

Curriculum Recommendations Using Transformer Base Model with InfoNCE Loss And Language Switching Method

1st Xiaonan Xu

Osaka, Japan

xiaonan.xu.academic@outlook.com

2nd Bin Yuan

Trine University

Phoenix, USA

binyuan2235@gmail.com

3rd Yongyao Mo

University of Sydney

Victoria, Australia

yomo0791@uni.sydney.edu.au

4th Tianbo Song*

Arizona State University

Phoenix, USA

tianbosong@aol.com

5th Shulin Li

Trine University

Phoenix, USA

liam.cool666@gmail.com

Abstract—The Curriculum Recommendations paradigm is dedicated to fostering learning equality within the ever-evolving realms of educational technology and curriculum development. In acknowledging the inherent obstacles posed by existing methodologies, such as content conflicts and disruptions from language translation, this paradigm aims to confront and overcome these challenges. Notably, it addresses content conflicts and disruptions introduced by language translation, hindrances that can impede the creation of an all-encompassing and personalized learning experience. The paradigm’s objective is to cultivate an educational environment that not only embraces diversity but also customizes learning experiences to suit the distinct needs of each learner. By proactively identifying and addressing these issues, the paradigm strives to pave the way for a more inclusive and responsive educational landscape, ensuring that learning opportunities are equitable and tailored to individual learners’ requirements.

To overcome these challenges, our approach builds upon notable contributions in curriculum development and personalized learning, introducing three key innovations. These include the integration of Transformer Base Model to enhance computational efficiency, the implementation of InfoNCE Loss for accurate content-topic matching, and the adoption of a language switching strategy to alleviate translation-related ambiguities. Together, these innovations aim to collectively tackle inherent challenges and contribute to forging a more equitable and effective learning journey for a diverse range of learners. Competitive cross-validation scores underscore the efficacy of sentence-transformers/LaBSE, achieving 0.66314, showcasing our methodology’s effectiveness in diverse linguistic nuances for content alignment prediction.

Index Terms—Curriculum Recommendation, Transformer model with InfoNCE Loss, Language Switching.

I. INTRODUCTION

In the ever-changing landscape of educational technology and curriculum development, the imperative pursuit of learning equality takes center stage. The commitment to delivering an inclusive and effective learning experience across diverse domains has spurred innovative approaches in curriculum recommendations. This article delves into the paradigm of

Learning Equality - Curriculum Recommendations, recognizing the profound impact of curriculum recommendations on shaping the learning journey. At its core, the paradigm strives to cultivate an educational environment that caters to the unique needs of learners, ensuring equal opportunities for all. The overarching goal is to render the learning experience not only personalized but also equitable, acknowledging and surmounting the challenges posed by existing methodologies. Through the introduction of novel strategies, the paradigm addresses these challenges head-on, aiming to revolutionize the educational landscape by fostering an environment where every individual’s educational journey is both personalized and marked by a commitment to equality.

In the realm of educational technology and curriculum development, a multitude of significant contributions have shaped the learning landscape. Tucker’s K-12 computer science model [1] and Chamunyonga et al. [2] call for enhanced medical curricula underscore impactful contributions in educational technology. Moreover, Bian et al. [3] addressed predictive model robustness, Wang et al. [4] tackled personalized learning nuances, Kumar et al. [5] navigated customized approaches, and Marras et al. [6] confronted bias and fairness challenges. In the latest research conducted in 2023, Sanusi et al. [7] provided a comprehensive overview of machine learning in K-12 education, Hassan et al. [8] used deep learning to enhance computing curriculum, and Atalla et al. [9] introduced an intelligent academic advising system. However, despite these impactful contributions, challenges such as potential biases, fairness issues, and the need for more tailored strategies persist in educational technology and curriculum development.

Despite these commendable efforts, there are inherent issues within existing methodologies. Challenges include handling related content conflicts, addressing noise in training data introduced by language translation, and ensuring effective sampling strategies to mitigate false positives in content-topic associations. Our approach introduces three key innovations to

tackle the identified challenges effectively. First, we employ Transformer Base Model [10] [11] with the InfoNCE Loss [12] as a symmetric contrastive loss-function. This facilitates precise loss calculation by emphasizing correct matches on the diagonal of the similarity matrix, reducing noise in the training process. Secondly, to mitigate issues introduced by language translation, we implement a language switching strategy after each epoch, involving alternating between languages. This strategy minimizes ambiguities during loss calculation and provides a score boost by diversifying the training data distribution.

In conclusion, the Learning Equality - Curriculum Recommendations paradigm builds upon the foundations laid by prior works while innovatively addressing their inherent challenges. By implementing Transformer Base Model with limited sequence length, utilizing InfoNCE Loss, and incorporating a language switching strategy, our approach aims to create a more equitable and effective learning environment for diverse learners. Competitive cross-validation scores underscore the efficacy of sentence-transformers/LaBSE, achieving 0.66314, showcasing our methodology's effectiveness in diverse linguistic nuances for content alignment prediction.

II. RELATED WORK

In the realm of educational technology and curriculum development, several significant contributions have been made to enhance the learning experience across various domains. Tucker [1] proposed a model curriculum for K-12 computer science, providing a foundational framework for computer science education at the primary and secondary levels. Meanwhile, Chamunyonga et al. [2] explored the impact of artificial intelligence and machine learning in radiation therapy, emphasizing the need for curriculum enhancement to address evolving technological landscapes in the medical field.

Due to the rapid advancements in deep learning, an increasing number of studies are being employed in the real of Curriculum Recommendations research. Bian et al. [3] introduced the concept of contrastive curriculum learning for sequential user behavior modeling, employing data augmentation techniques to improve the robustness of predictive models. Wang et al. [4] delved into online learner modeling and course recommendation based on emotional factors, adding a nuanced dimension to personalized learning approaches. Kumar et al. [5] focused on customized curriculum and learning approach recommendation techniques, specifically in the application of virtual reality in medical education, highlighting the importance of tailoring educational strategies to emerging technologies. Marras et al. [6] investigated the equality of learning opportunities through individual fairness in personalized recommendations, addressing the challenges associated with bias and fairness in educational technology.

In the latest research conducted in 2023, the exploration of the Learning Equality - Curriculum Recommendations has continued to evolve. Sanusi et al. [7] conducted a systematic review of teaching and learning machine learning in K-12 education, providing a comprehensive overview of the current

landscape and identifying potential areas for improvement. Hassan et al. [8] leveraged deep learning and big data to enhance computing curriculum for industry-relevant skills, presenting a Norwegian case study that underscores the integration of cutting-edge technologies into educational practices. Finally, Atalla et al. [9] introduced an intelligent recommendation system for automating academic advising based on curriculum analysis and performance modeling, showcasing innovative approaches to support academic decision-making.

III. ALGORITHM AND MODEL

A. Transformer Base Model with InfoNCE Loss

In this section, we will delve into our points of innovation. By implementing a Transformer Base Model with a limited sequence length, utilizing InfoNCE Loss, and incorporating a language switching strategy, our approach aims to create a more equitable and effective learning environment for diverse learners.

To address challenges in topic-content matching, we employ a Transformer Base Model with the InfoNCE Loss, as detailed in Vaswani et al. [10] and Feng et al. [11]. Illustrated in Fig.1, this architecture enables nuanced learning of topic-content relationships. Leveraging Transformer Base Models to encode both topics and content, the design empowers the model to capture intricate patterns and dependencies within the input data.

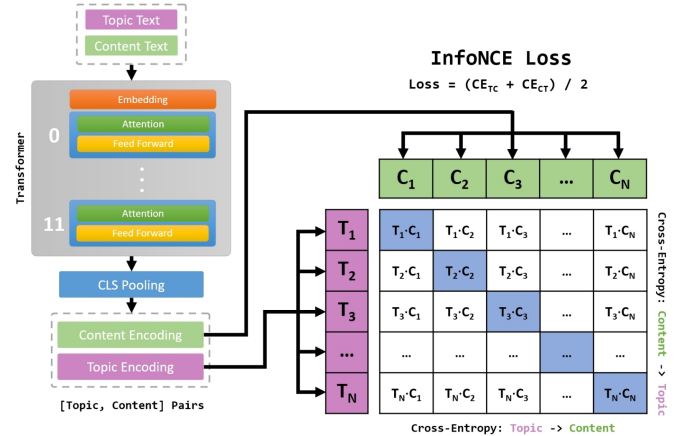


Fig. 1. InfoNCE Loss

By utilizing the InfoNCE Loss [12], our model aims to reduce noise in the training process, particularly when dealing with topics that share similar content. Unlike traditional cross-entropy loss, InfoNCE Loss emphasizes correct matches on the diagonal of the similarity matrix during loss calculation. This innovative loss function contributes to the model's ability to precisely identify and match topics with their corresponding content. The InfoNCE Loss is formulated as follows:

$$-\log \left(\frac{\exp(\text{sim}(x, y^+))}{\exp(\text{sim}(x, y^+)) + \sum_{k=1}^K \exp(\text{sim}(x, y_k^-))} \right) \quad (1)$$

Where:

- $\text{sim}(x, y)$ is a similarity measure between samples x and y .
- y^+ is the positive sample similar to x .
- y_k^- is the k -th negative sample dissimilar to x .
- K is the number of negative samples.

This loss function guarantees that, during the loss calculation, emphasis is placed solely on the diagonal of the similarity matrix. This is vital to prevent misinterpretation of elevated similarities between topics and content within the same batch. By focusing exclusively on the diagonal, the loss function ensures a more accurate assessment of matching within the context of the specific batch, mitigating the risk of erroneous interpretations of similarities between disparate elements in the dataset.

Moreover, we integrate a specialized shuffle function to construct batches that prevent the co-occurrence of topics and related content. This approach is essential for minimizing noise and conflicts in the training process. Additionally, we employ a sampling strategy for missing and incorrect content, ensuring that the model effectively learns from both positive and challenging negative samples. This comprehensive approach enhances the model's robustness by systematically addressing potential biases and difficulties inherent in the training data, promoting a more accurate and well-rounded learning experience.

B. Language Switching Method

we adopted a translation approach covering prevalent languages called language switching method. However, directly incorporating translated data into our training set presents challenges with the InfoNCE loss. For instance, translating an English topic and content item into French, where a corresponding topic and content already exist in the original French data, would introduce noise in loss calculation. To address this, we implement a careful filtering mechanism to exclude translated pairs with similar counterparts in the target language, ensuring that the augmented data contributes meaningfully to the training process without introducing redundancies or misleading information during loss calculation.

As shown in Fig.2, the switching strategy occurs every second epoch and is confined to languages like English (en), Spanish (es), Portuguese (pt), and French (fr). Despite anticipated noise in training, language switching contributes to a score increase of approximately 0.01–0.02. This modest gain implies that the introduced noise might influence training dynamics positively. The periodicity of the switching strategy, coupled with its language limitations, strategically introduces variability during training, and the observed score boost suggests a nuanced impact on the model's ability to adapt and generalize, showcasing the strategy's subtle yet meaningful role in enhancing the overall training process.

C. Datasets

The dataset originates from the Kolibri Studio curricular alignment tool, allowing users to create channels, construct

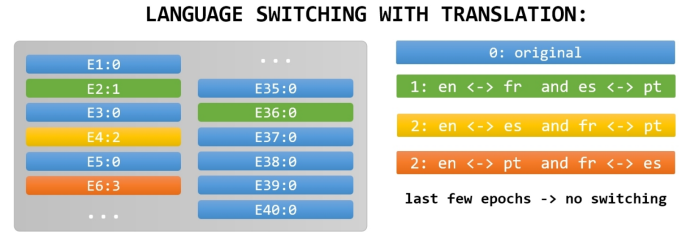


Fig. 2. Language Switching Method

topic trees representing curriculum taxonomies, and organize content items. Users can upload their content or import materials from the Kolibri Content Library. The challenge is to predict the optimal alignment of content items with topics, mimicking the choices made by curricular experts and Kolibri Studio users. The widespread integration of AI is evident [13]–[15]. The goal is to streamline the curator's task of discovering pertinent materials for each topic by recommending content items. The comprehensive test set comprises 10,000 additional topics (absent in the training set) and numerous extra content items exclusively correlated with the test set topics. This expansion aims to assess the model's ability to generalize and recommend relevant content beyond the training data, simulating real-world scenarios where novel topics and content continuously emerge.

D. Evaluation metrics

The F_2 score [16], also known as the F-beta score, serves as a valuable metric for assessing the accuracy of a test. It represents the weighted harmonic mean of precision and recall, with an emphasis on recall. The general formula for the F-beta score is expressed as:

$$F_\beta = \frac{(1 + \beta^2) \cdot \text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}} \quad (2)$$

Here, precision is defined as $\frac{TP}{TP+FP}$, and recall is defined as $\frac{TP}{TP+FN}$, where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives.

For the F_2 score, $\beta = 2$, indicating that recall is considered twice as important as precision. Substituting these values into the F-beta formula, we derive the following formula for the F_2 score:

$$F_2 = \frac{5 \cdot TP^2}{TP^2 + 5 \cdot TP \cdot FP + TP \cdot FN} \quad (3)$$

To obtain the mean F_2 score across all predictions in a dataset, calculate the F_2 score for each row and then compute the average. Additionally, a 10-fold cross-validation (CV) [17] split strategy is employed, aiming to minimize the overlap of content relations between different folds. This involves creating 10 buckets and assigning topics to the bucket where all attached content results in the least overlap with the same attached content in other buckets. While achieving a perfect split aligned with the leader board is challenging due to the

n x m relation of Topic x Content, efforts have been made to minimize the overlap during CV-split creation.

E. Results

In this section, we evaluate the performance of our model using a diverse set of pre-trained models to predict content alignments for given topics. Notable models such as sentence-transformers/LaBSE [11], facebook/mcontriever-msmarco [18], sentence-transformers/stsb-xlm-r-multilingual, and sentence-transformers/paraphrase-multilingual-mpnet-base-v2 [19] were employed. Training involved fold 0 of 10 Folds, with a max sequence length of 96 and a batch size of 768 pairs of [topic, content]. Utilizing polynomial decay with warmup (2 epochs) for 40 epochs and a learning rate of 0.0003. Table I presents the cross-validation scores, showcasing competitive performance across the models.

TABLE I
MODEL RESULTS

Model_Name	CV Score
sentence-transformers/LaBSE	0.66314
facebook/mcontriever-msmarco	0.66148
sentence-transformers/stsb-xlm-r-multilingual	0.65697
sentence-transformers/paraphrase-multilingual-mpnet-base-v2	0.66035

The competitive cross-validation scores highlight the efficacy of our sentence-transformers/LaBSE employing the InfoNCE loss and Language Switching leading with a score of 0.66314. These outcomes underscore the effectiveness of our methodology in capturing diverse linguistic nuances for content alignment prediction. The ensemble approach, incorporating innovative techniques, proves to be a robust solution for the task, showcasing its ability to yield competitive results in comprehensively addressing the challenges associated with content alignment prediction in the context of diverse linguistic nuances.

IV. CONCLUSION

Amidst this swift advancement, our study explores the nuanced domain of the Learning Equality - Curriculum Recommendations paradigm, representing a transformative shift in educational technology committed to ensuring inclusive and effective learning experiences. Our innovative approach, integrating Transformer Base Models, the InfoNCE Loss, and a language switching strategy, directly addresses challenges in the paradigm. The competitive cross-validation scores highlight the efficacy of our sentence-transformers/LaBSE employing the InfoNCE loss and Language Switching leading with a score of 0.66314. These outcomes underscore the effectiveness of our methodology in capturing diverse linguistic nuances for content alignment prediction.

Our study contributes to the ongoing evolution of AI's role in fostering a progressive and inclusive educational landscape. Amalgamating for educational advancement, we ensure tech fosters fairness and efficacy in dynamic learning frontiers. Navigating curriculum complexities, Learning Equality's

paradigm is a beacon, championing personalized and equitable learning experiences in the educational landscape.

REFERENCES

- [1] A. Tucker, *A model curriculum for k-12 computer science: Final report of the acm k-12 task force curriculum committee.* ACM, 2003.
- [2] C. Chamunyonga, C. Edwards, P. Caldwell, P. Rutledge, and J. Burbury, "The impact of artificial intelligence and machine learning in radiation therapy: considerations for future curriculum enhancement," *Journal of Medical Imaging and Radiation Sciences*, vol. 51, no. 2, pp. 214–220, 2020.
- [3] S. Bian, W. X. Zhao, K. Zhou, J. Cai, Y. He, C. Yin, and J.-R. Wen, "Contrastive curriculum learning for sequential user behavior modeling via data augmentation," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 3737–3746.
- [4] Y. Wang *et al.*, "Research on online learner modeling and course recommendation based on emotional factors," *Scientific Programming*, vol. 2022, 2022.
- [5] A. Kumar, A. K. J. Saudagar, M. AlKhathami, B. Alsamani, M. B. Khan, M. H. A. Hasanat, and A. Kumar, "Customized curriculum and learning approach recommendation techniques in application of virtual reality in medical education." *JUCS: Journal of Universal Computer Science*, vol. 28, no. 9, 2022.
- [6] M. Marras, L. Boratto, G. Ramos, and G. Fenu, "Equality of learning opportunity via individual fairness in personalized recommendations," *International Journal of Artificial Intelligence in Education*, vol. 32, no. 3, pp. 636–684, 2022.
- [7] I. T. Sanusi, S. S. Oyelere, H. Vartiainen, J. Suhonen, and M. Tukiainen, "A systematic review of teaching and learning machine learning in k-12 education," *Education and Information Technologies*, vol. 28, no. 5, pp. 5967–5997, 2023.
- [8] M. U. Hassan, S. Alaliyat, R. Sarwar, R. Nawaz, and I. A. Hameed, "Leveraging deep learning and big data to enhance computing curriculum for industry-relevant skills: A norwegian case study," *Heliyon*, vol. 9, no. 4, 2023.
- [9] S. Atalla, M. Daradkeh, A. Gawanmeh, H. Khalil, W. Mansoor, S. Miniaoui, and Y. Himeur, "An intelligent recommendation system for automating academic advising based on curriculum analysis and performance modeling," *Mathematics*, vol. 11, no. 5, p. 1098, 2023.
- [10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, E. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [11] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, "Language-agnostic bert sentence embedding," *arXiv preprint arXiv:2007.01852*, 2020.
- [12] C. Wu, F. Wu, and Y. Huang, "Rethinking infonce: How many negative samples do you need?" *arXiv preprint arXiv:2105.13003*, 2021.
- [13] C. Mou, W. Dai, X. Ye, and J. Wu, "Research on method of user preference analysis based on entity similarity and semantic assessment," in *2023 8th International Conference on Signal and Image Processing (ICSIP)*. IEEE, 2023, pp. 1029–1033.
- [14] W. Dai, Y. Jiang, C. Mou, and C. Zhang, "An integrative paradigm for enhanced stroke prediction: Synergizing xgboost and xdeepfm algorithms," *arXiv preprint arXiv:2310.16430*, 2023.
- [15] C. Mou, X. Ye, J. Wu, and W. Dai, "Automated icd coding based on neural machine translation," in *2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*. IEEE, 2023, pp. 495–500.
- [16] B. Prasetyo, M. Muslim, N. Baroroh *et al.*, "Evaluation performance recall and f2 score of credit card fraud detection unbalanced dataset using smote oversampling technique," in *Journal of physics: conference series*, vol. 1918, no. 4. IOP Publishing, 2021, p. 042002.
- [17] M. W. Browne, "Cross-validation methods," *Journal of mathematical psychology*, vol. 44, no. 1, pp. 108–132, 2000.
- [18] G. Izacard, M. Caron, L. Hosseini, S. Riedel, P. Bojanowski, A. Joulin, and E. Grave, "Towards unsupervised dense information retrieval with contrastive learning," *arXiv preprint arXiv:2112.09118*, vol. 2, no. 3, 2021.
- [19] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019.