Recommendation System for Education: An Approach to Suggesting Learning Sequences

João Paulo Silva[1], Rodrigo Aleixo[1], Thiago Sernaglia[1], Marcos de Almeida[1], Rubens Massa[2], Everton Silva[3], and Douglas Castilho[1]

[1] Laboratory of Technology and Innovation (LATIN)
Federal Institute of South of Minas Gerais (IFSULDEMINAS)
Poços de Caldas, Brazil
[2] A Recreativa, Poços de Caldas, Brazil
[3] Federal Institute of São Paulo (IFSP), Campinas, Brazil

Abstract. Continuous lesson planning requires navigating large collections of pedagogical materials, making the identification of relevant resources a time-consuming process. Although digital platforms are allies in the search for teaching materials, they commonly face the cold-start problem, the initial difficulty in recommendation due to the lack of prior data for collaborative filtering. This study proposes and evaluates a content-based recommendation system to suggest educational activities, named Learning Sequences. We conducted a comprehensive analysis comparing traditional information retrieval techniques (Bag of Words, TF-IDF) and dense word embedding models, including static (Word2Vec, GloVe, FastText) and contextualized (BERT) representations. Additionally, the PageRank algorithm was adapted to operate on textual similarity graphs, based on the relevance of the documents, named as Global and Local PageRank. Experimental results show an efficient framework to recommend learning sequences using TF-IDF and cosine distance-based method.

Keywords: Recommendation System · Learning Sequence · Similarity Methods · Embedding Methods

1 Introduction

The lesson planning process in the Brazilian educational context tends to be complex and often laborious, especially due to the legal and pedagogical requirements established by regulatory documents such as the National Common Curriculum Base (BNCC) [3]. It sets out specific guidelines to be developed and implemented throughout basic education, which requires teachers to adapt their pedagogical activities to comply with these directives. In light of this context, digital educational platforms have emerged as important tools to support teachers in the creation and structuring of pedagogical activities. They contribute to making the teaching and learning process simpler, more efficient, and of higher quality, since the planned use of digital technologies enhances personalization and student engagement [15].

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_37

This study focuses on the development of methods based on text vectorization techniques for recommendation of pedagogical activities, named as Learning Sequences. The absence of user interaction logs or feedback of the pedagogical dataset precludes the application of traditional collaborative filtering approaches. Recommendation Systems (RS) play a central role in mitigating information overload, a common challenge in lesson planning, where teachers must select appropriate materials from a large pool of available resources [1]. In educational settings and newly launched platforms, collaborative filtering is particularly vulnerable to the cold-start problem , as it relies on historical user interaction data. In such scenarios, content-based filtering emerges as a more suitable alternative [10]. This was consolidated through Vector Space Models (VSM) [12], in which semantics are ignored, and documents are represented by the frequency of their terms. Term Frequency-Inverse Document Frequency (TF-IDF) scheme became widely adopted [5] [14]. Recent studies [16], demonstrate that TF-IDF remains a competitive and computationally efficient baseline for short-text recommendation tasks. Graph-based algorithms [9] have been widely applied in Information Retrieval (IR) models. The application of RS in educational settings requires richer semantic precision [6]. To address the limitations of traditional VSM in IR, the literature has adopted dense word embedding techniques, such as Word2Vec [7] and GloVe [11]. The effectiveness of embeddings in recommendation systems has been empirically supported by some studies [8]. The adoption of FastText [2] is motivated by its ability to mitigate the out-of-vocabulary (OOV) problem through subword modeling. Models such as BERT [4] have been successfully adapted to capture dynamic contexts in recommendation tasks.

Given this scenario, content-based recommendation methods were chosen, in which recommendations are generated based on the textual similarity with the pedagogical content. To enable the recommendation process, multiple textual representation methods and similarity measures were evaluated. The main contributions of this work are: *i)* educational recommendation method, designed to suggest Learning Sequences; *ii)* comprehensive comparative analysis of multiple textual representation techniques; *iii)* semantic PageRank adaptation, applied to textual similarity graphs.

2 Materials and Methods

The alignment of the lesson plan with the BNCC requires teachers to select pedagogical materials appropriate for the grade level, the age group of the students, and the subject area. This process becomes challenging on educational platforms. One major difficulty is the cold-start problem, in which the absence of prior usage history or user evaluations limits the effectiveness of traditional recommendation methods. In this context, the main research question of this work is: how can relevant learning sequences be recommended based solely on their textual characteristics?

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_37

Table 1: Attributes of the dataset and their type in the recommendation system

Type	Attribute	Description
Input	`description`	The teacher's raw input describing the desired activity.
Output	`title`	The title of the recommended learning sequence.
Output	`objectives`	The specific pedagogical goals of the lesson.
Output	`summary`	A brief overview of the learning sequence content.
Output	`resources`	The materials and tools required to execute the class.
Output	`steps`	The detailed, step-by-step instructional guide.
Metadata	`year`	The target educational grade level for the activity.
Metadata	`discipline`	The specific academic subject or discipline.
Metadata	`theme`	The central pedagogical topic addressed in the sequence.

2.1 Learning Sequence - Educational Data Set

The database used in this work was provided by the Brazilian company A Recreativa, which is focused on creating solutions for educational environments. The dataset is used to assist teachers and educators across Brazil in designing educational content for specific pedagogical topics, comprising basic descriptions and complete learning sequences. The dataset contains approximately $2,000$ learning sequences and it is structurally divided into three distinct subsets: *i)* Learning Sequence Description (Input), presenting the general description of the activity provided by the user. *ii)* Learning Sequence (Output), presenting the set of information related to the pedagogical content used by teachers; and *iii)* Metadata, used to categorize the learning sequences. Table 1 summarizes the main attributes of the dataset and their respective roles within the recommendation framework. The average length of the descriptions (input) is about 93 tokens, while the learning sequences (output) are close to $1,282$ tokens.

2.2 Tokenization and Embedding Methods

To convert textual data into computable vectors, we evaluated six distinct representation methods, ranging from classical sparse techniques to dense static and contextual embeddings. To ensure reproducibility, Table 2 summarizes the methods, specific configurations, and hyperparameters.

2.3 Recommendation Methods

This section presents the methods evaluated in this work for recommending learning sequences. We present two variations of the classic PageRank method, based on graphs, as well as a simple baseline for evaluating the analyzed methods in the context of recommending learning sequences: *i)* Distance Based Method - Cosine (COS), determines how similar the vectors are based on the angle between them; *ii)* Correlation Based Method - Pearson (COR), used to measure the strength of the linear relationship between continuous datasets. *iii)* Similarity Based Method - Jaccard (JAC), measures the occurrence of identical words between texts. This method is used only on BoW and TF-IDF. *iv)* PageRank

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_37

Table 2: Methods and parameters of text representation methods

Method	Configuration / Pre-trained Model	Dim.	Key Characteristics
Bag of Words [12]	`CountVectorizer (binary=True)`	$\lvert V\rvert$	Sparse, binary word occurrence.
TF-IDF [5, 14]	`sublinear_tf=True, ngram_range=(1,2), min_df=5`	$\lvert V_{ngrams}\rvert$	Sparse, penalizes highly frequent uninformative terms.
Word2Vec [7]	`nilc-nlp/ word2vec-skip-gram-1000d`	1000	Dense, mean pooling over word tokens, ignores OOV words.
GloVe [11]	`mteb-pt/ average_pt_nilc_glove_s1000`	1000	Dense, mean pooling, captures global co-occurrence statistics.
FastText [2]	`cc.pt.300.bin`	300	Dense, mean pooling, infers OOV representations via subword n-grams.
Sentence-BERT [4]	`paraphrase-multilingual-MiniLM-L12-v2`	384	Dense, dynamic contextual embeddings optimized for semantic similarity.

Method (PR), adapted in this work to recommend textual activities. To achieve this, a graph was built in which nodes represent the activities and edges represent the similarity between them, calculated using the TF-IDF model and cosine similarity. To validate these connections, a minimum similarity threshold was established using Kruskal's algorithm, and the third quartile of its edge weights was adopted as the threshold, ensuring that at least 75% of the graph's structure remains connected. With the network established, node rankings are calculated using the PageRank formula. The PageRank application can be global, covering the entire system (Global PageRank - GPR), or local, calculated only over the X filtered activities (Local PageRank - LPR); and *v)* Baseline Method, in addition to being used as an embedding, Bag of Words was also applied as a baseline method. Since the model records the number of word occurrences between two texts, it represents the simplest way to evaluate similarity and perform recommendations.

3 Experiments and Results

To evaluate the proposed methods, we used two sets of experiments: *i) expected learning sequences* - these experiments evaluate how the methods are able to recommend exactly the expected output from among the K recommendations; *ii) relevant learning sequences* - a set of experiments that evaluates the most relevant learning sequences, being relevant the set of learning sequences with same metadata as the expected output. Table 3 summarizes the descriptions and mathematical definitions of the metrics employed in this work [13].

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_37

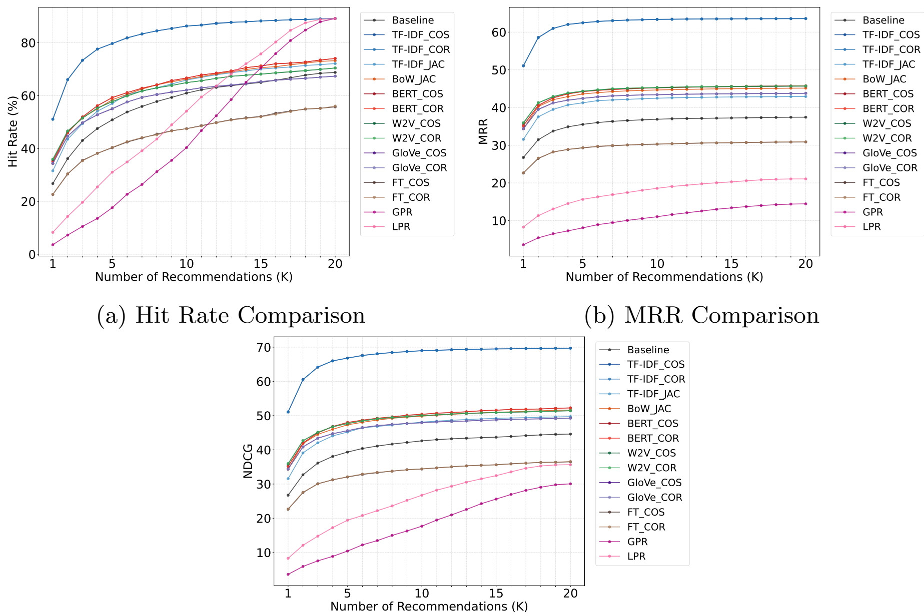
Table 3: Description and mathematical definitions of the evaluation metrics.

Metric	Description	Formula
Precision@K	Measures the proportion of relevant items among the Top-K recommended items.	$\frac{\lvert R_k \cap G \rvert}{k}$
Recall@K	Measures the coverage, revealing how many relevant items from the total set were successfully retrieved.	$\frac{\lvert R_k \cap G \rvert}{\lvert G \rvert}$
F1-Score@K	Balances Precision and Recall into a single value, relating accuracy to total coverage.	$2 \cdot \frac{\text{Prec@}k \cdot \text{Rec@}k}{\text{Prec@}k + \text{Rec@}k}$
Hit Rate@K	Averages the presence of at least one relevant item ("hit") within the Top-K recommendations.	$\frac{1}{N} \sum_{u=1}^{N} I(\lvert R_{k,u} \cap G_u \rvert > 0)$
MRR	Calculates the average of the reciprocal ranks of the first relevant item found.	$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\text{rank}_i}$
NDCG@K	Evaluates ranking quality by assigning higher logarithmic weights to items in better positions.	$\frac{\text{DCG@}K}{\text{IDCG@}K}$

Figure 1a presents the first experimental evaluation of Hit Rate metric for K ranging from 1 to 20. We can observe that there is a significant disparity between the various embedding techniques, similarity measures, and ranking algorithms to recommend learning sequences. The results demonstrate that the TF-IDF method combined with Cosine Similarity (TF-IDF_COS) consistently maintains the highest performance for small recommendation size K. The baseline method, despite its simplicity, presents robust results compared to other methods. Figure 1b presents results of the MRR, which provides quality measure of the recommendation ranking. The TF-IDF_COS maintains a dominant lead over all other methods. This results indicated that TF-IDF_COS can infer the correct learning sequence at the top of the suggested list, which is a requirement for user engagement in recommendation systems. Figure 1c presents the NDCG to measure the ranking quality by penalizing relevant items placed lower in the recommendation list. As previous results, the TF-IDF_COS method presents the highest performance. These results suggest that the TF-IDF_COS method identifies relevant learning sequences and places them in high positions for the user.

Figure 2a and 2b presents a comparison of classification and ranking metrics to evaluate the recommendation of relevant learning sequences, respectively. The classification analysis presented in Figure 2a show that the TF-IDF_COS and TF-IDF_COR achieve the highest Recall, reaching approximately 48%. Although Precision and F1 Score present low values across all methods (a common char-

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_37



(a) Hit Rate Comparison

(b) MRR Comparison

(c) NDCG Comparison

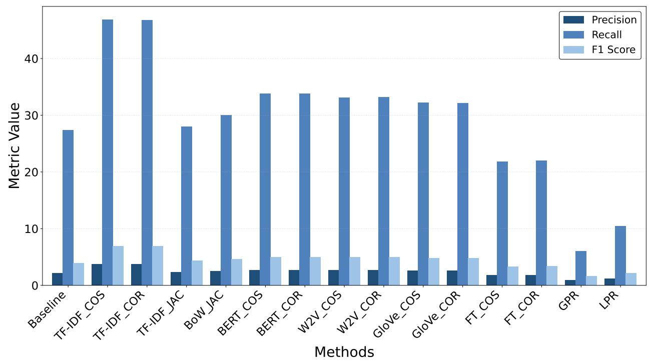
Fig. 1: Comparative analysis of recommendation performance for expected learning sequences using Hit Rate, MRR, and NDCG metrics.

acteristic in large-scale recommendation tasks where the relevant set is small compared to the total set), the TF-IDF methods presents the higher values. In contrast, graph-based methods like GPR and LPR, alongside FastText, exhibit significantly lower classification performance, with GPR and LPR failing to exceed a 10% Recall rate. The ranking metrics, presented in Figure 2b, corroborate the previous results. TF-IDF_COS and TF-IDF_COR outperform all other methods. The high MRR value specifically indicates that these methods are not only successful at finding the relevant learning sequence but also effective at positioning it at the top of the recommendation list. FastText and the graph-based algorithms GPR and LPR show the weakest ranking performance.

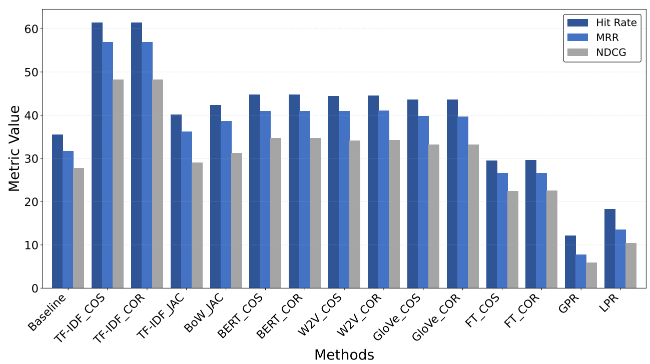
4 Conclusion

The experimental evaluatio revealed that exact term matching prevails in this specific domain. The TF-IDF model combined with Cosine Similarity significantly outperformed deep learning and structural ranking approaches. It demonstrated high immediate accuracy, achieving a Hit Rate of approximately 51% at the very first recommendation ($K = 1$). Maintaining its lead across larger windows, the model achieved the best performance in position-sensitive metrics, with an MRR above 0.63 and an NDCG above 0.68 for $K = 10$, evidencing

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_37



(a) Precision, Recall, and F1-Score



(b) Hit Rate, MRR, and NDCG

Fig. 2: Comparative analysis of recommendation performance for relevant learning sequences, evaluating both classification and ranking metrics.

its capability not only to retrieve the correct material but also to position it at the top of the recommendation list. Notably, the adaptation of the PageRank algorithm obtained a Hit Rate of only 31% for $K = 5$ and demonstrated lower efficacy in prioritizing the most relevant items in the top positions, reaching an MRR below 0.20 and an NDCG below 0.30 for $K = 10$. A current limitation is the handling of out-of-domain queries, as the system always returns the top-k results regardless of actual similarity. For future work, we plan to implement a minimum similarity threshold to detect when no suitable recommendations exist, preventing irrelevant suggestions.

5 Acknowledgments

The authors thank IFSULDEMINAS, FAPEMIG, and A Recreativa for their support. This study was financed by FAPEMIG (Research Foundation of the State of Minas Gerais, process APQ-01864-24) and A Recreativa.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_37

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE transactions on knowledge and data engineering **17**(6) (2005)
2. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Transactions of the association for computational linguistics **5** (2017)
3. Brasil. Ministério da Educação: Base Nacional Comum Curricular. MEC, Brasília (2018)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (2019)
5. Luhn, H.P.: A statistical approach to mechanized encoding and searching of literary information. IBM Journal of research and development **1**(4) (1957)
6. Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., Koper, R.: Recommender systems in technology enhanced learning. In: Recommender systems handbook. Springer, Boston, MA (2012)
7. Mikolov, T., Chen, K., Corrado, G.S., Dean, J.: Efficient estimation of word representations in vector space. In: International Conference on Learning Representations (2013)
8. Musto, C., Semeraro, G., de Gemmis, M., Lops, P.: Learning word embeddings from wikipedia for content-based recommender systems. In: Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20-23, 2016, Proceedings 38. Springer (2016)
9. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford InfoLab (1998)
10. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: The Adaptive Web (2007)
11. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (2014)
12. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Communications of the ACM **18**(11) (1975)
13. Shani, G., Gunawardana, A.: Evaluating recommendation systems. In: Recommender systems handbook. Springer (2010)
14. Spärck Jones, K.: A statistical interpretation of term specificity and its application in retrieval. Journal of documentation **28**(1) (1972)
15. Xie, H., Chu, H.C., Hwang, G.J., Wang, C.C.: Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. Computers & Education **140** (2019)
16. Yunanda, G., Nurjanah, D., Meliana, S.: Recommendation system from microsoft news data using tf-idf and cosine similarity methods. Building of Informatics, Technology and Science (BITS) **4**(1) (2022)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_37