Recommending Remedial Readings Using Student Knowledge State

Khushboo Thaker, Lei Zhang, Daqing He, Peter Brusilovsky
School of Computing and Information
University of Pittsburgh
Pittsburgh, PA, USA
k.thaker, lez39, dah44, peterb@pitt.edu

ABSTRACT

Assessment plays a vital role in learning, as it provides both instructors and students with feedback on the overall effectiveness of their teaching or learning. However, when a student fails to correctly answer certain questions in an assessment (such as a quiz), the student needs specific recommendations that are tailored to their learning needs and to the knowledge deficiency exposed by the assessment outcomes. In this paper, we explore the methods for automatically identifying the recommended textbook materials that are most relevant and suitable to the student. In particular, we conducted experiments on how to incorporate students' current knowledge state on domain concepts associated with the activity to recommend personalized remedial sections to each student. The results show that incorporating student knowledge states can significantly improve the quality of recommendations as compared to traditional content-based recommendations.

Keywords

Remedial Recommendation, student Modeling, domain concepts, dynamic student knowledge

1. INTRODUCTION

Along with the rapid development of internet and communication technologies, as well as the increasing amount of online materials in diverse formats, online learning and its supporting platforms have become vital for learning various new subjects. Regardless of if it is a self-regulated platform or instructor-regulated platform, learners are provided with diverse types of content, which may include notes, textbooks, videos, and other lecture material. Similar to traditional learning, in order to evaluate a learner's progress through course materials, various forms of assessments are embedded in the online learning process. For example, course platforms integrate quizzes and exams at the end of each learning module (section, subsection, or part). This is particularly important in self-regulated online courses, since these assessments help learners to reflect on the content and estimate their learning progress. A complete learning loop should incorporate the provision of providing learners' relevant remedial content materials to compensate for the knowledge deficit exposed by the assessment. In classic computer-assisted instruction (CAI), where the course content and assessments were created either by the same author or by a team, links to remedial content were created manually. However, modern online learning extensively uses open educational resources and question banks created by many independent authors. In this context, an automatically generated remedial recommendation of learning content after a failed assessment is vital to the success of online learning.

A natural approach for an educational recommender system is to use content similarity as the basis for remedial recommendation [36]. This approach recommends remedial content that is similar to the assessed content. However, deficiencies in student knowledge that are exposed by the assessment might not be limited to the most similar content. Thus, a content similarity-based approach could lead to a recommendation of materials that have either already been mastered by the student, or a recommendation of material for which students lack the prerequisite knowledge.

Q: A person searches for "Michael Jordan sport" in google search engine. Please mark all the possible information needs of the person:

A. A person who is part of some sport
B. A sport named Michael Jordan
C. Michael from country Jordan

Figure 1: Quiz example

The goal of this study is to explore the method of remedial recommendation that dynamically address student needs. The proposed approach is to focus on modeling the domain-relevant concepts that the student is learning. The motivation for using domain-specific concepts in representing recommended documents is illustrated by Figure 1. The figure shows an example question from the "information retrieval" course. A term-based recommendation (keyword-

Khushboo Thaker, Lei Zhang, Daqing He and Peter Brusilovsky "Recommending Remedial Readings Using Student's Knowledge state" In: *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, Anna N. Rafferty, Jacob Whitehill, Violetta Cavalli-Sforza, and Cristobal Romero (eds.) 2020, pp. 233 - 244

based) could count repeated words in the question and bring up documents related to "Michael Jordan" or "Sport". However, these documents may be irrelevant and thus would not be helpful for remediation. In contrast, a recommendation algorithm that attends to domain-specific concepts would select documents that are related to "Google search engine" or "information need", which can cause the recommended documents to be more directly relevant to the knowledge assessed in the question.

We believe that a beneficial remedial recommendation also needs to dynamically model the student's knowledge of the domain-specific concepts. A content-based or domain concept-based static remedial recommendation would provide the same recommendation to different students, regardless of the individual students' different levels of knowledge deficiency exposed by the assessment outcomes. For instance, still considering the example in Figure 1, students who are already familiar with the concept of "search engine" but are struggling with the concept of "information need" would need different remedial recommendations than those who are struggling with the concept of "search engine".

As a result, in this paper we will investigate the effects of incorporating both domain knowledge and student knowledge in remedial recommendations. More specifically, we aim to address the following research questions:

- **RQ 1**: Does the domain-based representation of educational content help perform remedial recommendation, either by acting alone or in combination with the content-based recommendation?
- **RQ 2**: Can we use automated keyphrase extraction techniques to generate domain-based representation ?
- **RQ 3**: Does the augmentation of student knowledge on domain-based recommendation help in providing dynamic remedial recommendations?

To address the research questions, we proposed a concept-level static remedial recommendation (StatRemRec) and a knowledge-level dynamic remedial recommendation (DynRemRec). To conduct this research, we used an online reading platform (ReadingCircle) [13]. The system provided the students with a platform to read their course textbook. In ReadingCircle, every subsection contains a quiz to test student performance (for details 4.1). Data obtained from the ReadingCircle platform was used to investigate and evaluate the proposed recommendation approach. We hypothesize that the StatRemRec and DynRemRec approaches will provide better recommendations than state-of-the-art recommendation models that completely ignore both the domain and student dynamic knowledge states.

We also released a dataset of annotated questions in our existing online textbook with relevant sections. This data can work as a benchmark for content recommendation and linking task[1]. The code and student model are available at: https://github.com/khushsi/RemRec

[1]https://pslcdatashop.web.cmu.edu/Project?id=637

2. RELATED WORK

Despite their overall similarity, the roots of both static and dynamic remedial recommendation approaches can be traced to two different research areas. The *static recommendation* of educational resources does not depend on the state of an individual student, and as a result, can be generated before a student starts working with learning content. Historically, static recommendations were explored in the field of educational hypermedia and called "intelligent hypertext", since this approach recommended resources that were not connected by a human-authored link. Research on intelligent hypertext started in the early days of this field and originally focused on linking resources using term-based resource similarity [22, 41]. Simple keyword-based approaches have been gradually replaced by semantic-level similarity, based on concepts of semantic web and domain ontology [8, 28], and later by modern text-processing approaches, such as topic modeling and concept extraction [24, 1, 37, 14].

The emergence of MOOCs and the online accumulation of large volumes of educational content encouraged a new wave of research on "intelligent" linking focused on connecting primary learning content, such as textbooks and MOOCs, with different external learning resources, such as videos, Wikipedia pages, or research papers [1, 18, 20].

In contrast, the *dynamic recommendation* model of educational content has to be generated on the fly, based on the current knowledge or interest of the learner. Dynamic recommendations could be traced back to the classic works on adaptive course "sequencing" [23] and generation [10]. The first generation of this work focused on adaptation to student levels of knowledge and used different student modeling approaches from the field of intelligent tutoring [35]. The emergence of recommender systems encouraged a different generation of research on dynamic recommendation that focused on learner interests and used techniques from the areas of recommender systems [21]. Due to its popularity, the term "recommendation" is now used to refer to both knowledge-based and interest-based recommendations. Recent work on educational recommendation frequently combines both knowledge and interest adaptation and supports a range of needs, such as fine-grained resource recommendation for practice activities [2, 37], reading materials [29], and videos, as well as coarse-grained recommendation of courses [30, 7] or textbooks [31].

The majority of research on educational recommendation has focused on recommending students' next thing to do and assumes that the student's overall progress is good. A different recommendation approach, known as remedial recommendation [3], has focused on recommending resources that can help a student to learn a concept in which a student is weak, in order to improve understanding or resolve misconceptions. Konstantin et al. [4] proposed a knowledge-gap based remedial recommendation approach. The method considers learners' previous success rates and categorize learners as expert, intermediate, or unknown. They found that this coarse-grained categorization may help in providing recommendations based on student needs. Although such a coarse-grained categorization is beneficial, it assumes that there is a single learning rate for all students. However, the existing advancement in education technologies have ways to infer students' individualized levels of knowledge [32] and learning rates [9, 11], called student models. In our work, we used student models to define fine-grained student knowledge states and explored the possibility of using them in remedial recommendations.

3. METHODOLOGY

In this paper, we investigated the effect of dynamically incorporating domain knowledge and student knowledge for remedial recommendation. The intuition is that this dynamic incorporation can enable more relevant and suitable resources to help students recover from failure within the assessment.

3.1 Problem Description

Formally, the research problem can be described as: for a given student S, who was recently assessed on question q; if student S failed on question q, we want to provide student S with a recommended reading from a content set $T = t_1, t_2, \ldots, t_c$ to help the student to grasp the knowledge that would be required to succeed on question q.

The vector representation of texts T and question q are constructed using domain concepts, the recommendation is computed based on the cosine similarity between the vectors of text T and question q, and the top five most similar texts are selected for the recommendation.

3.2 Static Remedial Recommendation (StatRemRec)

StatRemRec targets remedial recommendation based on incorporating semantic knowledge or domain knowledge of the content. To build a domain-based representation of education material, we used domain concepts. The approach of a domain concept-based representation of education material is commonly found in intelligent tutoring systems, which focus on problem-solving support and where every practice problem is associated with a set of domain knowledge components (concepts) [17]. In our case, concepts are expressed as key phrases. Each key phrase depicts a fragment of domain knowledge, a semantic entity, or a fine-grained topic. Each target education material, textbook section, and question (in our case) are annotated with domain concepts. Figure 2 shows an example output of these annotated domain concepts mapped to both a text and a question.

Once we have obtained a domain concept for both texts and questions, we build a representation of texts and questions as a frequency-based vector representation, based on the presence of domain concepts for each text and each question. For recommending texts for a particular question, we apply cosine similarity between question q and all the available sections in the text T and rank the top five most similar texts $\{R_q^1, R_q^2, R_q^3, R_q^4, R_q^5\}$ that share the same domain concepts with the questions q

3.3 Dynamic Remedial Recommendation (DynRemRec)

StatRemRec accounts for semantic knowledge in the document for recommending remedial materials. Although this is an improvement on a purely keyword-based content recommendation system, StatRemRec still recommends the same content for each student, regardless of the student's real-time content requirement. For example, a student failing on a question that asks about "Multiplication" will always be given recommendations for readings related to "Multiplication" with StatRemRec. For instance, if a student's skills are weak in a prerequisite concept, e.g. "Addition", it is crucial to support that student's current needs.

In education systems, intelligent tutoring systems account for this student-specific information to provide students with adaptive practices and has been shown to help with effective and efficient learning [38]. To provide adaptation, the tutors maintain dynamically changing student knowledge states while the student uses the tutoring system.

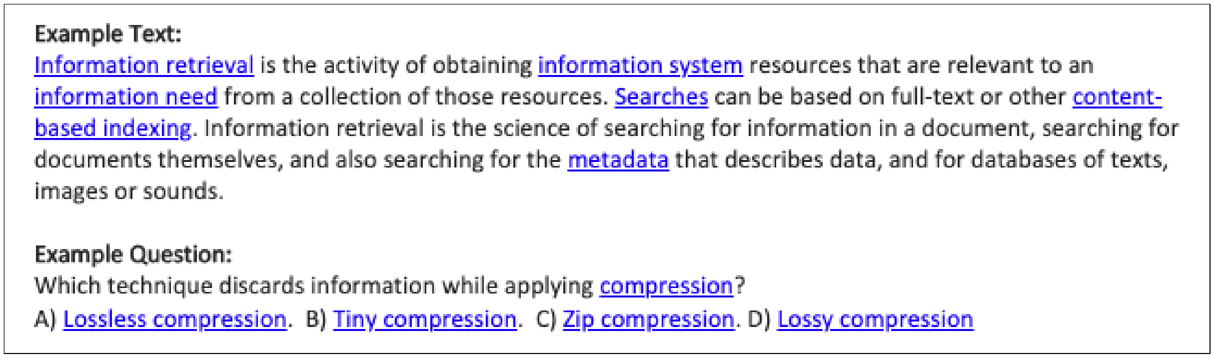
In DynRemRec, we maintained dynamically changing student knowledge states and used students' real-time knowledge states to generate a personalized remedial recommendation for each student. The following subsections provide details about student knowledge state generation and our approaches to integrating them into our remedial recommendation.

3.3.1 Student Knowledge State Generation

For generating students' knowledge state, we used a traditional and widely accepted student modeling framework, performance factor analysis (PFA) [34]. This model relies on expert annotated *skills* (also known as knowledge components or concepts). *Skills* are knowledge units associated with student activities, steps, and questions on which students' knowledge and performance are tested [17]. In our work, we considered *domain concepts* as *skills*, which has been shown to work in previous work on student modeling in online textbooks. [40, 16, 42]. At the base of the model is a Qmatrix, a binary matrix where columns represent *concepts or skills* and rows represent questions. Each cell is a binary value, where 1 in the cell with row r and column c represents that question r is an application of concept c. PFA represents the student's probability of success in answering a question as a function of the student's previous successful and failed attempts on the *concept* associated with the question, as shown in Equation 1

$$\textbf{PFA:} \quad \ln \frac{p_{sq}}{1 - p_{sq}} = \alpha_s + \sum_c \beta_c Q_{cq} + \sum_c Q_{cq}(\mu_c S_{sc} + \rho_c F_{sc}) \quad (1)$$

where, s is a student and q is a question. c is a *concept* (skill or knowledge component). α_s is a coefficient associated with learner s (regression intercept) and represents the proficiency of learner s. Q is a Qmatrix and Q_{cq} is the Qmatrix cell associated with question q and *Concept* c. β_c are coefficients associated with concept c. β_c represents the difficulty of concept c, while μ_c and ρ_c are coefficients associated with S_{sc} and F_{sc}. S_{sc} and F_{sc} are the number of success and failure attempts, respectively, of learner s on concept c. We consider PFA, as PFA provides granular evaluation based on individual students' prior success and failure on a particular skill [34].


Example Text:
Information retrieval is the activity of obtaining information system resources that are relevant to an information need from a collection of those resources. Searches can be based on full-text or other content-based indexing. Information retrieval is the science of searching for information in a document, searching for documents themselves, and also searching for the metadata that describes data, and for databases of texts, images or sounds.

Example Question:
Which technique discards information while applying compression?
A) Lossless compression. B) Tiny compression. C) Zip compression. D) Lossy compression

Figure 2: An example of domain concepts annotated to both a text and a question. This is just for example purposes and is not part of our study. The domain concepts are marked with a link

To obtain the knowledge of a student s on a concept c when the student failed on question q, we generate the probability of failure on a concept PF_{sc}. We assumed that there exists an item c annotated with concept c, and generated the probability of failure as:

$$PF_{sc} = \begin{cases} 1 - (\alpha_s + \beta_c + (\mu_c S_{sc} + \rho_c F_{sc})) & \text{c} \in q \\ 0 & c \notin q \end{cases} \tag{2}$$

where α_s is ability of student s and β_c is difficulty of concept c. S_{sc} and F_{sc} are previous success and failed attempts of student s on concept c, after the student failed on question q.

To generate the student knowledge state vector, we represented each domain concept associated with the question q with weight value PF_{sc}. The knowledge state consists of the probability of failure to make sure that a greater weight is given to domain concepts ($skills$) where the student has a high probability of failure (where they might lack sufficient knowledge). For the concepts that are not associated with question q, we made the probability to be zero, as the goal is to recommend material related to concepts that are associated with the question.

The representation of text (documents or textbook sections) is the same for both DynRemRec and StatRemRec (i.e. representation based on frequency on domain concepts, as discussed in Section 3.3). The change is in question representation, which is based on the presence of domain concepts in StatRemRec and is based on the dynamic knowledge state of domain concepts in DynRemRec.

4. EXPERIMENTS

The dataset from ReadingCircle [13] was used for exploring DynRemRec and StatRemRec. In this section, we will introduce the dataset before presenting the details of our experiments.

4.1 Student Dataset

ReadingCircle [13] is an online reading platform. It provides an online reading environment to students in a course where they read assigned textbook materials to prepare for class. There are quizzes of questions embedded in each section of the assigned readings to assess the progress of student learning on the content.

ReadingCircle keeps extensive logs for events associated with student reading and assessment. The dataset used in the experiments is collected from a version of ReadingCircle that has been adapted for supporting a graduate-level course on information retrieval at an University of Pittsburgh in spring 2016. There was no restriction on the number of attempts to the questions. ReadingCircle logs each and every attempt made by the student. This data set contains 9006 quiz interactions from 22 students and 4273 interactions of student failure on quizzes (for more details, refer to Table 1). The student dataset can be obtained from Datashop[2].

Table 1: ReadingCircle data details

Number of documents (sections)	66
Number of questions	89
Number of students	22
Average per student questions attempted	91
Student practice interactions	9006
Number of failure interactions	4273

4.1.1 Ground Truth

To evaluate the effect of recommendations on questions, we require a mapping of questions to sections. Each question maps to the section where it appears. This assumption holds as each quiz is created by subject experts (instructors and teaching assistants) to assess student knowledge in a particular section. In a few cases, a quiz will assess multiple sections. In this case, we map the questions appearing in those quiz to multiple sections. For more details on the question-to-section mapping dataset, please refer to Table 2

Table 2: Ground truth for recommendation

Number of (sections)	66
Number of questions	89
Number of questions per section	1.93
Number of questions linked to single section	81
Number of questions linked to multiple sections	8

[2]https://pslcdatashop.web.cmu.edu/Project?id=637

4.2 Domain Concept

Concept-based textbook representation was introduced by early projects that focused on adaptive textbooks [15, 43]. Adaptive textbooks associated every section of a digital textbook with a set of domain concepts (called outcomes) that are present in that section. In this work, we investigated both expert-annotated *domain concepts* [42] and automated extracted *domain concepts* [26, 6, 25].

- **Expert-based concepts (EBC):** For EBC, we used concepts that were generated by Wang et al [42]. Wang et al. [42] developed comprehensive expert-based annotation rules and proposed a two-step concept annotation system with three subject experts. [42]. The concepts are available for sections in the "Introduction of Information Retrieval" book, which is the same book that students are reading in the online course in ReadingCircle.

  In order to conduct the experiment, we want both questions and the text that are associated with the concepts. However, EBC is only available for textbook sections. In order to associate concepts with questions, we created a list of domain concepts using concepts in all sections, and performed a simple lookup on the concept list to find the domain concepts in the question and answer text. More details about the EBC concepts is mentioned in Table 3.

Table 3: EBC Concepts

Number of unique concepts	1047
Average number of concepts per section	30.83
Average number of concepts per quiz	6.52

To check if questions have concepts only from related sections, we plotted the distribution that depicts the number of unique sections that share concepts with questions, as shown in Figure 3. These statistics show that questions share concepts with an average of 24.3 sections.

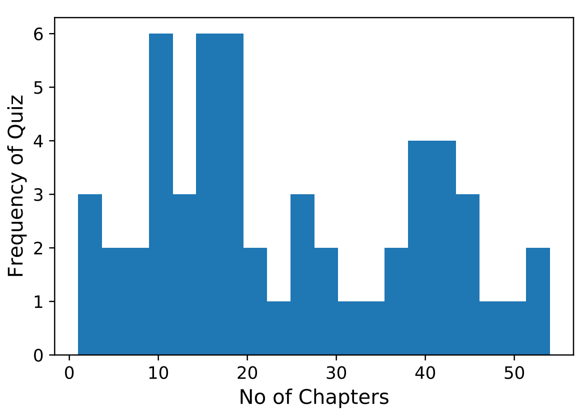


Figure 3: Distribution of number of unique sections sharing domain concepts per question.

- **Automated concept extraction (ACE):** StatRemRec and DynRemRec represent text with concept-level representation. In the case of DynRemRec, the student model is also trained using concepts associated with student practice activities. However, traditional expert-based concept generation is time consuming and hard to obtain with the incorporation of open-source course materials. Hence, it is impractical if time-consuming expert concepts becomes a necessary step in student recommendation on education resources. To prove the ease of incorporating StatRemRec and DynRemRec, we dedicated this set of experiments to test the feasibility of concept-level remedial recommendations (RQ 2). The experiment was designed to test StatRemRec and DynRemRec on concepts that were automatically extracted through keyphrase extraction techniques [6, 25, 26]. It is debatable if keyphrase extraction techniques extract domain concepts and could be used as knowledge components or *skills* on which students' knowledge is measured. For evidence of using concepts as *skills*, we relied on evidence from work by Thaker et al. [40] and Huang et al. [16], in which student models are trained on keyphrases in education content for adaptive textbooks.

  To perform the experiment and explore research question RQ-2, we selected three state-of-the art keyphrase extraction techniques, as discussed below:

  1. TextRank: TextRank [26] is a classic unsupervised keyphrase extraction technique. TextRank converts each document into a graph of words, based on word co-occurrence criteria. The algorithm then applies the page rank algorithm to the graph and extracts the important keyphrases.
  2. CopyRNN: CopyRNN [25] is a supervised deep-learning based sequence-to-sequence keyphrase generation technique. CopyRNN is one of the state-of-the-art supervised keyphrase extraction techniques. This will help us evaluate our model for supervised keyphrase extraction.
  3. TopicRank: TopicRank [6] is a graph-based unsupervised keyphrase extraction technique. The difference is TopicRank focuses on finding keyphrases that belong to the topic of the document. As a result, this technique can provide more insight into topic-based concept extraction.

  .

  Table 4 shows more details of the concepts extracted by different ACE methods. The table indicates that different algorithms will choose a different domain space for representing the domain.

4.3 Term-Based Recommendation Baseline (TextRec)

For our baseline, we used a simple term-based recommendation approach. TextRec finds the similarity between two documents (section text and quiz text) based on words that are present in the text. Each term in the document is used as a semantic unit and the document is represented as a vector of the TF-IDF weights [39]. Such a TF-IDF based document similarity was recently found to be effective for

Table 4: Statistics of ACE datasets

Model	Average no of Concepts Section	Quiz	Unique Concepts	Overlap Sections
CopyRNN	14.56	2.32	558	2.29
TextRank	27.39	7.53	698	28.88
TopicRank	96.01	7.85	2469	32.94

finding similar education resources [37]. Although state-of-the-art recommendation techniques use advanced semantic representations that use both word and document embeddings [27], we did not explore much in this area, as our focus is in understanding the effectiveness of including domain and student knowledge in recommending remedial resources.

4.4 Experiment Steps

To address the research questions mentioned in Section 1, we conducted the following experiments:

1. **Term vs Concept Representation:** To understand if domain-based representation is effective (RQ 1), we compared the concept-level techniques of StatRemRec and DynRemRec to the term-based approach TextRec.

2. **Fusion experiment:** As the TextRec approach has a term-based representation of education material, while StatRemRec and DynRemRec use concept-level representation, to leverage both types of representation (RQ 1), we fused the term-based approach with domain concept-based approaches. The fusion of term-based methods with concept-based methods is done by simple linear interpolation, as specified in Equation 3

$$Sim^{fused}_{q_i,t_j} = \alpha \cdot Sim^{concept}_{q_i,t_j} + (1-\alpha) \cdot Sim^{text}_{q_i,t_j} \quad (3)$$

   where Sim_{q_i,t_j} is the similarity between question q_i and section t_j. To determine the interpolation coefficient α in Equation 3, we selected the α that gave the best result for *TextRec + StatRemRec* on expert-based concepts and used it in all of our experiments.

3. **Experiment with ACE:** To address research question RQ 2, we performed remedial recommendation by using keyphrases as domain concepts.

4. **Knowledge Augmentation:** To address research question RQ 3, this experiment investigated differences in the recommendations generated from both StatRemRec and DynRemRec.

Figure 4 provides a complete picture of the experiment set up with all of the resources that were used for the experiment. Student interaction data is divided into students stratified in ten random folds. The training folds are used for training the student model, with available concept indexing for each question. The recommendation is evaluated on the test fold. The student model is used to generate dynamic knowledge-based concept representations for DynRemRec, and the results reported in Section 5 are averaged over 10 test folds.

4.5 Evaluation Metric

As discussed in Section 4.1.1, the question to section mapping is one to many, so we adopted mean reciprocal rank (MRR) and mean average precision (MAP) to evaluate the recommendations [33]. MRR is a good metric to understand, on average, the position on which a relevant recommendation is obtained, and MAP@5 will generally prefer the algorithm that recommends more relevant sections at the top of the list. Here, the wrong recommended section is considered not to be relevant and the correct section is considered to be relevant.

5. RESULTS AND DISCUSSION

5.1 Term vs Concept-Based Recommendation

Table 5 shows the performance of the term-based approach TextRec the and domain concept-based approaches StatRemRec and DynRemRec. Both StatRemRec and DynRemRec performed lower than the baseline TextRec. One potential reason for this finding is that TextRec relies on the keywords from the whole content of both the quiz and the section, while both StatRemRec and DynRemRec only index based on on a small number of identified concepts. Based on our calculation, the average length of sections in our dataset is $1,345$ words, whereas the average number of concepts annotated by experts in each section is only 13.5, which indicates that the concept-based representation relies on a comparatively few number of concepts.

These results show that, despite the importance of domain-specific concepts in explaining the content in education, confining the representation of text with only concept-level content could cause too much loss in useful textbook content.

5.2 Fusing Term and Concept-Based Recommendations

As Meng et al. [24] pointed out, term-based content representations of education materials can provide fine-grained term level information and statistics, while concept-based representation works on both a coarse-grained topic and semantic level. Consequently, it is beneficial to have information from both these representations when recommending a relevant document to a student. As term-based representation identifies the content similarity based on shared terms, concept-level representation provides emphasis on semantics and the knowledge that is represented by the concepts. To leverage the combined power of these two representations, we conducted fusion experiments on both the term- and concept-level approaches, as mentioned in Section 4.4. As Table 5 shows, there is a clear indication that fusion (*TextRec + StatRemRec*) surpasses the baseline TextRec and benefits from concept-level representation. To investigate this effect in detail, we plotted the performance of StatRem-

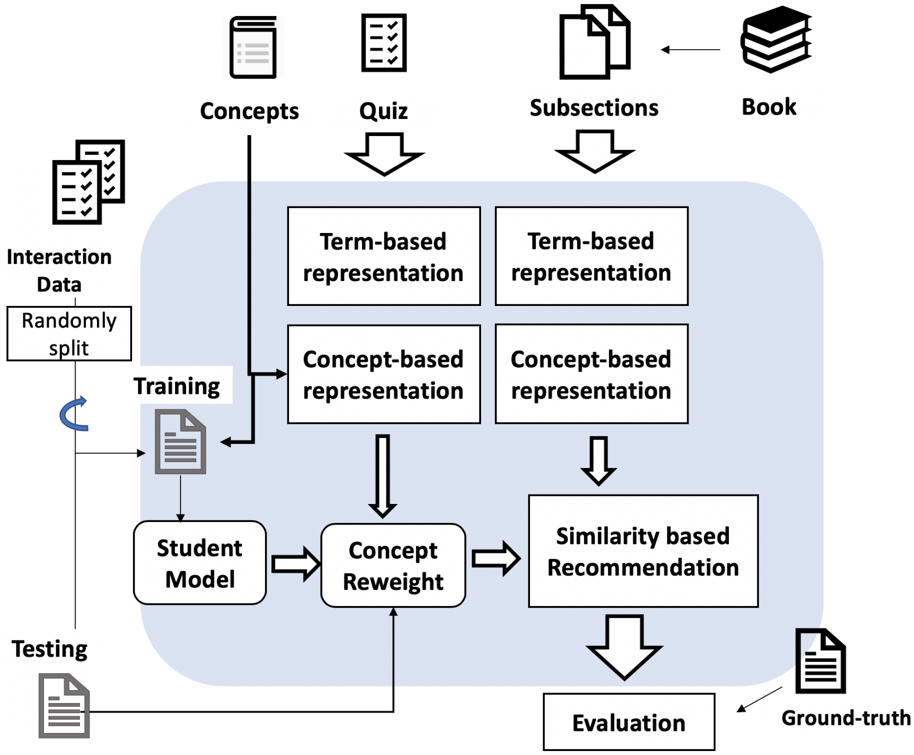


Figure 4: Experiment Setup

Table 5: Remedial recommendation performance on failed questions in ReadingCircle system in terms of MRR. * denotes a significant performance change of *metric* over text-based recommendation TextRec using a non-parametric Wilcoxon signed rank test. Numbers shown in bold indicate the top two best performance. The performance is based on expert annotated EBC. Parameter α for *TextRec + StatRemRec* and *TextRec + DynRemRec* is 0.60 based on *MAP@5*

Model	student knowledge	MRR	MAP@5
TextRec	-	83.00	74.01
StatRemRec	-	73.47	68.40
DynRemRec	✓	71.05	66.06
TextRec + StatRemRec	-	***91.01**	***86.18**
TextRec + DynRemRec	✓	***89.53**	***83.90**

Rec for different values of the interpolation co-efficient α, as shown in Figure 6.

Figure 6 and Figure 5 display some important characteristics of these recommendations. The curve of *TextRec + StatRemRec* starts with the performance of TextRec at the value of $\alpha = 0$. Initially, along with α increases from $\alpha = 0$ to $\alpha = 0.3$, the performance of the fusion-based approach *TextRec + StatRemRec* increases both in MRR and MAP@5, which shows the benefit from the inclusion of concept-based representation. Next, MRR performance stabilizes for a period from $\alpha = 0.3$ to $\alpha = 0.6$, which shows no obvious change in the position of the top-ranked relevant documents in rank lists. In this interval, MAP@5 (Figure 6) keeps increasing, which shows that concept-level representation are helping in either recommending new relevant documents or bringing already ranked relevant documents up at a higher ranked position. The performance of MAP, which looks at all the recommendations, is best at $\alpha = 0.60$. Since the performance improvement comes with the increasing weight on StatRemRec, it provides evidence that recommendation benefits more from the domain-specific concept-based representation. Fusion improved the performance of recommendation by 16% (significantly, with significance tested using Wilcoxon signed rank test), providing an answer to our research question RQ 1 that StatRemRec improves recommendation quality when augmented with a traditional content-based recommendation system.

5.3 Performance with ACE

In research question RQ 2, the goal is to understand the feasibility of using ACE as a domain concept and to use it in providing remedial recommendations. Table 6 compares the performance of some traditional ACE techniques.

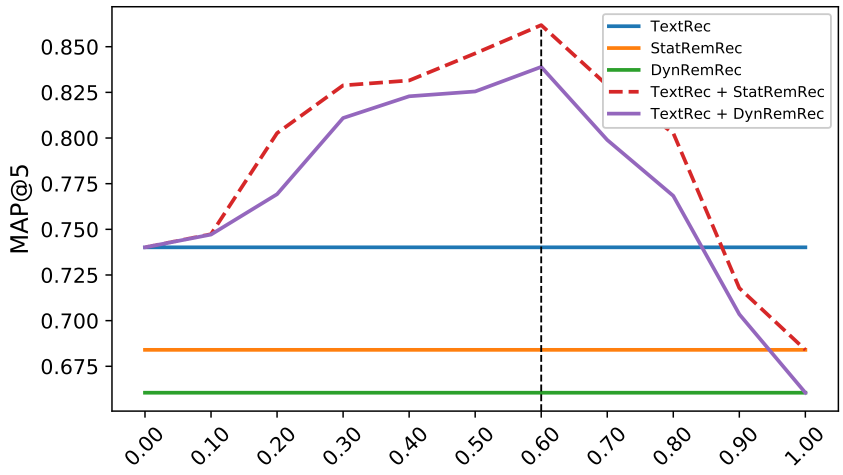


Figure 5: Threshold based recommendation values in terms of MAP with change in α values

We experimented with both supervised and unsupervised keyphrase extraction techniques. It is apparent from the performance of ACE that these techniques beat traditional content-based recommendation systems. TextRank does not improve much on content-based recommendation, but both TopicRank and CopyRNN surpass content-based recommendation. TopicRank is a winner among three methods and beats the recommendations that are based on EBC. A possible reason could be the difference between keyphrase extraction methods. TextRank and CopyRNN focus on extracting important keyphrases from a document, while TopicRank is focused on extracting keyphrases that are related to topics discussed in a document. Thus keyphrase extracted by TopicRank is more analogous to concepts discussed in the course. We experimented with comparatively few and somewhat simple keyphrase extraction techniques, as we aim to provide a piece of simple evidence for the feasibility of our approach (RQ 2). We leave for future work to perform a more comprehensive experiment with automated concept extraction, which extracts more advance domain knowledge like prerequisites and outcomes within a textbook [12, 19].

5.4 Augmenting Knowledge in Concept-Level Representation

The main goal of DynRemRec is to provide students with personalized remedial recommendations based on their real-time information needs. As Table 5 shows, DynRemRec performed worse than both TextRec and concept-based StatRemRec in MRR and MAP@5. As with StatRemRec, we fused DynRemRec with TextRec. The fusion of DynRemRec with term-based representation (TextRec + DynRemRec) revealed a similar output as TextRec + StatRemRec. This fusion improved the performance of recommendations by 13%, as compared to TextRec.

Although the fusion (TextRec + DynRemRec) improved the results in the case of DynRemRec, it is evident from Figure 6 that *TextRec + DynRemRec* was not able to improve over the performance of *TextRec + StatRemRec*. An explanation of this output is that DynRemRec addresses the need of students at each recommendation, while StatRemRec provides the same static recommendation to each student. This means that there may be cases in which experts think that a student will benefit from reading a particular section, but actual student needs might differ. Our current gold standard is expert-based, which does not target real-time student needs.

5.4.1 *Effectiveness of augmenting knowledge*

The goal of DynRemRec is to tailor the recommendations to student needs. Figure 7 shows the distribution of a unique set of recommendations generated against each question by *TextRec + DynRemRec*. As presented in the distribution in Figure 7, except for 12 questions, all the questions generated more than one distinct ranked list of recommendations. The results indicate that knowledge augmentation helps in providing an adaptive recommendation.

To understand the difference between StatRemRec and DynRemRec, we further investigated the cases where the recommendation of StatRemRec was different from DynRemRec. In our online textbooks, students read one to two chapters every week. The course instructors predefined the course sequence. We divided the course sections into three categories: *previous sections, current sections*, and *future sections*, based on the section of the question for which a student received the remedial recommendation. Previous sections can be considered as prerequisite sections, while the current section is the one for which the student is assessed. Future sections are advanced topics in which students lack complete knowledge. Figure 8 shows the distribution of recommended sections based on the three remedial recommendation techniques of TextRec, TextRec + StatRemRec, and TextRec + DynRemRec. A good remedial recommendation algorithm will recommend resources from current sections and previous sections, as understanding concepts explained in both previous and current sections will help students to solve the failed question. As Figure 8 shows, TextRec's recommendation is distributed in all the categories, while both TextRec + StatRemRec and TextRec + DynRemRec have

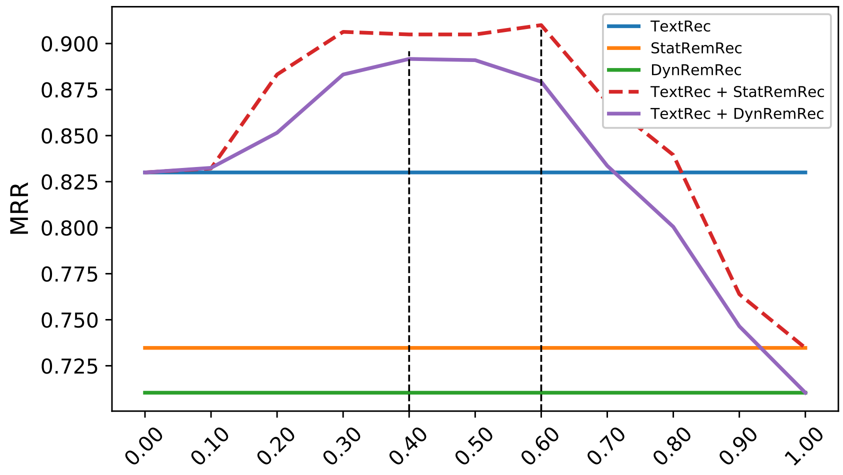


Figure 6: Threshold based recommendation metric in terms of MRR with change in α values

Table 6: Remedial recommendation performance on failed questions in an ReadingCircle system, in terms of MAP@5. The performance compares EBC with ACE-based methods for both StatRemRec and DynRemRec with an alpha value of 0.60, determined by experiment, in Figure 6. The performance is mean on 10 folds. Bold indicates the performance, which is as good as EBC.

Model	TextRec + StatRemRec	TextRec + DynRemRec
	MAP@5	MAP@5
ACE		
TextRank	83.14	83.97
CopyRNN	84.66	**84.05**
TopicRank	**89.90**	**88.85**
EBC	**86.81**	**84.11**

more recommendations in current sections and fewer recommendations in future sections. This result shows the clear benefit of the addition of domain knowledge, which helps in recommending the sections that are appropriate for the learner.

Augmenting student knowledge (TextRec + DynRemRec), on the other hand, further decreased the recommendation of future sections. However, DynRemRec also decreased the current section and provided more recommendations in previous sections than StatRemeRec. Recommendations on previous section can be the consequence of students' knowledge state. If a student is still weak in a prerequisite concept, StatRemRec will not consider those cases, while DynRemRec, which provides adaptive remedial recommendation, will make recommendations that are based on students' needs. This gives indirect evidence about the effectiveness of augmenting student knowledge in recommending resources.

6. CONCLUSIONS AND FUTURE WORK

This paper investigated the value of using domain and student knowledge for the remedial recommendation of reading resources.

We found that the use of domain knowledge significantly improves recommendation performance when fused with traditional content-based recommendations. The model *TextRec + StatRemRec*, which augments content-based recommendation with domain concept-based recommendations, significantly outperformed the traditional content-based recommender TextRec. Currently, fusion is achieved with a simple linear interpolation; we would like to investigate other fusion techniques in future studies.

While domain knowledge improves the quality of recommendation, it doesn't account for the knowledge and needs of individual students when recommending remedial reading. To address this, we tried to use dynamic student models that represent students' current knowledge state on domain concepts for providing truly personalized recommendations. *TextRec + DynRemRec*, which augments student knowledge with a content-based recommender, provided evidence to support the benefits of adding students' knowledge state for an adaptive recommendation. Although we provided some preliminary evidence for a personalized recommendation, it would be necessary to conduct a comprehensive study with real-time student feedback on recommendation. In future work, we will further investigate this phenomenon by incorporating different recommendation techniques to our online course platform. Such a study will provide a more accurate evaluation based on students' learning gain and overall system usage.

To address research questions RQ 1 and RQ 3, we used

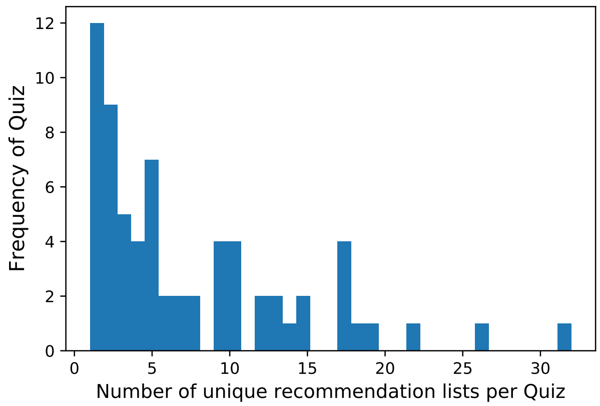


Figure 7: Distribution of unique lists of recommendations per quiz by *TextRec + DynRemRec*

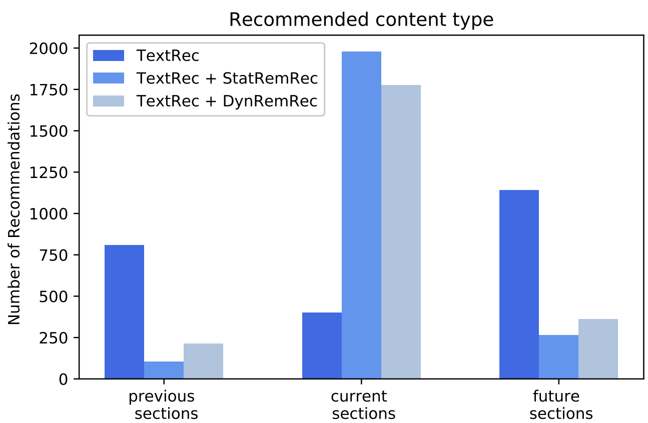


Figure 8: Category of recommended sections. The graph only plots for top recommended section. Previous, current, and future sections are categorized according to the sequence of the course.

expert-annotated domain knowledge (EBC) for building our recommender. As expert-provided concept indexing is expensive in terms of both time and resources, we further investigated traditional concept extraction approaches, such as ACE, to make our approach more feasible in practice. The performance of domain and knowledge augmented recommender on ACE proves that the technique is easy to adapt to new course content, for which expert-based concept indexing may not be available. A good future direction for this work is to investigate the importance of ACE by incorporating advanced semantic topic modeling [5] and prerequisite extraction techniques [19, 12]. A better representation of domain knowledge can lead to a more reliable knowledge unit generation for pedagogical design.

This work represents a first exploration of the power of considering students' knowledge state in recommending personalized remedial readings. The present work provides an interesting insight into automated remedial recommendation. We believe these types of models could play a more prominent role in future models of online learning where immediate or individualized instructor feedback is not available.

7. ACKNOWLEDGEMENTS

This paper was supported by the National Science Foundation Grants IIS-1525186 and Provost's Personalized Education Grants[3] by University of Pittsburgh. All data presented here is available from DataShop[4].

8. REFERENCES

[1] R. Agrawal, S. Gollapudi, A. Kannan, and K. Kenthapadi. Study navigator: An algorithmically generated aid for learning from electronic textbooks. *Journal of Educational Data Mining*, 6(1):53–75, 2014.

[2] F. Ai, Y. Chen, Y. Guo, Y. Zhao, Z. Wang, G. Fu, and G. Wang. Concept-aware deep knowledge tracing and exercise recommendation in an online learning system. In M. C. Desmarais, C. F. Lynch, A. Merceron, and R. Nkambou, editors, *Proceedings of the 12th International Conference on Educational Data Mining, EDM 2019, Montréal, Canada*. International Educational Data Mining Society (IEDMS), 2019.

[3] K. S. R. Anjaneyulu, R. A. Singer, and R. Harding. Usability studies of a remedial multimedia system. *Journal of Education Multimedia Hypermedia*, 7(2–3):207–236, June 1998.

[4] K. Bauman and A. Tuzhilin. Recommending remedial learning materials to students by filling their knowledge gaps. *MIS Quaterly*, 42(1), 2018.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[6] A. Bougouin, F. Boudin, and B. Daille. Topicrank: Graph-based topic ranking for keyphrase extraction. In *Sixth International Joint Conference on Natural Language Processing, IJCNLP 2013, Nagoya, Japan*, pages 543–551. Asian Federation of Natural Language Processing / ACL, 2013.

[7] H. Bydzovská. Course enrollment recommender system. In T. Barnes, M. Chi, and M. Feng, editors, *Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, North Carolina, USA*, pages 312–317. International Educational Data Mining Society (IEDMS), 2016.

[8] L. Carr, W. Hall, S. Bechhofer, and C. A. Goble. Conceptual linking: ontology-based open hypermedia. In V. Y. Shen, N. Saito, M. R. Lyu, and M. E. Zurko, editors, *Proceedings of the Tenth International World Wide Web Conference, WWW 10, Hong Kong, China*, pages 334–342. ACM, 2001.

[9] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modelling the acquisition of procedural knowledge. *User Modeling User-adapted Interaction*, 4(4):253–278, 1995.

[10] T. Diessel, A. Lehmann, and J. Vassileva. Individualized course generation: A marriage between cal and ical. In M. R. Kibby and J. R. Hartley,

[3]https://www.personalized.pitt.edu/content/iris-intelligent-recommender-instructors-and-students-completing-personalized-assessment

[4]https://pslcdatashop.web.cmu.edu

editors, *Computer Assisted Learning: Selected Contributions from the CAL '93 Symposium*, pages 57 – 64. Pergamon, Amsterdam, 1994.

[11] M. Dudík, S. J. Phillips, and R. E. Schapire. Performance guarantees for regularized maximum entropy density estimation. In J. Shawe-Taylor and Y. Singer, editors, *Learning Theory, 17th Annual Conference on Learning Theory, COLT 2004, Banff, Canada, Proceedings*, volume 3120 of *Lecture Notes in Computer Science*, pages 472–486. Springer, 2004.

[12] F. Gasparetti, C. D. Medio, C. Limongelli, F. Sciarrone, and M. Temperini. Prerequisites between learning objects: Automatic extraction based on a machine learning approach. *Telematics Informatics*, 35(3):595–610, 2018.

[13] J. Guerra, D. Parra, and P. Brusilovsky. Encouraging online student reading with social visualization. In E. Walker and C. Looi, editors, *Proceedings of the Workshops at the 16th International Conference on Artificial Intelligence in Education AIED 2013, Memphis, USA*, volume 1009 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013.

[14] J. Guerra, S. A. Sosnovsky, and P. Brusilovsky. When one textbook is not enough: Linking multiple textbooks using probabilistic topic models. In D. H. Leo, T. Ley, R. Klamma, and A. Harrer, editors, *Scaling up Learning for Sustained Impact - 8th European Conference, on Technology Enhanced Learning, EC-TEL 2013, Paphos, Cyprus.*, volume 8095 of *Lecture Notes in Computer Science*, pages 125–138. Springer, 2013.

[15] N. Henze, K. Naceur, W. Nejdl, and M. Wolpers. Adaptive hyperbooks for constructivist teaching. *Künstliche Intelligenz*, 13(4):26–31, 1999.

[16] Y. Huang, J. P. González-Brenes, R. Kumar, and P. Brusilovsky. A framework for multifaceted evaluation of student models. In O. C. Santos, J. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, M. C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, and M. C. Desmarais, editors, *Proceedings of the 8th International Conference on Educational Data Mining, EDM 2015, Madrid, Spain*, pages 203–210. International Educational Data Mining Society (IEDMS), 2015.

[17] K. R. Koedinger, A. T. Corbett, and C. Perfetti. The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive science*, 36(5):757–798, 2012.

[18] M. Kokkodis, A. Kannan, and K. Kenthapadi. Assigning educational videos at appropriate locations in textbooks. In J. C. Stamper, Z. A. Pardos, M. Mavrikis, and B. M. McLaren, editors, *Proceedings of the 7th International Conference on Educational Data Mining, EDM 2014, London, UK*, pages 201–204. International Educational Data Mining Society (IEDMS), 2014.

[19] I. Labutov, Y. Huang, P. Brusilovsky, and D. He. Semi-supervised techniques for mining learning outcomes and prerequisites. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada*, pages 907–915. ACM, 2017.

[20] X. Liu, Z. Jiang, and L. Gao. Scientific information understanding via open educational resources (OER). In R. Baeza-Yates, M. Lalmas, A. Moffat, and B. A. Ribeiro-Neto, editors, *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile*, pages 645–654. ACM, 2015.

[21] N. Manouselis, H. Drachsler, K. Verbert, and E. Duval. *Recommender Systems for Learning*. Springer Briefs in Electrical and Computer Engineering. Springer, 2013.

[22] J. T. Mayes, M. R. Kibby, and H. Watson. Strathtutor: The development and evaluation of a learning-by-browsing on the macintosh. *Computers and Education*, 12(1):221–229, 1988.

[23] D. McArthur, C. Stasz, J. Hotta, O. Peter, and C. Burdorf. Skill-oriented task sequencing in an intelligent tutor for basic algebra. *Instructional Science*, 17(4):281–307, 1988.

[24] R. Meng, Y. Huang, D. He, and P. Brusilovsky. Knowledge-based content linking for online textbooks. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2016, Omaha, NE, USA*, pages 18–25. IEEE Computer Society, 2016.

[25] R. Meng, S. Zhao, S. Han, D. He, P. Brusilovsky, and Y. Chi. Deep keyphrase generation. In R. Barzilay and M. Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, Volume 1: Long Papers*, pages 582–592. Association for Computational Linguistics, 2017.

[26] R. Mihalcea and P. Tarau. Textrank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing , EMNLP 2004, A meeting of SIGDAT, a Special Interest Group of the ACL, held in conjunction with ACL 2004, Barcelona, Spain*, pages 404–411. ACL, 2004.

[27] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. In Y. Bengio and Y. LeCun, editors, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA*, 2013.

[28] D. N. Milne and I. H. Witten. Learning to link with wikipedia. In J. G. Shanahan, S. Amer-Yahia, I. Manolescu, Y. Zhang, D. A. Evans, A. Kolcz, K. Choi, and A. Chowdhury, editors, *Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, USA*, pages 509–518. ACM, 2008.

[29] M. Mohseni, M. L. Maher, K. Grace, N. Najjar, F. Abbas, and O. Eltayeby. Pique: Recommending a personalized sequence of research papers to engage student curiosity. In S. Isotani, E. Millán, A. Ogan, P. M. Hastings, B. M. McLaren, and R. Luckin, editors, *Artificial Intelligence in Education - 20th International Conference, AIED 2019, Chicago, IL, USA, Proceedings, Part II*, volume 11626 of *Lecture Notes in Computer Science*, pages 201–205. Springer, 2019.

[30] R. Morsomme and S. V. Alferez. Content-based course recommender system for liberal arts education. In

M. C. Desmarais, C. F. Lynch, A. Merceron, and R. Nkambou, editors, *Proceedings of the 12th International Conference on Educational Data Mining, EDM 2019, Montréal, Canada, July 2-5, 2019*. International Educational Data Mining Society (IEDMS), 2019.

[31] R. Nagata, K. Takeda, K. Suda, J. Kakegawa, and K. Morihiro. Edu-mining for book recommendation for pupils. In T. Barnes, M. C. Desmarais, C. Romero, and S. Ventura, editors, *Educational Data Mining - EDM 2009, Cordoba, Spain. Proceedings of the 2nd International Conference on Educational Data Mining*, pages 91–100, 2009.

[32] Z. A. Pardos and N. T. Heffernan. Modeling individualization in a bayesian networks implementation of knowledge tracing. In P. D. Bra, A. Kobsa, and D. N. Chin, editors, *User Modeling, Adaptation, and Personalization, 18th International Conference, UMAP 2010, Big Island, HI, USA*, volume 6075 of *Lecture Notes in Computer Science*, pages 255–266. Springer, 2010.

[33] D. Parra and S. Sahebi. Recommender systems: Sources of knowledge and evaluation metrics. In *Advanced Techniques in Web Intelligence-2: Web User Browsing Behaviour and Preference Analysis*, pages 149–175, Berlin, Heidelberg, 2013. Springer.

[34] P. I. Pavlik, H. Cen, and K. R. Koedinger. Performance factors analysis - A new alternative to knowledge tracing. In V. Dimitrova, R. Mizoguchi, B. du Boulay, and A. C. Graesser, editors, *Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling, Proceedings of the 14th International Conference on Artificial Intelligence in Education, AIED, Brighton, UK*, volume 200 of *Frontiers in Artificial Intelligence and Applications*, pages 531–538. IOS Press, 2009.

[35] M. C. Polson and J. J. Richardson. *Foundations of intelligent tutoring systems*. Lawrence Erlbaum, 1988.

[36] B. Pursel, C. Liang, S. Wang, Z. Wu, K. Williams, B. Bräutigam, S. Saul, H. Williams, K. Bowen, and C. L. Giles. Bbookx: Design of an automated web-based recommender system for the creation of open learning content. In J. Bourdeau, J. Hendler, R. Nkambou, I. Horrocks, and B. Y. Zhao, editors, *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, Companion Volume*, pages 929–933. ACM, 2016.

[37] J. Rihák and R. Pelánek. Measuring similarity of educational items using data on learners' performance. In X. Hu, T. Barnes, A. Hershkovitz, and L. Paquette, editors, *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017, Wuhan, Hubei, China*. International Educational Data Mining Society (IEDMS), 2017.

[38] S. Ritter, J. R. Anderson, K. R. Koedinger, and A. Corbett. Cognitive tutor: Applied research in mathematics education. *Psychonomic bulletin & review*, 14(2):249–255, 2007.

[39] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523, 1988.

[40] K. Thaker, Y. Huang, P. Brusilovsky, and H. Daqing. Dynamic knowledge modeling with heterogeneous activities for adaptive textbooks. In K. E. Boyer and M. Yudelson, editors, *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, Buffalo, NY, USA*. International Educational Data Mining Society (IEDMS), 2018.

[41] D. Tudhope and C. Taylor. Navigation via similarity: Automatic linking based on semantic closeness. *Information Processing & Management*, 33(2):233–242, 1997.

[42] M. Wang, H. Chau, K. Thaker, P. Brusilovsky, and D. He. Concept annotation for intelligent textbooks. *CoRR*, abs/2005.11422, 2020.

[43] G. Weber and P. Brusilovsky. ELM-ART: An adaptive versatile system for web-based instruction. *International Journal of Artificial Intelligence in Education*, 12(4):351–384, 2001.