Predicting Prerequisite Relations for Unseen Concepts

Yaxin Zhu and **Hamed Zamani**
Center for Intelligent Information Retrieval
University of Massachusetts Amherst
{yaxinzhu, zamani}@cs.umass.edu

Abstract

Concept prerequisite learning (CPL) plays a key role in developing technologies that assist people to learn a new complex topic or concept. Previous work commonly assumes that all concepts are given at training time and solely focuses on predicting the unseen prerequisite relationships between them. However, many real-world scenarios deal with concepts that are left undiscovered at training time, which is relatively unexplored. This paper studies this problem and proposes a novel *alternating knowledge distillation* approach to take advantage of both content- and graph-based models for this task. Extensive experiments on three public benchmarks demonstrate up to 10% improvements in terms of F1 score.

1 Introduction

As the amount of online educational data rapidly grows, it is more important than ever to develop information access systems that assist users to learn new complex topics or concepts (Gwizdka et al., 2016; Collins-Thompson et al., 2017; Eickhoff et al., 2017; Urgo and Arguello, 2022). A fundamental step towards developing these systems is *concept prerequisite learning (CPL)*—the task of building a concept graph by structuring open knowledge in prerequisite relations. A prerequisite is a directed relation between two concepts, e.g., *Heap Tree* is a prerequisite of *Heap Sort*.

CPL was first introduced by Talukdar and Cohen (2012) with the aim of formulating probabilistic planning problems for machine learning solvers. A number of CPL models use various kinds external resources including Wikipedia links (Liang et al., 2015), textbook structures (Wang et al., 2016), and course dependencies (Liang et al., 2017; Liu et al., 2016) in two ways: content-based (Pan et al., 2017; Roy et al., 2019; Gasparetti, 2022) and graph-based (Liang et al., 2015). There also exists research on active learning (Liang et al., 2018a,b), unsupervised learning (Li et al., 2020), and domain adaptation (Li et al., 2021) to meet data insufficiency challenges in CPL problems. However, most existing approaches assume the system to reconstruct concept prerequisite paths with vague knowledge (i.e. incomplete relations) of each concept, or to transfer graph structure information to a new domain. In practical scenarios, knowledge is updated with both new concepts and relations introduced. To jump out of the offline setting of graph completion with given concepts, we define a new task - CPL for unseen concepts, i.e. predicting prerequisite relationships for concepts that never appear in the training set.

Most existing graph-based CPL approaches cannot be simply used for unseen concepts, because the randomly initialized concept representations do not get updated for unseen concepts. A simple solution would be initializing the concept embeddings based on content-based models (Li et al., 2019; Jia et al., 2021; Zhang et al., 2022; Sun et al., 2022). To better take advantage of content information, we propose a novel CPL model that consists of two components: one that solely focuses on textual content associated with each concept and another one that focuses on the concept graph structure. To train our model, we propose an iterative knowledge distillation approach by alternating between these two components as "teacher" and "student". Our main contributions include:

1. Exploring a new task to predict prerequisite relationships for unseen concepts.
2. Introducing a simple yet effective retrieval-augmented content-based approach for CPL.
3. Proposing an alternating knowledge distillation procedure that benefits both content- and graph-based models for CPL.
4. Advancing state-of-the-art on three public benchmarks. Extensive experiments shed light on the empirical contributions of each proposed

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 8542–8548

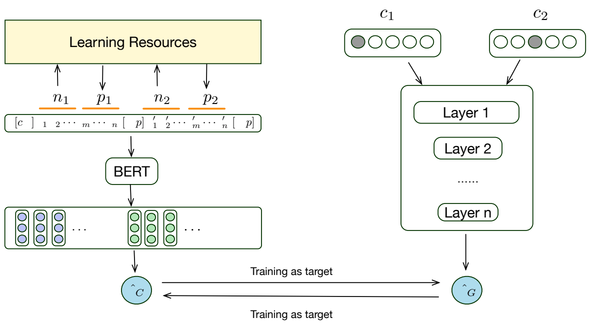

Figure 1: An illustration of the proposed model. c_i and c_j denote concepts, n_i and n_j are their names, & p_i and p_j are passages retrieved from a text corpus (i.e., Wikipedia).

component. Our code and model parameters are released for reproducibility purposes.[1]

2 Methodology

This section introduces a CPL model that uses two complementary components. The first component models prerequisite relations conditioned on textual content retrieved for each concept (i.e., a retrieval-augmented model), and the second component casts the problem as a link prediction task and models prerequisite relations conditioned on the graph structure. The proposed method uses a knowledge distillation approach that alters between these two components as the teacher and student models, iteratively. An overview of the proposed solution is presented in Figure 1.

Notation and Problem Statement: Let $G = (V, E)$ be a directed concept graph whose vertices are associated with concepts and edges represent prerequisite relations, V_h be a set of unseen concepts to dig out. The training set for concept prerequisite learning is equal to $D_{train} = \{(c_i, c_j, r_{ij}) : c_i, c_j \in V, r_{ij}\}$, where $r_{ij} = 1$ if $c_i \rightarrow c_j \in E$ and $r_{ij} = 0$ otherwise. The test set is $D_{test} = \{(c_i, c_j, r_{ij}) : c_i \in V_h \vee c_j \in V_h, r_{ij}\}$.

2.1 Alternating Knowledge Distillation for Concept Prerequisite Learning

There are two formulations of the concept prerequisite learning problem, as follows:

A content-based formulation for CPL: In this formulation, the aim is to develop a model that predicts prerequisite relations based on textual information associated with the concepts, as follows:

$$\arg\min_{\theta_C} \sum_{(c_i,c_j,r_{ij})\in D} L_C(p(c_i \rightarrow c_j|\phi, \theta_C), r_{ij}),$$

where L_C and θ_C denote the loss function and the content-based model parameters, respectively. $\phi(\cdot)$ is a function that takes a concept and provide a textual description of the concept.

A graph-based formulation for CPL: An alternative formulation of the CPL problem is to predict prerequisite relationships based on the prerequisite graph structure, as follows:

$$\arg\min_{\theta_G} \sum_{(c_i,c_j,r_{ij})\in D_{train}} L_G(p(c_i \rightarrow c_j|G, \theta_G), r_{ij}),$$

where L_G and θ_G denote the loss function and the graph-based model parameters, respectively.

Alternating knowledge distillation for training: θ_C and θ_G can be trained independently or jointly on the training set D_{train}. We introduce a more effective alternative optimization called alternating knowledge distillation (AKD), in which the roles of teacher and student models alternate between θ_C and θ_G repeatedly. Our motivation is to improve generalization in both of these models that are complementary. Therefore, we first train our content-based CPL model θ_C (see Section 2.2) using the ground-truth training data (i.e., D_{train}). Then, we consider θ_C as the teacher model and produce a pseudo-labeled training set $\hat{D}$ based on $p(c_i \rightarrow c_j|\phi, \theta_C)$ as follows: For every concept c_i, we consider k positive pseudo labels using $\text{topk}(\{p(c_i \rightarrow c_j|\phi, \theta_C) : \forall c_j \neq c_i\})$. Note that we exclude the concepts with less than 0.5 prerequisite probability, if any. We also take k' negative instances. These negative instances can be selected randomly or from the ones with the lowest probability. We then train the student model θ_G (see Section 2.3) on $D_{train} \cup \hat{D}$, use θ_G as the teacher, produce the pseudo-labeled training set and train the student model θ_C. We repeat this teacher-student alternation process for N steps. Empirically N is set to 4.

Since θ_C and θ_G are complementary , we interpolate their scores linearly to acquire the final probability:

$$\alpha p(c_i \rightarrow c_j|\phi, \theta_C) + (1 - \alpha)p(c_i \rightarrow c_j|G, \theta_G),$$

where $\alpha \in [0, 1]$ is a hyper-parameter. In the following two subsections, we describe how we model θ_C and θ_G, respectively.

2.2 Retrieval-Augmented Concept Prerequisite Learning

A simple approach for modeling $p(c_i \rightarrow c_j|\phi, \theta_C)$ is to use pre-trained language models (e.g.,

[1] https://github.com/yaxinzhuars/unseen_cpl/

	University Course			LectureBank			MOOC ML		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
VGAE (Li et al., 2019)	0.524	0.523	0.520	0.528	0.522	0.515	0.485	0.490	0.481
CPRL−−	0.554	0.566	0.540	0.543	0.545	0.543	0.498	0.496	0.479
BERT	0.771	0.771	0.767	0.729	0.738	0.732	0.746	0.753	0.761
NCF with BERT emds	0.668	0.676	0.663	0.519	0.519	0.517	0.640	0.666	0.649
Ours	**0.844**	**0.841**	**0.842**	**0.754**	**0.761**	**0.757**	**0.846**	**0.806**	**0.823**

Table 1: CPL prediction results obtained by the proposed model and the baselines for unseen concepts. The highest number in each column is bold-faced.

Dataset	#concepts	#p+	#p-
University Course	407	1005	996
LectureBank	205	904	967
Mooc ML	244	1737	4975

Table 2: Data statistics. p+ and p- represent positive pairs and sampled negative pairs, respectively. Unused concepts are excluded.

BERT (Devlin et al., 2018)) to represent the concept names. In our initial experiments, we observed that concept names are not sufficient for prerequisite prediction and more descriptive content should be produced by ϕ. Therefore, with the aim of taking advantage of a massive unstructured corpus from textual world knowledge, we augment the training data with passages retrieved from Wikipedia. To be concise, each concept name is regarded as a query in order to retrieve 100-token passages (similar to the Wikipedia DPR collection (Karpukhin et al., 2020)) using BM25. We use the first ranked passage for augmentation and compute $p(c_i \rightarrow c_j|\phi, \theta_C)$ by feeding "[CLS] n_i p_i [SEP] n_j p_j [SEP]" to BERT and using a fully-connected layer and sigmoid on top of the [CLS] representation. Note that, n_i and p_i are the concept name and the first retrieved passage for concept c_i, respectively. For the loss function, we use binary cross entropy.

2.3 Graph-based Concept Prerequisite Learning

For modeling θ_G, we aim at predicting missing links in a concept graph. Various approaches based on matrix factorization, geometry, and graph neural networks have been developed for the link prediction problem. In this work, we use neural collaborative filtering (NCF) (He et al., 2017) to obtain node representations and corresponding link existence likelihood. NCF is efficient, less prone to overfit, and can be used for directed graphs. It has demonstrated successful results in a number of recommendation problems. We use NCF to learn a representation for every concept. The representations are initialized randomly and we train the model using a binary cross entropy loss function.

3 Experiments

3.1 Data

We evaluate the effectiveness of our approach on the following three manually annotated benchmarks.

University Course (Liang et al., 2017): This dataset includes concepts from computer science course descriptions provided by 11 universities in the United States. The concepts were extracted using Wikipedia Miner (Milne and Witten, 2013).

LectureBank (Li et al., 2019): This dataset was constructed by collecting online lecture files from 60 courses covering NLP and related topics.

MOOC ML (Pan et al., 2017): This dataset contains concept prerequisite relations extracted from video subtitles of Coursera's machine learning courses using the approach presented by Parameswaran et al. (2010).

The statistics of these three benchmarks are presented in Table 2.

3.2 Experimental Setup

To evaluate the ability of predicting prerequisites of undiscovered concepts for our method, we randomly split the concept set of each dataset with a proportion of 9:1 as V and V_h, then reconstruct the training and test set by dropping any pair with an unseen concept into test set. Thus, the representation of implicit concepts will never be updated during training. The experiments are repeated for three times with different random splits and the average is reported. We use Precision, Recall, and F1 Score (macro averaged) as evaluation metrics.

We set the NCF's concept embedding dimensionality to 32, and the learning rates for θ_C and θ_G to

	University Course			LectureBank			MOOC ML		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Ours	**0.844**	**0.841**	**0.842**	**0.754**	**0.761**	**0.757**	**0.846**	0.806	**0.823**
Ours w/o AKD	0.797	0.798	0.796	0.753	0.753	0.753	0.815	0.790	0.801
Ours w/o Retr-Aug BERT	0.739	0.738	0.738	0.551	0.640	0.511	0.635	0.720	0.653
Ours w/o NCF	0.809	0.811	0.808	0.745	0.755	0.748	0.821	**0.825**	0.822

Table 3: The ablation study results. w/o AKD means interpolation only. w/o model means this model is only used for distillation but not interpolation. The highest number in each column is bold-faced.

	University Course			LectureBank			MOOC ML		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
NCF D_{train}	0.674	0.730	0.652	0.519	0.648	0.434	0.564	0.597	0.568
NCF $D_{train}, \hat{D}$	0.734	0.733	0.734	0.551	0.640	0.511	0.635	0.720	0.653
NCF $\hat{D}$	0.712	0.740	0.705	0.541	0.610	0.498	0.578	0.593	0.582

Table 4: NCF results with ground truth data D_{train}, first iteration distilled data $\hat{D}$, and both.

5e-5 and 1e-3, respectively. We use BERT-base in all experiments. For the AKD process, k is proportioned to the size of positive training instances for each concept (i.e., 10%). We also set $k' = k$. The hyper-parameter α was selected using grid search.

3.3 Baselines

We use the following competitive baselines: **VGAE** (Li et al., 2019) uses variational graph autoencoder to encode edges in the training set with a 2-layer Graph Convolutional Network (GCN), then adopts inner product to reconstruct the graph. **CPRL** (Jia et al., 2021) is the most recent CPL approach that produces state-of-the-art results by creating a heterogeneous graph for representing concepts in addition to learning objects. It uses a Relational Graph Convolutional Network (R-GCN) to encode nodes and a Siamese network to identify prerequisites. Note that CPRL uses an external resource. To provide a fairer comparison, we implemented a simplified version of CPRL that excludes the learning objects. We call this method **CPRL−−**. **BERT** (Devlin et al., 2018) takes a pair of concept names and is fine-tuned to classify prerequisite relations using the binary cross entropy loss. **NCF with BERT embs** is implemented to compare different combination methods of content and graph based models. We follow the strategies in (Li et al., 2019; Jia et al., 2021) that initialize NCF node representations with fine-tuned BERT embeddings.

3.4 Results

Comparison with the Baselines: According to Table 1, graph-based baselines perform poorly when dealing with unseen concepts. Unsurprisingly, carrying information from the pre-training step helps the BERT model produce the best results on both University Course and MOOC ML datasets. Implicitly using BERT representations is helpful, but the prediction ability of graph-based model is limited. Our method outperforms the baselines on all three datasets in terms of all metrics. The improvements come from the augmentation using passages retrieved from Wikipedia, the alternating knowledge distillation approach, and the explicit combination of complementary models.

Ablation Study: In our ablation study, we answer the following empirical research questions:

Q1: *Does retrieval augmentation improve the generalizability of the content-based model?* The F1 scores for the proposed retrieval-augmented BERT on University Course, LectureBank and MOOC ML are 0.796, 0.747, 0.800 respectively. Comparing them to the BERT's performance reported in Table 1 demonstrates the generalizability of retrieval augmentation for this task.

Q2: *How is the effect of different components in our AKD approach?* In Table 3 we eliminate each of components and demonstrate substantial drop in nearly all cases. The large performance drops by removing the retrieval-augmented BERT model is due to its role of capturing content information. This experiment demonstrates that all the components used in developing our approach contributes to the final performance.

Q3: *How does distilled data contribute to NCF training?* In Table 4, $\hat{D}$ acts as a better training set than D_{train}, indicating that even weakly anno-

ID	Source Concept	Target Concept	$\hat{p}_C$	$\hat{p}_G$	Ground-truth
1	approximation	numerical analysis	0.943	0.472	1
2	debugging	parallel computing	0.004	0.751	0
3	image processing	3D computer graphics	0.001	0.810	1
4	finite state machine	regular language	0.311	0.622	1

Table 5: Some example prerequisite pairs from the University Course dataset.

	University Course			**LectureBank**			**MOOC ML**		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score	Precision	Recall	F1 Score
VGAE (Li et al., 2019)	0.570	0.562	0.539	0.706	0.673	0.669	0.686	0.601	0.567
CPRL (Jia et al., 2021)	0.689	0.760	0.723	0.861	0.858	0.860	0.800	0.642	0.712
CPRL−−	0.660	0.692	0.646	0.717	0.716	0.711	0.496	0.496	0.495
BERT	0.838	0.835	0.835	0.838	0.837	0.838	0.895	0.878	0.886
NCF with BERT emds	0.722	0.721	0.716	0.545	0.547	0.533	0.631	0.626	0.629
Ours	**0.868**	**0.868**	**0.868**	**0.887**	**0.887**	**0.889**	**0.892**	**0.882**	**0.886**

Table 6: CPL prediction results obtained by the proposed model and the baselines for **seen concepts**. The highest number in each column is bold-faced.

tated unseen concept pairs can play an important role in guiding graph based models. A combination of ground truth D_{train} and $\hat{D}$ leads to strong improvement.

Learning Curve: We plot the learning curves of our model for all three datasets in Figure 2. The model's effectiveness is substantially improved by increasing the training data size. Given the slope of the learning curves, the proposed model is likely to achieve significantly higher F1 scores by increasing the training data size. This is an encouraging observation, especially given that the developed model is already very effective and obtains F1 scores of higher than 0.86 on all datasets.

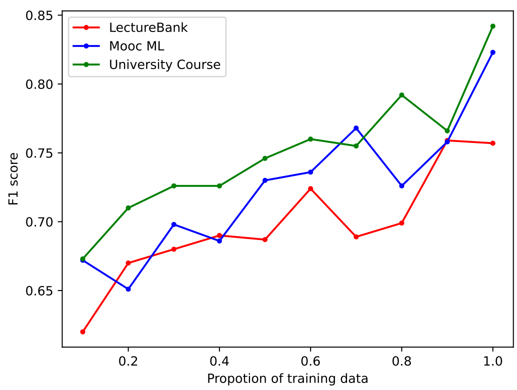

Figure 2: The learning curve for the proposed solution. The x-axis represents the proportion of the training data used and y-axis is the obtained F1 score on the test set.

Case Study: Table 5 demonstrates a few example concept pairs from the LectureBank dataset. The first two examples show that the graph-based model could not accurately predict the prerequisite relations, while the content-based model produces probabilities higher than 0.5. On the other hand, the last example shows that the graph-based model can sometimes correct the errors made by the content-based model. Of course, the interpolation of these two models can sometimes introduce noise (e.g., Example 3). In our ablation study (Table 3), we show that this interpolation leads to overall improvement.

Results for Seen Concepts: Even though this paper focuses on unseen concepts, we also compare our methods against the baselines for predicting prerequisite relations for seen concepts. Following previous work (Li et al., 2019; Jia et al., 2021), we split the concept pairs in each dataset into training and test sets with a proportion of 9:1 for LectureBank and 7:3 for others. Results in Table 6 show that our method outperforms baselines on all three datasets in terms of all metrics, indicating that our approach is equipped with the ability to deal with seen concepts.

4 Conclusions and Future Work

This paper explored the challenge of predicting prerequisites for unseen concepts in CPL. It proposed an alternating knowledge distillation approach that enables us to train more effective content-based and graph-based models, as well demonstrated that content-based CPL models can benefit from retrieval augmentation. In the future, we intent to extend the proposed solution to an online setting, where concept prerequisites can be extracted for every learning-oriented query in a search engine.

Limitations

One of the limitations of this work is that this work overlook the concept detection or extraction in the model design. Even though previous work also made similar assumptions (Liang et al., 2017; Li et al., 2019; Pan et al., 2017), we believe that this is an important aspect that should be considered in the future. Extracting concepts from unstructured data accurately can be challenging. Another limitation is related to the number of concepts in the datasets. In some real-world scenarios, the number of concepts would be significantly higher than those represented by the existing benchmarks. Increasing the number of concepts is likely to negatively impact the model's effectiveness or raise efficiency concerns.

Mistakes made by the CPL models, including ours, if they are used in learning-oriented search engines, are likely to negatively impact the learning outcome. For instance, missing a prerequisite relation during a learning session may lead to some misunderstanding about the concepts being learned. Therefore, we suggest raising awareness of such mistakes to the users so they can make wise decisions while learning online.

Acknowledgments

This work was supported in part by the Center for Intelligent Information Retrieval and in part by NSF grant #2106282. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the sponsor.

References

Kevyn Collins-Thompson, Preben Hansen, and Claudia Hauff. 2017. Search as learning (dagstuhl seminar 17092). In *Dagstuhl reports*, volume 7. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Carsten Eickhoff, Jacek Gwizdka, Claudia Hauff, and Jiyin He. 2017. Introduction to the special issue on search as learning. *Information Retrieval Journal*, 20(5):399–402.

Fabio Gasparetti. 2022. Discovering prerequisite relations from educational documents through word embeddings. *Future Generation Computer Systems*, 127:31–41.

Jacek Gwizdka, Preben Hansen, Claudia Hauff, Jiyin He, and Noriko Kando. 2016. Search as learning (sal) workshop 2016. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 1249–1250.

Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182.

Chenghao Jia, Yongliang Shen, Yechun Tang, Lu Sun, and Weiming Lu. 2021. Heterogeneous graph neural networks for concept prerequisite relation learning in educational data. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2036–2047.

Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense passage retrieval for open-domain question answering. *arXiv preprint arXiv:2004.04906*.

Irene Li, Alexander Fabbri, Swapnil Hingmire, and Dragomir Radev. 2020. R-vgae: Relational-variational graph autoencoder for unsupervised prerequisite chain learning. *arXiv preprint arXiv:2004.10610*.

Irene Li, Alexander R Fabbri, Robert R Tung, and Dragomir R Radev. 2019. What should i learn first: Introducing lecturebank for nlp education and prerequisite chain learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6674–6681.

Irene Li, Vanessa Yan, Tianxiao Li, Rihao Qu, and Dragomir Radev. 2021. Unsupervised cross-domain prerequisite chain learning using variational graph autoencoders. *arXiv preprint arXiv:2105.03505*.

Chen Liang, Zhaohui Wu, Wenyi Huang, and C Lee Giles. 2015. Measuring prerequisite relations among concepts. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 1668–1674.

Chen Liang, Jianbo Ye, Shuting Wang, Bart Pursel, and C Lee Giles. 2018a. Investigating active learning for concept prerequisite learning. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Chen Liang, Jianbo Ye, Zhaohui Wu, Bart Pursel, and C Lee Giles. 2017. Recovering concept prerequisite relations from university course dependencies. In *Thirty-First AAAI Conference on Artificial Intelligence*.

Chen Liang, Jianbo Ye, Han Zhao, Bart Pursel, and C Lee Giles. 2018b. Active learning of strict partial orders: A case study on concept prerequisite relations. *arXiv preprint arXiv:1801.06481*.

Hanxiao Liu, Wanli Ma, Yiming Yang, and Jaime Carbonell. 2016. Learning concept graphs from online educational data. *Journal of Artificial Intelligence Research*, 55:1059–1090.

David Milne and Ian H Witten. 2013. An open-source toolkit for mining wikipedia. *Artificial Intelligence*, 194:222–239.

Liangming Pan, Chengjiang Li, Juanzi Li, and Jie Tang. 2017. Prerequisite relation learning for concepts in moocs. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1447–1456.

Aditya Parameswaran, Hector Garcia-Molina, and Anand Rajaraman. 2010. Towards the web of concepts: Extracting concepts from large datasets. *Proceedings of the VLDB Endowment*, 3(1-2):566–577.

Sudeshna Roy, Meghana Madhyastha, Sheril Lawrence, and Vaibhav Rajan. 2019. Inferring concept prerequisite relations from online educational resources. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 9589–9594.

Hao Sun, Yuntao Li, and Yan Zhang. 2022. Conlearn: Contextual-knowledge-aware concept prerequisite relation learning with graph neural network. In *Proceedings of the 2022 SIAM International Conference on Data Mining (SDM)*, pages 118–126. SIAM.

Partha Talukdar and William Cohen. 2012. Crowdsourced comprehension: predicting prerequisite structure in wikipedia. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 307–315.

Kelsey Urgo and Jaime Arguello. 2022. Learning assessments in search-as-learning: A survey of prior work and opportunities for future research. *Information Processing & Management*, 59(2):102821.

Shuting Wang, Alexander Ororbia, Zhaohui Wu, Kyle Williams, Chen Liang, Bart Pursel, and C Lee Giles. 2016. Using prerequisites to extract concept maps fromtextbooks. In *Proceedings of the 25th acm international on conference on information and knowledge management*, pages 317–326.

Juntao Zhang, Nanzhou Lin, Xuelong Zhang, Wei Song, Xiandi Yang, and Zhiyong Peng. 2022. Learning concept prerequisite relations from educational data via multi-head attention variational graph auto-encoders. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 1377–1385.