Efficient Redundancy Detection in Large-Scale Examinations: A SRNN Model for Enhanced Test Quality and Resource Optimization

Najoua HRICH*
Professor Researcher, Department of Computer Science
Regional Center for Education and Training Professions,
Tangier, MOROCCO
nhrich@uae.ac.ma
Mohamed AZEKRI
Pedagogical Inspector of Computer Science & Expert on assessment systems
Regional Academy for Education and Training, Ministry of National Education,
MOROCCO
Charafeddin ELHADDOUCHI
Doctor researcher on Management of Information System
Regional Center for Education and Training Professions,
Tangier, MOROCCO

Publication History
Manuscript Reference No: IRJCS/RS/Vol.11/Issue05/MYCS10088
Research Article | Open Access | Double-Blind Peer-Reviewed | Article ID: IRJCS/RS/Vol.11/Issue05/MYCS10088
Received: 10, April 2024 | Revised: 11, May 2024 | Accepted: 23 May 2024 | Published Online: 30, May 2024
http://www.irjcs.com/volumes/Vol11/iss-05/04.MYCS10088.pdf
Article Citation: Najoua,Mohamed,Chrafeddin(2024). Efficient Redundancy Detection in Large-Scale Examinations: A SRNN Model for Enhanced Test Quality and Resource Optimization. IRJCS: International Research Journal of Computer Science, Volume 11, Issue 05 of 2024 pages 459-462 **doi:>** https://doi.org/10.26562/irjcs.2024.v1105.04
BibTeX Najoua@2024Efficient

Abstract: This project aims to develop an artificial intelligence model capable of efficiently detecting and eliminating redundancies in multiple-choice tests used in large-scale competitive examinations. By exploiting the data processing capabilities of deep learning, this model aims to improve the quality and reliability of the tests. This will ensure a fairer and more equitable assessment of candidates' skills but also optimize the time and resources needed to design and revise items. The project aims to facilitate the work of test designers by automating part of the revision process while maintaining high standards of item diversity and relevance.

Keywords: MCTs, items, distractor, cobaying, redundancy, deep learning, Natural language process

I. INTRODUCTION

In the context of competitive examinations, the number of items specified is often very high, sometimes reaching 100 to 120 items, with four distractors for each. This abundance of items increases the risk of redundancy, which can compromise the validity and reliability of assessment tests. Faced with this challenge, competition organizers and educators have developed an emerging practice called "cobaying", in which test designers take the same test as the candidates to assess the quality of the questions. Despite this method of validation, there is still a risk that redundancies will escape the designers' vigilance, which could lead to challenges during the examination process. To address this issue, this research explores an innovative approach based on the use of deep learning techniques to detect and solve redundancy problems in MCTs. In this paper, we present an intelligent solution aimed at overcoming this problem by developing a specific tool capable of automatically analyzing and correcting redundancies in multiple-choice exams. In addition, we propose a detailed methodology and empirical results demonstrating the effectiveness of our approach, particularly in contexts requiring a large number of items, such as competitions and competitive exams [1],[2].

II. BACKGROUND

A. MCTs Based Assessment

MCTs have been widely adopted in educational assessment due to their numerous advantages also present challenges. On the benefits side, MCTs offer a practical and objective assessment of learners' knowledge and skills, enabling a wide range of topics to be tested relatively quickly and cost-effectively. Their standardized format facilitates comparing learners' performance and simplifies the grading process. Moreover, MCTs enable reliable and fair assessment, reducing the potential biases associated with other assessment formats. However, despite their advantages, MCTs also present challenges. They can sometimes encourage memorization of information rather than deep comprehension, and some critics argue that they are not as effective at assessing complex skills or critical thinking abilities. Furthermore, the development of high-quality MCTs requires significant expertise in item writing and psychometrics [3],[4],[5] to ensure their validity and reliability.

Consequently, although MCTs are widely used in educational assessment, it is important to recognize their advantages and limitations and take action to alleviate the potential drawbacks of this assessment format [6].

B. *Deep learning models for text classification*

Advances in natural language processing (NLP) and machine learning (ML) have enabled the development of models capable of analyzing and understanding text in depth. Transformers, in particular, have shown remarkable performance in various NLP tasks. The use of deep learning models to detect redundancy in texts represents a significant advance in the field of NLP [7]. These models exploit deep neural architectures to extract relevant features from text data, enabling them to identify similarities and repetitions within text content. Architectures commonly used for this task include recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformers.

- RNNs adapt to model sequential data, making them relevant for detecting repetitive patterns in texts. For example, an RNN model can be used to identify sentences or passages that are repeated throughout a document. [8].
- CNNs extract local features from text data. They can detect similarities between text passages by analyzing local patterns. For example, a CNN can be used to identify similar sentences in a set of documents [9].
- Transformers have revolutionized the field of NLP by introducing attention mechanisms. They can consider relationships between words across the whole text, making it easier to detect redundancy across long distances. For example, a transform model can be used to identify text passages that are repeated at different points in a document. [10].

These deep learning models offer different yet complementary approaches to detecting redundancy in texts, by analyzing word sequences, local features, and long-distance relationships between words.

III. PROPOSED APPROACH

In this research, given the relative simplicity of the task of detecting redundancies in test items and distractors, we chose to use a Simple Recurrent Neural Network (SRNN) [11]. This architecture, with its simple structure and ease of implementation, offers an effective approach to this problem. Although it may have limitations in capturing long-term dependencies, these are less critical in the context of syntactic redundancy detection. By opting for this approach, we aim to take advantage of the training speed of the SRNN[12] while targeting adequate performance for the specific task of our research. The process of our proposal is shown in Fig 1:

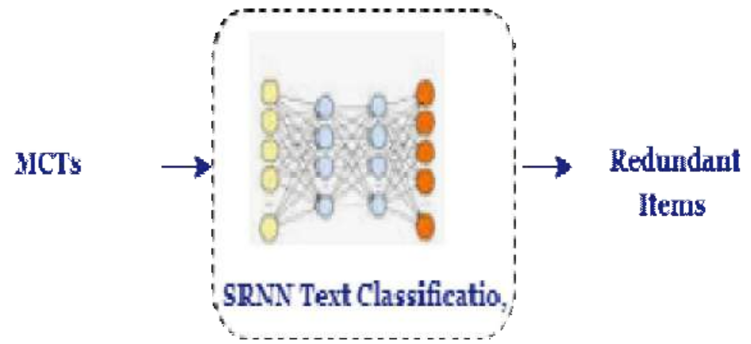

Fig. 1 Proposal approach

IV. METHODS AND MATERIALS

To contribute to improving item quality by identifying and eliminating duplications, the model is developed under Python, the most used and famous programming language in data science. Python is a high-level programming language, and its basic design philosophy is based on code readability and syntax that allows programmers to express concepts in just a few lines of code. Python is an open-source license, making it freely usable [13]. This model was designed to identify duplicate items within an MCT, using a Simple RNN, a specific variant of RNNs. then; we created a training dataset using MS Word. To optimize the quality of the information provided to the model, the items in this dataset are numbered from Q1 to Q1000, and each item is composed of 4 distractors A, B, C, and D. Finally, we used our model for 10 tests of 100 items each, to evaluate its effectiveness in detecting duplications and its practical application in a real assessment context. To train our deep learning model, we employed a standard method of splitting data into training and testing sets. Initially, we separated our training dataset, which comprises MCT questions and their corresponding distractors, into two parts: one for training the model and the other for validating the model. We allocated approximately 80% of the data for training and the remaining 20 % for validation. To evaluate our model, we opted for the use of two fundamental metrics in regression analysis: the Root Mean Square Error (RMSE)[14][15] and the Coefficient of Determination (R^2)[16][17]. The choice of these metrics stems from their ability to provide a comprehensive assessment of model performance in different respects. The RMSE quantifies the mean deviation between the values predicted by the model and the actual values of the target variable, providing a precise indication of the accuracy of predictions. In parallel, the R^2 measures the proportion of the variance of the dependent variable explained by the independent variables included in the model. This approach enables us to gain an in-depth understanding of the predictive and explanatory capacity of our regression model, facilitating informed decision-making in our scientific analyses.

V. RESULTS AND DISCUSSION

As previously mentioned, our model was subjected to a battery of 10 tests, each consisting of 100 questions. The details and performance of these assessments are given in Figures 2 and 3: figure 2 presents redundancy generated by the human process while Figure 3 presents redundancy generated by the AI-based model.

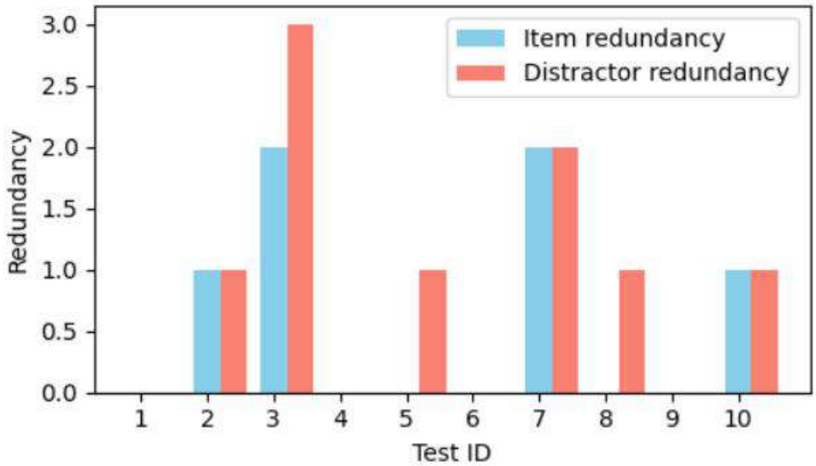

Fig. 2 Redundancy generated by human process

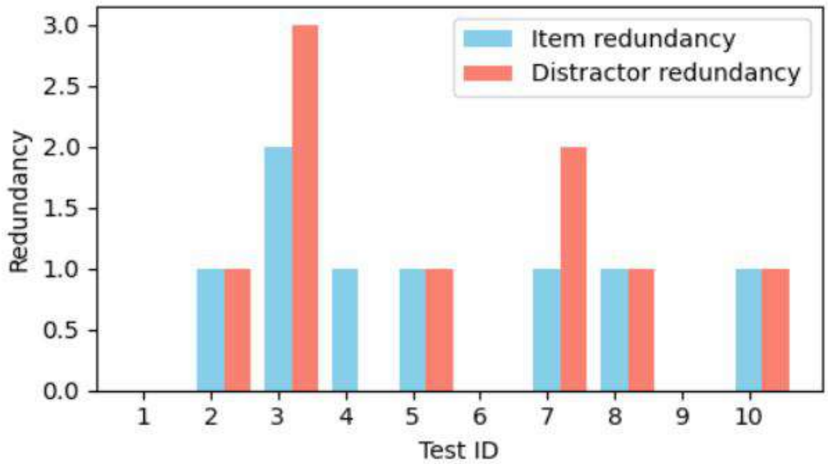

Fig. 3 Redundancy generated by AI-based model

To evaluate the effectiveness of our model, we implemented a Python script to calculate key performance measures for items and distractors. The execution of this script provides the subsequent outcomes as shown in figure 4.

```
Items :
   Correlation : 0.75
   RMSE : 0.55
   Coefficient of determination (R²) : 0.53

Distractors :
   Correlation : 1.00
   RMSE : 0.00
   Coefficient of determination (R²) : 1.00
```

Fig. 4 Statistical Results

For the items, the analysis revealed a correlation of 0.75, indicating a moderately strong linear relationship. The root-mean-square error (RMSE) of 0.55 suggests a moderate level of accuracy in predictions, and the coefficient of determination (R^2) of 0.53 highlights a reasonable degree of predictive power. Conversely, the distractors presented seemingly flawless results, with a perfect correlation of 1.00, an RMSE of 0.00, and an R^2 of 1.00. The performance indicators highlight a strong agreement between the data generated by humans and that of the proposed model. The high correlation, low Root Mean Square Error (RMSE), and a relatively high coefficient of determination R^2 testify to the notable effectiveness of the model in accurately predicting data compared to values generated by the human process.

ISSN: 2393-9842
https://www.irjcs.com/archives

VI. CONCLUSIONS

In conclusion, the article explored the innovative application of artificial intelligence models to improve the quality of assessment tests, focusing on the detection of redundancies in items and distractors. By integrating these models into the test development process, designers can not only speed up the piloting process but also guarantee greater validity and reliability of assessments. This approach also frees up valuable time and resources by automating tasks that were previously tedious and prone to human error. Ultimately, the integration of AI into test development marks a significant evolution, paving the way for more efficient and accurate methods of assessing learners' knowledge and skills.

REFERENCES

[1]. N. Hrich, M. Azekri, et M. Khaldi, « An AI Educational Tool for Detecting Redudancy in Distractors and Items within Multiples choice Tests. », INTED2024 Proceedings, p. 6454‑6458, 2024, https://doi:10.21125/inted.2024.1691.

[2]. N. Hrich, M. Azekri, et M. Khaldi, « An AI Educational Tool for Detecting Redudancy in Distractors and Items within Multiples choice Tests. », INTED2024 Proceedings, p. 6454‑6458, 2024, https://doi:10.21125/inted.2024.1691.

[3]. T. M. Haladyna et S. M. Downing, « A Taxonomy of Multiple-Choice Item-Writing Rules », Applied Measurement in Education, vol. 2, no 1, p. 37‑50, janv. 1989, https://doi:10.1207/s15324818ame0201_3.

[4]. S. M. Downing, « The Effects of Violating Standard Item Writing Principles on Tests and Students: The Consequences of Using Flawed Test Items on Achievement Examinations in Medical Education », Adv Health Sci Educ Theory Pract, vol. 10, no 2, p. 133‑143, juin 2005, https://doi:10.1007/s10459-004-4019-5.

[5]. M. C. Rodriguez, « Three Options Are Optimal for Multiple ‐ Choice Items: A Meta ‐ Analysis of 80 Years of Research », Educational Measurement, vol. 24, no 2, p. 3‑13, juin 2005, https://doi:10.1111/j.1745-3992.2005.00006.x.

[6]. M. J. Gierl, O. Bulut, Q. Guo, et X. Zhang, « Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review », Review of Educational Research, vol. 87, no 6, p. 1082‑1116, 2017, https://doi:10.3102/0034654317726529.

[7]. S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, et J. Gao, « Deep Learning Based Text Classification: A Comprehensive Review ». arXiv, 4 janvier 2021. https://doi:10.48550/arXiv.2004.03705.

[8]. X. Fu, E. Ch'ng, U. Aickelin, et S. See, « CRNN: A Joint Neural Network for Redundancy Detection ». arXiv, 4 juin 2017. https://doi:10.48550/arXiv.1706.01069.

[9]. Z. Wang, C. Li, et X. Wang, « Convolutional Neural Network Pruning with Structural Redundancy Reduction ». arXiv, 7 avril 2021. https://doi:10.48550/arXiv.2104.03438.

[10]. Z. Dai, G. Lai, Y. Yang, et Q. Le, « Funnel-Transformer: Filtering out Sequential Redundancy for Efficient Language Processing », in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2020, p. 4271‑4282.

[11]. J. Koutnik, K. Greff, F. Gomez, et J. Schmidhuber, « A Clockwork RNN », in Proceedings of the 31st International Conference on Machine Learning, PMLR, juin 2014, p. 1863‑1871.

[12]. Aurnhammer et S. L. Frank, « Comparing gated and simple recurrent neural network architectures as models of human sentence processing ». 19 septembre 2018. https://doi:10.31234/osf.io/wec74.

[13]. « Deep Learning with Python », Manning Publications. Consulté le: 4 janvier 2024. [En ligne]. Disponible sur: https://www.manning.com/books/deep-learning-with-python

[14]. M. Ćalasan, S. H. E. Abdel Aleem, et A. F. Zobaa, « On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: A novel exact analytical solution based on Lambert W function », Energy Conversion and Management, vol. 210, p. 112716, avr. 2020, https://doi:10.1016/j.enconman.2020.112716.

[15]. J. Willmott et K. Matsuura, « Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance », Climate Research, vol. 30, no 1, p. 79 ‑ 82, déc. 2005, https://doi:10.3354/cr030079.

[16]. What is R Squared? R2 Value Meaning and Definition », freeCodeCamp.org. [En ligne]. Disponible sur: https://www.freecodecamp.org/news/what-is-r-squared-r2-value-meaning-and-definition/

[17]. R-Squared, Corporate Finance Institute. [En ligne]. Disponible sur: https://corporatefinanceinstitute.com/resources/data-science/r-squared/