2021 International Conference on Emerging Smart Computing and Informatics (ESCI)
AISSMS Institute of Information Technology, Pune, India. Mar 5-7, 2021

Revolutionize Cosine Answer Matching Technique for Question Answering System

Vishaka Arjun Mandge
Department of Information Technology,
AISSMS Institute of Information Technology
Pune, Maharastra, India
mandge1999@gmail.com

Meenakshi A.Thalor
Department of Information Technology,
AISSMS Institute of Information Technology
Pune, Maharastra, India
meenakshi.thalor@aissmsioit.org

2021 International Conference on Emerging Smart Computing and Informatics (ESCI) | 978-1-7281-8519-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ESCI50559.2021.9396864

***Abstract*—Imitating humans by computers i.e. Knowledge Engineering has been into trend from past few years, resulting into reducing the unnecessary manpower and skilled workers in the field of education when it comes to computer assisted assessments. Various techniques developed to evaluate answers by computers are known to be non-performing resulting into undeveloped system that revolutionizes traditional method of assessment and instant evaluation of results. One such technique is the TF-IDF using cosine similarity which proves to be a bit lacked out when comes to synonym correction. This paper aims at revolutionizing the traditional assessment technique of pen paper, replacing it with modern solutions where computer imitates the skilled person. The main objective is to achieve an algorithm that best represents the or imitates the human when it comes to evaluating the subjective answers using Natural Language Processing techniques. This method proves to be efficient enough to withdraw the traditional system and develop a distinct solution that will benefit various Educational Institutions.**

Keywords— Knowledge Engineering, Natural Language Processing (NLP), Information Extraction, Revolutionize Assessment, Keyword Matching, Cosine Similarity, TF-IDF

I. INTRODUCTION

"Knowledge Engineering is a common term used for imitating how a human expert reacts/deals with situations and take efficient decisions". Computers had played a crucial role in the education system from a few decades. Despite the availability of enough resources, the utilization of these resources is inefficient which leads to redundant use of capital and manpower.

Artificial Intelligence and Machine Learning have given rise to new technologies and scientific advancements in the past few years. These advancements have helped many sectors of society to grow digitally. In the education sector, the subjective examinations are carried out by the traditional method of pen and paper. The papers are then checked manually by the teachers and the result is generated within weeks. Checking answer sheets manually is time consuming and inefficient.

To revolutionize this practice, we have proposed an algorithm that revolutionizes the existing cosine similarity algorithm in order to generate a better result. Several techniques have been utilized till date but no such algorithm yet developed that best imitates the human person i.e. on with a great Knowledge of any subject in order to evaluate the answers of the students. We propose to develop an algorithm that will best imitate the human expert.

II. LITERATURE REVIEW

The existing online systems have objective type questions and does not include any subjective answers and even if some of the online systems exist the answer matching accuracy is very less as compared to a human expert. In this paper we try to match the human expert when it comes to evaluating subjective answers. Following Table 1 includes the survey in order to develop an algorithm that increases the accuracy.

TABLE I. LITERATURE SURVEY TABLE

TITLE	METHODOLOGY	Limitation
Comparative Study of CNN and RNN for Natural Language Processing [1]	CNN are Hierarchical architectures that extracts position-invariant features. RNN are sequential architectures consisting of two types- LSTM and GRU. It is used for modelling units in sequence	GRU/LSTM surpass CNN in textual entailment while CNN dominates in answer selection. Optimization of hidden size and back size is crucial to obtain good performance of both CNN and RNN.
Knowledge base question answering with a matching-aggregation model and question-specific contextual relations [2]	It makes use of "Matching-Aggregation" framework, in which two sequences are matched at word-level and the matching results are aggregated for a final decision. It further makes use of question-specific contextual relations	The error analysis of Knowledge based question answering system suggest that complex questions are challenging and fail in case of ambiguous questions.
A systematic mapping study of language features identification from large text collection [3]	The research on NLP and ML ensures data extraction and complete mapping of studies, by identifying, analysing and interpreting the suitable evidence.	According to the analysed literature, NLP unsupervised learning is the most trending technique used but does not ensure large text collection.
Capturing the Semantics of Key Phrases Using Multiple Languages for Question retrieval [4]	Paraphrase based approach for word mismatch in question retrieval. For this it makes use of two approaches, key concept identification approach and pivot language translation based approach	This model outperforms most of the existing models by eliminating problems like word verboseness and word mismatch in question retrieval. Does not support multiple Linguistic resources.

Deep Learning Model used in text classification [5]	The deep learning model carries out text classification using three steps- text pre-processing, text feature extraction and classification model construction.	The deep learning model divides the text into long text and short text which the RNN model cannot parallelize well, CNN is best for long text.
Self-Learning System with Automatic Feedback for Text Answers [6]	It uses NLP tools and methodologies to automatically evaluate text answers. Compares semantic similarity between the model answer and the provided answer.	There are restrictions due to language complexity. System can only work with selected sentence pattern and grammar conditions.
Automatic Answer Validation System on English Language [7]	Combines the question and answer into Hypothesis(H) and the Supporting text(T) to check the entailment relation as either "VALIDATED" or "REJECTED"	The system uses Lexical information, Dependency, Chunking and Named Entities but does not use semantic features for answer validation.

III. METHODOLOGY-I

The Question Answering System, mainly uses NLP which is nothing but the ability of a computer to understand and interpret human language. In our model, NLP is used to analyse the textual data that will be provided by the student as well as the teacher.

A. Preprocessing

After the answer of the students as well as of the teacher is stored into the database, the comparison between the model answer of the student as well as of the teacher takes place. Firstly, tokenization takes place, tokenization is done both the ways by sentence tokenizing as well as word tokenizing. After tokenization, these individual words are parsed for Stop Words, Character Set to eliminate in order to perform keyword matching; this is basically the filtering step. For every individual token there is a categorization of weather it is a stop word, keyword or Charter set. The Stop words Such as "is, am, are" are ignored. Once the tokenization is done and stop words are eliminated, the character set put up by the teacher is also eliminated from the answer of the student. Once all the character set and stop words are removed the next step is to identify the keywords from the students answer (refer Table 2 and 3). Those keywords are then compared with the keywords present in the teacher's model answer.

TABLE II. MODEL FILE

SR.NO	QUESTION	ANSWER	KEYWORDS
1.	What is Machine Learning?	Machine learning is the ability of a computer to automatically learn from experience to perform certain tasks.	Machine, Learning, ability, computer, automatically, learn, experience, perform, tasks
2.	What is Knowledge Engineering?	Knowledge Engineering is to understand how a human would perform certain tasks in a certain domain.	Knowledge, Engineering, understand, human, perform, tasks, domain
3.	What is Natural Language Processing?	Natural Language Processing is the ability of computer to understand human language.	Natural, Language, Processing, ability, computer, understand, human, language

TABLE III. STUDENT FILE

SR.NO	QUESTION	ANSWER	KEYWORDS
1.	What is Machine Learning?	Machine learning is a subset of Artificial Intelligence where computers learn from past experience.	Machine, Learning, subset, Artificial, Intelligence, computers, learn, past, experience
2.	What is Knowledge Engineering?	Knowledge Engineering tries to imitate human behaviour.	Knowledge, Engineering, Tries, imitate, human, behaviour
3.	What is Natural Language Processing?	Natural Language Processing converts human language into machine language.	Natural, Language, Processing, converts, human, language, machine, language

B. Information Extraction

After identifying the keywords, the next step is to identify meaning of those keywords. Ranking of the keyword phrases takes place using various functionalities of Natural Language Processing. Feature extraction is used to reduce the complexity of the data during processing by eliminating the redundant features. The main process after feature extraction is to find out the importance of an individual word from the answer written by the student, through TF-IDF vectorization. This technique provides a statistical importance of a word in student's answer. To find out the similarity of student's answer with the teacher's answer the final concept used is cosine similarity. The cosine angle between the two vectors defines the similarity between two answers. The smaller the cosine angle is greater is the similarity among the two answers irrespective of the size of the solution itself. Refer Fig 1.

C. Score Generation

Score generation is the final step of our system. There are a few exceptional cases where if a student score is less the 35 percent and more the 90 percent then there is manual checking done by the teacher, who can view answers of the student from the database. They can only view the answer but cannot modify it.

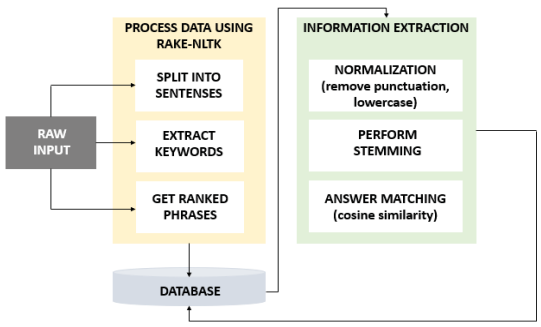

Fig. 1. Technical Flow Diagram

IV. CALCULATIONS

For example, if we consider the first question, 'What is Machine Learning?'. The model answer provided by the teacher and the answers given by the students is as follows,

Teacher : Machine learning is the ability of a computer to automatically learn from experience to perform certain tasks.

Student 1 : Machine learning is a subset of Artificial Intelligence where computers learn from past experience.

Student 2 : Machine learning is a part of Artificial Intelligence.

A. Term Frequency (TF) and Inverse Data Frequency (IDF):

TF stands for Term Frequency, which is nothing but the number of times a word has appeared in a particular document, i.e. the frequency of the word. For a word w, the term frequency can be calculated by,

$$TF = n_w / n$$

where,

n_w - No. of times the word w has appeared in the document.

n - No. of words in the document.

Inverse Data Frequency (IDF) is used to find out the frequency of the word across multiple documents.

$$IDF = 1 + \log (N / d_w)$$

where,

N - Total no. of documents.

d_w - No. of documents that contain the word w.

Further we calculate the weight of the word by,

$$Wt = TF * IDF$$

Table 4 shows a matrix with the occurrence of words in different answers against all the words present in different answers. These values are used for calculating TF, IDF and finally its weight.

TABLE IV. MATRICES

	Teacher	Student 1	Student 2
machine	1	1	1
learn	1	1	1
ability	1	0	0
computer	1	1	0
automatic	1	0	0
experience	1	1	0
perform	1	0	0
certain	1	0	0
task	1	0	0
subset	0	1	0
artificial	0	1	1
intelligence	0	1	1
past	0	1	0
part	0	0	1

The word count in the Teacher, Student 1and Student 2 is 17, 14 and 8. Consider for example, the word 'computer'. Term frequency will be,

TF = 1 / 17 = 0.588

IDF = 1 + log (3 / 2) = 1.176

Wt = * .176 = 0.691

Similarly, theft, IDF and the final weight for each word is calculated to be, refer Table 5, 6 and 7

TABLE V. TF

	Teacher	Student 1	Student 2
machine	0.588	0.0714	0.125
learn	0.588	0.0714	0.125
ability	0.588	0	0
computer	0.588	0.0714	0
automatic	0.588	0	0
experience	0.588	0.0714	0
perform	0.588	0	0
certain	0.588	0	0
task	0.588	0	0
subset	0	0.0714	0
artificial	0	0.0714	0.125
intelligence	0	0.0714	0.125
past	0	0.0714	0
part	0	0	0.125

TABLE VI. IDF

	IDF
machine	1
learn	1
ability	1.477
computer	0.691
automatic	1.477
experience	0.691
perform	1.477
certain	1.477
task	1.477
subset	1.477
artificial	0.691
intelligence	0.691
past	1.477
part	1.477

TABLE VII. TF*IDF

	Teacher	Student 1	Student 2
machine	0.588	0.0714	0.125
learn	0.588	0.0714	0.125
ability	0.868	0	0
computer	0.406	0.0493	0
automatic	0.868	0	0
experience	0.406	0.0493	0
perform	0.868	0	0
certain	0.868	0	0
task	0.868	0	0

subset	0	0.105	0
artificial	0	0.0493	0.863
intelligence	0	0.0493	0.863
past	0	0.105	0
part	0	0	0.185

B. Cosine Similarity:

Now we need to calculate the cosine similarity between the teachers' answer and the students answer. The cosine similarity between two documents calculates the cosine angle between them. The formula is as follows,

$$\text{Cosine Similarity} = \frac{\text{Dot product (teacher, student)}}{\| \text{teacher} \| * \| \text{student} \|}$$

Cosine similarity gives the measure of similarity between the teachers and students answer. If the value obtained by cosine similarity is small, then the similarity is high and vice versa.

V. METHODOLOGY-II

The methodology so used in order to generate a percentage match has a limitation that, it does not consider the synonyms of the words while calculating the term frequency. We propose to overcome this limitation by using a synonym dictionary in order to take into account the word that mean the same. For example, Consider the example in the calculations section which consist of the word '**subset**' and '**part**' (refer Table 8) which have the same meaning in the sentence. If they are considered to be one the accuracy in evaluating the answers will increase to a great extent.

TABLE VIII. MATRICES

	Teacher	**Student 1**	**Student 2**
machine	1	1	1
learn	1	1	1
ability	1	0	0
computer	1	1	0
automatic	1	0	0
experience	1	1	0
perform	1	0	0
certain	1	0	0
task	1	0	0
subset	0	1	0
artificial	0	1	1
intelligence	0	1	1
past	0	1	0
part	0	0	1

The synonyms of the keywords of the model answer are compared with the student's keywords and wise-versa can be done in order to eliminate the repetition of the keywords that mean the same (refer Table 9).

TABLE IX. MODIFIED MATRICES AFTER SYNONYM CORRECTION

	Teacher	**Student 1**	**Student 2**
machine	2	2	1
learn	1	1	1
ability	1	0	0
automatic	1	0	0
experience	1	1	0
perform	1	0	0
certain	1	0	0
task	1	0	0
subset	0	1	1
artificial	0	1	1
intelligence	0	1	1
past	0	1	0

So, the new calculations after the synonym correction done can be referred from the Table 10, 11 and 12.

TABLE X. TF WITH SYNONYM CORRECTION

	Teacher	**Student 1**	**Student 2**
machine	0.118	0.250	0.125
learn	0.588	0.0714	0.125
ability	0.588	0	0
automatic	0.588	0	0
experience	0.588	0.0714	0
perform	0.588	0	0
certain	0.588	0	0
task	0.588	0	0
subset	0	0.0714	0.125
artificial	0	0.0714	0.125
intelligence	0	0.0714	0.125
past	0	0.0714	0

TABLE XI. IDF WITH SYNONYM CORRECTION

	IDF
machine	1
learn	1
ability	1.477
automatic	1.477
experience	0.691
perform	1.477
certain	1.477
task	1.477
subset	1.477
artificial	0.691
intelligence	0.691
past	1.477

TABLE XII. TF*IDF WITH SYNONYM CORRECTION

	Teacher	**Student 1**	**Student 2**
machine	0.118	0.250	0.125
learn	0.588	0.0714	0.125
ability	0.868	0	0
automatic	0.868	0	0
experience	0.406	0.0493	0
perform	0.868	0	0
certain	0.868	0	0
task	0.868	0	0
subset	0	0.105	0.185
artificial	0	0.0493	0.863
intelligence	0	0.0493	0.863
past	0	0.105	0

Using the cosine similarity, we find the similarity between the model file and the student answers.

VI. RESULTS AND DISCUSSION

The cosine similarity for both the algorithms i.e. the algorithm without synonym correction and the algorithm after synonym correction is deduced and compared for accuracy as mentioned in Table 13 and 14.

TABLE XIII. . RESULT WITHOUT SYNONYM CORRECTION

	Document	Cosine Similarity with the Model Answer (Teacher)
Student 1	[0.0714, 0.0714, 0,0.493, 0, 0.0493, 0, 0, 0, 0.105, 0.0493, 0.0493, 0.105, 0]	0.327832128
Student 2	[0.125, 0.125, 0, 0, 0, 0, 0, 0, 0, 0, 0.185, 0.863, 0.863, 0, 0.105]	0.054823596

TABLE XIV. RESULT WITH SYNONYM CORRECTION

	Document	Cosine Similarity with the Model Answer (Teacher)
Student 1	[0.250, 0.0714, 0, 0, 0.0493, 0, 0, 0, 0.105, 0.0493, 0.0493, 0.105]	0.157148903
Student 2	[0.125, 0.125, 0, 0, 0, 0, 0, 0, 0.185, 0.863, 0.863, 0]	0.035626713

The cosine similarities for both the results is compared:

$$0.157148903 < 0.327832128, \text{ and}$$

$$0.035626713 < 0.054823596$$

The smaller the value is, the higher the similarity is. Hence, the proposed algorithm i.e. the algorithm with synonym correction yields better results as compared to the traditional one.

VII. CONCLUSION

Examination plays a very vital role in schools, colleges, universities and various other educational institutes. These educational institutes sometimes often conduct online examinations. But these exams only include multiple -choice questions, which are efficient in checking the student's aptitude skills, in contrast they fail to determine the theoretical knowledge a student possesses. Thus, subjective answers must be incorporated in online examinations. The proposed system attempts to calculate the subjective answers. The proposed system calculates the student's answer based on the Lexical similarity of the student's answer compared with teacher's answer with the help of the function cosine similarity () with synonym corrections included that overcomes the limitation of the existing algorithm.

REFERENCES

[1] Yin, W., Kann, K., Yu, M., & Schütze, H., "Comparative Study of CNN and RNN for Natural Language Processing". ArXiv, abs/1702.01923,2017.

[2] Lan, Y., Wang, S., & Jiang, J.," Knowledge Base Question Answering with a Matching-Aggregation Model and Question-Specific Contextual Relations", IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27, 1629-1638,2019.

[3] Diellza Nagavci Mati, Jaumin Ajdari, Bujar Raufi, Mentor Hamiti and Besnik Selimi, "A Systematic Mapping Study of Language Features Identification from Large Text Collection". In: 2019 8th Mediterranean Conference on embedded computing (MECO), 10-14 June 2019,

[4] W. Zhang, Z. Ming, Y. Zhang, T. Liu and T. Chua, "Capturing the Semantics of Key Phrases Using Multiple Languages for Question Retrieval," in IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 4, pp. 888-900, 1 April 2016.

[5] JingjingCai, Jianping Li, Wei Li, and JiWang, Deeplearning model used in text classification. In 15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), China, 2018, pp. 123-126, 2018.

[6] K.T. Kodithuwakku., W.S.J.M.C.D. Senevirathne,R.A.D.A. Randeniya, M.M.D.D.A. Wijewardane and M. P. A. W. Gamage, Self-Learning System with Automatic Feedback for Text Answers. In 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), pp. 1-7, 2017.

[7] Partha Pakray, Santanu Pal, Sivaji bandyopadhyay, Automatic Answer Validation System on English Language. In 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE),2010.