International Journal of Scientific & Engineering Research, Volume 11, Issue 12, December-2020
ISSN 2229-5518

Word Embedding Based Answer Evaluation System for Online Assessments

Pratham Sharma, Prakhar Baphna, Nannapaneni Akshaj

Abstract— With the COVID-19 pandemic changing the way we live our lives, the world is rapidly adopting a digital form of life. Digital platforms are increasingly being adopted by schools and universities for delivering lectures as well as conducting assessments. Online assessments greatly differ from their pen-paper counterparts in terms of the method of evaluation. Traditionally, pen-paper based assessments are evaluated manually by professors. This paper discusses a method based on word embedding and natural language processing that can be implemented in online assessments to emulate the traditional way of answer evaluation. Such a system can be used for evaluation of online assessments in the same way as traditional pen-paper based assessments, in a fraction of the time. With this system, reliability in answer evaluation is maintained while saving time and effort.

Index Terms— Assessment evaluation, Distance learning, Doc2Vec, Machine Learning, Natural Language Processing, Word2Vec, Word embedding model

1 Introduction

TRADITIONALLY, examinations in schools and universities have been conducted in pen-paper mode. Students are required to write the answer to each question in the answer sheet. Answer sheets are then given to the respective course instructor for evaluation. Instructors award marks to students based on how well the answer written by the student matches the answer expected by the faculty.

However, with classes being shifted to an online mode of course instruction, assessments are also being conducted in online mode. With this, the possibility of conducting pen-paper based examinations has greatly reduced. Usually, online exams are conducted in the form of Multiple-Choice Questions (MCQs). But online examinations fall behind offline (written) exams in terms of testing the conceptual understanding of students.

In order to overcome this pitfall of online exams, we propose a system that emulates the traditional manner of conducting examinations for schools and universities. The proposed system allows students to answer questions in the same manner as in offline examinations. The system then evaluates answer sheets for students based on how well the answer written matches the answer expected by the faculty conducting the test.

In order to do this, the proposed system makes use of word embedding models such as Word2Vec and Natural Language Processing to calculate the degree to which the student's response to a question matches the faculty's expected response. Marks are awarded based on this degree of similarity.

- *Pratham Sharma is currently pursuing bachelor's degree program in computer science and engineering in Vellore Institute of Technology, Vellore, India. E-mail: pratham.sharma2018@vitstudent.ac.in*
- *Prakhar Baphna is currently pursuing bachelor's degree program in information technology in Vellore Institute of Technology, Vellore, India. E-mail: prakhar.baphna2018@vitstudent.ac.in*
- *Nannapaneni Akshaj is currently pursuing bachelor's degree program in information technology in Vellore Institute of Technology, Vellore, India. E-mail: nannapaneni.akshaj2018@vitstudent.ac.in*

Hence, the proposed system is a way of tackling the pitfalls of online examinations by emulating the traditional manner of conducting and evaluating examinations.

2 Existing Methodologies

Currently, online examinations utilize forms such as Google Forms in order to create exam question papers consisting mainly of MCQs. Sometimes, very short answers may also be incorporated in the form of fill in the blank questions.

Answers for MCQs may be automatically evaluated by configuring the form appropriately. On doing so, students are awarded either 0 or 1 mark for each question based on whether the submitted response matches the correct response configured by the faculty. In some cases, very short answer questions may also be evaluated automatically, but often require a manual review to account for spelling errors, use of synonyms, abbreviations, capitalization, etc.

This brings us to the first drawback of online examinations – requirement of manual work for evaluation. Even with examinations being conducted online, some amount of manual work is required on the part of the course faculty to evaluate answer sheets and award marks.

However, the major drawback of the current method of conducting online examinations is the lack of ability of assessing a student's conceptual understanding of the subject. Multiple-Choice Questions do not adequately test a student's knowledge and understanding. The student may answer MCQs by guessing or simply by copying.

Traditional pen-paper based examinations fare better in testing the student's understanding and knowledge as they help capture the student's thought process while answering questions. With a short or long answer, a student can clearly demonstrate his/her understanding of the subject. The course faculty evaluates the answers and awards marks based on how much the student's answer matches the answer expected by the faculty.

Therefore, there is a great need for a system which closely emulates the traditional way of conducting and evaluating

International Journal of Scientific & Engineering Research, Volume 11, Issue 12, December-2020
ISSN 2229-5518

examinations. The system proposed in this paper is a novel system which aims to emulate the traditional manner of conducting and evaluating examinations. The system makes use of word embedding models such as Word2Vec and Natural Language Processing. Using these models, the system can automatically evaluate answers based on the degree of similarity between student's response and the response expected by the faculty.

Such a system will greatly reduce time and effort in conducting and evaluating online examinations while also being able to satisfactorily assess a student's knowledge and conceptual understanding. The system combines the reliability of the traditional method of conducting examinations with modern processing capabilities to reduce time and effort in evaluating answer sheets and awarding marks.

3 WORD EMBEDDING MODEL

In order to overcome the aforementioned drawbacks of the current system of online examinations, the proposed system, based on word embedding models such as Word2Vec and Doc2Vec, as well as Natural Language Prcessing (NLP), attempts to mimick and automate the manual method of evaluating answers.

In the manual method of answer evaluation, the faculty evaluates answers and awards marks based on how similar the answer submitted by the student is to the answer expected by the faculty semantically. The basic concept of the system proposed in this paper is to formulate a model to calculate the degree of similarity between two texts – in this case, the answer submitted by student and the answer expected by the course faculty. Marks are then awarded based on the degree of similarity calculated by the model.

The proposed system makes use of word and document embedding models which aim to express individual words in terms of multidimensional vectors. This is accomplished with the use of Word2Vec as well as Doc2Vec, that attempt to learn the "meaning" of a particular word based on its occurrence and usage with other words in a piece of text. This "meaning" of each word is represented mathematically as a multidimensional vector, with similar words having near-similar vectors.

The proposed system involves training a word embedding model on a large collection of structured text, known as a text corpus. A trained word embedding model is capable of representing words mathematically by generating multidimensional vectors for each word. By generating vectors for each word in a piece of text, a single vector can be generated for the entire text or paragraph. The system then computes the degree of similarity between two texts by calculating the cosine similarity between the vectors for the two texts. Finally, a score is given based on this degree of similarity.

3.1 Word2Vec and Doc2Vec

Word2Vec, developed by a team of reserachers at Google led by Tomas Mikolov in 2013, is a two-layer neural network that processes text by generating multidimensional vectors for words [1], [2]. It takes a large collection of structured text data, known as a text corpus, as input and produces vectors that represent words in that text corpus mathematically. The purpose and usefulness of Word2Vec is to group vectors of similar words together in a multidimensional vector space [2].

Generating such a mathematical representation of words makes it easier to detect similarity between words and texts as similar words will have near-similar vectors, located close to each other in the vector space. Calculating a numerical value for similarity then becomes as simple as calculating the cosine of the angle between the two vectors i.e., calculating cosine similarity.

Training a Word2Vec model requires a large enough text corpus. Given enough data, usage and contexts, Word2Vec can make highly accurate guesses about a word's meaning based on past occurrences [2]. This "meaning" of a word is represented in the form of a multidimensional vector. Hence, such a mathematical representation of words can be used to establish a word's association with other words.

Essentially, a word embedding model such as Word2Vec captures the semantics of text very well and serves as a great method to measure the semantic similarity between two pieces of text.

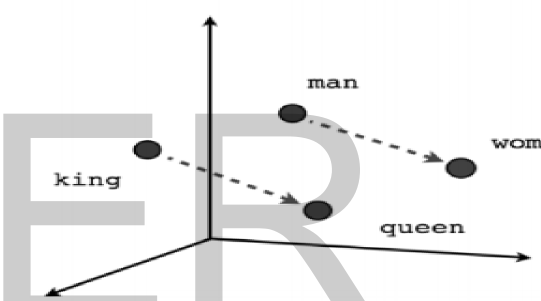

Fig. 1. "King" is to "Queen" as "Man" is to "Woman" – a classic example of how Word2Vec expresses words in the form of vectors in order to capture meaningful associations between them.

Doc2Vec extends Word2Vec to create a numerical representation of a document. Doc2Vec models generate a document vector for each document in a document corpus.

3.2 CBOW and Skip-gram Model

The two most popular algorithms used to generate word embeddings using Word2Vec are the Continuous Bag-of-Words model (CBOW) and the Skip-gram model. Both CBOW model and Ski-gram model are architectures for generating vector representation of words using a shallow neural network [2].

The CBOW model tries to understand the context of a given piece of text and takes the words in the text as input. It then predicts new words that are contextually relevant to the given text i.e., it tries to predict the tartget word by understanding the context of the surrounding words.

The skip-gram model tries to achieve the reverse of what the CBOW model does. To a certain extent, the skip-gram model is a mirror image of the CBOW model. It takes a target word as the input and tries to predict the source context words i.e., the possible surrounding words for the given target word.

International Journal of Scientific & Engineering Research, Volume 11, Issue 12, December-2020
ISSN 2229-5518

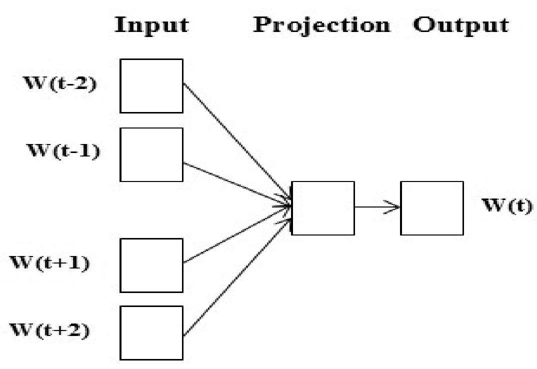

Fig. 2. Architecture for the Continuous Bag-of-Words (CBOW) Model.

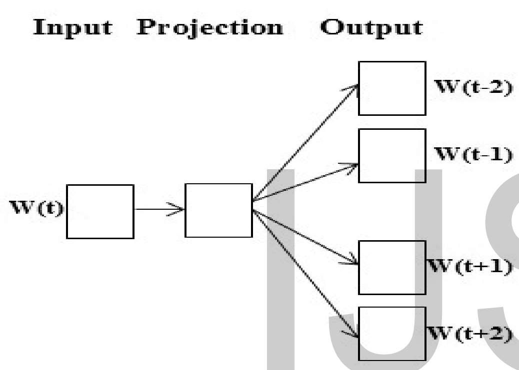

Fig. 3. Architecture for the Skip-gram Model.

Overall, the skip-gram model is found to work well with small amount of data and represents rare words well. On the other hand, the Continuous Bag-of-Words (CBOW) model is faster in training and gives a better representation for frequently used words.

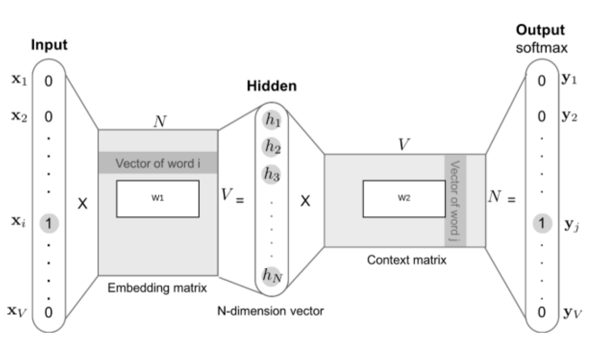

Fig. 4. Architecture for Word2Vec model. Input and output vectors are of size V while the hidden vector is of size N.

Word2Vec uses a shallow neural network with a single hidden layer based on the two models described above. On training on a text corpus, the Word2Vec model represents each word as a V-dimensional vector. By generating word embeddings and giving a mathematical representation to words, we can now perform text processing operations much more easily. Similarity measures can be applied easily in order to obtain the desired result.

3.3 Cosine Similarity

The proposed system works by calculating the degree of similarity between two texts – the answer submitted by student and the answer expected by the faculty conducting the test. A similarity measure must be used in order to perform this calculation.

With the help of the Word2Vec model, the system represents texts mathematically as multidimensional vectors. These multidimensional vectors capture the semantics of texts based on context and usage. Vectors for words are distributed in a multidimensional vector space. Words and texts having similar meanings are located close to each other in this vector space.

In order to calculate the degree of similarity between two such vectors, cosine similarity measure is used. Vectors for similar words are placed close together and have a small angle between them, thereby giving a high value for the cosine of this angle, and vice versa.

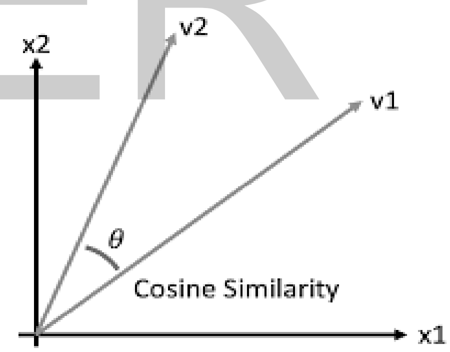

Fig. 5. Cosine similarity measure for two vectors, v1 and v2.

Cosine similarity for the vectors generated by the Word2Vec model for two texts can be calculated using scientific computation libraries in Python such as SciPy. This gives a numerical value between 0 and 1, with 1 indicating a high degree of similarity and 0 indicating very low similarity between the two texts.

3.4 System Implementation

The proposed system is implemented using Python programming language along with the following important libraries and packages:

1. Genism – library for training of vector embeddings containing the Word2Vec and Doc2Vec models.
2. SciPy – library for scientific and technical computa-

International Journal of Scientific & Engineering Research, Volume 11, Issue 12, December-2020
ISSN 2229-5518

tions.

The system uses the “wiki-english-20171001” text corpus for training the Word2Vec model. The wiki-english-20171001 text corpus consists of 6124 MB of extracted Wikipedia dump from Octuber 2017. Hence, the corpus provides a large enough dataset to train the Word2Vec model on.

Each document in the text corpus is tagged and yielded. A list of all tagged documents in the text corpus forms the training dataset for the Word2Vec model. The Word2Vec model is trained to represent each word in the corpus as a 200-dimensional vector. Vocabulary for the Word2Vec model is then constructed. The trained model can the be saved and reloaded again for use.

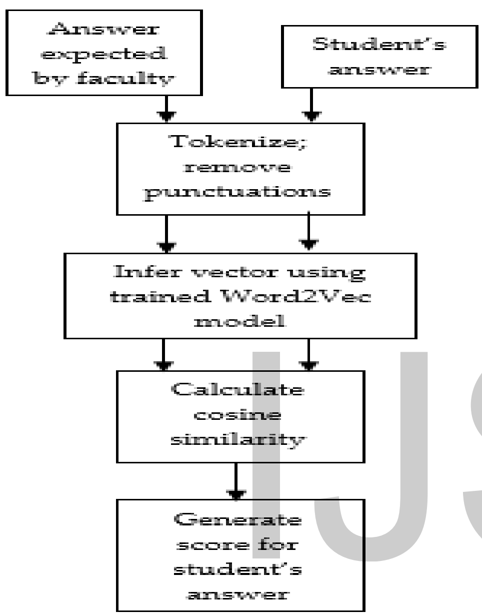

Fig. 6. Flowchart for the proposed system to award marks automatically based on cosine similarity measure.

Given two pieces of text – in the case of this paper, the student’s submitted answer and the answer expected by the faculty – similarity between them is calculated based on the cosine similarity measure using the SciPy library. Both texts are tokenized and punctuations are removed for better text processing. Using the trained Word2Vec model, the vector representation for each piece of text is inferred using the tokens in that text. Having represented both texts as vectors, similarity is calculated using built-in functions for cosine distance available in SciPy.

Subtracting the value for cosine distance between the vectors for the two texts from 1 gives the value for degree of similarity. This degree of similarity lies in the range of 0 to 1. Further, the similarity score can be multiplied with an appropriate factor such as 10 or 5 to award marks for a question. For example, for a question worth 10 marks, the degree of similarity calculated by the system can be multiplied with 10 to get the marks out of 10 for the answer.

4 RESULTS

The following table shows the results obtained by the proposed system in awarding marks automatically.

TABLE 1
RESULTS OBTAINED

S. No.	Student’s answer	Faculty’s expected answer	Score (out of 10)
1	“The President greets the press in Chicago”	“Barack Obama speaks in Illinois”	7.11
2	“Human Computer Interaction is the study of interaction between humans and computers”	“Human Computer Interaction is a discipline concerned with the interaction between computer systems and human users”	9.07
3	“The new farm bills are not acceptable to the farmers”	“Farmers are protesting in large numbers against the newly proposed farm bills”	8.13
4	“Machine Learning algorithms learn and improve through experience”	“Machine Learning can be used as a tool in the Data Science process”	6.90
5	“The dog is lazy so he lost the race”	“The quick brown fox jumps over the laxy dog”	5.83

Hence, the proposed system performs very well to emulate the manual method of answer evaluation. The system makes use of modern processing capabilities to perform the task of answer evaluation in minimal time, while also maintaining reliability. The new system greatly saves time and effort on part of the course faculty.

5 CONCLUSIONS

The drawbacks of the existing system of conducting online examinations are overcome by the system proposed in this paper. The new system based on word embedding models and NLP utilizes modern processing capabilities to save time and effort in evauating answers for online examinations.

Such a system can be implemented in a web-based application to provide a better platform to schools and universities to conduct online examinations.

REFERENCES

[1] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch and A. Joulin, “Advances in Pre-Training Distributed Word Representations”.

[2] H. Yousuf, A.Q. Al-Hamad, S. Salloum, “An Overview on CryptDb and Word2vec Approaches”, *Advamces in Science, Technology and Engineering Systems Journal,* vol. 5, no. 5, pp. 1282-1287 (2020).