

# Feature extraction and LSA in automatic essay evaluation

JOÃO SANTOS<sup>1</sup>, ELOI FAVERO<sup>2</sup>, SILVÉRIO SIROTHEAU<sup>3</sup>, AND SIMONE FREITAS<sup>4</sup>

<sup>1</sup>*Faculty of Mathematics – Federal University of Pará – Belém - Pará – Brasil, jcas@ufpa.br*

<sup>2</sup>*Faculty Computer Science – Federal University of Pará – Belém - Pará – Brasil, faver@ufpa.br*

<sup>3</sup>*Faculty Petroleum Engineering – Federal University of Pará – Salinópolis - Pará – Brasil, ssirotheau@gmail.com*

<sup>4</sup>*Faculty of letters – Federal University of Pará – Castanhal - Pará – Brazil, sinegrabr@gmail.com*

## 1 Abstract

The technological advances in the field of education have enabled a variety of online activities, among which are the answers to discursive questions [2]. Such questions serve the purpose of verifying and gauging students' learning, in particular the abilities to write and structure the argumentative discourse [4, 3, 7]. Within this context, the development of approaches to foster automate correction gains relevance, as it presents advantages such as:

- i) it reduces the correction time, proving more stable in the correction procedure, reducing the possibility of errors;
- ii) it provides immediate feedback which can support students in solving questions;
- iii) it allows multiple submissions at a time and place determined by the user;
- iv) it has low cost [6].

The main focus of the experiments here presented is the development of a set of algorithms for the automatic evaluation of written answers to discursive questions based on Natural Language Processing and Artificial Intelligence techniques. The main techniques used:

1. The initial work was to apply a latent semantic analysis (LSA) model for the automatic evaluation of short answers (25 to 70 words) to open-ended questions of the entrance examinations from the Federal University of Pará (Ufpa) [1];
2. Next, we conducted experiments using a method based on  $n$ -gram similarity and a categorization process [5];
3. Next, we conducted experiments using a method based on machine learning techniques to a corpus constituted of a sample covering 1000 essays from a public exam for admission to the career of Administrative Technician in Education.

A system is considered to perform well when the system versus human agreement index approaches or is higher than the human versus human agreement index. Considering a study corpus constituted of 1,000 (thousand) essays obtained from a selection process and previously evaluated by two human specialists, the quadratic kappa index (QK) among human evaluators was 0.68. For this same corpus, in our experiments, we obtained a system versus human index QK of 0.8412. In light of the results, we can consider that this technology is achieving a good performance, thus, the method can be used in combination with human evaluation in real tests with written answers to discursive questions.

## Bibliography

- [1] João Alves dos Santos and Eloi Favero. Practical use of a latent semantic analysis (lsa) model for automatic evaluation of written answers. *Journal of the Brazilian Computer Society*, 21, 12 2015.
- [2] Christine Chin and Jonathan Osborne. Students' questions: a potential resource for teaching and learning science. *Studies in Science Education*, 44(1):1–39, 2008.
- [3] M. Monjurul Islam and A. S. M. Latiful Hoque. Automated essay scoring using generalized latent semantic analysis. In *2010 13th International Conference on Computer and Information Technology (ICCIT)*, pages 358–363, 2010.
- [4] Ellis B. Page. The imminence of... grading essays by computer. *The Phi Delta Kappan*, 47(5):238–243, 1966.
- [5] Silverio Sirotheau, Joao Santos, Eloi Favero, and Simone Freitas. Automated evaluation of short answers using text similarity for the portuguese language. *Journal of Computer Science*, 15:1669–1677, 11 2019.
- [6] Pierpaolo Vittorini, Stefano Menini, and Sara Tonelli. An ai-based system for formative and summative assessment in data science course. *International Journal of Artificial Intelligence in Education*, 31(4):497 – 513, 2021.
- [7] K. Zupanc and Z. Bosnic. Automated essay evaluation augmented with semantic coherence measures. In *2014 IEEE International Conference on Data Mining*, pages 1133–1138, 2014.