A Model for Evaluating Numerical through Computer

Anurag Saxena, School of Management, IGNOU, INDIA

Abstract
In recent times the education world has seen the emergence of two efforts at using computers to evaluate the content of essays. The reason behind the emergence of these systems was the difficulty in managing the load of grading the essays and the tendency of the evaluators to shift the focus of assessment to multiple-choice questions. These systems generated a lot of enthusiasm as they not only save time but also provide consistency a computer can deliver.

Present communication tries to emphasize on the age-old concept of "marking schemes". It tries to build its arguments around two theories of learning. Firstly, structural learning theory which postulates that structural analysis is a methodology for identifying the rules to be learned for a given topic and breaking them down into their atomic components and secondly, script theory, which focused on the structure of knowledge and conceptual dependency theory. The key element of conceptual dependency theory is the idea that all the conceptualizations can be represented in terms of a small number of primitive acts performed.

Present paper attempts at building a system that will use the accuracy and consistency of a computer to achieve robust and unbiased student assessment.

INTRODUCTION

Technology has made it possible to think beyond multiple-choice objective test. Restrictions in including only right/wrong kind questions in online evaluation systems are no more in place. A beginning has been made to evaluate essays through computer. The pre-requisite is: The student has to key his own answer in the computer. In 1998, a professor of Psychology at University of Colorado at Boulder completed work on a software after a sustained effort of 10 years. The software shifted its focus on the content rather than writing and found out that these two are highly correlated. As a teacher one has to feed all the intelligence into the computer from the textbooks and other sources. The computer in turn forms relationships between the words. The teacher is expected to give computer a set of model answers that are graded by the human evaluators, which would eventually provide the computer a basis for comparison. It is not as simple as it looks but then once one is through one would be able to evaluate thousands of essays with the click of a button.

ELECTRONIC MANAGEMENT OF ESSAY TYPE QUESTIONS

Two approaches of evaluating essays by computer are discussed to show how these software works:

Intelligent Essay Assessor (IEA)

Intelligent Essay Assessor (IEA) is based on a statistical approach for analyzing the essays and content information. This statistical approach is called Latent Semantic Analysis, which is used primarily to extract semantic content of the text. Many scientists have proved that Latent semantic analysis captures the similarity of meanings expressed. This technique is successful in giving judgments very close to the human judgments. According to the developers of the IEA, the semantic analysis focuses on the conceptual content of the essay i.e. on the correctness and completeness of the content, the soundness of the arguments and the fluency and comprehensibility of the writing. This package basically permits the students to log on to the web and submit their essay to get a feedback on the missing points in their essays. After going through the feedback students can revise their essays and resubmit the work. The main benefit of this process is that on one hand the software is performing the function of summative evaluation and on the other hand it is also doing the function of formative evaluation.
E-rater is used by the Educational Testing Service (ETS) for the Graduate Management Admissions Test (GMAT) for testing the analytical writing assessments. E-rater is used as a second rater replacing the second human rater and thus reducing huge costs on evaluation. This system has been developed after 5 years of rigorous efforts in using the advanced computational linguistics techniques. Apart from GMAT test e-rater was also applied on Test of English as a Foreign Language (TOEFL) test. E-rater follows the scoring guide used by expert human evaluators for scoring. This scoring guide has a six-point scoring scale. It checks the essay for the argument structure, syntactic structure and vocabulary structure. The software is based on three general classes of features: Syntactic, rhetorical and topical content features. The features are extracted from the essay texts and quantified using computational linguistics techniques.

METHODOLOGIES OF ESSAY EVALUATORS

The two efforts in evaluating the essays through computer have added a new dimension to student assessment. Since the theories (Burstein, 1998; Foltz, 1996) behind these methodologies are a new in the educational setup and a little statistical in nature, an overview of them is given below:

Latent Semantic Analysis (LSA)

Latent Semantic Analysis (LSA) is a statistical model of word usage that permits comparisons of the semantic similarity between pieces of textual information. The method generates a depiction of words that are used in similar contexts and are more semantically associated. LSA generates a matrix of occurrences of each word in each essay and then decomposes these matrices into a set of hundreds of factors. Since the number of factors will be much smaller than the number of unique words, words will not be independent. If two terms are used in same context, they will have similar factors. Matching is then done between two pieces of textual information, even if they have no words in common.

Syntactic Structure Analysis

Syntactic variety is an important feature in evaluating essays. It defines the ratio and quantity of types of sentences, types of clauses and use of verbs. This analysis parses through each sentence in the essay and quantifies these features. This parsing is done on the basis of Microsoft's Natural Language Processing tool.

Rhetorical Structure Analysis

This analysis quantifies the evidence of organization of the essay. It quantifies the cue words and other structures of the essay. If an examinee is able to demonstrate that the essay he developed organizes the ideas logically and has connected them well, then he is ought to get a good score.

Topical Content Analysis

A good essay resembles other good essays in its vocabulary use patterns. This analysis evaluates the topical content of an essay by comparing it with the other model essays that are graded by the human evaluators. It uses two different measures of content similarity. First one is based on the vocabulary use in the essay as a whole and other is based on specific vocabulary content of the argument found in the essay.

VALIDITY OF AUTOMATED ESSAY SCORING

It has been proved that in assessments where stakes are high, automated essay scoring systems sometime fail to provide accurate assessments. In a study on the validity of automated essay scoring systems like e-rater, Powers et al (2001) compared the scores given by two expert human raters with the score of e-rater. These essays were written by experts called especially to contribute “typical” essays capable of achieving extreme scores by using “tricks”. Powers et al (2001) found that “Expert writers were more successful in tricking e-rater into assigning scores that were too high than in duping e-rater into awarding scores that
were too low”. Hearst (2000) asserted that scoring short answers is more challenging than scoring essays as they provide less textual evidence of the writer’s underlying meaning than essays. These findings showed that howsoever good these approaches are there is some element missing in them with respect to human evaluations. On the other side, human evaluations have their own shortcomings. It is therefore prudent to have a infallible system that overcomes these difficulties and is as unbiased and consistent as a computer is.

**A NEW APPROACH FOR EVALUATING NUMERICAL**

Let us first describe the two theories on which we will build our model.

**Script theory**

This theory was proposed by R Schank (1977, 1982, 1986). It is mainly concerned with the structure of knowledge. It postulated, “All conceptualizations can be represented in terms of a small number of primitive acts performed”. Memory is taken episode by episode which is called a “script”. These scripts facilitate inferences to be made by filling the missing information. As a script is divided into various schemes and various statements, a numerical can also be divided into steps. Each step can be assigned a script indicating what is to be filled in order to achieve the correct answer.

Mueller (2004) proposed use of commonsense reasoning to understand texts involving scripts. For example, “WalkUpStaircase(actor, staircase)” is an event signifying actor walks up staircase or “BuildingOf(room) = building” is a function meaning Building of room is building(Mueller, 2004). One can relate this with artificial intelligence that is given to a computer program. For example in MS-Excel the formulae are of the form COUNTIF(B2:C16,”k”) meaning count for all cells from B2 to C16 if the value of the cell is equal to “k”. The meanings of these formulae are interpreted in exactly the same manner as it is done in the earlier examples of scripts. Our intention of using script theory for evaluation is based on this property. We have divided the numerical into smaller scripts that are to be fed by the student in order to achieve the solution of the numerical. Computer can identify these scripts and thus evaluate them by assigning the awards to different steps (or a group of scripts).

**Structural Learning Theory**

J. Scandura (1977, 1980, 1984, 2001) proposed this theory. According to this theory, “Structural analysis is a methodology for identifying the rules to be learned for a given topic or class of tasks and breaking them into their atomic components”. According to structural learning theory a problem and its solution is identified first. The solution is then broken into atomic components. Each component represents a problem in itself. Solutions are found for newly created problems. In this manner one can get many higher order problems and their solutions. Iterative process of replacing lower order solutions with higher order solutions is continued to get more and more solutions.

Scandura (1977) as quoted in Kearsley (2005) gave an example in context of subtraction in which he illustrated the tasks in which subtraction can be learned:

- Recognize the digits 0-9, minus sign, columns and rows
- Learn the "borrowing" procedure that specifies if the top number is less than the bottom number in a column, the top number in the column to the right must be made smaller by 1.
- Replace a number of partial rules with a single rule for borrowing that covers all cases.
- Use problems with varying combinations of columns and perhaps different bases.

For example if one has taken a numerical that is to be evaluated by the computer. Using structural learning theory one can device a problem in such a way that it tests the learner from the point of view of simple tasks and then take the evaluation to difficult tasks. This on one hand will set the difficulty level of the evaluation as per the average learner and on the other hand pose problems that will only be attended by above average learners.

So the basic approach in our model is that we will divide the problem into atomic components and call them tasks. Each task will be associated with some marks and these tasks will be
structured from lower order to higher order problems. For each response of the student to the problems given in these tasks, computer will expect a script. If the script, which, computer expects, matches with the answer given by the student then marks are awarded otherwise not. To input these answers in the computer one can use Intelligent Character Readers (ICR) or better still take the test in an environment where the student himself keys the answer in the system.

**AN EXAMPLE TO HIGHLIGHT THE MODEL**

Let us see the evaluation of a problem from trigonometry given below with the help of the theories given above.

**Problem**

Given: A+B+C= 0x01 graphic
, Prove that cos(2A)+cos(2B)+cos(2C)+1= - 4cos(A)cos(B)cos(C)

**Solution**

LHS= cos(2A)+cos(2B)+cos(2C)+1
= 2cos(A+B)cos(A-B)+2cos2 (C) but A+B= -C so cos(A+B)= -cos(C)
= -2cos(C)cos(A-B)+2cos2(C)
= -2cos(C).(cos(A-B)-cos(C))
= -2cos(C).cos(A-B)+cos(A+B))
= -2cos(C).2cos(A)cos(B)
= -4cos(A)cos(B)cos(C)= RHS

This is a typical problem of trigonometrical identity. In this problem the learner is expected to use the formulae, use a trick from the information given in the question twice and apply formulae once again to reach the desired result. The learner has to use one set of rule till half way and then he has to repeat them in the opposite way. Let us see how a learner applies his knowledge to solve this. The “instructional activities” of the learner would be complete if and only if, he gets the capability of achieving the correct answer and that too by successfully traversing all the steps. Now let us assume that the question is of 10 marks. Here, we assign marks to the student if he successfully traverses a particular step.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Order</th>
<th>Hints that trigger action</th>
<th>Marks</th>
<th>Script</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Apply cos(C) + cos (D)</td>
<td>Low</td>
<td>The RHS has multiplication signs</td>
<td>2</td>
<td>PLUSCOS (A, B)</td>
</tr>
<tr>
<td>2 Apply the expansion of cos(2C)</td>
<td>Low</td>
<td>Since “1” is not there in RHS, answer should have “-1” in it</td>
<td>2</td>
<td>EXPCOS (2C)</td>
</tr>
<tr>
<td>3 Apply A+B+C</td>
<td>High</td>
<td>Why is it given in the question?</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>4 Substitute cos</td>
<td>High</td>
<td>Since there is a “-“ in the answer it should be “-“ and since there is a “cos” in the answer it should be “cos”</td>
<td>Same as step 2</td>
<td>PHICOS (C)</td>
</tr>
<tr>
<td>5 Take cos(C) common</td>
<td>Low</td>
<td>Take out whatever part of answer I can!</td>
<td>1</td>
<td>COMMONCOS (C)</td>
</tr>
<tr>
<td>6 Revert to cos(A+B)</td>
<td>High</td>
<td>If I can write A+B</td>
<td>2</td>
<td>PHICOS (A+B)</td>
</tr>
<tr>
<td>7 Know cos(C) + cos (D)</td>
<td>Low</td>
<td>Answer contain “cos” only</td>
<td>1</td>
<td>PLUSCOS (A-B, A+B)</td>
</tr>
</tbody>
</table>

Let us see how the computer will expect the student to answer this numerical.
LHS = \cos(2A)+\cos(2B)+\cos(2C)+1
= \text{PLUSCOS}(A, B)+\text{EXPCOS}(2C)
= \text{PHICOS}(C)\cos(A-B)+2\cos(2C)
= \text{COMMONCOS}(C)(\cos(A-B)\cos(C))
= \text{COMMONCOS}(C)(\cos(A-B)\cdot\text{PHICOS}(A+B))
= \text{COMMONCOS}(C).\text{PLUSCOS}(A-B, A+B)
= -4\cos(A)\cos(B)\cos(C)=\text{RHS}

The Solution will remain in the computer memory as above and it will match this with the student answer. The marks will be assigned in the manner in which it was depicted in the table given above. The problem of the sequencing of these steps now has to be thought over. For this consider the scripts used in the answer of this problem. There are four main scripts used in the answer. These are for performing the function of “plus”, “phi minus angle”, “expand” and “common” of cosine functions. There is a set space of sequential steps in which these scripts are to be performed. Thus it will be possible for the computer to identify these scripts in their appropriate place and award marks with the help of a suitable program.

CONCLUSION

We have tried to build our model on the basis of Script theory and Structural Learning theory and demonstrated the model through an example. This approach can certainly be extended to all problems that are mathematical in nature and also to numerical in other subjects. One can see that this approach has lot of similarities with the age-old concept of marking schemes where marks are awarded for each step. This model has implications for providing feedback too. Identifying the error that the student has made in a particular script and providing a correct version of that script to the student can do this.

Apart from this one can use this model for teaching also. Say if a student is well versed with the four scripts then there is no reason why he would not be able to solve the problem given in the example. Learning will be fun as for a given topic there will be a few scripts that will have to be mastered and what will be important here is that student makes an application of that script.

REFERENCES


Figures

Phi (a symbol)

π

Back to Papers