A Bi-level Individualized Adaptive Learning Recommendation System Based on Topic Modeling

Jiawei Xiong, Jordan M. Wheeler, Hye-Jeong Choi, and Allan S. Cohen

Abstract Adaptive learning offers real attention to individual students' differences and fits different needs from students. This study proposes a bi-level recommendation system with topic models, gradient descent, and a content-based filtering algorithm. In the first level, the learning materials were analyzed by a topic model, and topic proportions to each short item in each learning material were yielded as representation features. The second level contains a measurement component and a recommendation strategy component which employ gradient descent and content-based filtering algorithm to analyze personal profile vectors and make an individualized recommendation. An empirical data consists of cumulative assessments that were used as a demonstration of the recommendation process. Results have suggested that the distribution to the estimated values in the person profile vectors were related to the ability estimation from the Rasch model, and students with similar profile vectors could be recommended with the same learning material.

Keywords Individualized learning · Recommendation system · Topic model

1 Introduction

In recent years, especially during the pandemic, efforts have been made to expand online learning beyond the traditional classroom environment as it enables individuals to benefit from rich and high-quality learning resources (Dhawan, 2020; Liang & Hainan, 2019). The advantages of online learning have been recognized since it offers real attention to the individual differences and fits for different

J. Xiong (✉) · J. M. Wheeler · A. S. Cohen
University of Georgia, Athens, GA, USA
e-mail: jiawei.xiong@uga.edu

H.-J. Choi
The University of Georgia, Athens, GA, USA

The Human Resources Research Organization, Louisville, KY, USA

M. Wiberg et al. (eds.), *Quantitative Psychology*, Springer Proceedings
in Mathematics & Statistics 393, https://doi.org/10.1007/978-3-031-04572-1_10

needs of students (Imhof et al., 2020). More importantly, it makes it possible to analyze students' latent information, here referred to as profile information or profile, through human-computer interactions, and in particular, with the maturity of cutting-edge learning analytics, individualized adaptive learning provides students the prospects of access to tailored learning instructions, guidance, and content (Mavroudi et al., 2018). With the popularization of remote education and online learning, offering individualized and adaptive learning resources is an emerging research topic (Cheng & Bu, 2020). Individualized adaptive learning systems aim to provide learning materials fit to the current status of a student, and the pace and of learning and instruction approach are optimized for the need of each student (United States Department of Education, 2017). The individualized adaptive learning system provides flexible adaptation beyond what can be accomplished in traditional classroom settings in terms of learning resources (Koedinger et al., 2013).

The purpose of adaptive learning is realized by using a recommendation system, which may recommend the next learning materials based on the psychometric results and possibly other individual-level characteristics (Chen et al., 2018). Specifically, the recommendation system requires three components, an information learning component, a measurement component, and a recommendation strategy component. The information learning component employs a learning model to analyze features from the learning materials such that each learning material's features can be represented in a numerical space. The features can be used as representations of a series of skills or attributes that are available in the learning system (Chen et al., 2018). Traditional recommendation systems suggest online learning materials based on students' interests, knowledge, and data from other students with similar interests (Romero et al., 2007). These traditional methods, which utilize vector space models (Castells et al., 2006) in the information learning component, have disadvantages in both effectiveness and scalability (Kuang et al., 2011). In addition, with a heterogeneous student population and many learning materials, the learning model can be complex, and thus calibrating the model requires expensive computation (Tang et al., 2019). Topic modeling such as the Latent Dirichlet allocation (LDA; Blei et al., 2003), a hierarchical Bayesian topic model, was used to obtain a low dimensional vector that denotes each online learning activity in various adaptive learning scenarios, such as online course recommendations (Lin et al., 2021) and online documents recommendation (Kuang et al., 2011). Compared with traditional recommendation systems based on the student-item interactions and similarity, the topic model-based recommendation systems can consider the learning portfolios' textual features (Cao et al., 2019). With the learned features, the measurement component can find the profile vectors for students which may reveal students' proficiency on each attribute (Chen et al., 2018). Given the features and profiles, the prediction component uses a predicting model that predicts the outcomes of a student studying under a particular set of materials, and sequentially makes recommendations to each student on what to learn at the next step, based on the current information it obtained from the aforementioned two components. Content-based filtering (CB; Ghauth & Abdullah, 2010) algorithm, using information about students and/or learning materials, has been used as a prediction component in many recommendation systems (Bian & Xie, 2010; Romero et al., 2007). The CB focuses on the properties of learning materials, and learning materials' similarity is determined by the similarity in their features.

The cumulative assessment portfolio has been used as learning material in online learning (Pfennig, 2020). Cumulative assessments are widely used in many circumstances to determine at a particular time what students know and do not know and can help students get access to their learning achievements (Beagley & Capaldi, 2016; Ryan & Nykamp, 2000). The cumulative assessments are comprehensive and pre-assembled tests that assess students' knowledge of information from several didactic domains, and in which each assessment covers all previous contents. By using the cumulative assessments, instructors can identify a wide range of knowledge, skills, and concepts that students have mastered or not, so appropriate adjustments can be made to instructional practices and strategies toward the overall end-of-year expectations (den Boer et al., 2021). For example, some English language and arts (Georgia Center for Assessment, 2018) cumulative assessments were designed to collect evidence on student learning status, and serve as formative tools that can provide information on how well students understand concepts and their ability to demonstrate knowledge and skills in a particular content area or domain. In higher education, it is effective to intersperse several cumulative assessments throughout a course and the combined score on the assessments weighs in for the final course grade (den Boer et al., 2021). For example, the United States Medical Licensing Exam Step I assesses whether the examinees can successfully apply the knowledge of key concepts in basic sciences and is usually taken by medical school examinees at the end of the second year (USMLE, 2014). Some medical schools ask students to take the cumulative licensing examination before initiating clinical experiences (Cleghorn, 1986; Ryan & Nykamp, 2000). Given the fact that cumulative assessments have wide applications, this study selects a set of pre-assembled cumulative assessments as learning materials.

It is suggested that a good recommendation system should make full use of the information from both the students and the learning materials (Tang et al., 2019). Therefore, this study designs a bi-level structure. In the first level, the learning materials (i.e., cumulative assessments) were analyzed by a topic model and the topic proportions to each item stem in the cumulative assessment were yielded as representation features to the cumulative assessment. Although most educational applications with topic models adopt the LDA as a useful model (Wheeler et al., 2021; Xiong et al., 2019), the use of LDA as a topic model tool is useful for long documents such as the course syllabus (Apaza et al., 2014), and it suffers from the severe data sparsity in short text documents (Yan et al., 2013). For instance, the pre-assembled cumulative assessments may contain some short text items such as multiple-choice (MC) items, and which lengths are usually less than a passage or course syllabus content. Obviously, in such a circumstance, the use of LDA may cause sparse topic structures. To overcome the problem, in this first level, this study employs another topic model, called the bi-term topic model (BTM; Yan et al., 2013), which was designed to extract topic proportions for short context, to analyze the learning materials (i.e., each cumulative assessment) and obtain each short item's topic structure. The second level contains the measurement and recommendation strategy components which employ profile analysis and CB filtering algorithms. By proposing such a framework that applies both the BTM and CB filtering to recommend pre-assembled cumulative assessments with an empirical data demonstration, this recommendation system can analyze each student's profile components based on their response scores to the completed assessments and then predict rating scores for new assessments. The empirical results suggested this design can recommend relevant assessments for each student, and realize individualized recommendations based on the bi-level framework.

2 Method

2.1 Bi-term Topic Model

BTM generates the bi-terms in the whole corpus to reveal topics by considering the word-pair relation. The bi-term here was referred to as an unordered word-pair co-occurred in a short context such as the example given in Table 1. The texts given in the Table 1 are all simple examples of short texts. After removing stopwords such as "I", and stemming words into an original form such as changing from "apples" to "apple", from "eating" to "eat", the bi-terms were generated by construction word-pair combination in an unordered way.

The BTM graphical structure is represented in Fig. 1, and this generative process in the BTM can be described as:

1. Draw a topic distribution θ from Dirichlet distribution with parameter α, i.e., $\theta \sim Dir(\alpha)$.
2. For each topic $k \in [1, \cdots, K]$, draw a topic-specific word distribution ϕ_k from the Dirichlet distribution with parameter β, i.e., $\phi_k \sim Dir(\beta)$.
3. For each bi-term combination $b_i \in_B$:

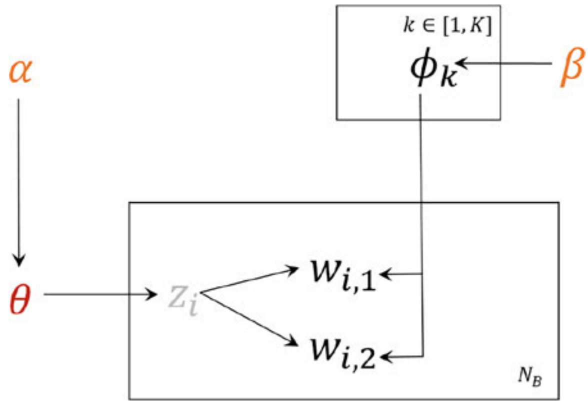
   (a) draw a topic assignment $z_i \sim Multinomial(\theta)$
   (b) draw two words, $w_{i,1}, w_{i,2} \sim Multinomial(\phi_{z_i})$

where N_B is the bi-term corpus which consists of all bi-terms given in a document collection, α is the prior distribution parameter to the topic distribution θ, β is the

Table 1 Simple bi-term examples

Text	Bi-terms
I visit an Apple store.	visit Apple, visit store, Apple store
I like eating apples.	like eat, like apple, eating apples
I love to watch Apple movies.	love watch, love Apple, love movie, watch Apple, watch movie, Apple movie

Fig. 1 BTM graphical structure



prior distribution parameter to the bi-term distribution ϕ_k. BTM directly models the word-co-occurrence pattern instead of a single word.

In this study, each pre-assembled cumulative assessment is treated as learning material, and adaptive learning happens after the completion of each assessment. Each MC item stem was modeled as a short text. Suppose the learning system consists of $i = 1, \ldots, I$ cumulative assessments, while each of which consists of n_i MC items, then a total of $I \times n_i$ items are treated as a large collection of short documents. Suppose k topics were determined for the learning system, and then each item can be represented by a k-dimensional feature vector. Therefore, each cumulative assessment is represented by a $n_i \times k$ dimensional matrix. To determine the optimal number of topics k for the corpus, average Jensen-Shannon (JSD; Tong & Zhang, 2016) was used as criteria. JSD is a popular method of measuring the similarity between two probability distributions and is also known as a total divergence to the average. Given two discrete topic distributions T_s and T_v, the JSD is defined as Eq. 1.

$$JSD(T_s \| T_v) = \frac{1}{2} KLD\left(T_s \| \frac{T_s + T_v}{2}\right) + \frac{1}{2} KLD\left(T_v \| \frac{T_s + T_v}{2}\right) \quad (1)$$

where $KLD\left(T_s \| \frac{T_s+T_v}{2}\right)$ is the Kullback-Leibler divergence (KLD; Mei et al., 2007) of T_s from $\frac{T_s+T_v}{2}$, and the KLD is defined as in the Eq. 2.

$$KLD(P \| Q) = \sum_s P_s \log \frac{P_s}{Q_v} \quad (2)$$

The average JSD shown in Eq. 3 is used to calculate the average similarity among all topic distributions.

$$\overline{JSD} = \frac{\sum_{s,v} JSD(T_s \| T_v)}{k} \tag{3}$$

By applying JSD to the topic assignment for each item in the learning system, it will measure the distance and similarity between each document. The topic model with minimal average JSD is used as the optimal topic model. With these k topics, each MC item n's features can be represented by a k-dimensional topic proportion vector $\boldsymbol{f_n} = (f_{n_1}, \cdots, f_{n_k})$.

2.2 Loss Function and Gradient Descent

Suppose a random assessment i is given at the initial stage to a total of $j = 1, \ldots J$ students, while the remaining $(I - 1)$ assessments were waiting in the system to be sequentially recommended to students. With the k-dimensional feature vector $\boldsymbol{f_{in}} = (f_{in_1}, \cdots, f_{in_k})$ for each MC item n in the assessment i, student j's profile vector can also be defined as a k-dimensional vector $\boldsymbol{\alpha_j} = (\alpha_{j_1}, \cdots, \alpha_{j_k})$. Each dimension to the profile vector serves as the weight or coefficient to items' feature vector. In addition, for each MC item in the assessment i, student j's response can be scored as either correct or incorrect (i.e., $t_{jn} = 1/0$), so that cross-entropy (De Boer et al., 2005) is used as a loss function with t_{jn} serving as guiding labels, which is defined in the Eq. 4

$$L(\boldsymbol{\alpha}) = -\frac{1}{Jn_i}\left[\sum_{j=1}^{J}\left[\sum_{n=1}^{n_i}\left[t_{jn}\log(p_{jin}) + (1 - t_{jn})\log(1 - p_{jin})\right]\right]\right] \tag{4}$$

where $p_{jin} = \sigma(\boldsymbol{\alpha_i f_{in}}) = \sigma\left(\alpha_{j1} f_{in_1} + \cdots + \alpha_{jk} f_{in_k}\right)$, and the $\sigma(\cdot)$ represents a sigmoid function.

The gradient descent (Amos & Yarats, 2020) is used to minimize the loss function until it reaches convergence. The process of finding the minimized loss function is described in Table 2, where ρ is a positively defined learning rate, and $\nabla J(\boldsymbol{\alpha_r})$ is the differential at $\boldsymbol{\alpha_r}$. The $\rho\nabla J(\boldsymbol{\alpha_r})$ is subtracted from α_r and moves toward the local minimum. So, a monotonic sequence $J(\alpha_r) \geq J(\alpha_{r+1}) \geq J(\alpha_{r+2}) \geq \cdots$ is obtained until convergence.

Table 2 Pseudo-code gradient descent algorithm

Algorithm	Gradient descent		
	For	$r = 1, 2, \cdots$	
		repeat	$\boldsymbol{\alpha_{r+1}} = \boldsymbol{r} - \rho\nabla J(\boldsymbol{\alpha_r})$
		until	convergence
	Output	$J(\boldsymbol{\alpha_{r+1}})\&\boldsymbol{\alpha}_{r+1}$	

3 Data and Analytic Framework

3.1 Data Description

Learning materials are used to construct a learning pool in which every learning material is pending to be recommended or not for each student. The learning pool in this study contains 8 science cumulative assessments (Georgia Center for Assessment, 2018) as learning materials, and each assessment contains 22 MC items. The assessments are designed to assess student learning on several sub-domains in science such as biology, physics, and chemistry. A link to the sample assessment was provided in the Appendix. Each MC item stem was pre-processed. This process includes stemming and lemmatization, and stop word removal. The stemming uses the stem of each word and cuts off the end or the beginning of the word such as the affixes of plural words. The lemmatization uses the context in which the word is being used and changes the word into the base forms such as the irregular verbs and irregular plural nouns. Stop words are high-frequency terms with little or no information and include words such as "the", "and", "is" etc. The cleaned item stems were treated as short texts and were modeled in BTM to extract representation vectors. The descriptive statistics to the length of clean item stems are listed in Table 3. The minimal length to the MC items is only 3, and the average length of these items is 10.530. Therefore, the lengths of items are relatively short and the use of BTM is appropriate.

Students' response to each MC item was scored as either correct (1) or incorrect (0). Students' responses t_{jn} to one learning material (i.e., one assessment containing 22 items) were used as guiding labels which are defined in Eq. 4. In this study, Assessment 4 was selected, and 492 students have responded to the 22 MC items in the assessment. All the response correctness t_{jn} given by these 492 students were used as the guiding labels to supervise the parameter estimation.

3.2 Bi-level Recommendation Framework

The bi-level recommendation system is shown in Fig. 2. The feature learning component in the first level employs the BTM described in Fig. 1 to extract feature matrix (dimension $n_i \times k$) for each assessment. The measurement component in the second level uses the J students' responses to each item in one selected assessment and employs the gradient descent algorithm described in Table 2 to minimize the loss function to obtain J's k-dimensional vectors as students' profile

Table 3 Descriptive statistics to MC items' length in the learning pool

Min.	Mean	Max.	SD
3.000	10.530	32.001	10.812

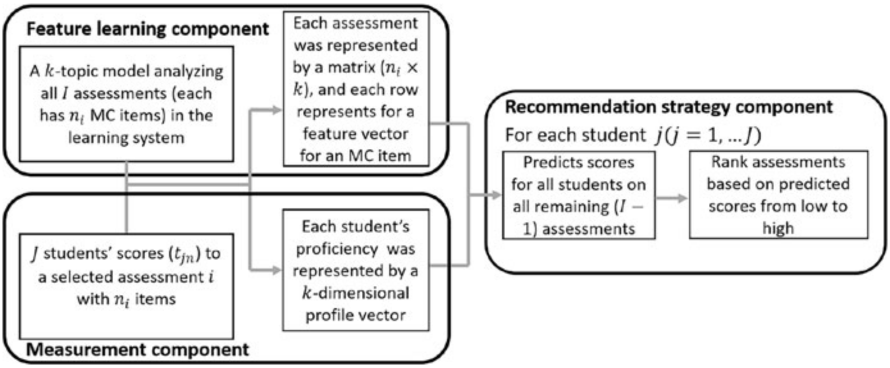


Fig. 2 Bi-level recommendation framework

vectors. The estimated profile vector was used as a quantified indicator of each student's understanding of each of the k topics at this moment. Furthermore, with the k-dimensional student j's profile vector and remaining $I - 1$ assessments' feature matrices, the recommendation strategy component predicts student j's score probability on each MC item in each remaining assessment with Eq. 5.

$$f(z_{jin}) = \frac{1}{1 + e^{-Z_{jin}}} \tag{5}$$

where $z_{jin} = \boldsymbol{\alpha}_{\boldsymbol{j}} \boldsymbol{f}_{\boldsymbol{in}}$, and student j will be predicted to get a score $g_{jin} = 1$ for item n in assessment i when $f(z_{jin}) \geq 0.5$, and score $g_{jin} = 0$ when $f(z_{jin}) < 0.5$.

3.3 Analytic Procedures

The first step of this recommendation system uses the feature learning component to construct a representation matrix for each learning material. That is, by modeling every learning material i in the feature learning component, a k-column feature matrix was extracted by the BTM with JSD. In this study, each extracted feature matrix contains 22 rows and k columns, and each feature matrix serves as a representation matrix of the corresponding learning material. Once the representation matrices for all learning materials were constructed, then one learning material was randomly selected for all students and the remaining $I - 1$ learning materials are still in the learning pool. The second step of this recommendation system uses all students' complete response patterns to estimate jth student's k-dimensional profile vector $\boldsymbol{\alpha}_{\boldsymbol{j}} == (\alpha_{j1}, \cdots, \alpha_{jk})$ in the measurement component with the loss function in Eq. 4 and gradient descent algorithm in Table 2. Finally, with the k-dimensional profile vectors for each student and remaining $I - 1$ assessments' feature matrices, the recommendation strategy component predicts jth student's correctness probability on each MC item in each remaining assessment with Eq. 5 and predicts a score g_{jin} for each item n of jth student. The summation $g_{ji} = \sum_n g_{jin}$ will be used as jth student's predicted total score on the assessment i in the learning system. After ranking the predicted total scores for all remaining $I - 1$ assessments from low to high, the system can make the next recommendation for each student.

4 Results

4.1 Step 1: Feature Learning Component with BTM and JSD

In the feature learning component, JSD was used as criteria for some exploratory topic models from 2 topics to 6 topics for all items in the learning system. Figure 3a is the average JSD values against the different number of topics, and the minimal JSD was achieved when the number of topics is four. So, it is suggested that the four-topic model fits best for all items in the learning system. After fitting a four-topic BTM, every item stem was characterized by a 4-dimensional vector and each dimension represents a topic's proportion in the item stem. For example, the item n in assessment i can be characterized by a vector of $\boldsymbol{f_{in}} = (0.1, 0.3, 0.1, 0.5)$ in which each value describes the topic distribution to this item such that 10% of words in the item belong to Topic 1 and 30% of words belong to Topic 2, etc. Table 4 lists the top 10 words under each of the four topics. Topic 1 can be described as words related to the chemistry process and natural resources, Topic 2 tends to employ words about the ecosystem, Topic 3 contains question words such as "select the class from the following samples", and Topic 4 shows words from astronomy and physics.

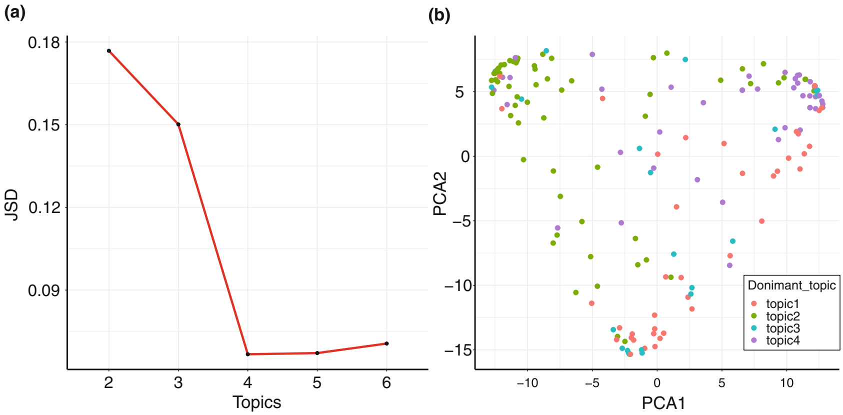


Fig. 3 (**a**) Average JSD against the number of topics; (**b**) PCA to the feature matrix

Table 4 Top 10 words from each of the four topics in the cumulative assessments

Topic 1		Topic 2		Topic 3		Topic 4	
water	0.034	model	0.019	class	0.021	earth	0.017
create	0.014	population	0.018	light	0.015	moon	0.015
rock	0.014	organism	0.016	give	0.015	sun	0.015
temperature	0.014	cell	0.016	form	0.012	weather	0.013
heat	0.013	base	0.015	see	0.011	layer	0.012
soil	0.011	show	0.014	sample	0.011	feather	0.011
hot	0.010	system	0.012	student	0.011	white	0.011
air	0.009	animal	0.012	leave	0.010	eye	0.010
plant	0.008	picture	0.010	need	0.009	gibbous	0.010
different	0.007	energy	0.009	question	0.009	move	0.010

Items may have different topic structures from each other; therefore, each item focuses on a domain combination in the cumulative assessment. If we denote the topic with the highest proportion as the dominant topic for an item, for instance, for the item with $\boldsymbol{f_{in}} = (0.1, 0.3, 0.1, 0.5)$, its dominant topic was denoted as Topic 4. Figure 3b shows the principal component analysis (PCA; Chou & Wang, 2010) to the feature matrix with identity information from dominant topics. The two principles, PCA1 and PCA2, are associated with test domains. PCA1 represents environment-associated contents, and PCA2 stands for biology-associated contents. Each point represents an item in the two-dimensional space and each color represents a dominant topic. In this figure, items with similar topic distributions could be closer to each other, which indicates the items with similar topic distributions may measure similar test domains. We also noticed that some items with different dominant topics are mixed, which is because these items' topic distributions are flat such that the dominant topic has a close proportion to other topics.

The analysis in the feature learning component yielded 8 feature matrices and each has a dimension of 22×4. In each feature matrix, every row represents an item feature vector f_{in} for nth item in ith assessment. Each dimension in the vector f_{in} represents for the topic proportion of nth item. Therefore, jth student's unknown profile vector is also 4-dimensional such that $\boldsymbol{\alpha_j} = (\alpha_{j_1}, \alpha_{j_2}, \alpha_{j_3}, \alpha_{j_4})$.

4.2 Step 2: Measurement Component with Gradient Descent

After constructing the feature matrices for all learning materials, Assessment 4 was randomly selected for all 492 students. With the obtained feature matrix for Assessment 4, the response patterns were used to estimate jth student's profile vector $\boldsymbol{\alpha_j} = (\alpha_{j_1}, \alpha_{j_2}, \alpha_{j_3}, \alpha_{j_4})$ in the measurement component. Students' profile vectors are obtained by gradient descent on the loss function. The gradient descent

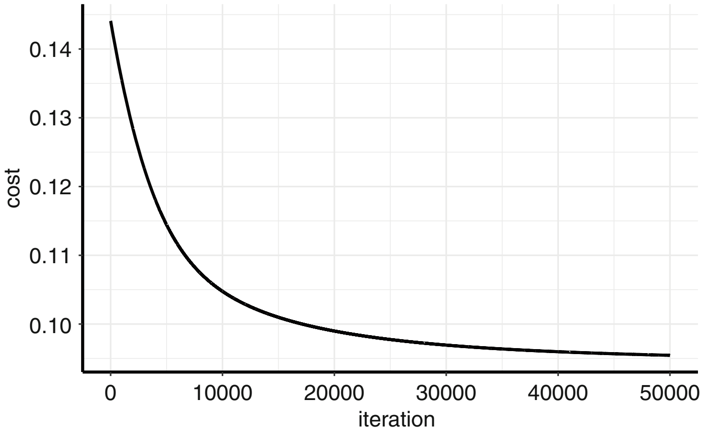


Fig. 4 Cost function against iterations

Table 5 Summary statistics on the estimated profile distribution

	Min.	Mean	Max.	SD
α_{j_1}	−1.045	−0.188	0.743	0.399
α_{j_2}	−1.039	0.201	1.498	0.364
α_{j_3}	−1.001	−0.017	0.859	0.339
α_{j_4}	−1.047	−0.084	0.686	0.364

changes are reflected in Fig. 4, which shows the cost decrease against each iteration. From Fig. 4, the estimation converged with increasing iterations finally and the cost was less than 0.1. The obtained estimations are student profile vectors that are coefficients to each dimension in the feature matrix.

The estimated profile vectors are summarized in Table 3. In Table 3, each row lists descriptive statistics of one dimension. As introduced, the estimated coefficient α_{j_k} can be interpreted as jth student's understanding of kth dimensional feature. For example, a student with the minimal value of $\alpha_{j_1} = -1.054$ may indicate that the student owns a relatively low understanding status of −1.054 on Topic 1, while a student with the maximal value of $\alpha_{j_1} = 0.743$ means that the student owns a relatively high understanding status to this topic. The standard deviations to the four dimensions were from 0.339 to 0.399. The α_{j_2} has the largest range from −1.039 to 1.498, while the other three dimensions are distributed between −1 and 1 (Table 5).

An item response analysis was conducted to help interpret profile vectors. Students' correctness responses were analyzed by a Rasch model, which can calibrate students' ability levels into logit scale and rank the logits on a one-dimensional continuum (Engelhard, 2013), to explore the relationship between students' profile vector and students' latent ability. By assuming there is a unidimensional ability of students for answering these items correctly, the density of calibrated Rasch ability is plotted in Fig. 5a, which is approximately normally distributed with a mean of 0.000. The minimal student ability value is −1.942 and the maximal ability value is

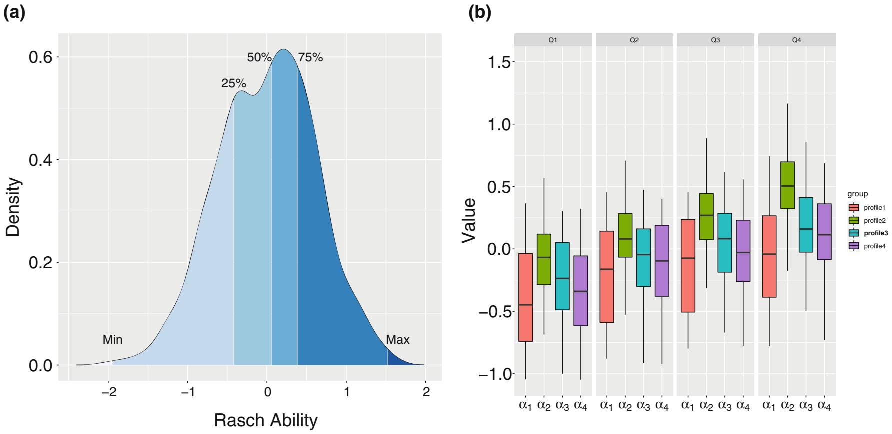


Fig. 5 (**a**) Rasch ability density with 4 quartiles; (**b**) Distribution of α in each quartile

1.521. The first quartile (the lowest 25%), second quartile (between 25.1 and 50%), and third quartile (50.1–75%) to the estimated Rasch ability are −0.413, 0.056, and 0.380, respectively. By labeling students into four categories according to their Rasch ability such that Q1: (−1.942, −0.413), Q2: (−0.413, 0.056), Q3: (0.056, 0.380), Q4: (0.380, 1.521), the order from Q1 to Q4 also represents both the ability levels and the probability of answering items correctly are increasing.

Figure 5b shows the distribution of α's based on every quartile category. It is clear to observe that each dimensional α value shows a trend of increasing from lower category to higher category and the α_2 exhibits the most obvious increase trend. Possibly Topic 2, containing words about the ecosystem, covers the main test sub-domain in the cumulative learning material. From Q1 to Q4, the Rasch estimates are increasing, and the probability of answering items correctly is increasing. Since each student's profile vector indicates the student's understanding status of certain topics, larger α will lead to a higher probability of taking each item correctly. This further verifies the homogeneity of Rasch analysis and profile analysis in terms of item correctness probability, and students with higher ability values are also likely to have higher profile values on each dimension.

4.3 Step 3: Recommendation Strategy Component with Predicted Total Scores

Given students' current understanding status of certain topics, their predicted total scores for all remaining assessments in the learning pool were calculated based on the profile vectors and learning matrices in the recommendation strategy component. Every two students tend to have different profile vectors unless they have the same responding pattern to the 22 MC items. Therefore, the predicted patterns on each of the remaining learning materials for every two students could be different.

All 492 profile vectors were multiplied to each of the learning matrices of the remaining assessments for predicting jth student's nth item score g_{jin} of learning material i using Eq. 5. For jth student, the summation of all item scores within ith assessment $g_{ji} = \sum_n g_{jin}$ is used as the predicted total score on the assessment i in the learning system. These scores indicate a student's predicted achievements on the remaining learning materials that the student may obtain based on their current α values. Learning materials with lower scores indicate that the student may perform comparatively worse on that materials than the ones with higher scores. For each student, the predicted scores on remaining learning materials were ranked from low to high, and the learning material with the lowest score was recommended to the student for next-step learning. Figure 6 shows students' predicted scores distribution for each of the remaining learning materials, where the vertical axis stands for the predicted scores. In this figure, the predicted scores for each assessment range from 0 to 22. The predicted scores of Assessment 1 have a relatively lower 1st quantile value, which indicates that more students were predicted to have a lower score on Assessment 1. The predicted scores of Assessment 7 have a higher 3rd quantile value than other assessments, which means that more students were predicted to have a higher score on Assessment 7.

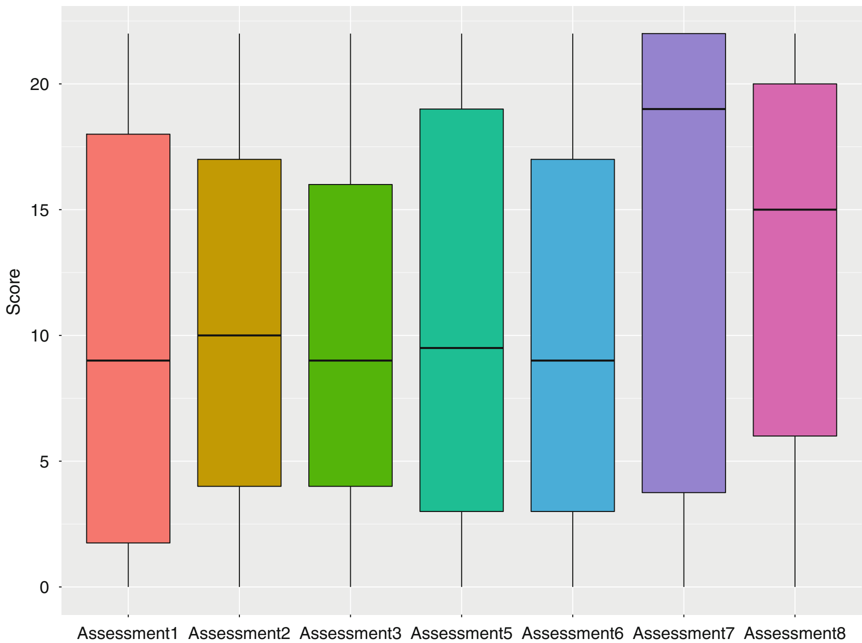


Fig. 6 Students' predicted scores distribution for each of the remaining learning materials

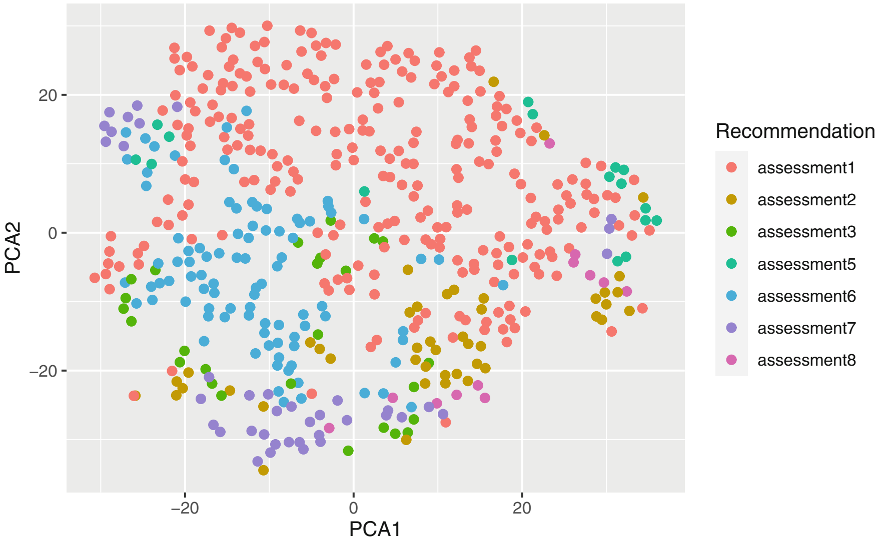


Fig. 7 PCA with recommended learning materials

Table 6 Correlations between each dimension of the profile vectors and each principal component

	PCA1	PCA2
α_{j_1}	**0.524**	0.229
α_{j_2}	**0.711**	**0.575**
α_{j_3}	0.190	0.207
α_{j_4}	0.250	**0.544**

Figure 7 shows the PCA on students' profile vectors with colors indicating their recommendation results on the 7 accumulative assessments. In this figure, each point represents a student, and students with similar profile vectors were close to each other on this lower-dimensional space. It can be seen that many students were recommended to take Assessment 1, and which is consistent with the observation from Fig. 6 that more students were predicted to have a lower score on Assessment 1. In addition, students with similar profile vectors could be recommended with the same learning material.

Table 6 listed the correlations between each dimension of the profile vectors and each principal component, and correlations higher than 0.500 were in bold. PCA1 is strongly correlated with two dimensions of the profile vectors and PCA1 increases with increasing α_{j_1} and α_{j_2}. This component can be viewed as a measure of the values to α_{j_1} and α_{j_2}. Furthermore, we see that the first principal component correlates most strongly with α_{j_2}. Considering the interpretation of α_{j_1} and α_{j_2}, PCA1 may indicate students' understanding status of concepts with the ecosystem and natural resources. PCA2 also correlates with two dimensions of the profile vectors, α_{j_2} and α_{j_4}. Similarly, this component can be viewed as a measure of the understanding of Topic 2 and Topic 4, so PCA2 primarily stands for students' understanding status of concepts with astronomy and physics.

5 Discussion and Conclusion

This study proposed a bi-level recommendation system consisting of three components. An empirical study shows that, by employing the topic model and gradient descent algorithm, students profile vectors can be extracted and individualized recommendations can be made based on the predicted scores for the learning materials. The analysis also suggested that the distribution to the estimated values in the person profile vectors were related to the ability estimation from the Rasch model. Future researches can focus on a simulation study that explores the recovery accuracy of the profile vectors.

Although the recommendation component shows interesting findings and predicts individual scores on each item in the learning material, one thing still needs to note is that, for each student, learning materials with different score patterns may be predicted with the same final scores. This is because the sum scores for each learning material were used. For example, suppose a four-item cumulative assessment of predicted score patterns of $s = (0, 1, 0, 1)$ has the same sum score as another assessment with pattern $s = (1, 0, 1, 0)$, then both assessments could be recommended next. In addition, this empirical data demonstration used assessments that have the same length. However, when learning materials consist of learning materials with different item numbers, a biased situation may be produced as short-length learning materials may be preferred if the sum score is still used. To better process situations with these problems, one possible solution is that the recommendation component design in the future study could assess the psychometric properties that each item has such as the item difficulties, and items with different difficulties can be assigned different weights when sum score was used.

Appendix

Data

Cumulative Assessments are aligned and assess a representation of the Georgia Standards of Excellence (GSE). These cumulative forms can help teachers to gather strong evidence on student learning toward the overall end-of-year expectations at each grade level. The sample assessment items were provided on this web site: https://www.lennections.com/assesslets-science

R Code

```
###########################################################
##########################read data#########################
###########################################################
data<-read.csv(file='data.csv', header=T, sep=",", fill=T,
      stringsAsFactors = F)
#processing
data2    <- udpipe(data, "english")
biterms <- as.data.table(data2)[, cooccurrence(x = lemma,
                                relevant = upos %in%
                                c("NOUN",
                                "ADJ", "VERB") &
                                nchar(lemma) > 2 & !lemma
                                %in%
                                stopwords("en"),
                                skipgram = 3),
                                by = list(doc_id)]
data3 <- data2[, c("doc_id", "lemma")]


###########################################################
###################decide optimal numbers##################
###########################################################
cd_k<-seq(2,10)
#JSD
model=NULL
for (i in cd_k) {
    model[[i]]   <- BTM(data3, biterms = biterms,
    k = i,
    alpha = 1,
    beta = 1,
    window = 3,
    iter = 5000, background = F,
    trace = F,detailed = F)
}

# Compute Jensen-Shannon Divergence for each value in model
scores <- predict(model[[1]], newdata = data3)
colnames(scores)<-c("topic1","topic2","topic3","topic4")
JSD <- function(p, q) {
m <- 0.5 * (p + q)
divergence <-
           0.5 * (sum(p * log(p / m)) + sum(q * log(q / m)))
return(divergence)
}

n <- dim(scores)[1]
X <- matrix(rep(0, n*n), nrow=n, ncol=n)
indexes <- t(combn(1:nrow(scores), m=2))
for (r in 1:nrow(indexes)) {
i <- indexes[r, ][1]
j <- indexes[r, ][2]
```

```
p <- scores[i, ]
q <- scores[j, ]
X[i, j] <- JSD(p,q)
}



############################################################
###################Estimation and predict##################
############################################################
#read students' response data
student = read.csv(file='4_Cumulative_Assesslet.csv',
                   header=T, sep=",",
                   fill=T,stringsAsFactors = F)
#M4 is feature matrix of 4th assessment
M4 = read.csv(file='M4.csv', header=T, sep=",",
                   fill=T,stringsAsFactors = F)

X = as.matrix(M4)
y= as.matrix(student)
N= dim(y)[1]*dim(y)[2]
theta.init = matrix(rnorm(n=dim(X)[2]*dim(y)[1],
                    mean=0,sd = 1),
nrow=dim(y)[1],ncol=dim(X)[2], byrow=T)
e = y - theta.init%*%t(X)
grad.init = -(2/N)*(e)%*%X
theta = theta.init - eta*(1/N)*grad.init
l2loss = c()
for(i in 1:iters){
myMatrix = y - theta%*%t(X)
# empty matrix for the results
squaredMatrix = matrix(nrow=dim(myMatrix)[1],
                                ncol=dim(myMatrix)[2])
for(i in 1:nrow(myMatrix)) {
for(j in 1:ncol(myMatrix)) {
squaredMatrix[i,j] = myMatrix[i,j]^2
}
}
l2loss = c(l2loss,sqrt(sum(squaredMatrix)))
e = y - theta%*%t(X)
grad = -(2/N)*e%*%X
theta = theta - eta*(2/N)*grad
# empty matrix for the results
squaredMatrix2 = matrix(nrow=dim(grad)[1],
                                    ncol=dim(grad)[2])
for(i in 1:nrow(grad)) {
for(j in 1:ncol(grad)) {
squaredMatrix2[i,j] = grad[i,j]^2
}
}
if(sqrt(sum(squaredMatrix2)) <= epsilon){
break
}
}
```

```
values<-list("coef" = theta, "l2loss" = l2loss)


h=sigmoid(X%*%t(theta.init))
sum(diag(-y%*%log(h)-(1-y)%*%log(1-h)))/m
#sigmoid function, inverse of logit
sigmoid <- function(z){1/(1+exp(-z))}

#initialize theta
theta <- matrix(rnorm(n=dim(X)[2]*dim(y)[1],
                                      mean=0,sd = 1),
nrow=dim(y)[1],ncol=dim(X)[2], byrow=T)
#comput GD
compCost<-function(para){
m <- dim(y)[1]*dim(y)[2]
j=0
for (i in seq(1,492*4,by=4)) {
k=match(i,seq(1,492*4,by=4))
l1_1=sigmoid(colSums(para[i:(i+3)]*t(X)))
l1 <- log(l1_1)
l2 <- log(1-l1_1)
j=j+sum(y[k,]*l1+(1-y[k,])*l2)
}
J=-j/m
}
```

References

Amos, B., & Yarats, D. (2020). *The differentiable cross-entropy method*. Paper presented at the International Conference on Machine Learning.

Apaza, R. G., Cervantes, E. V., Quispe, L. C., & Luna, J. O. (2014). *Online courses recommendation based on LDA*. Paper presented at the SIMBig.

Beagley, J. E., & Capaldi, M. (2016). The effect of cumulative tests on the final exam. *Primus, 26*(9), 878–888.

Bian, L., & Xie, Y. (2010). *Research on the adaptive strategy of adaptive learning system*. Paper presented at the International Conference on Technologies for E-Learning and Digital Entertainment.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research, 3*, 993–1022.

Cao, Y., Li, W., & Zheng, D. (2019). A hybrid recommendation approach using LDA and probabilistic matrix factorization. *Cluster Computing, 22*(4), 8811–8821.

Castells, P., Fernandez, M., & Vallet, D. (2006). An adaptation of the vector-space model for ontology-based information retrieval. *IEEE Transactions on Knowledge Data Engineering, 19*(2), 261–272.

Chen, Y., Li, X., Liu, J., & Ying, Z. (2018). Recommendation system for adaptive learning. *Applied Psychological Measurement, 42*(1), 24–41.

Cheng, Y., & Bu, X. (2020). *Research on key technologies of personalized education resource recommendation system based on big data environment*. Paper presented at the Journal of Physics: Conference Series.

Chou, Y.-T., & Wang, W.-C. (2010). Checking dimensionality in item response models with principal component analysis on standardized residuals. *Educational and Psychological Measurement, 70*(5), 717–731.

Cleghorn, G. D. (1986). Policies of US medical schools on the use of the NBME Part I and Part II examinations. *Journal of Medical Education, 61*(12), 954–957.

De Boer, P.-T., Kroese, D. P., Mannor, S., & Rubinstein, R. (2005). A tutorial on the cross-entropy method. *Annals of Operations Research, 134*(1), 19–67.

den Boer, A. W., Verkoeijen, P. P., & Heijltjes, A. E. (2021). Comparing formative and summative cumulative assessment: Two field experiments in an applied university engineering course. *Psychology Learning & Teaching, 20*(1), 128–143.

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems, 49*(1), 5–22.

Engelhard, G. Jr. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Routledge.

Georgia Center for Assessment. (2018). *English language and Arts cumulative Assessment*.

Ghauth, K. I., & Abdullah, N. A. (2010). Learning materials recommendation using good learners' ratings and content-based filtering. *Educational Technology Research and Development, 58*(6), 711–727.

Imhof, C., Bergamin, P., & McGarrity, S. (2020). Implementation of adaptive learning systems: Current state and potential. In *Online teaching and learning in higher education* (pp. 93–115). Springer.

Koedinger, K. R., Brunskill, E., Baker, R. S., McLaughlin, E. A., & Stamper, J. (2013). New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine, 34*(3), 27–41.

Kuang, W., Luo, N., & Sun, Z. (2011). *Resource recommendation based on topic model for educational system*. Paper presented at the 2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference.

Liang, Q., & Hainan, N. C. (2019). *Adaptive learning model and implementation based on big data*. Paper presented at the 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD).

Lin, Q., He, S., & Deng, Y. (2021). Method of personalized educational resource recommendation based on LDA and learner's behavior. *The International Journal of Electrical Engineering & Education*, 0020720920983511.

Mavroudi, A., Giannakos, M., & Krogstie, J. (2018). Supporting adaptive learning pathways through the use of learning analytics: Developments, challenges and future opportunities. *Interactive Learning Environments, 26*(2), 206–220.

Mei, Q., Shen, X., & Zhai, C. (2007). *Automatic labeling of multinomial topic models*. Paper presented at the Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Pfennig, A. (2020). *Improving learning outcome for GSL (German as a Second Language) students in a blended learning cumulative assessment material science course*. Paper presented at the Int. Conf. on Education and E-Learning ICEEL 2020.

Romero, C., Ventura, S., Delgado, J. A., & De Bra, P. (2007). *Personalized links recommendation based on data mining in adaptive educational hypermedia systems*. Paper presented at the European conference on technology enhanced learning.

Ryan, G. J., & Nykamp, D. (2000). Use of cumulative examinations at US schools of pharmacy. *American Journal of Pharmaceutical Education, 64*(4), 409–412.

Tang, X., Chen, Y., Li, X., Liu, J., & Ying, Z. (2019). A reinforcement learning approach to personalized learning recommendation systems. *British Journal of Mathematical and Statistical Psychology, 72*(1), 108–135.

Tong, Z., & Zhang, H. (2016). *A text mining research based on LDA topic modelling*. Paper presented at the International Conference on Computer Science, Engineering and Information Technology.

United States Department of Education. (2017). *Reimagining the role of technology in education: 2017 National Education Technology Plan update*. Washington, DC Retrieved from https://tech.ed.gov/files/2017/01/NETP17.pdf

USMLE. (2014). Federation of State Medical Boards of the United States and the National Board of Medical Examiners. *USMLE Bulletin of Information*.

Wheeler, J. M., Cohen, A. S., Xiong, J., Lee, J., & Choi, H.-J. (2021). Sample size for latent Dirichlet allocation of constructed-response items. In *Quantitative Psychology* (pp. 263–273). Springer.

Xiong, J., Choi, H.-J., Kim, S., Kwak, M., & Cohen, A. S. (2019). Topic Modeling of Constructed-Response Answers on Social Study Assessments. In *The annual meeting of the psychometric Society* (pp. 263–274). Springer, Cham.

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). *A biterm topic model for short texts*. Paper presented at the Proceedings of the 22nd International Conference on World Wide Web.