

ExpanRL: Hierarchical Reinforcement Learning for Course Concept Expansion in MOOCs

Jifan Yu¹, Chenyu Wang¹, Gan Luo¹,
Lei Hou^{1,2,3*}, Juanzi Li^{1,2,3}, Jie Tang^{1,2,3}, Minlie Huang^{1,2,3}, Zhiyuan Liu^{1,2,3}

¹Dept. of Computer SCi.& Tech., Tsinghua University, China 100084

²KIRC, Institute for Artificial Intelligence, Tsinghua University, China 100084

³Beijing National Research Center for Information Science and Technology, China 100084

{yujf18, luog18}@mails.tsinghua.edu.cn

{houlei, lijuanzi, jietang, aihuang, liuzy}@tsinghua.edu.cn

Abstract

Within the prosperity of Massive Open Online Courses (MOOCs), the education applications that automatically provide extracurricular knowledge for MOOC users have become rising research topics. However, MOOC courses’ diversity and rapid updates make it more challenging to find suitable new knowledge for students. In this paper, we present ExpanRL, an end-to-end hierarchical reinforcement learning (HRL) model for concept expansion in MOOCs. Employing a two-level HRL mechanism of seed selection and concept expansion, ExpanRL is more feasible to adjust the expansion strategy to find new concepts based on the students’ feedback on expansion results. Our experiments on nine novel datasets from real MOOCs show that ExpanRL achieves significant improvements over existing methods and maintain competitive performance under different settings.

1 Introduction

The cognitive-driven theory has been widely used in practical teaching since Ausubel firstly proposed it in (Ausubel, 1968), which suggests educators provide new knowledge for students to motivate their learning continuously. In fact, in addition to the concepts taught in course, many related concepts are also attractive and worthy of learning. As shown in Figure 1, when a student studies the concept *LSTM* in “Deep Learning” course from Coursera¹, many related concepts, including its prerequisite concepts (*RNN*), related scientists (*Jürgen Schmidhuber*) and its related applications (*Machine Translation*) can also benefit his/her further study. In traditional classrooms, these concepts are often considerably introduced by teachers.

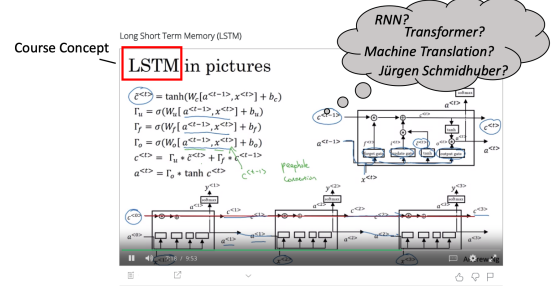


Figure 1: An example of course-related concepts in the “Deep Learning” course from Coursera.

However, in the era of Massive Open Online Courses (MOOCs), thousands of courses are pre-recorded for with millions of students with various backgrounds (Shah, 2019), which makes it infeasible to pick out these essential concepts manually. Therefore, there is a clear need to automatically discover course-related concepts so that they can easily acquire additional knowledge and achieve better educational outcomes.

This task is formally defined as **Course Concept Expansion** (Yu et al., 2019a), a special type of *Concept Expansion* or *Set Expansion* (Wang and Cohen, 2007), which refers to the task of expanding a small set of seed concepts into a complete set of concepts that belong to the same course or subject from external resources. Despite abundant efforts in related topics (He and Xin, 2011; Shen et al., 2017; Yan et al., 2019), existing methods still face three challenges when applied to MOOCs.

First, distinct from the task of enriching a certain concept set, the purpose of course concept expansion is to benefit students’ learning, making the context information insufficient to detect whether a concept is appropriate to be an expansion result. How to properly introduce student feedback in the model’s loop is a crucial challenge.

Second, unlike the set expansion for a clear general category (e.g., countries), courses are often the

*Corresponding author.

¹<https://www.coursera.org>

combinations of multiple categories, especially in interdisciplinary courses like *Mathematics for Computer Science*². Therefore, it isn't easy to model the course's semantic scope (Curran et al., 2007) when applying existing expansion methods.

Third, MOOCs are updated continuously, and numerous new courses arise everyday (Shah, 2019), which requires a good generalization ability of the expansion model; otherwise, the frequent model retraining will cause severe waste of resources.

To address the above problems, we construct a novel interactive environment on real MOOCs, which collects students' feedback on expansion results and provides new knowledge for MOOC students in an interesting way for better education. And based on the feedback, we propose ExpanRL, a hierarchical reinforcement learning framework for course concept expansion in MOOCs, which decomposes the concept expansion task into a hierarchy of two subtasks: high-level seed selection and low-level expansion.

Boosted by user feedback on expansion results, ExpanRL jointly learns how to select seed concepts to model the semantic scope of the course better, and whether a concept is beneficial for students. Moreover, the hierarchical reinforcement learning (HRL) structure enables ExpanRL to learn proper expansion strategies instead of the modeling of a particular course, making our model keep a high performance even in unobserved courses.

The evaluation is conducted on 9 datasets from real MOOC courses, compared with 5 representative baseline methods. We further conduct an online evaluation to investigate whether students admit the expanded concepts.

Our contributions include 1) an investigation on how to involve HRL framework into the task of concept expansion; 2) a paradigm that connects the NLP concept expansion task with the educational application; 3) an interactive MOOC environment, consisting of 9 novel datasets of different subjects, 6,553 extracted course concepts, and 495,324 user behaviors from a real MOOC website.

2 Preliminaries

2.1 Problem Formulation

Following (Yu et al., 2019a), **Course Concept Expansion** is formally defined as: given the course corpus \mathcal{D} , course concepts \mathcal{M} , and a knowledge

base \mathcal{KB} as an external source, the task is to return a ranked list of expanded concepts E_c .

In this formulation, a course corpus is defined as $\mathcal{D} = \{\mathcal{C}_j\}_{j=1}^{|n|}$, which is composed of n courses' video subtitles in the same subject area. Course concepts are the subjects taught in the course (such as *LSTM* in Figure 1), denoted as $\mathcal{M} = \{c_i\}_{i=1}^{|\mathcal{M}|}$. (Pan et al., 2017). Knowledge base $\mathcal{KB} = (E, R)$ is consist of concepts E and relations R , which is utilized as an external source to obtain expansion candidates. Though other source (such as Web tables) can also take on this role, we still employ a \mathcal{KB} to search for expansion candidates like the prior work, i.e., $E_c \subset E$.

2.2 Basic Model for Concept Expansion

The general idea of concept expansion is first to characterize the concept set according to its representative elements, then find new candidates and rank them to expand the set.

Seed Selection Stage. A group of representative concepts are called *seeds* and formalized to $K \subset E_c$ (Wang and Cohen, 2007; Mamou et al., 2018). While the expansion process is often carried out iteratively, we also formalize the expansion set of round t to E_c^t . Seed selection is to calculate the possibility that each concept in E_c^t becomes a seed, i.e., $P(c_i \in K^t \subset E_c^t | t)$, where K^t contains the seeds of t -th round.

Based on these *seeds*, we can extract features of the current set and search for candidate concepts for expansion from external sources.

Expansion Stage. After finding a new list of candidates $\mathcal{L}^t = \{c_1, \dots, c_{t'}, \dots, c_{|\mathcal{L}^t|}\}$, expansion stage aims to calculate the likelihood of $c_{t'}$ to be a expanded concept. The top candidates ranked by $c_{t'}$ are selected as new expanded concepts, denoted as N^t the likelihood can be formalized as $P(c_{t'} \in N^t \subset \mathcal{L}^t | K^t, t')$.

The expansion set is refreshed as $E_c^{t+1} = E_c^t \cup N^t$ until its size reaches the preset upper limit τ or cannot find new candidates (He and Xin, 2011).

2.3 Interactive MOOC Environment

The workflow above has been experimentally proven to be effective in many concept expansion tasks (Shen et al., 2018; Rastogi et al., 2019). However, such methods only consider the course concepts' semantic information, which makes their expansion results hard to match real learning needs, especially when dealing with the multi-category

²A course from the University of London in Coursera.

MOOC courses. Meanwhile, since the models are trained before launching, how to maintain high performance on new arisen courses is challenging. Yu et al. (2019a) designs an online game in MOOCs to collect user feedback on the expansion result, thereby employing an active pipeline model to face the above problems, which provides an interactive MOOC environment for reinforcement learning models.

However, the size of publicly published datasets (4 courses with 800 concepts in each course) is still insufficient to meet the need to train advanced deep learning models. Therefore, we extract 68 real MOOC courses of six subjects and build a large-scale MOOC interactive environment, which contains a gamefied interface for feedback collection and several course datasets: “Mathematics”, “Chemistry”, “Architecture”, “Psychology”, “Material Science” and “Computer Science”, covering diverse subjects of natural science, social science and engineering. The details of the datasets are presented in the experiment section.

We construct the environment through three stages. First, for each subject, we select its most relevant courses from a real MOOC website³. We use the method of Pan (2017) to extract the course concepts and manually select the high-quality ones as the course concepts \mathcal{M} . Second, we take XLORE (Jin et al., 2019) as \mathcal{KB} to search for candidate expansion concepts.

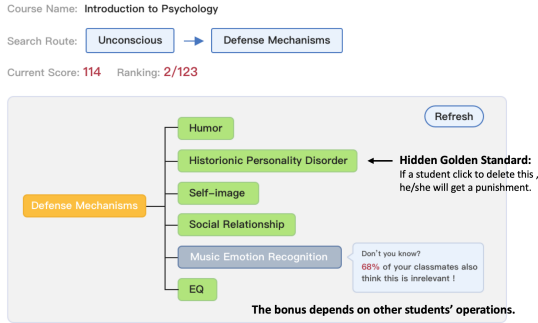


Figure 2: A demonstration of our interactive game in course *Introduction to psychology*. MOOC users can click irrelevant expansion candidates to get bonuses. The yellow concept on the left is from course, and the green concepts are expanded candidates.

Finally, we set up a game to present the expansion candidates. As shown in Figure 2, real MOOC users are drawn to pick out the *course-unrelated* ones to get bonuses. To ensure data quality, we set the game bonus depending on the group voting

³ Anonymous for blind review.

result. We also avoid their irresponsible operations by mixing some extracted course concepts among candidates to detect the spoilers. The operation records are employed to train our reinforcement learning model proposed in the next section.

3 The Proposed Model

In this section, we first introduce our hierarchical reinforcement concept expansion framework, ExpanRL, then present our high-level seed selection model and low-level expansion model separately.

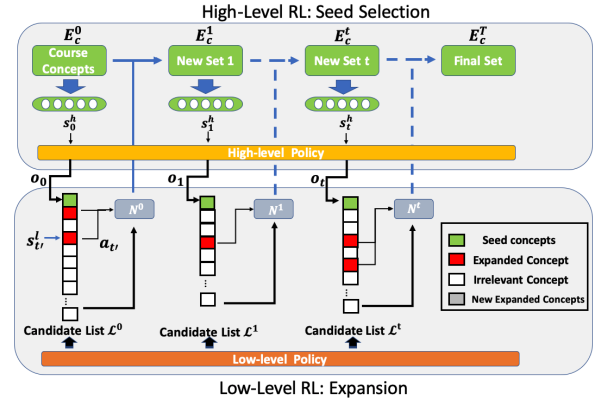


Figure 3: Framework of ExpanRL.

3.1 Overview

To obtain high-quality expanded course concepts for serving students in MOOCs, ExpanRL still needs to address three crucial problems. 1. How to properly utilize user feedback? 2. How to keep accurate modeling of the course during iterations? 3. How to keep a good generalization ability of the model when expanding in new MOOC courses?

Thanks to the interactive MOOC environment, we can deal with these issues by decomposing the basic concept expansion workflow into a hierarchical reinforcement learning framework. Figure 3 shows that the model can learn the complex connection between concepts and courses from user feedback instead of simple contextual information. The main idea of ExpanRL is to upgrade expanding strategies via such an end-to-end model, whose entire expansion process works as the basic concept expansion methods in Section 2.2, which can be naturally formulated as a semi-Markov decision process (Sutton et al., 1999) like : 1) a high-level RL process that selects *seeds* from E_c^t to search for a list of candidates \mathcal{L}^t ; 2) a low-level RL process that detect the high-quality expansion results among candidates and obtain N^t to refresh the set

to E_c^{t+1} . This process iterate until the size of the expansion set reaches the preset limit, τ .

Specially, before the whole process, we first utilize the method in (Pan et al., 2017) to extract course concepts \mathcal{M} from the given course corpus \mathcal{D} and initialize $E_c^0 = \mathcal{M}$.

3.2 Seed Selection with High-level RL

The high-level RL policy μ aims to select k seeds from the existing set E_c , which can be regarded as a conventional RL over options. An option refers to a high-level action, and a low-level RL will be launched once the agent executes an option. The high-level time step t is the expansion round.

Option: The option o_t is a vector consisting of 0 and 1, which represents the i -th concepts from expansion set E_c^t is or is not a selected seed for the current expansion round. Thus the dimension of o_t is the same as the size of E_c^t . When a low-level RL process enters a final state, the agent’s control will be taken over to the high-level RL process to execute the next options.

State: The state $\mathbf{s}_t^h \in \mathcal{S}^h$ of the high level RL process at time step t , is represented by a $k \times C$ matrix reshaped from the hidden state \mathbf{h}_t , where k is the size of seed set and C is the size of a compressed word embedding.

$$\mathbf{s}_t^h = \text{reshape}(\mathbf{h}_t) \quad (1)$$

To obtain the hidden state \mathbf{h}_t , we introduce a set representation RepSet (Skianis et al., 2019) to encode the current expansion set E_c^t . RepSet is unsupervised, order independent and can encode an $n \times V$ matrix to a V dimension vector. Note that $E_c^{t-1} \subset E_c^t$, so the current state is effected by the last state \mathbf{h}_{t-1} .

$$\mathbf{h}_t = \text{RepSet}(E_c^t). \quad (2)$$

Policy: The stochastic policy for seed selection $\mu : \mathcal{S} \rightarrow \mathcal{O}$ which specifies a probability distribution over options:

$$o_t \sim \mu(o_t | \mathbf{s}_t^h) = \mathbf{R}^t = \text{softmax}(\mathbf{s}_t^h \mathbf{W} (\mathbf{E}_c^t)^T). \quad (3)$$

where \mathbf{W} is a learnable parameter, which compresses a V length word embedding to a C length word embedding. \mathbf{E}_c^t is the matrix which consists of all course concepts’ word vector. \mathbf{R}^t is a matrix, while $\mathbf{R}_{j,i}^t$ indicates the possibility of the i -th

concept in E_c^t to be the j -th seed:

$$p(K_j^t = c_i, c_i \in E_c^t | t) = \begin{cases} \mathbf{R}_{j,i}^t \\ 0 \end{cases} \quad \text{if } c_i \text{ is selected before.} \quad (4)$$

And the possibility of the high-level RL to select K^t is shown below. Note that this possibility p is independent of i .

$$p^h(K^t) = \prod_{j=1}^k p(K_j^t = c_i, c_i \in E_c^t | t) \quad (5)$$

Reward: Then, the environment provides intermediate reward r_t^h to estimate the future return when executing o_t . The reward is given by the total reward of the last round of concept expansion.

$$r_t^h = \sum r_{t'}^l(o_t), \quad (6)$$

where $r_{t'}^l(o_t)$ is the low-level reward in time t' while the high-level option is o_t .

Candidate generation after high-level options: After the agent gives out an option o_t , we link the seed concepts from K^t into \mathcal{KB} and find their first-order neighbor concepts as the candidate list \mathcal{L}^t . Note that \mathcal{L}^t is sorted using the pairwise similarity between newly found candidates and seeds.

3.3 Concept Expansion with Low-level RL

Once the high-level policy has selected the seed set and generated a candidate list \mathcal{L}^t , the low-level policy π will scan the list and select high-quality expansion concepts from it to update E_c . The low-level policy over actions is formulated very similarly as the high-level policy over options. The option o_t and K^t from the high-level RL is taken as additional input throughout the low-level expansion process. The time step t' in low-level means the t' -th candidate in \mathcal{L}^t and the final expanded concepts in this round is N^t .

Action: The action at each time step is to assign a tag to the current candidate concept. The action space, i.e., $\mathcal{A} = \{1, 0\}$, where 1 represents the present concept is an expansion result of this set, 0 represents that the concept is not an expansion result.

State: The low-level intra-option state $\mathbf{s}_{t'}^l$ is represented by the word embedding of current expansion candidate $c_{t'}$.

$$\mathbf{s}_{t'}^l = \mathbf{c}_{t'} \quad (7)$$

Moreover, we use a Bi-LSTM (Huang et al., 2015) to provide a hidden state of current candidate list

h_t^l by encoding: 1) the selected seeds K^t , 2) a zero vector as a segmentation, 3) the candidate list \mathcal{L}^t , thereby utilizing the information of high-level option o_t to help low-level decisions.

$$\mathbf{h}_t^l = BiLSTM([K^t; \mathbf{0}; \mathcal{L}^t]) \quad (8)$$

Policy: The stochastic policy for expansion $\pi : \mathcal{S} \rightarrow \mathcal{A}$ outputs an action distribution given intra-option state \mathbf{s}_t^l and the high-level option $o_{t'}$ that launches the current subtask. Here \odot is the vector dot product.

$$\begin{aligned} a_{t'} \sim \pi(a_{t'} | \mathbf{s}_t^l; o_t) &= p^l(c_{t'}) = p(c_{t'} \in N^t | t') \\ &= \text{sigmoid}(\mathbf{h}_t^l \odot \mathbf{s}_{t'}), \end{aligned} \quad (9)$$

Reward: As introduced in section of *Preliminaries*, we construct an interactive game on the MOOC website to collect feedback from users on the expanded concepts. Users can pick out the unrelated concepts of the course, and the picked times of each expansion result c_i is recorded as $\varphi(c_i)$. Since such operations indicate the users' disagreements of the result, the low-level reward is designed to be negatively correlated with $\varphi(c_i)$ as follows:

$$r_{t'}^l = \begin{cases} -\varphi(c_i) / \max_{c_j \in \mathcal{L}^t} (\varphi(c_j)), & a_{t'} = 1 \\ \varphi(c_i) / \max_{c_j \in \mathcal{L}^t} (\varphi(c_j)), & a_{t'} = 0 \end{cases} \quad (10)$$

The count of user clicks determines the degree of relevance of each candidate to the course. It is worth noting that this degree is dynamic and depends on the concept that is mostly picked. This setting effectively controls the range of rewards.

Set refreshment after low-level actions: After the agent gives out an action $a_{t'}$, we can finally obtain the new expanded concepts N^t . The expansion set is updated as $E_c^{t+1} = E_c^t \cup N^t$ and the process turn to another round.

3.4 Hierarchical Policy Learning

To optimize the high level policy, we aim to maximize the expected cumulative rewards from the main task at each step t as the agent samples trajectories following the high-level policy μ , which can be computed as follows:

$$J(\theta_{\mu,t}) = \mathbb{E}_{\mathbf{s}^h, o, r^h \sim \mu(o|\mathbf{s}^h)} \left[\sum_{t=0}^T \log p^h(\mathbf{K}^t) \sum_{s=t}^T \gamma^{s-t} r_s^h \right], \quad (11)$$

where μ is parameterized by θ_{μ} , γ is a discount factor in RL, and the whole sampling process μ takes T time steps before it terminates.

Algorithm 1: Training Procedure of HRL

```

1 Extract course concepts from  $\mathcal{D}$  and initiate  $E_c^0 = \mathcal{M}$ ;
2 Initiate state  $\mathbf{s}_0^h \leftarrow \mathbf{0}$  and time step  $t \leftarrow 0$ ;
3 while  $|E_c| < \sigma$  do
4   Calculate  $\mathbf{s}_t^h$  by Eq.(1);
5   Sample  $o_t$  from  $\mathbf{s}_t^h$  by Eq.(3);
6   Search for candidates from  $\mathcal{KB}$  and generate a ranked candidate list  $\mathcal{L}$ ;
7   for  $j \leftarrow 1$  to  $|\mathcal{L}|$  do
8      $t' \leftarrow t' + 1$ ;
9     Calculate  $\mathbf{s}_{t'}^l$  by Eq.(7);
10    Sample  $a_{t'}$  from  $\mathbf{s}_{t'}^l$  by Eq.(9);
11    Add the expansion result into game and get feedback;
12    Obtain low-level reward  $r_{t'}^l$  by Eq.(10);
13  end
14   $t \leftarrow t + 1$ , refresh  $E_c$ ;
15  Obtain low-level final reward  $r_{fin}^l$ , high-level reward  $r_t^h$ ;
16 end
17 Obtain high-level final reward  $r_{fin}^h$  by Eq.(6);
18 Optimize the model with Eq.(11) and Eq.(12);

```

Similarly, we learn the low-level policy by maximizing the expected cumulative intra-option rewards from the sub task over option o_t when the agent samples along low-level policy $\pi(\cdot | o_t)$ at time step t :

$$J(\theta_{\pi,t}; o_{t'}) = \mathbb{E}_{\mathbf{s}^l, a, r^l \sim \pi(a|\mathbf{s}^l; o_{t'})} \left[\sum_{t'=0}^{T'} \log p^l(\mathbf{c}^{t'}) \sum_{s=t'}^{T'} \gamma^{s-t'} r_s^l \right], \quad (12)$$

if the subtask ends at time step T' .

Then we use policy gradient methods (Sutton et al., 2000) with the REINFORCE (Williams, 1992) algorithm to optimize both high-level and low-level policies. The entire training process is described at Algorithm 1.

4 Experiments

4.1 Experiment Setting

4.1.1 Datasets

We construct an interactive MOOC environment as Section 2.3 to collect user feedback on expansion results. To build a solid evaluation, we randomly selected 5% expanded concepts to be manually labeled benchmarks. For each concept, three annotators majoring in the corresponding domain are asked to label them as “0: Not helpful” or “1: Helpful” based on their knowledge. Thus, each dataset is triply annotated, and Pearson *correlation* coefficient is computed to assess the inter-annotator agreement. A candidate is labeled as a related concept when more than two annotators give positive

	MAT	CHEM	PSY	MS	ARC	CS	MAT+CS	CHEM+MS	MS+ARC
<i>#courses</i>	12	6	16	8	14	12	4	4	5
$ \mathcal{M} $	1,688	1,404	568	842	1,036	1,015	230	417	382
<i>#operations</i>	93,762	103,652	48,492	40,254	120,384	88,779	33,521	52,467	56,787
<i>0-Label</i>	24,278	15,796	13,245	11,876	33,127	17,775	7,092	9,367	7,898
<i>1-Label</i>	6,976	18,755	2,919	1,542	7,001	11,818	3,533	4,790	1,229
<i>correlation</i>	0.712	0.694	0.705	0.732	0.678	0.689	0.655	0.688	0.701

Table 1: Statistics of datasets

tags. Table 1 presents the detailed statistics, where *#courses*, $|\mathcal{M}|$, *1-Label* and *0-Label* are the number of courses, course concepts, positive and negative labels. *#operations* are user click times which is obtained from the game. *MAT*, *CHEM*, *PSY*, *MS*, *ARC* and *CS* correspond to Mathematics, Chemistry, Psychology, Material Science, Architecture and Computer Science.

In particular, we select 13 interdisciplinary courses⁴ and build three multi-category course datasets as *MAT+CS*, *CHEM+MS* and *MS+ARC* to further estimate the performance of ExpanRL on interdisciplinary courses. Note that these three datasets are subsets of the above six’s.

Dataset Usage. All the models are trained on the user operation data and evaluated on the expert annotated data. For the supervised learning baselines, we set the concepts with top 70% click records as negative, and the rest as positive samples.

4.1.2 Basic Settings

All hyper-parameters are tuned on the validation set. The dimension of word vectors in Eq. (2) is 768. The dimension of the compressed word vector C in Eq. (1) is 128. The word vectors of all baseline methods are initialized using BERT (Devlin et al., 2019). The learning rate is 1.0×10^{-4} for low-level RL, and 1.0×10^{-5} for high-level. The discount factor γ is 0.99. The seed size k is set to 10 and the upper limit τ of E_c is 20,000.

4.1.3 Baselines

We compare our hierarchical RL model (denoted as HRL) with five typical methods of set expansion. As these methods obtain expansion candidates from diverse resources, we mainly employ the different similarity metrics to rank the same expansion candidate list for evaluation. Especially to investigate the impact of seed selection strategies, we use a K-means clustering-based method and a pairwise similarity-based method to replace the high-level RL network, which are denoted as C-RL and P-RL.

⁴Course list is shown in Appendix.

- **PR.** Graph based method: We build the candidates and course concepts into a graph. When the similarity between two concepts exceeds a threshold⁵ σ_{PR} , there is a link between them. The PageRank score of each candidate is finally used for sorting. A most famous method employing graph based ranking is SEAL (Wang and Cohen, 2007)

- **SEISA.** SEISA (He and Xin, 2011) is an entity set expansion system developed by Microsoft after SEAL and outperforms traditional graph-based methods by an original unsupervised similarity metric. We implement its Dynamic Thresholding algorithm to sort expanded concepts.

- **EMB.** Embedding based method mainly utilizes context information to examine the similarity between expanded concepts and seeds according to (Mamou et al., 2018). For each expanded concept e , we calculate the sum of its cosine similarities with course concepts \mathcal{M} in BERT (Devlin et al., 2019) and use the average as golden standard to rank the expanded concept list.

- **PUL.** PU learning is a semi-supervised learning model regarding set expansion as a binary classification task. We employ the same setting as (Wang et al., 2017) to classify and sort concepts.

- **PIP.** It is a pipeline method for course concept expansion (Yu et al., 2019a), which first uses an on-line clustering method during candidate generation and then classify them to obtain final expansion results. We follow the workflow of this work to sort expanded concepts.

4.1.4 Evaluation Metrics

Our objective is to generate a ranked list of expanded concepts. Thus, we use the **Mean Average Precision**(MAP) as our evaluation metric, which is the preferred metric in information retrieval for evaluating ranked lists.

4.2 Overall Evaluation

Table 2 summarizes the comparing results of different methods on all datasets. The evaluation is

⁵ σ_{PR} is experimentally set to 0.5.

	MAT	CHEM	PSY	MS	ARC	CS	Avg	MAT+CS	CHEM+MS	MS+ARC	I-Avg
PR	0.763	0.705	0.482	0.470	0.300	0.690	0.568	0.659	0.664	0.401	0.575
SEISA	0.805	0.711	0.473	0.524	0.570	0.713	0.632	0.797	0.691	0.377	0.622
EMB	0.747	0.687	0.474	0.533	0.442	0.812	0.616	0.710	0.655	0.377	0.581
PUL	0.878	0.811	0.845	0.745	0.757	0.850	0.822	0.880	0.782	0.646	0.769
PIP	0.848	0.782	0.803	0.772	0.775	0.821	0.800	0.893	0.835	0.851	0.865
C-RL	0.902	0.795	0.818	0.753	0.716	0.800	0.797	0.851	0.849	0.758	0.820
P-RL	0.892	0.768	0.606	0.749	0.821	0.767	0.835	0.871	0.852	0.662	0.795
HRL	0.903	0.857	0.901	0.806	0.828	0.878	0.862	0.909	0.903	0.886	0.898

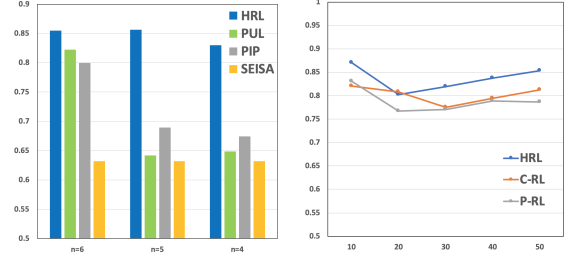
Table 2: MAP of different methods on datasets. (Seed set size = 10)

divided into two parts. The six datasets on the left are the performance of the model on various subjects, and **Avg** represents the average of their MAPs. The three datasets on the right are from the selected interdisciplinary courses, and **I-Avg** is the average of the model performance on them. We also divide the methods into unsupervised, supervised, and reinforcement learning models for further analysis. Overall, our approach HRL maintains an impressive performance (at 0.862 of Avg and 0.898 of I-Avg) over the existing methods, and unsupervised methods (such as SEISA, PR) are not so competitive when compared with methods with supervised information. We lead a detailed investigation to detect the performance among different datasets and the impact of seed selection in the following aspects:

For different datasets, our methods achieve robust results. It is worth noting that the range of the MAP of our method on these datasets does not exceed 0.097, while other baselines suffering from severe oscillations (SEISA of 0.428, EMB of 0.435, and PUL of 0.234). And these supervised methods (PUL, PIP) that perform well on a certain dataset are further analyzed in subsequent experiments.

For the performance on interdisciplinary courses. Most of the baselines meet a decline when turned to interdisciplinary courses. From this angle, PUL can not face this challenge. But PIP, C-RL, and HRL perform even better (with a lift of 0.04 on average), most likely because they all have a clustering-like seed selection process.

For different seed selection strategies. We also detect the impact of seed selection by replacing high-level RL. The comparison among three RL methods shows that: 1) P-RL performs better in one-category expansion tasks (beat C-RL at 0.038); 2) C-RL deal with interdisciplinary courses better than P-RL (as discussed above); 3) HRL is stronger than these two methods in all datasets. The results



(a) The MAP of different number of training sets. (b) The MAP of seed sizes.

Figure 4: Performance of different settings. (a) shows the average MAP when mask some of the datasets in training. (b) shows the MAP of different seed size.

exactly prove the superiority of HRL’s seed selection over rule-based strategies.

4.3 Result Analysis

Generalization Ability. Expansion models in MOOCs need to face with plenty of new courses every day. Thus we lead strict experiments to estimate the generalization ability of the model by masking training datasets. For example, the bar of $n = 5$ in Figure 4(a) indicates the average MAP when the models are trained on five subject datasets and tested on the other one. Thus $n = 6$ is the average MAP in Table 2 while $n = 5$ and $n = 4$ present the results of facing one or two kinds of new courses. Here we select HRL, PUL, and PIP for observation. Such an experiment shows that HRL still maintains an outstanding performance in new courses. Still, PIP and PUL suffer from a sharp decline in untrained new datasets (even at the same level as unsupervised methods).

The size of seed set k . For different settings of seed sizes, we compare the performance of ExpanRL with other RL based baselines. As shown in Figure 4(b), HRL keeps a high level of MAP among these settings (all over 0.8 on average). Meanwhile, we find that all these RL-based methods perform

	Cr@10	Cr@20	Cr@50
PR	0.097	0.182	0.425
SEISA	0.097	0.204	0.459
EMB	0.071	0.150	0.359
PUL	0.041	0.091	0.349
PIP	0.069	0.126	0.342
HRL	0.036	0.082	0.258

Table 3: Online Evaluation results.

better in small or large seed size (less than 10 or larger than 40), which requires future detection on this phenomenon.

Discussion. Based on the above experimental results, we summarize the analysis as follows: 1) the performance of unsupervised methods on different datasets is not as stable as the supervised or RL methods; 2) except for models that have a clustering-like seed selection process (PIP, C-RL, HRL), most models suffer from declines on interdisciplinary datasets; 3) although supervised models (PIP, PUL) perform well in some cases, they drastically decline in untrained new courses; 4) HRL, consisting of a feasible seed selection RL and expansion strategies from human efforts, keep a high performance under different settings. HRL deal with the challenges in MOOC expansion tasks, as claimed in the introduction.

4.4 MOOC Online Evaluation

Utilizing user feedback on the expansion results from our interactive MOOC environment, we also set up an online evaluation to detect whether users agree on the expansion results. Following the same evaluation metric in (Yu et al., 2019a), we denote **Click Rate** as $Cr@q$, which means the click rate of top q expanded concepts, i.e.,

$$Cr@q = \sum_{i=1}^q \varphi(c_i) / \sum_{j=1}^{|E_c|} \varphi(c_j) \quad (13)$$

A smaller $Cr@q$ indicates more users think the results are relevant to the course. We record the performance of each method in Table 3. Results show that ExpanRL obtains the best feedback from MOOC users under all three settings. It’s worth noting that the advantage of ExpanRL is evident while selecting larger-scale samples (The overlap rises from 0.005 to 0.091), which indicates that our model can provide more high-quality concepts.

5 Related Work

Our work follows the task of concept expansion in MOOCs (Yu et al., 2019a), a particular type of set

expansion problem, which takes several seeds as input and expands the entity set.

Set expansion was born to serve knowledge acquisition applications on the Internet. Google Sets was a pioneer which leded a series of early research, e.g. Bayesian Sets (Ghahramani and Heller, 2006), SEAL (Wang and Cohen, 2007), SEISA (He and Xin, 2011) and others (Sarmiento et al., 2007; Shi et al., 2010; Wang et al., 2015). These efforts utilize web tables as a resource and mainly serves for search engines. Recently, more related research has turned its attention to other application fields, such as news mining (Redondo-García et al., 2014), knowledge graphs (Zhang et al., 2017), education assistance (Yu et al., 2019a), etc. Meanwhile, corpus-based expansion methods snowball, and iterative bootstrapping became a common solution (Shen et al., 2017; Yu et al., 2019b; Yan et al., 2019), which expands the set in round and select high-quality results to extract feature iteratively. ExpanRL is inspired by this type of method and is designed to optimize the existing iterative process.

ExpanRL also benefits from hierarchical reinforcement learning (HRL), which has been employed in many NLP tasks (Zhang et al., 2019; Takanobu et al., 2019) and achieved impressive results. By decomposing complex tasks into multiple small tasks to reduce the complexity of decision making (Barto and Mahadevan, 2003), HRL naturally matches the iterative set expansion tasks.

6 Conclusion and Future Work

We investigate the task of course concept expansion, which utilizes the NLP approaches in improving MOOC education. After constructing a novel interactive MOOC environment to collect user feedback on expansion results, we design a paradigm, ExpanRL, which decomposes the concept expansion task into a hierarchy of two subtasks: high-level seed selection and low-level concept expansion. Experiment results on nine datasets from real MOOCs prove that ExpanRL can better serve students by recognizing the helpful expanded results and maintaining good performance in interdisciplinary courses and even new courses.

Promising future directions include detecting how to ensemble supervised learning and RL expansion models and applying the proposed model in related tasks. We also hope our design of interactive games can call for more fancy methods that utilize student feedback in NLP applications

in Education.

Acknowledgement

This work is supported by the National Key Research and Development Program of China (2018YFB1004503), NSFC Key Projects (U1736204, 61533018), grants from Beijing Academy of Artificial Intelligence (BAAI2019ZD0502), Institute for Guo Qiang, Tsinghua University (2019GQB0003), and XuetaoX.

References

- Ausubel. 1968. Educational psychology: A cognitive view.
- Andrew G Barto and Sridhar Mahadevan. 2003. Recent advances in hierarchical reinforcement learning. *Discrete event dynamic systems*, 13(1-2):41–77.
- James R Curran, Tara Murphy, and Bernhard Scholz. 2007. Minimising semantic drift with mutual exclusion bootstrapping. In *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*, volume 6, pages 172–180.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Zoubin Ghahramani and Katherine A Heller. 2006. Bayesian sets. In *Advances in neural information processing systems*, pages 435–442.
- Yeye He and Dong Xin. 2011. Seisa: set expansion by iterative similarity aggregation. In *Proceedings of the 20th international conference on World wide web*, pages 427–436.
- Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.
- Hailong Jin, Chengjiang Li, Jing Zhang, Lei Hou, Juanzi Li, and Peng Zhang. 2019. Xlore2: Large-scale cross-lingual knowledge graph construction and application. *Data Intelligence*, 1(1):77–98.
- Jonathan Mamou, Oren Pereg, Moshe Wasserblat, Alon Eirew, Yael Green, Shira Guskin, Peter Izsak, and Daniel Korat. 2018. Term set expansion based nlp architect by intel ai lab. *EMNLP 2018*, page 19.
- Liangming Pan, Xiaochen Wang, Chengjiang Li, Juanzi Li, and Jie Tang. 2017. Course concept extraction in moocs via embedding-based graph propagation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, pages 875–884.
- Pushpendre Rastogi, Adam Poliak, Vince Lyzinski, and Benjamin Van Durme. 2019. Neural variational entity set expansion for automatically populated knowledge graphs. *Information Retrieval Journal*, 22(3-4):232–255.
- José Luis Redondo-García, Michiel Hildebrand, Lilia Perez Romero, and Raphaël Troncy. 2014. Augmenting tv newscasts via entity expansion. In *European Semantic Web Conference*, pages 472–476. Springer.
- Luis Sarmento, Valentin Jijkuon, Maarten De Rijke, and Eugenio Oliveira. 2007. More like these: growing entity classes from seeds. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 959–962.
- D Shah. 2019. Year of mooc-based degrees: A review of mooc stats and trends in 2018. class central. *Class Central's MOOC Report*.
- Jiaming Shen, Zeqiu Wu, Dongming Lei, Jingbo Shang, Xiang Ren, and Jiawei Han. 2017. Setexpan: Corpus-based set expansion via context feature selection and rank ensemble. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 288–304.
- Jiaming Shen, Zeqiu Wu, Dongming Lei, Chao Zhang, Xiang Ren, Michelle T Vanni, Brian M Sadler, and Jiawei Han. 2018. Hiexpan: Task-guided taxonomy construction by hierarchical tree expansion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2180–2189.
- Shuming Shi, Huibin Zhang, Xiaojie Yuan, and Ji-Rong Wen. 2010. Corpus-based semantic class mining: distributional vs. pattern-based approaches. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 993–1001.
- Konstantinos Skianis, Giannis Nikolentzos, Stratis Limnios, and Michalis Vazirgiannis. 2019. Rep the set: Neural networks for learning set representations. *arXiv preprint arXiv:1904.01962*.
- Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. 2000. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*, pages 1057–1063.
- Richard S Sutton, Doina Precup, and Satinder Singh. 1999. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211.

Ryuichi Takanobu, Tianyang Zhang, Jiexi Liu, and Minlie Huang. 2019. A hierarchical framework for relation extraction with reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7072–7079.

Chi Wang, Kaushik Chakrabarti, Yeye He, Kris Ganjam, Zhimin Chen, and Philip A Bernstein. 2015. Concept expansion using web tables. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1198–1208.

Richard C Wang and William W Cohen. 2007. Language-independent set expansion of named entities using the web. In *Seventh IEEE international conference on data mining (ICDM 2007)*, pages 342–350. IEEE.

Yasheng Wang, Yang Zhang, and Bing Liu. 2017. Sentiment lexicon expansion based on neural pu learning, double dictionary lookup, and polarity association. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 553–563.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256.

Lingyong Yan, Xianpei Han, Le Sun, and Ben He. 2019. Learning to bootstrap for entity set expansion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 292–301, Hong Kong, China. Association for Computational Linguistics.

Jifan Yu, Chenyu Wang, Gan Luo, Lei Hou, Juanzi Li, Zhiyuan Liu, and Jie Tang. 2019a. Course concept expansion in moocs with external knowledge and interactive game. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 4292–4302.

Puxuan Yu, Zhiqi Huang, Razieh Rahimi, and James Allan. 2019b. Corpus-based set expansion with lexical features and distributed representations. In *Proceedings of the 42Nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR’19*, pages 1153–1156.

Jing Zhang, Bowen Hao, Bo Chen, Cuiping Li, Hong Chen, and Jimeng Sund. 2019. Hierarchical reinforcement learning for course recommendation in moocs. *Psychology*, 5(4.64):5–65.

Xiangling Zhang, Yueguo Chen, Jun Chen, Xiaoyong Du, Ke Wang, and Ji-Rong Wen. 2017. Entity set expansion via knowledge graphs. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1101–1104. ACM.

A Dataset Analysis & Case Study

We also analyze the characteristics of the datasets and do a case study to explore further the impact of different expansion tasks on the model, which will help choose the appropriate expansion model for various tasks.

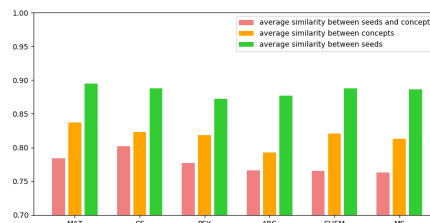


Figure 5: The average pairwise similarity of seeds, expanded concepts and seed-expand concept pairs.

We assess the degree of dispersion of the concepts from different subjects by calculating the pairwise average similarities. Combining the results in Overall Evaluation and Figure 5, we find the science subjects, MAT and CHEM, obtain the most aggregated concepts (Green bars in Figure), which also leads to a booming of all methods in this two datasets. Simultaneously, unsupervised models (SEISA, EMB) show significant performance degradation on PSY and ARC datasets, with the lowest average similarity of expansion results (Red and orange bars). This demonstrates the critical role of supervisory information in complex set expansion.

The contest between the supervised learning methods (PUL, PIP) and the RL methods can be observed more intuitively through the case study. We sample some errors from ARC and CHEM datasets in Figure 6. It is easy to find that the errors of supervised learning methods mainly come from some noise words, e.g., the word “architecture” in *computer architecture*. However, the errors of RL methods are mainly caused by classification, e.g., *electric potential energy* is highly relevant to chemistry, but it is a physics concept.

From this phenomenon, we speculate that SL knows more about the context of the concept, and RL understands the meaning of the concept better. Therefore, the joint method of combining supervised learning and RL is likely to be a promising research direction in expansion tasks.

	False Positive	False Negative
Chemistry	<div>Computer architecture</div> <div>Astronomical</div>	<div>Glass</div> <div>Richard B. Fuller</div> <div>Rubber bearings</div>
Architecture	<div>Matrix factorization</div> <div>Electric potential energy</div>	<div>Claude Ruggieri</div> <div>Resonance enhancement</div>

Figure 6: Some error cases in ARC and CHEM datasets. Blue concepts are errors from supervised methods, orange ones are from RL methods and black is the shared errors.

B List of interdisciplinary courses

In this section, we list the selected interdisciplinary courses to present this situation in real MOOCs. As shown in Table 4, many courses from MOOCs are related to more than one subject; this is a common phenomenon in practical teaching. The URLs of these courses are hidden for blind review.

Domain	CourseName
MAT+CS	Introduction to Data Science
	Computational Geometry
	Algorithm of Big Data
	Multivariate statistical analysis and R language modeling
CHEM+MS	Plant Fiber Chemistry
	Chemical Reaction Engineering
	Magical Material World
	Catalyst Design and Preparation
MS+ARC	Construction Materials
	Architecture Materials
	Explore the Materials Around You
	Road Engineering Materials
	Reinforced Concrete and Masonry Structures

Table 4: The list of selected interdisciplinary courses.