AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

A Similarity Comparison Model for Automated Assessment of Answers

[1]Madhumitha Ramamoorthy and [2]Dr. Ilango Krishnamurthi

[1]*Sri Krishna College of Engineering and Technology, Department of CSE, Madhumitha Ramamoorthy, Coimbatore-641008, India*
[2]*Sri Krishna College of Engineering and Technology, Department of CSE, Dr.Ilango Krishnamurthi, Coimbatore-641008, India*

ARTICLE INFO

Article history:
Received 19 September 2014
Received in revised form
19 November 2014
Accepted 29 November 2014
Available online 27 December 2014

Keywords:
Artificial Intelligence, Automatic assessment, Evaluation, N-gram similarity, cosine similarity.

ABSTRACT

Background: Automation is one of the areas in Artificial Intelligence where most of the research activities are going on. Automatic assessment of students' short and big answers is an example of an automated system. Many similarity measures help us to evaluate the similarity between the sentences. This paper proposes an algorithm for evaluating short and big answers of the students' using the existing semantic similarity measure i.e., cosine similarity and compares it with N-gram similarity. This paper evaluates the answers which belong to the information recall and comprehensive question categories. The comparison shows that the cosine similarity measure outperforms the n-gram similarity measure i.e., for information recall questions, the cosine similarity outperforms by 8.3% and for comprehension questions, the cosine similarity performs by 30.7%.

To Cite This Article: Madhumitha Ramamoorthy and Dr. Ilango Krishnamurthi, A Similarity Comparison Model for Automated Assessment of Answers. ***Aust. J. Basic & Appl. Sci.,*** *8(18): 648-656, 2014*

INTRODUCTION

Artificial intelligence deals with creating machines to exhibit human intelligence so that the machines works like humans. There are many applications of Artificial Intelligence which includes robotics, natural language understanding, expert systems, automation, speech recognition, game playing, learning, reasoning, etc. Among these areas automation is one important area where many research works takes place. The main advantage of automation is time management, reducing human labors, security and efficiency. Automated assessment of students' answers is one of the important researches going on in the educational sector. Many works have been done in automatic assessment of student's short answers and essays**.**

Computer Assisted Assessment (CAA) uses a computer in assessing the students' answers. The main advantage of this type of automatic evaluation is that it reduces teachers' work and makes them to concentrate on more teaching and research. This makes the teachers to concentrate on an individual student who is weak and help him/her to learn well and this helps the teachers to better understand the students' thinking process. This helps the teacher to learn in-depth the concepts and teach innovatively to the students'. This type of automatic evaluation will be more authentic since humans are not involved in evaluating task. The universities employ a number of staffs to evaluate the answer scripts of students. In CAA based evaluation many numbers of systems are employed to evaluate answer scripts. The morning session evaluation of a human will not be same as evening session evaluation. But the computer will not get tired, bored, irritated while correcting the answer scripts. All the answers will be assessed in an equal manner. The computers can analyze the answers in a detailed manner than humans. CAA also helps the students to get instant feedback immediately after they write the exams. The main disadvantage of this CAA is that the teachers cannot find the progress of the student whether he/she has written correctly or where the mistake is since the teachers are not reading the students' answers.

The assessment is divided into objective and subjective assessment. The objective assessment (www.selfgrowth.com) is the most common one because of the online exams. Multiple choice questions, true/false questions, fill in the blank questions and matching questions belong to objective assessment whereas short answers and long essay assessments belong to subjective assessment. In subjective assessment short answers have been concentrated more and there are a lot of approaches for assessment. But evaluating big answers is not focused fully.

There are different categories of questions to be evaluated by the human. Those categories of questions are

- Information Recall Questions

Corresponding Author: Madhumitha Ramamoorthy, Sri Krishna College of Engineering and Technology, Department of CSE, Madhumitha Ramamoorthy, Coimbatore-641008, India

Australian Journal of Basic and Applied Sciences, 8(18) December 2014, Pages: 648-656

- Comprehension Questions
- Application Questions
- Analysis Questions
- Synthesis Questions
- Evaluation Questions

Information Recall questions requires to recognize or recall the information that they have studied. The students are not asked to use the information, but asked to remember the information they know. To answer an information recall question, the students must remember facts, observations and definitions that they have studied before.

Comprehension questions make the students to write in their own words by organizing what they have learnt. The students will use the recalled or remembered information, but they have to show that they have the understood the recalled information by writing in their own words and also they should be able to write comparisons on the topics they have recalled or remembered.

Application questions make the students to apply or use a rule, theorem or definition that they have learned and apply them to solve the problem to find the correct answer. This is widely used in mathematics or science.

Analysis questions make the students to think and answer the questions asked by analyzing what they have learned and also they should give valid reasons to support their analysis. These types of questions help the students in a) identifying the reasons and causes for an occurrence, b) considering the information to come to a conclusion, c) finding evidence to support or refuse a conclusion.

Synthesis questions make the students to think creatively and innovatively to answer the questions. They have to produce their own explanations for making predictions and for solving the problems.

Evaluation questions make the students to judge their idea by explaining why one is better than the other and what criteria they have used to give judgments.

All these types of questions can be evaluated by the human. The Automated Assessment of answers should try to make the computers to evaluate these types of questions and their answers. The proposed model evaluates Information recall type questions and comprehensive questions automatically.

The rest of the paper is organized as follows. Section 2 outlines the related work. Section 3 describes the proposed System architecture. Section 4 gives the Data Sets used in this paper. Section 5 shows the Experimental Results and Discussion and Section 6 gives the Conclusion.

Related work:

Many works have been focused on the N-gram algorithm and cosine similarity algorithm for evaluating short answers.

N-gram algorithm:

In (K. Papineni, *et al*., 2001) the number of n-grams is counted from the text to be corrected and whether that text appears in any of the reference text. The frequency of each n-gram with which it is found appearing in any reference and combines the marks for each value of n.

In (P.Selvi, Dr.A.K.Bnerjee ,2010) the n-gram method is used to compute the score based on word-to-word match between students' and key answer. If more key answers (reference answers) are present similarity matching is done for each key answer with the student answer independently, then the best scoring pair is used to find the final score. This method uses three modules for matching the key answer and student answer. If the surface forms of the unigrams are matched it uses the exact module to score. If the two unigrams matches after preprocessing, the stemming module is used to score them. Then the heuristic rule based module is used to map two unigrams based on some heuristic rules like WordNet synonym match, numeric value matches, acronym match, derivational form match and country adjectival form/demonym match. This method is mainly used to score definitions, descriptions, yes/no questions and advantages/disadvantages questions.

In (Diana P´erez, *et al*., 2004) some parameters of the original Bleu algorithm is modified by the length of the maximum n-gram to look for coincidences in the key answer (reference text) and also modify the number of reference texts used. This method also works for definitions, descriptions, yes/no questions and advantages/disadvantages questions and checks the results for unigrams, bigrams, trigrams and four grams. The results were good until trigrams and four grams have no correlation.

In (Enrique Alfonseca and Diana P´erez, 2004) the original Bleu algorithm is modified by adding stemming, stop word removal and word sense disambiguation. WSD is done to find the sense of the words written in student answer and key answer. It uses SEMCOR CORPUS , the word sense of each word is found by using Wordnet. The student answer is sent to the Bleu algorithm where unigram co-occurrence metric is found and checks if two unigrams match. This system also finds the syntactic dependencies between the words, for which it uses a parser to extract the dependencies between the words. But the parser does not support prepositional phrase or co- reference resolution. This method improves the original algorithm but WSD has no proper algorithm, if present the results would be better.

Cosine Similarity:

In (Michael Mohler and Rada Mihalcea, 2009) knowledge based measures and corpus based measures are used to evaluate short answers written by students'. The cosine similarity is used to evaluate the correct answer and student answer.

In (Sachin Saxena *et al*.) [2] (Abigail R. Razon,*et al*.,) (Sargur Srihari , *et al*., 2008) the training answers are tokenized, lowercased and sorted alphabetically and using this, a matrix is constructed for the training and test essays. The rows of the matrix form the terms in reference and student answer and the column consists of documents in the training and test set. The matrix is filled by updating the number of times a term found in the document and a technique called singular value decomposition is done to decompose the matrix into three new matrixes. After this, dimensional reduction is performed on this matrix (decomposed matrix). The resulting matrixes after reducing the dimension are the new training vectors which forms the k-dimensional semantic space. The column vector of the test set matrix will project into the reduced dimensional space. Finally the cosine similarity between each test document vector and all training document vector is calculated and a score is assigned to test document which has the highest similarity to any one of the training document.

The disadvantage of most of the existing methods is that they use the training essays and many reference answers for evaluating the essays and answers. The method used in this paper uses only one key answer to automatically assess the students' answer.

System Architecture for Automated Assessment:

The system architecture given in Fig.1 of E-assessment system consists of Key answer, Student answer, matching model which uses cosine similarity measure for evaluating the student and key answer. The Scoring Model gives a score by comparing the student and key answer. The key answer for the given question is collected from the teachers and the student answer is compared with the key answer based on the cosine similarity measure and scores are allocated to the answers of the student based on the key points present in the answer. For evaluating the answer, first the key answer which is in the text file is loaded in the e-assessment system and then the student answer which is stored in the text file is also loaded in the e-assessment system and then it is scored.

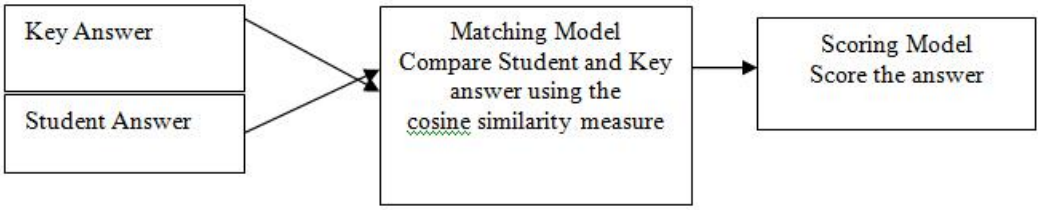

Fig. 1: System Architecture

Main Pseudo Code:

If fully scored return the total score
Else if not scored
Get the key point from the key answer
Search for that key point in the Student answer
If Student Answer! =null
Update the Score using Eq.1
Otherwise
Return the no Score
Else
Give the score fixed for that key point i.e returns the score for that key point
[Repeat for all key points]
Return the total score for that answer.

Key Answer and Student Answer:

The key answer for the question to be evaluated is prepared and stored in a text file. The key answer is stored in the form of key points. The important points in each of the questions to be evaluated is taken and stored in the text file. The student answer which is to be evaluated is also stored in a text file. The student answer is also considered to be written as points.

Matching Model:

The matching module is used to compare the key answer and student answer by using the cosine similarity measure. The key points written in the student answer are compared with the key point written in the key

Australian Journal of Basic and Applied Sciences, 8(18) December 2014, Pages: 648-656

answer. The first key point in the key answer is compared with all the key points in the student answer. Similarly each key point in the key answer is compared with all the key points in the student answer. Finally the scoring model gives the score (which is explained in the scoring model). The cosine similarity measure is described below:

Cosine Similarity:

Cosine Similarity measure finds the similarity between two vectors by measuring the cosine of the angle between the two vectors. Cosine similarity is represented by using a dot product and magnitude as

$$Cosine Similarity = \frac{A \bullet B}{\| A \| \| B \|} \quad (1)$$

Scoring Model:

The scores associated with each key point are stored in key answer text file. The total score is given by matching each key point in the key answer with all the key points in the student answer i.e., the first key point in the key answer will be matched with all the key points in the student answer and the maximum matched sentence will return the highest score. If the marks allotted for each key point is 1, and after evaluation using the cosine similarity measure, if it gets a similarity score more than 0.5, then the allotted mark is given for that key point. This procedure is repeated for all the key points and the total score is given for the answer.

Algorithm 3.3.1

Score Pseudo Code:

Read the Key Answer
If Key Answer! =null
    Check the Similarity using the cosine similarity measure using Eq.1
    If the Similarity Value>0.5 for each key point
        Return the assigned mark to the key point
        Else
        Return 0
Otherwise
Exit

Data Sets:

The dataset used for E-assessment of Answers in this paper is taken from the students' answer scripts written in an educational institution in Coimbatore, TamilNadu. The key answer is the teacher key points for the questions which are used for manual evaluation.

Data Sets for E-assessment of Answers - 2 marks question:

The data sets for E-assessment of Answers were also taken from the answer scripts of students written in an educational institution. The answers with different mark category were taken and compared with the two similarity measures. The datasets given below in the table belongs to 2 marks category questions and it falls in the information recall questions. 'Define', 'Recall', Recognize', 'What', 'Where', etc types of questions belong to Information Recall questions. This paper takes two subcategories of questions like 'Define' and 'What' to evaluate the answers. The questions considered for evaluation are 'Define rational agent' and 'What is a device controller?' The key answers to both the questions were prepared and they are compared with the students' answers of the respective questions. The key and the students' answers are shown in the Table 1 and Table 2.

Table 1: Sample Key and Student Answer for Question 1 -What is Device Controller?

Key Answer	Any I/O device connected to the CPU through a controller is called device controller.
	It controls the transfer of data from the computer to peripheral device and vice-versa.
S.No	Student Answers
Data Set 1	When an I/O is connected with a controller then it is called as device controller.
	The entire activity of that device is controlled by the device contoller.
Data Set 2	Device controller is a device which acts as an interface.
	It controls the transfer of data from the CPU and any peripheral I/O device.

Australian Journal of Basic and Applied Sciences, 8(18) December 2014, Pages: 648-656

Table 2: Sample Key and Student Answer for Question 2-Define Rational Agent.

Key Answer	For each possible percept sequence an ideal rational agent should select an action that is expected to maximize its performance measure given the evidence provided by percept sequence and built in knowledge agent has.
S.No	Student Answers
Data Set 3	An ideal rational agent is one which takes rational actions (i.e.,) right action always and it also has rational thinking.
Data Set 4	This agent should know which percept in the percept sequence will produce the optimal solution and choose that percept and always do right action.
Data Set 5	Ideal rational agent is an agent that does the right action and matches the set of percepts with the actions that gives the maximum possible outcome.
Data Set 6	It is the one that does the right actions even during uncertainties and maximizes its performance measure based on that input.

Data Sets for E-assessment of Answers - 13 marks question:

The data set given below in Table 3 is the sample key answer and Table 4 is the sample student answer written in an educational institution for a comprehensive question. Other datasets were also used for the evaluation of answers in the proposed E-assessment System.

Table 3: Sample Key Answer for Question 3.

S.No	Sentences
Sentence 1	Digital Computer is an electronic machine .It has the capability of high speed calculations.
Sentence 2	Computer can be used to solve problems by preparing and running a program.
Sentence 3	Computer program is a sequence of instructions.
Sentence 4	Each instruction specifies an operation to be performed.
Sentence 5	The functional blocks in a computer are
	1. ALU (Arithmetic and Logic Unit)
	2. CU(Control Unit)
	3. Memory
	4. Input Unit
	5. Output Unit
Sentence 6	The digital computer operates in the binary system. Data's inside the computer is in the binary form 0's and 1's.
Sentence 7	Instructions are also represented in 0's and 1's, such a program is called machine language program.
Sentence 8	Machine language are difficult from human so compilers are used to accept programs from high level language and translates into machine language.
Sentence 9	The input units are Keyboard, Floppy disk, magnetic tape, mouse, light pen, scanner, etc
Sentence 10	The output units are display terminal, printer, plotter, FDD(floppy disk drive), HDD(hard disk drive), magnetic tape drive, etc
Sentence 11	A Computer program consists of both the instructions and data. This program is fed into computer through input unit and stored in memory. The program consists of instructions and data.
Sentence 12	After fetching the instruction control unit issues control signals to the control unit.
Sentence 13	After the instructions are executed the result is stored in memory/temporarily in control unit and ALU, so that it can be used by the next instruction.

Table 4: Sample Student Answer for Question 3.

S.No	Sentences
Sentence 1	Computer is an electronic machine .It is used for high speed calculations.
Sentence 2	To solve a problem, programs are written and run.
Sentence 3	A program is a sequence of instructions.
Sentence 4	Instruction specifies the operations to be performed.
Sentence 5	The functional units in a computer are ALU , Control Unit, Memory, Input Unit , Output Unit
Sentence 6	The digital computer operates in the binary system. Data's inside the computer is in the binary form 0's and 1's.
Sentence 7	Instructions are also represented in 0's and 1's, such a program is called machine language program.
Sentence 8	Machine language are difficult from human so compilers are used to accept programs from high level language and translates into machine language.
Sentence 9	The input units are Keyboard, mouse, light pen, etc
Sentence 10	The output devices are display terminal, printer, plotter,
Sentence 11	The instructions are stored in memory and they are fetched by the control unit and is also decoded by the control unit. It is done by fetching the instructions from the memory one by one and this fetching is done by the control unit.
Sentence 12	According to the instruction control unit issues control signals to the control unit.
Sentence 13	The instruction is executed and the result is stored in memory/temporarily in control unit and ALU, so that it can be used by the next instruction.

Experimental results and discussion:

Comparison of key points in Key Vs Student Answer using cosine similarity and n-gram for Information Recall Questions for the Data Sets given in Table 1 and Table 2:

The key answer given Table 2 is compared with the two students answer in Data Set 1. The key answer in Table 2 consists of two key points namely k1-(a) and k1-(b). The Student answer also consists of two key points namely s1-(a) and s1-(b). The Table 5 shows the matrix form of evaluation of answers. The n-gram similarity values for the key points are given below in Table 5. When k1-(a) compared with s1-(a) and s1-(b), s1-(a) gives

Australian Journal of Basic and Applied Sciences, 8(18) December 2014, Pages: 648-656

the value more than 0.5, therefore the allotted mark for the key point i.e., 1 is given. But the second key point k1-(b), when compared to s1-(a) and s1-(b) gives n-gram similarity values less than 0.5, therefore allotted mark is not assigned to k1-(b) key point. Totally the Student answer i.e., Data Set 1 gets 1 mark out of 2 using n-gram similarity. Similarly the same Data set 1 is evaluated using cosine similarity measure and similarity values are given in Table 6. The total score for the Data Set 1 using cosine similarity measure is 1.

Table 5: N-gram Similarity Score for the answers in Table 1 for Data Set 1

Data Set 1	s1 -(a)	s1- (b)
K1 -(a)	0.51	0.3
K1 -(b)	0.08	0.11

Table 6: Cosine Similarity Score for the answers in Table 1 for Data Set 1.

Data Set 1	s1 -(a)	s1- (b)
K1 -(a)	0.59	0.2
K1 -(b)	0.07	0.25

All the data sets given in Table 1 and Table 2 are evaluated in the above matrix form and similarity values between the key points in the key answer vs. key points in the student answer for the information recall questions are obtained and are given in Table 7.

Table 7: Comparison of key points in Key Vs Student Answer using n-gram and cosine similarity for Information Recall Questions

Data Sets	Key Answer Key points Vs Student Answer Key points	N-gram Similarity Value	Cosine Similarity Value
Data Set 1	k1-(a) Vs s1-(a)	0.51	0.59
	k1-(a) Vs s1-(b)	0.3	0.32
	k1-(b) Vs s1-(a)	0.08	0.17
	k1-(b) Vs s1-(b)	0.11	0.25
Data Set 2	k1-(a) Vs s2-(a)	0.32	0.35
	k1-(a) Vs s2-(b)	0.16	0.2
	k1-(b) Vs s2-(a)	0.08	0.08
	k1-(b) Vs s2-(b)	0.4	0.6
Data Set 3	K1-(a) Vs s1-(a)	0.15	0.25
Data Set 4	K1-(a) Vs s2-(a)	0.23	0.32
Data Set 5	K1-(a) Vs s3-(a)	0.19	0.35
Data Set 6	K1-(a) Vs s4-(a)	0.1	0.25

The Fig.2 shows the comparison of key points in Key Vs Student Answer Using n-gram and cosine similarity for Information Recall Questions. The similarity values in Fig.2 show that the cosine similarity measure outperforms the n-gram similarity measure.

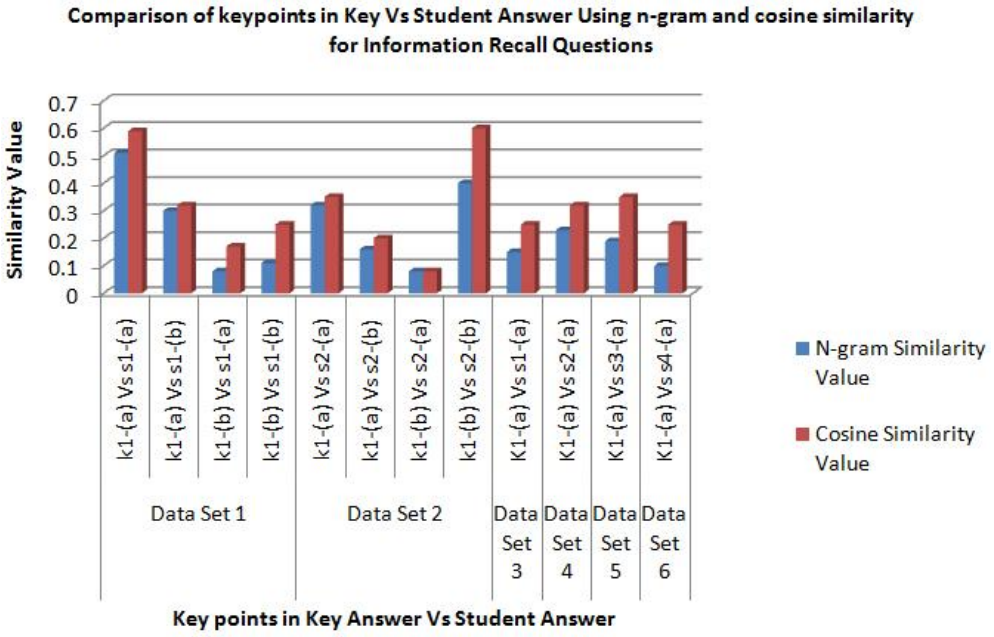

Fig. 2: Comparison of key points in Key Vs Student Answer using n-gram and cosine similarity for Information Recall Questions.

Australian Journal of Basic and Applied Sciences, 8(18) December 2014, Pages: 648-656

Comparison of key points in Key Vs Student Answer using n-gram and cosine similarity for Comprehensive Questions for the Data Sets given in Table 3 and Table 4:

The answers of the comprehensive question are compared using n-gram and cosine similarity measures which are given in Table 3 and Table 4. Table 3 shows the key answer and Table 4 shows the student answer. the similarity values for the sentences between key and student answer is given in the Table 8.

Table 8: Comparison of key points in Key Vs Student Answer using n-gram and cosine similarity for Comprehensive Questions.

S.No	Key Answer Key points Vs Student Answer Key points	N-Gram Similarity Value	Cosine Similarity Value
1	k1-(1) Vs s1-(1)	0.4634	0.7006
2	k1-(2) Vs s1-(2)	0.1637	0.2773
3	k1-(3) Vs s1-(3)	0.6611	0.8571
4	k1-(4) Vs s1-(4)	0.4523	0.5345
5	k1-(5) Vs s1-(5)	0.3777	0.5773
6	k1-(6) Vs s1-(6)	1	1
7	k1-(7) Vs s1-(7)	0.9909	1
8	k1-(8) Vs s1-(8)	1	1
9	k1-(9) Vs s1-(9)	0.3944	0.8017
10	k1-(10) Vs s1-(10)	0.2875	0.6666
11	k1-(11) Vs s1-(11)	0.5261	0.5786
12	k1-(12) Vs s1-(12)	1	0.8432
13	k1-(13) Vs s1-(13)	0.8914	0.8891

The Fig.3 shows the comparison of key points in Key Vs Student Answer using n-gram and cosine similarity for Comprehensive Questions. The similarity values in Fig.2 show that the cosine similarity measure outperforms the n-gram similarity measure.

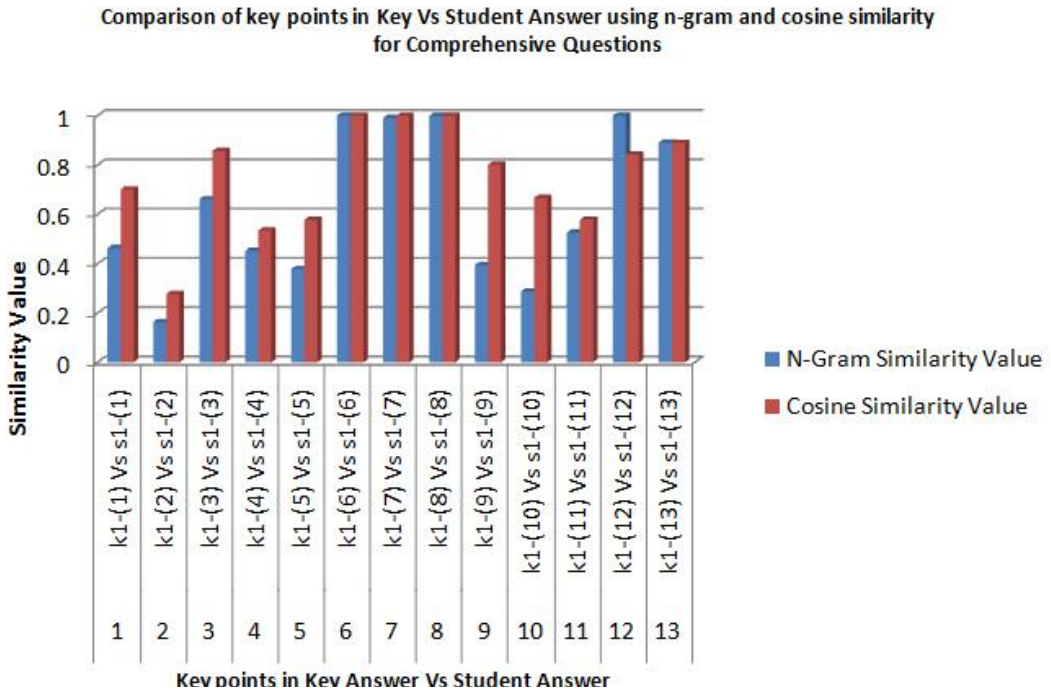

Fig. 3: Comparison of key points in Key Vs Student Answer using n-gram and cosine similarity for Comprehensive Questions

Comparison of the total score for the answers:

The similarity value of the separate key points using n-gram and cosine similarity measure was shown in Table 7 and Table 8. The total score for the answer as the whole for the data sets given above and also for the other datasets are given in the Table 9 and Fig 4 shows the performance comparison of the answers using n-gram and cosine similarity measure. The scores and the graph show that the cosine similarity measure outperforms the n-gram similarity measure.

Australian Journal of Basic and Applied Sciences, 8(18) December 2014, Pages: 648-656

Table 9: Comparison of the total score for the answers.

Data Sets	Marks Allotted for the question	N-gram Score	Cosine Score
Data Set 1	2	1	1
Data Set 2	2	0	1
Data Set 3	2	0	0
Data Set 4	2	0	0
Data Set 5	2	0	0
Data Set 6	2	0	0
Data Set 7	13	7	12
Data set 8	13	4	7

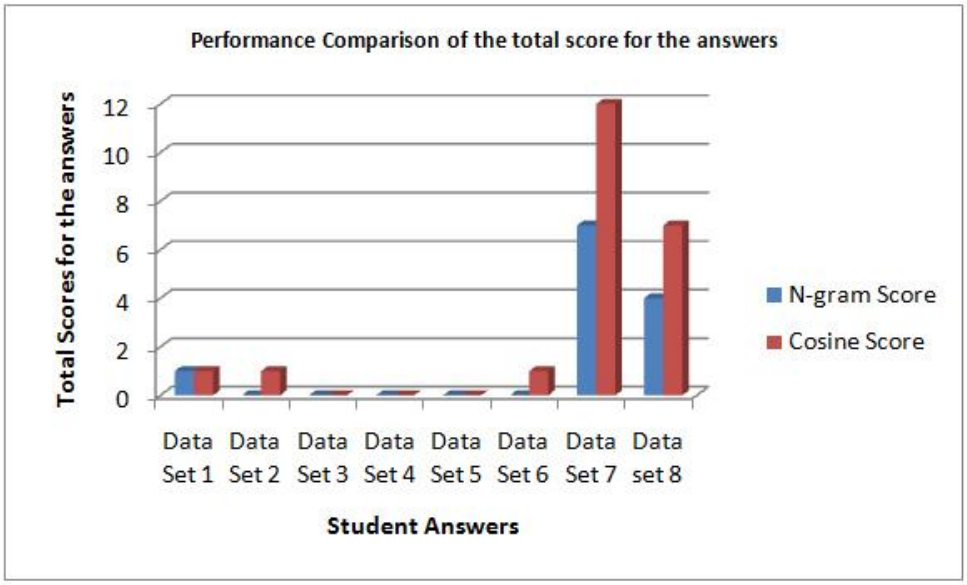

Fig. 3: Performance Comparison of the total score for the answers.

Percentage of Improvement from n-gram to cosine similarity:

The table 10 shows the percentage of improvement from n-gram to cosine similarity. The first six data sets shows the improvement for information recall questions and the last two datasets show the improvement for comprehension questions. The overall percentage of improvement for the information recall questions is 8.3% and the overall percentage of improvement for the comprehension questions is 30.7%. Therefore the cosine similarity measure outperforms the n-gram similarity measure. This is because the cosine similarity finds the similarity between two sentences by counting the number of times each word is appearing in that sentence. Then the cosine similarity finds the similarity value by measuring the dot product between the formed two vectors and finding the magnitude of the two formed vectors and multiplies them. Finally the dot product is divided by the product of the magnitude to obtain the cosine similarity value. The main advantage is that cosine similarity considers all the words in both the sentences to find the similarity value, whereas n-gram takes 4 or 5-grams to evaluate the similarity value , i.e., the n-gram similarity measure calculates the similarity value of finding the probability of the next word based on the previous n-words. Therefore n-gram method will be suitable for only very short sentences. Therefore the cosine similarity measure can be used to evaluate the short and big answers when compared to the n-gram similarity measure. This is the reason why the cosine similarity measure outperforms n-gram similarity measure by 8.3% in short answers and 30.7% in big answers.

Table 10: Percentage of improvement from n-gram to cosine similarity.

Data Sets	% of Improvement from n-gram to cosine similarity
Data Set 1	0
Data Set 2	50
Data Set 3	0
Data Set 4	0
Data Set 5	0
Data Set 6	0
Data Set 7	38.46153846
Data set 8	23.07692308

Australian Journal of Basic and Applied Sciences, 8(18) December 2014, Pages: 648-656

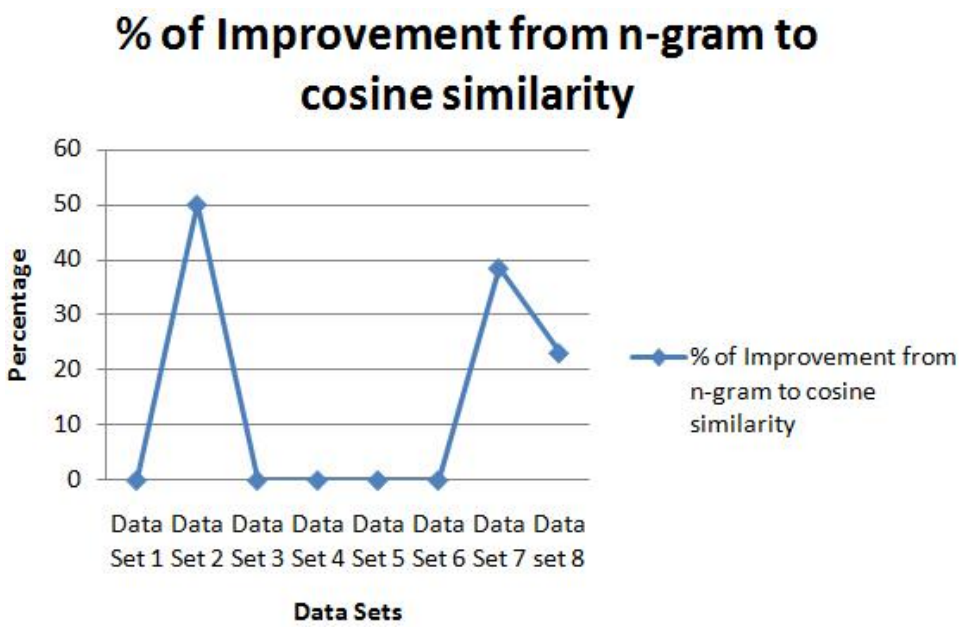

Fig. 4: Percentage of improvement from n-gram to cosine similarity

Conclusion:

This paper describes the importance of automatic assessment of students' answers. This type of automatic assessment is very important in educational sector since it helps the teachers to concentrate more on teaching and research. This helps in authentic and secured evaluation since there will be no human intervention. In most of the related work that we have surveyed, the techniques used may be different but the overall idea is to compare between the model answers and the students' answer in order to score them. The scores will depend upon how close both the answers match. The method used in this paper uses only one model answer for scoring the student answer. This is the main advantage of the method because most of the existing systems use a number of model answers to score. Therefore this method outperforms other methods even though the standard similarity measure is used for evaluating answers. The cosine similarity measure outperforms the n-gram measure by 8.3% for information recall questions and 30% for comprehension questions. Therefore in future, a new similarity measure which incorporates the cosine similarity measure and WordNet will be used to evaluate the answers and also other categories of questions other than information recall and comprehensive will be concentrated to evaluate.

REFERENCES

Abigail, R., Razon, C. Prospero Naval, Jr, D. Belen Calingacion, L. Rowena Cristina Guevara, Automated Essay Content Analysis using Concept Indexing.

Diana P´erez, Enrique Alfonseca and Pilar Rodr´ıguez, 2004. Application of the Bleu method for evaluating free text answers in an E-learning Environment.

Enrique Alfonseca and Diana P´erez, 2004. Automatic Assessment of Open Ended Questions with a Bleu inspired algorithm and shallow NLP, Advances in Natural Language Processing, Springer.

Michael Mohler and Rada Mihalcea, 2009. Text to Text Semantic Similarity for Automatic Short Answer Grading, EACL '09 Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics.

Papineni, K., S. Roukos, T. Ward and W. Zhu, 2001. Bleu:A method for automatic evaluation of machine translation.

Questions Skills: 6 Categories of Questions, KAREN TEACHER WORKING GROUP. http://ktwg.org/ktwg_texts.html.

Sargur Srihari, Jim Collins, Rohini Srihari, Harish Srinivasan, Shravya Shetty, Janina Brutt Griffler, 2008. Automated Scoring of Short Handwritten essays in reading comprehension tests, Elseiver, 172(2–3): 300–324.

Selvi, P., Dr. A.K. Bnerjee Automatic Short Answer Grading System (ASAGS), 2010. Inter JRI Computer Science and Networking, 2(1).

www.selfgrowth.com