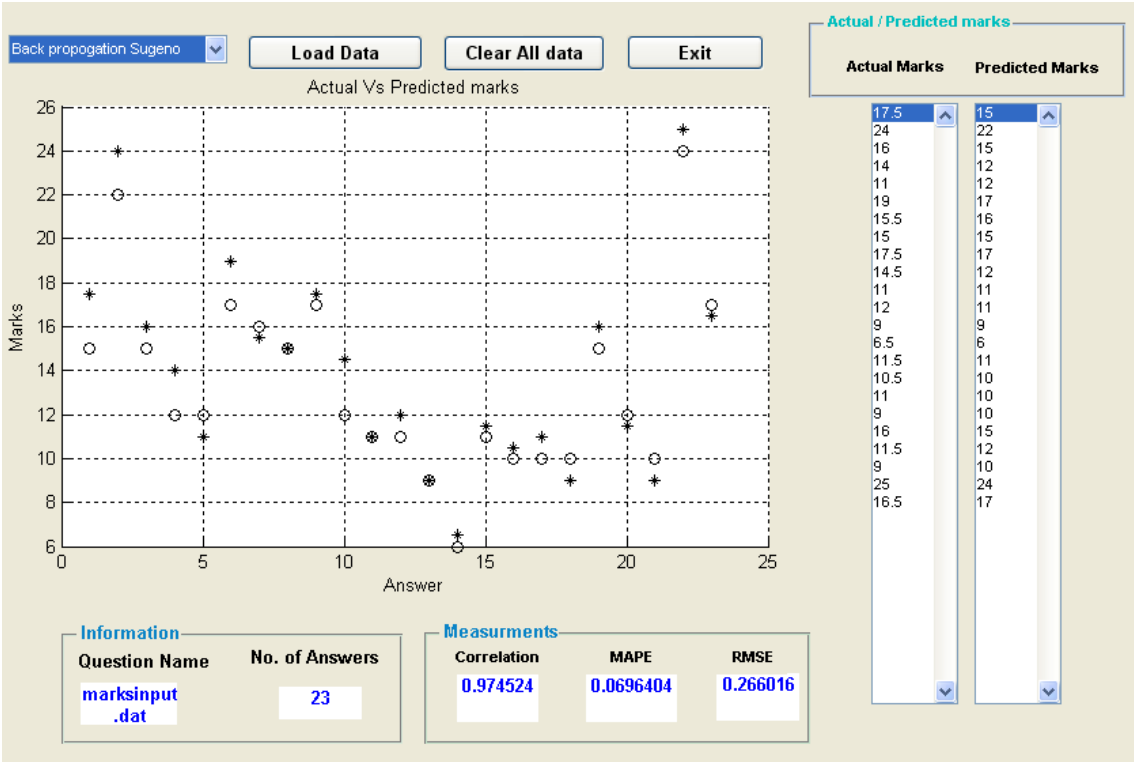


Fig. 5.7: Sample plot between actual and predicted marks for unseen data (Q1 untrained data PPU)

Despite the fact that this data is untrained data, the correlation in fig.(5.7) is high and equal to 0.9745; the MAPE is 0.0694 and RMSE is 0.266 which is small values and this indicates a promising result to use these kinds of modeling techniques in scoring essays.

Regarding the second question, the instructor in PPU asked the student to define the concept of "Programming". Table 5.5 shows the obtained results and the value of correlation is high in spite the fact that the data is untrained.

Table 5.5: Human scoring vs. computer scoring result (PPU Q2 untrained)

Answer No.	Human marks (0-30)	Computer score (0-30)	Percentage difference ((actual-predicted)/30)*100%	Correlation
1	18.5	17	5.0%	0.9755
2	9.5	8	5.0%	
3	11	10	3.3%	
4	8	9	3.3%	
5	15	14	3.3%	
6	17	17	0.0%	
7	9.5	9	1.7%	
8	7.5	8	1.7%	
9	6.5	5	5.0%	
10	12	11	3.3%	
11	7.5	8	1.7%	

Table 5.5 shows the value for untrained data from PPU(question two) and we can noticed that the percentage degree in differences between the actual and predicted marks is small and has the values less than 5% for all students' answers.

5.3.1 Unseen data from QOU

We have been in contact with Dr. Nadia Qawasmi from QOU; the data obtained from QOU is related to English subject and specifically in literature. From QOU we have a samples of answers related to one question, the instructor asked the student to define the concept of "Allegory". Table 5.6 shows the results obtained for QOU data (described in chapter four and appendix C).

Table 5.6: Human scoring vs. computer scoring result (QOU untrained data)

Answer No.	Human marks (0-30)	Computer score (0-30)	Percentage difference ((actual-predicted)/30)*100%	Correlation
1	17	15	6.7%	0.9954
2	10.5	8	8.3%	
3	14	13	3.3%	
4	12	11	3.3%	
5	10.5	9	5.0%	
6	7	5	6.7%	
7	5	4	3.3%	
8	2	2	0.0%	
9	11	9	6.7%	
10	13	12	3.3%	
11	12	11	3.3%	
12	11	10	3.3%	
13	13.5	12	5.0%	
14	6	5	3.3%	
15	10	9	3.3%	
16	8.5	7	5.0%	
17	6	5	3.3%	
18	30	29	3.3%	

The correlation value is high and equal to 0.9954. The value of MAPE and RMSE is 0.1292 and 0.348. This provides us with promising indication to adopt these models for scoring essays. Fig.(5.8) shows a sample plot between actual and predicted marks for unseen data from QOU.

From the tables 5.4, table 5.5 and table 5.6 we may notice that the differences between the actual (human) scoring the predicted one range between 0% to 8.3%. That is in marks, it ranges between 0 to 2.5 marks out of the total 30 marks. Although, the number of these samples are not so large (23, 11, and 18), we may conclude that this approach is adequate and suitable to address and solve the AES problem.

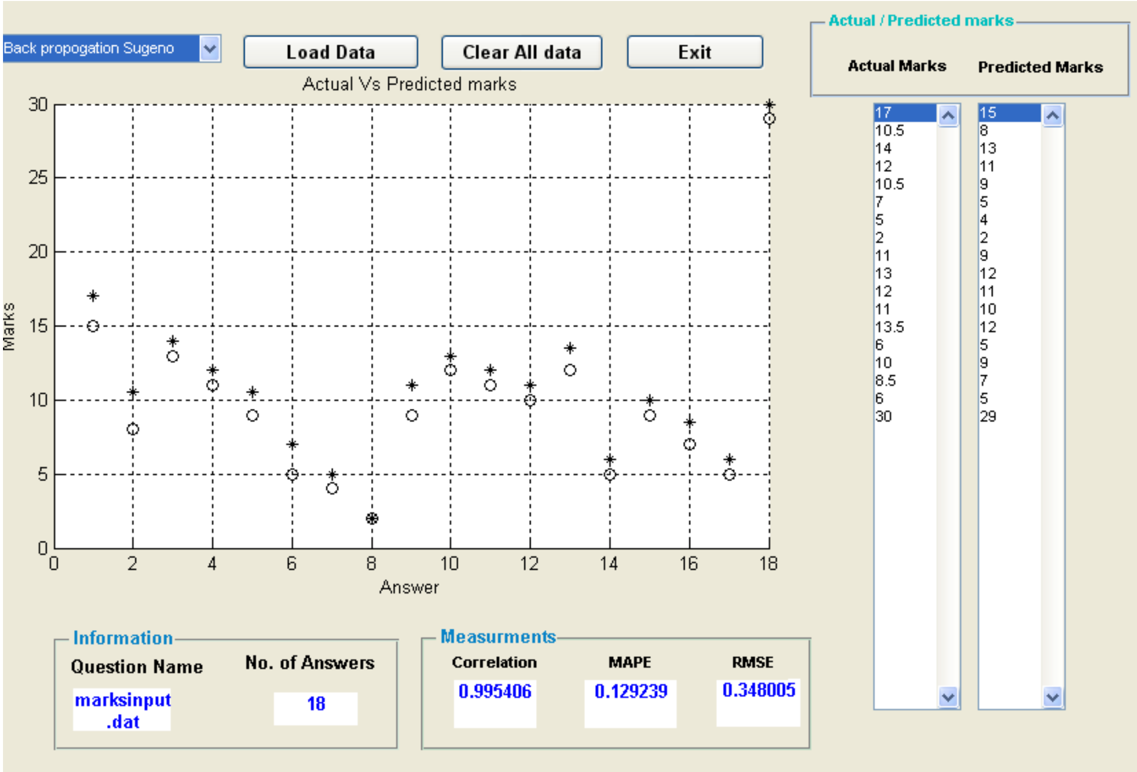


Fig. 5.8: sample plot between actual and predicted marks for unseen data (QOU)

5.4 The Developed Models: Comparison with other systems

One of the main advantages of AES system is that they can score essays instantly and provide immediate feedback. Most AES system tends to focus on product rather than process in writing (Dikli, 2006).Validation of automated essay scoring systems can be performed in various ways. A common way to validate computer –generated score is by comparing the correlation between computer-generated and human-rater scores to the correlation obtained between two human raters.

Although , all AES systems mentioned previously have been subject to different data set, but still we may use the obtained correlation coefficients between actual and predicted marks as a measure of comparison between their performance. An example about correlation coefficients measurement, Rudner and Liang reported that the Bayesian approach presented correlation accurate results as high as 0.80(Rudner and Liang, 2002).

The intelligent Essay Assessor (IEA) has been tested by comparing essays to one that have been previously scored; the correlation agreement ranging from 85% to 91% (average 88%). E-rater was tested on unseen data with correlation agreement between scores ranging from 84% to 91% (average 87.5%) (Sukkarieh, et al, 2003). The average correlation for our developed models is 0.9783. The underlying idea in IEA is to identify which of several calibration documents are most similar to the new document based on the most specific (i.e., least frequent) index terms. For essays, the average grade on the most similar calibration documents is assigned as the computer-generated score (Landauer, et al, 1998). A review of the research on IEA found that its scores typically correlate as well with human raters as the raters do with each other (Chung and O'Neil, 1997). PEG is also the better choice for evaluating writing style, as IEA returns grades that have literally nothing to do with writing style. IEA, e-rater and FAESS, however, appear to be the superior choice for grading content, as PEG relies on writing quality to determine grades. Table (5.7) summarizes the average correlation for PEG, IEA, modeling techniques applied to short essay auto-Grading problem, E-rater and our developed model (FAESS).

We can say that there are several reasons to compare our model wit other mode because there is similarity in some area. For example (IEA) has been tested by comparing essays to one that have been previously scored with average correlation equal to 88%. in our model we adopted this technique by tested our model with data that have been previously scored. Another important issue is E-rater; E-rater was tested on unseen data with average correlation equal to 87.5%. We, on our developed model used unseen data from both PPU and QOU. As a result of this reasons, we can compare our models with other AES despite the fact that there is difference in data used and techniques adopted as appear in table 5.7.

Table 5.7: Comparison between AES systems (Correlation coefficient).

System	Correlation
PEG	0.87
IEA	0.88
E-rater	0.88
FAESS	0.97
Modeling techniques applied to short answers	0.98

Fig.(5.9) shows the correlation plot for five Automatic Scoring System. We noticed from the graph that our developed model has the high correlation between actual and predicted marks which is 0.97 approximately.

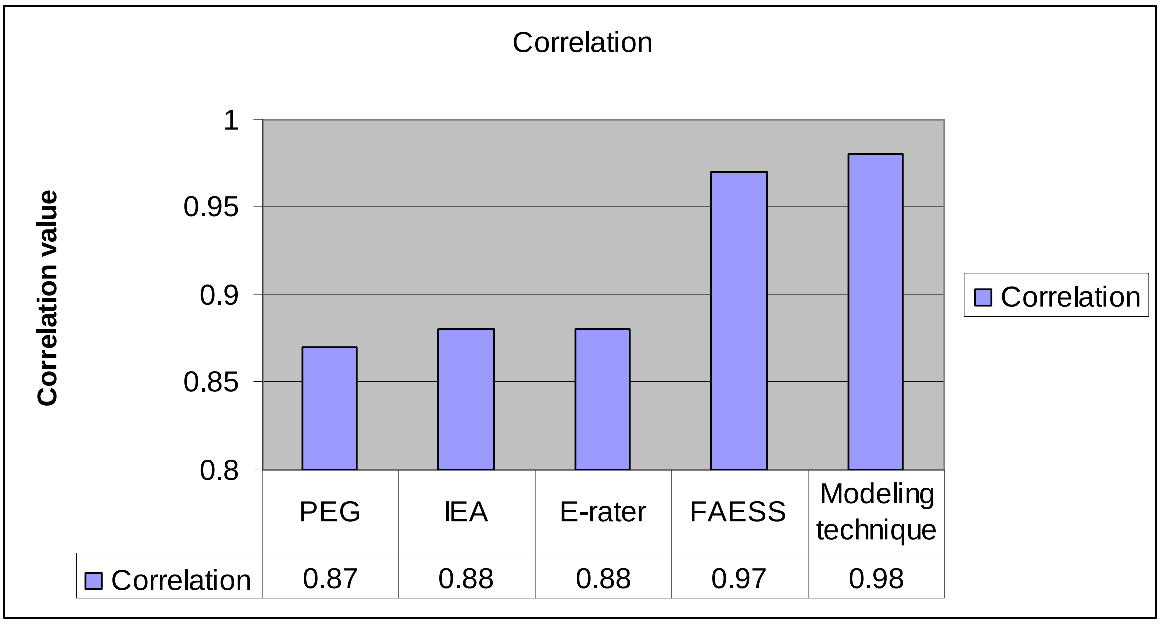


Fig. 5.9: The correlation comparison between AES systems

Chapter Six

Conclusion and further work

6.1 introductions

The demand for developing Automatic Essays Scoring systems in writing assessment is increasing. Assessment is considered to play a central role in the educational process. The interest in the development and in use of Computer-based scoring systems has grown exponentially in the last few years, due both to the increase of the number of students attending universities and to the possibilities provided by e-learning approaches to asynchronous and ubiquitous education. Some instructors and academic staffs fear that AES systems will eventually substitute humans, thus, the producer of these AES systems claim that the main role of these systems is not to replace teachers but to assist them. However, it is not clear to what extent AES can replace human raters in judging the quality of essay writing.

6.2 Conclusion

This study has presented a new approach in the field of AES. The main purpose of this study is to show that this approach will strengthen the current traditional arithmetic and statistical methods. The development of these models based on fuzzy and Neuro-Fuzzy techniques and we succeed to establish systematic approach to score students essays based on content's dimension.

During our literature review of the currently available AES, we have concentrated on the most widely used AES systems: Project Essay Grader (PEG), Intelligent Essay Assessor (IEA), E-rater and Bayesian Essay Test Scoring System (BETSY). Also we reviewed a model called "Modeling Techniques applied to Short Essays Auto-grading Problem". After further reading among these systems we may conclude the following:

- Most of AES systems are mainly developed based on English language. There are efforts to enable these systems to score essays in various languages.
- Among these previous systems, PEG is considered to be the best choice for evaluating writing style.
- IEA, E-rater and Oriqat model are suitable for scoring content dimension.
- BETSY combines some features from PEG and E-rater.

There are different advantages and disadvantages to use AES system. One of the main advantages of AES system is that they can score essays directly and provide immediate feedback to the students. The feedback from our developed models is the essay scoring result (mark). The teacher response is very important to the students in order to improve his writing ability and strengthen his answers in the future. The main disadvantages in using AES systems is that computers are not capable of scoring an essays as a human rater can do because the computer do what they are programmed to do.

Our developed FAESS has successfully scored the essays with agreement more than 97% with human grader scores. The preprocessing steps required for automatic scoring are mostly automated. Manual processing such as keywords selection could be automated in the future. Our method explicitly analyzes the content of the text. FAESS is based on input (keywords weight), process and output (predicted marks). In our process design, we have generated 1080 datasets of fifteen main keywords with specific constraints. In developing our modeling techniques, we used FIS and neural learning approaches to develop three models. Each model use different integrated and adaptive fuzzy inference systems with hybrid and back propagation optimization techniques. The first developed model was MISO Mamdani, in Mamdani model we trained the system manually (building the rules and selection membership functions, etc). The second model is MISO subtractive clustering Sugeno with hybrid optimization technique. Finally, the third model is MISO Grid partition Sugeno with back propagation optimization technique. The last two models were developed based on ANFIS.

Essays scoring based on content dimension have some difficulties when we are dealing with word's negation issue in the sentences. Most of previous studies and researches about AES based on content dimension ignore theses issue, in our approach, we have applied the following criteria: for any keywords or synonyms with negations before it, negative weights will assign. Later on fuzzy scoring method, the negative input was processed by suitable membership function to illustrate that there is a negation in the student answers. Using this criterion we can solve the problem of negation issue in AES based on content dimension. We have scored answers containing negation and the results are promising. Another important problem an AES based on content dimension is ignoring the order of the words in the sentences when we score the essays. In this study we succeed to identify the necessary procedures and steps required to solve the problem of words order in the sentence by using language parser.

The adequacy of the developed models has been checked using different measures including the correlation coefficient between theoretical (actual) and predicted marks, the MAPE and RMSE. For example, in subtractive clustering model, the value of correlation ranges between 0.9121 and 0.9977 with an average correlation 0.9948 for the training data and 0.9712 for the testing data. The value of MAPE ranges between 0.015 and 0.1516 with average MAPE 0.0399 for the training data and 0.1024 for the testing data. Finally, the value of RMSE ranges between 0.0328 and 0.2763 with average RMSE 0.0526 for the training data and 0.1837 for the testing data. From this result we conclude the ability of this model to score the essays in proper way since we had a high correlation coefficient between theoretical (actual) and predicted marks, and also the values of error measures (MAPE and RMSE) are small.

The results of the current study are feasibly accepted. The comparisons between the developed models using average results of correlation, MAPE and RMSE showed that there is a high degree of an agreement between actual and predicted marks. For further check, we have compared our results with other systems. The preliminary results are promising despite there are differences between the findings of the current study and those by the previous studies. The differences appears due to various factors, such as the difference in scoring dimensions(grammar, content, style, etc) or the differences in location, language, types of writing instruction and data used. For further investigation, we have checked unseen data from local universities (PPU and QOU). The obtained results are promising since we had a high correlation coefficient between actual and predicted marks and small values in the error measures (MAPE and RMSE).

Our developed models have been integrated with a stand alone application with GUI. The developed Automatic Essays Scoring System has the ability to be accessed on line through using web technology to score the written essays. The user can feed the system with keywords and synonyms of these keywords. After that, the user read the student answers stored in text format. Then, the user of the system can use normal score to grade the essay or he can used scoring method based on fuzzy and Neuro-fuzzy approach. The output from the system is the answer result (marks) together with a graphical plot between actual and predicted marks to show the degree of agreement between the two marks.

When we compared our developed models with other AES systems (i.e. PEG, IEA, E-rater, BETSY and Modeling Techniques applied to Short Essays Auto-grading Problem) we have some similar characteristics with other AES system (i.e. scoring essay based on English language). Also there are some differences between our developed models and other AES system this include the scoring dimension used (content, grammar, etc) and the number of samples required to train the models, etc.

The obtained results from unseen data from local universities support us with promising results since we had a high correlation coefficient between actual and predicted marks and small values in the error measures (MAPE and RMSE). We may notice that the differences between the actual (human) score and the predicted one range between 0% and 8.3% that is in marks, it ranges between 0 to 2.5 marks out of the total 30 marks. Although, the number of these samples are not so large (11, 23, and 18), we may conclude that this approach is adequate and suitable to address and solve the AES problem

6.3 Further work

There are several suggestions and issues in the field of AES that needs further research and investigations. These include:

- Further investigation with more samples of data using different types of questions from different fields.
- Having the ability to combine the AES system with e-learning environments such as MOODLE and other distance learning technology.
- Providing the students with feedback about the topics that the student did not cover in his answer by providing him with typical answers and other missing information in his answer.
- Developing a parser tree file to solve the problem of word’s order in content-based AES.
- Developing AES systems based on Arabic language since most AES systems re based on English language.

- GUI enhancement by enabling the teacher to select the weights of the keywords and by providing the teacher with some details about the answer (for example: the length of the answers).
- Generalizability of AES tools:
    - Using web technology to allow the teacher to use the system.
    - On the webpage where students submit their answers, students can click different buttons and submit their responses either for automated feedback or to their instructor as their final answer.
- Reducing the system constraints (for example: the use of mathematical equations).

References

1. Abdi, H. (2003): Neural Networks, Cognition and Neurosciences Program, the University of Texas at Dallas, 2003.

2. Abraham, A. (2004): Adaptation of Fuzzy Inference System Using Neural Learning, Chapter Three, pp. 53-83. Computer Science Department, Oklahoma State University, USA, 2004.

3. Anthony, M. (1999): Artificial Neural Networks. Cambridge University Press, 1999.

4. Arafeh, L., Singh, S., Putatunda, S. (1999): “A neuro fuzzy logic approach to material processing”, IEEE Transactions on Systems, Man, and Cybernetics, Part C, Vol. [29], No.(3), pp. 362-370, 1999.

5. Balog, E., Berta, I. (2001): “Fuzzy solutions in electrostatics”, Journal of Electrostatics, Vol. [51], pp. 409-415, 2001.

6. Birenbaum, M., Tatsuoka, K. (2008): Open-Ended Versus Multiple-Choice Response Formats, It Does Make a Difference for Diagnostic purpose, Applied Psychological Measurement. Vol. [11], No. (4), pp 385-395, 2008.

7. Burstein, J., Chodorow, M., Leacock, C. (2003): “Criterion[SM] Online essay Evaluation: An Application for Automated Evaluation of Student Essays”, Proceedings of the Fifteenth Annual Conference on Innovative Applications of Artificial Intelligence, vol. [2], pp. 1242-1249, Mexico , 2003.

8. Burstein, J., Chodorow, M. (1999): Automated Essay Scoring for Nonnative English Speakers. Workshop On Computer Mediated Language Assessment And Evaluation In Natural Language Processing, the Pennsylvania State University, 1999.

9. Castellano, G., Fanelli, A., Mencar, C. (2003): Design of Transparent Mamdani Fuzzy Inference Systems, CILAB: Computational Intelligence Laboratory, Department of Computer Science, University of Bari, Bari, Italy, pp. 468 – 476, 2003.

10. Chowdhury, G. (2002): Natural Language Processing, Department. of Computer and Information Sciences, University of Strathclyde, Glasgow, UK, Annual Review of Information Science and Technology, vol. [37], pp.51-89, 2002.

11. Chung, G., O’Neil, H. (1997): Methodological approaches to online Scoring of Essays, Center for the study of evaluation, University of California , Los Angelos. Vol. [418], 1997.

12. Cohen, Y., Ben-simon, A., Hovav, M. (2003.): “The effect of specific Language Features on the Complexity of Systems for Automated Essay Scoring”, the IAEA 29th Annual conference, Manchester, UK, October, 2003.

13. Dikli, S. (2006): “An Overview of Automated Scoring of Essays”, The Journal of Technology, Learning, and Assessment, Vol. [5], august, 2006.

14. Dikli, S. (2006): “Automated Essay Scoring”, Turkish Online Journal of Distance Education (TOJDE), Vol. [7], No. (1), Article: 5, January, 2006.

15. Dogan, E., Sasal, M., and Isik S. (2005): “Suspended Sediment Load Estimation in Lower Sakarya River by Using Soft Computational Methods”, Proceeding of the International Conference on Computational and Mathematical Methods in Science and Engineering, CMMSE 2005 Alicante, Spain, ,pp. 395-406, June, 2005.

16. Hearst, M. (2000): “The Debate on Automates Essay Grading”, University of California, Berkeley, California, USA. Vol.[15], Issue 5, pp. 22 – 37, Sep/Oct 2000.

17. Hélie, S., Chartier, S., Proulx, R. (2003): Applying Fuzzy Logic to Neural Modeling, Centre-Ville Montréal, Canada, 2003.

18. Hwang, Y., et al (2008): MUFIS: A neuro-Fuzzy Inference System using multiple types of fuzzy rules, IEEE World Congress on Computational Intelligence, pp.411-1414, June, 2008.

19. Imperial college, 2008, (http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol1/cs11/article1.html, 15.October. 2008).

20. Jang, J. (1993), ANFIS: Adaptive-network-based fuzzy inference system, IEEE Transactions on Systems and Machine Cybernetics, Vol. [23], No.(3), pp. 665–685, 1993.

21. Jain, A., et al (1999): Data Clustering: A Review, ACM Computing Surveys, Vol. [31], no. (3), 1999.

22. Jen-Yuan, et al (2004): “Text summarization using a trainable summarizer and latent semantic analysis”, Journal of information processing & management, vol. [41], no.(1) ,pp. 75-95, 2004.

23. Kanejia, D., Kumar, A., Prasad, D. (2003): “Automatic Evaluation of Student's answer using syntactically Enhanced LSA”, Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing – Vol. [2], pp. 53 – 60, 2003.

24. Knowledge technology (2008), http://www.knowledge-technologies.com/papers/IEA_FAQ.pdf, 11, Dec, 2008)

25. Landauer, K., Foltz, P., Laham, D. (1998): Introduction to Latent Semantic Analysis, Research supported in part by a contract from ARPA-CAETI, Department of Psychology, University of Colorado, 1998.

26. Landauer, T., et al(1997): "How Well Can Passage Meaning be Derived without Using Word Order?, A Comparison of Latent Semantic Analysis and Humans", Proceedings of the 19th annual meeting of the Cognitive Science Society, University of Colorado, pp.412-417, 1997.

27. Liddy, E. (2003): Encyclopedia of Library and Information Science, 2nd Ed. Marcel Decker Inc, 2003.

28. Mar´ın, D. (2004): Automatic evaluation of users' short essays by using statistical and shallow natural language processing techniques, Advanced Studies Diploma Work, University of Madrid, may, 2004.

29. Matlab, R14, (2004): The Language of technical computing, The Math works Inc,3 Apple Hill Drive, Natick, MA, (www.mathworks.com).

30. Millett, R. (2006): Automatic Holistic Scoring of ESL Essays using Linguistic Maturity Attributes, A thesis submitted to the faculty of Brigham Young University in partial fulfillment of the requirements for the degree of Masters of Arts, Department of Linguistics and English Language, Brigham Young University, August, 2006.

31. Oriqat, Y. (2007): Modeling Techniques Applied to short Essay Auto-Grading problem, M.Sc. Thesis, Al-Quds University, Jerusalem, Palestine, 2007.

32. Palmer, J., William, R., Dreher, H. (2002): Automated Essay Grading System Applied to a First Year University Subject, How Can We do it Better? , Curtin University of Technology, Perth, WA, Australia, June, 2002.

33. Pulman, S., Sukkarieh, J. (2005): Automatic Short Answer Marking, Computational Linguistics Group, Centre for Linguistics and Philology, University of Oxford, UK, 2005.

34. Rajagopalan ,A., et al (2003): Development of Fuzzy Logic and Neural Network Control and Advanced Emissions Modeling for Parallel Hybrid Vehicles, Center for Automotive Research, Intelligent Structures and Systems Laboratory ,The Ohio State University, Columbus, Ohio, 2003.

35. Rasmani, K. (2002): A Data-Driven Fuzzy Rule-Based approach for Student academic Performance Evaluation, The University of Edinburgh, PhD Thesis Proposal, Center of Intelligent Systems and Their applications, November, 2002.

36. Reja, U., et al (2003): Open-ended vs. Close-ended Questions in Web Questionnaires, Development in social science methodology, Slovenia, pp. 159-177, 2003.

37. Rudner, L., Liang, T.(2002):"Automated Essay Scoring Using Bayes' Theorem", The Journal of Technology, Learning, and Assessment, Vol.[1], No.(2), 2002.

38. Schnitma, L., et al (2001): "Fuzzy structures in dynamic system modeling", Proceedings of the IASTED International Conference on Control and Application (CA'2001), Banff, Canada, pp. 160-165, June, 2001.

39. Sharma, V. (2008): "Cutting tool wear estimation for turning", Journal of Intelligent Manufacturing. Vol.[19], No.(1), pp. 99-108, February, 2008.

40. Srihari, S., et al (2007): Automatic Scoring of Short Handwritten Essays in Reading Comprehension Tests, Center of Excellence for Document Analysis and Recognition (CEDAR),The state University of New York, June,2007.

41. Stergiou, C. and Siganos, D., 1996, "Neural networks," Surprise Journal, Vol. [14], (http://www.doc.ic.ac.uk/nd/surprise_96/journal/vol4/cs11/report.html, 22.August.2008).

42. Sukkarieh, J., Pulman, S., Raikes, N. (2004): "Auto-marking: using computational linguistics to score short, free text responses", Paper presented at the 30th annual conference of the International Association for Educational Assessment (IAEA), Philadelphia,USA, 2004.

43. The Math Works, (2008): (http://www.mathworks.com/access/helpdesk/help/pdf_doc/fuzzy/fuzzy.pdf, 16.October.2008).

44. Valenti,S., Neri, F., Cucchiarelli ,A. (2003): "An Overview of Current Research on Automated Essay Grading", Journal of Information Technology and Education, vol.[2], 2003.

45. Wagner, J., et al (2008):Automatic Grammaticality Judgments, National Centre for Language Technology ,School of Computing, Dublin City University, Ireland, 2008.

46. Wang, J., Brown, M. (2007): "Automated Essay Scoring Versus Human Scoring: A comparative Study", The Journal of Technology, Learning and Assessment, vol.[6], No.(2), 2007.

47. Wiemer-Hastings, P. (2004): Latent Semantic Analysis, 2nd edition, School of Computer Science, Telecommunications, and Information Systems, Oxford, UK, 2004.

48. Wild, F., et al (2005): Parameters Driving Effectiveness of Automated Essay Scoring with LSA, Department of Information Systems and New Media, Vienna University of Economics and Business Administration, Vienna, Austria, 2005.

49. William, R. (2006): "The Power Normalized Word Vectors for Automatically Grading Essays", Journal of Issues in informing science and information technology, vol.[3], pp. 721- 728 , 2006.

50. Williams, R., Dreher, H. (2005): Telecommunications Use in Education to Provide Interactive Visual Feedback on Automatically Graded Essays, School of Information Systems, Curtin University of Technology, 2005.

51. Zadeh, L. (1994): Fuzzy logic, Neural network and Soft computing, Communication of the ACM, Vol. [37], pp. 77-84, 1994.

Appendices

Appendix A: Glossary

AES	Automatic Essay Scoring
ANFIS	Adaptive Nuero-Fuzzy Inference System
ANN	Artificial Neural Network
BETSY	Bayesian Essay Test Scoring sYstem
ETS	Educational Testing Services
FAESS	Fuzzy Automatic Essay Scoring System
FIS	Fuzzy Inference System
FL	Fuzzy Logic
FRBS	Fuzzy Rule-base System
GMAT	General Management Admission Test
KAT	Knowledge Analysis Technology
IEA	Intelligent Essay Assessor
LSA	Latent Semantic Analysis
MAPE	Mean Absolute Percentage Error
MF	Membership Function
MISO	Multiple Input Single Output
MOODLE	Modular Object Oriented Dynamic Learning Environment
NLP	Natural Language Processing
PEG	Project Essay Grader
PPU	Palestine Polytechnic University
QOU	Al-Quds Open University
RMSE	Root Mean Square Error
SVD	Singular Value Decomposition
TWE	Test of Written English
Tr	Training set
Ts	Testing set
UN	United Nations

Appendix B: Matlab Code

***This function is used to calculate the values of correlation, MAPE and* RMSE.**

```
function[corrxy,mapexy,rmseyx]=eval_CorrMape(ActD,PredD)
yi=ActD;
xi=PredD;
[m n]=size(xi);% determine the size of predicted marks
[o p]=size(yi);%determine the size of actual marks
if(m~=o)|(n~=p) % size must be equal
   error('different size');
   end;
MAPE=0;
Nim=0;
Din=0;
y=mean(yi);
for i=1:m
   ni=(yi(i)-xi(i))^2;
   di=(yi(i)-y)^2;
   Nim=Nim+ni;
   Din=Din+di;
   merr1=(yi(i)-xi(i));
   if yi(i)==0
      merr=merr1;
   else
      merr=merr1/yi(i);
   end;
   MAPE=MAPE+abs(merr);
end;
corrxy=corr(yi,xi)              %Correlation
mapexy=((MAPE)/m)         %MAPE
rmseyx=(sqrt(Nim))/(m-1)  %RMSE
end; % the end of the function
```

This function is used to read the student answers and to prepare the input data to be scored using the fuzzy models

```
Function pushbutton11_Callback(hObject, eventdata, handles)
% hObject    handle to pushbutton11 (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)
[filename,pathname] = uigetfile('*.txt', 'Read the students answers');
cd(pathname);
finp=fopen(filename,'r');
c = fscanf(finp,'%s');

 global gmark;
 global notcounts;
```

```
 global arrayone
global allvalue
allvalue(:) = []
global counter
notcounts=0
counter=0;
allvalue = {'XXXX'}
% to read the main key words
kw1 = checkkw(char(handles.edit2));
kw2 = checkkw(char(handles.edit4));
kw3 = checkkw(char(handles.edit8));
kw4 = checkkw(char(handles.edit9));
kw5 = checkkw(char(handles.edit10));
kw6 = checkkw(char(handles.edit11));
kw7 = checkkw(char(handles.edit12));
kw8 = checkkw(char(handles.edit13));
kw9 = checkkw(char(handles.edit14));
kw10 = checkkw(char(handles.edit15));
kw11 = checkkw(char(handles.edit16));
kw12 = checkkw(char(handles.edit17));
kw13 = checkkw(char(handles.edit18));
kw14 = checkkw(char(handles.edit19));
kw15 = checkkw(char(handles.edit20));
% to read Synonym 1.6
kw1s1 = checkkw(char(handles.edit21));
kw2s1 = checkkw(char(handles.edit22));
kw3s1 = checkkw(char(handles.edit23));
kw4s1 = checkkw(char(handles.edit24));
kw5s1 = checkkw(char(handles.edit25));
kw6s1 = checkkw(char(handles.edit26));
kw7s1 = checkkw(char(handles.edit27));
kw8s1 = checkkw(char(handles.edit28));
kw9s1 = checkkw(char(handles.edit29));
kw10s1 = checkkw(char(handles.edit30));
kw11s1 = checkkw(char(handles.edit31));
kw12s1 = checkkw(char(handles.edit32));
kw13s1 = checkkw(char(handles.edit33));
kw14s1 = checkkw(char(handles.edit34));
kw15s1 = checkkw(char(handles.edit35));

% for reading synonyms 2 below
kw1s2 = checkkw(char(handles.edit36));
kw2s2= checkkw(char(handles.edit37));
kw3s2 = checkkw(char(handles.edit38));
kw4s2 = checkkw(char(handles.edit39));
kw5s2 = checkkw(char(handles.edit40));
kw6s2 = checkkw(char(handles.edit41));
kw7s2 = checkkw(char(handles.edit42));
kw8s2 = checkkw(char(handles.edit43));
kw9s2 = checkkw(char(handles.edit44));
```

```
kw10s2 = checkkw(char(handles.edit45));
kw11s2 = checkkw(char(handles.edit46));
kw12s2 = checkkw(char(handles.edit47));
kw13s2 = checkkw(char(handles.edit48));
kw14s2 = checkkw(char(handles.edit49));
kw15s2 = checkkw(char(handles.edit50));

% below for reading synonyms3
kw1s3 = checkkw(char(handles.edit51));
kw2s3= checkkw(char(handles.edit52));
kw3s3 = checkkw(char(handles.edit53));
kw4s3 = checkkw(char(handles.edit54));
kw5s3 = checkkw(char(handles.edit55));
kw6s3 = checkkw(char(handles.edit56));
kw7s3 = checkkw(char(handles.edit57));
kw8s3 = checkkw(char(handles.edit58));
kw9s3 = checkkw(char(handles.edit59));
kw10s3 = checkkw(char(handles.edit60));
kw11s3 = checkkw(char(handles.edit61));
kw12s3 = checkkw(char(handles.edit62));
kw13s3 = checkkw(char(handles.edit63));
kw14s3 = checkkw(char(handles.edit64));
kw15s3 = checkkw(char(handles.edit65));

 % below for reading synonyms 4
kw1s4 = checkkw(char(handles.edit66));
kw2s4= checkkw(char(handles.edit67));
kw3s4 = checkkw(char(handles.edit68));
kw4s4 = checkkw(char(handles.edit69));
kw5s4 = checkkw(char(handles.edit70));
kw6s4 = checkkw(char(handles.edit71));
kw7s4 = checkkw(char(handles.edit72));
kw8s4 = checkkw(char(handles.edit73));
kw9s4 = checkkw(char(handles.edit74));
kw10s4 = checkkw(char(handles.edit75));
kw11s4 = checkkw(char(handles.edit76));
kw12s4 = checkkw(char(handles.edit77));
kw13s4 = checkkw(char(handles.edit78));
kw14s4 = checkkw(char(handles.edit79));
kw15s4 = checkkw(char(handles.edit80));
% end of synonyms 4

 %  arrange each KW with its synonyms
r1={kw1;kw1s1;kw1s2;kw1s3;kw1s4}
r2={kw2;kw2s1;kw2s2;kw2s3;kw2s4};
r3={kw3;kw3s1;kw3s2;kw3s3;kw3s4};
r4={kw4;kw4s1;kw4s2;kw4s3;kw4s4};
r5={kw5;kw5s1;kw5s2;kw5s3;kw5s4};
r6={kw6;kw6s1;kw6s2;kw6s3;kw6s4};
r7={kw7;kw7s1;kw7s2;kw7s3;kw7s4};
```

```
r8={kw8;kw8s1;kw8s2;kw8s3;kw8s4};
r9={kw9;kw9s1;kw9s2;kw9s3;kw9s4};
r10={kw10;kw10s1;kw10s2;kw10s3;kw10s4};
r11={kw11;kw11s1;kw11s2;kw11s3;kw11s4};
r12={kw12;kw12s1;kw12s2;kw12s3;kw12s4};
r13={kw13;kw13s1;kw13s2;kw13s3;kw13s4};
r14={kw14;kw14s1;kw14s2;kw14s3;kw14s4};
r15={kw15;kw15s1;kw15s2;kw15s3;kw15s4};
 % search the first raw in GUI and assign input mark from 0-2
  gmark=0;
  findandmark(r1,c,1);
   findandmark(r2,c,2);
   findandmark(r3,c,3) ;
   findandmark(r4,c,4) ;
   findandmark(r5,c,5) ;
   findandmark(r6,c,6) ;
   findandmark(r7,c,7);
   findandmark(r8,c,8);
    findandmark(r9,c,9);
   findandmark(r10,c,10) ;
   findandmark(r11,c,11) ;
   findandmark(r12,c,12) ;
   findandmark(r13,c,13);
   findandmark(r14,c,14);
   findandmark(r15,c,15);
     notcounts = notcounts * 2
  if gmark >= notcounts
      gmark = gmark - notcounts
  else
        gmark = 0
  end
   set(handles.text27,'string',gmark);
     filewrite = fopen('c:\files\marksoutout.dat', 'a');
     fprintf(filewrite,'%1.1f\r\n',gmark);
     fclose(filewrite)
     filewritein = fopen('c:\files\marksinput.dat', 'a');
     fprintf(filewritein,'%1.1f\t',arrayone');
     fprintf(filewritein,'\r\n');
     fclose(filewritein)
% the end of this function
```

This function is used to read the input (KW's weight) and output (actual marks) file

```
function pushbutton1_Callback(hObject, eventdata, handles)
[filename,pathname] = uigetfile('*.dat', 'Please select the Training input marks');
cd(pathname);
finp=fopen(filename,'r');
set(handles.text12,'string',filename);
% the below is for actual marks output
```

```
[filename,pathname] = uigetfile('*.dat','please select the actual output marks');
cd(pathname);
finp2=fopen(filename,'r');
set(handles.popupmenu2,'String',{'General Mamdani ','Back propogation
Sugeno','Subclustering sugeno'});
xx=1;
while finp > 2
    Nbr33{xx}=fgetl(finp2);
        if (Nbr33{xx}==-1)
            Nbr33(xx)=[];
        break
        end
    xx=xx+1;
end
 Nbr333=char(Nbr33);
inputs22=str2num(Nbr333);
set(handles.text14,'string',xx-1);% to print number of answers
x=1;
while finp > 2
    Nbr{x}=fgetl(finp);
        if (Nbr{x}==-1)
            Nbr(x)=[];
        break
        end
    x=x+1;
end
Nbr2=char(Nbr);
set(handles.listbox3,'string',inputs22);
inputs=str2num(Nbr2);
handles.inputmarks=inputs22;
handles.outputmarks=inputs;
guidata(hObject,handles);
% end of function
```

This function is used to select and to read the Fuzzy inference system and to determine which FIS will be used to score the essay

```
function popupmenu2_Callback(hObject, eventdata, handles)

% Determine the selected data set.
inputs22=handles.inputmarks;
inputs=handles.outputmarks;
val = get(hObject,'Value');
      % Set current data to the selected data set.
      switch val
     case 1 % User selects Mamdani model
        fismat=readfis('GenMamdani');% reading FIS
        predtr=round(evalfis(inputs,fismat));
        set(handles.listbox2,'String',predtr);
```

```
  actoutTR=inputs22;% output data train
  hold on
  grid
  plot(handles.axes1,predtr,'ko');
  plot(handles.axes1,actoutTR,'k*');
  grid
  hold off
  [corrTR,mapeTR,rmseTR]=eval_CorrMape(actoutTR,predtr)
  set(handles.text8,'string',corrTR);
  set(handles.text9,'string',mapeTR);
 set(handles.text10,'string',rmseTR);
case 2 % User selects Grid partion sugeno.
 fismat=readfis('sugenogrid');% reading FIS
  predtr=round(evalfis(inputs,fismat));
  set(handles.listbox2,'String',predtr);
  actoutTR=inputs22;% output data trainin
  hold on
  grid
  plot(handles.axes1,predtr,'ko');
 plot(handles.axes1,actoutTR,'k*');
  grid
  hold off
  [corrTR,mapeTR,rmseTR]=eval_CorrMape(actoutTR,predtr)
  set(handles.text8,'string',corrTR);
  set(handles.text9,'string',mapeTR);
 set(handles.text10,'string',rmseTR);
case 3 % User selects Sugeno subtractive.
  fismat=readfis('sugSubCluster');% reading FIS
  predtr=round(evalfis(inputs,fismat));
  set(handles.listbox2,'String',predtr);
  actoutTR=inputs22;% output data trainin
  hold on
grid
  plot(handles.axes1,predtr,'ko');
   plot(handles.axes1,actoutTR,'k*');
  grid
  hold off
  [corrTR,mapeTR,rmseTR]=eval_CorrMape(actoutTR,predtr)
  set(handles.text8,'string',corrTR);
  set(handles.text9,'string',mapeTR);
 set(handles.text10,'string',rmseTR);
  end
```

Appendix C: Data sets used and data from local universities

1. QOU data:

Define :Allegory

Typical answer:

In literature, symbolic story that serves as a disguised representation for meanings other than those indicated on the surface. The characters in an allegory often have no individual personality, but are embodiments of moral qualities and other abstractions. The allegory is closely related to the parable, fable, and metaphor, differing from them largely in intricacy and length. The medieval morality play *Everyman,* personifying such abstractions as Fellowship and Good Deeds, recounts the death journey of Everyman. John Bunyan's *Pilgrim's Progress,* a prose narrative, is an allegory of people spiritual salvation. Although allegory is still used by some authors, its popularity as a literary form has declined in favor of a more personal form of symbolic expression.

Std#1:

Allegory is a symbolic narrative in which the surface details imply a secondary meaning. Allegory often takes the form of a story in which the characters represent moral qualities. The most famous example in English is John Bunyan’s *Pilgrim’s Progress*, in which the name of the central character, Pilgrim, epitomizes the book’s allegorical nature. Kay Boyle’s story "Astronomer’s Wife" and Christina Rossetti’s poem "Up-Hill" both contain allegorical elements.

Std#2:

allegory: "A story or visual image with a second distinct meaning partially hidden behind its literal or visible meaning. In written narrative, allegory involves a continuous parallel between two (or more) levels of meaning in a story, so that its persons and events correspond to their equivalents in a system of ideas or a chain of events external to the tale."

std#3

Allegory- Prose or verse in which the objects, events or people are presented symbolically, so that the story conveys a meaning other than and deeper than the actual incident or characters described. Often, the form is used to teach a moral lesson.

Std#4:
Allegory: A metaphorical narrative in prose or verse in which characters and parts of the narrative usually represent moral or spiritual values.

Std#5
Allegory: the saying of one thing and meaning another. Sometimes this trope works by an extended metaphor ('the ship of state foundered on the rocks of inflation, only to be salvaged by the tugs of monetarist policy'). More usually it is used of a story or fable that

has a clear secondary meaning beneath its literal sense. Orwell's *Animal Farm*, for example, is assumed to have an allegorical sense.

Std#7
Allegory
a narrative in which the agents and action — and sometimes the setting as well — are contrived not only to make sense in themselves, but also to signify a second, correlated order of persons, things, concepts, or events

Std#8:

Allegory--a universal symbol or personified abstraction. Example: Death portrayed as a cloaked "grim reaper" with scythe and hourglass, or Justice depicted as a blindfolded figure with a sword and balances. Also a literary work or genre (e.g., John Bunyan's *Pilgrim's Progress*) that makes widespread use of such devices.

Std#9:
An object or scene that is associated with a certain event or time of year. (ie. Grapes are allegorical of autumn for that is when they are harvested.) Used also in mythology to symbolize a god, (ie. Grapes symbolize Bacchus etc)

Std#10:

Allegory is a kind of extended metaphor in which an entire book, poem, or story can signify something other than the actual story that is being told. Many works of literature have several levels of meaning. George Orwells' *Animal Farm*, for example, can easily be read as the problems of a group of farm animals, but it is also an allegory of the political and social changes that occurred in the Soviet Union. Many allegories are not so obvious.

Std#11:

allegory
A narrative or other literary piece that is highly symbolic, such that the characters and events in the story represent either abstract ideas or characters and events in some other setting . Sometimes it is difficult to determine whether the original author meant a particular narrative to be an allegory, so that an “allegorical” interpretation is a possibility but not a certainty.

Std#12:

the representation of abstract ideas or principles by characters, figures, or events in narrative, dramatic, or pictorial form. A story, picture, or play employing such representation. John Bunyan's *Pilgrim's Progress* and Herman Melville's *Moby Dick* are allegories.

Std#13:
allegory, a story or visual image with a second distinct meaning partially hidden behind its literal or visible meaning. The principal technique of allegory is personification, whereby abstract qualities are given human shape—as in public statues of Liberty or Justice. An allegory may be conceived as a metaphor that is extended into a structured system. In

written narrative, allegory involves a continuous parallel between two (or more) levels of meaning in a story, so that its persons and events correspond to their equivalents in a system of ideas or a chain of events external to the tale: each character and episode in John Bunyan's *The Pilgrim's Progress* (1678), for example, embodies an idea within a pre existing Puritan doctrine of salvation.

Std#14:
Work of written, oral, or visual expression that uses symbolic figures, objects, and actions to convey truths or generalizations about human conduct or experience. It encompasses such forms as the fable and parable. Characters often personify abstract concepts or types, and the action of the narrative usually stands for something not explicitly stated. Symbolic allegories, in which characters may also have an identity apart from the message they convey, have frequently been used to represent political and historical situations and have long been popular as vehicles for satire.

Std#15:
Allegory in literature is the presentation of a subject under the guise of another suggestively similar. It was rarely written deliberately by the Greeks This kind of allegorizing was closely connected with etymology, at that time a pseudo-science which dealt in the ‘true’ meaning of words and names as revealed by assonances. Plato and the Alexandrian scholars rejected allegorical interpretation of literature. Its main proponents were the Stoic philosophers , who used it for the illustration and corroboration of their doctrines, and from them derive the surviving collections of allegorical interpretations of Homer.

Std#16:

Allegory is generally treated as a figure of rhetoric, but an allegory does not have to be expressed in language: it may be addressed to the eye, and is often found in realistic painting, sculpture or some other form of mimetic, or representative art.The etymological meaning of the word is broader than the common use of the word. Though it is similar to other rhetorical comparisons, an allegory is sustained longer and more fully in its details than a metaphor, and appeals to imagination, while an analogy appeals to reason or logic. The fable or parable is a short allegory with one definite moral.

Std#17:

A story that has a deeper or more general meaning in addition to its surface meaning. Allegories are composed of several symbols or metaphors. For example, in The Pilgrim's Progress, by John Bunyan, the character named Christian struggles to escape from a bog or swamp. The story of his difficulty is a symbol of the difficulty of leading a good life in the “bog” of this world. The “bog” is a metaphor or symbol of life's hardships and distractions. Similarly, when Christian loses a heavy pack that he has been carrying on his back, this symbolizes his freedom from the weight of sin that he has been carrying.

Std#18:
A figurative illustration of truths or generalizations about human conduct or experience in a narrative or description by the use of symbolic fictional figures and actions which resemble the subject's properties and circumstances.

2. PPU data set 1

What is the programming?

Student1
The process of writing, testing, debugging, and maintaining the source code of computer programs. This source code is written in a programming language .The code may be a modification of an existing source or something completely new. The purpose of programming is to create a program that exhibits a certain desired behavior (customization).(a set of instructions telling the computer what to do also referred as software).

Student2
Is the language of communication between man and machine language consists of numerous orders to implement a specific task What is not a simple process in which to write some words for the computer to implement as you wrote a letter to a friend and send him on the phone . And there are also amusing that definition: "Programming is stupid to learn how a friend you have to replace them as Person”.

Student3
Computer program is a set of instructions for a computer to perform a specific task. Programs generally fall into these categories applications, utilities or services. Programming is planning how to solve a problem. No matter what method is used pencil and paper, slide rule, adding machine, or computer -problem solving requires programming. Of course, how one programs depends on the device one uses in problem solving."

Student4
The act of creating software or some other set of instructions for a computer, is the process of writing a sequence of instructions to be executed by a computer to solve a problem. It is also considered as the act of writing computer programs. Computer programs are set of instructions that tell a computer to perform certain operations

Student5
An introduction to the terminology of computing plus some history and a brief look at the structure of a computer program. Programming is a creative process done by programmers to instruct a computer on how to do a task. Hollywood has helped instill an image of programmers as uber techies who can sit down at a computer and break any password in seconds or make highly tuned warp engines improve performance by 500% with just one tweak. Sadly the reality is far less interesting! A vocabulary and set of grammatical rules for instructing a computer to perform specific tasks.

Student6
Computer programming (often shortened to programming or coding), sometimes considered a branch of applied mathematics, is the process of writing, testing, debugging/troubleshooting, and maintaining the source code of computer programs. This source code is written in a programming language. The code may be a modification of an existing source or something completely new. The purpose of programming is to create a program that exhibits a certain desired behavior (customization). The process of writing source code requires expertise in many

different subjects, including knowledge of the application domain, specialized algorithms and formal logic.

Student8

is the art of making a computer do what you want it to do. or Programming is instructing a computer to do something for you with the help of a programming language. The role of a programming language can be described in two ways: Technical: It is a means for instructing a Computer to perform Tasks Conceptual: It is a framework within which we organize our ideas about things and processes

Student9

Computer Programming is the art of making a computer do what you want it to do Programming is a creative process done by programmers to instruct a computer on how to do a task. Hollywood has helped instill an image of programmers as uber techies who can sit down at a computer and break any password in seconds or make highly tuned warp engines improve performance by 500% with just one tweak. Sadly the reality is far less interesting!

Student10

a sequence of steps for each steps an arithmetic or logical operation is done for each operation a different set of control signals is needed. A basic command. The term instruction is often used to describe the most rudimentary programming commands

Student11

Computer Programming is the art of making a computer do what you want it to do.
At the very simplest level it consists of issuing a sequence of commands to a computer to achieve an objective. In the Microsoft world MS DOS users used to create text files with lists of commands called BAT files. These simply executed the sequence of commands as a BATCH, hence the name. You can still produce these in Windows environments today but in practice they are rarely seen. Programming is instructing a computer to do something for you with the help of a programming language.

PPU data set 2

Define the Operating System?

Student1

is the software component of a computer system that is responsible for the management and coordination of activities and the sharing of the resources of the computer. The operating system acts as a host for application programs that are run on the machine. As a host, one of the purposes of an operating system is to handle the details of the operation of the hardware .This relieves application programs from having to manage these details and makes it easier to write applications.

Student2

An operating system is the set of programs that controls a computer. Some examples of operating systems are UNIX, Mach, MS-DOS, MS-Windows, Windows/NT, OS/2, MacOS . Some of the operating system functions are: implementing the user interface, sharing hardware among users, allowing users to share data among themselves, preventing users from interfering with one another, scheduling resources among users, facilitating input/output, recovering from errors, accounting for resource usage, facilitating parallel operations, organizing data for secure and rapid access, and handling network communications.

Student3

An Operating System is a software program or set of programs that mediate access between physical devices (such as a keyboard, mouse, monitor, network connection) and application programs. a software program that provides the interface between a human and a computer. Software that shares a computer system's resources (processor, memory, disk space, network bandwidth, and so on) between users and the application programs they run. Controls access to the system to provide security

Student4
An Operating System is a computer program that manages the resources of a computer. It accepts keyboard or mouse inputs from users and displays the results of the actions and allows the user to run applications, or communicate with other computers via networked connections.

Student5
An Operating System is a computer program that manages the resources of a computer. It accepts keyboard or mouse inputs from users and displays the results of the actions and allows the user to run applications, or communicate with other computers via networked connections.

Student6

Software designed to handle basic elements of computer operation, such as sending instructions to hardware devices like disk drives and computer screens, and allocating system resources such as memory to different software applications being run. Given uniformly designed operating systems that run on many different computers, developers of software do not need to concern themselves with these problems, and are provided with a standard platform for new programs

Student8

is the software component of a computer system that is responsible for the management and coordination of activities and the sharing of the resources of the computer. The operating system acts as a host for application programs that are run on the machine. As a host, one of the purposes of an operating system is to handle the details of the operation of the hardware.

Student 9

An operating system is the infrastructure software component of a computer system; it is responsible for the management and coordination of activities and the sharing of the limited resources of the computer. The operating system acts as a host for applications that are run on the machine. As a host, one of the purposes of an operating system is to handle the details of the operation of the hardware. This relieves application programs from having to manage these details and makes it easier to write applications. Almost all computers, including handheld computers, desktop computers, supercomputers, and even video game consoles, use an operating system of some type. Some of the oldest models may however use an embedded operating system, that may be contained on a compact disk or other data storage device.

Student10
The software that manages hardware and resources on a computer. Applications use the operating system to make requests for services and interact with the computer's devices. A collection of software written to provide the fundamental instructions that a computer needs to manage resources, such as memory, the file system, and processes.

Student11

Software that controls the allocation and usage of hardware resources such as memory, CPU time, disk space, and peripheral devices. The operating system is the foundation upon which the applications are built. Popular operating systems include Windows, Mac OS and UNIX.

Student12

The master record control program that provides an interface for a user to communicate with the computer, manages hardware devices, manages and maintains disk file systems, and supports application programs.

Student13

The set of programs that control a computer system by controlling the execution of other programs and the use of resources such as disk space. System that consists of several programs that help the computer manage its own resources, such as manipulating files, running programs and controlling the keyboard and screen.
Student14
The operating system provides a set of interfaces that enables you to run applications on various hardware configurations. It contains the microkernel, board support package and higher level services, including I/O and file system support.

Student15

An operating system (abbreviated as 'OS') is the program that manages all the other programs in a computer.Refers to the code that operates a computer by managing its file systems, handling user input and output, and running programs. DOS, Windows, and UNIX are all operating systems.
Student16

Software that shares a computer system's resources (processor, memory, disk space, network bandwidth, and so on) between users and the application programs they run. Controls access to the system to provide security.

Student17

Windows and UNIX are the most popular operating systems which allow programs to run on a computer. Every general-purpose computer must have an operating system to run other programs. Operating systems perform basic tasks, such as recognizing input from the keyboard, sending output to the monitor, keeping track of files and directories on the disk, and controlling peripheral devices such as disk drives, printers, and scanners.

Student 18

A series of computer programs which control the operation of the computer itself. Application programs such as GIS software run under an operating system. Examples of operating systems include MS Windows 95, MS Windows NT, UNIX, VMS, DOS and OS/2.

Student 19

The operating system is the most basic program in a computer. All computers have an operating system that among other things is used for starting the computer and running other programs (application programs). The operating system performs important tasks like receiving input from the keyboard and mouse, sending information to the screen, keeping track of files and directories on the disk, as well as controlling the various units such as disks printers' etc-. An operating system also offers a user interface, giving the user the possibility to control the computer. Examples of operating systems are: Windows95/98, Windows NT/2000, Novell Netware, Mac OS, UNIX, Linux.

Student20

Computer program that allows users to enter and run their software packages. The operating system allows the machine to recognize and carry out the accountant's command. Further, there are built-in routines permitting the user's software to conduct input-output operations without specifying the exact hardware configuration. The operating system normally consists of the job control program, the input/output control system, and the processing program. If a computer operates under one system, it cannot use programs designated for a different operating system

Student21

The foundation software of a machine; that which schedules tasks, allocates storage, and presents a default interface to the user between applications. The facilities an operating system provides and its general design philosophy exert an extremely strong influence on programming style and on the technical cultures that grow up around its host machines.

Student22

Software that controls the operation of a computer, directs the input and output of data, keeps track of files, and controls the processing of computer programs. Its roles include managing the functioning of the computer hardware, running the applications programs, serving as an interface between the computer and the user, and allocating computer resources to various functions. When several jobs reside in the computer simultaneously and share resources (multitasking), the OS allocates fixed amounts of CPU time and memory in turn or allows one job to read data while another writes to a printer and still another performs computations. Through a process called time-sharing, a large computer can handle interaction with hundreds of users simultaneously, giving each the perception of being the sole user.

Student23

An operating system (commonly abbreviated *OS* and *O/S*) is the infrastructure software component of a computer system; it is responsible for the management and coordination of activities and the sharing of the limited resources of the computer. The operating system acts as a host for applications that are run on the machine. As a host, one of the purposes of an operating system is to handle the details of the operation of the hardware. This relieves application programs from having to manage these details and makes it easier to write applications. Almost all computers, including handheld computers, desktop computers, supercomputers, and even video game consoles, use an operating system of some type. Some of the oldest models may however use an embedded operating system, that may be contained on a compact disk or other data storage device.

More dataset:
Define United Nations UN

A1.1) The United Nations (UN) is an international organization whose stated aims are to facilitate cooperation in international law, international security, economic development, peace, and social progress and human rights issues. The United Nations was founded in 1945 to replace the League of Nations, in the hope that it would intervene in conflicts between nations and thereby avoid war. There are now 192 United Nations member states, including almost every recognized independent state. From its headquarters in New York City, Additional bodies deal with the governance of all other UN System agencies, such as the World Health Organization (WHO) and United Nations Children's Fund (UNICEF).

A1.2). United Nations (UN): Established in 1945, the United Nations is tasked with maintaining peace and stability in the world through cooperation among its member countries. The UN maintains peacekeeping and humanitarian missions around the world, as well as serving as a venue for mediation of international disputes and coordination of global initiatives on many issues.

A1.3).United Nations (U.N.) - international organization established in 1945. The U.N. supports cooperation among nations and the peaceful settlement of debates. The United States is one of the U.N.'s 183 member states.

A1.4) United Nations: The United Nations officially came into existence on 24 October 1945. The purposes of the United Nations, as set forth in the Charter, are to maintain international peace and security; to develop friendly relations among nations; to cooperate in solving international economic, social, cultural and humanitarian problems and in promoting respect for human rights and fundamental freedoms; and to be a centre for harmonizing the actions of nations in attaining these ends.

A1.5) The United Nations comprises 191 member countries (nearly every country in the world) and works toward maintaining international peace and security and finding solutions for global economic and humanitarian problems. The UN conducts its mission through six main organs, the Economic and Social Council, the General Assembly, the International Court of Justice, the Secretariat, the Security Council, and the Trusteeship Council. The United Nations was founded in 1945 when the UN Charter was ratified by China, France, the Soviet Union, the UK, the US, and a majority of other signatory countries. The UN is headquartered in New York and has offices in Geneva and Vienna.

A1.7). An international organization of countries set up in 1945, to promote international peace, security, and cooperation. Its members, originally the countries that fought against the Axis Powers in World War II, now number more than 150 and include most sovereign states of the world, the chief exceptions being Switzerland and North and South Korea. Administration is by a secretariat headed by the Secretary General. The chief deliberative body is the General Assembly. The Security Council bears the primary responsibility for the maintenance of peace and security. The headquarters of the United Nations are in New York.

A1.8).International organization founded (1945) at the end of World War II to maintain international peace and security, develop friendly relations among nations on equal terms, and encourage international cooperation in solving intractable human problems.

A1.9). The United Nations (U.N.) is an organization of 185 states that strives to attain international peace and security, promotes fundamental human rights and equal rights for men and women, and encourages social progress. By October 24, 1945, China, France, the United States, the Soviet Union, the United Kingdom, and a majority of the charter's other signatories had ratified it, and the United Nations was officially established. Shortly thereafter the U.S. Congress unanimously invited the United Nations to set up headquarters in the United States, and the organization chose New York City as its permanent home.

A1.10) An organization that includes virtually all countries in the world, with nearly 190 member nations. The Security Council is charged with solving crises and keeping peace. The United Nations also includes an Economic and Social Council; a Secretariat, or administrative division; and the International Court of Justice. World Court

Q2)Arab League

A2.1) Regional organization formed in 1945 and based in Cairo. It initially comprised Egypt, Syria, Lebanon, Iraq, Transjordan (now Jordan), Saudi Arabia, and Yemen; joining later were Libya, Sudan, Tunisia, Morocco, Kuwait, Algeria, Bahrain, Oman, Qatar, the United Arab Emirates, Mauritania, Somalia, the Palestine Liberation Organization, Djibouti, and Comoros. The league's original aims were to strengthen and coordinate political, cultural, economic, and social programs and to mediate disputes; a later aim was to coordinate military defense. Members have often split on political issues; Egypt was suspended for 10 years (1979 – 89) following its peace with Israel.

A2.2) popular name for the League of Arab States, formed in 1945 in an attempt to give political expression to the Arab nations. The original charter members were Egypt, Syria, Lebanon, Transjordan (now Jordan), Iraq, Saudi Arabia, and Yemen. A representative of Palestinian Arabs, although he did not sign the charter because he represented no recognized government, was given full status and a vote in the Arab League. The Palestine Liberation Organization (PLO) was granted full membership in 1976. Other current members include Algeria, Bahrain, Comoros, Djibouti, Eritrea (pending in 1999), Kuwait, Libya, Mauritania, Morocco, Oman, Qatar, Somalia, Sudan, Tunisia, and the United Arab Emirates.

A2.3) The League of Arab States, also known as the Arab League, is composed of twenty-two independent Arab states that have signed the Pact of the League of Arab States. Palestine, represented by the Palestinian Authority, is included as an independent state. The multipurpose League of Arab States seeks to promote Arab interests in general, but especially economic and security interests. It also works to resolve disputes among members and between member states and nonmember states. It has the image of unity in the protection of Arab independence and sovereignty. It promotes political, military, economic, social, cultural, and developmental cooperation among its members

A2.4) organization of Arab states in Southwest Asia, and North and Northeast Africa. It was formed in Cairo on March 22, 1945 with six members: Egypt, Iraq, Transjordan (renamed Jordan after 1946), Lebanon, Saudi Arabia, and Syria. Yemen joined as a member on May 5, 1945. The Arab League currently has 22 members. The Arab League is involved in political, economic, cultural, and social programs designed to promote the interests of its member states.

Q3)Palestine Liberation Organization:

A3.1) Umbrella political organization representing the Palestinian people in their drive for a Palestinian state. It was formed in 1964 to centralize the leadership of various groups. In 1969 Yasir 'Arafat, leader of Fatah, the PLO's largest faction, became its chairman. From the late 1960s the PLO engaged in guerrilla attacks on Israel from bases in Jordan, from which it was expelled in 1971. PLO headquarters moved to Lebanon. In 1974 'Arafat advocated limiting PLO activity to direct attacks against Israel, and the Arab community recognized the PLO as the sole legitimate representative of all Palestinians. It was admitted to the Arab League in 1976. In 1993 Israel recognized the PLO by signing an agreement with it granting Palestinian self-rule in parts of the West Bank and Gaza Strip.

A3.2) (PLO), coordinating council for Palestinian organizations, founded (1964) by Egypt and the Arab League and initially controlled by Egypt. Composed of various guerrilla groups and political factions, the PLO is dominated by Al Fatah, the largest group, whose leader, Yasir Arafat, was chairman of the PLO from 1969 to 2004 and established Palestinian control over the organization. Other groups in the PLO include the Syrian-backed As Saiqa and the Marxist-oriented Popular Front for the Liberation of Palestine (PFLP).

A3.3) A government that is recognized by the United Nations as the body that represents the people of Palestine displaced by the establishment of Israel. The PLO has long been led by Yasir Arafat and, at least officially, is committed to a compromise by which Israel would exchange portions of the Occupied Territories, areas including the Gaza Strip and the West Bank that Israel took from the Arabs in the Six-Day War. Arafat currently heads the Palestinian Authority, which has gained limited self-government over parts of the

Occupied Territories. He has been opposed by radical Arabs, especially by Hamas, an organization opposed to compromise with Israel. Israeli opinion on whether the PLO can be trusted to discharge its side of any compromise settlement is deeply divided.

A3.4) Founded by a meeting of 422 Palestinian national figures in Jerusalem in May 1964 following an earlier decision of the Arab League, its goal was the liberation of Palestine through armed struggle. These are often taken to mean the destruction of Israel. It also called for a right of return and self-determination for Palestinians. Palestinian statehood was not mentioned, although in 1974 the PLO called for an independent state in the territory of Mandate Palestine. In 1988, the PLO officially adopted a two-state solution, with Israel and Palestine living side by side contingent on specific terms such as making East Jerusalem capital of the Palestinian state and giving Palestinians right of return.

A3.5) Palestine Liberation Organization, PLO (a political movement uniting Palestinian Arabs in an effort to create an independent state of Palestine; when formed in 1964 it was a an organization dominated by Yasser Arafat's al-Fatah; in 1968 Arafat became chairman; received recognition by the United Nations and by Arab states in 1974 as a government in exile; has played a largely political role since the creation of the Palestine National Authority).

A3.6) Palestine Liberation Organisation. The umbrella group of political and militant groups headed by Yasser Arafat, whose Fatah faction is the biggest single component. Fatah in turn is subdivided into various groups. One of the most recent to emerge is the semi-autonomous Tanzim movement of young street fighters.

Q4)Palestine:

A4.1) A historical region of southwest Asia at the eastern end of the Mediterranean Sea Occupied since prehistoric times, it has been ruled by Hebrews, Egyptians, Romans, Byzantines, Arabs, and Turks. A British League of Nations mandate oversaw the affairs of the area from 1920 until 1948, when Israel declared itself a separate state and the West Bank territory was occupied by Jordan. The West Bank was subsequently annexed (1950) by Jordan and occupied (1967) by Israel. In 1988 the Palestine Liberation Organization under Yasir Arafat declared its intention of forming an Arab state of Palestine, probably including the West Bank, the Gaza Strip, and the Arab sector of Jerusalem.

A4.2) Region, at the eastern end of the Mediterranean Sea. It extends east to the Jordan River, north to the border between Israel and Lebanon, west to the Mediterranean, and south to the Negev desert, reaching the Gulf of Aqaba. A land of sharp contrasts, Palestine includes the Dead Sea, the lowest natural point of elevation on Earth, and mountain peaks higher than 2,000 ft (610 m) above sea level. It was governed by Britain under a League of Nations mandate from the end of World War I (1914 – 18) until 1948, when the State of Israel was proclaimed. Armies from Egypt, Transjordan, Syria, and Iraq attacked the next day.

A4.3) historic region on the eastern shore of the Mediterranean Sea, at various times comprising parts of modern Israel, the West Bank and Gaza, Jordan, and Egypt; also known as the Holy Land. The name is derived from a word meaning “land of the

Philistines.” This article discusses mainly the geography and the history of Palestine until the United Nations took up the Palestine problem in 1947; for the economy and later history.

A4.4) Palestine is a name which has been widely used since Roman times to refer to the region between the Mediterranean Sea and the Jordan River. In its broader meaning as a geographical term, Palestine can refer to an area that includes contemporary Israel and the Palestinian territories, parts of Jordan, and parts of Lebanon and Syria In its narrow meaning, it refers to the area within the boundaries of the former British Mandate of Palestine (1920-1948) west of the Jordan River.

A4.5): Palestine-- Land in the Middle East acquired by the British Empire after the defeat of the Ottoman Empire which, in 1947, was partitioned by the UN between Arabs and Jews, and which provoked an invasion by five Arab Armies in 1948. The defeat of these armies by the Jews led to the establishment of the State of Israel.

A4.6) Palestine, or the Holy Land. Corner of the Mediterranean, is bounded on the north by Lebanon, on the East. by the Jordan Valley, on the south by the Sinai Desert, and on the West by the sea; there is great diversity of climate throughout its extent owing to the great diversity of level, and its flora and fauna are of corresponding range; it suffered much during the wars between the Eastern monarchies and Egypt, and in the wars between the Crescent and the Cross, and is now by a strange fate in the hands of the Turk; it has in recent times been the theater of extensive exploring operations in the interest of its early history.

Q5. Security Council:

A5.1) A body of the United Nations tasked with keeping international peace. Located at U.N. headquarters in New York City, it was originally comprised of eleven members with five permanent members representing China, France, the Soviet Union, the United Kingdom, and the United States, and six nonpermanent representatives. In 1965 the body was amended to a fifteen-member council, composed of the same five permanent members and ten nonpermanent members. On substantive issues all five permanent members must be included in the affirmative vote, unless a member abstains, and permanent members have veto power. The Council may advise U.N. members to seek diplomatic or economic sanctions, and military action by U.N. forces may follow if sanctions prove inadequate.

A5.2)Cabinet of the United Nations Organization. It originally consisted of eleven members, expanded in 1965 to fifteen, of whom five (Britain, China, France, Russia, United States) were permanent members, the rest being elected by the General Assembly for a two-year period. In 1991 Russia was awarded the Soviet seat. The Security Council exercises primary responsibility within the UN for the maintenance of international peace and security. It can act only with the agreement of the five permanent members who exercise a veto; the lack of agreement on most issues throughout the Cold War severely restricted the role of the Security Council although since the late 1980s it has enjoyed a much more active role.

A5.3) Division of the United Nations whose primary purpose is to maintain international peace and security. The Security Council originally consisted of five permanent members China (represented by the government on Taiwan until 1971), France, the United Kingdom, the U.S., and the Soviet Union (succeeded in 1991 by Russia) and six rotating members elected by the United Nations General Assembly for two-year terms. In 1965 the number of nonpermanent members was increased to 10. UN members agree to abide by the Security Council's resolutions when they join. The Security Council investigates disputes that threaten international peace and advises on how to resolve them.

Q6) Jerusalem

A6.1) A city divided between east-central Israel and the Israeli-occupied West Bank. Jerusalem was founded as far back as the fourth millennium B.C. and was ruled by the Canaanites, Hebrews, Greeks, Romans, Persians, Arabs, Crusaders, Turks, and British before being divided in 1949 into eastern and western sectors under Israeli and Jordanian control. In 1967, Israeli forces captured the eastern sector from Jordan, later declaring the city as a whole to be the capital of the state of Israel. The legal status of Jerusalem, considered a holy city to Jews, Muslims, and Christians, remains fiercely disputed.

A6.2) City Located in the heart of historic Palestine, it is nestled between the West Bank and Israel. The Old City is a typical walled Middle Eastern enclosure; the modern city is an urban agglomeration of high-rises and housing complexes. It is holy to Judaism as the site of the Temple of Jerusalem, to Christianity because of its association with Jesus, and to Islam because of its connection with the Mi'raj (the Prophet Muhammad's ascension to heaven). Jewish shrines include the Western Wall. Islamic holy places include the Dome of the Rock. In 1000 BC David made it the capital of Israel. Razed by the Babylonians in the 6th century BC, it thereafter enjoyed only brief periods of independence. The Romans devastated it in the 1st and 2nd centuries AD, banishing the Jewish population.

A6.3) *Al Quds*, city (1994 pop. 578,800). It is situated on a ridge 2,500 ft (760 m) high that lies west of the Dead Sea and the Jordan River. Jerusalem is an administrative, religious, educational, cultural, and market center. Tourism and the construction of houses and hotels are the city's major industries. Manufactures include cut and polished diamonds, plastics, clothing, and shoes, and electronic printing and other high-technology industries have been developed. The city is served by road, rail, and air transport.

A6.4) Its largest city in both population and area, with a population of 747,600 residents over an area of 125. square kilometres (48.3 sq mi) if disputed East Jerusalem is included Located in the Judean Mountains, between the Mediterranean Sea and the northern tip of the Dead Sea, modern Jerusalem has grown up outside the Old City. The city has a history that goes back to the 4th millennium BCE, making it one of the oldest cities in the world.

Q7) Dome of the Rock

A7.1)Oldest existing Islamic monument. It is located on Temple Mount, previously the site of the Temple of Jerusalem. The rock over which it is built is sacred to both Muslims and Jews. In Islam, Muhammad is believed to have ascended into heaven from the site. In Judaism it is the site where Abraham prepared to sacrifice his son Isaac. Built in 685 – 91 as a place of pilgrimage, the octagonal building has richly decorated walls and a goldoverlaid dome mounted above a circle of piers and columns.

A7.2) The Dome of the Rock is an Islamic shrine and a major landmark located on the Haram al-Sharif in Jerusalem. It was completed in 691, making it the oldest extant Islamic building in the world. The Dome of the Rock is located at the visual center of an ancient man-made platform known as the Temple Mount (literally, the Mountain of the House) to the Jews and the Haram al-Sharif (Noble Sanctuary) to the Muslims.

Q8) Palestinian National Authority

A8.1) Palestinian National Authority, interim self-government body responsible for areas of the West Bank and Gaza Strip under Palestinian control. The PA was authorized by the Oslo Accords (1993) and subsequent Palestinian agreements with Israel, and was established in 1994. As now constituted the PA includes a president, prime minister and cabinet, a legislative council, and security forces. In 1994 Yasir Arafat, the leader of Fatah and the Palestine Liberation Organization, which was a party to the Oslo Accords, appointed an interim 19-member Palestinian National Authority, under his direction, to administer Palestinian affairs in the areas of self-rule.

A8.2)The September 1993 Oslo Accord between Israel and the Palestine Liberation Organization (PLO) called for the establishment of a Palestinian Interim Self-Governing Authority (PISGA) in those parts of the West Bank and Gaza from which Israeli forces would eventually withdraw. This concept was actualized by the subsequent May 1994 Gaza-Jericho Agreement that led to an initial Israeli withdrawal from those two areas and that created the Palestinian Authority (PA), which would exercise autonomous powers until an elected Palestinian council could replace it. PLO Chair Yasir Arafat and a body of PLO cadres were allowed to return from exile and form the 24-member council of ministers that made up the PA, which commenced functioning in July 1994

A8.3) A legal entity, established in 1994, by which Palestinians govern the parts of the Gaza Strip and West Bank that are not attached to Israel. Although divided on several issues, the leadership of the Palestinian Authority is committed to the establishment of a Palestinian state, which, it expects, will grow out of the Palestinian Authority

Q9) Six-Day War

A9.1) War between Israel and the Arab countries of Egypt, Syria, and Jordan. Palestinian guerrilla attacks on Israel from bases in Syria led to increased hostility between the two countries. Egypt answered by ordering the withdrawal of UN peacekeeping forces from the Sinai Peninsula and by moving troops into the area. Amid increasingly belligerent language from both sides, Egypt signed a mutual defense treaty with Jordan. Israel, surrounded and fearing an Arab attack was imminent, launched what it felt was a preemptive strike against the three Arab states on June 5, 1967. Israeli forces captured the

Sinai Peninsula, Gaza Strip, West Bank of the Jordan River, Old City of Jerusalem, and he Golan Heights. The status of these occupied territories subsequently became a major point of contention between the two sides.

A9.2) A war fought in 1967 by Israel on one side and Egypt, Syria, and Jordan on the other. Israel, victorious, took over the Golan Heights, the Jordanian portion of Jerusalem, the Jordanian West Bank of the Jordan River, and a large piece of territory in northeastern Egypt, including the Sinai Peninsula, which contains Mount Sinai. Israel still occupies all of these territories except the Sinai Peninsula, which it gave back to Egypt in 1982. Israel maintains that its security would be enormously endangered if it withdrew from the other places.

A9.3) The Six-Day, also known as the 1967 Arab-Israeli War, the Third Arab-Israeli War, Six Days' War, an-Naksah (The Setback), or the June War, was fought between Israel and Arab neighbors Egypt, Jordan, and Syria. The nations of Iraq, Saudi Arabia, Sudan, Tunisia, Morocco and Algeria also contributed troops and arms to the Arab forces.In May 1967, Egypt's president Nasser expelled the United Nations Emergency Force (UNEF) from the Sinai Peninsula. The peacekeeping force had been stationed there since 1957, following a British-French-Israeli invasion which was launched during the Suez Crisis. Egypt amassed 1,000 tanks and nearly 100,000 soldiers on the Israeli border and closed the Straits of Tiran to all ships flying Israeli flags or carrying strategic materials, receiving strong support from other Arab countries.

Q10) West Bank

A10.1) Area in Palestine, west of the Jordan River and east of Jerusalem. Covering an area of about 2,270 sq mi , excluding east Jerusalem. It is a region with deep history, forming the heart of historic Palestine. Populated areas include Nablus, Hebron, Bethlehem, and Jericho. Under a 1947 UN agreement, most of what is now the West Bank was to become part of a Palestinian state. When the State of Israel was formed, the Arabs attacked Israel , and the partition plan was never adopted. Following a truce, Jordan remained in control of the area and annexed it in 1950. Israel subsequently occupied it during the Six-Day War of 1967. During the 1970s and '80s Israel established settlements there.

A10.2) territory, formerly part of Palestine, after 1949 administered by Jordan, since 1967 largely occupied by Israel , 2,165 sq mi (5,607 sq km), west of the Jordan River, incorporating the northwest quadrant of the Dead Sea. Since mid-1994 limited Palestinian self-rule has existed in portions of the West Bank under the Palestinian Authority. Israelis who regard the area as properly Jewish territory often refer to it by the biblical names of Judaea and Samaria. The largest and most historically important cities are Hebron, Nablus, Bethlehem, and Jericho. East Jerusalem is regarded as part of the West Bank by Arabs.

A10.3) The West Bank refers to the territory situated west of the Jordan River that was not included as part of Israel following the establishment of the state after the Arab - Israel War of 1948. The West Bank's total area is 2,270 square miles (5,880 sq. km), smaller than the area that was originally allocated to a future Arab state by the United Nations partition resolution of November 1947. It is demarcated by the Green Line (the armistice line set by the 1949 Jordanian-Israeli talks at Rhodes) in the west and the Jordan River in the east.

Q11) *Intifada*

A11.1)Palestinian revolt against the Israeli occupation in the Gaza Strip and West Bank. Initially a spontaneous reaction to 20 years of occupation and worsening economic conditions, it was soon taken over by the Palestine Liberation Organization (PLO). Its tactics included strikes, boycotts, and confrontations with Israeli troops. The International Red Cross estimated that some 800 Palestinians, more than 200 under the age of 16, had been killed by Israeli security forces by 1990.. *Intifadah* pressure is credited with helping make possible the 1993 Israeli-PLO agreement on Palestinian self-rule. A breakdown in further negotiations in late 2000 led to another outburst of violence, which quickly became known as the Aqsa *intifadah*, named for the Aqsa Mosque in Jerusalem, where the fighting began.

A11.2) Intifada is Arabic for ‘a shaking off’. The term refers to the two Palestinian uprisings on the West Bank and Gaza (the territories occupied by Israel during the 1967 war). The first intifada arose spontaneously 1987 lasting until 1993 and the second began in 2000. It has been suggested that the emergence of the first intifada was a response to the realization that the Palestinian issue and the Arab-Israeli conflict was slipping as a key concern of Arab governments, and that Palestinians in the Occupied Territories would have to take matters into their own hands. The second intifada, known as the Al-Aqsa Intifada, erupted when Ariel Sharon visited the Al-Aqsa Mosque symbolizing a provocative violation of the holy site.

A11.3) The Palestinian uprising during the late 1980s and early 90s in the West Bank and Gaza Strip, areas that had been occupied by Israel since 1967. A vehicular accident that killed four Palestinians in the Gaza Strip in Dec., 1987, sparked immediate local protests that rapidly spread to the West Bank. The violence was marked by stone-throwing and the use of homemade explosive devices on behalf of the Arabs, and the use of tear gas, rubber bullets, and home demolition by Israeli troops attempting to quell the popular resistance. The conflict led to an Israeli military crackdown and the stagnation of the Arab economies in the occupied territories, but with the gradual establishment of Palestinian self-rule, beginning with the accord between Israel and the Palestine Liberation Organization in 1993, the violence eased significantly.

Q12) Hebron

A12.1) City in the West Bank, southwest of Jerusalem. It is a sacred city of Judaism and Islam as the home and burial place (at the Cave of Machpelah) of the patriarch Abraham. King David made Hebron his capital briefly in the 10th century BC. Except for a period of Crusader control in the 12th century, various Muslim dynasties ruled the city from AD 635 until after World War I (1914 – 18). It was part of the British mandate of Palestine from the early 1920s until the first Arab-Israeli war in 1948, when it came under the control of Transjordan (later Jordan). Along with the rest of the West Bank, it was annexed by Jordan in 1950 but was captured by Israel during the Six-Day War (1967). Bank cities.

A12.2) City (2003 est. pop. 155,000), the West Bank, called Al-Khalil in modern Arabic. Hebron is situated at an altitude of 3,000 ft (910 m) in a region where grapes, cereal grains, and vegetables are grown. Tanning, food processing, glassblowing, and the manufacture of sheepskin coats are the major industries. The city is also a road junction. Hebron has usually had a significant Jewish population, although following Arab riots in 1929 most

Jews left and did not return until after the Israeli occupation following the 1967 Arab-Israeli War, when numerous Jewish settlements were established outside Hebron. One of Judaism's four holy cities, Hebron is also a sacred place for Muslims

A12.3) City of the West Bank south-southwest of Jerusalem. Sacred to both Jews and Arabs as the home and burial place of Abraham and as King David's capital for seven years, the city has figured in every war in Palestine and has a history of Jewish-Arab violence. Occupied by Israel in 1967, Hebron came under Palestinian control in 1996. Population: 160,000.

A12.4) Hebron (in Arabic, *al-Khalil;* in Hebrew, *Hevron*) is an ancient city, holy to both Judaism and Islam, because it is the site of the Machpelah burial cave of the Biblical and Quranic figures Abraham, Isaac, and Jacob, and their respective wives Sarah, Rebekah, and Leah. Later, in the tenth century B.C.E., David was proclaimed king in Hebron when Saul died, and it became his first capital. Above the Machpelah cave is a mosque complex known as the al-Haram al-Ibrahimi.

Q13) Amnesty International

A13.1) International human-rights organization. It was founded in 1961 by Peter Benenson, a London lawyer who organized a letter-writing campaign calling for amnesty for "prisoners of conscience." AI seeks to inform the public about violations of human rights, especially abridgments of freedom of speech and religion and the imprisonment and torture of political dissidents. It actively seeks the release of political prisoners and support of their families when necessary. Its members and supporters are said to number one million people in some 140 countries. Its first director, Sean MacBride, won the 1974 Nobel Prize for Peace; AI itself won the award in 1977.

A13.2) With more than 40 years work behind it, Amnesty International strives to promote human rights around the world. It has nearly 2 million members, chapters in more than 60 countries, and supporters and donors from more than 100 countries. Having won the Nobel Peace Prize in 1977, the organization continues to campaign against such things as torture, the death penalty, and other human rights violations. British lawyer Peter Benenson, who died in 2005, founded Amnesty International as a letter-writing campaign in 1961 as a reaction to the incarceration of two Portuguese students who had toasted to freedom.

A13.3) AI human-rights organization founded in 1961 by Englishman Peter Benenson; it campaigns internationally against the detention of prisoners of conscience, for the fair trial of political prisoners, to abolish the death penalty and torture of prisoners, and to end extrajudicial executions and “disappearances” throughout the world. It was awarded the Nobel Peace Prize in 1977 for having aided in the release of more than 10,000 political prisoners worldwide. In 1998, the organization had over a million members worldwide.

A13.4) Amnesty International (commonly known as Amnesty or AI) is a UK-based international non-governmental organization which defines its mission as "to conduct research and generate action to prevent and end grave abuses of human rights and to demand justice for those whose rights have been violated." Founded in London in 1961, AI draws its attention to human rights abuses and campaigns for compliance with international standards. It works to mobilizes public opinion which exerts pressure on individuals who perpetrate abuses. The organization was awarded the 1977 Nobel Peace Prize for its "campaign against torture" and the UN Human Rights Prize in 1978, but has received criticism for both alleged anti-Western and alleged pro-Western bias

التصحيح التلقائي للأسئلة الإنشائية

إعداد: حمزة عبد الحميد مجاهد مجاهد

إشراف: الدكتور لبيب عرفة

ملخص:

قام الباحث في هذه الدراسة بتطوير نظام تصحيح تلقائي للمقالات الإنشائية . تتضمن الفكرة الرئيسية في النظام المقترح ، تصحيح المقالة عن طريق تحديد الكلمات الرئيسة في المقالة والمرادفات لكل كلمة رئيسة، والتي يحددها المعلم ، و يتم معالجة الكلمات الرئيسية ومرادفاتها باستخدام طرق الحوسبة البرمجية (Soft computing techniques) و بشكل ادق, تم استخدام نماذج تعتمد على المنطق الضبابي (Fuzzy Logic), و مزيج الخلايا العصبية و المنطق الضبابي (Neuro-Fuzzy Logic) . لقد تم تطوير ثلاثة نماذج تستطيع بدورها التصحيح التلقائي للمقالات الإنشائية و هي كما يلي: النموذج الأول هو " مامداني" (Mamdani) متعدد المدخلات مفرد المخرجات، و النموذج الثاني هو "سوجينو" (Sugeno) باستخدام تقنية التغذية الراجعة (Back propagation) , اما النموذج الاخير فهو "سوجينو" متعدد المدخلات مفرد المخرجات باستخدام التكتل الطرحي " Subtractive clustering" و التقنيات المركبة أو المهجنة(Hybrid technique). تستطيع هذه النماذج المطورة قراءة المقالة (النص) المرسلة و تحديد الكلمات الرئيسية وأربع مرادفات لها و إعطاء الوزن المناسب لكل كلمة بحيث يكون المدخل للنظام هو أوزان الكلمات و المخرجات من النظام هو العلامة المتنبأة . تم جمع 1080 إجابة لثلاثة عشر سؤالا تتعلق بموضوع التاريخ تم إنشاؤها باستخدام نظام مولد إجابات تلقائي و بعد ذلك تم تقسيم البيانات إلى قسمين باستخدام طريقة Cross-validation، قسم من البيانات يستخدم في تدريب أو تعليم النماذج (Learning) وعددها 718 إجابة وأخرى لفحص النتائج (Testing) وعددها 362 إجابة.

للتحقق من مدى ملاءمة وصلاحية النماذج، تم فحصها باستخدام معامل درجة التوافق (Correlation Coefficient) لقياس مدى تطابق العلامات الحقيقة مع المتنبأة, و كما تم استخدام طريقتين لحساب الخطأ هي متوسط نسبة القيمة المطلقة "ماب" (MAPE, Mean Absolute Percentage Error), ومتوسط مربع الجذر "ارمس" (RMSE, Root Mean Square Error), و أشارت النتائج الأولية إلى أن درجة التوافق بين العلامات الحقيقية و المتنبأة في النموذج الأول تتراوح ما بين 0.887 و 0.9969 بمعدل 0.9863 للإجابات المتعلقة بتدريب النظام و 0.9599 للأجوبة المتعلقة بفحص النظام. أما بالنسبة لقيم ماب فقد كانت تتراوح ما بين 0.017 و 0.1619 بمعدل 0.0854 للأجوبة الخاصة بالتدريب و 0.1189 للبيانات المتعلقة بفحص النظام. أخيرا النتائج المتعلقة بال " ارمس" كانت القيم تتراوح ما بين 0.0378 و 0.2947 بمعدل 0.087 للاجوبة المتعلقة بتدريب النظام و 0.2073 للأجوبة المتعلقة بفحص النظام.

بالنسبة للنموذج الثاني فقد كانت درجة التوافق بين العلامات الحقيفية و المتنبأة تتراوح ما بين 0.9347 و 0.9966 بمعدل 0.9906 للإجابات المتعلقة بتدريب النظام و 0.9675 للأجوبة المتعلقة بفحص النظام. أما بالنسبة لقيم ماب فقد كانت تتراوح ما بين 0.015 و 0.399 بمعدل 0.1039 للأجوبة الخاصة بالتدريب و 0.1071 للبيانات المتعلقة

بفحص النظام.أخيرا النتائج المتعلقة بال " ارمس" كانت القيم تتراوح ما بين 0.043 و 0.3775 بمعدل 0.0809 للاجوبة المتعلقة بتدريب النظام و 0.2069 للأجوبة المتعلقة بفحص النظام.

أما فيما يتعلق بالنموذج الأخير, سوجينو باستخدام التكتل الطرحي (subtractive clustering) فقد كانت النتـائج هي الافضل بين النماذج الاخرى حيث كانت درجة التوافق بين العلامات الحقيقية و المتنبأة تتراوح ما بين 0.9121 و 0.9977 بمعدل 0.9948 للإجابات المتعلقة بتدريب النظام و 0.9712 للأجوبة المتعلقة بفحـص النظـام. أمـا بالنسبة لقيم ماب فقد كانت تتراوح ما بين 0.015 و 0.1516 بمعدل 0.0399 للأجوبـة الخاصـة بالتـدريب و 0.1024 للبيانات المتعلقة بفحص النظام.أخيرا النتائج المتعلقة بال " ارمس" كانت القيم تتراوح ما بـين 0.0328 و 0.2763 بمعدل 0.0526 للأجوبة المتعلقة بتدريب النظام و 0.1837 للأجوبة المتعلقة بفحص النظـام. النتـائج الأولية هي نتائج واعدة وتظهر مدى قدرة النظام على تصحيح المقالات الإنشائية بدرجة توافق عالية.

لقد تم ترتيب النماذج التي تم بناؤها في تطبيق قائم بذاته (Stand Alone Application) ذي واجهـة رسـومية للمستخدم ليمثل نظاماً لتصحيح المقالات الإنشائية بحيث يستطيع المستخدم لهذا النظام من إدخال الكلمات الرئيـسية (Main Keywords) و مرادفات كل كلمة رئيسية , و بعد ذلك تتم عملية التصحيح التلقائي مـن خـلال قـراءة الأجوبة الإنشائية النصية باستخدام الواجهة الرسومية بطريقة سهلة و مرنة.

لمزيد من التحقق لمدى قدرة هذه النماذج على تصحيح المقالات قمنا بالحصول على بيانات من بعض الجامعات المحلية من اجل اختبارها على النظام. لقد حصلنا على عينة (حوالي 18 إجابة إنشائية) لسؤال يتعلق بموضوع الأدب الانجليزي من جامعة القدس المفتوحة, وقد كانت درجة التوافق لهذه العينة 0.9954 . و لقد قمنا أيضا بالاتصال بجامعة بوليتكنك فلسطين , و لقد حصلنا على عينتين لسؤالين اثنين(11 عينة من الأجوبة للسؤال الأول و 23 عينة من الأجوبة للسؤال الثاني) لموضوع تكنولوجيا المعلومات ، و كانت درجة التوافق بين العلامات الحقيقة والمتنبأة لأجوبة السؤال الأول تساوي 0.9755, في حين كانت نتائج التوافق لأجوبة السؤال الثاني تساوي 0.9745 , و كانت قيمة ماب تساوي 0.0696 و قيمة ارمس تساوي 0.266 . لقد كانت نسبة الفرق بين العلامة الحقيقية و المتنبأة تتراوح ما بين 0% إلى 8.3% أي بمعدل اختلاف بالعلامة يتراوح بين 0 و 2.5 علامة و تعتبر هذه النتائج واعدة .

إن نتائج التوافق التي حصلنا عليها من كلتا الجامعتين تعد نتائج واعدة, وأظهرت درجة توافق عالية بين العلامات الحقيقة والمتنبأة, وهذا يدل على مدى ملاءمة وكفاية هذه النماذج لتصحيح المقالات الإنشائية. و يقترح الباحث استخدام أسئلة من مجالات أخرى (علمية مثلا) و أيضا مزيدا من التحسينات على هذه النماذج و ذلك من خلال تطوير معرب لغوي (Language parser).