**NLP-based Automatic Answer Evaluation**

PAPER LINK: [NLP-based Automatic Answer Evaluation | IEEE](https://ieeexplore-ieee-org.ezaccess.libraries.psu.edu/document/9754052)

Q. What the aim of the research?  
  
Ans. The aim of this research paper is to understand how automatic evaluation of answer scripts using NLP can provide a quicker, more efficient, and unbiased method for assessing student performance. NLP technologies such as text summarization and similarity metrics can be used to evaluate student responses and determine score marks. These NLP based methods are timesaving, less dependent on the evaluator and can avoid any bias that the individual evaluator could have. The use of NLP in answer script assessment is a promising solution to the challenges of manual evaluation and can provide a more accurate and efficient assessment process.

Q. What are the scientific challenges?  
  
Ans. The approach used in the research paper is unsupervised and only a human can validate the accuracy and performance of the model, the final formula for calculating marks could simply be rectified to get the desired results. An unsupervised approach had been considered due to unavailability of a large dataset.

Q. Describe the contribution of this research?  
  
Ans. The research paper talks about how the data must be prepared to improve the similarity measure. One of the methods the author comes up with is the N-Gram format. This function forms the foundation of the algorithm. It performs the n-gram profiling of document centered about the words which are most used. The algorithm for different values of ‘n’ are tried and tested, and the best and most interpretable results are formed with bigrams (n=2). The function then returns a list of these bi-grams and trigrams. They are to be further used while evaluating similarity between two documents. This is a preprocessing step that could be useful for the research project we are working on as well.

Q. How the proposed contribution differs from related work published in different venues?  
  
A. The paper talks about four papers that focus on the development of automatic short question evaluation systems for grading objective or brief questions. Ahmed Magooda et al. (2014) developed a vector-based approach that utilized sentence representation strategies and a range of similarity metrics such as string similarity, knowledge-based similarity, and corpus-based similarity.

Md Arafat Sultan et al. (2016) created a simple brief question evaluation approach that focused on semantic similarity between the student answer and the correct answer. The performance was evaluated using the Mohler et al. dataset and a simpler bag-of-words model, and the results showed that the proposed model outperformed the bag-of-words model.  
Michael Mohler and colleagues (2015) created a methodology for automated objective answer marking using unsupervised methodologies. They considered knowledge-based and corpus-based similarity and integrated automated feedback from student responses to enhance performance. Although the model outperformed previous models, it did not consider grammatical and spelling errors.  
  
Jonathan Nau et al. (2017) proposed a system for answering brief questions automatically in Portuguese using linear regression with a combination of latent semantic analysis and a WordNet path-based similarity metric. The anticipated scores were compared to student marks and proved to be beneficial in their combination strategy.

Bibliographical references:

* A. Magooda, M.A. Zahran, M. Rashwan, H. Raafat and M.B. Fayek, “Vector Based Techniques for Short Answer Grading”, Proceedings of the Twenty Ninth International Florida Artificial Intelligence Research Society Conference, 2014, pp.238-243
* M.A. Sultan, C. Salazar, and T. Sumner, “Fast and Easy Short Answer Grading with High Accuracy,” Proceedings of NAACL-HLT, 2016, pp.1070–1075
* M. Mohler and R. Mihalcea, “Text-to-text Semantic Similarity for Automatic Short Answer Grading”, International Journal of Artificial Intelligence in Education 25, 2015, 118 – 125
* J. Nau, A. H. Filho, and G. Passero, “Evaluating Semantic Analysis Methods for Short Answer Grading Using Linear Regression, “PEOPLE International Journal of Social Sciences”, 2017, Volume 3 Issue 2, pp. 437 – 450.

Q. Assess the weakness and the strength of the contribution?  
  
A. The major weakness in the research paper is the lack of a large dataset because of which an unsupervised learning algorithm had to be used. Another weakness of the paper is how they conclude that cosine similarity gave the best approximation without showing a comparison of other similarity measures.  
  
Strengths of the research paper is that they can build a system that checks subjective answers and allocates marks like a human is created. The approach that was taken also has far reaching consequences and could be employed in other areas too. For example, plagiarism between two documents could also be calculated using the approach following fine tuning.

Q. Explain how these papers are relevant to your project? What are the learned lessons and how do they help you in developing, training, and testing your project's models.  
  
A. The research paper described above provides a pipeline for how the algorithm works and provides an insight on what the key processes that need to be taken care of for the algorithm to run efficiently. Even though the dataset on which the research paper was based on was unsupervised unlike the one that we propose on using, the preprocessing and cleaning steps are something that would be we could take inspiration from.