A Latent Factor Model For Instructor Content Preference Analysis

Jack Z. Wang
Rice University
jzwang@rice.edu

Andrew S. Lan
Princeton University
andrew.lan@princeton.edu

Phillip J. Grimaldi
Rice University
phillip.grimaldi@rice.edu

Richard G. Baraniuk
Rice University
richb@rice.edu

ABSTRACT
Existing personalized learning systems (PLSs) have primarily focused on providing learning analytics using data from learners. In this paper, we extend the capability of current PLSs by incorporating data from instructors. We propose a latent factor model that analyzes instructors' preferences in explicitly *excluding* particular questions from learners' assignments in a particular subject domain. We formulate the problem of predicting instructors' question exclusion preferences as a matrix factorization problem, and incorporate expert-labeled Bloom's Taxonomy tags on each question as a factor in our statistical model to improve model interpretability. Experimental results on a real-world educational dataset demonstrate that the proposed model achieves superior prediction performance compared to several other baseline methods commonly used in recommender systems. Additionally, by explicitly incorporating Bloom's Taxonomy, the model provides meaningful interpretations that help understand why instructors exclude certain questions. Since instructor preference data contains their insights after years of teaching experience, our proposed model has the potential to further improve the question recommendations that PLSs make for learners.

Keywords
personalized learning, educational data mining, latent factor model, Bloom's Taxonomy

1. INTRODUCTION
Today's education system has largely remained a "one-size-fits-all" learning experience in which the instructor selects a single learning action for all learners, ignoring their diverse backgrounds, interests, and goals. Modern machine learning (ML) techniques have led to a great acceleration in the development of personalized learning systems (PLSs) that have the potential to revolutionize education by delivering a high-quality and affordable personalized learning experience at large scale.

Current PLSs generally perform learning analytics using only learner data, overlooking data that instructors generate. However, when instructors are present in educational settings such as traditional classrooms, they generate important data that reveals how they prefer to interact with learning resources. Augmenting current learning analytics approaches by modeling instructors' preferences clearly provide advantages, since their preferences reflect years of teaching experience and thus provide valuable insights on how to utilize learning resources effectively. As a result, PLSs can refine their learning resource recommendations for learners using both learner data and these valuable insights. Additionally, analysis of instructor preferences for learning resources can serve as a starting point of recommending learning resource to learners when learner data is scarce such as at the beginning of a semester.

In this work, we focus on a specific instance of instructors' content[1] preferences. We collect instructors' preferences to exclude questions from being given to learners in their class via OpenStax Tutor[13], a personalized learning and teaching platform. OpenStax Tutor has a functionality to automatically *select* homework assignment questions for learners from a question corpus. At the same time, it allows instructors to *exclude* questions they do not want OpenStax Tutor to assign to learners in their classes from the corpus. While this exclusion option allows more flexibility for instructors to control homework assignment questions that learners receive, manually selecting questions to exclude from a (potentially huge) corpus is a labor-intensive process. As a result, analyzing instructors' question exclusion behavior has immediate utility in automating the question exclusion process.

1.1 Contributions

With the objective of analyzing instructors' preferences on assigning questions to learners on the OpenStax Tutor platform, we develop a novel latent factor model that predicts instructors' question preferences in a particular subject domain given previous records of whether instructors choose to *exclude* certain questions from homework assignments. The latent factor modeling approach is primarily inspired by SPARFA [10] which is a successful latent factor model for learner and content analysis. But more importantly, this approach allows flexible incorporation of prior knowledge in the form of meta-data into the model. Consequently, the model that we develop in this work can be easily extended to include additional information in the form of latent factors to explain instructors' question exclusion preferences,

[1] From now on, we will use the phrase "learning resources" and the word "content" interchangeably.

as well as be used in other educational data mining tasks where auxiliary information is available. Additionally, our proposed model incorporates expert-labeled Bloom's Taxonomy tags for each question to explain instructors' question exclusion preferences, based on the conjecture that instructors have varying inclinations towards different Bloom's Taxonomy tags[2].

Experimental results on a real-world educational dataset show that, compared to standard methods used in recommender systems, our model achieves higher overall accuracy in predicting instructors' question preferences. Additionally, we demonstrate that our model is highly interpretable in that the Bloom's Taxonomy explains question preferences of individual instructors, and reveals question preference patterns among instructors. Our analysis of the instructors' question exclusion preferences enables PLSs to incorporate instructors' insights on questions and potentially improve the quality of their personalized question recommendations.

We emphasize that our proposed model is not limited to analyzing instructors' question exclusion preferences; it can be easily modified to analyze instructors' preferences on a broader range of learning resources. Therefore, our work serves as an initial investigation into extending the capability of existing PLSs with the analysis of instructor learning resource interaction data.

1.2 Related Work

We formulate the problem of predicting instructors' question preferences as a matrix factorization problem underlying a recommender system. Recommender systems often rely on collaborative filtering (CF); the two most successful family of CF approaches to date are neighborhood-based methods and latent factor methods [4]. Neighborhood-based methods predict preferences based on neighbors chosen by some similarity measure. Latent factor methods, in particular, can be readily applied to education applications, resulting in tensor factorization for student modeling [15] and probabilistic models such as SPARFA [10], a primary source of inspiration for this work. However, these approaches, in their original form, do not have mechanisms to incorporate meta-data on learners and questions. Therefore, the explanatory power of these methods is usually limited. Our proposed model, on the other hand, extends the original latent factor model to explicitly include the Bloom's Taxonomy tag of each question as meta-data, providing additional interpretability and, at the same time, improves prediction accuracy.

Works including [6] and [12] incorporate external factors such as movie genres to improve users' movie rating prediction in the Netflix challenge [2], but their methods do not directly apply to education scenarios.

The work in [14] broadly describes a Bayesian approach to model instructors. While our work pursues a similar objective, we propose a concrete model with evaluations on a real-world dataset instead of a high-level overview. [11] uses the k-means clustering algorithm to recommend learning resources for instructors based on similar teaching styles among instructors. In addition to studying question type preferences, we approach the problem with a latent factor model instead of k-means clustering, yielding results that are more interpretable.

The work in [16] compares several models in predicting learners' next-term grades using various features including instructors' job title, rank, and tenure status. Our work, on the contrary, uses data that contains instructors' direct interaction with learning resources rather than simple demographic information.

[2]Bloom's Taxonomy hierarchically describes questions in terms of one of the six cognitive processes, including remembering, understanding, applying, analyzing, evaluating, and creating, in increasing cognitive complexity [9]. It describes the cognitive processes by which learners encounter and work with knowledge [1].

2. LATENT FACTOR MODEL

Let N, Q, K denote the total number of instructors, the total number of questions, and the total number of distinct Bloom's Taxonomy tags, respectively. Let $\mathbf{Y}$ be the binary-valued matrix of dimension N by Q that represents instructors' preference for a particular course, where $Y_{ij} = 1$ indicates instructor i explicitly denotes preference to exclude question j, and $Y_{ij} = 0$ indicates no preference. Also let $\mathbf{a}_j$ be a vector of dimension K that represents the question–Bloom's Taxonomy tag association for question j, where a_{jk} denotes the kth component of $\mathbf{a}_j$. $a_{jk} = 1$ indicates an association of question j with Bloom's Taxonomy tag k, and $a_{jk} = 0$ indicates no association.

With the above setup, we model $\mathbf{Y}$ as Bernoulli random variables:

$$Y_{ij} \sim \mathrm{Ber}\big(\phi(\mathbf{p}_i^T \mathbf{a}_j + \mathbf{g}_i^T \mathbf{h}_j)\big), \tag{1}$$

Where the function $\phi(\cdot)$ is the sigmoid function:

$$\phi(x) = \frac{1}{1+e^{-x}}.$$

In the model, $\mathbf{p}_i \in \mathbb{R}^K$, $\mathbf{g}_i \in \mathbb{R}^M$, $\mathbf{h}_j \in \mathbb{R}^M$ are model parameters to be estimated, where M is the dimension of $\mathbf{g}_i$ and $\mathbf{h}_j$ (we select the value of M via cross validation). Intuitively, the latent factor $\mathbf{p}_i$ represents the instructor Bloom's Taxonomy tag preference vector that reveals instructors' different preferences on each Bloom's Taxonomy tag. The latent factors $\mathbf{g}_i$ and $\mathbf{h}_j$ model additional factors that also contribute to explaining the observed data matrix $\mathbf{Y}$.

To compare the significance of the factor $\mathbf{p}_i$ against the factors $\mathbf{g}_i$ and $\mathbf{h}_j$, we use two simplified variants of the full model in Equation 1, namely P Model that involves only the factor $\mathbf{p}_i$, and GH Model that involves only factors $\mathbf{g}_i$ and $\mathbf{h}_j$:

$$\text{P Model:} \quad Y_{ij} \sim \mathrm{Ber}\big(\phi(\mathbf{p}_i^T \mathbf{a}_j)\big) \tag{2}$$

$$\text{GH Model:} \quad Y_{ij} \sim \mathrm{Ber}\big(\phi(\mathbf{g}_i^T \mathbf{h}_j)\big) \tag{3}$$

2.1 Optimization Algorithm

We formulate the maximum-likelihood parameter estimation problem for the proposed model as an optimization problem. The optimization objective is given by

$$\underset{\mathbf{P},\mathbf{G},\mathbf{H}}{\text{minimize}}\, f(\mathbf{P},\mathbf{G},\mathbf{H}),$$

where $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_N]$ denotes the matrix of instructor Bloom's Taxonomy tag preference associations by stacking the association vectors together. $\mathbf{G}$ and $\mathbf{H}$ are defined analogously. The cost function $f(\mathbf{P}, \mathbf{G}, \mathbf{H})$ is given by

$$f(\mathbf{P}, \mathbf{G}, \mathbf{H}) = \sum_{i=1}^{N} \sum_{j=1}^{Q} \log\Big(1 + \exp\big(-(\mathbf{p}_i^T \mathbf{a}_j + \mathbf{g}_i^T \mathbf{h}_j)\big)\Big) + \frac{\lambda}{2} \sum_{i=1}^{N} \|\mathbf{p}_i\|_2^2 + \frac{\gamma}{2} \sum_{i=1}^{N} \|\mathbf{g}_i\|_2^2 + \frac{\eta}{2} \sum_{j=1}^{Q} \|\mathbf{h}_j\|_2^2.$$

The last three terms in the cost function are regularization terms added to prevent overfitting. λ, γ, and η are regularization parameters for the factors $\mathbf{p}_i, \mathbf{g}_i, \mathbf{h}_j$, respectively.

The above optimization problem is non-convex, but the subproblems to optimize over each parameter while holding the others fixed are convex. We therefore employ block coordinate descent to efficiently find a local minima for the above optimization problem by iteratively updating each parameter in turn. The update equations for the parameters are given by

$$\begin{aligned}
\mathbf{p}_i^{\text{new}} &= \mathbf{p}_i^{\text{old}} - \delta \frac{\partial}{\partial \mathbf{p}_i} f(\mathbf{p}_i^{\text{old}}, \mathbf{g}_i^{\text{old}}, \mathbf{h}_j^{\text{old}}) \\
\mathbf{g}_i^{\text{new}} &= \mathbf{g}_i^{\text{old}} - \delta \frac{\partial}{\partial \mathbf{g}_i} f(\mathbf{p}_i^{\text{new}}, \mathbf{g}_i^{\text{old}}, \mathbf{h}_j^{\text{old}}) \\
\mathbf{h}_j^{\text{new}} &= \mathbf{h}_j^{\text{old}} - \delta \frac{\partial}{\partial \mathbf{h}_j} f(\mathbf{p}_i^{\text{new}}, \mathbf{g}_i^{\text{new}}, \mathbf{h}_j^{\text{old}}),
\end{aligned}$$

where δ is the step size. The gradients of the cost function with respect to each parameter are given by

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{p}_i} f(\mathbf{p}_i, \mathbf{g}_i, \mathbf{h}_j) &= -\sum_{j=1}^{Q} \frac{\mathbf{a}_j}{1 + e^{-(\mathbf{p}_i^T \mathbf{a}_j + \mathbf{g}_i^T \mathbf{h}_j)}} + \lambda \mathbf{p}_i \\
\frac{\partial}{\partial \mathbf{g}_i} f(\mathbf{p}_i, \mathbf{g}_i, \mathbf{h}_j) &= -\sum_{j=1}^{Q} \frac{\mathbf{h}_j}{1 + e^{-(\mathbf{p}_i^T \mathbf{a}_j + \mathbf{g}_i^T \mathbf{h}_j)}} + \gamma \mathbf{g}_i \\
\frac{\partial}{\partial \mathbf{h}_j} f(\mathbf{p}_i, \mathbf{g}_i, \mathbf{h}_j) &= -\sum_{i=1}^{N} \frac{\mathbf{g}_i}{1 + e^{-(\mathbf{p}_i^T \mathbf{a}_j + \mathbf{g}_i^T \mathbf{h}_j)}} + \eta \mathbf{h}_j.
\end{aligned}$$

At the beginning of optimization, we randomly initialize the model parameters $\mathbf{p}_i$, $\mathbf{g}_i$, $\mathbf{h}_j$ for all i, j. In each optimization iteration, we first loop over all i's to update all $\mathbf{p}_i$ and $\mathbf{g}_i$ while holding all $\mathbf{h}_j$'s fixed, and then loop over all j's to update $\mathbf{h}_j$ using the newly calculated $\mathbf{p}_i$'s and $\mathbf{g}_i$'s. We repeat the above iterations until convergence, i.e., the difference of the cost function between two iterations falls below a predefined threshold.

Note that the inference problem for the P Model in Equation 2 is convex, and optimization is straightforward via gradient descent. Since the GH Model in Equation 3 involves two sets of parameters and has a non-convex inference problem, we employ the same block coordinate descent method as in the full model.

2.2 Model Extensions

We now enumerate possible extensions to the proposed model. First, we can incorporate additional prior information as latent factors in the model by simply including other modalities of meta-data as an additional inner product terms of two more latent factors inside the $\phi(\cdot)$ function. In this way, in each inner product term, one factor denotes the newly included meta-data modality, and the other characterizes the instructor's exclusion preference in terms of that specific modality of meta-data. Concretely, the extension of the model in Equation 1 has the following form:

$$Y_{ij} \sim \text{Ber}\Bigg(\phi\bigg(\sum_{l=1}^{L} {\mathbf{u}_i^l}^T \mathbf{v}_j^l + \mathbf{g}_i^T \mathbf{h}_j\bigg)\Bigg), \tag{4}$$

where we have replaced the inner product term $\mathbf{p}_i^T \mathbf{a}_j$ in Equation 1 with a sum of L inner product terms. Each $\mathbf{u}_i^l$ and $\mathbf{v}_j^l$ model instructor and question association of a particular modality of meta-data. Additionally, the dimensions of $\mathbf{u}_i^l$ and $\mathbf{v}_j^l$ can vary for different l's depending on the mathematical representation of that meta-data modality.

Next, it is easy to see that the same approach can be applied to analyzing instructors' preferences on other learning resources. Although we specify in Equation 1 that Y_{ij} represents instructor i's preference for question j, Y_{ij} can naturally represent preferences to other contents types, by using j to index learning resources. Therefore, we can easily extend the proposed model in Equation 1 to analyze additional instructor preference data with a different preference data matrix $\mathbf{Y}$.

	Metrics	
Models	ACC	AUC
Proposed Model	**0.9033±0.0045**	**0.9592±0.0061**
P Model	0.8880±0.0047	0.8908±0.0064
GH Model	0.9026±0.0048	0.9254±0.0058

Table 1: Performance comparison between the proposed model and its variants, in terms of prediction accuracy (ACC) and area under operating characteristic curve (AUC). The proposed model achieves the best result among its two variants. The model involving the $\mathbf{g}_i$ and $\mathbf{h}_j$ factors achieves better performance than the model with the $\mathbf{p}_i$ factor alone.

3. EXPERIMENTS

We now evaluate the prediction performance of the proposed latent factor model using a real-world educational dataset. We further showcase the interpretability of the model by visualizing the instructor Bloom's Taxonomy tag preference vectors $\mathbf{p}_i$.

3.1 Dataset

We collect from OpenStax Tutor [13] 20 instructors' preferences on all 896 questions of the textbook "Concepts of Biology" that these instructors use in their classes, resulting in a fully observed data matrix $\mathbf{Y}$ of dimension 20 by 896. About 15% of all entries in $\mathbf{Y}$ have a value of 1, meaning that an instructor explicitly indicates to exclude a question, and the rest 0, meaning that there is no such indication. We remind the reader that excluding a question in OpenStax Tutor means that this question is excluded from the pool of questions that OpenStax Tutor selects from to assign to learners as personalized practice recommendations. We also collect the Bloom's Taxonomy tag for each question, labeled by domain experts, as meta-data on the questions. Since there are 6 distinct Bloom's Taxonomy tags

	Models/Methods			
Metric	**Full Model**	UBCF	IBCF	FSVD
ACC	**0.9033±0.0045**	0.8961±0.0048	0.8895±0.0048	0.8896±0.0045
F-1	**0.6483±0.0128**	0.6007±0.0158	0.5696±0.0137	0.6185±0.0158
Precision	**0.7163±0.0222**	0.7070±0.0214	0.6928±0.0254	0.6964±0.0236
Recall	**0.6153±0.0227**	0.5226±0.0190	0.4954±0.0159	0.5661±0.0248

Table 2: Performance comparison between the proposed model and existing collaborative filtering methods in terms of the four metrics. The proposed model shows superior prediction performance compared to the other methods on all metrics.

	Bloom's Taxonomy tag					
Instructor	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$
$i = 3$	0.9% 0.058	1.6% 0.083	0.5% 0.038	1.8% 0.216	0.0% 0.075	0.0% 0.084
$i = 5$	16.9% 0.441	16.3% 0.448	19.0% 0.501	5.5% 0.360	21.1% 1.000	33.3% 0.858
$i = 9$	63.1% 0.826	67.8% 1.000	72.4% 0.985	67.3% 0.924	42.1% 0.583	33.3% 0.215

Table 3: Comparison between $\mathbf{p}_{ik}$ (second row for each instructor) and the percentage of questions they actually excluded under each Bloom's taxonomy tag k (first row for each instructor), for selected instructors. The values of $\mathbf{p}_i$ estimated by the proposed model closely resemble the actual number of questions each instructor excluded.

in total, the dimension of the question–Bloom's Taxonomy tag association vector $\mathbf{a}_j$ is $K = 6$. The entries of $\mathbf{a}_j$ correspond to Bloom's Taxonomy tags in increasing levels of cognitive complexity, i.e., $k = 1$ represents "remembering", $k = 2$ represents "understanding", etc. Additionally, each question is only associated with one Bloom's Taxonomy in our dataset. Therefore, the values of $\mathbf{a}_j$ satisfy $a_{jk} \in \{0, 1\}$ and $\sum_k a_{jk} = 1$ for all j.

3.2 Experimental Setup

We compare our model and its variants against three methods frequently used in recommender systems: user-based collaborative filtering (UBCF), item-based collaborative filtering (IBCF), and funk singular value decomposition (FSVD). UBCF and IBCF use similarities among users (instructors) and items (questions), respectively, and predict a user's preference on an item based on the preferences of most similar users or items. FSVD makes the observation that the actual number of user and item types is much lower than the number of users and items, and therefore utilizes a low-rank model to model user–item interactions [4, 5]. [7] explain the detailed implementations and evaluation methods for UCBF, ICBF, and FSVD that we use in this paper.

We use a total of five metrics for model evaluation: (i) prediction accuracy (ACC), (ii) precision, (iii) recall, (iv) F-1 score, and (v) area under the receiver operating characteristic curve (AUC) of the resulting binary classifier [8]. Formulas for calculating metrics (i) through (iv) are shown below:

$$\begin{cases} \text{ACC} & = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{FP}+\text{TN}+\text{FN}} \\ \text{precision} & = \frac{\text{TP}}{\text{TP}+\text{FP}} \\ \text{recall} & = \frac{\text{TP}}{\text{TP}+\text{FN}} \\ \text{F-1} & = 2 \times \frac{\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}, \end{cases}$$

where TP denotes true positive, TN denotes true negative, FP denotes false positive, and FP denotes false negative. In the context of this paper, we treat preference for excluding a question, corresponding to $Y_{ij} = 1$, as the positive class. True positive means predicting the positive class when the ground truth is also positive. False positive means predicting the positive class when ground truth is negative, and the rest follows. All metrics take on values in $[0, 1]$, with larger values indicating better prediction performance. We perform two sets of comparisons, one between the full model and its two variants (the P and GH models) evaluated on the ACC and AUC metrics, and the other one between the full model and UCBF, IBCF, and FSVD using ACC, F-1, precision, and recall. Since the AUC metric is only appropriate for evaluating algorithms using probabilistic models, we do not evaluate the three CF methods that do not have an underlying probabilistic model.

We perform 5-fold cross validation for model selection, i.e. choosing the best set of parameters for each model, and model assessment, i.e. evaluating the best model on the test set, according to the train-validation-test split paradigm. First, we randomly select 20% of all observed data and set it aside as test set. We then randomly partition the remaining 80% of all data into four roughly equal-sized parts, fit the

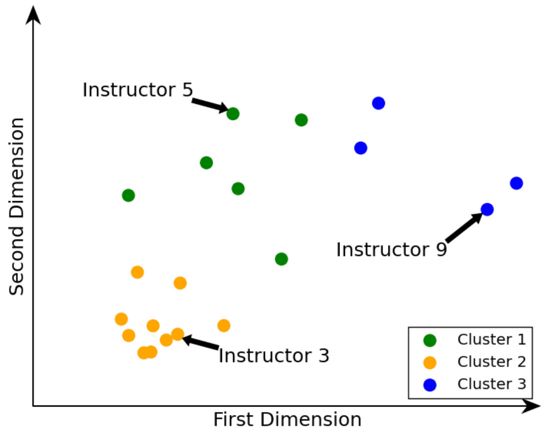

Figure 1: 2D projection of instructor Bloom's Taxonomy tag preference vectors using multidimensional scaling and clustering using k-means that shows instructors' diverse question exclusion preferences. Notice that instructors 3, 5, and 9 that we show to have very different question exclusion preferences also appear far apart in the plot.

model to first three of the four parts, and validate the fitted model using the fourth part of the data to select the values of the regularization parameters using grid-search. Finally, we select the best performing model, fit it on all data except for the test set, and evaluate its performance on the test set. We perform 20 random partitions of the data, average the evaluation results, and compare the best evaluation results of each method.

3.3 Results And Discussions

Table 2 shows results for the full model, UBCF, IBCF, and FSVD evaluated on the ACC, F-1, Precision, and Recall metrics. The relatively lower Recall scores of the full model compared to its ACC suggests that the proposed model still exhibits some albeit less tendency to avoid assigning an exclusion preference label than other methods. Nevertheless, comparing across columns, we see that the performance of the full model, regardless of the choice of metric, is significantly better than the rest of the models, showing promise for the proposed latent factor model in predicting instructors' question exclusion preferences.

Table 1 shows prediction performance results for the full model and its two variants evaluated on the ACC and AUC metrics. From the table, we observe that the full model achieves the best performance on both metrics. Further inspection of the results of the two variants reveals that the GH Model, which involves factors $\mathbf{g}_i$ and $\mathbf{h}_j$, achieves better results for both metrics than the P Model, which involves only factor $\mathbf{p}_i$. This implies that besides Bloom's Taxonomy, additional factors are needed in the latent factor model to better characterize instructors' question exclusion preferences. Even though Bloom's Taxonomy contribute only moderately to the prediction performance, the purpose of explicitly incorporating Bloom's Taxonomy, as stated earlier, is the power of interpretability it brings to the proposed model, which we demonstrate below.

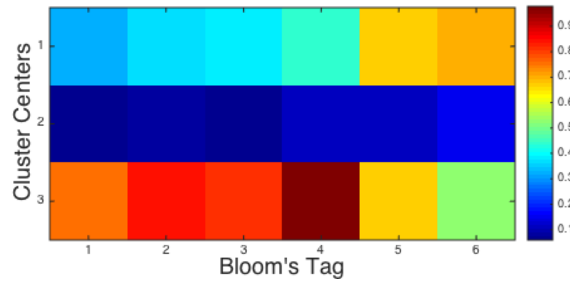

Figure 2: Heatmap visualization of the cluster centers that shows the radically different question exclusion preferences of each cluster of instructors.

First, we use the instructor Bloom's Taxonomy tag association vectors to interpret how instructors prefer to exclude certain questions in terms of Bloom's Taxonomy. Table 3 presents a comparison between the numerical values of entries in the instructor Bloom's Taxonomy tag preference vector $\mathbf{p}_i$ and the percentage of questions that the corresponding instructor excludes with each Bloom's Taxonomy tag, for a selected subset of instructors $i \in \{3, 5, 9\}$. Comparing the values in the two rows for each instructor i in the table, we observe that higher values of $\mathbf{p}_{ik}$ correspond to a higher percentage of the questions of Bloom's Taxonomy tag k that the instructor excludes. Therefore, $\mathbf{p}_{ik}$ reflects the degree to which instructor i prefers to exclude questions with Bloom's Taxonomy tag k. For example, we observe from the second row of instructor 5 that values of $\mathbf{p}_{ik}$ are high for $k = 5$ and $k = 6$, indicating that this instructor strongly prefers to exclude questions that involve more complex cognitive processes such as evaluating and creating. Second, the instructor Bloom's Taxonomy tag preference vectors uncover differences and patterns in instructors' Bloom's Taxonomy tag preferences. Comparing the second row of all instructors in Table 3, we see distinct preferences for different instructors. For example, values of $\mathbf{p}_{ik}$ for instructor 9 are high for $k = 1, 2, 3, 4$, indicating that this instructor strongly prefers to not assign questions that involve simpler cognitive processes such as remembering, understanding, applying and analyzing. Such preferences are opposite to those for instructor 5. Moreover, instructor 3 exhibits no obvious exclusion preference for any Bloom's Taxonomy tags by noting the small values of $\mathbf{p}_{ik}$ for $i = 3$, setting this instructor apart from both instructors 5 and 9.

We further visualize patterns in instructors' question preferences after projecting each $\mathbf{p}_i$ onto a 2-dimensional plane using multidimensional scaling [3]. We then run the K-means algorithm to group the instructors into 3 clusters. Figure 1 plots each $\mathbf{p}_i$ as a point in the 2-dimensional space, where the color of the point denotes the cluster that the point belongs to. The figure shows obvious clustering patterns, which means that instructors exhibit only a few patterns on their Bloom's Taxonomy tag preferences. Note that instructors 3, 5 and 9 are far apart in the figure and belong to different clusters. Figure 2 presents a heatmap visualization of the cluster centers that shows distinct Bloom's Taxonomy preferences across the three instructor clusters. For example, the first and third clusters demonstrate almost entirely opposite Bloom's Taxonomy preferences, where the first cluster tends to exclude questions with more complex cognitive process, whereas the third cluster tends to exclude questions with simpler cognitive processes. On the other hand, the second cluster does not exhibit strong exclusion preferences for any particular Bloom's Taxonomy tag. Such clustering could help a PLS to recommend questions to an instructor that they might want to exclude, based on instructors that have demonstrated similar Bloom's Taxonomy preferences.

4. CONCLUSIONS AND FUTURE WORK

We have presented a latent factor model that predicts instructors' question preferences, and explicitly incorporates questions' Bloom's Taxonomy tags to improve model interpretability. Evaluated on a real-world educational dataset, our proposed model shows superior prediction performance over popular collaborative filtering methods frequently used in recommender systems. Additionally, we demonstrated model interpretability by showing that the Bloom's Taxonomy captures each instructor's question preferences reasonably well, and also visualized different Bloom's Taxonomy preference patterns across instructors. These encouraging results show the promise of using latent factor approach for instructors' content preferences modeling to 1) potentially automate the question exclusion process in OpenStax Tutor, and 2) more broadly, to improve various aspects of personalized learning systems such as intelligent content recommendation that takes into account of instructors' preferences.

To achieve these goals, the following avenues of future research seem appropriate. First, we used only one source of meta-data, i.e., Bloom's Taxonomy tags, in the proposed model. We have shown that the proposed model is easily extendable to accommodate additional meta-data; moreover, the performance comparison between the P Model and the GH Model shows the need to incorporate additional factors. Therefore, we plan to extend the proposed model to include other sources of meta-data, such as the textbook chapter or section that each question belongs to, to improve both prediction accuracy and model interpretability. Second, we focused on instructors' preferences in a very specific content, i.e., question exclusion. We are interested to see how well the proposed modeling approach can be adapted to analyze instructors' preference for other learning resources. Third, we also plan to expand our experiments from a single textbook to multiple textbooks and domains, in order to validate the proposed approach for analyzing instructor preferences on a wide range of contents and across different subject domains.

5. REFERENCES

[1] P. Armstrong. *Bloom's Taxonomy*, 2014. `https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/`.

[2] R. M. Bell and Y. Koren. Lessons from the Netflix prize challenge. *ACM SIGKDD Explorations Newsletter*, 9(2):75–79, Dec. 2007.

[3] I. Borg and P. J. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2005.

[4] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan. Collaborative filtering recommender systems. *Foundations and Trends in Human–Computer Interaction*, 4(2):81–173, Feb. 2011.

[5] S. Funk. *Netflix update: Try this at home*, 2006. `http://sifter.org/~simon/journal/20061211.html`.

[6] Z. Gantner, L. Drumond, C. Freudenthaler, S. Rendle, and L. Schmidt-Thieme. Learning attribute-to-feature mappings for cold-start recommendations. In *Proceedings of the 10th IEEE International Conference on Data Mining*, pages 176–185, Dec. 2010.

[7] M. Hahsler. *recommenderlab: Lab for Developing and Testing Recommender Algorithms*, 2016. R package version 0.2-1.

[8] J. Huang and C. X. Ling. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering*, 17(3):299–310, Jan. 2005.

[9] D. R. Krathwohl. A revision of Bloom's taxonomy: An overview. *Theory into practice*, 41(4):212–218, Nov. 2002.

[10] A. S. Lan, A. E. Waters, C. Studer, and R. G. Baraniuk. Sparse factor analysis for learning and content analytics. *Journal of Machine Learning Research*, 15(1):1959–2008, Jan. 2014.

[11] C. Limongelli, M. Lombardi, A. Marani, and F. Sciarrone. A teaching-style based social network for didactic building and sharing. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education*, pages 774–777, Jul. 2013.

[12] M. G. Manzato. Discovering latent factors from movies genres for enhanced recommendation. In *Proceedings of the 6th ACM Conference on Recommender Systems*, pages 249–252, Sept. 2012.

[13] OpenStax Tutor. https://tutor.openstax.org/, 2016.

[14] Z. A. Pardos and N. T. Heffernan. Tutor modeling vs. student modeling. In *Proceedings of International Florida Artificial Intelligence Research Society Conference*, pages 420–425, May 2012.

[15] S. Sahebi, Y.-R. Lin, and P. Brusilovsky. Tensor factorization for student modeling and performance prediction in unstructured domain. In *Proceedings of the 9th International Conference on Educational Data Mining*, pages 502–506, June 2016.

[16] M. Sweeney, J. Lester, H. Rangwala, and A. Johri. Next-term student performance prediction: A recommender systems approach. *Journal of Educational Data Mining*, 26(1):33–68, Mar. 2016.