Towards Explainable Educational Recommendation through Path Reasoning Methods

Discussion Paper

Neda Afreen, Giacomo Balloccu*, Ludovico Boratto, Gianni Fenu and Mirko Marras

Department of Mathematics and Computer Science, University of Cagliari, V. Ospedale 72, 09124 Cagliari, Italy

Abstract

Current recommender systems in education lack explainability and interpretability, making it challenging for stakeholders to understand how the recommended content relates to them. Path reasoning methods are an emerging class of recommender systems that provides users with the reasoning behind a recommendation. While these methods have been shown to work well in several domains, there is no extensive research on their effectiveness in the context of education. In this ongoing project, we investigate the extent to which the existing path reasoning methods meet utility and beyond utility objectives in educational data. Experiments on two large-scale online course datasets show that this class of methods yields promising results and poses the ground for future advances.

Keywords

Recommender systems, Path reasoning, Recommendation utility, Beyond utility.

1. Introduction

Recommender systems (RSs) are being developed as a response to information overload, which has caused issues for the users while retrieving information that meets their needs. One prominent applicative area that is increasingly adopting this class of systems is education [1]. The effectiveness of RSs in education relies on the ability to provide learners with relevant and valuable educational resources that can enhance their learning experience [2]. As a result, it is crucial to ensure that the recommended content is appropriate, accurate, and reliable [3]. Being able to explain the reason a certain content has been recommended becomes therefore important to increase trust and acceptance of the system [4].

Massive open online courses (MOOCs) have pulled a large number of learners and teachers by providing an abundance of learning resources and recording learners' behavior on the platform [5]. Course recommendation is one among the several intelligent methods the MOOC context is benefiting from, in addition to knowledge tracing and intelligent tutoring, as examples [6]. Prior work has integrated collaborative filtering with deep learning to boost the model's ability [7]. Yet, such recommendations lack explainability, making it challenging for learners to comprehend

IIR2023: 13th Italian Information Retrieval Workshop, June 8th - 9th, 2023, Pisa, Italy

afreenneda1@gmail.com (N. Afreen); gballoccu@acm.org (G. Balloccu); ludovico.boratto@acm.org (L. Boratto); fenu@unica.it (G. Fenu); mirko.marras@acm.org (M. Marras)

0000-0002-4507-3516 (N. Afreen); 0000-0002-6857-7709 (G. Balloccu); 0000000260533015 (L. Boratto); 0000000346682476 (G. Fenu); 0000000319896057 (M. Marras)

how the recommended content relates to the material they have previously studied [8]. Graphs and Reinforcement Learning have been proven to be an effective and novel framework to counter this issue [9]. However, to the best of our knowledge, no study on methods that adopt such framework, namely path reasoning methods, exists in the context of education.

Having education as an application domain and course recommendation as a problem poses challenges with respect to considering other types of consumption items, such as movies or songs. We briefly summarize some of the key challenges in what follows.

- Education is characterized by an inherent *sparsity*, as users interact with much fewer items (the completion of an online course requires weeks/months, while that of a song or a movie requires minutes/hours);
- Online courses are less mainstream types of items w.r.t. consumption items. Hence, building *rich knowledge graphs*, to connect the items with the entities that characterize them, so as to shape effective reasoning paths, is complex and challenging;
- Course recommendation might have constraints in terms of prerequisites that are needed in order to follow a course. From an optimization point of view, these prerequisites *constraint the paths* we can follow in the knowledge graph, as not all the items can be reached as candidate recommendations;
- From an evaluation point of view, the recommendation of a course should not be just that of an effective item. Courses are expected to produce a *gain in the skills of a learner*; clearly, assessing recommendation effectiveness from this side is anything but trivial.

Considering the aforementioned challenges, in this ongoing project, we conduct a reproducibility study using path reasoning methods, proven to work well in other domains, in the context of course recommendation, aiming to move the first steps towards improving transparency and overall effectiveness in the recommendation process. Our choice was based on the fact that, during training, path reasoning methods leverage high-order relationships between courses and learners, modeled through paths that connect already experienced courses to those that are to be recommended. These paths are considered relevant and used to generate explanations for the recommendations. Experiments on two large-scale public course datasets show that these methods promisingly meet recommendation utility and beyond utility objectives.

2. Methods

Data Preparation. We conducted experiments on two state-of-the-art online education platform datasets, namely Xuetang and COCO. Xuetang [10] was collected from one of the largest MOOCs in China, whereas COCO [11] includes data from a popular worldwide online course platform. They both comprise very rich data resources. Xuetang includes 1,302 courses from 23 different categories, 82,535 users, and 458,454 user-course pairs. On the other hand, COCO consists of over 43K online courses, involving over 16K instructors and 2.5 million learners who provided over 4.5 million ratings. To prepare the data sets for analysis, we discarded users who evaluated less than 5 courses and sorted each user's interactions in chronological order. Next, we split the data into training, validation, and test sets, with 60% of the earliest interactions in the training set, 20% in the validation set, and the rest in the test set. The benchmarked models were optimized and tested using the same pre-processed data sets, to ensure a fair comparison.

Model Characterization. Path reasoning is an effective recommendation technique that considers complex connections between learners and courses. It identifies the relationship between recommended courses and previously attended courses by extracting reasoning paths and presents them to the learner in the form of textual explanations. Unlike regularization methods, which simply assign weight to product features without explanation, path reasoning provides meaningful justifications for the recommendations. PGPR [12] is a notable model belonging to this class that uses reinforcement learning to train an agent to find the most relevant paths for the learner. During training, the agent begins at a user and is trained to navigate to the correct courses using paths that result in high rewards. During inference, the agent can directly reach the correct courses, without exploring all possible paths between the learner and the courses, since it has learned an efficient route during training. On the other hand, CAFE [13] adopts a coarse-to-fine paradigm, which involves creating a learner profile initially to identify relevant patterns in the graph. To enable multi-hop path reasoning, the reasoner is broken down into a series of neural reasoning modules. These modules are combined in a way that aligns with the learner profile, allowing for efficient path reasoning that is guided by the learner's preferences. We based our choice on the findings from recent prior work [14], which highlighted that these two methods have been widely adopted and proven to lead to better recommendations, among knowledge-aware (explainable) methods.

Given the absence of rich and validated knowledge graphs for these datasets, the graph structure modeled by the methods in our experiments was just based on the users and courses as entities and the users' action of giving a rating to a course as a relationship. As a result, the baseline textual explanations the models will be able to deliver would be in the form "*course x is recommended to you because another user similar to you attended it*". Such type of explanation, usually referred to as *collaborative filtering-based explanation*, has been proven to be the most commonly returned by models in other domains, despite being the least appreciated by users, even in cases where such models were empowered with rich knowledge graphs [15]. We believe that creating baseline models based on the above assumptions will be important for setting the ground of the overall goal project. Once a (rich) knowledge graph is available and included in the data given to the model, we can observe the impact and power of the knowledge graph (and of the method being able to navigate it) by comparing our baseline results in this paper with those obtained with the models empowered with the richer knowledge graph, in terms of recommendation quality and explanation quality. This is our ambitious long-term goal.

Model Evaluation. The model's ability to recommend courses to learners was evaluated on a dataset by monitoring two widely-adopted evaluation metrics: Normalized Discounted Cumulative Gain (NDCG) and Mean Reciprocal Rank (MRR). Unlike recall and accuracy, NDCG considers the position of relevant courses in the recommended list, while MRR only looks at the position of the first relevant course. Additionally, we assessed the performance of the considered models in terms of relevant beyond-utility objectives [14]. First, we examine the coverage of the recommended courses, which measures the proportion of available courses recommended at least once by the model. Higher coverage can lead to increased learner satisfaction. We also evaluated the serendipity and diversity of the recommendations. Serendipity measures how

Table 1
Recommendation utility and beyond-utility evaluation metrics for path reasoning methods in education.

Method	Xuetang						COCO					
	NDCG	MRR	SER	DIV	NOV	COV	NDCG	MRR	SER	DIV	NOV	COV
PGPR	**0.23**	**0.12**	0.12	0.03	0.76	0.51	0.09	0.04	0.30	0.28	0.86	**0.59**
CAFE	0.18	0.07	**0.28**	0.02	**0.80**	**0.59**	0.09	0.04	**0.67**	0.25	**0.90**	0.58

much they differ from the benchmarked and baseline models; the greater the difference, the higher the level of surprise for the learner. Diversity measures the distinct product categories in the recommended list for better understanding and acceptance. Finally, we assessed recommendation novelty by calculating the inverse of the product's popularity, assuming that less popular courses are more likely to be surprising. Considering different (types of) evaluation metrics can allow us to understand how well the considered methods generalize to educational datasets.

3. Experimental Results

Table 1 collects the utility (NDCG and MRR) and beyond-utility metrics (serendipity, diversity, novelty, and coverage) obtained under our reproducibility protocol. It can be observed that, in terms of recommendation utility, PGPR achieved slightly higher estimates than CAFE for Xuetang (0.18 to 0.23 NDCG; 0.07 to 0.12 MMR); the estimates under the COCO dataset were almost identical between the methods. Concerning beyond-utility objectives, CAFE achieved a significantly higher serendipity than PGPR for both datasets (0.12 to 0.28 SER in Xuetang, 0.30 to 0.67 SER in COCO). Diversity scores for both datasets, consering both the models, are similar. Coverage, instead, depended on paths in the graph. Both methods have similar scores (0.59 in CAFE to 0.51 in PGPR for Xuetang, 0.59 in PGPR to 0.58 in CAFE for COCO). Finally, novelty scores were almost similar between the two classes of methods. Interestingly, on COCO, even though both methods resulted in similar recommendation utility, CAFE yielded significantly higher estimates for all the beyond-utility metrics than PGPR. This observation confirms that solely observing and selecting a method based on utility might lead to sub-optimal decisions.

4. Conclusions and Future Work

At the current stage, in this ongoing project, we investigated the extent to which the existing path reasoning methods meet utility and beyond utility objectives in educational data, with course recommendation as the case study. Our results show that path reasoning methods have promising performance in terms of utility and beyond-utility objectives, while being able to provide textual explanations. As outlined along the paper, in the next steps, we plan to devise rich knowledge graphs for leveraging the full power of path reasoning methods to meet both effectiveness and transparency. Integrating a rich knowledge graph, we also plan to introduce more explanation textual templates to make learners (but also teachers) understand the specific reasoning process; this has the potential to enhance learners' satisfaction and possibly better inform their decisions about learning. In addition to this, we will extend our evaluation protocol to monitor metrics specifically pertaining to the transparency dimension, such as the explanation quality and, finally, perform user studies with learners and teachers.

References

[1] P. A. O. Cano, E. C. P. Alarcón, Recommendation systems in education: A review of recommendation mechanisms in e-learning environments, Revista Ingenierias Universidad de Medellin 20 (2021) 147–158.

[2] M.-I. Dascalu, C.-N. Bodea, M. N. Mihailescu, E. A. Tanase, P. Ordoñez de Pablos, Educational recommender systems and their application in lifelong learning, Behaviour & information technology 35 (2016) 290–297.

[3] X. Chen, C. Huang, L. Yao, X. Wang, W. Zhang, et al., Knowledge-guided deep reinforcement learning for interactive recommendation, in: 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, 2020, pp. 1–8.

[4] N. Sonboli, J. J. Smith, F. Cabral Berenfus, R. Burke, C. Fiesler, Fairness and transparency in recommendation: The users' perspective, in: Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization, 2021, pp. 274–279.

[5] F. Bousbahi, H. Chorfi, Mooc-rec: a case based recommender system for moocs, Procedia-Social and Behavioral Sciences 195 (2015) 1813–1822.

[6] A. Khalid, K. Lundqvist, A. Yates, Recommender systems for moocs: A systematic literature survey (january 1, 2012–july 12, 2019), International Review of Research in Open and Distributed Learning 21 (2020) 255–291.

[7] G. Piao, Recommending knowledge concepts on mooc platforms with meta-path-based representation learning., International Educational Data Mining Society (2021).

[8] M. C. Urdaneta-Ponte, A. Mendez-Zorrilla, I. Oleagordia-Ruiz, Recommendation systems for education: systematic review, Electronics 10 (2021) 1611.

[9] G. Balloccu, L. Boratto, G. Fenu, M. Marras, Post processing recommender systems with knowledge graphs for recency, popularity, and diversity of explanations, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 646–656.

[10] J. Zhang, B. Hao, B. Chen, C. Li, H. Chen, J. Sun, Hierarchical reinforcement learning for course recommendation in moocs, in: Proceedings of the AAAI conference on artificial intelligence, volume 33, 2019, pp. 435–442.

[11] D. Dessì, G. Fenu, M. Marras, D. Reforgiato Recupero, Coco: Semantic-enriched collection of online courses at scale with experimental use cases, in: Trends and Advances in Information Systems and Technologies: Volume 2 6, Springer, 2018, pp. 1386–1396.

[12] Y. Xian, Z. Fu, S. Muthukrishnan, G. De Melo, Y. Zhang, Reinforcement knowledge graph reasoning for explainable recommendation, in: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, 2019, pp. 285–294.

[13] Y. Xian, Z. Fu, H. Zhao, Y. Ge, X. Chen, Q. Huang, S. Geng, Z. Qin, G. De Melo, S. Muthukrishnan, et al., Cafe: Coarse-to-fine neural symbolic reasoning for explainable recommendation, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 1645–1654.

[14] G. Balloccu, L. Boratto, C. Cancedda, G. Fenu, M. Marras, Knowledge is power, understanding is impact: Utility and beyond goals, explanation quality, and fairness in path reasoning recommendation, arXiv preprint arXiv:2301.05944 (2023).

[15] G. Balloccu, L. Boratto, G. Fenu, M. Marras, Reinforcement recommendation reasoning through knowledge graphs for explanation path quality, Knowledge-Based Systems 260 (2023) 110098.