e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

REVIEW ONLINE EXAMINATION SYSTEM USING ARTIFICIAL INTELLIGENCE

Vaibhav Gawali[*1], Nishant Patil[*2], Ritesh Deshmukh[*3], Ramdas Pawade[*4], Dr. Raise Khan[*5]

[*1,2,3,4]Under Graduate Student, B.Tech CSE SUN, Computer Science & Engineering, School Of Computer Science And Engineering, India.

[*5]Professor, Department of Computer Science & Engineering SUN, Computer Science & Engineering, School of Computer Science & Engineering, India.

DOI : https://www.doi.org/10.56726/IRJMETS63556

ABSTRACT

In today's world, online education has become very essential because of the fast growth in technology. Across the globe, as schools were shut down to contain the COVID-19 virus spread quickly led realization of the need for e-learning mediums too. While teaching in this rapid shift, the role of Artificial Intelligence (AI) comes into play to give academic integrity as an integral feature – especially when we come for online examinations that make one question about how not every student is cheating! Students taking online exams often try to cheat by referring to books, asking people in the room for answers, etc.

This can be implemented using AI tools and algorithms like a Local Binary Pattern Histogram, Dlib toolkit, and OpenCV library for facial detection and recognition, which offer prior experience in building smart invigilation systems. The aim of this system is to simplify examination processes by validating student identity and minimizing cases of impersonation or fraud.

One of the major obstacles in this regard is assessing subjective answers, which determines whether a student has understood and remembered that information to briefly express it. Where objective questions have fixed answers but subjective ones may be corrected differently. Grading these responses manually can be very time-consuming and automating this process is not always straightforward.

To address this, our system leverages the power of machine learning (ML) and natural language processing (NLP), to automatically grade subjective answers by matching a student's response with an ideal answer provided by the exam creator.

Keywords: Machine Learning, Natural Language Processing (NLP), Artificial Intelligence, Facial Detection

I. INTRODUCTION

Machine Learning falls within the field of AI, and it involves using algorithms to identify patterns in data. For instance, ML can be used to predict whether a human is healthy or if those two molecules bind to each other. Broadly speaking, One can categorize ML techniques into supervised learning which trains a model using labeled data in order to make predictions, and unsupervised learning which attempts to identify structure in unlabeled datasets. [Reinforcement learning — With reinforcement learning the model learns a set of steps to reach an ultimate goal, which we will not cover in detail here]

Natural language processing or NLP is the sub-field of Artificial intelligence (AI) to make computers understand and work with human languages, which are unstructured. Structured data, such as spreadsheets may be easier for computers to process automatically while analyzing human languages is much more difficult. NLP methods assist in machines understanding the text, voice recordings — language data that helps to communicate with us. Though tasks such as Sentence Meaning and Named-Entity Recognition are tricky problems to solve in NLP, every other standard Piece of Speech task relies on them.

Artificial intelligence AI, such as NLPs and talking tools (representative assistants) changed the method we have a tendency to act with technology. One of the things that start everything is "prompt engineering": creating queries or questions in a way to be understood and answered by an AI (or get back what you expect from it).

II. METHODOLOGY

For example, our grading software uses a model of similarity. Our subjective solution uses the BERT-base-nli-mean-tokens model, and on top of that, we apply Cosine Similarity for grading.

1. **BERT-base-nli-mean-tokens Model:** It is a model for understanding the relationships between sentences in language proposed (5) by Google. It employs an attention-based deep neural network which allows the system to pay different amounts of attention to parts of a sentence. The "mean-tokens" part tells the model that it should extract a fixed-length vector (i.e., summary representation) of this sentence.
2. **Cosine Similarity:** Cosine similarity measures the cosine of the angle between two vectors (or sentences). A score of 1 means that they are completely similar, while -1 stands for very different. It allows us to check how close a student's answer is to the correct answer.

How it works: Given student answers and reference answers, the BERT model embeddings these into vectors. The student's answer is then compared with the reference for which Cosine similarity looks at these vectors and tells how similar they are or in other words what percent directions of two vectors vary from each other. This seven is a proxy measure of the quality of that answer.

Short-answer grading: Pre-trained BERT model fine-tuned to assess short answers, which enables it to a broad language understanding of the type and meaning embedded in short-form text. Put together, we have a stable grading tool for educational applications using the powerful language representations of BERT and simple cosine similarity comparison.

III. PRELIMINARIES: HISTORY & BACKGROUND

It also adds more exams to be graded for teachers, making it that much harder and time-consuming to grade all of the papers. This problem can be resolved by using an online grading system so that teachers are able to concentrate all focus on more important things instead of wasting time.

AI works on the principles like Human Intelligence. A systematically integrated capability of Artificial Neural Network is necessary for Universal AI, which would include self-awareness. IBM Watson for example on the IBM Bluemix cloud platform is an ideal answer to implement intelligent agents that can automate communication between students and teachers in e-learning systems such as MOODLE.

Specifically, we have already tackled NLP in grading student answers. Md. Arafat Sultan et al presented an approach to grade short answers that employed similarities such as word meaning or sentence structure in responses for a better explanation of the human reference answer (ground truth). Similarly, further research such as that of Zhu [55] supports the use of pre-trained BERT for short answer grading while different language data is absent which makes its implementation limited. Leila Ouahrani desperately tried to seed a vocabulary in her head as she collected answers from students across all question types for Arabic, which then finely tuned the grading model for other languages.

Nuria et al. built a system to be able to recognize and classify facial expressions and posters. They succeeded — the new system could predict head movements and facial gestures with much better accuracy, using Hidden Markov Models for improved online exam monitoring.

IV. RESULTS & OBSERVATIONS

The authors tried different models and finally used BERT-base-nli-mean-tokens because it was the best on benchmark datasets for the task of understanding relations between two sentences. Cosine similarity along with this model proved to be a good approach for auto-grading short answers.

Suggest Different Solutions Based On A Comparative Analysis Of The Existing System

Input Encoding	Subword Tokens	Character Embeddings	Subword Tokens
Number of parameter	340 million	1.5 billion	1.5 bilion
Transfer learning	YES	YES	YES
Performance on benchmark datasets	State-of-the-art of the NLI tasks	Good but lower than BERT-BASE-NLI-TOKENS	Good but lower than BERT-BASE-NLI-TOKENS

Implementation feasibility	High	Good	Good
Library support	PyTorch tensor Flow transformers	TensorFlow	TensorFlow

Suitability for Text Data:

The cosine similarity is suitable for comparison of text because it preserves the meaning and structure of text in a way other methods like Euclidean distance and Jaccard similarity fail to preserve meaning contained in texts.

Interpretability:

All three methods—cosine similarity, Euclidean distance, and Jaccard similarity—are easy to understand and interpret

Computational-Complexity:

Cosine similarity is medium complexity; Euclidean distance and Jaccard similarity are relatively low.

V. CONCLUSION

The smart exam invigilation system successfully manages student enrolment and attendance for online exams. It accurately monitors students' eye movements and can detect when they speak or make gestures, even distinguishing whispers or lip movements. This system helps invigilators track multiple students at once and flags any suspicious activity for review, making online exams more secure and fair.

This paper introduces a foundation for future development with many possibilities to improve the system and make online invigilation more reliable and relaxed.

Some potential enhancements include:

1. **Student-Teacher Dashboard:** Teachers can create test questions and view students' scores graded by the BERT model.
2. **Manual Score Adjustment:** Since no ML algorithm is perfect, teachers can manually adjust scores if they feel it's necessary, blending machine learning with a human touch.
3. **Student Dashboard:** Students can view and answer test questions created by teachers and see their scores once graded by the model. If a teacher updates a score, it will reflect in real-time.

This online system could further evolve to use AI for creating questions, real-time facial recognition, and secure exam monitoring, making it suitable for university-level online exams.

VI. 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 openended 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).

[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.