IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 2, June 2023

Subjective Answer Evaluation using Machine Learning

Shijimol Shibu[1] and Sindhu Daniel[2]

Student, Department of computer Applications[1]

Assistant Professor, Department of computer Applications[2]

Musaliar College of Engineering and Technology, Pathanamthitta, Kerala, India

Abstract: *Subjective questions and answers can assess the performance and ability of a student in an open-ended manner. The answers, naturally, are not bound to any constraint, and students are free to write them according to their mindset and understanding of the concept. This paper proposes a novel approach that utilizes various machine learning, natural language processing techniques, and tools to evaluate descriptive answers automatically. Solution statements and keywords are used to evaluate answers, and a machine learning model is trained to predict the grades of answers. Subjective exams are considered more complex and scarier by both students and teachers due to their one fundamental feature, context. A subjective answer demands the checker check every word of the answer for scoring actively, and the checker's mental health, fatigue, and objectivity play a massive role in the overall result. Therefore, it is much more time and resource-efficient to let a system handle this tedious and somewhat critical task of evaluating subjective answers.*

Keywords: Subjective answers.

I. INTRODUCTION

Subjective questions and answers can assess the performance and ability of a student in an open-ended manner. The answers, naturally, are not bound to any constraint, and students are free to write them according to their mindset and understanding of the concept. With that said, several other vital differences separate subjective answers from their objective counterpart. For one, they are much longer than the objective questions. Secondly, they take more time to write. Moreover, they carry much more context and take a lot of concentration and objectivity from the teacher evaluating them. Evaluation of such questions using computers is a tricky task, mainly because natural language is ambiguous. Several pre-processing steps must be performed, such as cleaning the data and tokenization before working on it. Then the textual data can be compared using various techniques such as document similarity, latent semantic structures, concept graphs, ontologies. The final score can be evaluated based on Similarity, keywords presence, structure, language.

II. LITERATURE SURVEY

Automated Assessment System for Subjective Questions Based on LSI (Xinming Hu and Huosong Xia in 2010)

They used Chinese automatic segmentation techniques and subjective ontologies to make a kdimensionalLSI space matrix. The answers were presented in TF-IDF embedding matrices, and then Singular Value Decomposition (SVD) was applied to the term-document matrix, which formed a semantic space of vectors. LSI played the role of reducing problems with synonym and polysemy. At last, the Similarity between answers was calculated using cosine similarity. Dataset consisted of 35classes and 850 instances marked by teachers, and the results showed a 5% difference in grading done by teacher and the proposed system.

Automatic scoring system for short descriptive answer written in Korean using LSP (JeongEun Kim in 2017).

LSP can structure the semantic of the answer to help understand the user's intentions. A synonym list was also utilized to help expand the keywords so they match various answer styles. Dataset was obtained from 88 students and converted to LSP, which was later compared with the solution LSP to score the answer. As a result, the system performed better than the existing system

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 2, June 2023

Pairwise document similarity measure based on present term set (Oghbaie and Zanjireh in 2018).

pair-wise Similarity measure to measure the similarity between two documents based on the keywords which appear in at least one of the documents. The work proposed a new similarity measure called PDSM (pair-wise document similarity measure), a modified version of the preferable properties approach. The proposed similarity measure was applied to text mining applications such as documents detection, k Nearest Neighbours (KNN) for single-label classification, and K-means clustering. An evaluation measure of accuracy was used, and as a result, the PDSM method produced better results than other measures

III. PROPOSED SYSTEM

The proposed system consists of data collection and annotation, preprocessing module, similarity measurement module, model training module, results predicting module, machine learning model module, and final result predicting module. First, the inputs are being taken from the user, which consists of keywords, solutions, and answers. Figure 2 shows the proposed system.

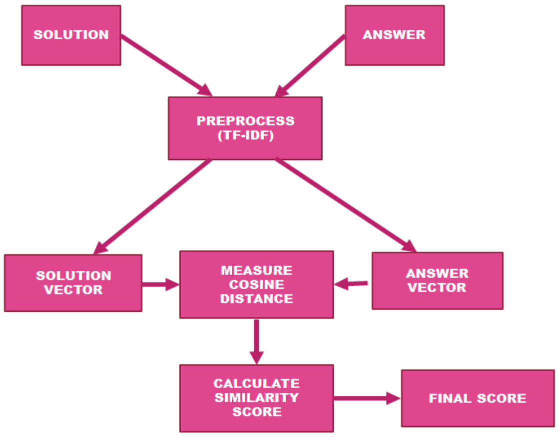

Fig 2 proposed system

IV. METHODOLOGY

Word2Vec

Word2Vec is a predictive word embedding technique which converts a word into a vector of numbers based on the context of the target word. Vectors out of the words are generated using the surrounding words to represent the target words. It uses Neural Network whose hidden layer encodes the word representation.

(Global Vector)

Glove is a count-based model that learns vectors or words from their co-occurrence information, i.e., how frequently they appear together in large text corpora. It is a high-speed word embedding technique which is trained on word-word co-occurrence probabilities and have the potential for achieve some form of meaning which can be encoded as vector differences.

Word Mover's Distance for Text Similarity

Introduction of the NLP (Natural Language Processing) revolutionized all the industries. So, NLP is a branch of AI (artificial Intelligence) that helps computer understand, interpret and manipulate human language. Now, with heaps of data available (thanks to big data) to us the major challenge that industries were facing was to communicate with computers. Our language system is astoundingly complex and diverse. We have the capability to express ourselves in infinite ways, may it be verbally, physically or written. The first challenge was written text. We have hundreds of languages and each with its unique set of grammar and syntax rules

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 2, June 2023

Comparison with Previous Methods

A pre-trained word2vec model from Google consisting of 300 dimensions of around 100 billion words vocabulary is used for this experiment. Corpus was divided into 8:2 ratio representing test and train data, respectively. Train data was used to calculate initial scores from the score prediction modules and train the machine learning model. Afterward, testing data was fed to the system one by one, updating the machine learning model. The results are obtained using cosine similarity and word mover's distance combined with a Multinomial Naive Bayes model. The score prediction module is working fairly accurately, achieving 88% accuracy. This much accuracy is significant because of word2vec in this case, and it can capture the semantic meaning of answers so well that it gives us very well Similarity among answers. Furthermore, if word2vec lacks inconsistent answers, keyword mapping and unmapped sentences thresholds still give a satisfactory score to the answers. The results obtained by cosine similarity are semantically weak, and the Subjective Answers Evaluation Using Machine Learning and Natural Language Processing model cannot get trained on the correct data as it does for the WDM case. Cosine similarity paired with a machine learning model yields 86% accuracy for this short datbase

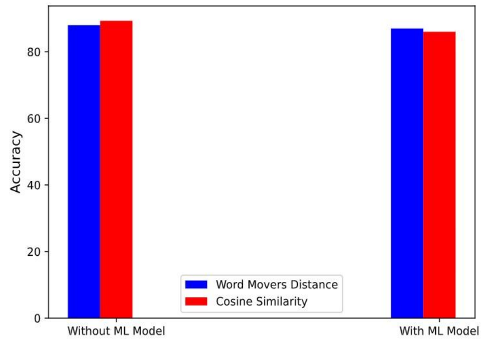

Fig 1 Accuracy comparison of different models.

V. CONCLUSION AND FUTURE SCOPE

This paper proposed a novel approach to subjective answers evaluation based on machine learning and natural language processing techniques. Two score prediction algorithms are proposed, which produce up to 88% accurate scores. Various similarity and dissimilarity thresholds are studied, and various other measures such as the keyword's presence and percentage mapping of sentences are utilized to overcome the abnormal cases of semantically loose answers. The experimentation results show that on average word2vec approach performs better than traditional word embedding techniques as it keeps the semantics intact. Furthermore, Word Mover's Distance performs better than Cosine Similarity in most cases and helps train the machine learning model faster. With enough training, the model can stand on its own and predict scores without the need for any semantics checking. We studied various methods used for subjective answer evaluation in the past and looked at their shortcomings. In this paper, we propose a new approach to solve this problem which consists of training a machine learning classification model with the help of results obtained from our result prediction module and then using our trained model to reinforce results from the prediction model, which can lead to a fully trained machine learning model.

REFERENCES

[1] J. Wang and Y. Dong, ''Measurement of text similarity: A survey,'' Information, vol. 11, no. 9, p. 421, Aug. 2020.

[2] M. Han, X. Zhang, X. Yuan, J. Jiang, W. Yun, and C. Gao, ''A survey on the techniques, applications, and performance of short text semantic similarity,'' Concurrency Compute, Pract. Expert., vol. 33, no. 5, Mar. 2021.

[3] M. S. M. Patil and M. S. Patil, ''Evaluating Student descriptive answers using natural language processing,'' Int. J. Eng. Res. Technol., vol. 3, no. 3, pp. 1716–1718, 2014.

[4] P. Patil, S. Patil, V. Miniyar, and A. Bandal, ''Subjective answer evaluation using machine learning,'' Int. J. Pure Appl. Math., vol. 118, no. 24, pp. 1–13, 2018.