IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁙ Peer-reviewed journal ⁙ Vol. 12, Issue 10, October 2023

DOI: 10.17148/IJARCCE.2023.121008

Review Online Examination System Using Artificial Intelligence

Rudresh Kale[1], Aniket Gaikwad[2], Shivani Gunjal[3], Prashant Gawali4, Abhay R. Gaidhani[5]

Under Graduate Student BE IT SITRC, Information Technology, Sandip Institute of Technology and Research Centre[1-4]

Professor Department of Information Technology SITRC, Information Technology, Sandip Institute of Technology and Research Centre[5]

Abstract: Modern society places a high importance on online education because of how quickly technology is developing and how education must change to keep up. E-learning is the only option left following the COVID-19 pandemic to keep instruction going during lockdowns, though. Artificial plays an important role in it. The avoidance of unfair means occurring during online exams is one of the most challenging circumstances exam invigilators encounter. Some of the issues can need consulting nearby references or perhaps getting assistance from neighbours. The principles of facial detection and recognition by Local Binary Pattern Histogram Algorithm, Dlib, toolkit, OpenCV library, and YOLOv3 are used in this research to offer a smart invigilation system that can facilitate exam enrollments and eliminate methods of impersonation and cheating. The evaluation of responses, particularly those of the subjective variety, is one of the main difficulties of online exams. Subjective responses gauge a student's capacity for information retention and verbal expression. Subjective questions, in contrast to objective questions, may have more than one valid response. These responses can state the same thing in a different language and grammatical structure. As a result, grading subjective questions manually takes a lot of time and is difficult to automate. This work uses machine learning (ML) and natural language processing (NLP) to automatically grade subjective questions. The objective response and the ideal response offered by the body that formulated the question were contrasted in the study.

Keywords: Machine Learning, Natural Language Processing (NLP), artificial intelligence, facial detection

I. INTRODUCTION

Machine learning (ML) approaches are a group of algorithms that aim to identify patterns in data and link those patterns to specific classes of samples in the data. For example, an ML model can predict whether a person is healthy or sick based on a set of features describing them, or it can predict whether an animal will receive treatment or not based on features describing it, or it can determine whether molecules have the potential to interact or not. Such patterns can also be found by ML techniques in an agnostic way, that is, without any knowledge of the classes. These techniques are known as supervised and unsupervised machine learning, respectively. Reinforcement learning is a third type of machine learning (ML) that looks for a series of steps that help achieve a certain objective together. Using cutting-edge computational techniques, we explain a number of supervised and unsupervised machine-learning strategies and illustrate a number of archetypal cases. Since reinforcement learning is complicated, it is not covered in full here; interested readers are directed to several very good reviews of the subject. Since our aim is to draw biomedical researchers' attention to the abundance of potent machine learning techniques and their potential to support both basic and applied research programs, we concentrate on concepts rather than processes. A statistical model is trained by an algorithm to generate predictions about an unlabeled instance in supervised learning.

Making computers understand natural language is the fundamental goal of natural language processing. But that's not a simple task. Structured data, such as spreadsheets and database tables, may be understood by computers. However, unstructured data, such as human languages, writings, and voice recordings, is more difficult for computers to grasp, necessitating the use of natural language processing. Natural language data is widely available in a variety of formats, and if computers could comprehend and interpret it, things would get a lot easier. There are various methods by which we can train the models to produce the desired results. It would be fantastic if computers could comprehend the vast amount of literature that has been produced by humans over thousands of years of writing. However, the task will never be simple. There are a lot of challenges out there, such as coreference resolution (which is, in my opinion, the most difficult thing), accurate Named-Entity Recognition (NER), accurate sentence comprehension, and accurate part-speech prediction. True human language understanding is impossible for computers. A well-trained model can distinguish and attempt to classify different parts of speech (noun, verb, adjective, supporter, etc.) based on inputted data and experiences

DOI: 10.17148/IJARCCE.2023.121008

if sufficient data is fed into it Artificial Intelligence (AI) has emerged as a transformative force across numerous domains, redefining how we interact with technology and information. Within AI, the role of natural language processing (NLP) and conversational agents, such as Platforms and virtual assistants, has gained significant prominence.

A key element that underpins the efficacy of these systems is prompt engineering, an advanced process that involves crafting input queries or prompts to solicit meaningful and contextually relevant responses from AI models.

II. HISTORY & BACKGROUND

Preliminaries: -

The number of exams and tests that must be graded likewise rises as student enrolment in educational institutions keeps rising, which adds time and effort to the professors' grading of the papers. Thus, it can be argued that an online grading system is a useful remedy for this issue, removing the extra time and effort that teachers could instead devote to other, more crucial areas.

The principles of human intellect that have been duplicated form the basis of artificial intelligence (AI) technology. Universal Artificial Intelligence, which cannot exist without combining self-awareness and self-cultivation components into Artificial Neural Networks, operates on the principle of dual contingency. It is recommended that intelligent agents be created on the IBM Bluemix platform using IBM Watson technology. Within the chatbot's Moodle learning management system, these agents must automate communication between the student and the teacher.

Md.Arafat Sultan in [12] and Bachman in [5] discuss the use of NLP to grade the responses submitted by the students. Md. Arataf and co-authors proposed a system wherein they provide a high-accuracy short answer grading system, by using various features like semantic vector similarity which enables integration of finer-grained lexical similarity measures, question decoding, term weighting, and length ratio which at last captures the comparative length of the answers provided by the student and model answer. Moreover, there is a large scope for improvement in factors like modality and polarity which can go undetected.

In [11], Xinhua Zhu suggests a mechanism for automatically scoring short answers. due to the absence of the ASAG corpus (Automatic Short Answer Grading). They suggested using a pre-trained BERT model to score the exams. Finding datasets in other foreign languages appears to be quite challenging because there aren't enough corpora or datasets to fully train a model. Therefore, Leila Ouahrani proposes a grading system for Arabic languages in [8] by gathering a dataset consisting of 5 different types of questions, 2133 student responses to each of these 5 types of questions, along with model answers, and using them to develop a grading system for foreign languages.

A real-time system that could track and categorise human face and mouth gestures was described by Nuria et al. in [4]. They implemented Hidden Markov Models, or HMMs, in their system and used 2-D blob characteristics to classify head movements and other expressions [4]. Their method [4] has undergone sufficient testing and has been able to classify objects with 100% accuracy.

III. METHODOLOGY

Software used for grading uses the model-and-similarity technique.

Two essential elements in the development of the suggested automatic subjective answer grading program utilising machine learning are the bert-base-nli-mean-tokens model and cosine similarity.

Algorithm 1 For the software, the authors employed the Cosine Similarity Technique and the Bert-Basenli-Mean-Tokens model. Google created a pre-trained language model called Bert-base-nli-meantokens. It is tailored for natural language inference (NLI) problems and is based on the transformer architecture.

Its underlying method makes use of a deep neural network with attention mechanisms, which enables the model to focus on various elements of the input sentence when producing a prediction. The model was trained to foretell whether the relationship between two sentences is one of entailment, contradiction, or neutrality during fine-tuning.

The "mean-tokens" part of the name refers to the fact that the model outputs a fixed-length vector representation (mean pooling) of the token-level representations for a given input sentence. This fixed-length representation was then used as input for the NLI task

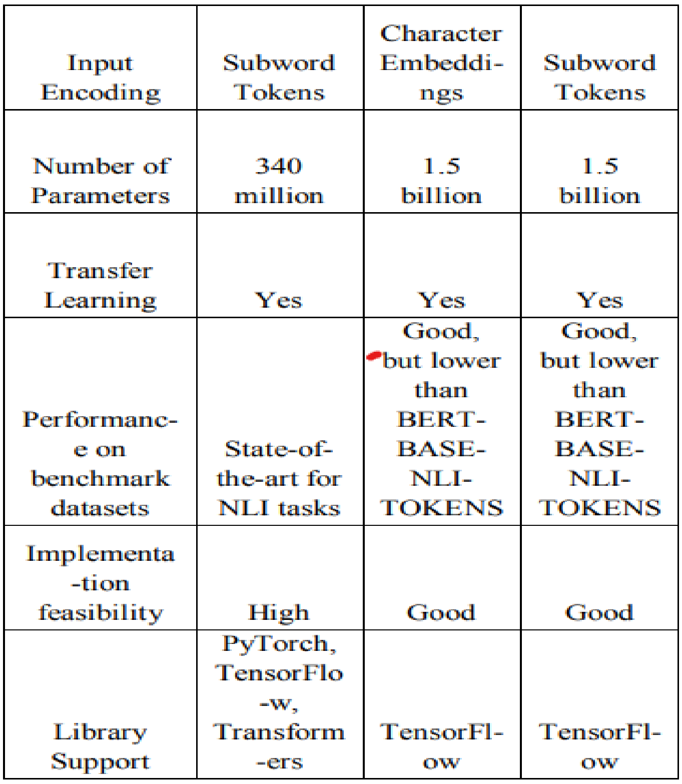
Cosine Similarity is used to measure the similarity between two vectors in an inner product space. It is used to determine whether two vectors are pointing in the same direction. In this paper's use case, it was used to measure the similarity between two texts. Given two vectors, the cosine similarity score was calculated as the dot product of the vectors divided by the magnitude of each vector. The resulting score ranges from -1 to 1, with a score of 1 indicating complete similarity and a score of -1 indicating complete dissimilarity. The formula to find the cosine similarity between two vectors 'x' and 'y' is – where x.y = dot product of the vectors 'x' and 'y'. ||x|| and ||y|| = length of the two vectors 'x' and 'y'. ||x||*||y|| = cross product of the two vectors 'x' and 'y'.

Working explained: The bert-base-nli-mean-tokens model was used to obtain vector representations of student responses and reference answers in the context of computerised short answer grading. Cosine similarity was then used to compare these representations to assess how closely the student responses matched those of the references. The cosine similarity score can be used to rate student answers automatically and as a proxy for answer qualityA model called Bert-basenli-mean-tokens was previously trained on a sizable corpus of text, enabling it to learn a general representation of language. The model is able to further specialise its representations during fine-tuning for the automatic short response grading assignment, better capturing the meaning of short answers. The model was able to make use of its pre-training and fine-tuning by comparing student answers to reference answers using these representations, which allowed it to carry out automatic short-answer grading with high accuracy. This makes it an effective tool for educational systems and applications that use machine learning.

Therefore, by incorporating cosine similarity into the automatic short answer grading system the authors leveraged the strengths of both the bert-base-nli-meantokens model and cosine similarity. The pre-trained BERT model provides strong representations of the student answers, while cosine similarity provides a simple and effective way to compare the similarity of these representations to reference answers. Machine learning is leveraged in this process by fine-tuning the BERT model on a short answer grading task, allowing it to learn to generate representations that are well-suited for comparison using cosine similarity.

IV. RESULTS & OBSERVATIONS A. Model Selection The authors compared a variety of models and then chose the bert-base-nli-mean-tokens model for their automatic short answer grading software based on its comparatively better performance on benchmark datasets for NLI Tasks.

IV. SUGGEST DIFFERENT SOLUTIONS BASED ON A COMPARATIVE ANALYSIS OF THE EXISTING SYSTEM



Input Encoding	Subword Tokens	Character Embeddi-ngs	Subword Tokens
Number of Parameters	340 million	1.5 billion	1.5 billion
Transfer Learning	Yes	Yes	Yes
Performanc-e on benchmark datasets	State-of-the-art for NLI tasks	Good, but lower than BERT-BASE-NLI-TOKENS	Good, but lower than BERT-BASE-NLI-TOKENS
Implementa-tion feasibility	High	Good	Good
Library Support	PyTorch, TensorFlo-w, Transform-ers	TensorFl-ow	TensorFl-ow

DOI: 10.17148/IJARCCE.2023.121008

1. Suitability for text data: Cosine similarity is well suited for finding similarity between texts as it considers both the meaning and structure of the texts. Euclidean distance and Jaccard similarity are not well-suited as they do not consider the meaning of the texts.

2. Interpretability: All three techniques have good interpretability as they are easy to understand and interpret.

3. Computational Complexity: Cosine similarity has medium computational complexity, while Euclidean distance and Jaccard similarity have low computational complexity.
Hence, cosine similarity was selected for the software as it is well-suited for text data and has good interpretability

V. CONCLUSION

In general, the smart test invigilation system was successful in enrolling and recording the attendance of students who must take part in online exams. It was quite accurate at tracking student eye contact and identifying when someone talked (while taking into account whether the exam was muted or whether someone tried to whisper or mouth something). People can rely on this system since it can warn authorities when specific student acts are suspected of being unfair, making it easy to monitor multiple pupils.

This paper has introduced 'base-level' work and concepts that have numerous opportunities for better performance and development in the future and can be upgraded to produce a relaxed, impartial online exam invigilation system In general, the smart test invigilation system was successful in enrolling and recording the attendance of students who must take part in online exams. It was quite accurate at tracking student eye contact and identifying when someone talked (while taking into account whether the exam was muted or whether someone tried to whisper or mouth something). People can rely on this system since it can warn authorities when specific student acts are suspected of being unfair, making it easy to monitor multiple pupils.

This paper has introduced 'base-level' work and concepts that have numerous opportunities for better performance and development in the future and can be upgraded to produce a relaxed, impartial online exam invigilation system.

1. A Student-Teacher Dashboard, where a teacher can create questions as a singleton or a group for a test and at the same time view scores of students graded by the BERT model.

2. Since no ML Algorithm is 100% accurate, the web app also allows teachers to edit scores graded by the model in case he/she feels the need to do so, thus getting a perfect amalgamation of Machine learning and human touch.

3. A Student dashboard is provided which allows students to answer the question(s) created by teachers and assess their grades once graded by the model. Real-time changes would be reflected if the teacher updates the student's score.

Creating an online system which can use AI for creating questions, monitoring students, using facial recognition on run time, and creating a secure online system for the university

REFERENCES

[1]. Burrows, S., Gurevych, I., Stein, B.: The eras and trends of automatic short answer grading. Int. J. Artif. Intell. Educ. 25, 60– 117 (2015).
[2]. Attali, Y., Powers, D., Freedman, M., Harrison, M., Obetz, S. (2008). Automated scoring of short-answer open-ended GRE subject test items. Technical Report RR-08-20. Princeton: Educational Testing Service.
[3]. Hasanah, U., Permanasari, A.E., Kusumawardani, S.S., Pribadi, F.S.: A review of an information extraction technique approach for automatic short answer grading. In: International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), pp. 192–196. IEEE (2016).
[4]. Dzikovska, M., et al.: SemEval-2013 task 7: the joint student response analysis and 8th recognizing textual entailment challenge. In: Seventh International Workshop on Semantic Evaluation, pp. 263–274 (2013).
[5]. Bachman, L.F., Carr, N., Kamei, G., Kim, M., Pan, M.J., Salvador, C., Sawaki, Y. (2002). A reliable approach to automatic assessment of short answer free responses. In S.C. Tseng, T.E. Chen, Y.F. Liu (Eds.), Proceedings of the 19th international conference on computational linguistics, volume 2 of COLING '02 (pp. 1–4). Taipei: Association for Computational Linguistics.
[6]. Magooda, A., Zahran, M.A., Rashwan, Raafat, H., Fayek, M.B.: Vector-based techniques for short answer grading. In: International Florida Artificial Intelligence Research Society Conference Ahmed, pp. 238–243 (2016).

DOI: 10.17148/IJARCCE.2023.121008

[7]. Mohler, M., Bunescu, R., Mihalcea, R.: Learning to grade short answer questions using semantic similarity measures and dependency graph alignments. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 752–762 (2011) 13. Passero, G., Haendchen Filho, A., Dazzi, R.: Avalia¸c˜ao d.

[8]. Leila Ouahrani and Djamal Bennouar. 2020. AR-ASAG An ARabic Dataset for Automatic Short Answer Grading Evaluation. In Proceedings of the Twelfth Language Resources and Evaluation Conference, pages 2634–2643, Marseille, France. European Language Resources Association

[9]. Linzhong XIA*, Jianfeng YE, De'an LUO, Mingxiang GUAN, Jun LIU, Xuemei CAO: Short text automatic scoring system based on BERT-BiLSTM model.In: Journal of Shenzhen University Science and Engineering, Volume 39, Issue 3: Pages 349-354 (2022)

[10]. Saha, T. I. Dhamecha, S. Marvaniya, R. Sindhgatta, and B. Sengupta, "Sentence level or token level features for automatic short answer grading?: Use both", Proc. Int. Conf. Artif. Intell. Educ., pp. 503-517, 2018.

[11]. X. Zhu, H. Wu and L. Zhang, "Automatic Short-Answer Grading via BERT-Based Deep Neural Networks," in IEEE Transactions on Learning Technologies, vol. 15, no. 3, pp. 364- 375, 1 June 2022.

[12]. . Md Arafat Sultan Cristobal Salazar Tamara Sumner "Fast and Easy Short Answer Grading with High Accuracy" in Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics.

[13]. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin "Attention Is All You Need" in 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

[14]. Ren, X., Liu, Y., & Wang, J. (2019). Automated ShortAnswer Grading Using BERT. arXiv preprint arXiv:1911.05507

[15]. . Chen, T., Yu, H., Lu, C., & Tsai, C. (2020). An Empirical Study on Automatic Short Answer Grading using a Pre-trained Language Model. arXiv preprint arXiv:2012.02634.

[16]. Nikhat Parveen, Saketh Ranga, Gouni Nishanth, Chaluvadi Sai Abhijith, Athmakur Harish Kumar Reddy (2021), Work Force Management System Using Face Recognition, Turkish Journal of Computer and Mathematics Education, Vol.12 No.9 (2021), 56-61 p

[17]. A. Ahmad, N. U. Khan, and A. W. Abbas, "PHP+MySQL based online examination system with power failure handling and dropbox capability," in Proceedings - 7th International Conference on Software Security and Reliability Companion, SERE-C 2013, 2013, pp. 21–25. doi: 10.1109/SERE-C.2013.27

[18]. Ugwitz, P., Kvarda, O., Juříková, Z., Šašinka, Č. and Tamm, S. (2022). Eye-Tracking in Interactive Virtual Environments: Implementation and Evaluation. Applied Sciences, 12(3), p.1027.

[19]. Meng J., WANG Y.H., "The Design of Intelligent Test Paper Generation for English Test Based on Genetic Algorithm," Theory and Practice of Education, 41(33):61-64,2021

[20]. S. Prathish, S. Athi Narayanan, and K. Bijlani, "An intelligent system for online exam monitoring," in Proceedings - 2016 International Conference on Information Science, ICIS 2016, Feb. 2017, pp. 138– 143. doi: 10.1109/INFOSCI.2016.7845315.